

명세서

발명의 명칭: 표적화된 탈아미노효소 및 이를 이용한 염기 교정

기술분야

- [1] 본 발명은 분리된 형태의 시토신 또는 아데닌 탈아미노효소 또는 이의 변이체, 비독성 전장 시토신 탈아미노효소 또는 이의 변이체, 이를 포함하는 융합단백질, 염기 교정용 조성물 및 이를 이용하여 염기를 교정하는 방법에 관한 것이다.

[2]

배경기술

- [3] DNA 결합 단백질과 탈아미노효소가 연결된 융합단백질은 DNA 이중 가닥 절단 (DSBs)을 생성하지 않고 유전체에서 표적화된 뉴클레오타이드 치환 또는 염기 교정(base editing), 유전적 장애를 유발하는 점 돌연변이를 교정하거나 원핵세포, 인간 및 다른 진핵세포에 목적하는 단일 뉴클레오타이드 변이를 도입하도록 표적화된 방식(targeted manner)으로 단일 뉴클레오타이드 전환을 가능하게 한다.
- [4] 표적 부위에 작은 삽입 또는 결실(indels)을 유도하는 CRISPR-Cas9와 같은 뉴클레아제와 달리 표적 부위에서의 수 개의 뉴클레오타이드 (window of several nucleotides) 내에서 단일 염기를 변환시킨다. 따라서, 배양 세포, 동물 및 식물에서 유전 질환을 유발하는 점 돌연변이를 교정하거나 단일 염기 다형성 (SNP)을 생성할 수 있다.
- [5] DNA 결합 단백질과 탈아미노효소가 연결된 융합단백질로 1) *S. pyogenes*에서 유래하는 촉매적으로 결핍된 Cas9 (catalytically-deficient Cas9; dCas9) 또는 D10A Cas9 니케이즈 (nCas9)와, 래트의 시토신 탈아미노효소인 rAPOBEC1를 포함하는 베이스 에디터 (Base Editors; BEs); 2) dCas9 또는 nCas9와, 바다칠성장어(sea lamprey)의 AID (activation-induced cytidine deaminase) ortholog인 PmCDA1 또는 인간 AID를 포함하는 Target-AID; 3) MS2-결합 단백질에 융합된 과활성화된 AID 변이체를 모집하기 위해 MS2 RNA 헤어핀에 연결된 sgRNAs와 dCas9를 포함하는 CRISPR-X 등이 예시될 수 있다.
- [6] 이와 같이 ZFN (zinc-finger nuclease), TALEN (transcription activator-like effector nuclease), CRISPR (clustered regularly interspaced short palindromic repeat) 시스템 및 핵산분해 효율이 없는 크리스퍼 연관 단백질 9 (Cas9) 변이체와 염기 탈아미노효소 단백질의 염기 교정 기술 등의 프로그램 (설정) 유전체 교정 도구들은 염기 서열의 변화를 통하여 식물의 유전적 연구와 작물 형질 개선에 발전되어 왔다. 그러나 이러한 도구들은 미토콘드리아와 엽록체를 포함한 식물 세포 기관의 DNA 서열을 교정하기에는 적합하지 않은데, 그 주된 이유는 가이드 RNA를 소기관으로 전달하거나, 두 화합물을 소기관에서 동시에 발현하는 것이 어렵기 때문이다. 식물 소기관은 광합성 및 호흡에 필요한 여러 필수적인

유전자를 암호화하고 있다. 이러한 소기관의 유전자를 교정하는 방법이나 도구는 이러한 유전자의 기능 연구나 작물 생산성과 형질 개선에 매우 필요하다. 예를 들어, 미토콘드리아 *atp6* 유전자의 표적 돌연변이는 번식에 유용한 특성인 응성불임을 유발할 수 있고, 엽록체 게놈의 16S rRNA 유전자의 특정 점 돌연변이는 항생제 내성을 유발할 수 있다.

- [7] 박테리아 독소 DddA_{tox}는 *Burkholderia cenocepacia*에서 유래된 박테리아 독소의 효소 기능을 맡는 부분으로, 이중 나선 DNA의 시토신을 탈아미노화 할 수 있다. 탈아미노효소의 예시로, DddA_{tox}는 세포에 독성이 있기 때문에 숙주세포에서의 독성을 피하기 위해, DddA_{tox}는 두 개의 비활성화된 분할체(split)로 나누어 사용하는데, 각각의 반체는 DNA에 결합할 수 있도록 고안된 DNA 결합 단백질에 연결하여 기능을 갖춘 DdCBE 쌍으로써 사용할 수 있다.
- [8] 원리적으로는 이러한 탈아미노화의 효소 반응은 두 개의 비활성화된 반체가 DNA 결합 단백질에 의해 목표 DNA 위에 가깝게 존재하여야 활성화되게 된다. 따라서 시토신에서 티민으로의 염기 교정은 두 개의 DNA 결합 단백질 결합 자리의 중간에서 일어나게 된다. DNA 결합 단백질로 TALE (transcription activator-like effector) 어레이에 결합된 두 개의 비활성 형태는 TALE에 의해 표적 DNA에 근처에서 결합될 때에 기능을 가진다. 시토신에서 티민으로 염기 변환은 두 TALE 결합 부위 사이의 14~18개 염기 사이의 영역에서 유도된다. 이러한 두개로 나눈 분할 형태 DddA_{tox}는 실험에 많은 제약이 따른다.
- [9] 전장 DddA_{tox}는 독성 때문에 일반적인 대장균을 이용한 클로닝은 불가능하고 독성을 방해하는 면역(immunity) 유전자를 함께 발현하는 대장균을 사용하여 클로닝한다.
- [10] 한편, 미토콘드리아 DNA는 세포 호흡에 아주 중요한 역할을 하는데, 미토콘드리아 산화적 인산화(OXPHOS) 기작을 통해 이루어진다. OXPHOS 기작은 생존에 필수적이기 때문에, 미토콘드리아 DNA의 돌연변이는 많은 에너지를 요구하는 조직과 같은 여러 장거나 근육에 심각한 기능 부전을 일으킬 수 있다. 주로 사람의 미토콘드리아 질병에서, 정상 미토콘드리아 DNA와 단일 염기 돌연변이가 있는 미토콘드리아 DNA는 공존하는데, 이는 미토콘드리아 DNA의 수적인 이형 상태(헤테로플라스미)를 일으킨다. 돌연변이와 정상 미토콘드리아 DNA의 균형이 임상적인 증상이 있는 미토콘드리아 질병의 발달을 결정한다. 생체 외 및 생체 내에서, 프로그래밍이 가능한 핵산가수분해효소를 돌연변이 미토콘드리아 DNA는 자르고 정상 미토콘드리아 DNA를 제거하지 않는 방식으로 사용하고 있었다. 그러나, 이러한 핵산가수분해효소는 특정한 돌연변이를 미토콘드리아에 도입하거나 되돌릴 수는 없는데, 미토콘드리아에서 DNA의 이중 가닥 절단이 핵과 같이 비상동 말단접합이나 상동 재조합을 통해 효율적으로 복구되지 않기 때문일 것이다.
- [11] 미토콘드리아 염기 교정을 통해 지금까지 시도되지 못했던 다양한 질병들에

대한 모델을 만들거나, 이를 치료하는 치료제를 제작하는데 도움을 줄 수 있다. 이러한 측면에서 고효율의 미토콘드리아 염기 교정 효소의 개발의 필요성이 증대되고 있다.

- [12] 이러한 기술적 배경하에서, 탈아미노화 효소의 잔기를 치환하여 비선택적인 염기 교정을 줄이거나, 신규 전장 탈아미노효소를 통해 독성을 가지지 않도록 하여, 목적하는 CBE(cytosine base editor) 또는 ABE(adenine base editor)로 사용할 수 있으며, DNA를 교정할 수 있음을 확인하고, 본 발명을 완성하였다.

[13]

[14] **발명의 요약**

- [15] 본 발명의 목적은 DNA 결합 단백질, 분리된 형태의 시토신 또는 아데닌 탈아미노효소 또는 이의 변이체, 비독성 전장 시토신 탈아미노효소 또는 이의 변이체를 포함하는 융합단백질을 제공하는데 있다.

- [16] 본 발명의 목적은 상기 융합단백질을 코딩하는 핵산을 제공하는데 있다.

- [17] 본 발명의 목적은 상기 융합단백질 또는 핵산을 포함하는 염기 교정용 조성물을 제공하는데 있다.

- [18] 본 발명의 목적은 상기 조성물을 세포에 처리하는 단계를 포함하는 염기 교정 방법을 제공하는데 있다.

- [19] 상기 목적을 달성하기 위하여, 본 발명은 (i) DNA 결합 단백질; 및 (ii) 시토신 탈아미노효소 (deaminase) 또는 이의 변이체 유래인 제1분할체 및 제2분할체를 포함하는 융합단백질로, 상기 제1분할체 및 제2분할체는 각각 상기 DNA 결합 단백질에 결합되는 형태인 융합단백질을 제공한다.

- [20] 본 발명은 (i) DNA 결합 단백질; 및 (ii) 시토신 탈아미노효소 또는 이의 변이체 유래 비독성 전장 시토신 탈아미노효소를 포함하는 융합단백질을 제공한다.

- [21] 본 발명은 (i) DNA 결합 단백질; (ii) 시토신 탈아미노효소 (deaminase) 또는 이의 변이체; 및 (iii) 아데닌 탈아미노효소를 포함하는 융합단백질로, 상기 시토신 탈아미노효소 또는 이의 변이체는 (a) 비독성 전장 시토신 탈아미노효소 또는 (b) 시토신 탈아미노효소 또는 이의 변이체 유래인 제1분할체 및 제2분할체를 포함하고, 상기 제1분할체 및 제2분할체는 각각 상기 DNA 결합 단백질에 결합되는 형태인, 융합단백질을 제공한다.

- [22] 본 발명은 상기 융합단백질을 코딩하는 핵산을 제공한다.

- [23] 본 발명은 상기 융합단백질 또는 핵산을 포함하는 염기 교정용 조성물을 제공한다.

- [24] 본 발명은 상기 융합단백질 또는 핵산을 포함하는 진핵세포 염기 교정용 조성물을 제공한다.

- [25] 본 발명은 또한, 상기 융합단백질 또는 핵산; 및 NLS (nuclear localization signal) 펩타이드 또는 이를 코딩하는 핵산을 포함하는 식물세포 염기 교정용 조성물을 제공한다.

- [26] 본 발명은 또한, 상기 융합단백질 또는 핵산; 및 엽록체 전달

- 웹타이드(chloroplast transit peptide) 또는 이를 코딩하는 핵산을 포함하는 식물세포 염기 교정용 조성물을 제공한다.
- [27] 본 발명은 또한, 상기 융합단백질 또는 핵산; 및 미토콘드리아 전달 단백질 (Mitochondrial Targeting Signal, MTS) 또는 이를 코딩하는 핵산을 포함하는 식물세포 염기 교정용 조성물을 제공한다.
- [28] 경우에 따라서 본 발명은 또한, 핵외 수송 신호 단백질 또는 이를 코딩하는 핵산을 추가로 포함하는, 식물세포 염기 교정용 조성물을 제공한다.
- [29] 본 발명은 또한, 상기 조성물을 식물세포에 처리하는 단계를 포함하는 식물세포 염기 교정방법을 제공한다.
- [30] 본 발명은 또한, 상기 융합단백질 또는 핵산; 및 NLS (nuclear localization signal) 웹타이드 또는 이를 코딩하는 핵산을 식물세포에 처리하는 단계를 포함하는 식물세포 염기 교정방법을 제공한다.
- [31] 본 발명은 또한, 상기 융합단백질 또는 핵산; 및 엽록체 전달 웹타이드(chloroplast transit peptide) 또는 이를 코딩하는 핵산을 식물세포에 처리하는 단계를 포함하는 식물세포 염기 교정방법을 제공한다.
- [32] 본 발명은 또한, 상기 융합단백질 또는 핵산; 및 미토콘드리아 전달 단백질 (Mitochondrial Targeting Signal, MTS) 또는 이를 코딩하는 핵산을 식물세포에 처리하는 단계를 포함하는 식물세포 염기 교정방법을 제공한다.
- [33] 본 발명은 또한, 상기 융합단백질 또는 핵산; 및 NLS (nuclear localization signal) 웹타이드 또는 이를 코딩하는 핵산을 포함하는, 동물세포 염기 교정용 조성물을 제공한다.
- [34] 본 발명은 또한, 상기 융합단백질 또는 핵산; 및 미토콘드리아 전달 단백질 (Mitochondrial Targeting Signal, MTS) 또는 이를 코딩하는 핵산을 포함하는, 동물세포 염기 교정용 조성물을 제공한다.
- [35] 본 발명은 또한, 경우에 따라서 핵외 수송 신호 단백질 또는 이를 코딩하는 핵산을 추가로 포함하는, 동물세포 염기 교정용 조성물을 제공한다.
- [36] 본 발명은 또한, 상기 조성물을 동물세포에 처리하는 단계를 포함하는 동물세포 염기 교정방법을 제공한다.
- [37] 본 발명은 또한, 상기 융합단백질 또는 핵산; 및 NLS (nuclear localization signal) 웹타이드 또는 이를 코딩하는 핵산을 동물세포에 처리하는 단계를 포함하는 동물세포의 염기 교정방법을 제공한다.
- [38] 본 발명은 또한, 상기 융합단백질 또는 핵산; 및 미토콘드리아 전달 단백질 (Mitochondrial Targeting Signal, MTS) 또는 이를 코딩하는 핵산을 동물세포에 처리하는 단계를 포함하는 동물세포의 염기 교정방법을 제공한다.
- [39] 본 발명은 또한, 상기 융합단백질 또는 이를 코딩하는 핵산을 포함하는 원핵세포 또는 진핵세포에서의 A-to-G 염기 교정용 조성물로, 상기 DNA 결합 단백질은 징크 핑거 단백질 (zinc finger protein), TALE 단백질 또는 CRISPR 연관 뉴클레아제이고, 상기 융합단백질의 시토신 탈아미노효소 또는 이의 변이체는

박테리아 유래이고, 이중 가닥 DNA 특이적인 것을 특징으로 하는 조성물을 제공한다.

- [40] 본 발명은 또한, 상기 융합단백질 또는 이를 코딩하는 핵산을 포함하는 원핵세포 또는 진핵세포에서의 A-to-G 염기 교정용 조성물로, 상기 DNA 결합 단백질은 징크 핑커 단백질 (zinc finger protein), TALE 단백질 또는 CRISPR 연관 뉴클레아제이고, 상기 융합단백질의 시토신 탈아미노효소 또는 이의 변이체는 박테리아 유래이고, 이중 가닥 DNA 특이적인 것이며, 상기 시토신 탈아미노효소 또는 이의 변이체의 N 말단에 DNA 결합 단백질이 결합되어 있고, 상기 융합단백질의 아데닌 탈아미노효소의 C 말단에 DNA 결합 단백질이 결합되어 있는 조성물을 제공한다.
- [41] 본 발명은 또한, 상기 융합단백질 또는 이를 코딩하는 핵산 및 UGI (uracil glycosylase inhibitor)를 포함하는 원핵세포 또는 진핵세포에서의 C-to-T 염기 교정용 조성물로, 상기 DNA 결합 단백질은 징크 핑커 단백질 (zinc finger protein), TALE 단백질 또는 CRISPR 연관 뉴클레아제이고, 상기 융합단백질의 시토신 탈아미노효소 또는 이의 변이체는 전장 비독성 시토신 탈아미노효소이고, 상기 융합단백질의 시토신 탈아미노효소 또는 이의 변이체는 박테리아 유래이고, 이중 가닥 DNA 특이적인 것을 특징으로 하는 조성물을 제공한다.
- [42] 본 발명은 또한, 상기 융합단백질 또는 이를 코딩하는 핵산 및 UGI (uracil glycosylase inhibitor)를 포함하는 원핵세포 또는 진핵세포에서의 C-to-T 염기 교정용 조성물로, 상기 DNA 결합 단백질은 징크 핑커 단백질 (zinc finger protein), TALE 단백질 또는 CRISPR 연관 뉴클레아제이고, 상기 융합단백질의 시토신 탈아미노효소 또는 이의 변이체는 제1분할체 및 제2분할체를 포함하는 분할 시토신 탈아미노효소이고, 상기 융합단백질의 시토신 탈아미노효소 또는 이의 변이체는 박테리아 유래이고, 이중 가닥 DNA 특이적인 것을 특징으로 하는 조성물을 제공한다.
- [43] 본 발명은 또한, 상기 융합단백질 또는 이를 코딩하는 핵산을 원핵세포 또는 진핵세포에 처리하는 단계를 포함하는 원핵세포 또는 진핵세포에서의 A-to-G 염기 교정 방법으로, 상기 DNA 결합 단백질은 징크 핑커 단백질 (zinc finger protein), TALE 단백질 또는 CRISPR 연관 뉴클레아제이고, 상기 융합단백질의 시토신 탈아미노효소 또는 이의 변이체는 박테리아 유래이고, 이중 가닥 DNA 특이적인 것을 특징으로 하는 방법을 제공한다.
- [44] 본 발명은 또한, 상기 융합단백질 또는 이를 코딩하는 핵산을 원핵세포 또는 진핵세포에 처리하는 단계를 포함하는 원핵세포 또는 진핵세포에서의 A-to-G 염기 교정 방법으로, 상기 DNA 결합 단백질은 징크 핑커 단백질 (zinc finger protein), TALE 단백질 또는 CRISPR 연관 뉴클레아제이고,
- [45] 상기 융합단백질의 시토신 탈아미노효소 또는 이의 변이체는 박테리아 유래이고, 이중 가닥 DNA 특이적인 것이며,
- [46] 상기 시토신 탈아미노효소 또는 이의 변이체의 N 말단에 DNA 결합 단백질이

결합되어 있고, 상기 융합단백질의 아데닌 탈아미노효소의 C 말단에 DNA 결합 단백질이 결합되어 있는, 방법을 제공한다.

- [47] 본 발명은 또한 상기 융합단백질 또는 이를 코딩하는 핵산 및 UGI (uracil glycosylase inhibitor)를 원핵세포 또는 진핵세포에 처리하는 단계를 포함하는 원핵세포 또는 진핵세포에서의 C-to-T 염기 교정 방법으로, 상기 DNA 결합 단백질은 징크 핑거 단백질 (zinc finger protein), TALE 단백질 또는 CRISPR 연관 뉴클레아제이고, 상기 융합단백질의 시토신 탈아미노효소 또는 이의 변이체는 박테리아 유래이고, 이중 가닥 DNA 특이적인 것을 특징으로 하는 방법을 제공한다.

[48]

도면의 간단한 설명

- [49] 도 1은 pTarget 플라스미드를 이용한 ZFD 최적화 결과를 나타낸 것이다.
- [50] a, ZFD 구조. Split-DddAtox 반쪽은 ZFP(Zinc Finger Protein: C 유형)의 C 말단에 융합된다. b, pTarget 라이브러리를 사용한 ZFD 플랫폼 최적화. pTarget 플라스미드는 크기가 1 ~ 24bp(빨간색으로 표시됨) 범위의 스페이서 및 ZFP DNA 결합 부위(녹색으로 표시됨) 영역을 포함한다. ZFD 구조에는 다양한 길이의 AA 링커(노란색 및 주황색으로 표시)와 다른 DddAtox 분할 사이트 및 방향(파란색으로 표시)이 포함되어 있다. c-d, ZFD 활성은 (b)에 설명된 변수들의 영향을 조사하기 위해 pTarget 라이브러리의 타겟 부위에서 측정되었다. 왼쪽 및 오른쪽 ZFD에서 동일한(c) 또는 다른(d) 길이의 링커가 있는 ZFD 쌍을 테스트했다. 염기 교정 빈도는 pTarget 플라스미드의 해당 영역에 targeted deep 시퀀싱에 의해 측정되었다. 데이터는 n=2 생물학적으로 독립적인 샘플에서 얻은 평균(s.e.m.)의 평균 $\pm$ 표준 오차로 표시된다.
- [51] 도 2는 pTarget 플라스미드에서의 다양한 링커를 가지는 ZFD 효율을 확인한 결과를 나타낸 것이다.
- [52] a, 히트 맵에 묘사된 1-24bps의 길이를 갖는 다양한 pTarget 스페이서에서 C/G에서 비 C/G로 교정되는 빈도. ZFP와 분할된 DddAtox 사이의 있는 여러가지 종류의 링커와 DddAtox가 분할된 부위 등 다양한 ZFD 구성이 테스트되었다. b, 각 ZFD 쌍의 전체적인 활성. x축에 사용된 명명법에서 밑에는 왼쪽 ZFD가 위는 오른쪽 ZFD가 표시되어있다. c, 스페이서 길이에 따른 염기 교정 효율. "AA"는 링커에 있는 아미노산의 수를 나타낸다. 데이터는 n=2 생물학적으로 독립적인 샘플에서 얻은 평균(s.e.m.)의 평균 $\pm$ 표준 오차로 표시.
- [53] 도 3은 24AA 링커와 다른 링커를 가지는 ZFD의 효율을 확인한 결과를 나타낸 것이다.
- [54] 히트 맵에 표시된 C/G에서 비 C/G로의 교정 효율에 대한 ZFD 링커 길이의 영향. ZFD 쌍의 왼쪽 ZFD는 24개의 AA 링커로 고정하고 오른쪽 ZFD는 가변 길이의 링커를 포함, 그 반대의 경우도 마찬가지. 오차 막대는 n = 2개의

생물학적으로 독립적인 샘플에 대한 평균(s.e.m.)의 표준 오차

- [55] 도 4는 생체내 핵을 표적하는 ZFD의 활성을 확인한 결과를 나타낸 것이다.
- [56] a, 핵 DNA 타겟 ZFD의 구조. Split-DddAtox 반쪽은 ZFP의 C 말단(C 유형) 또는 N 말단(N 유형)에 융합된다. ZFD 쌍은 CC 또는 NC 구성으로 설계되었으며, 각각 C형 왼쪽 ZFD와 C형 오른쪽 ZFD 또는 N형 왼쪽 ZFD와 C형 오른쪽 ZFD로 구성된다. b, HEK 293T 세포의 핵 DNA 타겟 부위에서 ZFD에 의해 유도된 염기 교정 빈도. 데이터는 $n=3$ 생물학적으로 독립적인 샘플에서 평균 $\pm$ s.e.m.으로 표시. c-f, HEK 293T 세포의 *NUMBL*(c), *INPP5D-2*(d), *TRAC-CC*(e) 및 *TRAC-NC*(f) 타겟 부위의 스페이서 내의 각 염기 위치에서 ZFD 유도 염기 교정 효율. 데이터는 $n=3$ 생물학적으로 독립적인 샘플에서 평균 $\pm$ s.e.m.으로 표시. g, 전기천공 또는 직접전달 방식으로 ZFD 단백질 또는 ZFD 인코딩 플라스미드를 전달 후 K562 세포에서 ZFD 유도 염기 교정 빈도. 하나 또는 네 개의 NLS가 있는 ZFD 단백질을 테스트했으며 왼쪽 및 오른쪽 ZFD를 등물랄로 사용했다. 전기천공 방식의 경우 Amaxa 4D-Nucleofector를 사용하여 수행했다. 직접적인 전달 방법의 경우, K562 세포를 왼쪽 및 오른쪽 ZFD 단백질을 함유하는 세포 배지와 함께 인큐베이션하였다. 세포는 동일한 방식으로 한 번(1x) 또는 두 번(2x) 처리되었다. 생물학적으로 독립적인 $n=2$ 샘플에서 데이터는 평균 $\pm$ s.e.m.으로 표시
- [57] 도 5는 핵 DNA를 표적하는 ZFD의 구성을 나타낸 모식도이다. a-d, 네 가지 가능한 ZFD 구성. NC 및 CN 구성은 구조적으로 동일하지만 왼쪽 및 오른쪽 ZFD 구성의 유형이 다르다.
- [58] 도 6은 생체내 핵 타겟하는 ZFD의 Indel률을 확인한 결과를 나타낸 것이다. 테스트된 모든 ZFD는 0.4% 미만의 빈도로 indel을 생성했다. 데이터는 $n=3$ 생물학적으로 독립적인 샘플에서 평균 $\pm$ s.e.m.으로 표시.
- [59] 도 7은 In-vitro ZFD 재조합 단백질에 대한 활성 테스트 결과를 나타낸 것이다.
- [60] a, TRAC 사이트를 대상으로 하는 ZFD 쌍의 정제 단계. GST 태그가 붙은 단백질은 글루타티온 세파로스 비드를 사용하여 E. coli 세포 용해물에서 정제되었다. 폴리아크릴아미드 겔 전기영동을 사용하여 정제 단계를 모니터링했다. 겔은 쿠마시 블루로 염색되었다. 레인 1, 분자량 마커. 레인 2, 단백질 발현이 IPTG로 유도되지 않은 세포의 샘플. 레인 3, 단백질 발현이 IPTG로 유도된 세포의 샘플. 레인 4, 초음파 처리 후 가용성 분획. 레인 5, 초음파 처리 후 불용성 분획. 레인 6, 컬럼의 통과 분획. 레인 6, 세척 분획. 레인 7, 용출 분획. 대표 마커의 크기는 왼쪽에 표시된다. 빨간색 상자는 ZFD 단백질을 나타낸다. b, 왼쪽 및 오른쪽 ZFD 결합 사이트. 빨간색 화살표는 ZFD 유도 탈아미노화에 대한 가능한 사이트를 나타냄. c, TRAC 부위를 포함하는 PCR 앰플리콘에 대한 ZFD 활성 개요. TRAC-NC ZFD 쌍은 먼저 시토신을 탈아미노화하여 우라실(빨간색으로 표시)을 생성한다. 그런 다음 USER 효소는 우라실을 절단하여 틈을 생성한다(빨간색 삼각형으로 표시). d, 아가로스 겔

전기영동으로 분석한 미처리 PCR 앰플리콘(왼쪽)과 ZFD 쌍으로 처리된 PCR 앰플리콘(오른쪽).

- [61] 도 8은 미토콘드리아 DNA 타겟 ZFD의 여러가지 배열 (configuration)에 대한 모식도이다. a-d, 네 가지 가능한 mitoZFD 구성. 기존 ZFD의 NLS는 MTS 및 NES로 대체되었다. NC 및 CN 구성은 구조적으로 동일하지만 왼쪽 및 오른쪽 ZFD 구성의 유형이 다르다.
- [62] 도 9는 mitoZFD의 미토콘드리아 유전자 염기 교정 효율을 확인한 결과를 나타낸 것이다.
- [63] a, HEK 293T 세포의 mitoZFD 및 TALE-DdCBE에 의해 유도된 mtDNA의 염기 교정 빈도. 데이터는 n=2 생물학적으로 독립적인 샘플에서 얻은 평균(s.e.m.)의 평균 $\pm$ 표준 오차로 표시. b-g, HEK 293T 세포에서, ND2(b), ND4L(c), COX2(d), ND6(e) 및 ND1(f) 타겟 부위의 스페이서 내 각 염기 위치에서의 mitoZFD 유도 염기 교정 효율 과 ND1(g) 표적 부위에서 TALE-DdCBE의 유도 염기 교정 효율. 데이터는 n=2 생물학적으로 독립적인 샘플에서 얻은 평균(s.e.m.)의 평균 $\pm$ 표준 오차로 표시. h, mitoZFD와 TALE-DdCBE에 의해 도입된 ND1 유전자의 DNA 변화와 아미노산 변화 비교. 각 돌연변이 대립유전자에 대한 시퀀싱 리드의 빈도(%)는 targeted deep 시퀀싱에 의해 측정되었다. ZFD 쌍 및 TALE-DdCBE 쌍에 대한 스페이서 영역은 파란색 점선으로 표시.
- [64] 도 10은 MT-ZFD가 처리된 HEK293T 세포에서 분리한 단세포 유래 클론 집단의 염기 교정 효율을 확인한 결과를 나타낸 것이다.
- [65] 대립유전자 분석을 위해 단일 세포 유래 클론을 얻었다. 각 단일 세포 유래 클론에서 C/G-to-non-C/G 교정 빈도는 targeted deep 시퀀싱으로 결정되었다. a, ND1 타겟 mitoZFD로 처리된 HEK 293T 세포 집단의 단일 세포 유래 클론. b, ND2 타겟 mitoZFD로 처리된 HEK 293T 세포 집단의 단일 세포 유래 클론. c, 미처리 HEK 293T 세포 집단의 단일 세포 유래 클론. ZFP 결합 사이트는 녹색으로 표시된다. mitoZFD 유도 교정을 거친 클론의 높은 교정 빈도는 빨간색으로 표시된다.
- [66] 도 11은 MT-ZFD가 처리된 HEK293T 세포에서 분리한 단세포 유래 클론 집단의 염기 교정 효율을 확인한 결과를 나타낸 것이다.
- [67] 높은 빈도의 염기 교정을 나타내는 단일 세포 유래 클론의 대립유전자 분석. 표는 ND1에서 염기 교정의 결과로 변경된 아미노산을 보여준다. 상단의 참조 시퀀스에서 빨간색 문자는 스페이서를 나타낸다. 대립 유전자에서 빨간색 문자는 아미노산 서열의 변화를 나타낸다. (*는 정지 코돈을 나타낸다.)
- [68] 도 12a 및 도 12b는 MT-ZFD가 처리된 HEK293T 세포에서 분리한 단세포 유래 클론 집단의 염기 교정 효율을 확인한 결과를 나타낸 것이다. 높은 빈도의 염기 교정을 나타내는 단일 세포 유래 클론의 대립유전자 분석. 표는 ND2에서 염기 교정의 결과로 변경된 아미노산을 보여준다. 상단의 참조 시퀀스에서 빨간색 문자는 스페이서를 나타낸다. 대립 유전자에서 빨간색 글자는 아미노산 서열의

변화를 나타낸다.

- [69] 도 13은 ZFD와 TALE-DdCBE의 조합에 의한 염기 교정 효율을 확인한 결과를 나타낸 것이다.
- [70] a, mitoZFD 및 TALE-DdCBE 쌍의 결합 영역의 DNA 서열. TALE-DdCBE에 의해 인식되는 사이트는 녹색으로 강조 표시되고 mitoZFD에 대해서는 파란색으로 강조 표시된다. 상단 서열은 mtDNA heavy strand를 나타내고 하단 서열은 mtDNA light strand를 나타낸다. b, ZFD, TALE-DdCBE 및 ZFD/DdCBE 혼합 쌍에 의해 교정된 시토신의 빈도. Targeted deep 시퀀싱을 사용하여 데이터를 얻었으며 데이터는 n=2 생물학적으로 독립적인 샘플에서 얻은 평균(s.e.m.)의 평균 $\pm$ 표준 오차로 표시. c, 각 염기 위치에서 염기 교정 활성의 히트 맵. 빨간색 상자는 각 구성에 대한 스페이서 영역을 나타낸다. 파란색 화살표는 mtDNA의 위치를 나타낸다.
- [71] 도 14는 ZFD 농도별, mRNA vs Plasmid에 따른 교정 효율을 확인한 결과를 나타낸 것이다. ND1 타겟 mitoZFD의 미토콘드리아 게놈 전체 타겟 특이성은 ZFD를 인코딩하는 mRNA 또는 플라스미드의 농도에 따라 다른 특이성을 가진다. mtDNA 전체 시퀀싱에 의해 결정된 on- 및 off- 타겟 염기 교정 빈도. 표시된 농도의 ND1-타겟 mitoZFD-인코딩 플라스미드 또는 mRNA로 트랜스펙션한 HEK 293T 세포의 결과가 그래프 점으로 표시됨. 빨간색 화살표는 타겟 사이트를 나타내고 빨간색 점은 타겟 사이트의 염기 교정 빈도를 나타내며 회색 점은 컨트롤에도 존재하는 SNP를 나타냄. 데이터는 n=2 생물학적으로 독립적인 샘플에서 얻은 평균(s.e.m.)의 평균 $\pm$ 표준 오차로 표시.
- [72] 도 15는 ZFD 농도별, mRNA vs Plasmid에 따른 교정 효율을 확인한 결과를 나타낸 것이다. a, 상단의 도면은 ND1 사이트에서 ZFD 결합을 보여준다. ZFD 결합 사이트는 녹색으로 표시된다. 스페이서 사이에 있는 타겟 시토신은 빨간색으로 표시된다. 도 14의 전체 mtDNA 시퀀싱 데이터에서 결정된 on-타겟 활성. 트랜스펙션된 mitoZFD를 인코딩하는 플라스미드 또는 mRNA의 양이 감소함에 따라 활성이 감소한다. b, 각 플라스미드 또는 mRNA 양에 대해 > 1%의 빈도로 교정되는 C/G 사이트의 수. c, 각 플라스미드 또는 mRNA 농도에 따른 미토콘드리아 게놈의 모든 C/G에 대한 평균 C/G-to-T/A 교정 빈도. 데이터는 n=2 생물학적으로 독립적인 샘플에서 얻은 평균(s.e.m.)의 평균 $\pm$ 표준 오차로 표시.
- [73] 도 16은 ZFD 농도별, mRNA vs Plasmid에 따른 교정 효율을 확인한 결과를 나타낸 것이다. mtDNA 전체 시퀀싱에 의해 결정된 on- 및 off- 타겟 염기 교정 빈도. 표시된 농도의 ND2-타겟 mitoZFD-인코딩 플라스미드 또는 mRNA로 트랜스펙션한 HEK 293T 세포의 결과가 그래프 점으로 표시됨. 빨간색 화살표는 타겟 사이트를 나타내고 빨간색 점은 타겟 사이트의 염기 교정 빈도를 나타내며 회색 점은 컨트롤에도 존재하는 SNP를 나타냄. 데이터는 n=2 생물학적으로 독립적인 샘플에서 얻은 평균(s.e.m.)의 평균 $\pm$ 표준 오차로 표시.
- [74] 도 17은 ZFD 농도별, mRNA vs Plasmid에 따른 교정 효율을 확인한 결과를

나타낸 것이다.

- [75] a, 상단의 도면은 ND2 사이트에서 ZFD 결합을 보여준다. ZFD 결합 사이트는 녹색으로 표시된다. 스페이서 사이에 있는 타겟 시토신은 빨간색으로 표시된다. 도 16의 전체 mtDNA 시퀀싱 데이터에서 결정된 on-타겟 활성. 트랜스펙션된 mitoZFD를 인코딩하는 플라스미드 또는 mRNA의 양이 감소함에 따라 활성이 감소한다. b, 각 플라스미드 또는 mRNA 양에 대해 > 1%의 빈도로 염기 교정되는 C/G 사이트의 수. c, 각 플라스미드 또는 mRNA 농도에 따른 미토콘드리아 게놈의 모든 C/G에 대한 평균 C/G-to-T/A 교정 빈도. 데이터는 n=2 생물학적으로 독립적인 샘플에서 얻은 평균(s.e.m.)의 평균 $\pm$ 표준 오차로 표시.
- [76] 도 18은 미토콘드리아 전체 시퀀싱/QQ 변이체를 제작하고 교정 효율을 확인한 결과를 나타낸 것이다.
- [77] a, QQ mitoZFD 변이체는 ZFD의 각 징크 핑거에 R(-5)Q 돌연변이를 포함하여 비특이적 DNA 접착을 제거한다. (징크 핑거 프레임워크의 위치 -5에 R이 없으면 근처에 있는 K 또는 R이 Q로 변환된다.) b, mitoZFD 처리 세포의 mtDNA 전체 시퀀싱. On- 및 off- 사이트의 교정 빈도는 각각 빨간색과 검은색 점으로 표시된다. 데이터는 n=2 생물학적으로 독립적인 샘플에서 얻은 평균(s.e.m.)의 평균 $\pm$ 표준 오차로 표시. >1%의 효율로 C/G에서 T/A로 염기 교정된 것은 모두 표시. c-d, 교정 효율 및 특이성은 전달된 ZFD 인코딩 mRNA의 용량에 따라 다르다. c, 미토콘드리아 게놈의 모든 C/G에 대한 평균 C/G-to-T/A 교정 빈도. d, 염기 교정 빈도가 1%를 초과하는 교정된 C/G 수.
- [78] 도 19. 식물 기반 염기 교정기의 골든 게이트 조립 시스템. cp-DdCBE 및 mt-DdCBE 구조에 대한 골든 게이트 어셈블리의 개략도. 대상 서열의 각 위치에 대해 총 424개의 세트 (= 6 x 64 3부 + 2 x 16 2부 + 2 x 4 1부)에서 TALE 서브 어레이 플라스미드를 선택하고 원하는 벡터와 혼합하여 특정 서열을 대상으로 하는 DdCBE를 코딩하는 플라스미드를 생성했다.
- [79] 도 20. 식물의 엽록체와 미토콘드리아 염기 교정. a, b, c, d, 16s rDNA (a, b) 및 psbA (c, d)에서 cp-DdCBE에 의해 유도된 엽록체 염기 교정 빈도와 패턴. 분할 DdCBE G1333과 G1397의 쌍을 양성추와 유채 원형질체에 형질감염 시켰다. e, fATP6 유전자에서 mt-DdCBE에 의해 유도된 미토콘드리아 염기 교정 효율과 패턴. 분할 DdCBE G1333과 G1397의 쌍을 양성추와 유채 원형질체에 형질감염 시켰다. a, c, e, TALE 결합 영역은 파란색으로 표시, 스페이서의 시토신은 주황색으로 표시. 모든 그래프에서 오차 막대는 평균 $\pm$ 표준 편차를 보여준다. 3 개의 독립적인 생물학적 복제물. b, d, f, 변환된 뉴클레오티드는 빨간색으로 표시. 교정된 대립 유전자 % (평균 $\pm$ 표준 편차)은 3 개의 독립적인 실험에서 얻었다.
- [80] 도 21. DdCBE를 통해 식물 기관 DNA 교정. a, 식물 기관 변이 유발 개략도. b, 대표적인 생어 시퀀싱 크로마토그램을 포함하여, 스펙티노마이신의 부재하에서 배양된 cp-DdCBE 형질감염 캘러스 중 C-G에서 T-A로 변환되는 것에 대한 효율.

변환된 뉴클레오티드 왼쪽에 빨간색으로 표시되어 있다. 화살표는 크로마토그램의 치환 뉴클레오티드를 보여준다. c, DdCBE 구동 식물 기관 변이 유발 요약. 변이 캘러스는 모의 처리된 캘러스에서의 빈도보다 훨씬 높은 교정 빈도를 가지는 것으로 나타난다. d, 16srDNA을 대상으로 cp-DdCBE를 코딩하는 mRNA를 상추 원형질체에 형질감염 후 유도되는 C에서 T로 변환 빈도. 오차 막대는 평균 $\pm$ s.d이다. n = 3의 독립적인 생물학적 복제물. e 2.5 개월차 스펙토마이신 내성 캘러스의 교정 빈도 및 패턴. f 대표적인 생어 시퀀싱 크로마토그램을 사용하여 DdCBE mRNA를 형질감염한 스트렙토 마이신 내성 식물 중 C·G에서 T·A로의 변환 효율. 화살표는 크로마토그램의 치환 뉴클레오티드를 보여준다. 스케일 바 : 1mm.

- [81] 도 22. 엽록체와 미토콘드리아 염기 교정 전략. cp-DdCBE 및 mt-DdCBE 전단백질 (preprotein, 전구체 단백질) 각각에 CTP (chloroplast transit peptide) 또는 MTS (mitochondrial targeting signal)가 포함되어 있기 때문에, 식물세포에서 번역된 후 엽록체 및 미토콘드리아로 운송된다. 전단백질은 세포 소기관의 외막과 내막을 통과하고, CTP 및 MTS는 각각 간질 프로세싱 펩티다제 (peptidase)와 미토콘드리아 프로세싱 펩티다제에 의해 절단된 다음, cp-DdCBE와 mt-DdCBE (성숙 단백질)이 최종 형태 (conformation)을 형성한다.
- [82] 도 23. 양상추 원형질체에서 DdCBE 플라스미드를 통해 교정하는 것에 대한 시간 경과. 형질감염된 원형질체를 각 시점에서 수집하고 타겟 딥 시퀀싱 (targeted deep sequencing)에 의해 교정 효율성을 분석했다. 빈도 (평균 $\pm$ 표준 편차)은 3 번의 독립적인 실험에서 얻었다.
- [83] 도 24. psbB 유전자의 염기 교정 빈도. 엽록체 psbB 유전자를 표적으로 하는 cp-DdCBE 쌍 Left-G1333-N + Right-G1333-C를 코딩하는 플라스미드를 유체 원형질체에 형질감염한 후 스페이서 영역의 염기 교정 효율을 타겟 딥 시퀀싱에 의해 분석 했다. TALE 결합 영역, 대상 시토신 및 변환된 뉴클레오티드는 각각 파랑, 오렌지 및 빨간색으로 표시되어 있다. 빈도 (평균 $\pm$ 표준 편차)은 n = 3 번의 독립적인 실험에서 계산되었다.
- [84] 도 25. 미토콘드리아 RPS14 유전자의 염기 교정 효율. RPS14 유전자를 대상으로 mt-DdCBE 쌍 Left-G1333-N + Right-G1333-C를 코딩하는 플라스미드를 유체 원형질체에 형질감염한 후 C에서 T로 변환 효율을 타겟 딥 시퀀싱에 의해 분석했다. TALE 결합 영역, 대상 시토신 및 변환된 뉴클레오티드는 각각 파랑, 오렌지 및 빨간색으로 표시되어 있다. 빈도 (평균 $\pm$ 표준 편차)은 n = 3 번의 독립적인 실험에서 계산되었다.
- [85] 도 26. 캘러스에서 엽록체 게놈을 대상으로 하는 염기 교정 효율. 4 주간 배양 후 양상추와 유체 캘러스에서 16srDNA과 psbA 중 표적 부위에 대하여 DdCBE을 통해 염기 교정한 빈도와 패턴. 스페이서 영역의 변환된 뉴클레오티드는 빨간색으로 표시되어 있다.
- [86] 도 27. 캘러스에서 미토콘드리아 게놈을 대상으로 하는 염기 교정 효율성. 유체

캘러스 중 ATP6 및 RPS14 유전자의 표적 부위에서 DdCBE을 통해 염기 교정의 빈도와 패턴이 타겟 딥 시퀀싱에 의해 확인되었다. 대상 스페이서 영역의 변환된 뉴클레오티드는 빨간색으로 표시되어 있다.

- [87] 도 28. DNA 미포함 염기 교정. 상추 원형질체에 DdCBE mRNA를 형질감염시킨 후 16srDNA 중 타겟 사이트에서의 엽록체 염기 교정의 빈도와 패턴. 원형질체를 7일간 배양한 후 타겟 딥 시퀀싱을 실행했다. 대상 스페이서 영역의 변환된 뉴클레오티드는 빨간색으로 표시되어 있다.
- [88] 도 29는 원형질체 및 캘러스 중 DdCBE mRNA 또는 DNA 서열이 존재하지 않음을 보여주는 겔 전기영동 결과를 나타낸 것이다 (M은 마커).
- [89] 도 30. 16srDNA 변이 선별. 빨간색 화살표는 스트렙토마이신 내성 녹색 캘러스를 보여준다.
- [90] 도 31. 항생제 내성 캘러스나 묘목에서 DdCBE 표적 부위 근처에 오프 타겟 변이는 발견되지 않았다. (a), (b) 오프 타겟 활성은 타겟 딥 시퀀싱에 의해 분석되었다. TALE 결합 부위와 스페이서 영역은 각각 녹색과 빨간색 밑줄로 표시되어 있다. (a) DdCBE 플라스미드를 형질감염한 상추 원형질체에서 배양한 스펙티노마이신 내성 캘러스. (b) 스트렙토마이신 내성의 묘목에서 얻은 슈트 (shoot).
- [91] 도 32. 온-타겟 사이트와 가장 상동성이 높은 5 개의 사이트에서의 오프 타겟 활성 분석. TALE 결합 부위에 최대 9 개의 불일치를 포함하여, 양상추 엽록체 게놈의 16s rRNA 유전자 특이적 DdCBE 5 개의 잠재적인 오프 타겟 부위와 선택되었다. TALE 결합 서열 및 불일치 뉴클레오티드는 각각 파란색과 빨간색으로 표시되어 있다. 오프 타겟 변이 빈도는 표적 딥 시퀀스를 사용하여 DdCBE 플라스미드 또는 DdCBE mRNA로 형질감염된 원형질체 및 약제 내성 캘러스 또는 슈트에서 측정되었다. 빈도 (평균 $\pm$ 표준 편차)은 3 번의 독립적인 실험에서 얻었다.
- [92] 도 33. DdCBE 조립 및 미토콘드리아 DNA 교정의 모식도. a 효율적인 DdCBE 구축을 위한 one-pot Golden-gate 조립의 설명. 최종 플라스미드 제작을 위한 왼쪽 및 오른쪽 모듈을 제작하기 위해 총 424개의 서열 (64개의 3개 서열 인식 서열 $\times$ 6 + 16개의 2개 서열 인식 서열 $\times$ 2 + 4개의 1개 서열 인식 서열 $\times$ 2) 및 발현 벡터를 섞었다. b 생쥐 미토콘드리아 DNA의 목표 유전자 ND5와 DdCBE가 상호작용하는 모식도. TALE 결합 부위는 회색으로, 염기 교정 범위는 검은색으로 표시되었다. 각각의 repeat variable diresidue module은 주황색, 파란색, 녹색 및 노란색으로 표시되었는데, 순서대로 아데닌 “NI”, 티민 “NG”, 구아닌 “NN”, 및 시토신 인식을 위한 “HD”이다.
- [93] 도 34. DdCBE로 인한 염기 교정으로 생긴 생쥐 미토콘드리아 ND5 점 돌연변이. a DdCBE 탈아미노화 효소로 매개된 시토신에서 티민으로의 염기 교정 목표 서열 및 NIH3T3 세포에서의 효율. 목표 서열에서 번역 코돈은 밑줄 표시되어 있고 교정이 가능한 부위는 빨간색으로 표기되어 있다. DdCBE 형질

감염을 위한 조합은 left 또는 right, -G1333 또는 -G1397, 및 -N 또는 -C로 표기되어 있다. left-G1333-N + right-G1333-C, left-G1333-C + right-G1333-N, left-G1397-N + right-G1397-C, 및 left-G1397-C + right-G1397-N의 C10 돌연변이에 대한 P 값은 각각 순서대로 0.0012, 0.0003, 0.0014, 및 0.0009 이고, C13 돌연변이는 순서대로 0.0116, 0.0076, 0.0030, 및 0.0003이다(* $p < 0.05$ 및 ** $p < 0.01$, Student's two-tailed t test 사용). b 이에 따른 생쥐 배반포에서의 염기 교정 효율. 서열 분석 데이터는 left-G1397-N과 right-G1397-C DdCBE mRNA를 미세주입한 접합체로부터 발달된 배반포로부터 얻었다. c 새로 태어난 새끼의 돌연변이 서열 정렬 도면. 타겟화 딥 시퀀싱 (Targeted deep sequencing)은 태어나자마자 얻은 꼬리 부위 및 생후 7일, 14일 후 발가락에서 얻은 조직으로부터 유전체 DNA를 추출하여 수행하였다. 교정된 염기는 빨간색으로 표시되었다. 돌연변이 미토콘드리아 유전체의 교정 빈도가 표시되었다. d 성체 F0 생쥐(sipup-1)의 다양한 조직에서의 교정 효율. 시퀀싱 데이터는 생후 50일에 각 조직에서 얻었다. 모든 그래프에서 진한 회색 막대와 밝은 회색 막대는 각각 m.C12539T(C10) 및 m.G12542A(C13) 돌연변이의 교정 빈도를 나타낸다. 오차 막대는 $n = 3$ 개의 생물학적으로 독립적인 샘플에 대한 평균(s.e.m.)의 표준 오차이다.

- [94] 도 35. 돌연변이 미토콘드리아 DNA의 생식 세포로의 전달. a mtDNA 돌연변이의 생식선 전달을 관찰하기 위해 암컷 F0(sipup-3) 생쥐와 야생형 C57BL/6J 수컷을 교배하여 F1 자손(101, 102)을 얻은 후, 타겟화 딥 시퀀싱 (Targeted deep sequencing)을 수행하였다. 교정된 염기는 빨간색으로 표시하였다. 돌연변이 미토콘드리아 유전체의 교정 빈도가 표시되어 있다. b 게놈 DNA의 targeted deep sequencing을 사용하여 얻은 F1 새끼(101)의 다양한 조직에서의 염기 교정 효율. 진하고 밝은 회색 막대는 각각 m.C12539T(C10) 및 m.G12542A(C13) 돌연변이의 빈도를 나타낸다. 오차 막대는 $n = 3$ 개의 생물학적으로 독립적인 샘플에 대한 평균(s.e.m.)의 표준 오차이다.
- [95] 도 36. DdCBE에 의해 생성된 생쥐 미토콘드리아 ND5 G12918A 돌연변이. a ND5 단백질에서 D393N 변화를 야기하는 m.G12918A 점 돌연변이를 만들기 위한 DdCBE 표적. 목표 코돈에 밑줄이 그어져 있고, 가능한 교정 부위가 빨간색으로 표시된다. b NIH3T3 세포에서 DdCBE를 사용한 시토신-티민 염기 교정의 효율. 형질 감염된 DdCBE 쌍의 조합이 표기되어 있다. 오차 막대는 s.e.m.이다. $n = 3$ 개의 생물학적으로 독립적인 샘플의 경우(n.s. 유의미하지 않음, * $p < 0.05$, ** $p < 0.01$, Student's two-tailed t test 사용). left-G1333-N + right-G1333-C, left-G1333-C + right-G1333-N, left-G1397-N + right-G1397-C, left-G1397-C + right-G1397-N의 C6 돌연변이에 대한 P 값은 각각 0.0052, 0.0099, 0.0027 및 0.0040이다. n.s에 대한 P 값은 0.4971이다. c m.G12918A 생쥐 배반포에서 점 돌연변이 염기 교정 효율. 시퀀싱 데이터는 1 세포기 배아에 left-G1397-C 및 right-G1397-N DdCBE를 코딩하는 mRNA를 미세주입한 후,

배양하여 발달 된 배반포에서 얻었다. d ND5 점 돌연변이를 지닌 생쥐(F0). DdCBE mRNA의 미세주입 후에 발달한 ND5 점 돌연변이를 보유하고 있는 F0 새끼. 갓 태어난 새끼에서 확인된 돌연변이 서열 정렬. 교정된 염기는 빨간색으로 표시되어 있고, 오른쪽은 돌연변이 미토콘드리아 유전자의 교정 빈도를 나타낸다.

- [96] 도 37. 시토신 탈아미노화 효소로 매개된 염기 교정을 통해 생성된 생쥐 미토콘드리아 ND5 년센스 돌연변이. a m.C12336T 년센스 돌연변이 및 m.G12341A 침묵 돌연변이를 생성하기 위한 DdCBE 목표 서열. m.C12336T(C9) 돌연변이는 ND5 단백질에서 Q199stop 돌연변이를 생성하는 반면, m.G12341A(C14)는 침묵하는 Q200Q 돌연변이를 일으킨다. 전사 트리플렛은 밑줄이 그어져 있고, 가능한 교정 위치는 빨간색으로 표시되어 있다. b NIH3T3 세포에서 년센스 돌연변이를 생성하는 시토신-티민 염기 교정의 효율성. 형질 감염된 DdCBE 쌍의 조합이 표기되어 있다. 진하고 밝은 회색 막대는 각각 m.C12336T(C9) 및 m.G12341A(C14) 돌연변이의 빈도를 나타낸다. 오차 막대는 s.e.m.을 나타낸다. n = 3개의 생물학적으로 독립적인 샘플의 경우(n.s. 유의미하지 않음, *p < 0.05, **p < 0.01, Student's two-tailed t test 사용). Left-G1333-N + right-G1333-C, left-G1333-C + right-G1333-N, left-G1397-N + right-G1397-C, left-G1397-C + right-G1397-N의 C9 돌연변이에 대한 P 값은 0.0065, 0.1143, 0.0266 및 0.0037이고, C14 돌연변이에 대한 P 값은 각각 0.0077, 0.0144, 0.0406 및 0.0214이다. c 생쥐 배반포에서의 교정 효율. 서열 분석 데이터는 접합체에 left-G1333-N 및 right-G1333-C DdCBE를 인코딩하는 mRNA를 미세주입한 후 발달한 배반포에서 얻었다. 진하고 밝은 회색 막대는 각각 C9 및 C14 돌연변이의 빈도를 나타낸다. d 갓 태어난 새끼의 돌연변이 서열 정렬. 교정된 염기는 빨간색으로 표시되었고, 오른쪽은 돌연변이 미토콘드리아 게놈의 교정 빈도를 나타낸다. e 야생형 및 교정된 생쥐의 생거 서열 분석 크로마토그램. 빨간색 화살표는 치환된 뉴클레오티드를 나타낸다.
- [97] 도 38. DdCBE 구조를 생성하는 골든 게이트 클로닝의 개략도. 모든 반응은 1 개의 튜브에서 동시에 발생한다. 화살표는 연속적인 반응 과정을 의미하지는 않는다. BsaI 효소를 사용하여 빈 발현 벡터 및 모듈 벡터를 절단하여 호환 응집 말단을 포함하여 선형화된 백본과 TALE 모듈 삽입물이 빠지도록 하였다. 다음으로, T4 DNA 리가아제가 백본과 6 개의 모듈 인서트를 결합하여 최종 DdCBE 구조를 만들었다. 8 개의 DdCBE 복제 백본 플라스미드를 사용했다. SOD2MTS의 경우 Left-G1333-N, Left-G1333-C, Left-G1397-N, Left-G1397-C이다. COX8A MTS의 경우 Right-G1333-N, Right-G1333-C, Right-G1397-N, Right-G1397-C이다.
- [98] 도 39. ND5 변이 마우스 (F0). (a) ND5 침묵 돌연변이 마우스, (b) ND5 G12918A 돌연변이 마우스, (c) DdCBE mRNA 마이크로 주입 후 발생한 ND5 년센스 돌연변이 마우스.

- [99] 도 40. (a) DdCBE-NES를 포함하는 벡터 모식도 및 NES 서열. (b) Mouse m.G12918 ND5 유전자 및 핵 내 4번 염색체의 ND5-like 서열. 미토콘드리아 TrnA 및 핵 내 5번 염색체 서열. 미토콘드리아 Rnr2 및 핵 내 6번 염색체 서열. (c) NIH3T3 세포주를 이용한 DdCBE 및 DdCBE-NES의 ND5 유전자에서의 교정 효율. (d) NIH3T3 세포주를 이용한 DdCBE 및 DdCBE-NES의 TrnA 유전자에서의 교정 효율. (e) NIH3T3 세포주를 이용한 DdCBE 및 DdCBE-NES의 Rnr2 유전자에서의 교정 효율. 주황색 그래프는 DdCBE, 회색 그래프는 DdCBE-NES의 교정 효율이다. (f) mitoTALEN TALE 배열의 DNA 인식 서열. (g) mitoTALEN을 처리하거나 처리하지 않은 실험군에서의 DdCBE 염기 교정 효율. 모든 그래프는 n=2이며 오차 막대는 평균 표준 오차이다.
- [100] 도 41. DdCBE-NES 및 mitoTALEN을 이용한 생쥐 배아 및 개체에서의 향상된 교정 효율. (a) DdCBE-NES를 이용한 배반포에서의 여러 미토콘드리아 DNA 표적(mtND5, mtTrnA, mtRNR2) 염기 교정 효율. (b) DdCBE, DdCBE-NES 및 mitoTALEN을 이용한 m.G12918A 염기 교정 효율 비교. (c) 생쥐 개체에서의 m.G12918A 염기 교정 효율 비교. 모든 그래프는 n>=3이며 오차 막대는 평균 표준 오차이다. (n.s.=통계적으로 무의미함, *p<0.05, **p<0.01, ***p<0.001을 Student's two-tailed t-test를 이용하여 얻었다.)
- [101] 도 42. (a) DdCBE 단백질 개량에 대한 모식도. (b,c) DddAtox 탈아미노화 효소의 크리스탈 구조. 결합 표면 자리의 잔기는 막대로 표현되었다. (b) G1397-N 분할 및 G1397-C 분할은 각각 순서대로 자주색과 하늘색으로 표기되었다. (b) G1333-N 분할 및 G1333-C 분할은 각각 순서대로 주황색과 녹색으로 표기되었다. (d, e) 결합 표면 돌연변이 중 G1397-N 및 G1397-C (d)와 G1333-N 및 G1333-C (e)는 빨간색으로 표시되었다.
- [102] 도 43. (a) G1397 결합 표면 돌연변이의 염기 교정 효율 그래프. 교정 범위 및 목표 시토신은 위쪽에 표시되었다. 돌연변이 및 야생형/TALE-free DddAtox 단백질은 옆에 표기한 바와 같이 co-transfection되었다. Left-DdCBE의 경우 TALE-free DddAtox 단백질은 G1397-N이며, Right-DdCBE의 경우 TALE-free DddAtox 단백질은 G1397-C이다. (b) DdCBE와 돌연변이의 목표 시토신-티민(구아닌-아데닌) 염기 교정 효율을 표기한 히트맵.
- [103] 도 44. (a) G1333 결합 표면 돌연변이의 염기 교정 효율 그래프. 교정 범위 및 목표 시토신은 위쪽에 표시되었다. 돌연변이 및 야생형/TALE-free DddAtox 단백질은 옆에 표기한 바와 같이 co-transfection 되었다. Left-DdCBE의 경우 TALE-free DddAtox 단백질은 G1333-N이며, Right-DdCBE의 경우 TALE-free DddAtox 단백질은 G1333-C이다. (b) DdCBE와 돌연변이의 목표 시토신-티민(구아닌-아데닌) 염기 교정 효율을 표기한 히트맵.
- [104] 도 45는 Wild type과 신규 full length DddA의 아미노산 염기서열 비교 결과를 나타낸 것이다.
- [105] 도 46은 전장 DddA를 동물 또는 식물세포에 전달한 형태를 나타낸 것이다.

