



등록특허 10-2212257



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년02월04일
(11) 등록번호 10-2212257
(24) 등록일자 2021년01월29일

- (51) 국제특허분류(Int. Cl.)
C12N 15/10 (2017.01) *C04B 40/06* (2006.01)
C12N 15/11 (2006.01) *C12Q 1/6806* (2018.01)
C12Q 1/6816 (2018.01)
- (52) CPC특허분류
C12N 15/1068 (2013.01)
C04B 40/06 (2013.01)
- (21) 출원번호 10-2019-7008458
- (22) 출원일자(국제) 2017년04월05일
심사청구일자 2020년04월03일
- (85) 번역문제출일자 2019년03월22일
- (65) 공개번호 10-2019-0041000
- (43) 공개일자 2019년04월19일
- (86) 국제출원번호 PCT/US2017/026232
- (87) 국제공개번호 WO 2018/038772
국제공개일자 2018년03월01일
- (30) 우선권주장
62/378,134 2016년08월22일 미국(US)
(뒷면에 계속)
- (56) 선행기술조사문헌
WO2015021080 A2*
WO2016011080 A2*
- *는 심사관에 의하여 인용된 문헌
- (73) 특허권자
트위스트 바이오사이언스 코포레이션
미국, 캘리포니아 94158, 샌프란시스코, 스위트 545, 사우스, 미션 베이 불러바드 455
- (72) 발명자
콕스 앤서니
미국 94040 캘리포니아주 마운틴 뷰 플라자 코트 1768
챈 쓰위엔
미국 94402 캘리포니아주 산 마테오 애버딘 드라이브 1302
- (74) 대리인
김진희, 김태홍

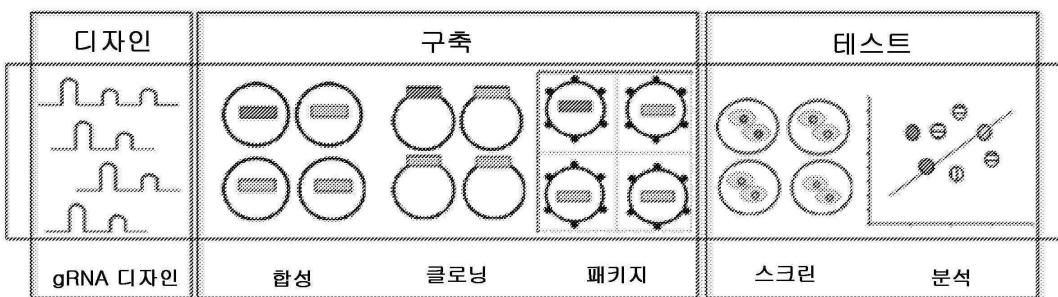
전체 청구항 수 : 총 26 항

심사관 : 한지혜

(54) 발명의 명칭 드 노보 합성된 핵산 라이브러리

(57) 요약

본원에서는 gRNA 서열을 코딩하는 올리고핵산 라이브러리 생성 방법을 개시한다. 본원에 기술된 방법에 의해 코딩된 gRNA는 단일 또는 이중 gRNA 서열일 수 있다. 기술된 방법은 정확성 및 균일성이 개선된, DNA 전구체로서의 또는 RNA 전사 생성물로서의 gRNA 라이브러리의 생성을 제공한다.

대 표 도

(52) CPC특허분류

C12N 15/1093 (2013.01)

C12N 15/111 (2013.01)

C12Q 1/6806 (2018.05)

C12Q 1/6816 (2018.05)

C12N 2310/10 (2013.01)

C12N 2310/20 (2017.05)

C12N 2330/31 (2013.01)

(30) 우선권주장

62/382,191 2016년08월31일 미국(US)

62/393,948 2016년09월13일 미국(US)

62/411,388 2016년10월21일 미국(US)

62/419,881 2016년11월09일 미국(US)

명세서

청구범위

청구항 1

핵산 라이브러리로서, 적어도 500개의 동일하지 않은 DNA 분자를 포함하고, 여기서, 각각의 동일하지 않은 DNA 분자는 상이한 gRNA 서열을 코딩하고, 각각의 gRNA 서열은 포유동물 유전자에 상보적인 표적화 도메인을 포함하며, 적어도 500개의 동일하지 않은 DNA 분자의 적어도 80%는 각각, 핵산 라이브러리 중 각각의 동일하지 않은 DNA 분자에 대한 평균 빈도의 2x 이내의 양으로 상기 핵산 라이브러리 중에 존재하는 것인 핵산 라이브러리.

청구항 2

제1항에 있어서, 각각의 동일하지 않은 DNA 분자의 GC 염기 함량이 20% 내지 85%인 핵산 라이브러리.

청구항 3

제1항에 있어서, 각각의 동일하지 않은 DNA 분자의 GC 염기 함량이 30% 내지 70%인 핵산 라이브러리.

청구항 4

제1항에 있어서, 적어도 500개의 동일하지 않은 DNA 분자의 적어도 90%가 각각, 핵산 라이브러리 중 각각의 동일하지 않은 DNA 분자에 대한 평균 빈도의 2x 이내의 양으로 상기 핵산 라이브러리 중에 존재하는 것인 핵산 라이브러리.

청구항 5

제1항에 있어서, 적어도 500개의 동일하지 않은 DNA 분자의 적어도 99%가 각각, 핵산 라이브러리 중 각각의 동일하지 않은 DNA 분자에 대한 평균 빈도의 2x 이내의 양으로 상기 핵산 라이브러리 중에 존재하는 것인 핵산 라이브러리.

청구항 6

제1항에 있어서, 적어도 500개의 동일하지 않은 DNA 분자가 적어도 2,000개의 동일하지 않은 DNA 분자를 포함하는 것인 핵산 라이브러리.

청구항 7

제1항에 있어서, 적어도 500개의 동일하지 않은 DNA 분자가 적어도 3,500개의 동일하지 않은 DNA 분자를 포함하는 것인 핵산 라이브러리.

청구항 8

제1항에 있어서, 적어도 500개의 동일하지 않은 DNA 분자가 적어도 100,000개의 동일하지 않은 DNA 분자를 포함하는 것인 핵산 라이브러리.

청구항 9

제1항에 있어서, 각각의 동일하지 않은 DNA 분자가 최대 200개의 염기 길이를 포함하는 것인 핵산 라이브러리.

청구항 10

제1항에 있어서, 각각의 동일하지 않은 DNA 분자가 100개 내지 200개의 염기 길이를 포함하는 것인 핵산 라이브러리.

청구항 11

제1항에 있어서, 적어도 500개의 동일하지 않은 DNA 분자가 생물학적 경로 중의 유전자를 표적화하는 gRNA 서열을 코딩하는 동일하지 않은 DNA 분자를 포함하는 것인 핵산 라이브러리.

청구항 12

제1항에 있어서, 적어도 500개의 동일하지 않은 DNA 분자가 전체 게놈 중의 유전자를 표적화하는 gRNA 서열을 코딩하는 동일하지 않은 DNA 분자를 포함하는 것인 핵산 라이브러리.

청구항 13

제1항에 있어서, gRNA가 단일 gRNA 또는 이중 gRNA인 핵산 라이브러리.

청구항 14

핵산 라이브러리로서, 적어도 2,000개의 동일하지 않은 핵산을 포함하고, 여기서, 각각의 동일하지 않은 핵산은 상이한 gRNA 서열을 코딩하고, 각각의 gRNA 서열은 진핵생물 유전자에 상보적인 표적화 도메인을 포함하며, 적어도 2,000개의 동일하지 않은 핵산의 적어도 80%는 핵산 라이브러리 중 각각의 동일하지 않은 핵산에 대한 평균 빈도의 2x 이내의 양으로 상기 핵산 라이브러리 중에 존재하는 것인 핵산 라이브러리.

청구항 15

제14항에 있어서, 각각의 동일하지 않은 핵산의 GC 염기 함량이 20% 내지 85%인 핵산 라이브러리.

청구항 16

제14항에 있어서, 각각의 동일하지 않은 핵산의 GC 염기 함량이 30% 내지 70%인 핵산 라이브러리.

청구항 17

제14항에 있어서, 적어도 2,000개의 동일하지 않은 핵산의 적어도 90%가 각각, 핵산 라이브러리 중 각각의 동일하지 않은 핵산에 대한 평균 빈도의 2x 이내의 양으로 상기 핵산 라이브러리 중에 존재하는 것인 핵산 라이브러리.

청구항 18

제14항에 있어서, 적어도 2,000개의 동일하지 않은 핵산의 적어도 99%가 각각, 핵산 라이브러리 중 각각의 동일하지 않은 핵산에 대한 평균 빈도의 2x 이내의 양으로 상기 핵산 라이브러리 중에 존재하는 것인 핵산 라이브러리.

청구항 19

제14항에 있어서, 각각의 동일하지 않은 핵산이 최대 200개의 염기 길이를 포함하는 것인 핵산 라이브러리.

청구항 20

제14항에 있어서, 각각의 동일하지 않은 핵산이 100개 내지 200개의 염기 길이를 포함하는 것인 핵산 라이브러리.

청구항 21

제14항에 있어서, 적어도 2,000개의 동일하지 않은 핵산이 생물학적 경로 중의 유전자를 표적화하는 gRNA 서열을 코딩하는 동일하지 않은 핵산을 포함하는 것인 핵산 라이브러리.

청구항 22

제14항에 있어서, 적어도 2,000개의 동일하지 않은 핵산이 전체 게놈 중의 유전자를 표적화하는 gRNA 서열을 코딩하는 동일하지 않은 핵산을 포함하는 것인 핵산 라이브러리.

청구항 23

제14항에 있어서, 각각의 동일하지 않은 핵산이 DNA 또는 RNA 분자를 포함하는 것인 핵산 라이브러리.

청구항 24

앰플리콘 라이브러리로서, 복수의 동일하지 않은 DNA 분자를 포함하고, 여기서, 각각의 동일하지 않은 DNA 분자는 증폭 생성물의 집단 중에 존재하고, 각각의 동일하지 않은 DNA 분자는 상이한 gRNA 서열을 코딩하며, 각각의 gRNA 서열은 진핵생물 유전자에 상보적인 표적화 도메인을 포함하고, 복수의 동일하지 않은 DNA 분자의 적어도 80%는 각각, 앰플리콘 라이브러리 중 각각의 동일하지 않은 DNA 분자에 대한 평균 빈도의 2x 이내의 양으로 상기 앰플리콘 라이브러리 중에 존재하는 것인 앰플리콘 라이브러리.

청구항 25

제24항에 있어서, 각각의 동일하지 않은 DNA 분자의 GC 염기 함량이 30% 내지 70%인 앰플리콘 라이브러리.

청구항 26

제24항에 있어서, gRNA가 단일 gRNA 또는 이중 gRNA인 앰플리콘 라이브러리.

청구항 27

삭제

청구항 28

삭제

청구항 29

삭제

청구항 30

삭제

청구항 31

삭제

청구항 32

삭제

청구항 33

삭제

청구항 34

삭제

청구항 35

삭제

청구항 36

삭제

청구항 37

삭제

청구항 38

삭제

청구항 39

삭제

청구항 40

삭제

청구항 41

삭제

청구항 42

삭제

청구항 43

삭제

청구항 44

삭제

청구항 45

삭제

청구항 46

삭제

청구항 47

삭제

청구항 48

삭제

청구항 49

삭제

청구항 50

삭제

청구항 51

삭제

청구항 52

삭제

청구항 53

삭제

청구항 54

삭제

청구항 55

삭제

청구항 56

삭제

청구항 57

삭제

청구항 58

삭제

청구항 59

삭제

청구항 60

삭제

청구항 61

삭제

청구항 62

삭제

발명의 설명

기술 분야

[0001] 상호 참조

[0002] 본 출원은 2016년 11월 9일 출원된 미국 가출원 번호 62/419,881; 2016년 10월 21일 출원된 미국 가출원 번호 62/411,388; 2016년 9월 13일 출원된 미국 가출원 번호 62/393,948; 2016년 8월 31일 출원된 미국 가출원 번호 62/382,191; 및 2016년 8월 22일 출원된 미국 가출원 번호 62/378,134의 이익을 주장하며, 상기 출원들은 각각 그 전문이 본원에서 참조로 포함된다.

[0003] 서열 목록

[0004] 본 출원은 ASCII 포맷으로 전자 제출된 서열 목록을 포함하고, 이 서열 목록은 그 전체가 본원에 참조로 포함된다. 2017년 4월 4일에 작성된 상기 ASCII 사본의 명칭은 44854-727_601_SL.txt이고, 그 크기는 13,567 바이트이다.

배경기술

[0005] 배경

[0006] 합성 생물학의 기초는, 맞춤식 경로 및 유기체의 신속하고 적절한 생성 및 최적화를 위해 접근가능화되어야 하는, DNA를 필요로 하는 반복 프로세스인 디자인, 구축, 및 테스트 프로세스이다. 디자인 단계에서, DNA를 구성하는 A, C, T 및 G 뉴클레오티드는 관심 영역을 포함하게 되는 다양한 서열로 제제화되는데, 여기서, 각 서열의 변이체가 테스트하고자 하는 구체적인 가설을 나타낸다. 이를 변이체 유전자 서열은, 진화 생물학에서 기원이 되고, 유전자, 게놈, 트랜스크립트 및 프로테옴을 구성하는 서열 전체에 관한 개념인 서열 공간의 서브세트를 나타낸다. 표적화된 게놈 편집과 관련하여, 유전자, 유전자 클러스터, 경로, 또는 전체 게놈의 효소적 편집을 특이적으로 지시하기 위한 고도로 정확하고, 균일한 핵산 라이브러리의 신속한 생성이 요구되고 있다.

[0007] 참고 문헌 인용

[0008] 본 명세서에서 언급된 모든 공개문헌, 특허, 및 특허 출원은, 마치 각각의 개별 공개문헌, 특허, 또는 특허 출

원이 참조로 인용되는 것으로 구체적으로 및 개별적으로 명시된 것과 같은 정도로 본원에서 참조로 인용된다.

발명의 내용

[0009] 본원에서는 고도로 정확한 핵산 라이브러리의 효율적인 드 노보 합성 및 스크리닝을 위한 시스템, 방법, 및 조성물을 제공한다. 본원에 기술된 핵산 라이브러리는 유전자, 유전자 클러스터, 생물학적 경로, 또는 전체 게놈을 특이적으로 표적화하고, 편집하기 위한 핵산을 포함한다.

[0010] 본원에서는 핵산 라이브러리로서, 적어도 500개의 동일하지 않은 DNA 분자를 포함하고, 여기서, 각각의 동일하지 않은 DNA 문자는 상이한 gRNA 서열을 코딩하고, 적어도 500개의 동일하지 않은 DNA 문자 중 적어도 약 80%는 각각, 핵산 라이브러리 중 각각의 동일하지 않은 DNA 문자에 대한 평균 빈도의 2x 이내의 양으로 상기 라이브러리 중에 존재하는 것인 핵산 라이브러리를 제공한다. 본원에서는 각각의 동일하지 않은 DNA 문자의 GC 염기 함량이 약 20% 내지 약 85%인 핵산 라이브러리를 제공한다. 본원에서는 각각의 동일하지 않은 DNA 문자의 GC 염기 함량이 약 30% 내지 약 70%인 핵산 라이브러리를 제공한다. 본원에서는 적어도 500개의 동일하지 않은 DNA 문자 중 적어도 약 90%는 각각, 핵산 라이브러리 중 각각의 동일하지 않은 DNA 문자에 대한 평균 빈도의 2x 이내의 양으로 상기 라이브러리 중에 존재하는 것인 핵산 라이브러리를 제공한다. 본원에서는 적어도 500개의 동일하지 않은 DNA 문자 중 적어도 약 99%는 각각, 핵산 라이브러리 중 각각의 동일하지 않은 DNA 문자에 대한 평균 빈도의 2x 이내의 양으로 상기 라이브러리 중에 존재하는 것인 핵산 라이브러리를 제공한다. 본원에서는 적어도 500개의 동일하지 않은 DNA 문자가 적어도 2,000개의 동일하지 않은 DNA 문자를 포함하는 것인 핵산 라이브러리를 제공한다. 본원에서는 적어도 500개의 동일하지 않은 DNA 문자가 적어도 3,500개의 동일하지 않은 DNA 문자를 포함하는 것인 핵산 라이브러리를 제공한다. 본원에서는 적어도 500개의 동일하지 않은 DNA 문자를 포함하는 것인 핵산 라이브러리를 제공한다. 본원에서는 적어도 500개의 동일하지 않은 DNA 문자가 적어도 100,000개의 동일하지 않은 DNA 문자를 포함하는 것인 핵산 라이브러리를 제공한다. 본원에서는 각각의 동일하지 않은 DNA 문자의 길이가 최대 200개의 염기 길이를 포함하는 것인 핵산 라이브러리를 제공한다. 본원에서는 각각의 동일하지 않은 DNA 문자의 길이가 약 100개 내지 약 200개의 염기 길이를 포함하는 것인 핵산 라이브러리를 제공한다. 본원에서는 적어도 500개의 동일하지 않은 DNA 문자가 생물학적 경로 중의 유전자를 표적화하는 gRNA 서열을 코딩하는 동일하지 않은 DNA 문자를 포함하는 것인 핵산 라이브러리를 제공한다. 본원에서는 적어도 500개의 동일하지 않은 DNA 문자가 전체 게놈 중의 유전자를 표적화하는 gRNA 서열을 코딩하는 동일하지 않은 DNA 문자를 포함하는 것인 핵산 라이브러리를 제공한다. 본원에서는 gRNA가 단일 gRNA 또는 이중 gRNA인 핵산 라이브러리를 제공한다.

[0011] 본원에서는 핵산 라이브러리로서, 적어도 2,000개의 동일하지 않은 핵산을 포함하고, 각각의 동일하지 않은 핵산은 상이한 sgRNA 서열을 코딩하고, 각각의 sgRNA 서열은 진핵생물 유전자에 상보적인 표적화 도메인을 포함하며, 적어도 2,000개의 동일하지 않은 핵산 중 적어도 약 80%는 핵산 라이브러리 중 각각의 동일하지 않은 핵산에 대한 평균 빈도의 2x 이내의 양으로 상기 라이브러리 중에 존재하는 것인 핵산 라이브러리를 제공한다. 본원에서는 각각의 동일하지 않은 핵산의 GC 염기 함량이 약 20% 내지 약 85%인 핵산 라이브러리를 제공한다. 본원에서는 각각의 동일하지 않은 핵산의 GC 염기 함량이 약 30% 내지 약 70%인 핵산 라이브러리를 제공한다. 본원에서는 적어도 2,000개의 동일하지 않은 핵산 중 적어도 약 90%는 각각, 핵산 라이브러리 중 각각의 동일하지 않은 핵산에 대한 평균 빈도의 2x 이내의 양으로 상기 라이브러리 중에 존재하는 것인 핵산 라이브러리를 제공한다. 본원에서는 적어도 2,000개의 동일하지 않은 핵산 중 적어도 약 99%는 각각, 핵산 라이브러리 중 각각의 동일하지 않은 핵산에 대한 평균 빈도의 2x 이내의 양으로 상기 라이브러리 중에 존재하는 것인 핵산 라이브러리를 제공한다. 본원에서는 각각의 동일하지 않은 핵산의 길이가 최대 200개의 염기 길이를 포함하는 것인 핵산 라이브러리를 제공한다. 본원에서는 각각의 동일하지 않은 핵산의 길이가 약 100개 내지 약 200개의 염기 길이를 포함하는 것인 핵산 라이브러리를 제공한다. 본원에서는 각각의 동일하지 않은 핵산이 생물학적 경로 중의 유전자를 표적화하는 sgRNA 서열을 코딩하는 동일하지 않은 핵산을 포함하는 것인 핵산 라이브러리를 제공한다. 본원에서는 적어도 2,000개의 동일하지 않은 핵산이 전체 게놈 중의 유전자를 표적화하는 sgRNA 서열을 코딩하는 동일하지 않은 핵산을 포함하는 핵산 라이브러리를 제공한다. 본원에서는 각각의 동일하지 않은 핵산이 DNA 또는 RNA 분자를 포함하는 것인 핵산 라이브러리를 제공한다.

[0012] 본원에서는 앰플리콘 라이브러리가 복수의 동일하지 않은 DNA 분자를 포함하고, 여기서, 각각의 동일하지 않은 DNA가 증폭 생성물의 집단 중에 존재하고, 각각의 동일하지 않은 DNA 문자는 상이한 gRNA 서열을 코딩하며, 복수의 동일하지 않은 DNA 문자 중 적어도 약 80%는 각각, 앰플리콘 라이브러리 중 각각의 동일하지 않은 DNA 문자에 대한 평균 빈도의 2x 이내의 양으로 상기 라이브러리 중에 존재하는 것인 앰플리콘 라이브러리를 제공한다. 본원에서는 각각의 동일하지 않은 DNA 문자의 GC 염기 함량이 약 30% 내지 약 70%인 앰플리콘 라이브

러리를 제공한다. 본원에서는 gRNA가 단일 gRNA 또는 이중 gRNA인 앰플리콘 라이브러리를 제공한다.

[0013]

본원에서는 세포 라이브러리로서, 복수의 세포 집단을 포함하고, 여기서, 각각의 세포 집단은 상이한 gRNA 서열을 코딩하는 DNA 분자를 포함하고, 각 gRNA 서열은 유전자에의 결합을 위한 표적화 영역을 포함하며, 세포 집단 중 적어도 15%는 유전자 발현에서 적어도 2배 감손(depletion)을 보이는 것인 세포 라이브러리를 제공한다. 본원에서는 세포 집단 중 적어도 45%는 유전자 발현에서 적어도 2배 감손을 보이는 것인 세포 라이브러리를 제공한다. 본원에서는 gRNA가 단일 gRNA 또는 이중 gRNA인 세포 라이브러리를 제공한다. 본원에서는 복수의 세포 집단이 단일 유전자당 적어도 3개의 상이한 gRNA 서열을 코딩하는 DNA 분자를 포함하는 것인 세포 라이브러리를 제공한다. 본원에서는 복수의 세포 집단이 단일 유전자당 적어도 5개의 상이한 gRNA 서열을 코딩하는 DNA 분자를 포함하는 것인 세포 라이브러리를 제공한다. 본원에서는 복수의 세포 집단이 적어도 2,000개의 세포 집단을 포함하는 것인 세포 라이브러리를 제공한다. 본원에서는 복수의 세포 집단이 생물학적 경로 중의 gRNA 서열을 코딩하는 DNA 분자를 포함하는 것인 세포 라이브러리를 제공한다. 본원에서는 복수의 세포 집단이 전체 게놈 중의 gRNA 서열을 코딩하는 DNA 분자를 포함하는 것인 세포 라이브러리를 제공한다. 본원에서는 게놈이 아라비듐시스 탈리아나(*Arabidopsis thaliana*), 카에노르합디티스 엘레강스(*Caenorhabditis elegans*), 캐니스 루푸스 패밀리아리스(*Canis lupus familiaris*), 클라미도모나스 레인하드티이(*Chlamydomonas reinhardtii*), 다니오 레리오(*Danio rerio*), 딕티오스텔리움 디스코이데움(*Dictyostelium discoideum*), 드로소필라 멜라노가스터(*Drosophila melanogaster*), 에스테리키아 콜라이(*Escherichia coli*), 호모 사피엔스(*Homo sapiens*), 마카카 물라타(*Macaca mulatta*), 무스 무스클루스(*Mus musculus*), 오릭토라구스 쿠니쿨루스(*Oryctolagus cuniculus*), 라투스 노르베기쿠스(*Rattus norvegicus*), 사카로마이세스 세레비시아에(*Saccharomyces cerevisiae*), 또는 수스 스크로파(*Sus scrofa*)인 세포 라이브러리를 제공한다. 본원에서는 각각의 세포 집단이 원핵 세포를 포함하는 것인 세포 라이브러리를 제공한다. 본원에서는 각각의 세포 집단이 진핵 세포를 포함하는 것인 세포 라이브러리를 제공한다. 본원에서는 각각의 세포 집단이 포유동물 세포를 포함하는 것인 세포 라이브러리를 제공한다. 본원에서는 각각의 세포 집단이 외인성 뉴클레아제 효소를 추가로 포함하는 것인 세포 라이브러리를 제공한다. 본원에서는 DNA 분자가 벡터 서열을 추가로 포함하는 것인 세포 라이브러리를 제공한다.

[0014]

본원에서는 세포 라이브러리로서, 복수의 세포 집단을 포함하고, 여기서, 각각의 세포 집단은 상이한 gRNA 서열을 코딩하는 DNA 분자를 포함하고, 각 gRNA 서열은 유전자에의 결합을 위한 표적화 영역을 포함하며, 세포 집단 중 최대 20%는 유전자 발현에서 0 또는 음의 감소를 보이는 것인 세포 라이브러리를 제공한다. 본원에서는 gRNA가 단일 gRNA 또는 이중 gRNA인 세포 라이브러리를 제공한다. 본원에서는 복수의 세포 집단이 단일 유전자당 적어도 3개의 상이한 gRNA 서열을 코딩하는 DNA 분자를 포함하는 것인 세포 라이브러리를 제공한다. 본원에서는 복수의 세포 집단이 단일 유전자당 적어도 5개의 상이한 gRNA 서열을 코딩하는 DNA 분자를 포함하는 것인 세포 라이브러리를 제공한다. 본원에서는 각각의 세포 집단이 포유동물 세포를 포함하는 것인 세포 라이브러리를 제공한다. 본원에서는 각각의 세포 집단이 적어도 2,000개의 세포 집단을 포함하는 것인 세포 라이브러리를 제공한다. 본원에서는 복수의 세포 집단이 적어도 적어도 10,000개의 세포 집단을 포함하는 것인 세포 라이브러리를 제공한다.

[0015]

본원에서는 적어도 500개의 동일하지 않은 DNA 분자에 대한 소정의 서열을 제공하는 단계로서, 여기서, 각각의 동일하지 않은 DNA 분자는 gRNA를 코딩하는 것인 단계; 적어도 500개의 동일하지 않은 DNA 분자를 합성하는 단계; 및 적어도 500개의 동일하지 않은 DNA 분자를 전사하여 gRNA의 라이브러리를 생성하는 단계를 포함하고, 여기서, gRNA의 라이브러리 중 적어도 약 75%의 gRNA는 적어도 500개의 동일하지 않은 DNA 분자에 대한 소정의 서열과 비교하여 오류가 없는 것인, gRNA 라이브러리를 합성하는 방법을 제공한다. 본원에서는 전사 단계 전에 적어도 500개의 동일하지 않은 DNA 분자를 세포 내로 전달하는 단계를 추가로 포함하는, gRNA 라이브러리를 합성하는 방법을 제공한다. 본원에서는 적어도 500개의 동일하지 않은 DNA 분자에 의해 코딩되는 gRNA 중 적어도 96%가 gRNA의 라이브러리 중에 존재하는 것인, gRNA 라이브러리를 합성하는 방법을 제공한다. 본원에서는 gRNA의 라이브러리 중 적어도 87%의 gRNA는 적어도 500개의 동일하지 않은 DNA 분자에 대한 소정의 서열과 비교하여 오류가 없는 것인, gRNA 라이브러리를 합성하는 방법을 제공한다. 본원에서는 적어도 500개의 동일하지 않은 DNA 분자를 벡터 내로 삽입하는 단계를 추가로 포함하는, gRNA 라이브러리를 합성하는 방법을 제공한다. 본원에서는 적어도 500개의 동일하지 않은 DNA 분자를 유기체의 세포로 전달하는 단계를 추가로 포함하는, gRNA 라이브러리를 합성하는 방법을 제공한다. 본원에서는 유기체가 아라비듐시스 탈리아나, 카에노르합디티스 엘레강스, 캐니스 루푸스 패밀리아리스, 클라미도모나스 레인하드티이, 다니오 레리오, 딕티오스텔리움 디스코이데움, 드로소필라 멜라노가스터, 에스테리키아 콜라이, 호모 사피엔스, 마카카 물라타, 무스 무스클루스, 오릭토라구스 쿠니쿨루스, 라투스 노르베기쿠스, 사카로마이세스 세레비시아에, 또는 수스 스크로파인, gRNA 라이브러리를 합성하는 방법을 제공한다. 본원에서는 각각의 동일하지 않은 DNA 분자가 단일 gRNA 또는 이중 gRNA를 코딩하는

것인, gRNA 라이브러리를 합성하는 방법을 제공한다.

[0016]

본원에서는 복수의 동일하지 않은 DNA 문자에 대한 소정의 서열을 제공하는 단계로서, 여기서, 각각의 동일하지 않은 DNA 문자는 gRNA를 코딩하는 것인 단계; 표면을 제공하는 단계로서, 여기서, 표면은 핵산 신장 반응을 위한 좌위의 클러스터를 포함하는 것인 단계; 복수의 동일하지 않은 DNA 문자를 합성하는 단계로서, 여기서, 각각의 동일하지 않은 DNA 문자는 표면으로부터 신장되는 것인 단계; 및 복수의 동일하지 않은 DNA 문자를 세포 내로 전달하는 단계를 포함하는, gRNA 라이브러리를 합성하는 방법을 제공한다. 본원에서는 각 클러스터가 약 50개 내지 약 500개의 좌위를 포함하는 것인, gRNA 라이브러리를 합성하는 방법을 제공한다. 본원에서는 각각의 동일하지 않은 DNA 문자의 길이가 최대 약 200개의 염기 길이를 포함하는 것인, gRNA 라이브러리를 합성하는 방법을 제공한다. 본원에서는 각각의 동일하지 않은 DNA 문자가 단일 gRNA 또는 이중 gRNA를 코딩하는 것인, gRNA 라이브러리를 합성하는 방법을 제공한다. 본원에서는 세포가 원핵 세포인, gRNA 라이브러리를 합성하는 방법을 제공한다. 본원에서는 세포가 진핵 세포인, gRNA 라이브러리를 합성하는 방법을 제공한다. 본원에서는 진핵 세포가 포유동물세포인, gRNA 라이브러리를 합성하는 방법을 제공한다. 본원에서는 각각의 세포가 외인성 뉴클레아제 효소를 포함하는 것인, gRNA 라이브러리를 합성하는 방법을 제공한다.

도면의 간단한 설명

[0017]

도 1a는 하기 성분: PAM, 표적 서열, CAS9 효소, 가이드 RNA(gRNA: Guide RNA), 및 도너 DNA를 포함하는, 주기적으로 간격을 띠고 분포하는 짧은 회문 구조의 반복부(CRISPR: Clustered Regularly Interspaced Short Palindromic Repeat) 복합체를 도시한 것이다.

도 1b는 하기 성분: PAM, 표적 서열, CAS9 효소, 가이드 RNA(gRNA), 및 비상동성 말단 연결 수복(NHEJ: non-homologous end joining repair) 경로를 위한 도너 DNA를 포함하는, 주기적으로 간격을 띠고 분포하는 짧은 회문 구조의 반복부(CRISPR) 복합체를 도시한 것이다.

도 2는 gRNA 라이브러리의 디자인, 합성, 클로닝, 패키징, 스크리닝 및 분석을 포함하는 gRNA 라이브러리 스크리닝 작업 흐름을 도시한 것이다.

도 3은 어레이 상에서 올리고핵산 라이브러리를 합성하고, 올리고핵산을 증폭시키고, 백터 내로 전달하고, gRNA 발현을 위한 발현 라이브러리를 형성하는 것을 포함하는, 라이브러리 구축을 위한 gRNA 라이브러리 스크리닝 작업 흐름을 도시한 것이다.

도 4a-4c는 다양한 gRNA의 다이어그램이다. 도 4a는 염기쌍 형성 영역, dCas9 헨들, 및 *S. pyogenes* 터미네이터 영역을 가지는 sgRNA 서열(서열 번호 40)의 다이어그램이다. 도 4b는 sgRNA 단독의 다이어그램이다. 도 4c는 dgRNA 단독의 다이어그램이다.

도 5a는 주형 가닥 표적화 배열로 된 sgRNA 서열의 다이어그램이다.

도 5b는 비주형 가닥 표적화 배열로 된 sgRNA 서열의 다이어그램이다.

도 6a는 전사되었을 때, 헤어핀 2차 구조를 형성하는 gRNA 서열을 생성하는, T7 프로모터와 gRNA 서열의 다이어그램이다.

도 6b는 전사되었을 때, 헤어핀 2차 구조를 형성하지 않는 gRNA 서열을 생성하는, T7 프로모터와 gRNA 서열의 다이어그램이다.

도 7은 표적 DNA의 시험판내 Cas9 매개 절단에 대한 작업 흐름을 도시한 것이다.

도 8은 컴퓨터 시스템의 예를 도시한 것이다.

도 9는 컴퓨터 시스템의 컴퓨터 아키텍처를 도시한 블록 다이어그램이다.

도 10은 복수의 컴퓨터 시스템, 복수의 휴대 전화 및 개인 정보 단말기, 및 네트워크 부착 스토리지(NAS: Network Attached Storage)를 도입하도록 구성된 네트워크를 나타내는 다이어그램이다.

도 11은 공유 가상 주소 메모리 공간을 사용하는 멀티프로세서 컴퓨터 시스템의 블록 다이어그램이다.

도 12는 4개의 sgRNA 디자인을 도시한 것이다. 도 12는 출현 순서대로 각각 서열 번호 20, 15, 21, 15, 22, 15, 41, 15, 42, 및 43을 개시한 것이다.

도 13a-13b는 X축 상의 뉴클레오티드 염기 및 Y축 상의 형광 단위를 포함하는, 바이오애널라이저(BioAnalyzer)

관독으로부터의 플롯이다.

도 14A-14J는 X축 상의 뉴클레오티드 염기 및 Y축 상의 형광 단위를 포함하는, 바이오애널라이저 관독으로부터의 플롯이다.

도 15는 각 클러스터가 올리고핵산의 신장 기점이 되는 121개의 좌위를 가지는 256개의 클러스터의 영상이다.

도 16a는 각 클러스터가 121개의 올리고핵산을 가지는 것인 240개의 클러스터로부터의 29,040개의 독특한 올리고핵산의 합성으로부터의 플레이트 간의 올리고핵산을 표현한 플롯(올리고핵산 빈도 대 존재도)이다.

도 16b는 각 개별 클러스터 간의 올리고핵산 빈도 대 존재도 측정을 나타낸 플롯이고, 여기서, 대조군 클러스터는 박스로 식별 표시되어 있다.

도 17은 4개의 개별 클러스터 간의 올리고핵산 빈도 대 존재도 측정을 나타낸 플롯이다.

도 18a는 각 클러스터가 121개의 올리고핵산을 가지는 것인 240개의 클러스터로부터의 29,040개의 독특한 올리고핵산의 합성으로부터의 플레이트 간의 오류율 대 빈도의 플롯이다.

도 18b는 각 개별 클러스터 간의 올리고핵산 오류율 대 빈도 측정을 나타낸 플롯이고, 여기서, 대조군 클러스터는 박스로 식별 표시되어 있다.

도 19는 4개의 클러스터 간의 올리고핵산 오류율 대 빈도 측정을 나타낸 플롯이다.

도 20은 올리고핵산당 비율의 척도로서의 GC 함량 대 올리고핵산 개수의 플롯이다.

도 21은 2개의 상이한 폴리머라제를 이용한 PCR로부터의 결과를 포함하는 플롯을 제공한다. 각 차트는 "빈도 관측치"(100,000개당 계수 "0" 내지 35"로 측정됨) 대 올리고핵산 개수(0 내지 2,000개)의 플롯이다.

도 22는 기록된 증폭 후의 올리고핵산 집단 균일성을 정량화한 차트를 제공하는 것이다.

도 23은 과다 증폭이 서열 드롭 아웃에 미치는 영향을 나타낸 플롯을 도시한 것이다.

도 24a-24b는 10,000개의 sgRNA 올리고핵산 CRISPR 라이브러리로부터 회수된 올리고핵산의 서열분석으로부터 얻은 결과를 도시한 것이다.

도 25는 101,000개의 sgRNA 올리고핵산 CRISPR 라이브러리로부터 회수된 올리고핵산의 서열분석으로부터 얻은 결과를 도시한 것이다.

도 26a는 적어도 2배 감소된 sgRNA의 비율(%)을 나타낸 그래프를 도시한 것이다.

도 26b는 0 또는 음의 값의 감소를 보이는 sgRNA의 비율(%)을 나타낸 그래프를 도시한 것이다.

발명을 실시하기 위한 구체적인 내용

[0018]

본원에서는 고도로 정확한 가이드 RNA("gRNA") 라이브러리의 효율적 합성 및 스크리닝을 위한 시스템, 방법, 및 조성물을 제공한다. 본원에 기술된 드 노보 합성 방법은 표적화된 유전자 편집을 위해 효소 시스템 내로 도입하기 위한 큰 gRNA 라이브러리의 신속하고, 고도로 정확한 생성을 제공한다.

[0019]

정의

[0020]

본 개시내용 전역에 걸쳐, 다양한 실시양태가 범위 형식으로 제시된다. 범위 형식의 설명은 단지 편의 및 간결함을 위한 것이며, 임의의 실시양태의 범주에 대한 변경불가능한 제한으로서 해석되어서는 안 된다는 것을 이해하여야 한다. 따라서, 문맥상 달리 분명하게 명시되지 않은 한, 범위에 관한 설명은 모든 가능한 하위범위 뿐만 아니라, 그 범위 내의 개별 수치 값도 하한 단위의 10분의 1까지 구체적으로 개시한 것으로서 간주되어야 한다. 예를 들어, 예컨대, 1 내지 6이라는 범위에 관한 설명은 하위범위, 예컨대, 1 내지 3, 1 내지 4, 1 내지 5, 2 내지 4, 2 내지 6, 3 내지 6 등 뿐만 아니라, 그 범위 내의 개별 값, 예를 들어, 1.1, 2, 2.3, 5, 및 5.9도 구체적으로 개시한 것으로서 간주되어야 한다. 이는 범위 폭과 상관없이 적용된다. 이러한 개체 범위의 상한 및 하한은 독립적으로 더 작은 범위 내에 포함될 수 있고, 이 또한 언급된 범위에서 임의의 구체적으로 배제된 한계를 제외하고 본 발명 내에 포함된다. 문맥상 달리 분명하게 명시되지 않은 한, 언급된 범위가 한계 중 하나 또는 둘 모두를 포함하는 경우, 포함된 한계 중 하나 또는 둘 모두를 배제한 범위 또한 본 개시내용 내에 포함된다.

[0021]

본원에서 사용된 용어는 단지 특정 실시양태를 기술하기 위한 것이며, 어느 실시양태도 제한하는 것으로 의도되

지 않는다. 본원에서 사용되는 바, 문맥상 달리 명시되지 않은 한, "하나"("a," "an") 및 "그"라는 단수 형태는 복수 형태도 포함하는 것으로 의도된다. 본 명세서에서 사용될 때, "포함하다" 및 "~을 포함하는"이라는 용어는 언급된 특징, 정수, 단계, 작업, 요소 및/또는 성분이 존재함을 명시하지만, 하나 이상의 다른 특징, 정수, 단계, 작업, 요소, 성분, 및/또는 그의 군의 존재 또는 부가를 배제시키는 것은 아님을 추가로 이해할 것이다. 본원에서 사용되는 바, "및/또는"이라는 용어는 연관된 열거 목록 중 하나 이상의 것의 임의의 및 모든 조합을 포함한다.

[0022] 문맥상 구체적으로 언급되지 않거나, 또는 문맥으로부터 명백해지지 않는 한, 수치 또는 수치 범위와 관련하여 본원에서 사용되는 "약"이라는 용어는 언급된 수치 및 수치 +/- 그의 10%, 또는 범위로 열거된 값인 경우, 열거된 하한 아래로 10% 내지 열거된 상한 위로 10%를 의미하는 것으로 이해된다.

[0023] 본원에서 지칭되는 "gRNA"라는 용어는 가이드 RNA 서열을 지칭하고, 단일 및 이중 가이드 RNA 서열, 둘 모두를 포함한다. 문맥상 구체적으로 언급되지 않거나, 또는 문맥으로부터 명백해지지 않는 한, 본원에서 사용되는 바, 본원에서 지칭되는 "dgRNA"라는 용어는 이중 가이드 RNA 서열: crRNA(표적 서열에 상보적인 시드 영역을 스페이서 서열) 및 RNA에 상보적인, 별개의 tracrRNA(전사활성화 서열)를 지칭한다. 문맥상 구체적으로 언급되지 않거나, 또는 문맥으로부터 명백해지지 않는 한, 본원에서 사용되는 바, 본원에서 지칭되는 "sgRNA"라는 용어는 융합된 crRNA 및 tracrRNA, 둘 모두를 포함하는, 단일 가이드 RNA 서열을 지칭한다.

[0024] 문맥상 구체적으로 언급되지 않거나, 또는 문맥으로부터 명백해지지 않는 한, 본원에서 사용되는 바, "올리고핵산" 및 "핵산"이라는 용어는 단일 가닥 문자 뿐만 아니라, 이중 또는 삼중 가닥 핵산도 포함한다. 이중 또는 삼중 가닥 핵산에서 핵산 가닥은 동연적(coextensive)일 필요는 없다(즉, 이중 가닥 핵산이 전장의 두 가닥 모두를 따라 이중 가닥이어야 할 필요는 없다). 달리 언급되지 않는 한, 제공될 때, 핵산 서열은 5'에서 3' 반향으로 열거된다. 본원에 기술된 방법은 단리된 핵산을 생성하는 것을 제공한다. 본원에 기술된 방법은 단리 및 정제된 핵산을 생성하는 것을 추가로 제공한다. 본원에서 지칭되는 "올리고핵산" 및 "핵산"은 적어도 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2,000개 이상의 염기 길이를 포함할 수 있다.

[0025] 문맥상 구체적으로 언급되지 않거나, 또는 문맥으로부터 명백해지지 않는 한, 본원에서 사용되는 바, 본원에서 사용되는 "앰플리콘"이라는 용어는 증폭 반응 생성물을 지칭한다.

[0026] 달리 명시되지 않는 한, 본 개시내용은 당업계의 기술 범위 내에 포함되어 있는 종래 분자 생물학 기술을 이용한다. 달리 정의되지 않는 한, 본원에서 사용된 모든 기술 용어 및 과학 용어는 본 발명이 속하는 분야의 숙련가가 보편적으로 이해하는 것과 동일한 의미를 가진다.

gRNA 라이브러리 스크리닝

[0028] 본원에서는 주기적으로 간격을 띠고 분포하는 짧은 회문 구조의 반복부(CRISPR)-효소 복합체에 도입하기 위한 고도로 정확한 gRNA의 라이브러리를 디자인, 구축 및 스크리닝하는 방법을 제공한다. 예컨대, 도 1a-ab를 참조한다. 본원에 기술된 방법을 사용하여 생성된 gRNA 라이브러리는 sgRNA 및 dgRNA 라이브러리, 둘 모두를 포함한다. 본원에서는 생성된 라이브러리에서 소정의 gRNA를 높게 출현시키는 고도로 균일한 합성 방법을 제공한다. 도 2를 참조한다. 디자인 전략법은 제한 없이, 유전자를 포함하는 gRNA를 디자인하는 것을 포함한다. 원하는 작업 흐름에 의존하여, 드 노보 합성된 핵산은 DNA 또는 RNA 염기이다.

