

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구

국제사무국

(43) 국제공개일

2018년 5월 31일 (31.05.2018)



(10) 국제공개번호

WO 2018/097657 A1

(51) 국제특허분류:

C12Q 1/68 (2006.01)

C12N 9/22 (2006.01)

(21) 국제출원번호:

PCT/KR2017/013547

(22) 국제출원일:

2017년 11월 24일 (24.11.2017)

(25) 출원언어:

한국어

(26) 공개언어:

한국어

(30) 우선권정보:

10-2016-0158438 2016년 11월 25일 (25.11.2016) KR

(71) 출원인: 주식회사 툴젠 (TOOLGEN INCORPORATED) [KR/KR]; 08501 서울시 금천구 가산디지털1로 219, 1204호, Seoul (KR). 서울대학교산학협력단 (SEOUL NATIONAL UNIVERSITY R&DB FOUNDATION) [KR/KR]; 08826 서울특별시 관악구 관악로 1, Seoul (KR).

(72) 발명자: 김대식 (KIM, Daesik); 08826 서울시 관악구 관악로 1, Seoul (KR).

(74) 대리인: 유미 특허법인 (YOU ME PATENT & 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, 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).

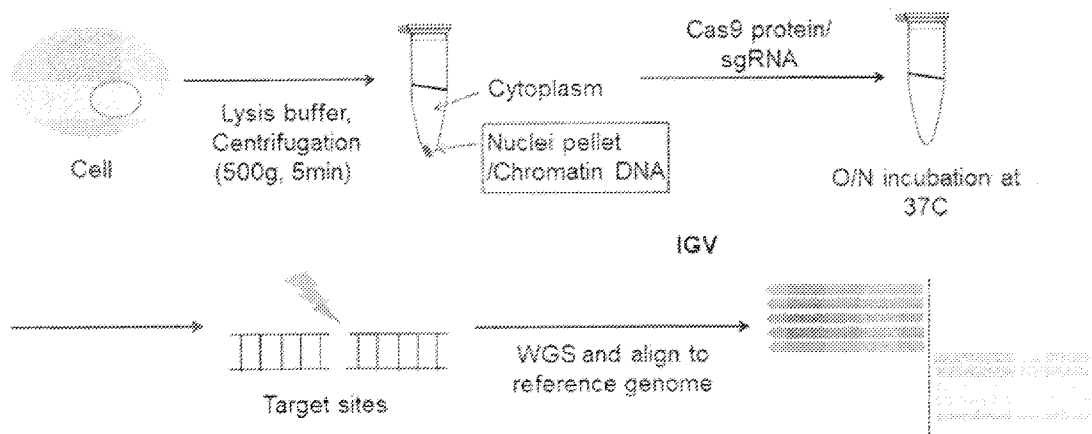
공개:

— 국제조사보고서와 함께 (조약 제21조(3))

— 명세서의 서열목록 부분과 함께 (규칙 5.2(a))

(54) Title: GENOME SEQUENCING METHOD AND GENOME EDITING IDENTIFICATION METHOD, WHICH USE CHROMATIN DNA

(54) 발명의 명칭: 크로마틴 DNA를 이용한 유전체 서열분석 방법 및 유전체 교정 확인 방법



(57) Abstract: The present invention relates to a method for detecting cleavage sites and/or off-target sites of a target-specific nuclease in genomic DNA and, particularly, treats chromatin protein-containing genomic DNA in vitro with a target-specific nuclease so as to cleave a genome, and then subjects the same to whole genome sequencing so as to identify off-target sites through data analysis.

(57) 요약서: 본 발명은 유전체 DNA에서의 표적 특이적 뉴클레아제의 절단 위치 및/또는 비표적 위치 (off-target site)를 검출하는 방법에 관한 것으로서, 구체적으로는 시험관 내(in vitro)에서 크로마틴 단백질을 포함하는 유전체 DNA에 표적 특이적 뉴클레아제를 처리하여 유전체를 절단한 뒤, 이를 전체 유전체 시퀀싱 (whole genome sequencing) 하여 데이터 분석을 통해 비표적 위치를 확인하는 것을 특징으로 한다.



WO 2018/097657 A1

【명세서】

【발명의 명칭】

크로마틴 DNA를 이용한 유전체 서열분석 방법 및 유전체 교정 확인 방법

5

【기술분야】

본 발명은 유전체 DNA에서의 표적 특이적 뉴클레아제의 절단 위치 및/또는 비표적 위치 (off-target site)를 검출하는 방법에 관한 것으로서, 구체적으로는 시험관 내 (*in vitro*)에서 크로마틴 단백질을 포함하는 유전체 DNA에 표적 특이적 뉴클레아제를 처리하여 유전체를 절단한 뒤, 이를 전체 유전체 시퀀싱 (whole genome sequencing)하여 데이터 분석을 통해 비표적 위치를 확인하는 것을 특징으로 한다.

10

【배경 기술】

RNA 유전자 가위 (RNA-guided engineered nucleases, RGEN)의 발전으로 인간 세포를 비롯한 다양한 동식물의 유전체 교정에 사용되어 왔다.

15

예컨대, ZFN (zinc finger nuclease), TALEN (transcriptional activator-like effector nuclease), 및 제2형 CRISPR / Cas (clustered regularly interspaced repeat / CRISPR-associated) 원핵생물 획득 면역 시스템 유래 RGEN (RNA-guided engineered nuclease) 등의 유전자 가위 (programmable nucleases)가 배양된 세포 및/또는 개체의 유전체 교정에 널리 사용되고 있다. 상기 유전자 가위를 이용한 유전체 교정 기술은 생명과학, 생명공학 및 의학분야 등에서 다양한 목적으로 이용될 수 있는 매우 유용한 기술이다. 예를 들어, 줄기세포 또는 체세포에서 표적화된 유전자 변형을 일으킴으로써 다양한 유전적 질환 또는 후천적 질환에 대한 유전자/세포 치료가 가능해졌다. 다만, 상기 유전자 가위들은 표적 위치 (on-target site) 뿐만 아니라 이와 상동성을 가지는 비표적 위치 (off-target site)를 절단하여, 원하지 않는 돌연변이를 일으킬 수 있다.

20

25

대표적인 예로, *S. pyogenes* 유래 Cas9 단백질 및 sgRNA (small guide RNA)로 구성되는 RGEN은 sgRNA와 혼성화 되는 20-bp (base pair) 서열 및 Cas9에 의해 인식되는 PAM (protospacer-adjacent motif) 서열인 5'-NGG-3'으로 구성되는 23-bp의 표적 DNA 서열을 인식하지만, 일부 뉴클레오티드 서열이 일치하지 않는 경우에도 작동될 수 있다. 나아가,

30

5 RGEN은 sgRNA 서열과 비교하여 추가 염기서열을 가지거나 (DNA bulge) 또는 하나 이상의 염기가 없는 (RNA bulge) 비표적 DNA 서열도 절단할 수 있다. 이와 유사하게 ZFN과 TALEN도 일부 염기가 다른 서열을 절단할 수 있다. 이는 유전체에 유전자 가위를 적용할 경우 표적 위치 이외에 상당한 수의 비표적 위치를 가질 수 있다는 것을 시사한다.

비표적 DNA 절단은 전암유전자 (proto-oncogene) 및 암억제유전자 (tumor suppressor gene)와 같이 원치 않는 유전자에서 돌연변이를 야기하거나, 및/또는 전위 (translocation), 결실 (deletion), 및 역위 (inversion)와 같은 유전체 재조합을 증가시킬 수 있어, 연구 분야 및 의학 분야 등에서 유전자 가위를 이용하는데 심각한 문제가 된다. 이러한 유전자 가위의 비표적 효과를 감소시키기 위해 다양한 전략이 보고되고는 있으나, 전체 유전체 수준에서 비표적 효과 없이 표적 위치에만 특이적으로 작동하는 유전자 가위는 아직까지 보고된 바 없다. 이러한 문제점을 다루기 위해, 유전자 가위의 특이성을 유전체 수준에서 확인할 수 있는 기술의 개발이 요구된다.

【발명의 상세한 설명】

【기술적 과제】

일 예는

20 (a) 분리된 유전체 (genomic) DNA 를 표적 특이적 뉴클레아제로 절단하는 단계; 및

(b) 상기 절단된 DNA 에 대한 전체 유전체 시퀀싱 (whole genome sequencing, WGS)을 수행하는 단계

를 포함하고,

25 상기 분리된 유전체 DNA 는 크로마틴 단백질을 포함하는 크로마틴 DNA 인,

유전체 DNA 서열 분석 방법을 제공한다.

다른 예는

30 (a) 분리된 유전체 (genomic) DNA 를 표적 특이적 뉴클레아제로 절단하는 단계; 및

(b) 상기 절단된 DNA 에 대한 전체 유전체 시퀀싱 (whole genome sequencing)을 수행하는 단계; 및

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

를 포함하고,

상기 분리된 유전체 DNA 는 크로마틴 단백질을 포함하는 크로마틴 DNA 인,

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

다른 예는 상기 표적 특이적 뉴클레아제의 절단 위치 또는 비표적 위치 (off-target site)를 검출하는 방법을 사용하여 비표적 위치가 적은 표적 부위를 선별하는 방법을 제공한다.

10 【기술적 해결방법】

본 명세서에서는 전체 유전체에서 RNA 유전자 가위 (programmable nuclease; RNA-guided engineered nucleases, RGEN)와 같은 표적 특이적 뉴클레아제의 비표적 절단 위치 (off-target site)를 확인하는 방법의 일환으로, 절단 유전체 시퀀싱 (digested genome sequencing, Digenome-seq; 표적 특이적 뉴클레아제 처리 전과 후를 한 눈에 파악해 잘린 위치를 구별하는 방식) 기술을 제공하며, 특히, 절단 유전체 시퀀싱 기술이 세포내의 크로마틴 단백질을 모두 제거한 DNA 를 대상으로 진행하기 때문에 실제 세포내의 크로마틴 상태를 고려하지 못하는 한계가 있음을 인식하여, 이러한 한계점을 보완하고자 크로마틴 구조가 보존 된 크로마틴 DNA 를 이용한 절단 유전체 시퀀싱 방법을 제안한다.

일 예는,

(a) 분리된 유전체 (genomic) DNA 를 표적 특이적 뉴클레아제로 절단하는 단계; 및

(b) 상기 절단된 DNA 에 대한 전체 유전체 시퀀싱 (whole genome sequencing)을 수행하는 단계

를 포함하고,

상기 분리된 유전체 DNA 는 크로마틴 단백질을 포함하는 크로마틴 DNA 인,

유전체 DNA (genome DNA) 서열 분석 방법을 제공한다.

다른 예는,

(a) 분리된 유전체 (genomic) DNA 를 표적 특이적 뉴클레아제로 절단하는 단계;

(b) 상기 절단된 DNA 에 대한 전체 유전체 시퀀싱 (whole genome sequencing)을 수행하는 단계; 및

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

를 포함하고,

상기 분리된 유전체 DNA 는 크로마틴 단백질을 포함하는 크로마틴
5 DNA 인,

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

다른 예는,

(a) 분리된 유전체 (genomic) DNA 를 표적 특이적 뉴클레아제로
절단하는 단계;

15 (b) 상기 절단된 DNA 에 대한 전체 유전체 시퀀싱 (whole genome sequencing)을 수행하는 단계; 및

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

를 포함하고,

20 상기 분리된 유전체 DNA 는 크로마틴 단백질을 포함하는 크로마틴
DNA 인,

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

상기 표적 특이적 뉴클레아제의 교정 효율 확인 방법은 상기 (c)
25 단계 이후에, 상기 확인된 절단된 위치가 표적 위치 (on-target site)가
아닌 경우, 비표적 위치 (off-target site)로 판단하는 단계 ((d) 단계) 및
상기 비표적 위치에서의 절단 정도 (비표적 위치 개수 및/또는 비표적
위치에서의 절단 빈도)를 측정하여 비교 대상의 절단 정도와 비교하는
단계(d-1)를 추가로 포함할 수 있으며, 이 경우, 비표적 위치에서의 절단
30 정도가 낮을수록 교정 효율 및/또는 정확도가 높다고 판단할 수 있다.
상기 비교 대상은 임의의 표적 DNA 의 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.

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

유전체 교정 및/또는 유전자 교정 기술은, 인간 세포를 비롯한 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 생체 또는 세포로부터 분리된 유전체 (genomic) DNA를 표적 특이적 뉴클레아제로 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 분리된 유전체 DNA 는 크로마틴 DNA 를 포함하는 것을 특징으로 한다. 본 명세서에서 크로마틴 DNA 라 함은 세포 (또는 핵)에서 히스톤, 비히스톤 단백질, RNA 등의 비-DNA 염색질 성분이 제거되지 않은, 즉, 5 상기한 비-DNA 염색질 성분들 중에서 선택된 하나 이상을 함께 포함하는 형태의 DNA 를 의미한다. 상기 크로마틴 DNA 는 세포의 세포막 제거 후 원심분리하여 얻어지는 세포질 층과 크로마틴 DNA 층 중에서 세포질 층을 제거하거나 크로마틴 DNA 층을 취하여 얻어질 수 있다 (도 1 참조). 상기 10 세포막 제거는 세포에 통상의 용해 버퍼 (lysis buffer)를 처리하여 수행될 수 있으나, 이에 제한되는 것은 아니다. 본 명세서에 기재된 분리된 유전체 DNA 는, 크로마틴으로부터 DNA 만 분리된 상태는 배제하는 의미일 수 있으며, 예컨대 세포 용해물에서 세포질을 제거하여 얻어지거나, 원심분리에 의하여 생성된 크로마틴 DNA 층을 취하여 얻어진 크로마틴 DNA 15 또는 크로마틴 DNA 과 세포질을 포함하는 것일 수 있다.

본 명세서에 사용된 바로서, 표적 특이적 뉴클레아제는, 유전자 가위 (programmable nuclease)라고도 불리며, 목적하는 유전체 DNA 상의 특정 위치를 인식하여 절단할 수 있는 모든 형태의 뉴클레아제 (예컨대, 엔도뉴클레아제)를 통칭한다.

20 예컨대, 상기 표적 특이적 뉴클레아제는 표적 유전자의 특정 서열을 인식하고 뉴클레오티드 절단 활성을 가져 표적 유전자에서 인델 (insertion and/or deletion, Indel)을 야기할 수 있는 모든 뉴클레아제에서 선택된 1 종 이상일 수 있다.

예컨대, 상기 표적 특이적 뉴클레아제는 25 유전체 상의 특정 표적 서열을 인식하는 도메인인 식물 병원성 유전자에서 유래한 TAL 작동자 (transcription activator-like effector) 도메인과 절단 도메인이 융합된 TALEN (transcription activator-like effector nuclease);

30 징크-핑거 뉴클레아제 (zinc-finger nuclease);

메가뉴클레아제 (megannuclease);

미생물 면역체계인 CRISPR 에서 유래한 RGEN (RNA-guided engineered nuclease; 예컨대, Cas 단백질 (예컨대, Cas9 등), Cpf1, 등);

아고 호몰로그 (Ago homolog, DNA-guided endonuclease)

등으로 이루어진 군에서 선택된 1 종 이상일 수 있으나, 이에 제한되는 것은 아니다.

상기 표적 특이적 뉴클레아제는 원핵 세포, 및/또는 인간 세포를 비롯한 동식물 세포 (예컨대, 진핵 세포)의 유전체에서 특정 염기서열을 인식해 이중나선절단 (double strand break, DSB)을 일으킬 수 있다. 상기 이중나선절단은 DNA 의 이중 나선을 잘라, 둔단 (blunt end) 또는 점착종단 (cohesive end)을 생성시킬 수 있다. DSB 는 세포 내에서 상동재조합 (homologous recombination) 또는 비상동재접합 (non-homologous end-joining, NHEJ) 기작에 의해 효율적으로 수선될 수 있는데, 이 과정에 소망하는 변이를 표적 위치에 도입할 수 있다.

상기 메가뉴클레아제는 이에 제한되는 것은 아니나, 자연-발생 메가뉴클레아제일 수 있고 이들은 15 - 40 개 염기쌍 절단 부위를 인식하는데, 이는 통상 4 개의 패밀리로 분류된다: LAGLIDADG 패밀리, GIY-YIG 패밀리, His-Cyst 박스 패밀리, 및 HNH 패밀리. 예시적인 메가뉴클레아제는 I-SceI, I-CeuI, PI-PspI, PI-SceI, I-SceIV, I-CsmI, I-PanI, I-SceII, I-PpoI, I-SceIII, I-CreI, I-TevI, I-TevII 및 I-TevIII 를 포함한다.

자연-발생 메가뉴클레아제, 주로 LAGLIDADG 패밀리로부터 유래하는 DNA 결합 도메인을 이용하여 식물, 효모, 초파리 (*Drosophila*), 포유동물 세포 및 마우스에서 위치-특이적 게놈 변형이 촉진되었으나, 이런 접근법은 메가뉴클레아제 표적 서열이 보존된 상동성 유전자의 변형 (Monet et al. (1999) *Biochem. Biophysics. Res. Common.* 255: 88-93)으로, 표적 서열이 도입되는 사전-조작된 게놈의 변형에는 한계가 있었다. 따라서, 의학적으로나 생명공학적으로 관련된 부위에서 신규한 결합 특이성을 나타내도록 메가뉴클레아제를 조작하려는 시도가 있었다. 또한, 메가뉴클레아제로부터 유래하는 자연-발생된 또는 조작된 DNA 결합 도메인이 이중성 뉴클레아제 (예, FokI)로부터 유래하는 절단 도메인에 작동 가능하게 연결되었다.

상기 ZFN 은 선택된 유전자, 및 절단 도메인 또는 절단 하프-도메인의 표적 부위에 결합하도록 조작된 징크-핑거 단백질을 포함한다. 상기 ZFN 은 징크-핑거 DNA 결합 도메인 및 DNA 절단 도메인을 포함하는 인공적인 제한효소일 수 있다. 여기서, 징크-핑거 DNA 결합 도메인은 선택된 서열에 결합하도록 조작된 것일 수 있다. 예를 들면, Beerli et al. (2002) *Nature Biotechnol.* 20:135-141; Pabo et al. (2001) *Ann. Rev.*

Biochem. 70:313-340; Isalan et al, (2001) Nature Biotechnol. 19: 656-660; Segal et al. (2001) Curr. Opin. Biotechnol. 12:632-637; Choo et al. (2000) Curr. Opin. Struct. Biol. 10:411-416 이 본 명세서
 5 참고자료로서 포함될 수 있다. 자연 발생된 징크 핑거 단백질과 비교하여, 조작된 징크 핑거 결합 도메인은 신규한 결합 특이성을 가질 수 있다. 조작 방법은 합리적 설계 및 다양한 타입의 선택을 포함하나 이에 국한되지는 않는다. 합리적 설계는, 예를 들어 삼중 (또는 사중) 뉴클레오티드 서열, 및 개별 징크 핑거 아미노산 서열을 포함하는 데이터베이스의 이용을 포함하며, 이때 각 삼중 또는 사중 뉴클레오티드 서열은 특정 삼중 또는
 10 사중 서열에 결합하는 징크 핑거의 하나 이상의 서열과 연합된다.

표적 서열의 선택, 융합 단백질 (및 그것을 암호화하는 폴리뉴클레오티드)의 설계 및 구성은 당업자에 공지되어 있으며, 참고자료로 미국특허출원 공개 2005/0064474 및 2006/0188987 의 전문에 상세하게 설명되며, 상기 공개특허의 전문이 본 발명의 참고자료로서 본
 15 명세서에 포함된다. 또한, 이러한 참고문헌 및 당업계의 다른 문헌에 개시된 대로, 징크 핑거 도메인 및/또는 다중-핑거 징크 핑거 단백질들이 임의의 적절한 링커 서열, 예를 들면 5 개 이상의 아미노산 길이의 링커를 포함하는 링커에 의해 함께 연결될 수 있다. 6 개 이상의 아미노산 길이의 링커 서열의 예는 미국등록특허 6,479,626; 6,903,185; 7,153,949 을
 20 참고한다. 여기 설명된 단백질들은 단백질의 각 징크 핑거 사이에 적절한 링커의 임의의 조합을 포함할 수 있다.

또한, ZFN 과 같은 뉴클레아제는 뉴클레아제 활성 부분 (절단 도메인, 절단 하프-도메인)을 포함한다. 주지된 대로, 예를 들면 징크 핑거 DNA 결합 도메인과 상이한 뉴클레아제로부터의 절단 도메인과 같이, 절단
 25 도메인은 DNA 결합 도메인에 이중성일 수 있다. 이중성 절단 도메인은 임의의 엔도뉴클레아제나 엑소뉴클레아제로부터 얻어질 수 있다. 절단 도메인이 유래할 수 있는 예시적인 엔도뉴클레아제는 제한 엔도뉴클레아제 및 메가뉴클레아제를 포함하나 이에 한정되지는 않는다.

