

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7558929号
(P7558929)

(45)発行日 令和6年10月1日(2024.10.1)

(24)登録日 令和6年9月20日(2024.9.20)

(51)国際特許分類

F I

C 1 2 N	15/09	(2006.01)	C 1 2 N	15/09	1 1 0
C 1 2 N	5/10	(2006.01)	C 1 2 N	5/10	Z N A
C 1 2 N	15/11	(2006.01)	C 1 2 N	15/11	Z
C 0 7 K	14/47	(2006.01)	C 0 7 K	14/47	
A 6 1 P	3/00	(2006.01)	A 6 1 P	3/00	

請求項の数 25 (全190頁) 最終頁に続く

(21)出願番号	特願2021-513762(P2021-513762)
(86)(22)出願日	令和1年5月11日(2019.5.11)
(65)公表番号	特表2021-523736(P2021-523736 A)
(43)公表日	令和3年9月9日(2021.9.9)
(86)国際出願番号	PCT/US2019/031896
(87)国際公開番号	WO2019/217941
(87)国際公開日	令和1年11月14日(2019.11.14)
審査請求日	令和4年5月11日(2022.5.11)
(31)優先権主張番号	62/670,498
(32)優先日	平成30年5月11日(2018.5.11)
(33)優先権主張国・地域又は機関	米国(US)
(31)優先権主張番号	62/780,864
(32)優先日	平成30年12月17日(2018.12.17)

最終頁に続く

(73)特許権者	520439645 ビーム セラピューティクス インク . アメリカ合衆国 0 2 1 4 2 マサチュー セッツ州 ケンブリッジ メイン ストリ ート 2 3 8 9 ス フロア
(74)代理人	100083806 弁理士 三好 秀和
(74)代理人	100095500 弁理士 伊藤 正和
(74)代理人	100111235 弁理士 原 裕子
(74)代理人	100195257 弁理士 大淵 一志
(72)発明者	エヴァンス、 ジョン アメリカ合衆国 0 2 1 3 9 マサチュー 最終頁に続く

(54)【発明の名称】 プログラム可能塩基エディターシステムを用いて病原性変異を抑制する方法

(57)【特許請求の範囲】

【請求項 1】

アルファ - 1 アンチトリプシン欠乏症 (A1AD) に関連する一塩基多型 (SNP) を含む SERPINA1 ポリヌクレオチドを編集するインビトロまたはエクスピボの方法であって、前記 SNP は SERPINA1 ポリヌクレオチド中にありアミノ酸位置 342 にリジン (R) を有するかまたはアミノ酸位置 264 にバリン (V) を有するアルファ - 1 アンチトリプシン (A1AT) タンパク質の発現をもたらすものであり、前記方法は、

前記 SERPINA1 ポリヌクレオチドを、1 つ以上のガイドポリヌクレオチドと複合体をなす塩基エディターと接触させることを含み、

前記塩基エディターは、核酸プログラム可能な DNA 結合タンパク質 (napDNAbp) ドメインとシチジンデアミナーゼドメインとを含み、前記 1 つ以上のガイドポリヌクレオチドは前記塩基エディターをターゲティングして位置 1455 における前記 SERPINA1 ポリヌクレオチドのシチジンを脱アミノ化し、それによって前記 SERPINA1 ポリヌクレオチドにコードされる A1AT タンパク質のアミノ酸位置 374 においてメチオニンからイソロイシンへの変異を誘導する、

方法。

【請求項 2】

前記接触させることは、細胞中である、請求項 1 に記載の方法。

【請求項 3】

前記 napDNAbp ドメインは、5'-NGG-3' または 5'-GGG-3' から選択される PAM 配列に 20

対する特異性を有するStreptococcus pyogenes Cas9 (SpCas9) ポリペプチドである、請求項 1 に記載の方法。

【請求項 4】

前記napDNAbpドメインはヌクレアーゼ不活性バリエントまたはニッカーゼバリエントである、請求項 1 または 2 に記載の方法。

【請求項 5】

前記シチジンデアミナーゼドメインはAPOBECデアミナーゼドメインであり、前記塩基エディターはBE4である、請求項 1 ~ 4 のいずれか一項に記載の方法。

【請求項 6】

前記 1 つ以上のガイドポリヌクレオチドが、CRISPR RNA (crRNA) とトランスコード化小RNA (tracrRNA) とを含み、前記crRNAは、A1ADに関連するSNPを含む前記SERPINA1核酸配列に対して相補的な核酸配列を含む、請求項 1 ~ 5 のいずれか一項に記載の方法。

10

【請求項 7】

インビトロまたはエクスピボの、細胞またはその前駆細胞であって、

アルファ - 1 アンチトリプシン欠乏症 (A1AD) に関連するSNPであって、SERPINA1ポリヌクレオチド中にありアミノ酸位置342にリジンを含みまたはアミノ酸位置264にバリンを含み、アルファ - 1 アンチトリプシン (A1AT) タンパク質の発現をもたらす、SNPと、

核酸プログラム可能なDNA結合タンパク質 (napDNAbp) ドメインおよびシチジンデアミナーゼドメインを含む塩基エディターまたはそれをコードするポリヌクレオチドと、

20

前記塩基エディターをターゲティングして前記SERPINA1ポリヌクレオチドの核酸位置1455におけるシチジンを脱アミノ化させ、それによって細胞中で前記SERPINA1ポリヌクレオチドにコードされるアルファ - 1 アンチトリプシン (A1AT) タンパク質のアミノ酸位置374においてメチオニンからイソロイシンへの変異を誘導する、1 つ以上のガイドポリヌクレオチドと

を含む、細胞またはその前駆細胞。

【請求項 8】

前記細胞はA1ADを有する対象からの肝細胞であり、前記肝細胞はA1ATポリペプチドを発現する、請求項 7 に記載の細胞。

30

【請求項 9】

前記napDNAbpドメインは、5'-NGG-3' または5'-GGG-3' から選択されるPAM配列に対する特異性を有するStreptococcus pyogenes Cas9 (SpCas9) ポリペプチドである、請求項 7 または 8 に記載の細胞。

【請求項 10】

前記napDNAbpドメインは、ヌクレアーゼ不活性バリエントまたはニッカーゼバリエントである、請求項 7 ~ 9 のいずれか一項に記載の細胞。

【請求項 11】

前記シチジンデアミナーゼドメインはAPOBECデアミナーゼドメインであり、前記塩基エディターはBE4である、請求項 7 ~ 10 のいずれか一項に記載の細胞。

40

【請求項 12】

前記 1 つ以上のガイドポリヌクレオチドはCRISPR RNA (crRNA) およびトランスコード化小RNA (tracrRNA) を含み、前記crRNAは、前記SERPINA1ポリヌクレオチドの核酸位置1455におけるシチジンに対して相補的な核酸配列を含む、請求項 7 ~ 11 のいずれか一項に記載の細胞。

【請求項 13】

対象におけるA1ADを治療する方法における使用のための、請求項 7 または 8 に記載の細胞を含む組成物であって、前記方法は、請求項 7 または 8 に記載の細胞を前記対象に投与することを含み、組成物。

【請求項 14】

50

細胞においてSERPINA1ポリヌクレオチドを編集するインビトロまたはエクスビボの方法における使用のための、塩基エディターまたはそれをコードするポリヌクレオチドを含む組成物であって、前記SERPINA1ポリヌクレオチドはアルファ - 1 アンチトリプシン欠乏症 (A1AD) に関連する一塩基多型 (SNP) を含むものであり、前記SNPはSERPINA1ポリヌクレオチド中にありアミノ酸位置342にリジンを有するかまたはアミノ酸位置264にバリンを有するアルファ - 1 アンチトリプシン (A1AT) タンパク質の発現をもたらすものであり、前記方法は、

核酸プログラム可能なDNA結合タンパク質 (napDNAbp) ドメインおよびシチジンデアミナーゼドメインを含む、塩基エディターまたはそれをコードするポリヌクレオチドと；

前記塩基エディターをターゲティングして前記SERPINA1ポリヌクレオチドの核酸位置1455におけるシチジンの改変をもたらす、それによって前記SERPINA1ポリヌクレオチドにコードされるアルファ - 1 アンチトリプシン (A1AT) タンパク質のアミノ酸位置374においてメチオニンからイソロイシンへの変異を誘導する、1つ以上のガイドポリヌクレオチドと

を、前記細胞に送達することを、含む、

組成物。

【請求項15】

前記細胞は、アミノ酸位置342にリジンを有するかまたはアミノ酸位置264にバリンを有するA1ATタンパク質を発現する対象から得られたものであり、前記細胞は、肝細胞であるかまたは肝細胞の前駆細胞である、請求項14に記載の組成物。

【請求項16】

前記napDNAbpドメインは、5'-NGG-3' または 5'-GGG-3' から選択されるPAM配列に対する特異性を有するStreptococcus pyogenes Cas9 (SpCas9) ポリペプチドを含む、請求項14または15に記載の組成物。

【請求項17】

前記napDNAbpドメインは、ヌクレアーゼ不活性バリエントまたはニックアーゼバリエントである、請求項14～16のいずれか一項に記載の組成物。

【請求項18】

前記塩基エディターはBE4である、請求項14～17のいずれか一項に記載の組成物。

【請求項19】

前記塩基エディターは、前記SERPINA1ポリヌクレオチドの核酸位置1455におけるシチジンに対して相補的な核酸配列を含む単一ガイドRNA (sgRNA) と複合体をなす、請求項14～18のいずれか一項に記載の組成物。

【請求項20】

肝細胞を産生するエクスビボまたはインビトロの方法であって、前記方法は、

(a) アルファ - 1 アンチトリプシン欠乏症 (A1AD) に関連するSNPであって、SERPINA1ポリヌクレオチド中にありアミノ酸位置342にリジンを有するかまたはアミノ酸位置264にバリンを有するアルファ - 1 アンチトリプシン (A1AT) タンパク質の発現をもたらすものである、SNPを含む肝前駆細胞に、

核酸プログラム可能なDNA結合タンパク質 (napDNAbp) ドメインおよびシチジンデアミナーゼドメインを含む塩基エディター、またはそれをコードするポリヌクレオチドと、

前記塩基エディターをターゲティングして前記SERPINA1ポリヌクレオチドの核酸位置1455におけるシチジンにおいてシチジン脱アミノ化をもたらす、それによって前記SERPINA1ポリヌクレオチドにコードされるアルファ - 1 アンチトリプシン (A1AT) タンパク質のアミノ酸位置374においてメチオニンからイソロイシンへの変異を誘導する、1つ以上のガイドポリヌクレオチドと

を導入すること；および

(b) 前記肝前駆細胞を肝細胞に分化させること

を含む、方法。

【請求項21】

10

20

30

40

50

前記肝前駆細胞はA1ATポリペプチドを発現し、A1ADを有する対象から得られるものである、請求項20に記載の方法。

【請求項22】

前記napDNAbpドメインは、5'-NGG-3' または 5'-GGG-3' から選択されるPAM配列に対する特異性を有するStreptococcus pyogenes Cas9 (SpCas9) ポリペプチドを含む、請求項20または21に記載の方法。

【請求項23】

前記napDNAbpドメインは、ヌクレアーゼ不活性バリエーションまたはニックアーゼバリエーションである、請求項20～22のいずれか一項に記載の方法。

【請求項24】

前記シチジンデアミナーゼドメインは、APOBECデアミナーゼドメインであり、前記塩基エディターはBE4である、請求項20～23のいずれか一項に記載の方法。

【請求項25】

前記塩基エディターおよび1つ以上のガイドポリヌクレオチドは、前記細胞中で複合体を形成し、前記塩基エディターは、前記SERPINA1ポリヌクレオチドの核酸位置1455におけるシチジンに対して相補的な核酸配列を含む単一ガイドRNA (sgRNA) と複合体をなす、請求項20～24のいずれか一項に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

関連出願

本出願は、2018年5月11日出願の米国仮出願第62/670,498号、および2018年12月17日出願の米国仮出願第62/780,864号の利益を主張するものであり、それぞれの内容全体が参照により本明細書に組み込まれる。

参照による組み込み

本明細書に記載されているすべての刊行物、特許、および特許出願は、それぞれの個々の刊行物、特許、または特許出願が具体的かつ個別に参照により組み込まれるように示されているのと同じ程度に、参照により本明細書に組み込まれる。別段の表示がない限り、本明細書に記載されている刊行物、特許および特許出願は、その全体が参照により本明細書に組み込まれる。

【0002】

ほとんどの既知の遺伝性疾患では、疾患の根底にある原因を研究または対処するために、遺伝子の確率的破壊ではなく、標的遺伝子座における点突然変異の補正が必要となる。

【背景技術】

【0003】

Clustered regularly interspaced short palindromic repeat (CRISPR) 系を用いた現在のゲノム編集技術は、遺伝子補正の第一段階として標的遺伝子座で二本鎖DNA切断を導入する。二本鎖DNA切断に反応して、細胞のDNA修復プロセスは、ほとんどの場合、DNA切断部位での非相同末端結合によるランダムな挿入または欠失 (インデル) をもたらす。ほとんどの遺伝性疾患は点突然変異から生じるが、点突然変異補正に対する現在のアプローチは非効率的であり、典型的には、dsDNA切断に対する細胞応答から生じる標的遺伝子座でのランダムな挿入および欠失 (インデル) を多く誘導する。したがって、より効率的で、確率的な挿入もしくは欠失 (インデル) または転座などの望ましくない産物のはるかに少ない、改良された形態のゲノム編集が必要とされている。

【0004】

1 - アンチトリプシン欠乏症 (A1AD) は、1 - アンチトリプシン (A1AT) タンパク質をコードするSERPINA1遺伝子における病理性変異が、疾患を有する個体におけるタンパク質産生の減少をもたらす遺伝的疾患である。A1ATは好中球エラスターゼの特に良好な阻害因子であり、肺のような組織および器官をエラスチン分解から保護する。結果として、A1ADを有する患者の肺におけるエラスチンは好中球エラスターゼによってより容

10

20

30

40

50

易に分解され、時間が経つと、肺弾性の消失が慢性閉塞性肺疾患（COPD）へと発達する。健常な個体においては、A1ATは肝臓内の肝細胞において産生されて、全身循環へと分泌され、そこでタンパク質はプロテアーゼ阻害因子として機能する。

【0005】

最も一般的な病原性A1ATバリエーションは、SERPINA1遺伝子におけるグアニンからアデニンへ（G A）の変異であり、これはA1ATタンパク質の第342アミノ酸におけるグルタミン酸からリジンへの置換をもたらす。この置換はタンパク質のミスフォールドと肝細胞内でのその重合を引き起こし、究極的には、毒性の凝集物が肝臓にダメージおよび硬変をもたらす。肝毒性は遺伝子ノックアウト（CRISPR/ZFN/TALEN）または遺伝子ノックダウン（siRNA）によって対処できる可能性があるが、これらいずれのアプローチも肺での病理に対する対処にはならない。肺病理はタンパク質補充療法によって対処し得るが、この療法は肝毒性に対する対処にはならない。遺伝子療法も、A1AT遺伝的欠損の対処としては不適切である。A1ADの患者の肝臓は内因性A1AT凝集により引き起こされる重い疾患負担をすでに抱えているため、肝臓においてA1ATを増加させる遺伝子療法は逆効果となる。従って、該疾患に伴う肺病理と肝毒性の両方に対処する、A1ADの患者を治療する方法が必要とされている。

【発明の概要】

【0006】

本明細書では、対象における遺伝的障害を治療する方法が提供され、この方法は、塩基エディター、または塩基エディターをコードするポリヌクレオチドを、必要とする対象に投与すること（ここで塩基エディターは、ポリヌクレオチドプログラム可能なヌクレオチド結合ドメインおよびデアミナーゼドメインを含む）と；ガイドポリヌクレオチドを対象に投与すること（ここで、ガイドポリヌクレオチドは、塩基エディターを対象の標的ヌクレオチド配列へとターゲティング（標的指向化）する）と；標的ヌクレオチド配列への塩基エディターのターゲティングに際して、核酸塩基を脱アミノ化することにより標的ヌクレオチド配列の核酸塩基を編集し、それにより核酸塩基を別の核酸塩基に変化させることにより遺伝的障害を治療することとを含み、ここで、該核酸塩基はポリヌクレオチドのタンパク質コード領域中にあり、該核酸塩基は該遺伝的障害の原因ではない（すなわち、該核酸塩基は、該遺伝的疾患を引き起こす変異をコードしてはいない）。

【0007】

本明細書では、必要とする対象において遺伝的障害を治療するための細胞、組織、または器官を産生する方法も提供され、該方法は、該細胞、組織、または器官を、塩基エディター、または該塩基エディターをコードするポリヌクレオチドと接触させること（ここで塩基エディターは、ポリヌクレオチドプログラム可能なヌクレオチド結合ドメインおよびデアミナーゼドメインを含む）と；該細胞、組織、または器官を、ガイドポリヌクレオチドに接触させること（ここでガイドポリヌクレオチドは、該細胞、組織、または器官の標的ヌクレオチド配列へと塩基エディターをターゲティングする）と；標的ヌクレオチド配列への塩基エディターのターゲティングに際して、核酸塩基を脱アミノ化することにより標的ヌクレオチド配列の核酸塩基を編集し、それにより、核酸塩基を別の核酸塩基に変化させることにより遺伝的障害を治療するための細胞、組織、または器官を産生することとを含み、ここで、該核酸塩基はポリヌクレオチドのタンパク質コード領域中にあり、該核酸塩基は該遺伝的障害の原因ではない。いくつかの実施形態では、方法はさらに、細胞、組織、または器官を対象に投与することを含む。いくつかの実施形態では、細胞、組織、または器官は、対象にとって自家系（autologous）である。ある態様において、細胞、組織、または器官は、対象にとって同種異系（allogeneic）である。ある態様において、細胞、組織、または器官は、対象にとって異種系（xenogenic）である。

【0008】

いくつかの実施形態では、核酸塩基を別の核酸塩基に変化させることは、ポリヌクレオチドによってコードされるタンパク質の活性の増加をもたらす。いくつかの実施形態では、核酸塩基を別の核酸塩基に変化させることは、ポリヌクレオチドによってコードされる

10

20

30

40

50

タンパク質の折り畳みの改善および/または安定性の増加をもたらす。いくつかの実施形態では、核酸塩基を別の核酸塩基に変化させることは、ポリヌクレオチドによってコードされるタンパク質の発現の増加をもたらす。いくつかの実施形態では、タンパク質の発現の増加は、タンパク質の翻訳率改善によるものである。いくつかの実施形態では、タンパク質の発現の増加は、そのタンパク質を含有する細胞小器官または細胞コンパートメントからの放出率上昇によるものである。いくつかの実施形態では、タンパク質の発現の増加は、そのタンパク質のシグナルペプチドのプロセッシング率改善によるものである。いくつかの実施形態では、タンパク質の発現の増加は、そのタンパク質と別のタンパク質との相互作用の変化によるものである。

【0009】

いくつかの実施形態では、該核酸塩基は、遺伝的障害の原因となる遺伝子中に位置する。いくつかの実施形態では、編集は、該遺伝子に位置する複数の核酸塩基を編集することを含み、その複数の核酸塩基は遺伝的障害の原因ではない。いくつかの実施形態では、編集は、少なくとも1つの他の遺伝子中に位置する1つ以上の追加的核酸塩基を編集することをさらに含む。いくつかの実施形態では、該遺伝子および少なくとも1つの他の遺伝子は、タンパク質の1つ以上のサブユニットをコードする。いくつかの実施形態では、該核酸塩基は、本明細書中の表3Aおよび3Bに列記された遺伝子中にあり、編集が、表3Aおよび3Bに示された遺伝子によってコードされるタンパク質のアミノ酸変化をもたらす。

【0010】

いくつかの実施形態では、遺伝的障害は、網膜色素変性症、アッシャー症候群、鎌状赤血球症、ベータサラセミア、 α 1-アンチトリプシン欠乏症(A1AD)、肝性ポルフィリン症、中鎖アシルCoA脱水素酵素(MCAD)欠乏症、ライソゾーム酸性リパーゼ(LAL)欠乏症、フェニルケトン尿症、ヘモクロマトーシス、フォン・ギールケ病、ポンペ病、ゴーシェ病、ハーラー症候群、嚢胞性線維症、または慢性疼痛である。いくつかの実施形態では、遺伝的障害は α 1-アンチトリプシン欠乏症(A1AD)である。いくつかの実施形態では、塩基編集は、F51L、M374I、A348V、A347V、K387R、T59A、T68Aからなる群から選択される、 α 1-アンチトリプシン(A1AT)タンパク質のアミノ酸変化をもたらす。いくつかの実施形態では、塩基編集は、A1ATにおいてM374Iアミノ酸変化をもたらす。

【0011】

ある実施形態では、遺伝的障害は鎌状赤血球症である。ある実施形態では、編集は、HbA/HbSテトラマーの重合能を低減させるアミノ酸変化をもたらす。ある実施形態では、該核酸塩基はヘモグロビンのサブユニット(HbB)をコードするHBB遺伝子に位置する。ある実施形態では、HBB遺伝子は、鎌状赤血球ヘモグロビンアリル(HbS)である。ある実施形態では、編集は、ヘモグロビンのサブユニットにおけるアミノ酸変化をもたらす。ある実施形態では、ヘモグロビンのサブユニットにおけるアミノ酸変化は、A70T、A70V、L88P、F85L、F85P、E22G、G16D、G16N、またはこれらの組合せを含む。ある実施形態では、該核酸塩基はヘモグロビンのサブユニット(HbA)をコードするHBA1またはHBA2遺伝子に位置する。ある実施形態では、編集は、ヘモグロビンのサブユニットにおけるアミノ酸変化をもたらす。ある実施形態では、サブユニットのアミノ酸変化は、鎌状赤血球ヘモグロビンのサブユニットとサブユニットの重合インターフェースに位置する。ある実施形態では、ヘモグロビンのサブユニットにおけるアミノ酸変化は、K11E、D47G、Q54R、N68D、E116K、H20Y、H50Yまたはこれらの組合せを含む。

【0012】

一態様において、プログラム可能な核酸塩基エディターを用いて病原性変異を抑制するための組成物および方法が提供される。本発明は、塩基エディター(例えばBE4)を用いて内因性SERPINA1遺伝子に改変を誘導する、A1ADを治療する方法を提供する。改変されたSERPINA1遺伝子は、 α 1-アンチトリプシンタンパク質におけるE342Kを安定化させるM374I変異をコードする。BE4を用いてM374Iを導入することは、肝毒性の緩和と

10

20

30

40

50

肺へのA1AT循環の増加を同時にもたらし、それによって有害なE342K変異の存在について補償し得る。この戦略は、肝臓に対する病原性タンパク質負荷を除去すると同時に肺への機能的タンパク質を回復させる。

【0013】

別の態様において、本発明は、1 - アンチトリプシン欠乏症 (A1AD) に関連する一塩基多型 (SNP) を含有するSERPINA1ポリヌクレオチドを編集する方法を提供し、この方法は、1つ以上のガイドポリヌクレオチドと複合体をなす塩基エディターとSERPINA1ポリヌクレオチドを接触させることを含み、ここで塩基エディターはポリヌクレオチドプログラム可能なDNA結合ドメインとシチジンデアミナーゼドメインとを含み、1つ以上のガイドポリヌクレオチドが、塩基エディターをターゲティングしてA1ADに関連する一塩基多型 (SNP) の改変をもたらす。一実施形態において、接触は、細胞中、真核細胞中、哺乳類細胞中、またはヒト細胞中である。別の実施形態では、細胞は、インビボまたはエクスピボである。

10

【0014】

別の態様において、本発明は、塩基エディター、塩基エディターをコードするポリヌクレオチド (ここで塩基エディターはポリヌクレオチドプログラム可能なDNA結合ドメインとシチジンデアミナーゼドメインとを含む) と; 塩基エディターをターゲティングしてSERPINA1ポリヌクレオチドの核酸位置1455におけるシチジンを脱アミノ化させる1つ以上のガイドポリヌクレオチドと; を該細胞またはその前駆体に導入することにより産生される細胞を提供する。一実施形態では、産生される細胞は肝細胞である。別の実施形態では、該細胞またはその前駆体は、胚細胞、人工多能性幹細胞、または肝細胞である。別の実施形態では、肝細胞はA1ATポリペプチドを発現する。別の実施形態では、細胞はA1ADを有する対象からのものである。別の実施形態では、細胞は哺乳類細胞またはヒト細胞である。

20

【0015】

別の態様において、本発明は、前述のいずれかの態様の細胞を対象に投与することを含む、対象におけるA1ADを治療する方法を提供する。一実施形態では、細胞は対象にとって自家系 (autologous) である。別の実施形態において、細胞は、対象にとって同種異系 (allogeneic) である。

【0016】

別の態様において、本発明は、前述のいずれかの態様の細胞から増殖または拡張された単離細胞または細胞集団を提供する。

30

【0017】

別の態様において、本発明は、対象におけるA1ADを治療する方法を提供し、該方法は、塩基エディター、または塩基エディターをコードするポリヌクレオチド (ここで塩基エディターはポリヌクレオチドプログラム可能なDNA結合ドメインとシチジンデアミナーゼドメインとを含む) と;

塩基エディターをターゲティングしてSERPINA1ポリヌクレオチドの核酸位置1455におけるシチジンを改変させる、1つ以上のガイドポリヌクレオチドとを対象に投与することを含む。

40

【0018】

上記で説明した複数の態様の一実施形態において、対象は哺乳類またはヒトである。別の実施形態では、方法は、塩基エディター、または塩基エディターをコードするポリヌクレオチドと、1つ以上のガイドポリヌクレオチドとを、対象の細胞に送達することを含む。別の実施形態において、細胞は肝細胞である。別の実施形態において、細胞は肝細胞の前駆細胞である。別の実施形態において、肝細胞はA1ATタンパク質を発現する。

【0019】

別の態様において、肝細胞またはその前駆細胞を産生する方法が提供され、該方法は、(a) A1ADに関連するSNPを含む肝前駆細胞に、塩基エディターまたはそれをコードするポリヌクレオチド (ここで、塩基エディターは、ポリヌクレオチドプログラム可能なヌク

50

レオチド結合ドメインとシチジンデアミナーゼドメインとを含む)と; 1つ以上のガイドポリヌクレオチド(ここで、1つ以上のガイドポリヌクレオチドは、塩基エディターをターゲティングしてSERPINA1ポリヌクレオチドの核酸位置1455におけるシチジンにおいてシチジン脱アミノ化をもたらす)とを導入すること、および

(b) 前記肝前駆細胞を幹細胞に分化させること

を含む。一実施形態では、方法は、肝前駆細胞を幹細胞に分化させることを含む。別の実施形態では、肝前駆細胞はA1ATポリペプチドを発現する。別の実施形態では、肝前駆細胞は、A1ADを有する対象から得られるものである。別の実施形態では、肝前駆細胞は哺乳類細胞またはヒト細胞である。

【0020】

別の態様において、本発明は、

5' -CAAUCAUUAAGAAGACAAAGGGUUU-3'
 5' -UCAUCAUUAAGAAGACAAAGGGUUU-3'
 5' -UUCAUCAUUAAGAAGACAAAGGGUUU-3'
 5' -GUUCAUCAUUAAGAAGACAAAGGGUUU-3'
 5' -UGUUCAUCAUUAAGAAGACAAAGGGUUU-3'
 5' -UUGUUCAUCAUUAAGAAGACAAAGGGUU-3'
 5' -UUCAUCAUUAAGAAGACAAAG-3'
 5' -UUCAUCAUUAAGAAGACAAAGG-3'
 5' -UCAUCAUUAAGAAGACAAAGGG-3'
 5' -AAUCAUUAAGAAGACAAAGGGU-3'

から選択される核酸配列を含む、ガイドRNAを提供する。

【0021】

別の態様において、本発明は、本明細書に記述またはその他の説明がされた態様のガイドRNAの18、19、20、21、または22ヌクレオチドを含むガイドRNAを提供する。

【0022】

別の態様において、本発明は、本明細書に記述された態様の塩基エディターと本明細書に記載されたガイドRNAとを含む、タンパク質核酸複合体を提供する。

【0023】

上記態様のいずれかまたは本明細書に記載される発明のいずれかの他の態様において、塩基エディターは、SERPINA1ポリヌクレオチドの位置1455におけるシチジンを脱アミノ化し、それによってA1ATタンパク質のアミノ酸位置374においてメチオニンからイソロイシンへの変異を誘導する。上記態様のいずれかまたは本明細書に記載される発明のいずれかの他の態様において、A1ATポリペプチドは、アミノ酸位置342においてリジンを含み、および/またはアミノ酸位置376においてリジンを含む。上記態様のいずれかまたは本明細書に記載される発明のいずれかの他の態様において、ポリヌクレオチドプログラム可能なDNA結合ドメインはStreptococcus pyogenes Cas9 (SpCas9)またはそのバリエーションである。上記態様のいずれかまたは本明細書に記載される発明のいずれかの他の態様において、SpCas9は、5' -NGG-3' または 5' -GGG-3' から選択されるPAM配列に対する特異性を有する。

【0024】

上記態様のいずれかまたは本明細書に記載される発明のいずれかの他の態様において、ポリヌクレオチドプログラム可能なDNA結合ドメインは、ヌクレアーゼ不活性バリエーションまたはニックーゼバリエーションである。上記態様のいずれかまたは本明細書に記載される発明のいずれかの他の態様において、ニックーゼバリエーションは、アミノ酸置換D10Aまたはそれに対応するアミノ酸置換を含む。上記態様のいずれかまたは本明細書に記載される発明のいずれかの他の態様において、シチジンデアミナーゼドメインは、デオキシリボ核酸(DNA)におけるシチジンを脱アミノ化することができるものである。上記態様のいずれかまたは本明細書に記載される発明のいずれかの他の態様において、シチジンデアミナーゼは、天然には存在しない、改変されたシチジンデアミナーゼである、上記態様のいずれ

10

20

30

40

50

かまたは本明細書に記載される発明のいずれかの他の態様において、シチジンデアミナーゼは、APOBECデアミナーゼである。上記態様のいずれかまたは本明細書に記載される発明のいずれかの他の態様において、塩基エディターはBE4である。上記態様のいずれかまたは本明細書に記載される発明のいずれかの他の態様において、1つ以上のガイドRNAは、CRISPR RNA (crRNA)およびトランスコード化小RNA (tracrRNA)を含み、crRNAは、A1ADに関連するSNPを含むSERPINA1核酸配列に対して相補的な核酸配列を含む。上記態様のいずれかまたは本明細書に記載される発明のいずれかの他の態様において、塩基エディターは、メチオニン374をコードするSERPINA1核酸配列に対して相補的な核酸配列を含む単一ガイドRNA (sgRNA)と複合体をなす。

【0025】

いくつかの実施形態において、本明細書に提供される方法のいずれかは、追加の核酸塩基の第2の編集をさらに含む。ある場合には、追加の核酸塩基は、遺伝的障害の原因ではないものである。ある場合には、追加の核酸塩基は、遺伝的障害の原因であるものである。

【0026】

いくつかの実施形態において、デアミナーゼドメインは、シチジンデアミナーゼドメインまたはアデノシンデアミナーゼドメインである。いくつかの実施形態において、デアミナーゼドメインはシチジンデアミナーゼドメインである。いくつかの実施形態において、デアミナーゼドメインはアデノシンデアミナーゼドメインである。いくつかの実施形態において、アデノシンデアミナーゼドメインは、デオキシリボ核酸 (DNA) におけるアデニンを脱アミノ化することができるものである。いくつかの実施形態において、ガイドポリヌクレオチドは、リボ核酸 (RNA)、またはデオキシリボ核酸 (DNA) を含む。いくつかの実施形態において、ガイドポリヌクレオチドは、CRISPR RNA (crRNA)配列、トランス活性化CRISPR RNA (tracrRNA)配列、またはそれらの組合せを含む。

【0027】

いくつかの実施形態において、本明細書に提供される方法のいずれかは、第2のガイドポリヌクレオチドをさらに含む。いくつかの実施形態において、第2のガイドポリヌクレオチドは、リボ核酸 (RNA)、またはデオキシリボ核酸 (DNA) を含む。いくつかの実施形態において、第2のガイドポリヌクレオチドは、CRISPR RNA (crRNA)配列、トランス活性化CRISPR RNA (tracrRNA)配列、またはそれらの組合せを含む。いくつかの実施形態において、第2のガイドポリヌクレオチドは、塩基エディターを第2の標的ヌクレオチド配列へとターゲティングする。いくつかの実施形態において、ポリヌクレオチドプログラム可能なDNA結合ドメインは、Cas9ドメイン、Cpf1ドメイン、CasXドメイン、CasYドメイン、Cas12b/C2c1ドメイン、またはCas12c/C2c3ドメインを含む。いくつかの実施形態において、ポリヌクレオチドプログラム可能なDNA結合ドメインは、ヌクレアーゼ不活 (dead) である。いくつかの実施形態において、ポリヌクレオチドプログラム可能なDNA結合ドメインはニッカーゼである。いくつかの実施形態において、ポリヌクレオチドプログラム可能なDNA結合ドメインはCas9ドメインを含む。いくつかの実施形態において、Cas9ドメインは、ヌクレアーゼ不活Cas9 (dCas9)、Cas9ニッカーゼ (nCas9)、またはヌクレアーゼ活性Cas9を含む。いくつかの実施形態において、Cas9ドメインはCas9ニッカーゼを含む。いくつかの実施形態において、ポリヌクレオチドプログラム可能なDNA結合ドメインは、操作または改変されたポリヌクレオチドプログラム可能なDNA結合ドメインである。

【0028】

いくつかの実施形態において、本明細書に提供される方法のいずれかは、第2の塩基エディターをさらに含む。いくつかの実施形態において、第2の塩基エディターは、第1のまたは一次的な塩基エディターとは異なるデアミナーゼドメインを含む。

【0029】

いくつかの実施形態において、塩基編集は、20%未満のインデル (indel) 形成をもたらす。いくつかの実施形態において、編集は、15%未満のインデル形成をもたらす。いくつかの実施形態において、編集は、10%未満のインデル形成をもたらす。いくつかの実施

10

20

30

40

50

形態において、編集は、5%未満のインデル形成をもたらす。いくつかの実施形態において、編集は、4%未満のインデル形成をもたらす。いくつかの実施形態において、編集は、3%未満のインデル形成をもたらす。いくつかの実施形態において、編集は、2%未満のインデル形成をもたらす。いくつかの実施形態において、編集は、1%未満のインデル形成をもたらす。いくつかの実施形態において、編集は、0.5%未満のインデル形成をもたらす。いくつかの実施形態において、編集は、0.1%未満のインデル形成をもたらす。いくつかの実施形態において、編集は、転座をもたらさない。

【0030】

本開示の特徴は、添付の特許請求の範囲に特に記載されている。本開示の特徴および利点のよりよい理解は、以下の詳細な説明を参照することによって得られる。この詳細な説明は、例示的な実施形態を示し、ここで、本開示の原理が利用され、添付の図面は以下の通りである。

10

【図面の簡単な説明】

【0031】

【図1】図1は、健常者とアンチトリプシン欠損症 (A1AD) 患者とを比較した模式図である。健常者では、 α -1アンチトリプシン (A1AT)はプロテアーゼのダメージから肺を保護し、肝臓は α -1アンチトリプシンを血液中に放出する。A1ADを有する患者において、正常に機能するA1ATタンパク質の欠損は肺組織損傷をもたらす。さらに、肝細胞に異常なA1ATが蓄積すると肝硬変になる。

【0032】

【図2】図2は、異なる遺伝子型についての血清 α -1アンチトリプシン (A1AT) レベルの典型的な範囲を示すグラフである (標準 (MM); α -1アンチトリプシン欠損症のヘテロ接合体キャリア (MZ, SZ); ホモ接合体欠損症 (SS, ZZ))。血清 α -1アンチトリプシン (AAT)濃度は、左「y」軸において μ Mで表されており、これが文献において一般的である。右の「y」軸は、臨床検査室および様々な測定技術 (比濁法や放射免疫拡散法)によって一般的に報告されるmg/dL単位への血清AAT濃度のおおよその換算を示す。

20

【0033】

【図3】図3は、SERPINA1にサプレッサー変異M374Iを導入するための標的部位の配列を示す。標準的spCas9 NGG PAM、および、その編集が所望のコドン変化M374Iをもたらす標的Cがハイライトされている。もしそれが編集されると望ましくないコドン変化E376KがもたらされるオフターゲットCも標識されている。

30

【0034】

【図4】図4は、A1ATタンパク質の異なるバリエーションをコードするプラスミドで一過的にトランスフェクトされたHEK293Tの培養上清における分泌タンパク質のレベルを示すグラフである。A1AT濃度は、Borel, Florie & Mueller, Christian. (2017) Alpha-1 Antitrypsin Deficiency: Methods and Protocols. 10.1007/978-1-4939-7163-3に刊行されているようにELISAによって決定した (その内容全体が参照により組み入れられる)。A1ATの最も一般的な2つの臨床的バリエーション (例えば病原性変異) はE264V (PiS アリル) と E342K (PiZ アリル) である。PiS および PiZ タンパク質は、野生型タンパク質よりも少ない量で産生される。図4において「補完的変異」と呼ばれるM374Iサプレッサー変異の追加は、分泌されるPiS および PiZ A1ATタンパク質のレベルを増大させると見られる。従って我々は、本明細書に記載される塩基エディターおよび塩基編集方法を用いたM374I変異の導入が肝細胞からのA1AT分泌を増加させ、肝毒性を緩和すると同時に肺へのA1ATの循環を増加させることができると仮定した。A1AT: アルファ-1 アンチトリプシン; A1AD: アルファ-1 アンチトリプシン欠損症; 「Z変異」はE342K (PiZ アリル) である; 「S変異」はE264V (PiS アリル) 変異である。

40

【0035】

【図5】図5は、HEK293TにおけるM374I変異の塩基編集の効率を示す棒グラフである。bpNLSの使用は、SV40核局在化シグナルよりも優れていた。最初のコドン使用と比べて、コドン最適化2は、プラスミドとしておよびmRNA+gRNAとして送達された場合の両

50

方において、より高い変種効率を生じている。

【0036】

【図6】図6は、TadAから出発してDNAデオキシアデノシンデアミナーゼを進化させるための戦略を示す概略図である。E. coliのライブラリーは、変異型ecTadA (TadA*)遺伝子がdCas9に融合されたもののプラスミドライブラリーと、抗生物質耐性遺伝子を修復するために標的化されたA・TからG・Cへの変異を必要とするセレクションプラスミドを含有している。生存したTadA*パリアントからの変異を、ヒトにおける塩基編集のためのABE構築物へと輸入した。

【0037】

【図7】図7は、予測塩基編集A1ATパリアントの機能的エラスターゼ活性を示すグラフを示す。野生型(WT) A1ATのエラスターゼ活性と比べた、E342K (PiZ)変異を有するA1ATパリアント；E342K変異と補完的M374I変異とを有するA1ATパリアント；E264V (PiS) 変異を有するA1ATパリアント；およびE264V変異と補完的M374I変異とを有するA1ATパリアントのエラスターゼ活性のパーセントをグラフ中に示している。

10

【0038】

【図8】図8A～8Cは、それぞれ塩基エディターBE4でトランスフェクトされたHEK293細胞(図8A)および人工多能性幹細胞(iPSC)(図8B)で観察された塩基編集のパーセンテージを示す3つのグラフを提供する。図8Cは、野生型初代肝細胞がトランスフェクトされた場合に達成された編集のパーセントを示す。

【0039】

【図9】図9は、BE4で編集されたIPSCに由来する肝細胞で達成された塩基編集のパーセントおよびA1AT分泌を示す。

20

【発明を実施するための形態】

【0040】

以下の説明および実施例は、本開示の実施形態を詳細に例示する。本開示は、本明細書に記載される特定の実施形態に限定されず、従って変化し得ることが理解されるべきである。本開示には、その範囲内に包含される多数の変形および修正があることを当業者は認識するであろう。

【0041】

全ての用語は、当業者によって理解されるものとして理解されることが意図されている。別段の定義がない限り、本明細書で使用される全ての技術的および科学的用語は、本開示が関係する分野の当業者によって一般的に理解されるものと同じ意味を有する。

30

【0042】

本明細書に開示されるいくつかの実施形態の実施は、別段の指示がない限り、当業者の技量の範囲内にある、免疫学、生化学、化学、分子生物学、微生物学、細胞生物学、ゲノミクスおよび組換えDNAの従来技術を使用する。例えば、Sambrook and Green, Molecular Cloning: A Laboratory Manual, 4th Edition (2012); the series Current Protocols in Molecular Biology (F. M. Ausubel, et al. eds.); the series Methods In Enzymology (Academic Press, Inc.), PCR 2: A Practical Approach (M.J. MacPherson, B.D. Hames and G.R. Taylor eds. (1995)), Harlow and Lane, eds. (1988) Antibodies, A Laboratory Manual, and Culture of Animal Cells: A Manual of Basic Technique and Specialized Applications, 6th Edition (R.I. Freshney, ed. (2010))参照。

40

【0043】

本明細書中で使用されるセクション見出しは、まとめる目的だけのためであり、記載される主題を限定するものと解釈されるべきではない。

【0044】

本開示の種々の特徴は、単一の実施形態に関連して説明され得るが、これらの特徴は、分離されて、または任意の適切な組み合わせで提供することもできる。逆に、本開示は、明確にするために別々の実施形態の文脈で説明され得るが、本開示はまた、単一の実施形

50

態で実施することもできる。

【0045】

定義

以下の定義は、本技術分野のものを補足するものであり、本出願に向けられるものであり、任意の関連する、または関連しない事例、例えば、任意の同時所有される特許または特許出願に帰属されるべきではない。本明細書に記載のものと類似するか、または等価である任意の方法および材料を、本開示の試験のための実施において使用することができるが、好ましい方法および材料を本明細書で説明する。したがって、本明細書で使用される用語は、特定の実施形態だけを説明するためのものであり、限定を意図するものではない。

【0046】

別段の定義がない限り、本明細書で使用されるすべての技術的および科学的用語は、本発明が属する分野の当業者によって一般的に理解される意味を有する。以下の参考文献は、当業者に、本発明において使用される多くの用語の一般的な定義を提供する: Singleton et al., Dictionary of Microbiology and Molecular Biology (2nd ed. 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988); The Glossary of Genetics, 5th Ed., R. Rieger et al. (eds.), Springer Verlag (1991); and Hale & Marham, The Harper Collins Dictionary of Biology (1991)。

【0047】

本出願において、単数形の使用は、特に断りのない限り、複数形を含む。本明細書で使用されるところの単数形「a」、「an」、および「the」は、文脈が明確にそうでないことを指示しない限り、複数の指示対象を含むことに留意されなければならない。本出願において、「又は」の使用は、別段の記載がない限り、「及び/又は」を意味する。さらに、用語「含む(including)」、ならびに「含む(include)」、「含む(includes)」、および「含まれる(included)」などの他の形態の使用は、非限定的である。

【0048】

本明細書及びクレームにおいて使用される場合、用語「含む(comprising)」（および「含む(comprise)」、および「含む(comprises)」などのそのあらゆる形態)、「有する(having)」（「有する(have)」、および「有する(has)」などのそのあらゆる形態)、「含む(including)」（「含む(include)」、および「含む(includes)」などのそのあらゆる形態)又は「含む(containing)」（「含む(contains)」、および「含む(contain)」などのそのあらゆる形態)は、包括的又は開放的であり、追加の、記載されていない要素又は方法工程を排除しない。本明細書で議論される任意の実施形態は、本開示の任意の方法または組成物に関して実施することができ、逆もまた同様であると考えられる。さらに、本開示の組成物を用いて、本開示の方法を達成することができる。

【0049】

用語「約(about)」または「およそ(approximately)」は、当業者によって決定されるように、特定の値について許容可能な誤差範囲内であることを意味し、これは、値がどのように測定または決定されるか、すなわち、測定システムの制限に部分的に依存する。例えば、「約」は、当該技術分野における実務によれば、1以内または1を超える標準偏差内であることを意味し得る。あるいは、「約」は、所定の値の最大20%、最大10%、最大5%、または最大1%の範囲を意味し得る。別法として、特に生物学的システムまたはプロセスに関して、この用語は、同じ桁以内、好ましくは5倍以内、より好ましくは2倍以内の値を意味することができる。特定の値が出願及び特許請求の範囲に記載されている場合、別段の記載がない限り、その特定の値について許容可能な誤差範囲内にあることを意味する「約」という用語が推定されるべきである。

【0050】

明細書における「いくつかの実施形態」、「ある実施形態」、「一実施形態」、または「他の実施形態」という言及は、その実施形態に関連して説明される特定の特徵、構造、または特性が、本開示の少なくともいくつかの実施形態に含まれるが、必ずしもすべての実施形態に含まれるとは限らないことを意味する。

10

20

30

40

50

【 0 0 5 1 】

「投与すること」は、本明細書に記載される1以上の組成物を患者または対象に提供することとして本明細書で言及される。例として、限定するものではないが、組成物の投与、例えば注射は、静脈内 (i.v.) 注射、皮下(s.c.) 注射、皮内 (i.d.) 注射、腹腔内(i.p.) 注射又は筋肉内(i.m.) 注射によって行われ得る。1つ以上のそのような経路を用いることができる。非経口投与は、例えば、ボーラス注射によって、または経時的に徐々に灌流することによって行うことができる。あるいは、または同時に、投与は経口経路によることができる。

【 0 0 5 2 】

「アデノシンデアミナーゼ」とは、アデニン (A) からイノシン (I) への加水分解的脱アミノ化を触媒するデアミナーゼを意味する。ある態様において、デアミナーゼまたはデアミナーゼドメインは、それぞれアデノシンまたはデオキシアデノシンからイノシンまたはデオキシイノシンへの加水分解的脱アミノ化を触媒するアデノシンデアミナーゼである。ある態様において、アデノシンデアミナーゼは、デオキシリボ核酸 (DNA) 中のアデノシンの加水分解的脱アミノ化を触媒する。本明細書中で提供されるアデノシンデアミナーゼ(例えば、遺伝子操作されたアデノシンデアミナーゼ、進化させたアデノシンデアミナーゼ)は、細菌などの任意の生物由来であり得る。いくつかの実施形態では、アデノシンデアミナーゼは、例えば*E. coli*、*S. aureus*、*S. typhi*、*S. putrefaciens*、*H. influenzae*、または*C. crescentus*のような細菌からのものである。いくつかの実施形態では、アデノシンデアミナーゼは*TadA*デアミナーゼである。いくつかの実施形態では、*TadA*デアミナーゼは、*E. coli TadA (ecTadA)* デアミナーゼまたはその断片である。

【 0 0 5 3 】

例えば、切り詰められた*ecTadA*は、全長*ecTadA*に対して一つ以上のN末端アミノ酸を欠失していてもよい。いくつかの実施形態において、切り詰められた*ecTadA*は、全長*ecTadA*に対して1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19または20個のN末端アミノ酸残基を欠失していてもよい。いくつかの実施形態において、切り詰められた*ecTadA*は、全長*ecTadA*に対して1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19または20個のC末端アミノ酸残基を欠失していてもよい。いくつかの実施形態において、*ecTadA*デアミナーゼは、N末端メチオニンを含まない。いくつかの実施形態において、*TadA*デアミナーゼは、N末端が切り詰められた*TadA*である。特定の实施形態において、*TadA*は、その全体が参照により本明細書に組み込まれるPCT/US2017/045381に記載されている*TadA*のいずれかである。

【 0 0 5 4 】

「薬剤 (agent)」とは、任意の小分子化合物、抗体、核酸分子、またはポリペプチド、またはそれらの断片を意味する。

【 0 0 5 5 】

「改善する (ameliorate)」とは、疾患の発生または進行を減少させる、抑制する、減衰させる、減少させる、停止させる、または安定させることを意味する。

【 0 0 5 6 】

「改変」とは、本明細書に記載されるような標準技術の公知の方法によって検出されるような、遺伝子またはポリペプチドの発現レベルまたは活性の変化 (増加または減少) を意味する。本明細書で使用される場合、改変は、発現レベルの10%の変化、好ましくは25%の変化、より好ましくは40%の変化、最も好ましくは50%以上の発現レベルの変化を含む。

【 0 0 5 7 】

「アナログ」とは、同一ではないが、類似の機能的または構造的な特徴を有する分子を意味する。例えば、ポリペプチドアナログは、対応する天然ポリペプチドの生物学的活性を保持しながら、天然ポリペプチドと比較してそのアナログの機能を増強させる特定の生化学的修飾を有する。そのような生化学的修飾は、例えばリガンド結合を変化させることなく、アナログのプロテアーゼ抵抗性、膜透過性、または半減期を増加させることができる

10

20

30

40

50

。アナログは、非天然アミノ酸を含み得る。

【 0 0 5 8 】

「アルファ - 1 アンチトリプシン (A1AT) タンパク質」は、UniProt受託番号P01009に対して少なくとも約95%のアミノ酸配列同一性を有するポリペプチドまたはその断片を意味する。特定の実施形態では、A1ATタンパク質は、以下の参照配列と比較して1つまたは複数の改変を含む。1つの特定の実施形態では、A1ADと関連するA1ATタンパク質は、E342K変異を含む。例示的A1ATアミノ酸配列が以下に提供される。

sp | P01009 | A1AT_HUMAN Alpha-1-antitrypsin OS=Homo sapiens OX=9606 GN=SERPINA1 PE=1 SV=3:

MPSSVSWGILLLAGLCCLVPVSLAEDPQGDAQAQKTDTSHTDQDHPTFNKITPNLAEFAFS
LYRQLAHQSNSTNIFFSPVSIATAFAMLSLGTKADTHDEILEGLNLFNLTEIPEAQIHEGF
QELLRTLNPDSQLQLTTGNGLFLSEGLKLVDFKLEVDVKKLYHSEAFVNFVFGDTEEAKKQ
INDYVEKGTQGGKIVDLVKELDRDTVFALVNYIFFKGGKWERPFVVDTEEDFHVDQVTTV
KVPMMKRLGMFNIQHCKKLSSWVLLMKYLGNATAIFFLPDEGKLQHLENELTHDIITKFL
ENEDRRSASLHLPKLSITGTLDKSVLGLGITKVFNSGADLSGVTEEAPLKLKSKAVHKA
VLTIDEKGTAAAGAMFLEAIPMSIPPEVKFNKPFVFLMIEQNTKSPLFMGKVVNPTQK

【 0 0 5 9 】

「塩基エディター (BE)」という用語は、核酸配列 (例えばDNAまたはRNA) 内の核酸塩基 (例えばA, T, C, G, またはU) に改変を与えることができるポリペプチドを含む薬剤を表す。いくつかの実施形態では、塩基エディターは、ガイドポリヌクレオチド (例えばガイドRNA) と共にポリヌクレオチドプログラム可能なヌクレオチド結合ドメインと核酸塩基編集ドメイン (例えばシチジンデアミナーゼドメインまたはアデノシンデアミナーゼドメイン) とを含む、融合タンパク質である。いくつかの実施形態では、塩基エディターは、シチジン塩基エディター (CBE) である。いくつかの実施形態では、塩基エディターは、アデノシン塩基エディター (ABE) である。いくつかの実施形態では、ポリヌクレオチドプログラム可能なDNA結合ドメインが、デアミナーゼドメインに融合または連結される。いくつかの実施形態では、塩基エディターは、ガイドポリヌクレオチド (例えばガイドRNA) と共にポリヌクレオチドプログラム可能なDNA結合ドメインとデアミナーゼドメインとを含む。いくつかの実施形態では、ポリヌクレオチドプログラム可能なDNA結合ドメインはCRISPR関連 (例えばCasまたはCpf1) 酵素である。いくつかの実施形態では、塩基エディターは、Cas9タンパク質がデアミナーゼドメイン (例えばアデノシンデアミナーゼまたはシチジンデアミナーゼ) に融合されたものである。いくつかの実施形態では、塩基エディターは、触媒的に不活 (dead) なCas9 (dCas9) がデアミナーゼドメインに融合されたものである。いくつかの実施形態では、塩基エディターは、Cas9ニックアーゼ (nCas9) がデアミナーゼドメインに融合されたものである。いくつかの実施形態では、塩基エディターは、塩基除去修復 (BER) の阻害因子に融合される。いくつかの実施形態では、塩基除去修復の阻害因子は、ウラシルDNAグリコシラーゼ阻害因子 (UGI) である。いくつかの実施形態では、塩基除去修復の阻害因子は、イノシン塩基除去修復阻害因子である。いくつかの実施形態では、塩基エディターは、核酸中の塩基を脱アミノ化することができる。いくつかの実施形態では、塩基エディターは、DNA分子中の塩基を脱アミノ化することができる。いくつかの実施形態では、塩基エディターは、RNA分子中の塩基を脱アミノ化することができる。いくつかの実施形態では、塩基エディターは、アデニン (A) を脱アミノ化することができる。いくつかの実施形態では、アデノシンデアミナーゼは、TadAから進化されたものである。いくつかの実施形態では、塩基エディターは、グアニン (G) を脱アミノ化することができる。いくつかの実施形態では、塩基エディターは、アデニン (A) を脱アミノ化することができる。いくつかの実施形態では、塩基エディターは、シトシン (C) を脱アミノ化することができる。塩基エディターの詳細は国際出願PCT/2017/045381 (WO2018/027078)およびPCT/US2016/058344 (WO2017/070632)に記載されており、その各々は参照によりその全体が本明細書に組み入れられる。

Komor, A.C., et al., "Programmable editing of a target base in genomic DNA w

ithout double-stranded DNA cleavage” Nature 533, 420-424 (2016); Gaudelli, N.M., et al., “Programmable base editing of A · T to G · C in genomic DNA without DNA cleavage” Nature 551, 464-471 (2017); およびKomor, A.C., et al., “Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity” Science Advances 3:eaao4774 (2017)も参照（その内容全体が参照によりその全体が本明細書に組み入れられる）。

【 0 0 6 0 】

いくつかの実施形態では、本明細書に記載される塩基編集の組成物、システム、および方法において使用されるシチジン塩基エディターBE4は、下に提供される下記の核酸配列（8877塩基対）を有する（Addgene, Watertown, MA.; Komor AC, et al., 2017, Sci Adv., 30;3(8):eaao4774. doi: 10.1126/sciadv.aao4774）。BE4核酸配列に対して少なくとも95%以上の同一性を有するポリヌクレオチド配列も包含される。

```

1  atatgccaag tacgccccct attgacgtca atgacggtaa atggcccgcc tggcattatg
61  cccagtacat gaccttatgg gactttccta ctggcagta catctacgta ttagtcatcg
121 ctattacat ggtgatgagg ttttggcagt acatcaatgg gcgtggatag cggtttgact
181 cacggggatt tccaagtctc caccccattg acgtcaatgg gagtttgatt tggcaccaaa
241 atcaacggga ctttcaaaaa tgtcgtaca actccgcccc attgacgcaa atgggaggga
301 ggctgttac gttggaggtc tatataagca gagctggatt agtgaaccgt cagatccgct
361 agagatccgc ggccgctaac acgactcact atagggagag ccgccaccat gagctcagag
421 actggcccag tggctgtgga cccacattg agacggcgga tcgagcccca tgagtttgag
481 gtattcttcg atccgagaga gctccgcaag gagacctgcc tgctttacga aattaattgg
541 gggggccggc actccatttg gcgacataca tcacagaaca ctaacaagca cgtcgaagtc
601 aacttcatcg agaagttcac gacagaaaga tatttctgtc cgaacacaag gtgcagcatt
661 acctggtttc tcagctggag cccatgcggc gaatgtagta gggccatcac tgaattcctg
721 tcaaggtatc cccacgtcac tctgtttatt tacatcgcaa ggctgtacca ccacgtgac
781 ccccgcaatc gacaaggcct gcgggatttg atctcttcag gtgtgactat ccaattatg
841 actgagcagg agtcaggata ctgctggaga aactttgtga attatagccc gagtaatgaa
901 gccactggc ctaggatcc ccatctgtgg gtacgactgt acgttcttga actgtactgc
961 atcatactgg gcctgcctcc ttgtctcaac attctgagaa ggaagcagcc acagctgaca
1021 ttctttacca tcgctcttca gtcttgcac taccagcgac tgccccaca cattctctgg
1081 gccaccgggt tgaatctgg tggttcttct ggtggttcta gcggcagcga gactcccggg
1141 acctcagagt ccgccacacc cgaaagtctt ggtggttctt ctggtggttc tgataaaaag
1201 tattctattg gtttagccat cggcactaat tccgttggat gggctgtcat aaccgatgaa
1261 tacaagtac cttcaagaa atttaagggtg ttggggaaca cagaccgtca ttcgattaaa
1321 agaatctta tcggtgccct cctattcgat agtggcgaaa cggcagaggc gactcgcctg
1381 aaacgaaccg ctgggagaag gtatacacgt cgcaagaacc gaatatgta cttacaagaa
1441 attttagca atgagatggc caaagttgac gattcttctt ttcaccgttt ggaagagtcc
1501 ttccttgtcg aagaggacaa gaaacatgaa cggcacccca tcttggaaa catagtagat
1561 gaggtggcat atcatgaaaa gtaccaacg atttatcacc tcagaaaaaa gctagttgac
1621 tcaactgata aagcggacct gaggttaac tacttggctc ttgccatat gataaagttc
1681 cgtgggacct ttctcattga ggggtatcta aatccggaca actcggatgt cgacaaactg
1741 ttcattcagt tagtacaac ctataatcag ttgtttgaag agaaccctat aatgcaagt
1801 ggcgtggatg cgaaggctat tcttagcgcc cgcctctcta aatcccgcg gctagaaaac
1861 ctgatcgcac aattaccggg agagaagaaa aatgggttgt tcggtaacct tatagcgctc
1921 tcaactaggc tgacacaaaa ttttaagtcg aacttcgact tagctgaaga tgccaaattg
1981 cagcttagta aggacacgta cgatgacgat ctgcacaatc tactggcaca aattggagat
2041 cagtatgcgg acttattttt ggctgcaaaa aacctagcg atgcaatcct cctatctgac
2101 atactgagag ttaatactga gattaccaag gcgcccgtat ccgcttcaat gatcaaaagg
2161 tacgatgaac atcaccaaga ctgacactt ctcaaggccc tagtccgtca gcaactgcct

```

2221 gagaaatata aggaaatatt ctttgatcag tcgaaaaacg ggtacgcagg ttatattgac
 2281 ggcggagcga gtcaagagga attctacaag tttatcaaac ccatattaga gaagatggat
 2341 gggacggaag agttgcttgt aaaactcaat cgcgaagatc tactgcgaaa gcagcggact
 2401 ttcgacaacg gtagcattcc acatcaaadc cacttaggag aattgcatgc tatacttaga
 2461 aggcaggagg atttttatcc gttcctcaaa gacaatcgtg aaaagattga gaaaatccta
 2521 acctttcgca taccttacta tgtgggaccc ctggcccag ggaactctcg gttcgcattg
 2581 atgacaagaa agtccgaaga aacgattact ccatggaatt ttgaggaagt tgtcgataaa
 2641 ggtgcgtcag ctcaatcgtt catcgagagg atgaccaact ttgacaagaa tttaccgaac
 2701 gaaaaagtat tgcctaagca cagtttactt tacgagtatt tcacagtgtc caatgaactc
 2761 acgaaagtta agtatgtcac tgagggcatg cgtaaaccg cttttctaag cggagaacag 10
 2821 aagaaagcaa tagtagatct gttattcaag accaaccgca aagtgcagct taagcaattg
 2881 aaagaggact actttaagaa aattgaatgc ttcgattctg tcgagatctc cgggtagaa
 2941 gatcgattta atgcgtcact tggtagctat catgacctcc taaagataat taaagataag
 3001 gacttcctgg ataacgaaga gaatgaagat atcttagaag atatagtgtt gactcttacc
 3061 ctctttgaag atcgggaaat gattgaggaa agactaaaaa catacgtca cctgttcgac
 3121 gataaggtta tgaacagtt aaagaggcgt cgctatacgg gctggggacg attgtcgcgg
 3181 aaacttatca acgggataag agacaagcaa agtggtaaaa ctattctga ttttctaag
 3241 agcgacggct tcgccaatag gaactttatg cagctgatcc atgatgactc ttaaccttc
 3301 aaagaggata taaaaaggc acaggtttcc ggacaagggg actcattgca cgaacatatt
 3361 gcgaatcttg ctggttcgcc agccatcaaa aagggcatac tccagacagt caaagtagtg 20
 3421 gatgagctag ttaaggtcat gggacgtcac aaaccgaaa acattgtaat cgagatggca
 3481 cgcgaaaatc aaacgactca gaaggggcaa aaaaacagtc gagagcggat gaagagaata
 3541 gaagagggta ttaaagaact gggcagccag atcttaaagg agcatcctgt gaaaaatacc
 3601 caattgcaga acgagaaact ttacctctat tacctacaaa atggaaggga catgtatgtt
 3661 gatcaggaac tggacataaa ccgtttatct gattacgacg tcgatcacat tgtaccccaa
 3721 tccttttga aggacgattc aatcgacaat aaagtgccta cacgctcggg taagaaccga
 3781 gggaaaagtg acaatgttcc aagcaggaa gtcgtaaaga aatgaagaa ctattggcgg
 3841 cagctcctaa atgcgaaact gataacgcaa agaaagtctg ataactaac taaagctgag
 3901 aggggtggct tgtctgaact tgacaaggcc ggatttatta aacgtcagct cgtggaacc
 3961 cgccaaatca caaagcatgt tgacagata ctgattccc gaatgaatac gaaatagcag 30
 4021 gagaacgata agctgattcg ggaagtcaaa gtaatcactt taaagtcaaa attggtgtcg
 4081 gacttcagaa aggattttca attctataaa gttagggaga taaataacta ccaccatgag
 4141 cacgacgctt atcttaatgc cgtcgtaggg accgactca ttaagaaata cccgaagcta
 4201 gaaagtgagt ttgtgtatgg tgattacaaa gtttatgacg tccgtaagat gatcgcgaaa
 4261 agcgaacagg agataggcaa ggctacagcc aaatacttct tttattctaa cattatgaat
 4321 ttctttaaga cggaaatcac tctggcaaac ggagagatac gcaaacgacc ttaattgaa
 4381 accaatgggg agacagggtga aatcgtatgg gataagggcc gggacttcgc gacggtgaga
 4441 aaagttttgt ccatgcccc agtcaacata gtaaagaaaa ctgaggtgca gaccggaggg
 4501 ttttcaaagg aatcgattct tccaaaagg aatagtata agctcatcgc tcgtaaaaag
 4561 gactgggacc cgaaaaagta cggtggttc gatagcccta cagttgccta ttctgtccta 40
 4621 gtagtggcaa aagttgagaa gggaaaatcc aagaaactga agtcagtcaa agaattattg
 4681 gggataacga ttatggagcg ctgcttttt gaaaagaacc ccatcgactt cttgaggcg
 4741 aaaggttaca aggaagtaaa aaaggatctc ataattaaac taccaaagta tagtctgttt
 4801 gagttagaaa atggccgaaa acggatgttg gctagcgcg gagagcttca aaaggggaaac
 4861 gaactcgac taccttctaa atacgtgaat ttctgtatt tagcgtccca ttacgagaag
 4921 ttgaaagggt cacctgaaga taacgaacag aagcaacttt ttgttgagca gcacaaacat
 4981 tatctcgacg aatcataga gcaaatttcg gaattcagta agagagtcac ctagctgat
 5041 gccaatctgg acaaagtatt aagcgcatac aacaagcaca gggataaacc catacgtgag
 5101 caggcggaaa atattatcca tttgtttact cttaccaacc tcggcgtcc agccgactc
 5161 aagtattttg acacaacgat agatcgcaaa cgatacactt ctaccaagga ggtgctagac 50

5221 gcgacactga ttcaccaatc catcacggga ttatatgaaa ctgggataga ttgtcacag
 5281 cttgggggtg actctgggtg ttctggagga tctgggtggtt ctactaatct gtcagatatt
 5341 attgaaaagg agaccggtaa gcaactgggt atccaggaat ccctctcat gctcccagag
 5401 gaggtggaag aagtcattgg gaacaagccg gaaagcgata tactcgtgca caccgcctac
 5461 gacgagagca ccgacgagaa tgcatgctt ctgactagcg acgcccctga atacaagcct
 5521 tgggctctgg tcatacagga tagcaacggt gagaacaaga ttaagatgct ctctgggtgt
 5581 tctggaggat ctgggtggtc tactaatctg tcagatatta ttgaaaagga gaccggtaag
 5641 caactgggta tccaggaatc catctctatg ctcccagagg aggtggaaga agtcattggg
 5701 aacaagccgg aaagcgatat actcgtgcac accgcctacg acgagagcac cgacgagaat
 5761 gtcatgcttc tgactagcga cggcccctgaa tacaagcctt gggctctggt catacaggat 10
 5821 agcaacggtg agaacaagat taagatgctc tctgggtggtt ctccaagaa gaagaggaaa
 5881 gtctaaccgg tcatcatcac catcaccatt gagtttaaac ccgctgatca gcctcgactg
 5941 tgccttctag ttgccagcca tctgtgttt gccctcccc cgtgccttcc ttgaccctgg
 6001 aaggtgccac tcccactgtc ctttctaata aaaatgagga aattgcatcg cattgtctga
 6061 gtaggtgca ttctattctg ggggggtggg tggggcagga cagcaagggg gaggattggg
 6121 aagacaatag caggcatgct ggggatgctg tgggctctat ggcttctgag gcggaagaa
 6181 ccagctgggg ctgataccg tcgacctta gctagagctt ggcgtaatca tggatcatagc
 6241 tgtttcctgt gtgaaattgt tatccgctca caattccaca caacatacga gccggaagca
 6301 taaagttaa agcctagggt gcctaagag tgagctaact cacattaatt gcgttgctgct
 6361 cactgcccgc ttccagctc ggaacactgt cgtgccagct gcattaatga atcggccaac 20
 6421 gcgcggggag aggcggttg cgtattgggc gctcttccgc ttctcgcctc actgactcgc
 6481 tgcgctcggc cgttcggctg cggcgagcgg taccagctca ctcaaaggcg gtaatacggc
 6541 tatccacaga atcaggggat aacgcaggaaga agaacatgtg agcaaaaggc cagcaaaagg
 6601 ccaggaaccg taaaaggcc gcgttgctgg cgtttttcca taggctccgc ccccctgacg
 6661 agcatcacia aatcgacgc tcaagtcaga ggtggcgaaa cccgacagga ctataaagat
 6721 accaggcgtt tccccctgga agctccctcg tgcgctctcc tgttccgacc ctgccgctta
 6781 ccgataacct gtccgccttt ctcccctcgg gaagcgtggc gctttctcat agctcacgct
 6841 gtaggtatct cagttcgggt taggtcgttc gctccaagct gggctgtgtg cacgaacccc
 6901 ccgttcagcc cgaccgctgc gccttatccg gtaactatcg tcttgatcc aaccggtaa
 6961 gacacgactt atcgccactg gcagcagcca ctggtaacag gattagcaga gcgaggtatg 30
 7021 taggcggtgc tacagagttc ttgaagtgg ggcctaacta cggctacact agaagaacag
 7081 tatttggtat ctgcgctctg ctgaagccag ttacctcgg aaaaagagtt ggtagctctt
 7141 gatccggcaa acaaaccacc gctggtagcg gtgggttttt tgtttgcaag cagcagatta
 7201 cgcgagaaa aaaaggatct caagaagatc ctttgatctt ttctacgggg tctgacgctc
 7261 agtggaaacga aaactcacgt taagggattt tggatcatgag attatcaaaa aggatcttca
 7321 cctagatcct tttaaattaa aatgaagtt ttaatcaat ctaaagtata tatgagtaaa
 7381 cttgggtctga cagttacca tgcttaatca gtgaggcacc tatctcagcg atctgtctat
 7441 ttcgttcac catagttgcc tgactccccg tctgttagat aactacgata cgggagggt
 7501 taccatctgg cccagtgct gcaatgatac cgcgagacc acgctaccg gctccagatt
 7561 tatcagcaat aaaccagcca gccggaagg cgcgagcag aagtggctc gcaactttat 40
 7621 ccgcctccat ccagtctatt aattgttgc ggaagctag agtaagtagt tcgccagtta
 7681 atagtttgc caacgttgt gccattgcta caggcatcgt ggtgtcacgc tctcgtttg
 7741 gtatggcttc attcagctcc ggttccaac gatcaaggcg agttacatga tccccatgt
 7801 tgtgcaaaaa agcggttagc tcttccgctc ctccgatcgt tgcagaagt aagtggccg
 7861 cagtgtatc actcatgggt atggcagcac tgataattc tcttactgtc atgccatccg
 7921 taagatgctt ttctgtgact ggtgagtact caaccaagtc attctgagaa tagtgtatgc
 7981 ggcgaccgag ttgctcttgc ccggcgtcaa tacgggataa taccgcgcca catagcagaa
 8041 ctttaaaagt gctcatcatt gaaaacgct cttcggggcg aaaactctca aggatcttac
 8101 cgctgtgag atccagttcg atgtaacca ctcgtgcacc caactgatct tcagcatctt
 8161 ttactttcac cagcgtttct gggtagcaaa aacaggaag gcaaaatgcc gcaaaaaagg 50

8221 gaataagggc gacacggaaa tgttgaatac tcatactctt ccttttcaa tattattgaa
 8281 gcatttatca gggttattgt ctcatgagcg gatacatatt tgaatgtatt tagaaaaata
 8341 aacaaatagg ggttccgcbc acatttcccc gaaaagtgcc acctgacgtc gacggatcgg
 8401 gagatcgatc tcccgatccc ctagggtcga ctctcagtac aatctgctct gatgccgcat
 8461 agttaagcca gtatctgctc cctgcttggt tgttgagggt cgctgagtag tgcgagca
 8521 aaatttaagc tacaacaagg caaggcttga ccgacaattg catgaagaat ctgcttaggg
 8581 ttaggcgttt tgcgctgctt cgcgatgtac gggccagata tacgcttga cattgattat
 8641 tgactagtta ttaatagtaa tcaattacgg ggtcattagt tcatagccca tatatggagt
 8701 tccgcttac ataacttacg gtaaatggcc cgcctggctg accgcccaac gacccccgcc
 8761 cattgacgtc aataatgacg tatgttcca tagtaacgcc aatagggact ttccattgac
 8821 gtcaatgggt ggagtattta cggtaaactg cccacttggc agtacatcaa gtgtatc

10

【 0 0 6 1 】

いくつかの実施形態では、シチジン塩基エディターは以下の配列を有する。

ATGagctcagagactggcccagtggtgtggaccccacattgagacggcggatcgagccccatgagtttgag
 gtattcttcgatccgagagagctccgcaaggagacctgcctgctttacgaaattaattgggggggcccggcact
 ccatttggcgacatacatcacagaacactaacaagcacgtcgaagtcaacttcatcgagaagttcacgacaga
 aagatatttctgtccgaacacaaggtgcagcattacctggtttctcagctggagccccatgcccgaatgtagta
 gggccatcactgaattctctgcaaggtatccccacgtcactctgtttatttacatcgcaaggctgtaccaccacg
 ctgacccccgcaatcgacaaggcctgcgggattgatctcttcaggtgtgactatccaaattatgactgagcag
 gactcaggatactgctggagaaactttgtgaattatagcccagtaataagcccactggcctaggtatcccc
 atctgtgggtacgactgtacgttcttgaactgtactgcatcactggtcctccttgtctcaacattctgag
 aaggaagcagccacagctgacattctttaccatcgctcttcagctcttgcattaccagcgactgccccacacat
 tctctggggccaccgggttgaatctggtggttctctggtggttctagcggcagcgagactccccgggacctcag
 agtccgccacacccgaaagtctggtggttctctggtggttctgataaaaagtattctattggtttagccatcg
 gcactaattccgttggatgggctgtcataaccgatgaatacaaagtaccttcaaagaaatthaagggtgtgggg
 aacacagaccgtcattcgattaaaaagaatcttatcgggtgccctcctattcgatagtgcgaaacggcagagg
 cgactcgctgaaacgaaccgctcggagaaggtatacacgtcgaagaaccgaatatgttacttacaagaat
 ttttagcaatgagatggccaaagttagcattcttcttccaccgtttggaagagtccttccctgtcgaagagga
 caagaacatgaacggcacccccatctttgaaacatagtagatgaggtggcatatcatgaaaagtaccaacg
 atttatcacctcagaaaaagctagttgactcaactgataaagcggacctgaggttaattctacttggctcttggc
 catatgataaagttccgtgggcactttctcattgaggggtgatctaaatccggacaactcggatgtcgacaaact
 gttcatccagttagtaaaaacctataatcagttgtttgaagagaacctataaatgcaagtggcgtggatgca
 aggctattcttagcggcccctctctaaatccccgacggctagaaaacctgatcgacacaaattaccggagagaa
 gaaaaatgggttgttcggtaaccttatagcgtctcactaggcctgacaccaaattttaagtcgaacttcgactt
 agctgaagatgccaaattgcagcttagtaaggacacgtacgatgacgatctcgacaatctactggcacaat
 ggagatcagatgcggacttattttggctgccaaaaaccttagcgtgcaatcctcctatctgacatactgaga
 gtttaactgagattaccaaggcgccgttatccgcttcaatgatcaaaaggtagcatgaacatccaagactt
 gacacttctcaaggccctagtcgctcagcaactgcctgagaaatataaggaaatattctttgatcagtcgaaaa
 acgggtacgcaggttatattgacggcggagcaggtcaagaggaattctacaagtttatcaaacccatattaga
 gaagatggatgggacggaagagttgcttgtaaaactcaatcggaagatctactgcaagcagcggactttc
 gacaacggtagcattccacatcaaatccacttaggcgaattgcatgctatacttagaaggcaggaggattttta
 tccgttctcaaaagacaatcgtgaaaagattgagaaaatcctaacctttcgcataccttactatgtgggacctt
 ggcccagggaaactctcggttcgcgatggatgacaagaaagtccgaagaaacgattactccatggaattttgag
 gaagttgtcgataaaggtgcgtcagctcaatcgttcatcgagaggatgaccaactttgacaagaattaccgaa
 cgaaaaagtattgcctaagcacagtttactttacgagtatctcacagtgtaaatgaactcacgaaagtttaagta
 tgtcactgagggcatgcgtaaacccgccttctaaagcggagaaacagaagaaagcaatagtagatctgttattc
 aagaccaaccgcaaagtgcagtttaagcaattgaaagaggactactttaagaaaattgaaatgcttcgattctgt
 cgagatctccggggtagaagatcgatttaatgcgtcacttggtagctatcatgacctcctaaagataattaaag
 ataaggacttctggataacgaagagaatgaagatatcttagaagatatagtttgactcttaccctcttgaag
 atcgggaaatgattgaggaaagactaaaaacatacgcctcacctgttcgacgataagggtatgaaacagttaaa

20

30

40

50

gaggcgtcgctatacgggctggggacgattgtcgcggaaacttatcaacgggataagagacaagcaaagtgg
 taaaactattctcgattttctaaagagcgacggcttcgccaataggaactttatgcagctgatccatgatgactc
 tttaaccttcaaagaggatatacaaaaaggcacaggtttccggacaaggggactcattgcacgaacatattgcg
 aatcttfgctggttcggcagccatcaaaaagggcatactccagacagtcaaagtagtgatgagctagttaagg
 tcatgggacgtcacaaccggaaaacattgtaatcgagatggcacgcgaaaatcaaacgactcagaagggggc
 aaaaaacagtcgagagcggatgaagagaatagaagaggggtattaaagaactgggcagccagatcttaag
 gagcatcctgtggaaaatacccaattgcagaacgagaaactttaccttattacctacaaaatggaagggaca
 tgtatgttgatcaggaactggacataaaccgtttatctgattacgacgtcgatcacattgtaccccaatcctttt
 gaaggacgattcaatcgacaataaagtgttacacgctcggataagaaccgagggaaaagtgacaatgttcc
 aagcggaggaagtgcgtaaagaaaatgaagaactattggcggcagctcctaaatgcgaaactgataacgcaaag
 aaagttcgataacttaactaaagctgagaggggtggcttgtctgaacttgacaaggccggatttattaacgctc
 agctcgtggaaccggccaaatcacaagcatgttgacagatactagattcccgaatgaatacgaatacga
 cgagaacgataagctgattcgggaagtcaaagtaatcactttaaagtcaaaattgggtgtcggacttcagaaag
 gattttcaattctataaagttagggagataaataactaccacatgcgacgacgcttatcttaatgccgtcgt
 gggaccgcactcattaagaaatacccgaaagctagaaagtgagtttgtgtatgggtgattacaaagttatgacgt
 ccgtaagatgatcgcgaaaagcgaacaggagataggcaaggctacagccaaatacttcttttattctaacatta
 tgaatttcttaagacggaaatcactctggcaaacggagagatacgaacgacctttaattgaaaccaatggg
 gagacaggtgaaatcgtatgggataagggccgggacttcgacgaggtgagaaaagtttgtccatgccccaa
 gtcaacatagtaaagaaaactgaggtgcagaccggagggttttcaaaggaatcgattcttccaaaaaggaata
 gtgataagctcatcgtcgttaaaaaggactgggaccggaaaagtagcgggtggcttcgatagccctacagttgc
 ctattctgtcctagtagtgcaaaagttgagaagggaaaatccaagaaactgaagtcagtcгаагааттатг
 gggataacgattatggagcgtcgtcttttgaagaacccccatcgacttcttgaggcgaaaggttacaagg
 aagtaaaaaaggatctcataattaactaccaaagtatagctgtttgagttagaaaatggccgaaaacggatg
 ttggctagcggcggagagcttcaaaaggggaacgaactcgcactaccgtctaaatacgtgaatttctgtattt
 agcgtcccattacgagaagttgaaaggttcacctgaagataacgaacagaagcaactttttgttgagcagcac
 aaacattatctcgacgaaatcatagagcaaatttcggaattcagtaagagagtcacctagctgatgccaatct
 ggacaaagtattaagcgcatacaacaagcacagggataaacccatacgtgagcagggcggaaaatattatcca
 tttgtttactcttaccacactcggcgtccagccgacattcaagatatttgacacaacgatagatcgaaaacgata
 cacttctaccaaggaggtgctagacgcgacactgattcaccaatccatcacgggattatagaaactcggata
 gatttgtcacagcttgggggtgactctggtggttctggaggatctggtggttctactaatctgtcagatattattg
 aaaaggagaccggtaagcaactggttatccaggaatccatcctcatgctcccagaggaggtggaagaagtca
 ttgggaacaagccggaaaagcgatatactcgtgcacaccgcctacgacgagagcaccgacgagaatgtcatgc
 ttctgactagcgcaccccctgaatacaagccttgggctctggtcatacaggatagcaacgggtgagaacaagat
 taagatgctctctggtggttctggaggatctggtggttctactaatctgtcagatattattgaaaaggagaccgg
 taagcaactggttatccaggaatccatcctcatgctcccagaggaggtggaagaagtcattgggaacaagccg
 gaaagcgatatactcgtgcacaccgcctacgacgagagcaccgacgagaatgtcatgcttctgactagcgac
 gcccctgaatacaagccttgggctctggtcatacaggatagcaacgggtgagaacaagattaagatgctctctg
 gtggttctAAAAGGACGGCGGACGGATCAGAGTTCGAGAGTCCGAAAAAAAAACGAAAGG
 TCGAAtaa

10

20

30

【 0 0 6 2 】

40

いくつかの実施形態では、シチジン塩基エディターは以下の配列を有する。

ATGTCATCCGAAACCGGGCCAGTGGCCGTAGACCCAACTCAGGAGGCGGATAGAACC
 CCATGAGTTTGAAGTGTTCCTTCGACCCAGAGAGCTGCGCAAAGAGACTTGCCTCCTGT
 ATGAAATAAATTGGGGGGTCCGCAATTCAATTTGGAGGCACACTAGCCAGAATACTAAC
 AAACACGTGGAGGTAAATTTTATCGAGAAGTTTACCACCGAAAGATACTTTTGCCCCAA
 TACACGGTGTTC AATTACCTGGTTTCTGT CATGGAGTCCATGTGGAGAATGTAGTAGAG
 CGATAACTGAGTTCCTGTCTCGATATCCTCACGTCACGTTGTTTATATACATCGCTCGG
 CTTTATCACCATGCGGACCCGCGGAACAGGCAAGGTCTTCGGGACCTCATATCCTCTGG
 GGTGACCATCCAGATAATGACGGAGCAAGAGAGCGGATACTGCTGGCGAAACTTTGTTA
 ACTACAGCCCAAGCAATGAGGCACACTGGCCTAGATATCCGCATCTCTGGGTTCTGACTG

50

TATGTCCTTGAACGTACTGCATAATTCTGGGACTTCCGCCATGCTTGAACATTCTGCG
 GCGGAAACAACCACAGCTGACCTTTTTTACGATTGCTCTCAAAGTTGCTACTACCAGC
 GATTGCCACCCACATCTTGTGGGCTACTGGACTCAAGTCTGGAGGAAGTTCAGGCGGA
 AGCAGCGGGTCTGAAACGCCCGGAACCTCAGAGAGCGCAACGCCCGAAAGCTCTGGAGG
 GTCAAGTGGTGGTAGTGATAAGAAATACTCCATCGGCCTCGCCATCGGTACGAATTCTG
 TCGGTTGGGCCGTTATCACCGATGAGTACAAGGTCCCTTCTAAGAAATTCAAGTTTTG
 GGCAATACAGACCGCCATTCTATAAAAAAAAAACCTGATCGGCGCCCTTTTGTGACAG
 TGGTGAGACTGCTGAAGCGACTCGCCTGAAGCGAACTGCCAGGAGGCGGTATACGAGGC
 GAAAAACCGAATTTGTTACCTCCAGGAGATTTTCTCAAATGAAATGGCCAAGGTAGAT
 GATAGTTTTTTTACCGCTTGGAAAGAAAGTTTTCTCGTTGAGGAGGACAAAAAGCACGA
 GAGGCACCCAATCTTTGGCAACATAGTCGATGAGGTGCGATACCATGAGAAATATCCTA
 CGATCTATCATCTCCGCAAGAAGCTGGTCGATAGCACGGATAAAGCTGACCTCCGGCTG
 ATCTACCTTGCTCTTGCTCACATGATTAAATTCAGGGGCCATTTCTGATAGAAGGAGA
 CCTCAATCCCGACAATTCTGATGTCGACAACTGTTTATTAGCTCGTTCAGACCTATA
 ATCAACTCTTTGAGGAGAACCCCATCAATGCTTCAGGGGTGGACGCAAAGGCCATTTTG
 TCCGCGCGCTTGAGTAAATCACGACGCCTCGAGAATTTGATAGCTCAACTGCCGGGTGA
 GAAGAAAAACGGGTTGTTTGGGAATCTCATAGCGTTGAGTTTGGGACTTACGCCAACT
 TTAAGTCTAACTTTGATTTGGCCGAAGATGCCAAATTGCAGCTGTCCAAAGATACCTAT
 GATGACGACTTGGATAACCTTCTTGCGCAGATTGGTGACCAATACGCGGATCTGTTTCT
 TGCCGCAAAAAATCTGTCCGACGCCATACTCTTGTCGGATATACTGCGCGTCAATACTG
 AGATAACTAAGGCTCCCTCAGCGCGTCCATGATTAAGATAACGATGAGCACCACCAA
 GATCTCACTCTGTTGAAAGCCCTGGTTCGCCAGCAGCTTCCAGAGAAGTATAAGGAGAT
 ATTTTTCGACCAATCTAAAAACGGCTATGCGGGTTACATTGACGGTGGCGCCTCTCAAG
 AAGAATTCTACAAGTTTATAAAGCCGATACTTGAGAAAATGGACGGTACAGAGGAATTG
 TTGGTTAAGCTCAATCGCGAGGACTTGTGAGAAAGCAGCGCACATTTGACAATGGTAG
 TATCCACACCAGATTCATCTGGGCGAGTTGCATGCCATTCTTAGAAGACAAGAAGATT
 TTTATCCGTTTCTGAAAGATAACAGAGAAAAGATTGAAAAGATACTTACCTTTTCGCATA
 CCGTATTATGTAGGTCCCCTGGCTAGAGGGAACAGTCGCTTCGCTTGGATGACTCGAAA
 ATCAGAAGAAACAATAACCCCTGGAATTTTGAAGAAGTGGTAGATAAAGGTGCGAGTG
 CCCAATCTTTTATTGAGCGGATGACAAATTTTGAACAAGAATCTGCCTAACGAAAAGGTG
 CTTCCCAAGCATTCCCTTTTGTATGAATACTTTACAGTATATAATGAACTGACTAAAGT
 GAAGTACGTTACCGAGGGGATGCGAAAGCCAGCTTTTCTCAGTGGCGAGCAGAAAAAAG
 CAATAGTTGACCTGCTGTTCAAGACGAATAGGAAGGTTACCGTCAAACAGCTCAAAGAA
 GATTACTTTAAAAGATCGAATGTTTTGATTTCAGTTGAGATAAGCGGAGTAGAGGATAG
 ATTTAACGCAAGTCTTGGAACTTATCATGACCTTTTGAAGATCATCAAGGATAAAGATT
 TTTTGGACAACGAGGAGAATGAAGATATCCTGGAAGATATAGTACTTACCTTGACGCTT
 TTTGAAGATCGAGAGATGATCGAGGAGCGACTTAAGACGTACGCACATCTCTTTGACGA
 TAAGGTTATGAAACAATTGAAACGCCGGCGGTATACTGGCTGGGGCAGGCTTTCTCGAA
 AGCTGATTAATGGTATCCGCGATAAGCAGTCTGGAAGACAATCCTTGACTTTCTGAAA
 AGTGATGGATTTGCAAATAGAACTTTATGCAGCTTATACATGATGACTCTTTGACGTT
 CAAGGAAGACATCCAGAAGGCACAGGTATCCGGCCAAGGGGATAGCCTCCATGAACACA
 TAGCCAACCTGGCCGGCTCACCAGCTATTA AAAAGGGAATATTGCAAACCGTTAAGGTT
 GTTGACGAACTCGTTAAGGTTATGGGCCGACACAAACCAGAGAATATCGTGATTGAGAT
 GGCTAGGGAGAATCAGACCACTCAAAAAGGTCAGAAAAATTCTCGCGAAAGGATGAAGC
 GAATTGAAGAGGGAATCAAAGA ACTTGGCTCTCAAATTTTGAAGAGCACC CGGTAGAA
 AACACTCAGCTGCAGAATGAAAAGCTGTATCTGTATTATCTGCAGAATGGTCGAGATAT
 GTACGTTGATCAGGAGCTGGATATCAATAGGCTCAGTGACTACGATGTGACCCACATCG
 TTCTCAATCTTTCTGAAAGATGACTCTATCGACAACAAAGTGTTGACGCGATCAGAT
 AAGAACCGGGGAAAATCCGACAATGTACCCTCAGAAGAAGTTGTCAAGAAGATGAAAAA
 CTATTGGAGACAATTGCTGAACGCCAAGCTCATAACACAACGCAAGTTCGATAACTTGA