[0029] 드 노보 합성된 DNA인 경우, 합성된 각 핵산이 전사 생성물로서 gRNA(예컨대, sgRNA) 서열을 코딩하는 DNA 서열인, 핵산을 포함하는 라이브러리를 합성한다. 일부 경우에서, 이어서, 합성된 핵산을 발현 벡터로 삽입한다. 한 예시적인 작업 흐름에서, 합성된 핵산을 바이러스 벡터로 삽입한 후, 세포로의 형질도입을 위해 패키징한 후, 스크리닝 및 분석을 수행한다. 도 2. 예시적인 세포로는 제한 없이, 원핵 및 진핵 세포를 포함한다. 예시적인 진핵 세포로는 제한 없이, 동물, 식물, 및 진균 세포를 포함한다. 예시적인 동물 세포로는 제한 없이, 곤충, 어류 및 포유동물 세포를 포함한다. 예시적인 포유동물 세포로는 마우스, 인간, 및 영장류 세포를 포함한다. 테스트되는 세포 기능의 예로는 제한 없이, 세포 증식, 이동/부착, 대사 활성, 및 세포 신호전달 활성 변화를 포함한다. 드 노보 합성된 RNA인 경우, gRNA 그 자체가 합성되고, 예컨대, 세포로의 형질감염과 같은 하류 적용에 이용가능하다.

[0030] 올리고핵산은 어레이(301) 상에서의 신장을 위해 클러스터(303)의 위치("좌우") 내에서 합성될 수 있다. 도 3 참조. 상기 배열은 클러스터링된 좌우 배열이 없는 전체 플레이트 전역에 걸쳐 합성된 올리고핵산의 증폭 생성

물과 비교하였을 때, "앰플리콘"으로 명명되는, 합성된 올리고핵산의 증폭 생성물의 개선된 올리고핵산 출현을 제공할 수 있다. 일부 경우에서, 단일 클러스터 내에서 합성된 올리고핵산의 증폭(310)은 증폭 반응 생성물에서 GC가 낮은 또는 GC가 높은 앰플리콘을 과소출현하는 것에 기인하여, 통상 "드리프트"로 명명되는, GC 함량이 상당한 올리고핵산을 가지는 큰 올리고핵산 집단의 반복된 합성에 기인하여 나타나는 출현에 대해 미치는 부정적 효과에 반작용을 일으킨다. 일부 경우에서, 본원에 기술된 단일 클러스터는 약 50-1,000, 75-900, 100-800, 125-700, 150-600, 200-500, 또는 300-400개의 별개의 좌위를 포함한다. 일부 경우에서, 단일 클러스터는 50-500개의 별개의 좌위를 포함한다. 일부 경우에서, 좌위는 스폽, 웨, 마이크로웨, 채널, 또는 포스트이다. 일부 경우에서, 각 클러스터는 동일한 서열을 가지는 올리고핵산의 신장을 지원하는 별개의 특징부가 적어도 1X, 2X, 3X, 4X, 5X, 6X, 7X, 8X, 9X, 10X 이상 반복된다.

[0031] 본원에서는 발현 벡터로의 삽입을 위한 gRNA 라이브러리를 제공한다. 도 3에서의 작업 흐름을 연속하여, 아래이 (301)는 올리고핵산 합성 및 신장을 위해 다중의 클러스터(303)의 좌위를 포함한다. 드 노보 DNA가 합성되고, 플레이트로부터 제거됨으로써 올리고핵산(305)(예컨대, sgRNA를 코딩하는 DNA)가 형성되고, 이를 증폭(310)를 형성시켜 벡터(330)으로의 삽입을 위한 올리고핵산의 라이브러리(320)를 형성함으로써 합성된 DNA를 포함하는 벡터의 라이브러리(335)를 형성한다. 일단 세포에서는 DNA는 gRNA(예컨대, sgRNA)로 전사되고, 게놈 편집 방식(예컨대, Cas9 기반 시스템)과의 결합을 위해 이용가능하다. 세포는 편집 효소(예컨대, Cas9)의 천연 또는 이소성 발현을 보일 수 있다. 편집 효소(예컨대, Cas9)는 이중 DNA 가닥 절단 활성, 또는 변형된 활성, 예컨대, 낙킹, 염기 스와핑 또는 서열 스와핑 활성을 가질 수 있다. 벡터로의 삽입을 위한 합성된 DNA는 sgRNA, dgRNA, 또는 그의 단편을 포함할 수 있다.

[0032] 본원에 개시된 핵산 라이브러리 삽입을 위한 발현 벡터로는 진핵 또는 원핵 발현 벡터를 포함한다. 예시적인 발현 벡터로는 제한 없이, 포유동물 발현 벡터: pSF-CMV-NEO-NH2-PPT-3XFLAG, pSF-CMV-NEO-COOH-3XFLAG, pSF-CMV-PURO-NH2-GST-TEV, pSF-OXB20-COOH-TEV-FLAG(R)-6His, pCEP4 pDEST27, pSF-CMV-Ub-KrYFP, pSF-CMV-FMDV-daGFP, pEF1a-m체리-N1 벡터(pEF1a-mCherry-N1 Vector), pEF1a-td토마토 벡터(pEF1a-tdTomato Vector), pSF-CMV-FMDV-하이그로(pSF-CMV-FMDV-Hygro), pSF-CMV-PGK-퓨로(pSF-CMV-PGK-Puro), pMCP-태그(m)(pMCP-tag(m)), 및 pSF-CMV-PURO-NH2-CMYC; 박테리아 발현 벡터: pSF-OXB20-베타Gal(pSF-OXB20-BetaGal), pSF-OXB20-Fluc, pSF-OXB20, 및 pSF-Tac; 식물 발현 벡터: pRI 101-AN DNA 및 pCambia2301; 및 효모 발현 벡터: pTYB21 및 pKLAC2, 및 곤충 벡터: pAc5.1/V5-His A 및 pDEST8을 포함한다.

[0033] 본원에 기술된 방법에 의해 합성된 드 노보 올리고핵산 라이브러리는 세포에서 발현될 수 있다. 일부 경우에서, 세포는 질환 상태와 연관이 있다. 예를 들어, 질환 상태와 연관이 있는 세포로는 세포주, 조직 샘플, 피험체로부터의 1차 세포, 피험체로부터 확장된 배양된 세포, 또는 모델 시스템 중의 세포를 포함하나, 이에 제한되지 않는다. 일부 경우에서, 모델 시스템은 식물 또는 동물 시스템이다. 일부 경우에서, 드 노보 올리고핵산 라이브러리는 세포 활성 변화를 사정하기 위해 세포에서 발현된다. 예시적인 세포 활성으로는 제한 없이, 증식, 사이클 진행, 세포 사멸, 부착, 이동, 생식, 세포 신호전달, 에너지 생산, 산소 이용률, 대사 활성, 및 노화, 자유 라디칼 손상에 대한 반응, 또는 그의 임의의 조합을 포함한다.

[0034] 본원에서는 gRNA 라이브러리가 유전자 1개당 복수의 동일하지 않은 gRNA를 포함하는 것인, gRNA 라이브러리(전사시, gRNA 라이브러리를 생성하는 DNA 라이브러리)를 합성하는 방법을 제공한다. gRNA는 sgRNA 또는 dgRNA를 코딩할 수 있다. 일부 경우에서, gRNA 라이브러리는 유전자 1개당 적어도 3, 4, 5, 6, 7, 8, 9, 10개, 또는 10개 초과의 동일하지 않은 gRNA를 포함한다. 일부 경우에서, gRNA 라이브러리는 하나 이상의 유전자를 표적화한다. 일부 경우에서, gRNA 라이브러리는 약 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100개, 또는 100개 초과의 유전자를 표적화한다. 일부 경우에서, gRNA 라이브러리는 약 1-100, 2-95, 5-90, 10-85, 15-80, 20-75, 25-70, 30-65, 35-60, 또는 40-50개의 유전자를 표적화한다. 일부 경우에서, 본원에 기술된 gRNA 라이브러리는 경로 내의 유전자를 표적화한다. 예시적인 경로로는 제한 없이, 대사, 세포 사멸, 세포 주기 진행, 면역 세포 활성화, 염증 반응, 혈관신생, 램프구신생, 저산소증 및 산화적 스트레스 반응, 세포 부착, 및 세포 이동 경로를 포함한다.

[0035] 본원에 기술된 바와 같은 gRNA 라이브러리를 합성하는 방법은 게놈의 일부에 상보적인 염기쌍 형성 영역인 게놈 표적 영역을 가지는 동일하지 않은 gRNA의 합성을 제공할 수 있다. 게놈 표적 영역은 엑손, 인트론, 코딩, 또는 비코딩 서열을 포함할 수 있다. 일부 경우에서, gRNA 라이브러리는 집단적으로 전체 게놈 중 적어도 또는 약 5%의 유전자에 상보적인 염기쌍 형성 영역을 가지는 동일하지 않은 gRNA를 포함한다. 일부 경우에서, gRNA 라이브러리는 집단적으로 전체 게놈 중 적어도 또는 약 80%의 유전자에 상보적인 염기쌍 형성 영역을 가지는 동일하지 않은 gRNA를 포함한다. 일부 경우에서, gRNA 라이브러리는 집단적으로 전체 게놈 중 적어도 또는 약 90%의 유전

자에 상보적인 염기쌍 형성 영역을 가지는 동일하지 않은 gRNA를 포함한다. 일부 경우에서, gRNA 라이브러리는 집단적으로 전체 게놈 중 적어도 또는 약 95%의 유전자에 상보적인 염기쌍 형성 영역을 가지는 동일하지 않은 gRNA를 포함한다. 일부 경우에서, gRNA 라이브러리는 집단적으로 전체 게놈 중 적어도 또는 약 100%의 유전자에 상보적인 염기쌍 형성 영역을 가지는 동일하지 않은 gRNA를 포함한다.

[0036] 본원에서는 상이한 세포 간에 적어도 2X의 유전자 감손을 보이는 gRNA를 생성하는 본원에 기술된 방법에 의해 합성된 gRNA 라이브러리를 제공한다. 일부 경우에서, gRNA 라이브러리는 세포 중에 또는 복수의 세포 집단 중에 존재할 때, 적어도 2X의 유전자 감손을 제공하는 gRNA를 적어도 또는 약 10%, 12%, 15%, 16%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65% 이상 포함한다. 일부 경우에서, 유전자는 필수 유전자, 즉, 세포 생존에 중요한 유전자이다. 예시적인 필수 유전자로는 제한 없이, PCNA, PSMA7, RPP21, 및 SF3B3을 포함한다. 일부 경우에서, gRNA 라이브러리는 세포 중에 존재할 때, 적어도 2X, 3X, 4X, 5X, 6X, 또는 6X 초과의 유전자 감손을 제공하는 gRNA를 포함한다. 일부 경우에서, gRNA 라이브러리는 세포 중에 또는 복수의 세포 집단 중에 존재할 때, 0 또는 음의 값의 유전자 감손을 보이는 gRNA를 최대 5%, 10%, 12%, 15%, 또는 20% 포함한다. 일부 경우에서, 복수의 세포 집단은 적어도 또는 약 2,000, 3,000, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 12,000, 15,000, 20,000, 25,000, 30,000개, 또는 30,000개 초과의 세포 집단을 포함한다. 일부 경우에서, gRNA 라이브러리는 복수의 유전자에 대하여 적어도 2X, 3X, 4X, 5X, 6X, 또는 6X 초과의 감손을 보이는 gRNA를 포함한다. 일부 경우에서, gRNA 라이브러리는 복수의 유전자에 대하여 적어도 2X의 감손을 보이는 gRNA를 평균 적어도 또는 약 40%, 45%, 50%, 60%, 65%, 70%, 75%, 80%, 또는 90% 초과로 포함한다. 상기 유전자 결실 프로파일을 제공하는 gRNA는 sgRNA 또는 dgRNA일 수 있다.

[0037] 본원에서는 고도로 균일한 올리고핵산의 라이브러리를 합성하는 방법을 제공한다. 일부 경우에서, 90% 초과의 합성된 올리고핵산(RNA 또는 DNA)는 핵산 라이브러리에 대한 올리고핵산 출현의 평균값의 4X 내로 나타난다. 일부 경우에서, 90% 초과의 올리고핵산은 라이브러리에 대한 올리고핵산 출현의 평균값의 2X 내로 나타난다. 일부 경우에서, 90% 초과의 올리고핵산은 라이브러리에 대한 올리고핵산 출현의 평균값의 1.5X 내로 나타난다. 일부 경우에서, 80% 초과의 올리고핵산은 라이브러리에 대한 올리고핵산 출현의 평균값의 1.5X 내로 나타난다.

[0038] 본원에 기술된 방법에 의해 드 노보 합성된 올리고핵산 라이브러리은 소정의 서열과 비교하여 정확한 서열을 고비율로 포함한다. 일부 경우에서, 본원에 개시된 드 노보 올리고핵산 라이브러리는 올리고핵산에 대하여 소정의 서열과 비교하여 70% 초과로 정확한 서열을 가진다. 일부 경우에서, 본원에 개시된 드 노보 올리고핵산 라이브러리는 올리고핵산에 대하여 소정의 서열과 비교하여 75% 초과로 정확한 서열을 가진다. 일부 경우에서, 본원에 개시된 드 노보 올리고핵산 라이브러리는 올리고핵산에 대하여 소정의 서열과 비교하여 80% 초과로 정확한 서열을 가진다. 일부 경우에서, 본원에 개시된 드 노보 올리고핵산 라이브러리는 올리고핵산에 대하여 소정의 서열과 비교하여 85% 초과로 정확한 서열을 가진다. 일부 경우에서, 본원에 개시된 드 노보 올리고핵산 라이브러리는 올리고핵산에 대하여 소정의 서열과 비교하여 90% 초과로 정확한 서열을 가진다. 일부 경우에서, 본원에 개시된 드 노보 올리고핵산 라이브러리는 올리고핵산에 대하여 소정의 서열과 비교하여 95% 초과로 정확한 서열을 가진다. 일부 경우에서, 본원에 개시된 드 노보 올리고핵산 라이브러리는 올리고핵산에 대하여 소정의 서열과 비교하여 100% 초과로 정확한 서열을 가진다.

[0039] 일부 경우에서, 본원에 개시된 드 노보 합성된 올리고핵산 라이브러리는 증폭 반응 후 올리고핵산에 대하여 소정의 서열과 비교하여 70% 초과로 정확한 서열을 가진다. 일부 경우에서, 본원에 개시된 드 노보 합성된 올리고핵산 라이브러리는 증폭 반응 후 올리고핵산에 대하여 소정의 서열과 비교하여 75% 초과로 정확한 서열을 가진다. 일부 경우에서, 본원에 개시된 드 노보 합성된 올리고핵산 라이브러리는 증폭 반응 후 올리고핵산에 대하여 소정의 서열과 비교하여 80% 초과로 정확한 서열을 가진다. 일부 경우에서, 본원에 개시된 드 노보 합성된 올리고핵산 라이브러리는 증폭 반응 후 올리고핵산에 대하여 소정의 서열과 비교하여 85% 초과로 정확한 서열을 가진다. 일부 경우에서, 본원에 개시된 드 노보 합성된 올리고핵산 라이브러리는 증폭 반응 후 올리고핵산에 대하여 소정의 서열과 비교하여 90% 초과로 정확한 서열을 가진다. 일부 경우에서, 본원에 개시된 드 노보 합성된 올리고핵산 라이브러리는 증폭 반응 후 올리고핵산에 대하여 소정의 서열과 비교하여 95% 초과로 정확한 서열을 가진다. 일부 경우에서, 본원에 개시된 드 노보 합성된 올리고핵산 라이브러리는 증폭 반응 후 올리고핵산에 대하여 소정의 서열과 비교하여 100% 초과로 정확한 서열을 가진다.

[0040] 일부 경우에서, 본원에 개시된 드 노보 합성된 올리고핵산 라이브러리는 세포 내로 전달되었을 때, 올리고핵산에 대하여 소정의 서열과 비교하여 80% 초과로 정확한 서열을 생성한다. 일부 경우에서, 본원에 개시된 드 노보 합성된 올리고핵산 라이브러리는 세포 내로 전달되었을 때, 올리고핵산에 대하여 소정의 서열과 비교하여 85% 초과로 정확한 서열을 생성한다. 일부 경우에서, 본원에 개시된 드 노보 합성된 올리고핵산 라이브러리는 세포

내로 전달되었을 때, 올리고핵산에 대하여 소정의 서열과 비교하여 90% 초과로 정확한 서열을 생성한다. 일부 경우에서, 본원에 개시된 드 노보 합성된 올리고핵산 라이브러리는 세포 내로 전달되었을 때, 올리고핵산에 대하여 소정의 서열과 비교하여 95% 초과로 정확한 서열을 생성한다. 일부 경우에서, 본원에 개시된 드 노보 합성된 올리고핵산 라이브러리는 세포 내로 전달되었을 때, 올리고핵산에 대하여 소정의 서열과 비교하여 100% 초과로 정확한 서열을 생성한다.

[0041] 일부 경우에서, 본원에 개시된 드 노보 합성된 올리고핵산 라이브러리는 세포 내로 전달되었을 때, 80% 초과로 서열을 나타낸다. 일부 경우에서, 본원에 개시된 드 노보 합성된 올리고핵산 라이브러리는 세포 내로 전달되었을 때, 90% 초과로 서열을 나타낸다. 일부 경우에서, 본원에 개시된 드 노보 합성된 올리고핵산 라이브러리는 세포 내로 전달되었을 때, 95% 초과로 서열을 나타낸다. 일부 경우에서, 본원에 개시된 드 노보 합성된 올리고핵산 라이브러리는 세포 내로 전달되었을 때, 100% 초과로 서열을 나타낸다.

[0042] 본원에 개시된 드 노보 올리고핵산 라이브러리에 대해 폴리머라제 효소 및 증폭 시약(예컨대, 완충제, 포스페이트, 및 dNTP) 첨가로 이루어지는 증폭 반응을 수행할 수 있다. 일부 경우에서, 드 노보 올리고핵산 라이브러리는 적어도 또는 약 6, 8, 10, 15, 20, 또는 20 초과의 사이클 동안 PCR에 의해 증폭된다. 일부 경우에서, 드 노보 올리고핵산 라이브러리는 약 6 내지 20, 7 내지 18, 8 내지 17, 9 내지 16, 또는 10 내지 15 사이클 범위로 PCR에 의해 증폭된다. 일부 경우에서, 드 노보 올리고핵산 라이브러리는 약 15 사이클 동안 PCR에 의해 증폭된다.

[0043] 일부 경우에서, 드 노보 올리고핵산 라이브러리의 증폭은 DNA 분자의 앰플리콘 라이브러리를 제공한다. 일부 경우에서, 앰플리콘 라이브러리는 gRNA 서열을 코딩하는 동일하지 않은 핵산을 포함한다. 일부 경우에서, gRNA 서열은 sgRNA 또는 dgRNA이다.

[0044] 일부 경우에서, 드 노보 올리고핵산 라이브러리는, 각각의 동일하지 않은 핵산이 DNA 분자를 포함하는 것인, 동일하지 않은 핵산을 포함한다. 일부 경우에서, DNA 분자의 개수는 약 500, 2,000, 3,500개 이상의 분자이다. 일부 경우에서, DNA 분자의 개수는 적어도 또는 약 250, 500, 1,000, 1,250, 2,000, 2,500, 3,000, 3,500, 4,000, 4,500, 5,000, 5,500, 6,000, 6,500, 7,000, 7,500, 8,000, 8,500, 9,000, 9,500, 10,000, 15,000, 20,000, 50,000, 100,000, 250,000, 500,000, 750,000, 1,000,000개, 또는 1,000,000개 초과의 분자이다. 일부 경우에서, DNA 분자의 개수는 최대 250, 500, 1,000, 1,250, 2,000, 2,500, 3,000, 3,500, 4,000, 4,500, 5,000, 5,500, 6,000, 6,500, 7,000, 7,500, 8,000, 8,500, 9,000, 9,500, 10,000, 15,000, 20,000, 50,000, 100,000, 250,000, 500,000, 750,000, 1,000,000개, 또는 1,000,000개 초과의 분자이다. 일부 경우에서, DNA 분자는 gRNA 서열을 코딩한다. 일부 경우에서, gRNA 서열은 sgRNA 또는 dgRNA이다.

[0045] 일부 경우에서, 드 노보 올리고핵산 라이브러리는, 각각의 동일하지 않은 핵산이 RNA 분자를 포함하는 것인, 동일하지 않은 핵산을 포함한다. 일부 경우에서, RNA 분자의 개수는 약 2,000개의 분자이다. 일부 경우에서, RNA 분자의 개수는 적어도 또는 약 250, 500, 1,000, 1,250, 2,000, 2,500, 3,000, 3,500, 4,000, 4,500, 5,000, 5,500, 6,000, 6,500, 7,000, 7,500, 8,000, 8,500, 9,000, 9,500, 10,000, 15,000, 20,000, 50,000, 100,000, 250,000, 500,000, 750,000, 1,000,000개, 또는 1,000,000개 초과의 분자이다. 일부 경우에서, RNA 분자의 개수는 최대 250, 500, 1,000, 1,250, 2,000, 2,500, 3,000, 3,500, 4,000, 4,500, 5,000, 5,500, 6,000, 6,500, 7,000, 7,500, 8,000, 8,500, 9,000, 9,500, 10,000, 15,000, 20,000, 50,000, 100,000, 250,000, 500,000, 750,000, 1,000,000개, 또는 1,000,000개 초과의 분자이다. 일부 경우에서, RNA 분자는 gRNA 서열을 코딩한다. 일부 경우에서, gRNA 서열은 sgRNA 또는 dgRNA이다.

[0046] 본원에서는 증폭 후 고도로 균일한 드 노보 올리고핵산 라이브러리를 제공한다. 일부 경우에서, 본원에 기술된 드 노보 올리고핵산 라이브러리 중 80% 초과의 올리고핵산은 증폭 후 전체 라이브러리에 대한 출현의 평균값의 적어도 약 1.5X 내로 나타난다. 일부 경우에서, 본원에 기술된 드 노보 올리고핵산 라이브러리 중 90% 초과의 올리고핵산은 증폭 후 전체 라이브러리에 대한 출현의 평균값의 적어도 약 1.5X 내로 나타난다. 일부 경우에서, 본원에 기술된 드 노보 올리고핵산 라이브러리 중 80% 초과의 올리고핵산은 증폭 후 전체 라이브러리에 대한 출현의 평균값의 적어도 약 2X 내로 나타난다. 일부 경우에서, 본원에 기술된 드 노보 올리고핵산 라이브러리 중 80% 초과의 올리고핵산은 증폭 후 전체 라이브러리에 대한 출현의 평균값의 적어도 약 2X 내로 나타난다.

[0047] 본원에 기술된 방법을 사용하여 드 노보 합성된 올리고핵산의 증폭되지 않은 집단은 동일하지 않은 올리고핵산 서열의 개수가 다를 수 있다. 일부 경우에서, 동일하지 않은 올리고핵산 서열의 개수는 약 2,000-1,000,000, 3,000 내지 900,000, 4,000-800,000, 5,000-700,000, 6,000-600,000, 7,000-500,000, 8,000-400,000, 9,000-300,000, 10,000-200,000, 11,000-100,000, 12,000-75,000, 14,000-60,000, 및 20,000-50,000개의 서열 범위

이다. 일부 경우에서, 동일하지 않은 올리고핵산 서열의 개수는 약 50-2,000, 75-1,800, 100-1,700, 150-1,600, 200-1,500, 250-1,400, 300-1,300, 400-1,200, 500-1,100, 600-1,000, 700-900개의 서열 범위이다. 일부 경우에서, 동일하지 않은 올리고핵산 서열의 개수는 2,000개의 서열이다. 일부 경우에서, 동일하지 않은 올리고핵산 서열의 개수는 1,000,000개 초과의 서열이다. 일부 경우에서, 동일하지 않은 올리고핵산 서열의 개수는 적어도 50, 100, 200, 400, 600, 800, 1,000, 1,200, 1,400, 1,600, 1,800, 2,000, 3,000, 5,000, 7,000, 10,000, 20,000, 30,000, 50,000, 100,000, 500,000, 700,000, 1,000,000, 10,000,000, 1,000,000,000개 이상의 서열이다. 일부 경우에서, 동일하지 않은 올리고핵산 서열의 개수는 최대 50, 100, 200, 400, 600, 800, 1,000, 1,200, 1,400, 1,600, 1,800, 2,000, 3,000, 5,000, 7,000, 10,000, 20,000, 30,000, 50,000, 100,000, 500,000, 700,000, 1,000,000개 이상의 서열이다. 일부 경우에서, 동일하지 않은 올리고핵산 서열의 개수는 최대 50, 100, 200, 400, 600, 800, 1,000, 1,200, 1,400, 1,600, 1,800, 2,000, 3,000, 5,000, 7,000, 10,000, 20,000, 30,000, 50,000, 100,000, 500,000, 700,000, 1,000,000개 이상의 서열이다.

[0048] 증폭되지 않은 집단의 올리고핵산은 상이한 양으로 존재할 수 있다. 일부 경우에서, 증폭되지 않은 집단의 올리고핵산은 적어도 또는 약 0.25 펫토몰인 양으로 존재한다. 일부 경우에서, 증폭되지 않은 집단의 올리고핵산은 적어도 또는 약 1 펫토몰인 양으로 존재한다. 일부 경우에서, 증폭되지 않은 집단의 올리고핵산은 적어도 0.25, 1, 10, 20, 30, 40, 50, 100, 250, 500, 750, 1,000 펫토몰, 또는 1,000 펫토몰 초과인 양으로 존재한다. 일부 경우에서, 증폭되지 않은 집단의 올리고핵산은 최대 0.25, 1, 10, 20, 30, 40, 50, 100, 250, 500, 750, 및 1,000 펫토몰인 양으로 존재한다.

[0049] 본원에서는 동일하지 않은 올리고핵산의 서열 길이 또는 평균 서열 길이가 상이한 것인, 동일하지 않은 올리고핵산의 라이브러리를 합성하는 방법을 제공한다. 일부 경우에서, 동일하지 않은 올리고핵산의 서열 길이 또는 평균 서열 길이는 최대 150개의 염기이다. 일부 경우에서, 동일하지 않은 올리고핵산의 서열 길이 또는 평균 서열 길이는 약 100개 내지 약 200개의 염기 범위이다. 일부 경우에서, 동일하지 않은 올리고핵산의 서열 길이 또는 평균 서열 길이는 적어도 30, 50, 100, 125, 150, 200, 250, 300, 350, 400, 450, 500개, 또는 500개 초과의 염기이다. 일부 경우에서, 동일하지 않은 올리고핵산의 서열 길이 또는 평균 서열 길이는 최대 150, 200, 250, 300, 350, 400, 450, 또는 500개의 염기이다. 동일하지 않은 올리고핵산의 예시적인 서열 길이는 약 25 내지 약 150개 또는 약 50 내지 약 200개의 염기 범위이다. 일부 경우에서, 동일하지 않은 올리고핵산의 서열 길이 또는 평균 서열 길이는 약 125 내지 약 200개 또는 약 150 내지 약 200개의 염기 범위이다.

가이드 RNA 서열

[0050] 본원에서는 게놈 서열 편집 효소(예컨대, Cas9)를 특정 표적 핵산 서열로 유도하기 위한 단일 가이드 RNA(sgRNA) 서열을 제공한다. Cas9 효소와의 복합체 형태인 sgRNA의 예는 **도 4a**에 도시되어 있고, 단독인 것의 예는 **도 4b**에 도시되어 있다. gRNA는 **도 4c**에 도시되어 있는 바와 같이, 이중 가이드 RNA일 수 있다. 본원에 개시된 가이드 서열은 염기쌍 형성 영역을 포함한다. 염기쌍 형성 영역은 표적 서열에의 결합을 위한 시드 영역, 및 임의적으로, 스페이서 영역을 포함한다. 염기쌍 형성 영역의 길이는 달라질 수 있다. 예를 들어, 염기쌍 형성 영역은 약 1 내지 10개, 1 내지 20개, 20 내지 25개, 또는 1 내지 30개의 염기 길이를 포함할 수 있다. 일부 경우에서, 염기쌍 형성 영역은 적어도 10, 15, 20, 25, 30개 이상의 염기 길이를 포함한다. 일부 경우에서, 염기쌍 형성 영역은 적어도 10개의 염기 길이의 시드 영역을 포함한다. 시드 영역은 약 8 내지 20개의 염기 길이를 포함할 수 있다. 일부 경우에서, 시드 영역의 길이는 약 12개의 염기 길이이다. 일부 경우에서, 본원에 기술된 염기쌍 형성 영역은 전사 동안 주형 가닥을 표적화하도록 디자인된다. **도 5a**. 일부 경우에서, 본원에 기술된 염기쌍 형성 영역은 전사 동안 비주형 가닥을 표적화하도록 디자인된다. **도 5b**.

[0052] 일부 경우에서, sgRNA의 염기쌍 형성 영역의 3'은 Cas9에의 결합을 위한 Cas9 핸들 영역이다. 일부 경우에서, Cas9 핸들 영역은 dCas9 효소에의 결합을 위한 dCas9 핸들 영역이다. 핸들 영역의 길이는 달라질 수 있다. 예를 들어, 핸들 영역은 약 1 내지 50개, 20 내지 45개, 또는 15 내지 60개의 염기 길이를 포함할 수 있다. 일부 경우에서, 핸들 영역은 적어도 35, 40, 45, 50개 이상의 염기 길이를 포함한다. 핸들 영역은 약 42개의 염기 길이를 포함할 수 있다.

[0053] 일부 경우에서, sgRNA의 핸들 영역의 3'은 터미네이터 영역이다. 일부 경우에서, 터미네이터 영역은 S. 피오제네스 터미네이터 영역이다. 일부 경우에서, 터미네이터 영역은 약 40개의 염기 길이를 포함한다. 일부 경우에서, 터미네이터 영역은 약 10 내지 50개, 20 내지 60개, 또는 30 내지 55개의 염기 길이를 포함한다.

[0054] 본원에 기술된 gRNA 서열에 대한 디자인 방식은 gRNA 서열을 코딩하는 DNA의 5' 상류에 DNA 의존성 RNA 폴리머라제 프로모터 영역의 포함을 포함할 수 있다. 예시적인 DNA 의존성 RNA 폴리머라제 프로모터 영역으로는 제한

없이, T3 및 T7 RNA 폴리머라제 프로모터 서열을 포함한다. 예를 들어, 도 6a는 T7 프로모터 영역이 gRNA의 5' 상류에 위치하고, gRNA가 헤어핀을 포함하는 것인, 전사된 생성된 gRNA가 생성되는 것인 배열을 도시한 것이다. 일부 배열에서, gRNA는 헤어핀 2차 구조를 형성하는 서열이 없도록 디자인된다(도 6b). 헤어핀 2차 구조는 Cas9 핸들 및/또는 터미네이터 영역에 존재하지 않을 수 있다.

[0055]

본원에서는 게놈 서열 편집 효소(예컨대, Cas9)를 특정 표적 핵산 서열로 유도하기 위한 dgRNA를 제공한다. 일부 경우에서, 라이브러리는 dgRNA에 대한 서열을 코딩하는 올리고핵산 서열을 포함한다. 일부 경우에서, 라이브러리는, 합성된 각 핵산이 전사 생성물로서 dgRNA 서열을 코딩하는 DNA 서열인, 핵산을 포함한다. 일부 경우에서, 라이브러리는 합성된 각 핵산이 RNA 서열이고, dgRNA 그 자체가 합성되는 것인, 핵산을 포함한다. 일부 경우에서, dgRNA의 라이브러리는 별개의 올리고핵산으로서 합성되는 crRNA 및 tracrRNA에 대한 올리고핵산 서열을 포함한다. 일부 경우에서, 올리고핵산 핵산은 별개로 crRNA 및 tracrRNA를 코딩한다. 일부 경우에서, 올리고핵산 핵산은 전사되었을 때, 별개의 crRNA 서열 및 별개의 tracrRNA 서열을 생성하는 단일 서열을 코딩한다. crRNA 및 tracrRNA에 대한 예시적인 서열은 하기 표 1에 제시되어 있다.

표 1

서열 번호	명칭	서열
5	crRNA-sp2	5'ATAACTCAATTGTAAAAAGTTTAGAGCTATGCT GTTTTG3'
6	tracrRNA	5'GGAACCATTCAAAACAGCATAGCAAGTAAATAA GGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCTTTTTT3'

[0056]

본원에 기술된 gRNA 라이브러리는 시험관내 스크리닝 및 분석을 위해 사용될 수 있다. 상기 배열을 예시한 것은 도 7에 도시되어 있으며, 여기서, 표적 이중 가닥 DNA 서열은 gRNA 서열 및 Cas9 효소와 함께 인큐베이션된다. 혼합 결과, 이중 가닥 DNA은 파단된다. DNA 파단으로 게놈 요소의 기능 또는 발현은 측정 가능한 정도로 변화가 일어날 수 있다. 본원에 기술된 gRNA, 또는 gRNA를 코딩하는 DNA는 제한 없이, 형질감염, 형질도입, 또는 전기 천공을 비롯한, 당업계에 공지된 각종 방법을 통해 세포에 부가될 수 있다.

[0058]

일부 경우에서, 본원에 기술된 gRNA 라이브러리는 생체내 또는 생체외 스크리닝 및 분석을 위해 사용된다. 스크리닝을 위해 세포로는 살아있는 피험체로부터 채취된 1차 세포 또는 세포주를 포함한다. 세포는 원핵생물(예컨대, 박테리아 및 진균) 또는 진핵생물(예컨대, 동물 및 식물)로부터의 것일 수 있다. 예시적인 동물 세포로는 제한 없이, 마우스, 토끼, 영장류, 및 곤충으로부터의 것을 포함한다. 일부 경우에서, 본원에 기술된 gRNA 라이브러리는 또한 다세포 유기체로 전달될 수 있다. 예시적인 다세포 유기체로는 제한 없이, 식물, 마우스, 토끼, 영장류, 및 곤충을 포함한다.

[0059]

게놈 조작

[0060]

본원에서는 특정 표적 핵산 서열의 뉴클레아제 표적화를 위한 핵산을 포함하는 라이브러리를 제공한다. 일부 경우에서, 본원에 기술된 라이브러리는, 핵산이 DNA, RNA, 그의 임의의 유사체, 또는 유도체인, 합성된 핵산을 포함한다. 일부 경우에서, 표적 핵산 서열은 DNA, RNA, 그의 임의의 유사체, 또는 유도체를 포함한다. 일부 경우에서, 뉴클레아제는 표적 핵산 서열을 절단한다. 일부 경우에서, 뉴클레아제는 표적 핵산에 결합하지만, 그를 절단하지는 않는다. 뉴클레아제의 유형으로는 전사 활성화인자 유사 이펙터 뉴클레아제(TALEN: Transcription Activator-Like Effector Nuclease), 아연 링거 뉴클레아제(ZFN: zinc finger nuclease), 메가뉴클레아제, 아르고노트(Argonaute), 및 주기적으로 간격을 띠고 분포하는 짧은 회문 구조의 반복부(CRISPR) 연관(Cas) 단백질을 포함하나, 이에 제한되지 않는다. 일부 경우에서, 뉴클레아제는 야생형 뉴클레아제, 유전적으로 변형된 뉴클레아제, 또는 재조합 뉴클레아제이다.

[0061]

표적화된 유전자 편집을 위한 모델 시스템은 Cas9 기반 접근법을 포함한다. gRNA와 함께 세포에서 발현되거나, 또는 세포로 전달되었을 때, Cas9는 mRNA의 주기적으로 간격을 띠고 분포하는 짧은 회문 구조의 반복부(CRISPR) 서열과의 복합체를 통해 유전자 정보를 표적화된 방식으로 도입 또는 결실시킬 수 있다. 도 1a-1b에 도시된 바와 같이, Cas9 복합체는 가이드 RNA("gRNA") 전사체와 체결된 Cas9 단백질을 포함한다. 예시된 gRNA는 표적 서열 영역, PAM 영역, 및 헤어핀 영역을 포함한다. CRISPR/Cas9 프로세스에서, gRNA는 Cas9 효소를 DNA의 특이적인 스트레치로 인도한다. 도시된 gRNA가 sgRNA(단일 가닥 가이드 RNA)인 경우, 복합체는 dgRNA(이중 가닥 가이

드 RNA)와 형성될 수 있다. 이어서, Cas9는 DNA를 절단하여 유전자를 불능화시키거나, 또는 수복시킨다. 상기 프로세스에 대한 예시적인 변형에 관한 비제한적인 목록이 본원에 기술되어 있다. CRISPR/dCas9 프로세스에서, 불능화된 또는 "데드" Cas9("dCas9": "dead" Cas9)는 더 이상 스플라이싱 기능을 발휘하지는 못하지만, 또 다른 효소 활성 부가시에는 상이한 표적 분자 변형 기능을 수행한다. 예를 들어, 시티딘 데아미나제를 dCas9에 테더링하면, C-G DNA 염기쌍은 T-A 염기쌍으로 전환된다. 대안적 dCas9 프로세스에서, dCas9에 테더링된 상이한 효소는 표적 DNA에서 염기 C를 T로, 또는 G를 A로 변이시킨다. 대안적으로, dCas9 프로세스는 RNA 폴리머라제 활성을 차단 또는 활성화시키는 전사 인자의 융합에 의해 변형될 수 있고, 이로써, 유전자 전사는 작동이 중단(CRISPRi) 또는 작동이 개시(CRISPRa)될 수 있고, 그에 따라 유전자 발현이 조절될 수 있다. 예를 들어, dCas9 프로세스는 전사 억제인자와의 융합에 의해 변형된다. 일부 경우에서, dCas9 프로세스는 전사 활성화인자와의 융합에 의해 변형된다. 일부 경우에서, dCas9 프로세스는 복수의 전사 억제인자 또는 전사 활성화인자와의 융합에 의해 변형된다. 대안적 방식에서, gRNA는 절단을 위한 다중 부위를 가지며, 이로써, 유전자 편집을 위한 다중 영역을 가지는 gRNA가 생성된다. Cas9n, 또는 "넉킹 Cas9"의 경우, RuvC 또는 HNH 절단 도메인은 불활성 상태로 변형된다. 이러한 불활성화를 통해 Cas9는 유일하게, 이중 가닥 파단이 아닌, DNA에 오직 단 하나의 가닥 파단(넉)만을 생성하게 된다. 일부 방식에서, 각 가닥마다 하나씩 2개의 Cas9n 효소가 이중 가닥 파단을 생성하는 데 사용된다. 이는 절단 부위의 상류 및 하류 영역, 둘 모두를 인식할 수 있기 때문에, 오프 타겟 효과는 제거된다. hfCas9의 경우, 오프 타겟 효과가 없는 Cas9 절단을 생성하기 위해 이중 Cas9n 단백질을 사용하는 대신, 변형된 Cas9 효소는 효소 활성 이전에 덜 완벽한 매치도 허용하도록 완화된 결합 표적 특이성을 염격성을 가진다. 일부 경우에서, dCas9 프로세스는 표적 핵산을 검출하기 위한 표지 또는 태그와의 융합으로 변형된다. 예를 들어, 표지는 표적 핵산을 검출하기 위한 형광 마커(예컨대, GFP)이다. 일부 경우에서, dCas9는 에피토프 태그에 융합되고, gRNA에 의해 특정되는 표적 핵산의 정제를 위해 사용된다.

[0062]

본원에서는 뉴클레아제를 특정 표적 핵산 서열로 지시하기 위한 핵산을 포함하는 라이브러리를 제공한다. 일부 경우에서, 표적 핵산 서열은 DNA를 포함한다. 일부 경우에서, 표적 핵산 서열은 RNA를 포함한다. 예를 들어, RNA 서열을 표적화하기 위해, C2c2를 지시하기 위한 핵산을 포함하는 라이브러리를 생성한다. 일부 경우에서, DNA 또는 RNA는 단일 가닥 또는 이중 가닥이다.

[0063]

본원에서는 특정 표적 핵산 서열의 뉴클레아제 표적화를 위한 핵산을 포함하는 라이브러리로서, 여기서, 뉴클레아제는 제한하는 것은 아니지만, 스트렙토코쿠스(*Streptococcus*), 캄필로박터(*Campylobacter*), 니트라티프락토르(*Nitratirfractor*), 스타필로코쿠스(*Staphylococcus*), 파르비바콜룸(*Parvibaculum*), 로세부리아(*Roseburia*), 네이세리아(*Neisseria*), 글루코나세토박터(*Gluconacetobacter*), 아조스피리움(*Azospirillum*), 스파에로카에타(*Sphaerochaeta*), 락토바실러스(*Lactobacillus*), 유박테리움(*Eubacterium*), 코리네박터(*Corynebacter*), 카르노박테리움(*Carnobacterium*), 로도박터(*Rhodobacter*), 리스테리아(*Listeria*), 팔루디박터(*Paludibacter*), 클로스트리디움 종(*Clostridium*), 라크노스피라세아에(*Lachnospiraceae*), 클로스트리디아리디움(*Clostridiariidium*), 렙토트리키아(*Leptotrichia*), 프란시셀라(*Francisella*), 레지오넬라(*Legionella*), 알리사이클로바실러스(*Alicyclobacillus*), 메타노메티오플러스(*Methanomethyophilus*), 포르피로모나스(*Porphyromonas*), 프레보텔라(*Prevotella*), 박테로이데테스(*Bacteroidetes*), 헬코코쿠스(*Helcoccoccus*), 레토스피라(*Letospira*), 데술포비브리오(*Desulfovibrio*), 데술포나트로눔(*Desulfonatronum*), 데술푸로코쿠스(*Desulfurococcus*), 오피투타세아에(*Opitutaceae*), 투베리바실러스(*Tuberibacillus*), 바실러스(*Bacillus*), 브레비바실러스(*Brevibacillus*), 메틸로박테리움(*Methylbacterium*), 나트로노박테리움(*Natronobacterium*), 플라보박테리움(*Flavobacterium*), 사카로마이세스(*Saccharomyces*), 클라미도모나스(*Chlamydomonas*), 씨무스(*Thermus*), 피로코쿠스(*Pyrococcus*), 미코플라스마(*Mycoplasma*), 또는 아시다미노코쿠스(*Acidaminococcus*) 종으로부터의 것인, 라이브러리를 제공한다. 예시적인 뉴클레아제는 하기 표 2a에 열거되어 있다. 본원에 기술된 gRNA는 상기 열거된 종들 중 임의의 것으로부터의 뉴클레아제, 또는 효소가 게놈 편집 기능을 허용하는 것인 추가의 종으로부터의 뉴클레아제의 터미네이터 서열에 결합할 수 있다. 예시적인 터미네이터 서열은 제한 없이, 하기 표 2b에 열거된 것을 포함한다. 예시적인 PAM 서열은 제한 없이, 하기 표 2c에 열거된 것을 포함한다.

표 2

표 2a.

