



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2021년01월29일

(11) 등록번호 10-2209636

(24) 등록일자 2021년01월25일

(51) 국제특허분류(Int. Cl.)

C12N 15/63 (2006.01) C12N 15/10 (2017.01)

C12N 15/11 (2006.01) C12N 15/86 (2006.01)

C12N 9/22 (2006.01) C40B 40/08 (2006.01)

(52) CPC특허분류

C12N 15/63 (2013.01)

C12N 15/102 (2013.01)

(21) 출원번호 10-2016-7024869

(22) 출원일자(국제) 2015년02월11일

심사청구일자 2020년01월08일

(85) 번역문제출일자 2016년09월08일

(65) 공개번호 10-2016-0122197

(43) 공개일자 2016년10월21일

(86) 국제출원번호 PCT/US2015/015476

(87) 국제공개번호 WO 2015/123339

국제공개일자 2015년08월20일

(30) 우선권주장

61/938,608 2014년02월11일 미국(US)

(56) 선행기술조사문헌

Science. 2014 January 3, 343(6166):84-87.*

(뒷면에 계속)

전체 청구항 수 : 총 20 항

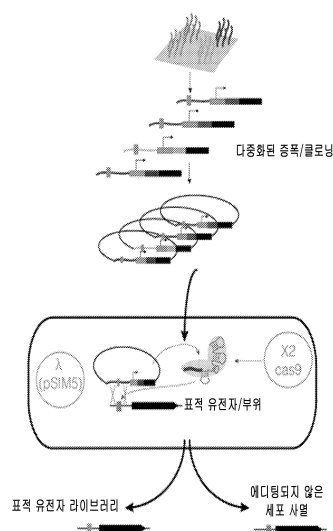
심사관 : 이현지

(54) 발명의 명칭 CRISPR 이용의 다중화된 게놈 조작

(57) 요약

다양한 CRISPR 시스템이 이용되는, 세포 또는 세포 집단 내의 염색체의 임의의 분절 또는 오픈 리딩 프레임(예를 들면, 단백질 라이브러리에서) 내에서 염색체의 합리적 다중화된 조작을 위한 방법 및 벡터가 개시되어 있다.

대표도 - 도11b



(52) CPC특허분류

C12N 15/111 (2013.01)

C12N 15/86 (2013.01)

C12N 9/22 (2013.01)

C40B 40/08 (2013.01)

C12N 2310/10 (2013.01)

C12N 2320/11 (2013.01)

C12N 2320/12 (2013.01)

C12N 2330/31 (2013.01)

(56) 선행기술조사문헌

Nature Biotechnology. 2013, 31(3):233-239.*

Nat Biotechnol. 2013 Sep;31(9):827-32.

W02014093622 A1

W02013176772 A1

*는 심사관에 의하여 인용된 문헌

명세서

청구범위

청구항 1

2 이상의 합성 올리고뉴클레오타드의 라이브러리로서, 각각의 합성 올리고뉴클레오타드가 하기의 공유 연결된 성분들:

(i) 세포 내의 표적 영역을 표적화하는 가이드(guide) RNA(gRNA) 서열을 코딩하는 핵산; (ii) 표적 영역에 대한 서열 변화를 포함하는 표적 영역에 상동성인 영역; 및 (iii) 뉴클레아제 매개 에디팅(editing)에 대한 면역성을 부여하는 부위를 포함하고,

뉴클레아제 매개 에디팅에 대한 면역성을 부여하는 부위가 프로토스페이스어(protospacer) 인접 모티프(PAM) 돌연변이인 2 이상의 합성 올리고뉴클레오타드의 라이브러리.

청구항 2

제1항에 있어서, (iv) 바코드를 더 포함하는 2 이상의 합성 올리고뉴클레오타드의 라이브러리.

청구항 3

제1항에 있어서, 표적 영역 중 적어도 하나가 비코딩 영역 내에 있는 것인 2 이상의 합성 올리고뉴클레오타드의 라이브러리.

청구항 4

제1항에 있어서, 표적 영역 중 적어도 하나가 관심 유전자 내에 있는 것인 2 이상의 합성 올리고뉴클레오타드의 라이브러리.

청구항 5

제4항에 있어서, 표적 영역에 대한 서열 변화가 관심 유전자 내에 있는 것인 2 이상의 합성 올리고뉴클레오타드의 라이브러리.

청구항 6

삭제

청구항 7

제1항에 있어서, 표적 영역에 대한 서열 변화가 PAM 돌연변이의 100개 뉴클레오타드 내에 있는 것인 2 이상의 합성 올리고뉴클레오타드의 라이브러리.

청구항 8

제1항에 있어서, gRNA가 단일 키메라 gRNA인 2 이상의 합성 올리고뉴클레오타드의 라이브러리.

청구항 9

제1항에 있어서, 적어도 10^4 - 10^6 개의 상이한 합성 올리고뉴클레오타드를 포함하는 2 이상의 합성 올리고뉴클레오타드의 라이브러리.

청구항 10

2 이상의 합성 올리고뉴클레오타드의 라이브러리로서, 각각의 합성 올리고뉴클레오타드가 하기의 공유 연결된 성분들:

(i) 세포 내의 표적 영역을 표적화하는 가이드 RNA(gRNA) 서열을 코딩하는 핵산; (ii) 표적 영역에 대한 서열

변화를 포함하는 표적 영역에 상동성인 영역; 및 (iii) 바코드를 포함하는 2 이상의 합성 올리고뉴클레오타드의 라이브러리.

청구항 11

제10항에 있어서, 적어도 하나의 합성 올리고뉴클레오타드가, 프로토스페이스 인접 모티프(PAM)가 뉴클레아제 인식에 대해 면역이 되게 하는 부위를 더 포함하고, PAM이 뉴클레아제 인식에 대해 면역이 되게 하는 부위가 PAM 돌연변이인 2 이상의 합성 올리고뉴클레오타드의 라이브러리.

청구항 12

제10항에 있어서, 표적 영역 중 적어도 하나가 비코딩 영역 내에 있는 것인 2 이상의 합성 올리고뉴클레오타드의 라이브러리.

청구항 13

제11항에 있어서, 표적 영역 중 적어도 하나가 관심 유전자 내에 있는 것인 2 이상의 합성 올리고뉴클레오타드의 라이브러리.

청구항 14

제13항에 있어서, 표적 영역에 대한 서열 변화가 관심 유전자 내에 있는 것인 2 이상의 합성 올리고뉴클레오타드의 라이브러리.

청구항 15

제1항에 있어서, 적어도 27,000개의 상이한 합성 올리고뉴클레오타드를 포함하는 2 이상의 합성 올리고뉴클레오타드의 라이브러리.

청구항 16

제1항에 있어서, 적어도 10^4 개의 상이한 합성 올리고뉴클레오타드를 포함하는 2 이상의 합성 올리고뉴클레오타드의 라이브러리.

청구항 17

제1항에 있어서, 적어도 10^5 개의 상이한 합성 올리고뉴클레오타드를 포함하는 2 이상의 합성 올리고뉴클레오타드의 라이브러리.

청구항 18

제1항에 있어서, 적어도 10^6 개의 상이한 합성 올리고뉴클레오타드를 포함하는 2 이상의 합성 올리고뉴클레오타드의 라이브러리.

청구항 19

제1항에 있어서, 뉴클레아제 매개 에디팅에 대한 면역성을 부여하는 부위가 표적 영역에 대한 하나 이상의 뉴클레오타드의 삽입을 포함하는 2 이상의 합성 올리고뉴클레오타드의 라이브러리.

청구항 20

제1항에 있어서, 뉴클레아제 매개 에디팅에 대한 면역성을 부여하는 부위가 표적 영역에 대한 하나 이상의 뉴클레오타드의 결실을 포함하는 2 이상의 합성 올리고뉴클레오타드의 라이브러리.

청구항 21

제1항에 있어서, 뉴클레아제 매개 에디팅에 대한 면역성을 부여하는 부위가 표적 영역에 대한 하나 이상의 뉴클레오타드의 치환을 포함하는 2 이상의 합성 올리고뉴클레오타드의 라이브러리.

청구항 22

삭제

청구항 23

삭제

청구항 24

삭제

청구항 25

삭제

청구항 26

삭제

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

발명의 설명

기술 분야

[0001] 관련 출원

[0002] 본원은 2014년 2월 11일자로 출원된 미국 가출원 제61/938,608호(이의 전체 교시는 본원에 참고로 도입됨)의 이익을 35 U.S.C. § 119 하에 주장한다.

배경 기술

[0003] 큰 DNA 구축물의 합리적(rational) 조작은 현재의 합성 생물학 및 게놈 조작 노력에 대한 주된 도전과제이다. 최근에, 이 도전과제를 다루고 돌연변이가 생성될 수 있는 특이성 및 속도를 증가시키기 위해 다양한 기술들이 개발되었다. 추가로, 유도적(adaptive) 돌연변이는 진화의 중추적인 유발자이지만, 그들의 존재도 및 세포 표현형에의 상대적 기여는 가장 잘 연구된 유기체에서조차도 잘 이해되어 있지 않다. 이것은 주로 이 유전형들의 관찰 및 재구축, 및 그들의 존재와 관심 있는 표현형의 상관성과 관련된 기술적 도전과제에 기인할 수 있다. 예를 들면, 무작위 돌연변이유발에 의존하는 게놈 에디팅(editing) 방법은 많은 돌연변이들로 구성된 복잡한 유전형을 유발하고, 이 돌연변이들 각각의 상대적 기여는 파악되기 어렵다. 더욱이, 대립형질들 사이의 상위(epistatic) 상호작용은 개별 돌연변이에 대한 정보의 결여로 인해 할당하기 어렵다.

발명의 내용

[0004] 주기적으로 간격을 두고 분포하는 짧은 회문구조 반복서열(Clustered regularly interspersed short palindromic repeats)(CRISPR)은 많은 세균 게놈들에 존재하고 적응 세균 면역에서 중요한 역할을 수행하는 것으로 발견되었다. 이 어레이의 전사는 유전자 억제 또는 DNA 절단을 위해 세포 내의 DNA 표적에 대한 CRISPR/cas 복합체의 서열 특이적 결합을 유도하는 CRISPR RNA들을 생성한다. 이 복합체의 특이성은 균주 조작을 위한 신규 생체내 적용을 가능하게 한다.

[0005] 염색체의 임의의 분절 내의 오픈 리딩 프레임(open reading frames)(예를 들면, 단백질 라이브러리를 생성하기 위함) 또는 다수의 유전자들 내에서 염색체의 합리적 다중화된(multiplexed) 조작을 수행하는 방법으로서, 다양한 CRISPR 시스템들을 사용하는 방법이 본원에 기재되어 있다. 이 방법은 종래 이용가능한 방법보다 더 효율적인 조합적 게놈 조작을 제공한다.

[0006] CRISPR의 다중화 성능의 확장은 현재의 기술적 도전과제를 제공하고 고처리율 포맷(high-throughput format)으로 합리적 라이브러리를 생성하기 위한 이 시스템의 사용을 가능하게 할 것이다. 이러한 진보는 최적 생산을 위해 복잡한 유전적 네트워크를 리팩토링(refactoring)하고자 하는 대사적 및 단백질 조작의 분야에 대한 넓은 파급 효과를 가진다.

[0007] 본 방법은 CRISPR 관련 뉴클레아제(nuclease) Cas9 및 서열 특이적 가이드(guide) RNA(gRNA)를 비롯한 CRISPR 시스템의 성분을 세포 내로 도입함으로써, 서열-유도된 이중 가닥 절단을 유도하는 CRISPR 시스템의 능력을 이용하여 상기 절단을 유도하는 단계를 포함한다. CRISPR 관련 뉴클레아제 Cas9 및 서열 특이적 가이드 RNA(gRNA)를 비롯한 CRISPR 시스템의 성분은 하나 이상의 벡터, 예컨대, 플라스미드 상에 코딩된 상태로 세포 내로 도입될 수 있다. DNA 리컴바이너링(recombineering) 카세트 또는 에디팅 올리고뉴클레오티드는 표적 유전자와 내부의 원하는 돌연변이, 및 CRISPR 시스템에 의해 인식될 수 있는 표적 유전자와 외부의 공통된 위치의 돌연변이를 포함하도록 합리적으로 디자인될 수 있다. 기재된 방법은 관심 있는 경로의 변경을 비롯한 많은 적용들을 위해 이용될 수 있다.

[0008] 한 실시양태에서, 본 방법은 (a) (i) 세포 내의 핵산의 표적 영역에 대해 상동성이고 표적 영역에 비해 적어도 하나의 뉴클레오티드의 돌연변이(원하는 돌연변이로서 지칭됨), 예컨대, 표적 영역에 비해 적어도 하나의 코돈에서 적어도 하나의 뉴클레오티드의 돌연변이를 포함하는 영역, 및 프로토스페이스(protospacer) 인접 모티프(PAM) 돌연변이를 포함하는 에디팅 카세트; (ii) 프로모터; 및 (iii) a) 표적 영역의 일부에 상보적인 영역(RNA) 및 b) Cas9 뉴클레아제를 동원하는 영역(RNA)을 포함하는 적어도 하나의 가이드 RNA(gRNA)를 코딩하는 벡터를 세포 내로 도입함으로써, 상기 벡터를 포함하는 세포를 생성하는 단계; (b) 상기 벡터 상에 코딩되어 있거나, 제2 벡터 상에 코딩되어 있거나 세포의 게놈 상에 코딩되어 있는 Cas9 뉴클레아제가 발현되는 조건 하에서 상기 벡터를 포함하는 세포를 유지함으로써, 상기 벡터를 포함하고 PAM 돌연변이를 포함하지 않는 세포, 및 상기 벡터 및 PAM 돌연변이를 포함하는 세포의 생성을 야기하는 단계; (c) 세포 생존능에 적절한 조건 하에서

(b)의 생성물을 배양함으로써 생존 세포를 생성하는 단계; (d) (c)에서 생성된 생존 세포를 얻는 단계; 및 (e) (d)에서 얻은 적어도 하나의 생존 세포의 벡터의 에디팅 올리고뉴클레오타이드를 서열결정하고 적어도 하나의 코돈의 돌연변이를 확인하는 단계를 포함하는 게놈 조작 방법이다.

[0009] 또 다른 실시양태에서, 본 방법은 (a) (i) a) 핵산의 표적 영역에 대해 상동성이고 표적 영역에 비해 적어도 하나의 뉴클레오타이드의 돌연변이, 예컨대, 표적 영역에 비해 적어도 하나의 코돈에서 적어도 하나의 뉴클레오타이드의 돌연변이를 포함하는 영역, 및 b) 프로토스페이서 인접 모티프(PAM) 돌연변이를 포함하는 적어도 하나의 에디팅 카세트; (ii) 적어도 하나의 프로모터; 및 (iii) a) 표적 영역의 일부에 상보적인 영역(RNA) 및 b) Cas9 뉴클레아제를 동원하는 영역(RNA)을 포함하는 적어도 하나의 가이드 RNA(gRNA)를 코딩하는 벡터를 제1 세포 집단 내로 도입함으로써, 상기 벡터를 포함하는 제2 세포 집단을 생성하는 단계; (b) 상기 벡터, 제2 벡터 또는 제2 세포 집단의 세포의 게놈 상에 코딩되어 있는 Cas9 뉴클레아제가 발현되는 조건 하에서 제2 세포 집단을 유지함으로써, PAM 돌연변이를 포함하지 않는 세포에서의 DNA 절단 및 이러한 세포의 사멸을 야기하는 단계; (c) (b)에서 생성된 생존 세포를 얻는 단계; 및 (d) 제2 세포 집단의 적어도 하나의 세포의 벡터의 에디팅 올리고뉴클레오타이드를 서열결정하여 적어도 하나의 코돈의 돌연변이를 확인하는 단계를 포함하는, 추적가능한 CRISPR 농축된 리컴바이니어링(trackable CRISPR enriched recombineering)에 의한 게놈 조작 방법이다.

