(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국

(43) 국제공개일
2024 년 4 월 25 일 (25.04.2024)

WIPO | PCT

(10) 국제공개번호
WO 2024/085674 A1

(51) 국제특허분류:
C12N 9/22 (2006.01) *C12N 15/10* (2006.01)
C12N 9/78 (2006.01) *C12N 15/113* (2010.01)
C12N 15/62 (2006.01)

(21) 국제출원번호: PCT/KR2023/016230

(22) 국제출원일: 2023 년 10 월 19 일 (19.10.2023)

(25) 출원언어: 한국어

(26) 공개언어: 한국어

(30) 우선권정보:
10-2022-0134734 2022 년 10 월 19 일 (19.10.2022) KR

(71) 출원인: 재단법인 아산사회복지재단 (**THE ASAN FOUNDATION**) [KR/KR]; 05505 서울특별시 송파구 올림픽로43길 88, Seoul (KR). 울산대학교 산학협력단 (**UNIVERSITY OF ULSAN FOUNDATION FOR INDUSTRY COOPERATION**) [KR/KR]; 44610 울산광역시 남구 대학로 93, Ulsan (KR). 성균관대학교산학협력단 (**RESEARCH&BUSINESS FOUNDATION SUNGKYUNKWAN UNIVERSITY**) [KR/KR]; 16419 경기도 수원시 장안구 서부로 2066, Gyeonggi-do (KR).

(72) 발명자: 김용섭 (**KIM, Yong Sub**); 05505 서울특별시 송파구 올림픽로43길 88, 융합연구관 10층, Seoul (KR). 권지연 (**KWON, Ji Yeon**); 05505 서울특별시 송파구 올림픽로43길 88, 융합연구관 9층, Seoul (KR). 김대식 (**KIM, Dae Sik**); 16419 경기도 수원시 장안구 서부로 2066, 의학관 6305호, Gyeonggi-do (KR).

(74) 대리인: 특허법인더웨이브 (**THEWAVE IP LAW FIRM**); 06097 서울특별시 강남구 선릉로 602 6층, Seoul (KR).

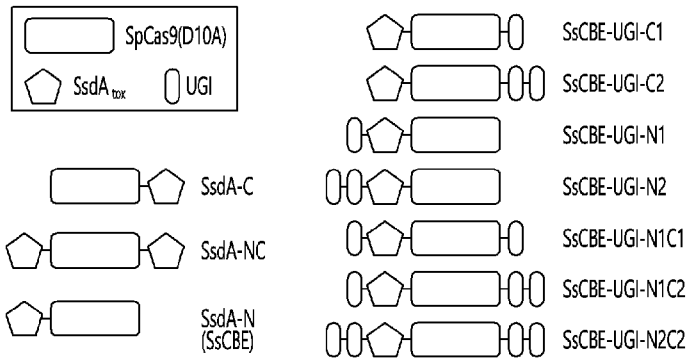
(81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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.

(84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, 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,

(54) **Title:** FUSION PROTEIN CONTAINING CAS PROTEIN AND BACTERIAL TOXIN AND USE THEREOF

(54) **발명의 명칭:** CAS 단백질 및 박테리아 톡신을 포함하는 융합 단백질 및 이의 용도

[도6]



(57) **Abstract:** The present invention relates to a fusion protein containing a CAS protein and a bacterial toxin, and a use thereof. A fusion protein, a polypeptide thereof, and a CRISPR-Cas system including same according to the present invention allow for effective base editing. In addition, because of its small size, the polypeptide facilitates easy delivery through vectors. Furthermore, when indels are induced therethrough, indel efficiency significantly increases, and for indel induction, the size of nucleotides where indels are formed also increases, whereby effective utilization can be made for gene knockout.

(57) **요약서:** Cas 단백질 및 박테리아 톡신을 포함하는 융합 단백질 및 이의 용도에 관한 것으로, 일 양상에 따른 융합 단백질, 이의 폴리펩타이드, 및 이를 포함하는 CRISPR-Cas 시스템에 의하면, 효과적인 염기 교정을 할 수 있다. 또한, 폴리펩타이드의 크기가 작아 벡터를 통한 전달이 용이한 효과가 있다. 또한, 이를 통해 인델을 유도하는 경우, 인델 효율이 현저히 증가하고 인델 유도의 경우 인델 효율이 현저히 상승하며, 인델이 형성되는 뉴클레오타이드의 크기도 증가하여 유전자의 넉아웃(knock-out)에 효과적으로 활용될 수 있다.

WO 2024/085674 A1

[다음 쪽 계속]

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

공개:

— 국제조사보고서와 함께 (조약 제21조(3))

— 명세서의 서열목록 부분과 함께 (규칙 5.2(a))

명세서

발명의 명칭: CAS 단백질 및 박테리아 톡신을 포함하는 융합 단백질 및 이의 용도

기술분야

[1] Cas 단백질 및 박테리아 톡신을 포함하는 융합 단백질 및 이의 용도에 관한 것이다.

배경기술

[2] 유전체 교정(Genome editing)이란 생명체의 유전정보를 자유롭게 교정하는 기술이다. 생명과학분야의 진보와 유전체 서열 분석 기술의 발전을 통해 우리는 다양한 유전정보에 대해 폭넓게 이해할 수 있게 되었다. 예를 들어 동식물의 번식, 질병과 성장, 다양한 인간 유전질병을 일으키는 유전자 변이, 바이오연료의 생산 등을 위한 유전자에 대한 이해는 이미 확보된 상황이지만 이를 직접적으로 활용하여 생명체를 개선하고, 인간 질병을 치료하는 수준에까지 이르기 위해서는 그 이상의 기술 진보가 필수적이다.

[3] 전체 교정 기술은 인간을 포함하여 동물, 식물, 미생물의 유전정보를 변화시켜 그 활용범위를 획기적으로 확장시킬 수 있다. 유전자가위는 원하는 유전정보를 정확히 자를 수 있도록 설계되어 만들어지는 분자 도구로 유전체 교정 기술에서 핵심역할을 하고 있다. 유전자 서열분석 분야를 한 차원 발전시켰던 차세대시퀀싱(Next generation sequencing) 기술과 같이, 유전자가위는 유전정보 활용의 속도와 그 범위를 확장시키고 새로운 산업분야를 창출해 내는 핵심 기술이 되고 있다.

[4] 그러나 Cas9을 활용한 시토신 염기 교정에 흔히 사용되는 단백질인 APOBEC 또는 AID 단백질은 크기가 상대적으로 커 세포치료제로 활용하기 위한 벡터로 제작하기 어렵다는 문제가 있다. 이에, 벡터로 제작하기 용이하면서도 염기 교정 효율이 우수한 CRISPR-Cas 시스템의 제작이 요구되고 있다.

발명의 상세한 설명

기술적 과제

[5] 일 양상은 Cas 단백질(CRISPR-associated protein) 및 박테리아 톡신(toxin)을 포함하는 융합 단백질을 제공하는 것이다.

[6] 다른 양상은 상기 융합 단백질을 암호화하는 폴리뉴클레오티드(polynucleotide)를 제공하는 것이다.

[7] 또 다른 양상은 상기 폴리뉴클레오티드를 포함하는 벡터를 제공하는 것이다.

[8] 또 다른 양상은 상기 융합 단백질 또는 이를 암호화하는 폴리뉴클레오티드; 및 가이드 폴리뉴클레오티드를 포함하는 CRISPR-Cas 시스템을 제공하는 것이다.

[9] 또 다른 양상은 핵산(nucleic acid) 분자를 상기 CRISPR-Cas 시스템과 접촉시키는 단계를 포함하는 핵산을 편집하는 방법을 제공하는 것이다.

과제 해결 수단

[10] 일 양상은 Cas 단백질(CRISPR-associated protein) 및 박테리아 톡신(toxin)을 포함하는 융합 단백질을 제공한다.

[11] 본 명세서에서 용어, "Cas 단백질(CRISPR-associated protein)"은 크리스퍼-결합 엔도뉴클레아제일 수 있다. 상기 Cas 단백질은 특정 표적 폴리뉴클레오티드 서열의 전부 또는 일부를 절단할 수 있다.

[12] 상기 Cas 단백질은 클래스 2 Cas(Class 2 CRISPR associated system) 단백질일 수 있다. 상기 클래스 2 Cas 단백질은 타입 II, 타입 V, 또는 타입 VI 시스템에 포함될 수 있다.

[13] 상기 타입 II 시스템은 cas1, cas2, 및 cas9 유전자를 가질 수 있다. 타입 II 시스템은 3개의 서브타입, 즉 서브타입 II-A, II-B 및 II-C로 추가로 분류될 수 있다. 서브타입 II-A는 추가의 유전자인 csn2를 포함할 수 있다. 서브타입 II-A 시스템을 갖는 유기체로 스트렙토코쿠스 써모필루스(*Streptococcus thermophilus*)가 있을 수 있다. 서브타입 II-B는 csn2가 결여되지만, cas4를 가질 수 있다. 서브타입 II-B 시스템을 갖는 유기체로는 레지오넬라 뉴모필라(*Legionella pneumophila*)가 있을 수 있다. 서브타입 II-C는 박테리아에서 발견되는 가장 일반적인 타입 II 시스템으로, Cas1, Cas2 및 Cas9의 3개의 단백질만을 가질 수 있다. 서브타입 II-C 시스템을 갖는 유기체로는 네이세리아 락타미카(*Neisseria lactamica*)가 있을 수 있다.

[14] 상기 타입 V 시스템은 cas12 유전자 및 cas1 및 cas2 유전자를 가질 수 있다. Cas12 유전자는 Cas9의 각각의 영역에 상동성인 RuvC-유사 뉴클레아제 도메인을 갖지만 Cas9 단백질에 존재하는 HNH 뉴클레아제 도메인이 결여된 단백질 Cas12을 코딩할 수 있다.

[15] 상기 타입 VI 시스템은 cas13 유전자 및 cas1 및 cas2 유전자를 가질 수 있다.

[16] 상기 타입 II 시스템에서, Cas9의 RuvC-유사 뉴클레아제 (RNase H 폴드 도메인 및 HNH (McrA-유사) 뉴클레아제 도메인은 각각 표적 핵산의 가닥 중 하나를 절단할 수 있다. 타입 II 시스템의 Cas9 절단 활성은 또한 Cas9에 의한 crRNA 및 표적 결합을 촉진시키는 듀플렉스를 형성하기 위해 crRNA의 tracrRNA의 혼성화를 필요로 할 수 있다.

[17] 상기 타입 V 시스템에서, Cas12의 RuvC-유사 뉴클레아제 도메인은 표적 핵산의 양쪽 가닥을 엇갈린 구성으로 절단하여 5' 오버행을 생성할 수 있다. 이러한 5' 오버행은 비-상동성 말단 연결 방법을 통한 DNA 삽입을 용이하게 할 수 있다. 타입 V 시스템의 Cas12 절단 활성은 또한 듀플렉스를 형성하기 위해 tracrRNA에 대한 crRNA의 혼성화를 필요로 하지 않고, 타입 V 시스템의 crRNA는 내부 듀플렉스를 형성하는 스템 루프 구조를 갖는 단일 crRNA를 사용할 수 있다. 상기 타입 V 시스템은 표적 서열의 위치에서 단일 또는 이중 가닥 파손을 유도할 수 있다. 가닥 파손은 5' 오버행을 이용한 엇갈린 절단일 수 있다.

[18] 상기 Cas 단백질은 Cas9 또는 Cas12를 포함할 수 있다.

[19] 상기 Cas12 단백질은 다양한 박테리아 종의 유래로부터 유래된 단백질을 의미할 수 있다. 상기 Cas 단백질은 스트렙토코커스(*Streptococcus*), 캄필로박터(*Campylobacter*), 니트라티프락토르(*Nitratifractor*), 스타필로코커스(*Staphylococcus*), 파르비바큘룸(*Parvibaculum*), 로세부리아(*Roseburia*), 네이세리아(*Neisseria*), 글루콘아세토박터(*Gluconacetobacter*), 아조스피릴룸(*Azospirillum*), 스파에로카에타(*Sphaerochaeta*), 락토바실러스(*Lactobacillus*), 유박테리움(*Eubacterium*), 코리네박터(*Corynebacter*), 카르노박테리움(*Carnobacterium*), 로도박터(*Rhodobacter*), 리스테리아(*Listeria*), 팔루디박터(*Paludibacter*), 클로스트리디움(*Clostridium*), 라크노스피라세아에(*Lachnospiraceae*), 클로스트리디아리디움(*Clostridiaridium*), 렙토트리키아(*Leptotrichia*), 프란시셀라(*Francisella*), 레지오넬라(*Legionella*), 알리사이클로바실러스(*Alicyclobacillus*), 메타노메티오필러스(*Methanomethyophilus*), 포르피로모나스(*Porphyromonas*), 프레보텔라(*Prevotella*), 박테로이데테스(*Bacteroidetes*), 헬코코커스(*Helcococcus*), 레토스피라(*Letospira*), 데설포비브리오(*Desulfovibrio*), 데설포나트로눔(*Desulfonatronum*), 오피투타세아에(*Opitutaceae*), 투베리바실러스(*Tuberibacillus*), 바실러스(*Bacillus*), 브레비바실러스(*Brevibacilus*), 메틸로박테리움(*Methylobacterium*) 또는 아시다미노코커스(*Acidaminococcus*) 속 유래의 것일 수 있다. 더욱 구체적으로, 상기 Cas12 단백질은 프란시셀라 툴라렌시스 1, 프란시셀라 툴라렌시스 아종 노비시다, 프레보텔라 알벤시스, 라크노스피라세아에 박테리움 MC2017 1, 부티리비브리오 프로테오클라스티쿠스, 페레그리니박테리아 박테리움 GW2011_GWA2_33_10, 파르쿠박테리아 박테리움 GW2011_GWC2_44_17, 스미텔라 종 SCADC, 아시다미노코커스 종 BV3L6, 라크노스피라세아에 박테리움 MA2020, 칸디다투스 메타노플라즈마 테르미툼, 유박테리움 엘리겐스, 모락셀라 보보쿨리 237, 렙토스피라 이나다이, 라크노스피라세아에 박테리움 ND2006, 포르피로모나스 크레비오리카니스 3, 프레보텔라 디시엔스 및 포르피로모나스 마카카에로 이루어진 군으로부터 선택되는 박테리아 종으로부터 유래된 것일 수 있다.

[20] 또한, 상기 Cas12 단백질은 Cas12a, mgCas12a, Cas12b, Cas12c, Cas12d, Cas12e, Cas12f, Cas12g, Cas12h, Cas12i, Cas12j로 이루어진 군으로부터 선택된 어느 하나일 수 있다. 상기 Cas12 단백질은 Cas12 단백질의 변형을 포함할 수 있다. 상기 Cas12 단백질이 뉴클레아제 활성을 갖는 경우, Cas12 단백질은 감소된 뉴클레아제 활성, 예를 들어 야생형 효소에 비하여 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95% 이상, 적어도 97%, 또는 100%의 뉴클레아제 비활성화를 갖도록 변형될 수 있다.

[21] 상기 Cas9은 사카로마이세스 뉴모니애(*S. pneumoniae*), 스트렙토코커스 피오게네스(*S. pyogenes*), 스트렙토코커스 써모필러스(*S. thermophilus*), 캄필로박터 제주니(*C. jejuni*) Cas9일 수 있으며, 이들 유기체로부터 유래된 돌연변이된 Cas9를 포함할 수 있다. 효소는 Cas9 상동체 또는 오솔로그(ortholog)일 수 있다. 일 구체예에서, CRISPR 효소는 진핵 세포에서의 발현을 위해 코돈-최적화될 수 있다. 일

구체예에서, CRISPR 효소는 표적 서열의 위치에서 1개 또는 2개 가닥의 절단을 유도할 수 있다. 상기 Cas9 단백질은 Cas9 단백질의 변형을 포함할 수 있다. 상기 Cas9 단백질이 뉴클레아제 활성을 갖는 경우, Cas9 단백질은 감소된 뉴클레아제 활성, 예를 들어 야생형 효소에 비하여 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95% 이상, 적어도 97%, 또는 100%의 뉴클레아제 비활성화를 갖도록 변형될 수 있다. 일 구체예에 따르면, 상기 Cas9은 Cas9 D10A일 수 있다.

[22] 일 구체예에 있어서 상기 Cas 단백질은 Cas9 단백질 또는 Cas12 단백질일 수 있다.

[23] 본 발명의 일 구현예에서, 적어도 하나의 핵 국소화 신호(nuclear localization signal; NLS)가 Cas 단백질을 인코딩하는 핵산 서열에 부착될 수 있다. 일 구체예에 있어서, 적어도 하나 이상의 C-말단 또는 N-말단 NLS가 부착될 수 있다. 하나 이상의 핵 국소화 서열(NLS), 예컨대 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 개 이상, 또는 이를 초과하는 NLS를 포함하는 Cas 단백질, 이의 오솔로그 또는 동족체를 인코딩할 수 있다. 본 명세서에 기재된 Cas 단백질 복합체의 바람직한 구현예에서, 코돈 최적화된 Cas 단백질은 단백질의 C-말단에 부착된 NLS를 포함할 수 있다. 특정 구현예에서, 예컨대 Cas를 세포 내 특정 부위, 예컨대 세포 소기관, 예를 들어 미토콘드리아. 플라스티드, 엽록체, 소포, 골지, (핵 또는 세포) 막, 리보솜, 핵소체, ER, 세포골격, 액포, 중심체, 뉴클레오솜, 과립, 중심립 등(이로 제한되지 않음)으로 국소화시키기 위한 기타 다른 국소화 태그를 Cas 단백질에 융합시킬 수 있다.

[24] 상기 Cas 단백질 및 박테리아 톡신은 링커(linker)를 통해 융합될 수 있다. 상기 링커는 Cas 단백질의 C-말단, N-말단, 또는 C-말단과 N-말단 모두 위치할 수 있으며, 상기 링커를 통해 박테리아 톡신이 Cas 단백질에 결합할 수 있다. 상기 적합한 링커 모티프 및 링커 배치형태는 문헌 [Chen et al., Fusion protein linkers: property, design and functionality. Adv Drug Deliv Rev. 2013; 65(10):1357-69]에 기재된 것들을 포함하며, 그의 전체 내용은 본원에 참조로 포함될 수 있다.

[25] 일 구체예에 있어서, 상기 박테리아 톡신은 SsdA(single-stranded DNA deaminase toxin A)일 수 있다.

[26] 상기 SsdA는 슈도모나스 속(*Pseudomonas SP.*) 균주 유래일 수 있다. 상기 슈도모나스 속 균주는 슈도모나스 시링가에(*Pseudomonas syringae*), 슈도모나스 콘겔랑스(*Pseudomonas congelans*), 슈도모나스 사바스타노이(*Pseudomonas savastanoi*), 슈도모나스 비리디플라바(*Pseudomonas viridiflava*), 슈도모나스 코로나파시엔스(*Pseudomonas coronafaciens*), 슈도모나스 플루오레센스 (*Pseudomonas fluorescens*), Pseudomonas sp. MPC6, Pseudomonas sp. GL-R-26, 또는 Pseudomonas sp. GL-RE-26일 수 있다.

[27] 상기 SsdA의 N 말단에는 PAAR 도메인이 있을 수 있으며, SsdA의 C 말단에는 DYW deaminase 도메인이 있을 수 있다. 상기 SsdA는 일반적인 아미노산 motif인 HxE, CxxC motif를 가지는 점에 있어서 기존 염기 교정에 사용된 탈아미노효소와 동일하나, SGW motif를 추가로 가지고 있는 점에서 기존 염기 교정에 사용

된 탈아미노효소와 차이가 있다. 상기 SsdA는 기존에 염기교정기술에 사용된 기타 다른 탈아미노효소와 구조적, 진화적으로 다른 탈아미노효소로 DYW-like deaminase로 분류이다. SsdA와 기존 염기교정기술에 사용된 기타 다른 탈아미노효소의 계통도 및 구성 도메인의 차이를 도 1에 나타내었다.

[28] 일 구체예에 따르면, 상기 SsdA는 서열번호 1의 아미노산 서열을 포함하는 것일 수 있다.

[29] 상기 SsdA는 톡신 도메인(domain)을 포함할 수 있다. 상기 SsdA의 톡신 도메인은 deaminase 활성을 가진 부분으로, 100 내지 200 개의 아미노산의 길이, 예를 들면 120 내지 180 아미노산의 길이, 120 내지 170 아미노산의 길이, 130 내지 160 아미노산의 길이, 또는 140 내지 160 아미노산의 길이를 가질 수 있다. 구체적으로, 상기 SsdA의 톡신 도메인은 서열번호 2의 아미노산 서열을 포함하는 것일 수 있다. 상기 톡신 도메인의 염기서열은 하기 표 1에 나타내었다.

[30] [표1]

	아미노산 서열	서열번호
톡신 도메인(domain)	KVSNIAESEAALGRASQARADLPQSKELKV KTVSSNDKKTLSGWGNKKPEGYERISAEQV KAKSEEIGHEVKSHPYDRDYKGQYFSSHAE KQMSIASPNHPLGVSKPMCTDCQGYFSQLA KYSKVEQTVADPKAIRIFKTDGSVETIMRSE	서열번호 2

[31] 상기 SsdA의 서열번호 2의 아미노산 서열(톡신 도메인)은 촉매활성부위(catalytic active site)를 포함하는 것일 수 있다. 상기 촉매활성부위(catalytic active site)는 HxE motif, CxxC motif 또는 SGW motif를 포함한 것일 수 있다. SGW motif는 APOBEC, AID 등 기존 탈아미노효소와 다르게 SsdA 효소만이 추가적으로 가지고 있는 모티프이다. 구체적으로 HxE motif는 서열번호 1(PAAR domaomn-containing protein)의 301 내지 303 번째 아미노산을 포함할 수 있고, HxE motif는 347 내지 349 번째 아미노산을 포함할 수 있으며, SGW motif는 301 내지 303번째 아미노산을 포함할 수 있다.

[32] 일 구체예에 있어서, 상기 SsdA는 서열번호 1(PAAR domaomn-containing protein)의 아미노산 서열과 적어도 80% 이상, 85% 이상, 90%이상, 91% 이상, 92% 이상, 93% 이상, 94% 이상, 95% 이상, 96% 이상, 97% 이상, 98% 이상 또는 99% 이상의 서열 동일성(sequence identity)을 가지는 것일 수 있다.

[33] 일 구체예에 있어서, 상기 융합 단백질은 Cas 단백질에 DYW deaminase 단백질이 결합한 것일 수 있다.