유사하게, 절단 하프-도메인은, 상기 제시된 바와 같이, 절단 활성을
 30 위하여 이량체화를 필요로 하는 임의의 뉴클레아제 또는 그것의 일부로부터 유래될 수 있다. 융합 단백질이 절단 하프-도메인을 포함하는 경우, 일반적으로 2 개의 융합 단백질이 절단에 필요하다. 대안으로, 2 개의 절단 하프-도메인을 포함하는 단일 단백질이 이용될 수도 있다. 2 개의 절단 하프-도메인은 동일한 엔도뉴클레아제 (또는 그것의 기능적 단편들)로부터

유래할 수도 있고, 또는 각 절단 하프-도메인이 상이한 엔도뉴클레아제 (또는 그것의 기능적 단편들)로부터 유래할 수도 있다. 또한, 2 개의 융합 단백질의 표적 부위는, 2 개의 융합 단백질과 그것의 각 표적 부위의 결합에 의해 절단-하프 도메인들이 서로에 대해 공간적으로 배향되어
 5 위치됨으로써, 절단 하프-도메인이, 예를 들어 이량체화에 의해 기능성 절단 도메인을 형성할 수 있도록 하는 관계로 배치되는 것이 바람직하다. 따라서, 일 구현예에서, 3 - 8 개 뉴클레오티드 또는 14 - 18 개 뉴클레오티드에 의해 표적 부위의 이웃 가장자리가 분리된다. 그러나, 임의의 정수의 뉴클레오티드 또는 뉴클레오티드 쌍이 2 개의 표적 부위
 10 사이에 개재될 수 있다 (예, 2 내지 50 개 뉴클레오티드 쌍 또는 그 이상). 일반적으로, 절단 부위는 표적 부위 사이에 놓인다.

제한 엔도뉴클레아제 (제한 효소)는 많은 종에 존재하며, DNA 에 서열-특이적으로 결합하여(표적 부위에서), 바로 결합 부위나 그 근처에서 DNA 를 절단할 수 있다. 어떤 제한 효소 (예, Type IIS)는 인식 부위로부터
 15 제거된 부위에서 DNA 를 절단하며, 분리 가능한 결합과 절단 가능한 도메인을 가진다. 예를 들면, Type IIS 효소 FokI 은 한 가닥 상의 인식 부위로부터 9 개 뉴클레오티드에서 그리고 나머지 한 가닥 상의 인식 부위로부터 13 개 뉴클레오티드에서 DNA 의 이중가닥 절단을 촉매한다. 따라서, 한 구현예에서, 융합 단백질은 최소 1 개의 Type IIS 제한
 20 효소로부터의 절단 도메인 (또는 절단 하프-도메인)과 하나 이상의 아연-핑거 결합 도메인 (조작될 수도 있고 그렇지 않을 수도 있는)을 포함한다.

"TALEN"은 DNA 의 타겟 영역을 인식 및 절단할 수 있는 뉴클레아제를 가리킨다. TALEN 은 TALE 도메인 및 뉴클레오티드 절단 도메인을 포함하는 융합 단백질을 가리킨다. 본 발명에서, "TAL 이펙터 뉴클레아제" 및
 25 "TALEN"이라는 용어는 호환이 가능하다. TAL 이펙터는 크산토모나스 (Xanthomonas) 박테리아가 다양한 식물 종에 감염될 때 이들의 타입 III 분비 시스템을 통해 분비되는 단백질로 알려져 있다. 상기 단백질은 숙주 식물 내의 프로모터 서열과 결합하여 박테리아 감염을 돕는 식물 유전자의 발현을 활성화시킬 수 있다. 상기 단백질은 34 개 이하의 다양한 수의
 30 아미노산 반복으로 구성된 중심 반복 도메인을 통해 식물 DNA 서열을 인식한다. 따라서, TALE 은 게놈 엔지니어링의 도구를 위한 신규 플랫폼이 될 수 있을 것으로 여겨진다. 다만 게놈-교정 활성을 갖는 기능 TALEN 을 제작하기 위해서 다음과 같이 현재까지 알려지지 않았던 소수의 주요 매개변수가 정의되어야 한다. i) TALE 의 최소 DNA-결합 도메인, ii) 하나의

타겟 영역을 구성하는 2 개의 절반-자리 사이의 스페이서의 길이, 및 iii) FokI 뉴클레아제 도메인을 dTALE 에 연결하는 링커 또는 융합 접합 (fusion junction).

본 발명의 TALE 도메인은 하나 이상의 TALE-반복 모듈을 통해 서열-
 5 특이적 방식으로 뉴클레오티드에 결합하는 단백질 도메인을 가리킨다. 상기 TALE 도메인은 적어도 하나의 TALE-반복 모듈, 보다 구체적으로는 1 내지 30 개의 TALE-반복 모듈을 포함하나 이에 한정되지 않는다. 본 발명에서, "TAL 이펙터 도메인" 및 "TALE 도메인"이라는 용어는 호환가능하다. 상기 TALE 도메인은 TALE-반복 모듈의 절반을 포함할 수 있다. 상기 TALEN 과
 10 관련하여 국제공개특허 WO/2012/093833 호 또는 미국공개특허 2013-0217131 호에 개시된 내용 전문이 본 명세서에 참고자료로서 포함된다.

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*
 15 and *Francisella* 1) 등과 같은 타입 II 및/또는 타입 V 의 CRISPR 시스템에 수반되는 뉴클레아제 (예컨대, 엔도뉴클레아제) 등으로 이루어진 군에서 선택된 1 종 이상일 수 있다. 이 경우, 상기 표적 특이적 뉴클레아제는 유전체 DNA 의 표적 부위로 안내하기 위한 표적 DNA 특이적 가이드 RNA 를 추가로 포함할 수 있다. 상기 가이드 RNA 는 생체 외 (in vitro)에서
 20 전사된(transcribed) 것일 수 있고, 예컨대 올리고뉴클레오티드 이중가닥 또는 플라스미드 주형으로부터 전사된 것일 수 있으나, 이에 제한되지 않는다. 상기 표적 특이적 뉴클레아제는 가이드 RNA 에 결합된 리보핵산-단백질 복합체를 형성(RNA-Guided Engineered Nuclease)하여 리보핵산 단백질 (RNP) 형태로 작용할 수 있다.

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

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

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

캠필로박터 속, 예컨대, 캠필로박터 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;

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

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

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

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

Cpf1 단백질은 상기 CRISPR/Cas 시스템과는 구별되는 새로운 CRISPR 시스템의 엔도뉴클레아제로서, Cas9 에 비해 상대적으로 크기가 작고 tracrRNA 가 필요 없으며, 단일 가이드 RNA 에 의해 작용할 수 있다. 또한, 티민 (thymine)이 풍부한 PAM (protospacer-adjacent motif) 서열을 인식하고 DNA 의 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 엔도뉴클레아제는 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 엔도뉴클레아제 (예컨대, Cas9, Cpf1, 등)은 재조합 DNA 에 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 DNA 를 적절한 유기체에서 발현시켜 표적 특이적 엔도뉴클레아제를 생산 (*in vivo* 또는 *in vitro*)하는 경우, 재조합 DNA 는 제조하고자 하는 단백질을 암호화 하는 코돈들 중에서 상기 유기체에 발현하기에 최적화된 코돈을 선택하여 재구성된 뉴클레오타이드 서열을 갖는 것일 수 있다.

상기 불활성화된 표적특이적 엔도뉴클레아제불활성화된 표적특이적 엔도뉴클레아제는 DNA 이중 가닥을 절단하는 엔도뉴클레아제 활성을 상실한 표적특이적 엔도뉴클레아제를 의미하는 것으로, 예컨대, 엔도뉴클레아제 활성을 상실하고 니케이즈 활성을 갖는 불활성화된 표적특이적 엔도뉴클레아제 및 엔도뉴클레아제 활성과 니케이즈 활성을 모두 상실한 불활성화된 표적특이적 엔도뉴클레아제 중에서 선택된 1 종 이상일 수 있다. 상기 불활성화된 표적특이적 엔도뉴클레아제가 니케이즈 활성을 갖는 것인 경우, 상기 시토신이 우라실로 변환되는 것과 동시 또는 순서와 무관하게 순차적으로, 시토신이 우라실로 변환된 가닥 또는 그 반대 가닥 (예컨대 반대 가닥)에서 nick 이 도입된다 (예컨대, PAM 서열의 5' 말단 방향으로 3 번째 뉴클레오타이드와 4 번째 뉴클레오타이드 사이에 nick 이 도입됨). 이와 같은 표적특이적 엔도뉴클레아제의 변형(돌연변이)는 적어도 촉매 활성을 갖는 아스파르트산 잔기 (catalytic aspartate residue; 예컨대, 스트렙토코커스 피요젠스 유래 Cas9 단백질의 경우 10 번째 위치의 아스파르트산 (D10) 잔기 등)가 임의의 다른 아미노산으로 치환된 Cas9 의 돌연변이를 포함하는 것일 수 있으며, 상기 다른 아미노산은 알라닌 (alanine)일 수 있지만, 이에 제한되지 않는다.

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 아미노산을 제외한 아미노산들 중에서 선택된 아미노산을 의미한다.

일 예에서, 상기 불활성화된 표적특이적 엔도뉴클레아제가 변형 Cas9 단백질인 경우, 변형 Cas9 단백질은 스트렙토코커스 피요젠스 (*Streptococcus pyogenes*) 유래의 Cas9 단백질(예컨대, SwissProt Accession number Q99ZW2(NP_269215.1))에 D10 위치에서의 돌연변이 (예컨대, 다른 아미노산으로의 치환)가 도입되어 엔도뉴클레아제 활성이 상실되고 니케이즈 활성을 갖는 변형 Cas9, 스트렙토코커스 피요젠스 (*Streptococcus pyogenes*) 유래의 Cas9 단백질에 D10 위치에서의 돌연변이(예컨대, 다른 아미노산으로의 치환)와 H840 위치에 돌연변이 (예컨대, 다른 아미노산으로의 치환)가 모두 도입되어 엔도뉴클레아제 활성 및 니케이즈 활성을 모두 상실한 변형 Cas9 단백질 등으로 이루어진 군에서 선택된 1 종 이상일 수 있다. 예컨대, 상기 Cas9 단백질의 D10 위치에서의 돌연변이는 D10A 돌연변이 (Cas9 단백질의 아미노산 중 10 번째 아미노산인 D가 A로 치환된 돌연변이를 의미함; 이하, Cas9에 도입된 돌연변이는 동일한 방법으로 표기됨)일 수 있고, 상기 H840 위치에서의 돌연변이는 H840A 돌연변이일 수 있다.

본 명세서에서, "뉴클레아제"는, 다른 언급이 없는 한, 앞서 설명된, 예컨대, Cas9, Cpf1, 등과 같은 "표적 특이적 뉴클레아제 (엔도뉴클레아제)"를 의미한다.

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 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 단백질 (엔도뉴클레아제)를 생산 (*in vivo* 또는 *in vitro*)하는 경우, 재조합 DNA는 제조하고자 하는 단백질을 암호화하는 코돈들 중에서 상기 유기체에 발현하기에 최적화된 코돈을 선택하여 재구성된 뉴클레오타이드 서열을 갖는 것일 수 있다.

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

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 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 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 서열을 포함할 수 있다. 상기 NLS 서열은 당업계에 잘 알려져 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 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, DNA 의 표적 부위와 혼성화 가능한 RNA)를 의미하며, Cas 단백질, Cpf1 등과 같은 뉴클레오타이드와 결합하여 표적 DNA 로 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).

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.

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.

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.

- 5 일 예에서, 표적특이적 엔도뉴클레아제로서 *Streptococcus pyogenes* 유래의 Cas9 단백질을 사용하는 경우, crRNA 는 다음의 일반식 1 로 표현될 수 있다:



상기 일반식 1 에서,

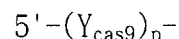
- 10 N_{cas9} 는 표적화 서열, 즉 표적 유전자(target gene)의 표적 부위(target site)의 서열에 따라서 결정되는 부위 (즉, 표적 부위의 서열과 혼성화 가능한 서열임)이며, 1 은 상기 표적화 서열에 포함된 뉴클레오타이드 수를 나타내는 것으로 17 내지 23 또는 18 내지 22 의 정수, 예컨대 20 일 수 있고;

- 15 상기 표적 서열의 3' 방향으로 인접하여 위치하는 연속하는 12 개의 뉴클레오타이드(GUUUUAGAGCUA) (서열번호 1)를 포함하는 부위는 crRNA 의 필수적 부분이고,

- 20 X_{cas9} 는 crRNA 의 3' 말단쪽에 위치하는 (즉, 상기 crRNA 의 필수적 부분의 3' 방향으로 인접하여 위치하는) m 개의 뉴클레오타이드를 포함하는 부위로, m 은 8 내지 12 의 정수, 예컨대 11 일 수 있으며, 상기 m 개의 뉴클레오타이드들은 서로 같거나 다를 수 있으며, 각각 독립적으로 A, U, C 및 G 로 이루어진 군에서 선택될 수 있다.

일 예에서, 상기 X_{cas9} 는 UGCUGUUUUG (서열번호 2)를 포함할 수 있으나 이에 제한되지 않는다.

- 25 또한, 상기 tracrRNA 는 다음의 일반식 2 로 표현될 수 있다:



(일반식 2)

상기 일반식 2 에서,

- 30 60 개의 뉴클레오타이드
(UAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC)
(서열번호 3)로 표시된 부위는 tracrRNA 의 필수적 부분이고,

Y_{cas9} 는 상기 tracrRNA 의 필수적 부분의 5' 말단에 인접하여 위치하는 p 개의 뉴클레오타이드를 포함하는 부위로, p 는 6 내지 20 의 정수,

예컨대 8 내지 19 의 정수일 수 있으며, 상기 p 개의 뉴클레오타이드들은 서로 같거나 다를 수 있고, A, U, C 및 G 로 이루어진 군에서 각각 독립적으로 선택될 수 있다.

- 또한, sgRNA 는 상기 crRNA 의 표적화 서열과 필수적 부위를 포함하는 crRNA 부분과 상기 tracrRNA 의 필수적 부분 (60 개 뉴클레오타이드)를 포함하는 tracrRNA 부분이 올리고뉴클레오타이드 링커를 통하여 헤어핀 구조 (stem-loop 구조)를 형성하는 것일 수 있다 (이 때, 올리고뉴클레오타이드 링커가 루프 구조에 해당함). 보다 구체적으로, 상기 sgRNA 는 crRNA 의 표적화 서열과 필수적 부분을 포함하는 crRNA 부분과 tracrRNA 의 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 있다.

일 예에서, sgRNA 는 다음의 일반식 3 으로 표현될 수 있다:

- 15 5'-(N_{Cas9})₁-(GUUUUAGAGCUA)-(올리고뉴클레오타이드 링커)-(UAGCAAGUUAUUUUAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC)-3' (일반식 3)

상기 일반식 3 에서, (N_{Cas9})₁ 는 표적화 서열로서 앞서 일반식 1 에서 설명한 바와 같다.

- 20 상기 sgRNA 에 포함되는 올리고뉴클레오타이드 링커는 3 내지 5 개, 예컨대 4 개의 뉴클레오타이드를 포함하는 것일 수 있으며, 상기 뉴클레오타이드들은 서로 같거나 다를 수 있고, A, U, C 및 G 로 이루어진 군에서 각각 독립적으로 선택될 수 있다.

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

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

- 상기 가이드 RNA 의 표적 서열은 표적 DNA 상의 PAM (Protospacer Adjacent Motif 서열(*S. pyogenes* Cas9 의 경우, 5'-NGG-3' (N 은 A, T, G, 또는 C 임))의 5'에 인접하여 위치하는 약 17 개 내지 약 23 개 또는 약 18 개 내지 약 22 개, 예컨대 20 개의 연속하는 핵산 서열일 수 있다.

상기 가이드 RNA 의 표적 서열과 혼성화 가능한 가이드 RNA 의 표적화 서열은 상기 표적 서열이 위치하는 DNA 가닥 (즉, PAM 서열(5'-NGG-3' (N 은

A, T, G, 또는 C 임)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
5 상보적 결합이 가능하다.

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 RNA 가 결합하는 DNA 가닥은 PAM 서열이 위치하는 가닥의 상보적 가닥이므로, 상기 가이드 RNA 에 포함된 표적화
10 서열은, RNA 특성상 T 를 U 로 변경하는 것을 제외하고, 표적 부위의 서열과 동일한 핵산 서열을 갖게 된다. 따라서, 본 명세서에서, 가이드 RNA 의 표적화 서열과 표적 부위의 서열 (또는 절단 부위의 서열)은 T 와 U 가 상호 변경되는 것을 제외하고 동일한 핵산 서열로 표시된다.

상기 가이드 RNA 는 RNA 형태로 사용 (또는 상기 조성물에
15 포함)되거나, 이를 암호화하는 DNA 를 포함하는 플라스미드 형태로 사용 (또는 상기 조성물에 포함)될 수 있다.

본 발명에서 용어 "표적 위치 (on-target site)" 란, 상기 표적 특이적 뉴클레아제를 이용하여 변이(절단, 삽입, 뺄/또는 결실)를 도입하고자 하는 위치를 의미하며, 그 목적에 따라 임의로 선택될 수 있는
20 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.

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
25 아니라 비표적 위치가 될 가능성이 있는 위치까지 포함하는 개념으로 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 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.
- 5 상기 비표적 위치는 이에 제한되는 것은 아니나, 표적 서열과 1 이상의 뉴클레오티드 불일치 (mismatch)를 가지는 위치일 수 있다.

- 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
- 10 비표적 효과 없이 표적 위치에만 특이적으로 작동하는 표적 특이적 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
- 15 유전체 DNA 의 비표적 위치를 검출하는데 사용될 수 있으며, 이를 생체 내에서 적용하였을 때 상기 검출된 비표적 위치와 동일한 위치에도 활성을 가질 것을 예상할 수 있다.

- 상기 (b) 단계는 상기 (a) 단계를 통해 절단된 DNA 를 이용해 전체 유전체 시퀀싱 (whole genome sequencing, WGS)을 수행하는 단계로서, 표적
- 20 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
- 25 반응 (hybridization)에 기초하여 초고속으로 시퀀싱을 수행하는 기술을 의미한다.

상기 (c) 단계는 상기 전체 유전체 시퀀싱으로 수득한 염기서열 데이터 (sequence read)에서 DNA 가 절단된 위치를 결정하는 단계로서, 시퀀싱 데이터를 분석하여 표적 특이적 뉴클레아제의 표적 위치 및 비표적 위치를 간편하게 검출할 수 있다. 상기 염기서열 데이터로부터 DNA 가

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.

- 5 예컨대, 상기 절단된 위치를 결정하기 위한 일례로서, 전체 유전체 시퀀싱을 통해 수득한 염기서열 데이터를 분석 프로그램 (예를 들어, BWA/GATK 또는 ISAAC 등)을 이용하여 유전체 상의 위치에 따라 정렬하였을 경우, 5' 말단이 수직 정렬된 위치가 DNA 가 절단된 위치를 의미할 수 있다. 본 명세서에 사용된 바로서, 상기 용어 "수직 정렬"이란, BWA/GATK 또는
- 10 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
- 15 시퀀싱되어 나타나게 된다.

- 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
- 20 않으므로 정렬 시 스테거드 (staggered) 방식으로 배열될 수 있다. 따라서, 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
- 25 데이터를 맵핑한 뒤, 유전체에서 동일 위치를 가지는 염기들을 각 위치에 맞게 배열하는 것을 의미한다. 따라서, 염기서열 데이터를 상기와 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 ISAAC 를
- 30 이용하여 정렬을 수행하였으나, 이에 제한되는 것은 아니다.

 정렬 결과, 상기 설명한 바와 같은 5' 말단이 수직 정렬된 위치를 찾는 등의 방법을 통해 표적 특이적 뉴클레아제에 의해 DNA 가 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 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, 1 개 내지 5 개, 1 개 내지 4 개, 1 개 내지 3 개, 1 개 내지 2 개, 또는 1 개의 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 RNA 와 상보적인 15 내지 30 뉴클레오티드 서열일 수 있고, 추가적으로 표적 특이적 뉴클레아제가 인식하는 서열 (예컨대, Cas9 의 경우 Cas9 이 인식하는 PAM 서열)을 포함할 수 있다.

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 각각의 가닥에 의해 나타나는 것이기 때문이다.

일 실시예에서, 유전체 DNA 를 표적 특이적 뉴클레아제(예컨대, RGEN)으로 절단한 뒤, 전체 유전체 분석 후 이를 ISAAC 로 정렬하여 절단된 위치에서는 수직 정렬, 절단되지 않은 위치에서는 스테거드 방식으로 정렬되는 패턴을 확인하였으며, 이를 5' 말단 플롯으로 나타내었을 때 절단 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 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 DNA 에 처리하여 수행될 수 있다. 이에 상기

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.

- 5 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 의미하는 것으로, 상기 표적 특이적 뉴클레아제에 의해 절단되는 위치를
- 10 말하는 것이나, 비표적 효과는 세포 내 비표적 위치에서 표적 특이적 뉴클레아제에 의해 인델 (Insertion and/or deletion)이 나타나는 효과를 의미한다.

- 상기 용어 "인델"은 DNA 의 염기 배열에서 일부 염기가 중간에 삽입되거나 (insertion) 및/또는 결실된 (deletion) 변이를 총칭한다.
- 15 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 유전자 가위에 의한 인델이 확인되어야 하는 것은 아니다. 한편, 본
- 20 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 비표적 위치에 대한 표적 특이적 뉴클레아제가 발현된 세포로부터 유전체
- 25 DNA 를 분리하고, 상기 DNA 의 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 확인하는 것일 수 있다. 상기 비표적 효과를 확인하는 단계는 비표적
- 30 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.

