



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2016-0102056
(43) 공개일자 2016년08월26일

(51) 국제특허분류(Int. Cl.)
C12N 15/113 (2010.01) C12N 9/22 (2006.01)
(52) CPC특허분류
C12N 15/113 (2013.01)
C12N 9/22 (2013.01)
(21) 출원번호 10-2016-7020111
(22) 출원일자(국제) 2014년09월18일
심사청구일자 없음
(85) 번역문제출일자 2016년07월22일
(86) 국제출원번호 PCT/US2014/056416
(87) 국제공개번호 WO 2015/099850
국제공개일자 2015년07월02일
(30) 우선권주장
61/921,007 2013년12월26일 미국(US)
(뒷면에 계속)

(71) 출원인
더 제너럴 하스피탈 코퍼레이션
미국, 메사추세츠 02114, 보스턴 프룻트 스트리트 55
(72) 발명자
차이, 쉐다
미국 02129 매사추세츠주 찰스타운 프로스펙트 스트리트 32 아파트먼트 1
정, 케이트 제이.
미국 01890 매사추세츠주 윈체스터 매그놀리아 웨이 1
(74) 대리인
양영준, 김영

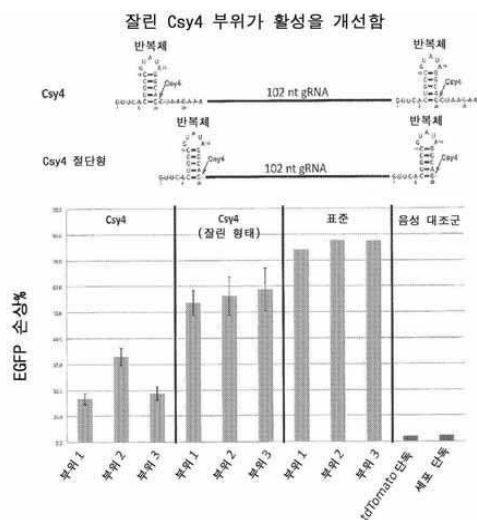
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 멀티플렉스 가이드 RNA

(57) 요약

선택적으로 포유류 세포에서, RNA 폴리머라제 II 및 III 프로모터로부터 고효성 CRISPR 가이드 RNA(gRNA)의 멀티플렉스 발현을 위한 방법 및 구축물. 본 발명은 적어도 부분적으로, 짧은 RNA 헤어핀 서열을 인식하는 엔도리보뉴클레아제인 Csy4를 이용하여 개별 gRNA가 Csy4 절단 부위에 의해 분리되는 하나의 더 긴 RNA 전사체(RNA pol II 또는 III 프로모터로부터 생성됨) 상에서 인코딩된 다중 기능적 gRNA를 절단할 수 있다는 발견에 기반한다.

대표도 - 도4



(52) CPC특허분류

C12N 2310/10 (2013.01)
C12N 2310/3519 (2013.01)
C12N 2310/51 (2013.01)

(30) 우선권주장

61/930,782 2014년01월23일 미국(US)
14/211,117 2014년03월14일 미국(US)
PCT/US2014/029068 2014년03월14일 미국(US)
PCT/2014/029304 2014년03월14일 미국(US)
PCT/2014/028630 2014년03월14일 미국(US)
PCT/US2014/035162 2014년04월23일 미국(US)

명세서

청구범위

청구항 1

가이드 RNA(gRNA)를 인코딩하는 복수의 서열을 포함하는 데옥시리보핵산(DNA) 분자로서, 각각의 gRNA에 서열 GTTCACTGCCGTATAGGCAG(SEQ ID NO:1) 또는 GTTCACTGCCGTATAGGCAGCTAAGAAA(SEQ ID NO:2)를 포함하거나 이로 구성되는 적어도 하나의 Csy4 절단 서열이 인접하는 DNA 분자.

청구항 2

제1항에 있어서, 프로모터 서열에 작동 가능하게 연결되는 DNA 분자.

청구항 3

제1항에 있어서, 2 개, 3 개, 또는 그 초과 gRNA 서열을 포함하는 DNA 분자.

청구항 4

제1항에 있어서, 프로모터 서열이 RNA 폴리머라제 II(Pol II) 프로모터 또는 Pol III 프로모터인 DNA 분자.

청구항 5

제1항에 있어서, 프로모터 서열이 RNA Pol II 프로모터인 DNA 분자.

청구항 6

제5항에 있어서, Pol II 프로모터가 CAG, EF1A, CAGGS, PGK, UbiC, CMV, B29, 테스민(Desmin), 엔도글린(Endoglin), FLT-1, GFPa 및 SYN1 프로모터로 구성된 군으로부터 선택되는 DNA 분자.

청구항 7

Csy4 절단 부위에 연결된 가이드 RNA(gRNA)를 인코딩하는 서열을 포함하는 1 개, 2 개, 3 개 또는 그 초과 카세트와 연결된 프로모터 서열을 포함하는 DNA 분자.

청구항 8

제7항에 있어서, Csy4 절단 부위에 연결된 제3 가이드 RNA를 인코딩하는 제3 서열에 연결되는, Csy4 절단 부위에 연결된 제2 가이드 RNA를 인코딩하는 제2 서열에 연결되는, Csy4 절단 부위에 연결된 제1 가이드 RNA를 인코딩하는 제1 서열에 작동 가능하게 연결되는 Pol II 프로모터를 포함하는 DNA 분자.

청구항 9

제7항에 있어서, 약 100 개 nt의 가이드 RNA 서열을 포함하는 DNA 분자.

청구항 10

제7항에 있어서, SEQ ID NO:1 또는 2의 Csy4 절단 부위를 포함하는 DNA 분자.

청구항 11

제1항 또는 제7항에 있어서, 기능적 Csy4 효소를 인코딩하는 서열을 추가로 포함하는 DNA 분자.

청구항 12

제1항 또는 제7항에 있어서, gRNA가 하기 서열:

(X₁₇₋₂₀)GUUUUAGAGCUAUGCUGUUUUG(X_N)(SEQ ID NO:4);

(X₁₇₋₂₀)GUUUUAGAGCUA(SEQ ID NO:5);

(X₁₇₋₂₀)GUUUUAGAGCUAUGCUGUUUUG(SEQ ID NO:6);

(X₁₇₋₂₀)GUUUUAGAGCUAUGCU(SEQ ID NO:7);

(X₁₇₋₂₀)GUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCG(X_N)(SEQ ID NO:8);

(X₁₇₋₂₀)GUUUUAGAGCUAUGCUGAAAAGCAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUC(X_N)(SEQ ID NO:9);

(X₁₇₋₂₀)GUUUUAGAGCUAUGCUGUUUUGGAAACAAAACAGCAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUC(X_N)(SEQ ID NO:10);

(X₁₇₋₂₀)GUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC(X_N)(SEQ ID NO:11);

(X₁₇₋₂₀)GUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC(SEQ ID NO:12);

(X₁₇₋₂₀)GUUUUAGAGCUAUGCUGGAAACAGCAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC(SEQ ID NO:13); 또는

(X₁₇₋₂₀)GUUUUAGAGCUAUGCUGGAAACAGCAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC(SEQ ID NO:14)

를 포함하며, 여기서 X₁₇₋₂₀은 표적 서열, 바람직하게는 프로토스페이서 인접 모티프(PAM)의 바로 5'의 표적 서열의 17 개 내지 20 개 연속 뉴클레오타이드의 상보적 가닥에 상보적인 서열이며, X_N은 Cas9에 대한 리보핵산의 결합을 방해하지 않는 임의의 서열인 DNA 분자.

청구항 13

제1항 또는 제7항에 있어서, 분자의 3' 말단에 하나 이상의 U를 포함하는 DNA 분자.

청구항 14

제1항 내지 제13항 중 어느 한 항의 DNA 분자에 의해 인코딩되는 RNA 분자.

청구항 15

제1항 내지 제13항 중 어느 한 항의 DNA 분자를 포함하는 벡터.

청구항 16

제15항에 있어서, 기능적 Csy4 효소를 인코딩하는 서열을 추가로 포함하는 벡터.

청구항 17

세포에서 복수의 가이드 RNA의 제조 방법으로서, 세포에서 제1항 내지 제13항 중 어느 한 항의 DNA 분자를 발현시키는 단계 또는 제14항의 RNA 분자와 세포를 접촉시키는 단계를 포함하는 방법.

청구항 18

제17항에 있어서, 세포가 포유류 세포이고, 세포가 외인성 기능적 Csy4 효소 서열을 또한 발현시키거나, 또는 방법이 기능적 Csy4 효소 또는 기능적 Csy4 효소를 인코딩하는 핵산을 투여하는 단계를 추가로 포함하는 방법.

청구항 19

세포에서 복수의 표적 유전자의 발현을 변형하는 방법으로서, 방법이 가이드 RNA(gRNA)를 인코딩하는 복수의 서열을 포함하는 DNA 분자를 발현시키는 단계를 포함하며, 각각의 gRNA는 표적 유전자의 적어도 17 개 nt 내지 20 개 nt에 상보적인 서열을 포함하고, 각각의 gRNA에는 서열 GTTCACTGCCGTATAGGCAG(SEQ ID NO:1) 또는 GTTCACTGCCGTATAGGCAGCTAAGAAA(SEQ ID NO:2)를 포함하거나 이로 구성되는 적어도 하나의 Csy4 절단 서열이 인접하는 방법.

청구항 20

세포에서 표적 유전자의 발현을 변형하는 방법으로서, 방법이 가이드 RNA(gRNA)를 인코딩하는 복수의 서열을 포함하는 DNA 분자를 발현시키는 단계를 포함하며, 각각의 gRNA는 표적 유전자의 적어도 17 개 nt 내지 20 개 nt에 상보적인 서열을 포함하고, 각각의 gRNA에는 서열 GTTCACTGCCGTATAGGCAG(SEQ ID NO:1) 또는 GTTCACTGCCGTATAGGCAGCTAAGAAA(SEQ ID NO:2)를 포함하거나 이로 구성되는 적어도 하나의 Csy4 절단 서열이 인접하는 방법.

발명의 설명**기술 분야**

[0001] 관련 출원에 대한 교차 참조

[0002] 본 출원은 2013년 12월 26일에 출원된 특허 가출원 61/921,007; 2014년 3월 14일에 출원된 14/211,117; 2014년 3월 14일에 출원된 PCT/US2014/029068; 2014년 3월 14일에 출원된 PCT/US2014/028630; 2014년 3월 23일에 출원된 PCT/US2014/035162; 및 2014년 3월 14일에 출원된 PCT/US2014/029304의 이익을 청구한다. 상기 모두가 본원에 참조로 포함된다.

[0003] 연방 지원 연구 또는 개발

[0004] 본 발명은 미국 국립 의료원이 수여하는 연구비 번호 DP1 GM105378 하의 정부 지원으로 수행되었다. 정부는 본 발명에 있어서 일정 권리를 갖는다.

[0005] 기술 분야

[0006] 선택적으로 포유류 세포에서, RNA 폴리머라제 II 및 III 프로모터로부터 고효율성 CRISPR 가이드 RNA(gRNA)의 멀티플렉스 발현을 위한 방법 및 구축물이 기재된다.

배경 기술

[0007] Cas9 뉴클레아제는 시험관내, 포유류 세포에서, 그리고 생체 모델 유기체, 예컨대 제브라다니오에서(Wang et al., Cell 153, 910-918(2013); Shen et al., Cell Res(2013); Dicarlo et al., Nucleic Acids Res(2013); Jiang et al., Nat Biotechnol 31, 233-239(2013); Jinek et al., Elife 2, e00471(2013); Hwang et al., Nat Biotechnol 31, 227-229(2013); Cong et al., Science 339, 819-823(2013); Mali et al., Science 339, 823-826(2013c); Cho et al., Nat Biotechnol 31, 230-232(2013); Gratz et al., Genetics 194(4):1029-35(2013)) 표적 DNA 서열에 부위-특이적 절단을 생성하기 위해 이용될 수 있는 프로그램 가능한 RNA로 가이드되고, 클러스터링되고, 규칙적으로 배치되는, 짧은 앞뒤상동 반복체(CRISPR)/CRISPR-연관된(Cas) 시스템의 기초를 형성한다(Wiedenheft et al., Nature 482, 331-338(2012); Horvath et al., Science 327, 167-170(2010); Terns et al., Curr Opin Microbiol 14, 321-327(2011)). 짧은 약 100 개 nt의 가이드 RNA(gRNA)는 Cas9와 복합체를 형성하며, 뉴클레아제를 특정한 표적 DNA 부위로 보낸다; 표적화는 gRNA의 5' 말단에서 적어도 17 개 내지 20 개 뉴클레오타이드(nt) 서열에 의해 매개되며, 이는 조작된 gRNA의 최초 17 개 내지 20 개 뉴클레오타이드 및 프로토스페이서 인접 모티프(PAM), 예로 서열 NGG 또는 NAG에 매칭하는 PAM 다음에 놓인 관심 표적 게놈 DNA 서열의 상보성 가닥 간 단순한 염기쌍 상보성을 통해 상호작용하고 상보적이도록 설계된다(Shen et al., Cell Res(2013); Dicarlo et al., Nucleic Acids Res(2013); Jiang et al., Nat Biotechnol 31, 233-239(2013); Jinek et al., Elife 2, e00471(2013); Hwang et al., Nat Biotechnol 31, 227-229(2013); Cong et al., Science 339, 819-823(2013); Mali et al., Science 339, 823-826(2013c); Cho et al., Nat Biotechnol 31, 230-232(2013); Jinek et al., Science 337, 816-821(2012)). gRNA는 다시 효과기 도메인(예로, 전사 활성화 도메인)에 융합되는 촉매적으로 불활성화된 Cas9 단백질(dCas9로 알려져 있음, 문헌[Jinek et al., Science 337:816-821(2012)]참고)을 보낼 수 있다, 예로 둘 다 본원에 참조로 포함되는 2013년 3월 15일에 출원된 USSN 61/799,647 및 2013년 6월 21일에 출원된 61/838,148 참고. 이들 후자 시스템은 관심 게놈 유전자위로 이중성 효과기 도메인의 RNA-가이드된 모집을 가능케 한다.

