



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2019-0044157
(43) 공개일자 2019년04월30일

(51) 국제특허분류(Int. Cl.)
C12N 9/78 (2006.01) C12N 15/10 (2017.01)
C12N 15/113 (2010.01) C12N 9/22 (2006.01)
C12N 9/96 (2006.01)
(52) CPC특허분류
C12N 9/78 (2013.01)
C12N 15/102 (2013.01)
(21) 출원번호 10-2017-0136195
(22) 출원일자 2017년10월20일
심사청구일자 없음

(71) 출원인
경상대학교산학협력단
경상남도 진주시 진주대로 501 (가좌동)
(72) 발명자
김재연
경상남도 진주시 새평거로 75, 406동 804호
트란티밀
경상남도 진주시 진주대로 501
부반티엔
경상남도 진주시 진주대로 501
(74) 대리인
최규환

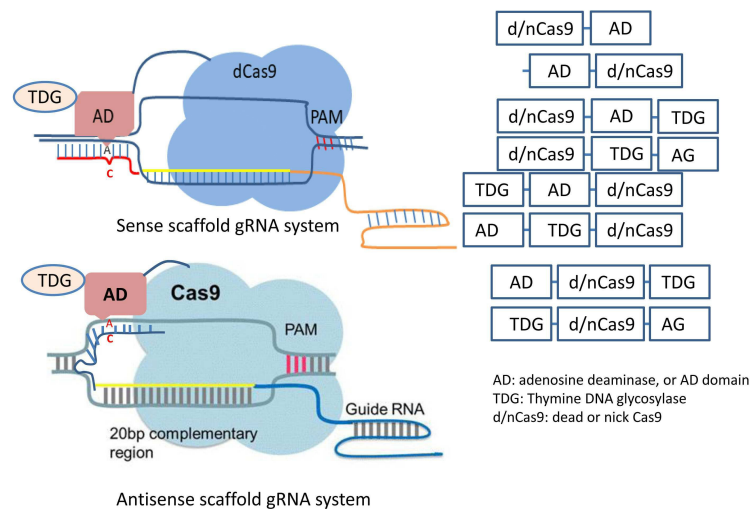
전체 청구항 수 : 총 5 항

(54) 발명의 명칭 아데닌 또는 아데노신 탈아미노효소를 유효성분으로 포함하는 단일 염기 편집용 조성물 및 이의 용도

(57) 요약

본 발명은 아데닌 또는 아데노신 탈아미노효소를 유효성분으로 포함하는 단일 염기 편집용 조성물 및 이의 용도에 관한 것으로, 본 발명의 염기 편집 방법은 DNA 염기 하나만을 표적으로 하여 특정 염기 하나를 다른 염기로 바로 전환하기 때문에 DNA 절단 없이도 교정할 수 있는 장점이 있어 돌연변이의 교정 및 암세포 치료제와 같은 다양한 방면에 활용가능성이 높을 것으로 기대된다.

대표도 - 도1



(52) CPC특허분류

C12N 15/113 (2013.01)

C12N 9/22 (2013.01)

C12N 9/96 (2013.01)

C12Y 305/04004 (2013.01)

C12N 2310/10 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 2017R1A4A1015515

부처명 과학기술정보통신부

연구관리전문기관 한국연구재단

연구사업명 기초연구

연구과제명 상동성재조합기반 식물유전체교정 원천기술 개발

기 여 율 1/2

주관기관 경상대학교

연구기간 2017.06.01 ~ 2020.02.28

이 발명을 지원한 국가연구개발사업

과제고유번호 PJ01184301

부처명 농촌진흥청

연구관리전문기관 시스템합성농생명공학사업단

연구사업명 차세대바이오그린21 사업

연구과제명 Supracellular 시스템합성생물학 기법을 이용한 자당이동경로 최적화 모듈 개발

기 여 율 1/2

주관기관 경상대학교

연구기간 2016.01.04 ~ 2017.12.31

명세서

청구범위

청구항 1

아데닌 탈아미노효소(deaminase) 또는 아데노신 탈아미노효소;와 Cas9(CRISPR associated protein 9) 단백질을 또는 이의 기능적 유사체;로 이루어진 융합 단백질.

청구항 2

아데닌 탈아미노효소(deaminase) 또는 아데노신 탈아미노효소; 티민 DNA 글리코실라제(thymine DNA glycosylase); 및 Cas9(CRISPR associated protein 9) 단백질 또는 이의 기능적 유사체;로 이루어진 융합 단백질.

청구항 3

제1항 또는 제2항의 융합 단백질;과 상기 융합 단백질의 Cas9 단백질 또는 이의 유사체 부분에 결합된 가이드 RNA;를 유효성분으로 포함하는 단일 염기 편집용 조성물.

청구항 4

제3항에 있어서, 상기 가이드 RNA는 PAM(protospacer adjacent motif) 자리와 인접한 10~20bp의 염기 서열과 상보적인 서열을 포함하며, 교정하고자 하는 표적 아데닌 자리와 미스매치가 생기게 하는 RNA 서열을 추가로 포함하는 것을 특징으로 하는 단일 염기 편집용 조성물.

청구항 5

제3항의 단일 염기 편집용 조성물을 핵산 분자와 접촉시키는 단계를 포함하는 단일 염기 편집 방법.

발명의 설명

기술 분야

[0001] 본 발명은 아데닌 또는 아데노신 탈아미노효소를 유효성분으로 포함하는 단일 염기 편집용 조성물 및 이의 용도에 관한 것으로, 더욱 구체적으로는, 데옥시-아데닌 또는 데옥시-아데노신(deoxy-adenine/deoxy-adenosine)을 데옥시-구아닌 또는 데옥시-구아노신(deoxy-guanine/deoxy-guanosine)으로 또는 데옥시-티민 또는 데옥시-티미딘(deoxy-thymine/deoxy-thymidine)을 데옥시-시토신 또는 데옥시-시티딘(deoxy-cytosine/deoxy-cytidine)으로 바꾸는 염기편집(base editing) 기술에 관한 것이다.

배경 기술

[0002] 현재의 유전체 교정(genome editing, genome engineering)은 교정하고자 하는 특정 유전자의 특정위치에 이중나선 절단(double strand break)을 도입하여 DNA 수선(DNA repair) 과정에서 유도되는 불완전 수선에 의한 삽입-결실(insertion-deletion; indel) 돌연변이를 유도시키는 NHEJ(non-homologous end joining) 기작에 의한 방법, 주변에 상동성 DNA가 존재할 때 일어나는 상동 재조합(homologous recombination)에 의한 방법이 있고 DNA의 이중나선 절단없이 또는 단일나선의 절단만 유도하되 시티딘 탈아미노효소(cytidine deaminase)를 사용하여 시토신(C)를 티민(T)로 바꾸는 시티딘(cytidine) 염기교정(base editing) 기술이 존재한다. 대부분의 유전병 돌연변이는 점돌연변이(point mutation)에 일어나며 자연상태에서 시티딘 탈아미노효소는 매우 빈번하게 일어나고 실제 과반의 점돌연변이는 C→T 돌연변이에 의해 일어난다고 알려졌다. 그러므로 유전병의 치료를 위해서나 좀 더 자유로운 염기교정 기술의 개발을 위해 T/A→C/G 염기교정 기술을 개발하는 것은 매우 유용한 유전자교정 기술의 확보 차원에서 중요한 이슈이다.

[0003] 자연계에는 현재까지 DNA 이중나선의 T/A를 기질로 인식하여 C/G로 바꾸는 효소나 공정은 알려져 있지 않다. 하지만 RNA의 아데노신(Adenosine, rA)을 탈아미노화시켜 이노신(Inosine, rI)으로 만들 수 있는 효소로 아데노신 탈아미노효소 패밀리(adenosine deaminase family)가 알려져 있다. rI은 전사, 번역, 수선 또는 스플라이싱

(splicing) 등에서 RNA의 구아닌(Guanine, rG)과 유사하게 기능하므로 RNA 서열이 변할 경우 단백질의 아미노산의 변화를 야기한다. 이 패밀리 구성원으로 tRNA에 작용하는 아데닌 탈아미노효소(adenine deaminase, ADAT)와 이중나선 RNA에 작용하는 아데노신 탈아미노효소(adenosine deaminase, ADAR)가 있다. ADAT의 경우 원핵생물과 효모 및 상위의 진핵생물에 공히 발견되나 ADAR은 동물에서만 발견된다. 또한 유리(free) 아데노신을 기질로 사용하는 아데노신 탈아미노효소(ADfA)가 있으나, 이는 ADAR 또는 ADAT와 유전자 서열상에서는 진화적으로 관련이 없다고 알려졌으나 촉매부위(catalytic site)에서 아연 이온 또는 글루탐산과 같은 핵심적인 아미노산이 공통적으로 발견되어지며 유사한 구조를 나타낸다. 최근 ADAR은 RNA/DNA 잡종(hybrid)에 작동하여 RNA/RNA에 비해 상대적으로 낮은 활성을 가지고 아데노신을 탈아미노화시킬 수 있다는 사실이 보고되었다.

[0004] 한편, 한국공개특허 제2015-0101446호에는 '표적 DNA에 특이적인 가이드 RNA 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는, 표적 DNA를 절단하기 위한 조성물 및 이의 용도'가 개시되어 있고, 한국공개특허 제2017-0065032호에는 '아데노신 디아미네이즈-2(ADA2)의 조성물, 이의 변이체 및 이를 사용하는 방법'이 개시되어 있으나, 본 발명의 아데닌 또는 아데노신 탈아미노효소를 유효성분으로 포함하는 단일 염기 편집용 조성물 및 이의 용도에 대해서는 기재된 바 없다.

발명의 내용

해결하려는 과제

[0005] 본 발명은 상기와 같은 요구에 의해 도출된 것으로서, 본 발명자들은 아데닌 또는 아데노신 탈아미노효소와 CRISPR/Cas9 시스템을 이용하여, 아데닌(A)을 구아닌(G)으로 직접적으로 전환할 수 있는 염기 편집 기술을 개발하고자, 촉매 활성 도메인 또는 전장의 아데닌/아데노신 탈아미노효소를 Cas9 또는 이의 변이체에 융합하고, 상기 융합된 단백질 및 염기 편집의 표적 서열에 상보적인 서열을 포함하는 가이드 RNA를 포함하는 단일 염기 편집 기술을 완성하였다.

과제의 해결 수단

[0006] 상기 과제를 해결하기 위해, 본 발명은 아데닌 탈아미노효소(deaminase) 또는 아데노신 탈아미노효소;와 Cas9(CRISPR associated protein 9) 단백질 또는 이의 기능적 유사체;로 이루어진 융합 단백질을 제공한다.

[0007] 또한, 본 발명은 아데닌 탈아미노효소(deaminase) 또는 아데노신 탈아미노효소; 티민 DNA 글리코실라제(thymine DNA glycosylase); 및 Cas9 단백질 또는 이의 기능적 유사체;로 이루어진 융합 단백질을 제공한다.

[0008] 또한, 본 발명은 상기 융합 단백질;과 상기 융합 단백질의 Cas9 단백질 또는 이의 유사체 부분에 결합된 가이드 RNA;를 유효성분으로 포함하는 단일 염기 편집용 조성물을 제공한다.

[0009] 또한, 본 발명은 상기 융합 단백질;과 상기 융합 단백질의 Cas9 단백질 또는 이의 유사체 부분에 결합된 가이드 RNA;를 유효성분으로 포함하는 단일 염기 편집용 조성물을 핵산 분자와 접촉시키는 단계를 포함하는 단일 염기 편집 방법을 제공한다.

[0010] 또한, 본 발명은 상기 융합 단백질과 상기 융합 단백질의 단일 또는 이중나선 DNA 상의 아데노신 뉴클레오타이드를 인식하여 보다 효과적인 탈아미노반응을 일으킬 수 있는 특이 돌연변이의 사용을 포함한다.

발명의 효과

[0011] 크리스퍼(CRISPR) 유전자 가위는 단일 염기의 문제인 점 돌연변이를 치료하는 데엔 어려움이 있다. 또한, 기존의 유전자 편집 기술은 염기를 바꿀 때 DNA 이중가닥의 절단에 의존하기 때문에, 표적이 되는 돌연변이도 고치지만 원치 않은 삽입/결실이 일어날 수 있는 문제점이 있다. 본 발명의 염기 편집 기법은 DNA 염기 하나만을 표적으로 하여 특정 염기 하나를 다른 염기로 바로 전환하기 때문에 DNA 절단 없이도 교정할 수 있는 장점이 있다. 본 발명의 기법은 유전병의 원인이 되는 돌연변이의 교정 및 암세포 치료제와 같은 다양한 방면에 활용가능성이 높을 것으로 기대된다.

도면의 간단한 설명

[0012] 도 1은 본 발명의 염기 교정 방법의 모식도이다.

도 2는 ScLEU2 표적 DNA 부위(Q180 주변)와 센스 스캐폴드 가이드 RNA(이하 gRNA) 1 결합부위를 나타낸다.

- 도 3은 ScLEU2 표적 DNA 부위(Q17 주변)와 안티센스 스캐폴드 gRNA 1 결합부위를 나타낸다.
- 도 4는 안티센스 스캐폴드 gRNA 1 염기서열과 접합구조를 나타낸다.
- 도 5는 ScLEU2 표적 DNA 부위(Q17 주변)과 안티센스 스캐폴드 gRNA 2 결합부위를 나타낸다.
- 도 6은 안티센스 스캐폴드 gRNA 2 염기서열과 접합구조를 나타낸다.
- 도 7은 d/nCas9 및 dCpf1 기반의 아데노신 염기 편집기(adenosine base editor) 개발 전략을 나타낸다.
- 도 8은 ADAR2-nCas9 염기 편집기와 ADAR2가 결합할 수 있는 RNA 결합도메인이 융합된 센스 스캐폴드 gRNA를 보인다.
- 도 9는 ADAR2-nCas9 염기 편집기와 ADAR2가 결합할 수 있는 RNA 결합도메인이 융합된 안티센스 스캐폴드 gRNA를 보여주며(좌), ADAR2가 결합할 수 있는 RNA 결합도메인이 포함된 안티센스 표적 RNA와 Cas9의 gRNA가 분리된 것을 보여준다(우).
- 도 10은 ADAR2-dLbCpf1 염기 편집기용 ADAR2 RNA 결합도메인을 갖는 센스 스캐폴드 gRNA를 보인다.
- 도 11은 ADAR2-dLbCpf1과 사용될 수 있는 다양한 센스(SE) 또는 안티센스(ASE) 스캐폴드 gRNA 패턴을 보여준다.
- 도 12는 유리(free) 아데닌 탈아미노효소에서 사용되는 촉매 부위의 핵심 아미노산을 보여준다.
- 도 13은 hADAR2d(E488Q)로부터 추가적인 돌연변이인 T490X(X=Q, N, D,E)의 활성을 예측한 것이다.
- 도 14는 hADAR2d-nSpCas9 기반 벡터시스템이다.
- 도 15는 nSpCas9-hADAR2d기반 벡터시스템이다.
- 도 16는 hADAR2d-dSpCas9 기반 벡터시스템이다.
- 도 17은 dSpCas9-hADAR2d기반 벡터시스템이다.