- 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 기반으로 하여, 각 뉴클레오타이드의 위치 i (즉 유전체 DNA 상의 뉴클레오타이드 위치)에 DNA 절단 점수를 산출하는 수식을 다음과 같이 제공한다 (도 4 참조):

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

- 10 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 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 3 일 수 있으나, 이에 제한되는 것은 아니다. 특히, 이에 제한되는 것은 아니나, 예컨대, 임의의 위치 (절단된 위치)의 염기서열에 있어서, C 값을 1 으로 하여 상기 산출된 점수가 2.5 점 이상이거나, 0.1 점 이상이며 On-target 서열과 homology 를 갖는 경우 (예컨대, On-target 서열과 10 개 이하의 미스매치를 가지고 PAM (5'-NGN-3' 또는 5'-NNG-3')을 포함하는 경우), 상기 임의의 위치(절단된

위치)를 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 RNA 를 다수 포함하는 표적 특이적 뉴클레아제)를 이용하여 수행될 수도 있으며, 본 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-seq 의 경우 각각의 표적 특이적 뉴클레아제에 의해 유전체 DNA 가 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-seq 를 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 RGEN 의 표적 위치로 선별하는 것이 비표적 효과를 최소화할 수 있음을 확인하였다. 이는 본 발명의 Digenome-seq 를 이용하여 표적 위치를 선별하는 바람직한 기준을 확립해가는 과정을 보여주는 일례로서, Digenome-seq 를 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-seq 를 통해 적은 비율로 검출되는 것을 확인하였다. 이는 RGEN 의 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 RGEN 의 표적 위치는 비표적 효과가 최소화된 것일 수 있다.

다른 예는 상기 표적 특이적 뉴클레아제의 절단 위치 또는 비표적 위치 (off-target site)를 검출하는 방법을 사용하여 비표적 위치가 적은 표적 부위 및/또는 상기 표적 부위를 표적으로 하는 표적 특이적 뉴클레아제를 선별하는 방법을 제공한다.

5 상기 선별 방법은,

(a) 분리된 유전체 (genomic) DNA 를 표적 특이적 뉴클레아제로 절단하는 단계;

(b) 상기 절단된 DNA 에 대한 전체 유전체 시퀀싱 (whole genome sequencing)을 수행하는 단계;

10 (c) 상기 시퀀싱으로 획득한 염기서열 데이터 (sequence read)에서 상기 절단된 위치를 확인하는 단계;

(d) 상기 확인된 절단된 위치가 표적 위치 (on-target site)가 아닌 경우, 비표적 위치 (off-target site)로 판단하는 단계; 및

(e) 상기 판단된 비표적 위치 데이터를 분석하여, 비표적 위치가
15 비교 대상보다 적게 나타난 경우 상기 표적 특이적 뉴클레아제가 표적으로 하는 표적 부위 및/또는 이를 표적으로 표적 특이적 뉴클레아제를 선택하는 단계

를 포함할 수 있다.

일 예에서, 상기 표적 특이적 뉴클레아제가 가이드 RNA에 의하여
20 표적 부위로 표적화되는 것인 경우, 상기 선별된 표적 부위 및/또는 표적 특이적 뉴클레아제는 상기 가이드 RNA의 표적 부위의 표적 DNA 서열과 혼성화 가능한 서열을 포함하는 것으로 특징화될 수 있다.

상기 비교 대상은 임의의 표적 DNA의 표적 서열에 대한 표적 특이적 뉴클레아제일 수 있으며, 일 예에서, 통상적으로 사용되거나 이미 알려진
25 표적 특이적 뉴클레아제 (예컨대, RGEN 및 가이드 RNA 조합)들 중 선택된 어느 하나일 수 있다.

【발명의 효과】

본 발명의 Digenome-seq 는 고도의 재현성으로 유전체 수준에서
30 유전자 가위의 비표적 위치를 검출할 수 있어, 표적 특이성이 높은 유전자 가위의 제작 및 이를 위한 연구에 사용될 수 있다.

【도면의 간단한 설명】

도 1은 크로마틴 DNA를 이용한 절단 유전체 시퀀싱 방법을

모식적으로 보여준다.

도 2a 내지 2c는 크로마틴 DNA와 크로마틴이 없는 DNA를 이용한 절단 유전체 시퀀싱 결과를 비교하여 보여주는 것으로,

도 2a는 크로마틴 단백질이 없는 DNA (분리된 DNA)에 RNA 유전자
5 가위를 처리하지 않고 전유전체를 시퀀싱한 경우(위), 크로마틴 단백질이 없는 DNA (분리된 DNA)에 RNA 유전자 가위를 처리한 후 전유전체를 시퀀싱한 경우 (중간), 및 크로마틴 DNA (크로마틴 단백질을 포함)에 RNA 유전자 가위를 처리한 후 전유전체를 시퀀싱한 경우 (아래)에서의 DNA 절단 점수 (DNA cleavage score)를 보여주는 그래프로서, 진한 색 막대 부분은
10 절단 유전체 시퀀싱 결과 측정 DNA 절단 점수(0.0001 - 10)를 넘는 위치의 전체 수를 나타내고, 연한 색 막대 부분은 이 중 Cas9이 결합하는 서열(NNG 또는 NGN)이 존재하고 20bp의 표적 위치의 서열과 10bp 이상의 서열이 일치하는 위치의 수를 나타내며;

도 2b는 크로마틴 단백질이 없는 DNA를 가지고 한 절단 유전체
15 시퀀싱과 크로마틴 단백질을 포함하는 DNA를 가지고 한 절단 유전체 시퀀싱에서 나온 비표적 위치의 수를 비교한 결과를 보여주는 밴다이어그램이고;

도 2c는 크로마틴 단백질이 없는 DNA를 가지고 한 절단 유전체
20 시퀀싱에서 나온 비표적 위치의 DNA 절단 점수와 크로마틴 단백질을 포함하는 DNA를 가지고 한 절단 유전체 시퀀싱에서 나온 비표적 위치의 절단 점수의 상관관계를 비교한 그래프이다.

도 3a 내지 3c는 크로마틴을 포함하는 DNA를 이용해 진행한 절단 유전체 시퀀싱 분석 결과를 보여주는 것으로,

도 3a는 세포 내에서의 돌연변이 비율과 크로마틴 단백질을 포함하는
25 DNA를 이용하여 절단 유전체를 시퀀싱한 경우의 DNA 절단 점수를 보여주는 그래프이고,

도 3b는 세포 내에서의 돌연변이 비율과 크로마틴 단백질이 없는 DNA를 이용하여 절단 유전체 시퀀싱을 했을 때의 DNA 절단 점수를 보여주는 그래프이며,

30 도 3c는 상기 도 3a와 3b에 나타난 돌연변이 비율(Indel 빈도)과 DNA 절단 점수와의 상관관계를 보여주는 그래프이다.

도 4는 Digenome-seq 분석에 대한 시험관 내 DNA 절단 점수 부여 시스템을 보여준다.

도 5a는 Hela cell에서 HBB 특이적 Cas9를 사용하여 Digenome 1.0에

의하여 확인된 비표적 위치 결과를 보여주는 밴다이어그램이다.

도 5b는 Hela cell에서 HBB 특이적 Cas9를 사용하여 Digenome 1.0에 의하여 확인된 비표적 위치에서의 핵 펠렛의 절단 점수와 native chromatin의 절단 점수의 상관관계를 보여주는 그래프이다.

- 5 도 6a는 Hela cell에서 HBB 특이적 Cas9를 사용하여 Digenome 2.0에 의하여 확인된 비표적 위치 결과를 보여주는 밴다이어그램이다.

도 6b는 Hela cell에서 HBB 특이적 Cas9를 사용하여 Digenome 2.0에 의하여 확인된 핵 펠렛 및 native chromatin에서의 비표적 위치의 절단 점수 간의 상관관계를 보여주는 그래프이다.

- 10 도 7a는 HEK293T cell에서 HBB 특이적 Cas9를 사용하여 Digenome 2.0에 의하여 확인된 비표적 위치 결과를 보여주는 밴다이어그램이다.

도 7b는 HEK293T cell에서 HBB 특이적 Cas9를 사용하여 Digenome 2.0에 의하여 확인된 비표적 위치에서의 핵 펠렛의 절단 점수와 native chromatin의 절단 점수의 상관관계를 보여주는 그래프이다.

- 15 도 8a는 Hela 세포의 핵 펠렛과 크로마틴이 제거된 DNA (chromatin free DNA)에 대하여 HBB 특이적 Cas9를 사용하여 Digenome 1.0를 수행하여 확인된 in vitro 절단 위치(비표적 위치)를 보여주는 밴다이어그램이다.

- 도 8b는 Hela cell에서 HBB 특이적 Cas9를 사용하여 Digenome 1.0에 의하여 확인된 핵 펠렛 및 크로마틴이 제거된 DNA에서의 비표적 위치의 절단 점수 간의 상관관계를 보여주는 그래프이다.도 9a는 Hela 세포의 핵 펠렛과 크로마틴이 제거된 DNA에 대하여 HBB 특이적 Cas9를 사용하여 Digenome 2.0를 수행하여 확인된 in vitro 절단 위치(비표적 위치)를 보여주는 밴다이어그램이다.
- 20

- 도 9b는 Digenome-capture sites에서의 DNA 서열들을 사용하여 WebLogo를 통해 얻은 서열 로고 (DNA 절단 점수 > 0.1)를 나타낸다.
- 25

도 9c는 Hela cell에서 HBB 특이적 Cas9를 사용하여 Digenome 2.0에 의하여 확인된 핵 펠렛 및 크로마틴이 제거된 DNA에서의 비표적 위치의 절단 점수 간의 상관관계를 보여주는 그래프이다.

- 도 9d는 크로마틴 DNA 및 크로마틴 제거 DNA에 대하여 HBB 표적 서열을 사용하여 Digenome-seq을 수행한 결과 얻어진 DNA 절단 점수와 Indel 빈도 사이의 상관관계를 보여주는 그래프이다.
- 30

도 9e는 크로마틴 DNA 및 크로마틴 제거 DNA에 대하여 HBB 표적 서열을 사용하여 Digenome-seq을 수행한 결과 얻어진 R^2 값을 보여주는 그래프이다.

【발명을 실시하기 위한 형태】

이하, 실시예를 통하여 본 발명을 더욱 상세히 설명하고자 한다. 이들 실시예는 오로지 본 발명을 보다 구체적으로 설명하기 위한 것으로서, 5 본 발명의 범위가 이들 실시예에 의해 제한되지 않는다는 것은 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에 있어 자명할 것이다.

실시예 1: 엔도뉴클레아제를 이용한 크로마틴 DNA 절단 (*in vitro* cleavage) 및 절단 DNA 의 서열 분석

10 유전체 DNA 절단 및 서열 분석에 사용하기 위하여, Hela (ATCC, CCL-2), HEK293T (ATCC, CRL-11268), K562 (ATCC, CCL-243), 및 T cell (공여자의 blood에서 추출)을 준비하였다. 상기 준비된 각각의 세포 (각각 5×10^5 개)에 Lysis buffer(1xPBS, 0.4% NP-40 (CAS 9016-45-9), 및 3mM $MgCl_2$)를 처리하여 세포막을 제거한 후, 500 x g에서 5분간 원심분리하여 15 세포질 층(상층액)과 크로마틴 DNA 층으로 분리하였다.

세포질 층을 제거한 후, 300nM 의 Cas9 단백질 (*Streptococcus pyogenes* 유래, SwissProt Accession number Q99ZW2; 서열번호 4)과 900nM 의 가이드 RNA 를 반응 buffer (100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 100 μ g/ml BSA, and pH 7.9)에 넣고 37°C에서 8 시간 동안 20 incubation 하였다. 상기 Cas9 및 가이드 RNA 처리에 의하여 DNA 절단이 일어난 반응물에 RNase A (50 μ g/mL)를 처리하여 가이드 RNA 를 제거하고 DNeasy Tissue kit (Qiagen)을 이용하여 전 유전체 DNA 를 정제하였다.

상기 분리된 전 유전체 DNA 를 1 μ g 의 양으로 취하여, Covaris system (Life Technologies)을 이용하여 400-500bp 정도로 절단한 후, 절단에 25 의하여 생성된 DNA 돌출부를 End Repair Mix 를 이용하여 제거하였다. 서열분석을 위한 라이브러리를 만들기 위하여, 상기 얻어진 절단된 DNA 를 어댑터 (절단된 DNA 에 결합하는 DNA 조각으로, Illumina 에서 제공하는 kit 를 이용함)와 결합을 시킨 후, HiSeq X Ten Sequencer (Macrogen)를 이용하여 전유전체 시퀀싱 (Whole genome sequencing)을 수행하였다.

30 본 실시예에 기재된 과정을 도 1에 모식적으로 나타내었다.

또한, 상기 사용된 가이드 RNA 는 다음의 뉴클레오타이드 서열을 갖는다:

5'-(표적 서열)-(GUUUUAGAGCUA; 서열번호 1)-(뉴클레오타이드 링커)-
(UAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC;
서열번호 3)-3'

(상기 표적 서열은 하기의 표 1 내지 표 8 에 기재된 DNA sequence
5 at target site (네모 박스로 표시하거나 on-target 으로 표시한 서열) 또는
Target sequence (on-target 서열)에서 3' 말단의 "NGG" (N 은 임의의
뉴클레오타이드로서, A, T, G, 또는 C 의 염기를 갖는 뉴클레오타이드임)를
제외한 뉴클레오타이드 서열에서 "T"를 "U"로 변환한 서열이며,

상기 뉴클레오타이드 링커는 GAAA 의 뉴클레오타이드 서열을 가짐.

10 이하 실시예에서, Cas9 단백질과 가이드 RNA 의 복합체를 유전자
가위라고 칭한다.

비교를 위하여, 상기 얻어진 크로마틴 DNA 에서 DNA 를 정제하여
얻어진 크로마틴 제거 DNA (chromatin free DNA)에 대하여 상기과 동일한
방법으로 유전체 절단 및 시퀀싱을 수행하였다. 상기 크로마틴 제거 DNA 는
15 상기 준비된 각각의 세포로부터 DNeasy tissue kit (Qiagen)를 이용하여
제조사의 사용설명서에 따라서 유전체 DNA 를 정제하여 준비하였다.

실시예 2: 크로마틴 DNA 와 크로마틴 제거 DNA 를 이용한 절단 유전체
서열 분석 결과 비교

20 실시예 1 에서 Hela, HEK293T, K562, 및 T cell 의 4 종의 세포에
대하여 얻어진 전 유전체 DNA 서열분석 결과를 이용하여, DNA 절단 점수
체계 (DNA Cleavage Scoring system)에 따라서 비표적 위치 (off-target
site)를 측정하였다 (절단 유전체 시퀀싱).

25 DNA 절단 점수 체계 (DNA Cleavage Scoring system)는 하기의 식에
따라서 구하였으며, 보다 상세한 사항은 "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) "를 참조로 하였다:

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 리드수)를 의미한다.

또한, 상기 C 값은 1로 하였다.

- 5 먼저, 세포 용해물에 아무 처리 없이 얻어진 유전체 (세포질+크로마틴)를 전 유전체 시퀀싱 한 것과 크로마틴을 포함하거나 제거된 상태의 DNA 에 유전자 가위를 처리하여 절단한 반응 결과물을 전 유전체 시퀀싱 한 결과를 이용하여 상기 방법으로 산출한 DNA 절단 점수를 서로 비교하였다. 이 중에서 표 1 의 조건 (Hela cell 및 표 1 의 target
- 10 sequence (HBB 표적 서열) 이용)에 해당하는 결과를 도 2a 내지 2c 에 나타내었다.

- 우선, 특정 DNA 절단 점수(0.0001 - 10)를 넘는 전체 위치의 수와 특정 DNA 절단 점수를 넘는 위치 중 Cas9 단백질이 결합하는 서열(NNG 또는 NGN; N 은 임의의 뉴클레오타이드로서, A, T(U), G, 또는 C 의 염기를 갖는
- 15 뉴클레오타이드임)이 존재하고 20bp 의 표적 위치의 서열과 10bp 이상의 서열이 일치하는 위치의 수를 구하여, 그 결과를 도 2a 에 나타내었다. 도 2a 는 크로마틴 제거 DNA(chromatin free DNA 로 표시)에서 유전자 가위를 처리하지 않고 전유전체 시퀀싱을 한 경우(상단), 크로마틴 제거 DNA 에서 유전자 가위를 처리한 후 전유전체 시퀀싱을 한 경우(중간), 및 크로마틴
- 20 DNA 에 유전자 가위를 처리한 경우(하단)에서의 DNA 절단 점수에 따른 위치 수 (number of sites)를 비교하여 보여주는 그래프이다. 도 2a 에서, 절단 유전체 시퀀싱 결과, 특정 DNA 절단 점수(0.0001 - 10)를 넘는 위치의 전체 수 (진한 색 막대)와 이 중 cas9 이 결합하는 서열(NNG 또는 NGN)이

존재하고 20bp 의 표적 위치의 서열과 10bp 이상의 서열이 일치하는 위치의 수(연한 색 막대)를 그래프로 나타내었다.

측정 DNA 절단 점수를 0.1 이상으로 하였을 때, Cas9 처리 없이 유전체를 전유전체 시퀀싱 한 것에서 나온 위치 중 Cas9 이 결합하는 서열(NNG 또는 NGN)이 존재하고 20bp 의 표적 위치와 서열과 10bp 이상의 서열이 일치하는 위치를 가지고 있는 곳이 없었기 때문에 DNA 절단 점수의 cutoff value 를 0.1 로 설정하였다.

이러한 기준을 통하여 크로마틴 DNA 에서 절단 유전체 시퀀싱을 한 경우 44 개의 비표적 절단 위치가 존재하였으며 이는 크로마틴 제거 DNA(chromatin free DNA)에서 절단 유전체 시퀀싱을 하였을 때 나온 119 개의 비표적 위치에 모두 포함이 되었다(도 2b 참조). 도 2b 는 밴다이어그램을 통하여 크로마틴 제거 DNA 에 대한 절단 유전체 시퀀싱과 크로마틴 DNA 에 대한 절단 유전체 시퀀싱 결과 얻어진 비표적 위치의 수를 비교하여 보여준다. 도 2b 는 크로마틴 DNA 를 사용하여 절단 유전체 시퀀싱을 수행함으로써 크로마틴을 제거 한 DNA 를 사용하는 경우와 비교하여 비표적 위치 수가 현저히 감소하여 보다 정확한 절단 유전체 시퀀싱이 가능함을 보여준다.

또한, 크로마틴 제거 DNA(chromatin free DNA)에 대한 절단 유전체 시퀀싱 결과 얻어진 비표적 위치의 DNA 절단 점수와 크로마틴 DNA 에 대한 절단 유전체 시퀀싱 결과 얻어진 비표적 위치의 절단 점수의 상관관계를 도 2c 에 나타내었다. 도 2c 에 나타난 바와 같이, 크로마틴 DNA 와 크로마틴 제거 DNA 각각에 대하여 절단 유전체 시퀀싱을 한 경우에 얻어진 DNA 절단 점수 사이의 상관관계를 보았을 때, 상관관계가 거의 없음을 볼 수 있다. 또한, 크로마틴 DNA 를 이용한 절단 유전체 시퀀싱을 Hela 세포 이외에도, HEK293T, K562, 및 T cell 과 같은 다양한 세포에 수행하여, 이들 세포에서도 적용 가능함을 입증하였다 (표 1 (표적서열들은 순서대로, 서열번호 5 내지 48, 이 중에서 서열번호 5 가 HBB on-target sequence 이며, 나머지는 off-target sequence 임) 및 표 2 (표적서열들은, 네모 박스로 표시된 on-target sequence 를 제외하고, 순서대로 서열번호 49 내지 84); 표 1 및 표 2 에서 소문자로 표시한 뉴클레오타이드가 mismatch nucleotide 임).

