

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국



(10) 국제공개번호

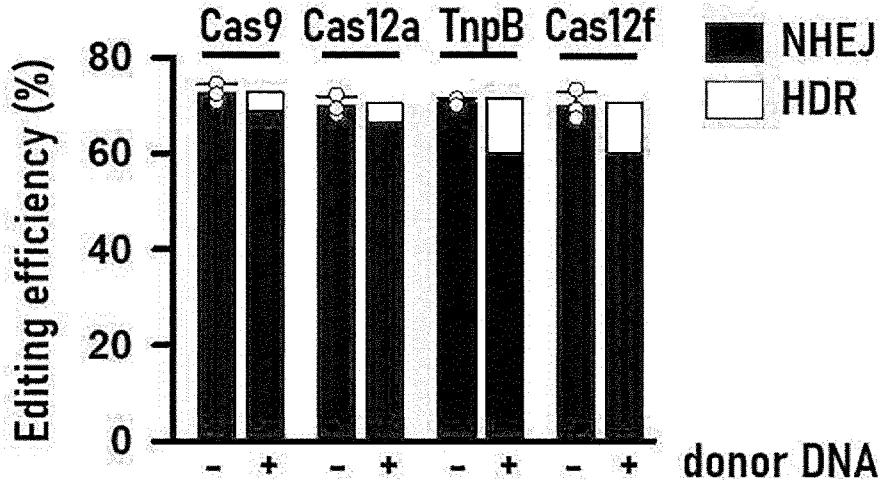
WO 2023/153845 A2

2023 년 8 월 17 일 (17.08.2023) WIPO | PCT

- (51) 국제특허분류: *C12N 15/63* (2006.01) *C12N 15/113* (2010.01) 34141 대전광역시 유성구 과학로 125, Daejeon (KR). 진현정 (CHIN, Hyun Jung); 34141 대전광역시 유성구 과학로 125, Daejeon (KR).
C12N 9/22 (2006.01)
- (21) 국제출원번호: PCT/KR2023/001945 (74) 대리인: 특허법인 세움 (SEUM IP); 06141 서울특별시 강남구 테헤란로 211, 13층, Seoul (KR).
- (22) 국제출원일: 2023 년 2 월 9 일 (09.02.2023)
- (25) 출원언어: 한국어 (81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.
- (26) 공개언어: 한국어
- (30) 우선권정보: 10-2022-0017179 2022 년 2 월 9 일 (09.02.2022) KR
- (71) 출원인: 주식회사 진코어 (GENKORE CO., LTD.) [KR/KR]; 34141 대전광역시 유성구 과학로 125, 303호, Daejeon (KR).
- (72) 발명자: 김용삼 (KIM, Yong Sam); 34118 대전광역시 유성구 가정로 43, 109동 802호, Daejeon (KR). 김도연 (KIM, Do Yon); 34069 대전광역시 유성구 반석동로 33, 505동 1802호, Daejeon (KR). 이정미 (LEE, Jeong Mi); (84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG,

(54) Title: TARGET SYSTEM FOR HOMOLOGY-DIRECTED REPAIR AND GENE EDITING METHOD USING SAME

(54) 발명의 명칭: 상동지정복구를 위한 TARGET 시스템 및 이를 이용한 유전자 편집 방법



(57) Abstract: The present invention relates to a novel target nucleic acid editing system including a miniaturized nucleic acid editing protein and an engineered guide RNA, uses of homology-directed repair (HDR) in a target gene thereof, and the like. According to one embodiment, homology-directed repair (HDR) using the target nucleic acid editing system has high HDR efficiency compared to other CRISPR/Cas systems due to a tendency to cut the back (outside) of a target nucleic acid, and also has the effect of enabling packaging of a gene editing system including a donor nucleic acid in a single vector, even when the packaging size is very limited like adeno-associated viruses (AAV), and maximizing the HDR efficiency by adding shRNA that inhibits a non-homologous end joining (NHEJ) process.

(57) 요약서: 본 발명은 초소형 핵산 편집 단백질 및 엔지니어링된 가이드 RNA를 포함하는 신규한 표적 핵산의 편집 시스템 및 이의 표적 유전자에서의 상동지정복구(HDR) 용도 등에 관한 것으로, 일 구현예에 따른 표적 핵산 편집 시스템을 이용한 상동지정복구에 의하면, 표적 핵산의 뒤쪽(바깥쪽)을 절단하는 경향성을 가져 다른 CRISPR/Cas 시스템에 비해 HDR 효율이 높을 뿐 아니라, 아데노-연관 바이러스(AAV)와 같이 패키징 사이즈가 매우 제한적인 경우에도 하나의 벡터에 공여자 핵산을 포함한 유전자 편집 시스템을 패키징할 수 있고, 비상동말단연결(NHEJ) 과정을 억제하는 shRNA를 추가하여 HDR 효율을 극대화할 수 있는 효과가 있다.

ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM),
유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,
FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME,
MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR),
OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM,
ML, MR, NE, SN, TD, TG).

공개:

- 국제조사보고서 없이 공개하며 보고서 접수 후 이를 별도 공개함 (규칙 48.2(g))
- 명세서의 서열목록 부분과 함께 (규칙 5.2(a))

명세서

발명의 명칭: 상동지정복구를 위한 TARGET 시스템 및 이를 이용한 유전자 편집 방법

기술분야

- [1] 본 발명은 상동지정복구에 의한 유전자 편집 효율이 개선된 새로운 초소형의 CRISPR/Cas12f 핵산 편집 시스템에 관한 것이다.
- [2] 본 출원은 2022년 2월 9일에 출원된 대한민국 특허출원 제10-2022-0017179호에 기초한 우선권을 주장하며, 해당 출원의 명세서 및 도면에 개시된 모든 내용은 본 출원에 원용된다.

배경기술

- [3] DNA의 이중가닥 절단은 비상동말단연결(non-homologous end joining, NHEJ) 또는 상동지정복구(homology-directed repair, HDR)이라는 DNA 복구 기작을 통해 복구될 수 있다. 비상동말단연결(NHEJ)에 의한 복구 과정에서는, DNA 절단 위치 사이에 무작위적 염기의 삽입(insertion) 또는 결실(deletion)이 일어나고(insertion and deletion, indel), 그 결과 DNA 이중가닥 절단이 일어난 유전자에 틀 이동 변이(frameshift mutation) 또는 조기 종결 변이(premature mutation)가 발생하여 해당 유전자가 제거(knock-out)된다. 반면, 상동지정복구(HDR)은 절단된 DNA를 복구하기 위하여 공여 DNA(donor DNA, homologous template)를 필요로 하고, 이 공여 DNA의 서열을 주형으로 하는 새로운 서열이 DNA 절단 위치에 도입될 수 있다. 이를 통해 상동지정복구(HDR)은 정교한 유전자 편집에 활용될 수 있다.
- [4] 현재 CRISPR/Cas 시스템으로 대표되는 유전자 편집 기술은 암, 유전 질환, 감염 질환 등에 대한 유전자 치료 개발을 위한 핵심 기술로서 유전자 편집의 효율성, 안전성, 전달성 등의 측면에서 다양한 기술개발이 이루어지고 있다.
- [5] 유전자 편집 기술은 (DNA 이중가닥 절단 및 Indel을 통해) 질환의 원인이 되는 병원성 유전자를 결실시킬 수 있기 때문에 많은 암, 유전 질환, 감염 질환 등 다양한 질병에 대한 유전자 치료제로서 가능성을 가진다. 그런데, 치료를 위한 목적 유전자를 도입할 수 있는 상동지정복구(HDR)를 가능하게 하거나 이의 효율을 개선할 수 있는 유전자 편집 기술은 보다 확장된 유용성을 가질 수 있다.
- [6] 또한, 유전자 치료를 위해서는 유전자 편집 시스템을 전신의 세포에 효율적으로 전달하는 것이 무엇보다도 중요하며 이를 위해 효율적인 매개체가 필요하다. 아데노 연관 바이러스(adeno-associated virus, AAV)는 그 안전성, 지속성 및 대량 생산과의 호환성으로 인해 유전자 치료제로서 FDA 승인을 받은 매개체이므로(비특허문헌 1), 하나의 AAV 벡터에 구성요소를 모두 포함할 수 있는 유전자가 위치 시스템은 유전 질환 치료에 중요한 도구로써 이용될 것으로 인정받고 있다(비특허문헌 2). 그러나 AAV의 패키징 용량(packaging capacity) 한계로 인하여 이를 통해 전달할 수 있는 유전자의 크기는 4.7kb 미만으로 제한적이다. 이는 AAV 벡

터를 세포 내 전달 매개체로 하여 유전자 치료에 사용하고자 할 때, 기존에 제작된 대부분의 유전자 가위(CRISPR/Cas) 시스템이 일반적으로 약 4.7kb 정도의 크기를 초과하기 때문에 임상 적용에 제약이 있음을 암시하는 것이다(비특허문헌 3). 이와 같은 이유로 분자량이 Cas9 보다는 작은 SaCas9(비특허문헌 4), CjCas9(비특허문헌 5) 등이 AAV를 매개체로 하여 세포 내로 전달될 수 있는 유전자 편집 도구로서 연구되고 있다.

- [7] 이와 같이, 분자량은 충분히 작지만 세포 내에서 높은 표적 효율 및 편집 활성을 가지는 새로운 시스템이 필요하다. 또한 지금까지의 유전자 편집 기술은 HDR 효율이 낮기 때문에 이를 획기적으로 향상시켜 정밀한 유전자 교정을 가능하게 하는 유전자 편집 도구가 필요한 실정이다.

발명의 상세한 설명

기술적 과제

- [8] 본 발명은 전술한 종래 기술의 문제점을 모두 해결하는 것을 그 목적으로 한다.
- [9] 또한, 본 발명은 개선된 상동지정복구(HDR) 효율을 나타내고/나타내거나 AAV 벡터를 비롯한 다양한 벡터 시스템에 수용가능한 초소형의 구조물로 구현될 수 있는 유전자 편집 기술을 제공하는 것을 목적으로 한다.
- [10] 본 발명의 목적은 이상에서 언급한 목적으로 제한되지 않는다. 본 발명의 목적은 이하의 설명으로 보다 분명해질 것이며, 청구범위에 기재된 수단 및 그 조합으로 실현될 것이다.

과제 해결 수단

- [11] 상기 목적을 달성하기 위한 본 발명의 대표적인 구성은 다음과 같다.
- [12] 본 발명의 일 태양에 따르면, Cas12f1, TnpB 또는 이의 변이체 단백질을 포함하는 엔도뉴클레아제 또는 상기 엔도뉴클레아제를 암호화하는 핵산; 가이드 서열을 포함하는 엔지니어링된 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 핵산; 및 공여자 핵산 분자 또는 상기 공여자 핵산 분자를 암호화하는 핵산을 포함하는, 표적 핵산의 편집 시스템 또는 표적 핵산의 유전자 편집 조성물이 제공된다.
- [13] 일 구현예에서, 상기 시스템은 표적 핵산에 이중가닥 절단(double-strand breaks)을 일으키는 것일 수 있다.
- [14] 다른 구현예에서, 상기 시스템은 표적 핵산 또는 그 인접 부위에서 공여자 핵산 분자를 주형으로 하여 이중가닥 절단의 상동지정복구(homology-directed repair) 과정에 의해 목적 서열이 도입되는 것일 수 있다.
- [15] 또 다른 구현예에서, 상기 Cas12f1, TnpB 또는 이의 변이체 단백질을 포함하는 엔도뉴클레아제; 및 상기 가이드 RNA를 리보뉴클레오단백질(ribonucleoprotein, RNP) 형태로 포함하는 것일 수 있다.

- [16] 또 다른 구현예에서, 상기 시스템 또는 조성물은 비상동말단연결(non-homologous end joining, NHEJ)에 관여하는 유전자의 발현 또는 그 발현물의 활성을 억제하는 분자를 더 포함할 수 있다.
- [17] 본 발명의 다른 태양에 따르면, Cas12f1, TnpB 또는 이의 변이체 단백질을 포함하는 엔도뉴클레아제를 암호화하는 뉴클레오티드 서열이 작동 가능하게 연결된 제1 핵산 구조물; 표적 핵산에 상보적으로 결합하는 가이드 서열을 포함하는 엔지니어링된 가이드 RNA를 암호화하는 뉴클레오티드 서열이 작동 가능하게 연결된 제2 핵산 구조물; 및 공여자 핵산 분자를 포함하는 제3 핵산 구조물을 포함하는 벡터 시스템이 제공된다.
- [18] 일 구현예에서, 상기 벡터 시스템에 포함하는 핵산 구조물은 동일하거나 상이한 벡터에 위치하는 것일 수 있다.
- [19] 다른 구현예에서, 상기 벡터 시스템은 비상동성말단(Non-homologous end joining DNA repair pathway; NHEJ)에 관여하는 유전자의 발현을 억제하는 분자를 암호화하는 뉴클레오티드 서열이 작동 가능하게 연결된 하나 이상의 핵산 구조물을 추가로 포함할 수 있다.
- [20] 또 다른 구현예에서, 상기 벡터 시스템은 벡터 내의 각 구성요소가 하나의 벡터에 포함되는 것일 수 있다.
- [21] 또 다른 구현예에서, 상기 벡터는 프로모터 또는 인핸서를 더 포함할 수 있다.
- [22] 또 다른 구현예에서, 상기 프로모터는 U6 프로모터, EFS 프로모터, EF1- α 프로모터, H1 프로모터, 7SK 프로모터, CMV 프로모터, LTR 프로모터, Ad MLP 프로모터, HSV 프로모터, SV40 프로모터, CBA 프로모터 또는 RSV 프로모터일 수 있다.
- [23] 또 다른 구현예에서, 상기 벡터는 레트로바이러스 벡터(retroviral(retrovirus) vector), 렌티바이러스 벡터(lentiviral(lentivirus) vector), 아데노바이러스 벡터(adenoviral(adenovirus vector), 아데노-연관 바이러스 벡터(adenovirus-associated viral(adenovirus-associated virus; AAV) vector), 백신시아바이러스 벡터(vaccinia viral(vaccinia virus) vector), 폭스바이러스 벡터(poxviral(poxvirus) vector), 단순포진 바이러스 벡터(herpes simplex viral(herpes simplex virus) vector) 및 파지미드 벡터(phagemid vector)로 구성된 군에서 선택되는 하나 이상의 바이러스 벡터일 수 있다.
- [24] 또 다른 구현예에서, 상기 벡터는 아데노-연관 바이러스 벡터이고, 상기 아데노-연관 바이러스 벡터는 상기 벡터 내의 구성요소 모두를 하나의 벡터에 포함할 수 있는 특징을 가질 수 있다.
- [25] 또 다른 구현예에서, 상기 벡터는 플라스미드, 네이키드 DNA, DNA 복합체, mRNA(전사물) 및 PCR 앰플리콘(amplicon)으로 이루어진 군에서 선택된 하나 이상의 비-바이러스 벡터일 수 있다.
- [26] 또 다른 구현예에서, 상기 플라스미드는 pcDNA 시리즈, pSC101, pGV1106, pACYC177, ColE1, pKT230, pME290, pBR322, pUC8/9, pUC6, pBD9, pHc79,

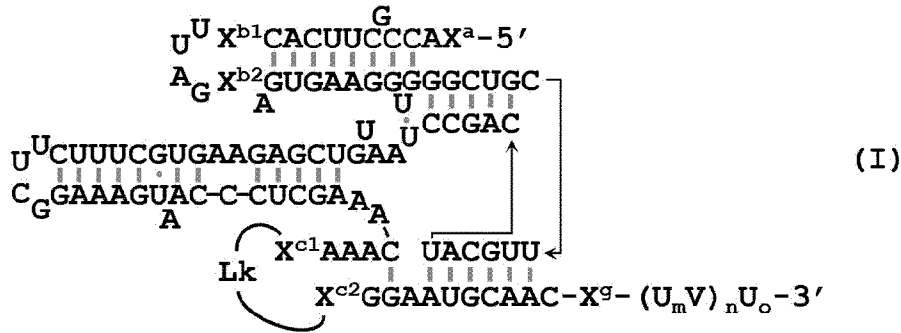
- pIJ61, pLAFR1, pHV14, pGEX 시리즈, pET 시리즈, 및 pUC19으로 이루어진 군에서 선택된 하나 이상일 수 있으나, 이에 제한되는 것은 아니다.
- [27] 본 발명의 또 다른 태양에 따르면, 상기 벡터 시스템에 의해 제조된 바이러스 또는 바이러스 입자 및 상기 바이러스 또는 바이러스 입자를 포함하는 조성물이 제공된다.
- [28] 일 구현예에서, 상기 바이러스는 레트로바이러스, 렌티바이러스, 아데노바이러스, 아데노-연관 바이러스, 백시니아바이러스, 폭스바이러스, 단순포진 바이러스 및 파지(phage)로 구성된 군에서 선택되는 것일 수 있다.
- [29] 본 발명의 또 다른 태양에 따르면, 상기 시스템 또는 조성물 또는 벡터 시스템을 세포와 접촉시키거나 세포 내에서 발현시키는 단계; 및 표적 핵산 또는 그 인접 부위에서 공여자 핵산 분자를 주형으로 하여 이중가닥 절단의 복구 과정에 의해 목적 서열이 도입되는 단계를 포함하는 세포 내 이중가닥 핵산 상의 표적 부위로 목적 서열을 도입하는 방법이 제공된다.
- [30] 일 구현예에서, 상기 이중가닥 절단의 복구 과정은 상동지정복구 기작에 의하는 것일 수 있다.
- [31] 다른 구현예에서, 상기 세포는 표적 핵산 또는 표적 유전자가 존재하는 원핵 세포 또는 진핵 세포일 수 있다.
- [32] 또 다른 구현예에서, 상기 진핵 세포는 효모(yeast), 곤충 세포, 식물 세포, 비인간-동물 세포 또는 인간 세포일 수 있다.
- [33] 또 다른 구현예에서, 상기 접촉 또는 발현은 생체 내 또는 생체 외에서 일어날 수 있다.
- [34] 또 다른 구현예에서, 상기 벡터 시스템은 레트로바이러스, 렌티바이러스, 아데노바이러스, 아데노-연관 바이러스, 백시니아바이러스, 폭스바이러스, 단순포진 바이러스 및 파지(phage)로 구성된 군에서 선택된 패키징 바이러스에 도입되어 패키징 바이러스에 의해 생성된 바이러스 형태로 원핵 세포 또는 진핵 세포 내로 전달되는 것일 수 있다.
- [35] 또 다른 구현예에서, 상기 벡터 시스템은 전기천공법, 유전자총, 초음파천공법, 자기주입법(magnetofection), 일시적인 세포 압축 또는 스퀴징 방법, 양이온성 리포솜법, 초산 리튬-DMSO, 지질-매개 형질감염(transfection), 인산칼슘 침전법(precipitation), 리포펙타민(lipofection), PEI(Polyethyleneimine)-매개 형질감염, DEAE-dextran 매개 형질감염 또는 나노파티클-매개 핵산 전달에 의해 원핵 세포 또는 진핵 세포 내로 전달될 수 있다.
- [36] 또 다른 구현예에서, 상기 벡터 시스템은 하나 이상의 지질 나노입자(lipid nanoparticles, LNP)를 통해 원핵 세포 또는 진핵 세포 내로 직접 전달될 수 있다.
- [37] 이하, 상기 본 발명의 복수 양태에 따른 각각의 시스템, 조성물, 벡터 시스템 및 방법에서 공통으로 적용되는 구현예를 기술한다.

- [38] 일 구현예에서, 상기 Cas12f1, TnpB 또는 이의 변이체 단백질은 서열번호 1 내지 서열번호 5로 이루어진 군에서 선택된 아미노산 서열과 70% 이상의 서열 동일성을 갖는 아미노산 서열을 포함할 수 있다.
- [39] 다른 구현예에서, 상기 TnpB 단백질은 서열번호 202 내지 서열번호 293으로 이루어진 군에서 선택된 아미노산 서열과 70% 이상의 서열 동일성을 갖는 아미노산 서열을 포함할 수 있다.
- [40] 또 다른 구현예에서, 상기 Cas12f1, TnpB 또는 이의 변이체 단백질은 (i) 서열번호 5의 아미노산 서열; (ii) 서열번호 1의 아미노산 서열; (iii) 서열번호 1의 아미노산 서열에서 N-말단의 1개 내지 28개의 아미노산이 제거 또는 치환된 아미노산 서열; 또는 (iv) 서열번호 1의 아미노산 서열의 N-말단 또는 C-말단에 1개 내지 600개의 아미노산이 추가된 아미노산 서열 중 하나의 서열을 포함할 수 있다.
- [41] 또 다른 구현예에서, 상기 추가된 1개 내지 600개의 아미노산은 서열번호 294 또는 서열번호 295의 아미노산 서열일 수 있다.
- [42] 또 다른 구현예에서, 상기 Cas12f1, TnpB 또는 이의 변이체 단백질은 서열번호 1 내지 서열번호 4로 이루어진 군에서 선택된 아미노산 서열과 70% 이상의 서열 동일성을 가질 수 있다.
- [43] 또 다른 구현예에서, 상기 Cas12f1, TnpB 또는 이의 변이체 단백질은 서열번호 1의 아미노산 서열과 70% 이상의 서열 동일성을 가질 수 있다.
- [44] 또 다른 구현예에서, 상기 Cas12f1, TnpB 또는 이의 변이체 단백질은 서열번호 5의 아미노산 서열과 70% 이상의 서열 동일성을 가질 수 있다.
- [45] 또 다른 구현예에서, 상기 엔지니어링된 가이드 RNA는 야생형 Cas12f1 가이드 RNA 서열에서 1개 이상의 뉴클레오티드가 치환, 결실, 삽입 또는 부가된 서열을 포함하고, 가이드 서열을 제외한 부분이 상기 야생형 Cas12f1 가이드 RNA와 50% 이상의 서열 동일성을 가질 수 있다.
- [46] 또 다른 구현예에서, 상기 야생형 Cas12f1 가이드 RNA는 (i) 하나 이상의 스템 (stem) 영역, (ii) tracrRNA-crRNA 상보성 영역 및 임의적으로 (iii) 연속되는 3개 이상의 유라실(U)을 포함하는 영역을 포함하는 tracrRNA(trans-activating CRISPR RNA) 및 crRNA(CRISPR RNA)를 포함하고, 상기 엔지니어링된 가이드 RNA는 다음 (a) 내지 (d)로 이루어진 군에서 선택되는 하나 이상의 변형을 포함할 수 있다: (a) 하나 이상의 스템 영역의 일부 또는 전부의 결실; (b) tracrRNA-crRNA 상보성 영역의 일부 또는 전부의 결실; (c) tracrRNA-crRNA 상보성 영역 내에 연속되는 3개 이상의 유라실(U)이 존재하는 경우 하나 이상의 U를 A, G 또는 C로 치환; 및 (d) crRNA 서열의 3'-말단에 하나 이상의 유리딘(uridine)의 부가.
- [47] 또 다른 구현예에서, 상기 야생형 Cas12f1 가이드 RNA는 서열번호 11의 핵산 서열로 이루어진 tracrRNA 및 서열번호 12의 핵산 서열로 이루어진 crRNA를 포함할 수 있다.
- [48] 또 다른 구현예에서, 상기 야생형 Cas12f1 가이드 RNA는 5'-말단부터 순차적으로 제1 스템 영역, 제2 스템 영역, 제3 스템 영역, 제4 스템 영역 및 tracrRNA-

crRNA 상보성 영역을 포함하는 tracrRNA 및 crRNA를 포함하고, 상기 엔지니어링된 가이드 RNA는 (a1) 제1 스템 영역의 일부 또는 전부의 결실; (a2) 제2 스템 영역의 일부 또는 전부의 결실; (b) tracrRNA-crRNA 상보성 영역의 일부 또는 전부의 결실; (c) tracrRNA-crRNA 상보성 영역 내에 연속되는 3개 이상의 유라실(U)이 존재하는 경우 하나 이상의 U를 A, G 또는 C로 치환; 및 (d1) crRNA 서열의 3'-말단에 U-rich tail의 부가(상기 U-rich tail의 서열은 5'-(U_mV)_nU_o-3'로 표시되고 여기서 상기 V는 각각 독립적으로 A, C 또는 G이고, m 및 o는 1 내지 20 사이의 정수이며, n은 0 내지 5 사이의 정수임)로 이루어진 군에서 선택되는 하나 이상의 변형을 포함할 수 있다.

- [49] 또 다른 구현예에서, 상기 엔지니어링된 가이드 RNA는 (d1) crRNA 서열의 3'-말단에 U-rich tail, (c) tracrRNA-crRNA 상보성 영역 내에 연속되는 3개 이상의 유라실(U)이 존재하는 경우 하나 이상의 U를 A, G 또는 C로 치환, 또는 상기 둘 모두의 변형을 포함할 수 있다.
- [50] 또 다른 구현예에서, 상기 엔지니어링된 가이드 RNA는 (a1) 제1 스템 영역의 일부 또는 전부의 결실; (a2) 제2 스템 영역의 일부 또는 전부의 결실; 및 (b) tracrRNA-crRNA 상보성 영역의 일부 또는 전부의 결실로 이루어진 군에서 선택된 하나 이상의 변형을 포함할 수 있다.
- [51] 또 다른 구현예에서, 상기 엔지니어링된 가이드 RNA는 (b1) tracrRNA-crRNA 상보성 영역의 일부의 결실을 포함할 수 있다. 구체적으로, 상기 상보성 영역의 일부는 1개 내지 54개의 뉴클레오티드일 수 있다.
- [52] 또 다른 구현예에서, 상기 엔지니어링된 가이드 RNA는 (b2) tracrRNA-crRNA 상보성 영역의 전부의 결실을 포함할 수 있다. 구체적으로, 상기 상보성 영역의 전부는 55개의 뉴클레오티드일 수 있다.
- [53] 또 다른 구현예에서, 상기 엔지니어링된 가이드 RNA는 (a1) 제1 스템 영역의 일부 또는 전부의 결실을 포함할 수 있다. 구체적으로, 상기 스템 영역의 일부 또는 전부는 1개 내지 20개의 뉴클레오티드일 수 있다.
- [54] 또 다른 구현예에서, 상기 엔지니어링된 가이드 RNA는 (a2) 제2 스템 영역의 일부 또는 전부의 결실을 포함할 수 있다. 구체적으로, 상기 스템 영역의 일부 또는 전부는 1 내지 27개의 뉴클레오티드일 수 있다.
- [55] 또 다른 구현예에서, 상기 엔지니어링된 가이드 RNA는 (a1) 제1 스템 영역의 일부 또는 전부의 결실; (d1) crRNA 서열의 3'-말단에 U-rich tail의 부가; 또는 상기 둘 모두의 변형을 포함할 수 있다.
- [56] 또 다른 구현예에서, 상기 엔지니어링된 가이드 RNA는 하기 식 (I)로 표시되는 서열로 이루어지거나 상기 서열과 80% 이상, 85% 이상, 90% 이상 또는 95% 이상의 서열 동일성을 가질 수 있다. 여기서 서열 동일성은 X^a, X^{b1}, X^{b2}, X^{c1} 및 X^{c2} 중 하나 이상을 포함하거나 포함하지 않은 서열을 기준으로 할 수 있다. 서열 동일성 산출에 Xg 및 (U_mV)_nU_o는 제외될 수 있다.

[57]



- [58] 상기 식 (I)에서, X^a , X^{b1} , X^{b2} , X^{c1} 및 X^{c2} 는 각각 독립적으로 0 내지 35개의 (폴리)뉴클레오티드로 이루어지고, X^g 는 10 내지 30개의 폴리뉴클레오티드로 이루어진 가이드 서열로서 표적 서열과 혼성화하거나 표적 서열에 상보적인 서열이고, Lk는 길이 2 내지 20의 폴리뉴클레오티드 링커이거나 부존재하고, $(U_mV)_nU_o$ 는 U-rich tail로서 존재하거나 부존재하고, 존재하는 경우 U는 유리딘이고, V는 각각 독립적으로 A, C 또는 G이고, m 및 o는 각각 독립적으로 1 내지 20 사이의 정수이며, n은 0 내지 5 사이의 정수이다.
- [59] 또 다른 구현예에서, 상기 X^a 는 서열번호 14의 핵산 서열을 포함하거나 상기 서열번호 14의 서열에서 1 내지 20개의 뉴클레오티드가 결실된 핵산 서열을 포함할 수 있다.
- [60] 또 다른 구현예에서, 상기 X^{b1} 은 서열번호 25의 핵산 서열을 포함하거나 상기 서열번호 25의 서열에서 1 내지 13개의 뉴클레오티드가 결실된 핵산 서열을 포함할 수 있다.
- [61] 또 다른 구현예에서, 상기 X^{b2} 은 서열번호 29의 핵산 서열을 포함하거나 상기 서열번호 29의 서열에서 1 내지 14개의 뉴클레오티드가 결실된 핵산 서열을 포함할 수 있다.
- [62] 또 다른 구현예에서, 상기 식 (I) 내의 서열 5'- X^{b1} UUAG X^{b2} -3'은 서열번호 34 내지 서열번호 38로 이루어진 군에서 선택된 어느 하나의 핵산 서열일 수 있다.
- [63] 또 다른 구현예에서, 상기 X^{c1} 은 서열번호 39의 핵산 서열을 포함하거나 상기 서열번호 39의 서열에서 1 내지 28개의 뉴클레오티드가 결실된 핵산 서열을 포함할 수 있다.
- [64] 또 다른 구현예에서, 상기 X^{c1} 서열 내에 연속되는 3개 이상의 유라실(U)이 존재하는 경우 이들 중 하나 이상의 U가 A, G 또는 C로 치환되는 변형을 포함할 수 있다.
- [65] 또 다른 구현예에서, 상기 X^{c2} 는 서열번호 58의 핵산 서열을 포함하거나 상기 서열번호 58의 서열에서 1 내지 27개의 뉴클레오티드가 결실된 핵산 서열을 포함할 수 있다.

- [66] 또 다른 구현예에서, 상기 X^{c2} 서열 내에 서열 5'-ACGAA-3'가 존재하는 경우 해당 서열이 5'-NGNNN-3'로 치환되고, 상기 N은 각각 독립적으로 A, C, G 또는 U일 수 있다.
- [67] 또 다른 구현예에서, 상기 식 (I) 내의 서열 5'- X^{c1} -Lk- X^{c2} -3'은 서열번호 80 내지 서열번호 86으로 이루어진 군에서 선택된 어느 하나의 핵산 서열일 수 있다.
- [68] 또 다른 구현예에서, 상기 Lk는 5'-GAAA-3', 5'-UUAG-3', 5'-UGAAAA-3', 5'-UUGAAAAA-3', 5'-UUCGAAAGAA-3'(서열번호 76), 5'-UUCAGAAAUGAA-3'(서열번호 77), 5'-UUCAUGAAAAUGAA-3'(서열번호 78) 및 5'-UUCAUUGAAAAAUGAA-3'(서열번호 79)로 이루어진 군에서 선택된 어느 하나의 핵산 서열을 포함할 수 있다.
- [69] 또 다른 구현예에서, 상기 $(U_mV)_nU_o$ 는 (i) n이 0이고, o가 1 내지 6 사이의 정수이거나 (ii) V가 A 또는 G이고, m 및 o는 각각 독립적으로 3 내지 6 사이의 정수이고, n은 1 내지 3 사이의 정수일 수 있다.
- [70] 또 다른 구현예에서, 상기 엔지니어링된 가이드 RNA는 서열번호 87 내지 서열번호 132로 이루어진 군에서 선택된 어느 하나의 핵산 서열로 이루어진 엔지니어링된 tracrRNA를 포함할 수 있다.
- [71] 또 다른 구현예에서, 상기 엔지니어링된 가이드 RNA는 서열번호 133 내지 서열번호 148로 이루어진 군에서 선택된 어느 하나의 핵산 서열로 이루어진 엔지니어링된 crRNA를 포함할 수 있다.
- [72] 또 다른 구현예에서, 상기 엔지니어링된 가이드 RNA는 듀얼 가이드 RNA 또는 싱글 가이드 RNA일 수 있다.
- [73] 또 다른 구현예에서, 상기 엔지니어링된 싱글 가이드 RNA는 서열번호 149 내지 서열번호 186으로 이루어진 군에서 선택된 어느 하나의 핵산 서열로 이루어진 것일 수 있다.
- [74] 또 다른 구현예에서, 상기 공여자 핵산 분자는 상동지정복구에서 주형으로 사용되는 서열로서 1 bp 내지 20 kb의 길이를 갖는 것일 수 있다.
- [75] 또 다른 구현예에서, 상기 비상동말단연결에 관여하는 유전자는 ATM1, XRCC4, XLF, XRCC6, LIG4 및 DCLRE1C로 이루어진 군으로부터 선택된 어느 하나 이상일 수 있으나, 이에 제한되지 않는다.
- [76] 또 다른 구현예에서, 상기 비상동말단연결에 관여하는 유전자의 발현을 억제하는 분자는 shRNA, siRNA, miRNA, 또는 안티센스 올리고뉴클레오타이드일 수 있으나, 이에 제한되지 않는다.

발명의 효과

- [77] 본 발명은 Cas12f1, 기존에 엔도뉴클레아제(endonuclease)로서 알려진 바 없는 새로운 TnpB 또는 이의 변이체 단백질 기반의 엔도뉴클레아제 및 엔도뉴클레아제와 함께 사용되어 우수한 상동지정복구 효율을 나타내는 엔지니어링된 가이드(engineered guide) RNA를 포함하는 표적 핵산 편집 시스템을 제공한다. 본 발

명의 표적 핵산 편집 시스템에 포함되는 Cas12f1, TnpB 또는 이의 변이체 단백질은 표적 핵산의 뒤쪽(바깥쪽)을 절단하는 경향성이 있어 다른 CRISPR/Cas 시스템에 비해 상동지정복구 효율이 매우 높다. 뿐만 아니라, 본 발명의 표적 핵산 편집 시스템은 초소형 엔도뉴클레아제 및 길이는 더 짧으면서도 우수한 편집 효율을 나타내는 엔지니어링된 가이드 RNA를 이용하므로, AAV와 같이 패키징 사이즈가 매우 제한적인 전달 매개체를 사용하는 경우에도 하나의 벡터에 공여자 핵산을 포함하여 상동지정복구를 위해 필요한 다양한 도구들을 탑재할 수 있으므로 상동지정복구를 포함한 다양한 형태의 유전자 편집을 위한 시스템을 구성할 수 있는 장점을 가진다. 특히, 비상동말단연결(NHEJ) 과정을 억제하는 분자, 예를 들어 shRNA를 추가할 수 있어 상동지정복구의 효율을 극대화할 수 있는 효과가 있다.

도면의 간단한 설명

- [78] 도 1은 일 실시예에 따른 표적 핵산 편집 시스템(TaRGET 시스템)의 엔지니어링된 가이드 RNA에서 각각의 변형부위를 도시한다(MS, modification site).
- [79] 도 2는 Cas9, Cas12a 및 TnpB(TaRGET)의 상동지정복구 효율을 비교하기 위한 공여자 핵산의 구조를 도시한다.
- [80] 도 3은 Cas9, Cas12a 및 TaRGET(Cas12f 및 TnpB)의 상동지정복구 효율을 비교한 결과를 도시한 그래프이다.
- [81] 도 4는 공여자 핵산의 길이에 따른 표적 핵산 편집 시스템의 상동지정복구 효율을 도시한 그래프이다.
- [82] 도 5는 상동지정복구를 위한 구성물을 HEK293T 세포에 주입한 후 Cas9 시스템과 표적 핵산 편집 시스템(TaRGET 시스템)의 비상동말단연결에 대한 상동지정복구 비율을 시간 경과에 따라 측정한 결과를 도시한 그래프이다.
- [83] 도 6은 표적 유전자 NLRC4, FUS 및 LOC105370393에 대해 Cas9, Cas12a 및 TaRGET(Cas12f 및 TnpB)의 상동지정복구 및 비상동말단연결의 비율을 도시한 그래프이다.
- [84] 도 7은 비상동말단연결 과정 또는 상동지정복구 과정에 관여하는 다양한 유전자의 발현을 억제하였을 때 표적 핵산 편집 시스템(TaRGET 시스템)의 상동지정복구 효율의 변화를 도시한 그래프이다.
- [85] 도 8은 *DCLRE1C* 유전자가 녹아웃된 HEK293T 세포와 야생형 HEK293T 세포에서 표적 핵산 편집 시스템(TaRGET 시스템)에 의한 비상동말단연결 및 상동지정복구 효율을 도시한 그래프이다.
- [86] 도 9는 *DCLRE1C* 유전자가 녹아웃된 HEK293T 세포와 야생형 HEK293T 세포에서 각각 Cas9, Cas12a 및 표적 핵산 편집 시스템(TaRGET 시스템)의 비상동말단연결에 대한 상동지정복구 비율을 도시한 그래프이다.
- [87] 도 10a 및 도 10b는 일 실시예에 따라 AAV에 패키징될 수 있는 4.7 kb 크기 내로 제작된 상동지정복구를 위한 구성물의 각 구성요소를 달리 하여 상동지정

복구 효율을 비교한 결과를 도시한다: 도 10a는 프로모터의 종류, 공여자 핵산의 길이 및 shDCLRE1C의 유무에 따라 각기 다르게 제작된 4종의 벡터를 도시한다; 도 10b는 4종 벡터의 상동지정복구 효율을 표적 유전자 NLRC4, FUS 및 LOC105370393에 대해 조사한 결과를 도시한 그래프이다.

발명의 실시를 위한 형태

- [88] 후술하는 본 발명에 대한 상세한 설명은, 본 발명이 실시될 수 있는 특정 구현에 관하여 특정 도면을 참조하여 기술될 것이지만, 본 발명은 이에 한정되지 않고, 적절하게 설명된다면, 그 청구항들이 주장하는 것과 균등한 모든 범위와 더불어 첨부된 청구항에 의해서만 한정된다. 본 발명의 다양한 구현예/실시예는 서로 다르지만 상호 배타적일 필요는 없음이 이해되어야 한다. 예를 들어, 본 명세서에 기재되어 있는 특정 형상, 구조 및 특성은 본 발명의 기술적 사상과 범위를 벗어나지 않으면서 일 구현예/실시예에서 다른 구현예/실시예로 변경되거나 구현예/실시예들이 조합되어 구현될 수 있다. 본 명세서에 사용된 기술 및 학술 용어들은, 달리 정의되지 않는 한, 본 발명이 속하는 분야에서 일반적으로 사용되는 것과 같은 의미를 갖는다. 본 명세서를 해석할 목적으로 하기 정의들이 적용될 것이고, 단수로 사용된 용어는 적절한 경우에는 복수형을 포함할 것이며 그 반대도 마찬가지이다.

[89] I. 정의

- [90] 본 명세서에서 사용된 용어 "표적 핵산(의) 편집 시스템", "유전자 편집 시스템", "유전자 복구 시스템" 또는 "상동지정복구 시스템"은 핵산 편집 단백질 또는 엔도뉴클레아제(endonuclease) 등의 핵산 분해효소, 상기 핵산 분해효소에 대응하는 핵산 표적화 분자 및 유전자 복구 기작에서 주형으로 사용되는 공여자 핵산이 포함된 시스템을 의미하는 것으로서, 표적 핵산 또는 표적 유전자에 결합 또는 상호작용하여 표적 핵산 또는 표적 유전자의 표적 부위를 절단, 편집, 수선 및/또는 복구할 수 있는 시스템을 의미한다. 여기서 핵산 표적화 분자는 엔지니어링된 가이드 RNA(gRNA)로 대표될 수 있으나 이에 제한되는 것은 아니다. 한편, 표적 핵산 편집 시스템은 표적 핵산 편집이 가능한 모든 형태로 존재할 수 있으며, 예를 들어, 핵산 분해효소와 핵산 표적화 분자를 포함하는 복합체 및 공여자 핵산을 포함하는 조성물 형태일 수 있으며, 상기 복합체와 상기 공여자 핵산을 각각 별개의 조성물에 포함된 키트 형태일 수 있으며, 핵산 분해효소를 암호화하는 핵산, 핵산 표적화 분자를 암호화하는 핵산 표적화 분자, 및 공여자 핵산을 암호화하는 핵산 표적화 분자를 포함하는 하나 이상의 벡터를 포함하는 벡터 시스템 또는 조성물일 수 있다.

- [91] 용어 "초소형 유전자 편집 시스템(Hypercompact TaRGET system)"은 초소형 CRISPR/Cas 단백질 또는 소형 엔도뉴클레아제 등(tiny endonuclease)(예컨대, Cas12f1, TnpB 또는 이들의 변이체)의 핵산 분해효소 및 상기 핵산 분해효소에 대응하는 핵산 표적화 분자가 포함된 유전자 편집 시스템으로서, 기존의 유전자 편

집 시스템과 차별되는 용어로 사용된다. 여기서 핵산 표적화 분자는 엔지니어링된 가이드 RNA(gRNA)로 대표될 수 있으나, 이에 제한되지는 않는다. 상기 시스템은 표적 핵산 또는 표적 유전자에 결합하여 표적 핵산 또는 유전자의 표적 부위를 절단, 편집, 수선 및/또는 복구할 수 있는 임의의 형태의 유전자 편집 시스템일 수 있다. 용어 "엔도뉴클레아제"는 "핵산 편집 단백질", "유전자 편집 단백질", "상동지정복구용 단백질" 또는 "핵산 분해 단백질"과 상호 교환적으로 사용될 수 있으며, 이들 엔도뉴클레아제 또는 단백질로 지칭되는 분자는 표적화하는 핵산인 DNA 또는 RNA, 또는 표적 유전자 내에 존재하는 프로토스페이스 인접 모티프(proto-spacer adjacent motif, PAM)를 인식한 후, 표적 핵산 서열의 내부 또는 외부 염기서열(sequence)에서 DNA 이중가닥 절단(double-strand breaks, DSBs)이 일어나게 할 수 있는 (엔도)뉴클레아제((endo)nuclease)를 의미한다. 또한, 상기 엔도뉴클레아제, 핵산 편집 단백질 등은 핵산 편집 시스템 또는 상동지정복구를 위한 핵산 구조물(construct)을 구성하는 효과기(effector) 단백질로도 지칭한다. 여기서 효과기(effector) 단백질은 가이드 RNA(gRNA) 또는 엔지니어링된 gRNA에 결합할 수 있는 핵산 분해 단백질이나, 표적 핵산 또는 표적 유전자에 결합할 수 있는 펩티드 단편일 수 있다.

- [92] 용어 "가이드 RNA(gRNA)"는 엔도뉴클레아제, 유전자 편집 단백질 또는 핵산 분해 단백질 등으로 지칭되는 분자와 복합체를 형성할 수 있고, 표적 핵산 서열과 상호작용(예컨대, 혼성화, 상보적 결합 또는 수소 결합 등)할 수 있으며, 표적 핵산 서열에 대한 복합체의 서열-특이적 결합(sequence-specific binding)을 야기하기에 충분한 정도로 표적 핵산 서열과 상보성을 갖는 가이드(guide) 서열을 포함하는 RNA를 의미한다. 본 명세서에서 가이드 RNA 또는 가이드 분자는 상호 교환적으로 사용될 수 있다.
- [93] 용어 "tracrRNA(trans-activating crRNA)" 및 "crRNA(CRISPR RNA)"는 유전자 편집 기술 분야에서 통상의 기술자가 인식할 수 있는 의미를 모두 포함한다. 이는 자연계에서 발견되는 듀얼 가이드 RNA(dual guide RNA)의 각 분자를 지칭하는 용어로 사용될 수 있고, 상기 tracrRNA 및 crRNA를 링커로 연결한 싱글 가이드 RNA(single guide RNA, sgRNA)의 각 해당 부분을 지칭하는데도 사용될 수 있다. 달리 서술하지 않는 한, tracrRNA 및 crRNA라고만 기재하는 경우 표적 핵산의 편집 시스템(또는 유전자 편집 시스템, 상동지정복구용 시스템 등)에서 가이드 RNA를 구성하는 tracrRNA 및 crRNA를 의미한다.
- [94] 용어 "스캐폴드(Scaffold) 영역"은 가이드 RNA(gRNA) 중 엔도뉴클레아제, 상동지정복구용 단백질, 유전자 편집 단백질 또는 핵산 분해 단백질 등으로 지칭되는 분자와 상호작용할 수 있는 부분을 통틀어 지칭하며, 자연계에서 발견되는 가이드 RNA의 부분 중 스페이스(spacer)를 제외한 나머지 부분을 지칭하는데 사용될 수 있다.
- [95] 용어 "가이드 서열(guide sequence)", "스페이스(space)" 또는 "스페이스 서열(spacer sequence)"은 상호 교환적으로 사용될 수 있으며, CRISPR/Cas 시스템에서

표적 서열 부분과 상호작용(예를 들어, 혼성화, 상보적 결합 또는 수소 결합 등)할 수 있는 폴리뉴클레오티드를 의미한다. 예컨대, 가이드 서열 또는 스페이서 서열은 표적 핵산 편집 시스템에서 가이드 RNA를 구성하는 crRNA의 3'-말단부 또는 3'-말단 부근에 직접 또는 링커 등을 통해 간접적으로 연결된 10개 내지 50개의 연속된 뉴클레오티드를 지칭한다.

- [96] 용어 "엔지니어링된(engineered)"는 "비-자연 발생적(non-naturally occurring)", "인공적(artificial)" 또는 "조작된(modified)"과 상호 교환적으로 사용될 수 있으며, 자연에서 발견되는 그대의 형태, 상태 등이 아님을 의미한다. 본 용어가 가이드 RNA, 가이드 폴리뉴클레오티드 또는 핵산 분자를 지시하는 경우, 가이드 RNA, 가이드 폴리뉴클레오티드 또는 핵산 분자는 자연에서 발견되거나 자연 발생적인 적어도 하나의 성분을 실질적으로 함유하지 않거나, 또는 자연에서 발견되지 않거나 비-자연 발생적인 적어도 하나의 성분을 실질적으로 함유함을 의미한다. 예를 들어, "엔지니어링된 가이드 RNA(engineered guide RNA)"는 자연계에 존재하는 가이드 RNA(gRNA)의 구성(예를 들어, 서열)에 인위적인 변형이 가해진 gRNA를 의미하며, 본 명세서 내에서 "augmented RNA"로 지칭될 수 있다.
- [97] 용어 "야생형"은 당업자에 의해 이해되는 해당 분야의 용어이며, 그것이 돌연변이체 또는 변이체 형태로부터 구별되는 정도로 천연에서 발생하는 것과 같은 전형적인 형태의 유기체, 균주, 유전자 또는 특징을 의미한다. 용어 "변이체"는 천연에서 발생하는 것에서 벗어난 패턴을 갖는 특성의 표현을 의미하는 것으로 이해해야 한다. 예컨대, Cas12f1, TnpB 또는 이의 변이체 단백질이라고 기재할 때, 상기 변이체 단백질은 (야생형) Cas12f1의 변이체 또는 (야생형) TnpB의 변이체를 의미할 수 있다.
- [98] 용어 "공여(자) 핵산(donor nucleic acid) 분자"는 "공여(자) DNA(donor DNA)", "공여자 폴리뉴클레오티드", "공여자 올리고뉴클레오티드" 및 "공여자 주형"과 상호 교환적으로 사용될 수 있으며, 핵산 서열의 적어도 일부가 선택된 표적 핵산 또는 표적 유전자의 표적 부위 내로 통합되는 것이 의도된 서열을 제공하는 핵산 또는 폴리뉴클레오티드를 의미한다. 전형적으로, 공여자 핵산은 단일가닥 폴리뉴클레오티드 또는 이중가닥 폴리뉴클레오티드이다. 예를 들어, 본 발명의 표적 핵산 편집 시스템은 공여자 핵산을 포함하여 게놈 DNA 내의 DNA 표적 서열을 변형 및/또는 복구시킬 수 있으며, 여기서 게놈 DNA는 DNA 표적 서열에서 공여자 핵산의 적어도 일부를 포함하도록 변형 및/또는 복구된다.
- [99] 용어 "표적 핵산(Target nucleic acid)" 또는 "표적 유전자(Target gene)"는 표적 핵산 편집 시스템(예를 들면, 상동지정복구용 시스템 또는 TaRGET 시스템)에 의한 유전자의 절단, 편집, 수선 및/또는 복구의 대상 또는 표적화 대상이 되는 유전자 또는 핵산을 의미한다. 표적 핵산 또는 표적 유전자는 혼용될 수 있으며, 서로 동일한 대상을 지칭할 수 있다. 상기 표적 유전자는 달리 기재되지 않은 한, 대상 세포가 가진 고유한 유전자 또는 핵산 혹은 외부 유래의 유전자 또는 핵산, 또는 인위적으로 합성된 핵산 또는 유전자일 수 있고, 단일가닥 DNA, 이중가닥 DNA 및/

또는 RNA 모두를 의미할 수 있다. 상기 표적 유전자 또는 표적 핵산은 본 발명에 따른 시스템에 의한 유전자 절단, 편집, 수선 등의 대상이 될 수 있다면 특별히 제한되지 않는다.

[100] 용어 "표적 부위(Target region)" 또는 "표적 서열(Target sequence)"은 표적 핵산 또는 표적 유전자 내 또는 그 주변에 존재하는 서열로, 본 발명의 초소형 핵산 편집 시스템이 표적 유전자 또는 표적 핵산을 절단하기 위해 인식하는 특정 서열을 의미한다. 상기 표적 부위 또는 표적 서열은 그 목적에 따라 적절히 선택될 수 있다.

[101] 용어 "상동지정복구(homology-directed repair, HDR)"는 이중가닥 DNA의 파손, 병변 등을 복구하는 세포 내 기작으로서, 상동지정복구의 가장 일반적인 형태는 상동재조합(homologous recombination)이다. 상동지정복구는 주로 세포 주기의 G2 및 S 기에서 DNA의 상동 조각이 핵에 존재하는 경우에 이중가닥 DNA의 파손, 병변 등을 복구하기 위한 세포 기작 중의 하나를 지칭한다. 상동지정복구는 복구를 프로그램하는 공여자 DNA를 주형으로 사용하며, 유전자의 의도된 부가를 비롯한 특정한 서열 변화를 게놈에 생성하는데 사용될 수 있다. 공여된 주형이 부위 특이적 뉴클레아제와 함께, 예컨대 본 발명의 시스템 또는 TaRGET 시스템과 함께 제공된다면, 세포 기구는 상동재조합에 의해 이중가닥 상의 파단을 복구할 것이며, 이 기작은 DNA 이중절단의 존재 하에 증진된다. 공여자 DNA가 존재하는 경우 상동지정복구와 비상동말단연결은 경합하여 동시에 발생하며 공여자 DNA가 부존재하는 경우에는 비상동말단연결만이 일어난다.

[102] 용어 "벡터"는 달리 특정되지 않는 한, 유전 물질을 세포 내로 운반할 수 있는 모든 물질을 통틀어 일컫는다. 예를 들어, 벡터는 전달 대상이 되는 유전 물질인 표적 핵산 편집 시스템의 효과기(effector) 단백질을 암호화하는 핵산 및/또는 가이드 RNA(gRNA)를 암호화하는 핵산을 포함하는 DNA 분자일 수 있으나, 이에 제한되는 것은 아니다. 또한, 본 발명에서 "벡터"는 삽입된 유전자가 정상적으로 발현되도록 작동 가능하게 연결된 필수적인 조절 요소를 포함하는 "발현 벡터"일 수 있다. 용어 "작동 가능하게 연결된(operably linked)"은 유전자 발현 기술에 있어서, 특정 구성이 다른 구성과 연결되어, 상기 특정 구성이 의도된 방식대로 기능할 수 있도록 연결되어 있는 것을 의미한다.

[103] 용어 "뉴클레오티드" 및 "핵산"은 상호 교환적으로 사용될 수 있고, 리보뉴클레오티드 또는 디옥시뉴클레오티드 중 하나의 임의의 길이 뉴클레오티드의 중합체 형태를 지칭한다. 따라서 이 용어는 단일-, 이중-, 또는 다중-가닥 DNA 또는 RNA, 게놈 DNA, cDNA, DNA-RNA 혼성체, 또는 퓨린 및 피리미딘 염기 또는 다른 천연, 화학적 또는 생화학적으로 변형된, 비천연 또는 유도체화된 뉴클레오티드 염기를 포함하는 중합체를 포함하지만, 이들로 제한되지 않는다. 용어 "폴리뉴클레오티드" 및 "핵산"은 본 명세서에 기재되는 구현예에 적용 가능한, 단일-가닥(예컨대 센스 또는 안티센스) 및 이중-가닥 폴리뉴클레오티드를 포함하는 것으로 이해되어야 한다.