명칭	수탁 번호	종
Cas9	Q99ZW2	스트렙토코쿠스 피오게네스 (<i>Streptococcus pyogenes</i>)
Cas9	J7RUA5	스타필로코쿠스 아우레우스 (<i>Staphylococcus aureus</i>)
Cas9	G3ECR1	스트렙토코쿠스 써모필러스 (<i>Streptococcus thermophilus</i>)
C2c2	P0DOC6	렙토트리키아 샤히이 (<i>Leptotrichia shahii</i>)
Cpf1	U2UMQ6	아시다미노코쿠스 종
C2c1	T0D7A2	알리사이클로바실러스 아시도테레스트리스 (<i>Alicyclobacterium acidoterrestris</i>)
FokI	P14870	플라보박테리움 오케아노코이테스 (<i>Flavobacterium okeanokoites</i>)
AciI	A0A0C5GQT3	락토바실러스 아시도필러스 (<i>Lactobacillus acidophilus</i>)
I-SceI	P03882	사카로마이세스 세레비시아에
I-CreI	P05725	클라미도모나스 레인하드타이
I-DmoI	1B24_A	데술푸로코쿠스 모빌리스 (<i>Desulfovibacterium mobilis</i>)
TtAgo	Q746M7	써무스 써모필러스 (<i>Thermus thermophilus</i>)
PfAgo	Q8U3D2	파로코쿠스 푸리오수스 (<i>Pyrococcus furiosus</i>)
NgAgo	A0A172MAH6	나트로노박테리움 그레고리이 (<i>Natronobacterium gregoryi</i>)

[0064]

표 2b

서열 번호	종	서열
7	스트렙토코쿠스 써모필러스	5' TACTCAACTTGAAAAGGTGGCACCGATTGGTGT 3'
8	스트렙토코쿠스 뮤탄스 (<i>Streptococcus mutans</i>)	5' TACACAACTTGAAAAAGTGCACCGATTGGTGCT 3'
9	리스테리아 인노쿠아 (<i>Listeria innocua</i>)	5' TTATCAACTTTAATTAAGTAGCGCTGTTGGCGCT 3'
10	미코플라스마 모빌레 (<i>Mycoplasma mobile</i>)	5' TATGCCGTAACTACTACTTATTTCAAAATAAGTAGTT 3'
11	캄필로박ter 제주니 (<i>Campylobacter jejuni</i>)	5' GACTCTGGGGTTACAATCCCCTAAACCGCT 3'

표 2c

종/Cas9의 변이체	PAM 서열
스트렙토코쿠스 피오게네스/ SpCas9	NGG
스트렙토코쿠스 피오게네스/SpCas9 D1135E 변이체	NGG
스트렙토코쿠스 피오게네스/SpCas9 VRER 변이체	NGCG
스트렙토코쿠스 피오게네스 /SpCas9 EQR 변이체	NGAG
스트렙토코쿠스 피오게네스 /SpCas9 VQR 변이체	NGAN 또는 NGNG
스타필로코쿠스 아우레우스/SaCas9	NNGRRT 또는 NNGRR(N)
네이세리아 메닌기티디스 (<i>Neisseria meningitidis</i>)	NNNNGATT
스트렙토코쿠스 써모필러스	NNAGAAW
트레포네마 덴티콜라 (<i>Treponema denticola</i>)	NAAAAC

[0065]

본원에서는 하나 이상의 뉴클레아제(들)를 특정 핵산 서열로 표적화하기 위한 핵산을 포함하는 라이브러리를 제공한다. 일부 경우에서, 뉴클레아제는 TALEN, ZFN, 메가뉴클레아제, 아르고노트, 및 Cas 단백질 중 적어도 하나이다. 예를 들어, 1개 초과의 뉴클레아제가 큰 계획 결실을 생성하기 위해, 한 번에 다중 서열을 변형시키기 위해 다중화될 수 있거나, 또는 예컨대, 니카제와 같은 다른 효소와 함께 사용될 수 있다. 일부 경우에서, 뉴클레아제의 개수는 표적 핵산 서열에 대하여 적어도 2개의 뉴클레아제이다. 일부 경우에서, 뉴클레아제의 개수는 표적 핵산 서열에 대하여 약 2 내지 3개, 2 내지 4개, 2 내지 5개, 2 내지 6개, 2 내지 7개, 2 내지 8개, 2 내지 9개, 또는 2 내지 10개 범위의 뉴클레아제이다.

[0067]

본원에서는 특정 핵산 서열의 니카제 표적화를 위한 합성된 핵산을 포함하는 라이브러리를 제공한다. 니카제는 핵산 서열에 단일 가닥 파단을 생성하는 효소이다. 일부 경우에서, 합성된 핵산은 DNA, RNA, 그의 임의의 유사체, 또는 유도체이다. 일부 경우에서, 특정 핵산 서열은 DNA, RNA, 그의 임의의 유사체, 또는 유도체를 포함한다. 일부 경우에서, 니카제는 특정 핵산 서열을 절단한다. 일부 경우에서, 니카제는 특정 핵산에 결합하지만, 그를 절단하지는 않는다. 일부 경우에서, 니카제는, 뉴클레아제가 TALEN, ZFN, 메가뉴클레아제, 아르고노트, 또는 Cas 단백질인, 변경된 뉴클레아제이다. 일부 경우에서, 니카제는 TALEN, ZFN, 메가뉴클레아제, 아르고노트, 또는 Cas 단백질의 뉴클레아제 도메인을 변경시킴으로써 생성된다. 일부 경우에서, 니카제는 Cas9의 뉴클레아제 도메인을 변경시킴으로써 생성된다.

- [0068] 일부 경우에서, 라이브러리는 특정 핵산 서열의 하나 이상의 니카제(들) 표적화를 위한 핵산을 포함한다. 일부 경우에서, 니카제의 개수는 특정 핵산 서열에 대하여 적어도 2개의 니카제이다. 일부 경우에서, 니카제의 개수는 특정 핵산 서열에 대하여 약 2 내지 3개, 2 내지 4개, 2 내지 5개, 2 내지 6개, 2 내지 7개, 2 내지 8개, 2 내지 9개, 또는 2 내지 10개 범위의 니카제이다. 일부 경우에서, 라이브러리는 하나 이상의 니카제 및 하나 이상의 뉴클레아제를 특정 핵산 서열으로 지시하기 위한 핵산을 포함한다.
- [0069] 본원에서 제공되는, 뉴클레아제를 특정 핵산 서열에 표적화시키기 위한 핵산을 포함하는 라이브러리는 특정 핵산 서열의 절단을 초래할 수 있다. 일부 경우에서, 뉴클레아제는 TALEN, ZFN, 메가뉴클레아제, 아르고노트, 및 Cas 단백질 중 적어도 하나이다. 일부 경우에서, 뉴클레아제는 절단 이외에도 특정 핵산 서열을 변형시키는 키메라 뉴클레아제이다. 예를 들어, 키메라 뉴클레아제는 메틸화, 탈메틸화, 폴리아데닐화, 탈아데닐화, 탈아민화, 또는 폴리우리디닐화를 초래한다.
- [0070] 전사 활성화인자 유사 이펙터 뉴클레아제
- [0071] 본원에서는 특정 핵산 서열의 전사 활성화인자 유사 이펙터 뉴클레아제(TALEN) 표적화를 위한 핵산을 포함하는 핵산 라이브러리를 합성하는 방법을 제공한다. TALEN은 특이적 표적 서열에서 이중 가닥 파단을 유도하는 테사용될 수 있는 조작된 서열 특이적 뉴클레아제 부류이다. TALEN은 전사 활성화인자 유사(TAL) 이펙터 DNA 결합 도메인, 또는 그의 기능성 일부를 뉴클레아제의 촉매 도메인에 융합시킴으로써 생성될 수 있다. TAL 이펙터 DNA 결합 도메인은, 각각이 반복 가변성 이중 잔기(RVD: repeat variable diresidue)로 공지된 고도의 가변성 12번째 및 13번째 아미노산을 포함하는 것인, 일반적으로 고도로 보존되는 33 또는 34개의 아미노산 서열 세그먼트인 TAL 반복부 시리즈를 포함한다. 각각의 RVD는 특이적 뉴클레오티드를 인식하고, 그에 결합할 수 있다. 따라서, TAL 이펙터 DNA 결합 도메인은 적절한 RVD를 포함하는 TAL 반복부를 조합함으로써 뉴클레오티드의 특이적 서열을 인식하도록 조작될 수 있다.
- [0072] 본원에서는 TAL 이펙터 DNA 결합 도메인을 코딩하는 동일하지 않은 핵산을 포함하는 핵산 라이브러리를 합성하는 방법을 제공한다. 일부 경우에서, TAL 이펙터 DNA 결합 도메인은 특정 표적 핵산 서열을 인식하고, 특정 부위에서 이중 가닥 파단을 유도하도록 디자인된다. 일부 경우에서, TAL 이펙터 DNA 결합 도메인은 특정 핵산 서열을 인식하고, 그에 결합하도록 디자인된 TAL 반복부를 다수 포함한다. 일부 경우에서, TAL 반복부의 개수는 적어도 또는 약 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35개 이상의 TAL 반복부이다.
- [0073] 일부 경우에서, TAL 이펙터 DNA 결합 도메인을 코딩하는 동일하지 않은 핵산을 포함하는 핵산 라이브러리가 합성된다. 일부 경우에서, 본원에 기술된 핵산 라이브러리는 번역되었을 때, 단백질 라이브러리를 코딩한다. 일부 경우에서, 핵산 라이브러리는 세포에서 발현되고, 단백질 라이브러리가 생성된다. 일부 경우에서, 합성된 핵산 라이브러리가 발현 백터 내로 삽입된다. 일부 경우에서, 합성된 핵산 라이브러리가 발현 백터 내로 삽입되고, 세포에서 발현된다.
- [0074] 본원에 기술된 방법에 의해 생성된 TAL 이펙터 DNA 결합 도메인을 코딩하는 핵산을 포함하는 핵산 라이브러리는 TALEN을 생성하는 데 사용될 수 있다. 일부 경우에서, 이는 백터에서 클로닝 및 발현된 TAL 이펙터 결합 도메인 라이브러리를 뉴클레아제와 혼합함으로써 달성된다. 예시적인 뉴클레아제로는 *Aci*I, *Acu*I, *Alw*I, *Bbv*I, *Bcc*I, *Bce*AI, *Bci*VI, *Bfu*AI, *Bmg*BI, *Bmr*I, *Bpm*I, *Bpu*EI, *Bsa*I, *Bsm*AI, *Bsm*FI, *Bse*RI, *Bsp*CNI, *Bsr*I, *Bsg*I, *Bsm*I, *Bsp*MI, *Bsr*BI, *Bsr*DI, *Btg*ZI, *Bts*I, *Bts*CI, *Ear*I, *Eci*I, *Fok*I, *Hga*I, *Hph*I, *Hpy*AV, *Mbo*II, *Mly*I, *Mme*I, *Mnl*I, *Nme*AlII, *Ple*I, *Sfa*NI, *Bbv*CI, *Bpu*10I, *Bsp*QI, *Sap*I, *Bae*I, *Bsa*XI, 또는 *Csp*CI를 포함하나, 이에 제한되지 않는다. 일부 경우에서, 혼합은 라이제이션에 의해 이루어진다. 예시적인 리가제로는 *E. coli* 리가제, T4 리가제, 포유동물 리가제(예컨대, DNA 리가제 I, DNA 리가제 II, DNA 리가제 III, DNA 리가제 IV), 열안정성 리가제, 및 패스트 리가제를 포함하나, 이에 제한되지 않는다. 본원에 기술된 방법에 의해 생성된 TALEN은 발현 백터 내로 삽입될 수 있다. 일부 경우에서, TALEN은 발현 백터 내로 삽입되고, 세포에서 발현된다.
- [0075] 본원에서는 원핵 또는 진핵 유기체의 계놈에서 유전자에 대하여 동일하지 않은 핵산 서열을 포함하는 TAL 이펙터 DNA 결합 도메인 라이브러리를 합성하는 방법을 제공한다. 일부 경우에서, TAL 이펙터 DNA 결합 도메인 라이브러리는 계놈 중 적어도 5%에 대한 유전자에 대하여 적어도 3, 4, 5, 6, 7, 8, 9, 10개, 또는 10개 초과의 동일하지 않은 핵산 서열을 포함한다. 일부 경우에서, TAL 이펙터 DNA 결합 도메인 라이브러리는 계놈 중 적어도 5%에 대한 하나 이상의 유전자에 대하여 동일하지 않은 핵산 서열을 포함한다. 일부 경우에서, TAL 이펙터 DNA 결합 도메인 라이브러리는 계놈 중 적어도 5%에 대한 약 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100개, 또는 100개 초과의 유전자에 대하여 동일하지 않은 핵산 서열을 포함한다. 일부 경우에서, TAL 이펙터 DNA 결합

도메인 라이브러리는 계놈 중 적어도 5%에 대한 약 1-100개, 2-95개, 5-90개, 10-85개, 15-80개, 20-75개, 25-70개, 30-65개, 35-60개, 또는 40-50개의 유전자에 대하여 동일하지 않은 핵산 서열을 포함한다.

[0076] 아연 평거 뉴클레아제

본원에서는 특정 핵산 서열의 아연 평거 뉴클레아제(ZFN) 표적화를 위한 핵산을 포함하는 핵산 라이브러리를 합성하는 방법을 제공한다. ZFN은 DNA 결합 아연 평거 도메인(ZFD: zinc finger domain)과 뉴클레아제의 융합에 의해 생성될 수 있다. ZFD는 하나 이상의 아연 평거를 통해 표적 핵산 서열에 결합할 수 있다. 일부 경우에서, ZFD는 적어도 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 아연 평거를 포함한다. 일부 경우에서, ZFD는 최대 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 아연 평거를 포함한다. 일부 경우에서, ZFD는 특정 표적 핵산 서열을 인식하고, 특정 부위에서 이중 가닥 파단을 유도하도록 디자인된다.

[0078] 본원에서는 전사 및 번역되었을 때, ZFD를 코딩하는 핵산을 포함하는 핵산 라이브러리를 합성하는 방법을 제공한다. 일부 경우에서, 핵산 라이브러리는 번역시, 단백질 라이브러리를 코딩한다. 일부 경우에서, 핵산 라이브러리는 세포에서 발현되고, 단백질 라이브러리가 생성된다. 일부 경우에서, 합성된 핵산 라이브러리가 발현 벡터 내로 삽입된다. 일부 경우에서, 합성된 핵산 라이브러리가 발현 벡터 내로 삽입되고, 세포에서 발현된다.

[0079] 본원에 기술된 방법에 의해 생성된 ZFD를 코딩하는 핵산을 포함하는 핵산 라이브러리는 ZFN을 생성하는 데 사용될 수 있다. 일부 경우에서, 이는 벡터에서 클로닝 및 발현된 ZFD를 뉴클레아제와 혼합함으로써 달성된다. 예시적인 뉴클레아제로는 *AcI*I, *Acu*I, *Alw*I, *Bbv*I, *Bcc*I, *Bce*AI, *Bci*VI, *Bfu*AI, *Bmg*BI, *Bmr*I, *Bpm*I, *Bpu*EI, *Bsa*I, *Bsm*AI, *Bsm*FI, *Bse*RI, *Bsp*CNI, *Bsr*I, *Bsg*I, *Bsm*I, *Bsp*MI, *Bsr*BI, *Bsr*DI, *Btg*ZI, *Bts*I, *Bts*CI, *Ear*I, *Eci*I, *Fok*I, *Hga*I, *Hph*I, *Hpy*AV, *Mbo*II, *Mly*I, *Mme*I, *Mn*II, *Nme*AlII, *Ple*I, *Sfa*NI, *Bbv*CI, *Bpu*10I, *Bsp*QI, *Sap*I, *Bae*I, *Bsa*XI, 또는 *Csp*CI를 포함하나, 이에 제한되지 않는다. 일부 경우에서, 혼합은 라이게이션에 의해 이루어진다. 예시적인 리가제로는 *E. coli* 리가제, *T4* 리가제, 포유동물 리가제 (예컨대, DNA 리가제 I, DNA 리가제 II, DNA 리가제 III, DNA 리가제 IV), 열안정성 리가제, 및 패스트 리가제를 포함하나, 이에 제한되지 않는다. 본원에 기술된 방법에 의해 생성된 ZFN은 발현 벡터 내로 삽입될 수 있다. 일부 경우에서, ZFN은 발현 벡터 내로 삽입되고, 세포에서 발현된다.

[0080] 본원에서는 원핵 또는 진핵 유기체의 계놈에서 유전자에 대하여 동일하지 않은 핵산 서열을 포함하는 ZFD 라이브러리를 합성하는 방법을 제공한다. 일부 경우에서, ZFD 라이브러리는 계놈 중 적어도 5%에 대한 유전자에 대하여 적어도 3, 4, 5, 6, 7, 8, 9, 10개, 또는 10개 초과의 동일하지 않은 핵산 서열을 포함한다. 일부 경우에서, ZFD 라이브러리는 계놈 중 적어도 5%에 대한 하나 이상의 유전자에 대하여 동일하지 않은 핵산 서열을 포함한다. 일부 경우에서, ZFD 라이브러리는 계놈 중 적어도 5%에 대한 약 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100개, 또는 100개 초과의 유전자에 대하여 동일하지 않은 핵산 서열을 포함한다. 일부 경우에서, ZFD 라이브러리는 계놈 중 적어도 5%에 대한 약 1-100개, 2-95개, 5-90개, 10-85개, 15-80개, 20-75개, 25-70개, 30-65개, 35-60개, 또는 40-50개의 유전자에 대하여 동일하지 않은 핵산 서열을 포함한다.

[0081] 메가뉴클레아제

[0082] 본원에서는 특정 핵산 서열의 메가뉴클레아제 표적화를 위한 핵산을 포함하는 핵산 라이브러리를 합성하는 방법을 제공한다. 메가뉴클레아제는 염기쌍(예컨대, 12-40개의 염기쌍) DNA 표적을 인식하고, 절단할 수 있는 효소이다. 일부 경우에서, 메가뉴클레아제는 표적 핵산 서열에 대한 특이성을 부여하는 다른 효소의 도메인을 포함하도록 조작된다. 예를 들어, 메가뉴클레아제는 TAL 이펙터 DNA 결합 도메인을 포함하도록 조작된다.

[0083] 본원에서는 전사 및 번역되었을 때, 메가뉴클레아제와 함께 사용하기 위한 결합 도메인을 코딩하는 핵산을 포함하는 핵산 라이브러리를 합성하는 방법을 제공한다. 일부 경우에서, 핵산 라이브러리는 번역시, 단백질 라이브러리를 코딩한다. 일부 경우에서, 핵산 라이브러리는 세포에서 발현되고, 단백질 라이브러리가 생성된다. 일부 경우에서, 합성된 핵산 라이브러리가 발현 벡터 내로 삽입된다. 일부 경우에서, 합성된 핵산 라이브러리가 발현 벡터 내로 삽입되고, 세포에서 발현된다.

[0084] 본원에 기술된 방법에 의해 생성된 도메인을 코딩하는 핵산을 포함하는 핵산 라이브러리는 특정 핵산 서열을 표적화하기 위한 메가뉴클레아제를 조작하는 데 사용될 수 있다. 일부 경우에서, 이는 벡터에서 클로닝 및 발현된 결합 도메인 라이브러리, 예컨대, TAL 이펙터 결합 도메인 라이브러리를 메가뉴클레아제와 혼합함으로써 달성된다. 본원에서는 제공된 방법과 함께 사용하기 위한 메가뉴클레아제의 예로는 *I-Sce*1, *I-Sce*11, *I-Sce*III, *I-Sce*IV, *I-Sce*V, *I-Sce*VI, *I-Sce*VII, *I-Ceu*1, *I-Ceu*AIIP, *I-Cre*1, *I-Crepsb*IP, *I-Crepsb*IIIP, *I-Crepsb*IIIP, *I-Crepsb*IVP, *I-Tli*1, *I-Ppol*, *PI-Psp*I, *F-Sce*1, *F-Sce*11, *F-Suv*1, *F-Tev*I, *F-Tev*II, *I-Amal*, *I-Anil*, *I-Chul*,

I-Cmoel, I-Cpa1, I-CpaII, I-Csm1, I-Cvul, I-CvuAIP, I-Ddil, I-DdiII, I-Dir1, I-Dmol, I-Hmul, I-HmuII, I-HsNIP, I-L1al, I-Msol, I-Naal, I-Nanl, I-NcIIP, I-NgrIP, I-Nit1, I-Njal, I-Nsp236IP, I-Pak1, I-PboIP, I-PcuIP, I-PcuAI, I-PcuVI, I-PgrIP, I-PobIP, I-PorIIP, I-PbpIP, I-SpBetaIP, I-Scal, I-SexIP, I-SneIP, I-Spom1, I-SpomCP, I-SpomIP, I-SpomIIP, I-SquIP, I-Ssp6803I, I-SthPhiJP, I-SthPhiST3P, I-SthPhiSTe3bP, I-TdeIP, I-Tev1, I-TevII, I-TevIII, I-UarAP, I-UarHGPAIP, I-UarHGPA13P, I-VinIP, I-ZbiIP, PI-MtuI, PI-MtuHIP PI-MtuHIIIP, PI-PfuI, PI-PfuII, PI-PkoI, PI-PkoII, PI-Rma43812IP, PI-SpBetaIP, PI-SceI, PI-TfuI, PI-TfuII, PI-Thyl, PI-TliI, PI-TliII, 또는 그의 단편을 포함하나, 이에 제한되지 않는다. 일부 경우에서, 혼합은 라이게이션에 의해 이루어진다. 예시적인 리가제로는 E. 콜라이 리가제, T4 리가제, 포유동물 리가제 (예컨대, DNA 리가제 I, DNA 리가제 II, DNA 리가제 III, DNA 리가제 IV), 열안정성 리가제, 및 패스트 리가제를 포함하나, 이에 제한되지 않는다. 본원에 기술된 방법에 의해 생성된 조작된 메가뉴클레아제는 발현 벡터 내로 삽입될 수 있다. 일부 경우에서, 조작된 메가뉴클레아제는 발현 벡터 내로 삽입되고, 세포에서 발현된다.

[0085]

본원에서는 원핵 또는 진핵 유기체의 게놈에서 유전자에 대하여 동일하지 않은 핵산 서열을 포함하는 메가뉴클레아제와 함께 사용하기 위한 결합 도메인 라이브러리를 합성하는 방법을 제공한다. 일부 경우에서, 도메인 라이브러리는 게놈 중 적어도 5%에 대한 유전자에 대하여 적어도 3, 4, 5, 6, 7, 8, 9, 10개, 또는 10개 초과의 동일하지 않은 핵산 서열을 포함한다. 일부 경우에서, 도메인 라이브러리는 게놈 중 적어도 5%에 대한 하나 이상의 유전자에 대하여 동일하지 않은 핵산 서열을 포함한다. 일부 경우에서, 도메인 라이브러리는 게놈 중 적어도 5%에 대한 약 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100개, 또는 100개 초과의 유전자에 대하여 동일하지 않은 핵산 서열을 포함한다. 일부 경우에서, 도메인 라이브러리는 게놈 중 적어도 5%에 대한 약 1-100개, 2-95개, 5-90개, 10-85개, 15-80개, 20-75개, 25-70개, 30-65개, 35-60개, 또는 40-50개의 유전자에 대하여 동일하지 않은 핵산 서열을 포함한다.

[0086]

아르고노트

[0087]

본원에서는 특정 핵산 서열의 아르고노트 표적화를 위한 핵산을 포함하는 핵산 라이브러리를 합성하는 방법을 제공한다. 아르고노트는 RNA 또는 DNA 가이드된 뉴클레아제 패밀리이다. 일부 경우에서, 아르고노트는 가이드 핵산을 이용하여 표적 핵산을 확인한다. 일부 경우에서, 가이드 핵산은 단일 가이드 RNA(sgRNA)이다. 일부 경우에서, 가이드 핵산은 가이드 DNA(gDNA)이다. 예시적인 아르고노트는 TtAgo, PfAgo, 및 NgAgo를 포함하나, 이에 제한되지 않는다. 일부 실시양태에서, 아르고노트는 NgAgo이다.

[0088]

본원에서는 원핵 또는 진핵 유기체의 게놈에서 유전자에 대하여 동일하지 않은 핵산 서열을 포함하는 가이드 핵산 라이브러리를 합성하는 방법을 제공한다. 일부 경우에서, 가이드 핵산 라이브러리는 sgRNA 라이브러리이다. 일부 경우에서, 가이드 핵산 라이브러리는 dgRNA 라이브러리이다. 일부 경우에서, 가이드 핵산 라이브러리는 게놈 중 적어도 5%에 대한 유전자에 대하여 적어도 3, 4, 5, 6, 7, 8, 9, 10개, 또는 10개 초과의 동일하지 않은 핵산 서열을 포함한다. 일부 경우에서, 가이드 핵산 라이브러리는 게놈 중 적어도 5%에 대한 하나 이상의 유전자에 대하여 동일하지 않은 핵산 서열을 포함한다. 일부 경우에서, 가이드 핵산 라이브러리는 게놈 중 적어도 5%에 대한 약 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100개, 또는 100개 초과의 유전자에 대하여 동일하지 않은 핵산 서열을 포함한다. 일부 경우에서, 가이드 핵산 라이브러리는 게놈 중 적어도 5%에 대한 약 1-100개, 2-95개, 5-90개, 10-85개, 15-80개, 20-75개, 25-70개, 30-65개, 35-60개, 또는 40-50개의 유전자에 대하여 동일하지 않은 핵산 서열을 포함한다.

[0089]

CRISPR 연관 단백질

[0090]

본원에서는 특정 핵산 서열의 CRISPR 연관(Cas) 단백질 표적화를 위한 gRNA를 코딩하는 핵산을 포함하는 핵산 라이브러리를 합성하는 방법을 제공한다. 일부 경우에서, Cas 단백질은 Cpf1, C2c1, C2c2, Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9(Csn1 또는 Cs12), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, 그의 동족체, 및 그의 변형된 버전 중 적어도 하나이다. 일부 경우에서, Cas 단백질은 Cas9이다.

[0091]

본원에서는 원핵 또는 진핵 유기체의 게놈에서 유전자에 대하여 동일하지 않은 핵산 서열을 포함하는 gRNA 라이브러리를 합성하는 방법을 제공한다. 일부 경우에서, gRNA 라이브러리는 게놈 중 적어도 5%에 대한 유전자에 대하여 적어도 3, 4, 5, 6, 7, 8, 9, 10개, 또는 10개 초과의 동일하지 않은 핵산 서열을 포함한다. 일부 경우에서, gRNA 라이브러리는 게놈 중 적어도 5%에 대한 하나 이상의 유전자에 대하여 동일하지 않은 핵산 서열을 포함한다.

함한다. 일부 경우에서, gRNA 라이브러리는 게놈 중 적어도 5%에 대한 약 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100개, 또는 100개 초과의 유전자에 대하여 동일하지 않은 핵산 서열을 포함한다. 일부 경우에서, gRNA 라이브러리는 게놈 중 적어도 5%에 대한 약 1-100개, 2-95개, 5-90개, 10-85개, 15-80개, 20-75개, 25-70개, 30-65개, 35-60개, 또는 40-50개의 유전자에 대하여 동일하지 않은 핵산 서열을 포함한다. gRNA 라이브러리는 sgRNA 또는 dgrNA를 코딩할 수 있다.

[0092] 변이체 라이브러리 합성

본원에서는 완전한 또는 부분적인 유전자 서열을 코딩하는 핵산과 gRNA 및 뉴클레아제, 예컨대, Cas9 효소 또는 Cas9 변이체 효소의 조합에 의해 생성되는 변이체 핵산 라이브러리의 합성 방법을 제공한다. 단편은 집단적으로 유전자의 전체 영역을 이격시킬 수 있다. 일부 경우에서, 라이브러리는 DNA 또는 RNA를 코딩한다. 일부 경우에서, 라이브러리는 단일 유전자를, 또는 최대로는 전체 게놈까지 코딩한다. 예를 들어, 약 20,000개의 유전자를 포함하는 게놈에 대하여 유전자 1개당 5개의 gRNA를 코딩하는 gRNA 라이브러리를 통해 약 100,000개의 gRNA가 생성될 것이다. 상기 라이브러리는 단일 유전자, 유전자의 경로, 또는 단일 게놈 중의 모든 유전자를 선택적으로 침묵화 또는 변형시키는 데 사용될 수 있다. 일부 방식에서, gRNA에는 상동성 서열이 없고, 무작위 말단 연결이 일어난다. 상기 프로세스를 통해 비상동성 말단 연결 ("NHEJ")이 이루어진다. 일부 경우에서, NHEJ 후에, 삽입, 결실, 프레임쉬프트, 또는 단일 염기 스와핑이 일어난다. **도 1b**를 참조한다.

본원에 기술된 합성된 라이브러리는 CRISPR-Cas9 작용에서의 적용을 위해 사용될 수 있으며, 여기서, 생성된 gRNA 서열은 표적 DNA 및 Cas9 효소를 포함하는 세포에서 또는 혼합물에서 표적 DNA 서열의 발현을 파괴시키거나, 또는 표적 DNA 서열의 발현 생성물 서열을 변경시키는 데 사용된다. 일부 실시양태에서, 각 변이체는 번역 동안 상이한 아미노산을 생성하는 코돈을 코딩한다. 하기 **표 3**은 변이체 부위에 대하여 가능한 각 코돈(및 대표적인 아미노산)의 목록을 제공한다.

표 3

코돈 및 아미노산 목록

아미노산	1 문자 코드	3 문자 코드	코돈					
알라닌	A	Ala	GCA	GCC	GCG	GCT		
시스테인	C	Cys	TGC	TGT				
아스파르트산	D	Asp	GAC	GAT				
글루탐산	E	Glu	GAA	GAG				
페닐알라닌	F	Phe	TTC	TTT				
글리신	G	Gly	GGA	GGC	GGG	GGT		
히스티딘	H	His	CAC	CAT				
이소류신	I	Iso	ATA	ATC	ATT			
리신	K	Lys	AAA	AAG				
류신	L	Leu	TTA	TTG	CTA	CTC	CTG	CTT
메티오닌	M	Met	ATG					
아스파라긴	N	Asn	AAC	AAT				
프롤린	P	Pro	CCA	CCC	CCG	CCT		
글루타민	Q	Gln	CAA	CAG				
아르기닌	R	Arg	AGA	AGG	CGA	CGC	CGG	CGT
세린	S	Ser	AGC	AGT	TCA	TCC	TCG	TCT
트레오닌	T	Thr	ACA	ACC	ACG	ACT		
발린	V	Val	GTA	GTC	GTG	GTT		
트립토판	W	Trp	TGG					
티로신	Y	Tyr	TAC	TAT				

[0095]

[0096] 본원에서는 완전한 또는 부분적인 유전자 서열을 코딩하는 핵산과 뉴클레아제의 조합에 의해 생성되는 변이체 핵산 라이브러리의 합성 방법으로서, 여기서, 뉴클레아제는 TALEN, ZFN, 또는 조작된 메가뉴클레아제인 변이체 핵산 라이브러리의 합성 방법을 제공한다. 일부 경우에서, 완전한 또는 부분적인 유전자 서열을 코딩하는 핵산과, 뉴클레아제와 함께, sgRNA와 같은 가이드 핵산의 조합에 의해 생성되는 변이체 핵산 라이브러리의 합성 방법으로서, 여기서, 뉴클레아제는 아르고노트, 또는 Cas 단백질인 변이체 핵산 라이브러리의 합성 방법도 제공한다. 본원에 기술된 합성된 라이브러리는 뉴클레아제 작용에서의 적용을 위해 사용될 수 있으며, 여기서, 생성된 핵산 서열은 표적 DNA 및 Cas9 효소를 포함하는 세포에서 또는 혼합물에서 표적 DNA 서열의 발현을 파괴시키거나, 또는 표적 DNA 서열의 발현 생성물 서열을 변경시키는 데 사용된다. 일부 실시양태에서, 각 변이체는 번역 동안 상이한 아미노산을 생성하는 코돈을 코딩한다.

[0097]

본원에 기술된 바와 같은 변이체 핵산 라이브러리는, 합성된 핵산에 의해 코딩된 후속 단백질 중의 단일 잔기의 복수의 상이한 변이체가 표준 번역 프로세스에 의해 생성되도록 적어도 단일 코돈에서 코딩하는 표적 핵산 서열

을 달리하기 위한 sgRNA 또는 dgRNA를 포함한다. 일부 경우에서, 각 변이체는 단백질 도메인의 상이한 아미노산을 생성하는 코돈을 코딩한다. 예를 들어, 단백질 도메인은 보존되는 도메인 또는 촉매 도메인이다. 일부 실시양태에서, 단백질 도메인은 제한하는 것은 아니지만, 키나제 도메인, ATP 결합 도메인, GTP 결합 도메인, 구아닌 뉴클레오티드 교환 인자(GEF: guanine nucleotide exchange factor) 도메인, GTP아제 활성화 단백질(GAP: GTPase activating protein) 도메인, 하이드롤라제 도메인, 엔도뉴클레아제 도메인, 엑소뉴클레아제 도메인, 프로테아제 도메인, 포스파타제 도메인, 포스포리파제 도메인, 플렉스트린 상동성 도메인, Src 상동성 도메인, 및 유비퀴틴 결합 도메인이다. 일부 경우에서, 변이체 핵산 라이브러리는 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개, 또는 10개 초과의 단백질 도메인에서의 변이를 코딩하는 핵산 서열을 표적화하기 위해 sgRNA 또는 dgRNA를 포함한다.

[0098] 일부 실시양태에서, 변이체는 특정 활성을 가지는 단백질에 대한 아미노산을 코딩한다. 예를 들어, 변이체는 메틸트랜스퍼라제 활성, 테메틸라제 활성, 디스뮤타제 활성, 알킬화 활성, 탈퓨린화 활성, 산화 활성, 피리미딘이량체 형성 활성, 인테그라제 활성, 트랜스포사제 활성, 레콤비나제 활성, 폴리머라제 활성, 리가제 활성, 헬리카제 활성, 포토리아제 활성, 글리코실라제 활성, 아세틸트랜스퍼라제 활성, 데아세틸라제 활성, 키나제 활성, 포스파타제 활성, 유비퀴틴 리가제 활성, 탈유비퀴틴화 활성, 아데닐화 활성, 탈아데닐화 활성, SUMO일화 활성, 탈SUMO일화 활성, 리보실화 활성, 탈리보실화 활성, 미리스토일화 활성, 리모델링화 활성, 프로테아제 활성, 옥시도리덕타제 활성, 트랜스퍼라제 활성, 하이드롤라제 활성, 리아제 활성, 이소머라제 활성, 신타제 활성, 신텐타제 활성, 또는 탈미리스토일화 활성을 포함하는 단백질을 코딩한다.

[0099] 상동성 유도 수복(HDR: homology-directed repair)에 의해 생성된 변이

[0100] 변이체 핵산 라이브러리 생성을 위한 예시적인 프로세스에서, 표적 DNA 라이브러리에서 변종을 생성하기 위해 Cas9 절단 및 상동성 재조합이 도입된다. 먼저, (RNA의 드 노보 합성, 또는 DNA의 드 노보 합성 후 (생체내 또는 시험관내에서) 전사에 의해 gRNA를 생성함으로써) gRNA의 라이브러리가 합성되고, 여기서, 라이브러리는 유전자 1개당 복수의 gRNA 분자를 포함한다. 예를 들어, gRNA 라이브러리는 유전자 1개당 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 gRNA를 포함할 수 있다. gRNA 라이브러리를 Cas9 효소 및 표적 DNA 라이브러리와 혼합하고, 여기서, 표적 DNA 라이브러리는 적어도 하나의 유전자 단편 또는 적어도 하나의 유전자를 코딩하는 핵산 서열을 포함한다. 예를 들어, 표적 DNA 라이브러리는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 유전자 또는 유전자 단편을 포함할 수 있다. 일부 경우에서, 표적 DNA 라이브러리는 경로 중의 다중의 유전자로부터의 서열, 또는 유기체 중의 모든 유전자로부터의 서열을 포함한다. 또한, 변이가 표적 DNA 가닥 내로 도입되도록 하기 위해 상동성 서열 및 변이체 핵산 서열을 포함하는 대체 서열도 혼합물에 첨가된다. 생성된 표적 DNA 라이브러리는 복수의 변이체 DNA 서열을 포함할 것이다. 일부 경우에서, 변이는 표적 DNA 서열 내로 결실, 프레임 쉬프트, 또는 삽입을 도입한다. 일부 경우에서, 변이체 DNA 서열을 통해 유전자 또는 유전자 단편 1개당 적어도 하나의 코돈에 대한 변이가 생성된다. 일부 경우에서, 유전자의 일부가 표적 DNA 내로 삽입되거나, 또는 대안적으로, 표적 DNA 서열의 일부(즉, 유전자 또는 전체 유전자의 단편)이 표적 DNA로부터 제거된다. 일부 경우에서, 변이체 DNA 서열을 통해 유전자 또는 유전자 단편과 연관된 적어도 하나의 전사 조절 서열, 예컨대, 프로모터, UTR, 또는 터미네이터 서열에 대한 변이가 생성된다.

[0101] 변이체 핵산 라이브러리 생성을 위한 일부 경우에서, 표적 DNA 라이브러리에서 변종을 생성하기 위해 뉴클레아제 절단 및 상동성 재조합이 도입되며, 여기서, 뉴클레아제는 TALEN, ZFN, 메가뉴클레아제, Cas, 또는 아르고노트이다. 일부 경우에서, 뉴클레아제가 TALEN인 경우, TAL 이펙터 DNA 결합 도메인의 라이브러리는 (RNA의 드 노보 합성, 또는 DNA의 드 노보 합성 후 (생체내 또는 시험관내에서) 전사 및 번역에 의해) 합성되고, 여기서, 라이브러리는 유전자 1개당 복수의 TAL 이펙터 DNA 결합 도메인 분자를 포함한다. 예를 들어, TAL 이펙터 DNA 결합 도메인 라이브러리는 유전자 1개당 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 TAL 이펙터 DNA 결합 도메인 분자를 포함할 수 있다. 이어서, TAL 이펙터 DNA 결합 도메인 라이브러리는 뉴클레아제 효소와 혼합되고, 이로써, TALEN을 생성할 수 있다. 일부 경우에서, TALEN은 표적 DNA 라이브러리와 조합되고, 여기서, 표적 DNA 라이브러리는 적어도 하나의 유전자 단편 또는 적어도 하나의 유전자를 코딩하는 핵산 서열을 포함한다. 예를 들어, 표적 DNA 라이브러리는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 유전자 또는 유전자 단편을 포함할 수 있다. 일부 경우에서, 표적 DNA 라이브러리는 경로 중의 다중의 유전자로부터의 서열, 또는 유기체 중의 모든 유전자로부터의 서열을 포함한다. 일부 경우에서, 또한, 변이가 표적 DNA 가닥 내로 도입되도록 하기 위해 상동성 서열 및 변이체 핵산 서열을 포함하는 대체 서열도 혼합물에 첨가된다. 생성된 표적 DNA 라이브러리는 복수의 변이체 DNA 서열을 포함할 것이다. 일부 경우에서, 변이는 표적 DNA 서열 내로 결실, 프레임 쉬프트, 또는 삽입을 도입한다. 일부 경우에서, 변이체 DNA 서열을 통해 유전자 또는 유전자 단편 1개당 적어도 하나의 코돈에 대한

변이가 생성된다. 일부 경우에서, 유전자의 일부가 표적 DNA 내로 삽입되거나, 또는 대안적으로, 표적 DNA 서열의 일부(즉, 유전자 또는 전체 유전자의 단편)이 표적 DNA로부터 제거된다. 일부 경우에서, 변이체 DNA 서열을 통해 유전자 또는 유전자 단편과 연관된 적어도 하나의 전사 조절 서열, 예컨대, 프로모터, UTR, 또는 터미네이터 서열에 대한 변이가 생성된다.

[0102] **변형된 Cas9 효소에 의해 생성된 변이**

[0103] 변이체 핵산 라이브러리 생성을 위한 제2 예시적인 프로세스에서, 변이체 표적 DNA 라이브러리를 생성하기 위해 변형된 Cas9 효소가 도입된다. 먼저, (RNA의 드 노보 합성, 또는 DNA의 드 노보 합성 후, 전사에 의해 gRNA를 생성함으로써) gRNA의 라이브러리가 합성되고, 여기서, 라이브러리는 유전자 1개당 복수의 gRNA 분자를 포함한다. 예를 들어, gRNA 라이브러리는 유전자 1개당 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 gRNA를 포함할 수 있다. gRNA 라이브러리를 변형된 Cas9 효소 및 표적 DNA 라이브러리와 혼합하고, 여기서, 표적 DNA 라이브러리는 적어도 하나의 유전자 단편 또는 적어도 하나의 유전자를 코딩하는 핵산 서열을 포함한다. 예를 들어, 표적 DNA 라이브러리는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 유전자 또는 유전자 단편을 포함할 수 있다. 일부 경우에서, 표적 DNA 라이브러리는 경로 중의 다중의 유전자로부터의 서열, 또는 유기체 중의 모든 유전자로부터의 서열을 포함한다. 변형된 Cas9 효소는 핵산 서열 변형 능력이 있는 또 다른 효소를 그에 테더링시킨다. 예시적인 변형된 Cas9 효소는 불능화된 또는 "데드" Cas9("dCas9")가 더 이상 스플라이싱 기능을 발휘하지는 못하지만, 또 다른 효소 활성 부가시에는 상이한 표적 분자 변형 기능을 수행하는 것인 dCas9 프로세스를 포함한다. 예를 들어, 시티딘 테아미나제를 dCas9에 테더링하면, C-G DNA 염기쌍은 T-A 염기쌍으로 전환된다. 대안적 dCas9 프로세스에서, dCas9에 테더링된 상이한 효소는 표적 DNA에서 염기 C를 T로, 또는 G를 A로 변이시킨다. 생성된 표적 DNA 라이브러리는 복수의 변이체 표적 DNA 서열을 포함한다. 일부 경우에서, 변이는 표적 DNA 서열 내로 결실, 프레임 쉬프트, 또는 삽입을 도입한다. 일부 경우에서, 변이체 DNA 서열을 통해 유전자 또는 유전자 단편 1개당 적어도 하나의 코돈에 대한 변이가 생성된다. 일부 경우에서, 변이체 DNA 서열을 통해 유전자 또는 유전자 단편과 연관된 적어도 하나의 전사 조절 서열, 예컨대, 프로모터, UTR, 또는 터미네이터 서열에 대한 변이가 생성된다.

[0104] **변형된 뉴클레아제에 의해 생성된 변이**

[0105] 본원에서는 변이체 표적 DNA 라이브러리를 생성하기 위해 도입된 변형된 뉴클레아제 효소를 포함하는 변이체 핵산 라이브러리 생성 방법을 제공한다. 일부 경우에서, 뉴클레아제는 TALEN이다. 일부 경우에서, TAL 이펙터 DNA 결합 도메인 라이브러리는 (RNA의 드 노보 합성, 또는 DNA의 드 노보 합성 후, 전사 및 번역에 의해 TAL 이펙터 DNA 결합 도메인 라이브러리를 생성함으로써) 합성되고, 여기서, 라이브러리는 유전자 1개당 복수의 동일하지 않은 핵산 서열을 포함한다. 예를 들어, TAL 이펙터 DNA 결합 도메인 라이브러리는 유전자 1개당 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 동일하지 않은 핵산 서열을 포함할 수 있다. 이어서, TAL 이펙터 DNA 결합 도메인 라이브러리를 뉴클레아제 효소와 혼합하여 TALEN을 생성할 수 있다. 일부 경우에서, 이어서, TALEN을 표적 DNA 라이브러리와 혼합하고, 여기서, 표적 DNA 라이브러리는 적어도 하나의 유전자 단편 또는 적어도 하나의 유전자를 코딩하는 핵산 서열을 포함한다. 예를 들어, 표적 DNA 라이브러리는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 유전자 또는 유전자 단편을 포함할 수 있다. 일부 경우에서, 표적 DNA 라이브러리는 경로 중의 다중의 유전자로부터의 서열, 또는 유기체 중의 모든 유전자로부터의 서열을 포함한다.