10

20

30

40

50

CGAAAGCCGAAAGAGGTTGGGTTGTCAGAATTGGACAAAGCTGGCTTTATTAAGCGCCAA
 TTGGTGGAGACCCGGCAGATTACGAAACACGTAGCACAAATTTTGGATTACGAATGAA
 TACCAAATACGACGAAAACGACAAATTGATACGCGAGGTGAAAGTGATTACGCTTAAGA
 GTAAGTTGGTTTCCGATTTTCAGGAAGGATTTTTCAGTTTTACAAAGTAAGAGAAATAAAC
 AACTACCACCACGCCCATGATGCTTACCTCAACGCGGTAGTTGGCACAGCTCTTATCAA
 AAAATATCCAAAGCTGGAAAGCGAGTTCGTTTACGGTGACTATAAAGTATACGACGTTT
 GGAAGATGATAGCCAAATCAGAGCAGGAAATTGGGAAGGCAACCGCAAATACTTCTTC
 TATTCAAACATCATGAACTTCTTTAAGACGGAGATTACGCTCGCGAACGGCGAAATACG
 CAAGAGGCCCTCATAGAGACTAACGGCGAAACCGGGGAGATCGTATGGGACAAAGGAC
 GGGACTTTGCGACCGTTAGAAAAGTACTTTCAATGCCACAAGTGAATATTGTTAAAAAG
 ACAGAAGTACAAACAGGGGGGTTTCAGTAAGGAATCCATTTTGGCCAAGCGGAACAGTGA
 TAAATTGATAGCAAGGAAAAAAGATTGGGACCCTAAGAAGTACGGTGGTTTTCGACTCTC
 CTACCGTTGCATATTCAGTCTTGTAGTTGCGAAAGTGGAAAAGGGGAAAAGTAAGAAG
 CTTAAGAGTGTTAAAGAGCTTCTGGGCATAACCATAATGGAACGGTCTAGCTTCGAGAA
 AAATCCAATTGACTTTTCTCGAGGCTAAAGGTTACAAGGAGGTAAAAAAGGACCTGATAA
 TTAAGTCCCAAAGTACAGTCTCTTCGAGTTGGAGAATGGGAGGAAGAGAATGTTGGCA
 TCTGCAGGGGAGCTCCAAAAGGGGAACGAGCTGGCTCTGCCTTCAAATAACGTGAACTT
 TCTGTACCTGGCCAGCCACTACGAGAAACTCAAGGGTTCTCCTGAGGATAACGAGCAGA
 AACAGCTGTTTGTAGAGCAGCACAAAGCATTACCTGGACGAGATAATTGAGCAAATTAGT
 GAGTTCTCAAAAAGAGTAATCCTTGCAGACGCGAATCTGGATAAAGTTCTTTCCGCCTA
 TAATAAGCACCCGGGACAAGCCTATACGAGAACAAGCCGAGAACATCATTACCTCTTTA
 CCCTTACTAATCTGGGCGCGCCGGCCGCCTTCAAATACTTCGACACCACGATAGACAGG
 AAAAGGTATACGAGTACCAAAGAAGTACTTGACGCCACTCTCATCCACCAGTCTATAAC
 AGGGTTGTACGAAACGAGGATAGATTTGTCCCAGCTCGGCGGGGACTCAGGAGGGTTCAG
 GCGGCTCCGGTGGATCAACGAATCTTTCCGACATAATCGAGAAAGAAACCGGCAAACAG
 TTGGTGTATCCAAGAATCAATCCTGATGCTGCCTGAAGAAGTAGAAGAGGTGATTGGCAA
 CAAACCTGAGTCTGACATTCTTGTCCACACCGCGTATGACGAGAGCACGGACGAGAACG
 TTATGCTTCTCACTAGCGACGCCCTGAGTATAAACCATGGGCGCTGGTCATCCAAGAT
 TCCAATGGGGAAAACAAGATTAAGATGCTTAGTGGTGGGTCTGGAGGGAGCGGTGGGTCT
 CACGAACCTCAGCGACATTATTGAAAAAGAGACTGGTAAACAACCTTGTAATACAAGAGT
 CTATTCTGATGTTGCCTGAAGAGGTGGAGGAGGTGATTGGGAACAAACCGGAGTCTGAT
 ATACTTGTTCATACCGCCTATGACGAATCTACTGATGAGAATGTGATGCTTTTACGTC
 AGACGCTCCCGAGTACAAACCCTGGGCTCTGGTATTTCAGGACAGCAATGGTGAGAATA
 AGATTAATAATGTTGAGTGGGGGCTCAAAGCGCACGGCTGACGGTAGCGAATTTGAGAGC
 CCCAAAAAAAACGAAAGGTCGAAAtaa

10

20

30

【 0 0 6 3 】

「塩基編集活性」とは、ポリヌクレオチド内の塩基を化学的に改変するように作用することを意味する。一実施形態では、第1の塩基が第2の塩基に変換される。一実施形態では、塩基編集活性は、例えば標的C・GをT・Aに変換する、シチジンデアミナーゼ活性である。別の実施形態では、塩基編集活性は、例えばA・TをG・Cに変換する、アデノシンデアミナーゼ活性である。

40

【 0 0 6 4 】

用語「塩基エディターシステム」は、標的ヌクレオチド配列の核酸塩基を編集するための系を指す。いくつかの実施形態において、塩基エディターシステムは、(1)ポリヌクレオチドプログラミング可能なヌクレオチド結合ドメインと、核酸塩基を脱アミノ化するためのデアミナーゼドメインとを含む塩基エディター (BE)、および(2) ポリヌクレオチドプログラム可能ヌクレオチド結合ドメインに付随するガイドポリヌクレオチド(例えばガイドRNA)を含む。いくつかの実施形態では、ポリヌクレオチドプログラミング可能なヌクレオチド結合ドメインは、ポリヌクレオチドプログラミング可能なDNA結合ドメインである。いくつかの実施形態では、塩基エディターは、シチジン塩基エディター (CBE) である。

50

いくつかの実施形態では、塩基エディターは、アデノシン塩基エディター（ABE）である。

【0065】

一部の実施形態では、核酸塩基エディターシステムは、複数の塩基編集コンポーネントを含むことができる。例えば、核酸塩基エディターシステムは、複数のデアミナーゼを含むことができる。ある態様において、ヌクレアーゼ塩基編集システムは、1以上のシチジンデアミナーゼおよび/または1以上のアデノシンデアミナーゼを含み得る。いくつかの実施形態において、単一のガイドポリヌクレオチドを利用して、異なるデアミナーゼを標的核酸配列にターゲティングすることができる。ある態様において、一对のガイドポリヌクレオチドが、異なるデアミナーゼを標的核酸配列にターゲティングするために利用され得る。

10

【0066】

塩基エディターシステムの核酸塩基成分およびポリヌクレオチドプログラミング可能なヌクレオチド結合成分は、互いに共有結合的または非共有結合的により結合され得る。例えば、いくつかの実施形態において、デアミナーゼドメインが、ポリヌクレオチドプログラミング可能なヌクレオチド結合ドメインによって標的ヌクレオチド配列にターゲティングされ得る。ある態様において、ポリヌクレオチドプログラミング可能なヌクレオチド結合ドメインは、デアミナーゼドメインに融合または連結され得る。いくつかの実施形態において、ポリヌクレオチドプログラミング可能なヌクレオチド結合ドメインは、デアミナーゼドメインと非共有結合的に相互作用または結合することによって、デアミナーゼドメインを標的ヌクレオチド配列にターゲティングすることができる。例えば、いくつかの実施形態において、核酸塩基編集成分、例えばデアミナーゼ成分は、ポリヌクレオチドプログラミング可能なヌクレオチド結合ドメインの一部をなすさらなる異種部分またはドメインと相互作用し、会合し、または複合体を形成することができる、さらなる異種部分またはドメインを含むことができる。いくつかの実施形態において、追加の異種部分は、ポリペプチドと結合し、相互作用し、会合し、または複合体を形成することができる。いくつかの実施形態において、追加の異種部分は、ガイドポリヌクレオチドに結合することができる。いくつかの実施形態において、追加の異種部分は、ポリペプチドリンカーに結合することができる。いくつかの実施形態において、追加の異種部分は、ポリヌクレオチドリンカーに結合することができる。追加の異種部分は、タンパク質ドメインであってもよい。いくつかの実施形態において、追加の異種部分は、K相同(KH)ドメイン、MS2コートタンパク質ドメイン、PP7コートタンパク質ドメイン、SfMu Comコートタンパク質ドメイン、ステリルモチーフ、テロメラーゼKu結合モチーフおよびKuタンパク質、テロメラーゼSm7結合モチーフおよびSm7タンパク質、またはRNA認識モチーフであり得る。

20

30

【0067】

塩基編集システムは、ガイドポリヌクレオチド成分をさらに含むことができる。塩基編集システムの構成要素は、共有結合、非共有結合的相互作用、またはそれらの結合および相互作用の任意の組合せを介して互いに結合され得ることが理解されるべきである。ある態様において、デアミナーゼドメインは、ガイドポリヌクレオチドによって標的ヌクレオチド配列にターゲティングされ得る。例えば、いくつかの実施形態において、塩基エディターシステムの核酸塩基編集成分、例えばデアミナーゼ成分は、ガイドポリヌクレオチドの一部またはセグメント(例えばポリヌクレオチドモチーフ)と相互作用し、結合し、または複合体を形成し得るさらなる異種部分またはドメイン(例えばRNAまたはDNA結合タンパク質のようなポリヌクレオチド結合ドメイン)を含み得る。いくつかの実施形態において、追加の異種部分またはドメイン(例えばRNAまたはDNA結合タンパク質などのポリヌクレオチド結合ドメイン)は、デアミナーゼドメインに融合または連結され得る。いくつかの実施形態において、追加の異種部分は、ポリペプチドと結合し、相互作用し、会合し、またはポリペプチドと複合体を形成することができる。いくつかの実施形態において、追加の異種部分は、ポリヌクレオチドと結合し、相互作用し、会合し、またはポリヌクレオチ

40

50

ドと複合体を形成することができる。いくつかの実施形態において、追加の異種部分は、ガイドポリヌクレオチドに結合することができる。いくつかの実施形態において、追加の異種部分は、ポリペプチドリinkerに結合することができる。いくつかの実施形態において、追加の異種部分は、ポリヌクレオチドリinkerに結合することができる。追加の異種部分は、タンパク質ドメインであってもよい。いくつかの実施形態において、追加の異種部分は、K相同(KH)ドメイン、MS2コートタンパク質ドメイン、PP7コートタンパク質ドメイン、SfMu Comコートタンパク質ドメイン、無菌アルファモチーフ、テロメラーゼKu結合モチーフおよびKuタンパク質、テロメラーゼSm7結合モチーフおよびSm7タンパク質、またはRNA認識モチーフであり得る。

【0068】

ある態様において、塩基編集システムは、塩基除去修復 (BER) コンポーネントの阻害因子をさらに含むことができる。塩基編集システムの構成要素は、共有結合、非共有結合的相互作用、またはそれらの結合および相互作用の任意の組み合わせを介して互いに結合され得ることが理解されるべきである。BER成分の阻害因子は、塩基除去修復阻害因子を含み得る。ある態様において、塩基除去修復の阻害因子は、ウラシルDNAグリコシラーゼ阻害因子 (UGI) であり得る。ある態様において、塩基除去修復の阻害因子は、イノシン塩基除去修復阻害因子であり得る。ある態様において、塩基除去修復の阻害因子は、ポリヌクレオチドプログラミング可能なヌクレオチド結合ドメインにより標的ヌクレオチド配列にターゲティングされ得る。ある態様において、ポリヌクレオチドプログラミング可能なヌクレオチド結合ドメインは、塩基除去修復の阻害因子に融合または連結され得る。ある態様において、ポリヌクレオチドプログラミング可能なヌクレオチド結合ドメインは、デアミナーゼドメインおよび塩基除去修復の阻害因子に融合または連結され得る。いくつかの実施形態において、ポリヌクレオチドプログラム可能なヌクレオチド結合ドメインは、塩基除去修復の阻害因子と非共有結合的に相互作用するか、または塩基除去修復の阻害因子と会合することによって、塩基除去修復の阻害因子を標的ヌクレオチド配列へとターゲティングすることができる。例えば、いくつかの実施形態において、塩基除去修復成分の阻害因子は、ポリヌクレオチドプログラミング可能なヌクレオチド結合ドメインの一部であるさらなる異種部分またはドメインと相互作用、会合、または複合体形成し得るさらなる異種部分またはドメインを含み得る。ある態様において、塩基除去修復の阻害因子は、ガイドポリヌクレオチドにより標的ヌクレオチド配列にターゲティングされ得る。例えば、いくつかの実施形態において、塩基除去修復の阻害因子は、ガイドポリヌクレオチドの一部またはセグメント(例えば、ポリヌクレオチドモチーフ)と相互作用、会合、または複合体形成し得るさらなる異種部分またはドメイン(例えば、RNAまたはDNA結合タンパク質のようなポリヌクレオチド結合ドメイン)を含み得る。いくつかの実施形態において、ガイドポリヌクレオチドのさらなる異種部分またはドメイン(例えば、RNAまたはDNA結合タンパク質のようなポリヌクレオチド結合ドメイン)は、塩基除去修復の阻害因子に融合または連結され得る。いくつかの実施形態において、追加の異種部分は、ポリヌクレオチドと結合、相互作用、会合、または複合体形成することができる。いくつかの実施形態において、追加の異種部分は、ガイドポリヌクレオチドに結合することができる。いくつかの実施形態において、追加の異種部分は、ポリペプチドリinkerに結合することができる。いくつかの実施形態において、追加の異種部分は、タンパク質ドメインであってもよい。いくつかの実施形態において、追加の異種部分は、K相同(KH)ドメイン、MS2コートタンパク質ドメイン、PP7コートタンパク質ドメイン、SfMu Comコートタンパク質ドメイン、無菌アルファモチーフ、テロメラーゼKu結合モチーフおよびKuタンパク質、テロメラーゼSm7結合モチーフおよびSm7タンパク質、またはRNA認識モチーフであり得る。

【0069】

用語「Cas9」または「Cas9ドメイン」とは、Cas9タンパク質またはその断片(例えばCas9の活性、不活性、または部分的に活性なDNA切断ドメイン、および/またはCas9のgRNA結合ドメインを含むタンパク質)を含むRNA誘導型ヌクレアーゼを表す。Cas9ヌクレア

10

20

30

40

50

ーゼは、casn1ヌクレアーゼまたはCRISPR (clustered regularly interspaced short palindromic repeat) 関連ヌクレアーゼとも呼ばれる。例示的なCas9は、アミノ酸配列が以下に提供されるStreptococcus pyogenes Cas9である :

MDKKYSIGLDIGTNSVGVAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGET
 AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSEFFHRLEESFLVEEDKKHERHPIFGNI
 VDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPNDSVDKLF
 QLVQIYNQLFEENPINASRVDAKAILSARLSKSRLENLIAQLPGEKRNGLFGNLIALSGLTP
 NFKSNFDLAEDAKLQLSKDQYDDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITK
 APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
 ILEKMDGTEELLVKNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEK
 ILTFRIPIYVVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK
 VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK
 KIECFDSVEISGVEDRFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEE
 RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIH
 DDSLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILOTVKIVDELVKVMGHKPENIVIEM
ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQE
LDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKL
ITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKV
ITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVR
KMIAKSEOEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATV
RKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
 KVEKGKSKKLKSVKELLGITIMERSSEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK
 RMLASAGELQKGNELALPSKYVNFLYLASHYEKLGKSPEDNEQKQLFVEQHKHYLDEIEEQISE
 FSKRVILADANLDKVL SAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTST
 KEVL DATLIHQ SITGLYETRIDLSQLGGD

10

20

30

(一本下線 : HNHドメイン ; 二重下線 : RuvCドメイン) 。

【 0 0 7 0 】

「保存的アミノ酸置換」または「保存的変異」という用語は、あるアミノ酸が共通の特性を有する別のアミノ酸に置き換わることを指す。個々のアミノ酸間の共通の性質を定義するための機能的な方法は、相同的な生物の対応するタンパク質間のアミノ酸変化の正規化された頻度を分析することである(Schulz, G. E. and Schirmer, R. H., Principles of Protein Structure, Springer-Verlag, New York (1979))。このような分析によれば、グループ内のアミノ酸が互いに優先的に交換され、それゆえに全体的なタンパク質構造に対するそれらの影響において互いに最も類似しているところアミノ酸のグループを定義することができる(上記Schulz, G. E. and Schirmer, R. H.)。保存変異の非限定的な例としては、例えば正電荷を維持することができるアミノ酸アルギニンからリジンおよびその逆 ; 負電荷を維持することができるアスパラギン酸からグルタミン酸およびその逆 ; 遊離の-OHが維持されるトレオニンからセリン ; および遊離-NH₂を維持できるアスパラギンからグルタミンのようなアミノ酸置換が挙げられる。

40

【 0 0 7 1 】

50

用語「Cas9」または「Cas9ドメイン」とは、Cas9タンパク質またはその断片(例えばCas9の活性、不活性、または部分的に活性なDNA切断ドメイン、および/またはCas9のgRNA結合ドメインを含むタンパク質)を含むRNA誘導型ヌクレアーゼを表す。Cas9ヌクレアーゼは、casn1ヌクレアーゼまたはCRISPR (clustered regularly interspaced short palindromic repeat) 関連ヌクレアーゼとも呼ばれる。例示的なCas9は、アミノ酸配列が以下に提供されるStreptococcus pyogenes Cas9である：

MDKKYSIGLDIGTNSVGWAVITDDYKVPSSKFKVLGNTDRHSIKKNLIGALLFGSGET
 AEATRLKRTARRRYTRRKNRICYLQEI FSNEMAKVDDSFHRLEESFLVEEDKKHERHPI FGNI
 VDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFI 10
 QLVQIYNQLFEENPINASRVDAKAILSARLSKSRLENLIAQLPGEKRNGLFGNLIALLSLGLTP
 NFKSNFDLAEDAKLQLSKDQYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITK
 APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
 ILEKMDGTEELLVKNLREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEK
 ILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK
 VLPKHSLLEYEFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK
 KIECFDSVEISGVEDRFNASLGAYHDLKI IKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEE 20
 RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDFANRNFQMQLIH
 DDSLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILOTVKIVDELVKVMGHKPENIVIEM
ARENQTTQKGQNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQE
LDINRLSDYDVDHIVPQSFIKDSDIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKL
ITQRKFDNLTKAERGGLSELDKAGFIKROLVETROI TKHVAQILD SRMNTKYDENDKLIREVKV
ITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVR
KMIKSEQIEGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATV
RKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA 30
 KVEKGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKDLI IKLPKYSLFELENGRK
 RMLASAGELQKGNELALPSKYVNFYLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEIIEQISE
 FSKRVILADANLDKVL SAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTST
 KEVL DATLIHQSI TGLYETRIDLSQLGGD

(一本下線：HNHドメイン；二重下線：RuvCドメイン)。

【0072】

用語「コード配列」または「タンパク質コード配列」は本明細書中で交換可能に使用され、タンパク質をコードするポリヌクレオチドのセグメントを表す。この領域または配列は、5'末端に近い方が開始コドンで境界され、3'末端に近い方が停止コドンで境界される。コード配列はオープンリーディングフレームとも呼ばれる。

【0073】

「保存的アミノ酸置換」または「保存的変異」という用語は、あるアミノ酸が共通の特性を有する別のアミノ酸に置き換わることを指す。個々のアミノ酸間の共通の性質を定義するための機能的な方法は、相同的な生物の対応するタンパク質間のアミノ酸変化の正規化された頻度を分析することである(Schulz, G. E. and Schirmer, R. H., Principles of Protein Structure, Springer-Verlag, New York (1979))。このような分析によれば

10

20

30

40

50

、グループ内のアミノ酸が互いに優先的に交換され、それゆえに全体的なタンパク質構造に対するそれらの影響において互いに最も類似しているところアミノ酸のグループを定義することができる(上記Schulz, G. E. and Schirmer, R. H.)。保存変異の非限定的な例としては、例えば正電荷を維持することができるアミノ酸アルギニンからリジンおよびその逆；負電荷を維持することができるアスパラギン酸からグルタミン酸およびその逆；遊離の-OHが維持されるトレオニンからセリン；および遊離-NH₂を維持できるアスパラギンからグルタミンのようなアミノ酸置換が挙げられる。

【0074】

「シチジンデアミナーゼ」とは、アミノ基をカルボニル基に変換する脱アミノ化反応を触媒することができるポリペプチドまたはその断片を意味する。1つの実施形態において、シチジンデアミナーゼは、シトシンをウラシルに、または5-メチルシトシンをチミンに変換する。Petromyzon marinus由来のPmCDA1 (Petromyzon marinus cytosine deaminase 1、「PmCDA1」)、哺乳動物(例えば、ヒト、ブタ、ウシ、ウマ、サル等)由来のAID(活性化誘導シチジンデアミナーゼ; AICDA)、およびAPOBECは、例示的シチジンデアミナーゼである。

10

【0075】

本明細書中で使用する場合、用語「デアミナーゼ」または「デアミナーゼドメイン」とは、脱アミノ化反応を触媒するタンパク質または酵素を指す。ある態様において、デアミナーゼまたはデアミナーゼドメインは、それぞれウリジンまたはデオキシウリジンへのシチジンまたはデオキシシチジンの加水分解的脱アミノ化を触媒するシチジンデアミナーゼである。ある態様において、デアミナーゼまたはデアミナーゼドメインはシトシンデアミナーゼであり、ウラシルへのシトシンの加水分解的脱アミノ化を触媒する。ある態様において、デアミナーゼは、ヒポキサンチンへのアデニンの加水分解的脱アミノ化を触媒するアデニンデアミナーゼである。

20

【0076】

ある態様において、デアミナーゼまたはデアミナーゼドメインは、例えばヒト、チンパンジー、ゴリラ、サル、ウシ、イヌ、ラット、またはマウスなどの生物由来の天然デアミナーゼのバリエーションである。ある態様において、デアミナーゼまたはデアミナーゼドメインは、非天然のものである。例えば、いくつかの実施形態において、デアミナーゼまたはデアミナーゼドメインは、天然のデアミナーゼに対して少なくとも50%、少なくとも55%、少なくとも60%、少なくとも65%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも91%、少なくとも92%、少なくとも93%、少なくとも94%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、少なくとも99.1%、少なくとも99.2%、少なくとも99.3%、少なくとも99.4%、少なくとも99.5%、少なくとも99.6%、少なくとも99.7%、少なくとも99.8%、または少なくとも99.9%の同一性を有する。例えば、デアミナーゼドメインは、国際PCT出願番号PCT/2017/045381(国際公開第2018/027078号)およびPCT/US 2016/058344(国際公開第2017/070632号)に記載されており、これらの各々は、その全体が参照により本明細書に組み込まれる。Komor, A.C., et al., "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., et al., "Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); Komor, A.C., et al., "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaa04774 (2017), and Rees, H.A., et al., "Base editing: precision chemistry on the genome and transcriptome of living cells." Nat Rev Genet. 2018 Dec;19(12):770-788. doi: 10.1038/s41576-018-0059-1も参照(その内容全体を参照により本明細書に組み込む)。

30

40

【0077】

「検出可能な標識」とは、目的の分子に連結されたときに、分光学的、光化学的、生化

50

学的、免疫化学的、または化学的手段を介してそれを検出可能にする組成物を意味する。例えば、有用な標識には、放射性同位体、磁気ビーズ、金属ビーズ、コロイド粒子、蛍光色素、高電子密度試薬、酵素(例えばELISAで一般的に使用されるもの)、ビオチン、ジゴキシゲニン、またはハプテンが含まれる。

【0078】

「疾患」とは、細胞、組織または器官の正常な機能を損傷または妨害するあらゆる状態または障害を意味する。疾患の例には、網膜色素変性症、Usher症候群、鎌状赤血球症、サラセミア、 α 1アンチトリプシン欠損症(A1AD)、肝ポルフィリン症、中鎖アシルCoA脱水素酵素(MCAD)欠損症、リソソーム酸性リパーゼ(LAL)欠損症、フェニルケトン尿症、ヘモクロマトーシス、Von Gierke病、Pompe病、Gaucher病、Hurler症候群、嚢胞性線維症、慢性疼痛などがある。特定の一実施形態では、疾患はA1ADである。

10

【0079】

「有効量」とは、未治療の患者と比較して、疾患の症状を改善するために必要な、薬剤または活性化化合物(例えば本明細書に記載の塩基エディター)の量を意味する。疾患の治療的処置のために本発明を実施するために使用される活性化化合物の有効量は、投与態様、対象の年齢、体重、および全般的健康状態に依存して変化する。最終的には、主治医または獣医が適切な量および用量を決定する。このような量は「有効」量と呼ばれる。一実施形態では、有効量は、細胞(例えば、インビトロまたはインビボの細胞)内の目的の遺伝子に改変を導入するのに十分な本発明の塩基エディターの量である。一実施形態では、有効量は、治療効果を達成する(例えば、網膜色素変性、アッシャー症候群、鎌状赤血球症、 α -サラセミア、 α -1アンチトリプシン欠損症(A1AD)、肝ポルフィリン症、中鎖アシルCoAデヒドロゲナーゼ(MCAD)欠損症、リソソーム酸性リパーゼ(LAL)欠損症、フェニルケトン尿症、ヘモクロマトーシス、フォン・ギールケ病、ポンペ病、ゴーシェ病、ハーラー症候群、嚢胞性線維症、または慢性疼痛を低減または制御する)のために必要な塩基エディターの量である。このような治療効果は、対象、組織または器官の全ての細胞において病原性遺伝子を変化させるのに十分である必要はなく、対象、組織または器官に存在する細胞の約1%、5%、10%、25%、50%、75%またはそれ以上において病原性遺伝子を変化させるだけでよい。一実施形態では、有効量は、疾患(例えば、網膜色素変性、Usher症候群、鎌状赤血球症、サラセミア、 α 1アンチトリプシン欠損症(A1AD)、肝ポルフィリン症、中鎖アシルCoA脱水素酵素(MCAD)欠損症、リソソーム酸性リパーゼ(LAL)欠損症、フェニルケトン尿症、ヘモクロマトーシス、フォン・ゲルケ病、ポンペ病、ゴーシェ病、ハーラー症候群、嚢胞性線維症、または慢性疼痛)の一つ以上の症状を改善するのに十分である。

20

30

【0080】

「断片」とは、ポリペプチドまたは核酸分子の一部を意味する。この部分は、好ましくは、参照核酸分子またはポリペプチドの全長の少なくとも10%、20%、30%、40%、50%、60%、70%、80%、または90%を含む。断片は、10、20、30、40、50、60、70、80、90、または100、200、300、400、500、600、700、800、900、または1000のヌクレオチドまたはアミノ酸を含み得る。

【0081】

「ハイブリダイゼーション」は、相補的な核酸塩基間の水素結合を意味し、ワトソン-クリック、フーグスティーンまたは逆フーグスティーン水素結合であり得る。たとえば、アデニンとチミンは相補的な核酸塩基で、水素結合を形成して対を形成する。

40

【0082】

用語「塩基修復の阻害因子」、「塩基修復阻害因子」、「IBR」または他の文法的等価物は、核酸修復酵素、例えば塩基除去修復酵素の活性を阻害することができるタンパク質を指す。塩基修復の阻害因子の非限定的な例としては、APE1、Endo III、Endo IV、Endo V、Endo VIII、Fpg、hOGGI、hNEIL1、T7 EndoI、T4 PDG、UDG、hSMUGLおよびhAAGの阻害因子が挙げられる。ある実施形態では、塩基修復阻害因子は、Endo VまたはhAAGの阻害因子である。ある実施形態では、塩基修復阻害因子は、触媒的に不活性なE

50

EndoVまたは触媒的に不活性なhAAGである。ある実施形態では、塩基修復阻害因子は、ウラシルグリコシラーゼ阻害因子 (UGI) である。UGIとは、ウラシル-DNAグリコシラーゼの塩基除去修復酵素を阻害することができるタンパク質を指す。いくつかの実施形態において、UGIドメインは、野生型UGIまたはその断片を含む。いくつかの実施形態において、本明細書で提供されるUGIタンパク質は、UGIの断片、および、UGIまたはUGI断片に相同的なタンパク質を含む。ある態様において、塩基修復阻害因子は、イノシン塩基除去修復の阻害因子である。いくつかの態様において、塩基修復阻害因子は、「触媒的に不活性なイノシン特異的ヌクレアーゼ」または「死んだイノシン特異的ヌクレアーゼ」である。

【0083】

いかなる特定の理論にも拘束されることを望まないが、触媒的に不活性なイノシリングリコシラーゼ (例えば、アルキルアデニングリコシラーゼ (AAG)) は、イノシンに結合することができるが、脱塩基部位を作ることイノシンを除去することもできず、それによって、新たに形成されるイノシン部分をDNA損傷/修復機構から立体的にブロックする。いくつかの実施形態において、触媒的に不活性なイノシン特異的ヌクレアーゼは、核酸中のイノシンに結合することができるが、核酸を切断しない。代表的な、触媒的に不活性なイノシン特異的ヌクレアーゼの非限定的例としては、(例えばヒト由来の) 触媒的に不活性なアルキルアデニングリコシラーゼ(AAGヌクレアーゼ)、および(例えばE.coli由来の) 触媒的に不活性なエンドヌクレアーゼV (EndoVヌクレアーゼ)が挙げられる。いくつかの実施形態において、触媒的に不活性なAAGヌクレアーゼは、E125Q突然変異または別のAAGヌクレアーゼにおける対応する突然変異を含む。

【0084】

用語「単離された」、「精製された」、または「生物学的に純粋な」とは、その天然状態で見出される場合に通常それに付随する成分が、様々な程度まで除去されている物質を指す。「単離」は、元の入手源又は周囲環境からの分離の程度を示す。「精製」は、単離よりも高い分離度を示す。「精製された」又は「生物学的に純粋な」タンパク質は、不純物がタンパク質の生物学的特性に実質的に影響を与えたり他の有害な結果を引き起こさないように、他の物質が十分に除去されている。すなわち、本発明の核酸またはペプチドは、組換えDNA技術によって産生された場合には細胞物質、ウイルス物質または培地を実質的に含まない場合、あるいは化学的に合成された場合には化学的前駆体その他の化学物質を実質的に含まない場合には、精製されている。純度および均一性は、典型的には、分析化学技術、例えば、ポリアクリルアミドゲル電気泳動または高速液体クロマトグラフィーを用いて決定される。用語「精製された」は、核酸またはタンパク質が電気泳動ゲルにおいて本質的に1つのバンドを生じることを意味し得る。修飾、例えば、リン酸化またはグリコシル化を受けることができるタンパク質については、異なる修飾は、別々に精製することができる異なる単離されたタンパク質を生じ得る。

【0085】

「単離ポリヌクレオチド」とは、本発明の核酸分子が由来する生物の天然ゲノムにおいて当該遺伝子に隣接する遺伝子を含まない核酸(例えばDNA)を意味する。したがって、この用語は、例えば、ベクターに組み込まれた；自律的に複製するプラスミドやウイルスに組み込まれた；原核生物や真核生物のゲノムDNAに組み込まれた；または他の配列とは独立した別の分子(例えば、PCRまたは制限エンドヌクレアーゼ消化によって生成されたcDNAまたはゲノムもしくはcDNA断片)として存在する組換えDNAを含む。さらに、この用語は、DNA分子から転写されるRNA分子、ならびに、さらなるポリペプチド配列をコードするハイブリッド遺伝子の一部である組換えDNAを含む。

【0086】

「単離ポリペプチド」とは、天然状態で付随する成分から分離された本発明のポリペプチドを意味する。典型的には、ポリペプチドは、それが天然状態で会合しているタンパク質および天然有機分子から重量で少なくとも60%フリーである場合に、単離されている。好ましくは、調製物は、重量で少なくとも75%、より好ましくは少なくとも90%、最も好ましくは少なくとも99%が本発明のポリペプチドである。本発明の単離されたポリペプチ

ドは、例えば、天然源からの抽出、そのようなポリペプチドをコードする組換え核酸の発現；または化学的にタンパク質を合成することにより得ることができる。純度は、任意の適切な方法、例えば、カラムクロマトグラフィー、ポリアクリルアミドゲル電気泳動、またはHPLC分析によって測定することができる。

【0087】

本明細書において使用される用語「リンカー」は、二つの分子もしくは部分（例えばタンパク質複合体もしくはリボ核複合体の二つの成分、または融合タンパク質の二つのドメイン、例えば、ポリヌクレオチドプログラミング可能DNA結合ドメイン(例dCas9)とデアミナーゼドメイン(例えばアデノシンデアミナーゼまたはシチジンデアミナーゼ)）を連結する共有結合リンカー(例えば共有結合)、非共有結合リンカー、化学基、または分子を指すことができる。リンカーは、塩基エディターシステムの異なるコンポーネントまたはコンポーネントの異なる部分を繋げることができる。例えば、いくつかの実施形態において、リンカーは、ポリヌクレオチドプログラミング可能ヌクレオチド結合ドメインのガイドポリヌクレオチド結合ドメインおよびデアミナーゼの触媒ドメインを繋げることができる。ある態様において、リンカーは、CRISPRポリペプチドおよびデアミナーゼを結合することができる。ある態様において、リンカーは、Cas9およびデアミナーゼを繋げることができる。ある態様において、リンカーは、dCas9およびデアミナーゼを繋げることができる。ある態様において、リンカーは、nCAs9およびデアミナーゼを繋げることができる。ある態様において、リンカーは、ガイドポリヌクレオチドおよびデアミナーゼを繋げることができる。いくつかの実施形態において、リンカーは、塩基エディターシステムの脱アミノ化成分およびポリヌクレオチドプログラミング可能なヌクレオチド結合成分を繋げることができる。いくつかの実施形態において、リンカーは、塩基エディターシステムの脱アミノ化成分のRNA結合部分と、ポリヌクレオチドプログラミング可能なヌクレオチド結合成分とを繋げることができる。いくつかの実施形態において、リンカーは、塩基エディターシステムの脱アミノ化成分のRNA結合部分と、ポリヌクレオチドプログラミング可能なヌクレオチド結合成分のRNA結合部分とを結合することができる。リンカーは、2つの基、分子、または他の部分の間に配置され、またはそれらによって挟まれ、共有結合または非共有結合相互作用を介してそれぞれに連結され、かくして2つを連結することができる。ある態様において、リンカーは、有機分子、基、ポリマー、または化学的部分であり得る。ある態様において、リンカーはポリヌクレオチドであり得る。ある態様において、リンカーはDNAリンカーであり得る。ある態様において、リンカーはRNAリンカーであり得る。ある態様において、リンカーは、リガンドに結合することができるアプタマーを含むことができる。ある態様において、リガンドは、炭水化物、ペプチド、タンパク質、または核酸であり得る。いくつかの実施形態において、リンカーは、リボスイッチに由来するアプタマーを含むことができる。アプタマーが由来するリボスイッチは、テオフィリンリボスイッチ、チアミンピロリン酸 (TPP) リボスイッチ、アデノシンコバラミン (AdoCbl) リボスイッチ、S-アデノシルメチオニン (SAM) リボスイッチ、SAHリボスイッチ、フラビンモノヌクレオチド (FMN) リボスイッチ、テトラヒドロ葉酸リボスイッチ、リジンリボスイッチ、グリシンリボスイッチ、プリンリボスイッチ、GlmSリボスイッチ、またはプレクエオシン1 (PreQ1) リボスイッチから選択され得る。ある態様において、リンカーは、ポリペプチドまたはポリペプチドリガンドなどのタンパク質ドメインに結合したアプタマーを含み得る。ある態様において、ポリペプチドリガンドは、K相同(KH)ドメイン、MS2コートタンパク質ドメイン、PP7コートタンパク質ドメイン、SfMu Comコートタンパク質ドメイン、無菌のモチーフ、テロメラーゼKu結合モチーフおよびKuタンパク質、テロメラーゼSm7結合モチーフおよびSm7タンパク質、またはRNA認識モチーフであり得る。ある態様において、ポリペプチドリガンドは、塩基エディターシステム成分の一部であり得る。例えば、核酸塩基編集成分は、デアミナーゼドメインおよびRNA認識モチーフを含み得る。

【0088】

ある態様において、リンカーは、アミノ酸または複数のアミノ酸(例えばペプチドまたは

10

20

30

40

50

タンパク質)であり得る。ある態様において、リンカーは、長さが約5~100アミノ酸、例えば、長さが約5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、20-30、30~40、40~50、50~60、60~70、70~80、80~90または90~100アミノ酸であり得る。ある態様において、リンカーは、長さが約100~150、150~200、200~250、250~300、300~350、350~400、400~450、または450~500アミノ酸であり得る。より長いまたはより短いリンカーも考えられる。

【0089】

ある態様において、リンカーは、Cas9ヌクレアーゼドメインを含むRNAプログラム可能ヌクレアーゼのgRNA結合ドメインと、核酸編集タンパク質(例えば、シチジンまたはアデノシンデアミナーゼ)の触媒ドメインとを繋げる。ある態様において、リンカーは、dCas9および核酸編集タンパク質を繋げる。例えば、リンカーは、2つの基、分子、または他の部分の間に配置されるか、または2つの基、分子、または他の部分によって隣接され、共有結合を介してそれぞれに連結され、したがって2つを連結する。ある態様において、リンカーは、アミノ酸または複数のアミノ酸(例えば、ペプチドまたはタンパク質)である。ある態様において、リンカーは、有機分子、基、ポリマー、または化学的部分である。ある態様において、リンカーは、長さが5~200アミノ酸、例えば、長さが5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、35、45、50、55、60、60、65、70、70、75、80、85、90、90、95、100、101、102、103、104、105、110、120、130、140、150、160、175、180、190、または200アミノ酸である。より長いまたはより短いリンカーも考えられる。ある態様において、リンカーは、XTENリンカーとも呼ばれ得るアミノ酸配列SGSETPGTSESATPESを含む。ある態様において、リンカーは、アミノ酸配列SGGSを含む。ある態様において、リンカーは、(SGGS)_n、(GGGS)_n、(GGGGGS)_n、(G)_n、(EAAAK)_n、(GGG)_n、SGSETPGTSESATPES、または(XP)_nモチーフ、またはこれらの任意の組合せを含み、ここで、nは独立して1~30の整数であり、Xは任意のアミノ酸である。いくつかの実施態様において、nは、1、2、3、4、5、6、7、8、9、10、11、12、13、14又は15である。ある態様において、リンカーは、複数のプロリン残基を含み、長さが5~21、5~14、5~9、5~7アミノ酸であり、例えば、PAPAP、PAPAPA、PAPAPAP、PAPAPAPA、P(AP)₄、P(AP)₇、P(AP)₁₀である。このようなプロリンに富むリンカーはまた、「リジッド(rigid)」リンカーとも呼ばれる。

【0090】

ある態様において、塩基エディターのドメインは、SGGSSGSETPGTSESATPESGGGS、SGGSSGGSSGSETPGTSESATPESGGSSGGGS、またはGGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTE PSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSSGGGSのアミノ酸配列を含むリンカーを介して融合される。ある態様において、塩基エディターのドメインは、XTENリンカーとも呼ばれ得るアミノ酸配列SGSETPGTSESATPESを含むリンカーを介して融合される。ある態様において、リンカーは、長さが24アミノ酸である。ある態様において、リンカーは、アミノ酸配列SGGSSGGSSGSETPGTSESATPESを含む。ある態様において、リンカーは、40アミノ酸長である。ある態様において、リンカーは、アミノ酸配列SGGSSGGSSGSETPGTSESATPESGGSSGGSSGSSGSSGGGSを含む。ある態様において、リンカーは、64アミノ酸長である。ある態様において、リンカーは、アミノ酸配列SGGSSGGSSGSETPGTSESATPESGGSSGGSSGSSGSSGSSGSETPGTSESATPESGGSSGGGSを含む。ある態様において、リンカーは、長さが92アミノ酸である。いくつかの実施形態では、リンカーは、アミノ酸配列PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSSGGGSを含む。

【0091】

本明細書中で使用される場合、用語「変異」とは、配列内、例えば核酸もしくはアミノ酸配列内の残基が、別の残基により置換されること、または配列内の1以上の残基の欠失もしくは挿入をいう。変異は、本明細書中において典型的には、元の残基を同定し、次いで配列内の残基の位置を同定し、そして新たに置換された残基を同定することによって、

記載される。本明細書中に提供されるアミノ酸置換(変異)を作製するための種々の方法は、当該技術分野においてよく知られており、例えば、Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012))によって提供されている。いくつかの態様において、本開示の塩基エディターは、有意な数の非意図的な変異、例えば非意図的な点突然変異を生成することなく、核酸(例えば対象のゲノム内の核酸)において「意図された変異」例えば点突然変異を効率的に生成することができる。いくつかの実施形態において、意図された変異は、その意図された変異を生じさせるように特に設計された、ガイドポリヌクレオチド(例えばgRNA)に結合した特定の塩基エディター(例えばシチジン塩基エディターまたはアデノシン塩基エディター)によって生じる変異である。

10

【0092】

一般に、配列(例えば、本明細書に記載のアミノ酸配列)においてなされるか又は同定される変異は、参照(または野生型)配列、すなわち、その変異を含まない配列に関連して番号付けされる。当業者は、参照配列に対するアミノ酸および核酸配列における突然変異の位置を決定する方法を容易に理解するであろう。

【0093】

「核局在化配列」、「核局在化シグナル」または「NLS」という用語は、タンパク質の細胞核への移入を促進するアミノ酸配列を意味する。核局在化配列は当該技術分野において公知であり、例えば、2000年11月23日に出願され2001年5月31日にWO/2001/038547として公開された国際PCT出願PCT/EP 2000/011690のPlank et al.に記載されており、その内容は、例示的な核局在化配列の開示について参照により本明細書に組み入れられる。他の実施形態では、NLSは、例えばKoblan et al., *Nature Biotech.* 2018 doi:10.1038/nbt.4172によって記載された、最適化されたNLSである。ある態様において、NLSは、アミノ酸配列KRTADGSEFESPKKKRKV, KRPAATKKAGQAKKKK, KKTELQT TNAENKTKKL, KRGINDRNFWRGENGRKTR, RKSGKIAAIVVKRPRK, PKKKRKV,またはMDSLLMNRKFLYQFKNVRWAKGRRETYLCを含む。

20

【0094】

本明細書中で使用される場合、用語「核酸」および「核酸分子」とは、核酸塩基および酸性部分を含む化合物、例えばヌクレオシド、ヌクレオチド、またはヌクレオチドのポリマーをいう。典型的には、ポリマー核酸、例えば3つ以上のヌクレオチドを含む核酸分子は、隣接するヌクレオチドがホスホジエステル結合を介して互いに連結されている直鎖分子である。ある態様において、「核酸」は、個々の核酸残基(例えばヌクレオチドおよび/またはヌクレオシド)を指す。ある態様において、「核酸」は、三つ以上の個々のヌクレオチド残基を含むオリゴヌクレオチド鎖をいう。本明細書中で使用される、用語「オリゴヌクレオチド」、「ポリヌクレオチド」、および「ポリ核酸」は、ヌクレオチドのポリマー(例えば少なくとも3つのヌクレオチドの鎖)を指すために交換可能に使用され得る。ある態様において、「核酸」は、RNAならびに一本鎖および/または二本鎖DNAを包含する。核酸は、例えば、ゲノム、転写物、mRNA、tRNA、rRNA、siRNA、snRNA、プラスミド、コスミド、染色体、染色分体、または他の天然に存在する核酸分子との関連において、天然に存在し得る。他方、核酸分子は、例えば非天然に存在する分子、組換えDNAもしくはRNA、人工染色体、操作されたゲノム、もしくはその断片、または合成DNA、RNA、DNA/RNAハイブリッド、または非天然に存在するヌクレオチドもしくはヌクレオシドを含む、非天然に存在する分子であり得る。さらに、用語「核酸」、「DNA」、「RNA」、および/または類似の用語は、核酸アナログ、例えばホスホジエステル骨格以外を有するアナログを含む。核酸は、天然源から精製される、組換え発現系を用いて産生され必要に応じて精製される、化学的に合成される、等が可能である。化学的に合成された分子の場合、核酸は、適切な場合には、例えば、化学的に修飾された塩基または糖、および骨格修飾を有するアナログなどのヌクレオシドアナログを含み得る。核酸配列は、特に示されない限り、5'~3'方向に示される。ある態様において、核酸は、天然ヌクレオシド(例えばアデノシン、チミジン、グアノシン、シチジン、ウリジン、デオキシアデノシン、デオキシチミジ

30

40

50

ン、デオキシグアノシン、およびデオキシシチジン)；ヌクレオシド類似体(例えば2-アミノアデノシン、2-チオチミジン、イノシン、ピロロピリミジン、3-メチルアデノシン、5-メチルシチジン、2-アミノアデノシン、C5-プロモウリジン、C5-フルオロウリジン、C5-ヨードウリジン、C5-プロピニル-ウリジン、C5-プロピニル-シチジン、C5-メチルシチジン、2-アミノアデノシン、7-デアザアデノシン、7-デアザグアノシン、8-オキソアデノシン、8-オキソグアニン、O⁶-メチルグアニン、および2-チオシチジン)；化学修飾塩基；生物学的に修飾された塩基(例えばメチル化塩基)；挿入塩基；修飾糖(例えば2'-フルオロリボース、リボース、2'-デオキシリボース、アラビノース、およびヘキソース)；および/または修飾リン酸基(例えばホスホロチオエートおよび5'-N-ホスホロアミダイト結合)であるか、またはそれらを含む。いくつかの実施形態では、RNAはCas9システムに関連するRNAである。例えば、RNAは、CRISPR RNA (crRNA)、トランスコード化小RNA (tracrRNA)、単一ガイドRNA (sgRNA)、またはガイドRNA (gRNA)であり得る。

【0095】

用語「核酸塩基」、「窒素塩基」、または「塩基」は、本明細書中で互換的に使用され、ヌクレオシドを形成する窒素含有生物学的化合物を指し、ヌクレオシドはヌクレオチドの成分である。塩基対を形成し、互いに積み重なる核酸塩基の能力は、直接、リボ核酸(RNA)およびデオキシリボ核酸(DNA)のような長鎖らせん構造をもたらす。アデニン(A)、シトシン(C)、グアニン(G)、チミン(T)、ウラシル(U)の5種類の核酸塩基は、一次塩基または標準塩基と呼ばれる。アデニンとグアニンはプリンに由来し、シトシン、ウラシル、チミンはピリミジンに由来する。DNAとRNAは修飾された他の(非一次)塩基を含むこともある。非限定的な例示的修飾核酸塩基としては、ヒポキサンチン、キサンチン、7-メチルグアニン、5,6-ジヒドロウラシル、5-メチルシトシン(m5C)、および5-ヒドロメチルシトシンが挙げられる。ヒポキサンチンとキサンチンは変異原の存在によって生成され得、どちらも脱アミノ化(アミン基のカルボニル基への置換)によって生成される。ヒポキサンチンはアデニンから修飾され得る。キサンチンはグアニンから修飾され得る。ウラシルはシトシンの脱アミノ化によって生じ得る。「ヌクレオシド」は核酸塩基と五炭糖(リボースまたはデオキシリボース)からなる。ヌクレオシドの例としては、アデノシン、グアノシン、ウリジン、シチジン、5-メチルウリジン(m5U)、デオキシアデノシン、デオキシグアノシン、チミジン、デオキシウリジン、およびデオキシシチジンが挙げられる。修飾核酸塩基を有するヌクレオシドとしては、イノシン(I)、キサントシン(X)、7-メチルグアノシン(m7G)、ジヒドロウリジン(D)、5-メチルシチジン(m5C)、プソイドウリジン()等が挙げられる。ヌクレオチドは、核酸塩基、五炭糖(リボースまたはデオキシリボース)、および少なくとも一つのリン酸基からなる。

【0096】

「核酸プログラム可能DNA結合タンパク質」または「napDNAbp」という用語は、「ポリヌクレオチドプログラム可能ヌクレオチド結合ドメイン」と交換可能に使用され、napDNAbpを特定の核酸配列にガイドする、ガイド核酸などの核酸(例えばDNAまたはRNA)に付随するタンパク質を意味する。例えば、Cas9タンパク質は、ガイドRNAに相補的な特定のDNA配列にCas9タンパク質をガイドするガイドRNAと結合することができる。ある態様において、napDNAbpは、Cas9ドメイン、例えばヌクレアーゼ活性Cas9、Cas9ニッカーゼ(ncas9)、またはヌクレアーゼ不活性Cas9(dCas9)である。核酸プログラム可能なDNA結合タンパク質の例としては、Cas9(例:dCas9およびncas9)、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h、およびCas12iが挙げられるが、これらに限定されない。他の核酸プログラム可能なDNA結合タンパク質もまた、本開示の範囲内であるが、それらは本開示に具体的に記載されていないこともあり得る。例えば、Makarova et al. "Classification and Nomenclature of CRISPR-Cas Systems: Where from Here?" CRISPR J. 2018 Oct;1:325-336. doi: 10.1089/crispr.2018.0033; Yan et al., "Functionally diverse type V CRISPR-Cas systems" Science. 2019 Jan 4;363(6422):88-91. doi: 10.1126/science.aav7271に記載されているとおりであり、それぞれの内容全体を参照により本明細書に組

10

20

30

40

50

み込む。

【0097】

「核酸塩基編集ドメイン」または「核酸塩基編集タンパク質」という用語は、本明細書において使用される場合、シトシン(もしくはシチジン)からウラシル(もしくはウリジン)またはチミン(もしくはチミジン)への、およびアデニン(もしくはアデノシン)からヒポキサンチン(もしくはイノシン)への脱アミノ化、ならびに非鋳型ヌクレオチド付加および挿入のような、RNAまたはDNAにおける核酸塩基修飾を触媒することができるタンパク質または酵素を指す。ある態様において、核酸塩基編集ドメインは、デアミナーゼドメイン(例えば、シチジンデアミナーゼ、シトシンデアミナーゼ、アデニンデアミナーゼ、またはアデノシンデアミナーゼ)である。ある態様において、核酸塩基編集ドメインは、天然に存在する核酸塩基編集ドメインであり得る。いくつかの実施形態において、核酸塩基編集ドメインは、天然に存在する核酸塩基編集ドメインから操作または進化された核酸塩基編集ドメインであり得る。核酸塩基編集ドメインは、細菌、ヒト、チンパンジー、ゴリラ、サル、ウシ、イヌ、ラット、またはマウスなどの任意の生物由来であり得る。例えば、核酸塩基編集タンパク質は、国際PCT出願番号PCT/2017/045381(国際公開第2018/027078号)およびPCT/US 2016/058344(国際公開第2017/070632号)に記載されており、これらの各々は、その全体が参照により本明細書に組み込まれている。Komor, A.C., et al., “Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage” Nature 533, 420-424 (2016); Gaudelli, N.M., et al., “Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage” Nature 551, 464-471 (2017); および Komor, A.C., et al., “Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity” Science Advances 3:eaa04774 (2017)も参照(その内容全体が参照により本明細書に組み込まれる)。

10

20

【0098】

本明細書で使用される場合、「薬剤を取得する」におけるような「取得する」は、その薬剤を合成すること、購入すること、または他の方法で獲得することを含む。

【0099】

本明細書で使用される「患者」または「対象」は、疾患または障害と診断されている、それらを生じるリスクがある、またはそれらを有するかもしれない疑いがある哺乳動物対象または個体を指す。ある態様において、用語「患者」は、疾患または障害を生ずる可能性が平均より高い哺乳動物対象を指す。例示的な患者は、ヒト、ヒト以外の霊長類、ネコ、イヌ、ブタ、ウシ、ネコ、ウマ、ラクダ、ラマ、ヤギ、ヒツジ、齧歯類(例えば、マウス、ウサギ、ラット、モルモット)および本明細書に開示される治療から利益を得ることができる他の哺乳動物であり得る。例示的なヒト患者は、男性および/または女性であり得る。

30

【0100】

「それを必要とする患者」または「それを必要とする対象」は、本明細書では、例えばアルファ-1アンチトリプシン欠損症(A1AD)であるがそれに限定されない、疾患または障害と診断された、またはその疾患または障害を有することが疑われる患者を表す。

40

【0101】

「病原性変異」、「病原性バリエーション」、「疾患原因変異」、「疾患原因バリエーション」、「有害変異」または「素因となる変異」という用語は、特定の疾患または障害に対する個体の感受性または素因を増大させる遺伝子変化または突然変異を指す。ある態様において、病原性変異は、遺伝子によってコードされるタンパク質中の少なくとも1つの野生型アミノ酸が少なくとも1つの病原性アミノ酸によって置換されたものを含む。

【0102】

用語「ペプチド」、「ポリペプチド」、「タンパク質」、およびそれらの文法的等価物は、本明細書において交換可能に使用され、ペプチド(アミド)結合によって互いに連結されたアミノ酸残基のポリマーを指す。この用語は、任意のサイズ、構造または機能のタン

50

パク質、ペプチドまたはポリペプチドを指す。典型的には、タンパク質、ペプチド、またはポリペプチドは、少なくとも3アミノ酸長である。タンパク質、ペプチド、またはポリペプチドは、個々のタンパク質または一群のタンパク質を指すことができる。タンパク質、ペプチド、またはポリペプチド中の1以上のアミノ酸は、例えば炭水化物基、ヒドロキシル基、リン酸基、ファルネシル基、イソファルネシル基、脂肪酸基、結合のためのリンカー、官能化、または他の修飾などの化学的実体の添加によって、修飾され得る。タンパク質、ペプチド、またはポリペプチドはまた、単一分子であってもよく、または多分子複合体であってもよい。タンパク質、ペプチド、またはポリペプチドは、天然に存在するタンパク質またはペプチドの単なる断片であり得る。タンパク質、ペプチド、またはポリペプチドは、天然に存在するもの、組換え体、もしくは合成のもの、またはそれらの任意の組み合わせであり得る。本明細書中で使用される場合、用語「融合タンパク質」とは、少なくとも二つの異なるタンパク質由来のタンパク質ドメインを含むハイブリッドポリペプチドをいう。1つのタンパク質は、融合タンパク質のアミノ末端(N末端)部分またはカルボキシ末端(C末端)タンパク質に位置することができ、かくして、それぞれ、アミノ末端融合タンパク質またはカルボキシ末端融合タンパク質を形成する。タンパク質は、異なるドメイン、例えば、核酸結合ドメイン(例えば、標的部位へのタンパク質の結合を誘導するCas9のgRNA結合ドメイン)および核酸切断ドメイン、あるいは核酸編集タンパク質の触媒ドメインを含み得る。ある態様において、タンパク質は、タンパク質性部分、例えば、核酸結合ドメインを構成するアミノ酸配列、および有機化合物、例えば、核酸切断剤として作用し得る化合物を含む。ある態様において、タンパク質は、核酸(例えば、RNAまたはDNA)と複合体化されているか、または核酸と会合している。本明細書で提供される任意のタンパク質は、当技術分野で公知の任意の方法によって生産することができる。例えば、本明細書で提供されるタンパク質は、組換えタンパク質発現および精製を介して生産することができ、これは、ペプチドリンカーを含む融合タンパク質に特に適している。組換えタンパク質の発現および精製のための方法はよく知られており、Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012))によって記載されているものを含み、その全内容は参照により本明細書に組み入れられる。

【0103】

本明細書に開示されるポリペプチドおよびタンパク質(機能的部分およびその機能的バリエーションを含む)は、一つ以上の天然に存在するアミノ酸の代わりに合成アミノ酸を含むことができる。そのような合成アミノ酸は当該分野で公知であり、例えば、アミノシクロヘキサンカルボン酸、ノルロイシン、 α -アミノ-n-デカン酸、ホモセリン、S-アセチルアミノメチル-システイン、trans-3-およびtrans-4-ヒドロキシプロリン、4-アミノフェニルアラニン、4-ニトロフェニルアラニン、4-クロロフェニルアラニン、4-カルボキシフェニルアラニン、 β -フェニルセリン、 β -ヒドロキシフェニルアラニン、フェニルグリシン、 β -ナフチルアラニン、シクロヘキシルアラニン、シクロヘキシルグリシン、インドリン-2-カルボン酸、1,2,3,4-テトラヒドロイソキノリン-3-カルボン酸、アミノマロン酸、アミノマロン酸モノアミド、N'-ベンジル-N'-メチルリジン、N',N'-ジベンジル-リジン、6-ヒドロキシリジン、オルニチン、 α -アミノシクロペンタンカルボン酸、アミノシクロヘキサンカルボン酸、 α -アミノシクロヘプタンカルボン酸、 β -(2-アミノ-2-ノルボルナン)-カルボン酸、 β -ジアミノ酪酸、 β -ジアミノプロピオン酸、ホモフェニルアラニン、および β -tert-ブチルグリシンが挙げられる。ポリペプチドおよびタンパク質は、ポリペプチド構築物の1つ以上のアミノ酸の翻訳後修飾と結合することができる。翻訳後修飾の非限定的な例としては、リン酸化、アセチル化およびホルミル化を含むアシル化、グリコシル化(N-リンクおよびO-リンクを含む)、アミド化、ヒドロキシル化、メチル化およびエチル化を含むアルキル化、ユビキチン化、ピロリドンカルボン酸の添加、ジスルフィド架橋の形成、硫酸化、ミリストイル化、パルミトイル化、イソプレニル化、ファルネシル化、ゲラニル化、グリピエーション、リポイル化およびヨード化が挙げられる。

10

20

30

40

50

【0104】

用語「ポリヌクレオチドプログラム可能ヌクレオチド結合ドメイン」とは、ポリヌクレオチドプログラム可能なDNA結合ドメインを特定の核酸配列にガイドするガイドポリヌクレオチド(例えばガイドRNA)のような核酸(DNAやRNAなど)に付随するタンパク質をいう。ある態様において、ポリヌクレオチドプログラム可能ヌクレオチド結合ドメインは、ポリヌクレオチドプログラム可能なDNA結合ドメインである。ある態様において、ポリヌクレオチドプログラム可能ヌクレオチド結合ドメインは、ポリヌクレオチドプログラム可能なRNA結合ドメインである。ある態様において、ポリヌクレオチドプログラム可能ヌクレオチド結合ドメインは、Cas9タンパク質である。Cas9タンパク質は、ガイドRNAに相補的な特定のDNA配列にCas9タンパク質をガイドするガイドRNAと結合することができる。ある態様において、ポリヌクレオチドプログラム可能ヌクレオチド結合ドメインは、Cas9ドメイン、例えばヌクレアーゼ活性Cas9、Cas9ニッカーゼ(nCas9)、またはヌクレアーゼ不活性Cas9(dCas9)である。核酸プログラム可能DNA結合タンパク質の非限定的な例は、Cas9(例:dCas9およびnCas9)、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h、およびCas12iを含む。Cas酵素の非限定的な例としては、Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas5d、Cas5t、Cas5h、Cas5a、Cas6、Cas7、Cas8、Cas8a、Cas8b、Cas8c、Cas9(Csn1またはCsn2とも呼ばれる)、Cas10、Cas10d、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h、Cas12i、Csy1、Csy2、Csy3、Csy4、Cse1、Csc1、Csc2、Csm2、Csm3、Csm4、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx1S、Csx11、Csf1、Csf2、CsO、Csf4、Csd1、Csd2、Cst1、Cst2、Csh1、Csh2、Csa1、Csa2、Csa3、Csa4、Csa5、タイプII Casエフェクタータンパク質、タイプV Casエフェクタータンパク質、タイプVI Casエフェクタータンパク質、CARF、DinG、それらのホモログ、またはそれらの修飾もしくは改変バージョンが挙げられる。他の核酸プログラム可能DNA結合タンパク質もまた、本開示の範囲内であるが、本開示に具体的に記載されていないこともあり得る。

10

20

【0105】

タンパク質または核酸に関連して本明細書中で使用される用語「組み換え体」とは、自然界には存在しないが、人間の工学の産物であるタンパク質または核酸を指す。例えば、いくつかの実施形態において、組換えタンパク質または核酸分子は、任意の天然に存在する配列と比較して、少なくとも1つ、少なくとも2つ、少なくとも3つ、少なくとも4つ、少なくとも5つ、少なくとも6つ、または少なくとも7つの変異を含むアミノ酸またはヌクレオチド配列を含む。

30

【0106】

「減少する」とは、少なくとも10%、25%、50%、75%、または100%の負の変化を意味する。

【0107】

「参照」は、標準又は対照条件を意味する。1つの実施形態において、参照は野生型または健常な細胞である。

40

【0108】

「参照配列」は、配列比較の基礎として使用される定義済み配列である。参照配列は、特定の配列のサブセット又は全体であり得る；例えば、完全長のcDNAもしくは遺伝子配列のセグメント、または完全なcDNAまたは遺伝子配列。ポリペプチドについては、参照ポリペプチド配列の長さは、一般に、少なくとも約16アミノ酸、好ましくは少なくとも約20アミノ酸、より好ましくは少なくとも約25アミノ酸、さらにより好ましくは約35アミノ酸、約50アミノ酸、または約100アミノ酸である。核酸については、参照核酸配列の長さは、一般に、少なくとも約50ヌクレオチド、好ましくは少なくとも約60ヌクレオチド、より好ましくは少なくとも約75ヌクレオチド、さらにより好ましくは約100ヌクレオチドまたは約300ヌクレオチドまたはそれらの周辺もしくはそれらの間の任意の整数である

50

【0109】

用語「RNAプログラム可能なヌクレアーゼ」および「RNA誘導ヌクレアーゼ」は、切断の標的ではない一つ以上のRNAをととも使用される(例えば、それに結合または付随する)。ある態様において、RNAプログラム可能ヌクレアーゼは、RNAと複合体である場合、ヌクレアーゼ:RNA複合体と称され得る。典型的には、結合したRNAはガイドRNA (gRNA) と呼ばれる。gRNAは2つ以上のRNAの複合体として存在することもあれば、1つのRNA分子として存在することもある。単一のRNA分子として存在するgRNAは、単一ガイドRNA (sgRNA) と呼ばれることがあるが、「gRNA」は、単一の分子として、または二以上の分子の複合体として存在するガイドRNAのいずれかを指すために互換的に使用される。典型的には、単一のRNA種として存在するgRNAは、二つのドメイン: (1) 標的核酸と相同性を共有する(例えば、標的へのCas9複合体の結合を誘導する)ドメイン; および (2) Cas9タンパク質に結合するドメインを含む。ある態様において、ドメイン (2) は、tracrRNAとして知られる配列に対応し、ステム-ループ構造を含む。例えば、いくつかの実施形態において、ドメイン (2) は、Jinek et al., *Science* 337:816-821(2012)に提供されるtracrRNAと同一または相同であり、その全内容を参照により本明細書に援用する。gRNA (例えばドメイン2を含むもの)の他の例は、2013年9月6日に出願された「Switchable Cas9 Nucleases And Uses Thereof」という名称の米国仮特許出願第61/874,682号および2013年9月6日に出願された「Delivery System For Functional Nucleases」という名称の米国仮特許出願第61/874,746号に見出され得る。いくつかの実施形態において、gRNAは、二つ以上のドメイン (1) および (2) を含み、「伸長された (extended) gRNA」と称され得る。例えば、伸長されたgRNAは、本明細書に記載されるように二つ以上のCas9タンパク質と結合し、二つ以上の異なる領域で標的核酸と結合する。gRNAは、ヌクレアーゼ/RNA複合体の標的部位への結合を媒介し、ヌクレアーゼ:RNA複合体の配列特異性を提供する、標的部位を相補するヌクレオチド配列を含む。ある態様において、RNAプログラム可能ヌクレアーゼは、(CRISPR関連システム) Cas9エンドヌクレアーゼ、例えば、*Streptococcus pyogenes*由来のCas9 (CsnI) である(例えば、"Complete genome sequence of an M1 strain of *Streptococcus pyogenes*." Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C, Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., *Proc. Natl. Acad. Sci. U.S.A.* 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma CM., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., *Nature* 471:602-607(2011)参照)。

【0110】

「SERPINA1ポリヌクレオチド」とは、A1ATタンパク質またはその断片をコードする核酸分子を意味する。NCBI Accession NO. NM_000295において入手可能な例示的SERPINA1ポリヌクレオチドの配列を以下に提供する：

```

1  acaatgactc ctttcggtaa gtgcagtgga agctgtacac tgcccaggca aagcgtccgg
61  gcagcgtagg cgggcgactc agatcccagc cagtggactt agcccctggt tgctcctccg
121 ataactgggg tgaccttggg taatattcac cagcagcctc ccccggtgcc cctctggatc
181 cactgcttaa atacggacga ggacagggcc ctgtctcctc agcttcaggc accaccactg
241 acctgggaca gtgaatcgac aatgccgtct tctgtctcgt ggggcatact cctgctggca
301 ggctgtgct gcctgggtccc tgtctccctg gctgaggatc cccagggaga tgctgcccag
361 aagacagata catcccacca tgatcaggat cacccaacct tcaacaagat ccccccaac
421 ctggctgagt tcgccttcag cctataccgc cagctggcac accagtccaa cagcaccaat

```

10

20

30

40

50

481 atctttcttct cccagtgag catcgctaca gcctttgcaa tgctctccct ggggaccaag
 541 gctgacactc acgatgaaat cctggagggc ctgaatttca acctcacgga gattccggag
 601 gctcagatcc atgaaggctt ccaggaactc ctccgtaccc tcaaccagcc agacagccag
 661 ctccagctga ccaccggcaa tggcctgttc ctcagcgagg gcctgaagct agtgataag
 721 tttttggagg atgttaaaaa gttgtaccac tcagaagcct tcaactgtcaa ctccggggac
 781 accgaagagg ccaagaaaca gatcaacgat tacgtggaga agggactca agggaaaatt
 841 gtggatttgg tcaaggagct tgacagagac acagtttttg ctctggtgaa ttacatcttc
 901 tttaaaggca aatgggagag accctttgaa gtcaaggaca ccgaggaaga ggacttccac
 961 gtggaccagg tgaccaccgt gaaggtgcct atgatgaagc gtttaggcat gtttaacatc
 1021 cagcactgta agaagctgtc cagctgggtg ctgctgatga aatacctggg caatgccacc
 1081 gccatcttct tctgcctga tgaggggaaa ctacagcacc tggaaaatga actcaccacc
 1141 gatatcatca ccaagttcct ggaaaatgaa gacagaaggt ctgccagctt acatttacc
 1201 aactgtcca ttactggaac ctatgatctg aagagcgtcc tgggtcaact gggcatcact
 1261 aaggtcttca gcaatggggc tgacctctcc ggggtcacag aggaggcacc cctgaagctc
 1321 tccaagggcg tgcataaggc tgtgctgacc atcgacgaga aagggactga agctgctggg
 1381 gccatgtttt tagagggcct acccatgtct atcccccccg aggtcaagtt caacaaaccc
 1441 tttgtcttct taatgattga acaaaatacc aagtctcccc tcttcatggg aaaagtggg
 1501 aatcccacc ccaaaataact gcctctcgtc cctcaacccc tccccctcat ccctggcccc
 1561 ctccctggat gacattaaag aagggttgag ctggtccctg cctgcatgtg actgtaaatc
 1621 cctcccatgt tttctctgag tctccctttg cctgctgagg ctgtatgtgg gctccaggta
 1681 acagtgtctg ctteggggcc cctgaactgt gttcatggag catctggctg ggtaggcaca
 1741 tgctgggctt gaatccaggg gggactgaat cctcagctta cggacctggg cccatctgtt
 1801 tctggagggc tccagtcttc cttgtcctgt cttggagtcc ccaagaagga atcacagggg
 1861 aggaaccaga taccagccat gaccccagge tccaccaagc atcttcatgt cccctgctc
 1921 atccccact cccccacc cagagttgct catcctgcca gggctggctg tgccccccc
 1981 aaggctgccc tccctggggc ccagaactg cctgatcgtg ccgtggccc gttttgtggc
 2041 atctgcagca acacaagaga gaggacaatg tctcctctt gacctcgtg cacctaacca
 2101 gactcgggcc ctgcacctc caggcacttc tggaaaatga ctgaggcaga ttcttctga
 2161 agcccattct ccattgggca acaaggacac ctattctgtc cttgtccttc catcgtgcc
 2221 ccagaagcc tcacatatct ccgttagaa tcaggtccct tctccccaga tgaagaggag
 2281 ggtctctgct ttgttttctc tatctctctc tcagacttga ccaggcccag caggccccag
 2341 aagaccatta ccctatatcc cttctctctc ctagtccat ggccataggc ctgctgatgg
 2401 ctcaggaag ccattgcaag gactcctcag ctatgggaga ggaagcacat caccattga
 2461 cccccgcaac ccctcccttt cctcctctga gtcccactg gggccacatg cagcctgact
 2521 tctttgtgcc tgttctgtc cctgcagtct tcagagggcc accgcagctc cagtgccacg
 2581 gcaggaggct gttcttgaat agcccctgtg gtaagggcca ggagagctct tccatcctc
 2641 aaggccctgc taaaggacac agcagccagg aagtccctg ggcccctagc tgaaggacag
 2701 cctgctocct ccgtctctac caggaatggc ctgtctctat ggaaggcact gcccacccc
 2761 aactaatct aggaatcact gtctaaccac tcaactgtcat gaatgtgtac ttaaaggatg
 2821 aggttgagtc ataccaaata gtgatttoga tagttcaaaa tgggtgaaatt agcaattcta
 2881 catgattcag tctaatacat ggataccgac tgtttccac acaagtctcc tgttctctta
 2941 agcttactca ctgacagcct tcaactctcc acaatacat taaagatatg gccatcacca

3001 agccccctag gatgacacca gacctgagag tctgaagacc tggatccaag ttctgacttt
 3061 tccccctgac agctgtgtga ccttcgtgaa gtcgccaac ctctctgagc cccagtcatt
 3121 gctagtaaga cctgcctttg agttggtatg atgttcaagt tagataacaa aatgtttata
 3181 cccattagaa cagagaataa atagaactac atttcttgca

PAM配列に対して相補的な塩基の位置を斜体、二重下線で示している。位置1455におけるGは、位置1455の標的Cに対して相補的なものであるが、太字、下線で示している。

【 0 1 1 1 】

用語「一塩基多型 (SNP)」は、ゲノム中の特定の位置で起こる単一ヌクレオチドの変異であり、ここで、各変異は集団内で認識できるある程度(例 1%)まで存在する。たとえば、ヒトゲノムの特定の塩基位置では、ほとんどの個体でCヌクレオチドが出現しうが、少数の個体ではその位置がAで占められている。これは、この特定の位置にSNPがあり、CまたはAという2つのヌクレオチドのバリエーションがこの位置のアリルであることを意味する。SNPは疾患に対する感受性の差；広範囲にわたるヒト疾患の根底にある。病気の重症度や治療に対する体の反応も、遺伝的バリエーションの表れである。SNPは、遺伝子のコード領域、遺伝子の非コード領域、または遺伝子間領域(遺伝子と遺伝子の間の領域)に存在しうる。ある態様において、コード配列内のSNPは、遺伝子コードの縮重のために、産生されるタンパク質のアミノ酸配列を必ずしも変化させない。コード領域のSNPには、同義SNPと非同義SNPの2種類がある。同義SNPはタンパク質配列に影響しないが、非同義SNPはタンパク質のアミノ酸配列を変化させる。非同義SNPにはミスセンスとナンセンスの2種類がある。タンパク質をコードする領域にないSNPは、遺伝子のスプライシング、転写因子の結合、メッセンジャーRNAの分解、または非コードRNAの配列に影響を与えることがある。この種のSNPによって影響を受ける遺伝子発現はeSNP(発現SNP)と呼ばれ、遺伝子の上流または下流にあり得る。一塩基バリエーション(SNV)は、頻度に制限のない一塩基のバリエーションであり、体細胞で生じ得る。体細胞一塩基バリエーション(例えば癌により引き起こされるもの)は一塩基改変とも呼ばれ得る。

【0112】

「特異的に結合する」とは、本発明のポリペプチドおよび/または核酸分子を認識および結合するが、試料(例えば生物学的試料)中の他の分子を実質的に認識および結合しない核酸分子、ポリペプチド、もしくはそれらの複合体(例えば、核酸プログラム可能なDNA結合ドメインおよびガイド核酸)、化合物、または分子を意味する。

【0113】

本発明の方法において有用な核酸分子は、本発明のポリペプチドまたはその断片をコードする任意の核酸分子を含む。そのような核酸分子は、内因性核酸配列と100%同一である必要はないが、典型的には実質的同一性を示す。内因性配列に対して「実質的同一性」を有するポリヌクレオチドは、典型的には、二本鎖核酸分子の少なくとも一つの鎖とハイブリダイズすることができる。本発明の方法において有用な核酸分子は、本発明のポリペプチドまたはその断片をコードする任意の核酸分子を含む。そのような核酸分子は、内因性核酸配列と100%同一である必要はないが、典型的には実質的同一性を示す。内因性配列に対して「実質的同一性」を有するポリヌクレオチドは、典型的には、二本鎖核酸分子の少なくとも一つの鎖とハイブリダイズすることができる。「ハイブリダイズする」とは、種々のストリンジェンシー条件下で相補的ポリヌクレオチド配列(例えば、本明細書に記載の遺伝子)の間、またはその一部の間二本鎖分子を形成する対をなすことを意味する。(例えばWahl, G. M. and S. L. Berger (1987) *Methods Enzymol.* 152:399; Kimmel, A. R. (1987) *Methods Enzymol.* 152:507参照)。

【0114】

例えば、ストリンジェントな塩濃度は、通常、約750 mM未満 NaClおよび75 mMクエン酸三ナトリウム、好ましくは約500 mM未満 NaClおよび50 mMクエン酸三ナトリウム、より好ましくは約250 mM未満 NaClおよび25 mMクエン酸三ナトリウムである。低ストリンジェンシーハイブリダイゼーションは、有機溶媒、例えばホルムアミドの非存在下で得ることができ、一方、高ストリンジェンシーハイブリダイゼーションは、少なくとも約35%ホルムアミド、より好ましくは少なくとも約50%ホルムアミドの存在下で得ることができる。ストリンジェントな温度条件は、通常、少なくとも約30℃、より好ましくは少なくとも約37℃、最も好ましくは少なくとも約42℃の温度を含むであろう。ハイブリダイゼーション時間、界面活性剤(例えば、ドデシル硫酸ナトリウム(SDS))の濃度、および担体DNAの含入または排除などの様々な追加のパラメーターは、当業者によく知られている。必要に応じてこれら様々な条件を組み合わせることによって、様々なレベルのストリンジェンシーが達成される。好ましい実施形態では、ハイブリダイゼーションは、30

で、750 mM NaCl、75 mMクエン酸三ナトリウムおよび1% SDS中で起こる。より好ましい実施形態では、ハイブリダイゼーションは、37 で、500 mM NaCl、50 mMクエン酸三ナトリウム、1% SDS、35%ホルムアミドおよび100 µg/ml変性サケ精子DNA (ssDNA) 中で起こる。最も好ましい実施形態において、ハイブリダイゼーションは、42 において、250 mM NaCl、25 mMクエン酸三ナトリウム、1% SDS、50%ホルムアミドおよび200 µg/ml ssDNA中で起こる。これらの条件の有用なバリエーションは、当業者には容易に明らかになるであろう。

【0115】

ほとんどの用途では、ハイブリダイゼーションに続く洗浄工程もまた、ストリンジェンシーが異なる。洗浄ストリンジェンシー条件は、塩濃度および温度によって定義することができる。上記のように、洗浄ストリンジェンシーは、塩濃度を低下させるか、または温度を上昇させることによって増加させることができる。例えば、洗浄工程のためのストリンジェントな塩濃度は、好ましくは約30 mM未満NaClおよび3 mMクエン酸三ナトリウムであり、最も好ましくは約15 mM未満 NaClおよび1.5 mMクエン酸三ナトリウムである。洗浄工程のためのストリンジェントな温度条件は、通常、少なくとも約25 、より好ましくは少なくとも約42 、さらにより好ましくは少なくとも約68 の温度を含む。好ましい実施形態において、洗浄工程は、25 で、30 mM NaCl、3 mMクエン酸三ナトリウム、および0.1% SDS中で行われる。より好ましい実施形態において、洗浄工程は、42

で、15 mM NaCl、1.5 mMクエン酸三ナトリウム、および0.1% SDS中で行われる。より好ましい実施形態において、洗浄工程は、68 で、15 mM NaCl、1.5 mMクエン酸三ナトリウム、および0.1% SDS中で行われる。これらの条件のさらなるバリエーションは、当業者には容易に明らかになるであろう。ハイブリダイゼーション技術は、当業者によく知られており、例えば、Benton and Davis (*Science* 196:180, 1977); Grunstein and Hogness (*Proc. Natl. Acad. Sci., USA* 72:3961, 1975); Ausubel et al. (*Current Protocols in Molecular Biology*, Wiley Interscience, New York, 2001); Berger and Kimmel (*Guide to Molecular Cloning Techniques*, 1987, Academic Press, New York); および Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New Yorkに記述されている。

【0116】

「対象」とは、限定されるものではないが、ウシ、ウマ、イヌ、ヒツジまたはネコなどのヒトまたは非ヒト哺乳動物を含む哺乳動物を意味する。

【0117】

「実質的に同一である」とは、参照アミノ酸配列(例えば、本明細書に記載のアミノ酸配列のいずれか1つ)または核酸配列(例えば、本明細書に記載の核酸配列のいずれか1つ)に対して少なくとも50%の同一性を示すポリペプチドまたは核酸分子を意味する。好ましくは、そのような配列は、比較のために使用される配列とアミノ酸レベルまたは核酸において少なくとも60%、より好ましくは80%または85%、より好ましくは90%、95%または99%までもの同一性を有する。

【0118】

配列同一性は、典型的には、配列解析ソフトウェア(例えば、Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705のSequence Analysis Software Package、BLAST、BESTFIT、GAP、またはPILEUP/PRETTYBOXプログラム)を用いて測定される。そのようなソフトウェアは、種々の置換、欠失、および/または他の改変に相同性の程度を割り当てることによって、同一または類似の配列をマッチさせる。保存的置換は、典型的には、以下の群内の置換を含む:グリシン、アラニン;バリン、イソロイシン、ロイシン;アスパラギン酸、グルタミン酸、アスパラギン、グルタミン;セリン、トレオニン;リジン、アルギニン;フェニルアラニン、チロシン。同一性の程度を決定するための例示的なアプローチにおいて、BLASTプログラムを使用することができ、 e^{-3} と e^{-100} の間の確率スコアが密接に関連した配列を示す。

10

20

30

40

50

【0119】

用語「標的部位」とは、核酸分子内の配列であって、核酸塩基エディターによって改変される配列をいう。一実施形態において、標的部位は、デアミナーゼまたはそれを含む融合タンパク質(例えば、シチジンまたはアデニンデアミナーゼ)によって脱アミノ化される。

【0120】

RNAプログラム可能なヌクレアーゼ(例:Cas9)は、DNA切断部位を標的とするためにRNA:A:DNAハイブリダイゼーションを使用するので、これらのタンパク質は、原理的に、ガイドRNAによって指定されるあらゆる配列を標的とすることができる。部位特異的切断のためにCas9のようなRNAプログラム可能なヌクレアーゼを使用する方法(例えばゲノムを改変するために)は、当該技術分野において公知である(例えば、Cong, L. et al., Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819-823 (2013); Mali, P. et al., RNA-guided human genome engineering via Cas9. *Science* 339, 823-826 (2013); Hwang, W.Y. et al., Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature biotechnology* 31, 227-229 (2013); Jinek, M. et al., RNA-programmed genome editing in human cells. *eLife* 2, e00471 (2013); Dicarlo, J.E. et al., Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic acids research* (2013); Jiang, W. et al., RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature biotechnology* 31, 233-239 (2013)を参照。これらの各々の全内容は参照により本明細書に組み入れられる)。

【0121】

本明細書中で使用される、用語「治療(treatment)」、「治療する(treating)」、またはその文法的等価物は、所望の薬理学的および/または生理学的効果を得ることを指す。いくつかの態様において、効果は治療的であり、すなわち、効果は、疾患および/またはそれに起因する有害症状を部分的または完全に治癒する。ある態様において、効果は予防的であり、すなわち効果は、疾患または状態の発生または再発を予防する。この目的のために、本開示の方法は、本明細書に記載されるような治療的に有効な量の組成物を投与することを含む。

【0122】

「ウラシルグリコシラーゼ阻害因子」とは、ウラシル除去修復系を阻害する因子を意味する。1つの実施形態において、該因子は、宿主ウラシル-DNAグリコシラーゼに結合してDNAからのウラシル残基の除去を妨げるタンパク質またはその断片である。

【0123】

本明細書で提供される範囲は、範囲内のすべての値についての省略形であると理解される。例えば、1~50の範囲は、1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, または50からなる群からの任意の数、数の組み合わせ、またはサブ範囲を含むと理解される。

【0124】

本明細書における変数の任意の定義における化学基のリストの記載は、リストされた基の任意の単一の基または組合せとしてのその変数の定義を含む。本明細書における変数または態様に関する実施形態の説明は、任意の単一の実施形態としての、または任意の他の実施形態もしくはその一部と組み合わせられた、その実施形態を含む。

【0125】

本明細書に提供される任意の組成物または方法は、本明細書に提供される他の任意の組成物および方法の1つ以上と組み合わせることができる。

【0126】

DNA編集は、遺伝子レベルで病原性変異を補正することにより疾患状態を改変する実行可能な手段として出現してきた。最近まで、全てのDNA編集プラットフォームは、特定のゲノム部位でDNA二本鎖切断(DSB)を誘導することにより機能し、内因性DNA修復経路に

10

20

30

40

50

依存して半確率的な態様で産物結果が決定され、遺伝的産物の複雑な集団を生じていた。相同性誘導修復 (HDR) 経路を介すれば、ユーザーにより定義される正確な修復結果を達成できるが、多くの課題が、治療に関連する細胞型でのHDRを用いた高効率修復を妨げている。実際的にはこの経路は、競合する、誤りを起こしやすい非相同末端結合経路と比較して効率が悪い。さらに、HDRは細胞周期のG1期とS期に厳密に限定されており、有糸分裂後の細胞におけるDSBの正確な修復を妨げる。その結果、これらの集団において高い効率で、ユーザー定義のプログラム可能な態様でゲノム配列を改変することは困難または不可能であることと判明してきた。

【 0 1 2 7 】

[核酸塩基エディター]

ポリヌクレオチドの標的ヌクレオチド配列を編集、修飾または改変するための塩基エディターまたは核酸塩基エディターが本明細書に開示される。本明細書に記載されるのは、ポリヌクレオチドプログラム可能なヌクレオチド結合ドメインおよび核酸塩基編集ドメインを含む核酸塩基エディターまたは塩基エディターである。ポリヌクレオチドプログラム可能なヌクレオチド結合ドメインは、結合されたガイドポリヌクレオチド(例えばgRNA)と一緒にある場合に、(結合されたガイド核酸の塩基と標的ポリヌクレオチド配列の塩基との間の相補的塩基対形成を介して)標的ポリヌクレオチド配列に特異的に結合することができ、それによって、編集されることが所望される標的核酸配列に塩基エディターを局在化させることができる。或る実施態様では、標的ポリヌクレオチド配列は一本鎖DNAまたは二本鎖DNAを含む。或る実施態様では、標的ポリヌクレオチド配列はRNAを含む。或る実施態様では、標的ポリヌクレオチド配列はDNA-RNAハイブリッドを含む。

【 0 1 2 8 】

[ポリヌクレオチドプログラム可能なヌクレオチド結合ドメイン]