[표 1]

Hela			
Chr	Location	DNA Cleavage Score	DNA sequence at target sites
chr11	5248215	14.20	CTTGCCCCACAGGGCAGTAACGG
chr9	104595883	8.07	tcaGCCCCACAGGGCAGTAAGGG
chr17	8370253	6.58	tTgtCCCCACAGGGCAGTAAACG
chr12	124803834	6.54	gcTGCCCCACAGGGCAGcAAAGG
chr14	94585327	3.96	aTgGCCCCACaAGGCAGaAATGG
chr22	17230623	3.43	tgTGCCCCACAGaGCaTAAGGG
chr2	121715240	3.04	gTgtCCCCACAGGGCAGgAAAGG
chr12	93549202	2.89	aTTGCCCCACgGGGCAGTgACGG
chr7	97874171	2.86	CcctCCCCACAGGGCAGTcATGG
chr3	19957634	2.82	gcTaCCCCACAGGGCAtTAgGGG
chrX	75006257	2.76	gTgGCCCCACAGGGCAGgAATGG
chr15	46598129	2.57	gTTGCCCCtCAGGaCAGTAcAGG
chr9	96637408	2.54	acaGCCCCACAGGGCataAACGG
chr5	131423385	2.53	tcTGCCCCACAGGcCAGgAAGGG
chr8	24931381	2.52	agTGCCaCACaCaGCAGTAAGGG
chr14	36889538	2.31	gTTatCCCCACAGGaCAGTgAGGG
chr20	39992928	2.31	agTGgCCCCcCAGGGCAGTgAGGG
chr1	177593980	2.15	tcTaCCCCrCAtGGCAGTAATGG
chr6	23709579	1.95	gaagCCCCtACAGGGCAGcAATGG
chr4	148531374	1.75	gTTaCCtCACAGaGCAGaAAGGG
chr10	72286450	1.47	CaaGCCCCACAGGGCAGacAGGG
chr19	8560462	1.31	aaatCCCCACAGGGCAGTAAGGC
chr11	76387498	1.18	CTgcCCctcagGGaCAGTAatGGG
chr5	14347051	1.08	CaTaCCCCACAGGtCAGTAAGGA
chr1	17346702	0.95	ggTcCCCacagGGtCAGTAAGGG
chr15	29983547	0.78	CcaGCCCCACAGGGCAGTAAAGC
chr19	34262012	0.78	gcTGctCCACAGGGCAGgtATGG
chr10	73555691	0.78	CagGCCCCACAGGaCAGgAAGGG
chr19	50010010	0.73	aTTGCCCCcCAGGtCAGTAgGGG
chr15	34059408	0.56	gTTaCCaCACAGaGCAGTtAAGG
chr13	44886376	0.54	ggaGCCCCACAGGGCAGagAGGG
chr9	134609673	0.50	tTTGCCCCtCAGGGCAGctAAGG
chr16	49082904	0.46	gcagGCCCCACAGGtCAGTgAGGG
chr3	128713852	0.45	aTaGCCCCACAGGGCAGgttAGG
chr12	27234755	0.44	gaTGCCtCACAGGaCAGgAAGGG
chr19	923900	0.37	aaTtaCCCCcCAGGGCAGgAAGGG
chr8	143699609	0.35	CcaGCCCCACcGGGCAGTAgATG
chr12	50056756	0.22	aagGCCCCgCAGGGCAGTttGGG
chr2	240591539	0.20	acaGCCCCACAGGGCAcTAgAGG
chr17	66624245	0.17	CcTctCCCCACAGGGCAGTAAAGG
chr9	134994964	0.15	CcTGCCCCACAGGGCAaTtATGG
chr2	43013308	0.15	CcTGCCcagtAaAGGCAGTAAGGG
chr8	41296595	0.14	tcaGCCCCACAGGtCAGcAATGG
chr22	35537395	0.14	agTGCCCCACAGGGgAGaAATGG

[표 2]

HEK293T			
Chr	Location	DNA Cleavage Score	DNA sequence at target sites
chr11	5248215	14.63	CTTGCCCCACAGGGCAGTAACGG
chr9	104595883	8.17	tcaGCCCCACAGGGCAGTAAGGG
chr17	8370253	3.65	tTgctCCACAGGGCAGTAAACG
chr12	124803834	2.96	gcTGCCCCACAGGGCAGcAAAGG
chr14	94585327	1.82	aTgGCCCCACAaGGCAGaAATGG
chr12	93549202	1.01	aTTGCCCCACgGGGCAGTgACGG
chr2	121715240	0.88	gTgtCCCCACAGGGCAGgAAAGG
chr12	27234755	0.70	gaTGCCtCACAGGaCAGgAAGGG
chrX	75006257	0.46	gTgGCCCCACAGGGCAGgAATGG
chr14	36889538	0.40	gTTatCCACAGGaCAGTgAGGG
chr6	23709579	0.29	gaaGCCCtACAGGGCAGcAATGG
chr7	97874171	0.28	CcctCCCCACAGGGCAGTcATGG
chr10	72286450	0.23	CaaGCCCCACAGGGCAGacAGGG
chr8	24931381	0.19	agTGCCaCACaGcAGTAAGGG
chr9	134994964	0.12	tTTGCCCCtCAGGGCAGctAAGG
K562			
Chr	Location	DNA Cleavage Score	DNA sequence at target sites
chr11	5248215	16.29	CTTGCCCCACAGGGCAGTAACGG
chr9	104595883	9.41	tcaGCCCCACAGGGCAGTAAGGG
chr17	8370253	3.80	tTgctCCACAGGGCAGTAAACG
chr12	124803834	2.65	gcTGCCCCACAGGGCAGcAAAGG
chrX	75006257	2.62	gTgGCCCCACAGGGCAGgAATGG
chr14	94585327	1.04	aTgGCCCCACAaGGCAGaAATGG
chr2	121715240	0.56	gTgtCCCCACAGGGCAGgAAAGG
chr22	17230623	0.30	tgTGCCCCACAGaGCaTAAGGG
chr12	93549202	0.28	aTTGCCCCACgGGGCAGTgACGG
chr1	17346702	0.15	ggTcCCcagGGtCAGTAAGGG
chr7	97874171	0.14	CcctCCCCACAGGGCAGTcATGG
chr6	23709579	0.12	gaaGCCCtACAGGGCAGcAATGG
T cell			
Chr	Location	DNA Cleavage Score	DNA sequence at target sites
chr11	5248215	15.29	CTTGCCCCACAGGGCAGTAACGG
chr9	104595883	5.64	tcaGCCCCACAGGGCAGTAAGGG
chr12	124803834	3.78	gcTGCCCCACAGGGCAGcAAAGG
chr17	8370253	3.27	tTgctCCACAGGGCAGTAAACG
chr14	94585327	3.03	aTgGCCCCACAaGGCAGaAATGG
chr12	93549202	1.94	aTTGCCCCACgGGGCAGTgACGG
chr22	17230623	1.52	tgTGCCCCACAGaGCaTAAGGG
chrX	75006257	1.46	gTgGCCCCACAGGGCAGgAATGG
chr6	23709579	1.23	gaaGCCCtACAGGGCAGcAATGG
chr2	121715240	1.06	gTgtCCCCACAGGGCAGgAAAGG
chr19	50010010	0.50	aTTGCCCCcCAGgtCAGTAAGGG
chr5	131423385	0.23	tcTGCCCCACAGGcCAGgAAGGG

실시에 3: 크로마틴 DNA를 이용한 절단 유전체 서열 분석

- 7 개의 다른 유전자 가위 (상이한 표적 서열을 포함하는 가이드 RNA 를 포함함)를 이용하여 크로마틴 DNA 를 이용한 절단 유전체 시퀀싱을 진행하여 세포 (Hela cell) 내에서의 돌연변이 (indel frequency)를 검증하여 그 결과를 표 3 (표적 서열들은 순서대로 서열번호 85 내지 115), 표 4 (표적 서열들은 순서대로 서열번호 116 내지 151), 표 5 (표적 서열들은 순서대로 서열번호 152 내지 164) 및 표 6 (표적 서열들은 순서대로 서열번호 165 내지 181)에 나타내었다 (표 3 내지 표 6 에서 소문자로 표시한 뉴클레오타이드가 mismatch nucleotide 임).
- 5

[H 3]

VEGFA1

	Chromosome	Location	DNA Cleavage score	Target sequence	Indel frequency (%)	
On-target	chr6	43737297	8.0	GGGTGGGGGGAGTITGCTCCAGG	(-)RGEN	(+)RGEN
VEGFA1-2	chr17	39796328	6.3	taGTGgaGGGAGcTTGCTCCTGG	0.01%	21.77%
VEGFA1-3	chr12	1988077	5.2	cGGgGGgaGGGAGTITGCTCCTGG	0.00%	16.90%
VEGFA1-4	chr6	14316373	4.5	GtGgGGGtaGAGTITGCTCCAGG	0.00%	11.73%
VEGFA1-5	chr15	65637537	3.7	GgaTGgaGGGAGTITGCTCCTGG	0.02%	8.10%
VEGFA1-6	chr16	8763213	3.5	aaGTaaGGGaAGTITGCTCCTGG	0.01%	25.28%
VEGFA1-7	chr11	71497119	3.3	aGGaaGGaGGAGTITGCTCCTGG	0.01%	0.01%
VEGFA1-8	chr12	26641302	3.1	aGtTtGGGGGAGTITGCTCCAGG	0.00%	0.02%
VEGFA1-9	chr5	139263024	3.1	ttGgGGGGGcAGTITGCTCCTGG	0.12%	0.12%
VEGFA1-10	chr1	99347651	2.9	GGGgaGGGGaAGTITGCTCCTGG	2.33%	7.19%
VEGFA1-11	chr11	3445204	2.8	aGGaaGGaGGAGTITGCTCCTGG	0.01%	13.84%
VEGFA1-12	chr20	56175356	2.7	aGGgaGGaGGaATITGCTCCAGG	0.02%	0.04%
VEGFA1-13	chr5	32945275	2.2	GcGTGGGGGtGTITGCTCCCGG	0.00%	0.72%
VEGFA1-14	chr17	47317539	2.1	ctGgtGGGGAGcTTGCTCCAGG	0.03%	1.00%
VEGFA1-15	chr3	125633992	2.1	aGGaaGGaGGAGTITGCTCCTGG	1.64%	4.14%
VEGFA1-16	chr12	131690199	1.9	GGGaGGGtGGAGTITGCTCCTGG	0.02%	0.01%
VEGFA1-17	chr13	31251013	1.8	tGtaGaGGGaGtTTGCTCCCGG	0.00%	6.70%
VEGFA1-18	chr7	17819097	1.8	acaactGGGGAGTITGCTCCTGG		Not tested
VEGFA1-19	chr22	18454623	1.6	GGaaaGGaGGAGcTTGCTCCAGG		Not tested
VEGFA1-20	chr3	195871264	0.9	GGtgGGGgaGAGcTaGCTCCGGG	0.00%	0.20%
VEGFA1-21	chr22	19698463	0.6	GaGgGGGaGcAGTITGCTCCAGG	0.01%	0.56%
VEGFA1-22	chr17	32986325	0.5	GGGdGtGGGGAGTITGCTCCAGG	0.04%	0.02%
VEGFA1-23	chr22	37215276	0.5	GGGTGGGGGGAGTITGCTCCAGG		Not tested
VEGFA1-24	chr13	28075024	0.4	atGaGgaGGGAGTITGCTCCAGG		Not tested
VEGFA1-25	chr6	50485682	0.4	atGTGtGGGGAGTITGCTCCAGG		Not tested
VEGFA1-26	chr5	7067159	0.3	GaGgGtGGGGAGTITGCTCCTGG		Not tested
VEGFA1-27	chr15	93140401	0.3	GGGdGaGGGaAGTITGCTCCAGG	0.02%	0.01%
VEGFA1-28	chr11	67574262	0.3	aGGaaGGaGGAGTITGCTCCTGG	0.01%	0.73%
VEGFA1-29	chr2	129199817	0.2	tccTGaGGGcAGTITGCTCCAGG		Not tested
VEGFA1-30	chr7	76022392	0.2	aGGgtGGGGcGtTTGCTCCCGG		Not tested
VEGFA1-31	chr13	26202812	0.2	GgtTGaGGGGAGTITGCTCCAGG	0.01%	0.17%

[4]

FANCF

	Chromosome	Location	DNA Cleavage score	Target sequence	Indel frequency (%)	
On-target	chr11	22647338	6.3	GGAAATCCCTTCTGAGCAGCCTGG	(-)RGEN	(+)RGEN
FANCF-2	chr7	102131659	5.5	GtctcCCCTTCTGAGCAGCAGG	0.06%	54.37%
FANCF-3	chr10	73463136	4.1	tGAATCCCaTCTcCAGCAGCAGG		
FANCF-4	chr7	102230832	3.8	GtctcCCCTTCTGAGCAGCAGG	0.01%	0.34%
FANCF-5	chr11	66475045	3.5	GGAAaCCCTTCTGAGCAGCAGG		
FANCF-6	chr16	28615201	3.2	GGctTCCCTTCTGAGCAGCAGG	0.00%	0.07%
FANCF-7	chr7	44076497	3.0	GtctcCCCTTCTGAGCAGCAGG	0.11%	0.12%
FANCF-8	chr18	8707528	1.7	GGAAcCCCTTCTGAGCAGCAGG		
FANCF-9	chr19	39405599	1.6	GctTCCCTTCTGAGCAGCAGG	0.05%	27.79%
FANCF-10	chr9	139891539	1.5	aGtctCCCaTCTGAGCAGCAGG		
FANCF-11	chr17	39675789	1.3	GGqAgTCCaTCTGAGCAGCAGG	0.01%	0.02%
FANCF-12	chrX	86355180	1.2	accATCCCTcTCTGAGCAGCAGG	0.02%	0.35%
FANCF-13	chr12	2719895	1.0	aAcTCCCTTCTGAGCAGCAGG	0.00%	0.01%
FANCF-14	chr10	43410031	0.7	GGATCCCTTcCTaCAGCAGCAGG	0.01%	5.41%
FANCF-15	chr17	78923978	0.7	aGAgTCCCTcTCTGAGCAGCAGG	0.01%	3.09%
FANCF-16	chr16	49671025	0.6	GGATCCCTTcCTGAGCAGCAGG	0.00%	0.82%
FANCF-17	chr10	37953200	0.6	GGATCCCTTcCTaCAGCAGCAGG	0.01%	2.75%

FANCF-18	chr7	5013586	0.6	GGTccCtCcTcGgCAGCACCcGG			Not tested
FANCF-19	chr2	133174786	0.5	atccTtCcTtTcTcGcAGCACCtGG			Not tested
FANCF-20	chr10	3151994	0.5	ctctgtCcTtTcTcGcAGCACCtGG			Not tested
FANCF-21	chr11	47554037	0.4	GGAAIcCcTtTcTcTaCAGCAtCcTG			Not tested
FANCF-22	chr17	30452575	0.4	ccgtgtgtCcTtTcTcGcAGCACCtGG			Not tested
FANCF-23	chrX	153002368	0.3	ccAcTcCcTtTcTcGcAGCACCtGC			Not tested
FANCF-24	chr4	53092937	0.2	aatAIcCcCtTcTcGcAGCACCAGG			Not tested
FANCF-25	chr17	65709672	0.2	attccCcCcTtTcTcGcAGCcCcTGG			Not tested
FANCF-26	chr22	45871981	0.2	ccccTcCcTcCtTcTcGcAGCACCcGG			Not tested
FANCF-27	chr12	115467808	0.2	aGggTcCcTtTcTcTcGcAGCcCcTGG	0.04%	0.06%	Invalidated
FANCF-28	chr19	10929976	0.2	aGtAaaggCtTcTcTcGcAGCACCtGG			Not tested
FANCF-29	chr3	35113165	0.2	tGAAtCcCtAaCtGcAGCACCAGG			Not tested
FANCF-30	chr2	242473665	0.1	GGgAaCcCcTcCcCAGCACCcGG			Not tested
FANCF-31	chr20	61912779	0.1	caacATcCcTtTcTcGcAGCcCcCAG			Not tested
FANCF-32	chr1	223900677	0.1	GGctTcCcCaCtGcAGCACCcGG			Not tested
FANCF-33	chr22	26091104	0.1	ccgtccCcCtTcTcTcGcAGCACCtCG			Not tested
FANCF-34	chr17	34985068	0.1	GGgtccCaCtTcTcTcGcAGCACCtGG	0.00%	0.00%	Invalidated
FANCF-35	chr1	27648497	0.1	aGcAtCcCcTtTcTcGcAGCcGgCGGG			Not tested
FANCF-36	chr13	114815604	0.1	GGgAaCcCcTtTcCgCAGCACCcCAG			Not tested

[5]

HEK3

	Chromosome	Location	DNA Cleavage score	Target sequence	Indel frequency (%)	
					(-)RGEN	(+)RGEN
On-target	chr9	110184637	5.8	GGCCAGACTGAGCACCIGATGG	0.01%	66.99%
HEK3-2	chr1	47005705	2.8	AGCtCAGACTGAGCAAGTGAGGG	0.01%	16.23%
HEK3-3	chr15	79749931	2.7	caCCCAGACTGAGCACCIGcTGG	0.00%	33.14%
HEK3-4	chr10	131593121	1.5	GagCCAGAA TGAGCACCIGTAGGG	0.00%	1.17%
HEK3-5	chr2	11810044	0.3	GgCCAGgCTGAGCACCIGgTGG		Not tested
HEK3-6	chr17	34954539	0.1	cGgCCcAACtGAGCAAGIGATGG		Not tested

RNF2

	Chromosome	Location	DNA Cleavage score	Target sequence	Indel frequency (%)	
					(-)RGEN	(+)RGEN
On-target	chr1	185056773	8.41	GTCATCTTAGTCATTACCTGAGG	0.01%	68.06%
RNF2-2	chr5	92036966	1.03	GgtATCTaAGTCATTACCTGTGG	0.07%	0.24%
RNF2-3	chr11	60883789	0.4	aTaaaCgTAGTCATTACCTGGGG		Not tested
RNF2-4	chr8	99451745	0.3	cgtgcattAGTCATTACCTGAGG	0.01%	0.02%
						Invalidated

HEK2

	Chromosome	Location	DNA Cleavage score	Target sequence	Indel frequency (%)	
					(-)RGEN	(+)RGEN
On-target	chr5	87240614	4.44	GAACACAAGCATAGACTGCGGG	0.01%	59.05%
HEK2-2	chr4	90522184	0.27	GAACACAatGCATAGatTGCCTGG	0.01%	16.33%
HEK2-3	chr5	142142325	0.2	tAAacactAaCATAGACTGCAGG		

[표 6]

EMX1						
	Chromosome	Location	DNA Cleavage score	Target sequence	Indel frequency (%)	
On-target						
EMX1-2	chr2	73160999	15.4	GAGTCCGAGCAGGAGGAGGAGG	(-)RGEN 0.23%	(+)RGEN 61.61%
EMX1-3	chr6	9118799	4.7	acGTCtGAGCAGAGGAGGATGG	0.03%	0.75%
EMX1-4	chr5	45359067	3.4	GAGTcaGAGCAGAGGAGGAGG	0.02%	47.11%
EMX1-5	chr15	44109764	1.9	GAGTCTaAGCAGAGGAGGAGG	0.42%	39.41%
EMX1-6	chr5	9227163	1.7	aAGTCtGAGCAGAGGAGGATGG		Not tested
EMX1-7	chr1	33606480	1.2	GAGcCtGAGCAGAGGAGGAGG	0.01%	0.08%
EMX1-8	chr18	34906762	0.8	GAGcCtGAGCgGAGAGGAGG		Not tested
EMX1-9	chr16	78848850	0.8	aAaTCCaACcAGAGAGGAGG		Not tested
EMX1-10	chr4	87256693	0.6	GAGTaaGAGaAGAGAGGAGG	0.08%	0.09%
EMX1-11	chr11	62365273	0.4	GAaTCCaAGCAGAGAGGAGG	0.02%	0.13%
EMX1-12	chr4	131662222	0.3	agaaTCCaagAGAGAGGATGG		Not tested
EMX1-13	chr15	61646877	0.2	aAGTCaGAGgAGAGAGGAGG	0.26%	0.47%
EMX1-14	chr19	1438827	0.2	GAaGtaGAGCAGAGAGAGG		Not tested
EMX1-15	chr17	8640230	0.2	GAGaCtGAGaAGAGAGAGG		Not tested
EMX1-16	chr14	75723909	0.1	aguTCCaAGCAGAGGAGAGG		Not tested
CCR5-9						
	Chromosome	Location	DNA Cleavage score	Target sequence		
On-target	chr3	46414766	12.2	TCATCCTCCTGACAAATCGATAGG		
CCR5-9-1	chr10	48385259	0.3	aCAICCTtCIGACAAICatTtGG		

세포 내에서, 상기 표 3 내지 6 에 나타난 절단 유전체 시퀀싱 결과 나온 비표적 위치에 대한 풀업변이 비율을 추정하여, 이를 그 위치에서의 DNA

절단 점수와 비교해 보았다. 상기 비교 결과를 도 3a 에 나타내었다. 비교를 위하여, 크로마틴 제거 DNA 에 대하여 동일한 시험을 수행하여, 그 결과를 도 3b 에 나타내고, 도 3a 및 3b 에 나타난 상관 관계 (평균 R^2 ; 1 에 가까울수록 상관성이 높음)를 비교하여 도 3c 에 나타내었다. 도 3a 내지 도 3c 에 나타난 바와 같이, 크로마틴 DNA 를 이용하여 절단 유전체 시퀀싱 하였을 경우, 특정 위치에서의 DNA 절단 점수와 돌연변이 비율의 상관 관계 (평균 $R^2 = 0.62$)가 크로마틴 제거 DNA 를 이용하여 절단 유전체 시퀀싱 하였을 때의 상관 관계 (평균 $R^2 = 0.20$) 보다 유의미하게 높은 것을 확인할 수 있다.