[34]

[35] 일 구체예에 있어서, 상기 SsdA는 불활성화된 SsdA일 수 있다.

[36] 상기 불활성화된 SsdA는 활성화된 SsdA의 촉매활성부위에서 아미노산 변이가 발생한 것일 수 있다. 상기 불활성화된 SsdA는 SsdA 대비 낮은 세포독성을 갖는 것일 수 있다.

[37] 일 구체예에 따르면, 상기 불활성화된 SsdA는 서열번호 1(PAAR domaomn-containing protein)의 아미노산 서열에서 G302, E349 위치에 아미노산 변이를 갖는 것일 수 있다. 상기 아미노산 변이는 야생형(wild type) 단백질이 G302, E349 위치에서 갖는 아미노산을 제외한 다른 아미노산으로 치환된 것을 의미한다. 상기 다른 아미노산은 아르기닌(R), 히스티딘(H), 라이신(K), 아스파르트산(D), 글루탐산(E), 세린(S), 트레오닌(T), 아스파라진(N), 글루타민(Q), 시스테인(C), 셀레노시스테인(U), 글리신(G), 프롤린(P), 알라닌(A), 발린(V), 이소류신(I), 류신(L), 메티오닌(M), 페닐알라닌(F), 티로신(Y), 트립토판(W) 및 상기 아미노산들의 모든 변형체로 이루어진 군에서 선택된 어느 하나의 것으로서 야생형 단백질이 변이 위치에서 갖는 아미노산을 제외한 아미노산인 것일 수 있다. 구체적으로, 상기 불활성화된 SsdA는 서열번호 1의 아미노산 서열에서 G302D, E349A 또는 이들의 상응하는 아미노산 변이를 갖는 것일 수 있다. 더욱 구체적으로, 상기 서열번호 1(PAAR domaomn-containing protein)의 아미노산 서열에서 G302D의 변이를 갖는 불활성화 된 SsdA는 서열번호 17인 것일 수 있다.

[38] 일 구체예에 있어서, 상기 불활성화 된 SsdA는 서열번호 17의 아미노산 서열과 적어도 80% 이상, 85% 이상, 90%이상, 91% 이상, 92% 이상, 93% 이상, 94% 이상, 95% 이상, 96% 이상, 97% 이상, 98% 이상 또는 99% 이상의 서열 동일성(sequence identity)을 가지는 것일 수 있다.

[39] 상기 SsdA는 단일가닥 DNA의 탈아미노화를 유도할 수 있으며, 상기 SsdA는 시티딘 디아미나제(cytidine deaminase)일 수 있다.

[40] 본 명세서에서 용어, "시티딘 디아미네이즈(cytidine deaminase)"는 시토신(cytosine), 시티딘(cytidine) 또는 데옥시시티딘(deoxycytidine)의 아미노(-NH2) 기를 제거하는 활성을 가지는 효소를 의미한다. 본 명세서에서 상기 시티딘(cytidine) 디아미네이즈는 시토신(cytosine) 디아미네이즈를 포함하는 개념으로 사용된다. 본 명세서에서 상기 시티딘(cytidine) 디아미네이즈는 상기 시토신(cytosine) 디아미네이즈와 혼용되어 사용될 수 있다.

[41] 상기 시티딘 디아미네이즈(cytidine deaminase)는 뉴클레오티드에 존재하는 염기인 시토신 (예컨대, DNA 또는 RNA에 존재하는 시토신)을 우라실로 변환 (C-to-U conversion or C-to-U editing)시키는 활성을 갖는 모든 효소를 의미하는 것으로, 표적 부위의 서열(표적 핵산서열)의 PAM 서열이 존재하는 가닥에 위치하는 시토신을 우라실로 변환시킨다.

[42] 상기 박테리아 톡신은 Cas 단백질의 말단에 결합된 것일 수 있다. 예를 들면, 상기 박테리아 톡신은 Cas 단백질의 C-말단, N-말단, 또는 C-말단과 N-말단 모두에 결합된 것일 수 있다.

[43] 일 구체예에 있어서, 상기 융합 단백질은 DNA 글리코실레이즈 억제제(DNA glycosylase inhibitor)를 더 포함할 수 있다.

[44] 상기 DNA 글리코실레이즈 억제제는 티민 글리코실레이즈 억제제(thymine glycosylase inhibitor), 우라실 글리코실레이즈 억제제(uracil glycosylase inhibitor), 옥소구아닌 글리코실레이즈 억제제(oxoguanine glycosylase inhibitor), 또는 알킬구아닌 DNA 글리코실레이즈 억제제(alkylguanine glycosylase inhibitor)일 수 있다.

[45] 상기 우라실 DNA 글리코실레이즈 억제제는 바실러스 서브틸리스 (Bacillus subtilis) 박테리오파지, PBS1로부터 유래한 우라실 DNA 글리코실라제 억제제, 바실러스 서브틸리스 박테리오파지, 또는 PBS2로부터 유래한 우라실 DNA 글리코실라제 억제제일 수 있으나, 이에 제한되는 것은 아니다.

[46]

[47] 다른 양상은 상기 융합 단백질을 암호화하는 폴리뉴클레오티드(polynucleotide)를 제공한다.

[48] 또 다른 양상은 상기 폴리뉴클레오티드를 포함하는 벡터를 제공한다.

[49] 본 명세서에서, 용어 "벡터"는 그것이 연결되어 있는 다른 핵산을 수송할 수 있는 핵산 분자를 의미할 수 있다. 벡터는, 단일-가닥, 이중-가닥, 또는 부분적 이중-가닥인 핵산 분자; 하나 이상의 자유 말단을 포함하는 핵산 분자, 자유 말단이 없는 핵산 분자(예를 들어, 원형); DNA, RNA 또는 이들 둘 다를 포함하는 핵산 분자; 및 당업계에서 알려진 기타 다른 다양한 폴리뉴클레오티드를 포함할 수 있다. 벡터의 일 종류로 "플라스미드"가 있으며, 이는 원형 이중 가닥 DNA 루프를 지칭할 수 있으며, 그 안으로 예컨대 표준 분자 클로닝 기법에 의해서 추가적인 DNA 분절이 삽입될 수 있다. 벡터의 또 다른 종류로 바이러스 벡터가 있으며, 바이러스-유래된 DNA 또는 RNA 서열이 바이러스(예를 들어, 레트로바이러스, 복제 결함 레트로바이러스, 아데노바이러스, 복제 결함 아데노바이러스, 및 아데노-연관 바이러스) 내로의 패키징을 위해 벡터에 존재할 수 있다. 재조합 발현 벡터는 숙주 세포에서 핵산의 발현에 적합한 형태로 본 발명의 핵산을 포함할 수 있는데, 이는 재조합 발현 벡터가 하나 이상의 조절 요소를 포함하는 것을 의미할 수 있으며, 하나 이상의 조절 요소는 발현에 사용될 숙주 세포에 기반하여 선택될 수 있고, 발현될 핵산 서열에 작동 가능하게 연결될 수 있다. 재조합 발현 벡터 내에서, "작동 가능하게 연결된"은 관심 뉴클레오티드 서열이 (예를 들어, 시험관 내 전사/번역 시스템에서 또는 벡터가 숙주 세포에 도입된 경우에는 숙주 세포에서) 상기 뉴클레오티드 서열의 발현을 허용하는 방식으로 조절 요소(들)에 연결됨을 의미할 수 있다.

[50] 본 명세서에서, 용어 "조절 요소"는 프로모터, 인핸서, 내부 리보솜 엔트리 부위(internal ribosomal 진입 site; IRES), 및 기타 다른 발현 제어 요소(예를 들어, 전사 종결 신호, 예컨대 폴리아데닐화 신호 및 폴리-U 서열)를 포함할 수 있다. 조절 요소는, 예를 들어 문헌[Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS

IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990)]에 기재되어 있다. 조절 요소는 많은 유형의 숙주 세포에서 뉴클레오티드 서열의 구성적 발현을 지시하는 것들과 특정 숙주 세포에서만 뉴클레오티드 서열의 발현을 지시하는 것들(예를 들어, 조직-특이적 조절 서열)을 포함할 수 있다. 일부 구체예에서, 벡터는 하나 이상의 pol III 프로모터 (예를 들어, 1, 2, 3, 4, 5 개 또는 그 이상의 pol III 프로모터), 하나 이상의 pol II 프로모터(예를 들어, 1, 2, 3, 4, 5 개 또는 그 이상의 pol II 프로모터), 하나 이상의 pol I 프로모터(예를 들어, 1, 2, 3, 4, 5 개 또는 그 이상의 pol I 프로모터), 또는 이들의 조합을 포함할 수 있다. pol III 프로모터의 예들은, 이로 제한되지는 않지만, U6 및 H1 프로모터를 포함할 수 있다. pol II 프로모터의 예들은, 이로 제한되지는 않지만, 레트로바이러스 라우스 육종 바이러스(Rous sarcoma virus; RSV) LTR 프로모터(선택적으로 RSV 인핸서와 함께), 사이토메갈로바이러스(CMV) 프로모터(선택적으로 CMV 인핸서와 함께), SV40 프로모터, 디하이드로폴레이트 환원효소 프로모터, β-액틴 프로모터, 포스포글리세롤 키나아제(phosphoglycerol kinase; PGK) 프로모터, 및 EF1α 프로모터를 포함할 수 있다. 예를 들면, 벡터는 렌티바이러스 및 아데노-연관 바이러스(AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, 또는 AAV9)를 포함할 수 있으며, 이와 같은 벡터의 유형은 또한 특정 유형의 세포들의 표적화를 위해 선택될 수 있다.

[51] 또한, 상기 벡터 시스템 내의 복수의 핵산 분자는 동일하거나 상이한 벡터 상에 위치할 수 있다.

[52] 일 구체예에서, 벡터, 예를 들어 플라스미드 또는 바이러스 벡터는, 예를 들어 근육내 주사에 의해 관심 조직으로 전달되는 한편, 기타 다른 경우에 전달은 정맥내, 경피, 비강내, 구강, 점막 또는 기타 다른 전달 방법을 통하여 이루어질 수 있다. 이와 같은 전달은 단일 용량 또는 다중 용량 중 어느 하나를 통하여 이루어질 수 있다. 당업자는, 본 명세서에서 전달되는 실제 투여량은 다양한 인자, 예컨대 벡터 선택, 표적 세포, 유기체, 또는 조직, 치료될 대상체의 일반 상태, 구하는 형질전환/변형의 정도, 투여 경로, 투여 방법, 구하는 형질전환/변형의 형태 등에 따라 크게 달라질 수 있다. 투여량은, 예를 들어 담체(물, 식염수, 에탄올, 글리세롤, 락토오스, 수크로오스, 인산칼슘, 젤라틴, 덱스트란, 한천, 펙틴, 땅콩유, 참기름 등), 희석제, 약학적으로 허용가능한 담체(예를 들어, 인산염 완충 식염수), 약학적으로 허용가능한 부형제, 및/또는 당업계에 알려진 기타 다른 화합물을 추가로 함유할 수 있다. 투여량은 하나 이상의 약학적으로 허용가능한 염, 예를 들어 무기산 염, 예컨대 염산염, 브롬화수소산염, 인산염, 황산염 등; 및 유기산 염, 예컨대 아세테이트, 프로피오네이트, 말로네이트, 벤조에이트 등을 추가로 함유할 수 있다. 추가적으로, 보조 성분, 예컨대 습윤 또는 유화제, pH 완충 성분, 겔 또는 겔화 물질, 풍미제, 착색제, 마이크로스피어, 중합체, 현탁제 등이 또한 본 명세서에서 제시될 수 있다. 추가적으로, 하나 이상의 기타 다른 통상의 약학 성분들, 예컨대 보존제, 보수제, 현탁화제, 계면활성제, 산화방지제, 충전제, 킬레이트화제,

코팅제, 화학 안정화제 등이 또한 존재할 수 있다. 적합한 예시적인 성분으로는, 미세결정성 셀룰로오스, 카르복시메틸셀룰로오스 나트륨, 폴리소르베이트 80, 페닐에틸 알코올, 클로로부탄올, 칼륨 소르베이트, 소르빈산, 이산화황, 프로필 갈레이트, 파라벤, 에틸 바닐린, 글리세린, 페놀, 파라클로로페놀, 젤라틴, 알부민 및 이의 조합을 포함한다. 예를 들면, 질병 치료를 위한 전달은 AAV를 통해 이루어질 수 있다. 인간에 대한 AAV의 생체 내 전달을 위한 치료적으로 유효한 투여량은, 용액 ml 당 약 1Ƴ1010 내지 약 1Ƴ10100의 AAV를 함유하는 약 20 ml 내지 약 50 ml 범위의 식염수 용액일 수 있다. 투여량은 임의의 부작용에 대하여 치료 이익의 균형을 맞추도록 조정될 수 있다.

[53]

[54] 또 다른 양상은 Cas 단백질 및 박테리아 톡신을 포함하는 융합 단백질 또는 상기 융합 단백질을 암호화하는 폴리뉴클레오티드; 및 가이드 폴리뉴클레오티드를 포함하는 CRISPR-Cas 시스템을 제공한다.

[55] 상기 융합 단백질 및 이를 암호화하는 폴리뉴클레오티드는 상기한 바와 같다.

[56] 상기 가이드 폴리뉴클레오티드는 표적화 서열 및/또는 활성화 서열을 포함하는 것일 수 있다.

[57] 본 명세서에서 사용되는 바와 같이, 용어 "표적화 서열"은 표적 핵산 내의 서열에 상보성인 DNA, 또는 DNA 및 RNA의 혼합물을 포함하는 폴리뉴클레오티드를 의미할 수 있다. 특정 실시양태에서, 표적화 서열은 또한 다른 핵산, 또는 핵산 유사체, 또는 이들의 조합물을 포함할 수 있다. 특정 실시양태에서, 표적화 서열은 이러한 구성이 숙주 세포의 내부에서 분해될 가능성이 보다 작기 때문에 단독으로 DNA로 이루어질 수 있다. 일부 실시양태에서, 이러한 구성은 표적 서열 인식 특이성을 증가시키고/시키거나 표적 벗어남 결합/혼성화의 발생을 감소시킬 수 있다. 상기 표적화 서열은 가이드 서열 또는 스페이서 서열을 포함하는 것일 수 있다. 상기 표적화 서열의 도메인의 길이는 적어도 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 23개, 24개, 25개, 26개, 27개, 28개, 29개 또는 30개 뉴클레오티드 길이일 수 있다.

[58] 본 명세서에서, 용어 "활성화 서열"은 Cas 단백질과 상호작용하거나, 회합하거나, 결합할 수 있는 RNA 또는 DNA, 또는 DNA 및 RNA의 혼합물을 포함하는 폴리뉴클레오티드의 일부를 의미할 수 있다. 일 구체예에 있어서, 활성화 영역은 또한 다른 핵산, 또는 핵산 유사체, 또는 이들의 조합물을 포함할 수 있다. 일 구체예에 있어서, 활성화 서열은 표적 서열에 인접하거나, 이에 연결된 것일 수 있다. 일 구체예에 있어서, 활성화 영역은 표적화 영역의 하류에 있을 수 있다. 일 구체예에 있어서, 활성화 영역은 표적화 영역의 상류에 있을 수 있다. 상기 활성화 서열은 직접 반복부 서열, 크리스퍼 RNA(CRISPR RNA: crRNA 및/또는 트랜스-활성화 RNA(trans-activating RNA: tracrRNA)를 포함하는 것일 수 있다.

[59] 일 구체예에 있어서, 가이드 폴리뉴클레오티드, 가이드 RNA, 성숙 crRNA, 및 미성숙 crRNA는 직접 반복부 서열 및 가이드 서열 또는 스페이서 서열을 포함

하거나, 이들로 이루어질 수 있다. 일 구체예에 있어서, 가이드 RNA 또는 성숙 crRNA는 가이드 서열 또는 스페이서 서열에 연결된 직접 반복부 서열을 포함하거나, 이로 이루어질 수 있다. 일 구체예에 있어서, 직접 반복부 서열은 가이드 서열 또는 스페이서 서열로부터 상류(즉, 5')에 위치할 수 있다.

[60] 일 구체예에 있어서, 상기 가이드 폴리뉴클레오티드는 crRNA 및 tracrRNA를 포함할 수 있다.

[61] 일 구체예에 있어서, 상기 가이드 폴리뉴클레오티드는 이중 가이드 RNA(dual guide RNA) 또는 단일-사슬 가이드 RNA(single-chain guide RNA: sgRNA)일 수 있다.

[62] 상기 시스템은 타겟 핵산(nucleic acid) 분자의 뉴클레오타이드(nucleotide) 서열 중 적어도 하나 이상의 뉴클레오타이드의 삭제(deletion), 삽입(insertion), 치환(substitution) 또는 인델(insertion and deletion; indel)을 형성할 수 있다.

[63] 상기 핵산은 RNA 또는 DNA일 수 있다.

[64] 일 구체예에 따르면, 상기 시스템는 타겟 핵산 분자의 뉴클레오타이드 서열에서 1 bp 내지 60 bp, 예를 들면, 1 bp 내지 55 bp, 1 bp 내지 50 bp, 1 bp 내지 45 bp, 1 bp 내지 40 bp, 1 bp 내지 35 bp, 1 bp 내지 30 bp, 1 bp 내지 25 bp, 1 bp 내지 20 bp, 1 bp 내지 15 bp, 1 bp 내지 10 bp, 1 bp 내지 5 bp, 5 bp 내지 60 bp, 5 bp 내지 55 bp, 5 bp 내지 50 bp, 5 bp 내지 45 bp, 5 bp 내지 40 bp, 5 bp 내지 35 bp, 5 bp 내지 30 bp, 5 bp 내지 25 bp, 5 bp 내지 20 bp, 5 bp 내지 15 bp, 5 bp 내지 10 bp, 10 bp 내지 60 bp, 10 bp 내지 55 bp, 10 bp 내지 50 bp, 10 bp 내지 45 bp, 10 bp 내지 40 bp, 10 bp 내지 35 bp, 10 bp 내지 30 bp, 10 bp 내지 25 bp, 10 bp 내지 20 bp, 10 bp 내지 15 bp, 15 bp 내지 60 bp, 15 bp 내지 55 bp, 15 bp 내지 50 bp, 15 bp 내지 45 bp, 15 bp 내지 40 bp, 15 bp 내지 35 bp, 15 bp 내지 30 bp, 15 bp 내지 25 bp, 15 bp 내지 20 bp, 20 bp 내지 60 bp, 20 bp 내지 55 bp, 20 bp 내지 50 bp, 20 bp 내지 45 bp, 20 bp 내지 40 bp, 20 bp 내지 35 bp, 20 bp 내지 30 bp, 20 bp 내지 25 bp, 25 bp 내지 60 bp, 25 bp 내지 55 bp, 25 bp 내지 50 bp, 25 bp 내지 45 bp, 25 bp 내지 40 bp, 25 bp 내지 35 bp, 25 bp 내지 30 bp, 30 bp 내지 60 bp, 30 bp 내지 55 bp, 30 bp 내지 50 bp, 30 bp 내지 45 bp, 30 bp 내지 40 bp, 30 bp 내지 35 bp, 35 bp 내지 60 bp, 35 bp 내지 55 bp, 35 bp 내지 50 bp, 35 bp 내지 45 bp, 35 bp 내지 40 bp, 40 bp 내지 60 bp, 40 bp 내지 55 bp, 40 bp 내지 50 bp, 40 bp 내지 45 bp, 45 bp 내지 60 bp, 45 bp 내지 55 bp, 45 bp 내지 50 bp, 50 bp 내지 60 bp, 50 bp 내지 55 bp 또는 55 bp 내지 60 bp의 뉴클레오타이드의 결실을 형성하는 것일 수 있다.

[65] 일 구체예에 따르면, 상기 시스템은 타겟 핵산 분자의 뉴클레오타이드 서열에서 1 bp 내지 60 bp, 예를 들면, 1 bp 내지 55 bp, 1 bp 내지 50 bp, 1 bp 내지 45 bp, 1 bp 내지 40 bp, 1 bp 내지 35 bp, 1 bp 내지 30 bp, 1 bp 내지 25 bp, 1 bp 내지 20 bp, 1 bp 내지 15 bp, 1 bp 내지 10 bp, 1 bp 내지 5 bp, 5 bp 내지 60 bp, 5 bp 내지 55 bp, 5 bp 내지 50 bp, 5 bp 내지 45 bp, 5 bp 내지 40 bp, 5 bp 내지 35 bp, 5 bp 내지 30 bp, 5 bp 내지 25 bp, 5 bp 내지 20 bp, 5 bp 내지 15 bp, 5 bp 내지 10 bp, 10 bp 내지 60 bp,

10 bp 내지 55 bp, 10 bp 내지 50 bp, 10 bp 내지 45 bp, 10 bp 내지 40 bp, 10 bp 내지 35 bp, 10 bp 내지 30 bp, 10 bp 내지 25 bp, 10 bp 내지 20 bp, 10 bp 내지 15 bp, 15 bp 내지 60 bp, 15 bp 내지 55 bp, 15 bp 내지 50 bp, 15 bp 내지 45 bp, 15 bp 내지 40 bp, 15 bp 내지 35 bp, 15 bp 내지 30 bp, 15 bp 내지 25 bp, 15 bp 내지 20 bp, 20 bp 내지 60 bp, 20 bp 내지 55 bp, 20 bp 내지 50 bp, 20 bp 내지 45 bp, 20 bp 내지 40 bp, 20 bp 내지 35 bp, 20 bp 내지 30 bp, 20 bp 내지 25 bp, 25 bp 내지 60 bp, 25 bp 내지 55 bp, 25 bp 내지 50 bp, 25 bp 내지 45 bp, 25 bp 내지 40 bp, 25 bp 내지 35 bp, 25 bp 내지 30 bp, 30 bp 내지 60 bp, 30 bp 내지 55 bp, 30 bp 내지 50 bp, 30 bp 내지 45 bp, 30 bp 내지 40 bp, 30 bp 내지 35 bp, 35 bp 내지 60 bp, 35 bp 내지 55 bp, 35 bp 내지 50 bp, 35 bp 내지 45 bp, 35 bp 내지 40 bp, 40 bp 내지 60 bp, 40 bp 내지 55 bp, 40 bp 내지 50 bp, 40 bp 내지 45 bp, 45 bp 내지 60 bp, 45 bp 내지 55 bp, 45 bp 내지 50 bp, 50 bp 내지 60 bp, 50 bp 내지 55 bp 또는 55 bp 내지 60 bp의 뉴클레오타이드의 삽입을 형성하는 것일 수 있다.