[0106] 일부 경우에서, 뉴클레아제는 ZFN이다. 일부 경우에서, ZFD 라이브러리는 (RNA의 드 노보 합성, 또는 DNA의 드 노보 합성 후, 전사 및 번역에 의해 ZFD 라이브러리를 생성함으로써) 합성되고, 여기서, 라이브러리는 유전자 1개당 복수의 동일하지 않은 핵산 서열을 포함한다. 예를 들어, ZFD 라이브러리는 유전자 1개당 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 동일하지 않은 핵산 서열을 포함할 수 있다. 이어서, ZFD 라이브러리를 뉴클레아제 효소와 혼합하여 ZFN을 생성할 수 있다. 일부 경우에서, 이어서, ZFN을 표적 DNA 라이브러리와 혼합하고, 여기서, 표적 DNA 라이브러리는 적어도 하나의 유전자 단편 또는 적어도 하나의 유전자를 코딩하는 핵산 서열을 포함한다. 예를 들어, 표적 DNA 라이브러리는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 유전자 또는 유전자 단편을 포함할 수 있다. 일부 경우에서, 표적 DNA 라이브러리는 경로 중의 다중의 유전자로부터의 서열, 또는 유기체 중의 모든 유전자로부터의 서열을 포함한다.

[0107] 일부 경우에서, 뉴클레아제는 메가뉴클레아제이다. 일부 경우에서, 메가뉴클레아제를 특정 핵산 서열에 표적화하기 위한 결합 도메인 라이브러리, 예컨대, TAL 이펙터 DNA 결합 도메인 라이브러리는 (RNA의 드 노보 합성, 또는 DNA의 드 노보 합성 후, 전사에 의해 결합 도메인 라이브러리를 생성함으로써) 합성되고, 여기서, 결합 도메인 라이브러리는 유전자 1개당 복수의 동일하지 않은 핵산 서열을 포함한다. 예를 들어, 결합 도메인 라이브

러리는 유전자 1개당 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 동일하지 않은 핵산 서열을 포함할 수 있다. 이어서, 결합 도메인 라이브러리를 메가뉴클레아제 효소와 혼합하여 조작된 메가뉴클레아제를 생성할 수 있다. 일부 경우에서, 이어서, 조작된 메가뉴클레아제를 표적 DNA 라이브러리와 혼합하고, 여기서, 표적 DNA 라이브러리는 적어도 하나의 유전자 단편 또는 적어도 하나의 유전자를 코딩하는 핵산 서열을 포함한다. 예를 들어, 표적 DNA 라이브러리는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 유전자 또는 유전자 단편을 포함할 수 있다. 일부 경우에서, 표적 DNA 라이브러리는 경로 중의 다중의 유전자로부터의 서열, 또는 유기체 중의 모든 유전자로부터의 서열을 포함한다.

[0108] 일부 경우에서, 뉴클레아제는 아르고노트이다. 일부 경우에서, 가이드 핵산 라이브러리(gRNA 또는 gDNA)는 (RNA의 드 노보 합성, 또는 DNA의 드 노보 합성 후, 전사에 의해 가이드 핵산 라이브러리를 생성함으로써) 합성되고, 여기서, 가이드 핵산 라이브러리는 유전자 1개당 복수의 동일하지 않은 핵산 서열을 포함한다. 예를 들어, 가이드 핵산 라이브러리는 유전자 1개당 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 동일하지 않은 핵산 서열을 포함할 수 있다. 가이드 핵산 라이브러리를 변형된 아르고노트 효소 및 표적 DNA 라이브러리와 혼합하고, 여기서, 표적 DNA 라이브러리는 적어도 하나의 유전자 단편 또는 적어도 하나의 유전자를 코딩하는 핵산 서열을 포함한다. 예를 들어, 표적 DNA 라이브러리는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 유전자 또는 유전자 단편을 포함할 수 있다. 일부 경우에서, 표적 DNA 라이브러리는 경로 중의 다중의 유전자로부터의 서열, 또는 유기체 중의 모든 유전자로부터의 서열을 포함한다.

[0109] 일부 경우에서, 변형된 뉴클레아제 효소는 핵산 서열 변형 능력이 있는 또 다른 효소를 그에 테더링시킨다. 예시적인 변형 능력으로는 메틸화, 탈메틸화, 폴리아데닐화, 탈아데닐화, 탈아민화, 및 폴리우리디닐화를 포함하나, 이에 제한되지 않는다. 일부 경우에서, 복수의 변이체 표적 DNA 서열을 포함하는 표적 DNA 라이브러리를 통해 변이가 생성된다. 일부 경우에서, 변이는 표적 DNA 서열 내로 결실, 프레임 쉬프트, 또는 삽입을 도입한다. 일부 경우에서, 변이체 DNA 서열을 통해 유전자 또는 유전자 단편 1개당 적어도 하나의 코돈에 대한 변이가 생성된다. 일부 경우에서, 변이체 DNA 서열을 통해 유전자 또는 유전자 단편과 연관된 적어도 하나의 전사 조절 서열, 예컨대, 프로모터, UTR, 또는 터미네이터 서열에 대한 변이가 생성된다.

[0110] 모델 시스템의 유전자를 표적화하기 위한 gRNA 라이브러리 합성

[0111] 본원에서는 본원에 기술된 핵산 라이브러리를 이용하여 모델 시스템을 스크리닝하는 방법을 제공한다. 일부 경우에서, 핵산 라이브러리는 본원에 기술된 gRNA 라이브러리이다. 일부 경우에서, 핵산 라이브러리는, 전사될 때, gRNA 서열의 전사가 이루어지는, 본원에 기술된 DNA 라이브러리이다. 모델 유기체의 비제한적인 예시적인 목록은 하기 표 4에서 제공된다.

표 4

유기체 및 유전자 개수

모델 시스템	단백질 코딩 유전자*
아라비돕시스 탈리아나	27000
카에노르합디티스 엘레강스	20000
캐나스 루푸스 패밀리아리스	19000
클라미도모나스 레인하드티이	14000
다니오 레리오	26000
덕티오스텔리움 디스코이데움	13000
드로소필라 멜라노가스터	14000
에스테리키아 콜라이	4300
마카카 물라타	22000
무스 무스쿨루스	20000
오릭토라구스 쿠니쿨루스	27000
라투스 노르베기쿠스	22000
사카로마이세스 세레비시아에	6600
수스 스크로파	21000
호모 사피엔스	21000

* 여기서, 개수는 단백질 코딩 유전자의 개수를 반영하며, tRNA 및 비코딩 RNA는 배제시킨다.
문현 [Ron Milo & Rob Phillips, *Cell Biology by the Numbers* 286 (2015)].

[0112]

[0113]

gRNA의 라이브러리는 (RNA의 드 노보 합성, 또는 DNA의 드 노보 합성 후, 전사에 의해 gRNA를 생성함으로써) 합성되고, 여기서, 라이브러리는 유전자 1개당 복수의 gRNA 분자를 포함한다. 예를 들어, 본원에 기술된 라이브러리는 유전자 1개당 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 gRNA를 포함할 수 있다. 일부 경우에서, 드 노보 합성된 라이브러리 내의 핵산은 단일 유전자당 적어도 또는 약 3개의 동일하지 않은 gRNA에 대한 서열을 코딩한다. 일부 경우에서, 핵산은 단일 유전자당 약 1 내지 약 10개 범위의 동일하지 않은 gRNA에 대한 서열을 코딩한다. 일부 경우에서, 핵산은 단일 유전자당 적어도 또는 약 1개의 동일하지 않은 gRNA에 대한 서열을 코딩한다. 일부 경우에서, 핵산은 단일 유전자당 최대 10개의 동일하지 않은 gRNA에 대한 서열을 코딩한다. 일부 경우에서, 핵산은 단일 유전자당 1 내지 2개, 1 내지 3개, 1 내지 4개, 1 내지 5개, 1 내지 6개, 1 내지 7개, 1 내지 8개, 1 내지 9개, 또는 1 내지 10개의 동일하지 않은 gRNA에 대한 서열을 코딩한다. 일부 경우에서, gRNA는 sgRNA이다. 일부 경우에서, gRNA는 dgRNA이다.

[0114]

일부 경우에서, 본원에 기술된 gRNA 라이브러리는 유기체의 유전자 1개당 하나 이상의 동일하지 않은 gRNA를 포함한다. 일부 경우에서, gRNA 라이브러리는 유기체에 대한 유전자 1개당 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 동일하지 않은 gRNA를 포함한다. 예시적인 유기체로는 제한 없이, 아라비돕시스 탈리아나, 카에노르합디티스 엘레강스, 캐나스 루푸스 패밀리아리스, 클라미도모나스 레인하드티이, 다니오 레리오, 덕티오스텔리움 디스코이데움, 드로소필라 멜라노가스터, 에스테리키아 콜라이, 마카카 물라타, 무스 무스쿨루스, 오릭토라구스 쿠니쿨루스, 라투스 노르베기쿠스, 사카로마이세스 세레비시아에, 수스 스크로파, 및 호모 사피엔스를 포함한다. 일부 경우에서, gRNA는 sgRNA이다. 일부 경우에서, gRNA는 dgRNA이다. 일부 경우에서, gRNA 라이브러리는 유기체의 전체 게놈 중 적어도 또는 약 5%에 대한 동일하지 않은 gRNA를 포함한다. 일부 경우에서, gRNA 라이브러리는 유기체의 전체 게놈 중 약 5% 내지 약 100%에 대한 동일하지 않은 gRNA를 포함한다. 일부 경우에서, gRNA 라이브러리는 유기체의 전체 게놈 중 적어도 또는 약 80%에 대한 동일하지 않은 gRNA를 포함한다. 일부 경우에서, sgRNA 라이브러리는 유기체의 전체 게놈 중 적어도 또는 약 90%에 대한 동일하지 않은 gRNA를 포함한다. 일부 경우에서, gRNA 라이브러리는 유기체의 전체 게놈 중 적어도 또는 약 95%에 대한 동일하지 않은 gRNA를 포함한다. 일부 경우에서, gRNA 라이브러리는 유기체의 전체 게놈 중 적어도 또는 약 100%에 대한 동일하지 않은 gRNA를 포함한다. 일부 경우에서, gRNA를 포함한다. 일부 경우에서, gRNA 라이브러리는 유기체의 전체 게놈 중 약 5% 내지 10%, 5% 내지 20%, 5% 내지 30%, 5% 내지 40%, 5% 내지 50%, 5% 내지 60%, 5% 내지 70%, 5% 내지 80%, 5% 내지 90%, 5% 내지 95%, 5% 내지

100%, 10% 내지 20%, 10% 내지 30%, 10% 내지 40%, 10% 내지 50%, 10% 내지 60%, 10% 내지 70%, 10% 내지 80%, 10% 내지 90%, 10% 내지 95%, 10% 내지 100%, 20% 내지 30%, 20% 내지 40%, 20% 내지 50%, 20% 내지 60%, 20% 내지 70%, 20% 내지 80%, 20% 내지 90%, 20% 내지 95%, 20% 내지 100%, 30% 내지 40%, 30% 내지 50%, 30% 내지 60%, 30% 내지 70%, 30% 내지 80%, 30% 내지 90%, 30% 내지 95%, 30% 내지 100%, 40% 내지 50%, 40% 내지 60%, 40% 내지 70%, 40% 내지 80%, 40% 내지 90%, 40% 내지 95%, 40% 내지 100%, 50% 내지 60%, 50% 내지 70%, 50% 내지 80%, 50% 내지 90%, 50% 내지 95%, 50% 내지 100%, 60% 내지 70%, 60% 내지 80%, 60% 내지 90%, 60% 내지 95%, 60% 내지 100%, 70% 내지 80%, 70% 내지 90%, 70% 내지 95%, 70% 내지 100%, 80% 내지 90%, 80% 내지 95%, 80% 내지 100%, 90% 내지 95%, 90% 내지 100%, 또는 95% 내지 100%에 대한 동일하지 않은 gRNA를 포함한다. 일부 경우에서, gRNA 라이브러리는 경로 중의 다중의 유전자로부터의 서열, 또는 유기체 중의 모든 유전자로부터의 서열을 포함한다. gRNA의 개수는 표 4에 열거된 유기체에서 유전자당 적어도 2X, 3X, 5X, 또는 10X를 포함할 수 있다. 일부 경우에서, gRNA 라이브러리는 유전자, 유전자의 군(예컨대, 3-10개의 유전자), 경로(예컨대, 10-100개의 유전자), 또는 챠시(chassis)(예컨대, 100-1000개의 유전자) 중 적어도 하나를 표적화한다.

[0115] 고도의 병렬식 드 노보 핵산 합성

본원에서는 혁신적인 합성 플랫폼을 생성하기 위하여 소형화, 병렬화, 및 규소 상의 나노웰 내에서 올리고핵산 합성에서부터 유전자 어셈블리까지의 종단 간 프로세스의 수직적 통합을 이용하는 플랫폼 접근법을 제공한다. 96 웰 플레이트와 동일한 풋프린트를 갖는 본원에 기술된 장치는, 단일의 고도로 병렬화된 작업에서 최대 대략 1,000,000개의 올리고핵산을 생성하는 종래의 합성 방법과 비교하여 100 내지 1,000배만큼 처리량을 증가시킬 수 있는 규소 합성 플랫폼을 제공한다. 일부 경우에서, 본원에 기술된 단일 규소 플레이트는 약 6,100개의 동일하지 않은 올리고핵산의 합성을 제공한다. 일부 경우에서, 동일하지 않은 올리고핵산은 각각 클러스터 내에 위치한다. 클러스터는 50 내지 500개의 동일하지 않은 올리고핵산을 포함할 수 있다.

[0117] 일부 경우에서, 본원에 기술된 gRNA 라이브러리를 코딩하는 DNA 라이브러리의 오류율은 DNA에 대한 소정의 서열과 비교하였을 때, 1:500 미만이다. 일부 경우에서, 본원에 개시된 드 노보 올리고핵산 라이브러리의 총 오류율은 DNA에 대한 소정의 서열과 비교하였을 때, 1:500, 1:1,000, 1:1,500, 1:2,000, 1:3,000, 1:5,000 미만, 또는 그 미만이다. 일부 경우에서, 총 오류율은 DNA에 대한 소정의 서열과 비교하였을 때, 1:1,000 미만이다. 오류율은 총 오류율 또는 평균 오류율일 수 있다.

[0118] 일부 경우에서, 본원에 기술된 gRNA 라이브러리를 코딩하는 RNA 라이브러리의 오류율은 RNA에 대한 소정의 서열과 비교하였을 때, 1:500 미만이다. 일부 경우에서, 본원에 개시된 드 노보 올리고핵산 라이브러리의 총 오류율은 RNA에 대한 소정의 서열과 비교하였을 때, 1:500, 1:1,000, 1:1,500, 1:2,000, 1:3,000, 1:5,000, 1:10,000 미만, 또는 그 미만이다. 일부 경우에서, 총 오류율은 RNA에 대한 소정의 서열과 비교하였을 때, 1:1,000 미만이다.

[0119] 기판

일부 경우에서, 본원에서는 각각의 클러스터가 올리고핵산의 부착 및 합성을 지원하는 복수의 좌위를 포함하는 것인 복수의 클러스터를 포함하는 기판을 제공한다. 본원에서 사용되는 바, "좌위"라는 용어는 단일의 소정의 서열을 코딩하는 올리고핵산이 표면으로부터 신장되는 것을 지원하는 구조 상의 별개의 영역을 지칭한다. 일부 경우에서, 좌위는 2차원 표면, 예컨대, 실질적으로 평면 상에 위치한다. 일부 실시양태에서, 좌위는 예컨대, 웰, 마이크로웰, 채널, 또는 포스트와 같은 3차원 표면 상에 위치한다. 일부 실시양태에서, 좌위의 표면은 올리고핵산 합성을 위해 적어도 하나의 뉴클레오티드에, 또는 바람직하게, 올리고핵산 집단의 합성을 위해 동일한 뉴클레오티드로 이루어진 집단에 부착되도록 능동적으로 작용화된 물질을 포함한다. 일부 실시양태에서, 올리고핵산은 동일한 핵산 서열을 코딩하는 올리고핵산 집단을 지칭한다. 일부 경우에서, 기판의 표면은 기판의 하나 또는 복수의 표면을 포함한다. 제공된 시스템 및 방법을 사용하여 라이브러리 내에서 합성된 올리고핵산에 대한 평균 오류율은 1,000개 중 1개 미만, 2,000개 중 약 1개 미만, 3,000개 중 약 1개 미만으로 빈번하거나, 또는 그보다 덜 빈번하다.

[0121] 일부 실시양태에서, 기판은 통상의 지지체 상의 주소지정가능한 위치에 상이한 소정의 서열을 갖는, 복수의 올리고핵산의 합성을 지원하는 표면을 포함한다. 일부 실시양태에서, 기판은 50, 100, 200, 400, 600, 800, 1,000, 1,200, 1,400, 1,600, 1,800, 2,000; 5,000; 10,000; 20,000; 50,000; 100,000; 200,000; 300,000; 400,000; 500,000; 600,000; 700,000; 800,000; 900,000; 1,000,000; 1,200,000; 1,400,000; 1,600,000; 1,800,000; 2,000,000; 2,500,000; 3,000,000; 3,500,000; 4,000,000; 4,500,000; 5,000,000; 10,000,000개 초과, 또는 그 초과의 동일하지 않은 올리고핵산의 합성을 지원한다. 일부 경우에서, 기판은 상이한 서열을 코딩하는, 50,

100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2,000; 5,000; 10,000; 20,000; 50,000; 100,000; 200,000; 300,000; 400,000; 500,000; 600,000; 700,000; 800,000; 900,000; 1,000,000; 1,200,000; 1,400,000; 1,600,000; 1,800,000; 2,000,000; 2,500,000; 3,000,000; 3,500,000; 4,000,000; 4,500,000; 5,000,000; 10,000,000개 초과, 또는 그 초과의 올리고핵산의 합성을 지원한다. 일부 실시양태에서, 올리고핵산 중 적어도 일부는 동일한 서열을 갖거나, 또는 동일한 서열로 합성되도록 구성된다. 일부 실시양태에서, 기관은 적어도 80, 90, 100, 120, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500개 이상의 염기를 갖는 올리고핵산의 성장을 위한 표면 환경을 제공한다.

[0122]

일부 실시양태에서, 올리고핵산은, 각 좌위가 올리고핵산 집단의 합성을 지원하는 것인, 기관의 상이한 좌위 상에서 합성된다. 일부 경우에서, 각 좌위는 또 다른 좌위 상에서 성장된 올리고핵산 집단과는 상이한 서열을 갖는 올리고핵산 집단의 합성을 지원한다. 일부 실시양태에서, 기관의 좌위는 복수의 클러스터 내에 위치한다. 일부 경우에서, 기관은 적어도 10, 500, 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 11,000, 12,000, 13,000, 14,000, 15,000, 20,000, 30,000, 40,000, 50,000개 이상의 클러스터를 포함한다. 일부 실시양태에서, 기관은 2,000; 5,000; 10,000; 100,000; 200,000; 300,000; 400,000; 500,000; 600,000; 700,000; 800,000; 900,000; 1,000,000; 1,100,000; 1,200,000; 1,300,000; 1,400,000; 1,500,000; 1,600,000; 1,700,000; 1,800,000; 1,900,000; 2,000,000; 300,000; 400,000; 500,000; 600,000; 700,000; 800,000; 900,000; 1,000,000; 1,200,000; 1,400,000; 1,600,000; 1,800,000; 2,000,000; 2,500,000; 3,000,000; 3,500,000; 4,000,000; 4,500,000; 5,000,000; 또는 10,000,000개 초과 또는 그 초과의 상이한 좌위를 포함한다. 기관은 약 10,000개의 상이한 좌위를 포함한다. 상이한 실시양태에서, 단일 클러스터 내의 좌위의 양은 다른 실시양태에서는 달라진다. 일부 경우에서, 각 클러스터는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 130, 150, 200, 300, 400, 500개 이상의 좌위를 포함한다. 일부 실시양태에서, 각 클러스터는 약 50-500개의 좌위를 포함한다. 일부 실시양태에서, 각 클러스터는 약 100-200개의 좌위를 포함한다. 일부 실시양태에서, 각 클러스터는 약 100-150개의 좌위를 포함한다. 일부 실시양태에서, 각 클러스터는 약 109, 121, 130 또는 137개의 좌위를 포함한다. 일부 실시양태에서, 각 클러스터는 약 19, 20, 61, 64개 또는 그 초과의 좌위를 포함한다.

[0123]

본원에서는 규소 플레이트 상에서 동일하지 않은 올리고핵산을 합성하는 방법을 제공한다. 일부 경우에서, 규소 플레이트는 약 1-10, 1-50, 또는 50-500개의 클러스터를 포함한다. 일부 경우에서, 규소 플레이트는 약 50, 100, 250, 500, 2,500, 5,000, 6,000, 6,150, 10,000개 초과 또는 그 초과의 클러스터를 포함한다. 일부 경우에서, 각 클러스터는 121개의 좌위를 포함한다. 일부 경우에서, 각 클러스터는 약 50-500, 50-200, 100-150개의 좌위를 포함한다. 일부 경우에서, 각 클러스터는 적어도 약 50, 100, 150, 200, 500, 1,000개 이상의 좌위를 포함한다. 일부 경우에서, 단일 플레이트는 100, 500, 10,000, 20,000, 30,000, 50,000, 100,000, 500,000, 700,000, 1,000,000개 이상의 좌위를 포함한다.

[0124]

일부 실시양태에서, 기관 상에서 합성된 상이한 올리고핵산의 개수는 기관에서 이용가능한 상이한 좌위의 개수에 의존한다. 일부 실시양태에서, 기관의 클러스터 내의 좌위의 밀도는 적어도 또는 약 1개의 좌위/ mm^2 , 10개의 좌위/ mm^2 , 25개의 좌위/ mm^2 , 50개의 좌위/ mm^2 , 65개의 좌위/ mm^2 , 75개의 좌위/ mm^2 , 100개의 좌위/ mm^2 , 130개의 좌위/ mm^2 , 150개의 좌위/ mm^2 , 175개의 좌위/ mm^2 , 200개의 좌위/ mm^2 , 300개의 좌위/ mm^2 , 400개의 좌위/ mm^2 , 500개의 좌위/ mm^2 , 1,000개의 좌위/ mm^2 이상이다. 일부 경우에서, 기관은 약 10개의 좌위/ mm^2 내지 약 10개의 좌위/약 500 mm^2 , 약 25개의 좌위/ mm^2 내지 약 25개의 좌위/약 400 mm^2 , 약 50개의 좌위/ mm^2 내지 약 50개의 좌위/약 500 mm^2 , 약 100개의 좌위/ mm^2 내지 약 100개의 좌위/약 500 mm^2 , 약 150개의 좌위/ mm^2 내지 약 150개의 좌위/약 500 mm^2 , 약 10개의 좌위/ mm^2 내지 약 10개의 좌위/약 250 mm^2 , 약 50개의 좌위/ mm^2 내지 약 50개의 좌위/약 250 mm^2 , 약 10개의 좌위/ mm^2 내지 약 10개의 좌위/약 200 mm^2 , 또는 약 50개의 좌위/ mm^2 내지 약 50개의 좌위/약 200 mm^2 를 포함한다. 일부 실시양태에서, 클러스터 내의 2개의 인접한 좌위의 중심부 사이의 거리는 약 10 μm 내지 약 500 μm , 약 10 μm 내지 약 200 μm , 또는 약 10 μm 내지 약 100 μm 이다. 일부 경우에서, 인접한 좌위의 두 중심부 사이의 거리는 약 10 μm , 20 μm , 30 μm , 40 μm , 50 μm , 60 μm , 70 μm , 80 μm , 90 μm 또는 100 μm 초과이다. 일부 경우에서, 2개의 인접한 좌위의 중심부 사이의 거리는 약 200 μm , 150 μm , 100 μm , 80 μm , 70 μm , 60 μm , 50 μm , 40 μm , 30 μm , 20 μm 또는 10 μm 미만이다. 일부 경우에서, 각 좌위의 너비는 약 0.5 μm , 1 μm , 2 μm , 3 μm , 4 μm , 5 μm , 6 μm , 7 μm , 8 μm , 9 μm , 10 μm , 20 μm , 30 μm , 40 μm , 50 μm , 60 μm , 70 μm , 80 μm , 90 μm 또는 100 μm 이다. 일부 경우에서, 각 좌위의 너비는 약 0.5 μm 내지 100 μm , 약 0.5 μm 내지 50 μm , 약 10 μm 내지 75 μm , 또는 약 0.5 μm 내지 50 μm 이다.

[0125]

일부 실시양태에서, 기관 내 클러스터의 밀도는 적어도 또는 약 1개의 클러스터/100 mm^2 , 1개의 클러스터/10 mm^2 ,

1개의 클러스터/5 mm^2 , 1개의 클러스터/4 mm^2 , 1개의 클러스터/3 mm^2 , 1개의 클러스터/2 mm^2 , 1개의 클러스터/1 mm^2 , 2개의 클러스터/1 mm^2 , 3개의 클러스터/1 mm^2 , 4개의 클러스터/1 mm^2 , 5개의 클러스터/1 mm^2 , 10개의 클러스터/1 mm^2 , 50개의 클러스터/1 mm^2 이상이다. 일부 실시양태에서, 기판은 약 1개의 클러스터/10 mm^2 내지 약 10개의 클러스터/1 mm^2 를 포함한다. 일부 실시양태에서, 2개의 인접한 클러스터의 중심부 사이의 거리는 약 50 μm , 100 μm , 200 μm , 500 μm , 1,000 μm , 또는 2,000 μm 또는 5,000 μm 미만이다. 일부 경우에서, 2개의 인접한 클러스터의 중심부 사이의 거리는 약 50 μm 내지 약 100 μm , 약 50 μm 내지 약 200 μm , 약 50 μm 내지 약 300 μm , 약 50 μm 내지 약 500 μm , 및 약 100 μm 내지 약 2,000 μm 이다. 일부 경우에서, 2개의 인접한 클러스터의 중심부 사이의 거리는 약 0.05 mm 내지 약 50 mm , 약 0.05 mm 내지 약 10 mm , 약 0.05 mm 내지 약 5 mm , 약 0.05 mm 내지 약 4 mm , 약 0.05 mm 내지 약 3 mm , 약 0.05 mm 내지 약 2 mm , 약 0.1 mm 및 10 mm , 약 0.2 mm 및 10 mm , 약 0.3 mm 내지 약 10 mm , 약 0.4 mm 내지 약 10 mm , 약 0.5 mm 및 10 mm , 약 0.5 mm 내지 약 5 mm , 또는 약 0.5 mm 내지 약 2 mm 이다. 일부 경우에서, 각 클러스터의 횡단면은 약 0.5 내지 2 mm , 약 0.5 내지 1 mm , 또는 약 1 내지 2 mm 이다. 일부 경우에서, 각 클러스터의 횡단면은 약 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9 또는 2 mm 이다. 일부 경우에서, 각 클러스터의 내부 횡단면은 약 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9 또는 2 mm 이다.

[0126] 일부 실시양태에서, 기판의 크기는 대략 표준 96 웰 플레이트의 크기, 예를 들어, 약 100 내지 200 $\text{mm} \times$ 약 50 내지 150 mm 이다. 일부 실시양태에서, 기판의 직경은 약 1,000 mm , 500 mm , 450 mm , 400 mm , 300 mm , 250 mm , 200 mm , 150 mm , 100 mm 또는 50 mm 이하이다. 일부 실시양태에서, 기판의 직경은 약 25 mm 내지 1,000 mm , 약 25 mm 내지 약 800 mm , 약 25 mm 내지 약 600 mm , 약 25 mm 내지 약 500 mm , 약 25 mm 내지 약 400 mm , 약 25 mm 내지 약 300 mm , 또는 약 25 mm 내지 약 200이다. 기판 크기의 비제한적인 예로는 약 300 mm , 200 mm , 150 mm , 130 mm , 100 mm , 76 mm , 51 mm 및 25 mm 를 포함한다. 일부 실시양태에서, 기판의 평면적은 적어도 약 100 mm^2 ; 200 mm^2 ; 500 mm^2 ; 1,000 mm^2 ; 2,000 mm^2 ; 5,000 mm^2 ; 10,000 mm^2 ; 12,000 mm^2 ; 15,000 mm^2 ; 20,000 mm^2 ; 30,000 mm^2 ; 40,000 mm^2 ; 50,000 mm^2 이상이다. 일부 실시양태에서, 기판의 두께는 약 50 mm 내지 약 2,000 mm , 약 50 mm 내지 약 1,000 mm , 약 100 mm 내지 약 1,000 mm , 약 200 mm 내지 약 1,000 mm , 또는 약 250 mm 내지 약 1,000 mm 이다. 기판 두께의 비제한적인 예로는 275 mm , 375 mm , 525 mm , 625 mm , 675 mm , 725 mm , 775 mm 및 925 mm 를 포함한다. 일부 경우에서, 기판의 두께는 직경에 따라 달라지며, 기판 조성에 의존한다. 예를 들어, 규소 이외의 다른 물질을 포함하는 기판은 직경이 동일한 규소 기판과 다른 두께를 가진다. 기판 두께는 사용되는 물질의 기계적 강도에 의해 측정될 수 있고, 기판은 취급하는 동안 균열 없이 그 자신의 중량을 지탱할 수 있을 정도로 충분히 두꺼워야 한다.

표면 물질

[0127] 본원에서는 본원에 기술된 방법 및 조성물에 적합한 임의의 다양한 물질로부터 제작된 기판, 기판 및 반응기를 제공한다. 특정 실시양태에서, 기판 물질은 낮은 수준의 뉴클레오티드 결합을 보이도록 제작된다. 일부 경우에서, 기판 물질은 높은 수준의 뉴클레오티드 결합을 보이는 상이한 표면을 생성하도록 개질된다. 일부 실시양태에서, 기판 물질은 가시광 및/또는 UV 광에 대하여 투과성을 나타낸다. 일부 실시양태에서, 기판 물질은 충분한 전도성을 띠고, 예컨대, 기판 전체 또는 그 일부에 걸쳐 균일한 전기장을 형성할 수 있다. 일부 실시양태에서, 전도성 물질은 전기 접지에 연결된다. 일부 경우에서, 기판은 열 전도성이거나, 또는 절연 처리가 된 것이다. 일부 경우에서, 물질은 화학적 또는 생화학적 반응, 예를 들어, 올리고핵산 합성 반응 프로세스를 견딜 수 있을 정도로 내화학성 및 내열성을 띤다. 일부 실시양태에서, 기판은 가요성 물질을 포함한다. 가요성 물질로는 제한 없이, 개질된 나일론, 개질되지 않은 나일론, 니트로셀룰로스, 폴리프로필렌 등을 포함한다. 일부 실시양태에서, 기판은 강성 물질을 포함한다. 강성 물질로는 제한 없이, 유리, 훈연 실리카, 규소, 이산화규소, 질화규소, 플라스틱(예를 들어, 폴리테트라플루오로에틸렌, 폴리프로필렌, 폴리스티렌, 폴리카보네이트, 및 그의 브랜드 등), 및 금속(예를 들어, 금, 백금 등)을 포함한다. 일부 실시양태에서, 기판은 규소, 폴리스티렌, 아가로스, 텍스트란, 셀룰로스계 중합체, 폴리아크릴아미드, 폴리디메틸실록산(PDMS), 유리, 또는 그의 임의의 조합을 포함하는 물질로부터 제작된다. 일부 경우에서, 기판은 본원에 열거된 물질, 또는 당업계에 공지된 임의의 다른 적합한 물질의 조합을 이용하여 제작된다.

표면 아키텍처

[0128] 다양한 실시양태에서, 기판은 돌출된 및/또는 함몰된 특징부를 포함한다. 상기와 같은 특징부를 갖는 것의 한가지 이점은 올리고핵산 합성을 지원할 수 있는 표면적이 증가한다는 점이다. 일부 실시양태에서, 돌출된 및/또는 함몰된 특징부를 갖는 기판은 3차원 기판으로 지칭된다. 일부 경우에서, 3차원 기판은 하나 이상의 채널을 포함한다. 일부 경우에서, 하나 이상의 좌위는 채널을 포함한다. 일부 경우에서, 채널은 증착 장치, 예컨대, 올리고

핵산 합성기를 통한 시약 중착에 접근가능하다. 일부 경우에서, 시약 및/또는 유체는 유체 소통하는 하나 이상의 채널 중 더욱 큰 웰 중에 수집된다. 예를 들어, 기판은 클러스터와 함께 복수의 좌위에 상응하는 복수의 채널을 포함하고, 복수의 채널은 클러스터의 한 웰과 유체 소통한다. 일부 방법에서, 올리고핵산 라이브러리는 클러스터의 복수의 좌위에서 합성된다.

[0131] 일부 실시양태에서, 구조물은 표면 상에서의 올리고핵산 합성을 위해 조절된 방식의 유동 및 물질 전달 경로를 허용하도록 구성된다. 일부 실시양태에서, 기판의 구성은 올리고핵산 합성 동안의 조절된 방식 및 고른 분포의 물질 전달 경로, 화학적 노출 시간, 및/또는 세척 효능을 허용한다. 일부 실시양태에서, 기판의 구성은 예를 들어, 성장하는 올리고핵산을 위해 충분한 부피를 제공함으로써 성장하는 올리고핵산에 의해 배제된 부피가, 올리고핵산을 성장시키는 데 이용가능하거나, 또는 그에 적합한 초기의 이용가능한 부피의 50, 45, 40, 35, 30, 25, 20, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1% 이하보다 더 큰 부피를 차지하지 않도록 함으로써 스윕 효율(sweep efficiency)을 증가시킬 수 있다. 일부 실시양태에서, 3차원 구조물은 관리된 방식의 유체 유동이 화학적 노출을 빠르게 교체할 수 있게 허용한다.

[0132] 일부 실시양태에서, 분리는 물리적 구조에 의해 달성된다. 일부 실시양태에서, 분리는 표면을 차별적으로 작용화하여 올리고핵산 합성을 위한 능동적 및 수동적 영역을 생성함으로써 달성된다. 차별적 작용화는 또한 기판 표면에 걸쳐 교대로 소수성을 띠게 만들어 중착된 시약의 비딩(beading) 또는 습윤화를 유발하는 물 접촉각 효과를 일으킴으로써 달성된다. 더욱 큰 구조물을 사용하면, 이웃 스폿의 시약에 의한 상이한 올리고핵산 합성 위치의 튀김(splashing) 또는 교차 오염을 감소시킬 수 있다. 일부 경우에서, 기판, 예컨대, 올리고핵산 합성기는 상이한 올리고핵산 합성 위치에 시약을 중착시키는 데 사용된다. 3차원 특징부를 갖는 기판은 낮은 오류율(예컨대, 약 1:500, 1:1000, 1:1500, 1:2,000; 1:3,000; 1:5,000; 또는 1:10,000 미만)로 다수의 올리고핵산(예컨대, 약 10,000개 초과)을 합성할 수 있도록 하는 방식으로 구성된다. 일부 경우에서, 기판은 1 μm 당 약 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 300, 400 또는 500개의 특징부, 또는 그 초과의 밀도로 특징부를 포함한다.

[0133] 기판의 한 웰은 기판의 또 다른 웰과 동일하거나, 또는 상이한 너비, 높이, 및/또는 부피를 가질 수 있다. 기판의 한 채널은 기판의 또 다른 채널과 동일하거나, 또는 상이한 너비, 높이, 및/또는 부피를 가질 수 있다. 일부 실시양태에서, 클러스터의 직경, 또는 클러스터를 포함하는 웰의 직경, 또는 그 둘 모두는 약 0.05 μm 내지 약 50 μm , 약 0.05 μm 내지 약 10 μm , 약 0.05 μm 내지 약 5 μm , 약 0.05 μm 내지 약 4 μm , 약 0.05 μm 내지 약 3 μm , 약 0.05 μm 내지 약 2 μm , 약 0.05 μm 내지 약 1 μm , 약 0.05 μm 내지 약 0.5 μm , 약 0.05 μm 내지 약 0.1 μm , 약 0.1 μm 및 10 μm , 약 0.2 μm 및 10 μm , 약 0.3 μm 내지 약 10 μm , 약 0.4 μm 내지 약 10 μm , 약 0.5 μm 및 10 μm , 약 0.5 μm 내지 약 5 μm , 또는 약 0.5 μm 내지 약 2 μm 이다. 일부 실시양태에서, 클러스터 또는 웰, 또는 그 둘 모두의 직경은 약 5 μm , 4 μm , 3 μm , 2 μm , 1 μm , 0.5 μm , 0.1 μm , 0.09 μm , 0.08 μm , 0.07 μm , 0.06 μm 또는 0.05 μm 이하이다. 일부 실시양태에서, 클러스터 또는 웰, 또는 그 둘 모두의 직경은 약 1.0 내지 1.3 μm 이다. 일부 실시양태에서, 클러스터 또는 웰, 또는 그 둘 모두의 직경은 약 1.150 μm 이다. 일부 실시양태에서, 클러스터 또는 웰, 또는 그 둘 모두의 직경은 약 0.08 μm 이다. 클러스터의 직경은 2차원 또는 3차원 기판 내의 클러스터를 지칭한다.

[0134] 일부 실시양태에서, 웰의 높이는 약 20 μm 내지 약 1,000 μm , 약 50 μm 내지 약 1,000 μm , 약 100 μm 내지 약 1,000 μm , 약 200 μm 내지 약 1,000 μm , 약 300 μm 내지 약 1,000 μm , 약 400 μm 내지 약 1,000 μm , 또는 약 500 μm 내지 약 1,000 μm 이다. 일부 경우에서, 웰의 높이는 약 1,000 μm 미만, 약 900 μm 미만, 약 800 μm 미만, 약 700 μm 미만, 또는 약 600 μm 미만이다.

[0135] 일부 실시양태에서, 기판은 클러스터 내의 복수의 좌위에 상응하는 복수의 채널을 포함하고, 여기서, 채널의 높이 또는 깊이는 약 5 μm 내지 약 500 μm , 약 5 μm 내지 약 400 μm , 약 5 μm 내지 약 300 μm , 약 5 μm 내지 약 200 μm , 약 5 μm 내지 약 100 μm , 약 5 μm 내지 약 50 μm , 또는 약 10 μm 내지 약 50 μm 이다. 일부 경우에서, 채널의 높이는 100 μm 미만, 80 μm 미만, 60 μm 미만, 40 μm 미만 또는 20 μm 미만이다.

[0136] 일부 실시양태에서, 채널, 좌위(예컨대, 실질적으로 평평한 기판에서), 또는 채널 및 좌위 둘 모두(예컨대, 좌위가 채널에 상응하는 것인 3차원 기판에서)의 직경은 약 1 μm 내지 약 1,000 μm , 약 1 μm 내지 약 500 μm , 약 1 μm 내지 약 200 μm , 약 1 μm 내지 약 100 μm , 약 5 μm 내지 약 100 μm , 또는 약 10 μm 내지 약 100 μm , 예를 들어, 약 90 μm , 80 μm , 70 μm , 60 μm , 50 μm , 40 μm , 30 μm , 20 μm 또는 10 μm 이다. 일부 실시양태에서, 채널, 좌위 또는 채널 및 좌위 둘 모두의 직경은 약 100 μm , 90 μm , 80 μm , 70 μm , 60 μm , 50 μm , 40 μm , 30 μm , 20 μm 또는 10 μm 미만이다. 일부 실시양태에서, 2개의 인접한 채널, 좌위, 또는 채널 및 좌위의 중심부 사

이의 거리는 약 1 μm 내지 약 500 μm , 약 1 μm 내지 약 200 μm , 약 1 μm 내지 약 100 μm , 약 5 μm 내지 약 200 μm , 약 5 μm 내지 약 100 μm , 약 5 μm 내지 약 50 μm , 또는 약 5 μm 내지 약 30 μm , 예를 들어, 약 20 μm 이다.

표면 개질

다양한 실시양태에서, 표면 개질은 기판 표면, 또는 장치 표면의 선택된 부위 또는 영역의 하나 이상의 화학적 및/또는 물리적 특성을 변화시키기 위해 추가 또는 공제 프로세스에 의해 표면을 화학적으로 및/또는 물리적으로 변경시키는 데 사용된다. 예를 들어, 표면 개질로는 제한 없이, (1) 표면의 습윤화 특성을 변화시키는 것, (2) 표면을 작용화하는 것, 즉, 표면 작용기를 제공하거나, 개질시키거나, 또는 치환하는 것, (3) 표면을 탈작용화하는 것, 즉, 표면 작용기를 제거하는 것, (4) 다르게는, 예컨대, 애칭을 통해 표면의 화학적 조성을 변경시키는 것, (5) 표면 조도를 증가 또는 감소시키는 것, (6) 표면 상에 코팅, 예컨대, 표면의 습윤화 특성과 상이한 습윤화 특성을 나타내는 코팅을 제공하는 것, 및/또는 (7) 표면 상에 미립자를 침착시키는 것을 포함한다.

일부 경우에서, 표면 위에 (부착 프로모터로 지칭되는) 화학층을 추가하는 것이 기판의 표면 상에의 좌위의 구조화된 패턴화를 촉진시킨다. 부착 프로모션 적용을 위한 예시적인 표면으로는 제한 없이, 유리, 규소, 이산화규소 및 질화규소를 포함한다. 일부 경우에서, 부착 프로모터는 표면 에너지가 높은 화학물질이다. 일부 경우에서, 제2 화학층을 기판의 표면 상에 증착시킨다. 일부 경우에서, 제2 화학층의 표면 에너지는 낮다. 일부 경우에서, 표면 상에 코팅된 화학층의 표면 에너지는 표면 상의 액적의 국재화를 지원한다. 선택된 패턴화 배열에 의존하여, 좌위의 근접성 및/또는 좌위에서의 유체 접촉 면적은 변경가능하다.

일부 실시양태에서, 예컨대, 올리고핵산 합성을 위해 그 위에 핵산 또는 다른 모이어티가 증착되어 있는 것인 기판 표면, 또는 분할된 좌위는 평활하거나, 또는 실질적으로 평평하거나(예컨대, 2차원), 또는 요철, 예컨대, 돌출된 또는 함몰된 특징부(예컨대, 3차원 특징부)를 가진다. 일부 실시양태에서, 기판 표면은 하나 이상의 상이한 화합물 층으로 개질된다. 관심의 대상이 되는 상기와 같은 개질 층으로는 제한 없이, 무기 및 유기 층, 예컨대, 금속, 금속 산화물, 중합체, 소형 유기 분자 등을 포함한다. 비제한적인 중합체 층은 웨티드, 단백질, 핵산 또는 그의 모방체(예컨대, 웨티드 핵산 등), 다당류, 인지질, 폴리우레탄, 폴리에스테르, 폴리카보네이트, 폴리우레아, 폴리아미드, 폴리에틸렌아민, 폴리아릴렌 술퍼드, 폴리실록산, 폴리이미드, 폴리아세테이트, 및 본원에 기술되거나, 또는 다르게는 당업계에 공지된 임의의 다른 적합한 화합물을 포함한다. 일부 경우에서, 중합체는 이종중합체이다. 일부 경우에서, 중합체는 동종중합체이다. 일부 경우에서, 중합체는 작용성 모이어티를 포함하거나, 또는 접합된 것이다.