[0010] 상기 실시양태들 중 어느 한 실시양태는 에디팅 올리고뉴클레오타이드 집단을 합성하고/하거나 얻는 단계도 포함할 수 있다. 어느 한 실시양태는 에디팅 올리고뉴클레오타이드 집단을 증폭하는 단계도 포함할 수 있다. 상기 실시양태들 중 임의의 실시양태에서, 벡터는 스페이서, 적어도 2개의 프라이밍(priming) 부위들, 또는 스페이서 및 적어도 2개의 프라이밍 부위들 둘 다를 추가로 포함할 수 있다. 일부 실시양태들에서, 에디팅 카세트는 PAM 돌연변이의 100개 뉴클레오타이드들 내에서 적어도 하나의 코돈의 돌연변이를 포함하는 표적 영역을 포함한다.

[0011] (i) 세포 내의 핵산의 표적 영역에 대해 상동성이고 표적 영역에 비해 적어도 하나의 뉴클레오타이드의 돌연변이(원하는 돌연변이로서 지칭됨)를 포함하는 영역, 및 프로토스페이서 인접 모티프(PAM) 돌연변이를 포함하는 에디팅 카세트;

[0012] (ii) 프로모터; 및

[0013] (iii) (a) 표적 영역의 일부에 상보적인 영역(RNA) 및 (b) Cas9 뉴클레아제를 동원하는 영역(RNA)을 포함하는 적어도 하나의 가이드 RNA(gRNA)

[0014] 를 포함하는 벡터도 기재되어 있다.

[0015] 추가 실시양태는

[0016] (i) 세포 내의 핵산의 표적 영역에 대해 상동성이고 표적 영역에 비해 적어도 하나의 코돈에서 적어도 하나의 뉴클레오타이드의 돌연변이(원하는 돌연변이로서 지칭됨)를 포함하는 영역, 및 프로토스페이서 인접 모티프(PAM) 돌연변이를 포함하는 에디팅 카세트;

[0017] (ii) 프로모터; 및

[0018] (iii) (a) 표적 영역의 일부에 상보적인 영역(RNA) 및 (b) Cas9 뉴클레아제를 동원하는 영역(RNA)을 포함하는 적어도 하나의 가이드 RNA(gRNA)

[0019] 를 포함하는 벡터이다.

[0020] 추가 실시양태는

[0021] (i) (a) 핵산의 표적 영역에 대해 상동성이고 표적 영역에 비해 적어도 하나의 뉴클레오타이드의 돌연변이를 포함하는 영역 및 (b) 프로토스페이서 인접 모티프(PAM) 돌연변이를 포함하는 적어도 하나의 에디팅 카세트;

[0022] (ii) 적어도 하나의 프로모터; 및

[0023] (iii) (a) 표적 영역의 일부에 상보적인 영역(RNA) 및 (b) Cas9 뉴클레아제를 동원하는 영역(RNA)을 포함하는 적어도 하나의 가이드 RNA(gRNA)

[0024] 를 포함하는 벡터이다.

[0025] 본 벡터의 또 다른 실시양태는

[0026] (i) (a) 핵산의 표적 영역에 대해 상동성이고 표적 영역에 비해 적어도 하나의 코돈에서 적어도 하나의 뉴클레

오티드의 돌연변이를 포함하는 영역 및 (b) 프로토스페이서 인접 모티프(PAM) 돌연변이를 포함하는 적어도 하나의 에디팅 카세트;

[0027] (ii) 적어도 하나의 프로모터; 및

[0028] (iii) (a) 표적 영역의 일부에 상보적인 영역(RNA) 및 (b) Cas9 뉴클레아제를 동원하는 영역(RNA)을 포함하는 적어도 하나의 가이드 RNA(gRNA)

[0029] 를 포함하는 벡터이다.

[0030] 상기 실시양태들 중 임의의 실시양태에서, 본 벡터는 스페이서; 적어도 2개의 프라이밍 부위들; 또는 스페이서 및 적어도 2개의 프라이밍 부위들도 포함할 수 있다. 돌연변이가 적어도 하나의 코돈에서 적어도 하나의 뉴클레오티드의 돌연변이인 벡터에서, 에디팅 카세트 돌연변이는 예를 들면, PAM 돌연변이의 100개 뉴클레오티드들 내에 존재할 수 있다.

[0031] 본원에 기재된 방법에 의해 생성된 세포 집단을 포함하는 라이브러리도 기재되어 있다. 세포 집단의 라이브러리는 본원에 기재된 벡터들 중 임의의 벡터를 가진 세포를 포함할 수 있다. 예를 들면, 세포 집단은

[0032] (i) 세포 내의 핵산의 표적 영역에 대해 상동성이고 표적 영역에 비해 적어도 하나의 뉴클레오티드의 돌연변이(원하는 돌연변이로서 지칭됨)를 포함하는 영역, 및 프로토스페이서 인접 모티프(PAM) 돌연변이를 포함하는 에디팅 카세트;

[0033] (ii) 프로모터; 및

[0034] (iii) (a) 표적 영역의 일부에 상보적인 영역(RNA) 및 (b) Cas9 뉴클레아제를 동원하는 영역(RNA)을 포함하는 적어도 하나의 가이드 RNA(gRNA)

[0035] 를 포함하는 벡터를 포함할 수 있다.

[0036] 추가 실시양태에서, 세포 집단은

[0037] (i) 세포 내의 핵산의 표적 영역에 대해 상동성이고 표적 영역에 비해 적어도 하나의 코돈에서 적어도 하나의 뉴클레오티드의 돌연변이(원하는 돌연변이로서 지칭됨)를 포함하는 영역, 및 프로토스페이서 인접 모티프(PAM) 돌연변이를 포함하는 에디팅 카세트;

[0038] (ii) 프로모터; 및

[0039] (iii) (a) 표적 영역의 일부에 상보적인 영역(RNA) 및 (b) Cas9 뉴클레아제를 동원하는 영역(RNA)을 포함하는 적어도 하나의 가이드 RNA(gRNA)

[0040] 를 포함하는 벡터를 포함할 수 있다.

[0041] 추가 실시양태에서, 본 방법은

[0042] (a) (i) 제1 단일 프로토스페이서 인접 모티프(PAM)의 결실을 PAM에 인접한 유전자(인접 유전자) 내의 적어도 하나의 코돈의 돌연변이와 커플링시키는 하나 이상의 에디팅 올리고뉴클레오티드, 예컨대, 합리적으로 디자인된 올리고뉴클레오티드 및 (ii) 염색체의 오픈 리딩 프레임의 뉴클레오티드 서열 5'을 표적화하는 가이드 RNA(gRNA)를, 예컨대, 동시형질전환(co-transformation)으로 제1 세포 집단 내로 도입하여, 재조합 DNA, 예컨대, 재조합 염색체 또는 플라스미드 내의 재조합 DNA를 포함하는 도너 라이브러리를 구축함으로써, 표적화된 코돈 돌연변이를 가진 재조합 염색체를 포함하는 제1 세포 집단을 포함하는 도너 라이브러리를 생성하는 단계;

[0043] (b) 예컨대, 에디팅 올리고뉴클레오티드로부터의 합성 특징을 사용하고 동시에 유전자의 3' 말단에서 제2 PAM 결실(목표(destination) PAM 결실)을 도입하는 재조합 염색체의 PCR 증폭으로 (a)에서 구축된 도너 라이브러리를 증폭함으로써, 표적화된 코돈 돌연변이를 목표 PAM 결실에 직접 커플링, 예컨대, 공유커플링시키고 목표 PAM 결실 및 표적화된 코돈 돌연변이를 보유하는 회수된 도너 라이브러리를 생성하는 단계; 및

[0044] (c) 목표 PAM 결실 및 표적화된 코돈 돌연변이를 보유하는 도너 라이브러리 및 목표 gRNA 플라스미드를 전형적으로 나이브(naive) 세포 집단인 제2 세포 집단 내로 도입(예를 들면, 동시형질전환)함으로써, 표적화된 코돈 돌연변이를 포함하는 목표 라이브러리를 생성하는 단계

[0045] 를 포함하는, CRISPR-보조의 합리적 단백질 조작(조합적 게놈 조작) 방법이다.

- [0046] 제1 세포 집단 및 제2 세포 집단(예를 들면, 나이브 세포 집단)은 전형적으로 세포들이 모두 동일한 유형의 세포들이고 원핵생물 또는 진핵생물, 예컨대, 세균, 포유동물 세포, 식물 세포 또는 곤충 세포(그러나, 이들로 한정되지 않음)일 수 있는 집단이다.
- [0047] 일부 실시양태들에서, 본 방법은 단백질이 생성되는 조건 하에서 목표 라이브러리를 유지하는 단계도 포함한다.
- [0048] 일부 실시양태들에서, 제1 세포는 Cas9 뉴클레아제 활성을 가진 폴리펩티드를 발현한다. 일부 실시양태들에서, Cas9 뉴클레아제 활성을 가진 폴리펩티드는 유도성 프로모터의 조절 하에서 발현된다.
- [0049] 일부 실시양태들에서, 에디팅 올리고뉴클레오티드는 제1 세포에 존재하는 (하나, 하나 이상 또는 적어도 하나의) 표적 핵산에 상보적이다. 일부 실시양태들에서, 에디팅 올리고뉴클레오티드는 제1 세포 내의 하나 초과 의 표적 부위 또는 유전자좌를 표적화한다. 일부 실시양태들에서, 에디팅 올리고뉴클레오티드의 핵산 서열[원하는 코돈]은 표적 핵산에 비해 하나 이상의 치환, 결실, 삽입, 또는 치환, 결실 및 삽입의 임의의 조합물을 포함 한다. 일부 실시양태들에서, 에디팅 올리고뉴클레오티드들은 합리적으로 디자인되고; 추가 실시양태에서, 이들은 무작위 돌연변이유발에 의해, 또는 축퇴성 프라이머 올리고뉴클레오티드의 사용에 의해 생성된다. 일부 실시 양태들에서, 에디팅 올리고뉴클레오티드는 핵산의 콜렉션(라이브러리)으로부터 유래한다.
- [0050] 일부 실시양태들에서, gRNA는 플라스미드 상에 코딩되어 있다. 일부 실시양태들에서, 에디팅 올리고뉴클레오티 드 및 gRNA는 에디팅 올리고뉴클레오티드 및 가이드 (g)RNA의 형질전환, 예컨대, 동시형질전환에 의해 제1 세포 내로 도입된다. 일부 실시양태들에서, 에디팅 올리고뉴클레오티드 및 gRNA는 제1 세포 내로 순차적으로 도입된 다. 다른 실시양태들에서, 에디팅 올리고뉴클레오티드 및 gRNA는 제1 세포 내로 동시에 도입된다.
- [0051] 일부 실시양태들에서, 도너 라이브러리의 회수는 (a) 에디팅 올리고뉴클레오티드의 도입에 대해 세포를 스크리 닝하는 단계 및 (b) 에디팅 올리고뉴클레오티드를 도입한 것으로 확인되는 세포를 선택하는 단계도 포함한다. 일부 실시양태들에서, 도너 라이브러리의 회수는 회수된 도너 라이브러리의 프로세싱도 포함한다.
- [0052] 일부 실시양태들에서, 목표 세포/나이브 세포는 Cas9 뉴클레아제 활성을 가진 폴리펩티드를 발현한다. 일부 실 시양태들에서, Cas9 뉴클레아제 활성을 가진 폴리펩티드는 유도성 프로모터의 조절 하에서 발현된다.
- [0053] (a) (i) 에디팅 올리고뉴클레오티드를 포함하는 합성 dsDNA 에디팅 카세트 및 (ii) 관심 있는 유전자의 바로 상 류(upstream)에 위치하는 게놈 서열을 표적화하는 가이드 RNA(gRNA)를 발현하는 벡터를, gRNA에 의한 에디팅 올 리고뉴클레오티드의 다중화된 리컴바이너링 및 선택적 농축이 일어나는 조건 하에서 제1 세포 집단 내로 도입 (예를 들면, 동시형질전환)함으로써, 도너 라이브러리를 생성하는 단계;
- [0054] (b) 관심 있는 유전자의 3' 말단에 인접한 프로토스페이서 인접 모티프(PAM)(목표 PAM)를 결실시키는 올리고뉴 클레오티드로 상기 도너 라이브러리를 증폭함으로써, 목표 PAM이 결실된(3' PAM 결실을 가진) dsDNA 에디팅 카 세트, 합리적 코돈 돌연변이 및 P1 부위를 포함하는 증폭된 도너 라이브러리를 생성하는 단계;
- [0055] (c) 상기 증폭된 도너 라이브러리를 효소, 예컨대, 제한효소(예를 들면, BsaI)로 프로세싱하여 P1 부위를 제거 하는 단계; 및
- [0056] (d) (c)에서 프로세싱된 증폭된 도너 라이브러리 및 목표 gRNA로 나이브 세포 집단을 동시형질전환함으로써, 목 표 PAM이 결실된(3' PAM 결실을 가진) dsDNA 에디팅 카세트, 합리적 코돈 돌연변이 및 목표 gRNA를 포함하는 동 시형질전환된 세포 집단을 생성하는 단계
- [0057] 를 포함하는, CRISPR-보조의 합리적 단백질 조작 방법도 기재되어 있다.
- [0058] 기재된 모든 실시양태들에서, 돌연변이는 임의의 유형의 원하는 돌연변이, 예컨대, 하나 이상의 삽입, 결실, 치 환 또는 이들 중 2개 또는 3개의 임의의 조합물(예를 들면, 삽입 및 결실; 삽입 및 치환; 결실 및 치환; 치환 및 삽입; 삽입, 결실 및 치환)일 수 있다. 삽입, 결실 및 치환은 임의의 수의 뉴클레오티드의 삽입, 결실 및 치 환일 수 있다. 이들은 코돈(코딩 영역) 및/또는 비코딩 영역에 존재할 수 있다.

도면의 간단한 설명

- [0059] 도 1A 및 1B는 CRISPR 보조의 합리적 단백질 조작(CARPE)의 개요를 제공한다. 도 1A는 도너 라이브러리 구축의 개략도를 보여준다. 합성 dsDNA 에디팅 카세트는 관심 있는 유전자의 상류에 위치하는 게놈 서열을 표적화하는 가이드 RNA(gRNA)를 발현하는 벡터와 함께 동시형질전환되었다. 동시형질전환은 gRNA에 의해 선택적으로 농축되 는 에디팅 올리고뉴클레오티드의 다중화된 리컴바이너링을 통해 도너 라이브러리를 생성하였다. 그 다음, 유

전자의 3' 말단에 인접한 PAM(목표 PAM)을 돌연변이시키는(결실시키는) 올리고뉴클레오타이드를 사용하여 도너 라이브러리를 증폭하였다. 도 1B는 최종 단백질 라이브러리 생성의 개략도를 보여준다. 도너 라이브러리를 BsaI로 프로세싱하여 P1 부위를 제거하였고, 합리적 코돈 돌연변이 및 3' PAM 결실을 가진 dsDNA 카세트의 라이브러리는 목표 gRNA와 함께 동시형질전환되어 최종 단백질 라이브러리를 생성하였다.

도 2는 에디팅 올리고뉴클레오타이드의 P1 특징의 고효율 도입뿐만 아니라 표적화된 코돈 위치(밑줄로 표시됨)에서의 돌연변이도 확인시켜주는, *galK* 도너 라이브러리 구축으로부터의 클론으로부터의 DNA 서열을 제공한다. P1의 서열은 서열번호 1로 제공되어 있다.

도 3A는 프라이머 디자인을 보여준다. 도 3B는 프라이머의 수에 대해 예측된 밀도를 보여준다.

도 4a는 링커 및 구축물 결과를 제공한다. 도 4b는 에멀전(emulsion) PCR 기반 추적과 관련된 10개의 에디팅부위들(edits)을 보여준다.

도 5는 대사적 조작을 위한 합리적 단백질 에디팅의 개략도이다.

도 6은 CRISPR 농축된 합리적 단백질 라이브러리의 생성의 개략도이다.

도 7은 CARPE의 셋업 및 입증의 개략도이다.

도 8은 반복적 CRISPR 동시선택(co-selection) 방법을 보여준다.

도 9는 CARPE를 이용한 다중화된 단백질 조작 방법을 제공한다.

도 10은 CARPE를 이용한 *galK* 도너 라이브러리의 구축을 보여준다.

도 11a는 CARPE를 이용한 다중화된 CRISPR 기반 에디팅의 개략도를 보여준다. 도 11b는 추적가능한 CRISPR 농축된 리컴바이너링에 의한 게놈 조작(GEn-TraCER)을 이용한 다중화된 CRISPR 기반 에디팅의 개략도를 보여준다.