- [104] 용어 "핵산 구조물(Nucleic acid construct)"은 엔도뉴클레아제, 핵산 편집 단백질 또는 핵산 분해 단백질 등을 암호화하는 뉴클레오티드 서열 및/또는 가이드 RNA를 암호화하는 뉴클레오티드 서열을 구성요소로 포함하는 구조물로서, 필요에 따라 다양한 종류의 (폴리)펩티드 또는 링커를 암호화하는 뉴클레오티드 서열을 추가로 포함할 수 있다. 상기 핵산 구조물은 본 발명의 상동지정복구를 위한 CRISPR/Cas 시스템, 벡터 시스템, 또는 초소형 유전자편집 시스템(Hypercompact TaRGET system)을 이루는 구성요소로 사용될 수 있다.
- [105] 용어 "단백질", "폴리펩티드" 및 "펩티드"는 상호 교환적으로 사용될 수 있으며, 유전적으로 암호화된 그리고 비유전적으로 암호화된 아미노산, 화학적 또는 생화학적으로 변형되거나 또는 유도체화된 아미노산, 및 변형된 펩티드 골격을 갖는 폴리펩티드를 포함할 수 있는 임의의 길이를 갖는 아미노산 중합체 형태를 지칭한다. 상기 용어는 N-말단의 메티오닌 잔기가 있거나 없는, 이중성 아미노산 서열과의 융합 단백질, 이중성 및 상동성 리더 서열과의 융합; 면역학적으로 태그된 단백질 등을 포함하지만, 이들로 제한되지 않는 융합 단백질을 모두 포괄한다.
- [106] 용어 "A, T, C, G 및 U"는 문맥 및 기술에 따라 DNA 또는 RNA 상에서 염기(base), 뉴클레오시드(nucleoside) 또는 뉴클레오티드(nucleotide)로 적절히 해석될 수 있다. 예를 들어, 염기를 의미하는 경우는 각각 아데닌(adenine; A), 구아닌(guanine; G), 시토신(cytosine; C), 티민(thymine; T) 및 유라실(uracil; U) 중 선택된 하나로 해석될 수 있다. 뉴클레오시드를 의미하는 경우는 각각 아데노신(adenosine; A), 티미딘(thymidine; T), 시티딘(cytidine; C), 구아노신(guanosine; G) 또는 유리딘(uridine; U)으로 해석될 수 있으며, 서열에서 뉴클레오티드를 의미하는 경우는 상기 각각의 뉴클레오시드를 포함하는 뉴클레오티드를 의미하는 것으로 해석되어야 한다.
- [107] 용어 "약"은 참조 양, 수준, 값, 수, 빈도, 퍼센트, 치수, 크기, 양, 중량 또는 길이에 대해 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 또는 1% 정도로 변하는 양, 수준, 값, 수, 빈도, 퍼센트, 치수, 크기, 양, 중량 또는 길이를 의미한다. 예를 들어, 용어 "약"은 숫자 또는 수치로 표현된 값 x 와 관련하여 사용될 때 $x \pm 5\%$ 를 의미할 수 있다.
- [108] 본 발명에서 사용되는 모든 기술용어는, 달리 정의되지 않는 이상 이 기술 분야의 통상의 기술자가 인식할 수 있는 의미를 모두 포함하고, 일반적으로 이해하는 바와 같은 의미로 사용되며, 문맥에 따라 적절히 해석될 수 있다. 또한, 본 명세서에는 바람직한 방법이나 시료가 기재되나, 이와 유사하거나 동등한 것들도 본 발명의 범주에 포함된다.
- [109] **II. 고효율의 상동지정복구를 위한 표적 핵산의 편집 시스템 및 이를 포함하는 조성물**
- [110] 본 발명자들은 TnpB(Transposon-associated transposase B) 단백질이 UnCas12f1 단백질과 유사한 아미노산 서열을 가지며(이에 따라, UnCas12f1 단백질과 유사

한 아미노산 서열을 갖는 TnpB는 CWCas12f1이라고도 명명된다), 현재까지 가장 많은 연구가 진행된 Cas9 단백질을 포함하는 기존 핵산 분해 단백질들보다 분자량은 1/3 정도로 작고, 표적 핵산 또는 표적 유전자에 대한 핵산 절단 효율이 월등히 높은 것을 확인하였다. 또한, Cas12f1, TnpB 및 이의 변이체 단백질이 고효율의 유전자 편집 단백질 활성을 나타내며, 특히 공여자 핵산이 존재하는 환경에서 상동지정복구를 유도하는 것을 처음으로 규명하였다.

- [111] 또한, 본 발명자들은 세포 내에서 표적 핵산 또는 표적 유전자를 절단, 편집, 수선 및/또는 복구하기 위해, 아데노 연관 바이러스(AAV) 벡터에 여윌롭게 탑재가 가능하며 효과적인 세포 내(*in vivo*) 전달이 가능한 시스템으로서 초소형의 핵산 편집 단백질인 Cas12f1, TnpB 또는 이들의 변이체 단백질을 포함하는 소형 엔도뉴클레아제; 상기 엔도뉴클레아제에 대해 높은 인델(indel; insertion or deletion) 효율을 나타내는 엔지니어링된 가이드 RNA; 및 공여자 핵산 분자를 포함하는 상동지정복구를 위한 새로운 초소형 표적 핵산의 편집 시스템을 제작하였다.
- [112] 본 발명자들은 기존에 알려진 Cas9 또는 Cas12a 등의 Cas 엔도뉴클레아제(endonuclease)가 아닌, 새로운 초소형 핵산 절단 단백질인 Cas12f1, TnpB 또는 이의 변이체 단백질을 이용함으로써 보다 효율적이고 응용 범위가 증대된 유전자 편집, 특히 상동지정복구의 유도가 가능함을 최초로 확인하였고, 이를 단일 아데노 연관 바이러스(AAV) 벡터에 모두 포함시켜 세포 내 전달이 가능하면서도 다양한 유전자에서 효율적으로 상동지정복구를 유도할 수 있는 새로운 초소형 핵산 편집 시스템을 구축하여 본 발명을 완성하였다.
- [113] 따라서 본 발명은 표적 핵산 또는 표적 유전자에서 부위(서열) 특이적으로 또한 고효율로 상동지정복구가 일어나도록 또는 유도되도록 하기 위한 Cas12f1, TnpB 또는 이의 변이체(variant) 단백질을 포함하는 엔도뉴클레아제 또는 상기 엔도뉴클레아제를 암호화하는 핵산; 가이드 서열을 포함하는 엔지니어링된 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 핵산; 및 공여자 핵산 분자 또는 상기 공여자 핵산 분자를 암호화하는 핵산을 포함하는 초소형 표적 핵산 편집 시스템(또는 TaRGET 시스템)에 관한 것이다.
- [114] 또한, 본 발명은 Cas12f1, TnpB 또는 이의 변이체(variant) 단백질을 포함하는 엔도뉴클레아제 또는 상기 엔도뉴클레아제를 암호화하는 핵산; 가이드 서열을 포함하는 엔지니어링된 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 핵산; 및 공여자 핵산 분자 또는 상기 공여자 핵산 분자를 암호화하는 핵산을 포함하는 표적 핵산의 유전자 편집 조성물, 상기 시스템 또는 조성물을 이용한 표적 핵산 또는 표적 유전자의 표적 부위로 목적 서열을 도입하는 방법에 관한 것이다.
- [115] 그 외에도, 본 발명을 구현하기 위한 다른 편집 시스템, 다른 방법, 조성물, 벡터 시스템, 바이러스 또는 바이러스 조성물에 관한 것이다.
- [116] 본 발명에 따른 상동지정복구를 위한 표적 핵산 편집 시스템은 기존에 연구된 대부분의 Cas 엔도뉴클레아제 및 이를 포함하는 유전자 편집 시스템이 그 크

기로 인하여 세포 내 전달 매개체로써 FDA 승인을 받은 아데노 연관 바이러스 (AAV) 벡터에 탑재할 수 없었던 제약을 해결한 의미 있는 결과이다.

- [117] 나아가, 상기 표적 핵산 편집 시스템은 고효율의 상동지정복구 유도에 필요한 충분한 크기의 공여자 핵산 분자를 포함하더라도 여전히 AAV 벡터의 패키징 한계인 4.7 kb 정도 크기로 제작될 수 있으며, 표적 핵산 또는 표적 유전자의 이중가닥 표적 부위를 절단하고 상동지정복구를 유도함으로써 원하는 표적 부위에 목적하는 서열의 도입이 가능하므로, 본 발명에 따른 표적 핵산 편집 시스템은 표적 핵산 또는 표적 유전자에서 상동지정복구를 포함한 다양한 형태의 유전자 편집을 위한 시스템을 구성할 수 있으며, 유전자 관련 질병에 대한 새로운 치료제로서도 폭넓게 적용할 수 있다.
- [118] 이하, 본 발명에서 제공되는 표적 핵산의 편집 시스템/조성물의 각 구성요소 및 이의 제조 방법을 상세히 설명한다.
- [119] **1. Cas12f1, TnpB 및 이의 변이체 단백질**
- [120] 본 발명의 일 태양에 따르면, 표적 핵산의 표적 부위를 절단하여 상동지정복구를 유도함에 있어 우수한 활성을 나타내고, 기존의 CRISPR/Cas9 시스템에 비하여 핵산 분해 단백질의 크기가 1/3 정도로 현저히 작은 것을 특징으로 하는 Cas12f1, TnpB 또는 이의 변이체(variant) 단백질을 포함하는 (소형) 엔도뉴클레아제가 제공된다.
- [121] 본 발명의 표적 핵산 편집 시스템에 포함되는 엔도뉴클레아제 또는 핵산 편집 단백질인 Cas12f1, TnpB 또는 이의 변이체 단백질은 자연계에서 발견되는 Cas12f1, Cas12f1 변이체, TnpB, TnpB 변이체, 또는 엔지니어링된 Cas12f1 또는 엔지니어링된 TnpB를 모두 포함한다. 상기 엔도뉴클레아제 또는 핵산 편집 단백질은 Cas12f1, TnpB 또는 이들의 변형된 핵산 편집 단백질, 예를 들어 상기 단백질에서 하나 이상의 아미노산이 결실, 추가 또는 치환된 단백질, dead 핵산 편집 단백질 또는 nick 핵산 편집 단백질일 수 있지만, 이들로 제한되지 않는다.
- [122] 본 발명의 구성 "Cas12f1 단백질"은 문헌[Harrington et al., Science, 362, 839-842 (2018)]에서 Cas14로 명명된 이펙터 단백질 중 하나로, Cas14a1 단백질로도 불린다. Cas12f1 단백질은 자연계에 존재하는 야생형(wildtype) Cas12f1 단백질(야생형 Cas14a1 단백질)일 수 있다. 또는, Cas12f1 단백질은 야생형 Cas12f1 단백질의 변이체(variant)일 수 있으며, 이때 상기 변이체는 "Cas12f1 변이체(Cas12f1 variant)" 또는 "Cas14a1 변이체(Cas14a1 variant)"로 지칭된다. 상기 Cas12f1 변이체는 야생형 Cas12f1 단백질과 동일한 기능을 가지는 변이체, 기능 일부 또는 전부가 변형된 변이체 및/또는 추가적인 기능이 부가된 변이체일 수 있다.
- [123] 일부 구현예에서, Cas12f1, TnpB 또는 이의 변이체 단백질은 서열번호 1 내지 서열번호 5로 이루어진 군에서 선택된 아미노산 서열과 70% 이상의 서열 동일성을 갖는 아미노산 서열을 포함할 수 있다.
- [124] 일부 구현예에서, TnpB 또는 이의 변이체 단백질은 Class 2, type V CRISPR/핵산 분해 단백질 중 V-F 서브타입에 속하는 Cas12f1 단백질과 크기가 유사한

IS200/IS605 패밀리의 전이효소 부속 단백질(transposase accessory protein) TnpB 단백질에서 유래된 서열이거나 이를 포함할 수 있다. TnpB 단백질은 종래에 전이효소(transposase)로 알려진 단백질이다. 현재까지 TnpB 단백질은 전이인자(transposon)을 암호화하는 핵산 분해 단백질(transposon-encoded nuclease)로만 알려져 있을 뿐, TnpB 단백질이 Cas 엔도뉴클레아제(endonuclease) 활성을 가지는지 알려진 바는 없다. 또한, TnpB 단백질에 대한 가이드 RNA도 알려진 바 없다. 본 발명은 부분적으로 TnpB 단백질 서열 기반의 TnpB 변이체 또는 엔지니어링된 TnpB가 핵산 분해 단백질 중 분자량이 가장 작은 그룹에 속하는 Cas12f1 단백질과 그 크기가 유사하면서, 표적 핵산 또는 표적 유전자를 표적화하여 표적 부위의 이중가닥 DNA를 절단하는 탁월한 엔도뉴클레아제(endonuclease) 활성을 가지고 있음을 처음으로 확인하고, TnpB 또는 이의 변이체 단백질과 함께 사용되어 우수한 상동지정복구 유도 효율을 나타내는 엔지니어링된 가이드 RNA(engineered guide RNA)를 제작함으로써 완성되었다. TnpB 또는 이의 변이체 단백질은 현존하는 핵산 분해 단백질 중 분자량이 가장 작은 그룹에 속하며, 본 발명의 엔지니어링된 짧은 가이드 RNA(gRNA)와 복합체를 형성하여 표적 핵산 또는 표적 유전자를 표적화하여 이중가닥을 절단하는 탁월한 효과를 가지고 있고, 공여자 핵산 분자가 함께 존재하는 경우 고효율의 상동지정복구 과정을 유도할 수 있어 세포 내 표적 유전자에 목적 서열을 도입하기 위한 초소형의 핵산 편집 시스템을 제작하는데 있어 큰 장점이 있다. 또한, 상기 TnpB 또는 이의 변이체 단백질은 5'-NGG-3'를 PAM으로 갖는 Cas9과 달리, 5'-TTTA-3' 또는 5'-TTTG-3' 등과 같은 T-rich PAM을 PAM으로 갖기 때문에, 티민(T)이 많은 서열을 표적 핵산 또는 표적 유전자로 선택할 수 있게 하여, 유전체 편집을 위한 핵산 분해 단백질의 선택 폭을 넓혀준다.

- [125] 일부 구현예에서, Cas12f1, TnpB 또는 이의 변이체 단백질은 (i) 서열번호 5의 아미노산 서열; (ii) 서열번호 1의 아미노산 서열; (iii) 서열번호 1의 아미노산 서열에서 N-말단의 1개 내지 28개의 아미노산이 제거 또는 치환된 아미노산 서열; 또는 (iv) 서열번호 1의 아미노산 서열의 N-말단 또는 C-말단에 1개 내지 600개의 아미노산이 추가된 아미노산 서열 중 하나를 포함하는 것일 수 있다.
- [126] 구체적으로, 일 구현예에 따르면 Cas12f1 또는 이의 변이체 단백질은 서열번호 5의 아미노산 서열을 포함하거나 이로 이루어진 단백질 또는 상기 서열번호 5의 아미노산 서열과 70% 이상의 서열 동일성을 갖는 아미노산 서열을 포함하거나 이로 이루어진 변이체 단백질일 수 있다.
- [127] 다른 구현예에 따르면, TnpB 또는 이의 변이체 단백질은 서열번호 1의 아미노산 서열을 포함하거나 이로 이루어진 단백질 또는 상기 서열번호 1의 아미노산 서열과 70% 이상의 서열 동일성을 갖는 아미노산 서열을 포함하거나 이로 이루어진 단백질일 수 있다. 또한, TnpB 변이체 단백질은 서열번호 1의 아미노산 서열에서 N-말단의 1개 내지 28개의 아미노산이 제거 또는 치환된 아미노산 서열을 포함하거나 이로 이루어진 TnpB 변이체 단백질일 수 있다. 이때, TnpB 변이체

단백질은 서열번호 5의 아미노산 서열로 이루어진 Cas12f1 단백질을 포함하지 않는다. 구체적으로, TnpB 또는 이의 변이체 단백질은 서열번호 1 내지 서열번호 4로 이루어진 군에서 선택된 아미노산 서열과 70% 이상의 서열 동일성을 갖는 단백질일 수 있다. 예를 들어, TnpB 변이체 단백질은 서열번호 1의 아미노산 서열과 70% 이상의 서열 동일성을 갖는 단백질을 포함한다.

[128] 다른 구현예에서, Cas12f1 변이체 단백질은 서열번호 5의 아미노산 서열로 이루어진 Cas12f1 단백질에 하나 이상의 아미노산이 더 포함된 것일 수 있다. 일부 구현예에서, Cas12f1 변이체 단백질은 TnpB 변이체 단백질을 포함한다. 예를 들어, Cas12f1 단백질의 N-말단에 CasX의 N-말단 26aa를 포함하는 TnpB-v1 단백질(서열번호 2), 28aa 무작위 서열을 포함하는 TnpB-v2 단백질(서열번호 3) 또는 26aa 무작위 서열을 포함하는 TnpB-v3 단백질(서열번호 4)을 포함하거나 이로 이루어질 수 있다.

[129] 일부 구현예에서, TnpB 또는 이의 변이체 단백질은 다른 생물 종에서 유래한 TnpB 단백질이거나 이로부터 유래된 변이체를 포함할 수 있다. 즉, TnpB 단백질은 유의미한 서열 동일성은 나타나지 않지만 다른 생물에서 동일한 기능을 하는 TnpB 동족체(homolog) 단백질을 포함한다. 구체적으로, TnpB 단백질 또는 이의 변이체 단백질은 서열번호 202 내지 서열번호 293으로 이루어진 군에서 선택된 어느 하나의 아미노산 서열을 포함하거나 이로 이루어진 것일 수 있다. TnpB 단백질(또는 TnpB 동족체 단백질)은 TnpB 단백질과 동일한 생체 내 활성(즉, 엔도뉴클레아제 활성)을 공유하는 단백질을 의미하는 것으로서 이들의 서열 유사성(또는 동일성)과는 무관하게 공통 조상으로부터 유래된 특징이 소실되지 않고 보존되어 있는 단백질을 의미한다.

[130] 또 다른 구현예에서, TnpB 또는 이의 변이체 단백질은 서열번호 1의 아미노산 서열의 N-말단 또는 C-말단에 1개 내지 600개의 아미노산이 추가된 아미노산 서열 중 하나를 포함하거나 이로 이루어진 단백질일 수 있다. 이때 추가된 1개 내지 600개의 아미노산 서열에는 제한이 없다. 일 예로, 상기 추가된 1개 내지 600개의 아미노산은 서열번호 294 또는 서열번호 295의 아미노산 서열일 수 있다. 추가 서열과 TnpB 변이체 단백질 사이에는 NLS 또는 NES 서열이 더 포함될 수 있다.

[131] 또한, Cas12f1, TnpB 또는 이의 변이체 단백질은 야생형의 Cas12f1 단백질과 동종의 기능을 가지거나, 야생형의 Cas12f1 단백질과 비교할 때, 기능이 변경된 것일 수 있다. 보다 구체적으로, 상기 변경은 전부 또는 일부 기능의 변형, 전부 또는 일부 기능의 상실 및/또는 부가적인 기능의 추가를 포함한다. Cas12f1, TnpB 또는 이의 변이체 단백질은 통상의 기술자가 초소형 핵산 편집 시스템의 핵산 분해 단백질에 적용할 수 있는 변경이라면, 특별히 제한 없이 임의의 변경을 포함할 수 있다. 예컨대, 상기 Cas12f1 변이체 단백질, TnpB 또는 이의 변이체 단백질은 DNA 이중가닥을 절단하는 활성뿐만 아니라, 단일가닥 DNA 또는 RNA, 또는 DNA 및 RNA의 혼성 이중가닥을 절단하는 활성, 염기 교정 및/또는 프라임 교정을 수행하기 위한 것일 수 있다.

- [132] 일부 구현예에서, 본 발명의 표적 핵산 편집 시스템은 표적 핵산 또는 표적 유전자의 표적 부위에서 핵산을 절단하는 것이므로, 표적 부위가 세포의 핵 내에 위치하는 것을 특징으로 할 수 있다. 이에, 본 발명의 표적 핵산 편집 시스템에 사용되는 Cas12f1, TnpB 또는 이의 변이체 단백질은 이를 핵 내로 위치시키는 핵 위치 신호(nuclear localization signal, NLS) 서열을 1개 또는 2개 이상 포함할 수 있다. 예컨대, 하나 이상의 핵 위치 신호 서열은 상기 Cas12f1, TnpB 또는 이의 변이체 단백질이 진핵세포(포유동물 세포 포함)의 핵에서 검출 가능한 양으로 핵 내로 표적화되도록 유도하는 데 충분한 양 또는 활성을 강도를 가질 수 있다. 예컨대, 그 활성의 강도 차이는 Cas12f1, TnpB 또는 이의 변이체 단백질 내에 포함되는 NLS의 수, 사용되는 특정 NLS(들)의 종류 또는 이들 인자의 조합으로부터 야기될 수 있다.
- [133] 또한, 다른 구현예에서, Cas12f1, TnpB 또는 이의 변이체 단백질에 포함되는 NLS는 N-말단에서 또는 그 근처에서 약 1, 2, 3, 4, 5, 6, 7, 8, 9 또는 10개 이상의 NLS, C-말단에서 또는 그 근처에서 약 1, 2, 3, 4, 5, 6, 7, 8, 9 또는 10개 이상의 NLS, 또는 이들의 조합이 다양하게 선택될 수 있다. 예컨대, N-말단에서 0 또는 적어도 하나 이상의 NLS 서열 및/또는 C-말단에서 0 또는 하나 이상의 NLS 서열을 포함할 수 있다. 하나 초과 NLS 서열이 존재할 때, 단일 NLS가 하나 초과의 복제물에 존재할 수 있고, 하나 초과 NLS 서열이 존재하는 하나 초과의 다른 NLS와 조합하여 존재할 수 있도록 각각의 NLS 서열은 다른 것과 독립적으로 선택될 수 있다.
- [134] 일부 구현예에서, NLS 서열은 단백질에 대해 이중성으로 하기의 NLS 서열이 예시되나 이에 제한되는 것은 아니다. 예컨대, 상기 NLS는 아미노산 서열 'PKKKRKV'를 갖는 SV40 바이러스 대형 T-항원의 NLS, 뉴클레오플라스민(nucleoplasmin)으로부터의 NLS 서열로서 'KRPAATKKAGQAKKKK'를 갖는 뉴클레오플라스민 이분(bipartite) NLS, 아미노산 서열 'PAAKRVKLD' 또는 'RQRRNELKRSP'를 갖는 c-myc NLS일 수 있다. 또한 hRNPA1 M9 NLS 서열, 임포틴-알파로부터의 IBB 도메인의 NLS 서열, 마이오마(myoma) T 단백질의 NLS 서열 및 인간 p53의 NLS 서열, 마우스 c-abl IV의 NLS 서열, 인플루엔자 바이러스 NS1의 NLS 서열, 간염 바이러스 델타 항원의 NLS 서열, 마우스 Mx1 단백질의 NLS 서열, 인간 폴리(ADP-리보스) 중합효소의 NLS 서열 또는 스테로이드 호르몬 수용체(인간) 글루코코르티코이드의 NLS 서열로부터 유래된 NLS 서열일 수 있다.
- [135] 또한, Cas12f1, TnpB 또는 이의 변이체 단백질은 세포 내의 유전자 발현 과정에 관여할 수 있는 다양한 효소(enzyme)가 융합된 것일 수 있다. 이때, 상기 효소가 융합된 Cas12f1, TnpB 또는 이의 변이체 단백질은 세포 내 유전자 발현에 다양한 양적 및/또는 질적 변화를 초래할 수 있다. 예컨대, 상기 추가적으로 결합되는 다양한 효소는 DNMT, TET, KRAB, DHAC, LSD, p300, Moloney Murine Leukemia Virus(M-MLV) 역전사 효소 또는 그 변이체일 수 있다. 이 때, 상기 역전사 효소가

융합된 Cas12f1, TnpB 또는 이의 변이체 단백질은 프라임 에디터(prime editor)로도 기능할 수 있다.

- [136] 일부 구현예에서, Cas12f1, TnpB 또는 이의 변이체 단백질은 표적 핵산 또는 표적 유전자의 이중가닥 절단(double-strand breaks)을 일으킬 수 있으며, 상기 이중가닥 절단은 Cas12f1, TnpB 또는 이의 변이체 단백질에 의해 표적 핵산의 바깥 부분이 절단되어 발생하는 것일 수 있다. 상기와 같이 Cas12f1, TnpB 또는 이의 변이체 단백질에 의해 표적 핵산 또는 표적 유전자에 이중가닥 절단이 발생하면 세포 내 기작에 의해 공여자 핵산 분자를 주형으로 하여 이중가닥 절단의 복구 과정이 진행될 수 있다. 이러한 이중가닥 절단의 복구에 의해 목적하는 서열이 표적 핵산 또는 표적 유전자에 도입된다.
- [137] **2. Cas12f1, TnpB 및 이의 변이체 단백질의 PAM 서열**
- [138] 일부 구현예에서, 초소형 핵산 편집 시스템이 표적 핵산 또는 표적 유전자의 표적 부위에 위치하고 정확하게 표적 부위 핵산을 절단하기 위해서는 하기의 두 가지 조건이 필요하다.
- [139] 먼저, 표적 핵산 또는 표적 유전자 내에 Cas12f1, TnpB 또는 이의 변이체 단백질이 인식할 수 있는 일정 길이의 염기서열이 있어야 한다. 또한, 상기 일정 길이의 염기서열 주변에 Cas12f1, TnpB 또는 이의 변이체 단백질에 대한 가이드 RNA(gRNA)에 포함된 가이드 서열(스페이서)과 상보적으로 결합할 수 있는 서열이 있어야 한다. 다시 말해, Cas12f1, TnpB 또는 이의 변이체 단백질이 상기 일정 길이의 염기서열을 인식하고, 가이드 RNA(gRNA)에 포함된 스페이서 서열 부분이 상기 일정 길이의 염기서열 주변 서열 부분과 상보적으로 결합할 때, 표적 핵산 또는 표적 유전자의 표적 부위 핵산을 정확하게 절단, 편집 및/또는 복구할 수 있다. 이때, Cas12f1, TnpB 또는 이의 변이체 단백질에 의해 인식되는 일정 길이의 염기 서열을 프로토스페이스 인접 모티프(Protospacer Adjacent Motif, PAM) 서열이라 한다. PAM 서열은 초소형 유전자 편집 단백질인 Cas12f1, TnpB 또는 이의 변이체 단백질에 따라 정해지는 고유한 서열이다. 이는 핵산 편집 시스템 내의 Cas12f1, TnpB 또는 이의 변이체 단백질과 gRNA 복합체의 표적 서열을 결정할 때, 상기 PAM 서열과 인접한 서열 내에서 표적 서열을 결정해야 하는 것을 의미한다.
- [140] Cas12f1, TnpB 또는 이의 변이체 단백질의 PAM 서열은 T-rich 서열일 수 있다. 보다 구체적으로, 상기 Cas12f1, TnpB 또는 이의 변이체 단백질의 PAM 서열은 5'-TTTN-3'일 수 있다. 이때, N은 디옥시티미딘(T), 디옥시아데노신(A), 디옥시사이티딘(C) 또는 디옥시구아노신(G) 중 하나이다.
- [141] 일부 구현예에서, Cas12f1, TnpB 또는 이의 변이체 단백질의 PAM 서열은 5'-TTTA-3', 5'-TTTT-3', 5'-TTTC-3' 또는 5'-TTTG-3'일 수 있다. 바람직하게, Cas12f1, TnpB 또는 이의 변이체 단백질의 PAM 서열은 5'-TTTA-3' 또는 5'-TTTG-3'일 수 있다.

- [142] 다른 구현예에서, Cas12f1, TnpB 또는 이의 변이체 단백질의 PAM 서열은 야생형 Cas12f1 단백질의 PAM 서열과는 다른 것일 수 있다.
- [143] **3. 상동지정복구용 Cas12f1, TnpB 또는 이의 변이체 단백질을 위한 가이드 RNA**
- [144] **(1) 개괄**
- [145] 본 발명의 구현예들은 종래 기술인 Cas9 시스템이 큰 단백질 분자량을 가짐으로 인해 발생하는 아데노 연관 바이러스(AAV)를 이용한 세포 내 전달의 한계점을 극복하기 위해 도출된 것이다. 따라서 본 발명의 표적 핵산 편집 시스템에 포함되는 상동지정복구를 위한 절단 활성을 갖는 단백질로서 분자량이 작은 Cas12f1, TnpB 또는 이의 변이체 단백질을 선택하는 것에 추가하여, 상기 Cas12f1, TnpB 또는 이의 변이체에 대한 가이드 RNA(gRNA)를 자연에 존재하는 것보다 훨씬 더 짧게 인위적으로 엔지니어링하여 크기의 최소화를 달성함과 동시에 표적에 대한 절단 및/또는 상동지정복구 효율은 증가된 엔지니어링된 가이드 RNA(augment RNA)를 제작하였다.
- [146] 본 발명의 일 구현예에 따른 초소형 엔도뉴클레아제인 TnpB 또는 이의 변이체에 대하여는 자연에 존재하는 gRNA가 발견되지 않았으므로, 상기 TnpB 또는 이의 변이체 단백질에 대한 고효율의 표적 및 편집 활성을 나타내는 최적의 gRNA를 제작하고자 하였다. 이러한 관점에서, TnpB 또는 이의 변이체 단백질에 대한 자연에 존재하는 gRNA는 TnpB 또는 이의 변이체 단백질과 크기가 유사한 Cas12f1에 대해 자연계에서 발견되는 야생형 gRNA일 수 있다. 즉, 본 발명에서 Cas12f1, TnpB 또는 이의 변이체 단백질에 대한 "야생형" gRNA는 "기본형" 또는 "원형(canonical)" gRNA의 의미로 사용되었다. 상기 야생형 gRNA는 (i) 하나 이상의 스템(stem) 영역, (ii) tracrRNA-crRNA 상보성 영역 및 임의적으로 (iii) 연속되는 3개 이상의 유라실(U)을 포함하는 영역을 포함하는, tracrRNA(trans-activating CRISPR RNA) 및 crRNA(CRISPR RNA)를 포함할 수 있다. 구체적으로, 야생형 gRNA는 5'-말단부터 순차적으로 제1 스템 영역, 제2 스템 영역, 제3 스템 영역, 제4 스템 영역 및 tracrRNA-crRNA 상보성 영역을 포함하는 tracrRNA 및 crRNA를 포함할 수 있다. 보다 구체적으로, 상기 야생형 gRNA는 서열번호 11의 염기서열을 갖는 야생형 tracrRNA를 포함하거나, 서열번호 12의 염기서열을 갖는 야생형 crRNA를 포함할 수 있다. 또한, 상기 야생형 gRNA는 싱글 가이드 RNA 형태로 융합되어 서열번호 13의 염기서열을 갖는 sgRNA일 수 있다.
- [147] 일 구현예로, Cas12f1, TnpB 또는 이의 변이체 단백질에 대한 gRNA는 자연계에서 발견되는 야생형 gRNA에 새로운 구성을 추가하거나, 기존의 구조를 제거 및/또는 치환하거나, 그 구조의 일부를 변형한 엔지니어링된 gRNA인 것을 특징으로 한다.
- [148] 일부 구체예에서, 엔지니어링된 gRNA는 야생형 gRNA 서열에서 1개 이상의 뉴클레오티드가 치환, 결실, 삽입 또는 부가된 서열을 포함하고, 가이드 서열을 제외한 부분이 상기 야생형 Cas12f1 gRNA와 적어도 50%, 50%, 55%, 60%, 65%,

70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 95% 서열 동일성을 갖는 엔지니어링된 gRNA이다. RNA, 핵산 또는 폴리펩티드의 문맥에서 용어 "서열 동일성"은 비교 범위에서 최적으로 정렬된 2개의 서열을 비교하여 결정된 값을 의미하며, 이때 비교 범위 내의 RNA, 핵산 등의 서열 부분은 최적의 정렬을 위해 기준 서열과 비교하여 삽입 또는 결실(즉, 갭)을 포함할 수 있다.

[149] 이하, 야생형과 엔지니어링된 gRNA의 구조 및 그의 변형에 대해 5개의 변형부위 별로 상세히 설명한다. 변형부위는 본 명세서 전체에 걸쳐 "MS(modification site)"로 약칭되었으며, "변형부위" 또는 "MS" 뒤의 숫자는 일 실시예에 따른 각 변형부위의 엔지니어링 흐름에 따라 순차적으로 부여한 것이나, 뒤의 숫자를 가지는 변형부위에서의 엔지니어링이 앞선 숫자의 변형부위에서의 엔지니어링을 반드시 포함한다는 의미는 아니다. 도 1은 본 발명의 구현예에 따른 엔지니어링된 가이드 RNA(engineered gRNA)가 포함하는 변형부위인 MS1 내지 MS5를 야생형 가이드 RNA 서열 상에 도시한 것이다.

[150] **(2) gRNA의 구조 및 정의**

[151] 본 발명의 가이드 RNA(gRNA)는 crRNA를 포함한다. crRNA는 tracrRNA 및/또는 이펙터 단백질과 결합 및/또는 상호작용하는 crRNA 내에 존재하는 일부 서열이다. 상기 crRNA는 야생형 crRNA 또는 엔지니어링된 crRNA 일 수 있다. 이때, 상기 crRNA는 직접반복부서열(direct repeat sequence) 및 가이드 서열(스페이서 서열)를 포함할 수 있고, 직접반복부서열은 가이드 서열의 5'말단에 위치할 수 있다. 또한, 상기 crRNA는 tracrRNA의 3' 말단에 위치할 수 있다.

[152] 또한, 가이드 RNA는 tracrRNA를 포함한다. 상기 tracrRNA 스캐폴드 서열은 crRNA 및/또는 이펙터 단백질과 결합 및/또는 상호작용하는 tracrRNA 전체 또는 일부 서열이다.

[153] 상기 tracrRNA는 야생형 tracrRNA 또는 엔지니어링된 tracrRNA일 수 있다. 상기 엔지니어링된 crRNA 또는 tracrRNA는 상기 야생형 crRNA 또는 tracrRNA의 일부 (뉴클레오타이드) 서열이 인위적으로 변형(치환, 결실 또는 삽입)되거나, 야생형 crRNA 또는 tracrRNA 서열보다 길이가 짧도록 변형된 서열일 수 있다.

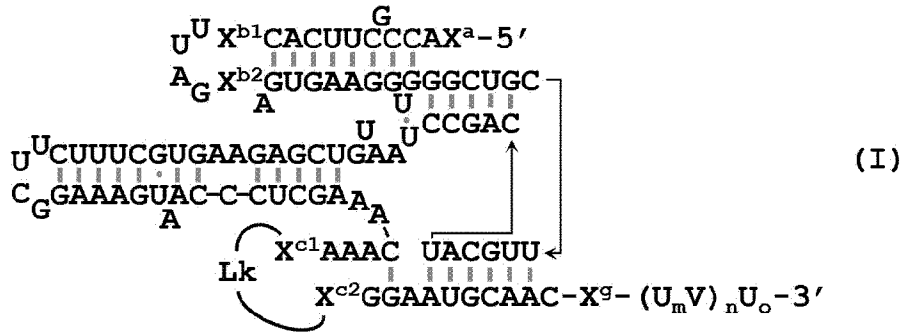
[154] **(2-1) 스캐폴드 서열(scaffold sequence)**

[155] 본 발명의 구현예에 따른 야생형 또는 엔지니어링된 가이드 RNA(gRNA)의 서열을 기능적으로 나누면, Cas12f1, TnpB 또는 이의 변이체 단백질과 상호작용하여 gRNA 및 상기 단백질이 복합체를 형성하도록 하는 서열 부분과 gRNA 및 상기 단백질 복합체가 표적 핵산을 찾아갈 수 있도록 하는 서열 부분으로 나눌 수 있다. 이때, 상기 Cas12f1, TnpB 또는 이의 변이체 단백질과 상호작용하여 gRNA 및 상기 단백질(엔도뉴클레아제) 복합체를 형성하도록 하는 서열 부분을 스캐폴드 서열이라 할 수 있다. 구체적으로, 상기 스캐폴드 서열은 tracrRNA 및 crRNA 인 두 분자 이상의 RNA의 서열을 포함할 수 있다.

- [156] 일 구현예에서, 엔지니어링된 gRNA가 듀얼 가이드 RNA인 경우, 스캐폴드 서열은 엔지니어링된 gRNA 서열 중 tracrRNA 서열 및 crRNA에 포함된 CRISPR RNA 반복 서열을 포함할 수 있다. 일 예로, tracrRNA 서열은 자연계에서 발견되는 tracrRNA 서열의 전부 또는 일부가 변형된 것일 수 있다. 또한, 상기 CRISPR RNA 반복 서열은 자연계에서 발견되는 CRISPR RNA 반복 서열의 전부 또는 일부가 변형된 것일 수 있다.
- [157] 다른 구현예에서, 엔지니어링된 가이드 RNA가 싱글 가이드 RNA(sgRNA)인 경우, 스캐폴드 서열은 엔지니어링된 tracrRNA 서열, 링커 서열 및 엔지니어링된 crRNA 서열에 포함된 CRISPR RNA 반복 서열을 포함할 수 있다. 일 구현예로, 상기 tracrRNA 서열은 자연계에서 발견되는 tracrRNA 서열의 전부 또는 일부가 변형된 것일 수 있다.
- [158] 또한, 일 구현예에서, 스캐폴드 서열은 tracrRNA 및 crRNA의 일부를 포함하며, 반드시 한 분자의 RNA를 지칭하는 것은 아니다. 상기 스캐폴드 서열은 다시 제 1 스템 영역, 제2 스템 영역, 제3 스템 영역, 제4 스템 영역 및 tracrRNA-crRNA 상보성 영역(제5 스템 영역으로 지칭될 수 있음)으로 세분화되는 영역을 포함할 수 있다. 본 발명에서, 상기 세분화된 영역 중 변형부위 3(MS3)을 포함하는 제1 스템 영역, 변형부위 5(MS5)를 포함하는 제2 스템 영역 및 변형부위 1(MS1)과 변형부위 4(MS4)를 포함하는 tracrRNA-crRNA 상보성 영역은 도 1에서 각기 다른 색의 음영으로 구분된 1점쇄선 박스로 표시된 영역에 대응되거나 이에 포함되는 영역으로 정의될 수 있다. 그 외, 제3 스템 영역은 도 1에서 G(-90)-C(-74) 서열에 대응되거나 이에 포함되는 영역이고, 제4 스템 영역은 도 1에서 U(-68)-A(-35) 서열에 대응되거나 이에 포함되는 영역으로 정의될 수 있다.
- [159] 한편, 본 발명에서 상기 스템 영역, tracrRNA-crRNA 상보성 영역 등으로 세분화된 영역은 스캐폴드 서열의 모든 영역을 포괄하는 것은 아니며, 스캐폴드 서열은 상기 세분화된 영역에 해당하지 않는 다른 영역 또는 서열을 포함할 수 있다.
- [160] 다른 구현예에서, 야생형의 Cas12f1 gRNA는 상술한 바와 같은 세분화된 영역을 가질 수 있는 스캐폴드 서열을 포함하며, (i) 하나 이상의 스템(stem) 영역, (ii) tracrRNA-crRNA 상보성 영역 및 임의적으로 (iii) 상기 tracrRNA-crRNA 상보성 영역 내에 연속되는 3개 이상의 유라실(U)을 포함하는 영역을 포함하는 스캐폴드 서열을 포함하는 특징이 있다. 구체적으로, 야생형 Cas12f1 gRNA는 하나 이상의 스템 영역을 포함하는 tracrRNA, 및 tracrRNA-crRNA 상보성 영역(또는 tracrRNA와 crRNA가 상보적으로 결합하여 형성된 또 다른 스템 영역)을 포함하는 tracrRNA 및/또는 crRNA를 포함할 수 있다. 야생형 type V-F CRISPR/Cas gRNA의 구조에 대한 상세한 정보는 문헌[Takeda et al., Structure of the miniature type V-F CRISPR-Cas effector enzyme, Molecular Cell 81, 1-13(2021)]를 참조한다. 일 구현예에서 상기 야생형 Cas12f1 가이드 RNA는 서열번호 11의 핵산 서열로 이루어진 tracrRNA 및 서열번호 12의 핵산 서열로 이루어진 crRNA를 포함할 수 있다.

- [161] 본 발명에서 상동지정복구를 위한 표적 핵산의 편집 시스템에 사용될 수 있는 엔지니어링된 gRNA는 (a) 하나 이상의 스템 영역의 일부 또는 전부의 결실, (b) tracrRNA-crRNA 상보성 영역의 일부 또는 전부의 결실 또는 (c) 연속되는 3 이상의 유라실(U)이 존재하는 경우에 이들 중 하나 이상의 U의 치환에 의한 변형을 포함할 수 있다. 상기 각각의 결실 및 치환에 의한 변형의 상세한 내용은 후술한다.
- [162] 한편, 본 발명의 일 구현예에 따른 엔지니어링된 crRNA는 스캐폴드 서열로서 (d) crRNA의 3'-말단에 하나 이상의 유리딘을 갖는 U-rich tail 영역을 추가로 포함할 수 있다. 이는 본 발명의 가이드 RNA 및 Cas12f1, TnpB 또는 이의 변이체 단백질 복합체를 포함하는 표적 핵산 편집 시스템의 상동지정복구 효율 향상을 위해 도입할 수 있는 엔지니어링된 스캐폴드 영역에 추가된 것이다. U-rich tail 영역에 대한 구체적인 내용은 후술한다.
- [163] 일 구현예에서, 엔지니어링된 스캐폴드 서열은, 자연계에서 발견되는 스캐폴드 서열에 전술한 (i) 하나 이상의 스템(stem) 영역, (ii) tracrRNA-crRNA 상보성 영역 및 (iii) 상기 tracrRNA-crRNA 상보성 영역 내에 연속되는 3개 이상의 유라실(U)을 포함하는 영역 중 어느 하나 이상의 영역에서의 변형이 조합된 것일 수 있다. 이 때, 엔지니어링된 tracrRNA는 야생형 tracrRNA보다 길이가 짧도록 변형된 tracrRNA일 수 있다. 또한, 엔지니어링된 tracrRNA는 연속된 네 개 또는 다섯 개 이상의 유리딘 서열을 포함하지 않도록 변형(MS1에서의 변형)된 tracrRNA일 수 있다. 또한, 엔지니어링된 tracrRNA는 연속된 다섯 개 이상의 유리딘 서열을 포함하지 않도록 변형되고, 야생형 tracrRNA보다 길이가 짧도록 변형된 tracrRNA일 수 있다. 또한, 엔지니어링된 tracrRNA는 제1 스템 영역, 제2 스템 영역, 제3 스템 영역, 제4 스템 영역 및 tracrRNA-crRNA 상보성 영역의 일부를 5'-말단에서 3'-말단 방향으로 순서대로 포함할 수 있다. 또한, 엔지니어링된 crRNA는 tracrRNA-crRNA 상보성 영역의 일부 및 가이드 서열인 스페이서 서열을 5'-말단에서 3'-말단 방향으로 순서대로 포함할 수 있다. 상기 tracrRNA의 tracrRNA-crRNA 상보성 영역은 crRNA의 직접반복서열(즉, crRNA의 tracrRNA-crRNA 상보성 영역)과 결합하기에 충분한 상보성을 갖는 임의의 폴리뉴클레오티드를 포함할 수 있다.
- [164] 다른 구현예에서, 엔지니어링된 gRNA는 하기 식 (I)로 표시되는 서열로 이루어지거나 상기 서열과 80% 이상, 85% 이상, 90% 이상 또는 95% 이상의 서열 동일성을 갖는 gRNA일 수 있다[하기 식 (I) 및 본 명세서 전체에 기재된 식 (I)에서, 검정색 실선은 뉴클레오티드 사이의 화학적 결합(예를 들어, 포스포다이에스터 결합)을 의미하고, 회색 굵은선은 뉴클레오티드 사이의 상보적 결합을 의미한다].

[165]



[166] 상기 식 (I)에서, X^a , X^{b1} , X^{b2} , X^{c1} 및 X^{c2} 는 각각 독립적으로 0 내지 35개의 (폴리)뉴클레오티드로 이루어지고, X^g 는 10 내지 30개의 폴리뉴클레오티드로 이루어진 가이드 서열로서 표적 서열과 혼성화하거나 표적 서열에 상보적인 서열이고, Lk는 길이 2 내지 20의 폴리뉴클레오티드 링커이거나 부존재하고, $(U_m V)_n U_o$ 는 U-rich tail로서 존재하거나 부존재하고, 존재하는 경우 U는 유리딘이고, V는 각각 독립적으로 A, C 또는 G이고, m 및 o는 각각 독립적으로 1 내지 20 사이의 정수이며, n은 0 내지 5 사이의 정수이다.

[167] 여기서, 상기 X^a , X^{b1} , X^{b2} , X^{c1} 또는 X^{c2} 가 0개의 뉴클레오티드로 이루어지는 경우는 X^a , X^{b1} , X^{b2} , X^{c1} 또는 X^{c2} 가 부존재한다는 의미로 해석된다.

[168] 또한, 식 (I)에서 상기 X^a , X^{b1} , X^{b2} , X^{c1} 또는 X^{c2} 가 0개의 뉴클레오티드로 이루어지거나 부존재하는 경우에는 X^a , X^{b1} , X^{b2} , X^{c1} 또는 X^{c2} 를 통해 연결된 2 이상의 뉴클레오티드가 존재하였을 경우 이들이 어떠한 방식으로든 직접 연결된 상태인 것으로 해석될 수 있다. 예를 들어, 식 (I)에서 X^{b1} 이 0개의 뉴클레오티드로 이루어지거나 부존재하는 경우 X^{b1} 의 5'-말단에 직접 연결된 뉴클레오티드와 X^{b1} 의 3'-말단에 직접 연결된 뉴클레오티드가 예를 들어, 포스포다이에스터 결합으로 직접 연결된 상태일 수 있다.

[169] 일부 구현예에서, 상기 X^a 는 0 내지 20개의 (폴리)뉴클레오티드로 이루어질 수 있고, X^{b1} 은 0 내지 13개의 (폴리)뉴클레오티드로 이루어질 수 있고, X^{b2} 는 0 내지 14개의 (폴리)뉴클레오티드로 이루어질 수 있고, X^{c1} 은 0 내지 28개의 (폴리)뉴클레오티드로 이루어질 수 있고, 또는 X^{c2} 는 0 내지 27개의 (폴리)뉴클레오티드로 이루어질 수 있다.

[170] 상기 스캐폴드 서열의 (i)의 변형은 X^a , X^{b1} , X^{b2} 로 표시된 폴리뉴클레오티드에 해당할 수 있고, (ii)의 변형은 X^{c1} 및 X^{c2} 로 표시된 폴리뉴클레오티드에 해당할 수 있고, (iii)의 변형은 X^{c1} 으로 표시된 폴리뉴클레오티드 내에 존재할 수 있다.

[171] 상기 스캐폴드 서열의 (i) 내지 (iii)의 변형에 대한 구체적인 내용은 하기 "(4) 고효율의 상동지정복구 효율 달성을 위한 gRNA의 변형" 항목을 참조한다.

[172] (2-2) 가이드 서열(guide sequence)

- [173] 본 발명의 구현예에 따른 야생형 또는 엔지니어링된 가이드 RNA(gRNA)는 표적 핵산을 찾아갈 수 있도록 하는 서열 부분, 즉 유전자 내의 표적 서열과 혼성화하거나 상보적 결합을 이루는 하나 이상의 가이드 서열을 포함할 수 있다.
- [174] 본원에서 "가이드 서열" 또는 "스페이스(spacer) 서열"로 지칭되는 서열은 표적 핵산 또는 표적 유전자 내의 표적 서열과 상보적인 서열이며, crRNA 반복 서열의 3'-말단 쪽에 연결된다. 가이드 서열은 Cas12f1, TnpB 또는 이의 변이체 단백질이 인식하는 PAM(Protospacer Adjacent Motif) 서열과 인접한 프로토스페이스 서열(protospacer sequence)과 상동성이 있는 서열로, 프로토스페이스 서열의 티미딘(T)이 유리딘(U)으로 치환된 서열을 가진다. 이때, 표적 서열 및 프로토스페이스 서열은 표적 핵산 내 포함된 상기 PAM 서열과 인접한 서열 내에서 결정되고, 이에 따라 가이드 서열이 결정된다.
- [175] 일 구현예로, crRNA의 가이드 서열 부분은 상기 표적 핵산과 상보적으로 결합할 수 있다. 일 구현예로, crRNA의 가이드 서열 부분은 상기 표적 핵산의 표적 서열 부분과 상보적으로 결합할 수 있다. 일 예로, 표적 핵산이 이중가닥 DNA인 경우, 가이드 서열은 이중가닥 DNA의 표적 가닥(Target strand)에 포함된 표적 서열과 상보적인 서열일 수 있다. 여기서, 표적 핵산이 이중가닥 DNA인 경우, 가이드 서열은 상기 이중가닥 DNA의 비-표적가닥(Non-target strand)에 포함된 프로토스페이스 서열과 상동성인 서열을 포함할 수 있다. 구체적으로, 가이드 서열은 프로토스페이스 서열과 동일한 염기 서열을 가지되, 상기 염기 서열에 포함된 티미딘(T) 각각이 모두 유리딘(U)으로 치환된 서열을 가질 수 있다. 일 예로, 가이드 서열은 프로토스페이스의 DNA 서열에 상응하는 RNA 서열을 포함할 수 있다. 상기 가이드 서열은 상류(upstream) 중 선택된 하나의 표적 서열 및 하류(downstream) 중 선택된 하나의 표적 서열과 혼성화 가능한 두 개의 가이드 서열의 조합을 포함할 수 있다.
- [176] 일 구현예로, 가이드 서열의 길이는 10 뉴클레오티드 내지 50 뉴클레오티드 길이일 수 있다. 바람직하게, 가이드 서열의 길이는 10 뉴클레오티드 내지 30 뉴클레오티드 길이일 수 있다. 보다 바람직하게, 스페이스 서열의 길이는 17 뉴클레오티드 내지 25 뉴클레오티드 길이일 수 있다.
- [177] 상기 "표적 서열(target sequence)"은 표적 핵산 또는 표적 유전자 내에 존재하는 서열로, 본원의 표적 핵산 편집 시스템 또는 TaRGET 시스템의 가이드 RNA에 의해 인식되는 서열 또는 표적 핵산 편집 시스템 또는 TaRGET 시스템에 의해 변형의 대상이 되는 서열을 의미한다. 구체적으로, 상기 표적 서열은 가이드 RNA에 포함된 가이드 서열에 상보성을 가지는 서열 또는 가이드 서열과 상보적으로 결합하는 서열을 의미한다. "표적 가닥(target strand)"은 표적 서열을 포함하는 가닥을 의미한다. 표적 핵산 또는 표적 유전자가 단일가닥인 경우, 해당 가닥은 표적 가닥일 수 있다. 또는, 표적 핵산 또는 표적 유전자가 이중가닥인 경우, 상기 이중가닥 중 하나는 표적 가닥일 수 있으며, 상기 표적 가닥에 상보적인 가닥이 존재할 수 있다. 이때, 상기 표적 가닥에 상보적인 가닥은 "비표적 가닥(non-target

strand)"으로 지칭된다. 비표적 가닥(non-target strand)은 PAM(Protospacer Adjacent Motif) 서열 및 프로토스페이스어(protospacer) 서열을 포함한다. 상기 PAM 서열은 본 발명의 표적 핵산 편집 시스템(또는 TaRGET 시스템)의 Cas12f1, TnpB 또는 이의 변이체 단백질이 인식하는 서열이다. 상기 프로토스페이스어 서열은 PAM 서열의 5'-말단 또는 3'-말단에 위치하는 서열로, 상기 프로토스페이스어 서열은 표적 서열에 상보성을 가지는 서열 또는 표적 서열과 상보적인 결합을 하는 서열이다. 프로토스페이스어 서열과 표적 서열 간의 상관관계는 표적 서열과 가이드 서열 간의 상관관계와 유사하다. 이러한 특징에 의해, 일반적으로 가이드 서열 설계시 프로토스페이스어 서열을 이용하여 설계할 수 있다. 즉, 표적 서열에 상보적으로 결합하는 가이드 서열을 설계시, 가이드 서열은 프로토스페이스어 서열과 동일한 염기서열을 가지는 뉴클레오티드 서열로 설계할 수 있다. 이때, 프로토스페이스어 서열의 염기서열 중 T는 U로 대체하여 가이드 서열을 설계한다.

[178] 상기 표적 서열은 15 내지 40개의 뉴클레오티드 서열일 수 있다. 일 예로, 상기 표적 서열은 15 내지 20개, 15 내지 25개, 15 내지 30개, 15 내지 35개 또는 15 내지 40개의 뉴클레오티드 서열일 수 있다. 또는 상기 표적 서열은 20 내지 25개, 20 내지 30개, 20 내지 35개 또는 20 내지 40개의 뉴클레오티드 서열일 수 있다. 또는 상기 표적 서열은 25 내지 30개, 25 내지 35개 또는 25 내지 40개의 뉴클레오티드 서열일 수 있다. 또는 상기 표적 서열은 30 내지 35개 또는 30 내지 40개의 뉴클레오티드 서열일 수 있다. 또는 상기 표적 서열은 35 내지 40개의 뉴클레오티드 서열일 수 있다. 다른 일 예로, 상기 표적 서열은 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 또는 40개의 뉴클레오티드 서열일 수 있다.

[179] 일부 구현예에서, 상기 가이드 서열은 표적 서열과 상보적인 결합을 하는 서열일 수 있다. 이때, 상기 상보적인 결합은 선택적으로 적어도 하나 이상의 미스매치(mismatch) 결합을 포함할 수 있다. 예를 들어, 상기 가이드 서열은 표적 서열과 혼성화하거나 상보적인 결합을 하는 서열로, 이때 상기 상보적인 결합은 0 내지 5개의 미스매치를 포함할 수 있다. 또는 상기 가이드 서열은 표적 서열에 대해 적어도 70% 이상 상보적인 뉴클레오티드 서열일 수 있다. 이때, 표적 서열이 DNA인 경우에 표적 서열 내에 존재하는 아데노신(A)에 대해, 상기 가이드 서열은 상기 아데노신(A)에 상보적인 결합을 형성할 수 있는 유리딘(U)을 포함할 수 있다.

[180] 일 구현예로서, 상기 가이드 서열은 표적 서열에 대해 적어도 70% 내지 75%, 적어도 70% 내지 80%, 적어도 70% 내지 85%, 적어도 70% 내지 90%, 적어도 70% 내지 95%, 적어도 70% 내지 100%, 적어도 75% 내지 80%, 적어도 75% 내지 85%, 적어도 75% 내지 90%, 적어도 75% 내지 95% 또는 적어도 75% 내지 100% 상보적인 서열일 수 있다. 또는 상기 가이드 서열은 표적 서열에 대해 적어도 80% 내지 85%, 적어도 80% 내지 90%, 적어도 80% 내지 95%, 적어도 80% 내지 100%, 적어도 85% 내지 90%, 적어도 85% 내지 95% 또는 적어도 85% 내지 100% 상보적인 서열일 수 있다. 또는 상기 가이드 서열은 표적 서열에 대해 적어도 90% 내지

95%, 적어도 90% 내지 100% 또는 적어도 95% 내지 100% 상보적인 서열일 수 있다. 또는 상기 가이드 서열은 표적 서열에 대해 적어도 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 또는 100% 상보적인 서열일 수 있다.

[181] 상기 가이드 서열은 프로토스페이서 서열과 동일한 또는 유사한 서열일 수 있다. 또는 상기 가이드 서열은 프로토스페이서 서열에 대해 서열 동일성 또는 서열 유사성을 가지는 서열일 수 있다. 이때, 상기 서열 동일성 또는 서열 유사성은 적어도 70% 이상인 것일 수 있다. 이때, 프로토스페이서 서열 내에 존재하는 티미딘(T)에 대해, 상기 가이드 서열은 티미딘(T) 대신에 유리딘(U)을 포함할 수 있다.