발명을 실시하기 위한 구체적인 내용

- [0013] 본 발명의 목적을 달성하기 위하여, 본 발명은 아데닌 탈아미노효소(deaminase) 또는 아데노신 탈아미노효소;와 Cas9(CRISPR associated protein 9) 단백질 또는 이의 기능적 유사체;로 이루어진 융합 단백질을 제공한다.
- [0014] 본 발명은 또한, 아데닌 탈아미노효소(deaminase) 또는 아데노신 탈아미노효소; 티민 DNA 글리코실라제(thymine DNA glycosylase); 및 Cas9(CRISPR associated protein 9) 단백질 또는 이의 기능적 유사체;로 이루어진 융합 단백질을 제공한다.
- [0015] 본 발명의 융합 단백질은 아데닌 또는 아데노신 탈아미노효소(이하, AD) 또는 AD 및 티민 DNA 글리코실라제를 사용하여 DNA 또는 원핵생물 및 진핵생물의 염색체의 특정 부분의 염기서열 중 A/T->G/C로 바꾸는 단일 염기 편집용 조성물의 구성요소이다.
- [0016] 본 발명의 융합 단백질에 있어서, 상기 아데닌 또는 아데노신 탈아미노효소는 Cas9 단백질 또는 이의 기능적 유사체의 아미노(N)-말단 또는 카복시(C)-말단에 위치할 수 있고, 효소 활성을 최적으로 유지할 수 있는 융합 단백질의 형태로 제조될 수 있다.
- [0017] 본 발명의 융합 단백질에 있어서, 상기 아데닌 또는 아데노신 탈아미노효소는 야생형 단백질의 전장, 야생형 단백질의 촉매 도메인, 촉매 활성이 증진된 돌연변이 단백질의 전장 또는 촉매 활성이 증진된 돌연변이 단백질의 촉매 도메인일 수 있으나, 이에 제한되지 않는다.
- [0018] 본 발명에 따른 아데닌 또는 아데노신 탈아미노효소는 서열번호 1 내지 서열번호 6의 아미노산 서열로 이루어질 수 있으나, 이에 제한되지 않는다.
- [0019] 본 발명의 아데닌 또는 아데노신 탈아미노효소는 서열번호 1 내지 서열번호 6으로 표시되는 아미노산 서열을 갖는 단백질 및 상기 단백질의 기능적 동등물을 포함한다. "기능적 동등물"이란 아미노산의 부가, 치환 또는 결실의 결과, 상기 서열번호 1 내지 서열번호 6으로 표시되는 아미노산 서열과 적어도 70% 이상, 바람직하게는 80% 이상, 더욱 바람직하게는 90% 이상, 더 더욱 바람직하게는 95% 이상의 서열 상동성을 갖는 것으로, 서열번호 1 내지 서열번호 6으로 표시되는 단백질과 실질적으로 동질의 생리활성을 나타내는 단백질을 말한다. 본 발명의 서열번호 1 및 2의 아미노산 서열은 인간(Homo sapiens)의 hADAR2 전장 단백질, 서열번호 3 및 4의 아미노산 서

열은 hADAR의 탈아미노 촉매 도메인(deaminase domain, hADAR2d), 서열번호 5 및 6의 아미노산 서열은 hADAR2의 2번째 RNA 결합 모티프부터 탈아미노 촉매 도메인까지 해당한다.

- [0020] 또한, 본 발명의 융합 단백질에 있어서, 상기 Cas9 단백질의 기능적 유사체는 Cpf1(Centromere and Promoter Factor 1), dCas9(catalytically-dead Cas9), nCas9(nickase Cas9), dCpf1(dead Cpf1), TALE(TAL effector), ZFP(Zinc Finger Proteins) 또는 NgAGO(Natronobacterium gregoryi Argonaute)일 수 있으나, 이에 제한되지 않으며, 다양한 미생물 유래의 또는 인위적으로 변형된 것으로 가이드(guide) RNA 또는 가이드 DNA를 사용하여 특정 DNA에 타겟팅할 수 있는 DNA 결합단백질을 포함할 수 있다.
- [0021] 본 발명에 따른 Cas9 단백질 또는 이의 기능적 유사체는 서열번호 7 내지 서열번호 10의 아미노산 서열로 이루어질 수 있으나, 이에 제한되지 않는다.
- [0022] 본 발명의 Cas9 단백질 또는 이의 기능적 유사체는 서열번호 7 내지 서열번호 10으로 표시되는 아미노산 서열을 갖는 단백질 및 상기 단백질의 기능적 동등물을 포함한다. 상기 단백질 및 단백질의 기능적 동등물의 범위는 전술한 것과 같다.
- [0023] Cpf1은 CRISPR/Cas 시스템의 한 종류로 Class2 Type5 CRISPR/Cas 시스템으로 분류된다. 그런데 Cpf1은 Cas9과 같이 단일 서브유닛 이펙터 모듈(single subunit effector module)로 작동하지만 기능적으로도 약간의 차이가 있다. Cas9이 유전자를 잘라내기 위해서는 tracrRNA를 필요로 하는데 반해 Cpf1은 tracrRNA를 필요하지 않는다. 또한, Cpf1은 유전자를 잘라낼 때 4~5개 뉴클레오타이드 유전자 오버행(overhang)을 발생시킨다는 특징을 가지고 있다. 이러한 여러 가지 특징들과 구조적인 차이 덕분에 Cpf1은 Cas9 보다 작동 효율은 다소 낮아도 더 높은 정확도를 나타낸다. 또 유전자 변이 유도시 DNA 플라스미드 형태가 아니라 RNP (ribonucleoprotein) 형태로 주입하면 더 정확도를 높일 수 있다고 보고되었다.
- [0024] 본 발명에 따른 티민 DNA 글리코실라제는 서열번호 11의 아미노산 서열로 이루어질 수 있으나, 이에 제한되지 않는다. 본 발명의 티민 DNA 글리코실라제는 서열번호 11의 아미노산 서열을 갖는 단백질 및 상기 단백질의 기능적 동등물을 포함한다. 상기 단백질 및 단백질의 기능적 동등물의 범위는 전술한 것과 같다.
- [0025] 본 발명은 또한, 본 발명의 융합 단백질;과 상기 융합 단백질의 Cas9 단백질 또는 이의 유사체 부분에 결합된 가이드 RNA;를 유효성분으로 포함하는 단일 염기 편집용 조성물을 제공한다.
- [0026] 본 발명의 용어 "염기 편집(base editing)"은 유전자 조작 기술인 유전자 가위, 크리스퍼(CRISPR)에서 더 발전한 형태의 기술로, 아데닌(A)-시토닌(C)-구아닌(G)-티민(T) 4개 염기의 근본적인 유전자 조합을 바꾸는 기술을 의미한다.
- [0027] 본 발명의 단일 염기 편집용 조성물에 있어서, 상기 융합 단백질은 아데닌 탈아미노효소(deaminase) 또는 아데노신 탈아미노효소;와 Cas9(CRISPR associated protein 9) 단백질 또는 이의 기능적 유사체;로 이루어진 융합 단백질, 또는 아데닌 탈아미노효소(deaminase) 또는 아데노신 탈아미노효소; 티민 DNA 글리코실라제(thymine DNA glycosylase); 및 Cas9(CRISPR associated protein 9) 단백질 또는 이의 기능적 유사체;로 이루어진 융합 단백질일 수 있으나, 이에 제한되지 않는다.
- [0028] 본 발명의 용어 "가이드 RNA(guide RNA)"는 교정하려는 DNA 위치를 찾아내는 역할을 수행하는 단일 가닥 RNA를 의미하며, 본 발명의 단일 염기 편집용 조성물에 있어서, 상기 가이드 RNA는 PAM(protospacer adjacent motif) 자리와 인접한 10~20bp의 염기 서열과 상보적인 서열을 포함하는 것일 수 있으나, 이에 제한되지 않으며, 상기 표적 서열로의 이동을 위한 상보적 서열 외에, 추가적으로 교정하고자 하는 표적 아데닌 자리와 미스매치가 생기게 하는 RNA 서열을 포함할 수 있으나, 이에 제한되지 않는다.
- [0029] 본 발명의 용어 "PAM(protospacer adjacent motif) 자리"는 Cas9 단백질이 표적 DNA에 결합하기 위해 필수적으로 필요한 서열로, 표적 서열에 뒤이어 위치하는 서열을 의미한다. 박테리아는 침입한 바이러스의 서열 일부를 박테리아 게놈 일부에 저장시켜 놓는데, 이 서열을 protospacer라 한다. 상기 protospacer 서열은 일부이므로, 그 인접 서열은 박테리아 본연의 서열이게 된다. PAM 자리의 역할은 박테리아 내에 많은 서열이 protospacer와 일치할 수 있지만, PAM이 그 옆에 붙어있지 않으면 해당 서열은 절단되지 않게 하는 기능을 한다. 일반적으로 Cas9의 경우 구아닌이 많은 PAM이 있는 부위의 유전자에 주로 작동하는 반면, Cpf1은 티민이 많은 PAM이 있는 부위의 유전자를 효율적으로 절단하는 것으로 알려져 있다.
- [0030] 본 발명의 단일 염기 편집용 조성물에 있어서, 상기 가이드 RNA는 아데닌 탈아미노효소의 활성과 선택성을 최적화하기 위해 변형된 1개 이상의 가이드 RNA 또는 스캐폴드(scaffold) RNA가 이용될 수 있으나, 이에 제한되지

않는다

- [0031] 본 발명의 단일 염기 편집용 조성물은 *in vitro* 및 *in vivo*, 원핵세포/생물 및 진핵세포/생물의 염색체, 또는 진핵세포의 미토콘드리아 또는 플라스티드의 DNA를 교정하는데 이용될 수 있으나, 이에 제한되지 않는다..
- [0032] 본 발명의 일 구현 예에 따른 단일 염기 편집용 조성물은, 서열번호 1 내지 6의 아미노산 서열로 이루어진 펩티드가 서열번호 7 내지 서열번호 10의 아미노산 서열로 이루어진 nSpCas9/dSpCas9의 N-말단 또는 C-말단에 융합된 융합 단백질 및 변형된 스캐폴드 RNA(도 2, 4, 6) 물질을 포함하는 복합체로 이루어져 있으나, 이에 제한되지 않는다.
- [0033] 또한, 상기 아데노신 염기 편집용 복합체는 hADAR2 탈아미노 촉매 도메인과 nSpCas9사이의 XTEN 링커(5'-TCCGGTAGCGAAACACCGGGGACTTCAGAATCGGCCACCCCGGAGTCT-3'; 서열번호 12) 및 2xSV40NLS(5'-GATCCAAAAAAGAAGAGAAAGGTAGATCCAAAAAAGAAGAGAAAGGTA-3'; 서열번호 13)를 C-말단에 포함한다. XTEN 링커 외, 융합 단백질의 연결을 위해 (GGGGGS)_n; n=1-5 등이 사용될 수 있다. 또한 A/T→I/T 전환 후 티민 DNA 글리코실라제를 사용하여 T를 제거하는 염기 삭제 복구 경로(base excision repair pathway)를 우선적으로 진행시켜 I→C(G-C)로의 변환을 극대화할 수 있다.
- [0034] 본 발명의 일 구현 예에 따른 단일 염기 편집용 조성물에 있어서, 서열번호 1 및 2의 아미노산 서열로 이루어진 아데닌 또는 아데노신 탈아미노효소를 사용할 경우에는 가이드 RNA 스캐폴드에 RNA 결합 모티프를 첨가하여 사용하고, 서열번호 4 내지 6의 아미노산 서열로 이루어진 아데닌 또는 아데노신 탈아미노효소를 사용할 경우에는 가이드 RNA 스캐폴드에 RNA 결합 모티프를 포함하지 않고 제작할 수 있다. 또한, 이 경우 표적 DNA 가닥(센스(S)-extended scaffold guide RNA 사용; 도 2) 및 비표적 DNA 가닥(안티센스(AS)-extended scaffold guide RNA 사용; 도 3 내지 6)의 두 DNA 나선의 아데노신 염기가 교정될 수 있다. 또한 n/dCas9을 쓰는 방식, dCpf1을 쓰는 방식과 다양한 스캐폴드 RNA를 센스(S) 또는 안티센스(AS)로 사용함은 물론 1개 또는 2개 이상으로 분리하되 2개로 나눌 경우는 ADAR2 결합 헤어핀 스템 루프(hairpin stem loop)를 포함하여 제조하는 방식이 조합될 수 있다(도 7 내지 11). DNA 가닥 상의 표적 아데닌은 스캐폴드 가이드 RNA와 결합시 상보위치에 시토신을 미스매치시켜 표적 선택성(selectivity)을 극대화할 수 있다(도 8, 9, 11; A-C mismatch). 또한 스캐폴드 가이드 RNA의 안정성 및 효율성을 증가시키기 위해 5' 말단에 망치머리형 리보자임(hammerhead ribozyme) 등의 변형이 사용될 수 있다.
- [0035] 상기 아데노신 염기 편집용 복합체는 hADAR2외에 hADAR1 등의 상동체 및 다양한 생물로부터 발견되는 ADAT를 사용하여 만들어질 수 있다. hADAR2의 E488Q, hADAR1 E1008Q는 효소활성이 증대된 돌연변이로 염기교정 효율을 증가시키지만, 상기 돌연변이체뿐만 아니라 야생형도 이용될 수 있다. hADAR2 및 hADAR1의 경우 RNA/DNA 잡종을 기질로 인식하므로 단일 가닥 DNA를 인지할 수 있는 아데노신 탈아미노효소는 유리 아데노신 탈아미노효소, 테옥시 시티딘 탈아미노효소(deoxy cytidine deaminase) 및 APOBEC(apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) 단백질의 촉매자리의 비교분석을 통해 단일 가닥 DNA를 효율적으로 인지할 수 있는 합성의 잡종 아데노신 탈아미노효소(synthetic hybrid adenosine deaminase)를 사용한다. 일례로 유리(free) 뉴클레오타이드 아데닌 탈아미노효소에서 중요한 잔기인 G184, H214, Q217, H238, D295, D296 및 주변부에서 중요하게 생각되는 H15, H17, D19 등의 조성을 APOBEC에 도입하여 시티딘 탈아미노효소를 아데노신 탈아미노효소로 개조하는 것을 포함한다(도 12). 또한 hADAR2d(E488Q)외에 T490을 알라닌, 글루타민, 글루탐산, 아스파라긴, 아스파르트산으로 바꾸어 단일 가닥 DNA 선택성을 증가시킨다(도 13). 또한 인간 hAPOBEC3A 구조와 유리(free) 뉴클레오타이드 아데닌 탈아미노효소의 구조분석을 통해 hAPOBEC3A 단백질의 T31G/A/S, Y130G/A 및 P100D 등의 점돌연변이를 통해 시토신 염기보다 큰 아데닌 염기가 촉매 자리(catalytic site)에 들어가게 하고 H⁷⁰AE⁷² 서열의 H70 과 E72 사이에 추가적인 글라이신을 (H⁷⁰AGE⁷² 또는 H⁷⁰GAE⁷²) 삽입하는 방법 및 촉매 포켓(catalytic pocket)에 위치한 루프(loop) 및 나선(helix)에서 하나 또는 그 이상의 아미노산을 제거하여 아데닌 기질을 위한 공간을 확보하는 돌연변이체(서열번호 14 및 서열번호 15)를 포함한다.
- [0036] 효모의 경우 대장균-효모 서플렉터를 기반으로 golden gate assembly를 통해 염기교정 모듈 복합체가 클로닝되고(플라스미드 코어 구조, 도 14 내지 17), 화학적인 쇼크 또는 전기천공법(electroporation) 방식을 통해 형질전환 되어졌다. 표적 유전자는 LEU2 영양요구성(auxotrophic) 마커 유전자 또는 플라스미드 장착 GFP 유전자가 사용되었다.
- [0037] 본 발명의 일 구현예는 LEU2 유전자의 17번째 또는 180번째 아미노산 대응서열인 CAA가 TAA(종결 코돈)으로 변형되어 A/T→G/C로 교정이 일어나면 Leu- 배지에서 살 수 있도록 설계되었다(도 1, 2). 설계된 제조합 벡터의