[66] 일 구체예에 따르면, 상기 시스템은 타겟 핵산 분자의 뉴클레오타이드 서열에서 1 bp 내지 60 bp, 예를 들면, 1 bp 내지 55 bp, 1 bp 내지 50 bp, 1 bp 내지 45 bp, 1 bp 내지 40 bp, 1 bp 내지 35 bp, 1 bp 내지 30 bp, 1 bp 내지 25 bp, 1 bp 내지 20 bp, 1 bp 내지 15 bp, 1 bp 내지 10 bp, 1 bp 내지 5 bp, 5 bp 내지 60 bp, 5 bp 내지 55 bp, 5 bp 내지 50 bp, 5 bp 내지 45 bp, 5 bp 내지 40 bp, 5 bp 내지 35 bp, 5 bp 내지 30 bp, 5 bp 내지 25 bp, 5 bp 내지 20 bp, 5 bp 내지 15 bp, 5 bp 내지 10 bp, 10 bp 내지 60 bp, 10 bp 내지 55 bp, 10 bp 내지 50 bp, 10 bp 내지 45 bp, 10 bp 내지 40 bp, 10 bp 내지 35 bp, 10 bp 내지 30 bp, 10 bp 내지 25 bp, 10 bp 내지 20 bp, 10 bp 내지 15 bp, 15 bp 내지 60 bp, 15 bp 내지 55 bp, 15 bp 내지 50 bp, 15 bp 내지 45 bp, 15 bp 내지 40 bp, 15 bp 내지 35 bp, 15 bp 내지 30 bp, 15 bp 내지 25 bp, 15 bp 내지 20 bp, 20 bp 내지 60 bp, 20 bp 내지 55 bp, 20 bp 내지 50 bp, 20 bp 내지 45 bp, 20 bp 내지 40 bp, 20 bp 내지 35 bp, 20 bp 내지 30 bp, 20 bp 내지 25 bp, 25 bp 내지 60 bp, 25 bp 내지 55 bp, 25 bp 내지 50 bp, 25 bp 내지 45 bp, 25 bp 내지 40 bp, 25 bp 내지 35 bp, 25 bp 내지 30 bp, 30 bp 내지 60 bp, 30 bp 내지 55 bp, 30 bp 내지 50 bp, 30 bp 내지 45 bp, 30 bp 내지 40 bp, 30 bp 내지 35 bp, 35 bp 내지 60 bp, 35 bp 내지 55 bp, 35 bp 내지 50 bp, 35 bp 내지 45 bp, 35 bp 내지 40 bp, 40 bp 내지 60 bp, 40 bp 내지 55 bp, 40 bp 내지 50 bp, 40 bp 내지 45 bp, 45 bp 내지 60 bp, 45 bp 내지 55 bp, 45 bp 내지 50 bp, 50 bp 내지 60 bp, 50 bp 내지 55 bp 또는 55 bp 내지 60 bp의 뉴클레오타이드의 인델을 형성하는 것일 수 있다.

[67] 일 구체예에 따르면, 상기 시스템은 뉴클레오타이드의 인델을 형성하는 효율이 5% 내지 50%, 예를 들면, 5% 내지 45%, 5% 내지 40%, 5% 내지 35%, 5% 내지 30%, 5% 내지 25%, 5% 내지 20%, 5% 내지 15%, 5% 내지 10%, 10% 내지 50%, 10% 내지 45%, 10% 내지 40%, 10% 내지 35%, 10% 내지 30%, 10% 내지 25%, 10% 내지 20%, 10% 내지 15%, 15% 내지 50%, 15% 내지 45%, 15% 내지 40%, 15% 내지 35%, 15% 내지 30%, 15% 내지 25%, 15% 내지 20%, 20% 내지 50%,

20% 내지 45%, 20% 내지 40%, 20% 내지 35%, 20% 내지 30%, 20% 내지 25%, 25% 내지 50%, 25% 내지 45%, 25% 내지 40%, 25% 내지 35%, 25% 내지 30%, 30% 내지 50%, 30% 내지 45%, 30% 내지 40%, 30% 내지 35%, 35% 내지 50%, 35% 내지 45%, 35% 내지 40%, 40% 내지 50%, 40% 내지 45%, 또는 45% 내지 50%일 수 있다.

[68] 일 구체예에 따르면, 상기 시스템은 뉴클레오타이드의 서열의 치환을 형성하는 효율이 1% 내지 20%, 예를 들면, 1% 내지 18%, 1% 내지 16%, 1% 내지 14%, 1% 내지 12%, 1% 내지 10%, 1% 내지 8%, 1% 내지 6%, 1% 내지 4%, 1% 내지 2%, 2% 내지 20%, 2% 내지 18%, 2% 내지 16%, 2% 내지 14%, 2% 내지 12%, 2% 내지 10%, 2% 내지 8%, 2% 내지 6%, 2% 내지 4%, 4% 내지 20%, 4% 내지 18%, 4% 내지 16%, 4% 내지 14%, 4% 내지 12%, 4% 내지 10%, 4% 내지 8%, 4% 내지 6%, 6% 내지 20%, 6% 내지 18%, 6% 내지 16%, 6% 내지 14%, 6% 내지 12%, 6% 내지 10%, 6% 내지 8%, 8% 내지 20%, 8% 내지 18%, 8% 내지 16%, 8% 내지 14%, 8% 내지 12%, 8% 내지 10%, 10% 내지 20%, 10% 내지 18%, 10% 내지 16%, 10% 내지 14%, 10% 내지 12%, 12% 내지 20%, 12% 내지 18%, 12% 내지 16%, 12% 내지 14%, 14% 내지 20%, 14% 내지 18%, 14% 내지 16%, 16% 내지 20%, 16% 내지 18%, 또는 18% 내지 20%일 수 있다.

[69] 상기 시스템은 타겟 핵산(nucleic acid) 분자의 뉴클레오타이드(nucleotide) 서열 중 적어도 4 뉴클레오타이드의 교정 윈도우(editing window)을 형성할 수 있다. 일 구체예에 따르면, 상기 시스템은 적어도 50 뉴클레오티드, 예를 들면, 적어도 49 뉴클레오티드, 적어도 48 뉴클레오티드, 적어도 47 뉴클레오티드, 적어도 46 뉴클레오티드, 적어도 45 뉴클레오티드, 적어도 44 뉴클레오티드, 적어도 43 뉴클레오티드, 적어도 42 뉴클레오티드, 적어도 41 뉴클레오티드, 적어도 40 뉴클레오티드, 적어도 39 뉴클레오티드, 적어도 38 뉴클레오티드, 적어도 37 뉴클레오티드, 적어도 36 뉴클레오티드, 적어도 35 뉴클레오티드, 적어도 34 뉴클레오티드, 적어도 33 뉴클레오티드, 적어도 32 뉴클레오티드, 적어도 31 뉴클레오티드, 적어도 30 뉴클레오티드, 적어도 29 뉴클레오티드, 적어도 28 뉴클레오티드, 적어도 27 뉴클레오티드, 적어도 26 뉴클레오티드, 적어도 25 뉴클레오티드, 적어도 24 뉴클레오티드, 적어도 23 뉴클레오티드, 적어도 22 뉴클레오티드, 적어도 21 뉴클레오티드, 적어도 20 뉴클레오티드, 적어도 19 뉴클레오티드, 적어도 18 뉴클레오티드, 적어도 17 뉴클레오티드, 적어도 16 뉴클레오티드, 적어도 15 뉴클레오티드, 적어도 14 뉴클레오티드, 적어도 13 뉴클레오티드, 적어도 12 뉴클레오티드, 적어도 11 뉴클레오티드, 적어도 10 뉴클레오티드, 적어도 9 뉴클레오티드, 적어도 8 뉴클레오티드, 적어도 7 뉴클레오티드, 적어도 6 뉴클레오티드, 적어도 5 뉴클레오티드 또는 적어도 4 뉴클레오티드의 교정 윈도우(editing window)를 갖는 것일 수 있다.

[70] 일 구체예에 따르면, 상기 시스템은 gRNA 표적 염기서열의 5' 말단으로부터 1 bp 내지 20 bp, 1 bp 내지 19 bp, 1 bp 내지 18 bp, 1 bp 내지 17 bp, 1 bp 내지 16 bp,

1 bp 내지 15 bp, 1 bp 내지 14 bp, 1 bp 내지 13 bp, 1 bp 내지 12 bp, 1 bp 내지 11 bp, 1 bp 내지 10 bp, 1 bp 내지 9 bp, 1 bp 내지 8 bp, 2 bp 내지 20 bp, 2 bp 내지 19 bp, 2 bp 내지 18 bp, 2 bp 내지 17 bp, 2 bp 내지 16 bp, 2 bp 내지 15 bp, 2 bp 내지 14 bp, 2 bp 내지 13 bp, 2 bp 내지 12 bp, 2 bp 내지 11 bp, 2 bp 내지 10 bp, 2 bp 내지 9 bp, 2 bp 내지 8 bp, 3 bp 내지 20 bp, 3 bp 내지 19 bp, 3 bp 내지 18 bp, 3 bp 내지 17 bp, 3 bp 내지 16 bp, 3 bp 내지 15 bp, 3 bp 내지 14 bp, 3 bp 내지 13 bp, 3 bp 내지 12 bp, 3 bp 내지 11 bp, 3 bp 내지 10 bp, 3 bp 내지 9 bp, 3 bp 내지 8 bp, 4 bp 내지 20 bp, 4 bp 내지 19 bp, 4 bp 내지 18 bp, 4 bp 내지 17 bp, 4 bp 내지 16 bp, 4 bp 내지 15 bp, 4 bp 내지 14 bp, 4 bp 내지 13 bp, 4 bp 내지 12 bp, 4 bp 내지 11 bp, 4 bp 내지 10 bp, 4 bp 내지 9 bp 또는 4 bp 내지 8 bp의 교정 윈도우(editing window)를 갖는 것일 수 있다.

[71] 일 구체예에 따르면, 상기 뉴클레오타이드 교정 윈도우(Editing window)는 gRNA 표적 염기서열의 5' 말단으로부터 첫번째 위치한 시토신(C_1) 내지 20번째 위치한 시토신(C_{20}), C_2 내지 C_{20}, C_3 내지 C_{20}, C_4 내지 C_{20}, C_1 내지 C_{19}, C_2 내지 C_{19}, C_3 내지 C_{19}, C_4 내지 C_{19}, C_1 내지 C_{18}, C_2 내지 C_{18}, C_3 내지 C_{18}, C_4 내지 C_{18}, C_1 내지 C_{17}, C_2 내지 C_{17}, C_3 내지 C_{17}, C_4 내지 C_{17}, C_1 내지 C_{16}, C_2 내지 C_{16}, C_3 내지 C_{16}, C_4 내지 C_{16}, C_1 내지 C_{15}, C_2 내지 C_{15}, C_3 내지 C_{15}, C_4 내지 C_{15}, C_1 내지 C_{14}, C_2 내지 C_{14}, C_3 내지 C_{14}, C_4 내지 C_{14}, C_1 내지 C_{13}, C_2 내지 C_{13}, C_3 내지 C_{13}, C_4 내지 C_{13}, C_1 내지 C_{12}, C_2 내지 C_{12}, C_3 내지 C_{12}, C_4 내지 C_{12}, C_1 내지 C_{11}, C_2 내지 C_{11}, C_3 내지 C_{11}, C_4 내지 C_{11}, C_1 내지 C_{10}, C_2 내지 C_{10}, C_3 내지 C_{10}, C_4 내지 C_{10}, C_1 내지 C_9, C_2 내지 C_9, C_3 내지 C_9, C_4 내지 C_9, C_1 내지 C_8, C_2 내지 C_8, C_3 내지 C_8 또는 C_4 내지 C_8,범위에서 염기 교정을 나타내는 것일 수 있다.

[72] 또 다른 양상은 핵산 분자를 상기 CRISPR-Cas 시스템과 접촉시키는 단계를 포함하는 핵산을 편집하는 방법을 제공한다.

[73] 상기 핵산, 및 CRISPR-Cas 시스템은 상기한 바와 같다.

[74] 일 구체예에 있어서, 상기 편집은 상기 핵산 분자의 뉴클레오타이드 서열 중 적어도 하나 이상의 뉴클레오타이드 서열의 결실, 삽입, 치환 또는 인델을 형성하는 것일 수 있다.

[75] 일 구체예에 따르면, 상기 핵산을 변형하는 방법에서 변형은 핵산 분자의 뉴클레오타이드 서열 중 1 bp 내지 60 bp, 예를 들면, 1 bp 내지 55 bp, 1 bp 내지 50 bp, 1 bp 내지 45 bp, 1 bp 내지 40 bp, 1 bp 내지 35 bp, 1 bp 내지 30 bp, 1 bp 내지 25 bp, 1 bp 내지 20 bp, 1 bp 내지 15 bp, 1 bp 내지 10 bp, 1 bp 내지 5 bp, 5 bp 내지 60 bp, 5 bp 내지 55 bp, 5 bp 내지 50 bp, 5 bp 내지 45 bp, 5 bp 내지 40 bp, 5 bp 내지 35 bp, 5 bp 내지 30 bp, 5 bp 내지 25 bp, 5 bp 내지 20 bp, 5 bp 내지 15 bp, 5 bp 내지 10 bp, 10 bp 내지 60 bp, 10 bp 내지 55 bp, 10 bp 내지 50 bp, 10 bp 내지 45 bp, 10 bp 내지 40 bp, 10 bp 내지 35 bp, 10 bp 내지 30 bp, 10 bp 내지 25 bp, 10 bp 내지 20 bp, 10 bp 내지 15 bp, 15 bp 내지 60 bp, 15 bp 내지 55 bp, 15 bp 내지 50 bp, 15 bp 내지 45

bp, 15 bp 내지 40 bp, 15 bp 내지 35 bp, 15 bp 내지 30 bp, 15 bp 내지 25 bp, 15 bp 내지 20 bp, 20 bp 내지 60 bp, 20 bp 내지 55 bp, 20 bp 내지 50 bp, 20 bp 내지 45 bp, 20 bp 내지 40 bp, 20 bp 내지 35 bp, 20 bp 내지 30 bp, 20 bp 내지 25 bp, 25 bp 내지 60 bp, 25 bp 내지 55 bp, 25 bp 내지 50 bp, 25 bp 내지 45 bp, 25 bp 내지 40 bp, 25 bp 내지 35 bp, 25 bp 내지 30 bp, 30 bp 내지 60 bp, 30 bp 내지 55 bp, 30 bp 내지 50 bp, 30 bp 내지 45 bp, 30 bp 내지 40 bp, 30 bp 내지 35 bp, 35 bp 내지 60 bp, 35 bp 내지 55 bp, 35 bp 내지 50 bp, 35 bp 내지 45 bp, 35 bp 내지 40 bp, 40 bp 내지 60 bp, 40 bp 내지 55 bp, 40 bp 내지 50 bp, 40 bp 내지 45 bp, 45 bp 내지 60 bp, 45 bp 내지 55 bp, 45 bp 내지 50 bp, 50 bp 내지 60 bp, 50 bp 내지 55 bp 또는 55 bp 내지 60 bp의 뉴클레오타이드의 결실을 형성하는 것일 수 있다.

[76] 일 구체예에 따르면, 상기 핵산을 변형하는 방법에서 변형은 핵산 분자의 뉴클레오타이드 서열 중 1 bp 내지 60 bp, 예를 들면, 1 bp 내지 55 bp, 1 bp 내지 50 bp, 1 bp 내지 45 bp, 1 bp 내지 40 bp, 1 bp 내지 35 bp, 1 bp 내지 30 bp, 1 bp 내지 25 bp, 1 bp 내지 20 bp, 1 bp 내지 15 bp, 1 bp 내지 10 bp, 1 bp 내지 5 bp, 5 bp 내지 60 bp, 5 bp 내지 55 bp, 5 bp 내지 50 bp, 5 bp 내지 45 bp, 5 bp 내지 40 bp, 5 bp 내지 35 bp, 5 bp 내지 30 bp, 5 bp 내지 25 bp, 5 bp 내지 20 bp, 5 bp 내지 15 bp, 5 bp 내지 10 bp, 10 bp 내지 60 bp, 10 bp 내지 55 bp, 10 bp 내지 50 bp, 10 bp 내지 45 bp, 10 bp 내지 40 bp, 10 bp 내지 35 bp, 10 bp 내지 30 bp, 10 bp 내지 25 bp, 10 bp 내지 20 bp, 10 bp 내지 15 bp, 15 bp 내지 60 bp, 15 bp 내지 55 bp, 15 bp 내지 50 bp, 15 bp 내지 45 bp, 15 bp 내지 40 bp, 15 bp 내지 35 bp, 15 bp 내지 30 bp, 15 bp 내지 25 bp, 15 bp 내지 20 bp, 20 bp 내지 60 bp, 20 bp 내지 55 bp, 20 bp 내지 50 bp, 20 bp 내지 45 bp, 20 bp 내지 40 bp, 20 bp 내지 35 bp, 20 bp 내지 30 bp, 20 bp 내지 25 bp, 25 bp 내지 60 bp, 25 bp 내지 55 bp, 25 bp 내지 50 bp, 25 bp 내지 45 bp, 25 bp 내지 40 bp, 25 bp 내지 35 bp, 25 bp 내지 30 bp, 30 bp 내지 60 bp, 30 bp 내지 55 bp, 30 bp 내지 50 bp, 30 bp 내지 45 bp, 30 bp 내지 40 bp, 30 bp 내지 35 bp, 35 bp 내지 60 bp, 35 bp 내지 55 bp, 35 bp 내지 50 bp, 35 bp 내지 45 bp, 35 bp 내지 40 bp, 40 bp 내지 60 bp, 40 bp 내지 55 bp, 40 bp 내지 50 bp, 40 bp 내지 45 bp, 45 bp 내지 60 bp, 45 bp 내지 55 bp, 45 bp 내지 50 bp, 50 bp 내지 60 bp, 50 bp 내지 55 bp 또는 55 bp 내지 60 bp의 뉴클레오타이드의 삽입을 형성하는 것일 수 있다.

[77] 일 구체예에 따르면, 상기 핵산을 변형하는 방법에서 변형은 핵산 분자의 뉴클레오타이드 서열 중 1 bp 내지 60 bp, 예를 들면, 1 bp 내지 55 bp, 1 bp 내지 50 bp, 1 bp 내지 45 bp, 1 bp 내지 40 bp, 1 bp 내지 35 bp, 1 bp 내지 30 bp, 1 bp 내지 25 bp, 1 bp 내지 20 bp, 1 bp 내지 15 bp, 1 bp 내지 10 bp, 1 bp 내지 5 bp, 5 bp 내지 60 bp, 5 bp 내지 55 bp, 5 bp 내지 50 bp, 5 bp 내지 45 bp, 5 bp 내지 40 bp, 5 bp 내지 35 bp, 5 bp 내지 30 bp, 5 bp 내지 25 bp, 5 bp 내지 20 bp, 5 bp 내지 15 bp, 5 bp 내지 10 bp, 10 bp 내지 60 bp, 10 bp 내지 55 bp, 10 bp 내지 50 bp, 10 bp 내지 45 bp, 10 bp 내지 40 bp, 10 bp 내지 35 bp, 10 bp 내지 30 bp, 10 bp 내지 25 bp, 10 bp 내지 20 bp, 10 bp 내지 15 bp, 15 bp 내지 60 bp, 15 bp 내지 55 bp, 15 bp 내지 50 bp, 15 bp 내지 45

bp, 15 bp 내지 40 bp, 15 bp 내지 35 bp, 15 bp 내지 30 bp, 15 bp 내지 25 bp, 15 bp 내지 20 bp, 20 bp 내지 60 bp, 20 bp 내지 55 bp, 20 bp 내지 50 bp, 20 bp 내지 45 bp, 20 bp 내지 40 bp, 20 bp 내지 35 bp, 20 bp 내지 30 bp, 20 bp 내지 25 bp, 25 bp 내지 60 bp, 25 bp 내지 55 bp, 25 bp 내지 50 bp, 25 bp 내지 45 bp, 25 bp 내지 40 bp, 25 bp 내지 35 bp, 25 bp 내지 30 bp, 30 bp 내지 60 bp, 30 bp 내지 55 bp, 30 bp 내지 50 bp, 30 bp 내지 45 bp, 30 bp 내지 40 bp, 30 bp 내지 35 bp, 35 bp 내지 60 bp, 35 bp 내지 55 bp, 35 bp 내지 50 bp, 35 bp 내지 45 bp, 35 bp 내지 40 bp, 40 bp 내지 60 bp, 40 bp 내지 55 bp, 40 bp 내지 50 bp, 40 bp 내지 45 bp, 45 bp 내지 60 bp, 45 bp 내지 55 bp, 45 bp 내지 50 bp, 50 bp 내지 60 bp, 50 bp 내지 55 bp 또는 55 bp 내지 60 bp의 뉴클레오타이드의 인델을 형성하는 것일 수 있다.

[78] 일 구체예에 따르면, 상기 핵산을 변형하는 방법은 뉴클레오타이드 서열의 인델을 형성하는 효율이 5% 내지 50%, 예를 들면, 5% 내지 45%, 5% 내지 40%, 5% 내지 35%, 5% 내지 30%, 5% 내지 25%, 5% 내지 20%, 5% 내지 15%, 5% 내지 10%, 10% 내지 50%, 10% 내지 45%, 10% 내지 40%, 10% 내지 35%, 10% 내지 30%, 10% 내지 25%, 10% 내지 20%, 10% 내지 15%, 15% 내지 50%, 15% 내지 45%, 15% 내지 40%, 15% 내지 35%, 15% 내지 30%, 15% 내지 25%, 15% 내지 20%, 20% 내지 50%, 20% 내지 45%, 20% 내지 40%, 20% 내지 35%, 20% 내지 30%, 20% 내지 25%, 25% 내지 50%, 25% 내지 45%, 25% 내지 40%, 25% 내지 35%, 25% 내지 30%, 30% 내지 50%, 30% 내지 45%, 30% 내지 40%, 30% 내지 35%, 35% 내지 50%, 35% 내지 45%, 35% 내지 40%, 40% 내지 50%, 40% 내지 45%, 또는 45% 내지 50%일 수 있다.