[182] 일 구현예로서, 상기 가이드 서열은 프로토스페이서 서열과 적어도 70% 내지 75%, 적어도 70% 내지 80%, 적어도 70% 내지 85%, 적어도 70% 내지 90%, 적어도 70% 내지 95%, 적어도 70% 내지 100%, 적어도 75% 내지 80%, 적어도 75% 내지 85%, 적어도 75% 내지 90%, 적어도 75% 내지 95% 또는 적어도 75% 내지 100% 동일한 또는 유사한 서열일 수 있다. 또는 상기 가이드 서열은 프로토스페이서 서열과 적어도 80% 내지 85%, 적어도 80% 내지 90%, 적어도 80% 내지 95%, 적어도 80% 내지 100%, 적어도 85% 내지 90%, 적어도 85% 내지 95% 또는 적어도 85% 내지 100% 동일한 또는 유사한 서열일 수 있다. 또는 상기 가이드 서열은 프로토스페이서 서열과 적어도 90% 내지 95%, 적어도 90% 내지 100% 또는 적어도 95% 내지 100% 동일한 또는 유사한 서열일 수 있다. 또는 상기 가이드 서열은 프로토스페이서 서열과 적어도 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 또는 100% 동일한 또는 유사한 서열일 수 있다.

[183] 다른 일 구현예로서, 상기 가이드 서열은 프로토스페이서 서열에 적어도 70% 내지 75%, 적어도 70% 내지 80%, 적어도 70% 내지 85%, 적어도 70% 내지 90%, 적어도 70% 내지 95%, 적어도 70% 내지 100%, 적어도 75% 내지 80%, 적어도 75% 내지 85%, 적어도 75% 내지 90%, 적어도 75% 내지 95% 또는 적어도 75% 내지 100%의 서열 동일성 또는 서열 유사성을 가지는 서열일 수 있다. 또는 상기 가이드 서열은 프로토스페이서 서열에 적어도 80% 내지 85%, 적어도 80% 내지 90%, 적어도 80% 내지 95%, 적어도 80% 내지 100%, 적어도 85% 내지 90%, 적어도 85% 내지 95% 또는 적어도 85% 내지 100%의 서열 동일성 또는 서열 유사성을 가지는 서열일 수 있다. 또는 상기 가이드 서열은 프로토스페이서 서열에 적어도 90% 내지 95%, 적어도 90% 내지 100% 또는 적어도 95% 내지 100%의 서열 동일성 또는 서열 유사성을 가지는 서열일 수 있다. 또는 상기 가이드 서열은 프로토스페이서 서열에 적어도 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 또는 100%의 서열 동일성 또는 서열 유사성을 가지는 서열일 수 있다.

[184] (3) 싱글 가이드 RNA 또는 듀얼 가이드 RNA

- [185] 본 발명의 구현예에 따른 엔지니어링된 가이드 RNA는 싱글 가이드 RNA 또는 듀얼 가이드 RNA일 수 있다. 듀얼 가이드 RNA는 가이드 RNA가 tracrRNA 및 crRNA의 두 분자 RNA로 구성된 것을 의미한다. 싱글 가이드 RNA(sgRNA)는 엔지니어링된 tracrRNA의 3'-말단 및 엔지니어링된 crRNA의 5'-말단이 링커를 통해 연결된 것을 의미한다.
- [186] 일 구현예에서, 엔지니어링된 싱글 가이드 RNA(sgRNA)는 링커 서열을 추가적으로 더 포함할 수 있고, tracrRNA 서열 및 crRNA 서열이 링커 서열을 통해 연결될 수 있다. 바람직하게, 엔지니어링된 스캐폴드 서열에 포함된 tracrRNA의 tracrRNA-crRNA 상보성 서열의 3'-말단 및 crRNA의 tracrRNA-crRNA 상보성 서열의 5'-말단이 링커를 통해 연결된 것을 포함할 수 있다. 보다 바람직하게, tracrRNA와 crRNA의 tracrRNA-crRNA 상보성 영역은 각각의 3'-말단 및 5'-말단이 링커 5'-GAAA-3'로 연결될 수 있다. 상기 링커에 대한 구체적인 내용은 식 (I)의 Lk에 대한 내용을 참조한다.
- [187] 일 구현예에서, 싱글 가이드 RNA의 서열은 5'-말단에서 3'-말단 방향으로, tracrRNA 서열, 링커 서열, crRNA 서열 및 U-rich tail 서열이 순차적으로 연결되어 있다. tracrRNA 서열의 일부 및 crRNA 서열에 포함된 CRISPR RNA 반복 서열의 전부 및 일부는 서로 상보적인 서열을 가진다.
- [188] 또한, 본 발명의 구현예에 따른 엔지니어링된 가이드 RNA는 tracrRNA 및 crRNA가 별개의 RNA 분자를 이루고 있는 듀얼 가이드 RNA일 수 있다. 이 때, tracrRNA의 일부 및 crRNA의 일부는 서로 상보적인 서열을 가져 이중가닥 RNA를 형성할 수 있다. 보다 구체적으로, 듀얼 가이드 RNA에서 tracrRNA의 3'-말단을 포함하는 일부 및 crRNA의 CRISPR RNA 반복 서열을 포함하는 일부가 이중가닥을 형성할 수 있다. 엔지니어링된 가이드 RNA는 Cas12f1, TnpB 또는 이의 변이체 단백질과 결합하여 가이드 RNA 및 상기 단백질의 복합체를 형성할 수 있으며, 상기 crRNA 서열에 포함된 가이드 서열과 상보적인 표적 서열을 인식하여 상기 표적 서열을 포함하는 표적 핵산을 편집할 수 있도록 한다.
- [189] 일 구현예로, tracrRNA의 서열은 상기 CRISPR RNA 반복 서열과 0개 내지 20개의 미스매치가 있는 상보적인 서열을 포함할 수 있다. 바람직하게, tracrRNA 서열은 CRISPR RNA 반복 서열과 0개 내지 8개 또는 8개 내지 12개의 미스매치가 있는 상보적인 서열을 포함할 수 있다.
- [190] **(4) 고효율의 상동지정복구 효율 달성을 위한 gRNA의 변형**
- [191] **(4-1) 개괄**
- [192] 본 발명의 엔지니어링된 가이드 RNA(gRNA)에 적용된 변형은 궁극적으로 높은 상동지정복구 효율을 달성하기 위한 목적을 가진다. 즉, 본 발명에서 개시하는 변형들은 길이가 더 긴 야생형의 gRNA와 비교하여 표적 핵산에 대한 인식/절단 효율이 유지 또는 향상된 더 짧은 길이의 엔지니어링된 gRNA를 제조함으로써, AAV 전달체의 패키징 한계치(약 4.7 kb) 내에서 더 많은 공간을 상동지정복구에 필요한 다른 필수 구성요소(예를 들어, 공여자 핵산 분자) 및/또는 부가

적 구성요소(예를 들어, 비상동말단연결 과정에 관여하는 유전자의 억제를 위한 shRNA)에 할당할 수 있도록 하여 기존의 CRISPR/Cas 시스템으로는 달성할 수 없었던 고효율의 상동지정복구 효과를 부여하고자 함에 있다.

- [193] 따라서 본 발명에서 제공하는 엔지니어링된 gRNA는 기본적으로 야생형 Cas12f1 gRNA 서열에서 1 이상의 뉴클레오티드가 치환, 결실, 삽입 또는 부가된 서열을 포함한다. 이때, 엔지니어링된 gRNA는 가이드 서열을 제외한 부분이 상기 야생형 Cas12f1 gRNA와 적어도 50%의 서열 동일성을 가지는 것일 수 있다.
- [194] 상술한 바와 같이, 야생형 Cas12f1 gRNA는 (i) 하나 이상의 스템(stem) 영역, (ii) tracrRNA-crRNA 상보성 영역 및 임의적으로 (iii) 연속되는 3개 이상의 유라실(U)을 포함하는 영역을 포함하는 tracrRNA(trans-activating CRISPR RNA) 및 crRNA(CRISPR RNA)를 포함할 수 있으며, 본 발명의 엔지니어링된 gRNA는 (a) 하나 이상의 스템 영역의 일부 또는 전부의 결실; (b) tracrRNA-crRNA 상보성 영역의 일부 또는 전부의 결실; (c) 연속되는 3개 이상의 유라실(U)이 존재하는 경우 그 중 하나 이상의 U의 치환; 및 (d) crRNA 서열의 3'-말단에 하나 이상의 유리딘(uridine)의 부가로 이루어진 군에서 선택되는 하나 이상의 변형을 포함할 수 있다.
- [195] 여기서, 상기 야생형 Cas12f1 gRNA가 포함하는 tracrRNA는 서열번호 11의 핵산 서열로 이루어지거나 이를 포함하는 것일 수 있다. 또한, 상기 야생형 Cas12f1 gRNA가 포함하는 crRNA는 서열번호 12의 핵산 서열로 이루어지거나 이를 포함하는 것일 수 있다. 일 구체예에서, 야생형 Cas12f1 가이드 RNA는 서열번호 11의 핵산 서열로 이루어진 tracrRNA 및 서열번호 12의 핵산 서열로 이루어진 crRNA를 포함하는 gRNA일 수 있다. 아래 표 1에 야생형 Cas12f1의 tracrRNA 및 crRNA의 서열 정보가 제공된다.

- [196] [표1]

명칭	염기서열	서열 번호
Wild-type tracrRNA	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAAA AGCUGUCCCUUAGGGGAUUAGAACUUGAGUGAAG GUGGGCUGCUUGCAUCAGCCUAAUGUCGAGAAGU GCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUC AUUUUCCUCUCCAAUUCUGCACAA	11
Wild-type crRNA	GUUGCAGAACCCGAAUAGACGAAUGAAGGAAUGC AAC	12

- [197] 일 구현예에서, 야생형 Cas12f1 gRNA는 5'-말단부터 순차적으로 제1 스템 영역, 제2 스템 영역, 제3 스템 영역, 제4 스템 영역 및 tracrRNA-crRNA 상보성 영역을 포함하는 tracrRNA 및 crRNA를 포함하고, 상기 엔지니어링된 가이드 RNA는 (a1) 제1 스템 영역의 일부 또는 전부의 결실; (a2) 제2 스템 영역의 일부 또는

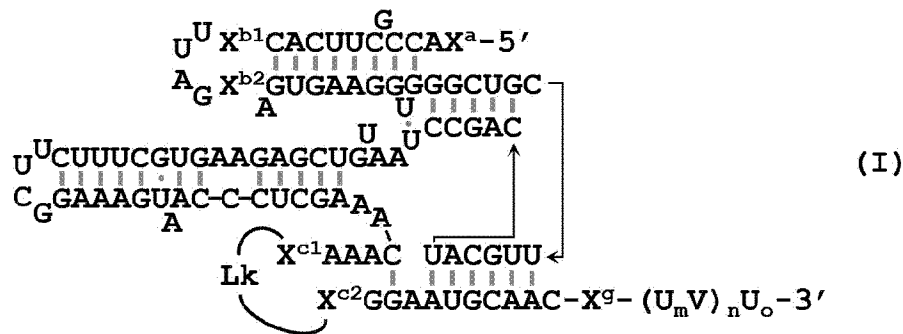
는 전부의 결실; (b) tracrRNA-crRNA 상보성 영역의 일부 또는 전부의 결실; (c) tracrRNA-crRNA 상보성 영역 내에 연속되는 3개 이상의 유라실(U)이 존재하는 경우 하나 이상의 U를 A, G 또는 C로 치환; 및 (d1) crRNA 서열의 3'-말단에 U-rich tail의 부가(상기 U-rich tail의 서열은 5'-(U_mV)_nU_o-3'로 표시되고 여기서 상기 V는 각각 독립적으로 A, C 또는 G이고, m 및 o는 1 내지 20 사이의 정수이며, n은 0 내지 5 사이의 정수임)로 이루어진 군에서 선택되는 하나 이상의 변형을 포함할 수 있다.

- [198] 여기서, 상기 야생형 Cas12f1 gRNA가 포함하는 tracrRNA는 서열번호 11의 핵산 서열로 이루어지거나 이를 포함하는 것일 수 있다. 또한, 상기 야생형 Cas12f1 gRNA가 포함하는 crRNA는 서열번호 12의 핵산 서열로 이루어지거나 이를 포함하는 것일 수 있다. 일 구체예에서, 야생형 Cas12f1 가이드 RNA는 서열번호 11의 핵산 서열로 이루어진 tracrRNA 및 서열번호 12의 핵산 서열로 이루어진 crRNA를 포함하는 gRNA일 수 있다.
- [199] 이하, 엔지니어링된 gRNA에서의 각 변형부위별 변형에 대해 자세히 설명한다.
- [200] **(4-2) 변형부위 1(modification site 1, MS1)에서의 변형**
- [201] 본 항목에서는 MS1에서의 변형을 기술한다. 일 구체예에서, 자연에 존재하는 가이드 RNA(gRNA)가 될 수 있는 야생형 tracrRNA(예컨대, 서열번호 11)는 서열 내에 연속된 다섯 개의 유라실(U)을 포함하는 서열을 가질 수 있다. 이는 상기 야생형 tracrRNA를 세포 내에서 벡터 등을 이용하여 발현시키고자 할 때, 특정 조건에서는 상기 서열이 전사종결신호로써 작용하여 의도하지 않은 전사의 조기 종결을 야기하는 문제를 안고 있다. 즉, 상기 연속된 다섯 개의 U를 포함하는 서열이 전사종결신호로써 작동하게 되는 경우에는 상기 tracrRNA의 정상적인 또는 완전한 발현이 억제되고, 정상적인 또는 완전한 gRNA의 형성 또한 저해되어 결과적으로 본 발명의 표적 핵산 편집 시스템의 표적 핵산 또는 표적 유전자의 절단 또는 상동지정복구 효율을 감소시킨다.
- [202] 따라서 상술한 문제점을 해결하기 위해, 엔지니어링된 gRNA는 야생형 tracrRNA(예컨대, 서열번호 11)의 연속된 세 개 이상, 네 개 이상, 다섯 개 이상의 U, 바람직하게는 네 개 또는 다섯 개의 U 중 적어도 하나의 U를 다른 뉴클레오티드인 A, C, T 또는 G로 인위적으로 변형시킨 것일 수 있다.
- [203] 일 구체예로, MS1으로 지칭되는 연속되는 3개 이상의 유라실(U)을 포함하는 영역에서 연속되는 3개 이상의 U 중 적어도 하나의 U를 다른 종류의 뉴클레오티드로 치환된 변형을 포함하는 엔지니어링된 gRNA가 제공된다. 일 예로, 상기 연속되는 3개 이상의 U는 tracrRNA의 tracrRNA-crRNA 상보성 영역 내에 존재할 수 있으며, 여기서 상기 연속되는 3개 이상의 U 중 하나 이상을 A, G 또는 C로 치환함으로써 3개 이상의 U가 연속되는 서열이 나타나지 않도록 변형될 수 있다.
- [204] 이때, 상기 변형되는 서열에 대응되는 crRNA의 tracrRNA-crRNA 상보성 영역 내 서열 또한 함께 변형되는 것이 바람직하다. 일 구체예로, tracrRNA의 tracrRNA-crRNA 상보성 영역 내에서 서열 5'-UUUUU-3'과 일부 상보적 결합을

이루는 crRNA의 tracrRNA-crRNA 상보성 영역 내에 서열 5'-ACGAA-3'가 존재하는 경우 해당 서열은 5'-NGNNN-3'로 치환될 수 있다. 여기서, N은 각각 독립적으로 A, C, G 또는 U이다.

- [205] 다른 구현예에서, 엔지니어링된 gRNA는 하기 식 (I)로 표시되는 서열로 이루어지거나 상기 서열과 80% 이상, 85% 이상, 90% 이상 또는 95% 이상 서열 동일성을 갖는 gRNA일 수 있다[하기 식 (I) 및 본 명세서 전체에 기재된 식 (I)에서, 검정색 실선은 뉴클레오티드 또는 특정 분자 사이의 화학적 결합(예를 들어, 포스포다이에스터 결합)을 의미하고, 회색 굵은선은 뉴클레오티드 사이의 상보적 결합을 의미한다]. 여기서 MS1은 식 (I)의 X^{c1} 및 X^{c2} 로 표시된 폴리뉴클레오티드 내에 존재할 수 있다.

[206]



- [207] 일 구현예로, 상기 식 (I)의 엔지니어링된 gRNA에서 X^{c1} 서열 내에 연속되는 3개 이상의 유라실(U)이 존재하는 경우 이들 중 하나 이상의 U가 A, G 또는 C로 치환되는 변형을 포함할 수 있다. 일 예로, X^{c1} 서열 내에 서열 5'-UUUUU-3'이 존재하는 경우 해당 서열은 5'-NNNCN-3'으로 치환될 수 있다. 여기서, N은 각각 독립적으로 A, C, G 또는 U이다. 보다 구체화된 예로, X^{c1} 서열 내의 서열 5'-UUUUU-3'은 하기 서열로 이루어진 군에서 선택된 어느 하나의 핵산 서열로 치환될 수 있으나, 연속되는 3개 이상의 U를 포함하는 서열을 나타나지 않게 하는 것이라면 하기 서열로 제한되지 않는다: 5'-UUUCU-3', 5'-GUUCU-3', 5'-UCUCU-3', 5'-UUGCU-3', 5'-UUUCC-3', 5'-GCUCU-3', 5'-GUUCC-3', 5'-UCGCU-3', 5'-UCUCC-3', 5'-UUGCC-3', 5'-GCGCU-3', 5'-GCUCC-3', 5'-GUGCC-3', 5'-UCGCC-3', 5'-GCGCC-3' 및 5'-GUGCU-3'.

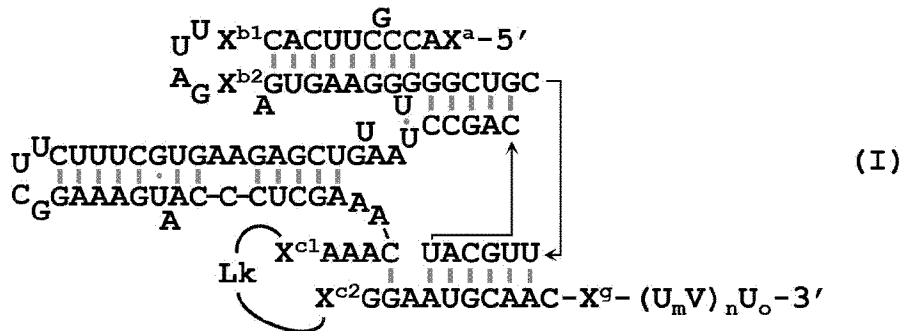
- [208] 다른 구현예로, 식 (I)의 엔지니어링된 gRNA에서 X^{c2} 서열은 X^{c1} 서열과 적어도 일부 서열이 상보적 결합을 이루는 영역을 포함하며(tracrRNA-crRNA 상보성 영역으로도 지칭됨), 이때 X^{c1} 서열 내에 존재하는 연속되는 3개 이상의 U와 적어도 하나의 상보성 결합을 형성하는 X^{c2} 서열 내의 대응 서열도 함께 변형될 수 있다. 일 예로, 상기 X^{c2} 서열 내에 서열 5'-ACGAA-3'가 존재하는 경우 해당 서열은 5'-NGNNN-3'로 치환될 수 있다. 여기서, N은 각각 독립적으로 A, C, G 또는 U이다. 보다 구체화된 예로, X^{c1} 서열 내의 서열 5'-ACGAA-3'은 하기 서열로 이루어진 군에서 선택된 어느 하나의 핵산 서열로 치환될 수 있으나 하기 서열에

제한되는 것은 아니다: 5'-AGGAA-3', 5'-AGCAA-3', 5'-AGAAA-3', 5'-AGCAU-3', 5'-AGCAG-3', 5'-AGCAC-3', 5'-AGCUA-3', 5'-AGCGA-3', 5'-AGCCA-3', 5'-UGCAA-3', 5'-UGCUA-3', 5'-UGCGA-3', 5'-UGCCA-3', 5'-GGCAA-3', 5'-GGCUA-3', 5'-GGCGA-3', 5'-GGCCA-3', 5'-CGCAA-3', 5'-CGCUA-3', 5'-CGCGA-3' 및 5'-CGCCA-3'.

- [209] 다른 구현예에서, X^c 서열 내의 연속되는 3개 이상의 U를 포함하는 서열이 다른 서열로 변형되는 경우, 이에 대응되는(즉, 적어도 일부가 상보적 결합을 형성하는) X^c 서열 내의 대응되는 뉴클레오티드는 변형된 뉴클레오티드와 상보적 결합을 이룰 수 있도록 변형되는 것이 바람직하다. 예를 들어, X^c 서열 내의 서열 5'-UUUUU-3'이 5'-GUGCU-3'으로 변형되는 경우 X^c 서열 내의 서열 5'-ACGAA-3'은 5'-AGCAA-3'로 변형되는 것이 바람직하나, 상보적 결합이 필수로 요구되는 것은 아니다.
- [210] **(4-3) 변형부위 2(modification site 2, MS2)에서의 변형**
- [211] 본 항목에서는 MS2에서의 변형을 기술한다. 일 구현예에서, 엔지니어링된 가이드 RNA(gRNA)는 자연계에서 발견되는 gRNA에 새로운 구성을 추가한 것으로서 crRNA 서열의 3'-말단에 하나 이상의 유리딘(uridine)이 부가된 것일 수 있다. 여기서, 상기 crRNA 서열의 3'-말단은 가이드 서열(스페이서)의 3'-말단일 수 있다. 본 명세서에서 상기 3'-말단에 부가된 하나 이상의 유리딘은 U-rich tail로도 지칭된다. 상기 3'-말단에 부가된 하나 이상의 유리딘 또는 U-rich tail을 포함하는 엔지니어링된 gRNA는 초소형 CRISPR/Cas12 시스템의 표적 핵산 또는 표적 유전자에 대한 핵산 절단 또는 상동지정복구 효율을 높이는 역할을 한다.
- [212] 본 명세서에서 사용되는 용어 "U-rich tail"은 유리딘(U)이 풍부하게 포함된 RNA 서열 그 자체뿐 아니라, 이를 암호화하는 DNA 서열을 의미할 수도 있으며, 이는 문맥에 따라서 적절하게 해석된다. 본 발명자들은 U-rich tail 서열의 구조 및 그 효과에 대해 실험적으로 자세히 밝혔으며, 이하 구체적인 구현예로 더 자세히 설명한다.
- [213] 일 구현예에서, U-rich tail 서열은 U_x 로 표현될 수 있다. 상기 x 는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20일 수 있다. 일 예로, x 는 상기 나열된 수치 중에서 선택된 두 수치 범위 내의 정수일 수 있다. 예를 들어, x 는 1 내지 6 사이의 정수일 수 있다. 또 다른 예를 들어, x 는 1 내지 20 사이의 정수일 수 있다. 일 구현예로, x 는 20 이상의 정수일 수 있다.
- [214] 다른 구현예에서, U-rich tail의 서열은 5'-(U_mV) $_nU_o$ -3'로 표시되고 여기서 상기 V 는 각각 독립적으로 A, C 또는 G이고, m 및 o 는 1 내지 20 사이의 정수이며, n 은 0 내지 5 사이의 정수일 수 있다. 일 예로, 상기 n 은 0, 1 또는 2일 수 있다. 일 예로, 상기 m 및 o 는 각각 독립적으로 1, 2, 3, 4, 5, 6, 7, 8, 9, 또는 10일 수 있다.
- [215] 다른 구현예에서, 엔지니어링된 gRNA는 하기 식 (I)로 표시되는 서열로 이루어지거나 상기 서열과 80% 이상, 85% 이상, 90% 이상 또는 95% 이상 서열 동일성

을 갖는 gRNA일 수 있고, 여기서 MS2는 식 (I)의 $(U_mV)_nU_o$ 에 대응되는 영역으로서 U는 유리딘이고, V, m, o 및 n은 위에 정의된 바와 같다.

[216]

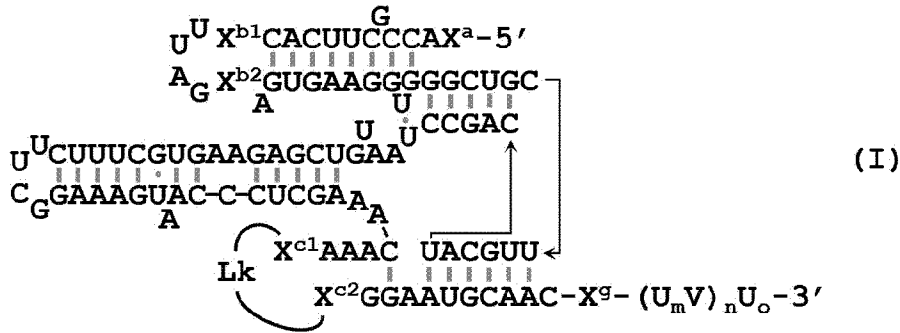


- [217] 바람직하게는, 상기 식 (I)로 표시되는 엔지니어링된 gRNA에서 $(U_mV)_nU_o$ 는 (i) n이 0이고, o가 1 내지 6 사이의 정수이거나 (ii) V가 A 또는 G이고, m 및 o는 각각 독립적으로 3 내지 6 사이의 정수이고, n은 1 내지 3 사이의 정수인 U-rich tail일 수 있다. 구체화된 예에서, 상기 식 (I)의 $(U_mV)_nU_o$ 는 5'-U-3', 5'-UU-3', 5'-UUU-3', 5'-UUUU-3', 5'-UUUUU-3', 5'-UUUUUU-3', 5'-UUUUUUU-3', 5'-UUUUUUUU-3', 5'-UUUUUUUUU-3', 5'-UUUUUUUUUU-3' 및 5'-UUUUUUUUUUU-3'로 이루어진 군에서 선택된 어느 하나의 서열로 이루어지고, 상기 R은 A 또는 G인 U-rich tail일 수 있다.
- [218] 또 다른 구현예에서, U-rich tail 서열은 유리딘이 1개 내지 5개 반복될 때마다 유리딘이 아닌 다른 리보뉴클레오타이드(A, C 또는 G)가 하나씩 포함된 변형된 유리딘 반복 서열을 포함할 수 있다. 상기 변형된 유리딘 연속 서열은 특히 엔지니어링된 crRNA를 발현하는 벡터를 설계할 때 유용하다. 일 구현예로, U-rich tail 서열은 UV, UUV, UUUV, UUUUV 및/또는 UUUUUUV가 하나 이상 반복된 서열을 포함할 수 있다. 이때, 상기 V는 A, C, G 중 하나이다.
- [219] 또한, 상기 U-rich tail 서열은 U_x 로 표현되는 서열 및 $(U_aV)_n$ 로 표현되는 서열이 조합된 형태일 수 있다. 일 구현예로, 상기 U-rich tail 서열은 $(U)^{n1}-V1-(U)^{n2}-V2-U_x$ 로 표현될 수 있다. 이때, V1 및 V2는 각각 아데닌(A), 시티딘(C), 구아닌(G) 중 하나이다. 이때, 상기 n1 및 n2는 각각 1 내지 4 사이의 정수일 수 있다. 이때, 상기 x는 1 내지 20 사이의 정수일 수 있다. 또한, 상기 U-rich tail 서열의 길이는 1nt, 2nt, 3nt, 4nt, 5nt, 6nt, 7nt, 8nt, 9nt, 10nt, 11nt, 12nt, 13nt, 14nt, 15nt, 16nt, 17nt, 18nt, 19nt, 또는 20nt일 수 있다. 일 구현예로, 상기 U-rich tail 서열의 길이는 20nt 이상일 수 있다.
- [220] 다른 구현예에서, 엔지니어링된 gRNA가 세포 내에서 발현될 경우 U-rich tail은 전사 조기 종결에 의해 복수의 형태로 존재할 수 있다. 예를 들어, 일 구현예에 따라 5'-UUUUUAUUUUUU-3' 서열의 U-rich tail이 포함되도록 의도한 gRNA가 세포 내에서 전사될 때 4개 이상 또는 5개 이상의 T는 종결 시퀀스로 작용할 수 있으므로, 5'-UUUUUAUUUU-3', 5'-UUUUUAUUUUU-3' 또는 5'-UUUUUAUUUUUU-3' 등의

U-rich tail을 포함하는 gRNA가 동시에 생성될 수 있다. 따라서, 본 발명에서 4개 이상의 U가 포함된 U-rich tail은 의도한 길이보다 더 짧은 길이의 U-rich tail 서열을 함께 포함하는 것으로 이해될 수 있다.

- [221] 또 다른 구현예에서, U-rich tail 서열은 본 발명의 상동지정복구를 위한 CRISPR/Cas12 시스템의 실사용 환경 및 발현 환경, 예를 들어 진핵 세포 또는 원핵 세포 내부 환경에 따라 유리딘 외에 추가적인 염기를 더 포함할 수 있다.
- [222] **(4-4) 변형부위 3(modification site 3, MS3)에서의 변형**
- [223] 본 항목에서는 MS3에서의 변형을 기술한다. 상술한 바와 같이 MS3은 gRNA 및 이펙터 단백질 복합체 내에서 스템 구조를 형성하는 뉴클레오티드의 일부 또는 전부를 포함하는 부위(제1 스템 영역으로 지칭될 수 있음)로서, 상기 MS3은 gRNA 및 이펙터 단백질이 복합체를 이룰 때 이펙터 단백질과 상호작용하지 않는 영역을 포함할 수 있다. MS3에서의 변형은 tracrRNA의 5'-말단 부근의 제1 스템 영역의 일부 또는 전부의 제거를 포함한다.
- [224] 일 구현예에서, 엔지니어링된 gRNA는 제1 스템 영역의 일부 또는 전부가 결실된 변형을 포함한다.
- [225] 다른 구현예에서, 엔지니어링된 gRNA는 tracrRNA 상의 제1 스템 영역의 일부 또는 전부가 결실된 변형을 포함하며, 이때 상기 결실되는 제1 스템 영역의 일부 또는 전부는 1개 내지 20개 뉴클레오티드일 수 있다. 구체적으로, 상기 제1 스템 영역의 일부 또는 전부는 2개 내지 20개, 3개 내지 20개, 4개 내지 20개, 5개 내지 20개, 6개 내지 20개, 7개 내지 20개, 8개 내지 20개, 9개 내지 20개, 10개 내지 20개, 11개 내지 20개, 12개 내지 20개, 13개 내지 20개, 14개 내지 20개, 15개 내지 20개, 16개 내지 20개, 17개 내지 20개, 18개 내지 20개, 19개 또는 20개 뉴클레오티드일 수 있다.
- [226] 또 다른 구현예에서, 엔지니어링된 gRNA는 하기 식 (I)로 표시되는 서열로 이루어지거나 상기 서열과 80% 이상, 85% 이상, 90% 이상 또는 95% 이상 서열 동일성을 갖는 gRNA일 수 있고, 여기서 MS3 또는 제1 스템 영역은 식 (I)의 X^a로 표시된 폴리뉴클레오티드에 대응되는 부위로서, 제1 스템 영역의 일부 또는 전부가 결실된 변형에 의해 상기 X^a는 0 내지 35개의 (폴리)뉴클레오티드로 이루어질 수 있고, 바람직하게는 0 내지 20개, 0 내지 19개, 0 내지 18개, 0 내지 17개, 0 내지 16개, 0 내지 15개, 0 내지 14개, 0 내지 13개, 0 내지 12개, 0 내지 11개, 0 내지 10개, 0 내지 9개, 0 내지 8개, 0 내지 7개, 0 내지 6개, 0 내지 5개, 0 내지 4개, 0 내지 3개, 0 내지 2개, 1개 또는 0개의 (폴리)뉴클레오티드로 이루어질 수 있다.

[227]



[228]

일 구현예로, 상기 식 (I)의 엔지니어링된 gRNA에서 X^a 는 서열번호 14의 핵산 서열을 포함하거나 상기 서열의 전부 또는 일부, 바람직하게는 상기 서열번호 14의 서열에서 1 내지 20개의 뉴클레오티드가 결실된 핵산 서열을 포함할 수 있다. 일 예로, 상기 뉴클레오티드의 결실은 서열번호 14의 서열에서 뉴클레오티드가 무작위로 적어도 1개, 2개, 3개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개 또는 20개가 결실된 것일 수 있다. 바람직한 예로, 상기 뉴클레오티드의 결실은 서열번호 14의 서열에서 5'-말단으로부터 적어도 1개, 2개, 3개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개 또는 20개의 뉴클레오티드가 5'-말단부터 순차적으로 결실된 것일 수 있다. 보다 구체적으로, 상기 X^a 는 5'-CUUCACUGAUAAAGUGGAGA-3'(서열번호 14), 5'-UUCACUGAUAAAGUGGAGA-3'(서열번호 15), 5'-UCACUGAUAAAGUGGAGA-3'(서열번호 16), 5'-CACUGAUAAAGUGGAGA-3'(서열번호 17), 5'-ACUGAUAAAGUGGAGA-3'(서열번호 18), 5'-CUGAUAAAGUGGAGA-3'(서열번호 19), 5'-UGAUAAAGUGGAGA-3'(서열번호 20), 5'-GAUAAAGUGGAGA-3'(서열번호 21), 5'-AUAAAGUGGAGA-3'(서열번호 22), 5'-UAAAGUGGAGA-3'(서열번호 23), 5'-AAAGUGGAGA-3'(서열번호 24), 5'-AAGUGGAGA-3', 5'-AGUGGAGA-3', 5'-GUGGAGA-3', 5'-UGGAGA-3', 5'-GGAGA-3', 5'-GAGA-3', 5'-AGA-3', 5'-GA-3' 또는 5'-A-3'의 서열을 포함하거나 이로 이루어질 수 있고, 또는 X^a 는 부존재할 수 있다.

[229]

(4-5) 변형부위 4(modification site 4, MS4)에서의 변형

[230]

본 항목에서는 MS4에서의 변형을 기술한다. MS4는 tracrRNA의 3'-말단부 및 crRNA의 5'-말단부에 걸쳐 위치한 부위, 또는 싱글 가이드 RNA 형태인 경우 tracrRNA에 해당하는 서열과 crRNA에 해당하는 서열이 적어도 일부 상보적 결합을 이루는 부위로서 tracrRNA-crRNA 상보성 영역(제5 스템 영역으로도 지칭될 수 있음)으로 지칭되는 서열의 일부 또는 전부를 포함할 수 있다. 본 발명에서 tracrRNA-crRNA 상보성 영역은 변형부위 1(MS1)과 변형부위 4(MS4)를 함께 포함할 수 있다. MS4에서의 변형은 tracrRNA-crRNA 상보성 영역의 일부 또는 전부의 결실을 포함한다. 상기 tracrRNA-crRNA 상보성 영역은 tracrRNA의 일부 및 crRNA의 일부를 포함하여, gRNA 및 핵산 분해 단백질의 복합체 내에서

tracrRNA에 포함된 일부 뉴클레오티드가 crRNA에 포함된 일부 뉴클레오티드와 상보적인 결합을 형성할 수 있는 뉴클레오티드를 포함하고, 이와 인접한 뉴클레오티드를 포함할 수 있다. tracrRNA의 tracrRNA-crRNA 상보성 영역은 gRNA와 핵산 분해 단백질 복합체 내에서 핵산 분해 단백질과 상호작용하지 않는 영역을 포함할 수 있다.

- [231] 일부 구현예에서, 엔지니어링된 gRNA는 tracrRNA에서의 tracrRNA-crRNA 상보성 영역의 일부 또는 전부의 결실, crRNA에서의 tracrRNA-crRNA 상보성 영역의 일부 또는 전부의 결실, 또는 상기 tracrRNA 및 crRNA 모두에서의 tracrRNA-crRNA 상보성 영역의 일부 또는 전부의 결실을 포함한다.
- [232] 다른 구현예에서, 엔지니어링된 gRNA는 tracrRNA-crRNA 상보성 영역의 일부가 결실된 변형을 포함하며, 이때 결실되는 상기 상보성 영역의 일부는 1개 내지 54개 뉴클레오티드일 수 있다.
- [233] 또 다른 구현예에서, 엔지니어링된 gRNA는 tracrRNA-crRNA 상보성 영역의 전부가 결실된 변형을 포함하며, 이때 결실되는 상기 상보성 영역의 전부는 55개 뉴클레오티드일 수 있다.
- [234] 구체적으로, 상기 tracrRNA-crRNA 상보성 영역의 일부 또는 전부는 3개 내지 55개, 5개 내지 55개, 7개 내지 55개, 9개 내지 55개, 11개 내지 55개, 13개 내지 55개, 15개 내지 55개, 17개 내지 55개, 19개 내지 55개, 21개 내지 55개, 23개 내지 55개, 25개 내지 55개, 27개 내지 55개, 29개 내지 55개, 31개 내지 55개, 33개 내지 55개, 35개 내지 55개, 37개 내지 55개, 39개 내지 55개 또는 41개 내지 55개 뉴클레오티드일 수 있으며, 바람직하게는 42개 내지 55개, 43개 내지 55개, 44개 내지 55개, 45개 내지 55개, 46개 내지 55개, 47개 내지 55개, 48개 내지 55개, 49개 내지 55개, 50개 내지 55개, 51개 내지 55개, 52개 내지 55개, 53개 내지 55개, 54개 또는 55개 뉴클레오티드일 수 있다.
- [235] 또 다른 구현예에서, 엔지니어링된 gRNA는 하기 식 (I)로 표시되는 서열로 이루어지거나 상기 서열과 80% 이상, 85% 이상, 90% 이상 또는 95% 이상 서열 동일성을 갖는 gRNA일 수 있고, 여기서 MS4 또는 tracrRNA-crRNA 상보성 영역은 식 (I)의 X^{c1} 및 X^{c2} 로 표시된 폴리뉴클레오티드에 대응되는 영역으로서, tracrRNA-crRNA 상보성 영역의 일부 또는 전부가 결실된 변형에 의해 상기 X^{c1} 및 X^{c2} 는 각각 독립적으로 0 내지 35개의 (폴리)뉴클레오티드로 이루어질 수 있다.

(I)

U^U X^{b1} CACU UCCCAX^a-5'

A_G X^{b2} GUGAAGGGGGCUGC

U^U CCGAC

U^U CUUUCGUGAAGAGCUGAA

C_G GAAAGUAC-C-CUCGAA_A

X^{c1} AAAC

UACGUU

X^{c2} GGAAUGCAAC-X^g-(U_m V)_n U_o-3'

Lk

[238] 일 구현예로, 상기 식 (I)의 엔지니어링된 gRNA에서 X^{cl}은 서열번호 39의 핵산 서열을 포함하거나 상기 서열번호 39의 서열에서 1 내지 28개의 뉴클레오티드가 결실된 핵산 서열을 포함할 수 있다. 바람직하게, 상기 뉴클레오티드의 결실은 서열번호 39의 서열에서 5'-말단으로부터 적어도 1개, 2개, 3개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 23개, 24개, 25개, 26개, 27개 또는 28개의 뉴클레오티드가 순차적으로 제거된 것일 수 있다. 보다 구체적으로, 상기 X^{cl}은 5'-UUCAUUUUUCCUCUCCAAUUCUGCACAA-3'(서열번호 39), 5'-UUCAUUUUUCCUCUCCAAUUCUGCACA-3'(서열번호 40), 5'-UUCAUUUUUCCUCUCCAAUUCUGCAC-3'(서열번호 41), 5'-UUCAUUUUUCCUCUCCAAUUCUGCA-3'(서열번호 42), 5'-UUCAUUUUUCCUCUCCAAUUCUGC-3'(서열번호 43), 5'-UUCAUUUUUCCUCUCCAAUUCUG-3'(서열번호 44), 5'-UUCAUUUUUCCUCUCCAAUUCU-3'(서열번호 45), 5'-UUCAUUUUUCCUCUCCAAUUC-3'(서열번호 46), 5'-UUCAUUUUUCCUCUCCAAUU-3'(서열번호 47), 5'-UUCAUUUUUCCUCUCCAAU-3'(서열번호 48), 5'-UUCAUUUUUCCUCUCCAA-3'(서열번호 49), 5'-UUCAUUUUUCCUCUCCA-3'(서열번호 50), 5'-UUCAUUUUUCCUCUCC-3'(서

열번호 51), 5'-UUCAUUUUUCCUCUC-3'(서열번호 52), 5'-UUCAUUUUUCCUCU-3'(서열번호 53), 5'-UUCAUUUUUCCUC-3'(서열번호 54), 5'-UUCAUUUUUCCU-3'(서열번호 55), 5'-UUCAUUUUUCC-3'(서열번호 56), 5'-UUCAUUUUUC-3'(서열번호 57), 5'-UUCAUUUUU-3', 5'-UUCAUUUU-3', 5'-UUCAUUU-3', 5'-UUCAUU-3', 5'-UUCAU-3', 5'-UUCA-3', 5'-UUC-3', 5'-UU-3' 또는 5'-U-3'의 서열을 포함하거나 이로 이루어질 수 있고, 또는 X¹은 부존재할 수 있다.

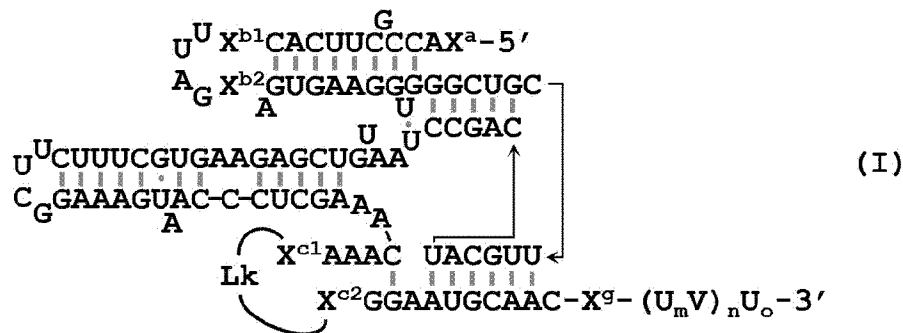
[239] 이때, 일부 뉴클레오티드가 제거된 X¹ 서열 내에 3개, 4개 또는 5개 이상의 유라실(U)을 포함하는 영역이 존재하는 경우에는 상술한 MS1에서의 변형이 또한 적용될 수 있다. MS1에 대한 구체적인 내용은 상기 "(4-2) 변형부위 1(modification site 1, MS1)에서의 변형" 항목을 참조한다.

[240] 또 다른 구현예로, 상기 식 (I)의 엔지니어링된 gRNA에서 X²는 서열번호 58의 핵산 서열을 포함하거나 상기 서열번호 58의 서열에서 1 내지 27개의 뉴클레오티드가 결실된 핵산 서열을 포함할 수 있다. 바람직하게, 상기 뉴클레오티드의 결실은 서열번호 58의 서열에서 5'-말단으로부터 적어도 1개, 2개, 3개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 23개, 24개, 25개, 26개 또는 27개의 뉴클레오티드가 순차적으로 제거된 것일 수 있다. 보다 구체적으로, 상기 X²는 5'-GUUGCAGAACCCGAAUAGACGAAUGAA-3'(서열번호 58), 5'-UUGCAGAACCCGAAUAGACGAAUGAA-3'(서열번호 59), 5'-UGCAGAACCCGAAUAGACGAAUGAA-3'(서열번호 60), 5'-GCAGAACCCGAAUAGACGAAUGAA-3'(서열번호 61), 5'-CAGAACCCGAAUAGACGAAUGAA-3'(서열번호 62), 5'-AGAACCCGAAUAGACGAAUGAA-3'(서열번호 63), 5'-GAACCCGAAUAGACGAAUGAA-3'(서열번호 64), 5'-AACCCGAAUAGACGAAUGAA-3'(서열번호 65), 5'-ACCCGAAUAGACGAAUGAA-3'(서열번호 66), 5'-CCCGAAUAGACGAAUGAA-3'(서열번호 67), 5'-CCGAAUAGACGAAUGAA-3'(서열번호 68), 5'-CGAAUAGACGAAUGAA-3'(서열번호 69), 5'-GAAUAGACGAAUGAA-3'(서열번호 70), 5'-AAUAGACGAAUGAA-3'(서열번호 71), 5'-AUAGACGAAUGAA-3'(서열번호 72), 5'-UAGACGAAUGAA-3'(서열번호 73), 5'-AGACGAAUGAA-3'(서열번호 74), 5'-GACGAAUGAA-3'(서열번호 75), 5'-ACGAAUGAA-3', 5'-CGAAUGAA-3', 5'-GAAUGAA-3', 5'-AAUGAA-3', 5'-AUGAA-3', 5'-UGAA-3', 5'-GAA-3', 5'-AA-3' 또는 5'-A-3'의 서열을 포함하거나 이로 이루어질 수 있고, 또는 X²는 부존재할 수 있다.

- [241] 이때, 일부 뉴클레오티드가 제거된 X^{c2} 서열 내에 X^{c1} 서열 내 3개 이상, 3개, 4개 또는 5개 이상의 U를 포함하는 서열에 대응되는 서열이 존재하는 경우에는 상술한 MS1에서의 변형이 또한 적용될 수 있다. MS1에 대한 구체적인 내용은 상기 "(4-2) 변형부위 1(modification site 1, MS1)에서의 변형" 항목을 참조한다.
- [242] 상술한 식 (I)의 엔지니어링된 gRNA에서 X^{c1} 과 X^{c2} 에 해당하는 영역은 각각 독립적으로 변형될 수 있으나, MS4 또는 tracrRNA-crRNA 상보성 영역은 tracrRNA와 crRNA가 상보적 결합을 이루는 영역으로서 듀얼 가이드 RNA로 작동하기 위해서는 X^{c1} 및 X^{c2} 각각에서 결실되는 뉴클레오티드의 위치와 개수를 동일하거나 유사하게 하는 것이 바람직하다. 즉, 상보성의 보존을 위해, MS4(tracrRNA-crRNA 상보성 영역)에서 tracrRNA의 3'-말단에 위치한 서열부터 순차적으로 결실시키는 경우 crRNA는 5'-말단 서열부터 순차적으로 결실시키는 것이 바람직하다.
- [243] 일부 구현예에서, 식 (I)의 엔지니어링된 gRNA에서 X^{c1} 의 3'-말단과 X^{c2} 의 5'-말단은 링커(Lk)로 연결되어 싱글 가이드 RNA(sgRNA) 형태로 변형될 수 있다. 상기 Lk는 tracrRNA 및 crRNA를 물리적 또는 화학적으로 연결하는 서열로서, 길이 1 내지 30개의 폴리뉴클레오티드 서열일 수 있다. 일 구현예로서, 상기 Lk는 1 내지 5개, 5 내지 10개, 10 내지 15개, 2 내지 20개, 15 내지 20개, 20개 내지 25개 또는 25 내지 30개의 뉴클레오티드 서열일 수 있다. 예를 들어, 상기 Lk는 5'-GAAA-3' 서열일 수 있으나, 이에 제한되는 것은 아니다. 다른 예로, 상기 Lk는 5'-UUAG-3', 5'-UGAAAA-3', 5'-UUGAAAAA-3', 5'-UUCGAAAGAA-3'(서열번호 76), 5'-UUCAGAAAUGAA-3'(서열번호 77), 5'-UUCAUGAAAAUGAA-3'(서열번호 78) 또는 5'-UUCAUUGAAAAAUGAA-3'(서열번호 79)의 서열을 포함하거나 이로 이루어진 링커일 수 있다.
- [244] 한편, sgRNA로 만들기 위해 링커(Lk)를 사용하는 것도 가능하지만, 3'-말단부의 일부 서열이 제거된 tracrRNA의 3' 말단부와 5'-말단부의 일부 서열이 제거된 crRNA의 3'-말단부를 직접 연결하는 것도 가능하다.
- [245] 또 다른 구현예로, 식 (I)의 엔지니어링된 gRNA에서 X^{c1} 과 X^{c2} 가 링커로 연결되는 경우, 식 (I)에서와 같이 5'- X^{c1} -Lk- X^{c2} -3'로 표현될 수 있으며, 이는 서열번호 80 내지 서열번호 86으로 이루어진 군에서 선택된 어느 하나의 핵산 서열일 수 있으나, 이에 제한되는 것은 아니다.
- [246] **(4-6) 변형부위 5(modification site 5, MS5)에서의 변형**
- [247] 본 항목에서는 MS5에서의 변형을 기술한다. 상술한 바와 같이, MS5는 제2 스템 영역으로 지칭되는 tracrRNA 내 3'-말단 방향에 위치한 영역에 대응된다. 상기 제2 스템 영역은 가이드 RNA(gRNA) 및 핵산 편집 단백질 복합체 내에서 스템 구조를 형성하는 뉴클레오티드를 포함하고, 이와 인접한 뉴클레오티드를 포함할 수 있다. 이때, 상기 스템 구조는 상술한 제1 스템 영역에 포함된 스템과는 구분되는 것이다.

- [248] 일 구현예에서, 엔지니어링된 gRNA는 제2 스템 영역의 일부 또는 전부가 결실된 변형을 포함한다.
- [249] 다른 구현예에서, 엔지니어링된 gRNA는 제2 스템 영역의 일부 또는 전부의 결실을 포함하고, 이때 상기 결실되는 제2 스템 영역의 일부 또는 전부는 1개 내지 27개 뉴클레오티드일 수 있다. 구체적으로, 상기 제2 스템 영역의 일부 또는 전부는 2개 내지 27개, 3개 내지 27개, 4개 내지 27개, 5개 내지 27개, 6개 내지 27개, 7개 내지 27개, 8개 내지 27개, 9개 내지 27개, 10개 내지 27개, 11개 내지 27개, 12개 내지 27개, 13개 내지 27개, 14개 내지 27개, 15개 내지 27개, 16개 내지 27개, 17개 내지 27개, 18개 내지 27개, 19개 내지 27개, 20개 내지 27개, 21개 내지 27개, 22개 내지 27개, 23개 내지 27개, 24개 내지 27개, 25개 내지 27개, 26개 또는 27개의 뉴클레오티드일 수 있다.
- [250] 또 다른 구현예에서, 엔지니어링된 gRNA는 하기 식 (I)로 표시되는 서열로 이루어지거나 상기 서열과 80% 이상, 85% 이상, 90% 이상 또는 95% 이상 서열 동일성을 갖는 gRNA일 수 있고, 여기서 MS5 또는 제2 스템 영역은 식 (I)의 X^{b1} 및 X^{b2} 로 표시된 폴리뉴클레오티드와 인접한 (폴리)뉴클레오티드(5'-UUAG-3' 서열의 루프 포함)를 포함하는 부위로서, 제2 스템 영역의 일부 또는 전부가 결실된 변형에 의해 상기 X^{b1} 및 X^{b2} 는 각각 독립적으로 0 내지 35개의 (폴리)뉴클레오티드로 이루어질 수 있다.

[251]



- [252] 바람직하게, 상기 X^{b1} 은 0 내지 13개, 0 내지 12개, 0 내지 11개, 0 내지 10개, 0 내지 9개, 0 내지 8개, 0 내지 7개, 0 내지 6개, 0 내지 5개, 0 내지 4개, 0 내지 3개, 0 내지 2개, 1개 또는 0개의 (폴리)뉴클레오티드로 이루어질 수 있다. 또한, 바람직하게, 상기 X^{b2} 는 0 내지 14개, 0 내지 13개, 0 내지 12개, 0 내지 11개, 0 내지 10개, 0 내지 9개, 0 내지 8개, 0 내지 7개, 0 내지 6개, 0 내지 5개, 0 내지 4개, 0 내지 3개, 0 내지 2개, 1개 또는 0개의 (폴리)뉴클레오티드로 이루어질 수 있다.
- [253] 일 구현예로, 상기 식 (I)의 엔지니어링된 gRNA에서 X^{b1} 은 서열번호 25의 핵산 서열을 포함하거나 상기 서열번호 25의 서열에서 1 내지 13개의 뉴클레오티드가 결실된 핵산 서열을 포함할 수 있다. 바람직하게, 상기 뉴클레오티드의 결실은 서열번호 25의 서열에서 5'-말단으로부터 적어도 1개, 2개, 3개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개 또는 13개의 뉴클레오티드가 순차적으로 제거된

것일 수 있다. 보다 구체적으로, 상기 X^{b1} 은 5'-CAAAAGCUGUCCC-3'(서열번호 25), 5'-CAAAAGCUGUCC-3'(서열번호 26), 5'-CAAAAGCUGUC-3'(서열번호 27), 5'-CAAAAGCUGU-3'(서열번호 28), 5'-CAAAAGCUG-3', 5'-CAAAAGCU-3', 5'-CAAAAGC-3', 5'-CAAAAG-3', 5'-CAAAA-3', 5'-CAAA-3', 5'-CAA-3', 5'-CA-3' 또는 5'-C-3'의 서열을 포함하거나 이로 이루어질 수 있고, 또는 X^{b1} 은 부존재할 수 있다.

[254] 다른 구현예로, 상기 식 (I)의 엔지니어링된 gRNA에서 X^{b2} 는 서열번호 29의 핵산 서열을 포함하거나 상기 서열번호 29의 서열에서 1 내지 14개의 뉴클레오티드가 결실된 핵산 서열을 포함할 수 있다. 바람직하게, 상기 뉴클레오티드의 결실은 서열번호 29의 서열에서 5'-말단으로부터 적어도 1개, 2개, 3개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개 또는 14개의 뉴클레오티드가 순차적으로 제거된 것일 수 있다. 보다 구체적으로, 상기 X^{b2} 는 5'-GGGAUUAGAACUUG-3' (서열번호 29), 5'-GGAUUAGAACUUG-3'(서열번호 30), 5'-GAUUAGAACUUG-3'(서열번호 31), 5'-AUUAGAACUUG-3'(서열번호 32), 5'-UUAGAACUUG-3'(서열번호 33), 5'-UAGAACUUG-3', 5'-AGAACUUG-3', 5'-GAACUUG-3', 5'-AACUUG-3', 5'-ACUUG-3', 5'-CUUG-3', 5'-UUG-3', 5'-UG-3' 또는 5'-G-3'의 서열을 포함하거나 이로 이루어질 수 있고, 또는 X^{b1} 은 부존재할 수 있다.

[255] 상술한 식 (I)의 엔지니어링된 gRNA에서 X^{b1} 과 X^{b2} 에 해당하는 영역은 각각 독립적으로 변형될 수 있으나, 정상적인 스템-루프 구조의 보존을 위해 X^{b1} 및 X^{b2} 각각에서 결실되는 뉴클레오티드의 위치와 개수를 동일하거나 유사하게 하는 것이 바람직하다. 예를 들어, X^{b1} 에서 5'-말단 방향의 서열부터 순차적으로 결실시키는 경우 X^{b2} 에서는 3'-말단 방향의 서열부터 순차적으로 결실시키는 것이 바람직하다.