기판의 분할된 좌위는 표면 에너지를 증가 및/또는 감소시키는 하나 이상의 모이어티로 작용화된다. 일부 경우에서, 모이어티는 화학적으로 불활성이다. 일부 경우에서, 모이어티는 원하는 화학 반응, 예를 들어, 올리고핵산 합성 반응 중 하나 이상의 프로세스를 지원하도록 구성된다. 표면의 표면 에너지, 또는 소수성을 표면 상에 부착되는 뉴클레오티드의 친화도를 결정짓는 인자이다. 일부 실시양태에서, 기판 작용화 방법은 (a) 이산화규소를 포함하는 표면을 갖는 기판을 제공하는 단계; 및 (b) 본원에 기술되어 있거나, 또는 다르게는 당업계에 공지된 적합한 실란화제, 예를 들어, 유기작용성 알콕시실란 분자를 사용하여 표면을 실란화시키는 단계를 포함할 수 있다. 일부 경우에서, 유기작용성 알콕시실란 분자로는 디메틸클로로-옥토데실-실란, 메틸디클로로-옥토데실-실란, 트리클로로-옥토데실-실란, 트리메틸-옥토데실-실란, 트리에틸-옥토데실-실란, 또는 그의 임의의 조합을 포함한다. 일부 실시양태에서, 기판 표면은 (감마 방사선 조사 또는 크롬산 산화, 및 하이드록시 알킬 표면으로의 환원에 의해 작용화된) 폴리에틸렌/폴리프로필렌, (클로로메틸화에 의해 유도체화되고, 벤질아민 작용성 표면으로 아민화된) 고도로 가교결합된 폴리스티렌-디비닐벤젠, 나일론(말단 아미노헥실 기는 직접적인 반응성을 나타냄)으로 작용화된 것, 또는 환원된 폴리테트라플루오로에틸렌으로 애칭된 것을 포함한다. 다른 방법 및 작용화제는 미국 특허 번호 5474796(상기 특허는 그 전문이 본원에서 참조로 포함된다)에 기술되어 있다.

일부 실시양태에서, 기판 표면은 전형적으로 기판 표면 상에 존재하는 반응성 친수성 모이어티를 통해 실란을 기판 표면에 커플링시키는 데 효과적인 반응 조건하에서 실란의 혼합물을 함유하는 유도체화 조성물과의 접촉에 의해 작용화된다. 실란화는 일반적으로 유기작용성 알콕시실란 분자와의 자가 조립을 통해 표면을 커버한다. 다양한 실록산 작용화 시약은 현재 당업계에 공지되어 있는 바와 같이, 예컨대, 표면 에너지를 낮추거나, 또는 증가시키기 위해 추가로 사용될 수 있다. 유기작용성 알콕시실란은 그의 유기 작용기에 따라 분류된다. 실록산 작용화 시약의 비제한적인 예로는 하이드록시알킬 실록산(실릴레이트 표면, 디보란으로 작용화 및 과산화수소에 의한 알콜 산화), 디올(디하이드록시알킬) 실록산(실릴레이트 표면, 및 디돌로 가수분해), 아미노알킬 실록산(아민은 중간 작용화 단계를 필요로 하지 않는다), 글리시독시실란(3-글리시독시프로필-디메틸-에톡시실란, 글리시독시-트리메톡시실란), 메르캅토실란(3-메르캅토프로필-트리메톡시실란, 3-4 에폭시사이클로헥실-에틸트리

메톡시실란 또는 3-메르캅토프로필-메틸-디메톡시실란), 비사이클로헵테닐-트리클로로실란, 부틸-알테하이드르-트리메톡시실란, 또는 이량체 2급 아미노알킬 실록산을 포함한다. 예시적인 하이드록시알킬 실록산으로는 3-하이드록시프로필로 전환되는 알릴 트리클로로실란, 또는 8-하이드록시옥틸로 전환되는 7-옥트-1-에닐 트리클로로실란을 포함한다. 디올(디하이드록시알킬) 실록산으로는 글리시딜 트리메톡시실란 유도된(2,3-디하이드록시프로필옥시)프로필(GOPS)을 포함한다. 아미노알킬 실록산으로는 3-아미노프로필로 전환되는 3-아미노프로필 트리메톡시실란(3-아미노프로필-트리에톡시실란, 3-아미노프로필-디에톡시-메틸실란, 3-아미노프로필-디메틸-에톡시실란, 또는 3-아미노프로필-트리메톡시실란)을 포함한다. 예시적인 이량체 2급 아미노알킬 실록산으로는 비스(실릴옥실프로필)아민으로 전환되는 비스 (3-트리메톡시실릴프로필) 아민을 포함한다. 일부 실시양태에서, 작용화제는 11-아세톡시운데실트리에톡시실란, n-데실트리에톡시실란, (3-아미노프로필)트리메톡시실란, (3-아미노프로필)트리에톡시실란, 글리시딜옥시프로필/트리메톡시실란 및 N-(3-트리에톡시실릴프로필)-4-하이드록시부티르아미드를 포함한다.

[0143] 올리고핵산 합성

[0144] 다양한 실시양태에서, 올리고핵산 합성 방법은 포스포라미다이트 화학법을 포함하는 프로세스를 포함한다. 일부 실시양태에서, 올리고핵산 합성은 염기를 포스포라미다이트와 커플링하는 것을 포함한다. 일부 실시양태에서, 올리고핵산 합성은 커플링 조건하에서 포스포라미다이트를 증착시켜 염기를 커플링하는 것을 포함하고, 여기서, 동일한 염기가 임의적으로 1회 초과로 포스포라미다이트로 증착되고, 즉, 이중으로 커플링된다. 일부 실시양태에서, 올리고핵산 합성은 비반응 부위의 캡핑을 포함한다. 일부 경우에서, 캡핑은 임의적이다. 일부 실시양태에서, 올리고핵산 합성은 산화를 포함한다. 일부 실시양태에서, 올리고핵산 합성은 탈차단, 또는 탈트리틸화를 포함한다. 일부 실시양태에서, 올리고핵산 합성은 황화를 포함한다. 일부 경우에서, 올리고핵산 합성은 산화 또는 황화 중 어느 하나를 포함한다. 일부 실시양태에서, 올리고핵산 합성 반응 동안 한 단계 또는 각 단계 동안 예를 들어, 테트라졸 또는 아세토니트릴을 사용하여 기판을 세척한다. 포스포라미다이트 합성 방법에서 어느 한 단계 동안의 시간은 약 2 min, 1 min, 50 sec, 40 sec, 30 sec, 20 sec 및 10 sec 미만인 시간을 포함한다.

[0145] 포스포라미다이트 방법을 사용하는 올리고핵산 합성은 포스파이트 트리에스테르 결합 형성을 위해 성장하는 올리고핵산 쇄에의 포스포라미다이트 빌딩 블록(예컨대, 뉴클레오시드 포스포라미다이트)의 잇따른 부가를 포함할 수 있다. 포스포라미다이트 올리고핵산 합성은 3'에서 5' 방향으로 진행된다. 포스포라미다이트 올리고핵산 합성을 통해 합성 사이클에 따라 하나의 뉴클레오티드를 성장하는 핵산 쇄에 조절된 방식으로 부가할 수 있다. 일부 실시양태에서, 각 합성 사이클은 커플링 단계를 포함한다. 포스포라미다이트 커플링은 활성화된 뉴클레오시드 포스포라미다이트와, 예를 들어, 링커를 통해 기판에 결합된 뉴클레오시드 사이의 포스파이트 트리에스테르 결합의 형성을 포함한다. 일부 실시양태에서, 뉴클레오시드 포스포라미다이트를 활성화된 기판에 제공한다. 일부 실시양태에서, 뉴클레오시드 포스포라미다이트를 기판에 결합된 뉴클레오시드 대비 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100배 이상의 과량으로 기판에 제공한다. 일부 실시양태에서, 뉴클레오시드 포스포라미다이트를 부가하는 것은 무수 환경에서, 예를 들어, 무수 아세토니트릴 중에서 수행된다. 뉴클레오시드 포스포라미다이트 부가 후, 기판을 임의적으로 세척한다. 일부 실시양태에서, 커플링 단계는 1회 이상의 추가 회차로 반복되고, 임의적으로, 세척 단계는 기판에 뉴클레오시드 포스포라미다이트를 부가하는 단계들 사이에 이루어진다. 일부 실시양태에서, 본원에서 사용되는 올리고핵산 합성 방법은 1, 2, 3회 이상의 순차적인 커플링 단계를 포함한다. 다수 경우에서, 커플링 이전에, 기판에 결합된 뉴클레오시드는 보호기 제거에 의해 탈보호화되고, 여기서, 보호기는 중합화를 막는 작용을 한다. 통상의 보호기는 4,4'-디메톡시트리틸(DMT)이다.

[0146] 커플링 후, 포스포라미다이트 올리고핵산 합성 방법은 임의적으로 캡핑 단계를 포함한다. 캡핑 단계에서, 성장하는 올리고핵산을 캡핑제로 처리한다. 캡핑 단계는, 내부 염기가 결실된 올리고핵산이 형성되지 못하게 막으면서, 커플링 후의 비반응 기판 결합 5'-OH 기의 추가의 쇄 신장을 차단하는 데 유용하다. 추가로, 1H-테트라졸로 활성화된 포스포라미다이트는 작게 구아노신의 06 위치와 반응할 수 있다. 이론으로 제한하지 않으면서, I₂/물에 의한 산화시, 상기 부산물은 가능하게는 06-N7 이동을 통해 탈퓨린화될 수 있다. 퓨린 결여(apurinic) 부위는 올리고뉴클레오티드의 최종 탈보호화의 과정에서 절단된 상태로 종결되어, 전장의 생성물의 수율을 감소시킬 수 있다. 06 변형은 I₂/물을 사용한 산화 이전에 캡핑 시약을 사용한 처리에 의해 제거될 수 있다. 일부 실시양태에서, 올리고핵산 합성 동안 캡핑 단계를 포함하는 것이 캡핑을 포함하지 않는 합성과 비교하여 오류율을 감소시킨다. 한 예로서, 캡핑 단계는 아세트산 무수물과 1-메틸이미다졸의 혼합물로 기판 결합 올리고핵산을 처리하

는 것을 포함한다. 캡핑 단계 후, 기판을 임의적으로 세척한다.

[0147]

일부 실시양태에서, 뉴클레오시드 포스포라미다이트의 부가 후, 몇 임의적으로 캡핑 및 1회 이상의 세척 단계 후에, 기판에 결합된, 성장하는 핵산을 산화시킨다. 산화 단계는 포스파이트 트리에스테르가, 자연적으로 발생된 포스페이트 디에스테르 뉴클레오시드간 결합의 보호된 전구체인 4 배위 포스페이트 트리에스테르로 산화되는 것을 포함한다. 일부 경우에서, 성장하는 올리고핵산의 산화는 임의적으로 약염기(예컨대, 피리딘, 루티딘 또는 콜리딘)의 존재하에서 아이오딘 및 물로 처리함으로써 달성된다. 산화는 무수 조건하에서, 예컨대, tert-부틸 하이드로페옥시드 또는 (1S)-(+)-(10-캄포슬포닐)-옥사지리딘(CSO)을 사용하여 수행될 수 있다. 일부 방법에서, 캡핑 단계는 산화 후에 수행된다. 지속될 수 있는, 산화로부터의 잔류성 물이 후속 커플링을 억제시킬 수 있는 바, 제2 캡핑 단계는 기판을 건조시킬 수 있다. 산화 후, 기판 및 성장하는 올리고핵산을 임의적으로 세척한다. 일부 실시양태에서, 올리고뉴클레오티드 포스포로티오에이트를 수득하기 위해 산화 단계를 황화 단계로 치환하고, 여기서, 임의의 캡핑 단계는 황화 후에 수행될 수 있다. 3-(디메틸아미노메틸리넨)아미노)-3H-1,2,4-디티아졸-3-티온, DDTT, 보카주(Beaucage) 시약으로도 공지되어 있는 3H-1,2-벤조디티올-3-온 1,1-디옥시드, 및 N,N,N',N'-테트라에틸티우람 디슬피드(TETD)를 포함하나, 이에 제한되지 않는 다수의 시약이 황을 효율적으로 전달할 수 있다.

[0148]

뉴클레오시드를 도입하는 후속 사이클이 커플링을 통해 이루어지도록 하기 위해, 기판에 결합된, 성장하는 올리고핵산의 보호된 5' 단부를 제거하면, 이로써, 1차 하이드록실 기는 다음 뉴클레오시드 포스포라미다이트와 반응한다. 일부 실시양태에서, 보호기는 DMT이고, 탈차단은 디클로로메탄 중 트리클로로아세트산으로 이루어진다. 연장된 시간 동안, 또는 권장된 산 용액보다 더욱 강한 산을 이용하여 탈트리틸화를 수행하면, 고체 지지체에 결합된 올리고뉴클레오티드의 탈퓨린화는 증가될 수 있고, 이로써 원하는 전장의 생성물의 수율은 감소된다. 본원에 기술된 본 발명의 방법 및 조성물은 원하지 않는 탈퓨린화 반응을 제한하는 조절된 탈차단 조건을 제공한다. 일부 경우에서, 탈차단 후에 기판에 결합된 올리고핵산을 세척한다. 일부 경우에서, 탈차단 후의 효율적인 세척이 합성된 올리고핵산이 낮은 오류율을 갖는 데 기여한다.

[0149]

올리고핵산 합성 방법은 전형적으로 하기 단계 순서를 반복하는 것을 포함한다: 보호된 단량체를 능동적으로 작용화된 표면(예컨대, 좌위)에 적용시켜 활성화된 표면과, 링커 또는 사전에 탈보호화된 단량체와 연결시키는 단계; 적용된 단량체를 탈보호화하여 후속하여 적용되는 보호된 단량체와 반응을 하도록 하는 단계; 및 연결을 위해 또 다른 보호된 단량체를 적용시키는 단계. 하나 이상의 중간 단계는 산화 또는 황화를 포함한다. 일부 경우에서, 1회 이상의 세척 단계가 상기 단계 중 한 단계 또는 그들 모두 전에 또는 그 후에 진행된다.

[0150]

포스포라미다이트 기반의 올리고핵산 합성 방법은 일련의 화학적 단계를 포함한다. 일부 실시양태에서, 합성 방법 중 하나 이상의 단계는 시약 사이클링을 포함하고, 여기서, 방법 중 하나 이상의 단계는 단계에 유용한 시약을 기판에 적용시키는 것을 포함한다. 예를 들어, 시약은 일련의 액체 중착 및 진공 건조 단계에 의해 사이클링된다. 3차원 특징부를 포함하는 기판, 예컨대, 웰, 마이크로웰, 채널 등의 경우, 시약이 임의적으로 웰 및/또는 채널을 통해 기판의 하나 이상의 영역을 통해 통과하도록 한다.

[0151]

다양한 실시양태에서, 본원에 기술된 방법 및/또는 기판을 사용하여 합성된 올리고핵산은 적어도 약 20, 30, 40, 50, 60, 70, 75, 80, 90, 100, 120, 150개 이상의 염기를 포함한다. 일부 실시양태에서, 적어도 약 1 pmol, 10 pmol, 20 pmol, 30 pmol, 40 pmol, 50 pmol, 60 pmol, 70 pmol, 80 pmol, 90 pmol, 100 pmol, 150 pmol, 200 pmol, 300 pmol, 400 pmol, 500 pmol, 600 pmol, 700 pmol, 800 pmol, 900 pmol, 1 nmol, 5 nmol, 10 nmol, 100 nmol 이상의 올리고핵산이 좌위 내에서 합성된다. 본원에서 제공된, 표면 상에서 올리고핵산을 합성하는 방법을 통해 빠른 속도로 합성할 수 있다. 한 예로서, 시간당 적어도 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 70, 80, 90, 100, 125, 150, 175, 200개 이상의 뉴클레오티드가 합성된다. 뉴클레오티드로는 아데닌, 구아닌, 티민, 시토신, 우리딘 빌딩 블록, 또는 그의 유사체/변형된 베전을 포함한다. 일부 실시양태에서, 올리고핵산 라이브러리는 기판 상에서 동시에 합성된다. 예를 들어, 약 또는 적어도 약 100; 1,000; 10,000; 100,000; 1,000,000; 2,000,000; 3,000,000; 4,000,000; 또는 5,000,000개의 분할된 좌위를 포함하는 기판이 적어도 동일한 개수의 상이한 올리고핵산의 합성을 지지하며, 여기서, 상이한 서열을 코딩하는 올리고핵산은 분할된 좌위 상에서 합성된다. 일부 실시양태에서, 올리고핵산 라이브러리는 약 3개월, 2개월, 1개월, 3주, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2일, 24시간 이하인 시간 이내에 본원에 기술된 낮은 오류율로 기판 상에서 합성된다. 일부 실시양태에서, 본원에 기술된 기판 및 방법을 사용하여 낮은 오류율로 합성된 올리고핵산 라이브러리로부터 어셈블리된 더욱 큰 핵산은 약 3개월, 2개월, 1개월, 3주, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5,

4, 3, 2일, 24시간 이하인 시간 이내에 제조된다.

[0152] 일단 생성을 위해 큰 올리고핵산을 선택하였으면, 드 노보 합성을 위해 소정의 올리고핵산 라이브러리를 디자인 한다. 고밀도 올리고핵산 어레이 생성을 위한 각종의 적합한 방법이 공지되어 있다. 작업 흐름 예에서, 기판 표면 층을 제공한다. 상기 예에서, 올리고핵산 합성 프로세스를 개선시키기 위해 표면의 화학적 성질을 변경시킨다. 액체에는 반발력을 가하기 위해 표면 에너지가 낮은 면적을 생성하고, 동시에, 액체에 인력을 가하기 위해 표면 에너지가 높은 면적을 생성한다. 표면 그 자체는 평면 형태일 수 있거나, 예컨대, 표면적을 증가시키는 돌출부 또는 마이크로웰과 같은 형상 변화를 함유할 수 있다. 작업 흐름 예에서, 선택된 고에너지 표면 문자는 국제 특허 출원 공개 WO/2015/021080(이 특허는 그 전문이 본원에서 참조로 포함된다)에 개시된 바와 같이, DNA 화학법을 지원하는 이중적 기능을 한다.

[0153] 올리고핵산 어레이의 반응계내 제조는 고체 지지체 상에서 생성되고, 이는 단일 뉴클레오티드 신장 프로세스를 이용하여 다중 올리고머를 동시에 신장시킨다. 증착 장치, 예컨대, 올리고핵산 합성기는 다중 올리고핵산이 동시에 잔기를 한 번에 하나씩 신장시켜 소정의 핵산 서열을 갖는 올리고머를 생성하게 하는 단계적 방식으로 시약을 이형시킬 수 있도록 디자인된다. 일부 경우에서, 이 단계에서 올리고핵산을 표면으로부터 절단한다. 절단은 예컨대, 암모니아 또는 메틸아민을 이용한 기체 절단을 포함한다.

컴퓨터 시스템

[0155] 본원에 기술된 시스템 중 임의의 시스템은 컴퓨터에 작동가능하게 연결될 수 있고, 근거리 방식으로 또는 원격적으로 컴퓨터를 통해 자동화될 수 있다. 다양한 실시양태에서, 본 발명의 방법 및 시스템은 컴퓨터 시스템 상의 소프트웨어 프로그램 및 그의 사용을 추가로 포함할 수 있다. 따라서, 분배/진공/재충전 기능의 동시화, 예컨대, 물질 증착 장치 이동, 분배 작용 및 진공 작동의 조직화 및 동시화를 위한 전산화된 조절은 본 개시내용의 범위 내에 있다. 컴퓨터 시스템은 사용자에 의해 특정된 염기 서열과 물질 증착 장치의 위치를 접속시켜 올바른 시약을 기판의 특정된 영역에 전달하도록 프로그래밍될 수 있다.

[0156] 도 8에 도시된 컴퓨터 시스템(800)은 고정된 매체(812)를 가진 서버(809)에 임의적으로 연결될 수 있는 매체(811) 및/또는 네트워크 포트(805)로부터의 지시를 판독할 수 있는 논리 장치로서 이해될 수 있다. 예컨대, 도 8에 제시된 시스템은 CPU(801), 디스크 드라이브(803), 임의적 입력 장치, 예컨대, 키보드(815) 및/또는 마우스(816), 및 임의적 모니터(807)를 포함할 수 있다. 데이터 통신은 제시된 통신 매체를 통해 근거리 또는 원거리에 위치하는 서버 쪽으로 달성될 수 있다. 통신 매체는 데이터를 전송하고/거나, 수신하는 임의의 수단을 포함할 수 있다. 예를 들어, 통신 매체는 네트워크 연결, 무선 연결 또는 인터넷 연결일 수 있다. 상기 연결은 월드 와이드 웹(World Wide Web)상에서의 통신을 제공할 수 있다. 본 개시내용과 관련된 데이터는 도 8에 도시된 바와 같이, 관계자(822)에 의한 접수 및/또는 검토를 위해 상기와 같은 네트워크 또는 연결 상에서 전송될 수 있다는 것이 구성된다.

[0157] 도 9는 본 발명의 예시적 실시양태와 관련하여 사용될 수 있는 컴퓨터 시스템(900)의 첫 번째 예시적 아키텍처를 도시한 블록 다이어그램이다. 도 9에 도시된 바와 같이, 예시적 컴퓨터 시스템은 지시를 프로세싱하기 위한 프로세서(902)를 포함할 수 있다. 프로세서의 비제한적인 예로는 인텔(Intel) 제온(Xeon)TM 프로세서, AMD 옵테론(Opteron)TM 프로세서, 삼성(Samsung) 32-비트 RISC ARM 1176JZ(F)-S v10TM 프로세서, ARM 코텍스(Cortex)-A8 삼성 S5PC100TM 프로세서, ARM 코텍스-A8 애플 A4(ARM Cortex-A8 Apple A4)TM 프로세서, 마벨(Marvell) PXA 930TM 프로세서, 또는 기능적으로 등가인 프로세서를 포함한다. 다중의 실행 쓰레드가 병렬식 프로세싱을 위해 사용될 수 있다. 일부 실시양태에서, 다중 프로세서 또는 다중 코어를 가진 프로세서 또한 단일 컴퓨터 시스템, 클러스터, 또는 복수의 컴퓨터, 휴대 전화, 및/또는 개인 정보 단말기 장치를 포함하는, 네트워크상에 걸쳐 분산된 시스템에서 사용될 수 있다.

[0158] 도 9에 도시된 바와 같이, 고속 캐시(904)를 프로세서(902)에 연결하거나, 또는 그에 도입하여 프로세서(902)에 의해 최근에 사용되었거나, 또는 빈번하게 사용되는 지시 또는 데이터를 위한 고속 메모리를 제공할 수 있다. 프로세서(902)는 프로세서 버스(908)에 의해 노스 브리지(906)에 연결된다. 노스 브리지(906)는 메모리 버스(912)에 의해 랜덤 액세스 메모리(RAM: random access memory)(910)에 연결되고, 프로세서(902)에 의해 RAM(910)에의 접근을 관리한다. 노스 브리지(906)는 또한 칩셋 버스(916)에 의해 사우스 브리지(914)에도 연결된다. 결국, 사우스 브리지(914)는 주변 장치 버스(918)에 연결된다. 주변 장치 버스는 예를 들면, PCI, PCI-X, PCI 익스프레스(Express) 또는 다른 주변 장치 버스일 수 있다. 노스 브리지 및 사우스 브리지는 종종 프로세서 칩셋으로 지칭되고, 프로세서, RAM, 및 주변 장치 버스(918) 상의 주변 구성 요소 사이의 데이터 전달을 관리한다. 일부 대안적 아키텍처에서, 별개의 노스 브리지 칩을 사용하는 대신, 노스 브리지의 기능성을 프로세서 내

로 도입할 수 있다. 일부 경우에서, 시스템(900)은 주변 장치 버스(918)에 부착된 액셀러레이터 카드(922)를 포함할 수 있다. 액셀러레이터는 필드 프로그램 가능 게이트 어레이(FPGA: field programmable gate array), 또는 특정 프로세싱을 가속화하기 위한 다른 하드웨어를 포함할 수 있다. 예를 들어, 액셀러레이터는 적응 데이터 재구조화를 위해, 또는 확장된 세트 프로세싱에서 사용된 대수식 평가를 위해 사용될 수 있다.

[0159] 소프트웨어 및 데이터는 외부 스토리지(924) 내에 저장되고, 프로세서에 의한 사용을 위해 RAM(910) 및/또는 캐시(904) 내에 로딩될 수 있다. 시스템(900)은 시스템 자원을 관리하기 위한 운용 시스템을 포함하고; 운용 시스템의 비제한적인 예로는 본 발명의 예시적인 실시양태에 따라 데이터 저장 및 최적화를 관리하기 위한 운용 시스템의 상부 상에서 실행되는 응용 소프트웨어 뿐만 아니라, 리눅스(Linux), 윈도우즈(Windows)TM, MACOSTM, 블랙베리(BlackBerry) OSTM, iOSTM 및 다른 기능적으로 등가인 운용 시스템도 포함한다. 본 예에서, 시스템(900)은 외부 스토리지, 예컨대, 네트워크 부착 스토리지(NAS)에의 네트워크 접속을 제공하기 위해 주변 장치 버스에 연결된 네트워크 접속 카드들(NIC: network interface card)((920) 및 (921)), 및 분산된 병렬식 프로세싱을 위해 이용될 수 있는 다른 컴퓨터 시스템도 포함한다.

[0160] 도 10은 복수의 컴퓨터 시스템((1002a) 및 (1002b)), 복수의 휴대 전화 및 개인 정보 단말기(1002c) 및 네트워크 부착 스토리지(NAS)((1004a) 및 (1004b))를 포함하는 네트워크(1000)를 보여주는 다이어그램이다. 예시적인 실시양태에서, 시스템((1002a), (1002b) 및 (1002c))은 데이터 저장을 관리할 수 있고, 네트워크 부착 스토리지(NAS)((1004a) 및 (1004b))에 저장된 데이터에 대한 데이터 접근을 최적화할 수 있다. 수학적 모델은 데이터를 위해 이용할 수 있고, 컴퓨터 시스템((1002a) 및 (1002b)), 및 휴대 전화 및 개인 정보 단말기 시스템(1002c)에 걸쳐 분산된 병렬식 프로세싱을 이용하여 평가될 수 있다. 컴퓨터 시스템((1002a) 및 (1002b)), 및 휴대 전화 및 개인 정보 단말기 시스템(1002c)은 또한 네트워크 부착 스토리지((1004a) 및 (1004b))에 저장된 데이터의 적응 데이터 재구조화를 위한 병렬식 프로세싱을 제공할 수 있다. 도 10은 단지 일례만을 보여주는 것이며, 매우 다양한 다른 컴퓨터 아키텍처 및 시스템이 본 발명의 다양한 실시양태와 함께 이용될 수 있다. 예를 들어, 블레이드 서버를 이용하여 병렬식 프로세싱을 제공할 수 있다. 후면판을 통해 프로세서 블레이드를 연결하여 병렬식 프로세싱을 제공할 수 있다. 스토리지 또한 후면판에 연결될 수 있거나, 또는 별개의 네트워크 접속을 통해 네트워크 부착 스토리지(NAS)로서 연결될 수 있다. 일부 예시적인 실시양태에서, 프로세서는 별개의 메모리 공간을 유지할 수 있고, 다른 프로세서에 의한 병렬식 프로세싱을 위해 네트워크 접속, 후면판 또는 다른 연결기를 통해 데이터를 전송할 수 있다. 다른 실시양태에서, 프로세서 중 일부 또는 그들 모두는 공유 가상 주소 메모리 공간을 이용할 수 있다.

[0161] 도 11은 예시적인 실시양태에 따라 공유 가상 주소 메모리 공간을 이용하는 멀티프로세서 컴퓨터 시스템의 블록 다이어그램이다. 상기 시스템은 공유된 메모리 서브시스템(1104)에 접근할 수 있는 복수의 프로세서(1102a-f)를 포함한다. 상기 시스템은 복수의 프로그램 가능 하드웨어 메모리 알고리즘 프로세서(MAP: memory algorithm processor)(1106a-f)를 메모리 서브시스템(1104)에 도입한다. 각각의 MAP(1106a-f)는 메모리(1108a-f) 및 하나 이상의 필드 프로그램 가능 게이트 어레이(FPGA)(1110a-f)를 포함할 수 있다. MAP는 설정 가능한 기능 단위를 제공하고, 특정 알고리즘 또는 알고리즘의 일부는 각각의 프로세서로 면밀하게 조정된 프로세싱을 위해 FPGA들(1110a-f)에게 제공될 수 있다. 예를 들어, 예시적 실시양태에서, MAP를 이용하여 데이터 모델에 관한 대수식을 평가하고, 적응 데이터 재구조화를 수행할 수 있다. 이러한 예에서, 각각의 MAP는 본 목적을 위해 모든 프로세서에 의해 전체적으로 접근 가능하다. 한 구성에서, 각각의 MAP는 직접 메모리 접근(DMA: Direct Memory Access)을 이용하여 연관된 메모리(1108a-f)에 접근함으로써, 상기 메모리가 각각의 마이크로프로세서(1102a-f)로부터 독립적으로 및 비동기적으로 과제를 실행하게 할 수 있다. 이러한 구성에서, MAP는 알고리즘의 파이프라인 및 병렬식 실행을 위해 결과를 또 다른 MAP에게 직접 공급할 수 있다.

[0162] 상기 컴퓨터 아키텍처 및 시스템은 단지 예일 뿐이고, 범용 프로세서, 보조프로세서, FPGA 및 다른 프로그램 가능 논리 장치의 임의의 조합을 이용하는 시스템, 시스템 온 칩(SOC: system on chip), 응용 주문형 접적 회로(ASIC: application specific integrated circuit), 및 다른 프로세싱 및 논리 소자를 비롯한, 매우 다양한 다른 컴퓨터, 휴대 전화 및 개인 정보 단말기 아키텍처 및 시스템이 예시적인 실시양태와 관련하여 사용될 수 있다. 일부 실시양태에서, 컴퓨터 시스템 모두 또는 그의 일부가 소프트웨어 또는 하드웨어에서 실행될 수 있다. 랜덤 액세스 메모리, 하드 드라이브, 플래쉬 메모리, 테이프 드라이브, 디스크 어레이, 네트워크 부착 스토리지(NAS), 및 다른 근거리 또는 분산된 데이터 스토리지 및 시스템을 비롯한, 임의의 다양한 데이터 저장 매체가 예시적 실시양태와 관련하여 사용될 수 있다.

[0163] 예시적인 실시양태에서, 상기 또는 다른 컴퓨터 아키텍처 및 시스템 중 임의의 컴퓨터 아키텍처 및 시스템 상에서 실행하는 소프트웨어 모듈을 이용하여 컴퓨터 시스템을 실행할 수 있다. 다른 실시양태에서, 펌웨어, 프로그

램 가능 논리 장치, 예컨대, 도 11에서 참조된 필드 프로그램 가능 게이트 어레이(FPGA), 시스템 온 칩(SOC), 응용 주문형 집적 회로(ASIC), 또는 다른 프로세싱 및 논리 소자에서 시스템의 기능을 부분적으로 또는 완전히 실행할 수 있다. 예를 들어, 하드웨어 액셀러레이터 카드, 예컨대, 도 9에 도시된 액셀러레이터 카드(922)의 사용을 통한 하드웨어 가속으로 세트 프로세서(Set Processor) 및 옵티마이저(Optimizer)를 실행할 수 있다.

[0164] 하기 실시예는 당업자에게 본원에 개시된 실시양태의 원리 및 실시를 더욱 명확하게 예시하기 위해 기술된 것이며, 임의의 청구된 실시양태의 범주를 제한하는 것으로 해석되지 않아야 한다. 달리 언급되지 않는 한, 모든 부(part) 및 백분율은 중량 기준으로 한다.

실시예

[0166] 하기 실시예는 본 발명의 다양한 실시양태를 예시하기 위한 목적으로 제공되는 것이며, 어느 방식으로든 본 발명을 제한하는 것으로 의도되지 않는다. 본원에 기술된 방법과 함께 본 실시예는 이제 바람직한 실시양태를 나타내는 것이며, 예시적인 것이고, 본 발명의 범주를 제한하는 것으로 의도되지 않는다. 특히 청구범위의 범주에 의해 정의되는 바와 같이, 본 발명의 정신 내에 포함되는 그 안에서의 변형 및 다른 사용이 당업자에게 일어날 것이다.

실시예 1: 기판 표면의 작용화

[0168] 올리고핵산 라이브러리의 부착 및 합성을 지원하기 위해 기판을 작용화시켰다. 먼저, 90% H_2SO_4 및 10% H_2O_2 를 포함하는 피라나 용액을 이용하여 기판 표면을 20분 동안 습식 방식으로 세정하였다. 수개의 비커에서 DI 물을 이용하여 기판을 헹구고, DI 물 거위목형 수도꼭지하에서 5 min 동안 유지시키고, N_2 를 이용하여 건조시켰다. 이어서, 기판을 NH_4OH (1:100; 3 mL:300 mL) 중에 5 min 동안 침지시키고, 핸드건(handgun)을 이용하여 DI 물로 헹구고, 3개의 연속된 비커에서 DI 물을 이용하여 매회 1 min 동안 침지시킨 후, 다시 핸드건을 이용하여 DI 물로 헹구었다. 이어서, 기판 표면을 O_2 에 노출시켜 기판을 플라즈마 세정하였다. SAMCO PC-300 기계를 사용하여 하류 모드에서 1 min 동안 250 와트에서 O_2 플라즈마 에칭을 실시하였다.

[0169] 세정된 기판 표면을, 하기 파라미터를 이용하여 YES-1224P 기상 증착 오븐 시스템을 사용함으로써 N-(3-트리에톡시실릴프로필)-4-하이드록시부티르아미드를 포함하는 용액으로 능동적으로 작용화시켰다: 0.5 내지 1 토르, 60 min, 70°C, 135°C 증발기. 브루어 사이언스(Brewer Science) 200X 스픬 코터를 이용하여 기판 표면을 레지스트 코팅하였다. SPR™ 3612 포토레지스트를 40 sec 동안 2,500 rpm으로 기판 상에 스픬 코팅하였다. 기판을 브루어 핫 플레이트 상에서 90°C에서 30 min 동안 미리 베이킹하였다. 칼 수스(Karl Suss) MA6 마스크 얼라이너 기계를 사용하여 기판에 대해 포토리소그래피를 실시하였다. 기판을 2.2 sec 동안 노출시키고, MSF 26A 중에서 1 min 동안 발색시켰다. 잔류 발색제는 핸드건으로 헹구고, 기판을 5 min 동안 물 중에 침지시켰다. 기판을 오븐에서 100°C에서 30 min 동안 베이킹한 후, 니콘(Nikon) L200을 이용하여 리소그래피 결함에 대하여 시각적으로 검사하였다. 세정 프로세스를 이용함으로써 SAMCO PC-300 기계를 사용하여 잔류 레지스트를 제거하여 1 min 동안 250 와트에서 O_2 플라즈마 에칭을 실시하였다.

[0170] 기판 표면을 10 μ l 경질 미네랄 오일과 함께 혼합된 퍼플루오로옥틸트리클로로실란 용액 100 μ l로 수동적으로 작용화시켰다. 기판을 챔버에 배치하고, 10 min 동안 펌핑한 후, 밸브를 펌프에 대해 닫고, 10 min 동안 정치시켰다. 챔버를 대기로 환기로 시켰다. 최대 전력으로 (Crest 시스템 상에서 9) 초음파처리하면서, 70°C에서 5 min 동안 500 mL NMP 중에 2회에 걸쳐 침지시킴으로써 기판을 레지스트 스트리핑시켰다. 이어서, 최대 전력으로 초음파처리하면서, 실온에서 5 min 동안 500 mL 이소프로판을 중에 기판을 침지시켰다. 기판을 200 푸르프 에탄올 300 mL 중에 디핑하고, N_2 를 이용하여 드라이 건조시켰다. 작용화된 표면을 활성화시켰고, 이는 올리고핵산 합성을 위한 지지체로서 작용하였다.

실시예 2: 올리고핵산 합성 장치 상에서의 50-mer 서열 합성

[0172] 2차원 올리고핵산 합성 장치를 플로우셀로 조립하고, 이를 플로우셀 (Applied Biosystems "ABI394 DNA Synthesizer")에 연결시켰다. 2차원 올리고핵산 합성 장치를 N-(3-트리에톡시실릴프로필)-4-하이드록시부티르아미드(Gelest)로 균일하게 작용화시키고, 이를 이용하여 본원에 기술된 올리고핵산 합성 방법을 사용함으로써 50 bp의 예시적인 올리고핵산("50-mer 올리고핵산")을 합성하였다.

[0173] 50-mer의 서열을 서열 번호 1로 기술하였다.

[0174] 5 'AGACAATCAACCATTGGGGTGGACAGCCTGACCTCTAGACTTCGGCAT##TTTTTTTTT3' (서열 번호 1), 여기서, #는 탈보호 동안 표면으로부터 올리고핵산을 유리시킬 수 있는 절단가능한 링커인, 티미딘-숙시닐 헥사미드 CED 포스포라미다이트(ChemGenes로부터의 CLP-2244)를 나타낸다.

[0175] 하기 표 5의 프로토콜에 따른 표준 DNA 합성 화학법(커플링, 캡핑, 산화, 및 탈차단) 및 ABI 합성기를 이용하여 합성을 수행하였다.

표 5

일반 DNA 합성 프로세스 명칭	표 5	
	프로세스 단계	시간 (sec)
세척 (아세토니트릴 세척 유동)	아세토니트릴 시스템 세척	4
	플로우셀로 향하는 아세토니트릴	23
	N2 시스템 세척	4
	아세토니트릴 시스템 세척	4
DNA 염기 첨가 (포스포라미다이트 + 활성화제 유동)	활성화제 매니폴드 세척	2
	플로우셀로 향하는 활성화제	6
	플로우셀로 향하는 활성화제 + 포스포라미다이트	6
	플로우셀로 향하는 활성화제	0.5

[0176]

표 5

일반 DNA 합성 프로세스 명칭	프로세스 단계	시간 (sec)
	플로우 셀로 향하는 활성화제 + 포스포라미다이트	5
	플로우 셀로 향하는 활성화제	0.5
	플로우 셀로 향하는 활성화제 + 포스포라미다이트	5
	플로우 셀로 향하는 활성화제	0.5
	플로우 셀로 향하는 활성화제 + 포스포라미다이트	5
	25sec 동안 인큐베이션	25
세척 (아세토니트릴 세척 유동)	아세토니트릴 시스템 세척	4
	플로우 셀로 향하는 아세토니트릴	15
	N2 시스템 세척	4
	아세토니트릴 시스템 세척	4
DNA 염기 증가 (포스포라미다이트 + 활성화제 유동)	활성화제 매니폴드 세척	2
	플로우 셀로 향하는 활성화제	5
	플로우 셀로 향하는 활성화제 + 포스포라미다이트	18
	25sec 동안 인큐베이션	25
세척 (아세토니트릴 세척 유동)	아세토니트릴 시스템 세척	4
	플로우 셀로 향하는 아세토니트릴	15
	N2 시스템 세척	4
	아세토니트릴 시스템 세척	4
캡핑(CapA+B, 1:1, 유동)	플로우 셀로 향하는 CapA+B	15
세척 (아세토니트릴 세척 유동)	아세토니트릴 시스템 세척	4
	플로우 셀로 향하는 아세토니트릴	15
	아세토니트릴 시스템 세척	4
	플로우 셀로 향하는 산화제	18
세척 (아세토니트릴 세척 유동)	아세토니트릴 시스템 세척	4
	N2 시스템 세척	4
	아세토니트릴 시스템 세척	4
	플로우 셀로 향하는 아세토니트릴	15
산화(산화제 유동)	아세토니트릴 시스템 세척	4
	플로우 셀로 향하는 아세토니트릴	15
	아세토니트릴 시스템 세척	4
	플로우 셀로 향하는 아세토니트릴	15
탈차단(탈차단제 유동)	N2 시스템 세척	4
	아세토니트릴 시스템 세척	4
	플로우 셀로 향하는 아세토니트릴	18
	N2 시스템 세척	4.13
세척 (아세토니트릴 세척 유동)	아세토니트릴 시스템 세척	4.13
	플로우 셀로 향하는 아세토니트릴	15

[0177]

플로우셀을 통한 별크 시약 전달과 유사한 방식으로 포스포라미다이트/활성화제 조합을 전달하였다. 전 기간 동안 시약으로 인해 환경은 "습윤" 상태로 유지되는 바, 건조 단계는 수행하지 않았다.

[0178]

더욱 빠르게 유동할 수 있도록 하기 위해 ABI 394 합성기로부터 유동 제한기를 제거하였다. (유동 제한기가 있는 경우, 모든 시약에 대하여 ~50 μ l/sec인 것과 비교하여) 유동 제한기가 없는 경우, 아미다이트(ACN 중 0.1 M), 활성화제, (ACN 중 0.25 M 벤조일티오테트라졸("BTT"; GlenResearch로부터의 30-3070-xx)), 및 Ox(20% 피리딘, 10% 물, 및 70% THF 중 0.02 M I2)에 대한 유동 속도는 대략 ~100 μ l/sec였고, 아세토니트릴("ACN") 및 캡핑 시약(CapA 및 CapB의 1:1 믹스, 여기서, CapA는 THF/피리딘 중의 아세트산 무수물이고, CapB는 THF 중의 16% 1-메틸이미디졸이다)에 대한 유동 속도는 대략 ~200 μ l/sec였고, 탈차단제(톨루엔 중 3% 디클로로아세트산) 대한 유동 속도는 대략 ~300 μ l/sec였다. 산화제를 완전히 압출시키는 데 소요되는 시간을 관찰하였고, 그에 맞춰 화학물질 유도 시간을 위한 타이밍을 조정하였고, 추가의 ACN 세척을 상이한 화학물질 사이에 도입하였다. 올리고핵산 합성 후, 칩을 75 psi에서 밤새도록 기체 암모니아 중에서 탈보호화하였다. 물 5 방울을 표면에 가

하여 올리고핵산을 회수하였다. 이어서, 회수된 올리고핵산을 바이오애널라이저 소형 RNA 칩 상에서 분석하였다 (데이터는 나타내지 않음).