발명의 내용

[0008] 요약

- [0009] 본 발명은 적어도 부분적으로, 짧은 RNA 헤어핀 서열을 인식하는 엔도리보뉴클레아제인 Csy4를 이용하여 개별 gRNA가 Csy4 절단 부위에 의해 분리되는 하나의 더 긴 RNA 전사체(RNA pol II 또는 III 프로모터로부터 생성됨) 상에서 인코딩된 다중 기능적 gRNA를 절단할 수 있다는 발견에 기반한다.
- [0010] 따라서 제1 양태에서, 본 발명은 가이드 RNA(gRNA)를 인코딩하는 복수의 서열을 포함하는 데옥시리보핵산(DNA) 분자를 제공하며, 여기서 각각의 gRNA에는 서열 GTTCACTGCCGTATAGGCAG(SEQ ID NO:1) 또는 GTTCACTGCCGTATAGGCAGTAAGAAA(SEQ ID NO:2)를 포함하거나 이로 구성되는 적어도 하나의 Csy4 절단 서열이 인접한다.
- [0011] 일부 구현예에서, DNA 분자는 프로모터 서열에 작동 가능하게 연결된다.
- [0012] 일부 구현예에서, DNA 분자에는 2 개, 3 개 또는 그 초과와 gRNA 서열이 포함되며, 각각에는 적어도 하나의 Csy4 절단 서열이 인접한다.
- [0013] 일부 구현예에서, 프로모터 서열은 RNA 폴리머라제 II(Pol II) 프로모터 또는 Pol III 프로모터, 바람직하게는 RNA Pol II 프로모터이다. 일부 구현예에서, Pol II 프로모터는 CAG, EF1A, CAGGS, PGK, UbiC, CMV, B29, 데스민(Desmin), 엔도글린(Endoglin), FLT-1, GFPa 및 SYN1 프로모터로 구성된 군으로부터 선택된다.
- [0014] 다른 양태에서, 본 발명은 Csy4 절단 부위, 예로 SEQ ID NO:1 또는 2에 연결된 가이드 RNA를 인코딩하는 서열, 즉 약 100 개 nt, 예로 95 개 내지 300 개 nt, 예로 *S. 피오제네스*-기반 시스템에 있어서 95 개 내지 105 개 nt의 서열을 포함하는 1 개, 2 개, 3 개 또는 그 초과와 카세트에 연결된 프로모터 서열을 포함하는 DNA 분자를 제공한다.
- [0015] 일부 구현예에서, DNA 분자는 Csy4 절단 부위에 연결된 제3 가이드 RNA를 인코딩하는 제3 서열에 연결된, Csy4 절단 부위에 연결된 제2 가이드 RNA를 인코딩하는 제2 서열에 연결된, Csy4 절단 부위에 연결된 제1 가이드 RNA를 인코딩하는 제1 서열에 작동 가능하게 연결된 Pol II 프로모터를 포함한다. 일부 구현예에서, Csy4 절단 부위에 연결된 추가 가이드 RNA가 포함된다. 예를 들어, DNA 분자는 하기 구조:
- [0016] 프로모터 - C4 - gRNA - C4 - gRNA - C4 - gRNA - C4
- [0017] 프로모터 - C4 - gRNA - C4 - gRNA - C4 - gRNA - C4 - gRNA - C4
- [0018] 프로모터 - C4 - gRNA - C4 - gRNA - C4 - gRNA - C4 - gRNA - C4 - gRNA C4 등을 가질 수 있다. 상기 예시에 서, C4는 Csy4 RNA 절단 부위를 인코딩하는 서열이고, gRNA는 가이드 RNA를 인코딩하는 서열이다.
- [0019] 일부 구현예에서, Cas9 sgRNA는 하기 서열:
- [0020] (X₁₇₋₂₀)GUUUUAGAGCUAUGCUGUUUUG(X_N)(SEQ ID NO:4);
- [0021] (X₁₇₋₂₀)GUUUUAGAGCUA(SEQ ID NO:5);
- [0022] (X₁₇₋₂₀)GUUUUAGAGCUAUGCUGUUUUG(SEQ ID NO:6);
- [0023] (X₁₇₋₂₀)GUUUUAGAGCUAUGC(SEQ ID NO:7);
- [0024] (X₁₇₋₂₀)GUUUUAGAGCUAGAAUAGCAAGUUAAAAUAGGCUAGUCCG(X_N)(SEQ ID NO:8);
- [0025] (X₁₇₋₂₀)GUUUUAGAGCUAUGCUGAAAAGCAUAGCAAGUUAAAAUAGGCUAGUCCGUUAUC(X_N)(SEQ ID NO:9);
- [0026] (X₁₇₋₂₀)GUUUUAGAGCUAUGCUGUUUUGGAAACAAAACAGCAUAGCAAGUUAAAAUAGGCUAGUCCGUUAUC(X_N)(SEQ ID NO:10);
- [0027] (X₁₇₋₂₀)GUUUUAGAGCUAGAAUAGCAAGUUAAAAUAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC(X_N)(SEQ ID NO:11);
- [0028] (X₁₇₋₂₀)GUUUUAGAGCUAGAAUAGCAAGUUAAAAUAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC(SEQ ID NO:12);
- [0029] (X₁₇₋₂₀)GUUUUAGAGCUAUGCUGGAAACAGCAUAGCAAGUUAAAAUAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC(SEQ ID NO:13); 또는
- [0030] (X₁₇₋₂₀)GUUUUAGAGCUAUGCUGGAAACAGCAUAGCAAGUUAAAAUAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC(SEQ ID NO:14)를 포함하며,

- [0031] 여기서 X_{17-20} 은 표적 서열, 바람직하게는 프로토스페이서 인접 모티프(PAM)의 바로 5'의 표적 서열의 17 개 내지 20 개 연속 뉴클레오타이드의 상보적 가닥에 상보적인 서열이며, X_N 은 Cas9에 대한 리보핵산의 결합을 방해하지 않는 임의의 서열이다. X_{17-20} 서열이 본원에서 *S. 피오제네스* Cas9 시스템으로 예시되지만, 더 긴 서열이, 예로 다른 시스템에 대해 적절하게 이용될 수 있다.
- [0032] 일부 구현예에서, DNA 분자에는 또한 기능적 Csy4 효소를 인코딩하는 서열이 포함된다.
- [0033] 또한, 본원에 기재된 DNA 분자를 포함하는, 예로 선택적으로 기능적 Csy4 효소를 인코딩하는 서열을 포함하는 벡터가 본원에서 제공된다. 또한, DNA 분자에 의해 생성되는 멀티플렉스 전사체, 예로 Csy4로 아직 절단되지 않은 온전한 RNA가 본원에서 제공된다.
- [0034] 다른 양태에서, 세포에서 복수의 가이드 RNA의 생성 방법이 본원에서 제공된다. 방법에는 세포에서 본원에 기재된 DNA 분자를 발현시키는 단계가 포함된다.
- [0035] 일부 구현예에서, 세포는 포유류 세포이며, 세포가 외인성의 기능적 Csy4 효소 서열을 또한 발현시키거나, 또는 방법은 기능적 Csy4 효소 또는 기능적 Csy4 효소를 인코딩하는 핵산을 투여하는 단계를 추가로 포함한다.
- [0036] 다른 양태에서 본 발명은 세포에서 하나 또는 복수의 표적 유전자의 발현을 변형하는 방법을 제공한다. 방법에는 본원에 기재된 DNA 분자, 예로 가이드 RNA(gRNA)를 인코딩하는 복수의 서열을 포함하는 DNA 분자를 발현시키는 단계가 포함되며, 여기서 각각의 gRNA는 하나 이상의 표적 유전자의 적어도 17 개 내지 20 개 nt와 상보적인 가변 서열을 포함하며, 각각의 gRNA에는 서열 GTTCACTGCCGTATAGGCAG(SEQ ID NO:1) 또는 GTTCACTGCCGTATAGGCAGCTAAGAAA(SEQ ID NO:2)를 포함하거나 이로 구성되는 적어도 하나의 Csy4 절단 서열이 인접한다.
- [0037] 본 발명의 방법 및 조성물에서, gRNA는 본원에 기재된 바와 같이 융합된 tracrRNA 및 crRNA를 포함하는 단일 가이드 RNA일 수도 있고, 또는 crRNA만 포함하고, tracrRNA는 동일하거나 상이한 DNA 분자로부터 발현될 수도 있다. 따라서 일부 구현예에서, 본원에 기재된 DNA 분자에는 또한 tracrRNA를 인코딩하는 서열이 포함된다. 일부 구현예에서, 방법에는 세포에서 별도의 tracrRNA를 발현시키는 단계, 예로 tracrRNA를 발현하는 DNA 분자 또는 벡터와 세포를 접촉시키는 단계가 포함된다.
- [0038] 달리 정의되지 않는 한, 본원에서 이용되는 모든 기술적 및 과학적 용어는 본 발명이 속하는 분야의 당업자가 일반적으로 이해하는 바와 동일한 의미를 갖는다. 본 발명에서 사용하기 위한 방법 및 물질이 본원에 기재된다; 당분야에 공지된 다른 적합한 방법 및 물질도 이용될 수 있다. 물질, 방법 및 실시예는 단지 예시이며, 제한하려는 것이 아니다. 본원에서 언급되는 모든 공보, 특허 출원, 특허, 서열, 데이터베이스 엔트리 및 다른 참고문헌은 이들의 전문이 참조로 포함된다. 상충 시에는 정의를 포함하는 본 명세서가 우선할 것이다.
- [0039] 본 발명의 다른 특징 및 장점은 하기 상세한 설명 및 도면, 그리고 청구범위로부터 자명할 것이다.

도면의 간단한 설명

- [0040] 도 1은 하기와 같이, 초기 멀티플렉스 실험에서 이용된 구축물을 예시하는 모식도이다:

1과 2: 별도의 tracrRNA를 갖는 직접 반복체 crRNA 어레이 및 Cas9

3과 4: Csy4, Cas9 및 별도의 tracrRNA를 갖고, Csy4 부위에 의해 분리되는 짧은 crRNA 어레이

5와 6: Csy4 부위에 의해 분리되는 전장 키메라 gRNA.

7: nls-FLAG 태그화된 Cas9

도 2는 도 1에 나타낸 구축물을 발현하는 세포에서의 실험 결과를 나타내는 막대 그래프이다. Csy4 부위와 전장 gRNA(구축물 5 및 6)가 가장 효율적인 멀티플렉스 프레임워크였다.

도 3은 예시적인 표준 및 멀티플렉스 Csy4-기반 gRNA 프레임워크 그리고 이들이 생성하는 전사체의 도식적 개요 및 비교이다. Csy4는 RNA Pol III 프로모터, U6에 대한 대안으로서 RNA Pol II 프로모터(예로, CAG)의 사용을 가능케 함을 주지하라.

도 4는 Csy4가 잘린 인식 부위를 절단하여 인간 세포에서 더 높은 활성을 갖는 gRNA를 생성함을 나타내는 막대 그래프이다. 잘린 부위의 가공은 또한 매끈한 5' 말단을 남기고, U6 프로모터에 의해 부여되는 gRNA 표적 서열

상의 5' G 제한부위를 효율적으로 제거한다.

도 5a 내지 5c는 단일 인간 세포에서 2-표적 멀티플렉스 편집의 증거를 나타내는 서열이다. 의도된 부위 2 또는 3에서 개별 결실이 관찰된다. 부위 2 및 3에 대해 동일한 서열 상에서 다중 결실이 관찰된다. 부위 2 및 3에 결친 결실도 관찰된다.

도 6은 Csy4-기반 시스템을 이용한 3 가지 gRNA의 성공적인 멀티플렉스 발현을 나타내는 모식도이다.

도 7은 RNA Pol II-전사된 mRNA로부터 Csy4에 의해 절제된 gRNA가 인간 세포에서 특이적인 표적으로 Cas9 뉴클레아제를 효율적으로 모집할 수 있음을 나타내는 막대 그래프이다. 이들 실험에서, gRNA는 RNA Pol II CAG 프로모터로부터 제조된 더 긴 mRNA 전사체에서 발현되었다.

발명을 실시하기 위한 구체적인 내용

[0041] 상세한 설명

[0042] Cas9 시스템의 하나의 잠재적인 장점은 세포에서 하나를 초과하는 게놈 유전자위 또는 표적 부위로 뉴클레아제 활성 또는 이중성 효과기 도메인을 모집하는 능력이다. 그러나 이러한 멀티플렉스 적용은 세포에서 하나를 초과하는 gRNA를 효율적으로 발현시키는 능력을 필요로 한다. 포유류 세포에 있어서, 하나의 짧은 gRNA를 발현시키기 위해 RNA 폴리머라제 III 프로모터(예로, U6 프로모터)를 이용하였다. 단일 전사체로부터 단일 세포에서 다중 gRNA 성분을 발현시키려는 이전 시도는 효율적인 것으로 입증되지 않았다.

[0043] Cas9 시스템에 있어서 추가적으로 바람직한 능력은 성분의 유도 가능한 버전을 생성하고, 성분의 조직 특이적 발현을 가능케 하는 것일 것이다. 유도 가능하고/하거나 조직 특이적인 RNA 폴리머라제 II 프로모터는 이전에 기재되었다. 그러나 Cas9 또는 dCas9 단백질이 이러한 RNA pol II 프로모터로부터 발현될 수 있지만, RNA pol II로부터의 전사 개시 및 중지 부위가 부정확하므로 이러한 방식으로 짧은 정의된 gRNA를 발현할 수 없다. 실제로 현재까지, 모든 gRNA는 RNA 폴리머라제 III 프로모터로부터 발현되었고, 이는 짧은 RNA의 발현을 위해 이상적으로 적합하다.

[0044] 본원에서 나타낸 바와 같이, 짧은 RNA 헤어핀 서열을 인식하는 엔도리보뉴클레아제, Csy4를 이용하여 개별 gRNA가 Csy4 절단 부위에 의해 구분되는 하나의 더 긴 RNA 전사체(RNA pol III 프로모터로부터 생성됨) 카세트 상에서 인코딩된 다중 기능적 gRNA를 절단할 수 있다. 기능적 gRNA는 RNA pol II 프로모터로부터 발현되는 더 긴 RNA 전사체로부터 성공적으로 절단될 수 있다.

[0045] gRNA/Csy4 다량체 카세트

[0046] 따라서 본원에서 다량체 카세트에 불리는 더 긴 RNA 전사체를 인코딩하는 DNA 분자가 본원에 기재되며, 여기에는 2 개 이상의 gRNA 서열이 포함되고, 각각의 gRNA에는 Csy4 절단 서열이 인접한다. DNA 분자에는 또한 프로모터가 포함될 수 있고, 선택적으로 하나 이상의 다른 전사 조절 서열, 예로, 인핸서, 사일런서, 인설레이터 및 폴리A 서열이 포함될 수 있다. 예로, [Xu et al., Gene. 2001 Jul 11;272(1-2):149-56]을 참고하라.

[0047] 프로모터

[0048] 본 방법에서 이용될 수 있는 여러 프로모터가 당분야에 공지되어 있다. 일부 구현예에서, 프로모터는 Pol III 또는 Pol III 프로모터, 바람직하게는 Pol II 프로모터이다. CAG 프로모터를 포함하는 다양한 Pol II 프로모터가 기재되었고, 본 발명의 조성물 및 방법에서 이용될 수 있다(예로, 문헌[Alexopoulou et al., BMC Cell Biology 9: 2, 2008; Miyazaki et al., Gene 79(2): 269-77(1989); Niwa et al., Gene 108(2): 193-9(1991)] 참고); 추가적인 프로모터에는 EF1A, CAGGS, PGK, UbiC 및 CMV 프로모터뿐만 아니라 조직 특이적 프로모터, 예컨대 B29, 테스민, 엔도글린, FLT-1, GFP, SYN1이 특히 포함된다; 여러 프로모터의 서열은 당분야에 공지되어 있다. 예를 들어, CMV 및 PGK 프로모터는 pSicoR 및 pSicoR PGK로부터 각각 증폭될 수 있고(Ventura et al., Proc Natl Acad Sci U S A 101: 10380-10385(2004)), UbiC 프로모터는 pDSL_hpUGIH(ATCC)로부터 증폭될 수 있고, CAGGS 프로모터는 pCAGGS(BCCM)로부터 증폭될 수 있고, EF1A 프로모터는 pEF6 벡터(Invitrogen)로부터 증폭될 수 있다. Pol II 코어 프로모터는 문헌[Butler and Kadonaga, Genes & Dev. 16: 2583-2592(2002)]에 기재되어 있다. Pol II 발현에 의해 유도된 더 큰 전사체로부터의 gRNA 절단은 가장-5' 위치에 임의의 뉴클레오티드를 갖는 gRNA를 생성할 수 있도록 한다(U6 또는 다른 RNA 폴리머라제 III 프로모터로부터의 표준 발현은 상기 뉴클레오티드의 정체에 제한을 부여한다).

