(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7191388号
(P7191388)

(45)発行日　令和4年12月19日(2022.12.19)

(24)登録日　令和4年12月9日(2022.12.9)

(51)国際特許分類		FI		
C12N 15/09 (2006.01)		C12N	15/09	110
C12N 15/63 (2006.01)		C12N	15/63	Z
C07K 19/00 (2006.01)		C07K	19/00	
C07K 14/195 (2006.01)		C07K	14/195	
C12N 15/60 (2006.01)		C12N	15/60	

請求項の数　52　（全323頁）　最終頁に続く

(21)出願番号	特願2019-552463(P2019-552463)
(86)(22)出願日	平成30年3月23日(2018.3.23)
(65)公表番号	特表2020-511970(P2020-511970A)
(43)公表日	令和2年4月23日(2020.4.23)
(86)国際出願番号	PCT/US2018/024208
(87)国際公開番号	WO2018/176009
(87)国際公開日	平成30年9月27日(2018.9.27)
審査請求日	令和3年3月23日(2021.3.23)
(31)優先権主張番号	62/490,587
(32)優先日	平成29年4月26日(2017.4.26)
(33)優先権主張国・地域又は機関	米国(US)
(31)優先権主張番号	62/511,934
(32)優先日	平成29年5月26日(2017.5.26)

最終頁に続く

(73)特許権者　507044516
プレジデント アンド フェローズ オブ ハーバード カレッジ
アメリカ合衆国 マサチューセッツ 02138， ケンブリッジ， クインシー ストリート 17

(74)代理人　100102842
弁理士　葛和 清司

(74)代理人　100189924
弁理士　小田切 美紗

(72)発明者　リュー，デイビッド，アール．
アメリカ合衆国 マサチューセッツ州 02420、レキシントン、ウィットマン サークル 3

(72)発明者　コモー，アレクシス，クリスティーン

最終頁に続く

(54)【発明の名称】　核酸によってプログラム可能なＤＮＡ結合蛋白質を含む核酸塩基編集因子

(57)【特許請求の範囲】

【請求項１】

（i）核酸によってプログラム可能なDNA結合蛋白質（napDNAbp）ドメイン；（ii）シチジンデアミナーゼドメイン；（iii）第1のウラシルグリコシラーゼ阻害因子（UGI）ドメイン；および（iv）第2のウラシルグリコシラーゼ阻害因子（UGI）ドメインを含み、ここで、核酸によってプログラム可能なDNA結合蛋白質（napDNAbp）ドメインが、CasX、CasY、Cpf1、Cpf1ニッカーゼ、dCpf1、Cpf1変異体蛋白質、C2c1、C2c2、C2c3、Cas9、dCas9、Cas9ニッカーゼ、もしくはアルゴノート(Ago)蛋白質であるか、または、配列番号2、4、6～41および47のいずれか1つと少なくとも90%の配列同一性を有する、融合蛋白質。

【請求項２】

第1のUGIドメインおよび／または第2のUGIドメインが、配列番号134と少なくとも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%同一であるアミノ酸配列、または、配列番号134によって規定されるアミノ酸配列を含む、請求項1に記載の融合蛋白質。

【請求項３】

融合蛋白質が、構造：

NH_2-[シチジンデアミナーゼドメイン]-[napDNAbpドメイン]-[第1のUGIドメイン]-[第2のUGIドメイン]-COOH；

NH_2-[第1のUGIドメイン]-[第2のUGIドメイン]-[シチジンデアミナーゼドメイン]-[n

apDNAbpドメイン]-COOH；
NH_2-[napDNAbpドメイン]-[シチジンデアミナーゼドメイン]-[第1のUGIドメイン]-[第2のUGIドメイン]-COOH；
NH_2-[第1のUGIドメイン]-[第2のUGIドメイン]-[napDNAbpドメイン]-[シチジンデアミナーゼドメイン]-COOH；
NH_2-[シチジンデアミナーゼドメイン]-[第1のUGIドメイン]-[第2のUGIドメイン]-[napDNAbpドメイン]-COOH；または
NH_2-[napDNAbpドメイン]-[第1のUGIドメイン]-[第2のUGIドメイン]-[シチジンデアミナーゼドメイン]-COOH；
を含み、
ここで、「]-[」の各場合が、任意のリンカーを含む、請求項1または2に記載の融合蛋白質。

【請求項４】
核局在配列（NLS）をさらに含む、請求項1～3のいずれか一項に記載の融合蛋白質。

【請求項５】
NLSが、二分NLSである、請求項4に記載の融合蛋白質。

【請求項６】
NLSが、アミノ酸配列：PKKKRKV（配列番号614）、MDSLLMNRRKFLYQFKNVRWAKGRRETYLC(配列番号615)、またはKRTADGSEFEPKKKRKV（配列番号740）を含む、請求項5に記載の融合蛋白質。

【請求項７】
(1)シチジンデアミナーゼドメインおよびnapDNAbpドメインが、リンカーを介して連結されている、および／または(2)第1のUGIドメインおよびNLSが、リンカーを介して連結されている、および／または(3)第2のUGIドメインおよびNLSが、リンカーを介して連結されている、および／または(4)napDNAbpドメインおよび第1のUGIドメインが、リンカーを介して連結されている、および／または(5)napDNAbpドメインおよび第2のUGIドメインが、リンカーを介して連結されている、および／または(6)第1のUGIドメインおよび第2のUGIドメインが、リンカーを介して連結されている、請求項4～6のいずれか一項に記載の融合蛋白質。

【請求項８】
融合蛋白質が、構造：NH_2-[シチジンデアミナーゼドメイン]-[napDNAbpドメイン]-[第1のUGIドメイン]-[第2のUGIドメイン]-[NLS]-COOHを含み、ここで「]-[」の各場合が、任意のリンカーを含む、
請求項4～7のいずれか一項に記載の融合蛋白質。

【請求項９】
第2のUGIドメインおよびNLSが、アミノ酸配列SGGS（配列番号606）を含むリンカーを介して連結されている、請求項7または8に記載の融合蛋白質。

【請求項１０】
napDNAbpドメインおよび第1のUGIドメインが、アミノ酸配列SGGS（配列番号606）を含むリンカーを介して連結されている、請求項7～9のいずれか一項に記載の融合蛋白質。

【請求項１１】
napDNAbpドメインおよび第1のUGIドメインが、アミノ酸配列$SGGS(GGS)_n$（配列番号612）を含むリンカーを介して連結されており、ここでnが、2である、請求項7～10のいずれか一項に記載の融合蛋白質。

【請求項１２】
第1のUGIドメインおよび第2のUGIドメインが、アミノ酸配列$SGGS(GGS)_n$（配列番号612）を含むリンカーを介して連結されており、ここでnが、2である、請求項7～11のいずれか一項に記載の融合蛋白質。

【請求項１３】
シチジンデアミナーゼドメインおよびnapDNAbpドメインが、アミノ酸配列SGSETPGTS

ESATPES（配列番号604）またはSGGSSGGSSGSETPGTSESATPESSGGSSGGS(配列番号605)を含むリンカーを介して連結されている、請求項7～12のいずれか一項に記載の融合蛋白質。

【請求項１４】

napDNAbpドメインが、(i)配列番号6、7、37、または38と少なくとも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%同一であるアミノ酸配列、または(ii)配列番号7または38のアミノ酸配列を含む、請求項1～13のいずれか一項に記載の融合蛋白質。

【請求項１５】

napDNAbpドメインが、ヌクレアーゼ不活性型Staphylococcus aureus Cas9（SaCas9d）蛋白質またはStaphylococcus aureus Cas9ニッカーゼ（SaCas9n）蛋白質である；請求項1～14のいずれか一項に記載の融合蛋白質。

【請求項１６】

SaCas9d蛋白質またはSaCas9n蛋白質が、(i)配列番号33～36のいずれか1つと少なくとも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%同一であるアミノ酸配列、または(ii)配列番号34または35のアミノ酸配列を含む、請求項15に記載の融合蛋白質。

【請求項１７】

napDNAbpドメインが、

(a)CasY蛋白質であり；ここで、CasY蛋白質が、(i)配列番号31と少なくとも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%同一であるアミノ酸配列、または(ii)配列番号31のアミノ酸配列を含む；

(b)Cpf1、dCpf1、Cpf1ニッカーゼ、またはCpf1変異体蛋白質であり；ここで、Cpf1ニッカーゼまたは変異体蛋白質が、(i)配列番号10～24のいずれか1つと少なくとも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%同一であるアミノ酸配列、(ii)配列番号10～16のいずれか1つと少なくとも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%同一であるアミノ酸配列、または(iii)配列番号10～24のいずれか1つのアミノ酸配列を含む；

(c)C2c1蛋白質であり；ここで、C2c1蛋白質が、(i)配列番号26と少なくとも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%同一であるアミノ酸配列、または(ii)配列番号26のアミノ酸配列を含む；

(d)C2c2蛋白質であり；ここで、C2c2蛋白質が、(i)配列番号27と少なくとも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%同一であるアミノ酸配列、または(ii)配列番号27のアミノ酸配列を含む；

(e)C2c3蛋白質であり；ここで、C2c3蛋白質が、(i)配列番号28と少なくとも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%同一であるアミノ酸配列、または(ii)配列番号28のアミノ酸配列を含む；

(f)アルゴノート蛋白質であり；ここで、アルゴノート蛋白質が、(i)配列番号25と少なくとも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%同一であるアミノ酸配列、または(ii)配列番号25のアミノ酸配列を含む；あるいは

(g)CasX蛋白質であり；ここで、CasX蛋白質が、(i)配列番号29または30と少なくとも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%同一であるアミノ酸配列、または(ii)配列番号29または30のアミノ酸配列を含む、

請求項1～13のいずれか一項に記載の融合蛋白質。
【請求項１８】
napDNAbpドメインが、Cpf1ニッカーゼ(nCpf1)である、請求項1～17のいずれか一項に記載の融合蛋白質。
【請求項１９】
Cpf1ニッカーゼ(nCpf1)が、Lachnospiraceae種からのCpf1ニッカーゼ（LbCpf1）または、Acidaminococcus種からのCpf1ニッカーゼ(AsCpf1)である、請求項18に記載の融合蛋白質。
【請求項２０】
リンカーのいずれか1つが、アミノ酸配列SGGS$(GGS)_n$(配列番号612)、$(GGGS)_n$(配列番号613)、$(GGGGS)_n$(配列番号607)、$(G)_n$(配列番号608)、$(EAAAK)_n$(配列番号609)、$(GGS)_n$(配列番号610)、$(SGGS)_n$(配列番号606)、SGSETPGTSESATPES(配列番号604)、SGGSSGGSSGSETPGTSESATPESSGGSSGGS(配列番号605)、$(XP)_n$(配列番号611)、またはそれらの組み合わせを含み、ここでnが、独立して1から30までの間の整数（両端含む）であり、およびXが、いずれかのアミノ酸である、請求項3～19のいずれか一項に記載の融合蛋白質。
【請求項２１】
シチジンデアミナーゼドメインが：
(i)アポリポ蛋白質B mRNA編集複合体（APOBEC）ファミリーデアミナーゼからのデアミナーゼである;
(ii)配列番号49～84のいずれか1つのアミノ酸配列と少なくとも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%同一であるアミノ酸配列を含む；
(iii)配列番号49～84のいずれか1つのアミノ酸配列を含む;
(iv)配列番号76のW90Y、R126E、およびR132Eからなる群から選択される1つ以上の変異、または別のAPOBECデアミナーゼの対応する1つ以上の変異を含む、ラットAPOBEC1（rAPOBEC1）デアミナーゼである；
(v)配列番号74のW90Y、Q126E、およびR132Eからなる群から選択される1つ以上の変異、または別のAPOBECデアミナーゼの対応する1つ以上の変異を含む、ヒトAPOBEC1（hAPOBEC1）デアミナーゼである；
(vi)配列番号60のW285Y、R320E、およびR326Eからなる群から選択される1つ以上の変異、または別のAPOBECデアミナーゼの対応する1つ以上の変異を含む、ヒトAPOBEC3G（hAPOBEC3G）デアミナーゼである；
(vii)活性化誘導デアミナーゼ（AID）である;または
(viii) Petromyzon marinusからのシチジンデアミナーゼ1（pmCDA1）である、
請求項1～20のいずれか一項に記載の融合蛋白質。
【請求項２２】
融合蛋白質が、2つ、3つ、4つ、または5つのUGIドメインを含む、請求項1～21のいずれか一項に記載の融合蛋白質。
【請求項２３】
請求項1～22のいずれか一項に記載の融合蛋白質と、融合蛋白質のnapDNAbpドメインに結合したガイドRNAとを含む、複合体。
【請求項２４】
ガイドRNAが、配列番号618の核酸配列を含む、請求項23に記載の複合体。
【請求項２５】
ガイドRNAが、10～100ヌクレオチド長であり、および標的配列に対して相補的である少なくとも10個の一続きのヌクレオチドの配列を含む、請求項23または24に記載の複合体。
【請求項２６】
標的配列が、哺乳動物のゲノム中にある、請求項25に記載の前記複合体。

【請求項２７】
哺乳動物が、ヒト、マウス、およびラットからなる群から選択される、請求項26に記載の前記複合体。
【請求項２８】
請求項1～13、17、および20～22のいずれか一項に記載の融合蛋白質と、融合蛋白質のnapDNAbpドメインに結合したガイドDNAとを含む、複合体。
【請求項２９】
請求項1～22のいずれか一項に記載の融合蛋白質をコードする、ポリヌクレオチド。
【請求項３０】
ガイドRNAまたはガイドDNAをさらにコードする、請求項29に記載のポリヌクレオチド。
【請求項３１】
請求項29または30に記載のポリヌクレオチドを含む、ベクター。
【請求項３２】
請求項1～22のいずれか一項に記載の融合蛋白質、請求項23～28のいずれか一項に記載の複合体、請求項29または30に記載のポリヌクレオチド、または請求項31に記載のベクターを含む、細胞。
【請求項３３】
請求項1～22のいずれか一項に記載の融合蛋白質、請求項23～28のいずれか一項に記載の複合体、請求項29または30に記載のポリヌクレオチド、請求項31に記載のベクター、または請求項32に記載の細胞を含む、医薬組成物。
【請求項３４】
薬物として用いるための、請求項1～22のいずれか一項に記載の融合蛋白質、請求項23～28のいずれか一項に記載の複合体、請求項29または30に記載のポリヌクレオチド、請求項31に記載のベクター、請求項32に記載の細胞、または請求項33に記載の医薬組成物。
【請求項３５】
疾患、異常、または状態の処置のための薬物の製造において用いるための、請求項1～22のいずれか一項に記載の融合蛋白質、請求項23～28のいずれか一項に記載の複合体、請求項29または30に記載のポリヌクレオチド、請求項31に記載のベクター、請求項32に記載の細胞、または請求項33に記載の医薬組成物。
【請求項３６】
疾患、異常、または状態の処置のための薬物の製造において用いるための、請求項1～22のいずれか一項に記載の融合蛋白質、請求項23～28のいずれか一項に記載の複合体、請求項29または30に記載のポリヌクレオチド、請求項31に記載のベクター、請求項32に記載の細胞、または請求項33に記載の医薬組成物であって、ここで、疾患、異常、または状態が、嚢胞性線維症、フェニルケトン尿症、表皮剥離性角化症（EHK）、4J型シャルコー・マリー・トゥース病、神経芽腫（NB）、フォン・ヴィレブランド病（vWD）、先天性ミオトニー、遺伝性腎アミロイドーシス、拡張型心筋症（DCM）、遺伝性リンパ浮腫、家族性アルツハイマー病、HIV、プリオン病、慢性乳児神経皮膚関節症候群（CINCA）、デスミン関連ミオパチー（DRM）、あるいは、変異体PI3KCA蛋白質、変異体CTNNB1蛋白質、変異体HRAS蛋白質、または変異体p53蛋白質に関連する新生物性疾患である、前記融合蛋白質、複合体、ポリヌクレオチド、ベクター、細胞、または医薬組成物。
【請求項３７】
核酸分子を、請求項1～22のいずれか一項に記載の融合蛋白質およびガイドRNAに接触させること（ここで、ガイドRNAが、生物のゲノム中の標的配列に対して相補的でありかつ標的塩基対を含む少なくとも10個の一続きのヌクレオチドの配列を含み）；または
核酸分子を、請求項23～28のいずれか一項に記載の複合体に接触させること、
を含む方法であって、ここで、方法が、in vitroでもしくはex vivoで、またはin vivoで非ヒト動物に行われる、前記方法。
【請求項３８】

標的配列が、疾患または異常に関連するT→C点変異を含み、およびここで、変異体C塩基の脱アミノ化が、疾患または異常に関連しない配列をもたらす、
請求項37に記載の方法。
【請求項３９】
点変異が、コドン内にあり、野生型コドンと比較して、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、請求項37または38に記載の方法。
【請求項４０】
リボヌクレオ蛋白質（RNP）複合体を生成するための方法であって、方法が、
（i）核酸塩基編集因子を、水溶液中でRNAと複合体化させ、それによって、水溶液中で核酸塩基編集因子およびRNAを含む複合体を形成すること；および
（ii）（i）の複合体をカチオン性脂質に接触させること
を含み；
ここで、核酸塩基編集因子が、請求項1～22のいずれか一項に記載の融合蛋白質であり、および／またはここで、（i）の複合体が、請求項23～28のいずれか一項に記載の複合体である、前記方法。
【請求項４１】
（i）の核酸塩基編集因子およびRNAが、1:1から1:1.5までのモル比で複合体化させられる、請求項40に記載の方法。
【請求項４２】
核酸塩基編集因子を精製するための方法であって、方法が、
（i）細胞において核酸塩基編集因子を発現させること、ここで、核酸塩基編集因子が、アフィニティータグを含む；
（ii）（i）の細胞を溶解し、それによって溶解物を創出すること；および
（iii）溶解物をアフィニティークロマトグラフィーに付し、それによって、精製された核酸塩基編集因子を含む溶出物を生成すること
を含み、
ここで、核酸塩基編集因子が、請求項1～22のいずれか一項に記載の融合蛋白質である、前記方法。
【請求項４３】
請求項1～22のいずれか一項に記載の融合蛋白質、請求項23～28のいずれか一項に記載の複合体、請求項29または30に記載のポリヌクレオチド、請求項31に記載のベクター、請求項32に記載の細胞、または請求項33に記載の医薬組成物を、対象の内耳へ送達することを含む方法であって、ここで、対象が、非ヒト動物である、前記方法。
【請求項４４】
二本鎖DNA配列の核酸塩基対を編集するための方法であって、方法が、
二本鎖DNA配列の標的領域を、核酸塩基編集因子およびガイド核酸を含む複合体に接触させること（ここで標的領域は、標的核酸塩基対を含む）；およびそれによって
該標的領域の鎖分離を誘導すること；
標的領域の一本鎖中の該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こと；
該標的領域の一方の鎖だけを切ること；および
第1の核酸塩基に相補的な第3の核酸塩基を、第2の核酸塩基に相補的な第4の核酸塩基で置き換えること；
を含み；
ここで、方法が、二本鎖DNA配列中に20%未満のインデル形成を引き起こし；
ここで、核酸塩基編集因子が、Cas9、dCas9、Cas9ニッカーゼ、CasX、CasY、Cpf1、dCpf1、Cpf1ニッカーゼ、C2c1、C2c2、C2c3、もしくはアルゴノートドメイン、または、配列番号2、4、6～41および47のいずれか1つと少なくとも90%の配列同一性を有するドメインを含み；
ここで、核酸塩基編集因子が、シチジンデアミナーゼドメイン、第1のUGIドメインおよび

第2のUGIドメインをさらに含み、
ここで、方法が、in vitroでもしくはex vivoで、またはin vivoで非ヒト動物に行われる、前記方法。
【請求項４５】
核酸塩基編集因子が、Cas9ニッカーゼドメインを含む、請求項44に記載の方法。
【請求項４６】
方法が、19%、18%、17%、16%、15%、14%、13%、12%、11%、10%、9%、8%、7%、6%、5%、4%、3%、2%、または1%未満のインデル形成を引き起こす、請求項44または45に記載の方法。
【請求項４７】
意図される編集された塩基対を創出する効率が、少なくとも5%、少なくとも10%、少なくとも20%、少なくとも30%、少なくとも40%、または少なくとも50%である、請求項44～46のいずれか一項に記載の方法。
【請求項４８】
意図される編集された塩基対：意図されない編集された塩基対の比が、2:1、3:1、4:1、5:1、6:1、7:1、または8:1である、請求項44～46のいずれか一項に記載の方法。
【請求項４９】
切られる一本鎖が、ガイド核酸にハイブリダイズさせられる、
請求項44～48のいずれか一項に記載の方法。
【請求項５０】
意図される編集された塩基対が、(a)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か；または(b)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、請求項44～49のいずれか一項に記載の方法。
【請求項５１】
核酸塩基編集因子が、請求項1～22のいずれか一項に記載の融合蛋白質を含む、請求項44～50のいずれか一項に記載の方法。
【請求項５２】
核酸分子を、請求項1～13、17および20～22のいずれか一項に記載の融合蛋白質およびガイドDNAに接触させること(ここで、ガイドDNAは、標的塩基対を含む生物のゲノム中の標的配列に対して相補的である少なくとも10個の一続きのヌクレオチドの配列を含む)；または
核酸分子を、請求項28に記載の複合体に接触させること
を含む、方法。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明の背景
核酸配列の標的化された編集、例えば標的化された切断またはゲノムDNAへの特異的な改変の標的化された導入は、遺伝子機能の研究にとって高度に有望なアプローチであり、ヒト遺伝子疾患の新たな治療を提供するポテンシャルをもまた有する[1]。理想的な核酸編集テクノロジーは3つの特徴を持つ：（1）所望の改変をインストールする高い効率；（2）最小限のオフターゲット活性；および（3）所与の核酸中のいずれかの部位、例えばヒトゲノム中のいずれかの部位を正確に編集するようにプログラムされる能力[2]。人工ジンクフィンガーヌクレアーゼ（ZFN）[3]、転写活性化因子様エフェクターヌクレアーゼ（TALEN）[4]、およびごく最近ではRNAによってガイドされるDNAエンドヌクレアーゼCas9[5]を包含する現行のゲノム工学ツールは、ゲノムの配列特異的DNA切断を成し遂げる。このプログラム可能な切断は、非相同末端結合（NHEJ）による切断部位のDNAの変異または相同組み換え修復（HDR）による切断部位の周辺のDNAの置き換えをもたらし得る[6,7]。

【０００２】
現行のテクノロジーの1つの欠点は、NHEJおよびHDRは両方とも確率的なプロセスであり、これらは、典型的には控えめな遺伝子編集効率、および所望の変改と競合し得る望まれない遺伝子変改をもたらすことである[8]。原理的には、多くの遺伝子疾患は、ゲノム中の特異的な場所における特異的なヌクレオチド変化（例えば、疾患に関連する遺伝子の特異的なコドンにおけるC→T変化）を成し遂げることによって処置され得るので[9]、かかる正確な遺伝子編集を達成するためのプログラム可能なやり方の開発は、強力な新たな研究ツールと遺伝子編集に基づくヒト治療法の見込まれる新たなアプローチと両方に相当するであろう。

【発明の概要】

【０００３】
本発明の概要
核酸によってプログラム可能なDNA結合蛋白質（napDNAbp）、例えばclustered regularly interspaced short palindromic repeat（CRISPR）システムは、最近発見された原核生物適応免疫システムであり[10]、これは、ロバストかつ一般的なゲノム工学を種々の生物および細胞株において可能にするように改変されている[11]。CRISPR-Cas（CRISPR関連）システムは蛋白質-RNA複合体であり、これらは塩基対形成を介して標的DNA配列に複合体を局在させるためのガイドとしてRNA分子（sgRNA）を用いる[12]。天然のシステムでは、それからCas蛋白質がエンドヌクレアーゼとして作用して、標的DNA配列を切断する[13]。システムが機能するためには、標的DNA配列はsgRNAに対して相補的でなければならず、かつ相補的な領域の3'末端に「プロトスペーサー隣接モチーフ」（PAM）をもまた含有しなければならない[14]。

【０００４】
公知のCas蛋白質のうち、S. pyogenes Cas9はゲノム工学のためのツールとして主に幅広く用いられている[15]。このCas9蛋白質は2つの別々のヌクレアーゼドメインを含有する大きいマルチドメイン蛋白質である。ヌクレアーゼ活性を廃絶するために点変異がCas9に導入され得、sgRNAによってプログラムされた様式でDNAに結合するその能力をなお保持するデッドCas9（dCas9）をもたらす[16]。原理的には、別の蛋白質またはドメインに融合させられたときに、dCas9は単純に適切なsgRNAとの共発現によってその蛋白質またはドメインを事実上いずれかのDNA配列へと標的化し得る。

【０００５】
ゲノム工学目的のためのdCas9複合体のポテンシャルは計り知れない。sgRNAによってプログラムされたゲノム中の特異的部位に蛋白質を持って行くそのユニークな能力は、理論上は、ヌクレアーゼを越えた種々の部位特異的ゲノム工学ツールへと発展させられ得、デアミナーゼ（例えばシチジンデアミナーゼ）、転写活性化因子、転写リプレッサー、ヒストン修飾蛋白質、インテグラーゼ、およびリコンビナーゼを包含する[11]。これらの見込まれる用途のいくつかは、転写活性化因子とのdCas9融合体によって最近実装されて、RNAによってガイドされる転写活性化因子[17,18]、転写リプレッサー[16,19,20]、およびクロマチン修飾酵素[21]を提供している。これらの融合体の種々のsgRNAとの単純な共発現は標的遺伝子の特異的発現をもたらす。これらの画期的研究は、ゲノムを正確に操るための容易くプログラム可能な配列特異的エフェクターの設計および構築のための道を敷いた。

【０００６】
本開示のいくつかの側面は、リンカーによって融合させられた核酸によってプログラム可能なDNA結合蛋白質（napDNAbp）、例えばCasX、CasY、Cpf1、C2c1、C2c2、C2c3、またはアルゴノート蛋白質と、シチジンデアミナーゼドメインとの一定の構成が、標的シチジン残基を効率的に脱アミノ化するために有用であるという認識に基づく。本開示の他の側面は、リンカーを介してnapDNAbpのN末端に融合させられたシチジンデアミナーゼドメインを有する核酸塩基編集融合蛋白質が、二本鎖DNA標的分子内の標的核酸を効率的に脱アミノ化することができたという認識に関する。例えば下の例3および4を見よ。これらは、本明細書においては塩基編集因子ともまた言われる融合蛋白質が、UGIドメイ

ンなしの塩基編集因子などの他の塩基編集因子よりも少ないインデルを創出し、効率的に標的核酸を脱アミノ化することを実証している。本開示の他の側面は、融合蛋白質が1つよりも多くのUGIドメインを含むときに、リンカーを介してnapDNAbpのN末端に融合させられたシチジンデアミナーゼドメインを有する核酸塩基編集融合蛋白質が、より高い効率および多大に改善された生成物純度を有する塩基編集を行うという認識に関する。例えば例１７を見よ。これは、2つのUGIドメインを含む融合蛋白質（例えば塩基編集因子）が、1つのUGIドメインを含むものなどの他の塩基編集因子よりも少ないインデルを創出し、効率的に標的核酸を脱アミノ化することを実証している。

【０００７】

いくつかの態様において、融合蛋白質は：（i）核酸によってプログラム可能なDNA結合蛋白質（napDNAbp）；（ii）シチジンデアミナーゼドメイン；および（iii）ウラシルグリコシラーゼ阻害因子（UGI）ドメインを含み、napDNAbpはCasX、CasY、Cpf1、C2c1、C2c2、C2c3、またはアルゴノート蛋白質である。いくつかの態様において、核酸によってプログラム可能なDNA結合蛋白質（napDNAbp）はCasX蛋白質である。いくつかの態様において、CasX蛋白質は配列番号29または30と少なくとも90%同一であるアミノ酸配列を含む。いくつかの態様において、CasX蛋白質は配列番号29または30のアミノ酸配列を含む。

【０００８】

いくつかの態様において、融合蛋白質は：（i）核酸によってプログラム可能なDNA結合蛋白質（napDNAbp）；（ii）シチジンデアミナーゼドメイン；（iii）第1のウラシルグリコシラーゼ阻害因子（UGI）ドメイン；および（iv）第2のウラシルグリコシラーゼ阻害因子（UGI）ドメインを含み、napDNAbpはCas9、dCas9、またはCas9ニッカーゼ蛋白質である。いくつかの態様において、napDNAbpはdCas9蛋白質である。いくつかの態様において、napDNAbpはCasX、CasY、Cpf1、C2c1、C2c2、C2c3、またはアルゴノート蛋白質である。いくつかの態様において、dCas9蛋白質はS. pyogenes dCas9（SpCas9d）である。いくつかの態様において、dCas9蛋白質はD10A変異を抱えるS. pyogenes dCas9である。いくつかの態様において、dCas9蛋白質は配列番号6または7と少なくとも90%同一であるアミノ酸配列を含む。いくつかの態様において、dCas9蛋白質は配列番号6または7のアミノ酸配列を含む。いくつかの態様において、dCas9蛋白質はS. aureus dCas9（SaCas9d）である。いくつかの態様において、dCas9蛋白質はD10A変異を抱えるS. aureus dCas9である。いくつかの態様において、dCas9蛋白質は配列番号33～36と少なくとも90%同一であるアミノ酸配列を含む。いくつかの態様において、dCas9蛋白質は配列番号33-36のアミノ酸配列を含む。

【０００９】

いくつかの態様において、核酸によってプログラム可能なDNA結合蛋白質（napDNAbp）はCasY蛋白質である。いくつかの態様において、CasY蛋白質は配列番号31と少なくとも90%同一であるアミノ酸配列を含む。いくつかの態様において、CasY蛋白質は配列番号31のアミノ酸配列を含む。

【００１０】

いくつかの態様において、核酸によってプログラム可能なDNA結合蛋白質（napDNAbp）はCpf1またはCpf1変異体蛋白質である。いくつかの態様において、Cpf1またはCpf1変異体蛋白質は配列番号9～24のいずれか1つと少なくとも90%同一であるアミノ酸配列を含む。いくつかの態様において、Cpf1またはCpf1変異体蛋白質は配列番号9～24のいずれか1つのアミノ酸配列を含む。

【００１１】

いくつかの態様において、核酸によってプログラム可能なDNA結合蛋白質（napDNAbp）はC2c1蛋白質である。いくつかの態様において、C2c1蛋白質は配列番号26と少なくとも90%同一であるアミノ酸配列を含む。いくつかの態様において、C2c1蛋白質は配列番号26のアミノ酸配列を含む。

【００１２】

いくつかの態様において、核酸によってプログラム可能なDNA結合蛋白質（napDNAbp）はC2c2蛋白質である。いくつかの態様において、C2c2蛋白質は配列番号27と少なくとも90%同一であるアミノ酸配列を含む。いくつかの態様において、C2c2蛋白質は配列番号27のアミノ酸配列を含む。

【００１３】

いくつかの態様において、核酸によってプログラム可能なDNA結合蛋白質（napDNAbp）はC2c3蛋白質である。いくつかの態様において、C2c3蛋白質は配列番号28と少なくとも90%同一であるアミノ酸配列を含む。いくつかの態様において、C2c3蛋白質は配列番号28のアミノ酸配列を含む。

【００１４】

いくつかの態様において、核酸によってプログラム可能なDNA結合蛋白質（napDNAbp）はアルゴノート蛋白質である。いくつかの態様において、アルゴノート蛋白質は配列番号25と少なくとも90%同一であるアミノ酸配列を含む。いくつかの態様において、アルゴノート蛋白質は配列番号25のアミノ酸配列を含む。

【００１５】

本開示のいくつかの側面は、本明細書において提供される融合蛋白質が、大きい割合のインデルを創出することなしに、1つ以上の変異（例えばC→T変異）を創出することができるという認識に基づく。いくつかの態様において、本明細書において提供される融合蛋白質（例えば塩基編集蛋白質）のいずれかは10%よりも少ないインデルを創出する。いくつかの態様において、本明細書において提供される融合蛋白質（例えば塩基編集蛋白質）のいずれかは10%、9%、8%、7%、6%、5.5%、5%、4.5%、4%、3.5%、3%、2.5%、2%、1.5%、1%、0.5%、または0.1%よりも少ないインデルを創出する。

【００１６】

いくつかの態様において、融合蛋白質はnapDNAbpおよびアポリポ蛋白質B mRNA編集複合体1（APOBEC1）デアミナーゼドメインを含み、ここで、デアミナーゼドメインは、アミノ酸配列SGSETPGTSESATPES（配列番号604）を含むリンカーを介してnapDNAbpドメインのN末端に融合させられる。いくつかの態様において、napDNAbpは本明細書において提供されるnapDNAbpのいずれかのアミノ酸配列を含む。いくつかの態様において、デアミナーゼはラットAPOBEC1（配列番号76）である。いくつかの態様において、デアミナーゼはヒトAPOBEC1（配列番号74）である。いくつかの態様において、デアミナーゼはpmCDA1（配列番号81）である。いくつかの態様において、デアミナーゼはヒトAPOBEC3G（配列番号60）である。いくつかの態様において、デアミナーゼは（配列番号82-84）のいずれか1つのヒトAPOBEC3Gバリアントである。いくつかの態様において、融合蛋白質はnapDNAbpとアポリポ蛋白質B mRNA編集複合体1触媒ポリペプチド様3G（APOBEC3G）デアミナーゼドメインとを含み、デアミナーゼドメインは、いずれかの長さまたは組成のリンカー（例えば、アミノ酸配列、ペプチド、ポリマー、または結合）を介してnapDNAbpドメインのN末端に融合させられる。いくつかの態様において、リンカーはアミノ酸配列SGSETPGTSESATPES（配列番号604）を含む。いくつかの態様において、リンカーはアミノ酸配列SGGSSGGSSGSETPGTSESATPESSGGSSGGS（配列番号605）を含む。

【００１７】

いくつかの態様において、融合蛋白質はnapDNAbpおよびシチジンデアミナーゼ1（CDA1）デアミナーゼドメインを含み、デアミナーゼドメインは、アミノ酸配列SGSETPGTSESATPES（配列番号604）を含むリンカーを介してnapDNAbpドメインのN末端に融合させられる。いくつかの態様において、リンカーはアミノ酸配列SGGSSGGSSGSETPGTSESATPESSGGSSGGS（配列番号605）を含む。いくつかの態様において、napDNAbpは本明細書において提供されるnapDNAbpのいずれかのアミノ酸配列を含む。

【００１８】

いくつかの態様において、融合蛋白質はnapDNAbpおよび活性化誘導シチジンデアミナーゼ（AID）デアミナーゼドメインを含み、デアミナーゼドメインは、アミノ酸配列SGSE

TPGTSESATPES（配列番号604）を含むリンカーを介してnapDNAbpドメインのN末端に融合させられる。いくつかの態様において、リンカーはアミノ酸配列SGGSSGGSSGSETPGTSESATPESSGGSSGGS（配列番号605）を含む。いくつかの態様において、napDNAbpは本明細書において提供されるnapDNAbpのいずれかのアミノ酸配列を含む。

【００１９】

本開示のいくつかの側面は、napDNAbpとリンカーによって融合させられたシチジンデアミナーゼドメインとの一定の構成が、標的シチジン残基を効率的に脱アミノ化するために有用であるという認識に基づく。本開示の他の側面は、アミノ酸配列SGSETPGTSESATPES（配列番号604）を含むリンカーを介してnapDNAbpのN末端に融合させられたアポリポ蛋白質B mRNA編集複合体1（APOBEC1）デアミナーゼドメインを有する核酸塩基編集融合蛋白質が、二本鎖DNA標的分子中の標的核酸を効率的に脱アミノ化することができたという認識に関する。いくつかの態様において、融合蛋白質はnapDNAbpドメインとアポリポ蛋白質B mRNA編集複合体1（APOBEC1）デアミナーゼドメインとを含み、デアミナーゼドメインは、アミノ酸配列SGSETPGTSESATPES（配列番号604）を含むリンカーを介してnapDNAbpのN末端に融合させられる。

【００２０】

本開示のいくつかの側面は、対象のゲノム、例えばヒトのゲノム中の単一部位を編集することを包含する、核酸の標的化された編集にとって有用である戦略、システム、試薬、方法、およびキットを提供する。いくつかの態様においては、napDNAbp（例えば、CasX、CasY、Cpf1、C2c1、C2c2、C2c3、またはアルゴノート蛋白質）およびデアミナーゼまたはデアミナーゼドメインの融合蛋白質が提供される。いくつかの態様においては、標的化された核酸編集のための方法が提供される。いくつかの態様においては、標的化された核酸編集蛋白質、例えばnapDNAbpとデアミナーゼまたはデアミナーゼドメインとの融合蛋白質の創出のための試薬およびキットが提供される。

【００２１】

本開示のいくつかの側面は、融合蛋白質を提供し、第2の蛋白質（例えば、シチジンデアミナーゼドメインなどの酵素ドメイン）に融合させられ、それゆえに融合蛋白質を形成する本明細書において提供されるnapDNAbpを含む。いくつかの態様において、第2の蛋白質は酵素ドメインまたは結合ドメインを含む。いくつかの態様において、酵素ドメインはヌクレアーゼ、ニッカーゼ、リコンビナーゼ、デアミナーゼ、メチルトランスフェラーゼ、メチラーゼ、アセチラーゼ、アセチルトランスフェラーゼ、転写活性化因子、または転写リプレッサードメインである。いくつかの態様において、酵素ドメインは核酸編集ドメインである。いくつかの態様において、核酸編集ドメインはデアミナーゼドメインである。いくつかの態様において、デアミナーゼ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る。いくつかの態様において、デアミナーゼはアポリポ蛋白質B mRNA編集複合体（APOBEC）ファミリーデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC1デアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC2デアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC3デアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC3Aデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC3Bデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC3Cデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC3Dデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC3Eデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC3Fデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC3Gデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC3Hデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC4デアミナーゼである。いくつかの態様において、デアミナーゼは活性化誘導デアミナーゼ（AID）である。デアミナーゼはいずれかの好適な生物（例えばヒトまたはラット）からであり得ることは了解されるべきである。いくつかの態様において、デアミナーゼはヒト、チンパンジー、ゴリラ、サル、ウシ(cow)、イヌ、ラット、またはマウスからである。いくつかの態様において、デアミナーゼは

ラットAPOBEC1（配列番号76）である。いくつかの態様において、デアミナーゼはヒトAPOBEC1（配列番号74）である。いくつかの態様において、デアミナーゼはpmCDA1である。

【００２２】

本開示のいくつかの側面は：（i）配列番号32のアミノ酸配列を含むCasX、CasY、Cpf1、C2c1、C2c2、C2c3、またはアルゴノート蛋白質ドメインと；（ii）アポリポ蛋白質BmRNA編集複合体1（APOBEC1）デアミナーゼドメインとを含む融合蛋白質を提供し、デアミナーゼドメインは、SGSETPGTSESATPES（配列番号604）のアミノ酸配列を含むリンカーを介してnapDNAbpのN末端に融合させられる。いくつかの態様において、デアミナーゼはラットAPOBEC1（配列番号76）である。いくつかの態様において、デアミナーゼはヒトAPOBEC1（配列番号74）である。いくつかの態様において、融合蛋白質は配列番号591のアミノ酸配列を含む。いくつかの態様において、融合蛋白質は配列番号5737のアミノ酸配列を含む。いくつかの態様において、デアミナーゼはpmCDA1（配列番号81）である。いくつかの態様において、デアミナーゼはヒトAPOBEC3G（配列番号60）である。いくつかの態様において、デアミナーゼは配列番号82～84のいずれか1つのヒトAPOBEC3Gバリアントである。

【００２３】

本開示の他の側面は、デアミナーゼドメインとnapDNAbpドメインとウラシルグリコシラーゼ阻害因子（UGI）ドメインとを含む融合蛋白質が、核酸分子中の標的ヌクレオチドを脱アミノ化することについて改善された効率を実証するという認識に関する。いずれの特定の理論によっても拘束されることを望むものではないが、U:Gヘテロ二重鎖DNAの存在に対する細胞性のDNA修復応答が、細胞の核酸塩基編集効率の減少を担い得る。ウラシルDNAグリコシラーゼ（UDG）は細胞のDNAからのUの取り除きを触媒し、これは塩基除去修復を開始し得、U:G対からC:G対への復帰を最も一般的な成果として有する。本明細書において実証される通り、ウラシルDNAグリコシラーゼ阻害因子（UGI）はヒトUDG活性を阻害し得る。いずれの特定の理論によっても拘束されることを望むものではないが、塩基除去修復は、一本鎖に結合する、編集された塩基をブロックする、UGIを阻害する、塩基除去修復を阻害する、編集された塩基を保護する、および／または編集されない鎖の「修理」を促進する分子などによって阻害され得る。それゆえに、本開示は、UGIドメインに融合させられたnapDNAbp-シチジンデアミナーゼドメインを含む融合蛋白質を企図する。

【００２４】

本開示のさらなる側面は、デアミナーゼドメインとnapDNAbpドメインと1つよりも多くのウラシルグリコシラーゼ阻害因子（UGI）ドメイン（例えば、1つ、2つ、3つ、4つ、5つ、またはより多くのUGIドメイン）とを含む融合蛋白質が、核酸分子中の標的ヌクレオチドを脱アミノ化することについて改善された効率および／または改善された核酸生成物純度を実証するという認識に関する。いずれの特定の理論によっても拘束されることを望むものではないが、第2のUGIドメインの追加はG:U塩基編集中間体へのUDGのアクセスを実質的に減少させ得、それによって塩基編集の効率を改善する。

【００２５】

本開示のいくつかの側面は、本明細書において提供される塩基編集因子のいずれかが、有意な割合のインデルを創出することなしに特異的なヌクレオチド塩基を改変することができるという認識に基づく。本明細書において用いられる「インデル」は核酸中のヌクレオチド塩基の挿入または欠失を言う。かかる挿入または欠失は遺伝子のコード領域内のフレームシフト変異に至り得る。いくつかの態様においては、核酸中に挿入または欠失（すなわちインデル）を創出することなしに核酸中の特異的なヌクレオチドを効率的に改変する（例えば変異させるかまたは脱アミノ化する）塩基編集因子を創出することが望ましい。ある態様において、本明細書において提供される塩基編集因子のいずれかは、インデルに対してより多大な割合の意図される改変（例えば点変異または脱アミノ化）を創出することができる。

【００２６】
ある態様において、本明細書において提供される塩基編集因子のいずれかは一定のパーセンテージの所望の変異を創出することができる。いくつかの態様において、所望の変異はC→T変異である。いくつかの態様において、所望の変異はC→A変異である。いくつかの態様において、所望の変異はC→G変異である。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編集因子のいずれかは少なくとも1%の所望の変異を創出することができる。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編集因子のいずれかは少なくとも1%、2%、3%、4%、5%、10%、15%、20%、25%、30%、40%、45%、50%、60%、70%、80%、90%、95%、または99%の所望の変異を創出することができる。
【００２７】
本開示のいくつかの側面は、本明細書において提供される塩基編集因子のいずれかが、有意な数の意図されない変異、例えば意図されない点変異を創出することなしに、核酸（例えば、対象のゲノム中の核酸）中に意図される変異、例えば点変異を効率的に創出することができるという認識に基づく。
【００２８】
いくつかの態様において、融合蛋白質のデアミナーゼドメインはnapDNAbpドメインのN末端に融合させられる。いくつかの態様において、UGIドメインはnapDNAbpドメインのC末端に融合させられる。いくつかの態様において、napDNAbpおよび核酸編集ドメインはリンカーを介して融合させられる。いくつかの態様において、napDNAbpドメインおよびUGIドメインはリンカーを介して融合させられる。いくつかの態様においては、第2のUGIドメインが第1のUGIドメインのC末端に融合させられる。いくつかの態様において、第1のUGIドメインおよび第2のUGIドメインはリンカーを介して融合させられる。
【００２９】
一定の態様においては、リンカーが、本発明のペプチドまたはペプチドドメインのいずれかを連結するために用いられ得る。リンカーは共有結合ほどに単純であり得るか、またはそれは長さが多くの原子のポリマー状リンカーであり得る。ある態様において、リンカーはポリペプチドであるか、またはアミノ酸に基づく。他の態様において、リンカーはペプチド様ではない。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、リンカーはアミド連結部の炭素-窒素結合である。ある態様において、リンカーは、環式または非環式の、置換または無置換の、分枝または非分枝の、脂肪族またはヘテロ脂肪族のリンカーである。ある態様において、リンカーはポリマー状である（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）。ある態様において、リンカーはアミノアルカン酸の単量体、二量体、またはポリマーを含む。ある態様において、リンカーはアミノアルカン酸を含む（例えば、グリシン、エタン酸、アラニン、ベータ-アラニン、3-アミノプロパン酸、4-アミノブタン酸、5-ペンタン酸など）。ある態様において、リンカーはアミノヘキサン酸（Ahx）の単量体、二量体、またはポリマーを含む。ある態様において、リンカーは炭素環式部分（例えばシクロペンタン、シクロヘキサン）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。他の態様において、リンカーはアミノ酸を含む。ある態様において、リンカーはペプチドを含む。ある態様において、リンカーはアリールまたはヘテロアリール部分を含む。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）の取り付けを容易化するための官能化された部分を包含し得る。いずれかの求電子剤がリンカーの一部として用いられ得る。例示的な求電子剤は、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、ハロゲン化アリール、ハロゲン化アシル、およびイソチオシアネートを包含するが、これらに限定されない。
【００３０】
いくつかの態様において、リンカーは、アミノ酸配列$(GGGGS)_n$（配列番号607）、$(G)_n$（配列番号608）、$(EAAAK)_n$（配列番号609）、$(GGS)_n$（配列番号610）、$(SGGS)_n$（配列番号606）、SGSETPGTSESATPES（配列番号604）、$(XP)_n$（配列番号611）、S

GGS(GGS)$_n$（配列番号612）、SGGSSGGSSGSETPGTSESATPESSGGSSGGS（配列番号605）、またはそれらのいずれかの組み合わせを含み、nは独立して1および30の間の整数であり、Xはいずれかのアミノ酸である。いくつかの態様において、リンカーはアミノ酸配列(GGS)$_n$（配列番号610）を含み、nは1、3、または7である。いくつかの態様において、リンカーはアミノ酸配列SGGS(GGS)$_n$（配列番号612）を含み、nは2である。いくつかの態様において、リンカーはアミノ酸配列SGSETPGTSESATPES（配列番号604）を含む。いくつかの態様において、リンカーはアミノ酸配列SGGSSGGSSGSETPGTSESATPESSGGSSGGS（配列番号605）を含む。

【００３１】

いくつかの態様において、融合蛋白質は、構造[核酸編集ドメイン]-[任意のリンカー配列]-[napDNAbp]-[任意のリンカー配列]-[UGI]を含む。いくつかの態様において、融合蛋白質は、構造[核酸編集ドメイン]-[任意のリンカー配列]-[UGI]-[任意のリンカー配列]-[napDNAbp]；[UGI]-[任意のリンカー配列]-[核酸編集ドメイン]-[任意のリンカー配列]-[napDNAbp]；[UGI]-[任意のリンカー配列]-[napDNAbp]-[任意のリンカー配列]-[核酸編集ドメイン]；[napDNAbp]-[任意のリンカー配列]-[UGI]-[任意のリンカー配列]-[核酸編集ドメイン]；[napDNAbp]-[任意のリンカー配列]-[核酸編集ドメイン]-[任意のリンカー配列]-[UGI]；または[核酸編集ドメイン]-[任意のリンカー配列]-[napDNAbp]-[任意のリンカー配列]-[第1のUGI]-[任意のリンカー配列]-[第2のUGI]を含む。

【００３２】

いくつかの態様において、核酸編集ドメインはデアミナーゼを含む。いくつかの態様において、核酸編集ドメインはデアミナーゼを含む。いくつかの態様において、デアミナーゼはシチジンデアミナーゼである。いくつかの態様において、デアミナーゼはアポリポ蛋白質B mRNA編集複合体（APOBEC）ファミリーデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC1デアミナーゼ、APOBEC2デアミナーゼ、APOBEC3Aデアミナーゼ、APOBEC3Bデアミナーゼ、APOBEC3Cデアミナーゼ、APOBEC3Dデアミナーゼ、APOBEC3Fデアミナーゼ、APOBEC3Gデアミナーゼ、APOBEC3Hデアミナーゼ、またはAPOBEC4デアミナーゼである。いくつかの態様において、デアミナーゼは活性化誘導デアミナーゼ（AID）である。いくつかの態様において、デアミナーゼはシチジンデアミナーゼ1（CDA1）である。いくつかの態様において、デアミナーゼはヤツメウナギCDA1（pmCDA1）デアミナーゼである。

【００３３】

いくつかの態様において、デアミナーゼはヒト、チンパンジー、ゴリラ、サル、ウシ(cow)、イヌ、ラット、またはマウスからである。いくつかの態様において、デアミナーゼはヒトからである。いくつかの態様において、デアミナーゼはラットからである。いくつかの態様において、デアミナーゼは、（配列番号76）によって規定されるアミノ酸配列を含むラットAPOBEC1デアミナーゼである。いくつかの態様において、デアミナーゼは、（配列番号74）によって規定されるアミノ酸配列を含むヒトAPOBEC1デアミナーゼである。いくつかの態様において、デアミナーゼはpmCDA1（配列番号81）である。いくつかの態様において、デアミナーゼはヒトAPOBEC3G（配列番号60）である。いくつかの態様において、デアミナーゼは（配列番号82～84）のいずれか1つのヒトAPOBEC3Gバリアントである。いくつかの態様において、デアミナーゼは、配列番号49～84によって規定されるアミノ酸配列のいずれか1つと少なくとも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%同一である。

【００３４】

いくつかの態様において、UGIドメインは、配列番号134と少なくとも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%同一であるアミノ酸配列を含む。いくつかの態様において、UGIドメインは配列番号134によって規定されるアミノ酸配列を含む。

【００３５】
本開示のいくつかの側面は、本明細書において提供されるnapDNAbp融合蛋白質とnapDNAbpに結合したガイドRNAとを含む複合体を提供する。
【００３６】
本開示のいくつかの側面は、本明細書において提供されるnapDNAbp、融合蛋白質、または複合体を用いる方法を提供する。例えば、本開示のいくつかの側面は、DNA分子を（a）本明細書において提供されるnapDNAbpまたは融合蛋白質およびガイドRNAに（ガイドRNAは約15～100ヌクレオチドの長さであり、標的配列に対して相補的である少なくとも10個の一続きのヌクレオチドの配列を含む）；あるいは（b）本明細書において提供されるnapDNAbp、napDNAbp融合蛋白質、またはgRNAとのnapDNAbpもしくはnapDNAbp複合体に接触させることを含む方法を提供する。
【００３７】
本開示のいくつかの側面は、核酸構築物を含むキットを提供し、（a）本明細書において提供されるnapDNAbpまたはnapDNAbp融合蛋白質をコードするヌクレオチド配列と；（b）（a）の配列の発現を駆動する異種プロモーターとを含む。いくつかの態様において、キットはガイドRNAバックボーンをコードする発現構築物をさらに含み、構築物は、ガイドRNAバックボーンへの標的配列と同一または相補的な核酸配列のクローニングを許すように位置したクローニング部位を含む。
【００３８】
本開示のいくつかの側面は、本明細書において提供される融合蛋白質のnapDNAbpをコードするポリヌクレオチドを提供する。本開示のいくつかの側面はかかるポリヌクレオチドを含むベクターを提供する。いくつかの態様において、ベクターはポリヌクレオチドの発現を駆動する異種プロモーターを含む。
【００３９】
本開示のいくつかの側面は、本明細書において提供されるnapDNAbp蛋白質、融合蛋白質、核酸分子、および／またはベクターを含む細胞を提供する。
【００４０】
Cas9ドメイン（例えばCas9、nCas9、またはdCas9）を包含する本明細書において提供される融合蛋白質のいずれかは、本明細書において提供されるnapDNAbpのいずれか、例えばCasX、CasY、Cpf1、C2c1、C2c2、C2c3、またはアルゴノート蛋白質によって置き換えられ得ることは了解されるはずである。
【００４１】
上のレポーターシステムの例示的な態様の記載は、例示目的のみのために提供されており、限定しようとすることを意味されていない。追加のレポーターシステム、例えば上で詳細に記載されている例示的なシステムの変形もまた本開示によって包摂される。
【００４２】
上の概要は、限定しない様式で、本明細書において開示されるテクノロジーの態様、利点、特質、および使用のいくつかを例示することを意味されている。本明細書において開示されるテクノロジーの他の態様、利点、特質、および使用は、発明を実施するための形態、図面、例、および特許請求の範囲から明らかであろう。
【図面の簡単な説明】
【００４３】
【図１】図1は、一本鎖DNA基質に対するデアミナーゼのデアミナーゼ活性を示す。ランダム化されたPAM配列（NNN PAM）を用いる一本鎖DNA基質を陰性対照として用いた。用いられた古典的(Canonical)PAM配列は（NGG PAM）を包含する。
【００４４】
【図２】図2は、一本鎖DNA基質に対するCas9:デアミナーゼ融合蛋白質の活性を示す。
【００４５】
【図３】図3は、Cas9:デアミナーゼ:sgRNA複合体による二本鎖DNA基質結合を例示する。

【００４６】
【図４】図4は、二本鎖DNA脱アミノ化アッセイを例示する。
【００４７】
【図５】図5は、Cas9融合体が二本鎖DNA標的配列の位置3～11（図5の模式図に従って付番される）を標的化し得ることを実証する。上側のゲル：1μM rAPOBEC1-GGS-dCas9、125nM dsDNA、1当量sgRNA。真中のゲル：1μM rAPOBEC1-$(GGS)_3$（配列番号610）-dCas9、125nM dsDNA、1当量sgRNA。下側のゲル：1.85μM rAPOBEC1-XTEN-dCas9、125nM dsDNA、1当量sgRNA。
【００４８】
【図６】図6は、正しいガイドRNA、例えば正しいsgRNAがデアミナーゼ活性に要求されることを実証する。
【００４９】
【図７】図7は、デアミナーゼ-dCas9:sgRNA複合体によるin vivo標的配列の標的DNA結合のメカニズムを例示する。
【００５０】
【図８】図8は、例示的な疾患関連標的配列の首尾良い脱アミノ化を示す。
【００５１】
【図９】図9は、His6-rAPOBEC1-XTEN-dCas9を用いるin vitroC→T編集効率を示す。
【００５２】
【図１０】図10は、HEK293T細胞におけるC→T編集効率がUGIとの融合によって多大に向上することを示す。
【００５３】
【図１１Ａ】図11A～11Cは、NBE1が、特異的なガイドRNAによってプログラムされるC→U変換をin vitroで媒介することを示す。図11A：核酸塩基編集戦略。ガイドRNAによって指定される座位に標的Cを有するDNAはdCas9によって結合され、これはDNA基質の局所的変性を媒介する。テザリングされたAPOBEC1酵素によるシチジン脱アミノ化は標的CをUに変換する。もたらされるG:Uヘテロ二重鎖はDNA複製または修復後にA:T塩基対へと永続的に変換され得る。Uが鋳型DNA鎖内である場合には、それは転写後にG→A変異を含有するRNA転写物をもまたもたらすであろう。
【図１１Ｂ－Ｃ】図11B：およそ5ヌクレオチドの活性ウィンドウを示す脱アミノ化アッセイ。2hに渡る37℃におけるdsDNA基質とのNBE1-sgRNA複合体のインキュベーション後に、5'フルオロフォア標識DNAを単離し、USER酵素（ウラシルDNAグリコシラーゼ・エンドヌクレアーゼVIII）と37℃で1hに渡ってインキュベーションして、いずれかのウラシルの部位におけるDNA切断を誘導した。もたらされたDNAを変性ポリアクリルアミドゲルによって分離し、いずれかのフルオロフォアが連結された鎖を視覚化した。各レーンは、プロトスペーサー内の標的Cの位置に従って、または標的Cが存在しない場合には「-」によって標識される。PAMから遠位の塩基を位置1としてカウントする。図11C：NBE1の配列特異性およびsgRNA依存性を示すデアミナーゼアッセイ。位置7に標的Cを有するDNA基質を、図11Bと同様に、正しいsgRNA、ミスマッチのsgRNA、またはsgRNAなしでどちらかで、NBE1とインキュベーションした。ミスマッチのsgRNAまたはsgRNAなしではC→U編集は観察されない。陽性対照サンプルは位置7に合成的に組み込まれたUを有するDNA配列を含有する。
【００５４】
【図１２Ａ】図12A～12Bは、in vitroの核酸塩基編集効率への配列コンテキストおよび標的C位置の効果を示す。図12A：in vitroの編集効率への標的Cの周辺の配列を変化させることの効果。プロトスペーサー配列5'-TTATTTCGTGGATTTATTTA-3'（配列番号591）のC_7について、標的鎖の80%（両方の鎖からの合計のシーケンシングリード(total sequencing reads)の40%）の脱アミノ化収率を1.0として定めた。位置1～6および8～13における全ての可能な1塩基変異を含有する基質の相対的な脱アミノ化効率が示される。値およびエラーバーは、異なる日に行われた2つ以上の独立した生物学的レプリケートの

平均および標準偏差を反映する。
【図１２Ｂ】図12B：in vitroの編集効率への各NCモチーフの位置効果。右に示される配列において指示される通り、プロトスペーサー内において、各NC標的モチーフを位置1から8まで変えた（PAMは赤で示される。プロトスペーサープラスプロトスペーサーの5'の1塩基もまた示される）。NBE1とのインキュベーション後に付番された標的C位置のそれぞれにTを含有する合計の配列リードのパーセンテージがグラフで示される。in vitroの最大の可能な脱アミノ化収率は合計のシーケンシングリードの50％（標的鎖の100％）であることに注意。値およびエラーバーは、異なる日に行われた2つまたは3つの独立した生物学的レプリケートの平均および標準偏差を反映する。図12Bはそれぞれ上から下まで配列番号619～626を図示する。
【００５５】
【図１３Ａ】図13A～13Cは、ヒト細胞の核酸塩基編集を示す。図13A：核酸塩基編集因子によって標的化される6つの哺乳類細胞ゲノム座位のプロトスペーサーおよびPAM配列。標的Cが、プロトスペーサー内におけるそれらの位置に対応する下付きの数によって指示される。図13Aはそれぞれ上から下まで配列番号127～132を図示する。
【図１３Ｂ】図13B：NBE1、NBE2、またはNBE3および適切なsgRNAを発現するプラスミドによって、HEK293T細胞をトランスフェクションした。トランスフェクションの3日後に、ゲノムDNAを抽出し、6つの座位において高スループットDNAシーケンシングによって分析した。指示される標的位置にTを有する合計のDNAシーケンシングリードのパーセンテージとして定められる細胞性のC→T変換パーセンテージが、NBE1、NBE2、およびNBE3については全ての6つのゲノム座位において、ドナーHDR鋳型によるwtCas9については6つの部位のうちの3つ（EMX1、HEK293部位3、およびHEK293部位4）において示される。値およびエラーバーは、異なる日に行われた3つの独立した生物学的レプリケートの平均および標準偏差を反映する。
【図１３Ｃ】図13C：全ての6つのゲノム座位についてのNBE2およびNBE3によるまたは6つの部位のうちの3つ（EMX1、HEK293部位3、およびHEK293部位4）におけるHDRのためのwtCas9および一本鎖DNA鋳型によるHEK293T細胞の処置後に、方法に記載される通り算出されたインデル形成の頻度が示される。値は、異なる日に行われた少なくとも3つの独立した生物学的レプリケートの平均を反映する。
【００５６】
【図１４】図14A～14Cは、哺乳類細胞の3つの疾患に該当する変異のNBE2およびNBE3によって媒介される修正を示す。各部位について、プロトスペーサーの配列は変異の名称の右に指示され、PAMと変異を担っている塩基とは、プロトスペーサー内のその位置に対応する下付きの数を有する太字によって指示される。各疾患関連アレルの上のアミノ酸配列は、赤色の核酸塩基編集後の修正されたアミノ酸配列と一緒に示される。各配列の真下は、対応する塩基を有する合計のシーケンシングリードのパーセンテージである。NBE2またはNBE3および適切なsgRNAをコードするプラスミドによって、細胞をヌクレオフェクションした。ヌクレオフェクションの2日後に、ゲノムDNAを抽出し、HTSによって分析して病原性変異修正を査定した。図14A：アルツハイマー病関連APOE4アレルは、マウスアストロサイトにおいて、合計のリードの11％（ヌクレオフェクションされたアストロサイトの44％）においてNBE3によってAPOE3'へと変換される。2つの近傍のCもまたTに変換されるが、もたらされる蛋白質の予測される配列（配列番号627）の変化を有さない。図14B：癌関連p53 N239D変異は、変異（配列番号628）についてヘテロ接合である処置されたヒトリンパ腫細胞の11％（ヌクレオフェクションされた細胞の12％）においてNBE2によって修正される。図14C：p53 Y163C変異はヌクレオフェクションされたヒト乳癌細胞の7.6％においてNBE3によって修正される（配列番号629）。
【００５７】
【図１５】図15A～15Dは、核酸塩基編集へのデアミナーゼ-dCas9リンカー長さおよび組成の効果を示す。GGS（図15A）、$(GGS)_3$（配列番号610）（図15B）、XTEN（図15C）、または$(GGS)_7$（配列番号610）（図15D）のデアミナーゼ-Cas9リンカーを有する

核酸塩基編集因子の脱アミノ化ウィンドウを示す、ゲルに基づくデアミナーゼアッセイ。2hに渡る37℃における125μM dsDNA基質との1.85μM編集因子-sgRNA複合体のインキュベーション後に、色素がコンジュゲーションされたDNAを単離し、USER酵素（ウラシルDNAグリコシラーゼ・エンドヌクレアーゼVIII）と37℃で追加の1時間に渡ってインキュベーションして、いずれかのウラシルの部位においてDNAバックボーンを切断した。もたらされたDNAを変性ポリアクリルアミドゲルによって分離し、色素がコンジュゲーションされた鎖をイメージングした。各レーンは、プロトスペーサー内の標的Cの位置に従ってまたは標的Cが存在しない場合には-によって付番される。8Uは、位置8に合成的に組み込まれたUを有する陽性対照配列である。

【００５８】

【図１６】図16A～16Bは、NBE1がin vitroで疾患に該当する変異を修正することができることをす。図16A：7つの疾患に該当する変異のプロトスペーサーおよびPAM配列。各ケースの疾患関連標的Cはプロトスペーサー内のその位置を反映する下付きの数によって指示される。両方のAPOE4 SNPを除く全ての変異では、標的Cは鋳型（非コード）鎖に所在する。図16Aはそれぞれ上から下まで配列番号631～636を図示する。図16B：NBE1とのインキュベーション、DNA単離、およびU含有位置においてDNAを切断するためのUSER酵素とのインキュベーションの前（-）および後（+）の各dsDNAオリゴヌクレオチドを示す、デアミナーゼアッセイ。USER酵素とのプロトスペーサー内の種々の位置にUを含有する合成オリゴヌクレオチドのインキュベーションからの陽性対照レーンは、Uの位置を指示する対応する数によって示される。

【００５９】

【図１７】図17は、NBE1の処理能力を示す。8つの一続きのCを含有する60mer DNAオリゴヌクレオチドのプロトスペーサーおよびPAMが上に示される。オリゴヌクレオチド（125μM）をNBE1（2μM）と2hに渡って37℃でインキュベーションした。DNAを単離し、高スループットシーケンシングによって分析した。観察された最も高頻度の9つの配列について、合計のリードのパーセントが示される。編集された鎖の大多数（＞93%）は、Tに変換された1つよりも多くのCを有する。この図は配列番号309を図示する。

【００６０】

【図１８Ａ－Ｄ】図18A～18Hは、UGIをNBE1に融合してNBE2を創出する効果を示す。図18A：核酸塩基編集因子によって標的化された6つの哺乳類細胞ゲノム座位のプロトスペーサーおよびPAM配列。編集可能なCはプロトスペーサー内のそれらの位置に対応する標識によって指示される。図18Aはそれぞれ上から下まで配列番号127～132を図示する。図18B～18G：NBE1、NBE2、またはNBE1およびUGIならびに適切なsgRNAを発現するプラスミドによって、HEK293T細胞をトランスフェクションした。トランスフェクションの3日後に、ゲノムDNAを抽出し、6つの座位における高スループットDNAシーケンシングによって分析した。指示される標的位置にTを有する合計のDNAシーケンシングリードのパーセンテージとして定められる細胞性のC→T変換パーセンテージが、全ての6つのゲノム座位におけるNBE1、NBE1およびUGI、ならびにNBE2について示される。

【図１８Ｅ－Ｈ】図18B～18G：NBE1、NBE2、またはNBE1およびUGIならびに適切なsgRNAを発現するプラスミドによって、HEK293T細胞をトランスフェクションした。トランスフェクションの3日後に、ゲノムDNAを抽出し、6つの座位における高スループットDNAシーケンシングによって分析した。指示される標的位置にTを有する合計のDNAシーケンシングリードのパーセンテージとして定められる細胞性のC→T変換パーセンテージが、全ての6つのゲノム座位におけるNBE1、NBE1およびUGI、ならびにNBE2について示される。図18H：NBE1、NBE1プラス別個のプラスミド上のUGI、NBE2、および非処置の細胞について、当該プロトスペーサーの周辺の510個のCにおけるC→T変異率が示される。データは、1.5×10^{6}細胞からの3,000,000個のDNAシーケンシングリードの結果を示す。値は、異なる日に執り行われた少なくとも2つの生物学実験の平均を反映する。

【００６１】

【図１９】図19は、U20SおよびHEK293T細胞におけるNBE2の核酸塩基編集効率を示す

。NBE2による細胞性のC→T変換パーセンテージがHEK293T細胞およびU20S細胞の6つの標的ゲノム座位のそれぞれについて示される。HEK293T細胞はlipofectamine 2000を用いてトランスフェクションし、U20S細胞はヌクレオフェクションした。U20Sヌクレオフェクション効率は74%であった。プラスミド送達の3日後に、ゲノムDNAを抽出し、HTSによって6つのゲノム座位における核酸塩基編集について分析した。値およびエラーバーは、異なる日にされた少なくとも2つの生物学実験の平均および標準偏差を反映する。