[256] 다른 구현예에서, 식 (I)의 엔지니어링된 gRNA의 X^{b1} 과 X^{b2} 를 연결하는 루프(Loop) 부분의 서열은 5'-UUAG-3'로 표시되어 있으나, 이는 필요에 따라 5'-NNNN-3', 5'-NNN-3' 등의 다른 서열로 치환될 수 있다. 여기서, N은 각각 독립적으로 A, C, G 또는 U이다. 예를 들면, 상기 5'-NNNN-3'는 5'-GAAA-3'일 수 있고, 상기 5'-NNN-3'은 5'-CGA-3'일 수 있다.

[257] 일 예로서, 식 (I)의 엔지니어링된 gRNA의 X^{b1} 과 X^{b2} 를 연결하는 루프(Loop) 부분의 서열은 5'-UUAG-3'이고, 상기 식 (I) 내의 서열 5'- X^{b1} UUAG X^{b2} -3'은 서열번호 34 내지 서열번호 38로 이루어진 군에서 선택된 어느 하나의 핵산 서열을 포함하거나 이로 이루어진 것일 수 있다.

[258] **(5) 추가 서열(additional sequence)**

[259] 본 발명의 상기 엔지니어링된 tracrRNA는 추가 서열(additional sequence)을 선택적으로 더 포함할 수 있다. 상기 추가 서열은 엔지니어링된 tracrRNA의 3'-말단에 위치할 수 있다. 또한, 상기 추가 서열은 또한 엔지니어링된 tracrRNA의 5'-말

단에 위치할 수도 있다. 예를 들어, 상기 추가 서열은 제1 스템 영역의 5'-말단에 위치할 수 있다.

- [260] 상기 추가 서열은 1개 내지 40개의 뉴클레오티드일 수 있다. 일 구현예로서, 상기 추가 서열은 임의의 뉴클레오티드 서열 또는 임의로 배열된 뉴클레오티드 서열일 수 있다. 예를 들어, 상기 추가 서열은 5'-AUAAAGGUGA-3'(서열번호 187) 서열일 수 있다.
- [261] 또한, 상기 추가 서열은 공지된 뉴클레오티드 서열일 수 있다. 일 예로, 상기 추가 서열은 망치머리형 리보자임(hammerhead ribozyme) 뉴클레오티드 서열일 수 있다. 여기서, 상기 망치머리형 리보자임의 뉴클레오티드 서열은 5'-CUGAUGAGUCCGUGAGGACGAAACGAGUAAGCUCGUC-3'(서열번호 188) 서열 또는 5'-CUGCUCGAAUGAGCAAAGCAGGAGUGCCUGAGUAGUC-3'(서열번호 189) 서열일 수 있다. 상기 열거한 서열들은 단순 예시로서, 추가 서열이 이에 제한되는 것은 아니다.
- [262] (5) 변형부위 1 내지 변형부위 5에서의 변형이 적용된 gRNA의 예시
- [263] 본 발명의 표적 핵산 편집 시스템에 포함되는 엔지니어링된 가이드 RNA(gRNA)는 상술한 변형부위 1(MS1) 내지 변형부위 5(MS5) 중 둘 이상의 변형부위에서의 변형을 포함하는 것일 수 있다.
- [264] 일부 구현예에서, 엔지니어링된 가이드 RNA는 (a1) 제1 스템 영역의 일부 또는 전부의 결실; (a2) 제2 스템 영역의 일부 또는 전부의 결실; (b) tracrRNA-crRNA 상보성 영역의 일부 또는 전부의 결실; (c) tracrRNA-crRNA 상보성 영역 내에 연속되는 3개 이상의 유라실(U)이 존재하는 경우 하나 이상의 U를 A, G 또는 C로 치환; 및 (d1) crRNA 서열의 3'-말단에 U-rich tail의 부가로 이루어진 군에서 선택되는 하나 이상의 변형을 포함하는 것일 수 있다. 상기 U-rich tail의 서열은 5'-(U_mV)_nU_o-3'로 표시될 수 있고, 여기서 상기 V는 각각 독립적으로 A, C 또는 G이고, m 및 o는 1 내지 20 사이의 정수이며, n은 0 내지 5 사이의 정수이다.
- [265] 일 예로, 상기 엔지니어링된 가이드 RNA는 (d1) crRNA 서열의 3'-말단에 U-rich tail의 부가 및 (c) tracrRNA-crRNA 상보성 영역 내에 연속되는 3개 이상의 유라실(U)이 존재하는 경우 하나 이상의 U를 A, G 또는 C로 치환을 포함하는 것일 수 있다.
- [266] 다른 예로, 상기 엔지니어링된 가이드 RNA는 (d1) crRNA 서열의 3'-말단에 U-rich tail의 부가, (c) tracrRNA-crRNA 상보성 영역 내에 연속되는 3개 이상의 유라실(U)이 존재하는 경우 하나 이상의 U를 A, G 또는 C로 치환 및 (a1) 제1 스템 영역의 일부 또는 전부의 결실을 포함하는 것일 수 있다.
- [267] 또 다른 예로, 상기 엔지니어링된 가이드 RNA는 (d1) crRNA 서열의 3'-말단에 U-rich tail의 부가, (c) tracrRNA-crRNA 상보성 영역 내에 연속되는 3개 이상의 유라실(U)이 존재하는 경우 하나 이상의 U를 A, G 또는 C로 치환 및 (a1) 제1 스템 영역의 일부 또는 전부의 결실을 포함하는 것일 수 있다.

- [268] 또 다른 예로, 상기 엔지니어링된 가이드 RNA는 (d1) crRNA 서열의 3'-말단에 U-rich tail의 부가, (a1) 제1 스템 영역의 일부 또는 전부의 결실 및 (b) tracrRNA-crRNA 상보성 영역의 일부 또는 전부의 결실을 포함하는 것일 수 있으며, 상기 일부 결실을 포함하는 tracrRNA-crRNA 상보성 영역 내에 연속되는 3개 이상의 유라실(U)이 존재하는 경우 하나 이상의 U를 A, G 또는 C로의 치환이 추가로 포함될 수 있다.
- [269] 또 다른 예로, 상기 엔지니어링된 가이드 RNA는 (d1) crRNA 서열의 3'-말단에 U-rich tail의 부가, (a1) 제1 스템 영역의 일부 또는 전부의 결실, (b) tracrRNA-crRNA 상보성 영역의 일부 또는 전부의 결실 및 (a2) 제2 스템 영역의 일부 또는 전부의 결실을 포함하는 것일 수 있으며, 상기 일부 결실을 포함하는 tracrRNA-crRNA 상보성 영역 내에 연속되는 3개 이상의 유라실(U)이 존재하는 경우 하나 이상의 U를 A, G 또는 C로의 치환이 추가로 포함될 수 있다.
- [270] 상술한 복수의 변형부위(MS)에서의 변형이 적용된 tracrRNA의 예시로서 서열번호 87 내지 서열번호 132의 뉴클레오티드 서열 포함하는 엔지니어링된 tracrRNA가 제공된다.
- [271] 구체적으로, 본 발명의 엔지니어링된 tracrRNA는 서열번호 87(MS1), 서열번호 88(MS1/MS3-1), 서열번호 89(MS1/MS3-2), 서열번호 90(MS1/MS3-3), 서열번호 91(MS1/MS4*-1), 서열번호 92 (MS1/MS4*-2), 서열번호 93(MS1/MS4*-3), 서열번호 94(MS1/MS5-1), 서열번호 95(MS1/MS5-2), 서열번호 96(MS1/MS5-3), 서열번호 97(MS1/MS3-3/MS4*-1), 서열번호 98(MS1/MS3-3/MS4*-2), 서열번호 99(MS1/MS3-3/MS4*-3), 서열번호 100(MS1/MS4*-2/MS5-1), 서열번호 101(MS1/MS4*-2/MS5-2), 서열번호 102(MS1/MS4*-2/MS5-3), 서열번호 103(MS1/MS3-3/MS5-1), 서열번호 104(MS1/MS3-3/MS5-2), 서열번호 105(MS1/MS3-3/MS5-3), 서열번호 106(MS1/MS3-3/MS4*-2/MS5-3), 서열번호 107(mature form, MF), 서열번호 108(MF/MS3-1), 서열번호 109(MF/MS3-2), 서열번호 110(MF/MS3-3), 서열번호 111(MF/MS4-1), 서열번호 112(MF/MS4-2), 서열번호 113(MF/MS4-3), 서열번호 114(MF/MS5-1), 서열번호 115(MF/MS5-2), 서열번호 116(MF/MS5-3), 서열번호 117(MF/MS5), 서열번호 118(MF/MS3-3/MS4-1), 서열번호 119(MF/MS3-3/MS4-2), 서열번호 120(MF/MS3-3/MS4-3), 서열번호 121(MF/MS4-3/MS5-1), 서열번호 122(MF/MS4-3/MS5-2), 서열번호 123(MF/MS4-3/MS5-3), 서열번호 124(MF/MS4-3/MS5-F), 서열번호 125(MF/MS3-3/MS5-1), 서열번호 126(MF/MS3-3/MS5-2), 서열번호 127(MF/MS3-3/MS5-3), 서열번호 128(MF/MS3-3/MS5), 서열번호 129(MF/MS3-3/MS4-3/MS5-3), 서열번호 130(MF/MS3-3/MS4-1/MS5), 서열번호 131(MF/MS3-3/MS4-2/MS5) 또는 서열번호 132(MF/MS3-3/MS4-3/MS5)의 뉴클레오티드 서열을 포함하거나 이로 이루어진 것일 수 있다.

[272] 일부 구현예로서, MS1, MS3, MS4 및 MS5에서 선택된 어느 하나 이상의 변형 부위에서 하나 이상의 변형을 갖는 엔지니어링된 tracrRNA의 예시적인 서열을 하기 표 2에 나타내었다.

[273] [표2]

tracrRNA	염기서열	서열 번호
MS1	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUGUCCCUUAGGGGAUUAGAACUUGAGUGA AGGUGGGCUGCUUGCAUCAGCCUAAUGUCGAGA AGUGCUUUCUUCGGAAAGUAACCCUCGAAACAA AUUCAGUGCUCUCCUCCAAUUCUGCACAA	87
MS1/MS3-1	GAUAAAGUGGAGAACCGCUUCACCAAAGCUGU CCCUUAGGGGAUUAGAACUUGAGUGAAGGUGG GCUGCUUGCAUCAGCCUAAUGUCGAGAAGUGCU UUCUUCGGAAAGUAACCCUCGAAACAAAUUCAG UGCUCUCCUCCAAUUCUGCACAA	88
MS1/MS3-2	UGGAGAACCGCUUCACCAAAGCUGUCCCUUAG GGGAUUAGAACUUGAGUGAAGGUGGGCUGCUU GCAUCAGCCUAAUGUCGAGAAGUGCUUUCUUCG GAAAGUAACCCUCGAAACAAAUUCAGUGCUCU CUCCAAUUCUGCACAA	89
MS1/MS3-3	ACCGCUUCACCAAAGCUGUCCCUUAGGGGAUU AGAACUUGAGUGAAGGUGGGCUGCUUGCAUCAG CCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUA ACCCUCGAAACAAAUUCAGUGCUCUCCUCCAAU UCUGCACAA	90
MS1/MS4*-1	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUGUCCCUUAGGGGAUUAGAACUUGAGUGA AGGUGGGCUGCUUGCAUCAGCCUAAUGUCGAGA AGUGCUUUCUUCGGAAAGUAACCCUCGAAACAA AUUCAGUGCUCUCCUCCAAUUC	91
MS1/MS4*-2	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUGUCCCUUAGGGGAUUAGAACUUGAGUGA AGGUGGGCUGCUUGCAUCAGCCUAAUGUCGAGA AGUGCUUUCUUCGGAAAGUAACCCUCGAAACAA AUUCAGUGCUCUCCUCUC	92

MS1/MS4*-3	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUGUCCCUUAGGGGAUUAGAACUUGAGUGA AGGUGGGCUGCUUGCAUCAGCCUAAUGUCGAGA AGUGCUUUCUUCGGAAAGUAACCCUCGAAACAA AUUCAGUGCU	93
MS1/MS5-1	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUGUUUAGAUUAGAACUUGAGUGAAGGUG GGCUGCUUGCAUCAGCCUAAUGUCGAGAAGUGC UUUCUUCGGAAAGUAACCCUCGAAACAAAUUCA GUGCUCCUCUCCAAUUCUGCACAA	94
MS1/MS5-2	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUUAGGAACUUGAGUGAAGGUGGGCUGCU UGCAUCAGCCUAAUGUCGAGAAGUGCUUUCUUC GGAAAGUAACCCUCGAAACAAAUUCAGUGCUCC UCUCCAAUUCUGCACAA	95
MS1/MS5-3	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA UUAGUUGAGUGAAGGUGGGCUGCUUGCAUCAGC CUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAA CCCUCGAAACAAAUUCAGUGCUCCUCUCCAAU CUGCACAA	96
MS1/MS3-3/ MS4*-1	ACCGCUUCACCAAAGCUGUCCCUUAGGGGAUU AGAACUUGAGUGAAGGUGGGCUGCUUGCAUCAG CCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUA ACCCUCGAAACAAAUUCAGUGCUCCUCUCCAAU UC	97
MS1/MS3-3/ MS4*-2	ACCGCUUCACCAAAGCUGUCCCUUAGGGGAUU AGAACUUGAGUGAAGGUGGGCUGCUUGCAUCAG CCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUA ACCCUCGAAACAAAUUCAGUGCUCCUCUC	98
MS1/MS3-3/ MS4*-3	ACCGCUUCACCAAAGCUGUCCCUUAGGGGAUU AGAACUUGAGUGAAGGUGGGCUGCUUGCAUCAG CCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUA ACCCUCGAAACAAAUUCAGUGCU	99
MS1/MS4*-2/ MS5-1	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUGUUUAGAUUAGAACUUGAGUGAAGGUG GGCUGCUUGCAUCAGCCUAAUGUCGAGAAGUGC	100

	UUUCUUCGGAAAGUAACCCUCGAAACAAAUUCA GUGCUCUCUC	
MS1/MS4*-2/ MS5-2	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUUAGGAACUUGAGUGAAGGUGGGCUGCU UGCAUCAGCCUAAUGUCGAGAAGUGCUUUCUUC GGAAAGUAACCCUCGAAACAAAUUCAGUGCUC UCUC	101
MS1/MS4*-2/ MS5-3	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA UUAGUUGAGUGAAGGUGGGCUGCUUGCAUCAGC CUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAA CCUCGAAACAAAUUCAGUGCUCUCUC	102
MS1/MS3-3/ MS5-1	ACCGCUUCACCAAAGCUGUUUAGAUUAGAACU UGAGUGAAGGUGGGCUGCUUGCAUCAGCCUAAU GUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUC GAAACAAAUUCAGUGCUCUCUCCAAUUCUGCA CAA	103
MS1/MS3-3/ MS5-2	ACCGCUUCACCAAAGCUGUAGGAACUUGAGUGA AGGUGGGCUGCUUGCAUCAGCCUAAUGUCGAGA AGUGCUUUCUUCGGAAAGUAACCCUCGAAACAA AUUCAGUGCUCUCUCCAAUUCUGCACAA	104
MS1/MS3-3/ MS5-3	ACCGCUUCACCAAUAGUUGAGUGAAGGUGGGC UGCUUGCAUCAGCCUAAUGUCGAGAAGUGCUUU CUUCGGAAAGUAACCCUCGAAACAAAUUCAGUG CUCCUCUCCAAUUCUGCACAA	105
MS1/MS3-3/ MS4*-2/MS5- 3	ACCGCUUCACCAAUAGUUGAGUGAAGGUGGGC UGCUUGCAUCAGCCUAAUGUCGAGAAGUGCUUU CUUCGGAAAGUAACCCUCGAAACAAAUUCAGUG CUCCUCUC	106
Mature Form(MF)	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUGUCCCUUAGGGGAUUAGAACUUGAGUGA AGGUGGGCUGCUUGCAUCAGCCUAAUGUCGAGA AGUGCUUUCUUCGGAAAGUAACCCUCGAAACAA AUUCAUUU	107
MF/MS3-1	GAUAAAGUGGAGAACCGCUUCACCAAAGCUGU CCCUUAGGGGAUUAGAACUUGAGUGAAGGUGG GCUGCUUGCAUCAGCCUAAUGUCGAGAAGUGCU	108

	UUCUUCGGAAAGUAACCCUCGAAACAAAUUCAU UU	
MF/MS3-2	UGGAGAACCGCUUCACCAAAAGCUGUCCCUUAG GGGAUUAGAACUUGAGUGAAGGUGGGCUGCUU GCAUCAGCCUAAUGUCGAGAAGUGCUUUCUUCG GAAAGUAACCCUCGAAACAAAUUCAUUU	109
MF/MS3-3	ACCGCUUCACCAAAAGCUGUCCCUUAGGGGAUU AGAACUUGAGUGAAGGUGGGCUGCUUGCAUCAG CCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUA ACCCUCGAAACAAAUUCAUUU	110
MF/MS4-1	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUGUCCCUUAGGGGAUUAGAACUUGAGUGA AGGUGGGCUGCUUGCAUCAGCCUAAUGUCGAGA AGUGCUUUCUUCGGAAAGUAACCCUCGAAACAA AUUCAU	111
MF/MS4-2	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUGUCCCUUAGGGGAUUAGAACUUGAGUGA AGGUGGGCUGCUUGCAUCAGCCUAAUGUCGAGA AGUGCUUUCUUCGGAAAGUAACCCUCGAAACAA AUUC	112
MF/MS4-3	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUGUCCCUUAGGGGAUUAGAACUUGAGUGA AGGUGGGCUGCUUGCAUCAGCCUAAUGUCGAGA AGUGCUUUCUUCGGAAAGUAACCCUCGAAACAA A	113
MF/MS5-1	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUGUUUAGAUUAGAACUUGAGUGAAGGUG GGCUGCUUGCAUCAGCCUAAUGUCGAGAAGUGC UUUCUUCGGAAAGUAACCCUCGAAACAAAUUCA UUU	114
MF/MS5-2	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUUUAGAGAACUUGAGUGAAGGUGGGCUG CUUGCAUCAGCCUAAUGUCGAGAAGUGCUUUCU UCGGAAAGUAACCCUCGAAACAAAUUCAUUU	115
MF/MS5-3	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA UUAGUUGAGUGAAGGUGGGCUGCUUGCAUCAGC	116

	CUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAA CCCUCGAAACAAAUUCAUUU	
MF/MS5	CUUCACUGAUAAAGUGGAGAACCGCUUCACUUA GAGUGAAGGUGGGCUGCUUGCAUCAGCCUAAUG UCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCG AAACAAAUUCAUUU	117
MF/MS3-3/M S4-1	ACCGCUUCACCAAAGCUGUCCCUUAGGGGAUU AGAACUUGAGUGAAGGUGGGCUGCUUGCAUCAG CCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUA ACCCUCGAAACAAAUUCAU	118
MF/MS3-3/M S4-2	ACCGCUUCACCAAAGCUGUCCCUUAGGGGAUU AGAACUUGAGUGAAGGUGGGCUGCUUGCAUCAG CCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUA ACCCUCGAAACAAAUUC	119
MF/MS3-3/M S4-3	ACCGCUUCACCAAAGCUGUCCCUUAGGGGAUU AGAACUUGAGUGAAGGUGGGCUGCUUGCAUCAG CCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUA ACCCUCGAAACAAA	120
MF/MS4-3/M S5-1	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUGUUUAGAUUAGAACUUGAGUGAAGGUG GGCUGCUUGCAUCAGCCUAAUGUCGAGAAGUGC UUUCUUCGGAAAGUAACCCUCGAAACAAA	121
MF/MS4-3/M S5-2	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUUUAGAGAACUUGAGUGAAGGUGGGCUG CUUGCAUCAGCCUAAUGUCGAGAAGUGCUUUCU UCGGAAAGUAACCCUCGAAACAAA	122
MF/MS4-3/M S5-3	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA UUAGUUGAGUGAAGGUGGGCUGCUUGCAUCAGC CUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAA CCCUCGAAACAAA	123
MF/MS4-3/M S5	CUUCACUGAUAAAGUGGAGAACCGCUUCACUUA GAGUGAAGGUGGGCUGCUUGCAUCAGCCUAAUG UCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCG AAACAAA	124

MF/MS3-3/M S5-1	ACCGCUUCACCAAAAGCUGUUUAGAUUAGAACU UGAGUGAAGGUGGGCUGCUUGCAUCAGCCUAAU GUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUC GAAACAAAUUCAUUU	125
MF/MS3-3/M S5-2	ACCGCUUCACCAAAAGCUUUAGAGAACUUGAGU GAAGGUGGGCUGCUUGCAUCAGCCUAAUGUCGA GAAGUGCUUUCUUCGGAAAGUAACCCUCGAAAC AAAUUCAUUU	126
MF/MS3-3/M S5-3	ACCGCUUCACCAAUUAGUUGAGUGAAGGUGGGC UGCUUGCAUCAGCCUAAUGUCGAGAAGUGCUUU CUUCGGAAAGUAACCCUCGAAACAAAUUCAUUU	127
MF/MS3-3/M S5	ACCGCUUCACUAGAGUGAAGGUGGGCUGCUUG CAUCAGCCUAAUGUCGAGAAGUGCUUUCUUCGG AAAGUAACCCUCGAAACAAAUUCAUUU	128
MF/MS3-3/M S4-3/MS5-3	ACCGCUUCACCAAUUAGUUGAGUGAAGGUGGGC UGCUUGCAUCAGCCUAAUGUCGAGAAGUGCUUU CUUCGGAAAGUAACCCUCGAAACAAA	129
MF/MS3-3/M S4-1/MS5	ACCGCUUCACUAGAGUGAAGGUGGGCUGCUUG CAUCAGCCUAAUGUCGAGAAGUGCUUUCUUCGG AAAGUAACCCUCGAAACAAAUUCAU	130
MF/MS3-3/M S4-2/MS5	ACCGCUUCACUAGAGUGAAGGUGGGCUGCUUG CAUCAGCCUAAUGUCGAGAAGUGCUUUCUUCGG AAAGUAACCCUCGAAACAAAUUC	131
MF/MS3-3/M S4-3/MS5	ACCGCUUCACUAGAGUGAAGGUGGGCUGCUUG CAUCAGCCUAAUGUCGAGAAGUGCUUUCUUCGG AAAGUAACCCUCGAAACAAA	132

[274] 또한, 상기 복수의 변형부위(MS)에서의 변형이 적용된 crRNA의 예시로서 서열번호 133 내지 서열번호 148의 뉴클레오티드 서열 포함하는 엔지니어링된 crRNA가 제공된다.

[275] 구체적으로, 본 발명의 엔지니어링된 crRNA는 서열번호 133(MS1), 서열번호 134(MS1/MS4^{*}-1), 서열번호 135(MS1/MS4^{*}-2), 서열번호 136(MS1/MS4^{*}-3), 서열번호 137(mature form; MF), 서열번호 138(MF/MS4-1), 서열번호 139(MF/MS4-2), 서열번호 140(MF/MS4-3), 서열번호 141(MS1/MS2), 서열번호 142(MS1/MS2/MS4^{*}-1), 서열번호 143(MS1/MS2/MS4^{*}-2), 서열번호 144(MS1/MS2/MS4^{*}-3), 서열번호 145(MF/MS2), 서열번호 146(MF/MS2/MS4-1), 서열번호 147(MF/MS2/

[277] [卅3]

crRNA	염기서열	서열 번호
MS1	GUUGCAGAACCCGAAUAGAGCAAUGAAGGAAUGCAAC	133
MS1/MS4 [*] -1	GAACCCGAAUAGAGCAAUGAAGGAAUGCAAC	134
MS1/MS4 [*] -2	GAAUAGAGCAAUGAAGGAAUGCAAC	135
MS1/MS4 [*] -3	AGCAAUGAAGGAAUGCAAC	136
MF	GAAUGAAGGAAUGCAAC	137
MF/MS4-1	AUGAAGGAAUGCAAC	138
MF/MS4-2	GAAGGAAUGCAAC	139
MF/MS4-3	GGAAUGCAAC	140
MS1/MS2	GUUGCAGAACCCGAAUAGAGCAAUGAAGGAAUGCAACNNNNNNNNNNNNNNNNNNNNUUUUUUUUUUUU	141
MS1/MS2/MS4 [*] -1	GAACCCGAAUAGAGCAAUGAAGGAAUGCAACNNNNNNNNNNNNNNNNNNNNUUUUUAUUUUUUU	142
MS1/MS2/MS4 [*] -2	GAAUAGAGCAAUGAAGGAAUGCAACNNNNNNNNNNNNNNNNNNNNUUUUUAUUUUUUU	143
MS1/MS2/MS4 [*] -3	AGCAAUGAAGGAAUGCAACNNNNNNNNNNNNNNNNNNNNNNNNNNNUUUUAUUUUUUU	144
MF/MS2	GAAUGAAGGAAUGCAACNNNNNNNNNNNNNNNNNNNNNNNNNNNUUUUAUUUUUUU	145
MF/MS2/MS4-1	AUGAAGGAAUGCAACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNUUUUAUUUUUUU	146
MF/MS2/MS4-2	GAAGGAAUGCAACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNUUUUAUUUUUUU	147

MF/MS2/MS4 -3	GGAAUGCAACNNNNNNNNNNNNNNNNNNNUU UUAUUUUUU	148
------------------	--	-----

- [278] 상기 표 3에서, 필요한 경우를 제외하고 모든 crRNA 서열에서 가이드 서열(스페이서)의 표시를 생략하였으며, 'NNNNNNNNNNNNNNNNNNNNNN'으로 표시된 서열은 표적 유전자 내의 표적 서열과 혼성화할 수 있는 임의의 가이드 서열(스페이서)을 의미한다. 상기 가이드 서열은, 상술한 바와 같이 목적하는 표적 유전자 및/또는 상기 표적 유전자 내 표적 서열에 따라 당업자에 의해 적절하게 설계될 수 있으며, 따라서 특정 길이의 특정 서열로 한정되는 것은 아니다.
- [279] 다른 구현예로, 엔지니어링된 gRNA는 서열번호 87 내지 서열번호 132로 이루어진 군에서 선택된 어느 하나의 뉴클레오티드 서열을 포함하거나 이로 이루어진 tracrRNA; 및 서열번호 133 내지 서열번호 148로 이루어진 군에서 선택된 어느 하나의 뉴클레오티드 서열을 포함하거나 이로 이루어진 crRNA를 포함하는 것일 수 있다.
- [280] 또 다른 구현예로, 본 발명의 엔지니어링된 gRNA가 싱글 가이드 RNA(sgRNA) 형태인 경우, 상기 엔지니어링된 sgRNA는 서열번호 149 내지 서열번호 186으로 이루어진 군에서 선택된 어느 하나의 뉴클레오티드 서열을 포함하거나 이로 이루어진 sgRNA일 수 있다.
- [281] 구체적으로, 상기 엔지니어링된 sgRNA는 MS1에서의 변형을 포함하는 서열번호 149의 sgRNA, MS1/MS2에서의 변형을 포함하는 서열번호 150의 sgRNA, MS1/MS2/MS3에서의 변형을 포함하는 서열번호 151의 sgRNA, MS2/MS3/MS4에서의 변형을 포함하는 서열번호 152의 sgRNA 또는 MS2/MS3/MS4/MS5에서 변형을 포함하는 서열번호 153의 sgRNA일 수 있다.
- [282] 또 다른 구체예로, 상기 엔지니어링된 sgRNA는 서열번호 154(MS1/MS3-1), 서열번호 155(MS1/MS3-2), 서열번호 156(MS1/MS3-3), 서열번호 157(MS1/MS4*-1), 서열번호 158(MS1/MS4*-2), 서열번호 159(MS1/MS4*-3), 서열번호 160(MS1/MS5-1), 서열번호 161(MS1/MS5-2), 서열번호 162(MS1/MS5-3), 서열번호 163(MS1/MS2/MS4*-2), 서열번호 164(MS1/MS3-3/MS4*-2), 서열번호 165(MS1/MS2/MS5-3), 서열번호 166(MS1/MS3-3/MS5-3), 서열번호 167(MS1/MS4*-2/MS5-3), 서열번호 168(MS1/MS2/MS3-3/MS4*-2), 서열번호 169(MS1/MS2/MS3-3/MS5-3), 서열번호 170(MS1/MS2/MS4*-2/MS5-3), 서열번호 171(MS1/MS3-3/MS4*-2/MS5-3) 또는 서열번호 172(MS1/MS2/MS3-3/MS4*-2/MS5-3)의 뉴클레오티드 서열을 포함하거나 이로 이루어진 sgRNA일 수 있다.
- [283] 또한, 상기 sgRNA는 발달된 형태(mature form, MF로 약칭됨)의 sgRNA인 서열번호 173의 뉴클레오티드 서열을 포함하거나 이로 이루어진 sgRNA일 수 있다.
- [284] 다른 구체예로, 상기 MF sgRNA에서 핵산 서열의 일부 변형을 포함하는 예시적인 sgRNA가 제공된다. 구체적으로, 상기 MF sgRNA는 서열번호 174(MS3-1), 서열번호 175(MS3-2), 서열번호 176(MS3-3), 서열번호 177(MS4-1), 서열번호

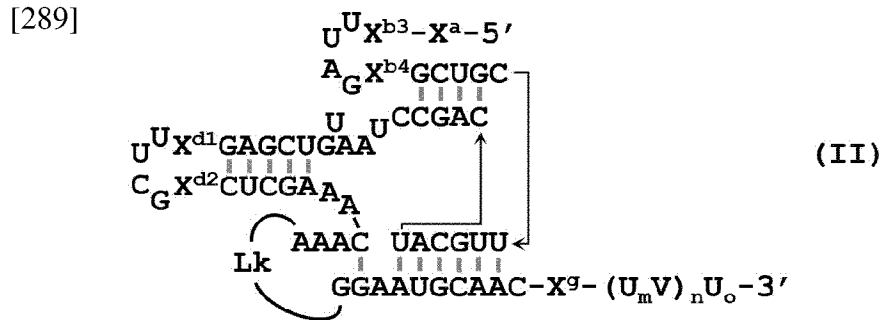
178(MS4-2), 서열번호 179(MS4-3), 서열번호 180(MS5-1), 서열번호 181(MS5-2), 서열번호 182(MS5-3), 서열번호 183(MS3-3/MS4-3), 서열번호 184(MS3-3/MS5-3), 서열번호 185(MS4-3/MS5-3) 또는 서열번호 186(MS3-3/MS4-3/MS5-3) 뉴클레오티드 서열을 포함하거나 이로 이루어진 sgRNA일 수 있다.

[285] 바람직한 구현예로, 엔지니어링된 sgRNA는 서열번호 150(ge3.0), 서열번호 151(ge4.0) 또는 서열번호 152(ge4.1)의 뉴클레오티드 서열로 이루어진 것일 수 있다.

[286] (6) 추가적인 변형의 예시

[287] 본 발명의 다른 태양에 따르면, 상술한 MS1 내지 MS5에서의 변형에 더하여 추가적인 gRNA의 변형이 적용될 수 있다.

[288] 일부 구현예에서, 엔지니어링된 가이드 RNA는 하기 식 (II)로 표시되는 서열로 이루어지거나 상기 서열과 90% 이상 서열 동일성을 갖는 것일 수 있다.



[290] 상기 식 (II)에서, X^a , X^{b3} , X^{b4} , X^{d1} 및 X^{d2} 는 각각 독립적으로 0 내지 35개의 (폴리) 뉴클레오티드로 이루어지고, X^g 는 10 내지 30개의 폴리뉴클레오티드로 이루어진 가이드 서열로서 표적 서열과 혼성화하거나 표적 서열에 상보적인 서열이고, Lk는 길이 2 내지 20의 폴리뉴클레오티드 링커이거나 부존재하고, $(U_m V)_n U_o$ 는 U-rich tail로서 존재하거나 부존재하고, 존재하는 경우 U는 유리딘이고, V는 각각 독립적으로 A, C 또는 G이고, m 및 o는 각각 독립적으로 1 내지 20 사이의 정수이며, n은 0 내지 5 사이의 정수이다.

[291] 일부 구현예에서, X^{b3} 은 서열 5'-ACCGCUUCAC-3'을 포함하거나 이로 이루어질 수 있다. 또한, X^{b3} 은 상기 서열에서 1개 내지 9개의 임의의 뉴클레오티드가 결실된 서열을 포함하거나 이로 이루어진 것일 수 있고, 또는 X^{b3} 은 부존재할 수 있다.

[292] 다른 구현예에서, X^{b4} 는 서열 5'-AGUGAAGGUGG-3'을 포함하거나 이로 이루어질 수 있다. 또한, X^{b4} 는 상기 서열에서 1개 내지 10개의 임의의 뉴클레오티드가 결실된 서열을 포함하거나 이로 이루어진 것일 수 있고, 또는 X^{b4} 는 부존재할 수 있다.

[293] 또 다른 구현예에서, X^{d1} 은 서열 5'-AAGUGCUUUC-3'을 포함하거나 이로 이루어질 수 있다. 또한, X^{d1} 은 상기 서열에서 1개 내지 9개의 임의의 뉴클레오티드가 결실된 것일 수 있고, 또는 X^{d1} 은 부존재할 수 있다.

- [294] 또 다른 구현예에서, X^{d2} 는 서열 5'-GAAAGUAACC-3'을 포함하거나 이로 이루어질 수 있다. 또한, X^{d2} 는 상기 서열에서 1개 내지 9개의 임의의 뉴클레오티드가 결실된 것일 수 있고, 또는 X^{d2} 는 부존재할 수 있다.
- [295] X^a , X^g , Lk 및 $(U_m V)_n U_o$ 에 대한 구체적인 내용은 상술한 바를 참조한다.
- [296] **(7) 화학적 변형(Chemical modification)**
- [297] 일부 구현예에서, 상기 엔지니어링된 gRNA에 포함되는 엔지니어링된 tracrRNA 또는 엔지니어링된 crRNA는 필요에 따라 적어도 하나 이상의 뉴클레오티드가 화학적 변형을 가질 수 있다. 이때, 상기 화학적 변형은 뉴클레오티드의 염기 및/또는 당에서 발생할 수 있는 다양한 공유 결합의 변형일 수 있다.
- [298] 일 예로, 상기 화학적 변형은 메틸화(methylation), 할로젠화(halogenation), 아세틸화(acetylation), 인산화(phosphorylation), PS(phosphorothioate) 연결, LNA(locked nucleic acid), 2'-O-methyl 3'phosphorothioate(MS) 또는 2'-O-methyl 3'thioPACE(MSP)일 수 있다. 상기 예시는 단순 예시로, 이에 제한되지 않는다.
- [299] 본 발명의 엔지니어링된 gRNA 및 Cas12f1, TnpB 또는 이의 변이체 단백질 복합체를 포함하는 초소형 핵산 편집 시스템을 사용하는 경우, 자연계에서 발견되는 가이드 RNA를 사용하는 경우에 비해 세포 내에서 표적 핵산 또는 표적 유전자의 절단 및/또는 상동지정복구 효율이 현저하게 향상되는 효과가 나타난다.
- [300] 무엇보다도 상기 엔지니어링된 gRNA는 고효율을 나타내는 길이의 최적화와 이에 따른 gRNA 합성 비용 절감, 바이러스 벡터에 삽입하는 경우에 추가 공간 또는 용량 확보, tracrRNA의 정상적인 발현, 작동가능한 gRNA 발현의 증가, gRNA의 안정성(stability) 증가, gRNA와 핵산 편집 단백질 복합체의 안정성 증가, 고효율의 gRNA 및 핵산 편집 단백질 복합체 형성 유도, gRNA 및 핵산 편집 단백질 복합체를 포함하는 초소형 표적 핵산 편집 시스템에 의한 표적 핵산의 절단 효율 증가 및 상기 시스템에 의한 표적 핵산의 상동지정복구 효율 증가 효과를 수반할 수 있다. 이에 따라, Cas12f1, TnpB 또는 이의 변이체 단백질에 대해 상술한 엔지니어링된 gRNA를 사용하면 전술한 종래 기술의 한계점을 극복하여 세포 내에서 높은 효율로 유전자를 절단 또는 편집할 수 있다.
- [301] 또한, 엔지니어링된 gRNA는 자연계에서 발견되는 gRNA와 비교하여 짧은 길이를 가지므로 유전자 편집 기술 분야에서 그 응용 가능성이 높다. 상기 엔지니어링된 gRNA를 사용하면 gRNA 및 핵산 편집 단백질 복합체를 포함하는 초소형 핵산 편집 시스템의 크기가 매우 작고, 편집 효율이 우수하다는 장점은 다양한 유전자 편집 기술에 활용될 수 있게 된다.
- [302] **3. 공여자 핵산 분자**
- [303] 상기 공여자 핵산 분자는 상동지정복구에서 주형으로 사용되는 서열이다. 공여자 핵산 분자는 최소한 하나의 공여자 서열을 포함한다. 일부 구현예에서, 공여자 핵산 분자의 목적 서열은 내인성 또는 선천적 염색체 서열에 상응할 수 있다. 목적 서열은 표적 부위(표적 서열)에서 또는 이의 근접한 부위와 본질적으로 동

일할 수 있지만, 최소한 하나의 뉴클레오티드 변화를 포함할 수 있다. 따라서, 목적 서열은 선천적 서열과의 통합 또는 교환 시에, 표적화된 위치에서 서열이 최소한 하나의 뉴클레오티드 변화를 포함하도록, 표적화된 부위에서 야생형 서열의 변형된 이형을 포함할 수 있다. 가령, 변화는 하나 또는 그 이상의 뉴클레오티드의 삽입, 하나 또는 그 이상의 뉴클레오티드의 결실, 하나 또는 그 이상의 뉴클레오티드의 치환, 또는 이들의 조합일 수 있다. 변형된 서열의 통합의 결과로서, 세포 또는 배아/동물은 표적화된 서열로부터 변형된 유전자 산물을 생산할 수 있다.

[304] 다른 구체예에서, 공여자 핵산 분자의 목적 서열은 외인성 서열에 상응한다. 본 발명의 실시예에서 이용된 바와 같이, "외인성" 서열은 세포 또는 배아에 선천적이지 않은 서열, 또는 세포의 유전체에서 선천적 위치가 상이한 위치에 있는 서열을 지칭한다. 가령, 외인성 서열은 유전체 내로 통합 시에 세포가 통합된 서열에 의해 코딩된 단백질을 발현할 수 있도록 외인성 프로모터 제어 서열에 작동 가능하게 연결될 수 있는 단백질 코딩 서열을 포함할 수 있다. 대안으로, 외인성 서열은 이의 발현이 내인성 프로모터 제어 서열에 의해 조절되도록 염색체 서열 내로 통합될 수 있다. 다른 반복에서, 외인성 서열은 전사 제어 서열, 다른 발현 제어 서열, RNA 코딩 서열, 기타 등등일 수 있다. 염색체 서열 내로 외인성 서열의 통합은 "녹인(knock-in)"으로 명명된다.

[305] 당업자에 의해 인지될 수 있는 바와 같이, 공여자 핵산 분자 또는 목적 서열의 길이는 변할 수 있고 변할 것이다. 가령, 공여자 핵산 분자 또는 목적 서열의 길이에서 여러 뉴클레오티드에서부터 수백 개의 뉴클레오티드 내지 수십만 개의 뉴클레오티드까지 변할 수 있다. 예를 들면, 공여자 핵산 분자 또는 목적 서열의 길이는 적어도 1 bp, 10 bp, 100 bp, 또는 200 bp 이상 일 수 있다. 예를 들면, 공여자 핵산 분자의 길이는 1 bp 내지 20kb, 100 bp 내지 50 kb, 100 bp 내지 30 kb, 100 bp 내지 10 kb, 100 bp 내지 10 kb, 200 bp 내지 50 kb, 200 bp 내지 30 kb, 또는 200 bp 내지 10 kb 일 수 있다. 일 구체예에서, 공여자 핵산 분자 또는 목적 서열의 길이는 1 bp 내지 20 kb, 1 bp 내지 10 kb, 1 bp 내지 8 kb, 1 bp 내지 6 kb, 또는 1 bp 내지 4 kb 일 수 있다.

[306] 공여자 핵산 분자는 상류(upstream)와 하류(downstream) 서열을 포함할 수 있다. 일부 구체예에서, 공여자 핵산 분자 내에 목적 서열은 각각, 염색체 서열 내에 표적화된 부위의 상류와 하류에 위치된 서열에 실제적인 서열 동일성을 갖는 상류 서열 및 하류 서열과 측면에서 접한다. 이들 서열 유사성 때문에, 공여자 핵산 분자의 상류와 하류 서열은 목적 서열이 염색체 서열 내로 통합(또는 이것과 교환)될 수 있도록, 공여자 핵산 분자와 표적화된 염색체 서열 사이에 상동성 재조합을 허용한다.

[307] 상류 서열은 본원에서 이용된 바와 같이, 표적화된 부위의 상류에 염색체 서열과 실제적인 서열 동일성을 공유하는 핵산 서열을 지칭한다. 유사하게, 하류 서열은 표적화된 부위의 하류에 염색체 서열과 실제적인 서열 동일성을 공유하는

핵산 서열을 지칭한다. 본 발명에서 이용된 바와 같이, 관용구 "실제적인 서열 동일성"은 최소한 약 75% 서열 동일성을 갖는 서열을 지칭한다. 따라서, 공여자 핵산 분자에서 상류와 하류 서열은 표적화된 부위의 상류 또는 하류에 서열과 약 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 가질 수 있다. 예시적인 구체예에서, 공여자 핵산 분자에서 상류와 하류 서열은 표적화된 부위의 상류 또는 하류에 염색체 서열과 약 95% 또는 100% 서열 동일성을 가질 수 있다. 일 구체예에서, 상류 서열은 표적화된 부위의 즉시 상류에 위치한 (즉, 표적화된 부위에 인접한) 염색체 서열과 실제적인 서열 동일성을 공유한다. 다른 구체예에서, 상류 서열은 표적화된 부위로부터 상류에 약 100개 뉴클레오타이드 내에 위치되는 염색체 서열과 실제적인 서열 동일성을 공유한다. 따라서, 예로서, 상류 서열은 표적화된 부위로부터 상류에 약 1 내지 약 20, 약 21 내지 약 40, 약 41 내지 약 60, 약 61 내지 약 80, 또는 약 81 내지 약 100개 뉴클레오타이드 내에 위치되는 염색체 서열과 실제적인 서열 동일성을 공유할 수 있다. 한 구체예에서, 하류 서열은 표적화된 부위의 즉시 하류에 위치한 (즉, 표적화된 부위에 인접한) 염색체 서열과 실제적인 서열 동일성을 공유한다. 다른 구체예에서, 하류 서열은 표적화된 부위로부터 하류에 약 100개 뉴클레오타이드 내에 위치되는 염색체 서열과 실제적인 서열 동일성을 공유한다. 따라서, 예로서, 하류 서열은 표적화된 부위로부터 하류에 약 1 내지 약 20, 약 21 내지 약 40, 약 41 내지 약 60, 약 61 내지 약 80, 또는 약 81 내지 약 100개 뉴클레오타이드 내에 위치되는 염색체 서열과 실제적인 서열 동일성을 공유할 수 있다.

[308] 각 상류 또는 하류 서열은 길이에서 약 20개 뉴클레오타이드 내지 약 5000개 뉴클레오타이드 범위에서 변할 수 있다. 일부 구체예에서, 상류와 하류 서열은 약 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2800, 3000, 3200, 3400, 3600, 3800, 4000, 4200, 4400, 4600, 4800, 또는 5000개 뉴클레오타이드를 포함할 수 있다. 예시적인 구체예에서, 상류와 하류 서열은 길이에서 약 50 내지 약 1500개 뉴클레오타이드 범위에서 변할 수 있다.

[309] 표적화된 염색체 서열에 서열 유사성을 갖는 상류와 하류 서열을 포함하는 공여자 핵산 분자는 선형 또는 환상일 수 있다. 공여자 핵산 분자가 환상인 구체예에서, 이것은 벡터의 부분일 수 있다. 가령, 벡터는 플라스미드 벡터일 수 있다.

[310] 공여자 핵산 분자는 표적화된 개월 부위를 포함할 수 있다. 다른 구체예에서, 공여자 핵산 분자는 RNA-유도된 엔도뉴클레아제에 의해 인식되는 최소한 하나의 표적화된 개월 부위를 부가적으로 포함할 수 있다. 공여자 핵산 분자에 부가된 표적화된 개월 부위는 목적 서열의 상류 또는 하류 또는 상류와 하류 둘 모두에 배치될 수 있다. 가령, 목적 서열은 RNA-유도된 엔도뉴클레아제에 의한 개월 시에, 목적 서열이 RNA-유도된 엔도뉴클레아제에 의한 개월 시에 산출된 염색체 서열 내에 것들과 양립하는 오버행과 측면에서 접하도록, 표적화된 개월 부위

와 측면에서 접할 수 있다. 따라서, 목적 서열은 비상동성 복구 과정에 의한 이중 가닥 절단의 복구 동안 개열된 염색체 서열과 결합될 수 있다. 일반적으로, 표적화된 개열 부위를 포함하는 공여자 핵산 분자는 환상일 것이다(가령, 플라스미드 벡터의 부분일 수 있다).

- [311] 공여자 핵산 분자는 임의선택적 오버행(overhang)을 갖는 짧은 목적 서열을 포함할 수 있다. 다른 대안적 구체예에서, 공여자 핵산 분자는 RNA-유도된 엔도뉴클레아제에 의해 산출된 오버행과 양립하는 임의선택적 짧은 오버행을 갖는 짧은 목적 서열을 포함하는 선형 분자일 수 있다. 이런 구체예에서, 목적 서열은 이중 가닥 절단의 복구 동안 개열된 염색체 서열과 직접적으로 결합될 수 있다. 일부 경우에, 목적 서열은 약 1,000개보다 적거나, 약 500개보다 적거나, 약 250개보다 적거나, 또는 약 100개보다 적은 뉴클레오티드일 수 있다. 일정한 경우에, 공여자 핵산 분자는 평활 말단을 갖는 짧은 공여자 서열을 포함하는 선형 분자일 수 있다. 다른 반복에서, 공여자 핵산 분자는 5' 및/또는 3' 오버행을 갖는 짧은 목적 서열을 포함하는 선형 분자일 수 있다. 오버행은 1, 2, 3, 4, 또는 5개 뉴클레오티드를 포함할 수 있다.
- [312] 전형적으로, 공여자 핵산 분자는 DNA일 것이다. DNA는 단일 가닥 또는 이중 가닥 및/또는 선형 또는 환상일 수 있다. 공여자 핵산 분자는 DNA 플라스미드, 세균 인공 염색체 (BAC), 효모 인공 염색체 (YAC), 바이러스 벡터, DNA의 선형 조각, PCR 단편, 나신 핵산, 또는 전달 운반체, 예를 들면, 리포솜 또는 폴록사머로 복합화된 핵산일 수 있다. 일 구체예에서, 목적 서열을 포함하는 공여자 핵산 분자는 플라스미드 벡터의 부분일 수 있다. 또한, 일 구체예에서, 목적 서열을 포함하는 공여자 핵산 분자는 최소한 하나의 추가 서열을 더욱 포함할 수 있다.
- [313] **4. 비상동말단연결 과정을 억제하는 분자**
- [314] 본 발명의 표적 핵산 편집 시스템은 상술한 핵산 편집 단백질(엔도뉴클레아제), 가이드 RNA 및 공여자 핵산 분자 외에도, 상동지정복구 효율 향상을 위한 다양한 종류의 분자를 추가로 포함할 수 있다.
- [315] 일부 구현예에서, 표적 핵산 편집 시스템은 비상동말단연결(non-homologous end joining, NHEJ)에 관여하는 유전자의 발현을 억제하는 분자를 더 포함할 수 있다. 구체적으로, 상기 시스템은 비상동말단연결에 관여하는 유전자 또는 단백질 등의 활성을 감소시킬 수 있는 인자, 예를 들면, 비상동말단연결에 관여하는 유전자의 발현을 억제하는 분자를 더 포함하는 것일 수 있다. 임의의 특정 이론에 구속됨이 없이, 예를 들면, 비상동말단연결 활성 감소는 상동지정복구-매개된 경로의 촉진을 일으킬 수 있다. 상기 억제제는 비상동말단연결 활성을 감소시키는데 혹은 상동지정복구 활성을 증가 또는 감소시키는데 사용될 수 있다. 이와 같은 억제제는, 예를 들어, 작은 분자 또는 억제성 핵산 예컨대 짧은 간섭 핵산(예를 들면, 짧은 간섭 RNA(siRNA), 이중-가닥 RNA(dsRNA), 마이크로-RNA(miRNA), 및 유전자 전사체에 특이적인 짧은 헤어핀 RNA(shRNA)) 또는 안티센스 올리고뉴클레오티드일 수 있다. 상기 억제제는, 예를 들어, 인산화, 유비

퀴틸화, 및 수모화를 통해 번역 후 변형에 의한 비상동말단연결 또는 상동지정복구 또는 그것의 상류 조절에 관여된 효소를 표적화하는 것일 수 있다.

[316] 일 구체예에 있어서, 상기 비상동말단연결에 관여하는 유전자는 ATM1, XRCC4, XLF, XRCC6, LIG4 및 DCLRE1C로 이루어진 군으로부터 선택된 어느 하나 이상일 수 있다.

[317] 본 발명의 일 실시예에 따르면, 비상동말단연결 과정의 억제를 위해 DCLRE1C, LIG4, XRCC4, XRCC6, XLF 및 ATM 유전자에 대한 shRNA를 본 발명의 핵산 편집 시스템에 포함시킨 결과, 상동지정복구 효율이 현저하게 상승하는 것을 확인하였다(실시예 3.1 참조).

[318] **5. 표적 핵산의 편집을 위한 시스템/조성물**

[319] 본 발명의 다른 태양에 따르면, 상술한 표적 핵산 편집 시스템을 포함하는 유전자 편집용 조성물이 제공된다. 또한, 후술한 벡터 시스템 또는 상기 표적 핵산 편집 시스템 및 벡터 시스템을 모두 포함하는 유전자 편집용 조성물이 제공된다.

[320] 일 구현예로, 본원의 유전자 편집(용) 조성물은 Cas12f1, TnpB 또는 이의 변이체(variant) 단백질을 포함하는 엔도뉴클레아제 또는 상기 엔도뉴클레아제를 암호화하는 핵산; 가이드 서열을 포함하는 엔지니어링된 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 핵산; 및 공여자 핵산 분자 또는 상기 공여자 핵산 분자를 암호화하는 핵산을 포함한다.

[321] 상기 "Cas12f1, TnpB 또는 이의 변이체(variant) 단백질", "엔지니어링된 가이드 RNA" 및 "공여자 핵산 분자"에 관한 사항은 상술한 내용을 참조한다.

[322] 다른 구현예로, 표적 핵산 편집 시스템 또는 유전자 편집 조성물에 포함되는 Cas12f1, TnpB 또는 이의 변이체 단백질을 포함하는 엔도뉴클레아제 및 상기 가이드 RNA는 리보뉴클레오단백질 입자(ribonucleoprotein particle, RNP) 형태로 포함될 수 있다.

[323] 한편, 본 발명의 유전자 편집용 조성물은 상기 본 발명에 따른 초소형 핵산 편집 시스템의 각 구성요소 이외에, 유전자 편집 용도에 필요한 적절한 물질을 추가로 포함할 수 있음은 자명하다.

[324] **III. 상동지정복구를 위한 표적 핵산 편집 시스템의 구성요소를 암호화하는 핵산**

[325] 본 발명에서 제공하는 표적 핵산 편집 시스템의 각 구성요소는 세포 내에서 발현되도록 하는 것이므로, 본 발명의 다른 태양에 따르면, 표적 핵산 편집 시스템의 각 구성요소를 암호화하는 핵산 또는 폴리뉴클레오티드가 제공된다.

[326] 구체적으로, 상기 핵산 또는 폴리뉴클레오티드는 발현하고자 하는 표적 핵산 편집 시스템에 포함된 핵산 편집 단백질, 가이드 RNA 및/또는 공여자 핵산 분자를 암호화하는 핵산 서열을 포함한다. 이때, 상기 핵산 또는 폴리뉴클레오티드의 서열은 야생형의 유전자 편집 단백질 및 야생형의 가이드 RNA를 암호화하는 핵산 서열뿐만 아니라, 그 목적에 따라 엔지니어링된 가이드 RNA 및/또는 코돈 최적화된 핵산 편집 단백질을 암호화하는 핵산 서열, 엔지니어링된 핵산 편집 단백

질을 암호화하는 핵산 서열, 또는 DNA 이중가닥 절단 활성이 상실되거나 감소된 유전자 편집 단백질을 암호화하는 핵산 서열을 포함할 수 있다.

[327] 본 발명에서, 상기 핵산 또는 폴리뉴클레오티드는 초소형 핵산 편집 단백질인 Cas12f1, TnpB 또는 이의 변이체 단백질을 발현하도록 구성된 서열을 포함할 수 있다. 여기서, 상기 Cas12f1, TnpB 또는 이의 변이체 단백질은 DNA 이중가닥 또는 단일가닥을 절단하는 활성을 가지는 단백질일 수 있다.

[328] 일 구현예로, 상기 핵산 또는 폴리뉴클레오티드는 Cas12f1, TnpB 또는 이의 변이체 단백질을 발현하도록 구성된 서열을 포함할 수 있다. 여기서, 상기 Cas12f1, TnpB 또는 이의 변이체 단백질은 서열번호 1 내지 서열번호 5로 이루어진 군에서 선택된 아미노산 서열과 70% 이상의 서열 동일성을 갖는 아미노산 서열을 포함하는 단백질일 수 있다. 이에 대한 상세한 내용은 상기를 참조한다.