- [106] 도 47은 human cell genomic context ROR1 site (a), HEK3 site (b), TYRO3 site (c) 에서 TC motif의 cytosine이 thymine으로 치환되는 활성을 확인한 결과를 나타낸 것이다.
- [107] 도 48은 전장 DddA의 장점을 나타낸 것이다.
- [108] 도 49는 Human cell genomic context TRAC site1 (a), TRAC site 2 (b), FANCF (c), HBB (d)에서 전장 DddA의 활성을 측정한 결과를 나타낸 것이다.
- [109] 도 50은 DddA-dCas9(D10A, H840A)-UGI를 이용한 Human cell genomic context TYRO3 (a), ROR1 (b), HEK3 (c), EMX1 site 2 (d), TRAC site 1 (e), HBB (f)에서 DddA의 활성을 측정한 결과를 나타낸 것이다.
- [110] 도 51. HEK293T 세포에서 전장 DddAtox의 염기교정 효율. (a) 구조 기반 방식으로 전장 DddAtox를 스크리닝하는 개략도. 적색 알라닌은 양전하를 띠는 아미노산 잔기가 알라닌으로 대체된 것으로 나타났다. (b) 알라닌으로 치환된 DddA 변이체의 E.coli 형질전환체를 나타내었고, E1347A는 대조군으로 활성부위 돌연변이체로 사용하였다. (c) TYRO3 사이트에서 DddA AAAAA 및 CBE의 교정 및 삽입결실 빈도. (d) TYRO3 부위의 대립 유전자 빈도가 C에서 T로의 치환을 빨간색으로 표시함. 프로토 스페이서는 파란색으로 표시되고 프로토 스페이서 인접 모티브(PAM)는 주황색으로 표시함
- [111] 도 52. 무독성 DddA GSVG. (a) 오류가 발생하기 쉬운 PCR을 기반으로 한 무독성 전장 DddAtox 변이체 스크리닝의 개략도. Cas9, nCas9(D10A), nCas9(H840A) 및 dCas9(D10A, H840A)의 N-말단 및 C-말단에 융합된 (b) 및 대립유전자 (c) 교정 빈도. 프로토 스페이서는 파란색으로 표시되고 프로토 스페이서 인접 모티브(PAM)는 주황색으로 표시함.
- [112] 도 53. TYRO3 부위(a), ROR1 부위 1(b), HEK3 부위(c)에서 양전하를 띠는 아미노산 잔기가 알라닌으로 치환된 DddAtox 변이체들이 nCas9(D10A)의 N-말단에 교정 빈도. 프로토 스페이서는 파란색으로 표시되고 프로토 스페이서 인접 모티브(PAM)는 주황색으로 표시함.
- [113] 도 54. ROR1 사이트 1(a), ROR1 사이트 2(b), ROR1 사이트 3(c), FANCF 사이트(d), HBB 사이트(e), HEK3 사이트(f), TRAC5 사이트 1(g) 및 EMX1 사이트(h, i, j)는 각각. 프로토 스페이서는 파란색으로 표시되고 프로토 스페이서 인접 모티브(PAM)는 주황색으로 표시함. C에서 T로의 치환은 빨간색으로 표시함. DddA의 타겟 윈도우는 프로토 스페이서 5' 업스트림으로 세어 음수로 표시함.
- [114] 도 55. HeLa 세포에서 AAAAA 및 E1347A의 TYRO3 사이트(a), ROR1 사이트 1(b), ROR1 사이트 2(c), ROR1 사이트 3(d), FANCF 사이트(e), HBB 사이트(f), HEK3 사이트(g), TRAC5 사이트 1(h), TRAC5 사이트 2(i) 및 EMX1 사이트 2(j)) 에서 교정 빈도. 프로토 스페이서는 파란색으로 표시되고 프로토 스페이서 인접 모티브(PAM)는 주황색으로 표시함. DddA의 타겟 윈도우는 proto-spacer 5' 업스트림으로 세어 음수로 표시함. 표적 시토신은 빨간색으로 표시함.

- [115] 도 56. TYRO3 사이트(a) 및 ROR1 사이트 1(b)에서 AAAAA 및 E1347A의 시간 종속 염기 교정 및 삽입결실 비율을 나타낸 것이다.
- [116] 도 57. EMX1 사이트 2(a), FANCF 사이트(b), TRAC5 사이트 1(c), TRAC5 사이트 2(d), ROR1 사이트 1(e), ROR1 사이트 2(f), ROR1 사이트 3(g) 및 HBB 사이트(h)에서 nCas9(D10A), nCas9(H840A) 및 dCas9의 N-말단에 융합된 GSVG의 교정, 삽입결실 및 대립유전자 빈도. 프로토 스페이서는 파란색으로 표시되고 프로토 스페이서 인접 모티브(PAM)는 주황색으로 표시함. C에서 T로의 치환은 빨간색으로 표시함. GSVG의 타겟 윈도우는 프로토 스페이서 5' 업스트림으로 세어 음수로 표시함.
- [117] 도 58. EMX1 사이트 2(a), EMX1 사이트 4(b), ROR1 사이트 2(c), 및 HBB 사이트(d)에서 nCas9(D10A), nCas9(H840A) 및 dCas9의 C-말단에 융합된 GSVG의 교정, 삽입결실 및 대립유전자 빈도. 프로토 스페이서는 파란색으로 표시되고 프로토 스페이서 인접 모티브(PAM)는 주황색으로 표시함. G에서 A로의 치환은 빨간색으로 표시함. GSVG의 타겟 윈도우는 프로토 스페이서의 첫 번째 위치에서 3' 다운스트림으로 세어 표시된다.
- [118] 도 59. TYRO3(a) 및 EMX1 사이트 2(b)에서 nCas9(H840A)의 C-말단에 융합된 E1347A, GSVG, SSVG, GSAG, GSVS의 시간 종속 교정 및 삽입결실 빈도.
- [119] 도 60. HEK293T 세포에서 mDdCBE의 미토콘드리아 염기 교정. ND4(a) 및 ND6(b)의 교정 효율성. 표적 시토신과 TALE 결합 부위는 각각 빨간색과 회색으로 표시함. DddAtox의 절반만 TALE 어레이에 융합되고 나머지 절반은 TALE이 없는 경우 ND4(c, d) 및 ND6(e, f)에서 교정 효율성. 왼쪽 TALE 배열과 오른쪽 TALE 배열은 각각 L과 R로 표시된다. 참조 계능과의 ND6 TALE 어레이 미스매치는 보라색으로 밑줄이 그어져 있음.
- [120] 도 61:
- [121] (a) 기존의 ZFP DNA 결합 단백질을 이용한 zinc finger cytosine deaminase(ZFD)의 모식도
- [122] (b) adenine deaminase를 ZFD에 삽입하는 곳 (빨간 화살표가 삽입위치이다)
- [123] (c) 만들어진 ZF-DdABE의 핵 DNA Trac site에서 염기 교정 효율. (C to T)
- [124] (d) 만들어진 ZF-DdABE의 핵 DNA Trac site에서 염기 교정 효율. (A to G)
- [125] (WT-ZFD는 adenine deaminase가 있지 않은 split DddAtox만 있는 C to T deaminase이다.)
- [126] (e) Mitochondria DNA를 target하는 ZF-DdABE의 ND1 site에서 효율. (C to T)
- [127] (f) Mitochondria DNA를 target하는 ZF-DdABE의 ND1 site에서 효율. (A to G)
- [128] 도 62:
- [129] (a) TALE과 split DddAtox를 이용한 DdABE의 모식도 (구성요소로는 split DddAtox, adenine deaminase 및 TALE array가 있다.)
- [130] (b) 미토콘드리아 ND4 site를 target하는 TALE에 Adenine deaminase만 붙였을 때 염기 교정 효율

- [131] (c) 미토콘드리아 ND1 site를 target하는 TALE-split DddAtox에 adenine deaminase를 붙였을 때 염기 교정 효율
- [132] (d) 왼쪽은 DdCBE의 pair를 오른쪽에는 TALE-split DddAtox에 adenine deaminase를 붙였을 때 염기 교정 효율을 한 개 뉴클레오타이드 단위로 본 것. (초록색 박스는 TALE이 붙는 부위이다.)
- [133] (e) 왼쪽 TALE-split DddAtox에 adenine deaminase를 붙이고, 오른쪽은 DdCBE의 pair를 이용 했을 때 염기 교정 효율을 한 개 뉴클레오타이드 단위로 본 것. (초록색 박스는 TALE이 붙는 부위이다.)
- [134] 도 63:
- [135] (a) UGI가 없을 때와 UGI가 있을 때의 미토콘드리아 ND1 site target DdABE의 C to T와 A to G의 염기 교정 효율 (빨간 박스는 adenine deaminase를 표시)
- [136] (b) UGI가 없을 때와 UGI가 있을 때의 미토콘드리아 ND4 site target DdABE의 C to T와 A to G의 염기 교정 효율 (빨간 박스는 adenine deaminase를 표시)
- [137] (c) 미토콘드리아 ND1 site target하는 DdABE의 구성중 가장 높은 효율을 가지는 것의 한 개 뉴클레오타이드 단위로 본 것.(초록색 박스는 TALE이 붙는 부위이다.)
- [138] (d) 미토콘드리아 ND4 site target하는 DdABE의 구성중 가장 높은 효율을 가지는 것의 한 개 뉴클레오타이드 단위로 본 것(초록색 박스는 TALE이 붙는 부위이다.)
- [139] 도 64:
- [140] (a) 위: 한 개 TALE module에 모든 construct가 있는 single tale module 모식도 (구성 요소로는 Full length DddAtox, adenine deaminase 및 TALE array가 있다.)
- [141] 아래: 두 개 TALE module을 사용하는 dual tale module (구성 요소로는 Full length DddAtox와 TALE array가 한쪽에, 다른 한쪽은 adenine deaminase와 TALE array가 있다.)
- [142] (b) 미토콘드리아 ND1 site를 target하는 single module 및 dual module DdABE의 염기 교정 효율
- [143] (c) 미토콘드리아 ND4 site를 target하는 single module 및 dual module DdABE의 염기 교정 효율
- [144] 도 65는 ND1 site를 타겟하는 single module의 염기 교정 효율 (구성 요소로는 TALE array, adenine deaminase(AD) 및 full length DddAtox (GSVG, AAAA, E1347A는 각각 variant이다.))을 확인한 결과를 나타낸 것이다.
- [145] 도 66은 ND1 site를 타겟하는 dual module의 염기 교정 효율 (구성 요소로는 TALE array, adenine deaminase(AD) 및 full length DddAtox (GSVG, AAAA, E1347A는 각각 variant이다.))을 확인한 결과를 나타낸 것이다.
- [146] 도 67. (a) ND1 site를 타겟하는 TALE 결합단백질에 TadA(AD) 아데닌 탈 아민 효소만 붙였을 때 염기 교정 효율을 나타낸 것이다.
- [147] (b) ND4 site를 타겟하는 TALE 결합단백질에 TadA(AD) 아데닌 탈 아민 효소만

붙었을 때 염기 교정 효율을 나타낸 것이다.

- [148] 도 68. ND1 site를 타겟하는 TALE에 dual module, single module 및 split-DddA-AD에 대한 아데닌과 시토신 염기 교정 효율을 나타낸 것이다. (맨 밑에서부터) 양쪽에 UGI가 있는 경우, 시토신 염기교정만 일어난다. 한쪽에 AD가 붙는 경우에는 시토신과 아데닌 염기교정이 모두 일어나게 되고, UGI가 없는 경우에는 선택적으로 아데닌 염기 교정만 일어나는 것을 볼 수 있다. 마찬가지로 dual module과 single module에서도 선택적으로 아데닌 염기 교정만 일어나는 것을 볼 수 있다.
- [149]
- [150] 발명을 실시하기 위한 구체적인 내용
- [151] 다른 식으로 정의되지 않는 한, 본 명세서에서 사용된 모든 기술적 및 과학적 용어들은 본 발명이 속하는 기술분야에서 숙련된 전문가에 의해서 통상적으로 이해되는 것과 동일한 의미를 갖는다. 일반적으로, 본 명세서에서 사용된 명명법은 본 기술분야에서 잘 알려져 있고 통상적으로 사용되는 것이다.
- [152] 본 명세서에서 사용되는 “교정”은 “편집”과 상호 호환적으로 사용 가능하고, 특이적 게놈 표적의 선택적 결실에 의해서 핵산 서열을 변경시키는 방법을 의미한다. 이러한 특이적 게놈 표적은 염색체 영역, 유전자, 프로모터, 오픈 리딩 프레임 또는 임의의 핵산 서열을 포함하지만, 이에 제한되지 않는다.
- [153] 본 명세서에서 사용되는 “단일 염기”는 핵산 서열 내 하나, 및 오직 하나의 뉴클레오티드를 의미한다. 단일 염기 교정의 맥락에서 사용될 때, 이것은 핵산 서열 내 특이적 위치의 염기가 상이한 염기로 치환되는 것을 의미한다. 이러한 치환은 제한없이 치환 또는 변형을 포함한, 많은 기전에 의해 일어날 수 있다.
- [154] 본 명세서에서 사용되는 “표적” 또는 “표적 부위”는 임의 조성 및/또는 길이의 사전에 확인된 핵산 서열을 의미한다. 이러한 표적 부위는 염색체 영역, 유전자, 프로모터, 오픈 리딩 프레임 또는 임의의 핵산 서열을 포함하지만, 이에 제한되지 않는다.
- [155] 본 명세서에서 사용되는 “온-타겟”은 프로그램 가능한 DNA 결합 영역 및/또는 단일 가이드 RNA 서열과 완전하게 상보성일 수 있는 특이적 게놈 표적의 하위 서열을 의미한다.
- [156] 본 명세서에서 사용되는 “오프-타겟”은 프로그램 가능한 DNA 결합 영역 및/또는 단일 가이드 RNA 서열과 부분적으로 상보성일 수 있는 특이적 게놈 표적의 하위 서열을 의미한다.
- [157]
- [158] 1. 분할 시토신 탈아미노효소 (Split Deaminase)
- [159]
- [160] 본 발명에 따른 융합단백질은 시토신 탈아미노효소 (deaminase) 또는 이의 변이체를 포함하는 융합단백질로, 상기 시토신 탈아미노효소 또는 이의 변이체는 시토신 탈아미노효소 또는 이의 변이체 유래인 제1분할체 및

- 제2분할체를 포함하고, 상기 제1분할체 및 제2분할체는 각각 상기 DNA 결합 단백질에 결합되는 형태를 가진다.
- [161] 상기 시토신 탈아미노효소는 아미노기 이탈 효소로, 시토신(C)을 우리딘(U)으로 전환시킬 수 있다.
- [162] 상기 시토신 탈아미노효소일 수 있다. 상기 시토신 탈아미노효소로 예를 들어, APOBEC1 (apolipoprotein B editing complex 1), AID (activation-induced deaminase)의 형태가 존재하나, 대부분의 DNA 탈아미노효소들은 단일 가닥 DNA에만 작동해서 DNA 결합 단백질에 연결하여 염기 교정하는 데에는 적합하지 않을 수 있다. 구체적으로, 상기 시토신 탈아미노효소는 상기 시토신 탈아미노효소는 이중 가닥 DNA에 작용하는 탈아미노효소 (DddA) 또는 이의 오솔로그 (orthologue)에서 유래한 것일 수 있다. 더욱 구체적으로, 상기 시토신 탈아미노효소는 이중 가닥 DNA 특이적 박테리아 시토신 탈아미노효소일 수 있다.
- [163] 상기 시토신 탈아미노효소는 분할된 형태로, 상기 시토신 탈아미노효소는 분리된 제1분할체 및 제2분할체를 포함하고, 상기 제1분할체와 제2분할체 각각은 탈아미노효소 활성이 없는 것이다.
- [164] 전장 시토신 탈아미노효소 중 tox 분할체에 해당하는 서열번호 1의 서열을 포함할 수 있다. 상기 시토신 탈아미노효소는 제1분할체 및 제2분할체를 포함하고, 상기 제1분할체와 제2분할체 각각은 탈아미노효소 활성이 없다.
- [165] [서열번호 1]
- [166] GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGGPTYPNYA
NAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMETLLPENAKMTVV
PPEGAIPVKRGATGETKVFTGNSNSPKSPTKGGC
- [167] 하나의 실시예에서, 상기 시토신 탈아미노효소의 제1분할체 또는 제2분할체는 서열번호 1의 서열 중 N 말단에서 G33, G44, A54, N68, G82, N98, G108로 구성된 군에서 선택되는 하나 이상까지의 서열을 포함할 수 있다. 상기 시토신 탈아미노효소의 제1분할체 또는 제2분할체는 서열번호 1의 서열 중 G34, P45, G55, N69, T83, A99, A109로 구성된 군에서 선택되는 하나 이상에서 C-말단까지의 서열을 포함할 수 있다.
- [168] 구체적으로, 상기 시토신 탈아미노효소는 서열번호 23의 제1분할체 (G1333-N) 및 서열번호 24의 제2분할체 (G1333-C)을 포함하거나, 서열번호 25의 제1분할체 (G1397-N) 및 서열번호 26의 제2분할체 (G1397-C)을 포함하거나, 서열번호 23의 제1분할체 (G1333-N) 및 서열번호 26의 제2분할체 (G1397-C)을 포함하거나, 서열번호 25의 제1분할체 (G1397-N) 및 서열번호 24의 제2분할체 (G1333-C)을 포함할 수 있다.
- [169] (서열번호 23) Wild-type DddAtox G1333-N
- [170] GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGG
- [171] (서열번호 27)

- [172] GGCTCTGGTTCCTACGCCCTGGGTCCATATCAGATTAGTGCTCCCCAACTC
CCCGCCTACAACGGTCAGACAGTGGGGACCTTTTACTATGTCAACGACGCC
GGGGGATTGGAATCCAAGGTTTTCTCTAGCGGTGGG
- [173] (서열번호 24) Wild-type G1333-C
- [174] PTPYPNYANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMETLL
PENAKMTVVPPEGAIPVKRGATGETKVFTGNSNSPKSPTKGGC
- [175] (서열번호 28)
- [176] CCAACACCTTATCCTAACTACGCTAACGCCGGGCACGTCGAGGGGCAGTC
AGCTCTTTTTATGAGAGATAACGGCATTAGCGAAGGGCTTGTGTTCCATAA
TAATCCTGAGGGCACCTGTGGCTTCTGTGTAAATATGACCGAAACACTTCT
GCCTGAGAACGCTAAAATGACTGTCTGTAACCCGAAGGCGCAATCCAG
TTAAACGGGGCGCAACCGGGCGAAACCAAAGTATTCACCGGAAACAGCAAT
AGTCCAAAGTCCCCACCAAGGGAGGTTGC
- [177] (서열번호 25) Wild-type DddAtox G1397-N
- [178] GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSGGPTYPNYA
NAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMETLLPENAKMTVV
PPEG
- [179] (서열번호 29)
- [180] GGTAGCTACGCACTTGGTCCTTACCAGATTAGCGCACCCCAACTCCCCGC
CTATAATGGTCAAACCGTCGGGACCTTTTACTACGTAAACGATGCTGGTGG
GCTGGAATCCAAAGTATTCTCCTCAGGGGGCCCTACACCCTACCCCAACTA
CGCCAATGCTGGTCATGTAGAAGGGCAGTCAGCACTGTTTATGCGCGATAA
TGGTATAAGCGAGGGGTGGTCTTCCATAACAACCCAGAGGGTACTTGTGG
CTTCTGTGTGAATATGACTGAAACCCTTCTGCCCCGAAAATGCCAAGATGAC
TGTCGTCCACCTGAAGGC
- [181] (서열번호 26) Wild-type DddAtox G1397-C
- [182] AIPVKRGATGETKVFTGNSNSPKSPTKGGC
- [183] (서열번호 30)
- [184] GCCATACCTGTGAAGCGGGGAGCAACAGGGGAGACAAAGGTGTTACAG
GCAACTCTAACAGTCCAAAGAGCCCCACCAAAGGCGGGTGT
- [185] G1333N, G1333C, G1397N 및 G1397C가 조합되어 분할 형태의 탈아미노효소로
사용될 수 있다. 구체적으로, Left-G1333-N + Right-G133-C, Left-G1397-N +
Right-G1397-C, Left-G1397-N + Right-G1333-C, 또는 Left-G1333-N +
Right-G1397-C의 형태로 사용될 수 있다.
- [186]
- [187] 2. 변이체
- [188]
- [189] 본 출원의 발명자들은 아미노산 잔기가 치환된 DdCBE 돌연변이를 통해, 상기

DdCBE 돌연변이가 원하지 않는 염기 교정을 억제하고자 하였다. DdCBE의 비표적 효과를 줄일 수 있는 고정밀의 DddA 유래 시토신 염기 교정기를 제시한다. 이러한 비표적 염기 교정 효과는 TALE과 DNA 사이의 상호작용과 무관하게 DddA_{tox} 탈아미노화 효소 분할체의 자발적인 결합으로 야기되는 현상이다.

- [190] 이 때 돌연변이를 도입하는 아미노산 잔기는 DddAtox의 이합체 (dimer)가 서로 결합하는 표면에 위치하는 접촉 부위이다. DddA_{tox} 분할체 간의 표면에 위치한 아미노산 잔기를 알라닌으로 치환하여 HF-DdCBE를 제작하였으며, HF-DdCBE는 TALE에 연결된 두개의 탈아미노화 효소 쌍이 DNA에 결합하지 못할 경우 제대로 된 기능을 하지 못하도록 하였다. 전 미토콘드리아 유전체 분석을 통해, 사람의 미토콘드리아 DNA에서 수많은 원치않는 비표적 C-to-T 전환을 일으키는 기존 DdCBE와는 달리, HF-DdCBE는 매우 효율적이고 정밀하다는 것을 확인하였다.
- [191] DddAtox의 경우 이론적으로 분할 형태의 이합체 (split dimer)가 모두 DNA의 표적 부위에 적용 (recruit) 되어야 염기 교정을 일으킬 수 있으나, 실제 실험 결과 한쪽 DdCBE는 DNA에 결합된 것과 다른 한쪽은 결합하지 않는 DdCBE를 사용할 경우에도 표적 염기 교정이 일어나게 된다. 이를 해결하고자 분할 형태의 이합체 (split dimer)가 서로 결합하는 단백질 표면의 잔기를 치환하여 원하지 않는 위치에서의 DdCBE pair의 결합을 방지하고자 한다.
- [192] 이를 바탕으로, 본 발명은 시토신 탈아미노화 효소 DddAtox의 아미노산 잔기를 치환하여 비선택적인 염기 교정을 줄이는 새로운 변이체에 관한 것이다.
- [193] 상기 시토신 탈아미노효소는 서열번호 23의 제1분할체 (G1333-N) 및 서열번호 24의 제2분할체 (G1333-C)을 포함하거나, 서열번호 25의 제1분할체 (G1397-N) 및 서열번호 26의 제2분할체 (G1397-C)을 포함한다.
- [194] 상기 시토신 탈아미노효소의 변이체는 서열번호 23의 제1분할체 중 3, 5, 10, 11, 13, 14, 15, 16, 17, 18, 19, 28, 30 및 31번 위치로 구성된 군에서 선택되는 하나 이상의 아미노산이 다른 아미노산으로 치환될 수 있다. 상기 서열번호 24의 제2분할체 중 13, 16, 17, 20, 21, 28, 29, 30, 31, 32, 33, 56, 57, 58 및 60번 위치로 구성된 군에서 선택되는 하나 이상의 아미노산이 다른 아미노산으로 치환된 것이다.
- [195] 본 발명에 따른 변이체는 상기 서열번호 25의 제1분할체 중 87, 88, 91, 92, 95, 100, 101, 102 및 103번 위치로 구성된 군에서 선택되는 하나 이상 또는 상기 서열번호 26의 제2분할체 중 13, 14, 15 및 16번 위치로 구성된 군에서 선택되는 하나 이상의 아미노산이 다른 아미노산으로 치환된 것이다.
- [196] 상기 '다른 아미노산'은, 알라닌, 이소류신, 류신, 메티오닌, 페닐알라닌, 프롤린, 트립토판, 발린, 아스파라긴산, 시스테인, 글루타민, 글리신, 세린, 트레오닌, 티로신, 아스파르트산, 글루탐산, 아르기닌, 히스티딘, 라이신, 상기 아미노산들의 공지된 모든 변형체 중에서, 야생형 단백질이 원래 변이 위치에

갖는 아미노산을 제외한 아미노산들 중에서 선택된 아미노산을 의미한다. 일 예에서, 상기 '다른 아미노산'은 알라닌일 수 있다.

- [197] 구체적으로, 서열번호 23의 제1분할체 중 Y3A, L5A, I10A, S11A, V13A, G14A, T15A, F16A, Y17A, Y18A, V19A, K28A, F30A, S31A 구성된 군에서 선택되는 하나 이상 (각각 Y1292A, L1294A, I1299A, S1300A, V1312A, G1313A, T1314A, F1315A, Y1316A, Y1317A, V1318A, K1327A, F1329A, S1330A에 해당)의 아미노산 치환을 포함할 수 있다.
- [198] 서열번호 24의 제2분할체 중 V13A, Q16A, S17A, F20A, M21A, E28A, G29A, L30A, V31A, F32A, H33A, K56A, M57A, T58A, V60A (각각 V1346A, Q1349A, S1350A, F1353A, M1354A, E1361A, G1362A, L1363A, V1364A, F1365A, H1366A, K1389A, M1390A, T1391A, V1393A 에 해당) 구성된 군에서 선택되는 하나 이상의 아미노산 치환을 포함할 수 있다.
- [199] 구체적으로, 서열번호 25의 제1분할체 중 C87A, V88A, T91A, E92A, L95A, K100A, M101A, T102A, V103A (각각 C1376A, V1377A, T1380A, E1381A, L1384A, K1389A, M1390A, T1391A, V1392A에 해당)으로 구성된 군에서 선택되는 하나 이상의 아미노산 치환을 포함할 수 있다.
- [200] 상기 서열번호 26의 제2분할체 중 K13A, V14A, F15A, T16A (각각 K1410A, V1411A, F1412A, T1413A에 해당)으로 구성된 군에서 선택되는 하나 이상의 아미노산 치환을 포함할 수 있다.
- [201]

[202] [DddAtox G1333-N 변이체]

Type	SEQ	No.
Y1292A	GSAALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKFSSGG	31
	GGCTCTGGTTCCgcaGCCCTGGGTCCATATCAGATTAGTGCTCCCCAACT CCCCGCCTACAACGGTCAGACAGTGGGGACCTTTTACTATGTCAACGAC CCGGGGGATTGGAATCCAAGGTTTTCTCTAGCGGTGGG	32
L1294A	GSYAAGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKFSSGG	33
	GGCTCTGGTTCCCTACGCCgcaGGTCCATATCAGATTAGTGCTCCCCAACTC CCCCGCCTACAACGGTCAGACAGTGGGGACCTTTTACTATGTCAACGACG CCGGGGGATTGGAATCCAAGGTTTTCTCTAGCGGTGGG	34
I1299A	GSYALGPYQASAPQLPAYNGQTVGTFYYVNDAGGLESKFSSGG	35
	GGCTCTGGTTCCCTACGCCCTGGGTCCATATCAGgcaAGTGCTCCCCAACT CCCCGCCTACAACGGTCAGACAGTGGGGACCTTTTACTATGTCAACGAC CCGGGGGATTGGAATCCAAGGTTTTCTCTAGCGGTGGG	36
S1300A	GSYALGPYQIAAPQLPAYNGQTVGTFYYVNDAGGLESKFSSGG	37
	GGCTCTGGTTCCCTACGCCCTGGGTCCATATCAGATTgcaGCTCCCCAACTC CCCCGCCTACAACGGTCAGACAGTGGGGACCTTTTACTATGTCAACGACG CCGGGGGATTGGAATCCAAGGTTTTCTCTAGCGGTGGG	38
V1312A	GSYALGPYQISAPQLPAYNGQTAGTFYYVNDAGGLESKFSSGG	39
	GGCTCTGGTTCCCTACGCCCTGGGTCCATATCAGATTAGTGCTCCCCAACT CCCCGCCTACAACGGTCAGACAgcaGGGACCTTTTACTATGTCAACGACG CCGGGGGATTGGAATCCAAGGTTTTCTCTAGCGGTGGG	40
G1313A	GSYALGPYQISAPQLPAYNGQTVATFYVNDAGGLESKFSSGG	41
	GGCTCTGGTTCCCTACGCCCTGGGTCCATATCAGATTAGTGCTCCCCAACT CCCCGCCTACAACGGTCAGACAGTGgcaACCTTTTACTATGTCAACGACG CCGGGGGATTGGAATCCAAGGTTTTCTCTAGCGGTGGG	42
T1314A	GSYALGPYQISAPQLPAYNGQTVGAFYYVNDAGGLESKFSSGG	43
	GGCTCTGGTTCCCTACGCCCTGGGTCCATATCAGATTAGTGCTCCCCAACT CCCCGCCTACAACGGTCAGACAGTGGGGgcaTTTTACTATGTCAACGACG CCGGGGGATTGGAATCCAAGGTTTTCTCTAGCGGTGGG	44
F1315A	GSYALGPYQISAPQLPAYNGQTVGTAYVNDAGGLESKFSSGG	45
	GGCTCTGGTTCCCTACGCCCTGGGTCCATATCAGATTAGTGCTCCCCAACT CCCCGCCTACAACGGTCAGACAGTGGGGACgcaTACTATGTCAACGACG CCGGGGGATTGGAATCCAAGGTTTTCTCTAGCGGTGGG	46

[203]

Y1316A	GSYALGPYQISAPQLPAYNGQTVGTFAYVNDAGGLESKVFSSGG	47
	GGCTCTGGTTCCTACGCCCTGGGTCCATATCAGATTAGTGCTCCCCAACT CCCCGCCTACAACGGTCAGACAGTGGGGACCTTTgcaTATGTCAACGACG CCGGGGGATTGGAATCCAAGGTTTTCTCTAGCGGTGGG	48
Y1317A	GSYALGPYQISAPQLPAYNGQTVGTFYAVNDAGGLESKVFSSGG	49
	GGCTCTGGTTCCTACGCCCTGGGTCCATATCAGATTAGTGCTCCCCAACT CCCCGCCTACAACGGTCAGACAGTGGGGACCTTTTACgcaGTCAACGACG CCGGGGGATTGGAATCCAAGGTTTTCTCTAGCGGTGGG	50
V1318A	GSYALGPYQISAPQLPAYNGQTVGTFYVNDAGGLESKVFSSGG	51
	GGCTCTGGTTCCTACGCCCTGGGTCCATATCAGATTAGTGCTCCCCAACT CCCCGCCTACAACGGTCAGACAGTGGGGACCTTTTACTATgcaAACGACG CCGGGGGATTGGAATCCAAGGTTTTCTCTAGCGGTGGG	52
K1327A	GSYALGPYQISAPQLPAYNGQTVGTFYVNDAGGLESVAVSSGG	53
	GGCTCTGGTTCCTACGCCCTGGGTCCATATCAGATTAGTGCTCCCCAACT CCCCGCCTACAACGGTCAGACAGTGGGGACCTTTTACTATGTCAACGAC GCCGGGGGATTGGAATCCgcaGTTTTCTCTAGCGGTGGG	54
F1329A	GSYALGPYQISAPQLPAYNGQTVGTFYVNDAGGLESKVAVSSGG	55
	GGCTCTGGTTCCTACGCCCTGGGTCCATATCAGATTAGTGCTCCCCAACT CCCCGCCTACAACGGTCAGACAGTGGGGACCTTTTACTATGTCAACGAC GCCGGGGGATTGGAATCCAAGGTTgcaTCTAGCGGTGGG	56
S1330A	GSYALGPYQISAPQLPAYNGQTVGTFYVNDAGGLESKVFASGG	57
	GGCTCTGGTTCCTACGCCCTGGGTCCATATCAGATTAGTGCTCCCCAACT CCCCGCCTACAACGGTCAGACAGTGGGGACCTTTTACTATGTCAACGAC GCCGGGGGATTGGAATCCAAGGTTTTgcaAGCGGTGGG	58

[204]

[205] [DddAtox G1333-C 변이체]

Type	SEQ	No.
V1346A	PTPYPNYANAGHAE G Q S ALFMRDNGISEGLVFHNNPEGTCGFCVNMTE L L PENAKMTVVPPEGAIPVKRGATGETKVFTGNSNSPKSPTKGGC	59
	CCAACACCTTATCCTAACTACGCTAACGCCGGGCACG ca GAGGGGCAGTC AGCTCTTTTATGAGAGATAACGGCATTAGCGAAGGGCTTGTGTTCCATAA TAATCCTGAGGGGCACCTGTGGCTTCTGTGTAAATATGACCGAAACACTTCT GCCTGAGAACGCTAAATGACTGTCGTACCACCCGAAGGCGCAATCCCA GTAAACGGGGCGCAACCGGCGAAACCAAAGTATTCACCGGAAACAGCA ATAGTCCAAAGTCCCCACCAAGGGAGGTTGC	60
Q1349A	PTPYPNYANAGHVEGA S ALFMRDNGISEGLVFHNNPEGTCGFCVNMTE L L PENAKMTVVPPEGAIPVKRGATGETKVFTGNSNSPKSPTKGGC	61
	CCAACACCTTATCCTAACTACGCTAACGCCGGGCACGTCGAGGGG Gca TC AGCTCTTTTATGAGAGATAACGGCATTAGCGAAGGGCTTGTGTTCCATAA TAATCCTGAGGGGCACCTGTGGCTTCTGTGTAAATATGACCGAAACACTTCT GCCTGAGAACGCTAAATGACTGTCGTACCACCCGAAGGCGCAATCCCA GTAAACGGGGCGCAACCGGCGAAACCAAAGTATTCACCGGAAACAGCA ATAGTCCAAAGTCCCCACCAAGGGAGGTTGC	62
S1350A	PTPYPNYANAGHVEGQ A ALFMRDNGISEGLVFHNNPEGTCGFCVNMTE L L PENAKMTVVPPEGAIPVKRGATGETKVFTGNSNSPKSPTKGGC	63
	CCAACACCTTATCCTAACTACGCTAACGCCGGGCACGTCGAGGGG CAGgc aGCTCTTTTATGAGAGATAACGGCATTAGCGAAGGGCTTGTGTTCCATAA TAATCCTGAGGGGCACCTGTGGCTTCTGTGTAAATATGACCGAAACACTTCT GCCTGAGAACGCTAAATGACTGTCGTACCACCCGAAGGCGCAATCCCA GTAAACGGGGCGCAACCGGCGAAACCAAAGTATTCACCGGAAACAGCA ATAGTCCAAAGTCCCCACCAAGGGAGGTTGC	64
F1353A	PTPYPNYANAGHVEGQSAL A MRDNGISEGLVFHNNPEGTCGFCVNMTE L L PENAKMTVVPPEGAIPVKRGATGETKVFTGNSNSPKSPTKGGC	65
	CCAACACCTTATCCTAACTACGCTAACGCCGGGCACGTCGAGGGG CAGT CAGCTCTT gca ATGAGAGATAACGGCATTAGCGAAGGGCTTGTGTTCCATA ATAATCCTGAGGGGCACCTGTGGCTTCTGTGTAAATATGACCGAAACACTTC TGCCTGAGAACGCTAAATGACTGTCGTACCACCCGAAGGCGCAATCCC AGTTAAACGGGGCGCAACCGGCGAAACCAAAGTATTCACCGGAAACAGC AATAGTCCAAAGTCCCCACCAAGGGAGGTTGC	66
M1354A	PTPYPNYANAGHVEGQSALF A RDNGISEGLVFHNNPEGTCGFCVNMTE L L PENAKMTVVPPEGAIPVKRGATGETKVFTGNSNSPKSPTKGGC	67
	CCAACACCTTATCCTAACTACGCTAACGCCGGGCACGTCGAGGGG CAGT CAGCTCTTTT gca AGAGATAACGGCATTAGCGAAGGGCTTGTGTTCCATA ATAATCCTGAGGGGCACCTGTGGCTTCTGTGTAAATATGACCGAAACACTTC TGCCTGAGAACGCTAAATGACTGTCGTACCACCCGAAGGCGCAATCCC AGTTAAACGGGGCGCAACCGGCGAAACCAAAGTATTCACCGGAAACAGC AATAGTCCAAAGTCCCCACCAAGGGAGGTTGC	68
E1361A	PTPYPNYANAGHVEGQSALFMRDNGIS A GLVFHNNPEGTCGFCVNMTE L L	69

[206]

	PENAKMTVVPPEGAIPVKRGATGETKVFTGNSNSPKSPTKGGC	
	CCAACACCTTATCCTAACTACGCTAACGCCGGGCACGTCGAGGGGCAGT CAGCTCTTTTATGAGAGATAACGGCATTAGCG _{gca} GGGCTTGTGTTCCATAA TAATCCTGAGGGCACCTGTGGCTTCTGTGTAAATATGACCGAAACACTTCT GCCTGAGAACGCTAAAATGACTGTCGTACCACCCGAAGGCGCAATCCCA GTAAACGGGGCGCAACCGGCGAAACCAAAGTATTACCGGAAACAGCA ATAGTCAAAGTCCCCACCAAGGGAGGTTGC	70
G1362A	PTPYPNYANAGHVEGQSALFMRDNGISE _{ALV} FHNNPEGTCGFCVNMTEILL PENAKMTVVPPEGAIPVKRGATGETKVFTGNSNSPKSPTKGGC	71
	CCAACACCTTATCCTAACTACGCTAACGCCGGGCACGTCGAGGGGCAGT CAGCTCTTTTATGAGAGATAACGGCATTAGCGAA _{gca} CTTGTGTTCCATAA TAATCCTGAGGGCACCTGTGGCTTCTGTGTAAATATGACCGAAACACTTCT GCCTGAGAACGCTAAAATGACTGTCGTACCACCCGAAGGCGCAATCCCA GTAAACGGGGCGCAACCGGCGAAACCAAAGTATTACCGGAAACAGCA ATAGTCAAAGTCCCCACCAAGGGAGGTTGC	72
L1363A	PTPYPNYANAGHVEGQSALFMRDNGISE _{AV} FHNNPEGTCGFCVNMTEILL PENAKMTVVPPEGAIPVKRGATGETKVFTGNSNSPKSPTKGGC	73
	CCAACACCTTATCCTAACTACGCTAACGCCGGGCACGTCGAGGGGCAGT CAGCTCTTTTATGAGAGATAACGGCATTAGCGAAGGG _{gca} GTGTTCCATA ATAATCCTGAGGGCACCTGTGGCTTCTGTGTAAATATGACCGAAACACTTC TGCCTGAGAACGCTAAAATGACTGTCGTACCACCCGAAGGCGCAATCCC AGTTAAACGGGGCGCAACCGGCGAAACCAAAGTATTACCGGAAACAGC AATAGTCAAAGTCCCCACCAAGGGAGGTTGC	74
V1364A	PTPYPNYANAGHVEGQSALFMRDNGISE _{GLA} FHNNPEGTCGFCVNMTEILL PENAKMTVVPPEGAIPVKRGATGETKVFTGNSNSPKSPTKGGC	75
	CCAACACCTTATCCTAACTACGCTAACGCCGGGCACGTCGAGGGGCAGT CAGCTCTTTTATGAGAGATAACGGCATTAGCGAAGGGCTT _{gca} TTCATAA TAATCCTGAGGGCACCTGTGGCTTCTGTGTAAATATGACCGAAACACTTCT GCCTGAGAACGCTAAAATGACTGTCGTACCACCCGAAGGCGCAATCCCA GTAAACGGGGCGCAACCGGCGAAACCAAAGTATTACCGGAAACAGCA ATAGTCAAAGTCCCCACCAAGGGAGGTTGC	76
F1365A	PTPYPNYANAGHVEGQSALFMRDNGISE _{GLV} AHNNPEGTCGFCVNMTEILL PENAKMTVVPPEGAIPVKRGATGETKVFTGNSNSPKSPTKGGC	77
	CCAACACCTTATCCTAACTACGCTAACGCCGGGCACGTCGAGGGGCAGT CAGCTCTTTTATGAGAGATAACGGCATTAGCGAAGGGCTTGTG _{gca} CATA ATAATCCTGAGGGCACCTGTGGCTTCTGTGTAAATATGACCGAAACACTTC TGCCTGAGAACGCTAAAATGACTGTCGTACCACCCGAAGGCGCAATCCC AGTTAAACGGGGCGCAACCGGCGAAACCAAAGTATTACCGGAAACAGC AATAGTCAAAGTCCCCACCAAGGGAGGTTGC	78
H1366A	PTPYPNYANAGHVEGQSALFMRDNGISE _{GLVF} ANNPEGTCGFCVNMTEILL PENAKMTVVPPEGAIPVKRGATGETKVFTGNSNSPKSPTKGGC	79
	CCAACACCTTATCCTAACTACGCTAACGCCGGGCACGTCGAGGGGCAGT CAGCTCTTTTATGAGAGATAACGGCATTAGCGAAGGGCTTGTGTT _{gca} A ATAATCCTGAGGGCACCTGTGGCTTCTGTGTAAATATGACCGAAACACTTC TGCCTGAGAACGCTAAAATGACTGTCGTACCACCCGAAGGCGCAATCCC	80

[207]

	AGTTAAACGGGGCGCAACCGGCGAAACCAAAGTATTCACCGGAAACAGC AATAGTCCAAAGTCCCCACCAAGGGAGGTTGC	
K1389A	PTPYPNYANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMTEILL PENAA M TVVPPEGAIPVKRGATGETKVFTGNSNSPKSPTKGGC	81
	CCAACACCTTATCCTAACTACGCTAACGCCGGGCACGTCGAGGGGCAGT CAGCTCTTTTTATGAGAGATAACGGCATTAGCGAAGGGCTTGTGTTCCATA ATAATCCTGAGGGCACCTGTGGCTTCTGTGTAATATGACCGAAACACTTC TGCCTGAGAACGCT gca ATGACTGTCGTACCACCCGAAGGGCGCAATCCCA GTAAACGGGGCGCAACCGGCGAAACCAAAGTATTCACCGGAAACAGCA ATAGTCCAAAGTCCCCACCAAGGGAGGTTGC	82
M1390A	PTPYPNYANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMTEILL PENAK A TVVPPEGAIPVKRGATGETKVFTGNSNSPKSPTKGGC	83
	CCAACACCTTATCCTAACTACGCTAACGCCGGGCACGTCGAGGGGCAGT CAGCTCTTTTTATGAGAGATAACGGCATTAGCGAAGGGCTTGTGTTCCATA ATAATCCTGAGGGCACCTGTGGCTTCTGTGTAATATGACCGAAACACTTC TGCCTGAGAACGCTAA A gca ACTGTCGTACCACCCGAAGGGCGCAATCCCA GTAAACGGGGCGCAACCGGCGAAACCAAAGTATTCACCGGAAACAGCA ATAGTCCAAAGTCCCCACCAAGGGAGGTTGC	84
T1391A	PTPYPNYANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMTEILL PENAKM A VVPPEGAIPVKRGATGETKVFTGNSNSPKSPTKGGC	85
	CCAACACCTTATCCTAACTACGCTAACGCCGGGCACGTCGAGGGGCAGT CAGCTCTTTTTATGAGAGATAACGGCATTAGCGAAGGGCTTGTGTTCCATA ATAATCCTGAGGGCACCTGTGGCTTCTGTGTAATATGACCGAAACACTTC TGCCTGAGAACGCTAA A AT gca GTCGTACCACCCGAAGGGCGCAATCCCA GTAAACGGGGCGCAACCGGCGAAACCAAAGTATTCACCGGAAACAGCA ATAGTCCAAAGTCCCCACCAAGGGAGGTTGC	86
V1393A	PTPYPNYANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMTEILL PENAKMTV A PPEGAIPVKRGATGETKVFTGNSNSPKSPTKGGC	87
	CCAACACCTTATCCTAACTACGCTAACGCCGGGCACGTCGAGGGGCAGT CAGCTCTTTTTATGAGAGATAACGGCATTAGCGAAGGGCTTGTGTTCCATA ATAATCCTGAGGGCACCTGTGGCTTCTGTGTAATATGACCGAAACACTTC TGCCTGAGAACGCTAA A AT gca CCACCCGAAGGGCGCAATCCCA GTAAACGGGGCGCAACCGGCGAAACCAAAGTATTCACCGGAAACAGCA ATAGTCCAAAGTCCCCACCAAGGGAGGTTGC	88

[208]

[209] [DddAtox G1397-N 변이체]

Type	SEQ	No.
C1376A	GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGGPTPYPNY ANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFC A VNMTETLLPENAKMT VVPPEG	89
	GGTAGCTACGCACTTGGTCCTTACCAGATTAGCGCACCCCAACTCCCCGC CTATAATGGTCAAACCGTCGGGACCTTTTACTACGTAAACGATGCTGGTG GGCTGGAATCAAAGTATTCTCCTCAGGGGGCCCTACACCCTACCCCAAC TACGCCAATGCTGGTCATGTAGAAGGGCAGTCAGCACTGTTTATGCGCGA TAATGGTATAAGCGAGGGGTTGGTCTTCCATAACAACCCAGAGGGTACTT GTGGCTTC G caGTGAATATGACTGAAACCCCTTCTGCCCGAAAATGCCAAGA TGA CT GTCTCGTCCACCTGAAGGC	90
V1377A	GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGGPTPYPNY ANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFC A NMTETLLPENAKM TVVPPEG	91
	GGTAGCTACGCACTTGGTCCTTACCAGATTAGCGCACCCCAACTCCCCGC CTATAATGGTCAAACCGTCGGGACCTTTTACTACGTAAACGATGCTGGTG GGCTGGAATCAAAGTATTCTCCTCAGGGGGCCCTACACCCTACCCCAAC TACGCCAATGCTGGTCATGTAGAAGGGCAGTCAGCACTGTTTATGCGCGA TAATGGTATAAGCGAGGGGTTGGTCTTCCATAACAACCCAGAGGGTACTT GTGGCTTC T GT g caAATATGACTGAAACCCCTTCTGCCCGAAAATGCCAAGA TGA CT GTCTCGTCCACCTGAAGGC	92
T1380A	GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGGPTPYPNY ANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNM A ETLLPENAKM TVVPPEG	93
	GGTAGCTACGCACTTGGTCCTTACCAGATTAGCGCACCCCAACTCCCCGC CTATAATGGTCAAACCGTCGGGACCTTTTACTACGTAAACGATGCTGGTG GGCTGGAATCAAAGTATTCTCCTCAGGGGGCCCTACACCCTACCCCAAC TACGCCAATGCTGGTCATGTAGAAGGGCAGTCAGCACTGTTTATGCGCGA TAATGGTATAAGCGAGGGGTTGGTCTTCCATAACAACCCAGAGGGTACTT GTGGCTTC T GTGTGAATATG g caGAAACCCCTTCTGCCCGAAAATGCCAAGA TGA CT GTCTCGTCCACCTGAAGGC	94
E1381A	GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGGPTPYPNY ANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMT A TLLPENAKMT VVPPEG	95
	GGTAGCTACGCACTTGGTCCTTACCAGATTAGCGCACCCCAACTCCCCGC CTATAATGGTCAAACCGTCGGGACCTTTTACTACGTAAACGATGCTGGTG GGCTGGAATCAAAGTATTCTCCTCAGGGGGCCCTACACCCTACCCCAAC TACGCCAATGCTGGTCATGTAGAAGGGCAGTCAGCACTGTTTATGCGCGA TAATGGTATAAGCGAGGGGTTGGTCTTCCATAACAACCCAGAGGGTACTT GTGGCTTC T GTGTGAATATGACT g caACCCTTCTGCCCGAAAATGCCAAGA TGA CT GTCTCGTCCACCTGAAGGC	96
L1384A	GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGGPTPYPNY ANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMTETL A PENAKM TVVPPEG	97
	GGTAGCTACGCACTTGGTCCTTACCAGATTAGCGCACCCCAACTCCCCGC CTATAATGGTCAAACCGTCGGGACCTTTTACTACGTAAACGATGCTGGTG GGCTGGAATCAAAGTATTCTCCTCAGGGGGCCCTACACCCTACCCCAAC TACGCCAATGCTGGTCATGTAGAAGGGCAGTCAGCACTGTTTATGCGCGA TAATGGTATAAGCGAGGGGTTGGTCTTCCATAACAACCCAGAGGGTACTT GTGGCTTC T GTGTGAATATGACTGAAACCCCTT g caCCCGAAAATGCCAAGA	98

[210]

	TGACTGTCGTCCCACCTGAAGGC	
K1389A	GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKFSSGGPTPYPNY ANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMTEILLPENAAAM TVVPPEG	99
	GGTAGCTACGCACTTGGTCCTTACCAGATTAGCGCACCCCAACTCCCCGC CTATAATGGTCAAACCGTCGGGACCTTTTACTACGTAAACGATGCTGGTG GGCTGGAATCAAAGTATTCTCCTCAGGGGGCCCTACACCCTACCCCAAC TACGCCAATGCTGGTCATGTAGAAGGGCAGTCAGCACTGTTTATGCGCGA TAATGGTATAAGCGAGGGGTTGGTCTTCCATAACAACCCAGAGGGTACTT GTGGCTTCTGTGTGAATATGACTGAAACCCCTTCTGCCCGAAAATGCCgcaA TGACTGTCGTCCCACCTGAAGGC	100
M1390A	GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKFSSGGPTPYPNY ANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMTEILLPENAKAT VVPPEG	101
	GGTAGCTACGCACTTGGTCCTTACCAGATTAGCGCACCCCAACTCCCCGC CTATAATGGTCAAACCGTCGGGACCTTTTACTACGTAAACGATGCTGGTG GGCTGGAATCAAAGTATTCTCCTCAGGGGGCCCTACACCCTACCCCAAC TACGCCAATGCTGGTCATGTAGAAGGGCAGTCAGCACTGTTTATGCGCGA TAATGGTATAAGCGAGGGGTTGGTCTTCCATAACAACCCAGAGGGTACTT GTGGCTTCTGTGTGAATATGACTGAAACCCCTTCTGCCCGAAAATGCCAAG gcaACTGTCGTCCCACCTGAAGGC	102
T1391A	GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKFSSGGPTPYPNY ANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMTEILLPENAKM AVPPEG	103
	GGTAGCTACGCACTTGGTCCTTACCAGATTAGCGCACCCCAACTCCCCGC CTATAATGGTCAAACCGTCGGGACCTTTTACTACGTAAACGATGCTGGTG GGCTGGAATCAAAGTATTCTCCTCAGGGGGCCCTACACCCTACCCCAAC TACGCCAATGCTGGTCATGTAGAAGGGCAGTCAGCACTGTTTATGCGCGA TAATGGTATAAGCGAGGGGTTGGTCTTCCATAACAACCCAGAGGGTACTT GTGGCTTCTGTGTGAATATGACTGAAACCCCTTCTGCCCGAAAATGCCAAG ATGgcaGTCGTCCCACCTGAAGGC	104
V1392A	GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKFSSGGPTPYPNY ANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMTEILLPENAKM TAVPPEG	105
	GGTAGCTACGCACTTGGTCCTTACCAGATTAGCGCACCCCAACTCCCCGC CTATAATGGTCAAACCGTCGGGACCTTTTACTACGTAAACGATGCTGGTG GGCTGGAATCAAAGTATTCTCCTCAGGGGGCCCTACACCCTACCCCAAC TACGCCAATGCTGGTCATGTAGAAGGGCAGTCAGCACTGTTTATGCGCGA TAATGGTATAAGCGAGGGGTTGGTCTTCCATAACAACCCAGAGGGTACTT GTGGCTTCTGTGTGAATATGACTGAAACCCCTTCTGCCCGAAAATGCCAAG ATGACTgcaGTCCCACCTGAAGGC	106

[211]

[212] [DddAtox G1397-C 변이체]

Type	SEQ	No.
K1410A	AIPVKRGATGETAVFTGNSNSPKSPTKGGC	107
	GCCATACCTGTGAAGCGGGGAGCAACAGGGGAGACAgcaGTGTTACAG GCAACTCTAACAGTCCAAAGAGCCCCACCAAAGGCGGGTGT	108
V1411A	AIPVKRGATGETKAFTGNSNSPKSPTKGGC	109
	GCCATACCTGTGAAGCGGGGAGCAACAGGGGAGACAAAGgcaTTCACAG GCAACTCTAACAGTCCAAAGAGCCCCACCAAAGGCGGGTGT	110
F1412A	AIPVKRGATGETKVATGNSNSPKSPTKGGC	111
	GCCATACCTGTGAAGCGGGGAGCAACAGGGGAGACAAAGGTGgcaACAG GCAACTCTAACAGTCCAAAGAGCCCCACCAAAGGCGGGTGT	112
T1413A	AIPVKRGATGETKVFAGNSNSPKSPTKGGC	113
	GCCATACCTGTGAAGCGGGGAGCAACAGGGGAGACAAAGGTGTTcgaG GCAACTCTAACAGTCCAAAGAGCCCCACCAAAGGCGGGTGT	114

[213]

[214] 본 발명에 따른 시토신 탈아미노효소 변이체는 위 표와 같이 기재된 아미노산 서열로 구성된 군에서 선택되는 하나 이상의 서열을 포함할 수 있다. 본 발명에 따른 시토신 탈아미노효소 변이체는 비특이적인 표적 부위 내 여러 염기에 원치 않는 교정을 일으키는 효과를 줄일 수 있는 가능성을 보여준다.

[215]

[216] 3. 전장 시토신 탈아미노효소

[217]

[218] 본 출원의 발명자들은 세포내 독성을 일으켜 분리된 형태 (split)로 사용하는 야생형 시토신 탈아미노효소 (Cytosine deaminase) DddA_{tox}의 양전하를 가지는 아미노산을 변형하여 만든 전장 DddA를 이용한 새로운 프로그램가능한 시토신 탈아미노효소를 개발하였다.

[219] 본 발명은 (i) DNA 결합 단백질; 및 (ii) 시토신 탈아미노효소 (deaminase) 또는 이의 변이체를 포함하는 융합단백질로, 상기 시토신 탈아미노효소 또는 이의 변이체는 비독성 전장 시토신 탈아미노효소인, 융합단백질에 관한 것이다.

[220] DddA_{tox}의 C-말단에는 양전하를 가지는 아미노산이(KRKKK) 특이적으로 몰려 있다. DNA는 음전하를 띄기 때문에 단백질의 양전하 성질을 가지는 아미노산과 결합한다. 이 양전하를 가지는 아미노산을 전극이 없는 아미노산으로 치환하여 DddA_{tox}가 DNA와 결합하는 힘을 약하게 하여 세포내 독성을 낮추거나 제거하려 한다. 즉, 양전하를 가지는 아미노산을 극성이 없는 아미노산으로 치환하여 독성을 가지지 않는 조합은 대장균을 이용한 클로닝이 가능하여 전장 DddA를 확보할 수 있다.

[221] 야생형 DddA_{tox}는 세포내 독성 때문에 두 개로 나눈 분리된 형태를 이용하는데 이는 실험에 많은 제약이 따른다. 특히, Cas9을 이용할 경우 다른 PAM을 인식하는 오르쏘고날 (orthogonal) Cas9 변이체를 이용한다. 이 경우 PAM이 제한적으로 존재하기 때문에 원하는 위치에 정확히 시토신 (cytosine)을 티민(thymine)으로 탈아미노화 (deamination) 시키는 것이 어려운 경우가 많다. 또한, 가장 높은 활성이 나타나는 타겟 윈도우 (target window)는 두개의 Cas9 변이체가 바인딩하는 사이의 40bp로 이 부위에 있는 원하지 않는 시토신까지 탈아미노화된다. 그러나, 전장 DddA는 분리되지 않기 때문에 PAM의 제약을 받지 않는다. 또한, 타겟 위치에서 10bp 안에 있는 TC motif에서 가장 활성이 좋기 때문에 정확성 또한 높다.

[222] Cas9이 타겟 사이트 (target site)에 결합해서 이루는 R-loop에 있는 ACA, GC, CC 모티프에 있는 시토신 (cytosine)을 티민(thymine)으로 탈아미노화하는 것을 확인하였다. 이는 분리된 형태에서는 확인이 되지 않는 활성이다.

[223] 전장 DddA는 Cas9뿐만 아니라, TALE 모듈, 징크 핑거 단백질을 이용하여 원하는 위치의 시토신 (cytosine)을 티민(thymine)으로 치환하는 것이 가능하다. 기존의 DddAtox는 분리된 형태라 쌍 (pair)으로 전달해야 했지만, 전장 DddA는

TALE 모듈, 징크 핑거 단백질의 한 모듈만을 이용할 수 있기 때문에 타겟 위치를 제약없이 선택할 수 있다. 또한, 게놈 사이트 (genomic site) 뿐 아니라 미토콘드리아 (mitochondria)나 식물 엽록체 (chloroplast) 또는 색소체 등의 DNA를 타겟하여 특정 DNA의 시토신 (cytosine)을 티민(thymine)으로 변환할 수 있다.