用語「ポリヌクレオチドプログラム可能なヌクレオチド結合ドメイン」とは、ガイドポリヌクレオチド(例えばガイドRNA)のような核酸(例えばDNAまたはRNA)と会合するタンパク質であって、その核酸がポリヌクレオチドプログラム可能なヌクレオチド結合ドメインを特定の核酸配列にガイドするところのものをいう。ある態様において、ポリヌクレオチドプログラム可能なヌクレオチド結合ドメインは、ポリヌクレオチドプログラム可能なDNA結合ドメインである。ある態様において、ポリヌクレオチドプログラム可能なヌクレオチド結合ドメインは、ポリヌクレオチドプログラム可能なRNA結合ドメインである。ある態様において、ポリヌクレオチドプログラム可能なヌクレオチド結合ドメインは、Cas9タンパク質である。ある態様において、ポリヌクレオチドプログラム可能なヌクレオチド結合ドメインは、Cpf1タンパク質である。

【 0 1 2 9 】

CRISPRは、可動遺伝要素(ウイルス、転移因子、接合プラスミド)に対する防御を提供する適応免疫系である。CRISPRクラスターは、スペーサー、先行する可動要素に相補的な配列、および標的侵入核酸を含む。CRISPRクラスターは転写され、CRISPR RNA (crRNA) にプロセシングされる。II型CRISPRシステムでは、pre-crRNAの正しいプロセシングはトランスコード小RNA (tracrRNA)、内因性リボヌクレアーゼ3 (rnc)、およびCas9タンパク質を必要とする。tracrRNAはリボヌクレアーゼ3によるpre-crRNAのプロセシングのガイドとなる。続いて、Cas9/crRNA/tracrRNAが、スペーサーに相補的な線状または環状のdsDNA標的をエンドヌクレアーゼで切断する。crRNAに相補的でない標的鎖は、最初にエンドヌクレアーゼ的に切断され、次にエキソヌクレアーゼ的に3' -5' にトリムされる。自然界では、DNA結合と切断にはタンパク質と両方のRNAが必要である。しかしながら、crRNAおよびtracrRNAの両方の側面を単一のRNA種に組み込むように、単一ガイドRNA(「sgRNA」、あるいは単に「gRNA」)を作製することができる。例えばJinek M., Chylinski K., Fonfara I., Hauer M., Doudna J. A., Charpentier E. Science 337:816-821(2012)を参照されたい(その内容全体が参照により本明細書に組み入れられる)。Cas9は、CRISPR反復配列中の短いモチーフ(PAMまたはプロトスペーサー隣接モチーフ)を認識して、自己と非自己を区別することを助ける。

10

20

30

40

50

【 0 1 3 0 】

Cas9ヌクレアーゼの配列および構造は、当業者によく知られている(例えば“Complete genome sequence of an M1 strain of *Streptococcus pyogenes*.” Ferretti et al., J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C, Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); “CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.” Deltcheva E., Chylinski K., Sharma CM., Gonzales K., Chao Y., Pirzada Z.A., Eckert M. R., Vogel J., Charpentier E., Nature 471:602-607(2011); および “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.” Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. Science 337:816-821(2012)参照。その内容全体が参照により本明細書に組み入れられる。)。Cas9オソログは、限定されるものではないが、*S.pyogenes*および*S.thermophilus*を含む種々の種において記述されてきた。さらなる適切なCas9ヌクレアーゼおよび配列は、本開示に基づいて当業者に明らかとなり得、そのようなCas9ヌクレアーゼおよび配列は、Chylinski, Rhun, and Charpentier, “The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems” (2013) RNA Biology 10:5, 726-737に開示されている生物および遺伝子由来のCas9配列を含む。その全内容は参照により本明細書に組み込まれる。

【 0 1 3 1 】

ある態様において、核酸プログラム可能なDNA結合タンパク質(napDNAbp)は、Cas9ドメインである。非限定的な例示的Cas9ドメインが本明細書に提供されている。Cas9ドメインは、ヌクレアーゼ活性Cas9ドメイン、ヌクレアーゼ不活性Cas9ドメイン、またはCas9ニッカーゼであり得る。いくつかの実施形態において、Cas9ドメインは、ヌクレアーゼ活性ドメインである。例えば、Cas9ドメインは、二本鎖核酸の両方の鎖(すなわち、二本鎖DNA分子の両方の鎖)を切断するCas9ドメインであり得る。いくつかの実施形態において、Cas9ドメインは、本明細書に記載のアミノ酸配列のいずれか1つを含む。いくつかの実施形態において、Cas9ドメインは、本明細書に記載されたアミノ酸配列のいずれかに対して少なくとも60%、少なくとも65%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、または少なくとも99.5%の同一性を有するアミノ酸配列を含む。いくつかの実施形態において、Cas9ドメインは、本明細書に記載されるアミノ酸配列のいずれかと比較して、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50またはそれ以上の変異を有するアミノ酸配列を含む。いくつかの実施形態において、Cas9ドメインは、本明細書に記載されたアミノ酸配列のいずれかと比較して、少なくとも10、少なくとも15、少なくとも20、少なくとも30、少なくとも40、少なくとも50、少なくとも60、少なくとも70、少なくとも80、少なくとも90、少なくとも100、少なくとも150、少なくとも200、少なくとも250、少なくとも300、少なくとも350、少なくとも400、少なくとも500、少なくとも600、少なくとも700、少なくとも800、少なくとも900、少なくとも1000、少なくとも1100、または少なくとも1200の同一の連続したアミノ酸残基を有するアミノ酸配列を含む。

【 0 1 3 2 】

ある態様において、Cas9ヌクレアーゼは、不活性な(例えば不活化された)DNA切断ドメインを有し、すなわち、Cas9はニッカーゼである。ヌクレアーゼ不活性化Cas9タンパク質は、互換的に「dCas9」タンパク質(nuclease-dead Cas9の意)とも称され得る。不活性なDNA切断ドメインを有するCas9タンパク質(又はその断片)を生成する方法は公知である(例えばJinek et al, Science. 337:816-821(2012); Qi et al, “Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Exp

ression” (2013) Cell. 28; 152(5): 1173-83参照 (各内容は参照により本明細書に組み込まれる))。例えば、Cas9のDNA切断ドメインは、HNHヌクレアーゼサブドメインとRuvC1サブドメインという2つのサブドメインを含むことが知られている。HNHサブドメインはgRNAに相補的な鎖を切断し、RuvC1サブドメインは非相補的な鎖を切断する。これらのサブドメイン内の変異はCas9のヌクレアーゼ活性を抑制し得る。例えば、突然変異D10AおよびH840Aは、*S. pyogenes* Cas9 のヌクレアーゼ活性を完全に不活性化する (Jinek et al, Science. 337:816-821(2012); Qi et al, Cell. 28;152(5): 1173-83 (2013))。ある実施形態において、Cas9ヌクレアーゼは、不活性な (例えば不活性化された) DNA切断ドメインを有し、すなわち、Cas9は「nCas9」タンパク質 (“nickase” Cas9の意) と呼ばれるニックアーゼである。ある態様において、Cas9の断片を含むタンパク質が提供される。例えば、いくつかの実施形態において、タンパク質は、以下の2つのCas9ドメインのうちの1つを含む: (1) Cas9のgRNA結合ドメイン;(2) Cas9のDNA切断ドメイン。ある態様において、Cas9またはその断片を含むタンパク質は、「Cas9パリアント」と称される。Cas9パリアントは、Cas9またはその断片と相同性を共有する。例えば、Cas9パリアントは、野生型Cas9と少なくとも約70%同一、少なくとも約80%同一、少なくとも約90%同一、少なくとも約95%同一、少なくとも約96%同一、少なくとも約97%同一、少なくとも約98%同一、少なくとも約99%同一、少なくとも約99.5%同一、または少なくとも約99.9%同一である。いくつかの実施形態において、Cas9変異体は、野生型Cas9と比較して、1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50またはそれ以上のアミノ酸変化を有し得る。いくつかの実施形態において、Cas9パリアントは、Cas9の断片 (例えばgRNA結合ドメインまたはDNA切断ドメイン)を含み、その断片は、野生型Cas9の対応する断片と少なくとも約70%同一であり、少なくとも約80%同一であり、少なくとも約90%同一であり、少なくとも約95%同一であり、少なくとも約96%同一であり、少なくとも約97%同一であり、少なくとも約98%同一であり、少なくとも約99%同一であり、少なくとも約99.5%同一であり、または少なくとも約99.9%同一である。ある態様において、断片は、対応する野生型Cas9のアミノ酸長の少なくとも30%、少なくとも35%、少なくとも40%、少なくとも45%、少なくとも50%、少なくとも55%、少なくとも60%、少なくとも65%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも95%同一、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、または少なくとも99.5%である。ある態様において、断片は、長さが少なくとも100アミノ酸である。ある態様において、断片は、少なくとも100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, または少なくとも1300アミノ酸の長さである。

【0133】

ある態様において、野生型Cas9は、*Streptococcus pyogenes*からのCas9に対応する (NCBI参照配列:NC_17053.1、以下のヌクレオチド及びアミノ酸配列):

【0134】

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGT
 GATCACTGATGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACC
 GCCACAGTATCAAAAAAAAAATCTTATAGGGGCTCTTTTATTTGGCAGTGGAGAGACAGCG
 GAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTAT
 TTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTTC
 ATCGACTTGAAGAGTCTTTTTTGGTGGAAAGAAGACAAGAAGCATGAACGTCATCCTATT
 TTTGGAAATATAGTAGATGAAGTTGCTTATCATGAGAAATATCCAACCTATCTATCATCT
 GCGAAAAAATTGGCAGATTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCT
 TAGCGCATATGATTAAGTTTCGTGGTCATTTTTTTGATTGAGGGAGATTTAAATCCTGAT
 AATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAATCTACAATCAATTATTTGA
 AGAAAACCCTATTAACGCAAGTAGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGA

10

20

30

40

50

GTAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAGAAATGGC
TTGTTTGGGAATCTCATTGCTTTGTCATTGGGATTGACCCCTAATTTTAAATCAAATTTT
GATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAGATACTTACGATGATGATTTAGA
TAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATT
TATCAGATGCTATTTTACTTTTACAGATATCCTAAGAGTAAATAGTGAAATAACTAAGGCT
CCCCTATCAGCTTCAATGATTAAGCGCTACGATGAACATCATCAAGACTTGACTCTTTT
AAAAGCTTTAGTTTCGACAACAACCTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAAT
CAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAA
TTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACTAAA
TCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCAA
TTCACCTGGGTGAGCTGCATGCTATTTTGAAGAACAAGAAGACTTTTATCCATTTTAA
AAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTGAATTCCTTATTATGTTGG
TCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAA
TTACCCCATGGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATT
GAACGCATGACAAACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAACATAG
TTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTG
AGGGAATGCGAAAACAGCATTTCCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTA
CTCTTCAAACAAATCGAAAAGTAAACCGTTAAGCAATTAAGAAGATTATTTCAAAAA
AATAGAATGTTTTGATAGTGTTGAAATTTTACGGAGTTGAAGATAGATTTAATGCTTCAT
TAGGCGCCTACCATGATTTGCTAAAAATTATTAAGATAAAGATTTTTTGGATAATGAA
GAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGG
GATGATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAAC
AGCTTAAACGTGCGCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGT
ATTAGGGATAAGCAATCTGGCAAACAATATTAGATTTTTTGAATCAGATGGTTTTGCT
CAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAGAAGATATTC
AAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGATTGCTAACTTAGCT
GGCAGTCCTGCTATTA AAAAAGGTATTTTACAGACTGTAAAAATTGTTGATGAACTGGT
CAAAGTAATGGGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATCAGA
CAACTCAAAGGGCCAGAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGGTATC
AAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAA
TGAAAAGCTCTATCTCTATTATCTACAAAATGGAAGAGACATGTATGTGGACCAAGAAT
TAGATATTAATCGTTTAAAGTATTATGATGTCGATCACATTGTTCCACAAAGTTTCATT
AAAGACGATTCAATAGACAATAAGGTAACCGGTTCTGATAAAAATCGTGGTAAATC
GGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACCTTC
TAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGA
GTTTTGAGTGAACCTTGATAAAGCTGGTTTTATCAAACGCCAATTGTTGAAACTCGCCA
AATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAA
ATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAATCTAAATTAGTTTCTGAC
TTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCA
TGATGCGTATCTAAATGCCGTCGTTGGAAGTCTTTGATTAAGAAATATCCAAAACCTTG
AATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTTCGTAAAATGATTGCTAAG
TCTGAGCAAGAAATAGGCAAAGCAACCGCAAATATTTCTTTTACTCTAATATCATGAA
CTTCTTCAAACAGAAATTACACTTGCAAATGGAGAGATTTCGAAACGCCCTCTAATCG
AAACTAATGGGGAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGGCACAGTG
CGCAAAGTATTGTCCATGCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGG
CGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTA
AAAAAGACTGGGATCCAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCA
GTCCTAGTGTTGCTAAGGTGGAAAAAGGAAATCGAAGAAGTTAAAATCCGTTAAAGA
GTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAATCCGATTGACTTTT
TAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAACCTACCTAAATAT

10

20

30

40

50

AGTCTTTTTGAGTTAGAAAACGGTCGTAACGGATGCTGGCTAGTGCCGGAGAATTACA
 AAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAGTC
 ATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAG
 CAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGT
 TATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACA
 AACCAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGA
 GCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTAC
 AAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACAC
 GCATTGATTTGAGTCAGCTAGGAGGTGACTGA

【 0 1 3 5 】

10

MDKKYSIGLDIGTNSVGWAVITDDYKVPSSKFKVLGNTDRHSIKKNLIGALLFGSGETAEATRL
 KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
 HEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQIY
 NQLFEENPINASRVDAKAILSARLSKSRLENLIAQLPGEKRNGLFGNLIALSGLTPNFKSNF
 DLAEADAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSAS
 MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
 GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRI
 PYYVGPLARGNSRFAMWTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKS
 LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
 SVEISGVEDRFNASLGAYHDLLKI IKDKDFLDNEENEDI LEDIVLTLTLFEDRGMIEERLKTYA
 HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI LDFLKS DGFANRNFMQLIHDDSLTF
 KEDIQKAQVSGQGHSLSHEQIANLAGSPAIKKGILOTVKIVDELVKVMGHKPENIVIEMARENQT
 TQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRL
SDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLI TQRKF
DNLTKAERGGLSELKAGFIKRQLVETROITKHVAQILDSRMNTKYDENDKLIREVKVI TLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI VWDKGRDFATVRKVL SM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKGFFDSPTVAYSVLVVAKVEKKG
 SKKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLI IKLPKYSLFELENGRKRMLASA
 GELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEIEEQISEFSKRVI
 LADANLDKVL SAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA
 TLIHQSI TGLYETRIDLSQLGGD

20

30

40

(単一下線 : HNHドメイン ; 二重下線 : RuvCドメイン) 。

【 0 1 3 6 】

いくつかの実施形態において、野生型Cas9は、以下のヌクレオチドおよび/またはアミノ酸配列に対応するか、またはこれらを含む:

ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCCGTTGGATGGGCTGT
 CATAACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACC
 GTCATTCGATTAAAAAGAATCTTATCGGTGCCCTCCTATTTCGATAGTGGCGAAACGGCA
 GAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACCGAAT

50

ATGTTACTTACAAGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTC
 ACCGTTTGAAGAGTCCTTCCTTGTCTGAAGAGGACAAGAAACATGAACGGCACCCCATC
 TTTGGAAACATAGTAGATGAGGTGGCATATCATGAAAAGTACCCAACGATTTATCACCT
 CAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGGACCTGAGGTTAATCTACTTGGCTC
 TTGCCCATATGATAAAGTTCCGTGGGCACCTTTCTCATTGAGGGTGATCTAAATCCGGAC
 AACTCGGATGTCGACAAACTGTTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGA
 AGAGAACCCTATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCT
 CTAATCCCGACGGCTAGAAAACCTGATCGCACAATTACCCGGAGAGAAGAAAAATGGG
 TTGTTCCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAATTTTAAGTCGAACTT
 CGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCG
 ACAATCTACTGGCACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAAC
 CTTAGCGATGCAATCCTCCTATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGC
 GCCGTTATCCGCTTCAATGATCAAAAGGTACGATGAACATCACCAAGACTTGACACTTC
 TCAAGGCCCTAGTCCGTGAGCAACTGCCTGAGAAATATAAGGAAATATTCTTTGATCAG
 TCGAAAAACGGGTACGCGAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAA
 GTTTATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAAACTCA
 ATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTTCGACAACGGTAGCATTCCACATCAA
 ATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCGTTCT
 CAAAGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTTCGCATACCTTACTATGTGG
 GACCCCTGGCCGAGGGAACCTCTCGGTTCCGCATGGATGACAAGAAAGTCCGAAGAAACG
 ATTACTCCATGGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCAT
 CGAGAGGATGACCAACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACA
 GTTTACTTTACGAGTATTTACAGTGTACAATGAACTCACGAAAGTTAAGTATGTCACT
 GAGGGCATGCGTAAACCCGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCT
 GTTATTC AAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGA
 AAATTGAATGCTTCGATTCTGTGAGATCTCCGGGGTAGAAGATCGATTTAATGCGTCA
 CTTGGTACGTATCATGACCTCCTAAAGATAATTAAGATAAGGACTTCCTGGATAACGA
 AGAGAATGAAGATATCTTAGAAGATATAGTGTGACTCTTACCCTCTTTGAAGATCGGG
 AAATGATTGAGGAAAGACTAAAAACATACGCTCACCTGTTTCGACGATAAGGTTATGAAA
 CAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGATTGTGCGGAAACTTATCAACGG
 GATAAGAGACAAGCAAAGTGGTAAAACCTATTCTCGATTTTCTAAAGAGCGACGGCTTCG
 CCAATAGGAACTTTATGCAGCTGATCCATGATGACTCTTTAACCTTCAAAGAGGATATA
 CAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATATTGCGAATCTTGC
 TGGTTCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTGGATGAGCTAG
 TTAAGGTCATGGGACGTCACAAACCGGAAAACATTGTAATCGAGATGGCACGCGAAAAT
 CAAACGACTCAGAAGGGGCAAAAAACAGTTCGAGAGCGGATGAAGAGAATAGAAGAGG
 GTATTAAGAAGACTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTG
 CAGAACGAGAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCA
 GGAACCTGGACATAAACCGTTTATCTGATTACGACGTCGATCACATTGTACCCCAATCCT
 TTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCGAGGG
 AAAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGA ACTATTGGCGGCA
 GCTCCTAAATGCGAAACTGATAACGCAAAGAAAGTTTCGATAACTTAACTAAAGCTGAGA
 GGGGTGGCTTGTCTGAACTTGACAAGGCCGGATTTATTAACGTCAGCTCGTGGAACC
 CGCCAAATCACAAAGCATGTTGCACAGATACTAGATTCCCGAATGAATACGAAATACGA
 CGAGAACGATAAGCTGATTCGGGAAGTCAAAGTAATCACTTTAAAGTCAA AATTGGTGT
 CGGACTTCAGAAAGGATTTTCAATTCTATAAAGTTAGGGAGATAAATAACTACCACCAT
 GCGCACGACGCTTATCTTAATGCCGTGCTAGGGACCGCACTCATTAAGAAATACCCGAA
 GCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCGGTAAGATGATCG
 CGAAAAGCGAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATT
 ATGAATTTCTTTAAGACGGAAATCACTCTGGCAAACGGAGAGATACGCAAACGACCTTT

10

20

30

40

50

AATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTCGCGA
 CGGTGAGAAAAGTTTTGTCCATGCCCAAGTCAACATAGTAAAGAAAAGTGGAGTGCAG
 ACCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAGGAATAGTGATAAGCTCATCGC
 TCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCT
 ATTCTGTCTAGTAGTGGCAAAGTTGAGAAGGGAAAATCCAAGAACTGAAGTCAGTC
 AAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAAAGAACCCCATCGA
 CTTCTTGAGGCGAAAGTTACAAGGAAGTAAAAAGGATCTCATAATTAACCTACCAA
 AGTATAGTCTGTTTGGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAG
 CTTCAAAGGGGAACGAACCTCGCACTACCGTCTAAATACGTGAATTTCTGTATTTAGC
 GTCCATTACGAGAAGTTGAAAGTTTACCTGAAGATAACGAACAGAAGCAACTTTTTG
 TTGAGCAGCACAAACATTATCTCGACGAAATCATAGAGCAAATTTTCGGAATTCAGTAAG
 AGAGTCATCCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATAACAACAGCACAG
 GGATAAACCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACC
 TCGGCGCTCCAGCCGCATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACT
 TCTACCAAGGAGGTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTATATGA
 AACTCGGATAGATTTGTACAGCTTGGGGGTGACGGATCCCCAAGAAGAAGAGGAAAG
 TCTCGAGCGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAG
 GATGACGATGACAAGGCTGCAGGA

10

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
 KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHHRLEESFLVEEDKKHERHPIFGNIVDEVAY
 HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTY
 NQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNLFGNLIALLSLGLTPNFKSNF
 DLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
 MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
 GTEELLVKLNREDLLRQRTFDNGSIPHQIHLGELHAILRRQEDFYPFKDNREKIEKILTFRI
 PYYVGPLARGNSRFAMTRKSEETITPWNFEEVVDKGASQSFIERMTNFDKNLPNEKVLPHKS
 LLEYEFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
 SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYA
 HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI LDFLKSDFANRNFQM LIHDDSLTF
 KEDIQKAQVSGQDLSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ
 TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
 LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKROLVETROI TKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTA LIKKYPKLESEFVYGDYKVYDVRKMIAK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
MPQVNIVKKTEVQTTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKG
 KSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKDLIIKLPKYSLELENGRKRMLAS
 AGELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRV
 IILADANLDKVL SAYNKHDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
 ATLIHQSI TGLYETRIDLSQLGGD

20

30

40

50

(単一下線 : HNHドメイン ; 二重下線 : RuvCドメイン) 。

【 0 1 3 7 】

ある態様において、野生型Cas9は、*Streptococcus pyogenes*由来のCas9に対応する (NCBI参照配列:NC_002737.2 (以下のヌクレオチド配列) ; およびUniprot参照配列 : Q99ZW2 (以下のアミノ酸配列)) 。

```

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGT
GATCACTGATGAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACC
GCCACAGTATCAAAAAAAAAATCTTATAGGGGCTCTTTTATTTGACAGTGGAGAGACAGCG
GAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTAT
TTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTC
ATCGACTTGAAGAGTCTTTTTTGGTGGAAAGAAGACAAGAAGCATGAACGTCATCCTATT
TTTGAAATATAGTAGATGAAGTTGCTTATCATGAGAAATATCCAACCTATCTATCATCT
GCGAAAAAAAAATTGGTAGATTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCT
TAGCGCATATGATTAAGTTTCGTGGTCATTTTTTTGATTGAGGGAGATTTAAATCCTGAT
AATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAACCTACAATCAATTATTTGA
AGAAAACCCTATTAACGCAAGTGGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGA
GTAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAAAAATGGC
TTATTTGGGAATCTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTTAAATCAAATTTT
GATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAGATACTTACGATGATGATTTAGA
TAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATT
TATCAGATGCTATTTTACTTTTACAGATATCCTAAGAGTAAATACTGAAATAACTAAGGCT
CCCCTATCAGCTTCAATGATTAACGCTACGATGAACATCATCAAGACTTGACTCTTTT
AAAAGCTTTAGTTTCGACAACAACCTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAAT
CAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAA
TTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACTAAA
TCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCAA
TTCACCTGGGTGAGCTGCATGCTATTTTGAAGAAGACAAGAAGACTTTTATCCATTTTAA
AAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGG
TCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAA
TTACCCCATGGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATT
GAACGCATGACAACTTTGATAAAAATCTTCCAATGAAAAAGTACTACCAAACATAG
TTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTG
AAGGAATGCGAAAACCAGCATTTCTTTTACGGTGAACAGAAGAAAGCCATTGTTGATTTA
CTCTTCAAACAAATCGAAAAGTAACCGTTAAGCAATTAAGAAGATTATTTCAAAAA
AATAGAATGTTTTGATAGTGTGAAATTTTACGGAGTTGAAGATAGATTTAATGCTTCAT
TAGGTACCTACCATGATTTGCTAAAAATTATTAAGATAAAGATTTTTTTGGATAATGAA
GAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGA
GATGATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAAC
AGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGT
ATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTTTTTGAAATCAGATGGTTTTGC
CAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGCATTTTAAAGAAGACATTC
AAAAAGCACAAAGTGTCTGGACAAGGCGATAGTTTACATGAACATATTGCAAATTTAGCT
GGTAGCCCTGCTATTTAAAAAAGGTATTTTACAGACTGTAAAAGTTGTTGATGAATTGGT
CAAAGTAATGGGGCGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATC
AGACAACCTCAAAGGGCCAGAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGGT
ATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCA
AAATGAAAAGCTCTATCTCTATTATCTCCAAAATGGAAGAGACATGTATGTGGACCAAG
AATTAGATATTAATCGTTTAAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTC
CTTAAAGACGATTCAATAGACAATAAGGTCTTAAACGCGTTCTGATAAAAATCGTGGTAA
ATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAAC

```

10

20

30

40

50

TTCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGT
GGAGGTTTGAGTGAACCTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCG
CCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATG
AAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCT
GACTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATGC
CCATGATGCGTATCTAAATGCCGTCGTTGGAAGTCTTTGATTAAGAAATATCCAAAAC
TTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTTCGTAAAATGATTGCT
AAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCAT
GAACTTCTTCAAAAACAGAAATTACACTTGCAAATGGAGAGATTTCGAAAACGCCCTCTAA
TCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGGCACA
GTGCGCAAAGTATTGTCCATGCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGAC
AGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTC
GTAAAAAAGACTGGGATCCAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTAT
TCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAATCCGTTAA
AGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAATCCGATTGACT
TTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAA
TATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATT
ACAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTA
GTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAACAATTGTTTGTG
GAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCG
TGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAG
ACAAACCAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTT
GGAGCTCCCGCTGCTTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTC
TACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAA
CACGCATTGATTTGAGTCAGCTAGGAGGTGACTGA

10

20

30

40

50

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
 KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
 HEKYPTIYHLRKKLVDSSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
 NQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
 DLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
 MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
 GTEELLVKNLREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRFI
 PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKS
 LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
 SVEISGVEDRFNASLGTYHDLKI IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
 HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDFANRNFQMQLIHDDSLTF
 KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILOTVKVDELVKVMGRHKPENIVIAMARENQ
 TTQKGQKNSRERMKRIE EGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINR
 LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRK
 FDNLTKAERGGLSELDKAGFIKRQLVETROI TKHVAQILDSRMNTKYDENDKLIREVKVITLKS
 KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVRKMIAK
SEOEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKGFFDSPTVAYSVLVVAKEVG
 KSKKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
 AGELQKGNELALPSKYVNFYLYLASHYEKLKGSPEQKQLFVEQHKHYLDEIEQISEFSKRV
 ILADANLDKVL SAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
 ATLIHQSI TGLYETRIDLSQLGGD

10

20

30

(単一下線 : HNHドメイン ; 二重下線 : RuvCドメイン) 。

【 0 1 3 8 】

いくつかの実施形態において、Cas9は、*Corynebacterium ulcerans* (NCBI Refs: NC_015683.1, NC_017317.1); *Corynebacterium diphtheria* (NCBI Refs: NC_016782.1, NC_016786.1); *Spiroplasma syrphidicola* (NCBI Ref: NC_021284.1); *Prevotella intermedia* (NCBI Ref: NC_017861.1); *Spiroplasma taiwanense* (NCBI Ref: NC_021846.1); *Streptococcus iniae* (NCBI Ref: NC_021314.1); *Belliella baltica* (NCBI Ref: NC_018010.1); *Psychroflexus torquisi* (NCBI Ref: NC_018721.1); *Streptococcus thermophilus* (NCBI Ref: YP_820832.1), *Listeria innocua* (NCBI Ref: NP_472073.1), *Campylobacter jejuni* (NCBI Ref: YP_002344900.1) もしくは *Neisseria meningitidis* (NCBI Ref: YP_002342100.1) からのCas9を表し、または他の生物からのCas9を表す。

40

【 0 1 3 9 】

いくつかの実施形態において、dCas9は、Cas9のヌクレアーゼ活性を不活性化する1つ以上の変異を有するCas9アミノ酸配列に対応するか、またはその一部もしくは全部を含む。特に言及されない限り、Cas9における変異は野生型参照配列に対して相対的に表される。例えば、いくつかの実施形態では、dCas9ドメインはD10AおよびH840A変異を含むか

50

、または別のCas9における対応する変異を含む。いくつかの実施形態では、dCas9はdCas9 (D10AおよびH840A) のアミノ酸配列を含む：

MDKKYSIGLAIGTNSVGVAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
 KRTARRRYTRRKNRICYLQEIFSNEMAKVDDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
 HEKYPTIYHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTY
 NQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSGLLTPNFKSNF
 DLAEADAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
 MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
 GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRI
 PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
 LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
 SVEISGVEDRFNASLGT YHDL LKI IKDKDFLDNEENEDI LEDIVLTLTLFEDREMIEERLKYA
 HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI LDFLKS DGFANRNF MQLIHDDSLTF
 KEDIQKAQVSGQGSLHEHIANLAGSPAIKKGILOTVKVDELVKVMGRHKPENIVIAMARENQ
 TTQKGQKNSRERMKRIEIEG IKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINR
 LSDYDVDAIVPQSFLKDDSIDNKVLT RSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
 FDNLTKAERGGLSELDKAGFIKRQLVETROI TKHVAQILDSRMNTKYDENDKLIREVKVITLKS
 KLVSDFRKDFQFYK VREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIAK
 SEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL S
 MPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK KYGGFDSPTVAYSVLVVAKVEKG
 KSKKLSVKELLGITIMERS SFEKNPIDFLEAKGYKEVKKDLI IKLPKYSLFELENGRKRMLAS
 AGELQKGNELALPSKYVNFLYLASHYEK LKGS PEDNEQKQLFVEQHKHYLDEIIEQISEF SKRV
 ILADANLDKVL SAYNKHRDKPIREQAENI IHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVLD
 ATLIHQSI TGLYETRIDLSQLGGD

10

20

30

(単一下線：HNHドメイン；二重下線：RuvCドメイン)。

【0140】

いくつかの実施形態において、Cas9ドメインはD10A突然変異を含み、一方、上記で提供したアミノ酸配列における位置840における残基、または本明細書で提供されるアミノ酸配列のいずれかにおける対応する位置における残基は、ヒスチジンのままである。

【0141】

他の実施形態において、例えばヌクレアーゼ不活性化Cas9 (dCas9) をもたらす、D10AおよびH840A以外の突然変異を有するdCas9バリエーションが提供される。このような突然変異は、例えば、D10およびH840における他のアミノ酸置換、またはCas9のヌクレアーゼドメイン内の他の置換(例えば、HNHヌクレアーゼサブドメインおよび/またはRuvC1サブドメインにおける置換)を含む。ある態様において、dCas9のバリエーションまたはホモログであって、少なくとも約70%同一、少なくとも約80%同一、少なくとも約90%同一、少なくとも約95%同一、少なくとも約98%同一、少なくとも約99%同一、少なくとも約99.5%同一、または少なくとも約99.9%の同一性を有するものが提供される。ある態様において、約5アミノ酸、約10アミノ酸、約15アミノ酸、約20アミノ酸、約25アミノ酸、約30アミノ酸、約40アミノ酸、約50アミノ酸、約75アミノ酸、約100アミノ酸またはそれ以上

40

50

だけ短いまたは長いアミノ酸配列を有するdCas9のバリエーションが提供される。

【0142】

いくつかの実施形態において、本明細書に提供されるCas9融合タンパク質は、Cas9タンパク質の全長アミノ酸配列、例えば、本明細書に提供されるCas9配列の1つを含む。しかしながら、他の実施形態において、本明細書に提供される融合タンパク質は、全長Cas9配列を含まず、その1つ以上の断片のみを含む。好適なCas9ドメインおよびCas9断片の例示的アミノ酸配列が本明細書に提供され、Cas9ドメインおよび断片のさらなる好適な配列は、当業者には明らかであろう。

【0143】

Cas9タンパク質は、そのガイドRNAに相補的な特定のDNA配列にCas9タンパク質をガイドする、ガイドRNAと結合することができる。ある態様において、ポリヌクレオチドプログラム可能なヌクレオチド結合ドメインは、Cas9ドメイン、例えばヌクレアーゼ活性Cas9、Cas9ニッカーゼ (nCas9)、またはヌクレアーゼ不活性Cas9 (dCas9) である。核酸プログラム可能なDNA結合タンパク質の例としては、Cas9 (例えばdCas9およびnCas9)、CasX、CasY、Cpf1、Cas12b/C2c1、およびCas12c/C2c3が挙げられるが、これらに限定されない。

10

【0144】

ヌクレアーゼ不活性化Cas9タンパク質は、互換的に、「dCas9」タンパク質(ヌクレアーゼ「dead」Cas9の意)、または触媒的に不活性なCas9と称され得る。不活性なDNA切断ドメインを有するCas9タンパク質(又はその断片)を生成する方法が知られている(例えば、Jinek et al., Science. 337:816-821(2012); Qi et al., "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression" (2013) Cell. 28;152(5):1173-83参照(各内容は参照により本明細書に組み込まれる))。例えば、Cas9のDNA切断ドメインは、HNHヌクレアーゼサブドメインとRuvC1サブドメインの2つのサブドメインを含むことが知られている。HNHサブドメインはgRNAに相補的な鎖を切断し、RuvC1サブドメインは非相補的な鎖を切断する。これらのサブドメイン内の変異はCas9のヌクレアーゼ活性を抑制する。例えば、突然変異D10AとH840Aは、S.pyogenes Cas9のヌクレアーゼ活性を完全に不活性化する(Jinek et al., Science. 337:816-821(2012); Qi et al., Cell. 28;152(5):1173-83 (2013))。一例として、ヌクレアーゼ不活性Cas9ドメインは、クローニングベクターpPlatTET-gRNA2 (Accession No. BAV54124) に提供されるアミノ酸配列を含む。

20

30

【0145】

例示的な、触媒的に不活なCas9 (dCas9)のアミノ酸配列は以下の通りである：
MDKKYSIGLAIGTNSVGWAVITDEYKVPSPKFKVLGNTDRHSIKKNLIGALLFDSGETAEA
TRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSDVD
KLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIAL
SLGLTPNFKSNFDLAEDAKLQLSKDITYDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDI
LRVNTTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG
GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ
EDFYFPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAMTRKSEETITPWNFEVVDKGGAS
AQSFIERMNTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLIN
GIRDKQSGKTILDFLKSDFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA
GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGI
KELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVAIVPQSFLK
DDSIDNKVLRSDKNRGSNDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRK
DFQFYKVVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVDVRKMIKSEQ

40

50

EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
 MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK
 VEK GKSKKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
 KRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGKSPEDNEQKQLFVEQHKHYLDEII
 EQISEFSKRVLADANLDKVL SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID
 RKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGD

【 0 1 4 6 】

例示的な、触媒的Cas9ニッカーゼ (nCas9) のアミノ酸配列は以下の通りである :

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKFKVLGNTDRHSIKKNLIGALLFDSGETAEA
 TRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNI 10
 VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVD
 KLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLI
 SLGLTPNFKSNFDLAEDAKLQLSKD TYDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDI
 LRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG
 GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ
 EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
 AQSFIERM TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
 IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLIN
 GIRDKQSGKTILDFLKSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA 20
 GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGI
 KELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK
 DDSIDNKVLRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
 LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRK
 DFQFYK VREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ
 EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
 MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK
 VEK GKSKKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
 KRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGKSPEDNEQKQLFVEQHKHYLDEII
 EQISEFSKRVLADANLDKVL SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID 30
 RKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGD

【 0 1 4 7 】

例示的な、触媒的に活性なCas9のアミノ酸配列は以下の通りである :

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKFKVLGNTDRHSIKKNLIGALLFDSGETAEA
 TRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNI
 VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVD
 KLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLI
 SLGLTPNFKSNFDLAEDAKLQLSKD TYDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDI
 LRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG
 GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ 40
 EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
 AQSFIERM TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
 IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLIN
 GIRDKQSGKTILDFLKSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA
 GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGI
 KELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK
 DDSIDNKVLRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
 LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRK
 DFQFYK VREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ 50

EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
 MPQVNIKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK
 VEK GKSKKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
 KRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGKSPEDNEQKQLFVEQHKHYLDEII
 EQISEFSKRVLADANLDKVL SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID
 RKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD.

【0148】

ある態様において、Cas9は、単細胞原核微生物のドメインおよび界を構成する古細菌(例えばナノアーキア)由来のCas9を指す。ある態様において、プログラム可能なヌクレオチド結合タンパク質は、例えば、Burstein et al., "New CRISPR-Cas systems from uncultivated microbes." Cell Res. 2017 Feb 21. doi: 10.1038/cr.2017.21に記載されているCasXまたはCasYタンパク質であり得、その全体の内容は参照により本明細書に組み込まれる。ゲノム分解メタゲノミクスを用いて、生命の古細菌ドメインにおいて最初に報告されたCas9を含め、多くのCRISPR Cas系が同定された。この分岐Cas9タンパク質は、ほとんど研究されていないナノアーキアにおいて、活性CRISPR Cas系の一部として発見された。細菌では、それまで知られていなかった二つの系、CRISPR-CasXとCRISPR-CasYが発見され、それらは、これまでに発見された中でも最もコンパクトな系に入る。いくつかの実施形態において、本明細書に記載される塩基エディターシステムにおいて、Cas9は、CasXまたはCasXのバリエーションによって置き換えられる。いくつかの実施形態において、本明細書に記載される塩基エディターシステムにおいて、Cas9は、CasYまたはCasYのバリエーションによって置き換えられる。核酸プログラム可能DNA結合タンパク質(napDNAbp)として他のRNA誘導DNA結合タンパク質も使用され得、本開示の範囲内であることが理解されるべきである。

10

20

【0149】

いくつかの実施形態において、プログラム可能なヌクレオチド結合タンパク質(本明細書において核酸プログラム可能なDNA結合タンパク質(napDNAbp)とも呼ばれる)は、CasXタンパク質である。いくつかの実施形態では、プログラム可能なヌクレオチド結合タンパク質はCasYタンパク質である。いくつかの実施形態において、プログラム可能なヌクレオチド結合タンパク質は、天然に存在するCasXまたはCasYタンパク質に対して少なくとも85%、少なくとも90%、少なくとも91%、少なくとも92%、少なくとも93%、少なくとも94%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、または少なくとも99.5%の同一性を有するアミノ酸配列を含む。いくつかの実施形態において、プログラム可能なヌクレオチド結合タンパク質は、天然に存在するCasXまたはCasYタンパク質である。いくつかの実施形態において、プログラム可能なヌクレオチド結合タンパク質は、本明細書に記載されるいずれかのCasXまたはCasYタンパク質に対して少なくとも85%、少なくとも90%、少なくとも91%、少なくとも92%、少なくとも93%、少なくとも94%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、または少なくとも99.5%の同一性を有するアミノ酸配列を含む。他の細菌種由来のCasXおよびCasYもまた、本開示に従って使用され得ることを理解されたい。

30

40

【0150】

例示的CasX ((uniprot.org/uniprot/F0NN87; uniprot.org/uniprot/F0NH53) tr | F0NN87 | F0NN87_SULIHCRISPR-associatedCasx protein OS = Sulfolobus islandicus (strain HVE10/4) GN = SiH_0402 PE=4 SV=1)のアミノ酸配列は以下の通りである。

MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRG
 KAKKKKGEEGETTTSNIILPLSGNDKNPWTTETLKCYNFPTTVALSEVFNFSQVKECEEVS
 APSFVKPEFYEFGRSPGMVERTRRVKLEVEPHYLIIAAGWVLRGLGKAKVSEG DYVGVN
 VFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPVSVVRIYTTISDAVGQN
 PTTINGGFSIDLTKLLEKRYLLSERLEAIARNALSISSNMRERYIVLANIYIYEYLTG SKRLE

50

DLLYFANRDLIMNLSDDGKVRDLKLISAYVNGELIRGEG

【 0 1 5 1 】

例示的CasX (tr | F0NH53 | F0NH53_SULIR CRISPR associated protein, CasX O S = Sulfolobus islandicus (strain REY15A) GN=SiRe_0771 PE=4 SV=1) のアミノ酸配列は以下の通りである。

MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVNLAYKIAKNNEDAAAERRG
KAKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFNFSQVKECEEVS
APSFVKPEFYKFGRRSPGMVERTRRVKLEVEPHYLIMAAAGWVLRGKAKVSEG DYVGV
NVFTPTRGILYSLIQNVNGIVPGIKPETAFLWIARKVVSSVTNPNVSVVSIY TISDAVGG
NPTTINGGFSIDLTKLLEKRDLLSERLEAIARNALSISSNMRERYIVLANYIYEYLTGSKRLE
DLLYFANRDLIMNLSDDGKVRDLKLISAYVNGELIRGEG

10

【 0 1 5 2 】

Deltaproteobacteria CasX

MEKRINKIRKKLSADNATKPVSRSGPMKTLLVRVMTDDLKRRLEKRRKKPEVMPQVISN
NAANNLRMLDDYTKMKEAILQVYWQEFKDDHVGLMCKFAQPASKKIDQNKLPKPEMDE
KGNLTTAGFACSCGQPLFVYKLEQVSEK GKAYTNYFGRCNVAEHEKLILLAQLKPKV KDS
DEAVTYSLGKFGQRALDFYSIHVTKESTHPVKPLAQIAGNRYASGPV GKALSDACMG TIA
SFLSKYQDIIIEHQKVVKGNQKRLESLRELAGKENLEYP SVTLPPQPHTKEGVDfAYNEVI
ARVRM WVNLNLWQK LKLSRDDAKPLRLKGFSPFPVVERRENEVDWWNTINEVKKLID
AKRDMGRVFWSGVTAEKRNTILEGYNYPNENDHKKREGSLENPKKPAKRQFGDLLLYL
EKKYAGDWGKVFDEAWERIDKKIAGLTSHIEREEARNAEDAQSKAVLTDWLRKASFVL
ERLKEMDEKEFYACEIQLQKWYGDLRGNPF AVEAENRVVDISGFSIGSDGHSIQYRNLLA
WKYLENGKREFYLLMNYGKKGRIRFTDGTDIKKS GKWQGLLYGGGKAKVIDLTFDPDDE
QLIILPLAFGTRQGREFIWNLLSLETGLIKLANGRVIEKTIYNKKIGRDEPALFVALTFER
REVVDPSNIKPVNLIGVARGENIPAVIALTDPEGCPLPEFKDSSGGPTDILRIGEGYKEKQR
AIQAAKEVEQRRAGGYSRKFASKSRNLADDMVRNSARDLFYHAVTHDAVLVFANLSRGF
GRQGKRTFMTERQYTKMEDWLTAKLAYEGLTSKTYLSKTLAQYTSKTCSNCGFTITYAD
MDVMLVRLKKTSDGWATT LNNKELKA EYQITYYNRYKRQTVEKELSAELDRLSEESGNN
DISKWTKGRRDEALFLLKRF SHRPVQE QFVCLDCGHEVHAAEQ AALNIARSWFLNSN
STEFKSYKSGKQPFVGAWQAFYKRRLKEVWKPNA

20

30

【 0 1 5 3 】

例示的CasY ((ncbi.nlm.nih.gov/protein/APG80656.1)

APG80656.1 CRISPR-associated protein CasY [uncultured Parcubacteria group bacterium]) のアミノ酸配列は以下の通りである :

MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIVSAINDDYV
GLYGLSNFDDLYNAEKRNEEKVYSVLD FVYDCVQYGAVFSY TAPGLLKNVAEVRGGSYE
LTKTLKGSHLYDELQIDKVIKFLNKKEISRANGSLDKLKKDIIDCFKAEYRERHKDQC NKL
ADDIKNAKKDAGASLGERQKKLFRDFFGISEQSENDKPSFTNPLNLTCLLPFDTVNNR
NRGEVLFNKLKEYAQKLDKNEGSLEMWEYIGIGNSGTAFSNFLGEGFLGRLRENKITELK
KAMMDITDAWRGQEQEEELEKRLRILAALTIKLRPKFDNHWGGYRSDINGKLSSWLQN
YINQTVKIKEDLKGHKKDLKKAKEMINRFGESDTKEEAVVSSLLESIEKIVPDD SADDEKP
DIPAIAIYRRFLSDGRLTLNRFVQREDVQEALIKERLEAEK KKKPKKRKKKSDAEDEKETID
FKELFPHLAKPLKLVNPFYGDSKRELYK KYKNAAIYTDALWKAVEKIYKSAFSSSLKNSFF
DTDFDKDFFIKRLQKIFS VYRRFNTDKWKPIVKN SFAPYCDIVSLAENEVLYKPKQSR SRK
SAAIDKNRVRLPSTENIAKAGIALARELSVAGFDWKDLLKKEEHEEYIDLIELHKTALALL
LAVTETQLDISALDFVENGTVKDFMKT RDGNLVLEGRFLEMFSQSIVFSELRGLAGLMSR
KEFITRSAIQTMNGKQAE LLYIPHEFQSAKITTPKEMSRAFLDLAPAEFATSLEPESEK S
LLK LKQMRYYPHYFGYELTRTGQ GIDGGVAENALRLEKSPVKKREIKCKQYKTLGRGQNK
IVLYVRSSYYQTQFLEWFLHRPKNVQTDVAVSGSFLIDEKKVKTRWNYDALTVALEPVS
GSERVFVSQPFTIFPEKSAEEEGQRYL GIDIGEYGIAYTALEITGDSAKILDQNFISDPQLKT

40

50

LREEVKGLKLDQRRGTFAMPSTKIARIRESLVHSLRNRIHHLALKHKAKIVYELEVSRLFEE
GKQKIKKVYATLKKADVSEIDADKNLQTTVWGKLVASEISASYTSQFCGACKKLWRAE
MQVDETITTQELIGTVRVIKGGTLIDAIKDFMRPPIFDENDTPFPKYRDFCDKHHISKKMR
GNSCLFICPFCRANADADIQASQTIALLRYVKEEKKVEDYFERFRKLKN IKVLGQMKKI

【0154】

ある態様において、核酸プログラム可能DNA結合タンパク質 (napDNAbp) は、微生物CRISPR-Cas系の単一のエフェクターである。微生物CRISPR-Cas系の単一エフェクターは、Cas9、Cpf1、Cas12b/C2c1、およびCas12c/C2c3を含むが、これらに限定されない。典型的には、微生物CRISPR-Cas系はクラス1およびクラス2系に分けられる。クラス1の系は多サブユニットエフェクター複合体をもち、クラス2の系は単一のタンパク質エフェクターをもち、たとえば、Cas9とCpf1はクラス2エフェクターである。Cas9およびCpf1に加えて、三つの異なるクラス2 CRISPR-Casシステム(Cas12b/C2c1およびCas12c/C2c3)がShmakov et al., “Discovery and Functional Characterization of Diverse Class 2 CRISPR Cas Systems”, Mol. Cell, 2015 Nov. 5; 60(3): 385-397によって記載されている(その全体の内容は参照により本明細書に組み込まれる)。2つの系、Cas12b/C2c1とCas12c/C2c3のエフェクターはCpf1に関連するRuvC様エンドヌクレアーゼ領域を含む。第3の系は、2つの予測されるHEPN RNaseドメインをもちエフェクターを含む。Cas12b/C2c1によるCRISPR RNAの産生とは異なり、成熟CRISPR RNAの産生がtracrRNA非依存性である。Cas12b/C2c1はDNA切断にCRISPRRNAとtracrRNAの両方に依存する。

【0155】

Alicyclobacillus acidoterrastris Cas12b/C2c1 (AacC2c1) の結晶構造が、キメラ単一分子ガイドRNA (sgRNA) との複合体として報告されている。例えば、Liu et al., “C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism”, Mol. Cell, 2017 Jan. 19; 65(2):310-322参照(その内容全体を参照により本明細書に組み込む)。また、三元複合体として標的DNAに結合した*Alicyclobacillus acidoterrastris* C2c1においても結晶構造が報告されている。例えば、Yang et al., “PAM-dependent Target DNA Recognition and Cleavage by C2C1 CRISPR-Cas endonuclease”, Cell, 2016 Dec. 15; 167(7):1814-1828参照(その内容全体が参照により本明細書に組み込まれる)。標的DNA鎖および非標的DNA鎖の両方と共に、AacC2c1の触媒的にコンピテントな立体配座は、1つのRuvC触媒ポケット内に独立して捕捉され、Cas12b/C2c1媒介切断は標的DNAのスタガード7ヌクレオチドの切断を生じる。Cas 12b/C2c1三元複合体と以前に同定されたCas9およびCpf1対応物の間の構造比較は、CRISPR Cas9システムにより使用される機構の多様性を示す。

【0156】

いくつかの実施形態において、本明細書で提供される融合タンパク質のいずれかの核酸プログラム可能DNA結合タンパク質 (napDNAbp) は、Cas12b/C2c1またはCas12c/C2c3タンパク質であり得る。いくつかの実施形態において、napDNAbpはCas12b/C2c1タンパク質である。いくつかの実施形態において、napDNAbpはCas12c/C2c3タンパク質である。いくつかの実施形態において、napDNAbpは、天然に存在するCas12b/C2c1またはCas12c/C2c3タンパク質に対して少なくとも85%、少なくとも90%、少なくとも91%、少なくとも92%、少なくとも93%、少なくとも94%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、または少なくとも99.5%の同一性を有するアミノ酸配列を含む。いくつかの実施形態において、napDNAbpは、天然に存在するCas12b/C2c1またはCas12c/C2c3タンパク質である。いくつかの実施形態において、napDNAbpは、本明細書に提供されるnapDNAbp配列のいずれかに対して少なくとも85%、少なくとも90%、少なくとも91%、少なくとも92%、少なくとも93%、少なくとも94%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、または少なくとも99.5%の同一性を有するアミノ酸配列を含む。他の細菌種由来のCas12b/C2c1またはCas12c/C2c3も本開示に従って使用することができることを理解さ

りたい。

【 0 1 5 7 】

Cas12b/C2c1 ((uniprot.org/uniprot/T0D7A2#2) sp | T0D7A2 | /C2C1_ALIAG C RISR- associated endonuclease C2c1 OS = Alicyclobacillus acido-terrestris (st rain ATCC 49025 / DSM 3922/ CIP 106132 / NCIMB 13137/GD3B) GN=c2c1 PE=1 SV=1) のアミノ酸配列は以下の通りである :

MAVKSIVKVLRLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQ
 ECDKTAEECKAELLERLRARQVENGHRGPAGSDDELLQLARQLYELLVPAIGAKGDAQ
 QIARKFLSPLADKDAVGGGLGIAGNKPVRMREAGEPGWEEKEKAETRKSADRTAD
 VLRALADFGLKPLMRVYTDSEMSSVEWKPLRKGQAVRTWDRDMFQQAERMMSWESW 10
 NQRVQGQYAKLVEQKNRFEQKNFVQGEHLVHLVNQLQQDMKEASPGLESKEQTAHYVT
 GRALRGSDKVFKEWGLAPDAPFDLYDAEIKNVQRRNTRRFSGHDLFAKLAPEYQALW
 REDASFLTRYAVYNSILRKLNHAKMFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEF
 GERRHAIRFHKLLKVENGVAREVDDVTVPIISMSEQLDNLLPRDPNEPIALYFRDYGAEQH
 FTGEFGGAKIQCRDQLAHMHRRRGARDVYLVNSVVRVQSQSEARGERRPPYAAVFRLVG
 DNHRAVHFHDKLSDYLAHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDEL
 KPNSKGRVPPFFPIKGNLNVAVHERSQLLKLPGETESKDLRAIREERQRTLRLRTQLAY
 LRLLVRCGSEDVGRRRERSWAKLIEQPVDAANHMTDPDWREAFENELQKLKSLHGICSDKE
 WMDAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYAKDVVGGNSIEQIEYLERQY 20
 KFLKSWFFGKVSQVIRAEKGSRAITLREHIDHAKEDRLKKLADRIIMEALGYVYALDE
 RGKGWVAKYPPCQLILLEELSEYQFNDRPPSENNQLMQWSHRGVFQELINQAQVHD
 LLVGTMYAAFSSRFDARTGAPGIRCRVPARCTQEHNPEPFPWWLNKFVVEHTLDACPL
 RADDLIPTGEGEIVSPFSAEEGDFHQIHADLNAAQNLQQRLWSDFDISQIRLRCDWGEV
 DGELVLIPRLTGKRTADSYSNKVFYTNVTGVTYERERERGGKRRKVFQAEKLSSEEAELLVEA
 DEAREKSVVLMRDPGIIINRGWTRQKEFWSMV NQRIEGLVKQIRSRVPLQDSACENT
 GDI

【 0 1 5 8 】

BhCas12b (Bacillus hisashii) NCBI Reference Sequence: WP_095142515

【 0 1 5 9 】

MAPKKKRKVGIIHGVPAATRSLKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQE 30
 AIYEHHEQDPKNPKKVSKAIEQAELWDFVLKMQKCNSTHEVDKDEVFNILRELYEELVP
 SSVEKKGEANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEEKKWE
 EDKKKDPLAKILGKLAEYGLIPLFIPYTDSENEPIVKEIKWMEKSRNQSVRRLDKDMFIQAL
 ERFLSWESWNLKVKEEYKVEKEYKTLEERIKEDIQALKALEQYKERQEQLLRDTLNTN
 EYRLSKRGLRGWREIIQKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKE
 NHFIWRNHPEYPYLYATFCEIDKKKDAKQATFTLADPINHPLWVRFEERSGSNLNKY
 RILTEQLHTEKLLKLVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGK
 HAFTYKDESIFPLKGTGGARVQFDRDHLRRYPHKVESGNVGRIFYNMTVNIEPTESPV
 SKSLKIHRDDFPKVVNFKPKELTEWIKDSKGGKLSGIESLEIGLRVMSIDLQQRQAAAASI
 FEVVDQKPDIEGKLPFKGTLYAVHRASFNIKLPGETLVKSREVLKAREDNLKL MNQ 40
 KLNFLRNVLHFQQFEDITEREKRVTKWISRQENS DVPLVYQDELIQIRELMYKPYKDWV
 AFLKQLHKKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPG
 EVRRLEPGQRFAIDQLNHLNALKEDRLK KMANTIIMHALGYCYDVRKKKWKAKNPACQI
 ILFEDLSNYPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAK
 TGSPGIRCSVVTKEKLQDNRRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRK
 CVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIIEFGEG
 YFILKDGVEWVNAGKLIKKGSSKQSSSELVDSIDLKDSFDLASELKGEKLMLYRDPSTGN
 VFPSDKWMAAGVFFGKLERILISKLTNQYSISTIEDSSKQSMKRPAATKKAGQAKKKK

【 0 1 6 0 】

いくつかの実施形態では、Cas12bはBvCas12Bであり、これはBhCas12bのバリエーション 50

トであってBhCas12Bに対して以下の変化を含む：S893R、K846R、およびE837G。

【0161】

BvCas12b (Bacillus sp. V3-13) NCBI Reference Sequence: WP_101661451.1

【0162】

MAIRSIKLMKMTNSGTDSIYLRKALWRTHQLINEGIAYYMNLLTLYRQEAIGDKTKEAYQ
 AELINIIRNQQRNNGSSEEHGSDQEILALLRQLYELIIPSSIGESGDANQLGNKFLYPLVDP
 NSQSGKGTSNAGRKPRWKRLKEEGNPDWELEKKKDEERKAKDPTVKIFDNLNKYGLLPL
 FPLFTNIQKDIEWLPLGKRQSVRKWDKDMFIQAIERLLSWESWNRRVADEYKQLKEKTE
 SYYKEHLTGGEWIEKIRKFEKERNMELEKNAFAPNDGYFITSRQIRGWDRVYEWKSKLP
 ESASPEELWKVVAEQQNKMSSEFGDPKVFSLANRENNDIWRGHSERIYHIAAYNGLQK
 KLSRTKEQATFTLPDAIEHPLWIRYESPGGTNLNLFKLEEKQKKNYYVTLISKIWPSEK
 WIEKENIEIPLPSIQFNRQIKLKQHVKGKQEISFSDYSSRISLDGVLGGSRIQFNRKYIKN
 HKELGEGDIGPVFFNLVVDVAPLQETRNGRLQSPIGKALKVISSDFSKVIDYKPKELMD
 WMNTGSASNSFGVASLLEGMRVMSIDMGQRTSASVSIFEVVKELPKDQEQKLFYSINDTE
 LFAIHKRSFLLNLPGEVVTKNKQQRQERRKKRQFVRSQIRMLANVLRLETKKTTPDERKK
 AIHKLMEIVQSYDSWTASQKEVWEKELNLLTNMAAFNDEIWKESLVELHHRIEYPVGGI
 VSKWRKGLSEGRKNLAGISMWNIDELEDTRLLISWSKRSRTPGEANRIETDEPFSSLL
 QHIQNVKDDRLKQMANLIIMTALGFKYDKEEKDRYKRWKETYACQIILFENLNRYLFNL
 DRSRRENSRLMKWAHRSIPRTVSMQGMFGLQVGDVRSEYSSRFHAKTGAPGIRCHALT
 EEDLKAGSNTLKRLEDGFINESELAYLKKGDIIPSQGGELFVTLKRYKKDSDNNELTVIH
 ADINAAQNLQKRFWQQNSEVYRVPCQLARMGEDKLYIPKSQTETIKKYFGKGSFVKNNT
 EQEVYKWEKSEKMKIKTDTTFDLQDLDFEDISKTIELAQQKKYLTMFRDPSGYFFNN
 ETWRPQKEYWSIVNIIKSKLKKKILSNKVEL

10

20

【0163】

ポリヌクレオチドプログラム可能ヌクレオチド結合ドメインはまた、RNAに結合する核酸プログラム可能タンパク質を含むことができることを理解されたい。例えば、ポリヌクレオチドプログラム可能ヌクレオチド結合ドメインは、ポリヌクレオチドプログラム可能ヌクレオチド結合ドメインをRNAにガイドする核酸と結合され得る。他の核酸プログラム可能DNA結合タンパク質もまた、本開示の範囲内にあるが、それらは本開示には特に列記されていない。

30

【0164】

本明細書中で用いることができるCAsタンパク質は、クラス1およびクラス2を含む。Casタンパク質の非限定的な例としては、Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5d、Cas5t、Cas5h、Cas5a、Cas6、Cas7、Cas8、Cas9(Csn1またはCsx12とも呼ばれる)、Cas10、Csy1、Csy2、Csy3、Csy4、Cse1、Cse2、Cse3、Cse4、Cse5、Csn1、Csn2、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Ccssx16、Cx16、Cx、Csx3、Csx1、Csx1S、Csf1、Csf2、CsO、Csf4、Csd1、Csd2、Cst1、Cst2、Csh1、Csh2、Csa1、Csa2、Csa3、Csa4、Csa5、Cas12a/Cpfl、Cas12b/C2cl、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h、およびCas12i、CARF、DinG、それらのホモログ、またはそれらの改変体が挙げられる。未改変のCRISPR酵素は、Cas9のように、2つの機能性エンドヌクレアーゼ領域、RuvCおよびHNHを有するDNA切断活性を有することができる。CRISPR酵素は、標的配列内および/または標的配列の相補鎖内などの標的配列において、一方または両方の鎖の切断を誘導することができる。例えば、CRISPR酵素は、標的配列の最初または最後のヌクレオチドから約1、2、3、4、5、6、7、8、9、10、15、20、25、50、100、200、500塩基対またはそれ以上にある一方または両方の鎖の切断を誘導することができる。

40

【0165】

標的配列を含む標的ポリヌクレオチドの一方または両方の鎖を切断する能力を欠失するように、対応する野生型酵素に対して変異されたCRISPR酵素をコードするベクターを用

50

いることができる。Cas9は、野生型の例示的なCas9ポリペプチド(例えば*S. pyogenes*からのCas9)と少なくとも、または少なくともおよそ、50%、60%、70%、80%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、または100%の配列同一性および/または配列相同性を有するポリペプチドを指すことができる。Cas9は、野生型の例示的なCas9ポリペプチド(例えば*S. pyogenes*からのもの)に対して、最大で、または最大でおよそ、約50%、60%、70%、80%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、または100%の配列同一性および/または配列相同性を有するポリペプチドを指すことができる。Cas9は、野生型、または欠失、挿入、置換、バリエーション、突然変異、融合、キメラ、またはそれらの任意の組合せなどのアミノ酸変化を含み得るCas9タンパク質の改変型を指すことができる。

10

【0166】

いくつかの実施形態において、本明細書に記載される方法は、組み換え操作された(engineered) Casタンパク質を利用することができる。ガイドRNA (gRNA) は、Cas結合に必要な足場配列と、修飾改変されるべきゲノム標的を規定するユーザー定義の約20塩基スペーサーとからなる短い合成RNAである。いくつかの実施形態において、その足場は、GUUUAGAGC UAGAAUAGC AAGUUAUUU AAGGCUAGUC CGUUAUCAAC UUGAAA AAGU GGCACCGAGU CGGUGCUUUUを含む。当業者がCasタンパク質特異性のゲノム標的を変えることができるかどうかは特に、gRNA標的化配列がゲノムの他の部分と比較してゲノム標的に対していかに特異的であるかによって決定される。

【0167】

20

Cas9ヌクレアーゼはRuvCとHNHという二つの機能性エンドヌクレアーゼ領域をもっている。Cas9は標的DNAに結合すると二番目のコンホメーション変化を起こし、これがヌクレアーゼドメインを位置付け、標的DNAの反対側の鎖を切断させる。Cas9を介したDNA切断の最終結果は、標的DNA内の二本鎖切断 (DSB) である(PAM配列の約3~4ヌクレオチド上流)。生じたDSBは、次の二つの一般的修復経路のうちの一つにより修復される: (1) 効率的だが誤りがちな非相同末端結合 (NHEJ) 経路; または (2) 効率は低いが高忠実度の高い相同性誘導修復 (HDR) 経路。

【0168】

非相同末端結合 (NHEJ) および/または相同性指向修復 (HDR) の「効率」は、任意の簡便な方法によって計算することができる。例えば、ある場合には、効率は、成功したHDRのパーセンテージで表すことができる。例えば、検査用ヌクレアーゼ分析を用いて切断産物を生成し、基質に対する産物の比を用いてパーセンテージを計算することができる。たとえば、HDRが成功した結果として新たに組み込まれた制限配列を含むDNAを直接切断する測定用ヌクレアーゼ酵素を使用することができる。切断される基質が多いほどHDRの割合が高かった(HDRの効率がより高かった)ことを示す。例示的な例として、HDRの割合(パーセンテージ)は、以下の式を用いて計算することができる[(切断産物)/(基質+切断産物)] (例えば、(b+c) / (a+b+c) ここで、「a」はDNA基質のバンド強度であり、「b」および「c」は切断生成物である。)

30

【0169】

場合によっては、効率はNHEJの成功率で表すことができる。例えば、T7エンドヌクレアーゼIアッセイを用いて切断産物を生成し、基質に対する産物の比を用いてNHEJのパーセンテージを計算することができる。T7エンドヌクレアーゼIは、野生型および突然変異体DNA鎖(NHEJは最初の切断部位に小さなランダムな挿入や欠失(インデル)を生じる)のハイブリダイゼーションから生じる mismatches のヘテロ二本鎖DNAを切断する。切断が多いほどNHEJの割合が高いこと(NHEJの効率が低いこと)を示す。例示的な例として、NHEJの割合(パーセンテージ)は、式 $(1 - (1 - (b+c)/(a+b+c))^{1/2}) \times 100$ を用いて計算することができ、ここで、「a」はDNA基質のバンド強度であり、「b」および「c」は切断生成物である(Ran et al., Cell. 2013 Sep. 12; 154(6):1380-9; and Ran et al., Nat Protoc. 2013 Nov.; 8(11): 2281-2308)。

40

【0170】

50

NHEJ修復経路は最も活性な修復機構であり、DSB部位での小ヌクレオチド挿入または欠失(インデル)を頻繁に引き起こす。NHEJ媒介DSB修復のランダム性は、Cas9およびgRNAまたはガイドポリヌクレオチドを発現する細胞集団が多様な突然変異のアレイをもたらすので、重要な実質的意味を有する。ほとんどの場合、NHEJは標的DNAに小さなインデルを生じさせ、その結果、標的遺伝子のオープンリーディングフレーム(ORF)内に未熟な終止コドンをもたらすアミノ酸欠失、挿入、またはフレームシフト突然変異が生じる。理想的な最終結果は、標的遺伝子内の機能喪失型変異である。

【0171】

NHEJ媒介DSB修復はしばしば遺伝子のオープンリーディングフレームを破壊するが、相同性指向修復(HDR)は単一ヌクレオチド変化からフルオロフォアまたはタグの付加のような大きな挿入までの範囲の特異的ヌクレオチド変化を生成するために使用できる。

10

【0172】

HDRを遺伝子編集に利用するために、所望の配列を含むDNA修復テンプレートを、gRNAおよびCas9またはCas9ニッカーゼと共に目的の細胞型に送達することができる。修復テンプレートは、所望の編集ならびにその標的のすぐ上流および下流(左右相同アームと呼ばれる)にさらなる相同配列を含むことができる。各相同アームの長さは導入される変化の大きさに依存し得、より大きな挿入はより長い相同アームを必要とする。修復鑄型は、一本鎖オリゴヌクレオチド、二本鎖オリゴヌクレオチド、または二本鎖DNAプラスミドであり得る。HDRの効率は、Cas9、gRNAおよび外因性修復テンプレートを発現する細胞においても、一般的に低い(10%未満の修正アリル)。HDRは細胞周期のS期とG2期の間に起こるので、細胞を同調させることによってHDRの効率を高めることができる。NHEJに關与する遺伝子を化学的または遺伝的に阻害することもHDR頻度を増加させ得る。

20

【0173】

いくつかの実施態様において、Cas9は、修飾Cas9である。所与のgRNA標的配列は、ゲノム全体にわたり、部分的な相同性が存在するさらなる部位を有し得る。これらの部位はオフターゲットと呼ばれ、gRNAを設計する際に考慮する必要があるが、gRNAの設計を最適化することに加えて、Cas9を修飾することによってCRISPRの特異性を高めることもできる。Cas9は2つのヌクレアーゼドメイン、RuvCおよびHNHの組合せ活性を介して二本鎖切断(DSB)を生成する。SpCas9のD10A変異体であるCas9ニッカーゼは1つのヌクレアーゼドメインを保持し、DSBではなくDNAニックを生成する。特異的遺伝子編集のためのHDR媒介遺伝子編集にニッカーゼを組み合わせることもできる。

30

【0174】

ある態様において、改変されたCas9は、高忠実度Cas9酵素である。いくつかの実施形態において、高忠実度Cas9酵素は、SpCas9(K855A)、eSpCas9(1.1)、SpCas9-HF1、または高精度Cas9バリエーション(HypaCas9)である。修飾Cas9eSpCas9(1.1)は、HNH/RuvCグループと非標的DNA鎖との間の相互作用を弱めるアラニン置換を含み、オフターゲット部位での鎖分離と切断を防止する。同様に、SpCas9-HF1は、DNAリン酸骨格とのCas9の相互作用を破壊するアラニン置換を介して、オフターゲット編集を低下させる。HypaCas9は、Cas9の校正と標的識別を増加させる突然変異をREC3ドメインに含む(SpCas9 N692A/M694A/Q695A/H698A)。3つの高忠実度酵素はすべて、野生型Cas9よりもオフターゲット編集が少ない。例示的な高忠実度Cas9のアミノ酸配列を下記に提供する。この配列において、参照Cas9に対する高忠実度Cas9ドメイン変異は太字、下線で示している。

40

MDKKYSIGLAIGTNSVGWAVITDEYKVPKFKFKVLGNTDRHSIKKNLIGALLFDSGETAEA
 TRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLVESFLVEEDKKHERHPIFGNI
 VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVD
 KLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIAL
 SLGLTPNFKSNFDLAEDAKLQLSKDITYDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDI
 LRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG
 GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRFTDNGSIPHQIHLGELHAILRRQ

50

EDFYPPFLKDNREKIEKILTFRIPYYVGLARGNSRFAMTRKSEETITPWNFEEVVDKGAS
 AQSFIERMTAFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
 IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGALSRKLIN
 GIRDKQSGKTILDFLKSDFANRNFMA^{LI}HDDSLTFKEDIQKAQVSGQGD^{SL}HEHIANLAG
 SPAIKKGILQTVKVVDELVKVMGRHKPENIVIE^{MARE}NQTTQKGQKNSRERMKRIE^{EGIK}
 ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD^{VDHIV}PQSFLK
 DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT^{KA}ERGG
 LSELDKAGFIKRQLVETRAITKHVAQILDSRMNTKYDENDKLIREVKVITL^{SKL}VSDFRK
 DFQFYK^{VREINNYHHA}HDAYLNAVVG^{TALIK}KYPKLESEFVYGDYK^{VYDVR}KMI^{AK}SEQ
 EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK^{VLS}
 MPQVNI^{VKKTEV}QTGGFSKESILPKRNSDKLIARKKDWDPK^{KYGG}FDSPTVAYS^{VLV}VAK
 VEK^{GKSK}KLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV^{KKDLII}KLPK^{YSL}FELENGR
 KRMLASAGELQKGNELALPSKYVNFLYLASHYEK^{LK}GSPE^{DNEQ}KQLFVEQH^{KHYL}DEII
 EQISEFSKRVILADANLDK^{VLSAY}NK^{HRDK}PIREQAENIIHLFTLTNLGAPAAFKY^{FD}TTID
 RKRYTSTKEVLDATLIHQ^{SITGLY}ETRIDL^{SQL}GGD.

10

【0175】

ある場合には、Cas9はバリエーションCas9タンパク質である。バリエーションCas9ポリペプチドは、野生型Cas9タンパク質のアミノ酸配列と比較して、一アミノ酸単位で異なる(例えば、欠失、挿入、置換、融合を有する)アミノ酸配列を有する。いくつかの例において、バリエーションCas9ポリペプチドは、Cas9ポリペプチドのヌクレアーゼ活性を低下させるアミノ酸変化(例えば欠失、挿入、または置換)を有する。例えば、いくつかの例において、バリエーションCas9ポリペプチドは、対応する野生型Cas9タンパク質のヌクレアーゼ活性の50%未満、40%未満、30%未満、20%未満、10%未満、5%未満、または1%未満を有する。ある場合には、バリエーションCas9タンパク質は実質的なヌクレアーゼ活性をもたない。対象Cas9タンパク質が、実質的なヌクレアーゼ活性を有さないバリエーションCas9タンパク質である場合、それは「dCas9」と称され得る。

20

【0176】

ある場合には、バリエーションCas9タンパク質はヌクレアーゼ活性を低下させる。例えば、バリエーションCas9タンパク質は、野生型Cas9タンパク質(例えば野生型Cas9タンパク質)のエンドヌクレアーゼ活性の約20%未満、約15%未満、約10%未満、約5%未満、約1%未満、または約0.1%未満を示す。

30

【0177】

ある場合には、バリエーションCas9タンパク質は、ガイド標的配列の相補鎖を切断することができるが、二本鎖ガイド標的配列の非相補鎖を切断する能力が低下している。例えば、バリエーションCas9タンパク質は、RuvCドメインの機能を低下させる突然変異(アミノ酸置換)を有することができる。非限定的な例として、いくつかの実施形態において、バリエーションCas9タンパク質は、D10A(アミノ酸位置10におけるアスパラギン酸からアラニン)を有し、したがって、二本鎖ガイド標的配列の相補鎖を切断することができるが、二本鎖ガイド標的配列の非相補鎖を切断する能力が低下している(したがって、このバリエーションCas9タンパク質が二本鎖標的核酸を切断するとき、二本鎖切断(DSB)の代わりに一本鎖切断(SSB)を生じる)(例えばJinek et al., Science. 2012 Aug. 17; 337(6096):816-21参照)。

40

【0178】

ある場合には、バリエーションCas9タンパク質は、二本鎖ガイド標的配列の非相補鎖を切断することができるが、ガイド標的配列の相補鎖を切断する能力が低下している。例えば、バリエーションCas9タンパク質は、HNHドメインの機能を低下させる変異(アミノ酸置換)を有することができる(RuvC/HNH/RuvCドメインモチーフ)。非限定的な例として、いくつかの実施形態において、バリエーションCas9タンパク質は、H840A(アミノ酸位置840におけるヒスチジンからアラニン)突然変異を有し、したがって、ガイド標的配列の非相補的スト

50

ランドを切断することができるが、ガイド標的配列の相補的ストランドを切断する能力が低下している(したがって、このバリエーションCas9タンパク質が二本鎖ガイド標的配列を切断すると、DSBの代わりにSSBが生じる)。このようなCas9タンパク質は、ガイド標的配列(例えば一本鎖ガイド標的配列)を切断する能力が低下しているが、ガイド標的配列(例えば一本鎖ガイド標的配列)に結合する能力を保持している。

【0179】

ある場合には、バリエーションCas9タンパク質は、二本鎖標的DNAの相補鎖および非相補鎖の両方を切断する能力が低下している。非限定的な例として、ある場合には、バリエーションCas9タンパク質は、D10AおよびH840A突然変異の両方を有し、その結果、ポリペプチドは、二本鎖標的DNAの相補鎖および非相補鎖の両方を切断する能力が低下している。この

10

【0180】

別の非限定的な例として、いくつかの場合において、バリエーションCas9タンパク質は、W476AおよびW1126A突然変異を有し、その結果、ポリペプチドは、標的DNA(例えば一本鎖標的DNA)を切断する能力が低下しているが、標的DNA(例えば一本鎖標的DNA)に結合する能力は保持している。

【0181】

別の非限定的な例として、いくつかの場合において、バリエーションCas9タンパク質は、P475A、W476A、N477A、D1125A、W1126A、およびD1127A突然変異を有し、その結果、ポリペプチドは、標的DNAを切断する能力が低下している。そのようなCas9タンパク質は、標的DNAを切断する能力を低下しているが(例えば一本鎖標的DNA)、標的DNA(例えば一本鎖標的DNA)に結合する能力を保持している。

20

【0182】

別の非限定的な例として、いくつかの場合において、バリエーションCas9タンパク質は、H840A、W476A、およびW1126A突然変異を有し、その結果、ポリペプチドは、標的DNA(例えば一本鎖標的DNA)に結合する能力は保持している。別の非限定的な例として、いくつかの場合において、バリエーションCas9タンパク質は、H840A、D10A、W476A、およびW1126A突然変異を有し、その結果、ポリペプチドは、標的DNAを切断する能力が低下している。このようなCas9タンパク質は、標的DNA(例えば一本鎖標的DNA)を切断する能力が低下しているが、標的DNA(例えば一本鎖標的DNA)に結合する能力は保持している。いくつかの実施形態において、バリエーションCas9は、Cas9 HNHドメインの位置840における触媒的His残基が回復されている(A840H)。

30

【0183】

別の非限定的な例として、いくつかの場合において、バリエーションCas9タンパク質は、H840A、P475A、W476A、N477A、D1125A、W1126A、およびD1127A突然変異を有し、その結果、ポリペプチドは、標的DNA(例えば一本鎖標的DNA)を切断する能力を低下させるが、標的DNA(例えば一本鎖標的DNA)に結合する能力を保持する。別の非限定的な例として、いくつかの場合において、バリエーションCas9タンパク質は、D10A、H840A、P475A、W476A、N477A、D1125A、W1126A、およびD1127A突然変異を有し、その結果、ポリペプチドは、標的DNAを切断する能力が低下している。そのようなCas9タンパク質は、標的DNA(例えば一本鎖標的DNA)を切断する能力を低下しているが、標的DNA(例えば一本鎖標的DNA)に結合する能力を保持している。バリエーションCas9タンパク質がW476AおよびW1126A変異を有する場合、またはバリエーションCas9タンパク質がP475A、W476A、N477A、D1125A、W1126A、およびD1127A変異を有する場合、バリエーションCas9タンパク質はPAM配列に効率的に結合しない。したがって、このような場合には、このようなバリエーションCas9タンパク質を結合の方法に用いると、この方法はPAM配列を必要としない。換言すれば、ある場合には、このようなバリエーションCas9タンパク質を結合の方法に用いる場合、この方法はガイドRNAを含み得るが、この方法は、PAM配列の

40

50

非存在下で行うことができる(したがって、結合の特異性はガイドRNAの標的セグメントによってもたらされる)。上記の効果を達成するために、他の残基を変異させ得る(すなわち一方または他方のヌクレアーゼ部分を不活性化する)。非限定的な例として、残基D10、G12、G17、E762、H840、N854、N863、H982、H983、A984、D986、および/またはA987を変異(すなわち置換)することができる。また、アラニン置換以外の変異も好適である。

【0184】

ある態様において、低下した触媒活性を有するバリエーションCas9タンパク質(例えばCas9タンパク質がD10、G12、G17、E762、H840、N854、N863、H982、H983、A984、D986、および/またはA987突然変異、例えばD10A、G12A、G17A、E762A、H840A、N854A、N863A、H982A、H983A、A984A、および/またはD986Aを有する場合)は、それがガイドRNAと相互作用する能力を保持する限り、部位特異的様式で標的DNAに結合することができる(ガイドRNAによって標的DNA配列に誘導されるからである)。

10

【0185】

S. pyogenes Cas9の代替としては、哺乳動物細胞において切断活性を示すCpf1ファミリー由来のRNA誘導エンドヌクレアーゼが挙げられ得る。PrevotellaおよびFrancisella 1由来のCRISPR (CRISPR/Cpf1) は、CRISPR/Cas9システムに類似したDNA編集技術である。Cpf1はクラスII CRISPR/Cas系のRNA誘導エンドヌクレアーゼである。この獲得免疫機構はPrevotellaやFrancisella細菌に見られる。Cpf1遺伝子はCRISPR遺伝子座に関連しており、ウイルスDNAを見出して切断するためにガイドRNAを用いるエンドヌクレアーゼをコードしている。Cpf1はCas9より小さく単純なエンドヌクレアーゼであるため、CRISPR/Cas9系の制限のいくつかを克服する。Cas9ヌクレアーゼとは異なり、Cpf1を介したDNA切断の結果は、短い3'突出を伴う二本鎖切断である。Cpf1の互い違いの切断パターンは、伝統的な制限酵素クローニングに類似した、方向性のある遺伝子導入の可能性を開くことができ、これは遺伝子編集の効率を高め得る。上述したCas9のバリエーションおよびオーソログと同様に、Cpf1は、CRISPRが標的とすることができる部位の数を、SpCas9が好むNGG PAM部位を欠くATに富む領域またはATに富むゲノムに拡大することもできる。Cpf1遺伝子座は / 混合ドメイン、RuvC Iとそれに続くらせん領域、RuvC IIおよびジンクフィンガー様ドメインを含む。Cpf1タンパク質は、Cas9のRuvCドメインに類似したRuvC様エンドヌクレアーゼドメインを有する。さらに、Cpf1はHNHエンドヌクレアーゼ領域をもたず、Cpf1のN末端はCas9のヘリックス認識ローブをもたない。Cpf1 CRISPR Casドメイン構成は、Cpf1が機能的にユニークであり、クラス2、タイプV CRISPRシステムとして分類されることを示した。Cpf1遺伝子座は、II型系よりもI型およびII型に類似したCas1、Cas2およびCas4タンパク質をコードしていた。機能的Cpf1はトランス活性化CRISPR RNA (tracrRNA) を必要としない;したがって、CRISPR (crRNA) だけを要する。Cpf1はCas9より小さいだけでなく、より小さいsgRNA分子(Cas9の約半分の数)のヌクレオチドを有するので、これはゲノム編集に有益である。Cas9が標的とするグリッチPAMとは対照的に、Cpf1-crRNA複合体は、モチーフ5'-YTN-3'に隣接するプロトSpacerの同定によって標的DNAまたはRNAを切断する。PAMの同定後、Cpf1は、4または5ヌクレオチドの突出を有するスティッキーエンド様のDNA二本鎖切断を導入する。

20

30

40

【0186】

本開示のいくつかの態様は、核酸プログラム可能なDNA結合タンパク質として作用するドメインを含む融合タンパク質を提供し、これは、例えば塩基エディターのようなタンパク質を特定の核酸(例えばDNAまたはRNA)配列へと誘導するために使用され得る。特定の実施形態において、融合タンパク質は、核酸プログラム可能なDNA結合タンパク質ドメインおよびデアミナーゼドメインを含む。DNA結合タンパク質としては、非限定的に、Cas9(例えばdCas9およびnCAs9)、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h、およびCas12iが挙げられる。Cas9とは異なるPAM特異性を有する核酸プログラム可能なDNA結合タンパク質の一例は、Prevotella および Francisella 1からのClustered Regularly Interspaced Short Palindromic

50

mic Repeatsである。Cas9と同様に、Cpf1はクラス2のCRISPRエフェクターである。Cpf1はCas9とは異なる特徴で頑強なDNA干渉を媒介することが示されている。Cpf1は、tracrRNAを欠く単一RNA誘導エンドヌクレアーゼであり、Tに富むプロトスペーサ隣接モチーフ(TTN、TTN、またはYTN)を利用する。さらに、Cpf1はDNAを互い違いの二本鎖切断で切断する。16のCpf1ファミリータンパク質のうち、AcidaminococcusとLachnospiraceaeからの二つの酵素がヒト細胞において効率的なゲノム編集活性を有することが示されている。Cpf1タンパク質は当技術分野で知られており、例えば以前にYamano et al., "Crystal structure of Cpf1 in complex with guide RNA and target DNA." Cell (165) 2016, p. 949-962により記述されており、その全体の内容は参照により本明細書に組み込まれる。

10

【0187】

また、ガイドヌクレオチド配列プログラム可能なDNA結合タンパク質ドメインとして使用され得るヌクレアーゼ不活性Cpf1 (dCpf1) バリエーションも、本発明の組成物および方法において有用である。Cpf1タンパク質は、Cas9のRuvCドメインに類似したRuvC様エンドヌクレアーゼドメインを有するが、HNHエンドヌクレアーゼドメインを有さず、Cpf1のN末端はCas9の α -ヘリックス認識ローブを有さない。Zetsche et al., Cell, 163, 759-771, 2015 (参照により本明細書に組み込まれる)は、Cpf1のRuvC様ドメインが両方のDNA鎖の切断を担い、RuvC様ドメインの不活性化はCpf1ヌクレアーゼ活性を不活化することを示した。例えば、Francisella novicida Cpf1におけるD917A、E1006A、またはD1255Aに対応する突然変異は、Cpf1ヌクレアーゼ活性を不活化する。いくつかの実施形態において、本開示のdCpf1は、D917A、E1006A、D1255A、D917A/E1006A、D917A/D1255A、E1006A/D1255A、またはD917A/E1006A/D1255Aに対応する突然変異を含む。Cpf1のRuvCドメインを不活性化する任意の突然変異、例えば、置換変異、欠失、または挿入が、本開示に従って使用され得ることを理解されたい。

20

【0188】

いくつかの実施形態において、本明細書で提供される融合タンパク質のいずれかの核酸プログラム可能DNA結合タンパク質 (napDNAbp) は、Cpf1タンパク質であり得る。いくつかの実施形態において、Cpf1タンパク質は、Cpf1ニッカーゼ (nCpf1) である。ある状態において、Cpf1タンパク質は、ヌクレアーゼ不活性Cpf1 (dCpf1) である。いくつかの実施形態において、Cpf1、nCpf1、またはdCpf1は、本明細書に開示されたCpf1配列に対して少なくとも85%、少なくとも90%、少なくとも91%、少なくとも92%、少なくとも93%、少なくとも94%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、または少なくとも99.5%の同一性を有するアミノ酸配列を含む。いくつかの実施形態において、dCpf1は、本明細書に開示されたCpf1配列に対して少なくとも85%、少なくとも90%、少なくとも91%、少なくとも92%、少なくとも93%、少なくとも94%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、または少なくとも99.5%の同一性を有するアミノ酸配列を含み、D917A、E1006A、D1255A、D917A/E1006A、D917A/D1255A、E1006A/D1255A、またはD917A/E1006A/D1255Aに対応する突然変異を含む。他の細菌種からのCpf1もまた、本開示に従って使用され得ることを理解されたい。

30

40

【0189】

野生型Francisella novicida Cpf1のアミノ酸配列は以下の通りである。D917、E1006、およびD1255は太字で下線を付している。

```
MSIYQEFVNKYSLSKTLRFELIPQGKTLLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
EEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFN
QNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIISFKGWTTYFKGFHEN
RKNVYSSNDIPTSIYRIVDDNLPKFLNKAKYESLKDKAPEAINYEQIKKDLAEELTFDID
YKTSEVNQRVFLSDEVFEIANFNLYLNQSGITKFNTIIGGKVFVNGENTKRKGINEYINLYS
QQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTMQSFYEQIAAFKTVEEKS
IKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLD
```

50

NPSKKEQELIAKKTEKAKYLSLETIKLALAEFNKHRDIDKQCRFEEILANFAAIPMIFDEIA
 QNKDNLAQISIKYQNQGKDLLQASAEDDVKAIKDLLDQTNLLHKLKIFHISQSEDKAN
 ILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKE
 PDNTAILFIKDDKYLLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFF
 SAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKD
 FGFRFSDTQRYNSIDFYREVENQGYKLTFFENISESYIDSVVNQGKLYLFQIYNKDFSAYS
 GRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNDNP
 KKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRG
 ERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKE
 MKEGYLSQVVHEIAKLVEIYNAIVVFEDLNF^UFGFKRGRFKVEKQVYQKLEKMLIEKLNLYLV
 FKDNFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKY
 ESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKN
 HNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSK
 TGTELDYLISPVADVNGNFFDSRQAPKNMPQDAD^UANGAYHIGLKGLMLLGRIKNNQEGK
 KLNLVIKNEEYFEFVQNRNN.