동식물 세포 및 개체로의 형질전환 방법은 PEG 기반 원형질체 형질전환(protoplast transformation), 미세주입법(microinjection), 전기천공법, 유전자총(gene gun, particle bombardment), 리포솜 매개 형질도입(transfection) 및 아그로박테리움(*Agrobacterium*) 매개 형질전환법이 이용될 수 있다. 이와 같은 형질전환 또는 형질도입 방법은 당업계에 잘 알려져 있다.

[0038] 본 발명은 또한, 본 발명의 융합 단백질;과 상기 융합 단백질의 Cas9 단백질 또는 이의 유사체 부분에 결합된 가이드 RNA;를 유효성분으로 포함하는 단일 염기 편집용 조성물을 핵산 분자와 접촉시키는 단계를 포함하는 염기 편집 방법을 제공한다.

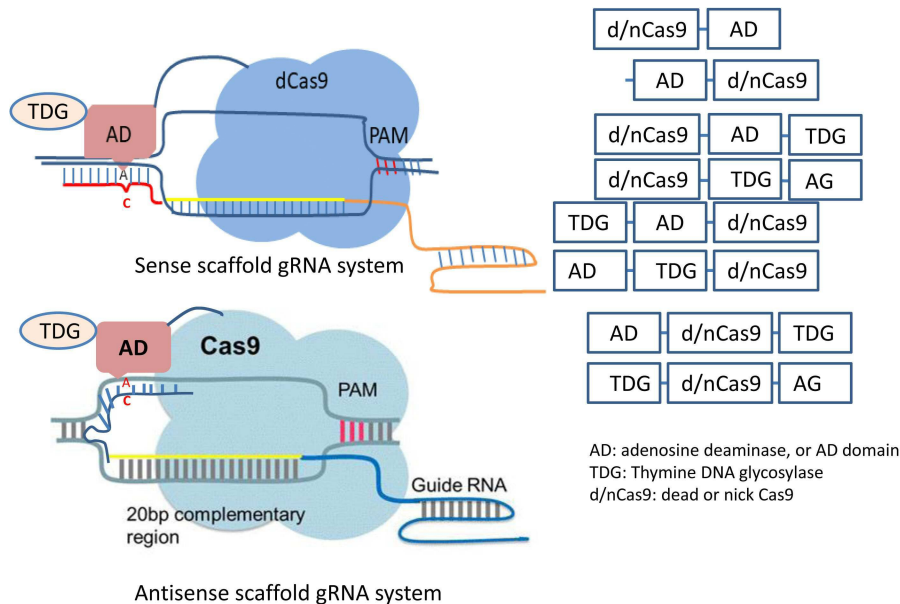
[0039] 본 발명의 염기 편집 방법에 있어서, 상기 핵산 분자는 교정하고자 하는 표적 염기를 포함하는 것일 수 있고, 상기 교정하고자 하는 표적 염기는 질병 또는 질환과 연관된 구아닌(guanine, G)이 아데닌(adenine, A)으로 점 돌연변이된 염기인 것일 수 있으나, 이에 제한되지 않는다.

[0040] 본 발명의 염기 편집 방법은, 구아닌(guanine, G)에서 아데닌(adenine, A)으로 점 돌연변이가 일어난 자리를 표적으로 하여, 본 발명의 아데닌 탈아미노효소(deaminase) 또는 아데노신 탈아미노효소;와 Cas9(CRISPR associated protein 9) 단백질 또는 이의 기능적 유사체;로 이루어진 융합 단백질과 가이드 RNA가 유효성분으로 포함된 조성물에 의해 A→G로 염기가 교정되는 방법이다.

[0041] 이상에서 설명한 바와 같이 본 발명에 아데닌 또는 아데노신 탈아미노효소를 유효성분으로 포함하는 단일 염기 편집용 조성물은 DNA 염기 하나만을 표적으로 하여 특정 염기 하나를 다른 염기로 바로 전환하기 때문에 DNA 절단 없이도 교정할 수 있는 장점이 있으므로, 농생명 및 의생명 분야에 다양한 용도로 활용될 수 있을 것이다.

도면

도면1

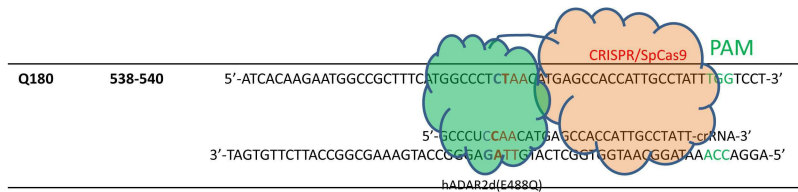


도면2

ScLEU2 target DNA region and sense (SE) scaffold gRNA1

No.	Gene	Target	Position	Sequence
1	ScLEU2	Q17	49-51	ATCGTCGTTTTGCGAGGTGACCACGTTGGT(c)CAAGAAATCACAGCCGAAGCCATTAAGGTTTC: 16, AS
		Q180	538-540	ATCACAGAATGGCCGCTTTCATGGCCCTA(c)CAACATGAGCCACCATTGCCTATTGGTCCT: 24, S

SE scaffold gRNA1: 5'-GCCCUCCAAACATGAGCCACCATTGCCTATT-crRNA-3'

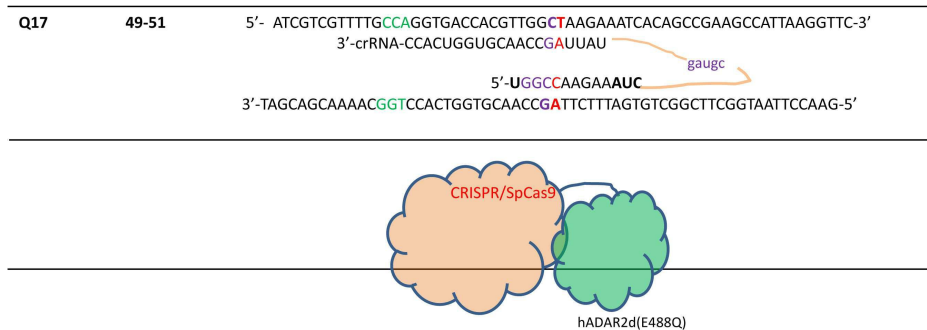


Glutamine codons (CAA) of ScLEU2 was modified to stop codon (TAA).
A/T-to G/C base editing rescues auxotrophic mutation.

도면3

ScLEU2 target DNA region and antisense (ASE) scaffold gRNA 1

ASE scaffold gRNA1: 5'-UGGCCAAGAAAUGCGUAGUAUUAGCCAACGUGGUVACC crRNA-3'



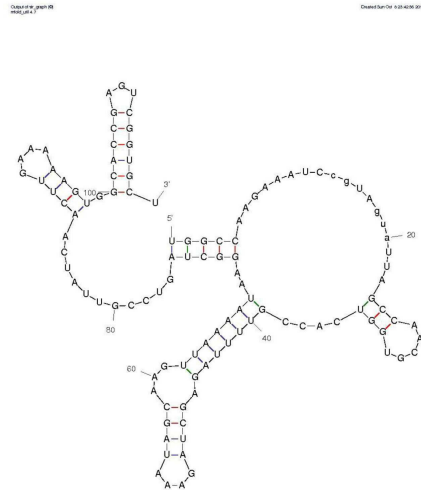
Glutamine codons (CAA) of ScLEU2 was modified to stop codon (TAA).
A/T-to G/C base editing rescues auxotrophic mutation.

도면4

antisense(ASE) scaffold gRNA 1 sequence and folding structure

5'-

U**GGCCAAGAA**AUCcgUAguuUAGCCAACGUGGUCACCGUUUUAGAGCUAGAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUU
GAAAAAGUGGCACCGAGUCGGUGCU-3'

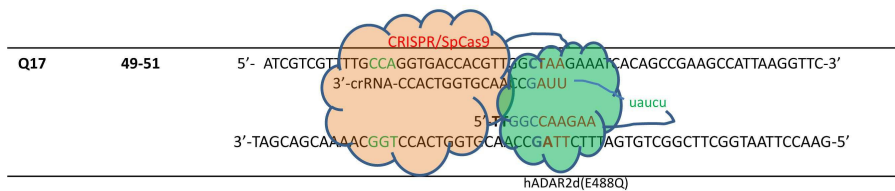


도면5

ScLEU2 target DNA region and antisense scaffold gRNA 2

➤ Add more nucleotides to the 5' end of the SE and ASE scaffolds to increase their life-time

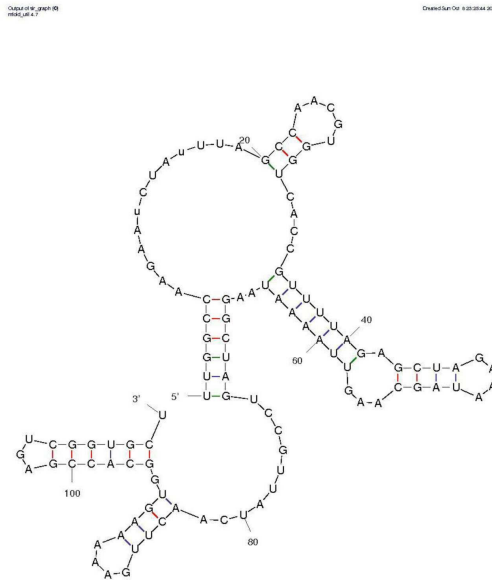
- AS-extended scaffolds for ADg-CRISPR/SpCas9g : **ASE-LEU2.3** scaffold with 2 nt left, 0 nt on the opposite of the right side of the TAG. Fixed the UU to AA. Using gAD stem-loop as a linker.



도면6

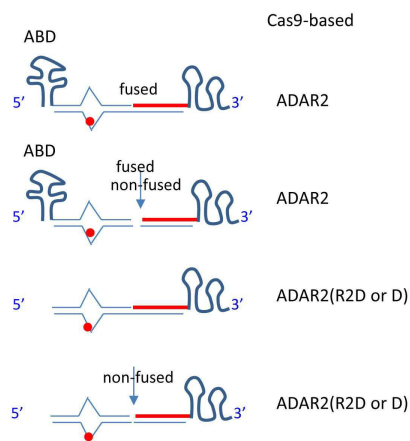
antisense scaffold gRNA 2 sequence and folding structure

5'-
TTGGCCAAGAAUCUAUUAGCCAACGTGGTCACCGTTTGTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC
CGTTATCAACTTGAAAAAGTGGCACCAGTGCCTGCT-3'



도면7

다양한 adenosine base editor 전략

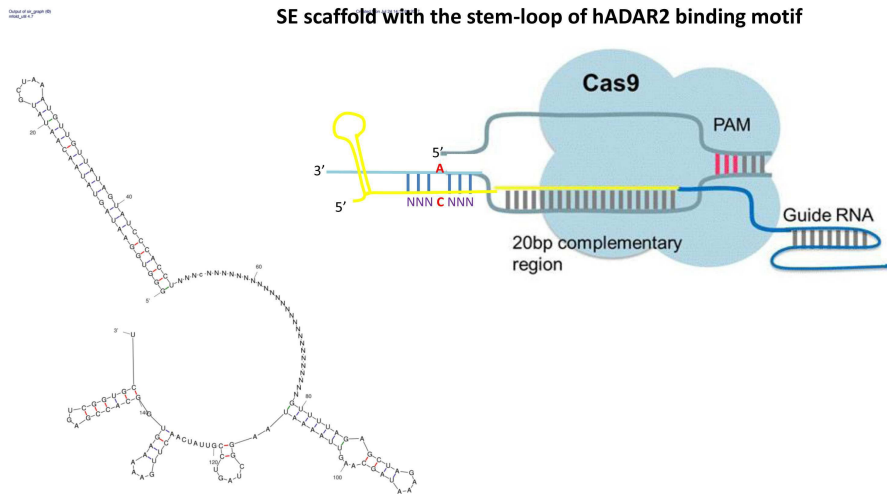


ABD: ADAR2 binding RNA domain

	Cas9 /Cpf 1	Adenosine deaminase	gRNA	A-target scaffold
Cas9-based	dCas9/nCas9	• ADAR1 • ADAR2 • ADAR1d(domain) • ADAR2d(domain)	16-25nt spacer	• Sense or antisense with ADAR1/2 binding RNA domain • Fused to gRNA or non-fused to gRNA • Without target scaffold
Cpf1-based	dCpf1	• Their more active mutated form • Synthetic adenosine deaminase		
NgAGO	NgAg o	Synthetic adenosine deaminase	16-25nt gDNA(m8 mismatch mutated form)	Without target scaffold