- [224] 또한, 크기가 작아 유전자 치료에 사용되는 바이러스 벡터인 AAV에 모든 구조체를 넣을 수 있다. 기존에 사용하는 CBE(cytosine base editor)는 Cas9이 타겟 사이트와 바인딩하여 이루는 R-loop에 있는 시토신 (cytosine)을 티민(thymine)으로 치환하는데, 이번에 발명된 전장 DddA는 R-loop의 밖에 있는 시토신을 탈아미노화한다. 따라서 기존의 CBE로는 교정에 제약이 있는 위치의 경우 시토신 (cytosine)을 티민(thymine)으로 변환하는 것이 가능하다.
- [225] 이에 기반하여, 상기 비독성 전장 시토신 탈아미노효소는 서열번호 1의 야생형 탈아미노효소 중 1개 이상, 2개 이상, 3개 이상, 4개 이상 또는 5개 이상의 아미노산이 다른 아미노산으로 치환될 수 있다.
- [226] 상기 '다른 아미노산'은, 알라닌, 이소류신, 류신, 메티오닌, 페닐알라닌, 프롤린, 트립토판, 발린, 아스파라긴산, 시스테인, 글루타민, 글리신, 세린, 트레오닌, 티로신, 아스파르트산, 글루탐산, 아르기닌, 히스티딘, 라이신, 상기 아미노산들의 공지된 모든 변형체 중에서, 야생형 단백질이 원래 변이 위치에 갖는 아미노산을 제외한 아미노산들 중에서 선택된 아미노산을 의미한다. 일 예에서, 상기 '다른 아미노산'은 알라닌일 수 있다.
- [227] 상기 비독성 전장 DddA는 종류에 따라 다음의 서열번호 12 내지 서열번호 18로 구성된 군에서 선택되는 서열을 포함할 수 있다.
- [228] Wild type (SEQ ID No: 1)
- [229] GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSGGPTYPNYA
NAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMETLLPENAKMTVV
PPEGAIPVKRGATGETKVFTGNSNSPKSPTKGGC
- [230] A1341D KRKKA (SEQ ID No: 12)
- [231] GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSGGPTYPNYD
NAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMETLLPENAKMTVV
PPEGAIPVKRGATGETKVFTGNSNSPKSPTAGGC
- [232] AAAAA (SEQ ID No: 13)
- [233] GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSGGPTYPNYA
NAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMETLLPENAKMTVV
PPEGAIPVAAGATGETAVFTGNSNSPASPTAGGC
- [234] AAAAK (SEQ ID No: 14)
- [235] GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSGGPTYPNYA
NAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMETLLPENAKMTVV
PPEGAIPVAAGATGETAVFTGNSNSPASPTKGGC

- [236] AAKAA (SEQ ID No: 15)
- [237] GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKFSSGGPTYPNYA
NAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMETLLPENAKMTVV
PPEGAIPVAAGATGETKVFTGNSNSPASPTAGGC
- [238] AAKAK (SEQ ID No: 16)
- [239] GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKFSSGGPTYPNYA
NAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMETLLPENAKMTVV
PPEGAIPVAAGATGETKVFTGNSNSPASPTKGGC
- [240] KAAAA (SEQ ID No: 17)
- [241] GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKFSSGGPTYPNYA
NAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMETLLPENAKMTVV
PPEGAIPVKAGATGETAVFTGNSNSPASPTAGGC
- [242] E1347A (SEQ ID No: 18)
- [243] GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKFSSGGPTYPNYA
NAGHVAGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMETLLPENAKMTVV
PPEGAIPVKRGATGETKVFTGNSNSPKSPTKGGC
- [244]
- [245] 상기 전장 탈아미노효소 변이체는 서열번호 1의 아미노산 서열 중 다음으로
구성되는 군에서 선택되는 하나 이상의 치환을 포함할 수 있다:
- [246] 37 위치의 S가 G로 치환;
- [247] 59번 위치의 G가 S로 치환;
- [248] 109번 위치의 A가 V로 치환; 및
- [249] 129번 위치의 S가 G로 치환.
- [250] 하나의 실시예에서, 상기 전장 탈아미노효소 변이체는 서열번호 1의 아미노산
서열 중 37 위치의 S가 G로 치환; 59번 위치의 G가 S로 치환; 109번 위치의 A가
V로 치환; 및 129번 위치의 S가 G로 치환을 포함하는, 서열번호 19의 서열을
포함할 수 있다.
- [251] GSVG (SEQ ID No: 19)
- [252] GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLEKVFSSGGPTYPNY
ANAGHVESQSALFMRDNGISEGLVFHNNPEGTCGFCVNMETLLPENAKMTV
VPPEGVIPVKRGATGETKVFTGNSNGPKSPTKGGC
- [253] 상기 전장 DddA GSVG는 일반적인 대장균을 이용한 클로닝이 가능하다. 전장
DddA GSVG가 타겟 사이트에 있는 TC motif의 시토신을 티민으로 탈아미노화
하는 것을 인간 세포 게놈 레벨 (human cell genomic context)에서 확인하였다.
전장 DddA GSVG을 Cas9의 N-terminal과 C-terminal에 각각 클로닝할 수 있다.
같은 타겟 사이트에서 Cas9의 N-terminal에 연결된 DddA GSVG는 시토신을
티민으로 치환할 수 있다. Cas9의 C-terminal에 연결된 DddA GSVG는 치환을
원하는 구아닌의 상보 서열에 있는 TC motif의 시토신을 티민으로 치환하여,

치환을 원하는 구아닌을 아데닌으로 치환하는 것을 인간 세포 게놈 레벨에서 확인하였다.

[254]

[255] 하나의 실시예에서, 상기 전장 탈아미노효소 변이체는 서열번호 20 내지 22로 구성된 군에서 선택되는 서열을 포함할 수 있다.

[256] SSVG (SEQ ID No: 20)

[257] GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSGGPTPYPNYA
NAGHVESQSALFMRDNGISEGLVFHNNPEGTCGFCVNMETLLPENAKMTVV
PPEGVIPVKRGATGETKVFTGNSNGPKSPTKGGC

[258] GSAG (SEQ ID No: 21)

[259] GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLEGKVFSGGPTPYPNY
ANAGHVESQSALFMRDNGISEGLVFHNNPEGTCGFCVNMETLLPENAKMTV
VPPEGAIPVKRGATGETKVFTGNSNGPKSPTKGGC

[260] GSVS (SEQ ID No: 22)

[261] GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLEGKVFSGGPTPYPNY
ANAGHVESQSALFMRDNGISEGLVFHNNPEGTCGFCVNMETLLPENAKMTV
VPPEGVIPVKRGATGETKVFTGNSNSPKSPTKGGC

[262]

[263] 4. DNA 결합 단백질

[264]

[265] 상기 DNA 결합 단백질은 예를 들어, 징크 핑거 단백질 (zinc finger protein), TALE (Transcription activator-like effector) 단백질, CRISPR 연관 뉴클레아제 또는 이들 중 2이상의 조합일 수 있다.

[266] 상기 징크 핑거 단백질의 징크 핑거 모티프는 DNA 결합 도메인을 가지고, 핑거의 c-말단 부분은 DNA 서열을 특이적으로 인식한다. 3 ~ 6 개의 징크 핑거 모티프를 포함하는 DNA 결합 단백질이 DNA 서열을 인식한다.

[267] 하나의 실시예에서, 상기 시토신 탈아미노효소의 제1분할체 및 시토신 탈아미노효소의 제2분할체 각각은 징크 핑거 단백질의 N-말단 또는 C-말단에 결합할 수 있다.

[268] 시토신 탈아미노효소의 제1분할체 N-말단에 징크 핑거 단백질 (ZF-Left)의 C-말단이 결합하고 시토신 탈아미노효소의 제2분할체 N-말단에 징크 핑거 단백질 (ZF-Right) C-말단이 결합 (CC configuration);

[269] 시토신 탈아미노효소의 제1분할체 C-말단에 징크 핑거 단백질 (ZF-Left)의 N-말단이 결합하고 시토신 탈아미노효소의 제2분할체 N-말단에 징크 핑거 단백질 (ZF-Right) C-말단이 결합 (NC configuration);

[270] 시토신 탈아미노효소의 제1분할체 N-말단에 징크 핑거 단백질 (ZF-Left)의 C-말단이 결합하고 시토신 탈아미노효소의 제2분할체 C-말단에 징크 핑거 단백질 (ZF-Right) N-말단이 결합 (CN configuration); 또는

- [271] 시토신 탈아미노효소의 제1분할체 C-말단에 징크 핑거 단백질 (ZF-Left)의 N-말단이 결합하고 시토신 탈아미노효소의 제2분할체 C-말단에 징크 핑거 단백질 (ZF-Right) N-말단이 결합 (NN configuration)할 수 있다.
- [272] 상기 ZF-Left은 다음의 서열번호 2의 서열을 포함할 수 있다:
- [273] [서열번호 2]
- [274] GIHGVPAAMAERPFQCRICMRNFSDRSNLSRHIRTHTGKPFACDICGRKFAI
SSNLNSHTKIHTGSQKPFQCRICMRNFSDNLSRHIRTHTGKPFACDICGRKF
ATSGNLTRHTKIHLR.
- [275] 상기 ZF-Right는 다음의 서열번호 3의 서열을 포함할 수 있다:
- [276] [서열번호 3]
- [277] GIHGVPAAMAERPFQCRICMRNFSDNLSVHIRTHTGKPFACDICGRKFAQ
KINLQVHTKIHTGKPFQCRICMRNFSDVLSEHIRTHTGKPFACDICGRKFA
QRNHRTHHTKIHLR.
- [278] ZF은 DNA 타겟에 따라 서열이 달라질 수 있다. ZF은 DNA 타겟 서열에 따라 맞춤형 제작될 수 있다. ZF는 DNA 3bp를 인식하기 때문에 보통 3~6개의 ZF를 이어 붙여 9~18bp DNA를 인식하는 ZF 조합을 제작할 수 있다. 예를 들어, GNN, TNN, CNN 또는 ANN 등의 모듈이 포함된 라이브러리를 이용하여 제작할 수 있다.
- [279] 경우에 따라서, 상기 징크 핑거 단백질은 링커를 통해 탈아미노효소와 연결될 수 있다. 상기 링커는 2 내지 40개 아미노산 잔기를 포함하는 펩타이드 링커일 수 있다. 상기 링커는 예를 들어, 2aa, 5aa, 10aa, 16aa, 24aa 또는 32aa 길이의 링커 등일 수 있으나, 이에 제한되는 것은 아니다.
- [280] 하나의 실시예에서, 상기 링커는 다음을 포함할 수 있다:
- [281] 2a.a 링커: GS
- [282] 5a.a 링커: TGEKQ (서열번호 8)
- [283] 10a.a 링커: SGAQGSTLDF (서열번호 9)
- [284] 16a.a 링커: SGSETPGTSESATPES (서열번호 10);
- [285] 24 a.a 링커: SGTPEHVG VYTLSGTPEHVG VYTL (서열번호 115); 또는
- [286] 32a.a 링커: GSGGSSGGSSGSETPGTSESATPESSGGSSGGS (서열번호 11).
- [287] 본 발명에 따른 구체적 실시예에서, 분할 탈아미노효소와 징크 핑거 단백질이 링커를 통해 연결될 수 있고, 제1분할체를 포함하는 반쪽 탈아미노효소 N-말단에 징크 핑거 단백질이 결합하고, 제2분할체를 포함하는 반쪽 탈아미노효소 N-말단에 징크 핑거 단백질이 결합한다. 이 때, 왼쪽과 오른쪽 ZFP가 붙는 곳 사이인 스페이스에서 C에서 T로의 염기 전환이 일어날 수 있다. 왼쪽과 오른쪽 ZFP 모두 24 a.a 링커를 통해 제1분할체를 포함하는 반쪽 탈아미노효소 및 제2분할체를 포함하는 반쪽 탈아미노효소 각각과 연결되는 경우, 높은 교정 효율을 보임을 확인하였다.
- [288] 상기 TAL 이펙터 (TALE)는 33개에서 34개의 반복된 아미노산 서열이 반복된

형태로 이루어져 있으며, 도메인(RVD, Repeated variant domain)이 약 9개 정도 반복된다. 도메인 하나당 한 개의 뉴클레오타이드를 인지 할 수 있으며, 12-13번째 아미노산 서열 에 따라 특이적인 DNA 서열과 결합 할 수 있다(HD->Cytosine, NI->Adenine, NG -> Thymine, NN -> Guanine). 탈이펙터(TALE)는 표적부위 내의 단일 DNA 가닥을 인식한다. 표적 부위 사이의 거리는 12-14 뉴클레오타이드일 수 있다.

- [289] 상기 TALE 도메인은 하나 이상의 TALE-Repeat의 조합에 의해 서열-특이적 방식으로 뉴클레오타이드에 결합하는 단백질 도메인을 의미한다. 적어도 하나 이상의 TALE-Repeat, 구체적으로 1 내지 30개의 TALE-Repeat를 포함하나 이에 한정되지 않는다. TALE-Repeat는 TALE 도메인 내 특정 뉴클레오타이드 서열을 인식하는 부위이다.
- [290] 상기 TALE 도메인에 백본 구조로 TALE의 N-말단 포함 영역 및 TALE의 C-말단 포함 영역이 포함된다. 상기 TALE의 N-말단 포함 제 1 TALE은 서열번호 4 또는 5에 의해 코딩될 수 있다. 상기 TALE의 C-말단 포함 제 2 TALE은 서열번호 6 또는 7에 의해 코딩될 수 있다.

[291]

Name	Sequence	No.
제 1 TALE N- term	GATATAGCTGATTTGAGGACCCTCGGGTACTCACAGCAGCAACAAGAAAA AATCAAACCAAAGGTACGTAGCACGGTTGCCCAACATCATGAAGCCCTAG TTGGTCACGGATTTACTCATGCCCATATTGTTGCACTGAGTCAACACCCTGC TGCATTAGGCACAGTGGCTGTGAAATATCAAGATATGATTGCTGCTCTTCCT GAGGCCACACACGAGGCAATTGTGGGTGTTGGAAAACAATGGTCTGGTG CGCGAGCGCTTGAAGCTCTCCTAACCGTAGCTGGAGAGTTGAGAGGCCCA CCGCTTCAGCTTGACACAGGGCAGCTGTTGAAGATAGCAAAGAGGGGGG GTGTAAGTCTGTTGAAGCAGTCCATGCATGGAGAAATGCATTAAGTGGAG CTCCA	4
	DIADLRTLGYQQQKEKIPKVRSTVAQHHEALVGHGFTHAHIVALSQHPA ALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPL QLDTGQLLKIARKGGVTAVEAVHAWRNALTGAP	269
제 2 TALE N- term	GATATTGCTGATCTCGAACATTGGGTATTCTCAGCAGCAACAAGAGAAA ATCAAACCGAAAGTCCGTTCAACCGTAGCTCAACATCATGAAGCCCTTGTT GGTCATGGATTTACTCATGCTCACATTGTTGCTCTAAGTCAGCACCCCTGCCG CGCTAGGCACAGTTGCCGTAAGTATCAAGATATGATAGCTGCTTTACCTG AAGCAACCCACGAGGCAATAGTGGGGGTGGGAAGAGGGGTGCAGGCG CCAGGGCACTGGAGGCACTCTTAACAGTAGCTGGAGAACTTCGGGGACC	5

[292]		ACCATTGCAGCTTGATACTGGACAACTCTTGAAGATCGCAAAAAGAGGGG GTGTGACTGCTGTGGAAGCTGTTCATGCATGGAGAAATGCTTTGACGGGA GCACCCCTG	
		DIADLRTLGYSSQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPA ALGTVAVKYQDMIAALPEATHEAIVGVGKRGAGARALEALLTVAGELRGPPL QLDTGQLLKIARKGGVTAVEAVHAWRNALTGAPL	270
제 1 TALE C- term		AGCATAGTTGCTCAGCTAAGTAGACCTGATCCAGCACTTGCTGCATTGACG AACGATCATCTCGTAGCGCTGGCTGTTTGGGCGGCAGGCCAGCCTTGA CGCCGTGAAGAAAGGATTAGGT	6
		SIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLG	271
제 2 TALE C- term		TCGATAGTAGCACAATTAAGCAGGCCTGACCCTGCACTTGCTGCTCTTACA AATGATCATCTTGTGCTCTGGCCTGTTTGGGTGGGCGTCCCGCGTTGGATG CTGTTAAGAAAGGACTTGGT	7
		SIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLG	271

[293] 절단 부위를 기준으로 상기 TALE 도메인이 결합하는 위치에 따라, 단일 TALE 어레이 (array) 또는 제 1 TALE 어레이 및 제 2 TALE 어레이가 각각 결합할 수 있다.

[294] 상기 시토신 탈아미노효소의 제1분할체에 제 1 TALE이 결합 (left TALE)되고, 상기 시토신 탈아미노효소의 제2분할체에 제 2 TALE (right TALE)이 결합될 수 있다. 각각 N'-TALE-제1분할체-C' 및 N'-TALE-제2분할체-C' 구조를 가진다.

[295] 상기 시토신 탈아미노효소가 전장인 경우 시토신 탈아미노효소의 N 말단에 단일 모듈 TALE이 결합될 수 있다. N-C 방향으로 단일 TALE 모듈 및 시토신 탈아미노효소가 포함된다. 이중 모듈 TALE이 포함될 수 있고, 이 때 전장 시토신 탈아미노효소의 N 말단에 제1 TALE이 결합되고, 제2 TALE이 별개로 포함될 수 있다. N-C 방향으로 제1 TALE 모듈 및 시토신 탈아미노효소가 포함되고, N'-TALE-시토신 탈아미노효소-C' 및 N'-TALE-C'의 구조를 가진다.

[296] TALE array는 타겟 DNA 서열에 따라 맞춤 제작할 수 있다. TALE array는 33~35 개의 아미노산 잔기로 구성된 모듈이 반복적으로 배열되어 구성되는 데 이들은 식물 병원균 *Xanthomonas*에서 유래한 것으로 모듈 하나가 A, C, G, T 염기

- 하나씩을 각각 인식하여 DNA에 결합한다. 각각의 모듈의 염기 특이성은 12, 13번째 아미노산 잔기에 의해 결정되는데 이들을 repeat variable di-residue(RVD)라고 한다. 예를 들어 RVD가 NN인 모듈은 G를 인식하고 NI는 A, HD는 C, NG는 T를 인식한다. TALE array는 최소 14 개에서 최대 18 개 이상의 모듈로 구성되어 타겟 DNA 서열 15bp~20bp를 인식할 수 있도록 설계할 수 있다.
- [297] 상기 CRISPR 연관 뉴클레아제와 관련하여, CRISPR array에는 두개의 RNA가 암호화 되어 있는데, 하나는 crRNA(CRISPR RNA)이고, 다른 하나는 tracrRNA(trans activating CRISPR RNA)이다. crRNA는 protospacer 부분에서 전사되며, tracrRNA와 결합하여 3차 구조를 형성한다. 두 종류의 RNA는 외부 DNA를 인지하고 자를 수 있도록 도와준다.
- [298] 상기 Cas 단백질은 Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9, Cas10, Cas12a, Cas12b, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, Cas12i, Cas12j, Cas13a, Cas13b, Cas13c, Cas13d, Cas14, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3 또는 Csf4의 엔도뉴클레아제일 수 있으나, 이에 제한되는 것은 아니다.
- [299] 상기 Cas 단백질은 코리네박터(*Corynebacter*), 수테렐라(*Sutterella*), 레지오넬라(*Legionella*), 트레포네마(*Treponema*), 피리팩터(*Filifactor*), 유박테리움(*Eubacterium*), 스트렙토코커스(*Streptococcus*: *Streptococcus pyogenes*), 락토바실러스(*Lactobacillus*), 미코플라즈마(*Mycoplasma*), 박테로이드(*Bacteroides*), 플라비플라비볼라(*Flaviivola*), 플라보박테리움(*Flavobacterium*), 아조스피릴럼(*Azospirillum*), 글루코나세토박터(*Gluconacetobacter*), 나이세리아(*Neisseria*), 로세부리아(*Roseburia*), 파비바큐럼(*Parvibaculum*), 스탕필로코커스(*Staphylococcus*: *Staphylococcus aureus*), 니트라티프랙터(*Nitratifractor*), 코리네박테리움(*Corynebacterium*) 및 캄필로박터(*Campylobacter*)로 이루어지는 군으로부터 선택되는 Cas 단백질의 오소로그(ortholog)를 포함하는 미생물 속으로부터 유래하고, 이들로부터 단순 분리된 것 또는 재조합된 것일 수 있다.
- [300] 변이된 형태의 Cas 단백질을 포함할 수 있다. DNA 이중 가닥을 절단하는 엔도뉴클레아제 활성을 상실하도록 변이된 것을 의미할 수 있으며, 예컨대, 엔도뉴클레아제 활성을 상실하고 니카아제 활성을 갖도록 변이된 변이 표적 특이적 뉴클레아제 및 엔도뉴클레아제 활성과 니카아제 활성을 모두 상실하도록 변이된 형태 중 1종 이상일 수 있다.
- [301] 니카아제 활성을 갖는 것인 경우, 상기 시토신 탈아미노효소에 의한 염기변환(예컨대, 시토신이 우라딘으로 변환)과 동시 또는 순서와 무관하게 순차적으로, 상기 염기 변환이 일어난 가닥 또는 그 반대 가닥 (예컨대, 염기 변환이 일어난 가닥의 반대 가닥)에서 nick이 도입될 수 있다 (예컨대, PAM이

위치하는 가닥의 반대가닥에서, PAM 서열의 5' 말단 방향으로 3번째 뉴클레오타이드와 4번째 뉴클레오타이드 사이에 해당하는 위치에 nick이 도입됨). 이와 같은 변이 (예컨대, 아미노산 치환 등)는 촉매 활성 도메인 (예컨대, Cas9의 경우 RuvC 촉매 도메인)에서 일어나는 것일 수 있다. 스트렙토코커스 피요젠스 유래 Cas9의 경우 변이는 촉매 활성을 갖는 아스파르트산 잔기 (catalytic aspartate residue; 10번째 위치의 아스파르트산 (D10) 등), 762번째 위치의 글루탐산 (E762), 840번째 위치의 히스티딘 (H840), 854번째 위치의 아스파라긴 (N854), 863번째 위치의 아스파라긴 (N863), 986번째 위치의 아스파르트산 (D986) 등으로 이루어진 군에서 선택된 하나 이상 임의의 다른 아미노산으로 치환된 돌연변이를 포함할 수 있다. 이 때, 치환되는 임의의 다른 아미노산은 알라닌 (alanine)일 수 있지만, 이에 제한되지 않는다.

- [302] 경우에 따라서, 스트렙토코커스 피요젠스 유래 Cas9 단백질의 1135번째 위치의 아스파르트산 (D1135), 1335번째 위치의 아르기닌 (R1335), 및 1337번째 위치의 트레오닌 (T1337) 중 하나 이상, 예컨대 3개 모두가 다른 아미노산으로 치환되어, 야생형 Cas9의 PAM 서열 (NGG)와 상이한 NGA (N은 A, T, G, 및 C 중에서 선택된 임의의 염기임)를 인식하도록 변이된 것일 수 있다.
- [303] 예를 들어, 스트렙토코커스 피요젠스 유래 Cas9 단백질의 아미노산 서열 중,
- [304] (1) D10, H840, 또는 D10 + H840;
- [305] (2) D1135, R1335, T1337, 또는 D1135 + R1335 + T1337; 또는
- [306] (3) (1)과 (2) 잔기 모두에서 아미노산 치환이 일어난 것일 수 있다.
- [307] 상기 '다른 아미노산'은, 알라닌, 이소류신, 류신, 메티오닌, 페닐알라닌, 프롤린, 트립토판, 발린, 아스파라긴산, 시스테인, 글루타민, 글리신, 세린, 트레오닌, 티로신, 아스파르트산, 글루탐산, 아르기닌, 히스티딘, 라이신, 상기 아미노산들의 공지된 모든 변형체 중에서, 야생형 단백질이 원래 변이 위치에 갖는 아미노산을 제외한 아미노산들 중에서 선택된 아미노산을 의미한다. 일 예에서, 상기 '다른 아미노산'은 알라닌, 발린, 글루타민, 또는 아르기닌일 수 있다.
- [308] 경우에 따라서, 가이드 RNA를 추가로 포함할 수 있다. 상기 가이드 RNA는 예를 들어, CRISPR RNA (crRNA), trans-activating crRNA (tracrRNA), 및 단일 가이드 RNA (single guide RNA; sgRNA)로 이루어진 군에서 선택된 1종 이상일 수 있으며, 구체적으로 crRNA와 tracrRNA가 서로 결합된 이중 가닥 crRNA:tracrRNA 복합체, 또는 crRNA 또는 그 일부와 tracrRNA 또는 그 일부가 올리고뉴클레오타이드 링커로 연결된 단일 가닥 가이드 RNA (sgRNA)일 수 있다.
- [309]
- [310] 5. 아데닌 탈아미노효소의 부가
- [311]
- [312] 본 발명은 (i) DNA 결합 단백질; (ii) 시토신 탈아미노효소 (deaminase) 또는 이의

변이체; 및 (iii) 아데닌 탈아미노효소를 포함하는 융합단백질로, 상기 시토신 탈아미노효소 또는 이의 변이체는 (a) 비독성 전장 시토신 탈아미노효소 또는 (b) 시토신 탈아미노효소 또는 이의 변이체 유래인 제1분할체 및 제2분할체를 포함하고, 상기 제1분할체 및 제2분할체는 각각 상기 DNA 결합 단백질에 결합되는 형태인, 융합단백질에 관한 것이다.

- [313] 본 출원의 발명자들은 DNA에 결합할 수 있는 TALE 또는 ZFP 단백질에 DddAtox 시토신 탈아미노효소와 A to G conversion을 일으킬 수 있는 아데닌 탈아미노효소를 연결하여 A 염기를 교정 가능한 염기 교정기 (base editor)를 제작하였다.
- [314] 기존의 DddAtox를 활용한 deaminase(DdCBE)의 경우 DNA binding module로써 TALE repeat을 사용하는 시토신 탈아미노효소이다. C to T 만을 일으키는 DdCBE와는 달리 DdABE는 A to G를 일으킬 수 있기 때문에 다른 변이 패턴을 만들어 낼 수 있다.
- [315] DdABE는 그 자체로 이중 가닥 DNA를 인식하여 탈아미노화를 일으키기 때문에 RNA 와 같은 부가적인 component가 없다. mitochondria나 chloroplast같은 경우 RNA를 delivery하는 기작이 알려져 있지 않아 Crispr system을 적용할 수가 없었다. 하지만 이러한 component가 없는 DdABE는 cell내의 genomic DNA를 target할 수 있는 것뿐만 아니라, mitochondria나 chloroplast와 같은 소기관의 DNA를 target하여 특정 DNA의 A to G conversion을 일으킬 수 있다.
- [316] 현재 미토콘드리아나 세포 소기관을 타겟하는 유전자 교정 기술은 DdCBE가 유일하다. 따라서 기존의 모든 기술을 통틀어 도입할 수 있는 변이는 C to T 밖에 할 수 없었는데, DdABE는 A to G를 일으킬 수 있기 때문에 변이를 도입할 수 있는 스펙트럼이 훨씬 다양해진다. 이는 현재까지 불가능했던 미토콘드리아 질병 모델 제작 또는 치료를 가능하게 한다.
- [317] 기존의 DdCBE 같은 경우 두 개의 TALE module(각각 왼쪽과 오른쪽에 붙는)필요하기 때문에 유전자 치료에 있어서 gene capacity 낮은 바이러스 벡터인 AAV에 탑재 할 수 없다. 하지만 DdABE는 한 개의 TALE module만을 사용할 수 있는 'single module'로 사용할 수 있기 때문에 AAV 바이러스에 탑재 할 수 있고, 유전자 치료에 강점을 가진다.
- [318] DdABE는 필요에 따라서 split 된 DddAtox를 사용해서 사용 할 수도 있고, full length DddAtox variant를 이용해서 사용할 수도 있기 때문에 호환성이 높다.
- [319] 상기 아데닌 탈아미노효소는 예를 들어, APOBEC1 (apolipoprotein B editing complex 1), AID (activation-induced deaminase) 및 tadA (tRNA-specific adenosine deaminase)으로 구성된 군에서 선택될 수 있으며, 구체적으로는 tadA (tRNA-specific adenosine deaminase)일 수 있다. 상기 아데닌 탈아미노효소는 예를 들어, 대장균 TadA의 변이체로서 디옥시-아데닌 탈아미노효소 (deoxy-adenine deaminase)일 수 있다.
- [320] 상기 시토신 탈아미노효소가 분할 형태로 포함되고, 상기 DNA 결합 단백질이

징크 핑거 단백질로, 시토신 탈아미노효소의 제1분할체 C-말단에 징크 핑거 단백질 (ZF-Left)의 N-말단이 결합하고 시토신 탈아미노효소의 제2분할체 N-말단에 징크 핑거 단백질 (ZF-Right) C-말단이 결합 (NC configuration)한 구조에서, 아데닌 탈아미노효소는 징크 핑거 단백질 (ZF-Left)의 C 말단, 시토신 탈아미노효소의 제1분할체의 N 말단 또는 C 말단, 징크 핑거 단백질 (ZF-Right)의 N 말단, 시토신 탈아미노효소의 제2분할체의 N 말단 또는 C 말단에 결합할 수 있다.

- [321] 아데닌 탈아미노효소는 시토신 탈아미노효소의 제1분할체 N-말단에 징크 핑거 단백질 (ZF-Left)의 C-말단이 결합하고 시토신 탈아미노효소의 제2분할체 N-말단에 징크 핑거 단백질 (ZF-Right) C-말단이 결합 (CC configuration); 시토신 탈아미노효소의 제1분할체 N-말단에 징크 핑거 단백질 (ZF-Left)의 C-말단이 결합하고 시토신 탈아미노효소의 제2분할체 C-말단에 징크 핑거 단백질 (ZF-Right) N-말단이 결합 (CN configuration); 또는 시토신 탈아미노효소의 제1분할체 C-말단에 징크 핑거 단백질 (ZF-Left)의 N-말단이 결합하고 시토신 탈아미노효소의 제2분할체 C-말단에 징크 핑거 단백질 (ZF-Right) N-말단이 결합 (NN configuration)에서도 징크 핑거 단백질 (ZF-Left)의 C 말단, 시토신 탈아미노효소의 제1분할체의 N 말단 또는 C 말단, 징크 핑거 단백질 (ZF-Right)의 N 말단, 시토신 탈아미노효소의 제2분할체의 N 말단 또는 C 말단에 결합할 수 있다.
- [322] 상기 시토신 탈아미노효소가 분할 형태로 포함되고, 상기 DNA 결합 단백질이 TALE인 경우 상기 시토신 탈아미노효소의 제1분할체에 제1 TALE 이 결합되고, 상기 시토신 탈아미노효소의 제2분할체에 제2 TALE이 결합될 수 있으며, 각각 N'-TALE-제1분할체DDDA-C' 및 N'-TALE-제2분할체DDDA-C' 구조를 가진다. 아데닌 탈아미노효소는 시토신 탈아미노효소의 제1분할체의 N 말단 또는 C 말단 또는 시토신 탈아미노효소의 제2분할체 N 말단 또는 C 말단에 결합할 수 있다.
- [323] 상기 시토신 탈아미노효소가 전장 형태로 포함되고, 상기 DNA 결합 단백질이 TALE인 경우 단일 TALE 모듈을 포함할 수 있고, N-C 방향으로 단일 TALE 모듈 및 시토신 탈아미노효소가 포함되며, 이 때 아데닌 탈아미노효소는 단일 TALE 모듈의 C 말단 방향, 시토신 탈아미노효소의 N 말단 또는 C 말단에 결합할 수 있다.
- [324] 상기 시토신 탈아미노효소가 전장 형태로 포함되고, 상기 DNA 결합 단백질이 TALE인 경우 듀얼 TALE 모듈을 포함할 수 있고, N-C 방향으로 제1 TALE 모듈 및 시토신 탈아미노효소가 포함되고, 아데닌 탈아미노효소 및 제2 TALE 포함 제2분할체가 추가로 포함될 수 있으며, N'-TALE-시토신 탈아미노효소-C' 및 N'-TALE-아데닌 탈아미노효소-C'의 구조로 아데닌 탈아미노효소는 TALE의 N 말단 또는 C 말단에 결합할 수 있다.
- [325] 경우에 따라서, 염기 교정 효율을 올려줄 수 있는 UGI (uracil DNA glycosylase

inhibitor)를 추가로 포함할 수 있다. UGI는 DNA에서 U를 제거하는 작용을 촉매하는 돌연변이 DNA를 복구하는 효소 UDG (Uracil DNA glycosylase) 활성을 억제시켜 염기 교정 효율을 증가시킬 수 있다.

- [326] 본 발명은 상기 융합단백질 또는 이를 코딩하는 핵산을 포함하는 원핵세포 또는 진핵세포에서의 A-to-G 염기 교정용 조성물로, 상기 DNA 결합 단백질은 징크 핑거 단백질 (zinc finger protein), TALE 단백질 또는 CRISPR 연관 뉴클레아제이고, 상기 융합단백질의 시토신 탈아미노효소 또는 이의 변이체는 박테리아 유래이고, 이중 가닥 DNA 특이적인 것을 특징으로 하는 조성물에 관한 것이다.
- [327] 본 발명은 상기 융합단백질 또는 이를 코딩하는 핵산을 포함하는 원핵세포 또는 진핵세포에서의 A-to-G 염기 교정용 조성물로, 상기 DNA 결합 단백질은 징크 핑거 단백질 (zinc finger protein), TALE 단백질 또는 CRISPR 연관 뉴클레아제이고,
- [328] 상기 융합단백질의 시토신 탈아미노효소 또는 이의 변이체는 박테리아 유래이고, 이중 가닥 DNA 특이적인 것이며, 상기 시토신 탈아미노효소 또는 이의 변이체의 N 말단에 DNA 결합 단백질이 결합되어 있고, 상기 융합단백질의 아데닌 탈아미노효소의 C 말단에 DNA 결합 단백질이 결합되어 있는, 조성물에 관한 것이다.
- [329] 본 발명은 상기 융합단백질 또는 이를 코딩하는 핵산 및 UGI (uracil glycosylase inhibitor)를 포함하는 원핵세포 또는 진핵세포에서의 C-to-T 염기 교정용 조성물로, 상기 DNA 결합 단백질은 징크 핑거 단백질 (zinc finger protein), TALE 단백질 또는 CRISPR 연관 뉴클레아제이고, 상기 융합단백질의 시토신 탈아미노효소 또는 이의 변이체는 박테리아 유래이고, 이중 가닥 DNA 특이적인 것을 특징으로 하는 조성물에 관한 것이다.
- [330] 구체적으로, 본 발명은 1) DNA 결합 단백질 2) 전장 이중 가닥 DNA 특이적 박테리아 시토신 탈아미노효소 또는 이의 변이체 및 3) E. coli TadA 유래 디옥시아데닌 탈아미노효소를 포함하는 원핵세포 및 진핵 세포에서의 A-to-G 염기 교정용 조성물(UGI 없음)로, 상기 DNA 결합 단백질은 징크 핑거 단백질 (ZFP) 또는 전사 활성화 인자 유사 이펙터 (TALE) 어레이 또는 촉매 결손 CRISPR-Cas9 (nCas9 또는 dCas9) 또는 Cas12a이고; 전장 이중 가닥 DNA 특이적 박테리아 시토신 탈아미노효소 *Burkholderia cenocepacia* 유래 DddAtox인 조성물이다.
- [331] 본 발명은 1) 전장 이중 가닥 DNA 특이적 박테리아 시토신 탈아미노효소 또는 이의 변이체에 작동 가능하게 연결된 왼쪽 DNA 결합 단백질 2) E. coli TadA 유래 디옥시아데닌 탈아미노효소에 작동 가능하게 연결된 오른쪽 DNA 결합 단백질을 포함하는 원핵세포 및 진핵 세포에서의 A-to-G 염기 교정용 조성물(UGI 없음)로, 상기 왼쪽 또는 오른쪽 DNA 결합 단백질은 징크 핑거 단백질 (ZFP) 또는 전사 활성화 인자 유사 이펙터 (TALE) 어레이 또는 촉매 결손

CRISPR-Cas9 (nCas9 또는 dCas9) 또는 Cas12a이고; 전장 이중 가닥 DNA 특이적 박테리아 시토신 탈아미노효소 *Burkholderia cenocepacia* 유래 DddAtox인 조성물이다.

- [332] 본 발명은 또한, 1) DNA 결합 단백질 2) 전장 이중 가닥 DNA 특이적 박테리아 시토신 탈아미노효소 또는 이의 변이체, 3) *E. coli* TadA 유래 디옥시아데닌 탈아미노효소 및 4) UGI (uracil glycosylase inhibitor)를 포함하는 원핵세포 및 진핵 세포에서의 A-to-G 및 C-to-T 염기 교정용 조성물로, 상기 DNA 결합 단백질은 징크 핑거 단백질 (ZFP) 또는 전사 활성화 인자 유사 이펙터 (TALE) 어레이 또는 촉매 결손 CRISPR-Cas9 (nCas9 또는 dCas9) 또는 Cas12a이고; 전장 이중 가닥 DNA 특이적 박테리아 시토신 탈아미노효소 *Burkholderia cenocepacia* 유래 DddAtox인 조성물이다.
- [333] 본 발명은 더욱이, 1) DNA 결합 단백질 2) 분할된 이중 가닥 DNA 특이적 박테리아 시토신 탈아미노효소 또는 이의 변이체 및 3) *E. coli* TadA 유래 디옥시아데닌 탈아미노효소를 포함하는 원핵세포 및 진핵 세포에서의 A-to-G 염기 교정용 조성물(UGI 없음)로, 상기 DNA 결합 단백질은 징크 핑거 단백질 (ZFP) 또는 전사 활성화 인자 유사 이펙터 (TALE) 어레이 또는 촉매 결손 CRISPR-Cas9 (nCas9 또는 dCas9) 또는 Cas12a이고; 분할된 이중 가닥 DNA 특이적 박테리아 시토신 탈아미노효소 *Burkholderia cenocepacia* 유래 DddAtox인 조성물이다.
- [334] 본 발명은 더욱이, 1) DNA 결합 단백질 2) 분할된 이중 가닥 DNA 특이적 박테리아 시토신 탈아미노효소 또는 이의 변이체 3) *E. coli* TadA 유래 디옥시아데닌 탈아미노효소 및 4) UGI (uracil glycosylase inhibitor)를 포함하는 원핵세포 및 진핵 세포에서의 A-to-G 및 C-to-T 염기 교정용 조성물로, 상기 DNA 결합 단백질은 징크 핑거 단백질 (ZFP) 또는 전사 활성화 인자 유사 이펙터 (TALE) 어레이 또는 촉매 결손 CRISPR-Cas9 (nCas9 또는 dCas9) 또는 Cas12a이고; 분할된 이중 가닥 DNA 특이적 박테리아 시토신 탈아미노효소 *Burkholderia cenocepacia* 유래 DddAtox인 조성물이다.
- [335] 본 발명은 상기 융합단백질 또는 이를 코딩하는 핵산을 원핵세포 또는 진핵세포에 처리하는 단계를 포함하는 원핵세포 또는 진핵세포에서의 A-to-G 염기 교정 방법으로, 상기 DNA 결합 단백질은 징크 핑거 단백질 (zinc finger protein), TALE 단백질 또는 CRISPR 연관 뉴클레아제이고, 상기 융합단백질의 시토신 탈아미노효소 또는 이의 변이체는 박테리아 유래이고, 이중 가닥 DNA 특이적인 것을 특징으로 하는 방법이다.
- [336] 본 발명은 상기 융합단백질 또는 이를 코딩하는 핵산을 원핵세포 또는 진핵세포에 처리하는 단계를 포함하는 원핵세포 또는 진핵세포에서의 A-to-G 염기 교정 방법으로, 상기 DNA 결합 단백질은 징크 핑거 단백질 (zinc finger protein), TALE 단백질 또는 CRISPR 연관 뉴클레아제이고, 상기 융합단백질의 시토신 탈아미노효소 또는 이의 변이체는 박테리아 유래이고, 이중 가닥 DNA

특이적인 것이며, 상기 시토신 탈아미노효소 또는 이의 변이체의 N 말단에 DNA 결합 단백질이 결합되어 있고, 상기 융합단백질의 아데닌 탈아미노효소의 C 말단에 DNA 결합 단백질이 결합되어 있는, 방법이다.

- [337] 본 발명은 상기 융합단백질 또는 이를 코딩하는 핵산 및 UGI (uracil glycosylase inhibitor)를 원핵세포 또는 진핵세포에 처리하는 단계를 포함하는 원핵세포 또는 진핵세포에서의 C-to-T 염기 교정 방법으로, 상기 DNA 결합 단백질은 징크 핑거 단백질 (zinc finger protein), TALE 단백질 또는 CRISPR 연관 뉴클레아제이고, 상기 융합단백질의 시토신 탈아미노효소 또는 이의 변이체는 박테리아 유래이고, 이중 가닥 DNA 특이적인 것을 특징으로 하는 방법이다.
- [338] 본 발명에 따른 조성물 또는 방법에 포함된 구성의 구체적 서열은 다음과 같다.
- [339] **ND1-ZFP-Right-1397C-AD (도.61f-g; SEQ ID No: 410)**
- [340] MLGFVGRVAAAPASGALRRLTPSASLPPAQLLLRAAPTAVHPVRDYAAQDY
KDDDDKVDEMTEKKFGTLTIHDTEKAAEFGIRIPGEKPFQCRICMRNFSDSGNLR
VHIRTHTGEKPYKCPDCGKSFSQSSSLIRHQRTHHTGEKPYECDHCGKSFSQSSH
LNVHKRTHHTGEKPYRCKYCDRSFSSNLQRHVRNIHLRSGTPHEVG VYTLTG
TPHEVG VYTLAIPVKGATGETKVFTGNSNSPKSPTKGGCSGSETPGTSESATP
ESSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGL
HDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVV
FGVRNSKRGAAAGSLMNVLNYPGMNHRVEITEGILADECAALLCDFYRMPRQV
FNAQKKAQSSIN
- [341]
- [342] **ND1-ZFP-Left-1397C-UGI (도.61f-g; SEQ ID No: 411)**
- [343] MLGFVGRVAAAPASGALRRLTPSASLPPAQLLLRAAPTAVHPVRDYAAQDY
KDDDDKVDEMTEKKFGTLTIHDTEKAAEFGIRIPGEKPFQCRICMRNFSDSGNLR
VHIRTHTGEKPYKCPDCGKSFSQSSSLIRHQRTHHTGEKPYECDHCGKSFSQSSH
LNVHKRTHHTGEKPYRCKYCDRSFSSNLQRHVRNIHLRSGTPHEVG VYTLTG
TPHEVG VYTLAIPVKGATGETKVFTGNSNSPKSPTKGGCSGGSTNLSDIIEKE
TGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEY
KPWALVIQDSNGENKIKML
- [344]
- [345] **ND1-ZFP-Right-1397N-UGI (도.61f-g; SEQ ID No: 412)**
- [346] MLGFVGRVAAAPASGALRRLTPSASLPPAQLLLRAAPTAVHPVRDYAAQYP
YDVPDYAVDEMTEKKFGTLTIHDTEKAAEFGIHGVPAAMGGSYALGPYQISAP
QLPAYNGQTVGTFYVNDAGGLESKVFSSGGPTYPNYANAGHVEGQSALFM
RDNGISEGLVFHNNPEGTCGFCVNMETLLPENAKMTVVPPEGSGTPHEVG VY
TLTGTPHEVG VYTLTKCPECCKSFSTKNSLTEHQRTHHTGEKPYKCECKGSFSS
KKALTEHQRTHHTGEKPYECNYCGKTFSVSSTLIRHQRHTGEKPYRCKYCDRS
FSSSNLQRHVRNIHLRSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNK

PESDILVHTAYDESTDENVMMLTSDAPEYKPWALVIQDSNGENKIKML

[347]

[348] **Trac-ZFP-LEFT-1397N-UGI (도.61d: SEQ ID No: 413)**

[349] MAPKKKRKVGIIHGVPAAAMGGSYALGPYQISAPQLPAYNGQTVGTFYYVND
AGGLESKVFSGGPTYPNYANAGHVEGQSALFMRDNGISEGLVFHNNPEGTC
GFCVNMETETLLPENAKMTVVPPEGSGTPHEVG VYTLSGTPHEVG VYTLFQCRI
CMRKFATSGSLTRHTKIHTGEKPFQCRICMRNFSRSDHLSTHIRTHTGEKPFAC
DICGRKFATSSNRTKHTKIHTHPRAPIKPFQCRICMRNFSRSDNLSHIRTHTG
EKPFACDICGRKFAWHSSLRVHTKIHLRSGGSTNLSHIEKETGKQLVIESILM
LPEEVEEVIGNKPESDILVHTAYDESTDENVMMLTSDAPEYKPWALVIQDSNGE
NKIKML

[350]

[351] **Trac-ZFP-right-1397C-AD (도.61d: SEQ ID No: 414)**

[352] MAPKKKRKVGIIHGVPAAAMAERPFQCRICMRNFSRSDHLSTHIRTHTGEKPFAC
CDICGRKFADRSHLARHTKIHTGSQKPFQCRICMRKFALKQHLNEHTKIHTGE
KPFQCRICMRNFSQSGNLARHIRTHTGEKPFACDICGRKFAHNSSLKDHTKIHL
RSGTPHEVG VYTLSGTPHEVG VYTLAIPVKRGATGETKVFTGNSNSPKSPTKG
GCSGSETPGTSESATPESSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVL
NNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCV
MCAGAMIHRSRIGRVVFGVRNSKRGAAAGSLMNVLNYPGMNHRVEITEGILADE
CAALLCDFYRMPRQVFNAQKKAQSSIN

[353]

[354] **ND1-Left-TALE-1397N-UGI (도 62: SEQ ID No: 415)**

[355] MALSRVCGTSRQLAPVLGYLGSRQKHSLPDYPYDVPDYAGYPYDVPDYA
GYPYDVPDYAGIRIQDLRTLGYSSQQQEKIKPKVRSTVAQHHEALVGHGFTH
AHIVALSQHPAALGTAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALL
TVAGELRGPPQLDGTGQLLKIAGRGGVTAVEAVHAWRNALTGAPLNLTDPQV
VAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQR
LLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAI
ASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLP
VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIAS
NNGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPV
CQAHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHD
GGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQ
AHGLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNNGG
KQALETVQRLLPVLCQAHGLTPAQVVAIASNNGGKQALETVQRLLPVLCQAH
GLTPEQVVAIASNNGGKQALESIVAQLSRPDALAALLVSGGSYALGPYQISAP
QLPAYNGQTVGTFYYVNDAGGLESKVFSGGPTYPNYANAGHVEGQSALFM

RDNGISEGLVFHNNPEGTCGFCVNMETLLPENAKMTVVPPEGSGGSTNLSDI
 EKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDA
 PEYKPWALVIQDSNGENKIKML

[356] **ND1-Left-TALE-1397C-UGI (㉠ 62: SEQ ID No: 416)**

[357] MALSRVCGTSRQLAPVLGYLGSRQKHSLPDYPYDVPDYAGYPYDVPDYA
 GYPYDVPDYAGIRIQDLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTH
 AHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALL
 TVAGELRGPPQLDGTGQLLKIAGRGGVTAVEAVHAWRNALTGAPLNLTDPQV
 VAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQR
 LLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAI
 ASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLP
 VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIAS
 NNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLP
 VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIAS
 NNGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPV
 LCQAHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHD
 GGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQ
 AHGLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNNGG
 KQALETVQRLLPVLCQAHGLTPAQVVAIASNNGGKQALETVQRLLPVLCQAH
 GLTPEQVVAIASNNGGKQALESIVAQLSRPDALAALLVGSAPVKRGATGET
 KVFTGNSNSPKSPTKGGCSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIG
 NKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML

[358] **ND1-Right-TALE-1397N-UGI (㉠ 62: SEQ ID No: 417)**

[359] MASVLTPLLLRGLTGSARRLPVPRAKIHSLDYKDHDGDYKDHDIDYKDDDD
 KGIRIQDLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHP
 AALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPP
 LQLDGTGQLLKIAGRGGVTAVEAVHAWRNALTGAPLNLTDPQVVAIASNNGG
 KQALETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQAH
 GLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQ
 ALETVQRLLPVLCQAHGLTPAQVVAIASNNGGKQALETVQRLLPVLCQAHGL
 TPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNNGGKQAL
 ETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPE
 QVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETV
 QRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQV
 VAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQR
 LLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAI
 ASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLP
 VLCQDHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIAS
 NNGGKQALESIVAQLSRPDALAALLVGSYSALGPYQISAPQLPAYNGQTVG
 TFYYVNDAGGLESKFSSGGPTYPNYANAGHVEGQSALFMRDNGISEGLVF

HNNPEGTCGFCVNMETETLLPENAKMTVVPPEGSGGSTNLSDIIEKETGKQLVIQ
 ESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQ
 DSNGENKIKML

[360] **ND1-Right-TALE-1397C-UGI (도 62: SEQ ID No: 418)**

[361] MASVLTPLLLRGLTGSARRLPVPRAKIHSLDYKDHDGDYKDHDIDYKDDDD
 KGIRIQDLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHP
 AALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPP
 LQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTDPQVVAIASNNGG
 KQALETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQAH
 GLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQ
 ALETVQRLLPVLCQAHGLTPAQVVAIASNNGGKQALETVQRLLPVLCQAHGL
 TPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNNGGKQAL
 ETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPE
 QVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETV
 QRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQV
 VAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQR
 LLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAI
 ASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLP
 VLCQDHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIAS
 NNGGKQALESIVAQLSRPDALAALLVGSAPVKGATGETKVFTGNSNSPKSP
 TKGGCSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAY
 DESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML

[362]

[363] **ND4-Left-TALE-AD (도. 62b: SEQ ID No: 419)**

[364] MALSRAVCGTSRQLAPVLGYLGSRQKHSLPDYPYDVPDYAGYPYDVPDYA
 GYPYDVPDYAMDIADLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTHA
 HIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLT
 VAGELRGPPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPEQVV
 AIASNNGGKQALETVQALLPVLCQAHGLTPQQVVAIASHDGGKQALETVQRL
 LPVLCQAHGLTPQQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAI
 ASNIGGKQALETVQALLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLP
 VLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIAS
 NIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLC
 QAHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDG
 GKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
 HGLTPEQVVAIASHDGGKQALETVQALLPVLCQAHGLTPEQVVAIASNNGGK
 QALETVQRLLPVLCQAHGLTPQQVVAIASNNGGKQALETVQRLLPVLCQAHG
 LTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASHDGGKQA

LETVQALLPVLCQAHGLTPQQVVAIASNGGGRPALESIVAQLSRPDPALAAALT
NDHLVALACLGGRPALDAVKKGLGGSSGSETPGTSESATPESSEVEFSHEYWM
RHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMALR
QGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNSKRGAAGS
LMNVLNYPGMNHRVEITEGILADECAALLCDFYRMPRQVFNAQKKAQSSIN

[365] **ND4-Right-TALE-AD (도 62b: SEQ ID No: 420)**

[366] MASVLTPLLLRGLTGSARRLPVPRAKIHSLDYKDHDGDYKDHDIDYKDDDD
KMDIADLRTLGYSSQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHP
AALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPP
LQLDTGQLLAKRGGVTAVEAVHAWRNALTGAPLNLTPEQVVAIASHDGGK
QALETVQALLPVLCQAHGLTPQQVVAIASHDGGKQALETVQRLLPVLCQAHG
LTPQQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNNGGKQA
LETVQALLPVLCQAHGLTPEQVVAIASNGGGKQALETVQALLPVLCQAHGLT
PEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALET
VQRLLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQ
VVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQR
LLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAI
ASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLP
VLCQAHGLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPQQVVAIAS
NIGGKQALETVQRLLPVLCQAHGLTPQQVVAIASNNGGKQALETVQRLLPV
CQAHGLTPQQVVAIASNGGGRPALESIVAQLSRPDPALAAALTNDHLVALACL
GGRPALDAVKKGLGGSSGSETPGTSESATPESSEVEFSHEYWMRHALTLAKR
ARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMALRQGGGLVMQNYR
LIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNSKRGAAGSLMNVLNYPGM
NHRVEITEGILADECAALLCDFYRMPRQVFNAQKKAQSSIN

[367]

[368] **ND1-Left-TALE-1333N-AD (도 63: SEQ ID No: 421)**

[369] MALSRVCGTSRQLAPVLGYLGSRQKHSLPDYPYDVPDYAGYPYDVPDYA
GYPYDVPDYAGIRIQDLRTLGYSSQQQEKIKPKVRSTVAQHHEALVGHGFTH
AHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALL
TVAGELRGPPQLDTGQLLAKRGGVTAVEAVHAWRNALTGAPLNLTDPQV
VAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQR
LLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAI
ASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLP
VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIAS
NGGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPV
CQAHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHD
GGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQ

AHGLTPEQVVAIASNNGGKQALETQVQRLLPVLCQAHGLTPAQVVAIASNNGG
 KQALETQVQRLLPVLCQAHGLTPAQVVAIASNNGGKQALETQVQRLLPVLCQAH
 GLTPEQVVAIASNNGGKQALESIVAQLSRPDPALAALLVSGSGSYALGPYQISAP
 QLPAYNGQTVGTFYVNDAGGLESKVFSGSGSETPGTSESATPESSEVEFSH
 EYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEI
 MALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNSKRG
 AAGSLMNVLNYPGMNHRVEITEGILADECAALLCDFYRMPRQVFNAQKKAQS
 SIN

[370]

[371] **ND1-Left-TALE-1333C-AD (㉮ 63: SEQ ID No: 422)**

[372] MALSRAVCGTSRQLAPVLGYLGSRQKHSLPDYPYDVPDYAGYPYDVPDYA
 GYPYDVPDYAGIRIQDLRTLGYSSQQQEKIKPKVRSTVAQHHEALVGHGFTH
 AHIVALSQHPAALGTAVVKYQDMIAALPEATHEAIVGVGKQWSGARALEALL
 TVAGELRGPPQLDGTGQLLKIAGRGGVTAVEAVHAWRNALTGAPLNLTDPQV
 VAIASHDGGKQALETQVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETQVQ
 RLLPVLCQDHGLTPAQVVAIASNIGGKQALETQVQRLLPVLCQDHGLTPDQVVAI
 ASNNGGKQALETQVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETQVQRLLP
 VLCQDHGLTPDQVVAIASHDGGKQALETQVQRLLPVLCQAHGLTPAQVVAIAS
 NNGGKQALETQVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETQVQRLLPVL
 CQAHGLTPDQVVAIASNNGGKQALETQVQRLLPVLCQAHGLTPAQVVAIASHD
 GGKQALETQVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETQVQRLLPVLCQ
 AHGLTPEQVVAIASNNGGKQALETQVQRLLPVLCQAHGLTPAQVVAIASNNGG
 KQALETQVQRLLPVLCQAHGLTPAQVVAIASNNGGKQALETQVQRLLPVLCQAH
 GLTPEQVVAIASNNGGKQALESIVAQLSRPDPALAALLVGSPTYPNYANAGH
 VEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMETLLPENAKMTVVPPEG
 AIPVKRGATGETKVFTGNSNSPKSPTKGGCSGSETPGTSESATPESSEVEFSHEY
 WMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIM
 ALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNSKRG
 AAGSLMNVLNYPGMNHRVEITEGILADECAALLCDFYRMPRQVFNAQKKAQSS
 IN

[373]

[374] **ND1-Right-TALE-1333N-AD (㉮ 63: SEQ ID No: 423)**

[375] RQVFNAQKKAQSSIN

[376]

[377] **ND1-Right-TALE-1333C-AD (㉮ 63: SEQ ID No: 424)**

[378] VFGVRNSKRGAAAGSLMNVLNYPGMNHRVEITEGILADECAALLCDFYRMPR
 QVFNAQKKAQSSIN

[379]

[380] **ND1-Left-TALE-1397C-AD (㉠ 62-63: SEQ ID No: 425)**

[381] MALSRVCGTSRQLAPVLGYLGSRQKHSLPDYPYDVPDYAGYPYDVPDYA
 GYPYDVPDYAGIRIQDLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTH
 AHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALL
 TVAGELRGPPQLDGTGQLLKIAGRGGVTAVEAVHAWRNALTGAPLNLTDPQV
 VAIASHDGGKQALETQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETQRL
 LLPVLCQDHGLTPAQVVAIASNIGGKQALETQRLLPVLCQDHGLTPDQVVAI
 ASNNGGKQALETQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETQRLLP
 VLCQDHGLTPDQVVAIASHDGGKQALETQRLLPVLCQAHGLTPAQVVAIAS
 NNGGKQALETQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETQRLLPVL
 CQAHGLTPDQVVAIASNNGGKQALETQRLLPVLCQAHGLTPAQVVAIASHD
 GGGKQALETQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETQRLLPVLCQ
 AHGLTPEQVVAIASNNGGKQALETQRLLPVLCQAHGLTPAQVVAIASNGGG
 KQALETQRLLPVLCQAHGLTPAQVVAIASNGGGKQALETQRLLPVLCQAH
 GLTPEQVVAIASNGGGKQALESIVAQLSRPDPALAALTNDHLVALACLGGRPA
 LDAVKKGLLVGSAIPVKRGATGETKVFTGNSNSPKSPTKGGCSGSETPGTSESA
 TPESSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNNRVIGEGWNRAI
 GLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHRSRIGR
 VVFGVRNSKRGAAGSLMNVLNYPGMNHRVEITEGILADECAALLCDFYRMPR
 QVFNAQKKAQSSIN