[329] 또한, 상기 핵산 또는 폴리뉴클레오티드는 Cas12f1, TnpB 또는 이의 변이체 단백질을 암호화하는 서열을 포함할 수 있다. 바람직하게, 상기 핵산 또는 폴리뉴클레오티드는 Cas12f1, TnpB 또는 이의 변이체 단백질을 암호화하는 인간 코돈 최적화된 핵산 서열을 포함할 수 있다. "코돈 최적화"는 고유 서열의 적어도 하나의 코돈을 대상 세포의 유전자에 더욱 빈번하게 또는 가장 빈번하게 사용되는 코돈으로 대체하면서, 고유 아미노산 서열을 유지함으로써 관심 대상 세포에서의 발현의 증진을 위해 핵산서열을 변형시키는 과정을 의미한다. 다양한 종은 특정 아미노산의 특정 코돈에 대한 특정 편향을 가지며, 코돈 편향(유기체 간의 코돈 사용의 차이)은 종종 mRNA의 번역의 효율과 상호관련되며, 이는 번역되는 코돈의 특성 및 특정 tRNA 분자의 이용가능성에 의해 좌우되는 것으로 여겨진다. 세포에서 선택된 tRNA의 우세는 일반적으로 펩티드 합성에 가장 빈번하게 사용되는 코돈을 반영한 것이다. 따라서, 유전자는 코돈 최적화에 기초하여 주어진 유기체에서 최적의 유전자 발현을 위해 맞춤화될 수 있다.

[330] 일 구현예로, 상기 TnpB 단백질을 암호화하는 핵산은 인간 코돈 최적화된 TnpB 단백질을 암호화하는 핵산일 수 있다. 예를 들어, 상기 인간 코돈 최적화된 TnpB 단백질을 암호화하는 핵산은 서열번호 6의 핵산 서열을 포함하는 것일 수 있다. 다른 예로, 상기 인간 코돈 최적화된 Cas12f1 단백질을 암호화하는 핵산은 서열번호 10의 핵산 서열을 포함하는 것일 수 있다. 또 다른 예로, 상기 인간 코돈 최적화된 Cas12f1의 변이체 또는 TnpB의 변이체 단백질을 암호화하는 핵산은 서열번호 7(TnpB-v1), 서열번호 8(TnpB-v2) 또는 서열번호 9(TnpB-v3)의 핵산 서열을 포함하는 것일 수 있다.

[331] 또한, 상기 핵산 또는 폴리뉴클레오티드는 변형된 Cas12f1, TnpB 또는 이의 변이체 단백질 또는 Cas12f1, TnpB 또는 이의 변이체의 융합 단백질을 암호화하는 서열을 포함할 수 있다. 일 구현예로, 상기 핵산 또는 폴리뉴클레오티드는 표적 핵산의 이중가닥 중 하나의 가닥만 절단하도록 변경된 Cas12f1, TnpB 또는 이의 변이체 단백질을 발현하도록 구성된 서열을 포함할 수 있다. 일 예로, 상기 변형된 Cas12f1, TnpB 또는 이의 변이체 단백질은 표적 핵산의 이중가닥 중 하나의 가

다만 절단할 수 있고, 절단하지 않는 가닥에 대해 염기 교정(Base editing) 또는 프라임 교정(Prime editing)을 할 수 있도록 변경된 것일 수 있다. 또는, 상기 핵산 또는 폴리뉴클레오티드는 표적 핵산에 대해 염기 교정(Base editing) 또는 프라임 교정(Prime editing) 또는 유전자 발현 조절 기능을 할 수 있도록 변경된 변이체 단백질을 암호화하는 서열을 포함할 수 있다.

- [332] 또한, 상기 핵산 또는 폴리뉴클레오티드는 Cas12f1, TnpB 또는 이의 변이체에 대해 최적의 표적 효율을 가지도록 엔지니어링된 가이드 RNA(augment RNA)를 발현하도록 구성된 것이거나 또는 하나 또는 둘 이상의 서로 다른 엔지니어링된 가이드 RNA를 발현하도록 구성된 서열을 포함할 수 있다. 일 예로, 상기 엔지니어링된 가이드 RNA 서열은 스캐폴드 서열, 스페이서 서열 및 U-rich tail 서열을 포함할 수 있다. 구체적으로, 상기 엔지니어링된 gRNA 서열은 변경된 tracrRNA 서열 및/또는 변경된 crRNA 서열을 포함하고, 여기에 U-rich tail 서열을 포함할 수 있다. U-rich tail을 포함한 엔지니어링된 gRNA 및 이의 변형에 대한 내용은 상기를 참조한다.
- [333] 또한, 상기 핵산 또는 폴리뉴클레오티드는 상동지정복구 과정에서 주형으로 사용하기 위한 공여자 핵산 분자를 포함할 수 있다. 공여자 핵산 분자에 관한 내용은 상기를 참조한다.
- [334] **IV. 표적 핵산 편집 시스템의 발현을 위한 벡터 시스템**
- [335] 본 발명의 또 다른 태양에 따르면, Cas12f1, TnpB 또는 이의 변이체 단백질을 포함하는 엔도뉴클레아제를 암호화하는 뉴클레오티드 서열이 작동 가능하게 연결된 제1 핵산 구조물; 표적 핵산에 상보적으로 결합하는 가이드 서열을 포함하는 엔지니어링된 가이드 RNA를 암호화하는 뉴클레오티드 서열이 작동 가능하게 연결된 제2 핵산 구조물; 및 공여자 핵산 분자를 포함하는 제3 핵산 구조물을 포함하는 하나 이상의 벡터를 포함하는 벡터 시스템이 제공된다.
- [336] 본 발명에서 제공하는 표적 핵산 편집 시스템을 상동지정복구를 포함하는 유전자 편집에 사용하기 위해서, 상기 표적 핵산 편집 시스템의 각 구성을 암호화하는 서열을 포함하는 벡터를 표적 세포 내로 직접 도입시키거나 바이러스 등의 매개체에 의해 전달하고, 표적 세포 내에서 상기 유전자 편집 시스템의 각 구성이 발현되도록 하는 방법이 이용될 수 있다.
- [337] 또한, 표적 핵산 또는 표적 유전자를 편집하기 위한 본 발명의 핵산 편집 시스템은 우수한 상동지정복구 효율을 달성하기 위해, 가이드 RNA 및 Cas12f1, TnpB 또는 이의 변이체 단백질 복합체의 각 구성성분은 작동가능하게 연결되어 하나의 벡터로 포함되는 것이 바람직하다. 여기서, 핵산 분해 단백질 또는 가이드 분자에는 필요에 따라 효과기 단백질이 연결되어 융합된 형태의 단백질을 이룰 수 있다.
- [338] 일 예로, 상기 융합된 형태의 단백질은 박테리오파지 외피 단백질 내에 존재하는 직교 RNA-결합 단백질 또는 어댑터 단백질을 포함할 수 있다. 여기서, 외피 단백질은 MS2, Q β , F2, GA, fr, JP501, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1,

TW18, VK, SP, FI, ID2, NL95, TW19, AP205, Φ Cb5, Φ Cb8r, Φ Cb12r, Φ Cb23r, 7s 및 PRR1등이 포함될 수 있다. 또한, 상기 융합된 형태의 단백질은 하나 이상의 지질 나노입자를 통해 전달될 수 있다.

- [339] 일 구현예로, 본 발명의 표적 핵산 편집 시스템의 구성요소에 해당하는 초소형 핵산 편집 단백질인 Cas12f1, TnpB 또는 이의 변이체 단백질 및 하나 이상의 가이드 RNA는 이를 암호화하는 하나 이상의 mRNA 분자로서 세포에 전달될 수 있다. 이때 상기 mRNA 분자는 하나 이상의 지질 나노입자를 통해 전달될 수 있다.
- [340] 또한, 본 발명의 표적 핵산 편집 시스템의 구성요소는 하나 이상의 DNA 분자 형태일 수 있다. 여기서, 하나 이상의 DNA 분자는 유전자 편집 단백질 또는 가이드 분자를 발현시키도록 작동가능하게 구성된 하나 이상의 조절 요소를 포함할 수 있다. 필요에 따라 하나 이상의 조절 요소는 유도성 프로모터를 포함할 수 있다.
- [341] 일 구현예에서, 상기 벡터 시스템에 포함되는 핵산 구조물은 동일하거나 상이한 벡터에 위치하는 것일 수 있다.
- [342] 상기 표적 핵산 편집 시스템을 구성하는 DNA 분자들은 하나 이상의 아데노-연관 바이러스(AAV) 벡터 내에 포함되어 세포 내로 전달될 수 있다. 바람직하게, 상기 DNA 분자들은 모두 하나의 아데노-연관 바이러스(AAV) 벡터 내에 포함되어 AAV 바이러스 내에 표적 핵산 편집 시스템을 구성하는 DNA 분자가 패키징된 형태로서 세포 내로 전달될 수 있다.
- [343] 보다 구체적으로, 본 발명의 초소형 유전자 편집 시스템이 세포 내에서 발현되도록 하는 벡터의 구성 요소는 다음을 포함한다.
- [344] **1. 표적 핵산 편집 시스템의 구성요소를 암호화하는 핵산 구조물**
- [345] 상기 벡터 시스템의 최종 목적은 본 발명의 표적 핵산 편집 시스템의 각 구성요소가 세포 내에서 발현되도록 하는 것이므로, 상기 벡터 시스템에 포함되는 서열은 표적 핵산 편집 시스템의 각 구성요소를 암호화하는 핵산 서열 중 하나 이상을 필수적으로 포함해야 한다.
- [346] 일 구현예로, 벡터 시스템은 Cas12f1, TnpB 또는 이의 변이체(variant) 단백질을 포함하는 엔도뉴클레아제를 암호화하는 뉴클레오티드 서열이 작동 가능하게 연결된 제1 핵산 구조물; 표적 핵산에 상보적으로 결합하는 가이드 서열을 포함하는 엔지니어링된 가이드 RNA를 암호화하는 뉴클레오티드 서열이 작동 가능하게 연결된 제2 핵산 구조물; 및 공여자 핵산 분자를 포함하는 제3 핵산 구조물을 포함할 수 있다. 이때, 상기 제1 핵산 구조물, 제2 핵산 구조물 및/또는 제3 핵산 구조물은 벡터 시스템의 동일한 벡터 상에 또는 상이한/별개의 벡터 상에 위치될 수 있다. 여기서, 상기 연결은 직접 또는 링커를 통해 연결될 수 있다.
- [347] 일 구현예로, 상기 핵산 구조물은 엔지니어링된 가이드 RNA(gRNA)를 암호화하는 핵산을 포함할 수 있다. 여기서, 엔지니어링된 gRNA는 엔지니어링된 tracrRNA 및/또는 엔지니어링된 crRNA를 포함할 수 있다. 이때, 상기 엔지니어링된 가이드 RNA는 앞서 설명한 엔지니어링된 가이드 RNA의 구현예와 동일한 구성을 가질 수 있다.

- [348] 또한, 상기 핵산 구조물에서 핵산 편집 단백질 Cas12f1, TnpB 또는 이의 변이체 단백질은 서열번호 1 내지 서열번호 5로 이루어진 군에서 선택된 아미노산 서열과 70% 이상, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 또는 95% 이상의 서열 동일성을 갖는 아미노산 서열을 포함하는 단백질일 수 있으며, 상기 핵산 구조물은 상기 단백질을 암호화하는 핵산 또는 상기 단백질의 코돈-최적화된 핵산을 포함할 수 있다. 일 예로, 초소형 핵산 편집 단백질은 서열번호 1 내지 서열번호 5로 이루어진 군에서 선택된 아미노산 서열과 70% 이상, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 또는 95% 이상의 서열 동일성을 갖는 것을 특징으로 하는 핵산 편집 단백질일 수 있고, 이를 암호화하는 코돈-최적화된 핵산은 인간 코돈-최적화된 핵산으로서, 서열번호 6 내지 서열번호 10 중 선택된 어느 하나의 뉴클레오티드 서열로 이루어진 것일 수 있다.
- [349] 또한, 상기 핵산 구조물은 N-말단 또는 C-말단에 핵 위치 신호(nuclear localization signal, NLS) 또는 핵 유출 신호(nuclear export signal, NES) 서열을 1개 이상 포함하는 것일 수 있다. NLS 서열은 핵 수송(nuclear transport) 작용으로 세포 핵 외부의 물질을 핵 내부로 수송할 때, 수송 대상인 단백질 등에 붙어 일종의 "태그" 역할을 하는 일정 길이의 펩티드 또는 그 서열을 의미한다. NES 서열은 핵 수송(nuclear transport) 작용으로 세포 핵 내부의 물질을 핵 외부로 수송할 때, 수송 대상인 단백질에 붙어 일종의 "태그" 역할을 하는 일정 길이의 펩티드 또는 그 서열을 의미한다. 예를 들어, 상기 NLS는 SV40 바이러스 대형 T-항원의 NLS; c-myc NLS; hRNPA1 M9 NLS; 뉴클레오플라스민(nucleoplasmin)으로부터의 NLS; 임포틴-알파로부터의 IBB 도메인의 서열; 마이오마(myoma) T 단백질의 서열; 인간 p53의 서열; 마우스 c-abl IV의 서열; 인플루엔자 바이러스 NS1의 서열; 간염 바이러스 델타 항원의 서열; 마우스 Mx1 단백질의 서열; 인간 폴리(ADP-리보스) 중합효소의 서열; 또는 스테로이드 호르몬 수용체(인간) 글루코코르티코이드의 서열로부터 유래된 NLS 서열일 수 있으나, 이에 제한되지 않는다.
- [350] 상기 벡터 시스템은 발현하고자 하는 핵산 편집 시스템에 포함된 가이드 RNA 및/또는 핵산 편집 단백질을 암호화하는 핵산 서열을 포함한다. 상기 핵산 서열에 관련된 내용은 상술한 내용을 참조한다.
- [351] 상기 벡터는 서로 다른 둘 이상의 엔지니어링된 가이드 RNA를 발현하도록 구성된 것일 수 있다. 일 구현예로, 상기 벡터는 엔지니어링된 제1 가이드 RNA 및 엔지니어링된 제2 가이드 RNA를 발현하도록 구성된 것일 수 있다. 일 구현예로, 상기 엔지니어링된 제1 가이드 RNA 서열은 제1 스캐폴드 서열, 제1 스페이서 서열 및 제1 U-rich tail 서열을 포함하고, 상기 엔지니어링된 제2 가이드 RNA 서열은 제2 스캐폴드 서열, 제2 스페이서 서열 및 제2 U-rich tail 서열을 포함할 수 있다.
- [352] 또한, 상기 벡터 시스템은 전술한 표적 핵산 편집 시스템의 구성요소 이외에, 이 기술 분야의 통상의 기술자가 필요에 의해 발현시키고자 하는 부가 발현 요소를 암호화하는 핵산 서열을 포함하고 있을 수 있다.

- [353] 일 예로, 상기 부가 발현 요소는 태그일 수 있다. 구체적으로, 상기 부가 발현 요소는, 글리포세이트(glyphosate), 글루포시네이트암모늄 (glufosinate ammonium) 또는 포스포노트리신(phosphinothricin)과 같은 제초제 저 항성 유전자, 암피실린(ampicillin), 카나마이신(kanamycin), G418, 블레오마이신(Bleomycin), 하이그로마이신(hygromycin), 클로람페니콜(chloramphenicol)과 같은 항생제 내성 유전자일 수 있다.
- [354] 다른 구현예에서, 상기 벡터 시스템은 비상동말단연결(Non-homologous end joining)에 관여하는 유전자의 발현을 억제하는 분자를 암호화하는 뉴클레오티드 서열이 작동 가능하게 연결된 하나 이상의 핵산 구조물을 추가로 포함할 수 있다. 여기서, 비상동말단연결에 관여하는 유전자는 ATM1, XRCC4, XLF, XRCC6, LIG4 및 DCLRE1C로 이루어진 군으로부터 선택된 어느 하나 이상일 수 있다. 또한, 상기 분자는 shRNA, siRNA, miRNA, 또는 안티센스 올리고뉴클레오티드일 수 있다. 상기 분자는 "4. 비상동말단연결 과정을 억제하는 분자" 항목을 참조하며, 상기 "비상동말단연결"과 관련된 내용은 후술한 내용을 참조한다.
- [355] **2. 조절 및/또는 제어 구성요소**
- [356] 상기 벡터 시스템을 직접 세포 내에서 발현시키기 위해서는 하나 이상의 조절 및/또는 제어 구성요소를 포함해야 한다. 구체적으로, 상기 조절 및/또는 제어 구성요소는 프로모터, 인핸서, 인트론, 폴리아데닐화 신호, 코작 공통(Kozak consensus) 서열, 내부 리보솜 유입 부위(IRES, Internal Ribosome Entry Site), 스플라이스 엑셉터, 2A 서열 및/또는 복제원점(replication origin)을 포함할 수 있으나, 이에 제한되는 것은 아니다. 여기서, 상기 복제원점은 f1 복제원점, SV40 복제원점, pMB1 복제원점, 아데노 복제원점, AAV 복제원점 및/또는 BBV 복제원점일 수 있으나, 이에 제한되는 것은 아니다.
- [357] **3. 프로모터**
- [358] 상기 벡터 시스템에 포함되어 있는 본 발명의 핵산 편집 시스템을 암호화하는 핵산 서열을 세포 내에서 발현시키기 위해서, 각 구성 요소를 암호화하는 서열에 프로모터 서열을 작동가능하게 연결시켜 세포 내에서 RNA 전사인자가 활성화될 수 있도록 해야 한다. 상기 프로모터 서열은 대응하는 RNA 전사인자 또는 발현 환경에 따라 달리 설계할 수 있으며, 본원의 핵산 편집 시스템(TaRGET system)의 구성요소를 세포 내에서 적절히 발현시킬 수 있는 것이라면 제한되지 않는다.
- [359] 일 예로, 상기 프로모터 서열은 RNA 중합효소 RNA Pol I, Pol II 또는 Pol III의 전사를 촉진시키는 프로모터일 수 있다. 구체적으로, 상기 프로모터는 U6 프로모터, EFS 프로모터, EF1- α 프로모터, H1 프로모터, 7SK 프로모터, CMV 프로모터, LTR 프로모터, Ad MLP 프로모터, HSV 프로모터, SV40 프로모터, CBA 프로모터 또는 RSV 프로모터 중 하나 수 있다.
- [360] **4. 종결 신호**

- [361] 상기 벡터 서열이 프로모터 서열을 포함하는 경우에 RNA 전사인자에 의해 상기 프로모터와 작동 가능하게 연결된 서열의 전사가 유도되는데, 이러한 RNA 전사 인자의 전사 종결을 유도하는 종결 신호가 포함될 수 있다. 상기 종결 신호는 프로모터 서열의 종류에 따라 달라질 수 있다. 구체적으로, 상기 프로모터가 U6, 또는 H1 프로모터일 경우, 상기 프로모터는 티미딘(T) 연속 서열인 TTTTT(T5) 또는 TTTTTT(T6) 서열을 종결 신호로 인식한다.
- [362] 본 발명에서 제공하는 엔지니어링된 가이드 RNA의 서열은 그 3'-말단에 U-rich tail 서열을 포함한다. 이에 따라, 상기 엔지니어링된 가이드 RNA를 암호화하는 서열은 그 3'-말단에 U-rich tail 서열에 대응하는 T-rich 서열을 포함하게 된다. 전술한 바, 일부 프로모터 서열은 티미딘(T) 연속 서열, 예를 들어 티미딘(T)이 5개 이상 연속으로 연결된 서열을 종결 신호로 인식하므로, 경우에 따라 상기 T-rich 서열을 종결 신호로 인식하게 될 수 있다.
- [363] 다시 말해, 본 명세서에서 제공하는 벡터 서열이 엔지니어링된 가이드 RNA를 암호화하는 서열을 포함하는 경우, 상기 엔지니어링된 gRNA 서열에 포함된 U-rich tail 서열을 암호화하는 서열이 종결 신호로 사용될 수 있다.
- [364] 일 구현예로, 상기 벡터 서열이 U6 또는 H1 프로모터 서열을 포함하고, 이와 작동 가능하게 연결된 엔지니어링된 가이드 RNA를 암호화하는 서열을 포함할 때, 상기 augment RNA 서열에 포함된 U-rich tail 서열을 암호화하는 서열 부분이 종결 신호로 인식될 수 있다. 이때, 상기 U-rich tail 서열은 유리딘(U)이 5개 이상 연속으로 연결된 서열을 포함한다.
- [365] **5. 부가 발현 요소**
- [366] 상기 벡터는 필요에 따라, NLS, NES 및/또는 태그 단백질 등의 부가 구성 요소를 발현하도록 구성된 것일 수 있다.
- [367] 일 구현예로, 상기 부가 구성 요소는 상기 Cas12f1, TnpB 또는 이의 변이체 단백질에 대한 엔지니어링된 가이드 RNA(gRNA)와는 독립적으로 발현될 수 있다.
- [368] 또 다른 구현예로, 상기 부가 구성 요소는 상기 Cas12f1, TnpB 또는 이의 변이체 단백질에 대한 엔지니어링된 가이드 RNA(gRNA)와 직접 또는 링커로 연결되어 발현될 수 있다.
- [369] 일 예로, 본 발명에 따른 초소형 유전자 편집 시스템의 구성요소를 암호화하는 핵산 구조물은 N-말단 또는 C-말단에 핵 위치 신호(nuclear localization sequences, NLS) 서열을 1개 이상 포함하는 것을 특징으로 하는 핵산 구조물일 수 있다. 여기서, 상기 부가 구성 요소는 본원의 핵산 편집 시스템을 발현시키고자 할 때 일반적으로 발현시키는 구성 요소일 수 있으며, 통상의 기술자에게 널리 인식되고 있는 공지기술을 참조할 수 있다.
- [370] 또한, 본 발명은 일 구현예로, 본 발명에 따른 엔지니어링된 가이드 RNA(gRNA) 또는 이를 암호화하는 핵산 및/또는 표적 핵산 편집 시스템의 구성 요소를 발현시키기 위해 벡터 등에 포함되는 핵산을 제공한다. 여기서, 상기 핵산은 자연계에 존재하는 DNA 또는 RNA일 수 있고, 상기 핵산의 일부 또는 전부

에 화학적 변형이 일어난 변형된 핵산일 수 있다. 예를 들어, 상기 핵산은 하나 이상의 뉴클레오타이드가 화학적으로 변형된 것일 수 있다. 이때, 상기 화학적 변형은 이 기술 분야의 통상의 기술자에게 알려진 핵산의 변형을 모두 포함할 수 있다.

[371] **6. 발현 벡터의 종류 및 형태**

[372] 본 발명에 따른 벡터는 바이러스 벡터일 수 있다. 보다 구체적으로, 상기 바이러스 벡터는 레트로바이러스 벡터, 렌티바이러스 벡터, 아데노바이러스 벡터, 아데노-연관 바이러스 벡터, 백시니아바이러스 벡터, 폭스바이러스 벡터, 단순포진 바이러스 벡터 및 파지미드 벡터로 구성된 군에서 선택되는 하나 이상일 수 있다. 일 구현예로, 상기 바이러스 벡터는 아데노-연관 바이러스 벡터일 수 있다.

[373] 또한, 본 발명에 따른 벡터는 비-바이러스 벡터일 수 있다. 보다 구체적으로, 상기 비-바이러스 벡터는 플라스미드, 네이키드 DNA, DNA 복합체, mRNA(전사물) 및 앰플리콘(amplicon)으로 구성된 군에서 선택되는 하나 이상일 수 있으나, 이에 제한되지 않는다. 일 구현예로, 상기 플라스미드는 pcDNA 시리즈, pSC101, pGV1106, pACYC177, ColE1, pKT230, pME290, pBR322, pUC8/9, pUC6, pBD9, pHc79, pIJ61, pLAFR1, pHV14, pGEX 시리즈, pET 시리즈, 및 pUC19으로 이루어진 군에서 선택된 것일 수 있다.

[374] 상기 용어 "네이키드 DNA"는 발현을 위해서 적절한 배향으로 적합한 발현 벡터(예를 들어, 플라스미드) 내에 클로닝된 단백질, 예컨대, 본 발명의 Cas12f1, TnpB 또는 이의 변이체를 암호화하는 DNA(예를 들어, 히스톤이 없는 DNA)를 지칭한다. 사용될 수 있는 바이러스 벡터는 SIN 렌티바이러스 벡터, 레트로바이러스 벡터, 폼(foamy) 바이러스 벡터, 아데노바이러스 벡터, 아데노-연관 바이러스(AAV) 벡터, 하이브리드 벡터 및/또는 플라스미드 트랜스포존(예를 들어, 슬리핑 뷰티 트랜스포존 시스템) 또는 인테그라제 기반 벡터 시스템을 포함하지만 이들로 제한되지 않는다.

[375] 상기 용어 "앰플리콘(amplicon)"은 핵산에 대해 이용되는 경우, 핵산 복제 산물을 의미하며, 여기서 산물은 핵산의 적어도 일부 뉴클레오타이드 서열과 동일하거나 상보적인 뉴클레오타이드 서열을 갖는다. 앰플리콘은, 예를 들어 폴리머라제 확장, 폴리머라제 연쇄 반응(PCR), 롤링 서클 증폭(RCA), 다중 변위 증폭(MDA), 결찰 확장, 또는 결찰 연쇄 반응을 포함하는, 주형으로서 핵산 또는 이들의 앰플리콘을 이용하는 다양한 임의의 증폭 방법에 의해 생성될 수 있다. 앰플리콘은 특정 뉴클레오타이드 서열의 단일 복사체(예를 들어, PCR 산물) 또는 뉴클레오타이드 서열의 다중 사본(예로 RCA의 콘카타머 산물)을 갖는 핵산 분자일 수 있다.

[376] 본 발명의 벡터는 선형 또는 원형 벡터 형태로 설계될 수 있다. 상기 벡터가 선형 벡터인 경우, 상기 선형 벡터 서열이 종결 신호를 따로 포함하지 않더라도, 그 3'-말단에서 RNA 전사가 종결된다. 그러나 상기 벡터가 원형 벡터인 경우, 상기 원형 벡터 서열이 종결 신호를 따로 포함하지 않는다면, RNA 전사가 종결되지

않게 된다. 그러므로 상기 벡터로 원형 벡터를 사용하는 경우에는 의도한 대상을 발현하기 위해서는 각 프로모터 서열과 관련된 전사 인자에 대응하는 종결 신호가 포함되어야 한다.

[377] 일 구현예에서, 바이러스 벡터 또는 비-바이러스 벡터는 리포솜, 폴리머 나노파티클(예컨대, 지질 나노파티클), 수중유 나노에멀전 또는 이들의 조합과 같은 전달 시스템에 의해 전달될 수 있거나, 바이러스 형태로 전달될 수 있다.

[378] **V. 표적 핵산 편집 시스템을 발현하는 바이러스**

[379] 본 발명의 또 다른 태양에 따르면, Cas12f1, TnpB 또는 이의 변이체(variant) 단백질질을 포함하는 엔도뉴클레아제를 암호화하는 뉴클레오티드 서열이 작동 가능하게 연결된 제1 핵산 구조물; 표적 핵산에 상보적으로 결합하는 가이드 서열을 포함하는 엔지니어링된 가이드 RNA를 암호화하는 뉴클레오티드 서열이 작동 가능하게 연결된 제2 핵산 구조물; 및 공여자 핵산 분자를 포함하는 제3 핵산 구조물을 포함하는 하나 이상의 벡터를 포함하는 바이러스 벡터 시스템에 의해 제조된 바이러스 또는 바이러스 입자가 제공된다.

[380] 일 구현예에서, 상기 바이러스 벡터는 예를 들어, 레트로바이러스 벡터(retroviral(retrovirus) vector), 렌티바이러스 벡터(lentiviral(lentivirus) vector), 아데노바이러스 벡터(adenoviral(adenovirus) vector), 아데노 연관 바이러스 벡터(adenovirus-associated viral (adenovirus-associated virus; AAV) vector), 백신시아바이러스 벡터(vaccinia viral(vaccinia virus) vector), 포스바이러스 벡터(poxviral(poxvirus) vector), 단순포진 바이러스 벡터(herpes simplex viral(herpes simplex virus) vector) 및 파지미드 벡터(phagemid vector)로 구성된 군에서 선택되는 하나 이상의 바이러스 벡터일 수 있다. 바람직하게, 상기 바이러스 벡터는 아데노 연관 바이러스 벡터일 수 있다.

[381] 다른 구현예에서, 상기 바이러스는 레트로바이러스, 렌티바이러스, 아데노바이러스, 아데노-연관 바이러스, 백신시아바이러스, 포스바이러스, 단순포진 바이러스 및 파지(phage)로 구성된 군에서 선택될 수 있다.

[382] 또 다른 구현예에서, 상기 파지는 λ gt4 λ B, λ -charon, λ Δ z1, 및 M13으로 이루어진 군에서 선택된 것일 수 있다.

[383] 상기 "Cas12f1, TnpB 또는 이의 변이체(variant) 단백질", "엔지니어링된 가이드 RNA" 및 "공여자 핵산 분자"에 관한 내용은 상기를 참조한다.

[384] 본 발명의 표적 핵산 편집 시스템을 바이러스, 특히 아데노 연관 바이러스(AAV)를 통해 표적 세포 또는 표적 부위로 효율적으로 전달하기 위해서는 상기 모든 구성요소를 암호화하는 뉴클레오티드 서열의 크기가 AAV의 패키징 한계인 4.7 kb 내로 설계하는 것이 중요하다. 본원의 초소형 핵산 편집 단백질 및 엔지니어링된 gRNA를 포함하더라도 그 크기가 매우 작기 때문에 더 긴 공여자 핵산 분자 및 추가적인 조절 분자를 더 포함하더라도 AAV에 의해 충분히 패키징될 수 있다는 이점이 있다.

- [385] 본 발명의 일 실시예에 따르면, 프로모터의 종류, 공여자 핵산의 길이 및 비상동말단연결 조절 유전자의 발현 억제를 위한 분자의 유무에 따라 4.7 kb 내 크기의 벡터를 제작하였고(도 10a 참조), 이들 4종 벡터 모두 3가지 표적 유전자 모두에서 높은 상동지정복구 효율을 나타냄을 확인하여, 본 발명에 따른 바이러스 벡터 시스템에 의해 발현된 AAV 등의 바이러스가 정상적으로 생성되어 효율적으로 본원의 핵산 편집 시스템을 세포 내로 전달 및 발현시킬 수 있음을 입증하였다(실시예 4, 도 10a 및 10b 참조).
- [386] 상기 바이러스 벡터는 선택적으로 조절/제어 구성요소, 프로모터 및/또는 부가 발현 요소를 추가로 포함할 수 있다. 상기 조절/제어 구성요소에 관한 내용은 상기를 참조한다.
- [387] **VI. 표적 핵산 편집 시스템을 이용한 상동지정복구 유도 방법**
- [388] 본 발명의 또 다른 태양에 따르면, 본 발명에 따른 시스템 또는 조성물, 또는 벡터 시스템을 세포와 접촉시키거나 세포 내에서 발현시키는 단계; 및 표적 핵산 또는 그 인접 부위에서 공여자 핵산 분자를 주형으로 하여 이중가닥 절단의 복구 과정에 의해 목적 서열이 도입되는 단계를 포함하는 세포 내 이중가닥 핵산 상의 표적 부위로 목적 서열을 도입하는 방법이 제공된다.
- [389] 일 구현예로, 상기 목적 서열을 도입하는 방법은 Cas12f1, TnpB 또는 이의 변이체에 대해 엔지니어링된 가이드 RNA(augment RNA), Cas12f1, TnpB 또는 이의 변이체 단백질 및 공여자 핵산 분자 또는 이들 각각을 암호화하는 핵산을 표적 핵산 또는 표적 유전자를 포함하고 있는 대상 세포 내에 전달하는 것을 포함한다. 그 결과, 상기 대상 세포 내에 엔지니어링된 가이드 RNA를 포함하는 가이드 RNA 및 Cas12f1, TnpB 또는 이의 변이체 단백질 복합체가 주입되거나, 상기 가이드 RNA 및 Cas12f1, TnpB 또는 이의 변이체 단백질 복합체의 형성이 유도되며, 상기 가이드 RNA 및 Cas12f1, TnpB 또는 이의 변이체 단백질 복합체에 의해 표적 유전자가 절단, 편집 및/또는 복구된다. 유전자 편집은 표적 유전자 또는 표적 핵산 내의 표적 서열을 가지는 이중가닥 DNA, 단일가닥 DNA, 또는 DNA와 RNA 혼성 이중가닥의 핵산 절단을 포함한다. 바람직하게는, 이중가닥 DNA의 핵산 절단을 포함한다. 여기서, 상기 Cas12f1, TnpB 변이체 단백질은 야생형의 Cas12f1 변이체 단백질, 엔지니어링된 Caf12f1 변이체 단백질, 변형된 Cas12f1 변이체 단백질 또는 Cas12f1 변이체의 동족체 단백질일 수 있다.
- [390] 세포에서의 DNA 파단(예를 들어, 이중가닥 절단)의 복구는 주로 2가지 DNA 복구 경로, 즉, 비상동말단연결(NHEJ, 예를 들면, C-NHEJ) 복구 경로 및 상동지정 복구(HDR) 경로를 통해 달성된다. 비상동말단연결이 일어나는 동안, Ku70/80 이종이량체는 DNA 말단에 결합하고, DNA 단백질 키나제(DNA-PK)를 동원한다 [문헌(Cannan & Pederson (2015) J Cell Physiol 231:3-14) 참조]. NHEJ와 연관된 분자들이 결합되면, DNA-PK는 그 자신의 촉매 서브유닛(DNA-PKcs)을 활성화시키고, 추가로 엔도뉴클레아제 아르테미스(Artemis; 또는 SNM1c로 공지됨)를 과정에 참여시킨다. 이중가닥 절단의 서브유닛에서, 아르테미스는 과량의 단일가

닥 DNA(ssDNA)를 제거하고, DNA 리가제 IV에 의해 라이게이션될 기질을 생성한다. 비상동말단연결에 의한 DNA 복구는 DNA-PKcs/Ku70/80 복합체를 통한 서열 상동성에 독립적인 블런트-말단 라이게이션 메커니즘을 포함한다. 세포 주기 동안, 비상동말단연결은 G0/G1 및 G2에서 우세하게 일어난다[문헌(Chiruvella et al., (2013) Cold Spring Harb Perspect Biol 5:a012757) 참조]. 현재의 연구는 비상동말단연결이 G0 및 G1에서 활성인 유일한 이중가닥 절단 복구 경로인 반면, 상동지정복구는 주로 S 및 G2 기에서 기능하여, 복제-연관된 이중가닥 절단의 복구에 주요한 역할을 함을 보였다[문헌 (Karanam et al., (2012) Mol Cell 47:320-329; Li and Xu (2016) Acta Biochim Biophys Sin 48(7):641-646) 참조]. 비상동말단연결은 상동지정복구와는 달리 분열 세포뿐만 아니라, 분열 및 비-분열 세포 둘 다에서 활성이다. 상동지정복구에 의한 DNA 복구 동안, 이중가닥이 절단된 유전자의 말단은 주로 MRN(MRE11-RAD50-NBS1) 복합체에 의해 절제되어 3'-ssDNA 꼬리를 노출한다[문헌(Heyer et al., (2010) Annu Rev Genet 44: 113-139) 참조]. 생리학적인 조건 하에서, 인접한 염색체는 상동성 서열을 제공하는 복구 주형으로서 사용될 것이다.

- [391] 제3 복구 메커니즘은 또한 "대안적 NHEJ(A-NHEJ)"로 지칭되는 미세상동성-매개된 말단 연결(MMEJ)이며, 여기서 유전적 결과는 작은 결실 및 삽입이 절단 부위에서 일어날 수 있다는 점에서 NHEJ와 유사하다. MMEJ는 보다 바람직한 DNA 말단 연결 복구 결과를 유도하는 DNA 파단 부위에 플랭킹된 소수의 뉴클레오타이드의 상동성 서열을 사용하며, 최근의 보고는 이 프로세스의 분자적 메커니즘을 추가로 설명하였다[문헌(Cho and Greenberg, (2015) Nature 518:174-176; Mateos-Gomez et al., (2015) Nature 518, 254-257; Ceccaldi et al., (2015) Nature 528, 258-262) 참조].
- [392] 포유류 세포에서, "표준적" 또는 "고전적" NHEJ 경로(C-NHEJ)는 핵산 또는 유전자에서 이중가닥 절단을 수복하기 위해 DNA-PK, Ku70-80, 아르테미스, 리가제 IV(Lig4), XRCC4, CLF, 및 Pol Mu를 포함하는 몇 개의 인자를 요구한다[문헌 (Kasperek & Humphrey (2011) Seminars in Cell & Dev. Biol. 22:886-897) 참조].
- [393] 따라서, 본 명세서에서 개시된 시스템, 조성물 또는 방법의 일부에서, 세포는 C-NHEJ에서 관여된 인자의 발현 또는 활성을 감소 또는 제거하도록 변형될 수 있다. 예를 들어, 일부 시스템 또는 방법은 MRE11, RAD50, NBS1, DNA-PK, CtIP, Ku70, Ku80, 아르테미스(DCLRE1C), 리가제 IV(Lig4), PNKP, XRCC4, XLF(XRCC4-like factor), ATM(ATM Serine/Threonine Kinase), CHK1/CHK2, CLF(CURLY LEAF), 및/또는 Pol Mu(POLM) 발현 또는 활성을 감소 또는 제거할 수 있는 인자를 더 포함할 수 있다.
- [394] 또한, 본 명세서에서 개시된 시스템, 조성물 또는 방법의 일부에서, 세포는 A-NHEJ에서 관여된 인자의 발현 또는 활성을 감소 또는 제거하도록 변형될 수 있다. 예를 들어, 일부 시스템, 조성물 또는 방법은 XRCC1, PARP(예를 들면,

PARP1), Lig1, 및/또는 Lig3 발현 또는 활성을 감소 또는 제거할 수 있는 인자를 더 포함할 수 있다.

- [395] 상동지정복구를 통한 DNA 복구는 자유 DNA 말단을 프로세싱하는 뉴클레아제 또는 헬리카제를 포함한 다수 클래스의 단백질, 그리고 보조 상동지정복구 인자를 위한 핵형성 부위로 작용하는 단백질 결합 도메인과 연관되어 있다. 일 구체예에서, 상동지정복구 단백질은 DNA 가닥 절단을 촉진하는 뉴클레아제 및/또는 헬리카제, 예컨대 MRE11, EXO1, DNA2, CtIP, TREX2 및 아폴로; 특이적 인자를 동원하거나 가닥 침입을 촉매하는 결합 인자/핵형성 단백질, 예컨대 BRCA1, BRCA2, PALB2, RAD50 또는 NBS1, RAD51, RAD52, RAD54, SRCAP, FANCI, FANCD2, BRIP1, SLX4, FANCA, FANCE 및 FANCL (이들 인자의 말단 절단되거나, 돌연변이되거나, 변형되거나, 또는 최적화된 버전 포함)로 이루어진 군으로부터 선택될 수 있다.
- [396] 일 구체예에 있어서, 상동지정복구를 강화하기 위하여, 본원에서 기술되는 상동지정복구 단백질의 중 어느 것의 전체 길이 또는 말단 절단된 단백질이 엔도뉴클레아제(예를 들면, Cas12f1, TnpB 또는 이의 변이체)와 결합될 수 있다. 따라서, 본 명세서에서는 Cas12f1, TnpB 또는 이의 변이체 단백질을 포함하는 융합 단백질을 추가로 제공한다. 상기 Cas12f1, TnpB 또는 이의 변이체 및 상동지정복구 단백질은 1-100, 1-50, 1-30 또는 1-20개의 임의의 아미노산 서열을 통해 연결된 것일 수 있다.
- [397] 일부 구체예에서, 상기 방법은 본 발명의 시스템, 조성물 또는 벡터 시스템을 세포와 접촉시키거나 세포 내에서 발현시키는 단계; 및 공여자 핵산 분자가 이중가닥 절단의 부위로 삽입되고 이중가닥 절단이 복구되기에 충분한 조건하에서 표적 부위에 이중가닥 절단을 유도함으로써, 표적 핵산 또는 그 인접 부위로 공여자 핵산 분자가 도입되는 단계를 포함하는, 세포 내 이중가닥 핵산 상의 표적 부위로 목적하는 서열을 도입하는 방법을 제공한다. 여기서, 이중가닥 절단의 복구 과정은 상동지정복구 기작에 의하는 것일 수 있다.
- [398] 상기 세포와 접촉시키는 단계는 상기 표적 핵산 편집 시스템 시스템의 세포 내로의 전달 또는 도입을 포함하는 것일 수 있다. 상기 세포에 표적 핵산 편집 시스템을 처리하는 것은 전기천공법, 유전자총, 초음파천공법, 자기주입법 (magnetofection), 나노파티클 방법 및/또는 일시적인 세포 압축 또는 스퀴징 방법을 이용한 것일 수 있다. 또는 상기 진핵 세포에 조성물을 처리하는 것은 양이온성 리포솜법, 초산 리튬-DMSO, 지질-매개 형질감염(transfection), 인산칼슘 침전법(precipitation), 리포펙타민(lipofection), PEI(Polyethyleneimine)-매개 형질감염, DEAE-dextran 매개 형질감염, 및/또는 나노파티클-매개 핵산 전달[문헌(Panyam et al., Adv Drug Deliv Rev. 2012 Sep 13. pii: S0169-409X(12)00283-9.) 참조]을 이용한 것일 수 있다. 상기 세포에 상기 시스템을 처리, 전달 또는 도입하는 것은 *in vitro*, *in vivo* 또는 *ex vivo*에서 수행될 수 있다.

- [399] 상기 벡터 시스템은 레트로바이러스, 렌티바이러스, 아데노바이러스, 아데노-연관 바이러스, 백시니아바이러스, 폭스바이러스, 단순포진 바이러스 및 파지(phage)로 구성된 군에서 선택된 패키징 바이러스에 도입되어 패키징 바이러스에 의해 생성된 바이러스 형태로 원핵 세포 또는 진핵 세포 내로 전달되는 것일 수 있다.
- [400] 상기 세포는 식물세포, 비인간 동물 세포 또는 인간 세포일 수 있다. 또한, 상기 세포는 진핵 세포 또는 원핵 세포일 수 있다.
- [401] 상기 "Cas12f1, TnpB 또는 이의 변이체(variant) 단백질", "엔지니어링된 가이드 RNA" 및 "공여자 핵산 분자"에 관한 내용은 상기를 참조한다.
- [402] 대표적인 구현예 또는 구현예의 조합은 다음과 같다.
- [403] [구현예 1]
- [404] Cas12f1, TnpB 또는 이의 변이체(variant) 단백질을 포함하는 엔도뉴클레아제 또는 상기 엔도뉴클레아제를 암호화하는 핵산;
- [405] 가이드 서열을 포함하는 엔지니어링된 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 핵산; 및
- [406] 공여자 핵산 분자 또는 상기 공여자 핵산 분자를 암호화하는 핵산
- [407] 을 포함하는 표적 핵산의 편집 시스템.
- [408] [구현예 2]
- [409] 전술한 구현예에 있어서,
- [410] 상기 시스템은 표적 핵산에 이중가닥 절단(double-strand breaks)을 일으키는 것인
- [411] 시스템.
- [412] [구현예 3]
- [413] 전술한 구현예 중 어느 한 구현예에 있어서,
- [414] 상기 시스템은 표적 핵산 또는 그 인접 부위에서 공여자 핵산 분자를 주형으로 하여 이중가닥 절단의 상동지정복구(homology-directed repair) 과정에 의해 목적 서열이 도입되는 것인
- [415] 시스템.
- [416] [구현예 4]
- [417] Cas12f1, TnpB 또는 이의 변이체(variant) 단백질을 포함하는 엔도뉴클레아제 또는 상기 엔도뉴클레아제를 암호화하는 핵산;
- [418] 가이드 서열을 포함하는 엔지니어링된 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 핵산; 및
- [419] 공여자 핵산 분자 또는 상기 공여자 핵산 분자를 암호화하는 핵산
- [420] 을 포함하는 표적 핵산의 유전자 편집 조성물.
- [421] [구현예 5]
- [422] 전술한 구현예 중 어느 한 구현예에 있어서,

- [423] 상기 Cas12f1, TnpB 또는 이의 변이체 단백질은 서열번호 1 내지 서열번호 5로 이루어진 군에서 선택된 아미노산 서열과 70% 이상의 서열 동일성을 갖는 아미노산 서열을 포함하는
- [424] 시스템 또는 조성물.
- [425] [구현예 6]
- [426] 전술한 구현예 중 어느 한 구현예에 있어서,
- [427] 상기 TnpB 단백질은 서열번호 202 내지 서열번호 293로 이루어진 군에서 선택된 아미노산 서열과 70% 이상의 서열 동일성을 갖는 아미노산 서열을 포함하는
- [428] 시스템 또는 조성물.
- [429] [구현예 7]
- [430] 전술한 구현예 중 어느 한 구현예에 있어서,
- [431] 상기 Cas12f1, TnpB 또는 이의 변이체 단백질은 하기 서열 중 하나를 포함하는
- [432] 시스템 또는 조성물:
- [433] (i) 서열번호 5의 아미노산 서열;
- [434] (ii) 서열번호 1의 아미노산 서열;
- [435] (iii) 서열번호 1의 아미노산 서열에서 N-말단의 1개 내지 28개의 아미노산이 제거 또는 치환된 아미노산 서열; 또는
- [436] (iv) 서열번호 1의 아미노산 서열의 N-말단 또는 C-말단에 1개 내지 600개의 아미노산이 추가된 아미노산 서열.
- [437] [구현예 8]
- [438] 전술한 구현예 중 어느 한 구현예에 있어서,
- [439] 상기 추가된 1개 내지 600개의 아미노산은 서열번호 294 또는 서열번호 295의 아미노산 서열인
- [440] 시스템 또는 조성물.
- [441] [구현예 9]
- [442] 전술한 구현예 중 어느 한 구현예에 있어서,
- [443] 상기 Cas12f1, TnpB 또는 이의 변이체 단백질은 서열번호 1 내지 서열번호 4로 이루어진 군에서 선택된 아미노산 서열과 70% 이상의 서열 동일성을 갖는
- [444] 시스템 또는 조성물.
- [445] [구현예 10]
- [446] 전술한 구현예 중 어느 한 구현예에 있어서,
- [447] 상기 Cas12f1, TnpB 또는 이의 변이체 단백질은 서열번호 1의 아미노산 서열과 70% 이상의 서열 동일성을 갖는
- [448] 시스템 또는 조성물.
- [449] [구현예 11]
- [450] 전술한 구현예 중 어느 한 구현예에 있어서,
- [451] 상기 엔지니어링된 가이드 RNA는 야생형 Cas12f1 가이드 RNA 서열에서 1개 이상의 뉴클레오티드가 치환, 결실, 삽입 또는 부가된 서열을 포함하고, 가이드

서열을 제외한 부분이 상기 야생형 Cas12f1 가이드 RNA와 50% 이상의 서열 동일성을 갖는

[452] 시스템 또는 조성물.

[453] [구현예 12]

[454] 전술한 구현예 중 어느 한 구현예에 있어서,

[455] 상기 야생형 Cas12f1 가이드 RNA는 (i) 하나 이상의 스템(stem) 영역, (ii) tracrRNA-crRNA 상보성 영역 및 임의적으로 (iii) 연속되는 3개 이상의 유라실(U)을 포함하는 영역을 포함하는 tracrRNA(trans-activating CRISPR RNA) 및 crRNA(CRISPR RNA)를 포함하고,

[456] 상기 엔지니어링된 가이드 RNA는 하기 (a) 내지 (d)로 이루어진 군에서 선택되는 하나 이상의 변형을 포함하는

[457] 시스템 또는 조성물:

[458] (a) 하나 이상의 스템 영역의 일부 또는 전부의 결실;

[459] (b) tracrRNA-crRNA 상보성 영역의 일부 또는 전부의 결실;

[460] (c) tracrRNA-crRNA 상보성 영역 내에 연속되는 3개 이상의 유라실(U)이 존재하는 경우 하나 이상의 U를 A, G 또는 C로 치환; 및

[461] (d) crRNA 서열의 3'-말단에 하나 이상의 유리딘(uridine)의 부가.

[462] [구현예 13]

[463] 전술한 구현예 중 어느 한 구현예에 있어서,

[464] 상기 야생형 Cas12f1 가이드 RNA는 서열번호 11의 핵산 서열로 이루어진 tracrRNA 및 서열번호 12의 핵산 서열로 이루어진 crRNA를 포함하는

[465] 시스템 또는 조성물.

[466] [구현예 14]

[467] 전술한 구현예 중 어느 한 구현예에 있어서,

[468] 상기 야생형 Cas12f1 가이드 RNA는 5'-말단부터 순차적으로 제1 스템 영역, 제2 스템 영역, 제3 스템 영역, 제4 스템 영역 및 tracrRNA-crRNA 상보성 영역을 포함하는 tracrRNA 및 crRNA를 포함하고,

[469] 상기 엔지니어링된 가이드 RNA는

[470] (a1) 제1 스템 영역의 일부 또는 전부의 결실;

[471] (a2) 제2 스템 영역의 일부 또는 전부의 결실;

[472] (b) tracrRNA-crRNA 상보성 영역의 일부 또는 전부의 결실;

[473] (c) tracrRNA-crRNA 상보성 영역 내에 연속되는 3개 이상의 유라실(U)이 존재하는 경우 하나 이상의 U를 A, G 또는 C로 치환; 및

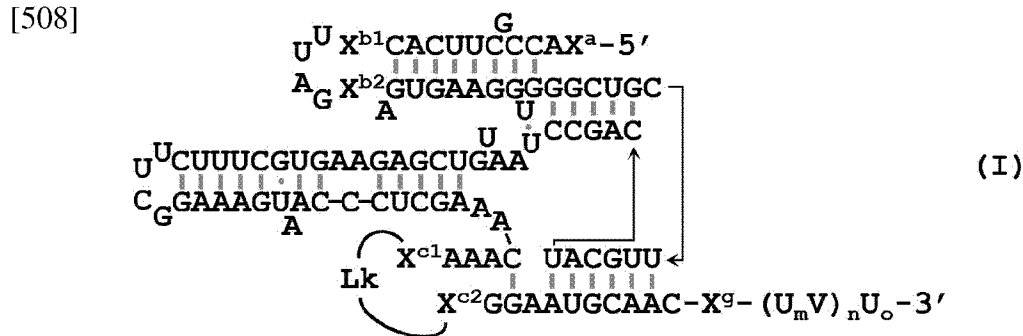
[474] (d1) crRNA 서열의 3'-말단에 U-rich tail의 부가(상기 U-rich tail의 서열은 5'-(U_mV)_nU_o-3'로 표시되고 여기서 상기 V는 각각 독립적으로 A, C 또는 G이고, m 및 o는 1 내지 20 사이의 정수이며, n은 0 내지 5 사이의 정수임)

[475] 로 이루어진 군에서 선택되는 하나 이상의 변형을 포함하는

[476] 시스템 또는 조성물.