[0180] **실시예 3: 올리고핵산 합성 장치 상에서의 100-mer 서열 합성**

[0181] 50-mer 서열 합성을 위해 실시예 2에 기술된 것과 동일한 프로세스를, 첫 번째 것은 N-(3-트리에톡시실릴프로필)-4-하이드록시부티르아미드로 균일하게 작용화된 것이고, 두 번째 것은 11-아세톡시운데실트리에톡시실란 및 n-데실트리에톡시실란의 5/95 믹스로 작용화된 것인, 2개의 상이한 규소 칩 상에서 100-mer 올리고핵산("100-mer" 올리고핵산"; 5' CGGGATCCTTATCGTCATCGTGTACAGATCCCGACCCATTGCTGTCCACCAGTCATGCTAGCCATACCATGATGATGATGATGATGAGAACCCCGCAT## TTTTTTTTTT3', 여기서, #는 티미딘-숙시닐 헥사미드 CED 포스포라미다이트(ChemGenes로부터의 CLP-2244)를 나타낸다; 서열 번호 2)를 합성하는 데 사용하였고, 표면으로부터 추출된 올리고핵산을 바이오애널라이저 기계 상에서 분석하였다(데이터는 나타내지 않음).

[0182] 하기 써멀 사이클링 프로그램을 사용하여 $50 \mu\text{l}$ PCR 믹스($25 \mu\text{l}$ NEB Q5 마스터믹스, $2.5 \mu\text{l}$ $10 \mu\text{M}$ 정방향 프라이머, $2.5 \mu\text{l}$ $10 \mu\text{M}$ 역방향 프라이머, 표면으로부터 추출된 $1 \mu\text{l}$ 올리고핵산, 및 전체가 $50 \mu\text{l}$ 가 될 때까지의 물) 중에서 정방향 프라이머($5'ATGCGGGTTCTCATCATC3'$; 서열 번호 3) 및 역방향 프라이머($5'CGGGATCCTTATCGTCATCG3'$; 서열 번호 4)를 이용함으로써 상기 두 칩으로부터의 10개의 샘플 모두 추가로 PCR 증폭시켰다:

[0183] 98°C , 30 sec

[0184] 98°C , 10 sec; 63°C , 10 sec; 72°C , 10 sec; 사이클 12회 반복

[0185] 72°C , 2 min.

[0186] PCR 생성물을 또한 바이오애널라이저 상에서 전개시켰고(데이터는 나타내지 않음), 이를 통해 100-mer 위치에 뾰족한 피크가 존재한다는 것이 입증되었다. 이어서, PCR 증폭 샘플을 클로닝하고, 생어(Sanger) 서열분석하였다. 하기 표 6에는 칩 1로부터의 스폿 1-5로부터 채취된 샘플, 및 칩 2로부터의 스폿 6-10으로부터 채취된 샘플에 대하여 수행된 생어 서열분석의 결과가 요약되어 있다.

표 6

스폿	오류율	사이클 효율
1	1/763 bp	99.87%
2	1/824 bp	99.88%
3	1/780 bp	99.87%
4	1/429 bp	99.77%
5	1/1525 bp	99.93%
6	1/1615 bp	99.94%
7	1/531 bp	99.81%
8	1/1769 bp	99.94%
9	1/854 bp	99.88%
10	1/1451 bp	99.93%

[0187]

[0188] 이로써, 상이한 표면 화학법을 통해 2개의 칩 상에서 고품질이고, 균일한 합성된 올리고핵산이 반복되었다. 전체적으로, 서열분석된 262개의 100-mer 중 233개에 상응하는 89%가 오류가 없는 완벽한 서열이었다.

[0189] 마지막으로, 하기 표 7에는 스폿 1-10으로부터의 올리고핵산 샘플로부터 수득된 서열에 대한 중요한 오류 특징이 요약되어 있다.

표 7

샘플 ID/스폿 번호	OSA_00 46/1	OSA_00 47/2	OSA_00 48/3	OSA_00 49/4	OSA_00 50/5	OSA_00 51/6	OSA_00 52/7	OSA_00 53/8	OSA_00 54/9	OSA_00 55/10
총 서열	32	32	32	32	32	32	32	32	32	32
서열분석 품질	25/28	27/27	26/30	21/23	25/26	29/30	27/31	29/31	28/29	25/28
올리고 품질	23/25	25/27	22/26	18/21	24/25	25/29	22/27	28/29	26/28	20/25
ROI 매치 계수	2500	2698	2561	2122	2499	2666	2625	2899	2798	2348
ROI 돌연변이	2	2	1	3	1	0	2	1	2	1
ROI 다중 염기 결실	0	0	0	0	0	0	0	0	0	0
ROI 작은 삽입	1	0	0	0	0	0	0	0	0	0
ROI 단일 염기 결실	0	0	0	0	0	0	0	0	0	0
큰 결실 계수	0	0	1	0	0	1	1	0	0	0
돌연변이: G>A	2	2	1	2	1	0	2	1	2	1
돌연변이: T>C	0	0	0	1	0	0	0	0	0	0
ROI 오류 계수	3	2	2	3	1	1	3	1	2	1
ROI 오류율	Err: ~1/834	Err: ~1/1350	Err: ~1/1282	Err: ~1/708	Err: ~1/2500	Err: ~1/2667	Err: ~1/876	Err: ~1/2900	Err: ~1/1400	Err: ~1/2349
ROI 마이너스 프라이머 오류율	MP Err: ~1/763	MP Err: ~1/824	MP Err: ~1/780	MP Err: ~1/429	MP Err: ~1/1525	MP Err: ~1/1615	MP Err: ~1/531	MP Err: ~1/1769	MP Err: ~1/854	MP Err: ~1/1451

[0190]

실시예 4: sgRNA 디자인

[0192]

Cas9 단백질에 의해 직접적인 서열 특이적 절단을 위해 5' 단부에 가변 영역을 가지는 키메라 sgRNA 서열을 디자인하였다. 도 4a를 참조한다. sgRNA 서열은 특이적인 DNA 결합을 위한, 20개의 염기로 이루어진 염기쌍 형성 영역을 가졌고, 상기 염기쌍 형성 영역은 12개의 염기로 이루어진 시드 영역을 포함하였다. 염기쌍 형성 영역의 5' 단부는 전사 개시 부위가 되도록 디자인하였다. 염기쌍 형성 영역에 대한 3' 인접부는 Cas9 결합을 위한 dCas9 핸들 영역으로서, 그의 길이는 42개의 염기 길이였다. dCas9 핸들 영역에 대한 3' 인접부는 S. 피오게네스 터미네이터 영역으로서, 그의 길이는 40개의 염기 길이였다. dCas9 핸들 영역 및 터미네이터 영역은 각각 헤어핀 구조를 생성하는 서열을 포함하도록 디자인되었다.

[0193]

sgRNA를 또한 주형(T) 또는 비주형(NT) DNA 가닥을 표적화하도록 디자인하였다(도 5a-5b). 주형 DNA 가닥을 표적화하도록 디자인된 sgRNA는 전사된 서열과 동일한 서열 동일성을 가지는 sgRNA의 염기쌍 형성 영역을 포함하였다. 비주형 DNA 가닥을 표적화하도록 디자인된 sgRNA는 전사된 서열의 역상보체인 sgRNA의 염기쌍 형성 영역을 포함하였다.

[0194]

추가 방식에서, 디자인하였다. 가변 염기쌍 형성 영역의 상류쪽으로 바로 앞에 T7 프로모터를 디자인하였다. 도 6a-6b를 참조한다. 시험관내에서 T7 폴리머라제를 이용하여 sgRNA를 제조할 수 있도록 하기 위해 T7 프로모터 영역을 부가하였다.

[0195]

실시예 5: sgRNA를 코딩하는 DNA 합성 - 디자인 및 폴리머라제 분석

[0196]

연결시 sgRNA 서열을 코딩하는 단편으로서 DNA 올리고핵산을 디자인하였다. 도 12. 가변 서열 영역(1233)의 상류쪽으로 바로 앞에 T7 프로모터를 포함하기 위해 sgRNA를 디자인하였다. DNA 올리고핵산의 드 노보 합성 후, 증폭 반응을 수행하여 중복 단편을 연결하고 신장시켰다.

[0197] DNA 올리고핵산(1201)의 전사를 통해 시험관내에서 T7 폴리머라제를 이용하여 DNA 주형으로부터 sgRNA가 제조되었다.

[0198] 하기 표 8에 명시된 바와 같이 디자인 1(1220), 디자인 2(1222), 디자인 3(1224), 및 디자인 4(1226)에 대한 서열을 디자인하였다. 디자인 1, 디자인 2, 디자인 3, 및 디자인 4, 각각에 대한 서열은 T7 프로모터, 가변 서열 부분, 및 불변 서열 영역(핸들 및 터미네이터)를 포함한다(표 8). 구체적으로, 도 12에 제시되어 있는 바와 같은 불변 서열 영역은 염기쌍 형성 영역((1211), (1213), (1215), (1217), (1223), 및 (1225))을 포함하는 Cas9 핸들 헤어핀, 및 염기쌍 형성 영역((1219), 및 (1221))을 포함하는 터미네이터 헤어핀을 포함한다.

표 8

서열 번호	명칭	서열
12	디자인 1	TAATACGACTCACTATAGGGGATGCGCGCAGTTGTCCGT TTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCC GTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTT
13	디자인 1의 T7 프로모터	TAATACGACTCACTATA
14	디자인 1의 가변 서열	GGATGCGCGCAGTTGTCC
15	디자인 1의 핸들 및 터미네이터	GTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGT CCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTT
44	디자인 2	GAAATTAAATACGACTCACTATAGGGGATGCGCGCAGTTG TCCGTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCT AGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTG CTTT
13	디자인 2의 T7 프로모터	TAATACGACTCACTATA
14	디자인 2의 가변 서열	GGATGCGCGCAGTTGTCC
15	디자인 2의 핸들 및 터미네이터	GTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGT CCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTT
16	디자인 3	GAGCTAATACGACTCACTATAGGGGATGCGCGCAGTTG CCGTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTA GTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTG CTTT
13	디자인 3의 T7 프로모터	TAATACGACTCACTATA
14	디자인 3의 가변 서열	GGATGCGCGCAGTTGTCC
15	디자인 3의 핸들 및 터미네이터	GTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGT CCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTT
17	디자인 4	CGAGCTAATACGACTCACTATAGGGGATGCGCGCAGTTG TCCGTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCT AGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTG CTTT
13	디자인 4의 T7 프로모터	TAATACGACTCACTATA
14	디자인 4의 가변 서열	GGATGCGCGCAGTTGTCC
15	디자인 4의 핸들 및 터미네이터	GTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGT CCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTT

[0199]

[0200] 시험관내 분석을 위해, T7 RNA 폴리머라제 프로모터 영역은 T7 RNA 폴리머라제에 의해 인식되도록 이중 가닥이어야 한다는 것에 주의한다. 하이브리드화를 위해 안티센스 올리고핵산이 사용되었다: 5'-TAATACGACTCACTATAGG-3'(서열 번호 18). 추가로, 하기 표 9는 주형 및 증폭 올리고핵산의 4개의 상이한 세트의 분석을 위해 사용된 프라이머 목록을 제공한다. 도 12를 참조한다.

표 9

서열 번호	명칭 (# 염기 - 용융점)	서열
19	SgR1-R1 (80bp)	AAAAGCACCGACTCGGTGCCACTTTCAAGTTGATA ACGGACTAGCCTTATTTAACTTGCTATTCTAGCTCT AAAAC
20	SgR1-F1 (58bp - 51°C/54°C)	TAATACGACTCACTATAAGGGATGCGCGCAGTTGTCC GTTTAGAGCTAGAAAATAGCA
21	SgR1-F2 (65bp - 56°C/56°C)	GAAATTAATACGACTCACTATAAGGGATGCGCGCAGT TGTCCGTTTAGAGCTAGAAAATAGCAAG
22	SgR1-F3 (66bp - 59°C/58°C)	GAGCTAATACGACTCACTATAAGGGATGCGCGCAGTT GTCCGTTTAGAGCTAGAAAATAGCAAGTT
23	SgR1-F4 (78bp - 62°C/62°C)	GAGCTAATACGACTCACTATAAGGGATGCGCGCAGTT GTCCGTTTAGAGCTAGAAAATAGCAAGTTAAAATAAG G
24	SgR1-AR1 (14bp - 52°C)	AAAAGCACCGACTC
25	SgR1-AR2 (15bp - 56°C)	AAAAGCACCGACTCG
26	SgR1-AR3 (16bp - 60°C)	AAAAGCACCGACTCGG
27	SgR1-AR4 (17bp - 62°C)	AAAAGCACCGACTCGGT
18	sgR1-AF1 (19bp - 51°C)	TAATACGACTCACTATAAGG
28	SgR1-AF2 (24bp - 56°C)	GAAATTAATACGACTCACTATAAGG
29	SgR1-AF3 (25bp - 59°C)	GAGCTAATACGACTCACTATAAGG
30	SgR1-AF4 (25bp - 62°C)	GCGAGCTAATACGACTCACTATAAGG

[0201]

[0202]

전장의 주형의 순도 및 수율을 최적화하는 다양한 조건하에서 주형 및 증폭 올리고핵산의 4개의 상이한 세트를 분석하였다. 100 fMol의 주형 올리고핵산(SgR1-R1 & SgR1-F1, SgR1-F2, SgR1-F3, SgR1-F4), 및 600 nMol의 폴리머라제-1 PCR 농도의 각 증폭 프라이머 세트를 이용하여 10 μ l PCR 반응을 수행하였다. 에펜도르프 마스터사이클러(Eppendorf Mastercycler) 상에서의 구배를 이용하여, 25 사이클 PCR에서 2개의 고충실도 DNA 폴리머라제(폴리머라제 1 및 3) 및 표준 DNA 폴리머라제(폴리머라제 2)를 사용하여 3개의 어닐링 온도(50°C, 55°C, 60°C)를 평가하였다. 하기 표 10은 반응 조건에 관한 요약을 제공하는 것이고, 하기 표 11은 증폭 프로토콜을 제공하는 것이다.

표 10

시약 (10ul rxns)	Pol. 1		Pol. 2		Pol. 3	
	1x	15x	1x	15x	1x	15x
Pol	0.1	1.5	0.1	1.5	0.1	1.5
완충제	2	30	1	15	2	30
dNTP's	0.2	3	0.2	3	0.2	3
Amp 프라이머 (10 uM)	0.6		0.6		0.6	
주형 올리고 (100 nM)	1		1		1	
H2O	6.1	91.5	7.1	106.5	6.1	91.5
총	10	150	10	150	10	150

[0203]

표 11

폴리머라제 1, 2			폴리머라제 3		
98	30sec		95	3min	
98	10sec	25x	98	10sec	25x
50 / 55 / 60	15sec		50 / 55 / 60	15sec	
72	10sec		72	10sec	
72	30sec		72	30sec	
4	유지		4	유지	

[0204]

폴리머라제-1 PCR 반응으로부터의 결과물을 바이오애널라이저 상에서 전개시켜(데이터는 나타내지 않음) 수율을 추정하였고, 이는 하기 표 12에 요약되었다. DNA 수율은 ng/u1로 제시되어 있다(표 12). 올리고핵산 디자인 3 및 4는 각각 올리고핵산 디자인 1 및 2보다 더 높은 DNA 수율을 얻었다. 또한 어닐링 온도가 높을수록, 증가된 수율을 얻었고, 여기서, 60°C일 때, 더 높은 수율을 가졌다.

표 12

폴리머라제 1	50 °C	55 °C	60 °C
올리고 디자인 1	0	0	0
올리고 디자인 2	2.9	4.5	3.7
올리고 디자인 3	6.8	9.2	10
올리고 디자인 4	9.9	13	15.3

열거된 수율(ng/u1).

[0206]

폴리머라제 2 PCR 반응으로부터의 결과물을 바이오애널라이저 상에서 전개시켜(데이터는 나타내지 않음) 수율을 추정하였고, 이는 하기 표 13에 요약되었다. DNA 수율은 ng/u1로 제시되어 있다(표 13). 이 역시, 올리고핵산 디자인 3 및 4는 각각 올리고핵산 디자인 1 및 2보다 더 높은 DNA 수율을 얻었다. 또한 어닐링 온도가 높을수록, 증가된 수율을 얻었고, 여기서, 60°C일 때, 더 높은 수율을 가졌다.

표 13

폴리머라제 2	50 °C	55 °C	60 °C
올리고 디자인 1	0	0	0
올리고 디자인 2	7.6	5.9	6.9
올리고 디자인 3	6.1	8.5	10.5
올리고 디자인 4	7.4	11.1	19.4

열거된 수율(ng/u1).

[0208]

폴리머라제 3 PCR 반응으로부터의 결과물을 바이오애널라이저 상에서 전개시켜(데이터는 나타내지 않음) 수율을 추정하였고, 이는 하기 표 14에 요약되었다. DNA 수율은 ng/u1로 제시되어 있다(표 14). 올리고핵산 디자인 3 및 4는 각각 올리고핵산 디자인 1 및 2보다 더 높은 DNA 수율을 얻었다. 또한 어닐링 온도가 높을수록, 증가된

수율을 얻었고, 여기서, 60°C일 때, 더 높은 수율을 가졌다.

표 14

폴리머라제 3	50 °C	55 °C	60 °C
올리고 디자인 1	10	13	12.1
올리고 디자인 2	12.4	14.3	15.9
올리고 디자인 3	13.2	26.1	28.8
올리고 디자인 4	16.1	13.2	18.5

열거된 수율(ng/uL).

[0210]

요약하면, 올리고핵산 디자인 3 및 4가 3개의 폴리머라제 모두에 대해 그 이용시 증가된 DNA 수율을 얻었다. 추가로, 60°C로 어닐링 온도가 높을수록, 증가된 DNA 수율을 얻었다.

[0211]

실시예 6: CRISPR sgRNA 합성 - 온도 분석

[0212]

실시예 5로부터의 올리고핵산 프라이머를 이용하여, 실시예 5에 기술된 바와 같은 PCR 반응 실행 후, 어닐링 온도 조건 증가가 미치는 영향을 분석하였다. 증폭 생성물을 바이오애널라이저 상에서 전개시켜(데이터는 나타내지 않음) 수율을 추정하였고, 이는 하기 표 15에 요약되었다. DNA 수율은 ng/uL로 제시되어 있다(표 15). 요약하면, 폴리머라제 3이 증가된 DNA 수율을 제공하고, 60°C의 어닐링 온도를 통해 증가된 DNA 수율을 얻었다.

표 15

	올리고 디자인 3				올리고 디자인 4			
	폴리머라제 3	폴리머라제 1	폴리머라제 3	폴리머라제 1	60 °C	65 °C	60 °C	65 °C
25 사이클	27.3	15.2	11.6	3.9	28.4	29.5	13.8	10.2

열거된 수율(ng/uL).

[0213]

실시예 7: sgRNA 생성 - 구조를 띠지 않는 RNA

[0214]

T7 프로모터 서열 및 터미네이터는 포함하지만, tracrRNA 헤어핀 함유 서열은 포함하지 않는, 변형된 sgRNA 주형(120 bp)을 생성하기 위해 2개의 어셈블리 올리고핵산을 디자인하였다. 하기 표 16을 참조한다.

표 16

서열 번호	명칭	서열
31	sgR2-주형	CGAGCTAATACGACTCACTATAGGGCACAACGTGGAGG ATGGCAGCGTGAGCTGGCTGATCACTACCAGCAAAACAC TCCAATCGGTGATGGTCTGTTGCACCGAGTCGGTGCTTTT
32	sgR2-F	CGAGCTAATACGACTCACTATAGGGCACAACGTGGAGG ATGGCAGCGTGAGCTGGCTGATCACTACCAG
33	sgR2-R	AAAGCACCGACTCGGTGCAACAGGACCATCACCGATTGG AGTGTGTTGCTGGTAGTGATCAGCCAGCTG

[0215]

실시예 5에서 sgRNA를 증폭시키기 위해 사용된 것과 동일한 프라이머 올리고핵산을 이용하여 어셈블리 올리고핵산을 증폭시켰다. 사용된 반응 조건은 하기 표 17에 요약되어 있다.

五 17

성분	1x	PCR 조건	시간	사이클
Pol 1	0.1	95°C	3 min	
완충제	2	98 °C	10 sec	25 x
dNTP's	0.2	65 °C	15 sec	
Amp 프라이머 (10um)	0.6	72 °C	10 sec	
주형 올리고 (100nm)	1	72 °C	30 sec	
H2O	6.1	4 °C	유지	
총	10			

[0219]

[0220]

T7 RNA 폴리머라제를 이용하는 전사를 통해 2차 구조를 띠지 않는, 80 bp의 RNA 생성물을 수득할 것으로 예측되었다. 시험관내 전사 키트(NEB HiScribe)를 이용하여 증폭 생성물의 전사를 수행하였다. 반응 혼합물을 바이오애널라이저 상에서 분석하였다. **도 13a-13b**를 참조한다. tracrRNA 헤어핀 함유 서열을 가지는 sgRNA(**도 13a**)보다 구조를 띠지 않는 디자인을 가지는 변형된 sgRNA 생성물(**도 13b**)이 더 순수하였다.

[0221]

실시예 8: sgRNA 지정 Cas9 절단

[0222]

T7 프로모터 영역을 포함하고, 각각은 720 bp GFP 코딩 서열 영역에 대한 상이한 인식 서열을 포함하는 것인 3 개의 sgRNA 서열을 디자인하였다. sgRNA 서열을 각각 2개의 올리고핵산의 PCR로부터 어셈블리시켰다. sgRNA 백본 및 프라이머는 하기 표 18에 제공되어 있다.

五 18

[0223]

[0224] 하기 표 19에 요약되어 있는 반응 조건하에서 어셈블리 올리고핵산을 증폭시켰다.

표 19

성분	1x		PCR 조건	시간	사이클
폴리머라제 3	0.1		95°C	3 min	
완충제	2		98 °C	10 sec	25 x
dNTP's	0.2		95 °C	15 sec	
Amp 프라이머(10 um)	0.6		60 °C	15 sec	
주형 올리고(100 nm)	1		72 °C	10 sec	
H2O	6.1		72 °C	30 sec	
총	10		4 °C	유지	

[0225]

[0226] 각 sgRNA 어셈블리 반응으로부터의 샘플을 바이오애널라이저 상에서 분석하였다(도 14A-14C). T7 RNA 폴리머라제 PCR 증폭 생성물을 이용하는 전사 반응을 수행하였다. 각 반응으로부터의 샘플을 바이오애널라이저 상에서 분석하였다(도 14D-14F).

[0227]

GFP 증폭 생성물, Cas9 및 전사된 sgRNA를 이용하여 Cas9 소화물을 제조하였다. 대조군의 경우, 단일 피크가 관찰된 것과 비교하여, 3개의 모든 소화물의 경우, 2개의 피크가 관찰되었다(도 14G-14J). 3개의 합성된 sgRNA를 이용하여 Cas9 절단으로부터의 예측된 단편 및 생성된 단편이 하기 표 20에 열거되어 있다.

표 20

sgRNA	예측 단편 1	생성된 단편 1	예측 단편 2	생성된 단편 2
sgR35	321	324	439	451
sgR36	342	350	418	430
sgR37	208	137	552	560

[0228]

[0229] Cas9 소화물 샘플을 정제하고, 바이오애널라이저 상에서 다시 분석하였다(데이터는 나타내지 않음). 정제된 샘플로부터의 결과는 하기 표 21에 요약되어 있다.

표 21

sgRNA	예측 단편 1	생성된 단편 1	예측 단편 2	생성된 단편 2
sgR35	321	323	439	451
sgR36	342	353	418	427
sgR37	208	220	552	560

[0230]

[0231] 실시예 9: 29,040개의 독특한 올리고핵산의 병렬식 어셈블리

[0232] 도 15에 제시된 바와 같은, 평평한 규조 플레이트 상에 각각이 121개의 좌위를 포함하는 것인 256개의 클러스터를 포함하는 구조물(1505)을 제작하였다. 클러스터의 확대도가 121개의 좌위를 포함하는 (1510)에 제시되어 있

다. 256개의 클러스터 중 240개로부터의 좌위는 상이한 서열을 가지는 올리고핵산의 합성을 위한 부착 및 지지를 제공하였다. 실시예 3으로부터의 일반 방법을 사용하여 포스포라미다이트 화학법에 의해 올리고핵산 합성을 수행하였다. 256개의 클러스터 중 16개로부터의 좌위는 대조군 클러스터였다. 합성된 29,040개의 독특한 올리고핵산(240개의 비대조군 클러스터 x 클러스터당 121개의 올리고핵산 집단)의 전반적인 분포는 도 16a에 제시되어 있다. NGS 서열분석 결과, 합성을 위해 선택된 디자인된 올리고핵산이 100% 나타나는 것으로 확인되었다. 도 16b에 제시된 바와 같은, 각 클러스터에 대한 분포가 측정되었다. 4개의 대표적인 클러스터에서 합성된 독특한 올리고핵산의 분포는 도 17에 제시되어 있다. 합성을 위해 디자인된 올리고핵산 모두 전반적인 수준으로 존재하였고, 올리고핵산 중 99%는 평균값의 2x 이내인 존재도를 가졌으며, 이는 합성 균일도가 높다는 것을 시사하는 것이다. 이러한 동일한 관찰 결과는 전체 클러스터 수준에서 일관되었다.

[0233] 일부미나 MiSeq(Illumina MiSeq) 유전자 서열분석기를 이용하여 각 올리고핵산에 대한 오류율을 측정하였다. 29,040개의 독특한 올리고핵산에 대한 오류율 분포는 도 18a에 제시되어 있고, 평균값은 약 500개의 염기 중 1개 정도인 값이며, 여기서, 일부 오류율은 800개의 염기 중 1개 정도로 낮았다. 도 18b에 제시된 바와 같은, 각 클러스터에 대한 분포가 측정되었다. 4개의 대표적인 클러스터에서의 독특한 올리고핵산에 대한 오류율 분포는 도 19에 제시되어 있다. 29,040개의 독특한 올리고핵산의 라이브러리는 20시간 미만의 시간이 경과한 후에 합성되었다. 29,040개의 독특한 올리고핵산 모두에 걸친 GC 백분율(%) 대 올리고핵산 출현의 분석 결과, GC 함량(올리고핵산당 대략 20% 내지 85% GC)에도 불구하고, 합성이 균일하였다는 것으로 나타났다(도 20).

실시예 10: sgRNA를 코딩하는 드 노보 합성된 DNA 라이브러리의 PCR 증폭 분석

[0235] 20-80% GC로 다양한 GC 함량을 가지며, 길이가 100개의 염기 길이인, 무작위화된 서열의 올리고핵산 9,996개를 디자인하고, 실시예 9에 기술된 것과 유사한 배열을 가지는 구조로 합성하였다. PCR 증폭이 GC 출현에 미치는 효과를 측정하기 위해, 올리고핵산 집단을 고충실도 DNA 폴리머라제(DNA 폴리머라제 1)를 이용하여 6 또는 20 사이클 동안 증폭시켰다. 대안적으로, 폴리머라제 선택이 증폭 후 전체 서열 출현에 영향을 주었는지 여부를 측정하기 위해, 올리고핵산 집단을 2개의 다른 고충실도 PCR 효소를 이용하여 6, 8, 10, 또는 15 사이클 동안 증폭시켰다. PCR 증폭 후, 차세대 서열분석을 위해 샘플을 준비하고, 일부미나 MiSeq 플랫폼에서 서열분석하였다. 대략적 리드 커버리지가 100X가 될 때까지 150개의 염기쌍 SE 리드를 생성하였다. 원시 FASTQ 파일을 분석하였다. 6, 10 또는 15 사이클에 대한 올리고핵산 출현은 어느 한 폴리머라제와 함께 도 21에 도시되어 있다. 리드 서열분석시 출현 빈도에 의해 측정된 올리고핵산 균일도를 각 조건에 대해 사정하였고, 이는 하기 표 22에 요약되어 있다.

표 22

	사이클	1.5x 이내인 것의 비율(%)	2x 이내인 것의 비율(%)
폴리머라제 1	6	72.1%	92.6%
	8	76.1%	90.3%
	10	70.9%	86.6%
	15	64.1%	82.7%
폴리머라제 2	6	91.9%	98.9%
	8	89.9%	98.1%
	10	90.1%	98.4%
	15	89.2%	97.9%

[0236]

[0237] 증폭 사이클 대 평균의 10%인 역치 미만인 집단의 비율인 도 22에 제시된 바와 같이, 각각의 증폭된 올리고핵산 집단에 대한 드롭 아웃의 개수가 정량화되었다. 폴리머라제 1 드롭 아웃은 빠르게 성장한 반면, 폴리머라제 2 드롭 아웃은 비교적 일정하게 그대로 유지되었다.

[0238] GC 분포가 과다증폭에 미치는 영향을 사정하였다(도 23). 일반적으로, GC 함량이 30% 내지 70%인 올리고핵산은 Y=X인 추세선을 따라 진행하였고, 더 많은 사이클이 진행될수록 빈도는 증가하였다. GC 함량이 70% 초과인 올리

고핵산은 일반적으로 20 사이클 후에 빈도는 약간 더 높았던 반면, GC 함량이 30% 미만인 올리고핵산은 일반적으로 6 사이클 후에 빈도가 약간 더 높았다.

[0239] 실시예 11: 인간 후생적 CRISPR 스크린

CRISPR-Cas9를 이용하여 기능성 도메인을 코딩하는 엑손 내로 돌연변이를 도입하기 위해 sgRNA 스크린을 수행하였다. 실시예 9에 기술된 것과 유사한 배열을 가지는 구조물 상에서 실시예 1에 기술된 바와 같은 규소 칩 상에서 실시예 3에 기술된 것과 유사한 방법을 이용하여 약 10,000개의 DNA 올리고핵산을 드 노보 합성하였다. 종합해 보면, 올리고핵산의 총 오류율은 약 1:500 이하였다. 각 올리고핵산의 길이는 최대 200개의 염기 길이였고, 올리고핵산 종 개당 적어도 1 fmole이 생성되었다. 올리고핵산을 PCR 증폭시키고, 백터로 클로닝시키고, sgRNA 전사를 위해 배양된 세포 내로 전기천공시켰다. 세포로부터 핵산을 단리시키고, 차세대 서열분석을 이용하여 서열분석하였다.

[0241] 서열분석 결과, sgRNA가 최소한의 편향성을 가지고, 고충실도로 제조되는, 고도로 정확하고, 균일한 라이브러리 합성인 것으로 나타났다. 경쟁자 풀과 비교하였을 때, 최소 서열분석 결과, 가이드 서열당 리드가 더 많고, 정확한 서열을 가지는 sgRNA의 회수율은 30% 더 높았다. 하기 표 23을 참조한다. 풀링된 서열분석 결과는 6 로그인 어레이 기반 경쟁자 풀과 비교하였을 때, 가이드 서열당 리드가 더 많고, 리드 분포가 훨씬 더 조밀(4 로그)한 것으로 나타났다. 도 24a-24b를 참조한다. 클론의 서열분석 입증 결과 sgRNA 회수율은 100%이고(도 24a), 상업적으로 이용가능한 어레이 기반 풀과 비교하였을 때, 서열 정확도는 더 높은 것으로 나타났다(도 24b). 서열분석된 클론 중, 정확한 sgRNA 서열의 경우, 유의적으로 더 많이 회수되었다. 하기 표 23을 참조한다. 올리고핵산 집단 중에서 소정의 서열이 100% 출현하였다. sgRNA 클론의 NGS 기반 입증 결과, sgRNA 회수율은 100%이고, 경쟁자 집단과 비교하였을 때, 클론당 합성의 정확도는 13% 더 높은 것으로 나타났다(데이터는 나타내지 않음).

표 23

	합성된 올리고 집단	상업적으로 이용가능한 올리고 집단
회수된 sgRNA 올리고	100%	>95.5%
정확한 서열 비율 (MiSeq)	약 87%	약 74%
정확한 서열 (생어 10 개의 클론)	약 100%	약 70%
클로닝된 올리고 집단 중 sgRNA 1 개당 평균 리드 (100x 정규화)	약 256	약 1024

[0242]

[0243] 실시예 12: 전체 계층 sgRNA 라이브러리

101,000개의 상이한 올리고핵산(20,200개의 유전자 표적당 5개의 sgRNA)에 대한 클론을 생성하기 위해 sgRNA를 코딩하는 DNA를 포함하도록 DNA 라이브러리를 디자인하였다. 실시예 9에 기술된 것과 유사한 배열을 가지는 구조물 상에서 실시예 1에 기술된 바와 같은 규소 칩 상에서 실시예 3에 기술된 것과 유사한 방법을 이용하여 101,000개의 올리고핵산을 드 노보 합성하였다. 합성된 올리고핵산을 PCR 증폭시키고, 소화시키고, 렌티바이러스 백터로 클로닝하고, 세포 내로 형질전환시켰다. 핵산을 세포로부터 단리시키고, 차세대 서열분석을 이용하여 서열분석하였다. 대안적으로, 합성된 올리고핵산을 PCR 증폭시켜 앰플리콘 기반 라이브러리를 형성하고, 서열분석하였다.

[0245] 차세대 서열분석 리드 대 회수된 sgRNA의 개수에 관한 플롯은, 올리고핵산 풀 크기가 증가함에 따라, 올리고핵산 집단은 상업적으로 이용가능한 어레이 기반 참조 올리고핵산 집단과 비교하였을 때, 최소한의 테일을 보이며, 전체 라이브러리에 걸쳐 더욱 균일하고, 더욱 조밀한 리드 분포를 유지하였다. 도 25.

[0246] 실시예 13. 표적화 및 활성이 개선된 sgRNA 라이브러리의 디자인

[0247] sgRNA 라이브러리를 디자인하고, 실시예 9에 기술된 것과 유사한 배열을 가지는 구조물 상에서 실시예 1에 기술된 바와 같은 규소 칩 상에서 실시예 3에 기술된 것과 유사한 방법을 이용하여 드 노보 합성하였다. 합성된 올리고핵산을 PCR 증폭시키고, 소화시키고, 백터로 클로닝하고, 스크리닝 및 분석을 비롯한 하류 적용에 사용하기 위해 세포 내로 전달하였다.

[0248] NAG PAM, NGG PAM, 고활성, 낮은 오프 타겟, 및 여과를 특징으로 하는 라이브러리를 포함하는 상이한 sgRNA 디

자인 파라미터를 비교하였다. 본원에 기술된 방법에 의해 디자인된 sgRNA 라이브러리는 다른 상업적으로 이용 가능한 gRNA 시스템과 비교하였을 때, 약 16%의 sgRNA로, 더 높은 비율의 sgRNA를 제공하였고, 그 결과, 유전자 발현은 적어도 2배 감소되었다. 도 26a. sgRNA 라이브러리는 또한 다른 상업적으로 이용 가능한 gRNA 시스템과 비교하였을 때, 약 17%로, 더 낮은 비율의 sgRNA를 제공하였고, 그 결과, 유전자 발현은 0 또는 음의 값의 감소를 나타내었다. 도 26b.

[0249] 또한, sgRNA 매개 감소를 필수 유전자 발현 수준에 대해 사정하였고, 여기서, 하기 유전자: PCNA, PSMA7, RPP21, 및 SF3B3은 sgRNA에 의해 표적화되는 것이었다. 적어도 2배 감소를 보인 sgRNA의 개수를 분석한 결과, 콤파레이터 1, 콤파레이터 2, 및 콤파레이터 3과 비교하였을 때, sgRNA 라이브러리가 더 높은 비율로 필수 유전자 감소 sgRNA를 가졌다. 하기 표 24를 참조한다.

표 24

유전자 명칭	콤파레이터 1	콤파레이터 2	콤파레이터 3	CRISPR 라이브러리
PCNA	1/5	2/6	5/9	5/5
PSMA7	1/5	0/6	2/9	3/5
RPP21	3/5	1/6	4/9	2/5
SF3B3	0/5	0/6	4/9	3/5
평균 (%)	25%	12.5%	42%	65%

[0250]

실시예 14. MS2에 대한 sgRNA 라이브러리

[0252] C2c2 단백질에 의한 서열 특이적 절단을 위해 sgRNA를 코딩하는 동일하지 않은 DNA 서열을 포함하는 DNA 라이브러리를 디자인하였다. 라이브러리는 박테리오파지 MS2 계놈의 C2c2 표적화를 위한 모든 가능한 스페이서 서열을 포함한다. 렙토트리키아 샤히이로부터의 C2c2의 성숙한 crRNA는 28개의 뉴클레오티드로 이루어진 최대 스페이서 길이를 포함하지 때문에, 박테리오파지 계놈 중 모든 가능한 28개의 뉴클레오티드 표적 부위를 타일링하였을 때, 약 3,500개의 스페이서 서열로 이루어진 라이브러리가 생성되었다.

[0253] 실시예 9에 기술된 것과 유사한 배열을 가지는 구조물 상에서 실시예 1에 기술된 바와 같은 규소 칩 상에서 실시예 3에 기술된 것과 유사한 방법을 이용하여 약 3,500개의 동일하지 않은 올리고핵산을 드 노보 합성하였다. 약 3,500개의 서열로 이루어진 라이브러리를 벡터 내로 삽입하고, E. 콜라이로 형질전환시켰다. 3개의 MS2 희석 액을 이용하여 MS2 박테리오파지로 E. 콜라이 세포를 감염시켰다. 이어서, MS2 감염에 대한 내성을 E. 콜라이에 부여한 서열에 대하여 라이브러리를 스크리닝하였다.

[0254] 다수의 스페이서 서열이 내성을 부여한 것으로 나타났다. 스페이서 출현(crRNA 빈도)과 비교하여, 다수의 스페이서 서열이 3개의 MS2 희석 액 감염에서 1.25 로그₂ 배 초과의 강화를 보인 반면, 비표적화 스페이서 서열은 강화되지 않은 것으로 나타났다.

실시예 15: 제브라파시에 대한 sgRNA 라이브러리

[0256] 약 130,000개의 sgRNA를 코딩하는 서열을 이용하여 DNA 라이브러리를 디자인한다. 평균적으로, 각 제브라파시 유전자에 대해 약 5개의 sgRNA 주형이 디자인된다. 실시예 9에 기술된 것과 유사한 배열을 가지는 구조물 상에서 실시예 1에 기술된 바와 같은 규소 칩 상에서 실시예 3에 기술된 것과 유사한 방법을 이용하여 올리고핵산을 드 노보 합성한다. 드 노보 합성으로, 각각은 규소 플레이트 표면 상의 상이한 좌위로부터 신장되는 130,000개의 올리고핵산이 제조된다. 올리고핵산을 플레이트로부터 제거하고, PCR에 의해 증폭시키고, 발현 벡터로 클로닝한다. 각 주형에 대해 서열분석을 수행한다. sgRNA 라이브러리를 제브라파시 배아 내로 주사한다. 제브라파시를 성체가 될 때까지 기른다. 이어서, 정자를 냉동보존하고, 서열분석하여 스크리닝함으로써 생식세포계열을 통해 전달된 삽입 및 결실의 서열을 확인한다. 생식세포계열 스크린 후, 경쟁적 대립유전자 특이적 PCR에 의해 정자의 유전자형을 분석한다.

실시예 16: 마우스에 대한 gRNA 라이브러리

[0258] 약 100,000개의 sgRNA를 코딩하는 서열을 이용하여 DNA 라이브러리를 디자인한다. 평균적으로, 마우스 유전자당 약 5개의 sgRNA 주형이 디자인된다. 실시예 9에 기술된 것과 유사한 배열을 가지는 구조물 상에서 실시예 1에 기술된 바와 같은 규소 칩 상에서 실시예 3에 기술된 것과 유사한 방법을 이용하여 올리고핵산을 드 노보 합성

한다. sgRNA 서열을 코딩하는 sgRNA 라이브러리를 드 노보 합성하여 100,000개의 올리고핵산을 생성한다. 드 노보 합성으로, 각각은 규소 플레이트 표면 상의 상이한 좌위로부터 신장되는 100,000개의 올리고핵산이 제조된다. 올리고핵산을 플레이트로부터 제거하고, PCR에 의해 증폭시키고, 벡터로 클로닝한다. 각 주형에 대해 서열분석을 수행한다. 서베이어 뉴클레아제 검정법 또는 서열분석에 의해 sgRNA 온 타겟 효율을 확인한다. 이어서, 원하는 유전적 배경을 가지는 마우스 접합체에 sgRNA를 미량주사한다. 대안적으로, sgRNA 효율 확인 후, sgRNA를 예컨대, 아데노 연관 바이러스(AAV: adeno-associated virus)와 같은 바이러스 벡터 내로 패키징한다. 이어서, sgRNA를 마우스의 원하는 위치에 주축성 방식으로 전달한다. 미리 선택된 표적 유전자에 대한 발현 수준을 마우스로부터 수집된 조직에서 관찰한다.

[0259] 실시예 17: 수용체 티로신 키나제에 대한 gRNA 라이브러리

총 290개의 상이한 DNA 올리고핵산에 달하는, 하기 표 25에 열거된 58개의 인간 수용체 티로신 키나제에 대한 유전자를 표적화하는 5개의 sgRNA를 코딩하는 서열을 이용하여 DNA 올리고핵산 라이브러리를 디자인한다. 실시예 9에 기술된 것과 유사한 배열을 가지는 구조물 상에서 실시예 1에 기술된 바와 같은 규소 칩 상에서 실시예 3에 기술된 것과 유사한 방법을 이용하여 올리고핵산을 드 노보 합성한다. 올리고핵산을 플레이트로부터 제거하고, PCR에 의해 증폭시키고, 벡터로 클로닝하고, 미리 선택된 세포 집단으로 전달한다. 각각의 미리 선택된 세포 집단에서의 표 25에 열거된 미리 선택된 유전자에 대한 발현 수준을 키나제 특이적 sgRNA가 없는 대조군 벡터에 노출된 대조군 세포 집단 대비로 비교한다.

표 25

유전자	Hs NT ACC #	Hs PROT ACC#
ALK	NM_004304	NP_004295
LTK	NM_002344	NP_002335
AXL	NM_001699	NP_001690
MER	NM_006343	NP_006334
TYRO3	NM_006293	NP_006284
DDR1	NM_013993	NP_001945
DDR2	NM_006182	NP_006173
EGFR	NM_005228	NP_005219

[0261]

ERBB2	NM_004448	NP_004439
ERBB3	NM_001982	NP_001973
ERBB4	NM_005235	NP_005226
EPHA1	NM_005232	NP_005223
EPHA2	NM_004431	NP_004422
EPHA3	NM_005233	NP_005224
EPHA4	NM_004438	NP_004429
EPHA5	L36644	P54756
EPHA6	AL133666	
EPHA7	NM_004440	NP_004431
EPHA8	AB040892	CAB81612
EPHB1	NM_004441	NP_004432
EPHB2	AF025304	AAB94602
EPHB3	NM_004443	NP_004434
EPHB4	NM_004444	NP_004435
EPHB6	NM_004445	NP_004436
EPHX		
FGFR1	M34641	AAA35835
FGFR2	NM_000141	NP_000132
FGFR3	NM_000142	NP_000133
FGFR4	NM_002011	NP_002002
IGF1R	NM_000875	NP_000866
INSR	NM_000208	NP_000199
INSR	J05046	AAC31759
MET	NM_000245	NP_000236
RON	NM_002447	NP_002438
MUSK	NM_005592	NP_005583
CSF1R	NM_005211	NP_005202
FLT3	NM_004119	NP_0041110
KIT	NM_000222	NP_000213
PDGFRA	NM_006206	NP_006197
PDGFRB	NM_002609	NP_002600
PTK7	NM_002821	NP_002812
RET	X12949	P07949
ROR1	NM_005012	NP_005003
ROR2	NM_004560	NP_004551
ROS1	NM_002944	NP_002935
RYK	S59184	AAB26341
TEK	NM_000459	NP_000450
TIE	NM_005424	NP_005415
NTRK1	NM_002529	NP_002520
NTRK2	NM_006180	NP_006171
NTRK3	NM_002530	NP_002521
VEGFR1	NM_002019	NP_002010
VEGFR2		AAB88005
VEGFR3	NM_002020	NP_002011
AATYK	NM_004920	NP_004911

[0262]

AATYK2	NM_014916	NP_055731
AATYK3		
DKFZp761P1010	NM_018423	NP_060893

[0263]

[0264] 실시예 18: 인간 키놈(kinome)에 대한 gRNA 라이브러리

총 2,590개의 상이한 DNA 올리고핵산에 달하는, 518개의 인간 키나제에 대한 유전자를 표적화하는 5개의 sgRNA를 코딩하는 서열을 이용하여 DNA 올리고핵산 라이브러리를 디자인한다. 올리고핵산을 플레이트로부터 제거하고, PCR에 의해 증폭시키고, 벡터로 클로닝하고, 미리 선택된 세포 집단으로 전달한다. 각각의 미리 선택된 세포 집단에서의 518개의 미리 선택된 유전자에 대한 발현 수준을 키나제 특이적 sgRNA가 없는 대조군 벡터에 노출된 대조군 세포 집단 대비로 비교한다.