[382]

[383] **ND1-Right-TALE-1397C-AD (㉠ 62-63: SEQ ID No: 426)**

[384] AALLCDFYRMPRQVFNAQKKAQSSIN

[385]

[386] **ND1-Left-TALE-1333N (㉠ 63: SEQ ID No: 427)**

[387] MALSRVCGTSRQLAPVLGYLGSRQKHSLPDYPYDVPDYAGYPYDVPDYA
 GYPYDVPDYAGIRIQDLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTH
 AHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALL
 TVAGELRGPPQLDGTGQLLKIAGRGGVTAVEAVHAWRNALTGAPLNLTDPQV
 VAIASHDGGKQALETQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETQRL
 LLPVLCQDHGLTPAQVVAIASNIGGKQALETQRLLPVLCQDHGLTPDQVVAI
 ASNNGGKQALETQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETQRLLP
 VLCQDHGLTPDQVVAIASHDGGKQALETQRLLPVLCQAHGLTPAQVVAIAS
 NNGGKQALETQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETQRLLPVL
 CQAHGLTPDQVVAIASNNGGKQALETQRLLPVLCQAHGLTPAQVVAIASHD
 GGGKQALETQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETQRLLPVLCQ
 AHGLTPEQVVAIASNNGGKQALETQRLLPVLCQAHGLTPAQVVAIASNGGG
 KQALETQRLLPVLCQAHGLTPAQVVAIASNGGGKQALETQRLLPVLCQAH

GLTPEQVVAIASNNGGKQALESIVAQLSRPDPALAALTNDHLVALACLGGRPA
LDAVKKGLGGSGSGSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKV
FSSGG

[388]

[389] **ND1-Left-TALE-1333C (도 63: SEQ ID No: 428)**

[390] MALSRAVCGTSRQLAPVLGYLGSRQKHSLPDYPYDVPDYAGYPYDVPDYA
GYPYDVPDYAGIRIQDLRTLGYSSQQQKEKIKPKVRSSTVAQHHEALVGHGFTH
AHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALL
TVAGELRGPPLQLDTGQLLKIARGGVTAVEAVHAWRNALTGAPLNLTDPQV
VAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQR
LLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAI
ASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLP
VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIAS
NNGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPV
CQAHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHD
GGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQ
AHGLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNNGG
KQALETVQRLLPVLCQAHGLTPAQVVAIASNNGGKQALETVQRLLPVLCQAH
GLTPEQVVAIASNNGGKQALESIVAQLSRPDPALAALTNDHLVALACLGGRPA
LDAVKKGLGGSGSPTYPNYANAGHVEGQSALFMRDNGISEGLVFHNNPEGT
CGFCVNMTETLLPENAKMTVVPPEGAIPVKRGATGETKVFTGNSNSPKSPTKG
GC

[391]

[392] **ND1-Left-TALE-1397N (도 62-63: SEQ ID No: 429)**

[393] MALSRAVCGTSRQLAPVLGYLGSRQKHSLPDYPYDVPDYAGYPYDVPDYA
GYPYDVPDYAGIRIQDLRTLGYSSQQQKEKIKPKVRSSTVAQHHEALVGHGFTH
AHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALL
TVAGELRGPPLQLDTGQLLKIARGGVTAVEAVHAWRNALTGAPLNLTDPQV
VAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQR
LLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAI
ASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLP
VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIAS
NNGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPV
CQAHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHD
GGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQ
AHGLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNNGG
KQALETVQRLLPVLCQAHGLTPAQVVAIASNNGGKQALETVQRLLPVLCQAH
GLTPEQVVAIASNNGGKQALESIVAQLSRPDPALAALTNDHLVALACLGGRPA

LDVKKGLGGSGSGSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKV
FSSGGPTYPNYANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMT
ETLLPENAKMTVVPPEG

[394]

[395] **ND1-Right-TALE-1333N (도 63: SEQ ID No: 430)**

[396] MASVLTPLLLRGLTGSARRLPVPRAKIHSLDYKDHDGDYKDHDIDYKDDDD
KGIRIQDLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHP
AALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPP
LQLDTGQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTDPQVVAIASNNGG
KQALETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQAH
GLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQ
ALETVQRLLPVLCQAHGLTPAQVVAIASNNGGKQALETVQRLLPVLCQAHGL
TPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNNGGKQAL
ETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPE
QVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETV
QRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQV
VAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQR
LLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAI
ASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLP
VLCQDHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIAS
NNGGKQALESIVAQLSRPDALAALLVSGSGSYALGPYQISAPQLPAYNGQTVG
TFYYVNDAGGLESKVFSGG

[397]

[398] **ND1-Right-TALE-1333C (도 63: SEQ ID No: 431)**

[399] MASVLTPLLLRGLTGSARRLPVPRAKIHSLDYKDHDGDYKDHDIDYKDDDD
KGIRIQDLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHP
AALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPP
LQLDTGQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTDPQVVAIASNNGG
KQALETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQAH
GLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQ
ALETVQRLLPVLCQAHGLTPAQVVAIASNNGGKQALETVQRLLPVLCQAHGL
TPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNNGGKQAL
ETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPE
QVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETV
QRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQV
VAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQR
LLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAI
ASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLP

VLCQDHGLTPAQVVAIASHDGGKQALETQVQRLLPVLCQAHGLTPEQVVAIAS
 NGGGKQALESIVAQLSRPDPALAALLVGSPTYPNYANAGHVEGQSALFMRD
 NGISEGLVFHNNPEGTCGFCVNMETLLPENAKMTVVPPEGAIPVKRGATGET
 KVFTGNSNSPKSPTKGGC

[400]

[401] **ND1-Right-TALE-1397N (도 62-63: SEQ ID No: 432)**

[402] MASVLTPLLLRGLTGSARRLPVPRAKIHSLDYKDHDGDYKDHDIDYKDDDD
 KGIRIQDLRTLGYSSQQQKEKIKPKVRSSTVAQHHEALVGHGFTHAHIVALSQHP
 AALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPP
 LQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTDPQVVAIASNNGG
 KQALETQVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETQVQRLLPVLCQAH
 GLTPEQVVAIASNNGGKQALETQVQRLLPVLCQAHGLTPDQVVAIASNNGGKQ
 ALETQVQRLLPVLCQAHGLTPAQVVAIASNNGGKQALETQVQRLLPVLCQAHGL
 TPEQVVAIASNNGGKQALETQVQRLLPVLCQAHGLTPAQVVAIASNNGGKQAL
 ETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETQVQRLLPVLCQAHGLTPE
 QVVAIASNNGGKQALETQVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETQ
 QRLLPVLCQAHGLTPAQVVAIASHDGGKQALETQVQRLLPVLCQDHGLTPDQV
 VAIASNNGGKQALETQVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETQVQ
 RLPVLCQAHGLTPDQVVAIASNIGGKQALETQVQRLLPVLCQAHGLTPAQVVAI
 ASHDGGKQALETQVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETQVQRLP
 VLCQDHGLTPAQVVAIASHDGGKQALETQVQRLLPVLCQAHGLTPEQVVAIAS
 NGGGKQALESIVAQLSRPDPALAALLVGSYSYALGPYQISAPQLPAYNGQTVG
 TFYYVNDAGGLESKFSSGGPTYPNYANAGHVEGQSALFMRDNGISEGLVF
 HNNPEGTCGFCVNMETLLPENAKMTVVPPEG

[403]

[404] **ND4-LEFT-TALE-1333N-AD (도 63: SEQ ID No: 433)**

[405] NAQKKAQSSIN

[406]

[407] **ND4-LEFT-TALE-1333C-AD (도 63: SEQ ID No: 434)**

[408] RNSKRGAAAGSLMNVNLNYPGMNHRVEITEGILADECAALLCDFYRMPRQVFN
 AQKKAQSSIN

[409]

[410] **ND4-LEFT-TALE-1397C-AD (도 63: SEQ ID No: 435)**

[411] MALSRVCGTSRQLAPVLGYLGSRQKHSLPDYPYDVPDYAGYPYDVPDYA
 GYPYDVPDYAMDIADLRTLGYSSQQQKEKIKPKVRSSTVAQHHEALVGHGFTHA
 HIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLT
 VAGELRGPPQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPEQVV
 AIASNNGGKQALETQVQALLPVLCQAHGLTPQQVVAIASHDGGKQALETQVQRL

LPVLCQAHGLTPQQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAI
 ASNIGGKQALETVQALLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLP
 VLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIAS
 NIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLC
 QAHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDG
 GKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
 HGLTPEQVVAIASHDGGKQALETVQALLPVLCQAHGLTPEQVVAIASNNGGK
 QALETVQRLLPVLCQAHGLTPQQVVAIASNNGGKQALETVQRLLPVLCQAHG
 LTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASHDGGKQA
 LETVQALLPVLCQAHGLTPQQVVAIASNNGGGRPALESIVAQLSRPDPALAAALT
 NDHLVALACLGGRPALDAVKKGLGGSGSAIPVKGATGETKVFTGNSNSPKS
 PTKGGCSGSETPGTSESATPESSEVEFSHEYWMRHALTLAKRARDEREVPVGA
 VLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTF
 EPCVMCAGAMIHSRIGRVVFGVRNSKRGAAAGSLMNVLNYPGMNHRVEITEGI
 LADECAALLCDFYRMPRQVFNAQKKAQSSIN

[412]

[413] **ND4-Right-TALE-1333N-AD (㉞ 63: SEQ ID No: 436)**

[414] SIN

[415]

[416] **ND4-Right-TALE-1333C-AD (㉞ 63: SEQ ID No: 437)**[417] GSLMNVLNYPGMNHRVEITEGILADECAALLCDFYRMPRQVFNAQKKAQSS
IN

[418]

[419] **ND4-Right-TALE-1397C-AD (㉞ 63: SEQ ID No: 438)**[420] MASVLTPLLLRGLTGSARRLPVPRAKIHSLDYKDHDGDYKDHDIDYKDDDD
 KMDIADLRTLGYSSQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHP
 AALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPP
 LQLDTGQLLKIARGGVTAVEAVHAWRNALTGAPLNLTPEQVVAIASHDGGK
 QALETVQALLPVLCQAHGLTPQQVVAIASHDGGKQALETVQRLLPVLCQAHG
 LTPQQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNNGGKQA
 LETVQALLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLT
 PEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALET
 VQRLLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQ
 VVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQR
 LLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAI
 ASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLP
 VLCQAHGLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPQQVVAIAS
 NIGGKQALETVQRLLPVLCQAHGLTPQQVVAIASNNGGKQALETVQRLLPVLC

CQAHGLTPQQVVAIASNNGGSRPALESIVAQLSRPDPALAAALTNDHLVALACLG
 GRPALDAVKKGLGGSGSAIPVKRGATGETKVFTGNSNSPKSPTKGGCSGSETP
 GTSESATPESSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEG
 WNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIH
 SRIGRVVFGVRNSKRGAAGSLMNVNLNYPGMNHRVEITEGILADECAALLCDF
 YRMPRQVFNAQKKAQSSIN

[421]

[422] **ND1-Left-TALE-AD-GSVG (도 64: SEQ ID No: 439)**

[423] VFTGNSNGPKSPTKGGC

[424] **ND1-Left-TALE-AD-E1347A (도 64: SEQ ID No: 440)**

[425] VFTGNSNSPKSPTKGGC

[426] **ND1-Left-TALE-AD-AAAAA (도 64: SEQ ID No: 441)**

[427] VFTGNSNSPASPTAGGC

[428] **ND1-Right-TALE-AD (도 64: SEQ ID No: 442)**

[429] MASVLTPLLLRGLTGSARRLPVPRAKIHSLDYKDHDGDYKDHDIDYKDDDD
 KGIRIQDLRTLGYSSQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHP
 AALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPP
 LQLDTGQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTDPQVVAIASNNGG
 KQALETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQAH
 GLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQ
 ALETVQRLLPVLCQAHGLTPAQVVAIASNNGGKQALETVQRLLPVLCQAHGL
 TPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNNGGKQAL
 ETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPE
 QVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETV
 QRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQV
 VAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQR
 LLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAI
 ASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLP
 VLCQDHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIAS
 NNGGKQALESIVAQLSRPDPALAAALLVSGSETPGTSESATPESSEVEFSHEYWM
 RHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMALR
 QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNSKRGAAGS
 LMNVNLNYPGMNHRVEITEGILADECAALLCDFYRMPRQVFNAQKKAQSSIN

[430] **ND1-Left-TALE-GSVG (도 64: SEQ ID No: 443)**

[431] MALSRAVCGTSRQLAPVLGYLGSRQKHSLPDYPYDVPDYAGYPYDVPDYA
 GYPYDVPDYAGIRIQDLRTLGYSSQQQEKIKPKVRSTVAQHHEALVGHGFTH
 AHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALL
 TVAGELRGPPPLQLDTGQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTDPQV

V
A I A S H D G G K Q A L E T V Q R L L P V L C Q A H G L T P D Q V V A I A S N G G G K Q A L E T V Q R
L L P V L C Q D H G L T P A Q V V A I A S N I G G K Q A L E T V Q R L L P V L C Q D H G L T P D Q V V A I
A S N N G G K Q A L E T V Q R L L P V L C Q A H G L T P D Q V V A I A S H D G G K Q A L E T V Q R L L P
V L C Q D H G L T P D Q V V A I A S H D G G K Q A L E T V Q R L L P V L C Q A H G L T P A Q V V A I A S
N G G G K Q A L E T V Q R L L P V L C Q D H G L T P A Q V V A I A S N I G G K Q A L E T V Q R L L P V L
C Q A H G L T P D Q V V A I A S N N G G K Q A L E T V Q R L L P V L C Q A H G L T P A Q V V A I A S H D
G G K Q A L E T V Q R L L P V L C Q D H G L T P D Q V V A I A S H D G G K Q A L E T V Q R L L P V L C Q
A H G L T P E Q V V A I A S N N G G K Q A L E T V Q R L L P V L C Q A H G L T P A Q V V A I A S N G G G
K Q A L E T V Q R L L P V L C Q A H G L T P A Q V V A I A S N G G G K Q A L E T V Q R L L P V L C Q A H
G L T P E Q V V A I A S N G G G K Q A L E S I V A Q L S R P D P A L A A L L V G S G S Y A L G P Y Q I S A P
Q L P A Y N G Q T V G T F Y Y V N D A G G L E G K V F S S G G P T P Y P N Y A N A G H V E S Q S A L F M
R D N G I S E G L V F H N N P E G T C G F C V N M T E T L L P E N A K M T V V P P E G V I P V K R G A T G
E T K V F T G N S N G P K S P T K G G C

[432] ND1-Left-TALE-E1347A (도 64: SEO ID No: 444)

[433] MALSRVCGTSRQLAPVLGYSRQKHSLPDYDVPDYAGYPYDVPDYA
GYPYDVPDYAGIRIQDLRTLGYSSQQQEKIKPKVRSVTAQHHHEALVGHGFTH
AHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALL
TVAGELRGPPQLQDGTGQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTDPQV
VAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQR
LLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAI
ASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLP
VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIAS
NGGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLC
QAHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHD
GGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQ
AHGLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGG
KQALETVQRLLPVLCQAHGLTPAQVVAIASNGGGKQALETVQRLLPVLCQAH
GLTPEQVVAIASNGGGKQALESIVAQLSRPDPALAALLVGSGSYALGPYQISAP
QLPAYNGQTVGTFYYVNDAGGLESKVFSGGPTPYPNYANAGHVAGQSALF
MRDNGISEGLVFHNNPEGTCGFCVNMTELLPENAKMTVVPPEGAIPVKRGAT
GETKVFTGNSNSPKSPTKGGC

[434] **ND1-Left-TALE-AAAAAA (㉔ 64 SEQ ID No: 445)**

[435] MALSRAVCGTSRQLAPVLGYLGSRQKHS�PDYPYDVPDYAGYPYDVPDYA
 GYPYDVPDYAGIRIQDLRTLGYSSQQQKEIKPKVRSTVAQHHEALVGHGFTH
 AHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALL
 TVAGELRGPPQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTDPQV
 VAIASHDGGKQALETVQRLLPVLCPAHGLTPDQVVAIASNGGGKQALETVQR
 LLPVLCODHGLTPAOVVAIASNIGGKQALETVORLLPVLCPDHGLTPDOVVAI

ASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLP
 VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIAS
 NNGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVL
 CQAHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHD
 GGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQ
 AHGLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNNGG
 KQALETVQRLLPVLCQAHGLTPAQVVAIASNNGGKQALETVQRLLPVLCQAH
 GLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNNGGKQ
 ALESIVAQLSRPDPALAALLVSGSYALGPYQISAP
 QLPAYNGQTVGTFYVNDAGGLESKVFSGGPTYPNYANAGHVEGQSALFM
 RDNGISEGLVFHNNPEGTCGFCVNMETLLPENAKMTVVPPEGAIPVAAGATG
 ETAVFTGNSNSPASPTAGGC

[436]

[437]

[438] **ND1-TALE-left (SEQ ID No: 446)**

[439] DLRTLGYSSQQQEKIKPKVVRSTVAQHHEALVGHGFTHAHIVALSQHPAALG
 TVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLD
 TGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTDPQVVAIASHDGGKQAL
 ETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTP
 AQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALET
 VQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQ
 VVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNNGGKQALETVQ
 RLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVA
 IASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLL
 PVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIAS
 NNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNNGGKQALETVQRLLPVL
 CQAHGLTPAQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNG
 GGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQRLLPVL

[440]

[441] **ND1-TALE-Right (SEQ ID No: 447)**

[442] DLRTLGYSSQQQEKIKPKVVRSTVAQHHEALVGHGFTHAHIVALSQHPAALG
 TVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLD
 TGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTDPQVVAIASNNGGKQAL
 ETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPE
 QVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETV
 QRLLPVLCQAHGLTPAQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQV
 VVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNNGGKQALETVQR
 LLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAI
 ASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQRLLP

[448] MDIADLRTLGYSSQQQKEIKPKVRSSTVAQHHEALVGHGFTHAHIVALSQHP
AALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPP
LQLDTGQLLKIAGRGGVTAVEAVHAWRNALTGAPLNLTPEQVVAIASHDGGK
QALETVQALLPVLCQAHGLTPQQVVAIASHDGGKQALETVQRLLPVLCQAHG
LTPQQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNNGGKQA
LETVQALLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLT
PEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALET
VQRLLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQ
VVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQR
LLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAI
ASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLP
VLCQAHGLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPQQVVAIAS
NIGGKQALETVORLLPVLCQAHGLTPOOVVAIASNNGGKQALETVORLLPVLC

CQAHGLTPQQVVAIASNGGGRPALESIVAQLSRPDPALAAALTNDHLVALACLG
GRPALDAVKKGLG

[449]

[450] **TRAC-ZFP-LEFT (SEQ ID No: 450)**

[451] FQCRICMRKFATSGSLTRHTKIHTGEKPFQCRICMRNFSRSDHLSTHIRTHTGE
KPFACDICGRKFATSSNRTKHTKIHTHPRAPIPKPFQCRICMRNFSRSDNLSEHI
RTHTGEKPFACDICGRKFAWHSSLRVHTKIHLR

[452] **TRAC-ZFP-Right (SEQ ID No: 451)**

[453] FQCRICMRNFSRSDHLSTHIRTHTGEKPFACDICGRKFADRSHLARHTKIHTG
SQKPFQCRICMRKFALKQHLNEHTKIHTGEKPFQCRICMRNFSQSGNLARHIRT
HTGEKPFACDICGRKFAHNSSLKDHTKIHLR

[454] **ND1-ZFP-Left (SEQ ID No: 452)**

[455] FQCRICMRNFSDSGNLRVHIRTHTGEKPYKCPDCGKSFSQSSSLIRHQRTHTG
EKPYECDHCGKSFSQSSHLNVHKRTHHTGEKPYRCKYCDRSFSISSNLQRHVRNI
HLR

[456] **ND1-ZFP-Right (SEQ ID No: 453)**

[457] YKPECGKSFSTKNSLTEHQRTHTGEKPYKPECGKSFSSKKALTEHQRTHT
GEKPYECNYCGKTFSVSSTLIRHQRIHTGEKPYRCKYCDRSFSISSNLQRHVRNI
HLR

[458]

[459] **G1333N-DddAtox (SEQ ID No: 454)**

[460] GSGSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSGG

[461] **G1333C-DddAtox (SEQ ID No: 455)**

[462] GSPTYPNYANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMET
LLPENAKMTVVPPEGAIPVKRGATGETKVFTGNSNSPKSPTKGGC

[463] **G1397N-DddAtox (SEQ ID No: 456)**

[464] GSGSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSGGPTYPN
YANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMETLLPENAKMT
VVPPEG

[465] **G1397C-DddAtox (SEQ ID No: 457)**

[466] GSAIPVKRGATGETKVFTGNSNSPKSPTKGGC

[467]

[468] **Adenine deaminase (AD: ABE 8e) (SEQ ID No: 458)**

[469] SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGL
HDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVV
FGVRNSKRGAAAGSLMNVLNYPGMNHRVEITEGILADECAALLCDFYRMPRQV
FNAQKKAQSSIN

[470]

- [471] **Full length DddAtox variant GSVG (SEQ ID No: 459)**
- [472] GSGSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLEKGVFSSGGPTYPN
YANAGHVESQSALFMRDNGISEGLVFHNNPEGTCGFCVNM TETLLPENAKMT
VVPPEGVIPVKRGATGETKVFTGNSNGPKSPTKGGC
- [473]
- [474] **Full length DddAtox variant E1347A (SEQ ID No: 460)**
- [475] GSGSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGGPTYPN
YANAGHVAGQSALFMRDNGISEGLVFHNNPEGTCGFCVNM TETLLPENAKMT
VVPPEGAIPVKRGATGETKVFTGNSNSPKSPTKGGC
- [476]
- [477] **Full length DddAtox variant AAAAA (SEQ ID No: 461)**
- [478] GSGSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGGPTYPN
YANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNM TETLLPENAKMT
VVPPEGAIPVAAGATGETAVFTGNSNSPASPTAGGC
- [479]
- [480] **UGI (SEQ ID No: 462)**
- [481] TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENV
MLLTSDAPEYKPWALVIQDSNGENKIKML
- [482]
- [483] **Single module**
- [484] **(TALE)-linker-AD-GSVG (SEQ ID No: 463)**
- [485] SGSETPGTSESATPESSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLN
NRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVM
CAGAMIHSRIGRVVFGVRNSKRGAAAGSLMNVLNYPGMNHRVEITEGILADEC
AALLCDFYRMPRQVFNAQKKAQSSINLVGSGSYALGPYQISAPQLPAYNGQTV
GTFYYVNDAGGLEKGVFSSGGPTYPN YANAGHVESQSALFMRDNGISEGLV
FHNNPEGTCGFCVNM TETLLPENAKMT VVPPEGVIPVKRGATGETKVFTGNSN
GPKSPTKGGC
- [486] **(TALE)-linker-AD-E1347A (SEQ ID No: 464)**
- [487] SGSETPGTSESATPESSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLN
NRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVM
CAGAMIHSRIGRVVFGVRNSKRGAAAGSLMNVLNYPGMNHRVEITEGILADEC
AALLCDFYRMPRQVFNAQKKAQSSINLVGSGSYALGPYQISAPQLPAYNGQTV
GTFYYVNDAGGLESKVFSSGGPTYPN YANAGHVAGQSALFMRDNGISEGLV
FHNNPEGTCGFCVNM TETLLPENAKMT VVPPEGAIPVKRGATGETKVFTGNSN
SPKSPTKGGC
- [488] **(TALE)-linker-AD-AAAAA (SEQ ID No: 465)**
- [489] SGSETPGTSESATPESSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLN

NRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVM
 CAGAMIHSRIGRVVFGVRNSKRGAAAGSLMNVLNYPGMNHRVEITEGILADEC
 AALLCDFYRMPRQVFNAQKKAQSSINLVGSGSYALGPYQISAPQLPAYNGQTV
 GTFYYVNDAGGLESKVFSGGPTYPNYANAGHVEGQSALFMRDNGISEGLV
 FHNNPEGTCGFCVNM TETLLPENAKMTVVPPEGAIPVAAGATGETAVFTGNS
 NSPASPTAGGC

[490]

[491] **Dual module**[492] **(TALE) -linker-AD (SEQ ID No: 466)**

[493] SGSETPGTSESATPESSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVN
 NRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVM
 CAGAMIHSRIGRVVFGVRNSKRGAAAGSLMNVLNYPGMNHRVEITEGILADEC
 AALLCDFYRMPRQVFNAQKKAQSSIN

[494] **(TALE)-GSVG (SEQ ID No: 467)**

[495] GSGSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLEKVFSSGGPTYPN
 YANAGHVESQSALFMRDNGISEGLVFHNNPEGTCGFCVNM TETLLPENAKMT
 VVPPEGVIPVKGATGETKVFTGNSNGPKSPTKGGC

[496] **(TALE)-E1347A (SEQ ID No: 468)**

[497] GSGSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSGGPTYPN
 YANAGHVAGQSALFMRDNGISEGLVFHNNPEGTCGFCVNM TETLLPENAKMT
 VVPPEGAIPVKGATGETKVFTGNSNSPKSPTKGGC

[498] **(TALE)-AAAAA (SEQ ID No: 469)**

[499] GSGSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSGGPTYPN
 YANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNM TETLLPENAKMT
 VVPPEGAIPVAAGATGETAVFTGNSNSPASPTAGGC

[500] **(TALE)-1397C-AD (SEQ ID No: 470)**

[501] GSAIPVKGATGETKVFTGNSNSPKSPTKGGCSGSETPGTSESATPESSEVEFS
 HEYWMRHALTLAKRARDEREVPVGAVLVLN NRVIGEGWNRAIGLHDPTAHA
 EIMALRQGGLVMQNYRLIDATLYVTFEPCVM CAGAMIHSRIGRVVFGVRNSK
 RGAAGSLMNVLNYPGMNHRVEITEGILADECAALLCDFYRMPRQVFNAQKK
 AQSSIN

[502] **(TALE)-1333N-AD (SEQ ID No: 471)**

[503] GSGSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSGGSGSETP
 GTSESATPESSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLN NRVIGEG
 WNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVM CAGAMIH
 SRIGRVVFGVRNSKRGAAAGSLMNVLNYPGMNHRVEITEGILADECAALLCDF
 YRMPRQVFNAQKKAQSSIN

[504] **(TALE)-1333C-AD (SEQ ID No: 472)**

- [505] GSPTPYPNYANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMETET
LLPENAKMTVVPPEGAIPVKRGATGETKVFTGNSNSPKSPTKGGCSGSETPGTS
ESATPESSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWN
RAIGLHDPTAHAEIMALRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRI
GRVVFVGVRNSKRGAAAGSLMNVLNYPGMNHRVEITEGILADECAALLCDFYR
MPRQVFNAQKKAQSSIN
- [506]
- [507] Linker(AA는 아미노산을 의미함)
- [508] 8AA: SGGGLGST (SEQ ID No: 473)
- [509] 16AA: SGSETPGTSESATPES (SEQ ID No: 474)
- [510] 32AA: SGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID No: 475)
- [511]
- [512] SOD2 MTS-3xHA:
MALSRVCGTSRQLAPVLGYLGSRQKHSLPDYPYDVPDYAGYPYDVPDYAG
YPYDVPDYA (SEQ ID No: 476)
- [513] COX8A MTS-3xFLAG:
- [514] MASVLTPLLLRGLTG SARRLPV PRAKIHSLDYKDHDGDYKDHDIDYKDDDD
K (SEQ ID No: 477)
- [515]
- [516] 5. 전달
- [517]
- [518] 본 발명에 따른 융합단백질은 예를 들어, 미세주입법 (microinjection),
전기천공법 (electroporation), DEAE-텍스트란 처리, 리포펙션 (lipofection),
나노파티클-매개 형질주입, 단백질 전달 도메인 매개 도입, 및 PEG-매개
형질주입 등과 같은 당업계의 다양한 방법에 의해 세포로 전달될 수 있지만,
이에 제한되는 것은 아니다.
- [519] 본 발명은 다른 관점에서, 상기 융합단백질을 코딩하는 핵산에 관한 것이다.
- [520] 상기 핵산과 관련하여, "폴리뉴클레오티드", "뉴클레오티드", "뉴클레오티드
서열", 및 "올리고뉴클레오티드"는 상호교환적으로 사용된다. 임의의 길이의
뉴클레오티드의 폴리머 형태, 테옥시리보뉴클레오티드 또는 리보뉴클레오티드,
또는 이들의 유사체를 포함할 수 있다. 폴리뉴클레오티드는 임의의 3차원
구조를 가질 수 있으며, 공지된 또는 비공지된 임의의 기능을 수행할 수 있다.
폴리뉴클레오티드는 하나 이상의 변형된 뉴클레오티드 예컨대, 메틸화된
뉴클레오티드 및 뉴클레오티드 유사체를 포함할 수 있다. 뉴클레오티드 구조에
대한 변형은 폴리머의 어셈블리 전 또는 후에 가능할 수 있다.
- [521] 상기 폴리뉴클레오티드는 RNA 서열, DNA 서열, 또는 이의 조합(RNA-DNA
조합 서열)일 수 있다.
- [522] 상기 융합단백질을 발현시키기 위한 수단으로서, 플라스미드 벡터, 코즈미드

백터, 박테리오파아지 백터 등 공지의 발현 백터를 사용할 수 있으며, 백터는 DNA 재조합 기술을 이용한 임의의 공지된 방법에 따라 당업자가 용이하게 제조할 수 있다.

- [523] 상기 백터는 플라스미드 백터 또는 바이러스 백터일 수 있으며, 상기 바이러스 백터는 구체적으로 아데노바이러스(Adenovirus), 아데노부속 바이러스(Adeno-associated virus), 렌티바이러스(Lentivirus), 레트로바이러스(Retrovirus) 백터일 수 있으나, 이에 한정된 것은 아니다.
- [524] 재조합 발현 백터는 숙주세포에서 핵산의 발현에 적합한 형태로 핵산을 포함할 수 있으며, 이는 재조합 발현 백터가 발현을 위해 사용되도록 즉, 발현될 핵산 서열에 작동 가능하게-연결되도록 숙주세포를 기반을 선택될 수 있는 하나 이상의 조절 요소를 포함함을 의미한다.
- [525] 재조합 발현 백터 내에서, "작동 가능하게 연결된"은 관심 뉴클레오티드 서열이 뉴클레오티드 서열의 발현을 허용하는 방식으로(예를 들어, 시험관내 전사/번역 시스템에서 또는 백터가 숙주세포 내로 도입되는 경우 숙주세포에서) 조절 요소에 연결됨을 의미한다.
- [526] 재조합 발현 백터는 T7 promoter를 포함하여 메신저 RNA 합성에 적합한 형태를 포함할 수 있으며, 이는 기내에서 mRNA 합성을 할 수 있도록, 즉, T7 polymerase에 의해 메신저 RNA가 합성될 수 있는 하나 이상의 조절 요소를 포함함을 의미한다.
- [527] "조절 요소"는 프로모터, 인핸서, 내부 리보솜 진입 부위(IRES), 및 기타 발현 조절 요소(예를 들어, 전사 종결 신호 예컨대, 폴리아데닐화 신호 및 폴리-U 서열)를 포함할 수 있다. 조절 요소는 많은 유형의 숙주세포에서 뉴클레오티드 서열의 유도 또는 항상 발현을 지시하는 요소 및 특정 숙주세포에서만 뉴클레오티드 서열의 발현을 지시하는 요소(예를 들어, 조직-특이적 조절 서열)를 포함한다. 조직-특이적 프로모터는 요망되는 관심 조직 예컨대, 근육, 뉴런, 뼈, 피부, 혈액, 특정 기관(예를 들어, 간, 췌장), 또는 특정 세포 유형(예를 들어, 림프구)에서 주로 발현을 지시할 수 있다. 조절 요소는 또한 조직 또는 세포-유형 특이적일 수 있거나 아닐 수 있는 세포-주기 의존성 또는 발달 단계-의존적 방식과 같은 일시적-의존적 방식으로 발현을 지시할 수 있다.
- [528] 경우에 따라서 백터는 하나 이상의 pol III 프로모터, 하나 이상의 pol II 프로모터, 하나 이상의 pol I 프로모터, 또는 이들의 조합을 포함한다. pol III 프로모터의 예는 비제한적으로 U6 및 H1 프로모터를 포함한다. pol II 프로모터의 예는 비제한적으로, 레트로바이러스 라우스 육종 바이러스(RSV) LTR 프로모터(임의로 RSV 인핸서를 가짐), 사이토메갈로바이러스(CMV) 프로모터(선택적으로 CMV 인핸서를 가짐)(예를 들어, Boshart et al(1985) Cell 41:521-530), SV40 프로모터, 디하이드로폴레이트 환원효소 프로모터, β -액틴 프로모터, 포스포글리세롤 키나제(PGK) 프로모터 및 EF1 α 프로모터를 포함한다.

- [529] "조절 요소"에는 인핸서 예를 들어, WPRE; CMV 인핸서; HTLV-I의 LTR에서 R-U5' 세그먼트; SV40 인핸서; 및 토끼 β -글로빈의 엑손 2 및 3 사이의 인트론 서열 등이 포함될 수 있다. 발현 벡터의 디자인은 형질전환될 숙주세포의 선택, 요망되는 발현 수준 등과 같은 인자에 좌우될 수 있음이 당업자에 의해 인식될 것이다. 벡터는 숙주세포 내로 도입되어 본원에 기재된 바와 같은 핵산에 의해 인코딩되는 융합 단백질 또는 펩티드를 포함하는 전사물, 단백질 또는 펩티드를 생성할 수 있다(예를 들어, 클러스터링된 규칙적 간격의 짧은 회문 반복부(CRISPR) 전사물, 단백질, 효소, 이의 돌연변이체, 이의 융합 단백질 등). 유사한 벡터는 렌티바이러스 및 아데노-관련 바이러스를 포함하고, 이러한 벡터의 유형은 또한 특정 유형의 세포를 표적화하기 위해 선택될 수 있다.
- [530] 상기 벡터는 각각 국소주입법 (예컨대, 병변 또는 표적 부위 직접 주입), 전기천공법 (electroporation), 리포펙션, 바이러스 벡터, 나노파티클 (nanoparticles) 뿐만 아니라, PTD (Protein translocation domain) 융합 단백질 방법 등을 통해 생체 내 또는 세포 내로 전달될 수 있다.
- [531] 상기 핵산은 리보핵산, 예를 들어 메신저 리보핵산 mRNA 형태로 주입하여 세포의 유전자 염기 교정, 예를 들어 동물세포 또는 식물세포에 제한없이 유전자 염기 교정이 가능하다.
- [532] 본 발명에 따른 핵산은 mRNA 형태일 수 있고, mRNA 형태로 전달하는 경우 DNA를 이용한 벡터 형태로 전달하는 것에 비해, mRNA로의 전사 과정이 불필요하여 유전자 교정 개시가 빠를 수 있다. 일시적 단백질 발현 (transient protein expression) 가능성이 높다.
- [533] 본 출원의 발명자들은 식물 소기관 유전자 교정을 위해 시토신 염기 교정기 (cytosine base editor)를 리보핵산, 예를 들어 메신저 리보핵산 형태로 식물 세포에 주입하였을 때 플라스미드로 전달하는 것에 비해 비표적 효과가 감소함을 확인하였다. 식물 소기관 유전자 교정에 있어 시토신 염기 교정기를 mRNA 형태로 식물세포에 형질전환 하였을 때, 플라스미드에 비해 비표적 효과에서 장점이 있음을 최초로 입증하였다.
- [534] 상기 mRNA는 직접 전달되거나, 캐리어를 통해 전달될 수 있다. 경우에 따라서, 핵산절단효소 및/또는 절단인자의 mRNA를 화학변형 (chemical modification) 하거나, 합성 자기-복제 RNA (synthetic self-replicative RNA) 형태로 직접 전달할 수 있다.
- [535] 세포로 mRNA를 전달하는 방법 또는 생체 내에서 mRNA를 인간, 동물과 같은 유기체의 세포로 전달하는 방법을 포함하여, 시험관 내 또는 생체 내에서 mRNA 분자를 세포로 전달할 수 있는 방법을 고려할 수 있다. 예를 들어, 지질(예를 들어, 리포솜, 마이셀 등)을 포함하거나, 나노파티클 또는 나노튜브를 포함하거나, 양이온성 화합물(예를 들어, 폴리에틸렌이민 또는 PEI)을 포함하거나, 양이온성 화합물(예를 들어, 폴리에틸렌이민 또는 PEI)을 포함하여, mRNA 분자를 세포로 전달할 수 있다. 경우에 따라서, mRNA를 세포로 전달하기

- 위해서 유전자 총 또는 바이오리스틱(biolistic) 입자 전달 시스템과 같은 볼리스틱 방법(bolistics method)을 사용할 수 있다.
- [536] 상기 캐리어는 예를 들어, 세포 투과성 펩타이드(CPP), 나노입자, 또는 중합체를 포함할 수 있으나, 이에 제한되는 것은 아니다.
- [537] 상기 CPP는 다양한 분자 카고(나노크기 입자로부터 작은 화학 분자 및 DNA의 큰 단편까지)의 세포 흡수를 용이하게 하는 짧은 펩타이드이다.
- [538] 상기 나노입자와 관련하여, 본 발명에 따른 조성물은 고분자 나노입자, 금속 나노입자, 금속/무기 나노입자 또는 지질 나노입자를 통해 전달될 수 있다. 상기 고분자 나노입자는 예를 들어, 롤링 씨클 증폭에 의해 합성된 DNA nanoclew, 실 유사 DNA 나노입자일 수 있다. DNA nanoclew, 실 유사 DNA 나노입자에 mRNA를 로딩하고, 엔도솜 탈출능을 향상시키기 위해 PEI를 코팅할 수 있다. 이러한 복합체가 세포막에 결합하여, 내재화된 다음 엔도솜 탈출을 통해 핵으로 이동하여, 전달되도록 할 수 있다.
- [539] 상기 금속 나노입자와 관련하여, 금입자를 연결시키고 양이온성 엔도솜 탈출 폴리머 (endosomal disruptive polymer)와 복합체를 형성하여 세포에 전달할 수 있다. 상기 양이온성 엔도솜 탈출 폴리머 (endosomal disruptive polymer)는 예를 들어, polyethylene imine, poly(arginine), poly(lysine), poly(histidine), poly-[2-((2-aminoethyl)amino)-ethyl-aspartamide] (pAsp(DET)), poly(ethylene glycol) (PEG)와 poly(arginine)의 block co-polymer, PEG와 poly(lysine)의 block co-polymer, 또는 PEG와 poly{N-[N-(2-aminoethyl)-2-aminoethyl]aspartamide} (PEG-pAsp(DET))의 block co-polymer일 수 있다.
- [540] 상기 금속/무기 나노입자와 관련하여, 예를 들어 ZIF-8 (zeolitic imidazolate framework-8)를 통해 mRNA를 캡슐화할 수 있다.
- [541] 경우에 따라서, 상기 mRNA는 음이온을 띠고(negatively charged), 양이온을 띠는(cationic) 물질들과 결합하여 나노입자 (nanoparticles)를 형성할 수 있고, 이는 세포로 수용체 매개 엔도사이토시스 (receptor-mediated endocytosis) 혹은 파고사이토시스 (phagocytosis) 등을 통해 침투할 수 있다.
- [542] 양이온성 중합체로 폴리알릴아민 (PAH); 폴리에틸렌이민 (PEI); 폴리(L-리신) (PLL); 폴리(L-아르기닌) (PLA); 폴리비닐아민 단일- 또는 공중합체; 폴리(비닐벤질-트리-C1-C4-알킬암모늄염); 지방족 또는 방향지방족 디할로젠화물 및 지방족 N,N,N',N'-테트라-C1-C4-알킬-알킬렌디아민의 중합체; 폴리(비닐피리딘) 또는 폴리(비닐피리디늄염); 폴리(N,N-디알릴-N,N-디-C1-C4-알킬-암모늄할로젠화물); 4차화된 디-C1-C4-알킬-아미노에틸 아크릴레이트 또는 메타크릴레이트의 단일- 또는 공중합체; 폴리쿠아드(POLYQUAD)TM; 폴리아미노아미드 등이 포함될 수 있다.
- [543] 양이온성 지질로 양이온성 리포솜 제제를 포함할 수 있다. 리포솜의 지질 이중층은 캡슐화된 핵산을 분해로부터 보호하며, 핵산에 결합할 수 있는 항체에 의한 특이적 중화를 방지할 수 있다. 엔도솜 성숙 동안 엔도솜 멤브레인

리포좀이 융합되어, 양이온성 지질-핵산절단효소의 효율적인 엔도솜 탈출이 가능하다. 양이온성 지질로 폴리에틸렌이민, 폴리아미도아민 (PAMAM) 성화 수지상체, 리포펙틴 (DOTMA와 DOPE의 조합), 리포펙타제, 리포펙타민(LIPOFECTAMINE)® (예컨대 리포펙타민® 2000, 리포펙타민® 3000, 리포펙타민® RNAiMAX, 리포펙타민® LTX), 세인트-레드(SAINT-RED) (시노블룩스 세라퓨틱스(Synvolux Therapeutics)사, 네덜란드 그로닝겐 소재), DOPE, 시토펙틴 (길레드 사이언시즈(Gilead Sciences) 사, 캘리포니아 포스터 시티 소재), 및 유펙틴 (JBL사, 캘리포니아 산 루이스 오비스포 소재)가 포함될 수 있다. 대표적인 양이온성 리포솜은

N-[1-(2,3-디올레올옥시)-프로필]-N,N,N-트리메틸암모늄 클로라이드 (DOTMA), N-[1-(2,3-디올레올옥시)-프로필]-N,N,N-트리메틸암모늄 메틸술페이트 (DOTAP), 3β-[N-(N',N'-디메틸아미노에탄)카르바모일]콜레스테롤 (DC-Chol), 2,3-디올레일옥시-N-[2(스페르민카르복사미도)에틸]-N,N-디메틸-1-프로판아미늄 트리플루오로아세테이트 (DOSPA),

1,2-디미리스틸옥시프로필-3-디메틸-히드록시에틸 암모늄 브로마이드; 또는 디메틸디옥타데실암모늄 브로마이드 (DDAB)으로부터 제조될 수 있다.

- [544] 지질 나노입자와 관련하여, 캐리어로 리포솜을 이용하여 전달할 수 있다. 리포솜은 내부 수성 구획을 둘러싸는 단일 또는 다중 라멜라 지질 이중층 및 상대적으로 불투과성인 외측 친지성 인지질 이중층으로 구성된 구체 소낭 구조이다. 리포솜 제형은 주로 천연 인지질 및 지질, 예컨대 1,2-디스테아로일-sn-글리세로-3-포스파티딜 콜린(DSPC), 스펡고미엘린, 포스파티딜콜린 또는 모노시알로강글리오사이드 등을 포함할 수 있다. 경우에 따라서, 혈장 내 불안정성을 해소하기 위하여, 지질막에 콜레스테롤 또는 1,2-디올레오일-sn-글리세로-3-포스포에탄올아민(DOPE)을 추가할 수 있다. 콜레스테롤의 첨가는 캡슐화된 생활성 화합물의 혈장 내로의 신속한 방출을 감소시키거나 또는 1,2-디올레오일-sn-글리세로-3-포스포에탄올아민(DOPE)은 안정성을 증가시킨다.

[545]

[546] 7. 염기 교정

[547]

[548] 본 발명은 다른 관점에서, 상기 융합단백질 또는 상기 핵산을 포함하는 염기 교정용 조성물에 관한 것이다.

[549] 본 발명은 다른 관점에서, 상기 조성물을 세포에 처리하는 단계를 포함하는 염기 교정방법에 관한 것이다.

[550] DNA 결합 단백질, 예를 들어 TALE 또는 ZFP (zinc finger protein)가 타겟 DNA에 결합한 후, 융합단백질의 시토신 탈아미노효소는 시토신의 아미노기를 가수분해하여 우라실의 형태로 변형시킨다. 우라실은 아데닌과 염기 쌍을 이룰 수 있기 때문에, 세포 내 DNA 복제 과정을 거쳐 시토신-구아닌 염기 쌍은

우라실-아데닌 염기 쌍을 거쳐 최종적으로 티민-아데닌 염기 쌍으로 교정될 수 있다. 또한, 융합단백질의 아데닌 탈아미노효소는 아데닌의 아미노기를 가수분해 하여 하이포잔틴(Hypoxanthine)의 형태로 변형시킨다. 하이포잔틴은 시토신과 염기쌍을 이룰 수 있기 때문에, 마찬가지로 세포 내 DNA 복제 과정에서 아데닌-티민 염기쌍은 하이포잔틴-시토신 염기 쌍을 거쳐 구아닌-시토신 염기 쌍으로 교정될 수 있다.

- [551] 상기 세포는 진핵세포 (예컨대, 효모 등의 균류, 진핵 동물 및 /또는 진핵 식물 유래 세포 (예컨대, 배아세포, 줄기세포, 체세포, 생식세포 등) 등), 진핵 동물 (예컨대, 인간, 원숭이 등의 영장류 개, 돼지, 소, 양, 염소, 마우스, 래트 등), 또는 진핵 식물 (예컨대, 녹조류 등의 조류, 옥수수, 콩, 밀, 벼 등)일 수 있으나, 이에 제한되는 것은 아니다.
- [552]
- [553] (1) 식물세포 DNA의 염기 교정
- [554]
- [555] 본 발명은 식물세포 DNA의 염기 교정용 조성물 또는 염기 교정방법에 관한 것이다. 상기 융합단백질 또는 이를 코딩하는 핵산; 및 NLS (nuclear localization signal) 펩타이드, 엽록체 전달 펩타이드(chloroplast transit peptide), 미토콘드리아 전달 단백질 (Mitochondrial Targeting Signal, MTS), 핵외 수송 신호 단백질 또는 이를 코딩하는 핵산을 포함하는, 식물세포 염기 교정용 조성물에 관한 것이다.
- [556] 본 발명은 또한, 상기 융합단백질 또는 핵산; 및 NLS (nuclear localization signal) 펩타이드 또는 이를 코딩하는 핵산을 포함하는 식물세포 염기 교정용 조성물을 제공한다.
- [557] 본 발명은 또한, 상기 융합단백질 또는 핵산; 및 엽록체 전달 펩타이드(chloroplast transit peptide) 또는 이를 코딩하는 핵산을 포함하는 식물세포 염기 교정용 조성물을 제공한다.
- [558] 본 발명은 또한, 상기 융합단백질 또는 핵산; 및 미토콘드리아 전달 단백질 (Mitochondrial Targeting Signal, MTS) 또는 이를 코딩하는 핵산을 포함하는 식물세포 염기 교정용 조성물을 제공한다.
- [559] 경우에 따라서 본 발명은 또한, 핵외 수송 신호 단백질 또는 이를 코딩하는 핵산을 추가로 포함하는, 식물세포 염기 교정용 조성물을 제공한다.
- [560] 구체적으로, 식물세포 핵 DNA, 미토콘드리아 DNA 또는 엽록체 DNA의 교정용 조성물 또는 염기 교정방법에 관한 것이다.
- [561] 구체적으로, 상기 융합단백질은 다음을 통해 식물세포에 전달될 수 있다:
- [562] Gene gun을 이용한 주입 (Bombardment);
- [563] PEG-매개 원형질체 형질감염 (Protoplast transfection, PEG-mediated);
- [564] 전기천공에 의한 원형질체 형질감염 (Protoplast transfection, electroporation); 또는
- [565] 미세주사에 의한 원형질체 주사 (Protoplast injection, microinjection).

- [566] 본 발명에 따른 융합단백질을 코딩하는 폴리뉴클레오티드 서열은 RNA 서열, DNA 서열, 또는 이의 조합(RNA-DNA 조합 서열)일 수 있다.
- [567] 상기 융합단백질을 코딩하는 폴리뉴클레오티드는 다음을 통해 식물세포에 전달될 수 있다:
- [568] *Agrobacterium*, 예를 들어 *Agrobacterium tumefaciens*, *Agrobacterium rhizogene* 등을 이용한 형질전환
- [569] - binary vector,
- [570] - Viral vector: Geminivirus, tobacco rattle virus (TRV), tomato mosaic virus (ToMV), foxtail mosaic virus (FoMV), barley yellow striate mosaic virus (BYSMV), *Sonchus yellow net rhabdovirus* (SYNV) 등;
- [571] 바이러스를 이용한 형질감염;
- [572] 주입 (Bombardment, Gene gun);
- [573] PEG-매개 원형질체 형질감염 (Protoplast transfection, PEG-mediated);
- [574] 전기천공에 의한 원형질체 형질감염 (Protoplast transfection, electroporation); 또는
- [575] 미세주사에 의한 원형질체 주사 (Protoplast injection, microinjection).
- [576] 상기 바이러스는 예를 들어, viral vector: Geminivirus, tobacco rattle virus (TRV), tomato mosaic virus (ToMV), foxtail mosaic virus (FoMV), barley yellow striate mosaic virus (BYSMV), *Sonchus yellow net rhabdovirus* (SYNV) 등을 예로 들 수 있다.
- [577] 상기 벡터는 각각 국소주입법 (예컨대, 병변 또는 표적 부위 직접 주입), 전기천공법 (electroporation), 리포펙션, 바이러스 벡터, 나노파티클 (nanoparticles) 뿐만 아니라, PTD (Protein translocation domain) 융합 단백질 방법 등을 통해 세포내로 전달될 수 있다.
- [578] 상기 식물 기관으로 수송되는 단백질 또는 이를 코딩하는 핵산과 관련하여, 상기 식물 기관은 미토콘드리아, 엽록체 또는 색소체 (plastid: 백색체, 유색체)일 수 있다.
- [579] 상기 식물 기관으로 수송되는 단백질은 예를 들어 엽록체 전달 펩타이드(chloroplast transit peptide) 또는 미토콘드리아 전달 단백질 (Mitochondrial Targeting Signal, MTS)일 수 있다.
- [580] 예를 들어, 엽록체 전달 단백질 (Chloroplast Transit Peptide, CTP) 또는 미토콘드리아 전달 신호 단백질 (Mitochondrial Targeting Signal, MTS)이 결합하여 식물 세포 내 엽록체와 미토콘드리아에 전달된다. 엽록체와 미토콘드리아로 전달될 때 N-말단의 CTP 또는 MTS 단백질을 제외한 나머지 부분은 전-단백질 (preprotein) 형태로 엽록체와 미토콘드리아 내부로 전달된다. 엽록체와 미토콘드리아 내부로 진입 과정에서 전달 단백질 부분이 떨어져 나가게 되고 엽록체와 미토콘드리아를 타겟팅하여 특정 부분을 염기 교정할 수 있다.

- [581] 상기 융합단백질 또는 이를 코딩하는 핵산 이외에, CTP (chloroplast transit peptide) 또는 이를 코딩하는 핵산 또는 미토콘드리아 전달 신호 단백질 (Mitochondrial Targeting Signal, MTS) 또는 이를 코딩하는 핵산을 식물세포에 처리 또는 전달하여 식물 미토콘드리아, 엽록체, 유색체 또는 백색체 DNA의 염기를 교정할 수 있다.
- [582] 미토콘드리아 유전자 교정시에 핵외 수송 서열을 염기 교정 단백질에 부착한 경우 보다 높은 효율로 염기가 교정될 수 있다. 상기 핵외 수송 신호 단백질은 예를 들어, MVM (Mirute virus of mice) 유래일 수 있으나, 이에 제한되는 것은 아니다. 상기 핵외 수송 신호 단백질은 예를 들어, 서열번호 31의 아미노산 서열을 포함할 수 있으나, 이에 제한되는 것은 아니다.
- [583] VDEMTKKFGTLTIHDEK (서열번호 31)
- [584] 본 발명은 또한, 야생형 DNA 염기서열을 자르지만 교정된 염기서열은 자르지 못하는 TAL (Transcription Activator-Like) 이펙터 (TALE) 도메인 및 FokI 핵산 가수분해효소(FokI Nuclease) 또는 이를 코딩하는 핵산 또는 ZFN(zinc finger nuclease) 또는 이를 코딩하는 핵산 즉, 미토콘드리아 핵산 가수분해 효소인 mitoTALEN (Mitochondrial TALE Nuclease) 또는 이를 코딩하는 핵산 또는 ZFN(zinc finger nuclease) 또는 이를 코딩하는 핵산을 추가로 포함함으로써, 미토콘드리아 서열 절단 단백질을 동시에 이용하는 경우에도 보다 높은 효율의 미토콘드리아 염기 교정을 기대할 수 있다.
- [585]
- [586]
- [587] (2) 동물세포 DNA의 염기 교정
- [588]
- [589] 본 발명은 동물세포 DNA의 염기 교정용 조성물 또는 염기 교정방법에 관한 것이다. 상기 융합단백질 또는 이를 코딩하는 핵산; 및 NLS (nuclear localization signal) 펩타이드, 미토콘드리아 전달 단백질 (Mitochondrial Targeting Signal, MTS), 핵외 수송 신호 단백질 또는 이를 코딩하는 핵산을 포함하는, 동물세포 염기 교정용 조성물에 관한 것이다.
- [590] 본 발명은 또한, 상기 융합단백질 또는 핵산; 및 NLS (nuclear localization signal) 펩타이드 또는 이를 코딩하는 핵산을 포함하는, 동물세포 염기 교정용 조성물을 제공한다.
- [591] 본 발명은 또한, 상기 융합단백질 또는 핵산; 및 미토콘드리아 전달 단백질 (Mitochondrial Targeting Signal, MTS) 또는 이를 코딩하는 핵산을 포함하는, 동물세포 염기 교정용 조성물을 제공한다.
- [592] 본 발명은 또한, 경우에 따라서 핵외 수송 신호 단백질 또는 이를 코딩하는 핵산을 추가로 포함하는, 동물세포 염기 교정용 조성물을 제공한다.
- [593] 상기 동물세포는 비인간 동물세포이고, 핵외 수송 신호 단백질 또는 이를 코딩하는 핵산; 및/또는 미토콘드리아 전달 신호 단백질 (Mitochondrial Targeting

Signal, MTS) 또는 이를 코딩하는 핵산을 처리하여 비인간 동물세포 미토콘드리아 DNA의 염기를 교정할 수 있다.

- [594] 예를 들어, 상기 융합단백질 또는 이를 코딩하는 핵산 이외에, 미토콘드리아 전달 단백질 (Mitochondrial Targeting Signal, MTS)이 결합하여 미토콘드리아에 전달된다. 미토콘드리아로 전달될 때 N-말단의 MTS 단백질을 제외한 나머지 부분은 전-단백질 (preprotein) 형태로 미토콘드리아 내부로 전달된다. 미토콘드리아 내부로 진입 과정에서 전달 단백질 부분이 떨어져 나가게 되고 미토콘드리아를 타겟팅하여 특정 부분을 염기 교정할 수 있다.
- [595] 미토콘드리아 전달 신호, TAL 이펙터 및 시토신 탈아미노화 효소 (cytosine deaminase, DddA_{tox})를 포함하는 TALE-DdCBE (TALE DddA-derived cytosine base editor) 또는 이를 코딩하는 핵산에 핵외 수송 신호(Nuclear Export Signal, NES) 또는 이를 코딩하는 핵산이 결합된 비인간 동물세포 미토콘드리아 DNA의 염기 교정용 조성물 또는 교정방법에 관한 것이다. 핵외 수송 신호로 인해 미토콘드리아-핵 유사 서열에 대한 핵 DNA 염기 교정을 감소시킬 수 있다.
- [596] 본 발명에 따르면, 핵외 수송 신호 단백질 또는 이를 코딩하는 핵산을 포함하여, 보다 효율적인 동물 미토콘드리아 DNA 염기 교정을 달성할 수 있다. 또한, 핵외 수송 신호로 인해 미토콘드리아-핵 유사 서열에 대한 핵 DNA 염기 교정을 감소시켜, 미토콘드리아 DNA만 교정될 수 있도록 할 수 있다.
- [597] 상기 핵외 수송 신호 단백질은 예를 들어, MVM (Mirute virus of mice) 유래일 수 있으나, 이에 제한되는 것은 아니다. 상기 핵외 수송 신호 단백질은 예를 들어, VDEMTKKFGTLTIHDTEK (서열번호 31)의 아미노산 서열을 포함할 수 있으나, 이에 제한되는 것은 아니다.
- [598] 상기 (1) 핵외 수송 신호 단백질 또는 이를 코딩하는 핵산 및 (2) DNA 결합 단백질, 탈아미노효소 (deaminase) 또는 이의 변이체 또는 이를 코딩하는 핵산과 동시에 또는 순차적으로 교정되기 전 야생형 DNA 염기서열을 자르지만 교정된 염기서열은 자르지 못하는 TAL (Transcription Activator-Like) 이펙터 (TALE) 도메인 및 FokI 핵산 가수분해효소(FokI Nuclease) 또는 이를 코딩하는 핵산 또는 ZFN(zinc finger nuclease) 또는 이를 코딩하는 핵산을 진핵세포에 주입할 수 있다.
- [599] 특히, 진핵세포 미토콘드리아 유전자의 염기 교정과 관련하여, 핵외 수송 신호 단백질 또는 이를 코딩하는 핵산 및/또는 미토콘드리아 전달 신호 단백질 (Mitochondrial Targeting Signal, MTS) 또는 이를 코딩하는 핵산을 포함할 수 있다.
- [600] 본 발명에 따르면, 동물 미토콘드리아 유전자 교정시에 핵외 수송 서열을 염기 교정 단백질에 부착한 경우 보다 높은 효율로 염기가 교정되며, 동물 배아에서는 특히 핵내 유사 서열의 비특이적 염기 교정 역시 억제되는 것으로 나타났다.
- [601] 본 발명은 또한, 미토콘드리아 핵산 가수분해 효소인 mitoTALEN (Mitochondrial TALE Nuclease) 또는 이를 코딩하는 핵산을 추가로 포함함으로써, 미토콘드리아 서열 절단 단백질을 동시에 이용하는 경우에도 보다 높은 효율의 미토콘드리아 염기 교정을 기대할 수 있다. 미토콘드리아 DNA 핵산 가수 분해

효소인 mitoTALEN (Mitochondrial TALE Nuclease)을 이용하여, 미토콘드리아 DNA를 절단할 수 있으며, 야생형(Wild-type) 미토콘드리아 유전체를 절단하여 동물에서 염기 교정된 유전체를 높은 효율로 얻을 수 있다.