[0049] 일부 구현예에서, 조직-특이적 프로모터가 이용되며, 짧고 정의된 gRNA 서열이 RNA-Pol II 전사체로부터 가공되

어 나올 수 있다.

- [0050] U6 소형 핵(sn) RNA 프로모터, 7SK 프로모터 및 H1 프로모터를 포함하는 여러 Pol III 프로모터가 당분야에 공지되어 있다. 예로, 문헌[Ro et al., BioTechniques, 38(4):625-627(2005)]를 참고하라.
- [0051] 가이드 RNA
- [0052] Cas9 뉴클레아제는 게놈 DNA 표적 부위에 대해 상보적인 그 5' 말단에 적어도 17 개 내지 20 개 nt를 보유하는 단일 gRNA를 이용해서 서열 NGG의 추가적인 근접 프로토스페이서 인접 모티프(PAM)를 보유하는 적어도 17 개 내지 20 개 nt의 특정한 게놈 표적으로 가이드될 수 있다.
- [0053] 따라서, 본원에 기재된 조성물에는 프로토스페이서 인접 모티프(PAM), 예로, NGG, NAG, 또는 NNGG의 바로 5' 표적 서열의 17 개 내지 20 개 뉴클레오티드(nt)에 상보적인 5' 말단 서열을 갖는, 보통 트랜스로 인코딩되는 tracrRNA에 융합된 crRNA를 포함하는 단일 가이드 RNA(sgRNA), 예로 문헌[Mali et al., Science 2013 Feb 15; 339(6121):823-6]에 기재된 바와 같은 단일 Cas9 가이드 RNA를 인코딩하는 서열이 포함될 수 있다.
- [0054] 가이드 RNA의 설계 및 발현 방법은 당분야에 공지되어 있다. 일반적으로 말하면, 가이드 RNA는 2 개의 상이한 시스템에서 나온다: 1) 시스템 1은 Cas9에 의한 절단을 가이드하기 위해 함께 작용하는 별도의 crRNA 및 tracrRNA를 이용하며; 2) 시스템 2는 단일 시스템에서 2 개의 별도의 가이드 RNA를 조합하는 키메라 crRNA-tracrRNA 하이브리드를 이용한다(Jinek et al. 2012). tracrRNA는 다양하게 잘릴 수 있고, 다양한 범위의 길이가 별도 시스템(시스템 1) 및 키메라 gRNA 시스템(시스템 2)에서 모두 작용하는 것으로 나타났다. 예로, 문헌[Jinek et al., Science 2012; 337:816-821; Mali et al., Science. 2013 Feb 15;339(6121):823-6; Cong et al., Science. 2013 Feb 15;339(6121):819-23; 및 Hwang and Fu et al., Nat Biotechnol. 2013 Mar;31(3):227-9; Jinek et al., Elife 2, e00471(2013)]를 참고하라. 시스템 2에 있어서, 일반적으로 더 긴 길이의 키메라 gRNA가 더 큰 표적 상 활성을 나타내었으나, 현재 다양한 길이의 gRNA의 상대적 특이성은 정의되지 않았고, 이에 따라 어떤 경우에는 더 짧은 gRNA를 이용하는 것이 바람직할 수 있다. 일부 구현예에서, gRNA는 전사 개시 부위의 약 100 bp 내지 800 bp 상류 이내, 예로 전사 개시 부위의 약 500 bp 상류 이내인 영역과 상보적이며, 전사 개시 부위를 포함하거나 전사 개시 부위의 하류 약 100 bp 내지 800 bp 이내, 예로 약 500 bp 이내이다. 일부 구현예에서, 하나를 초과하는 gRNA를 인코딩하는 벡터(예로, 플라스미드), 예로 표적 유전자의 동일한 영역 내의 상이한 부위에 대한 2 개, 3 개, 4 개, 5 개 또는 그 초과인 gRNA를 인코딩하는 플라스미드가 이용된다. 추가적인 가이드 RNA 및 게놈 편집 특이성을 증가시키는 방법은 INCREASING SPECIFICITY FOR RNA-GUIDED GENOME EDITING을 표제로 하는 특허 가출원 일련 번호 61/838,178에 기재된다.
- [0055] 일부 구현예에서, gRNA는 하기 서열:
- [0056] (X₁₇₋₂₀)GUUUUAGAGCUAUGCUGUUUG(X_N)(SEQ ID NO:4);
- [0057] (X₁₇₋₂₀)GUUUUAGAGCUA(SEQ ID NO:5);
- [0058] (X₁₇₋₂₀)GUUUUAGAGCUAUGCUGUUUG(SEQ ID NO:6);
- [0059] (X₁₇₋₂₀)GUUUUAGAGCUAUGC(SEQ ID NO:7);
- [0060] (X₁₇₋₂₀)GUUUUAGAGCUAGAAUAGCAAGUAAAAUAAGGCUAGUCCG(X_N)(SEQ ID NO:8);
- [0061] (X₁₇₋₂₀)GUUUUAGAGCUAUGCUGAAAAGCAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUC(X_N)(SEQ ID NO:9);
- [0062] (X₁₇₋₂₀)GUUUUAGAGCUAUGCUGUUUGGAAACAAAACAGCAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUC(X_N)(SEQ ID NO:10);
- [0063] (X₁₇₋₂₀)GUUUUAGAGCUAGAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC(X_N)(SEQ ID NO:11),
- [0064] (X₁₇₋₂₀)GUUUUAGAGCUAGAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC(SEQ ID NO:12);
- [0065] (X₁₇₋₂₀)GUUUUAGAGCUAUGCUGGAAACAGCAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC(SEQ ID NO:13); 또는
- [0066] (X₁₇₋₂₀)GUUUUAGAGCUAUGCUGGAAACAGCAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC(SEQ ID

NO:14)을 포함하거나 이로 구성되며,

[0067] 여기서, X_{17-20} 은 표적 서열, 바람직하게는 프로토스페이서 인접 모티프(PAM), 예로, NGG, NAG, 또는 NNGG의 바로 5'의 표적 서열의 적어도 17 개 내지 20 개 연속 뉴클레오타드의 상보적 가닥에 상보적인 서열이다(그러나 일부 구현예에서, 상기 상보적 영역은 20 개 nt 초과, 예로, 21 개 nt, 22 개 nt, 23 개 nt, 24 개 nt, 25 개 nt, 26 개 nt, 27 개 nt, 28 개 nt, 29 개 nt, 또는 30 개 nt 또는 그 초과 nt, 예로 17 개 nt 내지 30 개 nt로 더 길 수 있다). X_N 은 임의의 서열이며, 여기서 N(RNA에서)은 Cas9에 대한 리보핵산의 결합을 방해하지 않는 0 내지 300 또는 0 내지 200, 예로 0 내지 100, 0 내지 50, 또는 0 내지 20일 수 있다. 일부 구현예에서, RNA PolIII 전사를 종결시키기 위한 종결 신호로서 이용되는 하나 이상의 T의 선택적 존재의 결과, RNA에는 분자의 3' 말단에 하나 이상의 U, 예로, 1 개 내지 8 개 또는 그 초과 U(예로, U, UU, UUU, UUUU, UUUUU, UUUUUU, UUUUUUU)가 포함된다. 일부 구현예에서, RNA에는 표적 서열과 상보적이 아닌 RNA 분자의 5' 말단에 하나 이상의, 예로 최대 3 개의, 예로 1 개, 2 개 또는 3 개 또는 그 초과 뉴클레오타드가 추가로 포함된다. 선택적으로, 하나 이상의 RNA 뉴클레오타드, 예로 X_{17-20} 내의 하나 이상의 뉴클레오타드, 서열 X_N 내의 하나 이상의 뉴클레오타드, 또는 gRNA의 임의의 서열 내의 하나 이상의 뉴클레오타드가 개질되며, 예로, 잠기며(2'-O-4'-C 메틸렌 가교), 5'-메틸시티딘이거나, 2'-O-메틸-슈도우리딘이거나, 리보스 포스페이트 골격이 폴리아미드 사슬에 의해 대체된다.

[0068] 예를 들어, 일부 구현예에서, 문헌[Jinek et al. (Science. 337(6096):816-21(2012))]에 기재된 키메라 가이드 RNA, 예로 (X_{17-20})GUUUUAGAGCUAGAAUAGCAAGUAAAAUAAGGCUAGUCCG(SEQ ID NO:8); (X_{17-20})GUUUUAGAGCUAGAAUAGCAAGUAAAAUAAGG(SEQ ID NO:9)가 이용될 수 있고; 일부 구현예에서, 문헌[Jinek et al., Elife. 2:e00471(2013)]에 기재된 Cas9 결합을 위해 필요한 표적 DNA 서열에 대해 상보적인 5'-말단 17 개 내지 20 개 뉴클레오타드 서열 및 42 개 뉴클레오타드의 3'-말단 줄기 루프 구조를 보유하는 sgRNA, 예로 (X_{17-20})GUUUUAGAGCUAGAAUAGCAAGUAAAAUAAGGCUAGUCCG(SEQ ID NO:8)가 이용된다.

[0069] 일부 구현예에서, 가이드 RNA에는 3' 말단 상에 하나 이상의 아데닌(A) 또는 우라실(U) 뉴클레오타드가 포함된다.

[0070] 본원에 기재된 실시예는 단일 gRNA를 이용하지만, 방법은 이중 gRNA(예로, 천연 생성 시스템에서 확인된 crRNA 및 tracrRNA)와도 이용될 수 있다. 상기 경우, 단일 tracrRNA가 본 발명의 시스템을 이용해서 발현되는 여러 상이한 crRNA, 예로 하기(RNA에 있어서, T는 본원에서 U인 것으로 이해됨을 주지하라):

[0071] crRNA 서열: X_{17-20} -GTTTTAGAGCTAGAAA(SEQ ID NO:15)

[0072] tracrRNA 서열: TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC(SEQ ID NO:16)와 함께 이용될 것이다. 상기 경우, crRNA는 본원에 기재된 방법 및 분자에서 가이드 RNA로서 이용되며, tracrRNA는 동일하거나 상이한 DNA 분자로부터 발현될 수 있다.

[0073] 또한, 17 개 내지 20 개의 상보성 뉴클레오타드 서열을 갖는 가이드 RNA가 본원에 예시되지만, 일부 구현예에서 더 긴 서열, 예로 21 개 nt, 22 개 nt, 23 개 nt, 24 개 nt, 25 개 nt, 26 개 nt, 27 개 nt, 28 개 nt, 29 개 nt 또는 30 개 nt 또는 그 초과 nt, 예로 17 개 nt 내지 20 개 nt 대신 17 개 nt 내지 30 개 nt가 이용될 수 있다.

[0074] Csy4 절단 서열

[0075] 본원에 기재된 방법 및 조성물에서, Csy4 절단 서열은 각각의 가이드 RNA 간에 하나 또는 적어도 하나의 절단 서열을 가지며 각각의 가이드 RNA에 절단 서열이 인접하도록 DNA 분자 내로 삽입된다. 예시적인 Cys4 절단 서열에는 GTTCACTGCCGTATAGGCAG(잘린 20 개 nt)(SEQ ID NO:1) 및 GTTCACTGCCGTATAGGCAGCTAAGAAA(전장 28 개 nt)(SEQ ID NO:2)가 포함된다. 본원에서 나타난 바와 같이, 잘린 Csy4 절단 부위(SEQ ID NO:1)의 이용은 표준 부위의 이용보다 인간 세포에서 더 효율적이다. 본 발명자들의 최대 지식 범위에서, 이는 인간 세포에서 이용되는 Csy4 활성의 최초 실증이다.

[0076] 기능적 Csy4 효소 서열

[0077] 본원에 기재된 방법에서, Csy4 절단 부위에서 전사체를 절단할 수 있는 기능적 Csy4 효소가 또한 세포에서 발현된다.

[0078] 슈도모나스 애틀루기노사 UCBPP-PA14(Pa14), 예르시니아 페스티스 AAM85295(Yp), 에스체리키아 콜라이 UTI89(Ec89), 디켈로박터 노도수스 VCS1703A(Dn), 아시네토박터 바우만나이 AB0057(Ab), 모리텔라 sp. PE36(MP1, MP01), 쉬와넬라 sp. W3-18-1(SW), 파스퇴렐라 멀토시다 subsp. 멀토시다 Pm70(Pm), 펙토박테리움 와사비애(Pw) 및 디케이 다단티 Ech703(Dd)의 Csy4 동족체로부터의 예시적인 Csy4 서열은 문헌[Haurwitz et al., Science 329(5997): 1355-1358(2010)]의 도 S6에 나타난다. 바람직한 구현예에서, Csy4는 슈도모나스 애틀루기노사에서 유래된다.

[0079] 일부 구현예에서, Csy4는 또한 gRNA에 이중성 효과기 도메인을 공유 연결하기 위해 이용된다. Csy4는 단화-교체 효소인 것으로 여겨지며, 절단 후 그 표적 헤어핀 서열에 결합되어 유지된다(Sternberg et al., RNA. 2012 Apr;18(4):661-72). 따라서 Csy4는 각각의 절단된 gRNA의 3' 말단에 결합된 채 유지되는 것으로 예상된다. 본원에서 나타낸 바와 같이, 절단된 gRNA가 인간 세포에서 기능적인 것으로 나타났으므로, gRNA의 3' 말단 상에서 상기 Csy4 단백질의 존재는 gRNA가 Cas9와 복합체를 형성하고 그 활성을 지시하는 능력에 영향을 미치지 않는 것으로 나타난다. 따라서 이들 gRNA-Csy4 융합도 그 뉴클레아제 촉매 활성을 불활성화하는 돌연변이를 보유하는 Cas9 돌연변이체(dCas9 단백질)를 보낼 수 있을 것으로 추정된다. 따라서, Csy4에 이중성 기능적 도메인(HFD)을 융합시킬 수 있고(Csy4-HFD), 이어서 dCas9:sgRNA:Csy4-HFD 복합체는 이러한 도메인을 특정한 게놈 유전자위로 보낼 수 있었다. 이러한 HFD의 예에는 FokI 유래의 전사 활성화인자 또는 억제인자 도메인 또는 히스톤 또는 DNA 메틸화 상태를 개질하는 다른 도메인과 같은 다른 뉴클레아제 도메인이 포함될 수 있다.