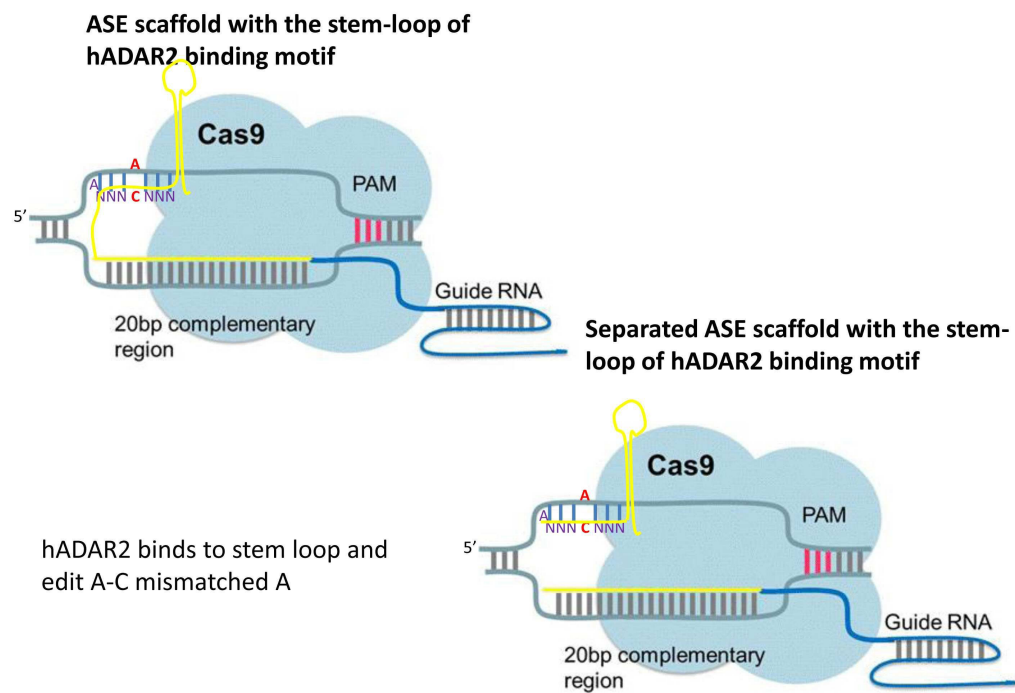
도면8

ADAR2-nCas9 and RNA motif for RNA binding domains of ADAR2 (sense gRNA)



도면9

ADAR2-nCas9 and RNA motif for RNA binding domains of ADAR2 (antisense gRNA)



도면10

ADAR2-dLbCpf1 and RNA motif for RNA binding domains (sense gRNA)

- Hybrid scaffolds for ADg-CRISPR/LbCpf1:

5'-

TAATTTCTACTAAGTGTAGATNNNNNNNNNNNNNNNNNNNN
 NNGGGTGGAATAGTATAACAATATGCTAAATGTTGTTATAGTATCC
 CACCT-3'

With cloning site BsmBI:

5'-

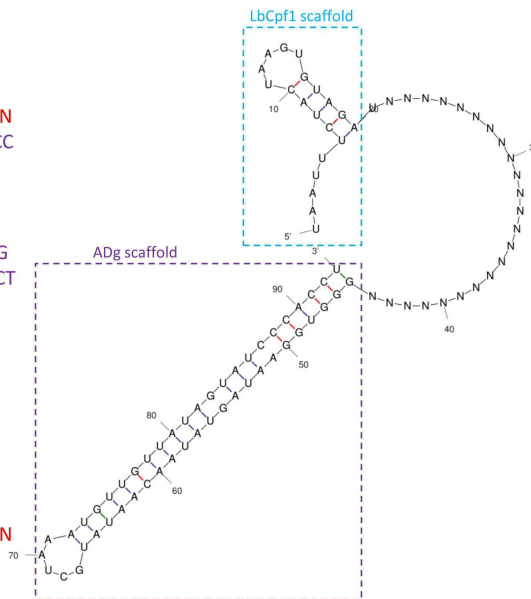
TAATTTCTACTAAGTGTAGATTGAGACGGCATGCATCGTCTCAGG
 GTGGAATAGTATAACAATATGCTAAATGTTGTTATAGTATCCACCT
 -3

Cloning strategy: Using ACCT-GTTT overhangs with BsmBI cutting sites for assembly of double-stranded oligos

- Simple scaffold for ADg-CRISPR/LbCpf1:

5'-

TAATTTCTACTAAGTGTAGATNNNNNNNNNNNNNNNNNNNN
 NNNNNNNNN



One of 'N' should be C(for C-A target mismatch)

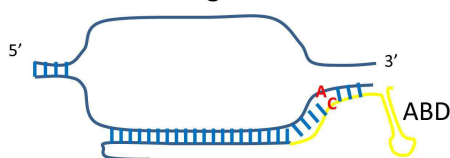
dG = -27.80 [Initially -28.00] ADg-gLbCpf1

Hybrid scaffolds for ADg-CRISPR/LbCpf1

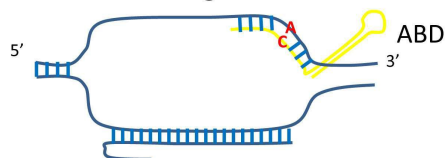
도면11

Sense(SE) or antisense(ASE) scaffold for ADAR2-dLbCpf1

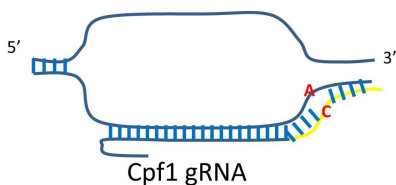
SE scaffold with the stem-loop of hADAR2 binding motif



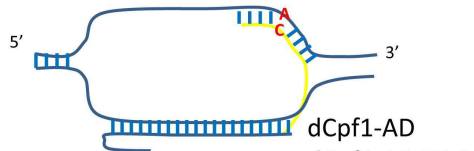
ASE scaffold with the stem-loop of hADAR2 binding motif



SE scaffold



ASE scaffold



AD: ADAR2 or its deaminase domain
 dCpf1: dead Cpf1
 TDG: Thymine DNA glycosylase

dCpf1-AD
 dCpf1-AD-TDG
 dCpf1-TDG-AD
 AD-dCpf1
 TDG-AD-dCpf1
 AD-TDG-dCpf1
 AD-dCpf1-TDG
 TDG-dCpf1-AD

도면12

1.2 The role of a.a. residues in the catalytic site: G184, H214, Q217, H238, D295, D296 Some neighbor residues need to be supplemented to mimic: H15, H17, D19

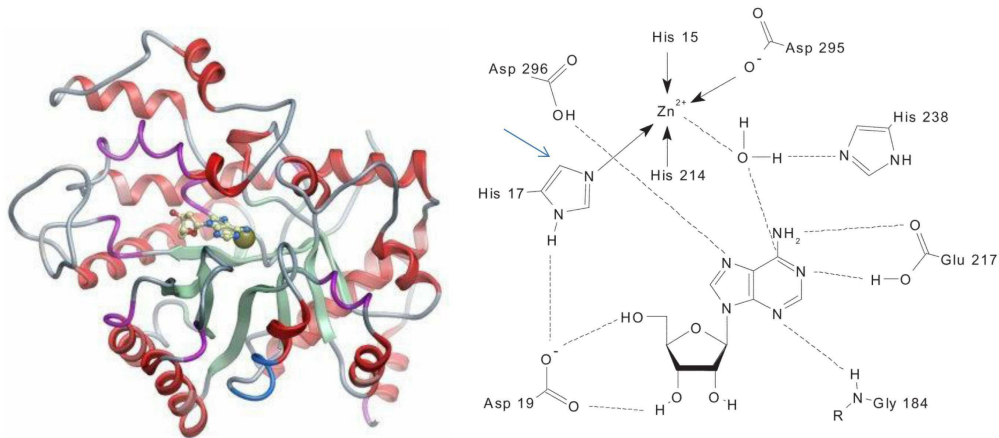


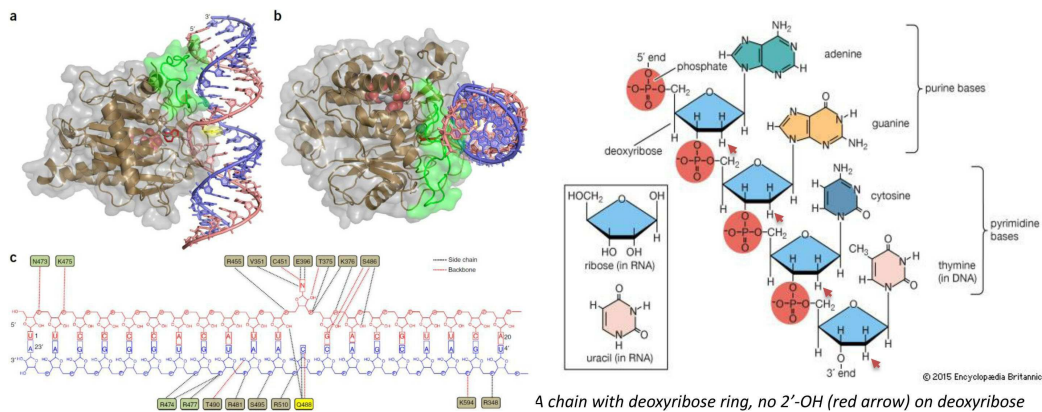
Fig. 1 ADA active site with its natural substrate. Only the important catalytic residues shown.

Gleeson et al., 2003

So we have a set of a.a. residues to engineer synthetic adenine deaminases working on ssDNAs or RNA-DNA hybrids: C262, G184, H214, Q217, H238, D295, D296
Some neighbor residues need to be supplemented to mimic: H15, H17, D19

도면13

Catalytic activities of point-mutated hADAR2d (E488Q) on ssDNAs:
Mode of interaction of R (Arginine) to a ribose/deoxyribose ring



A chain with deoxyribose ring, no 2'-OH (red arrow) on deoxyribose

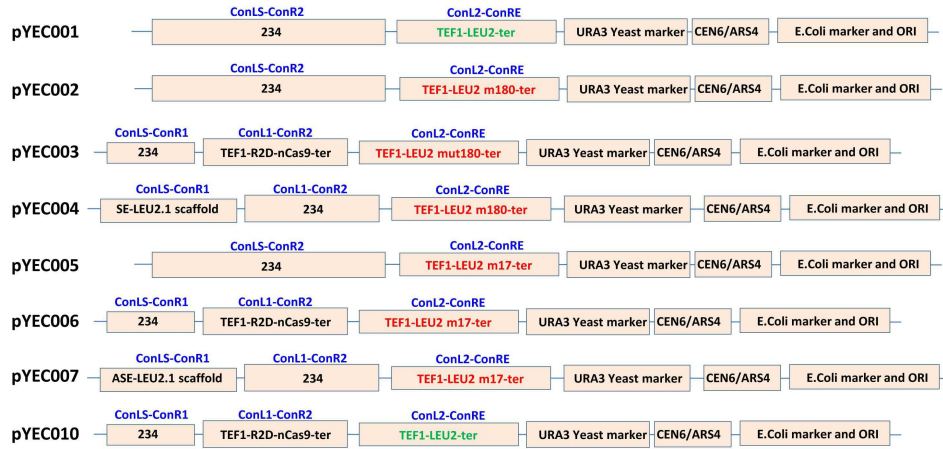
We can make point mutation at T490 so that it will not bind to the ribose -OH or alternative binding to other:

1. Replace T490 by a hydrophobic a.a. like Alanine (A) to make an inert residue.
2. Replace T490 by a polar side chain a.a. like Glutamine (Q) or Asparagine (N), Aspartic acid, Glutamic acid to make an -e acceptor and/or donor a.a. which may interact with deoxyribose 2'-H on ssDNAs.

도면14

Shuttle vector construction: R2D fused to N-terminal of nSpCas9

Control plasmids:



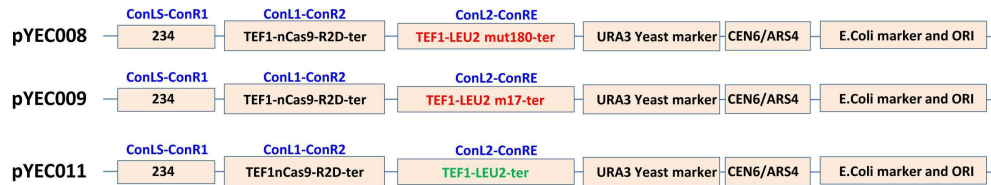
Test plasmids:



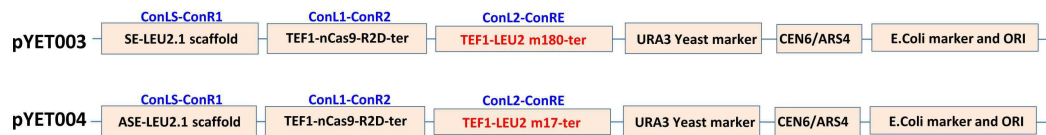
도면15

Shuttle vector construction: R2D fused to C-terminal of nSpCas9

Control plasmids:



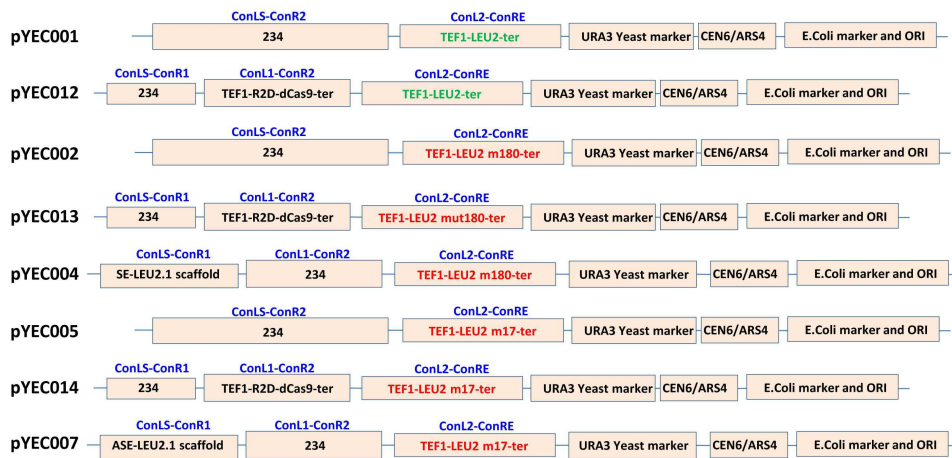
Test plasmids:



