

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국

(43) 국제공개일
2019년 8월 1일 (01.08.2019)



(10) 국제공개번호
WO 2019/147073 A1

(51) 국제특허분류:

C12Q 1/6869 (2018.01) *C12N 9/78* (2006.01)
C12Q 1/34 (2006.01) *C12N 9/96* (2006.01)
C12N 9/22 (2006.01)

(21) 국제출원번호: PCT/KR2019/001104

(22) 국제출원일: 2019년 1월 25일 (25.01.2019)

(25) 출원언어: 한국어

(26) 공개언어: 한국어

(30) 우선권정보:
10-2018-0009508 2018년 1월 25일 (25.01.2018) KR

(71) 출원인: 주식회사 툴젠 (**TOOLGEN INCORPORATED**) [KR/KR]; 08501 서울시 금천구 가산디지털1로 219, 제1204호, Seoul (KR). 서울대학교산학협력단 (**SEOUL NATIONAL UNIVERSITY R&DB FOUNDATION**) [KR/KR]; 08826 서울시 관악구 관악로 1, Seoul (KR).

(72) 발명자: 김대식 (**KIM, Dae-sik**); 08826 서울시 관악구 관악로 1, Seoul (KR).

(74) 대리인: 유미특허법인 (**YOU ME PATENT AND LAW FIRM**); 06134 서울시 강남구 테헤란로 115, 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, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

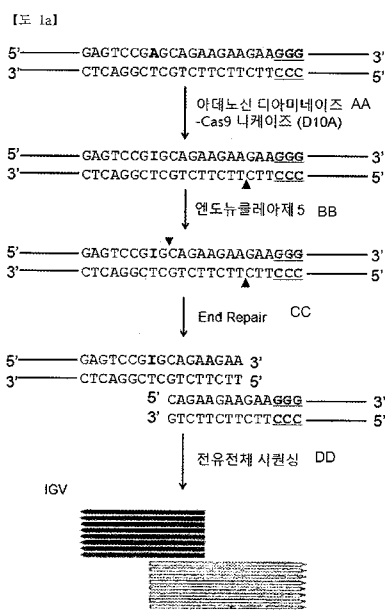
(84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

공개:

— 국제조사보고서와 함께 (조약 제21조(3))

(54) Title: METHOD FOR IDENTIFYING BASE EDITING BY USING ADENOSINE DEAMINASE

(54) 발명의 명칭: 아데노신 디아미나아제를 이용한 염기 교정 확인 방법



AA ... Adenosine deaminase-Cas9 nickase (D10A)
BB ... Endonuclease 5
CC ... End Repair
DD ... Genome-wide sequencing

(57) Abstract: Provided are: a base editing composition containing a deaminase and a target-specific nuclease; and a base editing method and a genetically modified animal preparation method, which use the base editing composition.

(57) 요약서: 디아미나아제 및 표적특이적 뉴클레아제를 포함하는 염기 교정용 조성물, 및 상기 염기 교정용 조성물을 이용한 염기 교정 방법 및 유전자 변형 동물 제조 방법이 제공된다.

WO 2019/147073 A1

【발명의 설명】

【발명의 명칭】

아데노신 디아미나아제를 이용한 염기 교정 확인 방법

5 【기술분야】

아데노신 디아미나아제 (adenosine deaminase)를 이용한 염기 교정 확인 기술과 관련된 것으로, 아데노신 디아미나아제를 포함하는 염기 교정 확인용 조성물 및 DNA에 아데노신 디아미나아제를 처리하는 단계를 포함하는 염기 교정 확인 방법이 제공된다.

10

【발명의 배경이 되는 기술】

CRISPR/Cas9 system은 인간세포를 비롯한 여러 동식물의 유전체 교정에 사용되어 왔지만 Cas9의 경우 DNA 절단으로 인해 표적부위에 DNA의 삽입이나 결실이 일어나게 된다. 이러한 단점을 보완하기 위해 최근에 Cas9
15 니케이즈 (Cas9 nickase)에 시토신 디아미나아제 (cystidine deaminase)를 연결시켜 DNA 절단 없이 C를 T(U)로 (또는 보다 낮은 빈도로, C를 G 또는 A로 변환) 변환시키는 방법이 개발되었다 (Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A. & Liu, D.R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533, 420-424
20 (2016) 참조).

하지만 시토신 디아미나아제 (cystidine deaminase)를 이용하여 C를 T로 (또는 보다 낮은 빈도로, C를 G 또는 A로 변환) 변환하거나 C에 대응하는 반대쪽 strand의 G를 A로 (또는 보다 낮은 빈도로, G를 C 또는 T로 변환)하는 것은 가능하지만, 퓨린 계열 염기인 A 또는 T를 다른 염기로
25 바꾸는 방법은 알려져 있지 않았다.

【발명의 내용】

【해결하고자 하는 과제】

본 명세서에서는 절단 유전체 시퀀싱 (digested genome sequencing, Digenome-seq; 표적 특이적 뉴클레아제 처리 전과 후를 한 눈에 파악해
30 잘린 위치를 구별하는 방식)을 이용하여 아데노신 디아미나아제 (adenosine deaminase)를 표적특이적 뉴클레아제 (예컨대, Cas9 니케이즈 또는 촉매적으로 결합된 Cas9 (dCas9) 등)에 결합시켜 만든 아데노신 염기교정

유전자 가위의 염기 교정을 확인하는 방안을 제공하며, 이를 통하여 상기 아데노신 염기교정 유전자 가위에 의한 염기 교정의 특이성을 평가할 수 있다.

S. pyogenes 유래 Cas9 니카아제 및 *E. coli* 유래 조작된 아데닌디아미나아제로 구성된 Adenine Base Editors (ABEs)는 가이드 RNA 표적 방식으로 효율적인 아데닌/구아닌 단일 뉴클레오티드 전환을 가능하게 하여, 동물 (예컨대, 인간 등의 포유류, 마우스 등의 설치류 등) 및/또는 식물의 세포에서 점돌연변이를 유도하거나 교정한다. ABE는 다른 유전자 교정 수단인 Cas9, BE (cytosine base editor) 등과 다른 off-target site를 인식하므로, 그 효율성 평가를 위하여 독립적인 방법을 개발하는 것이 필요하다. 본 명세서에서는 전체 유전체 시퀀싱 (whole genome sequencing: WGS)을 통하여, ABE의 유전체 (genome) DNA 전체에 대한 표적 특이성을 평가하는 방법을 제시한다. 상기 전체 유전체 시퀀싱은 구체적으로 Digenome-seq일 수 있다. 본 명세서에서 제공되는 Digenome-seq 기반 방법은 0.1%의 치환 빈도(substitution frequency)로 ABE 비표적 위치를 포획하기에 충분한 정도로 민감하다. 또한, 미리 조립된 ABE 리보핵산단백질, 변형된 가이드 RNA, 및 Sniper-ABE를 사용하여 인간 등의 동물 세포에서 ABE off-target 활성을 감소시킬 수 있음을 제안한다.

일 예는,

(a) (1) 아데노신 디아미나아제, 아데노신 디아미나아제를 암호화하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 플라스미드, 및 (2) 표적 특이적 뉴클레아제, 표적 특이적 뉴클레아제를 암호화하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 플라스미드를 분리된 세포에 도입시키거나 또는 분리된 DNA에 처리하여, DNA 이중가닥 중 하나를 절단하는 단계; 및

(b) 상기 DNA 이중가닥 중 하나가 절단된 DNA에 단일가닥 특이적 엔도뉴클레아제 (예컨대, 엔도뉴클레아제 V (endonuclease V) 또는, Alkyladenine DNA Glycosylase (AAG)와 엔도뉴클레아제 VIII (endonuclease VIII)의 조합)를 처리하여, 나머지 DNA 가닥을 절단하는 단계

를 포함하는, 아데닌 디아미나아제를 사용하여 DNA에 이중 가닥 절단 (double strand break)를 생성하는 방법을 제공한다.

다른 예는,

(a) (1) 아데노신 디아미나아제, 이를 암호화 하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 플라스미드, 및 (2) 표적 특이적 뉴클레아제,

이를 암호화하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 플라스미드를 분리된 세포에 도입시키거나 또는 분리된 DNA에 처리하여, DNA 이중가닥 중 하나를 절단하는 단계;

(b) 상기 DNA 이중가닥 중 하나가 절단된 DNA에 단일가닥 특이적
5 엔도뉴클레아제 (예컨대, 엔도뉴클레아제 V (endonuclease V) 또는, Alkyladenine DNA Glycosylase (AAG)와 엔도뉴클레아제 VIII (endonuclease VIII)의 조합)를 처리하여, 나머지 DNA 가닥을 절단하여, 이중 가닥 절단 (double-strand break)를 형성하는 단계; 및

(c) 상기 이중가닥 절단된 DNA에 대하여 전체 유전체 시퀀싱 (whole
10 genome sequencing)을 수행하는 단계

를 포함하는 DNA 서열 분석 방법을 제공한다.

다른 예는,

(a) (1) 아데노신 디아미나아제, 아데노신 디아미나아제를
암호화하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 플라스미드, 및 (2)
15 표적 특이적 뉴클레아제, 표적 특이적 뉴클레아제를 암호화하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 플라스미드를 분리된 세포에 도입시키거나 또는 분리된 DNA에 처리하여, DNA 이중가닥 중 하나를 절단하는 단계;

(b) 상기 DNA 이중가닥 중 하나가 절단된 DNA에 단일가닥 특이적
엔도뉴클레아제 (예컨대, 엔도뉴클레아제 V (endonuclease V) 또는,
20 Alkyladenine DNA Glycosylase (AAG)와 엔도뉴클레아제 VIII (endonuclease VIII)의 조합)를 처리하여, 나머지 DNA 가닥을 절단하여, 이중 가닥 절단 (double-strand break)를 형성하는 단계;

(c) 상기 이중가닥 절단된 DNA에 대하여 전체 유전체 시퀀싱 (whole
genome sequencing)을 수행하는 단계; 및

(d) 상기 시퀀싱으로 획득한 염기서열 데이터 (sequence read)에서
25 상기 절단된 위치를 확인하는 단계

를 포함하는,

표적 특이적 뉴클레아제의 절단 위치 또는 비표적 위치 (off-target
site)를 검출하는 방법을 제공한다.

30

【과제의 해결 수단】

아데노신 디아미나아제 (adenosine deaminase)를 방향 진화(directed
evolution)시킴으로써, DNA의 염기 중 A(adenosine)를

디아미네이션(deamination)시켜서 I(inosine)으로 치환시키는데 성공한 바 있다. 보다 구체적으로, Cas9 니케이즈 (nCas9) 또는 촉매적으로 결핍된 Cas9 (dCas9)에 DNA에 작동할 수 있도록 방향 진화된 아데노신 디아미나아제를 결합시킨 후, 세포 내로 도입하였을 때 DNA의 A가 G(I)로
 5 (또는 보다 낮은 빈도로, A가 C 또는 T로 변환) 변환하는 것을 관찰하였다.

이러한 사항에 기초하여, 아데노신 디아미나아제를 DNA 시퀀싱에 적용한 기술을 제공한다.

일 예는

(a) (1) 아데노신 디아미나아제, 아데노신 디아미나아제를
 10 암호화하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 플라스미드, 및 (2) 표적 특이적 뉴클레아제, 표적 특이적 뉴클레아제를 암호화하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 플라스미드를 분리된 세포에 도입시키거나 또는 분리된 DNA에 처리하여, DNA 이중가닥 중 하나를 절단하는 단계; 및

(b) 상기 DNA 이중가닥 중 하나가 절단된 DNA에 단일가닥 특이적
 15 엔도뉴클레아제 (예컨대, 엔도뉴클레아제 V (endonuclease V) 또는, Alkyladenine DNA Glycosylase (AAG)와 엔도뉴클레아제 VIII (endonuclease VIII)의 조합)를 처리하여, 나머지 DNA 가닥을 절단하는 단계

를 포함하는, 아데닌 디아미나아제를 사용하여 DNA에 이중 가닥 절단 (double strand break)를 생성하는 방법을 제공한다.

20 다른 예는,

(a) (1) 아데노신 디아미나아제, 아데노신 디아미나아제를
 암호화하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 플라스미드, 및 (2) 표적 특이적 뉴클레아제, 표적 특이적 뉴클레아제를 암호화하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 플라스미드를 분리된 세포에 도입시키거나
 25 또는 분리된 DNA에 처리하여, DNA 이중가닥 중 하나를 절단하는 단계;

(b) 상기 DNA 이중가닥 중 하나가 절단된 DNA에 단일가닥 특이적
 엔도뉴클레아제 (예컨대, 엔도뉴클레아제 V (endonuclease V) 또는, Alkyladenine DNA Glycosylase (AAG)와 엔도뉴클레아제 VIII (endonuclease VIII)의 조합)를 처리하여, 나머지 DNA 가닥을 절단하여, 이중 가닥 절단
 30 (double-strand break)를 형성하는 단계; 및

(c) 상기 이중가닥 절단된 DNA에 대하여 전체 유전체 시퀀싱 (whole genome sequencing)을 수행하는 단계

를 포함하는 DNA 서열 분석 방법을 제공한다.

다른 예는,

(a) (1) 아데노신 디아미나아제, 아데노신 디아미나아제를 암호화하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 플라스미드, 및 (2) 표적 특이적 뉴클레아제, 표적 특이적 뉴클레아제를 암호화하는 핵산 분자, 또는
5 상기 핵산 분자를 포함하는 플라스미드를 분리된 세포에 도입시키거나 또는 분리된 DNA에 처리하여, DNA 이중가닥 중 하나를 절단하는 단계;

(b) 상기 DNA 이중가닥 중 하나가 절단된 DNA에 단일가닥 특이적 엔도뉴클레아제 (예컨대, 엔도뉴클레아제 V (endonuclease V) 또는 Alkyladenine DNA Glycosylase (AAG)와 엔도뉴클레아제 VIII (endonuclease
10 VIII)의 조합)를 처리하여, 나머지 DNA 가닥을 절단하여, 이중 가닥 절단 (double-strand break)를 형성하는 단계;

(c) 상기 이중가닥 절단된 DNA에 대하여 전체 유전체 시퀀싱 (whole genome sequencing)을 수행하는 단계; 및

(d) 상기 시퀀싱으로 획득한 염기서열 데이터 (sequence read)에서
15 상기 절단된 위치를 확인하는 단계

를 포함하는,

표적 특이적 뉴클레아제의 절단 위치 또는 비표적 위치 (off-target site)를 검출하는 방법을 제공한다.

상기 비표적 위치 (off-target site)를 검출하는 방법은 상기 (d)
20 단계 이후에, 상기 확인된 절단된 위치가 표적 위치 (on-target site)가 아닌 경우, 비표적 위치 (off-target site)로 판단하는 단계 ((e) 단계)를 추가로 포함할 수 있다.

다른 예는,

(a) (1) 아데노신 디아미나아제, 아데노신 디아미나아제를
25 암호화하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 플라스미드, 및 (2) 표적 특이적 뉴클레아제, 표적 특이적 뉴클레아제를 암호화하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 플라스미드를 분리된 세포에 도입시키거나 또는 분리된 DNA에 처리하여, DNA 이중가닥 중 하나를 절단하는 단계;

(b) 상기 DNA 이중가닥 중 하나가 절단된 DNA에 단일가닥 특이적
30 엔도뉴클레아제 (예컨대, 엔도뉴클레아제 V (endonuclease V) 또는 Alkyladenine DNA Glycosylase (AAG)와 엔도뉴클레아제 VIII (endonuclease VIII)의 조합)를 처리하여, 나머지 DNA 가닥을 절단하여, 이중 가닥 절단 (double-strand break)를 형성하는 단계;

(c) 상기 이중가닥 절단된 DNA에 대하여 전체 유전체 시퀀싱 (whole genome sequencing)을 수행하는 단계; 및

(d) 상기 시퀀싱으로 수득한 염기서열 데이터 (sequence read)에서 상기 절단된 위치를 확인하는 단계

5 를 포함하는,

표적 특이적 뉴클레아제의 교정 효율 및/또는 정확도 확인 방법을 제공한다.

상기 표적 특이적 뉴클레아제의 교정 효율 확인 방법은 상기 (d) 단계 이후에, 상기 확인된 절단된 위치가 표적 위치 (on-target site)가
10 아닌 경우, 비표적 위치 (off-target site)로 판단하는 단계 ((e) 단계) 및
상기 비표적 위치에서의 절단 정도 (비표적 위치 개수 및/또는 비표적 위치에서의 절단 빈도)를 측정하여 비교 대상의 절단 정도와 비교하는 단계(e-1)를 추가로 포함할 수 있으며, 이 경우, 비표적 위치에서의 절단 정도가 낮을수록 교정 효율 및/또는 정확도가 높다고 판단할 수 있다.
15 상기 비교 대상은 임의의 표적 DNA의 표적 서열에 대한 표적 특이적 뉴클레아제일 수 있으며, 일 예에서, 통상적으로 사용되거나 이미 알려진 표적 특이적 뉴클레아제 (예컨대, RGEN 및 가이드 RNA 조합)들 중 선택된 어느 하나일 수 있다.

상기 단계 (a)에 있어서, (1) 아데노신 디아미나아제, 아데노신
20 디아미나아제를 암호화하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 플라스미드, 및 (2) 표적 특이적 뉴클레아제, 표적 특이적 뉴클레아제를 암호화하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 플라스미드는 분리된 세포 또는 분리된 DNA에 동시에 (함께) 처리될 수 있으며, 일
예에서, 아데노신 디아미나아제와 표적 특이적 뉴클레아제가 융합된 융합
25 단백질 (예컨대, 아데노신 디아미나아제가 N 말단쪽에 위치하고 표적 특이적 뉴클레아제가 C 말단쪽에 위치할 수 있음), 또는 상기 융합 단백질을 암호화하는 핵산 서열 또는 상기 핵산 서열을 포함하는 플라스미드 형태로 사용될 수 있다. 이 때, (3) 표적 특이적 뉴클레아제에 따라 적절한 가이드 RNA, 또는 이를 암호화하는 핵산 분자 또는 이를
30 포함하는 플라스미드를 추가로 포함할 수 있다.

상기 단계 (a)에 있어서, 아데노신 디아미나아제에 의하여, DNA의 이중가닥 중 PAM 서열이 존재하는 가닥의 표적 부위 (PAM의 5' 말단쪽으로 17 내지 23 nt 부위) 내의 아데노신의 이노신으로의 전환이 일어나고,

표적특이적 뉴클레아제에 의하여 상기 아데노신-이노신 전환이 일어난 가닥 (즉, PAM 서열이 존재하는 가닥)의 반대 가닥 (상보적 가닥)에서 단일 가닥 절단 (nick)이 발생한다.

상기 단계 (b)에 있어서, 엔도뉴클레아제 V를 사용하는 경우, 상기
 5 단계 (a)에 의하여 단일 가닥 절단이 일어난 가닥의 반대 가닥 (즉, 아데노신-이노신 전환이 일어난 가닥)의 상기 아데노신이 이노신으로 변환된 변이 위치에서 5' 방향 및/또는 3' 방향으로 5nt 이내, 4nt 이내, 3nt 이내, 2nt 이내, 또는 1nt 이내에 위치하는 뉴클레오타이드의 3' 말단이 절단된다. 상기 단계 (b)에 있어서, Alkyladenine DNA Glycosylase (AAG)와
 10 엔도뉴클레아제 VIII (endonuclease VIII)의 조합을 사용하는 경우, AAG에 의하여 상기 단계 (a)에서 전환된 이노신이 제거되어 AP 부위 (apurinic site)이 생성되고, 엔도뉴클레아제 VIII는 상기 AP 부위를 인식하여 AP 부위의 3' 및 5' 말단을 절단한다. 이러한 단계 (b)를 수행함으로써, 단계 (a)에서 단일 가닥 절단이 일어난 가닥과 반대 가닥에도 단일 가닥 절단이
 15 일어나, 결과적으로 이중 가닥 절단이 생성된다.

다른 예는

- (1) 아데노신 디아미나아제, 아데노신 디아미나아제를 암호화하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 플라스미드,
 - (2) 표적 특이적 뉴클레아제, 표적 특이적 뉴클레아제를 암호화하는
 20 핵산 분자, 또는 상기 핵산 분자를 포함하는 플라스미드, 및
 - (3) 단일가닥 특이적 엔도뉴클레아제 (예컨대, 엔도뉴클레아제 V (endonuclease V) 또는, Alkyladenine DNA Glycosylase (AAG)와 엔도뉴클레아제 VIII (endonuclease VIII)의 조합)
- 을 포함하는, DNA 이중 가닥 절단 (double strand breaks; DSBs)용
 25 조성물을 제공한다.

본 명세서에서 제공된 방법 및 조성물들에 있어서, 유전체 교정 (염기 교정) 및/또는 절단 대상이 되는 DNA는 유전체(genomic) DNA 또는 DNA 단편 (PCT 산물)일 수 있으며, 세포 내에 존재하거나 세포로부터 분리된 것일 수 있다. 상기 세포는 인간, 마우스 등의 포유류 등과 같은
 30 진핵 세포일 수 있다. 또한, 본 명세서에서 제공된 방법 및 조성물들에 있어서, 아데노신 디아미나아제 및 표적 특이적 뉴클레아제는 직접적 또는 펩타이드 링커를 통하여 간접적으로 연결된 융합 단백질 또는 이를 암호화하는 핵산 분자 또는 상기 핵산 분자를 포함하는 플라스미드 형태로

사용될 수 있다.

또한, 본 명세서에서 제공된 방법 및 조성물들에 있어서, 표적 특이적 뉴클레아제는 가이드 RNA와 함께 사용 또는 포함될 수 있다.

이 경우, 아데노신 디아미나아제, 표적 특이적 뉴클레아제, 및
5 가이드 RNA는

(i) 아데노신 디아미나아제, 표적 특이적 뉴클레아제, 및 가이드 RNA의 혼합물,

(ii) 아데노신 디아미나아제 암호화 핵산 서열 또는 이를 포함하는 플라스미드, 표적 특이적 뉴클레아제 암호화 핵산 서열 또는 이를 포함하는
10 플라스미드, 및 가이드 RNA 또는 이의 암호화 핵산 분자를 포함하는 플라스미드의 혼합물,

(iii) 아데노신 디아미나아제 및 표적 특이적 뉴클레아제가 연결된 융합 단백질 및 가이드 RNA의 혼합물,

(iv) 아데노신 디아미나아제 및 표적 특이적 뉴클레아제가 연결된
15 융합 단백질 암호화 핵산 분자 또는 상기 핵산 분자를 포함하는 플라스미드, 및 가이드 RNA 또는 이의 암호화 핵산 분자를 포함하는 플라스미드의 혼합물, 및

(v) 아데노신 디아미나아제 및 표적 특이적 뉴클레아제가 연결된 융합 단백질 및 가이드 RNA의 복합체 (리보핵산단백질, RNP) 또는 혼합물

20 로 이루어진 군에서 선택된 하나 이상의 형태로 사용 (예컨대, 분리된 세포 내로 도입 또는 분리된 DNA에 처리 또는 조성물에 포함)될 수 있다.

상기 플라스미드는 상기 아데노신 디아미나아제 암호화 유전자 및/또는 불활성화된 표적특이적 뉴클레아제 암호화 유전자를 삽입하고 이를
25 숙주세포 내에서 발현시킬 수 있는 발현 시스템을 포함하는 모든 플라스미드일 수 있다. 상기 플라스미드는 목적 유전자 발현을 위한 요소 (elements)를 포함하는 것으로, 복제원점 (replication origin), 프로모터, 작동 유전자 (operator), 전사 종결 서열 (terminator) 등을 포함할 수 있고, 숙주 세포의 게놈 내로의 도입을 위한 적절한 효소 부위 (예컨대,
30 제한 효소 부위) 및/또는 임의로 숙주 세포 내로의 성공적인 도입을 확인하기 위한 선별 마커 및/또는 단백질로의 번역을 위한 리보솜 결합 부위 (ribosome binding site; RBS) 및/또는 전자 조절 인자 등을 추가로 포함할 수 있다. 상기 플라스미드는 당업계에서 사용되는 플라스미드,

예컨대, pcDNA 시리즈, pSC101, pGV1106, pACYC177, ColE1, pKT230, pME290, pBR322, pUC8/9, pUC6, pBD9, pHc79, pIJ61, pLAFR1, pHV14, pGEX 시리즈, pET 시리즈, pUC19 등으로 이루어진 군에서 선택된 1종 이상일 수 있으나, 이에 제한되는 것은 아니다. 상기 숙주세포는 상기 아데노신

5 디아미나제에 의하여 염기 교정 또는 이중 가닥 절단을 도입하고자 하는 세포 (예컨대, 인간 세포 등의 포유류 등과 같은 진핵 세포) 또는 상기 아데노신 디아미나아제 암호화 유전자 및/또는 불활성화된 표적특이적 뉴클레아제 암호화 유전자를 발현하여 아데노신 디아미나아제 및 불활성화된 표적특이적 뉴클레아제를 발현할 수 있는 모든 세포 (예컨대, E.

10 coli 등) 들 중에서 선택될 수 있다.

본 명세서에서 절단 유전체 시퀀싱 (digested genome sequencing, Digenome-seq) 기술이라 함은 뉴클레아제에 의해 절단된 유전체의 서열 분석을 의미하는 것으로, 세포에서의 전체 유전체에서의 뉴클레아제 비표적 효과 (off-target effect)를 분석하기 위한 *in vitro* nuclease-digested

15 whole-genome sequencing에 적용될 수 있다. 뉴클레아제의 절단 부위에서 동일한 5' 말단을 갖는 sequence reads를 생성하며, 이들은 적절한 프로그램 (예컨대, Digenome program)에 의하여 컴퓨터로 확인 가능하다. 일 예에서, 절단 유전체 시퀀싱은 앞서 기재한 유전체 DNA (genome DNA) 서열 분석 방법 또는 뉴클레아제의 비표적 위치 (off-target site)를

20 검출하는 방법에서의 단계 (a), (b), 및 (c), 또는 단계 (a), (b), (c), 및 (d)를 포함하는 것으로 정의될 수 있다. 이를 다르게 설명하면, 상기 단계 (a), (b), 및 (c), 또는 단계 (a), (b), (c), 및 (d)는 절단 유전체 시퀀싱에 의하여 수행될 수 있다.

본 명세서에 사용된 바로서, 용어 "염기 변이 (또는 염기 치환)"은

25 해당 염기를 포함하는 뉴클레오타이드에 변이 (예컨대, 치환)이 일어난 것을 의미하는 것으로, "뉴클레오타이드 변이 (또는 뉴클레오타이드 치환)"과 동일한 의미로 사용될 수 있으며, 이러한 염기 변이는 대립유전자 중 하나 또는 두 개 모두에서 일어날 수 있다.

유전체 교정 및/또는 유전자 교정 기술은, 인간 세포를 비롯한

30 동식물 세포의 유전체 염기서열에 표적 지향형 변이를 도입할 수 있는 기술로서, 특정 유전자를 녹아웃 (knock-out) 또는 녹인 (knock-in)하거나, 단백질을 생성하지 않는 비-코딩 DNA 서열에 변이를 도입하는 등 다양한 형태로 수행될 수 있다. 본 명세서에서 제안되는 방법은 상기 유전체 교정

및/또는 유전자 교정 기술에 사용되는 표적 특이적 뉴클레아제의 비표적 위치를 검출하는 것으로, 이는 표적 위치에만 특이적으로 작동하는 표적 특이적 뉴클레아제 시스템을 개발하는데 유용하게 이용될 수 있다.

상기 (a) 단계는 분리된 세포 또는 생체 또는 세포로부터 분리된 유전체 (genomic) DNA에 아데노신 디아미나아제와 표적 특이적 뉴클레아제를 처리하여 DNA 이중가닥 중 한 가닥을 절단하는 단계로서, 분리된 유전체 DNA를 시험관 내 (*in vitro*)에서 특정 표적에 특이적으로 작용하는 뉴클레아제로 절단하는 단계이다. 상기 뉴클레아제는 표적 특이적으로 제작하였더라도 특이성에 따라 다른 부위, 즉 비표적 위치를 절단할 가능성을 갖는다. 따라서 결과적으로 상기 (a) 단계에 의해, 사용된 표적 특이적 뉴클레아제가 유전체 DNA에 대하여 활성을 가질 수 있는 위치인 표적 위치 또는 다수의 비표적 위치를 절단함으로써 특정 위치가 절단된 유전체 DNA 단편(lead)을 얻을 수 있다.

상기 분리된 유전체 DNA는 비형질전환 세포 (야생형 세포) 및/또는 표적 특이적 뉴클레아제가 발현되거나 뉴클레아제 활성을 갖도록 형질전환된 세포로부터 분리된 것일 수 있으며, 표적 특이적 뉴클레아제의 비표적 위치를 검출하고자 하는 목적에 따라 그 유래에 제한 없이 사용될 수 있다. 상기 유전체가 분리되는 세포는 모든 원핵 세포 및 진핵세포 (예컨대, 인간 등의 포유동물 세포, 진핵식물 세포 등)들 중에서 선택된 것일 수 있다.

상기 아데노신 디아미나아제는 퓨린 대사에 관여하는 효소(EC 3.5.4.4)로서, 아데노신을 디아미네이션(deamination; 아미노기를 케토기로 치환)시켜서 이노신 (inosine)으로 변환시키는 기능하는 모든 효소를 총칭한다. 일 예에서, 아데노신 디아미나아제는 원핵세포 또는 진핵 세포, 예컨대, 진핵 동물 (예컨대, 어류, 양서류, 파충류, 조류, 포유류 등), 또는 진핵 식물 유래의 것일 수 있으며, 예컨대, 인간 아데노신 디아미나아제 (예컨대, GenBank Accession No. NP_000013.2 (코딩 유전자: NM_000022.3), NP_001308979.1 (코딩 유전자: NM_001322050.1), NP_001308980.1 (코딩 유전자: NM_001322051.1) 등), 마우스 아데노신 디아미나아제 (예컨대, GenBank Accession No. NP_001258981.1 (코딩 유전자: NM_001272052.1), NP_031424.1 (코딩 유전자: NM_007398.4) 등), *E. coli* TadA (NP_417054.2) 등으로 이루어진 군에서 선택된 1종 이상일 수 있다. 상기 아데노신 디아미나아제는 단백질 또는 이를 암호화하는 DNA

(임의로 적절한 재조합 벡터에 포함될 수 있음), 또는 이를 암호화하는 mRNA 형태로 사용될 수 있다 또한 아데노신 디아미나아제는 기존 아데노신 디아미나아제가, 예컨대, *E. coli* 등의 원핵세포에서, 인위적 진화 (directed evolution; DE) 된 것일 수 있다. 인위적 진화 (directed evolution; DE)는 단백질 엔지니어링 방법 중 하나로, 돌연변이유발 과정 (변이체 라이브러리 생성), 선별 과정 (상기 변이체를 발현시키고 이 중에서 목적하는 기능(변이)을 갖는 변이체를 분리), 및 증폭 또는 재생산 과정 (다음 라운드를 위한 템플릿 생성)를 포함할 수 있고, in vivo (생체 내) 또는 in vitro에서 수행될 수 있다. 일 예에서, 상기 인위적 진화된 아데노신 디아미나아제는 *E. coli* TadA (NP_417054.2)를 *E. coli*에서 인위적 진화시킨 것일 수 있다. 일 예에서, 상기 인위적 진화된 아데노신 디아미나아제는 NP_417054.2의 N 말단의 메티오닌(M)을 제외한 아미노산 서열 중, 22번째 아미노산 잔기 (W), 35번째 아미노산 잔기 (H), 47번째 아미노산 잔기 (P), 50번째 아미노산 잔기 (R), 83번째 아미노산 잔기 (L), 105번째 아미노산 잔기 (A), 107번째 아미노산 잔기(D), 122번째 아미노산 잔기 (H), 145번째 아미노산 잔기 (S), 146번째 아미노산 잔기 (D), 151번째 아미노산 잔기 (R), 154번째 아미노산 잔기 (E), 155번째 아미노산 잔기 (I), 156번째 아미노산 잔기 (K) (이상, 아미노산 위치는 NP_417054.2의 두 번째 아미노산부터 기산함) 등으로 이루어진 군에서 선택된 하나 이상의 아미노산이 원래(야생형)와 다른 아미노산으로 치환 또는 결실된 것일 수 있으나 (참고예 2 참조), 이에 제한되는 것은 아니다.

상기 표적 특이적 뉴클레아제는 유전자 가위 (programmable nuclease)라고도 불리며, 목적하는 유전체 DNA 상의 특정 위치를 인식하여 절단할 수 있는 모든 형태의 뉴클레아제를 통칭한다.

본 명세서에서 사용된 표적 특이적 뉴클레아제는 DNA 이중가닥 중 한 가닥을 절단하는 활성을 갖는 모든 뉴클레아제일 수 있다.

일 구체예에서, 상기 표적 특이적 뉴클레아제는 Cas 단백질 (예컨대, Cas9 단백질(CRISPR (Clustered regularly interspaced short palindromic repeats) associated protein 9)), Cpf1 단백질 (CRISPR from *Prevotella* and *Francisella* 1) 등과 같은 타입 II 및/또는 타입 V의 CRISPR 시스템에 수반되는 뉴클레아제 (예컨대, 엔도뉴클레아제) 등으로 이루어진 군에서 선택된 1종 이상일 수 있다. 이 경우, 상기 표적 특이적 뉴클레아제는 유전체 DNA의 표적 부위로 안내하기 위한 표적 DNA 특이적 가이드 RNA를

추가로 포함할 수 있다. 상기 가이드 RNA는 생체 외 (in vitro)에서 전사된(transcribed) 것일 수 있고, 예컨대 올리고뉴클레오타이드 이중가닥 또는 플라스미드 주형으로부터 전사된 것일 수 있으나, 이에 제한되지 않는다. 상기 표적 특이적 뉴클레아제는 가이드 RNA에 결합된 리보핵산-
 5 단백질 복합체를 형성(RNA-Guided Engineered Nuclease)하여 리보핵산 단백질 (RNP) 형태로 작용할 수 있다.

Cas 단백질은 CRISPR/Cas 시스템의 주요 단백질 구성 요소로, 활성화된 엔도뉴클레아제 또는 nickase를 형성할 수 있는 단백질이다.

Cas 단백질 또는 유전자 정보는 NCBI (National Center for
 10 Biotechnology Information)의 GenBank와 같은 공지의 데이터 베이스에서 얻을 수 있다. 예컨대, 상기 Cas 단백질은,

스트렙토코커스 sp. (*Streptococcus* sp.), 예컨대, 스트렙토코커스 피요게네스 (*Streptococcus pyogenes*) 유래의 Cas 단백질, 예컨대, Cas9 단백질 (예컨대, SwissProt Accession number Q99ZW2(NP_269215.1));

15 캄필로박터 속, 예컨대, 캄필로박터 제주니 (*Campylobacter jejuni*) 유래의 Cas 단백질, 예컨대, Cas9 단백질;

스트렙토코커스 속, 예컨대, 스트렙토코커스 써모필러스 (*Streptococcus thermophiles*) 또는 스트렙토코커스 아우레우스 (*Streptococcus aureus*) 유래의 Cas 단백질, 예컨대, Cas9 단백질;

20 네이세리아 메닝기디티스 (*Neisseria meningitidis*) 유래의 Cas 단백질, 예컨대, Cas9 단백질;

파스테우렐라 (*Pasteurella*) 속, 예컨대, 파스테우렐라 물토시다 (*Pasteurella multocida*) 유래의 Cas 단백질, 예컨대 Cas9 단백질;

25 프란시셀라 (*Francisella*) 속, 예컨대, 프란시셀라 노비시다 (*Francisella novicida*) 유래의 Cas 단백질, 예컨대 Cas9 단백질

등으로 이루어진 군에서 선택된 하나 이상일 수 있으나, 이에 제한되는 것은 아니다.

Cpf1 단백질은 상기 CRISPR/Cas 시스템과는 구별되는 새로운 CRISPR 시스템의 엔도뉴클레아제로서, Cas9에 비해 상대적으로 크기가 작고
 30 tracrRNA가 필요 없으며, 단일 가이드 RNA에 의해 작용할 수 있다. 또한, 티민 (thymine)이 풍부한 PAM (protospacer-adjacent motif) 서열을 인식하고 DNA의 이중 사슬을 잘라 점착종단 (cohesive end; cohesive double-strand break)을 생성한다.

예컨대, 상기 Cpf1 단백질은 캔디다투스 (*Candidatus*) 속, 라치노스피라 (*Lachnospira*) 속, 뷰티리비브리오 (*Butyrivibrio*) 속, 페레그리니박테리아 (*Peregrinibacteria*), 액시도미노코쿠스 (*Acidaminococcus*) 속, 포르파이로모나스 (*Porphyromonas*) 속, 프레보텔라 (*Prevotella*) 속, 프란시셀라 (*Francisella*) 속, 캔디다투스 메타노플라스마 (*Candidatus Methanoplasma*), 또는 유박테리움 (*Eubacterium*) 속 유래의 것일 수 있고, 예컨대, *Parcubacteria bacterium* (GWC2011_GWC2_44_17), *Lachnospiraceae bacterium* (MC2017), *Butyrivibrio proteoclasticus*, *Peregrinibacteria bacterium* (GW2011_GWA_33_10), *Acidaminococcus* sp. (BV3L6), *Porphyromonas macacae*, *Lachnospiraceae bacterium* (ND2006), *Porphyromonas crevioricanis*, *Prevotella disiens*, *Moraxella bovoculi* (237), *Smiihella* sp. (SC_K08D17), *Leptospira inadai*, *Lachnospiraceae bacterium* (MA2020), *Francisella novicida* (U112), *Candidatus Methanoplasma termitum*, *Candidatus Paceibacter*, *Eubacterium eligens* 등의 미생물 유래의 것일 수 있으나, 이에 제한되는 것은 아니다.

상기 표적 특이적 뉴클레아제는 미생물에서 분리된 것 또는 재조합적 방법 또는 합성적 방법 등과 같이 인위적 또는 비자연적 생산된 것(non-naturally occurring)일 수 있다. 일 예에서, 상기 표적 특이적 뉴클레아제 (예컨대, Cas9, Cpf1, 등)은 재조합 DNA에 의하여 만들어진 재조합 단백질일 수 있다. 재조합 DNA(Recombinant DNA; rDNA)는 다양한 유기체로부터 얻어진 이종 또는 동종 유전 물질을 포함하기 위하여 분자 클로닝과 같은 유전자 재조합 방법에 의하여 인공적으로 만들어진 DNA 분자를 의미한다. 예컨대, 재조합 DNA를 적절한 유기체에서 발현시켜 표적 특이적 뉴클레아제를 생산 (*in vivo* 또는 *in vitro*)하는 경우, 재조합 DNA는 제조하고자 하는 단백질을 암호화 하는 코돈들 중에서 상기 유기체에 발현하기에 최적화된 코돈을 선택하여 재구성된 뉴클레오타이드 서열을 갖는 것일 수 있다.

본 명세서에서 사용된 상기 표적특이적 뉴클레아제는 변이된(불활성화된) 형태의 변이 표적특이적 뉴클레아제일 수 있다. 상기 변이(불활성화) 표적특이적 뉴클레아제는 DNA 이중 가닥을 모두 절단하는 엔도뉴클레아제 활성을 상실하도록 변이된 것을 의미할 수 있으며, 예컨대, DNA 이중 가닥을 모두 절단하는 엔도뉴클레아제 활성을 상실하고 니카아제 활성을 갖도록 변이된 변이 표적특이적 뉴클레아제 및 DNA 이중 가닥을

모두 절단하는 엔도뉴클레아제 활성과 니카아제 활성을 모두 상실하도록
 변이된 변이 표적특이적 뉴클레아제 중에서 선택된 1종 이상일 수 있다.
 상기 변이 표적특이적 뉴클레아제가 니카아제 활성을 갖는 것인 경우, 상기
 디아미나제에 의한 염기 변환(예컨대, 아데노신이 이노신으로 변환)과 동시
 5 또는 순서와 무관하게 순차적으로, 상기 염기 변환이 일어난 가닥 또는 그
 반대 가닥 (예컨대, 염기 변환이 일어난 가닥의 반대 가닥)에서 단일 가닥
 절단(nick)을 도입시킬 수 있다 (예컨대, PAM이 위치하는 가닥의
 반대가닥에서, PAM 서열의 5' 말단 방향으로 3번째 뉴클레오타이드와 4번째
 뉴클레오타이드 사이에 해당하는 위치에 nick이 도입됨). 이와 같은
 10 표적특이적 뉴클레아제의 변이 (예컨대, 아미노산 치환 등)는 적어도
 뉴클레아제의 촉매 활성 도메인 (예컨대, Cas9의 경우 RuvC 촉매
 도메인)에서 일어나는 것일 수 있다. 일 예에서, 상기 표적특이적
 뉴클레아제가 스트렙토코커스 피요젠스 유래 Cas9 단백질 (SwissProt
 Accession number Q99ZW2(NP_269215.1))인 경우, 상기 변이는 촉매 활성을
 15 갖는 아스파르트산 잔기 (catalytic aspartate residue; 10번째 위치의
 아스파르트산 (D10) 등), 762번째 위치의 글루탐산 (E762), 840번째 위치의
 히스티딘 (H840), 854번째 위치의 아스파라긴 (N854), 863번째 위치의
 아스파라긴 (N863), 986번째 위치의 아스파르트산 (D986), 539번째 위치의
 페닐알라닌(F539), 763번째 위치의 메티오닌 (M763), 890번째 위치의
 20 라이신 (K890) 등으로 이루어진 군에서 선택된 하나 이상 임의의 다른
 아미노산으로 치환된 돌연변이를 포함할 수 있다. 이 때, 치환되는 임의의
 다른 아미노산은 알라닌 (alanine)일 수 있지만, 이에 제한되지 않는다.

다른 예에서, 상기 변이된 표적특이적 뉴클레아제는
 특이도(정확도)이 개선되도록 변이가 일어난 것일 수 있다. 일 예에서,
 25 상기 변이된 표적특이적 뉴클레아제는 스트렙토코커스 피요젠스
 (*Streptococcus pyogenes*) 유래의 Cas9 단백질(예컨대, SwissProt
 Accession number Q99ZW2(NP_269215.1))의 F539, M763, 및 K890로 이루어진
 군에서 선택된 하나 이상의 아미노산 잔기가 원래 아미노산과 상이한
 아미노산으로 치환 또는 결실된 변이를 갖는 변이체 (보다 구체적으로,
 30 F539S, M763I, 및 K890N로 이루어진 군에서 선택된 하나 이상의 변이를
 포함하는 변이체, 예컨대, F539S, M763I, 및 K890N 변이를 모두 갖는
 변이체 (Sniper-Cas9))일 수 있다.

다른 예에서, 상기 변이된 (불활성화된) 표적특이적 뉴클레아제는

스트렙토코커스 피요젠스 (*Streptococcus pyogenes*) 유래의 Cas9 단백질(예컨대, SwissProt Accession number Q99ZW2(NP_269215.1))의 D10이 원래와 다른 아미노산 (예컨대, 알라닌(A))으로 치환 변이되어 니케이즈 활성을 갖고, 여기에 더하여, F539, M763, 및 K890로 이루어진 군에서
 5 선택된 하나 이상의 아미노산 잔기가 원래와 상이한 아미노산으로 치환 변이(예컨대, F539S, M763I, K890N)되어 특이도 (정확도)가 개선된 것일 수 있다.

다른 예에서, 상기 변이 표적특이적 뉴클레아제는 야생형 Cas9 단백질과 상이한 PAM 서열을 인식하도록 변이된 것일 수 있다. 예컨대,
 10 상기 변이 표적특이적 뉴클레아제는 스트렙토코커스 피요젠스 유래 Cas9 단백질의 1135번째 위치의 아스파르트산 (D1135), 1335번째 위치의 아르기닌 (R1335), 및 1337번째 위치의 트레오닌 (T1337) 중 하나 이상, 예컨대 3개 모두가 다른 아미노산으로 치환되어, 야생형 Cas9의 PAM 서열 (NGG)와 상이한 NGA (N은 A, T, G, 및 C 중에서 선택된 임의의 염기임)을
 15 인식하도록 변이된 것일 수 있다.

일 예에서, 상기 변이 표적특이적 뉴클레아제는 스트렙토코커스 피요젠스 유래 Cas9 단백질의 아미노산 서열 중,

- (1) D10, H840, 또는 이들의 조합 (D10 + H840);
- (2) D10와, F539, M763, 및 K890로 이루어진 군에서 선택된 하나
 20 이상의 조합 (예컨대, D10 + F539 + M763 + K890);
- (3) D1135, R1335, T1337, 또는 이들의 조합 (예컨대, D1135 + R1335 + T1337); 또는
- (5) (1) 내지 (3) 중에서 선택된 둘 이상의 조합
 에서 아미노산 치환이 일어난 것일 수 있다.

본 명세서에 사용된 바로서, 상기 '다른 아미노산'은, 알라닌, 이소류신, 류신, 메티오닌, 페닐알라닌, 프롤린, 트립토판, 발린, 아스파라긴산, 시스테인, 글루타민, 글리신, 세린, 트레오닌, 티로신, 아스파르트산, 글루탐산, 아르기닌, 히스티딘, 라이신, 상기 아미노산들의 공지된 모든 변형체 중에서, 야생형 단백질이 원래 변이 위치에 갖는
 25 아미노산을 제외한 아미노산들 중에서 선택된 아미노산을 의미한다. 일 예에서, 상기 '다른 아미노산'은 알라닌, 발린, 글루타민, 또는 아르기닌일 수 있다.

일 예에서, 상기 변이 표적특이적 뉴클레아제는 엔도뉴클레아제

활성을 상실(예컨대, 니카아제 활성을 갖거나, 엔도뉴클레아제 활성 및 니카아제 활성을 모두 상실)한 변형 Cas9 단백질, 또는 야생형 Cas9과 상이한 PAM 서열을 인식하는 것일 수 있다. 예컨대, 상기 변형 Cas9 단백질은, 스트렙토코커스 피요제네스 (*Streptococcus pyogenes*) 유래의

5 Cas9 단백질에 있어서,

(1) D10 또는 H840 위치에 돌연변이 (예컨대, 다른 아미노산으로의 치환)가 도입되어 이중 가닥 절단 엔도뉴클레아제 활성이 상실되고 니카아제 활성을 갖는 변형 Cas9, 또는 스트렙토코커스 피요제네스 (*Streptococcus pyogenes*) 유래의 Cas9 단백질에 D10 및 H840 위치에 모두

10 돌연변이 (예컨대, 다른 아미노산으로의 치환)가 도입되어 이중 가닥 절단 엔도뉴클레아제 활성 및 니카아제 활성을 모두 상실한 변형 Cas9 단백질;

(2) D10 위치에 더하여 F539, M763, 및 K890로 이루어진 군에서 선택된 하나 이상의 위치에 돌연변이 (예컨대, 다른 아미노산으로의 치환)가 도입되어 이중 가닥 절단 엔도뉴클레아제 활성이 상실되고

15 니카아제 활성을 가지면서 특이도가 향상된 변형 Cas9 단백질;

(3) D1135, R1335 및 T1337 중에서 하나 이상 또는 이들 모두에 돌연변이(예컨대, 다른 아미노산으로의 치환)가 도입되어 야생형과 상이한 PAM 서열을 인식하는 변형 Cas9 단백질; 또는

(3) (1) 내지 (3) 중 2 이상이 도입된 변형 Cas9 단백질

20 일 수 있다.

예컨대, 상기 Cas9 단백질의 D10 위치에서의 돌연변이는 D10A 돌연변이 (Cas9 단백질의 아미노산 중 10번째 아미노산인 D가 A로 치환된 돌연변이를 의미함; 이하, Cas9에 도입된 돌연변이는 동일한 방법으로 표기됨)일 수 있고, 상기 H840 위치에서의 돌연변이는 H840A 돌연변이일 수

25 있으며, D1135, R1335, 및 T1337 위치에서의 돌연변이는 각각 D1135V, R1335Q, 및 T1337R일 수 있고, F539S, M763I, 및 K890N 위치에서의 돌연변이는 각각 F539S, M763I, 및 K890N일 수 있다.

상기 뉴클레아제는 미생물에서 분리된 것 또는 재조합적 방법 또는 합성적 방법 등과 같이 인위적 또는 비자연적 생산된 것(non-naturally occurring)일 수 있다. 일 예에서, 상기 뉴클레아제는 재조합 DNA에

30 의하여 만들어진 재조합 단백질일 수 있다. 재조합 DNA(Recombinant DNA; rDNA)는 다양한 유기체로부터 얻어진 이중 또는 동종 유전 물질을 포함하기 위하여 분자 클로닝과 같은 유전자 재조합 방법에 의하여 인공적으로

만들어진 DNA 분자를 의미한다. 예컨대, 재조합 DNA를 적절한 유기체에서 발현시켜 단백질을 생산 (*in vivo* 또는 *in vitro*)하는 경우, 재조합 DNA는 제조하고자 하는 단백질을 암호화 하는 코돈들 중에서 상기 유기체에 발현하기에 최적화된 코돈을 선택하여 재구성된 뉴클레오타이드 서열을
5 갖는 것일 수 있다.

상기 뉴클레아제는 단백질, 이를 암호화하는 핵산 분자 (DNA 또는 mRNA), 가이드 RNA와 결합된 리보핵산 단백질, 상기 리보핵산 단백질을 암호화하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 재조합 벡터의 형태로 사용될 수 있다.

10 상기 뉴클레아제 또는 이를 코딩하는 핵산 분자는 핵 내로 전달, 작용, 및/또는 발현될 수 있는 형태일 수 있다.

상기 뉴클레아제는 세포 내로 도입되기에 용이한 형태일 수 있다. 일 예로, 상기 뉴클레아제는 세포 침투 펩타이드 및/또는 단백질 전달 도메인 (protein transduction domain)과 연결될 수 있다. 상기 단백질
15 전달 도메인은 폴리-아르기닌 또는 HIV 유래의 TAT 단백질일 수 있으나, 이에 제한되지 않는다. 세포 침투 펩타이드 또는 단백질 전달 도메인은 상기 기술된 예 외에도 다양한 종류가 당업계에 공지되어 있으므로, 당업자는 상기 예에 제한되지 않고 다양한 예를 적용할 수 있다.

또한, 상기 뉴클레아제 또는 암호화하는 핵산 분자는 핵 위치 신호 (nuclear localization signal, NLS) 서열 또는 이를 암호화하는 서열을
20 추가로 포함할 수 있다. 따라서, 상기 뉴클레아제를 암호화하는 핵산 분자를 포함하는 발현 카세트는 상기 뉴클레아제를 발현시키기 위한 프로모터 서열 등의 조절 서열, 또는 여기에 더하여, NLS 서열 (예컨대, SV40 NLS 등)을 포함할 수 있다. 상기 NLS 서열은 당업계에 잘 알려져
25 있다.

상기 뉴클레아제 또는 이를 암호화하는 핵산 분자는 분리 및/또는 정제를 위한 태그 또는 상기 태그를 암호화하는 핵산 서열과 연결될 수 있다. 일 예로, 상기 태그는 His 태그, Flag 태그, S 태그 등과 같은 작은 펩타이드 태그, GST (Glutathione S-transferase) 태그, MBP (Maltose
30 binding protein) 태그 등으로 이루어진 군에서 적절하게 선택될 수 있으나, 이에 제한되지 않는다.

본 발명에서, 용어 "가이드 RNA (guide RNA)"는 표적 DNA 특이적인 RNA (예컨대, DNA의 표적 부위와 혼성화 가능한 RNA)를 의미하며, Cas

단백질, Cpf1 등과 같은 뉴클레오타이드와 결합하여 표적 DNA로 인도하는 역할을 한다.

상기 가이드 RNA는 복합체를 형성할 뉴클레아제의 종류 및/또는 그 유래 미생물에 따라서 적절히 선택될 수 있다. 예컨대, 상기 가이드 RNA는 DNA 표적 부위와 혼성화 가능한 부위를 포함하는 CRISPR RNA (crRNA); Cas 단백질, Cpf1 등과 같은 엔도뉴클레오타이드와 상호작용하는 부위를 포함하는 *trans*-activating crRNA (tracrRNA); 및

상기 crRNA 및 tracrRNA의 주요 부위 (예컨대, crRNA의 혼성화 부위 및 tracrRNA의 상호작용 부위)가 융합된 형태의 단일 가이드 RNA (single guide RNA; sgRNA)

로 이루어진 군에서 선택된 1종 이상일 수 있으며,

구체적으로 CRISPR RNA (crRNA) 및 *trans*-activating crRNA (tracrRNA)를 포함하는 이중 RNA (dual RNA), 또는 crRNA 및 tracrRNA의 주요 부위를 포함하는 단일 가이드 RNA (sgRNA)일 수 있다.

상기 sgRNA는 표적 DNA 내 서열과 상보적인 서열을 가지는 부분 (이를 Spacer region, Target DNA recognition sequence, base pairing region 등으로도 명명함) 및 Cas 단백질 결합을 위한 hairpin 구조를 포함할 수 있다. 보다 구체적으로, 표적 DNA 내 서열과 상보적인 서열을 가지는 부분, Cas 단백질 결합을 위한 hairpin 구조 및 Terminator 서열을 포함할 수 있다. 상기 기술된 구조는 5'에서 3' 순으로 순차적으로 존재하는 것일 수 있으나, 이에 제한되는 것은 아니다. 상기 가이드 RNA가 crRNA 및 tracrRNA의 주요 부분 및 표적 DNA의 상보적인 부분을 포함하는 경우라면 어떠한 형태의 가이드 RNA도 본 발명에서 사용될 수 있다.

예컨대, Cas9 단백질을 타겟 유전자 교정을 위하여 두 개의 가이드 RNA, 즉, 표적 유전자의 표적 서열 부위와 혼성화 가능한 뉴클레오타이드 서열을 갖는 CRISPR RNA (crRNA)와 Cas9 단백질과 상호작용하는 *trans*-activating crRNA (tracrRNA; Cas9 단백질과 상호작용함)를 필요로 하며, 이들 crRNA와 tracrRNA는 서로 결합된 이중 가닥 crRNA:tracrRNA 복합체 형태, 또는 링커를 통하여 연결되어 단일 가이드 RNA (single guide RNA; sgRNA) 형태로 사용될 수 있다. 일 예에서, *Streptococcus pyogenes* 유래의 Cas9 단백질을 사용하는 경우, sgRNA는 상기 Cas9의 crRNA의 혼성화 가능한 뉴클레오타이드 서열을 적어도 포함하는 crRNA 일부 또는 전부와 상기 Cas9의 tracrRNA의 Cas9 단백질과 상호작용하는 부위를 적어도

포함하는 tracrRNA 일부 또는 전부가 뉴클레오타이드 링커를 통하여 헤어핀 구조 (stem-loop 구조)를 형성하는 것일 수 있다 (이 때 뉴클레오타이드 링커가 루프 구조에 해당할 수 있음).

5 상기 가이드 RNA, 구체적으로 crRNA 또는 sgRNA는 표적 DNA 내 서열과 상보적인 서열을 포함하며, crRNA 또는 sgRNA의 업스트림 부위, 구체적으로 sgRNA 또는 dualRNA의 crRNA의 5' 말단에 하나 이상, 예컨대, 1-10개, 1-5개, 또는 1-3개의 추가의 뉴클레오타이드를 포함할 수 있다. 상기 추가의 뉴클레오타이드는 구아닌 (guanine, G)일 수 있으나, 이에 제한되는 것은 아니다.

10 다른 예에서, 상기 뉴클레아제가 Cpf1인 경우, 상기 가이드 RNA는 crRNA를 포함하는 것일 수 있으며, 복합체를 형성할 Cpf1 단백질 종류 및/또는 그 유래 미생물에 따라서 적절히 선택될 수 있다.

15 상기 가이드 RNA의 구체적 서열은 뉴클레아제 (Cas9 단백질 또는 Cpf1)의 종류 (즉, 유래 미생물)에 따라서 적절히 선택할 수 있으며, 이는 이 발명이 속하는 기술 분야의 통상의 지식을 가진 자가 용이하게 알 수 있는 사항이다.