10

실시예 4: 크로마틴 DNA (핵펠렛 및 native chromatin DNA)의 절단 시험

앞선 실시예에서, Cas9 뉴클레아제 활성화에 크로마틴이 중요한 역할을 함을 확인할 수 있었으므로, 게놈 전체의 Cas9 비표적(Genome-wide Cas9 off-target)의 프로파일 시에 크로마틴 상태를 반영하기 위해 본래의 크로마틴 DNA(native chromatin DNA)를 사용하는 Digenome-seq (절단 유전체 시퀀싱) 방법이 이점을 가짐을 보이기 위하여 하기의 시험을 수행하였다.

게놈 전체의 Cas9 비표적 부위를 확인하기 위하여, 크로마틴 DNA 를 사용하여 Digenome-seq 을 수행하였다. 실시예 1 및 하기의 표 7 의 표적 서열 (박스 내 서열 (HBB 표적 서열))을 참조하여, Hela 세포 또는 HEK293T 세포로부터 분리된 핵 펠렛 (nuclei pellet) 및 native chromatin 을 미리 조립된 Cas9 단백질 (300nM) 및 sgRNA (900nM)를 포함하는 리보핵산 단백질과 함께 배양하고, DNA 정제 후, whole-genome sequencing (WGS)을 수행하였다 (도 1 참조). 상기 핵 펠렛은 세포에서 세포질 제거 후 핵막이 있는 상태로, 세포에 lysis buffer (1x PBS, 0.4% NP-40, 1 mM EDTA) 처리 후 500 x g 에서 5 분간 원심분리하고 세포질층(상층액)을 제거하여 준비하고 (실시예 1 내지 실시예 3 에서 사용된 크로마틴 DNA 와 동일함), native chromatin 는 핵막이 제거 된 상태로, 세포에 lysis buffer (1x PBS, 0.4% NP-40, 1 mM EDTA) 처리 후 500 x g 에서 5 분간 원심분리하고 세포질층(상층액)을 제거한 후, Nuc-lysis solution (10 mM EDTA, 0.5 mM EGTA, 0.1% Triton X100)을 처리하여 Nucleoplasm(상층액)을 제거하여 준비하였다.

인간 참조 게놈 (hg19)에 대한 서열 정렬 후, integrative genomics viewer (IGV)를 이용하여, 표적 부위에서의 Cas9-매개 DNA 소화에 의해 유도된 직선 정렬이 관찰되었다.

Digenome-seq 과 관련된 이전 연구(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))에서 사용된 DNA 절단 점수(DNA cleavage score)를 통해 Cas9
 5 비표적 위치를 확인하였다. 상기 Cas9 비표적 위치 확인은 Digenome 1.0 (DNA cleavage score 가 2.5 이상인 site 를 비표적 위치 후보군으로 결정함) 및 Digenome 2.0 (DNA cleavage score 가 0.1 이상이며 10 개 이하의 미스매치를 가지고 PAM (5'-NGN-3' 또는 5'-NNG-3')을 가지고 있는 site 를 비표적 위치로 결정)에 의하여 수행하였다.

10 상기 얻어진 결과를 도 5a (Hela cell 에서 HBB 특이적 Cas9 를 사용하여 Digenome 1.0 에 의하여 얻어진 비표적 위치 결과를 보여주는 벤다이어그램), 5b (Hela cell 에서 HBB 특이적 Cas9 를 사용하여 Digenome 1.0 에 의하여 얻어진 비표적 위치에서의 핵 펠렛의 절단 점수와 native chromatin 의 절단 점수의 상관관계를 보여주는 그래프), 6a (Hela cell 에서
 15 HBB 특이적 Cas9 를 사용하여 Digenome 2.0 에 의하여 얻어진 비표적 위치 결과를 보여주는 벤다이어그램), 6b (Hela cell 에서 HBB 특이적 Cas9 를 사용하여 Digenome 2.0 에 의하여 얻어진 핵 펠렛 및 native chromatin에서의 비표적 위치의 절단 점수 간의 상관관계를 보여주는 그래프), 7a (HEK293T cell 에서 HBB 특이적 Cas9 를 사용하여 Digenome 2.0 에 의하여 얻어진
 20 비표적 위치 결과를 보여주는 벤다이어그램), 및 7b (HEK293T cell 에서 HBB 특이적 Cas9 를 사용하여 Digenome 2.0 에 의하여 얻어진 비표적 위치에서의 핵 펠렛의 절단 점수와 native chromatin 의 절단 점수의 상관관계를 보여주는 그래프)에 각각 나타내었다.

도 5a 에서 확인되는 바와 같이, Hela cell 의 핵 펠렛 및 native
 25 chromatin 에서 Digenome 1.0 (컷오프: 2.5)을 통해 HBB 특이적 Cas9 를 사용하여 각각 15 및 11 개의 in vitro 절단 위치를 관찰하였다. 이 때, native chromatin 에서 확인된 in vitro 절단 위치 (11 개) 중 10 개 (91%; 10/11)는 핵 펠렛에서 확인된 절단 위치 (15 개) 중 10 개와 중첩되었다. 또한, 도 5b 에서 확인되는 바와 같이, Hela cell 에서, native
 30 chromatin 에서 확인된 in vitro 절단 위치의 DNA 절단 점수는 핵 펠렛에서 확인된 in vitro 절단 위치의 DNA 절단 점수와 높은 상관 관계를 보였다 ($R^2 = 0.97$). 이러한 결과는 핵 펠렛 또는 native chromatin 을 사용한 Digenome-seq 이 높은 재현성을 가짐을 보여준다.

도 6a 및 6b 에서 확인되는 바와 같이, Hela cell 의 핵 펠렛 및 native chromatin 에서 Digenome 2.0 (컷오프: 1.0)를 통해 HBB 특이적 Cas9 를 사용하여 각각 44 개와 37 개의 in vitro 절단 위치가 관찰되었으며, 이 중에서 34 개가 서로 중첩되었고, 이들 간 높은 상관 관계가 관찰되었다
5 ($R^2 = 0.97$).

도 7a 및 7b 에서 확인되는 바와 같이, HEK293T cell 의 핵 펠렛 및 native chromatin 에 대하여 Digenome 2.0 (컷오프: 1.0)를 통해 HBB 특이적 Cas9 를 사용하여 각각 12 개와 7 개의 in vitro 절단 위치가 관찰되었으며, 이 중에서 6 개가 서로 중첩되었고, 이들 간 높은 상관 관계가 관찰되었다
10 ($R^2 = 0.88$).

이러한 결과는 핵 펠렛과 native chromatin 을 사용하는 Digenome-seq 이 서로 상호 연관성이 높고 핵 펠렛을 사용한 Digenome-seq 이 native chromatin 을 사용하는 Digenome-seq 보다 더 많은 in vitro 절단 위치를 확인했으며, 이후 시험(표 7)에서는 핵 펠렛을 사용하여 Digenome-seq 를
15 수행하였다. 크로마틴 DNA 를 사용하는 Digenome-seq 이 일반적으로 여러 세포 유형에 적용 가능한지 확인하기 위하여, 이 방법을 K562 및 primary T cell 과 같은 다양한 세포에 적용하여 그 결과를 상기 Hela cell 및 HEK293T cell 에서의 결과와 함께 아래의 표 7 에 나타내었다 (표 7 (HBB on target sequence(네모 박스): 서열번호 5; 이를 제외한 서열들: 순서대로 서열번호
20 6 내지 84)에서 소문자로 표시한 뉴클레오타이드가 mismatch nucleotide 임).

[표 7]

Hela			
Chr	Location	DNA Cleavage Score	DNA sequence at target sites
chr11	5248215	14.20	CTTGCCCCACAGGGCAGTAACGG
chr9	104595883	8.07	tcaGCCCCACAGGGCAGTAAGGG
chr17	8370253	6.58	tTgctCCACAGGGCAGTAAACG
chr12	124803834	6.54	gcTGGCCACAGGGCAGcAAAGG
chr14	94585327	3.96	aTgGCCCCACaAGGCAGaAATGG
chr22	17230623	3.43	tgTGGCCACAGaGCACTAAGGG
chr2	121715240	3.04	gTgtCCCCACAGGGCAGgAAAGG
chr12	93549202	2.89	aTTGCCCCACgGGGCAGTgACGG
chr7	97874171	2.86	CcctCCCCACAGGGCAGTcATGG
chr3	19957634	2.82	gcTaCCCCACAGGGCAtTAAGGG
chrX	75006257	2.76	gTgGCCCCACAGGGCAGgAATGG
chr15	46598129	2.57	gTTGCCCCtCAGGaCAGTAcAGG
chr9	96637408	2.54	acaGCCCCACAGGGCAtaAAACGG
chr5	131423385	2.53	tcTGGCCACAGGcCAGgAAAGGG
chr8	24931381	2.52	agTGCcACACaGCAGTAAGGG
chr14	36889538	2.31	gTtAtCCACAGGaCAGTgAGGG
chr20	39992928	2.31	agTGgCCCcCAGGGCAGTgAGGG
chr1	177593980	2.15	tcTaCCCCACAtGGCAGTAATGG
chr6	23709579	1.95	gaaGCCCCtACAGGGCAGcAATGG
chr4	148531374	1.75	gTtACcTACAGaGCAGaAAAGGG
chr10	72286450	1.47	CaaGCCCCACAGGGCAGacAGGG
chr19	8560462	1.31	aaatCCCCACAGGGCAGTAAGGC
chr11	76387498	1.18	CTgcCCctcagGGaCAGTAAGGG
chr5	14347051	1.08	CaTaCCCCACAGGtCAGTAAGGA
chr1	17346702	0.95	ggTcCCCaagGGtCAGTAAGGG
chr15	29983547	0.78	CcaGCCCCACAGGGCAGTAAGGC
chr19	34262012	0.78	gcTGCtCCACAGGGCAGgtATGG
chr10	73555691	0.78	CagGCCCCACAGGaCAGcAAAGGG
chr19	50010010	0.73	aTTGCCCCcCAGGtCAGTAAGGG
chr15	34059408	0.56	gTtACCaCACAGaGCAGTtAAGG
chr13	44886376	0.54	ggaGCCCCACAGGGCAGagAGGG
chr9	134609673	0.50	tTTGCCCCtCAGGGCAGctAAGG
chr16	49082904	0.46	gcaGCCCCACAGGtCAGTgAGGG
chr3	128713852	0.45	aTaGCCCCACAGGGCAGgtTAGG
chr12	27234755	0.44	gaTGCCtCACAGGaCAGgAAAGGG
chr19	923900	0.37	aaTtaCCCcCAGGGCAGgAAAGGG
chr8	143699609	0.35	CcaGCCCCACcGGGCAGTAAGTG
chr12	50056756	0.22	aagGCCCCgCAGGGCAGTtGGG
chr2	240591539	0.20	acaGCCCCACAGGGCAcTAAGGG
chr17	66624245	0.17	CcTctCCACAGGGCAGTAAAGG
chr9	134994964	0.15	CcTGGCCCCACAGGGCAaTtATGG
chr2	43013308	0.15	CcTGGCCagtAaGGCAGTAAGGG
chr8	41296595	0.14	tcaGCCCCACAGGtCAGcAATGG
chr22	35537395	0.14	agTGGCCCCACAGGGgAGaAATGG

HEK293T			
Chr	Location	DNA Cleavage Score	DNA sequence at target sites
chr11	5248215	14.63	CTTGCCCCACAGGGCAGTAACGG
chr9	104595883	8.17	tcaGCCCCACAGGGCAGTAAGGG
chr17	8370253	3.65	tTgtCCCCACAGGGCAGTAAACG
chr12	124803834	2.96	gcTGCCCCACAGGGCAGcAAAGG
chr14	94585327	1.82	aTgGCCCCACAaGGCAGaAATGG
chr12	93549202	1.01	aTTGCCCCACgGGGCAGTgACGG
chr2	121715240	0.88	gTgtCCCCACAGGGCAGgAAAGG
chr12	27234755	0.70	gaTGCCtCACAGGaCAGgAAGGG
chrX	75006257	0.46	gTgGCCCCACAGGGCAGgAATGG
chr14	36889538	0.40	gTTatCCCCACAGGaCAGTgAGGG
chr6	23709579	0.29	gaaGCCCtACAGGGCAGcAATGG
chr7	97874171	0.28	CcctCCCCACAGGGCAGTcATGG
chr10	72286450	0.23	CaaGCCCCACAGGGCAGacAGGG
chr8	24931381	0.19	agTGCCaCACaGcAGTAAGGG
chr9	134994964	0.12	tTTGCCCCtCAGGGCAGctAAGG
K562			
Chr	Location	DNA Cleavage Score	DNA sequence at target sites
chr11	5248215	16.29	CTTGCCCCACAGGGCAGTAACGG
chr9	104595883	9.41	tcaGCCCCACAGGGCAGTAAGGG
chr17	8370253	3.80	tTgtCCCCACAGGGCAGTAAACG
chr12	124803834	2.65	gcTGCCCCACAGGGCAGcAAAGG
chrX	75006257	2.62	gTgGCCCCACAGGGCAGgAATGG
chr14	94585327	1.04	aTgGCCCCACAaGGCAGaAATGG
chr2	121715240	0.56	gTgtCCCCACAGGGCAGgAAAGG
chr22	17230623	0.30	tgTGCCCCACAGaGCAcTAAGGG
chr12	93549202	0.28	aTTGCCCCACgGGGCAGTgACGG
chr1	17346702	0.15	ggTcCCCaagGGtCAGTAAGGG
chr7	97874171	0.14	CcctCCCCACAGGGCAGTcATGG
chr6	23709579	0.12	gaaGCCCtACAGGGCAGcAATGG
T cell			
Chr	Location	DNA Cleavage Score	DNA sequence at target sites
chr11	5248215	15.29	CTTGCCCCACAGGGCAGTAACGG
chr9	104595883	5.64	tcaGCCCCACAGGGCAGTAAGGG
chr12	124803834	3.78	gcTGCCCCACAGGGCAGcAAAGG
chr17	8370253	3.27	tTgtCCCCACAGGGCAGTAAACG
chr14	94585327	3.03	aTgGCCCCACAaGGCAGaAATGG
chr12	93549202	1.94	aTTGCCCCACgGGGCAGTgACGG
chr22	17230623	1.52	tgTGCCCCACAGaGCAcTAAGGG
chrX	75006257	1.46	gTgGCCCCACAGGGCAGgAATGG
chr6	23709579	1.23	gaaGCCCtACAGGGCAGcAATGG
chr2	121715240	1.06	gTgtCCCCACAGGGCAGgAAAGG
chr19	50010010	0.50	aTTGCCCCcCAGGtCAGTAAGGG
chr5	131423385	0.23	tcTGCCCCACAGGcCAGgAAGGG

표 7 에서와 같이, 각 세포에서 정의되는 in vitro 절단 위치의 대부분이 서로 중첩되어 있음을 확인할 수 있다.

- 5 Hela 세포의 핵 펠렛과 크로마틴이 제거된 DNA (chromatin free DNA)에 대하여 HBB 특이적 Cas9를 사용하여 Digenome-seq (Digenome 1.0)를 수행하여 in vitro 절단 위치(on target site를 포함한 모든 절단위치)를

측정 및 비교하여, 그 결과를 도 8a 및 8b에 나타내었다. 도 8a에 나타난 바와 같이, Digenome 1.0을 사용하여 핵 펠렛과 크로마틴 제거 DNA에 대하여 Digenome-seq를 수행하여 각각 15와 48 개의 in vitro 절단 위치를 관찰하였으며, 핵 펠렛에서 관찰된 대부분의 in vitro 절단 부위 (93%)는 크로마틴 제거 DNA 사용 digenome-seq로 관찰된 in vitro 절단 부위와 중첩되었다.

Hela 세포의 핵 펠렛과 크로마틴이 제거된 DNA (chromatin free DNA)에 대하여 Digenome-seq (Digenome 2.0)를 수행하여 in vitro 절단 위치(On target site를 포함한 모든 절단위치)를 측정 및 비교하여, 그 결과를 도 9a 내지 도 9c에 나타내었다. 도 9a 및 도 9b에 나타난 바와 같이, Digenome 2.0을 사용하여 핵 펠렛과 크로마틴 제거 DNA에 대하여 Digenome-seq를 수행하여 각각 44 및 97 개의 in vitro 절단 위치가 관찰되었고, 핵 펠렛에서 관찰된 대부분의 in vitro 절단 위치부위 (98 %)는 크로마틴 제거 DNA 사용 digenome-seq로 관찰된 in vitro 절단 부위와 중첩되었다. 도 8b 및 도 9c에 나타난 바와 같이, 핵 펠렛과 크로마틴 제거 DNA에서 확인된 DNA 절단 점수는 서로 상관 관계를 거의 보이지 않았다 (Digenome 1.0의 경우 $R^2 = 0.22$, Digenome 2.0의 경우, $R^2 = 0.19$).

실시예 5: DNA 절단 점수(DNA cleavage score)와 Indel 빈도 (Indel frequencies) 간의 상관관계 확인

크로마틴 DNA (실시예 1 참조) 및 크로마틴 제거 DNA (실시예 1 참조)에 대하여 Digenome-seq (상기 표 7의 HBB 표적 서열을 사용)을 수행하고, DNA 절단 점수와 Indel 빈도 사이의 상관관계를 조사하여, 그 결과를 도 9d에 나타내었다. 크로마틴 제거 DNA를 사용한 Digenome-seq로부터 얻은 비표적 후보 위치들의 DNA 절단 점수와 indel 빈도간에는 낮은 상관관계를 보였으나 ($R^2 = 0.10$; 도 9e 참조), 크로마틴 DNA를 사용한 Digenome-seq로부터 얻은 비표적 후보 위치들의 DNA 절단 점수와 indel 빈도는 비교적 높은 상관관계를 보였다 ($R^2 = 0.72$) (도 9d 참조).

다음으로, HBB 표적 서열 (표 7) 및 다른 표적 서열 (표 8 참조)에 대해 핵 펠렛을 사용하여 Digenome-seq를 수행하고, DNA 절단 점수와 Indel 빈도를 측정하고, R^2 값을 계산하여, DNA 절단 점수와 Indel 빈도 간의 상관관계를 측정하여, 그 결과를 표 7 (HBB 표적 결과), 표 8 (표적 서열들은 순서대로 서열번호 85 내지 181), 및 도 9e (HBB, VEGFA1, HEK3,

EMX1, 및 FANCF 표적 결과) 나타내었다 (표 8에서 소문자로 표시한 뉴클레오타이드가 mismatch nucleotide임).