【００６２】

【図２０】図20は、核酸塩基編集が複数の細胞分裂に渡って存続することを示す。NBE2による細胞性のC→T変換パーセンテージが、細胞を継代する前および後のHEK293T細胞の2つのゲノム座位において表示される。HEK293T細胞はLipofectamine 2000を用いてトランスフェクションした。トランスフェクションの3日後に、細胞を収穫し、半分に分割した。1つの半分をHTS分析に付し、他の半分はおよそ5つの細胞分裂に渡って繁殖することを許し、それから収穫し、HTS分析に付した。

【００６３】

【図２１】図21は、原理的に核酸塩基編集によって修正され得るClinVarからの遺伝子バリアントを示す。ヒト遺伝子多様性およびそれらの対応する表現型のNCBI ClinVarデータベース[68]を、現行の核酸塩基編集テクノロジーによって修正され得る遺伝子疾患について検索した。左にリスト化されている逐次的制限を課すことによって、結果をフィルタリングした。x軸は、その制限および全ての上の制限を満足させる事象の数を対数スケールで示す。

【００６４】

【図２２】図22は、6つのゲノム座位の編集可能なCのin vitro同定を示す。6つの異なるゲノム部位にマッチする配列を有する合成80merをNBE1とインキュベーションし、それからHTSによって核酸塩基編集について分析した。各部位について、プロトスペーサーの配列は部位の名称の右に指示され、PAMは赤色で強調される。各配列の真下は、対応する塩基を有する合計のDNAシーケンシングリードのパーセンテージである。in vitro変換効率が＞10%である場合には、標的Cは「編集可能」であると考えられた。非標的鎖は核酸塩基編集の基質ではないので、最大収率は合計のDNAシーケンシングリードの50%であることに注意。この図はそれぞれ上から下まで配列番号127～132を図示する。

【００６５】

【図２３】図23は、EMX1オフターゲットにおけるNBE1、NBE2、およびNBE3の活性を示す。Lipofectamine 2000を用いて、NBE1、NBE2、またはNBE3およびEMX1配列にマッチするsgRNAを発現するプラスミドによって、HEK293T細胞をトランスフェクションした。トランスフェクションの3日後に、ゲノムDNAを抽出し、PCRによって増幅し、GUIDE-seq法を用いて先に決定されたEMX1 sgRNAのオンターゲット座位プラス上位10個の公知のCas9オフターゲット座位における高スループットDNAシーケンシングによって分析した[55]。EMX1オフターゲット5座位は増幅せず、示されていない。オンターゲットおよびオフターゲットプロトスペーサーおよびプロトスペーサー隣接モチーフ（PAM）の配列が表示される。プロトスペーサー内の元々のCの各位置にTを有する合計のDNAシーケンシングリードのパーセンテージとして定められる細胞性のC→T変換パーセンテージが、NBE1、NBE2、およびNBE3について示される。一番右には、各配列について報告されたシーケンシングリードの総数が表示される。この図はそれぞれ上から下まで配列番号127および637～645を図示する。

【００６６】

【図２４】図24は、FANCFオフターゲットにおけるNBE1、NBE2、およびNBE3の活性を示す。Lipofectamine 2000を用いて、NBE1、NBE2、またはNBE3およびFANCF配列にマッチするsgRNAを発現するプラスミドによって、HEK293T細胞をトランスフェクションした。トランスフェクションの3日後に、ゲノムDNAを抽出し、PCRによって増幅し、GUIDE-seq法を用いて先に決定されたFANCF sgRNAのオンターゲット座位プラス公知のCas9オフターゲット座位の全てにおける高スループットDNAシーケンシングによって

分析した[55]。オンターゲットおよびオフターゲットプロトスペーサーおよびプロトスペーサー隣接モチーフ（PAM）の配列が表示される。プロトスペーサー内の元々のCの各位置にTを有する合計のDNAシーケンシングリードのパーセンテージとして定められる細胞性のC→T変換パーセンテージが、NBE1、NBE2、およびNBE3について示される。一番右には、各配列について報告されたシーケンシングリードの総数が表示される。この図はそれぞれ上から下まで配列番号128および646～653を図示する。

【００６７】

【図２５】図25は、HEK293部位2オフターゲットにおけるNBE1、NBE2、およびNBE3の活性を示す。Lipofectamine 2000を用いて、NBE1、NBE2、またはNBE3およびHEK293部位2配列にマッチするsgRNAを発現するプラスミドによって、HEK293T細胞をトランスフェクションした。トランスフェクションの3日後に、ゲノムDNAを抽出し、PCRによって増幅し、GUIDE-seq法を用いて先に決定されたHEK293部位2 sgRNAのオンターゲット座位プラス公知のCas9オフターゲット座位の全てにおける高スループットDNAシーケンシングによって分析した[55]。オンターゲットおよびオフターゲットプロトスペーサーおよびプロトスペーサー隣接モチーフ（PAM）の配列が表示される。プロトスペーサー内の元々のCの各位置にTを有する合計のDNAシーケンシングリードのパーセンテージとして定められる細胞性のC→T変換パーセンテージが、NBE1、NBE2、およびNBE3について示される。一番右には、各配列について報告されたシーケンシングリードの総数が表示される。この図はそれぞれ上から下まで配列番号129、654、および655を図示する。

【００６８】

【図２６】図26は、HEK293部位3オフターゲットにおけるNBE1、NBE2、およびNBE3の活性を示す。Lipofectamine 2000を用いて、NBE1、NBE2、またはNBE3およびHEK293部位3配列にマッチするsgRNAを発現するプラスミドによって、HEK293T細胞をトランスフェクションした。トランスフェクションの3日後に、ゲノムDNAを抽出し、PCRによって増幅し、GUIDE-seq法を用いて先に決定されたHEK293部位3 sgRNAのオンターゲット座位プラス公知のCas9オフターゲット座位の全てにおける高スループットDNAシーケンシングによって分析した[55]。オンターゲットおよびオフターゲットプロトスペーサーおよびプロトスペーサー隣接モチーフ（PAM）の配列が表示される。プロトスペーサー内の元々のCの各位置にTを有する合計のDNAシーケンシングリードのパーセンテージとして定められる細胞性のC→T変換パーセンテージが、NBE1、NBE2、およびNBE3について示される。一番右には、各配列について報告されたシーケンシングリードの総数が表示される。この図はそれぞれ上から下まで配列番号130および656～660を図示する。

【００６９】

【図２７】図27は、HEK293部位4オフターゲットにおけるNBE1、NBE2、およびNBE3の活性を示す。Lipofectamine 2000を用いて、NBE1、NBE2、またはNBE3およびHEK293部位4配列にマッチするsgRNAを発現するプラスミドによって、HEK293T細胞をトランスフェクションした。トランスフェクションの3日後に、ゲノムDNAを抽出し、PCRによって増幅し、GUIDE-seq法を用いて先に決定されたHEK293部位4 sgRNAのオンターゲット座位プラス上位10個の公知のCas9オフターゲット座位における高スループットDNAシーケンシングによって分析した[55]。オンターゲットおよびオフターゲットプロトスペーサーおよびプロトスペーサー隣接モチーフ（PAM）の配列が表示される。プロトスペーサー内の元々のCの各位置にTを有する合計のDNAシーケンシングリードのパーセンテージとして定められる細胞性のC→T変換パーセンテージが、NBE1、NBE2、およびNBE3について示される。一番右には、各配列について報告されたシーケンシングリードの総数が表示される。この図はそれぞれ上から下まで配列番号131および661～670を図示する。

【００７０】

【図２８】図28は、非標的C変異率を示す。ここでは、試験された6つのオンターゲットおよび34個のオフターゲット座位の周辺の2,500個の別々のシトシンにおけるC→T変異率が示され、およそ1.8×10^{6}細胞に由来する合計で14,700,000個の配列リードに相当する。

【００７１】
【図２９Ａ】図29A～29Cは、ヒト細胞の塩基編集を示す。図29Aは哺乳類細胞における可能な塩基編集成果を示す。最初の編集はU:Gミスマッチをもたらした。ウラシルDNAグリコシラーゼ（UDG）によるUの認識および切り出しは塩基除去修復（BER）を開始した。これはC:G出発状態への復帰に至る。BERはUDGを阻害するBE2およびBE3によって妨げられた。U:Gミスマッチはミスマッチ修復（ＭＭＲ）によってもまたプロセシングされ、これはミスマッチのニックが入った鎖(nicked strand)を選好的に修復した。BE3はGを含有する編集されない鎖にニックを入れ(nicked)、U:Gミスマッチから所望のU:AまたはT:A成果への解消を選好した。
【図２９Ｂ】図29Bは、下の例の材料と方法に記載される通り処置されたHEK293T細胞を示す。指示される標的位置にTを有する合計のDNAシーケンシングリードのパーセンテージは、BE1、BE2、もしくはBE3による処置、またはドナーHDR鋳型とのwtCas9による処置を（for）示す。
【図２９Ｃ】図29Cは、図29Bの処置後のインデル形成の頻度を示す。値は図34にリスト化されている。図29Bおよび29Cでは、値およびエラーバーは異なる日に行われた3つの独立した生物学的レプリケートの平均およびs.d.を反映する。
【００７２】
【図３０Ａ】図30A～30Bは、哺乳類細胞の2つの疾患に該当する変異のBE3によって媒介される修正を示す。プロトスペーサーの配列は変異の右に示され、PAMおよび標的塩基は赤色であり、プロトスペーサー内のその位置を指示する下付きの数を有する。各配列の真下は、対応する塩基を有する合計のシーケンシングリードのパーセンテージである。細胞は材料と方法に記載される通りに処置した。図30Aは、マウスアストロサイトにおいて合計のリードの74.9%においてBE3によってAPOE3rに変換されたアルツハイマー病関連APOE4アレルを示す。2つの近傍のCもまたTに変換されたが、もたらされる蛋白質の予測される配列の変化を有さなかった。wtCas9およびドナーssDNAによるこれらの細胞の同一の処置は0.3%のみの修正をもたらし、26.1%のインデル形成を有する。この図は配列番号671および627を図示する。
【図３０Ｂ】図30Bは、ヌクレオフェクションされたヒト乳癌細胞の7.6%においてBE3によって修正された癌関連p53 Y163C変異を示し、0.7%のインデル形成を有する。wtCas9およびドナーssDNAによるこれらの細胞の同一の処置は変異修正をもたらさず、6.1%のインデル形成を有する。この図は配列番号672および629を図示する。
【００７３】
【図３１】図31は、HEK293部位2オフターゲットにおけるBE1、BE2、およびBE3の活性を示す。Lipofectamine 2000を用いて、BE1、BE2、またはBE3およびHEK293部位2配列にマッチするsgRNAを発現するプラスミドによって、HEK293T細胞をトランスフェクションした。トランスフェクションの3日後に、ゲノムDNAを抽出し、PCRによって増幅し、GUIDE-seq法を用いてJoungら（63）およびクロマチン免疫沈降高スループットシーケンシング（ChIP-seq）実験を用いてAdliら（18）によって先に決定されたHEK293部位2 sgRNAのオンターゲット座位プラス公知のCas9およびdCas9オフターゲット座位の全てにおける高スループットDNAシーケンシングによって分析した。オンターゲットおよびオフターゲットプロトスペーサーおよびプロトスペーサー隣接モチーフ（PAM）の配列が表示される。プロトスペーサー内の元々のCの各位置にTを有する合計のDNAシーケンシングリードのパーセンテージとして定められる細胞性のC→T変換パーセンテージが、BE1、BE2、およびBE3について示される。一番右には、報告されたシーケンシングリードの総数および報告されたChIP-seqシグナル強度が各配列について表示される。この図はそれぞれ上から下まで配列番号129、654、655、および673～677を図示する。
【００７４】
【図３２】図32は、HEK293部位3オフターゲットにおけるBE1、BE2、およびBE3の活性を示す。Lipofectamine 2000を用いて、BE1、BE2、またはBE3およびHEK293部位3 配列にマッチするsgRNAを発現するプラスミドによって、HEK293T細胞をトランスフ

ェクションした。トランスフェクションの3日後に、ゲノムDNAを抽出し、PCRによって増幅し、JoungらによってGUIDE-seq法を用いて[54]およびクロマチン免疫沈降高スループットシーケンシング（ChIP-seq）実験を用いて[61]先に決定されたHEK293部位3 sgRNAのオンターゲット座位プラス公知のCas9オフターゲット座位および上位5つの公知のdCas9オフターゲット座位の全てにおける高スループットDNAシーケンシングによって分析した。オンターゲットおよびオフターゲットプロトスペーサーおよびプロトスペーサー隣接モチーフ（PAM）の配列が表示される。プロトスペーサー内の元々のCの各位置にTを有する合計のDNAシーケンシングリードのパーセンテージとして定められる細胞性のC→T変換パーセンテージが、BE1、BE2、およびBE3について示される。一番右には、報告されたシーケンシングリードの総数および報告されたChIP-seqシグナル強度が各配列について表示される。この図はそれぞれ上から下まで配列番号130、656～660、および678～682を図示する。

【００７５】

【図３３】図33は、HEK293部位4オフターゲットにおけるBE1、BE2、およびBE3の活性を示す。Lipofectamine 2000を用いて、BE1、BE2、またはBE3およびHEK293部位4配列にマッチするsgRNAを発現するプラスミドによって、HEK293T細胞をトランスフェクションした。トランスフェクションの3日後に、ゲノムDNAを抽出し、PCRによって増幅し、GUIDE-seq法を用いて[54]およびクロマチン免疫沈降高スループットシーケンシング（ChIP-seq）実験を用いて[61]先に決定されたHEK293部位4 sgRNAのオンターゲット座位プラス上位10個の公知のCas9オフターゲット座位および上位5つの公知のdCas9オフターゲット座位における高スループットDNAシーケンシングによって分析した。オンターゲットおよびオフターゲットプロトスペーサーおよびプロトスペーサー隣接モチーフ（PAM）の配列が表示される。プロトスペーサー内の元々のCの各位置にTを有する合計のDNAシーケンシングリードのパーセンテージとして定められる細胞性のC→T変換パーセンテージが、BE1、BE2、およびBE3について示される。一番右には、報告されたシーケンシングリードの総数および報告されたChIP-seqシグナル強度が各配列について表示される。この図はそれぞれ上から下まで配列番号131、661から670、683、および684を図示する。

【００７６】

【図３４】図34は、マウスアストロサイトにおけるアルツハイマー病関連APOE4アレルからAPOE3rへのBE3によって媒介される修正後の非プロトスペーサー塩基の変異率を示す。図30Aおよび図34Bからのプロトスペーサーの両側の50塩基のDNA配列が、プロトスペーサーに対する相対的な各塩基の位置によって示される。PAMから遠位のプロトスペーサーの側は正の数によって指名され、PAMを包含する側は、PAMと共に、負の数によって指名される。各配列の真下は、非処置の細胞、BE3およびAPOE4 C158R変異を標的化するsgRNAによって処置された細胞、またはBE3およびVEGFA座位を標的化するsgRNAによって処置された細胞について、対応する塩基を有する合計のDNAシーケンシングリードのパーセンテージである。BE3によって処置されたサンプルはどちらも非処置の対照のものよりも上の変異率をもたらさなかった。この図はそれぞれ上から下まで配列番号685～688を図示する。

【００７７】

【図３５】図35は、HCC1954ヒト細胞における癌関連p53 Y163C変異のBE3によって媒介される修正後の非プロトスペーサー塩基の変異率を示す。図30Bおよび図39Bからのプロトスペーサーの両側の50塩基のDNA配列が、プロトスペーサーに対する相対的な各塩基の位置と共に示される。PAMから遠位のプロトスペーサーの側は正の数によって指名され、PAMを包含する側は、PAMと共に、負の数によって指名される。各配列の真下は、非処置の細胞、BE3およびTP53 Y163C変異を標的化するsgRNAによって処置された細胞、またはBE3およびVEGFA座位を標的化するsgRNAによって処置された細胞について、対応する塩基を有する合計のシーケンシングリードのパーセンテージである。BE3によって処置されたサンプルはどちらも非処置の対照のものよりも上の変異率をもたらさなかった

。この図はそれぞれ上から下まで配列番号689～692を図示する。
【００７８】
【図３６Ａ－Ｃ】図36A～36Fは、塩基編集へのデアミナーゼ、リンカー長さ、およびリンカー組成の効果を示す。図36Aはゲルに基づくデアミナーゼアッセイを示しており、rAPOBEC1、pmCDA1、hAID、hAPOBEC3G、rAPOBEC1-GGS-dCas9、rAPOBEC1-$(GGS)_3$（配列番号610）-dCas9、および dCas9-$(GGS)_3$（配列番号610）-rAPOBEC1のssDNAに対する活性を示す。酵素を哺乳類細胞溶解物由来のin vitro転写翻訳システムによって発現し、1.8μMの色素がコンジュゲーションされたssDNAおよびUSER酵素（ウラシルDNAグリコシラーゼ・エンドヌクレアーゼVIII）と37℃で2時間に渡ってインキュベーションした。もたらされたDNAを変性ポリアクリルアミドゲルによって分離し、イメージングした。陽性対照は標的Cと同じ位置に合成的に組み込まれたUを有する配列である。図36Bは、図36C～36Fに用いられた発現および精製された蛋白質のクマシー染色した変性PAGEゲルを示す。図36C～36Fは、ゲルに基づくデアミナーゼアッセイを示しており、GGS（図36C）、$(GGS)_3$（配列番号610）（図36D）、XTEN（図36E）、または$(GGS)_7$（配列番号610）（図36F）のデアミナーゼ-Cas9リンカーを有する塩基編集因子の脱アミノ化ウィンドウを示す。sgRNAと複合体化された1.85μMデアミナーゼ-dCas9融合体の125nM dsDNA基質との37℃における2時間のインキュベーション後に、色素がコンジュゲーションされたDNAを単離し、USER酵素と37℃で1時間に渡ってインキュベーションして、いずれかのウラシルの部位においてDNAバックボーンを切断した。もたらされたDNAを変性ポリアクリルアミドゲルによって分離し、色素がコンジュゲーションされた鎖をイメージングした。各レーンは、プロトスペーサー内の標的Cの位置に従って、または標的Cが存在しない場合には-によって付番される。8Uは位置8に合成的に組み込まれたUを有する陽性対照配列である。
【図３６Ｄ－Ｆ】図36C～36Fは、ゲルに基づくデアミナーゼアッセイを示しており、GGS（図36C）、$(GGS)_3$（配列番号610）（図36D）、XTEN（図36E）、または$(GGS)_7$（配列番号610）（図36F）のデアミナーゼ-Cas9リンカーを有する塩基編集因子の脱アミノ化ウィンドウを示す。sgRNAと複合体化された1.85μMデアミナーゼ-dCas9融合体の125nM dsDNA基質との37℃における2時間のインキュベーション後に、色素がコンジュゲーションされたDNAを単離し、USER酵素と37℃で1時間に渡ってインキュベーションして、いずれかのウラシルの部位においてDNAバックボーンを切断した。もたらされたDNAを変性ポリアクリルアミドゲルによって分離し、色素がコンジュゲーションされた鎖をイメージングした。各レーンは、プロトスペーサー内の標的Cの位置に従って、または標的Cが存在しない場合には-によって付番される。8Uは位置8に合成的に組み込まれたUを有する陽性対照配列である。
【００７９】
【図３７Ａ】図37A～37Cは、BE1塩基編集効率が哺乳類細胞において劇的に減少することを示す。図37A：塩基編集因子によって標的化された6つの哺乳類細胞ゲノム座位のプロトスペーサーおよびPAM配列。標的Cは、プロトスペーサー内のそれらの位置に対応する下付きの数によって赤色で指示される。
【図３７Ｂ】図37Bは、6つの異なるゲノム部位にマッチする配列を有する合成80merをBE1とインキュベーションし、それから塩基編集についてHTSによって分析したことを示す。各部位について、プロトスペーサーの配列がPAMと共に部位の名称の右に指示される。各配列の真下は、対応する塩基を有する合計のDNAシーケンシングリードのパーセンテージである。in vitro変換効率が＞10%である場合には、我々は標的Cを「編集可能」と考えた。非標的鎖はBE1によって影響されないので、最大収率は合計のDNAシーケンシングリードの50%であることに注意。値は単一の実験から示される。
【図３７Ｃ】図37Cは、BE1および適切なsgRNAを発現するプラスミドによって、HEK293T細胞をトランスフェクションしたことを示す。トランスフェクションの3日後に、ゲノムDNAを抽出し、6つの座位における高スループットDNAシーケンシングによって分析した。指示される標的位置にTを有する合計のDNAシーケンシングリードのパーセンテー

ジとして定められる細胞性のC→T変換パーセンテージが、全ての6つのゲノム座位におけるBE1について示される。HEK293T細胞からの全てのデータの値およびエラーバーは、異なる日に行われた3つの独立した生物学的レプリケートの平均および標準偏差を反映する。図37Aはそれぞれ上から下まで配列番号127～132を図示する。図37Bはそれぞれ上から下まで配列番号127～132を図示する。

【００８０】

【図３８】図38は、塩基編集が複数の細胞分裂に渡って存続することを示す。BE2およびBE3による細胞性のC→T変換パーセンテージが、細胞を継代する前および後のHEK293T細胞のHEK293部位3および4について示される。BE2またはBE3およびHEK293部位3または4を標的化するsgRNAを発現するプラスミドによって、HEK293T細胞をヌクレオフェクションした。ヌクレオフェクションの3日後に、細胞を収穫し、半分に分割した。1つの半分はHTS分析に付し、他の半分はおよそ5つの細胞分裂に渡って繁殖することを許し、それから収穫し、HTS分析に付した。値およびエラーバーは、少なくとも2つの生物学実験の平均および標準偏差を反映する。

【００８１】

【図３９Ａ】図39A～39Cは非標的C/G変異率を示す。ここでは、試験された6つのオンターゲットおよび34個のオフターゲット座位の周辺の2,500個の別々のシトシンおよびグアニンにおけるC→TおよびG→A変異率が示されており、およそ1.8×10^{6}細胞に由来する合計で14,700,000個の配列リードに相当する。図39Aおよび39Bは、BE1、BE2、およびBE3による細胞性の非標的C→TおよびG→A変換パーセンテージが、全ての2,500個のシトシン／グアニンについて、プロトスペーサーに対する相対的なそれらの位置に対して個々にプロットされることを示す。PAMから遠位のプロトスペーサーの側は正の数によって指名され、PAMを包含する側は負の数によって指名される。

【図３９Ｂ】図39Aおよび39Bは、BE1、BE2、およびBE3による細胞性の非標的C→TおよびG→A変換パーセンテージが、全ての2,500個のシトシン／グアニンについて、プロトスペーサーに対する相対的なそれらの位置に対して個々にプロットされることを示す。PAMから遠位のプロトスペーサーの側は正の数によって指名され、PAMを包含する側は負の数によって指名される。

【図３９Ｃ】図39Cは、BE1、BE2、およびBE3による平均の細胞性の非標的C→TおよびG→A変換パーセンテージ、ならびに最も高いおよび最も低い個々の変換パーセンテージが示されることを示す。

【００８２】

【図４０Ａ】図40A～40Bは、哺乳類細胞における2つの疾患に該当する変異のBE3によって媒介される修正の追加のデータセットを示す。各部位について、プロトスペーサーの配列は変異の名称の右に指示されており、PAMおよび変異を担っている塩基は赤色の太字で指示されており、プロトスペーサー内のその位置に対応する下付きの数を有する。各疾患関連アレルの上のアミノ酸配列は塩基編集後の修正されたアミノ酸配列と一緒に示される。各配列の真下は、対応する塩基を有する合計のシーケンシングリードのパーセンテージである。BE3および適切なsgRNAをコードするプラスミドによって、細胞をヌクレオフェクションした。ヌクレオフェクションの2日後に、ゲノムDNAをヌクレオフェクションされた細胞から抽出し、HTSによって分析して病原性変異修正を査定した。図40Aは、正しいsgRNAによって処置されたときにのみ、アルツハイマー病関連APOE4アレルがマウスアストロサイトにおいてBE3によって合計のリードの58.3%においてAPOE3rに変換されることを示す。2つの近傍のCもまたTに変換されるが、もたらされる蛋白質の予測される配列の変化を有さない。wtCas9およびドナーssDNAによるこれらの細胞の同一の処置は0.2%の修正をもたらし、26.7%のインデル形成を有する。

【図４０Ｂ】図40Bは、正しいsgRNAによって処置されたときにのみ、癌関連p53 Y163C変異が、ヌクレオフェクションされたヒト乳癌細胞の3.3%においてBE3によって修正されることを示す。wtCas9およびドナーssDNAによるこれらの細胞の同一の処置は検出可能な変異修正をもたらさず、8.0%のインデル形成を有する。図40Aから40Bは配列番号6

71、627、672、および629を図示する。
【００８３】
【図４１】図41は、例示的なUSER（ウラシル特異的切り出し試薬）酵素に基づくアッセイの模式図を示しており、これは一本鎖DNA（ssDNA）基質に対する種々のデアミナーゼの活性を試験するために用いられ得る。
【００８４】
【図４２】図42は、pmCDA-nCas9-UGI-NLS構築物の模式図、ならびに塩基編集因子（rAPOBEC1）および陰性対照（非処置）に対して相対的なHeK-3部位におけるその活性である。この図は配列番号693を図示する。
【００８５】
【図４３】図43は、pmCDA1-XTEN-nCas9-UGI-NLS構築物の模式図、ならびに塩基編集因子（rAPOBEC1）および陰性対照（非処置）に対して相対的なHeK-3部位におけるその活性である。この図は配列番号694を図示す。
【００８６】
【図４４】図44は、シチジンデアミナーゼ（CDA）またはAPOBECを用いてTに変換された標的Cを有する合計のシーケンシングリードのパーセントを示す。
【００８７】
【図４５】図45は、デアミナーゼ（CDA）またはAPOBECを用いてAに変換された標的Cを有する合計のシーケンシングリードのパーセントを示す。
【００８８】
【図４６】図46は、デアミナーゼ（CDA）またはAPOBECを用いてGに変換された標的Cを有する合計のシーケンシングリードのパーセントを示す。
【００８９】
【図４７】図47は、huAPOBEC3G-XTEN-nCas9-UGI-NLS構築物の模式図、ならびに変異形態（huAPOBEC3G*(D316R_D317R)-XTEN-nCas9-UGI-NLS、塩基編集因子（rAPOBEC1）、および陰性対照（非処置）に対して相対的なHeK-2部位におけるその活性である。この図は配列番号695を図示する。
【００９０】
【図４８】図48は、例7のセレクションアッセイに用いられたLacZ構築物の模式図を示す。
【００９１】
【図４９】図49は、異なるプラスミドおよび構築物からの復帰データを示す。
【００９２】
【図５０】図50は、lacZ復帰の検証および復帰したクローンの精製を示す。
【００９３】
【図５１】図51は、例7に用いられた脱アミノ化セレクションプラスミドを図示する模式図である。
【００９４】
【図５２】図52は、クロラムフェニコール復帰アッセイの結果を示す（pmCDA1融合体）。
【００９５】
【図５３Ａ】図53A～53Bは、2つの構築物のDNA修正誘導を実証した。
【図５３Ｂ】図53A～53Bは、2つの構築物のDNA修正誘導を実証した。
【００９６】
【図５４】図54は、クロラムフェニコール復帰アッセイの結果を示す（huAPOBEC3G融合体）。
【００９７】
【図５５】図55は、EMX1オフターゲットにおけるBE3およびHF-BE3の活性を示す。配列は上から下まで配列番号127および637～645に対応する。
【００９８】

【図５６】図56は、BE3およびHF-BE3のオンターゲット塩基編集効率を示す。
【００９９】
【図５７】図57は、変異が様々な程度でシチジン脱アミノ化に影響することを実証するグラフである。それぞれが触媒作用を軽度に棄損する変異の組み合わせは、他のものと比べて1つの位置における選択的脱アミノ化を許す。FANCF部位は$GGAATC_6C_7C_8TTC_{11}TGCAGCACCTGG$（配列番号128）であった。
【０１００】
【図５８】図58は、次世代の塩基編集因子を図示する模式図である。
【０１０１】
【図５９】図59は、Cas9バリアントから作られた新たな塩基編集因子を例示する模式図である。
【０１０２】
【図６０】図60は、異なるNGA PAM部位の塩基編集されたパーセンテージを示す。
【０１０３】
【図６１】図61は、NGCG PAM EMX（VRER BE3）および$C_1TC_3C_4C_5ATC_8AC_{10}ATCAACCGGT$（配列番号696）スペーサーを用いて、シチジンの塩基編集されたパーセンテージを示す。
【０１０４】
【図６２】図62は、異なるNNGRRT PAM部位からもたらされる塩基編集されたパーセンテージを示す。
【０１０５】
【図６３】図63は、異なるNNHRRT PAM部位からもたらされる塩基編集されたパーセンテージを示す。
【０１０６】
【図６４】図64A～64Cは、Cpf1 BE2を用いて異なるTTTN PAM部位からもたらされる塩基編集されたパーセンテージを示す。用いられたスペーサーは：$TTTCCTC_3C_4C_5C_6C_7C_8C_9AC_{11}AGGTAGAACAT$（図64A、配列番号697）、$TTTCC_1C_2TC_4TGTC_8C_9AC_{11}ACCCTCATCCTG$（図64B、配列番号698）、および$TTTCC_1C_2C_3AGTC_7C_8TC_{10}C_{11}AC_{13}AC_{15}C_{16}C_{17}TGAAAC$（図64C、配列番号699）であった。
【０１０７】
【図６５】図65は、シチジンデアミナーゼ点変異導入の速度論的調節によって達成される選択的脱アミノ化を図示する模式図である。
【０１０８】
【図６６】図66は、スペーサー中に複数のシチジンを有する細胞培養によって探られた脱アミノ化ウィンドウへの種々の変異の効果を示すグラフである。用いられたスペーサーは：$TGC_3C_4C_5C_6TC_8C_9C_{10}TC_{12}C_{13}C_{14}TGGCCC$（配列番号700）であった。
【０１０９】
【図６７】図67は、スペーサー中に複数のシチジンを有する細胞培養によって探られた脱アミノ化ウィンドウへの種々の変異の効果を示すグラフである。用いられたスペーサーは：$AGAGC_5C_6C_7C_8C_9C_{10}C_{11}TC_{13}AAAGAGA$（配列番号701）であった。
【０１１０】
【図６８】図68は、限定された数のシチジンを有するFANCF部位への種々の変異の効果を示すグラフである。用いられたスペーサーは：$GGAATC_6C_7C_8TTC_{11}TGCAGCACCTGG$（配列番号128）であった。三重変異体（W90Y,R126E,R132E）は第6の位置のシチジンを選好的に編集することに注意。
【０１１１】
【図６９】図69は、限定された数のシチジンを有するHEK3部位への種々の変異の効果を示すグラフである。用いられたスペーサーは：$GGCC_4C_5AGACTGAGCACGTGATGG$（配列番号702）であった。二重および三重変異体は第4の位置のシチジンと比べて第5の位置のシチジンを選好的に編集することに注意。

【０１１２】
【図７０】図70は、限定された数のシチジンを有するEMX1部位への種々の変異の効果を示すグラフである。用いられたスペーサーは：GAGTC$_5$C$_6$GAGCAGAAGAAGAAGGG（配列番号703）であった。三重変異体は第6ではなく第5の位置のシチジンのみを編集することに注意。
【０１１３】
【図７１】図71は、限定された数のシチジンを有するHEK2部位への種々の変異の効果を示すグラフである。用いられたスペーサーは：GAAC$_4$AC$_6$AAAGCATAGACTGCGGG（配列番号704）であった。
【０１１４】
【図７２】図72は、不死化アストロサイトにおけるBE3および変異W90Y R132Eを含むBE3のオンターゲット塩基編集効率を示す。
【０１１５】
【図７３】図73は、3つのCpf1融合体構築物の模式図を図示する。
【０１１６】
【図７４】図74は、BE3およびHF-BE3のプラスミド送達の比較を示す（EMX1、FANCF、およびRNF2）。
【０１１７】
【図７５】図75は、BE3およびHF-BE3のプラスミド送達の比較を示す（HEK3およびHEK4）。
【０１１８】
【図７６】図76は、全10個の部位におけるEMX-1のオフターゲット編集を示す。この図は配列番号127および637～645を図示する。
【０１１９】
【図７７】図77は、GAGTCCGAGCAGAAGAAGAAG（配列番号705）スペーサーを用いるHEK細胞へのデアミナーゼ蛋白質リポフェクションを示す。EMX-1オンターゲットおよびEMX-1オフターゲット部位2を調べた。
【０１２０】
【図７８】図78は、GGAATCCCTTCTGCAGCACCTGG（配列番号706）スペーサーを用いるHEK細胞へのデアミナーゼ蛋白質リポフェクションを示す。FANCFオンターゲットおよびFANCFオフターゲット部位1を調べた。
【０１２１】
【図７９】図79は、GGCCCAGACTGAGCACGTGA（配列番号707）スペーサーを用いるHEK細胞へのデアミナーゼ蛋白質リポフェクションを示す。HEK-3オンターゲット部位を調べた。
【０１２２】
【図８０】図80は、GGCACTGCGGCTGGAGGTGGGGG（配列番号708）スペーサーを用いるHEK細胞へのデアミナーゼ蛋白質リポフェクションを示す。HEK-4オンターゲット、オフターゲット部位1、部位3、および部位4。
【０１２３】
【図８１】図81は、sgHR_13（GTCAGGTCGAGGGTTCTGTC（配列番号709）スペーサー；C8標的：G51→終止）、sgHR_14（GGGCCGCAGTATCCTCACTC（配列番号710）スペーサー；C7標的；C7標的：Q68→終止）、ならびにsgHR_15（CCGCCAGTCCCAGTACGGGA（配列番号711）スペーサー；C10およびC11は標的である：W239またはW237→終止）についてsgRNA活性のin vitroアッセイの結果を示す。
【０１２４】
【図８２】図82は、sgHR_17（CAACCACTGCTCAAAGATGC（配列番号712）スペーサー；C4およびC5は標的である：W410→終止）、およびsgHR_16（CTTCCAGGATGAGAACACAG（配列番号713）スペーサー；C4およびC5は標的である：W273→終止）についてin vitroアッセイの結果を示す。

【０１２５】
【図８３】図83は、ゼブラフィッシュ胚におけるsgHR_13と複合体化されたBE3蛋白質の直接的なインジェクションを示す。
【０１２６】
【図８４】図84は、ゼブラフィッシュ胚におけるsgHR_16と複合体化されたBE3蛋白質の直接的なインジェクションを示す。
【０１２７】
【図８５】図85は、ゼブラフィッシュ胚におけるsgHR_17と複合体化されたBE3蛋白質の直接的なインジェクションを示す。
【０１２８】
【図８６】図86は、シトシン→チミン変化を作ることができる塩基編集因子を用いてなされ得る例示的な核酸変化を示す。
【０１２９】
【図８７】図87は、1つのAPOEアイソフォーム（例えばAPOE4）をアルツハイマー病の減少したリスクに関連する別のAPOEアイソフォーム（例えばAPOE3r）へと編集するために、塩基編集因子（例えばBE3）がいかに用いられ得るかを実証する、アポリポ蛋白質E（APOE）アイソフォームの例示を示す。
【０１３０】
【図８８】図88は、マウスアストロサイトにおけるAPOE4からAPOE3rへの塩基編集を示す。この図は配列番号671および627を図示する。
【０１３１】
【図８９】図89は、アルギニン残基37における蛋白質の早い短縮を引き起こすためのPRNPの塩基編集を示す。この図は配列番号577および714を図示する。
【０１３２】
【図９０】図90は、UDG（これをUGIが阻害する）をノックアウトすることが、C→T塩基編集の効率のクリーンさを劇的に改善することを示す。
【０１３３】
【図９１】図91は、ニッカーゼを有するがUGIなしの塩基編集因子の使用が成果の混合物に至り、非常に高いインデル率を有することを示す。
【０１３４】
【図９２Ａ－Ｂ】図92A～92Gは、SaBE3、SaKKH-BE3、VQR-BE3、EQR-BE3、およびVRER-BE3が、ヒト細胞において非NGG PAMを含有する標的部位における効率的な塩基編集を媒介することを示す。図92AはS. pyogenesおよびS. aureus Cas9を用いる塩基編集因子アーキテクチャを示す。図92Bは、替わりのまたは緩和されたPAM要件を有する最近特徴付けられたCas9バリアントを示す。
【図９２Ｃ】図92Cおよび92Dは、例12に記載される通りに示される塩基編集因子バリアントによって処置されたHEK293T細胞を示す。指示される標的位置にTに変換されたCを有する合計のDNAシーケンシングリード（トランスフェクションされた細胞の濃縮なし）のパーセンテージが示される。試験された各標的のPAM配列がX軸の下に示される。チャートは、NNGRRT PAMを有するゲノム座位におけるSaBE3およびSaKKH-BE3（図92C）、NNNRRT PAMを有するゲノム座位におけるSaBE3およびSaKKH-BE3（図92D）、NGAG PAM（図92E）およびNGAH PAM（図92F）を有するゲノム座位におけるVQR-BE3およびEQR-BE3、ならびにNGCG PAMを有するゲノム座位におけるVRER-BE3（図92G）の結果を示す。値およびエラーバーは少なくとも2つの生物学的レプリケートの平均および標準偏差を反映する。
【図９２Ｄ】図92Cおよび92Dは、例12に記載される通りに示される塩基編集因子バリアントによって処置されたHEK293T細胞を示す。指示される標的位置にTに変換されたCを有する合計のDNAシーケンシングリード（トランスフェクションされた細胞の濃縮なし）のパーセンテージが示される。試験された各標的のPAM配列がX軸の下に示される。チャートは、NNGRRT PAMを有するゲノム座位におけるSaBE3およびSaKKH-BE3（図92C

）、NNNRRT PAMを有するゲノム座位におけるSaBE3およびSaKKH-BE3（図92D）、NGAG PAM（図92E）およびNGAH PAM（図92F）を有するゲノム座位におけるVQR-BE3およびEQR-BE3、ならびにNGCG PAMを有するゲノム座位におけるVRER-BE3（図92G）の結果を示す。値およびエラーバーは少なくとも2つの生物学的レプリケートの平均および標準偏差を反映する。

【図９２Ｅ－Ｆ】チャートは、NNGRRT PAMを有するゲノム座位におけるSaBE3およびSaKKH-BE3（図92C）、NNNRRT PAMを有するゲノム座位におけるSaBE3およびSaKKH-BE3（図92D）、NGAG PAM（図92E）およびNGAH PAM（図92F）を有するゲノム座位におけるVQR-BE3およびEQR-BE3、ならびにNGCG PAMを有するゲノム座位におけるVRER-BE3（図92G）の結果を示す。

【図９２Ｇ】チャートは、NNGRRT PAMを有するゲノム座位におけるSaBE3およびSaKKH-BE3（図92C）、NNNRRT PAMを有するゲノム座位におけるSaBE3およびSaKKH-BE3（図92D）、NGAG PAM（図92E）およびNGAH PAM（図92F）を有するゲノム座位におけるVQR-BE3およびEQR-BE3、ならびにNGCG PAMを有するゲノム座位におけるVRER-BE3（図92G）の結果を示す。

【０１３５】

【図９３Ａ】図93A～93Cは、シチジンデアミナーゼドメインに変異を有する塩基編集因子が、狭められた編集ウィンドウを見せることを実証する。図93A～93Cは、変異体塩基編集因子および適切なsgRNAを発現するプラスミドによってトランスフェクションされたHEK293T細胞を示す。トランスフェクションの3日後に、ゲノムDNAを抽出し、指示される座位における高スループットDNAシーケンシングによって分析した。指示される標的位置にTに変化したCを有する合計のDNAシーケンシングリード（トランスフェクションされた細胞の濃縮なし）のパーセンテージが、EMXl部位（配列番号721）、HEK293部位3（配列番号719）、FANCF部位（配列番号722）、HEK293部位2（配列番号720）、部位A（配列番号715）、および部位B（配列番号718）座位について示される。図93Aは、塩基編集ウィンドウを狭める一定のシチジンデアミナーゼ変異を例示する。追加の変異の特徴付けについては図98を見よ。

【図９３Ｂ】図93Bは、ゲノム座位の編集ウィンドウ幅を成し遂げるシチジンデアミナーゼ変異の効果を示す。有益な変異同士を組み合わせることは、編集ウィンドウを狭めることへの相加的効果を有する。

【図９３Ｃ】図93Cは、YE1-BE3、YE2-BE3、EE-BE3、およびYEE-BE3が、BE3とは対照的に、一重改変された生成物を圧倒的に生成する塩基編集の生成物分布を成し遂げることを示す。値およびエラーバーは少なくとも2つの生物学的レプリケートの平均および標準偏差を反映する。

【０１３６】

【図９４】図94Aおよび94Bは、原理的にはこの研究において開発された塩基編集因子によって修正され得るClinVarからの遺伝子バリアントを示す。ヒト遺伝子多様性およびそれらの対応する表現型のNCBI ClinVarデータベースを、理論上塩基編集によって修正され得る遺伝子疾患について検索した。図94Aは、変改されたPAM特異性を有する塩基編集因子の使用による、ClinVarデータベースの全ての病原性T→C変異における塩基編集標的化範囲の改善を実証する。白色の画分は、BE3、またはこの研究において開発された5つの改変されたPAM塩基編集因子と一緒のBE3どちらかのPAM要件に基づく、アクセス可能な病原性T→C変異の割合を示す。図94Bは、狭められた活性ウィンドウを有する塩基編集因子の使用による、ClinVarデータベース中の全ての病原性T→C変異における塩基編集の標的化範囲の改善を示す。図93A～93Cに示される通り、BE3は同等の効率で位置4-8のCを編集することが想定された。YEE-BE3はその活性ウィンドウ内においてC5＞C6＞C7＞他のものの選好性によって編集することが想定された。白色の画分は、他のCの同等の編集なしにBE3によって編集され得る（左）または他のCの同等の編集なしにBE3またはYEE-BE3によって編集され得る（右）病原性T→C変異の割合を示す。

【０１３７】

【図９５Ａ】図95A～95Bは、塩基編集ウィンドウ幅への短縮されたガイドRNAの効果を示す。BE3および異なる5'短縮長さのsgRNAを発現するプラスミドによって、HEK293T細胞をトランスフェクションした。処置された細胞は例に記載される通りに分析した。図95Aは、EMX1ゲノム座位内の部位における、プロトスペーサーおよびPAM配列（上、配列番号715）、ならびに指示される標的位置にTを有する合計のDNAシーケンシングリードのパーセンテージとして定められる細胞性のC→T変換パーセンテージを示す。この部位においては、塩基編集ウィンドウは17ntの短縮gRNAの使用によって変改された。

【図９５Ｂ】図95Bは、HEK部位3ならびに部位4ゲノム座位内の部位における、プロトスペーサーおよびPAM配列（上、配列番号715および716）、ならびに指示される標的位置にTを有する合計のDNAシーケンシングリードのパーセンテージとして定められる細胞性のC→T変換パーセンテージを示す。これらの部位では、塩基編集ウィンドウの変化は観察されなかったが、sgRNAが短縮されると全ての基質塩基の編集効率の線形的な減少が注意された。

【０１３８】

【図９６】図96は、塩基編集ウィンドウ幅へのAPOBEC1-Cas9リンカー長さの効果を示す。XTEN、GGS、$(GGS)_3$（配列番号610）、$(GGS)_5$（配列番号610）、または$(GGS)_7$（配列番号610）のrAPOBEC1-Cas9リンカーを有する塩基編集因子およびsgRNAを発現するプラスミドによって、HEK293T細胞をトランスフェクションした。処置された細胞を例に記載される通りに分析した。異なるリンカーを有する種々の塩基編集因子について、指示される標的位置にTを有する合計のDNAシーケンシングリードのパーセンテージとして定められる細胞性のC→T変換パーセンテージが示される。

【０１３９】

【図９７Ａ－Ｂ】図97A～97Cは、塩基編集ウィンドウ幅へのrAPOBEC変異の効果を示す。

【図９７Ｃ】図97Cは、部位Aまたは部位Bどちらかを標的化するsgRNAおよび指示されるBE3点変異体を発現するプラスミドによってトランスフェクションされたHEK293T細胞を示す。処置された細胞は例に記載される通りに分析した。プロトスペーサー内およびプロトスペーサーから3塩基対以内の全てのCが表示され、細胞性のC→T変換パーセンテージが示される。編集効率が半最大値を超えるヌクレオチドの算出された数として定められる「編集ウィンドウ幅」が、全ての試験された変異体について表示される。

【０１４０】

【図９８】図98は、哺乳類細胞における塩基編集の生成物分布へのAPOBEC1変異の効果を示す。BE3またはその変異体および適切なsgRNAを発現するプラスミドによって、HEK293T細胞をトランスフェクションした。処置された細胞は例に記載される通り分析した。指示される標的位置にTを有する合計のDNAシーケンシングリードのパーセンテージとして定められる細胞性のC→T変換パーセンテージが示される（左）。C→T変換を含有する合計のシーケンシングリードのパーセントが右に示される。BE3点変異体は、1つの標的シチジンのみを有する部位、HEK部位4における塩基編集効率に有意には影響しない。

【０１４１】

【図９９】図99は、BE3およびHF-BE3におけるオンターゲット編集プラスミド送達の比較を示す。

【０１４２】

【図１００】図100は、BE3の蛋白質およびプラスミド送達におけるオンターゲット編集の比較を示す。

【０１４３】

【図１０１】図101は、HF-BE3の蛋白質およびプラスミド送達におけるオンターゲット編集の比較を示す。

【０１４４】

【図１０２】図102は、リポフェクションおよびHF変異をインストールすること両方が、オフターゲット脱アミノ化イベントを減少させることを示す。菱形は、オフターゲットが

検出されず、特異性比を100に設定したことを指示する。

【０１４５】

【図１０３】図103は、プロトスペーサー内の偶数位置に置かれたCを有する合成基質のin vitro C→T編集を示す（NNNNTC$_2$TC$_4$TC$_6$TC$_8$TC$_{10}$TC$_{12}$TC$_{14}$TC$_{16}$TC$_{18}$TC$_{20}$NGG、配列番号723）。

【０１４６】

【図１０４】図104は、プロトスペーサー内の奇数位置に置かれたCを有する合成基質のin vitro C→T編集を示す（NNNNTC$_2$TC$_4$TC$_6$TC$_8$TC$_{10}$TC$_{12}$TC$_{14}$TC$_{16}$TC$_{18}$TC$_{20}$NGG、配列番号723）。

【０１４７】

【図１０５】図105は、プラスミドvs.蛋白質送達による塩基編集の特異性比を図示する2つのグラフを包含する。

【０１４８】

【図１０６Ａ】図106A～106Bは、非NGG PAM部位に対するBE3活性を示す。BE3および適切なsgRNAを発現するプラスミドによって、HEK293T細胞をトランスフェクションした。処置された細胞は例に記載される通りに分析した。図106Aは、部位に対するBE3活性がSaBE3またはSaKKH-BE3によって効率的に標的化され得ることを示す。BE3はNAG PAMに対して低いが有意な活性を示す。この図は配列番号728および729を図示する。

【図１０６Ｂ】図106Bは、BE3が、VQR-BE3またはVRER-BE3とは対照的に、NGAまたはNGCG PAMを有する部位において有意に縮減された編集を有することを示す。この図は配列番号730および731を図示する。

【０１４９】

【図１０７Ａ】図107A～107Bは、VQR-BE3およびSaKKH-BE3へのAPOBEC1変異の効果を示す。VQR-BE3、SaKKH-BE3、またはその変異体および適切なsgRNAを発現するプラスミドによって、HEK293T細胞をトランスフェクションした。処置された細胞は下の例に記載される通りに分析した。指示される標的位置にTを有する合計のDNAシーケンシングリードのパーセンテージとして定められる細胞性のC→T変換パーセンテージが示される。図107Aは、ウィンドウ調節変異がVQR-BE3に適用され、NGA PAMによって標的化可能な部位における選択的な塩基編集を可能にし得ることを示す。この図は配列番号732および733を図示する。

【図１０７Ｂ】図107Bは、SaKKH-BE3に適用されたときに、変異が、標的ウィンドウ内の塩基選択性を授けることなく、塩基編集効率の総体的な減少を引き起こすことを示す。この図は配列番号728および734を図示する。

【０１５０】

【図１０８】図108は、ヌクレオチド編集の模式図を示す。次の略語が用いられている：（MMR）-ミスマッチ修復、（BE3ニッカーゼ）-塩基編集因子3を言い、これはCas9ニッカーゼドメインを含む、（UGI）-ウラシルグリコシラーゼ阻害因子、（UDG）-ウラシルDNAグリコシラーゼ、（APOBEC）-APOBECシチジンデアミナーゼを言う。

【０１５１】

【図１０９】図109は、例示的な塩基編集構築物の模式図を示す。塩基編集構築物の構造配置がBE3、BE4-pmCDA1、BE4-hAID、BE4-3G、BE4-N、BE4-SSB、BE4-(GGS)$_3$、BE4-XTEN、BE4-32aa、BE4-2xUGI、およびBE4について示される。リンカーは灰色で示される（XTEN、SGGS（配列番号606）、(GGS)$_3$（配列番号610）、および32aa）。デアミナーゼが示される（rAPOBEC1、pmCDA1、hAID、およびhAPOBEC3G）。ウラシルDNAグリコシラーゼ阻害因子（UGI）が示される。一本鎖DNA結合蛋白質（SSB）は紫色で示される。Cas9ニッカーゼ、dCas9(A840H)は赤色で示される。図109は次の標的配列をもまた示す：EMX1、FANCF、HEK2、HEK3、HEK4、およびRNF2。アミノ酸配列は上から下まで配列番号127～132によって指示される。PAM配列は終わりの3ヌクレオチドである。標的シトシン（C）は付番され、赤色で指示される。

【０１５２】

【図１１０】図110は、EMXl配列GAGTC$_5$CGAGCAGAAGAAGAAGGG（配列番号127）の標的シトシン（C$_5$）に対する、指示される塩基編集構築物（BE3、pmCDA1、hAID、hAPOBEC3G、BE4-N、BE4-SSB、BE4-(GGS)$_3$、BE-XTEN、BE4-32aa、およびBE4-2xUGI）の塩基編集結果を示す。変異させられた標的シトシン（C$_5$）の合計のパーセンテージが、各塩基編集構築物について「C$_5$」の下に指示される。インデルの合計のパーセンテージが、各塩基編集構築物について「インデル」の下に指示される。アデニン（A）、グアニン（G）、またはチミン（T）に変異させられた変異シトシンの割合が各塩基編集構築物について円グラフによって指示される。

【０１５３】

【図１１１】図111は、FANCF配列GGAATCCC$_8$TTCTGCAGCACCTGG（配列番号128）の標的シトシン（C$_8$）に対する、指示される塩基編集構築物（BE3、pmCDA1、hAID、hAPOBEC3G、BE4-N、BE4-SSB、BE4-(GGS)$_3$、BE-XTEN、BE4-32aa、およびBE4-2xUGI）の塩基編集結果を示す。変異させられた標的シトシン（C$_8$）の合計のパーセンテージが各塩基編集構築物について「C$_8$」の下に指示される。インデルの合計のパーセンテージが各塩基編集構築物について「インデル」の下に指示される。アデニン（A）、グアニン（G）、またはチミン（T）に変異させられた変異シトシンの割合が各塩基編集構築物について円グラフによって指示される。

【０１５４】

【図１１２】図112は、HEK2配列GAACAC$_6$AAAGCATAGACTGCGGG（配列番号129）の標的シトシン（C$_6$）に対する、指示される塩基編集構築物（BE3、pmCDA1、hAID、hAPOBEC3G、BE4-N、BE4-SSB、BE4-(GGS)$_3$、BE-XTEN、BE4-32aa、およびBE4-2xUGI）の塩基編集結果を示す。変異させられた標的シトシン（C$_6$）の合計のパーセンテージが各塩基編集構築物について「C$_6$」の下に指示される。インデルの合計のパーセンテージが各塩基編集構築物について「インデル」の下に指示される。アデニン（A）、グアニン（G）、またはチミン（T）に変異させられた変異シトシンの割合が各塩基編集構築物について円グラフによって指示される。

【０１５５】

【図１１３】図113は、HEK3配列GGCCC$_5$AGACTGAGCACGTGATGG（配列番号130）の標的シトシン（C$_5$）に対する、指示される塩基編集構築物（BE3、pmCDA1、hAID、hAPOBEC3G、BE4-N、BE4-SSB、BE4-(GGS)$_3$、BE-XTEN、BE4-32aa、およびBE4-2xUGI）の塩基編集結果を示す。変異させられた標的シトシン（C$_5$）の合計のパーセンテージが各塩基編集構築物について「C$_5$」の下に指示される。インデルの合計のパーセンテージが各塩基編集構築物について「インデル」の下に指示される。アデニン（A）、グアニン（G）、またはチミン（T）に変異させられた変異シトシンの割合が各塩基編集構築物について円グラフによって指示される。

【０１５６】

【図１１４】図114は、HEK4配列GGCAC$_5$TGCGGCTGGAGGTCCGGG（配列番号131）の標的シトシン（C$_5$）に対する、指示される塩基編集構築物（BE3、pmCDA1、hAID、hAPOBEC3G、BE4-N、BE4-SSB、BE4-(GGS)$_3$、BE-XTEN、BE4-32aa、およびBE4-2xUGI）の塩基編集結果を示す。変異させられた標的シトシン（C$_5$）の合計のパーセンテージが各塩基編集構築物について「C$_5$」の下に指示される。インデルの合計のパーセンテージが各塩基編集構築物について「インデル」の下に指示される。アデニン（A）、グアニン（G）、またはチミン（T）に変異させられた変異シトシンの割合が各塩基編集構築物について円グラフによって指示される。

【０１５７】

【図１１５】図115は、RNF2配列GTCATC$_6$TTAGTCATTACCTGAGG（配列番号132）の標的シトシン（C$_6$）に対する、指示される塩基編集構築物（BE3、pmCDA1、hAID、hAPOBEC3G、BE4-N、BE4-SSB、BE4-(GGS)$_3$、BE-XTEN、BE4-32aa、およびBE4-2xUGI）の塩基編集結果を示す。変異させられた標的シトシン（C$_6$）の合計のパーセンテージが各塩基編集構築物について「C$_6$」の下に指示される。インデルの合計のパーセンテ

ージが各塩基編集構築物について「インデル」の下に指示される。アデニン（A）、グアニン（G）、またはチミン（T）に変異させられた変異シトシンの割合が各塩基編集構築物について円グラフによって指示される。

【０１５８】

【図１１６】図116は、標的鎖にニックを入れるるCpf1変異体について試験するために用いられた、例示的な蛍光標識された（Cy3標識された）DNA構築物を示す。1のDNA構築物では、非標的鎖（上の鎖）および標的鎖（下の鎖）両方が蛍光標識される。2のDNA構築物では、非標的鎖（上の鎖）は蛍光標識され、標的鎖（下の鎖）は蛍光標識されない。3のDNA構築物では、非標的鎖（上の鎖）は蛍光標識されず、標的鎖（下の鎖）は蛍光標識される。

【０１５９】

【図１１７】図117は、30分間（30min）または2時間よりも多大（2h+）どちらかの反応時間をかけて図116に示されるDNA構築物の標的および非標的鎖を切断する、種々のCpf1構築物（例えばR836A、R1138A、野生型）の能力を実証するデータを示す。

【０１６０】

【図１１８】図118は、アーキテクチャAPOBEC-AsCpf1(R912A)-UGIを有する塩基編集因子が、標的配列の第7から第11の塩基までのウィンドウを有する（例えば、標的配列FANCF1、FANCF2、HEK3-3、およびHEK3-4の）C残基を編集することができることを実証するデータを示す。BGはバックグラウンド変異レベル（非処置）を指示する。AsCpf1はAsCpf1のみによって処置されたことを指示し（対照）、APOBEC-AsCpf1(R912A)-UGIは、標的鎖を選好的に切るCpf1を含有する塩基編集因子を指示し、APOBEC-AsCpf1(R1225A)-UGIは、非標的鎖を切るCpf1を含有する自滅型の塩基編集因子を指示する。FANCF1、FANCF2、HEK3-3、およびHEK3-4の標的配列は次の通りである： FANCF1 GCGGATGTTCCAATCAGTACGCA（配列番号724） FANCF2 CGAGCTTCTGGCGGTCTCAAGCA（配列番号725） HEK3-3 TGCTTCTCCAGCCCTGGCCTGG（配列番号726） HEK3-4 AGACTGAGCACGTGATGGCAGAG（配列番号727）

【０１６１】

【図１１９－１２１】図119は、Cpf1蛋白質（例えばAsCpf1またはLbCpf1）を含む塩基編集因子の模式図を示す。異なるリンカー配列（例えば、XTEN、GGS、$(GGS)_3$（配列番号610）、$(GGS)_5$（配列番号610）、および$(GGS)_7$（配列番号610））を、「リンカー」と標識される部分として試験した。これの結果は図120に示される。

【０１６２】

図120は、HEK3部位TGCTTCTC$_8$CAGCCCTGGCCTGG（配列番号592）のC$_8$残基を編集する、図119に示される構築物の能力を実証するデータを示す。APOBECドメインをCpf1ドメイン（例えばLbCpf1(R836A)またはAsCpf1(R912A)）に連結する異なるリンカー配列を試験した。試験された例示的なリンカーはXTEN、GGS、$(GGS)_3$（配列番号610）、$(GGS)_5$（配列番号610）、および$(GGS)_7$（配列番号610）を包含する。

【０１６３】

図121は、HEK3のTGCTTCTC$_8$C$_9$AGCCCTGGCCTGG（配列番号592）のC$_8$およびC$_9$残基を編集する、LbCpf1ドメインを有する図119に示される構築物の能力を実証するデータを示す。APOBECドメインをLbCpf1ドメインに連結するCentre of Integrative Bioinformatics VUによって維持されるデータベースからの異なるリンカー配列を試験した。試験された例示的なリンカーは1au7、1clk、1c20、1ee8、1flz、1ign、1jmc、1sfe、2ezx、および2rebを包含する。

【０１６４】

【図１２２－１２３】図122は、AsCpf1の構造の模式図を示しており、ここではNおよびC末端が指示される。

【０１６５】

図123は、SpCas9の構造の模式図を示しており、ここではNおよびC末端が指示される。

【０１６６】
【図１２４】図124は、AsCpf1の模式図を示しており、ここでは赤色の円は編集ウィンドウがある予測されるエリアを指示する。四角形はAPOBEC活性を妨害しようとし得るヘリックス状領域を指示する。
【０１６７】
【図１２５】図125Aおよび125Bは、高フィデリティ塩基編集因子（HF-BE3）の操作およびin vitroでの特徴付けを示す。図125AはHF-BE3の模式図を示す。HF-BE3を創出するためにBE3に導入された点変異が示される。図はPDB構造4UN3（Cas9）、4ROV（シチジンデアミナーゼ）、および1UGI（ウラシルDNAグリコシラーゼ阻害因子）を用いた。図125Bは、「TC」反復プロトスペーサーを含有する合成基質のin vitro脱アミノ化を示す。値およびエラーバーは異なる日に行われた2つの独立したレプリケートの平均および範囲を反映する。
【０１６８】
【図１２６Ａ－Ｂ】図126A～126Cは塩基編集因子蛋白質の精製を示す。図126Aは塩基編集因子発現のための最適なE. coli株のセレクションを示す。18℃における16hのIPTGによって誘導された蛋白質発現の後に、粗製の細胞溶解物を蛋白質含量について分析した。BL21 Star（DE3）（Thermo Fisher）細胞はBE3およびHF-BE3両方の最も有望な発現後レベルを示し、塩基編集因子の発現に用いた。図126Bは発現された塩基編集因子蛋白質の精製を示す。His6タグを塩基編集因子のC末端に置くことは、BE3およびHF-BE3両方について短縮生成物の生成に至る（レーン1および2）。意外なことに、この短縮生成物は蛋白質のN末端にHis6タグを置くことによって取り除かれた（レーン3～6）。Cas9発現にとって最適である（OD600=0.4）1よりも遅い、OD600=0.7の細胞密度で塩基編集因子の発現を誘導することは（レーン4～5）、塩基編集因子蛋白質の収率を改善する。精製は、マニュアルHisPur樹脂カラム、次にカチオン交換FPLC（Akta）を用いて行った。
【図１２６Ｃ】図126Cは精製されたBE3およびHF-BE3を示す。異なる濃度の精製されたBE3およびHF-BE3を熱およびLDSを用いて変性させ、ポリアクリルアミドゲルにローディングした。蛋白質サンプルはこの研究に用いられた蛋白質を代表する。図126A～126CのゲルはBOLT Bis-Tris Plus 4-12%ポリアクリルアミド（Thermo Fisher）である。電気泳動および染色は方法に記載される通りに行った。
【０１６９】
【図１２７Ａ】図127A～127Dは、ヒト細胞における高フィデリティ塩基編集因子（HF-BE3）の活性を示す。図127A～127Cは、BE3およびHF-BE3のプラスミドトランスフェクションに関連するオンおよびオフターゲット編集を、非反復的なゲノム座位EMX1（図127A）、FANCF（図127B）、およびHEK293部位3（図127C）を標的化するsgRNAによって処置されたHEK293T細胞からのゲノムDNAの高スループットシーケンシングを用いてアッセイしたことを示す。各sgRNAに関連するオンおよびオフターゲット座位は垂直線によって分離される。
【図１２７Ｂ－Ｃ】図127A～127Cは、BE3およびHF-BE3のプラスミドトランスフェクションに関連するオンおよびオフターゲット編集を、非反復的なゲノム座位EMX1（図127A）、FANCF（図127B）、およびHEK293部位3（図127C）を標的化するsgRNAによって処置されたHEK293T細胞からのゲノムDNAの高スループットシーケンシングを用いてアッセイしたことを示す。各sgRNAに関連するオンおよびオフターゲット座位は垂直線によって分離される。
【図１２７Ｄ】図127Dは、VEGFA部位2を標的化する高度に反復的なsgRNAに関連するオンおよびオフターゲット編集を示す。値およびエラーバーは異なる日に行われた3つの独立した生物学的レプリケートの平均±S.D.を反映する。図127A～127Cでは、星印は、処置されたサンプルおよび非処置の対照の間の比較に基づく有意な編集を指示する。*p≦0.05、**p≦0.01、および***p≦0.001（スチューデントの両側t検定）。図127Dでは、全ての処置されたサンプルは対照に対して相対的に有意な編集を表示したので、アステ

リスクは示されていない。個々のp値は表16にリスト化されている。
【０１７０】
【図１２８Ａ】図128A～128Cは、培養ヒト細胞におけるオンターゲットおよびオフターゲット塩基編集の効率へのBE3蛋白質またはプラスミドの用量の効果を示す。図128Aは、4つのゲノム座位のそれぞれにおけるオンターゲット編集効率が、各sgRNAについて、活性ウィンドウ内の全ての編集されたシトシンについて平均されたことを示す。値およびエラーバーは異なる日に行われた3つの独立した生物学的レプリケートの平均±S.E.Mを反映する。
【図１２８Ｂ－Ｃ】図128Bおよび128Cは、HEK293T細胞においてBE3プラスミドタイトレーション（図128B）またはBE3蛋白質タイトレーション（図128C）から生じたEMX1部位におけるオンおよびオフターゲット編集を示す。値およびエラーバーは異なる日に行われた3つの独立した生物学的レプリケートの平均±S.D.を反映する。
【０１７１】
【図１２９】図129A～129Bは、BE3およびHF-BE3のプラスミドおよび蛋白質送達についてオンターゲット：オフターゲット塩基編集頻度比を示す。塩基編集オンターゲット：オフターゲット特異性比は、活性ウィンドウ内の特定のシトシンのオンターゲット編集パーセンテージを、指示されるオフターゲット座位の対応するシトシンのオフターゲット編集パーセンテージによって割ることによって算出した（方法を見よ）。オフターゲット編集が検出の閾値（シーケンシングリードの0.025%）よりも下であるときには、我々はオフターゲット編集を検出の限界（0.025%）に設定し、オンターゲット編集パーセンテージをこの上限によって割った。◆によって示されるこれらのケースでは、示されている特異性比は下限に相当する。特異性比が、非反復的sgRNA FANCF、HEK293部位3、およびFANCF（図129A）ならびに高度に反復的なsgRNA VEGFA部位2（図129B）について示される。値およびエラーバーは、異なる日に行われた3つの独立した生物学的レプリケートの平均±S.D.を反映する。
【０１７２】
【図１３０Ａ－Ｂ】図130A～130Dは、培養ヒト細胞への塩基編集因子の蛋白質送達を示す。図130A～130Dは、EMX1（図130A）、FANCF（図130B）、HEK293部位3（図130C）、およびVEGFA部位2（図130D）を標的化するsgRNAと複合体化させられた塩基編集因子のRNP送達に関連するオンおよびオフターゲット編集を示す。オフターゲット塩基編集は、非反復的なsgRNAではシーケンシングされた座位の全てにおいて検出不可能であった。値およびエラーバーは異なる日に行われた3つの独立した生物学的レプリケートの平均±S.D.を反映する。星印は、処置されたサンプルおよび非処置の対照の間の比較に基づく有意な編集を指示する。*p≦0.05、**p≦0.01、および***p≦0.001（スチューデントの両側t検定）。
【図１３０Ｃ－Ｄ】図130A～130Dは、EMX1（図130A）、FANCF（図130B）、HEK293部位3（図130C）、およびVEGFA部位2（図130D）を標的化するsgRNAと複合体化させられた塩基編集因子のRNP送達に関連するオンおよびオフターゲット編集を示す。オフターゲット塩基編集は、非反復的なsgRNAではシーケンシングされた座位の全てにおいて検出不可能であった。値およびエラーバーは異なる日に行われた3つの独立した生物学的レプリケートの平均±S.D.を反映する。星印は、処置されたサンプルおよび非処置の対照の間の比較に基づく有意な編集を指示する。*p≦0.05、**p≦0.01、および***p≦0.001（スチューデントの両側t検定）。
【０１７３】
【図１３１Ａ－Ｂ】図131A～131Cは、ゲノム座位における塩基編集に関連するインデル形成を示す。図131Aは、VEGFA部位2、EMX1、FANCF、およびHEK293部位3 sgRNAのオンターゲット座位におけるインデル頻度を示す。図131Bは塩基編集：インデル形成の比を示す。菱形（◆）はインデルが検出されなかったことを指示する（処置されたサンプルおよび非処置の対照においてインデル頻度の有意な違いなし）。
【図１３１Ｃ】図131Cは、図131Aにおいて取り調べられたオンターゲット部位に関連す