일 예에서, 표적특이적 뉴클레아제로서 *Streptococcus pyogenes* 유래의 Cas9 단백질을 사용하는 경우, crRNA는 다음의 일반식 1로 표현될 수 있다:

20 $5'-(N_{\text{cas9}})_1-(\text{GUUUUAGAGCUA})-(X_{\text{cas9}})_m-3'$ (일반식 1)

상기 일반식 1에서,

25 N_{cas9} 는 표적화 서열, 즉 표적 유전자(target gene)의 표적 부위(target site)의 서열에 따라서 결정되는 부위 (즉, 표적 부위의 서열과 혼성화 가능한 서열임)이며, 1은 상기 표적화 서열에 포함된 뉴클레오타이드 수를 나타내는 것으로 17 내지 23, 17 내지 22, 18 내지 23, 또는 18 내지 22의 정수, 예컨대 17, 18, 19, 20, 21, 또는 22일 수 있고;

상기 표적 서열의 3' 방향으로 인접하여 위치하는 연속하는 12개의 뉴클레오타이드(GUUUUAGAGCUA)를 포함하는 부위는 crRNA의 필수적 부분이고,

30 X_{cas9} 는 crRNA의 3' 말단쪽에 위치하는 (즉, 상기 crRNA의 필수적 부분의 3' 방향으로 인접하여 위치하는) m개의 뉴클레오타이드를 포함하는 부위로, m은 8 내지 12의 정수, 예컨대 11일 수 있으며, 상기 m개의 뉴클레오타이드들은 서로 같거나 다를 수 있으며, 각각 독립적으로 A, U, C 및 G로 이루어진 군에서 선택될 수 있다.

일 예에서, 상기 X_{cas9} 는 UGCUGUUUUG를 포함할 수 있으나 이에 제한되지 않는다.

또한, 상기 tracrRNA는 다음의 일반식 2로 표현될 수 있다:

5'-(Y_{cas9})_p-

5 (UAGCAAGUUAUUAAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC)-3'

(일반식 2)

상기 일반식 2에서,

60개의

뉴클레오타이드

(UAGCAAGUUAUUAAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC)로

10 표시된 부위는 tracrRNA의 필수적 부분이고,

Y_{cas9} 는 상기 tracrRNA의 필수적 부분의 5' 말단에 인접하여 위치하는 p개의 뉴클레오타이드를 포함하는 부위로, p는 6 내지 20의 정수, 예컨대 8 내지 19의 정수일 수 있으며, 상기 p개의 뉴클레오타이드들은 서로 같거나 다를 수 있고, A, U, C 및 G로 이루어진 군에서 각각 독립적으로 선택될 수
15 있다.

또한, sgRNA는 상기 crRNA의 표적화 서열과 필수적 부위를 포함하는 crRNA 부분과 상기 tracrRNA의 필수적 부분 (60개 뉴클레오타이드)를 포함하는 tracrRNA 부분이 올리고뉴클레오타이드 링커를 통하여 헤어핀 구조 (stem-loop 구조)를 형성하는 것일 수 있다 (이 때,
20 올리고뉴클레오타이드 링커가 루프 구조에 해당함). 보다 구체적으로, 상기 sgRNA는 crRNA의 표적화 서열과 필수적 부분을 포함하는 crRNA 부분과 tracrRNA의 필수적 부분을 포함하는 tracrRNA 부분이 서로 결합된 이중 가닥 RNA 분자에서, crRNA 부위의 3' 말단과 tracrRNA 부위의 5' 말단이 올리고뉴클레오타이드 링커를 통하여 연결된 헤어핀 구조를 갖는 것일 수
25 있다.

일 예에서, sgRNA는 다음의 일반식 3으로 표현될 수 있다:

5'-(N_{cas9})₁-(GUUUUAGAGCUA)-(올리고뉴클레오타이드 링커)-

(UAGCAAGUUAUUAAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC)-3'

(일반식 3)

30 상기 일반식 3에서, (N_{cas9})₁는 표적화 서열로서 앞서 일반식 1에서 설명한 바와 같다.

상기 sgRNA에 포함되는 올리고뉴클레오타이드 링커는 3 내지 5개, 예컨대 4개의 뉴클레오타이드를 포함하는 것일 수 있으며, 상기

뉴클레오타이드들은 서로 같거나 다를 수 있고, A, U, C 및 G로 이루어진 군에서 각각 독립적으로 선택될 수 있다.

상기 crRNA 또는 sgRNA는 5' 말단 (즉, crRNA의 타겟팅 서열 부위의 5' 말단)에 1 내지 3개의 구아닌(G)을 추가로 포함할 수 있다.

- 5 상기 tracrRNA 또는 sgRNA는 tracrRNA의 필수적 부분(60nt)의 3' 말단에 5개 내지 7개의 우라실 (U)을 포함하는 종결부위를 추가로 포함할 수 있다.

- 10 상기 가이드 RNA의 표적 서열은 표적 DNA 상의 PAM (Protospacer Adjacent Motif 서열(*S. pyogenes* Cas9의 경우, 5'-NGG-3' (N은 A, T, G, 또는 C임))의 5'에 인접하여 위치하는 약 17개 내지 약 23개 또는 약 18개 내지 약 22개, 예컨대 20개의 연속하는 핵산 서열일 수 있다.

- 15 상기 가이드 RNA의 표적 서열과 혼성화 가능한 가이드 RNA의 표적화 서열은 상기 표적 서열이 위치하는 DNA 가닥 (즉, PAM 서열(5'-NGG-3' (N은 A, T, G, 또는 C임)이 위치하는 DNA 가닥)의 상보적인 가닥의 뉴클레오타이드 서열과 50% 이상, 60% 이상, 70% 이상, 80% 이상, 90% 이상, 95% 이상, 99% 이상, 또는 100%의 서열 상보성을 갖는 뉴클레오타이드 서열을 의미하는 것으로, 상기 상보적 가닥의 뉴클레오타이드 서열과 상보적 결합이 가능하다.

- 20 본 명세서에서, 표적 부위의 핵산 서열은 표적 유전자의 해당 유전자 부위의 두 개의 DNA 가닥 중 PAM 서열이 위치하는 가닥의 핵산 서열로 표시된다. 이 때, 실제로 가이드 RNA가 결합하는 DNA 가닥은 PAM 서열이 위치하는 가닥의 상보적 가닥이므로, 상기 가이드 RNA에 포함된 표적화 서열은, RNA 특성상 T를 U로 변경하는 것을 제외하고, 표적 부위의 서열과 동일한 핵산 서열을 갖게 된다. 따라서, 본 명세서에서, 가이드 RNA의
25 표적화 서열과 표적 부위의 서열 (또는 절단 부위의 서열)은 T와 U가 상호 변경되는 것을 제외하고 동일한 핵산 서열로 표시된다.

상기 가이드 RNA는 RNA 형태로 사용 (또는 상기 조성물에 포함)되거나, 이를 암호화하는 DNA를 포함하는 플라스미드 형태로 사용 (또는 상기 조성물에 포함)될 수 있다.

- 30 본 발명에서 용어 "표적 위치 (on-target site)"란, 상기 표적 특이적 뉴클레아제를 이용하여 변이(절단, 삽입, 및/또는 결실)를 도입하고자 하는 위치를 의미하며, 그 목적에 따라 임의로 선택될 수 있는 것으로 특정 유전자의 코딩 서열 내부에 존재할 수 있을 뿐만 아니라,

단백질을 생성하지 않는 비-코딩 DNA 서열에 존재할 수도 있다.

상기 표적 특이적 뉴클레아제는 서열 특이성 (specificity)을 가지므로 표적 위치에 작용하는 것이나, 표적 서열에 따라 비표적 위치 (off-target site)에 작용하는 부작용이 발생할 수도 있다.

- 5 본 명세서에서, 비표적 위치 (off-target site)라 함은 표적 특이적 뉴클레아제의 표적 서열과 동일하지 않은 서열을 갖지만 상기 표적 특이적 뉴클레아제가 활성을 가지는 위치를 말한다. 즉, 표적 위치 이외의, 표적 특이적 뉴클레아제에 의해 절단되는 위치를 말한다. 일 예에서, 상기 비표적 위치는 특정 표적 특이적 뉴클레아제에 대한 실제 비표적 위치뿐만 아니라 비표적 위치가 될 가능성이 있는 위치까지 포함하는 개념으로
10 사용될 수 있다. 상기 비표적 위치는 이에 제한되는 것은 아니나, 시험관 내 (*in vitro*)에서 표적 특이적 뉴클레아제에 의해 절단되는 표적 위치 이외의 모든 위치일 수 있다.

- 유전자 가위가 표적 위치 이외의 위치에서도 활성을 가지는 것은
15 다양한 원인에 의해 야기될 수 있다. 예컨대, 표적 위치에 대하여 설계된 표적 서열과 뉴클레오티드 불일치 (mismatch)를 가지는, 표적 위치와 서열 상동성이 높은 비표적 서열의 경우 유전자 가위가 작동할 가능성이 있다. 상기 비표적 위치는 이에 제한되는 것은 아니나, 표적 서열과 1 이상의 뉴클레오티드 불일치 (mismatch)를 가지는 위치일 수 있다.

- 20 이는 유전체 내에서 원치 않는 유전자의 돌연변이를 야기할 수 있어 상기 표적 특이적 뉴클레아제를 사용하는데 심각한 문제가 될 수 있다. 이에, 표적 특이적 뉴클레아제의 표적 위치에서의 활성 못지 않게 비표적 위치를 정확히 검출하여 분석하는 과정 또한 매우 중요할 수 있으며, 이는 비표적 효과 없이 표적 위치에만 특이적으로 작동하는 표적 특이적
25 뉴클레아제를 개발하는데 유용하게 사용될 수 있을 것이다.

- 본 발명의 목적상 상기 뉴클레아제는 생체 내 (*in vivo*) 및 시험관 내 (*in vitro*)에서 뉴클레아제 활성을 가질 수 있으므로, 시험관 내에서 유전체 DNA의 비표적 위치를 검출하는데 사용될 수 있으며, 이를 생체 내에서 적용하였을 때 상기 검출된 비표적 위치와 동일한 위치에도 활성을
30 가질 것을 예상할 수 있다.

 상기 (b) 단계는 상기 (a) 단계를 통하여 한 가닥이 절단된 DNA에 단일가닥 특이적 엔도뉴클레아제를 처리하여, 나머지 DNA 가닥을 절단하는 단계이다.

상기 단일가닥 특이적 엔도뉴클레아제는 단일가닥 DNA를 특이적으로
 절단하는 활성을 갖는 모든 엔도뉴클레아제를 총칭하는 것일 수 있다. 일
 예에서, 상기 단일가닥 특이적 엔도뉴클레아제는 엔도뉴클레아제 V, 또는
 엔도뉴클레아제 VIII와 알킬아데닌 DNA 글라이코실라아제 (Alkyladenine
 5 DNA Glycosylase; hAAG)와의 조합 등으로 이루어진 군에서 선택된 1종
 이상일 수 있다. 상기 엔도뉴클레아제 V는 원핵 세포 또는 진핵 세포,
 예컨대, 균류, 진핵 동물 (예컨대, 어류, 양서류, 파충류, 조류, 포유류
 등), 또는 진핵 식물 유래의 것일 수 있으며, 예컨대, 대장균 유래의
 endonuclease V (NP_418426.2; *E. coli* ENDOV) 등으로 이루어진 군에서
 10 선택된 1종 이상일 수 있다. 상기 단일가닥 특이적 엔도뉴클레아제 (예컨대,
 엔도뉴클레아제 V)는 아데노신이 이노신으로 변환된 가닥을 절단하는 것일
 수 있으며, 상기 아데노신이 이노신으로 변환된 변이 위치에서 5' 방향
 및/또는 3' 방향으로 5nt 이내, 4nt 이내, 3 nt 이내, 2nt 이내, 또는 1nt
 이내에 위치하는 뉴클레오타이드의 3' 말단을 절단하는 것이 수 있다.
 15 상기 엔도뉴클레아제 VIII는 아데닌 디아미나아제에 의하여 아데닌이
 이노신으로 변환된 뉴클레오타이드를 제거하는 역할을 하는 것으로,
 이노신을 이중 가닥 DNA로부터 제거하는 N-glycosylase 활성과 상기 이노신
 제거로 발생한 apurinic site (AP site)의 3' 및 5' 말단을 절단하는 AP-
 lyase 활성을 모두 갖는 모든 효소들 중에서 1종 이상 선택될 수 있다.
 20 예컨대, 상기 엔도뉴클레아제 VIII는 *E. coli* 엔도뉴클레아제 VIII (예컨대,
 GenBank Accession No. NP_415242.1 등), 인간 엔도뉴클레아제 VIII
 (예컨대, GenBank Accession Nos. BAC06476.1, NP_001339449.1,
 NP_001243481.1, NP_078884.2, NP_001339448.1 등), 마우스 엔도뉴클레아제
 VIII (예컨대, GenBank Accession Nos. BAC06477.1, NP_082623.1 등),
 25 *Escherichia coli* 엔도뉴클레아제 VIII (예컨대, GenBank Accession Nos.
 OBZ49008.1, OBZ43214.1, OBZ42025.1, ANJ41661.1, KYL40995.1, KMV55034.1,
 KMV53379.1, KMV50038.1, KMV40847.1, AQW72152.1 등) 등으로 이루어진
 군에서 선택된 1종 이상일 수 있으나, 이에 제한되는 것은 아니다. 상기
 알킬아데닌 DNA 글라이코실라아제는 N-glycosidic bond의 가수분해를
 30 촉매하여 alkylation-damaged 퓨린 염기를 제거하여 DNA를 repair하는
 효소로, 특히 3-methyladenine, 7-methyladenine, 7-methylguanine, 1N-
 ethenoadenine and hypoxanthine 등을 제거한다. 사용 가능한 알킬아데닌

DNA 클라이코실라아제는 인간 알킬아데닌 DNA 클라이코실라아제 (hAAG; GenBank Accession no. NP_001015052.1 (코딩 핵산 서열 NM_001015052.2), NP_001015054.1 (코딩 핵산 서열 NM_001015054.2), NP_002425.2 (코딩 핵산서열 NM_002434.3) 등으로 이루어진 군에서 선택된 1종 이상일 수 있으나 이에 제한되는 것은 아니다.

상기 (c) 단계는 상기 (b) 단계를 통해 절단된 DNA를 이용해 전체 유전체 시퀀싱 (whole genome sequencing, WGS)을 수행하는 단계로서, 표적 위치의 서열과 상동성을 가지는 서열을 찾아 비표적 위치일 것으로 예측하는 간접적인 방법과 달리 전체 유전체 수준에서 실질적으로 표적 특이적 뉴클레아제에 의해 절단되는 비표적 위치를 검출하기 위해 수행되는 것이다.

본 발명에서 용어 "전체 유전체 시퀀싱 (whole genome sequencing; WGS)"은 차세대 시퀀싱 (next generation sequencing)에 의한 전장 유전체 시퀀싱을 10 X, 20 X, 40 X 형식으로 여러 배수로 유전체를 읽는 방법을 의미한다. "차세대 시퀀싱"은 칩 (Chip) 기반 및 PCR 기반 페어드엔드 (paired end) 형식으로 전장 유전체를 조각내고, 상기 조각을 화학적인 반응 (hybridization)에 기초하여 초고속으로 시퀀싱을 수행하는 기술을 의미한다.

상기 (d) 단계는 상기 전체 유전체 시퀀싱으로 획득한 염기서열 데이터 (sequence read)에서 DNA가 절단된 위치를 결정하는 단계로서, 시퀀싱 데이터를 분석하여 표적 특이적 뉴클레아제의 표적 위치 및 비표적 위치를 간편하게 검출할 수 있다. 상기 염기서열 데이터로부터 DNA가 절단된 특정 위치를 결정하는 것은 다양한 접근 방법으로 수행될 수 있으며, 본 발명에서는 상기 위치를 결정하기 위한 여러 가지의 합리적인 방법들을 제공한다. 그러나 이는 본 발명의 기술적 사상에 포함되는 예시에 불과하며, 본 발명의 범위가 이들 방법에 의해 제한되는 것은 아니다.

예컨대, 상기 절단된 위치를 결정하기 위한 일례로서, 전체 유전체 시퀀싱을 통해 획득한 염기서열 데이터를 분석 프로그램 (예를 들어, BWA/GATK 또는 ISAAC 등)을 이용하여 유전체 상의 위치에 따라 정렬하였을 경우, 5' 말단이 수직 정렬된 위치가 DNA가 절단된 위치를 의미할 수 있다. 본 명세서에 사용된 바로서, 상기 용어 "수직 정렬"이란, BWA/GATK 또는 ISAAC 등의 프로그램으로 전체 유전체 시퀀싱 결과를 분석할 때, 인접한 왓슨 가닥 (Watson strand)과 크릭 가닥 (Crick strand) 각각에 대해, 2 개

이상의 염기서열 데이터의 5' 말단이 유전체 상의 동일한 위치 (nucleotide position)에서 시작되는 배열을 말한다. 이로 인하여, 표적 특이적 뉴클레아제에 의해 잘려 동일한 5' 말단을 갖게 되는 DNA 단편들이 각각 시퀀싱되어 나타나게 된다.

- 5 즉, 표적 특이적 뉴클레아제가 표적 위치 및 비표적 위치에 뉴클레아제 활성을 나타내 상기 위치를 절단하는 경우, 염기서열 데이터를 정렬하게 되면 공통적으로 절단된 부위는 각각 그 위치가 5' 말단으로 시작되므로 수직 정렬되나, 절단되지 않은 부위에는 5' 말단이 존재하지 않으므로 정렬 시 스테거드 (staggered) 방식으로 배열될 수 있다. 따라서,
- 10 수직 정렬된 위치를 표적 특이적 뉴클레아제에 의해 절단된 부위로 볼 수 있으며, 이는 곧 표적 특이적 뉴클레아제의 표적 위치 또는 비표적 위치를 의미하는 것일 수 있다.

- 상기 "정렬"은 표준 염기서열 (reference genome)로 염기서열 데이터를 맵핑한 뒤, 유전체에서 동일 위치를 가지는 염기들을 각 위치에
- 15 맞게 배열하는 것을 의미한다. 따라서, 염기서열 데이터를 상기과 같은 방식으로 정렬할 수 있다면 어떠한 컴퓨터 프로그램도 이용될 수 있으며, 이는 당업계에 이미 알려진 공지 프로그램이거나 또는 목적에 맞게 제작된 프로그램들 중에서 선택될 수 있다. 일 실시예에서는 ISAAC를 이용하여 정렬을 수행하였으나, 이에 제한되는 것은 아니다.

- 20 정렬 결과, 상기 설명한 바와 같은 5' 말단이 수직 정렬된 위치를 찾는 등의 방법을 통해 표적 특이적 뉴클레아제에 의해 DNA가 절단된 위치를 결정할 수 있고, 상기 절단된 위치가 표적 위치 (on-target site)가 아니라면, 비표적 위치 (off-target site)로 판단할 수 있다. 다시 말해, 표적 특이적 뉴클레아제의 표적 위치로 설계한 염기 서열과 동일한 서열은
- 25 표적 위치이고, 상기 염기 서열과 동일하지 않은 서열은 비표적 위치로 볼 수 있다. 이는 상기 기술한 비표적 위치의 정의상 자명한 것이다. 상기 비표적 위치는 특히, 표적 위치의 서열과 상동성을 가지는 서열로 구성된 것일 수 있고, 구체적으로 표적 서열과 1 개 이상의 뉴클레오티드 불일치 (mismatch)를 가지는 서열, 더욱 구체적으로 표적 위치와 1 내지 6 개, 1개
- 30 내지 5개, 1개 내지 4개, 1개 내지 3개, 1개 내지 2개, 또는 1개의 뉴클레오티드 불일치를 가지는 것일 수 있으나, 이에 특별히 제한되는 것은 아니고 표적 특이적 뉴클레아제가 절단할 수 있는 위치라면 본 발명의 범위에 포함될 수 있다. 이때, 상기 표적 위치는 가이드 RNA와 상보적인

15 내지 30 뉴클레오티드 서열일 수 있고, 추가적으로 표적 특이적 뉴클레아제가 인식하는 서열 (예컨대, Cas9의 경우 Cas9이 인식하는 PAM 서열)을 포함할 수 있다.

5 다른 예에서, 5' 말단이 수직 정렬된 위치를 찾는 방법 이외에도, 5' 말단 플롯에서 이중 피크 패턴을 보이는 경우 그 위치가 표적 위치가 아니라면 비표적 위치로 판단할 수 있다. 유전체 DNA 내의 각 위치에 대하여 동일한 염기의 5' 말단을 구성하고 있는 뉴클레오티드 수를 세어 그래프를 그릴 경우, 특정 위치에서 이중 피크 패턴이 나타나게 되는데, 상기 이중 피크는 표적 특이적 뉴클레아제에 의해 절단된 이중 가닥의
10 각각의 가닥에 의해 나타나는 것이기 때문이다.

일 실시예에서, 유전체 DNA를 표적 특이적 뉴클레아제(예컨대, RGEN)으로 절단한 뒤, 전체 유전체 분석 후 이를 ISAAC로 정렬하여 절단된 위치에서는 수직 정렬, 절단되지 않은 위치에서는 스테거드 방식으로 정렬되는 패턴을 확인하였으며, 이를 5' 말단 플롯으로 나타내었을 때 절단
15 위치에서 이중 피크의 독특한 패턴이 나타나는 것을 확인하였다.

나아가 이에 제한되는 것은 아니나, 구체적인 일례로 왓슨 가닥 (Watson strand)과 크릭 가닥 (Crick strand)에 해당하는 염기서열 데이터 (sequence read)가 각각 두 개 이상씩 수직으로 정렬되는 위치를 비표적 위치인 것으로 판단할 수 있고, 또한 20 % 이상의 염기서열 데이터가
20 수직으로 정렬되고, 각각의 왓슨 가닥 및 크릭 가닥에서 동일한 5' 말단을 가진 염기서열 데이터의 수가 10 이상인 위치가 비표적 위치, 즉 절단되는 위치인 것으로 판단할 수 있다.

상기 비표적 위치 확인(검출)은 시험관 내 (*in vitro*)에서 표적 특이적 뉴클레아제를 유전체 DNA에 처리하여 수행될 수 있다. 이에 상기
25 방법을 통해 확인(검출)된 비표적 위치에 대하여 실질적으로 생체 내 (*in vivo*)에서도 비표적 효과가 나타나는지 확인해볼 수 있다. 다만 이는 추가적인 검증 과정에 불과하므로 본 발명의 범위에 필수적으로 수반되는 단계는 아니며, 필요에 따라 추가적으로 수행될 수 있는 단계에 불과하다.

본 명세서에 사용된 바로서, 용어 "비표적 효과 (off-target effect)"는 비표적 위치 (off-target site)와는 구별되는 개념일 수 있다. 즉, 상기 설명한 바와 같이 본 발명에서 비표적 위치라는 개념은 표적 특이적 뉴클레아제가 작동할 수 있는 위치 중 표적 위치가 아닌 위치를
30 의미하는 것으로, 상기 표적 특이적 뉴클레아제에 의해 절단되는 위치를

말하는 것이나, 비표적 효과는 세포 내 비표적 위치에서 표적 특이적 뉴클레아제에 의해 인델 (Insertion and/or deletion)이 나타나는 효과를 의미한다.

5 상기 용어 "인델"은 DNA의 염기 배열에서 일부 염기가 중간에 삽입되거나 (insertion) 및/또는 결실된 (deletion) 변이를 총칭한다. 또한, 표적 특이적 뉴클레아제에 의해 상기 인델이 일어난 비표적 위치를 비표적 인델 위치라고 한다. 결론적으로, 본 명세서의 비표적 위치는 비표적 인델 위치를 포함하는 개념으로 볼 수 있으며, 표적 특이적 뉴클레아제가 활성을 가질 수 있는 가능성이 있는 위치로 족하며, 반드시
10 유전자 가위에 의한 인델이 확인되어야 하는 것은 아니다. 한편, 본 발명에서의 비표적 위치는 비표적 후보 위치 (candidate off-target site)로, 비표적 인델 위치는 검증된 비표적 위치 (validated off-target site)로도 명명될 수 있다.

구체적으로 상기 검증 과정은, 이에 제한되는 것은 아니나, 상기
15 비표적 위치에 대한 표적 특이적 뉴클레아제가 발현된 세포로부터 유전체 DNA를 분리하고, 상기 DNA의 비표적 위치에서 인델을 확인하여 비표적 위치에서의 비표적 효과를 확인하는 것일 수 있다. 이는, T7E1 분석, Cel-I 효소를 이용한 돌연변이 검출 분석 또는 표적화 딥시퀀싱 (targeted deep sequencing) 등 당업계에 공지된 인델 확인 방법을 수행하여 비표적 효과를
20 확인하는 것일 수 있다. 상기 비표적 효과를 확인하는 단계는 비표적 위치에서 인델이 일어났는지를 직접적으로 확인하는 것일 수 있다. 다만, 이러한 생체 내 검증 과정에서 인델이 일어나지 않았다고 하더라도, 이는 검출할 수 있는 수준 이하의 빈도로 인델이 일어날 경우까지 확인한 것은 아니므로 어디까지나 보조적인 수단으로 보아야 한다.

25 상기 기술한 것과 같이 수직 정렬된 위치를 확인하거나, 또는 5' 말단 플롯에서 이중 피크를 확인하는 것만으로도 비표적 위치를 충분히 검출할 수 있고 이는 고도의 재현성을 가지는 것이나, 불균일 절단 패턴 또는 낮은 시퀀싱 깊이 (depth)를 가지는 일부 위치가 누락될 수 있다는 문제가 있다. 이에 본 명세서에서는 염기서열 데이터의 정렬 패턴을
30 기반으로 하여, 각 뉴클레오타이드의 위치 i (즉 유전체 DNA 상의 뉴클레오타이드 위치)에 DNA 절단 점수를 산출하는 수식을 다음과 같이 제공한다:

i 위치에서의 점수 =

$$\sum_{a=1}^5 \frac{C(F_i-1)}{D_i} \times \frac{C(R_{i-4+a}-1)}{D_{i-4+a}} \times (F_i + R_{i-4+a} - 2) \\ + \sum_{a=1}^5 \frac{C(R_{i-1}-1)}{D_{i-1}} \times \frac{C(F_{i-3+a}-1)}{D_{i-3+a}} \times (R_{i-a} + F_{i-3+a} - 2)$$

F_i : i 위치에서 시작하는 정방향 염기서열 데이터의 수

R_i : i 위치에서 시작하는 역방향 염기서열 데이터의 수

D_i : i 위치에서의 시퀀싱 깊이 (depth)

C : 임의의 상수

상기 수식에서 염기서열 데이터의 수는 뉴클레오타이드 리드 수를 의미하고, 시퀀싱 깊이는 특정 위치에서의 시퀀싱 리드수를 의미한다.

- 또한, 상기 수식을 통해 기존의 Digenome-seq에서는 검출되지 않았던
- 5 다수의 추가적인 위치를 검출할 수 있으며, 이를 통해 거짓-양성 위치를 손쉽게 걸러낼 수 있다. 상기 수식에서 C 값은 당업자가 임의의 상수를 적용할 수 있는 것으로 본 발명의 실시예에 의해 제한되는 것은 아니다. 일 예에서, 상기 C 는 1 내지 1000, 1 내지 500, 1 내지 100, 1 내지 50, 1 내지 10, 1 내지 5, 또는 1 내지 3일 수 있으나, 이에 제한되는 것은
- 10 아니다. 특히, 이에 제한되는 것은 아니나, 예컨대, 임의의 위치 (절단된 위치)의 염기서열에 있어서, C 값을 1으로 하여 상기 산출된 점수가 2.5점 이상이거나, 0.1 점 이상이며 On-target 서열과 homology를 갖는 경우 (예컨대, On-target 서열과 10개 이하의 미스매치를 가지고 PAM (5'-NGN-3' 또는 5'-NNG-3')을 포함하는 경우), 상기 임의의 위치(절단된 위치)를
- 15 비표적 위치로 판단할 수 있다. 다만, 상기 점수의 기준은 목적에 따라 당업자에 의해 적절히 조정, 변경될 수 있다.

- 일 예에서, 본 명세서에서 제공되는 Digenome-seq 방법은 복수의 표적 특이적 뉴클레아제 (예컨대, 표적 부위가 상이한 가이드 RNA를 다수 포함하는 표적 특이적 뉴클레아제)를 이용하여 수행될 수도 있으며, 본
- 20 명세서에서는 이를 "복합 Digenome-seq"로 명명한다. 이 경우, 상기 표적 특이적 뉴클레아제는 2 개 이상, 구체적으로 2 내지 100 개의 표적에 대한 표적 특이적 뉴클레아제를 혼합한 것일 수 있으나, 이에 제한되는 것은

아니다.

상기 복합 Digenome-seq의 경우 각각의 표적 특이적 뉴클레아제에 의해 유전체 DNA가 절단되므로 절단 위치가 어느 유전자 가위에 의해 절단되었는지를 확인하는 것이 중요하다. 이는 표적 위치와의 편집 거리 (edit distance)에 따라 비표적 위치를 분류함으로써 달성될 수 있으며, 비표적 위치의 염기 서열이 표적 위치와 상동성을 가진다는 것을 전제로 한다. 이를 통해 각각의 유전자 가위에 대한 표적 및 비표적 위치가 명확하게 구분될 수 있다.

본 발명의 구체적인 일 실시예에서는 특정 위치를 표적으로 하는 RGEN (RNA-guided engineered nuclease)에 대하여, 전체 유전체에서 Digenome-seq를 통해 검출된 비표적 위치 중 표적 위치와의 뉴클레오티드 불일치가 6 개 이하인 상동성 위치가 13,000 개 이하이고, 뉴클레오티드 불일치가 2 개 이하인 상동성 위치를 가지지 않는 경우, 상기 특정 위치를 RGEN의 표적 위치로 선별하는 것이 비표적 효과를 최소화할 수 있음을 확인하였다. 이는 본 발명의 Digenome-seq를 이용하여 표적 위치를 선별하는 바람직한 기준을 확립해가는 과정을 보여주는 일례로서, Digenome-seq를 통해 유전자 가위의 비표적 효과를 최소화 시킬 수 있을 것으로 기대된다.

한편, 표적 위치의 서열과 상동성을 가지는 위치의 수는 뉴클레오티드 불일치 수준이 증가할 수록 Digenome-seq를 통해 적은 비율로 검출되는 것을 확인하였다. 이는 RGEN의 표적 위치를 선별함에 있어서, 표적 서열과 유전체 내에서 상동성을 가지는 뉴클레오티드 서열이 많을 수록, 특히 고도의 상동성을 가지는 뉴클레오티드 서열이 많을 수록 상대적으로 더욱 특이적이기 때문이다. 이를 통해 선별된 RGEN의 표적 위치는 비표적 효과가 최소화될 것일 수 있다.

다른 예에서, 상기 off-target 확인 방법은 CIRCLE-seq (circularization for in vitro reporting of cleavage effects by sequencing; Nature Methods 14, 607-614 (2017) 참조) 및/또는 SITE-seq (Cameron, P. et al. Mapping the genomic landscape of CRISPR-Cas9 cleavage. Nature methods 14, 600-606 (2017) 참조)을 통하여도 수행 가능하다 (도 2a 및 2b 참조).

【발명의 효과】

본 발명에서 제시되는 유전체 서열 분석 기술에 의하여, 고도의 재현성으로 유전체 수준에서 유전자 가위의 비표적 위치를 검출할 수 있어, 표적 특이성이 높은 유전자 가위의 제작 및 이를 위한 연구에 사용될 수 있다.

5

【도면의 간단한 설명】

도 1a는 아데노신 디아미나아제와 Cas9 니케이즈 (D10A) (ABE)와 엔도뉴클레아제 V를 사용하는 유전체 DNA 서열 분석 과정을 모식적으로 보여준다.

10 도 1b는 아데노신 디아미나아제와 Cas9 니케이즈 (D10A) (ABE)와 Alkyladenine DNA Glycosylase (AAG) 및 엔도뉴클레아제 VIII를 사용하는 유전체 DNA 서열 분석 과정을 모식적으로 보여준다.

도 2a는 ABE 및 엔도뉴클레아제 V를 처리한 후 real-time PCR을 통해 DSB(Double Strand Break)의 생성 여부를 확인한 결과를 보여준다.

15 도 2b는 ABE, Alkyladenine DNA Glycosylase (AAG), 및 엔도뉴클레아제 VIII를 처리한 후 real-time PCR을 통해 DSB의 생성 여부를 확인한 결과를 보여준다.

도 2c는 ABE 및 엔도뉴클레아제 V를 처리하거나 ABE, Alkyladenine DNA Glycosylase (AAG), 및 엔도뉴클레아제 VIII를 처리한 후의 염기 교정을 확인한 결과이다.

도 3a 및 3b는 ABE 및 엔도뉴클레아제 V를 처리하거나 ABE, Alkyladenine DNA Glycosylase (AAG), 및 엔도뉴클레아제 VIII를 처리한 후의, 전유전체 서열분석 (WGS) 결과를 보여준다.

25 도 4는 ABE 및 엔도뉴클레아제 V를 처리하거나 ABE, Alkyladenine DNA Glycosylase (AAG), 및 엔도뉴클레아제 VIII를 처리한 후의, 절단 위치를 확인한 결과이다.

도 5는 CIRCLE-seq 및 SITE-seq에 의한 유전체 DNA 서열 분석 과정을 모식적으로 보여준다.

30 도 6a 내지 6c는 on target site에 대하여 0 내지 4개의 mismatched 염기를 갖는 sgRNA를 사용하여 얻어진 염기교정 빈도 (%) 및 삽입/결실 빈도(%)를 보여주는 그래프로, 별표는 하나의 효소에 대한 relative activity ([mismatched sgRNA 결과값]/[matched sgRNA 결과값])가 나머지

두 개의 효소보다 3배 이상 높은 mismatched sgRNA를 나타내고, 3' 말단의 마지막 3개 뉴클레오타이드는 PAM 서열을, 소문자는 mismatched 뉴클레오타이드를 각각 나타낸다 (Means $\pm$ s.e.m. were from three independent experiments).

5 도 7a는 ABE7.10 및 Endo V를 사용하는 Digenome-seq 분석법의 overview로서, ABE7.10가 구아닌의 데옥시이노신(deoxyinosine)으로의 변환을 촉매하고, Endo V이 데옥시이노신을 인식하여 상기 데옥시이노신에 대하여 두 번째 phosphodiester bond 3'를 절단함을 보여주며, 삼각형 꼭지점 위치는 ABE7.10 nickase 및 Endo V에 의하여 절단되는
10 phosphodiester bond 위치를 보여준다.

도 7b는 ABE-매개 Digenome-seq workflow의 개략도이고,

도 7c는 ABE7.10 및 hAAG/Endo VIII를 사용하는 Digenome-seq 분석법의 개략도로서, ABE7.10가 하나의 가닥에서 adenine-to-inosine 변환을 매개하고 다른 가닥에서 nick을 생성시키며, hAAG이 이노신을
15 제거하여 AP site (apurinic/aprimidinic site)을 생성하고, Endo VIII (DNA glycosylase and AP-lyase)가 상기 AP site을 절단함을 보여주며, 삼각형 꼭지점 위치는 ABE7.10 nickase 및 hAAG/Endo VIII에 의하여 절단되는 site을 보여준다.

도 7d는 ABE7.10 표적 서열을 포함하는 PCR 산물에 ABE7.10 및 Endo
20 V를 처리한 결과를 보여주는 것으로, ABE7.10와 Endo V가 모두 처리된 경우에만 DNA에 double strand break가 형성됨을 보여준다.

도 7e는 ABE7.10 표적 서열을 포함하는 PCR 산물에 ABE7.10, hAAG, 및 Endo VIII를 처리한 결과를 보여주는 것으로, ABE7.10, hAAG, 및 Endo VIII가 모두 처리된 경우에만 DNA에 double strand break가 형성됨을
25 보여준다.도 7f는 Sanger sequencing 결과를 나타낸 것으로, ABE7.10에 의한 A-to-G conversion과 Endo V에 의한 DNA 절단을 보여준다.

도 7g는 qRT-PCR 결과를 나타낸 그래프로서, ABE7.10 nickase와 Endo V에 의한 DNA 절단을 보여준다.

도 7h는 Sanger sequencing 결과를 나타낸 것으로, 유전체 DNA를
30 ABE7.10로 처리한 경우의 A-to-G conversion, 및 유전체 DNA를 ABE7.10, hAAG, 및 Endo VIII로 처리한 경우의 G-to-A conversion을 보여준다.

도 7i는 qRT-PCR 결과를 나타낸 그래프로서, 유전체 DNA를 ABE7.10, hAAG, 및 Endo VIII로 분해한 결과를 보여준다.

도 7j는 WGS data를 사용하여 얻어진 *HEK2* target site에서의 sequence reads 의 수직 정렬(straight alignment) 결과를 보여주는 IGV 이미지이다.

도 7k는 ABE7.10, hAAG, 및 Endo VIII 처리 후 관찰되는 sequencing
5 reads의 수직 정렬(straight alignment) 결과를 보여주는 IGV 이미지이다.

도 8은 nickel affinity chromatography 및 heparin bead chromatography를 사용하여 정제된 ABE7.10의 SDS-PAGE 분석 결과를 보여준다 (M: marker, I(-): cell lysate before IPTG (isopropyl-1-thio- β -D-galactopyranoside) induction, I(+): cell lysate after IPTG
10 induction, S: soluble lysate fraction, IS: insoluble lysate fraction, FT: flow-through, W1, W2: waste after washing, Ni: Ni-NTA agarose beads after elution of bound protein, NE: protein fraction eluted from nickel beads, Hp: Heparin Sepharose 6 Fast Flow affinity resins after elution of bound protein, HE: protein fraction after purification
15 using heparin beads. 박스로 표시된 부분은 ABE7.10 단백질 밴드를 나타냄).

도 9는 ABE의 Digenome-seq 분석을 위한 *In vitro* DNA 절단 점수 시스템을 개략적으로 보여준다.

도 10a 및 10b는 Digenome-seq에 의해 확인된 유전체 전역의 ABE7.10
20 off-target site를 보여주는 것으로, 10a는 온전한 유전체 DNA (회색; 중심원으로부터 첫 번째 영역) 및 ABE7.10 + Endo V 처리된 유전체 DNA (파란색; 중심원으로부터 두 번째 영역), 및 ABE7.10 + hAAG + Endo VIII 처리된 유전체 DNA (붉은색; 중심원으로부터 세 번째 영역)을 사용하여 WGS를 통하여 얻어진 DNA 절단 점수를 보여주는 Genome-wide Circos plots이고 (화살표는 on-target site을 나타냄), 10b는 ABE7.10 + Endo V 또는 ABE7.10 + hAAG + Endo VIII를 사용하여 확인된 인간 유전체에서의 *in vitro* 절단 부위의 개수를 보여주는 벤다이어그램이다.

도 10c는 Digenome-captured sites에서 DNA 서열을 비교하고 WebLogo를 사용하여 얻어진 Sequence logos이다.

도 11은 RNF2, TYRO3, WEE1, EphB4, HPRT-exon6, 및 HPRT-exon8
30 부위에서의 sequencing reads의 수직 정렬 결과를 보여주는 IGV 이미지이다.

도 12a 및 도 12b는 온전한 유전체 DNA (gray; 중심에서 첫 번째 영역) 및 genomic DNA treated with ABE7.10/Endo V (blue; 중심에서 두

번째 영역), BE3 Δ UGI/USER (red; 중심원으로부터 3번째 영역), 또는 Cas9 (Green; 중심원으로부터 4번째 영역)으로 처리된 유전체 DNA를 사용하여 *HEK2* (12a) 및 *RNF2* (12b) 부위에 대하여 얻어진 Genome-wide Circos plots을 보여주며, 화살표는 on-target 부위를 나타낸다.

5 도 12c 및 도 12d는 Digenome-captured sites에서의 DNA 서열을 비교하고 WebLogo를 사용하여 얻어진 Sequence logos이다.

 도 12e는 targeted deep sequencing을 사용하여 측정된 ABE7.10 off-target sites에서의 염기 교정 효율을 나타낸 그래프로서, 3' 말단 끝의 3개 뉴클레오타이드는 PAM 서열을 나타내고, 소문자는 Mismatched base를
10 나타낸다(Means $\pm$ s.e.m. were from three independent experiments).

 도 12f는 ABE7.10-매개 치환 빈도와 Cas9-매개 indel 빈도 간의 상관성 (위쪽) 및 ABE7.10-매개 치환 빈도와 BE3-매개 치환 빈도 간의 상관성 (아래쪽)을 보여주는 Scatterplots이다.

 도 13a는 일반적으로 사용되는 형태 (GX₁₉), 절단된 형태 (gX₁₈ or
15 gX₁₇), 및 연장된 형태 (gX₂₀ 또는 ggX₂₀)의 sgRNAs의 표적 DNA-상보적 RNA 서열을 예시적으로 보여준다.

 도 13b는 *HEK2*, *RNF2*, *TK_EphB4*, *TYRO3*, *WEE1*, *HPRT-exon6*, 및 *HPRT-exon8*를 표적으로 하는 변형 sgRNA 사용시 targeted deep sequencing에 의하여 측정된 각 변형 sgRNA 별 ABE7.10의 염기 교정 빈도 평균을 나타낸
20 그래프이다.

 도 13c 및 13d는 HEK293T 세포에서 targeted deep sequencing으로 측정된 *HPRT-Exon 6* (13c) 및 *TYRO3*와 *HPRT-exon8* (13d)의 on- 또는 off-target sites에서의 ABE7.10-매개 염기 교정 빈도(%)를 결과를 보여주는 그래프 및 히트맵으로서, 히트맵은 GX₁₉ sgRNA를 사용한 경우와 비교한 변형
25 sgRNA의 상대적 특이도 (relative specificities)를 보여주며, 상대적 특이도 (Specificity ratio)는 측정 대상 sgRNA의 특이도(on-target frequency/off-target frequency)를 GX₁₉ sgRNA의 특이도 (on-target frequency/off-target frequency)로 나누어 계산된 값이고, 3' 말단 끝의 3개 뉴클레오타이드는 PAM 서열을 나타내고, 소문자는 Mismatched base를
30 나타낸다 (Means $\pm$ s.e.m. were from three independent experiments).

 도 13e 및 13f는 ABE7.10 및 Sniper ABE7.10을 사용한 경우의 *HPRT-Exon 6* (13e) 및 *HPRT-exon8*, *HEK2*, *EphB4* 및 *TYRO3* (이상, 13f)에 부위에서의 on- 및 off-target 활성 및 특이도 비율 (specificity ratio)을

보여주는 그래프 및 히트맵으로, 3' 말단 끝의 3개 뉴클레오타이드는 PAM 서열을 나타내고, 소문자는 Mismatched base를 나타내고, specificity ratio는 다음의 수식에 의하여 계산된다: Sniper ABE7.10의 specificity (on-target frequency/off-target frequency)/ABE7.10의 specificity (on-target frequency/off-target frequency) (Means $\pm$ s.e.m. were from three independent experiments).

도 13g 및 13h는 플라스미드를 통하여 세포내로 전달되는 경우와 RNP 형태로 세포내로 전달되는 경우의 HPRT-Exon 8 (13g) 및 HPRT-exon6, HEK2, EphB4 및 TYRO3 (이상, 13h)에 부위에서의 on- 및 off-target 활성 및 특이도 비율 (specificity ratio)을 보여주는 그래프 및 히트맵으로, 3' 말단 끝의 3개 뉴클레오타이드는 PAM 서열을 나타내고, 소문자는 Mismatched base를 나타내고, specificity ratio는 다음의 수식에 의하여 계산된다: RNP 전달시의 specificity (on-target frequency/off-target frequency)/플라스미드 전달시의 specificity (on-target frequency/off-target frequency) (Means $\pm$ s.e.m. were from three independent experiments).

도 13i 및 13j는 ABE7.10 또는 Sniper ABE7.10을 변형 sgRNA와 조합하여 사용한 경우의 HPRT-exon6 부위 (13i) 및 HPRT-exon8 부위 (13j)에서의 on- 및 off-target 활성 및 특이도 비율 (specificity ratio)을 보여주는 그래프 및 히트맵으로, 3' 말단 끝의 3개 뉴클레오타이드는 PAM 서열을 나타내고, 소문자는 Mismatched base를 나타내고, specificity ratio는 다음의 수식에 의하여 계산된다: Sniper ABE7.10와 변형 sgRNA의 조합의 specificity (on-target frequency/off-target frequency)/ABE7.10의 specificity (on-target frequency/off-target frequency) (Means $\pm$ s.e.m. were from three independent experiments).

【발명을 실시하기 위한 구체적인 내용】

이하에서는 실시예를 들어 본 발명을 더욱 구체적으로 설명하고자 하나, 이는 예시적인 것에 불과할 뿐 본 발명의 범위를 제한하고자 함이 아니다. 아래 기재된 실시예들은 발명의 본질적인 요지를 벗어나지 않는 범위에서 변형될 수 있음은 당 업자들에게 있어 자명하다.

실시예 1: 아데노신 디아미나아제와 Cas9 니케이즈 (D10A) 처리 후, 엔도뉴클레아제 V를 처리한 경우의 유전체 DNA 서열 분석

세포에서 추출한 DNA를 아데노신 디아미나아제-Cas9 니케이즈 (D10A, Cas9의 10번째 아미노산이 D에서 A로 변이된 것)와 guide RNA 복합체와
 5 혼합하게 되면 DNA의 한쪽 가닥이 절단되고 반대쪽 가닥의 A가 I로 바뀌게 된다. 반응시킨 DNA에 엔도뉴클레아제 V (endonuclease 5)를 처리하게 되면 I를 인식하여 절단을 유도한다. 결국 디아미나아제-Cas9 니케이즈 (D10A)와 엔도뉴클레아제 V에 의해 DNA의 두 가닥 파손 (DSB; double strand break)이 유도된다 (도 1a 참조).

10 이 후 DNA를 정제한 후, DNA의 두 가닥 파손(DSB)이 발생한 DNA를 end repair 과정을 거쳐 전유전체 시퀀싱을 진행한다. DNA의 두 가닥 파손(DSB)이 발생한 DNA의 경우 잘린 부분에서 시퀀싱에 의한 서열이 on-target 위치에서 직선나열 (straight alignment)이 됨을 알 수 있으며, 전 유전체에서 직선나열 (straight alignment)이 된 부분을 찾음으로써
 15 아데노신 디아미나아제-Cas9 니케이즈의 off-target을 검출 할 수 있다 (도 1a).

실시예 2: 아데노신 디아미나아제와 Cas9 니케이즈 처리 후, Alkyladenine DNA Glycosylase (AAG)를 처리한 경우의 유전체 DNA 서열 분석

20 또한, 아데노신 디아미나아제-Cas9 니케이즈가 처리된 DNA에 Alkyladenine DNA Glycosylase (AAG)를 처리하게 되면 I이 제거되면서 AP-site (apurinic/apyrimidinic site)가 만들어 지게 된다. 그 후 AP-site를 절단할 수 있는 엔도뉴클레아제 VIII (endonuclease 8)을 처리하면 I가 제거되면서 DNA의 두 가닥 파손 (DSB; double strand break)이 유도된다
 25 (도 1b).

이 후 DNA를 정제한 후, DNA의 두 가닥 파손(DSB)이 발생한 DNA를 end repair 과정을 거쳐 전유전체 시퀀싱을 진행한다. DNA의 두 가닥 파손(DSB)이 발생한 DNA의 경우 잘린 부분에서 시퀀싱에 의한 서열이 on-target 위치에서 직선나열 (straight alignment)이 됨을 알 수 있으며, 전 유전체에서 직선나열 (straight alignment)이 된 부분을 찾음으로써
 30 아데노신 디아미나아제-Cas9 니케이즈의 off-target을 검출 할 수 있다 (도 1b).

실시예 3: 아데노신 디아미나아제와 Cas9 니케이즈 처리 후, 엔도뉴클레아제 V 또는 AAG를 처리한 경우의 유전체 DNA 서열 분석

ABE와 엔도뉴클레아제 V 또는 ABE, Alkyladenine DNA Glycosylase (AAG), 엔도뉴클레아제 VIII을 처리한 후 real-time PCR을 통해 DSB가 일어나는 지를 확인하여 그 결과를 도 2a 및 2b에 나타내었다. 도 2a 및 2b에 나타난 바와 같이, ABE만 처리하였을 경우에는 50% 정도 절단 되는 것을 알 수 있었지만, ABE와 엔도뉴클레아제 V 또는 ABE, Alkyladenine DNA Glycosylase (AAG), 엔도뉴클레아제 VIII을 처리한 후에는 90% 이상의 DNA target site가 절단 된 것을 알 수 있다. 이러한 현상을 다시 한번 검증하기 위해서 sanger sequencing을 진행하였다. ABE를 처리하였을 때는 A가 G로 바뀌는 것을 볼 수 있었으며, ABE와 엔도뉴클레아제 V 또는 ABE, Alkyladenine DNA Glycosylase (AAG), 엔도뉴클레아제 VIII을 처리한 후에는 G로 변했던 부분이 거의 전부 (ABE와 엔도뉴클레아제 V) 또는 일부(ABE, Alkyladenine DNA Glycosylase (AAG), 엔도뉴클레아제 VIII)가 A로 바뀌는 것을 확인할 수 있다.

실시예 4: Whole Genome Sequencing (WGS) 수행

상기 실시예 3의 ABE와 엔도뉴클레아제 V 또는 ABE, Alkyladenine DNA Glycosylase (AAG), 엔도뉴클레아제 VIII을 처리한 DNA를 WGS를 진행하여 그 결과를 도 3a, 3b, 및 도 4에 나타내었다. 도 3a에서와 같이, 서열 분석 결과에서 straight alignment를 확인 할 수 있다. 도 3b 및 도 4에서와 같이, DNA cleavage scoring system을 이용하여 genome-wide off-target candidates를 찾을 수 있다.

아데노신 염기교정 유전자 가위의 전유전체에서의 off-target을 프로파일링 하는 방법은 CIRCLE-seq¹⁷과 SITE-seq¹⁸을 이용하여 응용이 가능하다 (도 5).

아데노신 염기교정 유전자 가위의 전유전체에서의 off-target을 프로파일링 하는 방법은 CIRCLE-seq(Tsai, S.Q. et al. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. Nature methods 14, 607-614 (2017))과 SITE-seq (Cameron, P. et al. Mapping the genomic landscape of CRISPR-Cas9 cleavage. Nature methods 14, 600-606 (2017))을 이용하여 응용이 가능하다 (도 5 참조).

참고예

1. 세포 배양 및 형질 감염

HEK293T 세포 (ATCC CRL-11268)를 10%(v/v) FBS 및 1%(v/v) 페니실린/스트렙토 마이신 (Welgene) 보충된 DMEM 배지에서 배양하고,
 5 HEK293T 세포(1.5×10^5)를 24-웰 플레이트에 접종하고, Lipofectamine 2000 (Invitrogen; ABE 및 BE3의 경우 3ml, SpCas9의 경우 $2 \mu\text{l}$ 사용)을 사용하여, ABE (Addgene plasmid #113128; $1.5 \mu\text{g}$; DNA 중량), BE3 (Addgene plasmid #73021; rAPOBEC1-XTEN-Cas9n-UGI-NLS; $1.5 \mu\text{g}$) 또는 SpCas9 (Addgene plasmid #43945 NP_269215.1; $0.5 \mu\text{g}$)를 코딩하는 플라스미드를 sgRNA 코딩
 10 플라스미드($0.5 \mu\text{g}$)와 함께 70-80% confluency로 형질감염시켰다.

하기 실시예에서 상기 ABE 또는 ABE7.10는 서로 동일한 단백질을 지칭하기 위하여 사용되며, 구체적으로 아데노신 디아미나아제 (인위적 진화 (directed evolution) 된 *E. coli* TadA (NP_417054.2))와 SpCas9 니케이즈 (D10A)(SpCas9 (NP_269215.1)의 10번째 아미노산 잔기 D가 A로
 15 치환된 변형 Cas9)가 연결된 융합 단백질을 의미한다 (참고예 2 참조).

ABE RNP-매개 유전체 교정의 경우, ABE7.10 단백질 ($10 \mu\text{g}$) 및 in vitro 전사된 sgRNA ($6 \mu\text{g}$), 또는 ABE7.10 단백질 ($0.6 \mu\text{g}$) 및 sgRNA ($0.2 \mu\text{g}$) 코딩 플라스미드를 Neon™ Transfection System(Thermo Fisher Scientific)을 통한 전기천공에 의하여 HEK293T 세포 (1×10^5)에
 20 형질감염시켰다. 72 h post-transfection 시점에 DNeasy Blood & Tissue Kit (Qiagen)를 사용하여 유전체 DNA를 추출하였다.

하기하는 실시예에 사용된 sgRNA는 표적 부위 (표 8 내지 18에서, 'sgRNA 표적 부위' 또는 'DNA seq at a cleavage sites'로 표시) 중에서 3' 말단의 PAM 서열 (5'-NGG-3' (N은 A, T, G, 또는 C임))을 제외한 서열에서
 25 T를 U로 바꾼 서열을 아래의 일반식 3 중의 표적화 서열 '(N_{Cas9})₁'로 하여 제작된 것을 사용하였다:

5'-(N_{Cas9})₁-(GUUUUAGAGCUA)-(GAAA)-
 (UAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC)-3'
 (일반식 3; 올리고뉴클레오타이드 링커: GAAA).

30

2. ABE7.10 단백질의 발현 및 정제

ABE7.10을 다음과 같이 설계하였다 (N 말단에서 C 말단 방향):

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQ

GGLVMQNYRLIDATLYVILEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGIL
 ADECAALLSDFFRMRRQEIKAQKKAQSSTDSSGSSGGSSGSETPGTSESATPESSGGSSGGSSEVEFSHEY
 WMRHALTLAKRRDEREVPVGA VLNNRVIGEGWNR AIGLHDPTAHAEIMALRQGGLVMQNYRLIDATL
 YVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFR
 5 MPRQVFNAQKKAQSSTDSSGSSGGSSGSETPGTSESATPESSGGSSGGSDDKKYSIGLAIGTNSVGWAVIT
DEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAK
VDDSFHRL EESFLVEEDKKHERHPIFGNI VDEVAYHEKYPTIYHLRKKLVDSTDKADRLIYLALAHMI
KFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGE
KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKQLSKD TYDDDLNLLAQIGDQYADFLAAKNLSDAILL
 10 SDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKF IKP ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHA I LRRQEDFY PFLKDNREKIEK
ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASQSFIERMTNFDKNLPNEKVL PKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGT YHDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIERLKT YAHLFDDKVMKQLK
 15 RRRYTGWGRLSRKLINGIRDKQSGKITLDFLKSDFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHE
HIANLAGSPA I KKGILQTVKVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG
SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLRSDK
NRGKSDNPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVA
QILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYK VREINNYHHAHDAYLNAVVG TALIKKYP
 20 KLESEFVYGDYKVYDVRKMI AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI
VWDKGRDFATVRKVL SMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS
VLVVAKEKGKSKKLKSVKELLGITIMERSSEKNPIDFLEAKGYKEVKKDLIKLPKYSLFELENGRKR
MLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIEQISEFSKRVIL
ADANLDKVL SAYNKH RDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI
 25 TGLYETRIDLSQLGGDPKKKRKV

(상기 아미노산 서열에서,

(1) 볼드체로 표시된 폴리펩타이드 부분: 야생형 *E. coli* TadA (NP_417054.2);

30 (2) 이탤릭체로 표시된 폴리펩타이드 부분: directed evolution된 *E. coli* TadA (NP_417054.2) (야생형(NP_417054.2)에서 변이된 아미노산 잔기를 이탤릭체+밑줄로 표시함);

(3) 밑줄로 표시된 폴리펩타이드 부분: SpCas9 nickase (D10A);

(4) 폴리펩타이드 (1)과 (2) 사이 및 폴리펩타이드 (2)와 (3) 사이의

올리고펩타이드 (SGGSSGGSSGSETPGTSESATPESSGGSSGGS): 펩타이드 링커; 및

(5) 폴리펩타이드 (3)의 C 말단에 연결된 올리고펩타이드 (PKKKRKV): SV40 NLS)

상기와 같이 설계된 ABE7.10(N-terminal His purification tag 포함)를 암호화하는 human codon-optimized 유전자가 pET28(Addgene)에 클로닝된 ABE7.10 발현용 플라스미드를 준비하고, 상기 플라스미드를 BL21 Star (DE3)-competent *E. coli* 세포 (NEB)에 형질전환시켰다. 형질전환 후, fresh single colony를 37℃에서 진탕하면서 카나마이신 50μg/mL 함유 Luria-Bertani (LB) 배지 (Thermo Fisher Scientific) 에서 밤새 배양하였다. 1:50으로 희석시킨 pre-cultures 8 mL를 카나마이신 50 μg/mL이 보충된 LB 배지 400 mL가 들어있는 각각의 플라스크에 접종하한 후, OD₆₀₀=0.65~0.70이 될 때까지 37℃에서 진탕 배양하고, 얻어진 배양물을 얼음에 보관하였다. IPTG (isopropyl-b-D-1-thiogalactopyranoside; GoldBio) 1mM를 사용하여 단백질 발현을 유도한 후, 배양물을 18 ℃에서 16 시간 동안 진탕 배양하였다.