[0080] Csy4-HFD는 Csy4 단백질의 N-말단 또는 C-말단으로 이중성 기능적 도메인(예로, VP64 또는 NF- κ B p65 유래의 전사 활성화 도메인)를 개재 링커, 예로 약 5 개 내지 20 개 또는 13 개 내지 18 개 뉴클레오티드의 링커를 포함하거나 포함하지 않고 융합시켜 생성된다. 전사 활성화 도메인은 Csy4의 N 또는 C 말단 상에 융합될 수 있다. 전사 활성화 도메인에 부가하여, 당분야에 공지된 다른 이중성 기능적 도메인(예로, 전사 억제인자(예로, KRAB, SID 등) 또는 사일런서, 예컨대 헤테로크로마틴 단백질 1(HP1, swi6으로도 알려져 있음), 예로 HP1 α 또는 HP1 β ; MS2 피막 단백질, 엔도리보뉴클레아제 Csy4, 또는 람다 N 단백질에 결합된 것들과 같이 고정된 RNA 결합 서열에 융합된 긴 비-코딩 RNA(lncRNA)를 모집할 수 있는 단백질 또는 펩티드; DNA의 메틸화 상태를 개질하는 효소(예로, DNA 메틸트랜스퍼라제(DNMT) 또는 TET 단백질); 또는 히스톤 서브유닛을 개질하는 효소(예로, 히스톤 아세틸트랜스퍼라제(HAT), 히스톤 탈아세틸라제(HDAC), 히스톤 메틸트랜스퍼라제(예로, 라이신 또는 아르기닌 잔기의 메틸화를 위한) 또는 히스톤 탈메틸화효소(예로, 라이신 또는 아르기닌 잔기의 탈메틸화를 위해))이 또한 이용될 수 있다. 이러한 도메인에 대한 여러 서열, 예로 DNA에서 메틸화된 시토신의 하이드록실화를 촉매하는 도메인은 공지되어 있다. 예시적인 단백질에는 DNA에서 5-메틸시토신(5-mC)을 5-하이드록시메틸시토신(5-hmC)으로 전환하는 효소인 Ten-Eleven-Translocation(TET)1-3 패밀리가 포함된다. 예로, WO/2014/144761을 참고하라.

[0081] 인간 TET1-3에 대한 서열은 당분야에 공지되어 있고, 하기 표에 나타난다:

유전자	GenBank 수탁 번호	
	아미노산	핵산
TET1	NP_085128.2	NM_030625.2
TET2*	NP_001120680.1(변이체 1) NP_060098.3(변이체 2)	NM_001127208.2 NM_017628.4
TET3	NP_659430.1	NM_144993.1

[0082]

[0083] * 변이체 (1)은 더 긴 전사체를 나타내며, 더 긴 이소형 (a)를 인코딩한다. 변이체 (2)는 변이체 1에 비해 5' UTR 및 3' UTR 그리고 코딩 서열이 상이하다. 생성 이소형 (b)는 더 짧고, 이소형 a에 비해 구별되는 C-말단을 갖는다.

[0084] 일부 구현예에서, 촉매 도메인의 전장 서열의 전부 또는 일부, 예로 시스테인-풍부 연장부를 포함하는 촉매 모듈 및 7 개의 고도 보존된 엑손에 의해 인코딩되는 20GFeDO 도메인, 예로 아미노산 1580-2052를 포함하는 Tet1 촉매 도메인, 아미노산 1290-1905를 포함하는 Tet2, 아미노산 966-1678를 포함하는 Tet3이 포함될 수 있다. 예로, 전체 3 개 Tet 단백질에서 주요 촉매 잔기를 예시하는 정렬에 대해, 문헌[Iyer et al., Cell Cycle. 2009 Jun 1;8(11):1698-710. Epub 2009 Jun 27]의 도 1 및 전장 서열에 대해 이들의 보충 자료(ftp 사이트 ftp.ncbi.nih.gov/pub/aravind/DONS/supplementary_material_DONS.html 에서 이용 가능)(예로, 서열 2c 참고)를 참고하라; 일부 구현예에서, 서열에는 Tet1의 아미노산 1418-2136 또는 Tet2/3에서의 대응 영역이 포함된다.

[0085] 다른 촉매 모듈은, 예로 문헌[Iyer et al., 2009]에서 확인된 단백질에서 유래될 수 있다.

[0086] 일부 구현예에서, 융합 단백질에는 Csy4 및 이중성 기능적 도메인 간에 링커가 포함된다. 이들 융합 단백질에서

(또는 연결된 구조의 융합 단백질 간에) 이용될 수 있는 링커에는 융합 단백질의 기능을 방해하지 않는 임의의 서열이 포함될 수 있다. 바람직한 구현예에서, 링커는 짧고, 예로 2 개 내지 20 개 아미노산이며, 전형적으로 가요성이다(즉, 높은 자유도를 갖는 아미노산, 예컨대 글리신, 알라닌 및 세린을 포함함). 일부 구현예에서, 링커는 GGGG(SEQ ID NO:3) 또는 GGGGS(SEQ ID NO:17)로 구성된 하나 이상의 단위, 예로 GGGG(SEQ ID NO:3) 또는 GGGGS(SEQ ID NO:17) 단위의 2 개, 3 개, 4 개 또는 그 초과와 반복체를 포함한다. 다른 링커 서열, 예로 GGS, GSG(SEQ ID NO:22), SGSETPGTSESA(SEQ ID NO:23), SGSETPGTSESATPES(SEQ ID NO:24), 또는 SGSETPGTSESATPEGSGGS(SEQ ID NO:25)가 또한 이용될 수 있다.

[0087] Cas9

[0088] 본원에 기재된 방법에서, Cas9가 또한 세포에서 발현된다. 여러 박테리아가 Cas9 단백질 변이체를 발현한다. *스트렙토코커스 피오케네스* 유래의 Cas9가 현재 가장 일반적으로 이용된다; 일부 다른 Cas9 단백질은 *S. 피오케네스* Cas9와 높은 수준의 서열 동일성을 가지며, 동일한 가이드 RNA를 이용한다. 다른 것들은 더욱 다양하며, 상이한 gRNA를 이용하고, 상이한 PAM 서열(RNA에 의해 특정되는 서열과 인접한 단백질에 의해 특정되는 2 개 내지 5 개의 뉴클레오티드 서열)을 또한 인식한다. Chylinski 등은 Cas9 단백질을 큰 그룹의 박테리아로부터 분류하였고(RNA Biology 10:5, 1-12; 2013), 다수의 Cas9 단백질이 이들의 보충 도 1 및 보충 표 1에 기재되며, 본원에 참조로 포함된다. 본원에 기재된 구축물 및 방법에는 임의의 Cas9 단백질 및 이들의 대응하는 가이드 RNA 또는 상용성인 다른 가이드 RNA의 이용이 포함될 수 있다. *스트렙토코커스 쉐모필리스* LMD-9 CRISPR1 시스템 유래의 Cas9는 문헌[Cong et al (Science 339, 819(2013))]에서 인간 세포에서도 기능하는 것으로 나타났다. 추가적으로, Jinek 등은 시험관내에서 *S. 쉐모필리스* 및 *L. 인노쿠아* 유래의(그러나 상이한 가이드 RNA를 이용할 가능성이 있는 *N. 매닝기티디스* 또는 *C. 제주니* 유래는 아님) Cas9 오르쏘로그가 이중 *S. 피오케네스* gRNA에 의해 가이드되어 약간 감소된 효율을 갖기는 하지만 표적 플라스미드 DNA를 절단할 수 있음을 나타내었다.

[0089] 일부 구현예에서, 본 발명의 시스템은 박테리아에서 인코딩되거나 포유류 세포에서의 발현을 위해 코돈-최적화된 *S. 피오케네스* 유래의 Cas9 단백질을 이용한다. HA 에피토프(아미노산 서열 DAYPYDVPDYASL(SEQ ID NO:18)) 및 핵 편재 신호(아미노산 서열 PKKKRKVEDPKKKRKVD(SEQ ID NO:19))에 융합된 *스트렙토코커스 피오케네스* Cas9 (잔기 1-1368)의 예시적인 서열은 하기와 같다:

[0090] MDKKYSIGLDIGTNSVGVAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNR
ICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSIDKADLRILIY
ALAHMIKFRGHFLIEGDLNPDNSDVKLFQIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKQLSKDQYDQYADLFLAAKNLSDAILLSDILRVNTEITKAP
LSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKNLRE
DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASASQSFIERMTNFDKNLPNEKVLPHKSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTN
RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWRSLRKLINGIRDKQSGKTILDFLKSDFANRNFQMLIHDDSLTFKEDIQKAQVSGQ
GDSLHEHIANLAGSPAICKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEIEGKELGSQL
KEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDQVHIVPQSFLKDDSIDNKVLTSDKNRGKSDNVPSEEVV

[0091] KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVI
TLKSKLVSDFRKDFQFYKVFREINNYHHADAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAY
FFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKDLII
KLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLLASHYKELKGSPEQNEQKQLFVEQHKHYLDEIIEQISEF
SKRVILADANLDKVLASAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQISITGLY

[0092] ETRIDLSQLGGDAYPYDVPDYASLGGSGSPKKRKVEDPKKKRKVD (SEQ ID NO:20)

[0093] 문헌[Jinek et al, 2013, 상기]를 참고하라.

[0094] 일부 구현예에서, 뉴클레아제를 니카아제로 만드는 D10A 및 H840A 돌연변이 중 하나 또는 단백질의 뉴클레아제 부분을 촉매적으로 불활성으로 만드는 D10A 및 H840A 돌연변이 둘 다를 함유하는 Cas9 서열이 이용된다. 본원에 기재된 방법 및 조성물에서 이용될 수 있는 촉매적으로 불활성인 *S. 피오케네스* Cas9(dCas9)의 서열은 하기와 같다; 돌연변이를 진한 글씨체로 밑줄 쳐서 나타낸다.

10 20 30 40 50 60
 MDKKYSIGLA IGTSVGVAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE
 70 80 90 100 110 120
 ATRLKRTARR RYTRRKNRIC YLQEIFSNEK AKVDDSFHHR LEESFLVEED KKHHRHPFG
 130 140 150 160 170 180
 NIVDEVAYHE KYPTIYHLRK KLVDSTDKAD LRLIYLALAH MIKFRGHFLI EGDLPDNDSD
 190 200 210 220 230 240
 VDKLFIQLVQ TYNQLFEENP INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN
 250 260 270 280 290 300
 LIALSLGLTP NFKSINFDAE DAKLQLSKDT YDDDLNLLA QIGDQYADLF LAAKNLSDAI
 310 320 330 340 350 360
 LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLLKALVR QQLPEKYKEI FFDQSKNGYA
 370 380 390 400 410 420
 GYIDGGASQE EFYKFIKPIE EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH
 430 440 450 460 470 480
 AILRRQEDFY PFLKDNREKI EKILTFRIPY YVGPLARGNS RFAWMTRKSE ETITPWNFEE
 490 500 510 520 530 540
 VVDKGASAQE FIERMTNFDK NLPNEKVLPE HSLLYEYFTV YNELTKVKYV TEGMRKPAFL
 550 560 570 580 590 600
 SGEQKKAIVD LLFKTNRKVT VKQLKEDYFK KIECFDSVEI SGVEDRFNAS LGTYHDLKI
 610 620 630 640 650 660
 IKDKDFLDNE ENEDILEDIV LTLTLFEDRE MIEERLKYA HLFDDKVMKQ LKRRRYTGWG
 670 680 690 700 710 720
 RLSRKLINGI RDKQSGKTIL DFLKSDGFAN RNFMQLIHDD SLTFKEDIQK AQVSGQGDSL
 730 740 750 760 770 780
 HEHIANLAGS PAIKKGILQT VKVDELVKV MGRHKPENIV IEMARENQTT QKGQKNSRER
 [0095]
 790 800 810 820 830 840
 MKRIEEGIKE LGSQILKEHP VENTQLQNEK LYLYYLQNGR DMYVDQELDI NRLSDYDVA
 850 860 870 880 890 900
 IVPQSFLKDD SIDKNVLTRS DKNRGKSDNV PSEEVVKMK NYWRQLLNAK LITQRKFDNL
 910 920 930 940 950 960
 TKAERGGLSE LDKAGFIKQ LVETRQITKH VAQILDSRMN TKYDENDKLI REVKVITLKS
 970 980 990 1000 1010 1020
 KLVSDFRKDF QFYKVVREINN YHHAHDAYLN AVVGTALIKK YPKLESEFVY GDYKVYDVRK
 1030 1040 1050 1060 1070 1080
 MIAKSEQEIG KATAKYFFYS NIMNFFKTEI TLANGKIRK PLIETNGETG EIVWDKGRDF
 1090 1100 1110 1120 1130 1140
 ATVRKVLSP QVNIVKKTEV QTGGFSKESI LPKRNSDKLI ARKKDWDPPK YGGFDSPTVA
 1150 1160 1170 1180 1190 1200
 YSVLVVAKVE KGKSKKLKSV KELLGITIME RSSFEKNPID FLEAKGYKEV KKDIIKLKP
 1210 1220 1230 1240 1250 1260
 YSLFELENGR KRMLASAGEL QKGNELALPS KYVNFYLLAS HYEKLKGSPE DNEQKQLFVE
 1270 1280 1290 1300 1310 1320
 QHKHYLDEII EQISEFSKRV ILADANLDKV LSAYNKHDK PIREQAENII HLFTLTNLGA
 [0096]