るオフターゲット座位において観察されたインデルを示す。値およびエラーバーは、異なる日に行われた3つの独立した生物学的レプリケートの平均±S.D.を反映する。

【０１７４】

【図１３２Ａ】図132A～132Dは、BE3のRNP送達を用いるゼブラフィッシュ胚および生体マウスの内耳におけるDNA不使用のin vivo塩基編集を示す。図132Aは、指示されるsgRNAと複合体化させられたBE3のインジェクションの4日後に収穫されたゼブラフィッシュにおけるオンターゲットゲノム編集を示す。値およびエラーバーは3匹のインジェクションおよび3匹の対照ゼブラフィッシュの平均±S.D.を反映する。対照は非関連sgRNAと複合体化させられたBE3をインジェクションした。

【図１３２Ｂ－Ｃ】図132Bは、カチオン性脂質ナノ粒子に封入されたBE3:sgRNA複合体のin vivoインジェクションを示す模式図を示す。図132Cは、VEGFA部位2ゲノム座位の塩基編集因子ウィンドウ内のシトシン残基の塩基編集を示す。

【図１３２Ｄ】図132Dは、VEGFA部位2標的座位の塩基編集ウィンドウ内の各シトシンにおけるオンターゲット編集を示す。図132D（図132Cおよび132D）は、値およびエラーバーが、VEGFA部位2を標的化するsgRNAを注射された3匹のマウス、3匹の注射されていないマウス、および非関連sgRNAを注射された1匹のマウスの平均±S.E.M.を反映することを示す。

【０１７５】

【図１３３Ａ】図133A～133Eは、マウスNIH/3T3細胞におけるオンおよびオフターゲット塩基編集を示す。図133Aは「VEGFA部位2」sgRNAに関連するオンターゲット塩基編集を示す（配列については図132Eを見よ）。陰性対照は、BE3をコードするプラスミドによってsgRNAなしで処置された細胞に対応する。値およびエラーバーは、異なる日に行われた3つの独立した生物学的レプリケートの平均±S.D.を反映する。

【図１３３Ｂ－Ｃ】図133Bから133Eは、この部位に関連するオフターゲット編集が、このsgRNAの上位4つの予測されるオフターゲット座位における高スループットDNAシーケンシングを用いて測定されたことを示す（配列は図132Eに示される）。図133Bはオフターゲット2を示し、図133Cはオフターゲット1を示し、図133Dはオフターゲット3を示し、図133Eはオフターゲット4を示す。値およびエラーバーは、異なる日に行われた3つの独立した生物学的レプリケートの平均±S.D.を反映する。

【図１３３Ｄ－Ｅ】図133Dはオフターゲット3を示し、図133Eはオフターゲット4を示す。値およびエラーバーは、異なる日に行われた3つの独立した生物学的レプリケートの平均±S.D.を反映する。

【０１７６】

【図１３４】図134A～134Bは、in vivo編集されたマウス組織からのオフターゲット塩基編集およびオンターゲットインデル分析を示す。図134Aは、VEGFA部位2に関連するオフターゲット座位の塩基編集ウィンドウ内の各シトシンについてプロットされた編集を示す。図134Bはオンターゲット塩基編集因子座位におけるインデル率を示す。値およびエラーバーは3匹の注射されたおよび3匹の対照マウスの平均±S.E.Mを反映する。

【０１７７】

【図１３５】図135A～135Cは、塩基編集生成物純度へのUNGをノックアウトすることの効果を示す。図135Aは、例17の材料と方法に記載される通りBE3によって処置されたHAP1（UNG^{+}）およびHAP1 UNG^{-}細胞を示す。編集されたDNAシーケンシングリード（標的Cが変異しているリード）の生成物分布が示される。図135Bは試験されたゲノム座位のプロトスペーサーおよびPAM配列を示しており、図135Aにおいて分析された標的Cは赤色で示される。図135Cは、HAP1細胞またはHAP1 UNG^{-}細胞におけるBE3による処置後のインデル形成の頻度を示す。値およびエラーバーは、異なる日に行われた3つの独立した生物学的レプリケートの平均±S.D.を反映する。

【０１７８】

【図１３６Ａ】図136A～136Dは、生成物純度への複数C塩基編集の効果を示す。図136Aは、非処置の、BE3によって処置された、およびAID-BE3によって処置されたヒトHEK

293T細胞の代表的な高スループットシーケンシングデータを示す。プロトスペーサーの配列が上に示され、PAMおよび標的Cは赤色であり、プロトスペーサー内のそれらの位置を指示する下付きの数を有する。各配列の真下は、対応する塩基を有する合計のシーケンシングリードのパーセンテージである。非T塩基よりもむしろTへとクリーンに編集される標的Cの相対パーセンテージは、この座位の3つのCを編集するAID-BE3によって処置された細胞では、1つのCのみを編集するBE3によって処置された細胞よりもずっと高い。

【図１３６Ｂ－Ｄ】図136Bは、例17の材料と方法に記載される通りBE3、CDA1-BE3、およびAID-BE3によって処置されたHEK293T細胞を示す。編集されたDNAシーケンシングリード（標的Cが変異しているリード）の生成物分布が示される。図136Cは研究されたゲノム座位のプロトスペーサーおよびPAM配列を示し、図136Bにおいて分析される標的Cは赤色で示される。図136Dは図136Aに示される処置後のインデル形成の頻度を示す。値およびエラーバーは異なる日に行われた3つの独立した生物学的レプリケートの平均±S.D.を反映する。

【０１７９】

【図１３７Ａ－Ｂ】図137A～137Cは、BE3のアーキテクチャを変化させることのC→T編集効率および生成物純度への効果を示す。図137Aは、研究されたゲノム座位のプロトスペーサーおよびPAM配列を示し、図137Cの標的Cは紫色および赤色で示され、図137Bの標的Cは赤色で示される。図137Bは、例17の材料と方法に記載される通りBE3、SSB-BE3、N-UGI-BE3、およびBE3-2xUGIによって処置されたHEK293T細胞を示す。編集されたDNAシーケンシングリード（標的Cが変異しているリード）の生成物分布がBE3、N-UGI-BE3、およびBE3-2xUGIについて示される。

【図１３７Ｃ】図137CはC→T塩基編集効率を示す。値およびエラーバーは異なる日に行われた3つの独立した生物学的レプリケートの平均±S.D.を反映する。

【０１８０】

【図１３８Ａ－Ｃ】図138A～138Dは、C→T編集効率および生成物純度へのBE3のリンカー長さの変形の効果を示す。図138AはBE3、BE3C、BE3D、およびBE3Eのアーキテクチャを示す。図138Bは、研究されたゲノム座位のプロトスペーサーおよびPAM配列を示し、図138Cの標的Cは紫色および赤色で示され、図138Dの標的Cは赤色で示される。図138Cは、例17の材料と方法に記載される通りBE3、BE3C、BE3D、またはBE3Eによって処置されたHEK293T細胞を示す。C→T塩基編集効率が示される。

【図１３８Ｄ】図138Dは編集されたDNAシーケンシングリード（標的Cが変異しているリード）の生成物分布をBE3、BE3C、BE3D、およびBE3Eについて示す。値およびエラーバーは異なる日に行われた3つの独立した生物学的レプリケートの平均±S.D.を反映する。

【０１８１】

【図１３９Ａ－Ｃ】図139A～139Dは、BE4がBE3と比較して塩基編集効率および生成物純度を増大させることを示す。図139AはBE3、BE4、およびTarget-AIDのアーキテクチャを示す。図139Bは研究されたゲノム座位のプロトスペーサーおよびPAM配列を示し、図139Cの標的Cは紫色および赤色で示され、図139Dの標的Cは赤色で示される。図139Cは、例17の材料と方法に記載される通りにBE3、BE4、またはTarget-AIDによって処置されたHEK293T細胞を示す。C→T塩基編集効率が示される。

【図１３９Ｄ】図139Dは編集されたDNAシーケンシングリード（標的Cが変異しているリード）の生成物分布をBE3およびBE4について示す。値およびエラーバーは、異なる日に行われた3つの独立した生物学的レプリケートの平均±S.D.を反映する。

【０１８２】

【図１４０】図140A～140Cは、CDA1-BE3およびAID-BE3が標的Gに後続するCをBE3よりも効率的に編集することを示す。図140Aは、研究されたゲノム座位のプロトスペーサーおよびPAM配列を示し、BE3、CDA1-BE3、およびAID-BE3によって編集される標的Cは赤色で示され、CDA1-BE3およびAID-BE3のみによって編集される標的C（Gに後続する）は紫色で示される。図140Bは、例17の材料と方法に記載される通りBE3、CDA1-BE3、AID-BE3、またはAPOBEC3G-BE3によって処置されたHEK293T細胞を示す。C→T

塩基編集効率が示される。図140Cは、HEK2座位を標的化してBE3、CDA1-BE3、またはAID-BE3によって処置されたHEK293T細胞からの、プロトスペーサーの配列に従ってビニングされ、分析された個々のDNAシーケンシングリードを示し、CDA1-BE3およびAID-BE3によるクリーンなC→T編集物を有するシーケンシングリードの＞85%が、Tへと編集された両方のCを有することを明らかにする（図140C）。

【０１８３】

【図１４１Ａ－Ｂ】図141A～141Cは、複数の編集可能なCを有する部位における不均一な編集がより低い生成物純度をもたらすことを示す。図141Aは研究されたゲノム座位のプロトスペーサーおよびPAM配列を示し、図141Cの標的Cは紫色および赤色で示され、図141Bの標的Cは赤色で示される。図141Bおよび141Cは、例17の材料と方法に記載される通りBE3によって処置されたHEK293T細胞を示す。編集されたDNAシーケンシングリード（標的Cが変異しているリード）の生成物分布が示される。同じ座位内の2つのCについて編集効率が不等であるときには、C→非T編集がより高頻度である。値およびエラーバーは異なる日に行われた3つの独立した生物学的レプリケートの平均±S.D.を反映する。

【図１４１Ｃ】図141Bおよび141Cは、例17の材料と方法に記載される通りBE3によって処置されたHEK293T細胞を示す。編集されたDNAシーケンシングリード（標的Cが変異しているリード）の生成物分布が示される。同じ座位内の2つのCについて編集効率が不等であるときには、C→非T編集がより高頻度である。値およびエラーバーは異なる日に行われた3つの独立した生物学的レプリケートの平均±S.D.を反映する。

【０１８４】

【図１４２Ａ－Ｂ】図142A～142Dは、複数のCの塩基編集がより高い塩基編集生成物純度をもたらすことを示す。図142Aは、研究されたゲノム座位のプロトスペーサーおよびPAM配列を示し、図142Bにおいて検討された標的Cは赤色で示される。図142Bは、例17の材料と方法に記載される通りBE3またはBE3B（これはUGIを欠く）によって処置されたHEK293T細胞を示す。編集されたDNAシーケンシングリード（標的Cが変異しているリード）の生成物分布が示される。

【図１４２Ｃ－Ｄ】図142Cは、HEK2座位を標的化してBE3またはBE3B（これはUGIを欠く）によって処置されたHEK293T細胞からのHTSリードを、位置6の一次標的Cのアイデンティティに従ってビニングしたことを示す。それから、もたらされたリードを位置4の二次標的Cの塩基のアイデンティティについて分析した。そのリードに1つの編集イベントのみがあるときには、C_6は非Tへと正しくなく編集されることがより蓋然的である。図142Dは、HEK4座位（1つの編集可能なCのみを有する部位）を標的化してBE3またはBE3Bによって処置された細胞のC_5にA、G、およびTを有する編集されたリードの分布を示し、単一のG:Uミスマッチが UNGによって開始される塩基除去修復によってプロセシングされて生成物の混合物を与えることを例示する。値およびエラーバーは異なる日に行われた3つの独立した生物学的レプリケートの平均±S.D.を反映する。

【０１８５】

【図１４３】図143は、複数のCの塩基編集がHEK3およびRNF2座位におけるより高い塩基編集生成物純度をもたらすことを示す。HEK3およびRNF2座位を標的化してBE3またはBE3B（UGIなし）によって処置されたHEK293T細胞からのDNAシーケンシングリードを、一次標的C位置（赤色）の塩基のアイデンティティに従って分離した。それから、シーケンシングリードの4つの群を二次標的C位置（紫色）の塩基のアイデンティティについて取り調べた。BE3では、一次標的C（赤色）がGへと正しくなく編集されるときには、二次標的CはCのままに留まることがより蓋然的である。逆に、一次標的C（赤色）がTに変換されるときには、二次標的Cもまた同じシーケンシングリードにおいてTへと編集されることが蓋然的である。値およびエラーバーは異なる日に行われた3つの独立した生物学的レプリケートの平均±S.D.を反映する。

【０１８６】

【図１４４】図144A～144Cは、BE4がBE3よりも低いインデル頻度を誘導し、Target-AIDがCDA1-BE3と類似の生成物純度を見せることを示す。図144Aは、例17の材料と方

法に記載される通りBE3、BE4、またはTarget-AIDによって処置されたHEK293T細胞を示す。インデル形成の頻度（例17の材料と方法を見よ）が示される。図144Bは、例17の材料と方法に記載される通りCDA1-BE3またはTarget-AIDによって処置されたHEK293T細胞を示す。編集されたDNAシーケンシングリード（標的Cが変異しているリード）の生成物分布が示される。図144Cは研究されたゲノム座位のプロトスペーサーおよびPAM配列を示しており、図144Bにおいて検討される標的Cは赤色で示される。値およびエラーバーは、異なる日に行われた3つの独立した生物学的レプリケートの平均±S.D.を反映する。

【０１８７】

【図１４５】図145A～145Cは、SaBE4がSaBE3と比較して増大した塩基編集収率および生成物純度を見せることを示す。図145Aは、例17の材料と方法に記載される通りSaBE3およびSaBE4によって処置されたHEK293T細胞を示す。指示される標的位置にTを有する合計のDNAシーケンシングリードのパーセンテージが示される。図145Bは研究されたゲノム座位のプロトスペーサーおよびPAM配列を示しており、図145Aの標的Cは紫色および赤色で示されており、図145Cにおいて検討される標的Cは赤色で示される。図145Cは編集されたDNAシーケンシングリード（標的Cが変異しているリード）の生成物分布を示す。値およびエラーバーは異なる日に行われた3つの独立した生物学的レプリケートの平均±S.D.を反映する。

【０１８８】

【図１４６】図146は、EMX1座位におけるBE3、CDA1-BE3、AID-BE3、またはAPOBEC3G-BE3による処置からの塩基編集成果を示す。プロトスペーサーの配列は上に示されており、PAMおよび標的塩基は赤色であり、プロトスペーサー内のそれらの位置を指示する下付きの数を有する。配列の真下は、対応する塩基を有する合計のシーケンシングリードのパーセンテージである。細胞は例17の材料と方法に記載される通りに処置した。示される値は1つの代表的な実験からである。

【０１８９】

【図１４７】図147は、FANCF座位におけるBE3、CDA1-BE3、AID-BE3、またはAPOBEC3G-BE3による処置からの塩基編集成果を示す。プロトスペーサーの配列は上に示されており、PAMおよび標的塩基は赤色であり、プロトスペーサー内のそれらの位置を指示する下付きの数を有する。配列の真下は、対応する塩基を有する合計のシーケンシングリードのパーセンテージである。細胞は例17の材料と方法に記載される通りに処置した。示される値は1つの代表的な実験からである。

【０１９０】

【図１４８】図148は、HEK2座位におけるBE3、CDA1-BE3、AID-BE3、またはAPOBEC3G-BE3による処置からの塩基編集成果を示す。プロトスペーサーの配列は上に示されており、PAMおよび標的塩基は赤色であり、プロトスペーサー内のそれらの位置を指示する下付きの数を有する。配列の真下は、対応する塩基を有する合計のシーケンシングリードのパーセンテージである。細胞は例17の材料と方法に記載される通りに処置した。示される値は1つの代表的な実験からである。

【０１９１】

【図１４９】図149は、HEK3座位におけるBE3、CDA1-BE3、AID-BE3、またはAPOBEC3G-BE3による処置からの塩基編集成果を示す。プロトスペーサーの配列は上に示されており、PAMおよび標的塩基は赤色であり、プロトスペーサー内のそれらの位置を指示する下付きの数を有する。配列の真下は、対応する塩基を有する合計のシーケンシングリードのパーセンテージである。細胞は例17の材料と方法に記載される通りに処置した。示される値は1つの代表的な実験からである。

【０１９２】

【図１５０】図150は、HEK4座位におけるBE3、CDA1-BE3、AID-BE3、またはAPOBEC3G-BE3による処置からの塩基編集成果を示す。プロトスペーサーの配列は上に示されており、PAMおよび標的塩基は赤色であり、プロトスペーサー内のそれらの位置を指示する下付きの数を有する。配列の真下は、対応する塩基を有する合計のシーケンシングリード

のパーセンテージである。細胞は例17の材料と方法に記載される通りに処置した。示される値は1つの代表的な実験からである。

【０１９３】

【図１５１】図151は、RNF2座位におけるBE3、CDA1-BE3、AID-BE3、またはAPOBEC3G-BE3による処置からの塩基編集成果を示す。プロトスペーサーの配列は上に示されており、PAMおよび標的塩基は赤色であり、プロトスペーサー内のそれらの位置を指示する下付きの数を有する。配列の真下は、対応する塩基を有する合計のシーケンシングリードのパーセンテージである。細胞は例17の材料と方法に記載される通りに処置した。示される値は1つの代表的な実験からである。

【０１９４】

【図１５２】図152は、LBCpf1融合体構築物の模式図を示す。構築物10は[Apobec]-[LbCpf1]-[UGI]-[UGI]のドメイン配置を有する；構築物11は[Apobec]-[LbCpf1]-[UGI]のドメイン配置を有する；構築物12は[UGI]-[Apobec]-[LbCpf1]のドメイン配置を有する；構築物13は[Apobec]-[UGI]-[LbCpf1]のドメイン配置を有する；構築物14は[LbCpf1]-[UGI]-[Apobec]のドメイン配置を有する；構築物15は[LbCpf1]-[Apobec]-[UGI]のドメイン配置を有する。各構築物について、3つの異なるLbCpf1蛋白質を用いた（D/N/A。これはヌクレアーゼデッドLbCpf1（D）；LbCpf1ニッカーゼ（N）、およびヌクレアーゼ活性型LbCpf1（A）を言う）。

【０１９５】

【図１５３】図153は、長さ19ヌクレオチドのガイドすなわちEMX19：TACTTTGTCCTCCGGTTCT（配列番号744）を用いるEMX標的TTTGTAC$_3$TTTGTC$_9$C$_{10}$TC$_{12}$C$_{13}$GGTTC$_{18}$TG（配列番号738）の6つのC残基のC→T編集のパーセンテージを示す。編集は図152に示される構築物のいくつかについて試験した。

【０１９６】

【図１５４】図154は、18ヌクレオチド長のガイドすなわちEMX18：TACTTTGTCCTCCGGTTC（配列番号745）を用いるEMX標的TTTGTAC$_3$TTTGTC$_9$C$_{10}$TC$_{12}$C$_{13}$GGTTC$_{18}$TG（配列番号738）の6つのC残基のC→T編集のパーセンテージを示す。編集は図152に示される構築物のいくつかについて試験した。

【０１９７】

【図１５５】図155は、17ヌクレオチド長のガイドすなわちEMX17：TACTTTGTCCTCCGGTT（配列番号746）を用いるEMX標的TTTGTAC$_3$TTTGTC$_9$C$_{10}$TC$_{12}$C$_{13}$GGTTC$_{18}$TG（配列番号738）の6つのC残基のC→T編集のパーセンテージを示す。編集は図152に示される構築物のいくつかについて試験した。

【０１９８】

【図１５６】図156は、23ヌクレオチド長のガイドすなわちHek2_23：CAGCCCGCTGGCCCTGTAAAGGA（配列番号747）を用いるHEK2標的TTTCC$_1$AGC$_4$C$_5$C$_6$GC$_8$TGGC$_{12}$C$_{13}$C$_{14}$TGTAAA（配列番号739）の8つのC残基のC→T編集のパーセンテージを示す。編集は図152に示される構築物のいくつかについて試験した。

【０１９９】

【図１５７】図157は、20ヌクレオチド長のガイドすなわちHek2_20：CAGCCCGCTGGCCCTGTAAA（配列番号748）を用いるHEK2標的TTTCC$_1$AGC$_4$C$_5$C$_6$GC$_8$TGGC$_{12}$C$_{13}$C$_{14}$TGTAAA（配列番号739）の8つのC残基のC→T編集のパーセンテージを示す。編集は図152に示される構築物のいくつかについて試験した。

【０２００】

【図１５８】図158は、19ヌクレオチド長のガイドすなわちHek2_19：CAGCCCGCTGGCCCTGTAA（配列番号749）を用いるHEK2標的TTTCC$_1$AGC$_4$C$_5$C$_6$GC$_8$TGGC$_{12}$C$_{13}$C$_{14}$TGTAAA（配列番号739）の8つのC残基のC→T編集のパーセンテージを示す。編集は図152に示される構築物のいくつかについて試験した。

【０２０１】

【図１５９】図159は、18ヌクレオチド長のガイドすなわちHek2_18：CAGCCCGCTGG

CCCTGTA（配列番号750）を用いるHEK2標的TTTCC$_1$AGC$_4$C$_5$C$_6$GC$_8$TGGC$_{12}$C$_{13}$C$_{14}$TGTAAA（配列番号739）の8つのC残基のC→T編集のパーセンテージを示す。編集は図152に示される構築物のいくつかについて試験した。

【０２０２】

【図１６０】図160は、図153に図示される実験の編集パーセンテージ値（インデルカウントに基づく調整後）およびインデルのパーセンテージを示す。

【０２０３】

【図１６１】図161は、図154に図示される実験の編集パーセンテージ値（インデルカウントに基づく調整後）およびインデルのパーセンテージを示す。

【０２０４】

【図１６２】図162は、図155に図示される実験の編集パーセンテージ値（インデルカウントに基づく調整後）およびインデルのパーセンテージを示す。

【０２０５】

【図１６３】図163は、図156に図示される実験の編集パーセンテージ値（インデルカウントに基づく調整後）およびインデルのパーセンテージを示す。

【０２０６】

【図１６４】図164は、図157に図示される実験の編集パーセンテージ値（インデルカウントに基づく調整後）およびインデルのパーセンテージを示す。

【０２０７】

【図１６５】図165は、図158に図示される実験の編集パーセンテージ値（インデルカウントに基づく調整後）およびインデルのパーセンテージを示す。

【０２０８】

【図１６６】図166は、図159に図示される実験の編集パーセンテージ値（インデルカウントに基づく調整後）およびインデルのパーセンテージを示す。

【０２０９】

定義

コンテキストが明瞭に別様に指示しない限り、本明細書および請求項において用いられる単数形「a」、「an」および「the」は単数および複数の参照を包含する。それゆえに、例えば「薬剤」の参照は単一の薬剤および複数のかかる薬剤を包含する。

【０２１０】

用語「核酸によってプログラム可能なDNA結合蛋白質」または「napDNAbp」は、例えば標的核酸配列にハイブリダイズすることによってnapDNAbpを特異的な核酸配列までガイドするガイド核酸（例えばgRNA）などの核酸（例えばDNAまたはRNA）と結び付く(with associate)蛋白質を言う。例えば、Cas9蛋白質は、ガイドRNAに対して相補的である特異的なDNA配列までCas9蛋白質をガイドするガイドRNAと結び付き得る。いくつかの態様において、napDNAbpはクラス2細菌CRISPR-Casエフェクターである。いくつかの態様において、napDNAbpは、Cas9ドメイン、例えばヌクレアーゼ活性型Cas9、Cas9ニッカーゼ（nCas9）、またはヌクレアーゼ不活性型Cas9（dCas9）である。核酸によってプログラム可能なDNA結合蛋白質の例は、限定なしに、Cas9（例えばdCas9およびnCas9）、CasX、CasY、Cpf1、C2c1、C2c2、C2c3、およびアルゴノートを包含する。しかしながら、核酸によってプログラム可能なDNA結合蛋白質は、RNAに結合する核酸によってプログラム可能な蛋白質をもまた包含することは了解されるべきである。例えば、napDNAbpはnapDNAbpをRNAへとガイドする核酸と結び付けられ得る。他の核酸によってプログラム可能なDNA結合蛋白質もまた本開示の範囲内であるが、それらは本開示では具体的に記載されずにあり得る。

【０２１１】

いくつかの態様において、napDNAbpは「RNAによってプログラム可能なヌクレアーゼ」または「RNAによってガイドされるヌクレアーゼ」である。用語は本明細書においては交換可能に用いられ、切断の標的ではない1つ以上のRNA（単数または複数）と複合体を形成する（例えば、前記RNAに結合するか、または前記RNAと結び付ける）ヌクレアーゼ

を言う。いくつかの態様において、RNAによってプログラム可能なヌクレアーゼは、RNAとの複合体であるときには、ヌクレアーゼ:RNA複合体と言われ得る。典型的には、結合されるRNA（単数または複数）はガイドRNA（gRNA）と言われる。gRNAは2つ以上のRNAの複合体としてまたは1つのRNA分子として存在し得る。1つのRNA分子として存在するgRNAはシングルガイドRNA（sgRNA）と言われ得るが、「gRNA」は単分子としてまたは2つ以上の分子の複合体としてどちらかで存在するガイドRNAを言うためにもまた用いられる。典型的には、単一のRNA種として存在するgRNAは2つのドメインを含む：（1）標的核酸に対して相同性を共有する（すなわち標的へのCas9複合体の結合を導く）ドメイン；および（2）Cas9蛋白質に結合するドメイン。いくつかの態様において、ドメイン（2）はtracrRNAとして公知の配列に対応し、ステムループ構造を含む。いくつかの態様において、ドメイン（2）はJinek et al., Science 337:816-821 (2012)において提供されているtracrRNAと同一または相同である。これの内容全体は参照によって本明細書に組み込まれる。gRNAの他の例（例えば、ドメイン2を包含するもの）は、「Switchable Cas9 Nucleases And Uses Thereof」と題する2013年9月6日出願のU.S.仮特許出願U.S.S.N.61/874,682および「Delivery System For Functional Nucleases」と題する2013年9月6日出願のU.S.仮特許出願U.S.S.N.61/874,746に見出され得る。それぞれの内容全体はそれらの全体が参照によってここに組み込まれる。いくつかの態様において、gRNAは2つ以上のドメイン（1）および（2）を含み、「延長型gRNA」と言われ得る。例えば、延長型gRNAは2つ以上のCas9蛋白質に結合し、本明細書に記載される通り2つ以上の別々の領域において標的核酸に結合するであろう。gRNAは標的部位を相補するヌクレオチド配列を含み、これは前記標的部位へのヌクレアーゼ/RNA複合体の結合を媒介し、ヌクレアーゼ:RNA複合体の配列特異性を提供する。いくつかの態様において、RNAによってプログラム可能なヌクレアーゼは、（CRISPR関連システム(CRISPR-associated system)）Cas9エンドヌクレアーゼ、例えばStreptococcus pyogenesからのCas9（Csn1）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., Najar 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); and "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)を見よ。これらのそれぞれの内容全体は参照によって本明細書に組み込まれる。

【０２１２】

いくつかの態様において、本明細書において提供されるsgRNAのいずれかは、napDNAbpに結合する配列、例えばsgRNAバックボーン配列を含む。例えば、sgRNAは、Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, and Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science, 337, 816-812; Mali P, Esvelt KM, Church GM (2013) Cas9 as a versatile tool for engineering biology. Nature Methods, 10, 957-963; Li JF, Norville JE, Aach J, McCromack M, Zhang D, Bush J, Church GM, and Sheen J (2013) Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. Nature Biotech, 31, 688-691; Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JRJ, Joung JK (2013) Efficient in vivo genome editing using RNA-guided nucleases. Nat Biotechnol, 31, 227-229; Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (20

13) Multiplex genome engineering using CRIPSR/Cas systems. Science, 339, 819-823; Cho SW, Kim S, Kim JM, Kim JS (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol, 31, 230-232; Jinek MJ, East A, Cheng A, Lin S, Ma E, Doudna J (2013) RNA-programmed genome editing in human cells. eLIFE, 2:e00471; DiCarlo JE, Norville JE, Mali P, Rios, Aach J, Church GM (2013) Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucl Acids Res, 41, 4336-4343; Briner AE, Donohoue PD, Gomaa AA, Selle K, Slorach EM, Nye CH, Haurwitz RE, Beisel CL, May AP, and Barrangou R (2014) Guide RNA functional modules direct Cas9 activity and orthogonality. Mol Cell, 56, 333-339に記載されており；これらのそれぞれの内容は参照によって本明細書に組み込まれる。いくつかの態様において、本明細書において提供されるgRNA（例えばsgRNA）のいずれかは、配列番号741のTのそれぞれがウラシル（U）であるGTAATTTCTACTAAGTGTAGAT（配列番号741）、すなわちGUAAUUUCUACUAAGUGUAGAU、または配列GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU-3'（配列番号618）の核酸配列を含む。

【０２１３】

RNAによってプログラム可能なヌクレアーゼ（例えばCas9）はDNA切断部位を標的化するためにRNA:DNAハイブリダイゼーションを用いるので、これらの蛋白質は、原理的には、ガイドRNAによって指定されるいずれかの配列を標的化する能力がある。RNAによってプログラム可能なヌクレアーゼ、例えばCas9を部位特異的切断のために（例えば、ゲノムを改変するために）用いる方法は当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)を見よ；これらのそれぞれの内容全体は参照によって本明細書に組み込まれる）。

【０２１４】

用語「Cas9」または「Cas9ヌクレアーゼ」は、Cas9蛋白質またはその断片を含むRNAによってガイドされるヌクレアーゼを言う（例えば、Cas9の活性な、不活性な、もしくは部分的に活性な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蛋白質）。Cas9ヌクレアーゼは場合によっては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また言われる。CRISPRは可動性遺伝子エレメント（ウイルス、転移エレメント、および接合性プラスミド）に対する防御を提供する適応免疫システムである。CRISPRクラスターは、スペーサー、祖先由来の可動性エレメントに対して相補的な配列を含有し、侵入核酸を標的化する。CRISPRクラスターは転写され、CRISPR RNA（crRNA）へとプロセシングされる。II型CRISPRシステムでは、プレcrRNAの正しいプロセシングは、trans-encoded small RNA（tracrRNA）、内在性リボヌクレアーゼ3（rnc）、およびCas9蛋白質を要求する。tracrRNAは、プレcrRNAのリボヌクレアーゼ3によって支援されるプロセシングのためのガイドとしての用をなす。爾後に、Cas9/crRNA/tracrRNAはスペーサーに対して相補的な線状または環状dsDNA標的をエンド的に切断する。crRNAに対して相補的でない標的鎖は第1にエンド的に切られ、それから3'-5'にエキソ的にトリミングされる。天然には、DNA結合および切断は典型的には蛋白質および両方のRNAを要求する。しかしながら、crRNAおよびtracrRNA両方の側面を単一のRNA種に組み込むように、シングルガイドRN

A（「sgRNA」または単純に「gNRA」）が操作され得る。例えばJinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. Science 337:816-821 (2012)を見よ。これの内容全体は参照によってここに組み込まれる。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、非自己に対して自己を見分けることを助ける。Cas9ヌクレアーゼ配列および構造は当業者に周知である（例えば、"Complete genome sequence of an Ml 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., Najar 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)；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)を見よ。これらのそれぞれの内容全体は参照によって本明細書に組み込まれる）。Cas9オーソログはS. pyogenesおよびS. thermophilusを包含するがこれらに限定されない種々の種において記載されている。追加の好適な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は明らかであろう。かかる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座位からのCas9配列を包含し；これの内容全体は参照によって本明細書に組み込まれる。いくつかの態様において、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、つまりCas9はニッカーゼである。

【０２１５】

ヌクレアーゼを不活性化されたCas9蛋白質は交換可能に（ヌクレアーゼ「デッド」なCas9として）「dCas9」蛋白質と言われ得る。不活性なDNA切断ドメインを有するCas9蛋白質（またはその断片）を創出するための方法は公知である（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 Expression" (2013) Cell. 28；152(5): 1173-83を見よ。これらのそれぞれの内容全体は参照によって本明細書に組み込まれる）。例えば、Cas9のDNA切断ドメインは2つのサブドメインHNHヌクレアーゼサブドメインおよびRuvC1サブドメインを包含することが公知である。HNHサブドメインはgRNAに対して相補的な鎖を切断し、RuvC1サブドメインは非相補鎖を切断する。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サイレンシングし得る。例えば、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)）。いくつかの態様においては、Cas9の断片を含む蛋白質が提供される。例えば、いくつかの態様において、蛋白質は2つのCas9ドメインの1つを含む：（1）Cas9のgRNA結合ドメイン；または（2）Cas9のDNA切断ドメイン。いくつかの態様において、Cas9またはその断片を含む蛋白質は「Cas9バリアント」と言われる。Cas9バリアントはCas9またはその断片に対して相同性を共有する。例えば、Cas9バリアントは野生型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。いくつかの態様において、Cas9バリアントは、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、またはより多くのアミノ酸変化を有し得る。いくつかの態様において、Cas9バリアントは、断片が野生型Cas9の対応する断片と少な

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%同一であるようなCas9の断片（例えば、gRNA結合ドメインまたはDNA切断ドメイン）を含む。いくつかの態様において、断片は、対応する野生型Cas9のアミノ酸長さの少なくとも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。

【０２１６】

いくつかの態様において、断片は長さ少なくとも100アミノ酸である。いくつかの態様において、断片は長さ少なくとも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アミノ酸である。いくつかの態様において、野生型Cas9はStreptococcus pyogenesからのCas9に対応する（NCBI参照配列：NC_017053.1、配列番号1（ヌクレオチド）；配列番号2（アミノ酸））。

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGT
GATCACTGATGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACC
GCCACAGTATCAAAAAAAATCTTATAGGGGCTCTTTTATTTGGCAGTGGAGAGACAGCG
GAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTAT
TTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTC
ATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATT
TTTGGAAATATAGTAGATGAAGTTGCTTATCATGAGAAATATCCAACTATCTATCATCT
GCGAAAAAAATTGGCAGATTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCT
TAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGAT
AATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAATCTACAATCAATTATTTGA
AGAAAACCCTATTAACGCAAGTAGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGA
GTAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAGAAATGGC
TTGTTTGGGAATCTCATTGCTTTGTCATTGGGATTGACCCCTAATTTTAAATCAAATTTT
GATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGA
TAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATT
TATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATAGTGAAATAACTAAGGCT
CCCCTATCAGCTTCAATGATTAAGCGCTACGATGAACATCATCAAGACTTGACTCTTTT
AAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAAT
CAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAA
TTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACTAAA
TCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCAAA
TTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTA
AAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGG
TCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAA
TTACCCCATGGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATT
GAACGCATGACAAACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAG
TTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTG
AGGGAATGCGAAAACCAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTA
CTCTTCAAAACAAATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAA
AATAGAATGTTTTGATAGTGTTGAAATTTCAGGAGTTGAAGATAGATTTAATGCTTCAT
TAGGCGCCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATTTTTTGGATAATGAA
GAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGG
GATGATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAAC
AGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGT

ATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTTTTGAAATCAGATGGTTTTGC
CAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGATATTC
AAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGATTGCTAACTTAGCT
GGCAGTCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAATTGTTGATGAACTGGT
CAAAGTAATGGGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATCAGA
CAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGGTATC
AAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAA
TGAAAAGCTCTATCTCTATTATCTACAAAATGGAAGAGACATGTATGTGGACCAAGAAT
TAGATATTAATCGTTTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCATT
AAAGACGATTCAATAGACAATAAGGTACTAACGCGTTCTGATAAAAATCGTGGTAAATC
GGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACTTC
TAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGA
GGTTTGAGTGAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCA
AATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAA
ATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGAC
TTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCA
TGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAAACTTG
AATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAG
TCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAA
CTTCTTCAAAACAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCG
AAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTG
CGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGG
CGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTA
AAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCA
GTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAATCCGTTAAAGA
GTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATTGACTTTT
TAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAATAT
AGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTACA
AAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAGTC
ATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAG
CAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGT
TATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACA
AACCAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGGA
GCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTAC
AAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACAC
GCATTGATTTGAGTCAGCTAGGAGGTGACTGA（配列番号1）

【化１】

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKR
NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLA
AKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIF
FDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGS
IPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVI
EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVV
KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVA
QILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL

【化２】

NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFF
KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVK
ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGEL
QKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSK
RVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYT
STKEVLDATLIHQSITGLYETRIDLSQLGGD

（配列番号2）（一重下線：HNHドメイン；二重下線：RuvCドメイン）

【０２１７】

いくつかの態様において、野生型Cas9は配列番号3（ヌクレオチド）および／または配列番号4（アミノ酸）に対応するかまたはそれを含む：

ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCCGTTGGATGGGCTGT
CATAACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACC
GTCATTCGATTAAAAAGAATCTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACGGCA
GAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACCGAAT
ATGTTACTTACAAGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTC
ACCGTTTGGAAGAGTCCTTCCTTGTCGAAGAGGACAAGAAACATGAACGGCACCCCATC
TTTGGAAACATAGTAGATGAGGTGGCATATCATGAAAAGTACCCAACGATTTATCACCT
CAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGGACCTGAGGTTAATCTACTTGGCTC
TTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAATCCGGAC
AACTCGGATGTCGACAAACTGTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGA
AGAGAACCCTATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCT
CTAAATCCCGACGGCTAGAAAACCTGATCGCACAATTACCCGGAGAGAAGAAAAATGGG
TTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAATTTTAAGTCGAACTT
CGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCG
ACAATCTACTGGCACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAAC
CTTAGCGATGCAATCCTCCTATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGC
GCCGTTATCCGCTTCAATGATCAAAAGGTACGATGAACATCACCAAGACTTGACACTTC
TCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATATTCTTTGATCAG

TCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAA
GTTTATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAAAACTCA
ATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCGACAACGGTAGCATTCCACATCAA
ATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCGTTCCT
CAAAGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGG
GACCCCTGGCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAACG
ATTACTCCATGGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCAT
CGAGAGGATGACCAACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACA
GTTTACTTTACGAGTATTTCACAGTGTACAATGAACTCACGAAAGTTAAGTATGTCACT
GAGGGCATGCGTAAACCCGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCT
GTTATTCAAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGA
AAATTGAATGCTTCGATTCTGTCGAGATCTCCGGGGTAGAAGATCGATTTAATGCGTCA
CTTGGTACGTATCATGACCTCCTAAAGATAATTAAAGATAAGGACTTCCTGGATAACGA
AGAGAATGAAGATATCTTAGAAGATATAGTGTTGACTCTTACCCTCTTTGAAGATCGGG
AAATGATTGAGGAAAGACTAAAAACATACGCTCACCTGTTCGACGATAAGGTTATGAAA
CAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGATTGTCGCGGAAACTTATCAACGG
GATAAGAGACAAGCAAAGTGGTAAAACTATTCTCGATTTTCTAAAGAGCGACGGCTTCG
CCAATAGGAACTTTATGCAGCTGATCCATGATGACTCTTTAACCTTCAAAGAGGATATA
CAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATATTGCGAATCTTGC
TGGTTCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTGGATGAGCTAG
TTAAGGTCATGGGACGTCACAAACCGGAAAACATTGTAATCGAGATGGCACGCGAAAAT
CAAACGACTCAGAAGGGGCAAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAGAGG
GTATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTG
CAGAACGAGAAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCA
GGAACTGGACATAAACCGTTTATCTGATTACGACGTCGATCACATTGTACCCCAATCCT
TTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCGAGGG
AAAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGGCA
GCTCCTAAATGCGAAACTGATAACGCAAAGAAAGTTCGATAACTTAACTAAAGCTGAGA
GGGGTGGCTTGTCTGAACTTGACAAGGCCGGATTTATTAAACGTCAGCTCGTGGAAACC
CGCCAAATCACAAAGCATGTTGCACAGATACTAGATTCCCGAATGAATACGAAATACGA
CGAGAACGATAAGCTGATTCGGGAAGTCAAAGTAATCACTTTAAAGTCAAAATTGGTGT
CGGACTTCAGAAAGGATTTTCAATTCTATAAAGTTAGGGAGATAAATAACTACCACCAT
GCGCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTAAGAAATACCCGAA
GCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCCGTAAGATGATCG
CGAAAAGCGAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATT
ATGAATTTCTTTAAGACGGAAATCACTCTGGCAAACGGAGAGATACGCAAACGACCTTT
AATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTCGCGA
CGGTGAGAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTGCAG
ACCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGTGATAAGCTCATCGC
TCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCT
ATTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGGAAAATCCAAGAAACTGAAGTCAGTC
AAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAAAGAACCCCATCGA
CTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCATAATTAAACTACCAA
AGTATAGTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAG
CTTCAAAAGGGGAACGAACTCGCACTACCGTCTAAATACGTGAATTTCCTGTATTTAGC
GTCCCATTACGAGAAGTTGAAAGGTTCACCTGAAGATAACGAACAGAAGCAACTTTTTG
TTGAGCAGCACAAACATTATCTCGACGAAATCATAGAGCAAATTTCGGAATTCAGTAAG
AGAGTCATCCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAG
GGATAAACCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACC
TCGGCGCTCCAGCCGCATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACT

TCTACCAAGGAGGTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTATATGA
AACTCGGATAGATTTGTCACAGCTTGGGGGTGACGGATCCCCCAAGAAGAAGAGGAAAG
TCTCGAGCGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAG
GATGACGATGACAAGGCTGCAGGA（配列番号3）

【化３】

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK
NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLA
AKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIF
FDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGS
IPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS

【化４】

LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA
YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

（配列番号4）（一重下線：HNHドメイン；二重下線：RuvCドメイン）

【０２１８】

いくつかの態様において、野生型Cas9はStreptococcus pyogenesからのCas9に対応する（NCBI参照配列：NC_002737.2、配列番号5（ヌクレオチド）；およびUniport参照配列：Q99ZW2、配列番号6（アミノ酸）。

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGT
GATCACTGATGAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACC
GCCACAGTATCAAAAAAAATCTTATAGGGGCTCTTTTATTTGACAGTGGAGAGACAGCG
GAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTAT
TTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTC
ATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATT
TTTGGAAATATAGTAGATGAAGTTGCTTATCATGAGAAATATCCAACTATCTATCATCT
GCGAAAAAAATTGGTAGATTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCT
TAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGAT
AATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAACCTACAATCAATTATTTGA
AGAAAACCCTATTAACGCAAGTGGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGA
GTAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAAAAATGGC
TTATTTGGGAATCTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTTAAATCAAATTTT
GATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGA

TAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATT
TATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATACTGAAATAACTAAGGCT
CCCCTATCAGCTTCAATGATTAAACGCTACGATGAACATCATCAAGACTTGACTCTTTT
AAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAAT
CAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAA
TTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACTAAA
TCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCAAA
TTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTA
AAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGG
TCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAA
TTACCCCATGGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATT
GAACGCATGACAAACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAG
TTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTG
AAGGAATGCGAAAACCAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTA
CTCTTCAAAACAAATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAA
AATAGAATGTTTTGATAGTGTTGAAATTTCAGGAGTTGAAGATAGATTTAATGCTTCAT
TAGGTACCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATTTTTTGGATAATGAA
GAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGA
GATGATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAAC
AGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGT
ATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTTTTGAAATCAGATGGTTTTGC
CAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGACATTC
AAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTACATGAACATATTGCAAATTTAGCT
GGTAGCCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAGTTGTTGATGAATTGGT
CAAAGTAATGGGGCGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATC
AGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGGT
ATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCA
AAATGAAAAGCTCTATCTCTATTATCTCCAAAATGGAAGAGACATGTATGTGGACCAAG
AATTAGATATTAATCGTTTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTC
CTTAAAGACGATTCAATAGACAATAAGGTCTTAACGCGTTCTGATAAAAATCGTGGTAA
ATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAAC
TTCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGT
GGAGGTTTGAGTGAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCG
CCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATG
AAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCT
GACTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATGC
CCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAAAC
TTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAATGATTGCT
AAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCAT
GAACTTCTTCAAAACAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAA
TCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCCACA
GTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGAC
AGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTC
GTAAAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTAT
TCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAATCCGTTAA
AGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATTGACT
TTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAA
TATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATT
ACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTA
GTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTG

GAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCG
TGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAG
ACAAACCAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTT
GGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTC
TACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAA
CACGCATTGATTTGAGTCAGCTAGGAGGTGACTGA（配列番号5）

【化５】

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK
NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLA
AKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIF
FDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGS
IPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA
YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

（配列番号6）（一重下線：HNHドメイン；二重下線：RuvCドメイン）

【０２１９】

いくつかの態様において、Cas9は：Corynebacterium ulcerans（NCBI Ref：NC_015683.1, NC_017317.1）；Corynebacterium diphtheria（NCBI Ref：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 I（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、または例5においてリスト化されている生物のいずれかからのCas9を言う。

【０２２０】

いくつかの態様において、dCas9は、部分的にまたは全体的に、Cas9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れを含む。例えば、いくつかの態様において、dCas9ドメインはD10Aおよび／またはH840A変異

を含む。
dCas9（D10AおよびH840A）：
【化６】

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDS
GETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKH
ERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPG
EKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYAD
LFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYN
ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEI
SGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTY
AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQL
IHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRH
KPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNE
KLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNR
GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVR
EINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKAT
AKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQ
VNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK
VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN
GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF
KYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

（配列番号7）（一重下線：HNHドメイン；二重下線：RuvCドメイン）。
【０２２１】
いくつかの態様において、Cas9ドメインはD10A変異を含むが、位置840の残基は、配列番号6によって提供されるアミノ酸配列において、または別のCas9ドメインによって提供されるアミノ酸配列のいずれか、例えば本明細書において提供されるCas9蛋白質のいずれかの対応する位置において、ヒスチジンのままに留まる。いずれの特定の理論によっても拘束されることを望むものではないが、触媒残基H840の存在は、Cas9の活性を回復させて、標的Cに対向するGを含有する編集されない（例えば脱アミノ化されない）鎖を切断する。（例えばA840からの）H840の回復は、Cを含有する標的鎖の切断をもたらさない。かかるCas9バリアントは、gRNAによって定められる標的配列に基づいて特異的な場所において一本鎖DNA切断（ニック）を創出する能力があり、編集されない鎖の修復に至り、究極的には、編集されない鎖のG→A変化をもたらす。このプロセスの模式図が図108に示されている。簡潔には、C-G塩基対のCは、デアミナーゼ、例えばAPOBECデアミナーゼによってUへと脱アミノ化され得る。Gを有する編集されない鎖にニックを入れることは、ミスマッチ修復メカニズムによるGの取り除きを容易化する。UGIはUの取り除きを防止するUDGを阻害する。
【０２２２】
他の態様においては、D10AおよびH840A以外の変異を有するdCas9バリアントが提供され、これらは例えばヌクレアーゼを不活性化されたCas9（dCas9）をもたらす。例として、かかる変異は、D10およびH820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サブドメインの置換）を包含する。いくつかの態様においては、dCas9のバリアントま

たはホモログ（例えば配列番号6のバリアント）が提供され、これらは配列番号6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。いくつかの態様においては、dCas9のバリアント（例えば配列番号6のバリアント）が提供され、配列番号6よりも約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、またはより多くだけ短いかまたは長いアミノ酸配列を有する。

【０２２３】

いくつかの態様において、本明細書において提供されるCas9融合蛋白質は、Cas9蛋白質、例えば本明細書において提供されるCas9配列の1つの全長アミノ酸配列を含む。しかしながら、他の態様においては、本明細書において提供される融合蛋白質は全長Cas9配列ではなくその断片のみを含む。例えば、いくつかの態様において、本明細書において提供されるCas9融合蛋白質はCas9断片を含み、断片はcrRNAおよびtracrRNAまたはsgRNAに結合するが、例えばそれがヌクレアーゼドメインの短縮版のみを含むかまたは全くヌクレアーゼドメインを含まないことで、機能的なヌクレアーゼドメインを含まない。好適なCas9ドメインおよびCas9断片の例示的なアミノ酸配列が本明細書において提供されており、Cas9ドメインおよび断片の追加の好適な配列は当業者には明らかであろう。

【０２２４】

いくつかの態様において、Cas9は：Corynebacterium ulcerans（NCBI Ref：NC_015683.1、NC_017317.1）；Corynebacterium diphtheria（NCBI Ref：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 I（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）；またはNeisseria. meningitidis（NCBI Ref：YP_002342100.1）からのCas9を言う。

【０２２５】

本明細書において用いられる用語「デアミナーゼ」または「デアミナーゼドメイン」は、脱アミノ化反応を触媒する蛋白質または酵素を言う。いくつかの態様において、デアミナーゼまたはデアミナーゼドメインはシチジンデアミナーゼであり、シチジンまたはデオキシシチジンからそれぞれウリジンまたはデオキシウリジンへの加水分解的な脱アミノ化を触媒する。いくつかの態様において、デアミナーゼまたはデアミナーゼドメインはシチジンデアミナーゼドメインであり、シトシンからウラシルへの加水分解的な脱アミノ化を触媒する。いくつかの態様において、デアミナーゼまたはデアミナーゼドメインは、ヒト、チンパンジー、ゴリラ、サル、牛、イヌ、ラット、またはマウスなどの生物からの天然に生起するデアミナーゼである。いくつかの態様において、デアミナーゼまたはデアミナーゼドメインは生物からの天然に生起するデアミナーゼのバリアントであり、これは天然には生起しない。例えば、いくつかの態様において、デアミナーゼまたはデアミナーゼドメインは、ある生物からの天然に生起するデアミナーゼと少なくとも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%、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%同一である。

【０２２６】

本明細書において用いられる用語「有効量」は、所望の生物学的応答を誘起するために十分である生物学的に活性な薬剤の量を言う。例えば、いくつかの態様において、ヌクレアーゼの有効量は、ヌクレアーゼによって特異的に結合および切断される標的部位の切断を誘導するために十分であるヌクレアーゼの量を言い得る。いくつかの態様において、本

明細書において提供される融合蛋白質、例えばヌクレアーゼ不活性型Cas9ドメインと核酸編集ドメイン（例えばデアミナーゼドメイン）とを含む融合蛋白質の有効量は、融合蛋白質によって特異的に結合および編集される標的部位の編集を誘導するために十分である融合蛋白質の量を言い得る。当業者によって了解されるであろう通り、薬剤、例えば融合蛋白質、ヌクレアーゼ、デアミナーゼ、リコンビナーゼ、ハイブリッド蛋白質、蛋白質二量体、蛋白質（もしくは蛋白質二量体）およびポリヌクレオチドの複合体、またはポリヌクレオチドの有効量は、種々の因子、例えば所望の生物学的応答、例えば編集されるべき特異的なアレル、ゲノム、または標的部位、標的化されようとする細胞または組織、および用いられようとする薬剤に依存して変わり得る。

【０２２７】

本明細書において用いられる用語「リンカー」は、2つの分子または部分、例えば、例えばヌクレアーゼ不活性型Cas9ドメインおよび核酸編集ドメイン（例えばデアミナーゼドメイン）などの融合蛋白質の2つのドメインを連結する、化学基または分子を言う。リンカーは例えばいずれかの長さおよび組成のアミノ酸配列、ペプチド、またはポリマーであり得る。いくつかの態様において、リンカーは、Cas9ヌクレアーゼドメインを包含するRNAによってプログラム可能なヌクレアーゼのgRNA結合ドメインと、核酸編集蛋白質の触媒ドメインとをつなぎ合わせる。いくつかの態様において、リンカーはdCas9および核酸編集蛋白質をつなぎ合わせる。典型的には、リンカーは、2つの基、分子、または他の部分の間に位置するかまたはそれらによってフランキングされ、共有結合を介して各1つに接続され、それゆえに2つを接続する。いくつかの態様において、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蛋白質）である。いくつかの態様において、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。いくつかの態様において、リンカーは1～100アミノ酸長、例えば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、30～35、35～40、40～45、45～50、50～60、60～70、70～80、80～90、90～100、100～150、または150～200アミノ酸長である。より長いまたはより短いリンカーもまた企図される。

【０２２８】

本明細書において用いられる用語「変異」は、配列、例えば核酸もしくはアミノ酸配列内の残基の別の残基による置換、または配列内の1つ以上の残基の欠失もしくは挿入を言う。典型的には、変異は、元々の残基、次に配列内の残基の位置および新たに置換された残基のアイデンティティを同定することによって本明細書に記載される。本明細書において提供されるアミノ酸置換（変異）を作るための種々の方法は当分野において周知であり、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。

【０２２９】

本明細書において用いられる用語「核酸」および「核酸分子」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言う。典型的には、ポリマー状の核酸、例えば3つ以上のヌクレオチドを含む核酸分子は、線状分子であり、その中の隣接ヌクレオチド同士はホスホジエステル連結部を介して互いに連結される。いくつかの態様において、「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）を言う。いくつかの態様において、「核酸」は3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言う。本明細書において用いられる用語「オリゴヌクレオチド」および「ポリヌクレオチド」は、ヌクレオチドのポリマー（例えば、少なくとも3ヌクレオチドのストリング）を言うために交換可能に用いられ得る。いくつかの態様において、「核酸」はRNAならびに一本および／または二本鎖DNAを包摂する。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、クロマチド、または他の天然に生起する核酸分子のコンテキストにおいて、天然に生起し得る。他方で、核酸分子は、天然に生起しない分

子、例えば組換えDNAもしくはRNA、人工染色体、人工ゲノム、またはその断片、あるいは合成DNA、RNA、DNA/RNAハイブリッドであり得、あるいは天然に生起しないヌクレオチドまたはヌクレオシドを包含してい得る。さらにその上、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バックボーン以外を有するアナログを包含する。核酸は、天然のソースから精製、組換え発現システムを用いて生成および任意に精製、化学合成などされ得る。適切なところでは、例えば化学合成される分子のケースでは、核酸は、ヌクレオシドアナログ、例えば化学修飾された塩基または糖を有するアナログ、ならびにバックボーン修飾を含み得る。別様に指示されない限り、核酸配列は5'から3'の方向に提示される。いくつかの態様において、核酸は、天然のヌクレオシド（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）；ヌクレオシドアナログ（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-ブロモウリジン、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-オキソグアノシン、O（6）-メチルグアニン、および2-チオシチジン）；化学修飾塩基；生物学的修飾塩基（例えばメチル化塩基）；インターカレーション塩基；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連結部）であるか、またはそれを含み得る。

【０２３０】

本明細書において用いられる用語「核酸編集ドメイン」は、核酸（例えばDNAまたはRNA）に1つ以上の改変（例えば、シチジン残基の脱アミノ化）をなすことができる蛋白質または酵素を言う。例示的な核酸編集ドメインは、デアミナーゼ、ヌクレアーゼ、ニッカーゼ、リコンビナーゼ、メチルトランスフェラーゼ、メチラーゼ、アセチラーゼ、アセチルトランスフェラーゼ、転写活性化因子、または転写リプレッサードメインを包含するが、これらに限定されない。いくつかの態様において、核酸編集ドメインはデアミナーゼ（例えば、シチジンデアミナーゼ、例えばAPOBECまたはAIDデアミナーゼ）である。

【０２３１】

本明細書において用いられる用語「増殖性疾患」は、細胞または細胞集団が異常に上昇した増殖率を見せることにおいて、細胞または組織ホメオスタシスが撹乱されているいずれかの疾患を言う。増殖性疾患は、高増殖性疾患、例えば前新生物性の過形成状態および新生物性疾患を包含する。新生物性疾患は細胞の異常な増殖を特徴とし、良性および悪性新生物両方を包含する。悪性新生物は癌ともまた言われる。

【０２３２】

用語「蛋白質」、「ペプチド」、および「ポリペプチド」は本明細書において交換可能に用いられ、ペプチド（アミド）結合によって一緒に連結されたアミノ酸残基同士のポリマーを言う。用語は、いずれかのサイズ、構造、または機能の蛋白質、ペプチド、またはポリペプチドを言う。典型的には、蛋白質、ペプチド、またはポリペプチドは少なくとも3アミノ酸の長さであるであろう。蛋白質、ペプチド、またはポリペプチドは個々の蛋白質または蛋白質の集まりを言い得る。蛋白質、ペプチド、またはポリペプチド中のアミノ酸の1つ以上は修飾され得、例えば、化学的実体、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、コンジュゲーション、官能化、または他の改変のためのリンカーなどの追加による。蛋白質、ペプチド、またはポリペプチドは単分子でもまたあり得、または多分子複合体であり得る。蛋白質、ペプチド、またはポリペプチドは、天然に生起する蛋白質またはペプチドの単に断片であり得る。蛋白質、ペプチド、またはポリペプチドは、天然に生起するか、組換えか、もしくは合成は、またはそれらのいずれかの組み合わせであり得る。本明細書において用いられる用語「融合蛋白質」は、少なくとも2つの異なる蛋白質からの蛋白質ドメインを含むハイブリッ

ドポリペプチドを言う。1つの蛋白質は融合蛋白質のアミノ末端（N末端）部またはカルボキシ末端（C末端）部に配置され得、それゆえにそれぞれ「アミノ末端融合蛋白質」または「カルボキシ末端融合蛋白質」を形成する。蛋白質は、異なるドメイン、例えば核酸編集蛋白質の核酸結合ドメイン（例えば、標的部位への蛋白質の結合を導くCas9のgRNA結合ドメイン）および核酸切断ドメインまたは触媒ドメインを含み得る。いくつかの態様において、蛋白質は、蛋白質部、例えば核酸結合ドメインを構成するアミノ酸配列と、有機化合物、例えば核酸切断薬剤として作用し得る化合物とを含む。いくつかの態様において、蛋白質は、核酸、例えばRNAとの複合体にあるか、またはそれとの結び付きにある。本明細書において提供される蛋白質のいずれかは当分野において公知のいずれかの方法によって生成され得る。例えば、本明細書において提供される蛋白質は組換え蛋白質発現および精製によって生成され得、これはペプチドリンカーを含む融合蛋白質に特に適している。組換え蛋白質発現および精製のための方法は周知であり、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包含する。これの内容全体は参照によって本明細書に組み込まれる。

【０２３３】

本明細書において用いられる用語「対象」は、個々の生物、例えば個々の哺乳動物を言う。いくつかの態様において、対象はヒトである。いくつかの態様において、対象は非ヒトの哺乳動物である。いくつかの態様において、対象は非ヒトの霊長類である。いくつかの態様において、対象はげっ歯類である。いくつかの態様において、対象はヒツジ、ヤギ、ウシ(a cattle)、ネコ、またはイヌである。いくつかの態様において、対象は脊椎動物、両生類の動物、爬虫類の動物、魚類の動物、昆虫、ハエ、または線虫である。いくつかの態様において、対象は研究動物である。いくつかの態様において、対象は遺伝子組換えされ、例えば遺伝子組換え非ヒト対象である。対象はどちらかの性および発生のいずれかのステージであり得る。

【０２３４】

用語「標的部位」は、デアミナーゼ、またはデアミナーゼを含む融合蛋白質（例えば、本明細書において提供されるdCas9-デアミナーゼ融合蛋白質）によって脱アミノ化される核酸分子内の配列を言う。

【０２３５】

用語「処置」、「処置する」、および「処置すること」は、本明細書に記載される通り、疾患もしくは異常またはその1つ以上の症状を逆行させるか、軽減するか、その始まりを遅延させるか、またはその進行を阻害することを狙った臨床的介入を言う。本明細書において用いられる用語「処置」、「処置する」、および「処置すること」は、本明細書に記載される通り、疾患もしくは異常またはそれの1つ以上の症状を逆行させるか、軽減するか、その始まりを遅延させるか、またはその進行を阻害することを狙った臨床的介入を言う。いくつかの態様において、処置は、1つ以上の症状が発生した後におよび／または疾患が診断された後に投与され得る。他の態様においては、処置は、例えば症状の始まりを防止するかもしくは遅延させるかまたは疾患の始まりもしくは進行を阻害するために、症状の非存在下において投与され得る。例えば、処置は、（例えば、症状の既往に照らしておよび／または遺伝学的もしくは他の感受性因子に照らして）症状の始まりに先立って感受性の個体に投与され得る。処置は、例えばそれらの再発を防止するかまたは遅延させるために、症状が解消した後にもまた続けられ得る。

【０２３６】

蛋白質または核酸のコンテキストにおいて本明細書において用いられる用語「組換え」は、天然には生起しないがヒトの工学の生成物である蛋白質または核酸を言う。例えば、いくつかの態様において、組換え蛋白質または核酸分子は、いずれかの天然に生起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。

【０２３７】
本明細書において用いられる用語「医薬組成物」は、疾患または異常の処置のコンテキストにおいて対象に投与され得る組成物を言う。いくつかの態様において、医薬組成物は、活性成分、例えばヌクレアーゼまたはヌクレアーゼをコードする核酸と、薬学的に許容し得る添加剤とを含む。
【０２３８】
本明細書において用いられる用語「塩基編集因子（BE）」または「核酸塩基編集因子（NBE）」は、核酸配列（例えばDNAまたはRNA）中の塩基（例えば、A、T、C、G、またはU）に改変をなすことができるポリペプチドを含む薬剤を言う。いくつかの態様において、塩基編集因子は核酸の塩基を脱アミノ化することができる。いくつかの態様において、塩基編集因子はDNA分子の塩基を脱アミノ化することができる。いくつかの態様において、塩基編集因子はDNAのシトシン（C）を脱アミノ化することができる。いくつかの態様において、塩基編集因子は、シチジンデアミナーゼドメインに融合させられた核酸によってプログラム可能なDNA結合蛋白質（napDNAbp）を含む融合蛋白質である。いくつかの態様において、塩基編集因子は、シチジンデアミナーゼに融合させられたCas9（例えば、dCas9およびnCas9）、CasX、CasY、Cpf1、C2c1、C2c2、C2c3、またはアルゴノート蛋白質を含む。いくつかの態様において、塩基編集因子はシチジンデアミナーゼに融合させられたCas9ニッカーゼ（nCas9）を含む。いくつかの態様において、塩基編集因子はシチジンデアミナーゼに融合させられたヌクレアーゼ不活性型Cas9（dCas9）を含む。いくつかの態様において、塩基編集因子は、塩基除去修復の阻害因子、例えばUGIドメインに融合させられる。いくつかの態様において、塩基編集因子はシチジンデアミナーゼに融合させられたCasX蛋白質を含む。いくつかの態様において、塩基編集因子はシチジンデアミナーゼに融合させられたCasY蛋白質を含む。いくつかの態様において、塩基編集因子はシチジンデアミナーゼに融合させられたCpf1蛋白質を含む。いくつかの態様において、塩基編集因子はシチジンデアミナーゼに融合させられたC2c1蛋白質を含む。いくつかの態様において、塩基編集因子はシチジンデアミナーゼに融合させられたC2c2蛋白質を含む。いくつかの態様において、塩基編集因子はシチジンデアミナーゼに融合させられたC2c3蛋白質を含む。いくつかの態様において、塩基編集因子はシチジンデアミナーゼに融合させられたアルゴノート蛋白質を含む。
【０２３９】
本明細書において用いられる用語「ウラシルグリコシラーゼ阻害因子」または「UGI」はウラシルDNAグリコシラーゼ塩基除去修復酵素を阻害することができる蛋白質を言う。
【０２４０】
本明細書において用いられる用語「Cas9ニッカーゼ」は、二重鎖核酸分子（例えば二重鎖DNA分子）の1つの鎖のみを切断することができるCas9蛋白質を言う。いくつかの態様において、Cas9ニッカーゼはD10A変異を含み、配列番号6の位置H840のヒスチジン、または別のCas9ドメイン、例えば本明細書において提供されるCas9蛋白質のいずれかの対応する変異を有する。例えば、Cas9ニッカーゼは配列番号8によって規定されるアミノ酸配列を含み得る。かかるCas9ニッカーゼは活性なHNHヌクレアーゼドメインを有し、DNAの非標的鎖、すなわちgRNAによって結合される鎖を切断する能力がある。さらに、かかるCas9ニッカーゼは不活性なRuvCヌクレアーゼドメインを有し、DNAの標的鎖、すなわち塩基編集が所望である鎖を切断する能力がない。
【０２４１】
例示的なCas9ニッカーゼ（クローニングベクターpPlatTET-gRNA2；アクセッションNo.BAV54124）。
MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDI

LRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG
GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ
EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLIN
GIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA
GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGI
KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK
DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRK
DFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ
EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK
VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
KRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII
EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID
RKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD（配列番号8）

【発明を実施するための形態】

【０２４２】

本発明のある態様の詳細な説明

本開示のいくつかの側面は融合蛋白質を提供し、これらは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またはCpf1蛋白質）と酵素ドメイン、例えば、例えばデアミナーゼドメインなどのDNA編集ドメインとを含む。デアミナーゼによる核酸塩基の脱アミノ化はそれぞれの残基の点変異に至り得、これは本明細書において核酸編集と言われる。それゆえに、Cas9バリアントまたはドメインとDNA編集ドメインとを含む融合蛋白質は、核酸配列の標的化された編集のために用いられ得る。かかる融合蛋白質は、in vitroのDNAの標的化された編集のために、例えば変異体細胞または動物の創出のために；標的化された変異の導入のために、例えば、エクスビボの細胞、例えば爾後に同じまたは別の対象に再導入される対象から得られた細胞における遺伝子の欠陥の修正のために；および標的化された変異の導入、例えば対象の遺伝子の欠陥の修正または疾患関連遺伝子の脱活性化変異の導入のために、有用である。典型的には、本明細書に記載される融合蛋白質のCas9ドメインはいずれかのヌクレアーゼ活性を有さず、代わりに、Cas9断片またはdCas9蛋白質もしくはドメインである。本発明の他の側面は融合蛋白質を提供し、これらは（i）核酸配列に結合することができるドメイン（例えば、Cas9またはCpf1蛋白質）；（ii）酵素ドメイン、例えばDNA編集ドメイン（例えばデアミナーゼドメイン）；および（iii）1つ以上のウラシルグリコシラーゼ阻害因子（UGI）ドメインを含む。少なくとも1つのUGIドメインの存在は、UGIドメインなしの融合蛋白質と比較して塩基編集効率を増大させる。2つのUGIドメインを含む融合蛋白質は、1つのUGIドメインを有するかまたはUGIドメインなしの融合蛋白質と比較して塩基編集効率および生成物純度をさらに増大させる。本明細書に記載されるCas9融合蛋白質の使用のための方法もまた提供される。