도 12는 *galK*의 코돈 24를 에디팅하기 위한 에디팅 카세트, 프로모터 및 스페이서를 포함하는 대표적인 GEn-TraCER 벡터(구축물)를 보여준다.

도 13은 GEn-TraCER을 이용한 *galK* 에디팅의 결과를 보여준다. 상부 패널은 *galK* 코돈 24 에디팅 GEn-TraCER 벡터로 형질전환된 세포로부터의 염색체 및 벡터(플라스미드)의 DNA 서열결정 결과로서, 상기 벡터 상의 에디팅 카세트(올리고뉴클레오타이드)가 원하는 게놈 에디팅부위(돌연변이)의 고효율 추적을 가능하게 하는 "트랜스-바코드(trans-barcode)"로서 서열결정될 수 있다는 것을 시사하는 DNA 서열결정 결과를 보여준다. 하부 패널은 비에디팅된 야생형 표현형(적색)을 나타내는 세포의 DNA 서열결정 크로마토그래프를 보여준다. 본 방법은 비에디팅된 야생형 대립형질 및 에디팅된/돌연변이된 대립형질 둘 다를 보유하는 다수의 염색체들을 가진 세포의 확인을 가능하게 한다.

도 14a 내지 14c는 GEn-TraCER의 개략도를 보여준다. 도 14a는 디자인 성분들의 개요를 보여준다. GEn-TraCER 카세트는 세포 게놈 내의 특정 부위를 표적화하고 dsDNA 절단을 야기하도록 가이드 RNA(gRNA) 서열(들)을 함유한다. 표적 영역에 상보적인 상동성 영역은 PAM 및 다른 인접 원하는 부위를 돌연변이시킨다. 재조합을 겪는 세포는 매우 풍부하게 존재하도록 선택적으로 농축된다. 벡터 내의 GEn-TraCER 에디팅 카세트의 서열결정은 게놈 에디팅부위/돌연변이의 추적을 가능하게 한다. 도 14b는 코돈 145에서 이. 콜라이(*E. coli*) *galK* 유전자에 대한 에디팅 카세트 디자인의 일례를 보여준다. PAM은 가장 가까운 이용가능한 PAM 위치에서 동일한 의미의 변화를 위해 만들어질 수 있는 가장 가까운 이용가능한 PAM 돌연변이에 의해 결실된다. 이것은 PAM 결실 부위에서 1개 또는 2개의 뉴클레오타이드의 "침묵 반흔(silent scar)"으로 돌연변이유발을 가능하게 한다. 도 14c는 어레이 기반 합성 방법을 이용하여 GEn-TraCER 카세트를 합성함으로써, 게놈 전체 규모로 체계적인 표적화 및 수천 개의 돌연변이들에 대한 적합도(fitness)의 동시적인 평가를 위해 적어도 10^4 개 내지 10^6 개의 카세트들의 병행 합성을 가능하게 할 수 있다는 것을 보여준다.

도 15A는 GEn-TraCER 벡터의 개요를 보여준다. 도 15B는 PAM 돌연변이와 돌연변이되는 코돈이 17개의 뉴클레오타이드들에 의해 분리되어 있는 이. 콜라이 *galK* 유전자 내의 Y145* 돌연변이의 생성을 위한 대표적인 GEn-TraCER의 일부를 보여준다. 대표적인 GEn-TraCER의 일부의 핵산 서열은 서열번호 28로 제공되어 있고, 역 상보체는 서열번호 33으로 제공되어 있다.

도 16A 내지 16C는 GEn-TraCER 디자인을 위한 대조군을 제공한다. 도 16A는 본 방법의 효능에 대한 에디팅 카세트의 크기의 효과를 보여준다. 도 16B는 본 방법의 효능에 대한 PAM 돌연변이/결실과 원하는 돌연변이 사이의

거리의 효과를 보여준다. 도 16C는 본 방법의 효능에 대한 MutS 시스템의 존재 또는 부재의 효과를 보여준다.

발명을 실시하기 위한 구체적인 내용

- [0060] 세균 및 고세균 CRISPR 시스템은 정밀 게놈 에디팅을 위한 강력한 신규 수단으로서 등장하였다. 스트렙토코커스 피오게네스(*Streptococcus pyogenes*)(에스. 피오게네스)로부터의 II형 CRISPR 시스템이 특히 시험관내에서 잘 특징규명되어 있고, 그의 이중 가닥 DNA(dsDNA) 결합 활성을 재프로그래밍하기 위한 단순한 디자인 규칙이 확립되어 있다(Jinek et al. Science (2012) 337(6096): 816-821). 세균(Cong et al. Science (2013) 339 (6121): 819-823), 사카로마이세스 세레비시아(*Saccharomyces cerevisiae*)(DiCarlo et al. Nucleic Acids Res. (2013) 41:4336-4343), 캐노르하브디티스 엘레간스(*Caenorhabditis elegans*)(Waaaijers et al. Genetics (2013) 195: 1187-1191) 및 다양한 포유동물 세포주들(Cong et al. Science (2013) 339 (6121): 819-823; Wang et al. Cell (2013) 153:910-918)을 비롯한 매우 다양한 유기체들에서의 CRISPR 매개 게놈 에디팅 방법의 이용은 문헌에 급속히 축적되었다. 다른 엔도뉴클레아제(endonuclease) 기반 게놈 에디팅 기술, 예컨대, 징크-핑거(zinc-finger) 뉴클레아제들(ZFN들), 귀소(homing) 뉴클레아제들 및 TALENs와 유사하게, 정밀 게놈 에디팅을 매개하는 CRISPR 시스템의 능력은 표적 인식의 고도로 특이적인 성질에 기인한다. 예를 들면, 에스케리키아 콜라이(*Escherichia coli*)로부터의 I형 CRISPR 시스템 및 에스. 피오게네스 시스템은 CRISPR RNA(crRNA)와 14-15 염기쌍 인식 표적 사이의 완벽한 상보성을 요구하는데, 이것은 CRISPR 시스템의 면역 기능이 천연적으로 사용된다는 것을 암시한다(Jinek et al. Science (2012) 337(6096): 816-821; Brouns et al. Science (2008) 321:960-964; Semenova et al. PNAS (2011) 108:10098-10103).
- [0061] DNA, 예컨대, 게놈 DNA에서 유도된 게놈 진화의 수행/변화(결실, 치환, 추가)의 발생을 위해 엔도뉴클레아제, 예컨대, *cas9* 유전자에 의해 코딩된 Cas9 뉴클레아제를 사용하는 게놈 에디팅 방법이 본원에 기재되어 있다. *cas9* 유전자는 임의의 공급원, 예컨대, 세균, 예컨대, 세균 에스. 피오게네스로부터 수득될 수 있다. *cas9*의 핵산 서열 및/또는 Cas9의 아미노산 서열은 천연 생성 *cas9* 및/또는 Cas9의 서열에 비해 돌연변이될 수 있고; 돌연변이는 예를 들면, 하나 이상의 삽입, 결실, 치환, 또는 이들 중 2개 또는 3개의 임의의 조합물일 수 있다. 이러한 실시양태에서, 생성된 돌연변이된 Cas9는 천연 생성 Cas9에 비해 향상된 또는 감소된 뉴클레아제 활성을 가질 수 있다.
- [0062] 도 1A, 1B 및 11a는 CRISPR 보조의 합리적 단백질 조작(CARPE)으로서 지칭되는 CRISPR 매개 게놈 에디팅 방법을 제공한다. CARPE는 단일 가닥 DNA(ssDNA) 또는 이중 가닥 DNA(dsDNA) 에디팅 카세트로부터의 유도된 돌연변이를 게놈 내로 직접 도입하는 "도너" 라이브러리 및 "목표" 라이브러리의 생성에 의존하는 2 단계 구축 방법이다. 도너 구축의 제1 단계(도 1A)에서, 합리적으로 디자인된 에디팅 올리고들을, 표적 DNA 서열, 예컨대, 오픈 리딩 프레임의 서열 5' 또는 관심 있는 다른 서열에 혼성화하거나/이러한 서열을 표적화하는 가이드 RNA(gRNA)와 함께 세포 내로 동시형질전환한다. CARPE의 핵심 혁신은 단일 프로토스페이서 인접 모티프(PAM)의 결실 또는 돌연변이를 인접 유전자 내의 하나 이상의 원하는 코돈의 돌연변이와 커플링시킴으로써, 단일 형질전환에서 전체 도너 라이브러리의 생성을 가능하게 하는 에디팅 올리고뉴클레오타이드의 디자인에 있다. 그 다음, 에디팅 올리고뉴클레오타이드로부터의 합성 특징을 사용하여, 예를 들면, PCR 반응으로 재조합 염색체를 증폭함으로써 도너 라이브러리를 회수하고; 동시에 제2 PAM 결실 또는 돌연변이를 상기 유전자의 3' 말단에서 도입한다. 따라서, 이 방법은 코돈-표적화된 돌연변이를 PAM 결실에 직접 공유커플링시킨다. CARPE의 제2 단계(도 1B)에서, 목표 PAM 결실/돌연변이 및 표적화된 돌연변이(하나 이상의 뉴클레오타이드, 예컨대, 하나 이상의 코돈 내의 하나 이상의 뉴클레오타이드의 원하는 돌연변이(들))를 보유하는 PCR-증폭된 도너 라이브러리를 목표 gRNA 벡터와 함께 나이트 세포 내로 동시형질전환하여, 합리적으로 디자인된 단백질 라이브러리를 발현하는 세포 집단을 생성한다.
- [0063] CRISPR 시스템에서, CRISPR 트랜스활성화(tracrRNA) 및 스페이서 RNA(crRNA)는 표적 영역의 선택을 안내한다. 본원에서 사용된 바와 같이, 표적 영역은 적어도 하나의 뉴클레오타이드의 돌연변이, 예컨대, 적어도 하나의 코돈(하나 이상의 코돈) 내의 적어도 하나의 뉴클레오타이드의 돌연변이가 요구되는 세포 또는 세포 집단의 핵산 내의 임의의 유전자좌를 지칭한다. 표적 영역은 예를 들면, 게놈 유전자좌(표적 게놈 서열) 또는 염색체의 유전자좌일 수 있다. tracrRNA 및 crRNA는 단일 가이드 RNA, 가이드 RNA 또는 gRNA로서 지칭되는 단일 키메라 RNA 분자로서 발현될 수 있다. gRNA의 핵산 서열은 표적 영역의 한 영역에 상보적인, 제1 영역으로서도 지칭되는 제1 핵산 서열, 및 줄기 루프 구조를 형성하고 Cas9를 표적 영역에 동원하는 기능을 수행하는, 제2 영역으로서도 지칭되는 제2 핵산 서열을 포함한다. 일부 실시양태들에서, gRNA의 제1 영역은 표적 게놈 서열의 상류에 위치하는 영역에 상보적이다. 일부 실시양태들에서, gRNA의 제1 영역은 표적 영역의 적어도 일부에 상보적이다. gRNA의 제1 영역은 표적 게놈 서열에 완전히 상보적(100% 상보적)일 수 있거나, 특이적으로 혼성화하거나/Cas9를 안내

하고 동원할 정도로 충분히 표적 게놈 서열에 상보적인 한, 하나 이상의 불일치를 포함할 수 있다. 일부 실시양태들에서, gRNA의 제1 영역은 길이에서 적어도 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 23개, 24개, 25개, 26개, 27개, 28개, 29개 또는 적어도 30개의 뉴클레오타이드들이다. 일부 실시양태들에서, gRNA의 제1 영역은 길이에서 적어도 20개의 뉴클레오타이드들이다. 일부 실시양태들에서, 제2 핵산 서열에 의해 형성되는 줄기 루프 구조는 길이에서 적어도 50개, 51개, 52개, 53개, 54개, 55개, 56개, 57개, 58개, 59개, 60개, 61개, 62개, 63개, 64개, 65개, 66개, 67개, 68개, 69개, 70개, 71개, 72개, 73개, 74개, 75개, 76개, 77개, 78개, 79개, 80개, 81개, 82개, 83개, 84개, 85개, 86개, 87개, 88개, 89개, 90개, 91개, 92개, 93개, 94개, 95개, 96개, 97개, 98개, 99개 또는 100개의 뉴클레오타이드들이다. 특정 실시양태에서, 줄기 루프 구조는 길이에서 80개 내지 90개, 또는 82개 내지 85개의 뉴클레오타이드들이고, 추가 특정 실시양태에서, 줄기 루프 구조를 형성하는 gRNA의 제2 영역은 길이에서 83개의 뉴클레오타이드들이다.

[0064] 일부 실시양태들에서, CARPE 방법을 이용하여 제1 세포 내로 도입하는 (도너 라이브러리의) gRNA의 서열은 제2/나이프 세포 내로 도입되는 (목표 라이브러리의) gRNA의 서열과 동일하다. 일부 실시양태들에서, 하나 초과 gRNA가 제1 세포 집단 및/또는 제2 세포 집단 내로 도입된다. 일부 실시양태들에서, 하나 초과 gRNA 분자는 하나 초과 표적 영역에 상보적인 제1 핵산 서열을 포함한다.

[0065] CARPE 방법에서, 기재된 방법에서 사용될, 에디팅 올리고뉴클레오타이드로서도 지칭되는 이중 가닥 DNA 카세트는 많은 공급원들로부터 수득될 수 있거나 유래할 수 있다. 예를 들면, 일부 실시양태들에서, dsDNA 카세트는 비상동성 무작위 재조합(NRR)에 의해 다양화된 핵산 라이브러리로부터 유래하고; 이러한 라이브러리는 NRR 라이브러리로서 지칭된다. 일부 실시양태들에서, 에디팅 올리고뉴클레오타이드는 예를 들면, 어레이 기반 합성에 의해 합성된다. 에디팅 올리고뉴클레오타이드의 길이는 에디팅 올리고뉴클레오타이드를 얻는 데에 이용된 방법에 의해 좌우될 수 있다. 일부 실시양태들에서, 에디팅 올리고뉴클레오타이드는 길이에서 대략 50개 내지 200개의 뉴클레오타이드들, 75개 내지 150개의 뉴클레오타이드들, 또는 80개 내지 120개의 뉴클레오타이드들이다.

[0066] 에디팅 올리고뉴클레오타이드는 (a) 세포의 핵산의 표적 영역에 대해 상동성이고 표적 영역에 비해 적어도 하나의 코돈의 돌연변이(원하는 돌연변이로서 지칭됨)를 포함하는 영역, 및 (b) 프로토스페이서 인접 모티프(PAM) 돌연변이를 포함한다. PAM 돌연변이는 CRISPR 시스템에 의해 더 이상 인식되지 않을 정도로 PAM의 서열을 돌연변이시키는 하나 이상의 뉴클레오타이드의 임의의 삽입, 결실 또는 치환일 수 있다. 이러한 PAM 돌연변이를 포함하는 세포는 CRISPR 매개 사멸에 대한 "면역" 세포인 것으로 주장될 수 있다. 표적 영역의 서열에 대한 원하는 돌연변이는 표적 영역의 적어도 하나의 코돈에서의 하나 이상의 뉴클레오타이드의 삽입, 결실 및/또는 치환일 수 있다.