이어지는 단백질 정제 과정은 4℃에서 수행하였다. 상기 얻어진 세포 배양물을 6,000g로 10분 동안 원심분리하여 세포를 수확하고, 용해 완충액 (50 mM sodium phosphate (Sigma-Aldrich), 500 mM NaCl (Sigma-Aldrich), 10 mM imidazole (Sigma-Aldrich), 1% Triton x-100 (Sigma-Aldrich), 20% glycerol, 1 mM 1,4-dithiothreitol (DTT; GoldBio), 1 mM phenylmethylsulfonyl fluoride (PMSF; Sigma-Aldrich), concentration of 1 mg/mL lysozyme (Sigma-Aldrich), 10uM ZnCl₂ (Sigma-Aldrich), pH 8.0)에 재현탁시켰다. 그 후 현탁액을 액체 질소와 수조 (37 ℃)에서 교대로 3 회 반복 배양하였다. 9분 (5초 (on), 10초 (off)) 동안 초음파 처리하여 세포를 용해시키고, 15,000g에서 20분 동안 원심분리하여 용해물을 분리(cleared)하였다. 상층액을 니켈 아가로오스 비드 (Ni-NTA, QIAGEN)와 함께 60분 동안 4℃에서 진탕 배양하였다. 용해물-수지 혼합물을 폴리프로필렌 컬럼에 로딩하고 3 컬럼 부피의 세척 완충액 (50 mM sodium phosphate (Sigma-Aldrich), 150 mM NaCl (Sigma-Aldrich), 35 mM imidazole (Sigma-Aldrich), 20% glycerol, 1 mM DTT (GoldBio), 10 μM ZnCl₂ (Sigma-Aldrich), pH 8.0)으로 세척하고, 결합된 단백질을 용출 완충액 (50 mM sodium phosphate (Sigma-Aldrich), 150 mM NaCl (Sigma-Aldrich), 250 mM imidazole (Sigma-Aldrich), 20% glycerol, 1 mM DTT

(GoldBio), 10 μ M ZnCl₂ (Sigma-Aldrich), pH 8.0)로 용출시켰다. 용출된 단백질 분획을 헤파린 아가로오스 비드 (Heparin Sepharose 6 Fast Flow, GE Healthcare)가 들어있는 폴리프로필렌 컬럼에 넣고 3 컬럼 부피의 세척 완충액 (50 mM sodium phosphate (Sigma-Aldrich), 150 mM NaCl (Sigma-Aldrich), 20% glycerol, 1 mM DTT (GoldBio), 10 μ M ZnCl₂ (Sigma-Aldrich), pH 8.0)으로 세척하였다. 결합된 단백질을 용출 완충액 (50 mM sodium phosphate (Sigma-Aldrich), 750 mM NaCl (Sigma-Aldrich), 20% glycerol, 1 mM DTT (GoldBio), 10 μ M ZnCl₂ (Sigma-Aldrich), pH 8.0)으로 용출시켰다. 생성된 단백질 분획물은 원심분리 컬럼 (Amicon Ultra-4 Contrifugal Filter Devices, Millipore)을 사용하여 6,000g에서 농축시켜, ABE7.10 단백질을 정제하였다.

3. PCR amplicon의 ABE7.10-매개 시험관내 분해 (*in vitro* digestion)

표적 유전자의 표적 부위를 포함하는 (표 1 내지 11 참조) PCR amplicon을 37°C에서 1시간 동안 100 μ l의 반응 부피로 ABE7.10 단백질 및 시험관내 전사된 (*in vitro* transcribed) HEK2-표적 sgRNA와 함께 배양하였다. 상기 얻어진 디아미네이션(deamination)된 생성물을 PCR 정제 키트 (MGmed)를 사용하여 정제하였다. 디아미네이션된 생성물 2 μ g을 엔도뉴클레아제 V (Endo V; NP_418426.2) (40 units) (New England BioLabs) 또는 인간 Alkyladenine DNA Glycosylase (hAAG; NP_001015052.1, New England BioLabs) (10 units) (New England BioLabs) 및 엔도뉴클레아제 VIII (Endo VIII; NP_415242.1) (20 units) (New England BioLabs)와 함께 200 μ l의 반응 부피로 37 °C에서 30 분간 배양한 후, DNA를 PCR 정제 키트 (MGmed)를 사용하여 다시 정제하고 Bioanalyzer (Agilent)로 분석하여, *in vitro* digestion 결과를 측정하였다.

4. 유전체 DNA (genomic DNA) 의 ABE7.10-매개 시험관내 분해

제조사 설명서에 따라서 DNeasy Blood & Tissue Kit (Qiagen)을 사용하여 HEK293T 세포로부터 유전체 DNA를 분리하였다. 유전체 DNA 8 μ g을 정제된 ABE7.10 단백질 (300nM) 및 sgRNA (900 nM)와 함께 400 μ l의 반응부피로 37 °C에서 8시간 동안 반응 완충액 (50 mM Tris-HCl (pH 8.0), 25 mM KCl, 2.5 mM MgSO₄, 0.1 mM EDTA, 10% glycerol, 2 mM DTT, 10 μ M ZnCl₂)에서 배양하였다. RNase A (50 μ g/mL)를 이용하여 sgRNA를 분리한 후,

DNeasy Blood & Tissue Kit (Qiagen)를 사용하여 DNA를 정제하였다. 정제된 DNA 3 μ g를 Endo V (40 units) 또는 hAAG (10 units) 및 Endo VIII (20 units)와 함께 200 μ l의 반응부피로 37 °C에서 2시간 동안 배양한 후, DNeasy Blood & Tissue Kit (Qiagen)를 사용하여 다시 정제하였다. 정제된 DNA를 KAPA SYBR FAST qPCR Master Mix (Kapa Biosystems)와 함께 배양하고 analyzed by real-time quantitative PCR (qPCR)로 분석하여 ABE7.10- 및 Endo V- 또는 ABE7.10-, hAAG-, 및 Endo VIII-매개 DNA 이중가닥절단(DSB)을 확인하였다. 온전한 유전체 DNA 분석은 comparative C_T 분석법으로 측정하였다.

10

5. 전체 유전체 시퀀싱 (Whole genome sequencing; WGS) 및 digenome sequencing

상기 얻어진 시험관내 분해된 DNA 1 μ g를 Covaris 시스템 (Thermo Fisher Scientific)을 사용하여 400 내지 500 bp 범위로 단편화하고, End Repair Mix (Illumina)와 함께 배양하여 무딘 말단(blunt end)을 만들었다. 단편화된 DNA를 어댑터(illumina)로 연결하여 라이브러리를 생성하고, (주)마크로젠에 의뢰하여 HiSeq X Ten Sequencer (Illumina)를 사용하여 전체 유전체 시퀀싱 (WGS)을 수행하였다. WGS는 sequencing depth를 30-40x로 하여 수행하였다. 유전체 서열 맵핑에는 Isaac aligner를 사용하였다. DNA 절단 부위는 Digenome 1.0 및 Digenome 2.0 프로그램을 사용하여 확인하였다 (Kim, D., Kim, S., Kim, S., Park, J. & Kim, J.S. Genome-wide target specificities of CRISPR-Cas9 nucleases revealed by multiplex Digenome-seq. Genome research 26, 406-415 (2016)). 본 명세서에서 사용된 Digenome 2.0 버전의 소스 코드는 'https://github.com/chizksh/digenome-toolkit2'에서 확인할 수 있다.

20

25

6. 표적 심층 시퀀싱 (Targeted deep sequencing)

KAPA HiFi HotStart PCR polymerase (#KK2502; KAPA BIOSYSTEMS)를 사용하여 표적 부위 (on-target sites) 및 잠재적 비표적 (potential off-target sites) 부위를 증폭시켰다. TruSeq HT Dual index-containing primer를 사용하여 증폭된 앰플리콘을 다시 증폭하여, deep sequencing 라이브러리를 생성하였다.

30

상기 표적 심층 시퀀싱에 사용된 프라이머를 하기의 표 1 내지 7에

정리하였다:

[표 1]

HEK2에 대한 프라이머

Site No.	1st PCR_F	1st PCR_R	2nd PCR_F	2nd PCR_R
HEK2_001	CCTTTCTCCACAGGC AGAGG	CCATCACACCCAGCC TCATT	TCAGGCGCTGTTGCT AGAAA	CAGAGGAGGAAAGCA AGCTCA
HEK2_002	GTTCCCACTGGACAG TCGAG	GCAGGCAAATGAGGC TTCTG	GTTCCCACTGGACAG TCGAG	GGCTATTGCTGGGG GAACT
HEK2_003	ATGAACAGGGGGGCT TGATG	CTGAGCCACTCCTTT CCCTG	CCTCCATCTTCTCCG CAGAC	TTCTTCGCCCTCTCA GCTTC
HEK2_004	AGTCGTACTATGCAA GCCACA	GGGCACATGTGTTTG CGTAA	TGGGTTCAAAAACAA ACAGAGAAGA	TCTTGTGAAACAGAA ATGTCAGTT
HEK2_005	TGCTGTGGTCCTAGG TCTCA	CTCAAGCAATCCTCC TGCCA	AGTGAAATCAGTGG GCCTG	TGGAAGAACCACCT TGTGAGT
HEK2_006	AGGCTGGGTGCAATT TGTTT	CCACAATTACATTTG CAGCAACC	TCCCATCACCATTTA CTGAGCA	ACACTGCCTGATTTT AAGCT
HEK2_007	GAGGCCCATTAACGT TTGGC	AGACTCAAAACCTGG CCCAC	ACCTGGCTGAGCTAA CTGTG	TCCAGCCCCATCTGT CAAAC
HEK2_008	TCCACCTGTTTCATCC TCCCT	ATGCCACCTCTGAAG CTGAG	TCTTTCATTGCATGG CTGTACC	TGAATTTGTGTTGGG CCACA
HEK2_009	TGGAAC TGGGAATT GGTGT	AGGGGCAGGAACACA AGAAA	CCTGATTTGATAGAC ATGACATTGT	CAAACAAACAAGCAA AGCCTGT
HEK2_010	GAATATCAGGGGATG GCGCA	AGAGCTAGGAGGGCC TGAAA	TGTCTTGAGTGGGCT GGTTC	GGCCTGAAATTTGCC CTGAC
HEK2_011	TCCTGGTGTCTCTT GTGCC	TCCACAGTTCCCTAG GGCAT	AATTCTCAAGAACAG TGAAAGTTGT	TCCACAGTTCCCTAG GGCAT
HEK2_012	TCACCTCCTCCTCT CACTT	ACAAGGACAGGCCAA GACTG	ACCCACAGAACAAA GAAATAACCC	AGTGGGACTATGCCA GAGAT
HEK2_013	CACTGGGGATGGCAA TCTGT	AAAGGCGGAAAGAGA CAGGG	ACCAGCAAAGAAAAG GGGGA	GGAAAGAGACAGGGT CCTGAG
HEK2_014	ATGGATGGTAGTGTC GTGGC	TGGAATAGACCTGGG CTCCA	CACACACACACAC ACACA	TGGAATAGACCTGGG CTCCA
HEK2_015	CACGCAGCCAAGTCA AACTC	GCAGAAGAATGGCGT GAACC	GCCACTGGCCATCCT AACAA	TTTCTGTGCCCTATG CTCCT
HEK2_016	CTGACAGGCATTAGG TGGGG	ATTATCTGCCTGCCT CAGCC	TCCCAGTCCAGCCTA TGC	GCACCTGGCCCACTA TATGT
HEK2_017	CATGTCCCCCTTTCC TCCAC	GTGGAAGAGCCTCTG GGAAC	CCTGCCCTTTCCTTC CTTGT	GGGCAGTTGTGAATC ATGGC

HEK2_018	GGTTGGGGTTGGGAT TAGGG	GGTGGAAGGACTCGC AGAAA	TGACAACCACGTTCA TGGGA	CCCATCTGTCTCTC TGGCC
HEK2_019	GATCAGCCCAGGTGC TCTTT	AGTTCAAGATCAGCC TGGCC	CTGCCTAGCCCAAAG ACCTC	TGGTCCCAGCTACTC AGGAA
HEK2_020	AAAGCAAGAGGGTGG GAGTG	TGACGACCCCTACT TCACT	CGTTGGCTTAGACCA GGGTA	TCCAGGATTTTGCCC CATCC
HEK2_021	CAGCCTGGCCAACAT AGTGA	AAACTGTGCTCCCTT AGGCC	ACCTCACTCTGTGTC CCCTT	AAACTGTGCTCCCTT AGGCC
HEK2_022	TATCAGCAGTTCCCC ATCGC	GGCTCTTACCAGTTT CCCCC	ACATCTGCAGCCCAG TTTCT	CACAGTGCTTTGCCC CATTG
HEK2_023	CTGCCACTAAGCCAC CTTCA	AGGCTGAGGCAGGAG AATTG	TCCCATGGATGAAAA GCCCCA	GTGACAGAGTGAGAC TCCGC
HEK2_024	GTGGAGGGTGCATGT GTACA	CTAGGACACGGCTAA GGCAG	TTTTCTGTGGGGCCA GTGTT	CCAAGAGGAAACCTA CCCCA
HEK2_025	TTCGGCGTGTGAAGT TAGCT	CGATCTTCCCACCTT GGCTT	CTGCCGTGGGAGACA ATTCA	ACTGGTTCTGTTTGT GGCCA
HEK2_026	AGCTTGCCAGATTGG TGGA	GAACACCATGCACGT TCACC	AGAGAACAGAGAACC TGCAAGG	CTGGCTTTGCTTCTC CCGTA
HEK2_027	ATGGGCTCAGCTACG TCATG	ACACAGAGTACACAG GCACG	GGGCCCAGGGTAAGA CTCTA	CTTGCACTGAGCCGA GATCA
HEK2_028	CAGCCTATGCCCAGG AATCC	CTAACCAAGAGCCCT GGTGT	CTGCTTAGGTCCAGC CAACA	TCTTGAAACTCCATG GCTGAA
HEK2_029	CACTCCAGCTCAGGT GACAG	TTACAGACGTGAGCC ACCAC	CATGTGAGGATGCAA TGGGC	GAGGCAGTTTGAGCT GAGCT
HEK2_030	GACCCTGAGCTGCTA CCATC	AGAAATGCTGCTCTG GCCAT	ACGAGTAAATGTTAA GCTTCTCCCT	ATGCTGCTCTGGCCA TAGAC
HEK2_031	ACAATGGTCAGGTAC GGCAG	TCCAGTGCAACACAG TCTGG	ACTGGTGTCTGAAG TCAAGTC	GAGCAGAGCCTCACA CTTGT
HEK2_032	GAACTCACAGAGCGG TGGA	CACCCAGGAAGACAT GGTCC	GTGTCTGGCGGAGAG AGAAC	ACAAATCCTGGCACC AGACC
HEK2_033	ATGGGAGAGCAATCC AGTGC	AAAACCTATGCCCC CCTGC	TGGCCCCAAGATCAA AAGCT	TGAGGGTAAGGACTG AGGCA
HEK2_034	GGCTTGTAGGATGAG CTGCA	TTCTTCTGCTTGCTC CAGCC	CTACCTCTCCAGCAA TGGTCC	TGCTTGGTCCAGCCT ATTGA
HEK2_042	GAGACCTTATGCCCT TCACCA	CGGTGAATTGGGTAA TGGCC	CCAAAGGCATGGTCC ATGAA	TTGGTGAGGCACTGT TTGGA
HEK2_044	GCACTTTGGATGTGG AAGGG	ACAAGGTCATTGCCT TCACA	AGACCTACACTGACC TGAGGT	ACAAGGTCATTGCCT TCACA
HEK2_050	CCCAACACAATCCCC	TCTGGATGTCTACCT	CCCTACCCAACACAA	TCTCTGGATTGCTTT

	TCCAG	CTCTAGCA	TCCCC	ACAGGTT
HEK2_052	AGCGGTGTAGGGATG AAAGC	GGCCAACTCCTCCTC ACTTC	GCTTCACCTCTCTGA GCCTC	TGGGCTGAAAACAAG AGGCA
HEK2_065	AGCATGTTGCCTAGG CTTGT	GCCCTGTATTTCAGAT CCGCT	TCCTCTGCCCTGACC ATGTA	GAGAGCCTCGGGTGA TTCTG
HEK2_079	CATGTGTCAGTTCCA GCCCT	GGCAAGAACTGGGT AGCCT	CGGTGCTTCAATGCT GTAAGG	AGAGGATCCCCAACG ACAGA
HEK2_082	TTCTCCAGCTGTGCA GAGTG	GAGCAGCTGTCCTCA ACCAT	TTCTGTACCTGCTGG GCATC	TCAAGGAAATCACCC TGCCC
HEK2_092	TCGACCTTGAAAAA CCTCCT	TGGAAGCTCTCCCTT CAGGA	TGGAGCTGCTCTTTT TTTCCTC	CACCTGAAGGCACC TTTTCA
HEK2_093	CCACCATTTCAGTGA AGGCC	TGCTCTTACCTGGCT CATGC	CCACCATTTCAGTGA AGGCC	AAGGTCAGGGGCCAT GTTTT
HEK2_099	GGAAGCTTTGCAAGT CAGGC	ACATACACTGGGGCC TGTTG	TGGGCACTGCTCAGT AAACA	GCACATCCTGTACAC ATACCCA
HEK2_101	AAGTGCTGGGATTAC AGGCC	GCCAGCAGGAAAGAG GACAT	CTGGCTCTCAACTTC CTCCG	CTAGTGCAAGGAAG CCTGG
HEK2_105	TGTGTGTCTTAAGCC GTGTC	CTTCTGGAAGATGC TGGGG	CCAGAAGTCTCCGTG GCTG	TTAATGCCGTATCCC CAGCC
HEK2_106	TTCCATAGAGGGGAG GCGAA	CAAGGGAGTTGCTAA GGGGG	TGGGACTGACCGAGA GACAA	CAAGGGAGTTGCTAA GGGGG
HEK2_110	GGGAGACACATGGGC TTTCA	CGCCCGCCAATTTT TTGTA	TCCCAAACTGGCTC TTGCA	ACCTGGCCCCCTTTA GATCT
HEK2_119	GCTGGATGGCAGCTC AAAAC	TTGGTCACATCACTC AGCCC	GCTGGATGGCAGCTC AAAAC	ACCTAGCACGCAATC AGTCA
HEK2_121	GCCACAACATCTGGT GGGTA	GAGGTGGAAGGTGG CTTGA	AGTTCTCCAGCAATG GTGTT	AGAGCCAGACCCTGT CTCAA
HEK2_139	AACCAGCATTATGGA CCGGG	TGGTGGCAATCTAGT GGAAGG	AACCAGCATTATGGA CCGGG	ACGATGAACCATCTG CTGAGA
HEK2_145	GGCATTCTTTGAGC AGGGC	AATCCCTTTTCCGTG GTGGG	GGCATTCTTTGAGC AGGGC	CATTCTGGCTCACA GTGGA

[표 2]

RNF2에 대한 프라이머

Site No.	1st PCR_F	1st PCR_R	2nd PCR_F	2nd PCR_R
RNF2-1	CTTTGCAGGTCAC AAGGCAC	TTTGGGTGGGTCT AAGTGGC	CACAGCTCAGGGT GGCTAAA	TGTGGCTTATGCT GCACTCA
RNF2-2	AGTGTCTGGTGAG GGCCTAT	ACCACTGTTACC CAGTACC	TCCAGCAATGTCT CAGGCTG	AGCCAACATACAG AAGTCAGGA

RNF2-3	AGAAGCAGAGGGT TGTGCT	GAGAGCTTGAGTG GGTCACC	CTCTGGCCATTGA TGCCAAA	CAGCTCATCTGAA AGCTCTGT
RNF2-4	TCTGGACCAAAAG AGCTTGCT	TGGAGTTGTCCCG AAGAGAAC	ACACCAGGTATCT GTCTTGACA	TGCAGACCCTATC TTTACAGCA
RNF2-5	GAGGGTCCCTCTC TGGAAC	CAGTGGCTGTGAG GATGTGT	CTTTGTAGCTGCA GGCCACA	GGGAGTCAGGAGC TTGCAG
RNF2-6	GGGTAAACCGAGG CATGTGA	CCCTGCTCATGCT CAAAGGA	GGCTTGAGGTAGT TGGCCAT	CTTGACACAGGAG AGGCAGG
RNF2-7	AGGGCCAGGCTAC CAGATAA	CCCCAGCTGCTA AATGTCA	TGGAGAGGGGGTA GAAGTGG	ATGTCACCTGTTT GCCTGCT
RNF2-8	AGGTCTTCTGACA ACCTGCA	AACTTAGTCAGGG GCCATGC	AGCTGGTTCCAAA CTTAGCCA	GGAACAAGAACAG TGAGGCC
RNF2-15	ACAAGCCTGAGTC ACTGAGC	ACCTTATGCACCT TGCCCTC	ACCTCATTTCTT GGAGTTCC	TTTATTGTCATT GTTACCATGACT
RNF2-19	AGAAAGAGCTCCC AATGGCC	CCCAGTGTTTTTC CCAGGGT	TGAAGGGTGTGGG GTAGGAA	CCCAGTGTTTTTC CCAGGGT
RNF2-20	GGACTGAGTCCCA GCCTCTA	GAGAGGTCAGGCA AACGACA	CACCTAGAACAAT GCCAGGC	CCACGCCATTG CTTTCAT
RNF2-24	AGGTAACAGTCCA GGCTGGA	ACAACGCACCTAG GGAAGT	TGACACCACCACA ATTCTTGG	AAAGCCACCGACC TTCAAGG
RNF2-27	GCCACTGAGAGCA GAGACAG	CTCTTTGAAGGTC TGGGGCA	GCCACTGAGAGCA GAGACAG	CTCTTTGAAGGTC TGGGGCA
RNF2-41	TACCCATCACAGT GGTCCCA	TGTTCCATAGTTC CTTTTTCCT	ACCACAATAGTAC ACAAGCCA	ACTTTTGTCTT CTTTGGTTCA

[표 3]

TYRO3에 대한 프라이머

Site No.	1st PCR_F	1st PCR_R	2nd PCR_F	2nd PCR_R
TYRO3-1	CCACTCCAAGATG AGGCCTG	TCCCCCTGAACCT GTTACCA	ACCACAACAGCCA GGACTTC	CCTGGCCTCTTCT CGCATAG
TYRO3-2	AGTTACACAGGGC CTTCGTG	CTGTCAACAAAGT GCTGGCC	CTTGACTCCCATG CCTCCTG	CTGTCAACAAAGT GCTGGCC
TYRO3-3	CAGACAAGCCTGC TTCTCCA	TCAGCAGTGTGAG AACGGAC	CAACCTTGTAACA GGGCCCT	AGCTTACTTTCTT GGGGAACA
TYRO3-4	GCCCTGTTAGACT GTTGCCT	TAGCTTAGTCTGG GCCACCT	GCTGCTATAGGCT GGGACTG	GCCCTTACCAGGA GCCAAAT
TYRO3-5	CAGCTGCCTTACA CACTGGA	TGCCCCAATCATG TCTCCAC	TTCGGAGTCATAG CCACTGC	GACCAGGTGTCAC GACTGTT
TYRO3-6	GCACCCATGATCT	TAGACAGGCCAGG	GATGACTTCTATC	AGCATGAAGAGCT

	CCAACCA	AGTTCGA	CGGGGCC	GGTCAGG
TYR03-7	GTGAGTCAGTGGT CCAGGTG	GAACAGGACTGTG GAGTGGG	CGAGGTGTTCTGT CCAGGAT	GGATGCAGTGAAC TTGTGGC
TYR03-8	TTTGGGGCTTTGA GTGTGGT	TACCATGTCCCA TGCCAAC	GGACCACCGATCA CAACACT	TAGAAGCCAACAC CCCACAC
TYR03-9	GCCTCACTTCACC TGGGATC	TCATGTTGGTCAG GGTGGTC	AGACTCCTTTCTC ACGCACT	TCCGTGGGGTCTA TTAGGCC
TYR03-10	TTGACCCTCTCTG AGCCTCA	TACACTCTGAGCC CTACCCC	TGGGAAGCACCTG GGATAGA	TACACTCTGAGCC CTACCCC
TYR03-11	AGGTGCTGGTGAT GAGTGTG	TGACTAGAGAAGC GGGTCCA	GCTGGTGATGAGT GTGTTGC	AGAGTTATGGTTA CTGCTCAGTT
TYR03-12	CTTCCCAGAAGCC TCTGCAA	GCTCATTTCCACT CCTGCCT	GGGCATCAGAAGT GGGACAA	TACCCACCTGTGT CCGTTTG

[표 4]

WEE1에 대한 프라이머

Site No.	1st PCR_F	1st PCR_R	2nd PCR_F	2nd PCR_R
WEE1-1	GAGCCATGTCTGC ACCTCTT	GGGAACAATGCCA CCCAAAC	TTCCCTGCCCCAT TGCTATC	ACACTCACAGGGC ATGTCTG
WEE1-2	TGGCATGCAAATA TCTCCCA	ATAAGCAGCACAC CACCACA	AGCACAGATGGCA AAAGCTG	GTCACAGTGTTC GGGGGAG
WEE1-3	ATGTGGCCAAGA TACCCAC	GCTACTTCCTACT GTCCCCC	CACTGGACATCTG AGGAGGC	CAGGCTGGAGTTC ACAGAGG
WEE1-4	TCATCAGAGGACC CCTGTGT	ATCTGTTTTGGAG GCAAGGT	ACTCCAAGAAGAT TACTTACTTCCA	ATCTGTTTTGGAG GCAAGGT
WEE1-5	CTGGCCAGACTT CTTGAGG	ATGGAGTGCATGG AGTGGTG	AGGTCCTGCTCCT CCTATGT	CTGTCCTTTGACC CAGTCCC
WEE1-6	ATGCTGTGTTCCA ACCCAGT	TGGCTCCACCATG TTGGAAG	AGGGTGACTTTGA CAGAACCA	CCTCTGGGCACTA TGCTAGG
WEE1-7	ATGTCTAGACCGT TGCTGGC	CGTTTGCTTTCAG TGACCCAG	ATGTCTAGACCGT TGCTGGC	CGTTTGCTTTCAG TGACCCAG
WEE1-8	GTGATTGGCTACC AACCCCA	CTTGAGGCCAGG ACACAAT	TGTGCCACTTAAC ATGGACTCT	GACCTAAAGACGT GCCACCT
WEE1-9	ATGTCCACACGAA CACCTGC	CCACCACATACAG CCTCTCC	TGGAAGTTACCAA GATGCCCC	AGTTAGTTTTCCA GAATGCCGT
WEE1-10	TGGGTGTTTTGAA GTGGCCT	CAGTGAGAGCCCT GAGGTTG	CCTCTGCTGAAAT CCGTGGT	TGGACCAGTACCA GTCCATG
WEE1-11	CGTGAGACATGGG GAGCTTT	GGGTTCCCTTCCC TGAGTGG	GTGGGTCTTTCTC TGGCCTC	CCTGGTTAGGCCC CATGATC

WEE1-12	TGGCCAACATGGT GAAACCT	TAGTGGTGTGACG AGGGCTA	GCAGTGAACCAAG ACCATGC	AAACTTGCTGTCC TTGCCCT
WEE1-13	AGCTGAAGTCTGG AAACGCA	GGTGACGAGTGAA ACCGTCT	TCCCACGACCCTA GTAGTCC	TTGCAGTGAGCCA AGATCGT
WEE1-14	TGGTGTGCTTAC CTGATTTGC	TGAAGAAAGGAGG TGTAGCTCT	TGCACTCATACAA TGTGGTCA	CCATCTGACCTGG GGGAAAC
WEE1-15	GA CTCCTGGGCTC AAGCAAT	TGTGCCCTGCTGA CATCATT	GA CTCCTGGGCTC AAGCAAT	TGCAGGCTATTGA CTCCTTCA
WEE1-16	TGAGCCCAGATTG TGCAACT	GCCTGCTCCTGAG ATCATGG	CAACAAAGCGAGG CTCTGC	CACACACACACAC ACACACA
WEE1-17	TTTGAAGCTCAG CCCTGTC	AATGGGAGATTGG TGGGCAG	AGTGCCAACAATT GTGAGTGT	GCTGTTGCTGGTC TTTGGT
WEE1-18	ACTCCTCACTCAA GCAAGCC	GTCCCTCAGGCCT CTCAAAC	GCCTCATCTACCT GCAAGGG	GTCCCTCAGGCCT CTCAAAC
WEE1-19	CTCAGCCTCCAA AGTGCTG	TAGGAGAAGGGGG AGAGCAC	AGCCTTGCTTTG AGTTCTGA	ACTCCTTCCCTGT TTCACCT
WEE1-20	GGGCCAGTCAAAT GAAGGGA	TTTGGGGGGCTG TCTCTTC	GGTCACTGACTGC GTCTCAA	TCAATCCTCCCTC ATTGTGGT
WEE1-21	TGCAGGGCCATTT GCATAGT	AGGATGGCAGATT TCTCATTGGA	CGGGTGGTATTTT TGCTAACCC	AGTTTGTAGAAAA GGTGCTGCC

[표 5]

EphB4에 대한 프라이머

Site No.	1st PCR_F	1st PCR_R	2nd PCR_F	2nd PCR_R
EphB4-01	GCTGGTCTCGAA CTCCTGAC	GTTCAATTGCTCA AAGCCGGG	ACCTCAAGTGATC TGCCCAC	TTCCTTTGGTGTGTG TGCCT
EphB4-02	GTGCCAGTGGA ATGGAGGA	CCACAACCTCCAC CTCTGCAA	TCCCAAAAGCCCA TAAGAGGAC	ACCGCTACCCATTTG GATCTC
EphB4-03	CCTTTCTTCCCC CTCCCCTA	GCTGCCCTAAG GTCACAAA	TGCCCCATAGTTT TGCCCAT	TCCTTCCCAAAAAGA TGTTGGT
EphB4-04	GTGGAGGTTCTT TTGGGGCT	ACCAGCCTAGAC AACATGGC	CCCCAACTCTGAG TCTGAGC	CCCCAATCTGCCCTA ACTCC
EphB4-05	AGGTTAGAAGCC AAGCCCTG	AGCCCAGGAACA GTAGGCTA	CTGCTACCTGCTC TTATGGCA	TCACACAACGATGAA ATCCCT
EphB4-06	GAGTAGCTGGGA TTACGGGC	TGGAGTGCAGTG GTGTGATC	GAGTAGCTGGGAT TACGGGC	AGCTCAGGAGTTCCA GACCA
EphB4-07	TGGTCAGGCTGG TCTTGAAC	AGGACAAGTCAC TGCTGCTG	CTGACCTCAGCCT CCCAAAG	TCATGTGCCATTATT GCCAC
EphB4-08	CTCAGCCTCCCA	AAATGGAGGCTG	ACCCAGCCTCCAT	AAATGGAGGCTGAGG

	AGTAGCTG	AGGGAACG	CAGAGAT	GAACG
EphB4-09	TCACAGGATTCT GGGCCTTG	CAAGTCCGGCTC TTGCTTCT	GCCTTGCTAACAT GAGCCAG	TGCGACCTCAAGTAA GTTGCT
EphB4-10	AACTTGTTGGCAC TCTAGGCA	TTTCACATGTGT CTCCCCGG	ACACAATGCATTG ATCGAGTTCT	ACACTGTGCCCTGGT CAAAT
EphB4-11	AGTGGATCCTCT GGGTTGGT	TTAGCTGGATGT GGTGGCAG	GGTTGGTTTGTCA GCAGCAG	GAGCTGGGTACTGGT TACACA
EphB4-12	CACTTACACAGA CGGGCAGT	GGGTCACTTTCC CCTCTTCG	AACTGTACTGAAA CTGAACTCTTTC	CAGCCCCTTCGGAGT CTTTT
EphB4-13	TACCAGGCTTTG CTTGAGGG	ACCATGTGGAAA TGTCAACCT	TACCAGGCTTTGC TTGAGGG	ACCATGTGGAAATGT CAACCT
EphB4-14	GCAGTGTGGTG GCAGATTC	TTTGGCGGACAG GAAGGTAC	ATTGTTTGCCT TACAGAGTTT	TGTCCAAAAGTAAA AACCTGGT
EphB4-15	CTCCTGACCTCG TGATCTGC	CTGGGAGGGAGG CAGAAATG	GGTCTCAAACCTCT GGGCCTC	TGGGGATTGTTTTGG GGATGA
EphB4-16	TAGGCCCTCCA ACACTGTA	AGGAAAATGGGC ATGGGCTT	TGCATCTGTTTCAT TTATTTTTGGCT	AGGAAAATGGGCATG GGCTT
EphB4-17	CAGAATGCAGTG ACCCCAGT	ACCATGCAAACA CTAGCCGA	CACCCTTCTGTCC CCATCAC	AGAAAGAACTAGGTG GATCCCT
EphB4-18	AAAGCATCACTC TCCAGCCC	TCACCATGCCCG GCTAATTT	GTCCCTCAAACG TCCCTG	TCACCATGCCCGCT AATTT
EphB4-19	GGAGGGAGGGGA AGACAGAT	CCTGAGACAAGA GCCACTGG	CTGAGGGGCATCA GGGAAAG	TCCCAGCATAGGACT CACCA
EphB4-20	TACATGTGCCAT GCTGGTGT	TGGAGTGCAGTG GTGTGATC	CCCAGAGTGTGAT GTTCCCC	GGTCTTGAACCTCTG GCCTC
EphB4-21	AAGCATCTGTAA CCCCAGCC	ATTTTGGGAGGC TGAGGTGG	TGCGCAACAGAGT GAGACTC	TGGGTGACAGAGCAA GACTC
EphB4-22	GGAGGGTTGCTC CACATTCT	TGAAGTGGTCGT GATCTTAAATGT	TGGCTGATAAGGC TGAGTGT	ATGAAGAGGGAAGCT ACAGC
EphB4-23	GCAGCTAAAGGG GTCCAGAG	TAACCATGCTCC TGCCACTG	AGGGGTGAGACTA AGAGCCC	CCACAGTGAACCTCT CAAAACC
EphB4-24	TGCCATTCAAAT GCAAGGCA	CTCTCTACCCAC CCAAGAAGG	TTGTTGCAGTATA AGGTGAGGT	TGTGACAGGCTTCAG TACACT
EphB4-25	CCTCCTGGGTTT AAGCGATT	TTTCCTTGTGT GCCCCCTT	TGCAACTGTTCTC ACCTGGT	CAGTGACCTGCTTTC CGTCA
EphB4-26	GCAAAGGCAAAC TGAGGGAC	GGATCCACTGAT GATTTTGCCT	ATGAATGGGGTCT GTGGGTT	GCTTCCCCACTCTCT GAACC
EphB4-27	GGGTTGAGGAAG GTTGAGCA	CAGCTCACTGGA AAAAAGCCA	AGGTTGAGCAAGG TCAGTGA	TGATGTTTGTGGCCA TGAGC

EphB4-28	TTCCTTGGCTTC TGGTTGCA	ATGCTGTGCTCA CTCACGAA	CACCCCTTCAAGA CCTGGAC	ATGCTGTGCTCACTC ACGAA
EphB4-29	TTGGCCTCCTAA AGTGCTGG	AGATGTGGCATC AAGGTGCA	TGCCCCCTATATG ACAAATGCA	TGTAAGCAGAGACCC CCAGT
EphB4-30	GTGGTAGACAGT AGCTACCAGC	ATATGAGCACCC ACCAGCAC	AGGTTTAGCTTCC TTTAAAGGCA	TGAGTTTCTTTGCTG TCCTCCA
EphB4-31	AGGGATATAGGG GGCACTTGA	TGAGGAGTGCTC ATCAGGTA	AGACAAATCTCCC ATGCACA	TGGCCTTGATCAACT GGATGA
EphB4-32	TGGTGCCACTAG CCTGAATG	CTTTCCACTGCC ATCTGGGT	GCCCTTTCTGCAT ATGGGGA	CCATTTCAGGGCCCT TCCTT
EphB4-33	CCAAGAGGATTG AGTTGAGGCA	AGCCCACCAGAT TTTAGGAGA	TGGGTAAGTGCTT GTCCCTG	AACCGATGGTGATAG GTTTGT
EphB4-34	TGCCTCACGATT TGATCACGA	CCCAGTGGCAGA AAATCAGT	TCAGCCCCGTAAA ACACACA	TCTGCCAAATTTTAA TGTTTCCCCA
EphB4-35	GCAAGCTCTGAC TGGTGAGT	CAGGGACTCCAG GAAAGCAG	AGTGCTTTTTCCTC CTGGTCC	GAGGAATTTGCAGGA CCCCA
EphB4-36	GGGCAGACTGAC ACTTCACA	CGTGCCATGGTT TTCAGCTC	GGCTGGGTACTCC TCTGAGA	GGGTTTGGTGTTGGA TGTCC
EphB4-37	GCTCTTTGGGTA GCAGTGGT	GGGAAAAGTGTA GGGGCACA	TTCAGGCTCATGT GCTTGGT	ACGCTTTTGGGAATG TTGGA
EphB4-38	CAAGGCCCAAC TAGGTGTT	AGCTGGGCTGGA AGAAAGTC	CATTGCTCTGTGT TGCAAGCT	AGCTGGGCTGGAAGA AAGTC
EphB4-39	ACCACTTACTAC GCCAAGCC	CCTGATTGAGG GGATGGGT	ACCACTTACTACG CCAAGCC	AGGCTGACATGTTCT TAGACTCT

[표 6]

HPRT-Exon6에 대한 프라이머

Site No.	1st PCR_F	1st PCR_R	2nd PCR_F	2nd PCR_R
HPRT_E6-01	TCTTACTGCTTGC TGAGGGC	CTCACCCGCTTTG GAATCCT	GCAGTTATACATGGG GGTTTTGG	AGCAATCACTTAA TCCCCCTTCA
HPRT_E6-02	TGAGGCCTAGGCT GAGAAGT	GTGCTTCAGGGGA GACTGTC	GGATGGAAGGAATCA TGGAGCT	CCCCTGCCCATCC TGTTTTA
HPRT_E6-03	GAACGTTGTTTTTC CCTGGGC	TGAGCCAAGATCG TGCCATT	GCCCTCTTCAGTGAG CTTCA	ATTGCACTCTAGC CTGGACG
HPRT_E6-04	TGTGTGCCTCCAA CTAGCTG	CCACTTTACCCTC CAGCCAG	AGCTGTCTCCTTTGC CTTCC	TTGGCAAAGTGAG CAACTGC
HPRT_E6-05	GATTCTCCTGCCT CAGCCTC	TCTTATGTTGCC AGGCTGG	AGGTGCCGATTATCA TGCCA	GTGTTGTGCAACC ATCACCA
HPRT_E6-06	CCTGGACTTGAGG	CGCCACCACACC	GAAGGTGACCTGGTA	TGGATGGATGACT

	CAGTGAG	TCTAATT	GGACA	TTGTCTACTT
HPRT_E6-07	TGGATTAGGGGCC CACCATA	ACGCCCAGCCTGT TTAACAT	AGCCAGCAGGTCATG AAGTC	TTGAGCCCAGGAG TCTGAGA
HPRT_E6-08	TGGTCCACACGA AAGAAGG	TCCATCCACCTAG CCATCCA	TGTCAGGAACACAGG CACTT	ATGTGAAGCAAGG GACAGGC
HPRT_E6-09	CCCCACCCCAAA AAATAGC	ACACCATGATGG AAGCCAA	TGTTGCATGCCAAGT ATTGTGA	GCCTGCTCAGAGT CAAGGAA
HPRT_E6-10	GAATGCATGAAAC ACCCCGG	TCAAGGATGATTG CCATTGCT	TCCCACAAATTAGGC CAGGAC	TGATTGCCATTGC TTGTTGGA
HPRT_E6-11	CCAAGAGACCATG GGCAGTT	TGAGCCGAGATTG TGCCATT	AATAGCAGGCCCTCT TCAGC	ATTGCACTCTAGC CTGGACG
HPRT_E6-12	CACACCCAGCCAA GATGGAT	CCAACCTCCCAAT CCCTGAC	AGCAGGCTATATTTT GTGGTTGC	ACCTGCAGAACTT AGTGAAACA
HPRT_E6-13	ACACACACAGACG AAGGCAT	CGACCTTCCTTTC CCGTGA	GCGGCTCCGTTTTGT TTGTT	TGGACATGTCACC CAAGCTG
HPRT_E6-14	TCCAGAGTGTAAC CGGAAAGT	TCAACTGAAGGTG GCAGGAA	ATAGCCTATTGCCCC TCCCA	GCCTCTAACTCTT TGTCTGGCT
HPRT_E6-15	ATGTTGGCCAGGA TGGTCTG	ATGATGTTGTGCT GCGTTGG	TCAGAAGCCTGCACT CTTCA	TCAAAGGTCACAA CACTTGCT
HPRT_E6-16	TTCAACCTCTGTC CTGCTGC	AGGCCTGGTGCTA TGAGAGA	TTCACTGCTCCAATG TGTGT	AGGTGAGTTAAAG GGGGGAGA
HPRT_E6-17	TGAACGAGGGTCA GTTTGGC	CTGTGTCAGTGGC CATGTCT	GCCTTTGCCTTATAT TCGGC	GTCAAACAGGTG TACATGATGCA
HPRT_E6-18	TCTTACCAACTGT GCCTTCCC	AACTGGGAGGGAA AAGTGCC	TGTCACGTGACCCTC TTCAT	TGAGTGACTACAG TGAACGTTGT
HPRT_E6-19	GGCAGGTGACTTA AGTCCCC	GCACATGATCCTG TGTGTGC	GGCAGGTGACTTAAG TCCCC	GCACATGATCCTG TGTGTGC
HPRT_E6-20	TGGCAGAATGGAT TGGTTGAGA	TGGAGACAGAGTC ATTGCAAA	TGGTCCCTACAGAGC TCTGA	GAAGAGACAGAGA CACAGAGCA
HPRT_E6-21	CAAAGTTCCAACA AGGGCCC	CTTCATGCTGGAG GCTGTCA	ACCAACTCCTTGCTG TTGGG	AGTGAAAGAAGGG AGCCAGC
HPRT_E6-22	GAGACCAGGACAC ACAACGT	TCGGCTCTCTTCA TCACTGT	CAAAATCCATTGCCC AGGCG	TCGGCTCTCTTCA TCACTGT
HPRT_E6-23	AATCGCTTGAACC CAGGAGG	CCAACCCTCCTCC TCTCAGA	TGCTAACACATGCCT CCCTG	CCAACCCTCCTCC TCTCAGA
HPRT_E6-24	CTTCCCGGGTTGA TCCTCAC	ACCAAACGACTTG CCTGAGA	CCTGCAACAAGTTG TCTGGA	TGGTGGCGATGAT CCTTCAA
HPRT_E6-25	TTTCATTGTCTGC TGCCAGA	CGTACTCTAACG GCCCAAA	AGGGAACACAGTCCT TATTTGGA	TGGAAATCACATC AAACCAACTAA

HPRT_E6-26	TCTCTCCAATGGC AGCAGTG	AGGCTGACAGTAG TTGCTTCC	TCTCTCCAATGGCAG CAGTG	CAGACAACAAAGT GTGGCCA
HPRT_E6-27	ATGGTGGTCTTGG TCACAGC	AAGATTCCCACAC AGGAGCG	GGTCTCAAACCTCTG GCCTC	ATACTGACTGGCT GGCTCTG
HPRT_E6-28	TCTGGCTTAATGT GGGGTGG	ATGCTCCCATGAG GCCAAAT	ATTGAGGGGGCTGGC TAAAC	ACCCTTCATCATC ACAGCTCT
HPRT_E6-29	AACCCACACAGTA GCAGGTA	ACAGCCGCAGTCC TATGTTC	AGCTTGTAGTGACCA GGTGA	ACACTATGGGCCA ATGTGAACT
HPRT_E6-30	TGGTTCCTGCTGG GTTCTTC	ACCCACACTATA CCCCCTC	CCAGCGCCTGTTTCC AAAAG	GCACACAAGATTC CAGCAGC
HPRT_E6-31	CCCCACTGGTGAC CTTACAC	ATGAGGGAAGGGG AAGACCT	CCACTGCAGGCTCAG GAATA	ACAAACTGGGGGT GCTTCTG
HPRT_E6-32	TGGTCGCCTCAGT CAGTTTC	GGGAACAGCAAGT GCAAAGG	TTACAGGCGTGAGTC ACCAC	AAGCCCAGGAATT TCGACGT
HPRT_E6-33	AGGGGTTTCTGCC TCTTTGAG	ACACCACCCAGTT CCACTTG	AGTCTTATGGCGGGA ATGTTT	GGCTATGTACACA TATTGTCCCA
HPRT_E6-34	AGCCATAACCACA ATCCCGT	GGAGAGGAGAGAA AGCACCG	AGCCATAACCACAAT CCCGT	AATGTGACCCTTG TGAGGCC
HPRT_E6-35	TCCCTTGAGCCC TACAGAA	GCTTGAGAGAGCC CTTTGAT	TCCCTTGAGCCCTA CAGAA	AGAACAGGCATGC CAGCTTT
HPRT_E6-36	GTTTCTCCTGACT CGGCCTC	CACCAAAAGCACA GGCAACA	TGGCCTTCCAAAGTG CTGG	GGGGTGTGTAAAG GCAACCT
HPRT_E6-37	GAGGCTGAGGCAG GAGAATC	AGGCTGGAGATA GAGCAGG	GCAACAGAGTGAGAC CCCAT	AAAGCCGCTGGAC TATGCTT
HPRT_E6-38	CGCTAAATGAGGG CAAGGGA	TGTCACGTCCTCC TAGGGTT	CGCTAAATGAGGGCA AGGGA	GACACCTTCCATG TCTCTGCA
HPRT_E6-39	GAAGATGATGGAG ACCGGCC	TGGGCAGTGGTCA CAAATGT	CATTGGCATGACAGG GGAGA	ACCTTCCCAAAC ACACTCCT
HPRT_E6-40	TGGTACCCTGTC TCTGTGT	ATTCCATGAGTCC CCGAGGA	TGGTCCCTCAAGAGC TAGCT	ATTCCATGAGTCC CCGAGGA
HPRT_E6-41	GCGGATCATGAGG TCAGGAG	TCTGAGCCAGGGA GCTTTTG	GACAGAGTGAGACCC CGTCT	TGTTTTTAAACTG CAGCAAATGCT
HPRT_E6-52	ATCAGGCCTGGCA TAGCATG	CCTCCTGAGTAGC TGGGACT	ATCAGGCCTGGCATA GCATG	CGGCTTTGACTCC ACCTCTT
HPRT_E6-81	GACACCATCATGC AGGGTCA	GGTCTTGAACCTCC TGGCCTC	TGCATTAGGAGAAGC TGGTTT	TGTTATGTGCATT TGGCCACA

[표 7]

HPRT-Exon8에 대한 프라이머

Site No.	1st PCR_F	1st PCR_R	2nd PCR_F	2nd PCR_R
HPRT_E8-1	GGTGAAGGACC CCACGAA	GAAGTGTACCCCT AGCCTGG	AGAGAATATTTGT AGAGAGGCACA	TGCTGGAAGGAGA AAACAATTC
HPRT_E8-2	CCCTTGGTGTCAG TAATCCCA	TGGCTGACAATGT CCTGCAT	TGAGCTCTTATAT GTGCAAGGCA	CCCTGACCATTCT CACCACC
HPRT_E8-3	TCTCCACGACCAT CCAGTCT	ATGTACCAAGCAC TGTCCCG	TAGCCTTTCTGAA GCAGGGC	GCCAGTCTGACTC TTGAGCT
HPRT_E8-4	TTAGACCTCCACT GGGTCGT	GCAGAGAGCAACA GTGGACT	GTCGTGTTGCACC CCATTTA	TCCATCTGAAGCA GACTAGCC
HPRT_E8-5	TCTGGGTGCCTCC ATTGAAC	CTGCCTGATGTCC ACTCTCC	GGAGACAGACACA TGCATGGA	TGCTTTCCACAT CCTGCAT
HPRT_E8-6	TGAGTTTAGGATG GCAGCAACT	CAATGCAAACCTG GCCTTCC	CCAGATGTTCCCA AACTCTGC	TGCCAGATGGATA TGCACCA
HPRT_E8-7	GAACCAGACCTCA CCCCATG	CCTGTGGCTCTGT CATGTGT	GTCATCCAGTTGG TGCCAGA	CCTGTGGCTCTGT CATGTGT
HPRT_E8-8	CAGAGTGGATGGT GTGCTGT	TAAGGCAGTGTGA AGCAGCA	CAGAGTGGATGGT GTGCTGT	TTGCCAGAGAGG GTCTCCT
HPRT_E8-9	TGCACGCCAGTCA AAGAGAT	TGACCTGGCAATT CCACTCC	AGCATGTGTCTGT ACTTCATTGT	TGACCTGGCAATT CCACTCC
HPRT_E8-10	CCAGTTTGAGTT CTGTGCC	CCTGAAGCATCCT GACTGCT	TGTTGACAGTCAA AAGTGAGGA	CCTGAAGCATCCT GACTGCT
HPRT_E8-11	ACTCTGGAGGCTC AGGTAGG	ATTACAGGCGTGA CCCACTG	CGTTATCACATCT TCAGCAAAGC	ATTACAGGCGTGA CCCACTG
HPRT_E8-12	CCCGACCTCCAGG AAAAAGA	GGTGGTGGGTGTT CAGATGT	TGAGGGCCAAACC AGAAGAG	CTCATTTTGAAC AACCCTCCC
HPRT_E8-13	GTCTCAAACCTCGT GGGCTCA	ACATGTCTGGCCT GCACATT	TCCATTCTGCTA AGTGCCC	ACGACACTTGTGC ACCTTTT
HPRT_E8-14	TGTACCTCCTACC TGCTCCC	GTGTGTGTGTGTG TGTGTGT	CAAAGAAACAGAA CCAATTGCATAT	GCAGGTAGTTGTA ACACAATGCT
HPRT_E8-15	TCTTCGGACTGAA CAAGCATCT	TTTCAGATGAAGG CCGACCC	CAGCCATGCCACT AAAGAGC	GAGGTTTCTTTCA GCAGCCC
HPRT_E8-16	GCCCATGCAATTG CTTAGGG	CACTCCAAGGACA GTCCACC	TCCCTTGGGATTT CAGGCTC	GCATGAGCCAACT GCATTGT
HPRT_E8-17	CCCACAAGCTTCA TTCATGGC	GCCTCTCAGGTTC ATGCCAT	CCCACAAGCTTCA TTCATGGC	ACGGACTGAGTTG GTACAGC
HPRT_E8-18	TTGGGGGTCTCAT TTGCCTC	CCCTTACCCCTTC TCTCCCA	TCCTTCTGCCTGA AAACAGACA	CCAAGGGCAATTT CCAGTGG
HPRT_E8-19	CACCACTGCACTC TACCCTG	CAAGACCAGCCTA GGCAACA	CACCACTGCACTC TACCCTG	ACTTGTGGATTTT GCATATAGCCT

HPRT_E8-20	AGTAGAGCTCCAG GGAAGGG	GCCAAGAGCTCTG GAACCTT	AGTAGAGCTCCAG GGAAGGG	GCCCCTTCACTTG TTCCTCA
------------	--------------------------	--------------------------	--------------------------	--------------------------

paired-end sequencing system을 갖춘 Illumina MiniSeq을 사용하여, 상기 생성된 라이브러리를 시퀀싱하였다. 염기교정빈도 (Base editing frequencies)는 editing window (position 4-7) 내의 하나 이상의 edit로 교정된 변형 표적 부위의 빈도를 나타낸다. 치환 및 indel 빈도를 계산하기 위한 컴퓨터 프로그램 소스 코드는 'https://github.com/ibs-cge/maund'에서 확인할 수 있다.

실시에 5. ABE7.10, BE3, 및 Cas9의 mismatched sgRNAs에 대한 허용 오차 (Tolerance)

내재 유전체 부위 (endogenous genomic loci)를 표적으로 하는 일련의 mismatched sgRNAs를 사용하여, ABE7.10, BE3 (cytosine base editor) 및 Cas9의 off-target 활성이 서로 다른지 여부를 조사하였다. 보다 구체적으로, 표적 부위 (on target site)에 대하여 1 내지 4 개의 불일치 염기를 포함하는 sgRNA를 코딩하는 플라스미드와 인간 코돈에 최적화된 ABE7.10, BE3 (Base Editor plasmid; Addgene plasmid #73021; rAPOBEC1-XTEN-Cas9n-UGI-NLS) 또는 SpCas9(NP_269215.1)를 코딩하는 플라스미드를 HEK293T 세포에 트랜스펙션하고 (참고예 1 참조), 참고예 6의 표적 심층 시퀀싱을 수행하여, 염기 교정 빈도 및 삽입/결실 (Indel) 빈도를 측정하였다.