- [602] 본 발명은 또한, 미토콘드리아 핵산 가수분해 효소인 mitoTALEN (Mitochondrial TALE Nuclease) 또는 이를 코딩하는 핵산을 추가로 포함함으로써, 미토콘드리아 서열 절단 단백질을 동시에 이용하는 경우에도 보다 높은 효율의 미토콘드리아 염기 교정을 기대할 수 있다. 구체적으로, 미토콘드리아 전달 신호가 결합된 TAL 이펙터 도메인 및 FokI 핵산 가수분해효소(FokI Nuclease) 또는 ZFN 또는 이를 코딩하는 핵산을 갖는 융합 단백질(mitoTALEN) 또는 이를 코딩하는 핵산을 포함할 수 있다.
- [603] 미토콘드리아 DNA 핵산 가수 분해 효소인 mitoTALEN (Mitochondrial TALE Nuclease)을 이용하여, 미토콘드리아 DNA를 절단할 수 있으며, 야생형(Wild-type) 미토콘드리아 유전체를 절단하여 동물에서 염기 교정된 유전체를 높은 효율로 얻을 수 있다.
- [604] 경우에 따라서, 염기 교정 효율을 올려줄 수 있는 UGI (uracil DNA glycosylase inhibitor)를 추가로 포함할 수 있다. UGI는 DNA에서 U를 제거하는 작용을 촉매하는 돌연변이 DNA를 복구하는 효소 UDG (Uracil DNA glycosylase) 활성을 억제시켜 염기 교정 효율을 증가시킬 수 있다.
- [605] 구체적으로, 분할된 박테리아 독소 DddAtox, DNA에 결합할 수 있도록 고안 된 TALE (transcription activator-like effector) 및 우라실 글리코실레이즈 억제제 (UGI)로 이루어진 DddA-유래 시토신 염기 교정기 (DdCBEs)는 목표한 시토신-티민 염기 변형을 미토콘드리아 DNA에서 가능하게 하였다. 실시예에 따르면, 생쥐 배아에서 높은 효율의 미토콘드리아 DNA 교정이 가능함을 보여주었다. 미토콘드리아 유전자 중 NADH 탈수 및 유비퀴논으로의 전자전달을 촉매하는 NADH 탈수소화 효소의 서브유닛을 코딩하는 MT-ND5(ND5)를 목표로 하였는데, 이는 m.G12918A와 같이 사람의 미토콘드리아 질병과 관련된 돌연변이와 m.C12336T와 같이 이른 종결코돈을 만드는 돌연변이를 포함하였다. 이를 통해 미토콘드리아 질병 모델을 생쥐에서 제작할 수 있었고, 미토콘드리아 질병을 치료할 수 있는 가능성을 제시하였다.
- [606] 상기 (1) 핵외 수송 신호 단백질 또는 이를 코딩하는 핵산에 (2) DNA 결합 단백질, 탈아미노효소 (deaminase) 또는 이의 변이체 또는 이를 코딩하는 핵산이 연결되고, 상기 (2)에 핵산 가수분해 효소인 (3) mitoTALEN (Mitochondrial TALE Nuclease) 또는 이를 코딩하는 핵산을 연결할 수 있다. (1) 내지 (3)을 전달하기 위해 단일 또는 복수의 전달 수단을 동일 또는 상이한 구성으로 조합하여 사용할 수 있다.
- [607] 상기 (1)은 제1전달 수단에 포함되고, (2)는 제2전달 수단에 포함되며, (3)은 제3전달 수단에 포함될 수 있다. 각각의 전달 시스템은 동시에 바이러스 전달수단이거나, 하나는 바이러스 전달수단이고 다른 하나는 비-바이러스

전달수단이거나, 동시에 비-바이러스 전달수단일 수 있다.

- [608] 상기 (1) 내지 (3)의 핵외 수송 신호 단백질, DdCBE, mitoTALEN을 혼합하여 전달할 수 있다.
- [609] 상기 (1) 내지 (3) 중 하나 이상은 핵외 수송 신호 단백질, DdCBE, mitoTALEN으로 전달하고, 일부는 (1) 내지 (3)를 코딩하는 DNA 서열이 벡터 상에 위치하도록 하여 전달할 수 있다.
- [610] 상기 (1) 내지 (3)를 코딩하는 DNA 서열이 동일한 벡터 상에 위치하도록 하여 하나의 벡터를 통해 동시에 전달할 수 있거나, 별개의 다른 벡터상에 위치하도록 하여 전달할 수 있다.
- [611] 본 발명에 따른 동물은 인간 또는 비인간 동물을 포함할 수 있다. 상기 비인간 형질전환 동물은 곤충류, 환형동물, 연체동물, 완족류, 선형동물, 강장동물, 해면류, 척색류, 척추동물일 수 있고 상기 척추동물은 어류, 양서류, 파충류, 조류 또는 포유류일 수 있으며, 상기 곤충류는 초파리(*Drosophila*)일 수 있고, 상기 선형동물은 예쁜꼬마선충(*C. elegans*)일 수 있으며, 상기 어류는 지브라피쉬(*zebrafish*)일 수 있고, 상기 포유류는 영장목, 식육목, 식충목, 설치목, 우제목, 기제목 또는 장비목일 수 있고, 상기 설치목은 래트 또는 마우스일 수 있다.
- [612] 본 발명에 따른 조성물을 인간 또는 비인간 동물의 배아에 도입하고, 상기 배아를 대리 모체에서 이식하고 임신시켜 염기 교정된 동물을 제조할 수 있다. 동물의 수정란에 본 발명에 따른 조성물을 도입하고 배양할 수 있다.
- [613] 상기 수득한 수정란을 대리모에 이식하여 분만시킬 수 있다. 상기 비인간 형질전환 동물은 분만 이후 형질전환 여부를 확인하는 단계를 추가로 포함할 수 있다. 상기 비인간 형질전환 동물을 교배하여 자손 형질전환 동물을 생산할 수 있다.
- [614] 상기 "자손"은 상기 비인간 형질전환 동물과 교배하여 생존가능한 모든 새끼 형질전환 동물을 의미하며, 보다 구체적으로는 상기 형질전환 동물을 부모로 형질전환 동물끼리 또는 상기 형질전환 동물과 정상 동물을 교배하여 생산한 F1 세대, F1 세대의 동물과 정상 동물을 교배하여 생산한 F2세대 및 그 이후 세대를 의미하나, 이에 한정되는 것은 아니다.
- [615] 상기 교배는 상기 형질전환 동물 또는 정상 동물과 교배하는 것을 특징으로 할 수 있다. 본 발명은 상기 형질전환 동물 또는 자손 형질전환 동물로부터 분리한 세포, 조직 및 부산물을 포함할 수 있다. 상기 부산물은 상기 형질전환 토끼로부터 유래한 물질 모든 것을 의미하나 바람직하게는 혈액, 혈청, 소변, 대변, 침, 장기 및 피부로 이루어진 군에서 선택되는 것을 특징으로 할 수 있으나, 이에 한정되는 것은 아니다.
- [616]
- [617] 실시예
- [618] 이하, 실시예를 통하여 본 발명을 더욱 상세히 설명하고자 한다. 이들 실시예는

오로지 본 발명을 예시하기 위한 것으로서, 본 발명의 범위가 이들 실시예에 의해 제한되는 것으로 해석되지는 않는 것은 당업계에서 통상의 지식을 가진 자에게 있어서 자명할 것이다.

[619]

[620] 실시예 1. 징크 핑거 탈아미노효소 (Zinc Finger Deaminase: ZFD)

[621] 핵 DNA 또는 미토콘드리아 DNA의 염기 교정은 의생명과학연구, 의학 및 생명공학에 널리 유용하다. ZFD 플랫폼은 DNA 결합 단백질, 분할 박테리아 독성 탈아미노효소 DDDAtox(interbacterial toxin deaminase DDDAtox) 및 우라실 글리코실라제 억제제(uracil glycosylase inhibitor; UGI)를 포함한다. UGI는 인간 세포에서 원치 않는 작은 삽입 및 결실(indel)을 유도하지 않고 표적화된 C를 T로의 염기 전환을 촉매한다. 공개적으로 사용 가능한 징크 핑거 리소스를 사용하여 ZFD를 인코딩하는 플라스미드를 제작하여 핵 DNA에서 최대 60% 및 미토콘드리아 DNA에서 30% 빈도로 염기 교정을 달성했다. ZFD는 CRISPR 기반의 염기교정과는 다르게 DNA를 잘라 단일 또는 이중가닥 절단을 만들지 않기 때문에 표적 위치에서 비-상동말단부착 (error-prone non-homologous end joinig)에 의한 원하지 않는 삽입 및 결실(indel)이 발생하지 않는다. 더욱이, 대장균(*E. coli*)에서 정제된 재조합(recombinant)된 ZFD 단백질들은 자발적으로 사람 세포를 통과해서 타겟 염기를 다른 염기로 전환시킨다(base conversion). 이는 유전자 없는 유전자치료의 원리검증(proof-of-principle)을 입증한다.

[622] 진핵 세포 및 유기체의 게놈 교정을 위한 기술에는 zinc finger nucleases (ZFNs), transcription activator-like effector (TALE) nucleases (TALENs), TALE-linked split interbacterial deaminase toxin DddA-derived cytosine base editors (a.k.a. DdCBEs), CRISPR-Cas9, 및 자르는 활성이 없어진 Cas9-linked deaminases (a.k.a. base editors)가 포함 되어 있지만, 이것에만 국한되진 않는다. 이러한 도구들은 원칙적으로 두가지 기능 단위로 구성된다. DNA 결합부분과 촉매 부분이다. 따라서 징크 핑거 어레이 (zinc finger array) 또는 TALE 어레이 (array)들은 DNA 결합부분으로 기능하고, 반면에 뉴클레아제 (ZFN 및 TALEN의 경우 FokI) 또는 탈아미노효소 (DdCBE에서의 split DddA_{tox} 및 CBE에서의 APOBEC1)가 촉매 단위로 기능한다. Crispr-cas9은 뉴클레아제 기능과 RNA로 유도되는 DNA 결합단백질(RNA-guided DNA binding protein) 기능 모두를 가지고 있다. 맞춤형 설계된, 프로그램 가능한 (programmable)한 ZFN, TALEN 및 Cas9과 같은 뉴클레아제는 DNA를 잘라 두가닥 절단(double strand break)을 일으킨다. 이를 복구하면서 타겟화된 방식 (targeted manner)으로 유전자 knock-out 및 knock-in을 일으킨다. 그러나 프로그램 가능한 뉴클레아제가 유도한 두 가닥 절단은 표적 부위에서 원하지 않는 유전자의 큰 손실(large deletion), p53 활성화 및 표적부위와 표적부위가 아닌 곳의 동시 DSB 수선과정에서 염색체 재배열을 유발할 수 있다. 대조적으로, 시토신 및 아데닌 염기 교정기기(CBE and ABE)를 포함하는 프로그램 가능한 염기 교정기들은 DSB를 일으키지 않아 세포에서

이러한 원하지 않는 현상들을 피할 수가 있으며, 수선 템플릿 (repair template) 또는 공여자 DNA(donor DNA) 없이 효율적으로 한 개의 염기 치환(single nucleotide conversion)을 촉매할 수 있다. 그러나 CBE나 ABE는 Cas9 니케이즈 변이체들을 포함하기 때문에 DNA 한 쪽 가닥을 잘라 닉 (nick) 또는 단일 가닥 절단 (single-strand break)을 일으켜 여전히 유전자 표적부위에서 원하지 않는 삽입-결실(indel)이 일어난다.

- [623] 이는 세포안에서 핵 DNA와 미토콘드리아 DNA의 C를 T로 염기 치환을 촉매할 수 있다. 맞춤형(custom-designed) DdCBE를 사용하여 쥐의 미토콘드리아 DNA 교정과 식물에서 엽록체 DNA 교정을 보여주었다. 사람 세포와 다른 진핵세포에서 인델 (indel)이 없고, 정확한 염기 교정을 하는 징크 핑거 탈아미노효소(ZFD)를 분할된 DddAtox를 맞춤형 설계된 징크 핑거 단백질에 연결해서 만들었다. 징크 핑거 어레이 (Zinc finger array: 2 x 0.3~0.6 k base pair)는 TALE 어레이 (2 x 1.7~2k base pair)나 *S.pyogenes* Cas9(4.1 k base pair)와 비교해서 크기가 작기 때문에, 생체 내 연구 및 유전자 치료 응용을 위해 ZFD가 인코딩된 유전자는 쉽게 바이러스 제한된 탑재 공간(cargo space)를 가진 바이러스 벡터 예를 들면 AAV에 탑재할 수 있다. Tale 어레이와는 달리, 징크 핑거 어레이는 C 말단이나 N 말단 부분에 부피가 큰 도메인을 가지고 있지 않으므로 엔지니어링 친화적이다. 분할된 DddAtox 절반은 징크 핑거 단백질의 C 말단 부분 이나 N 말단 부분에 융합될 수 있다. 또한, 징크 핑거 단백질은 고유한 세포 투과 능력이 있다. 이는 사람 세포에 핵산 없이(nucleic acid-free) 유전자 교정을 할 수 있게 한다. 이러한 특성들은 징크 핑거 단백질로 하여금 DNA 결합 모듈로써 핵이나 다른 세포 소기관의 염기를 교정하는 이상적인 플랫폼이 되도록 한다.

[624]

[625] 1-1. 재료 및 방법

[626]

[627] 플라스미드 제작

[628] 포유류에서 발현을 위한 p3s-ZFD 플라스미드는 p3s-ABE7.10

플라스미드(addgene, #113128)를 HindIII 및 XhoI(NEB) 효소로 자른 후 변형하여 만들었다. 효소에 의해 잘려진 p3s 플라스미드와 합성된 삽입 DNA는 HiFi DNA 조립 키트(NEB)를 사용하여 조립되었다. MTS-, ZFP(Toolgen, Sangamo 및 Barbas 모듈)-, split DddA- 또는 UGI를 인코딩하는 모든 삽입 DNA는 IDT에 의해 합성되었다. pTarget 플라스미드는 ZFD 활성을 위한 스페이서 서열의 최적 길이를 결정하기 위해 설계되었다. 여러 길이의 스페이서와 2개의 ZFP 결합 부위를 포함하는 각 pTarget 플라스미드는 2개의 효소(EcoRI 및 BamHI, NEB)로 분해된 pRGS-CCR5-NHEJ 리포터 플라스미드에 ZFP 결합 서열과 스페이서 서열을 삽입하여 구성되었다. 대장균에서 단백질 생산을 위한 pET-ZFD 플라스미드는 pET-Hisx6-rAPOBEC1-XTEN-nCas9-UGI-NLS 플라스미드(addgene, #89508)를 NcoI 및 XhoI(NEB)로 자른 후 변형하여

생성되었다. ZFD 서열은 PCR을 이용하여 p3s-ZFD 플라스미드로부터 증폭되었고, Hisx6 태그 및 GST 태그 서열은 올리고뉴클레오티드(Macrogen)로 합성되었다. 단백질 정제를 위한 모든 플라스미드는 ZFD 및 태그를 인코딩하는 서열을 효소로 잘려진 pET 플라스미드에 HiFi DNA 어셈블리 키트(NEB)를 사용하여 삽입하여 제작하였다. 화학적으로 수용가능한 DH5 α 대장균 세포에 플라스미드 형질전환 하고, 플라스미드를 제조사의 프로토콜에 따라 AccuPrep Plasmid Mini Extraction Kit(Bioneer)로 정제했다. Sanger sequencing으로 전체 서열을 확인한 후 원하는 플라스미드를 선택하였다.

[629]

[630] HEK293T 세포 배양 및 트랜스펙션

[631] HEK 293T 세포(ATCC CRL-11268)를 10% 소태아혈청(Welgene)과 1% 항생제-항진균 용액(Welgene)이 첨가된 Dulbecco의 Modified Eagle Medium(Welgene)에서 배양했다. HEK 293T 세포(7.5×10^4)를 48-웰 플레이트에 시딩(seeding)해주었다. 18-24시간 후, 세포가 70~80% 정도로 자라게 되면 왼쪽 및 오른쪽 ZFD를 인코딩하고 있는 플라스미드(각 500ng)를 Lipofectamine 2000 (1.5 μ L, Invitrogen)을 이용해서 트랜스펙션을 해주거나, pTarget 플라스미드(10ng)와 함께 트랜스펙션해 주었다. 세포를 트랜스펙션 처리 후 96시간에 harvest한 후 세포 용해 완충액(50 mM Tris-HCl, pH 8.0(Sigma-Aldrich), 1 mM EDTA(Sigma-Aldrich), 0.005% 나트륨 도데실 설페이트(Sigma-Aldrich)), 5 μ L의 프로테이나제 K(Qiagen) 100 μ L 을 넣어준 뒤 55°C에서 1시간, 95°C에서 10분 동안 인큐베이션 해 주었다. 전체 mtDNA 시퀀싱을 하기 위해 HEK 293T 세포에 ND1 또는 ND2 타겟 mitoZFD 쌍을 인코딩하는 플라스미드 또는 mRNA를 연속적으로 희석한 농도로 트랜스펙션하였다. 구조체의 양(ng)은 7.5×10^4 당 전달된 것이다. mtDNA는 트랜스펙션 후 96시간에 세포로부터 추출되었다.

[632] K562 세포 배양 및 트랜스펙션

[633] K562 세포는 10% 소 태아 혈청(Welgene) 및 1% 항생제-항진균 용액(Welgene)이 첨가된 RPMI 1640 배양액에서 배양되었다. 전기천공(electroporation)에 의한 K562 세포로의 ZFD 전달을 위해 프로그램 FF-120(Lonza)이 있는 Amaxa 4D-Nucleofector™ X Unit 시스템을 사용했다. 16-well Nucleocuvette™ Strip을 사용할 때 각 샘플에 첨가된 기질 용액의 최대 부피는 2 μ L였다. 단백질 같은 경우 왼쪽 및 오른쪽 ZFD 각각 220 pmol(최대 용량의 경우) 또는 110 pmol(최대 용량의 절반)를, 플라스미드의 경우 왼쪽 및 오른쪽 ZFD를 인코딩하는 플라스미드 500ng을 K562 세포(1×10^5)에 트랜스펙션하였다. 처리 후 96시간에, 세포를 100g에서 5분 동안 원심분리하여 수집하고 100 μ L의 세포 용해 완충액(50 mM Tris-HCl, pH 8.0(Sigma-Aldrich), 1 M EDTA(Sigma-Aldrich), 0.005% 나트륨 도데실 설페이트(Sigma-Aldrich)), 5 μ L의 프로테이나제 K(Qiagen)를 첨가하여 55°C에서 1시간, 95°C에서 10분 동안

인큐베이션 하였다. ZFD 또는 ZFD 인코딩 플라스미드를 K562 세포로 직접 전달하기 위해 이전에 ZFN의 직접 전달에 사용된 방법을 참조했다. 왼쪽 및 오른쪽 ZFD 단백질의 혼합물(최종 농도 50 μ M) 또는 왼쪽 및 오른쪽 ZFD를 인코딩하는 플라스미드의 혼합물(각각 500ng)을 100mM L-아르기닌 및 90 μ M ZnCl₂를 함유하는 pH 7.4의 무혈청 배지로 희석하였고 최종 부피를 20 μ L로 해주었다. K562 세포(1 $\times$ 10⁵)를 100g에서 5분간 원심분리하고 상층액을 제거하였다. 그런 다음 세포를 희석된 ZFD 용액에 풀어주고 37°C에서 1시간 동안 인큐베이션했다. 인큐베이션 후, 세포를 100g에서 5분 동안 원심분리한 다음, 새로운 배양 배지를 넣어주었다. 세포를 30°C(일시적인 저체온 조건) 또는 37°C에서 18시간 동안 유지한 다음 37°C에서 이를 더 키워줬다. 일부 세포는 위의 과정에 따라 2번 처리되었다. 세포를 처리 96시간 후에 분석하였다.

[634]

[635] ZFD 단백질 발현 및 정제

[636] 각각 C-말단 GST 태그가 있는 ZFD 각 쌍을 인코딩하는 플라스미드를 Rosetta(DE3) 컴피턴트 세포로 형질전환시킨 다음, 이를 카나마이신을 함유하는 LB-한천 플레이트에서 배양하였다. 밤새 배양한 후, 단일 콜로니를 선택하고 37°C에서 50 μ g/ml 카나마이신 및 100 μ M ZnCl₂를 함유하는 액체 배지에서 밤새(전배양) 배양하였다 다음날, 전배양의 일부를 다량의 액체 배지에 옮기고, 흡광도 A_{600 nm}, = ~0.5-0.70이 될 때까지 37 °C에서 배양했다. 배양물을 약 1시간 동안 얼음 위에 두었고, 그 후 0.5mM의 이소프로필 β -D-1-티오갈락토포라노사이드(IPTG; GoldBio)를 첨가하여 ZFD 단백질 발현을 유도하고 배양물을 18°C에서 14시간 동안 인큐베이션했다.

[637] 단백질 정제 과정으로는, 용해 용액(50 mM Tris-HCl(Sigma-Aldrich), 500 mM NaCl(Sigma-Aldrich), 1 mM MgCl₂(Sigma-Aldrich), 10 mM 1,4-dithiothreitol(DTT; GoldBio), 1% Triton X-10 Sigma-Aldrich), 10% 글리세롤, 1 mM 페닐메틸설포닐 플루오라이드(Sigma-Aldrich), 1 mg/ml 닭고기 달걀 흰자위에서 추출한 라이소자임(Sigma-Aldrich), 100 μ M ZnCl₂(Sigma-Aldrich), 100mM Sigma-Aldrich), pH 8.0).으로 세포를 용해시키고, 세포를 더 용해시키기 위해서 초음파 처리(3 min total, 5 s on, 10 s off)를 해줬다. 이것들을 해준 뒤 용액은 원심분리(13,000 rpm)하여 상층액만 추출하였다. 상층액은 Glutathione Sepharose 4B (GE healthcare)를 첨가하여 1시간 동안 인큐베이션 해주었다. 인큐베이션 후, resin-lysate 혼합물은 컬럼에 부어줬고, 세척 완충액(50mM Tris-HCl (Sigma-Aldrich), 500mM NaCl (Sigma-Aldrich), 10mM DTT (GoldBio), 1 mM MgCl₂ (Sigma-Aldrich), 100 μ M ZnCl₂ (Sigma-Aldrich), 10% glycerol, 100 mM arginine (Sigma-Aldrich), pH 8.0)으로 3번 씻어줬다. 레진 (resin)에 붙어 있는 단백질은 용출버퍼(50mM Tris-HCl (Sigma-Aldrich), 500mM NaCl (Sigma-Aldrich), 1mM MgCl₂ (Sigma-Aldrich), 40 mM glutathione (Sigma-Aldrich), 10% glycerol, 1mM DTT (GoldBio), 100 μ M ZnCl₂ (Sigma-Aldrich), 100 mM

arginine (Sigma-Aldrich), pH 8.0)로 resin으로부터 분리해주었다. 마지막으로, 추출된 단백질은 ~15 ng/ μ L(200-240 pmol/ μ L, 단백질 크기에 따라서)의 농도로 농축되었다.

[638]

[639] ZFD에 의한 PCR 앰플리콘의 시험관내 탈아미노화

[640] TRAC 부위를 포함하는 앰플리콘은 PCR을 사용하여 제조되었다. 8 μ g의 앰플리콘을 37 °C에서 1-2시간 동안 100 μ M ZnCl₂를 포함하는 NEB3.1 완충액에서 2 μ g의 각 ZFD 단백질(Left-G1397N 및 Right-G1397C)과 함께 인큐베이션 했다. 반응 후, 4 μ L의 Proteinase K 용액(Qiagen)을 넣고 55°C에서 30분 동안 인큐베이션 하여 ZFD를 제거하였고, 앰플리콘은 PCR 정제 키트(MGmed)를 사용하여 정제되었다. 1 μ g의 정제된 앰플리콘을 37°C에서 1시간 동안 2개의 USER 효소(NEB)와 함께 인큐베이션했다. 그런 다음, 앰플리콘을 4 μ L의 프로테아제 K 용액(Qiagen)과 함께 인큐베이션하고 PCR 정제 키트(MGmed)를 사용하여 다시 정제하였다. 정제된 PCR을 아가로스 겔에서 전기영동하고 이미지화하였다.

[641]

[642] 표적 딥 시퀀싱.

[643] 표적 및 오프타겟 부위의 염기교정 비율 분석을 위해 표적 부위는 중첩된 1차 PCR, 2차 PCR 증폭 및 TruSeq HT Dual 인덱스 함유 프라이머를 사용한 3차 PCR을 PrimeSTAR® GXL DNA 중합효소(TAKARA)를 사용하여 딥 시퀀싱 라이브러리를 생성했다. 라이브러리는 Illumina MiniSeq를 사용하여 paired-end 시퀀싱 하였다.

[644]

[645] mRNA 준비

[646] ZFD 서열의 업스트림에 T7 RNA 중합효소 프로모터를 포함하는 DNA 주형은 정방향 및 역방향 프라이머(정방향: 5'-CATCAATGGGCGTGGATAG-3' SEQ ID No: 116, 역방향: 5'-CATCAATGGGCGTGGATAG-3' SEQ ID No: 117, 역방향: 5'-GACACCTACTCAGACAATGC-3' SEQ ID No: 118)에 의해서 PCR 제작되었다. mRNA는 mMESSAGE mMACHINE™ T7 ULTRA 전사 키트(Thermo Fisher)를 사용하여 시험관 내에서 합성되었다. 시험관 내 전사된 mRNA는 제조사의 프로토콜에 따라 MEGAclean™ Transcription Clean-Up Kit(Thermo Fisher)를 사용하여 정제되었다.

[647]

[648] 전체 미토콘드리아 게놈 시퀀싱

[649] 전체 미토콘드리아 게놈 시퀀싱의 경우 세 단계가 필요하다. 1. 분리된 미토콘드리아에서 mtDNA 추출: 먼저, 3×10⁵ HEK293T 세포를 트립신 처리하고 ND1 또는 ND2 표적 mitoZFD 쌍으로 트랜스펙션 후 96시간 후에 원심분리(500g, 4분, 4°C)에 의해 수집하였다. 그 다음, 세포를 인산염 완충 식염수(Welgene)로

세척하고 원심분리하여 다시 수집하였다. 상층액을 제거하고 제조사의 프로토콜에 따라 Mitochondria Isolation Kit for Cultured Cells(Thermo Fisher)의 시약 기반 방법을 사용하여 배양된 세포에서 미토콘드리아를 분리했다. mtDNA는 DNeasy Blood & Tissue Kit(Qiagen)를 사용하여 분리된 미토콘드리아에서 추출되었다. 2. NGS 라이브러리 생성: 추출된 mtDNA에서 NGS 라이브러리를 생성하기 위해 Nextera™ DNA CD Indexes(Illumina)가 포함된 Illumina DNA Prep 키트를 사용했다. 3. NGS: 라이브러리를 풀링하고 MiniSeq 시퀀서(Illumina)에 로드했다. 평균 시퀀싱 depth는 50보다 컸다.

[650]

[651] 미토콘드리아 게놈 전체 DNA 교정 분석

[652] 전체 미토콘드리아 게놈 시퀀싱의 NGS 데이터를 분석하기 위해 BWA를 사용하여 Fastq 파일을 GRCh38.p13(릴리스 v102) 참조 게놈에 정렬하고 읽기 페어링 정보 및 플래그를 수정하여 SAMtools(v.1.9)로 BAM 파일을 생성했다. 그런 다음 REDIttools(v.1.2.1)의 REDIttoolDenovo.py 스크립트를 사용하여 미토콘드리아 게놈의 모든 시토신 및 구아닌 중에서 염기 교정율이 1% 이상인 위치를 식별했다. 세포주의 염기 교정율이 50% 이상인 위치를 단일 뉴클레오타이드 변이로 간주하여 모든 샘플에서 제외했다. off-target 분석을 위해서 각 ZFD의 표적 부위를 제외했다. 교정 빈도가 $\geq 1\%$ 인 나머지 위치를 off-target으로 간주하고 교정된 C/G 뉴클레오타이드의 수를 계산했다. 미토콘드리아 게놈의 offtarget의 평균 C/G에서 T/A 염기 교정 빈도를 계산하기 모든 C/G에 대해서 평균화했다. 특이성 비율은 on-target 평균 수치를 off-target 평균 수치로 나뉘어서 계산하였다. 미토콘드리아 게놈 전체 그래프는 표적 및 표적을 벗어난 위치에서 염기 교정율 표시하여 생성되었다.

[653]

[654] 1-2. ZFD 구조의 최적화

[655] 인간 및 다른 진핵세포에서 염기 교정을 하기 위한 ZFD를 개발하기 위해서, 먼저 반쪽 DddAtox에 붙는 징크 핑거 단백질(ZFP)의 아미노산 링커와 스페이서(spacer)의 길이를 최적화 시키려고 했다. 왼쪽과 오른쪽 ZFP가 붙는 곳 사이인 스페이서에서 C에서 T로의 염기 전환이 일어난다. 잘 특성화된 인간 CCR5 유전자를 타겟하는 ZFN 쌍을 골랐다. 이를 이용해서 2, 5, 10, 16, 24 및 32의 여러 가지 아미노산 링커를 갖는 ZFD를 만들었고, ZFD의 왼쪽과 오른쪽 ZFP 결합부위가 있고 반복되는 TC서열이 존재하는 1부터 24 base pair의 길이를 갖는 다양한 스페이서가 있는 일련의 타겟 플라스미드를 만들었다(도 1a, b 및 표 1.).

[656]

[657] [표1]

pTarget 라이브러리 서열

Spacer Length (bp)	Left-ZFD DNA binding region	Spacer sequence	Right-ZFD DNA binding region
1	GTCATCCTCATC (SEQ ID NO.: 119)	C	AAACTGCAAAAG (SEQ ID NO.: 120)
2	GTCATCCTCATC	TC	AAACTGCAAAAG
3	GTCATCCTCATC	TCG (SEQ ID NO.: 121)	AAACTGCAAAAG
4	GTCATCCTCATC	CTGA (SEQ ID NO.: 122)	AAACTGCAAAAG
5	GTCATCCTCATC	CTGAT (SEQ ID NO.: 123)	AAACTGCAAAAG
6	GTCATCCTCATC	CTGATC (SEQ ID NO.: 124)	AAACTGCAAAAG
7	GTCATCCTCATC	CTGATCG (SEQ ID NO.: 125)	AAACTGCAAAAG
8	GTCATCCTCATC	CTGATCGA (SEQ ID NO.: 126)	AAACTGCAAAAG
9	GTCATCCTCATC	CTGATCGAT (SEQ ID NO.: 127)	AAACTGCAAAAG
10	GTCATCCTCATC	CTGATCGATC (SEQ ID NO.: 128)	AAACTGCAAAAG
11	GTCATCCTCATC	CTGATCGATCG (SEQ ID NO.: 129)	AAACTGCAAAAG
12	GTCATCCTCATC	CTGATCGATCGA (SEQ ID NO.: 130)	AAACTGCAAAAG
13	GTCATCCTCATC	CTGATCGATCGAT (SEQ ID NO.: 131)	AAACTGCAAAAG
14	GTCATCCTCATC	CTGATCGATCGATC (SEQ ID NO.: 132)	AAACTGCAAAAG
15	GTCATCCTCATC	CTGATCGATCGATCG (SEQ ID NO.: 133)	AAACTGCAAAAG
16	GTCATCCTCATC	CTGATCGATCGATCGA (SEQ ID NO.: 134)	AAACTGCAAAAG
17	GTCATCCTCATC	CTGATCGATCGATCGAT (SEQ ID NO.: 135)	AAACTGCAAAAG
18	GTCATCCTCATC	CTGATCGATCGATCGATC (SEQ ID NO.: 136)	AAACTGCAAAAG
19	GTCATCCTCATC	CTGATCGATCGATCGATCG (SEQ ID NO.: 137)	AAACTGCAAAAG
20	GTCATCCTCATC	CTGATCGATCGATCGATCGA (SEQ ID NO.: 138)	AAACTGCAAAAG
21	GTCATCCTCATC	CTGATCGATCGATCGATCGAT (SEQ ID NO.: 139)	AAACTGCAAAAG
22	GTCATCCTCATC	CTGATCGATCGATCGATCGATC (SEQ ID NO.: 140)	AAACTGCAAAAG
23	GTCATCCTCATC	CTGATCGATCGATCGATCGATCG (SEQ ID NO.: 141)	AAACTGCAAAAG
24	GTCATCCTCATC	CTGATCGATCGATCGATCGATCGA (SEQ ID NO.: 142)	AAACTGCAAAAG

[658]

[659] DddAtox는 두 위치(G1333 및 G1397)에서 분할될 수 있으며 각 절반은 왼쪽 또는 오른쪽 ZFP에 융합될 수 있다. 생성된 24개 (=6개 linker x 2개 분할위치 x ZFP 융합위치(왼쪽 또는 오른쪽)) ZFD 구성의 염기 교정 효율을 각 24개의 스페이서를 갖는 타겟 플라스미드 대해서 측정했다. Hek293T 세포에서 형질감염 (transfection) 4일차에 딥 시퀀싱 (deep sequencing)을 통해서 측정했다.

[660] [구조]

[661] * Left-ZFD: SV40 NLS-ZFP(S162-left)-linker-DddAtox half-4aa linker-UGI

[662] * Right-ZFD: SV40 NLS-ZFP(S162-right)-linker-DddAtox half-4aa linker-UGI

[663] - SV40 NLS:

[664] PKKKRKV (SEQ ID No: 478)

[665]

[666] - ZFP(S162-left)

[667] GIHGVPAAAMAERPFQCRICMRNFSDRSNLSRHIRTHTGKPFACDICGRKFAI

SSNLNSHTKIHTGSQKPFQCRICMRNFSRSDNLARHIRTHTGEKPFACDICGRKF
ATSGNLTRHTKIHRLR (SEQ ID NO: 2)

[668]

[669] - ZFP(S162-right)

[670] GIHGVPAAAMAERPFQCRICMRNFSRSDNLSVHIRTHTGEKPFACDICGRKFAQ
KINLQVHTKIHTGEKPFQCRICMRNFSRSDVLSEHIRTHTGEKPFACDICGRKFA
QRNHRRTHTKIHRLR (SEQ ID NO: 3)

[671]

[672] - 징크 핑거 단백질과 DddAtox half 사이의 링커:

[673] - 2aa: GS

[674] - 5aa: TGEKP (SEQ ID No: 479)

[675] - 10aa: SGAQGSTLDF (SEQ ID No: 9)

[676] - 16aa: SGSETPGTSESATPES (SEQ ID No: 10)

[677] - 24aa: SGTPHEVG VYTL SGTPHEVG VYTL (SEQ ID No: 115)

[678] - 32aa: GSGGSSGGSSGSETPGTSESATPESGSGSSGGS (SEQ ID No: 11)

[679]

[680] - Split-DddAtox G1333-N

[681] GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGG (SEQ ID
No: 27)

[682]

[683] - Split-DddAtox G1333-C

[684] PTPYPNYANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMETLL
PENAKMTVVPPEGAIPV KRGATGETKVFTGNSNSPKSPTKGGC (SEQ ID No:
272)

[685]

[686] - Split-DddAtox G1397-N

[687] GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGGPTYPNYA
NAGHVEGQSALFMRDN
GISEGLVFHNNPEGTCGFCVNMETLLPENAKMTVVPPEG (SEQ ID No: 273)

[688]

[689] - Split-DddAtox G1397-C

[690] AIPVKRGATGETKVFTGNSNSPKSPTKGGC (SEQ ID No: 26)

[691]

[692] - 4aa linker

[693] SGGS (SEQ ID NO: 480)

[694]

[695] - UGI

[696] TNLSDIIEKETGKQLVIQESILMLPEEVVEEVIGNKPESDILVHTAYDESTDENV

MLLTSDAPEYKPWA LVIQDSNGENKIKML (SEQ ID NO: 481)

[697]

[698] 짧은 링커를 갖는 ZFD(2 및 5 아미노산(AA) 링커)는 효율이 낮거나 없다. 반면에 10AA 이상의 링커를 갖는 ZFD에서는 4 base pair 이상의 스페이서에서 C에서 T로의 염기 교정 효율이 1%~24% 범위로 나왔다(도 1c 와 도 2a,c). ZFD쌍 중 24AA 링커를 가지는 것이 가장 높은 교정 효율을 보여주었다. 가장 좋은 링커 조합을 찾기 위해, ZFD쌍의 왼쪽 ZFD에 24AA 링커를 고정하고, 오른쪽을 다양한 링커 길이를 가지는 ZFD로 조합을 하여 염기 교정 효율을 보았고, 똑같이 반대로도 해봤다(도 1d 및 도 3). 양쪽 모두 24AA 링커를 가지는 것이 가장 효율이 좋다는 것을 알 수 있었다. 또한 DddAtox가 G1397부분에서 분할되었을 때가 G1333 부분에서 분할되었을 때보다 더 좋은 효율을 가진다는 것을 발견하였다(도 1c와 도 2a,b). 이러한 가장 높은 효율을 가지는 ZFD쌍은 7~21bp의 스페이서에서 >6.7%의 높은 효율로 사이토신을 교정한다(도 1c 및 도 2a, c).

[699]

[700] 1-3. 생체 내 핵 DNA 타겟에서 염기 교정

[701]

[702] 다음으로 사람 세포에서 24AA 링커를 갖는 ZFD가 생체내 염색체 타겟 위치에서도 C에서 T로의 염기 교정을 촉매할 수 있는지 조사하였다. 총 8개 유전자 안에 있는 11곳(2개 ZFD쌍이 한 곳)을 타겟하는 22쌍의 ZFD가 제작되었다(도 4). 그 중에서, 14쌍의 ZFD는 공개적으로 사용가능한 징크 핑거 리소스로부터 가져다 사용하였다. 이전에 특성화된(previously-characterized) ZFN (CCR5 및 TRAC에 특정한)을 개조하여 다른 8쌍의 ZFD를 만들었다. 이유는 ZFN 같은 경우는 스페이서 길이가 5~7bp에서 타겟 DNA를 자르는 반면에 ZFD는 최소 7 이상인 스페이서 작동하기 때문에, 이러한 ZFN쌍에서 하나 또는 두 개의 징크 핑거를 떼고 붙이며 작동할 수 있는 ZFD들을 만들었다. FokI 뉴클레아제는 부분은 ZFP에 N-말단 이나 C-말단 부분에 융합될 수 있어 4개의 다른 구성을 가지는 ZFN을 만들 수 있기 때문에, 또한 다른 구성을 가지는 2쌍의 ZFD (도 4b에서 Trac-NC) 만들어 분할된 DddAtox 반쪽들도 기존의 ZFP C-말단뿐만 아니라 ZFP의 N-말단 부분에도 융합될 수 있는지 테스트하였다(NC 구성은 도 4a 및 도 5에 도시).

[703] [구조]

[704] * C type: SV40 NLS-Zinc finger protein-24aa linker-DddA_{tox} half-4aa linker-UGI

[705] * N type: SV40 NLS-DddA_{tox} half-24aa linker-Zinc finger protein-4aa linker-UGI

[706]

[707] - SV40 NLS

[708] PKKKRKV (SEQ ID No: 478)

[709]

- [710] - 24aa linker
- [711] SGPHEVGVYTLSTPHEVGVYTL
- [712]
- [713] - Split-DddA_{tox} G1397-N
- [714] GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSGGPTYPNYA
NAGHVEGQSALFMRDN
GISEGLVFHNNPEGTCGFCVNMETLLPENAKMTVVPPEG
- [715]
- [716] - Split-DddA_{tox} G1397-C
- [717] AIPVKRGATGETKVFTGNSNSPKSPTKGGC
- [718]
- [719] - 4aa linker
- [720] SGGS
- [721]
- [722] - UGI
- [723] TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENV
MLLTSDAPEYKPWA LVIQDSNGENKIKML
- [724]
- [725]
- [726] - ZFP
- [727] CCR5-1 Left (C type) [S162 ZFN-Left]
- [728] GIHGVPAAAMAERPFQCRICMRNFSDRSNLSRHIRTHTGEKPFACDICGRKFAI
SSNLNSHTKIHTGSQKPFQCRICMRNFSRSNLARHIRTHTGEKPFACDICGRKF
ATSGNLTRHTKIHLR
- [729]
- [730] CCR5-1 Right (C type) [S162 ZFN-Right]
- [731] GIHGVPAAAMAERPFQCRICMRNFSRSNLSVHIRTHTGEKPFACDICGRKFAQ
KINLQVHTKIHTGEKPFQCRICMRNFSRSNLSEHIRTHTGEKPFACDICGRKFA
QRNHRTTHTKIHLR
- [732]
- [733] CCR5-2 Left (C type) [S162 ZFN-Left]
- [734] GIHGVPAAAMAERPFQCRICMRNFSDRSNLSRHIRTHTGEKPFACDICGRKFAI
SSNLNSHTKIHTGSQKPFQCRICMRNFSRSNLARHIRTHTGEKPFACDICGRKF
ATSGNLTRHTKIHLR (SEQ ID NO: 2)
- [735]
- [736]
- [737] CCR5-2 Right (C type) [Modifying S162 ZFN-Right with additional ZF using the
Barbas set of zinc finger modules]

- [738] GIHGVPAAMAERPFQCRICMRNFSQSGDLRRHIRTHTGEKPFACDICGRKFA
RSDNLSVHTKIHTGSQKPFQCRICMRNFSQKINLOVHIRTHTGEKPFACDICGR
 KFARSDVLSEHTKIHLR (SEQ ID NO: 482)
- [739]
- [740] TRAC-Left (C type) [Adapted from Paschon, D.E. et al., 2019]
- [741] GIHGVPAAMAERPFQCRICMRNFSDQSNLRAHIRTHTGEKPFACDICGRKFA
TSSNRKTHTKIHTGSQKPFQCRICMRNFSLQOTLADHIRTHTGEKPFACDICGR
 KFAQSGNLARHTKIHLR (SEQ ID NO: 483)
- [742]
- [743] TRAC-Left (N type) [Adapted from Paschon, D.E. et al., 2019]
- [744] FQCRICMRKFATSGSLTRHTKIHTGEKPFQCRICMRNFSRSDHLSTHIRTHTGE
 KPFACDICGRKFATTSSNRKTHTKIHTHPRAPIKPFQCRICMRNFSRSDNLSEHI
 RTHTGEKPFACDICGRKFAWHSSLRVHTKIHLR (SEQ ID NO: 484)
- [745]
- [746] TRAC-Right (C type) [From Paschon, D.E. et al., 2019]
- [747] GIHGVPAAMAERPFQCRICMRNFSRSDHLSTHIRTHTGEKPFACDICGRKFAD
RSHLARHTKIHTGSQKPFQCRICMRKFALKOHLNEHTKIHTGEKPFQCRICMR
 NFSQSGNLARHIRTHTGEKPFACDICGRKFAHNSSLKDHTKIHLR (SEQ ID NO:
 485)
- [748]
- [749] MFAP1 Left (C type) [De novo designed using the Toolgen set of zinc finger
 modules]
- [750] GIRIPGEKPYSCGICGKSFSDSSAKRRHCILHTGEKPYTCSDCGKAFRDKSCLN
RHRRHTHTGEKPYKMECGKAFNRRLSHLTRHQRIHTGEKPYECNYCGKTFVSS
TLIRHQRIHLR (SEQ ID NO: 486)
- [751]
- [752] MFAP1 Right (C type) [Modifying S162 ZFN-Right with additional ZF using the
 Barbas set of zinc finger modules]
- [753] GIRERPYPVESCRRFSTSGSLVRHRIHTGQKPFQCRICMRNFSRSDDELTR
 HIRTHTGEKPFACDICGRKFARSDHLTTHTKIHTGEKPFQCRICMRKFAQSSNL
VRHTKIHLR (SEQ ID NO: 487)
- [754]
- [755] CCDC28B Left (C type) [Modifying S162 ZFN-Right with additional ZF using the
 Barbas set of zinc finger modules]
- [756] GIRERPYPVESCRRFSDPGHLVRHRIHTGQKPFQCRICMRNFSRSDDELTR
 HIRTHTGEKPFACDICGRKFARSDHLTTHTKIHTGEKPFQCRICMRKFARSDKL
VRHTKIHLR (SEQ ID NO: 488)
- [757]

- [758] CCDC28B Right (C type) [De novo designed using the Toolgen set of zinc finger modules]
- [759] GIRIPGEKPYECNYCGKTFSVSSTLIRHQRIHTGEKPYHCDWDGCGWKFARSDELTRHYRKHTGEKPFECKDCGKAFIQKSNLIRHQRTHTGEKPYHCDWDGCGWKFARSDELTRHYRKHLR (SEQ ID NO: 489)
- [760]
- [761] KDM4B Left (C type) [Modifying S162 ZFN-Right with additional ZF using the Barbas set of zinc finger modules]
- [762] GIRERPYPVESCRRFSDCDRLARHIRIHTGQKPFQCRICMRNFSRSDELTRHIRTHTGEKPFACDICGRKFARSDHLTTHTKIHTGEKPFQCRICMRKFARSDKLVRHTKIHLR (SEQ ID NO: 490)
- [763]
- [764] KDM4B Right (C type) [De novo designed using the Toolgen set of zinc finger modules]
- [765] GIRIPGEKPFECKDCGKAFIQKSNLIRHQRTHTGEKPYRCEECGKAFRWPSNLTRHKRIHTGEKPYRCEECGKAFRWPSNLTRHKRIHTGEKPYSCGICGKSFSDSSAKRRHCILHLR (SEQ ID NO: 491)
- [766]
- [767] NUMBL Left (C type) [Modifying S162 ZFN-Right with additional ZF using the Barbas set of zinc finger modules]
- [768] GIRERPYPVESCRRFSDCDRLARHIRIHTGQKPFQCRICMRNFSRSDELTRHIRTHTGEKPFACDICGRKFARSDHLTTHTKIHTGEKPFQCRICMRKFARSDKLVRHTKIHLR (SEQ ID NO: 492)
- [769]
- [770] NUMBL Right (C type) [De novo designed using the Toolgen set of zinc finger modules]
- [771] GIRIPGEKPYTCSDCGKAFRDKSCLNRHRRTHHTGEKPYKCGQCGKFYSQVSHLTRHQKIHTGEKPFECKDCGKAFIQKSNLIRHQRTHTGEKPYHCDWDGCGWKFARSDELTRHYRKHLR (SEQ ID NO: 493)
- [772]
- [773] INPP5D-1 Left (C type) [Modifying S162 ZFN-Right with additional ZF using the Barbas set of zinc finger modules]
- [774] GIRERPYPVESCRRFSDKLVRHIRIHTGQKPFQCRICMRNFSRSDELTRHIRTHTGEKPFACDICGRKFARSDHLTTHTKIHTGEKPFQCRICMRKFARSDKLVRHTKIHLR (SEQ ID NO: 494)
- [775]
- [776] INPP5D-1 Right (C type) [De novo designed using the Toolgen set of zinc finger modules]

- [777] GIRIPGEKPYTCSDCGKAFRDDKSCLNRHRRTHTGEKPYECNYCGKTFSVSSTL
IRHQRIHTGEKPYTCSDCGKAFRDDKSCLNRHRRTHTGEKPYTCSDCGKAFRDDK
SCLNRHRRTHLR (SEQ ID NO: 495)
- [778]
- [779] INPP5D-2 Left (C type) [Modifying S162 ZFN-Right with additional ZF using the Barbas set of zinc finger modules]
- [780] GIRERPACPVESCDRRFSRSDKLVHRHIRTGQKPFQCRICMRNFSRSDELTR
HIRTHTGEKPFACDICGRKFARSDHLLTHTKIHTGEKPFQCRICMRKFARSDKL
VRHRTKIHLR (SEQ ID NO: 496)
- [781]
- [782] INPP5D-2 Right (C type) [De novo designed using the Toolgen set of zinc finger modules]
- [783] GIRIPGEKPYECNYCGKTFSVSSTLIRHQRIHTGEKPYTCSDCGKAFRDDKSCLN
RHRRTHTGEKPYTCSDCGKAFRDDKSCLNRHRRTHTGEKPYHCDWDGCGWKF
ARSDELTRHYRKH (SEQ ID NO: 497)
- [784]
- [785] DVL3 Left (C type) [De novo designed using Barbas zinc finger modules]
- [786] GIHGVPAAEAERPFQCRICMRNFSTSGHLVRHIRTHTGEKPFACDICGRKFAT
SGHLVRHRTKIHTGEKPFQCRICMRNFSTSGELVRHIRTHTGEKPFACDICGRKF
AQSSNLVRHTKIHLR (SEQ ID NO: 498)
- [787]
- [788] DVL3 Right (C type) [S162 ZFN-left]
- [789] GIHGVPAAEAERPFQCRICMRNFSDRSNLSRHIRTHTGEKPFACDICGRKFAI
SSNLNSHTKIHTGSQKPFQCRICMRNFSRSDNLARHIRTHTGEKPFACDICGRKF
ATSGNLTRHTKIHLR (SEQ ID NO: 499)
- [790]
- [791] Hek 293T 세포에서, NC 구성을 가지는 ZFD를 포함해서 ZFD의 C에서 T로의 염기 교정 효율은 1.0%~60%를 보인다. 반면에, 삽입-결실(indel)은 <0.4% 정도로 거의 일어나지 않았다(도 4b와 도6). 위에서 설명한 플라스미드 기반으로 한 실험에서 봤듯이, 스페이서 5bp를 가지는 CCR5타겟하는 ZFD는 효율이 굉장히 떨어졌다. 스페이서 길이가 최소 7이상인 타겟에서 다른 20개의 ZFD쌍은 평균 교정 효율이 12.0±3.4%을 보였으며, 이는 평균 8.3±2.2%을 보이는 cas9 기반 염기 교정 유전자 (cas9-derived Base Editor2)와 견줄 만하다. 추가로 C에서 T로의 염기교정은 TC context뿐만 아니라 AC나 GCC에서도 일어났다(도 4c~f). AC는 NUMBL의 C₆에서, GCC는 INPP5D-2의 C₇에서 각각 C염기 교정 효율이 4.58% 와 1.85%까지 나타났다.

[792]

	Left-ZFD binding region	Spacer region	Right-ZFD binding region	ZFD configuration
MFAP1	5'- GACGGCCCCAGC 3'- CTGCCGGGGTCG	CGTAGACT GCATCTGA	GAATGGGCGGTT -3' CTTACCCGCCAA -5'	C-C
CCDC28B	5'- TGCCGCCCAGTC 3'- ACGGCGGGTCAG	CTGGCTGGAG GACCGACCTC	GTGGAAGCGGCT -3' CACCTTCGCCGA -5'	C-C
KDM4B	5'- GGCCGCCCACCC 3'- CCGGCGGGTGGG	GGTCCCCACT CCAGGGGTGA	GTCGGTGGTGAA -3' CAGCCACCACTT -5'	C-C
NUMBL	5'- GGCCGCCCACCC 3'- CCGGCGGGTGGG	CCAGACAG GGTCTGTC	GCGGAAGGAGCC -3' CGCCTTCTCTCGG -5'	C-C
INPP5D-1	5'- CCCC GCCCACCC 3'- GGGGCGGGTGGG	CGACGCCGCG GCTGCGGCGC	GCCGCCGCTGCC -3' CGGCGGCGACGG -5'	C-C
INPP5D-2	5'- CCCC GCCCACCC 3'- GGGGCGGGTGGG	CGACGCC GCTGCGG	GCGGCCGCGCT -3' CGCCGGCGGCGA -5'	C-C
DVL3	5'- ACCACCAGTTT 3'- TGGTGGTCGAAG	TTTGACTCA AAACTGAGT	GATGAGGATGAC -3' CTACTCCTACTG -5'	C-C
CCR5-1	5'- GTCATCCTCATC 3'- CAGTAGGAGTAG	CTGAT GACTA	AAACTGCAAAAG -3' TTTGACGTTTTT -5'	C-C
CCR5-2	5'- GTCATCCTCATC 3'- CAGTAGGAGTAG	CTGATAAA GACTATTT	CTGCAAAAGGCT -3' GACGTTTTCCGA -5'	C-C
TRAC-CC	5'- GTGATTGGGTTC 3'- CACTAACCCAAG	CGAATCCTCCT GCTTAGGAGGA	CCTGAAAGTGGCCGG -3' GGACTTTCACCGGCC -5'	C-C
TRAC-NC	5'- TGTCAGTGATTGGGTT 3'- ACAGTCACTAACCCAA	CCGAATCCTCCT GGCTTAGGAGGA	CCTGAAAGTGGCCGG -3' GGACTTTCACCGGCC -5'	N-C

[793]

[794]

1-4. 정제된 ZFD 단백질을 인간 세포에 직접 전달

[795]

[796]

유전자 교정 단백질을 인코딩하고 있는 플라스미드 DNA를 전달하는 것 대신에 정제된 유전자 교정 단백질 자체를 세포에 전달하는 것이 오프-타겟 (off-target) 영향을 줄이고, 외부 DNA에 의한 선천적 면역 반응 (innate immune response) 피하며, 외부 플라스미드 DNA의 생체 내 게놈에 삽입해 들어가는 것을 방지할 수 있다. 다른 그룹은 in vitro 및 in vivo에서 ZFP가 포유동물 세포를 자발적으로 통과해 들어갈 수 있다는 것을 보였다. ZFD의 단백질 매개 염기 교정을 증명하기 위해, 효율이 높았던 TRAC 유전자를 타겟하는 ZFD 쌍을 고르고, 1개 또는 4개의 NLS (nuclear localization signal)가 달린 ZFD 재조합

단백질을 대장균에서부터 정제하였다. 첫째로 ZFD 단백질의 염기 교정 효율을 TRAC site를 가지고 있는 PCR 앰플리콘 (amplicon)을 사용해서 in vitro에서 실험을 하였고, 효율이 굉장히 높다는 것을 확인하였다. 효율은 Uracil DNA glycosylase 와 DNA glycosylase-lyase Endonuclease VIII의 혼합물인 우라실 특이적 절단 시약(USER)을 통해서 유전자가 잘리는 것을 통해 확인할 수 있었다(도 7). 형질감염하기 어려운 세포인 인간 백혈병 세포 K562 세포에 TRAC-NC ZFD 단백질을 두 가지 방식, electroporation 또는 electroporation 없이 직접전달(direct delivery)으로 전달하였다. ZFD 단백질은 굉장히 효율적이었다. C에서 T로의 염기 교정을 26.5%(electroporation) 및 17%(direct delivery)을 보였다(도 4g). 종합해보면, 이러한 결과들은 ZFD를 인코딩하는 플라스미드 또는 정제된 제조합 ZFD 단백질이 사람 세포의 핵 DNA를 염기 교정하는데 사용될 수 있다는 것을 보여준다.

[797]

[798] 1-5. ZFD의 미토콘드리아 DNA 염기교정

[799]

[800] 크리스퍼 기반 시스템과는 달리, 맞춤 설계된 (custom-designed) DNA 결합 단백질에 분할 DddAtox를 융합한 시스템의 큰 이점은 이런 프로그램 가능한 염기 교정 가위가 미토콘드리아 DNA와 같은 세포 소기관 DNA를 교정하는데 이용될 수 있다는 것이다. ZFD를 미토콘드리아에 전달하기 위해, MTS (mitochondrial targeting signal)과 NES (nuclear export signal)를 미토콘드리아 유전자를 타겟하도록 설계된 9개의 ZFD들 N말단 부분에 연결하여 mitoZFD들을 제작하였다(도 8). ZFD의 ZFP부분은 공개적으로 사용가능한 zinc finger 리소스에서 가져왔다. ZFD는 스페이서 길이를 7~15bp 가질 수 있도록 제작하였고, DNA 결합부위는 왼쪽과 오른쪽 모두 12bp가 되도록 하였다.