[0067] CARPE 방법은 예시 목적으로만 세균 유전자와 관련하여 이하에 기재되어 있다. 상기 방법은 세균 및 고세균을 비롯한 임의의 원핵생물로부터의 유전자, 또는 효모 및 포유동물(인간을 포함함)을 비롯한 임의의 진핵생물 유전자를 비롯한 관심 있는 임의의 유전자(들)에 적용될 수 있다. 이. 콜라이 게놈 내의 *galK* 유전자에 대해 CARPE 방법을 수행하였는데, 이는 부분적으로 이 유전자에 대한 활성 분석의 이용가능성 때문이다. 리컴바이너어링을 매개하기 위해 BW23115 모균주 및 pSIM5 벡터(Datta et al. Gene (2008) 379:109-115)를 사용하여 상기 방법을 수행하였다. 아라비노스의 첨가에 의한 절단 활성의 조절을 가능하게 하기 위해 pBAD 프로모터의 조절하에 있는 pBTBx-2 골격 내로 *cas9* 유전자를 클로닝하였다. 축퇴성 프라이머들로부터, 및/또는 마이크로레이 기술을 통해 27,000 구성원 라이브러리의 일부로서 합성된 합리적으로 디자인된 올리고뉴클레오타이드들(올리고들)로부터 구축된 NNK 라이브러리로부터의 dsDNA 카세트를 사용하여 합성 dsDNA 카세트(127 bp)를 선택적으로 도입하는 능력의 평가를 수행하였다. 두 경우들에서, 올리고뉴클레오타이드는 *galK* 유전자 생성물의 활성 부위 잔기를 돌연변이시키도록 디자인되었다. *galK* 유전자좌로 향하는 프라이머로 수득한 앰플리콘 크기의 변화를 기초로 도너 균주 라이브러리의 고효율적인 회수를 확인하였다. NRR 라이브러리로부터의 이 콜로니 PCR 생성물들의 서열결정은 dsDNA 카세트로부터의 합성 프라이밍 부위(P1)가 약 90% 내지 100% 효율로 도입되었다는 것을 시사하였다. 이것은 다른 리컴바이너어링 기반 에디팅 방법(Costantino et al. PNAS (2003) 100:15748-15753; Wang et al. Nature (2009) 460:894-898)에서 전형적으로 사용되고 있는 오류 유발 *mutS* 녹아웃(knockout) 균주에 의존하지 않으면서 고효율로 이 라이브러리를 생성할 수 있다는 것을 시사하였다. 대립형질 교체 동안 *mutS* 보정에 기인할 수 있는, 코돈 돌연변이의 효율의 감소(약 20%)가 있었다. 목표 라이브러리 내의 클론들의 예비 평가는 구축의 두 단계들이 *mutS*⁺ 배경에서 수행될 때 최종 코돈 에디팅 효율이 약 10%이었다는 것을 시사하였다.

[0068] PAM과 코돈 돌연변이를 공유연결하지 않는 대신에 복제 동안 서로에 대한 이들의 인접성에 의존하는 대안적인

프로토콜을 이용하여 다른 최근에 공개된 동시선택가능한(co-selectable) 에디팅 프로토콜과의 비교를 수행하였다(Wang et al. Nat. Methods (2012) 9:591-593). 이 비-공유 실험에서, 전술된 바와 동일한 에디팅 올리고들을 사용하였고 동일한 도너/목표 PAM 부위를 표적화하는 ssDNA 올리고들을 사용하여 이들의 삽입에 대해 동시선택하고자 노력하였다. 생성된 돌연변이체들의 콜로니 스크리닝은 PAM 돌연변이체의 회수에서 고효율을 보여준다. 그러나, dsDNA 에디팅 카세트의 삽입을 위한 강력한 동시선택인 듯하지는 않다. 이것은 상당한 크기의 염색체 결실을 생성하는, PAM 결실 올리고뉴클레오타이드 및 에디팅 카세트의 상대적인 리컴바이너링 효율에서의 큰 차이에 기인할 수 있다.

[0069] 도너 구축 단계 동안 돌연변이의 상실을 방지하기 위한 노력으로 야생형 도너 균주에게 전달하기 전에, 예컨대, *mutS* 결핍 균주에서 도너 구축을 수행함으로써 CARPE 방법의 최종 에디팅 효율을 개선하는 능력을 평가할 수 있다. 추가로, 예컨대, *dxs*, *metA* 및 *folA*를 비롯한 다수의 필수 유전자들에 대해 CARPE를 이용함으로써 CARPE 방법의 일반성을 평가할 수 있다. 기재된 gRNA 디자인 방법을 이용하여 필수 유전자를 효과적으로 표적화하였다. 결과는 도너 라이브러리 생성 동안 일어나는 유전자 파괴에도 불구하고 도너 라이브러리를 효과적으로 구축할 수 있고 리컴바이너링 후 1시간 내지 3시간 이내에 회수할 수 있다는 것도 시사한다.

[0070] 추적가능한 CRISPR 농축된 리컴바이너링에 의한 게놈 조작(GEn-TraCER)으로서 지칭되는, CRISPR 매개 시스템을 사용하는 추적가능한 정밀 게놈 에디팅 방법도 본원에 제공되어 있다. GEn-TraCER 방법은 에디팅 카세트 및 gRNA 둘 다를 코딩하는 단일 벡터를 사용하여 고효율 에디팅/돌연변이를 달성한다. GEn-TraCER은 병행 DNA 합성, 예컨대, 어레이 기반 DNA 합성과 함께 이용될 때 수천 개의 정밀 에디팅부위들/돌연변이들의 단일 단계 생성을 제공하고 세포의 게놈(게놈 DNA)의 서열결정보다는 오히려 벡터 상의 에디팅 카세트의 서열결정으로 돌연변이를 맵핑하는 것을 가능하게 만든다. 상기 방법은 단백질 및 게놈 조작 적용뿐만 아니라 돌연변이, 예컨대, 실험실 진화 실험에서 확인된 돌연변이의 재구축에서도 넓은 유용성을 가진다.

[0071] GEn-TraCER 방법 및 벡터는 단일 벡터 상에서 원하는 돌연변이 및 PAM 돌연변이를 포함하는 에디팅 카세트와 gRNA를 코딩하는 유전자를 병용하는데, 이것은 단일 반응에서 돌연변이의 라이브러리를 생성하는 것을 가능하게 만든다. 도 11b에 나타난 바와 같이, 상기 방법은 원하는 돌연변이 및 PAM 돌연변이를 포함하는 에디팅 카세트를 포함하는 벡터를 세포 또는 세포 집단 내로 도입하는 단계를 포함한다. 일부 실시양태들에서, 상기 벡터가 도입되는 세포는 Cas9도 코딩한다. 일부 실시양태들에서, Cas9를 코딩하는 유전자는 추후에 세포 또는 세포 집단 내로 도입된다. 세포 또는 세포 집단에서 Cas9 및 gRNA를 포함하는 CRISPR 시스템의 발현이 활성화되고; gRNA는 Cas9를 dsDNA 절단이 일어나는 표적 영역에 동원한다. 임의의 특정 이론에 의해 구속받으려 하지 않지만, 표적 영역에 상보적인 에디팅 카세트의 상동성 영역은 PAM 및 표적 영역의 하나 이상의 코돈을 돌연변이시킨다. PAM 돌연변이를 도입하지 않은 세포 집단의 세포들은 Cas9 매개 dsDNA 절단으로 인한 비에디팅된 세포 사멸을 겪는다. PAM 돌연변이를 도입한 세포 집단의 세포들은 세포 사멸을 겪지 않고; 이들은 생존 상태로 유지되고 매우 풍부하게 존재하도록 선택적으로 농축된다. 생존 세포는 수득되고 표적화된 돌연변이의 라이브러리를 제공한다.

[0072] GEn-TraCER을 이용하는 추적가능한 게놈 에디팅 방법은 (a) 적어도 하나의 에디팅 카세트, 프로모터 및 적어도 하나의 gRNA를 코딩하는 벡터를 세포 또는 세포 집단 내로 도입함으로써, 상기 벡터를 포함하는 세포 또는 세포 집단(제2 세포 집단)을 생성하는 단계; (b) 상기 벡터, 제2 벡터 또는 제2 세포 집단의 세포의 게놈 상에 코딩되어 있는 Cas9 뉴클레아제가 발현되는 조건 하에서 제2 세포 집단을 유지함으로써, PAM 돌연변이를 포함하지 않는 제2 세포 집단의 세포의 DNA 절단 및 사멸을 야기하는 반면, PAM 돌연변이를 포함하는 제2 세포 집단의 세포가 생존가능하게 하는 단계; (c) 생존 세포를 얻는 단계; 및 (d) 제2 세포 집단의 적어도 하나의 세포 내의 상기 벡터의 에디팅 카세트를 서열결정하여 적어도 하나의 코돈의 돌연변이를 확인하는 단계를 포함한다.

[0073] 일부 실시양태들에서, *cas9*를 코딩하는 별개의 벡터도 세포 또는 세포 집단 내로 도입된다. 당분야에서 공지된 임의의 방법 또는 기법을 이용하여 세포 또는 세포 집단 내로의 벡터의 도입을 수행할 수 있다. 예를 들면, 표준 프로토콜, 예컨대, 화학적 형질전환 및 전기천공, 형질도입 및 입자 충격(bombardment)을 비롯한 형질전환으로 벡터를 도입할 수 있다.

[0074] 에디팅 카세트는 (a) 세포 또는 세포 집단 내의 핵산의 표적 영역을 인식하고(이 영역에 혼성화하고) 세포의 핵산의 표적 영역에 대해 상동성이고 표적 영역에 비해 적어도 하나의 코돈에서 적어도 하나의 뉴클레오타이드의 돌연변이(원하는 돌연변이로서 지칭됨)를 포함하는 영역, 및 (b) 프로토스페이스 인접 모티프(PAM) 돌연변이를 포함한다. PAM 돌연변이는 돌연변이된 PAM(PAM 돌연변이)이 CRISPR 시스템에 의해 인식되지 않도록 PAM의 서열을 돌연변이시키는 하나 이상의 뉴클레오타이드의 임의의 삽입, 결실 또는 치환일 수 있다. 이러한 PAM 돌연변이를

포함하는 세포는 CRISPR 매개 사멸에 대한 "면역" 세포인 것으로 주장될 수 있다. 표적 영역의 서열에 대한 원하는 돌연변이는 표적 영역의 적어도 하나의 코돈에서 하나 이상의 뉴클레오타이드의 삽입, 결실 및/또는 치환일 수 있다. 일부 실시양태들에서, PAM 돌연변이와 원하는 돌연변이 사이의 거리는 에디팅 카세트 상의 적어도 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 23개, 24개, 25개, 26개, 27개, 28개, 29개 또는 30개 뉴클레오타이드들이다. 일부 실시양태들에서, PAM 돌연변이는 에디팅 카세트의 말단으로부터 적어도 9개의 뉴클레오타이드들에 위치한다. 일부 실시양태들에서, 원하는 돌연변이는 에디팅 카세트의 말단으로부터 적어도 9개의 뉴클레오타이드들에 위치한다.

[0075] 일부 실시양태들에서, 표적 영역의 서열에 대한 원하는 돌연변이는 핵산 서열의 삽입이다. 표적 영역 내로 삽입된 핵산 서열은 임의의 길이의 핵산 서열일 수 있다. 일부 실시양태들에서, 삽입된 핵산 서열은 길이에서 적어도 50개, 100개, 150개, 200개, 250개, 300개, 350개, 400개, 450개, 500개, 550개, 600개, 650개, 700개, 750개, 800개, 850개, 900개, 950개, 1000개, 1100개, 1200개, 1300개, 1400개, 1500개, 1600개, 1700개, 1800개, 1900개 또는 적어도 2000개의 뉴클레오타이드들이다. 핵산 서열이 표적 영역 내로 삽입되는 실시양태에서, 에디팅 카세트는 길이에서 적어도 30개, 31개, 32개, 33개, 34개, 35개, 36개, 37개, 38개, 39개, 40개, 41개, 42개, 43개, 44개, 45개, 46개, 47개, 48개, 49개, 50개, 51개, 52개, 53개, 54개, 55개, 56개, 57개, 58개, 59개 또는 적어도 60개의 뉴클레오타이드들이고 표적 영역에 대한 상동성을 가진 영역을 포함한다.

[0076] 용어 "GEn-TraCER 카세트"는 에디팅 카세트, 프로모터, 스페이서 서열 및 gRNA를 코딩하는 유전자의 적어도 일부를 지칭하기 위해 사용될 수 있다. 일부 실시양태들에서, GEn-TraCER 카세트 상에서 gRNA를 코딩하는 유전자의 일부는 표적 영역에 상보적인 gRNA의 일부를 코딩한다. 일부 실시양태들에서, 표적 영역에 상보적인 gRNA의 일부는 길이에서 적어도 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 23개, 24개, 25개, 26개, 27개, 28개, 29개 또는 적어도 30개의 뉴클레오타이드들이다. 일부 실시양태들에서, 표적 영역에 상보적인 gRNA의 일부는 길이에서 24개의 뉴클레오타이드들이다. 일부 실시양태들에서, GEn-TraCER 카세트는 적어도 2개의 프라이밍 부위들을 추가로 포함한다. 일부 실시양태들에서, 프라이밍 부위는 예를 들면, PCR로 GEn-TraCER 카세트를 증폭하는 데에 사용될 수 있다. 일부 실시양태들에서, 표적 영역에 상보적인 gRNA의 일부가 프라이밍 부위로서 사용된다.

[0077] GEn-TraCER 방법에서, 기재된 방법에서 사용될 에디팅 카세트 및 GEn-TraCER 카세트는 많은 공급원들로부터 수득될 수 있거나 유래할 수 있다. 예를 들면, 일부 실시양태들에서, 에디팅 카세트는 예를 들면, 어레이 기반 합성에 의해 합성된다. 일부 실시양태들에서, GEn-TraCER 카세트는 예를 들면, 어레이 기반 합성에 의해 합성된다. 에디팅 카세트 및/또는 GEn-TraCER 카세트의 길이는 에디팅 카세트 및/또는 GEn-TraCER 카세트를 얻는 데에 이용된 방법에 의해 좌우될 수 있다. 일부 실시양태들에서, 에디팅 카세트는 길이에서 대략 50개 내지 300개의 뉴클레오타이드들, 75개 내지 200개의 뉴클레오타이드들, 또는 80개 내지 120개의 뉴클레오타이드들이다. 일부 실시양태들에서, GEn-TraCER 카세트는 길이에서 대략 50개 내지 300개의 뉴클레오타이드들, 75개 내지 200개의 뉴클레오타이드들, 또는 80개 내지 120개의 뉴클레오타이드들이다.

[0078] 일부 실시양태들에서, 본 방법은 예를 들면, 어레이 기반 합성으로 GEn-TraCER 카세트를 얻는 단계 및 벡터를 구축하는 단계도 포함한다. 벡터를 구축하는 방법은 당분야에서 통상의 기술을 가진 자에게 공지되어 있을 것이고 GEn-TraCER 카세트를 벡터에 결합시키는(ligating) 단계를 포함할 수 있다. 일부 실시양태들에서, GEn-TraCER 카세트, 또는 GEn-TraCER 카세트들의 서브세트(폴)는 벡터의 구축 전에 예를 들면, PCR에 의해 증폭된다.

[0079] 상기 벡터를 포함하고 Cas9도 코딩하는 세포 또는 세포 집단은 Cas9가 발현되는 조건 하에서 유지되거나 배양된다. Cas9 발현은 조절될 수 있다. 본원에 기재된 방법은 Cas9 발현이 활성화되는 조건 하에서 세포를 유지하여 Cas9의 생성을 야기하는 단계를 포함한다. Cas9가 발현되는 구체적인 조건은 Cas9 발현을 조절하는 데에 사용된 프로모터의 성질과 같은 요인에 의해 좌우될 것이다. 일부 실시양태들에서, Cas9 발현은 유도제 분자, 예컨대, 아라비노스의 존재 하에서 유도된다. Cas9 코딩 DNA를 포함하는 세포 또는 세포 집단이 유도제 분자의 존재 하에 있을 때, Cas9의 발현이 일어난다. 일부 실시양태들에서, Cas9 발현은 억제제 분자의 존재 하에서 억제된다. Cas9 코딩 DNA를 포함하는 세포 또는 세포 집단이 Cas9의 발현을 억제하는 분자의 부재 하에 있을 때, Cas9의 발현이 일어난다.

[0080] 생존 상태로 유지되는 세포 또는 세포 집단은 Cas9 매개 사멸의 결과로서 비에디팅된 세포 사멸을 겪는 세포로부터 수득되거나 분리되고; 이것은 예를 들면, 세포 집단을 배양 표면 상에 분산시켜, 추후에 평가를 위해 사용될 수 있는 생존 세포의 성장을 가능하게 함으로써 수행될 수 있다.

[0081] 상기 집단의 생존 세포(PAM 돌연변이를 도입한 세포) 내의 벡터 상의 에디팅 카세트를 서열결정함으로써 GEn-

TraCER 방법을 이용하여 PAM 돌연변이에 커플링된 원하는 돌연변이를 추적할 수 있다. 이것은 세포의 게놈을 서열결정할 필요 없이 돌연변이의 용이한 확인을 가능하게 한다. 본 방법은 하나 이상의 코돈의 돌연변이를 확인하기 위한 에디팅 카세트의 서열결정을 포함한다. 에디팅 카세트의 서열결정은 벡터의 성분으로서 수행될 수 있거나, 벡터로부터의 그의 분리 및 임의적으로 증폭 후에 수행될 수 있다. 서열결정은 당분야에서 공지된 임의의 서열결정 방법, 예컨대, 생거(Sanger) 서열결정을 이용함으로써 수행될 수 있다.