상기 얻어진 염기교정 빈도 (%) 및 삽입/결실 빈도(%)를 도 6a 내지 6c, 및 표 8 내지 표 10에 나타내었다:

[표 8]

표적 유전자: HEK2(on target site: GAACACAAAGCATAGACTGCGGG)

sgRNA 표적 부위	(+) ABE7.10		(+) Cas9		(+) BE3	
	Base editing frequency (%)	Relative activity	Indel frequency (%)	Relative activity	Base editing frequency (%)	Relative activity
GggtgCAAAGCATAGACTGCGGG	1.33	0.02	0.03	0.00	4.53	0.05
GAACgtggAGCATAGACTGCGGG	12.76	0.17	30.83	0.48	15.00	0.18
GAACACAgatgTAGACTGCGGG	6.48	0.09	0.30	0.00	1.18	0.01
GAACACAAAGCAcgagCTGCGGG	0.70	0.01	0.00	0.00	0.18	0.00

GAACACAAAGCATAGAtcatGGG	0.86	0.01	0.03	0.00	0.19	0.00
GAgtgCAAAGCATAGACTGCGGG	2.40	0.03	2.60	0.04	13.68	0.16
GAACAtggAGCATAGACTGCGGG	75.71	1.02	76.37	1.18	74.30	0.88
GAACACAAgatATAGACTGCGGG	22.45	0.30	0.64	0.01	7.76	0.09
GAACACAAAGCgcgGACTGCGGG	1.33	0.02	0.00	0.00	0.75	0.01
GAACACAAAGCATAagtTGCGGG	1.16	0.02	0.05	0.00	0.19	0.00
GAACACAAAGCATAGACcatGGG	5.12	0.07	0.02	0.00	0.52	0.01
GAgtACAAAGCATAGACTGCGGG	61.71	0.83	63.07	0.97	69.18	0.82
GAACgtAAAGCATAGACTGCGGG	57.23	0.77	73.27	1.13	75.35	0.89
GAACACggAGCATAGACTGCGGG	76.76	1.04	82.69	1.28	80.95	0.96
GAACACAAgaCATAGACTGCGGG	68.19	0.92	69.70	1.08	71.33	0.85
GAACACAAAGtgTAGACTGCGGG	77.93	1.05	45.78	0.71	76.05	0.90
GAACACAAAGCacgGACTGCGGG	3.01	0.04	0.03	0.00	1.10	0.01
GAACACAAAGCATAagCTGCGGG	49.41	0.67	1.68	0.03	23.86	0.28
GAACACAAAGCATAGAtcGCGGG	53.12	0.72	0.96	0.01	14.18	0.17
GAACACAAAGCATAGACTatGGG	33.85	0.46	0.54	0.01	4.27	0.05
GgACACAAAGCATAGACTGCGGG	77.99	1.05	56.26	0.87	80.99	0.96
GAAtACAAAGCATAGACTGCGGG	77.47	1.04	56.28	0.87	81.51	0.97
GAACAtAAAGCATAGACTGCGGG	75.90	1.02	60.44	0.93	77.98	0.93
GAACACAgAGCATAGACTGCGGG	78.61	1.06	52.17	0.81	80.94	0.96
GAACACAAAaCATAGACTGCGGG	77.42	1.04	43.30	0.67	77.51	0.92
GAACACAAAGCgTAGACTGCGGG	75.62	1.02	53.51	0.83	81.23	0.96
GAACACAAAGCATgGACTGCGGG	67.88	0.92	11.14	0.17	66.25	0.79
GAACACAAAGCATAGgCTGCGGG	78.22	1.05	28.65	0.44	74.33	0.88
GAACACAAAGCATAGACcGCGGG	79.11	1.07	47.36	0.73	81.18	0.96
GAACACAAAGCATAGACTGtGGG	79.86	1.08	25.81	0.40	73.01	0.87
GAACACAAAGCATAGACTGCGGG	74.14	1.00	64.70	1.00	84.20	1.00
no sgRNA	0.63	0.01	0.00	0.00	0.07	0.00

(상기 표에서,

말단의 'NGG' 부분은 PAM 서열이고;

sgRNA 서열은 표적 부위 서열 중 PAM 서열을 제외한 서열에서 T를 U로 변환한 서열이고;

- 5 on-target sequence에 대한 mismatched base는 소문자로 표시함;
Relative activity = (mismatched sgRNA 결과값/matched sgRNA 결과값);

하기 모든 표에서 동일하게 적용됨)

[표 9]

표적 유전자: HEK3 (on target site: GGCCCAGACTGAGCACGTGATGG)

sgRNA 표적 부위	(+) ABE7.10		(+) Cas9		(+) BE3	
	Base editing frequency (%)	Relative activity	Indel frequency (%)	Relative activity	Base editing frequency (%)	Relative activity
GatttAGACTGAGCACGTGATGG	0.32	0.01	0.00	0.00	0.17	0.00
GGCCtgagCTGAGCACGTGATGG	0.41	0.01	0.03	0.00	0.13	0.00
GGCCCAGAtcagGCACGTGATGG	0.48	0.01	0.05	0.00	0.20	0.00
GGCCCAGACTGAatgtGTGATGG	0.31	0.01	0.09	0.00	0.12	0.00
GGCCCAGACTGAGCACacagTGG	0.39	0.01	0.02	0.00	0.25	0.00
GGtttAGACTGAGCACGTGATGG	0.33	0.01	0.02	0.00	0.15	0.00
GGCCCgagCTGAGCACGTGATGG	5.23	0.13	1.15	0.02	16.86	0.22
GGCCCAGAtcaAGCACGTGATGG	2.18	0.05	0.00	0.00	0.21	0.00
GGCCCAGACTGgatACGTGATGG	0.56	0.01	0.01	0.00	0.27	0.00
GGCCCAGACTGAGCgtaTGATGG	0.79	0.02	0.01	0.00	0.18	0.00
GGCCCAGACTGAGCACGcagTGG	1.54	0.04	0.18	0.00	0.22	0.00
GGttCAGACTGAGCACGTGATGG	6.02	0.15	49.87	0.66	2.11	0.03
GGCCtgGACTGAGCACGTGATGG	40.39	1.01	61.63	0.82	49.37	0.65
GGCCCAagCTGAGCACGTGATGG	12.81	0.32	58.23	0.77	6.68	0.09
GGCCCAGAtcGAGCACGTGATGG	15.88	0.40	46.57	0.62	12.28	0.16
GGCCCAGACTagGCACGTGATGG	17.80	0.45	19.65	0.26	3.98	0.05
GGCCCAGACTGAatACGTGATGG	0.46	0.01	0.04	0.00	0.19	0.00
GGCCCAGACTGAGCgtGTGATGG	3.63	0.09	1.66	0.02	0.64	0.01
GGCCCAGACTGAGCACacGATGG	9.28	0.23	0.51	0.01	1.53	0.02
GGCCCAGACTGAGCACGTagTGG	14.31	0.36	21.05	0.28	10.61	0.14
GaCCCAGACTGAGCACGTGATGG	50.32	1.26	62.84	0.83	67.85	0.89
GGCtCAGACTGAGCACGTGATGG	37.32	0.94	66.54	0.88	50.80	0.67
GGCCCgGACTGAGCACGTGATGG	39.21	0.98	76.56	1.02	74.92	0.98
GGCCCAGgCTGAGCACGTGATGG	41.09	1.03	63.21	0.84	71.77	0.94
GGCCCAGACcGAGCACGTGATGG	38.99	0.98	53.99	0.72	60.41	0.79
GGCCCAGACTggGCACGTGATGG	64.05	1.61	57.72	0.77	72.00	0.94
GGCCCAGACTGAGtACGTGATGG	25.02	0.63	41.86	0.56	23.39	0.31
GGCCCAGACTGAGCAtGTGATGG	30.04	0.75	47.29	0.63	43.94	0.58
GGCCCAGACTGAGCACGcGATGG	48.08	1.21	61.45	0.81	66.11	0.87
GGCCCAGACTGAGCACGTGgTGG	43.88	1.10	60.87	0.81	72.75	0.95
GGCCCAGACTGAGCACGTGATGG	39.87	1.00	75.41	1.00	76.33	1.00
no sgRNA	0.21	0.01	0.00	0.00	0.16	0.00

[표 10]

표적 유전자: RNF2 (on target site: GTCATCTTAGTCATTACCTGAGG)

sgRNA 표적 부위	(+) ABE7.10		(+) Cas9		(+) BE3	
	Base editing frequency (%)	Relative activity	Indel frequency (%)	Relative activity	Base editing frequency (%)	Relative activity
GctgcCTTAGTCATTACCTGAGG	0.13	0.00	0.01	0.00	0.27	0.00
GTCActccAGTCATTACCTGAGG	0.12	0.00	0.04	0.00	0.60	0.01
GTCATCTTgactATTACCTGAGG	0.12	0.00	0.01	0.00	0.11	0.00
GTCATCTTAGTCgccgCCTGAGG	0.10	0.00	0.02	0.00	0.20	0.00
GTCATCTTAGTCATTAttcaAGG	0.14	0.00	0.01	0.00	0.16	0.00
GTtgcCTTAGTCATTACCTGAGG	0.15	0.00	0.01	0.00	1.13	0.02
GTCATtccAGTCATTACCTGAGG	0.10	0.00	0.03	0.00	0.65	0.01
GTCATCTTgacCATTACCTGAGG	0.12	0.00	0.00	0.00	0.23	0.00
GTCATCTTAGTtgTACCTGAGG	0.12	0.00	0.06	0.00	0.27	0.00
GTCATCTTAGTCATcgtCTGAGG	0.14	0.00	0.01	0.00	0.16	0.00
GTCATCTTAGTCATTActcaAGG	0.10	0.00	0.01	0.00	0.20	0.00
GTtgTCTTAGTCATTACCTGAGG	1.22	0.03	12.77	0.23	47.48	0.66
GTCActTtagTCATTACCTGAGG	0.22	0.01	0.23	0.00	15.33	0.21
GTCATCccAGTCATTACCTGAGG	1.44	0.03	7.76	0.14	34.05	0.47
GTCATCTTgaTCATTACCTGAGG	0.20	0.00	0.10	0.00	5.97	0.08
GTCATCTTAGctATTACCTGAGG	0.38	0.01	5.16	0.09	2.19	0.03
GTCATCTTAGTCgcTACCTGAGG	0.21	0.00	2.74	0.05	1.67	0.02
GTCATCTTAGTCATcgCCTGAGG	0.17	0.00	0.02	0.00	0.33	0.00
GTCATCTTAGTCATTAttTGAGG	0.14	0.00	0.02	0.00	0.13	0.00
GTCATCTTAGTCATTACCcaAGG	0.12	0.00	0.08	0.00	0.37	0.01
GcCATCTTAGTCATTACCTGAGG	18.06	0.41	46.51	0.85	65.55	0.91
GTCgTCTTAGTCATTACCTGAGG	14.01	0.32	49.31	0.90	70.67	0.98
GTCATtTtagTCATTACCTGAGG	3.53	0.08	51.54	0.94	56.95	0.79
GTCATCTcAGTCATTACCTGAGG	9.98	0.23	41.89	0.76	62.12	0.86
GTCATCTTaaTCATTACCTGAGG	1.37	0.03	33.09	0.60	16.41	0.23
GTCATCTTAGTtATTACCTGAGG	23.31	0.53	43.94	0.80	60.08	0.83
GTCATCTTAGTCaCTACCTGAGG	4.99	0.11	29.11	0.53	25.20	0.35
GTCATCTTAGTCATTgCCTGAGG	11.45	0.26	35.75	0.65	36.77	0.51
GTCATCTTAGTCATTActTGAGG	13.85	0.31	23.51	0.43	19.42	0.27
GTCATCTTAGTCATTACCTaAGG	8.98	0.20	27.32	0.50	31.08	0.43
GTCATCTTAGTCATTACCTGAGG	44.00	1.00	54.83	1.00	72.17	1.00

no sgRNA	0.13	0.00	0.03	0.00	0.12	0.00
----------	------	------	------	------	------	------

도 6a 내지 6c 및 표 8 내지 10에 나타난 바와 같이, Base editor 및 Cas9 뉴클레아제가 on target site에 대하여 1개 또는 2개 염기의 mismatch를 갖는 sgRNA와 함께 사용되는 경우 대부분 활성을 유지하지만 (1개 또는 2개 염기의 mismatched sgRNA에 대하여 관용(tolerance)을 가짐), 3개 또는 4개의 불일치를 갖는 sgRNA와 결합한 경우에는 대부분 활성이 현저하게 감소하였다. 특히, Cas9, ABE7.10 및 BE3의 mismatched sgRNA에 대한 관용은 서로 상이하게 나타났다. 도 6a 및 표 8에 나타난 바와 같이, HEK2 유전자의 표적 부위의 경우, 2 내지 4 개 염기의 mismatch를 갖는 sgRNA 사용시, ABE7.10는 비교적 높은 관용을 갖지만 (활성 유지), Cas9 및 BE3는 활성이 낮거나 불활성이다. 도 6b 및 표 9에 나타난 바와 같이, HEK3 유전자의 표적 부위의 경우, Cas9는 3-4 위치 (5'에서 3'방향으로의 1에서 23까지의 numbering에 따름)에서 tandem mismatch를 갖는 sgRNA와 함께 (복합체 형성) 사용시 높은 활성을 나타내었다 (fully-matched sgRNA를 사용하는 경우와 비교하여 66%의 활성을 나타냄). 반면, ABE7.10의 경우에는 상기와 동일한 sgRNA와 복합체를 형성하는 경우, fully-matched sgRNA를 사용하는 경우와 비교하여 15%의 비교적 낮은 상대적 활성을 나타내었다. 이러한 결과는 ABE7.10, BE3 및 Cas9가 인간 유전체에서 개별 세트의 비표적 부위를 인식함을 제안하고, unbiased 방식으로 유전체 전반의 ABE 특이성을 결정할 수 있는 방법이 요구됨을 시사한다.

실시예 6. genome-wide off-target site 확인 (HEK2 gene)

시험관내(*in vitro*) 생성된 DNA 이중가닥절단 (DSB)에 의존하는 Digenome-seq을 통해 ABE7.10의 유전체 전역의 표적 부위를 확인하기 위하여, 아데닌-디아미네이션(adenine deamination)에 의하여 생성된 이노신(inosine)을 포함하는 부위에서 DSB를 생성시켰다. *E. coli* Uracil DNA glycosylase (UDG)와 DNA glycosylase-lyase Endonuclease VIII (Endo VIII)의 혼합물인 Uracil-Specific Excision Reagent (USER) (New England Biolabs; <https://www.neb.com/products/m5505-user-enzyme>; Uracil DNA glycosylase (UDG) 및 DNA glycosylase-lyase Endonuclease VIII 혼합물과 50 mM KCl, 5 mM NaCl, 10 mM Tris-HCl (pH 7.4), 0.1 mM EDTA, 1 mM DTT, 175 mg/ml BSA 및 50%(w/v) glycerol 포함) 및 재조합 BE3 delta UGI

단백질 (His6-rAPOBEC1-nCas9 단백질; UGI 도메인을 결여한 BE3 변이형; pET28b-BE1(Addgene plasmid #73018)에서 site directed mutagenesis를 이용하여 dCas9의 A840을 H840로 치환하여 제조)를 사용하여, 사이토신의 디아미네이션 결과물인 우라실을 포함하는 부위에서 DSB를 생성시켜 (*in vitro*), genome-wide BE3 off-target site을 확인하였다 (in an unbiased manner). 이와 유사하게, *E. coli* 엔도뉴클레아제 V (Endo V) 또는 인간 알킬아데닌 DNA 글라이코실라아제 (human Alkyladenine DNA Glycosylase; hAAG)와 Endo VIII의 조합이 이노신 포함 부위에서 포스포다이에스테르 결합을 절단하는데 사용될 수 있음을 추론하였다 (도 7a 내지 7c 참조).

10 Endo V는 DNA에서 이노신을 인식하고 이노신에 대하여 두 번째 phosphodiester bond 3'를 절단하는 반면, hAAG는 이노신을 소거하여 AP 부위 (apurinic/apyrimidinic site)를 생성하고 이 부위는 Endo VIII에 의해 인식 및 프로세싱되어 단일가닥 절단 (single-strand break)을 유도한다.

15 상기와 같은 추론을 시험하기 위하여, 표적 DNA 서열 (*HEK2* on-taeger sequence; 표 8 참조)을 포함하는 PCR amplicon을 *E. coli*에서 발현 및 정제된 재조합 ABE7.10 단백질 (정제 결과는 도 8 참조) 및 이들의 sgRNA로 처리하여 하나의 DNA 가닥에서 아데닌 디아미네이션을 촉매하고 다른 DNA 가닥에서는 nick을 생성하고(도 7a), 그 후, Endo V (도 7a 참조)

20 또는 hAAG/Endo VIII (도 7c 참조)로 처리하여, composite DSB를 생성시켰다. 상기 결과를 도 7d 및 도 7e에 나타내었다. 예상과 같이, PCR 증폭물은 ABE7.10 + Endo V (도 7d) 또는 ABE7.10 + hAAG/Endo VIII (도 7e)에 의해 절단됨을 확인할 수 있다.

 inosine-specific DNA repair enzyme을 사용하여, 이노신을 포함하는

25 ABE-표적 부위에서 DSB가 생성될 수 있음을 확인한 후, HEK293T 세포로부터 분리된 인간 유전체 DNA를 ABE7.10 + Endo V 또는 ABE7.10 + hAAG/Endo VIII로 *in vitro* 처리하고 Digenome-seq 분석하여 유전체 전역의 ABE off-target site를 확인할 수 있는지 여부를 조사하였다. EK293T 세포로부터 분리된 인간 유전체 DNA를 ABE7.10 ribonucleoprotein (RNP) (300 nM

30 ABE7.10 and 900 nM sgRNA; targeted to an endogenous chromosomal site (*HEK2* on-taeger sequence; 표 8 참조))와 함께 8 시간 동안 배양한 후, repair enzyme (Endo V 또는 hAAG/Endo VIII)과 함께 2시간 더 배양하였다. Sanger 시퀀싱을 이용하여 표적 부위 (nick을 갖는 반대 가닥은 시퀀싱

불가함)에서의 ABE-매개 adenine-to-inosine conversion (PCR 증폭동안 이노신은 구아닌으로 변경됨) 및 이노신-포함 부위 (nicked DNA 가닥은 증폭 불가함)에서의 Endo V-촉매 DNA 절단을 확인하였다 (도 7f). 유전체 DNA를 ABE7.10과 Endo V로 처리하면 이노신을 포함하는 DNA 가닥이 엔도 V에 의해 절단된다. Endo V가 절단한 DNA 가닥은 PCR로 증폭될 수 없지만 이노신을 갖지 않는 DNA 가닥은 증폭된다. 그 결과, Sanger sequencing에 의해 관찰된 교정 부위에서는 아데닌만 관찰되고 구아닌은 관찰되지 않는다 (도 7f).

또한 ABE7.10 + Endo V를 사용한 표적 DNA 절단을 검증하기 위해 정량적 실시간 PCR을 수행하여, 그 결과를 도 7g에 나타내었다. Endo V 대신 hAAG 및 Endo VIII를 사용한 경우에도 유사한 결과가 얻어졌다 (도 7h 및 도 7i). 얻어진 유전체 DNA 분해 샘플에 대하여 전체 유전체 시퀀싱 (WGS)을 수행하였다. human reference genome (hg19)에 대하여 sequencing reads를 맵핑한 후, Integrative Genomics Viewer (IGV)를 사용하여 표적 부위 (on-target site)에서의 sequence reads의 정렬을 모니터링하고, sequence reads가 무작위 정렬보다는 DSB의 특성인 일정한(uniform) 정렬을 나타냄을 확인하였다 (도 7j 및 도 7k). 이러한 정렬은 ABE7.10 단독 유전체 DNA 또는 미처리 유전체 DNA에서는 관찰되지 않았다 (도 7j 및 도 7k). 이러한 결과는 Digenome-seq이 유전체 범위의 ABE 표적 (on-target) 부위 및 비표적 (off-target) 부위를 포괄적으로 매핑하는데 사용될 수 있음을 보여준다.

인간 유전체에서 ABE7.10 off-target sites를 결정하기 위하여, WGS 데이터를 사용하여 유전체 전역에 걸쳐서 각각의 염기쌍 위치에 대한 DNA 절단 점수 (DNA cleavage score)를 할당하고 (도 9 참조), *in vitro* cleavage sites에 해당하는, 높은 점수를 갖는 부위를 나열하였다 (초기 컷오프 점수는 2.5로 설정됨(Digenome 1.0에 따름; Kim, D., Kim, S., Kim, S., Park, J. & Kim, J.S. Genome-wide target specificities of CRISPR-Cas9 nucleases revealed by multiplex Digenome-seq. *Genome research* 26, 406-415 (2016); Digenome 1.0 프로그램은 서열 상동성을 고려하지 않음)). Endo V를 사용하여 17개의 *in vitro* cleavage sites을, hAAG/Endo VIII를 사용하여 18 개의 *in vitro* cleavage sites을 얻었다. 표적 부위를 포함한 16 개의 부위가 서로 다른 두 가지 시험 (Endo V 사용 vs. hAAG/Endo VIII를 사용)에서 공통적으로 확인되었으며(도 10 a 및 도 10b), 이는

Digenome-seq 분석이 높은 재현성과 신뢰성을 가짐을 보여준다. 이들 부위의 DNA 서열을 비교하여 얻은 Sequence logos는 거의 모든 뉴클레오타이드 위치가 ABE7.10 디아미나제의 특이성에 기여함을 보여준다 (도 10c 참조).

5

실시예 7. 다양한 유전자 표적시의 genome-wide off-target site 확인

실시예 6에서 사용된 HEK2 유전자 표적 ABE7.10 RNP에 더하여, 추가의 6 개 ABE7.10 RNP를 각각 사용하여, 유전체 DNA에 대하여 Digenome-seq (Digenome-seq 1.0: DNA cleavage score가 2.5 이상인 site를 비표적 위치 후보군으로 결정함; (DNA cleavage score가 0.1 이상이며 10개 이하의 미스매치를 가지고 PAM (5'-NGN-3' 또는 5'-NNG-3')을 가지고 있는 site를 비표적 위치로 결정함)을 수행하여, 그 결과를 표 11 내지 표 18 및 도 11에 나타내었다.

15

[표 11]

The number of ABE on-target and off-target sites identified by Digenome-seq

표적 유전자	On-target DNA sequence	Number of in vitro cleavage sites	
		Digenome 1.0	Digenome 2.0
<i>HEK2</i>	GAACACAAAGCATAGACTGCGGG	17	152
<i>RNF2</i>	GTCATCTTAGTCATTACCTGAGG	5	50
<i>TYRO3</i>	GGCCACACTAGCGTTGCTGCTGG	2	12
<i>WEE1</i>	GTCGAGATGTTCTATTACTCTGG	1	21
<i>EphB4</i>	GCAGAATATTCGGACAAACACGG	3	39
<i>HPRT-Exon6</i>	GTATAATCCAAAGATGGTCAAGG	28	148
<i>HPRT-Exon8</i>	GAAGTATTCATTATAGTCAAGGG	1	20
Average $\pm$ S.E.M		8 $\pm$ 4	60 $\pm$ 20

[표 12]

20

Digenome-captured sites (HEK2)

Site No.	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites
HEK2_001	chr19	35505485	20.1	GAACAC-AAGCAcAGACTGaAGG
HEK2_002	chr15	93557679	15.1	GAACACA-tGCATAGACTGCTAG

HEK2_003	chr4	53536209	14.7	GAAtACtAAGCATAGACTcCAGG
HEK2_004	chr4	90522183	14.3	GAACACAAtGCATAGAtTGCCGG
HEK2_005	chr1	167742859	13.4	aAACACAgAGCAcAGACTGCTGA
HEK2_006	chr10	43159340	13.4	aAACACAAAGaCATAGACcaCTGG
HEK2_007	chr5	87240613	8.9	GAACACAAAGCATAGACTGCGGG
HEK2_008	chr9	290167	6.6	aAACAtAAAGaATAGACTGCAAG
HEK2_009	chr3	28560566	5.8	GAACACAAAGtATAGaATGCTAG
HEK2_010	chr1	175310138	5.1	GgACACAAAGCtTAGACTcCAGG
HEK2_011	chr13	87616351	4.9	GAAgAtAAAGCATAGACTctAGG
HEK2_012	chr10	87125365	4.6	aAACAtAAAGCATAGACTGCaAAG
HEK2_013	chr18	32279782	3.9	aAAtACAAAGCATAGACTaatATG
HEK2_014	chr2	192248363	3.5	GAACACAtA-CATAGACaGCTGG
HEK2_015	chr5	160202385	3.0	GAAaACAAAGCAaAGaAaGCAGG
HEK2_016	chr15	63236612	2.9	aAcacaAtAGCATAGACTGgaCTG
HEK2_017	chr20	23101377	2.7	atACACAgAGCAaAGACTGCAGG
HEK2_018	chrX	48317375	2.5	aAACAtAAAGCcTAGACTGaCGG
HEK2_019	chr19	49972515	2.2	aAAaACAAAGCgTAGACTGtGGG
HEK2_020	chr12	81785997	2.1	GgACaAaAAAGCATAGACTcaAGG
HEK2_021	chr1	40461424	2.1	atACACACAAGCAcAGACTGCAGG
HEK2_022	chr15	92968604	2.1	GAACACAgcaCATAaACTGCAGG
HEK2_023	chr5	142142325	2.0	aAACACtAA-CATAGACTGCAGG
HEK2_024	chr5	131174461	1.7	aAAtACAAtGCATAGACTGCTAG
HEK2_025	chr1	36097072	1.4	GtAaACAAAGCATAGACTGaGGG
HEK2_026	chrX	132724931	1.3	GAACA-tAAtCAcAGACTGCTGG
HEK2_027	chr6	139353017	1.3	ccAaACAAAaCATAGACTGCTGG
HEK2_028	chrX	36949816	1.3	GAAaACAAAaCATAGAgTGCTGG
HEK2_029	chr17	2670110	1.2	aAtCAaAtAGCATAGACTGCATG
HEK2_030	chr13	91254877	1.2	GAACACAAAcatagtACTGaAGG
HEK2_031	chrX	138877515	1.2	aAACAgAAAGCATgGACTGCGGA
HEK2_032	chr13	113428466	1.1	cAAtACAAAGgATAGACTGCAGG
HEK2_033	chr15	41796089	1.1	agACACAcAGCAcAGACTGCAGG
HEK2_034	chr11	128508576	1.0	GAAttCAAAGCATAGAtTGCAGG
HEK2_035	chr11	131870604	0.9	acACACAAAGCATAGACTatGTG
HEK2_036	chr4	91844238	0.9	GgAaACAAAGCATAGACattTGG
HEK2_037	chr3	105801924	0.8	tAAtACAAAGCATAGAtaGtTGG
HEK2_038	chr8	66280643	0.8	GAAtttAAAGCATAGACTGCAAG
HEK2_039	chr3	168165117	0.8	ctgaACAcAGCAaAGACTGCTGG
HEK2_040	chr15	65377019	0.8	GAgCgatAAGCAcAGACTGCTGG

HEK2_041	chr9	97332608	0.7	GtAattAAAGCAcAGACTGCTGG
HEK2_042	chr3	33744669	0.7	GAAaACAAAcCAcAGACTGgGGG
HEK2_043	chr2	19474362	0.7	ctAttgcAAGCAcAGACTGCTGG
HEK2_044	chr16	64198277	0.7	GAAtACAtAaCATAGACTGgGGG
HEK2_045	chr1	199217096	0.7	atACcatAAGCATAGACTGtTGG
HEK2_046	chr22	28895717	0.7	attaAgAtAGCATAGACTGCAGG
HEK2_047	chr13	94827431	0.6	aAtgACtAAtaATAGACTGCTGG
HEK2_048	chr20	895768	0.6	atttcaAcAGCATAGACTGtAGG
HEK2_049	chr1	108869934	0.5	agtatagcAGCATAGACTGCAGG
HEK2_050	chr20	97640	0.5	GAAttCAAAGCATAGAtTGCAGG
HEK2_051	chrX	58256414	0.5	GcAttCtAAtaATAGACTGCTGG
HEK2_052	chr3	10835885	0.5	aAACACAtAGCAcAGcCTGCAGG
HEK2_053	chr8	97317606	0.5	GAACACAgtaCATAGACTGgCAG
HEK2_054	chr7	1973215	0.5	GgctgCctAGCAcAGACTGCCGG
HEK2_055	chr5	145405225	0.5	tctCctAccGCATAGACTGCTGG
HEK2_056	chr14	106066714	0.4	GtcgcaAAcGCATAGACTtCCGG
HEK2_057	chr2	34403831	0.4	cttattAAAGCAcAGACTGCTGG
HEK2_058	chr7	83764326	0.4	actatataAGCATAGACTGtTGG
HEK2_059	chr14	106188442	0.4	GtcgcaAAcGCATAGACTtCCGG
HEK2_060	chr9	111564240	0.4	cgAtttcAAGCATAGACTGCTGG
HEK2_061	chr4	91393910	0.4	tAAtcataAAGCATAGAAaTGCTGG
HEK2_062	chr9	117268271	0.4	GAAtAC-AAGCATAGACTGCTGT
HEK2_063	chr1	208174511	0.4	tcACAaAtAaGCATAGACTGgCGG
HEK2_064	chr6	164073518	0.4	cttgctAAAGCAcAGACTGCTGG
HEK2_065	chr18	56307002	0.4	aAgaACAAAaCATAGACTGCAGG
HEK2_066	chrX	119188148	0.4	aAttAaAcAGaATAGACTGCTGG
HEK2_067	chr10	11472312	0.4	GAACACAgAaCcATAGACTGgGGG
HEK2_068	chrX	93938357	0.4	aAACaAAAGCATAGACcaCAAG
HEK2_069	chr3	33958845	0.4	GAACACA-gGCATAGACTatGGA
HEK2_070	chr9	90257919	0.4	ctcagagAAGCAcAGACTGCAGG
HEK2_071	chr2	144687377	0.4	atcatatAtGCATAaACTGCAGG
HEK2_072	chr1	191029339	0.4	caACaAAAtatATAGACTGCTGG
HEK2_073	chr3	124241469	0.3	attgttAAAGCAcAGACTGCAGG
HEK2_074	chr11	80540770	0.3	agAtcttAAGCATAGACTGtGGG
HEK2_075	chr4	146455378	0.3	tAACACAAtacCATAGACTGgCGG
HEK2_076	chr6	70806985	0.3	GAAtACAAAtaATAGACTatTGG
HEK2_077	chr3	10539384	0.3	GAAtttAAAGCATAGACTctGGG
HEK2_078	chr13	80619390	0.3	aAACaAtAaaATAGACTGCAGG

HEK2_079	chr5	126385455	0.3	ccACACcAAGCATAGACTtCTGG
HEK2_080	chr9	92683115	0.3	aAcaAaAcAGCATAGACTGCATG
HEK2_081	chr17	2810242	0.3	tAAttgcAAGCATAGACgGCAGG
HEK2_082	chr12	131431445	0.3	GgAgAgAgAGCATAGACTGCTGG
HEK2_083	chr11	105148271	0.3	agAaAtAAtatATAGACTGCAGG
HEK2_084	chr1	75043747	0.3	tggttactAGCATAGACTtCAGG
HEK2_085	chr20	31946581	0.3	aAACACTtAaCAcAGACTGCAGG
HEK2_086	chr8	58172690	0.3	ccACcagcAGCATAGACTcCAGG
HEK2_087	chr7	39877158	0.2	aAgacattttCATAGACTGtCGG
HEK2_088	chr3	65508026	0.2	tAAGcacAAGCATAGgCTGCTGG
HEK2_089	chr6	40925009	0.2	aAcaAgtAtGCATAGACTGCTGG
HEK2_090	chr15	81265977	0.2	aAAtACcAAcCAcAGACTGCAGG
HEK2_091	chr8	102768749	0.2	aAAtgCAtAGCATAGACTGCTGA
HEK2_092	chr1	237916998	0.2	tAACACAAAGacTAaACTGCAGG
HEK2_093	chr1	215323272	0.2	tAcCACgAAGCATAGACTGtAGG
HEK2_094	chr14	60249634	0.2	GAtCACTcAaCATAGACTGCATG
HEK2_095	chr3	29416510	0.2	GAACAg-AAaCATAGACcaaTGG
HEK2_096	chr11	79574188	0.2	tAcaAaAtAagATAGACTGCTGG
HEK2_097	chr10	96064388	0.2	acAgttAAAGaATAGACTGCAGG
HEK2_098	chr7	78443524	0.2	aAACACAAAgGCATAaACTcCAGG
HEK2_099	chr13	51093353	0.2	GAACACAcA-CAcAGACTGaGGG
HEK2_100	chr13	45266353	0.2	GgACAtAAAtCATAaACTGCTGG
HEK2_101	chr1	20011125	0.2	aAAgctAcAGCATAGACTGCAAG
HEK2_102	chr1	77190607	0.2	tcACACAAAcCATAGACTGaGGG
HEK2_103	chr11	68639975	0.2	GAtCtacctatATAGACTGCAGG
HEK2_104	chr9	31652144	0.2	tAtagaggAGCATAGACTGCTGG
HEK2_105	chr19	5068807	0.2	tAtCagAAAGCAcAGACTGCGGG
HEK2_106	chr12	54405073	0.2	GAACAacAAGgATAaACTGCCGG
HEK2_107	chr12	33389548	0.2	tAcaAacAcGCATAGaATGCCGG
HEK2_108	chr1	15519301	0.2	tAAGAaAtgGaATAGACTGCAGG
HEK2_109	chr20	21064111	0.2	aAAAtAtAAGCATAGACTaCGGG
HEK2_110	chr20	17801623	0.2	GAAtgCcAAGCATAaACTGCAGG
HEK2_111	chr2	19844956	0.2	aActcCAAAGCATAtACTGCTGG
HEK2_112	chr2	116312126	0.2	GtgaACAcAGCATAGACTGgGGC
HEK2_113	chr15	35048979	0.2	atcCtaAgAaCATAGACTGCAGG
HEK2_114	chr11	19349648	0.2	tttggtAAAaCATAGACTGCTGG
HEK2_115	chr4	170242082	0.2	tAACAAaAtGCATAGACTGCTAG
HEK2_116	chr1	66454108	0.2	tgtataAcAGCATAcACTGCTGG

HEK2_117	chr9	20221246	0.2	aAACACtAAtaATAGACTGtGGG
HEK2_118	chr12	2617156	0.2	aAgCAatcAGCAcAGACTGCGGG
HEK2_119	chr19	32225924	0.2	GgACACAgAgTATAGACTGaAGG
HEK2_120	chr6	123554091	0.1	cAcaAatcAaaATAGACTGCTGG
HEK2_121	chr13	58440671	0.1	GAAttCAAAGCATgGACTGCAGG
HEK2_122	chr19	18370525	0.1	ttcCAaAAAaCATAGACTGCTGG
HEK2_123	chr8	74980814	0.1	GtAgAgtAAGCATaACTGCTGG
HEK2_124	chrX	101465228	0.1	GgcCtttAgGCAcAGACTGCAGG
HEK2_125	chr21	16057037	0.1	aAgactttAGCATAGACTGCAAG
HEK2_126	chr5	76655640	0.1	GAAaAgtAAG-ATAGACTGCCGG
HEK2_127	chrX	37275566	0.1	aActcatcAGCATAGACTGtAGG
HEK2_128	chrX	39813705	0.1	aAatACAgAGCtTAGACTaCTGG
HEK2_129	chr3	99649762	0.1	aAACAtgAAGCATAGACaGCAAG
HEK2_130	chr3	86016290	0.1	acAtAgAtAttATAGACTGCTGG
HEK2_131	chr2	42080736	0.1	cAcatTAAGCAcAGACTGCTGG
HEK2_132	chr7	154994106	0.1	GAgCctAgAGCAcAGACTGCAGG
HEK2_133	chr10	85566693	0.1	aAgactTAAGCAaAGACTGCTGG
HEK2_134	chr1	176037520	0.1	ttACACAtAGCAcAGACTaCAGG
HEK2_135	chr7	56616286	0.1	agAagtTAAGCATAGACTGgTGG
HEK2_136	chrX	69481649	0.1	ctAtAttgAGaATAGACTGCAGG
HEK2_137	chr17	79546106	0.1	GtgCACAcAGCATAGACTGCATG
HEK2_138	chr9	74099753	0.1	ctAgcaAAAaCATAGACTGCTGG
HEK2_139	chr7	113235709	0.1	tAAaACAAAGaAcAGACTGCTGG
HEK2_140	chr5	165353994	0.1	tAcagatAgtaATAGACTGCAGG
HEK2_141	chr5	134255399	0.1	aAttAatAAcaATAGACTGCTGG
HEK2_142	chr6	39115051	0.1	aAgattgcAGCATAaACTGCTGG
HEK2_143	chr10	732884	0.1	aAgCACAtAaCATAGACTGaAGG
HEK2_144	chr14	88566211	0.1	GtAgAaAAAGtATAGACTGCAGG
HEK2_145	chr14	89446986	0.1	tActAtggtGCATAGACTGCTGG
HEK2_146	chr11	133057022	0.1	ctctACTgAGaATAGACTGCAGG
HEK2_147	chr13	47277564	0.1	tgAaAtAAtaCATAaACTGCCGG
HEK2_148	chr6	42655828	0.1	aAtaAacttGCATAGACTGtAGG
HEK2_149	chr10	72922827	0.1	agACcCtAAGCATAGACTGCAGA
HEK2_150	chr16	55456878	0.1	tAcCcaAAtGCAcAGACTGCTGG
HEK2_151	chr2	194416373	0.1	aAgatagAttATAGACTGCTGG
HEK2_152	chr4	177589824	0.1	aAACACAAAaCATAaAtTaCTGG

[표 13]

Digenome-captured sites (RNF2)

Site No.	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites
RNF2-01	chr1	210858376	7.3	tcaccaTTAGTCATTACCTGCTG
RNF2-02	chr1	185056773	4.4	GTCATCTTAGTCATTACCTGAGG
RNF2-03	chr17	53928598	4.3	GTCATCTTAGTCATTAC-TGAGG
RNF2-04	chr6	143212078	4.1	GTaATaTTAGTCATTACCgGTGG
RNF2-05	chr17	17072075	3.6	GTCATCcTAGTCATTtaCTGGGG
RNF2-06	chr12	131619804	2.2	aTCAcCTTAGcCATTACCgGGG
RNF2-07	chr9	102127746	1.4	GTCA-CTTAGTCATTgCCTGTGG
RNF2-08	chr15	92315580	1.0	GTCAcaTTAgcCATTACCTGTGA
RNF2-09	chr2	57559709	1.0	tgtATCTTAGaCATTACaTGTGG
RNF2-10	chr5	29673802	0.8	aTttTCTTAGTgATTaCTGGGG
RNF2-11	chr15	25135467	0.8	catcTaaTAGTaATTACCTGGGG
RNF2-12	chr3	60339996	0.6	aTagTaTTAGTCATTACCTGTGA
RNF2-13	chr12	12595622	0.4	acCATCTTAGTCATTaCTaATG
RNF2-14	chrX	131090978	0.4	aTttTCTTAGTCATTACCTaGAG
RNF2-15	chr2	51894204	0.4	aTCATCTTcaTCATTACaTGAGG
RNF2-16	chr2	134638146	0.4	cctATaTataTCATTACCTtTGG
RNF2-17	chr16	62531206	0.4	GTCATtTTAGTCATTaCTtGAG
RNF2-18	chr22	44137982	0.4	aTCAcCTgAGTCATTACCcaTGG
RNF2-19	chr4	125521739	0.4	tTaATCTTAGTCATTAcTtTGG
RNF2-20	chr2	102668931	0.3	aTCATCaTcGTCATTaCTGGGG
RNF2-21	chr14	88336132	0.3	aTCA--TTAGTCATTgCCTGAGG
RNF2-22	chr7	43967266	0.3	cTtAT-TTAGTCATTACCTGTAG
RNF2-23	chr2	41520178	0.3	GTCATaTTAaTCATTACaTaGAG
RNF2-24	chr1	111785389	0.3	tcCATCTcAcTCATTACCTGAGG
RNF2-25	chr8	51846780	0.3	GTtATCTTAGTcTTACCTGAGA
RNF2-26	chr2	24749046	0.3	caCATCTTAccCATTACaTGAGG
RNF2-27	chr6	5869239	0.3	GcCAcCTcAGTCATTAgCTGGGG
RNF2-28	chrX	133138307	0.3	acatTagTAaaCATTACCTaGGG
RNF2-29	chr1	59227209	0.3	taCcTacctGTCATTACCTaTGG
RNF2-30	chr2	177556597	0.3	GatATCTTAGcCATTACCTaGGA
RNF2-31	chr3	113984053	0.3	tatcTacTAcTCATTACCTGAGG
RNF2-32	chr2	66353253	0.2	agCATCTctGTCATTACCcaGGG
RNF2-33	chr8	99451745	0.2	cgtgcaTTAGTCATTACCTGAGG
RNF2-34	chr5	109304296	0.2	GTCtgtaTAGTCATTACCTtTGG
RNF2-35	chr6	155177402	0.2	taCATCTatGTtATTACCTaTGG
RNF2-36	chr17	58390142	0.2	tTCATCTTtGTCATTACCTaAAG

RNF2-37	chr11	8498488	0.2	tatActaTAtaCATTACCTGGGG
RNF2-38	chr6	3726562	0.2	GctcatTcAcTCATTACCTaTGG
RNF2-39	chr16	5977420	0.2	GTgATCTaaAGTCATTACCTtAGG
RNF2-40	chrX	76946717	0.2	tTCAataTAaTCATTACCTGTGG
RNF2-41	chr18	62061226	0.2	aTtATtTTAGTCATTACCTtTGG
RNF2-42	chr19	14867049	0.2	ccCATCTcAGcCATTACCTGGGT
RNF2-43	chr13	70399620	0.1	agaATaTTAGTCcTTACCTGGGG
RNF2-44	chr12	86219647	0.1	GatcctTTAaTCATTACCTtTGG
RNF2-45	chr13	103624473	0.1	acCATCTTAGTCaCATTACCTGGGC
RNF2-46	chr1	174478567	0.1	GTatTaTTgtaCATTACCTGAGG
RNF2-47	chr2	209871803	0.1	agCATCTTAGcCATTACCTcTAG
RNF2-48	chr5	91144278	0.1	agCATCTTAaTCcTTACCTcAGG
RNF2-49	chr4	160106364	0.1	GaatcCTcAtaCATTACCTGTGG
RNF2-50	chr4	37552932	0.1	aTCATCTTttTaATTACCTaTGG

[표 14]

Digenome-captured sites (TYRO3)

Site No.	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites
TYRO3-01	chr15	76553332	4.3	GGCCACACTAGtGTTGcCGCTGG
TYRO3-02	chr15	41857272	3.4	GGCCACACTAGCGTTGCTGCTGG
TYRO3-03	chr20	45312757	2.0	GGCCACACcAGCcTTGCTGtCGG
TYRO3-04	chr5	173133468	0.8	GGCCAC--TAGCGTTGCTcCAGG
TYRO3-05	chr22	50686658	0.6	GGCCACACTgaCcTTGCTGCTGG
TYRO3-06	chr9	140109027	0.4	ttCCACAC-AGCtTTGCTGCTGG
TYRO3-07	chr17	17207899	0.3	GGCCAC-CTAGgTTGCTGCTGG
TYRO3-08	chr11	99666699	0.2	ccaCcCtCTAGCaTTGCTGCTGG
TYRO3-09	chr11	27495183	0.2	cctaACACcAaCGTTGCTGCTGG
TYRO3-10	chr1	38684664	0.2	ctgagCtCTAGCaTTGCTGCTGG
TYRO3-11	chr3	22936710	0.1	GGCatataTAGCaTTGCTGCTGG
TYRO3-12	chr6	150629124	0.1	GaaCACACTAaCaTTGCTGtGGG

[표 15]

Digenome-captured sites (WEE1)

Site No.	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites
WEE1-01	chr1	234241713	11.1	tTaGAGATGTTCTATTA tTCCGG
WEE1-02	chr11	9610086	2.0	GTCGAGATGTTCTATTACTCTGG
WEE1-03	chr4	18855558	1.9	GagGAGATGcTCTATTACTCCGG

WEE1-04	chrX	42756785	0.7	aTCaAGATGaaCTATTACTCTGG
WEE1-05	chr1	215805743	0.4	GTgGAGATGTTaTgTTACTCTGG
WEE1-06	chr5	57863350	0.4	GaCataATGTTCTATTACTCAAG
WEE1-07	chr4	119010382	0.3	t aatAGATGTTCTATTACTaAGG
WEE1-08	chr15	80845387	0.3	aTtaAGATGTTCTATTaTaGGG
WEE1-09	chr13	87867165	0.2	cagctaATGTaaTATTACTCAGG
WEE1-10	chr4	107037600	0.2	aTCaAGATaTTCTATTACTgGGG
WEE1-11	chrX	136297269	0.2	t c t t A t A t T T C T A T T A C T C A G G
WEE1-12	chr1	8500986	0.2	aTaGAGcaGTTaTATTACTCTGG
WEE1-13	chr11	71806684	0.2	GTaGAGATGTTCTcTaACTCAGG
WEE1-14	chr12	83599167	0.2	cTCt t t t a t T T C T A T T A C T C T G G
WEE1-15	chr8	71394697	0.1	aataaATaTTCTATTACTCAGG
WEE1-16	chr8	86942319	0.1	caactataaTTCTATTACTCTGG
WEE1-17	chr1	61165563	0.1	GgCtActgtcTCTATTACTCAGG
WEE1-18	chr20	45730192	0.1	aTttAGATaTTCTATTACTCTGA
WEE1-19	chr8	134795474	0.1	cTttAagataTCTATTACTCTGG
WEE1-20	chr2	128238175	0.1	caaGAGAcGcTCTATcACTCAGG
WEE1-21	chr7	31677975	0.1	aaacAGcTGTTCTATTACTCAGG

[표 16]

Digenome-captured sites (EphB4)

Site No.	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites
EphB4-01	chr7	100411300	10.3	GCAGAATATTCCGGACAAACACGG
EphB4-02	chr5	81021421	8.0	GCAGAATATTaCaGACAAACtAGG
EphB4-03	chr15	47712791	4.1	GCAGAATATcaGGACAAACAATG
EphB4-04	chr11	123329727	2.4	aCAGAATATTCaGACAAtCACAG
EphB4-05	chr1	93047514	2.2	GCAGtA-AaTaaGACAAACAAGG
EphB4-06	chr15	51198151	1.9	aCAaAATATTCCGGACaggCACGG
EphB4-07	chr18	29712895	1.8	GCAGAATtTaaGGACAAgCAAGG
EphB4-08	chr3	55670295	1.8	GCAGAATA--aGGACAAACATGG
EphB4-09	chr13	25919264	1.4	cagtAATATTCaGACtAAACATGG
EphB4-10	chr3	122057177	0.9	t t A a t A T A T a a G G A C A A A C A T G G
EphB4-11	chr1	26891812	0.8	aCAtAATATTCaaACAAACAGGG
EphB4-12	chr10	131074805	0.6	aatGAATaagGGACAAACAGGG
EphB4-13	chr4	123425007	0.5	tCAG-ATATaaGGACAAACATGG
EphB4-14	chr5	60692862	0.4	GCtGAATATTgGGACAAtCATGG
EphB4-15	chr9	930844	0.4	GCAGt t A T A a a a a G A C A A A C A G G G
EphB4-16	chr9	29344386	0.4	aCAGAATATaaGaACAAAtATGG

EphB4-17	chr 4	160752620	0.4	atAGAATAaTCaGACAAAaAAGG
EphB4-18	chr 16	23706770	0.4	GCAG-ATAaagGGACAAACATGG
EphB4-19	chr 11	117354211	0.3	GacagtGAAaTaGGACAAACATGG
EphB4-20	chr 14	90497703	0.3	tGAGAATATgCGGACAAACACAG
EphB4-21	chr 2	68459706	0.3	atAGAATATTaGGACAAAtAGAG
EphB4-22	chr 13	51951355	0.2	caAtAATAaaaGGACAAACATGG
EphB4-23	chr 17	7777056	0.2	aCAGgATAaaaGGACAAAtATGG
EphB4-24	chr 3	188431448	0.2	aCAGAtATAcgtAGACAAACAGAG
EphB4-25	chr 12	11705824	0.2	atAGAATAagaaGACAAACATGG
EphB4-26	chr 10	95697506	0.2	ttAGAATATTcCtACAAACAAGG
EphB4-27	chr X	123665071	0.2	cCAtgtTATTCaGACAAACATGG
EphB4-28	chr 16	75869847	0.2	GCAGAA-AaTaGGACAAAtCACGG
EphB4-29	chr 6	128784409	0.2	agAGAATATa-GGACAAACtTGG
EphB4-30	chr 5	144297821	0.2	GCAGAAataTttGACAAACAAGG
EphB4-31	chr 14	29760524	0.1	aaAGAATATa-GaACAAACAGGG
EphB4-32	chr 1	36914336	0.1	aCAGAA-ATaCGaACAAACATGG
EphB4-33	chr 21	26227853	0.1	aacaAATgcaaGGACAAACATGG
EphB4-34	chr 13	63528393	0.1	aCActAT-TTaaGACAAACAGGG
EphB4-35	chr 10	72516747	0.1	aCAGAActacgaGACAAACACGG
EphB4-36	chr 7	20108292	0.1	GCAGA-TaccCaGACAAACAGGG
EphB4-37	chr 1	89896856	0.1	taAGAATAT-aGGACAAACAGGA
EphB4-38	chr 4	145533104	0.1	cCAGAATATTa-GACAAACATGG
EphB4-39	chr 20	9784096	0.1	agAGAATATagGGACAAAtAAGG

[표 17]

Digenome-captured sites (HPRT-Exon6)

Site No.	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites
HPRT_E6-001	chrX	133627607	29.0	GTATAATCCAAAGATGGTCAAGG
HPRT_E6-002	chr 12	7468915	21.3	GTATAcatCcAAGATGGcCAGGG
HPRT_E6-003	chr 3	10057621	17.0	GTATtA-CCAAAGATGGTCtGGG
HPRT_E6-004	chr 22	39586062	12.7	GTATAATC-AAAGATGGcCcTGG
HPRT_E6-005	chr 7	13928792	10.5	GTATtAT-CAAAaATGGTCAAGG
HPRT_E6-006	chr 1	225831201	10.2	aTATAA-CCAAAGATGtTCACAG
HPRT_E6-007	chr 17	53575925	9.3	aTATAATCtTAAGATGGTCAAGG
HPRT_E6-008	chr 12	13406973	9.0	aTATAATCCtAAGATGtTCATGG
HPRT_E6-009	chr 14	27255077	6.1	GTATtATCCAAAGATGacCAGAG
HPRT_E6-010	chr 7	85914581	5.6	aTAcAATCCAAAGtTGGTCtGGG

HPRT_E6-011	chr3	9906633	4.7	aTATAttaCCAAAGATGGTCtGGG
HPRT_E6-012	chr8	125496169	4.5	GTAgaATCCAAtAGATGGaCAGGG
HPRT_E6-013	chr17	10064482	4.3	GTATtA-CCAAAGATGGTCATGG
HPRT_E6-014	chr2	53374352	4.3	GTAcAATCaAAAGATGaaCAAGG
HPRT_E6-015	chr4	102026801	4.1	aTATtATCCAAAGATGGagtTGG
HPRT_E6-016	chr6	19070067	3.8	aTgTAatTCCAAAGATGGTCAGGG
HPRT_E6-017	chr6	7761387	3.7	aagTAATCCAAAGATaGTCtTGG
HPRT_E6-018	chr12	59613203	3.6	GTATAAcCCAAAGaagTaATGG
HPRT_E6-019	chr16	20336811	3.6	aTATAtATtCcAAGATGGTtcTGG
HPRT_E6-020	chr4	136809212	3.5	tTATAATCCAAAGaAGGcCAAGG
HPRT_E6-021	chr12	117915656	3.5	GTATctAcCCAAAGATgtTCATGG
HPRT_E6-022	chr14	45005861	3.4	GTATAAT--AAAGATGGTCAAGG
HPRT_E6-023	chr14	104033102	3.3	aTAgaTtTCCAAAGATGGaCATGG
HPRT_E6-024	chr11	93731924	3.1	cTATAATCtAAAaATGGTCAAGG
HPRT_E6-025	chr13	75257165	2.9	GTATAtTagAAAGATGGTCATGA
HPRT_E6-026	chr15	54911620	2.8	GgATAATCaAAAGATGGaCtTGG
HPRT_E6-027	chrX	6849511	2.6	aTATAATgCAAAGcTGGTCACGG
HPRT_E6-028	chr11	87526371	2.5	aTAcAATCCAAAGATatTCtAGG
HPRT_E6-029	chr3	19652583	2.4	GTATAATCCAAAGgTGGcCtAGG
HPRT_E6-030	chr17	45009128	2.4	aTATAATCCAAtAGaAGGTCAAGA
HPRT_E6-031	chr8	9090663	2.3	GatataAatacAAGATGGaCAAGG
HPRT_E6-032	chr12	6810501	2.2	GgATAATCCAAAGATGGTgcAAG
HPRT_E6-033	chr6	118233456	1.5	aTATAAaCCAAAGATcGtTAAGG
HPRT_E6-034	chr14	85667729	1.4	tctcccTCCAAAGaAGGTCACGG
HPRT_E6-035	chrX	110551547	1.4	tTATAgTCCAAAGgaaGTCAGGG
HPRT_E6-036	chr8	18630339	1.2	tatTtATCCAAAGATGGTCtAGA
HPRT_E6-037	chr12	45338334	1.1	aTATAtaCCAAAGgTaGTCATGG
HPRT_E6-038	chr7	135056752	1.1	aggagATCCAAAGATGGTCAAGG
HPRT_E6-039	chr9	82090375	1.1	acATAgTCCAAAtATGGaCAGGG
HPRT_E6-040	chr12	103187212	1.0	GTATAtAattcAAGATGGTCAGGA
HPRT_E6-041	chr12	58996606	1.0	GaATAAgCCAAAGATGGTCAGTG
HPRT_E6-042	chr12	49657468	1.0	GTATctcCCAgAGATGGTCAAGG
HPRT_E6-043	chr7	120024149	0.9	aTgTAATCtAAAGATGaTtATGG
HPRT_E6-044	chr15	101375955	0.9	GTATAA-CCAAAtGATGGTCAGGA
HPRT_E6-045	chr10	13873222	0.9	aTgTcATCCAAAGaAGGTCAGAG
HPRT_E6-046	chr21	41428166	0.9	agATAtTCCAAAGATGGTgAGGG
HPRT_E6-047	chr1	7314483	0.8	tTtccAcCCAAAGATGGTCcAGG
HPRT_E6-048	chr8	133182863	0.8	tTATAgctCcAAGATGGTCAAGG

HPRT_E6-049	chr6	1253997	0.8	t cAgAATCCtAAGATGGTCAGGA
HPRT_E6-050	chr12	15783424	0.8	GgATAATCCAAAGtTGGTCATAG
HPRT_E6-051	chr9	15160111	0.7	GTATAATCC-tAGATGGTgAGGG
HPRT_E6-052	chr12	29153372	0.7	GTATt tTCCAAcGATGGTCATGG
HPRT_E6-053	chr3	1123204	0.7	GaATAATtCAAAGATaaTCAAGG
HPRT_E6-054	chr2	50999543	0.7	aaATAATa-AAAGATGGTCATGG
HPRT_E6-055	chr2	36768513	0.7	cTcatATCCAAAGATGGTaAAGG
HPRT_E6-056	chr6	4210565	0.7	tTATAA-CCAAAGATGGTaATGG
HPRT_E6-057	chr14	48260581	0.7	GccTAATCCAAAGATGagCAGGG
HPRT_E6-058	chr8	117575569	0.7	cTATAATgCCAAAGATGGTCATGG
HPRT_E6-059	chr5	160988739	0.7	GTATAgatCAAAGATGccCACGG
HPRT_E6-060	chr14	94364216	0.6	tctcctTCCAAAGATGGTCtGGG
HPRT_E6-061	chr9	125328946	0.6	aTAatcTCCAAAGATGtTCAGGG
HPRT_E6-062	chr9	134083092	0.6	aTATt tTCCAAAGATGaTCACAG
HPRT_E6-063	chr16	6159271	0.6	t atTgATCCAAAGATGaTCAAGA
HPRT_E6-064	chr22	27822987	0.6	aTATAtTtCcAAGATGGgCAGGG
HPRT_E6-065	chr19	48640167	0.6	tTcTcATCCAAgGATGGTCAGGG
HPRT_E6-066	chr11	107913428	0.6	aTATAaCaAAAGATGGTt tAGG
HPRT_E6-067	chr22	25519897	0.6	aaATAgTCCAAgGATGGcCAGGG
HPRT_E6-068	chr2	222954701	0.5	GTATt tTCtAAAGATGGTCATGA
HPRT_E6-069	chr1	21736973	0.5	GgATAAatCcAAGtTGGTCATGG
HPRT_E6-070	chr20	41218527	0.5	aTATt tTCCAAAGATGGTCATGA
HPRT_E6-071	chr16	55855323	0.5	tgActcTCCAAAGATGGTCACAG
HPRT_E6-072	chr3	39725483	0.5	caATAtTCCAAAGATGGTtATGG
HPRT_E6-073	chrX	141390834	0.5	GcAgcATCCAAAGATGGgCAGGG
HPRT_E6-074	chr4	23070418	0.5	cTATAATtCAAAGATGtTCtTGG
HPRT_E6-075	chr8	9653862	0.5	cacatATatAAAGATGGTCATGG
HPRT_E6-076	chr22	32085052	0.5	GTATctTCCAAAGATGGcCAATG
HPRT_E6-077	chr16	8011143	0.5	aatatt aCCtAAGATGGTCATGG
HPRT_E6-078	chr13	21574593	0.5	GTATAaCCAAAaATatTCAGGG
HPRT_E6-079	chr12	58995581	0.4	GaATAAgCCAAAGATGGTCAGTG
HPRT_E6-080	chr10	107263858	0.4	aTATAATgCAAGaGaaGTCAAAG
HPRT_E6-081	chr11	44024764	0.4	GTATAAgCCAAAGcaGGTCACGG
HPRT_E6-082	chr6	138575959	0.4	aTATAaCCAAAGAcGGcCcTGG
HPRT_E6-083	chr3	98960373	0.4	agATAATCCAAGGATGGTagAGG
HPRT_E6-084	chr8	43181791	0.4	aTATAATCCAgAaATGtTCACGG
HPRT_E6-085	chr5	118103678	0.4	actcttTtCcAAGATGGTCAGGG
HPRT_E6-086	chr7	129222414	0.4	cTgatgTCCAAAGATGGTCtCGG

HPRT_E6-087	chr3	160736673	0.4	GTgaAtTCCAAAGaAGGTCATGG
HPRT_E6-088	chrX	125678217	0.3	catTAATCCAAAGATGaTaAAGG
HPRT_E6-089	chr6	152915585	0.3	GctctATCCAAAGATGGaCAGGG
HPRT_E6-090	chr9	128839841	0.3	cTgTAATCCAgAaATaGTCAGGG
HPRT_E6-091	chr14	96980434	0.3	GTATAtgTCCAAAGATGaTCATGC
HPRT_E6-092	chr13	80168480	0.3	cTATAATCCAAgGAaGGcCAAGG
HPRT_E6-093	chr8	74179696	0.3	GTATccTCCAAAGATGagtAAGG
HPRT_E6-094	chr4	112559644	0.3	aaATAATaCAAAGATGGTtgGGG
HPRT_E6-095	chr13	74505576	0.3	caATAATCCAAAGaGaTtAAGG
HPRT_E6-096	chr19	32267421	0.3	aTATAATCaCAAAGATGGcaATGG
HPRT_E6-097	chr14	95877075	0.3	GTtTAAGCCAAAGtTGGTCgGGG
HPRT_E6-098	chr8	36263307	0.3	GTAgcATCCAAAGATGGTggAGG
HPRT_E6-099	chr5	28287911	0.3	GTAgaATatcAAGATGGTCAAGG
HPRT_E6-100	chr5	98310114	0.3	agAatgcCCtAAGATGGTCtGGG
HPRT_E6-101	chr6	76199451	0.3	aTATAAcCagAAGATGtTCATGG
HPRT_E6-102	chr10	83966021	0.3	GTATtATCtAAAGgaGGTCATGG
HPRT_E6-103	chr1	55051337	0.2	agATAtaCagAAGATGGTCgGGG
HPRT_E6-104	chr8	91475188	0.2	aaATAATCCAAAaATaGTCATGG
HPRT_E6-105	chr9	20225753	0.2	cTtTAtTCCAAAGATaGTCATGG
HPRT_E6-106	chr6	81303564	0.2	GcATAttTCCAAAcATGGTCAAGG
HPRT_E6-107	chr1	241707903	0.2	GTATttTtCAAAGATGGTCATGA
HPRT_E6-108	chr9	28194167	0.2	GTATttTCCAAAGATGGcCACGA
HPRT_E6-109	chr3	17000484	0.2	tataAATCCAAgGATGGTCtGAG
HPRT_E6-110	chr16	73853834	0.2	tTATttTCCAAAGATGGTCACGA
HPRT_E6-111	chr15	29999660	0.2	tgcacATCCAcAGATGGTCAGGG
HPRT_E6-112	chr5	157715478	0.2	GTgTAATCCAgAGATGaTCACAG
HPRT_E6-113	chr10	2711644	0.2	tctTAATCCAtAGaAGGTCAAGG
HPRT_E6-114	chr17	61689903	0.2	tTcaAAT--AAAGATGGTCACGG
HPRT_E6-115	chrX	125134402	0.2	tcATAATCCAAAaGATGGaCAGGG
HPRT_E6-116	chr1	94251248	0.2	aTATAATaCAAAataAGGTCAGGG
HPRT_E6-117	chr22	35100488	0.2	cTAgAtaCCAAAGATGGTCAGAG
HPRT_E6-118	chr2	38407889	0.2	aTtTAgtTCCAAAGaGGTCAGAG
HPRT_E6-119	chr7	112015061	0.2	aTATAcAaCCAAAGATGtTcTAgG
HPRT_E6-120	chr20	2163127	0.2	GTATAtAcataAAGaAGGTCATGG
HPRT_E6-121	chr8	100230333	0.2	aTtatATCCAAAtATGGTCAGAG
HPRT_E6-122	chr2	206041685	0.2	tTATAATCCAtAAaATGGTCATAG
HPRT_E6-123	chr1	153531237	0.2	tTAgtcTCCAAAGATGGTCACTG
HPRT_E6-124	chrX	118562021	0.2	tatcttTCCAAAGATGGTCtATG

HPRT_E6-125	chr10	26718853	0.2	GTATttTCCAAAGcTaGTCATGG
HPRT_E6-126	chr17	61180425	0.2	GcATAATCCAAgGAaGGTCcTGG
HPRT_E6-127	chr2	125810813	0.2	aTATttTCCAAAaATGGTCACAG
HPRT_E6-128	chr7	28299854	0.2	aTATAATCCAAAagaaGgaAAAG
HPRT_E6-129	chr11	23995596	0.1	aTATAtcCaAAAGATGGTCtGGA
HPRT_E6-130	chr1	34307709	0.1	GatattaaCAAAGATGGTCtGGG
HPRT_E6-131	chr5	173751304	0.1	GaAgAATCCAAcGATGGcCACGG
HPRT_E6-132	chrX	14442705	0.1	acATAATCaCAAAtATGGTCACGG
HPRT_E6-133	chr20	15867186	0.1	GTATttTCCAAAGATGGcCACAG
HPRT_E6-134	chr5	157484648	0.1	tctaAtgaatAAGATGGTCgAGG
HPRT_E6-135	chr1	203269364	0.1	agATAATgCAAAGAcGGTtAGGG
HPRT_E6-136	chr12	28980303	0.1	GTATAtaCCAAAaATGGTgAAGG
HPRT_E6-137	chr18	52818481	0.1	aTATAc-CCAAAGATGGTCaAAG
HPRT_E6-138	chr10	131217430	0.1	tctgAcTCCcAAGATGGTCtTGG
HPRT_E6-139	chr3	109018997	0.1	GgAgAtaCCAAAGATGGTCAGGA
HPRT_E6-140	chr21	20933404	0.1	GTATAtTtCAAAtATGGTCATAG
HPRT_E6-141	chr2	220647283	0.1	aTATAA-CCAAAGATtGTCAGAG
HPRT_E6-142	chr22	23199165	0.1	cTggAgTCCAgAGATGGTCATGG
HPRT_E6-143	chr6	17237684	0.1	GTAgaAAcCAgAGATGGaCAAGG
HPRT_E6-144	chr10	82243028	0.1	tTATAATaCAAAGgTGGTaATGG
HPRT_E6-145	chr15	43004167	0.1	cTATAATCaAAAGtaGGTCATAG
HPRT_E6-146	chr7	25485727	0.1	aTATttTCCAAAaATGGTCATAG
HPRT_E6-147	chr2	71429668	0.1	GTAAaATCCAAAGAgGacCAGGG
HPRT_E6-148	chr3	141311087	0.1	cctaAATCtAAAGATGGTCgGGG

[표 18]

Digenome-captured sites (HPRT-Exon8)

Site No.	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites
HPRT_E8-01	chrX	133632686	2.6	GAAGTATTCATTATAGTCAAGGG
HPRT_E8-02	chr11	93382674	1.3	GAAGcATTTCATTATAGTCAAAGG
HPRT_E8-03	chr21	34111536	0.6	GAAGaATTTCATTATAGaCAATGG
HPRT_E8-04	chr11	100880874	0.6	aAgcTATTCATTATAGcaAATGG
HPRT_E8-05	chr22	29285654	0.6	GAAGTtaATTTCATTATAGaCAATGG
HPRT_E8-06	chr4	15866630	0.5	aAAGTATTaATTATAGTCAAGGA
HPRT_E8-07	chr2	53543615	0.5	GAAGTAT-CcTTATAGTCaAAG
HPRT_E8-08	chr3	149876617	0.4	GAAGTAT-CATTATAGTCttGGG
HPRT_E8-09	chr8	71429233	0.4	GcAGTATTCATaATAGTCAAAAG
HPRT_E8-10	chrX	146032194	0.4	aAAGTATTtcaTATAGcCAAAGG

HPRT_E8-11	chr5	132311070	0.4	GAAtTATTCATcATAGcCAAAGG
HPRT_E8-12	chr3	36349740	0.3	tAAGTAaTCATTATAGTCaGATG
HPRT_E8-13	chr3	181655989	0.3	atAGTATTCATTATAGTaAcAGG
HPRT_E8-14	chr11	116029576	0.3	aAAaTATcCATTATAGTCttACG
HPRT_E8-15	chr10	131905650	0.3	GAAcTATTCATTATtTCAACGG
HPRT_E8-16	chr9	89312308	0.2	GttGTATTCATTATAGTCAAGAG
HPRT_E8-17	chr4	178543184	0.2	atAGTATTCATTATAGTaAcAGG
HPRT_E8-18	chr12	72733535	0.2	agAGTATcCAccATAGTCAAGGG
HPRT_E8-19	chr1	45916291	0.1	GAAGTAcatcTTATAaTCAATGG
HPRT_E8-20	chr4	153599265	0.1	aAtGTATTCATTATAGgCAAGAG

도 11 및 표 11 내지 18에서와 같이, Digenome 1.0 (DNA Cleavage Score cutoff: 2.5)으로 분석시, ABE7.10 + Endo V와 복합체를 이루는 총 7 개의 sgRNA는 표적 부위를 포함하여 1 내지 28 개 (sgRNA 당 8 ± 4 개) 부위에서 인간 유전체 DNA를 절단하였다. 이러한 결과는 Cas9 뉴클레아제와 BE3 Δ UGI 디아미나아제 (+ USER)가 각각 인간 유전체에서 70 ± 30 개 부위 (Kim, D., Kim, S., Kim, S., Park, J. & Kim, J.S. Genome-wide target specificities of CRISPR-Cas9 nucleases revealed by multiplex Digenome-seq. *Genome research* 26, 406-415 (2016)) 및 8 ± 3 부위에서 유전체 DNA를 절단한 결과와 비교될 수 있다.