10

【 0 1 9 0 】

Francisella novicida Cpf1 D917Aのアミノ酸配列は以下の通りである。(A917、E1006、およびD1255には太字で下線を付している。)

MSIYQEFVNKYLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
 EEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFN
 QNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIISFKGWTTYFKGFHEN
 RKNVYSSNDIPTSIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDID
 YKTSEVNQRVFSLDEVFEIANFN^ULNQSGITKFNTIIGGK^UFVNGENTKRKGINEYINLYS
 QQINDKTLK^UKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKS
 IKETLSLLFDDLKAQKLDLSKIYFKNDKSLDLSQQVFDDYSVIGTAVLEYITQQIAPKNLD
 NPSKKEQELIAKKTEKAKYLSLETIKLALAEFNKHRDIDKQCRFEEILANFAAIPMIFDEIA
 QNKDNLAQISIKYQNQGKDLLQASAEDDVKAIKDLLDQTNLLHKLKIFHISQSEDKAN
 ILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKE
 PDNTAILFIKDDKYLLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFF
 SAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKD
 FGFRFSDTQRYNSIDFYREVENQGYKLTFFENISESYIDSVVNQGKLYLFQIYNKDFSAYS
 GRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNDNP
 KKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIARG
 ERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKE
 MKEGYLSQVVHEIAKLVEIYNAIVVFEDLNF^UFGFKRGRFKVEKQVYQKLEKMLIEKLNLYLV
 FKDNFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKY
 ESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKN
 HNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSK
 TGTELDYLISPVADVNGNFFDSRQAPKNMPQDAD^UANGAYHIGLKGLMLLGRIKNNQEGK
 KLNLVIKNEEYFEFVQNRNN.

20

30

40

【 0 1 9 1 】

Francisella novicida Cpf1 E1006Aのアミノ酸配列は以下の通りである。(D917、A1006、およびD1255には太字で下線を付している。)

MSIYQEFVNKYLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
 EEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFN
 QNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIISFKGWTTYFKGFHEN
 RKNVYSSNDIPTSIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDID
 YKTSEVNQRVFSLDEVFEIANFN^ULNQSGITKFNTIIGGK^UFVNGENTKRKGINEYINLYS
 QQINDKTLK^UKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKS
 IKETLSLLFDDLKAQKLDLSKIYFKNDKSLDLSQQVFDDYSVIGTAVLEYITQQIAPKNLD

50

NPSKKEQELIAKKTEKAKYLSLETIKLALAEFNKHRDIDKQCRFEEILANFAAIPMIFDEIA
 QNKDNLAQISIKYQNQGKDLLQASAEDDVKAIKDLLDQTNLLHKLKIFHISQSEDKAN
 ILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKE
 PDNTAILFIKDDKYLLGVMNKKNNKIFDDKAIKENKGEQYKIVYKLLPGANKMLPKVFF
 SAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKD
 FGFRFSDTQRYNSIDFYREVENQGYKLTFFENISESYIDSVVNQGKLYLFQIYNKDFSAYS
 GRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNDNP
 KKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRG
 ERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKE
 MKEGYLSQVVHEIAKLVEIYNAIVVF~~AD~~LNFGFKRGRFKVEKQVYQKLEKMLIEKLNLYLV
 FKDNFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKY
 ESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKN
 HNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSK
 TGTELDYLISPVADVNGNFFDSRQAPKNMPQDAD~~ANG~~AYHIGLKGLMLLGRIKNNQEGK
 KLNLVIKNEEYFEFVQNRNN.

10

【 0 1 9 2 】

Francisella novicida Cpf1 D1255Aのアミノ酸配列は以下の通りである。(D917、E1006、およびA1255の変異位置には太字で下線を付している。)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
 EEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFN
 QNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIISFKGWTTYFKGFHEN
 RKNVYSSNDIPTSIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDID
 YKTSEVNQRVFSLDEVFEIANFNLYLNQSGITKFNTIIGGKRVNGENTKRKGINEYINLYS
 QQINDKTLKYYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKS
 IKETLSLLFDDLKAQKLDLSKIYFKNDKSLDLSQQVFDDYSVIGTAVLEYITQQIAPKNLD
 NPSKKEQELIAKKTEKAKYLSLETIKLALAEFNKHRDIDKQCRFEEILANFAAIPMIFDEIA
 QNKDNLAQISIKYQNQGKDLLQASAEDDVKAIKDLLDQTNLLHKLKIFHISQSEDKAN
 ILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKE
 PDNTAILFIKDDKYLLGVMNKKNNKIFDDKAIKENKGEQYKIVYKLLPGANKMLPKVFF
 SAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKD
 FGFRFSDTQRYNSIDFYREVENQGYKLTFFENISESYIDSVVNQGKLYLFQIYNKDFSAYS
 GRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNDNP
 KKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRG
 ERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKE
 MKEGYLSQVVHEIAKLVEIYNAIVVF~~ED~~LNFGFKRGRFKVEKQVYQKLEKMLIEKLNLYLV
 FKDNFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKY
 ESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKN
 HNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSK
 TGTELDYLISPVADVNGNFFDSRQAPKNMPQDA~~ANG~~AYHIGLKGLMLLGRIKNNQEGK
 KLNLVIKNEEYFEFVQNRNN.

20

30

40

【 0 1 9 3 】

Francisella novicida Cpf1 D917A/E1006Aのアミノ酸配列は以下の通りである。(A917、A1006、およびD1255には太字で下線を付している。)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
 EEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFN
 QNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIISFKGWTTYFKGFHEN
 RKNVYSSNDIPTSIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDID
 YKTSEVNQRVFSLDEVFEIANFNLYLNQSGITKFNTIIGGKRVNGENTKRKGINEYINLYS
 QQINDKTLKYYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKS
 IKETLSLLFDDLKAQKLDLSKIYFKNDKSLDLSQQVFDDYSVIGTAVLEYITQQIAPKNLD

50

NPSKKEQELIAKKTEKAKYLSLETIKLALAEFNKHRDIDKQCRFEEILANFAAIPMIFDEIA
 QNKDNLAQISIKYQNQGKDLLQASAEDDVKAIKDLLDQTNLLHKLKIFHISQSEDKAN
 ILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKE
 PDNTAILFIKDDKYLLGVMNKKNNKIFDDKAIKENKGEQYKIVYKLLPGANKMLPKVFF
 SAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKD
 FGFRFSDTQRYNSIDFYREVENQGYKLTFFENISESYIDSVVNQGKLYLFQIYNKDFSAYS
 GRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNDNP
 KKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIARG
 ERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKE
 MKEGYLSQVVHEIAKLVEIYNAIVVFADLNFGFKRGRFKVEKQVYQKLEKMLIEKLNLYLV
 FKNDFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKY
 ESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKN
 HNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSK
 TGTELDYLISPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGK
 KLNLVIKNEEYFEFVQNRNN.

10

【 0 1 9 4 】

Francisella novicida Cpf1 D917A/D1255Aのアミノ酸配列を以下の通りである。

(A917、E1006、およびA1255には太字で下線を付している。)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
 EEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFN
 QNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIISFKGWTTYFKGFHEN
 RKNVYSSNDIPTSIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDID
 YKTSEVNQRVFSLDEVFEIANFNLYLNQSGITKFNTIIGGKRVNGENTKRKGINEYINLYS
 QQINDKTLKYYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKS
 IKETLSLLFDDLKAQKLDLSKIYFKNDKSLDLSQQVFDDYSVIGTAVLEYITQQIAPKNLD
 NPSKKEQELIAKKTEKAKYLSLETIKLALAEFNKHRDIDKQCRFEEILANFAAIPMIFDEIA
 QNKDNLAQISIKYQNQGKDLLQASAEDDVKAIKDLLDQTNLLHKLKIFHISQSEDKAN
 ILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKE
 PDNTAILFIKDDKYLLGVMNKKNNKIFDDKAIKENKGEQYKIVYKLLPGANKMLPKVFF
 SAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKD
 FGFRFSDTQRYNSIDFYREVENQGYKLTFFENISESYIDSVVNQGKLYLFQIYNKDFSAYS
 GRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNDNP
 KKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIARG
 ERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKE
 MKEGYLSQVVHEIAKLVEIYNAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNLYLV
 FKNDFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKY
 ESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKN
 HNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSK
 TGTELDYLISPVADVNGNFFDSRQAPKNMPQDAAANGAYHIGLKGLMLLGRIKNNQEGK
 KLNLVIKNEEYFEFVQNRNN.

20

30

40

【 0 1 9 5 】

Francisella novicida Cpf1 E1006A/D1255Aのアミノ酸配列は以下の通りである。

(D917、A1006、およびA1255には太字で下線を付している。)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
 EEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFN
 QNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIISFKGWTTYFKGFHEN
 RKNVYSSNDIPTSIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDID
 YKTSEVNQRVFSLDEVFEIANFNLYLNQSGITKFNTIIGGKRVNGENTKRKGINEYINLYS
 QQINDKTLKYYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKS
 IKETLSLLFDDLKAQKLDLSKIYFKNDKSLDLSQQVFDDYSVIGTAVLEYITQQIAPKNLD

50

NPSKKEQELIAKKTEKAKYLSLETIKLAL EEFNKHRDIDKQCRFEEILANFAAIPMIFDEIA
 QNKDNLAQISIKYQNQGKDLLQASAEDDVKAIKDLLDQTNLLHKLKIFHISQSEDKAN
 ILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKE
 PDNTAILFIKDDKYLLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFF
 SAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKD
 FGFRFSDTQRYNSIDEFYREVENQGYKLT FENISESYIDSVVNQGKLYLFQIYNKDFSAYS
 KGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNDNP
 KKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRG
 ERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKE
 MKEGYLSQVVHEIAKLVIEYNAIVVFADLNF^{AD}FGFKRGRFKVEKQVYQKLEKMLIEKLNLYLV
 FKNDFDKTGGVLRAYQLTAPFETF^{AD}KKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKY
 ESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKN
 HNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSK
 TGTELDYLISPVADVNGNFFDSRQAPKNMPQDA^{AD}ANGAYHIGLKGLMLLGRIKNNQEGK
 KLNLVIKNEEYFEFVQNRNN.

10

【 0 1 9 6 】

Francisella novicida Cpf1 D917A/E1006A/D1255Aのアミノ酸配列は以下の通り
 である。(A917、A1006、およびA1255には太字で下線を付している。)

MSIQQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
 EEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFN
 QNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIISFKGWTTYFKGFHEN
 RKNVYSSNDIPTSIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDID
 YKTSEVNQRVFSLDEVFEIANFNLYLNQSGITKFNTIIGGKRVNGENTKRKGINEYINLYS
 QQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKS
 IKETLSLLFDDLKAQKLDLSKIYFKNDKSLDLSQQVFDDYSVIGTAVLEYITQQIAPKNLD
 NPSKKEQELIAKKTEKAKYLSLETIKLAL EEFNKHRDIDKQCRFEEILANFAAIPMIFDEIA
 QNKDNLAQISIKYQNQGKDLLQASAEDDVKAIKDLLDQTNLLHKLKIFHISQSEDKAN
 ILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKE
 PDNTAILFIKDDKYLLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFF
 SAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKD
 FGFRFSDTQRYNSIDEFYREVENQGYKLT FENISESYIDSVVNQGKLYLFQIYNKDFSAYS
 KGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNDNP
 KKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRG
 ERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKE
 MKEGYLSQVVHEIAKLVIEYNAIVVFADLNF^{AD}FGFKRGRFKVEKQVYQKLEKMLIEKLNLYLV
 FKNDFDKTGGVLRAYQLTAPFETF^{AD}KKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKY
 ESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKN
 HNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSK
 TGTELDYLISPVADVNGNFFDSRQAPKNMPQDA^{AD}ANGAYHIGLKGLMLLGRIKNNQEGK
 KLNLVIKNEEYFEFVQNRNN.

20

30

40

【 0 1 9 7 】

いくつかの実施形態では、バリエーションCasタンパク質は、spCas9、spCas9-VRQR、sp
 Cas9-VRER、xCas9 (sp)、saCas9、saCas9-KKH、spCas9-MQKSER、spCas9-LRK
 IQK、またはspCas9-LRVSQLであり得る。

【 0 1 9 8 】

例示的SaCas9のアミノ酸配列は以下の通りである：
 KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRH
 RIQRVVKLLFDYNLLTDHSELGINPYEARVKGLSQKLSSEEFSAALLHLAKRRGVHNVNE
 VEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSYVKEAKQLL
 KVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEKSPFGWKDIKEWYEMLMGHCTYFPEE

50

LRSVKYAYNADLYNALNDLNNLVITRDENEKLEYEYKFKQIIENVFKQKKKPTLKQIAKEIL
 VNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIEAELLDDQIAKILTIYQSSEDIQEE
 LTNLNSELQTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKV
 DLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIKKYGLPNDIIIELAREKNSKDAQMI
 NEMQKRNRQTNERIEEIIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFN
 YEVDHIIIPRSVSFDNSFNNKVLVKQEENSSKKGNRTPFQYLSSSDSKISYETFKKHILNLAK
 GKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLRLSYFRVNNLDVKV
 KSINGGFTSFLRRKWKFKKERNKGYKHHAEALIANADFIKFKWKLDKAKKVMENQM
 FEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKNRELINDTLYSTRK
 DDKGNTLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHPQTYQKLKLIMEQYGDEKN
 PLYKYYEETGNLYTKYSKKDNGPVIKKIKYYGNKLNLAHLDITDDYPNSRNKVVKLSLKP
 YRFDVYLDNGVYKFVTVKNLVDVIKKENYYEVNSKCYEEAKKLLKISNQAEFIASFYNNDLIK
 INGELYRVIGVNNDLLNRIEVMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDIL
 GNLYEVKSKKHPQIIKKG.

10

この配列において、太字で下線を付している残基N579は、（例えばA579に）変異されてSaCas9ニッカーゼを生じることができる。

【0199】

例示的SaCas9nのアミノ酸配列は以下の通りである：

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRH
 RIQRVKKLLFDYNLLTDHSELGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNE
 VEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSYVKEAKQLL
 KVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEKSPFGWKDIKEWYEMLMGHCTYFPEE
 LRSVKYAYNADLYNALNDLNNLVITRDENEKLEYEYKFKQIIENVFKQKKKPTLKQIAKEIL
 VNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIEAELLDDQIAKILTIYQSSEDIQEE
 LTNLNSELQTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKV
 DLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIKKYGLPNDIIIELAREKNSKDAQMI
 NEMQKRNRQTNERIEEIIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFN
 YEVDHIIIPRSVSFDNSFNNKVLVKQEEASSKKGNRTPFQYLSSSDSKISYETFKKHILNLAK
 GKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLRLSYFRVNNLDVKV
 KSINGGFTSFLRRKWKFKKERNKGYKHHAEALIANADFIKFKWKLDKAKKVMENQM
 FEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKNRELINDTLYSTRK
 DDKGNTLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHPQTYQKLKLIMEQYGDEKN
 PLYKYYEETGNLYTKYSKKDNGPVIKKIKYYGNKLNLAHLDITDDYPNSRNKVVKLSLKP
 YRFDVYLDNGVYKFVTVKNLVDVIKKENYYEVNSKCYEEAKKLLKISNQAEFIASFYNNDLIK
 INGELYRVIGVNNDLLNRIEVMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDIL
 GNLYEVKSKKHPQIIKKG.

20

30

この配列において、N579から変異されてSaCas9ニッカーゼを生じることができる残基A579には太字で下線を付している。

【0200】

例示的SaKKH Cas9のアミノ酸配列は以下の通りである：

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRH
 RIQRVKKLLFDYNLLTDHSELGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNE
 VEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSYVKEAKQLL
 KVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEKSPFGWKDIKEWYEMLMGHCTYFPEE
 LRSVKYAYNADLYNALNDLNNLVITRDENEKLEYEYKFKQIIENVFKQKKKPTLKQIAKEIL
 VNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIEAELLDDQIAKILTIYQSSEDIQEE
 LTNLNSELQTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKV
 DLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIKKYGLPNDIIIELAREKNSKDAQMI
 NEMQKRNRQTNERIEEIIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFN
 YEVDHIIIPRSVSFDNSFNNKVLVKQEEASSKKGNRTPFQYLSSSDSKISYETFKKHILNLAK

40

50

GKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLRLSYFRVNNLDVKV
 KSINGGFTSFLRRKWKFKKERNKGYKHHAAEDALIIANADFIFKEWKKLDKAKKVMENQM
 FEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKPKNRKLINDTLYSTRK
 DDKGNTLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHDHPQTYQKLKLIMEQYGDEKN
 PLYKYYEETGNYLTKYSSKKNPVIKIKIYYGNKLNALDITDDYPNSRNKVVKLSLKPYP
 RFDVYLDNGVYKFVTVKNLVDVIKKENYYEVNSKCYEEAKLKKISNQAEFIASFYKNDLIK
 INGELYRVIGVNNDLLNRIEVMIDITYREYLENMNDKRPPHIKIASKTQSIKKYSTDIL
 GNLYEVKSKKHPQIIKKG.

【0201】

N579から変異されてSaCas9ニッカーゼを生じることができる上記の残基A579には太
 字で下線を付している。E781、N967、およびR1014から変異されてSaKKH Cas9を生
 じることができる上記の残基K781、K967、およびH1014には斜体で下線を付している。

10

【0202】

塩基エディターのポリヌクレオチドプログラム可能ヌクレオチド結合ドメインは、それ
 自体が1つ以上のドメインを含むことができる。例えば、ポリヌクレオチドプログラム可
 能なヌクレオチド結合ドメインは、1つ以上のヌクレアーゼドメインを含むことができ
 る。ある態様において、ポリヌクレオチドプログラム可能ヌクレオチド結合ドメインのヌク
 レアーゼドメインは、エンドヌクレアーゼまたはエキソヌクレアーゼを含むことができ
 る。本明細書において、用語「エキソヌクレアーゼ」は、核酸(例えばRNAまたはDNA)を遊
 離末端から消化することができるタンパク質またはポリペプチドを指し、用語「エンドヌ
 クレアーゼ」は、核酸(例えばDNAまたはRNA)の内部領域を触媒(例えば劈開)すること
 ができるタンパク質またはポリペプチドを指す。ある態様において、エンドヌクレアーゼは
 、二本鎖核酸の一本鎖を切断することができる。ある態様において、エンドヌクレアーゼ
 は、二本鎖核酸分子の両方の鎖を切断することができる。ある態様において、ポリヌクレ
 オチドプログラム可能ヌクレオチド結合ドメインは、デオキシリボヌクレアーゼであり得
 る。ある態様において、ポリヌクレオチドプログラム可能ヌクレオチド結合ドメインは、
 リボヌクレアーゼであり得る。

20

【0203】

ある態様において、ポリヌクレオチドプログラム可能なヌクレオチド結合ドメインのヌ
 クレアーゼドメインは、標的ポリヌクレオチドの0本、1本または2本の鎖を切断すること
 ができる。ある場合には、ポリヌクレオチドプログラム可能なヌクレオチド結合ドメイン
 は、ニッカーゼドメインを含むことができる。本明細書において、用語「ニッカーゼ」は
 、二本鎖核酸分子(例えばDNA)中の二本鎖の一方の鎖のみを切断することができるヌク
 レアーゼドメインを含むポリヌクレオチドプログラム可能ヌクレオチド結合ドメインを指す
 。ある態様において、ニッカーゼは、活性なポリヌクレオチドプログラム可能ヌクレオチ
 ド結合ドメインに一つ以上の突然変異を導入することによって、ポリヌクレオチドプログ
 ラム可能ヌクレオチド結合ドメインの完全に触媒活性な(例えば天然の)形態から誘導す
 ることができる。例えば、ポリヌクレオチドプログラム可能なヌクレオチド結合ドメインが
 Cas9に由来するニッカーゼドメインを含む場合、Cas9に由来するニッカーゼドメインは
 、D10A突然変異および位置840におけるヒスチジンを含むことができる。そのような場
 合、残基H 840は触媒活性を保持し、それによって核酸二本鎖の一本鎖を切断すること
 ができる。別の例において、Cas9由来ニッカーゼドメインは、H840A突然変異を含むこと
 ができ、一方、位置10におけるアミノ酸残基は、Dのままである。いくつかの実施形態に
 おいて、ニッカーゼは、ニッカーゼ活性に必要ではないヌクレアーゼドメインの全部また
 は一部を除去することによって、ポリヌクレオチドプログラム可能ヌクレオチド結合ドメ
 インの完全に触媒活性な(例えば天然の)形態から誘導することができる。例えば、ポリヌ
 クレオチドプログラム可能なヌクレオチド結合ドメインがCas9に由来するニッカーゼドメ
 インを含む場合、Cas9に由来するニッカーゼドメインは、RuvCドメインまたはHNHドメ
 インの全部または一部の欠失を含むことができる。

30

40

【0204】

50

したがって、ニッカーゼドメインを含むポリヌクレオチドプログラム可能なヌクレオチド結合ドメインを含む塩基エディターは、特定のポリヌクレオチド標的配列(例えば結合したガイド核酸の相補的配列によって決定される)において一本鎖DNA切断(ニック)を生成することができる。ある態様において、ニッカーゼドメイン(例えばCas9由来のニッカーゼドメイン)を含む塩基エディターによって切断される核酸二本鎖標的ポリヌクレオチド配列の鎖は、塩基エディターによって編集されない鎖である(すなわち、塩基エディターによって切断される鎖は、編集される塩基を含む鎖とは反対の鎖である)。他の態様において、ニッカーゼドメイン(例えばCas9由来のニッカーゼドメイン)を含む塩基エディターは、編集のために標的とされるDNA分子の鎖を切断することができる。このような場合、非標的鎖は切断されない。

10

【0205】

触媒的に死んだ(すなわち標的ポリヌクレオチド配列を切断することができない)ポリヌクレオチドプログラム可能なヌクレオチド結合ドメインを含む塩基エディターも本明細書中に提供される。本明細書において、用語「触媒的に死んだ」および「死んだヌクレアーゼ」は、核酸の鎖を切断することができない結果となる一つ以上の突然変異および/または欠失を有するポリヌクレオチドプログラム可能なヌクレオチド結合ドメインを指すために交換可能に使用される。いくつかの実施形態において、触媒的に死んだポリヌクレオチドプログラム可能なヌクレオチド結合ドメイン塩基エディターは、一つ以上のヌクレアーゼドメインにおける特定の点突然変異の結果としてヌクレアーゼ活性を欠くことができる。例えば、Cas9ドメインを含む塩基エディターの場合、Cas9は、D10A突然変異およびH840A突然変異の両方を含むことができる。このような変異は両方のヌクレアーゼドメインを不活性化し、その結果ヌクレアーゼ活性を失う。他の実施形態において、触媒的に死んだポリヌクレオチドプログラム可能なヌクレオチド結合ドメインは、触媒ドメイン(例えばRuvC1および/またはHNHドメイン)の全部又は一部の一つ以上の欠失を含むことができる。さらなる態様において、触媒的に死んだポリヌクレオチドプログラム可能なヌクレオチド結合ドメインは、点突然変異(例えばD10AまたはH840A)ならびにヌクレアーゼドメインの全部または一部の欠失を含む。

20

【0206】

また、本明細書では、ポリヌクレオチドプログラム可能なヌクレオチド結合ドメインの以前に機能していたバージョンから、触媒的に死んだポリヌクレオチドプログラム可能なヌクレオチド結合ドメインを生成することができる突然変異も企図される。例えば、触媒的に死んだCas9(「dCas9」)の場合、D10AおよびH840A以外の突然変異を有してヌクレアーゼ不活性化Cas9をもたらすバリエーションが提供される。このような突然変異は、例えば、D10およびH840における他のアミノ酸置換、またはCas9のヌクレアーゼドメイン内の他の置換(例えば、HNHヌクレアーゼサブドメインおよび/またはRuvC1サブドメインにおける置換)を含む。さらなる適切なヌクレアーゼ不活性化dCas9ドメインは、本開示および当該分野における知識に基づいて当業者に明らかとなり得、本開示の範囲内である。このようなさらなる例示的な適切なヌクレアーゼ不活性化Cas9ドメインには、限定されるものではないが、D10A/H840A、D10A/D839A/H840A、およびD10A/D839A/H840A/N863A突然変異体ドメインが含まれる(例えば、Prashant et al., CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nature Biotechnology. 2013; 31(9): 833-838を参照されたい(その全内容は参照により本明細書に組み込まれる))。いくつかの実施形態において、dCas9ドメインは、本明細書に提供されるdCas9ドメインのいずれかに対して少なくとも60%、少なくとも65%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、または少なくとも99.5%の同一性を有するアミノ酸配列を含む。いくつかの実施形態において、Cas9ドメインは、本明細書に記載されるアミノ酸配列のいずれかと比較して1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35

30

40

50

、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50またはそれ以上の突然変異を有するアミノ酸配列を含む。いくつかの実施形態において、Cas9ドメインは、本明細書に記載されたアミノ酸配列のいずれかと比較して、少なくとも10、少なくとも15、少なくとも20、少なくとも30、少なくとも40、少なくとも50、少なくとも60、少なくとも70、少なくとも80、少なくとも90、少なくとも100、少なくとも150、少なくとも200、少なくとも250、少なくとも300、少なくとも350、少なくとも400、少なくとも500、少なくとも600、少なくとも700、少なくとも800、少なくとも900、少なくとも1000、少なくとも1100、または少なくとも1200の同一の連続したアミノ酸残基を有するアミノ酸配列を含む。

【0207】

塩基エディターに組み込むことができるポリヌクレオチドプログラム可能ヌクレオチド結合ドメインの非限定的な例としては、CRISPRタンパク質由来ドメイン、制限ヌクレアーゼ、メガヌクレアーゼ、TALヌクレアーゼ(TALEN)、およびジンクフィンガーヌクレアーゼ(ZFN)が挙げられる。いくつかの場合において、塩基エディターは、核酸のCRISPR(すなわちClustered Regularly Interspaced Short Palindromic Repeats)媒介修飾の際に、結合ガイド核酸を介して核酸配列に結合することができる天然もしくは修飾されたタンパク質またはその一部を含むポリヌクレオチドプログラム可能ヌクレオチド結合ドメインを含む。そのようなタンパク質は、本明細書において「CRISPRタンパク質」と呼ばれる。従って、本明細書に開示されているのは、CRISPRタンパク質の全部または一部を含むポリヌクレオチドプログラム可能ヌクレオチド結合ドメインを含む塩基エディター(すなわち、CRISPRタンパク質の全部または一部をドメインとして含む塩基エディター(それは塩基エディターの「CRISPRタンパク質由来ドメイン」とも呼ばれる))である。塩基エディターに組み込まれたCRISPRタンパク質由来ドメインは、野生型または天然型のCRISPRタンパク質と比較して改変され得る。例えば、以下に記載するように、CRISPRタンパク質由来ドメインは、野生型または天然型のCRISPRタンパク質と比較して、1以上の突然変異、挿入、欠失、再配列および/または組換えを含み得る。

【0208】

いくつかの実施形態において、塩基エディターに組み込まれたCRISPRタンパク質由来ドメインは、結合ガイド核酸と組み合わせられた場合に標的ポリヌクレオチドに結合することができるエンドヌクレアーゼ(例えばデオキシリボヌクレアーゼまたはリボヌクレアーゼ)である。いくつかの実施形態において、塩基エディターに組み込まれたCRISPRタンパク質由来ドメインは、結合ガイド核酸と組み合わせられた場合に標的ポリヌクレオチドに結合することができるニッカーゼである。いくつかの実施形態において、塩基エディターに組み込まれたCRISPRタンパク質由来ドメインは、結合ガイド核酸と組み合わせられた場合に標的ポリヌクレオチドに結合することができる触媒的に死んだドメインである。或る実施態様では、塩基エディターのCRISPRタンパク質由来ドメインに結合する標的ポリヌクレオチドはDNAであり、或る実施態様では、塩基エディターのCRISPRタンパク質由来ドメインに結合する標的ポリヌクレオチドはRNAである。

【0209】

ある態様において、塩基エディターのCRISPRタンパク質由来ドメインは、*Corynebacterium ulcerans* (NCBI Refs: NC_015683.1, NC_017317.1); *Corynebacterium diphtheria* (NCBI Refs: NC_016782.1, NC_016786.1); *Spiroplasma syrphidicola* (NCBI Ref: NC_021284.1); *Prevotella intermedia* (NCBI Ref: NC_017861.1); *Spiroplasma taiwanense* (NCBI Ref: NC_021846.1); *Streptococcus iniae* (NCBI Ref: NC_021314.1); *Belliella baltica* (NCBI Ref: NC_018010.1); *Psychroflexus torquis* (NCBI Ref: NC_018721.1); *Streptococcus thermophilus* (NCBI Ref: YP_820832.1); *Listeria innocua* (NCBI Ref: NP_472073.1); *Campylobacter jejuni* (NCBI Ref: YP_002344900.1); *Neisseria meningitidis* (NCBI Ref: YP_002342100.1), *Streptococcus pyogenes*, または *Staphylococcus aureus* 由来のCas9の全部または一部を含むことができる。

10

20

30

40

50

【0210】

ある態様において、Cas9ドメインは、Staphylococcus aureus由来のCas9ドメイン(SaCas9)である。ある態様において、SaCas9ドメインは、ヌクレアーゼ活性SaCas9、ヌクレアーゼ不活性SaCas9 (SaCas9d)、またはSaCas9ニッカーゼ (SaCas9n) である。いくつかの実施形態において、SaCas9は、N579A突然変異、または本明細書に提供されるアミノ酸配列のいずれかにおける対応する突然変異を含む。

【0211】

いくつかの実施形態において、SaCas9ドメイン、SaCas9dドメイン、またはSaCas9nドメインは、非標準PAMを有する核酸配列に結合することができ、いくつかの実施形態において、SaCas9ドメイン、SaCas9dドメイン、またはSaCas9nドメインは、NNGRRTまたはNNNRRT PAM配列を有する核酸配列に結合することができる。いくつかの実施形態において、SaCas9ドメインは、E781X、N967X、およびR1014X突然変異の1以上、または本明細書に提供されるアミノ酸配列のいずれかにおける対応する突然変異を含み、ここでXは任意のアミノ酸である。いくつかの実施形態において、SaCas9ドメインは、E781K、N967K、およびR1014H突然変異のうちの一つ以上、または本明細書に提供されるアミノ酸配列のいずれかにおける一つ以上の対応する突然変異を含む。いくつかの実施形態において、SaCas9ドメインは、E781K、N967K、またはR1014H突然変異、または本明細書に提供されるアミノ酸配列のいずれかにおける対応する突然変異を含む。

【0212】

塩基エディターは、高忠実度Cas9であるCas9の全部又は一部に由来するドメインを含むことができる。いくつかの実施形態において、塩基エディターの高忠実度Cas9ドメインは、対応する野生型Cas9ドメインと比較して、Cas9ドメインとDNAの糖-リン酸骨格との間の静電相互作用を減少させる一つ以上の突然変異を含む、操作されたCas9ドメインである。DNAの糖-リン酸骨格との静電相互作用を減少させた高忠実度Cas9ドメインは、より少ないオフターゲット効果を有することができる。いくつかの実施形態において、Cas9ドメイン(例えば、野生型Cas9ドメイン)は、Cas9ドメインとDNAの糖-リン酸主鎖との間の結合を減少させる一つ以上の突然変異を含む。いくつかの実施形態において、Cas9ドメインは、Cas9ドメインとDNAの糖-リン酸主鎖との間の結合を少なくとも1%、少なくとも2%、少なくとも3%、少なくとも4%、少なくとも5%、少なくとも10%、少なくとも15%、少なくとも20%、少なくとも25%、少なくとも30%、少なくとも35%、少なくとも40%、少なくとも45%、少なくとも50%、少なくとも55%、少なくとも60%、少なくとも65%、少なくとも70%以上減少させる一つ以上の突然変異を含む。

【0213】

[ガイドポリヌクレオチド]

本明細書中で使用される、用語「ガイドポリヌクレオチド」とは、標的配列に特異的であり得かつポリヌクレオチドプログラム可能ヌクレオチド結合ドメインタンパク質(例えばCas9またはCpf1)と複合体を形成し得るポリヌクレオチドをいう。一実施形態では、ガイドポリヌクレオチドはガイドRNAである。本明細書において使用される用語「ガイドRNA (gRNA)」およびその文法的等価物は、標的DNAに対して特異的であり得かつCasタンパク質と複合体を形成し得るRNAを表し得る。RNA/Cas複合体は、Casタンパク質を標的DNAに「ガイド」するのを補助することができる。Cas9/crRNA/tracrRNAは、スペーサーに相補的な線状または環状dsDNA標的をエンドヌクレアーゼ的に切断する。crRNAに相補的でない標的鎖が最初にエンドヌクレアーゼで切断され、次いでエキソヌクレアーゼ的に3' - 5' 方向にトリムされる。自然界では、DNA結合と切断にはタンパク質と両方のRNAが通常必要とされる。しかしながら、crRNAおよびtracrRNAの両方の側面を単一のRNA種に組み込むように、単一ガイドRNA(「sgRNA」、または単に「gRNA」)を作製することができる。例えば、Jinek M. et al., Science 337:816-821(2012)を参照されたい(その内容全体が参照により本明細書に組み入れられる)。Cas9は、CRISPR反復配列中の短いモチーフ(PAMまたはプロトスペーサー隣接モチーフ)を認識して、自己と非自己を区別するのを助ける。

10

20

30

40

50

【0214】

ある態様において、ガイドポリヌクレオチドは、少なくとも一つの単一ガイドRNA（「sgRNA」または「gRNA」）である。ある態様において、ガイドポリヌクレオチドは、少なくとも一つのtracrRNAであり、ある態様において、ガイドポリヌクレオチドは、ポリヌクレオチドプログラム可能なDNA結合ドメイン(例:Cas9またはCpf1)を標的ヌクレオチド配列に導くためにPAM配列を必要としない。

【0215】

本願明細書に開示される塩基エディターのポリヌクレオチドプログラム可能ヌクレオチド結合ドメイン(例えばCRISPR由来ドメイン)は、ガイドポリヌクレオチドと会合することによって標的ポリヌクレオチド配列を認識することができる。ガイドポリヌクレオチド(例えばgRNA)は、典型的には一本鎖であり、ポリヌクレオチドの標的配列に部位特異的に結合(すなわち相補的塩基対形成を介して)するようにプログラムすることができ、それによって、ガイド核酸を伴った塩基エディターを標的配列に導く。ガイドポリヌクレオチドはDNAであり得る。ガイドポリヌクレオチドはRNAであり得る。場合によっては、ガイドポリヌクレオチドは天然ヌクレオチド(例えばアデノシン)を含む。場合によっては、ガイドポリヌクレオチドは、非天然(または不自然)ヌクレオチド(例えばペプチド核酸またはヌクレオチドアナログ)を含む。ある場合には、ガイド核酸配列の標的化領域は、長さが少なくとも15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、または30ヌクレオチドであり得る。ガイド核酸の標的化領域は、長さが10~30ヌクレオチドの間、または長さが15~25ヌクレオチドの間、または長さが15~20ヌクレオチドの間であり得る。

【0216】

いくつかの実施形態において、ガイドポリヌクレオチドは、例えば相補的塩基対形成を介して互いに相互作用できる、二つ以上の個別のポリヌクレオチドを含む(例えば二重ガイドポリヌクレオチド)。例えば、ガイドポリヌクレオチドは、CRISPR RNA (crRNA) およびトランス活性化CRISPR RNA (tracrRNA) を含むことができる。例えば、ガイドポリヌクレオチドは、1以上のトランス活性化CRISPR RNA (tracrRNA) を含むことができる。

【0217】

II型CRISPRシステムにおいて、CRISPRタンパク質(例Cas9)による核酸の標的化は、典型的には、標的配列を認識する配列を含む第一のRNA分子(crRNA)と、ガイドRNA-CRISPRタンパク質複合体を安定化させる足場領域を形成する反復配列を含む第二のRNA分子(tracrRNA)との間の相補的塩基対形成を必要とする。このような二重ガイドRNAシステムは、本明細書に開示された塩基エディターを標的ポリヌクレオチド配列に導くためのガイドポリヌクレオチドとして利用され得る。

【0218】

いくつかの実施形態において、本明細書で提供される塩基エディターは、単一ガイドポリヌクレオチド(例えばgRNA)を利用する。いくつかの実施形態において、本明細書で提供される塩基エディターは、デュアル(二重)ガイドポリヌクレオチド(例えばデュアルgRNA)を利用する。いくつかの実施形態において、本明細書で提供される塩基エディターは、一つ以上のガイドポリヌクレオチド(例えば複数のgRNA)を利用する。いくつかの実施形態において、単一ガイドポリヌクレオチドが、本明細書に記載される異なる塩基エディターのために利用される。例えば、単一ガイドポリヌクレオチドを、シチジン塩基エディターおよびアデノシン塩基エディターのために利用することができる。

【0219】

他の実施形態では、ガイドポリヌクレオチドは、核酸のポリヌクレオチド標的化部分および核酸の足場部分の両方を単一分子(すなわち単一分子ガイド核酸)中に含むことができる。例えば、単一分子ガイドポリヌクレオチドは、単一ガイドRNA (sgRNAまたはgRNA)であり得る。本明細書において、「ガイドポリヌクレオチド配列」という用語は、塩基エディターと相互作用し標的ポリヌクレオチド配列に塩基エディターを導くことができる任意の単一、二重または複数分子核酸を企図する。

10

20

30

40

50

【0220】

典型的には、ガイドポリヌクレオチド(例えばcrRNA/trRNA複合体またはgRNA)は、標的ポリヌクレオチド配列を認識しそれに結合することができる配列を含む「ポリヌクレオチド標的セグメント」と、塩基エディターのポリヌクレオチドプログラム可能ヌクレオチド結合ドメイン成分内でガイドポリヌクレオチドを安定化させる「タンパク質結合セグメント」とを含む。或る実施態様では、ガイドポリヌクレオチドのポリヌクレオチド標的セグメントは、DNAポリヌクレオチドを認識してそれに結合し、それによりDNA中の塩基の編集を促進する。他の実施態様では、ガイドポリヌクレオチドのポリヌクレオチド標的セグメントは、RNAポリヌクレオチドを認識しそれに結合し、それによりRNA中の塩基の編集を促進する。ここで、「セグメント」とは、分子の一部分または領域、例えばガイドポリヌクレオチド中の連続したヌクレオチドのストレッチをいう。また、セグメントは、複合体の領域/セクションも表し得、従ってセグメントは2つ以上の分子の領域を含むことができる。例えば、ガイドポリヌクレオチドが複数の核酸分子を含む場合、タンパク質結合セグメントは、例えば相補性の領域に沿ってハイブリダイズした複数の別個の分子の全てまたは一部を含むことができる。ある態様において、二つの別個の分子を含むDNA標的化RNAのタンパク質結合セグメントは、(i) 長さが100塩基対である第一のRNA分子の塩基対40~75; および(ii) 長さが50塩基対である第二のRNA分子の10~25塩基対を含み得る。「セグメント」の定義は、特定の文脈において特に定義されない限り、全塩基対のうち特定の数に限定されず、所与のRNA分子からの塩基対の特定の数に限定されず、複合体内の別個の分子の特定の数に限定されず、任意の全長のRNA分子の領域を含み得、他の分子に対する相補性を有する領域を含み得る。

10

20

【0221】

ガイドRNAまたはガイドポリヌクレオチドは、2つ以上のRNA、例えばCRISPR RNA (crRNA) およびトランス活性化crRNA (tracrRNA) を含むことができる。ガイドRNAまたはガイドポリヌクレオチドは、時に、crRNAおよびtracrRNAの一部(例えば機能的部分)の融合によって形成される単鎖RNA、あるいは単一ガイドRNA (sgRNA) を含み得る。ガイドRNAまたはガイドポリヌクレオチドは、crRNAおよびtracrRNAを含む二重RNAであってもよい。さらに、crRNAは標的DNAとハイブリダイズし得る。

【0222】

上述のように、ガイドRNAまたはガイドポリヌクレオチドは、発現産物であり得る。例えば、ガイドRNAをコードするDNAは、ガイドRNAをコードする配列を含むベクターであり得る。ガイドRNAまたはガイドポリヌクレオチドは、単離されたガイドRNA、またはガイドRNAをコードする配列およびプロモーターを含むプラスミドDNAを細胞にトランスフェクトすることによって、細胞に導入することができる。ガイドRNAまたはガイドポリヌクレオチドは、ウイルス媒介遺伝子送達の使用など、他の方法で細胞に導入することもできる。

30

【0223】

ガイドRNAまたはガイドポリヌクレオチドを単離することができる。例えば、ガイドRNAは、単離されたRNAの形態で細胞または生物にトランスフェクトされ得る。ガイドRNAは、当技術分野で公知の任意のインビトロ転写系を用いたインビトロ転写によって調製することができる。ガイドRNAは、ガイドRNAをコードする配列を含むプラスミドの形態ではなく、単離されたRNAの形態で細胞に移入され得る。

40

【0224】

ガイドRNAまたはガイドポリヌクレオチドは、以下の3つの領域を含み得る：染色体配列中の標的部位に相補的であり得る5'末端における第1の領域、ステムループ構造を形成し得る第2の内部領域、および一本鎖であり得る第3の3'領域。各ガイドRNAが融合タンパク質を特異的な標的部位にガイドするように、各ガイドRNAの第1の領域を異ならせることもできる。さらに、各ガイドRNAの第2および第3の領域は、すべてのガイドRNAにおいて同一であり得る。

【0225】

50

ガイドRNAまたはガイドポリヌクレオチドの第1の領域は、そのガイドRNAの第1の領域が標的部位と塩基対を形成できるように、染色体配列の標的部位における配列に相補的であり得る。場合によっては、ガイドRNAの第一の領域は、約10ヌクレオチド~25ヌクレオチド(すなわち、10ヌクレオチドからヌクレオチドまで;または約10ヌクレオチドから約25ヌクレオチドまで;または10ヌクレオチドから約25ヌクレオチドまで;または約10ヌクレオチドから25ヌクレオチドまで)またはそれ以上を含むことができる。例えば、ガイドRNAの第一の領域と染色体配列中の標的部位との間の塩基対形成の領域は、長さが約10、11、12、13、14、15、16、17、18、19、20、22、23、24、25またはそれ以上のヌクレオチドであり得るか、またはおよそその長さであり得る。ときに、ガイドRNAの第一の領域は、長さが19、20、もしくは21ヌクレオチド、または約19、20、もしくは21ヌクレオチドであり得る。

10

【0226】

ガイドRNAまたはガイドポリヌクレオチドは、二次構造を形成する第2の領域を含むこともできる。例えば、ガイドRNAによって形成される二次構造は、ステム(またはヘアピン)およびループを含むことができる。ループとステムの長さはさまざまであり得る。例えば、ループの長さは(約)3から10ヌクレオチドの範囲であり得、ステムの長さは(約)6から20塩基対の範囲であり得る。ステムは、1から10または約10ヌクレオチドの一つ以上のバルジを含むことができる。第二の領域の全長は、約16から60ヌクレオチド長の範囲であり得る。例えば、ループは長さが(約)4ヌクレオチドであり得、ステムは(約)12塩基対であり得る。

20

【0227】

ガイドRNAまたはガイドポリヌクレオチドは、本質的に一本鎖状態であり得る3'末端の第3の領域も含むこともできる。たとえば、第3の領域は、対象とする細胞のいずれの染色体配列とも相補的でないこともあれば、ガイドRNAの残りの部分と相補的でないこともある。さらに第3の領域の長さはさまざまであり得る。第3の領域は、長さが約4ヌクレオチド以上であり得る。例えば、第3の領域の長さは、(約)5から60ヌクレオチド長の範囲であり得る。

【0228】

ガイドRNAまたはガイドポリヌクレオチドは、遺伝子標的の任意のエクソンまたはイントロンを標的とすることができる。場合によっては、ガイドは遺伝子のエキソン1または2を標的にすることができる;他の場合には、ガイドは遺伝子のエキソン3または4を標的にすることができる。組成物は、全てが同じエクソンを標的とする複数のガイドRNA、または場合によっては、異なるエクソンを標的とする複数のガイドRNAを含むことができる。遺伝子のエキソンとイントロンが標的とされ得る。

30

【0229】

ガイドRNAまたはガイドポリヌクレオチドは、(約)20ヌクレオチドの核酸配列を標的とし得る。標的核酸は、(約)20ヌクレオチド未満であり得る。標的核酸は、長さが少なくとも(約)5、10、15、16、17、18、19、20、21、22、23、24、25、30、または1~100ヌクレオチドの間の任意の長さであり得る。標的核酸は、長さが多くとも約5、10、15、16、17、18、19、20、21、22、23、24、25、30、40、50ヌクレオチド、または1~100ヌクレオチドの間の任意の長さであり得る。標的核酸配列は、PAMの最初のヌクレオチドの5'直近の(約)20塩基であり得る。ガイドRNAは、核酸配列を標的化し得る。標的核酸は、少なくとも(約)1~10、1~20、1~30、1~40、1~50、1~60、1~70、1~80、1~90、または1~100ヌクレオチドであり得る。

40

【0230】

ガイドポリヌクレオチド、例えば、ガイドRNAは、別の核酸、例えば、細胞のゲノム中の標的核酸またはプロトスペーサーにハイブリダイズし得る核酸を指すことができる。ガイドポリヌクレオチドはRNAであり得る。ガイドポリヌクレオチドはDNAであり得る。ガイドポリヌクレオチドは、核酸の配列に部位特異的に結合するようにプログラムまたは設計することができる。ガイドポリヌクレオチドは、一ポリヌクレオチド鎖を含み得、単一

50

ガイドポリヌクレオチドと呼ばれ得る。ガイドポリヌクレオチドは、2つのポリヌクレオチド鎖を含み得、二重ガイドポリヌクレオチドと呼ばれ得る。ガイドRNAはRNA分子として細胞や胚に導入され得る。例えば、RNA分子は、インビトロで転写され得、および/または化学的に合成され得る。RNAは、合成DNA分子、例えばgBlocks（登録商標）遺伝子断片から転写され得る。ガイドRNAはRNA分子として細胞や胚に導入され得る。ガイドRNAはまた、非RNA核酸分子、例えばDNA分子の形態で細胞または胚に導入され得る。例えば、ガイドRNAをコードするDNAが、対象の細胞または胚におけるガイドRNAの発現のためにプロモーター制御配列に作動可能に連結され得る。RNAコード配列は、RNAポリメラーゼIII (Pol III)によって認識されるプロモーター配列に作動可能に連結され得る。ガイドRNAを発現するために使用することができるプラスミドベクターは、px330ベクターおよびpx333ベクターを含むが、これらに限定されない。場合によっては、プラスミドベクター(例えばpx333ベクター)は、少なくとも二つのガイドRNAをコードするDNA配列を含むことができる。

10

【0231】

ガイドポリヌクレオチド(例えばガイドRNA)およびターゲティング配列を選択、設計および検証するための方法は、本明細書中に記載され、当業者に知られている。例えば、核酸塩基エディターシステムにおけるデアミナーゼドメイン(例えばAIDドメイン)の潜在的基質混合性(promiscuity)の影響を最小限にするために、意図せず脱アミノ化の標的となり得る残基(例えば、標的核酸遺伝子座内のssDNA上に潜在的に位置し得るオフターゲットC残基)の数を最小限にすることができる。さらに、ソフトウェアツールを使用して、標的核酸配列に対応するgRNAを最適化することができる。例えば、ゲノム全体の総オフターゲット活性を最小限化することができる。例えば、*S. pyogenes* Cas9を用いた標的化ドメイン選択肢の各可能性について、(例えばNAGまたはNGGなどの選択されたPAMに先行する)ある数(例えば1、2、3、4、5、6、7、8、9、または10)までのミスマッチ塩基対を含むすべてのオフターゲット配列をゲノムにわたって同定することができる。標的部位に相補的なgRNAの第一の領域を同定することができ、すべての第一の領域(例:crRNA)をその総予測オフターゲットスコアに従ってランク付けすることができ;上位にランクされた標的化ドメインは、最大のオンターゲット活性と最小のオフターゲット活性を持つ可能性が高いものを表す。候補標的化gRNAは、当技術分野で公知の方法および/または本明細書に記載の方法を用いて機能的に評価することができる。

20

30

【0232】

非限定的な例として、Cas9と共に使用するためのガイドRNAのcrRNAにおける標的DNAハイブリダイズ配列は、DNA配列探索アルゴリズムを用いて同定することができる。gRNA設計は、Bae S., Park J., & Kim J.-S. Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* 30, 1473-1475 (2014)に記載されているような公開ツールcas-offinderに基づくカスタムgRNA設計ソフトウェアを用いて実施することができる。このソフトウェアは、ガイドのゲノム全体的オフターゲット傾向を計算した後にガイドにスコアを付ける。通常、完全マッチから7個のミスマッチまでの範囲の一致が、17から24までの長さの範囲のガイドについて考慮される。いったんオフターゲットサイトが計算的に決定されると、各ガイドについて集約スコアが計算され、ウェブインターフェースを使用した表形式の出力に要約される。PAM配列に隣接する潜在的な標的部位を同定することに加えて、ソフトウェアは、選択された標的部位から1、2、3または3超のヌクレオチドだけ異なるすべてのPAM隣接配列も同定する。標的核酸配列、例えば標的遺伝子についてのゲノムDNA配列を得て、反復エレメントを、公に入手可能なツール、例えば、RepeatMaskerプログラムを用いてスクリーニングすることができる。RepeatMaskerは、入力されたDNA配列から、反復エレメントや低複雑性領域を探し出す。所与のクエリー配列に存在する反復の詳細な注釈が出力となる。

40

【0233】

同定後、ガイドRNA(例えばcrRNA)の第1の領域が、標的部位へのそれらの距離、そ

50

これらの直交性 (orthogonality)、および関連するPAM配列との密接な一致のための5'ヌクレオチドの存在(例えば、関連するPAM(例えば、化膿レンサ球菌についてのNGG PAM、黄色ブドウ球菌についてのNNGRRTまたはNNGRRV PAM)を含むヒトゲノムにおける密接な一致の同定に基づく5'G)に基づいて、階層にランク付けされ得る。本明細書において使用される場合、直交性とは、標的配列に対して最小数のミスマッチを含む、ヒトゲノムにおける配列の数を表す。「高レベルの直交性」または「良好な直交性」は、例えば、意図された標的以外にヒトゲノムにおいて同一の配列を有さず、標的配列において一または二のミスマッチを含有する配列も有さない20マー標的化ドメインを指し得る。良好な直交性を有する標的化ドメインは、オフターゲットDNA切断を最小化するように選択され得る。

10

【0234】

いくつかの実施形態において、塩基編集活性を検出すること、および候補ガイドポリヌクレオチドを試験することのためにレポーターシステムが使用され得る。いくつかの態様において、レポーターシステムは、塩基編集活性がレポーター遺伝子の発現をもたらす、レポーター遺伝子ベースのアッセイを含むことができる。例えば、レポーターシステムは、不活性化開始コドン、例えば、鋳型鎖上の3'-TAC-5'から3'-CAC-5'への突然変異を含むレポーター遺伝子を含み得る。標的Cの脱アミノ化に成功すると、対応するmRNAは5'-GUG-3'ではなく5'-AUG-3'として転写され、レポーター遺伝子の翻訳を可能にする。適切なレポーター遺伝子は、当業者には明らかであろう。レポーター遺伝子の非限定的な例としては、緑色蛍光タンパク質 (GFP)、赤色蛍光タンパク質 (RFP)、ルシフェラーゼ、分泌型アルカリホスファターゼ (SEAP)、または、発現が検出可能であり当業者には明らかである他の任意の遺伝子をコードする遺伝子が挙げられる。レポーター系を用いて、多くの異なるgRNAを試験することができ、例えば、標的DNA配列に関してどの残基を、それぞれの対応デアミナーゼが標的とするかを決定することができる。非鋳型鎖を標的とするsgRNAも、特定の塩基編集タンパク質(例えばCas9デアミナーゼ融合タンパク質)のオフターゲット効果を評価するために試験することができる。いくつかの実施形態において、そのようなgRNAは、変異開始コドンがgRNAと塩基対を形成しないように設計することができる。ガイドポリヌクレオチドは、標準リボヌクレオチド、修飾リボヌクレオチド(例えば、プソイドウリジン)、リボヌクレオチド異性体、および/またはリボヌクレオチド類似体を含むことができる。いくつかの実施形態において、ガイドポリヌクレオチドは、少なくとも1つの検出可能な標識を含むことができる。検出可能な標識は、フルオロフォア(例えば、FAM、TMR、Cy3、Cy5、テキサスレッド、オレゴングリーン、Alexa Fluor、Haloタグ、または適切な蛍光色素)、検出タグ(例えば、ビオチン、ジゴキシゲニン等)、量子ドット、または金粒子であり得る。

20

30

【0235】

ガイドポリヌクレオチドは、化学的に合成され得るか、酵素的に合成され得るか、またはそれらの組み合わせであり得る。例えば、ガイドRNAは、標準的なホスホロアミダイトベースの固相合成法を用いて合成することができる。あるいは、ガイドRNAをコードするDNAを、ファージRNAポリメラーゼにより認識されるプロモーター制御配列に作動可能に連結することによって、ガイドRNAをインビトロで合成することができる。適切なファージプロモーター配列の例は、T7、T3、SP6プロモーター配列、またはそれらのバリエーションを含む。ガイドRNAが二つの別々の分子(例えば、crRNAとtracrRNA)を含む実施形態において、crRNAは化学的に合成することができ、tracrRNAは酵素的に合成することができる。

40

【0236】

ある態様において、塩基エディターシステムは、複数のガイドポリヌクレオチド、例えばgRNAを含み得る。例えば、gRNAは、塩基エディターシステムに含まれる一つ以上の標的遺伝子座(例えば少なくとも1 gRNA、少なくとも2 gRNA、少なくとも5 gRNA、少なくとも10 gRNA、少なくとも20 gRNA、少なくとも30 gRNA、少なくとも50 gRNA)を標的とすることができる。前記複数のgRNA配列はタンデムに配置され得、好ましくは直

50

接反復によって分離される。

【0237】

ガイドRNAまたはガイドポリヌクレオチドをコードするDNA配列がベクターの一部であってもよい。さらに、ベクターは、追加の発現制御配列(例えばエンハンサー配列、Kozak配列、ポリアデニル化配列、転写終結配列など)、選択可能なマーカー配列(例えばGFPまたはピューロマイシンなどの抗生物質耐性遺伝子)、複製起点などを含むことができる。ガイドRNAをコードするDNA分子は線状であってもよい。ガイドRNAまたはガイドポリヌクレオチドをコードするDNA分子も環状であってもよい。

【0238】

いくつかの態様において、塩基エディターシステムの1つ以上の構成要素は、DNA配列によってコードされ得る。このようなDNA配列は、発現系(例えば細胞)と一緒に、または別々に導入することができる。例えば、ポリヌクレオチドプログラム可能なヌクレオチド結合ドメインおよびガイドRNAをコードするDNA配列を細胞に導入することができ、各DNA配列は別個の分子の一部であることができ(例えばポリヌクレオチドプログラム可能なヌクレオチド結合ドメインのコード配列を含む1つのベクターおよびガイドRNAコード配列を含む第2のベクター)、または両方が同じ分子の一部であることができる(例えばポリヌクレオチドプログラム可能なヌクレオチド結合ドメイン及びガイドRNAの両方のためのコード(および調節)配列を含む一つのベクター)。

10

【0239】

ガイドポリヌクレオチドは、新しいまたは増強された特徴を有する核酸を提供するための1以上の改変を含むことができる。ガイドポリヌクレオチドは、核酸親和性タグを含むことができる。ガイドポリヌクレオチドは、合成ヌクレオチド、合成ヌクレオチドアナログ、ヌクレオチド誘導体、および/または修飾ヌクレオチドを含むことができる。

20

【0240】

場合によっては、gRNAまたはガイドポリヌクレオチドは修飾(改変)を含むことができる。修飾は、gRNAまたはガイドポリヌクレオチドの任意の位置に施され得る。単一のgRNAまたはガイドポリヌクレオチドに対して複数の修飾を行うことができる。gRNAまたはガイドポリヌクレオチドは、修飾後に品質管理を受けることができる。場合によっては、品質管理は、PAGE、HPLC、MS、またはそれらの任意の組み合わせを含むことができる。

30

【0241】

gRNAまたはガイドポリヌクレオチドの修飾は、置換、挿入、欠失、化学的修飾、物理的修飾、安定化、精製、またはそれらの任意の組合せであり得る。

【0242】

gRNAまたはガイドポリヌクレオチドはまた、5' アデニル酸、5' グアノシン三リン酸キャップ、5' N7-メチルグアノシン三リン酸キャップ、5' 三リン酸キャップ、3' リン酸、3' チオリン酸、5' リン酸、5' チオリン酸、Cis-Synチミジン二量体、三量体、C12スペーサー、C3スペーサー、C6スペーサー、dスペーサー、PCスペーサー、rスペーサー、スペーサー18、スペーサー9、3' -3' 修飾、5' -5' 修飾、塩基脱落、アクリジン、アゾベンゼン、ビオチン、ビオチンBB、ビオチンTEG、コレステロールTEG、デスチオビオチンTEG、DNP TEG、DNP-X、DOTA、dT-ビオチン、二重ビオチン、ソラレン、ソラレンC2、ソラレンC6、TINA、3' DABCYL、ブラックホールクエンチャー1、ブラックホールクエンチャー2、DABCYL SE、dT-DABCYL、IRDye QC-1、QSY-21、QSY-35、QSY-7、QSY-9、カルボキシリリンカー、チオールリンカー、2'-デオキシリボヌクレオシド類似体プリン、2'-デオキシリボヌクレオシド類似体ピリミジン、リボヌクレオシド類似体、2'-O-メチルリボヌクレオシド類似体、糖修飾類似体、ウォブル/ユニバーサル塩基、蛍光色素標識、2'-フルオロRNA、2'-O-メチルRNA、メチルホスホネート、ホスホジエステルDNA、ホスホジエステルRNA、ホスホチオエステルDNA、ホスホリボチオネートRNA、UNA、プソイドリジン-5'-トリホスフェート、5'-メチルシジン-5'-トリホスフェート、またはそれらの任意の組合せでも修飾され得る。

40

50

【0243】

場合によっては、修飾は永続的である。その他の場合、修飾は一時的である。場合によっては、gRNAまたはガイドポリヌクレオチドに対して複数の修飾が行われる。gRNAまたはガイドポリヌクレオチドの修飾は、それらのコンホメーション、極性、疎水性、化学反応性、塩基対形成相互作用、またはそれらの組み合わせなどの、ヌクレオチドの物理化学的特性を変化させることができる。

【0244】

修飾はホスホロチオエート置換でもあり得る。ある場合には、天然のホスホジエステル結合は細胞のヌクレアーゼによって急速に分解されやすく、ホスホロチオエート (PS) 結合置換体を用いたヌクレオチド間結合の修飾は、細胞分解による加水分解に対してより安定である。修飾は、gRNAまたはガイドポリヌクレオチドの安定性を増大させることができる。修飾は生物学的活性を増強させることもできる。ある場合には、ホスホロチオエート強化RNA gRNAは、RNアーゼA、RNアーゼT1、子牛血清ヌクレアーゼ、またはそれらの組み合わせを阻害することができる。これらの特性は、in vivoまたはin vitroでヌクレアーゼへの暴露がある可能性が高いアプリケーションにおいて、PS-RNA gRNAの使用を可能にする。例えば、ホスホロチオエート (PS) 結合をgRNAの5'末端または3'末端における最後の3~5ヌクレオチドの間に導入することができ、それはエキソヌクレアーゼ分解を阻害し得る。ある場合には、ホスホロチオエート結合をgRNA全体に加えてエンドヌクレアーゼによる攻撃を減らすことができる。

【0245】

[プロトスペーサー隣接モチーフ]

用語「プロトスペーサー隣接モチーフ(PAM)」またはPAM様モチーフは、CRISPR細菌適応免疫系においてCas9ヌクレアーゼによって標的化されるDNA配列の直後の2~6塩基対DNA配列を指す。いくつかの実施形態では、PAMは5' PAM (すなわちプロトスペーサーの5'末端の上流に位置する)であり得る。他の実施形態では、PAMは3' PAM (すなわちプロトスペーサーの5'末端の下流に位置する)であり得る。

【0246】

プロトスペーサー隣接モチーフ(PAM)またはPAM様モチーフは、CRISPR細菌適応免疫系においてCas9ヌクレアーゼによって標的化されるDNA配列の直後の2~6塩基対DNA配列を指す。いくつかの実施形態では、PAMは5' PAM (すなわちプロトスペーサーの5'末端の上流に位置する)であり得る。他の実施形態では、PAMは3' PAM (すなわちプロトスペーサーの5'末端の下流に位置する)であり得る。PAM配列は標的結合に必須であるが、正確な配列はCasタンパク質の種類に依存する。

本明細書で提供される塩基エディターは、標準的または非標準的プロトスペーサー隣接モチーフ (PAM) 配列を含むヌクレオチド配列に結合することができるCRISPRタンパク質由来ドメインを含むことができる。PAM部位は、標的ポリヌクレオチド配列に近接するヌクレオチド配列である。本開示のいくつかの側面は、異なるPAM特異性を有するCRISPRタンパク質の全部または一部を含む塩基エディターを提供する。例えば、*S. pyogenes*由来のCas9 (spCas9) などのCas9タンパク質は、典型的に、特定の核酸領域に結合するために標準的なNGG PAM配列を必要とし、ここで「NGG」中の「N」はアデニン (A)、チミン(T)、グアニン (G)、またはシトシン (C) であり、Gはグアニンである。PAMはCRISPRタンパク質特異的であり得、異なるCRISPRタンパク質由来ドメインを含む異なる塩基エディター間で異なり得る。PAMは標的配列の5'または3'にあり得る。PAMは、標的配列の上流または下流にあり得る。PAMは、1、2、3、4、5、6、7、8、9、10またはそれ以上のヌクレオチドの長さであり得る。多くの場合、PAMは2~6ヌクレオチドの長さである。

【0247】

ある態様において、Cas9ドメインは、*Streptococcus pyogenes*由来のCas9ドメインである (SpCas9)。ある態様において、SpCas9ドメインは、ヌクレアーゼ活性SpCas9、ヌクレアーゼ不活性SpCas9 (SpCas9d)、またはSpCas9ニッカーゼ (SpCas9n) である

。いくつかの実施形態において、SpCas9は、D9X突然変異、または本明細書に提供されるアミノ酸配列のいずれかにおける対応する突然変異を含み、ここでXはD以外の任意のアミノ酸である。いくつかの実施形態において、SpCas9は、D9A突然変異、または本明細書に提供されるアミノ酸配列のいずれかにおける対応する突然変異を含む。ある態様において、SpCas9ドメイン、SpCas9dドメインまたはSpCas9nドメインは、非標準PAMを有する核酸配列に結合することができる。ある態様において、SpCas9ドメイン、SpCas9dドメインまたはSpCas9nドメインは、NGG、NGAまたはNGCG PAM配列を有する核酸配列に結合することができる。いくつかの実施形態において、SpCas9ドメインは、D1135X、R1335X、およびT1337X突然変異の1つ以上、または本明細書に提供されるアミノ酸配列のいずれかにおける対応する突然変異突然変異を含み、ここでXは任意のアミノ酸である。いくつかの実施形態において、SpCas9ドメインは、D1135E、R1335Q、およびT1337R突然変異の1つ以上、または本明細書に提供されるアミノ酸配列のいずれかにおける対応する突然変異を含む。いくつかの実施形態において、SpCas9ドメインは、D1135E、R1335Q、およびT1337R突然変異、または本明細書に提供されるアミノ酸配列のいずれかにおける対応する突然変異を含む。いくつかの実施形態において、SpCas9ドメインは、D1135X、R1335X、およびT1337X突然変異のうちの一つ以上、または本明細書において提供されるアミノ酸配列のいずれかにおける対応する突然変異を含み、ここでXは任意のアミノ酸である。いくつかの実施形態において、SpCas9ドメインは、D1135V、R1335Q、およびT1337R突然変異の1以上、または本明細書に提供されるアミノ酸配列のいずれかにおける対応する突然変異を含む。いくつかの実施形態において、SpCas9ドメインは、D1135V、R1335Q、およびT1337R突然変異、または本明細書に提供されるアミノ酸配列のいずれかにおける対応する突然変異を含む。いくつかの実施形態において、SpCas9ドメインは、D1135X、G1218X、R1335X、およびT1337X突然変異の1以上、または本明細書に提供されるアミノ酸配列のいずれかにおける対応する突然変異を含み、ここでXは任意のアミノ酸である。いくつかの実施形態において、SpCas9ドメインは、D1135V、G1218R、R1335Q、およびT1337R突然変異の1つ以上、または本明細書に提供されるアミノ酸配列のいずれかにおける対応する突然変異を含む。いくつかの実施形態において、SpCas9ドメインは、D1135V、G1218R、R1335Q、およびT1337R突然変異、または本明細書において提供されるアミノ酸配列のいずれかにおける対応する突然変異を含む。

10

20

30

【0248】

いくつかの実施形態において、本明細書に提供される融合タンパク質のいずれかのCas9ドメインは、本明細書に記載されるCas9ポリペプチドに対して少なくとも60%、少なくとも65%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、または少なくとも99.5%の同一性であるアミノ酸配列を含む。いくつかの実施形態において、本明細書に提供される融合タンパク質のいずれかのCas9ドメインは、本明細書に記載される任意のCas9ポリペプチドのアミノ酸配列を含む。いくつかの実施形態において、本明細書に提供される融合タンパク質のいずれかのCas9ドメインは、本明細書に記載される任意のCas9ポリペプチドのアミノ酸配列からなる。

40

【0249】

PAM配列に結合することができる例示的なSpCas9タンパク質のアミノ酸配列を以下に提供する：

【0250】

例示的なPAM結合SpCas9のアミノ酸配列は以下の通りである：

```
MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEA
TRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLLESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSDVD
KLFQILVQTYNQLFEENPINASGVDKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIAL
SLGLTPNFKSNFDLAEDAKLQLSKDQYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDI
```

50

LRVNT EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG
 GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ
 EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
 AQSFIERMTNFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
 IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLIN
 GIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA
 GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGI
 KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK
 DDSIDNKVLRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
 LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRK
 DFQFYKVVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQ
 EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
 MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK
 VEKGSKSKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
 KRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEII
 EQISEFSKRVLADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID
 RKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGD.

10

【 0 2 5 1 】

例示的なPAM結合SpCas9nのアミノ酸配列は以下の通りである：

20

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEA
 TRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNI
 VDEVAYHEKYPTIYHLRKKLVDSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSDVD
 KLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIAL
 SLGLTPNFKSNFDLAEDAKLQLSKD TYDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDI
 LRVNT EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG
 GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ
 EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
 AQSFIERMTNFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
 IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLIN
 GIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA
 GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGI
 KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK
 DDSIDNKVLRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
 LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRK
 DFQFYKVVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQ
 EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
 MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK
 VEKGSKSKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
 KRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEII
 EQISEFSKRVLADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID
 RKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGD.

30

40

【 0 2 5 2 】

例示的なPAM結合SpEQR Cas9のアミノ酸配列は以下の通りである：

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEA
 TRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNI
 VDEVAYHEKYPTIYHLRKKLVDSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSDVD
 KLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIAL
 SLGLTPNFKSNFDLAEDAKLQLSKD TYDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDI

50

LRVNT EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG
 GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ
 EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
 AQSFIERMTNFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
 IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLIN
 GIRDKQSGKTILDFLKSDFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA
 GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGI
 KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK
 DDSIDNKVLRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG 10
 LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRK
 DFQFYKVVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQ
 EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
 MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFESPTVAYSVLVVAK
 VEKGSKSKLKS VKELLGITIMERS SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
 KRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEII
 EQISEFSKR VILADANLDKVL SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID
 RKQYRSTKEVL DATLIHQ SITGLYETRIDLSQLGGD. この配列において、D1135、R13
 35、およびT1337から変異されてSpEQR Cas9を生じることができる残基E1135、Q13
 35、およびR1337には太字で下線を付している。 20

【0253】

例示的なPAM結合SpVQR Cas9のアミノ酸配列は以下の通りである。

MDKKYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEA
 TRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNI
 VDEVAYHEKYPTIYHLRKKLVDSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSDVD
 KLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIAL
 SLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDI
 LRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG
 GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ
 EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS 30
 AQSFIERMTNFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
 IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLIN
 GIRDKQSGKTILDFLKSDFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA
 GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGI
 KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK
 DDSIDNKVLRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
 LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRK
 DFQFYKVVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQ
 EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS 40
 MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAK
 VEKGSKSKLKS VKELLGITIMERS SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
 KRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEII
 EQISEFSKR VILADANLDKVL SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID
 RKQYRSTKEVL DATLIHQ SITGLYETRIDLSQLGGD. この配列において、D1135、R13
 35、およびT1337から変異されてSpVQR Cas9を生じることができる残基V1135、Q13
 35、およびR1337には太字で下線を付している。

【0254】

例示的なPAM結合SpVRER Cas9のアミノ酸配列は以下の通りである。

MDKKYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEA 50

TRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLVESFLVEEDKKHERHPIFGNI
 VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVD
 KLFQILVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIAL
 SLGLTPNFKSNFDLAEDAQLQLSKDQYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDI
 LRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG
 GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ
 EDFYPFLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
 AQSFIERM TNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
 IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWRLSRKLIN 10
 GIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA
 GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGI
 KELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK
 DDSIDNKVLRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
 LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRK
 DFQFYKVRINNYHHAHDAYLNAVVGTAIKKYPKLESEFVYGDYKVYDVRKMIKSEQ
 EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
 MPQVNIVKKTEVQTTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAK
 VEKGSKSKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
 KRMLASARELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEII 20
 EQISEFSKRVLADANLDKVL SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID
 RKEYRSTKEVLDATLIHQSI TGLYETRIDLSQLGGD.

【 0 2 5 5 】

例示的なPAM結合SpVRQR Cas9のアミノ酸配列は以下の通りである。

MDKKYSIGLAIGTNSVGWAVITDEYKVPKSKFKVLGNTDRHSIKKNLIGALLFDSGETAEA
 TRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLVESFLVEEDKKHERHPIFGNI
 VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVD
 KLFQILVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIAL
 SLGLTPNFKSNFDLAEDAQLQLSKDQYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDI
 LRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG 30
 GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ
 EDFYPFLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
 AQSFIERM TNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
 IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWRLSRKLIN
 GIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA
 GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGI
 KELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK
 DDSIDNKVLRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
 LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRK 40
 DFQFYKVRINNYHHAHDAYLNAVVGTAIKKYPKLESEFVYGDYKVYDVRKMIKSEQ
 EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
 MPQVNIVKKTEVQTTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAK
 VEKGSKSKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
 KRMLASARELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEII
 EQISEFSKRVLADANLDKVL SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID
 RKQYRSTKEVLDATLIHQSI TGLYETRIDLSQLGGD.

1135D1135、G1218、R1335、およびT1337から変異されてSpVRQR Cas9を生じることができる上記残基V1135、R1218、Q1335、およびR1337には太字で下線を付している。

10

20

30

40

50

【 0 2 5 6 】

ある態様において、Cas9ドメインは、組換えCas9ドメインである。ある態様において、組換えCas9ドメインは、SpyMacCas9ドメインである。いくつかの実施形態において、SpyMacCas9ドメインは、ヌクレアーゼ活性SpyMacCas9、ヌクレアーゼ不活性SpyMacCas9 (SpyMacCas9d)、またはSpyMacCas9ニッカーゼ(SpyMacCas9n)である。いくつかの実施形態において、SaCas9ドメイン、SaCas9dドメイン、またはSaCas9nドメインは、非標準PAMを有する核酸配列に結合することができる。いくつかの実施形態において、SpyMacCas9ドメイン、SpCas9dドメイン、またはSpCas9nドメインは、NAA PAM配列を有する核酸配列に結合することができる。

例示的SpyMacCas9

```
MDKKYSIGLDIGTNSVGWAVITDDYKVPKFKVVLGNTDRHSIKKNLIGALLFGSGETAEA
TRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLADSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSDVD
KLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSRLENLIAQLPGEKRNGLFGNLIAL
SLGLTPNFKSNFDLAEDAKLQLSKDITYDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDI
LRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGG
ASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQE
DFYPFLKDNREKIEKILTFRIPIYYVGPLARGNSRFAMTRKSEETITPWNFEVVDKGGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTRNKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDRGMIEERLKYAHLFDDKVMKQLKRRRYTGWRLSRKLIN
GIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGHSLEHQIANLA
GSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE
LGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDD
SIDNKVLRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLS
ELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDF
QFYKVINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIG
KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQ
VNIVKKTEIQTVGQNGGLFDDNPKSPLEVTSPKLVPLKELNPKKYGGYQKPTTAYPVLL
ITDTKQLIPISVMNKKQFEQNPVKFLRDRGYQQVGKNDFIKLPKYTLVDIGDGIKRLWAS
SKEIHKGNQLVVSCKSQILLYHAHHLSDLSNDYLQNHNNQFDVLFNEIISFSKCKLKGK
EHIQKIENVYSNKKNSASIEELAESFIKLLGFTQLGATSPFNFLGVKLNQKQYKGGKDYILP
CTEGLIRQSITGLYETRVDSLKIGED.
```

【 0 2 5 7 】

[高忠実度Cas9ドメイン]

本開示のいくつかの態様は、高忠実度Cas9ドメインを提供する。いくつかの実施形態において、高忠実度Cas9ドメインは、対応する野生型Cas9ドメインと比較して、Cas9ドメインとDNAの糖-リン酸骨格との間の静電相互作用を減少させる一つ以上の突然変異を含む人工Cas9ドメインである。特定の理論に縛られることを望まないが、DNAの糖-リン酸骨格との静電相互作用を減少させた高忠実度Cas9ドメインは、より少ないオフターゲット効果を有し得る。ある態様において、Cas9ドメイン(例えば、野生型Cas9ドメイン)は、Cas9ドメインとDNAの糖-リン酸骨格との間の結合を低減させる一つ以上の突然変異を含む。ある態様において、Cas9ドメインは、Cas9ドメインとDNAの糖-リン酸骨格との間の結合を少なくとも1%、少なくとも2%、少なくとも3%、少なくとも4%、少なくとも5%、少なくとも10%、少なくとも15%、少なくとも20%、少なくとも25%、少なくとも30%、少なくとも35%、少なくとも40%、少なくとも45%、少なくとも50%、少なくとも55%、少なくとも60%、少なくとも65%、または少なくとも70%だけ低減させる一つ以上の突然変異を含む。

【 0 2 5 8 】

いくつかの実施形態において、本明細書で提供されるCas9融合タンパク質のいずれかは

、N497X、R661X、Q695X、および/またはQ926X突然変異の1以上、または本明細書で提供されるアミノ酸配列のいずれかにおける対応する突然変異を含み、ここでXは任意のアミノ酸である。いくつかの実施形態において、本明細書で提供されるCas9融合タンパク質のいずれかは、N497A、R661A、Q695A、および/またはQ926A突然変異の1以上、または本明細書で提供されるアミノ酸配列のいずれかにおける対応する突然変異を含む。いくつかの実施形態において、Cas9ドメインは、D10A突然変異、または本明細書に提供されるアミノ酸配列のいずれかにおける対応する突然変異を含む。高い忠実度を有するCas9ドメインは当技術分野で公知であり、当業者には明らかであろう。例えば、高い忠実度を有するCas9ドメインは、Kleinstiver, B.P., et al. “High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects.” *Nature* 529, 490-495 (2016); およびSlaymaker, I.M., et al. “Rationally engineered Cas9 nucleases with improved specificity.” *Science* 351, 84-88 (2015)に記述されており、それぞれの全内容は参照により本明細書に組み入れられる。

では、Cas9と比べた高忠実度Cas9ドメイン突然変異は太字および下線で示されている。
 MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAETA
 TRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNI
 VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVD
 KLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIAL
 SLGLTPNFKSNFDLAEDAKLQLSKDITYDDLDNLLAQIGDQYADFLAAKNLSDAILLSDI
 LRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG
 GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ
 EDFYPFLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
 AQSFIERMTAFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
 IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGALSRKLIN
 GIRDKQSGKTILDFLKSDGFANRNFMALIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAG
 SPAIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIK
 ELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVHDHVPQSFLK
 DDSIDNKVLRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
 LSELDKAGFIKRQLVETRAITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRK
 DFQFYKVVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQ
 EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
 MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK
 VEKGSKSKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
 KRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGKSPEDNEQKQLFVEQHKHYLDEII
 EQISEFSKRVLADANLDKVL SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID
 RKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGD.