【０２４３】

核酸によってプログラム可能なDNA結合蛋白質

本開示のいくつかの側面は核酸によってプログラム可能なDNA結合蛋白質を提供し、これらは塩基編集因子などの蛋白質を特異的な核酸（例えばDNAまたはRNA）配列までガイドするために用いられ得る。本明細書において提供される融合蛋白質（例えば塩基編集因子）のいずれかはいずれかの核酸によってプログラム可能なDNA結合蛋白質（napDNAbp）を包含し得ることは了解されるべきである。例えば、Cas9ドメインを包含する本明細書に記載される融合蛋白質のいずれかは、Cas9ドメインの代わりに別のnapDNAbp、例えばCasX、CasY、Cpf1、C2c1、C2c2、C2c3、およびアルゴノートを用い得る。核酸に

よってプログラム可能なDNA結合蛋白質は、限定なしに、Cas9（例えばdCas9およびnCas9）、CasX、CasY、Cpf1、C2c1、C2c2、C2C3、およびアルゴノートを包含する。Cas9とは異なるPAM特異性を有する核酸によってプログラム可能なDNA結合蛋白質の1つの例はClustered Regularly Interspaced Short Palindromic Repeats from Prevotella and Francisella 1（Cpf1）である。Cas9と類似に、Cpf1もまたクラス2 CRISPRエフェクターである。Cpf1はロバストなDNA干渉を媒介し、Cas9とは別々の特質を有することが示されている。Cpf1は、tracrRNAを欠く1つのRNAによってガイドされるエンドヌクレアーゼであり、それは、Tリッチなプロトスペーサー隣接モチーフ（TTN、TTTN、またはYTN）を利用する。その上、Cpf1は互い違いのDNA二本鎖切断によってDNAを切断する。16個のCpf1ファミリー蛋白質のうち、AcidaminococcusおよびLachnospiraceaeからの2つの酵素は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蛋白質は当分野において公知であり、先に記載されている。例えばYamano et al., "Crystal structure of Cpfl in complex with guide RNA and target DNA." Cell (165) 2016, p. 949-962；これの内容全体は参照によって本明細書に組み込まれる。

【０２４４】

ガイドヌクレオチド配列によってプログラム可能なDNA結合蛋白質ドメインとして用いられ得るヌクレアーゼ不活性型Cpf1（dCpf1）バリアントもまた本組成物および方法に有用である。Cpf1蛋白質は、Cas9のRuvCドメインに類似である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のアルファヘリックス状の認識ローブを有さない。Zetsche et al., Cell, 163, 759-771, 2015（これは参照によって本明細書に組み込まれる）においては、Cpf1のRuvC様ドメインは両方のDNA鎖を切断することを担っており、RuvC様ドメインの不活性化はCpf1ヌクレアーゼ活性を不活性化することが示されている。例えば、Francisella novicida Cpf1（配列番号15）のD917A、E1006A、またはD1255Aに対応する変異はCpf1ヌクレアーゼ活性を不活性化する。いくつかの態様において、デッドなCpf1（dCpf1）は配列番号9のD917A、E1006A、D1255A、D917A/E1006A、D917A/D1255A、E1006A/D1255A、またはD917A/E1006A/D1255Aに対応する変異を含む。Cpf1のRuvCドメインを不活性化するいずれかの変異、例えば、置換変異、欠失、または挿入が本開示に従って用いられ得ることは理解されるべきである。

【０２４５】

いくつかの態様において、本明細書において提供される融合蛋白質のいずれかの核酸によってプログラム可能なDNA結合蛋白質（napDNAbp）はCpf1蛋白質である。いくつかの態様において、Cpf1蛋白質はCpf1ニッカーゼ（nCpf1）である。いくつかの態様において、Cpf1蛋白質はヌクレアーゼ不活性型Cpf1（dCpf1）である。いくつかの態様において、Cpf1、nCpf1、またはdCpf1は、配列番号9～24のいずれか1つと少なくとも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%同一であるアミノ酸配列を含む。いくつかの態様において、dCpf1は、配列番号9～16のいずれか1つと少なくとも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%同一であるアミノ酸配列を含み、配列番号9のD917A、E1006A、D1255A、D917A/E1006A、D917A/D1255A、E1006A/D1255A、またはD917A/E1006A/D1255Aに対応する変異を含む。いくつかの態様において、dCpf1蛋白質はいずれか1つの配列番号9～16のアミノ酸配列を含む。他の種からのCpf1もまた本開示に従って用いられ得ることは了解されるべきである。

野生型Francisella novicida Cpf1（配列番号9）（D917、E1006、およびD1255は太字および下線付きである）

【化７】

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKF
KNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTY
FKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKD
LAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK

【化８】

RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIG
TAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQC
RFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQ
TNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYS
DEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN
KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGY
EKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTF
ENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYK
LNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHC
PITINFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFN
IIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEY
NAIVVF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAY
QLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDK
ICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPT
KELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLI
SPVADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVI
KNEEYFEFVQNRNN

（配列番号９）

Francisella novicida Cpf1 D917A（配列番号10）（A917、E1006、およびD1255は太字および下線付きである）

【化９】

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKF
KNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTY
FKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKD
LAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIG
TAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQC
RFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQ
TNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYS
DEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN
KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGY
EKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTF

【化１０】

ENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYK
LNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHC
PITINFKSSGANKFNDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFN
IIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEY
NAIVVF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAY
QLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDK
ICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPT
KELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLI
SPVADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVI
KNEEYFEFVQNRNN

（配列番号10）

Francisella novicida Cpf1 E1006A（配列番号11）（D917、A1006、およびD1255は太字および下線付きである）

【化１１】

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKF
KNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTY
FKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKD
LAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIG
TAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQC
RFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQ
TNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYS
DEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN
KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGY
EKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTF
ENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYK
LNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHC
PITINFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFN
IIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEY
NAIVVF**A**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAY
QLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDK
ICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPT
KELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLI
SPVADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVI
KNEEYFEFVQNRNN

（配列番号11）
Francisella novicida Cpf1 D1255A（配列番号12）（D917、E1006、およびA1255は太字および下線付きである）

【化１３】

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKF
KNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTY
FKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKD
LAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIG
TAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQC
RFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQ
TNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYS
DEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN
KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGY
EKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTF
ENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYK
LNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHC
PITINFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFN
IIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEY
NAIVVF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAY
QLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDK
ICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPT
KELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLI
SPVADVNGNFFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVI
KNEEYFEFVQNRNN

（配列番号12）

Francisella novicida Cpf1 D917A/E1006A（配列番号13）（A917、A1006、およびD1255は太字および下線付きである）

【化１４】

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKF
KNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTY

【化１５】

FKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKD
LAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIG
TAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQC
RFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQ
TNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYS
DEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN
KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGY
EKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTF
ENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYK
LNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHC
PITINFKSSGANKFNDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFN
IIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEY
NAIVVF**A**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAY
QLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDK
ICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPT
KELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLI
SPVADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVI
KNEEYFEFVQNRNN

（配列番号13）

Francisella novicida Cpf1 D917A/D1255A（配列番号14）（A917、E1006、およびA1255は太字および下線付きである）

【化１６】

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKF
KNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTY
FKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKD
LAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIG
TAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQC
RFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQ
TNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYS
DEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN

【化１７】

KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGY
EKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTF
ENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYK
LNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHC
PITINFKSSGANKFNDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFN
IIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEY
NAIVVF**E**EDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAY
QLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDK
ICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPT
KELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLI
SPVADVNGNFFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVI
KNEEYFEFVQNRNN

（配列番号14）

Francisella novicida Cpf1 E1006A/D1255A（配列番号15）（D917、A1006、およびA1255は太字および下線付きである）

【化１８】

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKF
KNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTY
FKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKD
LAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIG
TAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQC
RFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQ
TNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYS
DEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN
KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGY
EKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTF
ENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYK
LNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHC
PITINFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFN
IIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEY
NAIVVF**A**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAY
QLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDK

【化１９】

ICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPT
KELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLI
SPVADVNGNFFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVI
KNEEYFEFVQNRNN

（配列番号15）

Francisella novicida Cpf1 D917A/E1006A/D1255A（配列番号16）（A917、A1006、およびA1255は太字および下線付きである）

【化２０】

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKF
KNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTY
FKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKD
LAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIG
TAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQC
RFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQ
TNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYS
DEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN
KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGY
EKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTF
ENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYK
LNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHC
PITINFKSSGANKFNDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFN
IIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEY
NAIVVF**A**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAY
QLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDK
ICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPT
KELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLI
SPVADVNGNFFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVI
KNEEYFEFVQNRNN

（配列番号16）

【０２４６】

いくつかの態様において、核酸によってプログラム可能なDNA結合蛋白質はAcidaminococcus種からのCpf1蛋白質（AsCpf1）である。Acidaminococcus種からの（form）Cpf1蛋白質は先に記載されており、当業者には明らかであろう。例示的なAcidaminococcus Cpf1蛋白質（AsCpf1）は、限定なしに、本明細書において提供されるAsCpf1蛋白質

のいずれかを包含する。

【０２４７】

野生型AsCpf1-残基R912は太字の下線付きで指示されており、残基661～667は斜体および下線付きで指示されている。

【化２１】

TQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTY
ADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTD
AINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRK
NVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEE
VFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLP
HRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSID
LTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINL
QEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTMLKKQEEKEILKSQLDSLLGLY
HLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNF
QMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFD
KMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEK
EPKKFQTAYA*KKTGDQK*GYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLG
EYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWT
GLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLY
QELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAA
NSPSKFNQRVNAYLKEHPETPIIGIDRGE**R**NLIYITVIDSTGKILEQRSLNTIQQFDYQKK
LDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFK
SKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGT
QSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILH
FKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYR
DLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATG
EDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKL
QNGISNQDWLAYIQELRN

（配列番号17）

【０２４８】

AsCpf1（R912A）- 残基A912は太字の下線付きで指示されており、残基661～667は斜体および下線付きで指示されている。

【０２４９】

【化２２】

TQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPII
DRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIG
RTDNLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFS

【化２３】

GFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGI
FVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDET
AHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALF
NELNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSL
KHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTMLKKQEEKEILKSQL
DSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVE
KFKLNFQMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEK
TSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYD
LNNPEKEPKKFQTAYA*KKTGDQK*GYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSS
QYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNL
HTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKT
PIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPIT
LNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGE**A**NLIYITVIDSTGKILEQRSLNTIQQ
FDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLE
NLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFT
SFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDV
KTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIEN
HRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQM
RNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHL
KESKDLKLQNGISNQDWLAYIQELRN

（配列番号19）

【０２５０】

いくつかの態様において、核酸によってプログラム可能なDNA結合蛋白質はLachnospiraceae種からのCpf1蛋白質（LbCpf1）である。Lachnospiraceae種からの（form）Cpf1蛋白質は先に記載されており、当業者には明らかであろう。例示的なLachnospiraceae Cpf1蛋白質（LbCpf1）は、限定なしに、本明細書において提供されるLbCpf1蛋白質のいずれかを包含する。

【０２５１】

いくつかの態様において、LbCpf1はニッカーゼである。いくつかの態様において、LbCpf1ニッカーゼは配列番号18に対して相対的にR836X変異体を含み、XはRを除くいずれかのアミノ酸である。いくつかの態様において、LbCpf1ニッカーゼは配列番号18に対して相対的にR836A変異体を含む。いくつかの態様において、LbCpf1はヌクレアーゼ不活性型LbCpf1（dLbCpf1）である。いくつかの態様において、dLbCpf1は配列番号18に対して相対的にD832X変異体を含み、XはDを除くいずれかのアミノ酸である。いくつかの態様において、dLbCpf1は配列番号18に対して相対的にD832A変異体を含む。追加のdCpf1蛋白質は当分野において例えばLi et al. "Base editing with a Cpfl-cytidine deaminase fusion" Nature Biotechnology; March 2018 DOI: 10.1038/nbt.4102に記載されており；これの内容全体は参照によって本明細書に組み込まれる。いくつかの態様において、dCpf1はLi et al.に記載されているCpf1の点変異D832A、E1006A、D1125Aの1つ、2つ、または3つを含む。

【０２５２】

野生型LbCpf1-残基R836およびR1138は太字の下線付きで指示されている。
【化２４】

MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKGVKKLLDRY
YLSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKGNEGYK
SLFKKDIIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAFRCIN
ENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDV
YNAIIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRESLSFYGEGYT
SDEEVLEVFRNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWN
VIRDKWNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYADADLSVVEK
LKEIIIQKVDEIYKVYGSSEKLFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKA
FFGEGKETNRDESFYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQF
MGGWDKDKETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKL
LPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFKDS
ISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKKEVDKLVEEGKL
YMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEEL
VVHPANSPIANKNPDNPKKTTTLSYDVYKDKRFSEDQYELHIPIAINKCPKNIFKINTE
VRVLLKHDDNPYVIGIDRGE**R**NLLYIVVVDGKGNIVEQYSLNEIINNFNGIRIKTDYHS
LLDKKEKERFEARQNWTSIENIKELKAGYISQVVHKICELVEKYDAVIALEDLNSGFK
NSRVKVEKQVYQKFEKMLIDKLNYMVDKKSNPCATGGALKGYQITNKFESFKSMST
QNGFIFYIPAWLTSKIDPSTGFVNLLKTKYTSIADSKKFISSFDRIMYVPEEDLFEFALD
YKNFSRTDADYIKKWKLYSYGNRIRIFRNPKKNNVFDWEEVCLTSAYKELFNKYGIN
YQQGDIRALLCEQSDKAFYSSFMALMSLMLQM**R**NSITGRTDVDFLISPVKNSDGIFYD
SRNYEAQENAILPKNADANGAYNIARKVLWAIGQFKKAEDEKLDKVKIAISNKEWLE
YAQTSVKH

（配列番号18）
【０２５３】
LbCpf1（R836A）-残基A836は太字の下線付きで指示されている。
【化２５】

MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKGVKKLLDRY
YLSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKGNEGYK
SLFKKDIIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAFRCIN
ENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDV
YNAIIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRESLSFYGEGYT

【化２６】

SDEEVLEVFRNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWN
VIRDKWNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYADADLSVVEK
LKEIIIQKVDEIYKVYGSSEKLFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKA
FFGEGKETNRDESFYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQF
MGGWDKDKETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKL
LPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFKDS
ISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKKEVDKLVEEGKL
YMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEEL
VVHPANSPIANKNPDNPKKTTTLSYDVYKDKRFSEDQYELHIPIAINKCPKNIFKINTE
VRVLLKHDDNPYVIGIDRGE**A**NLLYIVVVDGKGNIVEQYSLNEIINNFNGIRIKTDYHS
LLDKKEKERFEARQNWTSIENIKELKAGYISQVVHKICELVEKYDAVIALEDLNSGFK
NSRVKVEKQVYQKFEKMLIDKLNYMVDKKSNPCATGGALKGYQITNKFESFKSMST
QNGFIFYIPAWLTSKIDPSTGFVNLLKTKYTSIADSKKFISSFDRIMYVPEEDLFEFALD
YKNFSRTDADYIKKWKLYSYGNRIRIFRNPKKNNVFDWEEVCLTSAYKELFNKYGIN
YQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDFLISPVKNSDGIFYD
SRNYEAQENAILPKNADANGAYNIARKVLWAIGQFKKAEDEKLDKVKIAISNKEWLE
YAQTSVKH

（配列番号20）

【０２５４】

LbCpf1（R1138A）-残基A1138は太字の下線付きで指示されている。

【化２７】

MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKGVKKLLDRY
YLSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKGNEGYK
SLFKKDIIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAFRCIN
ENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDV
YNAIIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRESLSFYGEGYT
SDEEVLEVFRNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWN
VIRDKWNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYADADLSVVEK
LKEIIIQKVDEIYKVYGSSEKLFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKA
FFGEGKETNRDESFYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQF
MGGWDKDKETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKL
LPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFKDS
ISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKKEVDKLVEEGKL
YMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEEL
VVHPANSPIANKNPDNPKKTTTLSYDVYKDKRFSEDQYELHIPIAINKCPKNIFKINTE
VRVLLKHDDNPYVIGIDRGERNLLYIVVVDGKGNIVEQYSLNEIINNFNGIRIKTDYHS

【化２８】

LLDKKEKERFEARQNWTSIENIKELKAGYISQVVHKICELVEKYDAVIALEDLNSGFK
NSRVKVEKQVYQKFEKMLIDKLNYMVDKKSNPCATGGALKGYQITNKFESFKSMST
QNGFIFYIPAWLTSKIDPSTGFVNLLKTKYTSIADSKKFISSFDRIMYVPEEDLFEFALD
YKNFSRTDADYIKKWKLYSYGNRIRIFRNPKKNNVFDWEEVCLTSAYKELFNKYGIN
YQQGDIRALLCEQSDKAFYSSFMALMSLMLQMANSITGRTDVDFLISPVKNSDGIFYD
SRNYEAQENAILPKNADANGAYNIARKVLWAIGQFKKAEDEKLDKVKIAISNKEWLE
YAQTSVKH

（配列番号21）

【０２５５】

いくつかの態様において、Cpf1蛋白質は不具化されたCpf1蛋白質である。本明細書において用いられる「不具化されたCpf1」蛋白質は、野生型Cpf1蛋白質と比較して低減されたヌクレアーゼ活性を有するCpf1蛋白質である。いくつかの態様において、不具化されたCpf1蛋白質は非標的鎖よりも効率的に標的鎖を選好的に切る。例えば、Cpf1蛋白質は、編集されるべきヌクレオチドが所在する二重鎖核酸分子の鎖を選好的に切る。いくつかの態様において、不具化されたCpf1蛋白質は標的鎖よりも効率的に非標的鎖を選好的に切る。例えば、Cpf1蛋白質は、編集されるべきヌクレオチドが所在しない二重鎖核酸分子の鎖を選好的に切る。いくつかの態様において、不具化されたCpf1蛋白質は、それが非標的鎖を切るよりも少なくとも5%効率的に標的鎖を選好的に切る。いくつかの態様において、不具化されたCpf1蛋白質は、それが非標的鎖を切るよりも少なくとも5%、10%、15%、20%、25%、30%、35%、40%、50%、60%、70%、80%、90%、または少なくとも100%効率的に、標的鎖を選好的に切る。

【０２５６】

いくつかの態様において、不具化されたCpf1蛋白質は天然に生起しないCpf1蛋白質である。いくつかの態様において、不具化されたCpf1蛋白質は野生型Cpf1蛋白質に対して相対的に1つ以上の変異を含む。いくつかの態様において、不具化されたCpf1蛋白質は、野生型Cpf1蛋白質に対して相対的に1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、または20個の変異を含む。いくつかの態様において、不具化されたCpf1蛋白質は、配列番号18によって規定される、または別のCpf1蛋白質の対応するアミノ酸のR836A変異を含む。配列番号18のR836Aに対して相同な残基（例えば対応するアミノ酸）を含むCpf1もまた、類似の結果を達成するように変異させられ得ることは了解されるべきである。いくつかの態様において、不具化されたCpf1蛋白質は、配列番号18によって規定される、または別のCpf1蛋白質の対応するアミノ酸のR1138A変異を含む。いくつかの態様において、不具化されたCpf1蛋白質は、配列番号17によって規定される、または別のCpf1蛋白質の対応するアミノ酸のR912A変異を含む。いずれの特定の理論によっても拘束されることを望むものではないが、配列番号18（LbCpf1）の残基R836および配列番号17（AsCpf1）の残基R912は対応する（例えば相同な）残基の例である。例えば、配列番号17および18間のアライメントのある部分は、R912およびR836が対応する残基であることを示す。

【化２９】

```
AsCpf1      YQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQ--
LbCpf1      KCPKN-IFKINTEVRVLLKHDDNPYVIGIDRGERNLLYIVVVDGKGNIVEQYSLNEIINN
                 *   *:* .*.. **.. :  :**********:**.*:*..*:*:** *** *
```

【０２５７】

いくつかの態様において、本明細書において提供されるCpf1蛋白質のいずれかは1つ以上のアミノ酸欠失を含む。いくつかの態様において、本明細書において提供されるCpf1蛋白質のいずれかは1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、または20個のアミノ酸欠失を含む。いずれのの特定の理論によっても拘束されることを望むものではないが、Cpf1には、Cpf1に融合させられているデアミナーゼ（例えばAPOBEC）の機能を妨害し得るヘリックス状領域があり、これはAsCpf1（配列番号17）の残基661～667を包含する。この領域はアミノ酸配列KKTGDQKを含む。従って、本開示の側面は、Cpf1のこのヘリックス状領域を破壊する変異（例えば欠失）を含むCpf1蛋白質を提供する。いくつかの態様において、Cpf1蛋白質は、配列番号17の次の残基の1つ以上の欠失、または別のCpf1蛋白質の1つ以上の対応する欠失を含む：K661、K662、T663、G664、D665、Q666、およびK667。いくつかの態様において、Cpf1蛋白質は、配列番号17のT663およびD665欠失、または別のCpf1蛋白質の対応する欠失を含む。いくつかの態様において、Cpf1蛋白質は、配列番号17のK662、T663、D665、およびQ666欠失、または別のCpf1蛋白質の対応する欠失を含む。いくつかの態様において、Cpf1蛋白質は、配列番号17のK661、K662、T663、D665、Q666、およびK667欠失、または別のCpf1蛋白質の対応する欠失を含む。

【０２５８】

AsCpf1（T663およびD665を欠失）

TQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYAD
QCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINK
RHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAE
DISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSFPFY
NQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQ
ILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISHKKL
ETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSE
AFKQKTSEILSHAHAALDQPLPTTMLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEV
DPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVNKEKN
NGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPDAAKMIPKCSTQ
LKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKGQKGYREALC
KWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVE
TGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMK
RMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVS
HEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYI
TVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIH
EIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVG
GVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHF
LEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAG
KRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALI
RSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLL
LNHLKESKDLKLQNGISNQDWLAYIQELRN（配列番号22）

【０２５９】

AsCpf1（K662、T663、D665、およびQ666を欠失）

TQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYAD
QCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINK
RHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAE
DISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSFPFY
NQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQ
ILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISHKKL
ETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSE
AFKQKTSEILSHAHAALDQPLPTTMLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEV

DPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVNKEKN
NGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPDAAKMIPKCSTQ
LKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAKGKGYREALCKW
IDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETG
KLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKR
MAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSH
EIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYIT
VIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEI
VDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGG
VLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFL
EGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGK
RIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIR
SVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLL
NHLKESKDLKLQNGISNQDWLAYIQELRN（配列番号23）

【０２６０】

AsCpf1（K661、K662、T663、D665、Q666、およびK667を欠失）
TQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYAD
QCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINK
RHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAE
DISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSFPFY
NQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQ
ILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISHKKL
ETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSE
AFKQKTSEILSHAHAALDQPLPTTMLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEV
DPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVNKEKN
NGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPDAAKMIPKCSTQ
LKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAGGYREALCKWID
FTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKL
YLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMA
HRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEII
KDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVID
STGKILEQRSLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVD
LMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVL
NPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEG
FDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRI
VPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRS
VLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLN
HLKESKDLKLQNGISNQDWLAYIQELRN（配列番号24）

【０２６１】

いくつかの態様において、核酸によってプログラム可能なDNA結合蛋白質（napDNAbp）は、標的配列中に古典的（NGG）PAM配列を要求しない核酸によってプログラム可能なDNA結合蛋白質である。いくつかの態様において、napDNAbpはアルゴノート蛋白質である。かかる核酸によってプログラム可能なDNA結合蛋白質の1つの例はNatronobacterium gregoryiからのアルゴノート蛋白質（NgAgo）である。NgAgoはssDNAによってガイドされるエンドヌクレアーゼである。NgAgoは、それを標的部位までガイドする長さ～24ヌクレオチドの5'リン酸化されたssDNA（gDNA）に結合し、gDNA部位においてDNA二本鎖切断をなす。Cas9とは対照的に、NgAgo-gDNAシステムはプロトスペーサー隣接モチーフ（PAM）を要求しない。ヌクレアーゼ不活性型NgAgo（dNgAgo）を用いることは、標的化され得る塩基を多大に拡大し得る。NgAgoの特徴付けおよび使用はGao et al., Nat. Biotechnol., 2016 Jul；34(7):768-73. PubMed PMID: 27136078；Swarts

et al., Nature 507(7491)(2014):258-61；およびSwarts et al, Nucleic Acids Res. 43(10)(2015):5120-9に記載されており、これらのそれぞれは参照によって本明細書に組み込まれる。Natronobacterium gregoryi アルゴノートの配列は配列番号25によって提供される。

【０２６２】

いくつかの態様において、napDNAbpはアルゴノート蛋白質である。いくつかの態様において、napDNAbpは、天然に生起するアルゴノート蛋白質と少なくとも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%同一であるアミノ酸配列を含む。いくつかの態様において、napDNAbpは天然に生起するアルゴノート蛋白質である。いくつかの態様において、napDNAbpは、配列番号25のいずれか1つと少なくとも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%同一であるアミノ酸配列を含む。いくつかの態様において、napDNAbpはいずれか1つの配列番号25アミノ酸配列を含む。

野生型Natronobacterium gregoryi アルゴノート（配列番号25）

MTVIDLDSTTTADELTSGHTYDISVTLTGVYDNTDEQHPRMSLAFEQDNGERRYITLWK
NTTPKDVFTYDYATGSTYIFTNIDYEVKDGYENLTATYQTTVENATAQEVGTTDEDETFA
GGEPLDHHLDDALNETPDDAETESDSGHVMTSFASRDQLPEWTLHTYTLTATDGAKTDT
EYARRTLAYTVRQELYTDHDAAPVATDGLMLLTPEPLGETPLDLDCGVRVEADETRTLD
YTTAKDRLLARELVEEGLKRSLWDDYLVRGIDEVLSKEPVLTCDEFDLHERYDLSVEVGH
SGRAYLHINFRHRFVPKLTLADIDDDNIYPGLRVKTTYRPRRGHIVWGLRDECATDSLNT
LGNQSVVAYHRNNQTPINTDLLDAIEAADRRVVETRRQGHGDDAVSFPQELLAVEPNTH
QIKQFASDGFHQQARSKTRLSASRCSEKAQAFAERLDPVRLNGSTVEFSSEFFTGNNEQQ
LRLLYENGESVLTFRDGARGAHPDETFSKGIVNPPESFEVAVVLPEQQADTCKAQWDTM
ADLLNQAGAPPTRSETVQYDAFSSPESISLNVAGAIDPSEVDAAFVVLPPDQEGFADLASP
TETYDELKKALANMGIYSQMAYFDRFRDAKIFYTRNVALGLLAAAGGVAFTTEHAMPGD
ADMFIGIDVSRSYPEDGASGQINIAATATAVYKDGTILGHSSTRPQLGEKLQSTDVRDIMK
NAILGYQQVTGESPTHIVIHRDGFMNEDLDPATEFLNEQGVEYDIVEIRKQPQTRLLAVS
DVQYDTPVKSIAAINQNEPRATVATFGAPEYLATRDGGGLPRPIQIERVAGETDIETLTRQ
VYLLSQSHIQVHNSTARLPITTAYADQASTHATKGYLVQTGAFESNVGFL（配列番号25）

【０２６３】

いくつかの態様において、napDNAbpはアルゴノート蛋白質の原核生物ホモログである。アルゴノート蛋白質の原核生物ホモログは公知であり、例えばMakarova K., et al., "Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements", Biol. Direct. 2009 Aug 25；4:29. doi: 10.1186/1745-6150-4-29に記載されており、参照によって本明細書に組み込まれる。いくつかの態様において、napDNAbpはMarinitoga piezophila Argunaute（MpAgo）蛋白質である。CRISPR関連Marinitoga piezophila アルゴノート（MpAgo）蛋白質は5'リン酸化されたガイドを用いて一本鎖標的配列を切断する。5'ガイドは全ての公知のアルゴノートによって用いられる。MpAgo-RNA複合体の結晶構造は、5'リン酸相互作用をブロックする残基を含むガイド鎖結合部位を示す。このデータは、5'ヒドロキシル化されたガイドについて非古典的な特異性を有するアルゴノートサブクラスの進化を示唆している。例えばKaya et al., "A bacterial Argonaute with noncanonical guide RNA specificity", Proc Natl Acad Sci U S A. 2016 Apr 12；113（15）:4057-62を見よ。これの内容全体は参照によってここに組み込まれる）。例えばデアミナーゼ（例えばシチジンデアミナーゼ）を標的核酸（例えばssRNA）までガイドするために、他のアルゴノート蛋白質が本明細書に記載される融合蛋白質（例えば塩基編集因子）のいずれかに用いられ得ることは了解されるべきである。

【０２６４】
いくつかの態様において、核酸によってプログラム可能なDNA結合蛋白質（napDNAbp）は微生物CRISPR-Casシステムの単一のエフェクターである。微生物CRISPR-Casシステムの単一のエフェクターは、限定なしに、Cas9、Cpf1、C2c1、C2c2、およびC2c3を包含する。典型的には、微生物CRISPR-Casシステムはクラス1およびクラス2システムに分けられる。クラス1システムはマルチサブユニットエフェクター複合体を有し、クラス2システムは単一の蛋白質エフェクターを有する。Cas9およびCpf1はクラス2エフェクターである。Cas9およびCpf1に加えて、3つの別々のクラス2 CRISPR-Casシステム（C2c1、C2c2、および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おり、これの内容全体は参照によって本明細書に組み込まれる。システムの2つC2c1およびC2c3のエフェクターは、Cpf1に近縁のRuvC様エンドヌクレアーゼドメインを含有する。第3のシステムC2c2は、2つの予測されるHEPN RNaseドメインを有するエフェクターを含有する。C2c1によるCRISPR RNAの生成とは違って、成熟CRISPR RNAの生成はtracrRNA非依存的である。C2c1はDNA切断のためにはCRISPR RNAおよびtracrRNA両方に依存する。細菌C2c2は、そのRNAによって活性化される一本鎖RNA分解活性とは別々に、CRISPR RNA成熟のためのユニークなRNase活性を持つことが示されている。これらのRNase機能は互いおよびCpf1のCRISPR RNAプロセシング挙動とは異なる。例えば、East-Seletsky, et al., "Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection", Nature, 2016 Oct 13; 538(7624): 270-273を見よ。これの内容全体は参照によってここに組み込まれる。Leptotrichia shahiiのC2c2のin vitro生化学分析は、C2c2が単一のCRISPR RNAによってガイドされ、相補的なプロトスペーサーを持つssRNA標的を切断するようにプログラムされ得ることを示している。2つの保存されたHEPNドメインの触媒残基は切断を媒介する。触媒残基の変異は触媒不活性型RNA結合蛋白質を創出する。例えば、Abudayyeh et al., "C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector", Science, 2016 Aug 5; 353(6299)を見よ。これの内容全体は参照によってここに組み込まれる。

【０２６５】
Alicyclobaccillus acidoterrastris C2c1（AacC2c1）の結晶構造が、キメラ型の単分子ガイドRNA（sgRNA）との複合体として報告されている。例えば、参照によって本明細書に組み込まれるLiu et al., "C2cl-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism", Mol. Cell, 2017 Jan 19；65(2): 310-322を見よ。結晶構造は3者複合体として標的DNAに結合したAlicyclobacillus acidoterrestris C2c1についてもまた報告されている。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を見よ。これの内容全体は参照によってここに組み込まれる。標的および非標的DNA鎖両方によるAacC2c1の触媒的にコンピテントな立体配座が、独立して単一のRuvC触媒ポケット内に位置してキャプチャーされており、C2c1によって媒介される切断は標的DNAの互い違いの7ヌクレオチドの切断をもたらす。C2c1の3者複合体ならびに先に同定されたCas9およびCpf1カウンターパートの間の構造比較は、CRISPR-Cas9システムによって用いられるメカニズムの多様さを実証している。

【０２６６】
いくつかの態様において、本明細書において提供される融合蛋白質のいずれかの核酸によってプログラム可能なDNA結合蛋白質（napDNAbp）は、C2c1、C2c2、またはC2c3蛋白質である。いくつかの態様において、napDNAbpはC2c1蛋白質である。いくつかの態様において、napDNAbpはC2c2蛋白質である。いくつかの態様において、napDNAbpはC2c3蛋白質である。いくつかの態様において、napDNAbpは、天然に生起するC2c1、C2c2、またはC2c3蛋白質と少なくとも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%同一であるアミノ酸配列を含む。いくつかの態様において、napDNAbpは天然に生起するC2c1、C2c2、またはC2c3蛋白質である。いくつかの態様において、napDNAbpは、配列番号26～28のいずれか1つと少なくとも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%同一であるアミノ酸配列を含む。いくつかの態様において、napDNAbpはいずれか1つの配列番号26～28のアミノ酸配列を含む。他の細菌種からのC2c1、C2c2、またはC2c3もまた本開示に従って用いられ得ることは了解されるべきである。

C2c1（uniprot.org/uniprot/T0D7 A2#）

sp|T0D7A2|C2Cl_ALIAG CRISPR関連エンドヌクレアーゼC2c1 OS=Alicyclobacillus acidoterrestris（株ATCC 49025 / DSM 3922 / CIP 106132 / NCIMB 13137 / GD3B）GN=c2cl PE=1 SV=1

MAVKSIKVKLRLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQ
ECDKTAEECKAELLERLRARQVENGHRGPAGSDDELLQLARQLYELLVPQAIGAKGDAQ
QIARKFLSPLADKDAVGGLGIAKAGNKPRWVRMREAGEPGWEEEKEKAETRKSADRTAD
VLRALADFGLKPLMRVYTDSEMSSVEWKPLRKGQAVRTWDRDMFQQAIERMMSWESW
NQRVGQEYAKLVEQKNRFEQKNFVGQEHLVHLVNQLQQDMKEASPGLESKEQTAHYVT
GRALRGSDKVFEKWGKLAPDAPFDLYDAEIKNVQRRNTRRFGSHDLFAKLAEPEYQALW
REDASFLTRYAVYNSILRKLNHAKMFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEF
GERRHAIRFHKLLKVENGVAREVDDVTVPISMSEQLDNLLPRDPNEPIALYFRDYGAEQH
FTGEFGGAKIQCRRDQLAHMHRRRGARDVYLNVSVRVQSQSEARGERRPPYAAVFRLVG
DNHRAFVHFDKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDEL
KPNSKGRVPFFFPIKGNDNLVAVHERSQLLKLPGETESKDLRAIREERQRTLRQLRTQLAY
LRLLVRCGSEDVGRRERSWAKLIEQPVDAANHMTPDWREAFENELQKLKSLHGICSDKE
WMDAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYAKDVVGGNSIEQIEYLERQY
KFLKSWSFFGKVSGQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRIIMEALGYVYALDE
RGKGKWVAKYPPCQLILLEELSEYQFNNDRPPSENNQLMQWSHRGVFQELINQAQVHD
LLVGTMYAAFSSRFDARTGAPGIRCRRVPARCTQEHNPEPFPWWLNKFVVEHTLDACPL
RADDLIPTGEGEIFVSPFSAEEGDFHQIHADLNAAQNLQQRLWSDFDISQIRLRCDWGEV
DGELVLIPRLTGKRTADSYSNKVFYTNTGVTYYERERGKKRRKVFAQEKLSEEEAELLVEA
DEAREKSVVLMRDPSGIINRGNWTRQKEFWSMVNQRIEGYLVKQIRSRVPLQDSACENT
GDI（配列番号26）

C2c2（uniprot.org/uniprot/P0DOC6）

＞sp|P0DOC6|C2C2_LEPSD CRISPR関連エンドリボヌクレアーゼC2c2 OS=Leptotrichia shahii（株DSM 19757 / CCUG 47503 / CIP 107916 / JCM 16776 / LB37）GN=c2c2 PE=1 SV=1

【０２６７】

[00267] MGNLFGHKRWYEVRDKKDFKIKRKVKVKRNYDGNKYILNINENNNKEKID
NNKFIRKYINYKKNDNILKEFTRKFHAGNILFKLKGKEGIIRIENNDDFLETEEVVLYIEAY
GKSEKLKALGITKKKIIDEAIRQGITKDDKKIEIKRQENEEEIEIDIRDEYTNKTLNDCSIILR
IIENDELETKKSIYEIFKNINMSLYKIIEKIIENETEKVFENRYYEEHLREKLLKDDKIDVILT
NFMEIREKIKSNLEILGFVKFYLNVGGDKKKSKNKKMLVEKILNINVDLTVEDIADFVIKE
LEFWNITKRIEKVKKVNNEFLEKRRNRTYIKSYVLLDKHEKFKIERENKKDKIVKFFVENI
KNNSIKEKIEKILAEFKIDELIKKLEKELKKGNCDTEIFGIFKKHYKVNFDSKKFSKKSDEEK
ELYKIIYRYLKGRIEKILVNEQKVRLKKMEKIEIEKILNESILSEKILKRVKQYTLEHIMYLG
KLRHNDIDMTTVNTDDFSRLHAKEELDLELITFFASTNMELNKIFSRENINNDENIDFFGG
DREKNYVLDKKILNSKIKIIRDLDFIDNKNNITNNFIRKFTKIGTNERNRILHAISKERDLQG
TQDDYNKVINIIQNLKISDEEVSKALNLDVVFKDKKNIITKINDIKISEENNNDIKYLPSFSK
VLPEILNLYRNNPKNEPFDTIETEKIVLNALIYVNKELYKKLILEDDLEENESKNIFLQELK

KTLGNIDEIDENIIENYYKNAQISASKGNNKAIKKYQKKVIECYIGYLRKNYEELFDFSDFK
MNIQEIKKQIKDINDNKTYERITVKTSDKTIVINDDFEYIISIFALLNSNAVINKIRNRFFAT
SVWLNTSEYQNIIDILDEIMQLNTLRNECITENWNLNLEEFIQKMKEIEKDFDDFKIQTKK
EIFNNYYEDIKNNILTEFKDDINGCDVLEKKLEKIVIFDDETKFEIDKKSNILQDEQRKLSNI
NKKDLKKKVDQYIKDKDQEIKSKILCRIIFNSDFLKKYKKEIDNLIEDMESENENKFQEIYY
PKERKNELYIYKKNLFLNIGNPNFDKIYGLISNDIKMADAKFLFNIDGKNIRKNKISEIDAIL
KNLNDKLNGYSKEYKEKYIKKLKENDDFFAKNIQNKNYKSFEKDYNRVSEYKKIRDLVEF
NYLNKIESYLIDINWKLAIQMARFERDMHYIVNGLRELGIIKLSGYNTGISRAYPKRNGSD
GFYTTTAYYKFFDEESYKKFEKICYGFGIDLSENSEINKPENESIRNYISHFYIVRNPFADYS
IAEQIDRVSNLLSYSTRYNNSTYASVFEVFKKDVNLDYDELKKKFKLIGNNDILERLMKPK
KVSVLELESYNSDYIKNLIIELLTKIENTNDTL（配列番号27）

【０２６８】

C2c3, ＞CEPX01008730.1 海洋メタゲノムゲノムアセンブリTARA_037_MES_0.1-0.22, contig TARA_037_MES_0.1-0.22_scaffold22115_l, 全ゲノムショットガン配列から翻訳。

【０２６９】

[00269] MRSNYHGGRNARQWRKQISGLARRTKETVFTYKFPLETDAAEIDFDKAVQT
YGIAEGVGHGSLIGLVCAFHLSGFRLFSKAGEAMAFRNRSRYPTDAFAEKLSAIMGIQLPT
LSPEGLDLIFQSPPRSRDGIAPVWSENEVRNRLYTNWTGRGPANKPDEHLLEIAGEIAKQ
VFPKFGGWDDLASDPDKALAAADKYFQSQGDFPSIASLPAAIMLSPANSTVDFEGDYIAI
DPAAETLLHQAVSRCAARLGRERPDLDQNKGPFVSSLQDALVSSQNNGLSWLFGVGFQH
WKEKSPKELIDEYKVPADQHGAVTQVKSFVDAIPLNPLFDTTHYGEFRASVAGKVRSWV
ANYWKRLLDLKSLLATTEFTLPESISDPKAVSLFSGLLVDPQGLKKVADSLPARLVSAEEA
IDRLMGVGIPTAADIAQVERVADEIGAFIGQVQQFNNQVKQKLENLQDADDEEFLKGLKI
ELPSGDKEPPAINRISGGAPDAAAEISELEEKLQRLLDARSEHFQTISEWAEENAVTLDPIA
AMVELERLRLAERGATGDPEEYALRLLLQRIGRLANRVSPVSAGSIRELLKPVFMEEREFN
LFFHNRLGSLYRSPYSTSRHQPFSIDVGKAKAIDWIAGLDQISSDIEKALSGAGEALGDQL
RDWINLAGFAISQRLRGLPDTVPNALAQVRCPDDVRIPPLLAMLLEEDDIARDVCLKAFN
LYVSAINGCLFGALREGFIVRTRFQRIGTDQIHYVPKDKAWEYPDRLNTAKGPINAAVSSD
WIEKDGAVIKPVETVRNLSSTGFAGAGVSEYLVQAPHDWYTPLDLRDVAHLVTGLPVEK
NITKLKRLTNRTAFRMVGASSFKTHLDSVLLSDKIKLGDFTIIIDQHYRQSVTYGGKVKISY
EPERLQVEAAVPVVDTRDRTVPEPDTLFDHIVAIDLGERSVGFAVFDIKSCLRTGEVKPIH
DNNGNPVVGTVAVPSIRRLMKAVRSHRRRRQPNQKVNQTYSTALQNYRENVIGDVCNRI
DTLMERYNAFPVLEFQIKNFQAGAKQLEIVYGS（配列番号28）

【０２７０】

いくつかの態様において、本明細書において提供される融合蛋白質のいずれかの核酸によってプログラム可能なDNA結合蛋白質（napDNAbp）は古細菌（例えばナノ古細菌）からのCas9であり、これらは単細胞原核微生物のドメインおよび界を構成する。いくつかの態様において、napDNAbpはCasXまたはCasYである。これは例えばBurstein et al., "New CRISPR-Cas systems from uncultivated microbes." Cell Res. 2017 Feb 21. doi: 10.1038/cr.2017.21に記載されており、これは参照によって本明細書に組み込まれる。ゲノム分解的なメタゲノミクスを用いて、生き物の古細菌ドメインにおいて始めに報告されたCas9を包含するいくつものCRISPR-Casシステムが同定されている。この分岐したCas9蛋白質は活性なCRISPR-Casシステムの一部としてナノ古細菌において見出された。細菌では、2つの先に不明なシステムCRISPR-CasXおよびCRISPR-CasYが発見された。これらはかつて発見された最もコンパクトなシステムのうちである。いくつかの態様において、Cas9はCasXまたはCasXのバリアントを言う。いくつかの態様において、Cas9はCasYまたはCasYのバリアントを言う。他のRNAによってガイドされるDNA結合蛋白質が核酸によってプログラム可能なDNA結合蛋白質（napDNAbp）として用いられ得、本開示の範囲内であることは了解されるはずである。

【０２７１】

いくつかの態様において、本明細書において提供される融合蛋白質のいずれかの核酸によってプログラム可能なDNA結合蛋白質（napDNAbp）はCasXまたはCasY蛋白質である。いくつかの態様において、napDNAbpはCasX蛋白質である。いくつかの態様において、napDNAbpはCasY蛋白質である。いくつかの態様において、napDNAbpは、天然に生起するCasXまたはCasY蛋白質と少なくとも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%同一であるアミノ酸配列を含む。いくつかの態様において、napDNAbpは天然に生起するCasXまたはCasY蛋白質である。いくつかの態様において、napDNAbpは、配列番号29～31のいずれか1つと少なくとも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%同一であるアミノ酸配列を含む。いくつかの態様において、napDNAbpはいずれか1つの配列番号29～31のアミノ酸配列を含む。他の細菌種からのCasXおよびCasYもまた本開示に従って用いられ得ることは了解されるべきである。

CasX（uniprot.org/uniprot/F0NN87；uniprot.org/uniprot/F0NH53）

＞tr｜F0NN87｜F0NN87_SULIH CRISPR関連Casx蛋白質OS=Sulfolobus islandicus（株HVE10/4）GN=SiH_0402 PE=4 SV=1

MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRG
KAKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVS
APSFVKPEFYEFGRSPGMVERTRRVKLEVEPHYLIIAAAGWVLTRLGKAKVSEGDYVGVN
VFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSVVRIYTISDAVGQN
PTTINGGFSIDLTKLLEKRYLLSERLEAIARNALSISSNMRERYIVLANYIYEYLTGSKRLED
LLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG（配列番号29）

＞tr｜F0NH53｜F0NH53_SULIR CRISPR関連蛋白質Casx OS=Sulfolobus islandicus（株REY15A）GN=SiRe_0771 PE=4 SV=1

MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRG
KAKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVS
APSFVKPEFYKFGRSPGMVERTRRVKLEVEPHYLIMAAAGWVLTRLGKAKVSEGDYVGV
NVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSVVSIYTISDAVGQ
NPTTINGGFSIDLTKLLEKRDLLSERLEAIARNALSISSNMRERYIVLANYIYEYLTGSKRLE
DLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG（配列番号30）

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

＞APG80656.1 CRISPR関連蛋白質CasY [非培養Parcubacteria群細菌]

【０２７２】

[00272] MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIV
SAINDDYVGLYGLSNFDDLYNAEKRNEEKVYSVLDFWYDCVQYGAVFSYTAPGLLKNVA
EVRGGSYELTKTLKGSHLYDELQIDKVIKFLNKKEISRANGSLDKLKKDIIDCFKAEYRERH
KDQCNKLADDIKNAKKDAGASLGERQKKLFRDFFGISEQSENDKPSFTNPLNLTCCLLPF
DTVNNNRNRGEVLFNKLKEYAQKLDKNEGSLEMWEYIGIGNSGTAFSNFLGEGFLGRLRE
NKITELKKAMMDITDAWRGQEQEEELEKRLRILAALTIKLREPKFDNHWGGYRSDINGKL
SSWLQNYINQTVKIKEDLKGHKKDLKKAKEMINRFGESDTKEEAVVSSLLESIEKIVPDDS
ADDEKPDIPAIAIYRRFLSDGRLTLNRFVQREDVQEALIKERLEAEKKKKPKKRKKKSDAE
DEKETIDFKELFPHLAKPLKLVPNFYGDSKRELYKKYKNAAIYTDALWKAVEKIYKSAFSS
SLKNSFFDTDFDKDFFIKRLQKIFSVYRRFNTDKWKPIVKNSFAPYCDIVSLAENEVLYKP
KQSRSRKSAAIDKNRVRLPSTENIAKAGIALARELSVAGFDWKDLLKKEEHEEYIDLIELH
KTALALLLAVTETQLDISALDFVENGTVKDFMKTRDGNLVLEGRFLEMFSQSIVFSELRGL
AGLMSRKEFITRSAIQTMNGKQAELLYIPHEFQSAKITTPKEMSRAFLDLAPAEFATSLEP
ESLSEKSLLKLKQMRYYPHYFGYELTRTGQGIDGGVAENALRLEKSPVKKREIKCKQYKT

LGRGQNKIVLYVRSSYYQTQFLEWFLHRPKNVQTDVAVSGSFLIDEKKVKTRWNYDALT
VALEPVSGSERVFVSQPFTIFPEKSAEEEGQRYLGIDIGEYGIAYTALEITGDSAKILDQNFI
SDPQLKTLREEVKGLKLDQRRGTFAMPSTKIARIRESLVHSLRNRIHHLALKHKAKIVYEL
EVSRFEEGKQKIKKVYATLKKADVYSEIDADKNLQTTVWGKLAVASEISASYTSQFCGAC
KKLWRAEMQVDETITTQELIGTVRVIKGGTLIDAIKDFMRPPIFDENDTPFPKYRDFCDK
HHISKKMRGNSCLFICPFCRANADADIQASQTIALLRYVKEEKKVEDYFERFRKLKNIKVL
GQMKKI（配列番号31）

【０２７３】

核酸塩基編集因子のCas9ドメイン

限定しない例示的なCas9ドメインが本明細書において提供される。Cas9ドメインはヌクレアーゼ活性型Cas9ドメイン、ヌクレアーゼ不活性型Cas9ドメイン、またはCas9ニッカーゼであり得る。いくつかの態様において、Cas9ドメインはヌクレアーゼ活性型ドメインである。例えば、Cas9ドメインは二重鎖核酸の両方の鎖（例えば、二重鎖DNA分子の両方の鎖）を切るCas9ドメインであり得る。いくつかの態様において、Cas9ドメインは、いずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つを含む。いくつかの態様において、Cas9ドメインは、いずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つと少なくとも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%同一であるアミノ酸配列を含む。いくつかの態様において、Cas9ドメインは、いずれかのCas9蛋白質、例えば本明細書において提供されるCas9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、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、またはより多くの変異を有するアミノ酸配列を含む。いくつかの態様において、Cas9ドメインは、いずれかのCas9蛋白質、例えば本明細書において提供されるCas9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個の同一の一続きのアミノ酸残基を有するアミノ酸配列を含む。

【０２７４】

いくつかの態様において、Cas9ドメインはヌクレアーゼ不活性型Cas9ドメイン（dCas9）である。例えば、dCas9ドメインは、二重鎖核酸分子のどちらかの鎖を切断することなしに二重鎖核酸分子に（例えばgRNA分子を介して）結合し得る。いくつかの態様において、ヌクレアーゼ不活性型dCas9ドメインは、配列番号6によって規定されるアミノ酸配列のD10X変異およびH840X変異、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異を含み、Xはいずれかのアミノ酸変化である。いくつかの態様において、ヌクレアーゼ不活性型dCas9ドメインは、配列番号6によって規定されるアミノ酸配列のD10A変異およびH840A変異、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異を含む。1つの例として、ヌクレアーゼ不活性型Cas9ドメインは配列番号32によって規定されるアミノ酸配列を含む（クローニングベクターpPlatTET-gRNA2、アクセッションNo.BAV54124）。

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEA
TRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVD
KLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIAL

SLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDI
LRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG
GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ
EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLIN
GIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA
GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGI
KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLK
DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRK
DFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ
EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK
VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
KRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII
EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID
RKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD（配列番号32；例えば、Qi et al, Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell. 2013; 152(5): 1173-83を見よ。これの内容全体は参照によって本明細書に組み込まれる）。

【０２７５】

追加の好適なヌクレアーゼ不活性型dCas9ドメインは本開示の範囲内であり、本開示および当分野の知識に基づいて当業者には明らかであろう。かかる追加の例示的な好適なヌクレアーゼ不活性型Cas9ドメインは D10A/H840A、D10A/D839A/H840A、およびD10A/D839A/H840A/N863A変異体ドメインを包含するが、これらに限定されない（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を見よ。これの内容全体は参照によって本明細書に組み込まれる）。いくつかの態様において、dCas9ドメインは、本明細書において提供されるdCas9ドメインのいずれか1つと少なくとも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%同一であるアミノ酸配列を含む。いくつかの態様において、Cas9ドメインは、いずれかのCas9蛋白質、例えば本明細書において提供されるCas9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、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以上、またはより多くの変異を有するアミノ酸配列を含む。いくつかの態様において、Cas9ドメインは、いずれかのCas9蛋白質、例えば本明細書において提供されるCas9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個の同一の一続きのアミノ酸残基を有するアミノ酸配列を含む。

【０２７６】

いくつかの態様において、Cas9ドメインはCas9ニッカーゼである。Cas9ニッカーゼは二重鎖核酸分子（例えば二重鎖DNA分子）の1つの鎖のみを切断することができるCas9蛋

白質であり得る。いくつかの態様において、Cas9ニッカーゼは二重鎖核酸分子の標的鎖を切断し、Cas9ニッカーゼが、Cas9に結合しているgRNA（例えばsgRNA）と塩基対形成している（に対して相補的である）鎖を切断することを意味する。いくつかの態様において、Cas9ニッカーゼはD10A変異を含み、配列番号6の位置840のヒスチジン、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの変異を有する。例えば、Cas9ニッカーゼは配列番号8によって規定されるアミノ酸配列を含み得る。いくつかの態様において、Cas9ニッカーゼは二重鎖の核酸分子の塩基編集されない非標的鎖を切断し、Cas9ニッカーゼが、Cas9に結合しているgRNA（例えばsgRNA）と塩基対形成していない鎖を切断することを意味する。いくつかの態様において、Cas9ニッカーゼはH840A変異を含み、配列番号6の位置10のアスパラギン酸残基、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異を有する。いくつかの態様において、Cas9ニッカーゼは、本明細書において提供されるCas9ニッカーゼのいずれか1つと少なくとも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%同一であるアミノ酸配列を含む。追加の好適なCas9ニッカーゼは本開示の範囲内であり、本開示および当分野の知識に基づいて当業者には明らかであろう。

【０２７７】

縮減されたPAM排他性を有するCas9ドメイン

本開示のいくつかの側面は異なるPAM特異性を有するCas9ドメインを提供する。典型的には、S. pyogenesからのCas9（spCas9）などのCas9蛋白質は特定の核酸領域に結合するためには古典的NGG PAM配列を要求する。これはゲノム中の所望の塩基を編集する能力を限定し得る。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編集融合蛋白質は正確な場所に置かれることを必要とし得、例えば、そこでは、標的塩基は、PAMのおよそ15塩基上流である4塩基領域（例えば「脱アミノ化ウィンドウ」）内に置かれる。Komor, A.C., et al., "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016)を見よ。これの内容全体はここに参照によって組み込まれる。従って、いくつかの態様において、本明細書において提供される融合蛋白質のいずれかは、古典的（例えばNGG）PAM配列を含有しないヌクレオチド配列に結合することができるCas9ドメインを含有し得る。非古典的PAM配列に結合するCas9ドメインは当分野において記載されており、当業者には明らかであろう。例えば、非古典的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)に記載されており；それぞれの内容全体は参照によってここに組み込まれる。

【０２７８】

いくつかの態様において、Cas9ドメインは Staphylococcus aureusからのCas9ドメイン（SaCas9）である。いくつかの態様において、SaCas9ドメインはヌクレアーゼ活性型SaCas9、ヌクレアーゼ不活性型SaCas9（SaCas9d）、またはSaCas9ニッカーゼ（SaCas9n）である。いくつかの態様において、SaCas9はアミノ酸配列配列番号33を含む。いくつかの態様において、SaCas9は、配列番号33のN579X変異、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異を含み、XはNを除くいずれかのアミノ酸である。いくつかの態様において、SaCas9は、配列番号33のN579A変異、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異を含む。いくつかの態様において、SaCas9ドメイン、SaCas9dドメイン、またはSaCas9nドメインは非古典的PAMを有する核酸配列に結合し得る。いくつかの態様において、SaCas9ドメイン、SaCas9dドメイン、またはSaCas9nドメインは、NNGRRT PAM配列を有する核酸配列に結合し得

る。いくつかの態様において、SaCas9ドメインは、配列番号33のE781X、N967X、およびR1014X変異の1つ以上、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異を含み、Xはいずれかのアミノ酸である。いくつかの態様において、SaCas9ドメインは、配列番号33のE781K、N967K、およびR1014H変異の1つ以上、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの1つ以上の対応する変異を含む。いくつかの態様において、SaCas9ドメインは、配列番号33のE781K、N967K、もしくはR1014H変異、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異を含む。

【０２７９】

いくつかの態様において、本明細書において提供される融合蛋白質のいずれかのCas9ドメインは、配列番号33～36のいずれか1つと少なくとも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%同一であるアミノ酸配列を含む。いくつかの態様において、本明細書において提供される融合蛋白質のいずれかのCas9ドメインは配列番号33～36のいずれか1つのアミノ酸配列を含む。いくつかの態様において、本明細書において提供される融合蛋白質のいずれかのCas9ドメインは配列番号33～36のいずれか1つのアミノ酸配列からなる。

例示的なSaCas9配列

【化３０】

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRR
RRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRG
VHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSD
YVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEM
LMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFK
QKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLD
QIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHT
NDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLP
NDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQ
EGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPF
QYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVD
TRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAED
ALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHI
KDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLI
NKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPV
IKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDV
IKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRI
EVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKK
G

（配列番号33）

下線付きかつ太字である配列番号33の残基N579は、SaCas9ニッカーゼを生むために（例えばA579に）変異させられ得る。

例示的なSaCas9d配列

【化３１】

KRNYILGL**A**IGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRR
RRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRG
VHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSD
YVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEM
LMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFK
QKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLD
QIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHT
NDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLP
NDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQ
EGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPF
QYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVD
TRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAED
ALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHI
KDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLI
NKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPV
IKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDV
IKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRI
EVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKK
G

（配列番号34）

下線付きかつ太字である配列番号34の残基D10は、ヌクレアーゼ不活性型SaCas9dを生むために（例えばA10に）変異させられ得る。

例示的なSaCas9n配列

【化３２】

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRR
RRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRG
VHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSD
YVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEM
LMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFK
QKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLD
QIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHT
NDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLP
NDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQ
EGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEE**A**SKKGNRTPF
QYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVD

【化３３】

TRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAED
ALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHI
KDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLI
NKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPV
IKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDV
IKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRI
EVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKK
G

（配列番号35）。

配列番号35の残基A579は下線付きかつ太字であり、これはSaCas9ニッカーゼを生むために配列番号33のN579から変異させられ得る。

例示的なSaKKH Cas9

【化３４】

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRR
RRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRG
VHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSD
YVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEM
LMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFK
QKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLD
QIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHT
NDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLP
NDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQ
EGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEE**A**SKKGNRTPF
QYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVD
TRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAED
ALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHI
KDFKDYKYSHRVDKKPNR*K*LINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLI
NKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPV
IKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDV
IKKENYYEVNSKCYEEAKKLKKISNQAEFIASFY*K*NDLIKINGELYRVIGVNNDLLNRI
EVNMIDITYREYLENMNDKRPP*H*IIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKK
G

（配列番号36）。

配列番号36の残基A579は下線付きかつ太字であり、これはSaCas9ニッカーゼを生むために配列番号36のN579から変異させられ得る。配列番号36の残基K781、K967、およびH1014は下線付きかつ斜体であり、これらはSaKKH Cas9を生むために配列番号36のE781、N967、およびR1014から変異させられ得る。

【０２８０】

いくつかの態様において、Cas9ドメインは*Streptococcus pyogenes*からのCas9ドメ

イン（SpCas9）である。いくつかの態様において、SpCas9ドメインはヌクレアーゼ活性型SpCas9、ヌクレアーゼ不活性型SpCas9（SpCas9d）、またはSpCas9ニッカーゼ（SpCas9n）である。いくつかの態様において、SpCas9はアミノ酸配列配列番号37を含む。いくつかの態様において、SpCas9は、配列番号37のD9X変異、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異を含み、XはDを除くいずれかのアミノ酸である。いくつかの態様において、SpCas9は、配列番号37のD9A変異、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異を含む。いくつかの態様において、SpCas9ドメイン、SpCas9dドメイン、またはSpCas9nドメインは非古典的PAMを有する核酸配列に結合し得る。いくつかの態様において、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し得る。いくつかの態様において、SpCas9ドメインは、配列番号37のD1134X、R1334X、およびT1336X変異の1つ以上、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異を含み、Xはいずれかのアミノ酸である。いくつかの態様において、SpCas9ドメインは、配列番号37のD1134E、R1334Q、およびT1336R変異の1つ以上、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異を含む。いくつかの態様において、SpCas9ドメインは、配列番号37のD1134E、R1334Q、およびT1336R変異、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異を含む。いくつかの態様において、SpCas9ドメインは、配列番号37のD1134X、R1334X、およびT1336X変異の1つ以上、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異を含み、Xはいずれかのアミノ酸である。いくつかの態様において、SpCas9ドメインは、配列番号37のD1134V、R1334Q、およびT1336R変異の1つ以上、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異を含む。いくつかの態様において、SpCas9ドメインは、配列番号37のD1134V、R1334Q、およびT1336R変異、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異を含む。いくつかの態様において、SpCas9ドメインは、配列番号37のD1134X、G1217X、R1334X、およびT1336X変異の1つ以上、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異を含み、Xはいずれかのアミノ酸である。いくつかの態様において、SpCas9ドメインは、配列番号37のD1134V、G1217R、R1334Q、およびT1336R変異の1つ以上、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異を含む。いくつかの態様において、SpCas9ドメインは、配列番号37のD1134V、G1217R、R1334Q、およびT1336R変異、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異を含む。

【０２８１】

いくつかの態様において、本明細書において提供される融合蛋白質のいずれかのCas9ドメインは、配列番号37～41のいずれか1つと少なくとも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%同一であるアミノ酸配列を含む。いくつかの態様において、本明細書において提供される融合蛋白質のいずれかのCas9ドメインは配列番号37～41のいずれか1つのアミノ酸配列を含む。いくつかの態様において、本明細書において提供される融合蛋白質のいずれかのCas9ドメインは配列番号37～41のいずれか1つのアミノ酸配列からなる。

例示的なSpCas9

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIV
DEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDK

LFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL
RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGG
ASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQE
DFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLIN
GIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA
GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGI
KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK
DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRK
DFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ
EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK
VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
KRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII
EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID
RKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD（配列番号37）

例示的なSpCas9n

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIV
DEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDK
LFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL
RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGG
ASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQE
DFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLIN
GIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA
GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGI
KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK
DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRK
DFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ
EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK
VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
KRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII
EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID
RKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD（配列番号38）

例示的なSpEQR Cas9

【化３５】

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP
DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN
GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA
KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA
YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGF**E**SPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**Q**
Y**R**STKEVLDATLIHQSITGLYETRIDLSQLGGD

（配列番号39）

配列番号39の残基E1134、Q1334、およびR1336は下線付きかつ太字であり、これらはSpEQR Cas9を生むために配列番号39のD1134、R1334、およびT1336から変異させられ得る。

例示的なSpVQR Cas9

【化３６】

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP
DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN

【化３７】

GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA
KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA
YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGF**V**SPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**Q**
Y**R**STKEVLDATLIHQSITGLYETRIDLSQLGGD

（配列番号４０）

配列番号４０の残基Ｖ１１３４、Ｑ１３３４、およびＲ１３３６は下線付きかつ太字であり、これらはＳｐＶＱＲ Ｃａｓ９を生むために配列番号４０のＤ１１３４、Ｒ１３３４、およびＴ１３３６から変異させられ得る。

例示的なＳｐＶＲＥＲ Ｃａｓ９

【化３８】

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP
DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN
GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA
KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED

【化３９】

RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA
YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGF**V**SPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA**R**
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**E**
Y**R**STKEVLDATLIHQSITGLYETRIDLSQLGGD

（配列番号41）

配列番号41の残基V1134、R1217、Q1334、およびR1336は下線付きかつ太字であり、これらはSpVRER Cas9を生むために配列番号41のD1134、G1217、R1334、およびT1336から変異させられ得る。

【０２８２】

次は、非古典的（例えば非NGG）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な融合蛋白質（例えば塩基編集蛋白質）である：

例示的なSaBE3 (rAPOBEC1-XTEN-SaCas9n-UGI-NLS)

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLE
LYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESATP
ESKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRR
HRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNV
NEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQ
LLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFP
EELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKE
ILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQ
EELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPK
KVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQK
MINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNP
FNYEVDHIIPRSVSFDNSFNNKVLVKQEEASKKGNRTPFQYLSSSDSKISYETFKKHILNLA
KGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVK
VKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQ
MFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTR
KDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEK
NPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKP
YRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLI

KINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDI
LGNLYEVKSKKHPQIIKKGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDI
LVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV（配列
番号42）
例示的なSaKKH-BE3 (rAPOBEC1-XTEN-SaCas9n-UGI-NLS)
MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLE
LYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESATP
ESKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRR
HRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNV
NEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQ
LLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFP
EELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKE
ILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQ
EELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPK
KVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQK
MINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNP
FNYEVDHIIPRSVSFDNSFNNKVLVKQEEASKKGNRTPFQYLSSSDSKISYETFKKHILNLA
KGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVK
VKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQ
MFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTR
KDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEK
NPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKP
YRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYKNDLI
KINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDI
LGNLYEVKSKKHPQIIKKGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDI
LVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV（配列
番号43）
例示的なEQR-BE3 (rAPOBEC1-XTEN-Cas9n-UGI-NLS)
MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLE
LYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESATP
ESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGN
IVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV
DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIA
LSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYID
GGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRR
QEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKG
ASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQK
KAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLI
NGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL
AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEG
IKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG
GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR

10 20 30 40 50

KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFESPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESIL
MLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKM
LSGGSPKKKRKV（配列番号44）

VQR-BE3 (rAPOBEC1-XTEN-Cas9n-UGI-NLS)

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLE
LYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESATP
ESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGN
IVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV
DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIA
LSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYID
GGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRR
QEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKG
ASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQK
KAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLI
NGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL
AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEG
IKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG
GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESIL
MLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKM
LSGGSPKKKRKV（配列番号45）

VRER-BE3 (rAPOBEC1-XTEN-Cas9n-UGI-NLS)

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLE
LYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESATP
ESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGN
IVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV
DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIA
LSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYID

GGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRR
QEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKG
ASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQK
KAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLI
NGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL
AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEG
IKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG
GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESIL
MLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKM
LSGGSPKKKRKV（配列番号46）

【０２８３】

高フィデリティ塩基編集因子

本開示のいくつかの側面は、高いフィデリティを有するCas9ドメインを含むCas9融合蛋白質（例えば本明細書において提供される融合蛋白質のいずれか）を提供する。本開示の追加の側面は、野生型Cas9ドメインと比較して、Cas9ドメインとDNAの糖-リン酸バックボーンとの間の減少した静電相互作用を有するCas9ドメインを含むCas9融合蛋白質（例えば本明細書において提供される融合蛋白質のいずれか）を提供する。いくつかの態様において、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-リン酸バックボーンとの間の結びつきを減少させる1つ以上の変異を含む。いくつかの態様において、本明細書において提供されるCas9融合蛋白質のいずれかは、配列番号6によって提供されるアミノ酸配列のN497X、R661X、Q695X、および／またはQ926X変異の1つ以上、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異を含み、Xはいずれかのアミノ酸である。いくつかの態様において、本明細書において提供されるCas9融合蛋白質のいずれかは、配列番号6によって提供されるアミノ酸配列N497A、R661A、Q695A、および／またはQ926A変異の1つ以上、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異を含む。いくつかの態様において、Cas9ドメインは、配列番号6によって提供されるアミノ酸配列のD10A変異、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異を含む。いくつかの態様において、（例えば、本明細書において提供される融合蛋白質のいずれかの）Cas9ドメインは配列番号47によって規定されるアミノ酸配列を含む。いくつかの態様において、融合蛋白質は配列番号48によって規定されるアミノ酸配列を含む。高いフィデリティを有するCas9ドメインは当分野において公知であり、当業者には明らかであろう。例えば、高いフィデリティを有する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）に記載されており；それぞれの内容全体は参照によって本明細書に組み込まれる。