[801] [구조]

[802] * C type: MTS-FLAG tag-NES-Zinc finger protein-24aa linker-DddAtox half-4aa linker-UGI

[803] * N type: MTS-HA tag-NES-DddAtox half-24aa linker-Zinc finger protein-4aa linker-UGI

[804]

[805] - MTS (Mitochondrial Targeting Sequence of human mitochondrial ATP synthase F1 β subunit)

[806] MLGFVGRVAAAPASGALRRLTPSASLPPAQLLLRAAPTAVHPVRDYAAQ
(SEQ ID No:274)

[807]

[808] - FLAG tag (C type)

[809] DYKDDDDK (SEQ ID No:275)

[810]

- [811] - HA tag (N type)
- [812] YPYDVDPYA (SEQ ID No:276)
- [813]
- [814] - NES (Nuclear export signal)
- [815] VDEMTKKF (Minute virus if mice; MVM NES)
- [816]
- [817] - Split-DddAtox G1397-N
- [818] GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSGGPTYPNYA
NAGHVEGQSALFMRDN
GISEGLVFHNNPEGTCGFCVNMETLLPENAKMTVVPPEG
- [819]
- [820] - Split-DddAtox G1397-C
- [821] AIPVKRGATGETKVFTGNSNSPKSPTKGGC
- [822]
- [823] - 4aa linker
- [824] SGGS
- [825]
- [826] - UGI
- [827] TNLSDIIEKETGKQLVIQESILMLPEEVVEEVIGNKPESDILVHTAYDESTDENV
MLLTSDAPEYKPWA LVIQDSNGENKIKML
- [828]
- [829] - ZFP
- [830] * 징크핑거들은TGEKP 링커들로 연결되어 있다.
(ZF1-linker-ZF2-linker-ZF3-linker-ZF4).
- [831] * ND2-타겟 mitoZFD는 QQ 변이체를 가지고 있다. 이 변이체는 R또는 K
아미노산을 Q로 치환한 것이다. 빨간 글자로 표시되어 있다.

[832]

	Type	ZF1	No.	ZF2	No.	ZF3	No.	ZF4	No.
ND1_Left	C	FQORCM RNFSDSG NLRVHIRT H	500	YKCPDGG KSFSOSS SLRHQRT H	518	YECDHCG KSFSOSS HLNVHKR TH	536	YRCKYCD RSFSISSN LQRHVRNI H	554
ND1_Right	N	YKPECG KSFSSTN SLTEHQT H	501	YKPECG KSFSKK ALTEHQT H	519	YECNVOG KTFVSST LRHQBH H	537	YRCKYCD RSFSISSN LQRHVRNI H	555
ND2_Left	C	YKPECG KSFSRED NUHQT H	502	YHCDWD GCGWKF ARSDLT RHYRKH H	520	YKPECG KSFSRED NUHQT H	538	YRCKYCD RSFSISSN LQRHVRNI H	556
ND2_Right	N	YECDHCG KSFSOSS HLNVHKR TH	503	YHCDWD GCGWKF ARSDLT RHYRKH H	521	YKCPDGG KSFSOSS SLRHQRT H	539	YKCGQCG KFYSQVS HLTRHOKI H	557
Cox1_Left	C	YECDHCG KSFRQST HLTHRR H	504	YSCGOG KSFSOSS AKRRHCL H	522	YECDHCG KSFRQST HLTHRR H	540	YKCPDGG KSFSOSS SLRHQRT H	558
Cox1_Right	N	YKCPDGG KSFSOSS SLRHQRT H	505	YKCGQCG KFYSQVS HLTRHOKI H	523	YKCPDGG KSFSOSS SLRHQRT H	541	YECDHCG KSFSOSS HLNVHKR TH	559
Cox2_Left	C	YKPECG KSFSDPG HLVRHQR TH	506	YRCKYCD RSFSISSN LQRHVRNI H	524	YRCKYCD RSFSISSN LQRHVRNI H	542	YKPECG KSFSRED NUHQT H	560
Cox2_Right	N	YGCILCG KAFSKSS NLRRHEMI H	507	FQCKTCQ RKFSRSD HLKTHRT H	525	YKCGQCG KFYSQVS HLTRHOKI H	543	YKPECG KSFSRED NUHQT H	561
ND4_Left	C	YKPECG KSFSDPG HLVRHQR TH	508	YKPECG KSFSRED NUHQT H	526	YKPECG KSFSRED NUHQT H	544	YRCKYCD RSFSOSS NLQRHVR NIH	562
ND4_Right	C	YKPECG KSFSDPG HLVRHQR TH	509	FHCGYCE KSFSVKD YLTKHRT H	527	YKPECG KSFSSPA DLTRHQT H	545	FQORCM RNFSDSG NLRVHIRT H	563
ND4_Left	C	FHCGYCE KSFSVKD YLTKHRT H	510	YECVQCG KGFTQSS NLTHQRT H	528	YTCSYOG KSFTQSN TLKQHTRI H	546	YRCKYCD RSFSISSN LQRHVRNI H	564
ND4_Right	C	FHCGYCE KSFSVKD YLTKHRT H	511	YKPECG KSFSSTLD LIRHQT H	529	YKPECG KSFSKK ALTEHQT H	547	FQORCM RNFSDSG NLRVHIRT H	565
ND5_Left	C	FHCGYCE KSFSVKD YLTKHRT H	512	YECDHCG KSFSOSS HLNVHKR TH	530	FQORCM RNFSDSG NLRVHIRT H	548	YECDHCG KSFSOSS HLNVHKR TH	566
ND5_Right	N	FACPECP KRFMRSD NLTHIKT H	513	YECNVOG KTFVSST LRHQBH H	531	YECDHCG KAFSVSS NLNVHRR H	549	YECDHCG KAFSVSS NLNVHRR H	567
ND6_Left	C	YECDHCG KSFRQST HLTHRR H	514	YECDHCG KSFRQST HLTHRR H	532	YKCGQCG KFYSQVS HLTRHOKI H	550	YHCDWD GCGWKF ARSDLT RHYRKH H	568
ND6_Right	N	YRCEECG KAFRWPS NLTRHKRI H	515	YKPECG KSFSRED NUHQT H	533	YGCILCG KAFSKSS NLRRHEMI H	551	YKPECG KSFSRED NUHQT H	569
CYB_Left	C	FECKDGG KAFKOSN LIRHQT H	516	FHCGYCE KSFSVKD YLTKHRT H	534	YTCSYOG KAFRDKS CLNRHRR TH	552	YTCSYOG KSFTQSN TLKQHTRI H	570
CYB_Right	C	YTCSYOG KSFTQSN TLKQHTRI H	517	YKDECG KNFTOSS NLVHKRH H	535	YKPECG KSFSDPG HLVRHQR TH	553	YECDHCG KSFSOSS HLNVHKR TH	571

[833]

	Left-ZFD binding region	Spacer region	Right-ZFD binding region	ZFD configuration
ND1	5'- GTTTACTCAATC 3'- CAAATGAGTTAG	CTCTGATC GAGACTAG	AGGGTGAGCATC TCCCACTCGTAG	-3' -5' C-N
ND2	5'- CTACGCCTAATC 3'- GATGCGGATTAG	TACTCCACCTCAA ATGAGGTGGAGTT	TCACACTACTCC AGTGTGATGAGG	-3' -5' C-N
Cox1	5'- TCTGACTCTTAC 3'- AGACTGAGAATG	CTCCCTCTCTCC GAGGGAGAGAGG	TACTCCTGCTCG ATGAGGACGAGC	-3' -5' C-N
Cox2	5'- GCCATCATCCTA 3'- CGGTAGTAGGAT	GTCTCATCGCC CAGGAGTAGCGG	CTCCCATCCCTA GAGGGTAGGGAT	-3' -5' C-N
ND4L	5'- GCCCTACTAGTC 3'- CGGGATGATCAG	TCAATCTCC AGTTAGAGG	AACACATATGSC TTGTGTATACCG	-3' -5' C-C
ND4	5'- ATATTTTATATC 3'- TATAAAATATAG	TTCTTCGA AAGAAGCT	AACCACACTTAT TTGGTGTGAATA	-3' -5' C-C
ND5	5'- ATATCGGTTTCA 3'- TATAGCCAAAGT	TCCTCGC AGGAGCG	CTTAGCATGATT GAATCGTACTAA	-3' -5' C-N
ND6	5'- TCTTCTTCCAC 3'- AGAAGAAGGGTG	TCATCCTA AGTAGGAT	ACCCTACTCCTA TGGGATGAGGAT	-3' -5' C-N
CYB	5'- TTCATAGGCTAT 3'- AAGTATCCGATA	GTCTCCCG CAGGAGGGC	TGAGGCCAAATA ACTCCGGTTTAT	-3' -5' C-C

[834] HEK293T세포에서 mitoZFD에 의한 미토콘드리아 DNA 염기 교정 효율은 2.6%~30%(평균 $14 \pm 3\%$)이다(도 9a). NC 구성을 가지는 mitoZFD($16 \pm 4.5\%$, n=6)는 CC구성($8.3 \pm 2.7\%$, n=3)을 가지는 것만큼 효율이 나왔다. 스페이서 부분에서 TC context를 가지는 대부분의 사이토신은 다양한 효율로 염기 교정되었다(도 9b~g). 추가로, ND2 부위에 있는 ACC(C₈ 및 C₉) context를 가지는 사이토신도 각각 7.4% and 19.9% 정도되는 효율을 보였다(도 9b). 이것은 ZFD를 매개로 하는 C에서 T로의 염기 교정이 TC motif에 국한되지 않는다는 것을 보여준다.

[835] 다음으로 미토콘드리아 DNA(mtDNA) 변이 세포에서 단세포 유래 클론집단(single-cell derived clonal populations)을 분리해 mitoZFD가 독성이 있지 않고, 변이 mtDNA가 클론집단에서도 유지된다는 것을 증명하려고 했다. ND1-specific mitoZFD를 처리한 HEK293T 세포에서 분리한 단세포 유래 30개의 클론 집단 중, 5개는 ND1 유전자의 염기 교정 효율이 35%에서 98%로 나왔다(도 10a). 마찬가지로, ND2-specific mitoZFD를 처리한 HEK293T 세포에서 분리한 단세포 유래 36개 클론 집단 중, 7개는 ND2 유전자의 염기 교정 효율이 26%에서 76%로 나왔다(도 10b). 다른 모든 클론집단에 대해서는 0.4%-1.0%로 낮은 효율을 보였고, 대부분은 시퀀싱 에러처럼 보인다. ZFD를 처리하지 않은 세포에서도

비슷한 효율이 나왔다(도 10c). 이러한 결과는 mitoZFD가 heteroplasmic한 변이를 세포 개체에서 고르게 일으키지 않는다는 것을 보여준다. ZFD를 처리한 세포에서 대부분은 wild-type이고, 반면에 heteroplasmic 변이를 가지는 것은 98%까지의 변이율을 가진다. 이러한 변이는 clonal expansion을 해도 계속 유지된다(도 11 및 도 12).

[836]

[837] 1-6. mitoZFDs 및 TALE-기반 DdCBEs

[838]

[839] 제작한 *ND1*-specific mitoZFD가 동일한 유전자를 타겟하는 TALE 기반 DdCBE와 변이 패턴이 다른 것을 알 수 있었다(도 9f-h). 2개의 mitoZFD 같은 경우 C₅ 또는 C₈ 위치의 사이토신을 C에서 T로의 염기교정을 하는 반면(도 9f), DdCBE는 C₈, C₉, 및 C₁₁ 부위를 염기 교정을 한다(도 9g). 결과적으로, mitoZFD에 의한 아미노산 변화는 완전하게 DdCBE에 의한 아미노산 변화와 다르다(도 9h). mitoZFD가 붙는 왼쪽과 오른쪽 부분이 스페이서 8-bp 길이로 나뉜 반면, DdCBE 같은 경우 스페이서 길이 16-bp 길이로 나뉘어져 있기 때문에 이러한 차이로 인해 다른 변이 패턴을 가질 수 있다. 이러한 결과는 미토콘드리아 DNA에서 mizoZFD와 DdCBE가 서로 보완적으로 다양한 변이를 만들어 낼 수 있다는 것을 보여준다.

[840] 변이 패턴을 더 많이 만들기 위해, ZFD 단일체와 DdCBE의 단일체가 혼합되어 혼합쌍(hybrid pair)으로 사용될 수 있는지 테스트하였다. *ND1* 유전자를 타겟하는 10개의 혼합쌍은 HEK293T 세포에서 평균 17±3.4%의 염기 교정 효율로 좋은 활성을 보였다(도 13). 실제로, 혼합쌍 중 하나(TALE-L/ZFD-R1)는 두개의 DdCBE쌍 및 10개의 같은 곳을 타겟하는 ZFD 쌍보다 더 좋은 효율을 보였으며 가장 높은 41%의 염기 교정 효율을 보였다(도 13b). 더욱이, 혼합쌍은 DdCBE 및 mitoZFD와는 다른 변이 패턴을 보였다(도 13c). 몇 개(예를 들어, ZFD-L1/TALE-R 및 ZFD-L2/TALE-R)는 다른 부수적인 변이 없이 한 곳에서만 C에서 T로의 변이를 일으켰다. 대조적으로, 대부분의 mitoZFD쌍과 DdCBE쌍은 스페이서 내 여러 곳에서 C에서 T로의 변이를 일으킨다. 이러한 결과는 ZFD/DdCBE 혼합쌍이 특이적인 변이 패턴을 만들 수 있고, ZFD쌍이나 DdCBE쌍으로는 만들 수 없는 특정한 변이를 만들어 낼 수 있다는 것을 보여준다.

[841]

[842] 1-7. mitoZFD의 미토콘드리아 genome-wide 타겟 특이성

[843]

[844] mitoZFD가 미토콘드리아 DNA에서 off-target을 일으키는지 확인하기 위해서 *ND1* 또는 *ND2* 유전자를 타겟하는 두 쌍의 mitoZFD를 처리한 세포에서 유전자를 추출하고 미토콘드리아 전체 유전자 시퀀싱(whole mitochondrial genome sequencing)을 하였다. 다양한 양(5~500ng)의 mitoZFD 쌍을 인코딩하는

mRNA 또는 plasmid을 HEK293T 세포에 형질감염 하였다. 예측한대로, on-target 효율은 양에 의존적이었다(dose dependent). 높은 농도(100, 200 및 500ng)의 mRNA 또는 plasmid는 on-target에서 >30%의 효율을 보였지만, 수백개의 1%가 넘는 off-target도 일으켰다(도 14-17). 낮은 농도(5 및 10ng)에서는 이러한 off-target들이 거의 없어졌지만, on-target에서의 효율 역시 확연히 낮아졌다. 중간 농도인 50ng양의 mRNA가 가장 적절하였다. 수백개의 off-target을 일으키지 않으면서, 높은 on-target 효율을 유지하였다. 좀 더 남은 off-target을 없애기 위해, R(-5)Q 변이를 각 zinc finger 부분에 도입해서 비 특이적 DNA 접촉을 없애고자 하였다. 제작된 ZFD 변이체 (도 18에서 QQ로 나타냄)는 높은 on-target 활성은 유지하면서 off-target이 ZFD를 처리하지 않은 세포의 mtDNA에 비교해서 거의 없는 정교한 특이성을 가졌다(도 18 a 및 b).

- [845] 염기 교정은 DNA 두 가닥 절단을 일으키지 않고 DNA 수선 템플릿 없이 타겟 염기를 교정할 수 있는 비교적 새로운 기술이다. 염기교정은 세포, 동물 및 식물에서 C에서 T로의 또는 A에서 G로의 염기 교정을 일으켜 단일 염기 변이(SNP)의 기능적 영향을 연구할 수 있고, 치료 어플리케이션으로 질병을 일으키는 point mutation을 고치는 것을 할 수 있다. 두가지 종류의 염기 교정 기술이 개발되었다. CRISPR 기반인 아데닌 및 사이토신 염기 교정 가위와 DddA 기반의 염기 교정 기술이다. CRISPR 기반의 염기 교정 가위는 DNA 결합 유닛으로 catalytically-impaired Cas9 또는 Cas12a를 구성하고 있고, deaminase는 rat 또는 E.coli 유래의 단일 나선 DNA 특이적인(single-strand DNA specific) 것을 가지고 있다. 반면에, DdCBE 같은 경우 TALE DNA 결합 array와 이중 나선 DNA 특이적인(double-strand DNA specific) DddAtox를 구성하고 있다.
- [846] DdCBE와 비교해서 ZFD는 크기가 더 작다. 왜냐하면 ZFD에 포함된 zinc finger 단백질의 크기가 작는데 반해 DdCBE에 포함된 TALE array는 크기가 크기 때문이다. 결과적으로, 작은 탑재 공간을 가지고 있는 AAV 벡터에 ZFD 쌍을 인코딩하고 있는 유전자는 쉽게 탑재할 수 있는 반면에 DdCBE 쌍을 인코딩하는 유전자는 그렇지 못하다. 추가로, 간소한 ZFP 특성으로 인해 엔지니어링 하기가 쉽다. 반쪽 DddAtox를 ZFP의 N-말단 또는 C-말단 부분에 융합할 수 있기 때문에 ZFP 결합 부위의 상류 또는 하류에 작동하도록 만들 수 있다. 더욱이, ZFD 재조합 단백질은 전기천공이나 리포펙션 없이 자발적으로 사람 세포에 관통해서 들어갈 수 있다. 이는 gene-free 유전자 치료를 가능할 수 있게 한다. ZFD 쌍 또는 ZFD/DdCBE 혼합쌍은 고유한 변이 패턴을 만들 수 있다. 이는 DdCBE만 사용해서는 얻을 수 없는 패턴이었다. 이러한 특성들은 ZFD가 미토콘드리아 질병의 모델링 및 치료에 새로운 강력한 플랫폼이 되도록 한다.
- [847]
- [848] 실시예 2. TALE-DdCBE 식물 엽록체, 미토콘드리아 유전자 교정
- [849]
- [850] 미토콘드리아와 엽록체를 포함하는 식물 소기관에는 각각 호흡과 광합성을

위한 필수적인 많은 유전자를 암호화하는 자체 게놈이 있다. 식물 유전학과 생명공학에 대한 충족되지 않은 조건인 식물 소기관 유전자 교정은 이러한 세포 기관의 DNA를 표적화하기 위한 적절한 도구의 부족으로 인해 제한되어 왔다. DddA-유래 시토신 염기 교정 플라스미드 (DdCBE)를 조립하기 위해 16개의 발현 플라스미드 (생성 단백질을 미토콘드리아 전달하기 위한 8개, 및 엽록체로 전달하기 위한 다른 8개)와 424개의 TALE 서브 어레이 플라스미드로 구성된 골든 게이트 클로닝 시스템을 개발했고, 미토콘드리아와 엽록체에 효율적으로 증진된 점 돌연변이를 만들기 위해 완성된 DdCBE 플라스미드를 사용하였다. DdCBE 염기 교정은 상추 또는 유채의 칼리에서 최대 25% (미토콘드리아)와 38% (엽록체)의 효율을 유도하였다. DdCBE를 암호화하는 플라스미드에서 원인이 되는 비표적변이를 피하고자 DdCBE 메신저 RNA를 상추 원형질체에 전달하여 DNA 없이 염기교정을 엽록체에서 보여줬다. 게다가, 엽록체 16S rRNA 유전자의 점돌연변이를 도입하여 스트렙토마이신 또는 스펙티노마이신에 저항성이 있는 최대 99%의 교정 효율을 가지는 상추 칼리와 유식물체를 만들었다.

- [851] DdCBEs는 박테리아 시토신 디아미나아제 독성 물질 DddAtox에서 유래된 분리된 비독성 도메인과 특정 위치에 맞게 디자인된 TALE array 및 UGI로 구성되어 헤테로다имер로 타겟 DNA의 TALE 단백질 결합 위치 사이 스페이스 공간에서 시토신을 티민으로 치환하는 시토신 디아미네이션 반응을 일으키는 기능을 한다. 미토콘드리아와 엽록체로 발현하는 DdCBE 플라스미드의 결합을 빠르고 편리하게 나타내며 높고 효율적인 식물에서의 소기관 염기교정을 입증하는 것을 DdCBEs 결과로 사용하였다.
- [852]
- [853] 2-1. 방법
- [854]
- [855] 식물 원형질체에서의 발현 플라스미드 제작
- [856]
- [857] DdCBE 골든 게이트 데스티네이션 벡터는 김슨 어셈블리 방법을 이용하여 제작되었다. TAL N-터미널 도메인, HA tag, FLAG tag, TAL C-터미널 도메인, 분할 DddAtox, 및 UGI를 인코딩 하는 서열은 쌍자엽식물 (애기장대) 발현에 코돈 최적화하였고, Integrated DNA Technology를 통해 합성되었다. AtinfA 와 AtRbcS로부터의 CTP와 ATPase 델타 서브유닛과 ATPase 감마 서브유닛으로부터의 MTS를 인코딩하는 서열은 애기장대 cDNA로부터 증폭되었다. 식물 발현을 위해 백본 플라스미드에서 포유류 CMV 프로모터를 PcUbi 프로모터와 pea3A 터미네이터로 대체하여 사용하였다. 기내 DdCBE mRNA 전사를 위해 벡터 제작에는, T7 프로모터 카세트가 DdCBE 골든 게이트 데스티네이션 벡터에서 PcUbi 프로모터와 DdCBE 인코딩 지역 사이에 도입되었다.
- [858] TALE 어레이 유전자는 단일 골든 게이트 어셈블리를 이용하여 제작되었다.

DdCBE 발현 플라스미드는 424개의 TALE 어레이 플라스미드와 테스트네이션 벡터를 이용하여 BsaI 분해와 T4 라이게이션의 골든 게이트 어셈블리로 제작되었다. 단일 골든 게이트 클로닝은 다음과 같이 수행되었다. 20 반복의 37°C와 50°C 각 5분, 최종 50°C 15분 반응 후 80°C 5분. 모든 식물 원형질체 형질주입 벡터는 플라스미드 플러스 미니프랩 키트 (쿼아젠)을 이용하여 정제되었다. 벡터의 DNA와 아미노산 서열은 다음과 같다.

[859]

[860] [표2]

Name	Sequence	No.
파슬리 유비쿼틴 프로모터	CTAGCAACGATTGTACAATTGCTTCTTTAAAAAAGGAAGAAAGAAA GAAAGAAAAGAATCAACATCAGCGTTAACAAACGCCCCGTTACG GCCCCAACGGTCATATAGAGTAACGGCGTTAAGCGTTGAAAGACTC CTATCGAAATACGTAACCGCAAACGTGTCATAGTCAGATCCCCTCTT CCTTCACCGCCTCAAACACAAAATAATCTTCTACAGCCTATATATA CAACCCCCCTTCTATCTCTCTCTCTCTCTCTCTGTTATTCCTTGTTTAAATTA CTCTACCCCCAATTTAAGAAATCCTCTCTCTCTCTCTCTCTTCAATTTCAA GGTAAATCTCTCTCTCTCTCTCTCTCTCTCTGTTATTCCTTGTTTAAATTA GGTATGTATTATTGCTAGTTTGTTAATCTGCTTATCTTATGTATGCCT TATGTGAATATCTTATCTTGTTTCATCTCATCCGTTTAGAAGCTATAA ATTTGTTGATTTGACTGTGTATCTACACGTGGTTATGTTTATATCTAA TCAGATATGAATTTCTTCATATTGTTGCGTTTGTGTGTACCAATCCG AAATCGTTGATTTTTTTCATTTAATCGTGTAGCTAATTGTACGTATAC ATATGGATCTACGTATCAATTGTTTCATCTGTTTGTGTTTGTATGTATA CAGATCTGAAAACATCACTTCTCTCATCTGATTGTGTTGTTACATAC ATAGATATAGATCTGTTATATCATTTTTTTTATTAATTGTGTATATAT ATATGTGCATAGATCTGGATTACATGATTGTGATTATTTACATGATT TTGTTATTTACGTATGTATATATGTAGATCTGGACTTTTTGGAGTTGT TGACTTGATTGTATTTGTGTGTGTATATGTGTGTTCTGATCTTGATAT GTTATGTATGTGCAGC	162
완두콩 3A 터미네이터	CAGGCCTCCAGCTTTCGTCCGTATCATCGGTTTCGACAACGTTTCGT CAAGTTCAATGCATCAGTTTCATTGCCACACACCAGAATCCTACTA AGTTTGAGTATTATGGCATTGGAAAAGCTGTTTCTTCTATCATTG TTCTGCTTGTAATTTACTGTGTTCTTTTCAGTTTTTGTTCGGACATC AAAATGCAAAATGGATGGATAAGAGTTAATAAATGATATGGTCCTTT TGTTTCATTCTCAAATTATTATTATCTGTTGTTTTTACTTTAATGGGTT GAATTTAAGTAAGAAAGGAACAAACAGTGTGATATTAAGGTGCAAT GTTAGACATATAAAACAGTCTTTACCTCTCTTTGGTTATGCTTGA ATTGGTTTGTTCCTTCACTTATCTGTGTAATCAAGTTTACTATGAGTC TATGATCAAGTAATTATGCAATCAAGTTAAGTACAGTATAGGCTT	163
T7 promoter	TAATACGACTCACTATAGG	164

[861] DdCBE의 구성에 대한 구체적 아미노산 서열 및 TALE repeat에 대한 구체적 아미노산 서열은 다음과 같다.

[862]

[863] [33]

Name	Sequence	No.
AtinfA CTP-3xHA-N terminal domain	MLQLCSTFRPQLLLPCQFRFTNGVLIPQINYVASNSVNVNIRPMIRCQRAS GGRGGANRSKPAKPQVKEGSNKTVEGLVTESLPNGMFRVDLENGDNI LGYICGKIRKNFIRILPGDKVKVEMSVYDSTKGRIIFRMSSRDYPYDVPD YAGYPYDVPDYAGYPYDVPDYAMDIADLRTLGYSSQQQKEKIKPKVRS TVAQHHEALVGHGFTTHAHIVALSQHPAALGTVAVKYQDMIAALPEAT HEAIVGVGKQWSGARALEALLTVAGELRGPPQLDGTGQLLKIAGRGG VTAVEAVHAWRNALTGAPL	165
AtRbcS CTP-3xFLAG-N terminal domain	MASSMLSSATMVASPAQATMVAPFNGLKSSAAFPAATRKANNDITSITS NGGRVNCMQVWPPIGKKKFETLSYLPDLTSEDYKDHDGDYKDHDID YKDDDDKMDIADLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFT HAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKRGAGAR ALEALLTVAGELRGPPQLDGTGQLLKIAGRGGVTAVEAVHAWRNALT GAPL	166
AtATPase gamma subunit MTS-3xHA-N terminal domain	MAMAVFRREGRRLLPSIAARPIAAIRSPSSDQEEGLLGVRISISTQVVRN RMKSVKNIQKITKAMKMAASKLRAVQYPYDVPDYAGYPYDVPDYA GYPYDVPDYAMDIADLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGH GFTTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWS GARALEALLTVAGELRGPPQLDGTGQLLKIAGRGGVTAVEAVHAWRN ALTGAPL	167
AtATPase delta subunit MTS-3xFLAG-N terminal domain	MFQKASRLLSRVAAASSKSVTTTRAFSTELPSTLSDSYKDHDGDYKDH DIDYKDDDDKMDIADLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGH GFTTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKRGAR GARALEALLTVAGELRGPPQLDGTGQLLKIAGRGGVTAVEAVHAWRN ALTGAPL	168
Half (NG) domain-C-terminal domain-G1333-N-UGI	GLTPEQVVAIASNNGGKQALESIVAQLSRPDPALAALTNDHLVALACL GGRPALDAVKKKGLGGSGSYALGPYQISAPQLPAYNGQTVGTFYYVND AGGLESKFSSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGN KPESDILVHTAYDESTDENVMLLTSDAPEYKWPALVIQDSNGENKIKM L*	169
Half (NG) domain-C-terminal domain-G1333-C-UGI	GLTPEQVVAIASNNGGKQALESIVAQLSRPDPALAALTNDHLVALACL GGRPALDAVKKKGLGGSPYPYNYANAGHVEGQSALFMRDNGISEGLV FHNNPEGTCGFCVNMTEILLPENAKMTVVPPEGAIKPKRGATGETKVF TGNSNSPKSPTKGGCSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVI GNKPESDILVHTAYDESTDENVMLLTSDAPEYKWPALVIQDSNGENKI KML*	170
Half (NG) domain-C-terminal domain-	GLTPEQVVAIASNNGGKQALESIVAQLSRPDPALAALTNDHLVALACL GGRPALDAVKKKGLGGSGSYALGPYQISAPQLPAYNGQTVGTFYYVND AGGLESKFSSGGSPYPYNYANAGHVEGQSALFMRDNGISEGLVFHNN PEGTCGFCVNMTEILLPENAKMTVVPPEGSGGSTNLSDIIEKETGKQLV	171

[864]

G1397-N-UGI	IQESILMLPEEVVEEVIGNKPESDILVHTAYDESTDENVMMLLTSDAPEYKP WALVIQDSNGENKIKML*	
Half (NG) domain-C- terminal domain- G1397-C- UGI	GLTPEQVVAIASNNGGKQALESTVAQLSRDPALAAALTNDHLVALACL GGRPALDAVKKGLGGSAPVKRGATGETKVFTGNSNSPKSPTKGGCSG GSTNLSDIIEKETGKQLVIQESILMLPEEVVEEVIGNKPESDILVHTAYDES TDENVMLLTSDAPEYKPWALVIQDSNGENKIKML*	172
16s rDNA Left TALE repeat - GGGGAGT ACGTTTCGC AAG	NLTPAQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNN GGKQALETVQRLLPVLCQDH	173
	GLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNN GGKQALETVQRLLPVLCQDH	174
	GLTPAQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNN GGKQALETVQRLLPVLCQAH	175
	GLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNN GGKQALETVQRLLPVLCQAH	176
	GLTPAQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNN GGKQALETVQRLLPVLCQDH	177
	GLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNG GGKQALETVQRLLPVLCQDH	178
	GLTPAQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNN GGKQALETVQRLLPVLCQDH	179
	GLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNN GGKQALETVQRLLPVLCQDH	180
	GLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNN GGKQALETVQRLLPVLCQAH	181
16s rDNA Right TALE repeat - TGTGCGGG CCCCCGTC AA	NLTPAQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNN GGKQALETVQRLLPVLCQAH	182
	GLTPAQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNN GGKQALETVQRLLPVLCQAH	183
	GLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNN GGKQALETVQRLLPVLCQAH	184
	GLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNN GGKQALETVQRLLPVLCQDH	185
	GLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHD GGKQALETVQRLLPVLCQAH	186
	GLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHD GGKQALETVQRLLPVLCQDH	187
	GLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNN GGKQALETVQRLLPVLCQDH	188
	GLTPAQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHD GGKQALETVQRLLPVLCQDH	189

[865]

	GLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIG GKQALETVQRLLPVLCQDH	190
psbA Left TALE repeat - GCAACAG GAGCTGA ATAT	NLTPAQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHD GGKQALETVQRLLPVLCQDH	191
	GLTPAQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIG GKQALETVQRLLPVLCQDH	192
	GLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNI GGKQALETVQRLLPVLCQAH	193
	GLTPAQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNN GGKQALETVQRLLPVLCQDH	194
	GLTPAQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNN GGKQALETVQRLLPVLCQAH	195
	GLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNG GGKQALETVQRLLPVLCQAH	196
	GLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNI GGKQALETVQRLLPVLCQDH	197
	GLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNG GGKQALETVQRLLPVLCQAH	198
	GLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNG GGKQALETVQRLLPVLCQAH	199
	NLTPAQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNN GGKQALETVQRLLPVLCQAH	200
psbA Right TALE repeat - AGTTTCCG TCTGGGTA TG	GLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNG GGKQALETVQRLLPVLCQDH	201
	GLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHD GGKQALETVQRLLPVLCQAH	202
	GLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNN GGKQALETVQRLLPVLCQDH	203
	GLTPAQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHD GGKQALETVQRLLPVLCQAH	204
	GLTPAQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNN GGKQALETVQRLLPVLCQDH	205
	GLTPAQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNN GGKQALETVQRLLPVLCQDH	206
	GLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNI GGKQALETVQRLLPVLCQAH	207
	GLTPAQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNN GGKQALETVQRLLPVLCQDH	208
	NLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIG GKQALETVQRLLPVLCQDH	209
psbB Left TALE repeat - AAATGAA TTCCAAAA	GLTPAQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNG GGKQALETVQRLLPVLCQAH	210
	GLTPAQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNI	211

[866]

ATC	GGKQALETQVRLPVLCDH	
	GLTPAQVVAIASNIGGKQALETQVRLPVLCAHGLTPDQVVAIASNG GGKQALETQVRLPVLCAH	212
	GLTPDQVVAIASNIGGKQALETQVRLPVLCAHGLTPAQVVAIASHD GGKQALETQVRLPVLCDH	213
	GLTPDQVVAIASHDGGKQALETQVRLPVLCAHGLTPDQVVAIASNI GGKQALETQVRLPVLCAH	214
	GLTPDQVVAIASNIGGKQALETQVRLPVLCAHGLTPAQVVAIASNIG GKQALETQVRLPVLCDH	215
	GLTPAQVVAIASNIGGKQALETQVRLPVLCAHGLTPDQVVAIASNIG GKQALETQVRLPVLCDH	216
	GLTPAQVVAIASNIGGKQALETQVRLPVLCAHGLTPAQVVAIASHD GGKQALETQVRLPVLCDH	217
psbB Right TALE repeat - GAACGTA CAGGAAA ACC	NLTPEQVVAIASNIGGKQALETQVRLPVLCAHGLTPDQVVAIASNI GGKQALETQVRLPVLCDH	218
	GLTPDQVVAIASNIGGKQALETQVRLPVLCAHGLTPDQVVAIASHD GGKQALETQVRLPVLCDH	219
	GLTPDQVVAIASNIGGKQALETQVRLPVLCDHGLTPAQVVAIASNG GGKQALETQVRLPVLCAH	220
	GLTPDQVVAIASNIGGKQALETQVRLPVLCDHGLTPDQVVAIASHD GGKQALETQVRLPVLCDH	221
	GLTPAQVVAIASNIGGKQALETQVRLPVLCAHGLTPAQVVAIASNN GGKQALETQVRLPVLCAH	222
	GLTPDQVVAIASNIGGKQALETQVRLPVLCDHGLTPAQVVAIASNI GGKQALETQVRLPVLCAH	223
	GLTPDQVVAIASNIGGKQALETQVRLPVLCAHGLTPAQVVAIASNIG GKQALETQVRLPVLCDH	224
	GLTPAQVVAIASNIGGKQALETQVRLPVLCAHGLTPAQVVAIASHD GGKQALETQVRLPVLCDH	225
	GLTPAQVVAIASHDGGKQALETQVRLPVLCAH	226
ATP6 (lettuce, rapeseed) Left TALE repeat - TATTGGCA TTACTATA G	NLTPDQVVAIASNIGGKQALETQVRLPVLCAHGLTPDQVVAIASNI GGKQALETQVRLPVLCDH	227
	GLTPDQVVAIASNIGGKQALETQVRLPVLCAHGLTPEQVVAIASNG GGKQALETQVRLPVLCAH	228
	GLTPAQVVAIASNIGGKQALETQVRLPVLCDHGLTPDQVVAIASNN GGKQALETQVRLPVLCDH	229
	GLTPAQVVAIASNIGGKQALETQVRLPVLCAHGLTPEQVVAIASNG GGKQALETQVRLPVLCAH	230
	GLTPDQVVAIASNIGGKQALETQVRLPVLCAHGLTPAQVVAIASHD GGKQALETQVRLPVLCDH	231
	GLTPDQVVAIASNIGGKQALETQVRLPVLCAHGLTPDQVVAIASNI GGKQALETQVRLPVLCDH	232

[867]

	GLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNI GGKQALETVQRLLPVLCQDH	233
	GLTPAQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNN GGKQALETVQRLLPVLCQDH	234
	GLTPDQVVAIASNNGGKQALETVQRLLPVLCQDH	235
ATP6 (lettuce) Right TALE repeat - GAAAAAA TGAAGCCC A	NLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNI GGKQALETVQRLLPVLCQDH	236
	GLTPAQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIG GKQALETVQRLLPVLCQAH	237
	GLTPAQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNIG GKQALETVQRLLPVLCQAH	238
	GLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNG GGKQALETVQRLLPVLCQAH	239
	GLTPAQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNI GGKQALETVQRLLPVLCQDH	240
	GLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNN GGKQALETVQRLLPVLCQAH	241
	GLTPAQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHD GGKQALETVQRLLPVLCQDH	242
	GLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNI GGKQALETVQRLLPVLCQAH	243
ATP6 (rapeseed) Right TALE repeat - TAAAAAA TGAAGCCC A	NLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNI GGKQALETVQRLLPVLCQDH	244
	GLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIG GKQALETVQRLLPVLCQAH	245
	GLTPAQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNIG GKQALETVQRLLPVLCQAH	246
	GLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNG GGKQALETVQRLLPVLCQAH	247
	GLTPAQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNI GGKQALETVQRLLPVLCQDH	248
	GLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNN GGKQALETVQRLLPVLCQAH	249
	GLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPAQVVAIASHD GGKQALETVQRLLPVLCQAH	250
	GLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNI GGKQALETVQRLLPVLCQAH	251
RPS14 Left TALE repeat - CGTTCTGT ATCTGAG	NLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNN GGKQALETVQRLLPVLCQDH	252
	GLTPAQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNG GGKQALETVQRLLPVLCQAH	253
	GLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNG	254

[868]		GGKQALETVQRLLPVLCQAH	
		GLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNG GGKQALETVQRLLPVLCQDHG	255
		LTPAQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNNG GKQALETVQRLLPVLCQAH	256
		GLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNG GGKQALETVQRLLPVLCQAH	257
		GLTPAQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNI GGKQALETVQRLLPVLCQAH	258
		GLTPDQVVAIASNNGGKQALETVQRLLPVLCQAH	259
[869]	RPS14 Right TALE repeat - AATCCACG AAAAACG A	NLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNI GKQALETVQRLLPVLCQDH	260
		GLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHD GGKQALETVQRLLPVLCQDH	261
		GLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNI GGKQALETVQRLLPVLCQAH	262
		GLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNN GGKQALETVQRLLPVLCQAH	263
		GLTPAQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNI GKQALETVQRLLPVLCQAH	264
		GLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNI GKQALETVQRLLPVLCQAH	265
		GLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHD GGKQALETVQRLLPVLCQAH	266
		GLTPAQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNI GGKQALETVQRLLPVLCQAH	267

[869]

[870] mRNA 기내 전사

[871]

[872] DdCBE DNA 템플릿은 퓨전 DNA 폴리머라아제를 이용한 PCR에 의해 준비되었다 (서모 피셔). DdCBE mRNA는 기내 mRNA 합성 키트를 이용하여 합성 정제하였다 (엔지노믹스).

[873]

[874] 원형질체 추출과 형질도입

[875]

[876] 상추 종자를 70% 에탄올에 30초 및 0.4% 락스 용액에 15분 표면 살균 후 멸균증류수로 3번 행군다. 상추 종자는 0.5배 MS 배지와 2% 수크로스가 포함된 배지에 16시간 명조건 8시간 암조건에 25°C에서 파종한다. 유채 종자는 70% 에탄올 3분 및 1.0% 락스 용액에 30분 표면 살균 후 멸균증류수로 3번 행군다. 유채 종자는 1배 MS 배지와 3% 수크로스가 포함된 배지에 16시간 명조건 8시간 암조건에 25°C에서 파종한다.

[877] 원형질체 추출과 형질도입은 이전 연구를 따라 수행한다. 7일된 상추와 14일된 유채의 떡잎을 효소 용액에 흔들리는 (40rpm) 암조건에서 3시간 처리한다. 원형질체-효소 혼합물은 동일량의 W5 용액을 이용하여 위시 후, 온전한

원형질체를 80g 7분 원심분리를 하여 수크로스 용액으로부터 얻는다.

원형질체는 W5 용액에서 4°C 1시간 처리 후 폴리에틸렌 글라이콜을 이용한 형질도입을 한다.

- [878] 상추 원형질체와 유채 원형질체는 MMG 용액에 재현탁 후 플라스미드 또는 mRNA를 PEG를 이용한 형질도입을 진행하고 20분간 상온에서 배양한다. PEG-원형질체 혼합물은 동일량의 W5 용액을 이용하여 3번 행구는데 매번 천천히 뒤집으면서 그 이후 10분간 배양한다. 원형질체는 그 이후 100g 5분간 원심분리하여 펠렛형태로 만든다.

[879]

[880] 원형질체 배양

[881]

- [882] 상추 원형질체는 DdCBE를 인코딩하는 플라스미드가 도입된 후 상추 원형질체 배양 배지 (LPCM)에 의해 재부유하였다. 배지를 포함한 원형질체는 배지와 로우멜팅아가로스 2.4%가 포함된 배지와 1:1로 섞은 다음 즉시 6웰 접시에 올린다. 이후 혼합물은 고체화되면, 임베딩된 원형질체 위로 1ml의 액체배지를 올린 뒤 25°C 암조건에서 1주 동안 배양한다. 초기 배양 이후, 위에 올려진 액체배지를 신선한 배지로 바꾸 갈아주고, 임베딩된 원형질체는 16시간 약한 명조건 8시간 암조건에서 1주일 배양한 뒤 16시간 명조건 8시간 암조건에서 2주간 배양한다. 원형질체로부터 유도된 작은 캘러스는 재분화배지에 4주간 25°C 16시간 명조건 8시간 암조건에서 배양한다. 염기 교정 효율의 분석을 준비할 때는, 원형질체는 임베딩 없이 액체배지에서 25°C에 암조건에서 1주간 배양했다. 항생제 저항성을 실험하기 위해, 한달 임베딩 된 작은 캘러스는 50mg/L 농도의 스트렙토마이신 또는 50mg/L 농도의 스펙티노마이신이 포함된 재분화배지에서 25°C 16시간 명조건 8시간 암조건에서 4주간 배양했다. 4주 뒤, 항생제 저항성 녹색 캘러스나 부정지는 200mg/L 농도의 스트렙토마이신 또는 50mg/L 농도의 스펙티노마이신이 포함된 새로운 재분화 배지로 옮겼다.

- [883] DdCBE 인코딩 플라스미드가 도입된 유채 원형질체는 유채 배양 배지에 재부유하였다. 원형질체-배지 혼합물은 6 웰 접시에 옮긴 뒤 25°C 암조건에서 2주 동안 배양한다. 2주 뒤, 원형질체는 16시간 약한 명조건 8시간 암조건에서 3주 배양했다. 배지는 매우 신선한 배지로 교체하였다.

[884]

[885] DNA와 RNA 추출

- [886] 액체 배지에서 배양된 세포나 형질전환 된 캘러스의 총 DNA 또는 RNA는 디엔이지 식물 미니 키트 또는 알엔이지 식물 미니 키트를 이용하여 추출하였다. 배양된 세포나 캘러스는 10,000rpm에서 1분간 원심분리하여 수확하였다. 총 RNA로부터의 cDNA는 RNA to cDNA EcoDry Premix (타카라)를 이용하여 역전사되었다.

[887]

[888] 딥시퀀싱

[889]

[890] 타겟 지역은 퓨전 효소와 적절한 프라이머를 이용하여 증폭하였다 (서플 테이블 1). DNA 서열 분석 라이브러리를 만들기 위해 세 라운드의 PCR (첫째, nested PCR, 둘째, PCR, 셋째, 인덱싱 PCR)을 수행하였다. 동일 양의 DNA를 모은 뒤 MiniSeq (일루미나) 장비를 이용하여 서열분석하였다. 페어 인드 시퀀싱 파일은 Cas-analyzer와 컴퓨터 프로그램 소스코드에 의해 분석되었다.

[891]

[892] 2-2. 결과

[893]

[894] 엽록체를 타겟으로 하는 DdCBE (cp-DdCBE)와 미토콘드리아 타겟으로 하는 DdCBE (mt-DdCBE)를 만들기 위해 골든 게이트 어셈블리 시스템을 개발하였다 (도 19). 발현 플라스미드는 결합된 단백질로 엽록체 전달 펩타이드 또는 미토콘드리아 타겟팅 서열, TALE의 N- 또는 C- 터미널 도메인, 분할 DddAtox (G1333N, G1333C, G1397N, 및 G1397C)와 UGI를 암호화하며 이것은 쌍떡잎 식물의 발현에 최적화되어 있으며, 그것은 파슬리 유비퀴틴 (PcUbi) 프로모터와 pea3A 터미네이터로 조절된다. 커스텀 디자인된 TALE DNA 결합 서열과 DdCBE 플라스미드는 발현용 벡터와 여섯 개의 TALE 서브어레이 플라스미드를 E-tube에 섞어 단일 서브클로닝 단계로 만들 수 있다. 총 424 개의 (6X64 3 서열을 인식 + 2X16 2 서열을 인식 + 2 X 4 1 서열을 인식) 모듈형 TALE 서브어레이 플라스미드들은 5' 끝에 보존된 T를 포함하여 16-20 염기서열 길이를 인식하는 cp-DdCBE와 mt-DdCBE를 만들 수 있다. 결과적으로, 기능적으로 DdCBE 헤테로다이머는 32에서 40 bp DNA 염기서열을 인식한다.

[895]

[896] DdCBEs가 엽록체 염기교정에 축진이 가능한지 알기 위해, 30S 리보솜 서브유닛의 RNA 화합물의 암호화 하는 엽록체 16S rRNA 유전자에 알맞은 cp-DdCBEs 플라스미드 4쌍을 제작하였고, 상추와 유채 원형질체에 각각의 쌍을 같이 형질주입한 뒤, 7일 후 딥시퀀싱을 통해 염기 교정 효율을 측정하였다 (도 20a, b). 가장 높은 효율을 보인 cp-DdCBE 쌍은 (Left-G1397-N+Right-G1397-C) 두 개의 TALE 결합 부위 사이의 15bp의 스페이서 지역에 상추 원형질체에서는 30%, 유채 원형질체에서는 15%의 효율로 C*G를 T*A로 치환을 유도하였다 (도 20b). 이전 포유류 세포와 생쥐의 결과와 마찬가지로, cp-DdCBE에 의해 5'-TC 모티프인 시토신 (C9, C13)은 티민으로 우선적으로 치환되었다. 흥미롭게, 5'-AC 모티프인 시토신 (C7)은 상추 원형질체에서 다른 cp-DdCBE (Left-G1333-N+Right-G1333-C)에 의해 4.2%의 효율로 티민으로 교체되었다. 또한, 상추 원형질체에서 cp-DdCBE에 의한 염기교정의 14일 배양과정에서 지속됨을 조사하였다 (도 24). 교정 효율은 최대 10일까지 지속적으로 증가되었고 배양 기간 내내 유지되었다.

[897]

[898] 광계 II 에서 각각 D1와 CP-47 광합성 단백질 암호화하는 엽록체 유전자인 psbA와 psbB를 두 개의 추가적으로 실험하였다 (도 20c, d 및 도 25). Psb 유전자를 타겟으로 하는 cp-DdCBE 중에, 가장 효율이 높은 것은 (Left-G1397-C+Right-G1397-N) 상추 원형질체에서 C*G를 T*A로 치환을 최대 25%까지 유도 가능하였다 (도 20d). 이 염기교정 가위에 의해 오직 5'-TCC 두 개의 시토신 (C11, C12)만 효율적으로 티민으로 치환되었다. 이것은 5'-TCC가 처음 5'-TTC로 치환되었다가 그 다음 5'-TTT 될 가능성이 있다. 유채 원형질체에서, 다른 조합 (Left-G1333-N+Right-G1333-C)가 네 개의 시토신 위치 (C3, C4, C11, C12) 에서 가장 높은 효율을 보였고 최대 3.5% (C3)까지 효율을 보였다. 유채의 유전자에는 C3와 C4는 5'-TCC 서열이지만, 상추에서는 상대적으로 단일 염기 다형성 때문에 5'-ACC 서열을 가지고 있어 DdCBE에 의해 유채 유전자에서는 두 개의 시토신 (C3, C4)에서 효율적인 교정 효율의 원인이 되지만 상추 유전자에서는 그렇지 못하다. 또한, 유채 원형질체의 psbB 유전자를 타겟으로하는 cp-DdCBE 조합은 TCC 서열이 있는 두 개의 시토신을 0.36%에서 4.1%의 효율로 치환하였다 (도 25). 종합해보면, 이러한 결과는 DddAtox 나뉜 위치 (G1333 vs. G1397)과 방향 (Left-G1333-N vs. Left-G1333-C)을 포함하는 시토신 위치와 서열에 따라 교정 효율이 결정되며 cp-DdCBE는 식물의 엽록체 유전자에 효율적인 염기 교정이 가능한 것을 입증하였다.

[899]

[900] 다음으로, 맞춤 설계된 mt-DdCBE를 이용하여 식물의 미토콘드리아 DNA의 염기 교정을 달성하려고 하였다. 이를 위해 상추와 유채의 atp6 유전자를 타겟으로 하고 유채의 rps14 유전자를 타겟으로 하는 mt-DdCBE를 인코딩하는 플라스미드 (골든 게이트 클로닝 시스템을 이용)를 제작하고 플라스미드를 상추와 유채의 원형질체에 도입시킨 다음, 도입 7일 후 덤시퀀싱을 통해 염기 교정 효율을 측정했다 (도 20e, f, 도 26). 가장 효율이 높은 mt-DdCBE 조합은 (상추에서 Left-G1397-N+Right-G1397-C, 유채에서 Left-G1397-C+Right-G1397-N) atp6 유전자 타겟위치에서 상추 원형질체는 23%, 유채 원형질체는 23%의 C*G를 T*A로 치환을 변화하였다 (도 20). 또한, mt-DdCBE 조합은 rps14 타겟위치에서 유채 원형질체로부터 11%의 C*G를 T*A로 치환을 유도하였다. 이러한 결과들은 mt-DdCBE가 식물 미토콘드리아 DNA에 염기 교정이 잘 이루어진다고 보여진다.

[901]

[902] DdCBE에 의한 cpDNA와 mtDNA의 교정이 재분화 과정 중에 유지되는지 조사하기 위해, 주입 후 4주가 지나 DdCBE가 처리된 원형질체로부터 재분화된 상추와 유채 캘러스를 수집하여 (도 21a), 덤시퀀싱과 생거시퀀싱을 이용하여 각각의 캘러스의 염기교정 효율을 측정하였다 (도 21b, 도 27). DdCBE에 의해 염기 교정이 유도된 엽록체 또는 미토콘드리아의 유전자는 26개 중 22개의 상추

캘러스와 14개 중 7개의 유채 캘러스가 각각 최대 38%와 25%의 효율로 측정되었다 (도 21c). 또한 엽록체 *psbA* 유전자의 염기 교정은 상추 캘러스에서 최대 3.9%의 효율을 보였다 (도 27). 또한, 유채 캘러스에서 미토콘드리아 염기 교정은 *atp6*와 *rps14*에서 각각 최대 25%와 1.9%로 측정되었다 (도 27). 이러한 결과는 식물 원형질체에서 DdCBE의 발현을 견딜 수 있으며 원형질체에 재분화 과정에서 DdCBE에 의해 소기관 염기 교정을 유도하였다.

[903]

[904] 다음으로 플라스미드가 아닌 기내 전사된 cp-DdCBE mRNA를 이용하여 소기관에서 DNA-free 염기교정을 입증하려고 하였다. 상추에서 16S rRNA 유전자를 타겟으로 하는 cp-DdCBE를 인코딩하는 기내 전사물을 원형질체에 주입한 뒤 타겟 위치의 염기 교정 효율을 분석하였다 (도 21a). 원형질체에서 C-to-T 변이는 최대 25%까지 측정되었다 (도 21d, 도 28). 예상대로 DdCBE mRNA와 DNA 서열을 원형질체 도입 7일 후 사라졌다 (도 29). 이 방법은 호스트 지놈에 플라스미드 DNA 일부분이 잠재적으로 통합되는 것을 피할 수 있을 것이다.

[905]

[906] 원형질체로부터 재분화된 캘러스에서 소기관 교정이 안정적으로 유지되는 것에 힘 입어, 엽록체 DNA에서 16S rRNA 유전자의 교정을 통해 16S rRNA에 비가역적으로 결합하여 단백질 합성을 저해하는 스트렙토마이신과 스펙티노마이신 항생제의 저항성을 조사하였다. 16S rRNA 유전자의 몇 군데 단일 염기 다형성은 원핵생물과 진핵생물에서 스트렙토마이신 저항성을 일반적으로 관찰되며 특히, 16S rRNA C860T (대장균은 C912 위치)의 돌연변이는 담배에서 스트렙토마이신 저항성을 가지게 한다. 담배의 C860T 점 돌연변이는 상추에서 C9 위치와 동일하다 (도 20a, 21b, 22b, 22d). DdCBE가 처리된 원형질체로부터 재분화하여 얻어진 상추 캘러스를 스트렙토마이신과 스펙티노마이신이 첨가된 배지에 옮겼다. Mock 처리군은 항생제에 노출되었을 때 하얗게 변화였고, 이는 캘러스의 원형질체 기능장애를 나타낸다. 반면에, DdCBE 처리된 캘러스는 녹색을 유지하였고 이러한 항생제에 저항성을 보여준다. 저항성을 가지는 상추 캘리와 유식물체에서 DdCBE에 의한 교정효율을 분석하였다. C860T 변이와 같은, C9 위치의 C에서 T 치환은 약물 처리로부터 얻어진 캘러스와 슈트 (shoot)에서 최대 98.6% 관찰되었다 (도 21e,f). 흥미롭게도, 근처의 C13위치의 C에서 T 교정이 스펙티노마이신의 없을 경우 최대 20%의 효율을 보였지만 항생제가 있을 경우 보이지 않았는데, 약물 처리에 따라 이 변이가 선발된 것을 입증한다. 종합하였을 때, 이러한 결과는 원형질체에서 DdCBE에 의해 식물 소기관 변이가 유도된 것은, 세포 분열과 식물 발달 뒤에도 유지될 수 있으며, 엽록체 교정이 약물 선발로 호모플라스미를 이룰 수 있을 것이다.

[907]

- [908] 또한 16S rRNA 위치를 타겟으로 하는 TALE 디아미나아제의 오프타겟을 원형질체, 캘러스, 슈트에서 분석하였다. 단일세포에서 유래되어 항생제 저항성을 가지는 캘러스와 슈트에서 타겟사이트의 타겟주변 (양쪽 50 염기쌍) (도 31) 또는 엽록체 유전자에서 서열을 유사함을 기본으로 하여 선택한 5개의 잠재적 오프타겟 위치 (도 32) 오프타겟이 보이지 않는 것이 측정되었다. 반면에 DdCBE를 인코딩하는 플라스미드를 원형질체에 도입하였을 때, TC가 TT로 변하는 오프타겟이 5개의 잠재적 오프 타겟 중 3 군데에서 1.2%에서 4.1%의 낮은 효율로 유도되었다. TALE 디아미나아제를 인코딩하는 플라스미드 대신 기내 전사물 (mRNA)을 사용하였을 때 원형질체에서 오프 타겟의 효율이 크게 감소하였다 (도 22). 이러한 결과로 보면 과발현되거나 장기적인 플라스미드 유래 DdCBE의 발현은 오프타겟의 변이를 증가시킬 수 있고 일시적인, mRNA를 사용하는 mRNA 유래 발현은 오프타겟 염기 교정을 피하는데 바람직하다.
- [909]
- [910] 요약하면, 424 TALE 서브어레이 플라스미드와 16개의 발현 플라스미드를 사용하는 골든 게이트 클로닝 시스템을 개발하여 식물에서 소기관 염기교정을 하기 위해 DdCBE를 인코딩하는 플라스미드를 조립할 수 있다. 엽록체 DNA의 3 유전자, 미토콘드리아 DNA 2 유전자를 타겟하는 맞춤형 디자인 된 DdCBE는 상추와 유채 원형질체에서 높은 효율의 C에서 T 치환을 이루었다. 특히나, 식물의 소기관 교정을 세포분열과 식물 발달에서 유지가 교정이 유지가 되었다. 게다가, 엽록체 16S rRNA 유전자의 변이를 통해 거의 호모플라스미 (99%) 한 항생제 저항성 상추 캘러스와 식물체를 확보하였다. 항생제 선발 없이는, 미토콘드리아에서 25%, 엽록체에서 38%의 교정 효율을 보였다. 골든 게이트 클로닝 시스템이 식물의 소기관 DNA 교정에 가치있는 자원이 될 것이라 기대한다.
- [911]
- [912] 실시예 3. TALE-DdCBE 동물 DNA 교정
- [913]
- [914] 분할된 박테리아 독소 DddAtox, DNA에 결합할 수 있도록 고안된 TALE (transcription activator-like effector) 및 우라실 글리코실레이즈 억제제 (UGI)를 포함하는 DddA-유래 시토신 염기 교정기 (DdCBEs)는 목표한 시토신-티민 염기 변형을 미토콘드리아 DNA에서 가능하게 하였다. 생쥐 배아에서 높은 효율의 미토콘드리아 DNA 교정이 가능함을 보여주었다. 미토콘드리아 유전자 중 NADH 탈수 및 유비퀴논으로의 전자전달을 촉매하는 NADH 탈수소화 효소의 서브유닛을 코딩하는 MT-ND5(ND5)를 목표로 하였는데, 이는 m.G12918A와 같이 사람의 미토콘드리아 질병과 관련 된 돌연변이와 m.C12336T와 같이 이른 종결코돈을 만드는 돌연변이를 포함하였다. 이를 통해 미토콘드리아 질병 모델을 생쥐에서 제작할 수 있었고, 미토콘드리아 질병을 치료할 수 있는 가능성을 제시하였다.