[표 8]

	Chromosome	Location	DNA Cleavage score	Target sequence	Indel frequency (%)	
					(-)RGEN	(+)RGEN
On-target						
VEGFA1-2	chr6	43737297	8.0	GGGTGGGGGAGTTGCTCCAGG	0.01%	21.77%
VEGFA1-3	chr17	39796328	6.3	taGTGgaGGAGaTtGCTCCTGG	0.00%	16.90%
VEGFA1-4	chr12	1988077	5.2	cGGGgaGGGAGtTtGCTCCTGG	0.00%	11.73%
VEGFA1-5	chr6	14316373	4.5	GtGgGGGaGAGtTtGCTCCTGG	0.02%	8.10%
VEGFA1-6	chr15	65637537	3.7	GGaTGGaGGGAGtTtGCTCCTGG	0.01%	25.28%
VEGFA1-7	chr16	8763213	3.5	aaGTaaGGGaAGtTtGCTCCTGG	0.01%	0.01%
VEGFA1-8	chr11	71497119	3.3	aGgaGGaGGAGtTtGCTCCTGG	0.00%	0.02%
VEGFA1-9	chr12	26641302	3.1	aGtTtGGGGGAGtTtGCTCCTGG	0.12%	0.12%
VEGFA1-10	chr5	139263024	3.1	ttGgGGGGGaAGtTtGCTCCTGG	2.33%	7.19%
VEGFA1-11	chr1	99347651	2.9	GGGaGGGGaAGtTtGCTCCTGG	0.01%	13.84%
VEGFA1-12	chr11	3445204	2.8	aGgaGGaGGAGtTtGCTCCTGG	0.02%	0.04%
VEGFA1-13	chr20	56175356	2.7	aGgaGGaGGaTtTtGCTCCTGG	0.00%	0.72%
VEGFA1-14	chr5	32945275	2.2	GcGTGGGGGtGtTtGCTCCTGG	0.03%	1.00%
VEGFA1-15	chr17	47317539	2.1	ctGtLGGGGGAGcTtGCTCCTGG	1.64%	4.14%
VEGFA1-16	chr3	125633992	2.1	aGgaGGaGGAGtTtGCTCCTGG	0.02%	0.01%
VEGFA1-17	chr12	131690199	1.9	GGGaGGGtGGAGtTtGCTCCTGG	0.00%	6.70%
VEGFA1-18	chr13	31251013	1.8	tGtaGaGGGaGtTtTtGCTCCTGG		Not tested
VEGFA1-19	chr7	17819097	1.8	acaacLGGGGAGtTtGCTCCTGG		Not tested
VEGFA1-20	chr22	18454623	1.6	GGaaGGaGGAGcTtGCTCCTGG		Not tested
VEGFA1-21	chr3	195871264	0.9	GGtgGGGaGAGcTtGCTCCTGG	0.00%	0.20%
VEGFA1-22	chr22	19698463	0.6	GaGGGGGaGcAGtTtGCTCCTGG	0.01%	0.56%
VEGFA1-23	chr17	32986325	0.5	GGGtGtGGGaGcTtGCTCCTGG	0.04%	0.02%
VEGFA1-24	chr22	37215276	0.5	GGGTGGGGGAGtTtGCTCCTGG		Invalidated
VEGFA1-25	chr13	28075024	0.4	atGaGgaGGAGtTtGCTCCTGG		Not tested
VEGFA1-26	chr6	50485682	0.4	atGTtGGGGaTtTtGCTCCTGG		Not tested
VEGFA1-27	chr5	7067159	0.3	GaGgGtGGGGAGtTtGCTCCTGG		Not tested
VEGFA1-28	chr15	93140401	0.3	GGGtGaGGGaAGtTtGCTCCTGG	0.02%	0.01%
VEGFA1-29	chr11	67574262	0.3	aGgaGGaGGAGtTtGCTCCTGG	0.01%	0.73%
VEGFA1-30	chr2	129199817	0.2	tccTgaGGGaAGtTtGCTCCTGG		Not tested
VEGFA1-31	chr7	76022392	0.2	aGgaGGGGcGtTtGCTCCTGG		Not tested
VEGFA1-32	chr13	26202812	0.2	GGTtGaGGGGAGtTtGCTCCTGG	0.01%	0.17%

FANCF

	Chromosome	Location	DNA Cleavage score	Target sequence	(-)RGEN	(+)RGEN	Indel frequency (%)
On-target	chr11	22647338	6.3	GGAAICGCTTICGAGCACCCTGG	0.06%	54.37%	validated
FANCF-2	chr7	102131659	5.5	GctctcGCTTICGAGCACCAGG			Not tested
FANCF-3	chr10	73463136	4.1	TGAATCCaTCTCAGCACCAGG	0.01%	0.34%	validated
FANCF-4	chr7	102230832	3.8	GctctcGCTTICGAGCACCAGG			Not tested
FANCF-5	chr11	66475045	3.5	GGAAcGCTTICGAGCACCAGG	0.00%	0.07%	validated
FANCF-6	chr16	28615201	3.2	GGctTCCCTTICGAGCACCAGG	0.11%	0.12%	Invalidated
FANCF-7	chr7	44076497	3.0	GctctcGCTTICGAGCACCAGG			Not tested
FANCF-8	chr18	8707528	1.7	GGAAcGCGaTCTCAGCACCAGG	0.05%	27.79%	validated
FANCF-9	chr19	39405599	1.6	GctgTCCCTTICGAGCACCAGG			Not tested
FANCF-10	chr9	139891539	1.5	aGtTcGCTTICGAGCACCAGG			Not tested
FANCF-11	chr17	39675789	1.3	GGaATCCaTCTCAGCACCAGG	0.01%	0.02%	Invalidated
FANCF-12	chrX	86355180	1.2	accATCCCTCTCAGCACCAGG	0.02%	0.35%	validated
FANCF-13	chr12	2719895	1.0	aGcTCCCTTICGAGCACCAGG	0.00%	0.01%	Invalidated
FANCF-14	chr10	43410031	0.7	GGaTCCCTCTCAGCACCAGG	0.01%	5.41%	validated
FANCF-15	chr17	78923978	0.7	aGAgGCGCcTCTCAGCACCAGG	0.01%	3.09%	validated
FANCF-16	chr16	49671025	0.6	GGaTCCCTCTCAGCACCAGG	0.00%	0.82%	validated
FANCF-17	chr10	37953200	0.6	GGaTCCCTCTCAGCACCAGG	0.01%	2.75%	validated
FANCF-18	chr7	5013586	0.6	GGTccCTCTCAGCACCAGG			Not tested
FANCF-19	chr2	133174786	0.5	atccTctCTTICGAGCACCAGG			Not tested
FANCF-20	chr10	3151994	0.5	ctctgTCCCTTICGAGCACCAGG			Not tested
FANCF-21	chr11	47554037	0.4	GGAAICGCTTCTaCAGCAcCTTG			Not tested
FANCF-22	chr17	30452575	0.4	cctgctcCTTICGAGCACCAGG			Not tested
FANCF-23	chrX	153002368	0.3	ccAcTCCCTTICGAGCACCAGG			Not tested
FANCF-24	chr4	53092937	0.2	aatATCCcTCTCAGCACCAGG			Not tested
FANCF-25	chr17	65709672	0.2	atttcGCTTCTCAGCACCAGG			Not tested
FANCF-26	chr22	45871981	0.2	ccccTCCCTCTCAGCACCAGG			Not tested
FANCF-27	chr12	115467808	0.2	aGgTICCTTICGAGCACCAGG	0.04%	0.06%	Invalidated
FANCF-28	chr19	10929976	0.2	aGtAagGCTTCTCAGCACCAGG			Not tested
FANCF-29	chr3	35113165	0.2	TGAATCCtaeTCTCAGCACCAGG			Not tested
FANCF-30	chr2	242473665	0.1	GGgAaTCCcTCTCAGCACCAGG			Not tested
FANCF-31	chr20	61912779	0.1	cacATCCCTTCTCAGCACCAGG			Not tested
FANCF-32	chr1	223900677	0.1	GGctTCCcCaTCTCAGCACCAGG			Not tested
FANCF-33	chr22	26091104	0.1	cGgtcGCTTCTCAGCACCAGG			Not tested
FANCF-34	chr17	34985068	0.1	GGgTcCaCTTCTCAGCACCAGG	0.00%	0.00%	Invalidated
FANCF-35	chr1	27648497	0.1	aGcATCCCTTCTCAGCACCAGG			Not tested
FANCF-36	chr13	114815604	0.1	GGgAaTCCCTTCTCAGCACCAGG			Not tested

HEK3						
	Chromosome	Location	DNA Cleavage score	Target sequence	Indel frequency (%)	
On-target	chr9	110184637	5.8	GGCCAGACTGACGACGGAIGG	(-)RGEN 0.01% (+)RGEN 66.99%	validated
HEK3-2	chr1	47005705	2.8	agctcagactgagcaagtigaggg	0.01%	validated
HEK3-3	chr15	79749931	2.7	caCCAGACTGACGACGIGIGG	0.00%	validated
HEK3-4	chr10	131593121	1.5	GagCCAGAaTGACACGIGAGGG	0.00%	validated
HEK3-5	chr2	11810044	0.3	GcgCCAGGCTGAGCACGIGIGG		Not tested
HEK3-6	chr17	34954539	0.1	cGgCCcaACIGAGCAaGIGAIGG		Not tested
RNF2						
	Chromosome	Location	DNA Cleavage score	Target sequence	Indel frequency (%)	
On-target	chr1	185056773	8.41	GTCATCTTAGTCATTACCTGAGG	(-)RGEN 0.01% (+)RGEN 68.06%	validated
RNF2-2	chr5	92036966	1.03	GgtAICTaAGTCATTACCTGIGG	0.07%	validated
RNF2-3	chr11	60883789	0.4	aTaaaCqTAGTCATTACCTGGGG		Not tested
RNF2-4	chr8	99451745	0.3	cgtgcaTTAGTCATTACCTGAGG	0.01%	Invalidated
HEK2						
	Chromosome	Location	DNA Cleavage score	Target sequence	Indel frequency (%)	
On-target	chr5	87240614	4.44	GAACACAAAGCAATAGACTGCGGG	(-)RGEN 0.01% (+)RGEN 59.05%	validated
HEK2-2	chr4	90522184	0.27	GAACACAAAGCAATAGACTGCGGG	0.01%	validated
HEK2-3	chr5	142142325	0.2	tAAa ca ct a a CAATAGACTGCGAGG	16.33%	

EMX1						
	Chromosome	Location	DNA Cleavage score	Target sequence	Indel frequency (%)	
On-target	chr2	73160999	15.4	GAGTCCGAGCAGAGAAGAAGGG	(-)RGEN 0.23%	validated
EMX1-2	chr6	9118799	4.7	aCGTCTGAGCAGAGAAGAATGG	0.03%	validated
EMX1-3	chr5	45359067	3.4	GACTTaGAGCAGAGAAGAAGGG	0.02%	validated
EMX1-4	chr15	44109764	1.9	GAGTCTaAGCAGAGAAGAAGAG	0.42%	validated
EMX1-5	chr5	9227163	1.7	aAGTCTGAGCAGAGAAGAATGG		Not tested
EMX1-6	chr1	33606480	1.2	GAGCCTGAGCAGAGAAGAAGGG	0.01%	validated
EMX1-7	chr18	34906762	0.8	GAGCCTGAGCAGAGAAGAAGGG		Not tested
EMX1-8	chr16	78848850	0.8	aATCCaACcAGAGAAGAAGAAGG		Not tested
EMX1-9	chr4	87256693	0.6	GAGTaaGAGaAGAGAAGAAGGGG	0.08%	Invalidated
EMX1-10	chr11	62365273	0.4	GATTCaAGCAGAGAAGAAGAGaG	0.02%	validated
EMX1-11	chr4	131662222	0.3	aqaatCCaagAGAGAAGAATGG		Not tested
EMX1-12	chr15	61646877	0.2	aAGTCaGAGgAGAGAAGAAGGGG	0.26%	validated
EMX1-13	chr19	1438827	0.2	GaaqLaGAGCAGAGAAGAAGCcG		Not tested
EMX1-14	chr17	8640230	0.2	GAGaCTGAGaAGAGAAGAAGGG		Not tested
EMX1-15	chr14	75723909	0.1	aATTCaAGCAGAGAAGAAGGGG		Not tested
CCR5-9						
	Chromosome	Location	DNA Cleavage score	Target sequence		
On-target	chr3	46414766	12.2	TCATCTCTCTCACAATCGAATAGG		
CCR5-9-1	chr10	48385259	0.3	aCATCTTCTCTCACAATCatTTGG		

표 8 및 도 9e에서와 같이, 크로마틴 DNA를 사용하는 Digenome-seq로부터 측정된 R^2 값은 크로마틴 제거 DNA를 사용하여 측정된 R^2 값과 비교하여 높게 나타났다. 이러한 결과는 크로마틴 DNA를 사용하여 Digenome-seq를 수행하는 경우 세포에서의 Indel 빈도를 보다 정확하게 예측할 수 있음을 보여준다.

【청구범위】

【청구항 1】

(a) 분리된 유전체 (genomic) DNA 를 표적 특이적 뉴클레아제로 절단하는 단계;

5 (b) 상기 절단된 DNA 에 대한 전체 유전체 시퀀싱 (whole genome sequencing)을 수행하는 단계; 및

(c) 상기 시퀀싱으로 획득한 염기서열 데이터 (sequence read)에서 상기 절단된 위치를 결정하는 단계를 포함하고,

상기 분리된 유전체 DNA 는 크로마틴 DNA 를 포함하는 것인,

10 표적 특이적 뉴클레아제의 비표적 위치 (off-target site)를 검출하는 방법.

【청구항 2】

제 1 항에 있어서, 상기 절단된 위치가 표적 위치 (on-target site)가 아닌 경우, 비표적 위치 (off-target site)로 판단하는 단계를 추가로
15 포함하는 것인, 방법.

【청구항 3】

제 1 항에 있어서, 상기 절단된 위치는 획득한 염기서열 데이터 (sequence read)를 정렬하여 5' 말단이 수직 정렬된 위치, 또는 5' 말단 플롯에서 이중 피크 패턴을 보이는 위치인 것인, 방법.

20 【청구항 4】

제 1 항에 있어서, 상기 유전체 DNA 는 표적 특이적 뉴클레아제가 발현되는 세포 또는 발현되지 않는 세포로부터 분리된 것인, 방법.

【청구항 5】

제 3 항에 있어서, 상기 정렬은 표준 염기서열 (reference genome)로
25 염기서열 데이터를 맵핑한 뒤, BWA/GATK 또는 ISAAC 을 이용하여 수행되는 것인, 방법.

【청구항 6】

제 3 항에 있어서, 왓슨 가닥 (Watson strand)과 크릭 가닥 (Crick strand)에 해당하는 염기서열 데이터 (sequence read)가 각각 두 개 이상씩
30 수직으로 정렬되는 위치를 비표적 위치인 것으로 판단하는 단계를 추가로 포함하는 것인, 방법.

【청구항 7】

제 3 항에 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 것인, 방법.

【청구항 8】

제 1 항에 있어서, 분리된 유전체 DNA 는 표적 특이적 뉴클레아제가 발현된 세포로부터 분리된 것이고, 상기 DNA 의 비표적 위치에서 인델 (Insertion and deletion)을 확인하여 비표적 효과를 확인하는 단계를 추가로 포함하는 것인, 방법.

【청구항 9】

제 8 항에 있어서, 상기 인델을 확인하는 것은 상기 비표적 위치에 대한 T7E1 분석, Cel-I 효소를 이용한 돌연변이 검출 분석 또는 표적화 딥시퀀싱 (targeted deep sequencing)을 수행하여 이루어지는 것인, 방법.

【청구항 10】

제 1 항에 있어서, 상기 비표적 위치는 표적 위치와 1 개 이상의 뉴클레오티드 불일치 (mismatch)를 가지는 것인, 방법.

【청구항 11】

제 1 항에 있어서, 상기 비표적 위치는 표적 위치와 1 내지 6 개의 뉴클레오티드 불일치 (mismatch)를 가지는 것인, 방법.

【청구항 12】

제 1 항에 있어서, 상기 (c) 단계는 각각의 절단된 위치에 하기 식을 적용하여 절단 점수를 산출하여 수행되는 것인, 표적 특이적 뉴클레아제의 비표적 위치 (off-target site)를 검출하는 방법.

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

【청구항 13】

제 12 항에 있어서, 상기 식에서 상수 C 가 1 일 때 산출된 점수가 2.5 점 이상이거나, 상기 점수가 0.1 점 이상이고 표적 서열과 10 개 이하의 미스매치를 가지고 PAM 서열을 포함하는 경우, 그 절단된 위치를 비표적 위치로 판단하는 단계를 추가로 포함하는 것인, 방법.

5 **【청구항 14】**

제 1 항에 있어서, 상기 표적 특이적 뉴클레아제는 2 개 이상의 표적 서열에 대한 표적 특이적 뉴클레아제를 혼합한 것인, 방법.

【청구항 15】

10 제 1 항에 있어서, 상기 표적 특이적 뉴클레아제는 2 내지 100 개의 표적 서열에 대한 표적 특이적 뉴클레아제를 혼합한 것인, 방법.

【청구항 16】

제 14 항에 있어서, 표적 위치와의 편집 거리 (edit distance)에 따라 비표적 위치를 분류하는 단계를 추가로 포함하는 것인, 방법.

【청구항 17】

15 제 1 항에 있어서, 상기 유전자 가위는 메가뉴클레아제 (meganuclease), ZFN (zinc finger nuclease), TALEN (transcription activator-like effector nuclease), 및 RGEN (RNA-guided engineered nuclease)로 이루어진 군에서 선택되는 것인, 방법.

【청구항 18】

20 제 17 항에 있어서, 상기 RGEN 은 Cas9 단백질 또는 Cpf1 단백질인, 방법.

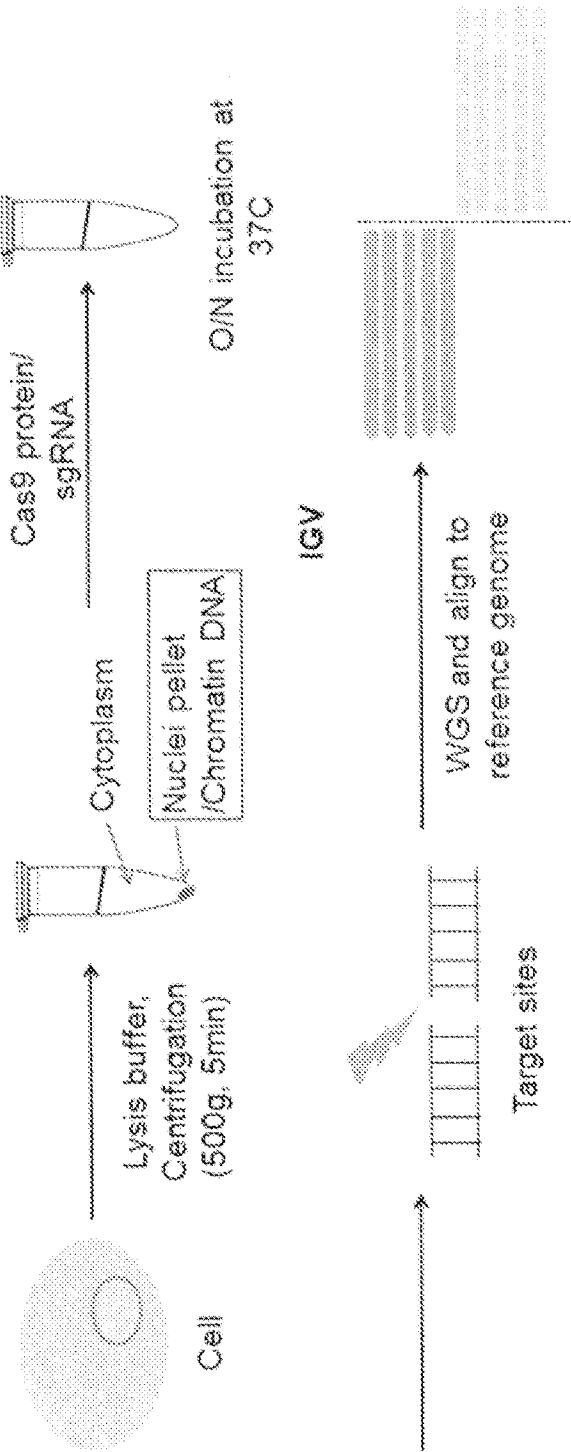
【청구항 19】

제 18 항에 있어서, 상기 RGEN 은 표적 유전자의 표적 서열에 혼성화 가능한 가이드 RNA 를 추가로 포함하는 것인, 방법.

25 **【청구항 20】**

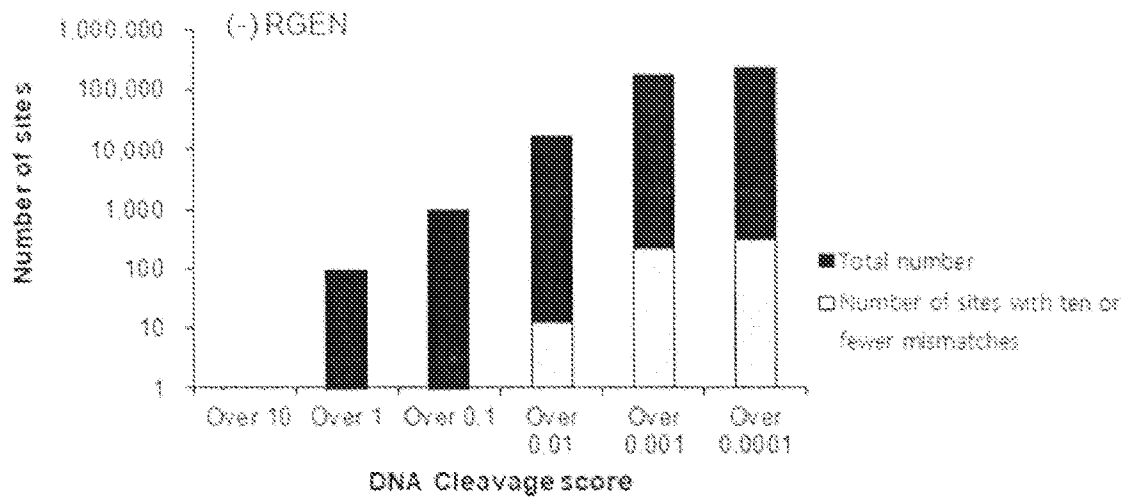
제 1 항 내지 제 19 항 중 어느 한 항에 있어서, 상기 크로마틴 DNA 는 히스톤, 비히스톤 단백질, 및 RNA 로 이루어진 비-DNA 염색질 성분 중에서 선택된 하나 이상을 함께 포함하는 형태의 DNA 인, 방법.