- [477] [구현예 15]
 [478] 전술한 구현예 중 어느 한 구현예에 있어서,
 [479] 상기 엔지니어링된 가이드 RNA는 (d1) crRNA 서열의 3'-말단에 U-rich tail, (c) tracrRNA-crRNA 상보성 영역 내에 연속되는 3개 이상의 유라실(U)이 존재하는 경우 하나 이상의 U를 A, G 또는 C로 치환, 또는 상기 둘 모두의 변형을 포함하는
 [480] 시스템 또는 조성물.
- [481] [구현예 16]
 [482] 전술한 구현예 중 어느 한 구현예에 있어서,
 [483] 상기 엔지니어링된 가이드 RNA는 (a1) 제1 스템 영역의 일부 또는 전부의 결실; (a2) 제2 스템 영역의 일부 또는 전부의 결실; 및 (b) tracrRNA-crRNA 상보성 영역의 일부 또는 전부의 결실로 이루어진 군에서 선택된 하나 이상의 변형을 포함하는
 [484] 시스템 또는 조성물.
- [485] [구현예 17]
 [486] 전술한 구현예 중 어느 한 구현예에 있어서,
 [487] 상기 엔지니어링된 가이드 RNA는 (b1) tracrRNA-crRNA 상보성 영역의 일부의 결실을 포함하고, 상기 상보성 영역의 일부는 1개 내지 54개의 뉴클레오티드인
 [488] 시스템 또는 조성물.
- [489] [구현예 18]
 [490] 전술한 구현예 중 어느 한 구현예에 있어서,
 [491] 상기 엔지니어링된 가이드 RNA는 (b2) tracrRNA-crRNA 상보성 영역의 전부의 결실을 포함하고, 상기 상보성 영역의 전부는 55개의 뉴클레오티드인
 [492] 시스템 또는 조성물.
- [493] [구현예 19]
 [494] 전술한 구현예 중 어느 한 구현예에 있어서,
 [495] 상기 엔지니어링된 가이드 RNA는 (a1) 제1 스템 영역의 일부 또는 전부의 결실을 포함하고, 상기 스템 영역의 일부 또는 전부는 1개 내지 20개의 뉴클레오티드인
 [496] 시스템 또는 조성물.
- [497] [구현예 20]
 [498] 전술한 구현예 중 어느 한 구현예에 있어서,
 [499] 상기 엔지니어링된 가이드 RNA는 (a2) 제2 스템 영역의 일부 또는 전부의 결실을 포함하고, 상기 스템 영역의 일부 또는 전부는 1 내지 27개의 뉴클레오티드인
 [500] 시스템 또는 조성물.
- [501] [구현예 21]
 [502] 전술한 구현예 중 어느 한 구현예에 있어서,

- [503] 상기 엔지니어링된 가이드 RNA는 (a1) 제1 스템 영역의 일부 또는 전부의 결실; (d1) crRNA 서열의 3'-말단에 U-rich tail의 부가; 또는 상기 둘 모두의 변형을 포함하는
- [504] 시스템 또는 조성물.
- [505] [구현예 22]
- [506] 전술한 구현예 중 어느 한 구현예에 있어서,
- [507] 상기 엔지니어링된 가이드 RNA는 하기 식 (I)로 표시되는 서열로 이루어지거나 상기 서열과 80% 이상 서열 동일성을 갖는 시스템 또는 조성물:



- [509] 식 (I)에서,
- [510] X^a , X^{b1} , X^{b2} , X^{c1} 및 X^{c2} 는 각각 독립적으로 0 내지 35개의 (폴리)뉴클레오티드로 이루어지고,
- [511] X^g 는 10 내지 30개의 폴리뉴클레오티드로 이루어진 가이드 서열로서 표적 서열과 혼성화하거나 표적 서열에 상보적인 서열이고,
- [512] Lk는 길이 2 내지 20의 폴리뉴클레오티드 링커이거나 부존재하고,
- [513] $(U_mV)_nU_o$ 는 U-rich tail로서 존재하거나 부존재하고, 존재하는 경우 U는 유리딘이고, V는 각각 독립적으로 A, C 또는 G이고, m 및 o는 각각 독립적으로 1 내지 20 사이의 정수이며, n은 0 내지 5 사이의 정수이다.
- [514] [구현예 23]
- [515] 전술한 구현예 중 어느 한 구현예에 있어서,
- [516] 상기 X^a 는 서열번호 14의 핵산 서열을 포함하거나 상기 서열번호 14의 서열에서 1 내지 20개의 뉴클레오티드가 결실된 핵산 서열을 포함하는
- [517] 시스템 또는 조성물.
- [518] [구현예 24]
- [519] 전술한 구현예 중 어느 한 구현예에 있어서,
- [520] 상기 X^{b1} 은 서열번호 25의 핵산 서열을 포함하거나 상기 서열번호 25의 서열에서 1 내지 13개의 뉴클레오티드가 결실된 핵산 서열을 포함하는
- [521] 시스템 또는 조성물.
- [522] [구현예 25]
- [523] 전술한 구현예 중 어느 한 구현예에 있어서,

- [524] 상기 X^{b2} 은 서열번호 29의 핵산 서열을 포함하거나 상기 서열번호 29의 서열에서 1 내지 14개의 뉴클레오티드가 결실된 핵산 서열을 포함하는
- [525] 시스템 또는 조성물.
- [526] [구현예 26]
- [527] 전술한 구현예 중 어느 한 구현예에 있어서,
- [528] 상기 식 (I) 내의 서열 5'- X^{b1} UUAGX ^{b2} -3'은 서열번호 34 내지 서열번호 38로 이루어진 군에서 선택된 어느 하나의 핵산 서열인
- [529] 시스템 또는 조성물.
- [530] [구현예 27]
- [531] 전술한 구현예 중 어느 한 구현예에 있어서,
- [532] 상기 X^{c1} 은 서열번호 39의 핵산 서열을 포함하거나 상기 서열번호 39의 서열에서 1 내지 28개의 뉴클레오티드가 결실된 핵산 서열을 포함하는
- [533] 시스템 또는 조성물.
- [534] [구현예 28]
- [535] 전술한 구현예 중 어느 한 구현예에 있어서,
- [536] 상기 X^{c1} 서열 내에 연속되는 3개 이상의 유라실(U)이 존재하는 경우 이들 중 하나 이상의 U가 A, G 또는 C로 치환되는 변형을 포함하는
- [537] 시스템 또는 조성물.
- [538] [구현예 29]
- [539] 전술한 구현예 중 어느 한 구현예에 있어서,
- [540] 상기 X^{c2} 는 서열번호 58의 핵산 서열을 포함하거나 상기 서열번호 58의 서열에서 1 내지 27개의 뉴클레오티드가 결실된 핵산 서열을 포함하는
- [541] 시스템 또는 조성물.
- [542] [구현예 30]
- [543] 전술한 구현예 중 어느 한 구현예에 있어서,
- [544] 상기 X^{c2} 서열 내에 서열 5'-ACGAA-3'가 존재하는 경우 해당 서열이 5'-NGNNN-3'로 치환되고, 상기 N은 각각 독립적으로 A, C, G 또는 U인
- [545] 시스템 또는 조성물.
- [546] [구현예 31]
- [547] 전술한 구현예 중 어느 한 구현예에 있어서,
- [548] 상기 식 (I) 내의 서열 5'- X^{c1} -Lk- X^{c2} -3'은 서열번호 80 내지 서열번호 86으로 이루어진 군에서 선택된 어느 하나의 핵산 서열인
- [549] 시스템 또는 조성물.
- [550] [구현예 32]
- [551] 전술한 구현예 중 어느 한 구현예에 있어서,
- [552] 상기 Lk는 5'-GAAA-3', 5'-UUAG-3', 5'-UGAAAA-3', 5'-UUGAAAAA-3', 5'-UUCGAAAGAA-3'(서열번호 76), 5'-UUCAGAAAUGAA-3'(서열번호 77), 5'-

UUCAUGAAAAUGAA-3'(서열번호 78) 및 5'-UUCAUUGAAAAUGAA-3'(서열번호 79)로 이루어진 군에서 선택된 어느 하나의 핵산 서열을 포함하는

[553] 시스템 또는 조성물.

[554] [구현예 33]

[555] 전술한 구현예 중 어느 한 구현예에 있어서,

[556] 상기 $(U_mV)_nU_o$ 는 (i) n 이 0이고, o 가 1 내지 6 사이의 정수이거나 (ii) V 가 A 또는 G이고, m 및 o 는 각각 독립적으로 3 내지 6 사이의 정수이고, n 은 1 내지 3 사이의 정수인

[557] 시스템 또는 조성물.

[558] [구현예 34]

[559] 전술한 구현예 중 어느 한 구현예에 있어서,

[560] 상기 엔지니어링된 가이드 RNA는 서열번호 87 내지 서열번호 132로 이루어진 군에서 선택된 어느 하나의 핵산 서열로 이루어진 엔지니어링된 tracrRNA를 포함하는

[561] 시스템 또는 조성물.

[562] [구현예 35]

[563] 전술한 구현예 중 어느 한 구현예에 있어서,

[564] 상기 엔지니어링된 가이드 RNA는 서열번호 133 내지 서열번호 148로 이루어진 군에서 선택된 어느 하나의 핵산 서열로 이루어진 엔지니어링된 crRNA를 포함하는

[565] 시스템 또는 조성물.

[566] [구현예 36]

[567] 전술한 구현예 중 어느 한 구현예에 있어서,

[568] 상기 엔지니어링된 가이드 RNA는 듀얼 가이드 RNA 또는 싱글 가이드 RNA인

[569] 시스템 또는 조성물.

[570] [구현예 37]

[571] 전술한 구현예 중 어느 한 구현예에 있어서,

[572] 상기 엔지니어링된 싱글 가이드 RNA는 서열번호 149 내지 서열번호 186으로 이루어진 군에서 선택된 어느 하나의 핵산 서열로 이루어진

[573] 시스템 또는 조성물.

[574] [구현예 38]

[575] 전술한 구현예 중 어느 한 구현예에 있어서,

[576] 상기 Cas12f1, TnpB 또는 이의 변이체 단백질을 포함하는 엔도뉴클레아제; 및 상기 가이드 RNA를 리보뉴클레오단백질(ribonucleoprotein, RNP) 형태로 포함하는

[577] 시스템 또는 조성물.

[578] [구현예 39]

[579] 전술한 구현예 중 어느 한 구현예에 있어서,

- [580] 상기 공여자 핵산 분자는 상동지정복구에서 주형으로 사용되는 서열로서 1 bp 내지 20 kb의 길이를 갖는
- [581] 시스템 또는 조성물.
- [582] [구현예 40]
- [583] 전술한 구현예 중 어느 한 구현예에 있어서,
- [584] 상기 시스템 또는 조성물은 비상동말단연결(non-homologous end joining, NHEJ)에 관여하는 유전자의 발현을 억제하는 분자를 더 포함하는
- [585] 시스템 또는 조성물.
- [586] [구현예 41]
- [587] 전술한 구현예 중 어느 한 구현예에 있어서,
- [588] 상기 비상동말단연결에 관여하는 유전자는 ATM1, XRCC4, XLF, XRCC6, LIG4 및 DCLRE1C로 이루어진 군으로부터 선택된 어느 하나 이상인
- [589] 시스템 또는 조성물.
- [590] [구현예 42]
- [591] 전술한 구현예 중 어느 한 구현예에 있어서,
- [592] 상기 분자는 shRNA, siRNA, miRNA, 또는 안티센스 올리고뉴클레오티드인
- [593] 시스템 또는 조성물.
- [594] [구현예 43]
- [595] Cas12f1, TnpB 또는 이의 변이체(variant) 단백질을 포함하는 엔도뉴클레아제를 암호화하는 뉴클레오티드 서열이 작동 가능하게 연결된 제1 핵산 구조물;
- [596] 표적 핵산에 상보적으로 결합하는 가이드 서열을 포함하는 엔지니어링된 가이드 RNA를 암호화하는 뉴클레오티드 서열이 작동 가능하게 연결된 제2 핵산 구조물; 및
- [597] 공여자 핵산 분자를 포함하는 제3 핵산 구조물을 포함하는 하나 이상의 벡터를 포함하는
- [598] 벡터 시스템.
- [599] [구현예 44]
- [600] 전술한 구현예 중 어느 한 구현예에 있어서,
- [601] 상기 벡터 시스템에 포함되는 핵산 구조물은 동일하거나 상이한 벡터에 위치하는
- [602] 벡터 시스템.
- [603] [구현예 45]
- [604] 전술한 구현예 중 어느 한 구현예에 있어서,
- [605] 비상동말단연결(Non-homologous end joining; NHEJ)에 관여하는 유전자의 발현을 억제하는 분자를 암호화하는 뉴클레오티드 서열이 작동 가능하게 연결된 하나 이상의 핵산 구조물을 추가로 포함하는
- [606] 벡터 시스템.
- [607] [구현예 46]

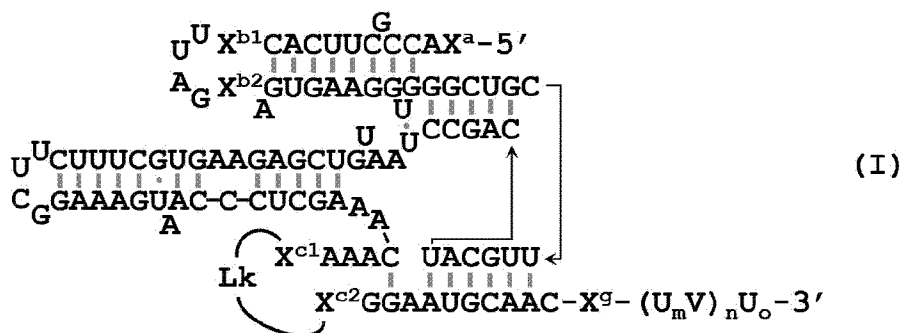
- [608] 전술한 구현예 중 어느 한 구현예에 있어서,
 [609] 상기 비상동말단연결에 관여하는 유전자는 ATM1, XRCC4, XLF, XRCC6, LIG4 및 DCLRE1C로 이루어진 군으로부터 선택된 어느 하나 이상인
 [610] 벡터 시스템.
 [611] [구현예 47]
 [612] 전술한 구현예 중 어느 한 구현예에 있어서,
 [613] 상기 분자는 shRNA, siRNA, miRNA, 또는 안티센스 올리고뉴클레오티드인
 [614] 벡터 시스템.
 [615] [구현예 48]
 [616] 전술한 구현예 중 어느 한 구현예에 있어서,
 [617] 벡터 내의 각 구성요소가 하나의 벡터에 포함되는 것인
 [618] 벡터 시스템.
 [619] [구현예 49]
 [620] 전술한 구현예 중 어느 한 구현예에 있어서,
 [621] 상기 공여자 핵산 분자의 길이는 1 bp 내지 20 kb인
 [622] 벡터 시스템.
 [623] [구현예 50]
 [624] 전술한 구현예 중 어느 한 구현예에 있어서,
 [625] 상기 벡터가 프로모터 또는 인핸서를 더 포함하는
 [626] 벡터 시스템.
 [627] [구현예 51]
 [628] 전술한 구현예 중 어느 한 구현예에 있어서,
 [629] 상기 프로모터는 U6 프로모터, EFS 프로모터, EF1- α 프로모터, H1 프로모터, 7SK 프로모터, CMV 프로모터, LTR 프로모터, Ad MLP 프로모터, HSV 프로모터, SV40 프로모터, CBA 프로모터 또는 RSV 프로모터인
 [630] 벡터 시스템.
 [631] [구현예 52]
 [632] 전술한 구현예 중 어느 한 구현예에 있어서,
 [633] 상기 벡터는 레트로바이러스 벡터(retroviral(retrovirus) vector), 렌티바이러스 벡터(lentiviral(lentivirus) vector), 아데노바이러스 벡터(adenoviral(adenovirus) vector), 아데노-연관 바이러스 벡터(adenovirus-associated viral(adenovirus-associated virus; AAV) vector), 백신시아바이러스 벡터(vaccinia viral(vaccinia virus) vector), 포스 바이러스 벡터(poxviral(poxvirus) vector), 단순포진 바이러스 벡터(herpes simplex viral(herpes simplex virus) vector) 및 파지미드 벡터(phagemid vector)로 구성된 군에서 선택되는 하나 이상의 바이러스 벡터인
 [634] 벡터 시스템.
 [635] [구현예 53]
 [636] 전술한 구현예 중 어느 한 구현예에 있어서,

- [637] 상기 벡터는 아데노-연관 바이러스 벡터이고, 상기 아데노-연관 바이러스 벡터는 상기 벡터 내의 구성요소 모두를 하나의 벡터에 포함할 수 있는
- [638] 벡터 시스템.
- [639] [구현예 54]
- [640] 전술한 구현예 중 어느 한 구현예에 있어서,
- [641] 상기 벡터는 플라스미드, 네이키드 DNA, DNA 복합체, mRNA(전사물) 및 앰플리콘(amplicon)으로 이루어진 군에서 선택된 하나 이상의 비-바이러스 벡터인
- [642] 벡터 시스템.
- [643] [구현예 55]
- [644] 전술한 구현예 중 어느 한 구현예에 있어서,
- [645] 상기 플라스미드는 pcDNA 시리즈, pSC101, pGV1106, pACYC177, ColE1, pKT230, pME290, pBR322, pUC8/9, pUC6, pBD9, pHC79, pIJ61, pLAFR1, pHV14, pGEX 시리즈, pET 시리즈, 및 pUC19으로 이루어진 군에서 선택된 하나 이상인
- [646] 벡터 시스템.
- [647] [구현예 56]
- [648] 전술한 구현예 중 어느 한 구현예에 있어서,
- [649] 상기 Cas12f1, TnpB 또는 이의 변이체 단백질은 서열번호 1 내지 5로 이루어진 군에서 선택된 아미노산 서열과 70% 이상의 서열 동일성을 갖는 아미노산 서열을 포함하는
- [650] 벡터 시스템.
- [651] [구현예 57]
- [652] 전술한 구현예 중 어느 한 구현예에 있어서,
- [653] 상기 TnpB 단백질은 서열번호 202 내지 서열번호 293로 이루어진 군에서 선택된 아미노산 서열과 70% 이상의 서열 동일성을 갖는 아미노산 서열을 포함하는
- [654] 벡터 시스템.
- [655] [구현예 58]
- [656] 전술한 구현예 중 어느 한 구현예에 있어서,
- [657] 상기 Cas12f1, TnpB 또는 이의 변이체 단백질은 하기 서열 중 하나를 포함하는
- [658] 벡터 시스템:
- [659] (i) 서열번호 5의 아미노산 서열;
- [660] (ii) 서열번호 1의 아미노산 서열;
- [661] (iii) 서열번호 1의 아미노산 서열에서 N-말단의 1개 내지 28개의 아미노산이 제거 또는 치환된 아미노산 서열; 또는
- [662] (iv) 서열번호 1의 아미노산 서열의 N-말단 또는 C-말단에 1개 내지 600개의 아미노산이 추가된 아미노산 서열.
- [663] [구현예 59]
- [664] 전술한 구현예 중 어느 한 구현예에 있어서,

- [665] 상기 추가된 1개 내지 600개의 아미노산은 서열번호 294 또는 서열번호 295의 아미노산 서열인
- [666] 벡터 시스템.
- [667] [구현예 60]
- [668] 전술한 구현예 중 어느 한 구현예에 있어서,
- [669] 상기 Cas12f1, TnpB 또는 이의 변이체 단백질은 서열번호 1 내지 서열번호 4로 이루어진 군에서 선택된 아미노산 서열과 70% 이상의 서열 동일성을 갖는
- [670] 벡터 시스템.
- [671] [구현예 61]
- [672] 전술한 구현예 중 어느 한 구현예에 있어서,
- [673] 상기 Cas12f1, TnpB 또는 이의 변이체 단백질은 서열번호 1의 아미노산 서열과 70% 이상의 서열 동일성을 갖는
- [674] 벡터 시스템.
- [675] [구현예 62]
- [676] 전술한 구현예 중 어느 한 구현예에 있어서,
- [677] 상기 엔지니어링된 가이드 RNA는 야생형 Cas12f1 가이드 RNA 서열에서 1개 이상의 뉴클레오티드가 치환, 결실, 삽입 또는 부가된 서열을 포함하고, 가이드 서열을 제외한 부분이 상기 야생형 Cas12f1 가이드 RNA와 50% 이상의 서열 동일성을 갖는
- [678] 벡터 시스템.
- [679] [구현예 63]
- [680] 전술한 구현예 중 어느 한 구현예에 있어서,
- [681] 상기 야생형 Cas12f1 가이드 RNA는 (i) 하나 이상의 스템(stem) 영역, (ii) tracrRNA-crRNA 상보성 영역 및 임의적으로 (iii) 연속되는 3개 이상의 유라실(U)을 포함하는 영역을 포함하는 tracrRNA(trans-activating CRISPR RNA) 및 crRNA(CRISPR RNA)를 포함하고,
- [682] 상기 엔지니어링된 가이드 RNA는 하기 (a) 내지 (d)로 이루어진 군에서 선택되는 하나 이상의 변형을 포함하는
- [683] 벡터 시스템:
- [684] (a) 하나 이상의 스템 영역의 일부 또는 전부의 결실;
- [685] (b) tracrRNA-crRNA 상보성 영역의 일부 또는 전부의 결실;
- [686] (c) tracrRNA-crRNA 상보성 영역 내에 연속되는 3개 이상의 유라실(U)이 존재하는 경우 하나 이상의 U를 A, G 또는 C로 치환; 및
- [687] (d) crRNA 서열의 3'-말단에 하나 이상의 유리딘(uridine)의 부가.
- [688] [구현예 64]
- [689] 전술한 구현예 중 어느 한 구현예에 있어서,
- [690] 상기 야생형 Cas12f1 가이드 RNA는 서열번호 11의 핵산 서열로 이루어진 tracrRNA 및 서열번호 12의 핵산 서열로 이루어진 crRNA를 포함하는

- [691] 벡터 시스템.
- [692] [구현예 65]
- [693] 전술한 구현예 중 어느 한 구현예에 있어서,
- [694] 상기 야생형 Cas12f1 가이드 RNA는 5'-말단부터 순차적으로 제1 스템 영역, 제2 스템 영역, 제3 스템 영역, 제4 스템 영역 및 tracrRNA-crRNA 상보성 영역을 포함하는 tracrRNA 및 crRNA를 포함하고,
- [695] 상기 엔지니어링된 가이드 RNA는
- [696] (a1) 제1 스템 영역의 일부 또는 전부의 결실;
- [697] (a2) 제2 스템 영역의 일부 또는 전부의 결실;
- [698] (b) tracrRNA-crRNA 상보성 영역의 일부 또는 전부의 결실;
- [699] (c) tracrRNA-crRNA 상보성 영역 내에 연속되는 3개 이상의 유라실(U)이 존재하는 경우 하나 이상의 U를 A, G 또는 C로 치환; 및
- [700] (d1) crRNA 서열의 3'-말단에 U-rich tail의 부가(상기 U-rich tail의 서열은 5'-(U_mV)_nU_o-3'로 표시되고 여기서 상기 V는 각각 독립적으로 A, C 또는 G이고, m 및 o는 1 내지 20 사이의 정수이며, n은 0 내지 5 사이의 정수임)
- [701] 로 이루어진 군에서 선택되는 하나 이상의 변형을 포함하는
- [702] 벡터 시스템.
- [703] [구현예 66]
- [704] 전술한 구현예 중 어느 한 구현예에 있어서,
- [705] 상기 엔지니어링된 가이드 RNA는 (d1) crRNA 서열의 3'-말단에 U-rich tail, (c) tracrRNA-crRNA 상보성 영역 내에 연속되는 3개 이상의 유라실(U)이 존재하는 경우 하나 이상의 U를 A, G 또는 C로 치환, 또는 상기 둘 모두의 변형을 포함하는
- [706] 벡터 시스템.
- [707] [구현예 67]
- [708] 전술한 구현예 중 어느 한 구현예에 있어서,
- [709] 상기 엔지니어링된 가이드 RNA는 (a1) 제1 스템 영역의 일부 또는 전부의 결실; (a2) 제2 스템 영역의 일부 또는 전부의 결실; 및 (b) tracrRNA-crRNA 상보성 영역의 일부 또는 전부의 결실로 이루어진 군에서 선택된 하나 이상의 변형을 포함하는
- [710] 벡터 시스템.
- [711] [구현예 68]
- [712] 전술한 구현예 중 어느 한 구현예에 있어서,
- [713] 상기 엔지니어링된 가이드 RNA는 (b1) tracrRNA-crRNA 상보성 영역의 일부의 결실을 포함하고, 상기 상보성 영역의 일부는 1개 내지 54개의 뉴클레오티드인
- [714] 벡터 시스템.
- [715] [구현예 69]
- [716] 전술한 구현예 중 어느 한 구현예에 있어서,

- [717] 상기 엔지니어링된 가이드 RNA는 (b2) tracrRNA-crRNA 상보성 영역의 전부의 결실을 포함하고, 상기 상보성 영역의 전부는 55개의 뉴클레오티드인
- [718] 시스템 또는 조성물.
- [719] [구현예 70]
- [720] 전술한 구현예 중 어느 한 구현예에 있어서,
- [721] 상기 엔지니어링된 가이드 RNA는 (a1) 제1 스템 영역의 일부 또는 전부의 결실을 포함하고, 상기 스템 영역의 일부 또는 전부는 1개 내지 20개의 뉴클레오티드인
- [722] 벡터 시스템.
- [723] [구현예 71]
- [724] 전술한 구현예 중 어느 한 구현예에 있어서,
- [725] 상기 엔지니어링된 가이드 RNA는 (a2) 제2 스템 영역의 일부 또는 전부의 결실을 포함하고, 상기 스템 영역의 일부 또는 전부는 1개 내지 27개의 뉴클레오티드인
- [726] 벡터 시스템.
- [727] [구현예 72]
- [728] 전술한 구현예 중 어느 한 구현예에 있어서,
- [729] 상기 엔지니어링된 가이드 RNA는 (a1) 제1 스템 영역의 일부 또는 전부의 결실; (d1) crRNA 서열의 3'-말단에 U-rich tail의 부가; 또는 상기 둘 모두의 변형을 포함하는
- [730] 벡터 시스템.
- [731] [구현예 73]
- [732] 전술한 구현예 중 어느 한 구현예에 있어서,
- [733] 상기 엔지니어링된 가이드 RNA는 하기 식 (I)로 표시되는 서열로 이루어지거나 상기 서열과 80% 이상 서열 동일성을 갖는 시스템 또는 조성물:
- [734]



- [735] 식 (I)에서,
- [736] X^{a} , X^{b1} , X^{b2} , X^{c1} 및 X^{c2} 는 각각 독립적으로 0 내지 35개의 (폴리)뉴클레오티드로 이루어지고,
- [737] X^{g} 는 10 내지 30개의 폴리뉴클레오티드로 이루어진 가이드 서열로서 표적 서열과 혼성화하거나 표적 서열에 상보적인 서열이고,

- [738] Lk는 길이 2 내지 20의 폴리뉴클레오티드 링커이거나 부존재하고,
 [739] $(U_mV)_nU_o$ 는 U-rich tail로서 존재하거나 부존재하고, 존재하는 경우 U는 유리딘이고, V는 각각 독립적으로 A, C 또는 G이고, m 및 o는 각각 독립적으로 1 내지 20 사이의 정수이며, n은 0 내지 5 사이의 정수이다.
- [740] [구현예 74]
 [741] 전술한 구현예 중 어느 한 구현예에 있어서,
 [742] 상기 X^a 는 서열번호 14의 핵산 서열을 포함하거나 상기 서열번호 14의 서열에서 1 내지 20개의 뉴클레오티드가 결실된 핵산 서열을 포함하는
 [743] 벡터 시스템.
- [744] [구현예 75]
 [745] 전술한 구현예 중 어느 한 구현예에 있어서,
 [746] 상기 X^{b1} 은 서열번호 25의 핵산 서열을 포함하거나 상기 서열번호 25의 서열에서 1 내지 13개의 뉴클레오티드가 결실된 핵산 서열을 포함하는
 [747] 벡터 시스템.
- [748] [구현예 76]
 [749] 전술한 구현예 중 어느 한 구현예에 있어서,
 [750] 상기 X^{b2} 은 서열번호 29의 핵산 서열을 포함하거나 상기 서열번호 29의 서열에서 1 내지 14개의 뉴클레오티드가 결실된 핵산 서열을 포함하는
 [751] 벡터 시스템.
- [752] [구현예 77]
 [753] 전술한 구현예 중 어느 한 구현예에 있어서,
 [754] 상기 식 (I) 내의 서열 5'- X^{b1} UUAGX ^{b2} -3'은 서열번호 34 내지 서열번호 38로 이루어진 군에서 선택된 어느 하나의 핵산 서열인
 [755] 벡터 시스템.
- [756] [구현예 78]
 [757] 전술한 구현예 중 어느 한 구현예에 있어서,
 [758] 상기 X^{c1} 은 서열번호 39의 핵산 서열을 포함하거나 상기 서열번호 39의 서열에서 1 내지 28개의 뉴클레오티드가 결실된 핵산 서열을 포함하는
 [759] 벡터 시스템.
- [760] [구현예 79]
 [761] 전술한 구현예 중 어느 한 구현예에 있어서,
 [762] 상기 X^{c1} 서열 내에 연속되는 3개 이상의 유라실(U)이 존재하는 경우 이들 중 하나 이상의 U가 A, G 또는 C로 치환되는 변형을 포함하는
 [763] 벡터 시스템.
- [764] [구현예 80]
 [765] 전술한 구현예 중 어느 한 구현예에 있어서,

- [766] 상기 X^{c2} 는 서열번호 58의 핵산 서열을 포함하거나 상기 서열번호 58의 서열에서 1 내지 27개의 뉴클레오티드가 결실된 핵산 서열을 포함하는
- [767] 벡터 시스템.
- [768] [구현예 81]
- [769] 전술한 구현예 중 어느 한 구현예에 있어서,
- [770] 상기 X^{c2} 서열 내에 서열 5'-ACGAA-3'가 존재하는 경우 해당 서열이 5'-NGNNN-3'로 치환되고, 상기 N은 각각 독립적으로 A, C, G 또는 U인
- [771] 벡터 시스템.
- [772] [구현예 82]
- [773] 전술한 구현예 중 어느 한 구현예에 있어서,
- [774] 상기 식 (I) 내의 서열 5'- X^{c1} -Lk- X^{c2} -3'은 서열번호 80 내지 서열번호 86으로 이루어진 군에서 선택된 어느 하나의 핵산 서열인
- [775] 벡터 시스템.
- [776] [구현예 83]
- [777] 전술한 구현예 중 어느 한 구현예에 있어서,
- [778] 상기 Lk는 5'-GAAA-3', 5'-UUAG-3', 5'-UGAAAA-3', 5'-UUGAAAAA-3', 5'-UUCGAAAGAA-3'(서열번호 76), 5'-UUCAGAAAUGAA-3'(서열번호 77), 5'-UUCAUGAAAAUGAA-3'(서열번호 78) 및 5'-UUCAUUGAAAAAUGAA-3'(서열번호 79)로 이루어진 군에서 선택된 어느 하나의 핵산 서열을 포함하는
- [779] 벡터 시스템.
- [780] [구현예 84]
- [781] 전술한 구현예 중 어느 한 구현예에 있어서,
- [782] 상기 $(U_mV)_nU_o$ 는 (i) n이 0이고, o가 1 내지 6 사이의 정수이거나 (ii) V가 A 또는 G이고, m 및 o는 각각 독립적으로 3 내지 6 사이의 정수이고, n은 1 내지 3 사이의 정수인
- [783] 벡터 시스템.
- [784] [구현예 85]
- [785] 전술한 구현예 중 어느 한 구현예에 있어서,
- [786] 상기 엔지니어링된 가이드 RNA는 서열번호 87 내지 서열번호 132로 이루어진 군에서 선택된 어느 하나의 핵산 서열로 이루어진 엔지니어링된 tracrRNA를 포함하는
- [787] 벡터 시스템.
- [788] [구현예 86]
- [789] 전술한 구현예 중 어느 한 구현예에 있어서,
- [790] 상기 엔지니어링된 가이드 RNA는 서열번호 133 내지 서열번호 148로 이루어진 군에서 선택된 어느 하나의 핵산 서열로 이루어진 엔지니어링된 crRNA를 포함하는

- [791] 벡터 시스템.
- [792] [구현예 87]
- [793] 전술한 구현예 중 어느 한 구현예에 있어서,
- [794] 상기 엔지니어링된 가이드 RNA는 듀얼 가이드 RNA 또는 싱글 가이드 RNA인
- [795] 벡터 시스템.
- [796] [구현예 88]
- [797] 전술한 구현예 중 어느 한 구현예에 있어서,
- [798] 상기 엔지니어링된 싱글 가이드 RNA는 서열번호 149 내지 서열번호 186으로 이루어진 군에서 선택된 어느 하나의 핵산 서열로 이루어진
- [799] 벡터 시스템.
- [800] [구현예 89]
- [801] 전술한 구현예 중 어느 한 구현예에 따른 벡터 시스템에 의해 제조된 바이러스.
- [802] [구현예 90]
- [803] 전술한 구현예 중 어느 한 구현예에 있어서,
- [804] 상기 바이러스는 레트로바이러스, 렌티바이러스, 아데노바이러스, 아데노-연관 바이러스, 백시니아바이러스, 폭스바이러스, 단순포진 바이러스 및 파지(phage)로 구성된 군에서 선택되는 것인
- [805] 바이러스.
- [806] [구현예 91]
- [807] 전술한 구현예 중 어느 한 구현예에 따른 바이러스를 포함하는 조성물.
- [808] [구현예 92]
- [809] 전술한 구현예 중 어느 한 구현예의 시스템 또는 조성물, 또는 전술한 구현예 중 어느 한 구현예의 벡터 시스템을 세포와 접촉시키거나 세포 내에서 발현시키는 단계; 및
- [810] 표적 핵산 또는 그 인접 부위에서 공여자 핵산 분자를 주형으로 하여 이중가닥 절단의 복구 과정에 의해 목적 서열이 도입되는 단계를 포함하는
- [811] 세포 내 이중가닥 핵산 상의 표적 부위로 목적 서열을 도입하는 방법.
- [812] [구현예 93]
- [813] 전술한 구현예 중 어느 한 구현예에 있어서,
- [814] 상기 이중가닥 절단의 복구 과정은 상동지정복구 기작에 의하는 것인
- [815] 방법.
- [816] [구현예 94]
- [817] 전술한 구현예 중 어느 한 구현예에 있어서,
- [818] 상기 세포는 표적 핵산 또는 표적 유전자가 존재하는 원핵 세포 또는 진핵 세포인
- [819] 방법.
- [820] [구현예 95]
- [821] 전술한 구현예 중 어느 한 구현예에 있어서,

- [822] 상기 진핵 세포는 효모(yeast), 곤충 세포, 식물 세포, 비인간-동물 세포 또는 인간 세포인
- [823] 방법.
- [824] [구현예 96]
- [825] 전술한 구현예 중 어느 한 구현예에 있어서,
- [826] 상기 벡터 시스템은 레트로바이러스, 렌티바이러스, 아데노바이러스, 아데노-연관 바이러스, 백시니아바이러스, 폭스바이러스, 단순포진 바이러스 및 파지(phage)로 구성된 군에서 선택된 패키징 바이러스에 도입되어 패키징 바이러스에 의해 생성된 바이러스 형태로 원핵 세포 또는 진핵 세포 내로 전달되는
- [827] 방법.
- [828] [구현예 97]
- [829] 전술한 구현예 중 어느 한 구현예에 있어서,
- [830] 상기 벡터 시스템은 전기천공법, 유전자총, 초음파천공법, 자기주입법(magnetofection), 일시적인 세포 압축 또는 스퀴징 방법, 양이온성 리포솜법, 초산 리튬-DMSO, 지질-매개 형질감염(transfection), 인산칼슘 침전법(precipitation), 리포펙타민(lipofection), PEI(Polyethyleneimine)-매개 형질감염, DEAE-dextran 매개 형질감염 또는 나노파티클-매개 핵산 전달에 의해 원핵 세포 또는 진핵 세포 내로 전달되는
- [831] 방법.
- [832] [구현예 98]
- [833] 전술한 구현예 중 어느 한 구현예에 있어서,
- [834] 상기 벡터 시스템은 하나 이상의 지질 나노입자(LNP)를 통해 원핵 세포 또는 진핵 세포 내로 직접 전달되는
- [835] 방법.
- [836] [구현예 99]
- [837] 전술한 구현예 중 어느 한 구현예에 있어서,
- [838] 상기 접촉 또는 발현은 생체 내 또는 생체 외에서 일어나는
- [839] 방법.
- [840] 이하, 실시예를 통해 본 명세서가 제공하는 발명에 대해 더욱 상세히 설명한다. 이들 실시예는 오로지 본 명세서에 의해 개시되는 내용을 예시하기 위한 것으로, 본 명세서에 의해 개시되는 내용의 범위가 이들 실시예에 의해 제한되는 것으로 해석되지 않는 것은 이 기술 분야에서 통상의 지식을 가진 자에게 있어서 자명할 것이다.
- [841] 실시예
- [842] 실시예 1. 상동지정복구를 위한 핵산 편집 시스템의 구성요소 제작
- [843] 실시예 1.1. Cas12f1, TnpB 또는 이의 변이체 단백질을 암호화하는 인간 코돈-최적화된 핵산

- [844] 본 발명의 상동지정복구를 위한 핵산 편집 시스템은 일 구성요소로서 Cas12f1, TnpB 또는 이의 변이체 단백질(즉, Cas12f1 변이체 단백질 또는 TnpB 변이체 단백질)을 포함한다. Cas12f1, TnpB 또는 이의 변이체 단백질은 서열번호 1 내지 서열번호 5로 이루어진 군에서 선택된 아미노산 서열과 70% 이상의 서열 동일성을 갖는 아미노산 서열을 포함한다. 바람직하게, Cas12f1 단백질은 서열번호 5의 아미노산 서열을 포함하거나 이로 이루어진 단백질을 포함하며, TnpB 단백질은 서열번호 1의 아미노산을 포함하거나 이로 이루어진 단백질을 포함한다(이 때, TnpB는 CWCas12f1으로도 분류 및 명명될 수 있다).
- [845] 또한, Cas12f1의 변이체 또는 TnpB의 변이체 단백질은 서열번호 1의 아미노산 서열을 기준으로 N-말단으로부터 1개 내지 28개의 아미노산이 제거 또는 치환된 아미노산 서열을 포함하거나 이로 이루어진 단백질을 포함한다. 상기 서열번호 1의 아미노산 서열을 기준으로 N-말단으로부터 1개 내지 28개의 아미노산이 제거 또는 치환된 아미노산 서열로 이루어진 Cas12f1 변이체 또는 TnpB 변이체 단백질의 대표예로서 본 명세서에서는 Cas12f1의 N-말단에 CasX의 N-말단 26aa를 포함하는 TnpB-v1 단백질(서열번호 2), 28aa 무작위 서열을 포함하는 TnpB-v2 단백질(서열번호 3) 또는 26aa 무작위 서열을 포함하는 TnpB-v3 단백질(서열번호 4)이 제공된다. 또한, Cas12f1의 변이체 또는 TnpB의 변이체 단백질은 서열번호 1의 아미노산 서열을 포함하거나 이로 이루어진 TnpB의 N-말단 또는 C-말단에 1개 내지 600개의 아미노산이 추가된 아미노산 서열로 이루어진 단백질을 포함한다. 여기서, N-말단 또는 C-말단에 추가되는 1개 내지 600개의 아미노산은 서열번호 294 또는 서열번호 295의 아미노산 서열을 포함하거나 이로 이루어질 수 있으며, 상기 추가되는 서열과 상기 변이체 단백질 사이에는 1개 이상의 NLS 서열이 더 포함될 수 있다. 또한, 다른 종에서 유래한 TnpB에 해당하는 TnpB 변이체 단백질은 서열번호 202 내지 서열번호 293 중에서 선택된 어느 하나의 아미노산 서열을 포함하거나 이로 이루어진 단백질일 수 있다.
- [846] 인간 세포에서 발현하는 표적 핵산 편집 시스템 및 핵산 절단을 위한 표적 핵산 편집을 암호화하는 핵산 구조물을 구축하기 위해서, 코돈 최적화 프로그램을 이용하여 Cas12f1, TnpB 또는 이의 변이체 단백질에 대한 인간 코돈-최적화된 유전자를 얻었다. 이와 같이 제작한 Cas12f1, TnpB 또는 이의 변이체 단백질을 암호화하는 인간 코돈-최적화된 염기서열은 각각 서열번호 6(TnpB), 서열번호 7(TnpB-v1), 서열번호 8(TnpB-v2), 서열번호 9(TnpB-v3) 및 서열번호 10(Cas12f1)에 나타내었다.
- [847] 하기 표 4에는 위에서 제작된 Cas12f1, TnpB 또는 이의 변이체 단백질의 아미노산 서열을 나타냈다. 또한, 표 5에는 Cas12f1, TnpB 또는 이의 변이체 단백질을 각각 암호화하는 인간 코돈-최적화된 핵산의 염기서열을 나타냈다. 이들은 실시예에서 표적 핵산 편집 시스템을 구성하는 핵산 편집 단백질을 암호화하는 핵산으로 사용되었다.

[848] [표4]

명칭	아미노산 서열	서열 번호
TnpB 단백질	MGEKSSRRRRNGKSGAWTAAITSCVGGKMAKNTI TKTLKLRIVRPYNsAEVEKIVADEKNNREKIALEKN KDKVKEACSKHLKVAAYCTTQVERNACLFCKARK LDDKFYQKLRGQFPDAVFWQEISEIFRQLQKQAAEI YNQSLIELYYEIFIKGKGIANASSVEHYLSDVCYTR AAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPT TKSDNFPIPLVKQKGGQYTGFESNHNSDFIIKIPFGR WQVKKEIDKYRPWEKFDFEQVQKSPKPIsLLLSTQ RRKRNGWSKDEGTEAEIKKVMNGDYQTSYIEVK RGSKIGEKSAWMLNLSIDVPKIDKGVDPSIIGGIDVG VKSPLVCAINNAFSRYSISDNDLFHFNKKMFARRRI LLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLI ERWACEIADFFIKNKVGTVQMENLESMKRKEDSYF NIRLRGFWPYAEMQNKIEFKLKQYGIEIRKVAPNNT SKTCSKCGHLNNYFNFEYRKKKNKFPHPKCEKCNFK ENADYNAALNISNPCLKSTKEEP	1
TnpB-v1 단백질	MEKRINKIRKKLSADNATKPVSRSGPMAKNTITKTL KLRIVRPYNsAEVEKIVADEKNNREKIALEKNKDK VKEACSKHLKVAAYCTTQVERNACLFCKARKLDD KFYQKLRGQFPDAVFWQEISEIFRQLQKQAAEIYNQ SLIELYYEIFIKGKGIANASSVEHYLSDVCYTRAEL FKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSD NFPIPLVKQKGGQYTGFESNHNSDFIIKIPFGRWQV KKEIDKYRPWEKFDFEQVQKSPKPIsLLLSTQRRKR NKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKI GEKSAWMLNLSIDVPKIDKGVDPSIIGGIDVGVKSP LVCAINNAFSRYSISDNDLFHFNKKMFARRRILLKK NRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWA CEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLR GFWPYAEMQNKIEFKLKQYGIEIRKVAPNNTSKTC SKCGHLNNYFNFEYRKKKNKFPHPKCEKCNFKENA DYNAALNISNPCLKSTKEEP	2
TnpB-v2 단백질	MAGGPGAGSAAPVSSTSSLPLAALNMRVMAKNTIT KTLKLRIVRPYNsAEVEKIVADEKNNREKIALEKNK DKVKEACSKHLKVAAYCTTQVERNACLFCKARKL	3

	DDKFYQKLRGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSDVCYTRA AELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGFESNHNSDFIIKIPFGR WQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQRRKRNGWSKDEGTEAEIKKVMNGDYQTSYIEVK RGSKIGEKS AWMLNLSIDVPKIDKGVDPSIIGGIDVG VKSPLVCAINNAFSRYSISDNDLFHFNKKMFARRRI LLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIER WACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRKVAPNNT SKTCSKCGHLNNYFNFEYRKKNKFPHPKCEKCNFK ENADYNAALNISNPKLKSTKEEP	
TnpB-v3 단백질	MAGGPGAGS AAPVSSTSSLPLAALNMMAKNTITKT LKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKD KVKEACSKHLKVAAYCTTQVERNACLFCKARKLD DDKFYQKLRGQFPDAVFWQEISEIFRQLQKQAAEIYN QSLIELYYEIFIKGKGIANASSVEHYLSDVCYTRA AE LFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKS DNFPIPLVKQKGGQYTGFESNHNSDFIIKIPFGRWQ VKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQRRK RNGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGS KIGEKS AWMLNLSIDVPKIDKGVDPSIIGGIDVG GVKSPLVCAINNAFSRYSISDNDLFHFNKKMFARRRIL LLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIER WACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIR LRGFWPYAEMQNKIEFKLKQYGIEIRKVAPNNTS KTCSKCGHLNNYFNFEYRKKNKFPHPKCEKCNFKE NADYNAALNISNPKLKSTKEEP	4
Cas12f1 단백질	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNRE KIALEKNKDKVKEACSKHLKVAAYCTTQVERNAC LFCKARKLDDKFYQKLRGQFPDAVFWQEISEIFRQL QKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLS DVCYTRA AE LFKNAAIASGLRSKIKSNFRLKELKN MKSGLPTTKSDNFPIPLVKQKGGQYTGFESNHNSD FIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPI SLLLSTQRRKRNGWSKDEGTEAEIKKVMNGDYQ	5

TSYIEVKRGSKIGEKS IGGIDVGVSPLVCAINNAFSRYSISDNDLFHFNKK MFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSE RFRKKLIERWACEIADFFIKNKVGTVQMENLESMK RKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIR KVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFK CEKCNFKENADYNAALNISNPCLKSTKEEP

[849]

[850] [5]

명칭	염기서열 (5' to 3')	서열 번호
TnpB 단백질을 암호화하 는 인간 코돈- 최적화된 핵산	ATGGGGGAGAAAAGTTCCCGCCGCCGACGGAATGGAAAAAGCGGTGCGTGACTGCTG CTATAACAAGCTGTGTTGGGGGTAAAGATGGCCAAGAACAATTACAAAGACACTGAA GCTGAGGATCGTGAGACCATAACAACAGCGCTGAGGTGCGAGAAGATTGTGGCTGATGAA AAGAACAACAGGGAAAAGATCGCCCTCGAGAAGAACAAGGATAAGGTGAAGGAGGCCT GCTCTAAGCACCTGAAAGTGGCCGCCTACTGCACCACACAGGTGGAGAGGAACGCCTG TCTGTTTTGTAAAGCTCGGAAGCTGGATGATAAGTTTTACCAGAAGCTGCGGGGCCAG TTCCCGATGCCGTCTTTTGGCAGGAGATTAGCGAGATCTTCAGACAGCTGCAGAAGC AGGCCGCCGAGATCTACAACAGAGCCTGATCGAGCTCTACTACGAGATCTTCATCAA GGGCAAGGGCATTGCCAACGCCTCCTCCGTGGAGCACTACCTGAGCGACGTGTGCTAC ACAAGAGCCGCCGAGCTCTTTAAGAACGCCGCTATCGCTTCCGGGTGAGGAGCAAGA TTAAGAGTAACTTCCGGCTCAAGGAGCTGAAGAACATGAAGAGCGGCCTGCCACTAC AAAGAGCGACAACCTCCCAATTCCACTGGTGAAGCAGAAGGGGGGCCAGTACACAGGG TTCGAGATTTCCAACCAACAGCGACTTTATTATTAAGATCCCCCTTGGCAGGTGGC AGGTCAAGAAGGAGATTGACAAGTACAGGCCCTGGGAGAAGTTTGATTTTCGAGCAGGT GCAGAAGAGCCCCAAGCCTATTTCCCTGCTGCTGTCCACACAGCGGCGGAAGAGGAAC AAGGGGTGGTCTAAGGATGAGGGACCGAGGCCGAGATTAAGAAAGTGATGAACGGCG ACTACCAGACAAGCTACATCGAGGTCAAGCGGGGCAGTAAGATTGGCGAGAAGAGCGC CTGGATGCTGAACCTGAGCATTGACGTGCCAAAGATTGATAAGGGCGTGATCCAGC ATCATCGGAGGGATCGATGTGGGGGTCAAGAGCCCCCTCGTGTGCGCCATCAACAACG CCTTCAGCAGGTACAGCATCTCCGATAACGACCTGTTCCACTTTAACAGAAGATGTT CGCCCGCGGAGGATTTTGCTCAAGAAGAACCAGCACAGCGGGCCGACACGGGGCC AAGAACAAGCTCAAGCCCATCACTATCCTGACCGAGAAGAGCGAGAGGTTTCAGGAAGA AGCTCATCGAGAGATGGGCCTGCGAGATCGCCGATTTCTTTATTAAGAACAAGTTCGG AACAGTGCAGATGGAGAACCTCGAGAGCATGAAGAGGAAGGAGGATTCTACTTCAAC ATTCGGCTGAGGGGGTTCTGGCCCTACGCTGAGATGCAGAACAAGATTGAGTTTAAGC TGAAGCAGTACGGGATTGAGATCCGAAGGTGGCCCCAACAAACACCAGCAAGACCTG CAGCAAGTGCGGGCACCTCAACAATACTTCAACTTCGAGTACCGGAAGAAGAACAAG TTCCCACACTTCAAGTGCGAGAAGTGCAACTTTAAGGAGAACGCCGATTACAACGCCG CCCTGAACATCAGCAACCCTAAGCTGAAGAGCACTAAGGAGGAGCCC	6
	ATGGAAGAGAGAATCAACAAGATCAGGAAGAAGCTGAGCGCCGACAACGCCACCAAGC CTGTGTCTAGGAGTGGCCCCATGGCCAAGAACAATTACAAAGACACTGAAGCTGA GGATCGTGAGACCATAACAACAGCGCTGAGGTGCGAGAAGATTGTGGCTGATGAAAAGAA CAACAGGGAAAAGATCGCCCTCGAGAAGAACAAGGATAAGGTGAAGGAGGCCTGCTCT AAGCACCTGAAAGTGGCCGCCTACTGCACCACACAGGTGGAGAGGAACGCCTGTCTGT TTTGTAAGCTCGGAAGCTGGATGATAAGTTTTACCAGAAGCTGCGGGGCCAGTTCCC	

[851]

TnpB-v1 단백질을 암호화하 는 인간 코돈- 최적화된 핵산	CGATGCCGTCTTTTGGCAGGAGATTAGCGAGATCTTCAGACAGCTGCAGAAGCAGGCC GCCGAGATCTACAACCAGAGCCTGATCGAGCTCTACTACGAGATCTTCATCAAGGGCA AGGGCATTGCCAACGCCTCCTCCGTGGAGCACTACCTGAGCGACGTGTGCTACACAAG AGCCGCCGAGCTCTTTAAGAACGCCGCTATCGCTTCCGGGCTGAGGAGCAAGATTAAG AGTAACTTCCGGCTCAAGGAGCTGAAGAACATGAAGAGCGGCCTGCCACTACAAAGA GCGACAACCTTCCCAATTCCACTGGTGAAGCAGAAGGGGGGCCAGTACACAGGGTTTCA GATTTCCAACCACAACAGCGACTTTATTATTAAGATCCCCCTTTGGCAGGTGGCAGGTC AAGAAGGAGATTGACAAGTACAGGCCCTGGGAGAAGTTTGATTTCGAGCAGGTGCAGA AGAGCCCCAAGCCTATTTCCCTGCTGCTGTCCACACAGCGGCGGAAGAGGAACAAGGG GTGGTCTAAGGATGAGGGGACCGAGGCCGAGATTAAGAAAGTGATGAACGGCGACTAC CAGACAAGCTACATCGAGGTCAAGCGGGCAGTAAGATTGGCGAGAAGAGCGCCTGGA TGCTGAACCTGAGCATTGACGTGCCAAAGATTGATAAGGGCGTGGATCCCAGCATCAT CGGAGGGATCGATGTGGGGGTCAAGAGCCCCCTCGTGTGCCCATCAACAACGCCTTC AGCAGGTACAGCATCTCCGATAACGACCTGTTCCACTTTAACAAGAAGATGTTGCCCC GGCGGAGGATTTTGCTCAAGAAGAACCGGCACAAGCGGGCCGACACGGGGCCAAGAA CAAGCTCAAGCCCATCACTATCCTGACCGAGAAGAGCGAGAGGTTTCAGGAAGAAGCTC ATCGAGAGATGGGCCTGCGAGATCGCCGATTTCTTTATTAAGAACAAGGTCCGAACAG TGCAGATGGAGAACCTCGAGAGCATGAAGAGGAAGGAGGATTCTACTTCAACATTG GCTGAGGGGGTTCTGGCCCTACGCTGAGATGCAGAACAAAGATTGAGTTTAAGCTGAAG CAGTACGGGATTGAGATCCGGAAGGTGGCCCCCAACAACACAGCAAGACCTGCAGCA AGTGCGGGCACCTCAACAACCTACTTCAACTTCGAGTACCGGAAGAAGAACAAGTTCCC ACACTTCAAGTGCGAGAAGTGCAACTTTAAGGAGAAGCGCGATTACAACGCCGCCCTG AACATCAGCAACCCCTAAGCTGAAGAGCACTAAGGAGGAGCCC	7
TnpB-v2 단백질을 암호화하 는 인간 코돈- 최적화된 핵산	ATGGCTGGCGGACCAGGCGCAGGTAGTGCTGCGCCAGTTCTTCAACTTCCTCCCTGC CCCTGGCTGCGCTTAACATGCGCGTGATGGCCAAGAACAATTACAAAGACACTGAA GCTGAGGATCGTGAGACCATACAACAGCGCTGAGGTCCGAGAAGATTGTGGCTGATGAA AAGAACAACAGGGAAAAGATCGCCCTCGAGAAGAACAAGGATAAGGTGAAGGAGGCCT GCTCTAAGCACCTGAAAGTGGCCGCTACTGCACCACACAGGTGGAGAGGAACGCCTG TCTGTTTTGTAAAGCTCGGAAGCTGGATGATAAGTTTTACCAGAAGCTGCGGGGCCAG TTCCCCGATGCCGTCTTTTGGCAGGAGATTAGCGAGATCTTCAGACAGCTGCAGAAGC AGGCCGCCGAGATCTACAACCAGAGCCTGATCGAGCTCTACTACGAGATCTTCATCAA GGGCAAGGGCATTGCCAACGCCTCCTCCGTGGAGCACTACCTGAGCGACGTGTGCTAC ACAAGAGCCGCCGAGCTCTTTAAGAACGCCGCTATCGCTTCCGGGCTGAGGAGCAAGA TTAAGAGTAACTTCCGGCTCAAGGAGCTGAAGAACATGAAGAGCGGCCTGCCACTAC AAAGAGCGACAACCTTCCCAATTCCACTGGTGAAGCAGAAGGGGGGCCAGTACACAGGG TTTCGAGATTTCCAACCACAACAGCGACTTTATTATTAAGATCCCCCTTTGGCAGGTGGC AGGTCAAGAAGGAGATTGACAAGTACAGGCCCTGGGAGAAGTTTGATTTTCGAGCAGGT GCAGAAGAGCCCCAAGCCTATTTCCTGCTGCTGTCCACACAGCGCGGAAGAGGAAC AAGGGGTGGTCTAAGGATGAGGGGACCGAGGCCGAGATTAGAAAGTGATGAACGGCG ACTACCAGACAAGCTACATCGAGGTCAAGCGGGGCAGTAAGATTGGCGAGAAGAGCGC CTGGATGCTGAACCTGAGCATTGACGTGCCAAAGATTGATAAGGGCGTGGATCCCAGC	8

[852]

TnpB-v3 단백질을 암호화하 는 인간 코돈- 최적화된 핵산	ATCATCGGAGGGATCGATGTGGGGGTCAAGAGCCCCCTCGTGTGCGCCATCAACAACG CCTTCAGCAGGTACAGCATCTCCGATAACGACCTGTTCCACTTTAACAAGAAGATGTT CGCCCGGCGGAGGATTTTGCTCAAGAAGAACCGGCACAAGCGGGCCGGACACGGGGCC AAGAACAAGCTCAAGCCCATCACTATCCTGACCGAGAAGAGCGAGAGGTTTCAGGAAGA AGCTCATCGAGAGATGGGCCCTGCGAGATCGCCGATTTCTTTATTAAGAACAAGGTCCG AACAGTGCAGATGGAGAACCCTCGAGAGCATGAAGAGGAAGGAGGATTCTACTTCAAC ATTGGCTGAGGGGGTTCTGGCCCTACGCTGAGATGCAGAACAAGATTGAGTTTAAGC TGAAGCAGTACGGGATTGAGATCCGGAAGGTGGCCCCAACAACACCAGCAAGACCTG CAGCAAGTGCGGGCACCTCAACAACCTACTTCAACTTCGAGTACCGGAAGAAGAACAAG TTCCCACACTTCAAGTGCAGAGAAGTGCAACTTTAAGGAGAACGCCGATTACAACGCCG CCCTGAACATCAGCAACCCTAAGCTGAAGAGCACTAAGGAGGAGCCC ATGGCTGGCGGACCAGGCGCAGGTAGTGCTGCGCCAGTTTCTTCAACTTCCTCCCTGC CCCTGGCTGCGCTTAACATGATGGCCAAGAACAATTACAAAGACACTGAAGCTGAG GATCGTGAGACCATAACAACAGCGCTGAGGTCGAGAAGATTGTGGCTGATGAAAAAAC AACAGGGAAGAATCGCCCTCGAGAAGAACAAGGATAAGGTGAAGGAGGCCCTGCTCTA AGCACCTGAAAGTGGCCGCCCTACTGCACCACACAGGTGGAGAGGAACGCCCTGTCGT TTGTAAAGCTCGGAAGCTGGATGATAAGTTTACCAGAAGCTGCGGGGCCAGTTCCCC GATGCGGTCTTTTGGCAGGAGATTAGCGAGATCTTCAGACAGCTGCAGAAGCAGGCCG CCGAGATCTACAACCAGAGCCTGATCGAGCTCTACTACGAGATCTTCATCAAGGGCAA GGGCATTGCCAACGCCCTCCTCCGTGGAGCACTACCTGAGCGACGTGTGCTACACAAGA GCCGCCGAGCTCTTTAAGAACGCCGCTATCGCTTCCGGGCTGAGGAGCAAGATTAAGA GTAACCTTCCGGCTCAAGGAGCTGAAGAACATGAAGAGCGGCCTGCCCACTACAAAGAG CGACAACCTTCCCAATTCCACTGGTGAAGCAGAAGGGGGGCCAGTACACAGGGTTCGAG ATTTCCAACCACAACAGCGACTTTATTATTAAGATCCCCTTTGGCAGGTGGCAGGTCA AGAAGGAGATTGACAAGTACAGGCCCTGGGAGAAGTTTGATTTTCGAGCAGGTGCAGAA GAGCCCCAAGCCTATTTCCCTGTGTGTCCACACAGCGGCGGAAGAGGAACAAGGGG TGGTCTAAGGATGAGGGGACCAGGCCGAGATTAAGAAAGTGATGAACGGCGACTACC AGACAAGCTACATCGAGGTCAAGCGGGCAGTAAGATTGGCGAGAAGAGCGCCTGGAT GCTGAACCTGAGCATTGACGTGCCAAAGATTGATAAGGGCGTGGATCCCAGCATCATC GGAGGGATCGATGTGGGGGTCAAGAGCCCCCTCGTGTGCGCCATCAACAACGCCCTTCA GCAGGTACAGCATCTCCGATAACGACCTGTTCCACTTTAACAAGAAGATGTTGCCCCG GCGGAGGATTTTGCTCAAGAAGAACCGGCACAAGCGGGCCGGACACGGGGCCAAGAAC AAGCTCAAGCCCATCACTATCCTGACCGAGAAGAGCGAGAGGTTTCAGGAAGAAGCTCA TCGAGAGATGGGCCCTGCGAGATCGCCGATTTCTTTATTAAGAACAAGGTCCGGAACAGT GCAGATGGAGAACCCTCGAGAGCATGAAGAGGAAGGAGGATTCTACTTCAACATTCCG CTGAGGGGGTTCTGGCCCTACGCTGAGATGCAGAACAAGATTGAGTTTAAGCTGAAGC AGTACGGGATTGAGATCCGGAAGGTGGCCCCAACAACACCAGCAAGACCTGCAGCAA GTGCGGGCACCTCAACAACCTACTTCAACTTCGAGTACCGGAAGAAGAACAAGTTCCCA CACTTCAAGTGCAGAGAAGTGCAACTTTAAGGAGAACGCCGATTACAACGCCGCCCTGA ACATCAGCAACCCTAAGCTGAAGAGCACTAAGGAGGAGCCC	9
---	---	----------

[853]

Cas12f1 단백질을 암호화하 는 인간 코돈- 최적화된 핵산	ATGGCCAAGAACAATTACAAAGACACTGAAGCTGAGGATCGTGAGACCATACAACA GCGCTGAGGTCGAGAAGATTGTGGCTGATGAAAAGAACACAGGGAAAAGATCGCCCT CGAGAAGAACAAGGATAAGGTGAAGGAGGCCTGCTCTAAGCACCTGAAAGTGGCCGCC TACTGCACCACACAGGTGGAGAGGAACGCCTGTCTGTTTTGTAAAGCTCGGAAGCTGG ATGATAAGTTTTACCAGAAGCTGCGGGGCCAGTTCCCGATGCCGTCTTTTGGCAGGA GATTAGCGAGATCTTCAGACAGCTGCAGAAGCAGGCCGCCGAGATCTACAACCAGAGC CTGATCGAGCTCTACTACGAGATCTTCATCAAGGGCAAGGGCATTGCCAACGCCTCCT CCGTGGAGCACTACCTGAGCGACGTGTGCTACACAAGAGCCGCCGAGCTCTTTAAGAA CGCCGCTATCGCTTCCGGGCTGAGGAGCAAGATTAAGAGTAACCTCCGGCTCAAGGAG CTGAAGACATGAAGAGCGGCCTGCCACTACAAAGAGCGACAACCTCCCAATTCCAC TGGTGAAGCAGAAGGGGGGCCAGTACACAGGGTTCGAGATTTCCAACCACAACAGCGA CTTTATTATTAAGATCCCTTTGGCAGGTGGCAGGTCAAGAAGGAGATTGACAAGTAC AGGCCCTGGGAGAAGTTTGATTTTCGAGCAGGTGCAGAAGAGCCCCAAGCCTATTTCCC TGCTGCTGTCCACACAGCGCGGAAGAGGAACAAGGGGTGGTCTAAGGATGAGGGGAC CGAGGCCGAGATTAAGAAAGTGATGAACGGCGACTACCAGACAAGCTACATCGAGGTC AAGCGGGCAGTAAGATTGGCGAGAAGAGCGCCTGGATGCTGAACCTGAGCATTGACG TGCCAAAGATTGATAAGGGCGTGGATCCAGCATCATCGGAGGGATCGATGTGGGGGT CAAGAGCCCCCTCGTGTGCGCCATCAACAACGCCTTCAGCAGGTACAGCATCTCCGAT AACGACCTGTTCCACTTTAACAAGAAGATGTTCCGCCGCGGAGGATTTTGCTCAAGA AGAACCAGCACAAGCGGGCCGACACGGGGCCAAGAACAAGCTCAAGCCCATCACTAT CCTGACCGAGAAGAGCGAGAGGTTTCAGGAAGAAGCTCATCGAGAGATGGGCCTGCGAG ATCGCCGATTTCTTTATTAAGAACAAGGTTCGGAACAGTGCAGATGGAGAACCTCGAGA GCATGAAGAGGAAGGAGGATTCCTACTTCAACATTCGGCTGAGGGGGTCTGTGCCCTA CGCTGAGATGCAGAACAAGATTGAGTTTAAGCTGAAGCAGTACGGGATTGAGATCCGG AAGGTGGCCCCCAACAACACCAGCAAGACCTGCAGCAAGTGCGGGCACCTCAACAAC ACTTCAACTTCGAGTACCGGAAGAAGAACAAGTTCACACTTCAAGTGCAGAGAAGTG CAACTTAAGGAGAACGCCGATTACAACGCCGCCCTGAACATCAGCAACCCTAAGCTG AAGAGCACTAAGGAGGAGCCC	10
--	---	----

[854] 상기 제작된 초소형 유전자 편집 핵산 구조물을 다음의 방법으로 제조하였다. 본 발명에 사용된 상기 핵산 구조물은 인간 코돈-최적화된 Cas12f1, TnpB 또는 이의 변이체(엔지니어링된 변이체 포함)의 유전자 서열을 포함한다. 상기 유전자 서열을 주형으로 PCR 증폭을 진행하고, Gibson assembly 방법에 의해 진핵 세포 시스템(eukaryotic cell system)에서 발현이 가능한 프로모터와 poly(A) 신호 서열(signal sequence)을 가지는 벡터에 원하는 클로닝(cloning) 서열에 맞게 클로닝을 진행하였다. 클로닝 후, 얻어진 재조합 플라스미드 벡터의 서열은 생거 시퀀싱(Sanger sequencing) 방법을 통하여 최종 확인하였다.