【０２８４】

本明細書において提供される塩基編集因子、例えば塩基編集因子2（BE2）または塩基

編集因子3（BE3）は、本明細書に記載される通りCas9ドメインを改変することによって高フィデリティ塩基編集因子へと変換されて、高フィデリティ塩基編集因子、例えば高フィデリティ塩基編集因子2（HF-BE2）または高フィデリティ塩基編集因子3（HF-BE3）を創出し得ることは了解されるべきである。いくつかの態様において、塩基編集因子2（BE2）はデアミナーゼドメイン、dCas9、およびUGIドメインを含む。いくつかの態様において、塩基編集因子3（BE3）はデアミナーゼドメイン、nCas9ドメイン、およびUGIドメインを含む。

配列番号6のCas9に対して相対的な変異が太字および下線で示されているCas9ドメイン

【化４０】

DKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR
YTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKK
LVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAI
LSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQ
IGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL
RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIER
MT**A**FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVK
QLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWG**A**LSRKLINGIRDKQSGKTILDFLKSDGFANRNFM**A**LIHDDSL
TFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQ
KGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG

【化４１】

LSELDKAGFIKRQLVETR**A**ITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVRE
INNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNF
FKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAK
GYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQ
KQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

（配列番号47）

HF-BE3

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLE
LYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESATP
ESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGN
IVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV
DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIA
LSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYID
GGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRR

QEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKG
ASAQSFIERMTAFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQK
KAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGALSRKLI
NGIRDKQSGKTILDFLKSDGFANRNFMALIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL
AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEG
IKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG
GLSELDKAGFIKRQLVETRAITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD（配列番号48）

【０２８５】

Cas9融合蛋白質

本明細書において開示されるCas9ドメインのいずれか（例えば、ヌクレアーゼ活性型Cas9蛋白質、ヌクレアーゼ不活性型dCas9蛋白質、またはCas9ニッカーゼ蛋白質）は第2の蛋白質に融合させられ得、それゆえに、本明細書において提供される融合蛋白質は、本明細書において提供されるCas9ドメインと第2の蛋白質または「融合パートナー」とを含む。いくつかの態様において、第2の蛋白質はCas9ドメインのN末端に融合させられる。しかしながら、他の態様においては、第2の蛋白質はCas9ドメインのC末端に融合させられる。いくつかの態様において、Cas9ドメインに融合させられる第2の蛋白質は核酸編集ドメインである。いくつかの態様において、Cas9ドメインおよび核酸編集ドメインはリンカーを介して融合させられるが、他の態様においては、Cas9ドメインおよび核酸編集ドメインは直接的に互いに融合させられる。いくつかの態様において、Cas9ドメインおよび核酸編集ドメインはいずれかの長さまたは組成のリンカーを介して融合させられる。例えば、リンカーはいずれかの長さおよび組成の結合、1つ以上のアミノ酸、ペプチド、またはポリマーであり得る。いくつかの態様において、リンカーは$(GGGS)_n$（配列番号613）、$(GGGGS)_n$（配列番号607）、$(G)_n$（配列番号608）、$(EAAAK)_n$（配列番号609）、$(GGS)_n$（配列番号610）、$(SGGS)_n$（配列番号606）、SGSETPGTSESATPES（配列番号604）、$SGGS(GGS)_n$（配列番号612）、SGGSSGGSSGSETPGTSESATPESSGGSSGGS（配列番号605）、もしくは$(XP)_n$（配列番号611）モチーフ、またはこれらのいずれかの組み合わせを含み、nは独立して1および30の間の整数であり、Xはいずれかのアミノ酸である。いくつかの態様において、リンカーは$(GGS)_n$モチーフを含み、nは1、3、または7である。いくつかの態様において、リンカーは$(GGS)_n$（配列番号610）モチーフを含み、nは1、2、3、4、5、6、7、8、9、10、11、12、13、14、または15である。いくつかの態様において、リンカーはアミノ酸配列$SGGS(GGS)_n$（配列番号612）を含み、nは1、2、3、4、5、6、7、8、9、または10である。いくつかの態様において、リンカーはアミノ酸配列$SGGS(GGS)_n$（配列番号612）を含み、nは2である。いくつかの態様において、リンカーはSGSETPGTSESATPES（配列番号604）のアミノ酸配列を含み、例においてはXTENリンカーともまた言われる）。いくつかの態様において、リンカーは、例において32アミノ酸リンカーともまた言われるアミノ酸配列SGGSSGGSSGSETPGTSESATPESSGGSSGGS（配列番号605）を含む。例において例示される通り、リンカーの長さは編集されるべき塩基に影響を及ぼし得る。例えば、3アミノ酸の長さのリンカー（例えば$(GGS)_1$）はPAM配列に対して相対的に2～5、2～4、2～3、3～4塩基編集ウィンドウを与え得、9アミノ酸リンカー（例えば$(GGS)_3$（配列番号610））はPAM配列に対して相対的に2～6、2～5、2～4、2～3、3～6、3～5、3～4、4～6、4～5、5～6塩基編集ウィンド

ウを与え得る。16アミノ酸リンカー（例えばXTENリンカー）はPAM配列に対して相対的に2～7、2～6、2～5、2～4、2～3、3～7、3～6、3～5、3～4、4～7、4～6、4～5、5～7、5～6、6～7塩基ウィンドウを与え得、例外的に強い活性を有し、21アミノ酸リンカー（例えば$(GGS)_7$（配列番号610））はPAM配列に対して相対的に3～8、3～7、3～6、3～5、3～4、4～8、4～7、4～6、4～5、5～8、5～7、5～6、6～8、6～7、7～8塩基編集ウィンドウを与え得る。様々なリンカー長さは、本開示のdCas9融合蛋白質がPAM配列から異なる距離の核酸塩基を編集することを許し得るという新規の知見は、有意な臨床的な重要性を提供する。なぜなら、PAM配列は、遺伝子内の修正されるべき疾患を引き起こす変異から様々な距離であり得るからである。ここで例として記載されているリンカー長さは限定的であることを意味されないことは理解されるべきである。

【０２８６】

いくつかの態様において、第2の蛋白質は酵素ドメインを含む。いくつかの態様において、酵素ドメインは核酸編集ドメインである。かかる核酸編集ドメインは、限定なしに、ヌクレアーゼ、ニッカーゼ、リコンビナーゼ、デアミナーゼ、メチルトランスフェラーゼ、メチラーゼ、アセチラーゼ、またはアセチルトランスフェラーゼであり得る。本開示に従って用いられ得る限定しない例示的な結合ドメインは、転写活性化因子ドメインおよび転写リプレッサードメインを包含する。

【０２８７】

デアミナーゼドメイン

いくつかの態様において、第2の蛋白質は核酸編集ドメインを含む。いくつかの態様において、核酸編集ドメインはC→U塩基変化を触媒し得る。いくつかの態様において、核酸編集ドメインはデアミナーゼドメインである。いくつかの態様において、デアミナーゼはシチジンデアミナーゼまたはシチジンデアミナーゼである。いくつかの態様において、デアミナーゼはアポリポ蛋白質B mRNA編集複合体（APOBEC）ファミリーデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC1デアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC2デアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC3デアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC3Aデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC3Bデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC3Cデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC3Dデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC3Eデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC3Fデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC3Gデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC3Hデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC4デアミナーゼである。いくつかの態様において、デアミナーゼは活性化誘導デアミナーゼ（AID）である。いくつかの態様において、デアミナーゼは脊椎動物デアミナーゼである。いくつかの態様において、デアミナーゼは無脊椎動物デアミナーゼである。いくつかの態様において、デアミナーゼはヒト、チンパンジー、ゴリラ、サル、ウシ(cow)、イヌ、ラット、またはマウスデアミナーゼである。いくつかの態様において、デアミナーゼはヒトデアミナーゼである。いくつかの態様において、デアミナーゼはラットデアミナーゼ、例えばrAPOBEC1である。いくつかの態様において、デアミナーゼは活性化誘導シチジンデアミナーゼ（AID）である。いくつかの態様において、デアミナーゼはシチジンデアミナーゼ1（CDA1）である。いくつかの態様において、デアミナーゼはPetromyzon marinusシチジンデアミナーゼ1（pmCDA1）である。いくつかの態様において、デアミナーゼはヒトAPOBEC3G（配列番号60）である。いくつかの態様において、デアミナーゼはヒトAPOBEC3Gの断片（配列番号83）である。いくつかの態様において、デアミナーゼはD316R_D317R変異を含むヒトAPOBEC3Gバリアント（配列番号82）である。いくつかの態様において、デアミナーゼはヒトAPOBEC3Gの断片であり、配列番号60のD316R_D317R変異に対応する変異を含んでいる（配列番号84）。

【０２８８】

いくつかの態様において、核酸編集ドメインは、配列番号49～84のいずれか1つのデアミナーゼドメインと少なくとも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%同一である。いくつかの態様において、核酸編集ドメインは配列番号49～84のいずれか1つのアミノ酸配列を含む。

【０２８９】

塩基編集因子の編集ウィンドウを調節するデアミナーゼドメイン

本開示のいくつかの側面は、例えばデアミナーゼドメインに点変異を作ることによって、本明細書において提供される融合蛋白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、融合蛋白質（例えば塩基編集因子）の処理能力に影響するという認識に基づく。例えば、塩基編集融合蛋白質内のデアミナーゼドメインの触媒活性を縮減するが消去はしない変異は、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ことをより蓋然的でなくし得、それによって脱アミノ化ウィンドウを狭める。脱アミノ化ウィンドウを狭める能力は、特異的な標的残基に隣接する残基の望まれない脱アミノ化を防止し得、これはオフターゲット効果を減少させるかまたは防止し得る。

【０２９０】

いくつかの態様において、本明細書において提供される融合蛋白質のいずれかは、縮減されたデアミナーゼ触媒活性を有するデアミナーゼドメイン（例えばシチジンデアミナーゼドメイン）を含む。いくつかの態様において、本明細書において提供される融合蛋白質のいずれかは、適切な対照と比較して縮減されたデアミナーゼ触媒活性を有するデアミナーゼドメイン（例えばシチジンデアミナーゼドメイン）を含む。例えば、適切な対照は、1つ以上の変異をデアミナーゼに導入することに先立つデアミナーゼのデアミナーゼ活性であり得る。他の態様において、適切な対照は野生型デアミナーゼであり得る。いくつかの態様において、適切な対照は野生型のアポリポ蛋白質B mRNA編集複合体（APOBEC）ファミリーデアミナーゼである。いくつかの態様において、適切な対照はAPOBEC1デアミナーゼ、APOBEC2デアミナーゼ、APOBEC3Aデアミナーゼ、APOBEC3Bデアミナーゼ、APOBEC3Cデアミナーゼ、APOBEC3Dデアミナーゼ、APOBEC3Fデアミナーゼ、APOBEC3Gデアミナーゼ、またはAPOBEC3Hデアミナーゼである。いくつかの態様において、適切な対照は活性化誘導デアミナーゼ（AID）である。いくつかの態様において、適切な対照はPetromyzon marinusからのシチジンデアミナーゼ1（pmCDA1）である。いくつかの態様において、デアミナーゼドメインは、適切な対照と比較して、少なくとも1%、少なくとも5%、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40%、少なくとも50%、少なくとも60%、少なくと70%、少なくとも80%、少なくとも90%、または少なくとも95%少ないデアミナーゼ触媒活性を有するデアミナーゼドメインであり得る。

【０２９１】

いくつかの態様において、本明細書において提供される融合蛋白質のいずれかは、rAPOBEC1（配列番号76）のH121X、H122X、R126X、R126X、R118X、W90X、W90X、およびR132Xからなる群から選択される1つ以上の変異、または別のAPOBECデアミナーゼの1つ以上の対応する変異を含むAPOBECデアミナーゼを含み、Xはいずれかのアミノ酸である。いくつかの態様において、本明細書において提供される融合蛋白質のいずれかは、rAPOBEC1（配列番号76）のH121R、H122R、R126A、R126E、R118A、W90A、W90Y、およびR132Eからなる群から選択される1つ以上の変異、または別のAPOBECデアミナーゼの1つ以上の対応する変異を含むAPOBECデアミナーゼを含む。

【０２９２】

いくつかの態様において、本明細書において提供される融合蛋白質のいずれかは、hAPOBEC3G（配列番号60）のD316X、D317X、R320X、R320X、R313X、W285X、W285X、R326Xからなる群から選択される1つ以上の変異、または別のAPOBECデアミナーゼの1つ以上の対応する変異を含むAPOBECデアミナーゼを含み、Xはいずれかのアミノ酸である。いくつかの態様において、本明細書において提供される融合蛋白質のいずれか

は、hAPOBEC3G（配列番号60）のD316R、D317R、R320A、R320E、R313A、W285A、W285Y、R326Eからなる群から選択される1つ以上の変異、または別のAPOBECデアミナーゼの1つ以上の対応する変異を含むAPOBECデアミナーゼを含む。

【０２９３】

いくつかの態様において、本明細書において提供される融合蛋白質のいずれかは、rAPOBEC1（配列番号76）のH121RおよびH122R変異、または別のAPOBECデアミナーゼの1つ以上の対応する変異を含むAPOBECデアミナーゼを含む。いくつかの態様において、本明細書において提供される融合蛋白質のいずれかは、rAPOBEC1（配列番号76）のR126A変異、または別のAPOBECデアミナーゼの1つ以上の対応する変異を含むAPOBECデアミナーゼを含む。いくつかの態様において、本明細書において提供される融合蛋白質のいずれかは、rAPOBEC1（配列番号76）のR126E変異、または別のAPOBECデアミナーゼの1つ以上の対応する変異を含むAPOBECデアミナーゼを含む。いくつかの態様において、本明細書において提供される融合蛋白質のいずれかは、rAPOBEC1（配列番号76）のR118A変異、または別のAPOBECデアミナーゼの1つ以上の対応する変異を含むAPOBECデアミナーゼを含む。いくつかの態様において、本明細書において提供される融合蛋白質のいずれかは、rAPOBEC1（配列番号76）のW90A変異、または別のAPOBECデアミナーゼの1つ以上の対応する変異を含むAPOBECデアミナーゼを含む。いくつかの態様において、本明細書において提供される融合蛋白質のいずれかは、rAPOBEC1（配列番号76）のW90Y変異、または別のAPOBECデアミナーゼの1つ以上の対応する変異を含むAPOBECデアミナーゼを含む。いくつかの態様において、本明細書において提供される融合蛋白質のいずれかは、rAPOBEC1（配列番号76）のR132E変異、または別のAPOBECデアミナーゼの1つ以上の対応する変異を含むAPOBECデアミナーゼを含む。いくつかの態様において、本明細書において提供される融合蛋白質のいずれかは、rAPOBEC1（配列番号76）のW90YおよびR126E変異、または別のAPOBECデアミナーゼの1つ以上の対応する変異を含むAPOBECデアミナーゼを含む。いくつかの態様において、本明細書において提供される融合蛋白質のいずれかは、rAPOBEC1（配列番号76）のR126EおよびR132E変異、または別のAPOBECデアミナーゼの1つ以上の対応する変異を含むAPOBECデアミナーゼを含む。いくつかの態様において、本明細書において提供される融合蛋白質のいずれかは、rAPOBEC1（配列番号76）のW90YおよびR132E変異、または別のAPOBECデアミナーゼの1つ以上の対応する変異を含むAPOBECデアミナーゼを含む。いくつかの態様において、本明細書において提供される融合蛋白質のいずれかは、rAPOBEC1（配列番号76）のW90Y、R126E、およびR132E変異、または別のAPOBECデアミナーゼの1つ以上の対応する変異を含むAPOBECデアミナーゼを含む。

【０２９４】

いくつかの態様において、本明細書において提供される融合蛋白質のいずれかは、hAPOBEC3G（配列番号60）のD316RおよびD317R変異、または別のAPOBECデアミナーゼの1つ以上の対応する変異を含むAPOBECデアミナーゼを含む。いくつかの態様において、本明細書において提供される融合蛋白質のいずれかは、hAPOBEC3G（配列番号60）のR320A変異、または別のAPOBECデアミナーゼの1つ以上の対応する変異を含むAPOBECデアミナーゼを含む。いくつかの態様において、本明細書において提供される融合蛋白質のいずれかは、hAPOBEC3G（配列番号60）のR320E変異、または別のAPOBECデアミナーゼの1つ以上の対応する変異を含むAPOBECデアミナーゼを含む。いくつかの態様において、本明細書において提供される融合蛋白質のいずれかは、hAPOBEC3G（配列番号60）のR313A変異、または別のAPOBECデアミナーゼの1つ以上の対応する変異を含むAPOBECデアミナーゼを含む。いくつかの態様において、本明細書において提供される融合蛋白質のいずれかは、hAPOBEC3G（配列番号60）のW285A変異、または別のAPOBECデアミナーゼの1つ以上の対応する変異を含むAPOBECデアミナーゼを含む。いくつかの態様において、本明細書において提供される融合蛋白質のいずれかは、hAPOBEC3G（配列番号60）のW285Y変異、または別のAPOBECデアミナーゼの1つ以上の対応する変異を含むAPOBECデアミナーゼを含む。いくつかの態様において、本明細書において提供さ

れる融合蛋白質のいずれかは、hAPOBEC3G（配列番号60）のR326E変異、または別のAPOBECデアミナーゼの1つ以上の対応する変異を含むAPOBECデアミナーゼを含む。いくつかの態様において、本明細書において提供される融合蛋白質のいずれかは、hAPOBEC3G（配列番号60）のW285YおよびR320E変異、または別のAPOBECデアミナーゼの1つ以上の対応する変異を含むAPOBECデアミナーゼを含む。いくつかの態様において、本明細書において提供される融合蛋白質のいずれかは、hAPOBEC3G（配列番号60）のR320EおよびR326E変異、または別のAPOBECデアミナーゼの1つ以上の対応する変異を含むAPOBECデアミナーゼを含む。いくつかの態様において、本明細書において提供される融合蛋白質のいずれかは、hAPOBEC3G（配列番号60）のW285YおよびR326E変異、または別のAPOBECデアミナーゼの1つ以上の対応する変異を含むAPOBECデアミナーゼを含む。いくつかの態様において、本明細書において提供される融合蛋白質のいずれかは、hAPOBEC3G（配列番号60）のW285Y、R320E、およびR326E変異、または別のAPOBECデアミナーゼの1つ以上の対応する変異を含むAPOBECデアミナーゼを含む。

【０２９５】

本開示のいくつかの側面は、（i）ヌクレアーゼ不活性型Cas9ドメイン；および（ii）核酸編集ドメインを含む融合蛋白質を提供する。いくつかの態様において、ヌクレアーゼ不活性型Cas9ドメイン（dCas9）は、いずれかのCas9蛋白質によって提供されるCas9のアミノ酸配列、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つと少なくとも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%同一であるアミノ酸配列を含み、Cas9のヌクレアーゼ活性を不活性化する変異を含む。Cas9のヌクレアーゼドメインを不活性にする変異は当分野において周知である。例えば、Cas9のDNA切断ドメインは、2つのサブドメイン、HNHヌクレアーゼサブドメインおよびRuvC1サブドメインを包含することが公知である。HNHサブドメインはgRNAに対して相補的な鎖を切断し、RuvC1サブドメインは非相補鎖を切断する。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サイレンシングし得る。例えば、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)）。いくつかの態様において、本開示のdCas9は、配列番号6によって提供されるアミノ酸配列のD10A変異、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異を含む。いくつかの態様において、本開示のdCas9は、配列番号6によって提供されるアミノ酸配列のH840A変異、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異を含む。いくつかの態様において、本開示のdCas9は、配列番号6によって提供されるアミノ酸配列のD10AおよびH840A変異両方、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異を含む。いくつかの態様において、Cas9は、配列番号6によって提供されるアミノ酸配列の位置840のヒスチジン残基、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異をさらに含む。触媒残基H840の存在は、標的Cに対向するGを含有する編集されない鎖を切断するCas9の活性を回復させる。H840の回復は、Cを含有する標的鎖の切断をもたらさない。いくつかの態様において、dCas9は配列番号32のアミノ酸配列を含む。Cas9のヌクレアーゼドメインを不活性化する他の変異もまた本開示のdCas9に包含され得ることは理解されるであろう。

【０２９６】

本明細書において開示される変異を含むCas9またはdCas9ドメインは、全長Cas9またはその断片であり得る。いくつかの態様において、Cas9またはその断片を含む蛋白質は「Cas9バリアント」と言われる。Cas9バリアントはCas9またはその断片に対して相同性を共有する。例えば、Cas9バリアントは野生型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%である。いくつかの態様において、Cas9バリアントは、断片が野生型Cas9、例えば配列番号6のアミノ酸配列を含むCas9の対応する断片と少なくとも約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%同一であるようなCas9の断片（例えば、gRNA結合ドメインまたはDNA切断ドメイン）を含む。

【０２９７】

本開示のCas9融合蛋白質のいずれかは核酸編集ドメイン（例えば、核酸を改変することができる酵素、例えばデアミナーゼ）をさらに含み得る。いくつかの態様において、核酸編集ドメインはDNA編集ドメインである。いくつかの態様において、核酸編集ドメインはデアミナーゼ活性を有する。いくつかの態様において、核酸編集ドメインはデアミナーゼまたはデアミナーゼドメインを含むかまたはそれからなる。いくつかの態様において、デアミナーゼはシチジンデアミナーゼである。いくつかの態様において、デアミナーゼはアポリポ蛋白質B mRNA編集複合体（APOBEC）ファミリーデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC1ファミリーデアミナーゼである。いくつかの態様において、デアミナーゼは活性化誘導シチジンデアミナーゼ（AID）である。いくつかの核酸編集ドメイン、およびかかるドメインを包含するCas9融合蛋白質は、本明細書に詳細に記載される。追加の好適な核酸編集ドメインは本開示および当分野の知識に基づいて当業者には明らかであろう。

【０２９８】

本開示のいくつかの側面は、核酸編集ドメインに融合させられたCas9ドメインを含む融合蛋白質を提供し、核酸編集ドメインはCas9ドメインのN末端に融合させられる。いくつかの態様において、Cas9ドメインおよび核酸編集-編集ドメインはリンカーを介して融合させられる。いくつかの態様において、リンカーは$(GGGS)_n$（配列番号613）、$(GGGGS)_n$（配列番号607）、$(G)_n$（配列番号608）、$(EAAAK)_n$（配列番号609）、$(GGS)_n$（配列番号610）、$(SGGS)_n$（配列番号606）、SGSETPGTSESATPES（配列番号604）モチーフ（例えば、Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to Fokl nuclease improves the specificity of genome modification. Nat. Biotechnol. 2014；32(6): 577-82を見よ；内容全体は参照によって本明細書に組み込まれる）、もしくは$(XP)_n$（配列番号611）モチーフ、またはこれらのいずれかの組み合わせを含み、nは独立して1および30の間の整数である。いくつかの態様において、nは、独立して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、もしくは30、または1つよりも多くのリンカーもしくは1つよりも多くのリンカーモチーフが存在する場合にはこれらのいずれかの組み合わせである。いくつかの態様において、リンカーは$(GGS)_n$（配列番号610）モチーフを含み、nは1、2、3、4、5、6、7、8、9、10、11、12、13、14、または15である。いくつかの態様において、リンカーは$(GGS)_n$（配列番号610）モチーフを含み、nは1、3、または7である。いくつかの態様において、リンカーはアミノ酸配列SGSETPGTSESATPES（配列番号604）を含む。追加の好適なリンカーモチーフおよびリンカー構成は当業者には明らかであろう。いくつかの態様において、好適なリンカーモチーフおよび構成はChen et al., Fusion protein linkers: property, design and functionality. Adv Drug Deliv Rev. 2013；65(10): 1357-69に記載されているものを包含し、これの内容全体は参照によって本明細書に組み込まれる。追加の好適なリンカー配列は本開示に基づいて当業者には明らかであろう。いくつかの態様において、本明細書において提供される例示的なCas9融合蛋白質の一般的アーキテクチャは、構造：

$[NH_2]$-[核酸編集ドメイン]-[Cas9]-[COOH]または

$[NH_2]$-[核酸編集ドメイン]-[リンカー]-[Cas9]-[COOH]

を含み、NH_2は融合蛋白質のN末端であり、COOHは融合蛋白質のC末端である。

【０２９９】

本開示の融合蛋白質は1つ以上の追加の特質を含み得る。例えば、いくつかの態様にお

いて、融合蛋白質は核局在配列（NLS）を含む。いくつかの態様において、融合蛋白質のNLSは核酸編集ドメインおよびCas9ドメインの間に局在する。いくつかの態様において、融合蛋白質のNLSはCas9ドメインのC末端に局在する。

【０３００】

存在し得る他の例示的な特質は、局在配列、例えば細胞質局在配列、排出配列、例えば核外排出配列、または他の局在配列、および融合蛋白質の可溶化、精製、または検出にとって有用である配列タグである。本明細書において提供される好適な蛋白質タグは、ビオチンカルボキシラーゼキャリア蛋白質（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タグともまた言われるポリヒスチジンタグ、マルトース結合蛋白質（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、緑色蛍光蛋白質（GFP）タグ、チオレドキシンタグ、Sタグ、Softag（例えばSoftag 1、Softag 3）、strepタグ、ビオチンリガーゼタグ、FlAsHタグ、V5タグ、およびSBPタグを包含するが、これらに限定されない。追加の好適な配列は当業者には明らかであろう。いくつかの態様において、融合蛋白質は1つ以上のHisタグを含む。

【０３０１】

いくつかの態様において、核酸編集ドメインはデアミナーゼである。例えば、いくつかの態様において、デアミナーゼドメインを有する例示的なCas9融合蛋白質の一般的アーキテクチャは、構造：

$[NH_2]$-[NLS]-[デアミナーゼ]-[Cas9]-[COOH]、

$[NH_2]$-[Cas9]-[デアミナーゼ]-[COOH]、

$[NH_2]$-[デアミナーゼ]-[Cas9]-[COOH]、または

$[NH_2]$-[デアミナーゼ]-[Cas9]-[NLS]-[COOH]

を含み；ここで、NLSは核局在配列であり、NH_2は融合蛋白質のN末端であり、COOHは融合蛋白質のC末端である。核局在配列は当分野において公知であり、当業者には明らかであろう。例えば、NLS配列はPlank et al., PCT/EP2000/011690に記載されており、これの内容は例示的な核局在配列のそれらの開示について参照によって本明細書に組み込まれる。いくつかの態様において、NLSはアミノ酸配列PKKKRKV（配列番号614）またはMDSLLMNRRKFLYQFKNVRWAKGRRETYLC（配列番号615）を含む。いくつかの態様においては、リンカーがCas9およびデアミナーゼの間に挿入される。いくつかの態様において、NLSはCas9ドメインのC末端に配置される。いくつかの態様において、NLSはCas9ドメインのN末端に配置される。いくつかの態様において、NLSはデアミナーゼおよびCas9ドメインの間に配置される。いくつかの態様において、NLSはデアミナーゼドメインのN末端に配置される。いくつかの態様において、NLSはデアミナーゼドメインのC末端に配置される。

【０３０２】

1つの例示的な好適な型の核酸編集ドメインは例えばAPOBECファミリーのシチジンデアミナーゼである。シチジンデアミナーゼ酵素のアポリポ蛋白質B mRNA編集複合体（APOBEC）ファミリーは、コントロールされたかつ有益な様式で変異導入を開始するための用をなす11個の蛋白質を包摂する[29]。1つのファミリーメンバー、活性化誘導シチジンデアミナーゼ（AID）は、転写依存的な鎖偏向的な様式でssDNA中のシトシンをウラシルに変換することによって抗体の成熟を担っている[30]。アポリポ蛋白質B編集複合体3（APOBEC3）酵素は、逆転写されたウイルスssDNAのシトシンの脱アミノ化によって、一定のHIV-1株に対する防御をヒト細胞に提供する[31]。これらの蛋白質は全てZn^{2+}配位モチーフ（His-X-Glu-X_{23-26}-Pro-Cys-X_{2-4}-Cys；配列番号616）および結合した水分子を触媒活性のために要求する。Glu残基は、脱アミノ化反応における求核攻撃のために水分子を水酸化亜鉛へと活性化するように作用する。各ファミリーメンバーは、hAIDのWRC（WはAまたはTであり、RはAまたはGである）からhAPOBEC3FのTTCまでの範囲であるその特有の特定の「ホットスポット」において選好的に脱アミノ化する[32]。APOBEC3Gの触媒ドメインの最近の結晶構造は、6つのαヘリックスによってフランキングされる5鎖βシート

コアを含む二次構造を明らかにした。これはファミリー全体で保存されていると信じられている[33]。活性中心ループはssDNA結合と「ホットスポット」のアイデンティティを決定することとを両方担っていることが示されている[34]。これらの酵素の過剰発現はゲノム不安定性および癌にリンクされており、それゆえに配列特異的標的化の重要性を強調している[35]。

【０３０３】

本開示のいくつかの側面は、APOBEC酵素などのシチジンデアミナーゼ酵素の活性がゲノムDNA中の特異的な部位へと導かれ得るという認識に関する。いずれの特定の理論によっても拘束されることを望むものではないが、認識薬剤としてCas9を用いる利点は、（１）Cas9の配列特異性が単純にsgRNA配列を変化させることによって容易に変改され得ることと；（２）Cas9がdsDNAを変性させることによってその標的配列に結合し、一本鎖、よってデアミナーゼの現実的な基質であるDNAのストレッチをもたらすこととを包含する。他の触媒ドメイン、または他のデアミナーゼからの触媒ドメインもまたCas9との融合蛋白質を創出するために用いられ得ることと、本開示はこれについて限定されないこととは理解されるべきである。

【０３０４】

本開示のいくつかの側面は、Cas9:デアミナーゼ融合蛋白質が図3の付番スキームに従う位置3～11のヌクレオチドを効率的に脱アミノ化し得るという認識に基づく。Cas9:デアミナーゼ融合蛋白質によって標的化され得るヌクレオチドについて本明細書において提供される結果から判断して、当業者は、融合蛋白質を脱アミノ化されるべきヌクレオチドを含む標的配列へと標的化するための好適なガイドRNAを設計する能力があるであろう。

【０３０５】

いくつかの態様において、デアミナーゼドメインおよびCas9ドメインはリンカーを介して互いに融合させられる。具体的な用途のためのデアミナーゼ活性にとって最適な長さを達成するために、デアミナーゼドメイン（例えばAID）およびCas9ドメイン間の種々のリンカー長さおよび柔軟性が採用され得る（例えば、形態$(GGGGS)_n$（配列番号607）、$(GGS)_n$（配列番号610）、および$(G)_n$（配列番号608）の非常に柔軟なリンカーから形態$(EAAAK)_n$（配列番号609）、$(SGGS)_n$（配列番号606）、SGSETPGTSESATPES（配列番号604）（例えばGuilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to Fokl nuclease improves the specificity of genome modification. Nat. Biotechnol. 2014; 32(6): 577-82を見よ；内容全体は参照によって本明細書に組み込まれる）、および$(XP)_n$（配列番号611）のより剛直なリンカーまでの範囲である）[36]。いくつかの態様において、リンカーは$(GGS)_n$（配列番号610）モチーフを含み、nは1、3、または7である。いくつかの態様において、リンカーは（SGSETPGTSESATPES（配列番号604）モチーフを含む。

【０３０６】

本開示の側面に従ってCas9ドメインに融合させられ得るいくつかの例示的な好適な核酸編集ドメイン、例えばデアミナーゼおよびデアミナーゼドメインが下で提供されている。いくつかの態様においては、それぞれの配列の活性ドメイン、例えば局在シグナル（核局在配列、核外排出シグナル、細胞質局在シグナル）なしのドメインが用いられ得ることは理解されるべきである。

【０３０７】

ヒトAID：

【化４２】

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGC
HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTAR
LYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENS
VRLSRQLRRILLPLYEVDDLRDAFRTLGL

（配列番号49）
（下線：核局在配列；二重下線：核外排出シグナル）
【０３０８】
マウスAID：
【化４３】

MDSLLMKQKKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSCSLDFGHLRNKSGC
HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPNLSLRIFTAR
LYFCEDRKAEPEGLRRLHRAGVQIGIMTFKDYFYCWNTFVENRERTFKAWEGLHENS
VRLTRQLRRILLPLYEVDDLRDAFRMLGF

（配列番号51）
（下線：核局在配列；二重下線：核外排出シグナル）
【０３０９】
イヌAID：
【化４４】

MDSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGC
HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAAR

【化４５】

LYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENREKTFKAWEGLHENS
VRLSRQLRRILLPLYEVDDLRDAFRTLGL

（配列番号52）
（下線：核局在配列；二重下線：核外排出シグナル）
【０３１０】
ウシ(Bovine)AID：
【化４６】

MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGCH
VELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARL
YFCDKERKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENS
VRLSRQLRRILLPLYEVDDLRDAFRTLGL

（配列番号53）
（下線：核局在配列；二重下線：核外排出シグナル）
【０３１１】
ラット: AID：
【化４７】

MAVGSKPKAALVGPHWERERIWCFLCSTGLGTQQTGQTSRW
LRPAATQDPVSPPRSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLD
FGYLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGN
PNLSLRIFTARLTGWGALPAGLMSPARPSDYFYCWNTFVENHERTFKAWEGLHENSV
RLSRRLRRILLPLYEVDDLRDAFRTLGL

（配列番号５４）
（下線：核局在配列；二重下線：核外排出シグナル）
【０３１２】
マウスAPOBEC-3：
【化４８】

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVS
LHHGVFKNKDNI*HAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFEC*AEQIVRFLAT
HHNLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGR
RFRPWKRLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETRFCVEGRRMD
PLSEEEFYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGKQ*HAEI*
*LFLDKIRSMELSQVTITCYLTWSPCPNC*AWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQ
KGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLRRIKES
WGLQDLVNDFGNLQLGPPMS

（配列番号５５）
（斜体：核酸編集ドメイン）
【０３１３】
ラットAPOBEC-3：
【化４９】

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLRYAIDRKDTFLCYEVTRKDCDSPVSL
HHGVFKNKDNI*HAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFEC*AEQVLRFLAT
HHNLSLDIFSSRLYNIRDPENQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGR
RFRPWKKLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETRFCVERRRVH
LLSEEEFYSQFYNQRVKHLCYYHGVKPYLCYQLEQFNGQAPLKGCLLSEKGKQ*HAEI*
*LFLDKIRSMELSQVIITCYLTWSPCPNC*AWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQ
KGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLHRIKES
WGLQDLVNDFGNLQLGPPMS

（配列番号５６）
（斜体：核酸編集ドメイン）
【０３１４】
アカゲザルAPOBEC-3G：
【化５０】

MVEPMDPRTFVSNFNNRPILSGLNTVWLCCEVKTKDPSGPPLDAKIFQGKVYSKAKY
*HPEM*RFLRWFHKWRQLHHDQEYKVTWYVSWSPCTRC*ANSVATFLAKDPKVTLTIFVAR
LYYFWKPDYQQALRILCQKRGGPHATMKIMNYNEFQDCWNKFVDGRGKPFKPRNN
LPKHYTLLQATLGELLRHLMDPGTFTSNFNNKPWVSGQHETYLCYKVERLHNDTWV
PLNQHRGFLRNQAPNIHGFPKGR*HAELCFLDLIPFWKLDGQQYRVTCFTSWSPCFSC*AQ
EMAKFISNNEHVSLCIFAARIYDDQGRYQEGLRALHRDGAKIAMMNYSEFEYCWDTF
VDRQGRPFQPWDGLDEHSQALSGRLRAI

（配列番号57）
（斜体：核酸編集ドメイン；下線：細胞質局在シグナル）
【０３１５】
チンパンジーAPOBEC-3G：
【化５１】

MKPHFRNPVERMYQDTFSDNFYNRPILSHRNTVWLCYEVKTKGPSRPPLDAKIFRGQ
VYSKLKY*HPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKC*TRDVATFLAEDPKVT
LTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELF
EPWNNLPKYYILLHIMLGEILRHSMDPPTFTSNFNNELWVRGRHETYLCYEVERLHND
TWVLLNQRRGFLCNQAPHKHGFLEGR*HAELCFLDVIPFWKLDLHQDYRVTCFTSWSPC*
*FSC*AQEMAKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLAKAGAKISIMTYSEFKHC
WDTFVDHQGCPFQPWDGLEEHSQALSGRLRAILQNQGN

（配列番号58）
（斜体：核酸編集ドメイン；下線：細胞質局在シグナル）
【０３１６】
ミドリザルAPOBEC-3G：
【化５２】

MNPQIRNMVEQMEPDIFVYYFNNRPILSGRNTVWLCYEVKTKDPSGPPLDANIFQGKL
YPEAKD*HPEMKFLHWFRKWRQLHRDQEYEVTWYVSWSPCTRC*ANSVATFLAEDPKVTL
TIFVARLYYFWKPDYQQALRILCQERGGPHATMKIMNYNEFQHCWNEFVDGQGKPF
KPRKNLPKHYTLLHATLGELLRHVMDPGTFTSNFNNKPWVSGQRETYLCYKVERSH
NDTWVLLNQHRGFLRNQAPDRHGFPKGR*HAELCFLDLIPFWKLDDQQYRVTCFTSWSP*
*CFSC*AQKMAKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLHRDGAKIAVMNYSEFE
YCWDTFVDRQGRPFQPWDGLDEHSQALSGRLRAI

（配列番号59）
（斜体：核酸編集ドメイン；下線：細胞質局在シグナル）
【０３１７】
ヒトAPOBEC-3G：
【化５３】

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQ
VYSELKY*HPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKC*TRDMATFLAEDPKVT
LTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELF
EPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERMHN
DTWVLLNQRRGFLCNQAPHKHGFLEGR*HAELCFLDVIPFWKLDLDQDYRVTCFTSWSP*

【化５４】

*CFSC*AQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHC
WDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN

（配列番号60）
（斜体：核酸編集ドメイン；下線：細胞質局在シグナル）
【０３１８】
ヒトAPOBEC-3F：
【化５５】

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQ
VYSQPEH*HAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDC*VAKLAEFLAEHPNVTL
TISAARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYSEGQPFMPWY
KFDDNYAFLHRTLKEILRNPMEAMYPHIFYFHFKNLRKAYGRNESWLCFTMEVVKH
HSPVSWKRGVFRNQVDPETHC*HAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPEC*AGE
VAEFLARHSNVNLTIFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCWENF
VYNDDEPFKPWKGLKYNFLFLDSKLQEILE

（配列番号61）
（斜体：核酸編集ドメイン）
【０３１９】
ヒトAPOBEC-3B：
【化５６】

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRG
QVYFKPQY*HAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDC*VAKLAEFLSEHPNVT
LTISAARLYYYWERDYRRALCRLSQAGARVTIMDYEEFAYCWENFVYNEGQQFMPW
YKFDENYAFLHRTLKEILRYLMDPDTFTFNFNNDPLVLRRRQTYLCYEVERLDNGTW
VLMDQHMGFLCNEAKNLLCGFY*GRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSW
GC*AGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYC
WDTFVYRQGCPFQPWDGLEEHSQALSGRLRAILQNQGN

（配列番号62）
（斜体：核酸編集ドメイン）
【０３２０】
ラットAPOBEC-3B：
MQPQGLGPNAGMGPVCLGCSHRRPYSPIRNPLKKLYQQTFYFHFKNVRYAWGRKNNFL
CYEVNGMDCALPVPLRQGVFRKQGHIHAELCFIYWFHDKVLRVLSPMEEFKVTWYMSW
SPCSKCAEQVARFLAAHRNLSLAIFSSRLYYYLRNPNYQQKLCRLIQEGVHVAAMDLPEF
KKCWNKFVDNDGQPFRPWMRLRINFSFYDCKLQEIFSRMNLLREDVFYLQFNNSHRVKP
VQNRYYRRKSYLCYQLERANGQEPLKGYLLYKKGEQHVEILFLEKMRSMELSQVRITCYL
TWSPCPNCARQLAAFKKDHPDLILRIYTSRLYFYWRKKFQKGLCTLWRSGIHVDVMDLP
QFADCWTNFVNPQRPFRPWNELEKNSWRIQRRLRRIKESWGL（配列番号63）
【０３２１】
ウシ(Bovine)APOBEC-3B：
DGWEVAFRSGTVLKAGVLGVSMTEGWAGSGHPGQGACVWTPGTRNTMNLLREVLFKQ
QFGNQPRVPAPYYRRKTYLCYQLKQRNDLTLDRGCFRNKKQRHAEIRFIDKINSLDLNPS
QSYKIICYITWSPCPNCANELVNFITRNNHLKLEIFASRLYFHWIKSFKMGLQDLQNAGIS
VAVMTHTEFEDCWEQFVDNQSRPFQPWDKLEQYSASIRRRLQRILTAPI（配列番号64）
【０３２２】
チンパンジーAPOBEC-3B：

MNPQIRNPMEWMYQRTFYYNFENEPILYGRSYTWLCYEVKIRRGHSNLLWDTGVFRGQ
MYSQPEHHAEMCFLSWFCGNQLSAYKCFQITWFVSWTPCPDCVAKLAKFLAEHPNVTL
TISAARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYNEGQPFMPWYKF
DDNYAFLHRTLKEIIRHLMDPDTFTFNFNNDPLVLRRHQTYLCYEVERLDNGTWVLMD
QHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAG
QVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVY
RQGCPFQPWDGLEEHSQALSGRLRAILQVRASSLCMVPHRPPPPPQSPGPCLPLCSEPPL
GSLLPTGRPAPSLPFLLTASFSFPPPASLPPLPSLSLSPGHLPVPSFHSLTSCSIQPPCSSRI
RETEGWASVSKEGRDLG（配列番号65）

【０３２３】

ヒトAPOBEC-3C：

【化５７】

MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFR
NQVDSETH*CHAERCFLSWFCDDILSPNTKYQVTWYTSWSPCPDC*AGEVAEFLARHSNVN
LTIFTARLYYFQYPCYQEGLRSLSQEGVAVEIMDYEDFKYCWENFVYNDNEPFKPWK
GLKTNFRLLKRRLRESLQ

（配列番号66）

（斜体：核酸編集ドメイン）

【０３２４】

ゴリラAPOBEC3C：

【化５８】

MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFR
NQVDSETH*CHAERCFLSWFCDDILSPNTNYQVTWYTSWSPCPEC*AGEVAEFLARHSNVN
LTIFTARLYYFQDTDYQEGLRSLSQEGVAVKIMDYKDFKYCWENFVYNDDEPFKPW
KGLKYNFRFLKRRLQEILE

（配列番号67）

（斜体：核酸編集ドメイン）

【０３２５】

ヒトAPOBEC-3A:

【化５９】

MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHN
QAKNLLCGFYGR*HAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGC*AGEVRAFLQE
NTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCP
FQPWDGLDEHSQALSGRLRAILQNQGN

（配列番号68）

（斜体：核酸編集ドメイン）

【０３２６】

アカゲザルAPOBEC-3A：

【化６０】

MDGSPASRPRHLMDPNTFTFNFNNDLSVRGRHQTYLCYEVERLDNGTWVPMDERRG
FLCNKAKNVPCGDYGC*HVELRFLCEVPSWQLDPAQTYRVTWFISWSPC*FRRGCAGQVR
VFLQENKHVRLRIFAARIYDYDPLYQEALRTLRDAGAQVSIMTYEEFKHCWDTFVDR
QGRPFQPWDGLDEHSQALSGRLRAILQNQGN

（配列番号69）
（斜体：核酸編集ドメイン）

【０３２７】

ウシ(Bovien)APOBEC-3A：

【化６１】

MDEYTFTENFNNQGWPSKTYLCYEMERLDGDATIPLDEYKGFVRNKGLDQPEKPC*H*
*AELYFLGKIHSWNLDRNQHYRLTCFISWSPC*YDCAQKLTTFLKENHHISLHILASRIYTHN
RFGCHQSGLCELQAAGARITIMTFEDFKHCWETFVDHKGKPFQPWEGLNVKSQALCT
ELQAILKTQQN

（配列番号70）
（斜体：核酸編集ドメイン）

【０３２８】

ヒトAPOBEC-3H：

【化６２】

MALLTAETFRLQFNNKRRLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKC*HAEIC*
*FINEIKSMGLDETQCYQVTCYLTWSPCSSC*AWELVDFIKAHDHLNLGIFASRLYYHWCK
PQQKGLRLLCGSQVPVEVMGFPKFADCWENFVDHEKPLSFNPYKMLEELDKNSRAIK
RRLERIKIPGVRAQGRYMDILCDAEV

（配列番号71）
（斜体：核酸編集ドメイン）

【０３２９】

アカゲザルAPOBEC-3H：

MALLTAKTFSLQFNNKRRVNKPYYPRKALLCYQLTPQNGSTPTRGHLKNKKKDHAEIRFI
NKIKSMGLDETQCYQVTCYLTWSPCPSCAGELVDFIKAHRHLNLRIFASRLYYHWRPNY
QEGLLLLCGSQVPVEVMGLPEFTDCWENFVDHKEPPSFNPSEKLEELDKNSQAIKRRLER
IKSRSVDVLENGLRSLQLGPVTPSSSIRNSR（配列番号72）

【０３３０】

ヒトAPOBEC-3D：

【化６３】

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRG
PVLPKRQSNHRQEVYFRFEN*HAEMCFLSWFCGNRLPANRRFQITWFVSWNPCLPC*VVK
VTKFLAEHPNVTLTISAARLYYYRDRDWRWVLLRLHKAGARVKIMDYEDFAYCWE
NFVCNEGQPFMPWYKFDDNYASLHRTLKEILRNPMEAMYPHIFYFHFKNLLKACGRN
ESWLCFTMEVTKHHSAVFRKRGVFRNQVDPETHC*HAERCFLSWFCDDILSPNTNYEVT*
*WYTSWSPCPEC*AGEVAEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGASVKI
MGYKDFVSCWKNFVYSDDEPFKPWKGLQTNFRLLKRRLREILQ

（配列番号73）
（斜体：核酸編集ドメイン）
【０３３１】
ヒトAPOBEC-1：
MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNTTN
HVEVNFIKKFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVARLF
WHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMM
LYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPPHILLATGLIHPSVAWR
（配列番号74）
【０３３２】
マウスAPOBEC-1：
MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSVWRHTSQNTSN
HVEVNFLEKFTTERYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPYVTLFIYIARLYH
HTDQRNRQGLRDLISSGVTIQIMTEQEYCYCWRNFVNYPPSNEAYWPRYPHLWVKLYV
LELYCIILGLPPCLKILRRKQPQLTFFTITLQTCHYQRIPPHLLWATGLK（配列番号75）
【０３３３】
ラットAPOBEC-1：
MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLE
LYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK（配列番号76）
【０３３４】
ヒトAPOBEC-2：
MAQKEEAAVATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRNVE
YSSGRNKTFLCYVVEAQGKGGQVQASRGYLEDEHAAAHAEEAFFNTILPAFDPALRYNVT
WYVSSSPCAACADRIIKTLSKTKNLRLLILVGRLFMWEEPEIQAALKKLKEAGCKLRIMKP
QDFEYVWQNFVEQEEGESKAFQPWEDIQENFLYYEEKLADILK（配列番号77）
【０３３５】
マウスAPOBEC-2：
MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVE
YSSGRNKTFLCYVVEVQSKGGQAQATQGYLEDEHAGAHAEEAFFNTILPAFDPALKYNVT
WYVSSSPCAACADRILKTLSKTKNLRLLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMK
PQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADILK（配列番号78）
【０３３６】
ラットAPOBEC-2：
MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVE
YSSGRNKTFLCYVVEAQSKGGQVQATQGYLEDEHAGAHAEEAFFNTILPAFDPALKYNVT
WYVSSSPCAACADRILKTLSKTKNLRLLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMK

PQDFEYLWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADILK（配列番号79）
【０３３７】
ウシ(Bovine)APOBEC-2：
MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRNVE
YSSGRNKTFLCYVVEAQSKGGQVQASRGYLEDEHATNHAEEAFFNSIMPTFDPALRYMV
TWYVSSSPCAACADRIVKTLNKTKNLRLLILVGRLFMWEEPEIQAALRKLKEAGCRLRIM
KPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADILK（配列番号80）
【０３３８】
Petromyzon marinus CDA1（pmCDA1）
MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQS
GTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTL
KIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWL
EKTLKRAEKRRSELSIMIQVKILHTTKSPAV（配列番号81）
【０３３９】
ヒトAPOBEC3G D316R_D317R
MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVY
SELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLT
IFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPW
NNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWV
LLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCA
QEMAKFISKNKHVSLCIFTARIYRRQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVD
HQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN（配列番号82）
【０３４０】
ヒトAPOBEC3G A鎖
MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFL
EGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTAR
IYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLS
GRLRAILQ（配列番号83）
【０３４１】
ヒトAPOBEC3G A鎖 D120R_D121R
MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFL
EGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTAR
IYRRQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLS
GRLRAILQ（配列番号84）
【０３４２】
いくつかの態様において、本明細書において提供される融合蛋白質は、核酸編集酵素の全長アミノ酸、例えば上で提供されている配列の1つを含む。しかしながら、他の態様においては、本明細書において提供される融合蛋白質は核酸編集酵素の全長配列ではなくその断片のみを含む。例えば、いくつかの態様において、本明細書において提供される融合蛋白質はCas9ドメインと核酸編集酵素の断片とを含み、例えば断片は核酸編集ドメインを含む。核酸編集ドメインの例示的なアミノ酸配列は上の配列中において斜体化された文字として示されており、かかるドメインの追加の好適な配列は当業者には明らかであろう。
【０３４３】
本発明の側面に従って用いられ得る、例えばヌクレアーゼ不活性型Cas9ドメインに融合させられ得る追加の好適な核酸編集酵素配列、例えばデアミナーゼ酵素およびドメイン配列は、本開示に基づいて当業者には明らかであろう。いくつかの態様において、かかる追加の酵素配列は、本明細書において提供される配列に少なくとも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%、または少なくとも99%類似であるデアミナーゼ酵素またはデアミナーゼドメイン配列を包含する。追加の好適なCas9ドメイン、バリアント、お

よび配列もまた当業者には明らかであろう。かかる追加の好適なCas9ドメインの例は、D10A、D10A/D839A/H840A、およびD10A/D839A/H840A/N863A変異体ドメインを包含するが、これらに限定されない（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を見よ。これの内容全体は参照によって本明細書に組み込まれる）。いくつかの態様において、Cas9は、配列番号6によって提供されるアミノ酸配列の位置840のヒスチジン残基、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異を含む。触媒残基H840の存在は、標的Cに対向するGを含有する編集されない鎖を切断するCas9の活性を回復させる。H840の回復は、Cを含有する標的鎖の切断をもたらさない。

【０３４４】

Cas9ドメインとデアミナーゼドメインとを含む融合蛋白質を創出するための追加の好適な戦略は、当分野の一般的知識との組み合わせによって、本開示に基づいて当業者には明らかであろう。本開示および当分野の知識から判断して、リンカーを用いてまたはリンカーの使用なしで本開示の側面に従う融合蛋白質を創出するための好適な戦略もまた当業者には明らかであろう。例えば、Gilbert et al., CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell. 2013; 154(2): 442-51は、2つのNLSをリンカー（SPKKKRKVEAS、配列番号617）として用いるVP64とのCas9のC末端融合体が転写活性化のために採用され得ることを示した。Mali et al., CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol. 2013; 31(9): 833-8は、リンカーなしのVP64とのC末端融合体が転写活性化のために採用され得ることを報告している。そして、Maeder et al., CRISPR RNA-guided activation of endogenous human genes. Nat Methods. 2013; 10: 977-979は、Gly_4Ser（配列番号613）リンカーを用いるVP64とのC末端融合体が転写活性化因子として用いられ得ることを報告している。最近、dCas9-Fok1ヌクレアーゼ融合体が首尾良く創出されており、親のCas9酵素と比較して改善された酵素特異性を見せる（Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to Fokl nuclease improves the specificity of genome modification. Nat. Biotechnol. 2014; 32(6): 577-82およびTsai SQ, Wyvekens N, Khayter C, Foden JA, Thapar V, Reyon D, Goodwin MJ, Aryee MJ, Joung JK. Dimeric CRISPR RNA-guided Fokl nucleases for highly specific genome editing. Nat Biotechnol. 2014; 32(6):569-76. PMID: 24770325では、それぞれSGSETPGTSESATPES（配列番号604）またはGGGGS（配列番号607）リンカーがFokI-dCas9融合蛋白質に用いられた）。

【０３４５】

本開示のいくつかの側面は、（i）Cas9酵素またはドメイン（例えば第1の蛋白質）；および（ii）核酸編集酵素またはドメイン（例えば第2の蛋白質）を含む融合蛋白質を提供する。いくつかの側面において、本明細書において提供される融合蛋白質は、さらに、（iii）プログラム可能なDNA結合蛋白質、例えばジンクフィンガードメイン、TALE、または第2のCas9蛋白質（例えば第3の蛋白質）を包含する。いずれの特定の理論によっても拘束されることを望むものではないが、プログラム可能なDNA結合蛋白質（例えば第2のCas9蛋白質）を、（i）Cas9酵素またはドメイン（例えば第1の蛋白質）；および（ii）核酸編集酵素またはドメイン（例えば第2の蛋白質）を含む融合蛋白質に融合させることは、標的核酸配列に対する融合蛋白質の特異性を改善するために、または古典的PAM（NGG）配列を含有しない標的核酸配列に結合するように融合蛋白質の特異性もしくは結合親和性を改善するために有用であり得る。いくつかの態様において、第3の蛋白質はCas9蛋白質（例えば第2のCas9蛋白質）である。いくつかの態様において、第3の蛋白質は本明細書において提供されるCas9蛋白質のいずれかである。いくつかの態様において、第3の蛋白質はCas9蛋白質（例えば第1の蛋白質）のN末端において融合蛋白質に融合させられる。

いくつかの態様において、第3の蛋白質はCas9蛋白質（例えば第1の蛋白質）のC末端において融合蛋白質に融合させられる。いくつかの態様において、Cas9ドメイン（例えば第1の蛋白質）および第3の蛋白質（例えば第2のCas9蛋白質）はリンカー（例えば第2のリンカー）を介して融合させられる。いくつかの態様において、リンカーは$(GGGGS)_n$（配列番号607）、$(G)_n$（配列番号608）、$(EAAAK)_n$（配列番号609）、$(GGS)_n$（配列番号610）、$(SGGS)_n$（配列番号606）、SGSETPGTSESATPES（配列番号604）、$SGGS(GGS)_n$（配列番号612）、SGGSSGGSSGSETPGTSESATPESSGGSSGGS（配列番号605）、もしくは$(XP)_n$（配列番号611）モチーフ、またはこれらのいずれかの組み合わせを含み、nは独立して1および30の間の整数である。いくつかの態様において、本明細書において提供される例示的なnapDNAbp融合蛋白質の一般的アーキテクチャは、構造：

[NH2]-[核酸編集酵素もしくはドメイン]-[napDNAbp]-[第3の蛋白質]-[COOH]；

[NH2]-[第3の蛋白質]-[napDNAbp]-[核酸編集酵素もしくはドメイン]-[COOH]；

[NH2]-[napDNAbp]-[核酸編集酵素もしくはドメイン]-[第3の蛋白質]-[COOH]；

[NH2]-[第3の蛋白質]-[核酸編集酵素もしくはドメイン]-[napDNAbp]-[COOH]；

[NH2]-[UGI]-[核酸編集酵素もしくはドメイン]-napDNAbp]-[第3の蛋白質]-[COOH]；

[NH2]-[UGI]-[第3の蛋白質]-[napDNAbp]-[核酸編集酵素もしくはドメイン]-[COOH]；

[NH2]-[UGI]-[napDNAbp]-[核酸編集酵素もしくはドメイン]-[第3の蛋白質]-[COOH]；

[NH2]-[UGI]-[第3の蛋白質]-[核酸編集酵素もしくはドメイン]-[napDNAbp]-[COOH]；

[NH2]-[核酸編集酵素もしくはドメイン]-[napDNAbp]-[第3の蛋白質]-[UGI]-[COOH]；

[NH2]-[第3の蛋白質]-[napDNAbp]-[核酸編集酵素もしくはドメイン]-[UGI]-[COOH]；

[NH2]-[NapDNAbp]-[核酸編集酵素もしくはドメイン]-[第3の蛋白質]-[UGI]-[COOH]；

[NH2]-[第3の蛋白質]-[核酸編集酵素もしくはドメイン]-[NapDNAbp]-[UGI]-[COOH]；または

[NH2]-[核酸編集酵素もしくはドメイン]-[NapDNAbp]-[第1のUGIドメイン]-[第2のUGIドメイン]-[COOH]

を含み；ここで、NH2は融合蛋白質のN末端であり、COOHは融合蛋白質のC末端である。いくつかの態様において、上の一般的アーキテクチャに用いられている「]-[」は任意のリンカー配列の存在を指示する。他の例では、本明細書において提供される例示的なNapDNAbp融合蛋白質の一般的アーキテクチャは、構造：

[NH2]-[核酸編集酵素もしくはドメイン]-[NapDNAbp]-[第2のNapDNAbp蛋白質]-[COOH]；

[NH2]-[第2のNapDNAbp蛋白質]-[NapDNAbp]-[核酸編集酵素もしくはドメイン]-[COOH]；

[NH2]-[NapDNAbp]-[核酸編集酵素もしくはドメイン]-[第2のNapDNAbp蛋白質]-[COOH]；

[NH2]-[第2のNapDNAbp蛋白質]-[核酸編集酵素もしくはドメイン]-[NapDNAbp]-[COOH]；

[NH2]-[UGI]-[核酸編集酵素もしくはドメイン]-[NapDNAbp]-[第2のNapDNAbp蛋白質]-[COOH]、

[NH2]-[UGI]-[第2のNapDNAbp蛋白質]-[NapDNAbp]-[核酸編集酵素もしくはドメイン]-[COOH]；

[NH2]-[UGI]-[NapDNAbp]-[核酸編集酵素もしくはドメイン]-[第2のNapDNAbp蛋白質]-[COOH]；

[NH2]-[UGI]-[第2のNapDNAbp蛋白質]-[核酸編集酵素もしくはドメイン]-[NapDNAbpHCOOH]；
[NH2]-[核酸編集酵素もしくはドメイン]-[NapDNAbp]-[第2のNapDNAbp蛋白質]-[UGI]-[COOH]；
[NH2]-[第2のNapDNAbp蛋白質]-[NapDNAbp]-[核酸編集酵素もしくはドメイン]-[UGI]-[COOH]；
[NH2]-[NapDNAbp]-[核酸編集酵素もしくはドメイン]-[第2のNapDNAbp蛋白質]-[UGI]-[COOH]；または
[NH2]-[第2のNapDNAbp蛋白質]-[核酸編集酵素もしくはドメイン]-[NapDNAbp]-[UGI]-[COOH]
を含み；ここで、NH_2は融合蛋白質のN末端であり、COOHは融合蛋白質のC末端である。いくつかの態様において、上の一般的アーキテクチャに用いられている「]-[」は任意のリンカー配列の存在を指示する。いくつかの態様において、第2のNapDNAbpはdCas9蛋白質である。いくつかの例では、本明細書において提供される例示的なCas9融合蛋白質の一般的アーキテクチャは図3に示される構造を含む。例示的なCas9融合蛋白質の一般的アーキテクチャのいずれかによって提供される蛋白質のいずれかは、本明細書において提供されるリンカーの1つ以上によって接続され得ることは了解されるべきである。いくつかの態様において、リンカー同士は同じである。いくつかの態様において、リンカー同士は異なる。いくつかの態様において、例示的なCas9融合蛋白質の一般的アーキテクチャのいずれかによって提供される蛋白質の1つ以上は、リンカーを介して融合させられない。いくつかの態様において、融合蛋白質は核標的化配列、例えば核局在配列をさらに含む。いくつかの態様において、本明細書において提供される融合蛋白質は核局在配列（NLS）をさらに含む。いくつかの態様において、NLSは融合蛋白質のN末端に融合させられる。いくつかの態様において、NLSは融合蛋白質のC末端に融合させられる。いくつかの態様において、NLSは第3の蛋白質のN末端に融合させられる。いくつかの態様において、NLSは第3の蛋白質のC末端に融合させられる。いくつかの態様において、NLSはCas9蛋白質のN末端に融合させられる。いくつかの態様において、NLSはCas9蛋白質のC末端に融合させられる。いくつかの態様において、NLSは核酸編集酵素またはドメインのN末端に融合させられる。いくつかの態様において、NLSは核酸編集酵素またはドメインのC末端に融合させられる。いくつかの態様において、NLSはUGI蛋白質のN末端に融合させられる。いくつかの態様において、NLSはUGI蛋白質のC末端に融合させられる。いくつかの態様において、NLSは1つ以上のリンカーを介して融合蛋白質に融合させられる。いくつかの態様において、NLSはリンカーなしで融合蛋白質に融合させられる。

【０３４６】

ウラシルグリコシラーゼ阻害因子融合蛋白質

本開示のいくつかの側面はウラシルグリコシラーゼ阻害因子（UGI）ドメインを含む融合蛋白質に関する。いくつかの態様において、Cas9ドメイン（例えば、ヌクレアーゼ活性型Cas9ドメイン、ヌクレアーゼ不活性型dCas9ドメイン、またはCas9ニッカーゼ）を含む本明細書において提供される融合蛋白質のいずれかは、さらに、直接的にまたはリンカーを介してどちらかでUGIドメインに融合させられ得る。本開示のいくつかの側面は、増大した核酸塩基編集効率を有する、デアミナーゼ-dCas9融合蛋白質、デアミナーゼ-ヌクレアーゼ活性型Cas9融合蛋白質、およびデアミナーゼ-Cas9ニッカーゼ融合蛋白質を提供する。いずれの特定の理論によっても拘束されることを望むものではないが、U:Gヘテロ二重鎖DNAの存在に対する細胞性のDNA修復応答は、細胞における核酸塩基編集効率の減少を担ってい得る。例えば、ウラシルDNAグリコシラーゼ（UDG）は細胞のDNAからのUの取り除きを触媒し、これは、U:G対からC:G対への復帰を最も一般的な成果として有する塩基除去修復を開始し得る。下の例で実証されている通り、ウラシルDNAグリコシラーゼ阻害因子（UGI）はヒトUDG活性を阻害し得る。それゆえに、本開示は、さらにUGIドメインに融合させられたdCas9-核酸編集ドメインを含む融合蛋白質を企図する。本開示は、さらにUGIドメインに融合させられたCas9ニッカーゼ-核酸編集ドメインを含む融合蛋白

質をもまた企図する。UGIドメインの使用は、C→U変化を触媒することができる核酸編集ドメインの編集効率を増大させ得ることは理解されるべきである。例えば、UGIドメインを含む融合蛋白質はC残基を脱アミノ化することがより効率的であり得る。いくつかの態様において、融合蛋白質は、構造：

[デアミナーゼ]-[任意のリンカー配列]-[dCas9]-[任意のリンカー配列]-[UGI]；

[デアミナーゼ]-[任意のリンカー配列]-[UGI]-[任意のリンカー配列]-[dCas9]；

[UGI]-[任意のリンカー配列]-[デアミナーゼ]-[任意のリンカー配列]-[dCas9]；

[UGI]-[任意のリンカー配列]-[dCas9]-[任意のリンカー配列]-[デアミナーゼ]；

[dCas9]-[任意のリンカー配列]-[デアミナーゼ]-[任意のリンカー配列]-[UGI]；

[dCas9]-[任意のリンカー配列]-[UGI]-[任意のリンカー配列]-[デアミナーゼ]；

[デアミナーゼ]-[任意のリンカー配列]-[dCas9]-[任意のリンカー配列]-[第1のUGI]-[任意のリンカー配列]-[第2のUGI]；

[デアミナーゼ]-[任意のリンカー配列]-[第1のUGI]-[任意のリンカー配列]-[第2のUGI]-[任意のリンカー配列]-[dCas9]；

[第1のUGI]-[任意のリンカー配列]-[第2のUGI]-[任意のリンカー配列]-[デアミナーゼ]-[任意のリンカー配列]-[dCas9]；

[第1のUGI]-[任意のリンカー配列]-[第2のUGI]-[任意のリンカー配列]-[dCas9]-[任意のリンカー配列]-[デアミナーゼ]；

[dCas9]-[任意のリンカー配列]-[デアミナーゼ]-[任意のリンカー配列]-[第1のUGI]-[任意のリンカー配列]-[第2のUGI]；または

[dCas9]-[任意のリンカー配列]-[第1のUGI]-[任意のリンカー配列]-[第2のUGI]-[任意のリンカー配列]-[デアミナーゼ]

を含む。

【０３４７】

他の態様において、融合蛋白質は、構造：

[デアミナーゼ]-[任意のリンカー配列]-[Cas9ニッカーゼ]-[任意のリンカー配列]-[UGI]；

[デアミナーゼ]-[任意のリンカー配列]-[UGI]-[任意のリンカー配列]-[Cas9ニッカーゼ]；

[UGI]-[任意のリンカー配列]-[デアミナーゼ]-[任意のリンカー配列]-[Cas9ニッカーゼ]；

[UGI]-[任意のリンカー配列]-[Cas9ニッカーゼ]-[任意のリンカー配列]-[デアミナーゼ]；

[Cas9ニッカーゼ]-[任意のリンカー配列]-[デアミナーゼ]-[任意のリンカー配列]-[UGI]；

[Cas9ニッカーゼ]-[任意のリンカー配列]-[UGI]-[任意のリンカー配列]-[デアミナーゼ]