[0082] 본원에 기재된 방법은 원핵세포 및 진핵세포를 비롯한, CRISPR 시스템이 작용할 수 있는(예를 들면, DNA를 표적화하고 절단할 수 있는) 임의의 유형의 세포에서 수행될 수 있다. 일부 실시양태들에서, 세포는 세균 세포, 예컨대, 에스케리키아 종(예를 들면, 이. 콜라이)이다. 다른 실시양태에서, 세포는 진균 세포, 예컨대, 효모 세포, 예를 들면, 사카로마이세스 종이다. 다른 실시양태에서, 세포는 조류 세포, 식물 세포, 곤충 세포, 또는 인간 세포를 비롯한 포유동물 세포이다.

[0083] "벡터"는 세포에게 전달되거나 세포에서 발현될 원하는 서열 또는 서열들을 포함하는 다양한 핵산들 중 임의의 핵산이다. 원하는 서열(들)은 예컨대, 제한 및 절찰, 또는 재조합에 의해 벡터 내에 포함될 수 있다. RNA 벡터도 사용가능하지만, 벡터는 전형적으로 DNA로 구성된다. 벡터는 플라스미드, 포스미드(fosmids), 파지미드(phagemids), 바이러스 게놈 및 인공 염색체를 포함하나 이들로 한정되지 않는다.

[0084] GEN-TraCER 방법에서 유용한 벡터는 본원에 기재된 적어도 하나의 에디팅 카세트, 프로모터 및 gRNA를 코딩하는 적어도 하나의 유전자를 포함한다. 일부 실시양태들에서, 하나 초과(예를 들면, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개 또는 10개 이상의 에디팅 카세트들)이 벡터 상에 포함된다. 일부 실시양태들에서, 하나 초과(예를 들면, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개 또는 10개 이상의 gRNA들)를 코딩하는 하나 초과(예를 들면, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개 또는 10개 이상의 gRNA들)의 일부에 상보적인 영역을 함유한다(예를 들면, 상이한 gRNA들이 존재하고, 이들 각각은 상이한 표적 영역의 일부에 상보적이다).

[0085] 일부 실시양태들에서, 적어도 하나의 에디팅 카세트, 프로모터 및 gRNA의 일부를 코딩하는 유전자를 포함하는 GEN-TraCER 카세트는 gRNA의 또 다른 일부를 코딩하는 벡터 내로 결합된다. 결합 시, GEN-TraCER 카세트로부터의 gRNA의 일부와 gRNA의 다른 일부는 결합되어 기능성 gRNA를 형성한다.

[0086] 프로모터와 gRNA를 코딩하는 유전자는 작동가능하게 연결된다. 일부 실시양태들에서, 본 방법은 Cas9를 코딩하는 제2 벡터의 도입을 포함한다. 이러한 실시양태에서, 벡터는 Cas9를 코딩하는 유전자에 작동가능하게 연결된 하나 이상의 프로모터도 포함할 수 있다. 본원에서 사용된 바와 같이, "작동가능하게" 연결된 프로모터가 유전자, 예컨대, gRNA를 코딩하는 유전자 또는 Cas9를 코딩하는 유전자를 코딩하는 DNA의 전사에 영향을 미치거나 이러한 전사를 조절한다는 것을 의미한다. 프로모터는 천연 프로모터(벡터가 도입되는 세포에 존재하는 프로모터)일 수 있다. 일부 실시양태들에서, 프로모터는 유도가능한 또는 억제가능한 프로모터(이 프로모터는 조절되어, 유전자, 예컨대, gRNA를 코딩하는 유전자 또는 Cas9를 코딩하는 유전자의 유도가능한 또는 억제가능한 전사를 가능하게 함), 예컨대, 분자(예를 들면, 유도제 또는 억제제)의 존재 또는 부재에 의해 조절되는 프로모터이다. gRNA의 발현에 필요한 프로모터의 성질은 종 또는 세포 유형에 따라 달라질 수 있고 당분야에서 통상의 기술을 가진 자에 의해 인식될 것이다.

[0087] 일부 실시양태들에서, 본 방법은 본원에 기재된 적어도 하나의 에디팅 카세트, 프로모터 및 적어도 하나의 gRNA를 포함하는 벡터의 도입 전에 또는 도입과 동시에 Cas9를 코딩하는 별개의 벡터를 세포 또는 세포 집단 내로 도입하는 단계를 포함한다. 일부 실시양태들에서, Cas9를 코딩하는 유전자는 세포 또는 세포 집단의 게놈 내로 삽입된다. Cas9 코딩 DNA는 본원에 기재된 적어도 하나의 에디팅 카세트, 프로모터 및 적어도 하나의 gRNA를 포함하는 벡터의 도입 전에 또는 본원에 기재된 적어도 하나의 에디팅 카세트, 프로모터 및 적어도 하나의 gRNA를 포함하는 벡터의 도입 후에 세포 게놈 내로 삽입될 수 있다. 대안적으로, 핵산 분자, 예컨대, Cas9를 코딩하는 DNA는 게놈 내로 삽입된 DNA로부터 발현될 수 있다. 일부 실시양태들에서, Cas9를 코딩하는 유전자는 세포의 게놈 내로 삽입된다.

[0088] 본원에 기재된 GEN-TraCER 방법에서 유용한 벡터는 스페이서 서열, 2개 이상의 프라이밍 부위들, 또는 스페이서 서열 및 2개 이상의 프라이밍 부위들 둘 다를 추가로 포함할 수 있다. 일부 실시양태들에서, GEN-TraCER 카세트를 플랜킹하는 프라이밍 부위들의 존재는 에디팅 카세트, 프로모터 및 gRNA 핵산 서열의 증폭을 가능하게 한다.

- [0089] 실시예
- [0090] 실시예 1: *galK*를 에디팅하기 위한 CARPE 방법의 이용
- [0091] 이. 콜라이 게놈 내의 갈락토키나제(galactokinase) 유전자인 *galK*에 대해 CARPE 방법을 수행하였고; 유전자 생성물의 활성을 평가하기 위한 많은 이용가능한 분석들이 있다. 리컴바이니어링을 매개하기 위해 이. 콜라이 BW23115 모균주 및 pSIM5 벡터(Datta et al. Gene (2008) 379:109-115)를 사용하여 실험을 수행하였다. Cas9를 코딩하는 유전자를 pBAD 프로모터의 조절 하에 있는 pBTBX-2 골격 내로 클로닝함으로써, 아라비노스를 배양 배지에 첨가하여 Cas9 절단 활성을 조절할 수 있게 하였다.
- [0092] 먼저, 합성 dsDNA 카세트(127 bp)를 선택적으로 도입하는 능력을 시험하였다. 합성 dsDNA 카세트는 축퇴성 프라이머들로부터, 또는 마이크로레이 기술을 통해 27,000 구성원 라이브러리의 일부로서 합성된 합리적으로 디자인된 올리고들로부터 구축된 NNR 라이브러리로부터 유래하였다. 두 경우들에서, 올리고뉴클레오티드는 *galK* 유전자 생성물의 활성 부위 잔기를 돌연변이시킬뿐만 아니라 합성 프라이밍 부위인 P1(서열번호 1)도 함유하도록 디자인되었다. *galK* 유전자좌로 향하는 프라이머들을 사용한 콜로니 PCR에 의해 얻은 앰플리콘 크기의 변화를 기초로 도너 균주 라이브러리의 고효율 회수를 확인하였다. NNR 라이브러리로부터의 콜로니 PCR 생성물의 서열 결정은 dsDNA 카세트로부터의 합성 프라이밍 부위(P1)가 약 90% 내지 100% 효율로 도입되었다는 것을 시사하였다(도 2). 이 놀라운 예상외의 결과는 라이브러리가 다른 리컴바이니어링 기반 에디팅 방법(Constantino, et al. PNAS (2003) 100:15748-15753; Wang et al. Nature (2009) 460: 894-898)에서 전형적으로 사용되고 있는 오류 유발 *mutS* 결핍 균주에 의존하지 않으면서 고효율로 생성될 수 있다는 것을 암시한다. 그러나, 대립형질 교체 동안 *MutS*에 의한 보정에 기인할 수 있는, 코돈 돌연변이 효율의 감소(약 20%)가 있었다. 이 연구에서, 구축의 두 단계들이 *mutS* 배경에서 수행되었을 때 최종 코돈 에디팅 효율은 약 10%이었다.
- [0093] CARPE 방법의 최종 에디팅 효율 및 일반성을 향상시키기 위해, 도너 구축 단계 동안 돌연변이의 상실을 방지하기 위한 노력으로 *mutS* 도너 균주에게 전달하기 전에 도너 구축을 *mutS* 결핍 균주에서 수행할 수 있다.
- [0094] 실시예 2: 필수 유전자를 표적화하기 위한 CARPE 방법의 이용
- [0095] CARPE 방법의 일반성을 시험하기 위해, 전술된 바와 같이, *dxs*, *metA* 및 *folA*를 비롯한 다수의 필수 유전자들에 대해 상기 방법을 이용하였다. gRNA 디자인 방법을 이용하여 필수 유전자를 표적화할 수 있다(도 3).
- [0096] *dxs* 유전자를 표적화하는 CARPE 실험으로부터의 데이터도 도너 라이브러리 생성 동안 일어나는 유전자 파괴에도 불구하고 리컴바이니어링 후 1시간 내지 3시간 이내에 도너 라이브러리를 효율적으로 구축하고 회수할 수 있다는 것을 암시한다.
- [0097] 실시예 3: 이소펜텐올(isopentenol)의 생성을 조절하기 위한 CARPE 방법의 이용
- [0098] 세균 생산을 통한 산업적 제조를 위한 보다 우수한 생물연료에 대한 탐구는 원하는 생성물을 위해 최신기술의 게놈 디자인, 조작 및 스크리닝을 수행하는 능력을 요구한다. 이전에, 본 발명자들은 이. 콜라이 게놈 내의 모든 유전자들의 발현 수준을 개별적으로 변경시키는 능력을 입증하였다(Warner et al. Nat. Biotechnol. (2010) 28:856-862). 추적가능한 다중체 리컴바이니어링(TRMR)으로서 지칭되는 이 방법은 약 8000개의 게놈적으로 변경된 세포들(약 4000개의 과다발현된 유전자들 및 약 4000개의 녹다운된 유전자들)의 라이브러리를 생성하였다. 이 라이브러리는 나중에 상이한 조건들 하에서 스크리닝되었는데, 이것은 유전자 생성물의 활성의 보다 더 깊은 이해를 가능하게 하였고 이 선택 하에서 보다 더 우수한 성능의 균주를 발생시켰다. TRMR은 두 수준들(과다발현 및 녹다운)에 대한 단백질 발현의 변경을 가능하게 하였으나 오픈 리딩 프레임(ORF)의 변경을 가능하게 하지 못하였다. 여기서, 본 발명자들은 생물연료의 최적 생산을 위해 ORF 변경 및 전체 대사 경로의 조작의 큰 라이브러리를 생성하는 것을 목표로 한다.
- [0099] (무작위 돌연변이유발과 대조적으로) 합리적으로 디자인되는 이러한 라이브러리의 제조에 있어서 주된 어려움은 표적 세포 내로의 원하는 돌연변이의 삽입 효율이다. 이. 콜라이에서 게놈을 변경시키는 표준 방법인 리컴바이니어링은 숙주 게놈 내로의 외래 DNA의 삽입을 용이하게 하기 위해 람다 파지로부터의 재조합 유전자를 사용한다. 그러나, 이 과정은 저효율이라는 문제점을 갖고 (TRMR에서와 같이) 항생제 내성 유전자의 추가에 이은 선택에 의해, 또는 재조합 사건의 반복적 유도에 의해(즉, MAGE(Wang et al. Nature (2008) 460:894-898)에 의해) 극복될 수 있다. 본원에 기재된 CARPE 방법은 집단으로부터 모든 비-재조합 세포들을 제거하기 위해 CRISPR 시스템의 사용을 수반하는 리컴바이니어링 효율을 증가시킨다. CRISPR은 침입 파지 및 플라스미드에 대항하는 세균 및 고세균의 최근에 발견된 RNA 기반 적응 방어 기작이다(Bhaya et al. Ann. Rev. of Genetics (2011))

45:273-297). 이 시스템은 2개의 플라스미드들, 즉 CRISPR 관련 뉴클레아제 Cas9를 코딩하는 한 플라스미드, 및 Cas9를 그의 유일무이한 위치로 안내하는 서열 특이적 가이드 RNA(gRNA)를 코딩하는 제2 플라스미드를 사용한 서열-유도된 이중 가닥 절단을 가능하게 하기 위해 대규모 조작을 받았다(Qi et al. Cell (2013) 45:273-297). CARPE 방법은 서열 의존적 방식으로 DNA 절단 및 결과적으로 세포 사멸을 유도하기 위해 CRISPR 시스템의 능력을 이용한다. 본 발명자들은 ORF 내의 원하는 돌연변이 이외에 CRISPR 기구에 의해 표적화되는 유전자의 오픈 리딩 프레임 외부의 공통된 위치에서 돌연변이를 포함하는 DNA 리컴바이너리링 카세트를 생성하였다. 원하는 돌연변이를 PAM 돌연변이/결실로 인한 CRISPR 매개 사멸로부터의 회피와 연관짓는/커플링시키는 이 방법은 전체 세포 집단 내에서의 조작된 세포의 현저한 농축을 가능하게 한다.

[0100] 본 방법은 DXS 경로를 사용함으로써도 입증된다. DSX 경로는 테르펜(terpenes) 및 테르페노이드(terpenoids)의 생합성을 야기하는 이소펜테닐 피로포스페이트(IPP)의 생성을 야기한다. 흥미롭게도, 요구된 유전자가 추가되면, IPP는 라이코펜(lycopene) 또는 이소펜테놀의 전구체일 수도 있다. 라이코펜이 세균 콜로니를 적색으로 만들어 용이하게 스크리닝될 수 있게 하므로, 이소펜테놀은 에탄올보다 더 높은 에너지 밀도 및 더 낮은 물 혼화성을 가진 '차세대' 생물연료인 것으로 간주된다. 조작을 위해 3종의 단백질을 선택하였다: 1) 상기 경로의 첫 번째 속도 제한 효소인 DSX, 2) 대사적 유동을 DXS 경로로부터 우회시키는 IspB, 및 3) 이. 콜라이 및 비. 서브틸리스(*B. subtilis*) 둘 다에서 IPP를 이소펜테놀로 전환시키는 것으로 밝혀진 NudF(Withers et al. App. Environ. Microbiol (2007) 73: 6277-6283; Zheng et al. Biotechnol. for biofuels (2013)6:57). 콜로니 색채 정량을 위해 개발된 신규 영상 분석 수단으로 DXS 및 IspB를 코딩하는 유전자들에서의 돌연변이를 증가된 라이코펜 생성에 대해 스크리닝할 것이다. GC/MS로 이소펜테놀 수준을 측정함으로써 NudF 활성을 직접적으로 분석할 것이고 바이오센서로서 사용될 이소펜테놀 영양요구성 세포로 NudF 활성을 간접적으로 분석할 것이다. 이 방법은 큰 돌연변이 라이브러리를 높은 정확성 및 효율로 이. 콜라이 개능 내로 합리적으로 조작하는 능력 및 고수율의 이소펜테놀을 생성하는 균주를 제공한다.

[0101] 실시예 4: *galK*를 에디팅하기 위한 GEn-TraCER 방법의 이용

[0102] 이. 콜라이에서 리컴바이너리링을 위한 모델 시스템으로서 사용되는 *galK* 유전자를 에디팅하기 위해 GEn-TraCER 방법을 이용하였다(Yu et al. 2000). 구축된 첫 번째 GEn-TraCER 카세트는 *galK*_Q24로서 지칭되는, *galK*의 코돈 24에서 인프레임(inframe) PAM 대신에 정지 코돈을 도입하도록 디자인되었다(도 12). 필요한 돌연변이를 고 처리율로 생성하기 위해 맞춤 파이톤 스크립트(custom python script)를 사용하여 구축물 및 벡터를 디자인하였다.