[855] 실시예 1.2. Cas12f1 및 TnpB 단백질의 발현 및 정제

[856] 상기 실시예 1.1에서 제조한 유전자를 발현시키고, 단백질을 정제하였다.

[857] 먼저 상기 핵산 구조물을 pMAL-c2 플라스미드 벡터에 클로닝하여 BL21(DE3) *E. coli* 세포에 형질전환하였다. 상기 형질전환된 *E. coli* 콜로니를 광학 밀도가 0.7에 도달할 때까지 37°C의 LB broth에서 성장시켰다. 상기 형질전환된 *E. coli* 세포들은 0.1 mM isopropylthio-β-D-galactoside 존재 하 18°C에서 하룻밤 배양되었다. 그 후, 상기 배양된 세포들을 3,500g에서 30분간 원심분리하여 수집하고, 수집된 세포들을 20 mM Tris-HCl(pH 7.6), 500 mM NaCl, 5 mM β-mercaptoethanol,

5% glycerol에 재현탁하였다. 상기 세포를 용해 버퍼에서 용해한 후, 음파처리(sonication)에 의해 파쇄하였다. 파쇄된 세포가 포함된 샘플을 15,000g로 30분간 원심분리하여 수득한 상층액을 0.45 μ m 주사기 필터(Millipore)를 통해 여과하고, 여과된 상층액을 FPLC 정제 시스템(KTA Purifier, GE Healthcare)을 사용하여, Ni²⁺-친화성 컬럼에 로드하였다. 결합 분획(bound fractions)은 80-400 mM imidazole, 20 mM Tris-HCl(pH 7.5) 구배에서 용출되었다.

[858] 상기 용출된 단백질을 TEV 프로테아제로 16시간 동안 처리하여 절단하였다. 절단된 단백질을 0.15-1.6 M NaCl 선형 농도구배의 Heparin 컬럼에서 정제하였다. Heparin 컬럼에서 정제된 재조합 Cas12f1 변이체 단백질은 20 mM Tris pH 7.6, 150 mM NaCl, 5 mM β -mercaptoethanol, 5% glycerol의 용액에서 투석되었다. 상기 투석된 단백질을 MBP 컬럼을 통과시켜 정제한 후, 0.5-1.2 M NaCl의 선형 구배로 monoS 컬럼(GE Healthcare) 또는 EnrichS에서 재정제하였다.

[859] 상기 재정제된 단백질들을 모아, 20 mM Tris pH 7.6, 150 mM NaCl, 5 mM β -mercaptoethanol, 5% glycerol의 용액으로 투석하여 본 발명에서 사용되는 초소형 유전자 편집 단백질(소형 엔도뉴클레아제)을 정제하였다. 상기 생산된 초소형 유전자 편집 단백질의 농도는 소 혈청 알부민(BSA)을 표준으로 사용하는 Bradford 정량법을 이용하여 정량하여 coomassie blue-stained SDS-PAGE 겔에서 전기영동적(electrophoretically)으로 측정되었다.

[860] 실시예 1.3. 상동지정복구를 위한 핵산 편집 시스템에 사용되는 가이드 RNA의 구축

[861] 상동지정복구에 사용하기 위한 핵산 편집 시스템(TaRGET 시스템)의 일 구성으로서 가이드 RNA(guide RNA, gRNA)는 도 1에 도시한 바와 같이 크게 5개 영역(각각 MS1 내지 MS5)에서 변형을 갖는 복수의 엔지니어링된 gRNA가 시험되었다. gRNA의 예시적인 서열을 하기 표 6에 나타내었다.

[862] [표6]

gRNA	Sequence(5' to 3')	서열 번호
Canonical sgRNA	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUGUCCCUuagGGGAUUAGAACUUGAGUGAA GGUGGGCUGCUUGCAUCAGCCUAAUGUCGAGAA GUGCUUUCUUCGGAAAGUAACCCUCGAAACAAA UUCAUUUUUCCUCUCCAAUUCUGCACAAgaaaGU UGCAGAACCCGAAUAGacgaaUGAAGGAAUGCAA CNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	13
MS1	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUGUCCCUUAGGGGAUUAGAACUUGAGUGA AGGUGGGCUGCUUGCAUCAGCCUAAUGUCGAGA	149

MS1/MS3-2	UGGAGAACCGCUUCACCAAAGCUGUCCCUUAG GGGAUUAGAACUUGAGUGAAGGUGGGCUGCUU GCAUCAGCCUAAUGUCGAGAAGUGCUUUCUUCG GAAAGUAACCCUCGAAACAAAUUCAGUGCUCU CUCCAAUUCUGCACAAGaaaGUUGCAGAACCCGA AUAGAGCAAUGAAGGAAUGCAACNNNNNNNNNN NNNNNNNNNNNN	155
MS1/MS3-3	ACCGCUUCACCAAAGCUGUCCCUUAGGGGAUU AGAACUUGAGUGAAGGUGGGCUGCUUGCAUCAG CCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUA ACCCUCGAAACAAAUUCAGUGCUCUCCUCUCCAAU UCUGCACAAGaaaGUUGCAGAACCCGAAUAGAGC AAUGAAGGAAUGCAACNNNNNNNNNNNNNNNNNN NNNN	156
MS1/MS4*-1	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUGUCCCUUAGGGGAUUAGAACUUGAGUGA AGGUGGGCUGCUUGCAUCAGCCUAAUGUCGAGA AGUGCUUUCUUCGGAAAGUAACCCUCGAAACAA AUUCAGUGCUCUCCUCUCCAAUUCgaaaGAACCCGA AUAGAGCAAUGAAGGAAUGCAACNNNNNNNNNN NNNNNNNNNNNN	157
MS1/MS4*-2	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUGUCCCUUAGGGGAUUAGAACUUGAGUGA AGGUGGGCUGCUUGCAUCAGCCUAAUGUCGAGA AGUGCUUUCUUCGGAAAGUAACCCUCGAAACAA AUUCAGUGCUCUCCUCUCgaaaGAAUAGAGCAAUGA AGGAAUGCAACNNNNNNNNNNNNNNNNNNNNNNNN	158
MS1/MS4*-3	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUGUCCCUUAGGGGAUUAGAACUUGAGUGA AGGUGGGCUGCUUGCAUCAGCCUAAUGUCGAGA AGUGCUUUCUUCGGAAAGUAACCCUCGAAACAA AUUCAGUGCUGgaaaAGCAAUGAAGGAAUGCAACN NNNNNNNNNNNNNNNNNNNNNNNN	159
MS1/MS5-1	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUGUuuagAUUAGAACUUGAGUGAAGGUGGG CUGCUUGCAUCAGCCUAAUGUCGAGAAGUGCUU	160

	UCUUCGGAAAGUAACCCUCGAAACAAAUUCAGU GCUCCUCUCCAAUUCUGCACAAGaaaGUUGCAGA ACCCGAAUAGAGCAAUGAAGGAAUGCAACNNNN NNNNNNNNNNNNNNNNNN	
MS1/MS5-2	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCuuagGAACUUGAGUGAAGGUGGGCUGCUUG CAUCAGCCUAAUGUCGAGAAGUGCUUUCUUCGG AAAGUAACCCUCGAAACAAAUUCAGUGCUCCUC UCCAAUUCUGCACAAGaaaGUUGCAGAACCCGAA UAGAGCAAUGAAGGAAUGCAACNNNNNNNNNN NNNNNNNNNN	161
MS1/MS5-3	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA uuagUUGAGUGAAGGUGGGCUGCUUGCAUCAGCC UAAUGUCGAGAAGUGCUUUCUUCGAAAGUAAAC CCUCGAAACAAAUUCAGUGCUCCUCUCCAAUUC UGCACAAGaaaGUUGCAGAACCCGAAUAGAGCAA UGAAGGAAUGCAACNNNNNNNNNNNNNNNNNN NN	162
MS1/MS2/M S4*-2	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUGUCCCUUAGGGGAUUAGAACUUGAGUGA AGGUGGGCUGCUUGCAUCAGCCUAAUGUCGAGA AGUGCUUUCUUCGAAAGUAACCCUCGAAACAA AUUCAGUGCUCCUCUCgaaaGAAUAGAGCAAUGA AGGAAUGCAACNNNNNNNNNNNNNNNNNNNN UUUAUUUU	163
MS1/MS3-3/ MS4*-2	ACCGCUUCACCAAAGCUGUCCCUUAGGGGAUU AGAACUUGAGUGAAGGUGGGCUGCUUGCAUCAG CCUAAUGUCGAGAAGUGCUUUCUUCGAAAGUA ACCCUCGAAACAAAUUCAGUGCUCCUCUCgaaaG AAUAGAGCAAUGAAGGAAUGCAACNNNNNNNN NNNNNNNNNNNN	164
MS1/MS2/M S5-3	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA uuagUUGAGUGAAGGUGGGCUGCUUGCAUCAGCC UAAUGUCGAGAAGUGCUUUCUUCGAAAGUAAAC CCUCGAAACAAAUUCAGUGCUCCUCUCCAAUUC UGCACAAGaaaGUUGCAGAACCCGAAUAGAGCAA	165

	UGAAGGAAUGCAACNNNNNNNNNNNNNNNNNNNN NNUUUUAUUUU	
MS1/MS3-3/ MS5-3	ACCGCUUCACCAAuuagUUGAGUGAAGGUGGGGCU GCUUGCAUCAGCCUAAUGUCGAGAAGUGCUUUC UUCGGAAAGUAACCCUCGAAACAAAUUCAGUGC <u>UCCUCUCCA</u> AUUCUGCACAAgaaaGUUGCAGAAC CCGA <u>UAGAGCAA</u> UGAAGGAAUGCAACNNNNNN NNNNNNNNNNNNNNNN	166
MS1/MS4*-2 /MS5-3	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA uuagUUGAGUGAAGGUGGGGCUGCUUGCAUCAGCC UAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAAC CCUCGAAACAAAUUCAGUGC <u>UCCUCUC</u> gaaaGAA UAGAGCAAUGAAGGAAUGCAACNNNNNNNNNN NNNNNNNNNNNN	167
MS1/MS2/M S3-3/MS4*-2	ACCGCUUCACCAAAGCUGUCCCuuagGGGAUUA GAACUUGAGUGAAGGUGGGCUGCUUGCAUCAGC CUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAA CCCUCGAAACAAAUUCAGUGC <u>UCCUCUC</u> gaaaGA AUAGAGCAAUGAAGGAAUGCAACNNNNNNNNNN NNNNNNNNNNNNUUUUUAUUUU	168
MS1/MS2/M S3-3/MS5-3	ACCGCUUCACCAAuuagUUGAGUGAAGGUGGGGCU GCUUGCAUCAGCCUAAUGUCGAGAAGUGCUUUC UUCGGAAAGUAACCCUCGAAACAAAUUCAGUGC <u>UCCUCUCCA</u> AUUCUGCACAAgaaaGUUGCAGAAC CCGA <u>UAGAGCAA</u> UGAAGGAAUGCAACNNNNNN NNNNNNNNNNNNNNUUUUUAUUUU	169
MS1/MS2/M S4*-2/MS5-3	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA uuagUUGAGUGAAGGUGGGGCUGCUUGCAUCAGCC UAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAAC CCUCGAAACAAAUUCAGUGC <u>UCCUCUC</u> gaaaGAA UAGAGCAAUGAAGGAAUGCAACNNNNNNNNNN NNNNNNNNNNNNUUUUUAUUUU	170
MS1/MS3-3/ MS4*-2/MS5 -3	ACCGCUUCACCAAuuagUUGAGUGAAGGUGGGGCU GCUUGCAUCAGCCUAAUGUCGAGAAGUGCUUUC UUCGGAAAGUAACCCUCGAAACAAAUUCAGUGC	171

	<u>UCCUCUCgaaaGAAUAGAGCAAUGAAGGAAUGCA</u> ACNNNNNNNNNNNNNNNNNNNNNNNNNNNN	
MS1/MS2/M S3-3/MS4*-2 /MS5-3	ACCGCUUCACCAAuuagUUGAGUGAAGGUGGGCU GCUUGCAUCAGCCUAAUGUCGAGAAGUGCUUUC UUCGGAAAGUAACCCUCGAAACAAAUUCAG <u>UGC</u> <u>UCCUCUCgaaaGAAUAGAGCAAUGAAGGAAUGCA</u> ACNNNNNNNNNNNNNNNNNNNNNNNNNNUUUUUAUUUU	172

[863] 또한, 상기 Canonical sgRNA에서 변형부위 중 하나인 MS1 서열이 제거된 성숙된 형태의 gRNA(mature form gRNA)를 제작하였다. 성숙된 형태의 gRNA의 예시적인 서열은 하기 표 7에 나타내었다.

[864] [표7]

gRNA	Sequence(5' to 3')	서열 번호
Mature form gRNA	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUGUCCCuagGGGAUUAGAACUUGAGUGAA GGUGGGCUGCUUGCAUCAGCCUAAUGUCGAGAA GUGCUUUCUUCGGAAAGUAACCCUCGAAACAAA UUCAUUUgaaaGAAUGAAGGAAUGCAACNNNNNN NNNNNNNNNNNNNNNNNN	173
MS3-1	GAUAAAGUGGAGAACCGCUUCACCAAAGCUGU CCCuagGGGAUUAGAACUUGAGUGAAGGUGGGC UGCUUGCAUCAGCCUAAUGUCGAGAAGUGCUUU CUUCGGAAAGUAACCCUCGAAACAAAUUCAUUU gaaaGAAUGAAGGAAUGCAACNNNNNNNNNNNN NNNNNNNN	174
MS3-2	UGGAGAACCGCUUCACCAAAGCUGUCCCuagG GGAUUAGAACUUGAGUGAAGGUGGGCUGCUUG CAUCAGCCUAAUGUCGAGAAGUGCUUUCUUCGG AAAGUAACCCUCGAAACAAAUUCAUUUgaaaGAA UGAAGGAAUGCAACNNNNNNNNNNNNNNNNNN NN	175
MS3-3	ACCGCUUCACCAAAGCUGUCCCuagGGGAUUA GAACUUGAGUGAAGGUGGGCUGCUUGCAUCAGC CUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAA CCCUCGAAACAAAUUCAUUUgaaaGAAUGAAGGA AUGCAACNNNNNNNNNNNNNNNNNNNNNNNN	176

MS4-1	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUGUCCC <u>Cuuag</u> GGGAUUAGAACUUGAGUGAA GGUGGGCUGCUUGCAUCAGCCUAAUGUCGAGAA GUGCUUUCUUCGGAAAGUAACCCUCGAAACAAA UUCA <u>Ugaaa</u> AUGAAGGAAUGCAACNNNNNNNNNN NNNNNNNNNN	177
MS4-2	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUGUCCC <u>Cuuag</u> GGGAUUAGAACUUGAGUGAA GGUGGGCUGCUUGCAUCAGCCUAAUGUCGAGAA GUGCUUUCUUCGGAAAGUAACCCUCGAAACAAA UUC <u>gaaa</u> GAAGGAAUGCAACNNNNNNNNNNNNNN NNNNNN	178
MS4-3	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUGUCCC <u>Cuuag</u> GGGAUUAGAACUUGAGUGAA GGUGGGCUGCUUGCAUCAGCCUAAUGUCGAGAA GUGCUUUCUUCGGAAAGUAACCCUCGAAACAAA <u>gaaa</u> GGAAGCAACNNNNNNNNNNNNNNNNNNNN	179
MS5-1	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCUGU <u>uuag</u> AUUAGAACUUGAGUGAAGGUGGG CUGCUUGCAUCAGCCUAAUGUCGAGAAGUGCUU UCUUCGGAAAGUAACCCUCGAAACAAAUUCAUU <u>Ugaaa</u> <u>GAA</u> AUGAAGGAAUGCAACNNNNNNNNNN NNNNNNNN	180
MS5-2	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA AAGCU <u>uuag</u> AGAACUUGAGUGAAGGUGGGCUGCU UGCAUCAGCCUAAUGUCGAGAAGUGCUUUCUUC GGAAAGUAACCCUCGAAACAAAUUCAUUU <u>Ugaaa</u> <u>G</u> <u>AA</u> AUGAAGGAAUGCAACNNNNNNNNNNNNNNNN NNNN	181
MS5-3	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA <u>uuag</u> UUGAGUGAAGGUGGGCUGCUUGCAUCAGCC UAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAAC CCUCGAAACAAAUUCAUUU <u>Ugaaa</u> <u>GAA</u> AUGAAGGAA UGCAACNNNNNNNNNNNNNNNNNNNNNN	182
MS3-3/MS4-3	ACCGCUUCACCAA <u>AAAGCUGUCCC</u> <u>Cuuag</u> GGGAUUAA GAACUUGAGUGAAGGUGGGCUGCUUGCAUCAGC	183

	CUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAA CCCUCGAAACAAAgaaaGGAAUGCAACNNNNNNNN NNNNNNNNNNNNNNNN	
MS3-3/MS5-3	ACCGCUUCACCAAuuagUUGAGUGAAGGUGGGCU GCUUGCAUCAGCCUAAUGUCGAGAAGUGCUUUC UUCGGAAAGUAACCCUCGAAACAAAUUCAUUUg aaaGAAUGAAGGAAUGCAACNNNNNNNNNNNNNNNN NNNNNNNN	184
MS4-3/MS5-3	CUUCACUGAUAAAGUGGAGAACCGCUUCACCAA uuagUUGAGUGAAGGUGGGCUGCUUGCAUCAGCC UAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAAAC CCUCGAAACAAAgaaaGGAAUGCAACNNNNNNNNNN NNNNNNNNNNNNNNNN	185
MS3-3/MS4-3/MS5-3	ACCGCUUCACCAAuuagUUGAGUGAAGGUGGGCU GCUUGCAUCAGCCUAAUGUCGAGAAGUGCUUUC UUCGGAAAGUAACCCUCGAAACAAAgaaaGGAU GCAACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	186

[865] 상기 표 6 및 표 7에서 'NNNNNNNNNNNNNNNNNNNNNNNNNN'으로 표시된 서열은 표적 유전자 내의 표적 서열과 혼성화할 수 있는 임의의 가이드 서열(스페이서 서열)을 의미한다. 상기 가이드 서열은 목적하는 표적 유전자 및/또는 상기 표적 유전자 내 표적 서열에 따라 당업자에 의해 적절하게 설계될 수 있으며, 따라서 특정 길이의 특정 서열로 한정되는 것은 아니다.

[866] 실시예 2. 핵산 편집 시스템을 이용한 상동지정복구(HDR)

[867] 실시예 2.1. 핵산 편집 시스템의 상동지정복구 효율 분석

[868] 상동지정복구를 위해서는 DNA의 이중가닥 절단을 위한 핵산 분해효소와 공여자(donor) 핵산이 필요하다. Cas12f1, TnpB 또는 이의 변이체 단백질을 포함하는 본원의 핵산 편집 시스템(TaRGET 시스템)의 상동지정복구 효율을 분석하고, 이를 기존의 Cas9과 Cas12a의 상동지정복구 효율과 비교하기 위한 연구를 수행하였다. 도 2는 예시적인 공여자 핵산의 구조 및 이를 이용한 유전자 편집(비상동말단연결 또는 상동지정복구에 의한 유전자 편집) 과정을 나타낸다. 도 2에서, 표적 핵산 서열은 5'-TTTAGAGGGAGACACAAGTTGATAGGG-3'(서열번호 296)를 사용하였다.

[869] 우선, 예시적인 표적 유전자로서 NLRC4(NLR Family CARD Domain Containing 4)에 대해 Cas9, Cas12a, Cas12f 및 TnpB의 상동지정복구 및 비상동말단연결 효율을 조사하였다. TnpB는 서열번호 1의 아미노산 서열을 갖는 단백질을 사용하였고, Cas12f는 서열번호 5의 아미노산 서열을 갖는 단백질을 사용하였으며, Cas9 및 Cas12a의 구체적인 서열 정보는 하기 표 8에 제공된다.

[870] [표8]

명칭	아미노산 서열	서열번호
Cas9 단백질	DKKYSIGLDIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEA TRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRIEESFLVEEDKKHERHPIFGN IVDEVAYHEKYPTIYHLRKKLVSDTKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV DKLFIQLVQTYNQLEENPINASGVDAKAILSARLSKSRRIENLI AQLPGEKKNGLFGNL IALSLGLTPNFKSNFDLAEDAKLQLSKDYYDDLDNLLAQIGDQYADFLAAKNLSDAIL LSDILRVNTEITKAPLSASMIKRYDEIHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAG YIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHA ILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEV VDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYEFTVYNELTKVKYVTEGMRKPAFLS GEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKII KDKDFLDNEENEDILEDIVLTLTFEDREMIERLKYAHLFDDKVMKQLKRRRYTGWR LSRKLINGIRDKQSGKTILDFLKSDFANRNFMLIHDDSLTFKEDIQKAQVSGQDLSH EHIANLAGSPAIKKGI LQTVKVDELVKVMGRHKPENIVIEARENQTTQKGQKNSRERM KRIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYVDHI VPQSFLKDDSIDNKVLRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNLT KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKIREVKVITLKS I.VSDFRKIDFQFYKVREINNYHHAHDAYLNAVVGTAI IKKYPKLESEFVYGDYKYVDVRKM IAKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIWDKGRDFA TVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAY SVLVVAKVEKGKSKKLKSVKELLGITIMERSSEKNPIDFLEAKGYKEVKKDLIKLPKY SLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKSPEDNEQKQLFVEQ HKHYLDEIEQISEFSKRVILADANLDKVL SAYNKHDKPIREQAENI IHLFTLTNLGAP AAFKYFDTTIDRKRYTSTKEVLDTL IHQSI TGLYETRIDLSQLGGD	297
Cas12a 단백질	TQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQQGFI EEDKARNHYKELKPIIDRIYKTY ADQCLQLVQLDWEENLSAAIDSYRKEKTEETRNLIEEQATYRNAIHDFYIGRTDNLDAI NKRHAEIYKGLFKAELFNGKVLKQLGTVTTEHENALLRSFDKFTTYFSGFYENRKNVFS AEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSI EEFV SFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHI IASLPHR FIPLFKQILSDRNTLSFILLEEFKSDEEVIQSFCYKTLRNENVLETAELFNELENSIDL THIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQ EII SAAGKELSEAFKQKTSEILSHAHAAALDQPLPTTLKKQEEKEILKSQDLSLLGLYHLL DWFVAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFLNFQMPTLA SGWDVNKEKNNGAILFVKNGLYLGLIMPQKQGRYKALSFEPTEKTSEGFDKMYDYFPDA AKMIPKCSTQLKAVTAHFQTHHTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAK KTGDKQGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHI SFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKL	298

[871]

NGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDE ARALLPNVITKEVSHEI IKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPE TPIIGIDRGERNLIYITVIDSTGKILEQSLNTIQQFDYQKKLDNREKERVAAARQAWSVV GTIKDLKQGYLSQVHIEIVDLMIHYQAVVLENLNFQFKSKRTGIAEKAVYQQFEKMLID KLNCLVLKDYPAEKVGGVLPYQLTDQFTSFAKMGTSQGFIFYVPAPYTSKIDPLTGFVD PFVWKTIKNHESRKHFLLEGDFLHYDVKTGDFILHFKMNRNLSFQRLPGFMPAWDIVFE KNETQFDAQGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGI VFRDGSNILP KLENDSDHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPPEWPM ADANGAYHI ALKGQLLLNLHLEKSKDLKLQNGISNQDWLAYIQELRN	
--	--

[872] 상기 Cas9 및 Cas12a 각각에 대해 사용된 가이드 RNA의 서열 정보는 하기 표 9에서 제공된다.

[873] [표9]

구분	염기서열	서열 번호
Cas9 gRNA	GAGGGAGACACAAGTTGATAgttttagagctagaaatagcaagttaa aataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgcttttt	299
Cas9 gRN A scaffold	gttttagagctagaaatagcaagttaaaataaggctagtcggttatcaactgaaaaagtg gcaccgagtcggtgc	300
Cas12a gRNA	taatttctactctttagatGAGGGAGACACAAGTTGATA	301
Cas12a crRNA	taatttctactctttagat	302

[874] 상기 표적 유전자 NLRC4의 서열은 서열번호 190의 서열 (GAGGGAGACACAAGTTGATA)을 사용하였다. 상기 표적 유전자 서열에서 Cas12a, Cas12f1의 PAM 서열은 표적 유전자 서열의 5' 말단부 방향에 위치한 5'-TTTA-3'이고, Cas9의 PAM 서열은 표적 유전자 서열의 3' 말단부 방향에 위치한 5'-GGG-3'이다. 각각의 핵산 분해효소와 gRNA를 암호화하는 핵산 서열을 포함하는 벡터 1 µg 및 공여자 핵산 1 µg을 HEK293T 세포에 형질감염시켰다. 그 후 유전체(genomic) DNA를 추출하고, 해당 NLRC4의 표적 부위를 PCR로 증폭한 다음 딥시퀀싱(deep sequencing, Illumina iSeq 100) 분석을 수행함으로써 상동지정복구 및 비상동말단연결의 효율을 조사하였다. 본 실험 결과는 도 3에 도시하였다.

[875] 도 3에서 확인할 수 있는 바와 같이, 공여자 핵산(도 3의 donor DNA)이 존재하지 않는 조건에서는 상기 3종(Cas9, Cas12a, Cas12f)의 서로 다른 Cas 시스템에서 모두 비상동말단연결(NHEJ)이 거의 비슷한 70% 정도의 효율로 발생하였다. 한편, 공여자 핵산이 존재하는 상태에서만 상동지정복구가 일어났다. Cas9 및 Cas12a의 경우 각각 4.13%, 4.26%의 낮은 상동지정복구 효율을 보인 반면, TnpB의 경우에는 11.87%의 높은 상동지정복구 효율을 보였다. 이러한 결과는 본

발명의 표적 핵산 편집 시스템이 Cas9 및 Cas12a 대비 상동지정복구 효율이 현저하게 높음을 입증한다.

[876] 실시예 2.2. 공여자 핵산의 길이에 따른 상동지정복구 효율 분석

[877] 공여자 핵산의 길이에 따른 상동지정복구 효율을 분석하였다. 구체적으로, 상기 실시예 2.1에서는 공여자 핵산을 600 bp의 단일한 길이로 한 것과 달리, 본 실시예에서는 공여 서열의 길이를 각각 8 kb, 6 kb, 5 kb, 4 kb, 3 kb, 2.5 kb, 2 kb, 1.6 kb, 1 kb, 800 bp, 600 bp, 400 bp 및 200 bp로 다양하게 변화시켰다. 상동지정복구 효율은 상기 실시예 2.1과 동일한 방법으로 표적 유전자 NLRC4를 대상으로 하여 분석하였다. 본 실험 결과는 도 4에 도시하였다.

[878] 도 4에 나타낸 바와 같이, 일 실시예에 따른 표적 핵산 편집 시스템은 공여자 핵산의 길이가 약 4 kb에 다다를 때까지 상동지정복구 효율이 길이에 비례하여 증가함을 확인하였다.

[879] 실시예 2.3. 시간에 따른 상동지정복구 및 비상동말단연결 효율의 분석

[880] 형질감염 후 경과된 시간에 따라 상동지정복구 및 비상동말단연결의 효율이 변화하는지 여부를 분석하였다. 구체적으로, HEK293T 세포에 Cas9 또는 TaRGET 시스템(TnpB)을 암호화하는 핵산 서열을 포함하는 벡터 1 µg 및 공여자 핵산 1 µg을 각각 형질감염시켰다. 표적 유전자는 상기 실시예 2.1과 동일하게 NLRC4를 사용하였으며, 형질감염 후 일정한 시간이 경과한 뒤에 세포를 수득하여 표적 부위를 증폭하여 딥시퀀싱 분석을 통해 상동지정복구 및 비상동말단연결 효율을 분석하였다. 형질감염된 HEK293T 세포는 최초 형질감염일을 기준으로 5일 후 계대배양을 하였으며, 계대배양 직후에는 위와 동일하게 Cas9 또는 TaRGET(TnpB) 시스템을 암호화하는 핵산 서열을 포함하는 벡터 1 µg 및 공여자 핵산 1 µg을 다시 형질감염시켰다. 이러한 과정을 최초 형질감염 10일 후에도 수행하였으며, 형질감염 후 14일까지의 유전자 편집 효율을 분석하였다. 본 실험 결과는 도 5에 도시하였다.

[881] 도 5에 나타낸 바와 같이, Cas9 및 TaRGET(TnpB) 시스템 모두 시간이 경과함에 따라 상동지정복구 효율이 증가하는 것으로 확인되었다. 특히, TaRGET 시스템은 Cas9 시스템과 비교하여 더욱 현격한 상동지정복구 효율의 증가를 보였다. 이와 동시에, 비상동말단연결 효율은 Cas9에 비해 TaRGET 시스템에서 더 억제됨을 알 수 있었다. 즉, 시간이 경과함에 따라 TaRGET 시스템의 경우 비상동말단연결에 비해 상동지정복구의 효율이 현저하게 증가함을 확인하였다.

[882] 실시예 2.4. 다양한 표적 유전자에서 Cas9, Cas12a 및 TaRGET 시스템의 상동지정복구 효율 분석

[883] 표적 유전자로서 NLRC4, FUS 및 LOC105370393을 대상으로 상동지정복구 효율을 분석하였다. 상기 각 표적 유전자의 표적 서열은 순서대로 각각 서열번호 190의 서열(GAGGGAGACACAAGTTGATA), 서열번호 191의 서열(GTGGGTAGGTCCAGTTTGGG) 및 서열번호 192의 서열(GCAGTACACCTGAGGGAACA)을 사용하였다. 상기 표적 유전자 서열에서

Cas12a 및 Cas12f 변이체의 PAM 서열은 표적 유전자 서열의 5' 말단부 방향에 위치한 5'-TTTA-3'이고, Cas9의 PAM 서열은 표적 유전자 서열의 3' 말단부 방향에 위치한 5'-GGG-3'이다. 목적 서열로는 길이 27 bp의 서열번호 193의 서열 (AACGTGACACGACGCGTTTCGGAGAAC)을 이용하였으며, gRNA는 상기 표 3 및 표 4의 ge_4.0(표적 유전자가 LOC105370393인 경우) 또는 ge_4.1(표적 유전자가 NLRC4 또는 FUS인 경우)을 이용하였다. HEK293T 세포에 Cas9, Cas12a 또는 TaRGET(Cas12f, TnpB) 시스템을 암호화하는 핵산 서열을 포함하는 벡터 1 µg 및 공여자 핵산 1 µg을 각각 형질감염시켰다. 형질감염 후 5일이 경과한 시점에 세포를 수집하여 표적 부위를 증폭하고 딥시퀀싱 분석을 통해 상동지정복구 및 비상동말단연결 효율을 분석하였다. 본 실험 결과는 도 6에 도시하였다.

[884] 도 6에서 확인할 수 있는 바와 같이, Cas9 및 Cas12a와 비교할 때 TaRGET 시스템(Cas12f, TnpB)은 세 표적 유전자 모두에서 현저히 높은 비상동말단연결 대비 상동지정복구 효율을 나타내었다. 세 표적 유전자에서 모두 Cas9과 Cas12a 사이의 편집 효율 차이는 크지 않았으나, TaRGET 시스템은 Cas9 및 Cas12a와 비교하여 3배에 가까운 비상동말단연결 대비 상동지정복구 효율을 나타내었다.

[885] 실시예 3. 유전자 복구 기작에 관여하는 유전자의 발현 억제를 통한 상동지정복구 효율 분석

[886] 실시예 3.1. 비상동말단연결(NHEJ) 관련 유전자의 발현 억제를 통한 상동지정복구 효율 분석

[887] 전술한 바와 같이, 공여자 DNA가 존재하는 상태에서는 상동지정복구와 비상동말단연결이 모두 가능하므로 두 유전자 수선 기작이 경합하여 발생하게 된다. 따라서 상동지정복구로의 유도를 향상시키기 위해 비상동말단연결 과정에 관여하는 것으로 알려진 다양한 유전자들의 발현을 shRNA를 사용하여 억제하였다. 비교를 위해, 상동지정복구 과정에 관여하는 유전자인 Rad51의 발현 억제 실험도 함께 진행되었다.

[888] 구체적으로, DCLRE1C, LIG4, XRCC4, KU70, XLF, ATM 및 Rad51 유전자를 각각 표적으로 하는 shRNA(순서대로 서열번호 194 내지 200) 및 대조군(Scrambled) shRNA(서열번호 201)를 제작하였다. 상기 shRNA의 구체적인 서열 정보는 하기 표 10에서 제공된다.

[889] [표10]

명칭	염기서열	서열 번호
shRNA for DC LRE1C	GCAGAGCTCTCGTTTCACATTCAAGAGATGTGAAAC GAGAGCTCTGC	194
shRNA for LIG4	GCA TGATCCTTCTGTAGGATTCAAGAGATCCTACAGA AGGATCATGC	195

shRNA for XRCC4	GAATCCACCTTGTCTTCTGATTCAAGAGATCAGAAACA AGGTGGATTC	196
shRNA for KU70	GCAGCATTGTGCAGATACATTCAAGAGATGTATCTG CACAATGCTGC	197
shRNA for XLF	GCATGAGTCTGGCATTACATTCAAGAGATGTAATGC CAGACTCATGC	198
shRNA for ATM	GCAAGCAGCTGAAACAAATTTCAAGAGAATTTGTTT CAGCTGCTTGC	199
shRNA for Rad51	CGCCAAAGAAGGAGCTAATAATTCAAGAGATTATTA GCTCCTTCTTTGGCG	200
shRNA for Scrambled	CAGAGCTAACTCAGATAGTACTTTCAAGAGAAGTAC TATCTGAGTTAGCTCTG	201

- [890] 상기 shRNA는 CBA 프로모터와 ge_4.1 gRNA를 포함하는 TaRGET 시스템 및 1.6 kb 크기의 공여자 핵산(27 bp 치환)과 함께 실시예 2.1에서와 동일한 방법으로 세포에 형질전환시키고, 이후 상동지정복구 효율을 분석하였다. 본 실험 결과는 도 7에 도시하였다. 도 7에서 확인할 수 있는 바와 같이, 비상동말단연결에 관여하는 유전자의 발현을 억제하는 경우 TaRGET 시스템의 상동지정복구 효율이 유의하게 증가하였다. 한편, 상동지정복구를 유도하는 Rad51 유전자의 발현을 억제할 경우에는 상동지정복구 효율이 감소함을 알 수 있었다.
- [891] 이와 같은 결과는 비상동말단연결에 관여하는 유전자의 발현을 억제함으로써 TaRGET 시스템을 이용한 상동지정복구 효율을 증가시킬 수 있음을 의미한다. 특히, TaRGET 시스템은 Cas 단백질 및 gRNA의 크기가 매우 작기 때문에 하나의 전달 벡터 내에 shRNA 분자를 함께 도입함으로써 상동지정복구 효율을 증가시킬 수 있는 장점이 있다.
- [892] 실시예 3.2. DCLRE1C 녹아웃 세포주에서의 상동지정복구 효율 검증
- [893] 실시예 3.1에서 상동지정복구 효율의 증가가 가장 컸던 DCLRE1C 유전자 억제의 효과를 추가로 검증하기 위해, DCLRE1C가 녹아웃된 HEK293T 세포주를 제작하였다. DCLRE1C^{-/-} 세포주 및 야생형 HEK293T 세포주에서 표적 유전자 DCLR4에 대한 비상동말단연결 대비 상동지정복구 효율을 조사하였다. 본 실험 결과는 도 8에 도시하였다.
- [894] 도 8에서 확인할 수 있는 바와 같이, DCLRE1C가 녹아웃된 세포에서는 야생형 대비 비상동말단연결 대비 상동지정복구 효율이 현저하게 증가되었다.
- [895] 이어서, 상기 두 종류의 세포에 Cas9, Cas12a 또는 TaRGET(Cas12f1 변이체) 시스템을 암호화하는 핵산 서열을 포함하는 벡터 1 µg 및 공여자 핵산 1 µg을 각각 형질감염시켰다. 형질감염 후 5일이 경과한 시점에서 세포를 수집한 다음 표적

부위를 증폭하여 덤시퀀싱 분석을 통해 상동지정복구 및 비상동말단연결 효율을 분석하였다. 본 실험 결과는 도 9에 도시하였다.

- [896] 도 9에서 확인할 수 있는 바와 같이, Cas9 및 Cas12a의 경우와 비교하여 TaRGET(Cas12f1 변이체) 시스템은 야생형 세포주에서 높은 비상동말단연결 대비 상동지정복구 효율을 나타내었다. 이와 같은 TaRGET 시스템의 비교 우위는 DCLRE1C 유전자가 녹아웃된 세포주에서 더 현격하게 관찰되었다. 구체적으로, Cas9과 Cas12a 시스템의 경우에도 DCLRE1C 유전자 녹아웃에 따른 상동지정복구 효율의 증가가 관찰되었으나 그 증가폭은 미미하였다. 대조적으로 TaRGET 시스템에서는 상동지정복구 효율의 증가폭이 월등하게 큰 것을 확인하였다. 이러한 결과는 TaRGET 시스템에서 DCLRE1C 유전자 발현을 억제할 수 있는 모듈을 추가로 장착할 경우 Cas9과 Cas12a에서 구현할 수 없는 월등한 수준의 상동지정복구 효율을 달성할 수 있음을 증명한 것이다.
- [897] 실시예 4. AAV 전달체를 사용한 상동지정복구 효율의 최적화
- [898] AAV 전달체를 사용하여 상동지정복구를 위한 표적 핵산 편집 시스템을 효율적으로 전달하기 위해서는 AAV 패키징 한계인 약 4.7 kb 내로 모든 구성물이 포함되어야 한다. 상기 구성물은 가이드 RNA를 포함한 TaRGET 시스템 및 공여자 핵산을 필수 구성요소로 포함하며, 실시예 3에서 확인된 비상동말단연결에 관여하는 유전자의 발현 조절을 위한 shRNA 등이 추가로 포함될 수 있다. 한편, 실시예 2.2에서 확인된 바와 같이 공여자 핵산은 약 4 kb에 다다를 때까지 그 길이에 비례하여 상동지정복구 효율이 증가하였으므로, 프로모터의 종류와 shRNA의 유무에 따라 최대 포함될 수 있는 길이의 공여자 핵산을 사용하였다.
- [899] 구체적으로, 상대적으로 더 긴 길이의 CBA(chicken β -actin) 프로모터 또는 더 짧은 길이의 EFS(elongation factor 1 α short) 프로모터를 사용하는 경우 및 shDCLRE1C를 사용하거나 사용하지 않는 경우의 4가지 TaRGET 시스템 벡터 조성물을 제작하였다. 그리고 각 벡터 조성물의 상동지정복구 효율을 NLRC4, FUS, LOC105370393 유전자에 대해서 조사하였다. 상기와 같이 제작한 벡터 조성물 2 μ g을 야생형 HEK293T 세포에 형질감염시키고, 5일이 경과한 후 세포를 수집하여 표적 부위의 증폭 및 덤시퀀싱에 의해 상동지정복구 효율을 분석하였다. 본 실험 결과는 도 10에 도시하였다.
- [900] 도 10a에 도시한 바와 같이, 각각의 벡터 조성물을 통해 TnpB 발현을 위한 프로모터로서 CBA 프로모터와 EFS 프로모터를 비교하였으며(도 10a에서 A와 C, B와 D의 비교), 이들 각각에 대해 shDCLRE1C가 존재하는 경우와 존재하지 않는 경우의 4가지 벡터 조성물을 구축하였다. 모든 벡터 조성물은 프로모터의 길이와 shDCLRE1C 존재 여부에 따라 공여자 핵산의 길이를 최대 조절함으로써 전체 구성물의 길이가 4.7 kb 내로 통일되도록 하였다. 이로써 AAV를 이용한 전달이 가능하도록 하였다.
- [901] 실험 결과 도 10b에 나타난 바와 같이, 길이가 더 긴 CBA 프로모터를 사용한 경우에는 shDCLRE1C를 추가하더라도 상동지정복구 효율의 증가가 크지 않음을

확인하였다(도 10b에서 A 및 B의 비교). 이러한 결과는 shDCLRE1C의 추가에 따라 공여자 핵산의 길이가 감소되고, 이로 인한 상동지정복구 효율 감소에 의해 shDCLRE1C에 의한 효과가 일부 상쇄되기 때문인 것으로 사료된다. 한편, EFS 프로모터를 사용한 경우에는 shDCLRE1C 추가에 따른 상동지정복구 효율의 증가가 크게 나타났다(도 10b에서 C 및 D의 비교). 이는 shDCLRE1Cf를 추가하더라도 CBA 프로모터를 사용한 경우와 비교하여 더 긴 공여자 핵산을 사용할 수 있기 때문에 가장 높은 상동지정복구 효율을 달성할 수 있는 것이다.

[902] 상기와 같은 결과는 프로모터의 종류, 공여자 핵산의 길이 및 비상동말단 연결 조절 유전자의 발현 억제를 위한 분자의 사용 여부 등을 조절, 선택함으로써 AAV 벡터를 이용한 전달시 최적의 상동지정복구 효율을 추구할 수 있음을 보여준다.

[903] 상기 진술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술 분야의 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다.

청구범위

- [청구항 1] Cas12f1, TnpB 또는 이의 변이체(variant) 단백질을 포함하는 엔도뉴클레아제 또는 상기 엔도뉴클레아제를 암호화하는 핵산; 가이드 서열을 포함하는 엔지니어링된 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 핵산; 및 공여자 핵산 분자 또는 상기 공여자 핵산 분자를 암호화하는 핵산을 포함하는 표적 핵산의 편집 시스템.
- [청구항 2] 제1항에 있어서, 상기 시스템은 표적 핵산에 이중가닥 절단(double-strand breaks)을 일으키는 것인 시스템.
- [청구항 3] 제1항에 있어서, 상기 시스템은 표적 핵산 또는 그 인접 부위에서 공여자 핵산 분자를 주형으로 하여 이중가닥 절단의 상동지정복구(homology-directed repair) 과정에 의해 목적 서열이 도입되는 것인 시스템.
- [청구항 4] Cas12f1, TnpB 또는 이의 변이체(variant) 단백질을 포함하는 엔도뉴클레아제 또는 상기 엔도뉴클레아제를 암호화하는 핵산; 가이드 서열을 포함하는 엔지니어링된 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 핵산; 및 공여자 핵산 분자 또는 상기 공여자 핵산 분자를 암호화하는 핵산을 포함하는 표적 핵산의 유전자 편집 조성물.
- [청구항 5] 제1항 또는 제4항에 있어서, 상기 Cas12f1, TnpB 또는 이의 변이체 단백질은 서열번호 1 내지 서열번호 5로 이루어진 군에서 선택된 아미노산 서열과 70% 이상의 서열 동일성을 갖는 아미노산 서열을 포함하는 시스템 또는 조성물.
- [청구항 6] 제1항 또는 제4항에 있어서, 상기 TnpB 단백질은 서열번호 202 내지 서열번호 293로 이루어진 군에서 선택된 아미노산 서열과 70% 이상의 서열 동일성을 갖는 아미노산 서열을 포함하는 시스템 또는 조성물.
- [청구항 7] 제1항 또는 제4항에 있어서, 상기 Cas12f1, TnpB 또는 이의 변이체 단백질은 하기 서열 중 하나를 포함하는 시스템 또는 조성물:
(i) 서열번호 5의 아미노산 서열;

- (ii) 서열번호 1의 아미노산 서열;
- (iii) 서열번호 1의 아미노산 서열에서 N-말단의 1개 내지 28개의 아미노산이 제거 또는 치환된 아미노산 서열; 또는
- (iv) 서열번호 1의 아미노산 서열의 N-말단 또는 C-말단에 1개 내지 600개의 아미노산이 추가된 아미노산 서열.

- [청구항 8] 제7항에 있어서,
상기 추가된 1개 내지 600개의 아미노산은 서열번호 294 또는 서열번호 295의 아미노산 서열인
시스템 또는 조성물.
- [청구항 9] 제1항 또는 제4항에 있어서,
상기 Cas12f1, TnpB 또는 이의 변이체 단백질은 서열번호 1 내지 서열번호 4로 이루어진 군에서 선택된 아미노산 서열과 70% 이상의 서열 동일성을 갖는
시스템 또는 조성물.
- [청구항 10] 제1항 또는 제4항에 있어서,
상기 Cas12f1, TnpB 또는 이의 변이체 단백질은 서열번호 1의 아미노산 서열과 70% 이상의 서열 동일성을 갖는
시스템 또는 조성물.
- [청구항 11] 제1항 또는 제4항에 있어서,
상기 엔지니어링된 가이드 RNA는 야생형 Cas12f1 가이드 RNA 서열에서 1개 이상의 뉴클레오티드가 치환, 결실, 삽입 또는 부가된 서열을 포함하고, 가이드 서열을 제외한 부분이 상기 야생형 Cas12f1 가이드 RNA와 50% 이상의 서열 동일성을 갖는
시스템 또는 조성물.
- [청구항 12] 제1항 또는 제4항에 있어서,
상기 야생형 Cas12f1 가이드 RNA는 (i) 하나 이상의 스템(stem) 영역, (ii) tracrRNA-crRNA 상보성 영역 및 임의적으로 (iii) 연속되는 3개 이상의 유라실(U)을 포함하는 영역을 포함하는 tracrRNA(trans-activating CRISPR RNA) 및 crRNA(CRISPR RNA)를 포함하고,
상기 엔지니어링된 가이드 RNA는 하기 (a) 내지 (d)로 이루어진 군에서 선택되는 하나 이상의 변형을 포함하는
시스템 또는 조성물:
(a) 하나 이상의 스템 영역의 일부 또는 전부의 결실;
(b) tracrRNA-crRNA 상보성 영역의 일부 또는 전부의 결실;
(c) tracrRNA-crRNA 상보성 영역 내에 연속되는 3개 이상의 유라실(U)이 존재하는 경우 하나 이상의 U를 A, G 또는 C로 치환; 및
(d) crRNA 서열의 3'-말단에 하나 이상의 유리딘(uridine)의 부가.
- [청구항 13] 제12항에 있어서,

상기 야생형 Cas12f1 가이드 RNA는 서열번호 11의 핵산 서열로 이루어진 tracrRNA 및 서열번호 12의 핵산 서열로 이루어진 crRNA를 포함하는 시스템 또는 조성물.