[915]

[916] 3-1. 방법

[917] 플라스미드 조립. TALEN (transcription activator-like effector nucleases) 시스템을 DddA 반체가 포함된 발현 플라스미드 및 최종 TALE-DddAtox 구조를 만들기 위해 차용하였다. TALEN 시스템의 발현 플라스미드에서, 핵내 수송 신호와 FokI 이량체 중 단량체를 미토콘드리아 전달 신호(MTS), DddA 탈아미노화 효소 반체 및 우라실 글리코실레이즈 억제제(UGI)로 치환하였다. MTS, DddA, 및 UGI를 인코딩하는 서열은 IDT에서 합성하였다. 발현 벡터를 제작하기 위해, Gibson assembly에 필요한 DNA 조각을 Q5 DNA 중합효소(NEB)를 이용해 증폭하였고, 이후 정제하였다. 정제된 유전자 조각은 HiFi DNA assembly kit(NEB)를 이용해 조립하였으며, E coli DH5a(Enzymomics)에 화학적 형질전환 시킨 후 생거 서열분석법으로 확인하였다. 따라서, 8개의 각기 다른 발현 플라스미드를 얻었는데, 여기에는 Golden-gate 클로닝을 위한 BsaI 제한효소 자리가 N 말단 도메인과 C 말단 도메인을 인코딩하는 서열 사이에 있다. DdCBE 플라스미드 조립을 위해, 발현 플라스미드에 모듈 벡터(각각은 TALE 서열을 코딩함), BsaI-HFv2 (10 U), T4 DNA ligase (200 U) 및 반응 버퍼를 한 튜브에 담았다. 이후 제한 효소 및 라이게이즈 반응은 서모사이클러에서 37°C 5분, 50°C 5분의 20 반복 후 50°C를 15분, 80°C를 5분동안 처리하였다. 접합된 플라스미드는 화학적 형질전환을 통해 E coli DH5α에 도입되었고, 최종 구조체는 생거 서열분석법으로 확인되었다. 세포주 도입을 위해 플라스미드는 미디프랩 하였다.

[918]

[919] 포유류 세포주 배양 및 형질 감염. NIH3T3 (CRL-1658, American Type Culture Collection (ATCC)) 세포주는 37°C의 5 % CO₂ 환경에서 배양되었다. 세포주는 10 % (v/v) 우아 혈청 보강한 DMEM (Gibco)의 배지에서 항생제 없이 성장하였고 마이코플라즈마 테스트는 하지 않았다. Lipofection을 위해 세포는 12-well cell culture plates (SPL, Seoul, Korea)에 1.5 × 10⁴ 세포 밀도로 형질 감염 18-24 시간 전 성장을 시작하였다. Lipofectamine 3000 (Invitrogen)을 이용하여 각각의 DdCBE 분할체를 500 ng 사용하여 총 1,000 ng의 플라스미드 DNA를 도입하였다. 세포는 형질 감염 4일 후 회수되었다.

[920]

[921] mRNA 제작. mRNA의 주형은 Q5 DNA 중합효소(NEB)를 이용해 PCR 증폭하였고, 다음 프라이머를 사용하였다 (F: 5'-CATCAA TGGGCGTGGATAG-3' SEQ ID No. 268, R: 5'-GACACCTACTCAGACAATGC-3' SEQ ID No. 269). DdCBE mRNA는 in vitro RNA transcription kit (mMESSAGE mMACHINE T7 Ultra kit, Ambion)를 사용하여 합성한 후 MEGAclear kit(Ambion)을 이용해 정제하였다.

[922]

- [923] 동물. 쥐를 포함한 실험은 모두 기초과학연구원의 동물 관리 및 사용 위원회의 승인을 받고 진행하였다. 과배란된 C57BL/6J 암컷을 C57BL/6J 수컷과 교배시켰고, ICR strain 암컷을 대리모로 사용하였다. 생쥐는 특정 병원체가 없는 시설에서 12 시간의 낮-밤 사이클을 지키는 조건 및 항온 및 항습 조건에서 사육되었다(20-26 °C, 40-60%).
- [924]
- [925] 생쥐 접합체에 미세주입. 미세주입 직전의 과정인 과배란 및 배아 채집, 미세주입은 지난 논문에서 기재된 대로 진행하였다. 미세주입을 위해서, left DdCBE mRNA (300 ng/μl) 및 right DdCBE mRNA (300 ng/μl)가 포함된 혼합물은 DEPC처리 된 주입 버퍼 (0.25 mM EDTA, 10 mM Tris, pH 7.4)에 풀렸으며, 접합체 세포질에 Nikon ECLIPSE Ti micromanipulator와 a FemtoJet 4i microinjector (Eppendorf)를 이용하여 주입되었다. 미세주입 후, 배아는 KSOM + AA (Millipore) 마이크로 드롭에 넣어져 37 °C, 5% CO₂ 조건에서 4일동안 배양되었다. 2세포기의 배아는 0.5-d.p.c. pseudo-pregnant 대리모의 난관에 이식되었다.
- [926]
- [927] 유전형 분석. 배반포 단계의 배아 및 조직은 분해 버퍼 (25 mM NaOH, 0.2 mM EDTA, pH 10)에 넣고 95°C 에서 20 분 동안 배양 후, HEPES(free acids, pH 조절 안함)를 이용해 최종 농도가 50 mM이 되도록 pH를 7.4로 조절하였다. 생쥐 새끼의 유전체 DNA 분리를 위해서는 DNeasy Blood & Tissue Kits (Qiagen)를 이용하였고, 생거 서열 분석법 및 targeted deep sequencing으로 분석하였다.
- [928]
- [929] 고처리 서열 분석을 위한 미토콘드리아 DNA 분리. 12-well plate에 배양 된 NIH3T3세포의 미토콘드리아를 분리하기 위해, 세포 배양 배지를 제거한 후 200 μl의 Mitochondrial isolation buffer A (ScienCell)를 배양 플레이트에 넣었다. 세포를 cell lifter를 이용해 긁은 후 마이크로튜브에 담고 일회용 막자를 이용해 세포를 갈아주었다. 15회의 분쇄 후 잘 갈린 균질액은 1,000×g, 4°C에서 5분동안 원심분리 되었다. 상층액을 새로운 마이크로튜브에 담고 10,000×g, 4°C에서 20분동안 원심분리하였다. 침전물을 20 μl 의 용해액(25 mM NaOH, 0.2 mM EDTA, pH 10)에 풀어준 후 95°C에서 20분간 끓여주었다. pH를 낮추기 위해 2 μl의 1 M HEPES (free acids, without pH adjustment)를 미토콘드리아 용해액에 추가하였다. 이와같이 준비 된 용액의 1 μl를 고처리 서열 분석을 위한 PCR 주형 가닥으로 사용하였다.
- [930]
- [931] 고처리 서열 분석. 딥 시퀀싱 라이브러리를 제작하기 위해 Q5 DNA Polymerase를 사용하여 nested 1차 PCR 및 2차 PCR을 진행하였고, 최종 인덱스 서열을 추가하였다. 라이브러리는 MiniSeq (Illumina)를 이용해 페어드 엔드 리드 서열 분석에 사용하였다. 전 미토콘드리아 유전체 분석을 위해서는 분리된

미토콘드리아 DNA를 tagmentation DNA prep kit (Illumina)을 이용하여 제조사 프로토콜을 따라 준비하였다. 모든 분석에서 페어드 엔드 서열분석 결과는 하나의 fastqjoin 파일을 이용해서 합쳐졌으며 CRISPR RGEN Tools (<http://www.rgenome.net/>)을 이용해 분석하였다.

[932]

[933] 데이터 분석 및 표시. 마이크로소프트 엑셀 (2019) 및 파워포인트 (2019)는 그림, 그래프, 및 표 제작에 사용되었다. 유전체 서열 정렬, 프라이머 제작 및 클로닝 고안에는 Geneious (version 2021.0.1) 및 Snapgene 5.2.3을 사용하였고, 기준 서열로 NC_005089를 사용하였다.

[934]

[935] 3-2. 결과

[936]

[937] DdCBE 플라스미드 조립. DdCBE에서 맞춤 설계된 TALE 서열의 조립을 보다 쉽게 하기 위해, 분할된 DddAtox 반쪽을 코딩하는 발현 플라스미드를 구성하고, 총 424개(6×64개의 3개 서열 인식 + 2×16개의 2개 서열 인식 + 2×4개의 1개 서열 인식)를 사용하는 Golden-Gate 클로닝 시스템을 사용하였다 (도 33a). 다음 표 4와 같이 6개의 TALE 플라스미드와 발현 플라스미드를 같은 튜브에 섞어서 바로 사용할 수 있는 DdCBE 플라스미드를 제작하였고, 15.5-18.5개의 repeat variable diresidue 서열을 갖도록 하였다(도 38).

[938]

[939] [표4]

	Left TALE Target Sequence	Right TALE Target Sequence
ND5 Silent Mutation	5'-T TTTCTACTGGTCCGAT T-3' SEQ ID NO: 277	5'-T TAAATAAAGTTATTAT T-3' SEQ ID NO: 280
ND5 G12918A Mutation	5'-T TGCAGGTATTAATTGCT T-3' SEQ ID NO: 278	5'-T TCCTAACAGGGTTCTAC T-3' SEQ ID NO: 281
ND5 Nonsense Mutation	5'-T TCCCTAAACATAAACTCA T-3' SEQ ID NO: 279	5'-T TGTTGTTGGAGAATA T-3' SEQ ID NO: 282

[940] DdCBE 구성에 대한 서열은 다음 표 5와 같다. 결과적으로 DdCBE는 5' 말단의 보존된 티민 서열을 포함하여 17-20개의 DNA 서열을 인식한다. 따라서 기능이 갖춰진 DdCBE 쌍은 총 32개에서 40개의 DNA 염기 서열을 인식한다.

[941]

[942] [표5]

Left-SOD2 MTS-3×HA-N terminal domain

MALSRVCGTSRQLAPVLGYLGSRQKHSLPDYPYDVPDYAGYPYDVPDYAGYPY
DVPDYAGIRIQDLRTLGYSSQQQEKIKPKVRSTVAQHHEALVGHGFTAHIVALSQ
HPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQ
LDTGQLLKIARGGVTAVEAVHAWRNALTGAPL SEQ ID No: 283

Right-COX8A MTS-3×FLAG-N terminal domain

MASVLTPLLLRGLTGSARRLPVPRAKIHSLDYKDHDGDYKDHDIDYKDDDDKIRIQ
DLRTLGYSSQQQEKIKPKVRSTVAQHHEALVGHGFTAHIVALSQHPAALGTVAVK
YQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQDLDTGQLLKIAR
GGVTAVEAVHAWRNALTGAPL SEQ ID No: 284

C terminal half domain (NG)-G1333-N-UGI

GLTPEQVVAIASNNGGKQALESIVAQLSRPDPALAALTNDHLVALACLGGRPALDA
VKKGLGGSGSGSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSGGG
GGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVML
LTSDAPEYKPPWALVIQDSNGENKIKML SEQ ID No: 285

C terminal half domain (NG)-G1333-C-UGI

GLTPEQVVAIASNNGGKQALESIVAQLSRPDPALAALTNDHLVALACLGGRPALDA
VKKGLGGSGSPTPYPNYANAGHVEGQSALFMRDNGISEGLVFHNNPEGTGCFV
NMTETLLPENAKMTVVPPEGAIPVKRGATGETKVFTGNSNSPKSPTKGGCSGGST
NLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLTSD
APEYKPPWALVIQDSNGENKIKML SEQ ID No: 286

C terminal half domain (NG)-G1397-N-UGI

GLTPEQVVAIASNNGGKQALESIVAQLSRPDPALAALTNDHLVALACLGGRPALDA
VKKGLGGSGSGSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSGGP
TPYPNYANAGHVEGQSALFMRDNGISEGLVFHNNPEGTGCFVNMETLLPENAK
MTVVPPEGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAY
DESTDENVMLTSDAPEYKPPWALVIQDSNGENKIKML SEQ ID No: 287

C terminal half domain (NG)-G1397-C-UGI

GLTPEQVVAIASNNGGKQALESIVAQLSRPDPALAALTNDHLVALACLGGRPALDA
VKKGLGGSGSAIPVKRGATGETKVFTGNSNSPKSPTKGGCSGGSTNLSDIIEKETG
KQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLTSDAPEYKPPWALV
IQDSNGENKIKML SEQ ID No: 288

ND5 silent mutation Left TALE repeat

NLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNNGGKQALET
VQRLLPVLCQAHGLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAI
ASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLC
QDHGLTPAQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNNGGKQA
LETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQV
VAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNNGGKQALETVQRLLPV
LCQDHGLTPAQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGG
KQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTP

DQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNNGGKQALETVQRL
LPVLCQAHGLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGG
KQALETVQRLLPVLCQD SEQ ID No: 289

ND5 silent mutation Right TALE repeat
NLTPEQVVAIASNGGGKQALETVQRLLPVLCOAHGLTPAQVVAIASNIGGKQALETV
QRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCOAHGLTPDQVVAIAS
NIGGKQALETVQRLLPVLCOAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDH
GLTPDQVVAIASNGGGKQALETVQRLLPVLCOAHGLTPDQVVAIASNIGGKQALET
VQRLLPVLCOAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPAQVVAIA
SNIGGKQALETVQRLLPVLCOAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQD
HGLTPEQVVAIASNGGGKQALETVQRLLPVLCOAHGLTPAQVVAIASNGGGKQALET
TVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCOAHGLTPDQVVAI
ASNGGGKQALETVQRLLPVLCOAHGLTPDQVVAIASNIGGGKQALETVQRLLPVLC
QAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCOAHGLTPDQVVAIASNGGGKQA
LETVQRLLPVLCQD SEQ ID No: 290

ND5 point mutation Left TALE repeat
 NLTPAQVVAIASNNGGKQALETVQRLLPVLCAQHGLTPDQVVAIASNNGGKQALET
 VQRLLPVLCAQHGLTPAQVVAIASHDGGKQALETVQRLLPVLCAQHGLTPAQVVAI
 ASNIGGKQALETVQRLLPVLCAQHGLTPAQVVAIASNNGGKQALETVQRLLPVLCO
 DHGLTPAQVVAIASNNGGKQALETVQRLLPVLCAQHGLTPDQVVAIASNNGGKQAL
 ETVQRLLPVLCAQHGLTPDQVVAIASNIGGKQALETVQRLLPVLCDHGLTPDQV
 AIASNNGGKQALETVQRLLPVLCAQHGLTPEQVVAIASNNGGKQALETVQRLLPVL
 CAQHGLTPAQVVAIASNIGGKQALETVQRLLPVLCDHGLTPAQVVAIASNIGGKQA
 LETVQRLLPVLCAQHGLTPEQVVAIASNNGGKQALETVQRLLPVLCAQHGLTPAQV
 VAIASNNGGKQALETVQRLLPVLCAQHGLTPDQVVAIASNNGGKQALETVQRLLPV
 LCDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCAQHGLTPAQVVAIASNNGG
 KQALETVQRLLPVLCD SEQ ID No: 291

ND5 point mutation Right TALE repeat
 NLTPAQVVAIASNNGGKQALETVQRLLPVLCDHGLTPAQVVAIASHDGGKQALET
 VQRLLPVLCAHGLTPEQVVAIASHDGGKQALETVQRLLPVLCAHGLTPEQVVAI
 ASNNGGKQALETVQRLLPVLCAHGLTPAQVVAIASNNGGKQALETVQRLLPVLCD
 DHGLTPAQVVAIASNNGGKQALETVQRLLPVLCAHGLTPEQVVAIASHDGGKQAL
 ETVQRLLPVLCAHGLTPAQVVAIASNNGGKQALETVQRLLPVLCAHGLTPAQV
 AIASNNGGKQALETVQRLLPVLCAHGLTPAQVVAIASNNGGKQALETVQRLLPVL
 CDHGLTPAQVVAIASNNGGKQALETVQRLLPVLCDHGLTPDQVVAIASNNGGK
 QALETVQRLLPVLCAHGLTPEQVVAIASNNGGKQALETVQRLLPVLCAHGLTPD
 QVVAIASHDGGKQALETVQRLLPVLCDHGLTPDQVVAIASNNGGKQALETVQRLL
 PVLCAHGLTPEQVVAIASNNGGKQALETVQRLLPVLCAHGLTPAQVVAIASHDG
 GKQALETVQRLLPVLCD SEQ ID No: 292

ND5 nonsense mutation Left TALE repeat
NLTPAQVVAIASNGGGKQALETVQRLLPVLCDHGLTPAQVVAIASHDGGKQALET
VQRLLPVLCOAHGLTPEQVVAIASHDGGKQALETVQRLLPVLCOAHGLTPDQVVAI
ASHDGGKQALETVQRLLPVLCOAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCOAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCOAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCOAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCOAHGLTPAQV

[944]

```

AIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLC
QDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQ
LETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQV
VAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPV
CQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGK
QALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQD SEQ ID No: 293

ND5 nonsense mutation Right TALE repeat
NLTPAQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALE
TVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPAQVVA
IASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLC
QAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQ
ALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQ
VVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNIGGK
QALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPD
QVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLL
PVLCQA SEQ ID No: 294

```

[945]

[946] 생체 외에서의 미토콘드리아 염기 교정. Golden-Gate 클로닝 시스템을 이용하여 생체 내 미토콘드리아 DNA 교정을 시도하기 위해 *Mus musculus* 미토콘드리아 NADH-유비퀴논 산화환원효소 사슬 5 단백질을 인코딩하는 ND5 유전자를 선정하였다. ND5 단백질은 NADH 탈수소효소 (유비퀴논)의 핵심 서브유닛으로, NADH에서 호흡 사슬로 전자를 전달하는 것을 촉매한다. 인간에서 ND5 유전자의 돌연변이는 MELAS(mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes)와 연관이 있다고 알려져 있으며, Leigh 증후군 또는 LHON(Leber's hereditary optic neuropathy)의 일부 증상과도 연관이 있는 것으로 알려져 있다. 인간에서 기능 부전을 모사할 수 있도록 미토콘드리아 유전자에 유전적 변형이 있는 생쥐 모델을 만들고자 하였다.

[947] 첫번째로, 몇 종류의 DdCBE 플라스미드를 조립하였는데, 이것은 두 가지의 침묵 돌연변이인 m.C12539T와 m.G12542A를 일으키도록 고안되었다. 이러한 플라스미드를 NIH3T3 생쥐 세포주에 형질감염 시키고 3일 후 염기 교정 빈도를 측정하였다. 예상한 것과 같이, 목표 범위의 시토신 염기는 최대 19%의 효율로 티민으로 교정되었다 (도 34a). DddAtox는 "TC" 서열의 시토신만을 배타적으로 탈아미노화 하는 것으로 기존에 보고되었는데, 실험 결과에서도 TC 맥락의 두개의 시토신만이 교정되었다. 인텔이나 다른 종류의 점돌연변이는 교정 목표 범위에서 유의미하게 생성되지 않았다.

[948] 생체 내에서의 미토콘드리아 염기 교정. 가장 효능이 뛰어난 DdCBE 쌍(left-G1397-N 및 right-G1397-C)를 생체 외 실험에 사용하였다. 이 DdCBE 쌍을 인코딩하는 생체 외 전사체를 C57BL6/J 배아의 1세포기에 미세주입한 4일 후, 총 32개의 배아 중 9개의 교정 성공된 배아를 얻을 수 있었다 (28%, 표 6).

[949]

[950] [표6]

Type of mutagenesis	Number of examined embryos	Number of blastocysts (%)	Number of transferred embryos	Number of offspring (%)	Number of edited/total blastocysts (%)	Number of edited/total offspring (%)
N/A (buffer injection)	20	11 (55)	NA	NA	0/11 (0)	NA
ND5 silent	56	32 (57)	30	4 (13)	9 (28)	3 (75)
ND5 G12918A	79	44 (55)	50	11 (22)	11 (25)	4 (36)
ND5 STOP	68	37 (54)	120	27 (23)	19 (51)	9 (33)

[951]

[952] TALE-DddAtox 탈아미노화 효소는 C * G 에서 T * A의 염기 전위를 효율적으로 생성하였는데, m.C12539에서는 2.2 에서 25%의 효율로, m.G12542에서는 0.63에서 5.8%로 그 효율이 나타났다. 이후 DdCBE가 주입된 배아를 대리모에 이식하여 m.C12539T와 m.G12542T를 가진 자손을 얻었다 (도 39). 4마리의 새끼(F0) 중 3마리가 C * G 에서 T * A 교정이 되었고 그 범위는 1에서 27% 까지였다(도 34c). 두 마리의 새끼는 발가락과 꼬리에서 비슷한 돌연변이 수치를 보였고, 태어난지 14일 후에도 그 효율을 유지하고 있었다. 또한, 이러한 미토콘드리아 DNA 돌연변이가 생후 50일이 지난 성체 F0 생쥐의 다양한 조직에서 검출되었다(도 34d). 이러한 결과는 1 세포기 접합체에서 DdCBE로 인해 생긴 미토콘드리아 DNA의 헤테로플라스미가 발달과 분화 과정동안 유지되는 것을 시사한다.

[953] DdCBE로 인해 생긴 돌연변이가 다음세대로 전달되는지 확인하기 위해, 암컷 F0 생쥐와 야생형 C57BL6/J 수컷을 교배시켜 F1 자손을 만들었다. m.C12539T와 m.G12542T 돌연변이는 두 마리의 새끼에서 6-26% 효율로 관찰되었다. 또한 이러한 미토콘드리아 교정이 11개의 각기 다른 조직에서 비슷한 수치로 관찰되었다(도 35b).

[954]

[955] DdCBE로 매개 된 MT-ND5 G12918A 돌연변이. 다음으로, m.G12918A 돌연변이를 제작하고자 했는데, 이 돌연변이는 사람에게서 미토콘드리아 질병을 일으키기도 한다. 주목할 점은 이 돌연변이는 여러 미토콘드리아 질병, 예를들어 Leigh 증후군, MELAS 증후군, 및 LHON 증후군 등을 일으킨다. 이 위치의 시토신 염기는 옆에 티민이 있기 때문에 DdCBE를 이용한 염기 교정이 가능하다 (도 36a). 4쌍의 DdCBE를 조립하고 NIH3T3에서 최대 6.4%까지 교정이

가능함을 확인하였다(도 36b). 이후 가장 효율이 좋은 DdCBE조합을 생쥐 접합체에 미세주입하여 그 효율을 배반포에서 관찰하였다. 44개 중 11개(25%)의 배아가 m.G12918A 돌연변이를 가지고 있었으며 그 효율은 0.25에서 23%까지였다(도 36c). 다음으로, DdCBE가 미세주입된 배아를 대리모에 이식하여 G12918A 돌연변이를 가진 자손을 얻었다(도 39b). 11 마리의 새로 태어난 생쥐 중 4마리가 이 돌연변이를 3.9-31.6% 정도 가진것을 확인하였다(도 36d). 태어난 직후 보여지는 표현형은 없었지만, 아마도 새끼가 매우 어리고 또한 정상 미토콘드리아 DNA와 돌연변이 DNA가 이형 상태로 존재하기 때문에, 이러한 결과는 DdCBE가 미토콘드리아 질병을 가진 동물 모델을 만들 수 있음을 시사한다.

[956]

[957] MT-ND5 년센스 돌연변이. 마지막으로, ND5의 기능 상실 돌연변이가 생쥐에서 유지가 가능한지 유전자에 년센스 돌연변이를 만들어 확인하였다. m.C12336을 목표 시토신으로 하여 이른 종결 코돈을 ND5 단백질의 199번째 위치에 도입하였다(Q199*; 도 37a). 우선 4개의 DdCBE 조합을 NIH3T3 세포주에 형질 감염시켜 염기 교정 효율을 확인하였고, 가장 효과적인 DdCBE 쌍이 년센스 돌연변이를 약 5.7% 효율로 일으키는 것을 확인하였다(도 37b). 이 DdCBE는 시토신-티민 교정을 일으키고, 약간 효율이 낮지만 m.G12341A인 침묵 돌연변이를 생성하는 돌연변이 (Q200Q) 역시 목표 범위에서 교정되는 것을 확인하였다. 37개 중 19개 (=51%)의 생쥐 배아에서 역시 m.C12336T와 m.G12341A 돌연변이를 순서대로 모두 32% 및 23%의 효율로 확인하였다(도 37c).

[958] 이 결과를 바탕으로, 생쥐 배아를 대리모에 이식하고, m.C12336T와 m.G12341A 돌연변이를 갖는 후손을 얻을 수 있었다(도 39c). 총 27개 중 9마리의 F0 생쥐(23%)가 C * G to T * A 교정을 보였으며 그 효율 범위는 0.22에서 57% 인데(도 37d,e), 이를 통해 ND5의 년센스 돌연변이가 배아 도태를 일으키지 않음을 알 수 있다.

[959]

[960] 실시예 4. 동물 미토콘드리아 DNA 교정

[961]

[962] 4-1. 핵외 수송 신호가 결합된 동물 미토콘드리아 염기 교정용 발현 벡터 제작

[963] TALE-DdCBE에 핵외 수송 신호가 결합된 단백질을 동물 세포에서 발현시킬 수 있도록 하는 벡터(도 40)를 제작하였다. 벡터에는 사이토메갈로바이러스 프로모터(CMV promoter)를 사용하였다. N 말단부터 미토콘드리아 전달 신호, 단백질 정제/검출 태그, TALE array N 말단 도메인, 반복 부위, C 말단 도메인, DddA 시토신 탈아미노화 효소 반체(split half), 우라실 글리코실레이즈 억제제 및 핵외수송신호를 포함한다(도 40a). 핵외수송신호는 예를 들어 MVM (Mirute virus of mice)의 NS2 단백질 유래를 사용할 수 있으나 다른 서열도 가능할 것이다.

이를 통해 발현된 단백질은 핵 외부로 방출되며, 미토콘드리아로 전달되어 염기 교정을 하게 된다. 이때 사용된 타겟 DNA는 미토콘드리아 ND5 유전자-4번 염색체 ND5-like 유전자, 미토콘드리아 TrnA-5번 염색체, 미토콘드리아 Rnr2-6번 염색체이다(도 40b).

[964]

[965] 4-2. 동물 세포주에서의 DdCBE-NES

[966] NIH3T3 세포주 (ATCC CRL-1658)를 transfection 하루 전날 오후 1.5×10^4 /well 로 1 ml의 세포 생장 배지(DMEM + 10% Bovine Calf Serum)가 담긴 12-well plate 에 분주하였다. 다음날 아침 DNA를 처리하지 않은 Mock, DdCBE, DdCBE-MVM NES 플라스미드 투입 실험군을 Lipofectamine 3000의 제조사 프로토콜에 따라 세포에 Transfection 하였다. 이후 3일간 이산화탄소 배양기(37°C, 5 % CO₂)에서 배양 후 세포를 걷어서 Qiagen Blood & Tissue Kit을 이용해 전체 DNA를 정제하였다. 이후 미토콘드리아 유전자 특이적인 PCR 프라이머를 이용해 증폭하고, 일루미나 Miniseq 장비를 이용하여 차세대 서열 분석(Next-generation Sequencing)을 수행한 다음 결과를 Cas-analyzer (www.rgenome.net)를 이용하여 염기교정 효율을 확인하였다.

[967] 도 40c,d,e는 DdCBE-NES를 NIH3T3 mouse 세포주에 Transfection 하여 생쥐 미토콘드리아 유전자인 ND5, TrnA, Rnr2에 돌연변이를 만든 것이다. DdCBE의 조합 별로 효율이 달라짐을 알 수 있다.

[968]

[969] 4-3. 동물 세포주에서의 mitoTALEN

[970] 도 40f에 표시된 서열을 인식하는 TALEN을 제작하였고, 이 구조체에 MTS를 연결하여 미토콘드리아에 DdCBE와 함께 도입하였다. 실험방법은 실시예 4-2와 같다. TALE 인식부위에 의도적으로 미스매치를 주어 야생형 mtDNA와는 1개의 미스매치로, 돌연변이 mtDNA와는 2개의 미스매치가 일어나도록 하였다. TALE은 1개의 서열 미스매치는 인식하지 못하는 경우가 있기 때문이다. 그 결과 DdCBE만 처리한 것에 비해 TALEN과 같이 처리한 경우 +2 미스매치 실험군에서 효율이 조금 더 상승하는 것을 확인하였다.

[971]

[972] 4-4. 동물 배아에서의 DdCBE-NES

[973] DdCBE 발현 벡터와 DdCBE-NES 발현 벡터를 주형으로 PCR 기법을 통해 T7 promoter와 DdCBE 또는 DdCBE-NES 발현 부분을 증폭시킨 DNA를 확보한다. 이 DNA를 주형으로 T7 polymerase를 이용하여 mRNA를 합성한다.

[974] 한쌍의 DdCBE mRNA 또는 한쌍의 DdCBE-NES mRNA를 미세주입용 용액에 섞어 생쥐 수정란에 미세주입한다.

[975] 4일 배양한 수정란은 배반포상태가 되고 이 배반포를 용해한다. 이를 주형으로 사용하여 미토콘드리아 DNA에서 핵 DNA와는 구분되는 표적 위치 부분을 PCR을 이용하여 증폭한 다음 추가적으로 인덱스와 시퀀싱 어댑터를 추가적인

PCR을 통해 증폭하였다. 일루미나 Miniseq 장비를 이용하여 고처리량 시퀀싱 (high-throughput sequencing)을 수행한 다음 결과를 Cas-analyzer (www.rgenome.net)를 이용하여 염기교정 효율을 확인하였다. 또한 미토콘드리아 표적부분이 유사한 서열을 가진 핵 속 DNA 를 PCR을 이용하여 증폭, sequencing을 수행하였다.

- [976] 그 결과 DdCBE는 미토콘드리아 DNA뿐만 아니라 핵속의 유사 DNA 서열에서도 변이를 유발하였다. (미토콘드리아 : 13.1 %, 핵 : 3.2 %) DdCBE-NES의 경우 미토콘드리아 표적의 변이 효율을 18.2 %로 높임과 동시에 핵 DNA 변이 효율을 0.2 %로 낮추었다. (도 41a). 다른 타겟인 TrnA 또는 Rnr2의 경우 핵 DNA 돌연변이는 일어나지 않았으나, 미토콘드리아 DNA의 경우 통계적으로 유의미하게 염기 교정 효율이 증가함을 알 수 있었다. (*p<0.05, **p<0.01, n.s. 통계적으로 유의미하지 않음)

[977]

[978] 4-5: 동물 배아에서의 DdCBE 및 mitoTALEN

- [979] DdCBE를 이용한 ND5 유전자 위치에 C to T conversion 이후 교정된 미토콘드리아 DNA가 세포내 차지하는 비중을 높이기 위하여 교정되지 않은 미토콘드리아 DNA 서열을 자르는 TALEN을 함께 주입하였다. 미세주입 방법과 sequencing 서열 확인 방법은 실시예 4-2 및 4-4와 동일하다. DdCBE를 단독으로 미세주입한 그룹은 11 %의 교정 효율이 보였고, mitoTALEN과 동시에 처리한 경우 33.3%로 상승하여 통계적으로 유의미한 교정 효율의 상승을 얻었다. 또한 DdCBE-NES를 단독으로 미세주입한 그룹은 20.5%의 교정 효율을 보였고, mitoTALEN과 동시에 처리한 경우 36.8%의 효율을 관찰하였다. 이 또한 통계적으로 유의미하였다(도 41b).

- [980] 미세주입된 수정란을 대리모에 이식하여 새로 태어난 생쥐 새끼에서도 마찬가지로 DdCBE는 10.9%의 교정 효율을 보였으나 DdCBE-NES 및 mitoTALEN을 같이 처리한 경우 23.4%의 효율을 얻을 수 있었다 (도 41c).

- [981] 동물 미토콘드리아 유전자 교정시에 핵외 수송 서열을 염기 교정 단백질에 부착한 경우 보다 높은 효율로 염기가 교정되며, 동물 배아에서는 특히 핵내 유사 서열의 비특이적 염기 교정 역시 억제되는 것으로 나타났다. 또한 미토콘드리아 서열 절단 단백질을 동시에 이용하는 경우에도 보다 높은 효율의 미토콘드리아 염기 교정을 기대할 수 있다.

[982]

[983] 실시예 5. 분할 DddA_{tox} 탈아미노화 효소 변이체

[984]

- [985] DdCBE의 비표적 효과를 줄일 수 있는 고정밀의 DddA 유래 시토신 염기 편집기를 제시한다. 이러한 비표적 염기 편집 효과는 TALE과 DNA 사이의 상호작용과 무관하게 DddAtox 탈아미노화 효소 분할체의 자발적인 결합으로 야기되는 현상이다. 따라서 DddAtox 분할체 간의 표면에 위치한 아미노산

잔기를 알라닌으로 치환하여 HF-DdCBE를 제작하였으며, HF-DdCBE는 TALE에 연결된 두개의 탈아미노화 효소 쌍이 DNA에 결합하지 못할 경우 제대로 된 기능을 하지 못하도록 하였다. 전 미토콘드리아 유전체 분석을 통해, 사람의 미토콘드리아 DNA에서 수많은 원치않는 비표적 C-to-T 전환을 일으키는 기존 DdCBE와는 달리, HF-DdCBE는 매우 효율적이고 정밀하다는 것을 확인하였다.

[986]

[987] 5-1. 방법

[988] 플라스미드 제작. DdCBE 발현 플라스미드에 점돌연변이를 도입하였다.

플라스미드를 Q5 Site-Directed Mutagenesis (NEB)용 돌연변이 제작 프라이머 (표 7)를 이용해 증폭하였고, 결과는 생거 서열분석법으로 확인하였다.

[989]

[990] [표 7]

Name	Sequence	No.
Left-G1333-N_F1315A_F	AGTGGGGACCGCATACTATGTCAACGAC	295
Left-G1333-N_F1315A_R	GTCTGACCGTTGTAGGCG	296
Left-G1333-N_F1329A_F	ATCCAAGGTTGCATCTAGCGGTGGGTCCGG	297
Left-G1333-N_F1329A_R	TCCAATCCCCGGCGTCG	298
Left-G1333-N_G1313A_F	CAGACAGTGGCAACCTTTTACTATGTCAACGAC	299
Left-G1333-N_G1313A_R	ACCGTTGTAGGCGGGGAG	300
Left-G1333-N_I1299A_F	TCCATATCAGGCAAGTGCTCCCCAACTCCC	301
Left-G1333-N_I1299A_R	CCCAGGGCGTAGGAACCA	302
Left-G1333-N_K1327A_F	ATTGGAATCCGCAGTTTTCTCTAGCGGTGGGTCCGG	303
Left-G1333-N_K1327A_R	CCCCGGCGTCGTTGACA	304
Left-G1333-N_L1294A_F	CTCTGGTTCCTACGCCGCAGGTCCATATCAGATTAG	305
Left-G1333-N_L1294A_R	CTAATCTGATATGGACCTGCGGCGTAGGAACCAGAG	306
Left-G1333-N_S1300A_F	ATATCAGATTGCAGCTCCCCAACTCCCCGC	307
Left-G1333-N_S1300A_R	GGACCCAGGGCGTAGGAA	308
Left-G1333-N_S1330A_F	CAAGGTTTTCGCAAGCGGTGGGTC	309
Left-G1333-N_S1330A_R	GATTCCAATCCCCGGCG	310
Left-G1333-N_T1314A_F	GGTCAGACAGTGGGGGCATTTTACTATGTC	311
Left-G1333-N_T1314A_R	GACATAGTAAATGCCCCACTGTCTGACC	312
Left-G1333-N_V1312A_F	GGTCAGACAGCAGGGACCTTTTACTATG	313
Left-G1333-N_V1312A_R	GTTGTAGGCGGGGAGTTG	314
Left-G1333-N_V1318A_F	TTTTACTATGCAAACGACGCCGGGG	315
Left-G1333-N_V1318A_R	GGTCCCCACTGTCTGACC	316
Left-G1333-N_Y1292A_F	CTCTGGTTCGCGAGCCCTGGGTCCATATCAGATTAGTGCTCC	317
Left-G1333-N_Y1292A_R	CCAGAGCCCCCAACCCC	318
Left-G1333-N_Y1316A_F	CAGTGGGGACCTTTGCATATGTCAACGACGC	319
Left-G1333-N_Y1316A_R	GCGTCGTTGACATATGCAAAGGTCCCCACTG	320
Left-G1333-N_Y1317A_F	GACCTTTTACGCAGTCAACGACGCCGGG	321
Left-G1333-N_Y1317A_R	CCCACTGTCTGACCGTTG	322
Left-G1397-N_C1376A_F	TGTGGCTTCGCAGTGAATATGACTGAAACCCTTCTG	323
Left-G1397-N_C1376A_R	CATATTCAGTCCGAAGCCACAAGTACCCTCTG	324
Left-G1397-N_E1381A_F	GTGTGAATATGACTGCAACCCTTCTGCCGAAAATG	325
Left-G1397-N_E1381A_R	CATTTTCGGGCAGAAGGGTTGCAGTCATATTCACAC	326
Left-G1397-N_K1389A_F	GAAAATGCCGCAATGACTGTCGTCACCTGAAGG	327

[991]

Left-G1397-N_K1389A_R	GACAGTCATTGCGGCATTTTCGGGCAGAAGG	328
Left-G1397-N_L1384A_F	GAAACCCTTGACCCGAAAATGCCAAGATGAC	329
Left-G1397-N_L1384A_R	ATTTTCGGGTGCAAGGGTTTCAGTCATATTCACACAGAAG	330
Left-G1397-N_M1390A_F	AATGCCAAGGCAACTGTCGTCCACCTGAAGG	331
Left-G1397-N_M1390A_R	GACGACAGTTGCCTTGGCATTTTCGGGCAGAAGG	332
Left-G1397-N_T1380A_F	GTGAATATGGCAGAAACCTTCTGCCCGAAAATG	333
Left-G1397-N_T1380A_R	AAGGGTTTCTGCCATATTCACACAGAAGCCACAAG	334
Left-G1397-N_T1391A_F	CGAAAATGCCAAGATGGCAGTCGTCCACCTGAAG	335
Left-G1397-N_T1391A_R	CTTCAGGTGGGACGACTGCCATCTTGGCATTTTCG	336
Left-G1397-N_V1377A_F	GGTACTTGTGGCTTCTGTGCAAATATGACTGAAACC	337
Left-G1397-N_V1377A_R	GGTTTCAGTCATATTTGCACAGAAGCCACAAGTACC	338
Left-G1397-N_V1392A_F	GAAAATGCCAAGATGACTGCAGTCCACCTGAAGG	339
Left-G1397-N_V1392A_R	CCTTCAGGTGGGACTGCAGTCATCTTGGCATTTTC	340
Right-G1333-C_E1361A_F	GGCATTAGCGCAGGGCTTGTGTTCCATAATAATCC	341
Right-G1333-C_E1361A_R	CACAAGCCCTGCGCTAATGCCGTTATCTCTCATAAAAAGAGC	342
Right-G1333-C_F1353A_F	GTCAGCTCTTGCAATGAGAGATAACGGCATTAGCGAAGGG	343
Right-G1333-C_F1353A_R	TGCCCCTCGACGTGCCCG	344
Right-G1333-C_F1365A_F	GGGCTTGTGGCACATAATAATCCTGAGGGCACCTG	345
Right-G1333-C_F1365A_R	GATTATTATGTGCCACAAGCCCTTCGCTAATG	346
Right-G1333-C_G1362A_F	CATTAGCGAAGCACTTGTGTTCC	347
Right-G1333-C_G1362A_R	CCGTTATCTCTCATAAAAAG	348
Right-G1333-C_H1366A_F	GCTTGTGTTGCAAATAATCCTGAGGGCAC	349
Right-G1333-C_H1366A_R	CCTTCGCTAATGCCGTTA	350
Right-G1333-C_K1389A_F	TGAGAACGCTGCAATGACTGTCTG	351
Right-G1333-C_K1389A_R	GGCAGAAGTGTTTCGGTC	352
Right-G1333-C_L1363A_F	TAGCGAAGGGGCGAGTGTTCATAATAATCCTGAGGGC	353
Right-G1333-C_L1363A_R	ATGGAACACTGCCCCCTTCGCTAATGCCGTTATCTC	354
Right-G1333-C_M1354A_F	AGCTCTTTTGAAGAGATAACGGCATTAGCGAAGG	355
Right-G1333-C_M1354A_R	GACTGCCCCCTCGACGTGC	356
Right-G1333-C_M1390A_F	GAACGCTAAAGCAACTGTCGTACCAC	357
Right-G1333-C_M1390A_R	TCAGGCAGAAGTGTTTCG	358
Right-G1333-C_Q1349A_F	CGTCGAGGGGGCATCAGCTCTTTTATGAGAGATAACGGC	359
Right-G1333-C_Q1349A_R	TGCCCCGGCGTTAGCGTAG	360
Right-G1333-C_S1350A_F	CGAGGGGCAGGCAGCTCTTTTATG	361
Right-G1333-C_S1350A_R	ACGTGCCCCGGCGTTAGCG	362
Right-G1333-C_T1391A_F	CGCTAAAATGGCAGTCGTACCAC	363
Right-G1333-C_T1391A_R	TTCTCAGGCAGAAGTGTTTC	364

[992]	Right-G1333-C_V1346A_F	GCCGGGCACGCAGAGGGGCAGT	365
	Right-G1333-C_V1346A_R	GTTAGCGTAGTTAGGATAAGGTGTTG	366
	Right-G1333-C_V1364A_F	CGAAGGGCTTGCATTCCATAATAATCC	367
	Right-G1333-C_V1364A_R	CTAATGCCGTTATCTCTC	368
	Right-G1333-C_V1393A_F	AATGACTGTGCGACCAACCCGAAG	369
	Right-G1333-C_V1393A_R	TTAGCGTTCTCAGGCAGAAG	370
	Right-G1397-C_F1412A_F	ACAAAGGTGGCAACAGGCAACTCTAACAGTCCAAAG	371
	Right-G1397-C_F1412A_R	GTTGCCTGTTGCCACCTTTGTCTCCCCTGTTGCTCC	372
	Right-G1397-C_K1410A_F	CAACAGGGGAGACAGCAGTGTTACAGGCAA	373
	Right-G1397-C_K1410A_R	TTGCCTGTGAACACTGCTGTCTCCCCTGTTG	374
	Right-G1397-C_T1413A_F	GGAGACAAAGGTGTTGCGCAGGCAACTCTAACAGTCC	375
	Right-G1397-C_T1413A_R	GGACTGTTAGAGTTGCCCTGCGAACACCTTTGTCTCC	376
	Right-G1397-C_V1411A_F	GAGACAAAGGCATTACAGGCAACTCTAACAGTC	377
	Right-G1397-C_V1411A_R	GCCTGTGAATGCCTTTGTCTCCCCTGTTGCTCC	378

[993]

[994] 결합 표면 돌연변이의 조립을 위해 미니프랩 된 돌연변이 발현 플라스미드에 모듈 벡터(각각은 TALE 배열을 코딩함), BsaI-HFv2 (10 U), T4 DNA ligase (200 U) 및 반응 버퍼를 한 튜브에 담았다. 이후 제한 효소 및 라이게이즈 반응은 서모사이클러에서 37°C 5분, 50°C 5분의 20 반복 후 50°C를 15분, 80°C를 5분동안 처리하였다. 접합 된 플라스미드는 화학적 형질전환을 통해 E coli DH5α에 도입되었고, 최종 구조체는 생거 서열분석법으로 확인되었다. 세포주 도입을 위해 플라스미드는 미니프랩 하였다.

[995]

[996] 포유류 세포주 배양 및 형질 감염. HEK 293T/17 (CRL-11268, American Type Culture Collection (ATCC)) 세포주는 37 °C의 5 % CO₂ 환경에서 배양되었다. 세포주는 DMEM supplemented with 10 % (v/v) fetal bovine serum (Gibco)의 배지에서 항생제 없이 성장하였고 마이코플라즈마 테스트는 하지 않았다. Lipofection을 위해 세포는 24-well cell culture plates (SPL, Seoul, Korea)에 1 X10⁵ 세포 밀도로 형질 감염 18-24 시간 전 성장을 시작하였다. Lipofectamine 2000 (Invitrogen)을 이용하여 각각의 DdCBE 분할체를 500 ng 사용하여 총 1,000 ng 의 플라스미드 DNA를 도입하였다. 세포는 형질감염 4일 후 회수되었다.

[997]

고처리 서열 분석을 위한 유전체 및 미토콘드리아 DNA 분리. 유전체 DNA를 분리하기 위해 세포 배양 배지를 제거한 후, DNeasy Blood & Tissue Kit (Qiagen)의 Proteinase K가 첨가된 lysis buffer를 세포를 바닥에서 분리하기 위해 세포 배양 플레이트에 넣어주었다. 이후 제조사의 프로토콜을 따라 유전체 DNA를 분리하였다. 전 미토콘드리아 유전체 서열 분석을 위해서는 200 μl의 Mitochondrial isolation buffer A (ScienCell)를 세포 배양 배지가 제거 된 배양 플레이트에 넣었다. 세포를 cell lifter를 이용해 긁은 후 마이크로튜브에 담고 일회용 막자를 이용해 세포를 갈아주었다. 20회의 분쇄 후 잘 갈린 균질액은 1,000×g, 4°C에서 5분동안 원심분리 되었다. 상층액을 새로운 마이크로튜브에

담고 10,000×g, 4°C에서 20분동안 원심분리하였다. 침전물을 10 µl 의 용해액(25 mM NaOH, 0.2 mM EDTA, pH 10)에 풀어준 후 95°C에서 20분간 끓여주었다. pH를 낮추기 위해 1 µl의 1 M HEPES (free acids, without pH adjustment)를 미토콘드리아 용해액에 추가하였다. 이와 같이 준비된 용액의 1 µl를 고처리 서열 분석을 위한 PCR 주형 가닥으로 사용하였다.

[998]

[999] 고처리 서열 분석. 딥 시퀀싱 라이브러리를 제작하기 위해 Q5 DNA Polymerase를 사용하여 nested 1차 PCR 및 2차 PCR을 진행하였고, 최종 인덱스 서열을 추가하였다. 라이브러리는 MiniSeq (Illumina)를 이용해 페어드 엔드 리드 서열 분석에 사용하였다. 전 미토콘드리아 유전체 분석을 위해서는 분리된 미토콘드리아 DNA를 tagmentation DNA prep kit (Illumina)을 이용하여 제조사 프로토콜을 따라 준비하였다. 모든 분석에서 페어드 엔드 서열분석 결과는 하나의 fastqjoin 파일을 이용해서 합쳐졌으며 CRISPR RGEN Tools (<http://www.rgenome.net/>)을 이용해 분석하였다.

[1000]

[1001] 5-2. 결과

[1002]

[1003] 식물체 안에서 엽록체 교정을 시도할 경우 비표적 염기 돌연변이가 엽록체 유전체 상에서 나타나게 되었고, 따라서 DdCBE의 정확성에 대한 의문을 제기하게 되었다. DdCBE의 비표적 염기 교정에 대해 두 가지 이유를 생각해 보았다. 첫번째로는 TALE 단백질과 DNA 사이의 비특이적인 결합이고, 다른 하나는 자연적이며 의도치 않게 일어나는 DddAtox 반체끼리의 결합이다 (도 42a). 본 연구에서는 DddAtox 분할 반체에 초점을 맞추고, 분할된 두 개의 단백질의 결합 표면을 개선하여 원치 않는 DddAtox 반체의 결합을 방지하도록 하였다.

[1004] 첫번째로, 미토콘드리아의 ND1 (mtND1) 유전자를 대상으로 각각의 서브유닛(Left-TALE 혹은 Right-TALE)이 각각 혼자서도 DNA에 결합, TALE이 없는 TALE-free DddAtox 또는 결합 서열이 없는 TALE-DddAtox의 다른 반체와 상호작용하여 시토신-티민 염기 교정을 일으키는지 확인해 보고자하였다. 인간 신장 배아 세포주 (HEK293T)에서 인간 미토콘드리아 ND1 (*mtND1*) 유전자를 목표 서열로 하는 DdCBE 쌍 (Left-TALE:G1397N (N-말단 쪽 G1397 DddAtox 반체로 TALE 배열의 C 말단에 결합하여 전체 인식 서열 중 왼쪽 부위를 맡는 구조) + Right-TALE:G1397C (C 말단 쪽 G1397 DddAtox 반체로 TALE 배열의 C 말단에 결합하여 전체 인식 서열 중 오른쪽 부위를 맡는 구조))은 60.7 %의 효율로 표적 서열의 C11을 효과적으로 시토신에서 티민으로 교정하였다(도 43a). 또한, 각각의 서브유닛과 쌍을 이룰 수 있는 TALE-free DddAtox 반체와 함께 처리하였을 때, 비록 원래의 DdCBE 쌍보다는 못하지만 염기 교정이 일어나는 것을 관찰하였다. 따라서, Left-TALE과 Right-TALE 각각은 목표 ND1

서열에 결합하여 31% 혹은 8.1%의 효율로 염기를 교정함을 알 수 있었다(도 43). 다시 말하면, 두 TALE 단백질과 각각 융합된 DdCBE 쌍은 한쪽의 TALE과만 결합하였을 때보다 2.0 배 ($= 60.7\%/31\%$) 또는 7.5 배 ($= 60.7\%/8.1\%$) 효율이 높음을 할 수 있다. 명백하게도 TALE과 결합된 DddAtox의 N 말단 반체는 TALE 서열이 없는 DddAtox의 C 말단 반체와 결합하여 탈아미노화 반응할 수 있도록 하며, 반대의 경우도 마찬가지였다.

- [1005] DddAtox는 두 위치(G1333 및 G1397)에서 나누어져 반체를 형성하므로, G1333 위치에서 mtND1 유전자를 목표로 하는 DdCBE 쌍을 제작(Left-TALE:G1333-N 및 Right-TALE:G1333-C) 하였고, Left-TALE:G1333-N 및 Right-TALE:G1333-C 구조체를 쌍을 이루는 TALE-free DddAtox 반체와 결합하도록 각각 독립적으로 이용하여 시토신-티민 교정을 일으키는 지 확인하였다. 예상했던 바와 같이 각각의 TALE 융합 DdCBE는 C8 위치에서 32.7 % (왼쪽 TALE 결합체) 또는 18.1% (오른쪽 TALE 결합체)의 염기 교정 효율을 보였고, 이는 원래의 DdCBE 쌍이 보여준 56.1%와 비교할 수 있다. 따라서 TALE이 결합된 DdCBE 쌍은 각각 따로 작용할 때에 비해 1.7배 ($= 56.1\%/32.7\%$) 혹은 3.1배 ($= 56.1\%/18.1\%$)의 효율로 작용한다는 것을 알 수 있었다. 종합하자면 이러한 결과는 DdCBE는 한쪽 TALE 배열만 결합하는 위치에서도 원치 않는 비표적 염기 교정을 일으킨다는 사실을 제시한다. TALE 단백질은 몇 개 가량의 불일치 서열이 있어도 DNA와 결합하기 때문에 DdCBE는 세포 소기관 또는 핵 내 유전자에 비표적 염기 교정을 일으킬 수 있는 가능성이 존재한다.
- [1006] 분할된 DddAtox 반체가 서로 결합함으로써 생기는 비표적 염기 교정을 줄이고자 고정밀의 DdCBE를 개발하고자 하였다. 분할 이합체의 결합 표면 부위를 개량한다면 자가 결합을 방지하거나 저해할 수 있을 것이라 생각하였다. 따라서 PyMOL 소프트웨어에서 파이썬 코드(InterfaceResidues.py)를 이용하여 1 평방 옹스트롬의 범위 안에서 두 분할체 DddAtox (분할 위치 G1333 및 G1397)에서 결합 표면의 아미노산 잔기를 발견하였다. 그 결과 G1397-N(G1397위치에서 분할한 N 말단 DddAtox 반체)에서 9개의 아미노산 잔기를, G1397-C(G1397위치에서 분할한 C 말단 DddAtox 반체)에서 4개의 잔기를, G1333-N(G1333위치에서 분할한 N 말단 DddAtox 반체)에서 14개의 아미노산을, G1333-C(G1333위치에서 분할한 C 말단 DddAtox 반체)에서 15개의 아미노산 잔기를 찾을 수 있었다(도 42b 및 42c).
- [1007] 이후 찾은 아미노산 잔기를 알라닌으로 치환한 여러개의 돌연변이 DddAtox 반체를 제작하고, 이러한 결합 표면 돌연변이 DdCBE를 각각 야생형 DdCBE 짝 혹은 TALE이 없는 DddAtox 짝과 함께 HEK293T 세포주에 처리하여 염기 교정 효율을 관찰하였다. G1397 분할의 여러 결합 표면 돌연변이, C1376A, M1390A 및 F1412A를 포함하는, 에서는 TALE-free DddAtox 뿐만 아니라 야생형과 짝지었을 때에도 시토신-티민 교정 효율을 목표 위치에서 관찰할 수 없었는데, 이는 이러한 돌연변이들은 다른 DddAtox 반체와 가까이 위치하여도

상호작용하지 못한다는 것을 의미한다. 몇몇의 돌연변이들, 예를 들어 E1381A 및 V1377A, 의 경우 TALE-free 반체와 결합하여 높은 효율의 염기 교정을 보였는데, 이는 이러한 돌연변이는 이합체의 상호작용과는 무관한 변화임을 보여준다.

- [1008] 중요하게도 몇몇 돌연변이, 예를 들어 K1389A, K1410A, 및 T1413A의 경우 야생형의 DdCBE 짝과 결합 시에는 높은 활성을 보였지만 TALE-free 짝과 사용 시에는 활성이 낮았다. 일례로 K1410A 돌연변이는 53.2%의 효율을 보여 야생형 DdCBE 짝(60.7%)에 비해 비슷한 결과를 얻었으나, TALE-free 와 함께 사용시에는 0.9%의 효율을 보여주어 59.1배의 차이(= 53.2%/0.9%)를 나타내었다. 상기했던 것과 같이 야생형의 경우 7.5 배 (= 60.7%/8.1%)의 차이를 보인다. 또한, 이런 돌연변이들은 야생형의 DdCBE보다 선택적으로 염기를 교정하였다. 변이체는 교정 범위 내에서 C₈, C₉, 또는 C₁₃보다 C₁₁에 대해 강한 선택성을 보여줬는데, 이는 야생형에서 4개의 시토신을 고르게 6.7% 이상으로 모두 교정한 것과 비교될 수 있다(도 43b).
- [1009] 또한 G1333에서 29개의 돌연변이(14 개 G1333N 및 15 개의 G1333C)를 스크리닝 하여 몇몇의 적합한 결합 표면 돌연변이를 얻을 수 있었다(도 44). 다수의 돌연변이들은 야생형과 짝지어졌을 때 효율이 저조하거나(I1299A, Y1316A, Y1317A, 및 F1329A), TALE-free와 짝지어졌을 때 효율이 증가한 것들(S1300A 및 T1314A) 이었다. 하지만 주목할만한 점은 K1389A, T1391A, 및 V1393A와 같은 돌연변이는 야생형과의 결합에서는 염기 교정이 잘 일어나며, TALE-free와는 저조한 효율을 나타내었다. 일례로 K1389A의 경우 38배의 차이(= 45.4%/1.2%)를 나타내는 반면 야생형 DdCBE 쌍은 약 3.1배의 차이(= 56.1%/18.1%)에 불과하다. 또한 K1389A의 경우 야생형보다 효율이 좋은 것처럼 관찰되기도 하였다. 또한 이 돌연변이들은 C₁₁이나 C₁₃이 아닌 C₈ 위치에 교정 효율이 집중되는 것으로 관찰되었는데, 이는 야생형 DdCBE의 19% 이상의 모든 시토신 위치에 교정 효율이 올라가는 것과 비교할 수 있다. 또한 주목해야 할 점은 G1397-C의 K1410A는 C8을 우선적으로 교정하는 반면에 G1333-C의 K1389A의 경우 C11을 선택적으로 더 고효율로 교정하는 것을 알 수 있다. 반면에 야생형의 DdCBE 쌍들(G1333 또는 G1397 DddA_{tox} 분할)은 모두 이러한 선택성이 저조함을 알 수 있다. 이러한 결과는 위에 기재된 결합 표면 돌연변이들은 DdCBE에서 표적 부위 내 여러 염기에 원치 않는 교정을 일으키는 효과를 줄일 수 있는 가능성을 보여준다.

[1010]

[1011] 실시예 6. 전장 탈아미노화 효소

[1012]

[1013] 분할된 세균 독소 DddAtox, TALE array, 우라실 글리코실라제 억제제(UGI)를 포함하는 DddA 유래 시토신 염기교정효소(DdCBE)는 진핵세포의 핵, 미토콘드리아 DNA(mtDNA), 식물의 엽록체 DNA의 표적 시토신을 티민으로

변환 가능하게 한다. 박테리아에 독성을 유발하는 DddAtox는 *Burkholderia cenocepacia*에서 유래했으며, 효소로 이중 가닥 DNA(double-strand DNA) 내에서 시토신을 탈아미노화한다. 숙주 세포의 독성을 피하기 위해 DddAtox는 비활성을 가지는 반으로 나누어, 각각 TALE DNA 결합 단백질에 결합하여 DdCBE를 만든다. TALE array에 결합된 두 개의 비활성 형태는 TALE에 의해 표적 DNA에 근처에서 결합될 때에 기능을 가진다. 시토신에서 티민으로 염기 변환은 두 TALE 결합 부위 사이의 14~18개 염기 사이의 영역에서 유도된다.