[0103] 원형 중합효소 클로닝(CPEC) 방법을 이용하여 대조군 카세트를 문헌(Qi et al. Cell (2013))에 기재된 gRNA 벡터 내로 클로닝하였다. 골격을 하기 프라이머들로 선형화하였다: CCAGAAATCATCCTTAGCGAAAGCTAAGGAT(서열번호 29) 및 GTTTTAGAGCTAGAAATAGCAAGTAAAATAAGGCT(서열번호 30).

[0104] GenTRACER 카세트를 지블록(gblocks)으로서 정돈하고 하기 프라이머들을 사용하여 증폭하였다:

[0105] ATCACGAGGCAGAAATTCAGATAAAAAAATCCTTAGCTTTCGCTAAGGATGATTCTGG(서열번호 31),

[0106] ACTTTTCAAGTTGATAACGACTAGCCTTATTTTAACTTGCTATTCTAGCTCTAAAC(서열번호 32).

[0107] CPEC를 이용하여 성분들을 함께 연결하고 이. 콜라이 내로 형질전환하여 벡터를 생성하였다. 이 절차는 대략 10^4 내지 10^5 CFU/ μ g의 클로닝 효율로 풀링된 올리고뉴클레오타이드 라이브러리들을 사용함으로써 다중으로 수행되어야 한다.

[0108] pSIM5(람다-RED 플라스미드) 및 X2-*cas9* 플라스미드를 보유하는 이. 콜라이 MG1655 세포를, 50 μ g/ml 카나마이신 및 34 μ g/ml 클로람페니콜을 함유하는 30°C의 LB에서 중간 대수기(mid log phase)(0.4 내지 0.7 OD)까지 생장시켰다. pSIM5 벡터의 리컴바이너리링 기능을 15분 동안 42°C에서 유도한 후 10분 동안 얼음 위에 놓아두었다. 그 다음, 펠렛화 및 10 ml 냉각된 H₂O를 사용한 2회 세척으로 세포를 일렉트로컴피턴트(electrocompetent) 세포로 만들었다. 세포를 100 ng의 GEn-TraCER 플라스미드(카베니실린 내성도 코딩함)로 형질전환하고 37°C에서 3시간 동안 회수하였다. 50 내지 100 μ l의 세포를, 50 μ g/ml 카나마이신 및 100 μ g/ml 카베니실린을 함유하는 적절한 배지에 플레이팅하여 CRISPR 에디팅된 균주를 선택적으로 농축하였다. 갈락토스로 보충된 맥콘키(MacConkey) 한천 상에서의 적색/백색 스크리닝을 이용하여 *galK* 유전자에 대한 에디팅 효율을 계산하였다.

[0109] 맥콘키 한천 상에서의 스크리닝을 기초로, *galK*_Q24* 디자인으로 약 100%의 에디팅 효율을 관찰하였다. 흥미롭

게도, 고효율을 달성하기 위해 불일치 복구 너아웃을 요구하는 올리고 매개 리컴바이니어링 방법(Li et al. 2003; Sawitzke et al. 2011; Wang et al. 2011))과 달리, 온전한 상태로 불일치 복구 기구를 갖거나 갖지 않은 균주에서 영향이 전혀 없었다.

[0110] 그 다음, 염색체 및 벡터 서열을 생거 서열결정으로 확인하였다.

[0111] 예측된 바와 같이, 벡터 내의 디자인된 돌연변이는 염색체 상에 반영되었고(도 13), 이것은 상기 돌연변이가 위치들 둘 다에 존재하고 플라스미드가 트랜스작용(transacting) 바코드(트랜스-바코드) 또는 게놈 에디팅부위의 기록으로서 사용된다는 것을 시사하였다.

[0112] Cas9 매개 절단으로부터 세포를 "면역화"시키되, 번역 생성물을 교란되지 않은 상태로 남기기 위해 동일한 의미의 PAM 돌연변이(도 14b, Δ PAM)로 구성된 "침묵 선택가능한 반흔(scars)"을 생성함으로써 게놈 규모로 단백질 코딩 프레임의 합리적 돌연변이유발을 위해 상기 디자인을 채택하였다. 본 발명자들은 침묵 흔적이 코돈에서의 인접 에디팅부위들 또는 관심 있는 다른 특징에 대한 동시선택을 고효율로 가능하게 할 수 있다고 추론하였다. PAM 돌연변이/결실과 *galK* 내의 원하는 돌연변이 사이의 상동성 아암(arm) 길이 및 거리의 효과를 평가하였고 효율을 비교하였다(도 16B). 상기 상동성 아암 길이가 동일한 PAM 에디팅부위를 가진 80개 뉴클레오티드들부터 100개 뉴클레오티드들(각각 약 5% 및 45%)까지 연장되었을 때 *galK* 위치 145에서의 돌연변이 효율의 상당한 증가가 관찰되었다.

[0113] 실시예 5: 돌연변이를 재구축하기 위한 GEN-TraCER 방법의 이용

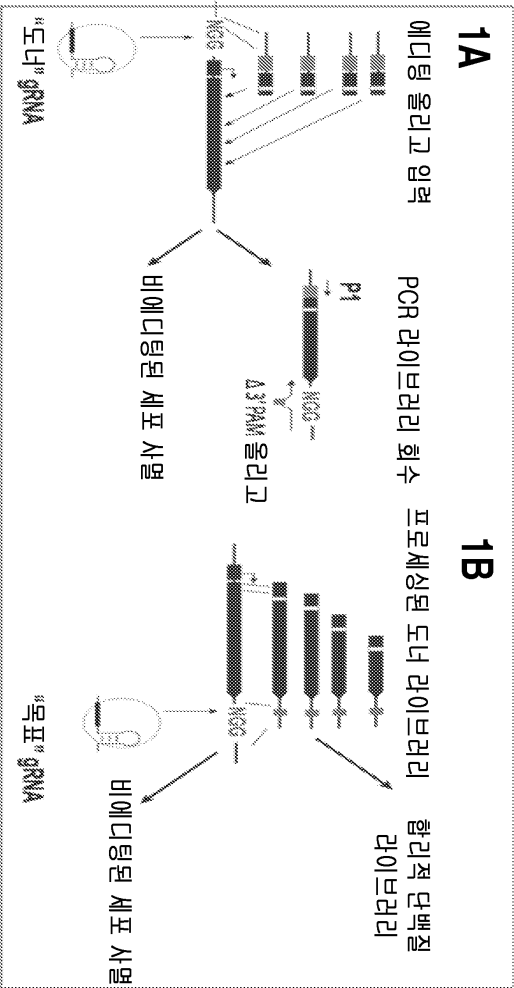
[0114] 단순한 사용자 입력 정의를 사용한 게놈 주변 부위의 표적화를 가능하게 하기 위해 맞춤 자동화된 디자인 소프트웨어를 이용하여 GEN-TraCER 방법을 게놈 규모까지 확장하였다. 이, 콜라이에서의 열적 적응의 최근 보고된 연구(Tenaillon et al. 2012)로부터 동일하지 않은 의미의 점 돌연변이들 전부를 구축함으로써 상기 방법을 시험하였다. 이 연구는 독립적으로 증식된 균주들로부터의 115개 단리물들에서 발생한 돌연변이들의 완전한 세트를 특징구명하였다. 이 데이터세트는 이 복잡한 표현형의 기작적 기반에 대한 추가 정보를 제공하는 개별 적합도(fitness) 효과를 가진 돌연변이들의 다양한 공급원을 제공한다. 다운스트림 적합도 분석에서 PAM 및 표적 코돈 돌연변이 둘 다에 대한 통계학적 보정을 가능하게 하기 위해 코돈 사용 및 Δ PAM(가능한 경우)에서 2배 중복으로 이 돌연변이들 각각을 재구축하였다.

[0115] 실시예 6: 유전적 상호작용을 조절하기 위한 GEN-TraCER 방법의 이용

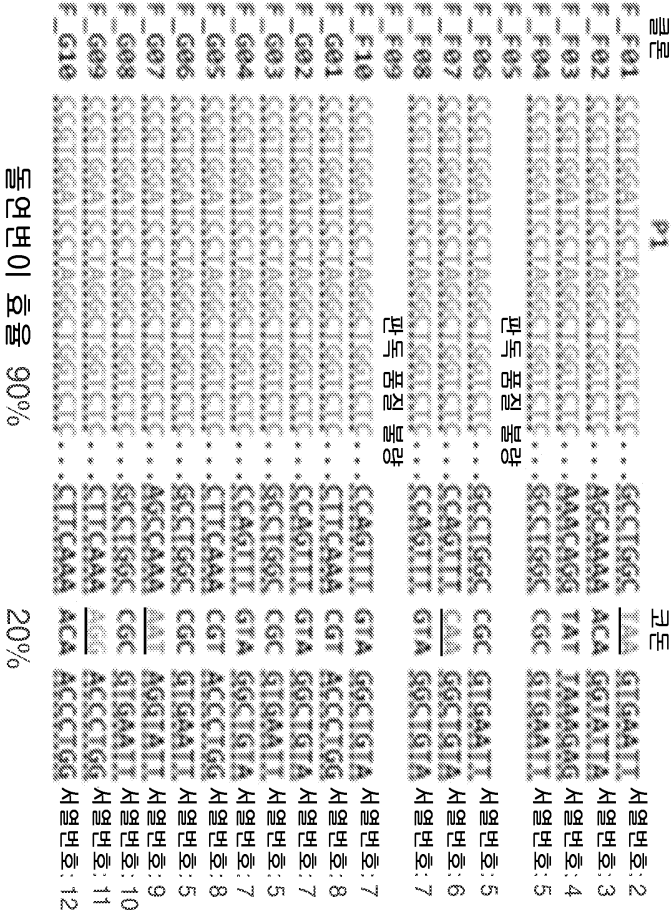
[0116] 이, 콜라이 게놈 내의 각각의 유전자의 상류에서 환경적인 신호(산소 수준, 탄소 공급원, 스트레스)에 의해 동적으로 조절되는 프로모터를 삽입함으로써 프로모터 리와이어링(rewiring) 라이브러리를 생성한다. GEN-TraCER 방법을 이용하여, 예를 들면, 생성할 관심 있는 화학물질에 대한 내성에 유리할 수 있는 리와이어링된(rewired) 유전형을 가진 균주를 생성한다.

도면

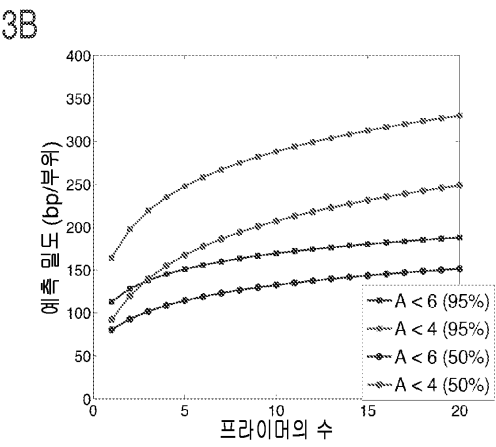
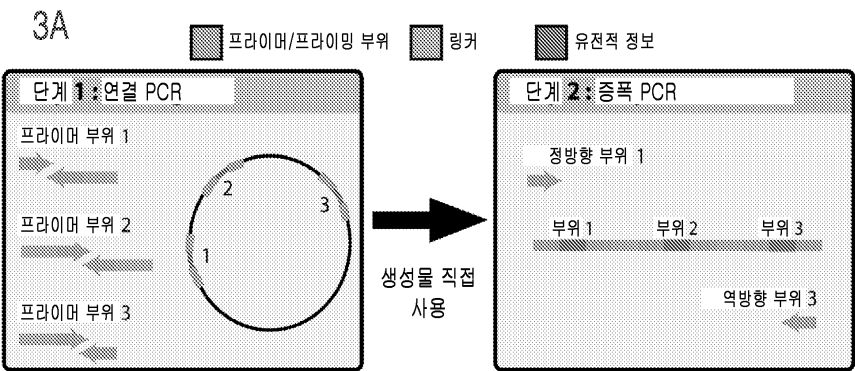
도면1