【도면 1】

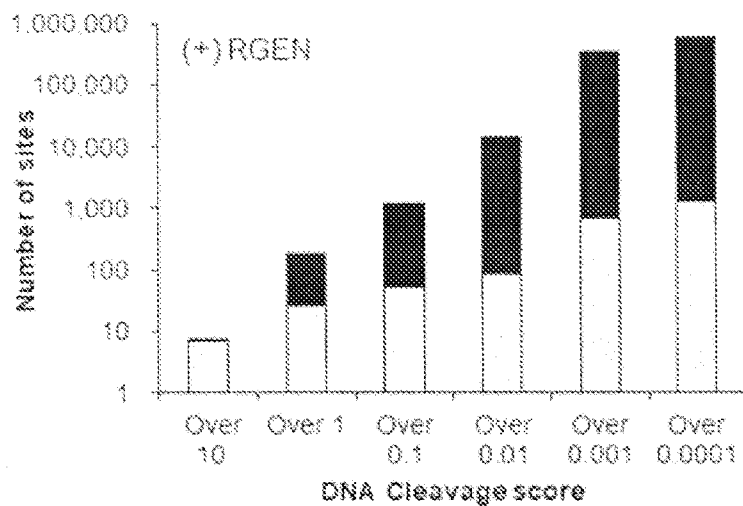


【도면 2】

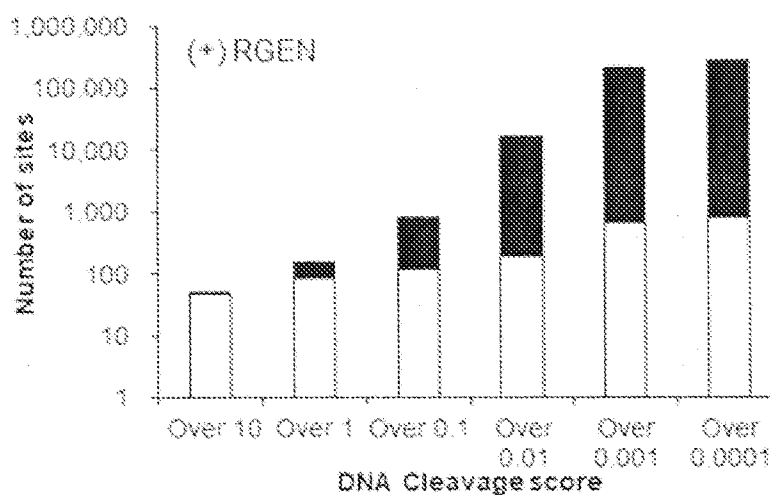
【図 2a】



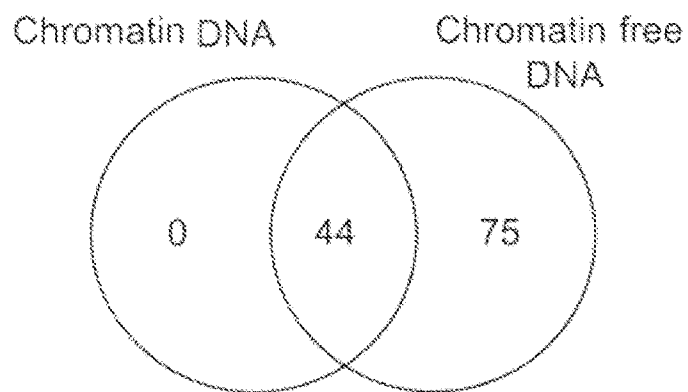
Chromatin free DNA



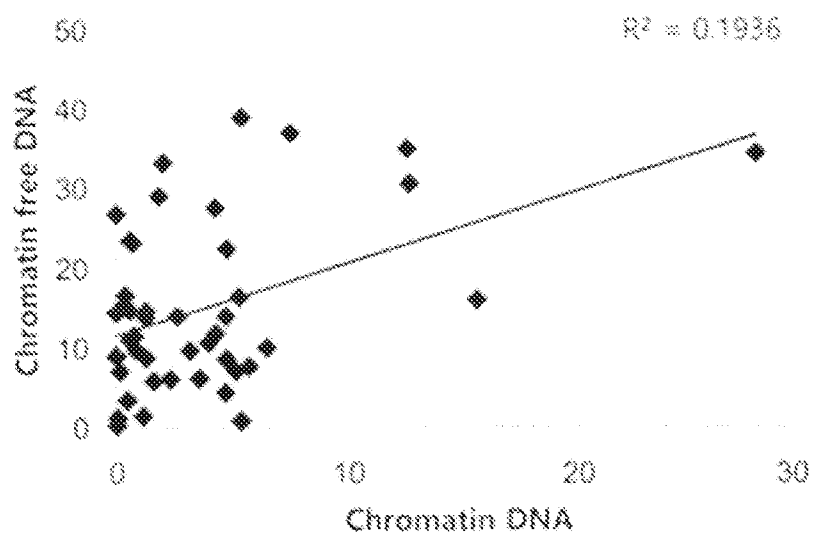
Chromatin DNA



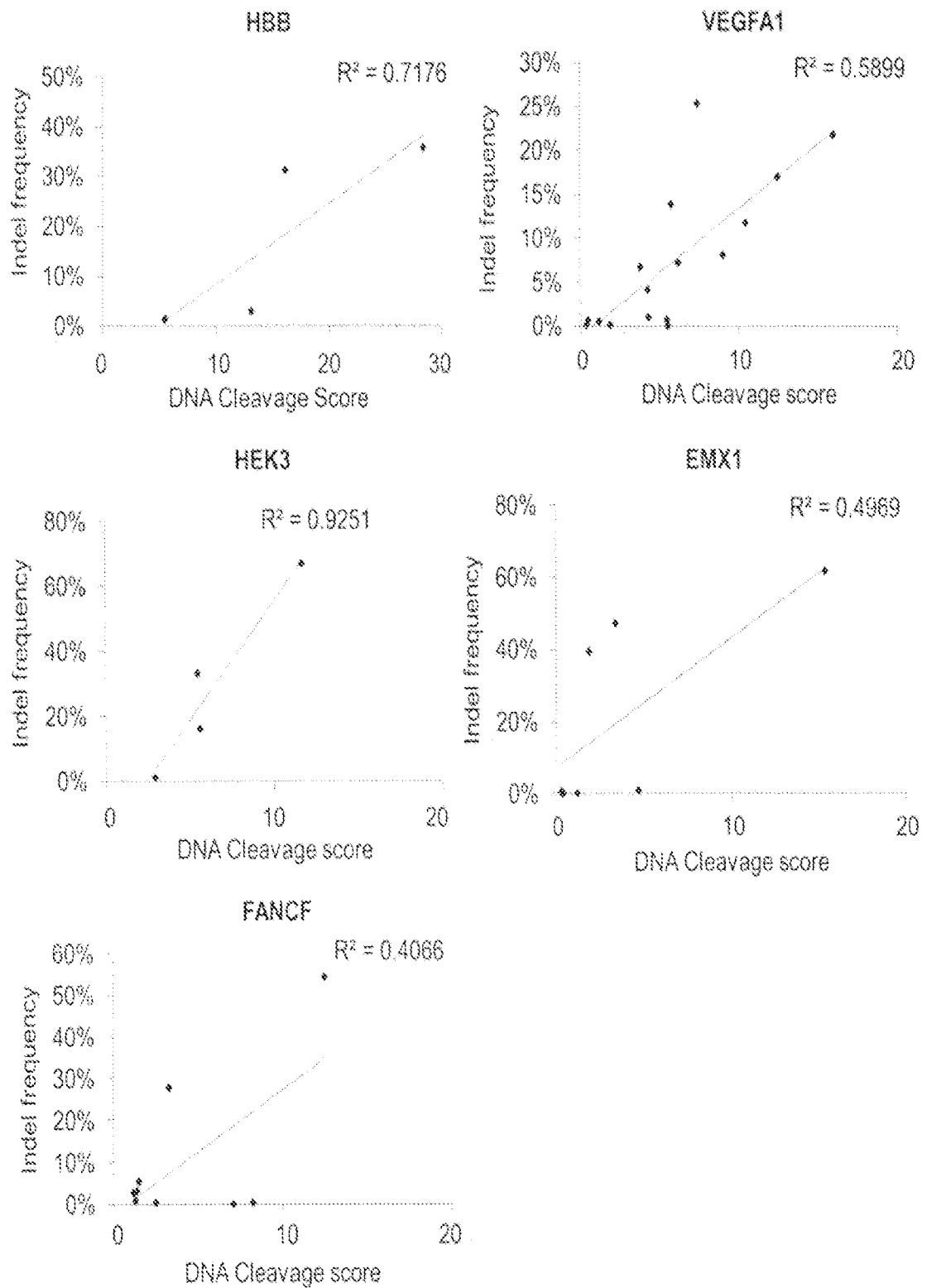
【図 2b】



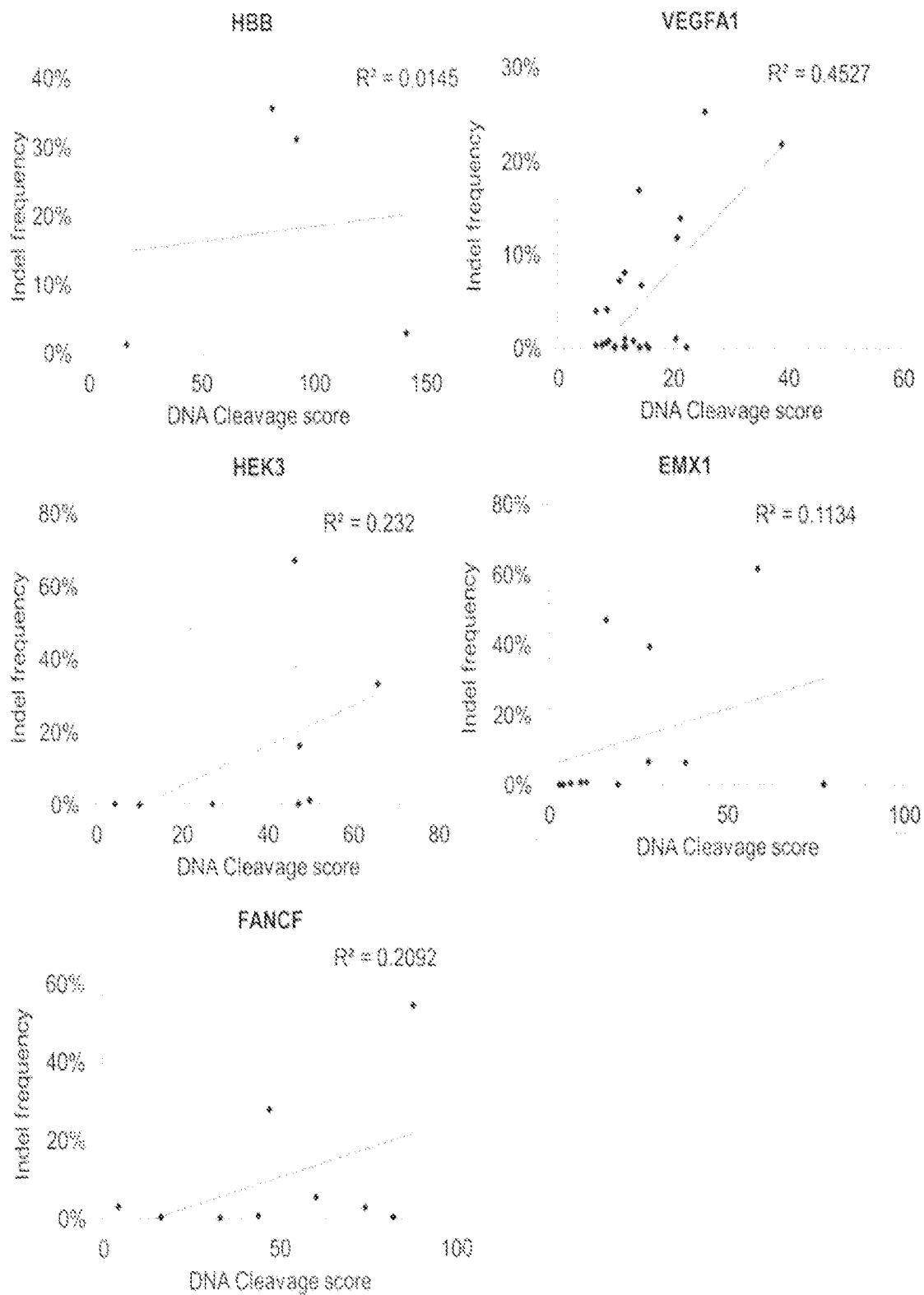
【図 2c】

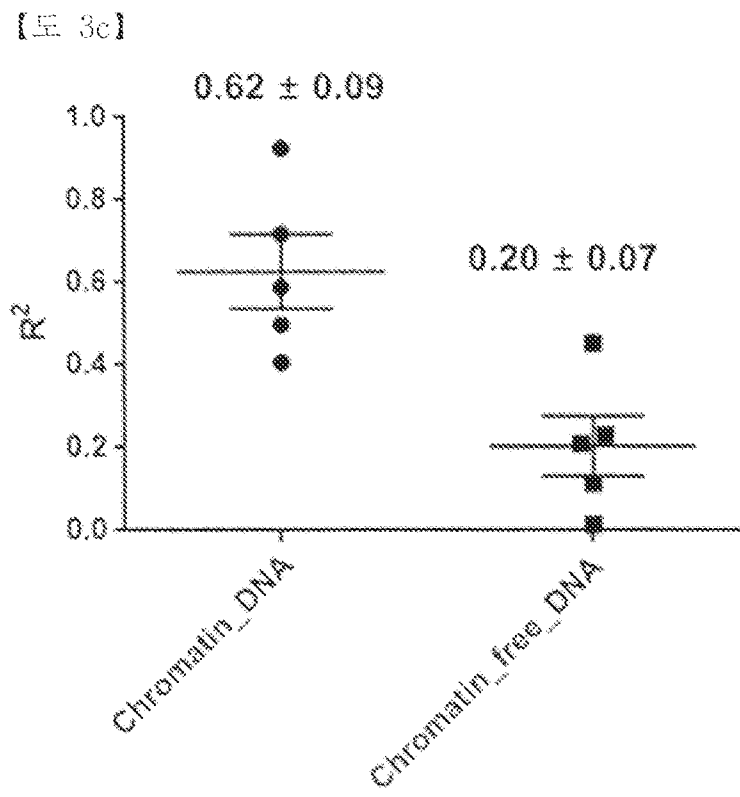


【図 3a】



【図 3b】





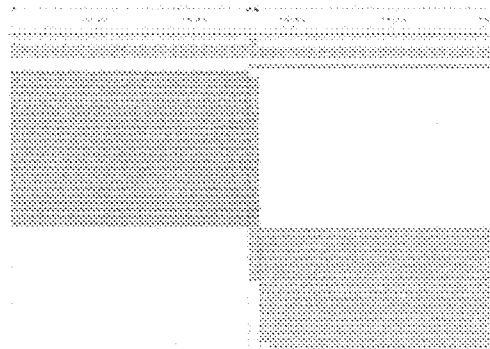
【도 4】

Score at position $i =$

$$\sum_{a=1}^5 \frac{100(F_i - 1)}{D_i} \times \frac{100(R_{i-4+a} - 1)}{D_{i-4+a}} \times (F_i + R_{i-4+a} - 2) \\ + \sum_{a=1}^5 \frac{100(R_{i-1} - 1)}{D_{i-1}} \times \frac{100(F_{i-3+a} - 1)}{D_{i-3+a}} \times (R_{i-1} + F_{i-3+a} - 2)$$

 F_i : Number of forward sequence reads starting at position i R_i : Number of reverse sequence reads starting at position i D_i : Sequencing depth at position i

Example :



		75006253	75006254	75006255	75006256	75006257	75006258	75006259
Count	Reverse	0	0	1	23	0	0	0
	Forward	0	0	0	9	9	1	0
Depth		23	23	23	31	18	19	19

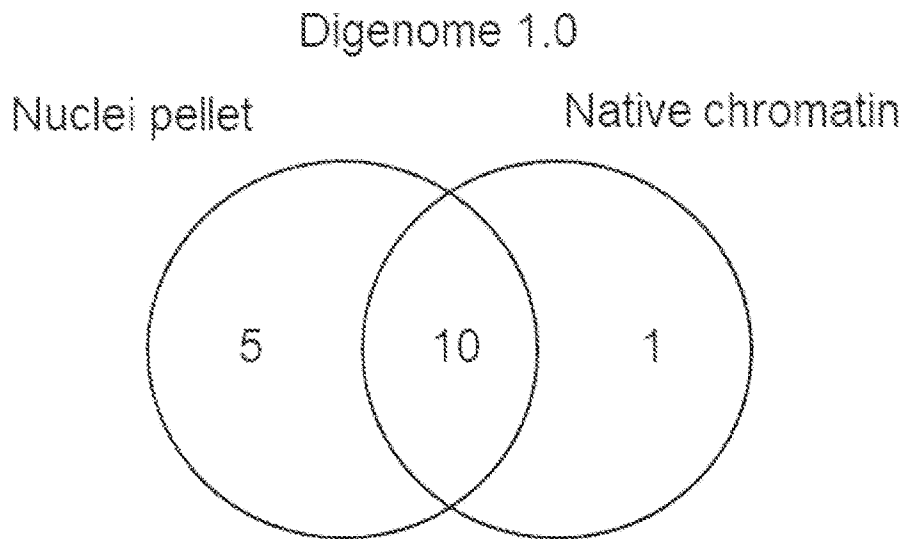
Score at position 75006256

$$= 10000 \times (22-1)/31 \times \{[(0-1)/23 \times (22+0-2) + (0-1)/31 \times (22+0-2) + (0-1)/18 \times (22+0-2) + (1-1)/19 \times (22+1-2) + (0-1)/19 \times (22+0-2)]$$

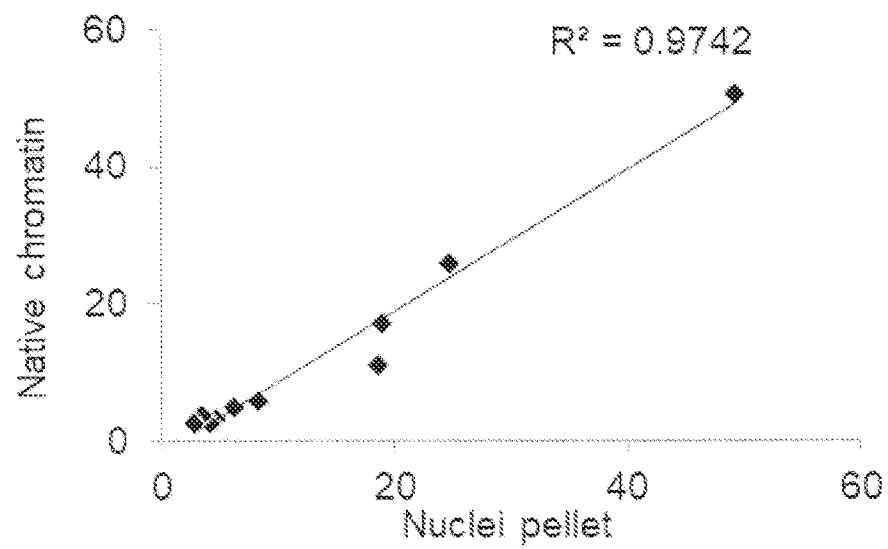
$$+ 10000 \times (9-1)/18 \times \{[(0-1)/23 \times (9+0-2) + (1-1)/23 \times (9+1-2) + (22-1)/31 \times (9+22-2) + (0-1)/18 \times (9+0-2) + (0-1)/19 \times (9+0-2)]$$