본 실시예에서 분석된 7 개의 sgRNA 중 두 개 (HEK2 및 RNF2)의 sgRNA를 각각 BE3 Δ UGI 또는 Cas9와 함께 사용하여, Digenome 분석을 수행하고, 그 결과를 ABE7.10와 함께 사용하여 얻어진 결과와 비교하여, 도 12a 및 도 12b에 나타내었다. HEK2 sgRNA를 ABE7.10, BE3 Δ UGI, 및 Cas9와 각각 결합하여 사용한 경우, 각각 17개, 2개, 및 24개의 in vitro 절단 부위를 생성하였다 (도 12c 참조). RNF2 sgRNA는 ABE7.10, BE3 Δ UGI, 및 Cas9와 각각 결합하여 사용된 경우, 각각 5개, 1개, 및 13개의 절단 부위를 생성하였다 (도 12d 참조). 2 개의 BE3 디아미나아제를 사용하여 확인된 3 개의 in vitro 절단 부위는 모두 각각의 ABE7.10 디아미나아제에 의해서도 확인되었다. 대조적으로, ABE7.10와 Cas9를 사용하여 얻은 in vitro 절단 부위는 크게 상이하였다. HEK2 ABE7.10으로 얻은 17개 중 10개 부위 (59%)는 Cas9를 사용한 경우에는 동정되지 않았다. RNF2 ABE7.10으로 얻은 5 개 중 2개 부위 (40%)는 Cas9를 사용한 경우에는 동정되지 않았다. 이러한 결과는 ABE가 일반적으로 Cas9보다 더 특이적이면서도

Cas9와 상이한 off-target site 세트를 인식할 수 있음을 시사한다.

상기한 7개 sgRNA를 사용하여 targeted amplicon sequencing를 통하여 확인된 57개 *in vitro* cleavage sites에서의 ABE off-target base editing을 검증하였다. 또한, Digenome 2.0 프로그램을 사용하여, PAM과 유사한 서열 (5'-NGN-3' 또는 5'-NNG-3')을 포함하고 cutoff score가 >0.1이며, 각각의 on-target sequences와 비교하여 10개 이하의 mismatches를 포함하는 후보 off-target sites를 추가로 확인하였다.

ABE7.10과 복합체를 형성한 7 개의 sgRNA를 이용하여 12 내지 152개 사이트 (60 ± 20 , on average)를 얻었으며, 이는 Digenome 1.0으로 얻은 7개 on-target 부위와 57개 *in vitro* 절단 부위 (표 12 내지 18)를 모두 포함하였다. deep sequencing을 통하여 얻어진 절단 점수가 1 이상이거나 서열 상동성이 높은 부위를 포함하여 총 193개 부위에서 7 개의 ABE7.10 디아미나아제의 염기 교정 또는 유전자 교정 빈도를 측정하여, 그 결과를 표 19 내지 표 25 및 도 12e에 나타내었다.

[표 19]

Mutation frequencies of ABE7.10, BE3, and Cas9 at on-target and off-target sites captured by Digenome-seq (HEK2; 'DNA sequence at a cleavage sites'은 표 12 참조)

Site No.	DNA Cleavage Score	Base editing frequency (%)				Indel frequency (%)	
		Untreated	(+) ABE7.10	Untreated	(+) BE3	Untreated	(+) Cas9
HEK2_001	20.1	0.12	0.11	0.06	0.04	0.01	0.01
HEK2_002	15.1	0.00	0.00	0.04	0.04	0.02	0.01
HEK2_003	14.7	0.06	0.09	0.03	0.04	0.00	0.01
HEK2_004	14.3	0.10	0.31	0.07	0.18	0.02	3.09
HEK2_005	13.4	0.15	0.13	0.00	0.00	0.00	0.01
HEK2_006	13.4	0.00	0.00	0.00	0.00	0.06	0.11
HEK2_007	8.9	0.06	63.19	0.06	74.03	0.00	87.57
HEK2_008	6.6	0.06	0.12	0.03	0.03	0.02	0.01
HEK2_009	5.8	0.10	0.09	0.00	0.00	0.00	0.00
HEK2_010	5.1	0.00	0.00	0.03	0.03	0.00	0.01
HEK2_011	4.9	0.09	0.08	0.00	0.00	0.00	0.00
HEK2_012	4.6	0.05	0.06	0.00	0.00	0.02	0.00
HEK2_013	3.9	0.00	0.00	0.00	0.00	0.01	0.00

HEK2_014	3.5	0.08	0.08	0.18	0.11	0.00	0.00
HEK2_015	3.0	0.11	0.13	0.00	0.00	0.00	0.00
HEK2_016	2.9	No PCR					
HEK2_017	2.7	0.13	0.07	0.00	0.00	0.01	0.00
HEK2_018	2.5	0.18	0.20	0.08	0.06	0.00	0.00
HEK2_019	2.2	0.12	0.12	0.00	0.00	0.00	0.00
HEK2_020	2.1	0.12	0.20	0.02	0.03	0.09	0.08
HEK2_021	2.1	0.17	0.19	0.23	0.23	0.01	0.08
HEK2_022	2.1	0.13	0.11	0.08	0.08	0.00	0.01
HEK2_023	2.0	0.09	0.09	0.00	0.00	0.09	0.06
HEK2_024	1.7	0.00	0.00	0.08	0.06	0.00	0.00
HEK2_025	1.4	0.13	0.14	0.03	0.05	0.01	0.01
HEK2_026	1.3	0.13	0.16	0.08	0.08	0.00	0.00
HEK2_027	1.3	0.15	0.16	0.06	0.03	0.00	0.00
HEK2_028	1.3	0.00	0.00	0.03	0.02	0.01	0.02
HEK2_029	1.2	0.23	0.15	0.00	0.00	0.00	0.00
HEK2_030	1.2	0.09	0.11	0.00	0.00	0.01	0.02
HEK2_031	1.2	0.09	0.12	0.02	0.03	0.01	0.00
HEK2_032	1.1	0.05	0.09	0.04	0.06	0.00	0.00
HEK2_033	1.1	0.11	0.22	0.07	0.04	0.00	0.00
HEK2_034	1.0	0.05	0.05	0.07	0.13	2.42	4.10
HEK2_042	0.7	0.03	0.04	0.07	0.11	0.03	0.07
HEK2_044	0.7	0.00	0.00	0.08	0.03	0.00	0.00
HEK2_050	0.5	0.22	0.19	0.00	0.00	0.22	0.23
HEK2_052	0.5	0.13	0.11	0.03	0.02	0.13	0.11
HEK2_065	0.4	0.11	0.15	0.15	0.08	0.11	0.07
HEK2_079	0.3	0.00	0.00	0.00	0.00	0.00	0.00
HEK2_082	0.3	0.00	0.00	0.34	0.24	0.00	0.00
HEK2_092	0.2	0.11	0.08	0.12	0.11	0.11	0.09
HEK2_093	0.2	0.02	0.03	0.14	0.13	0.02	0.05
HEK2_099	0.2	0.06	0.08	0.00	0.00	0.13	0.08
HEK2_101	0.2	0.00	0.00	0.07	0.06	0.00	0.00
HEK2_105	0.2	0.00	0.00	0.08	0.07	0.00	0.00
HEK2_106	0.2	0.21	0.21	0.06	0.03	0.21	0.15
HEK2_110	0.2	0.00	0.00	0.04	0.02	0.00	0.00
HEK2_119	0.2	0.00	0.00	0.05	0.05	0.00	0.00
HEK2_121	0.1	0.00	0.02	0.05	0.10	0.06	0.07
HEK2_139	0.1	0.11	0.09	0.02	0.03	0.11	0.11

HEK2_145	0.1	0.00	0.00	0.00	0.00	0.00	0.00
----------	-----	------	------	------	------	------	------

[표 20]

Mutation frequencies of ABE7.10, BE3, and Cas9 at on-target and off-target sites captured by Digenome-seq (RNF2; 'DNA sequence at a cleavage sites'은 표 13 참조)

Site No.	DNA Cleavage Score	Base editing frequency (%)				Indel frequency (%)	
		Untreated	(+) ABE7.10	Untreated	(+) BE3	Untreated	(+) Cas9
RNF2-1	7.3	0.03	0.02	0.14	0.09	0.00	0.01
RNF2-2	4.4	0.02	26.26	0.05	78.82	0.01	88.29
RNF2-3	4.3	0.05	0.07	0.29	0.25	0.00	0.01
RNF2-4	4.1	0.05	0.05	0.00	0.00	0.01	0.01
RNF2-5	3.6	0.06	0.07	0.10	0.12	0.00	0.01
RNF2-6	2.2	0.06	0.09	0.09	0.10	0.02	0.02
RNF2-7	1.4	0.13	0.08	0.05	0.07	0.01	0.01
RNF2-8	1.0	0.08	0.10	0.04	0.05	0.01	0.00
RNF2-15	0.4	0.02	0.01	0.11	0.09	0.01	0.02
RNF2-19	0.4	0.00	0.02	0.06	0.06	0.07	0.09
RNF2-20	0.3	0.08	0.12	0.06	0.05	0.04	0.04
RNF2-24	0.3	0.03	0.03	0.02	0.05	0.29	0.35
RNF2-27	0.3	0.02	0.03	0.06	0.06	0.01	0.00
RNF2-41	0.2	0.03	0.01	0.00	0.00	0.03	0.03

5

[표 21]

Mutation frequencies of ABE7.10, BE3, and Cas9 at on-target and off-target sites captured by Digenome-seq (TYRO3; 'DNA sequence at a cleavage sites'은 표 14 참조)

Site No.	DNA Cleavage Score	Base editing frequency (%)				Indel frequency (%)	
		Untreated	(+) ABE7.10	Untreated	(+) BE3	Untreated	(+) Cas9
TYRO3-1	4.3	0.30	7.77	0.08	3.84	0.06	12.94
TYRO3-2	3.4	0.48	71.60	0.07	39.78	0.05	49.28
TYRO3-3	2.0	3.38	3.51	0.18	0.27	0.00	0.14
TYRO3-4	0.8	0.26	0.21	0.15	0.16	0.00	0.00
TYRO3-5	0.6	0.50	0.38	0.09	0.10	0.00	0.02
TYRO3-6	0.4	0.25	0.24	0.08	0.06	0.01	0.00
TYRO3-7	0.3	0.27	0.27	0.51	0.38	0.01	0.01

TYR03-8	0.2	0.00	0.00	0.26	0.19	0.00	0.00
TYR03-9	0.2	0.26	0.23	0.06	0.10	0.00	0.00
TYR03-10	0.2	0.08	0.10	0.05	0.02	0.00	0.01
TYR03-11	0.1	0.07	0.11	0.00	0.00	0.00	0.00
TYR03-12	0.1	0.33	0.35	0.19	0.16	0.00	0.00

[표 22]

Mutation frequencies of ABE7.10, BE3, and Cas9 at on-target and off-target sites captured by Digenome-seq (WEE1; 'DNA sequence at a cleavage sites'은 표 15 참조)

Site No.	DNA Cleavage Score	Base editing frequency (%)				Indel frequency (%)	
		Untreated	(+) ABE7.10	Untreated	(+) BE3	Untreated	(+) Cas9
WEE1-1	11.1	0.26	0.26	0.00	0.00	0.00	0.06
WEE1-2	2.0	0.05	62.59	0.11	9.95	0.00	56.24
WEE1-3	1.9	0.38	0.42	0.00	0.00	0.00	4.13
WEE1-4	0.7	0.13	0.11	0.06	0.03	0.00	0.01
WEE1-5	0.4	0.08	0.11	0.00	0.00	0.00	0.01
WEE1-6	0.4	0.10	0.10	0.06	0.05	0.00	0.01
WEE1-7	0.3	0.04	0.06	0.00	0.00	0.00	0.02
WEE1-8	0.3	0.25	0.24	0.00	0.00	0.01	0.00
WEE1-9	0.2	0.05	0.04	0.12	0.17	0.01	0.01
WEE1-10	0.2	0.11	0.07	0.17	0.15	0.04	0.05
WEE1-11	0.2	0.24	0.29	0.00	0.00	0.46	0.43
WEE1-12	0.2	No PCR					
WEE1-13	0.2	0.67	0.69	0.00	0.00	0.40	0.29
WEE1-14	0.2	0.00	0.00	0.13	0.13	0.00	0.01
WEE1-15	0.1	0.14	0.11	0.00	0.00	0.01	0.01
WEE1-16	0.1	0.00	0.03	1.37	1.29	0.00	0.00
WEE1-17	0.1	0.14	0.08	0.13	0.20	0.02	0.00
WEE1-18	0.1	0.09	0.11	0.00	0.00	0.00	0.01
WEE1-19	0.1	0.11	0.12	0.00	0.00	0.01	0.00
WEE1-20	0.1	0.26	0.25	0.00	0.00	0.00	0.00
WEE1-21	0.1	0.09	0.12	0.27	0.19	0.02	0.00

[표 23]

Mutation frequencies of ABE7.10, BE3, and Cas9 at on-target and off-target sites captured by Digenome-seq (EphB4; 'DNA sequence at a

cleavage sites'은 표 16 참조)

	DNA Cleavage Score	Base editing frequency (%)				Indel frequency (%)	
		Untreated	(+) ABE7.10	Untreated	(+) BE3	Untreated	(+) Cas9
EphB4-01	10.3	0.12	57.62	0.14	0.17	0.01	74.73
EphB4-02	8.0	0.15	0.23	0.00	0.00	0.00	1.02
EphB4-03	4.1	0.09	0.11	0.26	0.30	0.01	0.14
EphB4-04	2.4	0.09	0.10	0.05	0.06	0.01	0.01
EphB4-05	2.2	0.14	0.13	0.08	0.09	0.01	0.00
EphB4-06	1.9	No PCR					
EphB4-07	1.8	0.33	0.30	0.09	0.10	0.00	0.01
EphB4-08	1.8	0.12	0.15	0.07	0.14	0.00	0.00
EphB4-09	1.4	0.50	0.42	0.00	0.00	0.01	0.01
EphB4-10	0.9	0.07	0.08	0.00	0.00	0.01	0.00
EphB4-11	0.8	0.07	0.08	0.16	0.16	0.01	0.02
EphB4-12	0.6	0.52	0.54	0.00	0.00	0.01	0.01
EphB4-13	0.5	0.15	0.14	0.11	0.09	0.01	0.01
EphB4-14	0.4	0.15	0.14	0.10	0.13	0.01	0.00
EphB4-15	0.4	0.12	0.10	0.00	0.00	0.01	0.01
EphB4-16	0.4	0.06	0.08	0.13	0.11	0.01	0.00
EphB4-17	0.4	0.20	0.20	0.00	0.00	0.32	0.33
EphB4-18	0.4	0.60	0.69	0.35	0.30	0.00	0.01
EphB4-19	0.3	0.11	0.09	0.10	0.10	0.00	0.00
EphB4-20	0.3	0.23	0.20	0.00	0.00	0.98	1.10
EphB4-21	0.3	No PCR					
EphB4-22	0.2	0.04	0.04	0.00	0.00	0.00	0.00
EphB4-23	0.2	0.05	0.06	0.05	0.05	0.12	0.11
EphB4-24	0.2	0.27	0.31	99.98	99.98	0.00	0.01
EphB4-25	0.2	0.06	0.06	0.00	0.00	0.00	0.00
EphB4-26	0.2	0.13	0.10	0.00	0.00	0.00	0.00
EphB4-27	0.2	No PCR					
EphB4-28	0.2	0.29	0.27	0.12	0.13	0.00	0.00
EphB4-29	0.2	0.55	0.48	0.00	0.00	0.00	0.01
EphB4-30	0.2	0.10	0.08	0.03	0.04	0.00	0.00
EphB4-31	0.1	0.15	0.18	0.00	0.00	0.24	0.21
EphB4-32	0.1	0.54	0.45	0.03	0.01	0.00	0.00
EphB4-33	0.1	0.23	0.15	0.15	0.11	0.03	0.01

EphB4-34	0.1	0.15	0.13	0.21	0.27	0.01	0.00
EphB4-35	0.1	6.14	6.02	0.20	0.31	0.00	0.00
EphB4-36	0.1	No PCR					
EphB4-37	0.1	0.22	0.19	0.00	0.00	0.00	0.00
EphB4-38	0.1	0.12	0.12	0.43	0.40	0.11	0.13
EphB4-39	0.1	0.11	0.11	0.00	0.00	0.01	0.00

[표 24]

Mutation frequencies of ABE7.10, BE3, and Cas9 at on-target and off-target sites captured by Digenome-seq (HPRT-Exon6; 'DNA sequence at a cleavage sites'은 표 17 참조)

Site No.	DNA Cleavage Score	Base editing frequency (%)				Indel frequency (%)	
		Untreated	(+) ABE7.10	Untreated	(+) BE3	Untreated	(+) Cas9
HPRT_E6-01	29.0	0.11	62.07	0.05	29.25	0.00	71.69
HPRT_E6-02	21.3	0.22	0.21	0.01	0.01	0.00	0.00
HPRT_E6-03	17.0	0.08	0.18	0.03	0.01	0.00	0.00
HPRT_E6-04	12.7	0.38	0.40	0.00	0.00	0.00	0.00
HPRT_E6-05	10.5	0.15	0.14	0.00	0.00	0.00	0.00
HPRT_E6-06	10.2	0.30	0.31	0.59	0.44	0.00	0.00
HPRT_E6-07	9.3	0.02	1.59	0.28	0.28	0.81	17.56
HPRT_E6-08	9.0	0.12	0.12	0.04	0.03	0.00	0.17
HPRT_E6-09	6.1	0.07	0.05	0.04	0.06	0.00	0.00
HPRT_E6-10	5.6	0.24	0.14	0.13	0.12	0.00	0.01
HPRT_E6-11	4.7	0.15	0.17	0.01	0.02	0.00	0.01
HPRT_E6-12	4.5	0.26	0.31	0.07	0.06	0.00	0.27
HPRT_E6-13	4.3	0.35	1.89	0.05	0.29	0.00	0.00
HPRT_E6-14	4.3	0.10	0.11	0.15	0.17	0.00	0.00
HPRT_E6-15	4.1	0.07	0.06	0.11	0.07	0.00	0.00
HPRT_E6-16	3.8	0.10	0.34	0.41	1.74	0.01	0.00
HPRT_E6-17	3.7	0.08	0.09	0.10	0.08	0.00	0.00
HPRT_E6-18	3.6	0.26	0.31	0.36	0.39	0.00	0.00
HPRT_E6-19	3.6	0.06	0.05	0.00	0.00	0.00	0.00
HPRT_E6-20	3.5	0.00	0.00	0.05	0.08	0.00	0.15
HPRT_E6-21	3.5	0.07	0.08	0.30	0.32	0.03	0.03
HPRT_E6-22	3.4	0.16	0.21	0.00	0.00	0.00	0.01
HPRT_E6-23	3.3	0.33	0.22	0.07	0.08	0.00	0.01

HPRT_E6-24	3.1	0.09	0.06	0.06	0.05	0.00	0.01
HPRT_E6-25	2.9	0.13	0.10	0.00	0.00	0.01	0.00
HPRT_E6-26	2.8	0.11	0.07	0.04	0.07	0.00	0.00
HPRT_E6-27	2.6	0.22	0.27	0.00	0.00	0.00	0.00
HPRT_E6-28	2.5	0.33	0.29	0.27	0.20	0.00	0.01
HPRT_E6-29	2.4	0.23	0.24	0.04	0.03	0.07	0.09
HPRT_E6-30	2.4	0.50	0.49	0.03	0.04	0.00	0.00
HPRT_E6-31	2.3	0.21	0.24	0.00	0.00	0.00	0.00
HPRT_E6-32	2.2	0.24	0.29	0.05	0.07	0.01	0.00
HPRT_E6-33	1.5	0.22	0.23	0.08	0.07	0.01	0.01
HPRT_E6-34	1.4	0.00	0.00	0.63	0.54	0.00	0.00
HPRT_E6-35	1.4	0.20	0.22	0.02	0.05	0.00	0.00
HPRT_E6-36	1.2	0.05	0.07	0.05	0.03	0.15	0.10
HPRT_E6-37	1.1	0.02	0.01	0.07	0.07	0.00	0.00
HPRT_E6-38	1.1	0.16	0.25	0.06	0.89	0.01	0.00
HPRT_E6-39	1.1	0.04	0.04	0.11	0.12	0.00	0.00
HPRT_E6-40	1.0	0.12	0.09	0.00	0.00	0.00	0.01
HPRT_E6-41	1.0	0.25	0.30	0.00	0.00	0.00	0.00
HPRT_E6-52	0.7	0.00	0.00	0.01	0.06	0.01	0.00
HPRT_E6-81	0.4	0.21	0.23	0.08	0.08	0.00	0.00

[표 25]

Mutation frequencies of ABE7.10, BE3, and Cas9 at on-target and off-target sites captured by Digenome-seq (HPRT-Exon8; 'DNA sequence at a cleavage sites'은 표 18 참조)

Site No.	DNA Cleavage Score	Base editing frequency (%)				Indel frequency (%)	
		Untreated	(+) ABE7.10	Untreated	(+) BE3	Untreated	(+) Cas9
HPRT_E8-1	2.4	0.02	70.04	0.05	27.31	0.07	64.71
HPRT_E8-2	1.3	0.20	5.77	0.09	0.58	0.07	50.40
HPRT_E8-3	0.6	0.16	0.16	0.05	0.02	0.01	0.04
HPRT_E8-4	0.6	0.04	0.06	63.15	63.00	0.01	0.01
HPRT_E8-5	0.6	0.10	0.08	0.10	0.11	0.00	0.02
HPRT_E8-6	0.5	0.03	0.07	N/A	N/A	0.01	0.19
HPRT_E8-7	0.5	0.16	0.15	0.07	0.11	0.00	0.00
HPRT_E8-8	0.4	0.07	0.07	0.08	0.05	0.03	0.04
HPRT_E8-9	0.4	0.02	0.03	0.06	0.09	0.02	0.04

HPRT_E8-10	0.4	0.01	0.01	N/A	N/A	0.01	0.03
HPRT_E8-11	0.4	0.08	0.07	0.09	0.09	0.04	0.03
HPRT_E8-12	0.3	0.10	0.09	0.05	0.05	0.00	0.01
HPRT_E8-13	0.3	0.07	0.03	0.06	0.05	0.00	0.03
HPRT_E8-14	0.3	0.08	0.14	0.67	0.48	0.02	0.00
HPRT_E8-15	0.3	0.10	0.07	0.09	0.07	0.00	0.01
HPRT_E8-16	0.2	0.02	0.02	0.08	0.09	0.02	0.05
HPRT_E8-17	0.2	0.10	0.11	0.05	0.05	0.00	0.02
HPRT_E8-18	0.2	0.05	0.08	0.14	0.10	0.01	0.20
HPRT_E8-19	0.1	0.05	0.06	0.06	0.03	0.03	0.01
HPRT_E8-20	0.1	0.15	0.17	0.04	0.04	0.01	0.00

상기 표 19 내지 25에 나타난 바와 같이, 173개 부위에서, 단일 뉴클레오타이드 치환은 시퀀싱 오류로 인해 발생하는 노이즈 수준 (일반적으로 0.1~1% 범위) 이상의 빈도로 검출되지 않았다. 다른 20개 validated sites 중에서 7개 on-target 부위는 29% 내지 72% 사이의 염기 교정 빈도를 나타내었다. 다른 13개의 validated off-target site은 0.1% 내지 7.8% 범위의 염기교정 빈도를 나타내서, 매우 낮은 교정 효율을 보였다. 이러한 결과는 Digenome-seq 분석법은 검출 한계 근처의 매우 낮은 교정 빈도를 갖는 ABE off-target site까지도 검출 가능할 정도로 충분히 민감한 분석법임을 보여준다.

상기 시험과 병행하여, 193 부위에서 Cas9-유도 indel 빈도와 BE3-유도 치환 빈도를 측정하여 ABE7.10와 비교하여 그 결과를 도 12f에 나타내었다. 도 12f에 나타난 바와 같이, ABE7.10 활성은, BE3 활성($R^2=0.39$)보다, Cas9 활성($R^2=0.77$)과 보다 강하게 연관성을 갖는 것으로 확인되었다. 도 12f의 위쪽 그래프에 있어서, ABE7.10에 의해서는 염기 교정 빈도가 비교적 높지만 BE3에 의해서는 염기교정빈도가 매우 낮은 (0.001% 이하) 6개 부위 (속이 빈 점으로 표시) 중에서, 2 개 부위는 BE3 editing window 내에 시토신을 포함하지 않는다. 도 12f의 아래쪽 그래프에 있어서, 6 개의 부위(속이 빈 점으로 표시)에서 ABE7.10에 의한 염기 교정 빈도가 높지만 Cas9에 의한 indel 빈도는 매우 낮게 (indel 빈도 0.001% 이하) 나타났으며, 이 중 3 개 부위(50%=3/6)는 각각의 on-target 부위와 비교하여 하나의 삽입 또는 결실된 뉴클레오타이드를 가지며, sgRNA와 혼성화 시 DNA 또는 RNA bulge를 형성할 수 있다. 이러한 bulge-형성

부위는 BE3의 off-target site의 공통적인 특징이지만 Cas9 off-target site에서는 거의 관찰되지 않는다. Digenome-seq을 통해 ABE7.10을 사용하여 확인된 193 개의 부위 중 3 개 또는 2 개의 부위는 각각 Cas9 또는 BE3에 의하여 세포 내에서 높은 빈도로 염기 교정 또는 유전자 교정이 일어났지만, ABE7.10에 의해서는 매우 낮은 염기 교정 빈도 (치환 빈도 0.001% 이하)를 나타냈다 (도 12f에서 회색 점으로 표시). 이러한 결과는 ABE7.10, BE3, 및 Cas9가 인간 유전체에서 상이한 세트의 off-target site을 인식할 수 있음을 재차 확인시켜 준다.

10 실시예 8. ABE off-target 효과 시험

ABE의 off-target 효과를 최소화하거나 제거하기 위하여, 다음의 세 가지 방법으로 시험하였다: sgRNA 변형, 플라스미드 대신 RNP 세포내 전달, 및 Sniper-Cas9 (*E. coli* 에서 directed evolution를 통하여 얻어진 engineered Cas9 변이체; 야생형 SpCas9 아미노산 서열(NP_269215.1)에 대하여 E539S, M761I, 및 K890N 변이를 가지며, D10A 변이를 추가로 가짐) 사용.

첫째로 (sgRNA 변형), sgRNA의 표적화 서열이 20개 뉴클레오타이드로 이루어진 표준 GX₁₉ sgRNA (X₁₉: 19개의 임의의 뉴클레오타이드 (각각 독립적으로 A, U(T), C, 및 G 중에서 선택됨)를 의미함; X 뒤의 숫자는 임의의 뉴클레오타이드 개수를 의미함, 이하 동일함)의 'X'의 5' 말단 부위로부터 1 또는 2개 뉴클레오타이드를 제거한 절단 sgRNA (gX₁₈ 또는 gX₁₇로 표시) (g는 표적 부위의 대응 서열과 불일치하는 구아닌을, G는 표적 부위의 대응 서열과 일치하는 구아닌을 나타냄) 또는 5' 말단에 1 또는 2개의 여분의 구아닌을 추가로 포함하는 연장된 sgRNA(gX₂₀ 또는 ggX₂₀이라고 함)를 사용하여 HEK293T 세포에서의 염기 교정 빈도를 측정하였다. 상기 사용된 변형 sgRNA를 HPRT exon 6 target sequence를 대표로 예시하면 도 13a와 같다. 다른 유전자에 대한 변형 sgRNA는 표 11에 기재된 각 유전자의 On-target DNA sequence을 기초로 도 13a를 참조하여 제작하였다.

ABE7.10 및 상기 변형 sgRNA를 사용하여 얻어진 HEK293T 세포에서의 염기 교정 빈도 결과(targeted deep sequencing으로 측정됨)를 도 13b 내지 도 13d 및 표 26 내지 표 32에 나타내었다:

[표 26]

변형 sgRNAs를 사용한 경우의 ABE7.10 off-target 효과 분석 (HEK2;

'DNA sequence at a cleavage sites'은 표 12 참조)

Site No.	DNA Cleavage Score	Base editing frequency (%)					
		Untreated	(+) ABE7.10				
			gX17	gX18	GX19	gX20	ggX20
HEK2_001	20.1	0.12	0.22	0.15	0.11	0.12	0.13
HEK2_002	15.1	0.00	0.00	0.00	0.00	0.00	0.00
HEK2_003	14.7	0.06	0.07	0.04	0.09	0.08	0.07
HEK2_004	14.3	0.10	0.10	0.11	0.31	0.21	0.11
HEK2_005	13.4	0.15	0.14	0.16	0.13	0.13	0.12
HEK2_006	13.4	0.00	0.00	0.00	0.00	0.00	0.00
HEK2_007	8.9	0.06	9.80	47.71	63.19	61.77	55.40
HEK2_008	6.6	0.06	0.07	0.05	0.12	0.12	0.08
HEK2_009	5.8	0.10	0.08	0.08	0.09	0.11	0.12
HEK2_010	5.1	0.00	0.00	0.00	0.00	0.00	0.00
HEK2_011	4.9	0.09	0.10	0.07	0.08	0.09	0.09
HEK2_012	4.6	0.05	0.05	0.03	0.06	0.04	0.04
HEK2_013	3.9	0.00	0.00	0.00	0.00	0.00	0.00
HEK2_014	3.5	0.08	0.09	0.13	0.08	0.05	0.07
HEK2_015	3.0	0.11	0.09	0.14	0.13	0.13	0.14
HEK2_016	2.9	No PCR					
HEK2_017	2.7	0.13	0.07	0.15	0.07	0.12	0.10
HEK2_018	2.5	0.18	0.25	0.21	0.20	0.23	0.23
HEK2_019	2.2	0.12	0.13	0.13	0.12	0.11	0.12
HEK2_020	2.1	0.12	0.20	0.11	0.20	0.14	0.14
HEK2_021	2.1	0.17	0.18	0.18	0.19	0.21	0.20
HEK2_022	2.1	0.13	0.10	0.09	0.11	0.14	0.13
HEK2_023	2.0	0.09	0.13	0.05	0.09	0.12	0.08
HEK2_024	1.7	0.00	0.00	0.00	0.00	0.00	0.00
HEK2_025	1.4	0.13	0.14	0.13	0.14	0.11	0.13
HEK2_026	1.3	0.13	0.12	0.15	0.16	0.11	0.16
HEK2_027	1.3	0.15	0.19	0.21	0.16	0.18	0.18
HEK2_028	1.3	0.00	0.00	0.00	0.00	0.00	0.00
HEK2_029	1.2	0.23	0.20	0.16	0.15	0.18	0.14
HEK2_030	1.2	0.09	0.11	0.10	0.11	0.08	0.09
HEK2_031	1.2	0.09	0.12	0.10	0.12	0.08	0.09
HEK2_032	1.1	0.05	0.07	0.12	0.09	0.10	0.06
HEK2_033	1.1	0.11	0.11	0.14	0.22	0.14	0.16
HEK2_034	1.0	0.05	0.05	0.06	0.05	0.05	0.05

HEK2_042	0.7	0.03	0.07	0.04	0.04	0.05	0.04
HEK2_044	0.7	0.00	0.00	0.00	0.00	0.00	0.00
HEK2_050	0.5	0.22	0.23	0.13	0.19	0.16	0.19
HEK2_052	0.5	0.13	0.10	0.09	0.11	0.07	0.11
HEK2_065	0.4	0.11	0.12	0.07	0.15	0.14	0.12
HEK2_079	0.3	0.00	0.00	0.00	0.00	0.00	0.00
HEK2_082	0.3	0.00	0.00	0.00	0.00	0.00	0.00
HEK2_092	0.2	0.11	0.08	0.11	0.08	0.06	0.10
HEK2_093	0.2	0.02	0.00	0.02	0.03	0.03	0.04
HEK2_099	0.2	0.06	0.15	0.13	0.08	0.10	0.13
HEK2_101	0.2	0.00	0.00	0.00	0.00	0.00	0.00
HEK2_105	0.2	0.00	0.00	0.00	0.00	0.00	0.00
HEK2_106	0.2	0.21	0.23	0.26	0.21	0.27	0.19
HEK2_110	0.2	0.00	0.00	0.00	0.00	0.00	0.00
HEK2_119	0.2	0.00	0.00	0.00	0.00	0.00	0.00
HEK2_121	0.1	0.00	0.02	0.04	0.02	0.07	0.02
HEK2_139	0.1	0.11	0.12	0.09	0.09	0.17	0.14
HEK2_145	0.1	0.00	0.00	0.00	0.00	0.00	0.00

[표 27]

변형 sgRNAs를 사용한 경우의 ABE7.10 off-target 효과 분석 (RNF2; 'DNA sequence at a cleavage sites'은 표 13 참조)

Site No.	DNA Cleavage Score	Base editing frequency (%)					
		Untreated	(+) ABE7.10				
			gX17	gX18	GX19	gX20	ggX20
RNF2-1	7.3	0.03	0.01	0.02	0.02	0.01	0.03
RNF2-2	4.4	0.02	0.07	0.63	26.26	29.43	16.19
RNF2-3	4.3	0.05	0.07	0.07	0.07	0.05	0.06
RNF2-4	4.1	0.05	0.05	0.05	0.05	0.04	0.05
RNF2-5	3.6	0.06	0.06	0.08	0.07	0.07	0.08
RNF2-6	2.2	0.06	0.08	0.04	0.09	0.06	0.05
RNF2-7	1.4	0.13	0.15	0.09	0.08	0.17	0.16
RNF2-8	1.0	0.08	0.11	0.10	0.10	0.08	0.07
RNF2-15	0.4	0.02	0.02	0.03	0.01	0.05	0.02
RNF2-19	0.4	0.00	0.02	0.02	0.02	0.00	0.01
RNF2-20	0.3	0.08	0.10	0.10	0.12	0.10	0.10
RNF2-24	0.3	0.03	0.01	0.02	0.03	0.03	0.03
RNF2-27	0.3	0.02	0.03	0.05	0.03	0.03	0.01
RNF2-41	0.2	0.03	0.01	0.01	0.01	0.01	0.01

[표 28]

변형 sgRNAs를 사용한 경우의 ABE7.10 off-target 효과 분석 (TYRO3;
'DNA sequence at a cleavage sites'은 표 14 참조)

Site No.	DNA Cleavage Score	Base editing frequency (%)					
		Untreated	(+) ABE7.10				
			gX17	gX18	GX19	gX20	ggX20
TYRO3-1	4.3	0.30	0.26	5.32	7.77	9.26	8.32
TYRO3-2	3.4	0.48	0.58	36.53	71.60	67.78	54.99
TYRO3-3	2.0	3.38	3.44	3.40	3.51	3.34	3.53
TYRO3-4	0.8	0.26	0.20	0.21	0.21	0.24	0.24
TYRO3-5	0.6	0.50	0.41	0.50	0.38	0.51	0.42
TYRO3-6	0.4	0.25	0.25	0.24	0.24	0.26	0.25
TYRO3-7	0.3	0.27	0.31	0.25	0.27	0.31	0.27
TYRO3-8	0.2	0.00	0.00	0.00	0.00	0.00	0.00
TYRO3-9	0.2	0.26	0.21	0.24	0.23	0.21	0.21
TYRO3-10	0.2	0.08	0.06	0.07	0.10	0.05	0.08
TYRO3-11	0.1	0.07	0.10	0.10	0.11	0.09	0.08
TYRO3-12	0.1	0.33	0.30	0.37	0.35	0.32	0.30

[표 29]

5 변형 sgRNAs를 사용한 경우의 ABE7.10 off-target 효과 분석 (WEE1;
'DNA sequence at a cleavage sites'은 표 15 참조)

Site No.	DNA Cleavage Score	Base editing frequency (%)					
		Untreated	(+) ABE7.10				
			gX17	gX18	GX19	gX20	ggX20
WEE1-1	11.1	0.26	0.26	0.24	0.26	0.26	0.23
WEE1-2	2.0	0.05	0.25	0.95	62.59	37.80	26.30
WEE1-3	1.9	0.38	1.80	0.40	0.42	0.42	0.39
WEE1-4	0.7	0.13	0.08	0.09	0.11	0.11	0.11
WEE1-5	0.4	0.08	0.10	0.12	0.11	0.10	0.07
WEE1-6	0.4	0.10	0.12	0.12	0.10	0.07	0.10
WEE1-7	0.3	0.04	0.03	0.04	0.06	0.05	0.06
WEE1-8	0.3	0.25	0.23	0.22	0.24	0.20	0.29
WEE1-9	0.2	0.05	0.03	0.04	0.04	0.03	0.06
WEE1-10	0.2	0.11	0.10	0.14	0.07	0.11	0.10
WEE1-11	0.2	0.24	0.22	0.26	0.29	0.22	0.18
WEE1-12	0.2	No PCR					
WEE1-13	0.2	0.67	0.71	0.66	0.69	0.69	0.70

WEE1-14	0.2	0.00	0.00	0.00	0.00	0.00	0.00
WEE1-15	0.1	0.14	0.13	0.12	0.11	0.15	0.13
WEE1-16	0.1	0.00	0.01	0.03	0.03	0.01	0.02
WEE1-17	0.1	0.14	0.10	0.11	0.08	0.12	0.11
WEE1-18	0.1	0.09	0.09	0.06	0.11	0.11	0.08
WEE1-19	0.1	0.11	0.08	0.10	0.12	0.09	0.08
WEE1-20	0.1	0.26	0.26	0.24	0.25	0.27	0.24
WEE1-21	0.1	0.09	0.09	0.15	0.12	0.12	0.13

[표 30]

변형 sgRNAs를 사용한 경우의 ABE7.10 off-target 효과 분석 (EPHB4; 'DNA sequence at a cleavage sites'은 표 16 참조)

Site No.	DNA Cleavage Score	Base editing frequency (%)					
		Untreated	(+) ABE7.10				
			gX17	gX18	GX19	gX20	ggX20
EphB4-01	10.3	0.12	0.26	0.95	57.62	33.54	22.95
EphB4-02	8.0	0.15	0.16	0.13	0.23	0.14	0.13
EphB4-03	4.1	0.09	0.11	0.09	0.11	0.10	0.11
EphB4-04	2.4	0.09	0.06	0.11	0.10	0.10	0.10
EphB4-05	2.2	0.14	0.12	0.12	0.13	0.11	0.13
EphB4-06	1.9	No PCR					
EphB4-07	1.8	0.33	0.29	0.29	0.30	0.35	0.29
EphB4-08	1.8	0.12	0.12	0.09	0.15	0.19	5.82
EphB4-09	1.4	0.50	0.49	0.49	0.42	0.44	0.40
EphB4-10	0.9	0.07	0.07	0.09	0.08	0.09	0.07
EphB4-11	0.8	0.07	0.07	0.07	0.08	0.07	0.05
EphB4-12	0.6	0.52	0.57	0.57	0.54	0.53	0.58
EphB4-13	0.5	0.15	0.16	0.16	0.14	0.14	0.15
EphB4-14	0.4	0.15	0.14	0.12	0.14	0.13	0.14
EphB4-15	0.4	0.12	0.11	0.10	0.10	0.11	0.10
EphB4-16	0.4	0.06	0.06	0.06	0.08	0.05	0.06
EphB4-17	0.4	0.20	0.16	0.18	0.20	0.17	0.18
EphB4-18	0.4	0.60	0.64	0.60	0.69	0.74	0.64
EphB4-19	0.3	0.11	0.10	0.09	0.09	0.07	0.10
EphB4-20	0.3	0.23	0.11	0.19	0.20	0.19	0.18
EphB4-21	0.3	No PCR					
EphB4-22	0.2	0.04	0.04	0.06	0.04	0.05	0.03
EphB4-23	0.2	0.05	0.05	0.06	0.06	0.08	0.08
EphB4-24	0.2	0.27	0.22	0.30	0.31	0.35	0.32

EphB4-25	0.2	0.06	0.05	0.09	0.06	0.05	0.06
EphB4-26	0.2	0.13	0.10	0.10	0.10	0.10	0.11
EphB4-27	0.2	No PCR					
EphB4-28	0.2	0.29	0.26	0.33	0.27	0.31	0.31
EphB4-29	0.2	0.55	0.40	0.37	0.48	0.37	0.40
EphB4-30	0.2	0.10	0.05	0.05	0.08	0.07	0.07
EphB4-31	0.1	0.15	0.13	0.15	0.18	0.19	0.15
EphB4-32	0.1	0.54	0.44	0.36	0.45	0.46	0.48
EphB4-33	0.1	0.23	0.21	0.15	0.15	0.15	0.23
EphB4-34	0.1	0.15	0.14	0.14	0.13	0.14	0.14
EphB4-35	0.1	6.14	5.82	6.08	6.02	6.27	5.68
EphB4-36	0.1	No PCR					
EphB4-37	0.1	0.22	0.16	0.21	0.19	0.17	0.17
EphB4-38	0.1	0.12	0.16	0.14	0.12	0.14	0.13
EphB4-39	0.1	0.11	0.08	0.09	0.11	0.09	0.11

[표 31]

변형 sgRNAs를 사용한 경우의 ABE7.10 off-target 효과 분석 (HPRT-Exon6; 'DNA sequence at a cleavage sites'은 표 17 참조)

Site No.	DNA Cleavage Score	Base editing frequency (%)					
		Untreated	(+) ABE7.10				
			gX17	gX18	GX19	gX20	ggX20
HPRT_E6-01	29.0	0.11	16.94	29.23	62.07	60.39	58.46
HPRT_E6-02	21.3	0.22	0.15	0.17	0.21	0.18	0.17
HPRT_E6-03	17.0	0.08	0.11	0.10	0.18	0.10	0.07
HPRT_E6-04	12.7	0.38	0.45	0.42	0.40	0.43	0.39
HPRT_E6-05	10.5	0.15	0.22	0.14	0.14	0.15	0.15
HPRT_E6-06	10.2	0.30	0.23	0.35	0.31	0.36	0.28
HPRT_E6-07	9.3	0.02	0.03	0.05	1.59	0.06	0.06
HPRT_E6-08	9.0	0.12	0.10	0.08	0.12	0.11	0.10
HPRT_E6-09	6.1	0.07	0.03	0.06	0.05	0.04	0.06
HPRT_E6-10	5.6	0.24	0.19	0.15	0.14	0.18	0.12
HPRT_E6-11	4.7	0.15	0.22	0.25	0.17	0.15	0.16
HPRT_E6-12	4.5	0.26	0.22	0.26	0.31	0.26	0.26
HPRT_E6-13	4.3	0.35	0.55	0.45	1.89	1.26	0.59
HPRT_E6-14	4.3	0.10	0.11	0.13	0.11	0.08	0.09
HPRT_E6-15	4.1	0.07	0.05	0.09	0.06	0.04	0.08
HPRT_E6-16	3.8	0.10	6.37	0.26	0.34	0.08	0.08
HPRT_E6-17	3.7	0.08	0.17	0.10	0.09	0.10	0.12

HPRT_E6-18	3.6	0.26	0.31	0.37	0.31	0.37	0.39
HPRT_E6-19	3.6	0.06	0.06	0.09	0.05	0.04	0.05
HPRT_E6-20	3.5	0.00	0.09	0.08	0.00	0.09	0.00
HPRT_E6-21	3.5	0.07	0.06	0.11	0.08	0.09	0.08
HPRT_E6-22	3.4	0.16	0.15	0.19	0.21	0.17	0.14
HPRT_E6-23	3.3	0.33	0.26	0.27	0.22	0.25	0.28
HPRT_E6-24	3.1	0.09	0.07	0.07	0.06	0.09	0.11
HPRT_E6-25	2.9	0.13	0.11	0.11	0.10	0.16	0.12
HPRT_E6-26	2.8	0.11	0.08	0.05	0.07	0.12	0.09
HPRT_E6-27	2.6	0.22	0.28	0.27	0.27	0.27	0.22
HPRT_E6-28	2.5	0.33	0.39	0.32	0.29	0.35	0.29
HPRT_E6-29	2.4	0.23	0.21	0.16	0.24	0.20	0.19
HPRT_E6-30	2.4	0.50	0.46	0.50	0.49	0.51	0.45
HPRT_E6-31	2.3	0.21	0.25	0.23	0.24	0.24	0.17
HPRT_E6-32	2.2	0.24	0.27	2.20	0.29	0.25	0.28
HPRT_E6-33	1.5	0.22	0.24	0.17	0.23	0.22	0.17
HPRT_E6-34	1.4	0.00	0.00	0.00	0.00	0.00	0.00
HPRT_E6-35	1.4	0.20	0.16	0.24	0.22	0.23	0.23
HPRT_E6-36	1.2	0.05	0.04	0.03	0.07	0.05	0.05
HPRT_E6-37	1.1	0.02	0.08	0.03	0.01	0.03	0.08
HPRT_E6-38	1.1	0.16	0.33	0.22	0.25	0.22	0.19
HPRT_E6-39	1.1	0.04	0.06	0.04	0.04	0.04	0.04
HPRT_E6-40	1.0	0.12	0.13	0.12	0.09	0.10	0.12
HPRT_E6-41	1.0	0.25	0.19	0.11	0.30	0.40	0.10
HPRT_E6-52	0.7	0.00	0.00	0.00	0.00	0.00	0.00
HPRT_E6-81	0.4	0.21	0.20	0.19	0.23	0.22	0.23

[표 32]

변형 sgRNAs를 사용한 경우의 ABE7.10 off-target 효과 분석 (HPRT-Exon8; 'DNA sequence at a cleavage sites'은 표 18 참조)

Site No.	DNA Cleavage Score	Base editing frequency (%)					
		Untreated	(+) ABE7.10				
			gX17	gX18	GX19	gX20	ggX20
HPRT_E8-1	2.4	0.02	37.90	62.81	70.04	63.66	56.04
HPRT_E8-2	1.3	0.20	0.20	0.25	5.77	0.56	0.29
HPRT_E8-3	0.6	0.16	0.13	0.19	0.16	0.15	0.12
HPRT_E8-4	0.6	0.04	0.05	0.05	0.06	0.04	0.06
HPRT_E8-5	0.6	0.10	0.12	0.12	0.08	0.11	0.13
HPRT_E8-6	0.5	0.03	0.03	0.04	0.07	0.02	0.06

HPRT_E8-7	0.5	0.16	0.11	0.12	0.15	0.16	0.15
HPRT_E8-8	0.4	0.07	0.07	0.09	0.07	0.09	0.09
HPRT_E8-9	0.4	0.02	0.02	0.02	0.03	0.04	0.01
HPRT_E8-10	0.4	0.01	0.01	0.01	0.01	0.01	0.02
HPRT_E8-11	0.4	0.08	0.05	0.08	0.07	0.05	0.09
HPRT_E8-12	0.3	0.10	0.06	0.10	0.09	0.08	0.09
HPRT_E8-13	0.3	0.07	0.04	0.03	0.03	0.04	0.03
HPRT_E8-14	0.3	0.08	0.11	0.10	0.14	0.11	0.10
HPRT_E8-15	0.3	0.10	0.10	0.07	0.07	0.10	0.10
HPRT_E8-16	0.2	0.02	0.05	0.05	0.02	0.02	0.02
HPRT_E8-17	0.2	0.10	0.09	0.11	0.11	0.10	0.10
HPRT_E8-18	0.2	0.05	0.04	0.45	0.08	0.04	0.04
HPRT_E8-19	0.1	0.05	0.05	0.03	0.06	0.03	0.06
HPRT_E8-20	0.1	0.15	0.14	0.18	0.17	0.17	0.17

도 13b는 HEK2, RNF2, TK_EphB4, TYRO3, WEE1, HPRT-exon6, 및 HPRT-exon8를 표적으로 하는 변형 sgRNA 사용시 targeted deep sequencing에 의하여 측정된 각 변형 sgRNA 별 ABE7.10의 염기 교정 빈도 5 평균을 나타낸 그래프이다.

도 13c 및 13d에 나타난 바와 같이, 변형 sgRNA는 표적화 활성은 유지하면서 거의 모든 부위에서 ABE7.10의 off-target 활성을 감소시켰다 (도 13d에서, y축의 그래프는 위에서부터 순서대로 ggX20, gX20, GX19, gX18, gX17, 및 untreated의 결과를 보여준다). 예컨대, 도 13c에서와 10 같이, HPRT Exon 6을 표적으로 하는 GX₁₉ sgRNA를 사용하여 확인된 4 개의 검증된 off-target site에 대하여 gX₂₀ 또는 ggX₂₀ sgRNA는 기존 GX₁₉ sgRNA와 비교하여 specificity ratio가 2 내지 26배 향상시켰다. 도 13c 및 13d에 나타난 바와 같이, 절단 sgRNAs는 대부분의 부위에서 ABE7.10의 on target 및 off-target 활성을 감소시키고, 5' 말단 근처 또는 그 부근의 15 불일치가 있는 부위에서 off-target 효과를 악화시키는 것으로 나타났다.