【0259】

ある場合には、バリエーションCas9タンパク質はH840A、P475A、W476A、N477A、D1125A、W1126A、D1127Aの突然変異を有し、その結果、標的DNAまたはRNAを切断する能力が低下している。このようなCas9タンパク質は、標的DNA(例えば一本鎖標的DNA)を切断する能力は低下しているが、標的DNA(例えば一本鎖標的DNA)に結合する能力は保持している。別の非限定的な例として、いくつかの場合において、バリエーションCas9タンパク質は、D10A、H840A、P475A、W476A、N477A、D1125A、W1126A、およびD1127A突然変異を有し、その結果、ポリペプチドは、標的DNA(例えば一本鎖標的DNA)を切断する能力が低下している。そのようなCas9タンパク質は、標的DNA(例えば一本鎖標的DNA)を切断する能力を低下しているが、標的DNA(例えば一本鎖標的DNA)に結合する能力を保持している。バリエーションCas9タンパク質がW476AおよびW1126A変異を有する場合、またはバリエーションCas9タンパク質がP475A、W476A、N477A、D1125A、W1126A、およびD1127A変異を有する場合、バリエーションCas9タンパク質はPAM配列に

効率的に結合しない。したがって、このような場合には、このようなバリエーションCas9タンパク質を結合の方法に用いると、この方法はPAM配列を必要としない。換言すれば、ある場合には、このようなバリエーションCas9タンパク質を結合の方法に用いる場合、この方法はガイドRNAを含み得るが、この方法は、PAM配列の非存在下で行うことができる(したがって、結合の特異性はガイドRNAの標的セグメントによってもたらされる)。上記の効果を達成するために、他の残基が変異され得る(すなわち一方または他方のヌクレアーゼ部分を不活性化する)。非限定的な例として、残基D10、G12、G17、E762、H840、N854、N863、H982、H983、A984、D986、および/またはA987を変更(すなわち置換)することができる。また、アラニン置換以外の変異も好適である。

【0260】

ある態様において、塩基エディターのCRISPRタンパク質由来ドメインは、標準的PAM配列(NGG)を有するCas9タンパク質の全部または一部を含み得る。他の実施形態では、塩基エディターのCas9由来ドメインは、非標準的PAM配列を用いることができる。そのような配列は本技術分野で記述されており当業者には明らかであろう。例えば、非標準的PAM配列に結合するCas9ドメインは、Kleinstiver, B. P., et al., "Engineered CRISPR-Cas9 nucleases with altered PAM specificities" *Nature* 523, 481-485 (2015); および Kleinstiver, B. P., et al., "Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition" *Nature Biotechnology*, 33, 1293-1298 (2015)に記述されており、それぞれの全内容を参照によりここに組み込む。

【0261】

いくつかの例において、本明細書に開示される塩基エディターのCRISPRタンパク質由来ドメインによって認識されるPAMは、塩基エディターをコードするインサート(例えばAVインサート)とは別個のオリゴヌクレオチド上で細胞に提供され得る。そのような場合、別個のオリゴヌクレオチド上にPAMを提供することは、さもなくば標的配列と同じポリヌクレオチド上に隣接するPAMが存在しないために切断することができない標的配列の切断を可能にする。

【0262】

一実施形態において、*S. pyogenes* Cas9 (SpCas9) を、ゲノム工学のためのCRISPRエンドヌクレアーゼとして使用することができる。ただし、他のものも使用され得る。場合によっては、異なるエンドヌクレアーゼを用いて特定のゲノム標的を標的化することができる。場合によっては、非NGG PAM配列を有する合成SpCas9由来バリエーションを使用することができる。さらに、様々な種からの他のCas9オルソログが同定されており、これらの「非SpCas9」は、本開示でも有用になり得る種々のPAM配列に結合し得る。例えば、比較的大きなサイズのSpCas9 (約4 kbのコード配列)は、細胞内で効率的に発現することができないSpCas9 cDNAプラスミドをもたらすこともあり得る。逆に、*Staphylococcus aureus* Cas9 (SaCas9) のコード配列は、SpCas9よりも約1キロ塩基短く、細胞内で効率的に発現させ得る。SpCas9と同様に、SaCas9エンドヌクレアーゼは、*in vitro*の哺乳類細胞および*in vivo*のマウスにおいて標的遺伝子を修飾する能力がある。ある場合には、Casタンパク質は異なるPAM配列を標的とすることができる。いくつかの場合において、標的遺伝子は、Cas9 PAM、例えば、5'-NGGに隣接し得る。他の場合には、他のCas9オルソログは異なるPAM要件を有し得る。例えば、*S. thermophilus* のもののようなPAM (CRISPR1の場合は5'-NNAGAA、CRISPR3の場合は5'-NGGNG) および *Neisseria meningitidis* のもの (5'-NNNGATT) のような他のPAMも標的遺伝子に隣接して見出され得る。

【0263】

いくつかの実施形態において、*S. pyogenes*系について、標的遺伝子配列は、5'-NGG PAMの前(すなわち、その5'側)にあり得、20 ntガイドRNA配列が、反対側の鎖と塩基対を形成して、PAMに隣接するCas9切断を媒介することができる。ある場合には、隣接切断は、PAMの(約)3塩基対上流であり得る。ある場合には、隣接切断は、PAMの(約)

10

20

30

40

50

10塩基対上流であり得る。ある場合には、隣接切断は、PAMの(約)0~20塩基対上流であり得る。例えば、隣接切断は、PAM上流の1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、または30塩基対の隣であり得る。隣接切断はPAMの1~30塩基対下流でもあり得る。

【0264】

[核局在化配列(NLS)を含む融合タンパク質]

1以上の核局在化配列(NLS)を含むCRISPR酵素をコードするベクターを使用することができる。例えば、約1、2、3、4、5、6、7、8、9、10個のNLSを使用することができる。CRISPR酵素は、アミノ末端またはその近傍のNLS、カルボキシ末端またはその近傍の約1個、2個、3個、4個、5個、6個、7個、8個、9個、10個またはそれ以上のNLS、またはこれらの任意の組み合わせ(例えばアミノ末端における1以上のNLSおよびカルボキシ末端における1以上のNLS)を含むことができる。複数のNLSが存在する場合、各NLSは他から独立して選択することができ、したがって1つのNLSが1つ以上のコピーで存在することができ、および/または1つ以上のコピーで存在する1つ以上の他のNLSと組み合わせて存在することができる。

10

【0265】

法において使用されるCRISPR酵素は、約6個のNLSを含むことができる。NLSに最も近いアミノ酸がN-またはC-末端からポリペプチド鎖に沿って約50アミノ酸内(例えば1、2、3、4、5、10、15、20、25、30、40、または50アミノ酸内)にあるとき、そのNLSはN-またはC-末端の近傍にあると考えられる。

20

【0266】

ある態様において、NLSは、アミノ酸配列PKKKRKVEGADKRTADGSEFES PKKKRKV, KRTADGSEFESPCKKKRKV, KRPAATKKAGQAKKKK, KKTELQTTNAENKTKKL, KRGINDRNFWRGENGRKTR, RKSGKIAAIVVKRPRKPKKKRKV, またはMDSLLMNRRKFLYQFKNVRWAKGRRETYLCを含む。

【0267】

いくつかの実施形態において、NLSはリンカー中に存在するか、またはNLSはリンカー、例えば本明細書に記載されるリンカーによって隣接される。いくつかの実施形態において、N末端またはC末端のNLSは、二部NLSである。二部NLSは、比較的短いスペーサー配列によって分離される二つの塩基性アミノ酸クラスターを含む(それゆえにbipartite、2部分と呼ばれ、一部(monopartite)NLSは異なる)。ヌクレオプラスミンのNLSであるKR[PAATKKAGQA]KKKKは遍在的な二部シグナルのプロトタイプであり、塩基性アミノ酸の二つのクラスターが約10アミノ酸のスペーサーによって隔てられたものである。例示的な二部NLSの配列は、PKKKRKVEGADKRTADGSEFES PKKKRKVである。

30

【0268】

いくつかの態様において、本発明の融合タンパク質は、リンカー配列を含まない。ある態様において、1以上のドメインまたはタンパク質の間にリンカー配列が存在する。

【0269】

PAM配列は、当該技術分野で公知の任意のPAM配列であり得る。適切なPAM配列には、以下が含まれるが、これらに限定されない: NGG, NGA, NGC, NGN, NGT, NGCG, NGAG, NGAN, NGNG, NGCN, NGCG, NGTN, NNGRRT, NNNRRT, NNGRR(N), TTTV, TYCV, TYCV, TATV, NNNNGATT, NNAGAAW, または NAAAAC。Yはピリミジンであり、Nは任意のヌクレオチド塩基であり、WはAまたはTである。

40

【0270】

[核酸塩基編集ドメイン]

ポリヌクレオチドプログラム可能なヌクレオチド結合ドメインおよび核酸塩基編集ドメイン(例えばデアミナーゼドメイン)を含む融合タンパク質を含む塩基エディターを本明細書に記載する。塩基エディターは、標的配列を認識することができるガイドポリヌクレオチドと相互作用することによって、標的ポリヌクレオチド配列中の1以上の塩基を編集す

50

るようにプログラムすることができる。標的配列がいったん認識されると、編集が行われるポリヌクレオチド上に塩基エディターが固定され、次いで、塩基エディターのデアミナーゼドメイン成分が標的塩基を編集することができる。

【0271】

ある態様において、核酸塩基編集ドメインは、デアミナーゼドメインである。場合によっては、デアミナーゼドメインは、シトシンデアミナーゼまたはシチジンデアミナーゼであり得る。いくつかの実施形態では、「シトシンデアミナーゼ」および「シチジンデアミナーゼ」という用語は、交換可能に使用され得る。場合によっては、デアミナーゼドメインは、アデニンデアミナーゼまたはアデノシンデアミナーゼであり得る。いくつかの実施形態では、「アデニンデアミナーゼ」および「アデノシンデアミナーゼ」という用語は、交換可能に使用され得る。核酸塩基編集タンパク質の詳細は、国際PCT出願番号PCT/2017/045381 (WO2018/027078) およびPCT/US2016/058344 (WO2017/070632) に記載されており、これらの各々は、その全体が参照により本明細書に組み込まれる。Komor, A.C., et al., "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" *Nature* 533, 420-424 (2016); Gaudelli, N. M., et al., "Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage" *Nature* 551, 464-471 (2017); および Komor, A.C., et al., "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" *Science Advances* 3:eaao4774 (2017)も参照(その全内容が参照により本明細書に組み込まれる)。

10

20

【0272】

[CからTへの編集]

いくつかの実施形態において、本明細書に開示される塩基エディターは、ポリヌクレオチドの標的シチジン(C)塩基を脱アミノ化してチミンの塩基対形成特性を有するウリジン(U)を生成することができる、シチジンデアミナーゼを含む融合タンパク質を含む。いくつかの実施形態において、例えば、ポリヌクレオチドが二本鎖(例えばDNA)である場合、ウリジン塩基は次いで、チミジン塩基で置換されて(例えば細胞修復機構により)、C:GからT:Aへの転移を生じさせることができる。他の実施形態において、塩基エディターによる核酸中のCからUへの脱アミノ化は、UからTへの置換を伴うことができない。

30

【0273】

ポリヌクレオチド中の標的Cを脱アミノ化してUを生じさせることは、本明細書に記載の塩基エディターによって実行することができるタイプの塩基編集の非限定的な例である。別の例において、シチジンデアミナーゼドメインを含む塩基エディターは、シトシン(C)塩基からグアニン(G)塩基への変換を媒介することができる。例えば、塩基エディターのシチジンデアミナーゼドメインによるシチジンの脱アミノ化によって生成されたポリヌクレオチドのUは、塩基除去修復機構(例えばウラシルDNAグリコシラーゼ(UDG)ドメインによるもの)によってポリヌクレオチドから切り出され、脱塩基部位を生成することができる。次に、脱塩基部位の反対側の核酸塩基が、Cのような別の塩基で置換され得(例えば塩基修復機構により)、例えばこれは損傷乗り越えポリメラーゼによって行われる。脱塩基部位の反対側の核酸塩基がCで置換されるのが典型的であるが、他の置換(例えばA、G、T)も起こりうる。

40

【0274】

従って、いくつかの実施形態において、本明細書に記載される塩基エディターは、ポリヌクレオチド中の標的Cを脱アミノ化してUにすることができる脱アミノ化ドメイン(例えばシチジンデアミナーゼドメイン)を含む。さらに、以下に記載するように、塩基エディターは、脱アミノ化から生じるUの、ある態様においてはTまたはGへの、変換を促進させるさらなるドメインを含むことができる。例えば、シチジンデアミナーゼドメインを含む塩基エディターは、さらにウラシルグリコシラーゼ阻害因子(UGI)ドメインを含み、TによるUの置換を媒介し、CからTへの塩基編集事象を完了することができる。損傷乗り越えポ

50

リメラーゼは脱塩基部位の反対側でCの取り込みを促進することができるので(すなわち脱塩基部位でのGの取り込みをもたらし、CからGへの塩基編集事象を完了する)、別の例では塩基エディターは損傷乗り越えポリメラーゼを組み込んで、CからGへの塩基編集の効率を改善させることができる。

【0275】

ドメインとしてシチジンデアミナーゼを含む塩基エディターは、DNA、RNAおよびDNA-RNAハイブリッドを含む任意のポリヌクレオチドにおいて標的Cを脱アミノ化することができる。典型的には、シチジンデアミナーゼは、ポリヌクレオチドの一本鎖部分の文脈に配置されるC核酸塩基を触媒する。ある態様において、標的Cを含むポリヌクレオチド全体が一本鎖であり得る。例えば、塩基エディターに組み込まれたシチジンデアミナーゼは、一本鎖RNAポリヌクレオチド中の標的Cを脱アミノ化することができる。他の実施形態において、シチジンデアミナーゼドメインを含む塩基エディターは、二本鎖ポリヌクレオチドに作用することができるが、標的Cは、脱アミノ化反応の時点で一本鎖状態になるポリヌクレオチドの一部に位置し得る。例えば、NAGPBドメインがCas9ドメインを含む実施形態において、いくつかのヌクレオチドは、Cas9-gRNA-標的DNA複合体の形成の際に対合しないままにしておくことができ、その結果、Cas9「Rループ複合体」が形成される。これらの不對ヌクレオチドは、一本鎖特異的ヌクレオチドデアミナーゼ酵素(例えばシチジンデアミナーゼ)の基質として作用し得る一本鎖DNAのバブルを形成することができる。

【0276】

ある態様において、塩基エディターのシチジンデアミナーゼは、アポリポタンパク質B mRNA編集複合体(APOBEC)ファミリーデアミナーゼの全部または一部を含むことができる。APOBECは進化的に保存されたシチジンデアミナーゼファミリーである。このファミリーのメンバーはCからUへの編集酵素である。APOBEC様タンパク質のN末端ドメインは触媒ドメインであり、C末端ドメインは偽触媒ドメインである。より具体的には、触媒ドメインは亜鉛依存性のシチジンデアミナーゼドメインであり、シチジンの脱アミノ化にとって重要である。APOBECファミリーメンバーには、APOBEC1、APOBEC2、APOBEC3A、APOBEC3B、APOBEC3C、APOBEC3D(今では「APOBEC3E」はこれを指す)、APOBEC3F、APOBEC3G、APOBEC3H、APOBEC4、および活性化誘導(シチジン)デアミナーゼが含まれる。一部の実施形態では、塩基エディターに組み込まれたデアミナーゼは、APOBEC1デアミナーゼの全部または一部を含む。一部の実施形態では、塩基エディターに組み込まれたデアミナーゼは、APOBEC2デアミナーゼの全部または一部を含む。一部の実施形態では、塩基エディターに組み込まれたデアミナーゼは、APOBEC3デアミナーゼの全部または一部を含む。一部の実施形態では、塩基エディターに組み込まれたデアミナーゼは、APOBEC3Aデアミナーゼの全部または一部を含む。一部の実施形態では、塩基エディターに組み込まれたデアミナーゼは、APOBEC3Bデアミナーゼの全部または一部を含む。一部の実施形態では、塩基エディターに組み込まれたデアミナーゼは、APOBEC3Cデアミナーゼの全部または一部を含む。一部の実施形態では、塩基エディターに組み込まれたデアミナーゼは、APOBEC3Dデアミナーゼの全部または一部を含む。一部の実施形態では、塩基エディターに組み込まれたデアミナーゼは、APOBEC3Eデアミナーゼの全部または一部を含む。一部の実施形態では、塩基エディターに組み込まれたデアミナーゼは、APOBEC3Fデアミナーゼの全部または一部を含む。一部の実施形態では、塩基エディターに組み込まれたデアミナーゼは、APOBEC3Gデアミナーゼの全部または一部を含む。一部の実施形態では、塩基エディターに組み込まれたデアミナーゼは、APOBEC3Hデアミナーゼの全部または一部を含む。一部の実施形態では、塩基エディターに組み込まれたデアミナーゼは、APOBEC4デアミナーゼの全部または一部を含む。ある態様において、塩基エディターに組み込まれたデアミナーゼは、活性化誘導デアミナーゼ(AID)の全部または一部を含む。ある態様において、塩基エディターに組み込まれたデアミナーゼは、シチジンデアミナーゼ1(CDA1)の全部または一部を含む。塩基エディターは、任意の適切な生物(例えばヒトまたはラット)からのデアミナーゼを含むことができることを理解されるべきである。ある態様において、塩基エディターのデアミナーゼドメインは、ヒト、チン

パンジー、ゴリラ、サル、ウシ、イヌ、ラット、またはマウス由来である。ある態様において、塩基エディターのデアミナーゼドメインは、ラットに由来する(例えばラットAPOBEC1)。ある態様において、塩基エディターのデアミナーゼドメインは、ヒトAPOBEC1である。ある態様において、塩基エディターのデアミナーゼドメインは、PmCDA1である。

【0277】

PmCDA1のアミノ酸配列および核酸配列を下記に示す。

tr | A5H718 | A5H718_PETMA Cytosine deaminase OS=Petromyzon marinus OX=7757 PE=2 SV=1 アミノ酸配列:

MTDAEYVRIHEKLDIYTFKKQFFNNKKS VSHRCYVLFELKRRGERRACFWGYAVNKPQSG
TERGIHAEIFSIRKVEEYLRDNPQGFTINWYSSWSPCADCAEKILEWYNQELRGNHGLTK
IWACKLYYEKNARNQIGLWNLDRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLE
KT

10

LKRAEKRRSELSIMIQVKILHTTKSPAV

核酸配列: EF094822.1 Petromyzon marinus isolate PmCDA.21 cytosine deaminase mRNA, complete cds:

TGACACGACACAGCCGTGTATATGAGGAAGGGTAGCTGGATGGGGGGGGGGGAATACG
TTCAGAGAGGA

CATTAGCGAGCGTCTTGTGGTGGCCTTGAGTCTAGACACCTGCAGACATGACCGACGC
TGAGTACGTGA

GAATCCATGAGAAGTTGGACATCTACACGTTTAAGAAACAGTTTTTCAACAACAAAAA
TCCGTGTCGCA

20

TAGATGCTACGTTCTCTTTGAATTAACGACGGGGTGAACGTAGAGCGTGTGGG
GCTATGCTGTG

AATAAACACAGAGCGGGACAGAACGTGGAATTCACGCCGAAATCTTTAGCATTAGAAA
AGTCGAAGAAT

ACCTGCGCGACAACCCCGGACAATTCACGATAAATTGGTACTCATCCTGGAGTCCTTGT
GCAGATTGCGC

TGAAAAGATCTTAGAATGGTATAACCAGGAGCTGCGGGGGAACGGCCACACTTTGAAA
TCTGGGCTTGC

AAACTCTATTACGAGAAAAATGCGAGGAATCAAATTGGGCTGTGGAACCTCAGAGATAA
CGGGGTTGGGT

30

TGAATGTAATGGTAAGTGAACACTACCAATGTTGCAGGAAAATATTCATCCAATCGTCG
CACAATCAATT

GAATGAGAATAGATGGCTTGAGAAGACTTTGAAGCGAGCTGAAAAACGACGGAGCGAGT
TGTCATTATG

ATTCAGGTAAAAATACTCCACACCACTAAGAGTCCTGCTGTTTAAGAGGCTATGCGGAT
GGTTTTTC

ヒトの活性化誘導型シチジンデアミナーゼ(AID)のコード配列(CDS)のアミノ酸配列および核酸配列を下記に示す。

tr | Q6QJ80 | Q6QJ80_HUMAN Activation-induced cytidine deaminase OS=Homo sapiens OX=9606 GN=AICDA PE=2 SV=1 アミノ酸配:

40

MDSLMLNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNKGCHVEL
L

FLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRK
AEPEGLRRLHRAGVQIAIMTFKAPV

核酸配列: NG_011588.1:5001-15681 Homo sapiens activation induced cytidine deaminase (AICDA), RefSeqGene (LRG_17) on chromosome 12:

AGAGAACCATCATTAAATTGAAGTGAGATTTTTCTGGCCTGAGACTTGACAGGGAGGCAAG
AAGACACTCTG

GACACCACTATGGACAGGTAAAGAGGCAGTCTTCTCGTGGGTGATTGCACTGGCCTTCC

50

ATGTAAAAAGA
 TGTAGATTCCTCTGCCTTTCTCATCTACACAGCCCAGGAGGGTAAGTTAATAAAGAGG
 GATTTATTGGT
 AAGAGATGATGCTTAATCTGTTTAACTGGGCCTCAAAGAGAGAATTTCTTTTCTTCT
 GTACTTATTAA
 GCACCTATTATGTGTTGAGCTTATATATACAAAGGGTTATTATATGCTAATATAGTAAT
 AGTAATGGTGG
 TTGGTACTATGGTAATTACCATAAAAATTATTATCCTTTTAAAATAAAGCTAATTATTA
 TTGGATCTTTT
 TTAGTATTCATTTTATGTTTTTATGTTTTTGATTTTTTAAAAGACAATCTCACCCCTGTT 10
 ACCCAGGCTG
 GAGTGCAGTGGTGAATCATAGCTTTCTGCAGTCTTGAACCTCTGGGCTCAAGCAATCC
 TCCTGCCTTGG
 CCTCCCAAAGTGTTGGGATACAGTCATGAGCCACTGCATCTGGCCTAGGATCCATTTAG
 ATTAAAATATG
 CATTTTAAATTTTAAAATAATATGGCTAATTTTACCTTATGTAATGTGTATACTGGCA
 ATAAATCTAGT
 TTGCTGCCTAAAGTTTAAAGTGCTTTCCAGTAAGCTTCATGTACGTGAGGGGAGACATT
 TAAAGTGAAC
 AGACAGCCAGGTGTGGTGGCTCACGCCTGTAATCCCAGCACTCTGGGAGGCTGAGGTGG 20
 GTGGATCGCTT
 GAGCCCTGGAGTTCAAGACCAGCCTGAGCAACATGGCAAACGCTGTTTCTATAACAAA
 AATTAGCCGGG
 CATGGTGGCATGTGCCTGTGGTCCAGCTACTAGGGGGCTGAGGCAGGAGAATCGTTGG
 AGCCCAGGAGG
 TCAAGGCTGCACTGAGCAGTGCTTGCGCCACTGCACTCCAGCCTGGGTGACAGGACCAG
 ACCTTGCCTCA
 AAAAAATAAGAAGAAAAATTAAAAATAAATGGAAACAACCTACAAAGAGCTGTTGTCCTA
 GATGAGCTACT
 TAGTTAGGCTGATATTTTGGTATTTAACTTTTAAAGTCAGGGTCTGTCACCTGCACTAC 30
 ATTATTAATAAT
 ATCAATTCTCAATGTATATCCACACAAAGACTGGTACGTGAATGTTTCATAGTACCTTTA
 TTCACAAAACC
 CCAAAGTAGAGACTATCCAAATATCCATCAACAAGTGAACAAATAAACAAAATGTGCTA
 TATCCATGCAA
 TGGAATACCACCCTGCAGTACAAAGAAGCTACTTGGGGATGAATCCCAAAGTCATGACG
 CTAATGAAAG
 AGTCAGACATGAAGGAGGAGATAATGTATGCCATACGAAATTCTAGAAAATGAAAGTAA
 CTTATAGTTAC
 AGAAAGCAAATCAGGGCAGGCATAGAGGCTCACACCTGTAATCCCAGCACTTTGAGAGG 40
 CCACGTGGGAA
 GATTGCTAGAACTCAGGAGTTCAAGACCAGCCTGGGCAACACAGTGAAACTCCATTCTC
 CACAAAATGG
 GAAAAAAGAAAGCAAATCAGTGGTTGTCCTGTGGGGAGGGGAAGGACTGCAAAGAGG
 GAAGAAGCTCTG
 GTGGGGTGAGGGTGGTGATTCAGGTTCTGTATCCTGACTGTGGTAGCAGTTTGGGGTGT
 TTACATCCAAA
 AATATTCGTAGAATTATGCATCTTAAATGGGTGGAGTTTACTGTATGTAAATTATACCT
 CAATGTAAGAA
 AAAATAATGTGTAAGAAAACCTTTCAATTCTCTTGCCAGCAAACGTTATTCAAATTCCTG 50

AGCCCTTTACT
 TCGCAAATTCTCTGCACTTCTGCCCGTACCATTAGGTGACAGCACTAGCTCCACAAAT
 TGGATAAATGC
 ATTTCTGGAAAAGACTAGGGACAAAATCCAGGCATCACTTGTGCTTTCATATCAACCAT
 GCTGTACAGCT
 TGTGTTGCTGTCTGCAGCTGCAATGGGGACTCTTGATTTCTTTAAGGAACTTGGGTTA
 CCAGAGTATTT
 CCACAAATGCTATTCAAATTAGTGCTTATGATATGCAAGACACTGTGCTAGGAGCCAGA
 AAACAAAGAGG
 AGGAGAAATCAGTCATTATGTGGGAACAACATAGCAAGATATTTAGATCATTTTGACTA 10
 GTTAAAAAAGC
 AGCAGAGTACAAAATCACACATGCAATCAGTATAATCCAAATCATGTAAATATGTGCCT
 GTAGAAAGACT
 AGAGGAATAAACACAAGAATCTTAACAGTCATTGTCATTAGACACTAAGTCTAATTATT
 ATTATTAGACA
 CTATGATATTTGAGATTTAAAAAATCTTTAATATTTTAAAATTTAGAGCTCTTCTATTTT
 TCCATAGTAT
 TCAAGTTTGACAATGATCAAGTATTACTCTTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTGA
 GATGGAGTTT
 TGGTCTTGTTGCCCATGCTGGAGTGGAAATGGCATGACCATAGCTCACTGCAACCTCCAC 20
 CTCCTGGGTTC
 AAGCAAAGCTGTCGCCTCAGCCTCCCGGGTAGATGGGATTACAGGCGCCACCACCACA
 CTCGGCTAATG
 TTTGTATTTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACCTCCT
 GACCTCAGAGG
 ATCCACCTGCCTCAGCCTCCCAAAGTGCTGGGATTACAGATGTAGGCCACTGCGCCCGG
 CCAAGTATTGC
 TCTTATACATTAAAAAACAGGTGTGAGCCACTGCGCCAGCCAGGTATTGCTCTTATAC
 ATTA AAAAATA
 GGCCGGTGCAGTGGCTCACGCCTGTAATCCAGCACTTTGGGAAGCCAAGGCGGGCAGA 30
 ACACCCGAGGT
 CAGGAGTCCAAGGCCAGCCTGGCCAAGATGGTGAAACCCCGTCTCTATTA AAAAATA
 ACATTACCTGG
 GCATGATGGTGGGCGCCTGTAATCCAGCTACTCAGGAGGCTGAGGCAGGAGGATCCGC
 GGAGCCTGGCA
 GATCTGCCTGAGCCTGGGAGGTTGAGGCTACAGTAAGCCAAGATCATGCCAGTATACTT
 CAGCCTGGGCG
 ACAAGTGAGACCGTAACAAAAA AAAAAAATTTAAAAAAGAAATTTAGATCAAGATC
 CAACTGTAAAA
 AGTGGCCTAAACACCACATTAAGAGTTTGGAGTTTATTCTGCAGGCAGAAGAGAACCA 40
 TCAGGGGGTCT
 TCAGCATGGGAATGGCATGGTGCACCTGGTTTTTGTGAGATCATGGTGGTGACAGTGTG
 GGAATGTTAT
 TTTGGAGGGACTGGAGGCAGACAGACCGGTTAAAAGGCCAGCACACAGATAAGGAGG
 AAGAAGATGAGG
 GCTTGGACCGAAGCAGAGAAGAGCAAACAGGGAAGGTACAAATTCAAGAAATATTGGG
 GGGTTTGAATCA
 ACACATTTAGATGATTAATTAATATGAGGACTGAGGAATAAGAAATGAGTCAAGGATG
 GTTCCAGGCTG
 CTAGGCTGCTTACCTGAGGTGGCAAAGTCGGGAGGAGTGGCAGTTTAGGACAGGGGGCA 50

GTTGAGGAATA
 TTGTTTTGATCATTTTTGAGTTTGAGGTACAAGTTGGACACTTAGGTAAAGACTGGAGGG
 GAAATCTGAAT
 ATACAATTATGGGACTGAGGAACAAGTTTATTTTATTTTTGTTTCGTTTTCTTGTTGAA
 GAACAAATTT
 AATTGTAATCCCAAGTCATCAGCATCTAGAAGACAGTGGCAGGAGGTGACTGTCTTGTG
 GGTAAGGGTTT
 GGGGTCCTTGATGAGTATCTCTCAATTGGCCTTAAATATAAGCAGGAAAAGGAGTTTAT
 GATGGATTCCA
 GGCTCAGCAGGGCTCAGGAGGGCTCAGGCAGCCAGCAGAGGAAGTCAGAGCATCTTCTT 10
 TGGTTTAGCCC
 AAGTAATGACTTCTTAAAAAGCTGAAGGAAAATCCAGAGTGACCAGATTATAAACTGT
 ACTCTTGCAAT
 TTCTCTCCCTCCTCTCACCCACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATT
 CAAAAATGTC
 CGCTGGGCTAAGGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGGCGTGACAG
 TGCTACATCCT
 TTTCACTGGACTTTGGTTATCTTCGCAATAAGGTATCAATTAAGTCGGCTTTGCAAGC
 AGTTTAATGGT
 CAACTGTGAGTGCTTTTAGAGCCACCTGCTGATGGTATTACTTCCATCCTTTTTTTGGCAT 20
 TTGTGTCTCT
 ATCACATTCCTCAAATCCTTTTTTTTTATTTCTTTTTCCATGTCCATGCACCCATATTAGA
 CATGGCCCAA
 AATATGTGATTTAATTCCTCCCCAGTAATGCTGGGCACCCTAATACCACTCCTTCCTTC
 AGTGCCAAGAA
 CAACTGCTCCCAAATGTTTACCAGCTTTCCTCAGCATCTGAATTGCCTTTGAGATTAA
 TTAAGCTAAAA
 GCATTTTTATATGGGAGAATATTATCAGCTTGTCOAAGCAAAAATTTAAATGTGAAAA
 ACAAATTGTGT
 CTTAAGCATTTTTGAAAATTAAGGAAGAAGAATTTGGGAAAAAATTAACGGTGGCTCAA 30
 TTCTGTCTTCC
 AAATGATTTCTTTCCCTCCTACTCACATGGGTCGTAGGCCAGTGAATACATTCAACAT
 GGTGATCCCA
 GAAACTCAGAGAAGCCTCGGCTGATGATTAATTAATTGATCTTTCCGCTACCCGAGA
 GAATTACATTT
 CCAAGAGACTTCTTCACCAAATCCAGATGGGTTTACATAAACTTCTGCCACGGGTAT
 CTCCTCTCTCC
 TAACACGCTGTGACGTCTGGGCTTGGTGAATCTCAGGGAAGCATCCGTGGGGTGAAG
 GTCATCGTCTG
 GCTCGTTGTTTGATGGTTATATTACCATGCAATTTCTTTGCCTACATTTGTATTGAATA 40
 CATCCCAATC
 TCCTTCCTATTCGGTGACATGACACATTCTATTTCAGAAGGCTTTGATTTTATCAAGCA
 CTTTCATTTAC
 TTCTCATGGCAGTGCCTATTACTTCTCTTACAATAACCCATCTGTCTGCTTTACCAAATC
 TATTTCCCT
 TTTCAGATCCTCCCAAATGGTCCTCATAAACTGTCTGCCTCCACCTAGTGGTCCAGGT
 ATATTTCCACA
 ATGTTACATCAACAGGCACTTCTAGCCATTTTCTTCTCAAAGGTGCAAAAAGCAACT
 TCATAAACACA
 AATTAATCTTCGGTGAGGTAGTGTGATGCTGCTTCTCCCAACTCAGCGCACTTCGTC 50

TTCCTCATTCC
 ACAAAAACCCATAGCCTTCCTTCACTCTGCAGGACTAGTGCTGCCAAGGGTTCAGCTCT
 ACCTACTGGTG
 TGCTCTTTTGGAGCAAGTTGCTTAGCCTCTCTGTAACACAAGGACAATAGCTGCAAGCAT
 CCCCAAAGATC
 ATTGCAGGAGACAATGACTAAGGCTACCAGAGCCGCAATAAAAGTCAGTGAATTTTAGC
 GTGGTCCTCTC
 TGTCTCTCCAGAACGGCTGCCACGTGGAATTGCTCTTCCTCCGCTACATCTCGGACTGG
 GACCTAGACCC
 TGGCCGCTGCTACCGCGTCACCTGGTTACCTCCTGGAGCCCCTGCTACGACTGTGCC 10
 GACATGTGGCC
 GACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCCTCTACTT
 CTGTGAGGACC
 GCAAGGCTGAGCCCGAGGGGCTGCGGCGGCTGCACCGCGCCGGGGTGCAAATAGCCATC
 ATGACCTTCAA
 AGGTGCGAAAGGGCCTTCCGCGCAGGCGCAGTGCAGCAGCCCGCATTTCGGGATTGCGAT
 GCGGAATGAAT
 GAGTTAGTGGGGAAGCTCGAGGGGAAGAAGTGGGCGGGGATTCTGGTTCACCTCTGGAG
 CCGAAATTTAA
 GATTAGAAGCAGAGAAAAGAGTGAATGGCTCAGAGACAAGGCCCCGAGGAAATGAGAA 20
 AATGGGGCCAGG
 GTTGCTTCTTTCCCTCGATTTGGAACCTGAACTGTCTTCTACCCCATATCCCGCCTT
 TTTTTCCTTT
 TTTTTTTTTTGAAGATTATTTTTACTGCTGGAATACTTTTGTAGAAAACCACGAAAGAA
 CTTTCAAAGCC
 TGGGAAGGGCTGCATGAAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTTTT
 GGTAAGGGGCT
 TCCTCGCTTTTTAAATTTTCTTTCTTTCTCTACAGTCTTTTTTGGAGTTTCGTATATTT
 TTATATTTTC
 TTATTGTTCAATCACTCTCAGTTTTTCATCTGATGAAAACCTTTATTTCTCCTCCACATCAG 30
 CTTTTTCTTC
 TGCTGTTTCACCATTTCAGAGCCCTCTGCTAAGGTTCTTTTTCCCTCCCTTTTTCTTTCTTT
 TGTTGTTTCA
 CATCTTTAAATTTCTGTCTCTCCCCAGGGTTGCGTTTTCTTCTGTCAGAAATCTTTTC
 TCCTTTTTTTT
 TTTTTTTTTTTTTTTTTTTTTTAAACAAACAAACAAAAAACCCTTTTCCCAAT
 TTACTTTCTT
 CCAACATGTTACAAAGCCATCCACTCAGTTTAGAAGACTCTCCGGCCCCACCGACCCCC
 AACCTCGTTTT
 GAAGCCATTCACTCAATTTGCTTCTCTTTTCTCTACAGCCCCTGTATGAGGTTGATGAC 40
 TTACGAGACG
 CATTTCTGACTTTGGGACTTTGATAGCAACTTCCAGGAATGTCACACACGATGAAATAT
 CTCTGCTGAAG
 ACAGTGGATAAAAAACAGTCCTTCAAGTCTTCTCTGTTTTTATTCTTCAACTCTCACTTT
 CTTAGAGTTT
 ACAGAAAAAATATTTATATACGACTCTTTAAAAAGATCTATGTCTTGAAAATAGAGAAG
 GAACACAGGTC
 TGGCCAGGGACGTGCTGCAATTGGTGCAGTTTTGAATGCAACATTGTCCCCTACTGGGA
 ATAACAGAACT
 GCAGGACCTGGGAGCATCCTAAAGTGTCACAGTTTTTCTATGACTTTTGGTAGGATGA 50

GAGCAGAAGGT
AGATCCTAAAAAGCATGGTGAGAGGATCAAATGTTTTTATATCAACATCCTTTATTATT
TGATTCATTTG
AGTTAACAGTGGTGTTAGTGATAGATTTTTCTATTCTTTTCCCTTGACGTTTACTTTCAA
GTAACACAAA
CTCTTCCATCAGGCCATGATCTATAGGACCTCCTAATGAGAGTATCTGGGTGATTGTGA
CCCCAAACCAT
CTCTCAAAGCATTAAATATCCAATCATGCGCTGTATGTTTTAATCAGCAGAAGCATGTT
TTTATGTTTGT
ACAAAAGAAGATTGTTATGGGTGGGGATGGAGGTATAGACCATGCATGGTCACCTTCAA 10
GCTACTTTAAT
AAAGGATCTTAAAATGGGCAGGAGGACTGTGAACAAGACACCCTAATAATGGGTGATG
TCTGAAGTAGC
AAATCTTCTGGAAACGCAAACCTCTTTTAAGGAAGTCCCTAATTTAGAAACACCCACAAA
CTTCACATATC
ATAATTAGCAAACAATTGGAAGGAAGTTGCTTGAATGTTGGGGAGAGGAAAATCTATTG
GCTCTCGTGGG
TCTCTTCATCTCAGAAATGCCAATCAGGTCAAGGTTTGCTACATTTTGTATGTGTGTGA
TGCTTCTCCCA
AAGGTATATTA ACTATATAAGAGAGTTGTGACAAAACAGAATGATAAAGCTGCGAACCG 20
TGGCACACGCT
CATAGTTCTAGCTGCTTGGGAGGTTGAGGAGGGAGGATGGCTTGAACACAGGTGTTCAA
GGCCAGCCTGG
GCAACATAACAAGATCCTGTCTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAGAGAGAGGG
CCGGGCGTGGTG
GCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGCCGGGCGGATCACCTGTGGTCA
GGAGTTTGAGA
CCAGCCTGGCCAACATGGCAAACCCCGTCTGTACTCAAAATGCAAAAATTAGCCAGGC
GTGGTAGCAGG
CACCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCAGGAGG 30
TGGAGGTTGCA
GTAAGCTGAGATCGTGCCGTTGCACTCCAGCCTGGGCGACAAGAGCAAGACTCTGTCTC
AGAAAAAAAAA
AAAAAAGAGAGAGAGAGAGAAAGAGAACAATATTTGGGAGAGAAGGATGGGGAAGCA
TTGCAAGGAAAT
TGTGCTTTATCCAACAAAATGTAAGGAGCCAATAAGGGATCCCTATTTGTCTCTTTTGG
TGTCTATTTGT
CCCTAACAACTGTCTTTGACAGTGAGAAAAATATTCAGAATAACCATATCCCTGTGCCG
TTATTACCTAG
CAACCCTTGCAATGAAGATGAGCAGATCCACAGGAAAACCTTGAATGCACAACTGTCTTA 40
TTTTAATCTTA
TTGTACATAAGTTTGTAAGAGGTTAAAAATTGTTACTTCATGTATTCATTTATATTTT
ATATTATTTTGT
CGTCTAATGATTTTTTATTAACATGATTTTCTTTTCTGATATATTGAAATGGAGTCTCAA
AGCTTCATAA
ATTTATAACTTTAGAAATGATTCTAATAACAACGTATGTAATTGTAACATTGCAGTAAT
GGTGCTACGAA
GCCATTTCTCTTGATTTTTAGTAACTTTTATGACAGCAAATTTGCTTCTGGCTCACTTT
CAATCAGTTA
AATAAATGATAAATAATTTTGGAAAGCTGTGAAGATAAAATACCAAATAAAATAATATAA 50

AAGTGATTTAT
ATGAAGTTAAAATAAAAAATCAGTATGATGGAATAAACTTG

【 0 2 7 8 】

本開示の態様に従ってCas9に融合され得る他の例示的デアミナーゼが、以下に提供される。いくつかの実施形態において、それぞれの配列の活性ドメイン、例えば、局在化シグナル(核局在化配列、核外搬出シグナル、細胞質局在化シグナル)を伴わないドメインが使用されることが理解されるべきである。

【 0 2 7 9 】

Human AID:

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNNGCH
VELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGPNLSLRIFTARLYF
CEDRKAPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLS
RQLRRILLPLYEVDDLRLDAFRTLGL

10

(下線 : 核局在化配列 ; 二重下線 : 核外輸送シグナル)

【 0 2 8 0 】

Mouse AID:

MDSLLMKQKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSCSLDFGHLRNKSGCH
VELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWPNLSLRIFTARLYF
CEDRKAPEGLRRLHRAGVQIGIMTFKDYFYCWNTFVENRERTFKAWEGLHENSVRILT
RQLRRILLPLYEVDDLRLDAFRMLGF

20

(下線 : 核局在化配列 ; 二重下線 : 核外輸送シグナル)

【 0 2 8 1 】

Dog AID:

MDSLLMKQKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGCHV
ELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAARLYFC
EDRKAPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENREKTFKAWEGLHENSVRLSR
QLRRILLPLYEVDDLRLDAFRTLGL

30

(下線 : 核局在化配列 ; 二重下線 : 核外輸送シグナル)

【 0 2 8 2 】

Bovine AID:

MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGCHV
ELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFC
DKERKAPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLS
RQLRRILLPLYEVDDLRLDAFRTLGL

40

(下線 : 核局在化配列 ; 二重下線 : 核外輸送シグナル)

【 0 2 8 3 】

Rat AID

50

MAVGSKPKAALVGPHERERIWCFLCSTGLGTQQTGQTSRWLRPAATQDPVSPRSL
 MKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGYLRNKSCHVELLFL
 RYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGPNLSLRIFTARLTGWGALP
 AGLMSPARPSDYFYCWNTFVENHERTFKAWEGLHENSVRLSRRLRILLPLYEVDL
DAFRTLGL

(下線 : 核局在化配列 ; 二重下線 : 核外輸送シグナル)

【 0 2 8 4 】

10

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSL
 HHGVFKNKDNIAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHH
 NLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFR
 PWKRLLTNFRYQDSKLQEILRPCYIPVSSSSSTLSNICLTKGLPETRFCVEGRRMDPLSE
 EEFYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGGQHAEIFLDKI
 RSMELSQVITTCYLTWSPCPNCAWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSL
 WQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLRRIKESWGLQDL
 VNDFGNLQLGPPMS

20

(斜体 : 核酸編集ドメイン)

【 0 2 8 5 】

Rat APOBEC-3:

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNRLRYAIDRKDTFLCYEVTRKDCDSPVSL
 HHGVFKNKDNIAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQVLRFLATH
 HNLSLDIFSSRLYNIRDPENQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFR
 PWKLLTNFRYQDSKLQEILRPCYIPVSSSSSTLSNICLTKGLPETRFCVERRRVHLLSE
 EEFYSQFYNQRVKHLCYYHGVKPYLCYQLEQFNGQAPLKGCLLSEKGGQHAEIFLDKI
 RSMELSQVITTCYLTWSPCPNCAWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSL
 WQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLHRIKESWGLQDL
 VNDFGNLQLGPPMS

30

(斜体 : 核酸編集ドメイン)

【 0 2 8 6 】

Rhesus macaque APOBEC-3G:

40

MVEPMDPRTFVSNFNRPILSGLNTVWLCCEVKTKDPSGPPLDAKIFQGKVYSKAKYH
PEMRFLRWFHKWRQLHHDQEYKVTWYVSWSPCTRCANSVATFLAKDPKVTLTIFVARLY
 YFWKPDYQQALRILCQKRGGPHATMKIMNYNEFQDCWNKFVDGRGKPKPRNNLPKH
 YTLLQATLGELLRHLMDPGTFTSNFNKPVVSGQHETYLCYKVERLHNDTWVPLNQH
 RGFLRNQAPNIHGFPKGRHAELCFDLIPFWKLDGQQYRVTCFTSWSPCFSCAQEMAKFIS
 NNEHVSLCIFAARIYDDQGRYQEGLRALHRDGAKIAMMNYSEFEYCWDTFVDRQGRPF
 QPWDGLDEHSQALSGRLRAI

10

(斜体：核酸編集ドメイン；下線：細胞質局在化シグナル)

【0287】

Chimpanzee APOBEC-3G:

MKPHFRNPVERMYQDTFSDNFYNRPILSHRNTVWLCYEVKTKGSPRPPLDAKIFRGQV
YSKLYHPEMRFFHWFSKWRKLHRDQEYEVWYISWSPCTKCTRDVATFLAEDPKVTLTIF
 FVARLYYFWDPDYQEALRSLCQKRDGPRAATMKIMNYDEFQHCWSKFVYSQRELFEPW
 NNLPKYYILLHIMLGEILRHSMDPPTFTSNFNELWVRGRHETYLCYEVEERLHNDTWV
 LNQRGFLCNQAPHKHGFLEGRHAELCFDLVIPFWKLDLHQDYRVTCFTSWSPCFSCAQE
 MAKFISNNKHVSLCIFAARIYDDQGRQCQEGLRTLAKAGAKISIMTYSEFKHCWDTFVDH
 QGCPFQPWDGLEEHSQALSGRLRAILQNQGN

20

(斜体：核酸編集ドメイン；下線：細胞質局在化シグナル)

【0288】

Green monkey APOBEC-3G:

MNPQIRNMVEQMEPDIFVYYFNRPILSGRNTVWLCYEVKTKDPSGPPLDANIFQGKLY
PEAKDHPEMKFLHWFWRKWRQLHRDQEYEVWYVSWSPCTRCANSVATFLAEDPKVTLTIF
 VARLYYFWKPDYQQALRILCQERGGPHATMKIMNYNEFQHCWNEFVDGQGKPKPRK
 NLPKHYTLLHATLGELLRHVMDPGTFTSNFNKPVVSGQRETYLCYKVERSHNDTWV
 LLNQHRGFLRNQAPDRHGFPKGRHAELCFDLIPFWKLDGQQYRVTCFTSWSPCFSCAQK
 MAKFISNNKHVSLCIFAARIYDDQGRQCQEGLRTLHRDGAKIAVMNYSEFEYCWDTFVD
 RQGRPFQPWDGLDEHSQALSGRLRAI

30

(斜体：核酸編集ドメイン；下線：細胞質局在化シグナル)

【0289】

Human APOBEC-3G:

40

50

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQV
YSELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLTI
 FVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPW
 NNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWV
 LLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQ
 EMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVD
 HQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN

10

(斜体 : 核酸編集ドメイン ; 下線 : 細胞質局在化シグナル)

【 0 2 9 0 】

Human APOBEC-3F:

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQV
 YSQPEHHAEMCFLSWFCGNQLPAYKCFQITWFWVSWTPCPCVAKLAEFLAEHPNVTLTIS
 AARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYSEGQPFMPWYKFD
 DNYAFLHRTLKEILRNPMEAMYPHIFYHFKNLRKAYGRNESWLCFTMEVVKHHSPVS
 WKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCECAGEVAEF
 LARHSNVNLTIFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCWENFVYND
 DEPFKPWKGLKYNFLFLDSKLQEILE

20

(斜体 : 核酸編集ドメイン)

【 0 2 9 1 】

Human APOBEC-3B:

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGQ
 VYFKPQYHAEMCFLSWFCGNQLPAYKCFQITWFWVSWTPCPCVAKLAEFLSEHPNVTLTI
 SAARLYYYWERDYRRALCRLSQAGARVTIMDYEEFAYCWENFVYNEGQQFMPWYKF
 DENYAFLHRTLKEILRYLMDPDTFTFNFNNDPLVLRRRQTYLCYEVERLDNGTWVLM
 QHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFWISWSPCFSWGCAGE
 VRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVY
 RQGCPFQPWDGLEEHSQALSGLRAILQNQGN

30

(斜体 : 核酸編集ドメイン)

【 0 2 9 2 】

Rat APOBEC-3B:

MQPQGLGNAGMGPVCLGCSHRRPYSPIRNPLKKLYQQTFYFHFKNVRYAWGRKNNFL
 CYEVNGMDCALPVPLRQGVFRKQGHHAELCFIYWFHDKVLRVLSPEEFKVTWYMSW
 SPCSKCAEQVARFLAAHRNLSLAIFSSRLYYYLRNPNYQQKLCRLIQEGVHVAAMDPEF
 KKCWNKFVDNDGQPFRPWMLRINFSFYDCKLQEIFSRMNLREDVFYLQFNNSHRVKP
 VQNRYYRRKSYLCYQLERANGQEPLKGYLLYKKGEQHVEILFLEKMRSMELSQVRITCYL
 TWSPCPNCARQLAAFKKDHDPDLILRIYTSRLYFWRKKFQKGLCTLWRSGLHVDVMDLPQ
 FADCWTNFVNPQRPFPRWNELEKNSWRIQRRLRRIKESWGL

40

50

【 0 2 9 3 】

Bovine APOBEC-3B:

DGWEVAFRSGTVLKAGVLGVSMTEGWAGSGHPGGACVWTPGTRNTMNULLREVLFKQ
QFGNQPRVPAPYYRRKTYLCYQLKQRNDLTLDRGCFRNKKQRHAERFIDKINSLDLNPS
QSYKIICYITWSPCPNCANELVNFITRNNHLKLEIFASRLYFHWIKSFKMGLQDLQONAGIS
VAVMTHTEFEDCWEQFVDNQSRPFQPWDKLEQYSASIRRRLLQRILTAPI

【 0 2 9 4 】

Chimpanzee APOBEC-3B:

MNPQIRNPMEMYQRTFYFNENEPILYGRSYTWLCYEYVKIRRGHSNLLWDTGVFRGQ
MYSQPEHHAEMCFLSWFCGNQLSAYKCFQITWVFSWTPCPDCVAKLAKFLAEHPNVTL
TISAARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYNEGQPFMPWYKF
DDNYAFLHRTLKEIIRHLMDDPTFTFNFNNDPLVLRRHQTYLCYEYVERLDNGTWVLM
QHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTFWISWSPCFSWGCAG
QVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVY
RQGC PFQPWDGLEEHSQALSGRRLRAILQVRASSLCMVPHRPPPPQSPGPCPLPCSEPL
GSLLPTRPAPSLPFLLTASFSPPPASLPPLPSLSLSPGHLVPSFHSLSLSCSIQPPCSSRI
RETEGWASVSKEGRDLG

10

【 0 2 9 5 】

Human APOBEC-3C:

MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRN
QVDSETH*CHAERCFLSWFCDDILSPNTKYQVTWYTSWSPC*PDCAGEVAEFLARHSNVNLT
IFTARLYYFQYPCYQEGRLSLSQEGVAVEIMDYEDFKYCWENFVYNDNEPFKPKWKGLK
TNFRLLKRRLRESLQ

20

【 0 2 9 6 】

Gorilla APOBEC-3C

MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRN
QVDSETH*CHAERCFLSWECDDILSPNTNYQVTWYTSWSPCPEC*AGEVAEFLARHSNVNLT
FTARLYYFQDQDYQEGRLSLSQEGVAVKIMDYKDFKYCWENFVYNDDEPFKPKWKGLK
YNFRFLKRRLQEILE

30

【 0 2 9 7 】

Human APOBEC-3A:

MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEYVERLDNGTSVKMDQHRGFLHNQ
AKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTFWISWSPCFSWGCAGEVRAFLQENT
HVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQCPFQP
WDGLDEHSQALSGRRLRAILQNQGN

40

(斜体 : 核酸編集ドメイン)

【 0 2 9 8 】

Rhesus macaque APOBEC-3A:

50

MDGSPASRPRHLMDPNTFTFNFNNDLSVRGRHQTYLCYEVERLDNGTWVPMDERRGF
 LCNKAKNVPCGDYGCHVELRFLCEVPSWQLDPAQTYRVTWFWISWSPCFRRGCAGQVRVF
 LQENKHVRLRIFAARIYDYDPLYQEALRTRLRDAGAQVSIMTYEEFKHCWDTFVDRQGR
 PFQPWDGLDEHSQALSGRLRAILQNQGN

(斜体 : 核酸編集ドメイン)

【 0 2 9 9 】

Bovine APOBEC-3A:

10

MDEYTFTENFNQGWPSKTYLCYEMERLDGDATIPLDEYKGFVRNKGLDQPEKPC~~HAEL~~
~~LYFLGKIHSWNLDRNQHYRLTCFISWSPCYDCAQKLTTFLENHHISLHILASRIYTHNRF~~
 CHQSGLC~~ELQAAGARITIMTFEDFKHCWETFDHKGKPFQPWEGLNVKSQALCTELQA~~
 ILKTQQN

(斜体 : 核酸編集ドメイン)

【 0 3 0 0 】

MALLTAETFRLQFNNKRRRLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKK~~CHAEICF~~
~~INEIKSMGLDETQCYQVTCYLTWSPCSSCAWELVDFIKAHDHLNLGIFASRLYYHWCKPQ~~
 QKGLRLLCGSQVPVEVMGFPKFADCWENFVDHEKPLSFNPYKMLEELDKNSRAIKRRL
 ERIKIPGVRAQGRYMDILCDAEV

20

(斜体 : 核酸編集ドメイン)

【 0 3 0 1 】

Rhesus macaque APOBEC-3H:

30

MALLTAKTFLQFNNKRRVVKPYYPRKALLCYQLTPQNGSTPTRGHLKNNKKDHAEIRFI
 NKIKSMGLDETQCYQVTCYLTWSPCPCAGELVDFIKAHRHLNLRIFASRLYYHWRPNY
 QEGLLLCGSQVPVEVMGLPEFTDCWENFVDHKEPPSFNPSEKLEELDKNSQAIKRRLER
 IKRSVDVLENGLRSLQLGPVTPSSSIRNSR

【 0 3 0 2 】

Human APOBEC-3D:

40

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGP
 VLPKRQSNHRQEVYFRFEN~~HAEMCFLSWFCGNRLPANRRFQITWVSWNPCLPCVVKVT~~
 KFLAHPNVTLTISAARLYYYRDRDWRWVLLRLHKAGARVKIMDYEDFAYCWENFVC
 NEGQPFMPWYKFDDNYASLHRTLKEILRNPMEAMYPHIFYHFKNLLKACGRNESWLC
 FTMEVTKHHS~~AVFRKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSP~~
~~CPECAGEVAEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGASVKIMGYKDFV~~
 SCWKNFVYSDDEPFKPKWGLQTNFRLKRLREILQ

(斜体 : 核酸編集ドメイン)

【 0 3 0 3 】

Human APOBEC-1:

50

MTSEKGPSTGDPTLRRRIEPWFDVFDYDPRELRKEACLLYEIKWGM~~SRKIWRSSGKNTTN~~

HVEVNFIIKFTSERDFHPSMSCSITWFLSWSPCWECQAIREFLSRHPGVTLVIYVARLF
 WHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPGDEAHWPQYPPLWMM
 LYALELHCIILSLPPCLKISRWRQNHLTFFRLHLQNCHYQTIPPHILLATGLIHPSVAWR
 【 0 3 0 4 】

Mouse APOBEC-1:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSVWRHTSQNTSN
 HVEVNFLEKFTTERYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPYVTLFIYIARLYH
 HTDQRNRQGLRDLISSGVTIQIMTEQEYCYCWRNFVNYPSPNEAYWPRYPHLWVKLYV
 LELYCIILGLPPCLKILRRKQPQLTFFTITLQTCHYQRIPPHLLWATGLK
 【 0 3 0 5 】

10

Rat APOBEC-1:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
 HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
 ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSNEAHWPRYPHLWVRLYVLE
 LYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK
 【 0 3 0 6 】

Human APOBEC-2:

MAQKEEAAVATEAASQNGEDLENLDDPEKLELIEPPEIVTGERLPANFFKFQFRNVE
 YSSGRNKTFLCYVVEAQQKGGQVQASRGYLEDEHAAHAAEEAFFNTILPAFDPALRYNVT
 WYVSSSPCAACADRIIKTSLKTKNLRLLILVGRLFMWEEPEIQAALKKLKEAGCKLRIMKP
 QDFEYVWQNFVEQEEGESKAFQPWEDIQENFLYEEKLADILK
 【 0 3 0 7 】

20

Mouse APOBEC-2:

MAQKEEAAEAAAPASQNGDDLENLEDPEKLELIDLPPFEIVTGVRLPVNFFKFQFRNVE
 YSSGRNKTFLCYVVEVQSKGGQAQATQGYLEDEHAGAHAAEEAFFNTILPAFDPALKYNVT
 WYVSSSPCAACADRILKTKNLRLLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMK
 PQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYEEKLADILK
 【 0 3 0 8 】

Rat APOBEC-2:

MAQKEEAAEAAAPASQNGDDLENLEDPEKLELIDLPPFEIVTGVRLPVNFFKFQFRNVE
 YSSGRNKTFLCYVVEAQS KGGQVQATQGYLEDEHAGAHAAEEAFFNTILPAFDPALKYNVT
 WYVSSSPCAACADRILKTKNLRLLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMK
 PQDFEYLWQNFVEQEEGESKAFEPWEDIQENFLYEEKLADILK
 【 0 3 0 9 】

30

Bovine APOBEC-2:

MAQKEEAAAAAEPASQNGEEVENLEDPEKLELIEPPEIVTGERLPAHYFKFQFRNVE
 YSSGRNKTFLCYVVEAQS KGGQVQASRGYLEDEHATNHAAEEAFFNSIMPTFDPALRYMV
 TWYVSSSPCAACADRIVKTLNKTKNLRLLILVGRLFMWEEPEIQAALRKLKEAGCRLRIM
 KPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYEEKLADILK
 【 0 3 1 0 】

40

Petromyzon marinus CDA1 (pmCDA1):

MTDAEYVRIHEKLDIYTFKKQFFNNKKS VSHRCYVLFELKRRGERRACFWGYAVNK PQS
 GTERGIHAEIFSIRKVEEYLRDNPQGFTINWYSSWSPCADCAEKILEWYNQELRG NGHTL
 KIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHQ LNENRWL
 EKTLEKRAEKRRSELSFMIQVKILHTTKSPAV
 【 0 3 1 1 】

Human APOBEC3G D316R D317R:

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGSPRPPLDAKIFRGQ VY
 SELKYHPEMRFFHWFSKWRKLRDQEYEV TWYISWSPCTKCTRD MATFLAEDP KVTLT
 IFVARLYYFWDPDYQEALRSLCQKRDGPRATMKF NYDEFQHCWSKFVYSQ RELFEPW

50

NNLPKYYILLHFMLGEILRHSMDPPTFTFNFNNEPWVRRHETYLCEYEVER MHNDTWV
 LLNQRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTC FTSWSPCFSCA
 QEMAKFISK KHVSLCIFTARIYRRQGRQCQGLRRTLAEAGAKISF T YSEFKHCWDTFVDH
 QGCPFQPWDGLDEHSQDLSGRLRAILQNQEN

【0312】

Human APOBEC3G chain A:

MDPPTFTFNFNNEPWVRRHETYLCEYEVERMHNDTWVLLNQRGFLCNQAPHKHG FL
 EGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKKNKHVSLCI FTAR
 IYDDQGRQCQGLRRTLAEAGAKISF TYSEFKHCWDTFVDHQGCPFQPWDGLD EHSQDLS
 GRLRAILQ

10

【0313】

Human APOBEC3G chain A D120R D121R:

MDPPTFTFNFNNEPWVRRHETYLCEYEVERMHNDTWVLLNQRGFLCNQAPHKHG FL
 EGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKKNKHVSLCI FTAR
 IYRRQGRQCQGLRRTLAEAGAKISFMTYSEFKHCWDTFVDHQGCPFQPWDGLDE HSQDL
 SGRLRAILQ

【0314】

本開示のいくつかの態様は、例えばデアミナーゼドメインに点突然変異を生じさせること
 によって、本明細書に記載される融合タンパク質のいずれかのデアミナーゼドメイン触
 媒活性を調節することが、融合タンパク質(例えば塩基エディター)のプロセス特性に影響
 を及ぼすという認識に基づく。例えば、塩基編集融合タンパク質内のデアミナーゼドメイ
 ンの触媒活性を減少させるが除去はしない突然変異は、デアミナーゼドメインが標的残基
 に隣接する残基の脱アミノ化を触媒する可能性を低くし、それによって脱アミノ化のウィ
 ンドウを狭めることができる。脱アミノ化のウィンドウを狭める能力は、特定の標的残基
 に隣接する残基の望まれない脱アミノ化を防ぐことができ、それは、オフターゲット効果
 を減少させるか、または防ぐことができる。

20

【0315】

例えば、いくつかの実施形態において、塩基エディターに組み込まれたAPOBECデアミ
 ナーゼは、rAPOBEC1のH121X、H122X、R126X、R126X、R118X、W90X、W90X
 、およびR132Xからなる群より選択される1以上の突然変異、または別のAPOBECデアミ
 ナーゼにおける1以上の対応する突然変異を含むことができ、ここでXは任意のアミノ酸で
 ある。ある態様において、塩基エディターに組み込まれたAPOBECデアミナーゼは、rAP
 OBEC1のH121R、H122R、R126A、R126E、R118A、W90A、W90Y、およびR132
 Eからなる群より選択される1以上の突然変異、または別のAPOBECデアミナーゼにおける
 1以上の対応する突然変異を含むことができる。

30

【0316】

ある態様において、塩基エディターに組み込まれたAPOBECデアミナーゼは、hAPOBE
 C3GのD316X、D317X、R320X、R320X、R313X、W285X、W285X、R326Xから
 なる群より選択される1以上の突然変異、または別のAPOBECデアミナーゼにおける1以上
 の対応する突然変異を含むことができ、ここでXは任意のアミノ酸である。いくつかの実
 施形態において、本明細書で提供される融合タンパク質のいずれかは、hAPOBEC3GのD3
 16R、D317R、R320A、R320E、R313A、W285A、W285Y、R326Eからなる群から
 選択される1以上の突然変異、または別のAPOBECデアミナーゼにおける1以上の対応する
 突然変異を含むAPOBECデアミナーゼを含む。

40

【0317】

ある態様において、塩基エディターに組み込まれたAPOBECデアミナーゼは、rAPOBE
 C1のH121RおよびH122R突然変異、または別のAPOBECデアミナーゼにおける1つ以上
 の対応する突然変異を含むことができる。いくつかの実施形態において、塩基エディター
 に組み込まれたAPOBECデアミナーゼは、rAPOBEC1のR126A突然変異、または別のAP
 OBECデアミナーゼにおける1つ以上の対応する突然変異を含むAPOBECデアミナーゼを含

50

ミナーゼは、hAPOBEC3GのR320EおよびR326E突然変異、または別のAPOBECデアミナーゼにおける1つ以上の対応する突然変異を含むAPOBECデアミナーゼを含むことができる。いくつかの実施形態において、塩基エディターに組み込まれたAPOBECデアミナーゼは、hAPOBEC3GのW285YおよびR326E突然変異、または別のAPOBECデアミナーゼにおける1つ以上の対応する突然変異を含むAPOBECデアミナーゼを含むことができる。いくつかの実施形態において、塩基エディターに組み込まれたAPOBECデアミナーゼは、hAPOBEC3GのW285Y、R320E、およびR326E突然変異、または別のAPOBECデアミナーゼにおける1つ以上の対応する突然変異を含むAPOBECデアミナーゼを含むことができる。

【0319】

限定されるものではないが、SaBE3、SaKKH-BE3、VQR-BE3、EQR-BE3、VRER-BE3、YE1-BE3、EE-BE3、YE2-BE3、およびYE-BE3を含む、多数の改変シチジンデアミナーゼがAddgeneから市販されている(プラスミド85169、85170、85171、85172、85173、85174、85175、85176、85177)。

【0320】

CからTへの核酸塩基編集タンパク質の詳細は、国際PCT出願番号PCT/US2016/058344 (WO2017/070632) およびKomor, A.C., et al., "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016) に記載されており、これらの全内容は参照により本明細書に組み入れられる。

【0321】

[AからGへの編集]

いくつかの実施形態において、本明細書に記載される塩基エディターは、アデノシンデアミナーゼを含むデアミナーゼドメインを含むことができる。塩基エディターのこのようなアデノシンデアミナーゼドメインは、Aを脱アミノ化して、Gの塩基対形成特性を示すイノシン (I) を形成することによって、アデニン (A) 核酸塩基からグアニン (G) 核酸塩基への編集することを促進することができる。アデノシンデアミナーゼは、デオキシリボ核酸 (DNA) 中のデオキシアデノシン残基のアデニンを脱アミノ化すること(すなわちアミン基を除去すること)ができる。

【0322】

いくつかの実施形態において、本明細書において提供される核酸塩基エディターは、1つ以上のタンパク質ドメインを融合させ、それによって融合タンパク質を生成することによって作製することができる。ある態様において、本明細書において提供される融合タンパク質は、融合タンパク質の塩基編集活性(例えば効率、選択性、特異性)を改善する一つ以上の特徴を含む。例えば、本明細書中に提供される融合タンパク質は、ヌクレアーゼ活性が低下したCas9ドメインを含むことができる。いくつかの実施形態において、本明細書中で提供される融合タンパク質は、ヌクレアーゼ活性を有さないCas9ドメイン (dCas9)、または、Cas9ニッカーゼ (nCas9) と呼ばれる、二本鎖DNA分子の一鎖を切断するCas9ドメインを有することができる。いかなる特定の理論にも拘束されることを望まないが、触媒残基(例えばH840)の存在が、標的Aと反対のTを含有する非編集(例えば非脱アミノ化)鎖を切断するCas9の活性を維持する。Cas9の触媒残基の突然変異(例えばD10からA10)は、標的A残基を含有する編集鎖の切断を妨げる。このようなCas9バリエーションは、gRNAによって規定される標的配列に基づく特定の位置で単一鎖DNA切断(ニック)を生じさせることができ、非編集鎖の修復を導き、最終的には編集鎖においてTからCへの変化をもたらす。ある態様において、AからGへの塩基エディターは、イノシン塩基除去修復の阻害因子、例えば、ウラシルグリコシラーゼ阻害因子 (UGI) ドメインまたは触媒的に不活性なイノシン特異的ヌクレアーゼをさらに含む。いかなる特定の理論にも束縛されることを望まないが、UGIドメインまたは触媒的に不活性なイノシン特異的ヌクレアーゼは、脱アミノ化されたアデノシン残基(例えばイノシン)の塩基除去修復を阻害または防止することができる。これが、塩基エディターの活性または効率を改善させることができる。

10

20

30

40

50

【0323】

アデノシンデアミナーゼを含む塩基エディターは、DNA、RNAおよびDNA-RNAハイブリッドを含む任意のポリヌクレオチドに作用することができる。特定の実施形態では、アデノシンデアミナーゼを含む塩基エディターは、RNAを含むポリヌクレオチドの標的Aを脱アミノ化することができる。例えば、塩基エディターは、RNAポリヌクレオチドおよび/またはDNA-RNAハイブリッドポリヌクレオチドの標的Aを脱アミノ化することができるアデノシンデアミナーゼドメインを含むことができる。一実施形態によると、塩基エディターに組み込まれたアデノシンデアミナーゼは、RNAに作用するアデノシンデアミナーゼ(ADAR、例えばADAR1またはADAR2)の全部または一部を含む。別の実施形態では、塩基エディターに組み込まれたアデノシンデアミナーゼは、tRNAに作用するアデノシンデアミナーゼ(ADAT)の全部または一部を含む。アデノシンデアミナーゼドメインを含む塩基エディターはまた、DNAポリヌクレオチドのA核酸塩基を脱アミノ化する能力も有し得る。一実施形態では、塩基エディターのアデノシンデアミナーゼドメインは、1以上の突然変異を含むADATの全部または一部を含み、これがDNA中の標的AをADATが脱アミノ化することを可能にする。例えば、塩基エディターは、D108N、A106V、D147Y、E155V、L84F、H123Y、I157F、または別のアデノシンデアミナーゼにおける対応する突然変異の1以上を含む大腸菌由来のADAT (EcTadA)の全部または一部を含むことができる。

10

【0324】

アデノシンデアミナーゼは、任意の適切な生物(例えば大腸菌)に由来することができる。ある態様において、アデニンデアミナーゼは、天然に存在するアデノシンデアミナーゼが本明細書に提供される突然変異のいずれかに対応する一つ以上の突然変異を含むものである(例えば、ecTadAにおける突然変異)。任意の相同タンパク質中の対応する残基は、例えば、配列アラインメントおよび相同残基の決定によって同定することができる。本明細書中に記載された突然変異のいずれか(例えば、ecTadAで同定された突然変異のいずれか)に対応する、任意の天然アデノシンデアミナーゼ(例えば、ecTadAに対して相同性を有するもの)における突然変異を、それに応じて生成することができる。

20

【0325】

特定の実施形態では、TadAは、その全体が参照により本明細書に組み込まれるPCT/US 2017/045381(国際公開第2018/027078号)に記載されているTadAのいずれかである。

【0326】

ある実施形態において、アデノシンデアミナーゼは、アミノ酸配列：
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGS
LMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAKKAQSS
TD
を含み、これは「TadA参照配列」と呼ばれる。

30

【0327】

いくつかの実施形態において、TadAデアミナーゼは前腸E. coli TadAデアミナーゼである。例えば、ある実施形態において、アデノシンデアミナーゼは、アミノ酸配列：
MRRAFITGVFFLSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNR
PIGRHDPTAHAEIMALRQGGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGA
RDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAKKAQSS
TD
を含む。

40

【0328】

しかしながら、本願で有用になるさらなるアデノシンデアミナーゼは当業者にとって明らかであり本開示の範囲内に入ることが理解されるべきである。例えば、アデノシンデアミナーゼは、tRNAに作用するアデノシンデアミナーゼ(ADAT)のホモログであり得る。限定されないが、例示的なADATホモログのアミノ酸配列は以下のものを含む：

【0329】

Staphylococcus aureus TadA:

50

MGSHMTNDIYFMTLAIIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAHAEH
IAIERAAKVLGSRLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCS GS LMN
LLQQS NFNHRAIVDKG VLKE AC S TLLTTFKnlRANKKS TN
【 0 3 3 0 】

Bacillus subtilis TadA:

MTQDELYMKEAIKEAKKAEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEML VID
EACKALGTWRLEGATLYVTLEPCVMCAGAVVLSRVEKVVFGAFDPKGGC S GTLMN LL
QEERFNHQAEVVSGVLEEECGGMLSAFFRELKkkKAARKNLSE
【 0 3 3 1 】

Salmonella typhimurium (*S. typhimurium*) TadA:

MPPAFITGVTSLSDVELDHEYWMRHALTLAKRAWDEREVPVGAVLVHNHRVIGEGWNR
PIGRHDPTAHAEIMALRQGGLVLQNYRLLDttLYVTLEPCVMCAGAMVHSRIGRVVFGA
RDAKTGAAGSLIDVLHHPGMNHRVEIIEGVLrDECATLLSDFFRMRRQEIK ALKKADRA
EGAGPAV
【 0 3 3 2 】

Shewanella putrefaciens (*S. putrefaciens*) TadA:

MDE YWMQVAMQM AEKAEAGE VPVGA VLVKDGQQIATGYNLS IS QHDPT AHAEI
LCLRSAGKKLENYRLLDAtLYITLEPCAMCAGAMVHSRIARVVYGARDEKTGAAGTVVN
LLQHAPFNHQVEVTSGVLAeACSAQLSRFFKRRRDEKKALKLAQRAQQGIE
【 0 3 3 3 】

Haemophilus influenzae F3031 (*H. influenzae*) TadA:

MDAAKVRSEFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQSDPT
AEIIALRNGAKNIQNYRLLNSTLYVTLEPCTMCAGAILHS RIKRLVFG ASDYKTGAI
GSRFHFFDDYKMNHTLEITSGVLAEECSQKLSTFFQKRREEKKIEKALLKSLSDK
【 0 3 3 4 】

Caulobacter crescentus (*C. crescentus*) TadA:

MRTDESEDQDHRMMRLALDAARAAAEAGETPVGAVILDPSTGEVIATAGNGPIAAHDPT
AHAEIAAMRAAAAKLGNyRLTDLTlVVTLEPCAMCAGAISHARIGRVVFGADD PKGGAV
VHGPKFFAQPTCHWRPEVTGGVLADESADLLRGFFRARRKAKI
【 0 3 3 5 】

Geobacter sulfurreducens (*G. sulfurreducens*) TadA:

MSSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRGHNLREGSNDPSA
HAEMIAIRQAARRSANWRLTGATLYVTLEPCLMCMGAILARLERVVFGCYDPKGGGAAGS
LYDLSADPRLNHQVRLSPGVCQEecGTMLSDFFRDLRRRKKAKATPALF IDERKVPPEP
【 0 3 3 6 】

TadA7.10:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNrvIGEGWNRAIGLHDPTAHA
EIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSL
MDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD
【 0 3 3 7 】

ある態様において、アデノシンデアミナーゼは、原核生物由来である。ある態様において、アデノシンデアミナーゼは、細菌由来である。ある態様において、アデノシンデアミナーゼは、*Escherichia coli*、*Staphylococcus aureus*、*Salmonella typhi*、*Shewanella putrefaciens*、*Haemophilus influenzae*、*Caulobacter crescentus*、または*Bacillus subtilis*に由来する。ある態様において、アデノシンデアミナーゼは、大腸菌由来である。

【 0 3 3 8 】

1つの実施形態において、本発明の融合タンパク質は、TadA7.10に連結された野生型TadAがCas9ニッカーゼに連結されたものを含む。特定の実施形態では、融合タンパク質は、単一のTadA7.10ドメイン(例えば、モノマーとして提供される)を含む。他の実施形態

10

20

30

40

50

では、ABE7.10エディターは、TadA7.10およびTadA(wt) を含み、これらはヘテロ二量体を形成することができる。

【0339】

ある態様において、アデノシンデアミナーゼは、本明細書に提供されるアデノシンデアミナーゼのいずれかに記載のアミノ酸配列のいずれかに対して少なくとも60%、少なくとも65%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、または少なくとも99.5%の同一性であるアミノ酸配列を含む。本明細書において提供されるアデノシンデアミナーゼは、一つ以上の突然変異(例えば、本明細書に提供される突然変異のいずれか)を含み得ることが理解されるべきである。本開示は、特定のパーセント同一性を有するデアミナーゼドメインが加えて本明細書に記載される突然変異のいずれかまたはその組合せを含むものを提供する。いくつかの実施形態において、アデノシンデアミナーゼは、参照配列または本明細書に提供されるアデノシンデアミナーゼのいずれかと比較して1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、またはそれ以上の変異を有するアミノ酸配列を含む。いくつかの実施形態において、アデノシンデアミナーゼは、当該技術分野において既知であるかまたは本明細書に記載されたアミノ酸配列のいずれかと比較して、少なくとも5個、少なくとも10個、少なくとも15個、少なくとも20個、少なくとも25個、少なくとも30個、少なくとも35個、少なくとも40個、少なくとも45個、少なくとも50個、少なくとも60個、少なくとも70個、少なくとも80個、少なくとも90個、少なくとも100個、少なくとも110個、少なくとも120個、少なくとも130個、少なくとも140個、少なくとも150個、少なくとも160個、または少なくとも170個の同一の連続するアミノ酸残基を有するアミノ酸配列を含む。

【0340】

いくつかの実施形態では、アデノシンデアミナーゼは、TadA参照配列に対してD108Xの変異、または別のアデノシンデアミナーゼにおける対応する変異を含み、ここでXは野生型アデノシンデアミナーゼにおける対応するアミノ酸以外の任意のアミノ酸を示す。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてD108G、D108N、D108V、D108A、もしくはD108Yの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。しかし、さらなるデアミナーゼを同様に配列させて本明細書で提供したように変異させることのできる相同のアミノ酸残基を同定することができることを認識されたい。

【0341】

いくつかの実施形態では、アデノシンデアミナーゼは、TadA参照配列においてA106Xの変異、または別のアデノシンデアミナーゼにおける対応する変異を含み、ここでXは野生型アデノシンデアミナーゼにおける対応するアミノ酸以外の任意のアミノ酸を示す。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてA106Vの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。

【0342】

いくつかの実施形態では、アデノシンデアミナーゼは、TadA参照配列においてE155Xの変異、または別のアデノシンデアミナーゼにおける対応する変異を含み、ここでXの存在は野生型アデノシンデアミナーゼにおける対応するアミノ酸以外の任意のアミノ酸を示す。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてE155D、E155G、もしくはE155Vの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。

【0343】

いくつかの実施形態では、アデノシンデアミナーゼは、TadA参照配列においてD147Xの変異、または別のアデノシンデアミナーゼにおける対応する変異を含み、ここでXの存在は野生型アデノシンデアミナーゼにおける対応するアミノ酸以外の任意のアミノ酸を示

10

20

30

40

50

す。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてD147Yの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。

【0344】

本明細書で提供する変異（たとえばTadA参照配列のアミノ酸配列に基づくもの）のいずれも、*S. aureus* TadA (saTadA) またはその他のアデノシンデアミナーゼ（たとえば細菌のアデノシンデアミナーゼ）等の他のアデノシンデアミナーゼに導入できることが理解されるべきである。TadA参照配列において同定されるいずれの変異も、相同的なアミノ酸残基を有する他のアデノシンデアミナーゼにおいて作ることができる。本明細書で提供する変異のいずれも、TadA参照配列または別のアデノシンデアミナーゼにおいて個別にまたは任意の組合せで作成し得ることも理解されるべきである。

10

【0345】

たとえば、アデノシンデアミナーゼは、TadA参照配列に対してD108N、A106V、E155V、および/もしくはD147Yの変異、または別のアデノシンデアミナーゼにおける対応する変異を含み得る。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列に対して以下の変異の群（変異の群を「;」で分離する）または別のアデノシンデアミナーゼにおける対応する変異を含む。D108NおよびA106V; D108NおよびE155V; D108NおよびD147Y; A106VおよびE155V; A106VおよびD147Y; E155VおよびD147Y; D108N、A106VおよびE155V; D108N、A106VおよびD147Y; D108N、E155VおよびD147Y; A106V、E155VおよびD147Y; ならびにD108N、A106V、E155VおよびD147Y。しかし、本明細書で提供する対応する変異の任意の組合せをアデノシンデアミナーゼにおいて作成してよいことを認識されたい（たとえばecTadA）。

20

【0346】

いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列に対してH8X、T17X、L18X、W23X、L34X、W45X、R51X、A56X、E59X、E85X、M94X、I95X、V102X、F104X、A106X、R107X、D108X、K110X、M118X、N127X、A138X、F149X、M151X、R153X、Q154X、I156X、および/もしくはK157Xの変異の1つもしくは複数、または別のアデノシンデアミナーゼにおける1つもしくは複数の対応する変異を含み、ここでXの存在は野生型アデノシンデアミナーゼにおける対応するアミノ酸以外の任意のアミノ酸を示す。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列に対してH8Y、T17S、L18E、W23L、L34S、W45L、R51H、A56E、またはA56S、E59G、E85K、またはE85G、M94L、I95I、V102A、F104L、A106V、R107C、またはR107H、またはR107P、D108G、またはD108N、またはD108V、またはD108A、またはD108Y、K110I、M118K、N127S、A138V、F149Y、M151V、R153C、Q154L、I156D、および/もしくはK157Rの変異のうち1つもしくは複数、または別のアデノシンデアミナーゼにおける1つもしくは複数の対応する変異を含む。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列に対してH8X、D108X、および/またはN127Xの変異の1つもしくは複数、または別のアデノシンデアミナーゼにおける1つもしくは複数の対応する変異を含み、ここでXは任意のアミノ酸の存在を示す。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列に対してH8Y、D108N、および/もしくはN127Sの変異の1つもしくは複数、または別のアデノシンデアミナーゼにおける1つもしくは複数の対応する変異を含む。

30

40

【0347】

いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列に対してH8X、R26X、M61X、L68X、M70X、A106X、D108X、A109X、N127X、D147X、R152X、Q154X、E155X、K161X、Q163X、および/もしくはT166X変異のうち1つもしくは複数、または別のアデノシンデアミナーゼにおける1つもしくは複数の対応する変異を含み、ここでXは野生型アデノシンデアミナーゼにおける対応するアミノ酸以外の任意のアミノ酸の存在を示す。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列に対してH8Y、R26W、M61I、L68Q、M70V、A106T、D108N、A109T、N127S、D147Y、R152C、Q154HもしくはQ154R、E155GもしくはE155VもしくはE155D、K

50

161Q、Q163H、および/もしくはT166P変異のうち1つもしくは複数、または別のアデノシンデアミナーゼにおける1つもしくは複数の対応する変異を含む。

【0348】

いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列に対してH8X、D108X、N127X、D147X、R152X、およびQ154Xからなる群から選択される1つ、2つ、3つ、4つ、5つ、または6つの変異、または別のアデノシンデアミナーゼにおける対応する変異を含み、ここでXは野生型アデノシンデアミナーゼにおける対応するアミノ酸以外の任意のアミノ酸の存在を示す。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列に対してH8X、M61X、M70X、D108X、N127X、Q154X、E155X、およびQ163Xからなる群から選択される1つ、2つ、3つ、4つ、5つ、6つ、7つ、もしくは8つの変異、または別のアデノシンデアミナーゼにおける対応する変異を含み、ここでXは野生型アデノシンデアミナーゼにおける対応するアミノ酸以外の任意のアミノ酸の存在を示す。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列に対してH8X、D108X、N127X、E155X、およびT166Xからなる群から選択される1つ、2つ、3つ、4つ、もしくは5つの変異、または別のアデノシンデアミナーゼにおける対応する変異を含み、ここでXは野生型アデノシンデアミナーゼにおける対応するアミノ酸以外の任意のアミノ酸の存在を示す。

10

【0349】

いくつかの実施形態では、アデノシンデアミナーゼはH8X、A106X、D108Xからなる群から選択される1つ、2つ、3つ、4つ、5つ、または6つの変異、別のアデノシンデアミナーゼにおける変異を含み、ここでXは野生型アデノシンデアミナーゼにおける対応するアミノ酸以外の任意のアミノ酸の存在を示す。いくつかの実施形態では、H8X、R126X、L68X、D108X、N127X、D147X、およびE155Xからなる群から選択される1つ、2つ、3つ、4つ、5つ、6つ、7つ、もしくは8つの変異、または別のアデノシンデアミナーゼにおける対応する変異を含み、ここでXは野生型アデノシンデアミナーゼにおける対応するアミノ酸以外の任意のアミノ酸の存在を示す。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列に対してH8X、D108X、A109X、N127X、およびE155Xからなる群から選択される1つ、2つ、3つ、4つ、もしくは5つの変異、または別のアデノシンデアミナーゼにおける対応する変異を含み、ここでXは野生型アデノシンデアミナーゼにおける対応するアミノ酸以外の任意のアミノ酸の存在を示す。

20

30

【0350】

いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列に対してH8Y、D108N、N127S、D147Y、R152C、およびQ154Hからなる群から選択される1つ、2つ、3つ、4つ、5つ、もしくは6つの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列に対してH8Y、M61I、M70V、D108N、N127S、Q154R、E155G、およびQ163Hからなる群から選択される1つ、2つ、3つ、4つ、5つ、6つ、7つ、もしくは8つの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列に対してH8Y、D108N、N127S、E155V、およびT166Pからなる群から選択される1つ、2つ、3つ、4つ、もしくは5つの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列に対してH8Y、A106T、D108N、N127S、E155D、およびK161Qからなる群から選択される1つ、2つ、3つ、4つ、5つ、もしくは6つの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列に対してH8Y、R126W、L68Q、D108N、N127S、D147Y、およびE155Vからなる群から選択される1つ、2つ、3つ、4つ、5つ、6つ、7つ、もしくは8つの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列に対してH8Y、D108N、A109T、N127S、およびE155Gからなる群から選択される1つ、2つ、3つ、4つ、もしくは5つの変異、または別のアデノシンデアミナーゼにおける対応

40

50

する変異を含む。

【0351】

本明細書で提供する変異のいずれも、およびさらなる任意の変異（たとえばTadA参照配列アミノ酸配列に基づく）は、他の任意のアデノシンデアミナーゼに導入することができる。本明細書で提供する変異のいずれも、TadA参照配列または別のアデノシンデアミナーゼにおいて個別にまたは任意の組合せで作成することができる。

【0352】

AからGへの核酸塩基編集タンパク質の詳細は、その全体の内容が参照により本明細書に組み込まれる国際PCT出願PCT/2017/045381 (WO2018/027078)およびGaudelli, N.M., et al., “Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage” Nature, 551, 464-471 (2017)に記載されている。

【0353】

いくつかの実施形態では、アデノシンデアミナーゼは、別のアデノシンデアミナーゼにおける変異の1つもしくは複数を含む。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてD108N、D108G、もしくはD108Vの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてA106VおよびD108Nの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列におけるR107CおよびD108Nの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてH8Y、D108N、N127S、D147Y、およびQ154Hの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてH8Y、R24W、D108N、N127S、D147Y、およびE155Vの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてD108N、D147Y、およびE155Vの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてH8Y、D108N、およびS127Sの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてA106V、D108N、D147Y、およびE155Vの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。

【0354】

いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてS2X、H8X、I49X、L84X、H123X、N127X、I156Xおよび/またはK160Xの変異のうち1つもしくは複数、または別のアデノシンデアミナーゼにおける1つもしくは複数の対応する変異を含み、ここでXの存在は野生型アデノシンデアミナーゼにおける対応するアミノ酸以外の任意のアミノ酸を示す。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてS2A、H8Y、I49F、L84F、H123Y、N127S、I156Fおよび/またはK160Sの変異のうち1つもしくは複数、または別のアデノシンデアミナーゼにおける1つもしくは複数の対応する変異を含む。

【0355】

いくつかの実施形態では、アデノシンデアミナーゼはL84X変異アデノシンデアミナーゼを含み、ここでXは野生型アデノシンデアミナーゼにおける対応するアミノ酸以外の任意のアミノ酸の存在を示す。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてL84Fの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。

【0356】

いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてH123Xの変異、または別のアデノシンデアミナーゼにおける対応する変異を含み、ここでXは野生型アデノシンデアミナーゼにおける対応するアミノ酸以外の任意のアミノ酸を示す。いく

10

20

30

40

50

つかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてH123Yの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてI157Xの変異、または別のアデノシンデアミナーゼにおける対応する変異を含み、ここでXは野生型アデノシンデアミナーゼにおける対応するアミノ酸以外の任意のアミノ酸を示す。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてI157Fの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。

【0357】

いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてL84X、A106X、D108X、H123X、D147X、E155X、およびI156Xからなる群から選択される1つ、2つ、3つ、4つ、5つ、6つ、もしくは7つの変異、または別のアデノシンデアミナーゼにおける対応する変異を含み、ここでXは野生型アデノシンデアミナーゼにおける対応するアミノ酸以外の任意のアミノ酸の存在を示す。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてS2X、I49X、A106X、D108X、D147X、およびE155Xからなる群から選択される1つ、2つ、3つ、4つ、5つ、もしくは6つの変異、または別のアデノシンデアミナーゼにおける対応する変異を含み、ここでXは野生型アデノシンデアミナーゼにおける対応するアミノ酸以外の任意のアミノ酸の存在を示す。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてH8X、A106X、D108X、N127X、およびK160Xからなる群から選択される1つ、2つ、3つ、4つ、もしくは5つの変異、または別のアデノシンデアミナーゼにおける対応する変異を含み、ここでXは野生型アデノシンデアミナーゼにおける対応するアミノ酸以外の任意のアミノ酸の存在を示す。

【0358】

いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてL84F、A106V、D108N、H123Y、D147Y、E155V、およびI156Fからなる群から選択される1つ、2つ、3つ、4つ、5つ、6つ、もしくは7つの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてS2A、I49F、A106V、D108N、D147Y、およびE155Vからなる群から選択される1つ、2つ、3つ、4つ、5つ、もしくは6つの変異を含む。

【0359】

いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてH8Y、A106T、D108N、N127S、およびK160Sからなる群から選択される1つ、2つ、3つ、4つ、もしくは5つの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。

【0360】

いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてE25X、R26X、R107X、A142X、および/またはA143Xのうち1つもしくは複数の変異、または別のアデノシンデアミナーゼにおける1つもしくは複数の対応する変異を含み、ここでXの存在は野生型アデノシンデアミナーゼにおける対応するアミノ酸以外の任意のアミノ酸を示す。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてE25M、E25D、E25A、E25R、E25V、E25S、E25Y、R26G、R26N、R26Q、R26C、R26L、R26K、R107P、R07K、R107A、R107N、R107W、R107H、R107S、A142N、A142D、A142G、A143D、A143G、A143E、A143L、A143W、A143M、A143S、A143Qおよび/またはA143Rのうち1つもしくは複数の変異、または別のアデノシンデアミナーゼにおける1つもしくは複数の対応する変異を含む。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列に対応する本明細書に記載した変異のうち1つもしくは複数の、または別のアデノシンデアミナーゼにおける1つもしくは複数の対応する変異を含む。

【0361】

いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてE25Xの変異、または別のアデノシンデアミナーゼにおける対応する変異を含み、ここでXは野生型

10

20

30

40

50

アデノシンデアミナーゼにおける対応するアミノ酸以外の任意のアミノ酸を示す。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてE25M、E25D、E25A、E25R、E25V、E25S、もしくはE25Yの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。

【0362】

いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてR26Xの変異、または別のアデノシンデアミナーゼにおける対応する変異を含み、ここでXは野生型アデノシンデアミナーゼにおける対応するアミノ酸以外の任意のアミノ酸を示す。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてR26G、R26N、R26Q、R26C、R26L、もしくはR26Kの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。

10

【0363】

いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてR107Xの変異、または別のアデノシンデアミナーゼにおける対応する変異を含み、ここでXは野生型アデノシンデアミナーゼにおける対応するアミノ酸以外の任意のアミノ酸を示す。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてR107P、R07K、R107A、R107N、R107W、R107H、もしくはR107Sの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。

【0364】

いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてA142Xの変異、または別のアデノシンデアミナーゼにおける対応する変異を含み、ここでXは野生型アデノシンデアミナーゼにおける対応するアミノ酸以外の任意のアミノ酸を示す。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてA142N、A142D、A142Gの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。

20

【0365】

いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてA143Xの変異、または別のアデノシンデアミナーゼにおける対応する変異を含み、ここでXは野生型アデノシンデアミナーゼにおける対応するアミノ酸以外の任意のアミノ酸を示す。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてA143D、A143G、A143E、A143L、A143W、A143M、A143S、A143Qおよび/もしくはA143Rの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。

30

【0366】

いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてH36X、N37X、P48X、I49X、R51X、M70X、N72X、D77X、E134X、S146X、Q154X、K157X、および/もしくはK161Xのうち1つもしくは複数の変異、または別のアデノシンデアミナーゼにおける対応する1つもしくは複数の変異を含み、ここでXの存在は野生型アデノシンデアミナーゼにおける対応するアミノ酸以外の任意のアミノ酸を示す。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてH36L、N37T、N37S、P48T、P48L、I49V、R51H、R51L、M70L、N72S、D77G、E134G、S146R、S146C、Q154H、K157N、および/もしくはK161Tの1つもしくは複数の変異、または別のアデノシンデアミナーゼにおける1つもしくは複数の対応する変異を含む。

40

【0367】

いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてH36Xの変異、または別のアデノシンデアミナーゼにおける対応する変異を含み、ここでXは野生型アデノシンデアミナーゼにおける対応するアミノ酸以外の任意のアミノ酸を示す。いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてH36Lの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。

【0368】

いくつかの実施形態では、アデノシンデアミナーゼはTadA参照配列においてN37Xの変異、または別のアデノシンデアミナーゼにおける対応する変異を含み、ここでXは野生型

50

52Hの変異、または別のアデノシンデアミナーゼにおける対応する変異を含む。

【 0 3 7 7 】

一実施形態では、アデノシンデアミナーゼは、突然変異H36L、R51L、L84F、A106V、D108N、H123Y、S146C、D147Y、E155V、I156F、およびK157Nを含むことができる。一部の実施形態では、アデノシンデアミナーゼは、tadA基準配列に対して突然変異の以下の組合せを含み、ここで、組合せの各突然変異は「_」によって分離され、突然変異の各組合せは括弧内にある：

(A106V_D108N), (R107C_D108N), (H8Y_D108N_N127S_D147Y_Q154H), (H8Y_R24W_D108N_N127S_D147Y_E155V), (D108N_D147Y_E155V), (H8Y_D108N_N127S), (H8Y_D108N_N127S_D147Y_Q154H), (A106V_D108N_D147Y_E155V) (D108Q_D147Y_E155V) (D108M_D147Y_E155V), (D108L_D147Y_E155V), (D108K_D147Y_E155V), (D108I_D147Y_E155V), (D108F_D147Y_E155V), (A106V_D108N_D147Y), (A106V_D108M_D147Y_E155V), (E59A_A106V_D108N_D147Y_E155V), (E59A cat dead_A106V_D108N_D147Y_E155V), (L84F_A106V_D108N_H123Y_D147Y_E155V_I156Y), (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F), (D103A_D104N), (G22P_D103A_D104N), (G22P_D103A_D104N_S138 A), (D103 A_D104N_S138 A), (R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F), (E25G_R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F), (E25D_R26G_L84F_A106V_R107K_D108N_H123Y_A142N_A143G_D147Y_E155V_I156F), (R26Q_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F), (E25M_R26G_L84F_A106V_R107P_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F), (R26C_L84F_A106V_R107H_D108N_H123Y_A142N_D147Y_E155V_I156F), (L84F_A106V_D108N_H123Y_A142N_A143L_D147Y_E155V_I156F), (R26G_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F), (E25A_R26G_L84F_A106V_R107N_D108N_H123Y_A142N_A143E_D147Y_E155V_I156F), (R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F), (A106V_D108N_A142N_D147Y_E155V), (R26G_A106V_D108N_A142N_D147Y_E155V), (E25D_R26G_A106V_R107K_D108N_A142N_A143G_D147Y_E155V), (R26G_A106V_D108N_R107H_A142N_A143D_D147Y_E155V), (E25D_R26G_A106V_D108N_A142N_D147Y_E155V), (A106V_R107K_D108N_A142N_D147Y_E155V), (A106V_D108N_A142N_A143G_D147Y_E155V), (A106V_D108N_A142N_A143L_D147Y_E155V), (H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N), (N37T_P48T_M70L_L84F_A106V_D108N_H123Y_D147Y_I49V_E155V_I156F), (N37S_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K161T), (H36L_L84F_A106V_D108N_H123Y_D147Y_Q154H_E155V_I156F),

(N72S_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F),
(H36L_P48L_L84F_A106V_D108N_H123Y_E134G_D147Y_E155V_I156F),
(H36L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N), (H36L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F),
(L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T),
(N37S_R51H_D77G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(R51L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N),
(D24G_Q71R_L84F_H96L_A106V_D108N_H123Y_D147Y_E155V_I156F_K160E),
(H36L_G67V_L84F_A106V_D108N_H123Y_S146T_D147Y_E155V_I156F), 10
(Q71L_L84F_A106V_D108N_H123Y_L137M_A143E_D147Y_E155V_I156F),
(E25G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L),
(L84F_A91T_F104I_A106V_D108N_H123Y_D147Y_E155V_I156F),
(N72D_L84F_A106V_D108N_H123Y_G125A_D147Y_E155V_I156F),
(P48S_L84F_S97C_A106V_D108N_H123Y_D147Y_E155V_I156F),
(W23G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(D24G_P48L_Q71R_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L),
(L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),
(H36L_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N), (N37S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_K161T), 20
(L84F_A106V_D108N_D147Y_E155V_I156F),
(R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K161T),
(L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K161T),
(L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E_K161T),
(L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E), (R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F), 30
(L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(R74A_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(L84F_R98Q_A106V_D108N_H123Y_D147Y_E155V_I156F),
(L84F_A106V_D108N_H123Y_R129Q_D147Y_E155V_I156F),
(P48S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F), (P48S_A142N),
(P48T_I49V_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_L157N), 40
(P48T_I49V_A142N),
(H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N),
(H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F (H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N),
(H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F 50

_K157N),
 (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155
 V_I156F _K157N),
 (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155
 V_I156F _K157N),
 (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V
 _I156F _K157N),
 (W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V
 _I156F _K157N),
 (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V 10
 _I156F _K161T),
 (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152H_E155
 V_I156F _K157N),
 (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V
 _I156F _K157N),
 (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P
 _E155V _I156F _K157N),
 (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y
 _E155V
 _I156F _K157N), 20
 (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y
 _R152P _E155V_I156F _K157N),
 (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V
 _I156F _K161T),
 (W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P
 _E155V _I156F _K157N),
 (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_R152
 P_E155V
 _I156F _K157N).