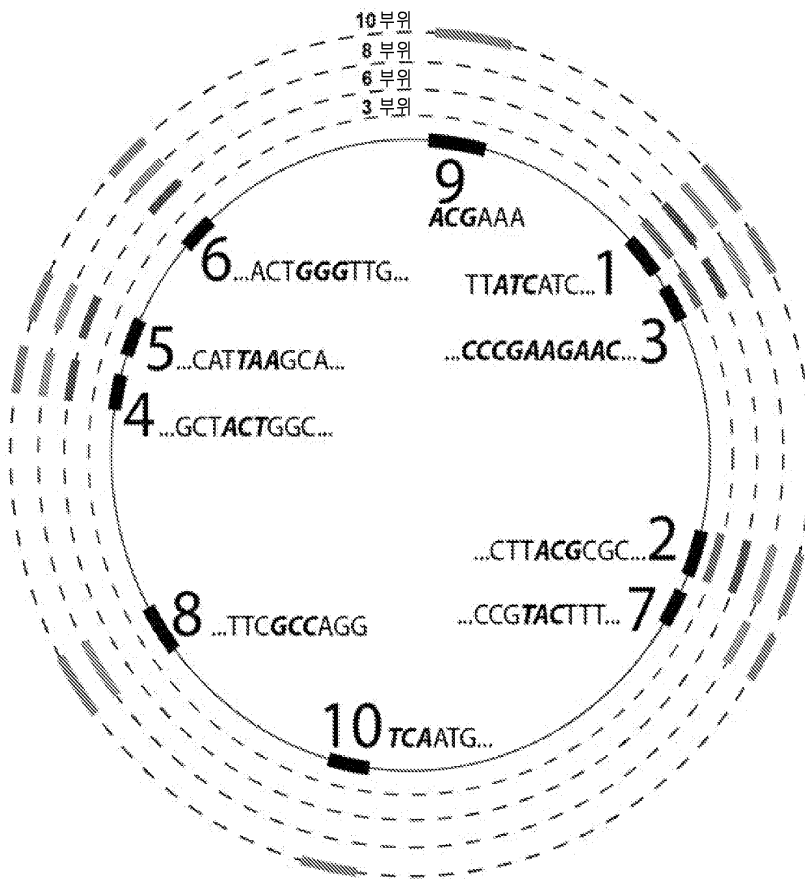
도면2



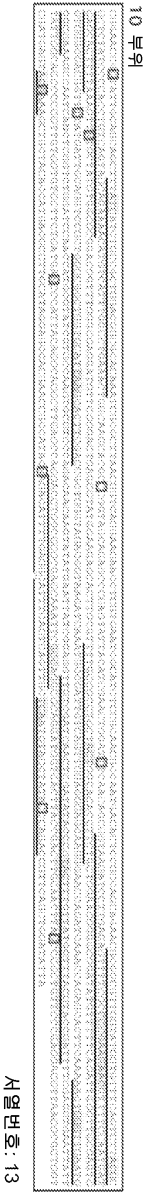
도면3



도면4a

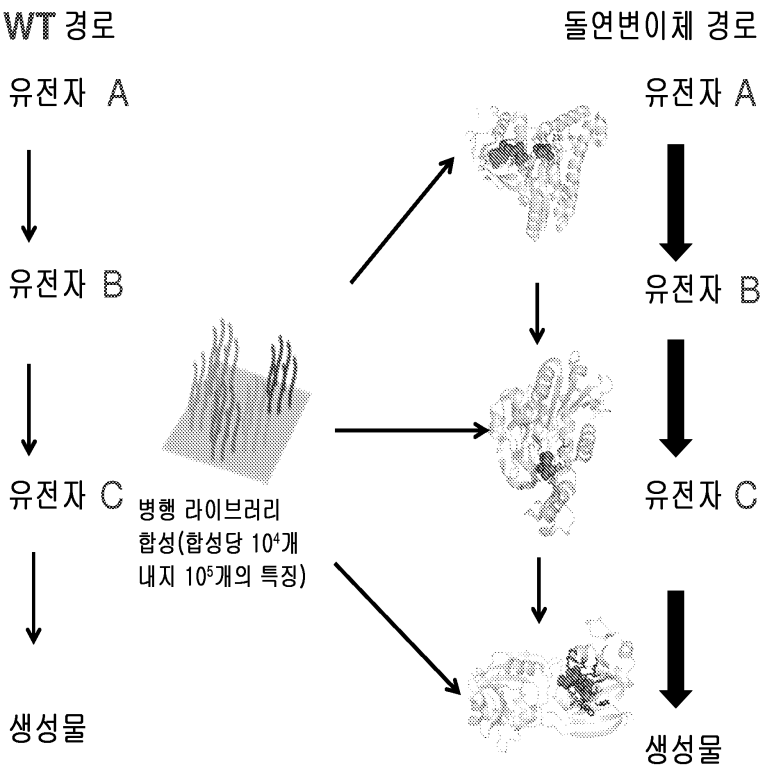


도면4b



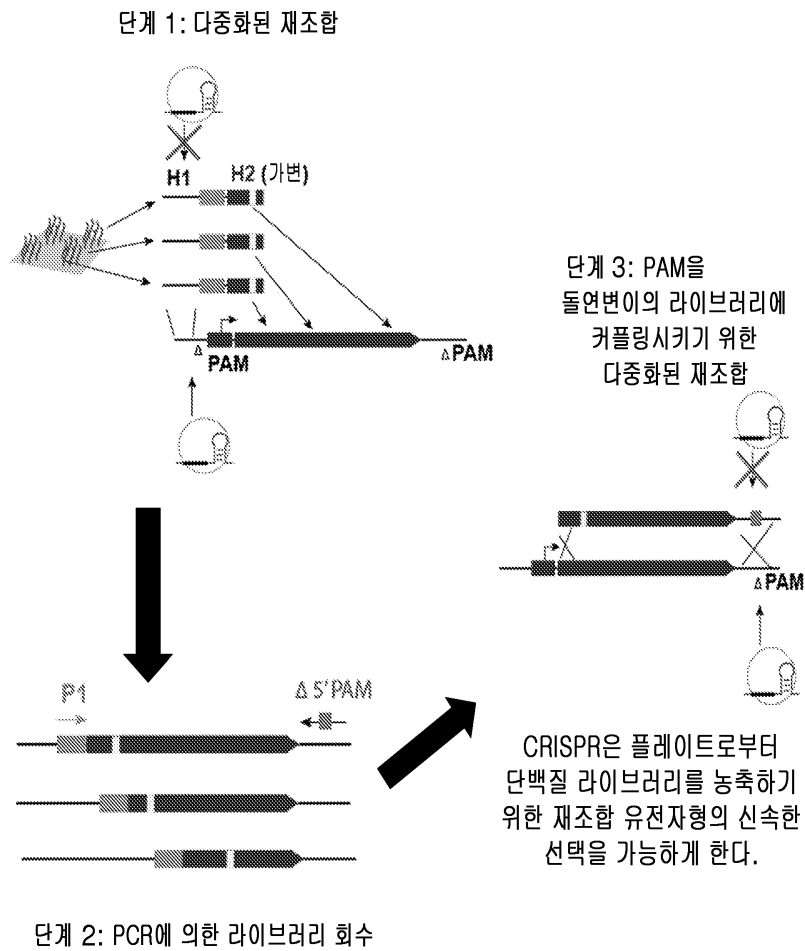
도면5

대사 조작을 위한 합리적 단백질 에디팅



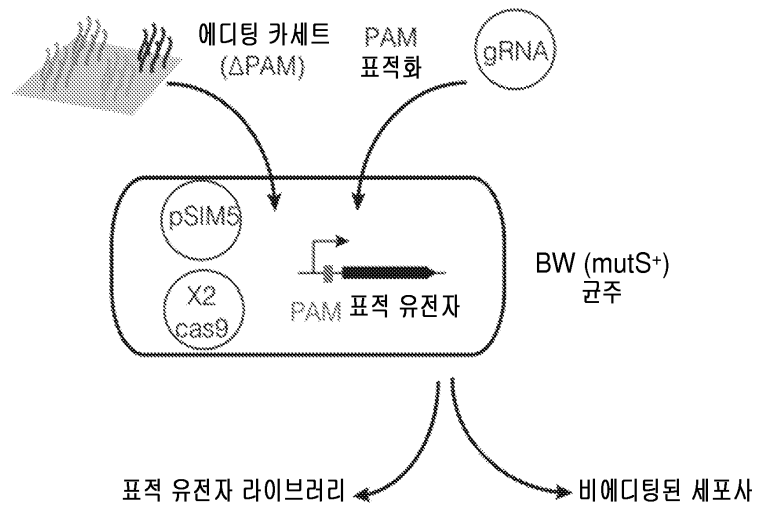
도면6

CRISPR 농축된 합리적 단백질 라이브러리

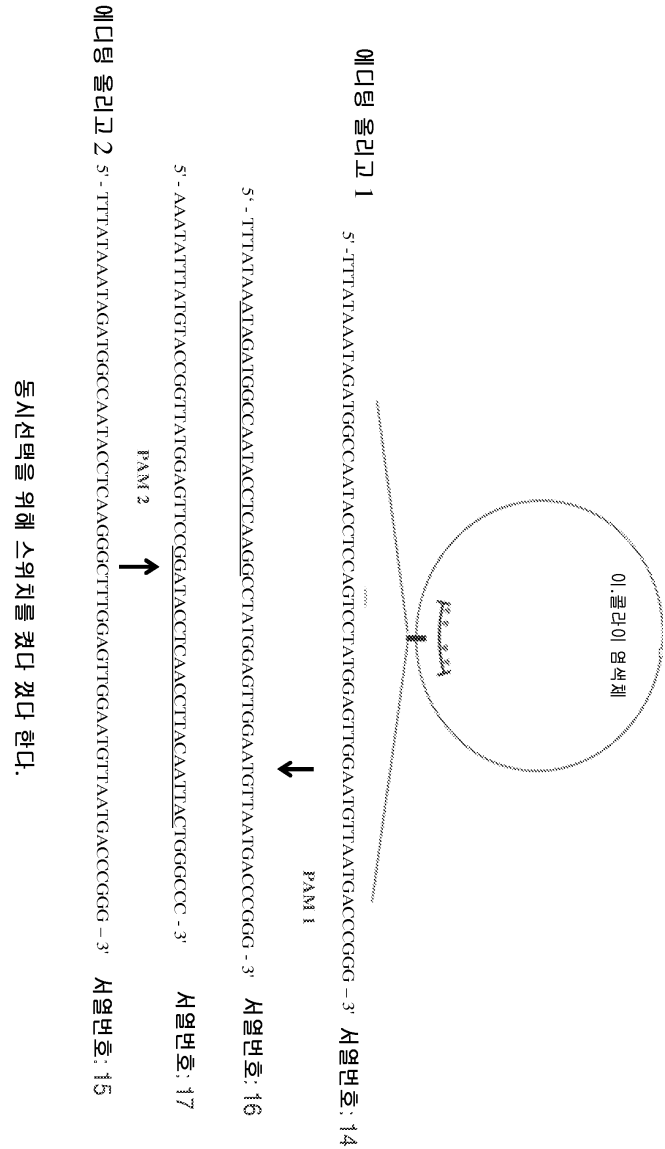


도면7

실험적 설정 및 개념 증명

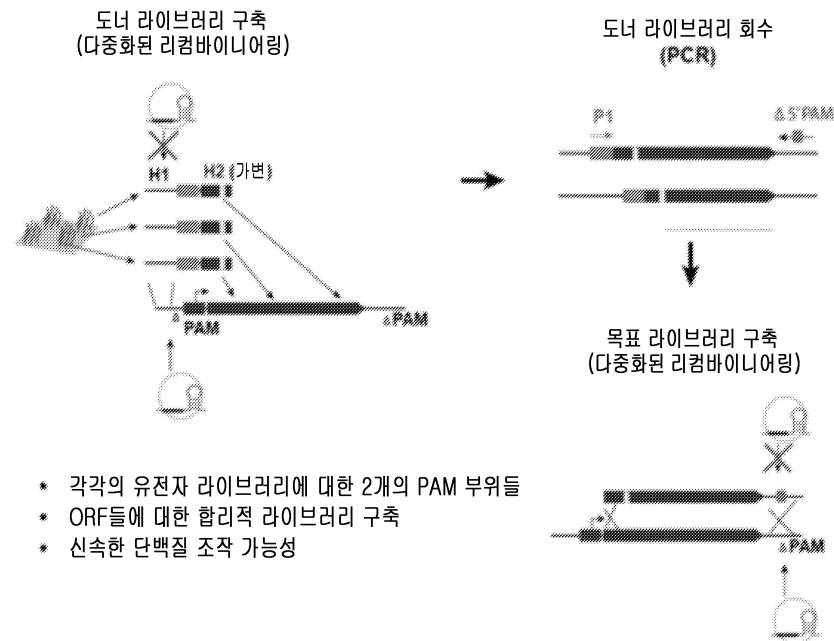


도면8



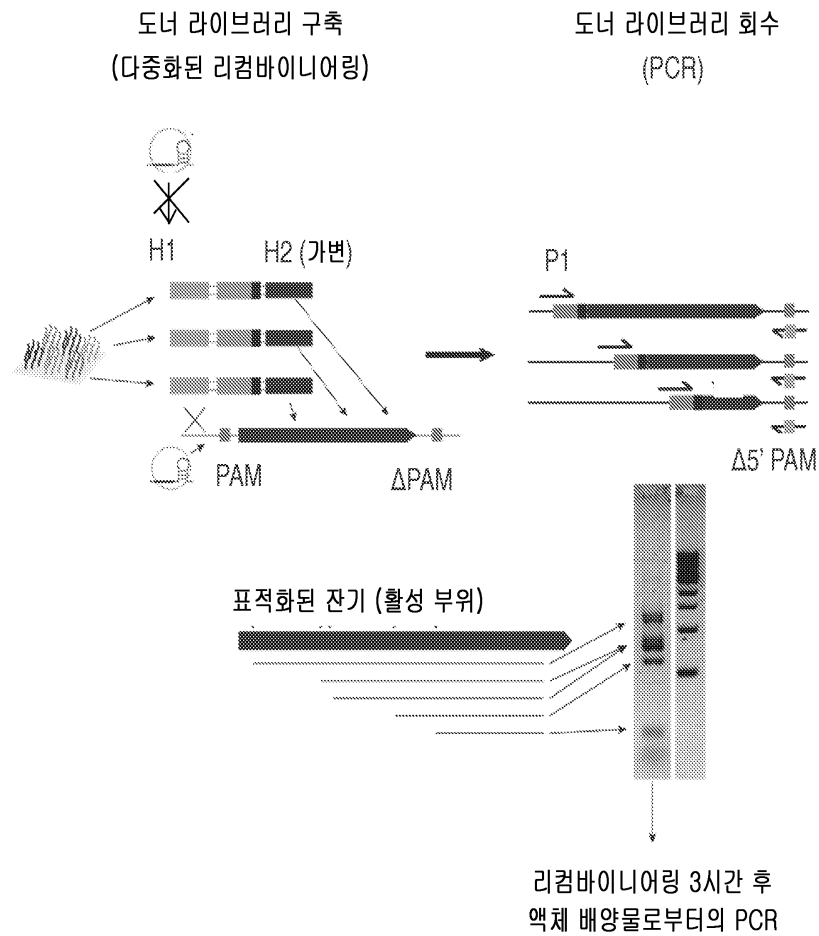
도면9

다중화된 단백질 조작 방법

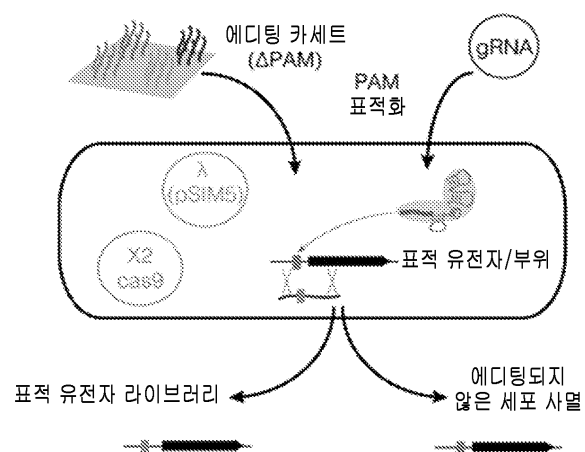


도면10

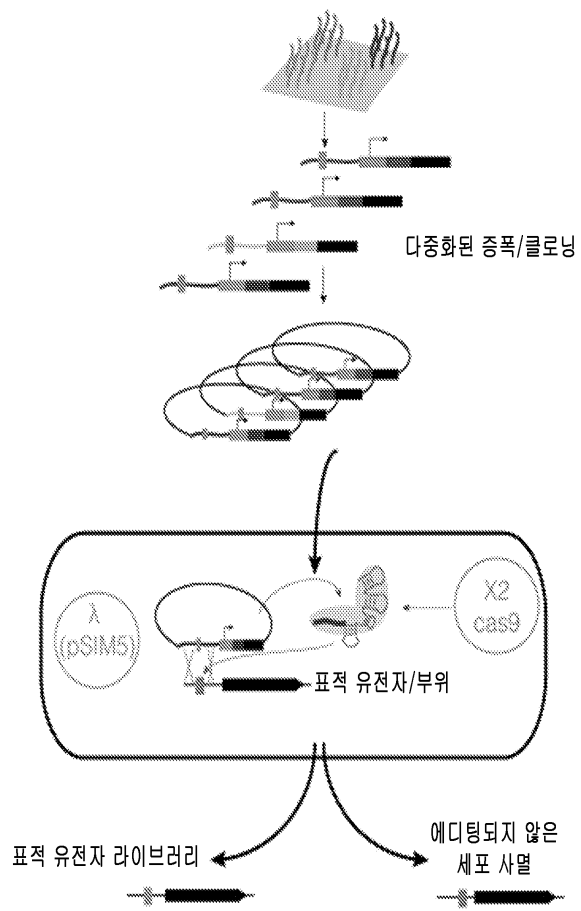
개념 증명 도너 라이브러리 구축 (galK)