[79] 일 구체예에 따르면, 상기 핵산을 변형하는 방법은 뉴클레오타이드 서열의 치환을 형성하는 효율이 1% 내지 20%, 예를 들면, 1% 내지 18%, 1% 내지 16%, 1% 내지 14%, 1% 내지 12%, 1% 내지 10%, 1% 내지 8%, 1% 내지 6%, 1% 내지 4%, 1% 내지 2%, 2% 내지 20%, 2% 내지 18%, 2% 내지 16%, 2% 내지 14%, 2% 내지 12%, 2% 내지 10%, 2% 내지 8%, 2% 내지 6%, 2% 내지 4%, 4% 내지 20%, 4% 내지 18%, 4% 내지 16%, 4% 내지 14%, 4% 내지 12%, 4% 내지 10%, 4% 내지 8%, 4% 내지 6%, 6% 내지 20%, 6% 내지 18%, 6% 내지 16%, 6% 내지 14%, 6% 내지 12%, 6% 내지 10%, 6% 내지 8%, 8% 내지 20%, 8% 내지 18%, 8% 내지 16%, 8% 내지 14%, 8% 내지 12%, 8% 내지 10%, 10% 내지 20%, 10% 내지 18%, 10% 내지 16%, 10% 내지 14%, 10% 내지 12%, 12% 내지 20%, 12% 내지 18%, 12% 내지 16%, 12% 내지 14%, 14% 내지 20%, 14% 내지 18%, 14% 내지 16%, 16% 내지 20%, 16% 내지 18%, 또는 18% 내지 20%일 수 있다.

[80] 상기 핵산을 변형하는 방법은 타겟 핵산(nucleic acid) 분자의 뉴클레오타이드(nucleotide) 서열 중 적어도 4 뉴클레오타이드의 교정 윈도우(editing window)을 형성할 수 있다. 일 구체예에 따르면, 상기 핵산을 변형하는 방법은 적어도 50 뉴클레오티드, 예를 들면, 적어도 49 뉴클레오티드, 적어도 48 뉴클레오티드, 적어도 47 뉴클레오티드, 적어도 46 뉴클레오티드, 적어도 45 뉴클레오티드, 적어도

44 뉴클레오티드, 적어도 43 뉴클레오티드, 적어도 42 뉴클레오티드, 적어도 41 뉴클레오티드, 적어도 40 뉴클레오티드, 적어도 39 뉴클레오티드, 적어도 38 뉴클레오티드, 적어도 37 뉴클레오티드, 적어도 36 뉴클레오티드, 적어도 35 뉴클레오티드, 적어도 34 뉴클레오티드, 적어도 33 뉴클레오티드, 적어도 32 뉴클레오티드, 적어도 31 뉴클레오티드, 적어도 30 뉴클레오티드, 적어도 29 뉴클레오티드, 적어도 28 뉴클레오티드, 적어도 27 뉴클레오티드, 적어도 26 뉴클레오티드, 적어도 25 뉴클레오티드, 적어도 24 뉴클레오티드, 적어도 23 뉴클레오티드, 적어도 22 뉴클레오티드, 적어도 21 뉴클레오티드, 적어도 20 뉴클레오티드, 적어도 19 뉴클레오티드, 적어도 18 뉴클레오티드, 적어도 17 뉴클레오티드, 적어도 16 뉴클레오티드, 적어도 15 뉴클레오티드, 적어도 14 뉴클레오티드, 적어도 13 뉴클레오티드, 적어도 12 뉴클레오티드, 적어도 11 뉴클레오티드, 적어도 10 뉴클레오티드, 적어도 9 뉴클레오티드, 적어도 8 뉴클레오티드, 적어도 7 뉴클레오티드, 적어도 6 뉴클레오티드 또는 적어도 5 뉴클레오티드의 교정 윈도우(editing window)를 갖는 것일 수 있다.

[81] 일 구체예에 따르면, 상기 핵산을 변형하는 방법은 gRNA 표적 염기서열의 5' 말단으로부터 1 bp 내지 20 bp, 1 bp 내지 19 bp, 1 bp 내지 18 bp, 1 bp 내지 17 bp, 1 bp 내지 16 bp, 1 bp 내지 15 bp, 1 bp 내지 14 bp, 1 bp 내지 13 bp, 1 bp 내지 12 bp, 1 bp 내지 11 bp, 1 bp 내지 10 bp, 1 bp 내지 9 bp, 1 bp 내지 8 bp, 2 bp 내지 20 bp, 2 bp 내지 19 bp, 2 bp 내지 18 bp, 2 bp 내지 17 bp, 2 bp 내지 16 bp, 2 bp 내지 15 bp, 2 bp 내지 14 bp, 2 bp 내지 13 bp, 2 bp 내지 12 bp, 2 bp 내지 11 bp, 2 bp 내지 10 bp, 2 bp 내지 9 bp, 2 bp 내지 8 bp, 3 bp 내지 20 bp, 3 bp 내지 19 bp, 3 bp 내지 18 bp, 3 bp 내지 17 bp, 3 bp 내지 16 bp, 3 bp 내지 15 bp, 3 bp 내지 14 bp, 3 bp 내지 13 bp, 3 bp 내지 12 bp, 3 bp 내지 11 bp, 3 bp 내지 10 bp, 3 bp 내지 9 bp, 3 bp 내지 8 bp, 4 bp 내지 20 bp, 4 bp 내지 19 bp, 4 bp 내지 18 bp, 4 bp 내지 17 bp, 4 bp 내지 16 bp, 4 bp 내지 15 bp, 4 bp 내지 14 bp, 4 bp 내지 13 bp, 4 bp 내지 12 bp, 4 bp 내지 11 bp, 4 bp 내지 10 bp, 4 bp 내지 9 bp 또는 4 bp 내지 8 bp의 교정 윈도우(editing window)를 갖는 것일 수 있다.

[82] 일 구체예에 따르면, 상기 핵산을 변형하는 방법의 교정 윈도우(Editing window)는 gRNA 표적 염기서열의 5' 말단으로부터 첫번째 위치한 시토신(C_1) 내지 20번째 위치한 시토신(C_{20}), C_2 내지 C_{20}, C_3 내지 C_{20}, C_4 내지 C_{20}, C_1 내지 C_{19}, C_2 내지 C_{19}, C_3 내지 C_{19}, C_4 내지 C_{19}, C_1 내지 C_{18}, C_2 내지 C_{18}, C_3 내지 C_{18}, C_4 내지 C_{18}, C_1 내지 C_{17}, C_2 내지 C_{17}, C_3 내지 C_{17}, C_4 내지 C_{17}, C_1 내지 C_{16}, C_2 내지 C_{16}, C_3 내지 C_{16}, C_4 내지 C_{16}, C_1 내지 C_{15}, C_2 내지 C_{15}, C_3 내지 C_{15}, C_4 내지 C_{15}, C_1 내지 C_{14}, C_2 내지 C_{14}, C_3 내지 C_{14}, C_4 내지 C_{14}, C_1 내지 C_{13}, C_2 내지 C_{13}, C_3 내지 C_{13}, C_4 내지 C_{13}, C_1 내지 C_{12}, C_2 내지 C_{12}, C_3 내지 C_{12}, C_4 내지 C_{12}, C_1 내지 C_{11}, C_2 내지 C_{11}, C_3 내지 C_{11}, C_4 내지 C_{11}, C_1 내지 C_{10}, C_2 내지 C_{10}, C_3 내지 C_{10}, C_4 내지 C_{10}, C_1 내지 C_9, C_2 내지 C_9

, C_3 내지 C_9, C_4 내지 C_9, C_1 내지 C_8, C_2 내지 C_8, C_3 내지 C_8 또는 C_4 내지 C_8,범위에서 염기 교정을 나타내는 것일 수 있다.

발명의 효과

[83] 일 양상에 따른 융합 단백질, 이의 폴리펩타이드, 및 이를 포함하는 CRISPR-Cas 시스템에 의하면, 효과적인 염기 교정을 할 수 있다. 또한, 폴리펩타이드의 크기가 작아 Cas 단백질과 결합이 용이하고, 세포치료제로 활용하는 경우 벡터를 통한 전달이 용이한 효과가 있다.

[84] 또한, 이를 통해 인델을 유도하는 경우, 인델 효율 및 인델이 형성되는 뉴클레오타이드의 크기도 증가하여 유전자의 넉아웃(knock-out)에 효과적으로 활용될 수 있다.

도면의 간단한 설명

[85] 도 1은 SsdA와 기존 염기교정기술에 사용되던 다른 탈아미노효소와 구조적/진화적으로 다름을 보여주는 모식도이다.

[86] 도 2는 서열번호 4의 염기서열을 갖는 단일가닥 DNA에 SsdA를 처리하여 탈아미노화 시키고, UDG와 NaOH를 처리하여 단일가닥 DNA를 절단 과정을 나타낸 그림 및 SsdA의 탈아미노화 여부를 확인한 웨스턴 블랏 결과를 나타낸 사진이다.

[87] 도 3는 표적 DNA의 타겟 사이트에서 Cas 단백질, SsdA 및 가이드 폴리뉴클레오티드를 포함하는 CRISPR-Cas 시스템의 염기 교정 효율을 나타낸 그래프이다:
도 3a는 RNF2 DNA의 타겟 사이트에서 Cas 단백질, SsdA 및 가이드 폴리뉴클레오티드를 포함하는 CRISPR-Cas 시스템의 염기 교정 효율을 나타낸 그래프이고,
도 3b는 HEK2 DNA의 타겟 사이트에서 Cas 단백질, SsdA 및 가이드 폴리뉴클레오티드를 포함하는 CRISPR-Cas 시스템의 염기 교정 효율을 나타낸 그래프이다.

[88] 도 4는 Cas9(D10A)-SsdA 융합 단백질, Cas9(D10A)-SsdA-UGI 융합 단백질, dCas9-SsdA 융합 단백질, 및 dCas9-SsdA-UGI 융합 단백질 각각의 세포 독성을 나타낸 그래프이다.

[89] 도 5는 Cas9 및 SsdA 융합 단백질을 포함하는 CRISPR-Cas 시스템의 염기 교정 효율 및 인델 형성 효율을 나타낸 그래프이다:

[90] 도 5a는 Cas9(D10A)-SsdA 융합 단백질, Cas9(D10A)-SsdA-UGI 융합 단백질, dCas9-SsdA 융합 단백질, 또는 dCas9-SsdA-UGI 융합 단백질을 포함하는 CRISPR-Cas 시스템 각각의 HEK2, HEK3, HEK4, 및 RNF2에서 염기 교정 효율 및 인델 형성 효율을 나타낸 그래프이고, 도 5b는 cjCas9(D8A)-SsdA 융합 단백질, cjCas9(D8A)-SsdA-UGI 융합 단백질, 또는 cjCas9(L58Y/D900K)(D8A)-SsdA-UGI 융합 단백질을 포함하는 CRISPR-Cas 시스템 각각의 EPAS1_e2, EPAS1_e5, HIF_e8, HIF_e9, 및 TFPi에서의 염기 교정 효율을 나타낸 그래프이며, 도 5c는 Uracil-DNA glycosylase (UDG) 발현 억제 세포주에서 cjCas9(D8A)-SsdA 융합 단백질, cjCas9(D8A)-SsdA-UGI 융합 단백질, 또는 cjCas9(L58Y/D900K)(D8A)-

SsdA-UGI 융합 단백질을 포함하는 CRISPR-Cas 시스템 각각의 EPAS1_e2, EPAS1_e5, HIF_e8, HIF_e9, 및 TFPi에서의 염기 교정 효율을 나타낸 그래프이다.

[91] 도 6는 SsdA와 Uracil glycosylase inhibitor (UGI)가 Cas 단백질의 C 말단, N 말단, 그리고 N, C 말단 양쪽에 결합된 융합 단백질을 제조하기 위한 플라스미드를 나타낸 모식도이다.

[92] 도 7은 SsdA와 Cas 단백질의 결합 위치에 따른 염기 교정 효율 확인한 그래프이다:

[93] 도 7a는 Cas9(D10A)-SsdA 융합 단백질(SsdA-C), SsdA-Cas9(D10A) 융합 단백질(SsdA-N, SsCBE), 또는 SsdA-Cas9(D10A)-SsdA(SsdA-NC) 융합 단백질을 포함하는 CRISPR-Cas 시스템 각각의 HEK2, HEK3, HEK4, RNF2, FANCF, TYRO3, CCR5, 또는 EMX1 에서 시토신 염기 교정 효율을 나타낸 그래프이고, 도 7b는 SsdA-Cas9(D10A) 융합단백질(SsdA-N, SsCBE)의 C 말단, N 말단, 그리고 C, N 말단 양쪽에 한 개 또는 두 개의 UGI를 결합한 융합단백질을 포함하는 CRISPR-Cas 시스템 각각의 HEK2, HEK3, HEK4, RNF2, FANCF, TYRO3, CCR5, 또는 EMX1 에서 시토신 염기 교정 효율을 나타낸 그래프이며, 도 7c는 염기 교정 효율이 현저히 높은 UGI를 N 말단에 두 개 결합한 UGI-UGI-SsdA-Cas9(D10A) 융합단백질(SsCBE-UGI-N2) 또는 UGI를 N, C 말단 양쪽에 두 개씩 결합한 UGI-UGI-SsdA-Cas9(D10A)-UGI-UGI 융합단백질(SsCBE-UGI-N2C2)에 대하여 정확함 염기 교정 효율 비교를 위해 표적 위치 20 종 (HEK2-1, HEK2-2, HEK2-3, HEK2-4, HEK3-1, HEK3-2, HEK3-3, HEK3-6, HEK3-7, HEK3-8, HEK4-1, HEK4-2, HEK4-3, HEK4-4, HEK4-5, HEK4-6, HEK4-7, HEK4-8, RNF2-3, RNF2-4)에서 시토신 염기 교정 효율을 나타낸 그래프이고, 도 7d는 도7c에서 보인 20 종의 표적 위치에 대한 염기 교정에 대하여 염기 교정 기술의 특징인 염기 교정 범위(Editing window)를 분석하여 나타낸 그래프이다.

[94] 도 8은 표적 위치(HEK2, HEK3, HEK2-2, HEK3-8, HEK4-2, 또는 HEK4-7 사이트)에서 UGI-UGI-SsdA-Cas9(D10A) 융합단백질(SsCBE-UGI-N2), BE3 및 BE4max의 치환 빈도(substitution frequency)(%)를 나타낸 그래프이다.

[95] 도 9는 Cas9(D10A)-SsdA(G54D)-UGI 융합 단백질, Cas9(D10A)-SsdA(G54D) 융합 단백질, dCas9-SsdA(G54D)-UGI 융합 단백질 또는 dCas9-SsdA(G54D) 융합 단백질을 포함하는 CRISPR-Cas 시스템 각각의 HEK2, HEK3, 및 HEK4에서 염기 교정 효율 및 인델 형성 효율을 나타낸 그래프이다.

발명의 실시를 위한 형태

[96] 이하 실시예를 통하여 보다 상세하게 설명한다. 그러나, 이들 실시예는 예시적으로 설명하기 위한 것으로 본 발명의 범위가 이들 실시예에 한정되는 것은 아니다.

[97]

[98] **실시예 1. 플라스미드 클로닝(plasmid cloning)**

[99] 서열번호 1의 아미노산 서열을 갖는 His6-SsdA 및 서열번호 3의 아미노산 서열을 갖는 SsdAI를 인코딩하는 gBlock 이중 가닥 DNA 단편 (Integrated DNA Technologies) 및 박테리아 발현 벡터 pET-28b(+) DNA(Novagen)을 XbaI 및 XhoI 제한 효소(New England Biolabs)로 각각 37 ℃에서 3시간 동안 처리하였다. 그 후, 선형화된 pET-28b 벡터를 아가로스 겔 추출(Geneall)을 사용하여 정제하고 퀵 리가제(New England Biolabs)를 사용하여 상기 gBlock 이중 가닥 DNA 단편과 연결하였다.

[100] 그리고, pCMV plamid DNA를 사용하여, PCR 증폭을 통해 Cas9와 UGI의 코딩 서열(coding sequence)을 얻고, Gibson Assembly Master Mix(New England Biolabs)를 사용하여 gBlock에서 SsdA 서열을 증폭하고 pCMV 플라스미드로 서브클론하였다.

[101] 서열번호 1의 PAAR-domin-containing protein 아미노산 서열 및 서열번호 3의 SsdAI 아미노산 서열을 표 2에 나타내었다.

[102] [표2]

	아미노산 서열	서열번호
SsdA protein(PAAR-domin-containing protein)	MSAAARVNDPIEHTGSLTGLLAGLAIGAIGAALVVGT GGLAAVAIVGASAATGAGVGQLIGSLSCCNHQTGQIV SGSSNVYINGEPAARAHADQAKCDEHTSRPQVIAQGS SNVYINGHPAARVGDRTACDAKIVVGSSNVFIGGGTE TTDPINPEVPELLERSILLVGLASAVVLASPVIVIAGLV GGIAGGTVGSMGGAQLFGEGTDGQKLMAFGGALLG GGLGAKGGKWFDTRYDIKVQGVGSNLGNLKITPKGA AKVSNIAESEAALGRASQARADLPQSKELKVKTVSSN DKKTLSGWGNKKPEGYERISAEQVKAKSEEIGHEVKS HPYDRDYKGQYFSSHAEKQMSIASPNHPLGVSKPMC TDCQGYFSQLAKYSKVEQTVADPKAIRIFKTDGSVETI MRSE	서열번호 1
SsdAI protein	MNNKSKVLIEKLLLEVAKSPEGELILPLRKLLWNTITE DETAAKKKAILTALDVMCVRQGVNFWIKKFGDNEPL NYILNIALETAEGKFDESKALGLRDEFYVSIVEDQEYE VEEYPAMFVGHAAANTIARAVDDFQFEPYDHRVDRD LDPEGFESSYLVASAFAGGLSEDGDPKLRRAFWEWY LSIAVPQVV	서열번호 3

[103] **실시예 2. SsdA의 정제(purification)**

[104] 실시예 1에서 클로닝한 SsdA 단백질을 대장균 BL21을 사용하여 정제하였다.

[105] 보다 구체적으로, pET-28b-His6-SsdA-SsdAI를 0.5mM IPTG를 사용하여 대장균 BL21에 투여하고, Ni-NTA 아가로스 비드(Qiagen)를 사용하여 His6-SsdA 및

SsdAI 복합 단백질을 정제하였다. 그리고 His6-SsdA를 SsdAI로부터 분리하기 위해, His6-SsdA 및 SsdAI 복합 단백질을 변성 완충액(8M urea, 50mM Tris-HCl pH 7.5, 500mM NaCl 및 1mM DTT)으로 변성시킨 후, 4 ℃에서 16 시간 동안 배양하였다. 변성된 단백질 복합체를 포함하는 현탁액 완충액을 Ni-NTA 아가로스 비드(Qiagen)와 혼합하고 중력 흐름 컬럼(gravity-flow column)에 로딩하여 결합되지 않은 SsdAI를 제거하였다. 그 후, 요소(6 M, 4 M, 2 M, 1 M 및 0 M)의 농도가 감소하는 변성 완충액을 SsdA의 리폴딩(refolding)을 위해 처리하였다. Ni-NTA 아가로스 비드에 결합된 리폴딩된 단백질을 300 mM 이미다졸을 함유하는 용출 완충액으로 용리시킨 후, 용출된 단백질을 20 mM Tris-HCl pH 7.5, 200 mM NaCl, 1 mM DTT, 및 40%(w/v) 글리세롤을 이용하여 투석(dialysis)하고, Amicon Ultra-15 Centrifugal Filter Unit(Millipore)을 사용하여 농축하였다. 그리고 His6-SsdA 단백질의 농도를 SDS-PAGE로 분석하였다.

[106]

[107] **실시예 3. 세포 배양 및 형질감염(transfection)**

[108] HEK293T 세포(ATCC CRL-11268)를 10% 소태아혈청(fetal bovine serum: FBS) 및 1% 페니실린/스트렙토마이신(Welgene)이 포함된 Dulbecco's modified Eagle's medium(DMEM)에서 보존한 뒤, 상기 HEK293T 세포를 웰 밀도당 6 x 10^4 세포로 TC 처리된 48웰 플레이트(Corning Life Sciences)에 시딩하였다. 시딩 후 24 시간이 경과한 뒤, 약 60% 세포 컨플루언시에서 500 ng의 플라스미드(250 ng의 Cas9-SsdA 발현 플라스미드 및 250 ng의 gRNA 발현 플라스미드) 및 1.5 uL의 Lipofectamine 2000(Thermo Fisher Scientific)을 사용하여 형질감염을 수행하였다. 그리고 형질감염된 세포를 37 ℃에서 3 일 동안 인큐베이션하고, 용해 완충액(pH 7.5의 10 mM Tris-HCl, 0.05% SDS, 100 mg/mL 프로테이나제 K; QIAGEN)을 사용하여 세포를 직접 용해하여 게놈 DNA를 제조하였다. 상기 세포 용해물을 56 ℃에서 30분 동안 인큐베이션하고, 99 ℃에서 15 분 동안 추가로 인큐베이션하여 Proteinase K를 비활성시켰다.

[109]

[110] **실시예 4. 타겟 딥 시퀀싱(deep sequencing) 및 데이터 분석**

[111] 표적 부위를 PCR을 수행하여 증폭하였고(총 2 또는 3 회), Illumina MiniSeq 또는 iSeq 100 시퀀싱 시스템을 통해 시퀀싱을 수행하였다.

[112] 보다 구체적으로, 3 mL의 세포 용해물 또는 1 mL의 분리된 게놈 DNA를 1 차 PCR에 적용한 후, 1차 PCR 생성물 1 mL를 2차 PCR에 사용하였다. Illumina TruSeq HT 이중 인덱스 어댑터 시퀀스를 1 mL의 2차 PCR 산물을 사용하여 인덱스 PCR 프라이머 쌍으로 부착하였다. PCR 앰플리콘의 크기를 2% 아가로스 겔에서 확인하였으며 앰플리콘은 Illumina MiniSeq 또는 iSeq 100 시퀀싱 시스템을 사용하였다. 표적 심층 시퀀싱 분석은 MAUND(https:/github.com/ibscge/maund)를 사용하여 수행하였으며 모든 결과는 Cas-Analyzer(http:/www.rgenome.net/cas-analyzer/)로 확인하였다.

[113]

[114] **실시예 5. SsdA의 단일 가닥 DNA에서의 시토신의 탈아미노화 확인**

[115] SsdA가 단일 가닥 DNA의 시토신(Cytosine) 염기를 우라실(Uracil)로 전환하는지 확인하기 위해, FAM을 포함하는 단일 가닥 DNA에 SsdA를 처리하였다.