【 0 3 7 8 】 30

特定の実施形態では、本明細書で提供される融合タンパク質は、融合タンパク質の塩基
 編集活性を改善する1つ以上の特徴を含む。例えば、本明細書に提供される融合タンパク
 質は、ヌクレアーゼ活性が低下したCas9ドメインを含み得る。いくつかの実施形態におい
 て、本明細書で提供される融合タンパク質は、ヌクレアーゼ活性を有さないCas9ドメイン
 (dCas9)、またはCas9ニッカーゼ (nCas9) と呼ばれる、二本鎖DNA分子の1鎖を切断す
 るCas9ドメインを有し得る。

【 0 3 7 9 】

[シチジンデアミナーゼ]

【 0 3 8 0 】

1つの実施形態において、本発明の融合タンパク質は、シチジンデアミナーゼを含む。 40
 ある態様において、本明細書において提供されるシチジンデアミナーゼは、シトシンまた
 は5-メチルシトシンを脱アミノ化してウラシルまたはチミンにすることができる。いくつ
 かの実施形態において、本明細書に提供されるシチジンデアミナーゼは、DNA中のシトシ
 ンを脱アミノ化することができる。シチジンデアミナーゼは、任意の適切な生物に由来す
 ることができる。いくつかの実施形態において、シチジンデアミナーゼは、天然に存在す
 るシチジンデアミナーゼが本明細書に提供される突然変異のいずれかに対応する1つ以上
 の突然変異を含むところのものである。当業者は、例えば、配列アラインメントおよび相
 同的残基の決定によって、任意の相同的タンパク質中の対応する残基を同定することがで
 きる。従って、当業者は、本明細書に記載された突然変異のいずれかに対応する突然変異
 を、任意の天然に存在するシチジンデアミナーゼにおいて生じさせることができる。 ある 50

態様において、シチジンデアミナーゼは、原核生物由来である。ある態様において、シチジンデアミナーゼは、細菌由来である。ある態様において、シチジンデアミナーゼは、哺乳動物(例えば、ヒト)由来である。

【0381】

いくつかの実施形態において、シチジンデアミナーゼは、本明細書に記載されるシチジンデアミナーゼアミノ酸配列のいずれかに対して少なくとも60%、少なくとも65%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、または少なくとも99.5%の同一性であるアミノ酸配列を含む。本明細書において提供されるシチジンデアミナーゼは、一つ以上の突然変異(例えば、本明細書に提供される突然変異のいずれか)を含み得ることを理解されたい。本開示は、特定のパーセント同一性を有する任意のデアミナーゼドメインに、本明細書に記載される突然変異またはその組合せのいずれかが加えられたものを提供する。いくつかの実施形態において、シチジンデアミナーゼは、参照配列または本明細書に提供されるいずれかのシチジンデアミナーゼと比較して、1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50個またはそれ以上の変異を有するアミノ酸配列を含む。いくつかの実施形態において、シチジンデアミナーゼは、当技術分野において既知であるかまたは本明細書に記載されたアミノ酸配列のいずれかと比較して、少なくとも5、少なくとも10、少なくとも15、少なくとも20、少なくとも25、少なくとも30、少なくとも35、少なくとも40、少なくとも45、少なくとも50、少なくとも60、少なくとも70、少なくとも80、少なくとも90、少なくとも100、少なくとも110、少なくとも120、少なくとも130、少なくとも140、少なくとも150、少なくとも160、または少なくとも170個の同一の連続するアミノ酸残基を有するアミノ酸配列を含む。

【0382】

本発明の融合タンパク質は、核酸編集ドメインを含む。ある態様において、核酸編集ドメインは、CからUへの塩基変化を触媒することができる。ある態様において、核酸編集ドメインは、デアミナーゼドメインである。ある態様において、デアミナーゼは、シチジンデアミナーゼまたはアデノシンデアミナーゼである。ある態様において、デアミナーゼは、アポリポタンパク質B mRNA編集複合体(APOBEC)ファミリーデアミナーゼである。ある態様において、デアミナーゼは、APOBEC1デアミナーゼである。ある態様において、デアミナーゼは、APOBEC2デアミナーゼである。ある態様において、デアミナーゼは、APOBEC3デアミナーゼである。ある態様において、デアミナーゼは、APOBEC3Aデアミナーゼである。ある態様において、デアミナーゼは、APOBEC3Bデアミナーゼである。ある態様において、デアミナーゼは、APOBEC3Cデアミナーゼである。ある態様において、デアミナーゼは、APOBEC3Dデアミナーゼである。ある態様において、デアミナーゼは、APOBEC3Eデアミナーゼである。ある態様において、デアミナーゼは、APOBEC3Fデアミナーゼである。ある態様において、デアミナーゼは、APOBEC3Gデアミナーゼである。ある態様において、デアミナーゼは、APOBEC3Hデアミナーゼである。ある態様において、デアミナーゼは、APOBEC4デアミナーゼである。ある態様において、デアミナーゼは、活性化誘導デアミナーゼ(AID)である。ある態様において、デアミナーゼは、脊椎動物デアミナーゼである。ある態様において、デアミナーゼは、無脊椎動物デアミナーゼである。ある態様において、デアミナーゼは、ヒト、チンパンジー、ゴリラ、サル、ウシ、イヌ、ラット、またはマウスデアミナーゼである。ある態様において、デアミナーゼは、ヒトデアミナーゼである。ある態様において、デアミナーゼは、ラットデアミナーゼ、例えば、rAPOBEC1である。ある態様において、デアミナーゼは、Petromyzon marinusシチジンデアミナーゼ1(pmCDA1)である。ある態様において、デアミナーゼは、ヒトAPOBEC3Gである。ある態様において、デアミナーゼは、ヒトAPOBEC3Gの断片である。ある態様において、デアミナーゼは、D316R D317R突然変異を含むヒトAPOBEC3Gバリエーションである。ある態様において、デアミナーゼは、ヒトAPOBEC3Gの断片であり、D316R

D317R突然変異に対応する突然変異を含む。ある態様において、核酸編集ドメインは、本明細書に記載される任意のデアミナーゼのデアミナーゼドメインに対して少なくとも80%、少なくとも85%、少なくとも90%、少なくとも92%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、または少なくとも99.5%の同一性である。

【0383】

[核酸塩基エディターのCas9ドメイン]

【0384】

いくつかの態様では、核酸プログラム可能なDNA結合タンパク質(napDNAbp)はCas9ドメインである。非限定的で例示的なCas9ドメインを本明細書で提供する。Cas9ドメインはヌクレアーゼ活性Cas9ドメイン、ヌクレアーゼ不活性Cas9ドメイン、またはCas9ニッカーゼであってよい。いくつかの実施形態では、Cas9ドメインはヌクレアーゼ活性ドメインである。たとえば、Cas9ドメインは二本鎖核酸の両方の鎖(たとえば二本鎖DNA分子の両方の鎖)を切断するCas9ドメインであってよい。いくつかの実施形態では、Cas9ドメインは本明細書で説明するアミノ酸配列のいずれか1つを含む。いくつかの実施形態では、Cas9ドメインは、本明細書で説明するアミノ酸配列のいずれか1つに対して少なくとも60%、少なくとも65%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、少なくとも99.5%の同一性であるアミノ酸配列を含む。いくつかの実施形態では、Cas9ドメインは、本明細書で説明するアミノ酸配列のいずれか1つと比較して1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、またはそれを超える変異を有するアミノ酸配列を含む。いくつかの実施形態では、Cas9ドメインは、本明細書で説明するアミノ酸配列のいずれか1つと比較して少なくとも10、少なくとも15、少なくとも20、少なくとも30、少なくとも40、少なくとも50、少なくとも60、少なくとも70、少なくとも80、少なくとも90、少なくとも100、少なくとも150、少なくとも200、少なくとも250、少なくとも300、少なくとも350、少なくとも400、少なくとも500、少なくとも600、少なくとも700、少なくとも800、少なくとも900、少なくとも1000、少なくとも1100、または少なくとも1200の同一の連続的なアミノ酸残基を有するアミノ酸配列を含む。

【0385】

ある態様において、Cas9ドメインは、ヌクレアーゼ不活性Cas9ドメイン(dCas9)である。例えば、dCas9ドメインは、二本鎖核酸分子のいずれの鎖も切断することなく、二本鎖核酸分子に結合し得る(例えば、gRNA分子を介して)。いくつかの実施形態において、ヌクレアーゼ不活性dCas9ドメインは、本明細書中に記載されたアミノ酸配列のD10X突然変異およびH840X突然変異、または本明細書中に提供されたアミノ酸配列のいずれかにおける対応する突然変異を含み、Xは任意のアミノ酸変化である。いくつかの実施形態において、ヌクレアーゼ不活性dCas9ドメインは、本明細書に記載のアミノ酸配列のD10A突然変異およびH840A突然変異、または本明細書に記載のアミノ酸配列のいずれかにおける対応する突然変異を含む。

【0386】

ある態様において、Cas9ドメインは、Cas9ニッカーゼである。Cas9ニッカーゼは、二本鎖核酸分子(例えば二本鎖DNA分子)の一方の鎖のみを切断することができるCas9タンパク質であり得る。いくつかの実施形態において、Cas9ニッカーゼは、二本鎖核酸分子の標的鎖を切断し、これは、Cas9ニッカーゼが、Cas9に結合しているgRNA(例えばsgRNA)と塩基対を形成している(相補的である)鎖を切断することを意味する。ある態様において、Cas9ニッカーゼは、D10A突然変異を含み、位置840にヒスチジンを有する。いくつかの実施形態において、Cas9ニッカーゼは、二本鎖核酸分子の非標的、非塩基編集鎖を切断し、これは、Cas9ニッカーゼが、Cas9に結合しているgRNA(例えばsgRNA)と塩

10

20

30

40

50

基対を形成していない鎖を切断することを意味する。ある態様において、Cas9ニッカーゼは、H840A突然変異を含み位置10にアスパラギン酸残基を有するか、または対応する突然変異。いくつかの実施形態において、Cas9ニッカーゼは、本明細書に提供されるCas9ニッカーゼのいずれかと少なくとも60%、少なくとも65%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、または少なくとも99.5%の同一性であるアミノ酸配列を含む。さらなる適切なCas9ニッカーゼは、本開示および当該分野における知識に基づいて当業者に明らかであり、本開示の範囲内である。

【0387】

[排他性が低減されたCas9ドメイン]

典型的には、*S. pyogenes*由来のCas9 (spCas9) などのCas9タンパク質は、特定の核酸領域に結合するために標準的なNGG PAM配列を必要とし、ここで「NGG」の「N」はアデノシン (A)、チミジン (T) またはシトシン (C) であり、Gはグアノシンである。これは、ゲノム内の所望の塩基を編集する能力を制限し得る。いくつかの実施形態において、本明細書に提供される塩基編集融合タンパク質は、正確な位置、例えばPAMの上流にある標的塩基を含む領域に配置することが必要になり得る。例えばKomor, A.C., et al., “Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage” *Nature* 533, 420-424 (2016)参照 (これらの内容全体は、参照により本明細書に組み込まれる)。従って、いくつかの実施形態において、本明細書で提供される融合タンパク質のいずれかは、標準的 (例えばNGG) PAM配列を含まないヌクレオチド配列に結合することができるCas9ドメインを含み得る。非標準的PAM配列に結合するCas9ドメインは本技術分野において記述されており当業者には明らかであろう。例えば、非標準PAM配列に結合するCas9ドメインは、Kleinstiver, B. P., et al., “Engineered CRISPR-Cas9 nucleases with altered PAM specificities” *Nature* 523, 481-485 (2015); およびKleinstiver, B. P., et al., “Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition” *Nature Biotechnology* 33, 1293-1298 (2015); Nishimasu, H., et al., “Engineered CRISPR-Cas9 nuclease with expanded targeting space” *Science*. 2018 Sep 21;361(6408):1259-1262, Chatterjee, P., et al., “Minimal PAM specificity of a highly similar SpCas9 ortholog” *Sci Adv*. 2018 Oct 24;4(10):eaau0766. doi: 10.1126/sciadv.aau0766に記述されており、それぞれの全内容を参照によりここに組み込む。次の表1に、いくつかのPAMバリエーションを示す。

10

20

30

40

50

【表 1】

バリエント	PAM
spCas9	NGG
spCas9-VRQR	NGA
spCas9-VRER	NGCG
xCas9 (sp)	NGN
saCas9	NNGRRT
saCas9-KKH	NNNRRT
spCas9-MQKSER	NGCG
spCas9-MQKSER	NGCN
spCas9-LRKIQK	NGTN
spCas9-LRVSQK	NGTN
spCas9-LRVSQL	NGTN
SpyMacCas9	NAA
Cpf1	5' (TTTV)

10

20

【0388】

[ガイドRNAを伴うCas9複合体]

【0389】

本開示のいくつかの態様は、本明細書に提供される融合タンパク質のいずれか、およびガイドRNA（例えば、目的遺伝子を標的化するガイド）を含む複合体を提供する。融合タンパク質ドメインを連結するための任意の方法を用いて（例えば、 $(GGGS)_n$ 、 $(GGGGS)_n$ 、および $(G)_n$ の形態の非常に柔軟なリンカーから、 $(EAAAK)_n$ 、 $(SGGS)_n$ 、SGSETPGTSE SATPESの形態のより剛性の高いリンカーまで（例えばGuilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nat. Biotechnol. 2014; 32(6): 577-82参照。その全内容は参照によりここに組み込まれる）、および $(XP)_n$ ）、核酸塩基エディターの活性のための最適な長さを達成することができる。いくつかの実施態様において、 n は、1、2、3、4、5、6、7、8、9、10、11、12、13、14又は15である。ある態様において、リンカーは、 $(GGS)_n$ モチーフを含み、ここで、 n は、1、3、または7である。いくつかの実施形態において、本明細書で提供される融合タンパク質のCas9ドメインは、アミノ酸配列SGSETPGTSESATPESを含むリンカーを介して融合される。

30

40

【0390】

ある態様において、ガイド核酸（例えばガイドRNA）は、15～100ヌクレオチド長であり、標的配列に相補的である少なくとも10個の連続するヌクレオチドの配列を含む。いくつかの実施形態において、ガイドRNAは、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、または50ヌクレオチドの長さである。いくつかの実施形態において、ガイドRNAは、標的配列に相補的な15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37

50

、38、39、または40個の連続したヌクレオチドの配列を含む。ある態様において、標的配列はDNA配列である。ある態様において、標的配列は、細菌、酵母、真菌、昆虫、植物または動物のゲノムにおける配列である。ある態様において、標的配列は、ヒトのゲノムにおける配列である。いくつかの実施形態において、標的配列の3'末端は、標準PAM配列(NGG)にすぐ隣接している。いくつかの実施形態において、標的配列の3'末端は、非標準PAM配列(例えば、表1に列挙されている配列又は5'-NAA-3')にすぐ隣接している。ある態様において、ガイド核酸(例えばガイドRNA)は、目的遺伝子における配列に対して相補的である。

【0391】

本開示のいくつかの態様は、本明細書に提供される融合タンパク質または複合体を使用する方法を提供する。例えば、本開示のいくつかの局面は、DNA分子を、本明細書中に提供される融合タンパク質のいずれか、および少なくとも一つのガイドRNAと接触させることを含む方法を提供し、ここで、ガイドRNAは、約15~100ヌクレオチド長であり、標的配列に相補的である少なくとも10個の連続したヌクレオチドの配列を含む。いくつかの実施形態において、標的配列の3'末端は、AGC、GAG、TTT、GTG、またはCAA配列にすぐ隣接している。いくつかの実施形態において、標的配列の3'末端は、NGA、NAA、NGCG、NGN、NNGRRT、NNRRRT、NGCG、NGCN、NGTN、NGTN、NGTN、または5'(TTTV)配列にすぐ隣接している。

【0392】

それぞれの配列における特定の位置または残基の番号付けは、使用される特定のタンパク質および番号付けスキームに依存することが理解されるであろう。たとえば、成熟タンパク質の前駆体と成熟タンパク質そのものとは番号付けが異なることがあり、種ごとの配列の違いが番号付けに影響することがある。当業者は、当業者に周知の方法、例えば、配列アラインメントおよび相同的残基の決定によって、任意の相同的タンパク質およびそれぞれのコード核酸中のそれぞれの残基を同定することができる。

【0393】

本明細書に開示された融合タンパク質のいずれかを標的部位、例えば、編集される突然変異を含む部位にターゲティングするためには、融合タンパク質をガイドRNAと共に共発現させることが典型的に必要であることは当業者には明らかであろう。本明細書の他の箇所より詳細に説明されるように、ガイドRNAは、典型的には、Cas9結合を可能にするtracrRNAフレームワークと、Cas9:核酸編集酵素/ドメイン融合タンパク質に配列特異性を付与するガイド配列とを含む。あるいは、ガイドRNAおよびtracrRNAは、2つの核酸分子として別々に提供され得る。いくつかの態様において、ガイドRNAは、ガイド配列が標的配列に相補的な配列を含むという構造を含む。ガイド配列は典型的には20ヌクレオチド長である。Cas9:核酸編集酵素/ドメイン融合タンパク質を特定のゲノム標的部位にターゲティングするための適切なガイドRNAの配列は、本開示に基づいて当業者に明らかであろう。そのような適切なガイドRNA配列は、典型的には、編集される標的ヌクレオチドの50ヌクレオチド以内の上流または下流内の核酸配列に相補的なガイド配列を含む。提供された融合タンパク質のいずれかを特定の標的配列に標的化するのに適したいくつかの例示的なガイドRNA配列が本明細書に提供される。

【0394】

[Cas9ドメインとシチジンデアミナーゼまたはアデノシンデアミナーゼとを含む融合タンパク質を使用する方法]

本開示のいくつかの態様は、本明細書に提供される融合タンパク質または複合体を使用する方法を提供する。例えば、本開示のいくつかの局面は、目的タンパク質をコードするDNA分子を、本明細書中に提供される融合タンパク質のいずれか、および少なくとも一つのガイドRNAと接触させることを含む方法を提供し、ここで、ガイドRNAは、約15~100ヌクレオチド長であり、標的配列に相補的である少なくとも10個の連続したヌクレオチドの配列を含む。いくつかの実施形態において、標的配列の3'末端は、標準的PAM配列(NGG)にすぐ隣接している。いくつかの実施形態において、標的配列の3'末端は、標準的P

10

20

30

40

50

くとも75%、80%、85%、90%、95%、96%、97%、98%、99%、または99.5%の同一性であるアミノ酸配列を含む。

【0399】

[塩基エディターシステム]

本明細書で提供される塩基エディターシステムは、(a) 対象のポリヌクレオチド(例えば、二本鎖DNAまたはRNA、一本鎖DNAまたはRNA)の標的ヌクレオチド配列を、核酸塩基エディター(例えば、アデノシン塩基エディターまたはシチジン塩基エディター)およびガイドポリ核酸(例えばgRNA)を含む塩基エディターシステムと接触させる工程であって、標的ヌクレオチド配列は標的核酸塩基対を含む、工程と；(b) 標的領域の鎖分離を誘導する工程と；(c) 標的領域の一本鎖における標的核酸塩基対の第1の核酸塩基を第2の核酸塩基に変換する工程と；(d) 標的領域の鎖を2本以上は切断することなく、第1の核酸塩基に相補的な第3の核酸塩基が、第2の核酸塩基に相補的な第4の核酸塩基に置換される工程とを含む。一部の実施形態では、ステップ(b)は省略されることを理解されたい。ある態様において、標的核酸塩基対は、1以上の遺伝子における複数の核酸塩基対である。いくつかの実施形態において、本明細書に提供される塩基エディターシステムは、1以上の遺伝子における複数の核酸塩基対の多重編集を可能にする。ある態様において、複数の核酸塩基対は、同一遺伝子内に位置する。いくつかの実施形態において、複数の核酸塩基対は、1またはそれより多い遺伝子に位置し、ここで、少なくとも1つの遺伝子は、異なる遺伝子座に位置する。

10

【0400】

いくつかの実施形態において、切断された一本鎖(ニック鎖)は、ガイド核酸にハイブリダイズされる。いくつかの実施形態において、切断された一本鎖は、第一の核酸塩基を含む鎖と反対である。一部の実施形態では、塩基エディターはCas9ドメインを含む。いくつかの実施形態において、第1の塩基はアデニンであり、第2の塩基はG、C、AまたはTではない。いくつかの実施形態において、第2の塩基はイノシンである。

20

【0401】

本明細書に提供される塩基編集システムは、触媒的に欠陥のあるStreptococcus pyogenes Cas9と、シチジンデアミナーゼと、塩基除去修復の阻害因子とを含む融合タンパク質を用いて、二本鎖DNA切断を生じることなく、ドナーDNA鋳型を必要とせず、かつ過剰な確率的挿入および欠失を誘導することなく、DNAにおけるプログラム可能な一塩基(C TまたはA G)変化を誘導する、ゲノム編集の新規なアプローチを提供する。

30

【0402】

本明細書では、塩基エディターシステムを用いて核酸塩基を編集するためのシステム、組成物、および方法を提供する。いくつかの実施形態において、塩基エディターシステムは、(1) 核酸塩基を編集するための、ポリヌクレオチドプログラム可能なヌクレオチド結合ドメインおよび核酸塩基編集ドメイン(例えばデアミナーゼドメイン)を含む塩基エディター(BE)と、(2) ポリヌクレオチドプログラム可能なヌクレオチド結合ドメインに付随するガイドポリヌクレオチド(例えばガイドRNA)とを含む。ある態様において、塩基エディターシステムは、シトシン塩基エディター(CBE)を含む。いくつかの実施形態では、塩基エディターシステムは、アデノシン塩基エディター(ABE)を含む。ある態様において、ポリヌクレオチドプログラム可能なヌクレオチド結合ドメインは、ポリヌクレオチドプログラム可能なDNA結合ドメインである。ある態様において、ポリヌクレオチドプログラム可能なヌクレオチド結合ドメインは、ポリヌクレオチドプログラム可能なRNA結合ドメインである。ある態様において、核酸塩基編集ドメインは、デアミナーゼドメインである。場合によっては、デアミナーゼドメインは、シトシンデアミナーゼまたはシチジンデアミナーゼであり得る。いくつかの実施形態では「シトシンデアミナーゼ」および「シチジンデアミナーゼ」という用語は交換可能に使用され得る。場合によっては、デアミナーゼドメインは、アデニンデアミナーゼまたはアデノシンデアミナーゼであり得る。いくつかの実施形態では「アデニンデアミナーゼ」および「アデノシンデアミナーゼ」という用語は交換可能に使用され得る。核酸塩基編集タンパク質の詳細は、国際PCT出願番号PCT/2017

40

50

/045381 (WO2018/027078) およびPCT/US2016/058344 (WO2017/070632) に記載されており、これらの各々は、その全体が参照により本明細書に組み込まれる。Komor, A.C., et al., “Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage” *Nature* 533, 420-424 (2016); Gaudelli, N. M., et al., “Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage” *Nature* 551, 464-471 (2017); および Komor, A.C., et al., “Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity” *Science Advances* 3:eaa04774 (2017)も参照のこと（その内容全体が参照により本明細書に組み入れられる）。

10

【0403】

ある態様において、塩基エディターは、編集された鎖の塩基除去修復を阻害する。いくつかの実施形態において、塩基エディターは、編集されていない鎖を保護または結合する。一部の実施形態では、塩基エディターはUGI活性を含む。ある態様において、塩基エディターは、触媒的に不活性なイノシン特異的ヌクレアーゼを含む。一部の実施形態では、塩基エディターはニッカーゼ活性を含む。ある態様において、塩基対の意図された編集は、PAM部位の上流にある。ある態様において、塩基対の意図された編集は、PAM部位の1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、または20ヌクレオチド上流にある。いくつかの実施形態において、塩基対の意図された編集は、PAM部位の下流にある。ある態様において、意図される編集された塩基対は、PAM部位の1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、または20ヌクレオチド下流にある。

20

【0404】

いくつかの実施形態では、本方法は、標準(例えばNGG) PAMサイトを必要としない。いくつかの実施形態において、核酸塩基エディターは、リンカー又はスペーサーを含む。ある態様において、リンカーまたはスペーサーは、長さが1~25アミノ酸である。ある態様において、リンカーまたはスペーサーは、長さが5~20アミノ酸である。ある態様において、リンカーまたはスペーサーは、長さが10、11、12、13、14、15、16、17、18、19、または20アミノ酸である。

【0405】

一部の実施形態では、標的領域は標的ウィンドウを含み、標的ウィンドウは標的核酸塩基対を含む。ある態様において、標的ウィンドウは、1~10ヌクレオチドを含む。ある態様において、標的ウィンドウは、長さが1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、または20ヌクレオチドである。いくつかの実施形態において、塩基対の意図された編集は、標的ウィンドウ内にある。いくつかの実施形態において、標的ウィンドウは、塩基対の意図された編集を含む。いくつかの実施形態において、本方法は、本明細書に提供される塩基エディターのいずれかを用いて実施される。いくつかの実施形態において、標的ウィンドウは、脱アミノ化ウィンドウである。

30

【0406】

いくつかの実施形態では、塩基エディターはシチジン塩基エディター (CBE) である。いくつかの実施形態において、非限定的な例示的CBEは、BE1 (APOBEC1-XTEN-dCas9)、BE2 (APOBEC1-XTEN-dCas9-UGI)、BE3 (APOBEC1-XTEN-dCas9(A840H)-UGI)、BE3-Gam、saBE3、saBE4-Gam、BE4、BE4-Gam、saBE4、またはsaB4E-Gamである。BE4はAPOBEC1 Cas9n(D10A)リンカーを32アミノ酸まで、Cas9n UGIリンカーを9アミノ酸まで拡張し、単一塩基エディター構築物のC末端に別の9アミノ酸リンカーと共にUGIの第二コピーを付加する。塩基性エディタsaBE3およびsaBE4では、*S. pyogenes* Cas9n(D10A)をより小さい*S. aureus* Cas9n(D10A)で置き換えた。BE3-Gam、saBE3-Gam、BE4-Gam、およびsaBE4-Gamは、174残基のGamタンパク質が、16アミノ酸のXTENリンカーを介してBE3、saBE3、BE4、およびsaBE4のN末端に融合されたものを有する。

40

50

【0407】

一部の実施形態では、塩基エディターはアデノシン塩基エディター (ABE) である。ある態様において、アデノシン塩基エディターは、DNA中のアデニンを脱アミノ化することができる。ある態様において、ABEは、BE3のAPOBEC1成分を、天然または操作された大腸菌TadA、ヒトADAR2、マウスADA、またはヒトADAT2で置換することによって生成される。いくつかの実施形態において、ABEは、進化されたTadAバリエーションを含む。いくつかの実施形態では、ABEはABE 1.2 (TadA*-XTEN-nCas9-NLS) である。ある態様において、TadA*はA106VおよびD108N突然変異を含む。

【0408】

ある態様において、ABEは第二世代ABEである。ある態様において、ABEはABE 2.1であり、これは、TadA* (TadA*2.1) において追加の突然変異D147YおよびE155Vを含むものである。いくつかの実施形態において、ABEはABE 2.2であり、これはABE 2.1が触媒的に不活性化されたバージョンのヒトアルキルアデニンDNAグリコシラーゼ(E125Q変異を伴うAAG)に融合されたものである。いくつかの実施形態において、ABEはABE 2.3であり、ここではABE 2.1が触媒的に不活性化された大腸菌Endo V (D35A変異で不活性化される)に融合されたものである。いくつかの態様において、ABEはABE 2.6であり、これはABE 2.1におけるリンカーの二倍の長さ(32アミノ酸、(SGGS)₂-XTEN-(SGGS)₂)のリンカーを有する。いくつかの実施形態において、ABEはABE 2.7であり、これはABE 2.1が追加の野生型TadAモノマーと連結されたものである。いくつかの実施形態では、ABEはABE 2.8であり、これはABE 2.1が追加のTadA*2.1モノマーでつながれたものである。いくつかの実施形態において、ABEはABE 2.9であり、これは進化されたTadA (TadA*2.1) とABE 2.1のN末端との直接融合体である。いくつかの実施形態において、ABEはABE 2.10であり、これは野生型TadAとABE 2.1のN末端との直接融合体である。いくつかの実施形態において、ABEはABE 2.11であり、これは、TadA*モノマーのN末端において不活性化E59A突然変異を有するABE 2.9である。いくつかの実施形態において、ABEはABE 2.12であり、これは、内部TadA*モノマーにおいて不活性化E59A突然変異を有するABE 2.9である。

【0409】

ある態様において、ABEは、第三世代ABEである。ある態様において、ABEはABE 3.1であり、これは、三つのさらなるTadA突然変異(L84F、H123Y、およびI157F)を伴うABE 2.3である。

【0410】

いくつかの実施形態において、ABEは第四世代ABEである。いくつかの実施形態において、ABEはABE4.3であり、これはさらなるTadA突然変異A142N (TadA*4.3) を伴うABE 3.1である。

【0411】

いくつかの実施形態において、ABEは第五世代ABEである。いくつかの実施形態において、ABEはABE 5.1であり、これは生存クローンからの突然変異のコンセンサセット(H36L、R51L、S146C、およびK157N) をABE 3.1に輸入することによって生成される。いくつかの実施形態において、ABEはABE 5.3であり、これは内部の進化されたTadA*に融合された野生型大腸菌TadAを含むヘテロ二量体構築物を有する。ある態様において、ABEは、以下の表2に示されるように、ABE 5.2、ABE 5.4、ABE 5.5、ABE 5.6、ABE 5.7、ABE 5.8、ABE 5.9、ABE 5.10、ABE 5.11、ABE 5.12、ABE 5.13、またはABE 5.14である。一部の実施形態では、ABEは第六世代ABEである。一部の実施形態では、ABEは、以下の表2に示されるように、ABE 6.1、ABE 6.2、ABE 6.3、ABE 6.4、ABE 6.5、またはABE 6.6である。いくつかの実施形態において、ABEは、第七世代ABEである。いくつかの実施形態において、ABEは、以下の表2に示されるように、ABE7.1、ABE7.2、ABE7.3、ABE7.4、ABE7.5、ABE7.6、ABE7.7、ABE7.8、ABE7.9、またはABE7.10である。

表2：ABEの遺伝子型

10

20

30

40

50

【表 2】

	23	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	147	152	155	156	157	16
ABE0.1	W	R	H	N	P		R	N	L	S	A	D	H	G	A	S	D	R	E	I	K	K
ABE0.2	W	R	H	N	P		R	N	L	S	A	D	H	G	A	S	D	R	E	I	K	K
ABE1.1	W	R	H	N	P		R	N	L	S	A	N	H	G	A	S	D	R	E	I	K	K
ABE1.2	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	D	R	E	I	K	K
ABE2.1	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.2	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.3	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.4	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.5	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.6	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.7	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.8	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.9	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.10	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.11	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.12	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE3.1	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.2	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.3	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.4	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.5	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.6	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.7	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.8	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE4.1	W	R	H	N	P		R	N	L	S	V	N	H	G	N	S	Y	R	V	I	K	K

10

20

30

40

50

	23	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	147	152	155	156	157	16
ABE4.2	W	G	H	N	P		R	N	L	S	V	N	H	G	N	S	Y	R	V	I	K	K
ABE4.3	W	R	H	N	P		R	N	F	S	V	N	Y	G	N	S	Y	R	V	F	K	K
ABE5.1	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.2	W	R	H	S	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	T
ABE5.3	W	R	L	N	P		L	N	I	S	V	N	Y	G	A	C	Y	R	V	I	N	K
ABE5.4	W	R	H	S	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	T
ABE5.5	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.6	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.7	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.8	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.9	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.10	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.11	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.12	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.13	W	R	H	N	P		L	D	F	S	V	N	Y	A	A	S	Y	R	V	F	K	K
ABE5.14	W	R	H	N	S		L	N	F	C	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE6.1	W	R	H	N	S		L	N	F	S	V	N	Y	G	N	S	Y	R	V	F	K	K
ABE6.2	W	R	H	N	T	V	L	N	F	S	V	N	Y	G	N	S	Y	R	V	F	N	K
ABE6.3	W	R	L	N	S		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE6.4	W	R	L	N	S		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE6.5	W	R	L	N	I	V	L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE6.6	W	R	L	N	T	V	L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE7.1	W	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE7.2	W	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE7.3	I	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE7.4	R	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE7.5	W	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	H	V	F	N	K
ABE7.6	W	R	L	N	A		L	N	I	S	V	N	Y	G	A	C	Y	P	V	I	N	K

10

20

30

40

50

	23	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	147	152	155	156	157	16
ABE7.7	L	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K
ABE7.8	I	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE7.9	L	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	P	V	F	N	K
ABE7.10	R	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K

10

【0412】

ある態様において、塩基エディターは、核酸塩基編集ドメイン(例えばデアミナーゼドメインの全部または一部)に融合されたポリヌクレオチドプログラム可能なヌクレオチド結合ドメイン(例えばCas9由来ドメイン)を含む融合タンパク質である。ある態様において、塩基エディターは、ウラシルグリコシラーゼ阻害因子(UGI)の全部または一部を含むドメインをさらに含む。ある態様において、塩基エディターは、ウラシルDNAグリコシラーゼ(UDG)などの、ウラシル結合タンパク質(UBP)の全部または一部を含むドメインを含む。ある態様において、塩基エディターは、核酸ポリメラーゼの全てまたは一部を含むドメインを含む。ある態様において、塩基エディターに組み込まれる核酸ポリメラーゼまたはその一部は、損傷乗り越えDNAポリメラーゼである。

20

【0413】

いくつかの実施形態では、塩基エディターのドメインは、複数のドメインを含むことができる。例えば、Cas9に由来するポリヌクレオチドプログラム可能なヌクレオチド結合ドメインを含む塩基エディターは、野生型または天然のCas9のRECローブおよびNUCローブに対応するRECローブおよびNUCローブを含むことができる。別の例では、塩基エディターは、RuvCIドメイン、BHドメイン、REC1ドメイン、REC2ドメイン、RuvCIIドメイン、L1ドメイン、HNHドメイン、L2ドメイン、RuvCIIIドメイン、WEDドメイン、TOPOドメインまたはCTDドメインのうちの一つ以上を含むことができる。ある態様において、塩基エディターの一つ以上のドメインは、そのドメインを含むポリペプチドの野生型バージョンに対する突然変異(例えば置換、挿入、削除)を含む。例えば、ポリヌクレオチドプログラム可能DNA結合ドメインのHNHドメインは、H840A置換を含むことができる。別の例では、ポリヌクレオチドプログラム可能なDNA結合ドメインのRuvCIドメインは、D10A置換を含むことができる。

30

【0414】

本明細書に開示される塩基エディターの異なるドメイン(例えば隣接ドメイン)は、一つ以上のリンカードメイン(例えばXTENリンカードメイン)を使用して、または使用せずに、互いに接続することができる。場合によっては、リンカードメインは、二つの分子もしくは部分(例えば、第一のドメイン(例えばCas9由来ドメイン)および第二のドメイン(例えばシチジンデアミナーゼドメインまたはアデノシンデアミナーゼドメイン)のような融合タンパク質の二つのドメイン)をつなげる結合(例えば共有結合)、化学基、または分子であり得る。ある態様において、リンカーは、共有結合(例えば炭素-炭素結合、ジスルフィド結合、炭素-ヘテロ原子結合など)である。ある態様において、リンカーは、アミド結合における炭素窒素結合である。特定の实施形態では、リンカーは、環状もしくは非環状、置換もしくは非置換、分枝もしくは非分枝の脂肪族もしくはヘテロ脂肪族リンカーである。ある態様において、リンカーは、ポリマー(例えばポリエチレン、ポリエチレングリコール、ポリアミド、ポリエステルなど)である。特定の实施態様では、リンカーは、アミノアルカン酸のモノマー、ダイマー又はポリマーを含む。ある態様において、リンカーは、アミノアルカン酸(例えばグリシン、エタン酸、アラニン、 β -アラニン、3-アミノプロパン酸、4-アミノブタン酸、5-ペンタン酸等)を含む。いくつかの実施態様において、リンカー

40

50

は、アミノヘキサ酸 (Ahx) のモノマー、ダイマー又はポリマーを含む。ある態様において、リンカーは炭素環部分(例えばシクロペンタン、シクロヘキサ)に基づく。他の態様において、リンカーはポリエチレングリコール部分 (PEG) を含む。ある態様において、リンカーは、アリアルまたはヘテロアリアル部分を含む。ある態様において、リンカーは、フェニル環に基づく。リンカーは、ペプチドからリンカーへの求核剤(例えばチオール、アミノ)の結合を促進させる官能化部分を含むことができる。リンカーの一部として任意の求電子剤を使用することができる。例示的な求電子剤としては、活性化エステル、活性化アミド、マイケル受容体、ハロゲン化アルキル、ハロゲン化アリアル、ハロゲン化アシル、およびイソチオシアナートが挙げられるが、これらに限定されない。ある態様において、リンカーは、Cas9ヌクレアーゼドメインを含むRNAプログラム可能ヌクレアーゼのgRNA

10

【0415】

典型的には、リンカーは2つの基、分子、またはその他の部分の間、またはその傍に位置して、共有結合を介して互いに連結し、それによって2つを連結する。いくつかの実施形態では、リンカーはアミノ酸または複数のアミノ酸(たとえばペプチドまたはタンパク質)である。いくつかの実施形態では、リンカーは有機分子、基、ポリマー、または化学部分である。いくつかの実施形態では、リンカーは約2~100アミノ酸の長さ、たとえば2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30~35、35~40、40~45、45~50、50

20

【0416】

本明細書に開示する塩基エディターのドメインは、任意の順序に配置することができる。たとえばポリヌクレオチドプログラム可能なヌクレオチド結合ドメインとデアミナーゼドメインを含む融合タンパク質を含む塩基エディターの非限定的な例は、以下のように配置することができる。

30

NH₂-[核酸塩基編集ドメイン]-リンカー1-[例えば Cas9 由来ドメイン]-COOH;

NH₂-[例えば シチジンデアミナーゼ]-リンカー1-[例えば Cas9 由来ドメイン]-COOH;

NH₂-[例えば シチジンデアミナーゼ]-リンカー1-[例えば Cas9 由来ドメイン]-リンカー2-[UGI]-COOH;

NH₂-[例えば APOBEC]-リンカー1-[例えば Cas9 由来ドメイン]-COOH;

NH₂-[例えば シチジンデアミナーゼ]-リンカー1-[例えば Cas9 由来ドメイン]-COOH;

NH₂-[例えば APOBEC]-リンカー1-[例えば Cas9 由来ドメイン]-COOH;

40

NH₂-[例えば APOBEC]-リンカー1-[例えば Cas9 由来ドメイン]-リンカー2-[UGI]-COOH

NH₂-[例えば アデノシンデアミナーゼ]-[例えば Cas9 由来ドメイン]-COOH;

NH₂-[例えば Cas9 由来ドメイン]-[例えば アデノシンデアミナーゼ]-COOH;

NH₂-[例えば アデノシンデアミナーゼ]-[例えば Cas9 由来ドメイン]-[イノシンBER阻害因子]-COOH;

NH₂-[例えば アデノシンデアミナーゼ]-[イノシンBER阻害因子]-[例えば Cas9 由来ドメイン]-COOH;

NH₂-[イノシンBER阻害因子]-[例えば アデノシンデアミナーゼ]-[例えば Cas9 由来ドメイン]-COOH;

50

NH₂-[例えば Cas9 由来ドメイン]-[例えば アデノシンデアミナーゼ]-[イノシンBER阻害因子]-COOH;

NH₂-[例えば Cas9 由来ドメイン]-[イノシンBER阻害因子]-[例えば アデノシンデアミナーゼ]-COOH; または

NH₂-[イノシンBER阻害因子]-[例えば Cas9 由来ドメイン]-[例えば アデノシンデアミナーゼ]-COOH

【 0 4 1 7 】

さらに、場合によっては、Gamタンパク質を塩基エディターのN末端に融合させることができる。場合によっては、Gamタンパク質を塩基エディターのC末端に融合させることができる。バクテリオファージMuのGamタンパク質は二本鎖切断 (DSB) の末端に結合し、それらを分解から保護することができる。いくつかの実施形態において、Gamを使用してDSBの遊離末端を結合することにより、塩基編集のプロセス中のインデル形成を低減することができる。いくつかの実施形態において、174残基のGamタンパク質は、塩基エディターのN末端に融合される。Komor, A.C., et al., “ Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity ” Science Advances 3:eaao4774 (2017) を参照されたい。場合によっては、1つまたは複数の突然変異が、野生型ドメインと比較して、塩基編集ドメインの長さを変化させることができる。例えば、少なくとも1つのドメインにおける少なくとも1つのアミノ酸の欠失は、塩基エディターの長さを短くすることができる。別の例では、1つまたは複数の突然変異は、野生型ドメインに対するドメインの長さを変化させない。例えば、いずれかのドメインにおける置換 (複数可) は塩基エディターの長さを変えない。全てのドメインの長さが野生型ドメインと同じであるこのような塩基エディターの非限定的な例は、以下を含むことができる:

NH₂-[APOBEC1]-リンカー1-[Cas9(D10A)]-リンカー2-[UGI]-COOH;
 NH₂-[CDA1]-リンカー1-[Cas9(D10A)]-リンカー2-[UGI]-COOH;
 NH₂-[AID]-リンカー1-[Cas9(D10A)]-リンカー2-[UGI]-COOH;
 NH₂-[APOBEC1]-リンカー1-[Cas9(D10A)]-リンカー2-[SSB]-COOH;
 NH₂-[UGI]-リンカー1-[APOBEC1]-リンカー2-[Cas9(D10A)]-COOH;
 NH₂-[APOBEC1]-リンカー1-[Cas9(D10A)]-リンカー2-[UGI]-リンカー3-[UGI]-COOH;

NH₂-[Cas9(D10A)]-リンカー1-[CDA1]-リンカー2-[UGI]-COOH;
 NH₂-[Gam]-リンカー1-[APOBEC1]-リンカー2-[Cas9(D10A)]-リンカー3-[UGI]-COOH;

NH₂-[Gam]-リンカー1-[APOBEC1]-リンカー2-[Cas9(D10A)]-リンカー3-[UGI]-リンカー4-[UGI]-COOH;

NH₂-[APOBEC1]-リンカー1-[dCas9(D10A, H840A)]-リンカー2-[UGI]-COOH; または

NH₂-[APOBEC1]-リンカー1-[dCas9(D10A, H840A)]-COOH

【 0 4 1 8 】

いくつかの実施形態において、本明細書中で提供される塩基編集融合タンパク質は、正確な位置、例えば、標的塩基が規定された領域(例えば「脱アミノ化ウィンドウ」)内に配置されるような位置に配置される必要がある。場合によっては、標的は4塩基領域内にあり得る。場合によっては、そのような規定された標的領域は、PAMの約15塩基上流にあり得る。Komor, A.C., et al., “ Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage ” Nature 533, 420-424 (2016); Gaudelli, N.M., et al., “ Programmable base editing of A · T to G · C in genomic DNA without DNA cleavage ” Nature 551, 464-471 (2017); および Komor, A.C., et al., “ Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity ” Science Advances 3:eaao4774 (2017) 参照; その全内容が参照により本明細書に組み

10

20

30

40

50

入れられる。

【0419】

定義された標的領域は、脱アミノ化ウィンドウであり得る。脱アミノ化ウィンドウは、塩基エディターが標的ヌクレオチドに作用して脱アミノ化する、規定された領域であり得る。いくつかの実施態様において、脱アミノ化ウィンドウは、2、3、4、5、6、7、8、9又は10塩基領域内にある。いくつかの実施態様において、脱アミノ化ウィンドウは、PAMの5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24又は25塩基上流にある。

【0420】

本開示の塩基エディターは、標的ポリヌクレオチド配列の編集を促進させる任意のドメイン、特徴またはアミノ酸配列を含むことができる。例えば、いくつかの実施形態において、塩基エディターは、核局在化配列 (NLS) を含む。いくつかの実施形態において、塩基エディターのNLSは、デアミナーゼドメインとポリヌクレオチドプログラム可能なヌクレオチド結合ドメインとの間に位置する。いくつかの実施形態において、塩基エディターのNLSは、ポリヌクレオチドプログラム可能なヌクレオチド結合ドメインに対してC末端に位置する。

10

【0421】

本開示の融合タンパク質は1つまたは複数の追加的特徴を含み得ることが理解されるべきである。本明細書に開示される塩基エディターに存在し得る他の例示的特徴は、細胞質局在化配列、核輸出配列などの輸出配列、または他の局在化配列などの局在化配列、ならびに融合タンパク質の可溶化、精製、または検出に有用な配列タグである。本明細書において提供される適切なタンパク質タグには、限定されるものではないが、ビオチンカルボキシラーゼキャリアータグ (BCCP) タグ、myc-タグ、カルモジュリンタグ、FLAG-タグ、ヘマグルチニン (HA) -タグ、ヒスチジンタグまたはHis-タグとも呼ばれるポリヒスチジンタグ、マルトース結合タンパク質 (MBP) -タグ、nus-タグ、グルタチオン-S-トランスフェラーゼ (GST) -タグ、緑色蛍光タンパク質 (GFP) -タグ、チオレドキシントグ、S-タグ、ソフタグ(例ソフタグ1、ソフタグ3)、ストレプトタグ、ビオチンリガーゼタグ、FLASHタグ、V5タグ、およびSBP-タグが含まれる。さらなる適切な配列は、当業者には明らかであろう。いくつかの実施形態において、融合タンパク質は、1つ以上のHisタグを含む。

20

【0422】

融合タンパク質に含まれ得るタンパク質ドメインの非限定的な例としては、デアミナーゼドメイン(例えばシチジンデアミナーゼおよび/またはアデノシンデアミナーゼ)、ウラシルグリコシラーゼ阻害因子 (UGI) ドメイン、エピトープタグ、レポーター遺伝子配列、および/または以下の活性の一つ以上を有するタンパク質ドメインが挙げられる:メチラーゼ活性、デメチラーゼ活性、転写活性化活性、転写抑制活性、転写放出因子活性、ヒストン修飾活性、RNA切断活性、および核酸結合活性。追加のドメインは、異種の機能的ドメインであり得る。そのような異種機能ドメインは、DNAメチル化、DNA損傷、DNA修復、標的DNAと結合した標的ポリペプチド(例えばヒストン、DNA結合タンパク質など)の修飾などの機能活性を付与し、例えばヒストンメチル化、ヒストンアセチル化、ヒストンユビキチン化などをもたらすことができる。

30

【0423】

付与される他の機能には、メチルトランスフェラーゼ活性、脱メチル化活性、脱アミノ化活性、ジスムターゼ活性、アルキル化活性、脱プリン化活性、酸化活性、ピリミジン二量体形成活性、インテグラーゼ活性、トランスポザーゼ活性、リコンビナーゼ活性、ポリメラーゼ活性、リガーゼ活性、ヘリカーゼ活性、フォトリアーゼ活性またはグリコシラーゼ活性、アセチルトランスフェラーゼ活性、デアセチラーゼ活性、キナーゼ活性、ホスファターゼ活性、ユビキチンリガーゼ活性、脱ユビキチン化活性、アデニル化活性、脱アデニル化活性、SUMO化活性、脱SUMO化活性、リボシル化活性、脱リボシル化活性、ミリスチル化活性、リモデリング活性、プロテアーゼ活性、酸化還元酵素活性、トランスフ

40

50

エラーゼ活性、ヒドロラーゼ活性、イソメラーゼ活性、シンターゼ活性、シンテターゼ活性、および脱ミリスチル化活性、またはこれらの任意の組合せが含まれる。

【0424】

エピトープタグの非限定的な例は、ヒスチジン (His) タグ、V5タグ、FLAGタグ、インフルエンザヘマグルチニン (HA) タグ、Mycタグ、VSV-Gタグ、およびチオレドキシン (Trx) タグを含む。レポーター遺伝子の例としては、グルタチオン-S-トランスフェラーゼ (GST)、西洋ワサビペルオキシダーゼ (HRP)、クロラムフェニコールアセチルトランスフェラーゼ (CAT)、 β -ガラクトシダーゼ、 β -グルクロニダーゼ、ルシフェラーゼ、緑色蛍光タンパク質 (GFP)、HcRed、DsRed、シアン蛍光タンパク質 (CFP)、黄色蛍光タンパク質 (YFP)、および青色蛍光タンパク質 (BFP) を含む自己蛍光タンパク質が挙げられるが、これらに限定されない。さらなるタンパク質配列には、DNA分子に結合するかまたは他の細胞分子に結合するアミノ酸配列が含まれ得、これらには、マルトース結合タンパク質 (MBP)、Sタグ、Lex A DNA結合ドメイン (DBD) 融合体、GAL4 DNA結合ドメイン融合体、および単純ヘルペスウイルス (HSV) BP16タンパク質融合体が挙げられるが、これらに限定されない。

10

【0425】

[塩基エディターの効率性]

CRISPR-Cas9ヌクレアーゼは、標的ゲノム編集を媒介するために広く使用されている。ほとんどのゲノム編集応用において、Cas9はガイドポリヌクレオチド(例えば単一ガイドRNA (sgRNA))と複合体を形成し、sgRNA配列により指定される標的部位で二本鎖DNA切断 (DSB) を誘導する。細胞は主に非相同末端結合 (NHEJ) 修復経路を介してこのDSBに应答し、遺伝子を破壊するフレームシフト変異を生じる確率的挿入または欠失 (インデル) を生じる。DSBに隣接する配列と高度の相同性を有するドナーDNA鋳型の存在下で、相同性指向性修復 (HDR) として知られる代替経路を介して遺伝子補正が達成され得る。残念ながら、ほとんどの非侵襲的条件下では、HDRは非効率であり、細胞状態および細胞型に依存し、より高いインデルの頻度によって支配される。ヒトの疾患に関連する既知の遺伝的変異の大部分は点突然変異であるため、より効率的かつクリーンに正確な点突然変異を作製できる方法が必要である。本明細書に提供される塩基編集システムは、二本鎖DNA切断を生じることなく、ドナーDNA鋳型を必要とせず、かつ過剰な確率的挿入および欠失を誘導することなく、ゲノム編集を編集するための新しい方法を提供する。

20

30

【0426】

本明細書に提供される塩基エディターは、著しい割合のインデルを生成することなく、特定のヌクレオチド塩基を改変することができる。「インデル (indel)」という用語は、本明細書において使用される場合、核酸内のヌクレオチド塩基の挿入または欠失を指す。このような挿入または欠失は、遺伝子のコード領域内でフレームシフト突然変異を引き起こす可能性がある。いくつかの実施形態において、標的ヌクレオチド配列において多数の挿入または欠失(すなわちインデル)を生じることなく、核酸内の特定のヌクレオチドを効率的に改変(例えば変異または脱アミノ化)する塩基エディターを生成することが望ましい。特定の実施形態では、本明細書に提供される塩基エディターのいずれも、インデルに対して意図された改変(例えば点突然変異または脱アミノ化)のより大きな割合を生成することができる。

40

【0427】

いくつかの実施形態において、本明細書で提供される塩基エディターシステムのいずれも、標的ポリヌクレオチド配列において、50%未満、40%未満、30%未満、20%未満、19%未満、18%未満、17%未満、16%未満、15%未満、14%未満、13%未満、12%未満、11%未満、10%未満、9%未満、8%未満、7%未満、6%未満、5%未満、4%未満、3%未満、2%未満、1%未満、0.9%未満、0.8%未満、未満、0.7%未満、0.6%未満、0.5%未満、0.4%未満、0.3%未満、0.2%未満、0.1%未満、0.09%未満、0.08%未満、0.07%未満、未満、0.06%未満、0.05%未満、0.04%未満、0.03%未満、0.02%未満、または0.01%未満のインデル形成をもたらす。

50

【0428】

本開示のいくつかの態様は、本明細書に提供される塩基エディターが、有意な数の非意図的な突然変異（例えば非意図的な点突然変異など）を生成することなく、核酸（例えば対象のゲノム内の核酸）において、意図する突然変異（例えば点突然変異など）を効率的に生成することができるという認識に基づく。

【0429】

いくつかの実施形態において、本明細書に提供される塩基エディターは、少なくとも0.01%の意図された突然変異（すなわち少なくとも0.01%の編集効率）を生じることができる。いくつかの実施形態において、本明細書に提供される塩基エディターは、少なくとも0.01%、1%、2%、3%、4%、5%、10%、15%、20%、25%、30%、40%、45%、50%、60%、70%、80%、90%、95%、または99%の意図された突然変異を生成することができる。

10

【0430】

いくつかの実施形態において、本明細書に提供される塩基エディターは、1:1より大きい、意図された点突然変異対インデルの比を生成することができる。いくつかの実施形態において、本明細書に提供される塩基エディターは、少なくとも1.5:1、少なくとも2:1、少なくとも2.5:1、少なくとも3:1、少なくとも3.5:1、少なくとも4:1、少なくとも4.5:1、少なくとも5:1、少なくとも5.5:1、少なくとも6:1、少なくとも6.5:1、少なくとも7:1、少なくとも7.5:1、少なくとも8:1、少なくとも8.5:1、少なくとも9:1、少なくとも10:1、少なくとも11:1、少なくとも12:1、少なくとも13:1、少なくとも14:1、少なくとも15:1、少なくとも20:1、少なくとも25:1、少なくとも30:1、少なくとも40:1、少なくとも50:1、少なくとも100:1、少なくとも200:1、少なくとも300:1、少なくとも400:1、少なくとも500:1、少なくとも600:1、少なくとも700:1、少なくとも800:1、少なくとも900:1、もしくは少なくとも1000:1、またはそれ以上の、意図される点突然変異対インデルの比を生じることができる。

20

【0431】

意図された突然変異およびインデルの数は、例えば、国際PCT出願番号PCT/2017/045381 (WO2030/027078) およびPCT/US2016/058344 (WO2017/070632); Komor, A.C., et al., "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" *Nature* 533, 420-424 (2016); Gaudelli, N. M., et al., "Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage" *Nature* 551, 464-471 (2017); および Komor, A.C., et al., "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity," *Science Advances* 3:eaao4774 (2017) (その全内容は参照によりここに組み込まれる)に記載されているもののような、任意の適切な方法を用いて決定することができる。

30

【0432】

いくつかの実施形態において、インデル頻度を計算するために、インデルが生じ得るウィンドウの両側に隣接する二つの10 bp配列に対する正確なマッチについて配列決定リードがスキャンされる。完全マッチが見つからない場合、そのリードは解析から除外される。このインデルウィンドウの長さが参照配列と完全にマッチする場合、リードはインデルを含まないものとして分類される。インデルウィンドウが参照配列よりも2塩基以上長いまたは短い場合、配列決定リードは、それぞれ挿入または欠失として分類される。いくつかの態様において、本明細書において提供される塩基エディターは、核酸の領域におけるインデルの形成を制限することができる。ある態様において、その領域は、塩基エディターによって標的化されるヌクレオチドのところにあるか、または塩基エディターによって標的化されるヌクレオチドの2、3、4、5、6、7、8、9または10ヌクレオチド以内の領域である。

40

【0433】

標的ヌクレオチド領域で形成されるインデルの数は、核酸（例えば細胞のゲノム内の核酸

50

)が塩基エディターに曝される時間の量に依存し得る。いくつかの実施形態において、インデルの数または割合は、標的ヌクレオチド配列(例えば細胞のゲノム内の核酸)を塩基エディターに曝してから少なくとも1時間、少なくとも2時間、少なくとも6時間、少なくとも12時間、少なくとも24時間、少なくとも36時間、少なくとも48時間、少なくとも3日、少なくとも4日、少なくとも5日、少なくとも7日、少なくとも10日、または少なくとも14日後に決定される。本明細書に記載される塩基エディターの特徴は、本明細書に提供される融合タンパク質、または融合タンパク質を使用する方法のいずれにも適用され得ることが理解されるべきである。

【0434】

[多重編集]

いくつかの実施形態において、本明細書に提供される塩基エディターシステムは、1以上の遺伝子における複数の核酸塩基対の多重編集を可能にする。ある態様において、複数の核酸塩基対は、同一遺伝子内に位置する。いくつかの実施形態において、複数の核酸塩基対は、1つまたはそれより多い遺伝子に位置し、ここで、少なくとも1つの遺伝子は、異なる遺伝子座に位置する。ある態様において、多重編集は、1以上のガイドポリヌクレオチドを含むことができる。いくつかの実施形態では、多重編集は、1つ以上の塩基エディターシステムを含むことができる。いくつかの実施形態において、多重編集は、単一のガイドポリヌクレオチドを有する1つ以上の塩基エディターシステムを含むことができる。いくつかの実施形態において、多重編集は、複数のガイドポリヌクレオチドを有する1つ以上の塩基エディターシステムを含むことができる。いくつかの実施形態において、多重編集は、単一の塩基エディターシステムを有する1以上のガイドポリヌクレオチドを含むことができる。いくつかの実施形態において、多重編集は、標的ポリヌクレオチド配列への結合を標的化するためにPAM配列を必要としない少なくとも1つのガイドポリヌクレオチドを含むことができる。いくつかの実施形態において、多重編集は、標的ポリヌクレオチド配列への結合を標的化するためにPAM配列を必要とする少なくとも1つのガイドポリヌクレオチドを含むことができる。いくつかの実施形態において、多重編集は、標的ポリヌクレオチド配列への結合を標的化するためにPAM配列を必要としない少なくとも1つのガイドポリヌクレオチドと、標的ポリヌクレオチド配列への結合を標的化するためにPAM配列を必要とする少なくとも1つのガイドポリヌクレオチドとの混合物を含むことができる。本明細書に記載される塩基エディターのいずれかを使用する多重編集の特徴は、本明細書に提供される塩基エディターのいずれかを使用する方法の任意の組み合わせに適用され得ることを理解されたい。また、本明細書に記載される塩基エディターのいずれかを使用する多重編集は、複数の核酸塩基対の順次的編集を含むことができることを理解されたい。

【0435】

本明細書で提供される方法は、(a) 対象のポリヌクレオチドの標的ヌクレオチド配列(例えば二本鎖DNA配列)を、核酸塩基エディター(例えばアデノシン塩基エディターまたはシチジン塩基エディター)およびガイドポリヌクレオチド酸(例えばgRNA)を含む塩基エディターシステムと接触させる工程であって、ここで、標的ヌクレオチド配列は標的核酸塩基対を含む、工程と；(b) 標的領域の鎖分離を誘導する工程と；(c) 標的領域の一本鎖における標的核酸塩基対の第一の核酸塩基を第二の核酸塩基に編集する工程と；(d) 標的領域の鎖を1本より多くは切断せず、第1の核酸塩基に相補的な第3の核酸塩基が第2の核酸塩基に相補的な第4の核酸塩基に置換される工程とを含む。

【0436】

いくつかの実施形態において、複数の核酸塩基対は、1つ以上の遺伝子に存在する。ある態様において、複数の核酸塩基対は、同じ遺伝子内にある。いくつかの実施形態において、1つ以上の遺伝子における少なくとも1つの遺伝子は、異なる遺伝子座に位置する。

【0437】

ある態様において、塩基編集は、少なくとも1つのタンパク質コード領域における複数の核酸塩基対の編集である。いくつかの実施形態において、編集は、少なくとも1つのタ

10

20

30

40

50

ンパク質非コード領域における複数の核酸塩基対の編集である。ある態様において、編集は、少なくとも1つのタンパク質コード領域および少なくとも1つのタンパク質非コード領域における複数の核酸塩基対の編集である。

【0438】

いくつかの態様において、編集は、1以上のガイドポリヌクレオチドを伴う。いくつかの実施形態では、塩基エディターシステムは、1つ以上の塩基エディターシステムを含むことができる。いくつかの実施形態では、塩基エディターシステムは、単一のガイドポリヌクレオチドとともに1つ以上の塩基エディターシステムを含むことができる。いくつかの実施形態において、塩基エディターシステムは、複数のガイドポリヌクレオチドとともに1つ以上の塩基エディターシステムを含むことができる。いくつかの実施形態において、編集は、単一の塩基エディターシステムを有する1以上のガイドポリヌクレオチドを伴う。いくつかの実施形態において、編集は、標的ポリヌクレオチド配列への結合を標的化するためにPAM配列を必要としない少なくとも1つのガイドポリヌクレオチドを伴う。いくつかの実施形態において、編集は、標的ポリヌクレオチド配列への結合を標的化するためにPAM配列を必要とする少なくとも1つのガイドポリヌクレオチドを伴う。いくつかの実施形態において、編集は、標的ポリヌクレオチド配列への結合を標的化するためにPAM配列を必要としない少なくとも1つのガイドポリヌクレオチドと、標的ポリヌクレオチド配列への結合を標的化するためにPAM配列を必要とする少なくとも1つのガイドポリヌクレオチドとの混合物を伴う。本明細書に記載される塩基エディターのいずれかを使用する多重編集の特徴は、本明細書に提供される塩基エディターのいずれかを使用する方法の任意の組み合わせに適用され得ることを理解されたい。また、編集は、複数の核酸塩基対の順次的編集を含むことができることを理解されたい。

【0439】

[塩基エディターを使用する方法]

疾患関連遺伝子およびアレルにおける点突然変異の補正は、治療および基礎研究への応用を有する遺伝子補正のための新しい戦略を切り開く。ここで開示される部位特異的単一塩基改変システムは、ある遺伝子の機能を意図的に抑制または除去する「リバーズ」遺伝子療法においても応用を有し得る。これらのケースでは、タンパク質において不活性化変異をもたらす残基を部位特異的に変異させること、またはタンパク質の機能を阻害する変異を用いて、*in vitro*、*ex vivo*、または*in vivo*でタンパク質機能を除去または阻害することができる。

【0440】

本開示は、本明細書に提供される塩基エディターシステムによって補正することができる点突然変異に関連するまたはその点突然変異によって引き起こされる疾患と診断された対象の治療方法を提供する。例えば、いくつかの態様において、疾患、例えば遺伝子突然変異によって引き起こされる疾患を有する対象に、疾患関連遺伝子に不活性化変異を導入する核酸塩基エディター(例えばアデノシンデアミナーゼ塩基エディターまたはシチジンデアミナーゼ塩基エディター)の有効量を投与する工程を含む方法が提供される。

【0441】

いくつかの実施形態では、疾患は増殖性疾患である。いくつかの実施形態では、疾患は遺伝的疾患である。いくつかの実施形態では、疾患は腫瘍性疾患である。いくつかの実施形態では、疾患は代謝疾患である。いくつかの実施形態では、疾患はリソソーム蓄積症である。好適な疾患および障害の例としては、限定されないが、鎌状赤血球症、ベータサラセミア、またはアルファ1アンチトリプシン欠損症(A1AD)が挙げられる。点突然変異を補正するかまたは疾患関連遺伝子に不活性化突然変異を導入することによって治療することができる他の疾患は、当業者にとって既知であり得、本開示はこの点において限定されない。本開示は、追加の疾患または障害の治療方法、例えばデアミナーゼ媒介遺伝子編集によって補正することができる点突然変異に関連する、またはそれによって引き起こされる疾患または障害の治療方法を提供する。いくつかのそのような疾患は本明細書に記載されており、そして本明細書に提供される戦略および融合タンパク質で治療することがで

きるさらなる適切な疾患は、本開示に基づいて当業者に明らかであろう。それぞれの配列における特定の位置または残基の番号付けは、使用される特定のタンパク質および番号付けスキームに依存することが理解されることができ、例えば成熟タンパク質の前駆体と成熟タンパク質そのものの番号付けは異なることがあり、種間の配列の違いが番号付けに影響することがある。当業者は、任意の相同的タンパク質およびそれぞれのコード核酸中のそれぞれの残基を、当技術分野で周知の方法、例えば配列アラインメントおよび相同的残基の決定によって同定することができる。

【0442】

本明細書では、疾患または障害に関連する標的ヌクレオチド配列中の核酸塩基を編集するために、塩基エディターまたは塩基エディターシステムを使用する方法が提供される。いくつかの実施形態において、塩基エディター(例えばアデノシンデアミナーゼおよびCas9ドメインを含むもの)の活性は、点突然変異の補正をもたらす。ある態様において、標的DNA配列は、疾患または障害と関連するG A点突然変異を含み、ここで、突然変異体A塩基の脱アミノ化は、疾患または障害と関連しない配列をもたらす。ある態様において、標的DNA配列は、疾患または障害と関連するT C点突然変異を含み、ここで、突然変異体C塩基の脱アミノ化は、疾患または障害と関連しない配列をもたらす。

【0443】

いくつかの実施形態において、標的DNA配列はタンパク質をコードし、点突然変異はコドン中にあり、野生型コドンと比較して突然変異コドンによってコードされるアミノ酸の変化をもたらす。いくつかの実施形態において、突然変異体Aの脱アミノ化は、突然変異体コドンによってコードされるアミノ酸の変化をもたらす。いくつかの実施形態において、突然変異体Aの脱アミノ化は、野生型アミノ酸をコードするコドンをもたらす。いくつかの実施形態において、突然変異体Cの脱アミノ化は、突然変異体コドンによってコードされるアミノ酸の変化をもたらす。いくつかの実施形態において、突然変異体Cの脱アミノ化は、野生型アミノ酸をコードするコドンをもたらす。ある態様において、対象は、疾患もしくは障害を有するか、または疾患もしくは障害と診断されている。

【0444】

ある態様において、本明細書に提供されるアデノシンデアミナーゼは、DNAのデオキシアデノシン残基のアデニンを脱アミノ化することができる。開示の他の態様は、アデノシンデアミナーゼ(例えば、本明細書に記載のDNA中のデオキシアデノシンを脱アミノ化するアデノシンデアミナーゼ)および特定のヌクレオチド配列に結合することができるドメイン(例えばCas9やCpf1タンパク質)を含む融合タンパク質を提供する。例えば、アデノシンは、典型的にはシトシン残基と塩基対を形成するイノシン残基に変換されることができ、このような融合タンパク質は、特に核酸配列の標的化編集に有用である。そのような融合タンパク質は、インビトロでのDNAの標的化された編集のために、例えば、突然変異体の細胞または動物の生成のために；標的化された突然変異の導入のために、例えば、*ex vivo*の細胞における遺伝的欠陥の補正のために、例えば、対象から得られその後同じまたは別の対象に再導入される細胞において；および*in vivo*での標的突然変異の導入のために使用することができ、例えば、GからAへ、またはTからCへの突然変異による疾患関連遺伝子における遺伝子欠陥の補正または不活性化突然変異の導入は、本明細書に提供される核酸塩基エディターを用いて処理することができる。本開示は、デアミナーゼおよび核酸塩基エディターを利用するデアミナーゼ、融合タンパク質、核酸、ベクター、細胞、組成物、方法、キット、システム等を提供する。

【0445】

[意図する変異の生成]

いくつかの態様において、本明細書において提供される方法の目的は、遺伝子編集を介して機能不全遺伝子の機能を回復させることである。いくつかの実施形態において、機能不全遺伝子の機能は、意図された突然変異を導入することによって回復される。本明細書において提供される核酸塩基編集タンパク質は、例えばヒト細胞培養物中の疾患関連突然変異を補正することによって、インビトロでの遺伝子編集に基づくヒト治療のために検証

10

20

30

40

50

され得る。本明細書において提供される核酸塩基編集タンパク質、例えば、ポリヌクレオチドプログラム可能なヌクレオチド結合ドメイン(例:Cas9)および核酸塩基編集ドメイン(例えば、アデノシンデアミナーゼドメインまたはシチジンデアミナーゼドメイン)を含む融合タンパク質は、任意のAからGまたはCからTへの単一点突然変異を補正するために使用することができる。これが当業者によって理解されよう。最初のケースでは、変異体AからIへの脱アミノ化が変異を修正し、後者のケースでは、変異体Tと塩基対を形成しているAの脱アミノ化とそれに続く一回の複製が変異を修正する。

【0446】

いくつかの態様において、本開示は、有意な数の非意図的な突然変異(例えば非意図的な点突然変異)を生成することなく、核酸(例えば対象のゲノム内の核酸)において、意図する突然変異(例えば点突然変異)を効率的に生成することができる塩基エディターを提供する。いくつかの実施形態において、意図された突然変異は、意図された突然変異を生じさせるように特に設計されたガイドポリヌクレオチド(例えばgRNA)に結合した特定の塩基エディター(例えばシチジン塩基エディターまたはアデノシン塩基エディター)によって生じる突然変異である。ある態様において、意図される突然変異は、疾患または障害に関連する突然変異である。ある態様において、意図される突然変異は、疾患または障害に関連するアデニン(A)からグアニン(G)への点突然変異である。ある態様において、意図される突然変異は、疾患または障害に関連するシトシン(C)からチミン(T)への点突然変異である。いくつかの実施形態において、意図される突然変異は、遺伝子のコード領域または非コード領域内のアデニン(A)からグアニン(G)への点突然変異である。ある態様において、意図される突然変異は、遺伝子のコード領域または非コード領域内のシトシン(C)からチミン(T)への点突然変異である。

【0447】

いくつかの実施形態において、本明細書に提供される塩基エディターのいずれかは、意図される突然変異対意図されない突然変異の比(例えば、意図される点突然変異:意図されない点突然変異)を1:1よりも大きくすることができる。いくつかの実施形態において、本明細書に提供される塩基エディターのいずれかは、意図される突然変異対意図されない突然変異の比(例えば意図される点突然変異:意図されない点突然変異)を、少なくとも1.5:1、少なくとも2:1、少なくとも2.5:1、少なくとも3:1、少なくとも3.5:1、少なくとも4:1、少なくとも4.5:1、少なくとも5:1、少なくとも5.5:1、少なくとも6:1、少なくとも6.5:1、少なくとも7:1、少なくとも7.5:1、少なくとも8:1、少なくとも10:1、少なくとも12:1、少なくとも15:1、少なくとも20:1、少なくとも25:1、少なくとも30:1、少なくとも40:1、少なくとも50:1、少なくとも100:1、少なくとも150:1、少なくとも200:1、少なくとも250:1、少なくとも500:1、または少なくとも1000:1以上とすることができる。

【0448】

塩基エディター効率の詳細は、国際PCT出願番号PCT/2017/045381(WO2030/027078)およびPCT/US2016/058344(WO2017/070632)に記載されており、これらの各々は、その全体が参照により本明細書に組み込まれる。Komor, A.C., et al., "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" *Nature* 533, 420-424 (2016); Gaudelli, N.M., et al., "Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage" *Nature* 551, 464-471 (2017); および Komor, A.C., et al., "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" *Science Advances* 3:eaa04774 (2017)も参照(その全体が参照により本明細書に組み込まれる)。

【0449】

いくつかの実施形態では、1つ以上の遺伝子における複数の核酸塩基の編集が少なくとも1つの意図された変異の形成をもたらす。いくつかの実施形態では、少なくとも1つの意図された変異の形成が、疾患表現型を抑制する補完的変異の導入をもたらす。本明細書

10

20

30

40

50

に記載される塩基エディターの多重編集の特徴は、本明細書に提供される塩基エディターを使用する方法の任意の組み合わせに適用され得ることが理解されるべきである。

【0450】

[補完的変異の導入]

いくつかの実施形態では、本明細書に提供される塩基エディターは、1つ以上の補完的 (compensatory) 変異を導入して、遺伝子のオープンリーディングフレームの変異を補正することができ、そしてそのことは、(1)活性部位変異を補正することにより、もしくは触媒活性もしくは基質親和性を増加させるアロステリック変異を導入することにより、タンパク質の活性を増加させるか、(2)タンパク質の安定性を増加させるか、または(3)翻訳率を向上させること、エンドソーム放出を増加させること、シグナルペプチドプロセッシングを改善すること、もしくは他のタンパク質 (例えばリプレッサーまたはシャペロン) との相互作用を増加/減少させることによりタンパク質の発現を増加させる。いくつかの実施形態では、補完的変異は疾患原因変異を無効にし得る。補完的変異導入の非限定的な例は表3Aおよび3Bに記載されている。突然変異および他の配列バリエーションを記述する命名法の詳細は、den Dunnen, J.T. and Antonarakis, S.E., "Mutation Nomenclature Extensions and Suggestions to Describe Complex Mutations: A Discussion." *Human Mutation* 15:712 (2000)に記載されており、その内容全体が参照により本明細書に組み入れられる。

10

【0451】

一態様において、疾患または障害はアルファ1アンチトリプシン欠損症 (A1AD) である。いくつかの実施形態では、病原性変異は、A1ATタンパク質をコードするSERPINA1遺伝子中にある。A1ATタンパク質における変異はA1ADに関連する。(表3A)。いくつかの実施形態では、SERPINA1の病原性変異はE342K (PiZアリル)である。いくつかの実施形態では、SERPINA1の病原性変異はE264V (PiSアリル)である。いくつかの実施形態では、A1ATのPiZ または PiSアリルの変異体作用を抑制する補完的変異はM374Iである (図3および図4)。いくつかの実施形態では、A1ATのPiZ または PiSアリルの変異体作用を抑制する補完的変異はF51Lである。いくつかの実施形態では、A1ATのPiZ または PiSアリルの変異体作用を抑制する補完的変異はA348V/A347Vである。いくつかの実施形態では、A1ATのPiZ または PiSアリルの変異体作用を抑制する補完的変異はK387Rである。いくつかの実施形態では、A1ATのPiZ または PiSアリルの変異体作用を抑制する補完的変異はT59Aである。いくつかの実施形態では、A1ATのPiZ または PiSアリルの変異体作用を抑制する補完的変異はT68Aである。

20

30

【0452】

別の態様において、疾患または障害は表3Bに記載されるものを表す。一実施形態では、疾患または障害は鎌状赤血球症である。ある実施形態では、1つ以上の補完的変異は、ヘモグロビンのサブユニットをコードする遺伝子に導入され得る。ある実施形態では、1つ以上の補完的変異は、ヘモグロビンのベータ () サブユニット (HbB) をコードするHBB遺伝子に導入され得る。ある実施形態では、HBB遺伝子は、鎌状赤血球ヘモグロビンアリル (HbS) である。ある実施形態では、HBB遺伝子に1つ以上の補完的変異を導入することは、ヘモグロビンのベータサブユニットのアミノ酸配列における変化をもたらす。ある実施形態では、ベータヘモグロビンサブユニットにおける変化は、A70T、A70V、L88P、F85L、F85P、E22G、G16D、G16N、またはそのいずれかの組合せである。ある実施形態では、HBA1またはHBA2遺伝子に1つ以上の補完的変異を導入することが、ヘモグロビンのアルファサブユニットのアミノ酸配列における変化をもたらす。ある実施形態では、塩基編集は、ヘモグロビンのアルファサブユニットのアミノ酸配列における変化をもたらし得る。ある実施形態では、アルファヘモグロビンサブユニットの該アミノ酸配列は、ヘモグロビンのアルファサブユニットとベータサブユニットの重合境界面に位置する。ある実施形態では、アルファサブユニットの該アミノ酸配列は、鎌状赤血球ヘモグロビンのアルファサブユニットとベータサブユニットの重合境界面に位置する。ある実施形態では、アルファサブユニットのアミノ酸配列における変化は、K11E、D47G、Q54R、N6

40

50

8D、E116K、H20Y、H50Y、またはそのいずれかの組合せである。ある実施形態では、これらの変化のいずれかが、HbA/HbSテトラマーを形成する重合能力を減少させ得る。ある実施形態では、これらの変化のいずれかは、ヘモグロビンの1つ以上のアロステリック部位にある。ある実施形態では、これらの変化のいずれかは、ヘモグロビンの1つ以上の非アロステリック部位にある。ある実施形態では、鎌状赤血球ヘモグロビンのアミノ酸配列におけるこれらの変化のいずれかは、HBA1 または HBA2 遺伝子に位置する追加的な核酸塩基の追加的な編集と共にマルチプレックス化され得る。ある実施形態では、疾患は嚢胞性線維症 (CF) であり、補完的変異 (例えばR555K, F409L, F433L, H667R, R1070W, R29K, R553Q, I539T, G550E, F429S, Q637R) は、脊椎動物においてCTRF膜タンパク質および塩素イオンチャネルをコードする嚢胞性線維症膜コンダクタンス制御因子 (CTRF) 遺伝子における変化を含む。ある実施形態では、疾患は、ミスフォールドまたはミスアセンブルされた (バリエーション) トランスサイレチンタンパク質により誘導されるトランスサイレチン (TTR) 心アミロイドーシスであり、補完的変異 (例えばA108V, R104H, T119M) は、ミスフォールドまたはミスアセンブルされたバリエーションについて補完する、TTRタンパク質における変化を含む。

【0453】

本明細書で提供される塩基編集システムは、他のヘモグロビンアリルのいずれかの病原性アミノ酸を抑制するために使用され得ることが理解されるべきである。ある実施形態では、上記変化がヘモグロビンの鎌状化を最小限化する。ある実施形態では、上記変化はヘモグロビンサブユニットの重合に関わる1つ以上のアミノ酸残基にある。ある実施形態では、上記変化がヘモグロビンの溶解性を向上させる。ヘモグロビンサブユニットの重合に関わる他のあらゆるアミノ酸残基が本明細書において企図される。

【表3A】

表 3A. <i>SERPINA1</i> 遺伝子における補完的変異の導入					
	遺伝子	補完的変異	塩基エディター	gRNA 標的化配列	PAM
1	<i>SERPINA1</i>	F51L	ABE	GAAGAAGAUUUGGUGCUGU	NGG
2	<i>SERPINA1</i>	M374I	CBE	UCAAUCAUUAAGAAGACAAA	NGG
3	<i>SERPINA1</i>	A348V/A347V	CBE		
4	<i>SERPINA1</i>	K387R	ABE	ACUUUCCCAUGAAGAGGGG	NGA
5	<i>SERPINA1</i>	T59A	ABE	CAUCGCUACAGCCUUUGCAA	NGC
6	<i>SERPINA1</i>	T68A	ABE	GGGACCAAGGCUGACACUCA	NGA

10

20

30

40

50

【表 3 B】

表 3B. 疾患原因遺伝子における補完的変異の導入					
	遺伝子	補完的変異	塩基エディター	gRNA 標的化配列	PAM
1.	<i>HBB</i>	A70T	CBE		
2.	<i>HBB</i>	A70V	CBE	CGGUGCCUUUAGUGAUGGCC	NGG
3.	<i>HBB</i>	L88P	ABE	UGCAGCUCACUCAGUGUGGC	NNNRRT
4.	<i>HBB</i>	F85L および/ または F85P	ABE	CAGUGUGGCAAAGGUGCCCU	NNNRRT
5.	<i>HBB</i>	E22G	ABE	CGUGGAUGAAGUUGGUGGUG	NGG
6.	<i>HBB</i>	G16D および/ または G16N	BE	CUUGCCCCACAGGGCAGUAA	NGG
7.	<i>CFTR</i>	R555K	CBE	CUAAAGAAAUUCUUGCUCGU	NGA
8.	<i>CFTR</i>	F409L	ABE	UUGCUUUCUCAAUAAUUC	NNNRRT
9.	<i>CFTR</i>	F433L	ABE	GUGAGAAAUACUGAAGAAG	NGG
10.	<i>CFTR</i>	H667R	ABE	UUACACCGUUUCUCAUAGA	NGG
11.	<i>CFTR</i>	R1070W	CBE	UUCGGACGGCAGCCUACU	NGA
12.	<i>CFTR</i>	R29K	CBE	CGCUGUCUGUAUCCUUCCU	NNNRRT
13.	<i>CFTR</i>	R553Q	CBE	GCUCGUUGACCUCCACUCAG	NNNRRT
14.	<i>CFTR</i>	I539T	ABE	AGAACUAUUAUUGUCUUUC	NGC
15.	<i>CFTR</i>	G550E	CBE	GCUCGUUGACCUCCACUCAG	NNNRRT
16.	<i>CFTR</i>	F429S	ABE		
17.	<i>CFTR</i>	Q637R	ABE	AAAACUACAGCCAGACUUU	NGC
18.	<i>TTR</i>	A108V	CBE	ACACCAUUGCCGCCUGCUG	NGC
19.	<i>TTR</i>	R104H	CBE	AAUGGUGUAGCGGCGGGGC	NNGRRT
20.	<i>TTR</i>	T119M	CBE		

【 0 4 5 4 】

[送達システム]

本開示による核酸塩基エディターをコードする核酸は、当業者に知られる方法によって、または本明細書に記載されるように、対象に投与され、またはインピットロで細胞に送達され得る。一実施形態では、核酸塩基エディターは、肝臓、肺、または他の任意の器官およびそれらの前駆体の細胞に選択的に送達される。特定の実施形態では、編集を経た細胞が、コードされるタンパク質の機能に対する遺伝子編集の機能的影響を抑制するために使用され得る。一実施形態では、核酸塩基エディターは、例えばベクター（例えばウイルス性ベクターまたは非ウイルス性ベクター）、ベクターに基づかない方法（例えば裸のDNA

10

20

30

40

50

、DNA複合体、脂質ナノ粒子を使用すること)、またはそれらの組合せによって送達され得る。

【0455】

核酸塩基エディターをコードする核酸は、裸のDNAまたはRNAとして、例えばトランスフェクションもしくは電気穿孔の手段によって、肝臓、肺、または他の任意の器官の細胞に直接送達され得、または標的細胞による取り込みを促進させる分子(例えばN-アセチルガラクトースアミン)に結合され得る。本明細書に記載されるベクターのような、核酸ベクターも使用され得る。

【0456】

本明細書に開示される塩基エディターを、ウイルスベクター中に含まれる核酸上にコードさせることができる。ウイルスベクターは、レンチウイルス、アデノウイルス、レトロウイルス、およびアデノ随伴ウイルス(AAV)を含んでもよい。ウイルスベクターを、用途に基づいて選択することができる。例えば、AAVは、その軽度の免疫原性に起因してin vivoでの遺伝子送達のために一般的に使用される。アデノウイルスは、それらが誘導する強い免疫原性応答のため、ワクチンとして一般的に使用される。ウイルスベクターのパッケージング能力は、ベクター中にパッケージングすることができる塩基エディターのサイズを制限し得る。例えば、AAVのパッケージング能力は、2つの134塩基の逆方向末端反復(ITR)を含む約4.5 kbである。