- [청구항 14] 제12항에 있어서,
 상기 야생형 Cas12f1 가이드 RNA는 5'-말단부터 순차적으로 제1 스템 영역, 제2 스템 영역, 제3 스템 영역, 제4 스템 영역 및 tracrRNA-crRNA 상보성 영역을 포함하는 tracrRNA 및 crRNA를 포함하고,
 상기 엔지니어링된 가이드 RNA는
 (a1) 제1 스템 영역의 일부 또는 전부의 결실;
 (a2) 제2 스템 영역의 일부 또는 전부의 결실;
 (b) tracrRNA-crRNA 상보성 영역의 일부 또는 전부의 결실;
 (c) tracrRNA-crRNA 상보성 영역 내에 연속되는 3개 이상의 유라실(U)이 존재하는 경우 하나 이상의 U를 A, G 또는 C로 치환; 및
 (d1) crRNA 서열의 3'-말단에 U-rich tail의 부가(상기 U-rich tail의 서열은 5'-(U_mV)_nU_o-3'로 표시되고 여기서 상기 V는 각각 독립적으로 A, C 또는 G 이고, m 및 o는 1 내지 20 사이의 정수이며, n은 0 내지 5 사이의 정수임)로 이루어진 군에서 선택되는 하나 이상의 변형을 포함하는 시스템 또는 조성물.
- [청구항 15] 제14항에 있어서,
 상기 엔지니어링된 가이드 RNA는 (d1) crRNA 서열의 3'-말단에 U-rich tail, (c) tracrRNA-crRNA 상보성 영역 내에 연속되는 3개 이상의 유라실(U)이 존재하는 경우 하나 이상의 U를 A, G 또는 C로 치환, 또는 상기 둘 모두의 변형을 포함하는 시스템 또는 조성물.
- [청구항 16] 제14항에 있어서,
 상기 엔지니어링된 가이드 RNA는 (a1) 제1 스템 영역의 일부 또는 전부의 결실; (a2) 제2 스템 영역의 일부 또는 전부의 결실; 및 (b) tracrRNA-crRNA 상보성 영역의 일부 또는 전부의 결실로 이루어진 군에서 선택된 하나 이상의 변형을 포함하는 시스템 또는 조성물.
- [청구항 17] 제16항에 있어서,
 상기 엔지니어링된 가이드 RNA는 (b1) tracrRNA-crRNA 상보성 영역의 일부의 결실을 포함하고, 상기 상보성 영역의 일부는 1개 내지 54개의 뉴클레오티드인 시스템 또는 조성물.
- [청구항 18] 제16항에 있어서,

상기 엔지니어링된 가이드 RNA는 (b2) tracrRNA-crRNA 상보성 영역의 전부의 결실을 포함하고, 상기 상보성 영역의 전부는 55개의 뉴클레오티드인

시스템 또는 조성물.

[청구항 19]

제14항에 있어서,

상기 엔지니어링된 가이드 RNA는 (a1) 제1 스템 영역의 일부 또는 전부의 결실을 포함하고, 상기 스템 영역의 일부 또는 전부는 1개 내지 20개의 뉴클레오티드인

시스템 또는 조성물.

[청구항 20]

제14항에 있어서,

상기 엔지니어링된 가이드 RNA는 (a2) 제2 스템 영역의 일부 또는 전부의 결실을 포함하고, 상기 스템 영역의 일부 또는 전부는 1 내지 27개의 뉴클레오티드인

시스템 또는 조성물.

[청구항 21]

제14항에 있어서,

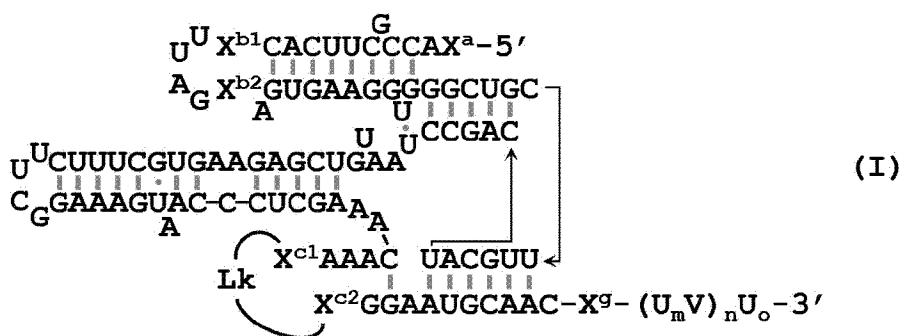
상기 엔지니어링된 가이드 RNA는 (a1) 제1 스템 영역의 일부 또는 전부의 결실; (d1) crRNA 서열의 3'-말단에 U-rich tail의 부가; 또는 상기 둘 모두의 변형을 포함하는

시스템 또는 조성물.

[청구항 22]

제1항 또는 제4항에 있어서,

상기 엔지니어링된 가이드 RNA는 하기 식 (I)로 표시되는 서열로 이루어지거나 상기 서열과 80% 이상 서열 동일성을 갖는 시스템 또는 조성물:



식 (I)에서,

X^{a} , X^{b1} , X^{b2} , X^{c1} 및 X^{c2} 는 각각 독립적으로 0 내지 35개의 (폴리)뉴클레오티드로 이루어지고,

X^{g} 는 10 내지 30개의 폴리뉴클레오티드로 이루어진 가이드 서열로서 표적 서열과 혼성화하거나 표적 서열에 상보적인 서열이고,

Lk는 길이 2 내지 20의 폴리뉴클레오티드 링커이거나 부존재하고,

$(U_m V)_n U_o$ 는 U-rich tail로서 존재하거나 부존재하고, 존재하는 경우 U는 유리딘이고, V는 각각 독립적으로 A, C 또는 G이고, m 및 o는 각각 독립적으로 1 내지 20 사이의 정수이며, n은 0 내지 5 사이의 정수이다.

- [청구항 23] 제22항에 있어서,
상기 X^a 는 서열번호 14의 핵산 서열을 포함하거나 상기 서열번호 14의 서열에서 1 내지 20개의 뉴클레오티드가 결실된 핵산 서열을 포함하는 시스템 또는 조성물.
- [청구항 24] 제22항에 있어서,
상기 X^{b1} 은 서열번호 25의 핵산 서열을 포함하거나 상기 서열번호 25의 서열에서 1 내지 13개의 뉴클레오티드가 결실된 핵산 서열을 포함하는 시스템 또는 조성물.
- [청구항 25] 제22항에 있어서,
상기 X^{b2} 은 서열번호 29의 핵산 서열을 포함하거나 상기 서열번호 29의 서열에서 1 내지 14개의 뉴클레오티드가 결실된 핵산 서열을 포함하는 시스템 또는 조성물.
- [청구항 26] 제22항에 있어서,
상기 식 (I) 내의 서열 5'- X^{b1} UUAG X^{b2} -3'은 서열번호 34 내지 서열번호 38로 이루어진 군에서 선택된 어느 하나의 핵산 서열인 시스템 또는 조성물.
- [청구항 27] 제22항에 있어서,
상기 X^{c1} 은 서열번호 39의 핵산 서열을 포함하거나 상기 서열번호 39의 서열에서 1 내지 28개의 뉴클레오티드가 결실된 핵산 서열을 포함하는 시스템 또는 조성물.
- [청구항 28] 제27항에 있어서,
상기 X^{c1} 서열 내에 연속되는 3개 이상의 유라실(U)이 존재하는 경우 이들 중 하나 이상의 U가 A, G 또는 C로 치환되는 변형을 포함하는 시스템 또는 조성물.
- [청구항 29] 제22항에 있어서,
상기 X^{c2} 는 서열번호 58의 핵산 서열을 포함하거나 상기 서열번호 58의 서열에서 1 내지 27개의 뉴클레오티드가 결실된 핵산 서열을 포함하는 시스템 또는 조성물.
- [청구항 30] 제29항에 있어서,
상기 X^{c2} 서열 내에 서열 5'-ACGAA-3'가 존재하는 경우 해당 서열이 5'-NGNNN-3'로 치환되고, 상기 N은 각각 독립적으로 A, C, G 또는 U인 시스템 또는 조성물.
- [청구항 31] 제22항에 있어서,

상기 식 (I) 내의 서열 5'-X^{c1}-Lk-X^{c2}-3'은 서열번호 80 내지 서열번호 86으로 이루어진 군에서 선택된 어느 하나의 핵산 서열인 시스템 또는 조성물.

- [청구항 32] 제22항에 있어서,
상기 Lk는 5'-GAAA-3', 5'-UUAG-3', 5'-UGAAA-3', 5'-UUGAAAAA-3', 5'-UUCGAAAGAA-3'(서열번호 76), 5'-UUCAGAAAUCAA-3'(서열번호 77), 5'-UUCAUGAAAAUGAA-3'(서열번호 78) 및 5'-UUCAUUGAAAAAUGAA-3'(서열번호 79)로 이루어진 군에서 선택된 어느 하나의 핵산 서열을 포함하는 시스템 또는 조성물.
- [청구항 33] 제22항에 있어서,
상기 (U_mV)_nU_o는 (i) n이 0이고, o가 1 내지 6 사이의 정수이거나 (ii) V가 A 또는 G이고, m 및 o는 각각 독립적으로 3 내지 6 사이의 정수이고, n은 1 내지 3 사이의 정수인 시스템 또는 조성물.
- [청구항 34] 제14항에 있어서,
상기 엔지니어링된 가이드 RNA는 서열번호 87 내지 서열번호 132로 이루어진 군에서 선택된 어느 하나의 핵산 서열로 이루어진 엔지니어링된 tracrRNA를 포함하는 시스템 또는 조성물.
- [청구항 35] 제14항에 있어서,
상기 엔지니어링된 가이드 RNA는 서열번호 133 내지 서열번호 148로 이루어진 군에서 선택된 어느 하나의 핵산 서열로 이루어진 엔지니어링된 crRNA를 포함하는 시스템 또는 조성물.
- [청구항 36] 제12항에 있어서,
상기 엔지니어링된 가이드 RNA는 듀얼 가이드 RNA 또는 싱글 가이드 RNA인 시스템 또는 조성물.
- [청구항 37] 제36항에 있어서,
상기 엔지니어링된 싱글 가이드 RNA는 서열번호 149 내지 서열번호 186으로 이루어진 군에서 선택된 어느 하나의 핵산 서열로 이루어진 시스템 또는 조성물.
- [청구항 38] 제1항 또는 제4항에 있어서,
상기 Cas12f1, TnpB 또는 이의 변이체 단백질을 포함하는 엔도뉴클레아제; 및 상기 가이드 RNA를 리보뉴클레오단백질(ribonucleoprotein, RNP) 형태로 포함하는 시스템 또는 조성물.

- [청구항 39] 제1항 또는 제4항에 있어서,
상기 공여자 핵산 분자는 상동지정복구에서 주형으로 사용되는 서열로서 1 bp 내지 20 kb의 길이를 갖는
시스템 또는 조성물.
- [청구항 40] 제1항 또는 제4항에 있어서,
상기 시스템 또는 조성물은 비상동말단연결(non-homologous end joining, NHEJ)에 관여하는 유전자의 발현을 억제하는 분자를 더 포함하는
시스템 또는 조성물.
- [청구항 41] 제40항에 있어서,
상기 비상동말단연결에 관여하는 유전자는 ATM1, XRCC4, XLF, XRCC6, LIG4 및 DCLRE1C로 이루어진 군으로부터 선택된 어느 하나 이상인
시스템 또는 조성물.
- [청구항 42] 제40항에 있어서,
상기 분자는 shRNA, siRNA, miRNA, 또는 안티센스 올리고뉴클레오티드인
시스템 또는 조성물.
- [청구항 43] Cas12f1, TnpB 또는 이의 변이체(variant) 단백질을 포함하는 엔도뉴클레아제를 암호화하는 뉴클레오티드 서열이 작동 가능하게 연결된 제1 핵산 구조물;
표적 핵산에 상보적으로 결합하는 가이드 서열을 포함하는 엔지니어링된 가이드 RNA를 암호화하는 뉴클레오티드 서열이 작동 가능하게 연결된 제2 핵산 구조물; 및
공여자 핵산 분자를 포함하는 제3 핵산 구조물을 포함하는 하나 이상의 벡터를 포함하는
벡터 시스템.
- [청구항 44] 제43항에 있어서,
상기 벡터 시스템에 포함되는 핵산 구조물은 동일하거나 상이한 벡터에 위치하는
벡터 시스템.
- [청구항 45] 제43항에 있어서,
비상동말단연결(Non-homologous end joining; NHEJ)에 관여하는 유전자의 발현을 억제하는 분자를 암호화하는 뉴클레오티드 서열이 작동 가능하게 연결된 하나 이상의 핵산 구조물을 추가로 포함하는
벡터 시스템.
- [청구항 46] 제45항에 있어서,
상기 비상동말단연결에 관여하는 유전자는 ATM1, XRCC4, XLF, XRCC6, LIG4 및 DCLRE1C로 이루어진 군으로부터 선택된 어느 하나 이상인
벡터 시스템.

- [청구항 47] 제45항에 있어서,
상기 분자는 shRNA, siRNA, miRNA, 또는 안티센스 올리고뉴클레오티드인
벡터 시스템.
- [청구항 48] 제43항에 있어서,
벡터 내의 각 구성요소가 하나의 벡터에 포함되는 것인
벡터 시스템.
- [청구항 49] 제43항에 있어서,
상기 공여자 핵산 분자의 길이는 1 bp 내지 20 kb인
벡터 시스템.
- [청구항 50] 제43항에 있어서,
상기 벡터가 프로모터 또는 인핸서를 더 포함하는
벡터 시스템.
- [청구항 51] 제50항에 있어서,
상기 프로모터는 U6 프로모터, EFS 프로모터, EF1- α 프로모터, H1 프로모터, 7SK 프로모터, CMV 프로모터, LTR 프로모터, Ad MLP 프로모터, HSV 프로모터, SV40 프로모터, CBA 프로모터 또는 RSV 프로모터인
벡터 시스템.
- [청구항 52] 제43항에 있어서,
상기 벡터는 레트로바이러스 벡터(retroviral(retrovirus) vector), 렌티바이러스 벡터(lentiviral(lentivirus) vector), 아데노바이러스 벡터(adenoviral(adenovirus vector), 아데노-연관 바이러스 벡터(adeno-associated viral(adeno-associated virus; AAV) vector), 백신시아바이러스 벡터(vaccinia viral(vaccinia virus) vector), 폭스바이러스 벡터(poxviral(poxvirus) vector), 단순포진 바이러스 벡터(herpes simplex viral(herpes simplex virus) vector) 및 파지미드 벡터(phagemid vector)로 구성된 군에서 선택되는 하나 이상의 바이러스 벡터인
벡터 시스템.
- [청구항 53] 제52항에 있어서,
상기 벡터는 아데노-연관 바이러스 벡터이고, 상기 아데노-연관 바이러스 벡터는 상기 벡터 내의 구성요소 모두를 하나의 벡터에 포함할 수 있는
벡터 시스템.
- [청구항 54] 제43항에 있어서,
상기 벡터는 플라스미드, 네이키드 DNA, DNA 복합체, mRNA(전사물) 및 앰플리콘(amplicon)으로 이루어진 군에서 선택된 하나 이상의 비-바이러스 벡터인
벡터 시스템.
- [청구항 55] 제54항에 있어서,

상기 플라스미드는 pcDNA 시리즈, pSC101, pGV1106, pACYC177, ColE1, pKT230, pME290, pBR322, pUC8/9, pUC6, pBD9, pHc79, pIJ61, pLAFR1, pHV14, pGEX 시리즈, pET 시리즈, 및 pUC19으로 이루어진 군에서 선택된 하나 이상인 벡터 시스템.

[청구항 56] 제43항에 있어서, 상기 Cas12f1, TnpB 또는 이의 변이체 단백질은 서열번호 1 내지 5로 이루어진 군에서 선택된 아미노산 서열과 70% 이상의 서열 동일성을 갖는 아미노산 서열을 포함하는

벡터 시스템.

[청구항 57] 제43항에 있어서, 상기 TnpB 단백질은 서열번호 202 내지 서열번호 293로 이루어진 군에서 선택된 아미노산 서열과 70% 이상의 서열 동일성을 갖는 아미노산 서열을 포함하는

벡터 시스템.

[청구항 58] 제43항에 있어서, 상기 Cas12f1, TnpB 또는 이의 변이체 단백질은 하기 서열 중 하나를 포함하는

벡터 시스템:

(i) 서열번호 5의 아미노산 서열;

(ii) 서열번호 1의 아미노산 서열;

(iii) 서열번호 1의 아미노산 서열에서 N-말단의 1개 내지 28개의 아미노산이 제거 또는 치환된 아미노산 서열; 또는

(iv) 서열번호 1의 아미노산 서열의 N-말단 또는 C-말단에 1개 내지 600개의 아미노산이 추가된 아미노산 서열.

[청구항 59] 제58항에 있어서, 상기 추가된 1개 내지 600개의 아미노산은 서열번호 294 또는 서열번호 295의 아미노산 서열인

벡터 시스템.

[청구항 60] 제43항에 있어서, 상기 Cas12f1, TnpB 또는 이의 변이체 단백질은 서열번호 1 내지 서열번호 4로 이루어진 군에서 선택된 아미노산 서열과 70% 이상의 서열 동일성을 갖는

벡터 시스템.

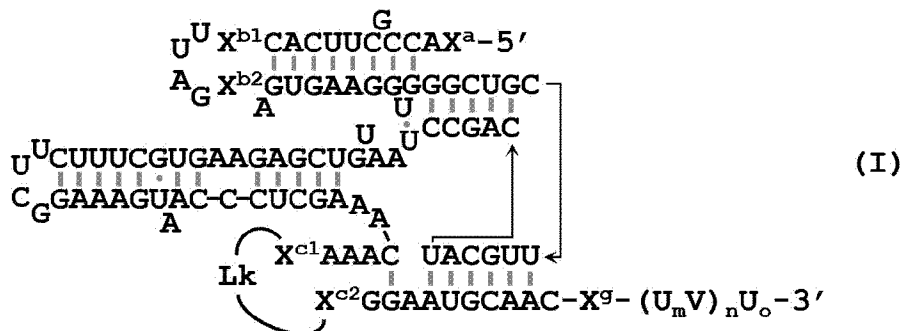
[청구항 61] 제43항에 있어서, 상기 Cas12f1, TnpB 또는 이의 변이체 단백질은 서열번호 1의 아미노산 서열과 70% 이상의 서열 동일성을 갖는

벡터 시스템.

- [청구항 62] 제43항에 있어서,
상기 엔지니어링된 가이드 RNA는 야생형 Cas12f1 가이드 RNA 서열에서 1개 이상의 뉴클레오티드가 치환, 결실, 삽입 또는 부가된 서열을 포함하고, 가이드 서열을 제외한 부분이 상기 야생형 Cas12f1 가이드 RNA와 50% 이상의 서열 동일성을 갖는
벡터 시스템.
- [청구항 63] 제43항에 있어서,
상기 야생형 Cas12f1 가이드 RNA는 (i) 하나 이상의 스템(stem) 영역, (ii) tracrRNA-crRNA 상보성 영역 및 임의적으로 (iii) 연속되는 3개 이상의 유라실(U)을 포함하는 영역을 포함하는 tracrRNA(trans-activating CRISPR RNA) 및 crRNA(CRISPR RNA)를 포함하고,
상기 엔지니어링된 가이드 RNA는 하기 (a) 내지 (d)로 이루어진 군에서 선택되는 하나 이상의 변형을 포함하는
벡터 시스템:
(a) 하나 이상의 스템 영역의 일부 또는 전부의 결실;
(b) tracrRNA-crRNA 상보성 영역의 일부 또는 전부의 결실;
(c) tracrRNA-crRNA 상보성 영역 내에 연속되는 3개 이상의 유라실(U)이 존재하는 경우 하나 이상의 U를 A, G 또는 C로 치환; 및
(d) crRNA 서열의 3'-말단에 하나 이상의 유리딘(uridine)의 부가.
- [청구항 64] 제63항에 있어서,
상기 야생형 Cas12f1 가이드 RNA는 서열번호 11의 핵산 서열로 이루어진 tracrRNA 및 서열번호 12의 핵산 서열로 이루어진 crRNA를 포함하는
벡터 시스템.
- [청구항 65] 제43항에 있어서,
상기 야생형 Cas12f1 가이드 RNA는 5'-말단부터 순차적으로 제1 스템 영역, 제2 스템 영역, 제3 스템 영역, 제4 스템 영역 및 tracrRNA-crRNA 상보성 영역을 포함하는 tracrRNA 및 crRNA를 포함하고,
상기 엔지니어링된 가이드 RNA는
(a1) 제1 스템 영역의 일부 또는 전부의 결실;
(a2) 제2 스템 영역의 일부 또는 전부의 결실;
(b) tracrRNA-crRNA 상보성 영역의 일부 또는 전부의 결실;
(c) tracrRNA-crRNA 상보성 영역 내에 연속되는 3개 이상의 유라실(U)이 존재하는 경우 하나 이상의 U를 A, G 또는 C로 치환; 및
(d1) crRNA 서열의 3'-말단에 U-rich tail의 부가(상기 U-rich tail의 서열은 5'-(U_mV)_nU-3'로 표시되고 여기서 상기 V는 각각 독립적으로 A, C 또는 G이고, m 및 n은 1 내지 20 사이의 정수이며, n은 0 내지 5 사이의 정수임)
로 이루어진 군에서 선택되는 하나 이상의 변형을 포함하는
벡터 시스템.

- [청구항 66] 제65항에 있어서,
상기 엔지니어링된 가이드 RNA는 (d1) crRNA 서열의 3'-말단에 U-rich tail, (c) tracrRNA-crRNA 상보성 영역 내에 연속되는 3개 이상의 유라실 (U)이 존재하는 경우 하나 이상의 U를 A, G 또는 C로 치환, 또는 상기 둘 모두의 변형을 포함하는
벡터 시스템.
- [청구항 67] 제65항에 있어서,
상기 엔지니어링된 가이드 RNA는 (a1) 제1 스템 영역의 일부 또는 전부의 결실; (a2) 제2 스템 영역의 일부 또는 전부의 결실; 및 (b) tracrRNA-crRNA 상보성 영역의 일부 또는 전부의 결실로 이루어진 군에서 선택된 하나 이상의 변형을 포함하는
벡터 시스템.
- [청구항 68] 제67항에 있어서,
상기 엔지니어링된 가이드 RNA는 (b1) tracrRNA-crRNA 상보성 영역의 일부의 결실을 포함하고, 상기 상보성 영역의 일부는 1개 내지 54개의 뉴클레오티드인
벡터 시스템.
- [청구항 69] 제67항에 있어서,
상기 엔지니어링된 가이드 RNA는 (b2) tracrRNA-crRNA 상보성 영역의 전부의 결실을 포함하고, 상기 상보성 영역의 전부는 55개의 뉴클레오티드인
시스템 또는 조성물.
- [청구항 70] 제67항에 있어서,
상기 엔지니어링된 가이드 RNA는 (a1) 제1 스템 영역의 일부 또는 전부의 결실을 포함하고, 상기 스템 영역의 일부 또는 전부는 1개 내지 20개의 뉴클레오티드인
벡터 시스템.
- [청구항 71] 제67항에 있어서,
상기 엔지니어링된 가이드 RNA는 (a2) 제2 스템 영역의 일부 또는 전부의 결실을 포함하고, 상기 스템 영역의 일부 또는 전부는 1개 내지 27개의 뉴클레오티드인
벡터 시스템.
- [청구항 72] 제67항에 있어서,
상기 엔지니어링된 가이드 RNA는 (a1) 제1 스템 영역의 일부 또는 전부의 결실; (d1) crRNA 서열의 3'-말단에 U-rich tail의 부가; 또는 상기 둘 모두의 변형을 포함하는
벡터 시스템.
- [청구항 73] 제43항에 있어서,

상기 엔지니어링된 가이드 RNA는 하기 식 (I)로 표시되는 서열로 이루어지거나 상기 서열과 80% 이상 서열 동일성을 갖는 시스템 또는 조성물:



식 (I)에서,

X^{a} , X^{b1} , X^{b2} , X^{c1} 및 X^{c2} 는 각각 독립적으로 0 내지 35개의 (폴리)뉴클레오티드로 이루어지고,

X^{g} 는 10 내지 30개의 폴리뉴클레오티드로 이루어진 가이드 서열로서 표적 서열과 혼성화하거나 표적 서열에 상보적인 서열이고,

Lk는 길이 2 내지 20의 폴리뉴클레오티드 링커이거나 부존재하고,

$(\text{U}_\text{m} \text{V})_\text{n} \text{U}_\text{o}$ 는 U-rich tail로서 존재하거나 부존재하고, 존재하는 경우 U는 유리딘이고, V는 각각 독립적으로 A, C 또는 G이고, m 및 o는 각각 독립적으로 1 내지 20 사이의 정수이며, n은 0 내지 5 사이의 정수이다.

[청구항 74] 제73항에 있어서,

상기 X^{a} 는 서열번호 14의 핵산 서열을 포함하거나 상기 서열번호 14의 서열에서 1 내지 20개의 뉴클레오티드가 결실된 핵산 서열을 포함하는 벡터 시스템.

[청구항 75] 제73항에 있어서,

상기 X^{b1} 은 서열번호 25의 핵산 서열을 포함하거나 상기 서열번호 25의 서열에서 1 내지 13개의 뉴클레오티드가 결실된 핵산 서열을 포함하는 벡터 시스템.

[청구항 76] 제73항에 있어서,

상기 X^{b2} 은 서열번호 29의 핵산 서열을 포함하거나 상기 서열번호 29의 서열에서 1 내지 14개의 뉴클레오티드가 결실된 핵산 서열을 포함하는 벡터 시스템.

[청구항 77] 제73항에 있어서,

상기 식 (I) 내의 서열 $5' - \text{X}^{\text{b1}} \text{UUAGX}^{\text{b2}} - 3'$ 은 서열번호 34 내지 서열번호 38로 이루어진 군에서 선택된 어느 하나의 핵산 서열인 벡터 시스템.

[청구항 78] 제73항에 있어서,

상기 X^1 은 서열번호 39의 핵산 서열을 포함하거나 상기 서열번호 39의 서열에서 1 내지 28개의 뉴클레오티드가 결실된 핵산 서열을 포함하는 벡터 시스템.

[청구항 79] 제78항에 있어서,
상기 X^1 서열 내에 연속되는 3개 이상의 유라실(U)이 존재하는 경우 이들 중 하나 이상의 U가 A, G 또는 C로 치환되는 변형을 포함하는 벡터 시스템.

[청구항 80] 제73항에 있어서,
상기 X^2 는 서열번호 58의 핵산 서열을 포함하거나 상기 서열번호 58의 서열에서 1 내지 27개의 뉴클레오티드가 결실된 핵산 서열을 포함하는 벡터 시스템.

[청구항 81] 제80항에 있어서,
상기 X^2 서열 내에 서열 5'-ACGAA-3'가 존재하는 경우 해당 서열이 5'-NGNNN-3'로 치환되고, 상기 N은 각각 독립적으로 A, C, G 또는 U인 벡터 시스템.

[청구항 82] 제73항에 있어서,
상기 식 (I) 내의 서열 5'- X^1 -Lk- X^2 -3'은 서열번호 80 내지 서열번호 86으로 이루어진 군에서 선택된 어느 하나의 핵산 서열인 벡터 시스템.

[청구항 83] 제73항에 있어서,
상기 Lk는 5'-GAAA-3', 5'-UUAG-3', 5'-UGAAAA-3', 5'-UUGAAAAA-3', 5'-UUCGAAAGAA-3'(서열번호 76), 5'-UUCAGAAAUGAA-3'(서열번호 77), 5'-UUCAUGAAAAUGAA-3'(서열번호 78) 및 5'-UUCAUUGAAAAAUGAA-3'(서열번호 79)로 이루어진 군에서 선택된 어느 하나의 핵산 서열을 포함하는 벡터 시스템.

[청구항 84] 제73항에 있어서,
상기 $(U_m V)_n U_o$ 는 (i) n 이 0이고, o 가 1 내지 6 사이의 정수이거나 (ii) V 가 A 또는 G이고, m 및 o 는 각각 독립적으로 3 내지 6 사이의 정수이고, n 은 1 내지 3 사이의 정수인 벡터 시스템.

[청구항 85] 제43항에 있어서,
상기 엔지니어링된 가이드 RNA는 서열번호 87 내지 서열번호 132로 이루어진 군에서 선택된 어느 하나의 핵산 서열로 이루어진 엔지니어링된 tracrRNA를 포함하는 벡터 시스템.

[청구항 86] 제43항에 있어서,

상기 엔지니어링된 가이드 RNA는 서열번호 133 내지 서열번호 148로 이루어진 군에서 선택된 어느 하나의 핵산 서열로 이루어진 엔지니어링된 crRNA를 포함하는

벡터 시스템.

[청구항 87] 제43항에 있어서,
상기 엔지니어링된 가이드 RNA는 듀얼 가이드 RNA 또는 싱글 가이드 RNA인

벡터 시스템.

[청구항 88] 제43항에 있어서,
상기 엔지니어링된 싱글 가이드 RNA는 서열번호 149 내지 서열번호 186으로 이루어진 군에서 선택된 어느 하나의 핵산 서열로 이루어진

벡터 시스템.

[청구항 89] 제43항 내지 제88항 중 어느 한 항에 따른 벡터 시스템에 의해 제조된 바이러스.

[청구항 90] 제89항에 있어서,
상기 바이러스는 레트로바이러스, 렌티바이러스, 아데노바이러스, 아데노-연관 바이러스, 백시니아바이러스, 폭스바이러스, 단순포진 바이러스 및 파지(phage)로 구성된 군에서 선택되는 것인

바이러스.

[청구항 91] 제89항에 따른 바이러스를 포함하는 조성물.

[청구항 92] 제1항 내지 제42항 중 어느 한 항의 시스템 또는 조성물, 또는 제43항 내지 제88항 중 어느 한 항의 벡터 시스템을 세포와 접촉시키거나 세포 내에서 발현시키는 단계; 및
표적 핵산 또는 그 인접 부위에서 공여자 핵산 분자를 주형으로 하여 이중가닥 절단의 복구 과정에 의해 목적 서열이 도입되는 단계를 포함하는 세포 내 이중가닥 핵산 상의 표적 부위로 목적 서열을 도입하는 방법.

[청구항 93] 제92항에 있어서,
상기 이중가닥 절단의 복구 과정은 상동지정복구 기작에 의하는 것인 방법.

[청구항 94] 제92항에 있어서,
상기 세포는 표적 핵산 또는 표적 유전자가 존재하는 원핵 세포 또는 진핵 세포인

방법.

[청구항 95] 제94항에 있어서,
상기 진핵 세포는 효모(yeast), 곤충 세포, 식물 세포, 비인간-동물 세포 또는 인간 세포인

방법.

[청구항 96] 제92항에 있어서,

상기 벡터 시스템은 레트로바이러스, 렌티바이러스, 아데노바이러스, 아데노-연관 바이러스, 백시니아바이러스, 폭스바이러스, 단순포진 바이러스 및 파지(phage)로 구성된 군에서 선택된 패키징 바이러스에 도입되어 패키징 바이러스에 의해 생성된 바이러스 형태로 원핵 세포 또는 진핵 세포 내로 전달되는

방법.

[청구항 97] 제92항에 있어서,
상기 벡터 시스템은 전기천공법, 유전자총, 초음파천공법, 자기주입법 (magnetofection), 일시적인 세포 압축 또는 스퀴징 방법, 양이온성 리포솜 법, 초산 리튬-DMSO, 지질-매개 형질감염(transfection), 인산칼슘 침전법 (precipitation), 리포펙타민(lipofection), PEI(Polyethyleneimine)-매개 형질 감염, DEAE-dextran 매개 형질감염 또는 나노파티클-매개 핵산 전달에 의해 원핵 세포 또는 진핵 세포 내로 전달되는

방법.

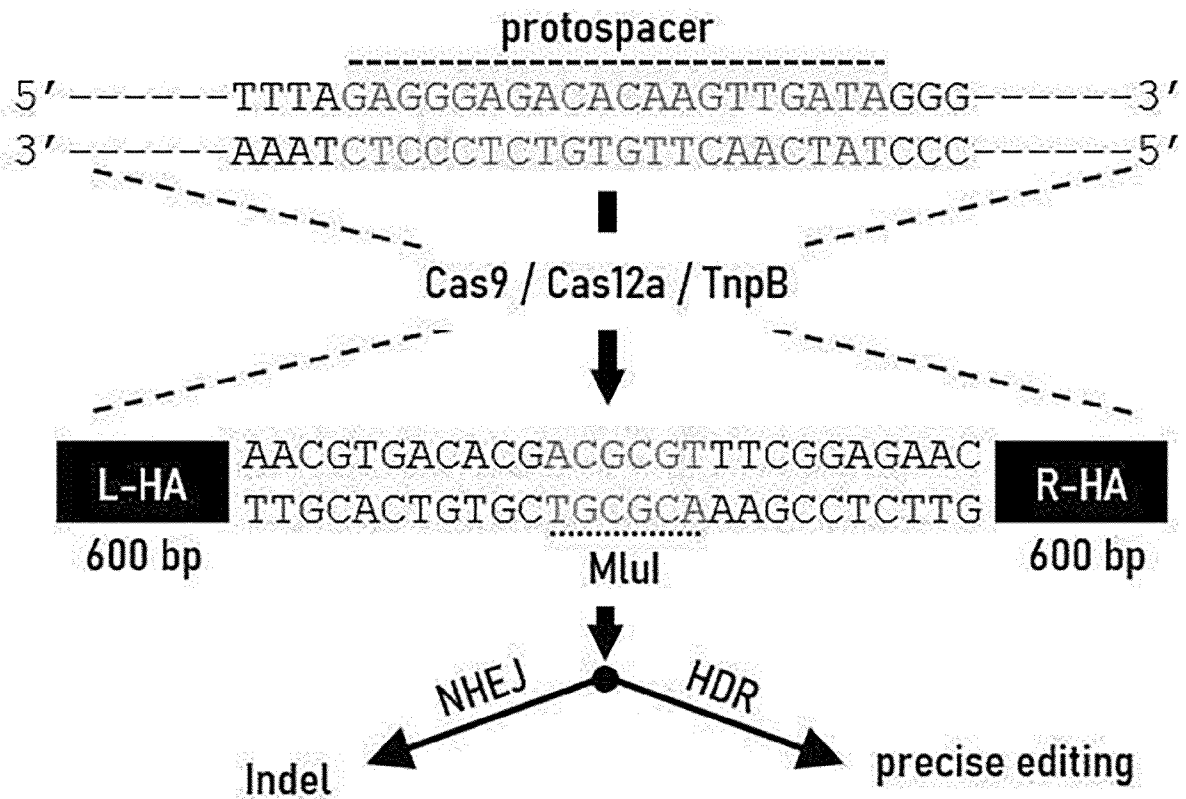
[청구항 98] 제92항에 있어서,
상기 벡터 시스템은 하나 이상의 지질 나노입자(LNP)를 통해 원핵 세포 또는 진핵 세포 내로 직접 전달되는

방법.

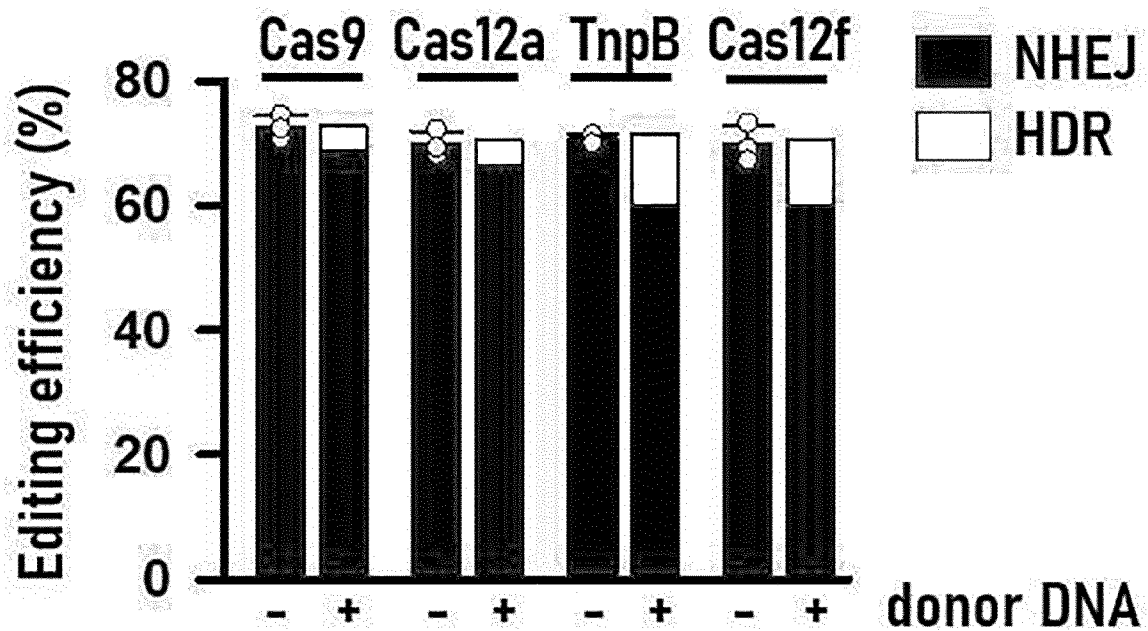
[청구항 99] 제92항에 있어서,
상기 접촉 또는 발현은 생체 내 또는 생체 외에서 일어나는

방법.

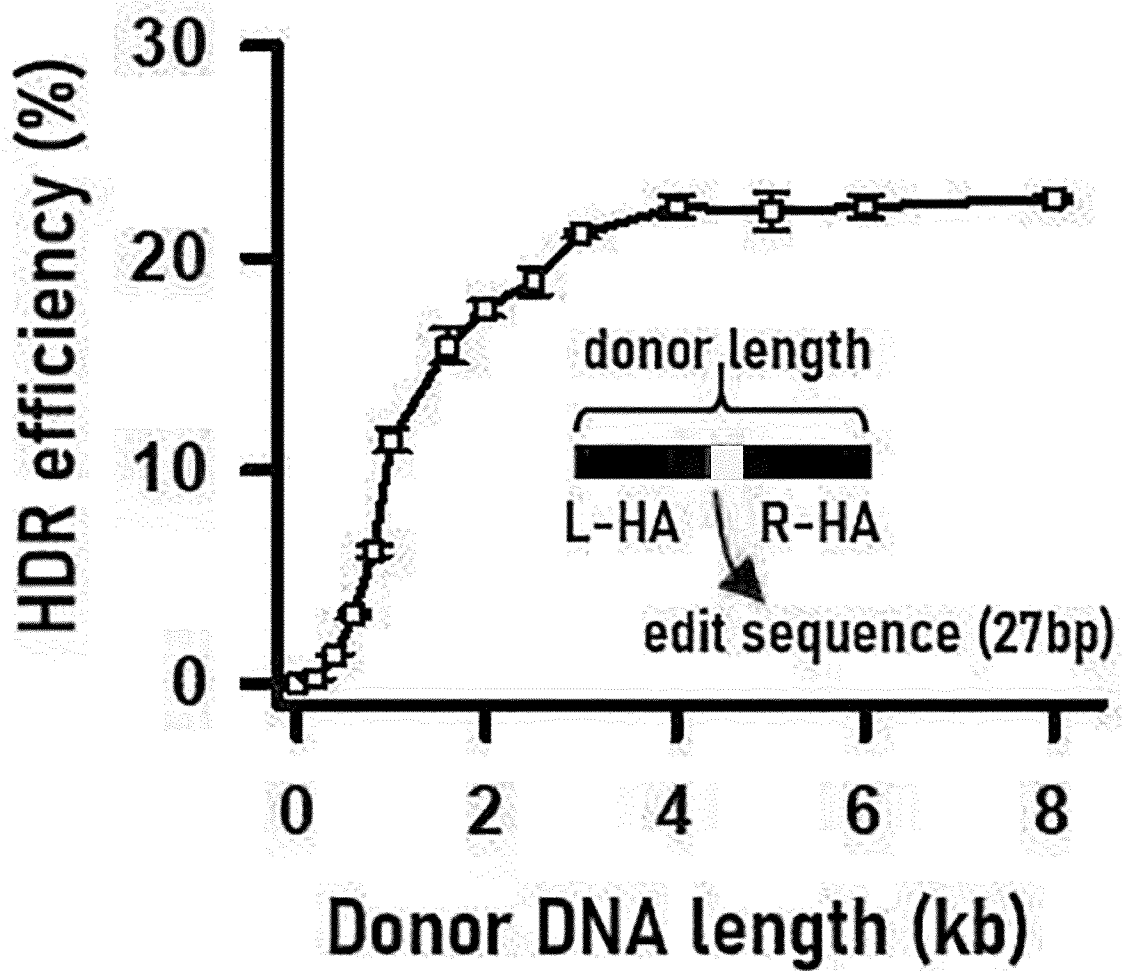
[도2]



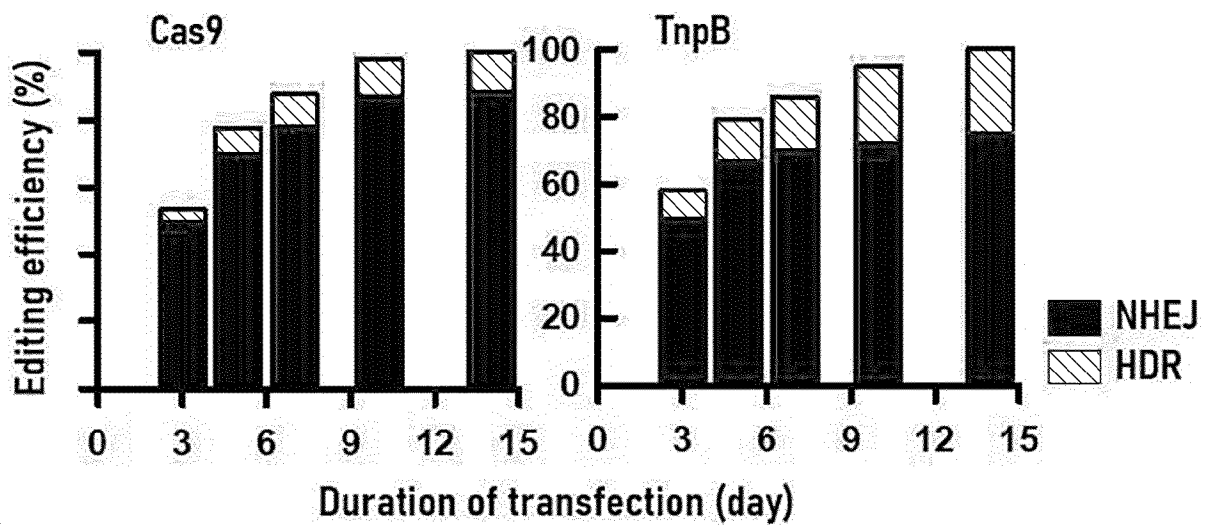
[도3]



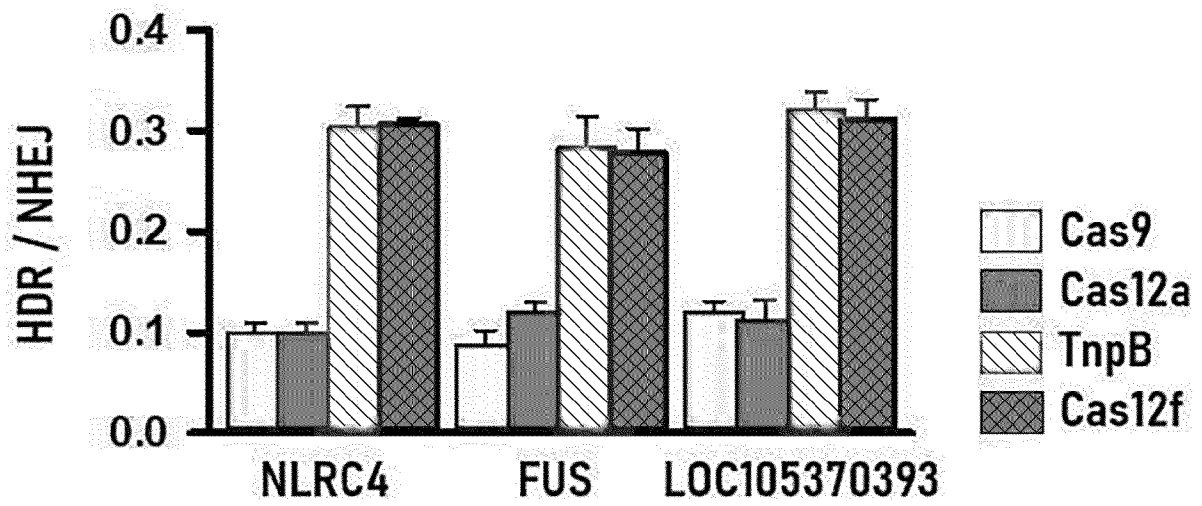
[도4]



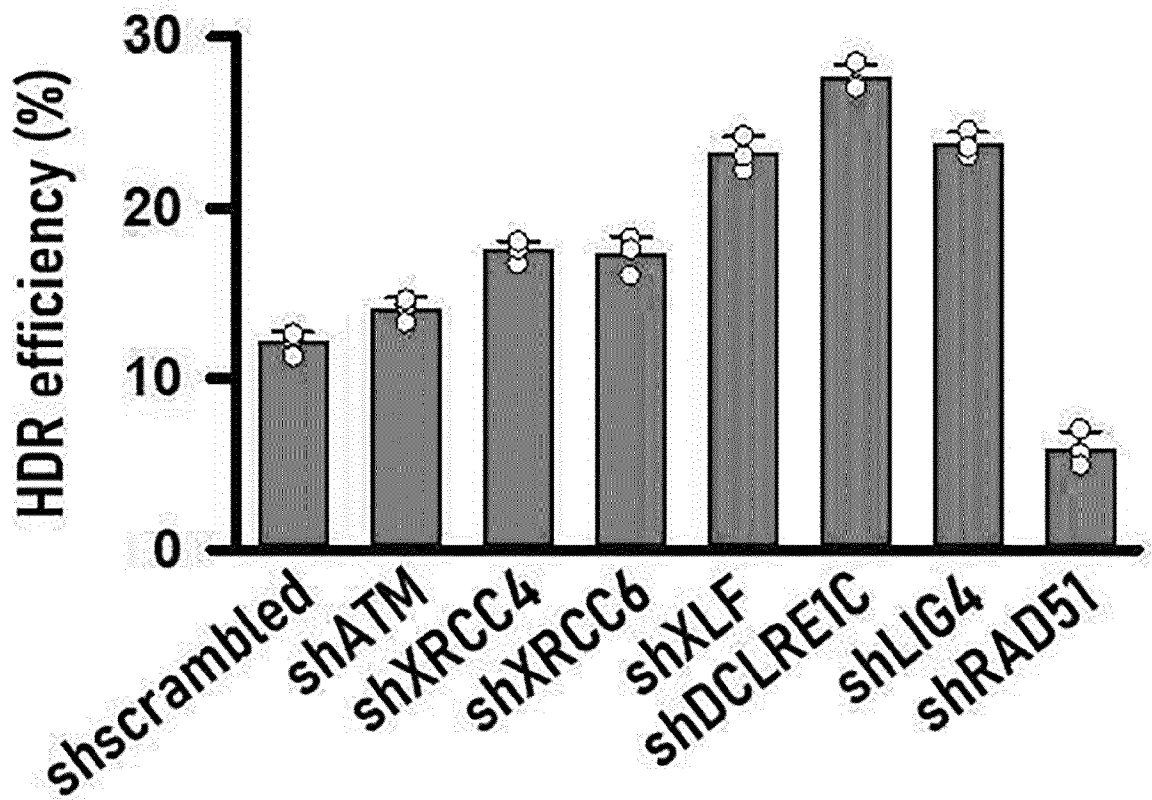
[도5]



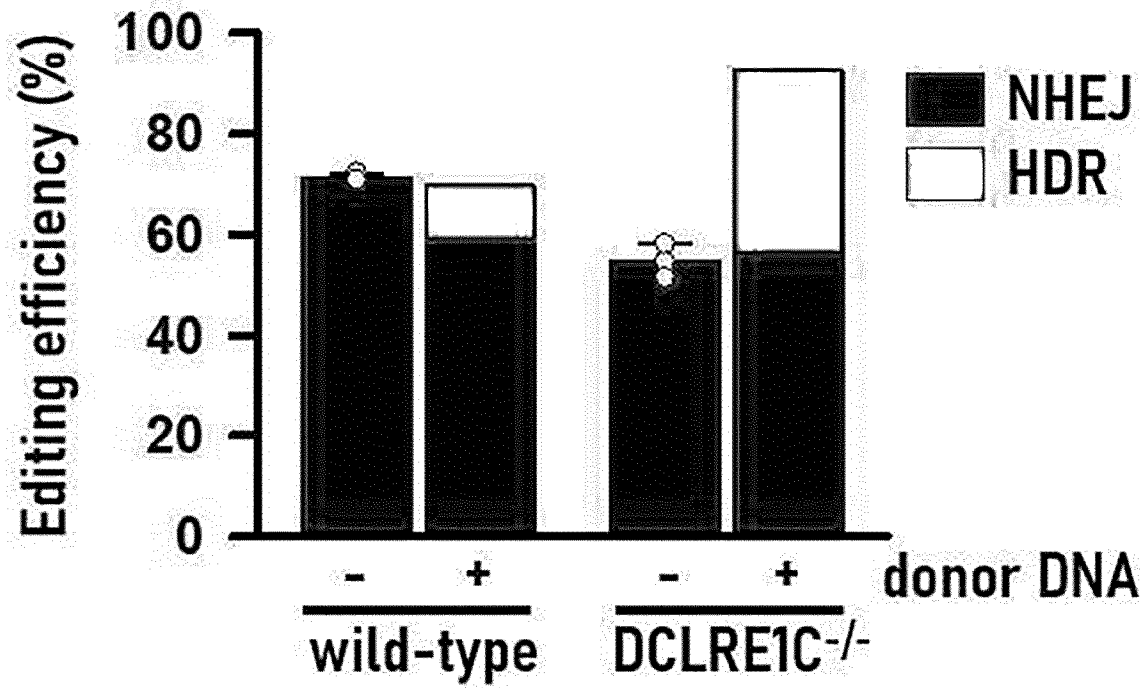
[도6]



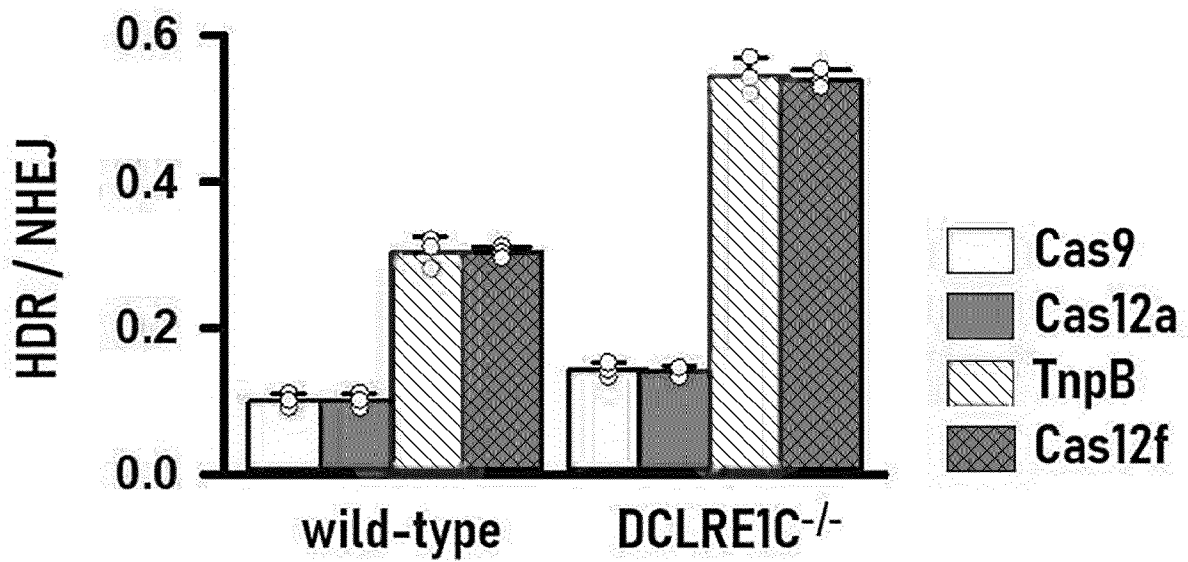
[도7]



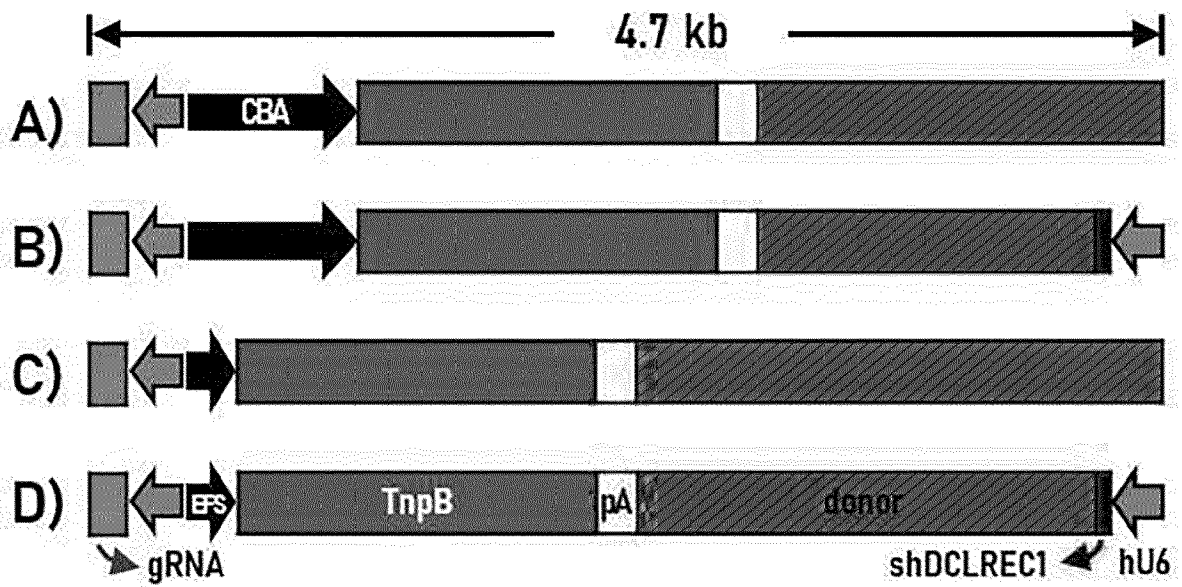
[도8]



[도9]



[도 10a]



[도 10b]