[116] 보다 구체적으로, 서열번호 4의 염기 서열(5'- Aaaaaaaaaaaaaaagcgaaaaaaaaaaaaaaaaa-3')을 갖는 단일 가닥 DNA에 1 내지 200 nM의 SsdA를 37 ℃에서 1 시간 처리한 뒤, 우라실 글리코실레이즈(Uracil DNA Glycosylase: UDG)를 37 ℃ 조건에서 30 분 동안 처리하여 우라실로 변화된 DNA 염기를 제거하여 abasic site를 만들었다. 그 후, 100 mM NaOH를 처리하고, 95 ℃ 조건에서 2 분 동안 인큐베이션하여 abasic site를 절단하였다. 그리고 그 절단 여부를 웨스턴 블랏을 통해 확인하고, 그 결과를 도 2에 나타내었다.

[117] 도 2는 서열번호 4의 염기서열을 갖는 단일가닥 DNA에 SsdA를 처리하여 탈아미노화 시키고, UDG 와 NaOH를 처리하여 단일가닥 DNA를 절단 과정을 나타낸 그림 및 SsdA의 탈아미노화 여부를 확인한 웨스턴 블랏 결과를 나타낸 사진이다.

[118] 도 2에 나타낸 바와 같이, 단일가닥 DNA에 SsdA, UDG 및 NaOH를 처리하였을 때만 DNA 절단이 발생하는 것을 확인하였다. 이는 SsdA가 단일 가닥 DNA의 시토신을 탈아미노화시켜 우라실로 효과적으로 전환시켜줄 수 있음을 의미한다.

[119]

[120] **실시예 6. Cas 단백질, SsdA 및 가이드 폴리뉴클레오티드를 포함하는 CRISPR-Cas 시스템의 염기 교정 효율 확인**

[121] Cas 단백질, SsdA 및 가이드 폴리뉴클레오티드를 포함하는 CRISPR-Cas 시스템의 염기 교정 효율을 확인하기 위해, 표적 DNA(RNF2 및 HEK2)에 SsdA, dCas9 및 gRNA를 포함하는 CRISPR-Cas 시스템을 처리하여 시토신의 탈아미노화의 형성 여부를 확인하였다. 상기 표적 DNA(RNF2 및 HEK2)의 염기서열은 표 3에 나타내었다.

[122] [표3]

표적 DNA	타겟 사이트(target site)	서열 번호
RNF2	GTC3ATC6TTAGTC12ATTACCTGAGG	서열번호 5
HEK2	GAAC4AC6AAAGC11ATAGACTGCGGG	서열번호 6

[123] 보다 구체적으로, RNF2 DNA 및 HEK2 DNA 각각에 100 nM Cas9, 300 nM sgRNA, 및 40 nM SsdA를 처리한 후 37 ℃에서 8 시간 동안 인큐베이션하고, 표적 DNA의 타겟 위치에서 시토신의 우라실로의 전환을 유도하였다. 그리고, 8 시간 후 RNase와 Protease K를 처리하여 sgRNA, Cas9, 및 SsdA를 제거하였다. 그리고 qiagen DNA extraction kit을 이용하여 DNA를 정제(DNA purification)하였다. 그 후 타겟 사이트를 포함하는 프라이머(primer)를 사용하여 PCR을 진행한 후 Deep sequencing을 통하여 염기 교정 효율을 측정하고 그 결과를 도 3에 나타내었다.

[124] 도 3은 표적 DNA의 타겟 사이트에서 Cas 단백질, SsdA 및 가이드 폴리뉴클레오티드를 포함하는 CRISPR-Cas 시스템의 염기 교정 효율을 나타낸 그래프이다:

[125] 도 3a는 RNF2 DNA의 타겟 사이트에서 Cas 단백질, SsdA 및 가이드 폴리뉴클레오티드를 포함하는 CRISPR-Cas 시스템의 염기 교정 효율을 나타낸 그래프이고, 도 3b는 HEK2 DNA의 타겟 사이트에서 Cas 단백질, SsdA 및 가이드 폴리뉴클레오티드를 포함하는 CRISPR-Cas 시스템의 염기 교정 효율을 나타낸 그래프이다.

[126] 도 3a에 나타낸 바와 같이, RNF2를 타겟하는 gRNA를 사용하였을 경우, 시토신의 우라실로의 전환 효율이 C_3에서는 약 7%, C_{12}에서는 약 11%로 나타남을 확인하였다. 또한, 도 3b에 나타낸 바와 같이, HEK2를 타겟하는 gRNA를 사용하였을 경우에는 C_4, C_6, 및 C_{11}에서 시토신의 우라실로의 전환 효율이 약 18 내지 19% 로 나타남을 확인하였다.

[127] 이는, 상기 CRISPR-Cas 시스템이 DNA 이중 가닥 중에 한쪽 가닥에서만 유전자 편집을 발생시킬 수 있고, 이에 따라 최대 염기 교정 효율이 50%임을 고려하였을 때, 이와 같은 결과는 상당히 높은 수준의 염기 교정 효율을 나타냄을 의미한다.

[128] 또한, 이에 추가로, 상기 Cas 단백질, SsdA 및 가이드 폴리뉴클레오티드를 포함하는 CRISPR-Cas 시스템에 의해 발생한 염기 전환이 시토신에서 우라실로의 전환이 맞는지 확인하기 위해, 탈아미노화된 DNA에 우라실 특이적 제거 효소인 Uracil-Specific Excision Reagent(USER) 효소를 처리하고, PCR을 통해 타겟 사이트를 증폭하였다. 그 결과, 시토신이 우라실로 전환된 리드(read)의 수가 없어지는 것을 확인하였다.

[129] 이와 같은 결과는, 상기 CRISPR-Cas 시스템이 시토신을 우라실로 효과적으로 전환시키는 것을 의미한다.

[130]

[131] **실시예 7. Cas 단백질 및 SsdA를 포함하는 융합 단백질의 세포 독성 확인**

[132] Cas 단백질, SsdA 및 UGI를 포함하는 융합 단백질의 세포 독성을 확인하고 그 결과를 도 4에 나타내었다.

[133] CRISPR-Cas 시스템의 SsdA가 진핵세포 내에서 독성이 있는지 확인하기 위해서 Cas9(D10A)와 결합한 단백질을 발현하는 플라스미드를 제작하였다. 그리고, HEK293 세포를 48 웰 플레이트(well plate)에 6 x 10^4 cell/well 농도로 분주한 후 상기 플라스미드를 HEK293 세포에 형질감염(transfection)하였다. 형질주입하고 48시간 후 살아있는 세포를 트립신화(trypsinization)하여 헤모사이토미터(hemacytometer)를 이용하여 HEK293 세포의 수를 카운팅하였다.

[134] 도 4는 Cas9(D10A)-SsdA 융합 단백질, Cas9(D10A)-SsdA-UGI 융합 단백질, dCas9-SsdA 융합 단백질, 및 dCas9-SsdA-UGI 융합 단백질 각각의 세포 독성을 나타낸 그래프이다.

[135] 도 4에 나타낸 바와 같이, Cas9, SsdA 및 UGI를 포함하는 융합 단백질의 경우, 세포 독성이 거의 없음을 확인하였다.

[136]

[137] **실시예 8. Cas 단백질 및 SsdA를 포함하는 융합 단백질을 포함하는 CRISPR-Cas 시스템의 염기 교정 효율 및 인델 형성 효율 확인**

[138] CRISPR-Cas 시스템의 SsdA가 진핵세포 내에서 시토신 탈아미노화를 효과적으로 일으키는지 확인하기 위해서 Cas9과 결합한 단백질을 발현하는 플라스미드를 제작하여 HEK293 세포에 형질감염(transfection)하였다. 그리고, 형질 주입된 세포로부터 표적 위치(HEK2, HEK3, HEK4, RNF2, EPAS1_e2, EPAS1_e5, HIF_e8, HIF_e9, 또는 TFPi 사이트)의 염기서열을 PCR로 증폭하고 차세대염기서열분석법(NGS)를 이용하여 변이의 도입 여부 및 인델 형성 여부를 분석하고, 그 결과를 도 5에 나타내었다.

[139] Cas9은 spCas9(Cas9(D10A)) 또는 cjCas9(D8A)을 사용하였다. Uracil-DNA glycosylase(UDG) 발현 억제 세포주(UNG KD)는 UDG를 넉다운(knockdown)하는 shRNA(5'-GTCTACAGACATAGAGGATTT -3: 서열번호 7) 발현 플라스미드 및 HIV 기반 패키징 플라스미드(유전자: Gag/Pol, Rev, VSV-G 포함)를 Opti MEM(Invitrogen)에서 Lipofectamine 3000(Invitrogen) 시약과 혼합하여 HEK293 세포를 형질감염시킴으로써 제조하였다.

[140] 도 5는 Cas9 및 SsdA 융합 단백질을 포함하는 CRISPR-Cas 시스템의 염기 교정 효율 및 인델 형성 효율을 나타낸 그래프이다:

[141] 도 5a는 Cas9(D10A)-SsdA 융합 단백질, Cas9(D10A)-SsdA-UGI 융합 단백질, dCas9-SsdA 융합 단백질, 또는 dCas9-SsdA-UGI 융합 단백질을 포함하는 CRISPR-Cas 시스템 각각의 HEK2, HEK3, HEK4, 및 RNF2에서의 염기 교정 효율 및 인델 형성 효율을 나타낸 그래프이고, 도 5b는 cjCas9(D8A)-SsdA 융합 단백질, cjCas9(D8A)-SsdA-UGI 융합 단백질, 또는 cjCas9(L58Y/D900K)(D8A)-SsdA-UGI 융합 단백질을 포함하는 CRISPR-Cas 시스템 각각의 EPAS1_e2, EPAS1_e5, HIF_e8, HIF_e9, 및 TFPi에서의 염기 교정 효율을 나타낸 그래프이며, 도 5c는 Uracil-DNA glycosylase (UDG) 발현 억제 세포주에서 cjCas9(D8A)-SsdA 융합 단백질, cjCas9(D8A)-SsdA-UGI 융합 단백질, 또는 cjCas9(L58Y/D900K)(D8A)-SsdA-UGI 융합 단백질을 포함하는 CRISPR-Cas 시스템 각각의 EPAS1_e2, EPAS1_e5, HIF_e8, HIF_e9, 및 TFPi에서의 염기 교정 효율을 나타낸 그래프이다.

[142] 도 5a에 나타낸 바와 같이, Cas9(D10A)-SsdA-UGI의 경우 HEK2, HEK3, HEK4, and RNF2 사이트 모두에서 4 내지 11%의 시토신 염기교정을 일으키는 것을 확인하였으며, 표적 위치에서 염기서열의 인델이 형성되는 것을 확인하였다. 또한, dCas9-SsdA-UGI의 경우 1 내지 3.5% 수준으로 시토신 염기교정이 일어나는 것을 확인하였으며, 인델은 형성되지 않음을 확인하였다.

[143] 도 5b에 나타낸 바와 같이, cjCas9(D8A)-SsdA-UGI, 및 cjCas9(L58Y/D900K)(D8A)-SsdA-UGI의 경우 모두 약 3%의 염기 교정을 일으키는 것을 확인하여, UGI가 결합하지 않은 융합 단백질 대비 향상된 염기교정 효율을 보임을 확인하였다.

[144] 도 5c에 나타낸 바와 같이, UDG 발현 억제 세포주에서는 cjCas9(D8A)-SsdA-UGI 및 cjCas9(L58Y/D900K)(D8A)-SsdA-UGI의 경우 모두 약 15 % 정도의 염기 교정 효율을 보여, UGI가 결합하지 않은 cjCas9(D8A)-SsdA 융합 단백질 대비 현저히 향상된 염기교정 효율을 보임을 확인하였다.

[145] 표 4는 Cas9(D10A)-Ssda-UGI와 HEK2 표적 gRNA에 의해 발생한 염기교정된 서열을 분석한 결과를나타낸 것이다.

[146] [표4]

서열	read	서열번호
GGAAACTGGAACACAAAGC ATAGACTGCGGGGCGGGCCA	3,912 (WT)	서열번호 8
GGAAACTGGAACATAAAGC ATAGACTGCGGGGCGGGCCA	237	서열번호 9
GGAAACTGGAATATAAAGC ATAGACTGCGGGGCGGGCCA	103	서열번호 10
GGAAACTGGAATACAAAGC ATAGACTGCGGGGCGGGCCA	61	서열번호 11
GGAAACTGGAACACAAAGT ATAGACTGCGGGGCGGGCCA	29	서열번호 12

[147] **실시예 9. SsdA와 Cas 단백질의 결합 위치에 따른 염기 교정 효율 확인**

[148] 세포 내 DNA에서 시토신 탈아미노화를 일으키면 이를 수선하는 단백질로 Uracil-DNA glycosylase(UDG)가 잘 알려져 있다. SsdA가 실제로 세포 내 DNA에서 시토신 탈아미노화를 일으키는 것을 정확히 확인하기 위해 먼저 UDG를 녹아웃한 HEK293 세포주(HEK293 UDG-KO)를 제작하였다.

[149] SsdA가 Cas 단백질의 C 말단, N 말단, 그리고 N, C 말단 양쪽에 결합된 융합 단백질의 염기 교정 효율 및 인델 형성 효율을 확인하기 위해, SsdA의 위치를 달리하여 융합 단백질을 제조하고, 이를 통해 염기 교정 효율 및 인델 형성 효율을 확인하였다.

[150] 보다 구체적으로, 도 6과 같은 pCMV 플라스미드를 제작하고, 이를 실시예 3에 따라 각 표적 위치에 대한 gRNA와 함께 HEK293 UDG-KO 세포로 도입하였다. 그리고, 형질 주입된 세포로부터 표적 위치(HEK2, HEK3, HEK4, RNF2, FANCF, TYRO3, CCR5, 또는 EMX1 사이트)의 염기서열을 PCR로 증폭하고 차세대염기서열분석법 (NGS)를 이용하여 변이의 도입여부 및 인델 형성 여부를 분석하고, 그 결과를 도 7에 나타내었다

[151] 도 6은 SsdA와 Uracil glycosylase inhibitor (UGI)가 Cas 단백질의 C 말단, N 말단, 그리고 N, C 말단 양쪽에 결합된 융합 단백질을 제조하기 위한 플라스미드를 나타낸 모식도이다.

[152] 도 7은 SsdA와 Cas 단백질의 결합 위치에 따른 염기 교정 효율 확인한 그래프이다:

[153] 도 7a는 Cas9(D10A)-SsdA 융합 단백질(SsdA-C), SsdA-Cas9(D10A) 융합 단백질(SsdA-N, SsCBE), 또는 SsdA-Cas9(D10A)-SsdA(SsdA-NC) 융합 단백질을 포함하는 CRISPR-Cas 시스템 각각의 HEK2, HEK3, HEK4, RNF2, FANCF, TYRO3, CCR5, 또는 EMX1 에서 시토신 염기 교정 효율을 나타낸 그래프이고,도 7b는 SsdA-Cas9(D10A) 융합단백질(SsdA-N, SsCBE)의 C 말단, N 말단, 그리고 C, N 말단 양쪽에 한 개 또는 두 개의 UGI를 결합한 융합단백질을 포함하는 CRISPR-Cas 시스템 각각의 HEK2, HEK3, HEK4, RNF2, FANCF, TYRO3, CCR5, 또는 EMX1 에서 시토신 염기 교정 효율을 나타낸 그래프이며, 도 7c는 염기 교정 효율이 현저히 높은 UGI를 N 말단에 두 개 결합한 UGI-UGI- SsdA-Cas9(D10A) 융합단백질(SsCBE-UGI-N2); 또는 UGI를 N, C 말단 양쪽에 두 개씩 결합한 UGI-UGI-SsdA-Cas9(D10A)-UGI-UGI 융합단백질(SsCBE-UGI-N2C2);에 대하여 정확함 염기 교정 효율 비교를 위해 표적 위치 20 종 (HEK2-1, HEK2-2, HEK2-3, HEK2-4, HEK3-1, HEK3-2, HEK3-3, HEK3-6, HEK3-7, HEK3-8, HEK4-1, HEK4-2, HEK4-3, HEK4-4, HEK4-5, HEK4-6, HEK4-7, HEK4-8, RNF2-3, RNF2-4)에서 시토신 염기 교정 효율을 나타낸 그래프이고, 도 7d는 도 7c에서 보인 20 종의 표적 위치에 대한 염기 교정에 대하여 염기 교정 기술의 특징인 염기 교정 범위(Editing window)를 분석하여 나타낸 그래프이다.

[154] 상기 표적 유전자의 타겟 사이트의 염기서열은 표 5에 나타내었다.

[155] [표5]

표적 DNA	타겟 사이트 (taregt site)	서열번호
HEK2	GAACACAAAGCATAGACTGC	18
HEK2-1	CCAGCCCGCTGGCCCTGTAA	19
HEK2-2	GCTGGCCCTGTAAAGGAAAC	20
HEK2-3	GTTTCCTTTACAGGGCCAGC	21
HEK2-4	GCACTTGTTTGCAGCTATTC	22
HEK3	GGCCCAGACTGAGCACGTGA	23
HEK3-1	CTGCTTCTCCAGCCCTGGCC	24
HEK3-2	CCCTGGCCTGGGTCAATCCT	25
HEK3-3	GACTGAGCACGTGATGGCAG	26
HEK3-6	CTTCCTCCAGAGGGCGTCGC	27
HEK3-7	CAGGACAGCTTTTCCTAGAC	28
HEK3-8	CAGCTCCTGCACCGGGATAC	29

HEK4	GGCACTGCGGCTGGAGGTGG	30
HEK4-1	GGGGCACCGCGGCGCCCCGG	31
HEK4-2	GCGGCGCCCCGGTGGCACTG	32
HEK4-3	CGCCCCGGTGGCACTGCGGC	33
HEK4-4	TCCCTTCCTTCCACCCAGCC	34
HEK4-5	CCCTGCCTGTCATCCTGCTT	35
HEK4-6	GCAGTGCCACCGGGGCGCCG	36
HEK4-7	CTCCAGCCGCAGTGCCACCG	37
HEK4-8	ACCTCCAGCCGCAGTGCCAC	38
RNF2	GTCATCTTAGTCATTACCTG	39
RNF2-3	TACACGTCTCATATGCCCCT	40
RNF2-4	TCAACCATTAAGCAAAACAT	41
EMX1	GTCACCTCCAATGACTAGGG	42
FANCF	GGAATCCCTTCTGCAGCACC	43
TYRO3	GGCCACACTAGCGTTGCTGC	44
CCR5	TGACATCAATTATTATACAT	45

[156] 도 7a에 나타낸 바와 같이, Cas의 C 말단에 SsdA를 결합한 SsdA-Cas9(D10A)(SsdA-C);의 경우 시토신 염기교정 효율은 N 말단 또는 N, C 말단 양쪽에 결합한 경우 대비 다소 감소하는 것을 확인하였다. 반면, Cas의 N 말단에 SsdA를 결합하는 경우(SsdA-N, SsCBE), 염기 교정 효율이 상승하는 것을 확인하였다.

[157] 도 7b에 나타난 바와 같이 염기 교정 효율이 높은 SsdA-Cas9(D10A)(SsdA-N, SsCBE);에 UGI를 추가로 결합한 경우 UGI가 없을 때보다 염기 교정 효율이 상승하였으며, UGI를 N 말단에 두 개 결합한 UGI-UGI-SsdA-Cas9(D10A) 융합단백질(SsCBE-UGI-N2); 또는 UGI를 N, C 말단 양쪽에 두 개씩 결합한 UGI-UGI-SsdA-Cas9(D10A)-UGI-UGI 융합단백질(SsCBE-UGI-N2C2);의 경우 염기교정 효율이 현저히 상승하는 것을 확인하였다.

[158] 도 7c에 나타난 바와 같이, 염기 교정 효율이 가장 높은 UGI를 N 말단에 두 개 결합한 UGI-UGI-SsdA-Cas9(D10A) 융합단백질(SsCBE-UGI-N2); 또는 UGI를 N, C 말단 양쪽에 두 개씩 결합한 UGI-UGI-SsdA-Cas9(D10A)-UGI-UGI 융합단백질(SsCBE-UGI-N2C2);에 대하여 20 종의 표적 위치에서 비교한 결과, UGI를 N 말단에 두 개 결합한 UGI-UGI-SsdA-Cas9(D10A) 융합단백질(SsCBE-UGI-N2);이 조금 더 상승된 염기 교정 효율을 나타내는 것을 확인하였다.

[159] 도 7d에 나타난 바와 같이, 염기 교정 효율이 가장 높은 UGI를 N 말단에 두 개 결합한 UGI-UGI-SsdA-Cas9(D10A) 융합단백질(SsCBE-UGI-N2); 또는 UGI를 N,

C 말단 양쪽에 두 개씩 결합한 UGI-UGI-SsdA-Cas9(D10A)-UGI-UGI 융합단백질 (SsCBE-UGI-N2C2);에 대하여 20 종의 표적 위치에서 염기 교정 범위 (Editing window)를 비교한 결과, gRNA 표적 염기서열의 5' 말단으로부터 4bp 내지 8 bp 사이의 위치에서 높은 염기 교정 효율을 나타내는 것을 확인하였다.

[160]

[161] **실시예 10. 야생형 HEK293 세포에서 UGI-UGI-SsdA-Cas9(D10A) 융합단백질 (SsCBE-UGI-N2)를 이용한 염기 교정능 확인**

[162] 염기 교정능이 가장 높은 것으로 확인된 UGI-UGI-SsdA-Cas9(D10A) 융합단백질(SsCBE-UGI-N2);에 따른 야생형 HEK293 세포에서도 염기 교정 효율이 나타나는지 확인하였다.

[163] 보다 구체적으로, UGI-UGI-SsdA-Cas9(D10A) 융합단백질(SsCBE-UGI-N2);과 기존의 BE3 또는 BE4max를 실시예 3에 따라 각 표적 위치에 대한 gRNA와 함께 HEK293 세포로 도입하였다. 그리고 형질 주입된 세포로부터 표적 위치(HEK2, HEK3, HEK2-2, HEK3-8, HEK4-2, 또는 HEK4-7 사이트)의 염기서열을 PCR로 증폭하고 차세대염기서열분석법 (NGS)를 이용하여 변이의 도입여부 및 인델 형성 여부를 분석하고, 그 결과를 도 8에 나타내었다.

[164] 도 8은 표적 위치(HEK2, HEK3, HEK2-2, HEK3-8, HEK4-2, 또는 HEK4-7 사이트)에서 UGI-UGI-SsdA-Cas9(D10A) 융합단백질(SsCBE-UGI-N2), BE3 및 BE4max의 치환 빈도(substitution frequency)(%)를 나타낸 그래프이다.