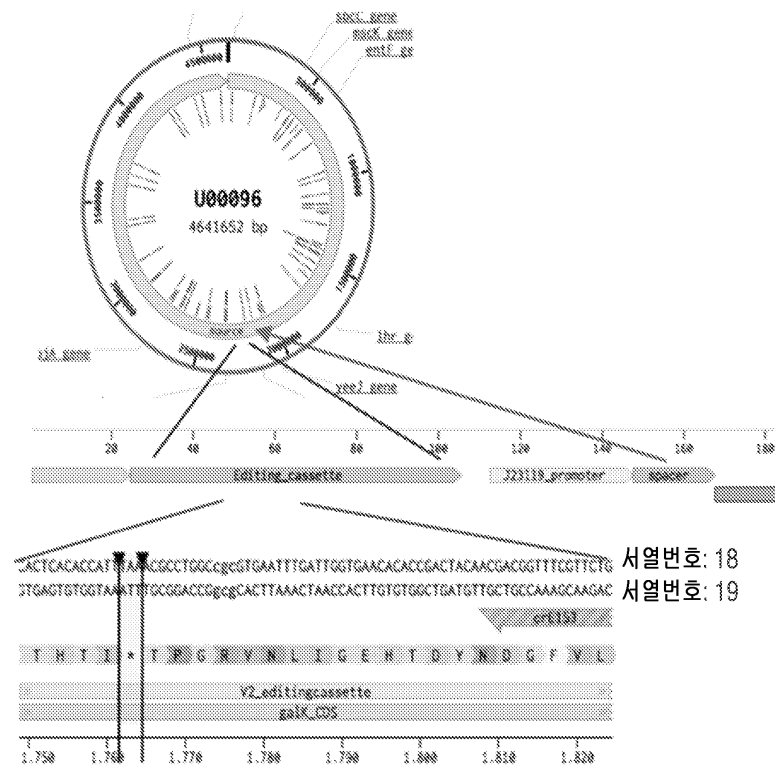
도면11a



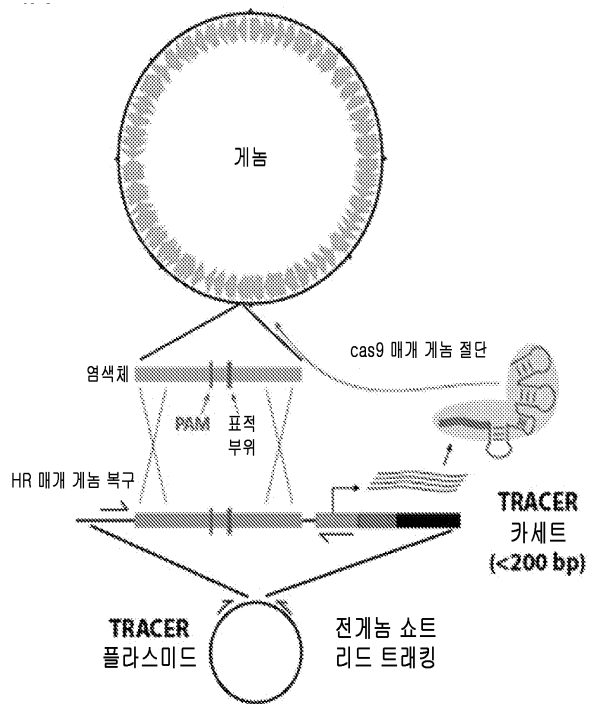
도면11b



도면 12



도면14a

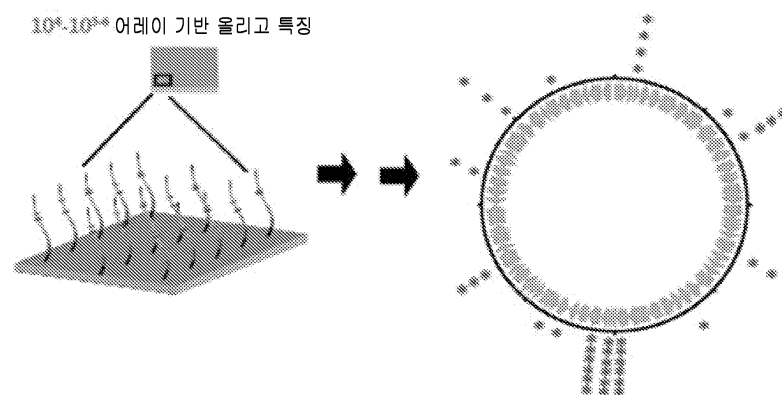


도면14b

ΔPAM	스페이서	표적 코돈	
GGA ACC GTA TTG CAG CAG CTT TAT CAT CTG CCG CTG GAC GGC GCA			서열번호: 23
G T V L Q Q L Y H L P L D G A			서열번호: 24

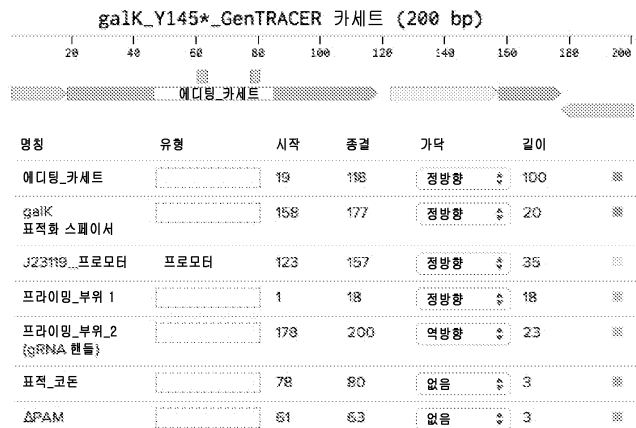
ΔPAM	스페이서	표적 코돈	
GGA ACC GTA TTG CAG CAG CTT TAA CAT CTG CCG CTG GAC GGC GCA			서열번호: 25
G T V L Q Q L * H L P L D G A			서열번호: 26
			서열번호: 27

도면14c

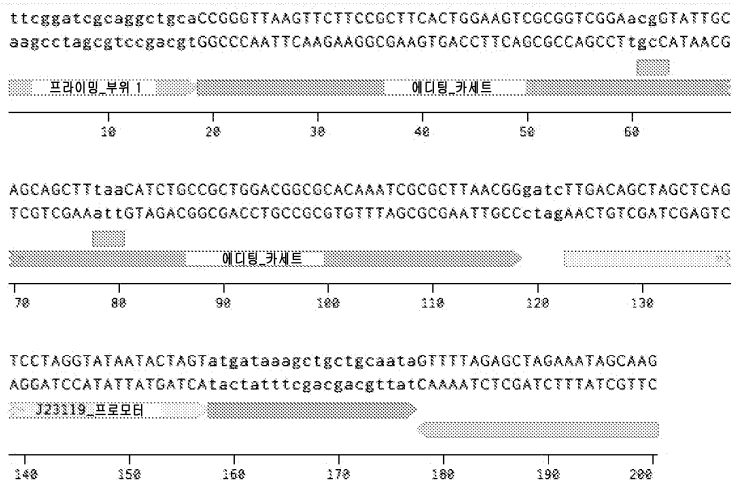


도면15

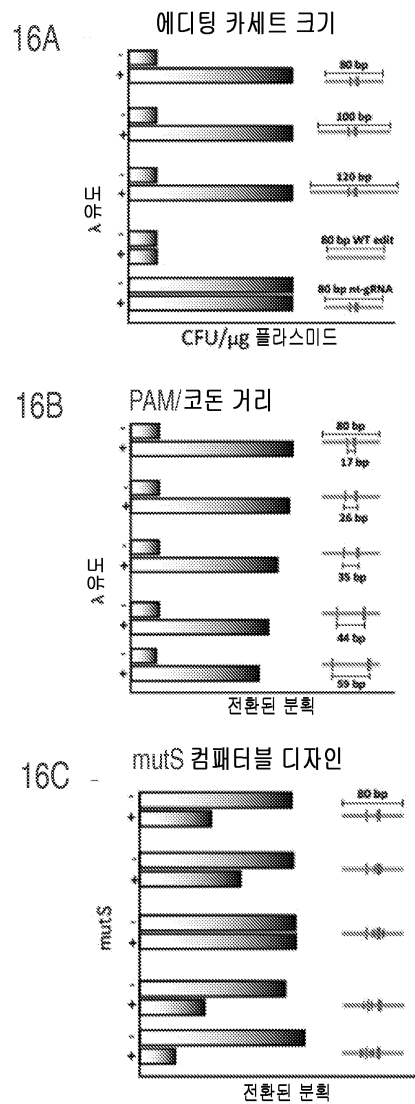
15A



15B



도면16



서열 목록

SEQUENCE LISTING

<110> The Regents of the University of Colorado, A Body Corporate

<120> CRISPR ENABLED MULTIPLEXED GENOME ENGINEERING

<130> C1102.70033W000

<140> Not Yet Assigned

<141> 2015-02-11

<150> 61/938,608

<151> 2014-02-11

<160> 33

<170> PatentIn version 3.5

<210> 1
 <211> 22
 <212> DNA
 <213> artificial sequence
 <220><223> synthetic polynucleotide
 <400> 1
 ccgtggatcc taggctggtc tc 22
 <210> 2
 <211> 17
 <212> DNA
 <213> artificial sequence
 <220><223> synthetic polynucleotide
 <400> 2
 gcctggctaa gtgaatt 17
 <210> 3
 <211> 17
 <212> DNA
 <213> artificial sequence
 <220><223> synthetic polynucleotide
 <400> 3
 agcaaaaaca ggtatta 17
 <210> 4
 <211> 17
 <212> DNA
 <213> artificial sequence
 <220><223> synthetic polynucleotide
 <400> 4
 aaacaggtat taaagag 17
 <210> 5
 <211> 17
 <212> DNA
 <213> artificial sequence
 <220><223> synthetic polynucleotide

<400> 5
gcctggccgc gtgaatt 17

<210> 6
<211> 17
<212> DNA
<213> artificial sequence
<220><223> synthetic polynucleotide
<400> 6
ccagtttcaa ggctgta 17

<210> 7
<211> 17
<212> DNA
<213> Artificial Sequence
<220><223> synthetic polynucleotide
<400> 7
ccagtttgta ggctgta 17

<210> 8
<211> 17
<212> DNA
<213> artificial sequence
<220><223> synthetic polynucleotide
<400> 8
cttcaaacgt accctgg 17

<210> 9
<211> 17
<212> DNA
<213> artificial sequence
<220><223> synthetic polynucleotide
<400> 9
agccaaaaat aggtatt 17

<210> 10
<211> 17
<212> DNA
<213> artificial sequence

<220><223> synthetic polynucleotide

<400> 10

gcctggccgc gtgaatt

17

<210> 11

<211> 17

<212> DNA

<213> artificial sequence

<220><223> synthetic polynucleotide

<400> 11

cttcaaaagg accctgg

17

<210> 12

<211> 17

<212> DNA

<213> artificial sequence

<220><223> synthetic polynucleotide

<400> 12

cttcaaaaca accctgg

17

<210> 13

<211> 1052

<212> DNA

<213> artificial sequence

<220><223> synthetic polynucleotide

<220><221> misc_feature

<222> (325)..(325)

<223> n is a, c, g, or t

<220><221> misc_feature

<222> (632)..(632)

<223> n is a, c, g, or t

<220><221> misc_feature

<222> (806)..(806)

<223> n is a, c, g, or t

<220><221> misc_feature

<222> (892)..(892)

<223> n is a, c, g, or t

<220><221> misc_feature

<222> (985)..(985)

<223> n is a, c, g, or t

<400> 13

cggaaccggt attgcagcag ctttatcatc tgccgctgga cggcgacaaa atcgcgctta 60

acggtcagga agaacctcgt gaatcgcatc tccgcaacgc caatgacact cgcgcagcag 120

aacgcggcgt tggatggtg tttcagtcctt acgcgctcta tccccacctg tcagtagcag 180

aaaacatgtc atttggcctg aaactggcaa gcacgcccgg tagcagagcc cgagtattac 240

atcgaactgg atttcaacag cggtaagatc cttgagagtt ttcgccccga agaacgtttt 300

ccaatgatga gcacttttaa agttntgcta tgtggcgcgg acggtttggg cgaggatggt 360

cgggactgag cgtatggaag agcacttgcg ttttgccgcc tgctactggc acaccttctg 420

ctggaacggg gcggatatgt ttttcagatt gggttgcgag tgggtggcga cggttttatg 480

attgcaggtc cgtcgggcgg ctgggtgcatt aagcacttcg accgctgggt agacggtaag 540

atcaaatgcc gattgtcagt tggagggagc aagggaacca gatcaaacca gagcacttca 600

aaaactgggt tgaatgggag aaagccaatc anctcggctt ggatccaacc atgcgtatcg 660

tgccaagtgc gtcccagaca acaagtatat gaattatcgg aggttgctga tcaactcgat 720

ataccgtac tttgttatgg tttacgtacc gattttcgag gtgaattatt tattggcagc 780

ccccgttgtg gtgtccttct gatganccca ttacaggcac gcttgagcca ggacctggcg 840

cgcgagcaaa ttcgccaggc gcaggatggt cacttaccga cttgcaactt ancccgtgt 900

tcccaggac gttaacgcgc tggtcgatga gtacaaaagc tgctacacca tgacgccttg 960

cataggagcc acgaaccgcc atacnagaca ttttgaggca tttcagtcag ttgctcaatg 1020

tacctatacc cagaccgttc agctggatat ta 1052

<210> 14

<211> 59

<212> DNA

<213> artificial sequence

<220><223> synthetic polynucleotide

<400> 14

tttataaata gatggccaat acctccagtc ctatggagtt ggaatgttaa tgaccggg 59

<210> 15

<211> 59

<212> DNA
 <213> artificial sequence
 <220><223> synthetic polynucleotide
 <400> 15
 ttataaata gatggccaat acctcaaggg ctttgagtt ggaatgttaa tgaccggg 59

<210> 16
 <211> 59
 <212> DNA
 <213> artificial sequence
 <220><223> synthetic polynucleotide
 <400> 16
 ttataaata gatggccaat acctcaaggc ctatggagtt ggaatgttaa tgaccggg 59

<210> 17
 <211> 59
 <212> DNA
 <213> artificial sequence
 <220><223> synthetic polynucleotide
 <400> 17
 aaatatttat gtaccggta tggagttccg gataacctcaa cttacaatt actgggcc 59

<210> 18
 <211> 76
 <212> DNA
 <213> artificial sequence
 <220><223> synthetic polynucleotide
 <400> 18
 cactcacacc atttaaagc ctggccgct gaatttgatt ggtgaacaca ccgactaca 60

cgacggtttc gttctg 76

<210> 19
 <211> 76
 <212> DNA
 <213> artificial sequence
 <220><223> synthetic polynucleotide
 <400> 19

cagaacgaaa ccgtcgttgt agtcggtgtg ttcaccaatc aaattcacgc ggccaggcgt 60
 ttaaattggtg tgagtg 76
 <210> 20
 <211> 24
 <212> DNA
 <213> artificial sequence
 <220><223> synthetic polynucleotide
 <400> 20
 cacaccattt aaacgcctgg ccgc 24
 <210> 21
 <211> 24
 <212> DNA
 <213> artificial sequence
 <220><223> synthetic polynucleotide
 <400> 21
 cacaccattt aagcgcctgg ccgc 24
 <210> 22
 <211> 24
 <212> DNA
 <213> artificial sequence
 <220><223> synthetic polynucleotide
 <400> 22
 cacaccattc aggcgcctgg ccgc 24
 <210> 23
 <211> 45
 <212> DNA
 <213> artificial sequence
 <220><223> synthetic polynucleotide
 <400> 23
 ggaaccgtat tgcagcagct ttatcatctg ccgctggacg gcgca 45
 <210> 24
 <211> 15
 <212> PRT

<213> artificial sequence

<220><223> synthetic polypeptide

<400> 24

Gly Thr Val Leu Gln Gln Leu Tyr His Leu Pro Leu Asp Gly Ala

1 5 10 15

<210> 25

<211> 45

<212> DNA

<213> artificial sequence

<220><223> synthetic polynucleotide

<400> 25

ggaacggtat tgcagcagct ttaacatctg ccgctggacg gcgca 45

<210> 26

<211> 7

<212> PRT

<213> artificial sequence

<220><223> synthetic polypeptide

<400> 26

Gly Thr Val Leu Gln Gln Leu

1 5

<210> 27

<211> 7

<212> PRT

<213> artificial sequence

<220><223> synthetic polypeptide

<400> 27

His Leu Pro Leu Asp Gly Ala

1 5

<210> 28

<211> 200

<212> DNA

<213> artificial sequence

<220><223> synthetic polynucleotide

<400> 28

ttcggatcgc aggcctgcacc gggttaagtt cttccgcttc actggaagtc gcggtcggaa 60
 cggtattgca gcagctttta catctgccgc tggacggcgc acaaatcgcg cttaacggga 120

tcttgacagc tagctcagtc ctaggtataa tactagtatg ataaagctgc tgcaatagtt 180
 ttagagctag aaatagcaag 200

<210> 29
 <211> 31
 <212> DNA
 <213> artificial sequence
 <220><223> synthetic polynucleotide
 <400> 29
 ccagaaatca tccttagcga aagctaagga t 31
 <210> 30
 <211> 36
 <212> DNA
 <213> artificial sequence
 <220><223> synthetic polynucleotide
 <400> 30
 gttttagagc tagaaatagc aagttaaaat aaggct 36

<210> 31
 <211> 60
 <212> DNA
 <213> artificial sequence
 <220><223> synthetic polynucleotide
 <400> 31
 atcacgaggc agaatttcag ataaaaaaaa tccttagctt tcgctaagga tgatttctgg 60
 <210> 32
 <211> 60
 <212> DNA
 <213> artificial sequence
 <220><223> synthetic polynucleotide
 <400> 32
 actttttcaa gttgataacg gactagcctt attttaactt gctatttcta gctctaaaac 60
 <210> 33

<211> 200

<212> DNA

<213> artificial sequence

<220><223> synthetic polynucleotide

<400> 33

cttgctatTT ctagctctaa aactattgca gcagctttat catactagta ttatacctag 60

gactgagcta gctgtcaaga tcccgTTaag cgcgatttgt gcgccgtcca gcggcagatg 120

ttaaagctgc tgcaataaccg ttccgaccgc gacttccagt gaagcgggaag aacttaaccc 180

ggtgcagcct gcgatccgaa 200