도면16

Shuttle vector construction: R2D fused to N-terminal of dSpCas9

Control plasmids:



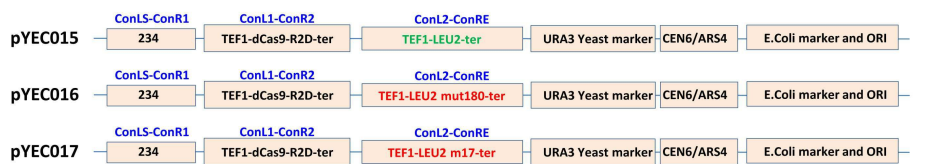
Test plasmids:



도면17

Shuttle vector construction: R2D fused to C-terminal of dSpCas9

Control plasmids:



Test plasmids:



서열 목록

- <110> INDUSTRY-ACADEMIC COOPERATION FOUNDATION GYEONGSANG NATIONAL UNIVERSITY
- <120> Composition for single base editing comprising adenine or adenosine deaminase as effective component and uses thereof
- <130> PN17397
- <160> 15
- <170> Kopatent In 2.0
- <210> 1

<211> 2169

<212> DNA

<213> Artificial Sequence

<220><223> hADAR2, E488Q mutation(AD), type 3a

<400> 1

cggctctcata tgatggatat agaagatgaa gaaaacatga gttccagcag cactgatgtg 60

aaggaaaacc gcaatctgga caacgtgtcc cccaaggatg gcagcacacc tgggcctggc 120

gagggtcttc agctctccaa tgggggtggt ggtggccccg gcagaaagcg gcccttgag 180

gagggcagca atggccactc caagtaccgc ctgaagaaaa ggaggaaaac accaggggccc 240

gtcctcccca agaacgcctt gatgcagctg aatgagatca agcctggttt gcagtacaca 300

ctcctgtccc agactgggcc cgtgcacgcg cctttgtttg tcatgtctgt ggaggatgaat 360

ggccagggttt ttgagggtc tgggtcccaca aaaaaaagg caaaactcca tgctgtctgag 420

aaggccttga ggtctttcgt tcagtttctt aatgcctctg agggccacct ggccatgggg 480

aggaccctgt ctgtcaacac ggacttcaca tctgaccagg ccgacttccc tgacacgctc 540

ttcaatggtt ttgaaactcc tgacaaggcg gagcctccct ttacgtggg ctccaatggg 600

gatgactcct tcagttccag cggggacctc agcttgtctg cttccccggt gcctgccagc 660

ctagcccagc ctctctccc tgtcttacca ccattccac ccccgagtgg gaagaatccc 720

gtgatgatct tgaacgaact gcgcccagga ctcaagtatg acttctctc cgagagcggg 780

gagagccatg ccaagagctt cgtcatgtct gtggtcgtgg atggtcagtt ctttgaaggc 840

tcggggagaa acaagaagct tgccaaggcc cgggctgcgc agtctgccct ggccgccatt 900

tttaacttgc acttggatca gacgccatct cgccagccta ttcccagtga gggccttcag 960

ctgcatttac cgcaggtttt agctgacgct gtctcacgcc tggtcctggg taagtttggt 1020

gacctgaccg acaacttctc ctcccctcac gctgcagaa aagtgtctgc tggagtcgtc 1080

atgacaacag gcacagatgt taaagatgcc aaggatgataa gtgtttctac aggaacaaaa 1140

tgtattaatg gtgaatacat gagtgatcgt ggccttgcat taaatgactg ccatgcagaa 1200

ataatatctc ggagatcctt gctcagattt ctttatacac aacttgagct ttacttaaat 1260

aacaaagatg atcaaaaaag atccatcttt cagaaatcag agcgaggggg gtttaggctg 1320

aaggagaatg tccagtttca tctgtacatc agcacctctc cctgtggaga tgccagaatc 1380

ttctcaccac atgagccaat cctggaagaa ccagcagata gacacccaaa tcgtaaagca 1440

agaggacagc tacggaccaa aatagagtct ggtcagggga cgattccagt gcgtccaat 1500

gcgagcatcc aaacgtggga cggggtgctg caaggggagc ggctgctcac catgtcctgc 1560

agtgacaaga ttgcacgctg gaacgtgggtg ggcatccagg gatccctgct cagcattttc 1620
gtggagccca ttacttctc gagcatcatc ctgggcagcc ttaccacgg ggaccacctt 1680
tccagggccca tgtaccagcg gatctccaac atagaggacc tgccacctct ctacaccctc 1740
aacaagcctt tgctcagtgg catcagcaat gcagaagcac ggcagccagg gaaggccccc 1800

aacttcagtg tcaactggac ggtaggcgac tccgtattg aggtcatcaa cgccacgact 1860
gggaaggatg agctgggccc cgctccccgc ctgtgtaagc acgcgttgta ctgtcgtgg 1920
atgcgtgtgc acggcaaggt tccctccac ttactacgt ccaagattac caaacccaac 1980
gtgtaccatg agtccaagct ggccgcaaag gattaccagg ccgccaaggc gcgtctgttc 2040
acagccttca tcaaggcggg gctgggggcc tgggtggaga agcccaccga gcaggaccag 2100
ttctcactca cgccctccgg tagcgaaca ccggggactt cagaatcggc cccccggag 2160
tctggttct 2169

<210> 2

<211> 2221

<212> DNA

<213> Artificial Sequence

<220><223> hADAR2 , E488Q mutation (AD), type 4a

<400> 2

cggtctcaat cctccggtag cgaaacaccg gggacttcag aatcgccac cccggagtct 60
atggatatag aagatgaaga aaacatgagt tccagcagca ctgatgtgaa ggaaaaccgc 120
aatctggaca acgtgtcccc caaggatggc agcacacctg ggcttgccga gggctctcag 180
ctctccaatg ggggtgggtg tggccccggc agaaagcggc ccctggagga gggcagcaat 240
ggccactcca agtaccgcct gaagaaaagg aggaaaacac cagggccctt cctccccaag 300

aacgccctga tgcagctgaa tgagatcaag cctggtttgc agtacacact cctgtcccag 360
actgggcccc tgcacgcgcc ttgtttgtc atgtctgtgg aggtgaatgg ccaggttttt 420
gagggtcttg gtcccacaaa gaaaaaggca aaactccatg ctgctgagaa ggccttgagg 480
tctttcgttc agtttcctaa tgctctgag gccacctgg ccatggggag gacctgtct 540
gtcaacacgg atttcacatc tgaccaggcc gacttccctg acacgtcttt caatggtttt 600
gaaactcctg acaaggcggg gcctcccttt tacgtgggct ccaatgggga tgactccttc 660
agttccagcg gggacctcag cttgtctgct tccccgtgct ctgccagcct agcccagcct 720

cctctccctg tcttaccacc attcccaccc ccgagtggga agaatccgt gatgatcttg 780
aacgaactgc gccaggact caagtatgac ttctctccg agagcgggga gagccatgcc 840

aagagcttcg tcatgtctgt ggtcgtggat ggtcagttct ttgaaggctc ggggagaaac 900
aagaagcttg ccaaggcccg ggctgcgcag tctgccctgg ccgccatttt taacttgac 960
ttggatcaga cgccatctcg ccagcctatt cccagtgagg gccttcagct gcatttaccg 1020
caggttttag ctgacgtgtg ctacgcctg gtccctgggtg agtttgggtga cctgaccgac 1080
aacttctcct cccctcacgc tcgcagaaaa gtgctggctg gagtctcat gacaacaggc 1140

acagatgtta aagatgccaa ggtgataagt gtttctacag gaacaaaatg tattaatggt 1200
gaatacatga gtgatcgtgg ccttgcatga aatgactgcc atgcagaaat aatatctcgg 1260
agatccttgc tcagatttct ttatacaca cttgagcttt acttaataa caaagatgat 1320
caaaaaagat ccatctttca gaaatcagag cgaggggggt ttaggctgaa ggagaatgtc 1380
cagtttcacg tgiacatcag caccctctcc tgtggagatg ccagaatctt ctaccacat 1440
gagccaatcc tggaagaacc agcagataga caccctaatc gtaaagcaag aggacagcta 1500
cggacaaaaa tagagtctgg tcaggggacg attccagtgc gctccaatgc gagcatcaa 1560

acgtgggacg ggggtgtgca aggggagcgg ctgctcacca tgtcctgcag tgacaagatt 1620
gcacgtctga acgtggtggg catccaggga tcctgtctca gcattttcgt ggagccatt 1680
tacttctcga gcatcatcct gggcagcctt taccacgggg accaccttcc cagggccatg 1740
taccagcgga tctccaacat agaggacctg ccacctctct acacctcaa caagccttg 1800
ctcagtggca tcagcaatgc agaagcacgg cagccaggga aggccccaa cttcagtgtc 1860
aactggacgg taggcgactc cgctattgag gtcataacg ccacgactgg gaaggatgag 1920
ctgggccgcg cgtcccgctt gtgtaagcac gcgttgtact gtcgctggat gcgtgtgcac 1980

ggcaaggttc cctccactt actacgtctc aagattacca aaccaacgt gtacatgag 2040
tccaagctgg cggcaaagga gtaccaggcc gccaaagcgc gtctgttcac agccttcac 2100
aaggcggggc tgggggcctg ggtggagaag cccaccgagc aggaccagt ctcaactcacg 2160
cccgatcaa aaaagaagag aaaggtagat ccaaaaaaga agagaaaggt ataatggctg 2220
a 2221

<210> 3

<211> 1224

<212> DNA

<213> Artificial Sequence

<220><223> hADAR2 domain, E488Q mutation (D), type 3a

<400> 3

cggtctcata tgatgtcga ttaccgcag gttttagctg acgtgtctc acgcctggc 60

ctgggtaagt ttggtgacct gaccgacaac ttctcctccc ctcacgctcg cagaaaagtg 120
ctggctggag tcgtcatgac aacaggcaca gatgttaaag atgccaaggt gataagtgtt 180
tctacaggaa caaatgtat taatggtgaa tacatgagtg atcgtggcct tgcattaaat 240
gactgccatg cagaaataat atctcggaga tccttgctca gatttcttta tacacaactt 300
gagctttact taaataacaa agatgatcaa aaaagatcca tctttcagaa atcagagcga 360
gggggggttta ggctgaagga gaatgtccag ttctatctgt acatcagcac ctctccctgt 420

ggagatgcca gaatcttctc accacatgag ccaatcctgg aagaaccagc agatagacac 480
ccaaatcgta aagcaagagg acagctacgg accaaaatag agtctggtca ggggacgatt 540
ccagtgcgct ccaatgcgag catccaaacg tgggacgggg tgctgcaagg ggagcggctg 600
ctcaccatgt cctgcagtga caagattgca cgctggaacg tgggtggcat ccagggatcc 660
ctgctcagca ttttctgga gccattttac ttctcagca tcctcctggg cagcctttac 720
cacggggacc acctttccag ggccatgtac cagcggatct ccaacataga ggacctgcca 780
cctctctaca cctcaacaa gcctttgctc agtggcatca gcaatgcaga agcacggcag 840

ccagggaagg cccccaactt cagtgtcaac tggacggtag gcgactccgc tattgaggtc 900
atcaacgcca cgactgggaa ggatgagctg ggccgcgcgt cccgcctgtg taagcacgcg 960
ttgtactgtc gctggatgcg tgtgcacggc aaggttcctt cccacttact acgtccaag 1020
attaccaaac ccaacgtgta ccatgagtcc aagctggcgg caaaggagta ccaggccgcc 1080
aaggcgcgtc tgttcacagc cttcatcaag gcggggctgg gggcctgggt ggagaagccc 1140
accgagcagg accagttctc actcacgcc tccggtagcg aaacaccggg gacttcagaa 1200
tcggccaccc cggagtctgg ttct 1224

<210> 4

<211> 1273

<212> DNA

<213> Artificial Sequence

<220><223> hADAR2 domain, E488Q mutation (D), type 4a

<400> 4

cggctcctaat cctccggtag cgaaacaccg gggacttcag aatcgccac cccggagtct 60
ctgcatttac cgcaggtttt agctgacgct gtctcacgcc tggctcctggg taagtttggg 120
gacctgaccg acaacttctc ctccctcac gctcgcagaa aagtgtggc tggagtcgtc 180
atgacaacag gcacagatgt taaagatgcc aaggtgataa gtgtttctac aggaacaaaa 240
tgtattaatg gtgaatacat gagtgatcgt ggccttgcct taaatgactg ccatgcagaa 300

ataatatctc ggagatcctt gctcagattt ctttatacac aacttgagct ttacttaaat 360
aacaagatg atcaaaaaat atccatcttt cagaaatcag agcgaggggg gtttaggctg 420
aaggagaatg tccagtttca tctgtacatc agcacctctc cctgtggaga tgccagaatc 480
ttctcaccac atgagccaat cctggaagaa ccagcagata gacacccaaa tcgtaaagca 540
agaggacagc tacggacca aatagagtct ggtcagggga cgattccagt gcgctccaat 600
gcgagcatcc aaacgtggga cggggtgctg caaggggagc ggctgctcac catgtcctgc 660
agtgacaaga ttgcacgctg gaacgtgggtg ggcatccagg gatccctgct cagcattttc 720

gtggagccca ttactttctc gagcatcatc ctgggcagcc tttaccacgg ggaccacctt 780
tccagggccca tgtaccacg gatctccaac atagaggacc tgccacctct ctacaccctc 840
aacaagcctt tgctcagtg catcagcaat gcagaagcac ggcagccagg gaaggccccc 900
aacttcagt tcaactggac ggtaggcgac tccgctattg aggtcatcaa cgccacgact 960
gggaaggatg agctgggccg cgcgtcccgc ctgtgtaagc acgcgttgta ctgtcgtg 1020
atgcgtgtgc acggcaaggt tccctccac ttactacgt ccaagattac caaaccaac 1080
gtgtaccatg agtccaagct ggcggaag gagtaccagg ccgccaaggc gcgtctgttc 1140

acagccttca tcaaggcggg gctgggggcc tgggtggaga agcccaccga gcaggaccag 1200
ttctcactca cgcccgatcc aaaaaagaag agaaagtag atcaaaaaa gaagagaaag 1260
gtataatggc tga 1273