[165] 도 8에 나타난 바와 같이 UGI-UGI-SsdA-Cas9(D10A) 융합단백질(SsCBE-UGI-N2);은 6 종의 표적 위치(HEK2, HEK3, HEK2-2, HEK3-8, HEK4-2, 또는 HEK4-7 사이트)에서 성공적으로 염기 교정 효능을 나타내었다.

[166]

[167] **실시예 11. SsdA 변이체를 이용한 CRISPR-Cas 시스템의 세포 독성 및 염기 교정능 확인**

[168] SsdA의 변이체에 따른 세포 독성 및 염기 교정능을 확인하기 위해, 불활성화된 SsdA를 제작하고, 이에 따른 효과를 확인하였다.

[169] 보다 구체적으로, SsdA의 촉매활성부위에 G54D 변이를 도입한 뒤, 상기 변이가 발생한 SsdA 단백질을 Cas9에 결합하고 세포 내 기능을 확인하였다. 그리고 그 결과를 도 9에 나타내었다. SsdA의 촉매활성부위의 G54D 변이는 서열번호 1의 아미노산 서열(PAAR domaomn-containing protein)에서 G302D 변이이다. 구체적으로, 상기 SsdA의 촉매활성부위의 G54D 변이는 서열번호 17의 서열을 가진 것 일 수 있다. 상기 SsdA의 촉매활성부위의 G54D 변이의 아미노산 서열은 표 6에 나타내었다.

[170] [표6]

	아미노산 서열	서열 번호

SsdA의 촉매활성부위의 G54D 변이 서열	KVSNIAESEAALGRASQARADLPQSKELKV KTVSSNDKKTLS**D**WGNKKPEGYERISAEQV KAKSEEIGHEVKSHPYDRDYKGQYFSSHAE KQMSIASPNHPLGVSKPMCTDCQGYFSQLA KYSKVEQTVADPKAIRIFKTDGSVETIMRSE	17

[171] 도 9는 Cas9(D10A)-SsdA(G54D)-UGI 융합 단백질, Cas9(D10A)-SsdA(G54D) 융합 단백질, dCas9-SsdA(G54D)-UGI 융합 단백질 또는 dCas9-SsdA(G54D) 융합 단백질을 포함하는 CRISPR-Cas 시스템 각각의 HEK2, HEK3, 및 HEK4에서 염기 교정 효율 및 인델 형성 효율을 나타낸 그래프이다.

[172] 도 9에 나타낸 바와 같이, SsdA의 G54D 변이는 Cas9(D10A)와 결합하는 경우 UGI 유무와 상관없이 인델이 약 18 내지 34%의 매우 높은 효율로 나타나는 것을 확인하였다. 이는 SsdA(G54D)는 UGI 기능에 상관없이 표적 위치에서 인델을 더 높은 효율로 일으킬 수 있음을 의미한다.

[173] 또한, SsdA(G54D)를 사용하는 경우, 세포 내 독성이 현저히 감소하였을 뿐만 아니라, 결실 되는 염기서열의 크기가 현저히 증가되는 것을 확인하였다.

[174] 표 7은 SsdA(G54D)-Cas9(D10A)에 의해 발생한 염기 교정된 서열을 분석한 결과를 나타낸 것이다.

[175] 표 7에 나타낸 바와 같이, SsdA(G54D)-Cas9(D10A)에 의해 발생한 indel 서열을 분석한 결과, 통상적으로 1bp의 결실을 형성하는 Cas9과는 달리, SsdA(G54D)-Cas9(D10A)의 경우 결실 되는 염기서열의 크기가 현저히 증가되는 것을 확인하였다. 이러한 결과는 SsdA(G54D)를 포함하는 CRISPR-Cas 시스템은 Cas9만을 이용한 시스템보다 유전자 녹아웃 효율이 증가할 수 있음을 의미한다.

[176] [표7]

서열	read	서열번호
AATTTTCCAGCCCGCTGGCCCTGTAAAGGAAACTGGAACACAAAGCATAGACTGCGGGGCGGGCCAGCCTGAATA	1,010 (WT)	서열번호 13
AATTTTCCAGCCTGAATA	503	서열번호 14
AATTTTCCAGCCCGCTGGCCCTGTAAAGGAAACTGGAACACAAAGCGGGGCGGGCCAGCCTGAATA	354	서열번호 15
AATTTTCCAGCCCGCTGGCCCTGTAAAGGAAACTGGAACACAAAGCATA	349	서열번호 16

청구범위

[청구항 1] Cas 단백질(CRISPR-associated protein) 및 박테리아 톡신(toxin)을 포함하고, 상기 박테리아 톡신은 SsdA(single-stranded DNA deaminase toxin A)인 것인, 융합 단백질.

[청구항 2] 청구항 1에 있어서,
상기 Cas 단백질은 Cas9 단백질 또는 Cas12 단백질인 것인, 융합 단백질.

[청구항 3] 청구항 1에 있어서,
상기 SsdA는 시티딘 디아미나제(cytidine deaminase)인 것인, 융합 단백질.

[청구항 4] 청구항 1에 있어서,
상기 SsdA는 불활성화된 SsdA인 것인, 융합 단백질.

[청구항 5] 청구항 1에 있어서,
상기 SsdA는 서열번호 1의 아미노산 서열을 포함하는 것인, 융합 단백질.

[청구항 6] 청구항 1에 있어서,
상기 SsdA는 서열번호 1의 아미노산 서열과 적어도 85%이상의 서열 동일성(sequence identity)를 갖는 서열을 포함하는 것인, 융합 단백질.

[청구항 7] 청구항 4에 있어서,
상기 불활성화된 SsdA는 SsdA의 촉매활성부위(catalytic active site)에서 아미노산 변이가 발생한 것인, 융합 단백질.

[청구항 8] 청구항 4에 있어서,
상기 불활성화된 SsdA는 서열번호 1의 아미노산 서열에서 G302D, E349A, 또는 이들의 상응하는 아미노산 변이를 갖는 것인, 융합 단백질.

[청구항 9] 청구항 4에 있어서,
상기 불활성화된 SsdA는 서열번호 17의 아미노산 서열을 갖는 것인, 융합 단백질.

[청구항 10] 청구항 4에 있어서,
상기 불활성화된 SsdA는 서열번호 17의 아미노산 서열과 적어도 85%이상의 서열 동일성(sequence identity)를 갖는 서열을 포함하는 것인, 융합 단백질.

[청구항 11] 청구항 1에 있어서,
상기 박테리아 톡신은 Cas 단백질의 C-말단, N-말단, 또는 C-말단과 N-말단 모두에 결합된 것인, 융합 단백질.

[청구항 12] 청구항 1에 있어서,
상기 융합 단백질은 DNA 글리코실레이즈 억제제(DNA glycosylase inhibitor)를 더 포함하는 것인, 융합 단백질.

[청구항 13] 청구항 12에 있어서,
상기 DNA 글리코실레이즈 억제제는 티민 글리코실레이즈 억제제(thymine glycosylase inhibitor), 우라실 글리코실레이즈 억제제(uracil

glycosylase inhibitor), 옥소구아닌 글리코실레이즈 억제제(oxoguanine glycosylase inhibitor), 또는 알킬구아닌 DNA 글리코실레이즈 억제제 (alkylguanine glycosylase inhibitor)인 것인, 융합 단백질.

[청구항 14] 청구항 1에 있어서,
상기 융합 단백질은 타겟 서열의 5'말단으로부터 1bp 내지 20bp 범위의 교정 윈도우 (editing window)를 갖는 것인, 융합 단백질.

[청구항 15] 청구항 1 내지 14 중 어느 한 항의 융합 단백질을 암호화하는 폴리뉴클레오티드(polynucleotide).

[청구항 16] 청구항 15의 폴리뉴클레오티드를 포함하는 벡터.

[청구항 17] Cas 단백질(CRISPR-associated protein) 및 박테리아 톡신(toxin)을 포함하는 융합 단백질 또는 상기 융합 단백질을 암호화하는 폴리뉴클레오티드; 및 가이드 폴리뉴클레오티드를 포함하고, 상기 박테리아 톡신은 SsdA(single-stranded DNA deaminase toxin A)인 것인, CRISPR-Cas 시스템.

[청구항 18] 청구항 17에 있어서,
상기 가이드 폴리뉴클레오티드는 크리스퍼 RNA(CRISPR RNA: crRNA) 및 트랜스-활성화 RNA(trans-activating RNA: tracrRNA)를 포함하고, 상기 가이드 폴리뉴클레오티드는 이중 가이드 RNA(dual guide RNA) 또는 단일-사슬 가이드 RNA(single-chain guide RNA: sgRNA)인 것인, CRISPR-Cas 시스템.

[청구항 19] 청구항 17에 있어서,
상기 시스템은 타겟 핵산(nucleic acid) 분자의 뉴클레오타이드(nucleotide) 서열 중 적어도 하나 이상의 뉴클레오타이드의 삭제(deletion), 삽입(insertion), 치환(substitution) 또는 인델(insertion and deletion; indel)을 형성하는 것인, CRISPR-Cas 시스템.

[청구항 20] 청구항 19에 있어서,
상기 시스템은 타겟 핵산 분자의 뉴클레오타이드 서열에서 1 bp 내지 60 bp의 뉴클레오타이드의 결실, 1 bp 내지 60 bp의 뉴클레오타이드의 삽입, 또는 1 bp 내지 60 bp의 뉴클레오타이드의 인델을 형성하는 것인, CRISPR-Cas 시스템.

[청구항 21] 청구항 17에 있어서,
상기 시스템은 타겟 서열의 5'말단으로부터 1bp 내지 20bp 범위의 교정 윈도우 (editing window)를 갖는 것인, CRISPR-Cas 시스템.

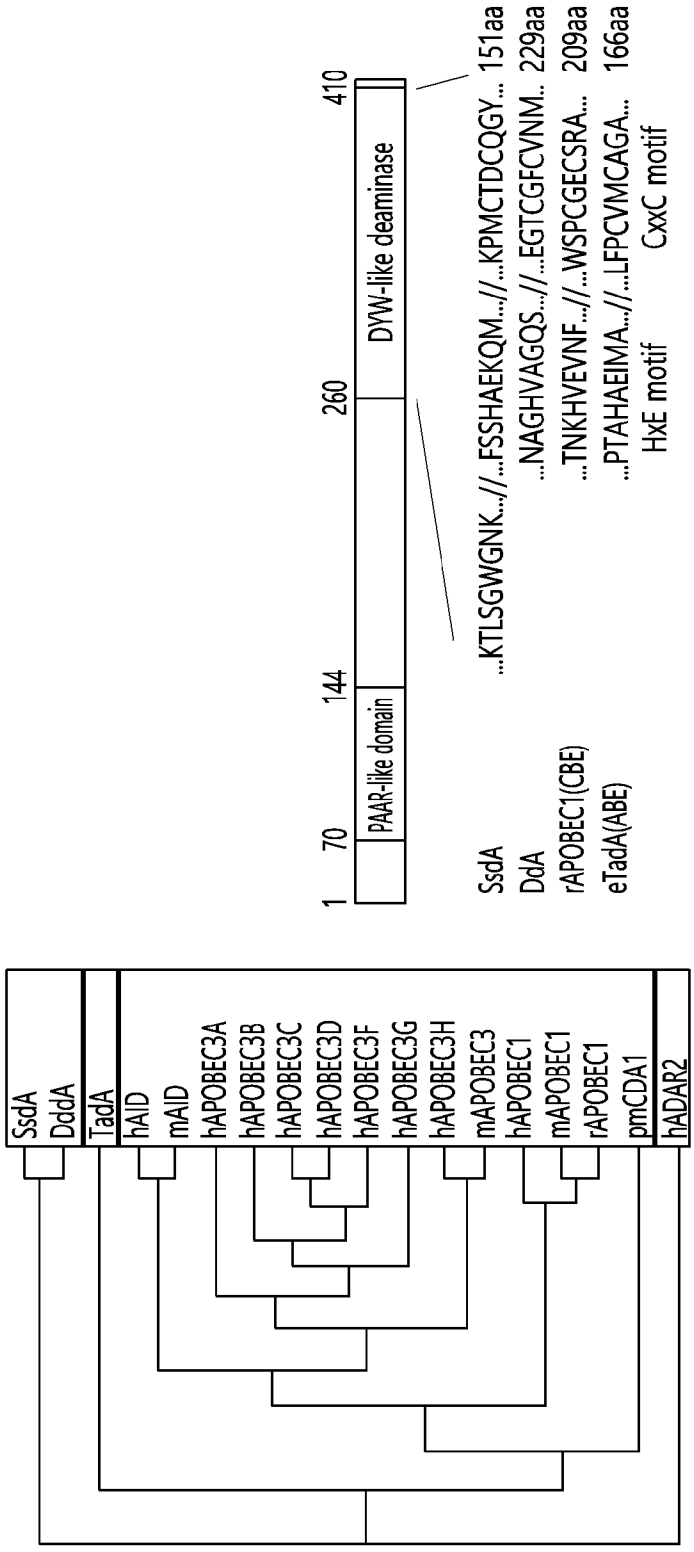
[청구항 22] 핵산(nucleic acid) 분자를 CRISPR-Cas 시스템과 접촉시키는 단계를 포함하는 핵산을 편집하는 방법으로,
상기 편집은 상기 핵산 분자의 뉴클레오타이드(nucleotide) 서열 중 적어도 하나 이상의 뉴클레오타이드 서열의 결실(deletion), 삽입(insertion), 치환(substitution) 또는 인델(insertion and deletion; indel)을 형성하는 것이고,

상기 CRISPR-Cas 시스템은 Cas 단백질(CRISPR-associated protein) 및 박테리아 독신(toxin)을 포함하는 융합 단백질 또는 상기 융합 단백질을 암호화하는 폴리뉴클레오티드; 및 가이드 폴리뉴클레오티드를 포함하고, 상기 박테리아 독신은 SsdA(single-stranded DNA deaminase toxin A)인 것인, 핵산을 편집하는 방법.

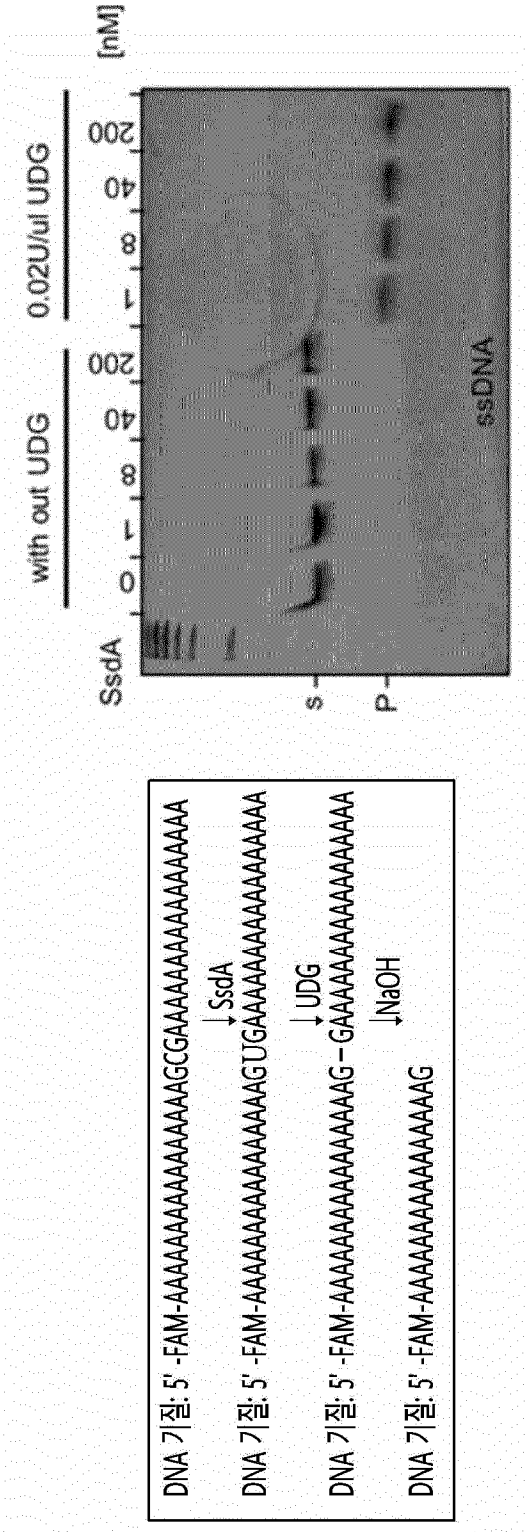
[청구항 23] 청구항 22에 있어서,

상기 편집은 상기 핵산 분자의 뉴클레오타이드 서열 중 1 bp 내지 60 bp의 뉴클레오타이드의 결실, 1 bp 내지 60 bp의 뉴클레오타이드의 삽입, 또는 1 bp 내지 60 bp의 뉴클레오타이드의 인델인 것인, 핵산을 편집하는 방법.

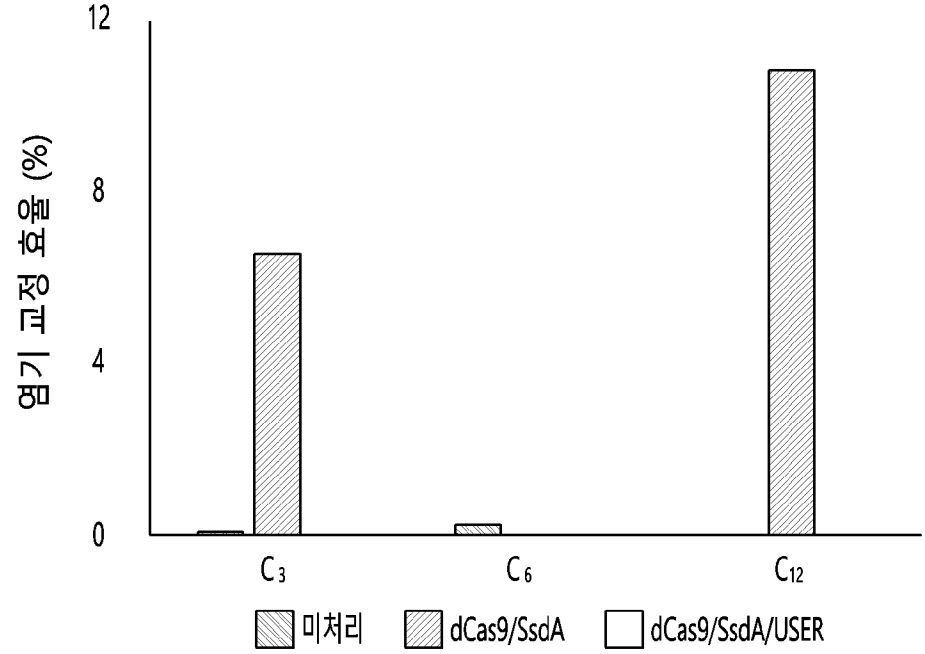
[도1]


1
70
144
260
410
PAAR-like domain
DYW-like deaminase
SsdA ...KTLSGWGNK...//...FSSHAEKQM.../...KPMCTDCQGY... 151aa
DdA ...NAGHVAGQS.../...EGTCGFCVNM.. 229aa
rAPOBEC1(CBE) ...TNKHVEVNF.../...WSPCGECSRA... 209aa
eTadA(ABE) ...PTAHAEIMA.../...LFPCVMCAGA... 166aa
HxE motif CxxC motif
SsdA
DddA
TadA
hAID
mAID
hAPOBEC3A
hAPOBEC3B
hAPOBEC3C
hAPOBEC3D
hAPOBEC3F
hAPOBEC3G
hAPOBEC3H
mAPOBEC3
hAPOBEC1
mAPOBEC1
rAPOBEC1
pmCDA1
hADAR2

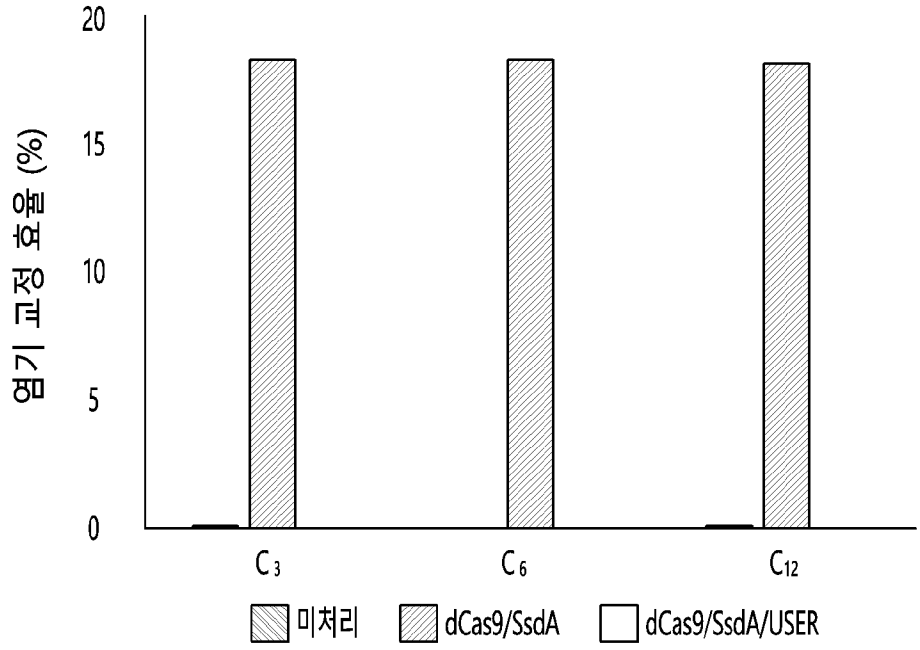
[도2]


[nM]
0.02U/ul UDG
with out UDG
0
1
8
40
200
200
40
8
1
0
SsdA
s
P
ssDNA
DNA 기질: 5' -FAM-AAAAAAAAAAAAAAAAAGCGAAAAAAAAAAAAAAAA
↓SsdA
DNA 기질: 5' -FAM-AAAAAAAAAAAAAAAAGUGAAAAAAAAAAAAAAAA
↓UDG
DNA 기질: 5' -FAM-AAAAAAAAAAAAAAAAG – GAAAAAAAAAAAAAAAA
↓NaOH
DNA 기질: 5' -FAM-AAAAAAAAAAAAAAAAG