또한, 야생형 SpCas9 대신에 Sniper-Cas9가 아데닌 디아미나아제 모이어티와 융합된 Sniper ABE7.10를 사용하여 상기와 동일한 시험을 수행하여, Digenome-seq로 측정된 결과를 도 13e 및 13f (도 13f에서, y축의 그래프는 위에서부터 순서대로 (+) SpCas9 ABE7.10, (+)ABE7.10, 및 20 untreated의 결과를 보여준다) 및 표 33 내지 표 39에 나타내었다:

[표 33]

Mutation frequencies of ABE7.10 and Sniper ABE7.10 at on-target and off-target sites captured by Digenome-seq (HEK2; 'DNA sequence at a cleavage sites'은 표 12 참조)

Site No.	DNA Cleavage Score	Base editing frequency (%)		
		Untreated	(+) ABE7.10	(+) Sniper ABE7.10
HEK2_001	20.1	0.17	0.17	0.16
HEK2_002	15.1	0.55	0.55	0.59
HEK2_003	14.7	0.02	0.05	0.07
HEK2_004	14.3	0.16	0.33	0.13
HEK2_005	13.4	0.16	0.22	0.22
HEK2_006	13.4	0.22	0.18	0.17
HEK2_007	8.9	0.31	72.76	75.92
HEK2_008	6.6	0.09	0.10	0.11
HEK2_009	5.8	0.11	0.10	0.09
HEK2_010	5.1	0.39	0.42	0.34
HEK2_011	4.9	0.12	0.13	0.11
HEK2_012	4.6	0.24	0.23	0.22
HEK2_013	3.9	0.20	0.24	0.20
HEK2_014	3.5	0.02	0.03	0.00
HEK2_015	3.0	0.17	0.21	0.22
HEK2_016	2.9	No PCR		
HEK2_017	2.7	0.20	0.22	0.18
HEK2_018	2.5	0.32	0.30	0.31
HEK2_019	2.2	0.37	0.36	0.43
HEK2_020	2.1	0.15	0.13	0.12
HEK2_021	2.1	0.50	0.42	0.47
HEK2_022	2.1	0.19	0.20	0.24
HEK2_023	2.0	6.76	6.74	7.27
HEK2_024	1.7	0.10	0.11	0.11
HEK2_025	1.4	0.22	0.24	0.23
HEK2_026	1.3	0.23	0.19	0.22
HEK2_027	1.3	0.40	0.39	0.41
HEK2_028	1.3	0.14	0.16	0.15
HEK2_029	1.2	0.22	0.20	0.23
HEK2_030	1.2	0.18	0.20	0.16
HEK2_031	1.2	0.10	0.16	0.13
HEK2_032	1.1	0.08	0.12	0.10

HEK2_033	1.1	0.18	0.21	0.20
HEK2_034	1.0	0.13	0.11	0.10
HEK2_042	0.7	0.20	0.21	0.24
HEK2_044	0.7	0.00	0.00	0.00
HEK2_050	0.5	0.12	0.13	0.13
HEK2_052	0.5	0.13	0.15	0.16
HEK2_065	0.4	0.23	0.26	0.22
HEK2_079	0.3	0.12	0.09	0.13
HEK2_082	0.3	0.21	0.22	0.21
HEK2_092	0.2	0.10	0.10	0.11
HEK2_093	0.2	0.12	0.10	0.09
HEK2_099	0.2	0.12	0.14	0.10
HEK2_101	0.2	0.19	0.18	0.17
HEK2_105	0.2	0.16	0.14	0.14
HEK2_106	0.2	0.27	0.29	0.30
HEK2_110	0.2	0.00	0.00	0.00
HEK2_119	0.2	0.10	0.12	0.13
HEK2_121	0.1	0.09	0.11	0.10
HEK2_139	0.1	0.25	0.31	0.28
HEK2_145	0.1	0.15	0.14	0.15

[표 34]

Mutation frequencies of ABE7.10 and Sniper ABE7.10 at on-target and off-target sites captured by Digenome-seq (RNF2; 'DNA sequence at a cleavage sites'은 표 13 참조)

Site No.	DNA Cleavage Score	Base editing frequency (%)		
		Untreated	(+) ABE7.10	(+) Sniper ABE7.10
RNF2-1	7.3	0.03	0.05	0.05
RNF2-2	4.4	0.22	32.79	16.24
RNF2-3	4.3	0.44	0.45	0.51
RNF2-4	4.1	0.18	0.17	0.19
RNF2-5	3.6	0.98	0.94	0.89
RNF2-6	2.2	0.35	0.33	0.42
RNF2-7	1.4	0.36	0.37	0.33
RNF2-8	1.0	0.32	0.35	0.27
RNF2-15	0.4	0.29	0.27	0.26
RNF2-19	0.4	0.22	0.23	0.24

RNF2-20	0.3	1.39	1.23	1.22
RNF2-24	0.3	0.11	0.18	0.18
RNF2-27	0.3	0.40	0.48	0.46
RNF2-41	0.2	0.02	0.02	0.04

[표 35]

Mutation frequencies of ABE7.10 and Sniper ABE7.10 at on-target and off-target sites captured by Digenome-seq (TYRO3; 'DNA sequence at a cleavage sites'은 표 14 참조)

Site No.	DNA Cleavage Score	Base editing frequency (%)		
		Untreated	(+) ABE7.10	(+) Sniper ABE7.10
TYRO3-1	4.3	0.64	3.04	3.12
TYRO3-2	3.4	1.17	72.66	68.95
TYRO3-3	2.0	11.24	11.54	11.48
TYRO3-4	0.8	0.36	0.36	0.32
TYRO3-5	0.6	1.02	1.07	1.02
TYRO3-6	0.4	0.68	0.63	0.61
TYRO3-7	0.3	0.54	0.44	0.76
TYRO3-8	0.2	0.00	0.00	0.00
TYRO3-9	0.2	0.98	0.85	0.95
TYRO3-10	0.2	0.16	0.18	0.07
TYRO3-11	0.1	0.32	0.22	0.28
TYRO3-12	0.1	0.70	0.88	0.82

5

[표 36]

Mutation frequencies of ABE7.10 and Sniper ABE7.10 at on-target and off-target sites captured by Digenome-seq (WEE1; 'DNA sequence at a cleavage sites'은 표 15 참조)

Site No.	DNA Cleavage Score	Base editing frequency (%)		
		Untreated	(+) ABE7.10	(+) Sniper ABE7.10
WEE1-1	11.1	0.80	0.73	0.76
WEE1-2	2.0	0.35	58.59	53.38
WEE1-3	1.9	2.38	2.32	2.08
WEE1-4	0.7	0.21	0.24	0.22
WEE1-5	0.4	0.18	0.17	0.17
WEE1-6	0.4	0.76	0.96	0.88
WEE1-7	0.3	0.12	0.12	0.11

WEE1-8	0.3	1.05	0.92	1.02
WEE1-9	0.2	0.59	0.61	0.58
WEE1-10	0.2	0.24	0.28	0.23
WEE1-11	0.2	0.23	0.21	0.13
WEE1-12	0.2	No PCR		
WEE1-13	0.2	1.45	1.50	1.65
WEE1-14	0.2	0.00	0.00	0.00
WEE1-15	0.1	0.46	0.41	0.56
WEE1-16	0.1	0.00	0.00	0.00
WEE1-17	0.1	0.18	0.17	0.13
WEE1-18	0.1	0.23	0.28	0.24
WEE1-19	0.1	0.32	0.45	0.32
WEE1-20	0.1	0.81	0.97	1.13
WEE1-21	0.1	0.21	0.24	0.25

[표 37]

Mutation frequencies of ABE7.10 and Sniper ABE7.10 at on-target and off-target sites captured by Digenome-seq (EphB4; 'DNA sequence at a cleavage sites'은 표 16 참조)

Site No.	DNA Cleavage Score	Base editing frequency (%)		
		Untreated	(+) ABE7.10	(+) Sniper ABE7.10
EphB4-01	10.3	0.13	52.42	54.66
EphB4-02	8.0	0.15	0.19	0.13
EphB4-03	4.1	0.16	0.16	0.16
EphB4-04	2.4	0.26	0.27	0.21
EphB4-05	2.2	0.12	0.14	0.12
EphB4-06	1.9	No PCR		
EphB4-07	1.8	0.13	0.14	0.14
EphB4-08	1.8	No PCR		
EphB4-09	1.4	0.27	0.24	0.26
EphB4-10	0.9	0.07	0.08	0.07
EphB4-11	0.8	0.14	0.13	0.14
EphB4-12	0.6	0.18	0.19	0.22
EphB4-13	0.5	0.18	0.19	0.17
EphB4-14	0.4	0.12	0.10	0.13
EphB4-15	0.4	0.02	0.02	0.03
EphB4-16	0.4	0.05	0.07	0.08

EphB4-17	0.4	0.09	0.08	0.08
EphB4-18	0.4	0.39	0.37	0.35
EphB4-19	0.3	0.06	0.07	0.10
EphB4-20	0.3	0.11	0.13	0.12
EphB4-21	0.3	0.00	0.16	0.15
EphB4-22	0.2	0.08	0.11	0.08
EphB4-23	0.2	0.06	0.10	0.07
EphB4-24	0.2	0.12	0.14	0.14
EphB4-25	0.2	0.30	0.38	0.35
EphB4-26	0.2	0.07	0.09	0.08
EphB4-27	0.2	No PCR		
EphB4-28	0.2	0.17	0.14	0.14
EphB4-29	0.2	0.22	0.22	0.21
EphB4-30	0.2	0.13	0.13	0.09
EphB4-31	0.1	0.24	0.20	0.22
EphB4-32	0.1	0.21	0.16	0.23
EphB4-33	0.1	0.17	0.18	0.12
EphB4-34	0.1	0.17	0.15	0.13
EphB4-35	0.1	2.32	2.36	2.33
EphB4-36	0.1	No PCR		
EphB4-37	0.1	0.13	0.12	0.12
EphB4-38	0.1	0.22	0.25	0.24
EphB4-39	0.1	0.19	0.22	0.22

[표 38]

Mutation frequencies of ABE7.10 and Sniper ABE7.10 at on-target and off-target sites captured by Digenome-seq (HPRT-Exon6; 'DNA sequence at a cleavage sites'은 표 17 참조)

Site No.	DNA Cleavage Score	Base editing frequency (%)		
		Untreated	(+) ABE7.10	(+) Sniper ABE7.10
HPRT_E6-01	29.0	0.73	58.63	63.33
HPRT_E6-02	21.3	0.14	0.17	0.15
HPRT_E6-03	17.0	0.22	0.34	0.28
HPRT_E6-04	12.7	0.34	0.36	0.35
HPRT_E6-05	10.5	0.12	0.12	0.12
HPRT_E6-06	10.2	0.26	0.29	0.23
HPRT_E6-07	9.3	0.12	1.45	0.19

HPRT_E6-08	9.0	0.12	0.11	0.10
HPRT_E6-09	6.1	0.12	0.12	0.15
HPRT_E6-10	5.6	0.09	0.11	0.11
HPRT_E6-11	4.7	0.26	0.32	0.24
HPRT_E6-12	4.5	0.18	0.18	0.16
HPRT_E6-13	4.3	0.12	1.19	0.23
HPRT_E6-14	4.3	0.14	0.12	0.16
HPRT_E6-15	4.1	0.04	0.03	0.04
HPRT_E6-16	3.8	0.07	0.19	0.07
HPRT_E6-17	3.7	0.12	0.12	0.12
HPRT_E6-18	3.6	0.25	0.26	0.20
HPRT_E6-19	3.6	0.07	0.07	0.08
HPRT_E6-20	3.5	0.14	0.10	0.13
HPRT_E6-21	3.5	0.07	0.08	0.07
HPRT_E6-22	3.4	0.12	0.15	0.12
HPRT_E6-23	3.3	0.08	0.08	0.07
HPRT_E6-24	3.1	0.11	0.12	0.12
HPRT_E6-25	2.9	0.09	0.08	0.09
HPRT_E6-26	2.8	0.09	0.07	0.07
HPRT_E6-27	2.6	0.46	0.45	0.45
HPRT_E6-28	2.5	0.13	0.13	0.14
HPRT_E6-29	2.4	0.28	0.27	0.29
HPRT_E6-30	2.4	0.27	0.30	0.30
HPRT_E6-31	2.3	0.15	0.18	0.12
HPRT_E6-32	2.2	0.23	0.24	0.20
HPRT_E6-33	1.5	0.22	0.24	0.27
HPRT_E6-34	1.4	0.00	0.00	0.00
HPRT_E6-35	1.4	0.10	0.09	0.09
HPRT_E6-36	1.2	0.09	0.11	0.08
HPRT_E6-37	1.1	0.02	0.06	0.09
HPRT_E6-38	1.1	0.21	0.22	0.21
HPRT_E6-39	1.1	0.04	0.04	0.02
HPRT_E6-40	1.0	0.23	0.23	0.28
HPRT_E6-41	1.0	0.06	0.13	0.16
HPRT_E6-52	0.7	0.00	0.00	0.00
HPRT_E6-81	0.4	0.13	0.18	0.15

[표 39]

Mutation frequencies of ABE7.10 and Sniper ABE7.10 at on-target

and off-target sites captured by Digenome-seq (HPRT-Exon8; 'DNA sequence at a cleavage sites'은 표 18 참조)

Site No.	DNA Cleavage Score	Base editing frequency (%)		
		Untreated	(+) ABE7.10	(+) Sniper ABE7.10
HPRT_E8-1	2.4	0.03	70.09	71.15
HPRT_E8-2	1.3	0.11	10.33	3.53
HPRT_E8-3	0.6	0.17	0.17	0.19
HPRT_E8-4	0.6	0.05	0.03	0.06
HPRT_E8-5	0.6	0.08	0.08	0.10
HPRT_E8-6	0.5	0.04	0.10	0.09
HPRT_E8-7	0.5	0.16	0.16	0.14
HPRT_E8-8	0.4	No PCR		
HPRT_E8-9	0.4	0.07	0.06	0.04
HPRT_E8-10	0.4	0.11	0.11	0.12
HPRT_E8-11	0.4	0.05	0.07	0.05
HPRT_E8-12	0.3	0.14	0.15	0.13
HPRT_E8-13	0.3	0.05	0.05	0.03
HPRT_E8-14	0.3	0.26	0.25	0.25
HPRT_E8-15	0.3	0.04	0.05	0.06
HPRT_E8-16	0.2	0.11	0.11	0.10
HPRT_E8-17	0.2	0.09	0.09	0.08
HPRT_E8-18	0.2	0.04	0.15	0.05
HPRT_E8-19	0.1	0.05	0.04	0.07
HPRT_E8-20	0.1	0.08	0.05	0.05

또한, ABE7.10와 sgRNA가 결합된 ABE7.10 RNP 형태로 세포에 도입(참고예 1 참조)된 경우의 염기 교정 빈도를 측정하여, ABE7.10와 sgRNA가 플라스미드를 통하여 세포에 도입(참조예 1 참조)된 경우와 비교하여, 도 13g 및 13h (도 13h에서, y축의 그래프는 위에서부터 순서대로 (+)ABE7.10_RNP, (+)ABE7.10_플라스미드, 및 untreated의 결과를 보여준다)에 나타내었다.

도 13e 내지 13h 및 표 26 내지 32에 나타난 바와 같이, Sniper ABE7.10의 사용하는 경우 및 ABE7.10 RNP를 사용하는 경우, 각각 ABE7.10의 사용하는 경우 및 ABE7.10 플라스미드를 사용하는 경우와 비교하여 염기 교정 특이성을 약 7 배까지 향상시킬 수 있다.

또한, Sniper ABE7.10를 변형 sgRNAs와 조합하여 사용함으로써, ABE7.10의 off-target 효과를 보다 감소시킬 수 있음을 확인하였다 (도 13i 및 13j 참조; 도 13i 및 13j에서, y축의 그래프는 위에서부터 순서대로 ggX20-Sniper ABE7.10, gX20-Sniper ABE7.10, GX19-Sniper ABE7.10, gX18-Sniper ABE7.10, gX17-Sniper ABE7.10, GX19-ABE7.10, 및 untreated의 결과를 보여준다).

본 실시예를 통하여, mismatched sgRNAs, Digenome-seq, 및 targeted amplicon sequencing을 사용하여, ABE7.10, BE3, 및 Cas9가 인간 유전체에서 서로 다른 세트의 off-target sites을 인식할 수 있고, 각각 독립적으로 평가 수단을 필요로 함을 확인하였다. 상기 시험된 7 개 유전자를 표적으로 하는 ABE7.10 디아미나아제는 인간 유전체의 제한된 수의 부위에서 매우 특이적으로 *in vitro* A-to-G conversion을 촉매함을 확인하였다. 또한, 변형 sgRNAs, ABE7.10 RNPs, 및/또는 Sniper ABE7.10를 사용함으로써, ABE7.10의 off-target 활성을 감소시키거나 제거할 수 있음을 확인하였다.

【청구범위】

【청구항 1】

(1) 아데노신 디아미나아제, 이를 암호화 하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 플라스미드,

5 (2) 불활성화된 표적 특이적 뉴클레아제, 이를 암호화하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 플라스미드,

(3) 가이드 RNA, 및

(4) Alkyladenine DNA Glycosylase (AAG)와 엔도뉴클레아제 VIII (endonuclease VIII)의 조합 또는 엔도뉴클레아제 V (endonuclease V)

10 을 포함하고,

상기 불활성화된 표적 특이적 뉴클레아제는 DNA 이중 가닥을 모두 절단하는 엔도뉴클레아제 활성을 상실한 것인,

아데닌 디아미나아제를 사용하는 DNA 이중 가닥 절단용 조성물.

【청구항 2】

15 제1항에 있어서, 상기 불활성화된 표적특이적 뉴클레아제는 DNA 이중 가닥을 절단하는 엔도뉴클레아제 활성을 상실한 Cas9 단백질 또는 Cpf1 단백질인, 아데닌 디아미나아제를 사용하는 DNA 이중 가닥 절단용 조성물.

【청구항 3】

20 제2항에 있어서, 상기 불활성화된 표적특이적 뉴클레아제는 스트렙토코커스 피요젠스 (*Streptococcus pyogenes*) 유래의 Cas9 단백질에 아미노산 잔기 D10, H840, F539, M763, 및 K890로 이루어진 군에서 선택된 하나 이상이 원래와 다른 아미노산으로 치환된 돌연변이가 도입된 것인, 아데닌 디아미나아제를 사용하는 DNA 이중 가닥 절단용 조성물.

【청구항 4】

25 제1항에 있어서,

상기 아데닌 디아미나아제, 불활성화된 표적특이적 뉴클레아제, 및 가이드 RNA는

(i) 아데노신 디아미나아제, 표적 특이적 뉴클레아제, 및 가이드 RNA의 혼합물, (ii) 아데노신 디아미나아제 암호화 핵산 서열 또는 이를
30 포함하는 플라스미드, 표적 특이적 뉴클레아제 암호화 핵산 서열 또는 이를 포함하는 플라스미드, 및 가이드 RNA 또는 이의 암호화 핵산 분자를 포함하는 플라스미드의 혼합물, (iii) 아데노신 디아미나아제 및 표적 특이적 뉴클레아제가 연결된 융합 단백질 및 가이드 RNA의 혼합물, (iv)

아데노신 디아미나아제 및 표적 특이적 뉴클레아제가 연결된 융합 단백질
 암호화 핵산 분자 또는 상기 핵산 분자를 포함하는 플라스미드, 및 가이드
 RNA 또는 이의 암호화 핵산 분자를 포함하는 플라스미드의 혼합물, 및 (v)
 아데노신 디아미나아제 및 표적 특이적 뉴클레아제가 연결된 융합 단백질
 5 및 가이드 RNA의 복합체 또는 혼합물

로 이루어진 군에서 선택된 하나 이상의 형태인, 아데닌
 디아미나아제를 사용하는 DNA 이중 가닥 절단용 조성물.

【청구항 5】

제1항에 있어서,
 10 상기 불활성화된 표적특이적 뉴클레아제는, 스트렙토코커스 피요젠스
 (*Streptococcus pyogenes*) 유래의 Cas9 단백질에,
 아미노산 잔기 D10가 원래 아미노산과 다른 아미노산으로 치환된
 돌연변이가 도입되거나,
 아미노산 잔기 D10 및 H840이 모두 원래 아미노산과 다른
 15 아미노산으로 치환된 돌연변이가 도입되거나
 아미노산 잔기 F539, M763, 및 K890이 모두 원래 아미노산과 다른
 아미노산으로 치환된 돌연변이가 도입된 것인,
 아데닌 디아미나아제를 사용하는 DNA 이중 가닥 절단용 조성물.

【청구항 6】

제1항 내지 제5항 중 어느 한 항에 있어서, 상기 가이드 RNA는
 crRNA와 tracrRNA가 서로 결합된 이중 가닥 crRNA:tracrRNA 복합체, 또는
 단일 가닥 가이드 RNA (sgRNA)인, 아데닌 디아미나아제를 사용하는 DNA
 이중 가닥 절단용 조성물.

【청구항 7】

(a) (1) 아데노신 디아미나아제, 아데노신 디아미나아제를
 암호화하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 플라스미드, 및 (2)
 불활성화된 표적 특이적 뉴클레아제, 불활성화된 표적 특이적 뉴클레아제를
 암호화하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 플라스미드를
 분리된 세포에 도입시키거나 또는 분리된 DNA에 처리하여, DNA 이중가닥 중
 30 하나를 절단하는 단계; 및

(b) 상기 DNA 이중가닥 중 하나가 절단된 DNA에 Alkyladenine DNA
 Glycosylase (AAG)와 엔도뉴클레아제 VIII (endonuclease VIII)의 조합,
 또는 엔도뉴클레아제 V (endonuclease V)을 처리하여, 나머지 DNA 가닥을

절단하는 단계

를 포함하는, 아데닌 디아미나아제를 사용하여 DNA에 이중 가닥 절단 (double strand break)를 생성하는 방법.

【청구항 8】

- 5 (a) (1) 아데노신 디아미나아제, 아데노신 디아미나아제를 암호화하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 플라스미드, 및 (2) 불활성화된 표적 특이적 뉴클레아제, 불활성화된 표적 특이적 뉴클레아제를 암호화하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 플라스미드를 분리된 세포에 도입시키거나 또는 분리된 DNA에 처리하여, DNA 이중가닥 중
10 하나를 절단하는 단계;

(b) 상기 DNA 이중가닥 중 하나가 절단된 DNA에 Alkyladenine DNA Glycosylase (AAG)와 엔도뉴클레아제 VIII (endonuclease VIII)의 조합, 또는 엔도뉴클레아제 V (endonuclease V)을 처리하여, 나머지 DNA 가닥을 절단하는 단계; 및

- 15 (c) 상기 이중가닥 절단된 DNA에 대하여 전체 유전체 시퀀싱 (whole genome sequencing)을 수행하는 단계

를 포함하는, 아데닌 디아미나아제에 의하여 염기 교정(base editing)이 도입된 DNA의 핵산 서열 분석 방법.

【청구항 9】

- 20 (a) (1) 아데노신 디아미나아제, 아데노신 디아미나아제를 암호화하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 플라스미드, 및 (2) 불활성화된 표적 특이적 뉴클레아제, 불활성화된 표적 특이적 뉴클레아제를 암호화하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 플라스미드를 분리된 세포에 도입시키거나 또는 분리된 DNA에 처리하여, DNA 이중가닥 중
25 하나를 절단하는 단계;

(b) 상기 DNA 이중가닥 중 하나가 절단된 DNA에 Alkyladenine DNA Glycosylase (AAG)와 엔도뉴클레아제 VIII (endonuclease VIII)의 조합, 또는 엔도뉴클레아제 V (endonuclease V)을 처리하여, 나머지 DNA 가닥을 절단하는 단계;

- 30 (c) 상기 이중가닥 절단된 DNA에 대하여 전체 유전체 시퀀싱 (whole genome sequencing)을 수행하는 단계; 및

(d) 상기 시퀀싱으로 획득한 염기서열 데이터 (sequence read)에서 상기 절단된 위치를 확인하는 단계

를 포함하는, 아데닌 디아미나아제의 염기 교정 위치 확인 방법.

【청구항 10】

(a) (1) 아데노신 디아미나아제, 아데노신 디아미나아제를
암호화하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 플라스미드, 및 (2)
5 불활성화된 표적 특이적 뉴클레아제, 불활성화된 표적 특이적 뉴클레아제를
암호화하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 플라스미드를
분리된 세포에 도입시키거나 또는 분리된 DNA에 처리하여, DNA 이중가닥 중
하나를 절단하는 단계;

(b) 상기 DNA 이중가닥 중 하나가 절단된 DNA에 Alkyladenine DNA
10 Glycosylase (AAG)와 엔도뉴클레아제 VIII (endonuclease VIII)의 조합,
또는 엔도뉴클레아제 V (endonuclease V)을 처리하여, 나머지 DNA 가닥을
절단하는 단계;

(c) 상기 이중가닥 절단된 DNA에 대하여 전체 유전체 시퀀싱 (whole
genome sequencing)을 수행하는 단계; 및

15 (d) 상기 시퀀싱으로 획득한 염기서열 데이터 (sequence read)에서
상기 절단된 위치를 확인하는 단계

를 포함하는, 아데닌 디아미나아제의 비표적 위치 (off-target site)
확인 방법.

【청구항 11】

20 제7항 내지 제10항 중 어느 한 항에 있어서, 상기 불활성화된
표적특이적 뉴클레아제는 DNA 이중 가닥을 절단하는 엔도뉴클레아제 활성을
상실한 Cas9 단백질 또는 Cpf1 단백질인, 방법.

【청구항 12】

25 제7항 내지 제10항 중 어느 한 항에 있어서, 상기 불활성화된
표적특이적 뉴클레아제는 스트렙토코커스 피요젠스 (*Streptococcus
pyogenes*) 유래의 Cas9 단백질에 아미노산 잔기 D10, H840, F539, M763, 및
K890로 이루어진 군에서 선택된 하나 이상이 원래와 다른 아미노산으로
치환된 돌연변이가 도입된 것인, 방법.

【청구항 13】

30 제7항 내지 제10항 중 어느 한 항에 있어서,
상기 아데닌 디아미나아제, 불활성화된 표적특이적 뉴클레아제, 및
가이드 RNA는

(i) 아데노신 디아미나아제, 표적 특이적 뉴클레아제, 및 가이드

- RNA의 혼합물, (ii) 아데노신 디아미나아제 암호화 핵산 서열 또는 이를 포함하는 플라스미드, 표적 특이적 뉴클레아제 암호화 핵산 서열 또는 이를 포함하는 플라스미드, 및 가이드 RNA 또는 이의 암호화 핵산 분자를 포함하는 플라스미드의 혼합물, (iii) 아데노신 디아미나아제 및 표적 특이적 뉴클레아제가 연결된 융합 단백질 및 가이드 RNA의 혼합물, (iv) 아데노신 디아미나아제 및 표적 특이적 뉴클레아제가 연결된 융합 단백질 암호화 핵산 분자 또는 상기 핵산 분자를 포함하는 플라스미드, 및 가이드 RNA 또는 이의 암호화 핵산 분자를 포함하는 플라스미드의 혼합물, 및 (v) 아데노신 디아미나아제 및 표적 특이적 뉴클레아제가 연결된 융합 단백질 및 가이드 RNA의 복합체 또는 혼합물

로 이루어진 군에서 선택된 하나 이상의 형태로 사용되는 것인, 방법.

【청구항 14】

- 제7항 내지 제10항 중 어느 한 항에 있어서,
상기 불활성화된 표적특이적 뉴클레아제는, 스트렙토코커스 피요젠스 (*Streptococcus pyogenes*) 유래의 Cas9 단백질에,
아미노산 잔기 D10가 원래 아미노산과 다른 아미노산으로 치환된 돌연변이가 도입되거나,
아미노산 잔기 D10 및 H840이 모두 원래 아미노산과 다른 아미노산으로 치환된 돌연변이가 도입되거나
아미노산 잔기 F539, M763, 및 K890이 모두 원래 아미노산과 다른 아미노산으로 치환된 돌연변이가 도입된 것인,
방법.

【청구항 15】

- 제7항 내지 제10항 중 어느 한 항에 있어서, 상기 가이드 RNA는 crRNA와 tracrRNA가 서로 결합된 이중 가닥 crRNA:tracrRNA 복합체, 또는 단일 가닥 가이드 RNA (sgRNA)인, 방법.

【청구항 16】

- 제7항 내지 제10항 중 어느 한 항에 있어서, 시험관 내 (in vitro)에서 수행되는 것인, 방법.

【청구항 17】

- 제7항 내지 제10항 중 어느 한 항에 있어서, 상기 단계 (a)의 분리된 DNA는 유전체 DNA인, 방법.

【청구항 18】

제10항에 있어서, 상기 단계 (d) 이후에,
상기 절단 위치가 표적 위치 (on-target site)가 아닌 경우, 비표적 위치 (off-target site)로 판단하는 단계를 추가로 포함하는, 방법.

【청구항 19】

- 5 제10항에 있어서,
상기 단계 (d)에서 확인된 절단 위치는 수득한 염기서열 데이터를 정렬하여 5' 말단이 수직 정렬된 위치, 또는 5' 말단 플롯에서 이중 피크 패턴을 보이는 위치인 것인, 방법.

【청구항 20】

- 10 제19항에 있어서, 상기 정렬은 표준 염기서열 (reference genome)로 염기서열 데이터를 맵핑한 뒤, BWA/GATK 또는 ISAAC을 이용하여 수행되는 것인, 방법.

【청구항 21】

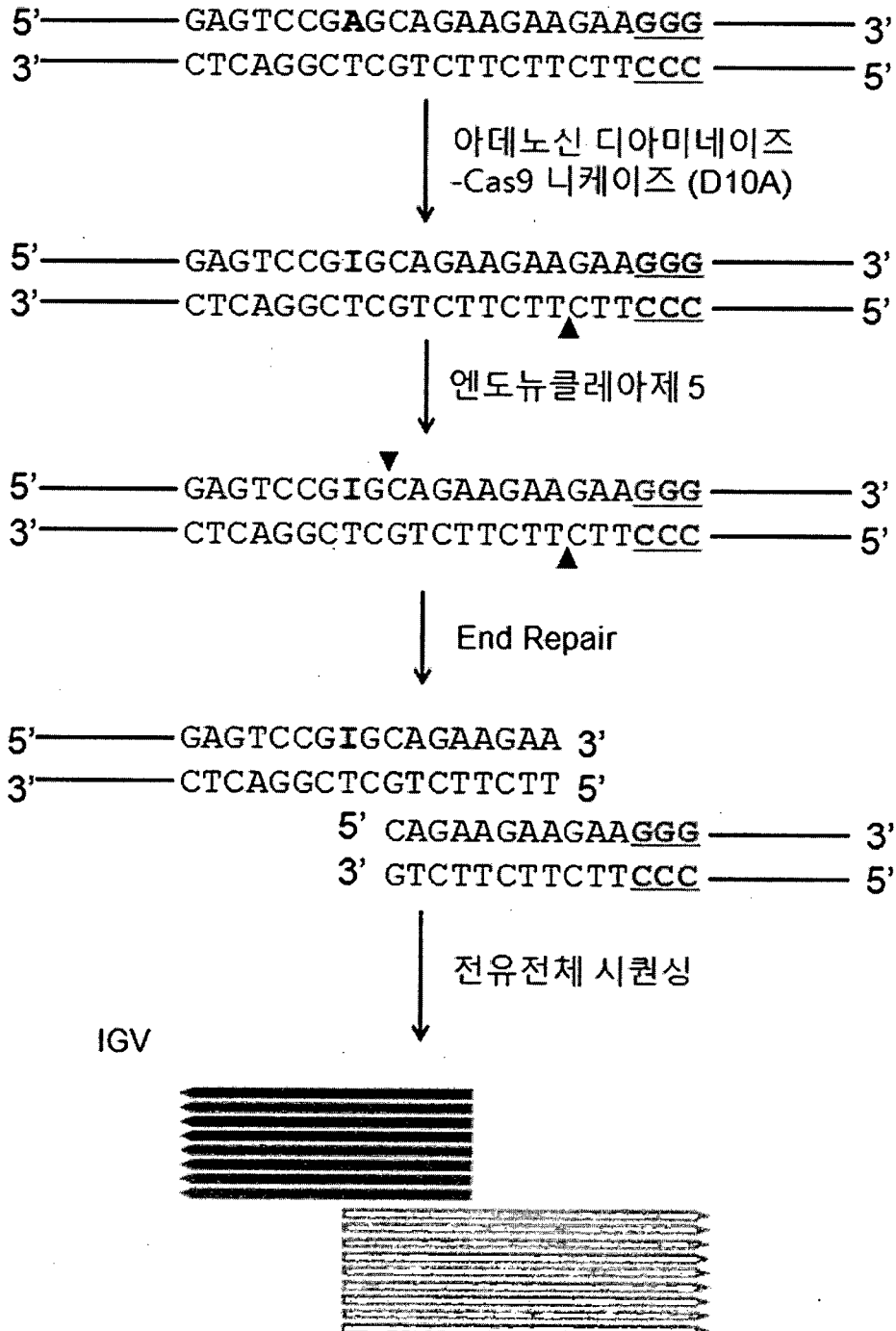
- 15 제19항에 있어서, 왓슨 가닥 (Watson strand)과 크릭 가닥 (Crick strand)에 해당하는 염기서열 데이터 (sequence read)가 각각 두 개 이상씩 수직으로 정렬되는 위치를 비표적 위치인 것으로 판단하는 단계를 추가로 포함하는, 방법.

【청구항 22】

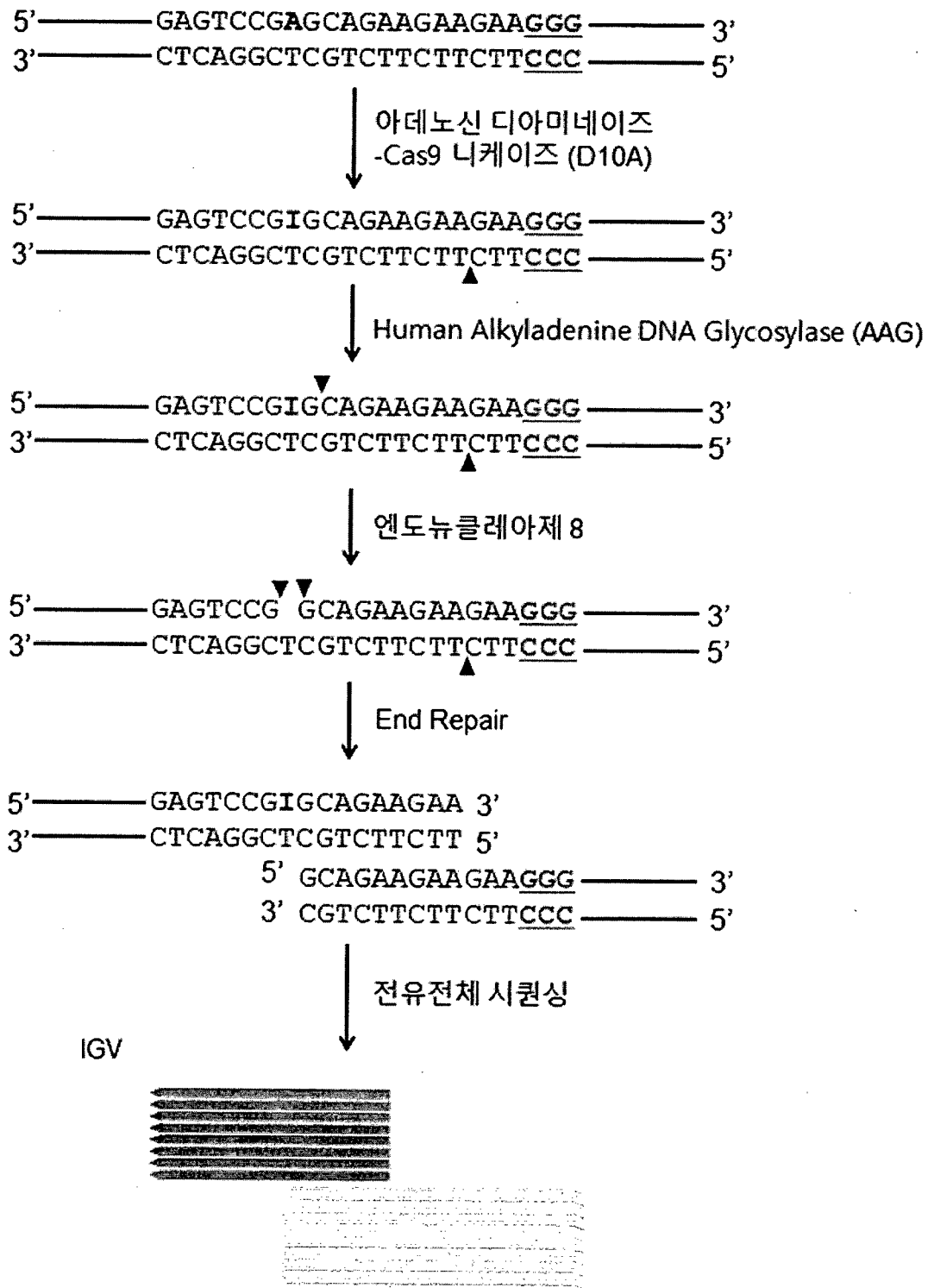
- 20 제19항에 있어서, 20 % 이상의 염기서열 데이터가 수직으로 정렬되고, 각각의 왓슨 가닥 및 크릭 가닥에서 동일한 5' 말단을 가진 염기서열 데이터의 수가 10 이상인 위치가 비표적 위치인 것으로 판단하는 단계를 추가로 포함하는, 방법.

【도면】

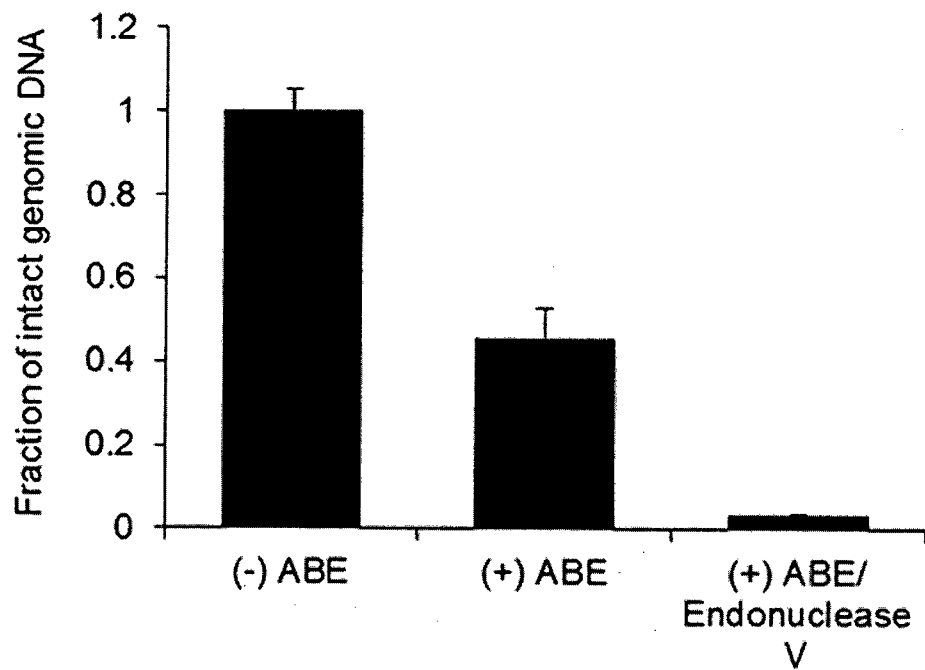
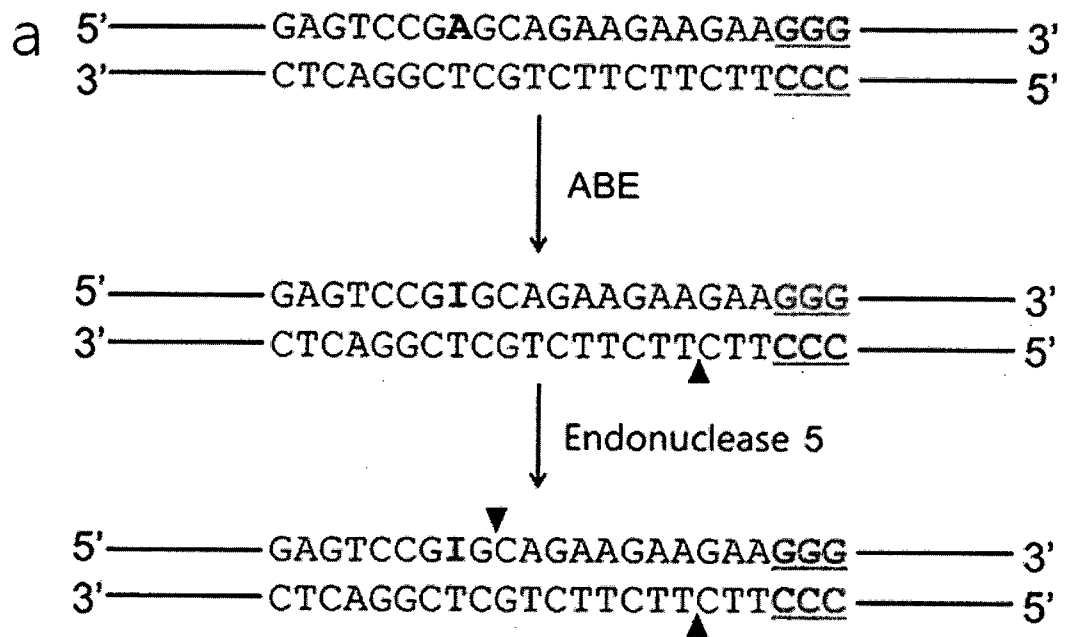
【도 1a】



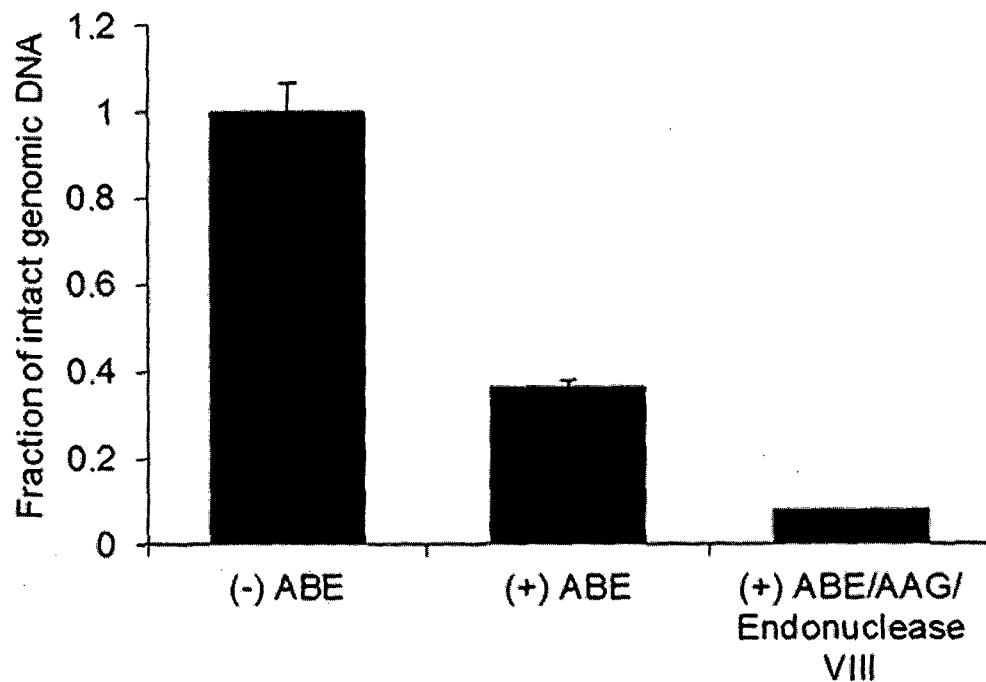
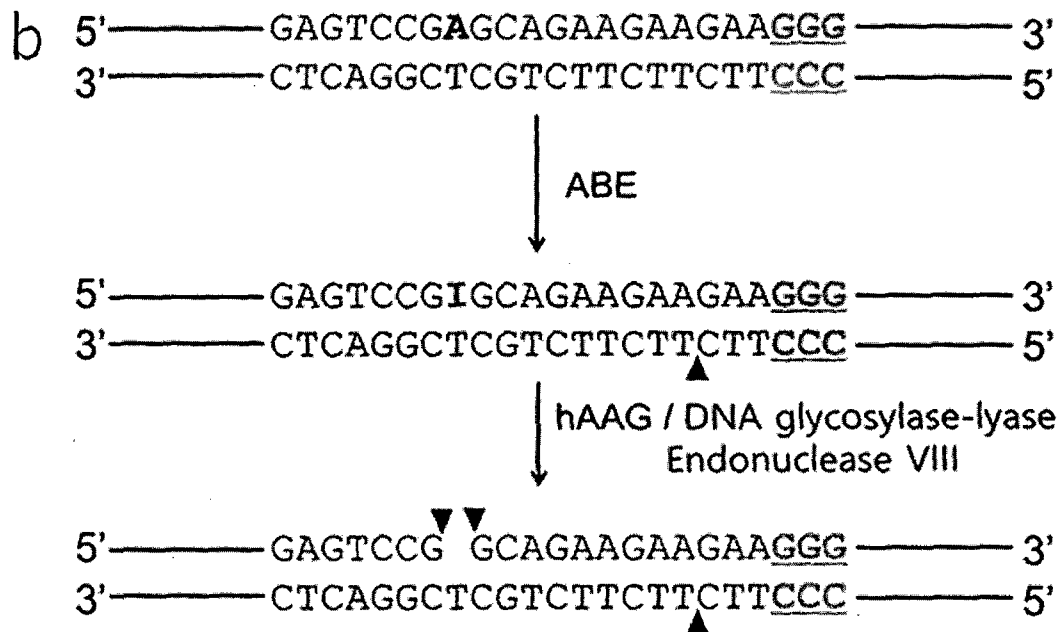
【도 1b】



【도 2a】

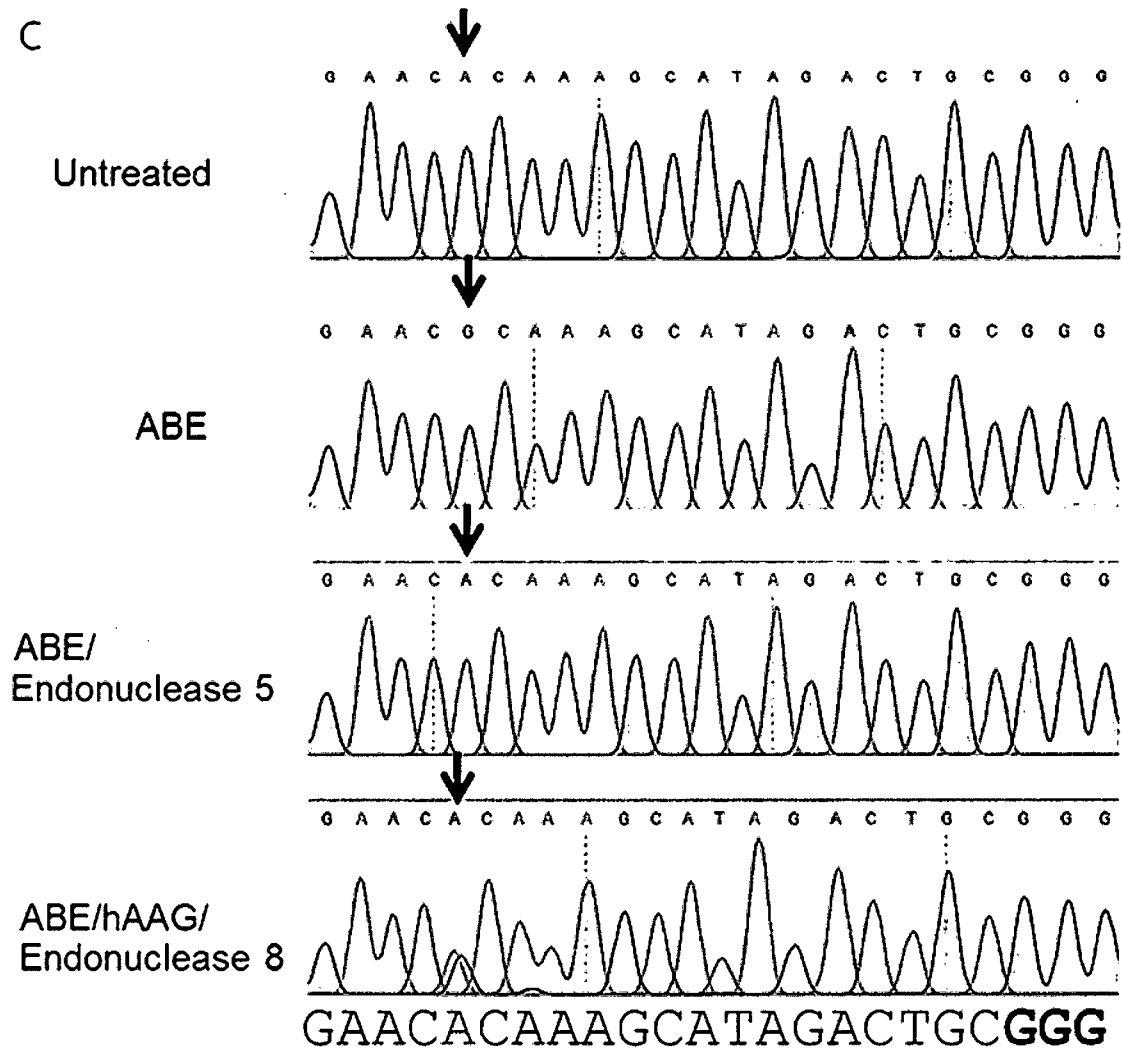


【도 2b】



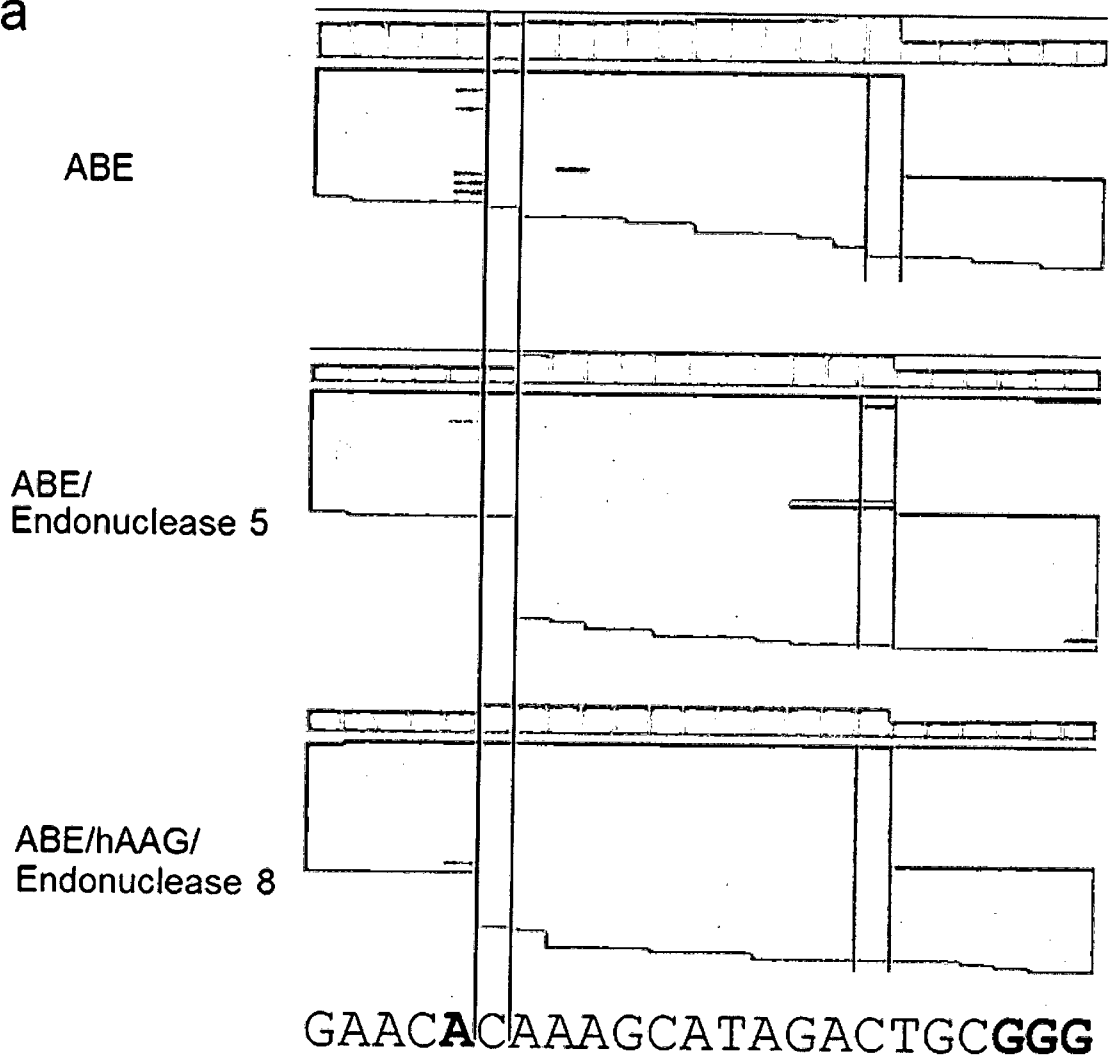
【도 2c】

C

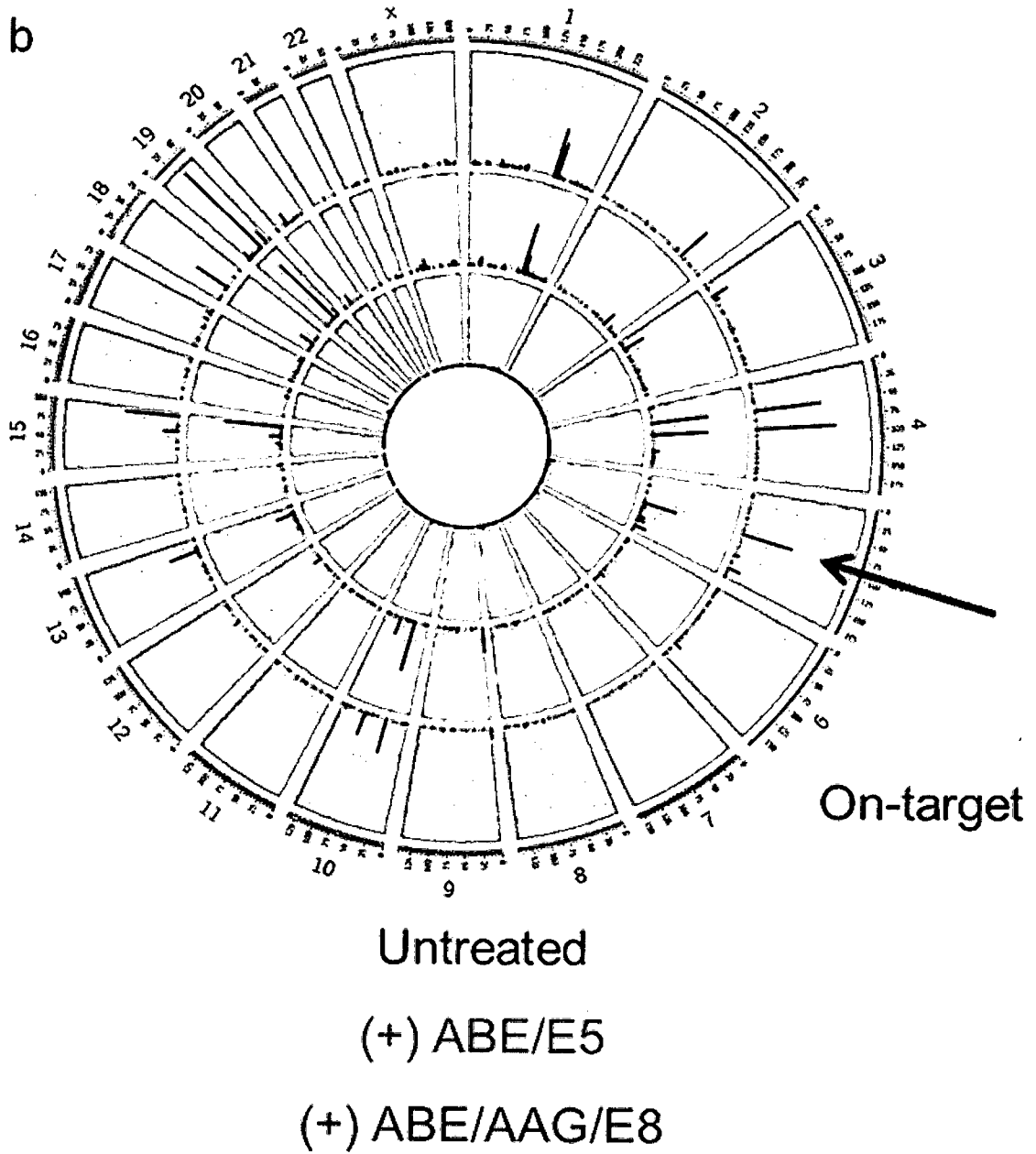


【도 3a】

a



【도 3b】



【도 4】

a ABE/
Endonuclease 5

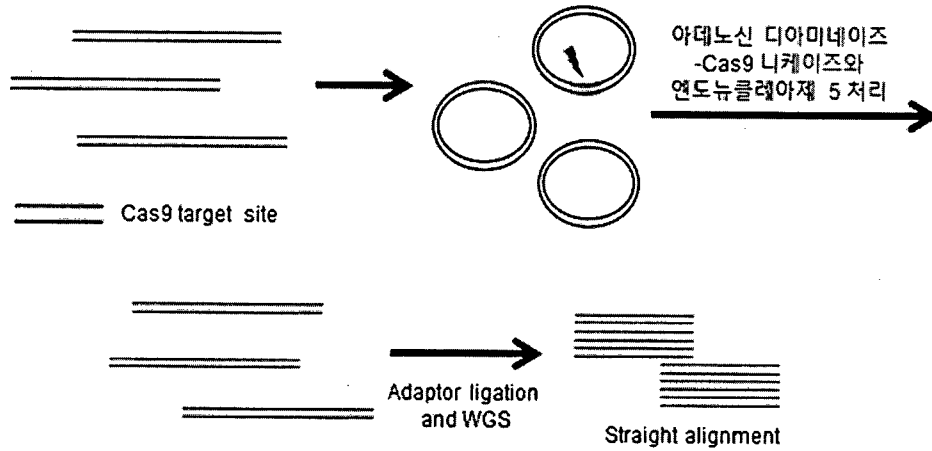
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites
HEK2_ABE_E5_001	chr19	35505485	20.1	GAACAC-AAGCacAGACTGaAGG
HEK2_ABE_E5_002	chr15	93557679	15.1	GAACACA-tGCATAGACTGCTAG
HEK2_ABE_E5_003	chr4	53536209	14.7	GAAtActAAGCATAGACTcCAGG
HEK2_ABE_E5_004	chr4	90522183	14.3	GAACACAAtGCATAGAtTGCCGG
HEK2_ABE_E5_005	chr1	167742859	13.4	aAACACAgAGCacAGACTGCTGA
HEK2_ABE_E5_006	chr10	43159340	13.4	aAACACAAAGaCATAGACCaCTGG
HEK2_ABE_E5_007	chr5	87240613	8.9	GAACACAAAGCATAGACTGCGGG
HEK2_ABE_E5_008	chr9	290167	6.6	aAACAtAAAGaATAGACTGCAAG
HEK2_ABE_E5_009	chr3	28560566	5.8	GAACACAAAGtATAGaATGCTAG
HEK2_ABE_E5_010	chr1	175310138	5.1	GgACACAAAGCtTAGACTcCAGG
HEK2_ABE_E5_011	chr13	87616351	4.9	GAAGAtAAAGCATAGACTctAGG
HEK2_ABE_E5_012	chr10	87125365	4.6	aAACAtAAAGCATAGACTGCAAG
HEK2_ABE_E5_013	chr18	32279782	3.9	aAAtACAAAGCATAGACTaAtATG
HEK2_ABE_E5_014	chr2	192248363	3.5	GAACACAtA-CATAGACaGCTGG
HEK2_ABE_E5_015	chr5	160202385	3.0	GAAaACAAAGCAaAGGaaGCAGG
HEK2_ABE_E5_016	chr15	63236612	2.9	aAcacaAtAGCATAGACTGgaCTG
HEK2_ABE_E5_017	chr20	23101377	2.7	atACACAgAGCAaAGACTGCAGG
HEK2_ABE_E5_018	chrX	48317375	2.5	aAACAtAAAGCctTAGACTGaCGG

b ABE/hAAG
/Endonuclease 8

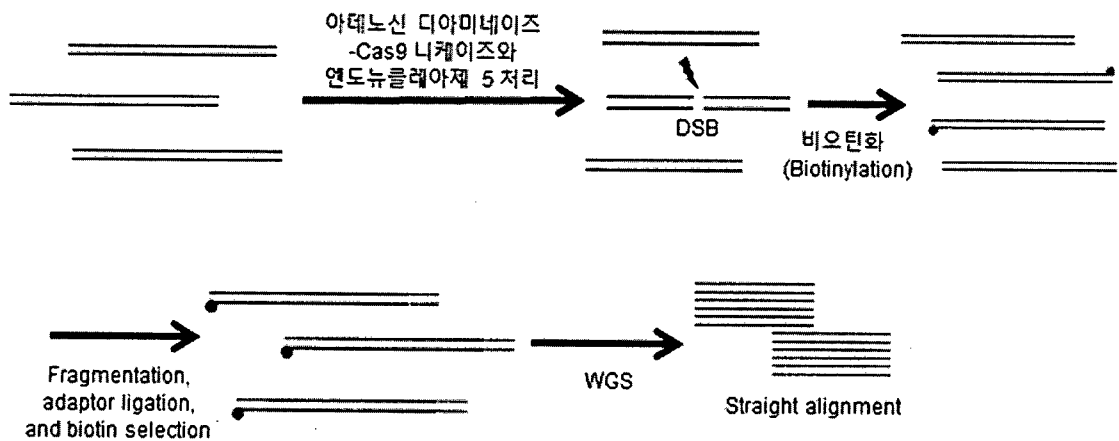
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites
HEK2_ABE_E8_001	chr19	35505485	23.1	GAACAC-AAGCacAGACTGaAGG
HEK2_ABE_E8_002	chr4	90522183	17.0	GAACACAAtGCATAGAtTGCCGG
HEK2_ABE_E8_003	chr4	53536209	14.2	GAAtActAAGCATAGACTcCAAG
HEK2_ABE_E8_004	chr15	93557679	11.6	GAACACA-tGCATAGACTGCTAG
HEK2_ABE_E8_005	chr5	87240613	11.2	GAACACAAAGCATAGACTGCGGG
HEK2_ABE_E8_006	chr1	167742861	10.8	aAACACAgAGCacAGACTGCTGA
HEK2_ABE_E8_007	chr18	32279782	8.0	aAAtACAAAGCATAGACTaAtATG
HEK2_ABE_E8_008	chr1	175310137	7.8	GgACACAAAGCtTAGACTcCAGG
HEK2_ABE_E8_009	chr10	43159340	7.7	aAACACAAAGaCATAGACCaCTGG
HEK2_ABE_E8_010	chr2	192248363	7.4	GAACACAtA-CATAGACaGCTGG
HEK2_ABE_E8_011	chr13	87616351	6.3	GAAGAtAAAGCATAGACTctAGG
HEK2_ABE_E8_012	chr10	87125365	5.2	aAACAtAAAGCATAGACTGCAAG
HEK2_ABE_E8_013	chr3	28560566	3.7	GAACACAAAGtATAGaATGCTAG
HEK2_ABE_E8_014	chr19	49972515	3.2	aAAaACAAAGCgTAGACTGtGGG
HEK2_ABE_E8_015	chr15	63236610	3.1	aAcacaAtAGCATAGACTGgaCTG
HEK2_ABE_E8_016	chr20	23101377	2.7	atACACAgAGCAaAGACTGCAGG