<210> 5

<211> 1530

<212> DNA

<213> Artificial Sequence

<220><223> hADAR2 domain, E488Q mutation (R2D), type 3a

<400> 5

cggctctcata tgatggccag cctagcccag cctcctctcc ctgtcttacc accattccca 60
ccccgagtg ggaagaatcc cgtgatgac ttgaacgaac tgcgcccagg actcaagtat 120

gacttctct ccgagagcgg ggagagccat gccaagagct tcgtcatgtc tgtggctgtg 180
gatggtcagt tctttgaagg ctcggggaga aacaagaagc ttgccaaggc cgggctgcg 240
cagtcgtccc tggccgcat tttaacttg cacttggatc agacgccatc tcgccagcct 300
attcccagt agggccttca gctgcattta ccgcaggttt tagctgacgc tgtctcacgc 360
ctggctctgg gtaagtttgg tgacctgacc gacaacttct cctccctca cgctcgaga 420
aaagtgtgg ctggagtcgt catgacaaca ggcacagatg ttaaagatgc caaggatgata 480

agtgttttcta caggaacaaa atgtattaat ggtgaatata tgagtgatcg tggccttgca 540

 ttaaattgact gccatgcaga aataatatct cggagatcct tgctcagatt tctttataca 600
 caacttgagc ttactttaaa taacaaagat gatcaaaaaa gatccatctt tcagaaatca 660
 gagcgagggg ggtttaggct gaaggagaat gtccagtttc atctgtacat cagcacctct 720
 ccctgtggag atgccagaat cttctcacca catgagccaa tcctggaaga accagcagat 780
 agacacccaa atcgtaaagc aagaggacag ctacggacca aaatagagtc tggtcagggg 840
 acgattccag tgcgtccaa tgcgagcatc caaacgtggg acggggtgct gcaaggggag 900
 cggtgctca ccatgtcctg cagtgacaag attgcacgt ggaacgtggt gggcatccag 960

 ggatccctgc tcagcatttt cgtggagccc atttacttct cgagcatcat cctgggcagc 1020
 ctttaccagc gggaccacct ttccagggcc atgtaccagc ggatctcaa catagaggac 1080
 ctgccacctc tctacacct caacaagcct ttgtcagtg gcatcagcaa tgcagaagca 1140
 cggcagccag ggaaggcccc caacttcagt gtcaactgga cggtaggcga ctccgctatt 1200
 gaggtcatca acgccacgac tgggaaggat gagctgggcc gcgcgtcccg cctgtgtaag 1260
 cagcgttgt actgtcgtg gatgcgtgtg cagggcaagg ttcctccca cttactacgc 1320
 tccaagatta ccaaacccaa cgtgtacat gatccaagc tggcggcaaa ggagtaccag 1380

 gccgccaagg cgcgtctgtt cacagccttc atcaaggcgg ggctgggggc ctgggtggag 1440
 aagcccaccg agcaggacca gttctcactc acgccctccg gtagcgaaac accggggact 1500
 tcagaatcgg ccaccccgga gtctggttct 1530

 <210> 6
 <211> 1582
 <212> DNA
 <213> Artificial Sequence
 <220><223> hADAR2 domain, E488Q mutation (R2D), type 4a
 <400> 6

 cggtctcaat cctccggtag cgaaacaccg gggacttcag aatcggccac cccggagtct 60
 atggccagcc tagcccagcc tctctccct gtcttaccac cattcccacc cccgagtggg 120

 aagaatcccg tgatgatctt gaacgaactg cgeccaggac tcaagtatga ctctctctcc 180
 gagagcgggg agagccatgc caagagcttc gtcatgtctg tggctgtgga tggtcagttc 240
 tttgaaggct cggggagaaa caagaagctt gccaaagccc gggctgcgca gtctgccctg 300
 gccgccattt ttaacttgca cttggatcag acgccatctc gccagcctat tcccagttag 360
 ggccttcagc tgcatctacc gcaggtttta gctgacgtg tctcacgcct ggtcctgggt 420

aagtttgggtg acctgaccga caacttctcc tcccctcacg ctgcgagaaa agtgctggct 480
ggagtcgtca tgacaacagg cacagatgtt aaagatgccca aggtgataag tgtttctaca 540

ggaacaaaat gtattaatgg tgaatacatg agtgatcgtg gccttgcat t aaatgactgc 600
catgcagaaa taatatctcg gagatccttg ctgagatttc tttatacaca acttgagctt 660
tacttaata acaaatatga tcaaaaaaga tccatctttc agaaatcaga gcgagggggg 720
tttaggctga aggagaatgt ccagtttcat ctgtacatca gcacctctcc ctgtggagat 780
gccagaatct tctcaccaca tgagccaatc ctggaagaac cagcagatag acacccaaat 840
cgtaaagcaa gaggacagct acggacccaa atagagtctg gtcaggggac gattccagt 900
cgctccaatg cgagcatcca aacgtgggac ggggtgctgc aaggggagcg gctgctcacc 960

atgtcctgca gtgacaagat tgcacgtggtg aacgtggtgg gcatccaggg atccctgctc 1020
agcatttttc tggagcccat ttacttctcg agcatcatcc tgggcagcct ttaccacggg 1080
gaccaccttt ccagggccat gtaccagcgg atctccaaca tagaggacct gccacctctc 1140
tacacctca acaagccttt gtcagtggc atcagcaatg cagaagcacg gcagccaggg 1200
aaggccccc acttcagtgt caactggacg gtaggcgact ccgtattga ggtcatcaac 1260
gccacgactg ggaaggatga gctgggccgc gcgtcccgc tgtglaagca cgcgttgtac 1320
tgtcgtgga tgcgtgtgca cggcaaggtt cctccact tactacgctc caagattacc 1380

aaaccaacg tgiaccatga gtccaagctg gcggcaaagg agtaccaggc cgccaaggcg 1440
cgtctgttca cagccttcat caagcgggg ctgggggcct ggggtgagaa gcccaccgag 1500
caggaccagt tctcactcac gcccgatcca aaaaagaaga gaaaggtaga tccaaaaaag 1560
aagagaaagg tataatggct ga 1582

<210> 7
<211> 4172
<212> DNA
<213> Artificial Sequence
<220><223> nSpCas9, type 3b
<400> 7

ttctgataaa aagtattcta ttggtttagc catcggcact aattccgttg gatgggctgt 60

cataaccgat gaatacaaag taccttcaaa gaaatttaag gtgttgggga acacagaccg 120
tcattcgatt aaaaagaatc ttatcgggtgc ctcctattc gatagtggcg aaacggcaga 180
ggcgactcgc ctgaaacgaa ccgctcggag aaggtatata cgtcgcaaga accgaatatg 240
ttacttaciaa gaaattttta gcaatgagat ggccaaagtt gacgattctt tctttcaccg 300

tttgggaagag tccttccttg tcgaagagga caagaaacat gaacggcacc ccatctttgg	360
aaacatagta gatgaggtag catatcatga aaagtaccca acgattttatc acctcagaaa	420
aaagctagtt gactcaactg ataaagcgga cctgaggtta atctacttgg ctcttgccca	480
tatgataaag ttccgtgggc actttctcat tgagggtgat ctaaatccgg acaactcgga	540
tgctgacaaa ctgttcaccc agttagtaca aacctataat cagttgtttg aagagaaccc	600
tataaatgca agtggcgtgg atgcgaaggc tattcttagc gcccgctct ctaaatcccg	660
acggctagaa aacctgatcg cacaattacc cggagagaag aaaaatgggt tgttcggtaa	720
ccttatagcg ctctcactag gcctgacacc aaattttaag tcgaacttcg acttagctga	780
agatgcaaaa ttgcagctta gtaaggacac gtacgatgac gatctcgaca atctactggc	840
acaaattgga gatcagtatg cggacttatt ttggctgcc aaaaacctta gcgatgcaat	900
cctcctatct gacatactga gagttaatac tgagattacc aaggcgccgt tatccgcttc	960
aatgatcaaa aggtacgatg aacatcacca agacttgaca cttctcaagg ccctagtccg	1020
tcagcaactg cctgagaaat ataaggaaat attctttgat cagtcgaaaa acgggtacgc	1080
aggttatatt gacggcggag cgagtcaaga ggaattctac aagtttatca aacctatatt	1140
agagaagatg gatgggacgg aagagttgct tgtaaaactc aatcgcaag atctactgcg	1200
aaagcagcgg actttcgaca acggtagcat tccacatcaa atccacttag gcgaattgca	1260
tgctatactt agaaggcagg aggattttta tccgttcctc aaagacaatc gtgaaaagat	1320
tgagaaaatc ctaacctttc gcatacctta ctatgtggga cccctggccc gagggaaactc	1380
tcggttcgca tggatgacaa gaaagtccga agaaacgatt actccatgga attttgagga	1440
agttgtcgat aaaggtgcgt cagctcaatc gtatcatcgag aggatgacca actttgacaa	1500
gaatttaccg aacgaaaaag tattgcctaa gcacagtta ctttacgagt atttcacagt	1560
gtacaatgaa ctacgaaaag ttaagtatgt cactgagggc atgcgtaaac ccgcctttct	1620
aagcggagaa cagaagaaag caatagtaga tctgttattc aagaccaacc gcaaagtgac	1680
agttaagcaa ttgaaagagg actactttaa gaaaattgaa tgcttcgatt ctgtcgagat	1740
ctccggggta gaagatcgat ttaatgcgtc acttggtagc tatcatgacc tcctaaagat	1800
aattaaagat aaggacttcc tggataacga agagaatgaa gatatcttag aagatatagt	1860
gttgactctt accctctttg aagatcgga aatgattgag gaaagactaa aaacatacgc	1920
tcacctgttc gacgataagg ttatgaaaca gttaaagagg cgtcgctata cgggctgggg	1980
acgattgtcg cggaactta tcaacgggat aagagacaag caaagtggta aaactattct	2040
cgattttcta aagagcgacg gcttcgccaa taggaacttt atgcagctga tccatgatga	2100

ctctttaacc ttcaaagagg atatacaaaa ggcacagggt tccggacaag gggactcatt	2160
gcacgaacat attgcgaatc ttgctggttc gccagccatc aaaaaggga tactccagac	2220
agtcaaagta gtggatgagc tagttaaggt catgggacgt cacaaccgg aaaacattgt	2280
aatcgagatg gcacgcgaaa atcaaacgac tcagaagggg caaaaaaca gtcgagagcg	2340
gatgaagaga atagaagagg gtattaaaga actgggcagc cagatcttaa aggagcatcc	2400
tgtggaaaat acccaattgc agaacgagaa actttacctc tattacctac aaaatggaag	2460
ggacatgtat gtgatcagg aactggacat aaaccgttta tctgattacg acgtcgatca	2520
cattgtaccc caatcctttt tgaaggacga ttcaatcgac aataaagtgc ttacacgctc	2580
ggataagaac cgagggaaaa gtgacaatgt tccaagcgag gaagtcgtaa agaaaatgaa	2640
gaactattgg cggcagctcc taaatgcgaa actgataacg caaagaaagt tcgataactt	2700
aactaaagct gagaggggtg gcttgtctga acttgacaag gccggattta ttaaactgca	2760
gctcgtggaa acccgccaaa tcacaaagca tgttgacag atactagatt cccgaatgaa	2820
tacgaaatac gacgagaacg ataagctgat tcgggaagtc aaagtaatca ctttaaagtc	2880
aaaattggtg tcggacttca gaaaggattt tcaattctat aaagttaggg agataaataa	2940
ctaccacat gcgcacgacg cttatcttaa tgccgtcgta gggaccgcac tcattaagaa	3000
ataccgaag ctagaaagtg agtttgtgta tggtgattac aaagtttatg acgtccgtaa	3060
gatgatcgcg aaaagcgaa aggagatagg caaggctaca gccaaatact tcttttattc	3120
taacattatg aatttcttta agacggaaat cactctggca aacggagaga tacgcaaacg	3180
acctttaatt gaaaccaatg gggagacagg tgaatcgta tgggataagg gccgggactt	3240
cgcgacggtg agaaaagttt tgtccatgcc ccaagtcaac atagtaaaga aaactgaggt	3300
gcagaccgga gggttttcaa aggaatcgat tcttccaaaa aggaatagt ataagctcat	3360
cgctcgtaaa aaggactggg acccgaaaaa gtacggtggc ttcgatagcc ctacagttgc	3420
ctattctgtc ctagtagtgg caaaagtga gaagggaata tccaagaaac tgaagtcagt	3480
caaagaatta ttggggataa cgattatgga gcgctcgtct tttgaaaaga accccatcga	3540
cttccttagg gcgaaaggtt acaaggaagt aaaaaaggat ctcataatta aactacaaa	3600
gtatagtctg tttgagttag aaaatggccg aaacggatg ttggctagcg ccggagagct	3660
tcaaaagggg aacgaactcg cactaccgtc taaatactg aatttctgt atttagctc	3720
ccattacgag aagttgaaag gttcacctga agataacgaa cagaagcaac tttttgtga	3780
gcagcacaaa cattatctcg acgaaatcat agagcaaatt tcggaattca gtaagagagt	3840

catcctagct gatgccaatc tggacaaagt attaagcgca tacaacaagc acagggataa 3900
 acccatacgt gagcaggcgg aaaatattat ccatttggtt actcttacca acctcggcgc 3960
 tccagccgca ttcaagtatt ttgacacaac gatagatcgc aaacgataca ctcttaccaa 4020
 ggaggtgcta gacgcgacac tgattcacca atccatcacg ggattatatg aaactcggat 4080
 agatttgta cagcttgggg gtgactctgg tggttctgat ccaaaaaaga agagaaaggt 4140
 agatccaaaa aagaagagaa aggtataaat cc 4172