[0266] 실시예 19: 인간 포스파토ーム(phosphatome)에 대한 gRNA 라이브러리

[0267]

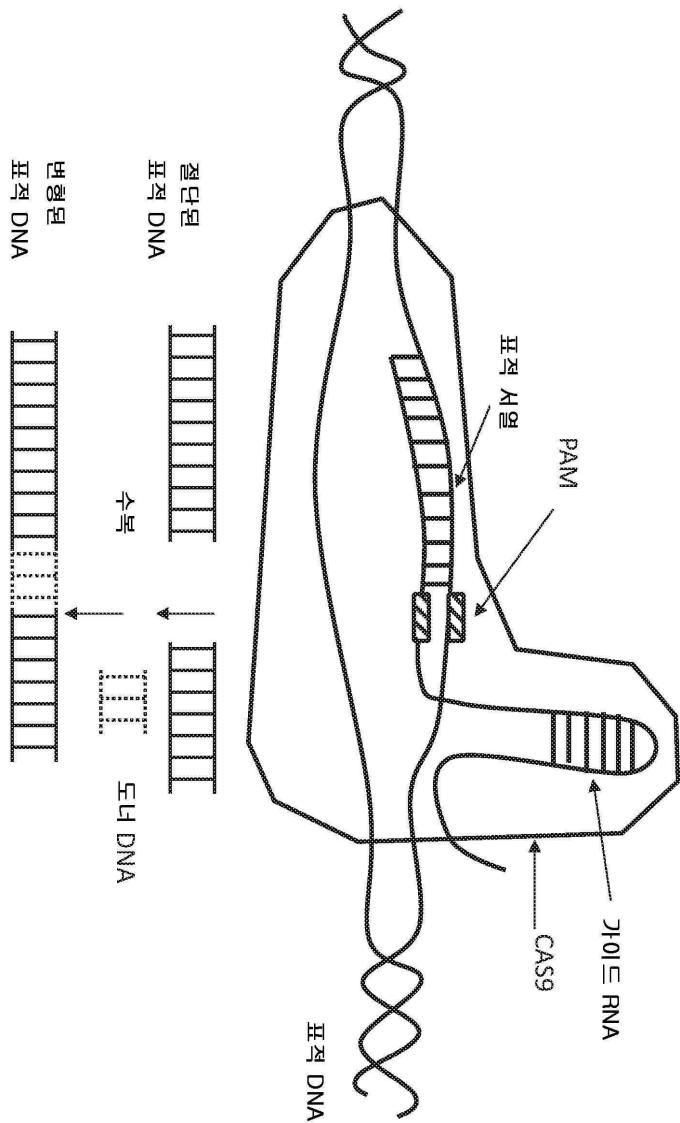
총 1,000개의 상이한 DNA 올리고핵산에 달하는, 200개의 인간 포스파타제에 대한 유전자를 표적화하는 5개의 sgRNA를 코딩하는 서열을 이용하여 DNA 올리고핵산 라이브러리를 디자인한다. 올리고핵산을 플레이트로부터 제거하고, PCR에 의해 증폭시키고, 벡터로 클로닝하고, 미리 선택된 세포 집단으로 전달한다. 각각의 미리 선택된 세포 집단에서의 200개의 미리 선택된 유전자에 대한 발현 수준을 키나제 특이적 sgRNA가 없는 대조군 벡터에 노출된 대조군 세포 집단 대비로 비교한다.

[0268]

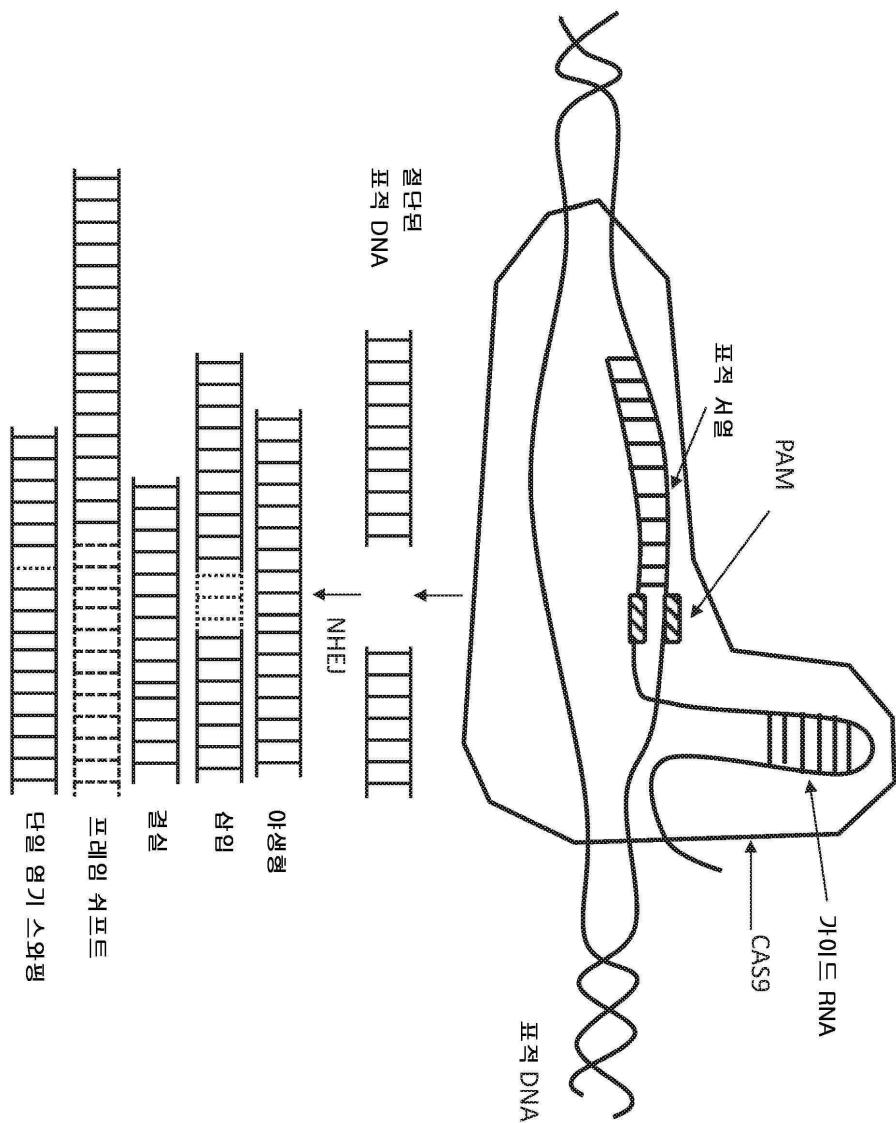
본 발명의 바람직한 실시양태가 본원에 제시 및 기술되어 있지만, 당업자에게는 상기 실시양태는 단지 예로서 제공된 것이라는 것이 자명할 것이다. 이에 본 발명을 벗어나지 않으면서 다수의 변형, 변화, 및 치환이 당업자에게 이루어질 수 있을 것이다. 본원에 기술된 본 발명의 실시양태에 대한 다양한 대안이 본 발명을 실시하는데 사용될 수 있다는 것을 이해하여야 한다. 하기 특허청구범위는 본 발명의 범주를 정의하고, 이들 특허청구범위의 내의 방법 및 구조, 및 그의 등가물이 특허청구범위에 의해 포함되는 것으로 의도된다.

도면

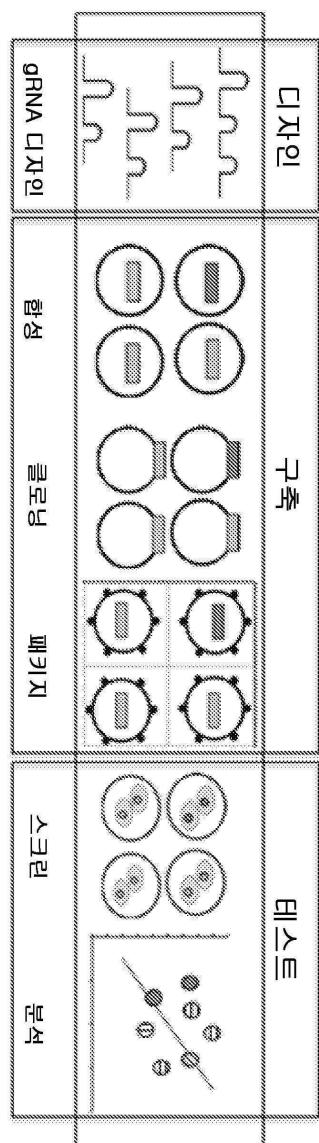
도면 1a



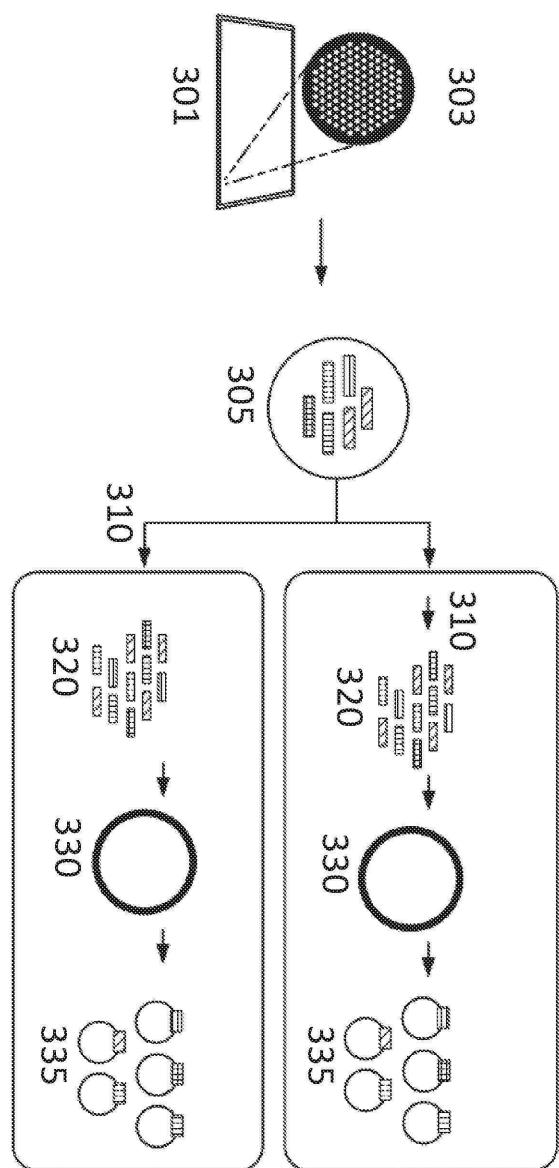
도면 1b



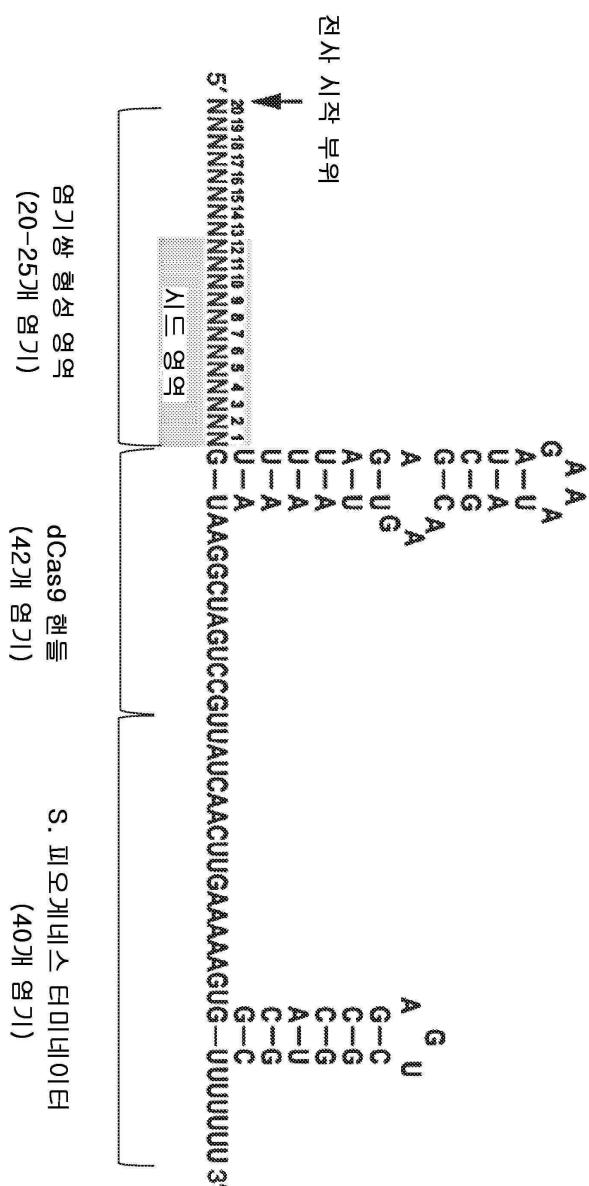
도면2



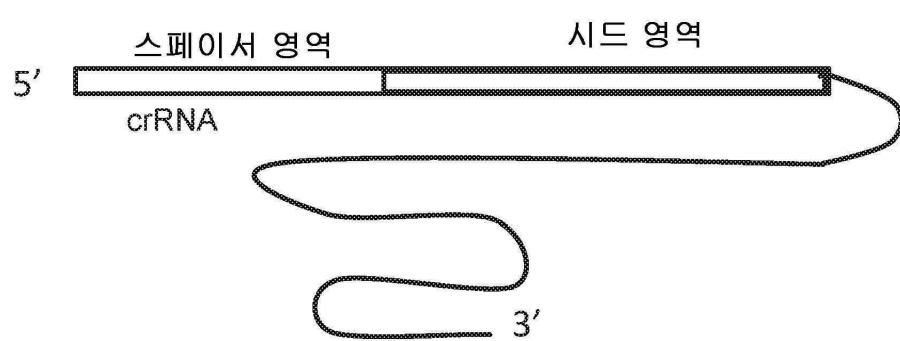
도면3



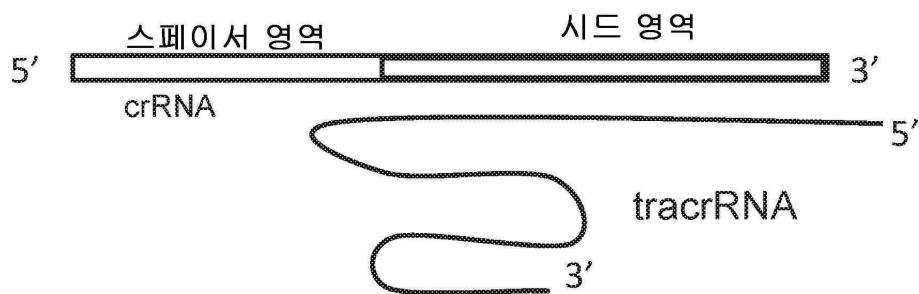
도면4a



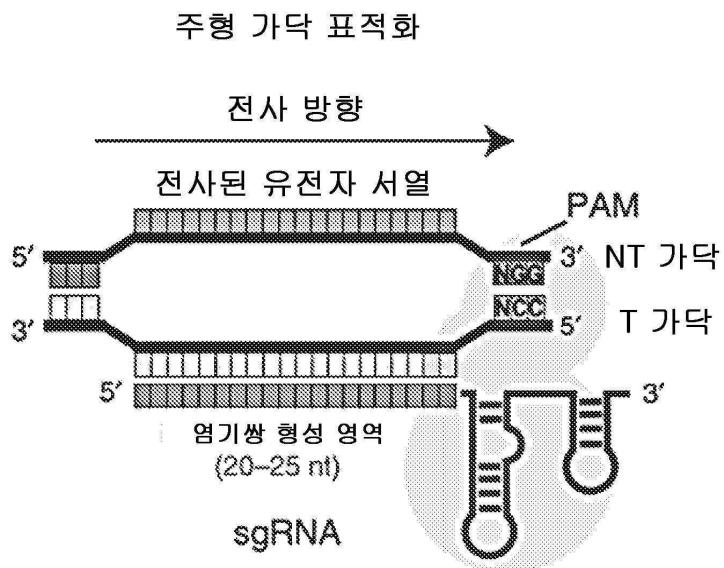
도면4b



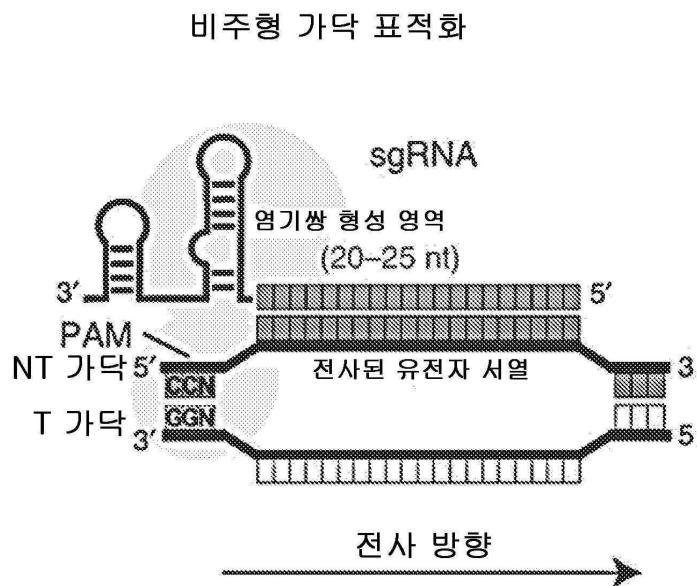
도면4c



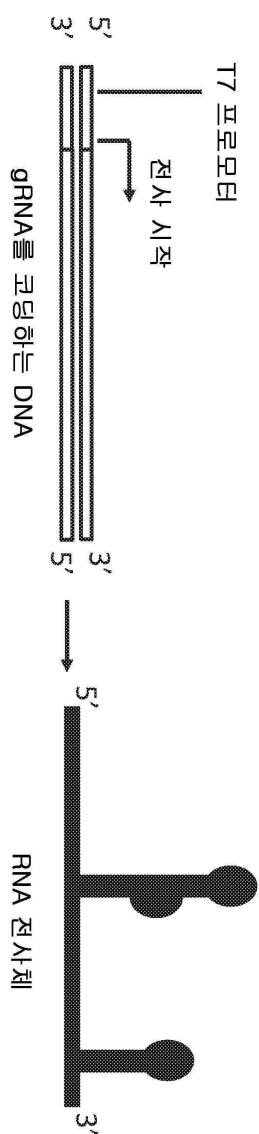
도면5a



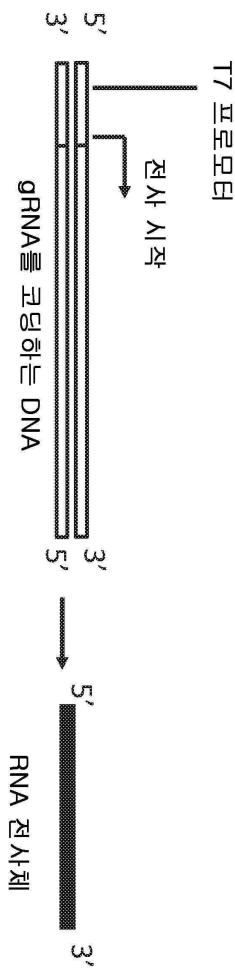
도면5b



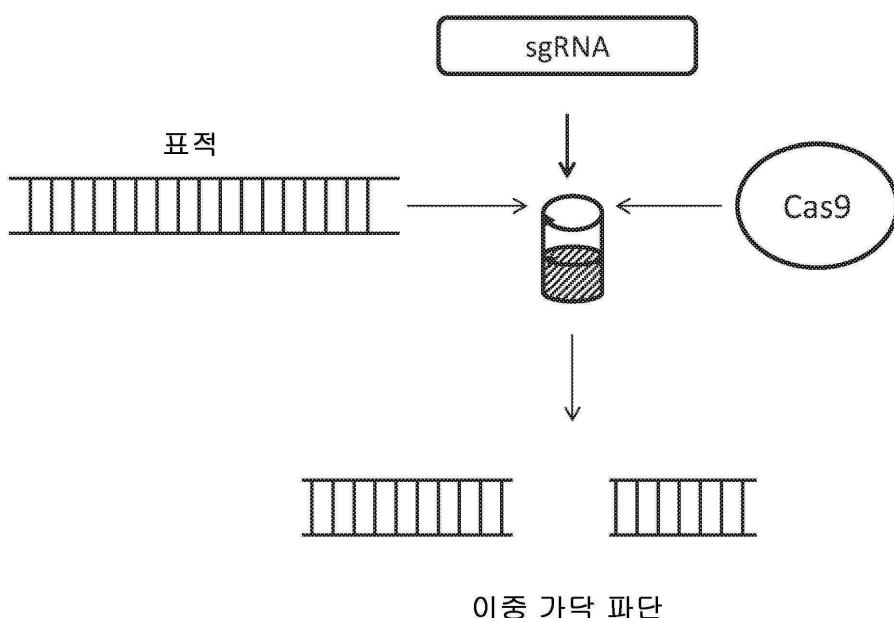
도면 6a



도면6b

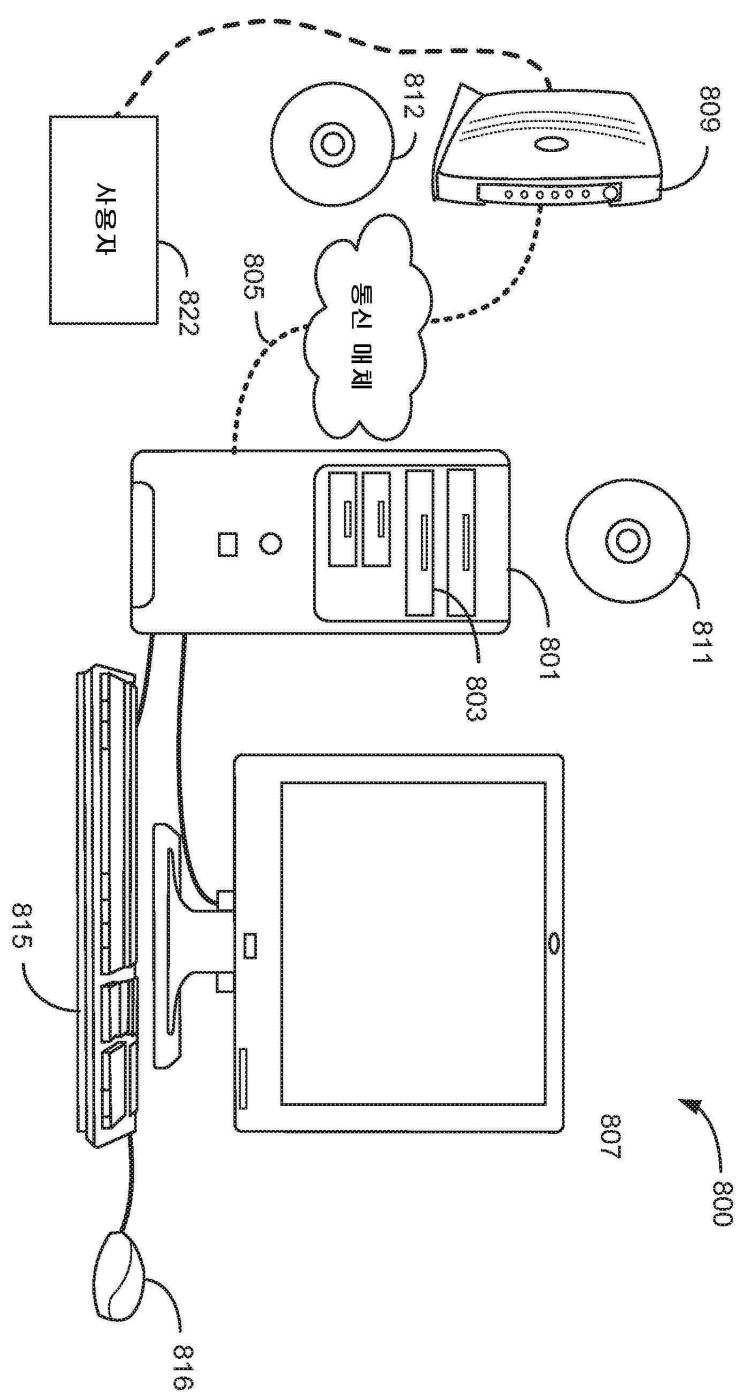


도면7

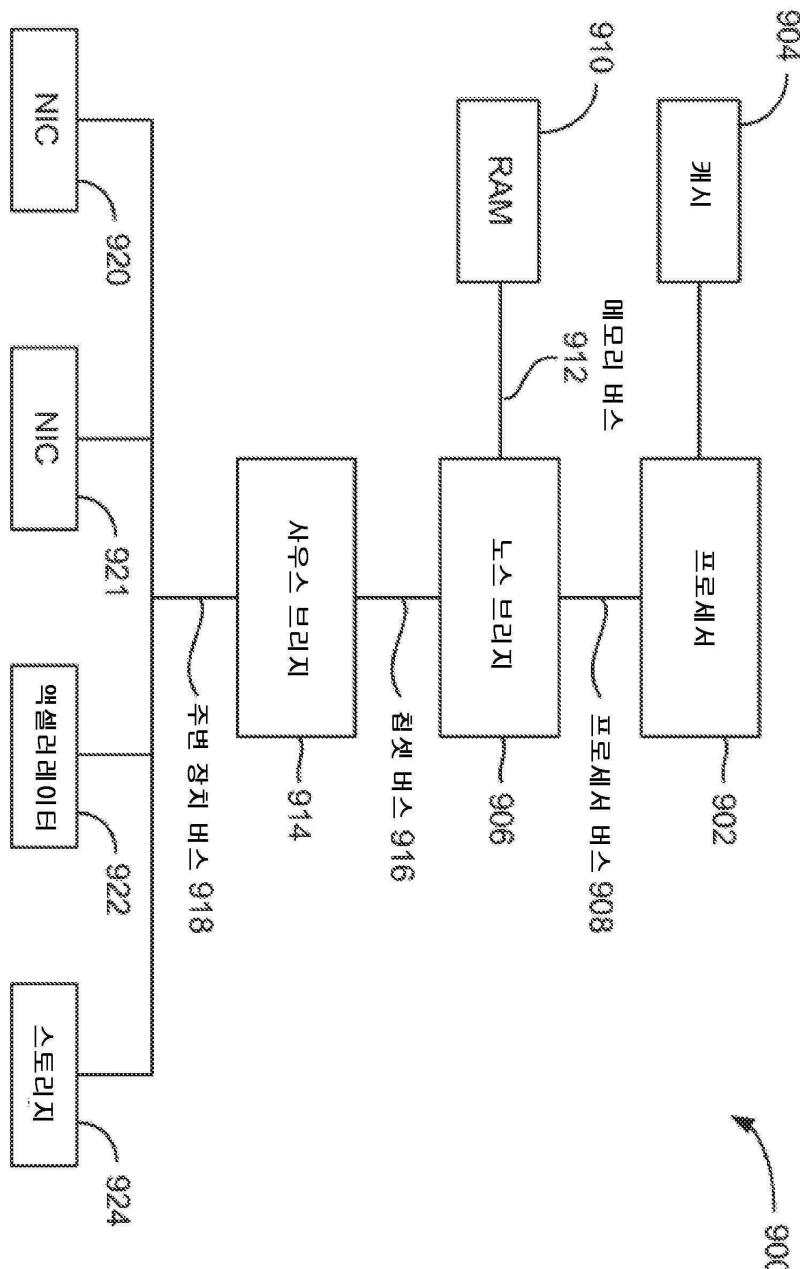


이중 가닥 파단

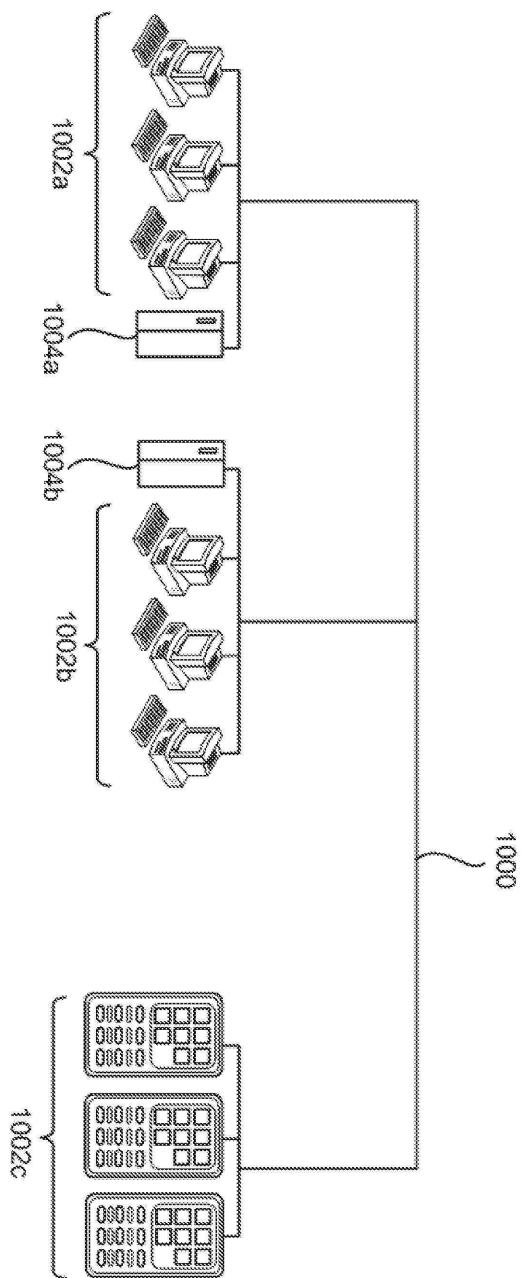
도면8



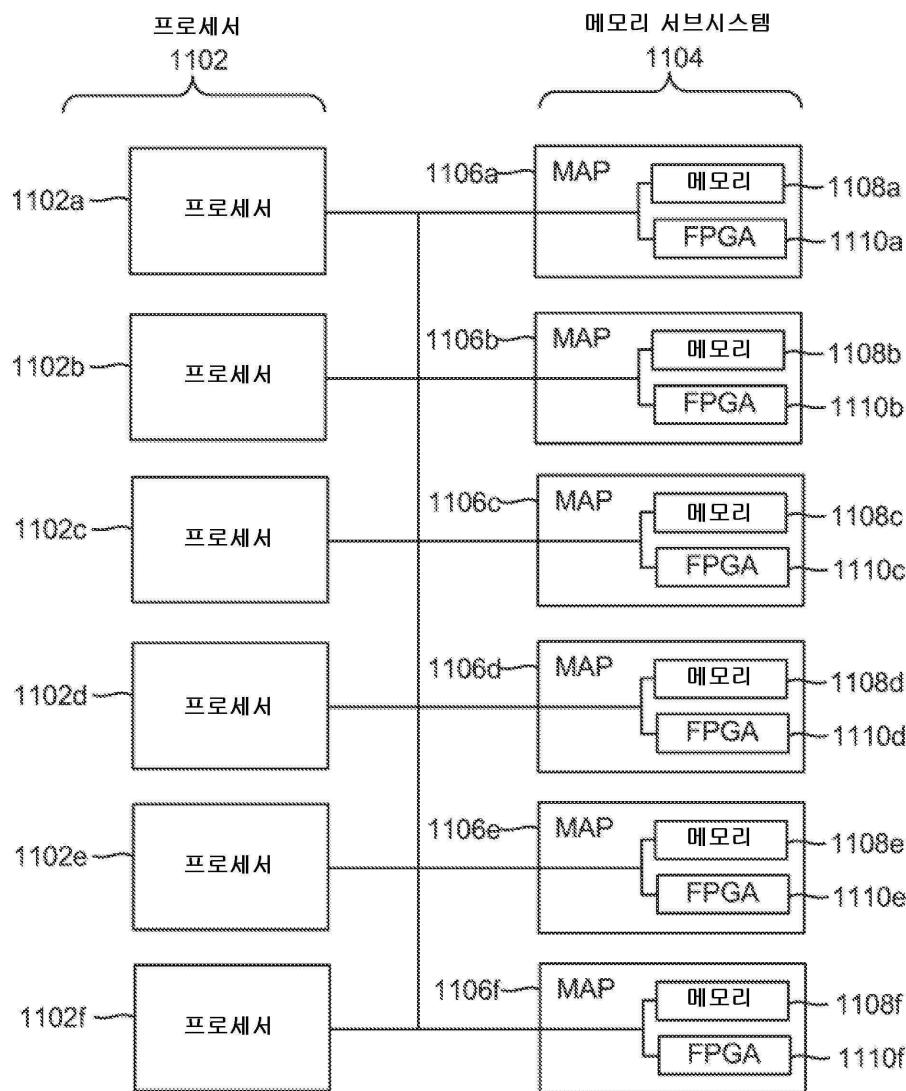
도면9



도면10



도면11



도면12

1차인

5' TAATACGACTCACTATAGGGGGATGCCGAGTGTCCGGTTAGAGCTAGAAATAGCA3'
 5' GTTTTAGAGCTAGAAATAGCAAGTAAATAGCAACCGAGTCGGCTT13

21자인

5'GAAATTAACTACGACTCTACTATAGGGATGCGCCGAGTGTCCGTTAGAGCTAGAAATAGCAAG' 5'GTTTAGAGCTAAATAGCAAGTTAAA[...]TGCCACCGAGTCGGTGTCTT3

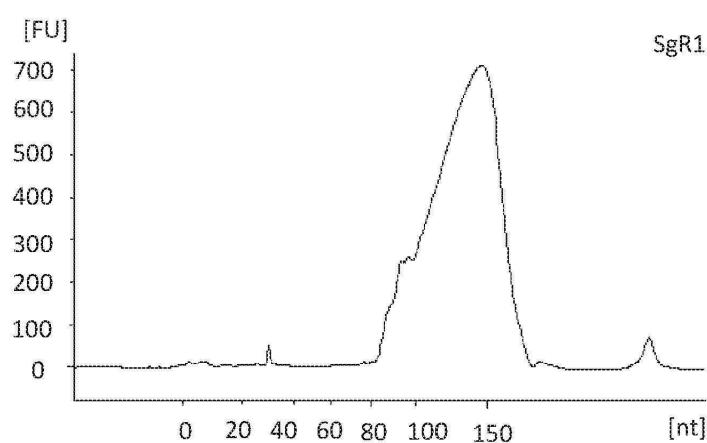
3인자

5'GAGCTAAATACGACTCACTATAGGGGATGCGCGAGTTGTCGTTAGAGCTAGAAATAGCAAGT3'
 5'GTTTAGAGCTAGAAATAGCAAGTAAAG...TTGGGACCGAGGTGGTGTCTT3'