- 1330 1340 1350 1360
PAAFKYFDTT IDRKYRTSTK EVLDATLIHQ SITGLYETRI DLSQLGGD (SEQ ID NO:21). 예로, 문헌[Mali et al., 2013, 상기; 및 Jinek et al., 2012, 상기]를 참고하라. 대안적으로, Cas9는 2013년 6월 21 일에 출원된 RNA-GUIDED TARGETING OF GENETIC AND EPIGENOMIC REGULATORY PROTEINS TO SPECIFIC GENOMIC LOCI를 표제로 하고 일련 번호 61/838,148이 할당된 미국 특허 가출원 및 PCT/US2014/027335에 기재된 바와 같은 dCas9-이종기능적 도메인 융합물(dCas9-HFD)일 수 있다.
- Cas9는 본원에 기재된 바와 같은 발현 벡터, 예로 Cas9를 인코딩하는 서열, 예로 Cas9 cDNA 또는 유전자를 포함하는 염색체외 플라스미드 또는 바이러스 벡터로부터 발현될 수 있고; 외인성 Cas9 cDNA 또는 세포 계능 내로 통합된 유전자로부터 발현될 수 있고; Cas9를 인코딩하는 mRNA; 실제 Cas9 단백질 자체; 또는 비-포유류의 세포의 경우, 외인성 Cas9일 수 있다.
- 발현 시스템**
- 예로, 본원에 기재된 Csy4/가이드 RNA 구축물의 생체내 또는 시험관내 발현을 위한, 핵산 분자를 포함하는 발현 벡터가 이용될 수 있다. 다중 gRNA 발현용 벡터(잠재적으로는 유도 가능한 또는 조직-/세포-형 특이적인 방식으로)가 연구 및 치료 적용을 위해 이용될 수 있다.
- 본원에 기재된 융합 단백질 및 다량체성 가이드 RNA 카세트를 이용하기 위해, 이들을 인코딩하는 핵산으로부터 이들을 발현시키는 것이 바람직할 수 있다. 이는 다양한 방식으로 수행될 수 있다. 예를 들어, 가이드 RNA 카세트 또는 Csy4 또는 Cas9 단백질을 인코딩하는 핵산이 복제 및/또는 발현용 원핵생물 또는 진핵생물 세포 내로의 형질전환을 위해 중간체 벡터 내로 클로닝될 수 있다. 중간체 벡터는 전형적으로 원핵생물 벡터, 예로 융합 단백질 생산을 위해 또는 융합 단백질을 인코딩하는 핵산의 보관 또는 조작을 위한 플라스미드, 또는 서플 벡터, 또는 곤충 벡터이다. 가이드 RNA 또는 융합 단백질을 인코딩하는 핵산은 또한 식물 세포, 동물 세포, 바람직하게는 포유류 세포 또는 인간 세포, 진균 세포, 박테리아 세포 또는 원생동물 세포로의 투여를 위해, 발현 벡터 내로 클로닝될 수 있다.
- 발현을 수득하기 위해, 가이드 RNA 또는 융합 단백질을 인코딩하는 서열은 전형적으로 전사를 지시하기 위한 프로모터를 함유하는 발현 벡터 내로 서브클로닝된다. 적합한 박테리아 및 진핵생물 프로모터는 당분야에 널리 공지되어 있고, 예로 문헌[Sambrook et al., Molecular Cloning, A Laboratory Manual(3d ed. 2001); Kriegler, Gene Transfer and Expression: A Laboratory Manual(1990); 및 Current Protocols in Molecular Biology(Ausubel et al., eds., 2010)]에 기재되어 있다. 조작된 단백질을 발현하기 위한 박테리아 발현 시스템은, 예로 *E. 콜라이*, *바실러스* sp. 및 *살모넬라*(Palva et al., 1983, Gene 22:229-235)에서 이용 가능하다. 이러한 발현 시스템을 위한 키트가 상업적으로 이용 가능하다. 포유류 세포, 효모 및 곤충 세포용 진핵생물 발현 시스템은 당분야에 널리 공지되어 있고, 또한 상업적으로 이용 가능하다.
- 여러 적합한 벡터, 예로 재조합 레트로바이러스, 렌티바이러스, 아데노바이러스, 아데노-연관 바이러스 및 단순 헤르페스 바이러스 1, 아데노바이러스-유래 벡터를 포함하는 바이러스 벡터 또는 재조합 박테리아 또는 진핵생물 플라스미드는 당분야에 공지되어 있다. 예를 들어, 발현 구축물에는 코딩 영역; 프로모터 서열, 예로 선택된 세포 유형으로 발현을 제한하는 프로모터 서열, 조건 프로모터, 또는 강력한 일반 프로모터; 인핸서 서열; 미변역 조절 서열, 예로 5' 미변역 영역(UTR), 3' UTR; 폴리아데닐화 부위; 및/또는 인설레이터 서열이 포함될 수 있다. 이러한 서열은 당분야에 공지되어 있고, 당업자는 적합한 서열을 선택할 수 있을 것이다. 예로, 문헌[Current Protocols in Molecular Biology, Ausubel, F.M. et al. (eds.) Greene Publishing Associates,(1989), Sections 9.10-9.14] 및 다른 표준 실험실 매뉴얼을 참고하라. 일부 구현예에서, 발현은 당분야에 공지된 바와 같은 조직-특이적 프로모터를 이용해서 특정한 세포 유형으로 제한될 수 있다.
- 상술된 바와 같이, 가이드 RNA를 발현하기 위한 벡터에는 가이드 RNA의 발현을 유도하기 위한 RNA Pol II 또는 Pol III 프로모터가 포함될 수 있다. 이들 인간 프로모터는 플라스미드 전달감염 후 포유류 세포에서 gRNA의 발현을 허용한다. 대안적으로, 예로 시험관내 전사를 위해, T7 프로모터가 이용될 수 있고, RNA가 시험관내 전사되고 정제될 수 있다. 핵산 발현을 지시하기 위해 이용되는 프로모터는 특정 적용에 의존한다. 예를 들어, 전형적으로 융합 단백질의 발현 및 정제를 위해 강력한 항상성 프로모터가 이용된다. 대조적으로, 융합 단백질이 유전자 조절을 위해 생체내 투여되어야 하는 경우, 융합 단백질의 특정 용도에 따라, 항상성 또는 유도성 프로모터가 이용될 수 있다. 또한, 융합 단백질의 투여를 위해 바람직한 프로모터는 약한 프로모터, 예컨대 HSV TK 또는 유사한 활성을 갖는 프로모터일 수 있다. 프로모터에는 또한 트랜스 활성화에 반응하는 요소, 예로 저산소증 반응 요소, Gal4 반응 요소, lac 억제인자 반응 요소 및 소분자 제어 시스템, 예컨대 테트라사이클린-조절 시스템

템 및 RU-486 시스템이 포함될 수 있다(예로, Gossen & Bujard, 1992, Proc. Natl. Acad. Sci. USA, 89:5547; Oligino et al., 1998, Gene Ther., 5:491-496; Wang et al., 1997, Gene Ther., 4:432-441; Neering et al., 1996, Blood, 88:1147-55; 및 Rendahl et al., 1998, Nat. Biotechnol., 16:757-761 참고).

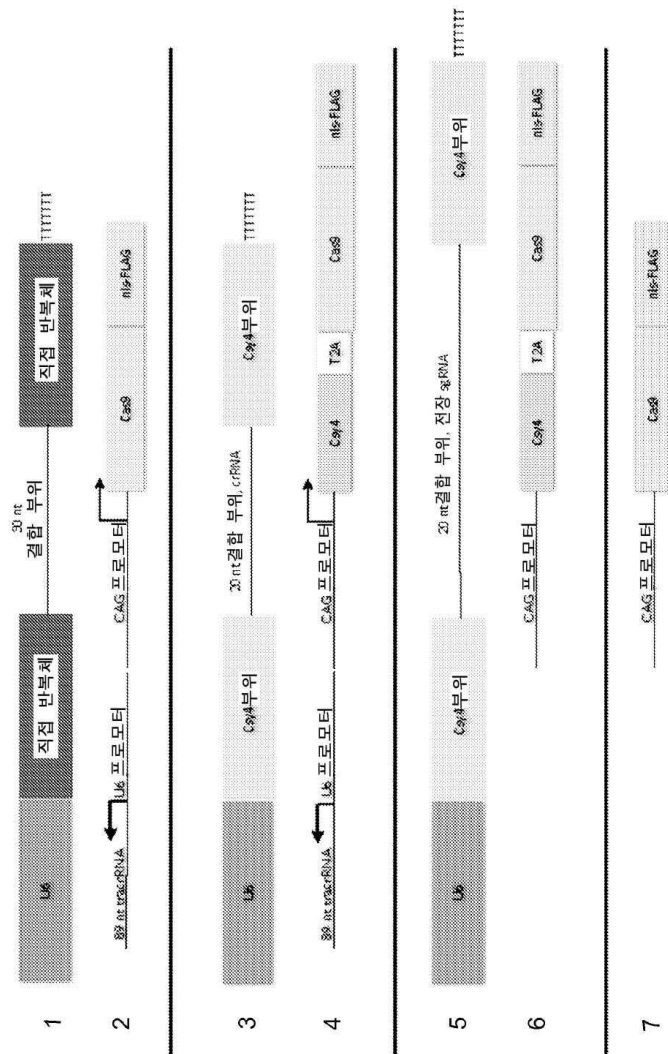
- [0105] 프로모터에 부가하여, 발현 벡터는 전형적으로 원핵생물 또는 진핵생물인 숙주 세포에서 핵산 발현을 위해 필요한 모든 추가 요소를 함유하는 발현 카세트 또는 전사 단위를 함유한다. 따라서 전형적인 발현 카세트는, 예로 융합 단백질을 인코딩하는 핵산 서열에 대해 작동 가능하게 연결된 프로모터 및 예로 전사체의 효율적인 폴리아데닐화, 전사 종결, 리보솜 결합 부위 또는 번역 종결을 위해 필요한 임의의 신호를 함유한다. 카세트의 추가 요소에는, 예로 인핸서 및 이중성의 스플라이스된 인트론 신호가 포함될 수 있다.
- [0106] 세포 내로 유전 정보를 수송하기 위해 이용되는 특정 발현 벡터는 융합 단백질의 의도되는 용도, 예로 식물, 동물, 박테리아, 진균, 원생동물 등에서의 발현에 대해 선택된다. 표준 박테리아 발현 벡터에는 플라스미드, 예컨대 pBR322 기반 플라스미드, pSKF, pET23D 및 상업적으로 이용 가능한 태그-융합 발현 시스템, 예컨대 GST 및 LacZ가 포함된다. 바람직한 태그-융합 단백질은 말토스 결합 단백질(MBP)이다. 이러한 태그-융합 단백질은 조작된 TALE 반복 단백질의 정제를 위해 이용될 수 있다. 에피토프 태그, 예로, c-myc 또는 FLAG가 또한 재조합 단백질에 부가되어 발현 모니터링을 위해, 그리고 세포 및 세포하 편재의 모니터링을 위해 편리한 단리 방법을 제공할 수 있다.
- [0107] 진핵생물 바이러스 유래의 조절 요소를 함유하는 발현 벡터, 예로 SV40 벡터, 파필로마 바이러스 벡터 및 엡스타인-바 바이러스에서 유래된 벡터가 종종 진핵생물 발현 벡터에서 이용된다. 다른 예시적인 진핵생물 벡터에는 pMSG, pAV009/A+, pMT010/A+, pMAMneo-5, 배칼로바이러스 pDSVE 및 SV40 초기 프로모터, SV40 후기 프로모터, 메탈로티오나인 프로모터, 쥐과 유방암 바이러스 프로모터, 라우스 육종 바이러스 프로모터, 폴리헤드린 프로모터, 또는 진핵생물 세포에서의 발현을 위해 효과적인 것으로 나타난 다른 프로모터의 지시 하에 단백질 발현을 허용하는 임의의 다른 벡터가 포함된다.
- [0108] 일부 발현 시스템은 안정적으로 전달감염된 세포주의 선택을 위한 마커, 예컨대 티미딘 키나아제, 하이그로마이신 B 포스포트랜스퍼라제 및 디하이드로폴레이트 환원효소를 갖는다. 폴리헤드린 프로모터 또는 다른 강력한 배칼로바이러스 프로모터의 지시 하에 융합 단백질 인코딩 서열을 갖는, 곤충 세포에서 배칼로바이러스 벡터를 이용하는 것과 같은 고수율 발현 시스템이 또한 적합하다.
- [0109] 전형적으로 발현 벡터에 포함되는 요소에는 또한 *E. 콜라이*에서 기능하는 레플리콘, 재조합 플라스미드를 보유하는 박테리아 선택을 허용하는 항생제 내성을 인코딩하는 유전자 및 재조합 서열의 삽입을 허용하는 플라스미드의 비필수 영역에 독특한 제한효소 부위가 포함된다.
- [0110] 표준 전달감염 방법을 이용하여 다량의 단백질을 발현하는 박테리아, 포유류, 효모 또는 곤충 세포주를 생성한 후 표준 기법을 이용해서 정제된다(예로, Colley et al., 1989, J. Biol. Chem., 264:17619-22; Guide to Protein Purification, in Methods in Enzymology, vol. 182(Deutscher, ed., 1990) 참고). 진핵생물 및 원핵생물 세포의 형질전환은 표준 기법에 따라 수행된다(예로, 문헌[Morrison, 1977, J. Bacteriol. 132:349-351; Clark-Curtiss & Curtiss, Methods in Enzymology 101:347-362(Wu et al., eds, 1983)] 참고).
- [0111] 숙주 세포 내로 외래 뉴클레오티드 서열을 도입하기 위한 임의의 공지된 절차가 이용될 수 있다. 이들에는 인산칼슘 전달감염, 폴리브렌, 원형질체 융합, 전기천공, 핵감염, 리포솜, 마이크로주사, 네이키드 DNA, 플라스미드 벡터, 바이러스 벡터, 에피솜 및 통합형 둘 다, 그리고 숙주 세포 내로 클로닝된 게놈 DNA, cDNA, 합성 DNA 또는 다른 외래 유전 물질을 도입하기 위한 임의의 다른 널리 공지된 방법(예로, 문헌[Sambrook et al., 상기] 참고)의 이용이 포함된다. 이용된 특정 유전 조작 절차가 선택 단백질을 발현할 수 있는 숙주 세포 내로 적어도 하나의 유전자를 성공적으로 도입할 수 있어야 하는 것만이 필요하다.
- [0112] 일부 구현예에서, Cas9 또는 Csy4 단백질에는 핵으로 전위될 단백질을 제공하는 핵 편재 도메인이 포함된다. 몇몇 핵 편재 서열(NLS)이 공지되어 있고, 임의의 적합한 NLS가 이용될 수 있다. 예를 들어, 여러 NLS는 이분 염기성 반복체(문헌[Garcia-Bustos et al, 1991, Biochim. Biophys. Acta, 1071:83-101]에서 검토됨)로 불리는 복수의 염기성 아미노산을 갖는다. 이분 염기성 반복체를 함유하는 NLS는 키메라 단백질의 임의의 부분에 배치되고 핵 내에 편재되는 키메라 단백질을 생성할 수 있다. 바람직한 구현예에서, 본원에 기재된 융합 단백질의 궁극적 기능은 전형적으로 단백질이 핵에 편재되는 것을 필요로 할 것이므로, 핵 편재 도메인이 최종 융합 단백질 내로 도입된다. 그러나, 단백질이 내재적 핵 전위 기능을 갖는 경우 별도의 핵 편재 도메인을 부가할 필요가 없을 수 있다.