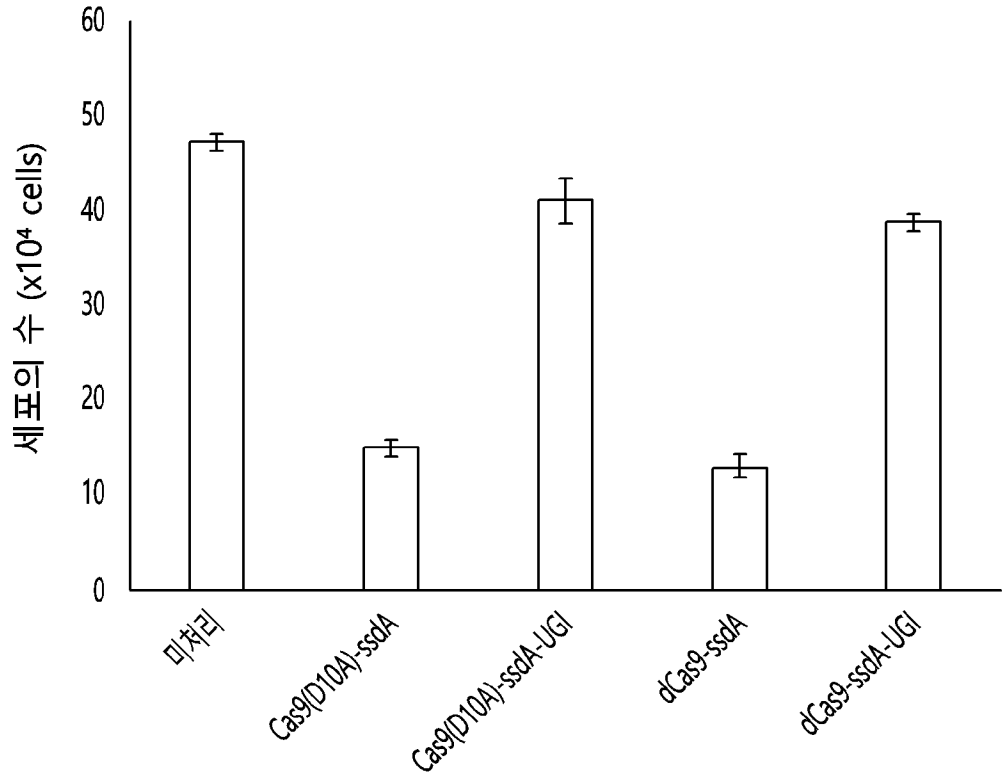
[도3a]


12
8
4
0
염기 교정 효율 (%)
C3
C6
C12
미처리
dCas9/SsdA
dCas9/SsdA/USER

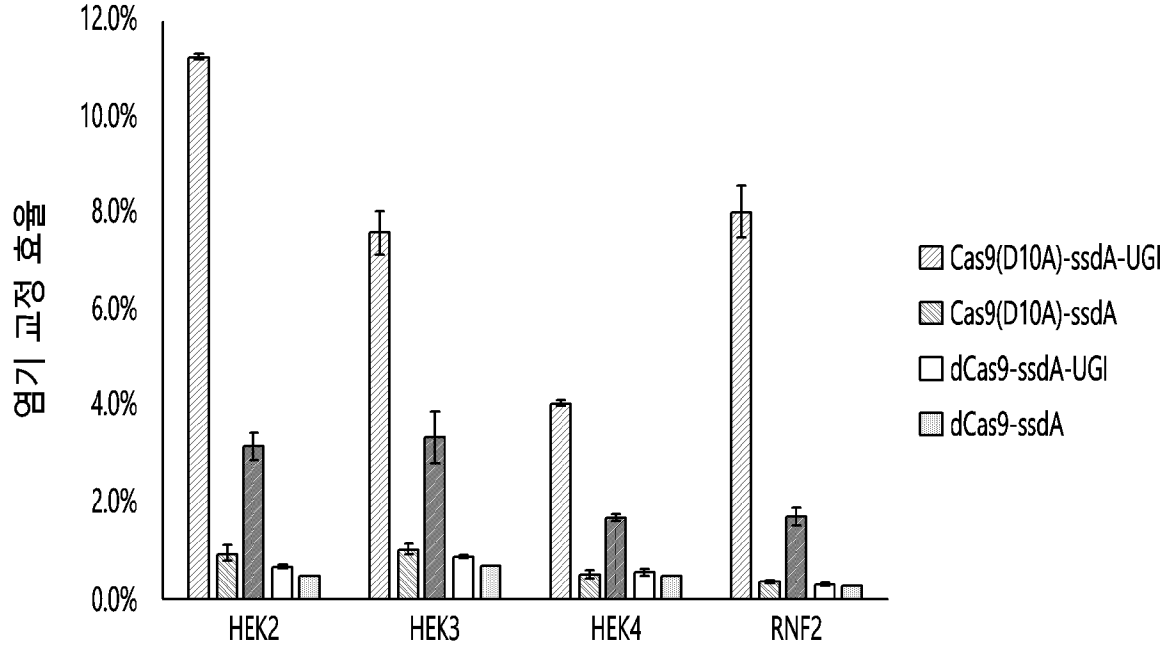
[도3b]

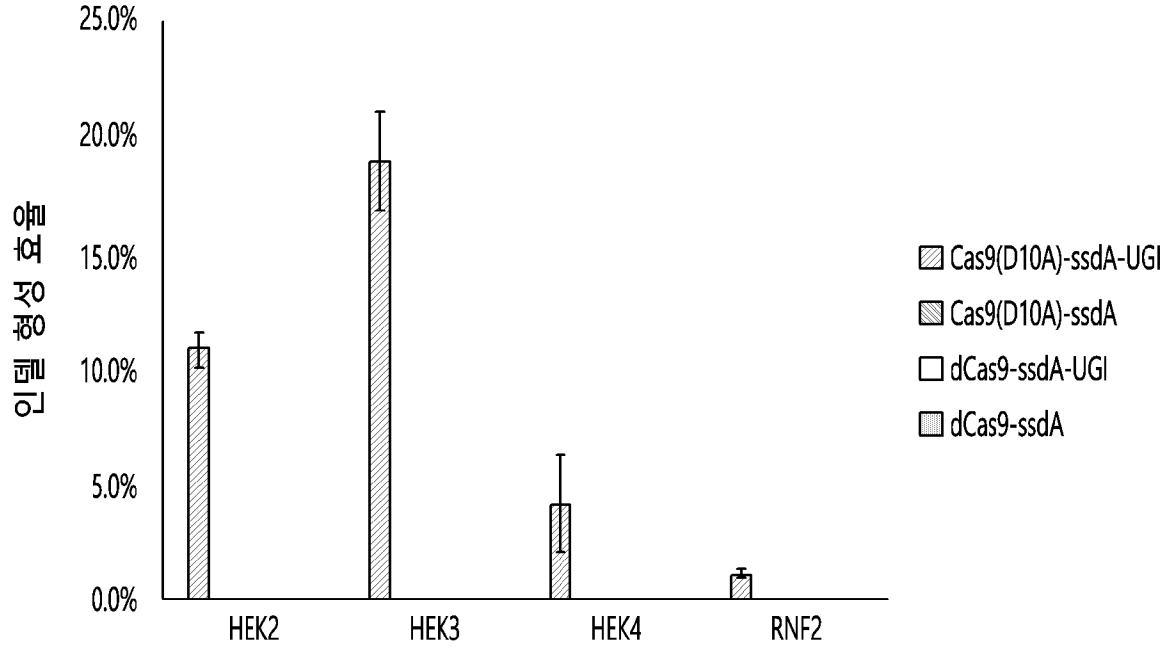

20
15
10
5
0
염기 교정 효율 (%)
C3
C6
C12
미처리
dCas9/SsdA
dCas9/SsdA/USER

[도4]

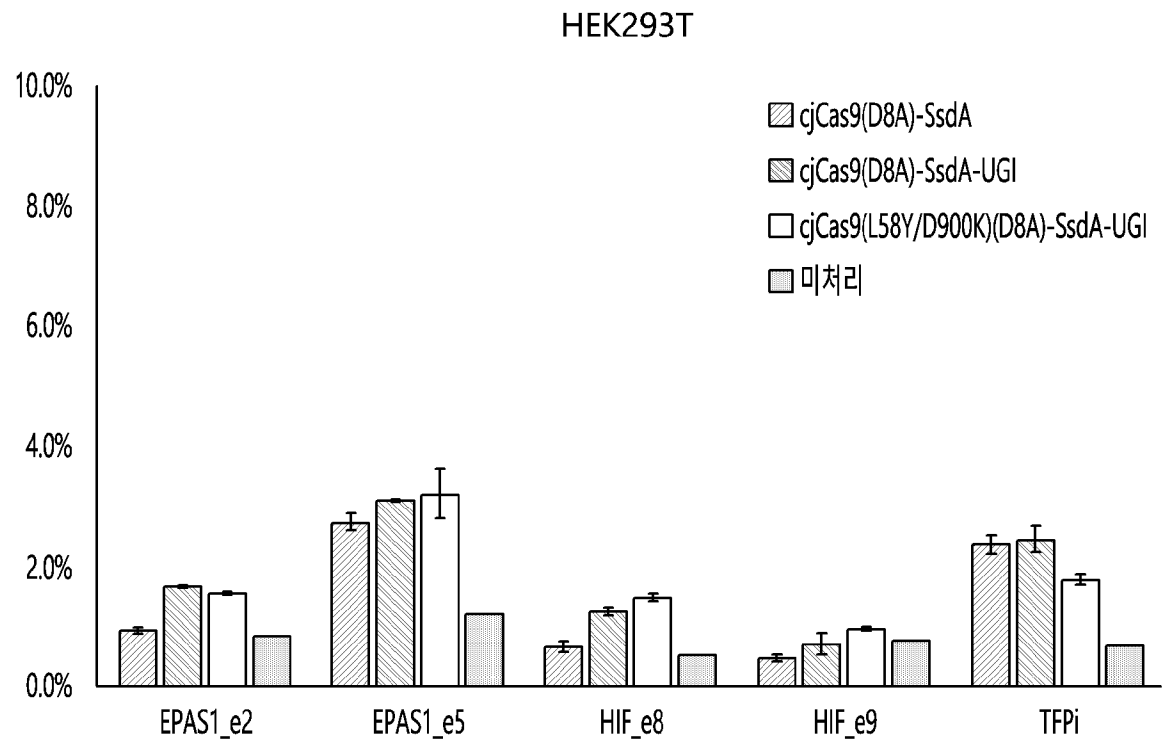

60
50
40
30
20
10
0
세포의 수 (x10⁴ cells)
미처리
Cas9(D10A)-ssdA
Cas9(D10A)-ssdA-UGI
dCas9-ssdA
dCas9-ssdA-UGI

[도5a]

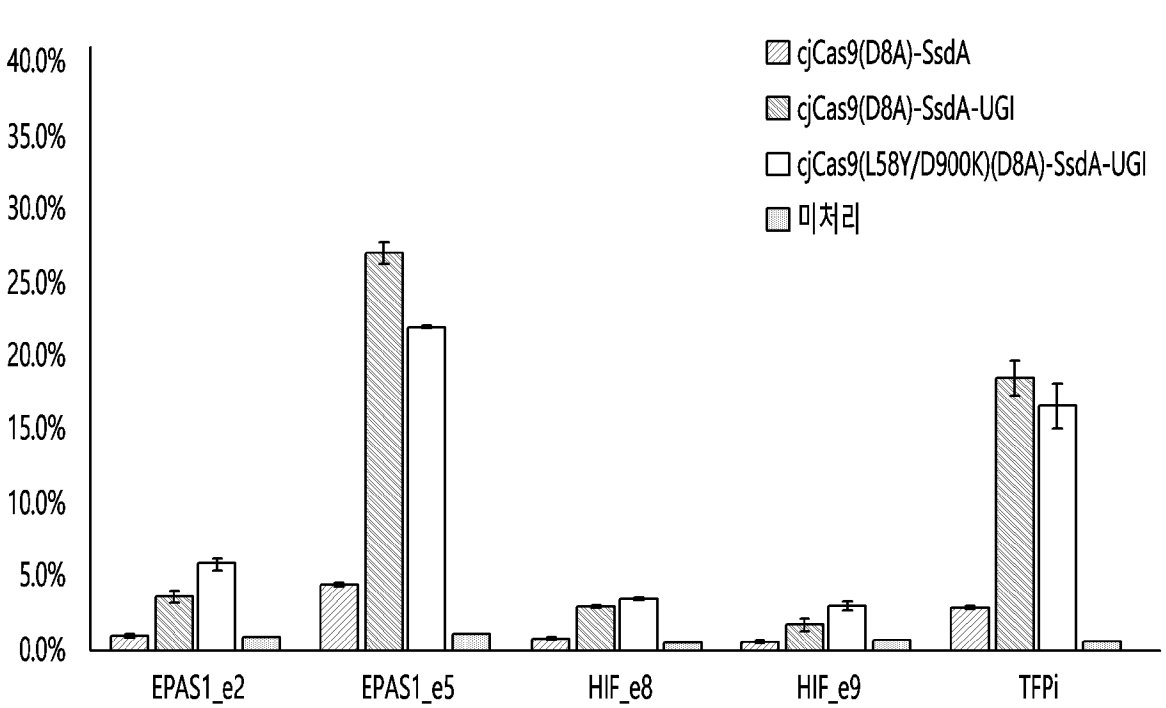

12.0%
10.0%
8.0%
6.0%
4.0%
2.0%
0.0%
염기 교정 효율
HEK2
HEK3
HEK4
RNF2
Cas9(D10A)-ssdA-UGI
Cas9(D10A)-ssdA
dCas9-ssdA-UGI
dCas9-ssdA


25.0%
20.0%
15.0%
10.0%
5.0%
0.0%
인델 형성 효율
HEK2
HEK3
HEK4
RNF2
Cas9(D10A)-ssdA-UGI
Cas9(D10A)-ssdA
dCas9-ssdA-UGI
dCas9-ssdA

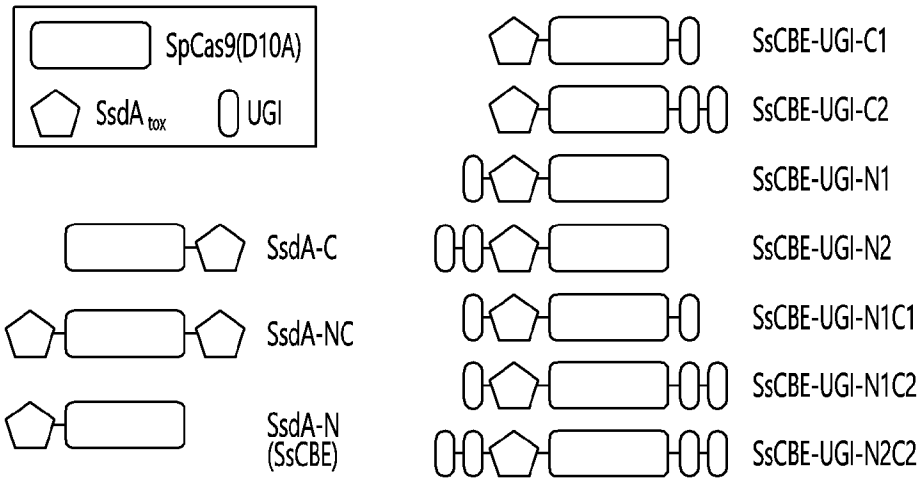
[도5b]


HEK293T
10.0%
8.0%
6.0%
4.0%
2.0%
0.0%
cjCas9(D8A)-SsdA
cjCas9(D8A)-SsdA-UGI
cjCas9(L58Y/D900K)(D8A)-SsdA-UGI
미처리
EPAS1_e2
EPAS1_e5
HIF_e8
HIF_e9
TFPi

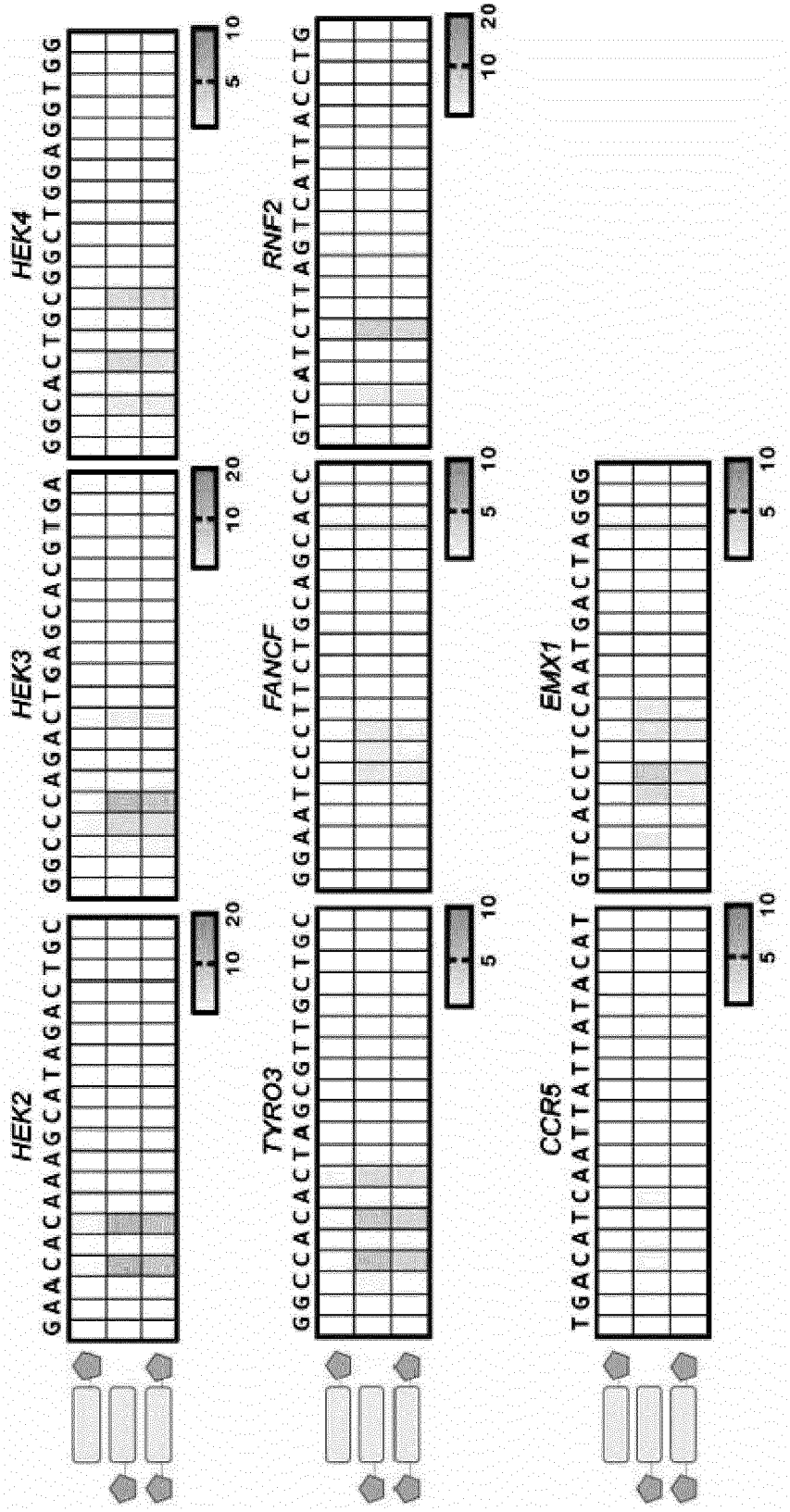
[도5c]


HEK293T-UNG KD
40.0%
35.0%
30.0%
25.0%
20.0%
15.0%
10.0%
5.0%
0.0%
cjCas9(D8A)-SsdA
cjCas9(D8A)-SsdA-UGI
cjCas9(L58Y/D900K)(D8A)-SsdA-UGI
미처리
EPAS1_e2
EPAS1_e5
HIF_e8
HIF_e9
TFPi

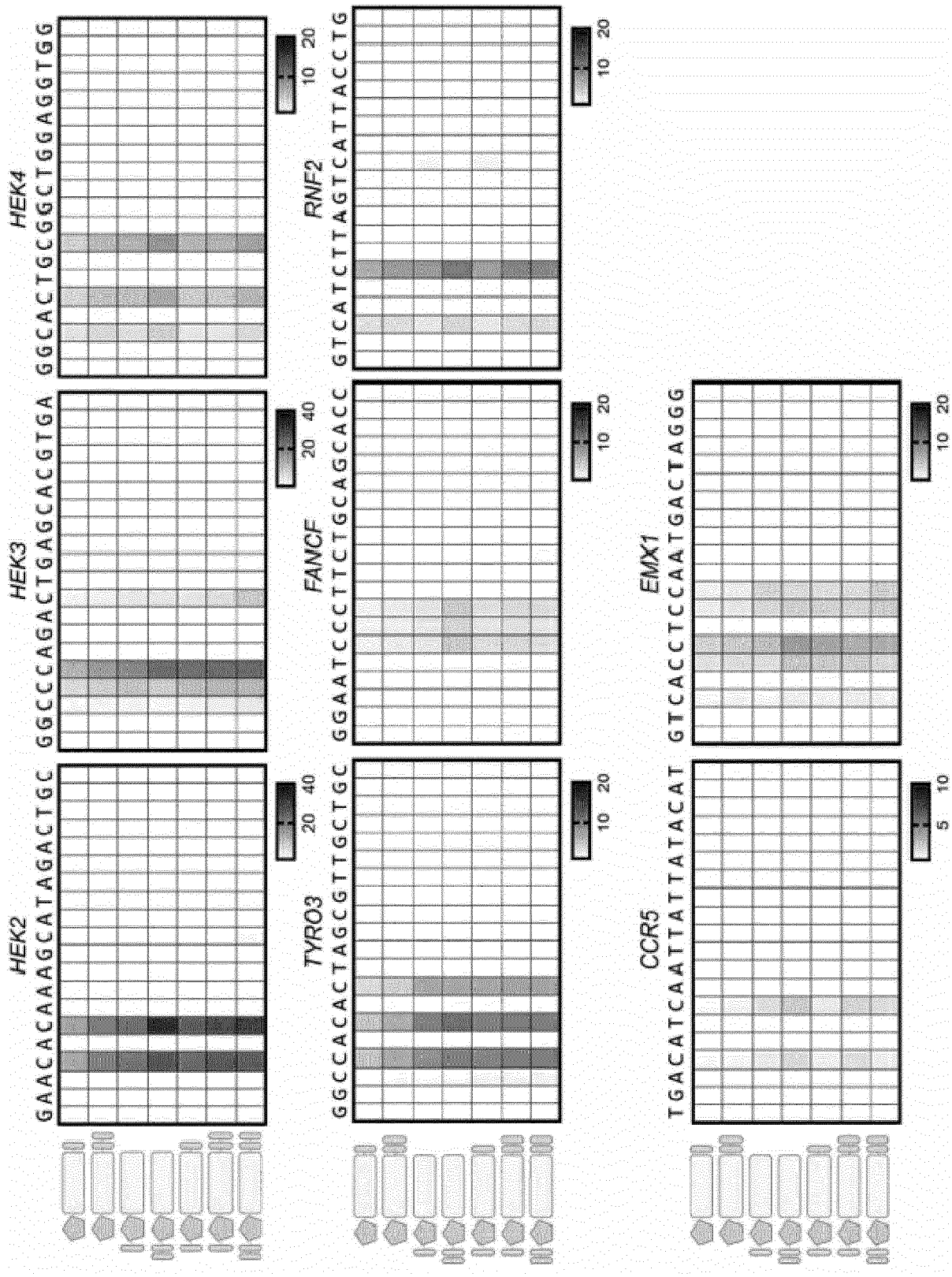
[도6]