$$= 208.600$$

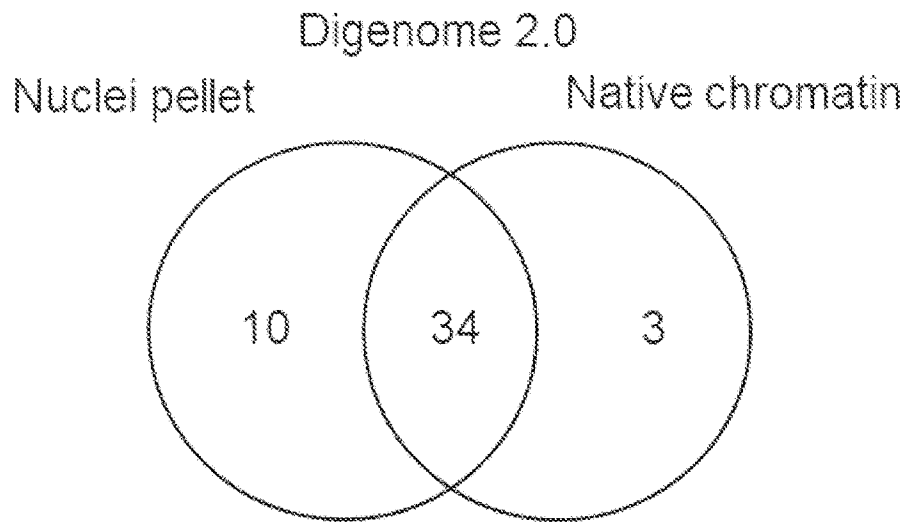
【図 5a】



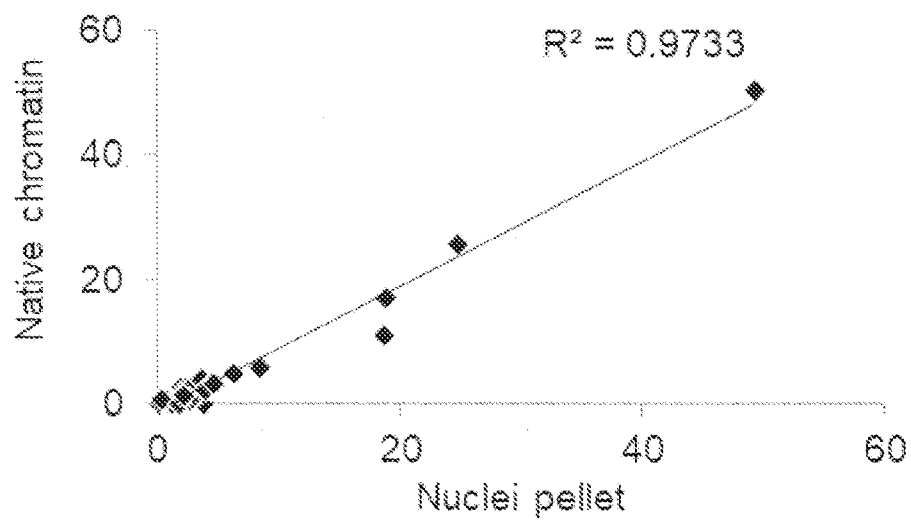
【図 5b】



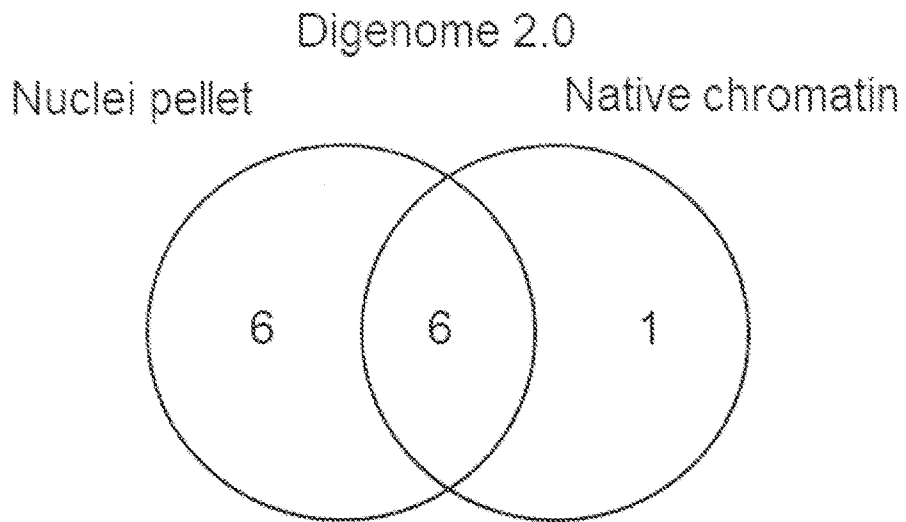
【도 6a】



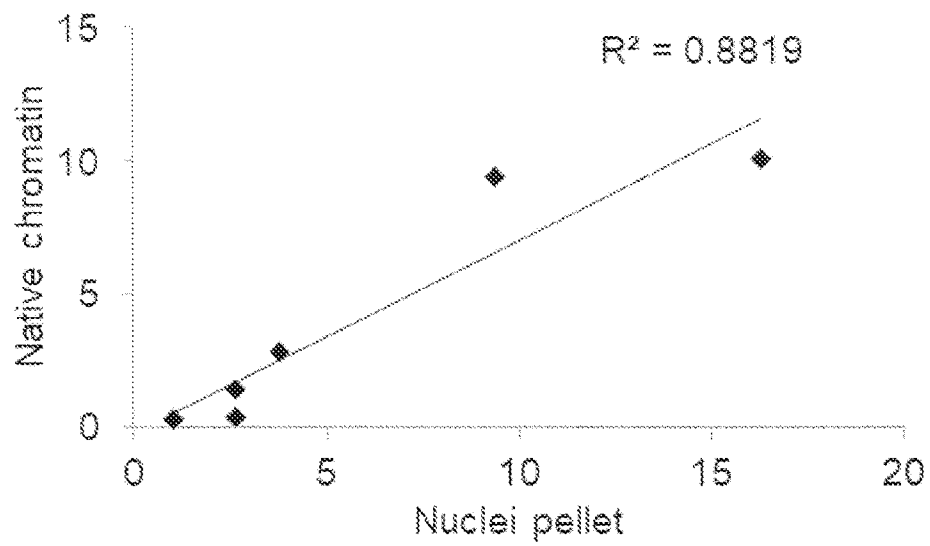
【도 6b】



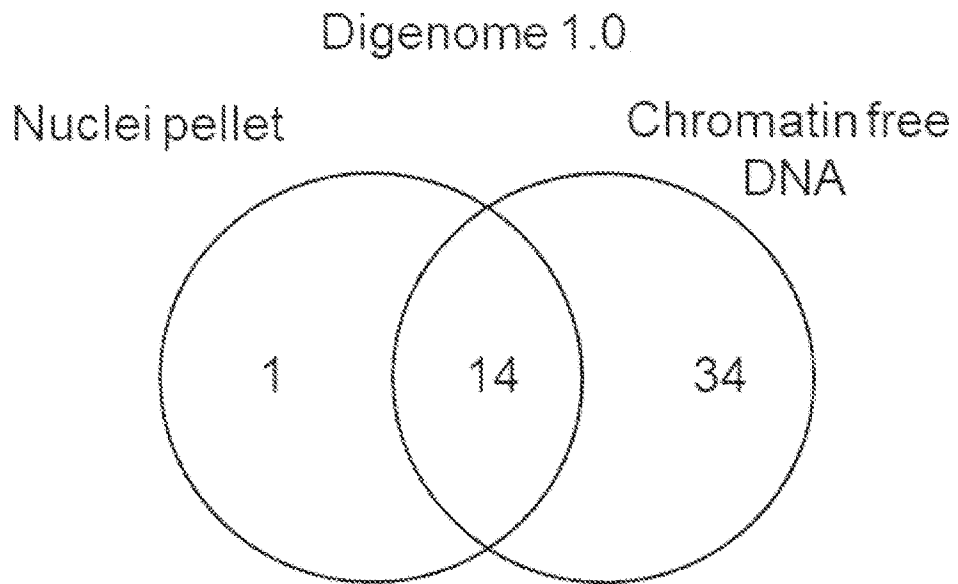
【図 7a】



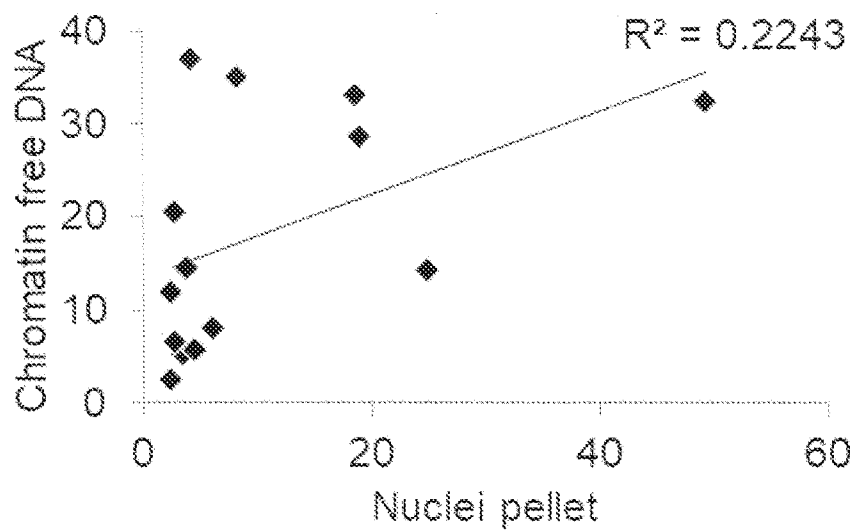
【図 7b】



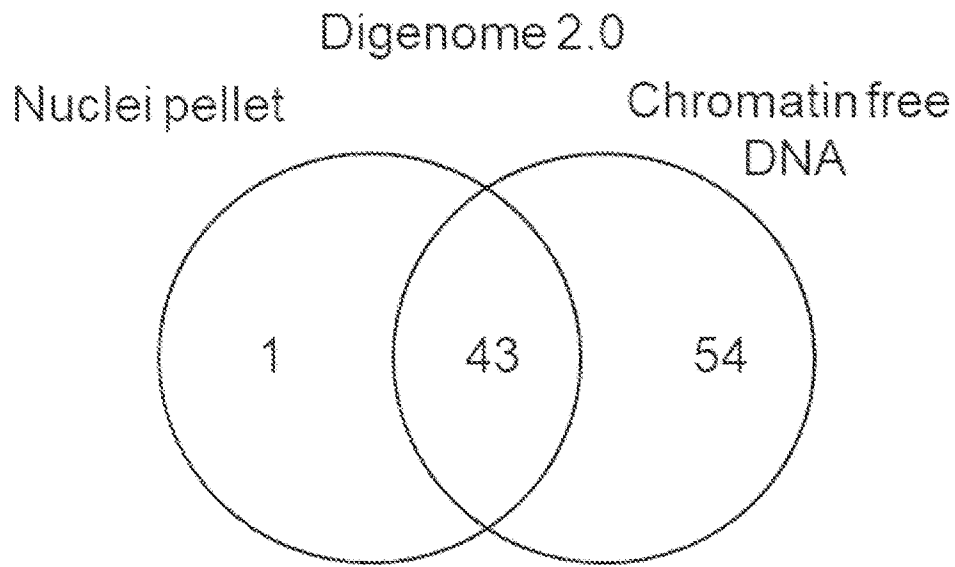
【図 8a】



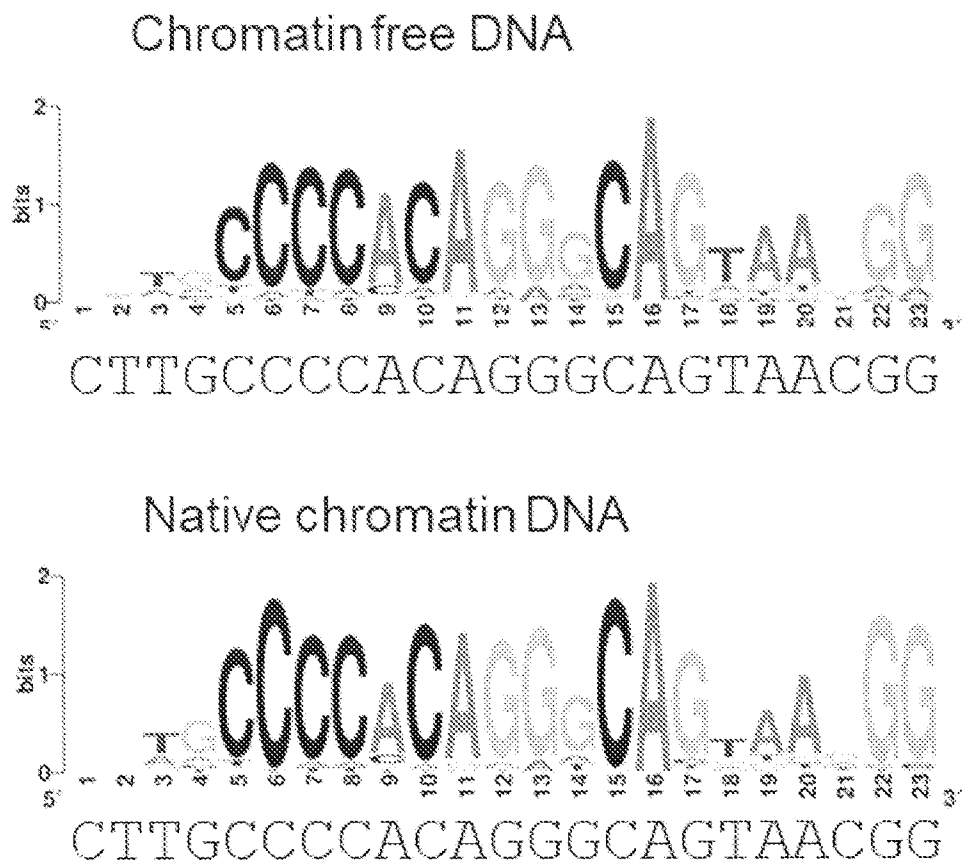
【図 8b】



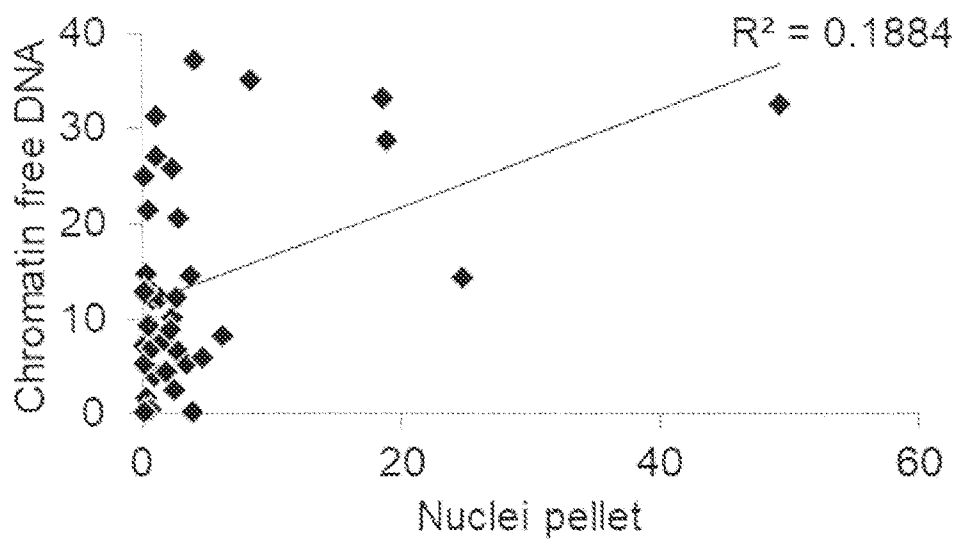
【도 9a】



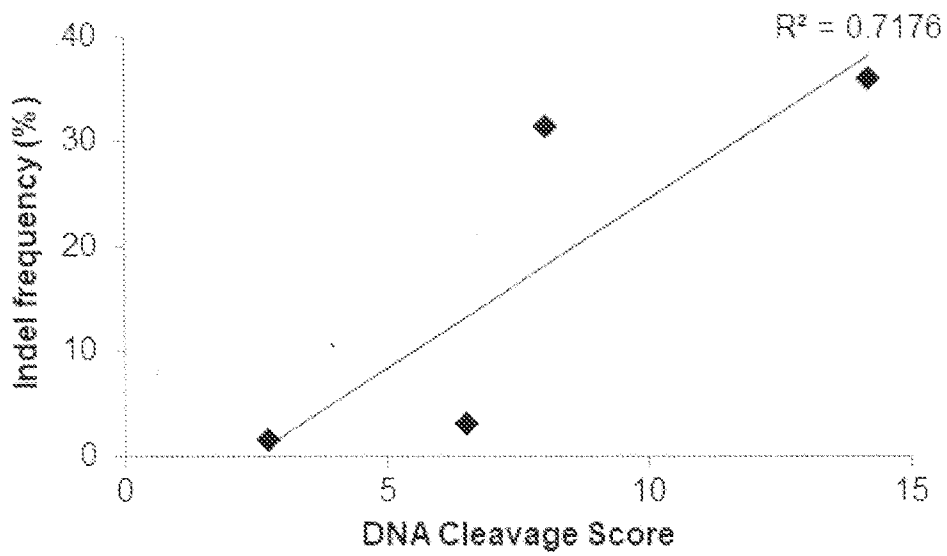
【도 9b】

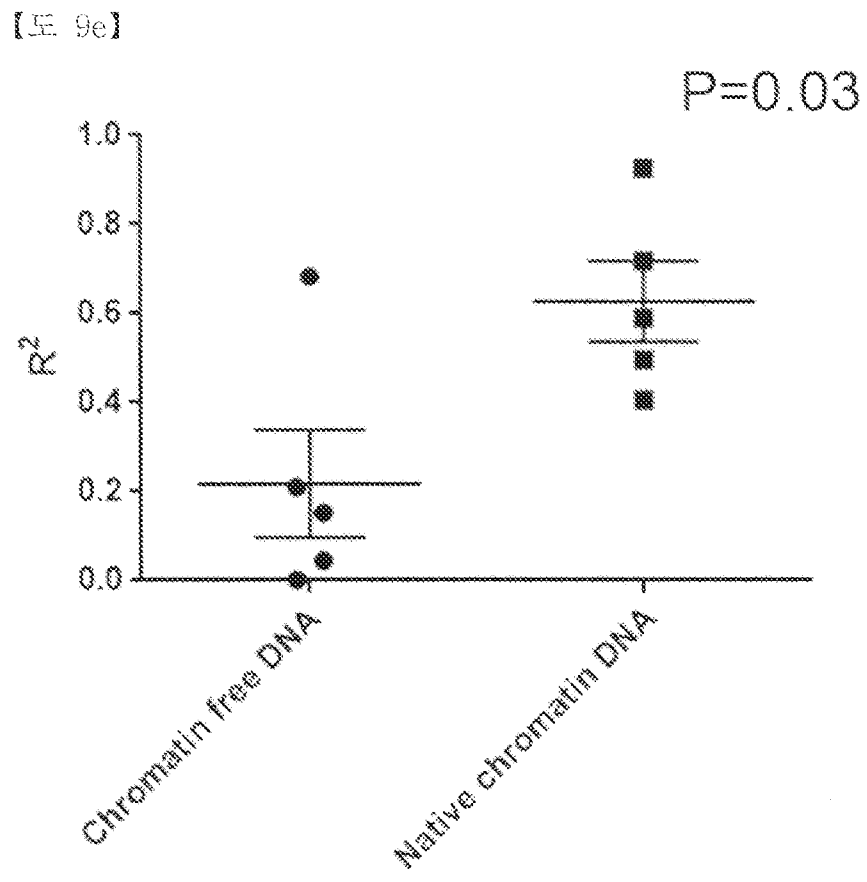


【図 9c】



【図 9d】





INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2017/013547

A. CLASSIFICATION OF SUBJECT MATTER

C12Q 1/68(2006.01)i, C12N 9/22(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 1/68; G06F 19/20; C12N 9/22

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: dielectric, nuclease, cutting, off-target, detection

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	KR 10-2016-0058703 A (INSTITUTE FOR BASIC SCIENCE) 25 May 2016 See claims 1-18.	1-20
Y	CHO, S. W. et al., "Analysis of Off-target Effects of CRISPR/Cas-derived RNA-guided Endonucleases and Nickases", Genome Research, 2014, vol. 24, pages 132-141 See pages 135-136.	1,2,4,8-11,14-19
Y	VERES, A. et al., "Low Incidence of Off-target Mutations in Individual CRISPR-Cas9 and TALEN Targeted Human Stem Cell Clones Detected by Whole-genome Sequencing", Cell Stem Cell, 2014, vol. 15, pages 27-30 See pages 27-29.	1,2,4,8-11,14-19
A	KUSCU, C. et al., "Genome-wide Analysis Reveals Characteristics of Off-target Sites Bound by the Cas9 Endonuclease", Nature Biotechnology, 2014, vol. 32, no. 7, pages 677-683 See the entire document.	1-20



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

22 FEBRUARY 2018 (22.02.2018)

Date of mailing of the international search report

22 FEBRUARY 2018 (22.02.2018)

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex-Daejeon, 189 Seonsa-ro, Daejeon 302-701,
Republic of Korea

Facsimile No. +82-42-481-8578

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2017/013547

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	PATTANAYAK, V. et al., "High-throughput Profiling of Off-target DNA Cleavage Reveals RNA-programmed Cas9 Nuclease Specificity", Nature Biotechnology, 2013, vol. 31, no. 9, pages 839-843 See the entire document.	1-20

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2017/013547

Patent document cited in search report	Publication date	Patent family member	Publication date
KR 10-2016-0058703 A	25/05/2016	CN 107109486 A	29/08/2017
		EP 3219810 A1	20/09/2017
		JP 2017-533724 A	16/11/2017
		WO 2016-076672 A1	19/05/2016

A. 발명이 속하는 기술분류(국제특허분류(IPC))

C12Q 1/68(2006.01)i, C12N 9/22(2006.01)i

B. 조사된 분야

조사된 최소문헌(국제특허분류를 기재)

C12Q 1/68; G06F 19/20; C12N 9/22

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: 유전체, 뉴클레아제, 절단, 비표적, 검출

C. 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
X	KR 10-2016-0058703 A (기초과학연구원) 2016.05.25 청구항 1-18 참조.	1-20
Y	CHO, S. W. 등, `Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases` Genome Research, 2014년, 24권, 페이지 132-141 페이지 135-136 참조.	1,2,4,8-11,14-19
Y	VERES, A. 등, `Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing` Cell Stem Cell, 2014년, 15권, 페이지 27-30 페이지 27-29 참조.	1,2,4,8-11,14-19
A	KUSCU, C. 등, `Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease` Nature Biotechnology, 2014년, 32권, 7호, 페이지 677-683 전체 문헌 참조.	1-20

☒ 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

2018년 02월 22일 (22.02.2018)

국제조사보고서 발송일

2018년 02월 22일 (22.02.2018)

ISA/KR의 명칭 및 우편주소



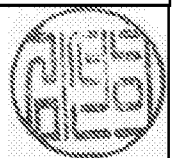
대한민국 특허청
(35208) 대전광역시 서구 청사로 189,
4동 (둔산동, 정부대전청사)

팩스 번호 +82-42-481-8578

심사관

김선희

전화번호 +82-42-481-5405



C(계속). 관련 문헌		
카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
A	PATTANAYAK V. 등, `High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity` Nature Biotechnology, 2013년, 31권, 9호, 페이지 839-843 전체 문헌 참조.	1-20

국제조사보고서에서
인용된 특허문헌

공개일

대응특허문헌

공개일

KR 10-2016-0058703 A

2016/05/25

CN 107109486 A

2017/08/29

EP 3219810 A1

2017/09/20

JP 2017-533724 A

2017/11/16

WO 2016-076672 A1

2016/05/19