- [0113] 본 발명에는 벡터 및 벡터를 포함하는 세포가 포함된다.
- [0114] *라이브러리*
- [0115] 또한, 예로 연구 적용을 위해, 예로 잠재적인 약물 표적에 대해 스크리닝하거나 유전 수준에서 상호작용을 정의하기 위해 유도 가능한, 조직- 또는 세포-형 특이적 멀티플렉스 벡터에서의 gRNA의 조합 라이브러리가 본원에 제공된다.
- [0116] **이용 방법**
- [0117] 기재된 방법에는 세포에서 발현시키는 단계 또는 본원에 기재된 바와 같은 다량체성 카세트와 단축된 gRNA에 의해 가이드될 수 있는 뉴클레아제, 예로 상술된 바와 같은 Cas9 뉴클레아제 및 상술된 바와 같은 Csy4 뉴클레아제와 세포를 접촉시키는 단계가 포함될 수 있다.
- [0118] 기재된 시스템은 여러 내인성 유전자의 발현을 동시에 개질하거나 단일 유전자의 여러 부분을 표적화하기 위해 유용한 다방면 도구이다. 이를 달성하기 위한 현재의 방법은 표적화될 각 부위에 대해 별도의 gRNA-인코딩 전사체의 이용을 필요로 한다. 각각의 세포가 각각의 gRNA를 발현하도록 보장하는 것이 불가능하므로, 별도의 gRNA는 세포 모집단의 멀티플렉스 게놈 편집을 위해 최적이지 않다; 다중 전사체로, 세포는 gRNA의 복잡하고 불균일한 무작위 혼합물을 얻는다. 그러나 본 발명의 시스템은 단일 전사체로부터 다중 gRNA의 발현을 허용하며, 이는 다중 gRNA의 발현에 의한 게놈에서의 다중 부위의 표적화를 허용한다. 또한, 단일-전사체 시스템으로, 각각의 세포는 유사한 화학양론으로 모든 gRNA를 발현할 것이다. 따라서 상기 시스템은 다수의 유전자 발현을 동시에 변형하거나 단일 유전자, 프로모터 또는 인핸서로 다중 Cas9 또는 HFD를 모집하기 위해 쉽게 이용될 수 있다. 상기 능력은, 예로 기초 생물학 연구를 위해 광범위한 유용성을 가질 것이며, 여기서 유전자 기능을 연구하고 단일 경로 및 합성 생물학에서 다중 유전자의 발현을 조작하기 위해 이용될 수 있고, 이는 연구자들로 하여금 다중 입력 신호에 반응하는 세포 내 회로를 생성할 수 있도록 할 것이다. 상기 기술이 구현되고 멀티플렉스화에서 채택될 수 있는 상대적인 용이성으로 인해, 이를 여러 광범위 적용을 갖는 널리 유용한 기술로 만들 것이다.
- [0119] 본원에 기재된 방법에는 하나 이상의 유전자 및 Csy4 및 Cas9를 인코딩하는 핵산에 대한 본원에 기재된 다량체성 gRNA 카세트를 인코딩하는 핵산과 세포를 접촉시켜서 하나 이상의 유전자의 발현을 조정하는 단계가 포함된다.
- [0120] **실시예**
- [0121] 본 발명은 하기 실시예에서 추가로 설명되며, 이는 청구범위에 기재된 본 발명의 범위를 제한하지 않는다.
- [0122] *실시예 1. CRISPR/Cas9를 이용한 멀티플렉스 편집*
- [0123] 단일 전사체로부터 발현된 crRNA 또는 sgRNA의 어레이로부터 CRISPR/Cas9를 이용한 멀티플렉스 편집을 수행하는 목적으로, 하기와 같이 3 가지 전략을 시도하였다:
- [0124] 1. 별도의 tracrRNA를 갖는 직접-반복체가 인접한 crRNA 어레이 및 Cas9
- [0125] 2. Csy4, Cas9 및 별도의 tracrRNA와 함께 발현되는, Csy4 부위에 의해 분리되는 짧은 crRNA 어레이
- [0126] 3. Csy4 부위에 의해 분리되는 전장 단일 가이드 RNA(sgRNA)
- [0127] 각 세트의 구축물(도 1에 예시됨)을 U2OS-EGFP 손상 검정에서 EGFP를 효율적으로 손상시키는 능력에 대해 시험하였다. 결과를 도 2에 나타낸다. 전략 1 및 2를 이용하여 설계된 구축물은 단일 표적에 대해서도 EGFP-손상 검정에서 최저 활성을 나타내었다; 따라서 추가 실험(후술됨)은 전략 3의 최적화에 집중하였다.
- [0128] *실시예 2. 포유류 세포에서 RNA 폴리머라제 II 및 III 프로모터로부터 고효율 CRISPR 가이드 RNA의 멀티플렉스 발현*
- [0129] Csy4 뉴클레아제를 이용하여 더 긴 전사체로부터 gRNA를 절단해내기 위한 예시적인 전략의 모식적 개요를 도 3에 나타낸다. 개념-증명을 실증하기 위한 초기 실험에서, 2 개 버전의 Csy4-절단된 RNA 헤어핀 부위를 인간 세포에서 절단에 대해 시험하였다. 이를 수행하기 위해, 2 개의 Csy4 절단 부위 중 하나가 인접한 gRNA를 발현시켰다:
- [0130] 1. GTTCACTGCCGTATAGGCAGCTAAGAAA(전체 28 개 nt)(SEQ ID NO:2)
- [0131] 2. GTTCACTGCCGTATAGGCAG(잘린 20 개 nt)(SEQ ID NO:1)

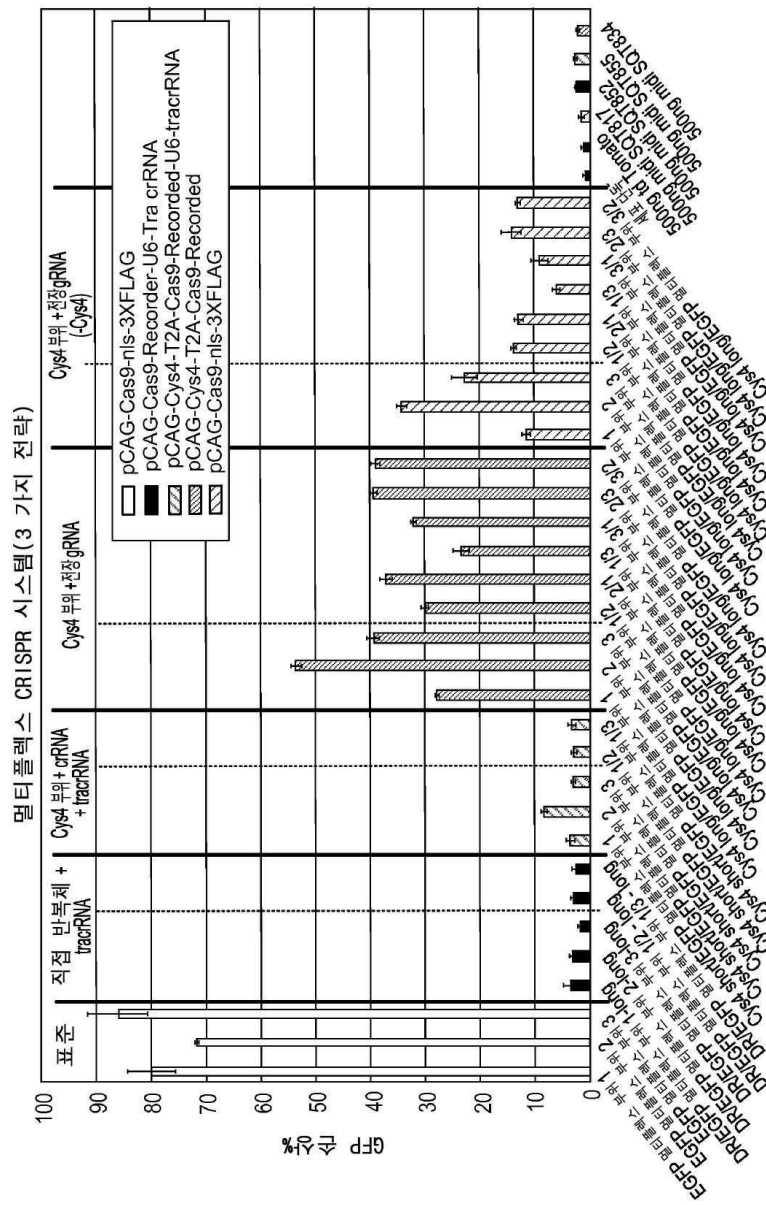
- [0132] 결과는 이들의 5' 및 3' 말단 상에 잘린 20 개 nt 서열이 인접한 gRNA가 더 긴 28 개 nt 서열이 인접한 것들보다 포유류 세포에서 더 활성이 있었음을 나타내었다(도 4). 본 발명자들의 최대 지식 범위에서, 이는 Csy4 뉴클레아제가 생활성 인간 세포에서 RNA 전사체를 가공하기 위해 이용될 수 있다는 최초 실증이다. 20 개 nt의 잘린 부위의 하나의 중요한 추가적 장점은, 더 긴 28 개 nt 서열과는 달리, 이것이 더 긴 전사체로부터 가공된 gRNA의 5' 말단 상에 임의의 추가적인 뉴클레오티드를 남기지 않는다는 것이다(도 4). 이는 가장-5' 위치에 임의의 원하는 뉴클레오티드를 갖는 gRNA의 발현을 가능케 한다. 이는 가장-5' 위치에 특정한 뉴클레오티드(들)에 대한 요건을 갖는 RNA 폴리머라제 III 프로모터로부터의 gRNA 발현에 대비하여 개선이다.
- [0133] 상기 Csy4-기반 시스템을 이용하여, 2 개 및 3 개의 상이한 gRNA(도 5 및 6)의 효율적인 발현이 실증되었다. 상기 접근을 이용해서 동시 발현된 gRNA는 인간 세포에서 예상된 부위에서 변형을 유도하였다.
- [0134] 이들 결과는 또한 상기 Csy4-기반 전략이 RNA Pol II 프로모터에 의해 생성된 더 긴 mRNA 상에 인코딩된 gRNA와 함께 이용될 수 있음을 실증하였다(도 7). 이들 실험에서, 잘린 Csy4 부위가 인접한 3 개의 상이한 단일 gRNA 중 하나가 CAG 프로모터(RNA Pol II 프로모터)에 의해 생성된 mRNA 상에 인코딩되었다. 도 7에 나타낸 바와 같이, 이들 구축물의 전체 3 개는 Csy4의 존재 하에서만 인간 세포에서 Cas9 뉴클레아제를 보낼 수 있는 기능적 gRNA를 생성할 수 있다. 관찰되는 표적화된 Cas9 활성 수준은 (비록 다소 더 낮지만) 이들 gRNA가 표준 RNA Pol III 프로모터를 이용해서 단독으로 발현되는 경우 관찰되는 것과 또는 RNA Pol III 프로모터로부터의 Csy4-인접한 전사체에서와 필적하였다.(도 7)
- [0135] 요약하면, 본 발명의 결과는 1) 인간 세포에서 Csy4의 존재 하에 그리고 Csy4 절단 부위에 의해 분리되는 경우, 최대 3 개의 기능적 gRNA가 단일 RNA pol III 전사체로부터 생성될 수 있고, 2) 다중 Csy4-가공된 gRNA가 단일 인간 세포에서 멀티플렉스 변화를 도입하도록 Cas9 뉴클레아제를 보내기 위해 이용될 수 있고, 3) Csy4 절단 부위가 인접한 기능적 gRNA가 RNA 폴리머라제 II 프로모터로부터 제조된 더 긴 mRNA 전사체로부터의 Csy4 뉴클레아제에 의해 절제될 수 있음을 실증하였다.
- [0136] **다른 구현예**
- [0137] 본 발명이 이들의 상세한 설명과 함께 기재되었으나, 상기 기재는 첨부된 청구범위의 범위에 의해 정의되는 본 발명의 범위를 제한하려는 것이 아니며 예시하려는 것임이 이해되어야 한다. 다른 양태, 장점 및 개질은 하기 청구범위의 범위 내에 속한다.

도면

도면1

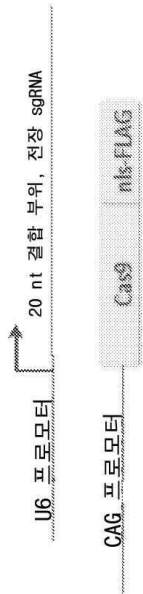


도면2

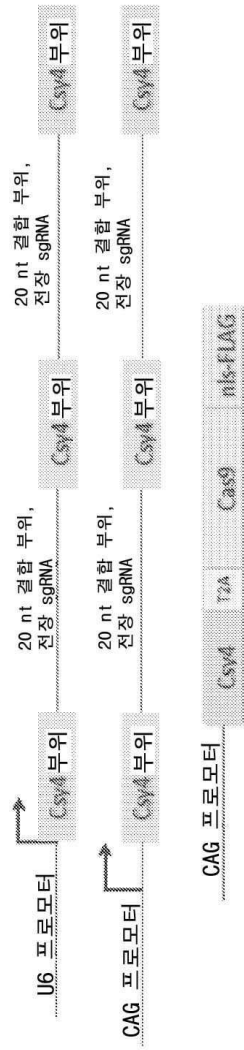


도면3

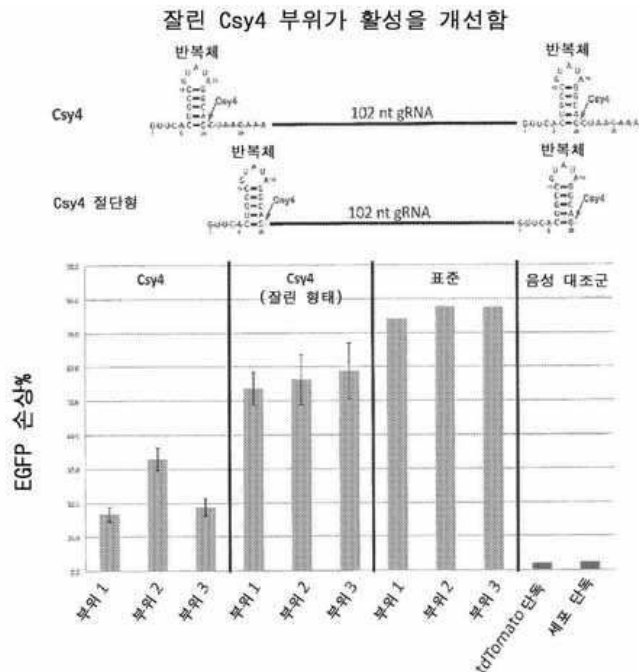
표준 gRNA 및 Cas9 발현 벡터



Csy4 부위가 인접한 멀티플렉스 gRNA 및 Csy4/Cas9 공동발현 벡터



도면4



도면5a

부위 2 GCGAGGTGAAGTTCGAGGGCGAC
부위 3 CCTACGGCGTGCAGTGCTTCAGC
(부위S3/2)

ACGTAACAGCGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC

A1 GTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC

A2 TCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC

A12 TCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC

B12 CCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC

C3 TCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC

C4GTAACAGCGCCACAAGTTCAGCGT-----

C8 CCAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC

D2 GGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC

D4 TAAACGCGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC

D12 CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC

E1 TAAACGCGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC

G1 AAACGCGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC

G7 ACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC

G9 AAACGCGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC

G12 AAACGCGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC

[illegible]

도면5b

```

GTGCAGTGCCTTCAGCCGCTACCCCGACCATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAG
GTGCAGTGCCTTCAGCCGCTACCCCGACCATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAG
-----
-TGCAGTGCCTTCAGCCGCTACCCCGACCATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAG
-TGCAGTGCCTTCAGCCGCTACCCCGACCATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAG
-----
---CAGTGCCTTCAGCCGCTACCCCGACCATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAG
GTGCAGTGCCTTCAGCCGCTACCCCGACCATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAG
-----
GTGCAGTGCCTTCAGCCGCTACCCCGACCATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAG
GTGCAGTGCCTTCAGCCGCTACCCCGACCATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAG
GTGCAGTGCCTTCAGCCGCTACCCCGACCATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAG
-TGCAGTGCCTTCAGCCGCTACCCCGACCATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAG
-----
-----TCCGCCATGCCCGAAG

GCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCGAGGTGAAGTT
GCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCG-----T
-----
GCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAG-----TT
GCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCGAGGTGAAGTT
-----
GCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGAC-----GGCAACTACAAGTT
GCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGAC-----
-----CATCAAGGACGACGGCAACTACAAGACCCGCGCGCA-----AGTT
GCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACC-----
GCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCGAGGT-----T
GCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCGAGG-----
GCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCGAGGTGAAGTT
-----TGAAGTT
-----TGAAGTT
GNTACGTCCAGGAGCGCACCATNTTCTTCAAGGACGACGGCAANTACAAGACCCGCGCGCA-----AGTT