<210> 8

<211> 4174

<212> DNA

<213> Artificial Sequence

<220><223> nSpCas9, type 3

<400> 8

tatgatggat aaaaagtatt ctattggttt agccatcggc actaattccg ttggatgggc 60
 tgtcataacc gatgaatata aagtaccttc aaagaaatit aaggtgttgg ggaacacaga 120
 ccgtcattcg attaaaaaga atcttatcgg tgccctccta ttcgatagtg gcgaaacggc 180
 agaggcgact cgctgaaac gaaccgctcg gagaaggtat acacgtcgca agaaccgaat 240
 atgttactta caagaaatit ttagcaatga gatggccaaa gttgacgatt ctctctttca 300
 ccgtttggaa gagtccctcc ttgtcgaaga ggacaagaaa catgaacggc accccatctt 360

tggaacata gtagatgagg tggcatatca tgaaaagtac ccaacgattt atcacctcag 420
 aaaaaagcta gttgactcaa ctgataaagc ggacctgagg ttaatctact tggctcttgc 480
 ccatatgata aagttccgtg ggcactttct cattgagggt gatctaaatc cggacaactc 540
 ggatgtcgac aaactgttca tccagttagt acaaacctat aatcagttgt ttgaagagaa 600
 ccctataaat gcaagtggcg tggatgcgaa ggctattctt agcgcccgcc tctctaaatc 660
 ccgacggcta gaaaacctga tcgcacaatt acccgagag aagaaaaatg ggttgttcgg 720
 taaccttata gcgtctcac taggcctgac accaaatit aagtcgaact tcgacttagc 780

tgaagatgcc aaattgcagc ttagtaagga cacgtacgat gacgatctcg acaatctact 840
 ggacaaaatt ggagatcagt atgcggactt atttttggct gccaaaaacc ttagcgatgc 900
 aatctccta tctgacatac tgagagttaa tactgagatt accaaggcgc cgttatccgc 960
 ttcaatgatc aaaaggtacg atgaacatca ccaagacttg acacttctca aggcctagt 1020
 ccgtcagcaa ctgcctgaga aatataagga aatattcttt gatcagtcga aaaacgggta 1080
 cgcaggttat attgacggcg gagcgagtca agaggaattc tacaagtta tcaaaccat 1140

attagagaag atggatggga cggaagagtt gcttgtaaaa ctcaatcgcg aagatctact	1200
gcgaaagcag cggactttcg acaacggttag cattccacat caaatccact taggcgaatt	1260
gcatgctata cttagaaggc aggaggattt ttatccgttc ctcaaagaca atcgtgaaaa	1320
gattgagaaa atcctaacct ttgcataacc ttactatgtg ggacccttg cccgagggaa	1380
ctctcgttgc gcatggatga caagaaagtc cgaagaaacg attactccat ggaattttga	1440
ggaagtgtgc gataaagggtg cgtcagctca atcgttcac gagaggatga ccaactttga	1500
caagaattta ccgaacgaaa aagtattgcc taagcacagt ttactttacg agtatttcac	1560
agtgtacaat gaactcacga aagttaagta tgtcactgag ggcatgcgta aaccgcctt	1620
tctaagcgga gaacagaaga aagcaatagt agatctgtta ttcaagacca accgcaaagt	1680
gacagttaag caattgaaag aggactactt taagaaaatt gaatgcttcg attctgtcga	1740
gatctccggg gtagaagatc gatttaatgc gtcacttggg acgtatcatg acctcctaaa	1800
gataattaaa gataaggact tcctggataa cgaagagaat gaagatatct tagaagatat	1860
agtgttgact cttaccctct ttgaagatcg ggaaatgatt gaggaaagac taaaaacata	1920
cgctcacctg ttcgacgata aggttatgaa acagttaaag aggcgtcgct atacgggctg	1980
gggacgattg tcgcggaac ttatcaacgg gataagagac aagcaaagtg gtaaaactat	2040
tctcgatttt ctaaagagcg acggcttcgc caataggaac tttatgcagc tgatccatga	2100
tgactcttta accttcaaag aggatataca aaaggcacag gtttccggac aaggggactc	2160
attgcacgaa catattgcga atcttgctgg ttgccagcc atcaaaaagg gcatactcca	2220
gacagtcaaa gtagtggatg agctagttaa ggtcatggga cgtcacaac cggaacaacat	2280
tgtaatcgag atggcacgcg aaaatcaaac gactcagaag gggcaaaaa acagtcgaga	2340
gcggatgaag agaatagaag agggatttaa agaactgggc agccagatct taaaggagca	2400
tcctgtggaa aataccaat tgcagaacga gaaactttac ctctattacc tacaaaatgg	2460
aagggacatg tatgttgatc aggaactgga cataaacctt ttatctgatt acgacgtcga	2520
tcacattgta cccaatcct tttgaagga cgattcaatc gacaataaag tgcttacacg	2580
ctcgataag aaccgagggg aaagtgacaa tgttccaagc gaggaagtcg taaagaaaat	2640
gaagaactat tggcggcagc tcctaaatgc gaaactgata acgcaaagaa agttcgataa	2700
cttaactaaa gctgagaggg gtggcttgtc tgaacttgac aaggccggat ttattaaacg	2760
tcagctcgtg gaaacccgcc aaatcacaaa gcatgttgca cagatactag attcccgaat	2820
gaatacgaag tacgacgaga acgataagct gattcgggaa gtcaaagtaa tcactttaaa	2880

gtcaaaattg gtgtcggact tcagaaagga ttttcaattc tataaagtta gggagataaa	2940
taactaccac catgcgcacg acgcttatct taatgccgtc gtagggaccg cactcattaa	3000
gaaatacccg aagctagaaa gtgagtttgt gtatggtgat taaaaagttt atgacgtccg	3060
taagatgac gcgaaaagcg aacaggagat aggcaaggct acagccaaat acttctttta	3120
ttctaacatt atgaatttct ttaagacgga aatcactctg gcaaacggag agatacgcaa	3180
acgaccttta attgaaacca atggggagac aggtgaaatc gtatgggata agggccggga	3240
cttcgcgacg gtgagaaaag ttttgtccat gcccgaagtc aacatagtaa agaaaactga	3300
gggtgcagacc ggaggggtttt caaaggaatc gattcttcca aaaaggaata gtgataagct	3360
catcgctcgt aaaaaggact gggacccgaa aaagtacggt ggcttcgata gccctacagt	3420
tcctatttct gtcctagtag tggcaaaagt tgagaaggga aaatccaaga aactgaagtc	3480
agtcaagaa ttattgggga taacgattat ggagcgctcg tcttttgaaa agaaccctat	3540
cgacttcctt gaggcgaaag gttacaagga agtaaaaaag gatctcataa ttaactacc	3600
aaagtatagt ctgtttgagt tagaaaatgg ccgaaaacgg atgttgcta gcgccggaga	3660
gcttcaaaag gggaacgaac tcgcactacc gtctaaatac gtgaatttcc tgtatttagc	3720
gtcccattac gagaagtga aaggttcacc tgaagataac gaacagaagc aactttttgt	3780
tgagcagcac aaacattatc tcgacgaaat catagagcaa atttcggaat tcagtaagag	3840
agtcaccta gctgatgcca atctggacaa agtattaagc gcatacaaca agcacaggga	3900
taaaccata cgtgagcagg cggaaaatat tatccatttg tttactctta ccaacctcgg	3960
cgtccagcc gcattcaagt attttgacac aacgatagat cgcaaacgat acacttctac	4020
caaggagggt ctagacgca cactgattca ccaatccatc acgggattat atgaaactcg	4080
gatagatttg tcacagcttg ggggtgactc tgggtgttct cccaagaaga agaagagaaa	4140
ggtagatcca aaaaagaaga gaaaggtagg atcc	4174
<210> 9	
<211> 4172	
<212> DNA	
<213> Artificial Sequence	
<220><223> dSpCas9, type 3b	
<400> 9	
ttctgataaa aagtattcta ttggtttagc catcggcact aattccgttg gatgggctgt	60
cataaccgat gaatacaaag taccttcaaa gaaatttaag gtgttgggga acacagaccg	120
tcattcgatt aaaaagaatc ttatcggtgc cctcctattc gatagtggcg aaacggcaga	180

ggcgactcgc ctgaaacgaa ccgctcggag aaggtatata cgtcgcaaga accgaatatg	240
ttacttacaa gaaattttta gcaatgagat ggccaaagtt gacgattctt tctttcaccg	300
tttggaagag tcttctcttg tcgaagagga caagaaacat gaacggcacc ccatcttttg	360
aaacatagta gatgaggtgg catacatga aaagtaccca acgatttatc acctcagaaa	420
aaagctagtt gactcaactg ataaagcgga cctgaggtta atctacttgg ctcttgccca	480
tatgataaag ttccgtgggc actttctcat tgagggtgat ctaaatccgg acaactcgga	540
tgtcgacaaa ctgttcatcc agttagtaca aacctataat cagtgtttg aagagaaccc	600
tataaatgca agtggcgtgg atgcgaaggc tattcttagc gccgcctct ctaaatcccg	660
acggctagaa aacctgatcg cacaattacc cggagagaag aaaaatgggt tgttcggtaa	720
ccttatagcg ctctcactag gcctgacacc aaattttaag tcgaacttcg acttagctga	780
agatgcaaaa ttgcagctta gtaaggacac gtacgatgac gatctcgaca atctactggc	840
acaaattgga gatcagtatg cggacttatt ttggctgcc aaaaacctta gcgatgcaat	900
cctcctatct gacatactga gagttaatac tgagattacc aaggcgccgt tatccgttc	960
aatgatcaaa aggtacgatg aacatcacca agacttgaca cttctcaagg ccctagtccg	1020
tcagcaactg cctgagaaat ataaggaaat attctttgat cagtcgaaaa acgggtacgc	1080
aggttatatt gacggcggag cgagtcaaga ggaattctac aagtttatca aacctatatt	1140
agagaagatg gatgggacgg aagagttgct tgtaaaactc aatcgcaag atctactgcg	1200
aaagcagcgg actttcgaca acggtagcat tccacatcaa atccacttag gcgaattgca	1260
tgctatactt agaaggcagg aggattttta tccgttcctc aaagacaatc gtgaaaagat	1320
tgagaaaaac ctaacctttc gcatacctta ctatgtggga cccctggccc gagggaaactc	1380
tcggttcgca tggatgacaa gaaagtcga agaaacgatt actccatgga attttgagga	1440
agttgtcgat aaaggtgcgt cagctcaatc gttcatcgag aggatgacca actttgacaa	1500
gaatttaccg aacgaaaaag tattgcctaa gcacagtta ctttacgagt atttcacagt	1560
gtacaatgaa ctacgaaaag ttaagtatgt cactgagggc atgcgtaaac ccgcctttct	1620
aagcggagaa cagaagaaag caatagtaga tctgttattc aagaccaacc gcaaagtac	1680
agttaagcaa ttgaaagagg actactttaa gaaaattgaa tgcttcgatt ctgtcgagat	1740
ctccggggta gaagatcgat ttaatgcgtc acttggtagc tatcatgacc tcctaaagat	1800
aattaaagat aaggacttcc tggataacga agagaatgaa gatatcttag aagatatagt	1860
gttgactctt accctctttg aagatcggga aatgattgag gaaagactaa aaacatacgc	1920

tcacctgttc gacgataagg ttatgaaaca gttaaagagg cgtcgctata cgggctgggg	1980
acgattgtcg cggaacctta tcaacgggat aagagacaag caaagtggta aaactattct	2040
cgattttcta aagagcgacg gcttcgccaa taggaacttt atgcagctga tccatgatga	2100
ctctttaacc ttcaaagagg atatacaaaa ggcacagggt tccggacaag gggactcatt	2160
gcacgaacat attgcgaatc ttgctgggtc gccagccatc aaaaaggga tactccagac	2220
agtcaaagta gtggatgagc tagttaaggt catgggacgt cacaaccgg aaaacattgt	2280
aatcgagatg gcacgcgaaa atcaaacgac tcagaagggg caaaaaaca gtcgagagcg	2340
gatgaagaga atagaagagg gtattaaaga actgggcagc cagatcttaa aggagcatcc	2400
tgtggaaaat acccaattgc agaacgagaa actttacctc tattacctac aaaatggaag	2460
ggacatgtat gtgatcagg aactggacat aaaccgttta tctgattacg acgtcgatgc	2520
cattgtacc ccaatctttt tgaaggacga ttcaatcgac aataaagtgc ttacacgctc	2580
ggataagaac cgagggaaaa gtgacaatgt tccaagcgag gaagtcgtaa agaaaatgaa	2640
gaactattgg cggcagctcc taaatgcgaa actgataacg caaagaaagt tcgataactt	2700
aactaaagct gagaggggtg gcttgtctga acttgacaag gccggattta ttaaactgca	2760
gctcgtggaa acccgccaaa tcacaaagca tgttgacag atactagatt cccgaatgaa	2820
tacgaaatac gacgagaacg ataagctgat tcgggaagtc aaagtaatca ctttaaagtc	2880
aaaattgggtg tcggacttca gaaaggattt tcaattctat aaagttaggg agataaataa	2940
ctaccacat gcgcacgacg cttatcttaa tgccgtcgta gggaccgcac tcattaagaa	3000
ataccggaag ctagaaagtg agtttgtgta tgggtattac aaagtttatg acgtccgtaa	3060
gatgatcgcg aaaagcgaa aggagatagg caaggctaca gccaataact tcttttattc	3120
taacattatg aatttcttta agacggaaat cactctggca aacggagaga tacgcaaacg	3180
acctttaatt gaaaccaatg gggagacagg tgaatcgta tgggataagg gccgggactt	3240
cgcgacgggtg agaaaagttt tgtccatgcc ccaagtcaac atagtaaaga aaactgaggt	3300
gcagaccgga gggttttcaa aggaatcgat tcttcaaaa aggaatagtg ataagctcat	3360
cgctcgtaaa aaggactggg acccgaaaaa gtacggtggc ttcgatagcc ctacagtgc	3420
ctattctgtc ctagtagtgg caaaagtga gaagggaata tccaagaaac tgaagtcagt	3480
caaagaatta ttggggataa cgattatgga gcgctcgtct tttgaaaaga accccatcga	3540
cttccttgag gcgaaaggtt acaaggaagt aaaaaaggat ctcataatta aactacaaa	3600
gtatagtctg tttgagttag aaaatggccg aaaacggatg ttggctagcg ccggagagct	3660
tcaaaagggg aacgaactcg cactaccgtc taaatactg aatttctgt atttagcgtc	3720
ccattacgag aagttgaaag gttcacctga agataacgaa cagaagcaac tttttgtga	3780

gcagcacaaa cattatctcg acgaaatcat agagcaaatt tcggaattca gtaagagagt 3840
catcctagct gatgccaatc tggacaaagt attaagcgca tacaacaagc acagggataa 3900
acccatacgt gagcaggcgg aaaatattat ccatttgttt actcttacca acctcggcgc 3960
tccagccgca ttcaagtatt ttgacacaac gatagatcgc aaacgataca ctcttaccaa 4020

ggaggtgcta gacgcgacac tgattcacca atccatcacg ggattatatg aaactcggat 4080
agatttgtca cagcttgggg gtgactctgg tggttctgat ccaaaaaaga agagaaaggt 4140
agatccaaaa aagaagagaa aggtataaat cc 4172