[デアミナーゼ]-[任意のリンカー配列]-[Cas9ニッカーゼ]-[任意のリンカー配列]-[第1のUGI]-[任意のリンカー配列]-[第2のUGI]；

[デアミナーゼ]-[任意のリンカー配列]-[第1のUGI]-[任意のリンカー配列]-[第2のUGI]-[任意のリンカー配列]-[Cas9ニッカーゼ]；

[第1のUGI]-[任意のリンカー配列]-[第2のUGI]-[任意のリンカー配列]-[デアミナーゼ]-[任意のリンカー配列]-[Cas9ニッカーゼ]；

[第1のUGI]-[任意のリンカー配列]-[第2のUGI]-[任意のリンカー配列]-[Cas9ニッカーゼ]-[任意のリンカー配列]-[デアミナーゼ]；

[Cas9ニッカーゼ]-[任意のリンカー配列]-[デアミナーゼ]-[任意のリンカー配列]-[第1のUGI]-[任意のリンカー配列]-[第2のUGI]；または

[Cas9ニッカーゼ]-[任意のリンカー配列]-[第1のUGI]-[任意のリンカー配列]-[第2のUGI]-[任意のリンカー配列]-[デアミナーゼ]

を含む。

【０３４８】
上に記載される融合蛋白質のいずれかは、（ｉ）核酸によってプログラム可能なDNA結合蛋白質（napDNAbp）；（ii）シチジンデアミナーゼドメイン；および（iii）2つ以上のUGIドメインを含み得、2つ以上のUGIドメインは構築物中において互いに隣接し得（例えば、[第1のUGI]-[第2のUGI]。「-」は任意のリンカーである）、または2つ以上のUGIドメインは（ｉ）のnapDNAbpおよび／または（ii）のシチジンデアミナーゼドメインによって分離され得る（例えば、[第1のUGI]-[デアミナーゼ]-[第2のUGI]、[第1のUGI]-[napDNAbp]-[第2のUGI]、[第1のUGI]-[デアミナーゼ]-[napDNAbp]-[第2のUGI]など。「-」は任意のリンカーである）ことは了解されるべきである。
【０３４９】
別の側面において、融合蛋白質は：（ｉ）Cas9酵素またはドメイン；（ii）核酸編集酵素またはドメイン（例えば第2の蛋白質）（例えばシチジンデアミナーゼドメイン）；（iii）第1のウラシルグリコシラーゼ阻害因子ドメイン（UGI）（例えば第3の蛋白質）；および（iv）第2のウラシルグリコシラーゼ阻害因子ドメイン（UGI）（例えば第4の蛋白質）を含む。第1および第2のウラシルグリコシラーゼ阻害因子ドメイン（UGI）は同じかまたは異なり得る。いくつかの態様において、Cas9ドメイン（例えば第1の蛋白質）およびデアミナーゼ（例えば第2の蛋白質）はリンカーを介して融合させられる。いくつかの態様において、Cas9ドメインはデアミナーゼのc末端に融合させられる。いくつかの態様において、Cas9蛋白質（例えば第1の蛋白質）および第1のUGIドメイン（例えば第3の蛋白質）はリンカー（例えば第2のリンカー）を介して融合させられる。いくつかの態様において、第1のUGIドメインはCas9蛋白質のc末端に融合させられる。いくつかの態様において、第1のUGIドメイン（例えば第3の蛋白質）および第2のUGIドメイン（例えば第4の蛋白質）はリンカー（例えば第3のリンカー）を介して融合させられる。いくつかの態様において、第2のUGIドメインは第1のUGIドメインのc末端に融合させられる。いくつかの態様において、リンカーは$(GGGGS)_n$（配列番号607）、$(G)_n$（配列番号608）、$(EAAAK)_n$（配列番号609）、$(GGS)_n$（配列番号610）、$(SGGS)_n$（配列番号606）、SGSETPGTSESATPES（配列番号604）、SGGS$(GGS)_n$（配列番号612）、SGGSSGGSSGSETPGTSESATPESSGGSSGGS（配列番号605）、もしくは$(XP)_n$（配列番号611）モチーフ、またはこれらのいずれかの組み合わせを含み、nは独立して1および30の間の整数である。いくつかの態様において、第1のリンカーは1～50アミノ酸のアミノ酸配列を含む。いくつかの態様において、第1のリンカーは1～40アミノ酸のアミノ酸配列を含む。いくつかの態様において、第1のリンカーは1～35アミノ酸のアミノ酸配列を含む。いくつかの態様において、第1のリンカーは1～30アミノ酸のアミノ酸配列を含む。いくつかの態様において、第1のリンカーは1～20アミノ酸のアミノ酸配列を含む。いくつかの態様において、第1のリンカーは10～20アミノ酸のアミノ酸配列を含む。いくつかの態様において、第1のリンカーは30～40アミノ酸のアミノ酸配列を含む。いくつかの態様において、第1のリンカーは14、16、または18アミノ酸のアミノ酸配列を含む。いくつかの態様において、第1のリンカーは16アミノ酸のアミノ酸配列を含む。いくつかの態様において、第1のリンカーは30、32、または34アミノ酸のアミノ酸配列を含む。いくつかの態様において、第1のリンカーは32アミノ酸のアミノ酸配列を含む。いくつかの態様において、第1のリンカーはSGSETPGTSESATPES（配列番号604）モチーフを含む。いくつかの態様において、第1のリンカーはSGGSSGGSSGSETPGTSESATPESSGGSSGGS（配列番号605）モチーフを含む。いくつかの態様において、第2のリンカーは1～50アミノ酸のアミノ酸配列を含む。いくつかの態様において、第2のリンカーは1～40アミノ酸のアミノ酸配列を含む。いくつかの態様において、第2のリンカーは1～35アミノ酸のアミノ酸配列を含む。いくつかの態様において、第2のリンカーは1～30アミノ酸のアミノ酸配列を含む。いくつかの態様において、第2のリンカーは1～20アミノ酸のアミノ酸配列を含む。いくつかの態様において、第2のリンカーは2～20アミノ酸のアミノ酸配列を含む。いくつかの態様において、第2のリンカーは2～10アミノ酸のアミノ酸配列を含む。いくつかの態様において、第2のリンカーは10～20アミノ酸のアミノ酸配列を含む。いくつかの態様にお

いて、第2のリンカーは2、4、または6アミノ酸のアミノ酸配列を含む。いくつかの態様において、第2のリンカーは7、9、または11アミノ酸のアミノ酸配列を含む。いくつかの態様において、第2のリンカーは14、16、または18アミノ酸のアミノ酸配列を含む。いくつかの態様において、第2のリンカーは4アミノ酸のアミノ酸配列を含む。いくつかの態様において、第2のリンカーは9アミノ酸のアミノ酸配列を含む。いくつかの態様において、第2のリンカーは16アミノ酸のアミノ酸配列を含む。いくつかの態様において、第2のリンカーは$(SGGS)_n$（配列番号606）モチーフを含み、nは1から30までの間の整数である。いくつかの態様において、第2のリンカーは$(SGGS)_n$（配列番号606）モチーフを含み、nは1である。いくつかの態様において、第2のリンカーは$SGGS(GGS)_n$（配列番号612）モチーフを含み、nは1から30までの間の整数である。いくつかの態様において、第2のリンカーは$SGGS(GGS)_n$（配列番号612）モチーフを含み、nは2である。いくつかの態様において、第3のリンカーは1～50アミノ酸のアミノ酸配列を含む。いくつかの態様において、第3のリンカーは1～40アミノ酸のアミノ酸配列を含む。いくつかの態様において、第3のリンカーは1～35アミノ酸のアミノ酸配列を含む。いくつかの態様において、第3のリンカーは1～30アミノ酸のアミノ酸配列を含む。いくつかの態様において、第3のリンカーは1～20アミノ酸のアミノ酸配列を含む。いくつかの態様において、第3のリンカーは2～20アミノ酸のアミノ酸配列を含む。いくつかの態様において、第3のリンカーは2～10アミノ酸のアミノ酸配列を含む。いくつかの態様において、第3のリンカーは10～20アミノ酸のアミノ酸配列を含む。いくつかの態様において、第3のリンカーは2、4、または6アミノ酸のアミノ酸配列を含む。いくつかの態様において、第3のリンカーは7、9、または11アミノ酸のアミノ酸配列を含む。いくつかの態様において、第3のリンカーは14、16、または18アミノ酸のアミノ酸配列を含む。いくつかの態様において、第3のリンカーは4アミノ酸のアミノ酸配列を含む。いくつかの態様において、第3のリンカーは9アミノ酸のアミノ酸配列を含む。いくつかの態様において、第3のリンカーは16アミノ酸のアミノ酸配列を含む。いくつかの態様において、第3のリンカーは$(SGGS)_n$（配列番号606）モチーフを含み、nは1から30までの間の整数である。いくつかの態様において、第3のリンカーは$(SGGS)_n$（配列番号606）モチーフを含み、nは1である。いくつかの態様において、第3のリンカーは$SGGS(GGS)_n$（配列番号612）モチーフを含み、nは1から30までの間の整数である。いくつかの態様において、第3のリンカーは$SGGS(GGS)_n$（配列番号612）モチーフを含み、nは2である。

【０３５０】

いくつかの態様において、融合蛋白質は、構造：

[デアミナーゼ]-[任意のリンカー配列]-[dCas9]-[任意のリンカー配列]-[第1のUGI]-[任意のリンカー配列]-[第2のUGI]；

[デアミナーゼ]-[任意のリンカー配列]-[Cas9ニッカーゼ]-[任意のリンカー配列]-[第1のUGI]-[任意のリンカー配列]-[第2のUGI]；または

[デアミナーゼ]-[任意のリンカー配列]-[Cas9]-[任意のリンカー配列]-[第1のUGI]-[任意のリンカー配列]-[第2のUGI]

を含む。

【０３５１】

別の側面において、融合蛋白質は：（i）Cas9酵素またはドメイン；（ii）核酸編集酵素またはドメイン（例えば第2の蛋白質）（例えばシチジンデアミナーゼドメイン）；（iii）2つよりも多くのウラシルグリコシラーゼ阻害因子（UGI）ドメインを含む。

【０３５２】

いくつかの態様において、本明細書において提供される融合蛋白質はリンカー配列を含まない。いくつかの態様においては、任意のリンカー配列の1つまたは両方が存在する。いくつかの態様においては、任意のリンカー配列の1つ、2つ、または3つが存在する。

【０３５３】

いくつかの態様において、上の一般的アーキテクチャに用いられている「-」は任意のリンカー配列の存在を指示する。いくつかの態様において、UGIを含む融合蛋白質は、核標

的化配列、例えば核局在配列をさらに含む。いくつかの態様において、本明細書において提供される融合蛋白質は核局在配列（NLS）をさらに含む。いくつかの態様において、NLSは融合蛋白質のN末端に融合させられる。いくつかの態様において、NLSは融合蛋白質のC末端に融合させられる。いくつかの態様において、NLSはUGI蛋白質のN末端に融合させられる。いくつかの態様において、NLSはUGI蛋白質のC末端に融合させられる。いくつかの態様において、NLSはCas9蛋白質のN末端に融合させられる。いくつかの態様において、NLSはCas9蛋白質のC末端に融合させられる。いくつかの態様において、NLSはデアミナーゼのN末端に融合させられる。いくつかの態様において、NLSはデアミナーゼのC末端に融合させられる。いくつかの態様において、NLSは第2のCas9のN末端に融合させられる。いくつかの態様において、NLSは第2のCas9のC末端に融合させられる。いくつかの態様において、NLSは1つ以上のリンカーを介して融合蛋白質に融合させられる。いくつかの態様において、NLSはリンカーなしで融合蛋白質に融合させられる。いくつかの態様において、NLSは本明細書において提供または参照されるNLS配列のいずれか1つのアミノ酸配列を含む。いくつかの態様において、NLSは配列番号614または配列番号615によって規定されるアミノ酸配列を含む。

【０３５４】

いくつかの態様において、UGIドメインは野生型UGIまたは配列番号134によって規定されるUGIを含む。いくつかの態様において、本明細書において提供されるUGI蛋白質は、UGIの断片、およびUGIまたはUGI断片に対して相同な蛋白質を包含する。例えば、いくつかの態様において、UGIドメインは配列番号134によって規定されるアミノ酸配列の断片を含む。いくつかの態様において、UGI断片は、配列番号134によって規定されるアミノ酸配列の少なくとも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%を含むアミノ酸配列を含む。いくつかの態様において、UGIは、配列番号134によって規定されるアミノ酸配列に対して相同なアミノ酸配列、または配列番号134によって規定されるアミノ酸配列の断片に対して相同なアミノ酸配列を含む。いくつかの態様において、UGIもしくはUGIの断片またはUGIもしくはUGI断片のホモログを含む蛋白質は「UGIバリアント」と言われる。UGIバリアントはUGIまたはその断片に対する相同性を共有する。例えば、UGIバリアントは、野生型UGIまたは配列番号134によって規定されるUGIと少なくとも70%同一、少なくとも75%同一、少なくとも80%同一、少なくとも85%同一、少なくとも90%同一、少なくとも95%同一、少なくとも96%同一、少なくとも97%同一、少なくとも98%同一、少なくとも99%同一、少なくとも99.5%同一、または少なくとも99.9%同一である。いくつかの態様において、UGIバリアントは、断片が野生型UGIまたは配列番号134によって規定されるUGIの対応する断片と少なくとも70%同一、少なくとも80%同一、少なくとも90%同一、少なくとも95%同一、少なくとも96%同一、少なくとも97%同一、少なくとも98%同一、少なくとも99%同一、少なくとも99.5%同一、または少なくとも99.9%であるようなUGIの断片を含む。いくつかの態様において、UGIは次のアミノ酸配列を含む：

＞sp┆P14739┆UNGI_BPPB2ウラシルDNAグリコシラーゼ阻害因子

MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSD
APEYKPWALVIQDSNGENKIKML（配列番号134）

【０３５５】

好適なUGI蛋白質およびヌクレオチド配列が本明細書において提供されており、追加の好適なUGI配列は当業者には公知であり、例えばWang et al., Uracil-DNA glycosylase inhibitor gene of bacteriophage PBS2 encodes a binding protein specific for uracil-DNA glycosylase. J. Biol. Chem. 264: 1163-1171 (1989)；Lundquist et al., Site-directed mutagenesis and characterization of uracil-DNA glycosylase inhibitor protein. Role of specific carboxylic amino acids in complex formation with Escherichia coli uracil-DNA glycosylase. J. Biol. Chem. 272:21408-21419 (1997)；Ravishankar et al., X-ray analysis of a complex of Escherichia coli

uracil DNA glycosylase (EcUDG) with a proteinaceous inhibitor. The structure elucidation of a prokaryotic UDG. Nucleic Acids Res. 26:4880-4887 (1998)；およびPutnam et al., Protein mimicry of DNA from crystal structures of the uracil-DNA glycosylase inhibitor protein and its complex with Escherichia coli uracil-DNA glycosylase. J. Mol. Biol. 287:331-346 (1999)に公開されるものを包含し、それぞれの内容全体は参照によって本明細書に組み込まれる.

【０３５６】

追加の蛋白質がウラシルグリコシラーゼ阻害因子であり得ることは了解されるべきである。例えば、ウラシルDNAグリコシラーゼ塩基除去修復酵素を阻害すること（例えば立体的にブロックすること）ができる他の蛋白質は本開示の範囲内である。加えて、塩基除去修復をブロックまたは阻害するいずれかの蛋白質もまた本開示の範囲内である。いくつかの態様において、本明細書に記載される融合蛋白質は１つのUGIドメインを含む。いくつかの態様において、本明細書に記載される融合蛋白質は1つよりも多くのUGIドメインを含む。いくつかの態様において、本明細書に記載される融合蛋白質は2つのUGIドメインを含む。いくつかの態様において、本明細書に記載される融合蛋白質は2つよりも多くのUGIドメインを含む。いくつかの態様においては、DNAに結合する蛋白質が用いられる。別の態様においては、UGIの代替物が用いられる。いくつかの態様において、ウラシルグリコシラーゼ阻害因子は一本鎖DNAに結合する蛋白質である。例えば、ウラシルグリコシラーゼ阻害因子はErwinia tasmaniensis一本鎖結合蛋白質であり得る。いくつかの態様において、一本鎖結合蛋白質はアミノ酸配列（配列番号135）を含む。いくつかの態様において、ウラシルグリコシラーゼ阻害因子はウラシルに結合する蛋白質である。いくつかの態様において、ウラシルグリコシラーゼ阻害因子はDNA中のウラシルに結合する蛋白質である。いくつかの態様において、ウラシルグリコシラーゼ阻害因子は触媒不活性型ウラシルDNAグリコシラーゼ蛋白質である。いくつかの態様において、ウラシルグリコシラーゼ阻害因子は、DNAからウラシルを切り出さない触媒不活性型ウラシルDNAグリコシラーゼ蛋白質である。例えば、ウラシルグリコシラーゼ阻害因子はUdgXである。いくつかの態様において、UdgXはアミノ酸配列（配列番号136）を含む。別の例として、ウラシルグリコシラーゼ阻害因子は触媒不活性型UDGである。いくつかの態様において、触媒不活性型UDGはアミノ酸配列（配列番号137）を含む。他のウラシルグリコシラーゼ阻害因子は本開示の範囲内であり、当業者には明らかであろうことは了解されるべきである。いくつかの態様において、ウラシルグリコシラーゼ阻害因子は、配列番号135～137または143～148のいずれか1つに対して相同である蛋白質である。いくつかの態様において、ウラシルグリコシラーゼ阻害因子は、配列番号135～137または143～148のいずれか1つと少なくとも50%同一、少なくとも55%同一、少なくとも60%同一、少なくとも65%同一、少なくとも70%同一、少なくとも75%同一、少なくとも80%同一、少なくとも85%同一、少なくとも90%同一、少なくとも95%同一、少なくとも96%同一、少なくとも98%同一、少なくとも99%同一、または少なくとも99.5%同一である蛋白質である。

Erwinia tasmaniensis SSB（耐熱性一本鎖DNA結合蛋白質）

MASRGVNKVILVGNLGQDPEVRYMPNGGAVANITLATSESWRDKQTGETKEKTEWHRV
VLFGKLAEVAGEYLRKGSQVYIEGALQTRKWTDQAGVEKYTTEVVVNVGGTMQMLGGR
SQGGGASAGGQNGGSNNGWGQPQQPQGGNQFSGGAQQQARPQQQPQQNNAPANNEPP
IDFDDDIP（配列番号135）

UdgX（DNA中のウラシルに結合するが、切り出さない）

MAGAQDFVPHTADLAELAAAAGECRGCGLYRDATQAVFGAGGRSARIMMIGEQPGDKED
LAGLPFVGPAGRLLDRALEAADIDRDALYVTNAVKHFKFTRAAGGKRRIHKTPSRTEVVA
CRPWLIAEMTSVEPDVVVLLGATAAKALLGNDFRVTQHRGEVLHVDDVPGDPALVATV
HPSSLLRGPKEERESAFAGLVDDLRVAADVRP（配列番号136）

UDG（触媒不活性型ヒトUDG。DNA中のウラシルに結合するが、切り出さない）

MIGQKTLYSFFSPSPARKRHAPSPEPAVQGTGVAGVPEESGDAAAIPAKKAPAGQEEPGT
PPSSPLSAEQLDRIQRNKAAALLRLAARNVPVGFGESWKKHLSGEFGKPYFIKLMGFVAE

ERKHYTVYPPPHQVFTWTQMCDIKDVKVVILGQEPYHGPNQAHGLCFSVQRPVPPPPSL
ENIYKELSTDIEDFVHPGHGDLSGWAKQGVLLLNAVLTVRAHQANSHKERGWEQFTDAV
VSWLNQNSNGLVFLLWGSYAQKKGSAIDRKRHHVLQTAHPSPLSVYRGFFGCRHFSKTN
ELLQKSGKKPIDWKEL（配列番号137）

【０３５７】

UGIとして用いられ得る追加の一本鎖DNA結合蛋白質は下に示されている。UGIとして、他の一本鎖結合蛋白質、例えばDickey TH, Altschuler SE, Wuttke DS. Single-stranded DNA-binding proteins :multiple domains for multiple functions. Structure. 2013 Jul 2;21(7): 1074-84. doi: 10.1016/j.str.2013.05.013. Review.; Marceau AH. Functions of single-strand DNA-binding proteins in DNA replication, recombination, and repair. Methods Mol Biol. 2012;922: 1-21. doi: 10.1007/978-1-62703-032-8_l.; Mijakovic, Ivan, et al., Bacterial single-stranded DNA-binding proteins are phosphorylated on tyrosine. Nucleic Acids Res 2006; 34 (5): 1588-1596. doi: 10.1093/nar/gkj514; Mumtsidu E, Makhov AM, Konarev PV, Svergun DI, Griffith JD, Tucker PA. Structural features of the single-stranded DNA-binding protein of Epstein-Barrvirus. J Struct Biol. 2008 Feb;161(2): 172-87. Epub 2007 Nov 1; Nowak M, Olszewski M, Spibida M, Kur J. Characterization of single-stranded DNA-binding proteins from the psychrophilic bacteria Desulfotalea psychrophila, Flavobacterium psychrophilum, Psychrobacter arcticus, Psychrobactercryohalolentis, Psychromonas ingrahamii, Psychroflexus torquis, and Photobacterium profundum. BMC Microbiol. 2014 Apr 14;14:91. doi: 10.1186/1471-2180-14-91; Tone T, Takeuchi A, Makino O. Single-stranded DNA binding protein Gp5 of Bacillus subtilis phage Φ29 is required for viral DNA replication in growth-temperature dependent fashion. Biosci Biotechnol Biochem. 2012;76(12):2351-3. Epub 2012 Dec 7; Wold. REPLICATION PROTEIN A: A Heterotrimeric, Single-Stranded DNA-Binding Protein Required for Eukaryotic DNA Metabolism. Annual Review of Biochem. 1997; 66:61-92. doi: 10.1146/annurev.biochem.66.1.61; Wu Y, Lu J, Kang T. Human single-stranded DNA binding proteins: guardians of genome stability. Acta Biochim Biophys Sin (Shanghai). 2016 Jul;48(7):671-7. doi: 10.1093/abbs/gmw044. Epub 2016 May 23. Reviewに記載されるものが用いられ得ることは了解されるべきである；それぞれの内容全体は参照によってここに組み込まれる。

mtSSB - SSBP1一本鎖DNA結合蛋白質1 [Homo sapiens (ヒト)]（UniProtKB: Q04837; NP_001243439.1）

MFRRPVLQVLRQFVRHESETTTSLVLERSLNRVHLLGRVGQDPVLRQVEGKNPVTIFSLA
TNEMWRSGDSEVYQLGDVSQKTTWHRISVFRPGLRDVAYQYVKKGSRIYLEGKIDYGEY
MDKNNVRRQATTIIADNIIFLSDQTKEKE（配列番号138）

一本鎖DNA結合蛋白質3アイソフォームA [Mus musculus]（UniProtKBQ9D032-1; NCBI Ref: NP_076161.2）

MFAKGKGSAVPSDGQAREKLALYVYEYLLHVGAQKSAQTFLSEIRWEKNITLGEPPGFLH
SWWCVFWDLYCAAPERRDTCEHSSEAKAFHDYSAAAAPSPVLGNIPPNDGMPGGPIPPG
FFQGPPGSQPSPHAQPPPHNPSSMMGPHSQPFMSPRYAGGPRPPIRMGNQPPGGVPGT
QPLLPNSMDPTRQQGHPNMGGSMQRMNPPRGMGPMGPGPQNYGSGMRPPPNSLGPA
MPGINMGPGAGRPWPNPNSANSIPYSSSSPGTYVGPPGGGGPPGTPIMPSPADSTNSSDN
IYTMINPVPPGGSRSNFPMGPGSDGPMGGMGGMEPHHMNGSLGSGDIDGLPKNSPNNIS
GISNPPGTPRDDGELGGNFLHSFQNDNYSPSMTMSV（配列番号139）

RPA1 - 複製蛋白質A 70kDa DNA結合サブユニット（UniProtKB: P27694；NCBI Ref: NM_002945.3）

MVGQLSEGAIAAIMQKGDTNIKPILQVINIRPITTGNSPPRYRLLMSDGLNTLSSFMLATQ
LNPLVEEEQLSSNCVCQIHRFIVNTLKDGRRVVILMELEVLKSAEAVGVKIGNPVPYNEGL
GQPQVAPPAPAASPAASSRPQPQNGSSGMGSTVSKAYGASKTFGKAAGPSLSHTSGGTQS
KVVPIASLTPYQSKWTICARVTNKSQIRTWSNSRGEGKLFSLELVDESGEIRATAFNEQVD
KFFPLIEVNKVYYFSKGTLKIANKQFTAVKNDYEMTFNNETSVMPCEDDHHLPTVQFDF
TGIDDLENKSKDSLVDIIGICKSYEDATKITVRSNNREVAKRNIYLMDTSGKVVTATLWGE
DADKFDGSRQPVLAIKGARVSDFGGRSLSVLSSSTIIANPDIPEAYKLRGWFDAEGQALDG
VSISDLKSGGVGGSNTNWKTLYEVKSENLGQGDKPDYFSSVATVVYLRKENCMYQACPT
QDCNKKVIDQQNGLYRCEKCDTEFPNFKYRMILSVNIADFQENQWVTCFQESAEAILGQ
NAAYLGELKDKNEQAFEEVFQNANFRSFIFRVRVKVETYNDESRIKATVMDVKPVDYREY
GRRLVMSIRRSALM（配列番号140）
RPA2 - 複製蛋白質A 32kDaサブユニット（UniProtKB: P15927；NCBI Ref: NM_00
2946）
MWNSGFESYGSSSYGGAGGYTQSPGGFGSPAPSQAEKKSRARAQHIVPCTISQLLSATLV
DEVFRIGNVEISQVTIVGIIRHAEKAPTNIVYKIDDMTAAPMDVRQWVDTDDTSSENTVV
PPETYVKVAGHLRSFQNKKSLVAFKIMPLEDMNEFTTHILEVINAHMVLSKANSQPSAGR
APISNPGMSEAGNFGGNSFMPANGLTVAQNQVLNLIKACPRPEGLNFQDLKNQLKHMSV
SSIKQAVDFLSNEGHIYSTVDDDHFKSTDAE（配列番号141）
RPA3 - 複製蛋白質A 14kDaサブユニット（UniProtKB: P35244；NCBI Ref:NM_002
947.4）
MVDMMDLPRSRINAGMLAQFIDKPVCFVGRLEKIHPTGKMFILSDGEGKNGTIELMEPL
DEEISGIVEVVGRVTAKATILCTSYVQFKEDSHPFDLGLYNEAVKIIHDFPQFYPLGIVQH
D（配列番号142）
<u>細菌一本鎖DNA結合蛋白質：</u>
ssbA - 一本鎖DNA結合蛋白質 [Bacillus subtilis subsp. subtilis str. 168]（UniPro
tKB: P37455；NCBI Ref: ）
MLNRVVLVGR LTKDPELRYT PNGAAVATFT LAVNRTFTNQ SGEREADFIN
CVTWRRQAEN VANFLKKGSL AGVDGRLQTR NYENQQGQRV FVTEVQAESV
QFLEPKNGGG SGSGGYNEGN SGGGQYFGGG QNDNPFGGNQ NNQRRNQGNS
FNDDPFANDG KPIDISDDDLPF（配列番号143）
一本鎖DNA結合蛋白質2 [Streptomyces coelicolor A3（2）]（UniProtKB: Q9X8U3
；NCBI Ref: NP_628093.1）
MAGETVITVVGNLVDDPELRFTPSGAAVAKFRVASTPRTFDRQTNEWKDGESLFLTCSV
WRQAAENVAESLQRGMRVIVQGRLKQRSYEDREGVKRTVYELDVDEVGASLRSATAKVT
KTSGQGRGGQGGYGGGGGGQGGGGWGGGPGGGQQGGGAPADDPWATGGAPAGGQQGG
GGQGGGGWGGGSGGGGGYSDEPPF（配列番号144）
一本鎖DNA結合蛋白質 [Streptococcus pneumoniae R6]（UniProtKB: P66855；NC
BI Ref: NP_358988.1）
MINNVVLVGRMTRDAELRYTPSNVAVATFTLAVNRTFKSQNGEREADFINVVMWRQQA
ENLANWAKKGSLIGVTGRIQTRSYDNQQGQRVYVTEVVAENFQMLESRSVREGHTGGAY
SAPTANYSAPTNSVPDFSRNENPFGATNPLDISDDDLPF（配列番号145）
<u>ウイルス一本鎖 DNA結合蛋白質：</u>
一本鎖DNA結合蛋白質 [ヒトアルファヘルペスウイルス1]（UniProtKB：P04296；NCB
I Ref：YP_009137104.1）
METKPKTATTIKVPPGPLGYVYARACPSEGIELLALLSARSGDSDVAVAPLVVGLTVESGF
EANVAVVVGSRTTGLGGTAVSLKLTPSHYSSSVYVFHGGRHLDPSTQAPNLTRLCERARR
HFGFSDYTPRPGDLKHETTGEALCERLGLDPDRALLYLVVTEGFKEAVCINNTFLHLGGS
DKVTIGGAEVHRIPVYPLQLFMPDFSRVIAEPFNANHRSIGENFTYPLPFFNRPLNRLLFE
AVVGPAAVALRCRNVDAVARAAAHLAFDENHEGAALPADITFTAFEASQGKTPRGGRDG
GGKGPAGGFEQRLASVMAGDAALALESIVSMAVFDEPPTDISAWPLFEGQDTAAARANA

VGAYLARAAGLVGAMVFSTNSALHLTEVDDAGPADPKDHSKPSFYRFFLVPGTHVAANP
QVDREGHVVPGFEGRPTAPLVGGTQEFAGEHLAMLCGFSPALLAKMLFYLERCDGGVIV
GRQEMDVFRYVADSNQTDVPCNLCTFDTRHACVHTTLMRLRARHPKFASAARGAIGVFG
TMNSMYSDCDVLGNYAAFSALKRADGSETARTIMQETYRAATERVMAELETLQYVDQA
VPTAMGRLETIITNREALHTVVNNVRQVVDREVEQLMRNLVEGRNFKFRDGLGEANHA
MSLTLDPYACGPCPLLQLLGRRSNLAVYQDLALSQCHGVFAGQSVEGRNFRNQFQPVLR
RRVMDMFNNGFLSAKTLTVALSEGAAICAPSLTAGQTAPAESSFEGDVARVTLGFPKELR
VKSRVLFAGASANASEAAKARVASLQSAYQKPDKRVDILLGPLGFLLKQFHAAIFPNGKP
PGSNQPNPQWFWTALQRNQLPARLLSREDIETIAFIKKFSLDYGAINFINLAPNNVSELA
MYYMANQILRYCDHSTYFINTLTAIIAGSRRPPSVQAAAAAWSAQGGAGLEAGARALMDA
VDAHPGAWTSMFASCNLLRPVMAARPMVVLGLSISKYYGMAGNDRVFQAGNWASLMG
GKNACPLLIFDRTRKFVLACPRAGFVCAASSLGGGAHESSLCEQLRGIISEGGAAVASSVF
VATVKSLGPRTQQLQIEDWLALLEDEYLSEEMMELTARALERGNGEWSTDAALEVAHEA
EALVSQLGNAGEVFNFGDFGCEDDNATPFGGPGAPGPAFAGRKRAFHGDDPFGEGPPDK
KGDLTLDML（配列番号146）
Bacillus virus phi29からの一本鎖DNA結合蛋白質（UniProtKB：Q38504.1；NCBI Ref：YP_002004532.1）
MENTNIVKATFDTETLEGQIKIFNAQTGGGQSFKNLPDGTIIEANAIAQYKQVSDTYGDAK
EETVTTIFAADGSLYSAISKTVAEAASDLIDLVTRHKLETFKVKVVQGTSSKGNVFFSLQLS
L（配列番号147）
一本鎖DNA結合蛋白質 [Burkholderia virus DC1]（UniProtKB：I6NRL7；NCBI Ref：YP_006589943.1）
MASVNKVILVGNLGADPETRYLPSGDAISNIRLATTDRYKDKASGEMKESTEWHRVSFFG
RLAEIVDEYLRKGAPVYIEGRIRTRKWQDNAGQDRYTTEIVAEKMQMLGDRRDGGERQQ
RAPQQQQQRTQRNGYADATGRAQPSQRPAAGGGFDEMDDDIPF（配列番号148）

【０３５８】

いくつかの態様において、核酸編集ドメインはデアミナーゼドメインである。いくつかの態様において、デアミナーゼ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る。いくつかの態様において、デアミナーゼはアポリポ蛋白質B mRNA編集複合体（APOBEC）ファミリーデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC1デアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC2デアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC3デアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC3Aデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC3Bデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC3Cデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC3Dデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC3Eデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC3Fデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC3Gデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC3Hデアミナーゼである。いくつかの態様において、デアミナーゼはAPOBEC4デアミナーゼである。いくつかの態様において、デアミナーゼは活性化誘導デアミナーゼ（AID）である。いくつかの態様において、デアミナーゼ（demianse）はラットAPOBEC1（配列番号74）である。いくつかの態様において、デアミナーゼはヒトAPOBEC1（配列番号76）である。いくつかの態様において、デアミナーゼはPetromyzon marinusシチジンデアミナーゼ1（pmCDA1）である。いくつかの態様において、デアミナーゼはヒトAPOBEC3G（配列番号60）である。いくつかの態様において、デアミナーゼはヒトAPOBEC3Gの断片（配列番号83）である。いくつかの態様において、デアミナーゼはD316R_D317R変異を含むヒトAPOBEC3Gバリアント（配列番号82）である。いくつかの態様において、デアミナーゼはヒトAPOBEC3Gの断片であり、配列番号60のD316R_D317R変異に対応する変異を含んでいる（配列番号84）。

【０３５９】
いくつかの態様において、リンカーは、$(GGGS)_n$（配列番号613）、$(GGGGS)_n$（配列番号607）、$(G)_n$（配列番号608）、$(EAAAK)_n$（配列番号609）、$(GGS)_n$（配列番号610）、SGSETPGTSESATPES（配列番号604）、もしくは$(XP)_n$（配列番号611）モチーフ、またはこれらのいずれかの組み合わせを含み、nは独立して1と30との間の整数である。
【０３６０】
好適なUGI蛋白質およびヌクレオチド配列が本明細書において提供されており、追加の好適なUGI配列は当業者に公知であり、例えば、Wang et al., Uracil-DNA glycosylase inhibitor gene of bacteriophage PBS2 encodes a binding protein specific for uracil-DNA glycosylase. J. Biol. Chem. 264: 1163-1171 (1989)；Lundquist et al., Site-directed mutagenesis and characterization of uracil-DNA glycosylase inhibitor protein. Role of specific carboxylic amino acids in complex formation with Escherichia coli uracil-DNA glycosylase. J. Biol. Chem. 272:21408-21419 (1997)；Ravishankar et al., X-ray analysis of a complex of Escherichia coli uracil DNA glycosylase (EcUDG) with a proteinaceous inhibitor. The structure elucidation of a prokaryotic UDG. Nucleic Acids Res. 26:4880-4887 (1998)；およびPutnam et al., Protein mimicry of DNA from crystal structures of the uracil-DNA glycosylase inhibitor protein and its complex with Escherichia coli uracil-DNA glycosylase. J. Mol. Biol. 287:331-346 (1999)に公開されるものを包含し、これらの内容全体は参照によって本明細書に組み込まれる。いくつかの態様において、任意のリンカーは$(GGS)_n$（配列番号610）モチーフを含み、nは1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、または20である。いくつかの態様において、任意のリンカーは$(GGS)_n$（配列番号610）モチーフを含み、nは1、3、または7である。いくつかの態様において、任意のリンカーはアミノ酸配列SGSETPGTSESATPES（配列番号604）を含み、これは例においてはXTENリンカーともまた言われる。
【０３６１】
いくつかの態様において、Cas9ニッカーゼは、ゲノムがin vivoで編集される生物の編集されない鎖上の塩基の取り除きをさらに容易化し得る。本明細書に記載される通り、Cas9ニッカーゼは、配列番号6のD10A変異、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異を含み得る。いくつかの態様において、本開示のCas9ニッカーゼは、配列番号6の変異840のヒスチジン、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する残基を含み得る。Cas9ニッカーゼを含むかかる融合蛋白質は、標的DNA配列の一本の鎖、例えば編集されようとしない鎖を切断し得る。いずれの特定の理論によっても拘束されることを望むものではないが、この切断は、デアミナーゼによってなされたC→U編集を逆行させるミスマッチ修復メカニズムを阻害し得る。
【０３６２】
ガイドRNAとのCas9複合体
本開示のいくつかの側面は複合体を提供し、本明細書において提供される融合蛋白質のいずれかと融合蛋白質のCas9ドメイン（例えば、dCas9、ヌクレアーゼ活性型Cas9、またはCas9ニッカーゼ）に結合したガイドRNAとを含む。
【０３６３】
いくつかの態様において、ガイドRNAは15～100ヌクレオチドの長さであり、標的配列に対して相補的である少なくとも10個の一続きのヌクレオチドの配列を含む。いくつかの態様において、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。いくつかの態様において、ガイドRNAは、標的配列に対して相補的である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、または40個の一続きのヌクレオチドの配列を含む。いくつかの態様において、標的配列はDNA配列である。いくつかの態様において、標的配列は哺乳動物のゲノム中の配列である。いくつかの態様において、標的配列はヒトのゲノム中の配列である。いくつかの態様において、標的配列の3'末端は古典的PAM配列（NGG）に直ちに隣接している。いくつかの態様において、ガイドRNAは疾患または異常に関連する配列に対して相補的である。

【０３６４】

Cas9融合蛋白質を用いる方法

本開示のいくつかの側面は、本明細書において提供されるCas9蛋白質、融合蛋白質、または複合体を用いる方法を提供する。例えば、本開示のいくつかの側面は、DNA分子を（a）本明細書において提供されるCas9蛋白質または融合蛋白質のいずれかおよび少なくとも1つのガイドRNAに（ガイドRNAは約15～100ヌクレオチドの長さであり、標的配列に対して相補的である少なくとも10個の一続きのヌクレオチドの配列を含む）；または（b）本明細書において提供される少なくとも1つのgRNAとのCas9蛋白質、Cas9融合蛋白質、またはCas9蛋白質もしくは融合蛋白質複合体に、接触させることを含む方法を提供する。いくつかの態様において、標的配列の3'末端は古典的PAM配列（NGG）に直ちに隣接はしない。いくつかの態様において、標的配列の3'末端はAGC、GAG、TTT、GTG、またはCAA配列に直ちに隣接する。

【０３６５】

いくつかの態様において、標的DNA配列は疾患または異常に関連する配列を含む。いくつかの態様において、標的DNA配列は疾患または異常に関連する点変異を含む。いくつかの態様において、Cas9蛋白質、Cas9融合蛋白質、または複合体の活性は点変異の修正をもたらす。いくつかの態様において、標的DNA配列は疾患または異常に関連するT→C点変異を含み、変異体C塩基の脱アミノ化は疾患または異常に関連しない配列をもたらす。いくつかの態様において、標的DNA配列は蛋白質をコードし、点変異はコドン内にあり、野生型コドンと比較して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。いくつかの態様において、変異体Cの脱アミノ化は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。いくつかの態様において、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。いくつかの態様において、接触させることは対象においてin vivoである。いくつかの態様において、対象は疾患または異常を有するか、またはそれと診断されている。いくつかの態様において、疾患または異常は、嚢胞性線維症、フェニルケトン尿症、表皮剥離性角化症（EHK）、4J型シャルコー・マリー・トゥース病、神経芽腫（NB）、フォン・ヴィレブランド病（vWD）、先天性ミオトニー、遺伝性腎アミロイドーシス、拡張型心筋症（DCM）、遺伝性リンパ浮腫、家族性アルツハイマー病、HIV、プリオン病、慢性乳児神経皮膚関節症候群（CINCA）、デスミン関連ミオパチー（DRM）、変異体PI3KCA蛋白質、変異体CTNNB1蛋白質、変異体HRAS蛋白質、または変異体p53蛋白質に関連する新生物性疾患である。

【０３６６】

いくつかの態様は、本明細書において提供されるCas9 DNA編集融合蛋白質を用いるための方法を提供する。いくつかの態様において、融合蛋白質は、標的核酸塩基、例えばC残基を脱アミノ化することによって核酸に点変異を導入するために用いられる。いくつかの態様において、標的核酸塩基の脱アミノ化は、遺伝子の欠陥の修正、例えば遺伝子産物の機能喪失に至る点変異の修正をもたらす。いくつかの態様において、遺伝子の欠陥は、疾患または異常、例えばリソソーム蓄積症または例えばI型糖尿病などの代謝疾患に関連する。いくつかの態様において、本明細書において提供される方法は、疾患または異常に関連する遺伝子産物をコードする遺伝子またはアレルに脱活性化点変異を導入するために用いられる。例えば、いくつかの態様においては、（例えば増殖性疾患の処置において）癌遺伝子に脱活性化点変異を導入するためにCas9 DNA編集融合蛋白質を採用する方法が本明細書において提供される。いくつかの態様において、脱活性化変異はコード配列中に未

成熟終止コドンを創出し得、これは、短縮された遺伝子産物、例えば全長蛋白質の機能を欠く短縮型蛋白質の発現をもたらす。

【０３６７】

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、ゲノム編集によって機能不全遺伝子の機能を回復させることである。本明細書において提供されるCas9デアミナーゼ融合蛋白質は、遺伝子編集に基づくヒト治療法について、in vitroで、例えばヒト細胞培養において疾患関連変異を修正することによって検証され得る。本明細書において提供される融合蛋白質、例えばCas9ドメインと核酸デアミナーゼドメインとを含む融合蛋白質が、いずれかの単一のT→CまたはA→G点変異を修正するために用いられ得ることは当業者によって理解されるであろう。第1のケースでは、変異体Cから再びU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Gと塩基対形成しているCの脱アミノ化、次に複製のラウンドが、変異を修正する。

【０３６８】

提供される融合蛋白質によってin vitroまたはin vivoで修正され得る例示的な疾患に該当する変異は、PI3KCA蛋白質のH1047R(A3140G)多型である。ホスホイノシチド-3-キナーゼ触媒アルファサブユニット（PI3KCA）蛋白質はホスファチジルイノシトールのイノシトール環の3-OH基をリン酸化するように作用する。PI3KCA遺伝子は多くの異なる癌腫において変異していることが見出されており、それゆえに、それは強力な癌遺伝子であると考えられている[37]。事実、A3140G変異は例えばHCT116、SKOV3、およびT47D細胞株などのいくつかのNCI-60癌細胞株に存在し、これらはアメリカンタイプカルチャーコレクション（ATCC）から容易く入手可能である[38]。

【０３６９】

いくつかの態様においては、修正されるべき変異を持つ細胞、例えば点変異、例えばPI3KCA蛋白質のH1047R置換をもたらすPI3KCA遺伝子のエキソン20のA3140G点変異を持つ細胞が、Cas9デアミナーゼ融合蛋白質と、コードするPI3KCA遺伝子のそれぞれの変異部位へと融合蛋白質を標的化する適切に設計されたsgRNAとをコードする発現構築物に接触させられる。対照実験が行われ得、ここでは、融合酵素をPI3KCA遺伝子内にある非C残基へと標的化するようにsgRNAが設計される。処置された細胞のゲノムDNAが抽出され得、PI3KCA遺伝子の該当する配列がPCR増幅およびシーケンシングされて、ヒト細胞培養における融合蛋白質の活性を査定し得る。

【０３７０】

PI3KCAの点変異を修正する例は例示目的のために提供されており、本開示を限定することを意味されないことは理解されるであろう。当業者は、本開示のDNA編集融合蛋白質が、他の癌および他の増殖性疾患を包含する癌以外の疾患に関連する他の点変異および変異を修正するために用いられ得ることを理解するであろう。

【０３７１】

疾患関連遺伝子およびアレルの点変異の首尾良い修正は、治療法および基礎研究における用途を有する遺伝子修正のための新たな戦略を開く。Cas9とデアミナーゼ酵素またはドメインとの開示される融合体のような部位特異的1塩基改変システムは、「逆」遺伝子治療にもまた用途を有し、ここでは、一定の遺伝子機能が意図的に抑圧または廃絶される。これらのケースでは、Trp（TGG）、Gln（CAAおよびCAG）、またはArg（CGA）残基を未成熟終止コドン（TAA、TAG、TGA）へと部位特異的に変異させることが、in vitro、ex vivo、またはin vivoで蛋白質機能を廃絶するために用いられ得る。

【０３７２】

本開示は、本明細書において提供されるCas9 DNA編集融合蛋白質によって修正され得る点変異に関連するかまたはそれによって引き起こされる疾患と診断された対象の処置のための方法を提供する。例えば、いくつかの態様においては、かかる疾患、例えば上に記載されているPI3KCA点変異に関連する癌を有する対象に、点変異を修正するかまたは疾患関連遺伝子に脱活性化変異を導入するCas9デアミナーゼ融合蛋白質の有効量を投与することを含む方法が提供される。いくつかの態様において、疾患は増殖性疾患である。いく

つかの態様において、疾患は遺伝子疾患である。いくつかの態様において、疾患は新生物性疾患である。いくつかの態様において、疾患は代謝疾患である。いくつかの態様において、疾患はリソソーム蓄積症である。点変異を修正するかまたは疾患関連遺伝子に脱活性化変異を導入することによって処置され得る他の疾患は当業者には公知であろう。本開示はこれについて限定されない。

【０３７３】

本開示は、追加の疾患または異常、例えば、デアミナーゼによって媒介される遺伝子編集によって修正され得る点変異に関連するかまたはそれによって引き起こされる疾患または異常の処置のための方法を提供する。いくつかのかかる疾患は本明細書に記載されており、本明細書において提供される戦略および融合蛋白質によって処置され得る追加の好適な疾患は本開示に基づいて当業者には明らかであろう。例示的な好適な疾患および異常は下でリスト化されている。それぞれの配列中の具体的な位置または残基の付番は用いられる特定の蛋白質および付番スキームに依存することは理解されるであろう。付番は例えば成熟蛋白質の前駆体および成熟蛋白質それ自体では異なり得、種から種への配列の違いが付番に影響し得る。当業者は、当分野において周知の方法によって、例えば相同な残基の配列アライメントおよび決定によって、いずれかの相同な蛋白質中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能力があるであろう。例示的な好適な疾患および異常は、限定なしに、次を包含する：嚢胞性線維症（例えば、Schwank et al., Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. Cell stem cell. 2013; 13: 653-658；およびWu et. al., Correction of a genetic disease in mouse via use of CRISPR-Cas9. Cell stem cell. 2013; 13: 659-662を見よ。これらのどちらも遺伝子の欠陥を修正するためにデアミナーゼ融合蛋白質を用いてはいない）；フェニルケトン尿症-例えば、フェニルアラニンヒドロキシラーゼ遺伝子の位置835（マウス）もしくは240（ヒト）または相同残基におけるフェニルアラニン→セリン変異（T＞C変異）- 例えばMcDonald et al, Genomics. 1997; 39:402-405を見よ；ベルナール・スーリエ症候群（BSS）-例えば、血小板膜糖蛋白質IXの位置55もしくは相同残基におけるフェニルアラニン→セリン変異または残基24もしくは相同残基におけるシステイン→アルギニン（T＞C変異）-例えば、Noris et al., British Journal of Haematology. 1997; 97: 312-320およびAli et al., Hematol. 2014; 93: 381-384を見よ；表皮剥離性角化症（EHK）-例えば、ケラチン1の位置160もしくは161（開始メチオニンをカウントする場合）または相同残基におけるロイシン→プロリン変異（T＞C変異）-例えばChipev et al., Cell. 1992; 70: 821-828を見よ。www.uniprot.orgにおけるUNIPROTデータベースのアクセッション番号P04264をもまた見よ；慢性閉塞性肺疾患（COPD）-例えば、α_1-アンチトリプシンのプロセシングされた形態の位置54もしくは55（開始メチオニンをカウントする場合）または相同残基あるいは未プロセシング形態の残基78または相同残基におけるロイシン→プロリン変異（T＞C変異）-例えばPoller et al., Genomics. 1993; 17: 740-743を見よ。UNIPROTデータベースのアクセッション番号P01011をもまた見よ；4J型シャルコー・マリー・トゥース病-例えばFIG4の位置41または相同残基におけるイソロイシン→トレオニン変異（T＞C変異）-例えばLenk et al., PLoS Genetics. 2011; 7: el002104を見よ；神経芽腫（NB）- 例えばCaspase-9の位置197または相同残基におけるロイシン→プロリン変異（T＞C変異）-例えばKundu et al., 3 Biotech. 2013, 3:225-234を見よ；フォン・ヴィレブランド病（vWD）-例えば、フォン・ヴィレブランド因子のプロセシングされた形態の位置509もしくは相同残基またはフォン・ヴィレブランド因子の未プロセシング形態の位置1272もしくは相同残基におけるシステイン→アルギニン変異（T＞C変異）-例えばLavergne et al., Br. J. Haematol. 1992を見よ。UNIPROTデータベースのアクセッション番号P04275; 82: 66-72をもまた見よ；先天性ミオトニー-例えば筋塩素チャネル遺伝子CLCN1の位置277または相同残基におけるシステイン→アルギニン変異（T＞C変異）-例えばWeinberger et al., The J. of Physiology. 2012; 590: 3449-3464を見よ；遺伝性腎アミロイドーシス-例えば、アポリポ蛋白質AIIのプロセシングされた形態の位置78もしくは相同残基または

未プロセシング形態の位置101もしくは相同残基における終止コドン→アルギニン変異（T＞C変異）-例えばYazaki et al., Kidney Int. 2003; 64: 11-16を見よ；拡張型心筋症（DCM）-例えばFOXD4遺伝子の位置148または相同残基におけるトリプトファン→アルギニン変異（T＞C変異）、例えばMinoretti et. al., Int. J. of Mol. Med. 2007; 19: 369-372を見よ；遺伝性リンパ浮腫-例えばVEGFR3チロシンキナーゼの位置1035または相同残基におけるヒスチジン→アルギニン変異（A＞G変異）、例えばIrrthum et al., Am. J. Hum. Genet. 2000; 67: 295-301を見よ；家族性アルツハイマー病-例えばプレセニリン1の位置143または相同残基におけるイソロイシン→バリン変異（A＞G変異）、例えばGallo et. al., J. Alzheimer's disease. 2011; 25: 425-431を見よ；プリオン病-例えばプリオン蛋白質の位置129または相同残基におけるメチオニン→バリン変異（A＞G変異）-例えばLewis et. al., J. of General Virology. 2006; 87: 2443-2449を見よ；慢性乳児神経皮膚関節症候群（CINCA）-例えばクリオピリンの位置570または相同残基におけるチロシン→システイン変異（A＞G変異）-例えばFujisawa et. al. Blood. 2007; 109: 2903-2911を見よ；ならびにデスミン関連ミオパチー（DRM）-例えばαβクリスタリンの位置120または相同残基におけるアルギニン→グリシン変異（A＞G変異）-例えばKumar et al., J. Biol. Chem. 1999; 274: 24137-24141を見よ。全ての参照およびデータベースエントリーの内容全体は参照によって本明細書に組み込まれる。

【０３７４】

本明細書において開示されるCas9:核酸編集酵素/ドメイン融合蛋白質を標的部位、例えば編集されるべき点変異を含む部位へと標的化するためには、典型的には、Cas9:核酸編集酵素/ドメイン融合蛋白質をガイドRNA、例えばsgRNAと一緒に共発現することが必要であることは当業者には明らかであろう。本明細書において他所でより詳細に説明されている通り、典型的には、ガイドRNAは、Cas9結合を許すtracrRNAフレームワークと、Cas9:核酸編集酵素/ドメイン融合蛋白質に配列特異性を授けるガイド配列とを含む。いくつかの態様において、ガイドRNAは構造5'-[ガイド配列]-guuuuagagcuagaaauagcaaguuaaaauaaggcuaguccguuaucaacuugaaaaaguggcaccgagucggugcuuuuu-3'（配列番号618）を含み、ガイド配列は、標的配列に対して相補的である配列を含む。ガイド配列は典型的には20ヌクレオチドの長さである。Cas9:核酸編集酵素/ドメイン融合蛋白質を特異的なゲノム標的部位へと標的化することにとって好適なガイドRNAの配列は、本開示に基づいて当業者には明らかであろう。典型的には、かかる好適なガイドRNA配列は、編集されるべき標的ヌクレオチドの上流または下流50ヌクレオチド以内の核酸配列に対して相補的であるガイド配列を含む。Cas9:核酸編集酵素/ドメイン融合蛋白質を特異的な標的配列へと標的化することにとって好適ないくつかの例示的なガイドRNA配列は下で提供されている。

【０３７５】

塩基編集因子効率

本開示のいくつかの側面は、本明細書において提供される塩基編集因子のいずれかが有意な割合のインデルを創出することなしに特異的なヌクレオチド塩基を改変することができるという認識に基づく。本明細書において用いられる「インデル」は核酸中のヌクレオチド塩基の挿入または欠失を言う。かかる挿入または欠失は遺伝子のコード領域内のフレームシフト変異に至り得る。いくつかの態様においては、核酸に多数の挿入または欠失（すなわちインデル）を創出することなしに核酸中の特異的なヌクレオチドを効率的に改変する（例えば、変異させるかまたは脱アミノ化する）塩基編集因子を創出することが望ましい。ある態様において、本明細書において提供される塩基編集因子のいずれかは、インデルに対してより多大な割合の意図される改変（例えば点変異または脱アミノ化）を創出することができる。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編集因子は、1:1よりも多大である意図される点変異対インデルの比を創出することができる。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編集因子は、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、または少なくとも1000:1、またはより多くである意図される点変異対インデルの比を創出することができる。意図される変異およびインデルの数は、いずれかの好適な方法、例えば下の例に用いられている方法を用いて決定され得る。

【０３７６】

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編集因子は、核酸のある領域におけるインデルの形成を限定することができる。いくつかの態様において、領域は、塩基編集因子によって標的化されるヌクレオチドか、または塩基編集因子によって標的化されるヌクレオチドから2、3、4、5、6、7、8、9、もしくは10ヌクレオチド以内の領域である。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編集因子のいずれかは、核酸のある領域におけるインデルの形成を、1%よりも少し、1.5%よりも少し、2%よりも少し、2.5%よりも少し、3%よりも少し、3.5%よりも少し、4%よりも少し、4.5%よりも少し、5%よりも少し、6%よりも少し、7%よりも少し、8%よりも少し、9%よりも少し、10%よりも少し、12%よりも少し、15%よりも少し、または20%よりも少しまでに限定することができる。ある核酸領域において形成されるインデルの数は、核酸（例えば、細胞のゲノム中の核酸）が塩基編集因子に暴露される時間の量に依存し得る。いくつかの態様において、インデルの数または割合は、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日の核酸（例えば、細胞のゲノム中の核酸）を塩基編集因子に暴露することの後に決定される。

【０３７７】

本開示のいくつかの側面は、本明細書において提供される塩基編集因子のいずれかが、有意な数の意図されない変異、例えば意図されない点変異を創出することなしに、核酸（例えば、対象のゲノム中の核酸）に点変異などの意図される変異を効率的に創出することができるという認識に基づく。いくつかの態様において、意図される変異は、意図される変異を創出するように特異的に設計されたgRNAに結合した特異的な塩基編集因子によって創出される変異である。いくつかの態様において、意図される変異は疾患または異常に関連する変異である。いくつかの態様において、意図される変異は、疾患または異常に関連するシトシン（C）→チミン（T）点変異である。いくつかの態様において、意図される変異は疾患または異常に関連するグアニン（G）→アデニン（A）点変異である。いくつかの態様において、意図される変異は遺伝子のコード領域内のシトシン（C）→チミン（T）点変異である。いくつかの態様において、意図される変異は遺伝子のコード領域内のグアニン（G）→アデニン（A）点変異である。いくつかの態様において、意図される変異は、終止コドン、例えば遺伝子のコード領域内の未成熟終止コドンを創出する点変異である。いくつかの態様において、意図される変異は終止コドンを消去する変異である。いくつかの態様において、意図される変異は遺伝子のスプライシングを変改する変異である。いくつかの態様において、意図される変異は、遺伝子の制御配列（例えば遺伝子プロモーターまたは遺伝子リプレッサー）を変改する変異である。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編集因子のいずれかは、1:1よりも多大である意図される変異対意図されない変異（例えば、意図される点変異：意図されない点変異）の比を創出することができる。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編集因子のいずれかは、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15

0:1、少なくとも200:1、少なくとも250:1、少なくとも500:1、もしくは少なくとも1000:1、またはより多くである意図される変異対意図されない変異（例えば、意図される点変異：意図されない点変異）の比を創出することができる。本明細書の「塩基編集因子効率」の項に記載されている塩基編集因子の特徴は、本明細書において提供される融合蛋白質、または融合蛋白質を用いる方法のいずれかに適用され得ることは了解されるべきである。

【０３７８】

核酸を編集するための方法

本開示のいくつかの側面は核酸を編集するための方法を提供する。いくつかの態様において、方法は、核酸の核酸塩基（例えば、二本鎖DNA配列の塩基対）を編集するための方法である。いくつかの態様において、方法は：a）核酸（例えば二本鎖DNA配列）の標的領域を塩基編集因子（例えば、シチジンデアミナーゼドメインに融合させられたCas9ドメイン）とガイド核酸（例えばgRNA）とを含む複合体に接触させるステップ（標的領域は標的核酸塩基対を含む）、b）前記標的領域の鎖分離を誘導するステップ、c）標的領域の1本の鎖の前記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ステップ、およびd）前記標的領域の1つの鎖だけを切るステップを含み、第1の核酸塩基の塩基に対して相補的な第3の核酸塩基が第2の核酸塩基に対して相補的な第4の核酸塩基によって置き換えられ；方法は20%よりも少ないインデル形成を核酸にもたらす。いくつかの態様において、ステップbは省かれることは了解されるべきである。いくつかの態様において、第1の核酸塩基はシトシンである。いくつかの態様において、第2の核酸塩基は脱アミノ化されたシトシンまたはウラシルである。いくつかの態様において、第3の核酸塩基はグアニンである。いくつかの態様において、第4の核酸塩基はアデニンである。いくつかの態様において、第1の核酸塩基はシトシンであり、第2の核酸塩基は脱アミノ化されたシトシンまたはウラシルであり、第3の核酸塩基はグアニンであり、第4の核酸塩基はアデニンである。いくつかの態様において、方法は、19%よりも少ない、18%、16%、14%、12%、10%、8%、6%、4%、2%、1%、0.5%、0.2%、または0.1%よりも少ないインデル形成をもたらす。いくつかの態様において、方法は、第4の核酸塩基に対して相補的である第5の核酸塩基によって第2の核酸塩基を置き換えることをさらに含み、それによって、意図される編集された塩基対を創出する（例えばC:G→T:A）。いくつかの態様において、第5の核酸塩基はチミンである。いくつかの態様においては、意図される塩基対の少なくとも5%が編集される。いくつかの態様においては、意図される塩基対の少なくとも10%、15%、20%、25%、30%、35%、40%、45%、または50%が編集される。

【０３７９】

いくつかの態様において、標的ヌクレオチド中の意図される生成物対意図されない生成物の比は少なくとも2:1、5:1、10:1、20:1、30:1、40:1、50:1、60:1、70:1、80:1、90:1、100:1、もしくは200:1、またはより多くである。いくつかの態様において、意図される点変異対インデル形成の比は1:1、10:1、50:1、100:1、500:1、もしくは1000:1よりも多大、またはより多くである。いくつかの態様において、切られる一本鎖（ニック鎖）はガイド核酸とハイブリダイズさせられる。いくつかの態様において、切られる一本鎖は、第1の核酸塩基を含む鎖に対向する。いくつかの態様において、塩基編集因子はCas9ドメインを含む。いくつかの態様において、第1の塩基はシトシンであり、第2の塩基はG、C、A、またはTではない。いくつかの態様において、第2の塩基はウラシルである。いくつかの態様において、第1の塩基はシトシンである。いくつかの態様において、第2の塩基はG、C、A、またはTではない。いくつかの態様において、第2の塩基はウラシルである。いくつかの態様において、塩基編集因子は編集された鎖の切断修復を阻害する。いくつかの態様において、塩基編集因子は編集されない鎖を保護または結合する。いくつかの態様において、塩基編集因子はUGI活性を含む。いくつかの態様において、塩基編集因子はニッカーゼ活性を含む。いくつかの態様において、意図される編集される塩基対はPAM部位の上流である。いくつかの態様において、意図される編集される塩基対は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ヌクレオチド上流である。いくつかの態様において、意図される編集される塩基対はPAM部位の下流である。いくつかの態様において、意図される編集される塩基対は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ヌクレオチド下流である。いくつかの態様において、方法は古典的（例えばNGG）PAM部位を要求しない。いくつかの態様において、核酸塩基編集因子はリンカーを含む。いくつかの態様において、リンカーは長さ1～25アミノ酸である。いくつかの態様において、リンカーは長さ5～20アミノ酸である。いくつかの態様において、リンカーは長さ10、11、12、13、14、15、16、17、18、19、または20アミノ酸である。いくつかの態様において、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。いくつかの態様において、標的ウィンドウは1～10ヌクレオチドを含む。いくつかの態様において、標的ウィンドウは長さ1～9、1～8、1～7、1～6、1～5、1～4、1～3、1～2、または1ヌクレオチドである。いくつかの態様において、標的ウィンドウは長さ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。いくつかの態様において、意図される編集される塩基対は標的ウィンドウ内である。いくつかの態様において、標的ウィンドウは意図される編集される塩基対を含む。いくつかの態様において、方法は本明細書において提供される塩基編集因子のいずれかを用いて行われる。いくつかの態様において、標的ウィンドウは脱アミノ化ウィンドウである。

【０３８０】

いくつかの態様において、本開示はヌクレオチドを編集するための方法を提供する。いくつかの態様において、本開示は二本鎖DNA配列の核酸塩基対を編集するための方法を提供する。いくつかの態様において、方法は、a）塩基編集因子とガイド核酸（例えばgRNA）とを含む複合体に二本鎖DNA配列の標的領域を接触させること（標的領域は標的核酸塩基対を含む）、b）前記標的領域の鎖分離を誘導すること、c）標的領域の一本鎖の前記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こと、d）前記標的領域の1つの鎖だけを切ること、を含み、第1の核酸塩基の塩基に対して相補的な第3の核酸塩基は、第2の核酸塩基に対して相補的な第4の核酸塩基によって置き換えられ、第2の核酸塩基は、第4の核酸塩基に対して相補的である第5の核酸塩基によって置き換えられ、それによって意図される編集された塩基対を創出し、意図される編集された塩基対を創出する効率は少なくとも5%である。いくつかの態様において、ステップbは省かれることは了解されるべきである。いくつかの態様においては、意図される塩基対の少なくとも5%が編集される。いくつかの態様においては、意図される塩基対の少なくとも10%、15%、20%、25%、30%、35%、40%、45%、または50%が編集される。いくつかの態様において、方法は、19%よりも少ない、18%、16%、14%、12%、10%、8%、6%、4%、2%、1%、0.5%、0.2%、または0.1%よりも少ないインデル形成を引き起こす。いくつかの態様において、標的ヌクレオチドにおける意図される生成物対意図されない生成物の比は、少なくとも2:1、5:1、10:1、20:1、30:1、40:1、50:1、60:1、70:1、80:1、90:1、100:1、もしくは200:1、またはより多くである。いくつかの態様において、意図される点変異対インデル形成の比は、1:1、10:1、50:1、100:1、500:1、もしくは1000:1よりも多大、またはより多くである。いくつかの態様において、切られる一本鎖はガイド核酸に対してハイブリダイゼズさせられる。いくつかの態様において、切られる一本鎖は、第1の核酸塩基を含む鎖に対向する。いくつかの態様において、第1の塩基はシトシンである。いくつかの態様において、第2の核酸塩基はG、C、A、またはTではない。いくつかの態様において、第2の塩基はウラシルである。いくつかの態様において、塩基編集因子は編集される鎖の塩基除去修復を阻害する。いくつかの態様において、塩基編集因子は編集されない鎖を保護するかまたは結合する。いくつかの態様において、核酸塩基編集因子はUGI活性を含む。いくつかの態様において、核酸塩基編集因子はニッカーゼ活性を含む。いくつかの態様において、意図される編集される塩基対はPAM部位の上流である。いくつかの態様において、意図される編集される塩基対は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ヌクレオチド上流である。いく

つかの態様において、意図される編集される塩基対はPAM部位の下流である。いくつかの態様において、意図される編集される塩基対は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ヌクレオチド下流である。いくつかの態様において、方法は古典的（例えばNGG）PAM部位を要求しない。いくつかの態様において、核酸塩基編集因子はリンカーを含む。いくつかの態様において、リンカーは長さ1～25アミノ酸である。いくつかの態様において、リンカーは長さ5～20アミノ酸である。いくつかの態様において、リンカーは長さ10、11、12、13、14、15、16、17、18、19、または20アミノ酸である。いくつかの態様において、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。いくつかの態様において、標的ウィンドウは1～10ヌクレオチドを含む。いくつかの態様において、標的ウィンドウは長さ1～9、1～8、1～7、1～6、1～5、1～4、1～3、1～2、または1ヌクレオチドである。いくつかの態様において、標的ウィンドウは長さ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。いくつかの態様において、意図される編集される塩基対は標的ウィンドウ内に生起する。いくつかの態様において、標的ウィンドウは意図される編集される塩基対を含む。いくつかの態様において、核酸塩基編集因子は本明細書において提供される塩基編集因子のいずれか1つである。

【０３８１】

医薬組成物

いくつかの態様において、本明細書に記載される融合蛋白質、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おいて提供される融合蛋白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おいて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との複合体を形成するRNAによってガイドされるヌクレアーゼ（例えばCas9）を含むリボヌクレオ蛋白質複合体とカチオン性脂質とを含む。いくつかの態様において、医薬組成物はgRNA、核酸によってプログラム可能なDNA結合蛋白質、カチオン性脂質、および薬学的に許容し得る添加剤を含む。医薬組成物は1つ以上の追加の治療上活性な物質を任意に含み得る。