```


도면5c

CGAGGGCGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAG	SEQ ID NO:
CGAGGGCGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAG	SEQ ID NO:
-GAGGGCGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAG	SEQ ID NO:
CGAGGGCGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAG	SEQ ID NO:
CGAGGGCGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAA	SEQ ID NO:
-----ACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAA	SEQ ID NO:
CGAGGGCGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAA	SEQ ID NO:
-----ACCGCATCGAGCTGAAGGGCATCGACTTCAA	SEQ ID NO:
CGAGGGCGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCA	SEQ ID NO:
-----TGGTGAACCGCATCGAGCTGAAGGGCATCGA	SEQ ID NO:
CGAGGGCGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCA	SEQ ID NO:
-----GCGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAA	SEQ ID NO:
CGAGGGCGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCGA	SEQ ID NO:
CGAGGGCGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCGAC	SEQ ID NO:
CGAGGGCGACACCCCTGGTGAACCGCATCGAGCTGAAGGG	SEQ ID NO:
CGAGGGCGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACT	SEQ ID NO:

도면6

egfp 부위 3

CACCTTGACCTACGGCGTGCAGTGTTCAGCGCTACCGGACACATCAAGCAGCAGCAGCTTCTTCAAGTCCGCCATGCCGAGGCTACGCCAGAGCGCACCA
 CACCTTGACCTACG-----
 CACCTTGACCTACGGCGTGCAGTGTTCAGCGCTACCGGACACATCAAGCAGCAGCAGCTTCTTCAAGTCCGCCATGCCGAGGCTACGCCAGAGCGCACCA
 CACCTTGACCT-----CAGCGCGCTACCGGACACATCAAGCAGCAGCAGCTTCTTCAAGTCCGCCATGCCGAGGCTACGCCAGAGCGCACCA

egfp 부위 2

TCTTCTTCAAGCAGCAGCACTACAGACCCGCGCGAGGTGGAAGTTCCGAGGGGACACCCCTGGTG
 -----TGAAGTTCCGAGGGGACACCCCTGGT
 TCTTCTTCAAGCAGCAGCACTACAGACCCGCGCGAGG-----GGACACCCCTGG
 TCTTCTTCAAGCAGCAGCACTACAGACCCGCGCGAGG-----CGACACCCCTGGT

CCR5

TTATTATACATCGGAGCCCTGCCAAAAATCAATGAGCAAAATCGAGCCGCTCTGCTTACTACTGCTTTCATCTTGGTTTGTGGCAACA
 TTATTATACATCGGAGCCCTGCCAAAAATCAATGAGCAAAATCGAGCCGCTCTC-----GCTCTACTACTGCTTTCATCTTGGTTTGTGGCAACA
 TTATTATACATCGGAGCCCTGCCAAAAATCAATGAGCAAAATCGAGCCGCTC-----A

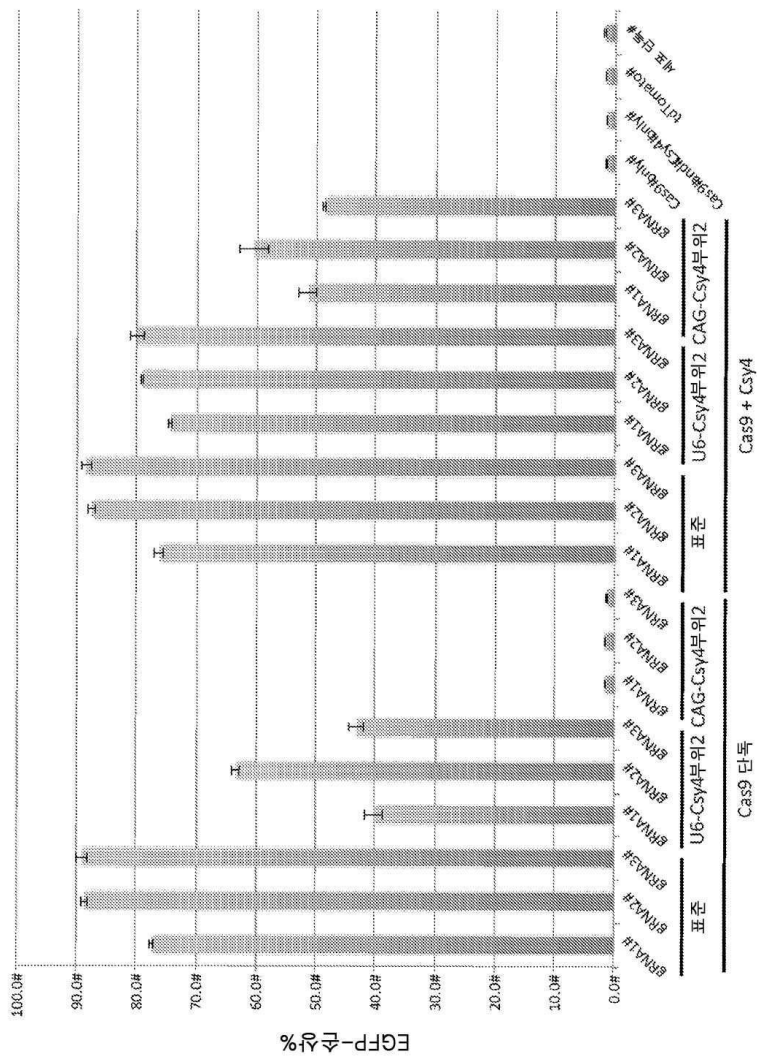
TGCTGGTCACTCCTCATCTCATAACTGCAAAAGGCTGAGAGCAT
 TGCTGGTCACTCT
 TGCTGGTCACTCTC

SEQ ID NO:
 SEQ ID NO:
 SEQ ID NO:
 SEQ ID NO:

SEQ ID NO:
 SEQ ID NO:
 SEQ ID NO:
 SEQ ID NO:

SEQ ID NO:
 SEQ ID NO:
 SEQ ID NO:
 SEQ ID NO:

도면7



서열 목록

SEQUENCE LISTING

<110> THE GENERAL HOSPITAL CORPORATION

<120> MULTIPLEX GUIDE RNAS

<130> 40174-0012W01

<140> PCT/US14/56416

<141> 2014-09-18

<150> PCT/US2014/035162

<151> 2014-03-23

<150> 14/211,117

<151> 2014-03-14

<150> PCT/US2014/029068

<151> 2014-03-14

<150> PCT/US2014/028630

<151> 2014-03-14

<150> PCT/US2014/029304

<151> 2014-03-14

<150> 61/921,007

<151> 2013-12-26

<160> 54

<170> PatentIn version 3.5

<210> 1

<211> 20

<212

> DNA

<213> Artificial Sequence

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

<400> 1

gttcactgcc gtataggcag

20

<210> 2

<211> 28

<212> DNA

<213> Artificial Sequence

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

<400> 2

gttcactgcc gtataggcag ctaagaaa

28

<210> 3

<211> 4

<212> PRT

<213> Artificial Sequence

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

peptide

<400> 3

Gly Gly Gly Ser

1

<210> 4

<211> 342

<212> RNA

<213> Artificial Sequence

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

polynucleotide

<220><221> modified_base

<222> (1)..(20)

<223> a, c, u or g and this region may encompass 17-20 nucleotides

wherein some positions may be absent

<220><221> modified_base

<222> (43)..(342)

<223> a, c, u or g and this region may encompass 0-300 nucleotides

wherein some positions may be absent

<400> 4

nnnnnnnnnn nnnnnnnnnn guuuuagagc uaugcuguuu ugnnnnnnnnn nnnnnnnnnn	60
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	120
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	240
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nn	342

<210> 5

<211> 32

<212> RNA

<213> Artificial Sequence

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

oligonucleotide

<220><221> modified_base

<222> (1)..(20)

<223> a, c, u or g and this region may encompass 17-20 nucleotides

wherein some positions may be absent

<400> 5

nnnnnnnnnn nnnnnnnnnn guuuuagagc ua	32
-------------------------------------	----

<210> 6
 <211> 42
 <212> RNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <220><221> modified_base
 <222> (1)..(20)

<223> a, c, u or g and this region may encompass 17-20 nucleotides
 wherein some positions may be absent

<400> 6
 nnnnnnnnnn nnnnnnnnnn guuuuagagc uaugcuguuu ug 42

<210> 7
 <211> 36
 <212> RNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <220><221> modified_base
 <222> (1)..(20)
 <223> a, c, u or g and this region may encompass 17-20 nucleotides
 wherein some positions may be absent
 <400> 7

nnnnnnnnnn nnnnnnnnnn guuuuagagc uaugcu 36

<210> 8
 <211> 362
 <212> RNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 polynucleotide
 <220><221> modified_base
 <222> (1)..(20)
 <223> a, c, u or g and this region may encompass 17-20 nucleotides

wherein some positions may be absent

<220><221> modified_base

<222> (63)..(362)

<223> a, c, u or g and this region may encompass 0-300 nucleotides

wherein some positions may be absent

<400> 8

nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc	60
cgnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	120
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	240
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	360
nn	362

<210> 9

<211> 375

<212> RNA

<213> Artificial Sequence

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

polynucleotide

<220><221> modified_base

<222> (1)..(20)

<223> a, c, u or g and this region may encompass 17-20 nucleotides

wherein some positions may be absent

<220><221> modified_base

<222> (76)..(375)

<223> a, c, u or g and this region may encompass 0-300 nucleotides

wherein some positions may be absent

<400> 9

nnnnnnnnnn nnnnnnnnnn guuuuagagc uaugcugaaa agcauagcaa guuaaaauaa	60
ggcuaguccg uuauccccnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	120
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	240

nnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn 300
 nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn 360
 nnnnnnnnnnn nnnnnnn 375

<210> 10

<211> 387

<212> RNA

<213> Artificial Sequence

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

polynucleotide

<220><221> modified_base

<222> (1)..(20)

<223> a, c, u or g and this region may encompass 17-20 nucleotides

wherein some positions may be absent

<220><221> modified_base

<222> (88)..(387)

<223> a, c, u or g and this region may encompass 0-300 nucleotides

wherein some positions may be absent

<400> 10

nnnnnnnnnn nnnnnnnnnnn guuuuagagc uaugcuguuu uggaacacaaa acagcauagc 60

aaguuaaaau aaggcuaguc cguaucnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn 120

nnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn 180

nnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn 240

nnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn 300

nnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn 360

nnnnnnnnnn nnnnnnnnnnn nnnnnnnnn 387

<210> 11

<211> 396

<212> RNA

<213> Artificial Sequence

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

polynucleotide

<220><221> modified_base

<222> (1)..(20)

<223> a, c, u or g and this region may encompass 17-20 nucleotides
wherein some positions may be absent

<220><221> modified_base

<222> (97)..(396)

<223> a, c, u or g and this region may encompass 0-300 nucleotides
wherein some positions may be absent

<400> 11

nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc	60
cguuaucaac uugaaaaagu ggcaccgagu cggugcnnnn nnnnnnnnnn nnnnnnnnnn	120
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	240
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnn	396

<210> 12

<211> 96

<212> RNA

<213> Artificial Sequence

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

<220><221> modified_base

<222> (1)..(20)

<223> a, c, u or g and this region may encompass 17-20 nucleotides
wherein some positions may be absent

<400> 12

nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc	60
cguuaucaac uugaaaaagu ggcaccgagu cggugc	96

<210> 13

<211> 106

<212> RNA

<213> Artificial Sequence

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

polynucleotide

<220><221> modified_base

<222> (1)..(20)

<223> a, c, u or g and this region may encompass 17-20 nucleotides
wherein some positions may be absent

<400> 13

nnnnnnnnnn nnnnnnnnnn guuuuagagc uaugcuggaa acagcauagc aaguuuuuuu 60

aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugc 106

<210> 14

<211> 106

<212> RNA

<213> Artificial Sequence

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

polynucleotide

<220><221> modified_base

<222> (1)..(20)

<223> a, c, u or g and this region may encompass 17-20 nucleotides
wherein some positions may be absent

<400> 14

nnnnnnnnnn nnnnnnnnnn guuuuagagc uaugcuggaa acagcauagc aaguuuuuuu 60

aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugc 106

<210> 15

<211> 36

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<220><223> Description of Combined DNA/RNA Molecule: Synthetic

oligonucleotide

<220><221> modified_base

<222> (1)..(20)

<223> a, c, u or g and this region may encompass 17-20 nucleotides
wherein some positions may be absent

<400> 15

nnnnnnnnnn nnnnnnnnnn gtttagagc tagaaa

36

<210> 16

<211> 60

<212> DNA

<213> Artificial Sequence

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

<400> 16

tagcaagtta aaataaggct agtccgttat caacttgaaa aagtggcacc gagtcggtgc

60

<210> 17

<211> 5

<212> PRT

<213> Artificial Sequence

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

<400> 17

Gly Gly Gly Gly Ser

1 5

<210> 18

<211> 13

<212> PRT

<213> Artificial Sequence

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

peptide

<400> 18

Asp Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Ser Leu

1 5 10

<210> 19

<211> 17

<212> PRT

<213> Artificial Sequence

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

peptide

<400> 19

Pro Lys Lys Lys Arg Lys Val Glu Asp Pro Lys Lys Lys Arg Lys Val

1 5 10 15

Asp

<210> 20

<211> 1401

<212> PRT

<213> Artificial Sequence

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

polypeptide

<400> 20

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160
 Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
 165 170 175
 Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
 180 185 190

 Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
 195 200 205
 Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
 210 215 220
 Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
 225 230 235 240
 Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
 245 250 255

 Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
 260 265 270
 Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
 275 280 285
 Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
 290 295 300
 Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
 305 310 315 320

 Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
 325 330 335
 Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
 340 345 350
 Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
 355 360 365
 Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
 370 375 380

 Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
 385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
 405 410 415
 Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
 420 425 430
 Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
 435 440 445

 Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
 450 455 460
 Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
 465 470 475 480
 Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
 485 490 495
 Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
 500 505 510

 Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
 515 520 525
 Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
 530 535 540
 Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
 545 550 555 560
 Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
 565 570 575

 Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
 580 585 590
 Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
 595 600 605
 Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
 610 615 620
 Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
 625 630 635 640

 His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645	650	655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp		
660	665	670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe		
675	680	685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe		
690	695	700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu		
705	710	715
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly		
725	730	735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly		
740	745	750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln		
755	760	765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile		
770	775	780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro		
785	790	795
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu		
805	810	815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg		
820	825	830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys		
835	840	845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg		
850	855	860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys		
865	870	875
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys		
885	890	895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
 900 905 910
 Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
 915 920 925
 Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
 930 935 940
 Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
 945 950 955 960

 Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
 965 970 975
 Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
 980 985 990
 Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
 995 1000 1005
 Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
 1010 1015 1020

 Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
 1025 1030 1035
 Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
 1040 1045 1050
 Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
 1055 1060 1065
 Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
 1070 1075 1080

 Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
 1085 1090 1095
 Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
 1100 1105 1110
 Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
 1115 1120 1125
 Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130	1135	1140
Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys		
1145	1150	1155
Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser		
1160	1165	1170
Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys		
1175	1180	1185
Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu		
1190	1195	1200
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly		
1205	1210	1215
Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val		
1220	1225	1230
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser		
1235	1240	1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys		
1250	1255	1260
His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys		
1265	1270	1275
Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala		
1280	1285	1290
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn		
1295	1300	1305
Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala		
1310	1315	1320
Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser		
1325	1330	1335
Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr		
1340	1345	1350
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp		
1355	1360	1365

Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Ser Leu Gly Ser Gly
1370 1375 1380

Ser Pro Lys Lys Lys Arg Lys Val Glu Asp Pro Lys Lys Lys Arg
1385 1390 1395

Lys Val Asp
1400

<210> 21

<211> 1368

<212> PRT

<213> Artificial Sequence

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

polypeptide

<400> 21

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400
 Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

 405 410 415
 Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
 420 425 430
 Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
 435 440 445
 Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
 450 455 460
 Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

 465 470 475 480
 Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
 485 490 495
 Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
 500 505 510
 Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
 515 520 525
 Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

 530 535 540
 Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
 545 550 555 560
 Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
 565 570 575
 Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
 580 585 590
 Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

 595 600 605
 Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
 610 615 620
 Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
 625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895
 Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
 900 905 910
 Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

 915 920 925
 Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
 930 935 940
 Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
 945 950 955 960
 Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
 965 970 975
 Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

 980 985 990
 Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
 995 1000 1005
 Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
 1010 1015 1020
 Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
 1025 1030 1035
 Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

 1040 1045 1050
 Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
 1055 1060 1065
 Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
 1070 1075 1080
 Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
 1085 1090 1095
 Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

 1100 1105 1110
 Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
 1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val		
1130	1135	1140
Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys		
1145	1150	1155
Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser		
1160	1165	1170
Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys		
1175	1180	1185
Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu		
1190	1195	1200
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly		
1205	1210	1215
Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val		
1220	1225	1230
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser		
1235	1240	1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys		
1250	1255	1260
His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys		
1265	1270	1275
Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala		
1280	1285	1290
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn		
1295	1300	1305
Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala		
1310	1315	1320
Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser		
1325	1330	1335
Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr		
1340	1345	1350
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp		

<p>1355</p> <p><210> 22</p> <p><211> 4</p> <p><212> PRT</p> <p><213> Artificial Sequence</p> <p><220><223> Description of Artificial Sequence: Synthetic peptide</p> <p><400> 22</p> <p>Gly Gly Ser Gly</p> <p>1</p> <p><210> 23</p> <p><211> 12</p> <p><212> PRT</p> <p><213> Artificial Sequence</p> <p><220><223> Description of Artificial Sequence: Synthetic peptide</p> <p><400> 23</p> <p>Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala</p> <p>1 5 10</p> <p><210> 24</p> <p><211> 16</p> <p><212> PRT</p> <p><213> Artificial Sequence</p> <p><220><223> Description of Artificial Sequence: Synthetic peptide</p> <p><400> 24</p> <p>Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser</p> <p>1 5 10 15</p> <p><210> 25</p> <p><211> 21</p> <p><212> PRT</p> <p><213> Artificial Sequence</p> <p><220><223> Description of Artificial Sequence: Synthetic</p>	<p>1360</p>	<p>1365</p>
--	-------------	-------------

peptide

<400> 25

Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Gly

1 5 10 15

Gly Ser Gly Gly Ser

20

<210> 26

<211> 24

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 26

gccgaggtga agttcgaggg cgac 24

<210> 27

<211> 23

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 27

cctacggcgt gcagtgttc agc 23

<210> 28

<211> 332

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 28

acgtaaaccg ccacaagttc agcgtgtccg gcgagggcga gggcgatgcc acctacggca 60

agctgacct gaagttcatc tgcaccaccg gcaagctgcc cgtgccctgg cccaccctcg 120

tgaccacct gacctacggc gtgcagtgtc tcagccgcta ccccgaccac atgaagcagc 180

acgacttctt caagtcgcc atgcccgaag gctacgtcca ggagcgcacc atcttcttca 240

aggacgacgg caactacaag acccgcgccg aggtgaagtt cgagggcgac accctggtga 300
accgcatcga gctgaagggc atcgattca ag 332
<210> 29
<211> 295
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polynucleotide
<400> 29
gtccggcgag ggcgagggcg atgccacctt cggcaagctg accctgaagt tcatctgcac 60
caccggcaag ctgcccgtgc cctggcccac cctcgtgacc accctgacct acggcgtgca 120
gtgcttcagc cgtacccccg accacatgaa gcagcacgac ttcttcaagt ccgccaatgcc 180

cgaaggctac gtccaggagc gcaccatctt cttaaggac gacggcaact acaagaccgc 240
cgtcgagggc gacacctgg tgaaccgat cgagctgaag ggcatcgact tcaag 295
<210> 30
<211> 158
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polynucleotide
<400> 30
tccggcgagg ggcgagggcg tgccacctac ggcaagctga ccctgaagtt catctgcacc 60
accggcaagc tgcccgtgcc ctggcccacc ctcgtgacca ccctgacgag ggcgacaccc 120
tggtgaaccg catcgagctg aaggcgatcg acttcaag 158

<210> 31
<211> 285
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polynucleotide
<400> 31
tccggcgagg ggcgagggcg tgccacctac ggcaagctga ccctgaagtt catctgcacc 60
accggcaagc tgcccgtgcc ctggcccacc ctcgtgaaca ccctgacct cgtgcagtgc 120

ttcagccgct accccgacca catgaagcag cagcattct tcaagtccgc catgcccga 180
 ggctacgtcc aggagcgcac catcttcttc aaggacgacg gcaactacaa gttcgagggc 240
 gacaccctgg tgaaccgcat cgagctgaag ggcatcgact tcaag 285

<210> 32

<211> 301

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 32

ccggcgaggg cgagggcgat gccacctacg gcaagctgac cctgaagttc atctgcacca 60
 ccggcaagct gccctgtccc tggcccaccc tcgtgaccac cctgacctac gtgcagtgt 120
 tcagccgcta ccccgaccac atgaagcagc acgacttctt caagtcgcc atgccgaag 180
 gctacgtcca ggagcgacc atcttcttca aggacgacgg caactacaag acccgcgccg 240
 aggtgaagtt cgagggcgac accctgggtga accgcatcga gctgaagggc atcgacttca 300

a 301

<210> 33

<211> 156

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 33

tccggcgagg gcgagggcga tgccacctac ggcaagctga ccctgaagtt catctgcacc 60
 accggcaagc tgccctgtcc ctggcccacc ctctgacca ccctgacct cggacacct 120
 ggtgaaccgc atcgagctga agggcatcga cttaa 156

<210> 34

<211> 197

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 34

glaaacggcc acaagttcag cgtgcagtgc ttcagccgct accccgacca catgaagcag	60
cacgacttct tcaagtccgc catgcccgaaggctacgtcc aggagcgacac catctttcttc	120
aaggacgacg gcaactacaa gacggcaact acaagttcga gggcgacacc ctggtgaacc	180
gcatcgagct gaagggc	197

<210> 35

<211> 283

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 35

ccacaagttc agcgtgtccg gcgagggcga gggcgatgcc acctacggca agctgaccct	60
gaagttcattc tgcaccaccg gcaagctgcc cgtgccctgg cccaccctcg tgaccaccct	120
gacctacggc gtgcagtgtc ttcagccgcta ccccgaccac atgaagcagc acgacttctt	180
caagtccgcc atgccgaag gctacgtcca ggagcgacac atctttcttca aggacgacgg	240
caactacaag acaccgcatt gagctgaagg gcatcgactt caa	283

<210> 36

<211> 175

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 36

ggccacaagt tcagcgtgtc cggcgagggc gagggcgatg ccacctacgg caagctgacc	60
ctgaagttca tctgcaccac cggcaccattc aaggacgacg gcaactacaa gacccgcgcc	120
gaagttcgag ggcgacaccc tgggtgaaccg catcgagctg aaggcgatcg acttc	175

<210> 37

<211> 291

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 37

taaacggcca caagttcagc gtgtccggcg agggcgaggg cgatgccacc tacggcaagc	60
tgacctgaa gttcatctgc accaccggca agctgcccgt gccctggccc accctcgtga	120
ccacctgac ctacggcgtg cagtgttca gccgtaccc cgaccacatg aagcagcacg	180
actttctcaa gtccgccatg cccgaaggct acgtccagga gcgcaccatc ttcttcaagg	240
acgacggcaa ctacaagacc tggatgaacc catcgagctg aaggcatcg a	291

<210> 38

<211> 314

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 38

cacaagtcca gcgtgtccgg cgagggcgag ggcatgccca cctacggcaa gctgacctg	60
aagttcatct gcaccaccgg caagctgccc gtgccctggc ccacctctgt gaccacctg	120
acctacggcg tgcagtgtt cagccgtac cccgaccaca tgaagcagca cgacttttc	180
aagtcgccca tgccgaagg ctacgtccag gagcgacca ttttttcaa ggacgacggc	240
aactacaaga cccgcgccga ggttcgaggg cgacacctg gtgaaccgca tcgagctgaa	300
gggcatcgac ttca	314

<210> 39

<211> 316

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 39

taaacggcca caagttcagc gtgtccggcg agggcgaggg cgatgccacc tacggcaagc	60
tgacctgaa gttcatctgc accaccggca agctgcccgt gccctggccc accctcgtga	120
ccacctgac ctacggcgtg cagtgttca gccgtaccc cgaccacatg aagcagcacg	180
actttctcaa gtccgccatg cccgaaggct acgtccagga gcgcaccatc ttcttcaagg	240
acgacggcaa ctacaagacc cgcgccgagg gcgacacct ggtgaaccgc atcgagctga	300
aggcatcga cttcaa	316

<210> 40

<211> 318

<212> DNA

<213> Artificial Sequence

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

<400> 40

aaacggccac aagttcagcg tgtccggcga gggcgagggc gatgccacct acggcaagct	60
gacctgaag ttcatctgca ccaccggcaa gctgcccggtg ccctggccca ccctcgtgac	120
cacctgacc tacgtgcagt gttcagccg ctaccccgac cacatgaagc agcacgactt	180
cttcaagtcc gccatgcccg aaggctacgt ccaggagcgc accatcttct tcaaggacga	240
cggcaactac aagaccgcg ccgaggtgaa gtctgagggc gacaccctgg tgaaccgcat	300
cgagctgaag ggcacga	318

<210> 41

<211> 185

<212> DNA

<213>

Artificial Sequence

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

<400> 41

acggccacaa gttagcgtg tccggcgagg gcgagggcga tgccacctac ggcaagctga	60
ccctgaagtt catctgcacc accggcaagc tgcccgtgcc ctggcccacc ctctgacca	120
ccctgaccta cgtaagttc gagggcgaca ccctggtgaa ccgcatcgag ctgaagggca	180
tcgac	185

<210> 42

<211> 180

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 42

aaacggccac aagttcagcg tgtccggcga gggcgagggc gatgccacct acggcaagct	60
---	----

gaccctgaag ttcatctgca ccaccggcaa gctgcccggtg ccctggccca ccctcgtgac 120
caccctgacc tacgtgaagt tcgagggcga caccctggtg aaccgcatcg agctgaaggg 180
<210> 43
<211> 242
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polynucleotide
<220><221> modified_base
<222> (132)..(132)
<223> a, c, t, g, unknown or other
<220><221> modified_base
<222> (153)..(153)
<223> a, c, t, g, unknown or other
<220><221> modified_base
<222> (174)..(174)
<223> a, c, t, g, unknown or other
<400> 43
aaacggccac aagttcagcg tgtccggcga gggcgagggc gatgccacct acggcaagct 60
gaccctgaag ttcatctgca ccaccggcaa gctgcccggtg ccctggccca cccttcgcc 120
atgccgaag gntacgtcca ggagcgacc atntttctca aggacgacgg caantacaag 180
acccgcgccg aagttcgagg gcgacacct ggtgaaccgc atcgagctga agggcatcga 240
ct 242
<210> 44
<211> 175
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polynucleotide
<400> 44
caccctgacc tacggcgtgc agtgettcag ccgctacccc gaccacatga agcagcacga 60
cttcttcaag tccgccatgc ccgaaggcta cgtccaggag cgcaccatct tcttcaagga 120
cgacggcaac tacaagacct gcgccgaggt gaagttcgag ggcgacaccc tgggtg 175

<210> 45

<211> 39

<212> DNA

<213> Artificial Sequence

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

<400> 45

caccctgacc tacgtgaagt tcgagggcga caccctggt 39

<210> 46

<211> 161

<212> DNA

<213> Artificial Sequence

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

<400> 46

caccctgacc tacggcgtgc agtgcttcag ccgctacccc gaccacatga agcagcacga 60

cttcttcaag tccgccatgc ccgaaggcta cgtccaggag cgcaccatct tcttcaagga 120

cgacggcaac tacaagacc gcgccgaggg cgacaccctg g 161

<210> 47

<211> 145

<212> DNA

<213>

> Artificial Sequence

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

<400> 47

caccctgacc tcagccgcta ccccgaccac atgaagcagc acgacttctt caagtccgcc 60

atgccgaag gctacgtcca ggagcgcacc atcttcttca aggacgacgg caactacaag 120

accgcgcgcg aggcgacacc ctggt 145

<210> 48

<211> 153

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 48

ttattataca tcggagccct gccaaaaaat caatgtgaag caaatcgag cccgcctcct 60

gcctccgctc tactcactgg tggtcatctt tggttttgtg ggcaacatgc tggtcaccc 120

catcctgata aactgcaaaa ggctgaagag cat 153

<210> 49

<211> 113

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 49

ttattataca tcggagccct gccaaaaaat caatgtgaag caaatcgag cccgcctccg 60

ctctactcac tgggtgttcac ctttggtttt gtgggcaaca tgctggtcat cct 113

<210> 50

<211> 70

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 50

ttattataca tcggagccct gccaaaaaat caatgtgaag caaatcgag cccgcacgct 60

ggtcacccctc 70

<210> 51

<211> 62

<212> RNA

<213> Artificial Sequence

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

oligonucleotide

<220><221> modified_base

<222> (1)..(20)

<223> a, c, u or g and this region may encompass 17-20 nucleotides

wherein some positions may be absent

<400> 51

nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cg 62

<210> 52

<211> 54

<212> RNA

<213> Artificial Sequence

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

oligonucleotide

<220><221> modified_base

<222> (1)..(20)

<223> a, c, u or g and this region may encompass 17-20 nucleotides

wherein some positions may be absent

<400> 52

nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuaaaau aagg 54

<210> 53

<211> 28

<212> RNA

<213> Artificial Sequence

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

oligonucleotide

<400> 53

guucacugcc guauaggcag cuaagaaa 28

<210> 54

<211> 20

<212> RNA

<213> Artificial Sequence

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

oligonucleotide

<400> 54

guucacugcc guauaggcag 20