<210> 10
<211> 4174
<212> DNA
<213> Artificial Sequence
<220><223> dSpCas9, type 3
<400> 10

tatgatggat aaaaagtatt ctattggttt agccatcggc actaattccg ttggatgggc 60
tgtcataacc gatgaataca aagtacctic aaagaaattt aaggtgttgg ggaacacaga 120

ccgtcattcg attaaaaaga atcttatcgg tgcctccta ttcgatagt gcgaaacggc 180
agaggcgact cgctgaaac gaaccgctcg gagaaggtat acacgtcgca agaaccgaat 240
atgttactta caagaaattt ttagcaatga gatggccaaa gttgacgatt ctttctttca 300
ccgtttggaa gattccttcc ttgtcgaaga ggacaagaaa catgaacggc accccatctt 360
tgaaacata gtagatgagg tggcatatca tgaaggtac ccaacgattt atcacctcag 420
aaaaagcta gttgactcaa ctgataaagc ggacctgagg ttaatctact tggctcttgc 480
ccatatgata aagttccgtg ggcactttct cattgagggt gatctaaatc cggacaactc 540

ggatgtcgac aaactgttca tccagttagt acaaacctat aatcagttgt ttgaagagaa 600
ccctataaat gcaagtggcg tggatgcgaa ggctattctt agcggccgcc tctctaaatc 660
ccgacggcta gaaaacctga tcgcacaatt acccggagag aagaaaaatg ggttgttcgg 720
taaccttata gcgtcttcac taggcctgac accaaatttt aagtcgaact tcgacttagc 780
tgaagatgcc aaattgcagc ttagtaagga cacgtacgat gacgatctcg acaatctact 840
ggcacaatt ggagatcagt atgcggactt atttttggct gccaaaaacc ttagcgatgc 900
aatcctccta tctgacatac tgagagttaa tactgagatt accaaggcgc cgttatccgc 960

ttcaatgac aaaaggtacg atgaacatca ccaagacttg acacttctca aggccctagt 1020
ccgtcagcaa ctgcctgaga aatataagga aatattcttt gatcagtcga aaaacgggta 1080

cgcaggttat attgacggcg gagcgagtca agaggaattc tacaagtta tcaaaccat	1140
attagagaag atggatggga cggaagagtt gcttgtaaaa ctcaatcgcg aagatctact	1200
gcgaaagcag cggactttcg acaacggtag cattccacat caaatccact taggcgaatt	1260
gcatgctata ctiagaaggc aggaggattt ttatccgttc ctcaaagaca atcgtgaaaa	1320
gattgagaaa atcctaacct ttgcatacc ttactatgtg ggaccctgg cccgaggga	1380
ctctcggttc gcatggatga caagaaagtc cgaagaaacg attactccat ggaattttga	1440
ggaagttgtc gataaagggtg cgtcagctca atcgttcadc gagaggatga ccaactttga	1500
caagaattta ccgaacgaaa aagtattgcc taagcacagt ttactttacg agtatttcac	1560
agtgtacaat gaactcacga aagttaagta tgcactgag ggcatgcgt aacccgcctt	1620
tctaagcgga gaacagaaga aagcaatagt agatctgtta ttcaagacca accgcaaagt	1680
gacagttaag caattgaaag aggactactt taagaaaatt gaatgcttcg attctgtcga	1740
gatctccggg gtagaagatc gatttaatgc gtcacttgg acgtatcatg acctcctaaa	1800
gataattaaa gataaggact tcctggataa cgaagagaat gaagatatct tagaagatat	1860
agtgttgact ctaccctct ttgaagatcg ggaatgatt gaggaagac taaaaacata	1920
cgctcacctg ttcgacgata aggttatgaa acagttaaag aggcgtcgct atacgggctg	1980
gggacgattg tcgcggaac ttatcaacgg gataagagac aagcaaagt gtaaaactat	2040
tctcgatttt ctaaagagcg acggcttcgc caataggaac tttatgcagc tgatecatga	2100
tgactcttta accttcaaag aggatataca aaaggcacag gtttccggac aaggggactc	2160
attgcacgaa catattgcga atcttgctgg ttcgccagcc atcaaaaagg gcatactcca	2220
gacagtcaaa gtagtggatg agctagttaa ggtcatggga cgtcacaac cggaaaacat	2280
tgtaatcgag atggcacgcg aaaatcaaac gactcagaag gggcaaaaaa acagtcgaga	2340
gcggatgaag agaatagaag agggatttaa agaactgggc agccagatct taaaggagca	2400
tcctgtggaa aataccaat tgcagaacga gaaactttac ctctattacc tacaaaatgg	2460
aaggacatg tatgttgatc aggaactgga cataaacctt ttatctgatt acgacgtcga	2520
tgccattgta cccaatcct ttttgaagga cgattcaatc gacaataaag tgcttacacg	2580
ctcgataag aaccgaggga aaagtgacaa tgttccaagc gaggaagtcg taaagaaat	2640
gaagaactat tggcggcagc tcctaaatgc gaaactgata acgcaaagaa agttcgataa	2700
cttaactaaa gctgagagggt gtggcttgic tgaacttgac aaggccggat ttattaaacg	2760
tcagctcgtg gaaacccgcc aaatcacaaa gcatgttgca cagatactag attcccgaat	2820
gaatacgaaa tacgacgaga acgataagct gattcgggaa gtcaaagtaa tcactttaaa	2880
gtcaaaattg gtgtcggact tcagaaagga ttttcaattc tataaagtta gggagataaa	2940

taactaccac catgcgcacg acgcttatct taatgccgtc gtagggaccg cactcattaa 3000
gaaataccg aagctagaaa gtgagtttgt gtatggtgat tacaagttt atgacgtccg 3060

taagatgac gcgaaaagcg aacaggagat aggcaaggct acagccaaat acttctttta 3120
ttctaacatt atgaatttct ttaagacgga aatcactctg gcaaaccgag agatacgcaa 3180
acgaccttta attgaaacca atggggagac aggtgaaatc gtatgggata agggccggga 3240
cttcgcgacg gtgagaaaaa ttttgtccat gccccaaagc aacatagtaa agaaaactga 3300
ggtgcagacc ggagggtttt caaaggaatc gattcttcca aaaaggaata gtgataagct 3360
catcgctcgt aaaaaggact gggaccgcaa aaagtacggt ggcttcgata gccctacagt 3420
tgcctattct gtccatagtag tggcaaaagt tgagaaggga aaatccaaga aactgaagtc 3480

agtcaaagaa ttattgggga taacgattat ggagcgctcg tcttttgaaa agaaccctat 3540
cgacttcctt gaggcgaaaag gttacaagga agtaaaaaag gatctcataa ttaaaactacc 3600
aaagtatagt ctgtttgagt tagaaaatgg ccgaaaacgg atgttggcta gcgccggaga 3660
gcttcaaaaag gggaacgaac tcgcactacc gtctaaatac gtgaatttcc tgtatttagc 3720
gtccatttac gagaagttga aaggttcacc tgaagataac gaacagaagc aactttttgt 3780
tgagcagcac aaacattatc tcgacgaaat catagagcaa atttcggaat tcagtaagag 3840
agtcaccta gctgatgcca atctggacaa agtattaagc gcatacaaca agcacaggga 3900

taaaccata cgtgagcagg cggaataat tatccatttg tttactctta ccaacctcgg 3960
cgctccagcc gcattcaagt attttgacac aacgatagat cgcaaacgat acacttctac 4020
caaggagggtg ctagacgga cactgattca ccaatccatc acgggattat atgaaactcg 4080
gatagatttg tcacagcttg ggggtgactc tgggtgttct cccaagaaga agaagagaaa 4140
ggtagatcca aaaaagaaga gaaagtagg atcc 4174

<210> 11
<211> 410
<212> PRT
<213> Homo sapiens
<400> 11

Gly Glu Ala Glu Asn Ala Gly Ser Tyr Ser Leu Gln Gln Ala Gln Ala

1 5 10 15
Phe Tyr Thr Phe Pro Phe Gln Gln Leu Met Ala Glu Ala Pro Asn Met
20 25 30
Ala Val Val Asn Glu Gln Gln Met Pro Glu Glu Val Pro Ala Pro Ala

35 40 45
 Pro Ala Gln Glu Pro Val Gln Glu Ala Pro Lys Gly Arg Lys Arg Lys
 50 55 60
 Pro Arg Thr Thr Glu Pro Lys Gln Pro Val Glu Pro Lys Lys Pro Val
 65 70 75 80

 Glu Ser Lys Lys Ser Gly Lys Ser Ala Lys Ser Lys Glu Lys Gln Glu
 85 90 95
 Lys Ile Thr Asp Thr Phe Lys Val Lys Arg Lys Val Asp Arg Phe Asn
 100 105 110
 Gly Val Ser Glu Ala Glu Leu Leu Thr Lys Thr Leu Pro Asp Ile Leu
 115 120 125
 Thr Phe Asn Leu Asp Ile Val Ile Ile Gly Ile Asn Pro Gly Leu Met
 130 135 140
 Ala Ala Tyr Lys Gly His His Tyr Pro Gly Pro Gly Asn His Phe Trp

 145 150 155 160
 Lys Cys Leu Phe Met Ser Gly Leu Ser Glu Val Gln Leu Asn His Met
 165 170 175
 Asp Asp His Thr Leu Pro Gly Lys Tyr Gly Ile Gly Phe Thr Asn Met
 180 185 190
 Val Glu Arg Thr Thr Pro Gly Ser Lys Asp Leu Ser Ser Lys Glu Phe
 195 200 205
 Arg Glu Gly Gly Arg Ile Leu Val Gln Lys Leu Gln Lys Tyr Gln Pro
 210 215 220

 Arg Ile Ala Val Phe Asn Gly Lys Cys Ile Tyr Glu Ile Phe Ser Lys
 225 230 235 240
 Glu Val Phe Gly Val Lys Val Lys Asn Leu Glu Phe Gly Leu Gln Pro
 245 250 255
 His Lys Ile Pro Asp Thr Glu Thr Leu Cys Tyr Val Met Pro Ser Ser
 260 265 270
 Ser Ala Arg Cys Ala Gln Phe Pro Arg Ala Gln Asp Lys Val His Tyr
 275 280 285

Tyr Ile Lys Leu Lys Asp Leu Arg Asp Gln Leu Lys Gly Ile Glu Arg

290 295 300
 Asn Met Asp Val Gln Glu Val Gln Tyr Thr Phe Asp Leu Gln Leu Ala
 305 310 315 320
 Gln Glu Asp Ala Lys Lys Met Ala Val Lys Glu Glu Lys Tyr Asp Pro
 325 330 335
 Gly Tyr Glu Ala Ala Tyr Gly Gly Ala Tyr Gly Glu Asn Pro Cys Ser
 340 345 350
 Ser Glu Pro Cys Gly Phe Ser Ser Asn Gly Leu Ile Glu Ser Val Glu
 355 360 365

Leu Arg Gly Glu Ser Ala Phe Ser Gly Ile Pro Asn Gly Gln Trp Met
 370 375 380
 Thr Gln Ser Phe Thr Asp Gln Ile Pro Ser Phe Ser Asn His Cys Gly
 385 390 395 400
 Thr Gln Glu Gln Glu Glu Glu Ser His Ala
 405 410

<210> 12
 <211> 48
 <212> DNA
 <213> Artificial Sequence
 <220><223> XTEN linker
 <400> 12

tccggtagcgc aaacaccggg gacttcagaa tcggccaccc cggagtct 48

<210> 13
 <211> 48
 <212> DNA
 <213> Artificial Sequence
 <220><223> 2xSV40NLS
 <400> 13

gatccaaaaa agaagagaaa ggtagatcca aaaaagaaga gaaaggta 48

<210> 14
 <211> 200

<212> PRT

<213> Artificial Sequence

<220><223> hAPOBEC3A mutation 1

<400> 14

Met Glu Ala Ser Pro Ala Ser Gly Pro Arg His Leu Met Asp Pro His

1 5 10 15

Ile Phe Thr Ser Asn Phe Asn Asn Gly Ile Gly Arg His Lys Gly Tyr

20 25 30

Leu Cys Tyr Glu Val Glu Arg Leu Asp Asn Gly Thr Ser Val Lys Met

35 40 45

Asp Gln His Arg Gly Phe Leu His Asn Gln Ala Lys Asn Leu Leu Cys

50 55 60

Gly Phe Tyr Gly Arg His Ala Gly Glu Leu Arg Phe Leu Asp Leu Val

65 70 75 80

Pro Ser Leu Gln Leu Asp Pro Ala Gln Ile Tyr Arg Val Thr Trp Phe

85 90 95

Ile Ser Trp Ser Asp Cys Phe Ser Trp Gly Cys Ala Gly Glu Val Arg

100 105 110

Ala Phe Leu Gln Glu Asn Thr His Val Arg Leu Arg Ile Phe Ala Ala

115 120 125

Arg Ile Gly Asp Tyr Asp Pro Leu Tyr Lys Glu Ala Leu Gln Met Leu

130 135 140

Arg Asp Ala Gly Ala Gln Val Ser Ile Met Thr Tyr Asp Glu Phe Lys

145 150 155 160

His Cys Trp Asp Thr Phe Val Asp His Gln Gly Cys Pro Phe Gln Pro

165 170 175

Trp Asp Gly Leu Asp Glu His Ser Gln Ala Leu Ser Gly Arg Leu Arg

180 185 190

Ala Ile Leu Gln Asn Gln Gly Asn

195 200

<210> 15

<211> 196

<212> PRT

<213> Artificial Sequence

<220><223> hAPOBEC3A mutation 2

<400> 15

Met Glu Ala Ser Pro Ala Ser Gly Pro Arg His Leu Met Asp Pro His

1 5 10 15

Ile Phe Thr Ser Asn Phe Asn Asn Gly Ile Gly Arg His Lys Gly Tyr

20 25 30

Leu Cys Tyr Glu Val Glu Arg Leu Asp Asn Gly Thr Ser Val Lys Met

35 40 45

Asp Gln His Arg Gly Phe Leu His Asn Gln Ala Lys Asn Leu Leu Cys

50 55 60

Gly Phe Tyr Gly Arg His Ala Gly Glu Leu Arg Phe Leu Asp Leu Val

65 70 75 80

Pro Ser Leu Gln Leu Asp Pro Ala Gln Ile Tyr Arg Val Thr Trp Phe

85 90 95

Ile Ser Trp Ser Asp Cys Ser Gly Cys Ala Val Arg Ala Phe Leu Gln

100 105 110

Glu Asn Thr His Val Arg Leu Arg Ile Phe Ala Ala Arg Ile Gly Asp

115 120 125

Tyr Asp Pro Leu Tyr Lys Glu Ala Leu Gln Met Leu Arg Asp Ala Gly

130 135 140

Ala Gln Val Ser Ile Met Thr Tyr Asp Glu Phe Lys His Cys Trp Asp

145 150 155 160

Thr Phe Val Asp His Gln Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu

165 170 175

Asp Glu His Ser Gln Ala Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln

180 185 190

Asn Gln Gly Asn

195