【０３８２】

いくつかの態様において、本明細書において提供される組成物は、対象内において標的化されたゲノム改変を成し遂げるために、対象、例えばヒト対象に投与される。いくつかの態様においては、細胞が対象から得られ、本明細書において提供される医薬組成物のいずれかに接触させられる。いくつかの態様においては、対象から取り出されエクスビボで医薬組成物に接触させられた細胞は、任意に所望のゲノム改変が細胞に成し遂げられたか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公知であり、例えばU.S.Pat.No.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；および7,163,824に記載されている。これらの全ての開示はそれらの全体が参照によって本明細書に組み込まれる。本明細書において提供される医薬組成物の記載はヒトへの投与にとって好適である医薬組成物に主として向けられているが、かかる組成物は全ての種類の動物または生物への投与にとって一般的に好適であることは当業者によって理解されるであろう。組成物を種々の動物への投与にとって好適にするための、ヒトへの投与にとって好適な医薬組成物の改変は良く理解されており、普通の技能の獣医薬理学者はせいぜいただ普通の実験手順によってかかる改変を設計し得、および／または行い得る。医薬組成物の投与が企図される対象は、ヒトおよび／もしくは他の霊長類の動物；哺乳動物、飼育動物、ペット、および商業的に該当する哺乳動物、例えばウシ(cattle)、ブタ、ウマ、ヒツジ、ネコ、イヌ、マウス、および／もしくはラット；ならびに／またはニワトリ、カモ、ガチョウ、および／もしくは七面鳥などの商業的に該当するトリを包含するトリを包含するが、これらに限定されない。

【０３８３】

本明細書に記載される医薬組成物の製剤は、薬理学の分野において公知のまたは以後に開発されるいずれかの方法によって調製され得る。一般的に、かかる調製方法は、活性成分（単数または複数）を添加剤および／または1つ以上の他の補助的成分と結びつけ、それから、必要なおよび／または望ましい場合には、生成物を所望のシングルまたはマルチドーズユニットへと成形および／またはパッケージングするステップを包含する。

【０３８４】

加えて、医薬製剤は薬学的に許容し得る添加剤を含み得、本明細書において用いられるこれは、所望の特定の剤形に適するいずれかおよび全ての溶媒、分散媒、希釈剤、または他の液体基剤、分散もしくは懸濁助剤、表面活性剤、等張化剤、増粘もしくは乳化剤、保存料、固体結合剤、潤滑剤、および同類を包含する。Remington's The Science and Practice of Pharmacy, 21st Edition, A. R. Gennaro（Lippincott, Williams & Wilkins, Baltimore, MD, 2006；その全体が参照によって本明細書に組み込まれる）は、医薬組成物を製剤することに用いられる種々の添加剤、およびそれらの調製のための公知技術を開示している。ヌクレアーゼを含む医薬組成物を生成するための追加の好適な方法、試薬、添加剤、および溶媒については、その全体が参照によって本明細書に組み込まれるPCT出願PCT/US2010/055131（2010年Nov.2出願の公開番号WO2011053982A8）をもまた見よ。例えば、いずれかの望ましくない生物学的効果を生成することまたはさもなければ有害な様式で医薬組成物のいずれかの他のコンポーネント（単数または複数）と相互作用することによって、いずれかの従来の添加剤媒体が物質またはその誘導体と不適合である限りを除いて、その使用は本開示の範囲内であることが企図される。

【０３８５】

いくつかの態様において、本発明に従う組成物は、次の1つ以上を包含するがこれらに限定されない種々の疾患、異常、および／または状態のいずれかの処置のために用いられ得る：自己免疫異常（例えば糖尿病、狼瘡、多発性硬化症、乾癬、関節リウマチ）；炎症性の異常（例えば関節炎、骨盤内炎症性疾患）；感染性疾患（例えばウイルス感染（例えば、HIV、HCV、RSV）、細菌感染、真菌感染、敗血症）；神経学的異常（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）；心血管系異常（例えば動脈硬化、高コレステロール血症、血栓症、血液凝固異常、血管新生異常、例えば黄斑変性）；増殖性の異常（例えば癌、良性新生物）；呼吸器異常（例えば慢性閉塞性肺疾患）；消化器異常（例えば炎症性腸疾患、潰瘍）；筋骨格系異常（例えば線維筋痛症、関節炎）；内分泌、代謝、および栄養異常（例えば糖尿病、骨粗鬆症）；泌尿器異常（例えば腎臓病）；精神異常（例えば抑うつ、統合失調症）；皮膚異常（例えば創傷、湿疹）；血液およびリンパ異常（例えば貧血、血友病）；など。

【０３８６】

キット、ベクター、細胞

本開示のいくつかの側面は、核酸構築物を含むキットを提供し、（a）本明細書において提供されるCas9蛋白質またはCas9融合蛋白質をコードするヌクレオチド配列と；（b）（a）の配列の発現を駆動する異種プロモーターとを含む。いくつかの態様において、キットはガイドRNAバックボーンをコードする発現構築物をさらに含み、構築物は、ガイドRNAバックボーンへの標的配列と同一または相補的な核酸配列のクローニングを許すように位置したクローニング部位を含む。

【０３８７】

本開示のいくつかの側面は、本明細書において提供される融合蛋白質のCas9蛋白質をコードするポリヌクレオチドを提供する。本開示のいくつかの側面はかかるポリヌクレオチドを含むベクターを提供する。いくつかの態様において、ベクターはポリヌクレオチドの発現を駆動する異種プロモーターを含む。

【０３８８】

本開示のいくつかの側面は、本明細書において提供されるCas9蛋白質、融合蛋白質、融合蛋白質をコードする核酸分子、Cas9蛋白質とgRNAとを含む複合体、および／またはベクターを含む、細胞を提供する。

【０３８９】
上のレポーターシステムの例示的な態様の記載は例示目的のみのために提供されており、限定しようとすることを意味されていない。追加のレポーターシステム、例えば、上に詳細に記載されている例示的なシステムの変形もまた本開示によって包摂される。
【０３９０】
例
例１：Cas9デアミナーゼ融合蛋白質
いくつものCas9:デアミナーゼ融合蛋白質を創出し、創出された融合体のデアミナーゼ活性を特徴付けた。次のデアミナーゼを試験した：
ヒトAID（hAID）：
MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVEL
LFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPYLSLRIFTARLYFCEDR
KAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRR
ILLPLYEVDDLRDAFRTLGLLD（配列番号49）
ヒトAID-DC（hAID-DC。7倍増大した活性を有するhAIDの短縮版）：
MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVEL
LFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDR
KAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRR
ILL（配列番号50）
ラットAPOBEC1（rAPOBEC1）：
MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLE
LYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK（配列番号76）
ヒトAPOBEC1（hAPOBEC1）
MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNTTN
HVEVNFIKKFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVARLF
WHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMM
LYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPPHILLATGLIH PSVAWR
（配列番号74）
Petromyzon marinus（ヤツメウナギ）CDA1（pmCDA1）：
MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQS
GTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTL
KIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWL
EKTLKRAEKRRSELSIMIQVKILHTTKSPAV（配列番号81）
ヒトAPOBEC3G（hAPOBEC3G）：
MELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLT
IFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPW
NNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWV
LLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCA
QEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVD
HQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN（配列番号60）
【０３９１】
ssDNAに対するデアミナーゼ活性．脱アミノ化のためのUSER（ウラシル特異的切り出し試薬）酵素に基づくアッセイを採用して、一本鎖DNA（ssDNA）基質に対する種々のデアミナーゼの活性を試験した。USER酵素はNew England Biolabsから得た。ssDNA基質には標的シトシン残基を異なる位置に提供した。ssDNAシトシン標的残基の脱アミノ化は標的シトシンからウラシルへの変換をもたらす。USER酵素はウラシル塩基を切り出し、ssDNAバックボーンをその位置において切断し、ssDNA基質をDNAの2つのより短い断片へと切る。いくつかのアッセイでは、ssDNA基質は、1つの末端を色素によって、例えば5

'Cy3標識（下のスキーム中の＊）によって標識される。鎖の脱アミノ化、切り出し、および切断によって、基質は電気泳動に付され得、基質およびそれから遊離したいずれかの断片は標識を検出することによって視覚化され得る。Cy5がイメージングされるところでは、標識を有する断片のみがイメージングによって可視であろう。

【０３９２】

1つのUSER酵素アッセイでは、試験された種々のデアミナーゼの標的配列にマッチするssDNA基質を用いた。試験されたデアミナーゼをコードする発現カセットを、先にラボで用いたCMVバックボーンプラスミド（Addgeneプラスミド52970）に挿入した。デアミナーゼ蛋白質を製造者の推奨に従ってTNT Quick Coupled転写/翻訳システム（Promega）を用いて発現した。90minのインキュベーション後に、5mLの溶解物を5'Cy3標識ssDNA基質および1単位のUSER酵素（NEB）と3時間に渡ってインキュベーションした。DNAを10%TBE PAGEゲルによって分離し、DNAはCy色素イメージングを用いてイメージングした。USER酵素アッセイの模式図が図41に示されている。

【０３９３】

図1は、ssDNA基質、例えばDoench 1、Doench 2、G7'、およびVEGF標的2に対する試験されたデアミナーゼのデアミナーゼ活性を示している。rAPOBEC1酵素は、古典的NGG PAMを有する一本鎖DNA基質では実質的な量の脱アミノ化を見せたが、陰性対照非古典的NNN PAMでは見せなかった。APOBECファミリーデアミナーゼを有するCas9融合蛋白質を創出した。次の融合アーキテクチャを構築し、ssDNAに対して試験した：

【化６４】

rAPOBEC1-GGS-*dCas9* 一次配列

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARL
YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVR

【化６５】

LYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKGGS*DKKY*
SIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK
NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQI
HLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI
TPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYV
TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFN
ASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDK
VMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLT
FKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVI
EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVV
KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ
ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNA
VVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE
SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL
DATLIHQSITGLYETRIDLSQLGGD

（配列番号１４９）

【化６６】

rAPOBEC1-$(GGS)_3$-*dCas9* －次配列

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARL
YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVR
LYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKGGSGGSG
GS*MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSG*
ETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGD
LNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE
KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLF
LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKE
IFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN
GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMT
RKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI
HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRH
KPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY
LYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNV
PSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIT
KHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAH
DAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIM

【化６７】

NFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF
SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

（配列番号１５０）

【化６８】

dCas9-GGS-rAPOBEC1

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP
DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN
GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA
KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRK
SEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTK
VKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE
DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHL
FDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHD
DSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPE
NIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY
LQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSE
EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAY
LNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFF
KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKE
LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQ
KGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKR
VILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS
*TKEVLDATLIHQSITGLYETRIDLSQLGGD*GGSMSSETGPVAVDPTLRRRIEPHEFEVF
FDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSIT
WFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMT
EQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQL
TFFTIALQSCHYQRLPPHILWATGLK

（配列番号１５１）

【化６９】

dCas9-GGS_3-rAPOBEC1

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP
DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN
GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA
KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRK
SEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTK
VKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE

【化７０】

DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHL
FDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHD
DSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPE
NIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY
LQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSE
EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAY
LNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFF
KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKE
LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQ
KGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKR
VILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS
TKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGGSGGSMSSETGPVAVDPTLRRRIEP
HEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCP
NTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSG
VTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNIL
RRKQPQLTFFTIALQSCHYQRLPPHILWATGLK

（配列番号１５２）

【化７１】

rAPOBEC1-XTEN-*dCas9* 一次配列

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARL
YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVR
LYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPG
TSESATPES*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL*
FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEED
KKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHF
LIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIA
QLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGD
QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL
PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK
QRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNS
RFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR
NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNR
GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL
VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN
NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIV
KKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGK
SKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKR
MLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFD
TTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

（配列番号153）

【０３９４】

図2は、N末端デアミナーゼ融合体が一本鎖DNA基質に対して有意な活性を示したことを示す。この理由から、N末端アーキテクチャのみをさらなる実験に選んだ。

【０３９５】

図3は、デアミナーゼ-dCas9:sgRNA複合体による二本鎖DNA基質を例示する。いくつもの二本鎖デアミナーゼ基質配列を創出した。配列は下に提供される。図3に従う構造はこれらの配列中に同定される（36bp：下線付き、sgRNA標的配列：太字；PAM：枠囲み；21b：斜体化）。全ての基質は5'Cy3標識によって標識した：

【化７２】

2:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGTCCCGCGGATTTATTTATTTA
ATGG*ATGACCTCTGGATCCATGGAC*-3’ (配列番号 85)
3:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCTTCCGCGGATTTATTTATTT
ATGG*ATGACCTCTGGATCCATGGAC*-3’ (配列番号 86)
4:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCTTCCGCGGATTTATTTATT
ATGG*ATGACCTCTGGATCCATGGAC*-3’ (配列番号 87)
5:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCATTCCGCGGATTTATTTAT
TTGG*ATGACCTCTGGATCCATGGAC*-3’ (配列番号 88)
6:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCTATTCCGCGGATTTATTTA
TTGG*ATGACCTCTGGATCCATGGAC*-3’ (配列番号 89)
7:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCTTATTCCGCGGATTTATTT
ATGG*ATGACCTCTGGATCCATGGAC*-3’ (配列番号 90)
8:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCATTATTCCGCGGATTTATT
TTGG*ATGACCTCTGGATCCATGGAC*-3’ (配列番号 91)
9:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCTATTATTCCGCGGATTTAT
TTGG*ATGACCTCTGGATCCATGGAC*-3’ (配列番号 92)
10:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCATTATATTCCGCGGATTT
ATTGG*ATGACCTCTGGATCCATGGAC*-3’ (配列番号 93)
11:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCTATTATATTCCGCGGATT
TATGG*ATGACCTCTGGATCCATGGAC*-3’ (配列番号 94)
12:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCTTATTATATTCCGCGGAT
TTTGG*ATGACCTCTGGATCCATGGAC*-3’ (配列番号 95)
13:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCATTATTATATTCCGCGGA
TTTGG*ATGACCTCTGGATCCATGGAC*-3’ (配列番号 96)
14:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCTATTATTATATTCCGCGG
ATTGG*ATGACCTCTGGATCCATGGAC*-3’ (配列番号 97)
15:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCATTATTATTATTACCGCG
GATGG*ATGACCTCTGGATCCATGGAC*-3’ (配列番号 98)
18:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCATTATTATTATTATTACCG
CTGG*ATGACCTCTGGATCCATGGAC*-3’ (配列番号 99)
“-“ :
GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGTAATATTAATTTATTTATTTAAT
GG*ATGACCTCTGGATCCATGGAC*-3’ (配列番号 100)
8U:GTAGGTAGTTAGGATGAATGGAAGGTTGGTGTAGATTATTATCUGCGGATTTAT
TGG*ATGACCTCTGGATCCATGGACAT*-3’ (配列番号 101)

*「8U」を除く全ての基質において、図3の上の鎖はここで指定される配列の相補物である。「8U」のケースでは、Uに対向して「G」がある。

【０３９６】

図4は、二本鎖DNA脱アミノ化アッセイの結果を示す。融合体を発現し、Ni-NTAおよびセファロースクロマトグラフィー両方によって、N末端His6タグによって精製した。dsDNA基質に対する脱アミノ化を査定するために、先のスライドに示されている種々のdsDNA基質を1:8のdsDNA：融合蛋白質比でインキュベーションし、37℃で2時間に渡ってインキュベーションした。ひとたび融合体のdCas9部分がDNAに結合すると、それはDNAへのUSER酵素のアクセスをブロックする。よって、融合蛋白質をインキュベーション後に変性させ、dsDNAをスピンカラムによって精製し、次に、USER酵素との45minのインキュベーション、およびもたらされたDNA基質および基質断片の10%TBE-尿素ゲルによる分離をした。

【０３９７】

図5は、Cas9融合体が二本鎖DNA標的配列の位置3～11を標的化し得ることを実証する（図3の模式図に従って付番される）。上側のゲル：1μM rAPOBEC1-GGS-dCas9、125nM dsDNA、1eq sgRNA。真中のゲル：1μM rAPOBEC1-$(GGS)_3$-dCas9、125nM d

sDNA、1eq sgRNA。下側のゲル：1.85μM rAPOBEC1-XTEN-dCas9、125nM dsDNA、1eq sgRNA。これらのゲルからのデータに基づくと、位置3～11（図3の付番に従う）はデアミナーゼの活性に十分に暴露されて、試験された融合蛋白質によって標的化される。他の位置へのデアミナーゼのアクセスは最も蓋然的にはdCas9蛋白質によってブロックされる。

【０３９８】

データは、3アミノ酸のみのリンカー（GGS）が、デアミナーゼがDNAの一本鎖部分にアクセスすることを許すためには最適ではないことをさらに指示する。9アミノ酸リンカー[$(GGS)_3$]（配列番号610）およびより構造化された16アミノ酸リンカー（XTEN）はより効率的な脱アミノ化を許す。

【０３９９】

図6は、正しいガイドRNA、例えば正しいsgRNAがデアミナーゼ活性に要求されることを実証する。ゲルは、デアミナーゼをdCas9に融合させると、デアミナーゼ酵素は配列特異的になり（例えば、eGFP sgRNAとの融合体を用いることは脱アミノ化をもたらさない）、dsDNAを脱アミノ化する力能をもまたデアミナーゼに授けることを示す。デアミナーゼ酵素の本来の基質はssDNAであり、sgRNAが追加されないときには、脱アミノ化は生起しなかった。これは、APOBECデアミナーゼがそれ自体ではdsDNAを脱アミノ化しないという報告された知識と整合する。データは、Cas9が短いウィンドウ内の二本鎖DNAらせんを開裂させ、一本鎖DNAを露出させ、これはそれからシチジン脱アミノ化のためにAPOBECデアミナーゼにとってアクセス可能であることを指示する。用いられたsgRNA配列は下に提供される。配列（36bp：下線付き、sgRNA標的配列：太字；PAM：枠囲み；21bp：斜体化）

DNA配列8：

【化７３】

5’-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCC**ATTATTCCGCGGATTTATT**
TTGG*ATGACCTCTGGATCCATGGAC*-3’

（配列番号102）

正しいsgRNA配列（部分的な3'配列）：

【化７４】

5’-**AUUAUUCCGCGGAUUUAUUU**GUUUUAGAGCUAG...-3’

（配列番号103）

eGFP sgRNA配列（部分的な3'配列）：

【化７５】

5’-**CGUAGGCCAGGGUGGUCACG**GUUUUAGAGCUAG...-3’

（配列番号104）

【０４００】

例2：DNA標的配列の脱アミノ化

例示的な脱アミノ化標的。

本明細書に記載されるdCas9:デアミナーゼ融合蛋白質は、in vitroもしくはエクスビボの細胞にまたはin vivoの対象に送達され得、標的ヌクレオチドがPAMにとって位置3～1

1にあるときにC→TまたはG→A転換を成し遂げるために用いられ得る。例示的な脱アミノ化標的は限定なしに次を包含する：CCR5短縮：CCR5のQ93、Q102、Q186、R225、W86、またはQ261をコードするコドンのいずれかは脱アミノ化されて終止コドンを創出し得、これはCCR5の非機能的な短縮をもたらし、HIV処置への用途を有する。APOE4変異：C11RおよびC57R変異体APOE4蛋白質をコードする変異体コドンは脱アミノ化されて野生型アミノ酸に復帰し得、アルツハイマーの処置への用途を有する。eGFP短縮：Q158、Q184、Q185をコードするコドンのいずれかは脱アミノ化されて終止コドンを創出し得、またはM1をコードするコドンは脱アミノ化されてIをコードし得、これらの全てはeGFP蛍光の喪失をもたらし、レポーターシステムへの用途を有する。eGFP回復：実質的な蛍光を見せないT65AまたはY66C変異体GFPをコードする変異体コドンは、脱アミノ化されて野生型アミノ酸を回復させ、蛍光を授け得る。PIK3CA変異：K111E変異体PIK3CAをコードする変異体コドンは脱アミノ化されて野生型アミノ酸残基を回復させ得、癌への用途を有する。CTNNB1変異：T41A変異体CTNNB1をコードする変異体コドンは脱アミノ化されて野生型アミノ酸残基を回復させ得、癌への用途を有する。HRAS変異：Q61R変異体HRASをコードする変異体コドンは脱アミノ化されて野生型アミノ酸残基を回復させ得、癌への用途を有する。P53変異：Y163C、Y236C、またはN239D変異体p53をコードする変異体コドンのいずれかは脱アミノ化されて野生型アミノ酸配列をコードし得、癌への用途を有する。二本鎖DNA中のこれらの標的配列を脱アミノ化することのフィージビリティは図7および8において実証される。図7はデアミナーゼ-dCas9:sgRNA複合体によるin vivo標的配列の標的DNA結合のメカニズムを例示する。

【０４０１】

図8は例示的な疾患関連標的配列の首尾良い脱アミノ化を示す。上側のゲル：CCR5 Q93：pos.10にコード鎖標的（位置2、5、6、8、9に見込まれるオフターゲット）；CCR5 Q102：pos.9にコード鎖標的（位置1、12、14に見込まれるオフターゲット）；CCR5 Q186：pos.9にコード鎖標的（位置1、5、15に見込まれるオフターゲット）；CCR5 R225：pos.6にコード鎖標的（見込まれるオフターゲットなし）；eGFP Q158：pos.5にコード鎖標的（位置1、13、16に見込まれるオフターゲット）；eGFP Q184/185：pos.4および7にコード鎖標的（位置3、12、14、15、16、17、18に見込まれるオフターゲット）；eGFP M1：pos.12に鋳型鎖標的（位置2、3、7、9、11に見込まれるオフターゲット）（位置7および9を小さい程度に標的化する）；eGFP T65A：pos.7に鋳型鎖標的（位置1、8、17に見込まれるオフターゲット）；PIK3CAK111E：pos.2に鋳型鎖標的（位置5、8、10、16、17に見込まれるオフターゲット）；PIK3CAK111E：pos.13に鋳型鎖標的（位置11、16、19に見込まれるオフターゲット）X。下側のゲル：CCR5 W86：pos.2および3に鋳型鎖標的（位置1、13に見込まれるオフターゲット）X；APOE4 C11R：pos.11にコード鎖標的（位置7、13、16、17に見込まれるオフターゲット）；APOE4 C57R：pos.5にコード鎖標的）（位置7、8、12に見込まれるオフターゲット）；eGFP Y66C：pos.11に鋳型鎖標的（位置1、4、6、8、9、16に見込まれるオフターゲット）；eGFP Y66C：pos.3に鋳型鎖標的（位置1、8、17に見込まれるオフターゲット）；CCR5 Q261：pos.10にコード鎖標的（位置3、5、6、9、18に見込まれるオフターゲット）；CTBNN1 T41A：pos.7に鋳型鎖標的（位置1、13、15、16に見込まれるオフターゲット）X；HRAS Q61R：pos.6に鋳型鎖標的（位置1、2、4、5、9、10、13に見込まれるオフターゲット）；p53 Y163C：pos.6に鋳型鎖標的（位置2、13、14に見込まれるオフターゲット）；p53 Y236C：pos.8に鋳型鎖標的（位置2、4に見込まれるオフターゲット）；p53 N239D：pos.4に鋳型鎖標的（位置6、8に見込まれるオフターゲット）。疾患標的の例示的なDNA配列は下で提供されている（PAM（5'-NGG-3'）および標的位置は枠囲みされている）：

【化７６】

CCR5 **Q93**: 5’-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAACTAT**GCTGCCGCC**
CAGTGGGACTTTGG*AAATACAATGTGTCAACTCTT*-3’ (配列番号 105)
CCR5 **Q102**: 5’-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAAAATA**CAATGTGT**
CAACTCTTGACAGGG*CTCTATTTTATAGGCTTCTTC*-3’ (配列番号 106)
CCR5 **Q186**: 5’-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTATTTTC**CATACAGT**
CAGTATCAATTCTGG*AAGAATTTCCAGACATTAAAG*-3’ (配列番号 107)
CCR5 **R225**: 5’-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGCTTC**GGTGTCGA**
AATGAGAAGAAGAGG*CACAGGGCTGTGAGGCTTATC*-3’ (配列番号 108)

【化７７】

CCR5 **W86**: 5’-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGTGAG**C**[**C**]**AGAAGG**
GGACAGTAAGA[AGG]*AAAAACAGGTCAGAGATGGCC*-3’ (配列番号 109)
CCR5 **Q261**: 5’-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTATCCTG**AACACCTT**
C[**C**]**AGGAATTCTT**[TGG]*CCTGAATAATTGCAGTAGCTC*-3’ (配列番号 110)
APOE4 **C11R**: 5’-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGACAT**GGAGGAC**
GTG[**C**]**GCGGCCGCC**[TGG]*TGCAGTACCGCGGCGAGGTGC*-3’ (配列番号 111)
APOE4 **C57R**: 5’-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTACTGCA**GAAG**[**C**]**GC**
CTGGCAGTGTACC[AGG]*CCGGGGCCCGCGAGGGCGCCG*-3’ (配列番号 112)
eGFP **Q158**: 5’-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGCCGA**CAAG**[**C**]**AGA**
AGAACGGCATCA[AGG]*TGAACTTCAAGATCCGCCACA*-3’ (配列番号 113)
eGFP **Q184/185**: 5’-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTAACCAC**TAC**[**C**]
AG[**C**]**AGAACACCCCCAT**[CGG]*CGACGGCCCCGTGCTGCTGCC*-3’ (配列番号 114)
eGFP **M1**: 5’-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTACCTCG**CCCTTGCTCA**
C[**C**]**ATCTCGAGT**[CGG]*CCGCCAGTGTGATGGATATCT*-3’ (配列番号 115)
eGFP **T65A**: 5’-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTACACGC**GTAGG**[**C**]**CA**
GGGTGGTCACG[AGG]*GTGGGCCAGGGCACGGGCAGC*-3’ (配列番号 116)
eGFP **Y66C**: 5’-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAAAGCA**CTGCACTC**
CG[**C**]**AGGTCAGGG**[TGG]*TCACGAGGGTTGGCCAGGGCA*-3’ (配列番号 117)
eGFP **Y66C**: 5’-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTACACTC**CG**[**C**]**AGGTC**
AGGGTGGTCACG[AGG]*GTTGGCCAGGGCACGGGCAGG*-3’ (配列番号 118)
PIK3CA **K111E**: 5’-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGGATC**T**[**C**]**TTC**
TTCACGGTTGCCTAC[TGG]*TTCAATTACTTTTAAAAATGG*-3’ (配列番号 119)
PIK3CA **K111E**: 5’-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTATTCTC**GATTG**
AGGATCT[**C**]**TTCTTCA**[CGG]*TTGCCTACTGGTTCAATTACT*-3’ (配列番号 120)
CTNNB1 **T41A**: 5’-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAAGGAG**CTGTGG**
[**C**]**AGTGGCACCAGAA**[TGG]*ATTCCAGAGTCCAGGTAAGAC*-3’ (配列番号 121)
HRAS **Q61R**: 5’-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGTACT**CCTCC**[**C**]**GG**
CCGGCGGTATCC[AGG]*ATGTCCAACAGGCACGTCTCC*-3’ (配列番号 122)
p53 **Y163C**: 5’-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTATGACT**GCTTG**[**C**]**AG**
ATGGCCATGGCG[CGG]*ACGCGGGTGCCGGGCGGGGGT*-3’ (配列番号 123)
p53 **Y236C**: 5’-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTACTGTT**ACACATG**[**C**]
AGTTGTAGTGGA[TGG]*TGGTACAGTCAGAGCCAACCT*-3’ (配列番号 124)
p53 **N239D**: 5’-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGGAAC**TGT**[**C**]**ACAC**
ATGTAGTTGTAG[TGG]*ATGGTGGTACAGTCAGAGCCA*-3’ (配列番号 125)

【０４０２】

例3：ウラシルグリコシラーゼ阻害因子融合は脱アミノ化効率を改善する

dCas9:デアミナーゼ融合蛋白質による哺乳類細胞における直接的なプログラム可能な核酸塩基編集効率は、ウラシルグリコシラーゼ阻害因子（UGI）をdCas9:デアミナーゼ融合蛋白質に融合させることによって有意に改善され得る。

【０４０３】

図9はrAPOBEC1-XTEN-dCas9を用いるヒトHEK293細胞のin vitro C→T編集効率を示している：

【化７８】

rAPOBEC1-**XTEN**-dCas9-*NLS* －次配列

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLY
HHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRL
YVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK**SGSETPGTS**
ESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF
DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIE
GDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEK
YKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWM
TRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNE
LTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS
GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI
HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRH
KPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL
YYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNV
PSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI
TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAH
DAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI
MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKL
KSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDR
KRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD SGGS*PKKKRKV*

【化７９】

EMX1: 5’- GAGT**C**$_5$**C**$_6$GAG**C**$_{10}$AGAAGAAGAAGGG -3’ (配列番号 127)

FANCF: 5’- GGAAT**C**$_6$**C**$_7$**C**$_8$TT**C**$_{11}$TGCAGCACCTGG -3’ (配列番号 128)

HEK293 部位 2: 5’- GAA**C**$_4$A**C**$_6$AAAG**C**$_{11}$ATAGACTGCGGG -3’ (配列番号 129)

HEK293 部位 3: 5’- GG**C**$_3$**C**$_4$**C**$_5$AGA**C**$_9$TGAGCACGTGATGG -3’ (配列番号 130)

【化８０】

HEK293 部位 4: 5’- GG**C**$_3$A**C**$_5$TG**C**$_8$GG**C**$_{11}$TGGAGGTGGGGG -3’ (配列番号 735)

RNF2: 5’- GT**C**$_3$AT**C**$_6$TTAGTCATTACCTGAGG -3’ (配列番号 132)

＊PAMは枠囲みされており、標的ウィンドウ（位置3～11）内のC残基は付番および太字化されている。

【０４０４】

図10は、UGIドメインがrAPOBEC1:dCas9融合蛋白質に融合させられるときに、HEK293T細胞における同じプロトスペーサー配列に対するC→T編集効率が多大に向上することを実証している。

【化８１】

rAPOBEC1-**XTEN**-dCas9-**UGI**-*NLS* 一次配列

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLY
HHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRL
YVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK**SGSETPGTS**
ESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF
DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIE
GDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEK
YKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWM
TRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNE
LTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS
GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI
HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRH
KPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL
YYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNV
PSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI
TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAH
DAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI
MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKL
KSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDR
KRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGS**TNLSDIIEKETGKQLVIQESI**
LMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGE
NKIKMLSGGS*PKKKRKV*

（配列番号133）

【０４０５】

図9および10のパーセンテージは、標的配列の両方の鎖をシーケンシングすることから示されている。鎖の1つのみが脱アミノ化の基質であるので、このアッセイにおける最大の可能な脱アミノ化値は50%である。従って、脱アミノ化効率は表に示されているパーセンテージの倍である。例えば、50%の値は二本鎖標的配列の100%の脱アミノ化に関する。ウラシルグリコシラーゼ阻害因子（UGI）をdCas9:デアミナーゼ融合蛋白質に融合させたときには（例えばrAPOBEC1-XTEN-dCas9-[UGI]-NLS）、細胞の編集効率の有意な増大が観察された。この結果は、哺乳動物の細胞では、U:G塩基対のウラシル塩基を切り出すDNA修復機構がDNA編集の律速プロセスであることを指示している。UGIをdCas9:デアミナーゼ融合蛋白質にテザリングすることは編集収率を多大に増大させる。

【０４０６】

UGIなしでは、ヒト細胞の典型的な編集効率は～2～14%の収率範囲であった（図9および図10、「XTEN」エントリー）。UGIでは（図10「UGI」エントリー）、編集は～6～40%の範囲で観察された。それゆえに、UGI融合体を用いることは、点変異を修正すること

に加えて過剰なインデルをもまた作出するHDRによって点変異を修正する現行の代替的な方法よりも効率的である。cas9:デアミナーゼ:UGI融合体による処置からもたらされるインデルは観察されなかった。

【０４０７】

例４：二本鎖DNA切断なしのゲノムDNA中の標的ヌクレオチドの直接的なプログラム可能な変換

現行のゲノム編集テクノロジーは遺伝子修正の第1ステップとして当該標的座位に二本鎖DNA切断を導入する[39,40]。ほとんどの遺伝子疾患は単一の核酸塩基から異なる核酸塩基への変異から生ずるが、かかる変化を復帰させるための現行のアプローチは非常に非効率的であり、典型的には、二本鎖DNA切断に対する細胞応答の帰結として標的座位において豊富なランダムな挿入および欠失（インデル）を誘導する[39,40]。本明細書においては、二本鎖DNAバックボーン切断を要求することなしに、プログラム可能な様式で1つの標的核酸塩基から別のものへの直接的変換を可能にするゲノム編集の新たな戦略、核酸塩基編集の開発を報告する。CRISPR/Cas9およびシチジンデアミナーゼ酵素APOBEC1の融合体を操作した。これらは、ガイドRNAによってプログラムされる能力を保持し、二本鎖DNA切断を誘導せず、シチジンからウラシルへの直接的変換を媒介し、それによって、DNA複製、DNA修復、または鋳型鎖が標的化される場合には転写の後に、C→T（またはG→A）置換を成し遂げる。もたらされた「核酸塩基編集因子」はおよそ5ヌクレオチドのウィンドウ内のシチジンを変換し、ヒト疾患に該当する種々の点変異をin vitroで効率的に修正し得る。4つの形質転換ヒトおよびマウスの細胞株において、それぞれ、ウラシルグリコシラーゼ阻害因子（UGI）を融合させる第2および編集されない鎖を標的化するCas9ニッカーゼを用いる第3世代核酸塩基編集因子は、核酸塩基編集に対する細胞性のDNA修復応答を克服し得、ヒト細胞において合計の細胞DNAの37%までまたは（～15～75%）の永続的な修正をもたらし、最小限の（典型的には≦1%）インデル形成を有する。対照的に、同じ標的における古典的Cas9によって媒介されるHDRは平均で0.7%の修正を生み、4%のインデル形成を有した。核酸塩基編集因子を用いて、ヒト乳癌およびリンパ腫細胞において2つの発癌性p53変異を野生型アレルに復帰させ、マウスアストロサイトにおいてApoE4のアルツハイマー病関連Argコドンを非疾患関連Cysコドンに変換した。塩基編集は点変異のゲノム編集の範囲および効率を拡大する。

【０４０８】

clustered regularly interspaced short palindromic repeat（CRISPR）システムは、種々の生物および細胞株におけるゲノム工学を媒介するために転用されてきた原核生物適応免疫システムである[41]。CRISPR/Cas9蛋白質-RNA複合体はガイドRNAとの塩基対形成によって標的DNA配列に局在し、本来的には、ガイドRNAによって指定される座位にDNA二本鎖切断（DSB）を作出する。DSBに応答して、内在性のDNA修復プロセスは、主に、非相同末端結合（NHEJ）によってDNA切断の部位にランダムな挿入または欠失（インデル）をもたらす。相同なDNA鋳型の存在下において、切断部位の周辺のDNAは相同組み換え修復（HDR）によって置き換えられ得る。疾患関連遺伝子の単純な破壊が十分であるときには（例えばいくつかの機能獲得型疾患を処置するためには）、標的化されたDNA切断、次にインデル形成が有効であり得る。しかしながら、ほとんどの公知の遺伝子疾患では、疾患の裏にある原因に対処または研究するためには、遺伝子の確率的な破壊よりもむしろ標的座位の点変異の修正が必要とされる[68]。

【０４０９】

この必要を動機として、研究者はHDRの効率を増大させNHEJを抑圧するために鋭意の作業を投入してきた。例えば、NHEJ経路の必須酵素リガーゼIVの小分子阻害剤はHDR効率を増大させることが示されている[42,43]。しかしながら、この戦略は、典型的にはHDRを下方制御する有糸分裂後の細胞においては難題であり、その治療上の妥当性は非標的細胞のリガーゼIVを阻害する見込まれるリスクによって限定される。HDR効率は高度に細胞周期依存的であるので、向上したHDR効率は、化学的に同調した細胞へのCas9-ガイドRNA複合体の時間内送達によってもまた達成され得る[44]。しかしながら、細胞を同調するこ

とは高度に破壊的であるので、かかるアプローチは細胞培養による研究用途に限定される。これらの発展にもかかわらず、ほとんどのコンテキストにおいてHDRを用いて点変異を置き換えるための現行の戦略は、特に未改変の非分裂細胞においては非常に非効率的である（典型的には～0.1から5%）[42,43,45,46,75]。加えて、HDRは二本鎖切断の解消の間にNHEJと競合し、インデルは一般的に遺伝子置き換えよりも豊富な成果である。これらの観察は、二本鎖DNA切断を作出することに依拠しないゲノムDNAの特異的な改変をインストールするための代替的なアプローチを開発する必要を強調している。NHEJ経路の必須酵素リガーゼIVの小分子阻害剤はHDR効率を増大させることが示されている[42,43]。しかしながら、この戦略は、典型的にはHDRを下方制御する有糸分裂後の細胞においては難題であり、その治療上の妥当性は非標的細胞のリガーゼIVを阻害する見込まれるリスクによって限定される。HDR効率は高度に細胞周期依存的であるので、向上したHDR効率は、化学的に同調した細胞へのCas9-ガイドRNA複合体の時間内送達によってもまた達成され得る[44]。しかしながら、細胞を同調することは高度に破壊的であるので、かかるアプローチは細胞培養による研究用途に限定される。いくつかのケースでは、首尾良いHDRの生成物がPAM配列中に変異を含有し、よってもはや爾後のCas9改変の基質ではないように、HDR鋳型を設計し、HDR生成物の総体的な収率を増大させることが可能であるが[75]、かかるアプローチは生成物配列に制約を課す。最近、この戦略は非標的鎖に対して相補的であるssDNAドナーおよび高効率リボヌクレオ蛋白質（RNP）送達の使用とカップリングされて、HDRの効率を実質的に増大させたが、これらのケースでさえも、HDR対NHEJ成果の比は相対的に低い（＜2）[83]。

【０４１０】

DNAバックボーン切断を要求することなしの、プログラム可能な標的座位における1つの核酸塩基から別のものへの変換の直接的な触媒作用は、当該座位に所望ではないランダムなインデルを導入することなしに、HDRに対して相対的に遺伝子修正の効率を増大させ得ることが想像された。そのヌクレアーゼ活性を不活性化するAsp10AlaおよびHis840Ala変異を含有する触媒的にデッドなCas9（dCas9）は、ガイドRNAによってプログラムされた様式でDNAに結合するその能力を保持するが、DNAバックボーンを切断しない[16,47]。原理的に、1つの核酸塩基から別のものへの直接的変換を媒介する酵素または化学触媒とのdCas9のコンジュゲーションは、RNAによってプログラムされた核酸塩基編集を可能にし得る。シトシン（C）の脱アミノ化はシチジンデアミナーゼによって触媒され[29]、ウラシル（U）をもたらし、これはチミン（T）の塩基対形成特性を有する。ガイドRNAによって指定されるDNA座位においてCをUに変換するそれらの能力を試験するために、dCas9をシチジンデアミナーゼ酵素に融合させた。ほとんどの公知のシチジンデアミナーゼはRNAに対して働き、DNAを受け入れることが公知である少数の例は一本鎖DNAを要求する[48]。dCas9-標的DNA複合体の最近の研究は、剥がれたDNA鎖の少なくとも9ヌクレオチドがCas9:ガイドRNA:DNA「Rループ」複合体の形成の際に対形成していないことを明らかにしている[12]。本当に、Cas9 Rループ複合体の構造においては、剥がれたDNA鎖上のプロトスペーサーの始めの11ヌクレオチドは無秩序的であり、それらの動きが高度に制限されてはいないことを示唆している[76]。非鋳型鎖シトシンにおけるCas9ニッカーゼによって誘導される変異は、細胞性のシチジンデアミナーゼ酵素によるそれらのアクセス可能性から生じ得ることが推測されている[77]。dCas9-標的DNA複合体の最近の研究は、非鋳型鎖上の少なくとも26塩基は、Cas9がその標的DNA配列に結合するときに対形成していないことを明らかにしている[49]。Rループ中の一本鎖DNAのこのストレッチの部分集合は、dCas9がテザリングされたシチジンデアミナーゼの基質としての用をなして、DNA中のCからUへの直接的なプログラム可能な変換を成し遂げ得ることが推論された（図11A）。

【０４１１】

4つの異なるシチジンデアミナーゼ酵素（hAID、hAPOBEC3G、rAPOBEC1、およびpmCDA1）を哺乳類細胞溶解物由来のin vitro転写-翻訳システムによって発現し、ssDNA脱アミノ化について評価した。4つの酵素のうち、rAPOBEC1は試験された条件において最も高いデアミナーゼ活性を示し、dCas9融合体実験に選ばれた（図36A）。rAPOBEC1

をdCas9のC末端に付加することはデアミナーゼ活性を廃絶するが、dCas9のN末端への融合は、未融合の酵素のものと同等のレベルでssDNAに対するデアミナーゼ活性を保存する。異なる長さおよび組成のリンカーによって4つのrAPOBEC1-dCas9融合体を発現および精製し（図36B）、in vitroのシングルガイドRNA（sgRNA）によってプログラムされるdsDNA脱アミノ化について各融合体を評価した（図11A～11Cおよび図15A～15D）。

【０４１２】

効率的な配列特異的sgRNA依存的なC→U変換がin vitroで観察された（図11A～11C）。変換効率は長さ9アミノ酸超のrAPOBEC1-dCas9リンカーを用いると最も多大であった。脱アミノ化に対して感受性の位置の数（脱アミノ化「活性ウィンドウ」）は、リンカー長さが3から21アミノ酸に延長されたときに増大する（図36C～36F、15A～15D）。16残基XTENリンカー[50]は、これらの2つの特徴の間の有望なバランスをオファーすることが見出され、プロトスペーサー隣接モチーフ（PAM）から遠位の末端を位置1としてカウントしてプロトスペーサー内の位置4から8までのおよそ5ヌクレオチドの効率的な脱アミノ化ウィンドウを有した。rAPOBEC1-XTEN-dCas9蛋白質は第1世代核酸塩基編集因子（NBE1）としての用をなす。

【０４１３】

C→T核酸塩基編集によって理論上修正され得るヒト疾患に該当する7つの変異を選出し、対応する配列の二本鎖DNA 80merを合成し、これらの変異をin vitroで修正するNBE1の能力を査定した（図16A～16B）。NBE1は、in vitroのこれらの7つの標的の6つにおいて、標的C、または複数のCが存在するときには活性ウィンドウ内の少なくとも1つのCの効率的な編集と整合する生成物を生み、44%という見かけ上の平均編集効率を有した（図16A～16B）。複数のCが脱アミノ化ウィンドウ内に存在する3つのケースでは、これらのシトシンのいくつかまたは全ての脱アミノ化の証拠が観察された。試験された7つのケースの1つのみにおいて、編集された生成物の実質的な収率が観察された（図16A～16B）。APOBEC1基質の好ましい配列コンテキストはCCまたはTCであることが報告されているが[51]、標的座位へのdCas9結合によって媒介されるデアミナーゼおよびその一本鎖DNA基質の増大した有効モル濃度は、この制限を緩和し得ることが期待された。NBE1の配列コンテキストの一般性を照らし出すために、プロトスペーサー内の位置7に単一の固定されたCを含有する60mer二本鎖DNAオリゴヌクレオチド、ならびにプロトスペーサー塩基1～6および8～13が個々に他の3塩基のそれぞれに変えられた全ての36個の一重変異したバリアントを編集するその能力をアッセイした。in vitro Cas9アッセイの標準的な条件に類似に、これらの37個の配列のそれぞれを、1.9μM NBE1、1.9μMの対応するsgRNA、および125nM DNAによって2hに渡って処置した[52]。高スループットDNAシーケンシング（HTS）は標的鎖の50から80%のC→U変換を明らかにした（両方のDNA鎖から生ずる合計の配列リードの25から40%。これらの1つはNBE1の基質ではない）（図12A）。標的Cの周辺のヌクレオチドは編集効率に少ない効果を有し、標的Cの直ちに5'の塩基がGでない限り（このケースでは編集効率は実質的により低かった）配列コンテキストとは独立であった（図12A～12B）。全ての4つのNCモチーフに対するin vitroのNBE1活性をプロトスペーサー内の位置1～8において査定した（図12A～12B）。一般的に、基質に対するNBE1活性は順序TC≧CC≧AC＞GCを踏襲することが観察され、最大編集効率は標的Cが位置7にまたは近くにあるときに達成された。加えて、核酸塩基編集因子は高処理能力であり、5塩基活性ウィンドウ内においては同じDNA鎖上の全てのCのほとんどをUに効率的に変換するであろうことが観察された（図17）。

【０４１４】

BE1は試験管内においては基質を効率的にプロセシングするが、細胞では、樹形図的な可能なDNA修復成果が塩基編集の最初のU:G生成物の運命を決定する（図29A）。ヒト細胞における核酸塩基編集の有効性を試験するために、NBE1コドン使用法を哺乳類発現のために最適化し、C末端核局在配列（NLS）を付加し[53]、ヒトゲノム中の6つの良く研究された標的部位における14個のCにおいてヒト細胞のCをTに変換するその能力をアッセイした（図37A）[54]。NBE1を6つの異なるゲノム部位に対応する合成80merとインキュベー

ションすること、次にHTSによって、編集可能なCをin vitroで各プロトスペーサー内に確認した（図13A～13C、図29Bおよび図25）。次に、NBE1および6つの標的sgRNAの1つをコードするプラスミドによってHEK293T細胞をトランスフェクションし、核酸塩基編集が生起するために3日を許し、ゲノムDNAを細胞から抽出し、座位をHTSによって分析した。標的座位における細胞のC→T編集が全ての6つのケースで観察されたが、核酸塩基編集の効率は、合計のDNA配列の1.1%～6.3%または0.8%～7.7%（標的鎖の2.2%～12.6%に対応する）、in vitro核酸塩基編集のものと比較して効率の6.3倍～37倍または5倍～36倍の減少であった（図13A～13C、図29Bおよび図25）。基質CがTによって先行されているときには、位置4から8の典型的なウィンドウの外にいくつかの塩基編集が観察された。我々はこれをTC基質についてのAPOBEC1の通常でなく高い活性に帰する[48]。

【０４１５】

U:Gヘテロ二重鎖DNAの存在に対する細胞のDNA修復応答が細胞における核酸塩基編集効率の大きい減少を担っているかどうかを問うた（図29A）。ウラシルDNAグリコシラーゼ（UDG）は細胞のDNAからのUの取り除きを触媒し、塩基除去修復（BER）を開始し、U:G対からC:G対への復帰を最も一般的な成果として有する（図29A）[55]。B. subtilisバクテリオファージPBS1からの83残基蛋白質、ウラシルDNAグリコシラーゼ阻害因子（UGI）は、ヒトUDG活性を強力にブロックする（IC_{50}=12pM）[56]。UGIをNBE1のC末端に融合させて第2世代核酸塩基編集因子NBE2を作出し、全ての6つのゲノム座位に対する編集アッセイを反復した。ヒト細胞における編集効率はNBE2ではNBE1のものよりも平均で3倍高く、合計のDNA配列の22.8%まで（標的鎖の45.6%まで）の遺伝子変換効率をもたらした（図13A～13Cおよび図29B）。塩基編集をヒト細胞において試験するために、BE1コドン使用法を哺乳類発現のために最適化し、C末端核局在配列（NLS）を付加した[53]。

【０４１６】

UGIを過剰発現する別個のプラスミドをNBE1とコトランスフェクションしたときには、類似の編集効率が観察された（図18A～18H）。しかしながら、NBE1へのUGIの直接的な融合は、モニタリングされた非標的のゲノムの場所においてC→T変異の有意な増大をもたらさなかったが、未融合のUGIの過剰発現はゲノム中の他所におけるC→T変異の頻度を検出可能に増大させた（図18A～18H）。U20S細胞において同じ6つのゲノム標的に対する編集効率を査定することによって、NBE2によって媒介される核酸塩基編集の一般性を確認し、HEK293T細胞のものに類似の結果を観察した（図19）。重要なことに、NBE2は典型的にはいずれかの検出可能なインデルをもたらさず（図13Cおよび図29C）、二本鎖DNA切断へのNHEJの公知のメカニズム的な依存性と整合した[57,78]。一緒になって、これらの結果は、UGIをNBE1にコンジュゲーションすることがヒト細胞の核酸塩基編集の効率を多大に増大させ得ることを指示する。

【０４１７】

HEK293T細胞の複数の細胞分裂に渡って試験されたゲノム座位の2つにおける編集効率をモニタリングすることによって、ヒト細胞における核酸塩基編集の永続性を確認した。ゲノムDNAを2つの時点において収穫した：NBE2および適切なsgRNAを発現するプラスミドによるトランスフェクションの3日後、ならびに細胞を継代し、それらを追加の4日に渡って成長させた後（およそ5つの爾後の細胞分裂）。非継代細胞（編集は3つの異なる標的Cについて標的鎖の4.6%～6.6%で観察された）および継代細胞（編集は同じ3つの標的Cについて標的鎖の4.6%～6.4%で観察された）の間に編集効率の有意な変化は観察されず、核酸塩基編集が細胞分裂後に永続的になることを確認した（図20）。稀に、インデルは細胞性の修復プロセスによるU:G損傷のプロセシングから生ずるであろう。これには、インデルに至ることが公知である一本鎖切断中間体が関与する[84]。数百の内在性のU:G損傷が偶発的なシチジンデアミナーゼから毎日ヒト細胞あたり創出されることに鑑みると[85]、U:G損傷修復からの合計のインデル頻度が単一の標的座位においてBE1またはBE2活性から増大することは非蓋然的であることが期待された。

【０４１８】

細胞の核酸塩基編集の効率をさらに増大させるために、編集されない鎖にニックを入れ

ることは、編集されたUのより小さい画分が細胞によって取り除かれることをもたらし得ることが期待された。なぜなら、真核生物のミスマッチ修復機構は、鎖の不連続性を用いて、ミスマッチの二重鎖のいずれかの切断された鎖にDNA修復を導くからである（図29A）[58,79,80]。触媒His残基をCas9 HNHドメインの位置840において回復させ[47,59]、標的Cに対向するGを含有する編集されない鎖にニックを入れるが、Cを含有する標的鎖を切断しない第3世代核酸塩基編集因子NBE3をもたらした。NBE3はCas9のAsp10Ala変異をなお含有するので、それは二本鎖DNA切断を誘導しない。編集されない鎖にニックを入れるこの戦略は、NBE2に対して相対的に、ヒト細胞における核酸塩基編集効率を追加の1.4～4.8倍増強し、合計のDNA配列の36.3%までがHEK293T細胞の同じ6つのヒトゲノム標的上に標的C→T変換を含有することをもたらした（図13A～13Cおよび図29B）。重要なことに、平均0.8%である（6つの異なる座位について0.2%～1.6%までの範囲の）小さい頻度のインデルのみが、NBE3処置からは観察された（図13C、図29C、および図34）。対照的に、これらの座位の3つにおけるHDRを媒介するように野生型Cas9、sgRNA、および一本鎖DNAドナー鋳型によって細胞を処置したときには、平均0.7%のみであるC→T変換効率が観察され、平均3.9%であるずっと高い相対的なインデル形成を有した（図13A～13Cおよび図29C）。アレル変換対NHEJ成果の比は、BE2では平均＞1,000、BE3では23、野生型Cas9では0.17であった（図3c）。我々は、HEK293T細胞の複数の細胞分裂に渡ってHEK293部位3および4ゲノム座位における編集効率をモニタリングすることによって、ヒト細胞における塩基編集の永続性を確認した（図38）。総合すると、これらの結果は、核酸塩基編集が、ヒト細胞において、Cas9によって媒介されるHDRよりもずっと効率的な標的化された1塩基編集を遂げ得、インデル形成がずっと少ないか（NBE3）またはない（NBE2）ことを確立する。

【０４１９】

次に、ヒト細胞におけるNBE1、NBE2、およびNBE3のオフターゲット活性を評価した。Cas9、dCas9、およびCas9ニッカーゼのオフターゲット活性は大がかりに研究されている（図23～24および31～33）[54,60,62]。図12A～12Bにおいて観察された配列コンテキスト非依存性と整合して、rAPOBEC1の配列選好性は標的のCから1よりも多い塩基にあるDNA塩基とは独立であることが示されているので[63]、核酸塩基編集因子の見込まれるオフターゲット活性はオフターゲットCas9結合から生ずることが想定された。Cas9オフターゲット部位のある画分のみが核酸塩基編集の活性ウィンドウ内にCを有するであろうから、オフターゲット核酸塩基編集部位は古典的Cas9バリアントのオフターゲット部位の部分集合であるはずである。研究された6つの部位のそれぞれについて、GUIDE-seq法を用いて先に決定されたヒト細胞の上位10個の公知のCas9オフターゲット座位をシーケンシングした（図23～27および31～33）[54,61]。公知のdCas9オフターゲット座位の部分集合のみにおける検出可能なオフターゲット核酸塩基編集（NBE1およびNBE2では16/34、47%、NBE3では17/34、50%）が観察された。全てのケースにおいて、オフターゲット塩基編集基質は5塩基標的ウィンドウ内にCを含有した。一般的に、オフターゲットC→T変換は、オフターゲットのCas9ヌクレアーゼによって媒介されるゲノム改変頻度と符合した（図23～27）。およそ1.8×10^{6}細胞に由来する合計で14,700,000個の配列リードに相当する、試験された6つのオンターゲットおよび34個のオフターゲット座位の周辺の2,500個の別々のシトシンにおけるC→T変換をもまたモニタリングした。非処置の細胞のものと比較して、NBE1、NBE2、またはNBE3処置によって、これらの他の部位のいずれかにおけるC→T変換の検出可能な増大は観察されなかった（図28）。一緒にすると、これらの知見は、核酸塩基編集因子のオフターゲット基質がCas9オフターゲット基質の部分集合を包含することと、ヒト細胞における核酸塩基編集因子が、ここで用いられる方法によって検出され得るレベルでは、標的化されていないC→T変換をゲノム中に誘導しないこととを示唆する。非継代HEK293T細胞（編集は、BE2では3つの標的Cについてシーケンシングされた鎖の1.8%から2.6%、BE3では6.2%～14.3%で観察された）および塩基編集後におよそ5つの細胞分裂を経た細胞（編集は、BE2では同じ標的Cについてシーケンシングされた鎖の1.9%～2.3%、BE3では6.4%～14.5%で観察された）の間には、編集効率の

実質的な変化は観察されず、これらの細胞における塩基編集は耐久性であることを確認した（拡張データ図6）。

【０４２０】

最後に、哺乳類細胞において3つの疾患に該当する変異を修正する核酸塩基編集のポテンシャルを試験した。アポリポ蛋白質E遺伝子バリアントAPOE4はアミノ酸位置112および158に2つのArg残基をコードし、遅発型アルツハイマー病の最も大きくかつ最も一般的な遺伝学的リスク因子である[64]。APOE2(Cys112/Cys158)、APOE3(Cys112/Arg158)、およびAPOE3'(Arg112/Cys158)を包含する位置112または158にCys残基を有するApoEバリアントは、APOE4よりも実質的に低いアルツハイマー病リスクを授けることが示されているか[65]または推定されている[81]。in vitroでAPOE4をAPOE3'に変換するNBE1の能力によって触発され（図16A～16B）、内在性のマウスAPOE遺伝子がヒトAPOE4によって置き換えられた不死化マウスアストロサイト（Taconic）において、この変換を試みた。NBE3および適切なsgRNAをコードするDNAをヌクレオフェクションによってこれらのアストロサイトに送達し（25%のヌクレオフェクション効率）、2日後に全ての処置された細胞からゲノムDNAを抽出し、編集効率をHTSによって測定した。Arg158からCys158への変換が合計のDNAシーケンシングリードの58～75%（ヌクレオフェクションされたアストロサイトの44%）において観察された（図14A～14Cおよび図30A）。予想される通り、コドン158の第3の位置における合計のDNAの36～50%の編集およびLeu159の第1の位置における合計のDNAの38～55%の編集もまた観察された。なぜなら、これらのCの全ての3つは活性な核酸塩基編集ウィンドウ内にあるからである。しかしながら、TGCおよびTGT両方はCysをコードし、CTGおよびTTG両方はLeuをコードするので、他の2つのC→T変換のどちらもApoE3'蛋白質のアミノ酸配列の変化をもたらさない。1×10^6細胞に由来する＞1,500,000のシーケンシングリードから、NBE3処置後の標的座位における1.7%のインデルの証拠が観察された（図35）。対照的に、wtCas9およびドナーssDNAによるアストロサイトの同一の処置は、Cas9およびHDRを用いる1塩基修正の先の報告と整合する効率、標的座位における0.1～0.3%のAPOE4修正および26～40%のインデルをもたらした[45,75]（図30Aおよび図40A）。同一にだがVEGFA座位を標的化するsgRNAによって処置されたアストロサイトは、APOE4塩基編集の証拠を表示しなかった（図34および図40A）。これらの結果は、それらの処理能力がゲノムDNA中に1つよりも多くのヌクレオチド変化をもたらすときでさえも、核酸塩基編集因子が編集の主要生成物として蛋白質のコード配列中に正確な1アミノ酸変化を成し遂げ得ることを実証している。Cas9、dCas9、およびCas9ニッカーゼのオフターゲット活性は大がかりに研究されている[54,60～62]。一般的に、BE1、BE2、およびBE3によるオフターゲットC→T変換は、オフターゲットのCas9ヌクレアーゼによって媒介されるゲノム改変頻度と符合した。

【０４２１】

ドミナントネガティブp53変異Tyrl63CysおよびAsn239Aspはいくつかの型の癌に強く関連している[66,67]。これらの変異の両方は鋳型鎖上のC→T変換によって修正され得る（図16A～16B）。NBE3およびTyrl63Cysを修正するようにプログラムされたsgRNAをコードするDNAによって、p53 Tyrl63Cys変異についてホモ接合であるヒト乳癌細胞株（HCC1954細胞）をヌクレオフェクションした。HCC1954細胞のヌクレオフェクション効率は＜10%であったので、IRFPを発現するプラスミドをこれらの細胞にコヌクレオフェクションして、処置の2日後の蛍光活性化セルソーティングによるヌクレオフェクションされた細胞の単離を可能にした。ゲノムDNAのHTSはヌクレオフェクションされたHCC1954細胞の7.6%におけるTyr163Cys変異の修正を明らかにした（図30Bおよび図40A～40B）。NBE2およびAsn239Aspを修正するようにプログラムされたsgRNAをコードするDNAによって、p53 Asn239Aspについてヘテロ接合であるヒトリンパ腫細胞株（ST486細胞）をもまたヌクレオフェクションし、92%のヌクレオフェクション効率を有した）。処置されたST486細胞の11%（ヌクレオフェクションされたST486細胞の12%）において、Asn239Asp変異の修正が観察された。HEK細胞における知見と整合して、インデルはNBE2によるST486細胞の処置からは観察されず、0.6%のインデル形成がNBE3によ

るHCC1954細胞の処置から観察された。2×10^{5}細胞に由来する＞2,000,000個のシーケンシングリードからは、プロトスペーサーの両側の少なくとも50塩基対以内において、他のDNA変化は非処置の対照のものよりも上の頻度では検出されなかった（図14A～14C、図30B）。総合すると、これらの結果は、我々の知識では、他の方法を用いる現行では達成可能でない効率と他のゲノム改変イベントの欠如とによる、ゲノムDNA中の3つの疾患関連アレルからそれらの野生型形態への変換に相当する。

【０４２２】

ヒト遺伝子疾患に対処するための核酸塩基編集因子の見込まれる妥当性を照らし出すために、NCBI ClinVarデータベース[68]を、原理的にこのアプローチによって修正され得る公知の遺伝子疾患について検索した。第1に一塩基多型（SNP）のみを調べ、それからいずれかの非病原性バリアントを取り除くことによって、ClinVarをフィルタリングした。24,670個の病原性SNPのうち、3,956個はT→CまたはA→G置換どちらかによって引き起こされる。SNPを脱アミノ化活性ウィンドウ内に位置させるであろう近傍のNGG PAMを有するバリアントのみを包含するように、このリストをさらにフィルタリングし、原理的にはここで記載される核酸塩基編集因子によって修正され得る1,089個の臨床的に該当する病原性遺伝子バリアントをもたらした（図21）。

【０４２３】

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編集因子のいずれかは疾患または異常を処置するために用いられ得る。例えば、本明細書において提供される塩基編集因子のいずれかは、本明細書において提供される疾患または異常のいずれかに関連する1つ以上の変異を修正するために用いられ得る。処置され得る例示的な疾患または異常は、限定なしに、3-メチルグルタコン酸尿症2型、46,XY性腺異形成症、4-アルファ-ヒドロキシフェニルピルビン酸ヒドロキシラーゼ欠乏症、6-ピルボイルテトラヒドロプテリンシンターゼ欠乏症、全色盲、酸不安定サブユニット欠乏症、先端異骨症、掌蹠角化症、ACTH抵抗性、ACTH非依存性大結節性副腎過形成、活性化PI3K-デルタ症候群、急性間欠性ポルフィリン症、急性骨髄性白血病、アダムズ・オリバー症候群1／5／6、アデニロコハク酸リアーゼ欠乏症、副腎白質ジストロフィー、成人型神経セロイドリポフスチン症、眼球運動失行を伴う成人発症型失調症、睡眠相前進症候群、加齢黄斑変性、アラジール症候群、アレキサンダー病、アラン・ハーンドン・ダドリー症候群、X連鎖劣性アルポート症候群、小児交互性片麻痺、肺静脈のアライメント不整を伴う肺胞毛細血管異形成、エナメル質形成不全症、アミロイド形成性トランスサイレチンアミロイドーシス、筋萎縮性側索硬化症、貧血（G6PD欠乏による非球状溶血性）、貧血（鉄芽球性、ピリドキシン不応性、常染色体劣性）、無爪症、アンチトロンビンIII欠乏症、大動脈瘤、再生不良性貧血、アポリポ蛋白質C2欠乏症、見かけ上のミネラルコルチコイド過剰、アロマターゼ欠乏症、不整脈原性右室心筋症、家族性肥大型心筋症、肥大型心筋症、先天性多発性関節拘縮症、アスパルチルグルコサミン尿症、窒息性胸郭ジストロフィー、ビタミンE欠損を伴う運動失調、運動失調（痙性）、心房細動、心房中隔欠損症、非典型溶血性尿毒症症候群、常染色体優性CD11C+/CD1C+樹状細胞欠乏症、ミトコンドリアDNA欠失を伴う常染色体優性進行性外眼筋麻痺、バライトサー・ウィンター症候群、バーター症候群、基底核石灰化、ベックウィズ・ヴィーデマン症候群、良性家族性新生児痙攣、良性型肩甲腓骨型筋ジストロフィー、ベルナール・スーリエ症候群、ベータサラセミア・インターメディア、ベータ-D-マンノシドーシス、ビエッティクリスタリン角膜網膜ジストロフィー、胆汁酸吸収不良、ビオチニダーゼ欠乏症、ベルエソン・フォルスマン・レーマン症候群、ブーシェ・ノイハウザー症候群、ボーエン・コンラディ症候群、短指症、ブラウン・ヴィアレット・ファンラール症候群、ブルガダ症候群、心臓不整脈、心臓顔皮膚症候群、心筋症、カルネヴァーレ症候群、カルニチンパルミトイルトランスフェラーゼII欠乏症、カーペンター症候群、白内障、カテコラミン誘発多形性心室頻拍、セントラルコア病、染色体1、9、および16のセントロメア不安定性および免疫不全症、常染色体優性脳動脈症、脳・眼・顔・骨格症候群、セロイドリポフスチン症、シャルコー・マリー・トゥース病、コレステロール蓄積症、軟骨石灰化症、軟骨異形成症、慢性進行性多発性硬化症、コエンザイムQ10欠乏症、コ

ーエン症候群、第V因子および第VIII因子複合欠乏症、複合免疫不全症、複合型酸化的リン酸化欠乏症、複合型部分17-アルファ-ヒドロキシラーゼ／17,20-リアーゼ欠乏症、補体因子d欠乏症、複合型完全17-アルファ-ヒドロキシラーゼ／17,20-リアーゼ欠乏症、桿体錐体ジストロフィー、先天性拘縮性クモ状指趾症、グリコシル化の先天性異常、先天性脂肪腫性過成長、卵巣の新生物、PIK3CA関連過成長スペクトラム、先天性QT延長症候群、先天性筋ジストロフィー、先天性筋肥大・脳症候群、先天性筋無力症候群、線維タイプ不均等症を伴う先天性ミオパチー、アイヒスフェルト型先天性筋ジストロフィー、先天停止性夜盲、角膜ジストロフィー、コルネリア・デランゲ症候群、頭蓋骨幹端骨異形成症、クリグラー・ナジャール症候群、クルーゾン症候群、骨ジストロフィーを伴う皮膚弛緩症、チアノーゼ、嚢胞性線維症、シスチン症、チトクロムcオキシダーゼ欠乏症、ミトコンドリア複合体I欠乏症、D-2-ヒドロキシグルタル酸尿症、ダノン病、迷路形成不全、小耳、および小歯症を伴う難聴（LAMM）、難聴、アセチルCoAアセチルトランスフェラーゼ欠乏症、フェロキシダーゼ欠乏症、UDPグルコース-ヘキソース-1-リン酸ウリジリルトランスフェラーゼ欠乏症、デジェリン・ソッタス病、デビュクオア症候群、DFNA、糖尿病2型、糖尿病・難聴症候群、ダイアモンド・ブラックファン貧血、捻曲性骨異形成症、ジヒドロプテリジン還元酵素欠損症、ジヒドロピリミジナーゼ欠損症、拡張型心筋症、播種性非定型マイコバクテリア感染、遠位型関節拘縮症、遠位遺伝性運動ニューロパチー、ドンナイ・バロー症候群、デュシェンヌ型筋ジストロフィー、ベッカー型筋ジストロフィー、遺伝性汎発性色素異常症、先天性角化不全症、ジストニア、早期乳児てんかん性脳症、エーラス・ダンロス症候群、アイヒスフェルト型先天性筋ジストロフィー、エメリー・ドレフュス型筋ジストロフィー、エナメル腎症候群、反対型栄養障害型表皮水疱症、疱疹状表皮水疱症、てんかん、反復発作性運動失調、変動性紅斑角皮症、赤芽球性プロトポルフィリン症、運動不耐、滲出性硝子体網膜症、ファブリー病、第V因子欠乏症、第VII因子欠乏症、第xiii因子欠乏症、家族性大腸腺腫症、乳癌、卵巣癌、寒冷蕁麻疹、慢性乳児神経皮膚関節症候群、片麻痺性片頭痛、高コレステロール血症、肥大型心筋症、低アルファリポ蛋白血症、低カリウム血症・低マグネシウム血症、若年性痛風、高リポ蛋白血症、限局性アミロイドーシス、ビタミンD不応性低リン血症性くる病、FG症候群、外眼筋線維症、フィンランド型先天性ネフローゼ症候群、焦点性てんかん、巣状分節性糸球体硬化症、前頭鼻異形成、前頭側頭型認知症、フルクトースビスホスファターゼ欠損症、ガムストープ・ウォールファールト症候群、ガングリオシドシアリダーゼ欠損症、GATA-1関連血小板減少症、ゴーシェ病、巨大軸索性ニューロパチー、グランツマン血小板無力症、糸球体嚢胞腎、糸球体症、グルココルチコイド不応症、グルコース-6-リン酸輸送欠乏症、グルタル酸尿症、糖原病、ゴーリン症候群、全前脳胞症、GRACILE症候群、出血性末梢血管拡張症、ヘモクロマトーシス、ヘモグロビンH症、溶血性貧血、血球貪食症候群、結腸の癌腫、マイアー症候群、白質脳症、遺伝性第IX因子欠乏症、遺伝性第VIII因子欠乏症、遺伝性第XI因子欠乏症、遺伝性フルクトース尿症、遺伝性非ポリポーシス大腸新生物、遺伝性膵炎、遺伝性熱変形赤血球症、楕円赤血球症、ヘテロタキシー、異所性灰白質、組織球・骨髄細網症、組織球症・リンパ節腫大プラス症候群、アルドラーゼA欠乏によるHNSHA、ホロカルボキシラーゼ合成酵素欠損症、ホモシステイン血症、ハウエル・エバンス症候群、胞状奇胎、高カルシウム尿性高カルシウム血症、高免疫グロブリンD、メバロン酸尿症、高インスリン血性低血糖症、高カリウム性周期性四肢麻痺、フォン・オイレンブルクの先天性パラミオトニー、高リポ蛋白血症、高マンガン血症、高メチオニン血症、高リン血症、高血圧、低マグネシウム血症、低ベータリポ蛋白血症、低カルシウム血症、低ゴナドトロピン性性腺機能低下症、低ゴナドトロピン性性腺機能低下症、低汗性外胚葉形成不全症、高IgM免疫不全、低汗性Ｘ連鎖外胚葉形成不全症、低マグネシウム血症、副甲状腺機能低下症、特発性肺線維症、免疫不全、免疫グロブリンA欠損症、乳児低ホスファターゼ症、乳児ジストニア・パーキンソニズム、インスリン依存型糖尿病、中間型メープルシロップ尿症、坐骨膝蓋骨異形成、膵島細胞過形成、成長ホルモン単独欠損症、ルトロピン単独欠損症、イソ吉草酸血症、ジュベール症候群、若年性ポリポーシス症候群、若年網膜分離症、カルマン症候群、カルタゲナー症候群、クーゲルベルグ・ウェランダー病、格子状角膜ジスト