SpCas9(D10A)
SsdA tox
UGI
SsdA-C
SsdA-NC
SsdA-N
(SsCBE)
SsCBE-UGI-C1
SsCBE-UGI-C2
SsCBE-UGI-N1
SsCBE-UGI-N2
SsCBE-UGI-N1C1
SsCBE-UGI-N1C2
SsCBE-UGI-N2C2

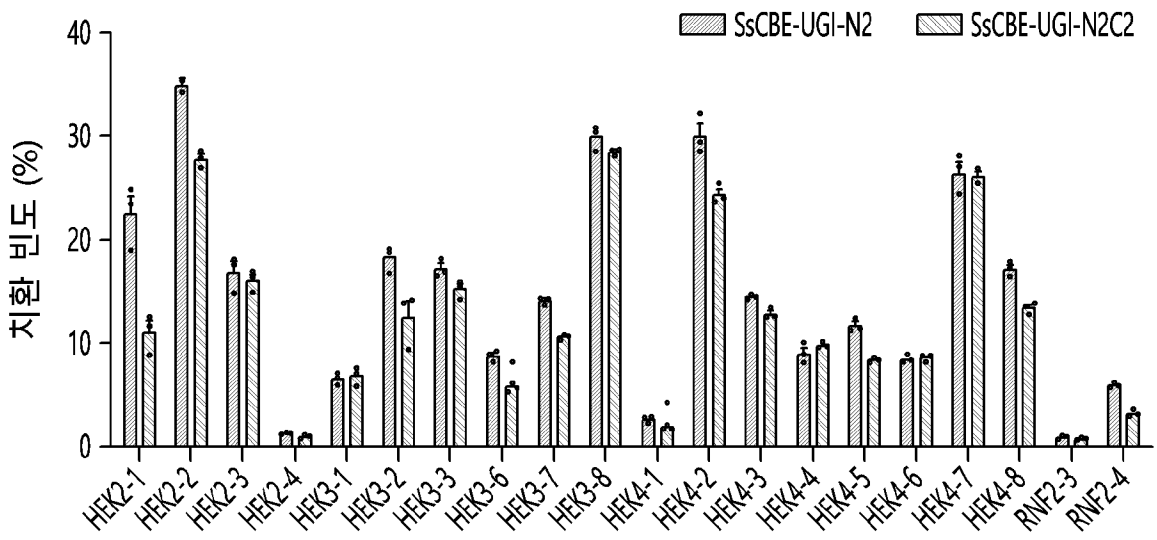
[도7a]


HEK4
GGCACTGCGGCTGGAGGTGG
HEK3
GGCCCAGACTGAGCACGTGA
HEK2
GAACACAAAGCATAGACTGC
RNF2
GTCATCTTAGTCATTACCTG
FANCF
GGAATCCCTTCTGCAGCACC
TYRO3
GGCCACACTAGCGTTGCTGC
EMX1
GTCACCTCCAATGACTAGGG
CCR5
TGACATCAATTATTATACAT
10
5
20
10

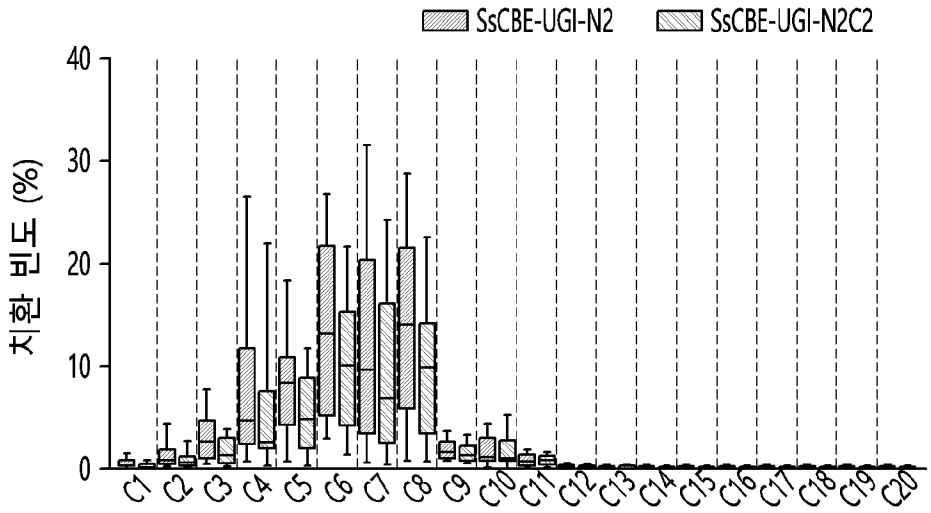
[도7b]


HEK4
GGCACTGCGGCTGGAGGTGG
HEK3
GGCCCAGACTGAGCACGTGA
HEK2
GAACACAAAGCATAGACTGC
RNF2
GTCATCTTAGTCATTACCTG
FANCF
GGAATCCCTTCTGCAGCACC
TYRO3
GGCCACACTAGCGTTGCTGC
EMX1
GTCACCTCCAATGACTAGGG
CCR5
TGACATCAATTATTATACAT
10 20
20 40
5 10

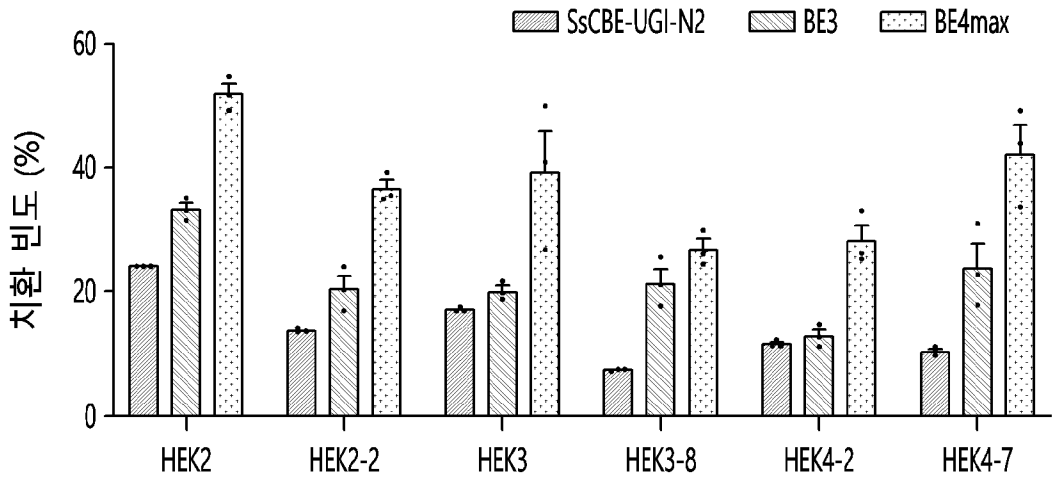
[도7c]


SsCBE-UGI-N2
SsCBE-UGI-N2C2
치환 빈도 (%)
0
10
20
30
40
HEK2-1
HEK2-2
HEK2-3
HEK2-4
HEK3-1
HEK3-2
HEK3-3
HEK3-6
HEK3-7
HEK3-8
HEK4-1
HEK4-2
HEK4-3
HEK4-4
HEK4-5
HEK4-6
HEK4-7
HEK4-8
RNF2-3
RNF2-4

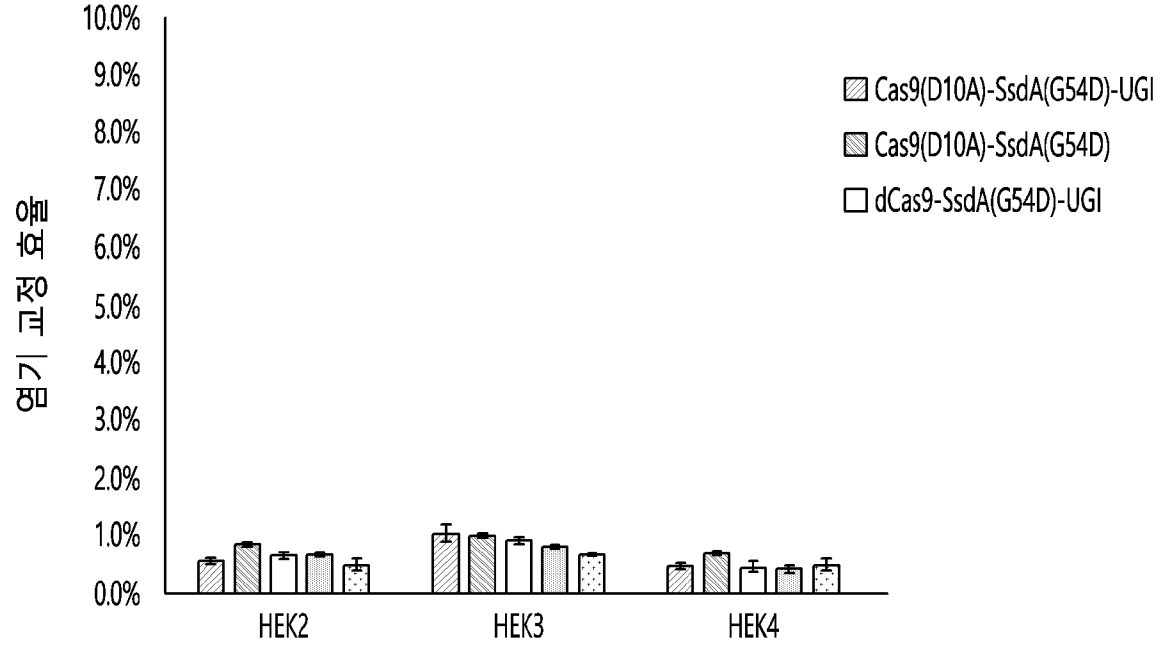
[도7d]

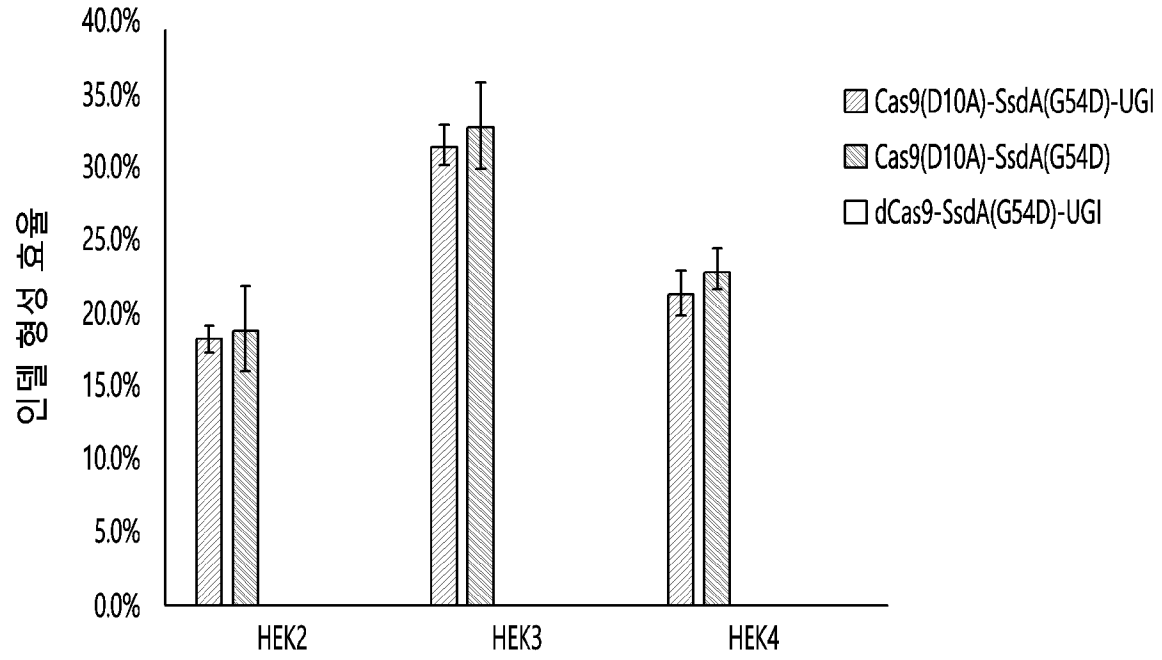

SsCBE-UGI-N2
SsCBE-UGI-N2C2
치환 빈도 (%)
0
10
20
30
40
C1
C2
C3
C4
C5
C6
C7
C8
C9
C10
C11
C12
C13
C14
C15
C16
C17
C18
C19
C20

[도8]


SsCBE-UGI-N2
BE3
BE4max
치환 빈도 (%)
0
20
40
60
HEK2
HEK2-2
HEK3
HEK3-8
HEK4-2
HEK4-7

[도9]


염기 교정 효율
10.0%
9.0%
8.0%
7.0%
6.0%
5.0%
4.0%
3.0%
2.0%
1.0%
0.0%
HEK2
HEK3
HEK4
Cas9(D10A)-SsdA(G54D)-UGI
Cas9(D10A)-SsdA(G54D)
dCas9-SsdA(G54D)-UGI


인델 형성 효율
40.0%
35.0%
30.0%
25.0%
20.0%
15.0%
10.0%
5.0%
0.0%
HEK2
HEK3
HEK4
Cas9(D10A)-SsdA(G54D)-UGI
Cas9(D10A)-SsdA(G54D)
dCas9-SsdA(G54D)-UGI

INTERNATIONAL SEARCH REPORT

International application No.
PCT/KR2023/016230

A. CLASSIFICATION OF SUBJECT MATTER

C12N 9/22(2006.01)i; **C12N 9/78**(2006.01)i; **C12N 15/62**(2006.01)i; **C12N 15/10**(2006.01)i; **C12N 15/113**(2010.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12N 9/22(2006.01); A61K 31/167(2006.01); A61K 48/00(2006.01); C07K 14/415(2006.01); C12N 15/10(2006.01); C12N 15/82(2006.01); C12N 9/78(2006.01)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models: IPC as above
Japanese utility models and applications for utility models: IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS (KIPO internal) & keywords: Cas 단백질(Cas protein), 유전자 가위(genetic scissors, CRISPR), Cas9, Cas12, 박테리아 톡신(bacteria toxin), SsdA(single-stranded DNA toxin A), 시티딘 탈아민효소(cytidine deaminase)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	VELAZQUEZ MUNOZ, Elena et al. Recombination-independent genomic editing and chromosomal site-focused diversification of Gram-negative bacteria. Dissertation, Departamento de Biología, Universidad Autonoma de Madrid. 2021, pp. 1-151. See pages 16-35.	1-23
A	MOK, B. Y. et al. A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. Nature. July 2020, vol. 583, thesis no. 7817, pp. 631-637. See abstract; and page 3.	1-23
A	WO 2022-212584 A1 (UNIVERSITY OF WASHINGTON et al.) 06 October 2022 (2022-10-06) See entire document.	1-23
A	KR 10-2019-0008548 A (TOOLGEN INCORPORATED) 24 January 2019 (2019-01-24) See entire document.	1-23
A	KR 10-2021-0005184 A (OREGON HEALTH & SCIENCE UNIVERSITY) 13 January 2021 (2021-01-13) See entire document.	1-23

☑ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 January 2024	**30 January 2024**

Name and mailing address of the ISA/KR	Authorized officer
Korean Intellectual Property Office **Government Complex-Daejeon Building 4, 189 Cheongsa-ro, Seo-gu, Daejeon 35208**	
Facsimile No. **+82-42-481-8578**	Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/KR2023/016230

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2021-048272 A1 (JOHN INNES CENTRE) 18 March 2021 (2021-03-18) See entire document.	1-23

INTERNATIONAL SEARCH REPORT

International application No.
PCT/KR2023/016230

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
   a. ☑ forming part of the international application as filed.
   b. ☐ furnished subsequent to the international filing date for the purposes of international search (Rule 13*ter*.1(a)),
      ☐ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.

2. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.

3. Additional comments:

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/KR2023/016230

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
WO 2022-212584 A1	06 October 2022	None	
KR 10-2019-0008548 A	24 January 2019	CN 109312386 A	05 February 2019
		CN 109312386 B	25 October 2022
		EP 3473728 A1	24 April 2019
		EP 3473728 B1	20 October 2021
		JP 2019-517802 A	27 June 2019
		JP 6940117 B2	22 September 2021
		KR 10-2523302 B1	20 April 2023
		US 2019-0177710 A1	13 June 2019
		WO 2017-217768 A1	21 December 2017
KR 10-2021-0005184 A	13 January 2021	CN 112135640 A	25 December 2020
		EP 3787693 A1	10 March 2021
		JP 2021-522787 A	02 September 2021
		US 2021-0130832 A1	06 May 2021
		WO 2019-213065 A1	07 November 2019
WO 2021-048272 A1	18 March 2021	None	

국 제 조 사 보 고 서

국제출원번호
PCT/KR2023/016230

A. 발명이 속하는 기술분류(국제특허분류(IPC))

C12N 9/22(2006.01)i; **C12N 9/78**(2006.01)i; **C12N 15/62**(2006.01)i; **C12N 15/10**(2006.01)i; **C12N 15/113**(2010.01)i

B. 조사된 분야

조사된 최소문헌(국제특허분류를 기재)

C12N 9/22(2006.01); A61K 31/167(2006.01); A61K 48/00(2006.01); C07K 14/415(2006.01); C12N 15/10(2006.01); C12N 15/82(2006.01); C12N 9/78(2006.01)

조사된 기술분야에 속하는 최소문헌 이외의 문헌

한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC
일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC

국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우))

eKOMPASS(특허청 내부 검색시스템) & 키워드: Cas 단백질(Cas protein), 유전자 가위(genetic scissors, CRISPR), Cas9, Cas12, 박테리아 톡신(bacteria toxin), SsdA(single-stranded DNA toxin A), 시티딘 탈아민효소(cytidine deaminase)

C. 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
A	MUNOZ, ELENA VELAZQUEZ, "Recombination-independent genomic editing and chromosomal site-focused diversification of Gram-negative bacteria", Dissertation. 2021., Universidad Autonoma de Madrid, Departamento de Biología, 페이지 1-151 페이지 16-35	1-23
A	MOK, B. Y. 등, 'A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing', Nature, 2020.07., 제583권, 논문번호 7817, 631-637 페이지 초록; 페이지 3	1-23
A	WO 2022-212584 A1 (UNIVERSITY OF WASHINGTON 등) 2022.10.06 전체 문헌	1-23
A	KR 10-2019-0008548 A (주식회사 툴젠) 2019.01.24 전체 문헌	1-23
A	KR 10-2021-0005184 A (오레곤 헬스 앤드 사이언스 유니버시티) 2021.01.13 전체 문헌	1-23

☑ 추가 문헌이 C(계속)에 기재되어 있습니다.

☑ 대응특허에 관한 별지를 참조하십시오.

* 인용된 문헌의 특별 카테고리:

"A" 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌
"D" 본 국제출원에서 출원인이 인용한 문헌
"E" 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후에 공개된 선출원 또는 특허 문헌
"L" 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌
"O" 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌
"P" 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌
"T" 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌
"X" 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다.
"Y" 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다.
"&" 동일한 대응특허문헌에 속하는 문헌

국제조사의 실제 완료일	국제조사보고서 발송일
2024년01월30일(30.01.2024)	**2024년01월30일(30.01.2024)**

ISA/KR의 명칭 및 우편주소

대한민국 특허청
(35208) 대전광역시 서구 청사로 189, 4동 (둔산동, 정부대전청사)

팩스 번호 **+82-42-481-8578**

심사관

허주형

전화번호 **+82-42-481-5373**

국 제 조 사 보 고 서

국제출원번호
PCT/KR2023/016230

C. 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
A	WO 2021-048272 A1 (JOHN INNES CENTRE) 2021.03.18 전체 문헌	1-23

국 제 조 사 보 고 서

국제출원번호
PCT/KR2023/016230

제1기재란 핵산염기 및/또는 아미노산 서열(첫 번째 용지의 1.c의 계속)

1. 국제출원에 개시된 핵산염기 및/또는 아미노산 서열과 관련하여, 국제조사는 다음에 기초하여 수행되었습니다.

   a. ☑ 출원시 국제출원의 일부를 구성하는 서열목록

   b. ☐ 국제조사를 목적으로 국제출원일 이후에 제출된 서열목록(규칙 13의3.1(a))

      ☐ 서열목록이 출원시 국제출원의 개시 범위를 넘지 않는다는 취지의 진술서를 첨부

2. ☐ 국제출원에 개시된 핵산염기 및/또는 아미노산 서열에 대해, 본 보고서는 WIPO 표준 ST.26을 준수하는 서열목록이 없이 유효한 조사를 할 수 있는 범위에서 작성되었습니다

3. 추가 의견:

국 제 조 사 보 고 서
대응특허에 관한 정보

국제출원번호
PCT/KR2023/016230

국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
WO 2022-212584 A1	2022/10/06	없음	
KR 10-2019-0008548 A	2019/01/24	CN 109312386 A	2019/02/05
		CN 109312386 B	2022/10/25
		EP 3473728 A1	2019/04/24
		EP 3473728 B1	2021/10/20
		JP 2019-517802 A	2019/06/27
		JP 6940117 B2	2021/09/22
		KR 10-2523302 B1	2023/04/20
		US 2019-0177710 A1	2019/06/13
		WO 2017-217768 A1	2017/12/21
KR 10-2021-0005184 A	2021/01/13	CN 112135640 A	2020/12/25
		EP 3787693 A1	2021/03/10
		JP 2021-522787 A	2021/09/02
		US 2021-0130832 A1	2021/05/06
		WO 2019-213065 A1	2019/11/07
WO 2021-048272 A1	2021/03/18	없음	