【0457】

AAVゲノムは、それぞれ四つの複製タンパク質および三つのキャプシドタンパク質をコードする二つの遺伝子からなり、両側に145 bpの逆方向末端反復配列(ITR)がある。ピリオンは三つのキャプシドタンパク質Vp1、Vp2およびVp3から成り、これらは同じオープンリーディングフレームから1:1:10比で産生されるが、異なるスプライシング(Vp1)および選択的翻訳開始部位(Vp2とVp3のそれぞれ)から産生される。Vp3はピリオン中に最も豊富に存在するサブユニットであり、ウイルスの指向性を規定する細胞表面での受容体認識に関与する。ウイルス感染性において機能するホスホリパーゼドメインがVp1のコニークなN末端に同定されている。

【0458】

組換えAAV(rAAV)は、wt AAVと同様に、ベクター導入遺伝子カセットを挟むシス作用性145 bp ITRを利用し、外来DNAのパッケージングのために最大4.5 kbを提供する。感染後、rAAVは本発明の融合タンパク質を発現することができ、環状の頭・尾コンカテマーの状態のエピソームとして存在することにより、宿主ゲノムに組み込まれることなく存続することができる。このシステムを用いたrAAVの成功例は数多くあるが、in vitroおよびin vivoでは、パッケージング能力が限られているため、遺伝子のコード配列の長さがwt AAVゲノムの長さ以上である場合にAAV媒介遺伝子送達の使用が制限されている。

【0459】

AAVベクターの小さいパッケージング能力は、このサイズを超える多数の遺伝子の送達および/または大きな生理学的調節エレメントの使用を困難にする。これらの課題は、例えば、送達されるタンパク質を2つ以上の断片に分割することによって対処することができる。ここで、N末端断片を分割インテイン-Nに融合し、C末端断片を分割インテイン-Cに融合する。そしてこれらの断片を2つ以上のAAVベクターにパッケージングする。本明細書中で使用される場合、「インテイン」とは、隣接するN末端エクステインおよびC末端エクステイン(例えば、連結されるべき断片)をライゲートする自己スプライシングタンパク質イントロン(例えばペプチド)を指す。異種タンパク質断片を連結するための特定のインテインの使用は、例えば、Wood et al., J. Biol. Chem. 289(21): 14512-9 (2014)に記載されている。例えば、インテインIntNおよびIntCは、別々のタンパク質断片に融合された場合、互いを認識して、自身をスプライスして排出し、それと同時に、融合している上記タンパク質断片の隣接するN-およびC-末端エクステインをライゲートして、それによってそれによって2つのタンパク質断片から完全長タンパク質を再構成する。他の適切なインテインは当業者に明らかであろう。

10

20

30

40

50

【0460】

本発明の融合タンパク質の断片は、長さを変えることができる。ある態様において、タンパク質断片は、長さが2アミノ酸～約1000アミノ酸の範囲である。ある態様において、タンパク質断片は、長さが約5アミノ酸～約500アミノ酸の範囲である。ある態様において、タンパク質断片は、長さが約20アミノ酸～約200アミノ酸の範囲である。ある態様において、タンパク質断片は、長さが約10アミノ酸～約100アミノ酸の範囲である。他の長さの適切なタンパク質断片は、当業者には明らかであろう。

【0461】

ある態様において、ヌクレアーゼ(例えばCas9)の一部または断片は、インテインに融合される。ヌクレアーゼは、インテインのN末端またはC末端に融合され得る。ある態様において、融合タンパク質の一部または断片は、インテインに融合され、AAVキャプシドタンパク質に融合される。インテイン、ヌクレアーゼおよびキャプシドタンパク質は、任意の配置(例えば、ヌクレアーゼ-インテイン-キャプシド、インテイン-ヌクレアーゼ-キャプシド、キャプシド-インテイン-ヌクレアーゼなど)と一緒に融合され得る。いくつかの実施形態において、インテインのN末端は融合タンパク質のC末端に融合され、インテインのC末端はAAVキャプシドタンパク質のN末端に融合される。

10

【0462】

一実施形態では、大きな導入遺伝子発現カセットを二つの別々の半分(5'および3'末端、またはヘッドおよびテール)に分割することによって二重AAVベクターが生成され、カセットの各半分が単一のAAVベクター(5 kb未満)内にパッケージングされる。次いで、両方の二重AAVベクターによる同一細胞の同時感染に続いて、(1) 5'および3'ゲノム(二重AAV重複ベクター)間の相同組換え(HR); (2) ITRを介した5'及び3'ゲノム(二重AAVトランススプライシングベクター)の尾・頭コンカテマー化; または(3) それら二つのメカニズムの組み合わせ(二重AAVハイブリッドベクター)により、完全長トランスジーン発現カセットの再構築が達成される。in vivoでの二重AAVベクターの使用は完全長タンパク質の発現をもたらす。二重AAVベクタープラットフォームの使用は、4.7 kbの大きさの導入遺伝子のための効率的で実行可能な遺伝子導入戦略を表す。

20

【0463】

塩基エディターを設計するための、開示される戦略は、ウイルスベクターにパッケージングされることができる塩基エディターを生成するために有用となり得る。塩基エディターの送達のためのRNAまたはDNAウイルスベースのシステムの使用は、培養中または宿主中の特定の細胞にウイルスをターゲティングし、ウイルスの積荷を核または宿主細胞ゲノムに輸送する、高度に進化したプロセスを利用する。ウイルスベクターは、培養中の細胞、患者に直接投与することができ(in vivo)、またはそれらを用いて細胞をin vitroで処理し得、改変された細胞を任意で患者に投与することができる(ex vivo)。従来のウイルスベースのシステムは、遺伝子導入のためのレトロウイルス、レンチウイルス、アデノウイルス、アデノ随伴及び単純ヘルペスウイルスベクターを含み得る。レトロウイルス、レンチウイルス、およびアデノ随伴ウイルスの遺伝子導入法では、宿主ゲノムへの組み込みが可能であり、しばしば挿入された導入遺伝子の長期発現をもたらす。さらに、多くの異なる細胞型および標的組織において高い形質導入効率が観察されている。

30

40

【0464】

レトロウイルスの指向性は、外来のエンベロープタンパク質を組み込み、標的細胞の潜在的な標的集団を拡大することによって変化させることができる。レンチウイルスベクターは、非分裂細胞を形質導入または感染させることができ、典型的には高いウイルス力価を産生するレトロウイルスベクターである。したがって、レトロウイルス遺伝子導入系の選択は標的組織に依存する。レトロウイルスベクターは、6～10 kbまでの外来配列のパッケージング能力を有するシス作用性の長い末端反復から構成される。最小シス作用性LTRは、ベクターの複製およびパッケージングのために十分であり、次いでそれを用いて治療遺伝子が標的細胞に組み込まれ、永続的な導入遺伝子発現が提供される。広く使用されているレトロウイルスベクターには、マウス白血病ウイルス(MuLV)、テナガザル白血病

50

ウイルス (GaLV)、サル免疫不全ウイルス (SIV)、ヒト免疫不全ウイルス (HIV)、およびそれらの組み合わせに基づくものが含まれる(例えば Buchscher et al., J. Virol. 66:2731-2739 (1992); Johann et al., J. Virol. 66:1635-1640 (1992); Sommerfeld et al., Virol. 176:58-59 (1990); Wilson et al., J. Virol. 63:2374-2378 (1989); Miller et al., J. Virol. 65:2220-2224 (1991); PCT/US94/05700参照)。

【0465】

レトロウイルスベクター、特にレンチウイルスベクターは、標的細胞への効率的な組み込みのために所定の長さより小さいポリヌクレオチド配列を必要とし得る。例えば、9 kb を超える長さのレトロウイルスベクターは、より小さいサイズのものと比較して低いウイルス力価をもたらし得る。いくつかの態様において、本開示の塩基エディターは、レトロウイルスベクターを介して効率的なパッケージングおよび標的細胞への送達を可能にするのに十分なサイズである。場合によっては、塩基エディターは、ガイド核酸および/またはターゲティング可能なヌクレアーゼ系の他の成分と共に発現された場合でも、効率的なパッケージングおよび送達を可能にするサイズである。

10

【0466】

一過性発現が好ましい用途では、アデノウイルスベースのシステムを使用することができる。アデノウイルスベクターは、多くの細胞型において非常に高い形質導入効率が可能であり、細胞分裂を必要としない。このようなベクターでは、高い力価および発現レベルが得られている。このベクターは比較的簡単な系で大量に生成できる。

【0467】

アデノ随伴ウイルス(「AAV」)ベクターは、標的核酸で細胞を形質導入するために使用することもでき、例えば、核酸およびペプチドのインビトロでの産生において、ならびにインビボおよびエクスピボでの遺伝子治療手順のために使用され得る(例えば、West et al., Virology 160:38-47 (1987); U.S. Patent No. 4,797,368; WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest. 94:1351 (1994))。組換えAAVベクターの構築は、米国特許第5,173,414号; Tratschin et al., Mol. Cell. Biol. 5:3251-3260 (1985); Tratschin, et al., Mol. Cell. Biol. 4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); および Samulski et al., J. Virol. 63:03822-3828 (1989)を含む多くの刊行物に記載されている。

20

30

【0468】

従って、本明細書に記載される塩基エディターは、ウイルスベクターを用いて送達され得る。塩基エディターシステムの1つ以上のコンポーネントは、1つ以上のウイルスベクター上にコードされ得る。例えば、塩基エディターおよびガイド核酸が、単一のウイルスベクター上にコードされ得る。他の場合には、塩基エディターおよびガイド核酸は、異なるウイルスベクター上にコードされる。いずれの場合にも、塩基エディターおよびガイド核酸は、各々、プロモーターおよびターミネーターに作動可能に連結され得る。

【0469】

ウイルスベクター上にコードされる成分の組み合わせは、選択されたウイルスベクターのカーゴサイズ制約によって決定され得る。

40

【0470】

塩基エディターのための非ウイルス性送達アプローチも利用可能である。非ウイルス核酸ベクターの1つの重要なカテゴリーは、有機または無機であり得るナノ粒子である。ナノ粒子は当技術分野でよく知られており、任意の適切なナノ粒子設計を用いて、ゲノム編集システム構成要素またはそのような構成要素をコードする核酸を送達することができる。例えば、有機(例えば、脂質および/またはポリマー)ナノ粒子は、本開示の特定の実施形態における送達ビヒクルとしての使用に好適であり得る。ナノ粒子製剤、および/または遺伝子導入において使用するための例示的な脂質を、表4に示す(下記)。

50

【表 4】

遺伝子移入のために用いられる脂質		
脂質	略語	特徴
1,2-Dioleoyl-sn-glycero-3-phosphatidylcholine	DOPC	Helper
1,2-Dioleoyl-sn-glycero-3-phosphatidylethanolamine	DOPE	Helper
Cholesterol		Helper
N-[1-(2,3-Dioleoyloxy)propyl]N,N,N-trimethylammonium chloride	DOTMA	Cationic
1,2-Dioleoyloxy-3-trimethylammonium-propane	DOTAP	Cationic
Diocetadecylamidoglycylspermine	DOGS	Cationic
N-(3-Aminopropyl)-N,N-dimethyl-2,3-bis(dodecyloxy)-1-propanaminium bromide	GAP-DLRIE	Cationic
Cetyltrimethylammonium bromide	CTAB	Cationic
6-Lauroxyhexyl ornithinate	LHON	Cationic
1-(2,3-Dioleoyloxypropyl)-2,4,6-trimethylpyridinium	2Oc	Cationic
2,3-Dioleoyloxy-N-[2(sperminecarboxamido-ethyl)-N,N-dimethyl-1-propanaminium trifluoroacetate	DOSPA	Cationic
1,2-Dioleoyl-3-trimethylammonium-propane	DOPA	Cationic
N-(2-Hydroxyethyl)-N,N-dimethyl-2,3-bis(tetradecyloxy)-1-propanaminium bromide	MDRIE	Cationic
Dimyristoxypropyl dimethyl hydroxyethyl ammonium bromide	DMRI	Cationic
3β-[N-(N',N'-Dimethylaminoethane)-carbamoyl]cholesterol	DC-Chol	Cationic
Bis-guanidium-tren-cholesterol	BGTC	Cationic
1,3-Dioleoyloxy-2-(6-carboxy-spermyl)-propylamide	DOSPER	Cationic
Dimethyloctadecylammonium bromide	DDAB	Cationic
Diocetadecylamidoglycylspermidin	DSL	Cationic
rac-[(2,3-Dioctadecyloxypropyl)(2-hydroxyethyl)]-dimethylammonium chloride	CLIP-1	Cationic
rac-[2(2,3-Dihexadecyloxypropyl-oxymethyloxy)ethyl]trimethylammonium bromide	CLIP-6	Cationic
Ethyl dimyristoylphosphatidylcholine	EDMPC	Cationic
1,2-Distearoyloxy-N,N-dimethyl-3-aminopropane	DSDMA	Cationic
1,2-Dimyristoyl-trimethylammonium propane	DMTAP	Cationic
O,O'-Dimyristyl-N-lysyl aspartate	DMKE	Cationic
1,2-Distearoyl-sn-glycero-3-ethylphosphocholine	DSEPC	Cationic
N-Palmitoyl D-erythro-sphingosyl carbamoyl-spermine	CCS	Cationic
N-t-Butyl-N0-tetradecyl-3-tetradecylaminopropionamidine	diC14-amidine	Cationic
Octadecenolyoxy[ethyl-2-heptadecenyl-3 hydroxyethyl]imidazolium chloride	DOTIM	Cationic
N1 -Cholesteryloxycarbonyl-3,7-diazanonane-1,9-diamine	CDAN	Cationic
2-(3-[Bis(3-amino-propyl)-amino]propylamino)-N-ditetradecylcarbamoylme-ethyl-acetamide	RPR209120	Cationic
1,2-dilinoleoyloxy-3-dimethylaminopropane	DLinDMA	Cationic
2,2-dilinoleyl-4-dimethylaminoethyl-[1,3]-dioxolane	DLin-KC2-	Cationic

遺伝子移入のために用いられる脂質

脂質	略語	特徴
dilinoleyl-methyl-4-dimethylaminobutyrate	DMA DLin-MC3- DMA	Cationic

10

20

30

40

50

表 5 は、遺伝子導入および/またはナノ粒子製剤における使用のための例示的ポリマーを
 列挙する。

【表 5】

遺伝子移入のために用いられるポリマー	
ポリマー	略語
Poly(ethylene)glycol	PEG
Polyethylenimine	PEI
Dithiobis (succinimidylpropionate)	DSP
Dimethyl-3,3'-dithiobispropionimide	DTBP
Poly(ethylene imine)biscarbamate	PEIC
Poly(L-lysine)	PLL
Histidine modified PLL	
Poly(N-vinylpyrrolidone)	PVP
Poly(propylenimine)	PPI
Poly(amidoamine)	PAMAM
Poly(amidoethylenimine)	SS-PAEI
Triethylenetetramine	TETA
Poly(β -aminoester)	
Poly(4-hydroxy-L-proline ester)	PHP
Poly(allylamine)	
Poly(α -[4-aminobutyl]-L-glycolic acid)	PAGA
Poly(D,L-lactic-co-glycolic acid)	PLGA
Poly(N-ethyl-4-vinylpyridinium bromide)	
Poly(phosphazene)s	PPZ
Poly(phosphoester)s	PPE
Poly(phosphoramidate)s	PPA
Poly(N-2-hydroxypropylmethacrylamide)	pHPMA
Poly (2-(dimethylamino)ethyl methacrylate)	pDMAEMA
Poly(2-aminoethyl propylene phosphate)	PPE-EA
Chitosan	
Galactosylated chitosan	
N-Dodacylated chitosan	
Histone	
Collagen	
Dextran-spermine	D-SPM

表 6 は、本明細書で記述される融合タンパク質をコードするポリヌクレオチドのための
 送達法を要約する。

40

50

【表 6】

送達	ベクター／モード	非分裂細胞への送達	発現の期間	ゲノム挿入	送達される分子の種類	
物理的	例えば電気穿孔、パーティクルガン、リン酸カルシウムトランスフェクション	YES	一過性	NO	核酸およびタンパク質	10
ウイルス性	レトロウイルス	NO	安定	YES	RNA	
	レンチウイルス	YES	安定	YES/NOだが改変を伴う	RNA	
	アデノウイルス	YES	一過性	NO	DNA	
	アデノ随伴ウイルス (AAV)	YES	安定	NO	DNA	
非ウイルス性	ワクシニアウイルス	YES	高度に一過性	NO	DNA	20
	単純ヘルペスウイルス	YES	安定	NO	DNA	
非ウイルス性	カチオン性リポソーム	YES	一過性	送達されるものによる	核酸およびタンパク質	
	ポリマーナノ粒子	YES	一過性	送達されるものによる	核酸およびタンパク質	
生物学的非ウイルス性	弱毒化細菌	YES	一過性	NO	核酸	
送達媒体	遺伝子操作バクテリオファージ	YES	一過性	NO	核酸	30
	哺乳類ウイルス様粒子	YES	一過性	NO	核酸	
	生物学的リポソーム：赤血球ゴーストおよびエキソソーム	YES	一過性	NO	核酸	

40

【0471】

別の局面において、例えばCas9またはそのバリエーションなどの核酸結合タンパク質および目的のゲノム核酸配列を標的とするgRNAのような、ゲノム編集システム構成要素またはそのような構成要素をコードする核酸の送達は、リボ核酸タンパク質 (RNP) を細胞に送達することによって達成され得る。RNPは、標的化gRNAと複合体を形成した核酸結合タンパク質、例えば、Cas9を含む。RNPは、エレクトロポレーション、ヌクレオフェクション、またはカチオン性脂質媒介方法、例えば、Zuris, J.A. et al., 2015, Nat. Biotechnology, 33(1):73-80によって報告されているもののような、公知の方法を用いて細胞に送達することができる。RNPは、CRISPR塩基編集システムにおける使用のために有利であ

50

り、特に一次細胞のようなトランスフェクトが困難な細胞のために有利である。さらに、RNPは、細胞におけるタンパク質発現で起こり得る困難を軽減することもでき、特に、CRISPRプラスミドにおいて使用され得るCMVまたはEF1Aなどの真核生物プロモーターがよく発現されない場合にはそうである。有利なことに、RNPの使用は、細胞への外来DNAの送達を必要としない。さらに、核酸結合タンパク質およびgRNA複合体を含むRNPは、経時的に分解されるので、RNPの使用は、オフターゲット効果を制限する可能性を有する。プラスミドベースの技術の場合と同様の方法で、RNPは、結合タンパク質(例えばCas9バリエーション)を送達するために、そして相同性依存修復(HDR)を導くために使用され得る。

【0472】

別の局面において、例えばCas9またはそのバリエーションなどの核酸結合タンパク質および目的のゲノム核酸配列を標的とするgRNAのような、ゲノム編集システム構成要素またはそのような構成要素をコードする核酸の送達は、リボ核タンパク質(RNP)を細胞に送達することによって達成され得る。RNPは、標的化gRNAと複合体を形成した核酸結合タンパク質、例えば、Cas9を含む。RNPは、エレクトロポレーション、ヌクレオフェクション、またはカチオン性脂質媒介方法、例えば、Zuris, J.A. et al., 2015, Nat. Biotechnology, 33(1):73-80によって報告されているもののような、公知の方法を用いて細胞に送達することができる。RNPは、CRISPR塩基編集システムにおける使用のために有利であり、特に一次細胞のようなトランスフェクトが困難な細胞のために有利である。さらに、RNPは、細胞におけるタンパク質発現で起こり得る困難を軽減することもでき、特に、CRISPRプラスミドにおいて使用され得るCMVまたはEF1Aなどの真核生物プロモーターがよく発現されない場合にはそうである。有利なことに、RNPの使用は、細胞への外来DNAの送達を必要としない。さらに、核酸結合タンパク質およびgRNA複合体を含むRNPは、経時的に分解されるので、RNPの使用は、オフターゲット効果を制限する可能性を有する。プラスミドベースの技術の場合と同様の方法で、RNPは、結合タンパク質(例えばCas9バリエーション)を送達するために、そして相同性依存修復(HDR)を導くために使用され得る。

【0473】

核酸分子発現をコードする塩基エディターを駆動するために使用されるプロモーターは、AAV ITRを含み得る。これは、ベクター中のスペースを占領してしまい得るさらなるプロモーター要素の必要性を排除するために有利であり得る。解放された追加のスペースは、ガイド核酸または選択マーカーなどの追加の要素の発現を駆動するために使用することができる。ITR活性は比較的弱いので、これは選択したヌクレアーゼの過剰発現による潜在的毒性を低減するために使用できる。

【0474】

任意の適切なプロモーターを使用して、塩基エディターおよび、適当な場合には、ガイド核酸の発現を誘導することができる。遍在性発現のために使用することができるプロモーターとしては、CMV、CAG、CBh、PGK、SV40、フェリチン重鎖または軽鎖などが挙げられる。脳または他のCNS細胞での発現については、適切なプロモーターとしては、全てのニューロンについてのシナプシンI、興奮性ニューロンについてのCaMKII、GABA作動性ニューロンについてのGAD67もしくはGAD65またはVGATなどが挙げられ得る。肝細胞での発現については、適切なプロモーターとしては、アルブミンプロモーターが挙げられる。肺細胞の発現については、適切なプロモーターはSP-Bを含むことができ、内皮細胞については、適切なプロモーターはICAMを含むことができ、造血細胞については、適切なプロモーターはIFN またはCD45を含むことができる。骨芽細胞用。適当なプロモーターにはOG-2が含まれる。

【0475】

いくつかの場合において、本開示の塩基エディターは、同じ核酸分子内で別々のプロモーターが塩基エディターおよび適合性ガイド核酸の発現を駆動することを可能にするのに十分な小ささである。例えば、ベクターまたはウイルスベクターは、塩基エディターをコードする核酸に作動可能に連結された第1のプロモーター、およびガイド核酸に作動可能に連結された第2のプロモーターを含むことができる。

10

20

30

40

50

【0476】

ガイド核酸の発現を駆動するために使用されるプロモーターには、U6またはH1などのPol IIIプロモーター、ならびにgRNAアデノ随伴ウイルス (AAV) を発現するためのPol IIプロモーターとイントロンのカセットの使用が含まれ得る。

【0477】

一以上のガイド核酸を伴うまたは伴わない本明細書に記載される塩基エディターは、アデノ随伴ウイルス (AAV)、レンチウイルス、アデノウイルス、または他のプラスミドもしくはウイルスベクター型を用いて、特に、例えば、米国特許第8,454,972号(アデノウイルスの製剤、用量)、米国特許第8,404,658号(AAVの製剤、用量)および米国特許第5,846,946号(DNAプラスミドの製剤、用量)からの製剤および用量、ならびにレンチウイルス、AAVおよびアデノウイルスが関わる臨床試験およびそれに関する出版物からの製剤および用量を用いて送達することができる。例えば、AAVについては、投与経路、製剤化および用量は、米国特許第8,454,972号およびAAVが関わる臨床試験におけるようにすることができる。アデノウイルスについては、投与経路、製剤化および用量は、米国特許第8,404,658号およびアデノウイルスが関わる臨床試験におけるようにすることができる。プラスミド送達については、投与経路、処方および用量は、米国特許第5,846,946号およびプラスミドが関わる臨床試験におけるようにすることができる。投与量は平均的な70 kgの個体(例えば成人男性)に基づいて算出するか、またはそれに外挿することができ、患者、対象、体重および動物種の異なる哺乳動物について調整することができる。投与の頻度は、年齢、性別、全般的な健康状態、患者又は対象の他の状態及び対処される特定の状態又は症状を含む通常の要因に依存して、医学又は獣医学の実施者(例えば医師、獣医師)の技量の範囲内である。ウイルスベクターを対象組織に注射することができる。細胞型特異的塩基編集のために、塩基エディターおよび任意のガイド核酸の発現を、細胞型特異的プロモーターによって駆動し得る。

【0478】

*in vivo*送達に関して、AAVは他のウイルスベクターよりも有利であり得る。場合によっては、AAVは低い毒性を可能にするが、これは、免疫応答を活性化させ得る細胞粒子の超遠心分離を必要としない精製方法に起因し得る。ある場合には、AAVは宿主ゲノムに組み込まれないため、挿入突然変異を引き起こす確率の低さを可能にする。

【0479】

AAVのパッケージング限界は4.5または4.75 Kbである。このことは、開示された塩基エディター並びにプロモーター及び転写ターミネーターは、単一のウイルスベクター中にフィットし得ることを意味する。4.5 Kbまたは4.75 Kbを超える構築物は、ウイルス産生を有意に減少させ得る。例えば、SpCas9は非常に大きく、遺伝子自体が4.1 Kbを超え、AAVに詰め込むことが困難であるため、本開示の実施形態は、従来の塩基エディターよりも短い長さの開示された塩基エディターを利用することを含む。いくつかの例では、塩基エディターは4 kb未満である。開示される塩基エディターは、4.5 kb、4.4 kb、4.3 kb、4.2 kb、4.1 kb、4 kb、3.9 kb、3.8 kb、3.7 kb、3.6 kb、3.5 kb、3.4 kb、3.3 kb、3.2 kb、3.1 kb、3 kb、2.9 kb、2.8 kb、2.7 kb、2.6 kb、2.5 kb、2 kb、または1.5 kb未満であり得る。いくつかの場合において、開示された塩基エディターは、長さが4.5 kb以下である。

【0480】

AAVは、AAV1、AAV2、AAV5、またはそれらの任意の組み合わせであり得る。標的とする細胞に関してAAVのタイプを選択することができる；例えば、AAV血清型1、2、5またはハイブリッドキャプシドAAV1、AAV2、AAV5またはそれらの任意の組合せを選択して、脳または神経細胞を標的化することができる；また、心臓組織を標的とするAAV4を選択することができる。AAV8は肝臓への送達に有用である。これらの細胞に関する特定のAAV血清型の表は、Grimm, D. et al, J. Virol. 82: 5887-5911 (2008)に見出され得る。

【0481】

10

20

30

40

50

レンチウイルスは複雑なレトロウイルスであり、有糸分裂細胞と有糸分裂後細胞の両方に感染してその遺伝子を発現する能力をもつ。最も一般的に知られているレンチウイルスはヒト免疫不全ウイルス (HIV) であり、これは広範囲の細胞型を標的とするために他のウイルスのエンベロープ糖タンパク質を使用する。

【0482】

レンチウイルスは以下のように調製できる。pCasES 10 (これはレンチウイルス伝達プラスミド骨格を含む)をクローニングした後、低継代(p=5)のHEK293FTを、T-75フラスコ中に播種し、抗生物質なしで、10%ウシ胎児血清を含むDMEM中でトランスフェクションの前日に50%コンフルエンスにした。20時間後、培地をOptiMEM (無血清) 培地に変え、トランスフェクションを4時間後に行った。10 µgのレンチウイルストランスファー

10

【0483】

レンチウイルスは以下のように精製できる。ウイルス上清を48時間後に回収する。上清をまずデブリから除去し、0.45 µm低タンパク質結合 (PVDF) フィルターを通して濾過する。次いで、これらを24,000 rpmで2時間超遠心機で遠心する。ウイルスペレットを50 µlのDMEM中に4 で一晩再懸濁する。次いでそれらを等分し、直ちに-80 で凍結する。

20

【0484】

別の実施形態では、ウマ伝染性貧血ウイルス (EIAV) に基づく最小の非霊長類レンチウイルスベクターも考えられる。別の実施形態では、血管静止タンパク質エンドスタチンおよびアンジオスタチンを発現するウマ感染性貧血ウイルスに基づくレンチウイルス遺伝子治療ベクター、RetinoStat RTMを、網膜下注射を介して送達することが考えられる。別の実施形態では、自己不活化レンチウイルスベクターの使用が考えられる。

【0485】

この系の任意のRNA、例えば、ガイドRNAまたは塩基エディターをコードするmRNAは、RNAの形態で送達することができる。塩基エディターをコードするmRNAは、インビトロ転写を用いて生成することができる。例えば、ヌクレアーゼmRNAは、以下の要素を含むPCRカセットを用いて合成することができる：T7プロモーター、任意でコザック配列 (GCCACC)、ヌクレアーゼ配列、および グロビン-ポリAテールからの3' UTRのような3' UTR。このカセットは、T7ポリメラーゼによる転写のために用いることができる。ガイドポリヌクレオチド(例えばgRNA)もまた、インビトロ転写を用いて、T7プロモーター、次いで配列「GG」、およびガイドポリヌクレオチド配列を含有するカセットから転写され得る。

30

発現を増強し、可能性のある毒性を低減するために、塩基エディターコード配列および/またはガイド核酸は、例えば擬似Uまたは5-メチル-Cを用いて、1以上の修飾ヌクレオシドを含むように修飾され得る。いくつかの実施形態では、gRNA分子は、最初と最後の3塩基についてホスホロチオエート結合および2' O-Me修飾を有する。

40

【0486】

いくつかの実施形態では、mRNAは、Cap 5' UTR ORF 3' UTRの形態を有する。いくつかの実施形態では、5' UTRは以下の通りである：

AGGAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAGAGCCACC

【0487】

いくつかの実施形態では、3' UTRは以下の通りである：

GCGGCCGCUUAAUUAAGCUGCCUUCUGCGGGGCUUGCCUUCUGGCCAUGCCCUUCUU
CUCUCCUUGCACCUGUACCUCUUGGUCUUUGAAUAAAGCCUGAGUAGGAAGAAAA
AA

50

AA

【 0 4 8 8 】

いくつかの実施形態では、塩基エディターは、以下の構造および配列を有する：

Cap-AGGAAUAAGAGAGAAAAGAAGAGUAAGAAGAAUAUAAGAGCCACCAUGAGCA
 GCGAGACAGGCCUGUGGCUGUGGAUCCUACACUGCGGAGAAGAAUCGAGCCCCACGA
 GUUCGAGGUGUUCUUCGACCCAGAGAGCUGCGGAAAGAGACAUGCCUGCUGUACGAG
 AUCAACUGGGGCGGCAGACACUCUAUCUGGCGGCACACAAGCCAGAACACCAACAAGC
 ACGUGGAAGUGAACUUUAUCGAGAAGUUACGACCGAGCGGUACUUCUGCCCCAACAC
 CAGAUGCAGCAUCACCGUUGUUUCUGAGCUGGUCCCCUUGCGGCGAGUGCAGCAGAGCC
 AUCACCGAGUUUCUGUCCAGAUAUCCCCACGUGACCCUGUUCAUCUAUAUCGCCCGGC
 UGUACCACCACGCCGAUCCUAGAAAUAGACAGGGACUGCGCGACCUGAUCAGCAGCGG
 AGUGACCAUCCAGAUCAUGACCGAGCAAGAGAGCGGCUACUGCUGGCGGAACUUCGUG
 AACUACAGCCCCAGCAACGAAGCCACUGGCCUAGAUAUCCUCACCUGUGGGUCCGAC
 UGUACGUGCUGGAACUGUACUGCAUUAUCCUGGGCCUGCCUCCAUGCCUGAACAUCCU
 GAGAAGAAAGCAGCCUCAGCUGACCUUCUUCACAAUCGCCUGCAGAGCUGCCACUAC
 CAGAGACUGCCUCCACACAUCCUGUGGGCCACCGGACUUAAGAGCGGAGGAUCUAGCG
 GCGGCUCUAGCGGAUCUGAGACACCUGGCACAAGCGAGUCUGCCACACCUGAGAGUAG
 CGGCGGAUCUUCUGGCGGCUCGACAAAGAAGUACUCUAUCGGACUGGCCAUCCGGCACC
 AACUCUGUUGGAUGGGCCGUGAUCACCGACGAGUACAAGGUGCCCAGCAAGAAAUUCA
 AGGUGCUGGGCAACACCGACCGGCACAGCAUCAAGAAGAAUCUGAUCGGCGCCUGCU
 GUUCGACUCUGGCGAAACAGCCGAAGCCACCAGACUGAAGAGAACCGCCAGGCGGAGA
 UACACCCGGCGGAAGAACCAGGAUCUGCUACCUGCAAGAGAUCUUCAGCAACGAGAUGG
 CCAAGGUGGACGACAGCUUCUCCACAGACUGGAAGAGUCCUUCUGGUGGAAGAGGA
 CAAGAAGCACGAGCGGCACCCCAUCUUCGGCAACAUCGUGGAUGAGGUGGGCCUACCAC
 GAGAAGUACCCACCAUCUACCACCUGAGAAAGAAACUGGUGGACAGCACCGACAAGG
 CCGACCUGAGACUGAUCUACCUGGCUCUGGCCACAUGAUCUAGUUCGGGGCCACUUC
 UCUGAUCGAGGGCGAUCUGAACCCCGACAACAGCGACGUGGACAAGCUGUUCAUCCAG
 CUGGUGCAGACCUACAACCAGCUGUUCGAGGAAAACCCCAUCAACGCCUCUGGGCGUGG
 ACGCCAAGGCUAUCCUGUCUGCCAGACUGAGCAAGAGCAGAAGGCUGGAAAACCUGAU
 CGCCCAGCUGCCUGGCGAGAAGAAGAAUGGCCUGUUCGGCAACCUGAUUGCCCUGAGC
 CUGGGACUGACCCCUAACUUCAAGAGCAACUUCGACCUGGGCCGAGGAUGCCAAACUGC
 AGCUGAGCAAGGACACCUACGACGACGACCUGGACAAUCUGCUGGCCCAGAUCCGGCGA
 UCAGUACGCCGACUUGUUUCUGGGCCGCAAGAACCUGUCCGACGCCAUCCUGCUGAGC
 GAUAUCCUGAGAGUGAACACCGAGAUCACAAAGGCCCCUCUGAGCGCCUCUAUGAUC
 AGAGAUACGACGAGCACCACCAGGAUCUGACCCUGCUGAAGGCCCCUCGUUAGACAGCA
 GCUGCCAGAGAAGUACAAAGAGAUUUUCUUCGAUCAGUCCAAGAACGGCUACGCCGGC
 UACAUUGAUGGGCGGAGCCAGCCAAGAGGAAUUCUACAAGUUCAUCAAAGCCCAUCCUGG
 AAAAGAUGGACGGCACCGAGGAACUGCUGGUCAAGCUGAACAGAGAGGACCUGCUGCG
 GAAGCAGCGGACCUUCGACAAUGGCUCUAUCCUCACCAGAUCCACCUGGGAGAGCUG
 CACGCCAUUCUGCGGAGACAAGAGGACUUUUACCCAUUCUGAAGGACAACCGGGAAA
 AGAUCGAGAAGAUCCUGACCUUCAGGAUCCCUACUACGUGGGACCACUGGCCAGAGG
 CAAUAGCAGAUUCGCCUGGAUGACCAGAAAGAGCGAGGAAACCAUCACACCCUGGAAC
 UUCGAGGAAGUGGUGGACAAGGGCGCCAGCGCUCAGUCCUUCAUCGAGCGGAUGACCA
 ACUUCGAUAAGAACCUGCCUAACGAGAAGGUGCUGCCAAGCACUCCUGCUGUAUGA
 GUACUUCACCGUGUACAACGAGCUGACCAAGUGAAAUAACGUGACCGAGGGAAUGAGA
 AAGCCCGCCUUCUGAGCGGCGAGCAGAAAAGGCCAUUGUGGAUCUGCUGUUCAAGA
 CCAACCGGAAAGUGACCGUGAAGCAGCUGAAGAGGACUACUUCAAGAAAUCGAGUG
 CUUCGACAGCGUGGAAAUCAGCGGCGUGGAAGAUCGGUCAAUGCCAGCCUGGGCACA
 UACCACGACCUGCUGA AAAUUAUCAAGGACAAGGACUUCUGGACAACGAAGAGAACG
 AGGACAUUCUCGAGGACAUCGUGCUGACCCUGACACUGUUUGAGGACAGAGAGAUGA

10

20

30

40

50

UCGAGGAACGGCUGAAAACAUACGCCACCUGUUCGACGACAAAGUGAUGAAGCAACU
GAAGCGGAGGCGGUACACAGGCUGGGGCAGACUGUCUCGGAAGCUGAUAACGGCAUC
CGGGUAAGCAGUCCGGCAAGACAAUCCUGGAUUUCCUGAAGUCCGACGGCUUCGCCA
ACAGAAACUUCAUGCAGCUGAUCCACGACGACAGCCUGACCUUUAAGAGGACAUCCA
GAAAGCCCAGGUGUCCGGCCAAGGCGAUUCUCUGCAGGACACAUUGCCAACCUGGCC
GGAUCUCCCGCCAUAAGAAGGGCAUCCUGCAGACAGUGAAGGUGGUGGACGAGCUUG
UGAAAGUGAUGGGCAGACACAAGCCCCGAGAACAUCGUGAUCGAAAUGGCCAGAGAGAA
CCAGACCACACAGAAGGGCCAGAAGAACAGCCGCGAGAGAAUGAAGCGGAUCGAAGAG
GGCAUCAAAAGAGCUGGGCAGCCAGAUCUGAAAAGAACCCCCUGGAAAACACCCAGC
UGCAGAACGAGAAGCUGUACCUGUACUACCUGCAGAAUGGACGGGAUAUGUACGUGG
ACCAAGAGCUGGACAUAACCGGCUGAGCGACUACGAUGUGGACCAUAUCGUGCCCCA
GAGCUUUCUGAAGGACGACUCCAUCGAUAACAAGGUCCUGACCAGAAGCGACAAGAAC
CGGGGCAAGAGCGAUAACGUGCCCUCGGAAGAGGUGGUCAGAAGAUGAAGAACUACU
GGCGACAGCUGCUGAACGCCAAGCUGAUUACCCAGCGGAAGUUCGAUAACCUGACCAA
GGCCGAGAGAGGGCGCCUGAGCGAACUUGAUAAGGCCGGCUUCAUUAAGCGGCAGCUG
GUGGAAACCCGGCAGAUACCCAACACGUGGCACAGAUUCUGGACUCCGGGAUGAACA
CUAAGUACGACGAGAAUGACAAGCUGAUCCGGGAAGUGAAGUCAUCACCCUGAAGUC
UAAGCUGGUGUCCGAUUUCCGGAAGGAUUUCCAGUUCUACAAGUGCGGGAAAUCAA
CAACUACCAUCACGCCACGACGCCUACCUGAAUGCCGUUGUUGGAACAGCCCUGAUC
AGAAGUAUCCCAAGCUGGAAAGCGAGUUCGUGUACGGCGACUACAAGGUGUACGACG
UGCGGAAGAUGAUCGCCAAGAGCGAACAAGAGAUCGGCAAGGCUACCGCCAAGUACUU
UUUCUACAGCAACAUCAUGAACUUUUUCAAGACAGAGAUCACCCUGGCCAACGGCGAG
AUCCGGAAAAGACCCCUGAUCGAGACAAACGGCGAAACCGGGGAGAUUCGUGUGGGUA
AGGGCAGAGAUUUUGCCACAGUGCGGAAAGUGCUGAGCAUGCCCCAAGUGAAUAUCGU
GAAGAAAACCGAGGUGCAGACAGGCGGCUUCAGCAAAGAGUCUAUCCUGCCUAAGCGG
AACAGCGAUAAGCUGAUCGCCAGAAAGAAGGACUGGGACCCUAAGAAGUACGGCGGCU
UCGAUAGCCCUACCGUGGCCUAUUCUGUGCUGGUGGUGGCCAAAGUGGAAAAGGGCAA
GUCCAAAAGCUCUAAAGAGCGUGAAAGAGCUGCUGGGGAUCACCAUCAUGGAAAGAAGC
AGCUUUGAGAAGAACCCGAUCGACUUUCUGGAAGCCAAGGGCUACAAGAAGUCAAGA
AGGACCUCAUCAUAAGCUCGCCAAGUACAGCCUGUUCGAGCUGGAAAUGGCCGGAA
GCGGAUGCUGGCCUCAGCAGGCGAACUGCAGAAAGGCAUAUGAACUGGCCCUAGC
AAAUACGUAACUUCUGUACCUGGCCAGCCACUAUGAGAAGCUGAAGGGCAGCCCCG
AGGACAAUGAGCAAAGCAGCUGUUUGUGGAACAGCACAAGCACUACCUGGACGAGAU
CAUCGAGCAGAUACGCGAGUUCUCCAAGAGAGUGAUCCUGGCCGACGCUAACCUGGAU
AAGGUGCUGUCUGCCUAUAACAAGCACCCGGGACAAGCCUAUCAGAGAGCAGGCCGAGA
AUAUCAUCCACCUUUUACCCUGACCAACCUGGGAGCCCCUGCCGCCUUAAGUACUU
CGACACCACCAUCGACCGGAAGAGGUACACCAGCACCAAGAGGUGCUGGACGCCACA
CUGAUCCACCAGUCUAUCACCGGCCUGUACGAAACCCGGAUCGACCUGUCUCAGCUCG
GCGGCGAUUCUGGUGGUUCUGGCGGAAGUGGCGGAUCCACCAUUCUGAGCGACAUCAU
CGAAAAGAGACAGGCAAGCAGCUCGUGAUCCAAGAAUCCAUCCUGAUGCUGCCUGAA
GAGGUUGAGGAAGUGAUCGGCAACAAGCCUGAGUCCGACAUCUGGUGCACACCGCCU
ACGAUGAGAGCACCGAUGAGAACGUCAUGCUGCUGACAAGCGACGCCCCUGAGUACAA
GCCUUGGGCUCUCGUGAUUCAGGACAGCAAUGGGGAGAACAAGAUAAGAUGCUGAGC
GGAGGUAGCGGAGGCAGUGGCGGAAGCACAAACCUGUCUGAUUAUCAUUGAAAAGAAA
CCGGGAAGCAACUGGUCAUUAAGAGUCCAUUCUCAUGCUCGCCGAAGAAGUCGAGGA
AGUCAUUGGAAACAAACCCGAGAGCGAUAUUCUGGUCCACACAGCCUAUGACGAGUCU
ACAGACGAAAACGUGAUGCUCCUGACCUCUGACGCUCCCAGUAUAAGCCCUGGGCAC
UUGUUAUCCAGGACUCUAACGGGGAAAACAAAUAUUAAGUUGUCCGGCGGCAGCAA
GCGGACAGCCGAUGGAUCUGAGUUCGAGAGCCCCAAGAAGAAACGGAAGGUgGAGUaa
GCGGCCGCUUAAUUAAGCUGCCUUCUGCGGGGCUUGCCUUCUGGCCAUGCCCUUCUU

10

20

30

40

50

CUCUCCUUGCACCCUGUACCCUCUUGGUCUUUGAAUAAAGCCUGAGUAGGAAGAAAAA
 AA
 AA
 【0489】

本開示はいくつかの態様において、細胞または生物を改変する方法を包含する。細胞は、原核細胞または真核細胞であり得る。細胞は哺乳動物細胞であり得る。哺乳動物細胞は、非ヒト霊長類、ウシ、ブタ、齧歯類またはマウス細胞であり得る。本開示の塩基エディター、組成物および方法によって細胞に導入される改変は、細胞および細胞の子孫が、抗体、澱粉、アルコールまたは他の所望の細胞アウトプットなどの生物学的産物の生産を改善するために改変されるようなものであり得る。本開示の方法によって細胞に導入される改変は、細胞およびその子孫が、産生される生物学的産物を変化させる改変を含むようなものであり得る。

10

【0490】

システムは、1つまたは複数の異なるベクターを含むことができる。一態様では、塩基エディターは、所望の細胞型、好ましくは真核細胞、好ましくは哺乳動物細胞またはヒト細胞の発現のためにコドン最適化される。

【0491】

一般に、コドン最適化とは、目的の宿主細胞における発現を増強するために、天然配列の少なくとも一つのコドン(例えば約1個、2個、3個、4個、5個、10個、15個、20個、25個、50個以上のコドン)を、天然アミノ酸配列を維持しながら、その宿主細胞の遺伝子においてより頻繁にまたは最も頻繁に使用されるコドンで置換することによって、核酸配列を改変するプロセスをいう。様々な種は特定のアミノ酸の特定のコドンについて特定の偏りを示す。コドンバイアス(生物間のコドン利用の違い)は、メッセンジャーRNA (mRNA)の翻訳効率としばしば相関し、それは特に、翻訳されるコドンの性質および特定のトランスファーRNA (tRNA)分子の利用可能性に依存すると考えられている。細胞内の選択されたtRNAの優位性は、一般にペプチド合成で最も頻繁に使われるコドンを反映している。従って、遺伝子は、コドン最適化に基づいて、所与の生物における最適な遺伝子発現に合わせて調整することができる。コドン使用表は、例えば、「コドン使用データベース」www.kazusa.or.jp/codon/ (2002年7月9日に訪問した)で容易に入手可能であり、これらの表は、多くの方法で適合させることができる。Nakamura, Y., et al. "Codon usage tabulated from the international DNA sequence databases: status for the year 2000" *Nucl. Acids Res.* 28:292 (2000)参照。例えばGene Forge (Aptagen; Jacobus, Pa.)のような、特定の宿主細胞における発現のために特定の配列をコドン最適化するためのコンピュータアルゴリズムも利用可能である。ある態様において、操作されたヌクレアーゼをコードする配列中の一つ以上のコドン(例えば1、2、3、4、5、10、15、20、25、50以上、または全てのコドン)は、特定のアミノ酸について最も頻繁に使用されるコドンに対応する。

20

30

【0492】

パッケージング細胞は、典型的には、宿主細胞に感染し得るウイルス粒子を形成するために使用される。そのような細胞は、アデノウイルスをパッケージングする293細胞、およびレトロウイルスをパッケージングするpsi.2細胞またはPA317細胞を含む。遺伝子治療に使用されるウイルスベクターは、通常、核酸ベクターをウイルス粒子にパッケージ化する細胞株を作製することによって生成される。ベクターは、典型的には、パッケージングおよびその後の宿主への組み込みに必要な最小のウイルス配列を含み、他のウイルス配列は、発現されるべきポリヌクレオチドのための発現カセットによって置き換えられる。欠けているウイルス機能は、典型的にはパッケージング細胞株によってトランスに供給される。例えば、遺伝子治療に使用されるAAVベクターは、典型的には、パッケージングおよび宿主ゲノムへの組み込みに必要なAAVゲノムからのITR配列のみを有する。ウイルスDNAは、他のAAV遺伝子、すなわちrepおよびcapをコードするがITR配列を欠くヘルパープラスミドを含む細胞株においてパッケージングされ得る。細胞株は、ヘルパーとしてのア

40

50

デノウイルスによっても感染され得る。ヘルパーウイルスは、AAVベクターの複製およびヘルパープラスミドからのAAV遺伝子の発現を促進することができる。ある場合には、ヘルパープラスミドは、ITR配列の欠如のために、有意な量ではパッケージングされない。アデノウイルスによる汚染は、例えばAAVよりもアデノウイルスの方が感受性である熱処理によって減少させることができる。

【0493】

[医薬組成物]

本開示の他の態様は、本明細書に記載される塩基エディター、融合タンパク質、または融合タンパク質-ガイドポリヌクレオチド複合体のいずれかを含む医薬組成物に関する。用語「医薬組成物」は、本明細書中で使用される場合、薬学的使用のために処方される組成物を指す。いくつかの態様において、医薬組成物は、薬学的に許容される担体をさらに含む。ある態様において、医薬組成物は、さらなる剤(例えば特異的送達、半減期の延長のためのもの、または他の治療化合物)を含む。

10

【0494】

ここで使用されるところの用語「薬学的に許容される担体」は、液体または固体の充填剤、希釈剤、賦形剤、製造助剤(例えば潤滑剤、タルクマグネシウム、ステアリン酸カルシウムまたはステアリン酸亜鉛、あるいはステアリン酸)、または溶媒カプセル化材料などの、身体のある部位(例えば送達部位)から別の部位(例えば器官、組織または身体の一部)への化合物の運搬または輸送に関する薬学的に許容される材料、組成物またはビヒクルを意味する。薬学的に許容される担体は、製剤の他の成分と適合し、対象の組織に有害でないという意味で許容される(例えば生理的適合性、無菌性、生理的pH等)。

20

【0495】

薬学的に許容される担体として役立つことができる物質のいくつかの非限定的な例は、以下を含む：(1) ラクトース、グルコースおよびスクロースのような糖；(2) コーンスターチ、ジャガイモでん粉等のでん粉；(3) セルロース及びその誘導体(カルボキシメチルセルロースナトリウム、メチルセルロース、エチルセルロース、微結晶セルロース、酢酸セルロース等)；(4) トラガント末；(5) モルト；(6) ゼラチン；(7) ステアリン酸マグネシウム、ラウリル硫酸ナトリウム、タルク等の潤滑剤；(8) ココアバター、坐剤用ワックス等の添加剤；(9) 落花生油、綿実油、ペニバナ油、ゴマ油、オリーブ油、コーン油、大豆油等の油；(10) プロピレングリコール等のグリコール；(11) ポリオール、例えばグリセリン、ソルビトール、マンニトール及びポリエチレングリコール(PEG)；(12) エステル、例えばオレイン酸エチルおよびラウリン酸エチルなど；(13) 寒天；(14) 水酸化マグネシウム、水酸化アルミニウム等の緩衝剤；(15) アルギン酸；(16) パイロジェンフリー水；(17) 生理食塩液；(18) リンガー液；(19) エチルアルコール；(20) pH緩衝液；(21) ポリエステル、ポリカーボネート及び/又はポリ無水物；(22) ポリペプチドおよびアミノ酸のような増量剤(23) エタノールのような血清アルコール；及び(23) 製剤に使用される他の非毒性適合性物質。湿潤剤、着色剤、離型剤、コーティング剤、甘味剤、風味料、香料、保存剤および抗酸化剤も製剤中に存在させることができる。「賦形剤」、「担体」、「薬学的に許容される担体」、「ベヒクル」などの用語は、本明細書では交換可能に使用される。

30

【0496】

いくつかの態様において、医薬組成物は、対象への送達のために、例えば遺伝子編集のために製剤化される。本明細書に記載の医薬組成物を投与する適切な経路は、限定されるものではないが、局所、皮下、経皮、皮内、病巣内、関節内、腹腔内、膀胱内、経粘膜、歯肉、歯内、蝸牛内、経鼓膜、器官内、硬膜外、髄腔内、筋肉内、静脈内、血管内、骨内、眼周囲、腫瘍内、脳内、および脳室内投与を含む。

40

【0497】

いくつかの態様において、本明細書に記載される医薬組成物は、患部(例えば腫瘍部位)に局所的に投与される。いくつかの実施形態において、本明細書に記載される医薬組成物は、注射、カテーテル、坐剤、またはインプラントによって、対象に投与され、インプラントは、多孔質、非多孔質、またはゼラチン質材料であり、例えば、シアラスティック膜

50

、または繊維などの膜を含む。

【0498】

他の態様において、本明細書に記載される医薬組成物は、制御放出システムにおいて送達される。一実施形態では、ポンプを使用することができる(例えばLanger, 1990, Science 249: 1527-1533; Sefton, 1989, CRC Crit. Ref. Biomed. Eng. 14:201; Buchwald et al., 1980, Surgery 88:507; Saudek et al, 1989, N. Engl. J. Med. 321:574参照)。別の実施形態では、ポリマー材料を使用することができる。(例えばMedical Applications of Controlled Release (Langer and Wise eds., CRC Press, Boca Raton, Fla., 1974); Controlled Drug Bioavailability, Drug Product Design and Performance (Smolen and Ball eds., Wiley, New York, 1984); Ranger and Peppas, 1983, Macromol. Sci. Rev. Macromol. Chem. 23:61. See also Levy et al., 1985, Science 228: 190; During et al., 1989, Ann. Neurol. 25:351; Howard et al., 1989, J. Neurosurg. 71: 105.)。他の制御放出システムは、例えば、上記Langerに記載されている。

10

【0499】

いくつかの態様において、医薬組成物は、対象、例えばヒトへの静脈内または皮下投与に適合された組成物として、ルーチンの手順に従って製剤化される。いくつかの態様において、注射による投与のための医薬組成物は、可溶化剤としての無菌等張使用の溶液および注射部位の痛みを緩和するためのリグノカインなどの局所麻酔薬である。一般に、成分は、単位投与形態、例えば、活性剤の量を示すアンプルまたはサシエットのよう密閉容器内の乾燥凍結乾燥粉末または水を含まない濃縮物として、別々にまたは一緒に供給される。薬剤が注入によって投与される場合には、無菌の医薬グレードの水または生理食塩水を含む注入ボトルを用いてそれを分注することができる。医薬組成物が注射によって投与される場合、投与前に成分を混合することができるように、注射用滅菌水または生理食塩水のアンプルを提供することができる。

20

【0500】

全身投与のための医薬組成物は、液体、例えば滅菌生理食塩水、乳酸化リンゲル液またはハンク液であり得る。さらに、医薬組成物は、固体形態であって、使用の直前に再溶解または懸濁され得る。凍結乾燥形態もまた考えられる。医薬組成物は、非経口投与にも適した、リポソームまたは微結晶などの脂質粒子または小胞内に含有することができる。粒子は、組成物がその中に含まれる限り、単ラメラまたは複数ラメラのような任意の適切な構造であり得る。化合物は、融合性脂質ジオレオイルホスファチジルエタノールアミン(DOPE)、低量(5-10モル%)のカチオン性脂質を含み、ポリエチレングリコール(PEG)コーティングにより安定化された、「安定化プラスミド脂質粒子」(SPLP)中に捕捉され得る(Zhang Y. P. et al, Gene Ther. 1999, 6: 1438-47)。このような粒子および小胞には、N-[1-(2,3-ジオレオイルキシ)プロピル]-N,N,N-トリメチル-アモニウムメチル硫酸塩、あるいは「DOTAP」のような正電荷脂質が特に好ましい。このような脂質粒子の調製はよく知られている。例えば、米国特許第4,880,635; 4,906,477; 4,911,928; 4,917,951; 4,920,016;および4,921,757号参照(その各々は、参照により本明細書に組み込まれる)。

30

40

【0501】

本明細書に記載の医薬組成物は、例えば、単位用量として投与または包装することができる。「単位用量」という用語は、本開示の医薬組成物に関して使用される場合、対象のための単一用量として適した物理的に個別の単位を指し、各単位は、必要な希釈剤(例えば担体、またはビヒクル)と合わせて所望の治療効果を生じるように計算された所定量の活性物質を含有する。

【0502】

さらに、この医薬組成物は、(a)凍結乾燥形態の本発明の化合物を含有する容器、および(b)本発明の凍結乾燥化合物の再構成または希釈のために使用される、薬学的に許容される希釈剤(例えば無菌のもの)を含有する第2の容器を含む薬学的キットとして提供する

50

ことができる。必要に応じて、そのような容器には、医薬品又は生物学的製剤の製造、使用又は販売を規制する政府機関によって規定された様式の通知であって、人に投与するための製造、使用又は販売の機関による承認を反映するものとすることができる。

【0503】

別の態様では、上記の疾患の治療にとって有用な材料を含有する製品が含まれる。一部の実施形態では、製品は、容器およびラベルを含む。好適な容器としては、例えば、ビン、バイアル、シリンジ、および試験管が挙げられる。容器を、ガラスまたはプラスチックなどの様々な材料から形成させることができる。一部の実施形態では、容器は、本明細書に記載の疾患を治療するのに有効である組成物を保持し、滅菌アクセスポートを有してもよい。例えば、容器は、皮下注射針によって貫通可能なストッパーを有する静脈内溶液バッグまたはバイアルであってもよい。組成物中の活性薬剤は、本発明の化合物である。一部の実施形態では、容器上のラベルまたは容器と会合したラベルは、組成物が選択された疾患を治療するために使用されることを示す。製品は、リン酸緩衝塩水、リンゲル溶液、またはデキストロス溶液などの薬学的に許容される緩衝剤を含む第2の容器をさらに含んでもよい。それは、他の緩衝剤、希釈剤、フィルター、針、シリンジ、使用説明書を含むパッケージインサートなどの、商業的観点およびユーザー観点から望ましい他の材料をさらに含んでもよい。

10

【0504】

いくつかの態様において、本明細書に記載される融合タンパク質、gRNA、および/または複合体のいずれかは、医薬組成物の一部として提供される。いくつかの態様において、医薬組成物は、本明細書に提供される融合タンパク質のいずれかを含む。いくつかの態様において、医薬組成物は、本明細書に提供される複合体のいずれかを含む。いくつかの態様において、医薬組成物は、gRNAおよびカチオン性脂質と複合体を形成するRNA-ガイドヌクレアーゼ(例えばCas9)を含むリボ核タンパク質複合体を含む。ある態様において、医薬組成物は、gRNA、核酸プログラム可能DNA結合タンパク質、カチオン性脂質、および薬学的に許容される賦形剤を含む。医薬組成物は、場合により、1つ以上の追加の治療活性物質を含むことができる。

20

【0505】

ある態様において、本明細書に提供される組成物は、対象内で標的化されたゲノム改変をもたらすために、対象、例えばヒト対象に投与される。ある態様において、細胞が対象から取得され、本明細書に提供される医薬組成物と接触される。いくつかの態様において、対象から取り出されてex vivoで医薬組成物と接触された細胞は、場合により、所望のゲノム改変が細胞において実行または検出された後に、対象に再導入される。ヌクレアーゼを含む医薬組成物を送達する方法は既知であり、例えば、米国特許第6,453,242; 6,503,717; 6,534,261; 6,599,692; 6,607,882; 6,689,558; 6,824,978; 6,933,113; 6,979,539; 7,013,219; and 7,163,824号に記載されており、その開示全体が参照により本明細書に組み込まれる。本明細書に提供される医薬組成物の説明は、主に、ヒトへの投与に適した医薬組成物に向けられているが、そのような組成物は、一般に、あらゆる種類の動物または生物への投与に適していることが、当業者によって理解されるであろう。

30

【0506】

種々の動物への投与に適した組成物を与えるための、ヒトへの投与に適した医薬組成物の改変は十分に理解されており、通常の熟練した獣医薬理学者は、もし必要だとしても単に通常の実験で、そのような改変を設計および/または実施することができる。薬学的組成物の投与が意図される対象には、限定されるものではないが、ヒトおよび/または非ヒト霊長類; 哺乳動物、家畜、ペット、およびウシ、ブタ、ウマ、ヒツジ、ネコ、イヌ、マウス、および/またはラットなどの商業的に関連のある哺乳動物; ニワトリ、カモ、ガチョウ及び/又は七面鳥のような商業的に関連のある鳥類が含まれる。

40

【0507】

本明細書に記載される薬学的組成物の製剤は、薬学の分野において公知の、または今後開発される任意の方法によって調製することができる。一般に、このような調製方法は、

50

活性成分を賦形剤および/または1つ以上の他の補助成分と会合させ、次いで、必要および/または所望であれば、製品を所望の単回または複数回投与単位に成形および/または包装する工程を含む。薬学的製剤はさらに、薬学的に許容される賦形剤を含むことができ、それは、本明細書で使用される場合、所望の特定の剤形に適した、溶媒、分散媒体、希釈剤、または他の液体ビヒクル、分散または懸濁助剤、界面活性剤、等張剤、増粘剤または乳化剤、保存剤、固体結合剤、潤滑剤などのいずれかおよび全てを含む。Remington's The Science and Practice of Pharmacy, 21st Edition, A. R. Gennaro (Lippincott, Williams & Wilkins, Baltimore, MD, 2006 (参照によりその全体が本明細書に組み込まれる))は、薬学的組成物を製剤化する際に使用される種々の賦形剤およびその調製のための公知の技術を開示する。ヌクレアーゼを含む医薬組成物を製造するためのさらなる適切な方法、試薬、賦形剤および溶媒については、参照によりその全体が本明細書に組み込まれるPCT出願PCT/US2010/055131(2010年11月2日出願、公開番号WO2011053982 A8)も参照のこと。

10

【0508】

あらゆる従来の賦形剤媒体は、望ましくない生物学的効果を生じさせること、または他のかたちで医薬組成物の何らかの他の成分と有害な様式で相互作用することなどによって、物質またはその誘導体と不適合である場合を除き、その使用が本開示の範囲内にあると考えられる。

【0509】

上記の組成物は、有効量で投与することができる。有効量は、投与方法、治療される特定の状態、および所望の結果に依存する。それはまた、状態のステージ、対象の年齢および身体的状態、もしあれば併用療法の性質、および医師によく知られた同様の因子に依存し得る。治療用途のためには、それは医学的に望ましい結果を達成するのに十分な量である。

20

【0510】

一部の実施形態では、本開示による組成物を、限定されるものではないが、以下のうちの1つまたは複数：自己免疫障害(例えば、糖尿病、狼瘡、多発性硬化症、乾癬、関節リウマチ)；炎症性障害(例えば、関節炎、骨盤内炎症性疾患)；感染症(例えば、ウイルス感染症(例えば、HIV、HCV、RSV)、細菌感染症、真菌感染症、敗血症)；神経障害(例えば、アルツハイマー病、ハンチントン病；自閉症；デュシェンヌ型筋ジストロフィー)；心血管障害(例えば、アテローム性動脈硬化症、高コレステロール血症、血栓症、凝固障害、黄斑変性などの血管新生障害)；増殖性障害(例えば、がん、良性新生物)；呼吸器障害(例えば、慢性閉塞性肺疾患)；消化器障害(例えば、炎症性腸疾患、潰瘍)；骨格筋障害(例えば、線維筋痛症、関節炎)；内分泌、代謝、および栄養障害(例えば、糖尿病、骨粗鬆症)；泌尿器系障害(例えば、腎疾患)；精神障害(例えば、抑鬱、統合失調症)；皮膚障害(例えば、創傷、湿疹)；血液およびリンパ管障害(例えば、貧血、血友病)などの様々な疾患、障害、および/または状態のいずれかの治療のために使用することができる。

30

【0511】

[キット]

本開示の様々な態様は、塩基エディターシステムを含むキットを提供する。一実施形態では、キットは、核酸塩基エディター融合タンパク質をコードするヌクレオチド配列を含む核酸構築物を含む。この融合タンパク質は、デアミナーゼ(例えば、シチジンデアミナーゼまたはアデニンデアミナーゼ)および核酸プログラム可能DNA結合タンパク質(napDNA bp)を含む。ある態様において、キットは、対象核酸分子、例えば表3Aおよび3Bに特定される遺伝子における疾患関連変異を標的とすることができる、少なくとも一つのガイドRNAを含む。いくつかの実施形態において、キットは、少なくとも一つのガイドRNAをコードするヌクレオチド配列を含む核酸構築物を含む。

40

【0512】

いくつかの実施形態において、キットは、それを使用して表3Aおよび3Bにおける一

50

つ以上の遺伝子の1つ以上の疾患関連変異を編集するための使用説明書を提供する。説明書は、一般に、核酸分子を編集するためのキットの使用に関する情報を含む。他の実施形態では、説明書は、以下のうちの少なくとも1つを含む：注意事項；警告；臨床試験；および/または参考資料。使用説明書は、容器（もしあれば）に直接印刷するか、容器に貼付するラベルとして印刷するか、または容器内にもしくは容器と共に提供される独立したシート、パンフレット、カードまたはフォルダーとして印刷され得る。さらなる実施形態では、キットは、適切な動作パラメータのためのラベルまたは別個のインサート（添付文書）の形態で説明書を含むことができる。さらに別の実施形態では、キットは、検出、校正、または正規化のための標準として使用される、適切なポジティブおよびネガティブコントロールまたは対照サンプルを有する1つ以上の容器を含むことができる。キットは、（滅菌）リン酸緩衝生理食塩水、リンゲル液、またはデキストロース溶液などの薬学的に許容される緩衝液を含む第2の容器をさらに含むことができる。さらに、他の緩衝剤、希釈剤、フィルター、針、注射器、および使用説明書付きの添付文書を含む、商業的観点および使用者の観点から望ましい他の物質を含むことができる。

【0513】

特定の実施形態では、キットは、アルファ-1アンチトリプシン欠損症（A1AD）を有する対象の治療のために有用である。

【0514】

塩基エディターシステムの方法および組成物ならびに使用を包含する、以下の番号付けされた追加の実施形態が本明細書で企図される：

1. 必要とする対象において疾患を治療する方法であって、

ガイドポリヌクレオチドまたはそれをコードする核酸；

ポリヌクレオチドプログラム可能なDNA結合ドメイン又はそれをコードする核酸、及びデアミナーゼドメインまたはそれをコードする核酸

を含む塩基エディターシステムを対象に投与することを含み、

該ポリヌクレオチドが、塩基エディターシステムをターゲティングして対象の細胞のSERPINA1ポリヌクレオチド中の核酸塩基の脱アミノ化をもたらすことができ、それによって疾患を治療し、

前記核酸塩基は前記疾患の原因ではない

方法。

2. 必要とする対象において疾患を治療する方法であって、

(a) 細胞に、

ガイドポリヌクレオチドまたはそれをコードする核酸；

ポリヌクレオチドプログラム可能なDNA結合ドメインまたはそれをコードする核酸、および

デアミナーゼドメインまたはそれをコードする核酸、

を含む塩基エディターシステムを導入することと、

(b) 前記細胞を前記対象に投与することと

を含み、前記ガイドポリヌクレオチドが、前記塩基エディターシステムをターゲティングして前記細胞中のSERPINA1ポリヌクレオチド中の核酸塩基の脱アミノ化をもたらすことができ、それによって前記疾患を治療し、

前記核酸塩基は前記疾患の原因ではない、

方法。

3. 前記細胞が肝細胞またはその前駆細胞である、実施形態2の方法。

4. 前駆細胞を分化させて肝細胞を生成することをさらに含む、実施形態3の方法。

5. 前記細胞が対象にとって自家性である、実施形態2～4のいずれか一つの方法。

6. 前記細胞が対象にとって同種異系である、実施形態2～4のいずれか一つの方法。

7. 前記細胞が対象にとって異種性である、実施形態2～4のいずれか一つの方法。

8. 前記対象が哺乳動物である、先の実施形態のいずれか一つの方法。

9. SERPINA1ポリヌクレオチドを編集する方法であって、SERPINA 1ポリヌクレオチ

ドを、

ガイドポリヌクレオチド；

ポリヌクレオチドプログラム可能なDNA結合ドメイン、および
デアミナーゼドメイン

を含む塩基エディターシステムと接触させることを含み、

前記ガイドポリヌクレオチドが、前記塩基エディターシステムをターゲティングしてSERPINA1ポリヌクレオチド中の核酸塩基の脱アミノ化をもたらすことができ、

ここで、核酸塩基は前記疾患の原因ではない、
方法。

10. 疾患の治療のための改変された細胞を産生する方法であって、

ガイドポリヌクレオチドまたは1以上のガイドポリヌクレオチドをコードする核酸；

ポリヌクレオチドプログラム可能なDNA結合ドメインまたはそれをコードする核酸、及
び

デアミナーゼドメインまたはそれをコードする核酸

を含む塩基エディターシステムを細胞に導入することをふくみ、

前記ガイドポリヌクレオチドが、前記塩基エディターシステムをターゲティングして前
記細胞中のSERPINA1ポリヌクレオチド中の核酸塩基の脱アミノ化をもたらすことができ

、
前記核酸塩基は前記疾患の原因ではない、

方法。

11. 前記導入がインビボである、実施形態10の方法。

12. 前記導入がex vivoである、実施形態10の方法。

13. 前記細胞が前記疾患を有する対象から得られるものである、実施形態12の方法。

14. 前記細胞が哺乳動物細胞である、実施形態10～13のいずれか1つの方法。

15. 前記細胞が肝細胞またはその前駆細胞である、実施形態14の方法。

16. 前駆細胞を分化させて肝細胞を産生することをさらに含む、実施形態15の方法。

17. 前記ポリヌクレオチドプログラム可能なDNA結合ドメインがCas9ドメインである、
前記実施形態のいずれか1つの方法。

18. Cas9ドメインがヌクレアーゼ不活性Cas9ドメインである、実施形態17の方法。

19. Cas9ドメインがCas9ニックアーゼドメインである、実施形態18の方法。

20. Cas9ドメインがSpCas9ドメインを含む、実施形態17～19のいずれか1つの方法。

21. SpCas9ドメインがD10Aおよび/またはH840Aアミノ酸置換またはそれらに対応す
るアミノ酸置換を含む、実施形態20の方法。

22. SpCas9ドメインがNGG PAMに対する特異性を有する、実施形態20または21の方法
。

23. SpCas9ドメインが、NGA PAM、NGT PAM、またはNGC PAMに対する特異性を有
する、実施形態20～22のいずれか1つの方法。

24. SpCas9ドメインが、アミノ酸置換L 1111 R、D 1135 V、G 1218 R、E 1219 F
、A 1322 R、R 1335 V、T 1337 R、ならびにL 1111、D 1135 L、S 1136 R、G 1
218 S、E 1219 V、D 1332 A、R 1335 Q、T 1337 I、T 1337 V、T 1337 F、およ
びT 1337 Mのうちの1つ以上、またはそれらに対応するアミノ酸置換を含む、実施形態2
0～23のいずれか1つの方法。

25. SpCas9ドメインが、アミノ酸置換L 1111 R、D 1135 V、G 1218 R、E 1219 F
、A 1322 R、R 1335 V、T 1337 R、ならびにL 1111、D 1135 L、S 1136 R、G 1
218 S、E 1219 V、D 1332 A、D 1332 S、D 1332 T、D 1332 V、D 1332 L、D
1332 K、D 1332 R、R 1335 Q、T 1337 I、T 1337 V、T 1337 F、T 1337 S、T
1337 N、T 1337 K、T 1337 R、T 1337 H、T 1337 Q、およびT 1337 Mのうちの
1つ以上、またはそれらに対応するアミノ酸置換を含む、実施形態20～23のいずれか1つ
の方法。

26. SpCas9ドメインが、アミノ酸置換D 1135 L、S 1136 R、G 1218 S、E 1219 V

10

20

30

40

50

、 A 1322 R、 R 1335 Q、 T 1337、 および A 1322 R、 ならびに L 1111、 D 1135 L、 S 1136 R、 G 1218 S、 E 1219 V、 D 1332 A、 D 1332 S、 D 1332 T、 D 1332 V、 D 1332 L、 D 1332 K、 D 1332 R、 R 1335 Q、 T 1337 I、 T 1337 V、 T 1337 F、 T 1337 S、 T 1337 N、 T 1337 K、 T 1337 R、 T 1337 H、 T 1337 Q、 および T 1337 M のうちの 1 つ以上、 またはそれらに対応するアミノ酸置換を含む、 実施形態 20 ~ 23 のいずれか 1 つの方法。

27. SpCas9 ドメインがアミノ酸置換 D 1135 M、 S 1136 Q、 G 1218 K、 E 1219 F、 A 1322 R、 D 1332 A、 R 1335 E、 および T 1337 R、 またはそれらに対応するアミノ酸置換を含む、 実施形態 20 ~ 23 のいずれか 1 つの方法。

28. SpCas9 ドメインが、 NG PAM、 NNG PAM、 GAA PAM、 GAT PAM、 または CAA PAM に対する特異性を有する、 実施形態 20 または 21 の方法。

10

29. SpCas9 ドメインがアミノ酸置換 E 480 K、 E 543 K、 および E 1219 V、 またはそれらに対応するアミノ酸置換を含む、 実施形態 28 の方法。

30. Cas 9 ドメインが SaCas9 ドメインを含む、 実施形態 17 ~ 19 のいずれか 1 つの方法。

31. SaCas9 ドメインは、 NNNRRT PAM に対する特異性を有する、 実施形態 30 の方法。

32. SaCas9 ドメインは、 NNGRRT PAM に対する特異性を有する、 実施形態 31 の方法。

33. SaCas9 ドメインがアミノ酸置換 N 579 A またはそれに対応するアミノ酸置換を含む、 実施形態 30 ~ 32 のいずれか 1 つの方法。

34. SaCas9 ドメインがアミノ酸置換 E 782 K、 N 968 K、 および R 1015 H、 またはそれらに対応するアミノ酸置換を含む、 実施形態 30 ~ 33 のいずれか 1 つの方法。

20

35. Cas9 ドメインが St1Cas9 ドメインを含む、 実施形態 17 ~ 19 のいずれか 1 つの方法。

36. St1Cas9 ドメインが NNACCA PAM に対する特異性を有する、 実施形態 35 の方法。

37. 前記デアミナーゼドメインがシチジンデアミナーゼドメインを含む、 前記実施形態のいずれか 1 つの方法。

38. シチジンデアミナーゼドメインが APOBEC ドメインを含む、 実施形態 31 の方法。

39. APOBEC ドメインが APOBEC1 ドメインを含む、 実施形態 32 の方法。

40. デアミナーゼドメインがアデノシンデアミナーゼドメインを含む、 実施形態 1 ~ 36 のいずれか 1 つの方法。

41. アデノシンデアミナーゼドメインが、 天然には存在しない改変アデノシンデアミナーゼドメインである、 実施形態 40 の方法。

30

42. アデノシンデアミナーゼドメインが TadA ドメインを含む、 実施形態 41 の方法。

43. TadA ドメインは、 TadA 7.10 のアミノ酸配列を含む、 実施形態 42 の方法。

44. 前記塩基エディターシステムが少なくとも 1 つの UGI ドメインをさらに含む、 前記実施形態のいずれか 1 つの方法。

45. 前記塩基エディターシステムが少なくとも二つの UGI ドメインを含む、 実施形態 44 の方法。

46. 前記塩基エディターシステムがジンクフィンガードメインをさらに含む、 前記実施形態のいずれか 1 つの方法。

47. ジンクフィンガードメインが、 認識ヘリックス配列 RNEHLEV、 QSTTLKR、 および RTEHLAR、 または認識ヘリックス配列 RGEHLRQ、 QSGTLKR、 および RNDKLVP を含む、 実施形態 46 の方法。

40

48. ジンクフィンガードメインが zf1ra または zf1rb である、 実施形態 46 または 47 の方法。

49. 前記塩基エディターシステムが核局在化シグナル (NLS) をさらに含む、 前記実施形態のいずれか 1 つの方法。

50. 前記塩基エディターシステムが 1 つ以上のリンカーをさらに含む、 前記実施形態のいずれか 1 つの方法。

51. ポリヌクレオチドプログラム可能な DNA 結合ドメイン、 デアミナーゼドメイン、 UGI ドメイン、 NLS、 及び/又はジンクフィンガードメインのうち二つ以上がリンカーを介して連結されている、 実施形態 50 の方法。

52. リンカーがペプチドリンカーであり、 それによって塩基編集融合タンパク質を形成し

50

ている、実施形態50の方法。

53. 前記ペプチドリンカーが、SGGSSGSETPGTSESATPESSGGS, SGGSSGGSSGSETPGTSESATPESSGGSSGGS, GSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSSGGS, SGGSSGGSSGSETPGTSESATPES, SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGSSGGS, SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGSSGSETPGTSESATPESSGGSSGGS, PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSSGGS, (SGGS)_n, (GGGS)_n, (GGGG)_n, (G)_n, (EAAAK)_n, (GGS)_n, GSETPGTSESATPES, および (XP)_nからなる群から選択されるアミノ酸配列含む、実施形態52の方法。

10

54. 塩基編集融合タンパク質がBE4のアミノ酸配列を含む、実施形態53の方法。

55. 前記塩基編集融合タンパク質は、

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
 AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGS
 LMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAKKKAQSSTDSGGSSGGSS
 GSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNN
 RVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHS
 RIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVF
 NAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSDKKYSIGLAIGTNSVGWA
 VITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRIC
 YLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKK
 LVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPIN
 ASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALLSLGLTPNFKSNFDLAEDAK
 LQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKR
 YDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
 GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTF
 RIPYYVGPLARGNSRFAMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNE
 KVLPHKSLLEYEFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK
 EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFE
 DREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG
 FANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
 VKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEELGKELGSQILKEHPVENTQLQ
 NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGS
 DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELKAGFIKRQLVETRQI
 TKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVVREINNYHHAHDA
 YLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEIQAQKATAKYFFYSNIMNFFK
 TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFSKE
 SILPKRNSDKLIARKKDWDPKKYGGFmqPTVAYSVLVVAKVEKGGKSKLKSVKELGITI
 MERSSFENPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASAKfLQKGNELALPS
 KYVNFYLYLASHYEKLGKSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL
 SAYNKHDKPIREQAENIIHLFTLTNLGAPrAFKYFDTTIaRKeYrSTKEVLDATLIHQSITEG
 LYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKY

20

30

40

のアミノ酸配列を含む、実施形態53の方法。

56. SERPINA1ポリヌクレオチドが、前記疾患の原因となる病原性一塩基多型 (SNP) を含む、先の実施形態のいずれか一つの方法。

57. 前記疾患がアルファ-1アンチトリプシン欠損症 (A1AD) である、実施形態56の方法。

58. SERPINA1ポリヌクレオチドは、病原性SNPから生じるアミノ酸変異を含むA1ATタンパク質をコードする、実施形態57の方法。

59. 前記アミノ酸変異が342 Lまたは376 L変異またはそれらに対応する位置のものである、実施形態58の方法。

50

60. 前記核酸塩基の脱アミノ化が、A1ATタンパク質の位置342もしくは376またはそれらに対応する位置以外の位置でのアミノ酸置換をもたらす、実施形態58または59の方法。
61. 前記核酸塩基の脱アミノ化が、A1ATタンパク質におけるF 51 L、M 374 I、A 348 V、A 347 V、K 387 R、T 59 A、およびT 68 Aからなる群より選択されるアミノ酸置換またはそれらに対応する置換をもたらす、実施形態60の方法。
62. 前記核酸塩基の脱アミノ化が、A1ATタンパク質の位置374またはそれに対応する位置におけるアミノ酸置換をもたらす、実施形態60の方法。
63. A1ATタンパク質におけるアミノ酸置換がM374I置換またはそれに対応する置換である、実施形態62の方法。
64. 前記核酸塩基がSERPINA1ポリヌクレオチドの1455の位置またはそれに対応する位置にある、実施形態63の方法。 10
65. 前記ガイドポリヌクレオチドが2つの個別のポリヌクレオチドを含み、前記2つの個別のポリヌクレオチドが2つのDNA、2つのRNA、またはDNAとRNAである、前記実施形態のいずれか1つの方法。
66. 前記ガイドポリヌクレオチドがcrRNAおよびtracrRNAを含み、前記crRNAはSERPINA1ポリヌクレオチド中の標的配列に対して相補的な核酸配列を含む、前記実施形態のいずれか1つの方法。
67. 標的配列がSERPINA1ポリヌクレオチドの位置1455を含む、実施形態66の方法。
68. 標的配列が、GAAGAAGATATTGGTGCTGT, TCAATCATTAGAAGACAAA, ACTTTTCCCATGAAGAGGGG, CATCGCTACAGCCTTTGCAA, および GGGACCAAGGCTGACAC TCA 20
から選択される配列を含む、実施形態66の方法。
69. 前記塩基エディターシステムが単一ガイドRNA (sgRNA) を含む、実施形態66または67の方法。
70. sgRNAが、5' -CAAUCAUUAAGAAGACAAAGGGUUU-3' ,
5' -UCAUCAUUAAGAAGACAAAGGGUUU-3' ,
5' -UUCAUCAUUAAGAAGACAAAGGGUUU-3' ,
5' -GUUCAUCAUUAAGAAGACAAAGGGUUU-3' ,
5' -UGUUCAUCAUUAAGAAGACAAAGGGUUU-3' ,
5' -UUGUUCAUCAUUAAGAAGACAAAGGGUUU-3' , 30
5' -UUCAUCAUUAAGAAGACAAAG-3' ,
5' -UUCAUCAUUAAGAAGACAAAGG-3' ,
5' -UCAUCAUUAAGAAGACAAAGGG-3' , および
5' -AAUCAUUAAGAAGACAAAGGGU-3' からなる群より選択される配列を含む、実施形態68の方法。
71. 必要とする対象においてアルファ-1アンチトリプシン欠損症 (A1AD) を治療する方法であって、
単一ガイドRNA (sgRNA)、
BE 4のアミノ酸配列を含む融合タンパク質
を含む塩基エディターシステムを前記対象に投与することを含み、 40
前記sgRNAは、前記塩基エディターシステムをターゲティングして前記対象の細胞中のSERPINA1ポリヌクレオチド中の位置1455またはそれに対応する位置におけるシチジンを脱アミノ化させ、それによってA1ADを治療し、
ここで、前記sgRNAは、
5' -CAAUCAUUAAGAAGACAAAGGGUUU-3'
5' -UCAUCAUUAAGAAGACAAAGGGUUU-3' ,
5' -UUCAUCAUUAAGAAGACAAAGGGUUU-3' ,
5' -GUUCAUCAUUAAGAAGACAAAGGGUUU-3' ,
5' -UGUUCAUCAUUAAGAAGACAAAGGGUUU-3' ,
5' -UUGUUCAUCAUUAAGAAGACAAAGGGUUU-3' , 50

5' -UUCAAUCAUUAAGAAGACAAAG-3' ,
 5' -UUCAAUCAUUAAGAAGACAAAGG-3' ,
 5' -UCAAUCAUUAAGAAGACAAAGGG-3' , および
 5' -AAUCAUUAAGAAGACAAAGGGU-3' からなる群から選択される配列を含む、

方法。

72. 必要とする対象においてアルファ - 1アンチトリプシン欠損症 (A1AD)を治療する方法であって、

(a) 細胞に、

単一ガイドRNA (sgRNA)、

BE4のアミノ酸配列を含む融合タンパク質を含む塩基エディターシステムを導入することと、

(b) 前記細胞を前記対象に投与することとを含み、

前記sgRNAは、前記塩基エディターシステムをターゲティングして前記細胞中のSERPINA1ポリヌクレオチドの位置1455またはそれに対応する位置におけるシチジンを脱アミノ化させ、それによってA1ADを治療し、

ここで、前記sgRNAは、

5' -CAAUCAUUAAGAAGACAAAGGGUUU-3' ,
 5' -UCAAUCAUUAAGAAGACAAAGGGUUU-3' ,
 5' -UUCAAUCAUUAAGAAGACAAAGGGUUU-3' ,
 5' -GUUCAAUCAUUAAGAAGACAAAGGGUUU-3' ,
 5' -UGUCAAUCAUUAAGAAGACAAAGGGUUU-3' ,
 5' -UUGUCAAUCAUUAAGAAGACAAAGGGUU-3' ,
 5' -UUCAAUCAUUAAGAAGACAAAG-3' ,
 5' -UUCAAUCAUUAAGAAGACAAAGG-3' ,
 5' -UCAAUCAUUAAGAAGACAAAGGG-3' , および
 5' -AAUCAUUAAGAAGACAAAGGGU-3' からなる群から選択される配列を含み、

前記細胞は前記対象から得られる肝細胞である、

方法。

73. 塩基エディターシステムを含む改変された細胞であって、

前記塩基エディターシステムが、

ガイドポリヌクレオチドまたはそれをコードする核酸；

ポリヌクレオチドプログラム可能なDNA結合ドメインまたはそれをコードする核酸、および

デアミナーゼドメインまたはそれをコードする核酸

を含み、

ここで、前記ガイドポリヌクレオチドは、前記塩基エディターシステムをターゲティングして前記細胞中のSERPINA1ポリヌクレオチドにおける核酸塩基の脱アミノ化をもたらすことができ、前記核酸塩基は疾患の原因ではない、

細胞。

74. 導入がインビボである、実施形態73の改変された細胞。

75. 導入がex vivoである、実施形態73の改変された細胞。

76. 前記細胞は前記疾患を有する対象から得られるものである、実施形態75の改変された細胞。

77. 前記細胞が哺乳動物細胞である、実施形態73~76のいずれか1つの改変された細胞。

78. 前記細胞が肝細胞またはその前駆細胞である、実施形態77の改変された細胞。

79. 前駆細胞を分化させて肝細胞を産生することをさらに含む、実施形態78の改変された細胞。

10

20

30

40

50

80. 前記ポリヌクレオチドプログラム可能なDNA結合ドメインがCas9ドメインである、実施形態73~79のいずれか1つの改変された細胞。
81. Cas9ドメインがヌクレアーゼ不活性Cas9ドメインである、実施形態80の改変された細胞。
82. Cas9ドメインがCas9ニッカーゼドメインである、実施形態80の改変された細胞。
83. Cas9ドメインがSpCas9ドメインを含む、実施形態80~82のいずれか1つの改変された細胞。
84. SpCas9ドメインがD10Aおよび/またはH840Aアミノ酸置換またはそれらに対応するアミノ酸置換を含む、実施形態83の改変された細胞。
85. SpCas9ドメインがNGG PAMに対する特異性を有する、実施形態83または84の改変された細胞。 10
86. SpCas9ドメインがNGA PAM、NGT PAM、またはNGC PAMに対する特異性を有する、実施形態83~85のいずれか1つの改変された細胞。
87. SpCas9ドメインが、アミノ酸置換L 1111 R、D 1135 V、G 1218 R、E 1219 F、A 1322 R、R 1335 V、T 1337 R、ならびにL 1111、D 1135 L、S 1136 R、G 1218 S、E 1219 V、D 1332 A、R 1335 Q、T 1337 I、T 1337 V、T 1337 F、およびT 1337 Mのうち1つ以上、またはそれらに対応するアミノ酸置換を含む、実施形態83~86のいずれか1つの改変された細胞。
88. SpCas9ドメインが、アミノ酸置換L 1111 R、D 1135 V、G 1218 R、E 1219 F、A 1322 R、R 1335 V、T 1337 R、ならびにL 1111、D 1135 L、S 1136 R、G 1218 S、E 1219 V、D 1332 A、D 1332 S、D 1332 T、D 1332 V、D 1332 L、D 1332 K、D 1332 R、R 1335 Q、T 1337 I、T 1337 V、T 1337 F、T 1337 S、T 1337 N、T 1337 K、T 1337 R、T 1337 H、T 1337 Q、およびT 1337 Mのうち1つ以上、またはそれらに対応するアミノ酸置換を含む、実施形態83~86のいずれか1つの改変された細胞。 20
89. SpCas9ドメインが、アミノ酸置換D 1135 L、S 1136 R、G 1218 S、E 1219 V、A 1322 R、R 1335 Q、T 1337、およびA 1322 R、ならびにL 1111、D 1135 L、S 1136 R、G 1218 S、E 1219 V、D 1332 A、D 1332 S、D 1332 T、D 1332 V、D 1332 L、D 1332 K、D 1332 R、R 1335 Q、T 1337 I、T 1337 V、T 1337 F、T 1337 S、T 1337 N、T 1337 K、T 1337 R、T 1337 H、T 1337 Q、およびT 1337 Mのうち1つ以上、またはそれらに対応するアミノ酸置換を含む、実施形態83~86のいずれか1つの改変された細胞。 30
90. SpCas9ドメインが、アミノ酸置換D 1135 M、S 1136 Q、G 1218 K、E 1219 F、A 1322 R、D 1332 A、R 1335 E、およびT 1337 R、またはそれらに対応するアミノ酸置換を含む、実施形態83~86のいずれか1つの改変された細胞。
91. SpCas9ドメインがNG PAM、NNG PAM、GAA PAM、GAT PAM、またはCAA PAMに対する特異性を有する、実施形態83または84の改変された細胞。
92. SpCas9ドメインがアミノ酸置換E 480 K、E 543 K、およびE 1219 V、またはそれらに対応するアミノ酸置換を含む、実施形態91の改変された細胞。
93. Cas9ドメインがSaCas9ドメインを含む、実施形態80~82のいずれか1つの改変された細胞。 40
94. SaCas9ドメインがNNRRT PAMに対する特異性を有する、実施形態93の改変された細胞。
95. SaCas9ドメインがNNGRRT PAMに対する特異性を有する、実施形態94の改変された細胞。
96. SaCas9ドメインがアミノ酸置換N 579 Aまたはそれらに対応するアミノ酸置換を含む、実施形態93~95のいずれか1つの改変された細胞。
97. SaCas9ドメインがアミノ酸置換E 782 K、N 968 K、およびR 1015 H、またはそれらに対応するアミノ酸置換を含む、実施形態93~96のいずれか1つの改変された細胞。
98. Cas9ドメインがSt1Cas9ドメインを含む、実施形態80~82のいずれか1つの改変 50