【도 5】

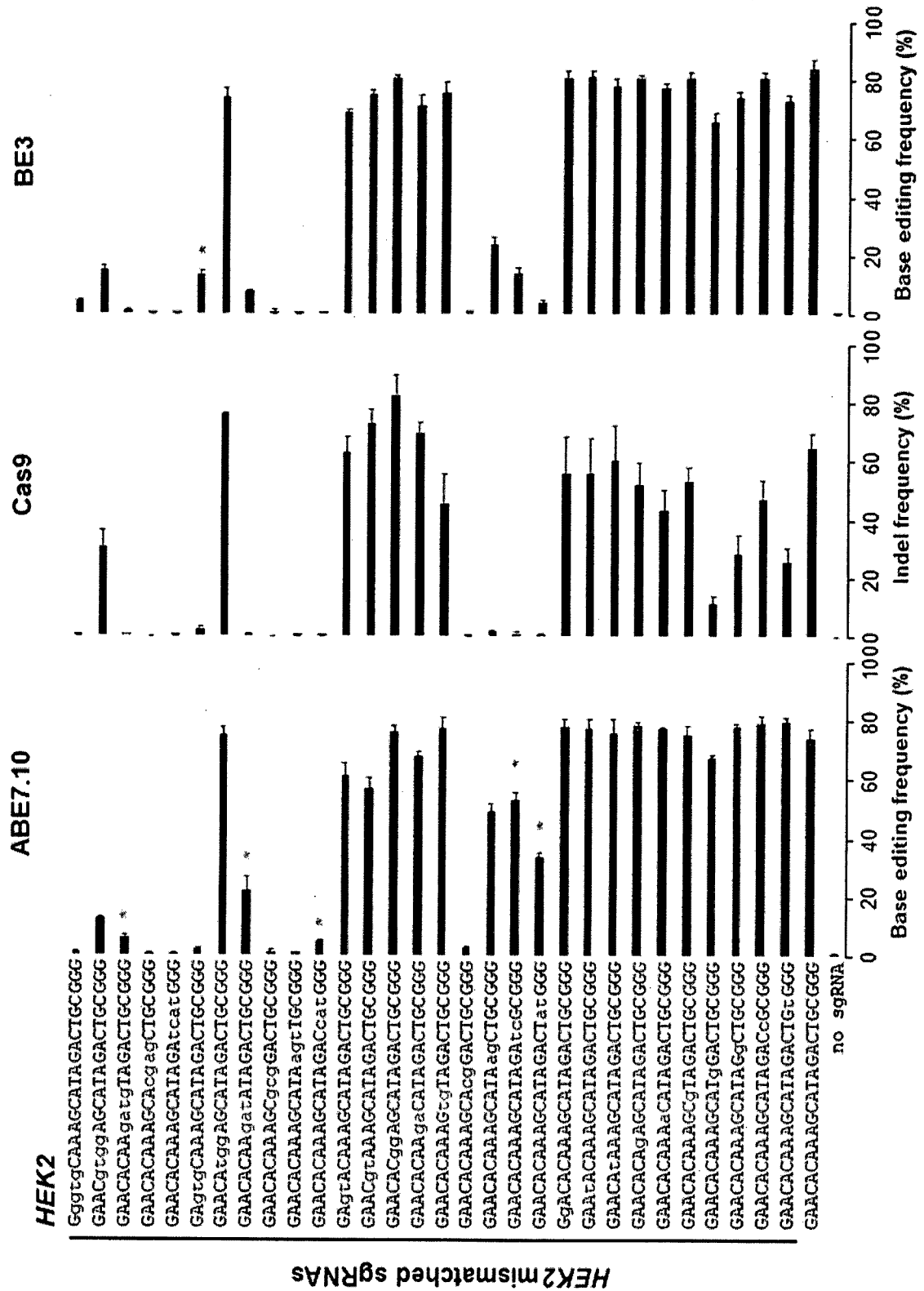
CIRCLE-seq



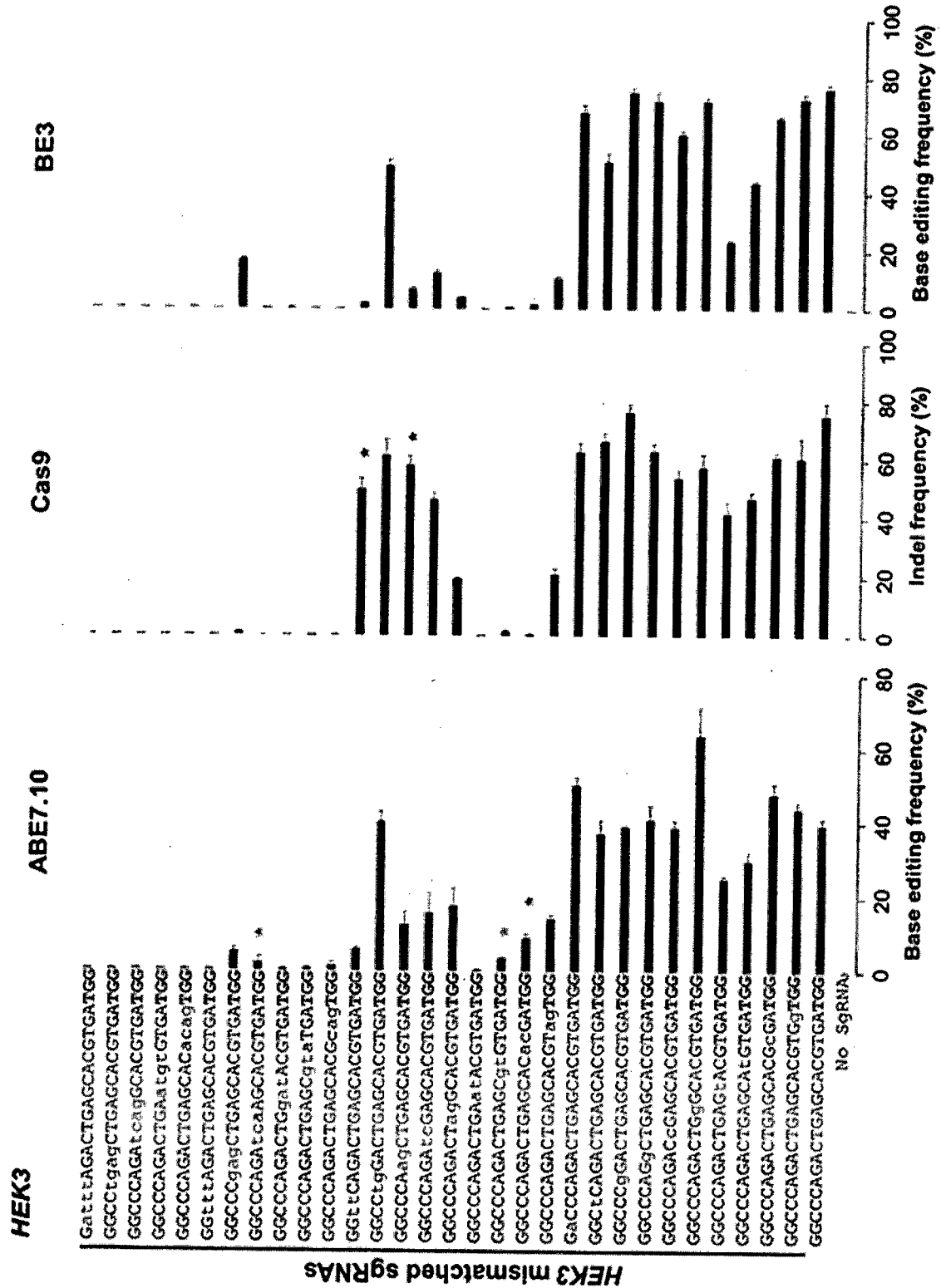
SITE-seq



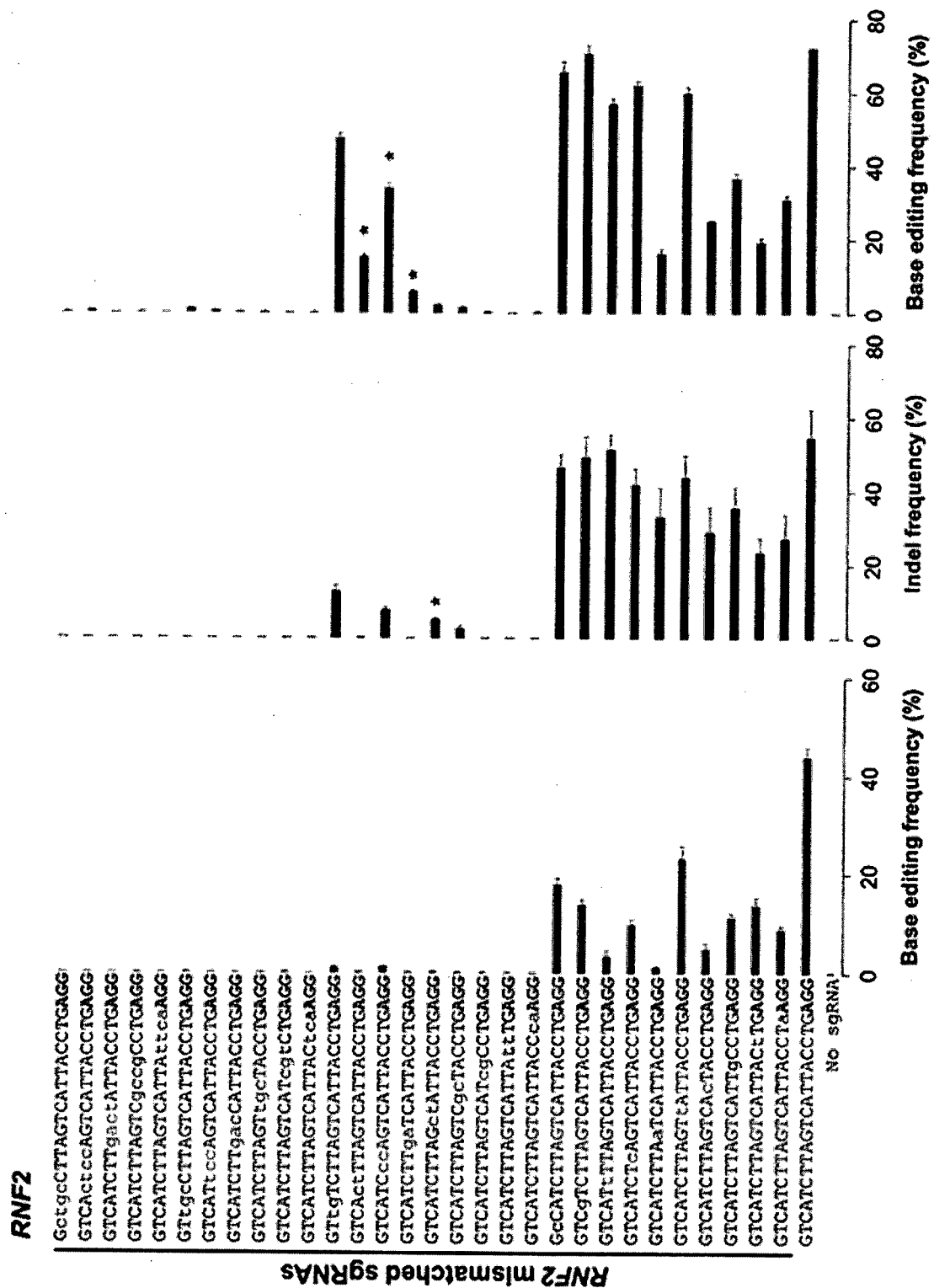
【도 6a】



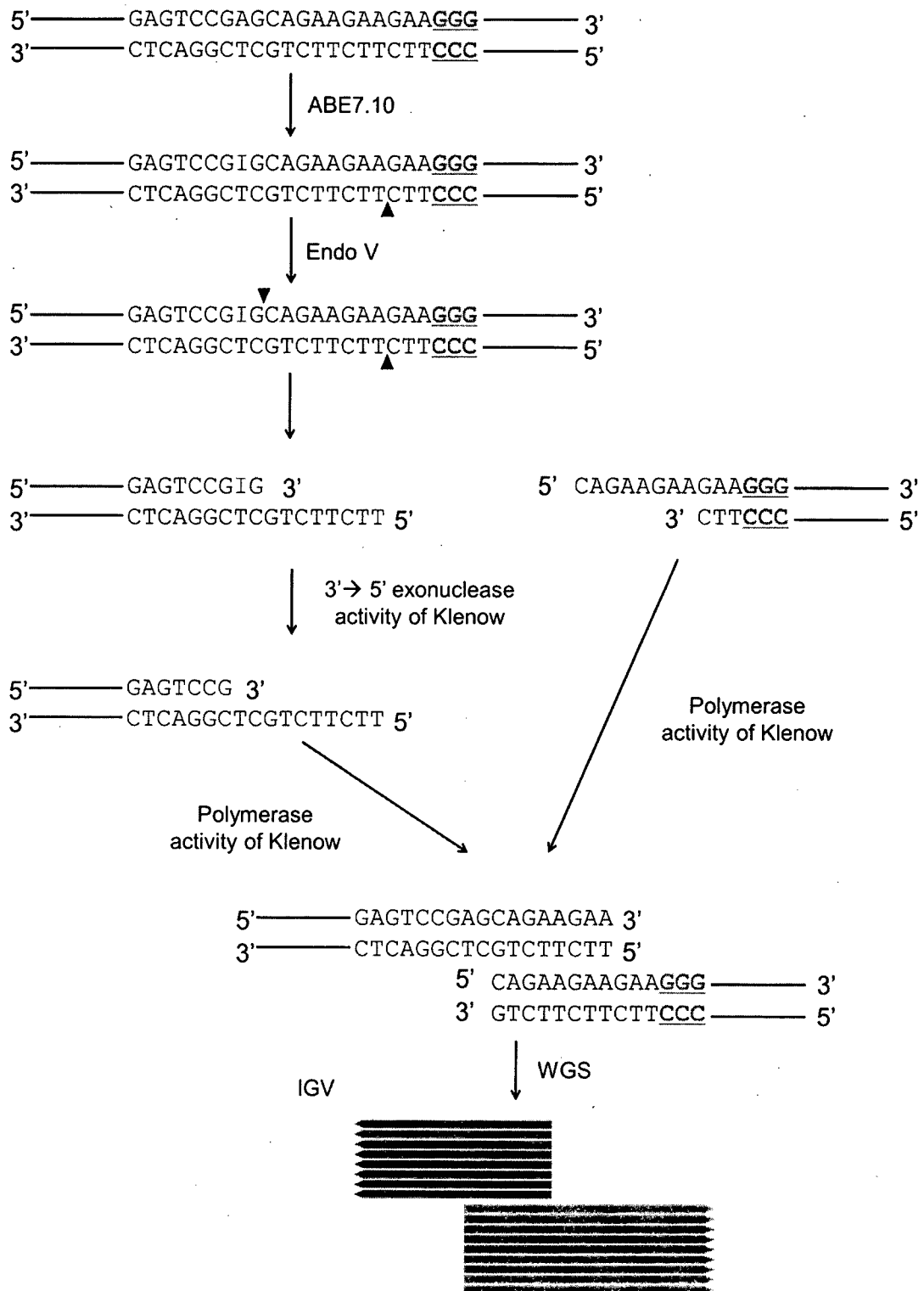
【도 6b】



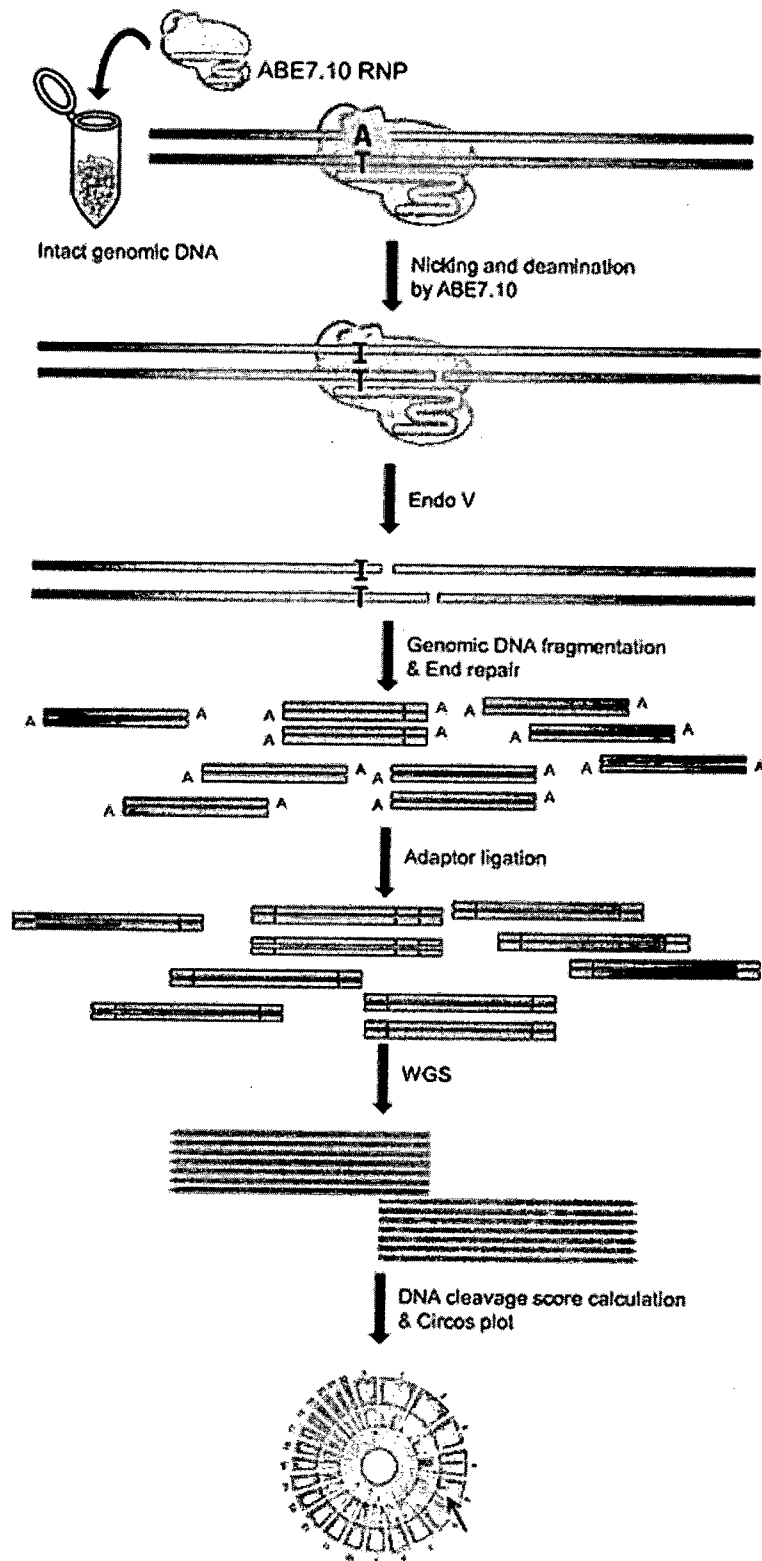
【도 6c】



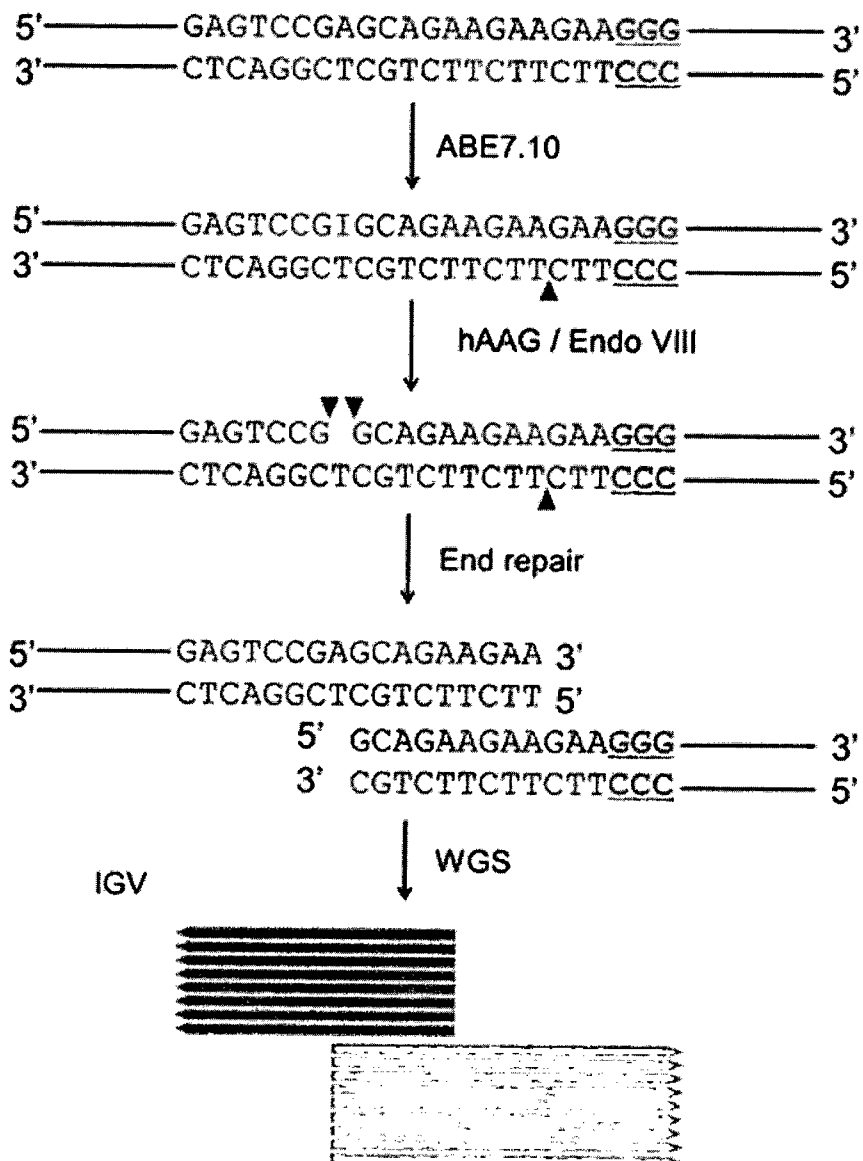
【도 7a】



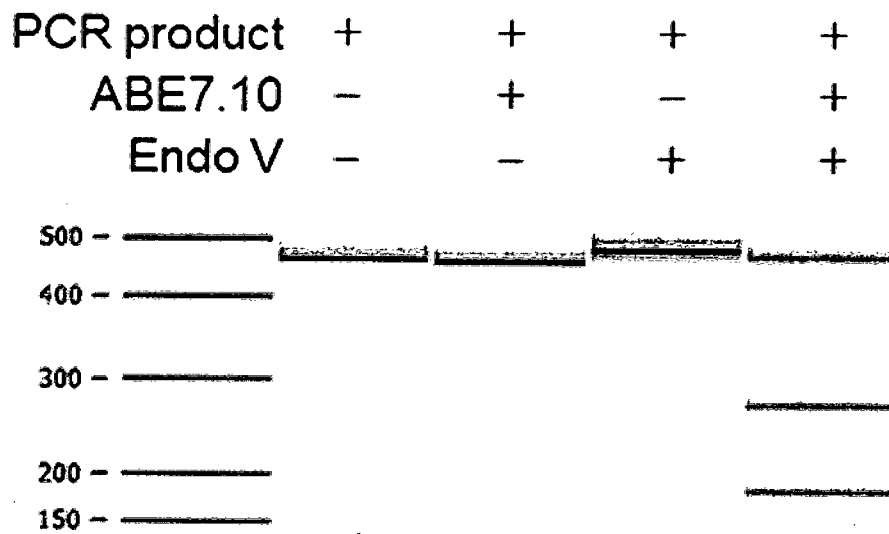
【도 7b】



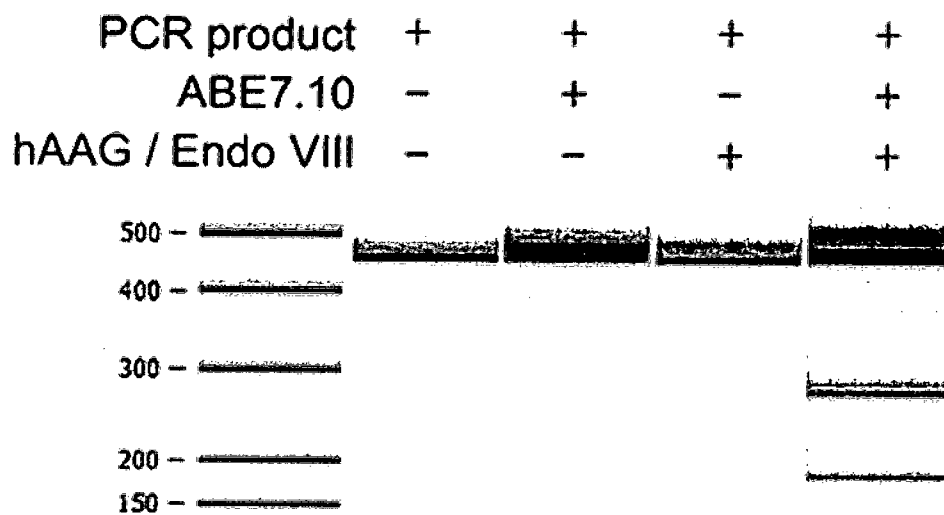
【도 7c】



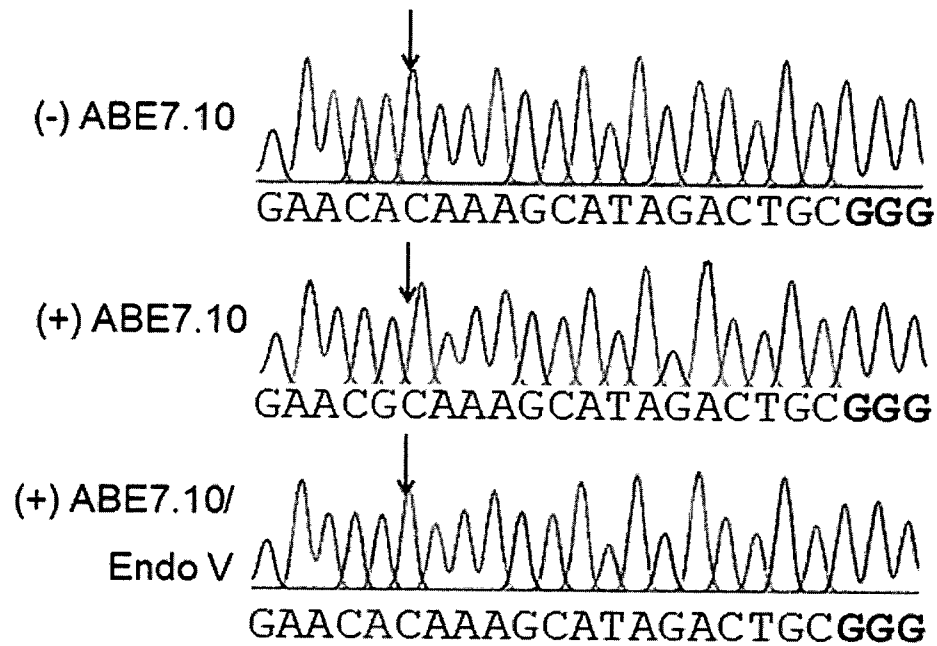
【도 7d】



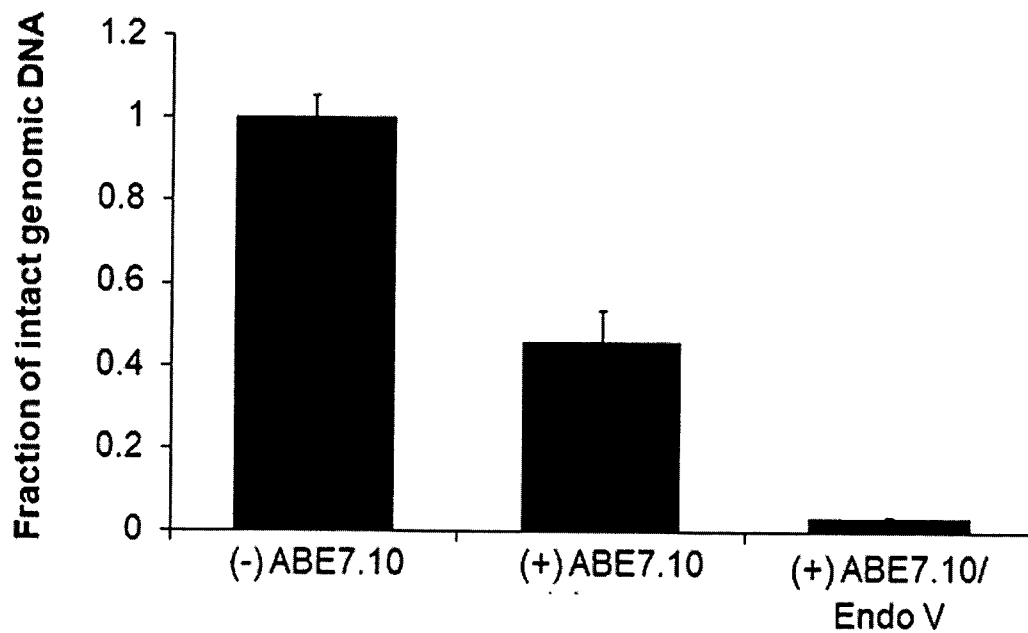
【도 7e】



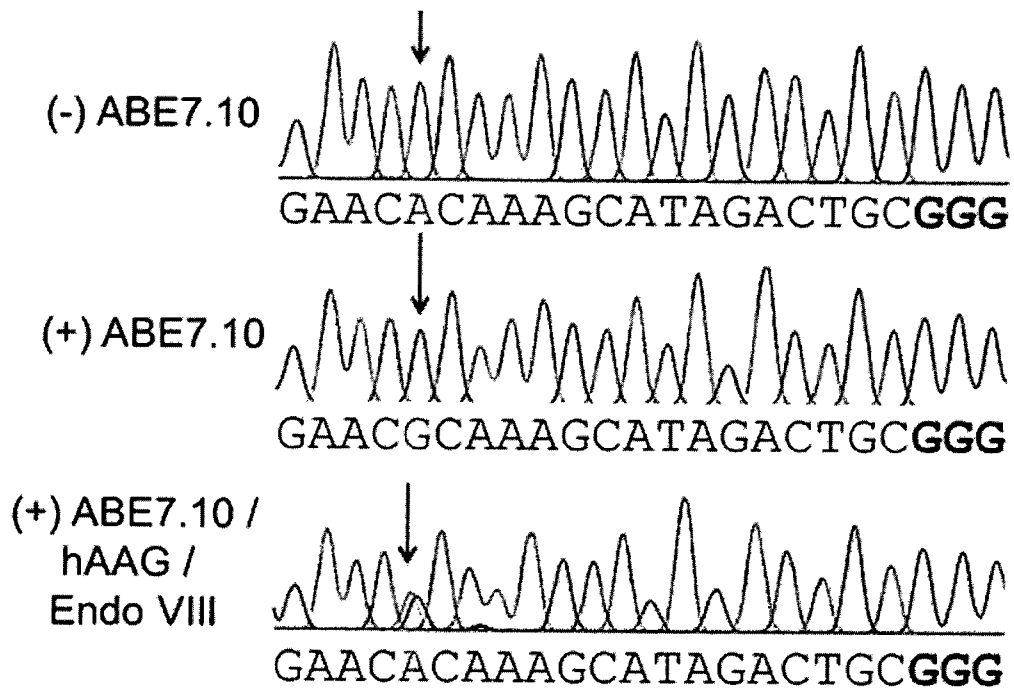
【도 7f】



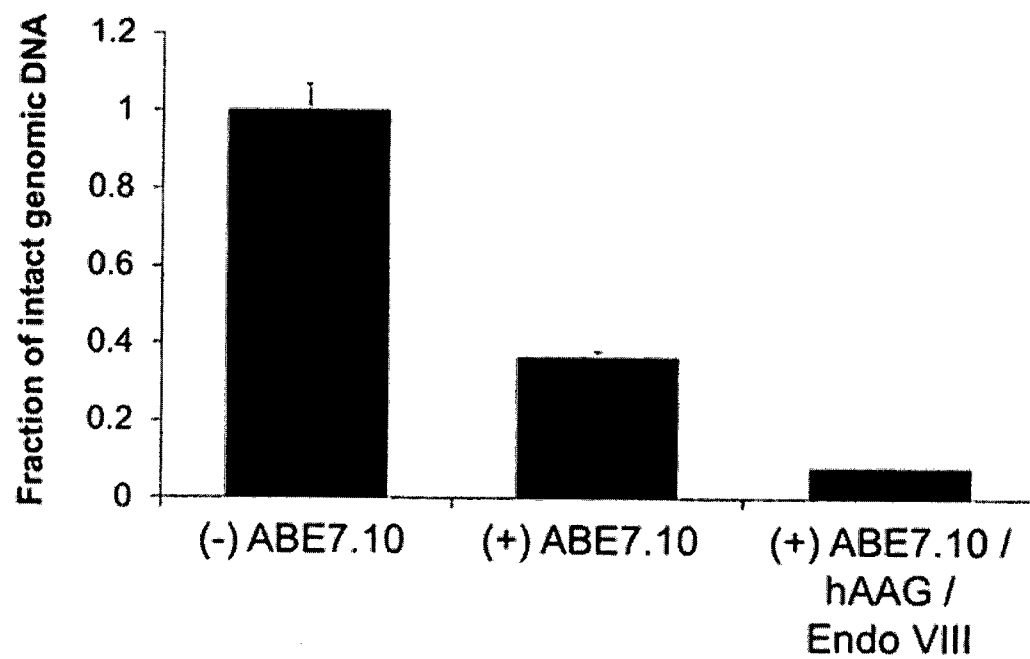
【도 7g】



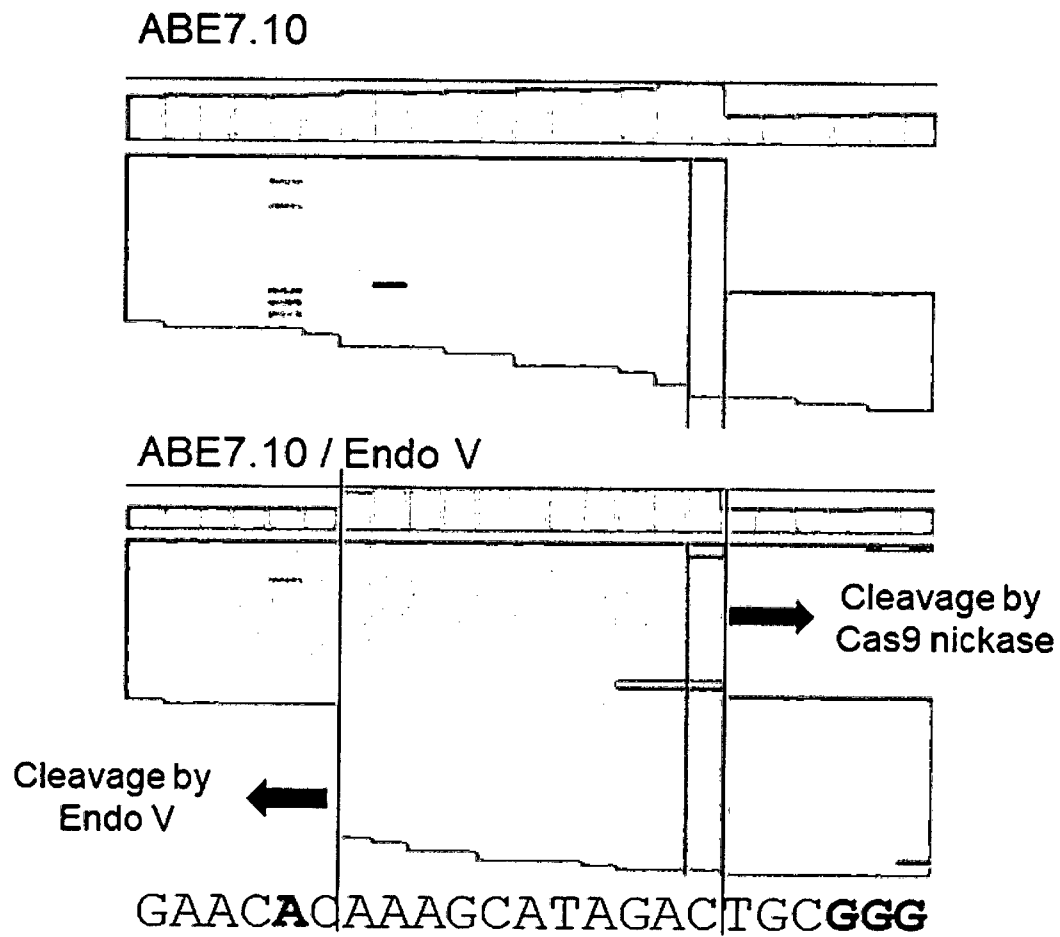
【도 7h】



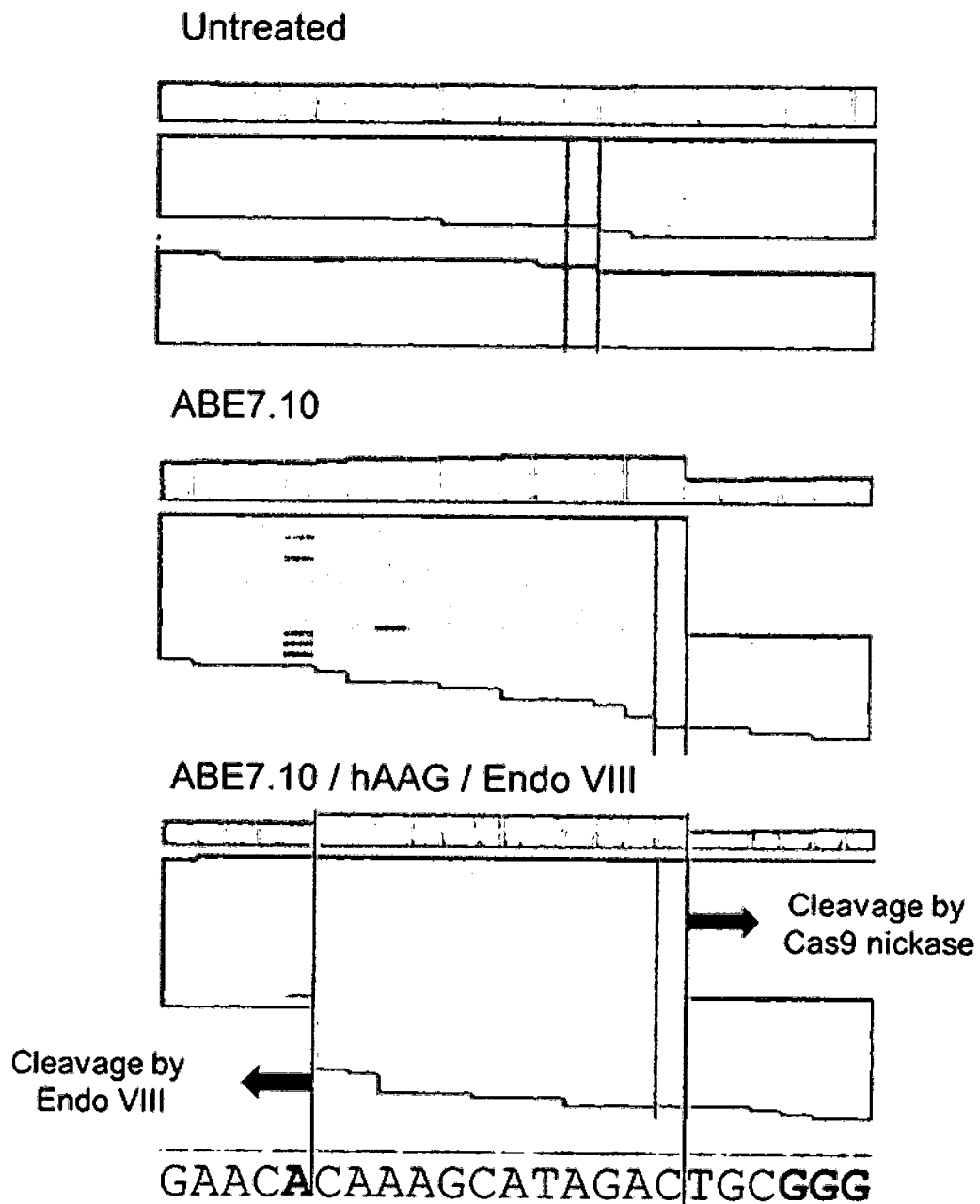
【도 7i】



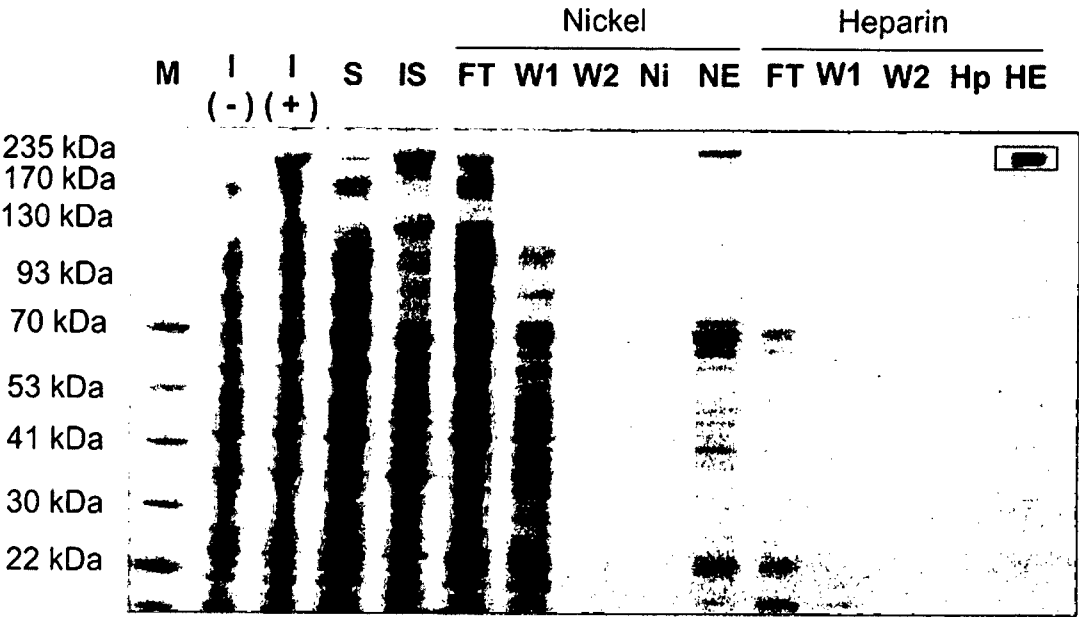
【도 7j】



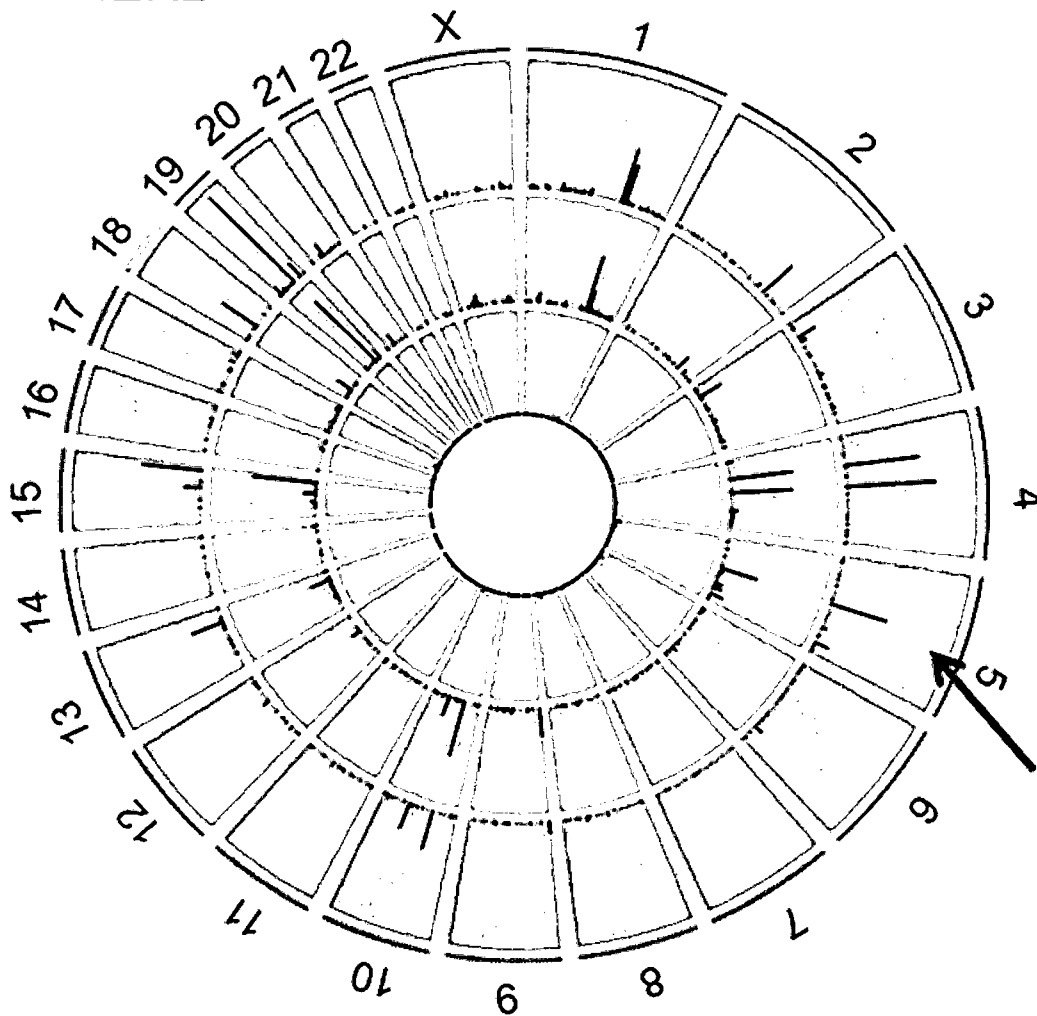
【도 7k】



【도 8】



【도 10a】

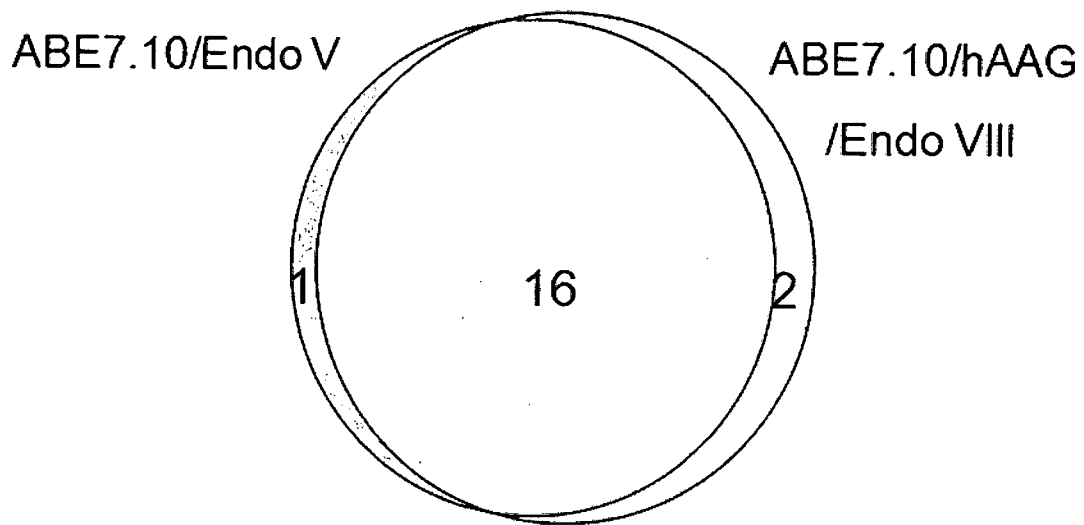
HEK2

Untreated

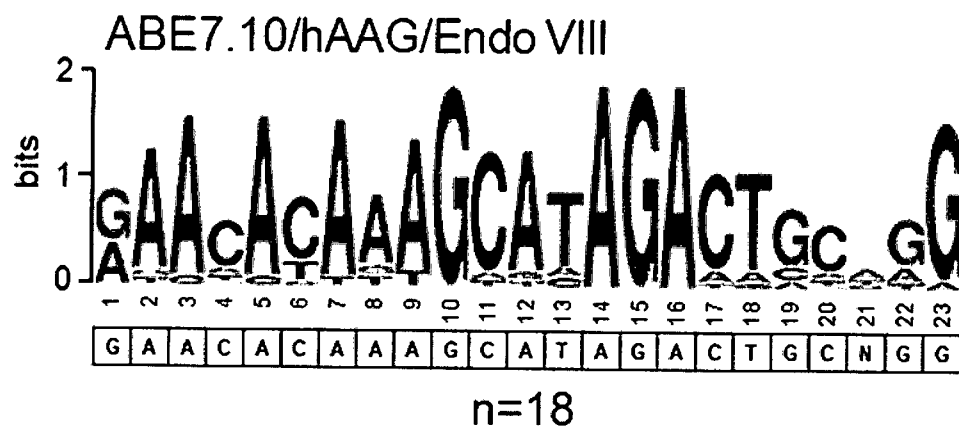
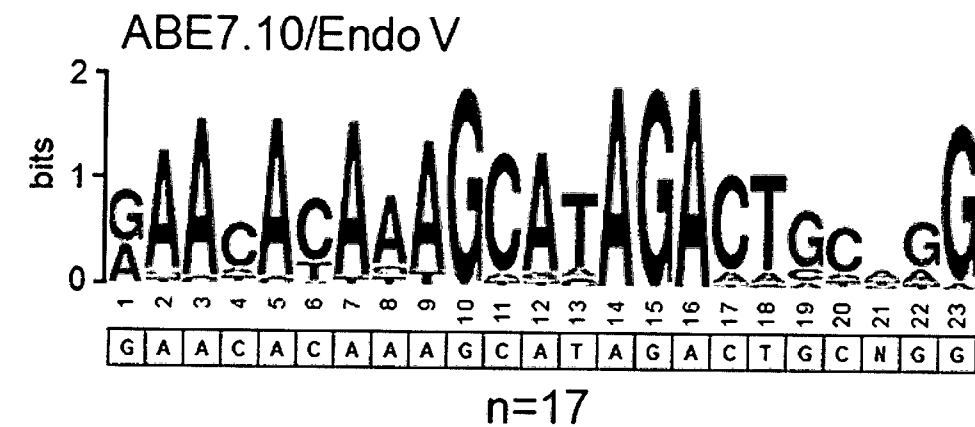
ABE7.10/Endo V