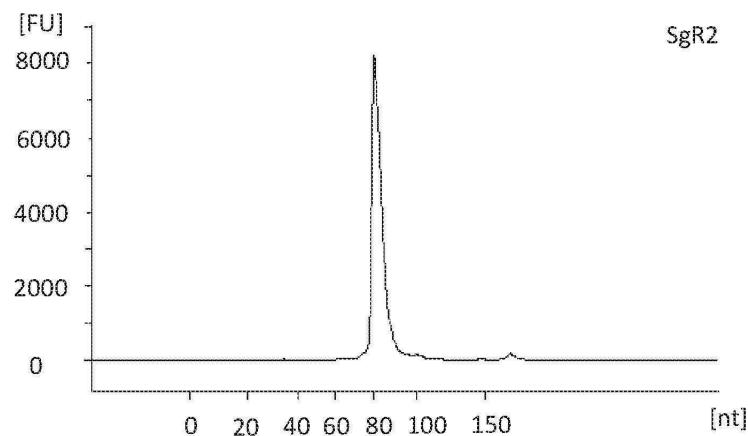
디자인 4

□ 자료 1, 2, 3: [...] = TAAGGGTAGTCCGGTATCAACTTGAAAAAA
□ 자료 4: [...] = CTAGTCCGGTTATCAACTTGAAAAAGTG

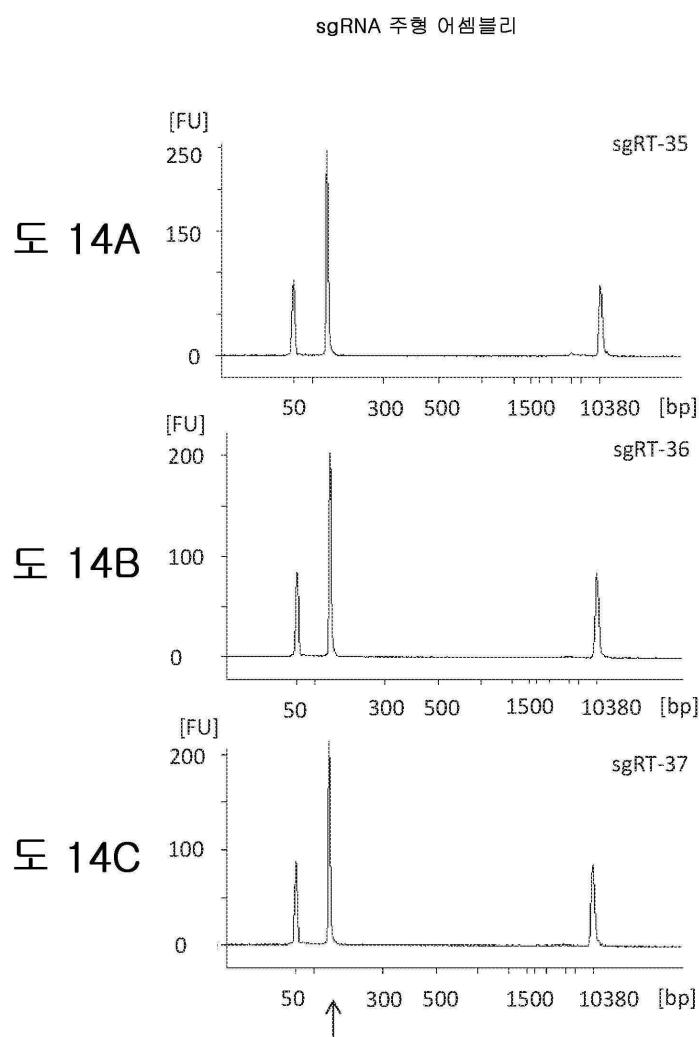
도면 13a



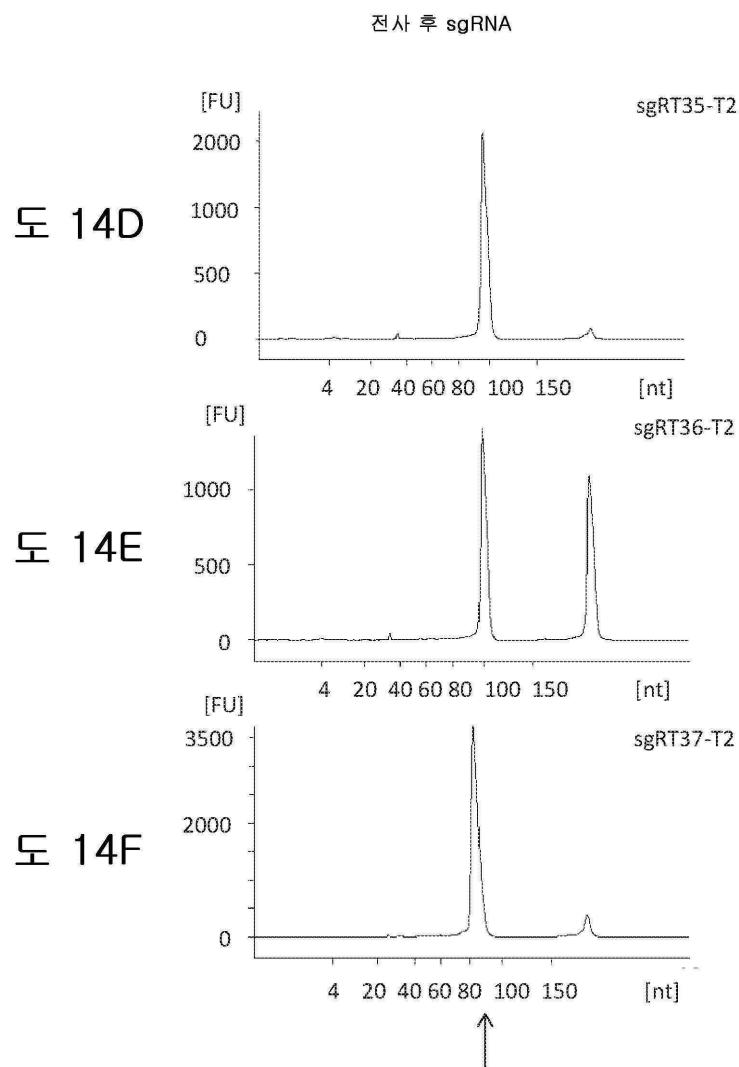
도면13b



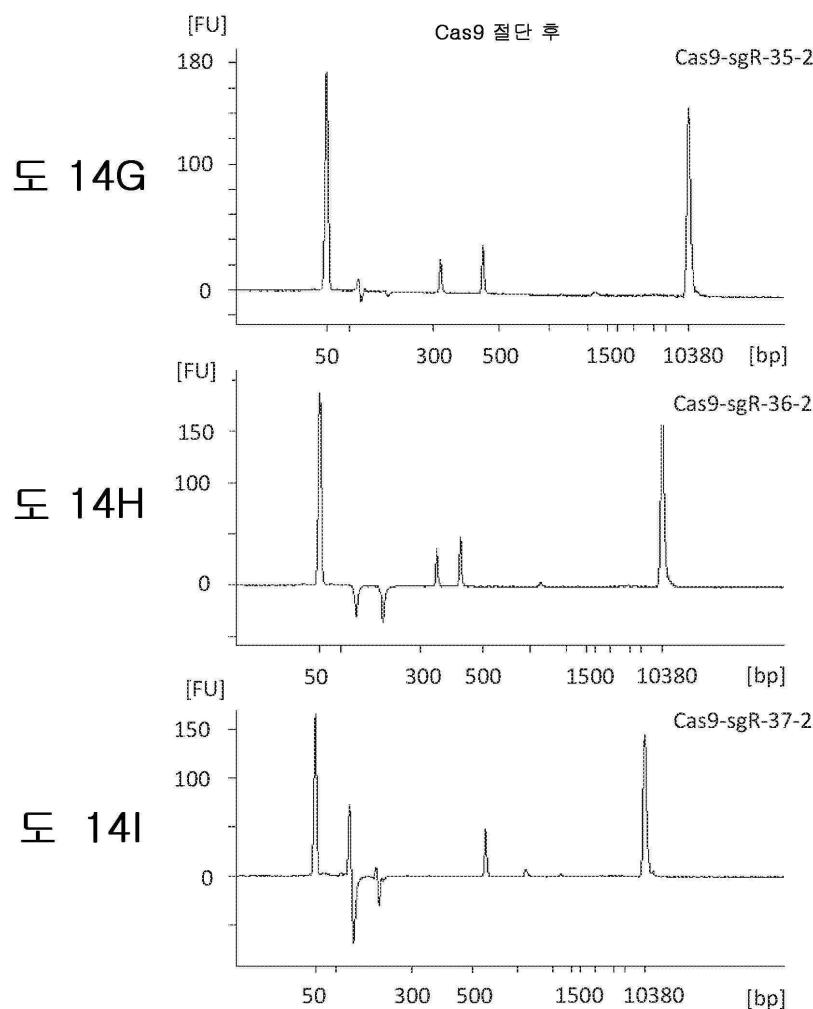
도면14a



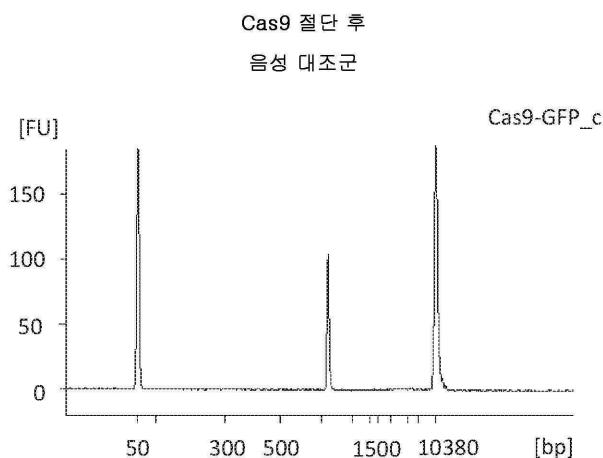
도면 14b



도면14c

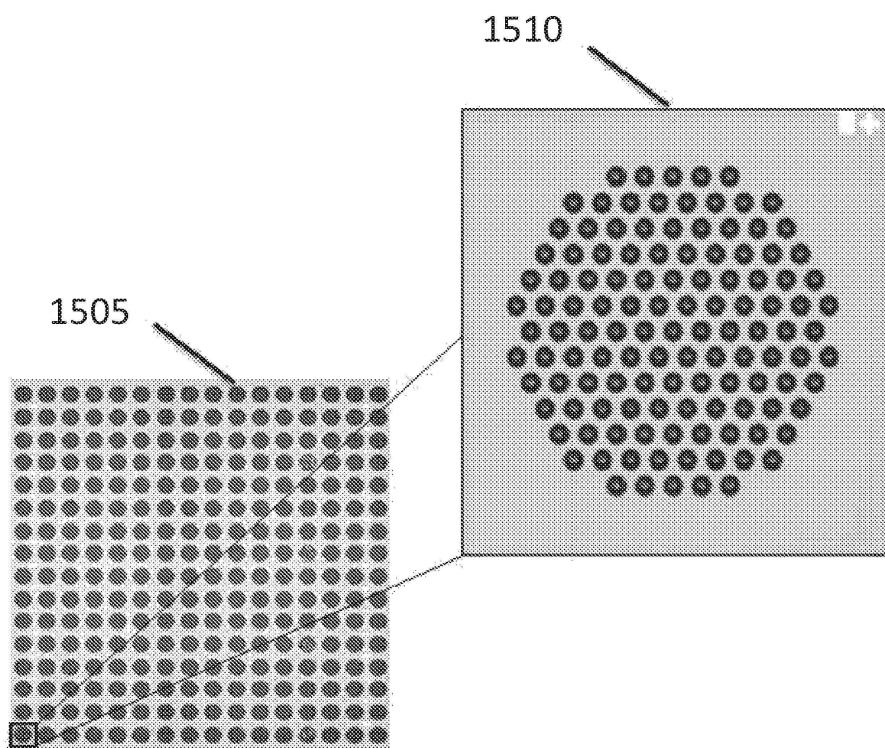


도면14d

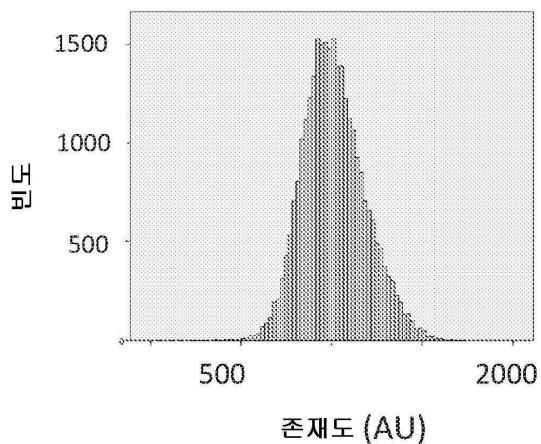


도 14J

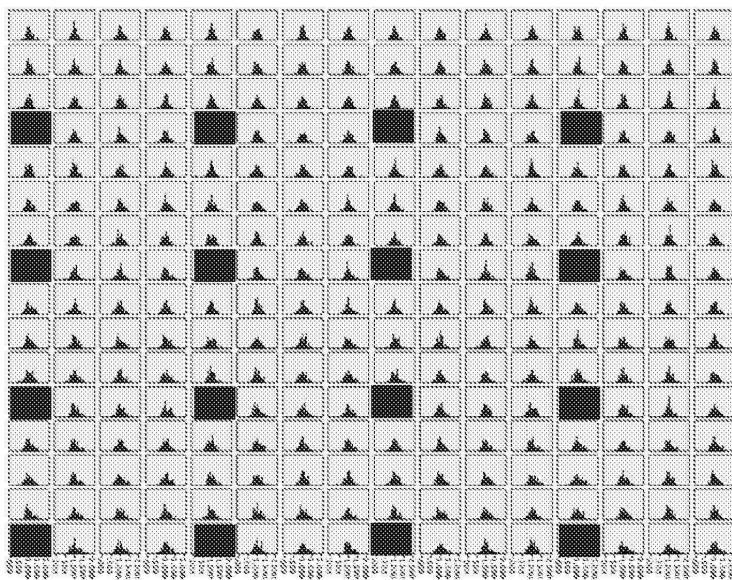
도면15



도면16a

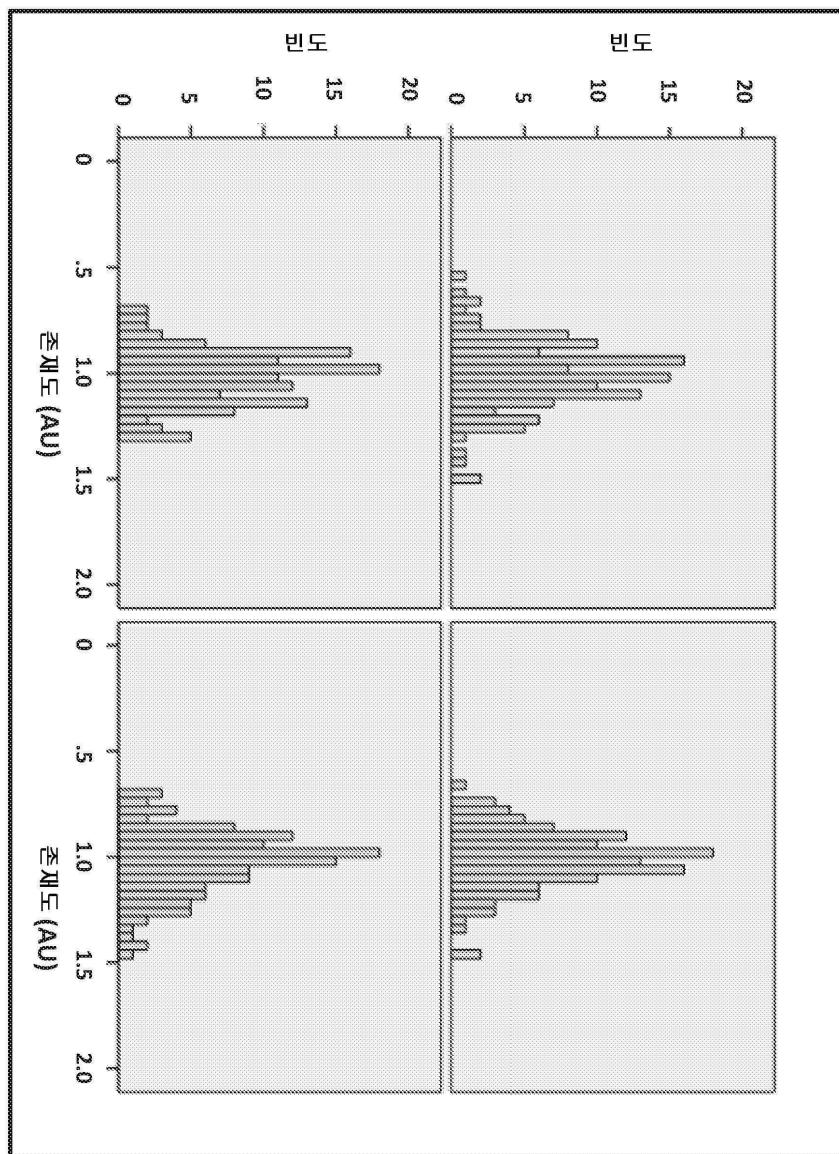


도면 16b

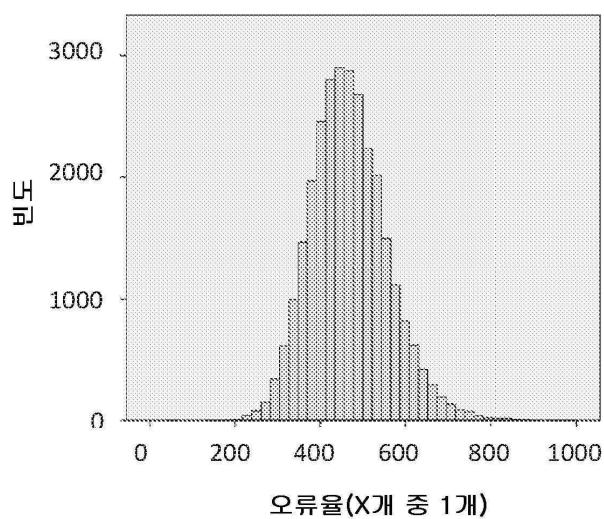


■ 대조군 클러스터

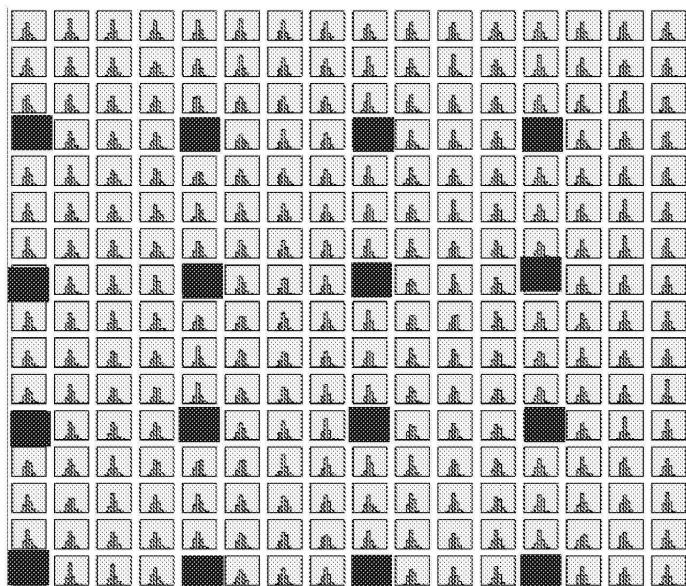
도면17



도면18a

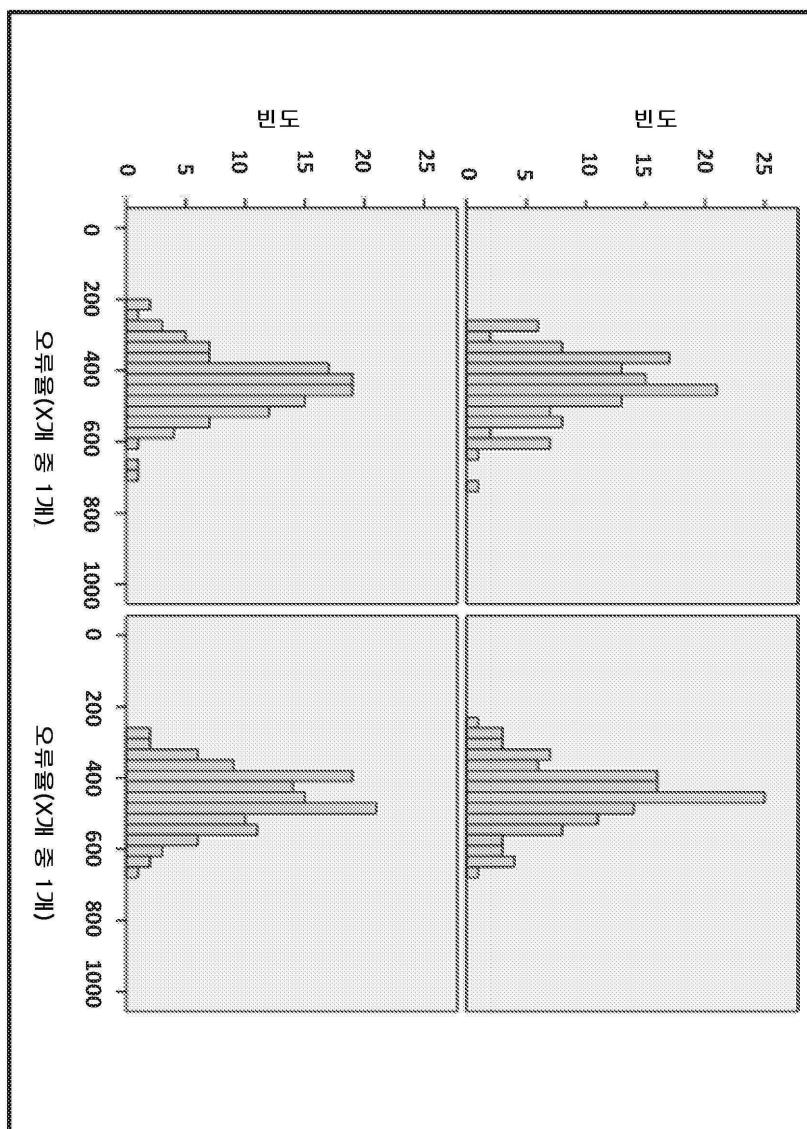


도면 18b

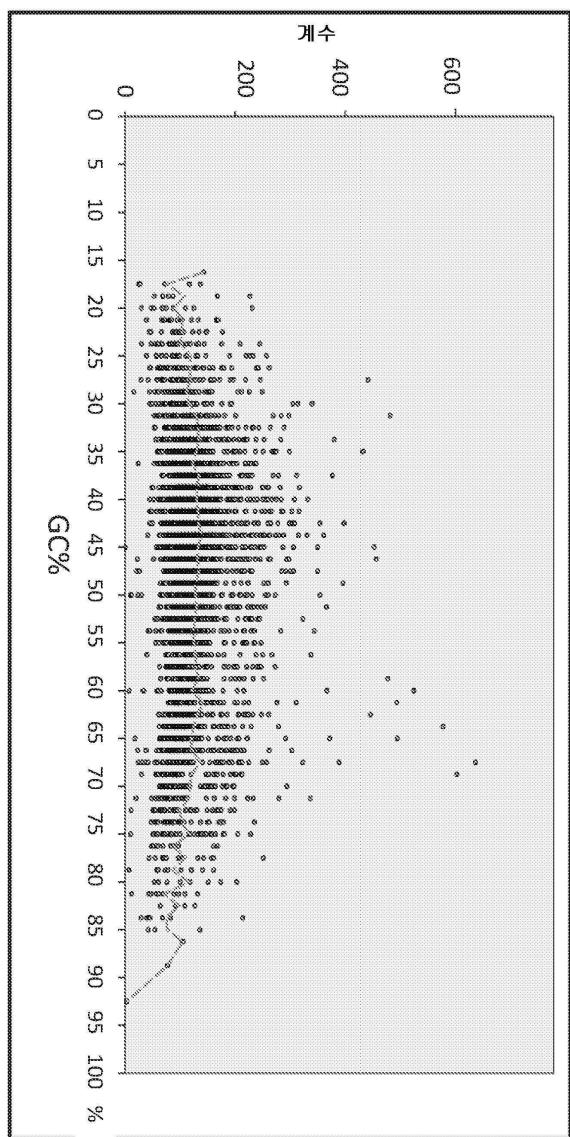


■ 대조군 클러스터

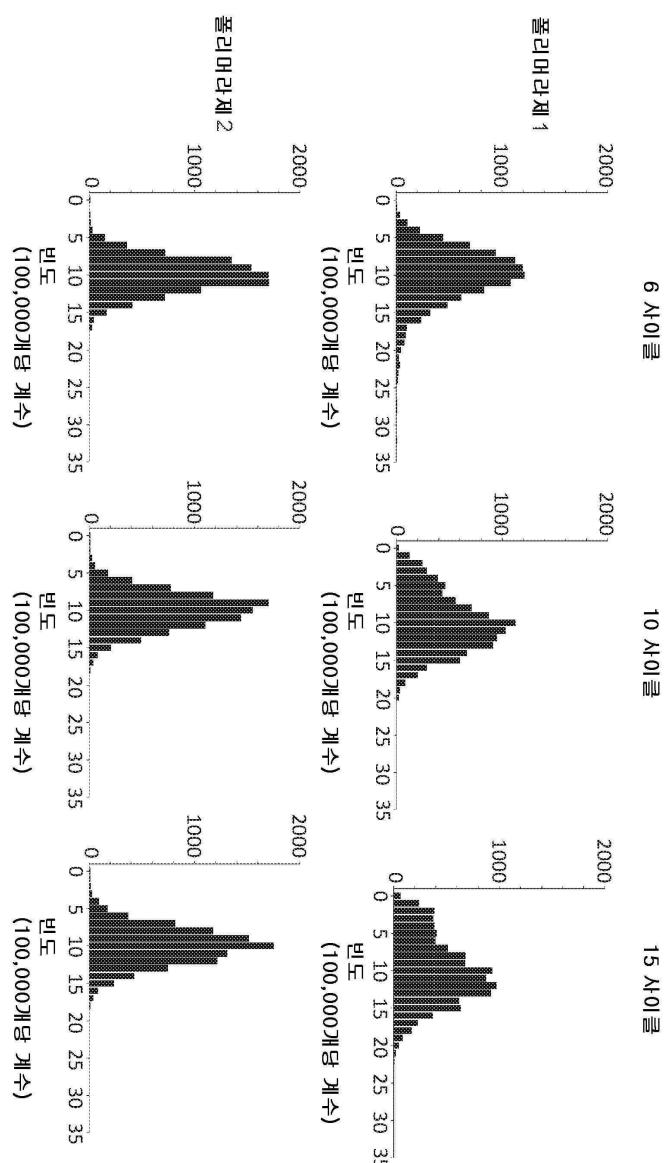
도면19



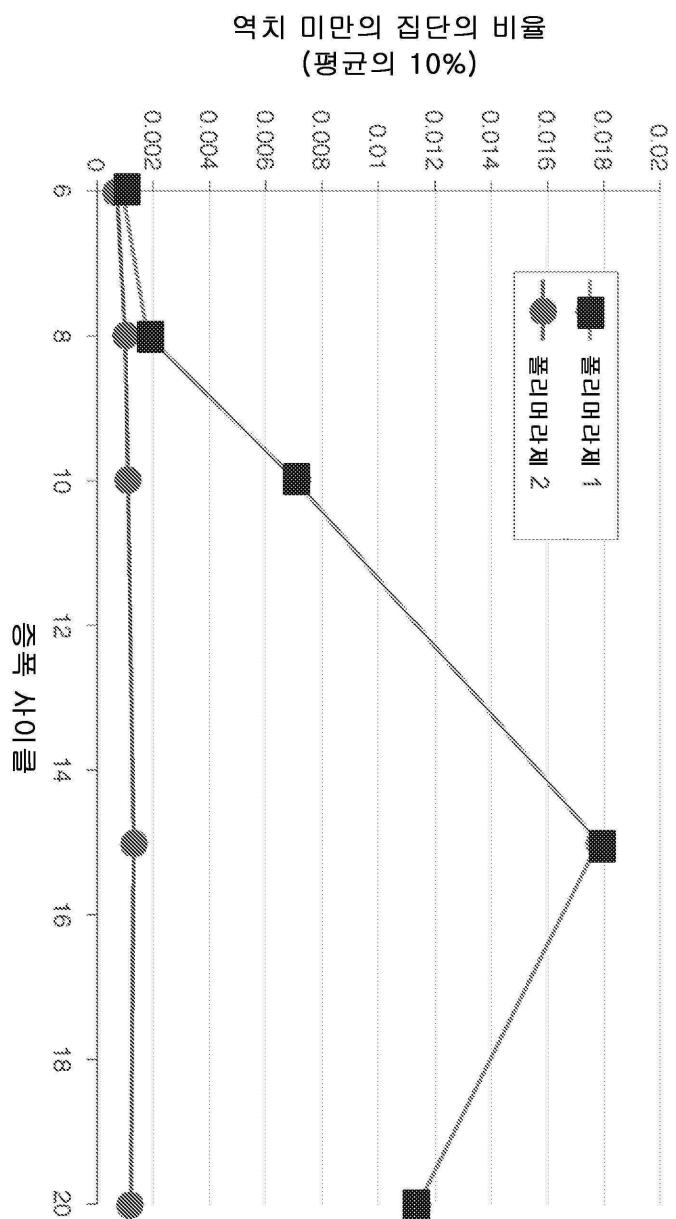
도면20



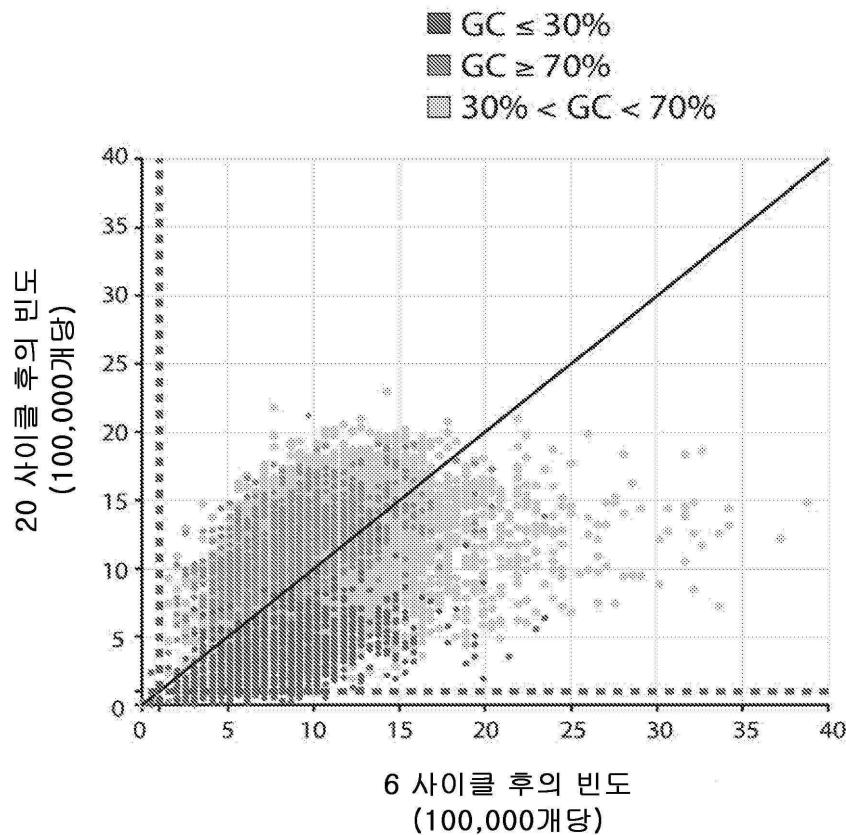
도면21



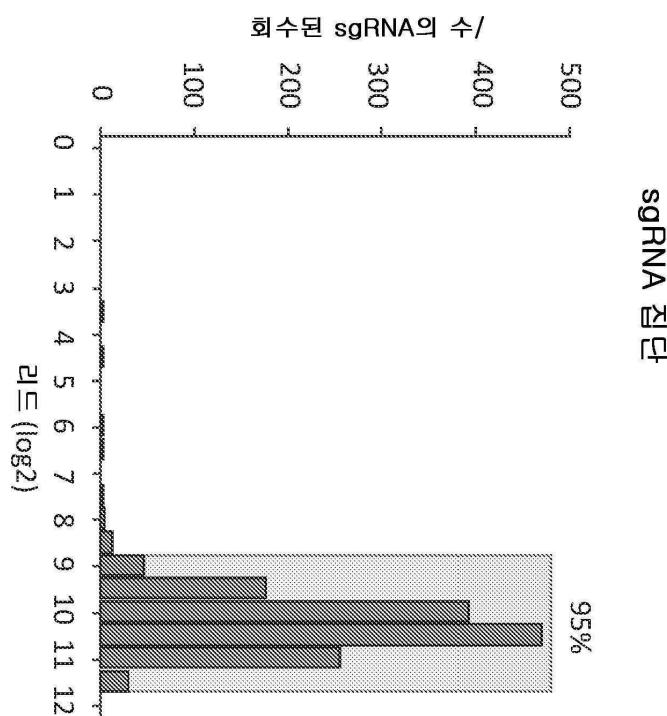
도면22



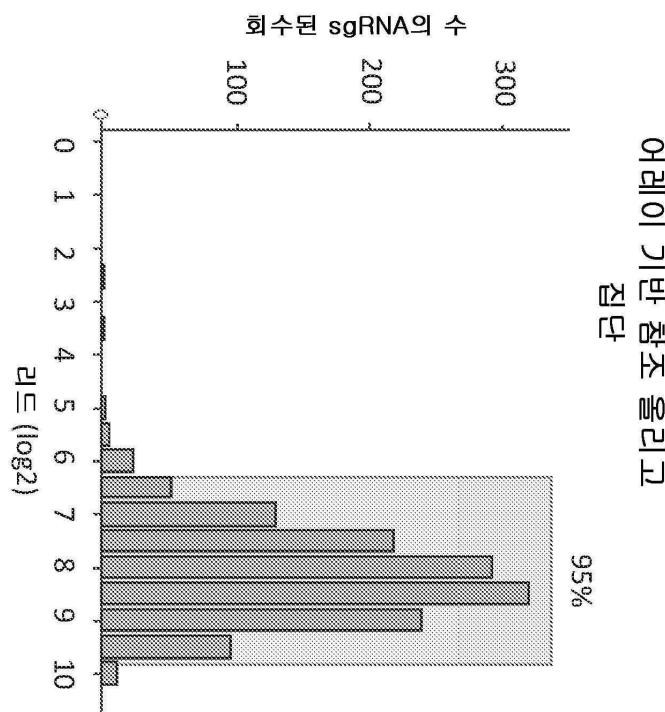
도면23



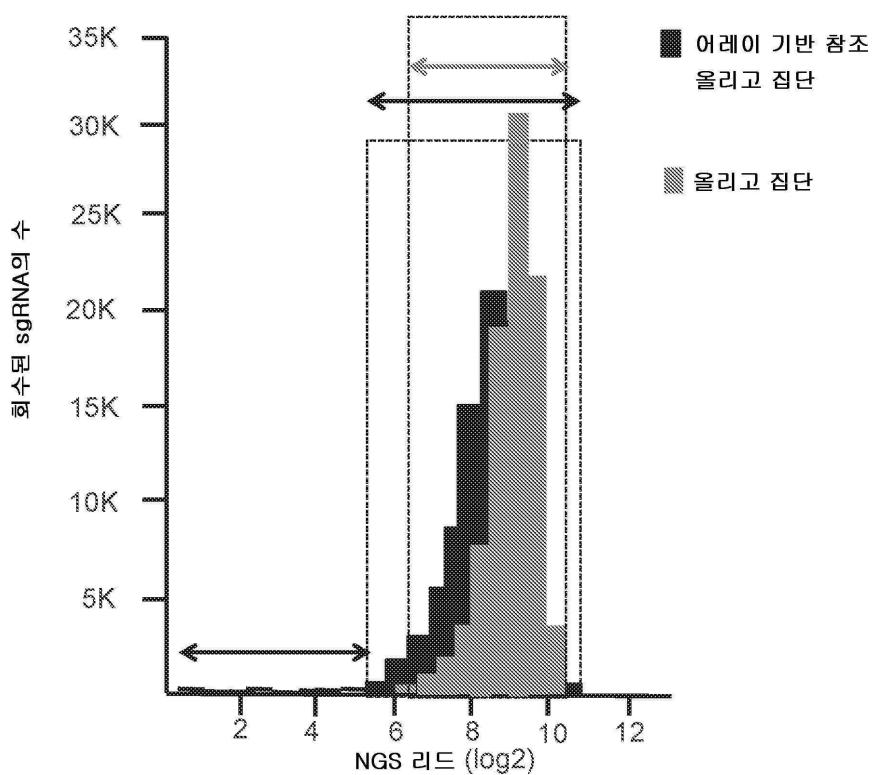
도면24a



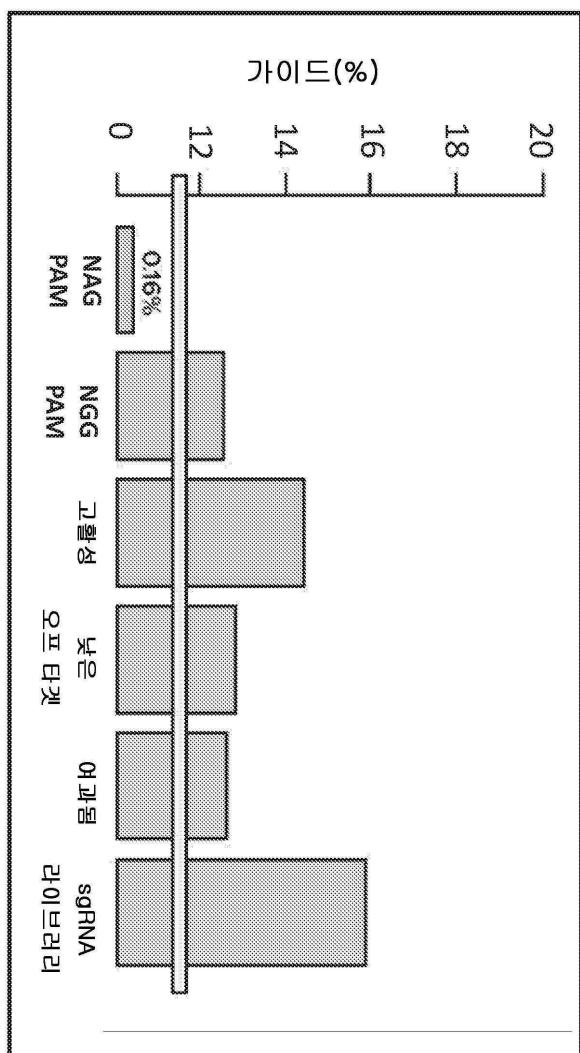
도면24b



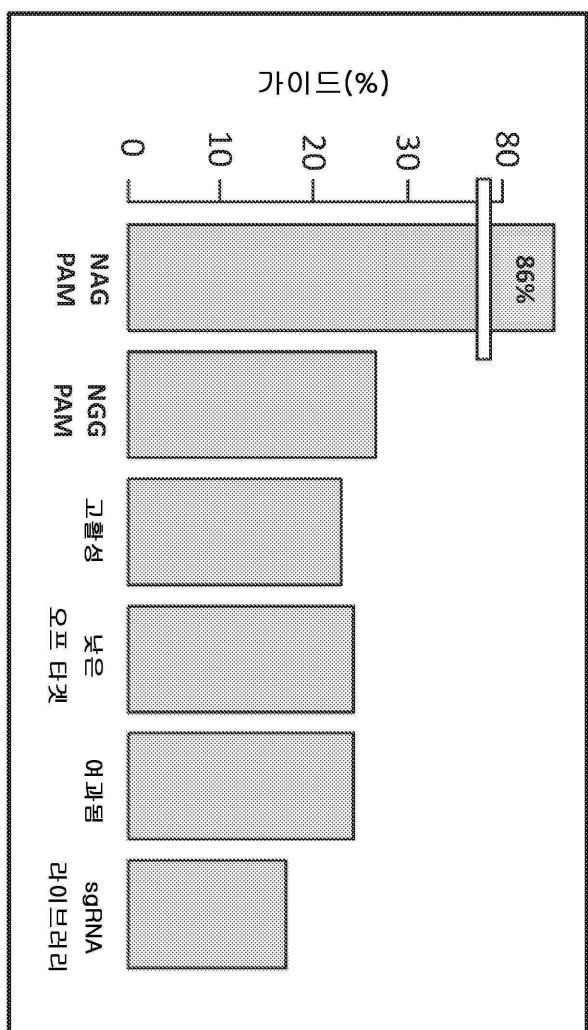
도면25



도면26a



도면26b



서 열 목 록

SEQUENCE LISTING

<110> TWIST BIOSCIENCE CORPORATION

<120> DE NOVO SYNTHESIZED NUCLEIC ACID LIBRARIES

<130> 44854-727.601

<140><141><150> 62/419,881

<151> 2016-11-09

<150> 62/411,338

<151> 2016-10-21

<150> 62/393,948

<151> 2016-09-13

<150> 62/382,191

<151> 2016-08-31

<150> 62/378,134

<151> 2016-08-22

<160> 44

<170> PatentIn version 3.5

<210> 1

<211> 62

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<220><221> modified_base

<222> (51)..(52)

<223> Thymidine-succinyl hexamide CED phosphoramidite

<400> 1

agacaatcaa ccatttgggg tggacagcct tgacctctag acttcggcat	tttttttt	60
tt		62

<210> 2

<211> 112

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<220><221> modified_base

<222> (101)..(102)

<223> Thymidine-succinyl hexamide CED phosphoramidite

<400> 2

cgggatcctt atcgtcatcg tcgtacagat cccgaccat ttgctgtcca ccagtcatgc	60
tagccatacc atgatgatga tcatgatgag aaccccgcat tttttttttt tt	112

<210> 3

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

primer
<400> 3

atgcgggtt ctcatcatc 19

<210> 4

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic primer

<400> 4

cgggatcctt atcgcatcg 20

<210> 5

<211> 42

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 5

ataactcaat ttgtaaaaaa gtttagagc tatgctgtt tg 42

<210> 6

<211> 86

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 6

ggaaccattc aaaacagcat agcaagttaa aataaggcta gtccgttac aacttgaaaa 60

agtggcaccg agtcggtgct tttttt 86

<210> 7

<211> 38

<212> DNA

<213> Streptococcus thermophilus

<400> 7

tactcaactt	gaaaagggtgg	caccgattcg	gtgttttt	38		
<210>	8					
<211>	39					
<212>	DNA					
<213>	Streptococcus mutans					
<400>	8					
tacacaactt	gaaaaagtgc	gcaccgattc	ggtgcttt	39		
<210>	9					
<211>	40					
<212>	DNA					
<213>	Listeria innocua					
<400>	9					
ttatcaactt	ttaattaagt	agcgctgttt	cggcgcttt	40		
<210>	10					
<211>	40					
<212>	DNA					
<213>	Mycoplasma mobile					
<400>	10					
tatgccgtaa	ctactactta	tttcaaaat	aagtagttt	40		
<210>	11					
<211>	36					
<212>	DNA					
<213>	Campylobacter jejuni					
<400>	11					
gactctgcgg	ggttacaatc	ccctaaaacc	gctttt	36		
<210>	12					
<211>	117					
<212>	DNA					
<213>	Artificial Sequence					
<220><223>	Description of Artificial Sequence: Synthetic					
	polynucleotide					
<400>	12					
taatacact	cactataggg	gatgcgcgca	gttgtccgtt	tttagagctag	aatagcaag	60

ttaaaataag gctagtcgt tatcaacttg aaaaagtggc accgagtcgg tgcttt	117
<210> 13	
<211> 17	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic	
oligonucleotide	
<400> 13	
taatacgact cactata	17
<210> 14	
<211> 18	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic	
oligonucleotide	
<400> 14	
ggatgcgcgc agttgtcc	18
<210> 15	
<211> 80	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic	
oligonucleotide	
<400> 15	
gttttagagc tagaaatagc aagttaaaat aaggctagtc cgtttatcaac ttgaaaaagt	60
ggcacccgagt cggtgcttt	80
<210> 16	
<211> 121	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic	
polynucleotide	
<400> 16	

gagctaatac gactcaactat aggggatgcg cgcaaggatgc cggttttagag ctagaaatag 60

caaggtaaaa taaggctagt ccgttatcaa cttgaaaaag tggcaccgag tcggtgctt 120

t 121

<210> 17

<211> 122

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 17

cggatcaata cgactcaacta taggggatgc ggcgcgttgtt ccgttttaga gctagaaata 60

gcaaggtaaaa ataaggcttag tccgttatca acttggaaaaa gtggcaccgaa gtcggtgctt 120

tt 122

<210> 18

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 18

taatacgact cactatagg 19

<210> 19

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 19

aaaaggcaccg actcggtgcc acttttcaa gttgataacg gactagcctt attttaactt 60

gctatttcta gctctaaaac 80

<210> 20

<211> 58

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 20

taatacact cactataggg gatgcgcga gttgtccgtt ttagagctag aaatagca 58

<210> 21

<211> 65

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 21

gaaattaata cgactcacta tagggatgc gcgcagttgt ccgttttaga gctagaaata 60

gcaag 65

<210> 22

<211> 66

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 22

gagctaatac gactcactat agggatgca cgcaatgtc cgtttagag ctagaaatag 60

caagtt 66

<210> 23

<211> 75

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 23

gagctaatac gactcactat agggatgca cgcaatgtc cgtttagag ctagaaatag 60

caagttaaa taagg 75

<210> 24	
<211> 14	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic	
oligonucleotide	
<400> 24	
aaaagcaccg actc	14
<210> 25	
<211> 15	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic	
oligonucleotide	
<400> 25	
aaaagcaccg actcg	15
<210> 26	
<211> 16	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic	
oligonucleotide	
<400> 26	
aaaagcaccg actcgg	16
<210> 27	
<211> 17	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic	
oligonucleotide	
<400> 27	
aaaagcaccg actcggt	17
<210> 28	
<211> 24	

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 28

gaaattaata cgactcacta tagg 24

<210> 29

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 29

gagctaatac gactcactat agg 23

<210> 30

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 30

gcgagctaat acgactcact atagg 25

<210> 31

<211> 120

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 31

cgagctataa cgactcacta tagggcaca acgtggagga tggcagcgtg cagctggctg 60

atcaactacca gcaaaacact ccaatcggtg atggcctgt tgcaccgagt cggtgcttt 120

<210> 32

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 32

cgagctaata	cgactcacta	tagggcaca	acgtggagga	tggcagcgtg	cagctggctg	60
atcactacca	g					71

<210> 33

<211> 69

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 33

aaagcaccga	ctcggtgcaa	caggaccatc	accgattgga	gtgtttgct	ggtagtgatc	60
agccagctg						69

<210> 34

<211

> 130

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<220><221> modified_base

<222> (25)..(42)

<223> a, c, t, g, unknown or other

<400> 34

cgagctaata	cgactcacta	taggnnnnn	nnnnnnnnnn	nngttttaga	gctatgtga	60
aaagcatagc	aagttaaat	aaggctagtc	cgttatcaac	ttgaaaaagt	ggcaccgagt	120
cggtgcttt						130

<210> 35

<211> 720

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 35

atgcgtaaag gcgaaagact gttcaactggt gtcgtcccta ttctggtgga actggatggt	60
gatgtcaacg gtcataagtt tccgtgcgt ggcgagggtg aaggtgacgc aactaatggt	120
aaactgacgc tgaagttcat ctgtactact ggttaaactgc cggtaccttg gccgactctg	180
gtaacgacgc tgacttatgg tggcgtgc tttgctcgat atccggacca tatgaaggcag	240
catgacttct tcaagtccgc catgccgaa ggctatgtgc aggaacgcac gatttcctt	300
aaggatgacg gcacgtacaa aacgcgtgacg gaagtgaaat ttgaaggcga taccctggta	360

aaccgcattt agctgaaagg cattgacttt aaagaagacg gcaatatcct gggccataag	420
ctggaaataca atttaacag ccacaatgtt tacatcaccg ccgataaaca aaaaaatggc	480
attaaaggcga attttaaat tcgcccacaac gtggaggatg gcagcgtgca gctggctgat	540
cactaccaggc aaaaacactcc aatcggtat ggtcctgttc tgcgtccaga caatcactat	600
ctgagcacgc aaagcggtct gtctaaagat ccgaacgaga aacgcgtatca tatggttctg	660
ctggagttcg taaccgcaggc gggcatcaggc catggatgg atgaactgtt caaatgataa	720

<210> 36

<211> 70

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 36

cggatataa cggactacta taggaacgca cgatttcctt tagttttaga gctatgtga	60
aaagcatagc	70

<210> 37

<211> 70

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 37

cgagctaata cgactcacta taggcattga cttaaagaa gagttttaga gctatgctga	60	
aaagcatagc	70	
<210> 38		
<211> 70		
<212> DNA		
<213> Artificial Sequence		
<220><223> Description of Artificial Sequence: Synthetic		
oligonucleotide		
<400> 38		
cgagctaata cgactcacta taggaggatg gcagcgtgca gcgttttaga gctatgctga	60	
aaagcatagc	70	
<210> 39		
<211> 88		
<212> DNA		
<213> Artificial Sequence		
<220><223> Description of Artificial Sequence: Synthetic		
oligonucleotide		
<400> 39		
aaaagcaccg actcggtgcc acttttcaa gttgataacg gactaggcctt atttaactt	60	
gctatgctt tcagcatagc tctaaaac		88
<210> 40		
<211> 103		
<212> RNA		
<213> Artificial Sequence		
<220><223> Description of Artificial Sequence: Synthetic		
polynucleotide		
<220><221> modified_base		
<222> (1)..(20)		
<223> a, c, u, g, unknown or other		
<400> 40		
nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc	60	
cguuaauaac uugaaaaagu ggcaccgagu cggugcuuuu uuu	103	
<210> 41		

<211> 76

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 41

cgagctaata cgactcacta tagggatgc gcgcagttgt ccgttttaga gctagaata 60

gcaagttaaa ataagg 76

<210> 42

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 42

taaggctagt ccgttatcaa cttgaaaaag 30

<210> 43

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 43

ctagtcgtt atcaacttga aaaagtg 27

<210> 44

<211> 122

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 44

gaaattaata cgactcacta tagggatgc gcgcagttgt ccgttttaga gctagaata 60

gcaagttaaa ataaggctag tcgttatca acttggaaaa gtggcaccga gtcggcgttt 120