P) を含む、実施形態71～118のいずれか一つの改変された細胞。

120. 前記疾患がアルファ-1アンチトリプシン欠損症 (A1AD) である、実施形態119の改変された細胞。

121. 前記SERPINA1ポリヌクレオチドが、病原性SNPから生じたアミノ酸変異を含むA1ATタンパク質をコードする、実施形態120の改変された細胞。

122. 前記アミノ酸変異が342 Lもしくは376 L変異またはそれに対応する位置である、実施形態121の改変された細胞。

123. 前記核酸塩基の脱アミノ化が、A1ATタンパク質の位置342もしくは376またはそれに対応する位置以外の位置でのアミノ酸置換をもたらす、実施形態121または122の改変された細胞。

10

124. 前記核酸塩基の脱アミノ化が、A1ATタンパク質におけるF 51 L、M 374 I、A 348 V、A 347 V、K 387 R、T 59 A、およびT 68 Aからなる群より選択されるアミノ酸置換、またはそれらに対応する置換をもたらす、実施形態123の改変された細胞。

125. 前記核酸塩基の脱アミノ化が、A1ATタンパク質の位置374またはそれに対応する位置におけるアミノ酸置換をもたらす、実施形態122の改変された細胞。

126. A1ATタンパク質におけるアミノ酸置換がM 374 I置換またはその対応する置換である、実施形態125の改変された細胞。

127. 前記核酸塩基がSERPINA1ポリヌクレオチドの位置1455またはそれに対応する位置にある、実施形態126の改変された細胞。

128. 前記ガイドポリヌクレオチドが2つの個別のポリヌクレオチドを含み、前記2つの個別のポリヌクレオチドが、2つのDNA、2つのRNA、またはDNAとRNAである、実施形態71～127のいずれか1つの改変された細胞。

20

129. 前記ガイドポリヌクレオチドがcrRNAおよびtracrRNAを含み、crRNAは、SERPINA1ポリヌクレオチド中の標的配列に対して相補的な核酸配列を含む、実施形態71～128のいずれか1つの改変された細胞。

130. 前記標的配列がSERPINA1ポリヌクレオチドの位置1455を含む、実施形態129の改変された細胞。

131. 前記標的配列が、GAAGAAGATATTGGTGCTGT, TCAATCATTAAAGAAGACAAA, ACTTTTCCCATGAAGAGGGG, CATCGCTACAGCCTTTGCAA, および GGGACCAAGGCTGACTCAから選択される配列を含む、実施形態130の改変された細胞。

30

132. 前記塩基エディターシステムが単一ガイドRNA (sgRNA) を含む、実施形態130または131の改変された細胞。

133. 前記sgRNAが、

5' -CAAUCAUUAAGAAGACAAAGGGUUU-3' ,

5' -UCAUCAUUAAGAAGACAAAGGGUUU-3' ,

5' -UUCAUCAUUAAGAAGACAAAGGGUUU-3' ,

5' -GUUCAUCAUUAAGAAGACAAAGGGUUU-3' ,

5' -UGUUCAUCAUUAAGAAGACAAAGGGUUU-3' ,

5' -UUGUUCAUCAUUAAGAAGACAAAGGGUUU-3' ,

5' -UCAUCAUUAAGAAGACAAAG-3' ,

5' -UUCAUCAUUAAGAAGACAAAGG-3' ,

5' -UCAUCAUUAAGAAGACAAAGGG-3' , および

5' -AAUCAUUAAGAAGACAAAGGGU-3'

40

からなる群より選択される配列を含む、実施形態132の改変された細胞。

134. 単一ガイドRNA (sgRNA)

BE4のアミノ酸配列を含む融合タンパク質

を含む塩基エディターシステムを含む、改変された細胞であって、

前記sgRNAは、前記塩基エディターシステムをターゲティングしてSERPINA1ポリヌクレオチド中の位置1455またはそれに対応する位置におけるシチジンを脱アミノ化させることができる。

50

135. ここで、前記sgRNAは、

5' -CAAUCAUUAAGAAGACAAAGGGUUU-3' ,
 5' -UCAUCAUUAAGAAGACAAAGGGUUU-3' ,
 5' -UUCAUCAUUAAGAAGACAAAGGGUUU-3' ,
 5' -GUUCAUCAUUAAGAAGACAAAGGGUUU-3' ,
 5' -UGUUCAUCAUUAAGAAGACAAAGGGUUU-3' ,
 5' -UUGUUCAUCAUUAAGAAGACAAAGGGUUU-3' ,
 5' -UUCAUCAUUAAGAAGACAAAG-3' ,
 5' -UUCAUCAUUAAGAAGACAAAGG-3' ,
 5' -UCAUCAUUAAGAAGACAAAGGG-3' , および
 5' -AAUCAUUAAGAAGACAAAGGGU-3'

からなる群から選択される配列を含み、

ここで、前記細胞は肝細胞である。

136. 塩基エディターシステムであって、

ガイドポリヌクレオチドまたはそれをコードする核酸;

ポリヌクレオチドプログラム可能なDNA結合ドメインまたはそれをコードする核酸、および

デアミナーゼドメインまたはそれをコードする核酸
 を含み、

ここで、前記ガイドポリヌクレオチドは、前記塩基エディターシステムをターゲティング
 してSERPINA1ポリヌクレオチド中の核酸塩基の脱アミノ化をもたらすことができ、

前記核酸塩基は疾患の原因ではない、

塩基エディターシステム。

137. Cas 9ドメインがヌクレアーゼ不活性Cas 9ドメインである、実施形態135の塩基
 エディターシステム。

138. Cas 9ドメインがCas9ニッカーゼドメインである、実施形態135の塩基エディター
 システム。

139. Cas 9ドメインがSpCas9ドメインを含む、実施形態135～137のいずれか1つの塩
 基エディターシステム。

140. SpCas9ドメインがD10Aおよび/またはH840Aアミノ酸置換またはそれらに対応す
 るアミノ酸置換を含む、実施形態138の塩基エディターシステム。

141. SpCas9ドメインがNGG PAMに対する特異性を有する、実施形態138または139の
 塩基エディターシステム。

142. SpCas9ドメインが、NGA PAM、NGT PAM、またはNGC PAMに対する特異性を
 有する、実施形態138～140のいずれか1つの塩基エディターシステム。

143. SpCas9ドメインが、アミノ酸置換L 1111 R、D 1135 V、G 1218 R、E 1219
 F、A 1322 R、R 1335 V、T 1337 R、ならびにL 1111、D 1135 L、S 1136 R、G
 1218 S、E 1219 V、D 1332 A、R 1335 Q、T 1337 I、T 1337 V、T 1337 F、お
 よびT 1337 Mのうちの1つ以上、またはそれらに対応するアミノ酸置換を含む、実施形
 態138～141のいずれか1つの塩基エディターシステム。

144. SpCas9ドメインが、アミノ酸置換L 1111 R、D 1135 V、G 1218 R、E 1219
 F、A 1322 R、R 1335 V、T 1337 R、ならびにL 1111、D 1135 L、S 1136 R、G
 1218 S、E 1219 V、D 1332 A、D 1332 S、D 1332 T、D 1332 V、D 1332 L、D
 1332 K、D 1332 R、R 1335 Q、T 1337 I、T 1337 V、T 1337 F、T 1337 S、T
 1337 N、T 1337 K、T 1337 R、T 1337 H、T 1337 Q、およびT 1337 Mのうちの
 1以上、またはそれらに対応するアミノ酸置換を含む、実施形態138～141のいずれか1つ
 の塩基エディターシステム。

145. SpCas9ドメインが、アミノ酸置換D 1135 L、S 1136 R、G 1218 S、E 1219
 V、A 1322 R、R 1335 Q、T 1337、およびA 1322 R、ならびにL 1111、D 1135 L
 、S 1136 R、G 1218 S、E 1219 V、D 1332 A、D 1332 S、D 1332 T、D 1332 V

10

20

30

40

50

、 D 1332 L、 D 1332 K、 D 1332 R、 R 1335 Q、 T 1337 I、 T 1337 V、 T 1337 F、 T 1337 S、 T 1337 N、 T 1337 K、 T 1337 R、 T 1337 H、 T 1337 Q、 および T 1337 M のうちの 1 つ以上、 またはそれらに対応するアミノ酸置換を含む、 実施形態 138 ~ 141 のいずれか 1 つの塩基エディターシステム。

146. SpCas9 ドメインが、 アミノ酸置換 D 1135 M、 S 1136 Q、 G 1218 K、 E 1219 F、 A 1322 R、 D 1332 A、 R 1335 E、 および T 1337 R、 またはそれらに対応するアミノ酸置換を含む、 実施形態 138 ~ 141 のいずれか 1 つの塩基エディターシステム。

147. SpCas9 ドメインが、 NG PAM、 NNG PAM、 GAA PAM、 GAT PAM、 または CAA PAM に対する特異性を有する、 実施形態 138 または 139 の塩基エディターシステム。

148. SpCas9 ドメインが、 アミノ酸置換 E 480 K、 E 543 K、 および E 1219 V、 またはそれらに対応するアミノ酸置換を含む、 実施形態 146 の塩基エディターシステム。

10

149. Cas9 ドメインが SaCas9 ドメインを含む、 実施形態 135 ~ 137 のいずれか 1 つの塩基エディターシステム。

150. SaCas9 ドメインは、 NNNRRT PAM に対する特異性を有する、 実施形態 148 の塩基エディターシステム。

151. SaCas9 ドメインが NNGRRT PAM に対する特異性を有する、 実施形態 149 の塩基エディターシステム。

152. SaCas9 ドメインがアミノ酸置換 N579A またはそれに対応するアミノ酸置換を含む、 実施形態 135 ~ 137 のいずれか 1 つの塩基エディターシステム。

153. SaCas9 ドメインがアミノ酸置換 E 782 K、 N 968 K、 および R 1015 H、 またはそれらに対応するアミノ酸置換を含む、 実施形態 148 ~ 151 のいずれか 1 つの塩基エディターシステム。

20

154. Cas9 ドメインが St1Cas9 ドメインを含む、 実施形態 135 ~ 137 のいずれか 1 つの塩基エディターシステム。

155. St1Cas9 ドメインが NNACCA PAM に対する特異性を有する、 実施形態 153 の塩基エディターシステム。

156. デアミナーゼドメインがシチジンデアミナーゼドメインを含む、 実施形態 134 ~ 154 のいずれか 1 つの塩基エディターシステム。

157. シチジンデアミナーゼドメインが APOBEC ドメインを含む、 実施形態 155 の塩基エディターシステム。

30

158. APOBEC ドメインが APOBEC1 ドメインを含む、 実施形態 156 の塩基エディターシステム。

159. デアミナーゼドメインがアデノシンデアミナーゼドメインを含む、 実施形態 134 ~ 157 のいずれか 1 つの塩基エディターシステム。

160. アデノシンデアミナーゼドメインが、 天然には存在しない改変されたアデノシンデアミナーゼドメインである、 実施形態 158 の塩基エディターシステム。

161. アデノシンデアミナーゼドメインが TadA ドメインを含む、 実施形態 159 の塩基エディターシステム。

162. TadA ドメインが TadA 7.10 のアミノ酸配列を含む、 実施形態 160 の塩基エディターシステム。

40

163. 前記塩基エディターシステムが少なくとも 1 つの UGI ドメインをさらに含む、 実施形態 134 ~ 161 のいずれか 1 つの塩基エディターシステム。

164. 前記塩基エディターシステムは、 少なくとも 2 つの UGI ドメインを含む、 実施形態 162 の塩基エディターシステム。

165. 前記塩基エディターシステムは、 ジンクフィンガードメインをさらに含む、 実施形態 134 ~ 163 のいずれか 1 つの塩基エディターシステム。

166. ジンクフィンガードメインが、 認識ヘリックス配列 RNEHLEV、 QSTTLKR、 および RTEHLAR、 または認識ヘリックス配列 RGEHLRQ、 QSGTLKR、 および RNDKLVLP を含む、 実施形態 164 の塩基エディターシステム。

167. ジンクフィンガードメインが zf1ra または zf1rb である、 実施形態 165 の塩基エディ

50

ターシステム。

168. 前記塩基エディターシステムは、核局在化シグナル (NLS) をさらに含む、実施形態134 ~ 166のいずれか一つの塩基エディターシステム。

169. 前記塩基エディターシステムは、1つ以上のリンカーをさらに含む、実施形態134 ~ 167のいずれか一つの塩基エディターシステム。

170. 前記ポリヌクレオチドプログラム可能なDNA結合ドメイン、デアミナーゼドメイン、UGIドメイン、NLS、及び/又はジンクフィンガードメインのうちの一つ以上がリンカーを介して接続されている、実施形態168の塩基エディターシステム。

171. リンカーがペプチドリンカーであり、それによって塩基編集融合タンパク質を形成している、実施形態169の塩基エディターシステム。

172. 前記170ペプチドリンカーが、SGGSSGSETPGTSESATPESSGGS, SGGSSGGSSGS ETPGTSESATPESSGGSSGGS, GSGSGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEG SAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSG GS, SGGSSGGSSGSETPGTSESATPES, SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGG SSGGS, SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGSSGSETPGTSESATPESS GGSSGGS, PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTE PSEGSAP GTSTEPSEGSAPGTSESATPESGPGSEPAT, (SGGS)n, (GGGS)n, (GGGS)n, (G)n, (EAAAK)n, (GGG)n, SGSETPGTSESATPES, および (XP)nからなる群から選択されるアミノ酸配列を含むペプチドリンカーである、実施形態170の塩基エディターシステム。

173. 塩基編集融合タンパク質がBE4のアミノ酸配列を含む、実施形態170の塩基エディターシステム。

174. 前記塩基編集融合タンパク質は、

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH AEIMALRQGGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGS LMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAKKAQSSTDSSGSSGGSS GSETPGTSESATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNN RVIGEGWNRAIGLHDPTAHAEIMALRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHS RIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVF NAQKKAQSSTDSSGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWA VITDEYKVPSPKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRIC YLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKK LVDSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSVDKLFQILVQTYNQLFEENPIN ASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAK LQLSKDQYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKR YDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPPFLKDNREKIEKILTF RIPYYVGPLARGNSRFAMTRKSEETITPWNFEEVVDKGASQSFIERMTNFDKNLPNE KVLPHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLEF DREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKILQTVKVVDEL VKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEELGKELGSQLKEHPVENTQLQ NEKLYLYYLQNGRDMYVDQELDINRLSDYDQVHIVPQSFLKDDSIDNKVLRSDKNRGS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELKAGFIKRQLVETRQI TKHVAQILD SRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVVREINNYHHAHDA YLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFK TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKE SILPKRNSDKLIARKKDWDPKKGmPTVAYSVLVVAKVEKKGSKKLKSVKELLGITI MERSSFENPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASAKfLQKGNELALPS

10

20

30

40

50

KYVNFLYLASHYEKLKGSPEDENEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL
SAYNKHRDKPIREQAENIIHLFTLTNLGAPrAFKYFDTTIaRKeYrSTKEVL DATLIHQSI
LYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

の amino 酸配列を含む、実施形態170の塩基エディターシステム。

175. SERPINA1ポリヌクレオチドが、疾患の原因となる病原性一塩基多型 (SNP) を含む、実施形態134～173のいずれか一つの塩基エディターシステム。

176. 疾患がアルファ-1アンチトリプシン欠損症 (A1AD) である、実施形態174の塩基エディターシステム。

177. SERPINA1ポリヌクレオチドが、病原性SNPから生じる amino 酸変異を含むA1ATタンパク質をコードする、実施形態175の塩基エディターシステム。

178. 前記 amino 酸変異が、342Lまたは376L変異またはそれに対応する位置のものである、実施形態176の塩基エディターシステム。

179. 前記核酸塩基の脱 amino 化が、A1ATタンパク質の位置342もしくは376またはそれに対応する位置以外の位置での amino 酸置換をもたらす、実施形態176または177の塩基エディターシステム。

180. 前記核酸塩基の脱 amino 化が、A1ATタンパク質におけるF 51 L、M 374 I、A 348 V、A 347 V、K 387 R、T 59 A、およびT 68 Aからなる群より選択される amino 酸置換、またはそれらに対応する置換をもたらす、実施形態178の塩基エディターシステム。

181. 前記核酸塩基の脱 amino 化が、A1ATタンパク質の位置374またはそれに対応する位置における amino 酸置換をもたらす、実施形態178の塩基エディターシステム。

182. A1ATタンパク質における amino 酸置換がM374I置換またはそれに対応する置換である、実施形態180の塩基エディターシステム。

183. 前記核酸塩基がSERPINA1ポリヌクレオチドの位置1455またはそれに対応する位置にある、実施形態126の塩基エディターシステム。

184. 前記ガイドポリヌクレオチドが2つの個別のポリヌクレオチドを含み、前記2つの個別のポリヌクレオチドが、2つのDNA、2つのRNA、またはDNAとRNAである、実施形態134～182のいずれか一つの塩基エディターシステム。

185. ガイドポリヌクレオチドがcrRNAおよびtracrRNAを含み、crRNAはSERPINA1ポリヌクレオチド中の標的配列に対して相補的な核酸配列を含む、実施形態183のいずれか一つの塩基エディターシステム。

186. 標的配列がSERPINA1ポリヌクレオチドの位置1455を含む、実施形態184の塩基エディターシステム。

187. 標的配列が、GAAGAAGATATTGGTGCTGT, TCAATCATTAAGAAGACAAA, ACTTTTCCCATGAAGAGGGG, CATCGCTACAGCCTTTGCAA, および GGGACCAAGGCTGACACTCAから選択される配列を含む、実施形態184の塩基エディターシステム。

188. 該塩基エディターシステムが単一ガイドRNA (sgRNA) を含む、実施形態185または186の塩基エディターシステム。

189. 前記sgRNAが、

5' -CAAUCAUUAAGAAGACAAAGGGUUU-3' ,
5' -UCAUCAUUAAGAAGACAAAGGGUUU-3' ,
5' -UUCAUCAUUAAGAAGACAAAGGGUUU-3' ,
5' -GUUCAUCAUUAAGAAGACAAAGGGUUU-3' ,
5' -UGUUCAUCAUUAAGAAGACAAAGGGUUU-3' ,
5' -UUGUUCAUCAUUAAGAAGACAAAGGGUU-3' ,
5' -UUCAUCAUUAAGAAGACAAAG-3' ,
5' -UUCAUCAUUAAGAAGACAAAGG-3' ,
5' -UCAUCAUUAAGAAGACAAAGGG-3' , および
5' -AAUCAUUAAGAAGACAAAGGGU-3'

からなる群より選択される配列を含む、実施形態187の塩基エディターシステム。

10

20

30

40

50

190. 単一ガイドRNA (sgRNA)、

BE4のアミノ酸配列を含む融合タンパク質
を含む塩基エディターシステムであって、

前記sgRNAは、前記塩基エディターシステムをターゲティングしてSERPINA1ポリヌクレオチド中の位置1455またはそれに対応する位置のシチジンを脱アミノ化させることができ、

191. ここで、前記sgRNAは、

5' -CAAUCAUUAAGAAGACAAAGGGUUU-3' ,
5' -UCAUCAUUAAGAAGACAAAGGGUUU-3' ,
5' -UUCAUCAUUAAGAAGACAAAGGGUUU-3' ,
5' -GUUCAUCAUUAAGAAGACAAAGGGUUU-3' ,
5' -UGUUCAUCAUUAAGAAGACAAAGGGUUU-3' ,
5' -UUGUUCAUCAUUAAGAAGACAAAGGGUU-3' ,
5' -UUCAUCAUUAAGAAGACAAAG-3' ,
5' -UUCAUCAUUAAGAAGACAAAGG-3' ,
5' -UCAUCAUUAAGAAGACAAAGGG-3' , および
5' -AAUCAUUAAGAAGACAAAGGGU-3'

10

からなる群から選択される配列を含む。

192. 必要とする対象において疾患を治療する方法であって、

ガイドポリヌクレオチドまたはそれをコードする核酸;

20

ポリヌクレオチドプログラム可能なDNA結合ドメインまたはそれをコードする核酸、および

デアミナーゼドメインまたはそれをコードする核酸

を含む塩基エディターシステムを前記対象に投与することを含み、

ここで、前記ガイドポリヌクレオチドは、前記塩基エディターシステムをターゲティングして前記対象の細胞の標的ポリヌクレオチドにおける核酸塩基の脱アミノ化をもたらすことができ、前記核酸塩基は前記疾患の原因ではない、

方法。

193. 必要とする対象において疾患を治療する方法であって、

(a) 細胞に、

30

ガイドポリヌクレオチドまたはそれをコードする核酸;

ポリヌクレオチドプログラム可能なDNA結合ドメインまたはそれをコードする核酸、および

デアミナーゼドメインまたはそれをコードする核酸

を含む塩基エディターシステムを導入することと、

(b) 前記細胞を前記対象に投与することと

を含み、

前記ガイドポリヌクレオチドは、前記塩基エディターシステムをターゲティングして前記対象の細胞の標的ポリヌクレオチド中の核酸塩基の脱アミノ化をもたらすことができ、それによって疾患を治療し、前記核酸塩基は前記疾患の原因ではない、

40

方法。

194. 疾患の治療のための改変された細胞を産生する方法であって、

ガイドポリヌクレオチドまたはそれをコードする核酸;

ポリヌクレオチドプログラム可能なDNA結合ドメインまたはそれをコードする核酸、および

デアミナーゼドメインまたはそれをコードする核酸

を含む塩基エディターシステムを細胞に導入することを含み、

前記ガイドポリヌクレオチドは前記塩基エディターシステムをターゲティングして前記細胞の標的ポリヌクレオチド中の核酸塩基の脱アミノ化をもたらすことができ、

前記核酸塩基は前記疾患の原因ではない、方法。

50

195. 前記導入することがインピボまたはエクスピボである、実施形態192の方法。
196. 前記細胞が肝細胞またはその前駆細胞である、実施形態192または193の方法。
197. 標的ポリヌクレオチドは、タンパク質をコードする遺伝子を含み、前記遺伝子は前記疾患の原因となる病原性一塩基多型 (SNP) を含む、実施形態190～194のいずれか一実施形態の方法。
198. 前記疾患が、鎌状赤血球症、ベータ-サラセミア、アルファ-1アンチトリプシン欠損症 (A1AD)、ATTRアミロイドーシス、または嚢胞性線維症である、実施形態95の方法。
199. 前記タンパク質が、病原性SNPに起因するアミノ酸変異を含む、実施形態195または196の方法。
200. 前記核酸塩基の脱アミノ化が、タンパク質の発現、活性、または安定性を改変する、実施形態197の方法。 10
201. 前記核酸塩基の脱アミノ化が、タンパク質の発現、活性、または安定性を増加させる、実施形態198の方法。
202. 前記遺伝子がCFTRであり、タンパク質がCFTRタンパク質である、実施形態195～199のいずれか1つの方法。
203. 前記脱アミノ化が、CFTRタンパク質におけるR 555 K、F 409 L、F 433 L、H 667 R、R 1070 W、R 29 K、R 553 Q、I 539 T、G 550 E、F 429 S、およびQ 637 Rからなる群より選択されるアミノ酸置換またはそれらに対応する置換をもたらす、実施形態200の方法。
204. 前記遺伝子がTTRであり、タンパク質がTTRタンパク質である、実施形態195～199のいずれか1つの方法。 20
205. 前記脱アミノ化が、TTRタンパク質におけるA 108 V、R 104 H、およびT 119 Mからなる群より選択されるアミノ酸置換またはそれらに対応する置換をもたらす、実施形態202の方法。
206. 前記遺伝子がHBBであり、タンパク質がヘモグロビンのベータサブユニット (HbB) である、実施形態195～199のいずれか1つの方法。
207. 前記脱アミノ化が、HbBのA 70 T、A 70 V、L 88 P、F 85 L、F 85 P、E 22 G、G 16 D、およびG 16 Nからなる群より選択されるアミノ酸置換またはそれらに対応する置換をもたらす、実施形態204の方法。
208. 前記ポリヌクレオチドプログラム可能なDNA結合ドメインがCas9ドメインである、実施形態189～205のいずれか1つの方法。 30
209. Cas9ドメインがヌクラーゼ不活性Cas9ドメインまたはCas9ニッカーゼドメインである、実施形態206の方法。
210. Cas9ドメインがSpCas9ドメインを含む、実施形態206または207の方法。
211. SpCas9ドメインがD10Aおよび/またはH840Aアミノ酸置換またはそれに対応するアミノ酸置換を含む、実施形態208の方法。
212. SpCas9ドメインがNGN PAMに対する特異性を有する、実施形態209の方法。
213. Cas9ドメインが、アミノ酸置換D 1135 M、S 1136 Q、G 1218 K、E 1219 F、A 1322 R、D 1332 A、R 1335 E、およびT 1337 R、またはそれらに対応するアミノ酸置換を含む、実施形態208～210のいずれか1つの方法。 40
214. Cas9ドメインがSaCas9ドメインを含む、実施形態206または207の方法。
215. SaCas9ドメインは、NNRRRT PAMに対する特異性を有する、実施形態212の方法。
216. SaCas9ドメインがアミノ酸置換N579Aまたはそれに対応するアミノ酸置換を含む、実施形態212または213の方法。
217. Cas9ドメインがアミノ酸置換E 782 K、N 968 K、およびR 1015 H、またはそれらに対応するアミノ酸置換を含む、実施形態212～214のいずれか1つの方法。
218. デアミナーゼドメインがシチジンデアミナーゼドメインを含む、実施形態189～215のいずれか1つの方法。
219. シチジンデアミナーゼドメインがAPOBEC1ドメインを含む、実施形態216の方法。 50

220. デアミナーゼドメインがアデノシンデアミナーゼドメインを含む、実施形態189～215のいずれか1つの方法。

221. アデノシンデアミナーゼドメインが、*TadA* 7.10を含む*TadA*ドメインを含む、実施形態218の方法。

222. 塩基エディターシステムが少なくとも1つのUGIドメインをさらに含む、実施形態189～219のいずれか1つの方法。

223. 塩基エディターシステムが少なくとも二つのUGIドメインを含む、実施形態220の方法。

224. 塩基エディターシステムが1つ以上のリンカーをさらに含む、実施形態189～221のいずれか1つの方法。

225. ポリヌクレオチドプログラム可能なDNA結合ドメインおよびデアミナーゼドメインがリンカーを介して連結される、実施形態222の方法。

226. UGIドメインおよびデアミナーゼドメインがリンカーを介して接続される、実施形態222または223の方法。

227. リンカーがペプチドリンカーであり、それによって塩基編集融合タンパク質を形成している、実施形態224の方法。

228. 塩基編集融合タンパク質がBE4のアミノ酸配列を含む、実施形態225の方法。

229. 前記塩基編集融合タンパク質は、

```

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGS
LMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAKKKAQSSTDSGGSSGGSS
GSETPGTSESATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNN
RVIGEGWNRAIGLHDPTAHAEIMALRQGGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHS
RIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVF
NAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWA
VITDEYKVPSPKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRIC
YLQEIFSNEMAKVDDSFHRLVESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKK
LVDSTDKADRLRIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPIN
ASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALLSLGLTPNFKSNFDLAEDAK
LQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKR
YDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTF
RIPYYVGPLARGNSRFAMTRKSEETITPWNFEEVVDKGGASAQSFIERMTNFDKNLPNE
KVLPHKSHLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK
EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLE
DREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG
FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIIEGKELGSQILKEHPVENTQLQ
NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLRSDKNRGKS
DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELKAGFIKRQLVETRQI
TKHVAQILD SRMNTKYDENDKLIREVKVITLKS KLVSDFRKFDFQFYKVREINNYHHAHDA
YLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFK
TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKE
SILPKRNSDKLIARKKDWDPKKYGGFmqPTVAYSVLVVAKVEKGKSKKLKSVKELLGITI
MERSSFENPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKfLQKGNELALPS
KYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL
SAYNKHRDKPIREQAENIIHLFTLTNLGAPrAFKYFDTTIaRKeYrSTKEVLDATLIHQ SITG
LYETRIDLSQLGGDEGADKRTADGSEFESP KKKRKY

```

のアミノ酸配列を含む、実施形態225の方法。

230. 脱アミノ化が10%未満のインデル形成をもたらす、実施形態159～197のいずれか

10

20

30

40

50

一つの方法。

231. ガイドポリヌクレオチドまたはそれをコードする核酸;

ポリヌクレオチドプログラム可能なDNA結合ドメインまたはそれをコードする核酸、および

デアミナーゼドメインまたはそれをコードする核酸

を含む、塩基エディターシステムであって、

前記ガイドポリヌクレオチドが、塩基エディターシステムをターゲティングして標的ポリヌクレオチド中の核酸塩基の脱アミノ化をもたらすことができ、

前記核酸塩基は疾患の原因ではなく、

前記標的ポリヌクレオチドは表3Aまたは表3Bの標的配列を含む、

塩基エディターシステム。

10

【実施例】

【0515】

以下の実施例は、説明目的のためだけに提供され、本明細書に提供される特許請求の範囲を限定することを意図するものではない。

【0516】

(実施例1)

塩基エディター中のPAMバリエーションの検証

新規CRISPR系およびPAMバリエーションにより、塩基エディターは標的SNPで精密修正を行うことができる。いくつかの新規PAMバリエーションを評価および検証した。PAM評価および塩基エディターの詳細は、例えば、国際出願PCT/2017/045381 (WO2018/027078) およびPCT/US2016/058344 (WO2017/070632) に記載されており、それぞれ、その全体が参照により本明細書に組み込まれる。また、Komor, A.C., et al., "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., et al., "Programmable base editing of A · T to G · C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); および Komor, A.C., et al., "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017) (それぞれの全内容が参照により本明細書に組み込まれる) も参照されたい。

20

30

【0517】

(実施例2)

アルファ - 1 アンチトリプシン欠損症 (A1AD) を修正するための遺伝子編集

アルファ - 1 アンチトリプシン (A1AまたはA1AT) は、第14染色体上のSERPINA1遺伝子によってコードされるプロテアーゼ阻害因子である。この糖タンパク質は、主に肝臓で合成され、血中に分泌され、その血清濃度は、健康な成人では1.5 ~ 3.0g/L (20 ~ 52 μmol/L) である。A1ATは、肺間質および肺胞上皮被覆液 (alveolar lining fluid) に拡散し、そこでそれは好中球エラスターゼを不活化することによって、プロテアーゼ媒介性損傷から肺組織を保護する。アルファ - 1 アンチトリプシン欠損症 (A1AD) は、常染色体共優性様式で遺伝する。

40

【0518】

SERPINA1遺伝子の100種を超える遺伝子バリエーションが記載されているが、全てが疾患と関連しているわけではない。これらのバリエーションのアルファベットによる命名は、ゲル電気泳動上でのその移動速度に基づいている。最も一般的なバリエーションは、M (中程度の移動度) アリルであり、2つの最も頻度が高い欠損アリルは、PiSおよびPiZ (後者が最も遅い移動速度を有する) である。測定可能な血清タンパク質を産生しないいくつかの変異が記載されている; これらのものは、「ヌル」アリルと呼ばれる。最も一般的な遺伝子型は、正常な血清レベルのアルファ - 1 アンチトリプシンを産生するMMである。重篤な欠損症を有する多くの人々は、Zアリルについてホモ接合性 (ZZ) である。Zタンパク質は、肝細胞の小胞体中でのその産生中に、ミスフォールディングし、重合する; これらの異常

50

なポリマーは、肝臓中で捕捉され、アルファ - 1 アンチトリプシンの血清レベルを大きく低下させる。アルファ - 1 アンチトリプシン欠損症を有する患者に見られる肝疾患は、肝細胞中での異常なアルファ - 1 アンチトリプシンタンパク質の蓄積ならびに自食作用、小胞体ストレス応答およびアポトーシスを含む、その後の細胞応答によって引き起こされる。図2は、アルファ - 1 アンチトリプシンの最も一般的な遺伝子型 (MM、MZ、SS、SZおよびZZ) と、それらに関連する対応血清レベルを示す。アルファ - 1 アンチトリプシンの循環レベルの低下は、肺における好中球エラスターゼ活性の増大をもたらす；プロテアーゼと抗プロテアーゼ活性とのこの不均衡は、この状態と関連する肺疾患をもたらす (図1)。

【0519】

アルファ - 1 アンチトリプシン欠損症 (A1AD) は、白人において最も一般的であり、それは肺および肝臓に最も頻繁に影響する。肺では、最も一般的な兆候は、肺底において最も顕著である早発型 (30代および40代の患者) 汎細葉性肺気腫である。しかしながら、びまん性または上葉性気腫が起こり得、気管支拡張症も起こり得る。最も頻繁に記載される症状としては、呼吸困難、喘鳴および咳が挙げられる。罹患した個体の肺機能試験は、COPDと一致する所見を示す；しかしながら、気管支拡張剤に対する応答性が観察され、喘息として誤診されることがある。

【0520】

ZZ遺伝子型によって引き起こされる肝疾患は、様々な形で現れる。罹患した幼児は、新生児期に胆汁鬱滞性黄疸を呈することがあり、無胆汁便 (薄い色または土色) および肝腫大を伴うこともある。血中の抱合型ビリルビン、トランスアミナーゼおよびガンマ-グルタミルトランスフェラーゼレベルが上昇する。年長児および成人における肝疾患は、トランスアミナーゼ上昇の偶発的所見または静脈瘤出血もしくは腹水を含む確立された肝硬変の兆候を呈し得る。アルファ - 1 アンチトリプシン欠損症はまた、患者に、肝細胞癌に罹りやすくさせる。ホモ接合性ZZ遺伝子型は肝疾患の発症にとって必要であるが、ヘテロ接合性Z変異は、C型肝炎感染および嚢胞性線維症肝疾患などにおいてより重篤な肝疾患のリスクをもたらすことによって、他の疾患についての遺伝学的修飾因子として作用し得る。

【0521】

A1ADの2つの最も一般的な臨床的バリエーションは、E264V (PiS) およびE342K (PiZ) アリルである。半数を超えるA1AD患者が、少なくとも1コピーのE342K変異を担持する。相同性誘導型修復 (HDR) によるヌクレアーゼゲノム編集は非効率的であり、豊富なインデルが、循環レベルを低下させ、肺症状を悪化させるであろう。AAVベクターを用いて肝臓細胞に形質導入することが関わる遺伝子療法は、さらなるミスフォールディングされたタンパク質のため肝臓病理を悪化させる。野生型A1ATと、E342K A1ATをノックダウンするsiRNAとの両方をコードするAAVは、両方の病理に対処するのに有望である。

【0522】

プラスミドトランスフェクションのために、HEK293細胞に最適化された高効率低毒性DNAトランスフェクション試薬であるMirus TransIT293を、U6プロモーターを有する250ngのgRNAプラスミドおよびCMVプロモーターを有する750ngの塩基エディタープラスミドと共に、3 μ l : 1 μ gの比で使用して、ヒト胚性腎細胞 (HEK293T) を一過的にトランスフェクトした。最適化されたBE4である塩基エディターは、以下の配列を有していた：

```
ATGAGCAGCGAGACAGGCCCTGTGGCTGTGGATCCTACACTGCGGAGAAGAATCGAGCC
CCACGAGTTCGAGGTGTTCTTCGACCCAGAGAGCTGCGGAAAGAGACATGCCTGCTGT
ACGAGATCAACTGGGGCGGCAGACACTCTATCTGGCGGCACACAAGCCAGAACACCAAC
AAGCACGTGGAAGTGAAC TTTATCGAGAAGTTTACGACCGAGCGGTA CTTCTGCCCCAA
CACCAGATGCAGCATCACCTGGTTTCTGAGCTGGTCCCCTTGCGGCGAGTGCAGCAGAG
CCATCACCGAGTTTCTGTCCAGATATCCCCACGTGACCCTGTTTCATCTATATCGCCCGG
CTGTACCACCGCCGATCCTAGAAATAGACAGGGACTGCGCGACCTGATCAGCAGCGG
AGTGACCATCCAGATCATGACCGAGCAAGAGAGCGGCTACTGCTGGCGGAAC TTCGTGA
```

10

20

30

40

50

ACTACAGCCCCAGCAACGAAGCCCCTGGCCTAGATATCCTCACCTGTGGGTCCGACTG
 TACGTGCTGGAAGTACTGCATCATCCTGGGCCTGCCTCCATGCCTGAACATCCTGAG
 AAGAAAGCAGCCTCAGCTGACCTTCTTCAACAATCGCCCTGCAGAGCTGCCACTACCAGA
 GACTGCCTCCACACATCCTGTGGGCCACCGGACTTAAGAGCGGAGGATCTAGCGGGCGG
 TCTAGCGGATCTGAGACACCTGGCACAAGCGAGTCTGCCACACCTGAGAGTAGCGGGCGG
 ATCTTCTGGCGGCTCCGACAAGAAGTACTCTATCGGACTGGCCATCGGCACCAACTCTG
 TTGGATGGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTG
 GGCAACACCGACCGGCACAGCATCAAGAAGAATCTGATCGGCGCCCTGCTGTTGACTC
 TGGCGAAACAGCCGAAGCCACCAGACTGAAGAGAACCGCCAGGCGGAGATACACCCGGC
 GGAAGAACCGGATCTGCTACCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGAC
 GACAGCTTCTTCCACAGACTGGAAGAGTCTTCCCTGGTGGAAAGAGGACAAGAAGCACGA
 GCGGCACCCCATCTTCCGCAACATCGTGGATGAGGTGGCCTACCACGAGAAGTACCCCA
 CCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAAGGCCGACCTGAGACTG
 ATCTACCTGGCTCTGGCCACATGATCAAGTTCCGGGGCCACTTTCTGATCGAGGGCGA
 TCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACA
 ACCAGCTGTTGAGGAAAACCCCATCAACGCCTCTGGCGTGGACGCCAAGGCTATCCTG
 TCTGCCAGACTGAGCAAGAGCAGAAGGCTGGAAAACCTGATCGCCAGCTGCCTGGCGA
 GAAGAAGAATGGCCTGTTCCGCAACCTGATTGCCCTGAGCCTGGGACTGACCCCTAACT
 TCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTAC
 GACGACGACCTGGACAATCTGCTGGCCAGATCGGCGATCAGTACGCCGACTTGTCTTCT
 GGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGATATCCTGAGAGTGAACACCG
 AGATCACAAAGGCCCTCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAG
 GATCTGACCCTGCTGAAGGCCCTCGTTAGACAGCAGCTGCCAGAGAAGTACAAAGAGAT
 TTTCTTCGATCAGTCCAAGAACGGCTACGCCGGCTACATTGATGGCGGAGCCAGCCAAG
 AGGAATTCTACAAGTTCATCAAGCCATCCTGGAAAAGATGGACGGCACCGAGGAACTG
 CTGGTCAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAATGGCTC
 TATCCCTCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGAGACAAGAGGACT
 TTTACCCATTCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCAGGATC
 CCCTACTACGTGGGACCACTGGCCAGAGGCAATAGCAGATTCGCCTGGATGACCAGAAA
 GAGCGAGGAAACCATCACACCCTGGAACCTTCGAGGAAGTGGTGGACAAGGGCGCCAGCG
 CTCAGTCTTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCTAACGAGAAGGTG
 CTGCCAAGCACTCCCTGCTGTATGAGTACTTCACCGTGTACAACGAGCTGACCAAAGT
 GAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTTCTGAGCGGCGAGCAGAAAAAGG
 CCATTGTGGATCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAG
 GACTACTTCAAGAAAATCGAGTGCTTCGACAGCGTGGAAATCAGCGGCGTGGAAAGATCG
 GTTCAATGCCAGCCTGGGCACATAACCACGACCTGCTGAAAATTATCAAGGACAAGGACT
 TCCTGGACAACGAAGAGAACGAGGACATTCTCGAGGACATCGTGCTGACCCTGACACTG
 TTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACATACGCCACCTGTTGACGGA
 CAAAGTGATGAAGCAACTGAAGCGGAGGCGGTACACAGGCTGGGGCAGACTGTCTCGGA
 AGCTGATCAACGGCATCCGGGATAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAAG
 TCCGACGGCTTCGCCAACAGAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTT
 TAAAGAGGACATCCAGAAAGCCAGGTGTCCGGCCAAGGCGATTCTCTGCACGAGCACA
 TTGCCAACCTGGCCGGATCTCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTG
 GTGGACGAGCTTGTGAAAGTGATGGGCAGACACAAGCCCGAGAACATCGTGATCGAAAT
 GGCCAGAGAGAACCAGACCACAGAAAGGGCCAGAAGAACAGCCGCGAGAGAATGAAG
 CGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGGA
 AAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAAATGGACGGGATA
 TGTACGTGGACCAAGAGCTGGACATCAACCGGCTGAGCGACTACGATGTGGACCATATC
 GTGCCCCAGAGCTTTCTGAAGGACGACTCCATCGATAACAAGGTCCTGACCAGAAGCGA
 CAAGAACCAGGGCAAGAGCGATAACGTGCCCTCCGAAGAGGTGGTCAAGAAGATGAAGA

10

20

30

40

50

ACTACTGGCGACAGCTGCTGAACGCCAAGCTGATTACCCAGCGGAAGTTCGATAACCTG
 ACCAAGGCCGAGAGAGGGCGCCTGAGCGAACTTGATAAGGCCGGCTTCATTAAGCGGCA
 GCTGGTGGAAACCCGGCAGATCACCAAACACGTGGCACAGATTCTGGACTCCCGGATGA
 AACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTCATCACCTGAAG
 TCTAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTCTACAAAGTGCGGGAAATCAA
 CAACTACCATCACGCCACGACGCCTACCTGAATGCCGTTGTTGGAACAGCCCTGATCA
 AGAAGTATCCCAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTG
 CGGAAGATGATCGCCAAGAGCGAACAAGAGATCGGCAAGGCTACCGCCAAGTACTTTTT
 CTACAGCAACATCATGAACTTTTTCAAGACAGAGATCACCTGGCCAACGGCGAGATCC
 GGAAAAGACCCCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGC
 AGAGATTTTGCCACAGTGCGGAAAGTGCTGAGCATGCCCAAGTGAATATCGTGAAGAA
 AACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCTAAGCGGAACAGCG
 ATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGATAGC
 CCTACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAGGGCAAGTCCAAAAA
 GCTCAAGAGCGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTTGAGA
 AGAACCCGATCGACTTTCTGGAAGCCAAGGGCTACAAGAAGTCAAGAAGGACCTCATC
 ATCAAGCTCCCCAAGTACAGCCTGTTTCGAGCTGGAAAATGGCCGGAAGCGGATGCTGGC
 CTCAGCAGGCCGAACTGCAGAAAGGCAATGAACTGGCCCTGCCTAGCAAATACGTCAACT
 TCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCAGCCCCGAGGACAATGAGCAA
 AAGCAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAG
 CGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAACCTGGATAAGGTGCTGTCTGCCT
 ATAACAAGCACCGGGACAAGCCTATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTT
 ACCCTGACCAACCTGGGAGCCCCTGCCGCCTTCAAGTACTTCGACACCACCATCGACCG
 GAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACACTGATCCACCAGTCTATCA
 CCGGCCTGTACGAAACCCGGATCGACCTGTCTCAGCTCGGCGGCGATTCTGGTGGTTCT
 GGCGGAAGTGGCGGATCCACCAATCTGAGCGACATCATCGAAAAAGAGACAGGCAAGCA
 GCTCGTGATCCAAGAATCCATCCTGATGCTGCCTGAAGAGGTTGAGGAAGTGATCGGCA
 ACAAGCCTGAGTCCGACATCCTGGTGCACACCGCCTACGATGAGAGCACCGATGAGAAC
 GTCATGCTGCTGACAAGCGACGCCCTGAGTACAAGCCTTGGGCTCTCGTGATTGAGGA
 CAGCAATGGGGAGAACAAGATCAAGATGCTGAGCGGAGGTAGCGGAGGCAGTGGCGGA
 AGCACAAACCTGTCTGATATCATTGAAAAAGAAACCGGGAAGCAACTGGTCAATCAAGA
 GTCCATTCTCATGCTCCCGGAAGAAGTCGAGGAAGTCATTGGAAACAAACCCGAGAGCG
 ATATTCTGGTCCACACAGCCTATGACGAGTCTACAGACGAAAACGTGATGCTCCTGACC
 TCTGACGCTCCCGAGTATAAGCCCTGGGCACTTGTTATCCAGGACTCTAACGGGGAAAA
 CAAAATCAAAATGTTGTCCGGCGGCAGCAAGCGGACAGCCGATGGATCTGAGTTCGAGA
 GCCCAAGAAGAAACGGAAGGTgGAGtaa

【0523】

mRNAトランスフェクションのためには、HEK293T細胞に、2つの20msパルスを用い
 1150VでNeon Systemを使用して3 μgの全RNAをエレクトロポレーションした。合成g
 RNA および mRNAのトランスフェクションのためには、最初と最後の3つの塩基にホス
 ホロチオエート結合および2OMe修飾を有する修飾gRNAを使用した。全てのNNGRRTお
 よびNNNRRT PAMについて、スペーサーとsaCas9スキャフォールドは、以下の配列を有
 する。

【0524】

GUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACUAAAACAAGGCAAAAUGCCGU
 GUUUAUCUCGUCAACUUGUUGGCGAGAUUUUUU

【0525】

プラスミドトランスフェクションについては4日後およびRNAエレクトロポレーション
 については2日後、ゲノムDNAを、0.05%SDS、25 μg/mlプロテイナーゼK、10mM Tris
 pH8.0の単純溶解緩衝液を用いて細胞から抽出した後、85 °Cで熱不活化した。ゲノム部位

10

20

30

40

50

をPCR増幅し、MiSeq上で配列決定した。それぞれの位置での塩基頻度について、およびインデルパーセントについて、以前に記載されたように結果を分析した。インデル算出の詳細は、国際出願PCT/2017/045381 (WO2018/027078) およびPCT/US2016/058344 (WO2017/070632) に記載されており、それぞれその全体が参照により本明細書に組み込まれる。また、Komor, A.C., et al., “Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage” Nature 533, 420-424 (2016); Gaudelli, N.M., et al., “Programmable base editing of A · T to G · C in genomic DNA without DNA cleavage” Nature 551, 464-471 (2017); および Komor, A.C., et al., “Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity” Science Advances 3:eaao4774 (2017) (それぞれの全内容が参照により本明細書に組み込まれる) も参照。

【0526】

図3は、SERPINA1遺伝子中の変異についての抑制性変異塩基編集の戦略を示す。BE4塩基エディターを用いたM374Iの導入は、肝臓毒性の緩和と肺へのA1ATの循環の増加との両方を同時にもたらし得る。図4に示すように、M374Iは、HEK293T細胞からのバリエーションPiZ A1ATタンパク質とバリエーションPiS A1ATタンパク質の分泌を増加させ、バリエーションE342K A1ATタンパク質およびE264V A1ATタンパク質の安定化を助けた。分泌されたA1ATの量は、PiM PiS PiZという臨床的パターンに従っていた。E376K変異からのオフターゲット効果はPiS または PiZのバリエーションA1ATタンパク質と組み合わせられた場合に有害であると見られた。分泌だけが求められる表現型ではない。編集された産物は野生型タンパク質ではないため、組み換え変異体A1ATを活性すなわち好中球エラスターゼの阻害についてアッセイした。

【0527】

分泌実験は、各A1ATバリエーションをコードする125ngのpCMVを用いて48ウェルプレートで一過的にトランスフェクトされたHEK293T細胞において行った。トランスフェクションは6つの重複実験で行い、トランスフェクションの24時間後に細胞培養上清を採取した。細胞上清中のA1ATの濃度を、A1ATに対する抗体を用いたELISAによってアッセイした。

【0528】

図5は、HEK293TにおけるM374Iの最適化された塩基編集を示す。コンストラクトの設計および送達パラメータを最適化した。望まれる結果：望まれない結果(M374I:E376Kまたはインデル)の比に対する影響はほとんど観察されなかった。

【0529】

図6は、TadA tRNAデアミナーゼから出発してDNAデオキシアデノシンデアミナーゼを進化させる戦略を提供する。

【0530】

塩基編集A1ATバリエーションのエラスターゼ活性のパーセントを図7に示す。補完的変異M374Iの存在が、A1ATタンパク質におけるE342K変異およびE264V変異のそれぞれの抑制的活性を緩和した。バイスタンダー編集が最小限であるM374Iの著しい塩基編集が、E342Kアリルを有するA1ATを含むiPSC由来肝細胞、および野生型(WT)ヒト肝細胞の両方において達成された(図8)。M374Iの塩基編集は、iPSC由来E342K肝細胞のA1AT分泌における顕著な(40%)増加と関連していた(図9)。BE4 RNAの量(用量)を増やすと編集が増えたが、A1AT分泌における対応する増加はもたらさなかった。理論に拘束されることは望まないが、トランスフェクションの際に高いRNA用量を使うと細胞毒性が起こる可能性がある。補完的変異M374Iの導入に際して、iPSC由来E342K肝細胞においてA1AT分泌の再現可能な増加が検出された。初代ヒト肝細胞(PHH)におけるパイロット実験評価では、A1AT分泌に対するネガティブな影響は示されなかった。

【0531】

[配列]

下記表 7 は、記述された実施形態において利用された野生型およびバリエーション (E342K) の SERPINA1 にコードされたアミノ酸配列ならびに野生型およびバリエーション (E342K) の SERPINA1 ポリヌクレオチドのオープンリーディングフレーム (ORF) 核酸配列の代表的リストを提供する。

表 7 . 例示的配列

【表 7】

	配列	
SERPINA1 アミノ酸	MPSSVSWGILLLAGLCCCLVPVSLAEDPQGDAQAQKTDTSHHDDQDHPTFNKIT PNLAEFASFSLYRQLAHQSNSTNIFFSPVSIATAFAMLSLGTKADTHDEILEGL NFNLTEIPEAQIHEGFQELLRTLNPDSQLQLTTGNGLFLSEGLKLVDFLE DVKKLYHSEAFVNFVGDTEEAKKQINDYVEKGTQGKIVDLVKELDRDVF ALVNYIFFKKGKWERPFVVDTEEEDFHVDQVTTVKVPMMLKRLGMFNIQHC KKLSSWVLLMKYLGNAIAIFFLPDEGKQLHLENELTHDIITKFLNEDRRSA SLHLPKLSITGTDLKSVLGLGITKVFNSGADLSGVTEEAPLKLKSKAVHKA VLTIDEKGTAAAGAMFLEAIPMSIPPEVKFNKPFVFLMIEQNTKSPLFMGKV VNPTQK	10
SERPINA1 ORF	ATGCCGTCTTCTGTCTCGTGGGGCATCCTCCTGCTGGCAGGCCTGTGCTG CCTGGTCCCTGTCTCCCTGGCTGAGGATCCCCAGGGAGATGCTGCCAG AAGACAGATACATCCCACCATGATCAGGATCACCAACCTTCAACAAG ATCACCCCAACCTGGCTGAGTTCGCCTCAGCCTATAACCGCCAGCTGG CACACCAGTCCAACAGCACCAATATCTTCTTCTCCCCAGTGAGCATCGC TACAGCCTTTGCAATGCTCTCCCTGGGGACCAAGGCTGACACTCACGAT GAAATCCTGGAGGGCCTGAATTTCAACCTCACGGAGATTCGGGAGGCTC AGATCCATGAAGGCTTCCAGGAACCTCCGTACCCTCAACCAGCCAGA CAGCCAGCTCCAGCTGACCACCGCAATGGCCTGTTCTCAGCGAGGGC CTGAAGCTAGTGGATAAGTTTTTGGAGGATGTTAAAAAGTTGTACCACT CAGAAGCCTTCACTGTCAACTTCGGGGACACCGAAGAGGCCAAGAAAC AGATCAACGATTACGTGGAGAAGGGTACTCAAGGGAAAATTGTGGATT TGGTCAAGGAGCTTGACAGAGACACAGTTTTTGTCTCTGGTGAATTACAT CTTCTTTAAAGGCAAATGGGAGAGACCCTTTGAAGTCAAGGACACCGA GGAAGAGGACTTCCACGTGGACCAGGTGACCACCGTGAAGGTGCCTAT GATGAAGCGTTTAGGCATGTTTAAACATCCAGCACTGTAAGAAGCTGTCC AGCTGGGTGCTGCTGATGAAATACCTGGGCAATGCCACCGCCATCTTCT TCCTGCCTGATGAGGGGAAACTACAGCACCTGGAAAATGAACTCACCC ACGATATCATCACCAAGTTCCTGGAAAATGAAGACAGAAGGTCTGCCA GCTTACATTTACCCAAACTGTCCATTACTGGAACCTATGATCTGAAGAG CGTCCTGGGTCAACTGGGCATCACTAAGGTCTTCAGCAATGGGGCTGAC CTCTCCGGGGTACAGAGGAGGCACCCCTGAAGCTCTCAAGGCCGTGC ATAAGGCTGTGCTGACCATCGACGAGAAAGGACTGAAGCTGCTGGGG CCATGTTTTTAGAGGCCATACCATGTCTATCCCCCGAGGTCAAGTTC AACAAACCCTTTGTCTTCTTAATGATTGAACAAAATACCAAGTCTCCCC TCTTCATGGGAAAAGTGGTGAATCCCACCCAAAA	20 30 40

SERPI NA1 E342K アミ ノ酸	MPSSVSWGILLLAGLCLVPVSLAEDPQGDAAQKTDTSHHDDQDHPTFNKIT PNLAEFASFSLYRQLAHQSNSTNIFFSPVSIATAFAMLSLGTKADTHDEILEGL NFNLTEIPEAQIHEGFQELLRRLNQPDSQLQLTTGNGLFLSEGLKLVDFKLE DVKKLYHSEAFVNFVFGDTEEAQKQINDYVEKGTQGGKIVDLVKELDRDRTVF ALVNYIFFKKGKWERPFVVDTEEEEDFHVDQVTTVKVPMMLKRLGMFNIQHC KKLSSWVLLMKYLGNAIAIFFLPDEGKLQHLENELTHDIITKFLNEDRRSA SLHLPKLSITGTDLKSVLGQLGITKVFSNGADLSGVTEEAAPLKLKSKAVHKA VLTIDKKGTEAAGAMFLEAIPMSIPPEVKFNKPFVFLMIEQNTKSPLFMGKV VNPTQK	10
SERPI NA1 E342K ORF	ATGCCGTCTTCTGTCTCGTGGGGCATCCTCCTGCTGGCAGGCCTGTGCTG CCTGGTCCCTGTCTCCCTGGCTGAGGATCCCCAGGGAGATGCTGCCAG AAGACAGATACATCCCACCATGATCAGGATCACCCAACCTTCAACAAG ATCACCCCAACCTGGCTGAGTTCGCCTTCAGCCTATACCGCCAGCTGG CACACCAGTCCAACAGCACCAATATCTTCTTCTCCCCAGTGAGCATCGC TACAGCCTTTGCAATGCTCTCCCTGGGGACCAAGGCTGACACTCACGAT GAAATCCTGGAGGGCCTGAATTTCAACCTCACGGAGATTCCGGAGGCTC AGATCCATGAAGGCTTCCAGGAACCTCCTCCGTACCCTCAACCAGCCAGA CAGCCAGCTCCAGCTGACCACCGCAATGGCCTGTTCCCTCAGCGAGGGC CTGAAGCTAGTGGATAAGTTTTTGGAGGATGTTAAAAAGTTGTACCACT CAGAAGCCTTCACTGTCAACTTCGGGGACACCGAAGAGGCCAAGAAAC AGATCAACGATTACGTGGAGAAGGGTACTCAAGGGAAAATTGTGGATT TGGTCAAGGAGCTTGACAGAGACACAGTTTTTGTCTCTGGTGAATTACAT CTTCTTTAAAGGCAAATGGGAGAGACCCTTTGAAGTCAAGGACACCGA GGAAGAGGACTTCCACGTGGACCAGGTGACCACCGTGAAGGTGCCTAT GATGAAGCGTTTAGGCATGTTTAAACATCCAGCACTGTAAGAAGCTGTCC AGCTGGGTGCTGCTGATGAAATACCTGGGCAATGCCACCGCCATCTTCT TCCTGCCTGATGAGGGGAAACTACAGCACCTGGAAAATGAACTCACCC ACGATATCATACCAAGTTCCTGGAAAATGAAGACAGAAGGTCTGCCA GCTTACATTTACCCAAACTGTCCATTACTGGAACCTATGATCTGAAGAG CGTCCTGGGTCAACTGGGCATCACTAAGGTCTTCAGCAATGGGGCTGAC CTCTCCGGGGTCACAGAGGAGGCACCCCTGAAGCTCTCCAAGGCCGTGC ATAAGGCTGTGCTGACCATCGACaAGAAAGGGACTGAAGCTGCTGGGG CCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCCGAGGTCAAGTTC AACAAACCCTTTGTCTTCTTAATGATTGAACAAAATACCAAGTCTCCCC TCTTCATGGGAAAAGTGGTGAATCCCACCCAAAAA	20
		30

【 0 5 3 2 】

[実施例 3 . 材料と方法]

本明細書に記載される実施例で提供された結果は、以下の材料および方法を用いて得られた。

40

[クローニング/トランスフェクション]

PCRはVeraSeq ULtra DNAポリメラーゼ (Enzymatics) またはQ5 Hot Start High-Fidelity DNAポリメラーゼ(New England Biolabs)を用いて行った。USERクローニング(New England Biolabs)を用いて塩基エディター (BE) プラスミドを構築した。デアミナーゼ遺伝子はgBlocks Gene Fragments (Integrated DNA Technologies)として合成された。用いたCas9遺伝子は下記に示す。Cas9遺伝子は以前に報告されたプラスミドから得られた。デアミナーゼと融合遺伝子をpCMV (哺乳類コドン最適化)またはpET28b (大腸菌コドン最適化)の骨格にクローニングした。sgRNA発現プラスミドは、部位特異的変異誘発を用いて構築した。

【 0 5 3 3 】

50

簡単に説明すると、上述のプライマーは、製造業者の指示に従ってT4ポリヌクレオチドキナーゼ(New England Biolabs)を用いて5'リン酸化した。次に、リン酸化プライマー、およびテンプレートとしての、A1AT sgRNA発現プラスミドをコードする核酸を含むプラスミドと共に、Q5 Hot Start High-Fidelity Polymerase (New England Biolabs)を用いて、製造業者の指示に従ってPCRを実施した。PCR産物をDpnI (20 U、New England Biolabs)と共に37℃で1時間インキュベートし、QIAprepスピンカラム (Qiagen) で精製し、製造業者の指示に従ってQuickLigase (New England Biolabs)を用いてライゲーションした。DNAベクター増幅はMach1コンピテント細胞(ThermoFisher Scientific)を用いて行った。

【0534】

gRNAについて、以下の足場配列を示す：GUUUUAGAGC UAGAAAUAGC AAGUAAAA AU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU。この足場は、本明細書の表に示されるPAM、例えばNGG、NGA、NGC、NGT PAMのために使用された；gRNAは、このスキャフォールド配列と、本明細書に提供されるか、または当業者の知識に基づいて決定されかつ当業者に理解されるような、疾患関連遺伝子(例えば表3A、3Bおよび4)のためのおよびスペーサー配列(標的配列)とを包含する。(例えば、Komor, A.C., et al., "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., et al., "Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); Komor, A.C., et al., "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaa04774 (2017), and Rees, H.A., et al., "Base editing: precision chemistry on the genome and transcriptome of living cells." Nat Rev Genet. 2018 Dec;19(12):770-788. doi: 10.1038/s41576-018-0059-1参照。)

【0535】

使用されたプライマーのDNA配列は以下の通りである：

BEAM53

ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNGCCGTGCATAAGGCTGTGCTG

BEAM54

TGGAGTTCAGACGTGTGCTCTTCCGATCTGGGTGGATTACCACTTTTCCCATG

BEAM1704

ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNN AGGTGTCCACGTGAGCCTTG

【0536】

[ssDNAに対するインビトロのデアミナーゼアッセイ]

すべてのssDNA基質の配列を下記に示す。Cy3標識基質はすべてIntegrated DNA Technologies (IDT) から入手した。デアミナーゼは、1 μgのプラスミドを使用して、製造業者の指示に従ってTNT T7 Quick Coupled Transcription/Translation Kit (Promega) を使用してin vitroで発現させた。タンパク質発現に続いて、5 μlの溶解物を、CutSmart緩衝液(New England Biolabs) (50 mM酢酸カリウム、29 mMトリス酢酸、10 mM酢酸マグネシウム、100 μg ml⁻¹ BSA、pH 7.9)中で35 μlのssDNA (1.8 μM)およびUSER酵素(1ユニット)と組み合わせ、37℃で2時間インキュベートした。切断されたU含有基質を、10% TBE 尿素ゲル (Bio Rad) 上で完全長未改変基質から分離した。

【0537】

[His6 - rAPOBEC1 - リンカー - dCas9融合体の発現および精製]

大腸菌BL21 STAR (DE 3) コンピテント細胞(ThermoFisher Scientific)を、pET28b-His6-rAPOBEC1-リンカー-dCas9をコードするプラスミドで形質転換した。得られた発現株を、カナマイシン100 μg ml⁻¹を含むLuria-Bertani (LB) プロス中で37℃で一晩増殖させた。細胞を同じ増殖培地に1:100希釈し、OD600= ~ 0.6になるまで37℃で増殖させた。培養物を4℃で2時間冷却し、イソプロピル-β-D-1-チオガラクトピラノシド (IP

10

20

30

40

50

TG) を0.5 mMで添加してタンパク質発現を誘導した。約16時間後、4,000 gで遠心分離して細胞を集め、溶解緩衝液(50 mMトリス(ヒドロキシメチル)アミノメタン(Tris) -HCl (pH 7.5)、1 M NaCl、20%グリセロール、10 mMトリス(2-カルボキシエチル)ホスフィン(TCEP、Soltec Ventures))に再懸濁した。細胞を超音波処理(20秒パルス・オン、20秒パルス・オフ、合計8分間、6 W出力)によって溶解し、25,000 gで15分間遠心分離した後に溶解物上清を単離した。溶解物をHis-Purニッケル-ニトリロ酢酸 (ニッケル-NTA) 樹脂(ThermoFisher Scientific)と共に4 で1時間インキュベートし、His標識融合タンパク質を捕捉した。樹脂をカラムに移し、溶解緩衝液40 mlで洗浄した。Hisタグ融合タンパク質を、285 mMイミダゾールを補充した溶解緩衝液中で溶出し、限外濾過(100 kDa分子量カットオフのAmicon-Millipore)によって全容量1 mlに濃縮した。タンパク質を、50 mMトリス(ヒドロキシメチル)-アミノメタン (Tris) -HCl (pH 7.0)、0.1 M NaCl、20%グリセロール、10 mM TCEPを含む低塩精製緩衝液中で20mlに希釈し、SPセファロースFast Flow樹脂(GE Life Sciences)上にロードした。樹脂をこの低塩緩衝液40 mlで洗浄し、50 mMトリス(ヒドロキシメチル)アミノメタン (Tris) -HCl (pH 7.0)、0.5 M NaCl、20%グリセロール、10 mM TCEPを含む活性緩衝液5 mlでタンパク質を溶出した。溶出したタンパク質をSDS-PAGEで定量した。

10

【0538】

[sgRNAのin vitro転写]

T7プロモーターに続いて20-bp sgRNA標的配列を含む直鎖DNA断片を、製造業者の指示に従ってTranscriptAid T7 High Yield Transcription Kit (ThermoFisher Scientific)を用いて、in vitroで転写した。製造業者の指示に従ってMEGAclear キット(ThermoFisher Scientific)を用いてsgRNA産物を精製し、UV吸収により定量した。

20

【0539】

[Cy3共役dsDNA基質の調製]

80 ntの無標識鎖を、PAGE精製オリゴヌクレオチドとしてIDT社にオーダーした。補足情報に列挙される25-nt Cy3標識プライマーは各80 nt基質の3'末端に対して相補的である。このプライマーは、HPLC精製オリゴヌクレオチドとしてIDTにオーダーした。Cy3標識dsDNA基質を生成するために、80 nt鎖(5 µlの100 µM溶液)を、Cy3標識プライマー(5 µlの100 µM溶液)とNEBuffer 2 (38.25 µlの50 mM NaCl、10 mM Tris-HCl、10 mM MgCl₂、1 mM DTT、pH 7.9溶液、New England Biolabs)中でdNTP (0.75 µlの100 mM溶液)と組み合わせ、95 で5分間加熱し、その後45 まで0.1 /秒の速度で徐々に冷却した。このアニーリング期間の後、Klenow exo-(5U、New England Biolabs)を加え、反応物を37 で1時間インキュベートした。溶液を緩衝液PB (250 µl、Qiagen)およびイソプロパノール(50 µl)で希釈し、QIAprepスピンカラム (Qiagen) で精製し、50 µlのTris緩衝液で溶出した。dsDNAに対するデアミナーゼアッセイ。精製した融合タンパク質(活性緩衝液中1.9 µMを20 µl)を1当量の適切なsgRNAと組み合わせ、周囲温度で5分間インキュベートした。Cy3標識dsDNA基質を125 nMの最終濃度で添加し、得られた溶液を37 で2時間インキュベートした。緩衝液PB (100 µl、Qiagen)およびイソプロパノール(25 µl)の添加により融合物からdsDNAを分離し、EconoSpinマイクロスピカラム(Epoch Life Science)で精製し、20 µlのCutSmart緩衝液(New England Biolabs)で溶出した。精製された編集dsDNAにUSER酵素(1 U、New England Biolabs)を添加し、37 で1時間インキュベートした。5 µlの反応溶液を15 µlのDMSOベースのローディング緩衝液(5 mM Tris、0.5 mM EDTA、12.5%グリセロール、0.02%プロモフェノールブルー、0.02%キシレンシアン、80% DMSO)と組み合わせることによって、Cy3標識鎖をその相補鎖から完全に変性させた。全長C含有基質を、10% TBE-尿素ゲル (Bio-Rad) 上で、切断されたU含有編集基質から分離し、GE Amersham Typhoonイメージャー上で画像化した。

30

40

【0540】

[ハイスループットシーケンシングのためのin vitro編集dsDNAの調製]

下記オリゴヌクレオチドはIDT社から入手した。相補的配列をTris緩衝液中で組み合わ

50

せ(5 μ lの100 μ M溶液)、95 $^{\circ}$ Cに5分間加熱し、続いて45 $^{\circ}$ Cまで0.1 $^{\circ}$ C/秒の速度で徐々に冷却することによってアニールし、60 bpのdsDNA基質を生成した。精製融合タンパク質(活性緩衝液中1.9 μ Mを20 μ l)を1当量の適切なsgRNAと組み合わせ、周囲温度で5分間インキュベートした。60マーdsDNA基質を125 nMの最終濃度で添加し、得られた溶液を37 $^{\circ}$ Cで2時間インキュベートした。緩衝液PB (100 μ l, Qiagen)およびイソプロパノール(25 μ l)の添加により融合物からdsDNAを分離し、EconoSpinマイクロスピncラム(Epoch Life Science)で精製し、20 μ lのTris緩衝液で溶出した。得られた編集されたDNA (1 μ lをテンプレートとして使用した)を、製造業者の指示に従って上記ハイスループット配列決定プライマー対およびVeraSeq Ultra (Enzymatics) を用いて13サイクルの増幅を伴うPCRによって増幅した。RapidTips (Diffinity Genomics)を用いてPCR反応生成物を精製し、精製したDNAを、シーケンシングアダプターを含むプライマーを用いてPCRにより増幅し、精製し、先に記載したようにMiSeqハイスループットDNAシーケンサー(Illumina) 上で配列決定した。

【0541】

[細胞培養]

HEK293T (ATCC CRL-3216)およびU2OS (ATCC HTB-96)を、10% (v/v) ウシ胎児血清 (FBS) を補充したダルベッコの改変イーグル培地プラスGlutaMax (ThermoFisher) 中で、37 $^{\circ}$ C、5% CO₂で維持した。HCC1954細胞(ATCC CRL-2338)を、上記のように補充したRPMI-1640培地(ThermoFisher Scientific)中で維持した。SERPINA1遺伝子を含む不死化細胞(Taconic Biosciences)を、10% (v/v) ウシ胎児血清 (FBS) および200 μ g ml⁻¹ゲネチシン(ThermoFisher Scientific)を補足したダルベッコの改変イーグル培地プラスGlutaMax (ThermoFisher Scientific)中で培養した。

【0542】

[トランスフェクション]

HEK293Tを、48ウェルのコラーゲン被覆BioCoatプレート (Corning) 上に播種し、約85%コンフルエントの状態トランスフェクトした。簡単に述べれば、750 ngのBEおよび250 ngのsgRNA発現プラスミドを、1.5 μ lのリポフェクタミン2000 (ThermoFisher Scientific)/ウェルを用いて、製造業者のプロトコルに従ってトランスフェクトした。HEK293T細胞を、メーカーの指示に従って、適切なAmaxa Nucleofector IIプログラムを用いてトランスフェクトした(HEK 293T細胞のためのプログラムQ-001を用いるVキット)。

【0543】

[ゲノムDNA試料のハイスループットDNA配列決定]

トランスフェクトした細胞を3日後に収穫し、Agencourt DNAdvance Genomic DNA Isolation Kit (Beckman Coulter)を製造業者の指示に従って用いてゲノムDNAを単離した。対象のオン・ターゲットおよびオフ・ターゲットゲノム領域を、隣接するハイスループット配列決定プライマー対BEAM53/BEAM54またはBEAM1704/BEAM54を用いたPCRによって増幅した。PCR増幅は、5 ngのゲノムDNAをテンプレートとして使用し、Phusion高忠実度DNAポリメラーゼ (ThermoFisher) を用いて、製造業者の指示に従って実施した。反応が直線的増幅範囲内で停止することを確実にするために、各プライマー対について別々にサイクル数を決定した。RapidTips (Diffinity Genomics)を用いてPCR産物を精製した。精製DNAを、配列決定アダプターを含むプライマーを用いたPCRにより増幅した。生成物をゲル精製し、Quant-iT PicoGreen dsDNA Assay Kit (ThermoFisher) およびKAPA Library Quantification Kit-Illumina (KAPA Biosystems)を用いて定量した。サンプルは、前述のようにIllumina MiSeq上で配列決定された(Pattanaya k, Nature Biotechnol. 31, 839-843 (2013))。

【0544】

[データ分析]

配列決定リード (read) は、MiSeq Reporter (Illumina) を使用して自動的に脱マルチプレックス化され、個々のFASTQファイルを、カスタムMatlabを使用して分析した。

10

20

30

40

50

各リードは、Smith-Watermanアルゴリズムを使用して、適切な参照配列にペアワイズでアラインメントされた。Qスコアが31未満の塩基コールはNに置き換え、ヌクレオチド頻度の計算から除外した。この処理により、予測されるMiSeq塩基呼び出しエラー率は約1,000分の1になる。リードと参照配列がギャップを含まないアラインメント配列は、アラインメントテーブルに保存し、そこから各遺伝子座について塩基頻度を表にすることができた。インデル頻度は、前述の基準 (Zuris, et al., Nature Biotechnol. 33, 73-80 (2015)) を用いて、カスタムのMatlabスクリプトで定量した。配列決定リードをスキャンして、インデル(indel)が生じる可能性のあるウィンドウの両側に隣接する二つの10 bp配列との正確なマッチを探した。完全なマッチが見つからなかった場合、そのリードは分析から除外した。このインデルウィンドウの長さが参照配列と正確に一致した場合、そのリードはインデルを含まないものとして分類された。インデルウィンドウが参照配列よりも2塩基以上長いまたは短い場合、その配列決定リードは、それぞれ挿入または欠失として分類された。

10

【図面】

【図1】

【図2】

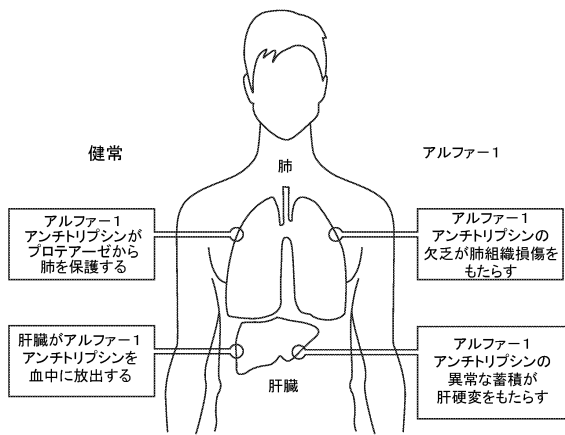


FIG. 1

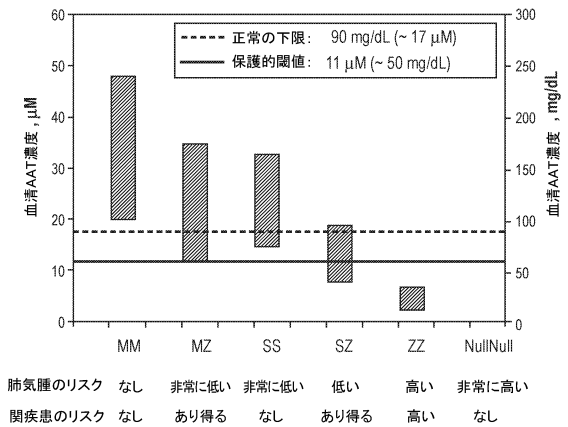


FIG. 2

20

30

40

50

【 図 3 】

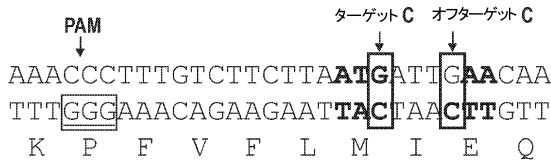


FIG. 3

【 図 4 】

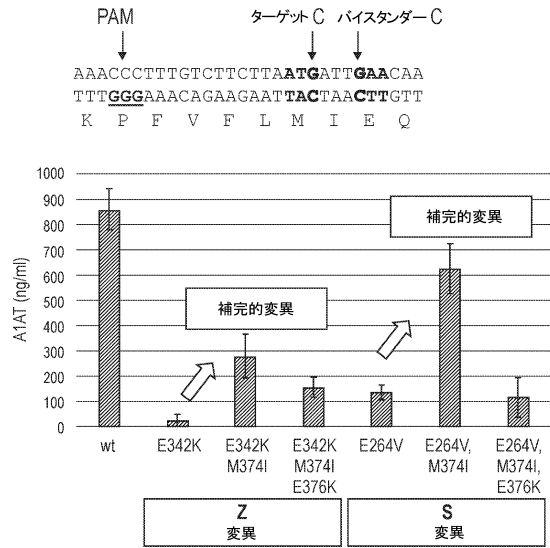


FIG. 4

【 図 5 】

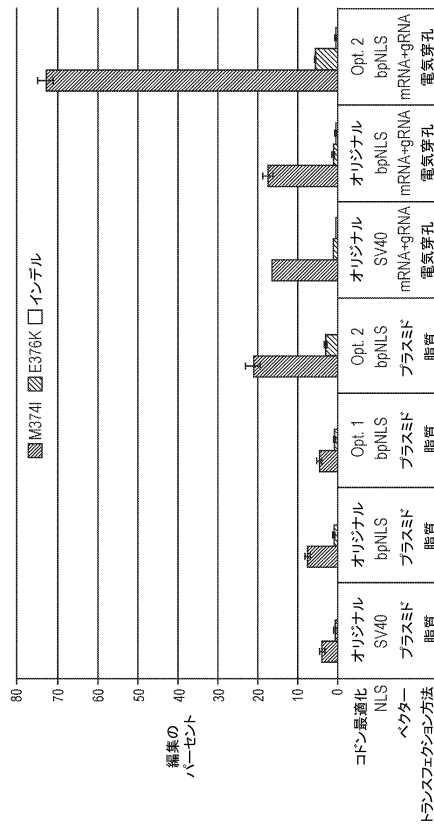


FIG. 5

【 図 6 】

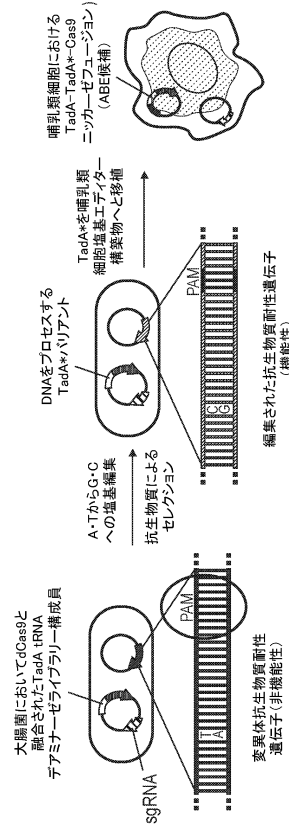


FIG. 6

10

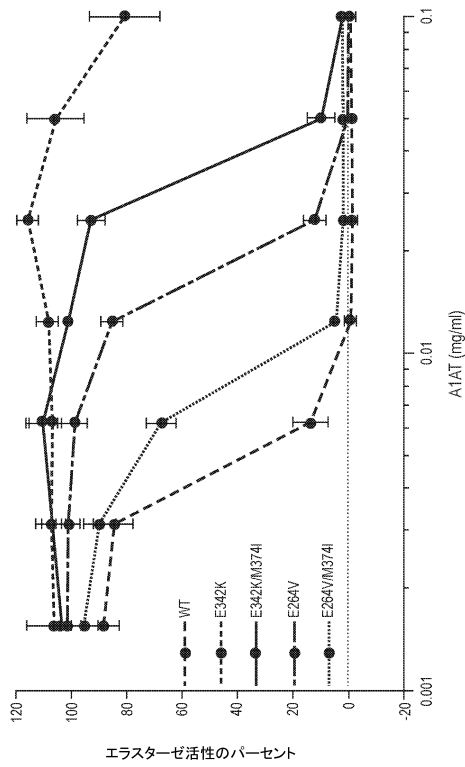
20

30

40

50

【 図 7 】



【 図 8 A 】

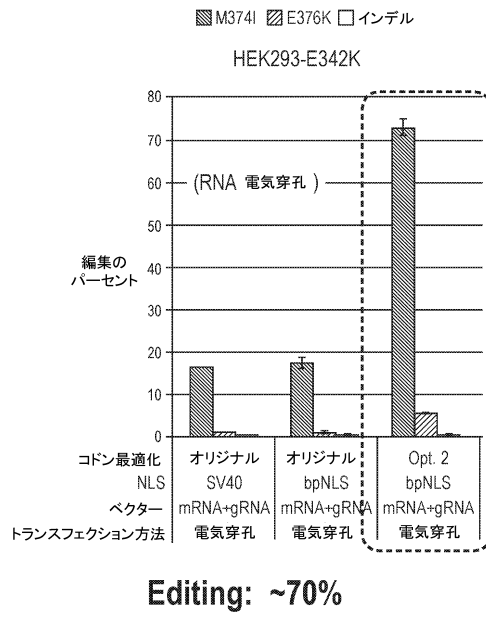
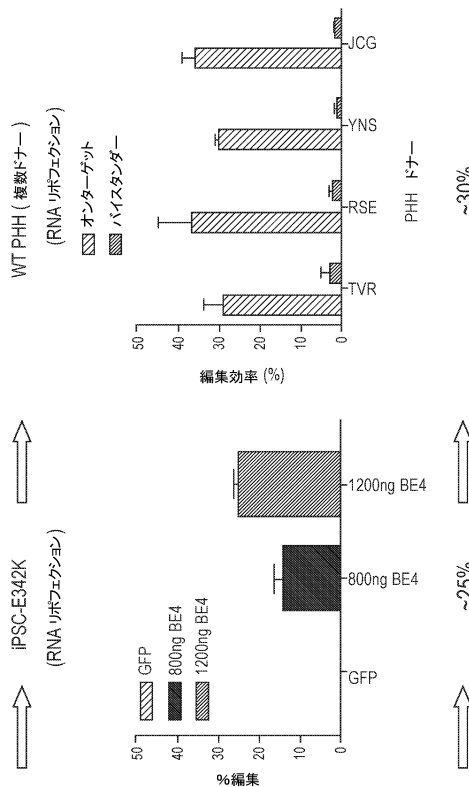


FIG. 8A

【 図 8 B - 8 C 】



【 図 9 】

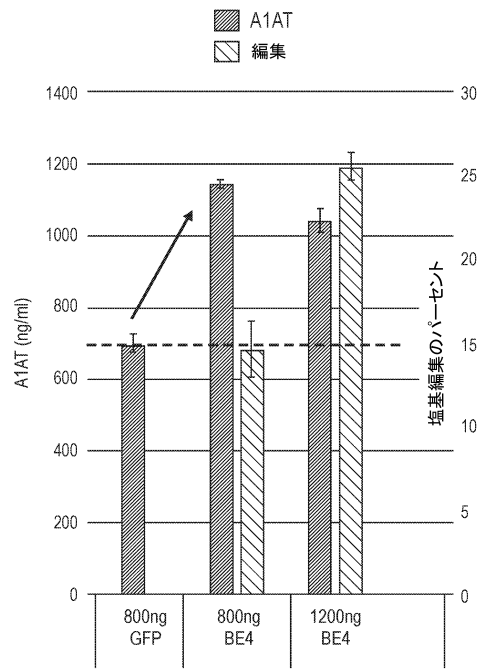


FIG. 9

10

20

30

40

50

【配列表】

0007558929000001.app

10

20

30

40

50

フロントページの続き

(51)国際特許分類

A 6 1 K	35/12 (2015.01)	F I	A 6 1 K	35/12	
A 6 1 K	35/407 (2015.01)		A 6 1 K	35/407	
A 6 1 K	35/545 (2015.01)		A 6 1 K	35/545	
A 6 1 K	38/55 (2006.01)		A 6 1 K	38/55	
A 6 1 P	43/00 (2006.01)		A 6 1 P	43/00	1 1 1
C 1 2 N	5/071 (2010.01)		C 1 2 N	5/071	
C 1 2 N	9/22 (2006.01)		C 1 2 N	9/22	
C 0 7 K	14/81 (2006.01)		C 0 7 K	14/81	
C 1 2 N	15/12 (2006.01)		C 1 2 N	15/12	

(33)優先権主張国・地域又は機関

米国(US)

セッツ州 ケンブリッジ ランズダウン ストリート 2 6 2nd フロア

(72)発明者 フ、 ヤンファン

アメリカ合衆国 0 2 1 3 9 マサチューセッツ州 ケンブリッジ ランズダウン ストリート 2 6
2nd フロア

(72)発明者 バッカー、 マイケル

アメリカ合衆国 0 2 1 3 9 マサチューセッツ州 ケンブリッジ ランズダウン ストリート 2 6
2nd フロア

審査官 松井 一泰

(56)参考文献 国際公開第 2 0 1 8 / 0 2 7 0 7 8 (W O , A 1)

(58)調査した分野 (Int.Cl. , D B 名)

C 1 2 N 1 5 / 0 0 - 1 5 / 9 0

C 1 2 N 1 / 0 0 - 7 / 0 8

C A p l u s / M E D L I N E / E M B A S E / B I O S I S (S T N)

U n i P r o t / G e n e S e q