ABE7.10/hAAG/Endo VIII

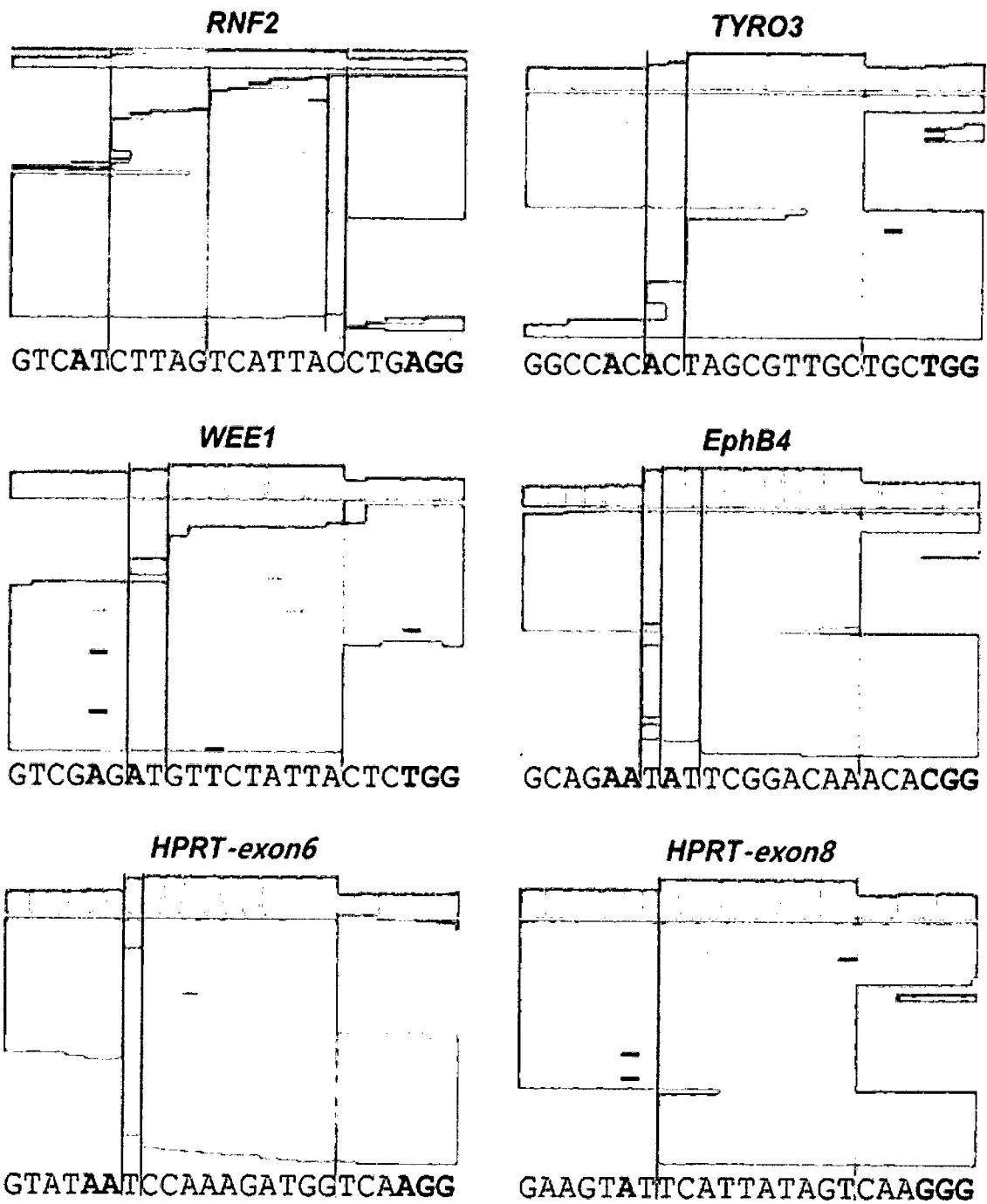
【도 10b】



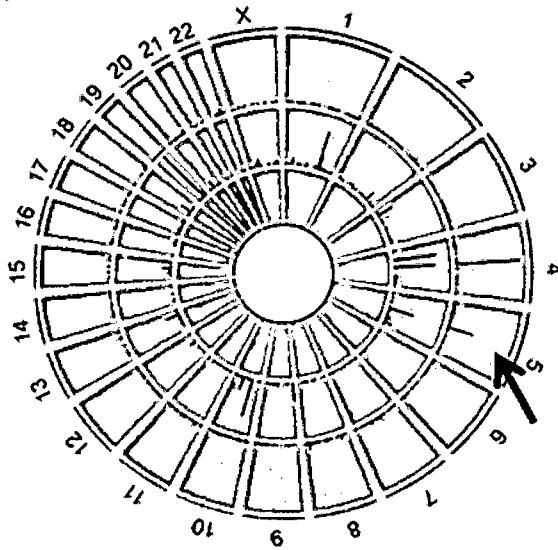
【도 10c】



【도 11】



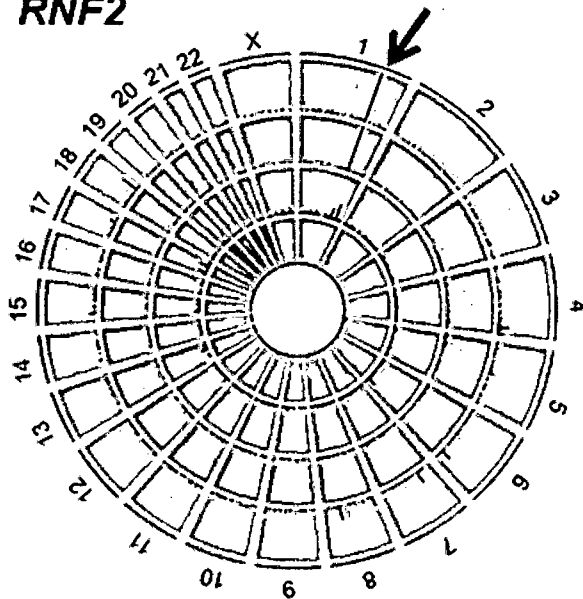
【도 12a】

HEK2

Untreated (+) ABE7.10

(+ BE3ΔUGI

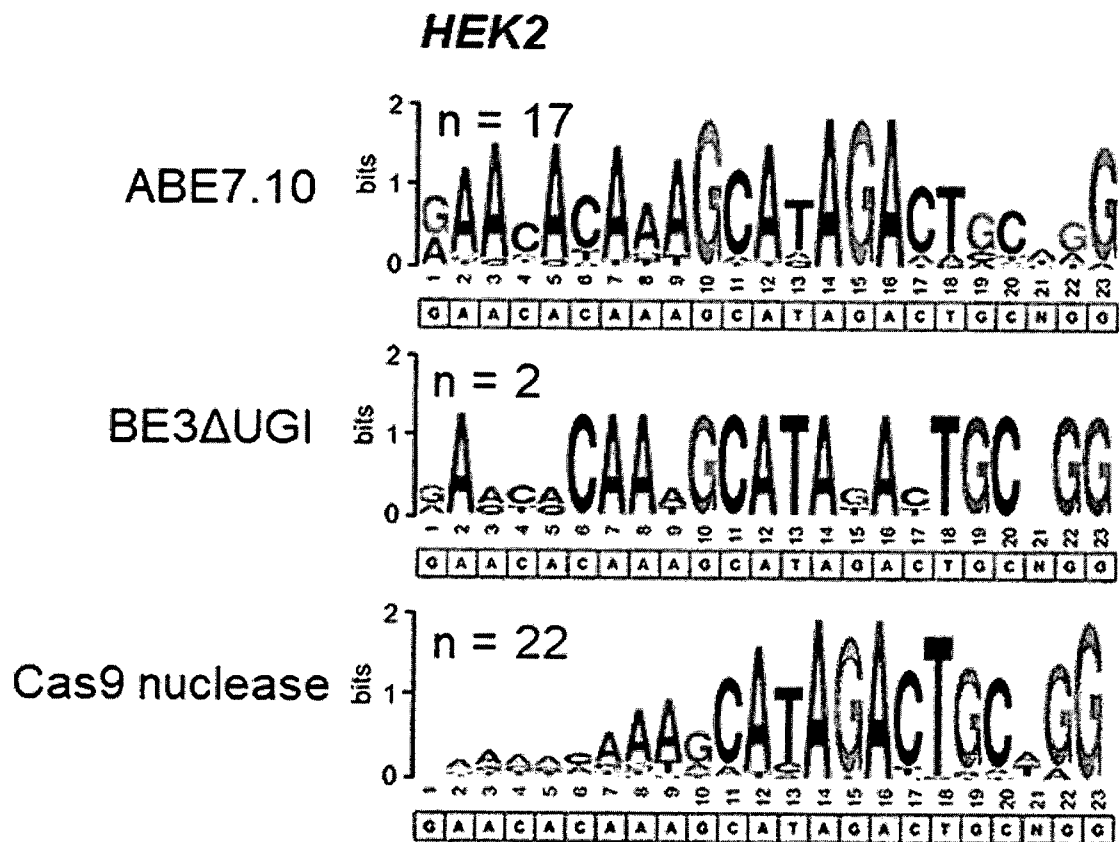
【도 12b】

RNF2

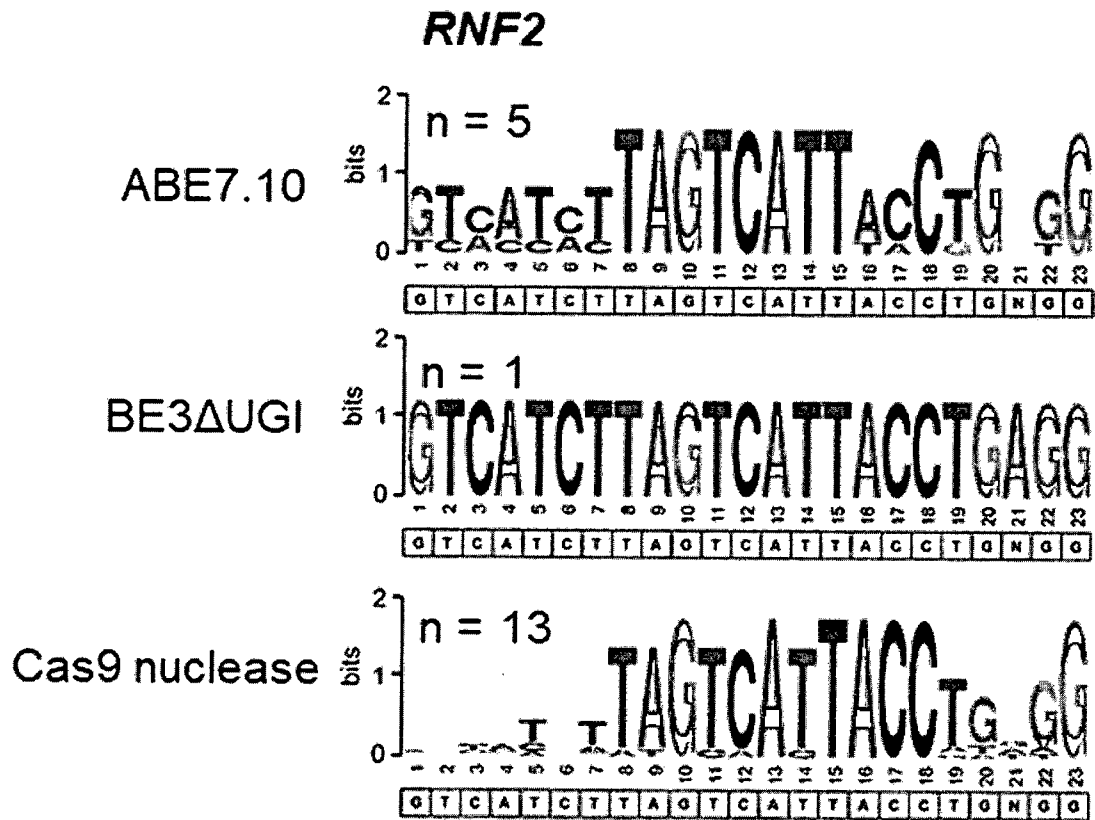
Untreated (+) ABE7.10

(+ BE3ΔUGI (+) Cas9 nuclease

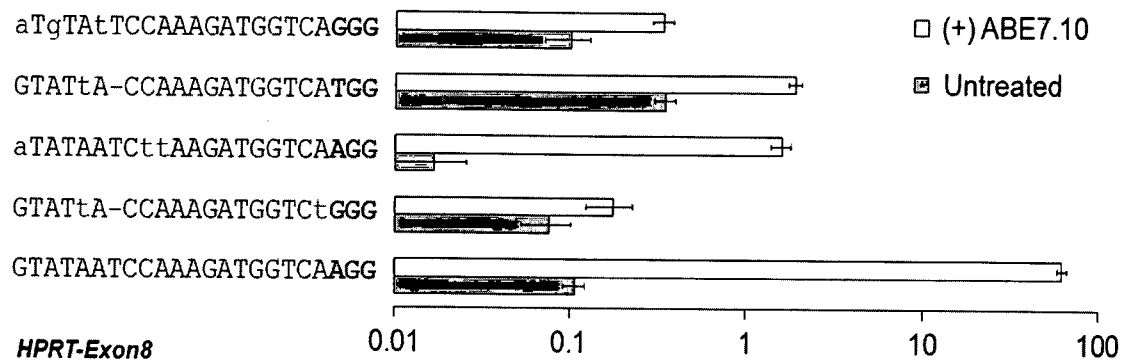
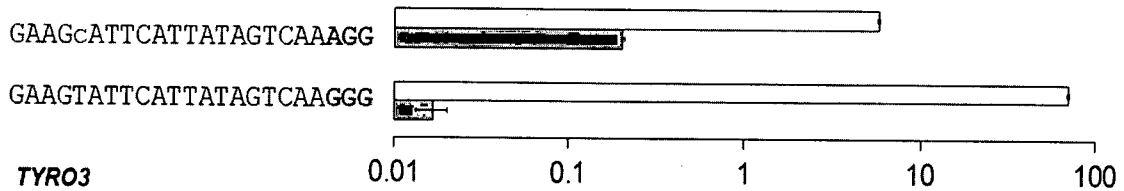
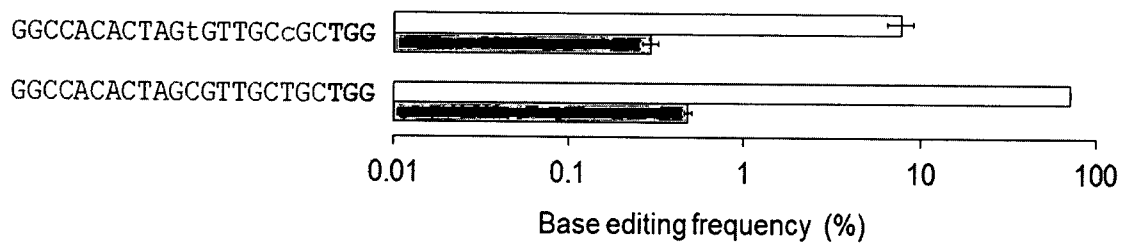
【도 12c】



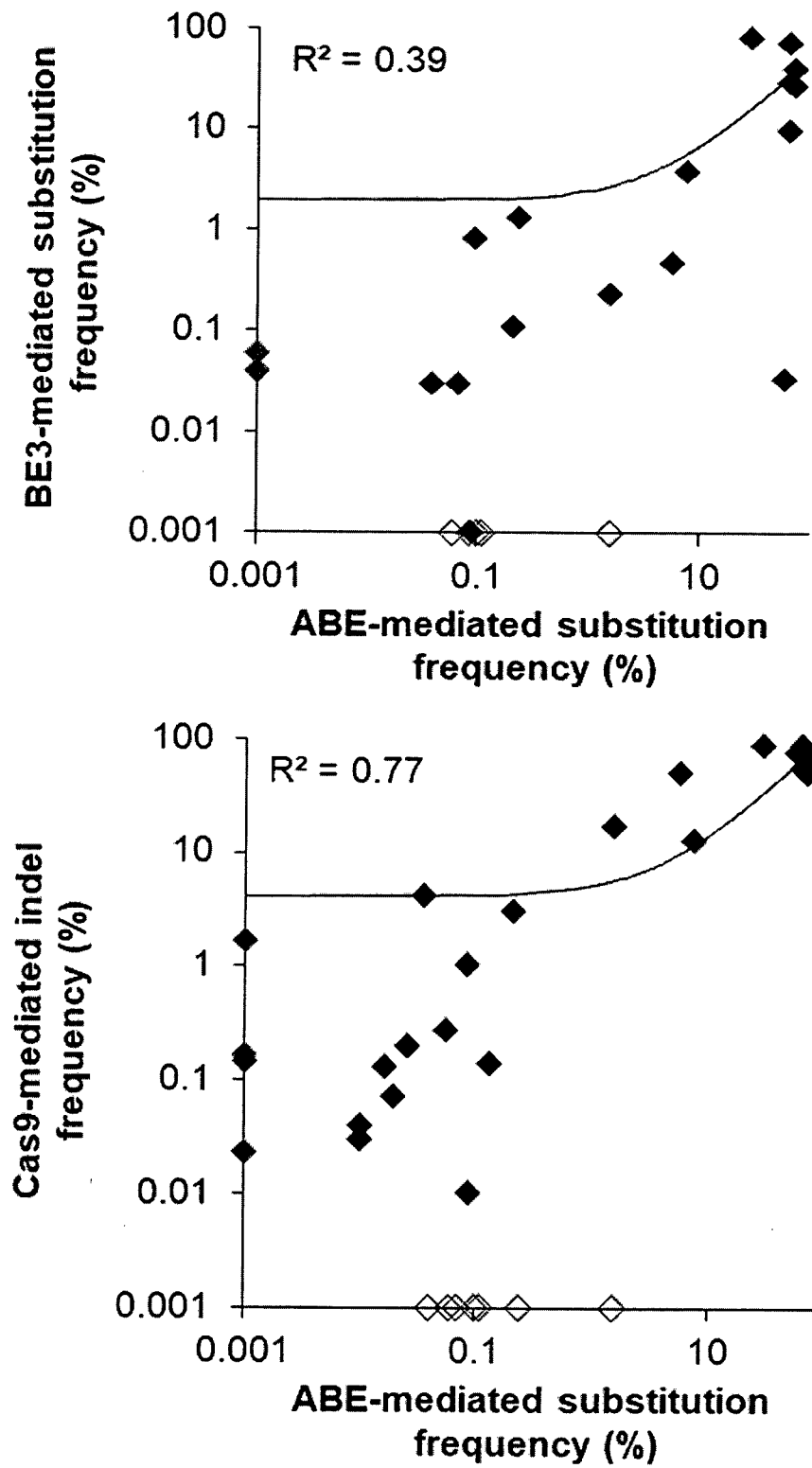
【도 12d】



【도 12e】

HPRT-Exon6**HPRT-Exon8****TYRO3**

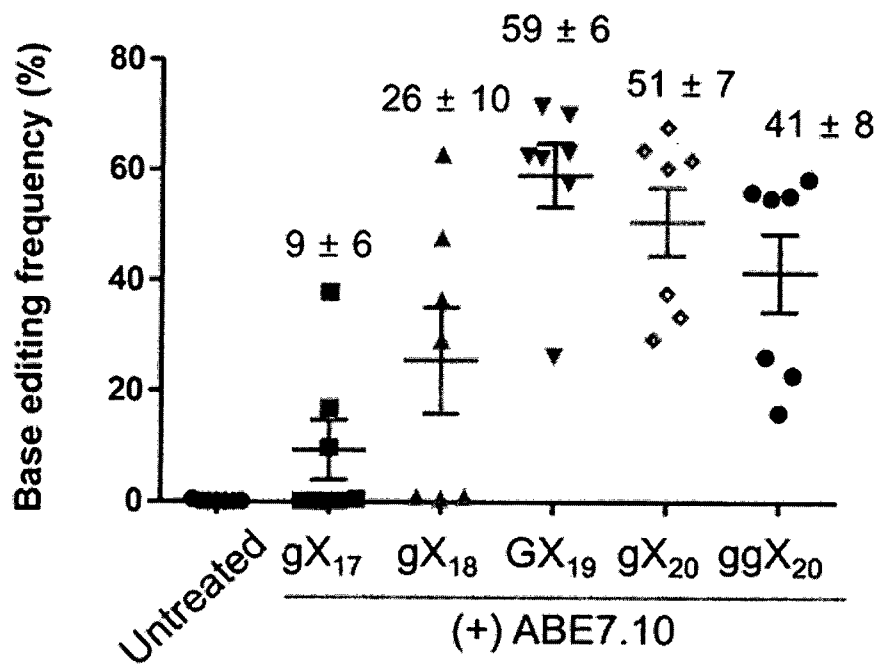
【도 12f】



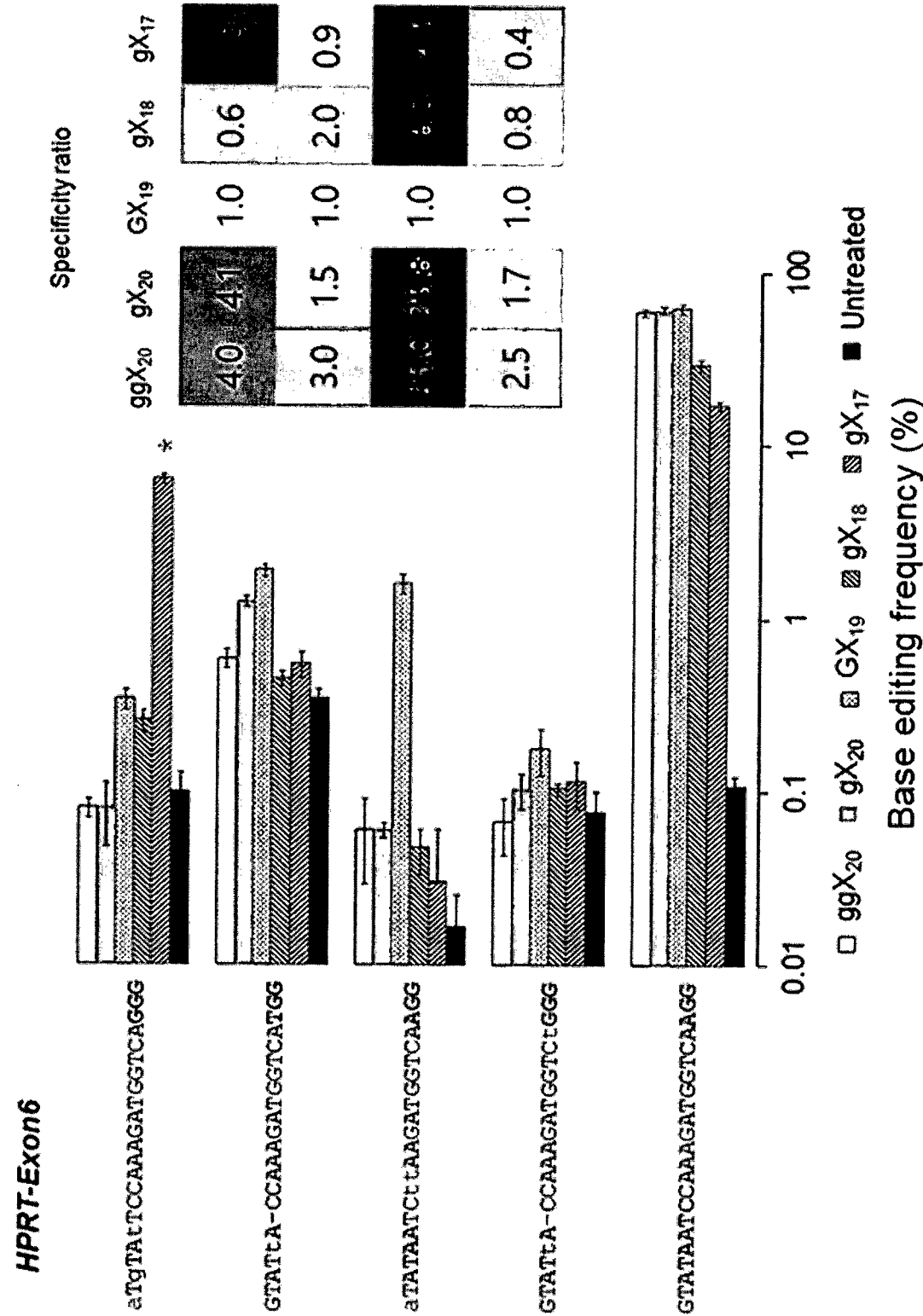
【도 13a】

5'-GTATAATCCAAAGATGGTCANGG-3'	<i>HPRT</i> exon 6 target sequence
5'-gTAATCCAAAGATGGTCA-3'	gX ₁₇ sgRNA
5'-gATAATCCAAAGATGGTCA-3'	gX ₁₈ sgRNA
5'-GTATAATCCAAAGATGGTCA-3'	GX ₁₉ sgRNA
5'-gGTATAATCCAAAGATGGTCA-3'	gX ₂₀ sgRNA
5'-ggGTATAATCCAAAGATGGTCA-3'	ggX ₂₀ sgRNA

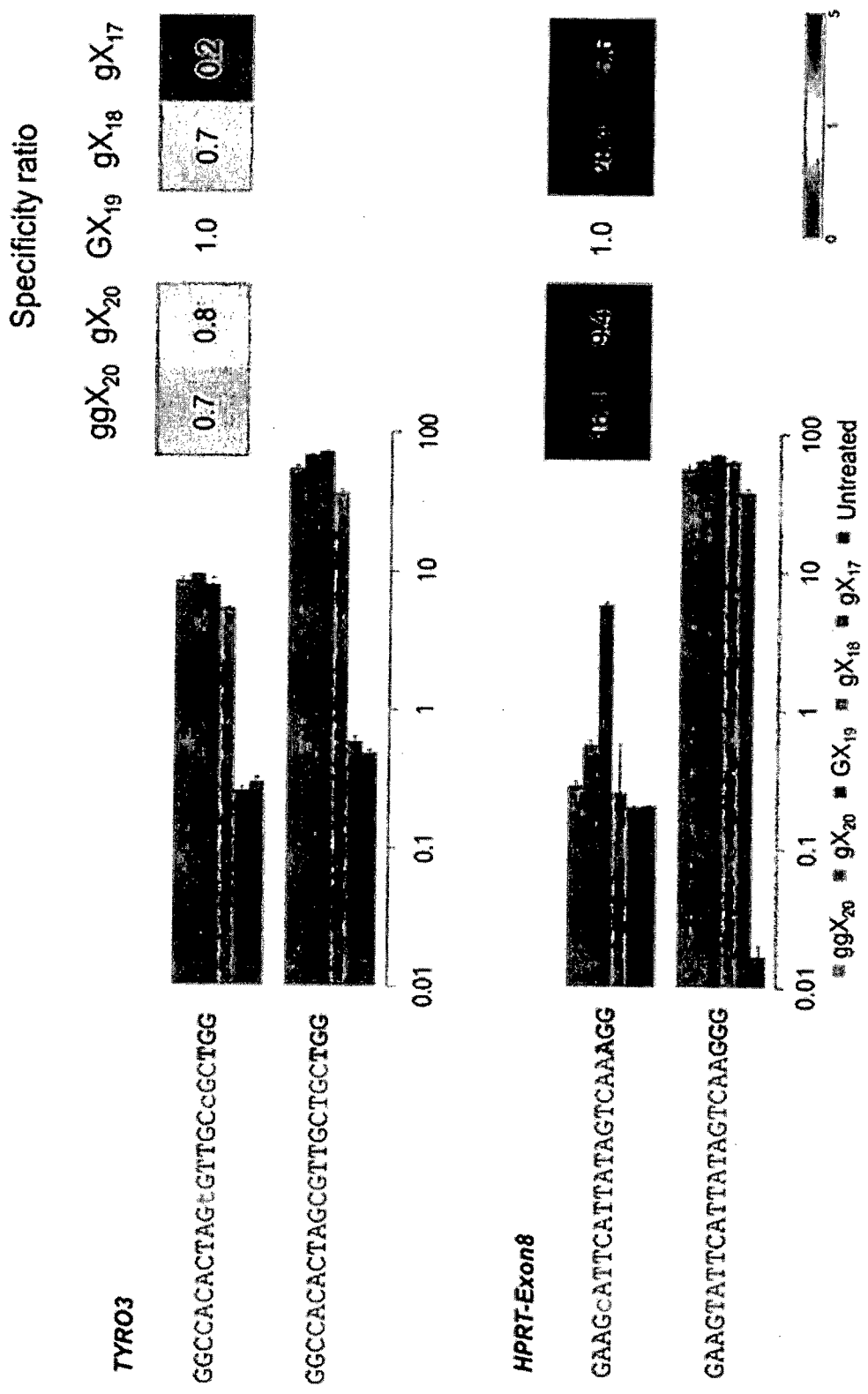
【도 13b】

GX₁₉ vs gX₁₈ : p-value 0.011GX₁₉ vs gX₁₇ : p-value < 0.0001

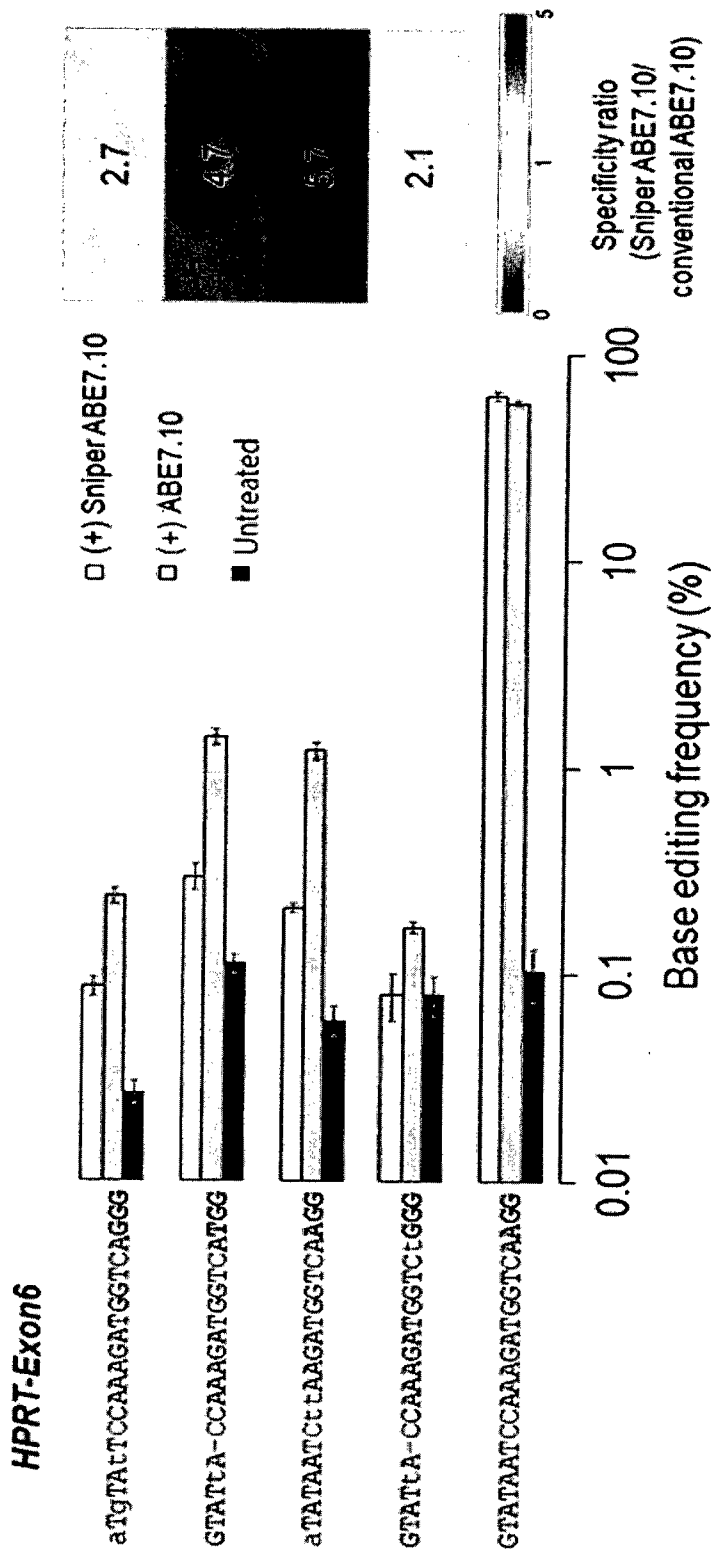
【도 13c】



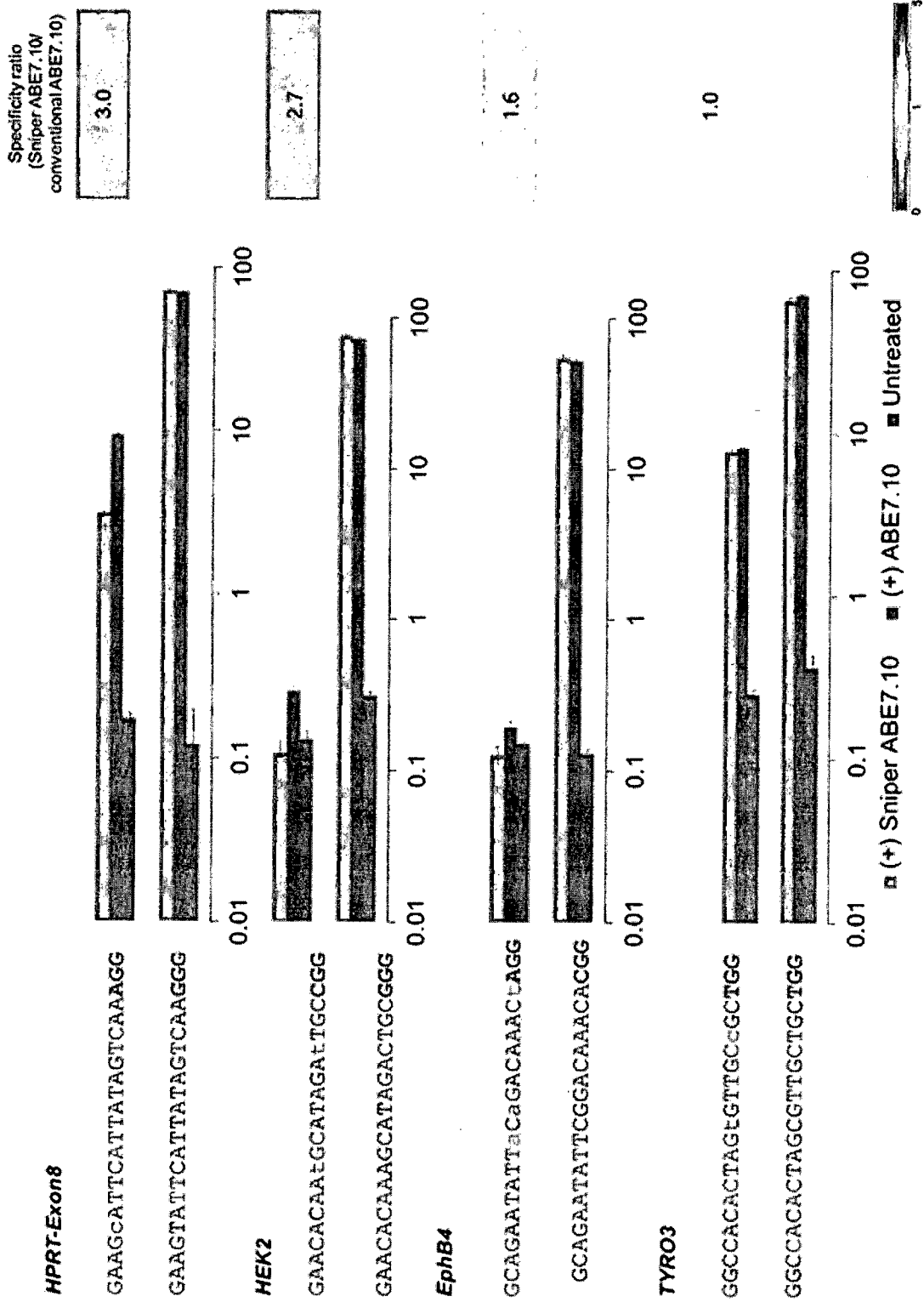
【図 13d】

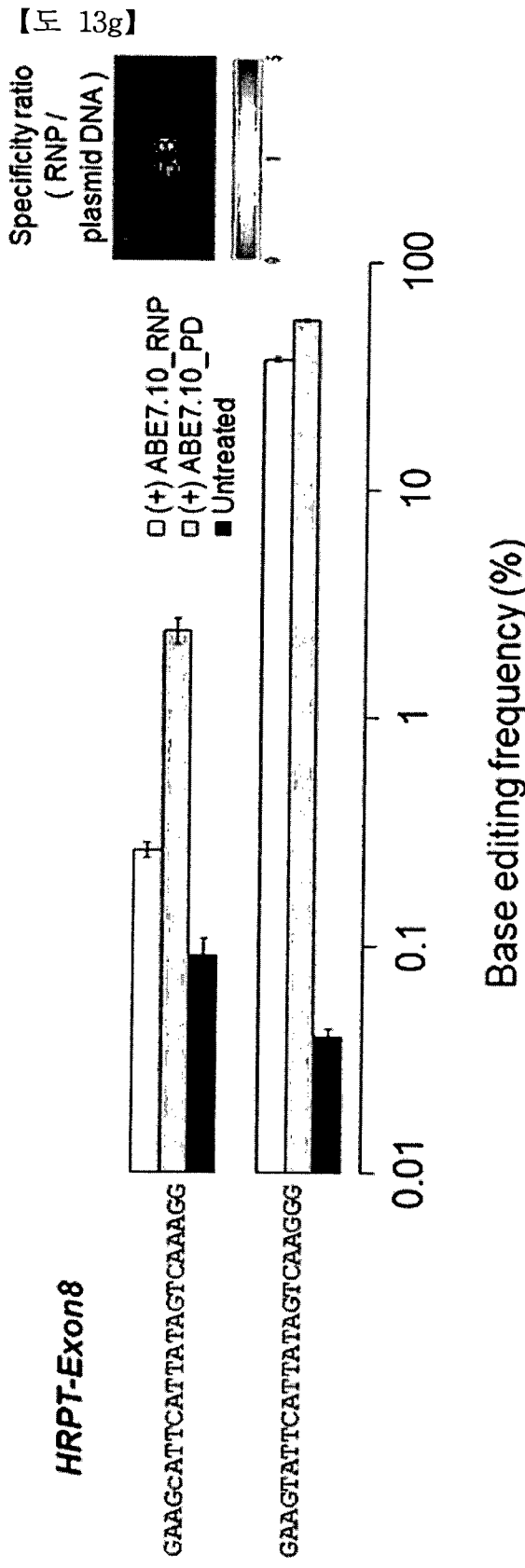


【도 13e】

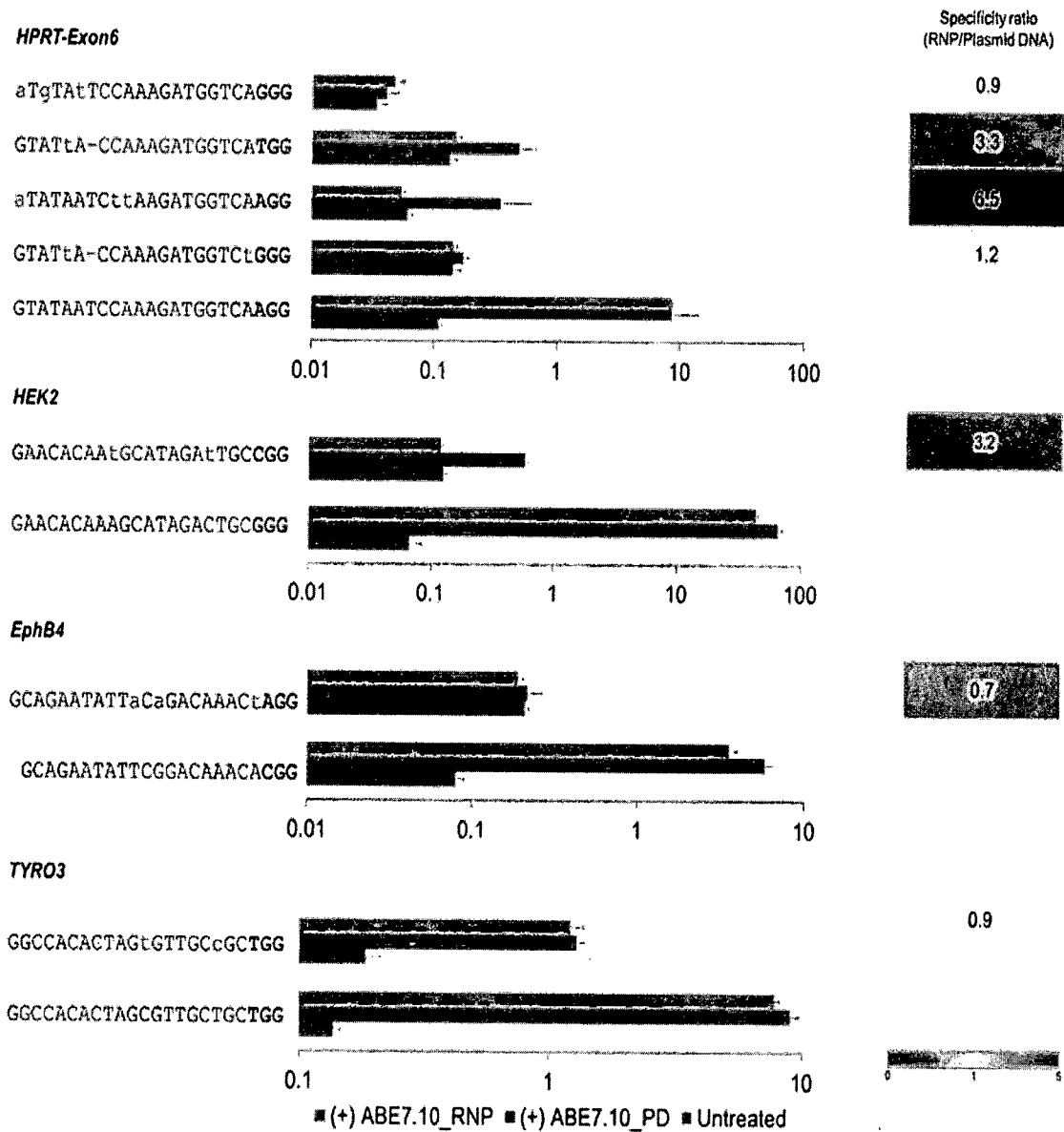


【图 13f】

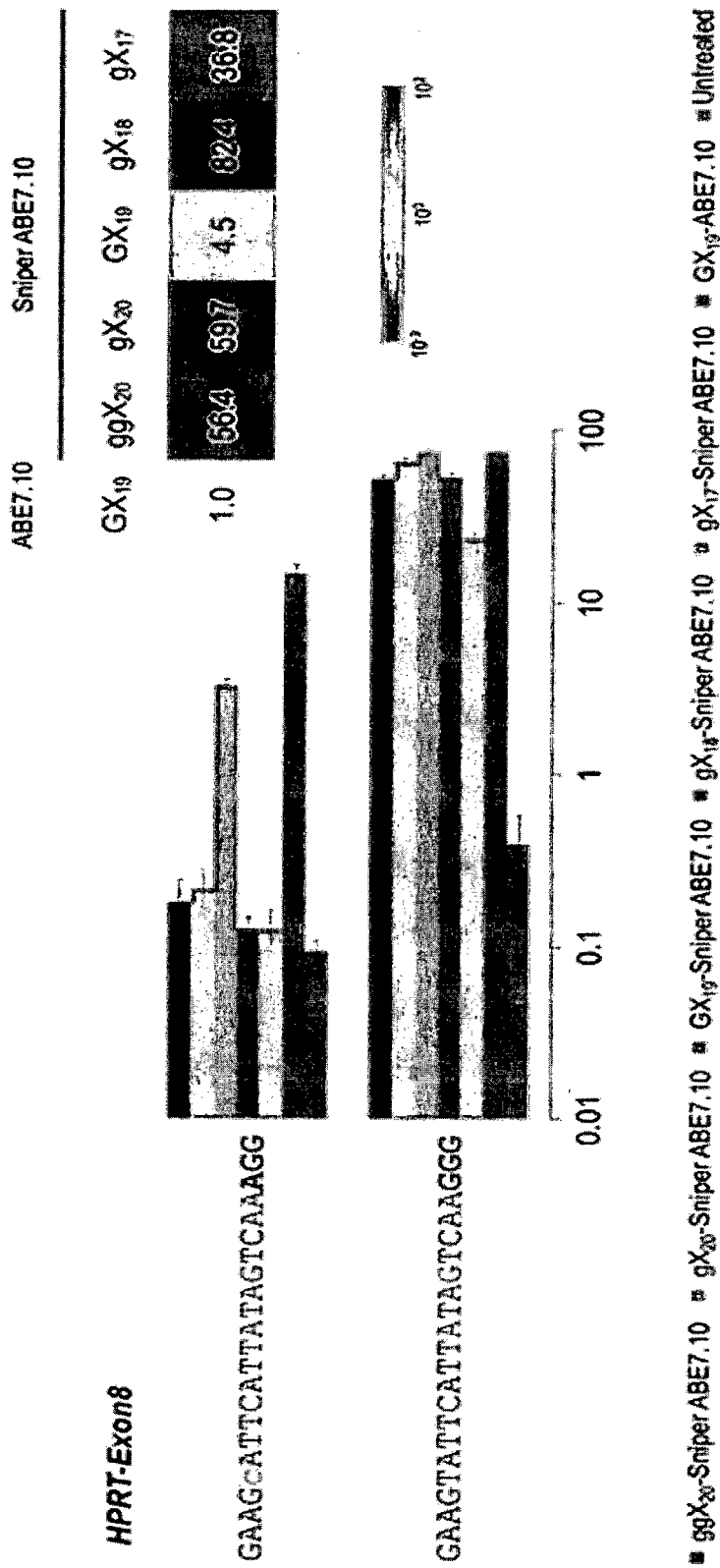




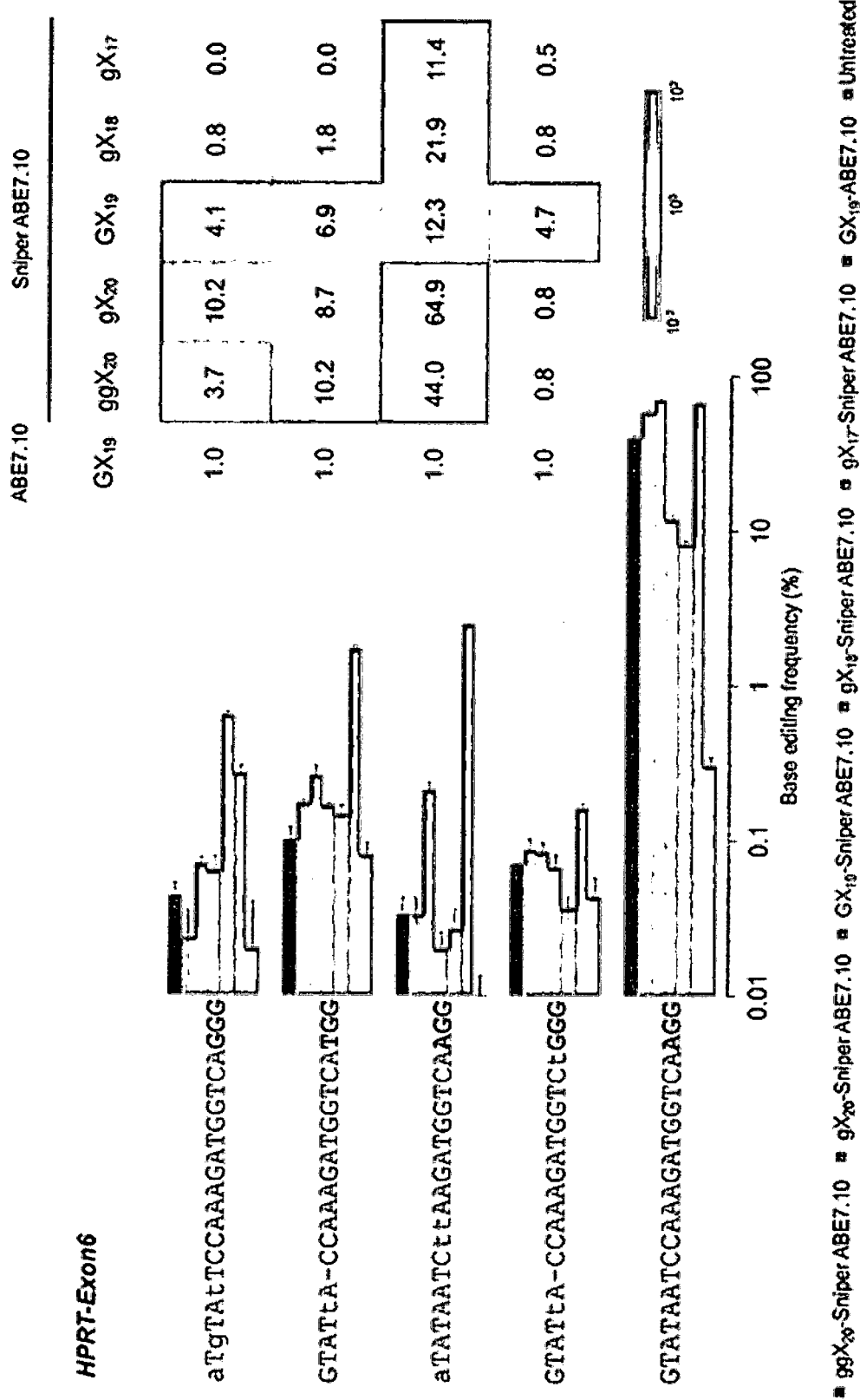
【도 13h】



【도 13i】



【도 13j】



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2019/001104

A. CLASSIFICATION OF SUBJECT MATTER

C12Q 1/6869(2018.01)i, C12Q 1/34(2006.01)i, C12N 9/22(2006.01)i, C12N 9/78(2006.01)i, C12N 9/96(2006.01)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)

C12Q 1/6869; C12N 15/01; C12N 9/22; C12N 9/78; C12Q 1/34; C12N 9/96

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: adenosine deaminase, plasmid, nuclease, guide RNA, Alkyladenine DNA Glycosylase (AAG), Cas9

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	GAUDELLI, N. M. et al. Programmable base editing of A · T to G · C in genomic DNA without DNA cleavage. Nature. 25 October 2017, vol. 551, pages 464-471 and Methods See abstract, page 464, right column, second paragraph, page 466, left column, first paragraph, figures 1, 2 and extended data, figure 2.	1-22
A	US 2015-0166980 A1 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) 18 June 2015 See the entire document.	1-22
A	WO 2016-022363 A2 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) 11 February 2016 See the entire document.	1-22
A	KOMOR, A. C. et al. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature. 19 May 2016, vol. 533, pages 420-424 See the entire document.	1-22
A	ZHENG, Y. et al. DNA editing in DNA/RNA hybrids by adenosine deaminases that act on RNA. Nucleic Acids Res. 28 January 2017, vol. 45, no. 6, pages 3369-3377 See the entire document.	1-22



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

“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

10 MAY 2019 (10.05.2019)

Date of mailing of the international search report

10 MAY 2019 (10.05.2019)

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex Daejeon Building 4, 189, Cheongsu-ro, Seo-gu,
Daejeon, 35208, Republic of Korea

Facsimile No. +82-42-481-8578

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2019/001104

Patent document cited in search report	Publication date	Patent family member	Publication date
US 2015-0166980 A1	18/06/2015	AU 2014-362208 A1	30/06/2016
		CA 2933625 A1	18/06/2015
		CN 105934516 A	07/09/2016
		EP 3080265 A1	19/10/2016
		JP 2017-500035 A	05/01/2017
		US 2015-0165054 A1	18/06/2015
		US 2015-0166981 A1	18/06/2015
		US 2015-0166982 A1	18/06/2015
		US 2015-0166983 A1	18/06/2015
		US 2015-0166984 A1	18/06/2015
		US 2015-0166985 A1	18/06/2015
		US 2016-0304846 A1	20/10/2016
		US 9068179 B1	30/06/2015
		US 9840699 B2	12/12/2017
		WO 2015-089406 A1	18/06/2015
WO 2016-022363 A2	11/02/2016	AU 2015-298571 A1	23/02/2017
		CA 2956224 A1	11/02/2016
		EP 3177718 A2	14/06/2017
		US 0077453 B2	18/09/2018
		US 2017-0268022 A1	21/09/2017
		WO 2016-022363 A3	26/05/2016
		WO 2016-022363 A9	11/02/2016

A. 발명이 속하는 기술분류(국제특허분류(IPC))

C12Q 1/6869(2018.01)i, C12Q 1/34(2006.01)i, C12N 9/22(2006.01)i, C12N 9/78(2006.01)i, C12N 9/96(2006.01)i

B. 조사된 분야

조사된 최소문헌(국제특허분류를 기재)

C12Q 1/6869; C12N 15/01; C12N 9/22; C12N 9/78; C12Q 1/34; C12N 9/96

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: 아데노신 디아미나아제, 플라스미드, 뉴클레아제, 가이드 RNA, Alkyladenine DNA Glycosylase(AAG), Cas9

C. 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
A	GAUDELLI, N. M. 등, `Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage`, Nature, 2017.10.25, 551권, 페이지 464-471 및 Methods 요약, 페이지 464 오른쪽 컬럼 두 번째 단락, 페이지 466 왼쪽 컬럼 첫 번째 단락, 도면 1,2 및 확장 자료 도면 2 참조.	1-22
A	US 2015-0166980 A1 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) 2015.06.18 전체 문헌 참조.	1-22
A	WO 2016-022363 A2 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) 2016.02.11 전체 문헌 참조.	1-22
A	KOMOR, A. C. 등, `Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage`, Nature, 2016.05.19, 533권, 페이지 420-424 전체 문헌 참조.	1-22
A	ZHENG, Y. 등, `DNA editing in DNA/RNA hybrids by adenosine deaminases that act on RNA`, Nucleic Acids Res., 2017.01.28, 45권, 6호, 페이지 3369-3377 전체 문헌 참조.	1-22

☐ 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

“T” 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌

“E” 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후에 공개된 선출원 또는 특허 문헌

“X” 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다.

“L” 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌

“Y” 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다.

“O” 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌

“&” 동일한 대응특허문헌에 속하는 문헌

“P” 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌

국제조사의 실제 완료일

2019년 05월 10일 (10.05.2019)

국제조사보고서 발송일

2019년 05월 10일 (10.05.2019)

ISA/KR의 명칭 및 우편주소

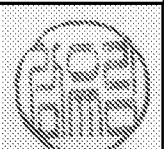
대한민국 특허청
(35208) 대전광역시 서구 청사로 189,
4동 (둔산동, 정부대전청사)

팩스 번호 +82-42-481-8578

심사관

감유림

전화번호 +82-42-481-3516



국제조사보고서에서
인용된 특허문헌

공개일

대응특허문헌

공개일

US 2015-0166980 A1

2015/06/18

AU 2014-362208 A1

2016/06/30

CA 2933625 A1

2015/06/18

CN 105934516 A

2016/09/07

EP 3080265 A1

2016/10/19

JP 2017-500035 A

2017/01/05

US 2015-0165054 A1

2015/06/18

US 2015-0166981 A1

2015/06/18

US 2015-0166982 A1

2015/06/18

US 2015-0166983 A1

2015/06/18

US 2015-0166984 A1

2015/06/18

US 2015-0166985 A1

2015/06/18

US 2016-0304846 A1

2016/10/20

US 9068179 B1

2015/06/30

US 9840699 B2

2017/12/12

WO 2015-089406 A1

2015/06/18

WO 2016-022363 A2

2016/02/11

AU 2015-298571 A1

2017/02/23

CA 2956224 A1

2016/02/11

EP 3177718 A2

2017/06/14

US 0077453 B2

2018/09/18

US 2017-0268022 A1

2017/09/21

WO 2016-022363 A3

2016/05/26

WO 2016-022363 A9

2016/02/11