- [1014] DdCBE는 세포 소기관 DNA를 교정하지 못하는 CRISPR 유래 염기 교정기와 달리 핵 및 세포 소기관 DNA 모두에서 표적화된 염기 교정을 가능하게 하지만, 이를 유도하기 위해 하나가 아닌 두 개의 TALE 구성이 필요하다는 단점이 있다. 첫번째 단점은, TALE이 5' 말단과 3' 말단 모두에서 티민으로 표적 DNA 부위에 결합해야 하기 때문에 두 개의 TALE array를 사용하면 표적 가능한 부위가 제한된다. 둘째, 하나가 아닌 두 개의 TALE 구성을 전달하는 것은 종종 비효율적이고 도전적이다. 유전자 치료에 널리 사용되는 아데노 관련 바이러스(AAV) 벡터(용량, ~4.7kbps)와 같이 용량이 제한된 바이러스 벡터는 이량체 DdCBE의 조합이 너무 커(2 X 4.1kbps, 프로모터 및 polyA 신호 포함) 두 개의 DdCBE 인코딩 서열을 수용할 수 없다. 더 큰 용량의 단일 벡터에 두 개의 TALE array DNA를 클로닝하는 것은 두 개의 TALE array 서열의 높은 유사성으로 인해 어려울 수 있다. 마지막으로 하나가 아닌 두 개의 TALE array를 사용하면 오프타겟 효과가 악화될 수 있다. 분할 DddAtox가 있는 이량체 DdCBE의 이러한 한계를 극복하기 위해서 무독성 전장 DddAtox 형태의 mDdCBE(단량체 DdCBE)를 핵 및 소기관의 표적 DNA에서 시토신을 티민으로 변환을 유도하는 DdCBE를 제시한다.

[1015]

[1016] 6-1. 방법

[1017]

- [1018] 플라스미드 제작. DddA 변이체는 합성된 전장 DddAtox(gBlock, IDT)를 주형으로 사용하고 표 8의 프라이머 및 Q5 DNA 중합효소(NEB)를 사용하여 PCR 증폭 하였다. 이러한 PCR 산물은 Apobec1이 BamHI과 Sma I(NEB)로 절단된 p3s-BE3 부위에서 Gibson 어셈블리(NEB)를 사용하여 클로닝 하였다. TALE-DddAtox(Addgene #158093, #158095, #157842, #157841)는 플라스미드를 BamHI과 Sma I로 절단하고 DddA 변이체를 표 8의 프라이머를 사용하여 PCR 증폭하고 Gibson 어셈블리를 사용하여 클로닝했다. 확보한 플라스미드를 화학적 방법으로 제조된 *E. coli* DH5α에 열 충격 방법으로 형질전환하고, 살아남은 콜로니의 플라스미드 염기서열을 Sanger 염기서열분석법으로 분석하였다. 최종 플라스미드는 세포 형질감염을 위해 midi-prepped(Macherey-Nagel) 하였다.

[1019]

[1020] [표 8]

Name	Sequence	No.
DddA F	cccaagcttgcaccatgggatccggcagctacgccctgggtccgtatcag	380
DddA R	cccgggagttctcgtgcccgtgcaaccgcctttggctcgggtcttctggg	381
K1402A R	cccgggagttctcgtgcccgtgcaaccgcctttggctcgggtcttctgggtgtgtgtaccggtaaaca tttggtttcaccggtagcggcgcGCcactggaatgg	382
R1403A R	cccgggagttctcgtgcccgtgcaaccgcctttggctcgggtcttctgggtgtgtgtaccggtaaaca tttggtttcaccggtagcggcGCccttctcactgg	383
K1402A R1403A R	cccgggagttctcgtgcccgtgcaaccgcctttggctcgggtcttctgggtgtgtgtaccggtaaaca tttggtttcaccggtagcggcGCCGCcactggaatgg	384
K1410A R	cccgggagttctcgtgcccgtgcaaccgcctttggctcgggtcttctgggtgtgtgtaccggtaaaca GGCggttcacc	385
K1420A R	cccgggagttctcgtgcccgtgcaaccgcctttggctcgggtcGCCgggtgtgtgtgt	386
K1424A R	cccgggagttctcgtgcccgtgcaaccgccGGCggtcgggtcttctggg	387
E1347 F	ctatgccaatgccgggtcatgtgGCCggtcagagcggcctgttcag	388
E1347 R	catgaacaggcgctctgaccGGCcacatgaccggcattggcatag	389
error_prone PCR F	tgagccagctggggcggcgacagcggcagcagactccgggacctcagagtcgccacacccgaaag tggcagctacgcccctgggtcc	390
error_prone PCR R	aatagggccctctagatgcatgctcagtttagcaaccgcctttgggtcggg	391
TAIL Right DddA F	tgaagaaaggtctggggcggatccggcagctacgccctgggtccgtatcag	392
TAIL_Left_D ddA_F	tgaagaaaggcctgggtggatccggcagctacgccctgggtccgtatcag	393
TAIL DddA R	tcagattagttgagccgccagagcaaccgcctttgggtcgggtcttctggg	394
TAIL UGI F	tctggcggctcaactaatctgagcgacatcattg	395
TAIL UGI R	cagatccgaaaatggatatacaagctccc	396

[1021] 무작위 돌연변이 유발. GeneMorph ii Random mutagenesis kit (Agilent)를 사용하여 제조업체의 프로토콜에 따라 합성된 전체 길이 DddAtox (gBlock, IDT)를 주형으로 하여 오류가 발생하기 쉬운 PCR을 하였다. 요약하면, DddAtox DNA 1ng, 100ng, 700ng을 주형으로 사용하여 각각 0~16 mutations/kb의 무작위 변이를 도입하였다. 전체 길이 DddAtox gBlock은 표 8의 프라이머를 사용하여 사전에 PCR 증폭 하였다. 전체 PCR 결과물을 합하여 SmaI 및 XhoI으로 절단된 p3s-UGI-Cas9(H840A)에 Gibson 어셈블리 (NEB)를 사용하여 클로닝하였다. 화학적 방법으로 제조된 E. coli DH5α에 열 충격 방법으로 세포에 플라스미드를 형질전환하고, 생존 한 콜로니가 가지고 있는 플라스미드 염기서열을 Sanger 염기서열분석법으로 분석하였다. 분석한 플라스미드 중 코딩 프레임을 가지는 p3s-UGI-nCas9(H840A)-DddAtox 플라스미드를 sgRNA와 함께 HEK293T 세포에

형질감염 후 표적 딥 시퀀싱을 통한 교정 활성 확인하였다.

[1022]

[1023] 포유류 세포 배양 및 형질감염. HEK293T(ATCC, CRL-11268) 세포 및 HeLa(ATCC, CCL-2) 세포를 5% CO₂에서 37°C로 유지하여 배양하였다. 세포를 10%(v/v) 태아 소 혈청(Welgene) 및 1% 페니실린/스트렙토마이신(Welgene)이 보충된 DMEM에서 배양하였다. 세포를 형질감염 24시간 전에 3X105 세포(HEK293T) 및 4x104 세포(HeLa)의 밀도로 48웰 플레이트(Corning)에 시딩하고 Lipofectamine 2000(Invitrogen) 및 Cas9 융합 DddA 플라스미드(750ng)와 sgRNA(250ng)로 형질감염 하였다. TALE-DddA는 200ng 플라스미드와 Lipofectamine 2000을 사용하여 HEK293T 세포에 형질감염 하였다. sgRNA 서열은 표 9에 나타내었다.

[1024]

[1025] [표9]

Name	sgRNA sequence 5' to 3'	No.
HEK3	GGCCCAGACTGAGCACGTGA	397
TYRO3	GGCCACACTAGCGTTGCTGC	398
ROR1 site1	GCCATAGATGGTGGACCGAA	399
ROR1 site2	CCATCTATGGCTCTCGGCTG	400
ROR1 site3	CCGCAGCCGAGAGCCATAGA	401
FANCF	GGAATCCCTTCTGCAGCACC	402
HBB	CTTGCCCCACAGGGCAGTAA	403
EMX1-1	TGCCCCCTCCCTCCCTGGCCC	404
EMX1-2	CCCTGGCCCAGGTGAAGGTG	405
EMX1-3	GTGAAGGTGTGGTCCAGAAC	406
EMX1-4	AAAGTACAAACGGCAGAAGC	407
TRAC5 site1	GTGGTAGCGGAACTCACTAAG	408
TRAC5 site2	CACCCAGCCTGCTCTGCCTT	409

[1026] 게놈 및 미토콘드리아 DNA 준비. Cas9 융합 DddA 변이체에 형질감염된 세포를 형질감염 2일 후에 수확하고, TALE-DddA에 형질감염된 세포를 3일 후에 수확하였다. 게놈 및 미토콘드리아 DNA는 DNeasy 혈액 및 조직 키트(Qiagen)를 사용하여 분리하였다. 대규모 분석을 위해 5 µL의 프로테이나제 K(Qiagen)가 포함된 100 µL의 세포 용해 완충액(50 mM Tris-HCl, pH 8.0(Sigma-Aldrich), 1 mM EDTA(Sigma-Aldrich), 0.005% sodium dodecyl sulfate(Sigma-Aldrich))을 사용하여 DNA를 추출했다. 용해물을 55°C에서 1시간 반응 후, 95°C에서 10분 동안 반응하였다.

[1027]

[1028] 6-2. 결과

[1029]

[1030] Wild type과 신규 전장 DddA의 아미노산 염기서열을 비교하고 도 45에 변경된 아미노산을 회색 박스로 표시하였다.

[1031] 도 46과 같이, DddA를 Cas9의 N-말단 앞에 16개의 아미노산으로 이루어진 linker를 이용하여 연결하였고, C-말단에는 UGI (uracil glycosylase inhibitor)와 NLS (nuclear localization signal)을 4개의 아미노산으로 이루어진 linker를 이용하여 연결하였다. 반대로 DddA를 Cas9의 C-말단 뒤에 16개의 아미노산으로 이루어진 linker를 이용하여 연결하였고, N-말단에는 UGI와 NLS를 4개의 아미노산으로 이루어진 linker를 이용하여 연결하였다.

[1032] 본 발명에서는 DddA-Cas9(D10A, D10A and H840A)-UGI를 사용하여 실험하였다. DddA를 DNA 결합 단백질인 Zinc finger protein, TALE 모듈에 결합하여 하나의 모듈만으로 시토신을 티민으로 치환할 수 있다. 기존의 split은 두개의 모듈이 있어야 하지만, full length DddA는 한 개의 모듈만 있어도 가능하다. 이 두 DNA 결합 단백질은 NLS (nuclear localization signal), MTS (mitochondrial targeting sequence), CTP (chloroplast transit peptide)를 연결하여 genome region 뿐만 아니라, Cas9으로는 불가능한 mitochondria와 식물의 chloroplast의 시토신을 티민으로 치환할 수 있다. 도 47에 나타낸 바와 같이, human cell genomic context ROR1 site (a), HEK3 site (b), TYRO3 site (c)에서 TC motif의 cytosine이 thymine으로 치환되는 활성을 확인하였다. Target 위치에서 25bp 떨어진 TC motif의 cytosine이 thymine으로 치환되는 활성을 확인하였다 (a). A1341D KRKKA는 CC motif의 두 번째 cytosine이 thymine으로 치환하는 활성을 확인하였다 (a, b). Catalytic mutant인 E1347A도 TC motif의 cytosine이 thymine으로 치환되는 활성을 확인함(a, b, c). 빨간색 밑줄은 Cas9이 binding하는 부위를 나타내었다. 효율은 전체 sequencing read중 indel이 없는 read 중에서 시토신이 티민으로 치환된 비율을 퍼센트로 나타내었다. 또한 전체 sequencing read들 중 indel의 비율을 퍼센트로 나타내었다.

[1033] 도 48에 나타낸 바와 같이, 빨간 네모 박스는 split해서 DddAtox의 활성을 확인한 부분을 나타낸 것이고, 같은 부분을 full length DddA를 이용하여 3개의 target site로 나눠서 활성을 측정하였다. Split은 PAM이 다른 orthogonal Cas9을 사용하여 두 Cas9 사이에 있는 시토신을 티민으로 치환하는데 이 경우 원하는 시토신을 티민으로 정확하게 치환하기 힘들다. 하지만 Full length DddA는 같은 target site에 Cas9이 binding 하는 부분을 3개로 나눠 target 할 수 있어 원하는 시토신을 티민으로 정확하게 치환 할 수 있다. 효율은 전체 sequencing read중 indel이 없는 reads 중에서 시토신이 티민으로 치환된 비율을 퍼센트로 나타내었다. 또한 전체 sequencing read들 중 indel의 비율을 퍼센트로 나타내었다.

[1034] 도 49에 나타낸 바와 같이, Human cell genomic context TRAC site1 (a), TRAC site 2 (b), FANCF (c), HBB (d)에서 full length DddA의 활성을 측정하였다. 빨간색 밑줄은 Cas9이 binding하는 부위를 나타내었다. 효율은 전체 sequencing read중

- indel이 없는 reads 중에서 시토신이 티민으로 치환된 비율을 퍼센트로 나타내었다. 또한 전체 sequencing read들 중 indel의 비율을 퍼센트로 나타내었다.
- [1035] 도 50에 나타난 바와 같이, DddA-dCas9(D10A, H840A)-UGI를 이용한 Human cell genomic context TYRO3 (a), ROR1 (b), HEK3 (c), EMX1 site 2 (d), TRAC site 1 (e), HBB (f)에서 DddA의 활성을 측정하였다. 효율은 전체 sequencing read중 시토신이 티민으로 치환된 비율을 퍼센트로 나타내었다. Indels은 관찰되지 않았다.
- [1036] 염기 교정에 유용한 무독성 전장 DddAtox 변이체를 얻기 위해 구조 기반 특정 부위 변이, 무작위 돌연변이 두 가지 방식을 취했다. 첫 번째 접근 방식에서는 DNA 결합이 감소하거나 촉매 활성이 감소된 DddAtox 변이체를 활성이 없는 CRISPR-Cas9(dCas9) 또는 nickase(nCas9) 변이체에 융합하여 사람 유래의 세포에서 표적 시토신이 티민으로 치환되는 새로운 염기 교정기를 개발했다. 이를 위해 DddAtox의 양전하를 따는 아미노산을 알라닌으로 치환하여 발현 벡터에 서브클로닝 하려고 했다 (도 51a). 이러한 변이체가 음전하를 띠는 dsDNA에 결합하는 것을 약화 하여 잠재적으로 독성을 피할 수 있다고 추론했다. 대부분의 알라닌 치환 변이체는 E. coli 형질전환체를 형성하지 못했다 (도 51b). 생성된 형질전환체로부터 분리된 플라스미드 DNA 염기분석결과 단백질을 형성하는 영역에서 다양한 프레임변이 돌연변이가 유도되었다. 이러한 전장 DddAtox 변이체는 포유류발현 프로모터의 조절하에 있지만 E. coli에서 약하게 발현되어 세포 사멸을 초래했다. 다행히도 프레임 시프트 돌연변이가 없는 여러 개의 삼중 또는 사중 또는 오중("AAAAA"라고 함) 알라닌 치환 변이체를 얻을 수 있었다. 또한 활성 부위 돌연변이 E1347A도 성공적으로 클로닝되었다.
- [1037] 다음으로 D10A nCas9 또는 dCas9 및 UGI에 융합된 AAAAA 변이체가 인간 배아 신장 293T(HEK293T) 세포에서 염기 편집을 유도할 수 있는지 여부를 조사했다(도 51c, d). 랫 APBEC1 탈아미노효소, 우라실 글리코실라제 억제제(UGI)와 dCas9(또는 D10A nCas9)로 구성된 염기교정기 2(또는 3)은 proto-spacer 영역 내의 좁은 영역에서 활성을 가지지만, AAAAA 변이체는 proto-spacer 영역의 바로 5'위쪽에서 시토신을 티민으로 최대 43%의 치환을 유도했다. 예상외로 E1347A 돌연변이는 37%(nCas9 융합) 또는 16%(dCas9 융합)의 빈도로 동일한 C-3 위치에서 염기 교정을 유도했으며(도 51c), 이는 E1347A 돌연변이가 DddAtox의 완전한 탈아미노 불활성화를 유도하지 않고, 인간 유래 세포에서 염기 교정을 달성하기에 충분히 높은 탈아미노 활성이 있는 것을 확인하였다. 그러나 5중 AAAAAA 돌연변이와 결합된 E1347A 변이체는 염기 교정을 유도하지 못했다. 또한 dCas9 또는 nCas9 및 UGI에 융합된 프레임 이동 돌연변이가 없는 E1347A, AAAAA 및 알라닌으로 치환된 다른 변이체 (도 53)가 proto-spacer 영역의 위쪽으로 최대 25에 베이스 떨어진 곳에서 교정이 일어나는 것을 확인했고, 다른 여러 사이트에서는 최대 26% 교정 효율이 있는 것을 확인하였다 (도 54). 또한, 최대 60%의 교정 빈도로 HeLa 세포에서 매우

효율적이었다(도 55). 이러한 융합단백질에 의해 유도된 염기 교정은 최대 21일 동안 세포에서 유지되었으며, 이는 이러한 염기 교정이 세포 독성이 없음을 시사한다(도 56).

[1038] 시토신 염기교정기의 교정 윈도우 변경을 위해, 알라닌으로 치환된 변이체들을 H840A nCas9의 C 터미널에 융합하려고 했다. 예기치 않게, 프레임시프트 돌연변이가 없는 손상되지 않은 구성물을 얻는 데 실패했다. 따라서 오류가 발생하기 쉬운 PCR을 수행하여 DddAtox 코딩 서열에 무작위 돌연변이를 도입했으며 4가지 점 돌연변이 S1326G, G1348S, A1398V, S1418G("GSVG"라고 함)가 있는 무독성 전장 DddAtox 변이체를 얻을 수 있었다(S1326G, G1348S, A1398V, S1418G 각각 서열번호 269의 아미노산 서열 중 37 위치의 S가 G로 치환; 59번 위치의 G가 S로 치환; 109번 위치의 A가 V로 치환; 및 129번 위치의 S가 G로 치환을 포함하는, 서열번호 276의 서열을 포함, 도 52a). 또한 이 변이체는 dCas9, D10A nCas9 및 Cas9의 C 말단과 dCas9, nCas9 및 Cas9의 N 말단에 융합되었다. 인간 유래 세포에서 야생형 Cas9을 제외한 이러한 융합 단백질은 최대 38%의 효율로 다양한 부위에서 시토신에서 티민으로의 전환을 유도했다(도 52b, 57, 58). 흥미롭게도, dCas9, D10A nCas9 및 H840A nCas9의 C 말단에 융합된 GSVG 변이체를 포함하는 융합 단백질은 proto-spacer 인접 모티프(PAM)의 3' 다운스트림에서 시토신 염기 교정을 나타내는 반면, N 말단에 융합된 동일한 변이체를 포함하는 융합 단백질인 dCas9 및 nCas9는 proto-spacer의 5' 업스트림에서 염기 교정을 유도했다(도 52c). 예상대로 Cas9를 포함하는 융합 단백질은 염기 치환보다는 삽입결실을 유도했다.

[1039] 어떤 돌연변이가 GSVG 변이체에서 중요한지를 찾아내기 위해, SSVG, GGVG, GSAG 및 GSVS의 4가지 복귀체를 사이트 지정 돌연변이 유발을 통해 만들려고 시도했다. SSVG, GSAG 및 GSVS 복귀체를 얻을 수 있었지만 nCas9의 C 말단에 융합된 GGVG 변이체를 얻지 못했다. G1348은 촉매 부위의 핵심 부위인 E1347 바로 옆에 있다. G1348S 돌연변이는 촉매 활성을 감소시켜 대장균의 세포독성을 피할 수 있었을 것이다. 최대 21일 동안 형질감염된 세포의 2개 표적 부위에서 3개의 복귀체와 GSVG 변이체의 교정 빈도를 측정했다. GSAG 및 GSVS에 의해 유도된 시토신에서 티민 교정의 빈도는 형질감염 후 3일부터 21일까지 점차적으로 ~2배 감소하여 이 두 복귀체가 다소 세포독성이 있는 반면 GSVG 및 SSVG은 유지되었다(도 59). 이러한 결과는 GSVG 변이체에서 G1348S가 필수적이며 S1326G가 중성인 반면 A1398V 및 S1418G는 세포독성을 감소시킨다는 것을 시사한다.

[1040] 종합하면, 위 결과는 dsDNA에 대한 친화도 감소(AAAAA) 또는 약화된 탈아미노활성(E1347A 및 가능하게는 GSVG) 또는 감소된 세포독성(GSVG)을 갖는 무독성, 전장 DddAtox 변이체가 dCas9 또는 nCas9에 융합되어 새로 변경된 교정 윈도우를 가지는 염기 교정기를 발명하였다. 이러한 염기교정기 이하 dCas9-mDdBE(dCas9의 C 말단에 융합된 전체 길이의 단량체 DddAtox 변이체로

구성된 DddA 파생 염기 교정기), nCas9-mDdBE, mDdCE-dCas9 및 mDdCE-nCas9로 명명될 수 있고, BE2 또는 BE3에서 도달할 수 없는 proto-spacer의 업스트림 또는 다운스트림 위치에서 염기 교정에 사용되는 BE2 또는 BE3 이후의 염기교정기이다.

- [1041] 또한, 무독성 전장 DddAtox 변이체가 미토콘드리아 DNA 교정에 사용될 수 있는지 여부를 조사했다. 여러 변이체 중에서 GSVG 및 E1347A 변이체 2개만 미토콘드리아 유전자, ND4 및 ND6에 결합하도록 설계된 TALE array의 C 말단에 성공적으로 융합되었다. GSVG 변이체를 포함하는 단량체 DdCBE(mDdCBE)는 원래 분할된 DdCBE 쌍과 동등한 최대 31% (ND4) (도 60a) 및 27% (ND6) (도 60b)의 표적 뉴클레오티드 위치에서 염기 교정을 달성했다. E1347A를 포함하는 mDdCBE도 최대 7.2%(ND4)와 8.9%(ND6)의 교정 비율로 효율성은 떨어지지만 목표 시토신을 티민으로 변환 하였다. 흥미롭게도 ND4 유전자에 특화된 원래의 DdCBE 쌍(G1333 분할)은 C4 위치에서 교정 효율이 0.8%인 반면, GSVG를 포함하는 2개의 mDdCBE는 26%와 31%의 높은 교정 효율을 나타내었다. 이 결과는 분할 이량체 DdCBE와 mDdCBE가 다른 돌연변이 패턴을 가질 수 있음을 보여주고 mDdCBE가 이량체 DdCBE와 상보적일 수 있어 주어진 표적 부위에서 다양한 돌연변이를 유도할 수 있음을 시사한다.

[1042]

- [1043] 분할 이량체 DdCBE에 비해 mDdCBE의 한 가지 잠재적인 이점은 비특이적 TALE-DNA 상호작용으로 인한 오프타겟 효과가 이량체 DdCBE에 비해 절반인 것이다. 분할 DddAtox가 있는 이량체 DdCBE는 하나의 서브유닛만 결합할 수 있는 절반 부위에서 작동할 수 있으며, 이는 원치 않는 오프타겟 돌연변이를 발생시킨다. DdCBE 쌍의 비활성 DddAtox 절반은 다른 비활성 절반을 모집하여 기능적 탈아미노효소를 형성할 수 있다. 이 가설을 확인하기 위해 이량체 DdCBE의 한 서브유닛을 인코딩하는 플라스미드와 TALE이 없는 DddAtox 절반을 인코딩하는 플라스미드를 HEK293T 세포에 공동 형질감염 시키고 두 미토콘드리아 표적 부위에서 교정 빈도를 측정했다. 예상대로 시토신에서 티민으로의 교정은 0.7에서 3.6% 빈도로 대상 사이트에서 관찰되었다(도 60c~f). 이 결과는 DdCBE 쌍에서 분할된 DddAtox 반쪽이 절반 부위에서 서로 상호작용하여 원치 않는 오프타겟 돌연변이를 일으킬 수 있으며 mDdCBE가 이량체 DdCBE에 의해 유도된 오프타겟 돌연변이의 절반을 피할 수 있음을 시사한다.

[1044]

- [1045] 실시예 7. DdABE를 이용한 사람 세포에서 높은 효율의 A-to-G 염기 교정

[1046]

- [1047] DddA 유래 시토신 염기 교정(DdCBE)에 의한 미토콘드리아 DNA 염기 교정은 여러가지 세포주(cell line)와 동물들의 질병 모델을 만들 수 있게 하였고, 미토콘드리아 유전병을 치료할 수 있는 새로운 길을 열었다. 하지만, DdCBE는

거의 TC-toTT 염기 교정만 한정적으로 일으키기 때문에 모든 경우의 수에서 1/8정도만 커버를 할 수 있다. 여기에, TALE (transcription activator-like effector)에 두 종류의 탈아미노효소를 연결한 TALED를 개발하였다. 여기에 TALE은 원하는 DNA 부분에 결합할 수 있도록 맞춤 설계(custom-designed) 하였고, 촉매 능력을 상실한 DddAtox 시토신 탈아미노효소 변이체 및 E.coli 유래의 DNA 아데닌 탈아미노효소인 TadA 단백질을 포함하고 있다. TALED는 사람 미토콘드리아에서 기존의 TC context에 관해서만 시토신 염기교정이 가능했던 이전 염기 교정 기술과 달리 모든 A-to-G에 대한 염기 교정이 가능하게 해준다. 실제로 맞춤 제작한 TALED는 사람 세포 여러 target에서 높은 효율(50% 정도까지)로 아데닌 염기 교정을 일으킬 수 있었다.

[1048]

[1049] 기존에는 없던 새로운 염기 교정 기술을 개발하기 위해, 여러 TadA 변이체 중에서 ABE8e의 TadA 변이체(TadA*)를 선택하였다. 이는 높은 효율로 아데닌 교정을 할 수 있을 뿐만 아니라 여러가지 다양한 DNA 결합 단백질과 호환될 수 있게 개량된 것이기 때문에 실제 TALE 또는 ZFP(zinc finger protein)에도 호환이 잘 될 수 있기 때문이다.

[1050] 처음으로 ND1 또는 ND4를 표적할 수 있게 맞춤 제작된 TALE에 TadA*와 MTS (mitochondria target sequence)를 융합하였고, 실제로 미토콘드리아 DNA에 염기교정을 일으킬 수 있는지 실험하였다. 표적 서열 심층 분석 (targeted deep sequencing) 결과 융합단백질의 아데노신 염기 교정 효율이 굉장히 낮지만 나오는 것을 볼 수 있었다. ND1 site에서는 최대 1.2% 효율(도 67a)로, ND4 site에서는 0.6% 효율(도 67b)로 아데닌 염기 교정을 일으켰다. 비록 굉장히 낮은 효율이지만, TadA*가 DNA 단일 가닥만 특이적으로 작동하는 것으로 알려진 것에 비해 TALE에 융합되어 있으면 이중가닥 표적DNA에서도 염기교정을 일으키는 것을 알 수 있었다.

[1051] 미토콘드리아 DNA에서 아데닌 염기교정이 일어날 수 있다는 결과와 함께, 이러한 효율을 더 올리기 위해서 이미 알려져있는 DddAtox 단백질을 융합하는 것을 생각하였다. DddAtox 단백질은 시토신 탈아민을 하는 단백질로 Burkholderia cenocepacia 유래 interbacterial toxin이다. 이 단백질은 DNA 이중가닥에 작동하는 단백질이기 때문에 이 단백질을 사용하게 되면 TadA*아데닌 탈아미노효소가 타겟 DNA에 더 잘 접근할 수 있도록 도와줄 것이기 때문이다. 기존의 DddAtox를 사용하는 DdCBE같은 경우 DddAtox 단백질을 반으로 나누어서 왼쪽 DNA부분에 붙는 TALE(Left-TALE (L-TALE))과 오른쪽 DNA부분에 붙는 Right-TALE (R-TALE))에 각각 나뉘어져서 split 형태로 융합되고, 시토신 염기 효율을 올려주는 uracil glycosylase inhibitor(UGI)까지 융합(TALE-split DddAtox-UGI)되어 있다. 이렇게 DddAtox를 분할해서 사용하는 이유는 전장의 단백질을 사용하게 되면 세포 독성이 발생하기 때문이다. 먼저 ND1 site를 타겟하는 DdCBE의 한쪽 부분에 UGI 대신에 TadA*를 달아서

L-TALE-split DddAtox- TadA*, R-TALE-1397C- TadA* 형태로 각각 만들어서 실험해보았다. 신기하게도, 한쪽 부분에는 TadA*가 있는 것을 다른 한쪽에는 1397N과 UGI가 있는 것을 쌍으로 해서 사람 세포에 전달해주었을 때 A-to-G와 C-to-T 염기 교정이 모두 일어나는 것을 확인할 수 있었다(도 62c). 기존의 DdCBE같은 경우 약 20% 시토신 염기 교정이 일어나고, 아데닌 염기교정은 전혀 일어나지 않는 반면 한쪽 부분에 TadA*가 있는 경우 시토신 염기 교정이 절반 수준으로 줄어들게 되고, 아데닌 염기 교정이 약 10%정도 일어나는 것을 확인하였다(도 62c). 아데닌 염기 교정과 시토신 염기 교정 효율이 비슷하게 일어났다(도 62c).

[1052] 물론 시토신 염기교정과 아데닌 염기교정이 모두 일어나는 것은 무작위로 돌연변이를 일으키는 것에는 유용할 지 몰라도, 질병을 치료하는 것에 있어서, 특히 C-to-T 돌연변이로 발생하는 LOHN 및 MEALS 같은 미토콘드리아 유전병을 치료하기 위해서는 아데닌만을 염기교정을 일으키는 것이 바람직하다. 따라서 이러한 동시에 발생하는 시토신 염기교정을 없애기 위해서 UGI를 없앴다. DdCBE에서, UGI의 경우 시토신 탈아미노효소 DddAtox가 C를 U로 탈아민화시키면 DNA 수선 과정 중 세포내 있는 복구 (repair) 단백질인 우라실 글리코실라제 (uracil glycosylase)에 의해 U가 다시 수선되는 것을 방지하고자 우라실 글리코실라제를 억제하기 위한 저해제 용도로 융합된 것이다. 따라서 이러한 UGI를 제거해주면, 아데닌 염기교정 효율은 유지하고, 시토신 염기교정을 억제 할 수 있을 것이라고 생각하였다. 놀랍게도, UGI가 없는 ND1표적 TALE 탈아미노효소 쌍은 시토신 염기교정을 거의 일으키지 않고(<0.5%) 높은 효율로 아데닌 염기 교정(약 50%)만을 일으키는 것을 확인하였다(도 63a,c). 이는 UGI가 있는 것보다도 훨씬 높은 효율이다. 더 나아가 ND4를 표적하는 TALE 탈아미노효소 쌍에서도 확인하였고, ND1 표적하는 것처럼 높은 효율(약 35%)로 아데닌 염기만을 교정하는 것을 확인하였다(도 63b,d). 이처럼 DddAtox시스템과 TadA*를 융합한 DNA 이중가닥에서 작동하는 새로운 아데닌deaminase인 TALED를 개발하였고, 세계 최초로 사람 미토콘드리아에서 아데닌 염기교정을 할 수 있었다. 또한 처음 TALE에 TadA*만을 달았던 것보다 최종적으로 50배 정도 높게 아데닌 염기를 교정할 수 있었다.

[1053] 다음으로 더 나아가서 촉매 활성을 없앤 전장 E1347A DddAtox 변이체 또는 개발한 촉매활성은 유지하고 세포독성은 없앤 변이체(AAAAA와GSVG)를 활용하여 아데닌 염기교정을 일으키려고 했다. 시토신 염기교정을 일으키는 것이 목적이 아니고, DNA 이중가닥에서 아데닌 염기 교정을 일으키는 것이 목적이기 때문에 시토신 염기 교정 활성이 없어진 전장 E1347A DddAtox변이체를 사용하여도 되었고, UGI가 없으면 시토신 염기교정이 효율이 없어지는 것을 확인하였기 때문에 동시에 세포독성만을 없앤 변이체도 이용할 수 있었다. 전장 변이체 포함하는 TALED를 두가지 타입으로 만들었다(도 64a).

첫번째는 하나의 TALE에 TadA*(AD)와 전장 DddAtox변이체를 모두 포함하는 것(mTALED), 두번째는 각각의 TALE에 TadA*(AD)와 전장 DddAtox변이체를 각각 따로 융합하는 것(dTALED), 두 가지 타입을 만들어서 테스트해보았다(도 64a). 놀랍게도, ND1을 표적하는 두가지 타입의 TALED 모두 높은 효율로 아데닌 염기교정을 일으키는 것을 확인하였다(도 64b). mTALED 경우 최대 약 45% 효율을 보여줬으며, dTALED 역시 약 50%정도의 아데닌 염기 교정 효율을 보여줬다(도 64b). ND1 site 뿐만 아니라 ND4 site에서도 마찬가지로 실험을 하였고, 비슷하게 높은 효율로 아데닌 염기교정을 일으켰다(도 64c). 또한 시토신 염기 교정 활성이 없어진 전장 E1347A DddAtox변이체를 사용했을 때도 높은 효율의 아데닌 염기 교정을 일으키는 것을 볼 수 있었다(도 64b,c). 이는 시토신 탈 아민 능력이 상실 되었음에도 충분히 TadA*를 DNA 이중가닥에 접근하도록 돕는 역할은 유지되는 것으로 보여진다. 위의 결과들을 자세하게 단일 염기 수준으로 봤을 때(도 65, 66), TALE이 DNA에 결합하는 바로 근처에서 아데닌 염기 교정이 일어나는 것을 볼 수 있다. 또한 두개의 TALE이 이용하는 경우 그 사이(Spacer)에서만 염기교정이 일어나게 되는데, 신기하게도 하나의 TALE이 이용하는 mTALED에서도 비슷한 타겟 길이를 가지는 것을 볼 수 있었다(도 65, 66).

- [1054] 또한 시스템이 zinc finger protein(ZFP) system에서도 작동이 되는지, 핵 DNA에서도 작동하는 지 궁금하였다. 따라서 핵 DNA를 타겟하는 NC 형태의 ZFP만들었고, 여기에 분할된 DddAtox와 TadA*를 융합하였다(도 61a). 이 때 ZFP에 여러가지 위치에 TadA*를 융합해주었다(도 61b). 여러가지 형태 중 핵 DNA에서 최대 10%정도 아데닌 염기를 교정하는 개체를 만들 수 있었다(도 61d). UGI가 한 쪽에 존재하기에 시토신 염기 교정 효율도 높은 효율로 존재하는 것을 볼 수 있었다(도 61c). 사람 세포 핵 DNA에서 ZFP-DddAtox-TadA* system이 작동하는 것을 볼 수 있었고, 더 나아가서 미토콘드리아에서도 작동하는 지 실험하였다. 이 때 핵 DNA에서 가장 높은 효율로 작동했던 구조를 참고하였다. 핵 DNA로 전달하게 하는 Nuclear Localization Signal(NLS) 대신 미토콘드리아로 향하게 하는 Mitochondrial Targeting Sequence(MTS)를 달아주고, ND1 site을 타겟하는 ZFP(NC타입)에 왼쪽 ZFP에는 1397N이, 오른쪽 ZFP에는 1397C와 더불어 TadA*가 융합된 형태로 실험을 하였다. 결과적으로 약 3% 효율로 아데닌 염기 교정이 일어나는 것을 볼 수 있었다(도 61.g). 비록 TALED 보다 아데닌 염기 교정 효율이 낮았지만, 단백질 사이를 연결해주는 linker등 여러가지 조건을 최적화 시켜준다면 이 ZFP system을 이용해서도 좋은 효율의 아데닌 염기 교정을 일으킬 수 있을 것이다.

[1055]

- [1056] 현재까지 유전자 교정 기술은 눈부신 발전을 해왔다. CRISPR 기반의 유전자 가위(CRISPR Cas9, base editor, prime editor 등)는 off-target을 개선하고, 효율을 높이며 다양하게 발전해왔다. 하지만 이러한 많은 발전에도 미토콘드리아

유전병에 대한 치료 방법은 한계가 있었다. 이유는 CRISPR 기반의 기술은 실제로 촉매하는 단백질과 표적으로 가이드 역할을 해주는 gRNA로 구성되어 있다. 하지만, 단백질과는 달리 gRNA를 미토콘드리아로 보내주는 방법의 부재로 아직까지 미토콘드리아 유전자를 다루는 기술은 DNA를 절단을 일으켜 미토콘드리아 DNA를 없애는 것 말고는 존재하지 않았다. 하지만, 미국의 David R. Liu 그룹에서 처음으로 미토콘드리아에서 염기교정을 일으킬 수 있는 DdCBE 선보였다. DdCBE는 DNA 이중가닥에서 작동하는 시토신 탈아미노효소 DddAtox를 포함하고 있기 때문에 DNA 결합 단백질 TALE에 융합하여 단백질만을 보내 염기 교정을 일으킬 수 있었다. 하지만, 이러한 DdCBE는 시토신 중에서도 TC 형태의 시토신에서만 제한적으로 염기교정을 일으키기 때문에 실제로 질병모델 제작 또는 유전병 치료에 있어서 한계점이 많았다. 이에 세계 최초로 미토콘드리아에서 아데닌 염기 교정을 일으킬 수 있는 TALED를 만들었고, TALED는 최대 50%정도의 높은 효율을 가지고 있으며, 주변 타겟 부분에서 다양한 종류의 아데닌에서 염기교정을 일으켰다. 또한 UGI가 있을 경우에는 시토신과 아데닌 염기를 동시에 교정을 할 수 있기 때문에 무작위적 변이를 일으킬 때 유용하게 사용할 수 있는 반면, UGI가 없을 경우에는 시토신 염기교정은 안 일어나고, 아데닌 염기교정만을 일으키기 때문에 특이적인 아데닌 염기 교정 기술로 이용될 수 있다. 또한 ZFP 시스템에서도 적용이 되며, 핵 DNA 관해서도 아데닌 염기교정이 가능함을 보였다. 이러한 TALED의 개발은 많은 미토콘드리아 유전병에 해결책을 제시해주며 그에 해당하는 질병 모델을 만들수 있게 해주고, 아직까지 개척되지 않은 많은 미토콘드리아 유전자 관련 연구들을 하는데 유용한 도구가 될 것이다.

[1057]

[1058] 이상으로 본 발명의 내용의 특정한 부분을 상세히 기술하였는바, 당업계의 통상의 지식을 가진 자에게 있어서, 이러한 구체적 기술은 단지 바람직한 실시양태일 뿐이며, 이에 의해 본 발명의 범위가 제한되는 것이 아닌 점은 명백할 것이다. 따라서, 본 발명의 실질적인 범위는 첨부된 청구항들과 그것들의 등가물에 의하여 정의된다고 할 것이다.

[1059]

산업상 이용가능성

[1060] 본 발명에 따르면, 유전자 교정시에 시토신 탈아미노효소 분할 형태 (split) 결합 표면의 특정 아미노산 잔기를 치환함으로써, 원하지 않는 시토신 탈아미노효소의 비선택성을 줄일 수 있다.

[1061] 전장 시토신 탈아미노효소와 관련하여, 기존에 사용되는 시토신 염기 교정 (cytosine base editor)로 교정 (editing)이 어려운 부분의 교정이 가능하다. 기존의 시토신 염기 교정 수단 중 탈아미노효소 (deaminase)로 사용하는 Apobec1은 암 유발 유전자로 알려져 있어서 치료목적으로 사용에 제약이 따르지만, 이번에

개발된 전장 탈아미노효소는 이러한 문제점이 없을 것으로 판단된다.

[1062] DNA 결합 단백질을 포함해서 2.5Kb 정도로 작아 AAV 벡터를 이용한 유전자 치료에 사용 가능하고, mRNA, RNP 전달이 용이하며, 원핵생물을 이용한 유용 물질 생산 등에 사용 가능하다.

[1063]

서열목록 Free Text

[1064] 전자파일 첨부하였음.

청구범위

- [청구항 1] (i) DNA 결합 단백질; 및
(ii) 시토신 탈아미노효소 (deaminase) 또는 이의 변이체 유래인 제1분할체 및 제2분할체를 포함하는 융합단백질로,
상기 제1분할체 및 제2분할체는 각각 상기 DNA 결합 단백질에 결합되는 형태인 융합단백질.
- [청구항 2] (i) DNA 결합 단백질; 및
(ii) 시토신 탈아미노효소 또는 이의 변이체 유래 비독성 전장 시토신 탈아미노효소를 포함하는 융합단백질.
- [청구항 3] 제1항에 있어서, 상기 제1분할체와 제2분할체 각각은 시토신 탈아미노효소 활성이 없는 것을 특징으로 하는, 융합단백질.
- [청구항 4] 제1항에 있어서, 상기 제1분할체는 서열번호 1의 서열 중 N 말단에서 G33, G44, A54, N68, G82, N98, G108로 구성된 군에서 선택되는 하나 이상까지의 서열을 포함하는 융합단백질.
- [청구항 5] 제1항 또는 제2항에 있어서 상기 시토신 탈아미노효소는 이중 가닥 DNA에 작용하는 탈아미노효소 (DddA) 또는 이의 오솔로그 (orthologue)에서 유래한 것을 특징으로 하는 융합단백질.
- [청구항 6] 제1항에 있어서, 상기 제2분할체는 서열번호 1의 서열 중 G34, P45, G55, N69, T83, A99, A109로 구성된 군에서 선택되는 하나 이상에서 C-말단까지의 서열을 포함하는 융합단백질.
- [청구항 7] 제1항에 있어서, 상기 시토신 탈아미노효소의 변이체는 서열번호 1의 서열 중 N 말단에서 G44까지의 서열을 포함하는 제1분할체 중 3, 5, 10, 11, 13, 14, 15, 16, 17, 18, 19, 28, 30 및 31번 위치로 구성된 군에서 선택되는 하나 이상의 아미노산이 다른 아미노산으로 치환된 융합단백질.
- [청구항 8] 제1항에 있어서, 상기 시토신 탈아미노효소의 변이체는 서열번호 1의 서열 중 P45에서 C-말단까지의 서열을 포함하는 제2분할체 중 13, 16, 17, 20, 21, 28, 29, 30, 31, 32, 33, 56, 57, 58 및 60번 위치로 구성된 군에서 선택되는 하나 이상의 아미노산이 다른 아미노산으로 치환된 융합단백질.
- [청구항 9] 제1항에 있어서, 상기 시토신 탈아미노효소의 변이체는 서열번호 1의 서열 중 N 말단에서 G108까지의 서열을 포함하는 제1분할체 중 87, 88, 91, 92, 95, 100, 101, 102 및 103번 위치로 구성된 군에서 선택되는 하나 이상의 아미노산이 다른 아미노산으로 치환된 융합단백질.
- [청구항 10] 제1항에 있어서, 상기 시토신 탈아미노효소의 변이체는 서열번호 1의 서열 중 A109에서 C-말단까지의 서열을 포함하는 제2분할체 중 13, 14, 15 및 16번 위치로 구성된 군에서 선택되는 하나 이상의 아미노산이 다른

- 아미노산으로 치환된 융합단백질.
- [청구항 11] 제2항에 있어서, 상기 비독성 전장 시토신 탈아미노효소는 서열번호 1의 야생형 시토신 탈아미노효소 중 37, 59, 109, 129로 구성되는 군에서 선택되는 하나 이상 위치의 아미노산이 다른 아미노산으로 치환된 융합단백질.
- [청구항 12] 제7항 내지 제11항 중 어느 한 항에 있어서, 상기 다른 아미노산은 알라닌인 융합단백질.
- [청구항 13] 제2항에 있어서, 상기 비독성 전장 시토신 탈아미노효소는 서열번호 12 내지 22로 구성된 군에서 선택되는 하나 이상인 융합단백질.
- [청구항 14] 제1항 또는 제2항에 있어서, 상기 DNA 결합 단백질은 징크 핑커 단백질 (zinc finger protein), TALE 단백질 또는 CRISPR 연관 뉴클레아제인 융합단백질.
- [청구항 15] 제1항 또는 제2항에 있어서, 상기 DNA 결합 단백질은 2 내지 40개 아미노산 잔기를 포함하는 펩타이드 링커를 통해 시토신 탈아미노효소 또는 이의 변이체와 연결되는 것을 특징으로 하는 융합단백질.
- [청구항 16] 제15항에 있어서, 상기 링커는 다음을 포함하는 융합단백질:
 2a.a 링커: GS
 5a.a 링커: TGEKQ (서열번호 8)
 10a.a 링커: SGAQGSTLDF (서열번호 9)
 16a.a 링커: SGSETPGTSESATPES (서열번호 10);
 24 a.a 링커: SGTPEHVG VYTL SGTPEHVG VYTL (서열번호 115); 또는
 32a.a 링커: GSGGSSGGSSGSETPGTSESATPESSGGSSGGSS (서열번호 11).
- [청구항 17] 제1항에 있어서, 상기 제1분할체 및 제2분할체 각각은 징크 핑거 단백질의 N-말단 또는 C-말단에 결합된 융합단백질.
- [청구항 18] 제1항 또는 제2항에 있어서, 상기 시토신 탈아미노효소에 단일 TALE 어레이 (array) 또는 제 1 TALE 어레이 및 제 2 TALE 어레이가 각각 결합하는 융합단백질.
- [청구항 19] 제1항 또는 제2항에 있어서, (iii) 아데닌 탈아미노효소를 추가로 포함하는 융합단백질.
- [청구항 20] 제19항에 있어서, 상기 아데닌 탈아미노효소는 대장균 TadA의 변이체로서 디옥시-아데닌 탈아미노효소 (deoxy-adenine deaminase)인, 융합단백질.
- [청구항 21] 제19항에 있어서, 상기 아데닌 탈아미노효소는 DNA 결합 단백질 또는 시토신 탈아미노효소 또는 이의 변이체의 N 또는 C 말단에 결합된 융합단백질.
- [청구항 22] 제1항 내지 제21항 중 어느 한 항의 융합단백질을 코딩하는 핵산.
- [청구항 23] 제22항에 있어서, 리보핵산 또는 DNA인 핵산.

- [청구항 24] 제1항 내지 제21항 중 어느 한 항의 융합단백질 또는 제22항의 핵산을 포함하는 염기 교정용 조성물.
- [청구항 25] 제24항에 있어서, UGI (uracil glycosylase inhibitor)를 추가로 포함하는 조성물.
- [청구항 26] 제1항 내지 제21항 중 어느 한 항의 융합단백질 또는 제22항의 핵산을 포함하는 진핵세포 염기 교정용 조성물.
- [청구항 27] 제1항 내지 제21항 중 어느 한 항의 융합단백질 또는 제22항의 핵산; 및 NLS (nuclear localization signal) 펩타이드 또는 이를 코딩하는 핵산을 포함하는 식물세포 염기 교정용 조성물.
- [청구항 28] 제1항 내지 제21항 중 어느 한 항의 융합단백질 또는 제22항의 핵산; 및 엽록체 전달 펩타이드(chloroplast transit peptide) 또는 이를 코딩하는 핵산을 포함하는 식물세포 염기 교정용 조성물.
- [청구항 29] 제1항 내지 제21항 중 어느 한 항의 융합단백질 또는 제22항의 핵산; 및 미토콘드리아 전달 단백질 (Mitochondrial Targeting Signal, MTS) 또는 이를 코딩하는 핵산을 포함하는 식물세포 염기 교정용 조성물.
- [청구항 30] 제29항에 있어서, 핵외 수출 신호 (Nuclear Export Signal, NES) 또는 이를 코딩하는 핵산을 추가로 포함하는, 조성물.
- [청구항 31] 제27항 내지 제30항 중 어느 한 항에 있어서, 상기 융합단백질은 다음을 통해 식물세포에 전달되는 것을 특징으로 하는 조성물:
Gene gun을 이용한 주입 (Bombardment);
PEG-매개 원형질체 형질감염 (Protoplast transfection, PEG-mediated);
전기천공에 의한 원형질체 형질감염 (Protoplast transfection, electroporation); 또는
미세주사에 의한 원형질체 주사 (Protoplast injection, microinjection).
- [청구항 32] 제31항에 있어서, 상기 핵산은 다음을 통해 식물세포에 전달되는 것을 특징으로 하는 조성물:
Agrobacterium 예를 들어 Agrobacterium tumefaciens, Agrobacterium rhizogene 등을 이용한 형질전환;
바이러스 형질감염;
Gene gun을 이용한 주입 (Bombardment);
PEG-매개 원형질체 형질감염 (Protoplast transfection, PEG-mediated);
전기천공에 의한 원형질체 형질감염 (Protoplast transfection, electroporation); 또는
미세주사에 의한 원형질체 주사 (Protoplast injection, microinjection).
- [청구항 33] 제27항 내지 제30항 중 어느 한 항에 있어서, 상기 식물의 미토콘드리아, 엽록체 또는 색소체 (plastid: 백색체, 유색체)의 염기 교정용 조성물.
- [청구항 34] 제27항 내지 제30항 중 어느 한 항에 있어서, 야생형 DNA 염기서열을 자르지만 교정된 염기서열은 자르지 못하는 TAL (Transcription

- Activator-Like) 이펙터 (TALE); 및 FokI 핵산 가수분해효소(FokI Nuclease) 또는 이를 코딩하는 핵산 또는 ZFN(zinc finger nuclease) 또는 이를 코딩하는 핵산을 추가로 포함하는 조성물.
- [청구항 35] 제27항 내지 제30항 중 어느 한 항의 조성물을 처리하는 단계를 포함하는 진핵세포의 핵, 미토콘드리아, 또는 색소체 DNA의 염기를 교정하는 방법.
- [청구항 36] 제35항에 있어서, 야생형 DNA 염기서열을 자르지만 교정된 염기서열은 자르지 못하는 TALEN 또는 ZFN 또는 이를 코딩하는 핵산을 추가로 포함해 염기교정의 효율을 높이는 방법.
- [청구항 37] 제27항 내지 제30항 중 어느 한 항의 조성물을 식물세포에 처리하는 단계를 포함하는 식물세포 염기 교정방법.
- [청구항 38] 제1항 내지 제21항 중 어느 한 항의 융합단백질 또는 제22항의 핵산; 및 NLS (nuclear localization signal) 펩타이드 또는 이를 코딩하는 핵산을 식물세포에 처리하는 단계를 포함하는 식물세포 염기 교정방법.
- [청구항 39] 제1항 내지 제21항 중 어느 한 항의 융합단백질 또는 제22항의 핵산; 및 엽록체 전달 펩타이드(chloroplast transit peptide) 또는 이를 코딩하는 핵산을 식물세포에 처리하는 단계를 포함하는 식물세포 염기 교정방법.
- [청구항 40] 제1항 내지 제21항 중 어느 한 항의 융합단백질 또는 제22항의 핵산; 및 미토콘드리아 전달 단백질 (Mitochondrial Targeting Signal, MTS) 또는 이를 코딩하는 핵산을 식물세포에 처리하는 단계를 포함하는 식물세포 염기 교정방법.
- [청구항 41] 제1항 내지 제21항 중 어느 한 항의 융합단백질 또는 제22항의 핵산; 및 NLS (nuclear localization signal) 펩타이드 또는 이를 코딩하는 핵산을 포함하는 동물세포 염기 교정용 조성물
- [청구항 42] 제1항 내지 제21항 중 어느 한 항의 융합단백질 또는 제22항의 핵산; 및 미토콘드리아 전달 단백질 (Mitochondrial Targeting Signal, MTS) 또는 이를 코딩하는 핵산을 포함하는, 동물세포 염기 교정용 조성물.
- [청구항 43] 제42항에 있어서, 핵외 수출 신호 (Nuclelar Export Signal) 또는 이를 코딩하는 핵산을 추가로 포함하는 조성물.
- [청구항 44] 제42항에 있어서, 야생형 DNA 염기서열을 자르지만 교정된 염기서열은 자르지 못하는 TAL (Transcription Activator-Like) 이펙터 (TALE); 및 FokI 핵산 가수분해효소(FokI Nuclease) 또는 이를 코딩하는 핵산 또는 ZFN 또는 이를 코딩하는 핵산을 추가로 포함하는 조성물.
- [청구항 45] 제41항 또는 제42항의 조성물을 동물세포에 처리하는 단계를 포함하는 동물세포의 염기 교정방법.
- [청구항 46] 제1항 내지 제21항 중 어느 한 항의 융합단백질 또는 제22항의 핵산; 및 NLS (nuclear localization signal) 펩타이드 또는 이를 코딩하는 핵산을 동물세포에 처리하는 단계를 포함하는 동물세포의 염기 교정방법.

- [청구항 47] 제1항 내지 제21항 중 어느 한 항의 융합단백질 또는 제22항의 핵산; 및 미토콘드리아 전달 단백질 (Mitochondrial Targeting Signal, MTS) 또는 이를 코딩하는 핵산을 동물세포에 처리하는 단계를 포함하는 동물세포의 염기 교정방법.
- [청구항 48] 제46항 또는 제47항에 있어서, 야생형 DNA 염기서열을 자르지만 교정된 염기서열은 자르지 못하는 TALEN 또는 ZFN 또는 이를 코딩하는 핵산을 추가로 포함해 염기교정의 효율을 높이는 방법.
- [청구항 49] 제19항의 융합단백질 또는 이를 코딩하는 핵산을 포함하는 원핵세포 또는 진핵세포에서의 A-to-G 염기 교정용 조성물로, 상기 DNA 결합 단백질은 징크 핑커 단백질 (zinc finger protein), TALE 단백질 또는 CRISPR 연관 뉴클레아제이고, 상기 융합단백질의 시토신 탈아미노효소 또는 이의 변이체는 박테리아 유래이고, 이중 가닥 DNA 특이적인 것을 특징으로 하는 조성물.
- [청구항 50] 제19항의 융합단백질 또는 이를 코딩하는 핵산을 포함하는 원핵세포 또는 진핵세포에서의 A-to-G 염기 교정용 조성물로, 상기 DNA 결합 단백질은 징크 핑커 단백질 (zinc finger protein), TALE 단백질 또는 CRISPR 연관 뉴클레아제이고, 상기 융합단백질의 시토신 탈아미노효소 또는 이의 변이체는 박테리아 유래이고, 이중 가닥 DNA 특이적인 것이며, 상기 시토신 탈아미노효소 또는 이의 변이체의 N 말단에 DNA 결합 단백질이 결합되어 있고, 상기 융합단백질의 아데닌 탈아미노효소의 C 말단에 DNA 결합 단백질이 결합되어 있는, 조성물.
- [청구항 51] 제19항의 융합단백질 또는 이를 코딩하는 핵산 및 UGI (uracil glycosylase inhibitor)를 포함하는 원핵세포 또는 진핵세포에서의 C-to-T 염기 교정용 조성물로, 상기 DNA 결합 단백질은 징크 핑커 단백질 (zinc finger protein), TALE 단백질 또는 CRISPR 연관 뉴클레아제이고, 상기 융합단백질의 시토신 탈아미노효소 또는 이의 변이체는 박테리아 유래이고, 이중 가닥 DNA 특이적인 것을 특징으로 하는 조성물.
- [청구항 52] 제19항의 융합단백질 또는 이를 코딩하는 핵산을 원핵세포 또는 진핵세포에 처리하는 단계를 포함하는 원핵세포 또는 진핵세포에서의 A-to-G 염기 교정 방법으로, 상기 DNA 결합 단백질은 징크 핑커 단백질 (zinc finger protein), TALE 단백질 또는 CRISPR 연관 뉴클레아제이고, 상기 융합단백질의 시토신 탈아미노효소 또는 이의 변이체는 박테리아 유래이고, 이중 가닥 DNA 특이적인 것을 특징으로 하는 방법.
- [청구항 53] 제19항의 융합단백질 또는 이를 코딩하는 핵산을 원핵세포 또는 진핵세포에 처리하는 단계를 포함하는 원핵세포 또는 진핵세포에서의 A-to-G 염기 교정 방법으로, 상기 DNA 결합 단백질은 징크 핑커 단백질 (zinc finger protein), TALE 단백질 또는 CRISPR 연관 뉴클레아제이고,

상기 융합단백질의 시토신 탈아미노효소 또는 이의 변이체는 박테리아 유래이고, 이중 가닥 DNA 특이적인 것이며,
 상기 시토신 탈아미노효소 또는 이의 변이체의 N 말단에 DNA 결합 단백질이 결합되어 있고, 상기 융합단백질의 아데닌 탈아미노효소의 C 말단에 DNA 결합 단백질이 결합되어 있는, 방법.

[청구항 54] 제19항의 융합단백질 또는 이를 코딩하는 핵산 및 UGI (uracil glycosylase inhibitor)를 원핵세포 또는 진핵세포에 처리하는 단계를 포함하는 원핵세포 또는 진핵세포에서의 C-to-T 염기 교정 방법으로, 상기 DNA 결합 단백질은 징크 핑커 단백질 (zinc finger protein), TALE 단백질 또는 CRISPR 연관 뉴클레아제이고, 상기 융합단백질의 시토신 탈아미노효소 또는 이의 변이체는 박테리아 유래이고, 이중 가닥 DNA 특이적인 것을 특징으로 하는 방법.