ロフィー、レーバー先天性黒内障、レーバー視神経萎縮症、左室緻密化障害、リー症候群、ミトコンドリア複合体Ｉ欠損症、妖精症、関節拘縮症、前角細胞病、白血球接着不全症、白質ジストロフィー、白質脳症、卵巣白質ジストロフィー、L-フェリチン欠損症、リ・フラウメニ症候群、肢帯型筋ジストロフィー-ジストログリカノパチー、ロイス・ディーツ症候群、QT延長症候群、大頭症／自閉症症候群、黄斑角膜ジストロフィー、黄斑ジストロフィー、悪性高熱症素因、前立腺の悪性腫瘍、メープルシロップ尿症、マーデンウォーカー様症候群、マルファン症候群、マリー・ウンナ遺伝性乏毛症、マスト細胞症、胎便関連性腸閉塞、中鎖アシル補酵素Aデヒドロゲナーゼ欠損症、メルニック・フレイザー症候群、精神遅滞、メロシン欠損型先天性筋ジストロフィー、中皮腫、異染性白質ジストロフィー、骨幹端軟骨異形成症、メトヘモグロビン血症、メチルマロン酸尿症、ホモシスチン尿症、小頭症、脈絡網膜症、リンパ浮腫、小眼球症、軽症型非PKU型高フェニルアラニン血症、ミッチェル・ライリー症候群、ミトコンドリア3-ヒドロキシ-3-メチルグルタリル-CoAシンターゼ欠損症、ミトコンドリア複合体I欠損症、ミトコンドリア複合体III欠損症、ミトコンドリアミオパチー、ムコリピドーシスIII、ムコ多糖症、多種スルファターゼ欠損症、筋無力症候群、結核菌、ミエロペルオキシダーゼ欠損症、マイアー症候群、ミオクロニーてんかん、筋原線維性ミオパチー、ミオグロビン尿症、ミオパチー、近視、先天性ミオトニー、ナバホ神経肝症、ネマリンミオパチー、胃の新生物、腎性尿崩症、ネフロン癆、ネフローゼ症候群、神経線維腫症、中性脂肪蓄積症、ニーマン・ピック病、非ケトーシス型高グリシン血症、ヌーナン症候群、ヌーナン症候様異常、ノーラム病、黄斑変性、N末端アセチルトランスフェラーゼ欠損症、眼皮膚白皮症、眼歯指異形成、大堂症候群、視神経無形成、オルニチンカルバモイルトランスフェラーゼ欠損症、口腔顔面指趾症候群、骨形成不全症、大理石骨病、卵巣形成不全、爪甲肥厚、非表皮融解性掌蹠角化症、パピヨン・ルフェーブル症候群、ハイム・ムンク症候群、歯周炎、ピーリングスキン症候群、ペンドレッド症候群、ペルオキシソーム脂肪酸アシルCoAレダクターゼ1異常、ペルオキシソーム生合成異常、ファイファー症候群、フェニルケトン尿症、フェニルケトン尿症、非PKU型高フェニルアラニン血症、下垂体ホルモン欠損症、毛孔性紅色粃糠疹、結節性多発動脈炎、多発性嚢胞腎、多嚢胞性脂肪膜性骨異形成症、多小脳回症、橋小脳低形成症、汗孔角化症、脊髄後索型運動失調、原発性肢端紅痛症、高シュウ酸尿症、進行性家族性肝内胆汁うっ滞症、進行性偽リウマチ性異形成症、プロピオン酸血症、仮性半陰陽、偽性低アルドステロン症、弾性線維性仮性黄色腫様異常、プリンヌクレオシドホスホリラーゼ欠損症、ピリドキサール-5-リン酸依存性てんかん、腎異形成、網膜色素変性、小脳性運動失調、骨異形成、細網異形成症、網膜色素変性、アッシャー症候群、網膜芽細胞腫、網膜症、RRM2B関連ミトコンドリア病、ルビンスタイン・テイビ症候群、シュナイダークリスタリン角膜ジストロフィー、脂腺腫瘍、重症先天性好中球減少症、乳児重症ミオクロニーてんかん、重症Ｘ連鎖ミオチュブラーミオパチー、指爪甲形成不全症、顔面異形、乏毛症、短肋骨胸郭異形成症、シアル酸蓄積症、シアリドーシス、鉄芽球性貧血、小径線維ニューロパチー、スミス・マゲニス症候群、ソースビー眼底ジストロフィー、痙性運動失調、痙性対麻痺、精子形成不全、球状赤血球症、スフィンゴミエリン／コレステロールリピドーシス、脊髄小脳性運動失調、裂手裂足症、脊椎骨端骨幹端異形成症、致死性扁平椎異形成症、頭頸部の扁平上皮癌、スターガルト病、スクラーゼ・イソマルターゼ欠損症、乳幼児突然死症候群、大動脈弁上部狭窄症、サーファクタント代謝異常、タンジール病、タットン・ブラウン・ラーマン症候群、胸部大動脈瘤および大動脈解離、血栓傾向、甲状腺ホルモン不応症、TNF受容体関連周期性発熱症候群（TRAPS）、歯数不足症、トルサード・ド・ポワント、大血管転位、トリーチャー・コリンズ症候群、結節性硬化症症候群、チロシナーゼ陰性型眼皮膚白皮症、チロシナーゼ陽性型眼皮膚白皮症、チロシン血症、UDPグルコース-4-エピメラーゼ欠損症、ウルリッヒ型先天
性筋ジストロフィー、ベスレムミオパチー、アッシャー症候群、UV高感受性症候群、ファンデルワウデ症候群、膝窩翼状片症候群、極長鎖アシルCoAデヒドロゲナーゼ欠損症、膀胱尿管逆流症、硝子体網膜脈絡膜症、フォン・ヒッペル・リンドウ病、フォン・ヴィレブランド病、ワールデンブルグ症候群、ワルシャワ染色体不安定症候群、WFS1関連異常、

ウィルソン病、色素性乾皮症、X連鎖無ガンマグロブリン血症、X連鎖遺伝性運動感覚ニューロパチー、X連鎖重症複合免疫不全症、およびツェルウェーガー症候群を包含する。

【０４２４】

核酸塩基編集の開発はゲノム編集の範囲および有効性両方を進歩させる。ここに記載される核酸塩基編集因子は、事実上インデル形成を有さない編集（NBE2）または低い頻度（ここでは典型的には≦1%）のインデル形成を有するより効率的な編集（NBE3）という選択肢を研究者にオファーする。塩基編集の生成物が定義上もはや基質ではないことは、蓋然的には、従来のCas9用途を邪魔し得る爾後の生成物転換を防止することによって編集効率に寄与する。二本鎖DNA切断と細胞状態および細胞型によって多大に変わる確率的なDNA修復プロセスとへの依拠を取り除くことによって、核酸塩基編集は、クリーンにインストールされ得るゲノム改変の型、これらの改変の効率、および編集が適合可能である細胞の型を拡大するポテンシャルを有する。改善されたDNA特異性を有する最近の人工Cas9バリアント[69,70,82]または送達方法[71]、および変改されたPAM特異性を有するCas9バリアント[72]をこの戦略にインテグレーションして、改善されたDNA特異性を有するかまたはより幅広い範囲の疾患関連変異さえも標的化し得る追加の核酸塩基編集因子を提供し得ることが蓋然的である。これらの知見は、追加の核酸塩基転換を触媒する酵素とのdCas9の追加の融合体を操作することが、核酸塩基編集によってなされ得る可能なDNA塩基変化の画分を増大させるであろうことをもまた示唆している。これらの結果は、プログラム可能なゲノムおよびエピゲノム塩基編集の追加の型を可能にし得る、メチラーゼおよび脱メチル化酵素を包含する他のDNA修飾酵素の融合のためのアーキテクチャをもまた示唆している。

【０４２５】

材料および方法

クローニング。

この文書において用いられる全ての構築物およびプライマーのDNA配列は補足的な配列にリスト化されている。NBE1、NBE2、およびNBE3をコードする遺伝子を含有するプラスミドはAddgeneから入手可能であろう。PCRはVeraSeq ULtra DNAポリメラーゼ（Enzymatics）またはQ5ホットスタート高フィデリティDNAポリメラーゼ（New England Biolabs）を用いて行った。NBEプラスミドはUSERクローニング（New England Biolabs）を用いて構築した。デアミナーゼ遺伝子はgBlocks遺伝子断片として合成し（Integrated DNA Technologies）、Cas9遺伝子は先に報告されているプラスミドから得た[18]。デアミナーゼおよび融合遺伝子をpCMV（哺乳動物コドンに最適化されている）またはpET28b（E. coliコドンに最適化されている）バックボーン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導入を用いて構築した。簡潔には、補足的な配列にリスト化されているプライマーを製造者の説明書に従って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、製造者の説明書に従って、リン酸化プライマーおよび鋳型としてのプラスミドpFYF1320（EGFP sgRNA発現プラスミド）によって、Q5ホットスタート高フィデリティポリメラーゼ（New England Biolabs）を用いて、PCRを行なった。PCR生成物をDpn1（20U，New England Biolabs）と37℃で1hに渡ってインキュベーションし、QIAprepスピンカラム（Qiagen）によって精製し、製造者の説明書に従って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コンピテントセル（Thermo Fisher Scientific）を用いて実施された。

【０４２６】

ssDNAに対するin vitroデアミナーゼアッセイ。

全てのssDNA基質の配列は補足的な配列にリスト化されている。全てのCy3標識基質はIntegrated DNA Technologies（IDT）から得た。デアミナーゼは、1μgのプラスミドを用いて、製造者の説明書に従ってTNT T7 Quick Coupled転写/翻訳キット（Promega）を用いてin vitro発現した。蛋白質発現後に、5μLの溶解物を35μLのssDNA（1.8μM）およびUSER酵素（1単位）とCutSmart緩衝液（New England Biolabs）（50mM

酢酸カリウム，29mMトリス酢酸，10mM酢酸マグネシウム，100ug/mL BSA，pH7.9）中で組み合わせ、37℃で2hに渡って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）上で未改変の全長基質から分離した。

【０４２７】

His_6-rAPOBEC1-リンカー-dCas9融合体の発現および精製。

GGS、$(GGS)_3$（配列番号610）XTEN、または$(GGS)_7$（配列番号610）リンカーを有する$pET28b\text{-}His_6$-rAPOBEC-リンカー-dCas9をコードするプラスミドによって、E. Coli BL21 STAR（DE3）コンピテントセル（Thermo Fisher Scientific）を形質転換した。もたらされた発現株を、100μg/mLのカナマイシンを含有するルリア・ベルターニ（LB）ブロスによって37℃で一晩成長させた。細胞を同じ成長培地によって1:100希釈し、37℃でOD_{600}=～0.6まで成長させた。培養物を2hの期間をかけて4℃に冷却し、イソプロピル-β-D-l-チオガラクトピラノシド（IPTG）を0.5mMで追加して蛋白質発現を誘導した。～16h後に、細胞を4,000gにおける遠心によって収集し、溶解緩衝液（50mMトリス（ヒドロキシメチル）アミノメタン（Tris）-HCl，pH7.0，1M NaCl，20%グリセロール，10mMトリス（2-カルボキシエチル）ホスフィン（TCEP，Soltec Ventures））中に再懸濁した。細胞をソニケーションによって溶解し（6W出力で合計8minに渡って20sパルスオン，20sパルスオフ）、15minの25,000gにおける遠心後に、溶解物の上清を単離した。溶解物をHis-Purニッケル-ニトリロ酢酸（ニッケル-NTA）樹脂（Thermo Fisher Scientific）と4℃で1hに渡ってインキュベーションして、Hisタグ付き融合蛋白質をキャプチャーした。樹脂をカラムに移し、40mLの溶解緩衝液によって洗浄した。Hisタグ付き融合蛋白質を、285mMイミダゾールを添加した溶解緩衝液によって溶出し、限外濾過（Amicon-Millipore，100kDa分子量カットオフ）によって1mL合計体積まで濃縮した。蛋白質を、50mMトリス（ヒドロキシメチル）アミノメタン（Tris）-HCl，pH7.0，0.1M NaCl，20%グリセロール，10mM TCEPを含有する低塩精製緩衝液によって20mLまで希釈し、SP Sepharose Fast Flow樹脂（GE Life Sciences）にローディングした。樹脂を40mLのこの低塩緩衝液によって洗浄し、50mMトリス（ヒドロキシメチル）アミノメタン（Tris）-HCl，pH7.0，0.5M NaCl，20%グリセロール，10mM TCEPを含有する5mLの活性緩衝液によって蛋白質を溶出した。溶出された蛋白質をSDSPAGEゲルによって定量した。

【０４２８】

sgRNAのin vitro転写。

20bp sgRNA標的配列によって後続されたT7プロモーターを含有する線状DNA断片を、補足的な配列にリスト化されているプライマーを用いて、製造者の説明書に従ってTranscriptAid T7高収率転写キット（Thermo Fisher Scientific）によってin vitro転写した。sgRNA生成物は製造者の説明書に従ってMEGAclearキット（Thermo Fisher Scientific）を用いて精製し、UV吸光度によって定量した。

【０４２９】

Cy3がコンジュゲーションされたdsDNA基質の調製。

80ヌクレオチド非標識鎖の配列が補足的な配列にリスト化されており、PAGE精製オリゴヌクレオチドとしてIDTにオーダーした。補足的な配列にリスト化されている25nt Cy3標識プライマーは各80nt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標識されたdsDNA基質を創出するために、80nt鎖（5μLの100μM溶液）をdNTP（0.75μLの100mM溶液）と共にNEBuffer 2（38.25μLの50mM NaCl，10mM Tris-HCl，10mM $MgCl_2$，1mM DTT，pH7.9溶液，New England Biolabs）中でCy3標識プライマー（5μLの100pM溶液）と組み合わせ、5minに渡って95℃に加熱し、次に0.1℃/sの速度における45℃への徐々の冷却をした。このアニーリング期間後に、Klenow exo$^-$（5U，New England Biolabs）を追加し、反応を37℃で1hに渡ってインキュベーションした。溶液をPB緩衝液（250μL，Qiagen）およびイソプロパノール（50μL）によって希釈し、QIAprepスピンカラム（Qiagen）によって精製し、50μLのTris緩衝液によって溶出した。

【０４３０】
dsDNAに対するデアミナーゼアッセイ。
精製された融合蛋白質（20μLの活性緩衝液中の1.9μM）を1当量の適切なsgRNAと組み合わせ、常温で5minに渡ってインキュベーションした。Cy3標識dsDNA基質を125nMの終濃度まで追加し、もたらされた溶液を37℃で2hに渡ってインキュベーションした。dsDNAをPB緩衝液（100μL，Qiagen）およびイソプロパノール（25μL）の追加によって融合体から分離し、EconoSpinマイクロスピンカラム（Epoch Life Science）によって精製し、20μLのCutSmart緩衝液（New England Biolabs）によって溶出した。USER酵素（1U，New England Biolabs）を精製された編集されたdsDNAに追加し、37℃で1hに渡ってインキュベーションした。5μLの反応溶液を15μLのDMSOに基づくローディング緩衝液（5mM Tris，0.5mM EDTA，12.5%グリセロール，0.02%ブ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標識鎖をその相補物からすっかり変性させた。全長のC含有基質を10%TBE-尿素ゲル（Bio-Rad）によっていずれかの切断されたUを含有する編集された基質から分離し、GE Amersham Typhoonイメージャーによってイメージングした。
【０４３１】
高スループットシーケンシング（HTS）のためのin vitro編集されたdsDNAの調製。
補足的な配列にリスト化されているオリゴヌクレオチドをIDTから得た。相補的配列同士をTris緩衝液中で組み合わせ（5μLの100μM溶液）、5minに渡る95℃への加熱、次に0.1℃/sの速度における45℃への徐々の冷却によってアニーリングして、60bp dsDNA基質を創出した。精製された融合蛋白質（20μLの活性緩衝液中の1.9μM）を1当量の適切なsgRNAと組み合わせ、常温で5minに渡ってインキュベーションした。60mer dsDNA基質を125nMの終濃度まで追加し、もたらされた溶液を37℃で2hに渡ってインキュベートした。dsDNAをPB緩衝液（100μL，Qiagen）およびイソプロパノール（25μL）の追加によって融合体から分離し、EconoSpinマイクロスピンカラム（Epoch Life Science）によって精製し、20μLのTris緩衝液によって溶出した。もたらされた編集されたDNA（1μLを鋳型として用いた）は、補足的な配列において指定されているHTSプライマーペアおよびVeraSeq Ultra（Enzymatics）を製造者の説明書に従って用いて、13サイクルの増幅によって、PCRによって増幅した。PCR反応生成物はRapidTips（Diffinity Genomics）を用いて精製し、精製されたDNAはシーケンシングアダプターを含有するプライマーによるPCRによって増幅し、精製し、MiSeq高スループットDNAシーケンサー（Illumina）によって先に記載されている通りシーケンシングした[73]。
【０４３２】
細胞培養。
HEK293T（ATCC CRL-3216）、U2OS（ATCC-HTB-96）、およびST486細胞（ATCC）は10%（v/v）ウシ胎児(fetal bovine)血清（FBS）およびペニシリン／ストレプトマイシン（1×，Amresco）を添加したダルベッコ改変イーグル培地プラスGlutaMax（Thermo Fisher）によって37℃において5%CO_2で維持した。HCC195細胞（ATCC CRL-2338）は上に記載される通り添加したRPMI-1640培地（Thermo Fisher Scientific）によって維持した。APOE遺伝子のApoE4アイソフォームを含有する不死化ラットアストロサイト（Taconic Biosciences）は、10%（v/v）ウシ胎児血清（FBS）および200μg/mLジェネティシン（Thermo Fisher Scientific）を添加したダルベッコ改変イーグル培地プラスGlutaMax（Thermo Fisher Scientific）によって培養した。
【０４３３】
トランスフェクション。
HEK293T細胞を48ウェルコラーゲンコートBioCoatプレート（Corning）に接種し、およそ85%コンフルエンシーでトランスフェクションした。簡潔には、製造者のプロトコールに従ってウェルあたり1.5μlのLipofectamine 2000（Thermo Fisher Scientific）を用いて、750ngのNBEおよび250ngのsgRNA発現プラスミドをトランスフェクションした。製造者の説明書に従って適切なAMAXA NUCLEOFECTOR（商標）IIプログラム

を用いて、アストロサイト、U2OS、HCC1954、HEK293T、およびST486細胞をトランスフェクションした。40ngのinfrared RFP（Addgeneプラスミド45457）[74]をヌクレオフェクション溶液に追加して、これらの細胞株のヌクレオフェクション効率を査定した。アストロサイト、U2OS、およびST486細胞では、ヌクレオフェクション効率はそれぞれ25%、74%、および92%であった。HCC1954細胞では、ヌクレオフェクション効率は＜10%であった。よって、トリプシン処理後に、HCC1954細胞を40ミクロンストレーナー（Fisher Scientific）を通して濾過し、ヌクレオフェクションされたHCC1954細胞をiRFPシグナル（abs 643nm，em 670nm）を用いてBeckman Coulter MoFlo XDPセルソーターによって収集した。他の細胞はヌクレオフェクションされた細胞の濃縮なしに用いた。

【０４３４】

ゲノムDNAサンプルの高スループットDNAシーケンシング。

トランスフェクションされた細胞を3d後に収穫し、製造者の説明書に従ってAgencourt DNAdvanceゲノムDNA単離キット（Beckman Coulter）を用いてゲノムDNAを単離した。補足的な配列にリスト化されているフランキングHTSプライマーペアによるPCRによって、当該オンターゲットおよびオフターゲットゲノム領域を増幅した。PCR増幅は5ngのゲノムDNAを鋳型として用いて製造者の説明書に従ってPhusion高フィデリティDNAポリメラーゼ（Thermo Fisher）によって実施した。サイクル数は、反応が増幅の直線範囲においてストップすることを保証するように各プライマーペアについて別個に決定した（EMX1、FANCF、HEK293部位2、HEK293部位3、HEK293部位4、およびRNF2プライマーについてそれぞれ30、28、28、28、32、および32サイクル）。PCR生成物をRapidTips（Diffinity Genomics）を用いて精製した。精製されたDNAを、シーケンシングアダプターを含有するプライマーによるPCRによって増幅した。生成物をゲル精製し、QUANT-IT（商標）PicoGreen dsDNAアッセイキット（Thermo Fisher）およびKAPAライブラリー定量キット-Illumina（KAPA Biosystems）を用いて定量した。サンプルを先に記載されている通りIllumina MiSeqによってシーケンシングした[73]。

【０４３５】

データ分析。

MiSeqレポーター（Illumina）を用いて、シーケンシングリードを自動デマルチプレックスし、補足的な注によって提供されるカスタムのMatlabスクリプトによって、個々のFASTQファイルを分析した。Smith-Watermanアルゴリズムを用いて、各リードを適切な参照配列とペアワイズアライメントした。31よりも下のQスコアを有するベースコールはNによって置き換え、それゆえにヌクレオチド頻度を算出することから除外した。この処置はおよそ1,000分の1という予想MiSeqベースコールエラー率を生む。リードおよび参照配列がギャップを含有しないアライメントされた配列をアライメント表に保存し、これから塩基頻度を各座位について集計し得た。

【０４３６】

補足的な注に示されているカスタムのMatlabスクリプトによって、先に記載された基準を用いて、インデル頻度を定量した[71]。インデルが生起し得るウィンドウの両側をフランキングする2つの10bp配列との厳密なマッチについて、シーケンシングリードをスキャンした。厳密なマッチが見つからない場合には、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厳密にマッチする場合には、リードはインデルを含有しないと分類した。インデルウィンドウが参照配列よりも2塩基以上長いかまたは短い場合には、シーケンシングリードをそれぞれ挿入または欠失として分類した。

【０４３７】

本明細書において、例えば背景技術、発明の概要、発明を実施するための形態、例、および／または参照の項において言及される全ての刊行物、特許、特許出願、公開、およびデータベースエントリー（例えば配列データベースエントリー）は、各個々の刊行物、特許、特許出願、公開、およびデータベースエントリーが具体的にかつ個々に参照によって本明細書に組み込まれる場合のように、それらの全体が参照によってここに組み込まれる

。相容れないケースでは、本明細書におけるいずれかの定義を包含して本願が対照する。

【０４３８】

補足的な配列

sgRNAトランスフェクションプラスミドを創出するために用いられたプライマー。

rev_sgRNA_plasmidを全てのケースに用いた。pFYF1320プラスミドを材料と方法の項に注記されている通り鋳型として用いた。配列番号187～196がそれぞれ下に上から下まで載せられている。

【化８２】

rev_sgRNA_plasmid	GGTGTTTCGTCCTTTCCACAAG
fwd _p53 _Y163C	GCTTGCAGATGGCCATGGCGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd_p53_N239D	TGTCACACATGTAGTTGTAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd_APOE4 _C158R	GAAGCGCCTGGCAGTGTACCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd_EMX1	GAGTCCGAGCAGAAGAAGAAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd_FANCF	GGAATCCCTTCTGCAGCACCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd_HEK293_2	GAACACAAAGCATAGACTGCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd_HEK293_3	GGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd_HEK293_4	GGCACTGCGGCTGGAGGTGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd_RNF2	GTCATCTTAGTCATTACCTGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC

【０４３９】

in vitroデアミナーゼアッセイに用いられた全てのssDNA基質の配列。

配列番号197～199がそれぞれ下に上から下まで載せられている。

【化８３】

rAPOBEC1 基質	Cy3-ATTATTATTATTCCGCGGATTTATTTATTTATTTATTTATTT
hAID/pmCDA1 基質	Cy3-ATTATTATTATTAGCTATTTATTTATTTATTTATTTATTT
hAPOBEC3G 基質	Cy3-ATTATTATTATTCCCGGATTTATTTATTTATTTATTTATTT

【０４４０】

ゲルに基づくデアミナーゼアッセイのためのsgRNAのT7転写の基質としての用をなすPCR生成物を創出するために用いられたプライマー。

rev_gRNA_T7は全てのケースに用いた。pFYF1320プラスミドを材料と方法の項に注記されている通り鋳型として用いた。配列番号200～223がそれぞれ下に上から下まで載せられている。

【化８４】

rev_sgRNA_T7	AAAAAAAGCACCGACTCGGTG
fwd_sgRNA_T7_dsDNA_2	TAATACGACTCACTATAGGCCGCGGATTTATTTATTTAAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_3	TAATACGACTCACTATAGGTCCGCGGATTTATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_4	TAATACGACTCACTATAGGTTCCGCGGATTTATTTATTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_5	TAATACGACTCACTATAGGATTCCGCGGATTTATTTATTGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_6	TAATACGACTCACTATAGGTATTCCGCGGATTTATTTATGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_7	TAATACGACTCACTATAGGTTATTCCGCGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_8	TAATACGACTCACTATAGGATTATTCCGCGGATTTATTTGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_9	TAATACGACTCACTATAGGTATTATTCCGCGGATTTATTGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_10	TAATACGACTCACTATAGGATTATTATCCGCGGATTTATGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_11	TAATACGACTCACTATAGGTATTATATTCCGCGGATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_12	TAATACGACTCACTATAGGTTATTATATTCCGCGGATTTGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_13	TAATACGACTCACTATAGGATTATTATATTCCGCGGATTGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_14	TAATACGACTCACTATAGGTATTATTATATTCCGCGGATGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_15	TAATACGACTCACTATAGGATTATTATTATTACCGCGGAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_18	TAATACGACTCACTATAGGATTATTATTATTATTACCGCGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_noC	TAATACGACTCACTATAGGATATTAATTTATTTATTTAAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_APOE4_C112R	TAATACGACTCACTATAGGGGAGGACGTGCGCGGCCGCCGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_APOE4_C158R	TAATACGACTCACTATAGGGAAGCGCCTGGCAGTGTACCGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_CTNNB1_T41A	TAATACGACTCACTATAGGCTGTGGCAGTGGCACCAGAAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_HRAS_Q61R	TAATACGACTCACTATAGGCCTCCCGGCCGGCGGTATCCGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_53_Y163C	TAATACGACTCACTATAGGGCTTGCAGATGGCCATGGCGGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_53_Y236C	TAATACGACTCACTATAGGACACATGCAGTTGTAGTGGAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_53_N239D	TAATACGACTCACTATAGGTGTCACACATGTAGTTGTAGGTTTTAGAGCTAGAAATAGCA

【０４４１】

ゲルに基づくdsDNAデアミナーゼアッセイに用いられた80ヌクレオチド非標識鎖およびCy3標識ユニバーサルプライマーの配列。

配列番号224～248がそれぞれ下に上から下まで載せられている。

【化８５】

Cy3-プライマー	Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTA
dsDNA_2	GTCCATGGATCCAGAGGTCATCCATTAAATAAATAAATCCGCGGGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_3	GTCCATGGATCCAGAGGTCATCCATAAATAAATAAATCCGCGGAAGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_4	GTCCATGGATCCAGAGGTCATCCATAATAAATAAATCCGCGGAAGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_5	GTCCATGGATCCAGAGGTCATCCAAATAAATAAATCCGCGGAATGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_6	GTCCATGGATCCAGAGGTCATCCAATAAATAAATCCGCGGAATAGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_7	GTCCATGGATCCAGAGGTCATCCATAAATAAATCCGCGGAATAAGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_8	GTCCATGGATCCAGAGGTCATCCAAAATAAATCCGCGGAATAATGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_9	GTCCATGGATCCAGAGGTCATCCAAATAAATCCGCGGAATAATAGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_10	GTCCATGGATCCAGAGGTCATCCAATAAATCCGCGGATAATAATGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_11	GTCCATGGATCCAGAGGTCATCCATAAATCCGCGGAATATAATAGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_12	GTCCATGGATCCAGAGGTCATCCAAAATCCGCGGAATATAATAAGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_13	GTCCATGGATCCAGAGGTCATCCAAATCCGCGGAATATAATAATGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_14	GTCCATGGATCCAGAGGTCATCCAATCCGCGGAATATAATAATAGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_15	GTCCATGGATCCAGAGGTCATCCATCCGCGGTAATAATAATAATGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_16	GTCCATGGATCCAGAGGTCATCCAGCGGTAATAATAATAATAATGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_noC	GTCCATGGATCCAGAGGTCATCCATTAAATAAATAAATTAATATTACTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_8U	5Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTGTAGATTATTATCUGCGGATTTATTGGATGACCTCTGGATCCATGGACAT
dsDNA_APOE_C112R	GCACCTCGCCGCGGTACTGCACCAGGCGGCCGCGCACGTCCTCCATGTCTACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_APOE_C158R	CGGCGCCCTCGCGGGCCCCGGCCTGGTACACTGCCAGGCGCTTCTGCAGTACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_CTNNB1_T41A	GTCTTACCTGGACTCTGGAATCCATTCTGGTGCCACTGCCACAGCTCCTTACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_HRAS_Q61R	GGAGACGTGCCTGTTGGACATCCTGGATACCGCCGGCCGGGAGGAGTACTACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_p53_Y163C	ACCCCCGCCCGGCACCCGCGTCCGCGCCATGGCCATCTGCAAGCAGTCATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_p53_Y236C	AGGTTGGCTCTGACTGTACCACCATCCACTACAACTGCATGTGTAACAGTACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_p53_N239D	TGGCTCTGACTGTACCACCATCCACTACAACTACATGTGTGACAGTTCCTACCAACCTTCCATTCATCCTAACTACCTAC

【０４４２】

高スループットシーケンシングのためのsgRNAのT7転写のための基質としての用をなすPCR生成物を創出するために用いられたプライマー。

rev_gRNA_T7（上）を全てのケースに用いた。pFYF1320プラスミドを材料と方法の項に注記されている通り鋳型として用いた。配列番号249～300がそれぞれ下に上から下まで載せられている。

【化８６】

fwd_sgRNA_T7_HTS_ base	TAATACGACTCACTATAGGTTATTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 1A	TAATACGACTCACTATAGGATATTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 1C	TAATACGACTCACTATAGGCTATTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 1G	TAATACGACTCACTATAGGGTATTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 2A	TAATACGACTCACTATAGGTAATTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 2C	TAATACGACTCACTATAGGTCATTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 2G	TAATACGACTCACTATAGGTGATTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 3T	TAATACGACTCACTATAGGTTTTTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 3C	TAATACGACTCACTATAGGTTCTTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 3G	TAATACGACTCACTATAGGTTGTTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 4A	TAATACGACTCACTATAGGTTAATTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 4C	TAATACGACTCACTATAGGTTACTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 4G	TAATACGACTCACTATAGGTTAGTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 5A	TAATACGACTCACTATAGGTTATATCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 5C	TAATACGACTCACTATAGGTTATCTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 5G	TAATACGACTCACTATAGGTTATGTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 6A	TAATACGACTCACTATAGGTTATTACGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 6C	TAATACGACTCACTATAGGTTATTCCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 6G	TAATACGACTCACTATAGGTTATTGCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 8A	TAATACGACTCACTATAGGTTATTTCATGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 8T	TAATACGACTCACTATAGGTTATTTCTTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 8C	TAATACGACTCACTATAGGTTATTTCCTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 9A	TAATACGACTCACTATAGGTTATTTCGAGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 9C	TAATACGACTCACTATAGGTTATTTCGCGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 9G	TAATACGACTCACTATAGGTTATTTCGGGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 10A	TAATACGACTCACTATAGGTTATTTCGTAGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 10T	TAATACGACTCACTATAGGTTATTTCGTTGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 10C	TAATACGACTCACTATAGGTTATTTCGTCGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 11A	TAATACGACTCACTATAGGTTATTTCGTGAATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 11T	TAATACGACTCACTATAGGTTATTTCGTGTATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 11C	TAATACGACTCACTATAGGTTATTTCGTGCATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 12T	TAATACGACTCACTATAGGTTATTTCGTGGTTTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 12C	TAATACGACTCACTATAGGTTATTTCGTGGCTTTATTTAGTTTTAGAGCTAGAAATAGCA

【化８７】

fwd_sgRNA_T7_HTS_ 12G	TAATACGACTCACTATAGGTTATTTCGTGGGTTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 13A	TAATACGACTCACTATAGGTTATTTCGTGGAATTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 13C	TAATACGACTCACTATAGGTTATTTCGTGGACTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 13G	TAATACGACTCACTATAGGTTATTTCGTGGAGTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ multiC	TAATACGACTCACTATAGGTTCCCCCCCCGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ TCGCACCC_odd	TAATACGACTCACTATAGGCGCACCCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ CCTCGCAC_odd	TAATACGACTCACTATAGGCTCGCACGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ ACCCTCGC_odd	TAATACGACTCACTATAGGCCCTCGCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ GCACCCTC_odd	TAATACGACTCACTATAGGCACCCTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ TCGCACCC_even	TAATACGACTCACTATAGGTCGCACCCGTGGATTTATTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ CCTCGCAC_even	TAATACGACTCACTATAGGCCTCGCACGTGGATTTATTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ ACCCTCGC_even	TAATACGACTCACTATAGGACCCTCGCGTGGATTTATTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ GCACCCTC_even	TAATACGACTCACTATAGGGCACCCTCGTGGATTTATTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ EMX1	TAATACGACTCACTATAGGGAGTCCGAGCAGAAGAAGAAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ FANCF	TAATACGACTCACTATAGGGGAATCCCTTCTGCAGCACCGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ HEK293_site2	TAATACGACTCACTATAGGGAACACAAAGCATAGACTGCGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ HEK293_site3	TAATACGACTCACTATAGGGGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ HEK293_site4	TAATACGACTCACTATAGGGGCACTGCGGCTGGAGGTGGGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ RNF2	TAATACGACTCACTATAGGGTCATCTTAGTCATTACCTGGTTTTAGAGCTAGAAATAGCA

【０４４３】

高スループットシーケンシング（HTS）のためのin vitro編集されたdsDNAの配列。

編集された鎖の配列が示されている。示されている全ての配列の逆相補物もまた得られた。材料と方法に記載されている通り相補鎖同士をアニーリングすることによって、dsDNA基質を得た。EMX1、FANCF、HEK293部位2、HEK293部位3、HEK293部位4、およびRNF2座位に相当するオリゴヌクレオチドは、元々はゲルに基づくデアミナーゼアッセイへの使用のために設計されており、よって、それらの5'末端に同じ25nt配列を有する（Cy3プライマーのものにマッチする）。配列番号301～352がそれぞれ下に上から下まで載せられている。

【化８８】

塩基配列	ACGTAAACGGCCACAAGTTCTTATTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
1A	ACGTAAACGGCCACAAGTTCATATTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
1C	ACGTAAACGGCCACAAGTTCCTATTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
1G	ACGTAAACGGCCACAAGTTCGTATTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
2A	ACGTAAACGGCCACAAGTTCTAATTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
2C	ACGTAAACGGCCACAAGTTCTCATTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
2G	ACGTAAACGGCCACAAGTTCTGATTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
3T	ACGTAAACGGCCACAAGTTCTTTTTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
3C	ACGTAAACGGCCACAAGTTCTTCTTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
3G	ACGTAAACGGCCACAAGTTCTTGTTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
4A	ACGTAAACGGCCACAAGTTCTTAATTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
4C	ACGTAAACGGCCACAAGTTCTTACTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
4G	ACGTAAACGGCCACAAGTTCTTAGTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
5A	ACGTAAACGGCCACAAGTTCTTATATCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
5C	ACGTAAACGGCCACAAGTTCTTATCTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
5G	ACGTAAACGGCCACAAGTTCTTATGTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
6A	ACGTAAACGGCCACAAGTTCTTATTACGTGGATTTATTTATGGCATCTTCTTCAAGGACG
6C	ACGTAAACGGCCACAAGTTCTTATTCCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
6G	ACGTAAACGGCCACAAGTTCTTATTGCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
8A	ACGTAAACGGCCACAAGTTCTTATTTCATGGATTTATTTATGGCATCTTCTTCAAGGACG
8T	ACGTAAACGGCCACAAGTTCTTATTTCTTGGATTTATTTATGGCATCTTCTTCAAGGACG
8C	ACGTAAACGGCCACAAGTTCTTATTTCCTGGATTTATTTATGGCATCTTCTTCAAGGACG
9A	ACGTAAACGGCCACAAGTTCTTATTTCGAGGATTTATTTATGGCATCTTCTTCAAGGACG
9C	ACGTAAACGGCCACAAGTTCTTATTTCGCGGATTTATTTATGGCATCTTCTTCAAGGACG
9G	ACGTAAACGGCCACAAGTTCTTATTTCGGGGATTTATTTATGGCATCTTCTTCAAGGACG
10A	ACGTAAACGGCCACAAGTTCTTATTTCGTAGATTTATTTATGGCATCTTCTTCAAGGACG
10T	ACGTAAACGGCCACAAGTTCTTATTTCGTTGATTTATTTATGGCATCTTCTTCAAGGACG
10C	ACGTAAACGGCCACAAGTTCTTATTTCGTCGATTTATTTATGGCATCTTCTTCAAGGACG
11A	ACGTAAACGGCCACAAGTTCTTATTTCGTGAATTTATTTATGGCATCTTCTTCAAGGACG
11T	ACGTAAACGGCCACAAGTTCTTATTTCGTGTATTTATTTATGGCATCTTCTTCAAGGACG
11C	ACGTAAACGGCCACAAGTTCTTATTTCGTGCATTTATTTATGGCATCTTCTTCAAGGACG

【化８９】

12T	ACGTAAACGGCCACAAGTTCTTATTTCGTGGTTTTATTTATGGCATCTTCTTCAAGGACG
12C	ACGTAAACGGCCACAAGTTCTTATTTCGTGGCTTTATTTATGGCATCTTCTTCAAGGACG
12G	ACGTAAACGGCCACAAGTTCTTATTTCGTGGGTTTATTTATGGCATCTTCTTCAAGGACG
13A	ACGTAAACGGCCACAAGTTCTTATTTCGTGGAATTATTTATGGCATCTTCTTCAAGGACG
13C	ACGTAAACGGCCACAAGTTCTTATTTCGTGGACTTATTTATGGCATCTTCTTCAAGGACG
13G	ACGTAAACGGCCACAAGTTCTTATTTCGTGGAGTTATTTATGGCATCTTCTTCAAGGACG
multiC	ACGTAAACGGCCACAAGTTCTTCCCCCCCCGATTTATTTATGGCATCTTCTTCAAGGACG
TCGCACCC_odd	ACGTAAACGGCCACAAGTTTCGCACCCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
CCTCGCAC_odd	ACGTAAACGGCCACAAGTTCCTCGCACGTGGATTTATTTATGGCATCTTCTTCAAGGACG
ACCCTCGC_odd	ACGTAAACGGCCACAAGTTACCCTCGCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
GCACCCTC_odd	ACGTAAACGGCCACAAGTTGCACCCTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
TCGCACCC_even	ACGTAAACGGCCACAAGTATTCGCACCCGTGGATTTATTATGGCATCTTCTTCAAGGACG
CCTCGCAC_even	ACGTAAACGGCCACAAGTATCCTCGCACGTGGATTTATTATGGCATCTTCTTCAAGGACG
ACCCTCGC_even	ACGTAAACGGCCACAAGTATACCCTCGCGTGGATTTATTATGGCATCTTCTTCAAGGACG
GCACCCTC_even	ACGTAAACGGCCACAAGTATGCACCCTCGTGGATTTATTATGGCATCTTCTTCAAGGACG
EMX1_invitro	GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCATCACATCAACCGGTG
FANCF_invitro	GTAGGTAGTTAGGATGAATGGAAGGTTGGTACTCATGGAATCCCTTCTGCAGCACCTGGATCGCTTTTCCGAGCTTCTGG
HEK293_site2_invitro	GTAGGTAGTTAGGATGAATGGAAGGTTGGTAAACTGGAACACAAAGCATAGACTGCGGGGCGGGCCAGCCTGAATAGCTG
HEK293_site3_invitro	GTAGGTAGTTAGGATGAATGGAAGGTTGGTACTTGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAGCCCTGCT
HEK293_site4_invitro	GTAGGTAGTTAGGATGAATGGAAGGTTGGTACCGGTGGCACTGCGGCTGGAGGTGGGGGTTAAAGCGGAGACTCTGGTGC
RNF2_invitro	GTAGGTAGTTAGGATGAATGGAAGGTTGGTATGGCAGTCATCTTAGTCATTACCTGAGGTGTTCGTTGTAACTCATATAA

【０４４４】

in vitro編集されたdsDNAのHTSのためのプライマー。

配列番号353～361がそれぞれ下に上から下まで載せられている。

【化９０】

fwd_invitro_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACGTAAACGGCCACAA
rev_invitro_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCGTCCTTGAAGAAGATGC
fwd_invitro_HEK_targets	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTAGGTAGTTAGGATGAATGGAA
rev_EMX1_invitro	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACCGGTTGATGTGATGG
rev_FANCF_invitro	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCAGAAGCTCGGAAAAGC
rev_HEK293_site2_invitro	TGGAGTTCAGACGTGTGCTCTTCCGATCTCAGCTATTCAGGCTGGC
rev_HEK293_site3_invitro	TGGAGTTCAGACGTGTGCTCTTCCGATCTAGCAGGGCTTCCTTTC
rev_HEK293_site4_invitro	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCACCAGAGTCTCCG
rev_RNF2_invitro	TGGAGTTCAGACGTGTGCTCTTCCGATCTTTATATGAGTTACAACGAACACC

【０４４５】

全ての哺乳類細胞培養実験からのオンターゲットおよびオフターゲット部位のHTSのためのプライマー。

配列番号362～469がそれぞれ下に上から下まで載せられている。

【化９１】

fwd_EMX1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCAGCTCAGCCTGAGTGTTGA
rev_EMX1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCGTGGGTTTGTGGTTGC
fwd_FANCF_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCATTGCAGAGAGGCGTATCA
rev_FANCF_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGTCCCAGGTGCTGAC
fwd_HEK293_site2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCAGCCCCATCTGTCAAACT
rev_HEK293_site2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGAATGGATTCCTTGGAAACAATGA
fwd_HEK293_site3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATGTGGGCTGCCTAGAAAGG
rev_HEK293_site3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCCAGCCAAACTTGTCAACC
fwd_HEK293_site4_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGAACCCAGGTAGCCAGAGAC
rev_HEK293_site4_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCTTTCAACCCGAACGGAG
fwd_RNF2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTCTTCTTTATTTCCAGCAATGT
rev_RNF2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTTTTCATGTTCTAAAAATGTATCCCA
fwd_p53_Y163C_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTACAGTACTCCCCTGCCCTC
rev_p53_Y163C_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCTGCTCACCATCGCTATCT
fwd_p53_N239D_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCTCATCTTGGGCCTGTGTT
rev_p53_N239D_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTAAATCGGTAAGAGGTGGGCC
fwd_APOE4_C158R_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCGGACATGGAGGACGTG
rev_APOE4_C158R_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGTTCCACCAGGGGCCC
fwd_EMX1_off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGCCCAATCATTGATGCTTTT
rev_EMX1_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTAGAAACATTTACCATAGACTATCACCT
fwd_EMX1_off2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGTAGCCTCTTTCTCAATGTGC
rev_EMX1_off2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCTTTCACAAGGATGCAGTCT
fwd_EMX1_off3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGAGCTAGACTCCGAGGGGA
rev_EMX1_off3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCTCGTCCTGCTCTCACTT
fwd_EMX1_off4_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGAGGCTGAAGAGGAAGACCA
rev_EMX1_off4_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGCCCAGCTGTGCATTCTAT

【化９２】

fwd_EMX1_off6_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCAAGAGGGCCAAGTCCTG
rev_EMX1_off6_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCAGCGAGGAGTGACAGCC
fwd_EMX1_off7_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCACTCCACCTGATCTCGGGG
rev_EMX1_off7_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCGAGGAGGGAGGGAGCAG
fwd_EMX1_off8_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACCACAAATGCCCAAGAGAC
rev_EMX1_off8_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGACACAGTCAAGGGCCGG
fwd_EMX1_off9_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCCACCTTTGAGGAGGCAAA
rev_EMX1_off9_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTTCCATCTGAGAAGAGAGTGGT
fwd_EMX1_off10_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTCATACCTTGGCCCTTCCT
rev_EMX1_off10_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCCTAGGCCCACACCAG
fwd_FANCF_off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAACCCACTGAAGAAGCAGGG
rev_FANCF_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGTGCTTAATCCGGCTCCAT
fwd_FANCF_off2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCAGTGTTTCCATCCCGAA
rev_FANCF_off2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCTGACCTCCACAACTCT
fwd_FANCF_off3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTGGGTACAGTTCTGCGTGT
rev_FANCF_off3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCACTCTGAGCATCGCCAAG
fwd_FANCF_off4_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGTTTAGAGCCAGTGAACTAGAG
rev_FANCF_off4_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCAAGACAAAATCCTCTTTATACTTTG
fwd_FANCF_off5_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGGAGGGGACGGCCTTAC
rev_FANCF_off5_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCCTCTGGCGAACATGGC
fwd_FANCF_off6_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCTGGTTAAGAGCATGGGC
rev_FANCF_off6_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGATTGAGTCCCCACAGCACA
fwd_FANCF_off7_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCAGTGTTTCCCATCCCCAA
rev_FANCF_off7_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGACCTCCACAACTGGAAAAT
fwd_FANCF_off8_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCTTCCAGACCCACCTGAAG
rev_FANCF_off8_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTACCGAGGAAAATTGCTTGTCG
fwd_HEK293_site2_off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTGTGGAGAGTGAGTAAGCCA
rev_HEK293_site2_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTACGGTAGGATGATTTCAGGCA
fwd_HEK293_site2_off2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCACAAAGCAGTGTAGCTCAGG
rev_HEK293_site2_off2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTTTTTTGGTACTCGAGTGTTATTCAG
fwd_HEK293_site3_off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCCCTGTTGACCTGGAGAA
rev_HEK293_site3_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACTGTACTTGCCCTGACCA
fwd_HEK293_site3_off2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTGGTGTTGACAGGGAGCAA
rev_HEK293_site3_off2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGAGATGTGGGCAGAAGGG
fwd_HEK293_site3_off3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGAGAGGGAACAGAAGGGCT
rev_HEK293_site3_off3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCCAAAGGCCCAAGAACCT
fwd_HEK293_site3_off4_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCTAGCACTTTGGAAGGTCG
rev_HEK293_site3_off4_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCTCATCTTAATCTGCTCAGCC
fwd_HEK293_site3_off5_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAAAGGAGCAGCTCTTCCTGG
rev_HEK293_site3_off5_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCTGCACCATCTCCCACAA

【化９３】

fwd_HEK293_site4_off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGCATGGCTTCTGAGACTCA
rev_HEK293_site4_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCTCCCTTGCACTCCCTGTCTTT
fwd_HEK293_site4_off2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTTGGCAATGGAGGCATTGG
rev_HEK293_site4_off2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGAAGAGGCTGCCCATGAGAG
fwd_HEK293_site4_off3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGTCTGAGGCTCGAATCCTG
rev_HEK293_site4_off3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGTGGCCTCCATATCCCTG
fwd_HEK293_site4_off4_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTTCCACCAGAACTCAGCCC
rev_HEK293_site4_off4_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCGGTTCCTCCACAACAC
fwd_HEK293_site4_off5_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCACGGGAAGGACAGGAGAAC
rev_HEK293_site4_off5_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCAGGGGAGGGATAAAGCAG
fwd_HEK293_site4_off6_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCACGGGAGATGGCTTATGT
rev_HEK293_site4_off6_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACATCCTCACTGTGCCACT
fwd_HEK293_site4_off7_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTCAGTCTCGGCCCCTCA
rev_HEK293_site4_off7_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCCACTGTAAAGCTCTTGGG
fwd_HEK293_site4_off8_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGGGTAGAGGGACAGAGCTG
rev_HEK293_site4_off8_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGACCCCACATAGTCAGTGC
fwd_HEK293_site4_off9_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCTGTCAGCCCTATCTCCATC
rev_HEK293_site4_off9_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGGGCAATTAGGACAGGGAC
fwd_HEK293_site4_off10_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCAGCGGAGGAGGTAGATTG
rev_HEK293_site4_off10_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCAGTACCTGGAGTCCCGA
fwd_HEK2_ChIP_off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGACAGGCTCAGGAAAGCTGT
rev_HEK2_ChIP_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTACACAAGCCTTTCTCCAGGG
fwd_HEK2_ChIP_off2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAATAGGGGGTGAGACTGGGG
rev_HEK2_ChIP_off2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCCTCAGACGAGACTTGAGG
fwd_HEK2_ChIP_off3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGCCAGCAGGAAAGGAATCT
rev_HEK2_ChIP_off3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGACTGCACCTGTAGCCATG
fwd_HEK2_ChIP_off4_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCAAGGAAATCACCCTGCCC
rev_HEK2_ChIP_off4_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTAACTTCCTTGGTGTGCAGCT
fwd_HEK2_ChIP_off5_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATGGGCTCAGCTACGTCATG
rev_HEK2_ChIP_off5_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTAATAGCAGTGTGGTGGGCAA
fwd_HEK3_ChIP_off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCGCACATCCCTTGTCTCTCT
rev_HEK3_ChIP_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTACTGGAGCACACCCCAAG
fwd_HEK3_ChIP_off2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGGGTCACGTAGCTTTGGTC
rev_HEK3_ChIP_off2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGGTGGCCATGTGCAACTAA
fwd_HEK3_ChIP_off3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTACTACGTGCCAGCTCAGG
rev_HEK3_ChIP_off3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTACCTCCCCTCCTCACTAACC
fwd_HEK3_ChIP_off4_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCCTCAGCTCCATTTCCTGT
rev_HEK3_ChIP_off4_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTAACCTTTATGGCACCAGGGG
fwd_HEK3_ChIP_off5_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGAGCTCAGCATTAGCAGGCT
rev_HEK3_ChIP_off5_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTTCCTGGCTTTCCGATTCCC

【化９４】

fwd_HEK4_ChIP_off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTGCAATTGGAGGAGGAGCT
rev_HEK4_ChIP_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACCAGCTACAGGCAGAACA
fwd_HEK4_ChIP_off3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCTACCCCCAACACAGATGG
rev_HEK4_ChIP_off3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCACACAACTCAGGTCCTCC

【０４４６】

HDR研究に用いられた一本鎖オリゴヌクレオチドドナー鋳型（ssODN）の配列。

EMZ1センス（配列番号470）

【化９５】

TCATCTGTGCCCCTCCCTCCCTGGCCCAGGTGAAGGTGTGGTTCCAGAACCGGAGGACAAAGTACA
AACGGCAGAAGCTGGAGGAGGAAGGGCCTGAGTTTGAGCAGAAGAAGAAGGGCTCCCATCACATC
AACCGGTGGCGCATTGCCACGAAGCAGGCCAATGGGGAGGACATCGATGTCACCTCCAATGACTAG
GGT

EMZ1アンチセンス（配列番号471）

【化９６】

ACCCTAGTCATTGGAGGTGACATCGATGTCCTCCCCATTGGCCTGCTTCGTGGCAATGCGCCACCG
GTTGATGTGATGGGAGCCCTTCTTCTTCTGCTCAAACTCAGGCCCTTCCTCCTCCAGCTTCTGCCGT
TTGTACTTTGTCCTCCGGTTCTGGAACCACACCTTCACCTGGGCCAGGGAGGGAGGGGCACAGATG
A

HEK293部位3センス（配列番号472）

【化９７】

CATGCAATTAGTCTATTTCTGCTGCAAGTAAGCATGCATTTGTAGGCTTGATGCTTTTTTTCTGCTTCT
CCAGCCCTGGCCTGGGTCAATCCTTGGGGCTTAGACTGAGCACGTGATGGCAGAGGAAAGGAAGC
CCTGCTTCCTCCAGAGGGCGTCGCAGGACAGCTTTTCCTAGACAGGGGCTAGTATGTGCAGCTCCT

HEK293部位3アンチセンス（配列番号473）

【化９８】

AGGAGCTGCACATACTAGCCCCTGTCTAGGAAAAGCTGTCCTGCGACGCCCTCTGGAGGAAGCAGG
GCTTCCTTTCCTCTGCCATCACGTGCTCAGTCTAAGCCCCAAGGATTGACCCAGGCCAGGGCTGGA
GAAGCAGAAAAAAAGCATCAAGCCTACAAATGCATGCTTACTTGCAGCAGAAATAGACTAATTGCATG

HEK293部位4センス（配列番号474）

【化９９】

GGCTGACAAAGGCCGGGCTGGGTGGAAGGAAGGGAGGAAGGGCGAGGCAGAGGGTCCAAAGCAG
GATGACAGGCAGGGGCACCGCGGCGCCCCGGTGGCATTGCGGCTGGAGGTGGGGGTTAAAGCGG
AGACTCTGGTGCTGTGTGACTACAGTGGGGGCCCTGCCCTCTCTGAGCCCCCGCCTCCAGGCCTGT
GTGTGT

HEK293部位4アンチセンス（配列番号475）

【化１００】

ACACACACAGGCCTGGAGGCGGGGGCTCAGAGAGGGCAGGGCCCCCACTGTAGTCACACAGCACC
AGAGTCTCCGCTTTAACCCCCACCTCCAGCCGCAATGCCACCGGGGCGCCGCGGTGCCCCTGCCT
GTCATCCTGCTTTGGACCCTCTGCCTCGCCCTTCCTCCCTTCCTTCCACCCAGCCCGGCCTTTGTCA
GCC

APO4Eセンス（配列番号476）

【化１０１】

AGCACCGAGGAGCTGCGGGTGCGCCTCGCCTCCCACCTGCGCAAGCTGCGTAAGCGGCTCCTCCG
CGATGCCGATGACCTGCAGAAGTGCCTGGCAGTGTACCAGGCCGGGGCCCGCGAGGGCGCCGAG
CGCGGCCTCAGCGCCATCCGCGAGCGCCTGGGGCCCCTGGTGGAACAGGGCCGCGTGCGGGCCG
CCACTGT

APO4Eアンチセンス（配列番号477）
【化１０２】

ACAGTGGCGGCCCGCACGCGGCCCTGTTCCACCAGGGGCCCCAGGCGCTCGCGGATGGCGCTGA
GGCCGCGCTCGGCGCCCTCGCGGGCCCCGGCCTGGTACACTGCCAGGCACTTCTGCAGGTCATCG
GCATCGCGGAGGAGCCGCTTACGCAGCTTGCGCAGGTGGGAGGCGAGGCGCACCCGCAGCTCCT
CGGTGCT

p53 Y163Cセンス（配列番号478）
【化１０３】

ACTCCCCTGCCCTCAACAAGATGTTTTGCCAACTGGCCAAGACCTGCCCTGTGCAGCTGTGGGTTGA
TTCCACACCCCCGCCCGGCACCCGCGTCCGCGCCATGGCCATCTACAAGCAGTCACAGCACATGAC
GGAGGTTGTGAGGCGCTGCCCCCACCATGAGCGCTGCTCAGATAGCGATGGTGAGCAGCTGGGGC
TG

p53 Y163Cアンチセンス（配列番号479）
【化１０４】

CAGCCCCAGCTGCTCACCATCGCTATCTGAGCAGCGCTCATGGTGGGGGCAGCGCCTCACAACCTC
CGTCATGTGCTGTGACTGCTTGTAGATGGCCATGGCGCGGACGCGGGTGCCGGGCGGGGGTGTGG
AATCAACCCACAGCTGCACAGGGCAGGTCTTGGCCAGTTGGCAAAACATCTTGTTGAGGGCAGGGG
AGT

【０４４７】
デアミナーゼ遺伝子gBlocks遺伝子断片
hAID（配列番号169）
rAPOBEC1（哺乳動物）（配列番号170）
【化１０５】

CATCCTTGGTACCGAGCTCGGATCCAGCCACCATGAGCTCAGAGACTGGCCCAGTGGCTGTGGACC
CCACATTGAGACGGCGGATCGAGCCCCATGAGTTTGAGGTATTCTTCGATCCGAGAGAGCTCCGCA
AGGAGACCTGCCTGCTTTACGAAATTAATTGGGGGGGCCGGCACTCCATTTGGCGACATACATCACA
GAACACTAACAAGCACGTCGAAGTCAACTTCATCGAGAAGTTCACGACAGAAAGATATTTCTGTCCG
AACACAAGGTGCAGCATTACCTGGTTTCTCAGCTGGAGCCCATGCGGCGAATGTAGTAGGGCCATC
ACTGAATTCCTGTCAAGGTATCCCCACGTCACTCTGTTTATTTACATCGCAAGGCTGTACCACCACGC
TGACCCCCGCAATCGACAAGGCCTGCGGGATTTGATCTCTTCAGGTGTGACTATCCAAATTATGACT
GAGCAGGAGTCAGGATACTGCTGGAGAAACTTTGTGAATTATAGCCCGAGTAATGAAGCCCACTGG
CCTAGGTATCCCCATCTGTGGGTACGACTGTACGTTCTTGAACTGTACTGCATCATACTGGGCCTGC
CTCCTTGTCTCAACATTCTGAGAAGGAAGCAGCCACAGCTGACATTCTTTACCATCGCTCTTCAGTCT
TGTCATTACCAGCGACTGCCCCCACACATTCTCTGGGCCACCGGGTTGAAATGAGCGGCCGCTCGA
TTGGTTTGGTGTGGCTCTAA

pmCDA1（配列番号171）
【化１０６】

CATCCTTGGTACCGAGCTCGGATCCAGCCACCATGACAGACGCTGAATATGTTAGGATCCATGAAAA
ACTGGATATCTATACATTTAAGAAGCAGTTCTTCAATAACAAAAAGTCAGTATCTCACAGATGCTATGT
CCTGTTCGAACTCAAGAGAAGAGGAGAAAGGCGGGCCTGTTTCTGGGGGTACGCGGTTAATAAACC
CCAGTCCGGGACCGAGAGGGGGATTCACGCCGAGATCTTTTCAATTAGGAAGGTTGAAGAGTATCT
TCGCGACAATCCCGGTCAGTTCACAATTAACTGGTACAGCTCCTGGAGCCCTTGCGCTGATTGCGCC
GAGAAAATACTCGAATGGTACAACCAGGAGTTGAGAGGCAATGGCCACACTCTCAAGATTTGGGCTT
GCAAGCTTTACTACGAGAAGAACGCGAGAAATCAGATTGGCTTGTGGAACCTCAGGGACAACGGGG
TCGGGTTGAATGTTATGGTGTCCGAACATTACCAGTGCTGTAGAAAGATCTTCATTCAGTCCAGTCAC
AATCAGCTGAACGAGAACAGATGGCTGGAGAAAACACTGAAACGGGCAGAGAAAAGGCGCTCAGAG
CTGAGTATCATGATCCAGGTCAAAATCCTGCATACAACCAAAAGCCCGGCTGTATAAGCGGCCGCTC
GATTGGTTTGGTGTGGCTCTAA

haPOBEC3G（配列番号172）

【化１０７】

CATCCTTGGTACCGAGCTCGGATCCAGCCACCATGGAGCTGAAGTATCACCCTGAGATGCGGTTTTT
CCACTGGTTTAGTAAGTGGCGCAAACTTCATCGGGATCAGGAGTATGAAGTGACCTGGTATATCTCT
TGGTCTCCCTGCACAAAATGTACACGCGACATGGCCACATTTCTGGCCGAGGATCCAAAGGTGACG
CTCACAATCTTTGTGGCCCGCCTGTATTATTTCTGGGACCCGGATTATCAGGAGGCACTTAGGTCAT
TGTGCCAAAAGCGCGACGGACCACGGGCGACTATGAAAATCATGAATTATGACGAATTCCAGCATTG
CTGGAGTAAGTTTGTGTACAGCCAGCGGGAGCTGTTCGAGCCCTGGAACAATCTTCCCAAGTACTAC
ATACTGCTTCACATTATGTTGGGGGAGATCCTTCGGCACTCTATGGATCCTCCTACCTTTACGTTTAA
CTTTAATAATGAGCCTTGGGTTCGCGGGCGCCATGAAACCTATTTGTGCTACGAGGTCGAGCGGATG
CATAATGATACGTGGGTCCTGCTGAATCAGAGGAGGGGGTTTCTGTGTAACCAGGCTCCACATAAAC
ATGGATTTCTCGAGGGGCGGCACGCCGAACTGTGTTTCCTTGATGTGATACCTTTCTGGAAGCTCGA
CCTTGATCAAGATTACAGGGTGACGTGTTTCACCTCCTGGTCACCCTGCTTCAGTTGCGCCCAAGAG
ATGGCTAAATTTATCAGTAAGAACAAGCATGTGTCCCTCTGTATTTTTACAGCCAGAATTTATGATGAC
CAGGGCCGGTGCCAGGAGGGGCTGCGGACACTCGCTGAGGCGGGCGCGAAGATCAGCATAATGA
CATACTCCGAATTCAAACACTGTTGGGACACTTTTGTGGACCACCAGGGCTGCCCATTTCAGCCGTG
GGATGGGCTCGACGAACATAGTCAGGATCTCTCAGGCCGGCTGCGAGCCATATTGCAGAACCAGGA
GAATTAGGCGGCCGCTCGATTGGTTTGGTGTGGCTCTAA

rAPOBEC1（E. coli）（配列番号173）

【化１０８】

GGCCGGGGATTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGATGTCTTCTGAAA
CCGGTCCGGTTGCGGTTGACCCGACCCTGCGTCGTCGTATCGAACCGCACGAATTCGAAGTTTTCT
TCGACCCGCGTGAACTGCGTAAAGAAACCTGCCTGCTGTACGAAATCAACTGGGGTGGTCGTCACT
CTATCTGGCGTCACACCTCTCAGAACACCAACAAACACGTTGAAGTTAACTTCATCGAAAAATTCACC
ACCGAACGTTACTTCTGCCCGAACACCCGTTGCTCTATCACCTGGTTCCTGTCTTGGTCTCCGTGCG
GTGAATGCTCTCGTGCGATCACCGAATTCCTGTCTCGTTACCCGCACGTTACCCTGTTCATCTACATC
GCGCGTCTGTACCACCACGCGGACCCGCGTAACCGTCAGGGTCTGCGTGACCTGATCTCTTCTGGT
GTTACCATCCAGATCATGACCGAACAGGAATCTGGTTACTGCTGGCGTAACTTCGTTAACTACTCTCC
GTCTAACGAAGCGCACTGGCCGCGTTACCCGCACCTGTGGGTTCGTCTGTACGTTCTGGAACTGTA
CTGCATCATCCTGGGTCTGCCGCCGTGCCTGAACATCCTGCGTCGTAAACAGCCGCAGCTGACCTT

CTTCACCATCGCGCTGCAGTCTTGCCACTACCAGCGTCTGCCGCCGCACATCCTGTGGGCGACCGG
TCTGAAAGGTGGTAGTGGAGGGAGCGGCGGTTCAATGGATAAGAAATAC

【０４４８】

NBE1、NBE2、およびNBE3のアミノ酸配列

E. coli発現のためのNBE1（His_6-rAPOBEC1-XTEN-dCAS9）（配列番号154）

【化１０９】

MGSSHHHHHHMSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGL
RDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQ
LTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSK
KFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRL
EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSL
GLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSA
SMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG
KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSS
FEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKL
KGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNL
GAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

哺乳動物発現のためのNBE1（rAPOBEC1-XTEN-dCas9-NLS）（配列番号155）

【化１１０】

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKF
TTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQ
IMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSC
HYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDR
HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLF
IQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDL
AEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQ
DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILE
DIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVI
EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIM
NFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKG
YKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

ヒトAPOBEC1を有する哺乳動物発現のための代替的なNBE1（hAPOBEC1-XTEN-dCas9-NLS）（配列番号158）

**MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNTTN
HVEVNFIKKFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVARLF
WHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMM
LYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPPHILLATGLIHPSVAWRG
SETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLI
GALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEE
DKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLI
EGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPG
EKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFL
AAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQ
IHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM
RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH
DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRY
TGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ
GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK
NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSD
YDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQ
RKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK
VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYK
VYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD
KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD
SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLG
APAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV**

NBE2（rAPOBEC1-XTEN-dCas9-UGI-NLS）（配列番号156）

【化１１１】

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKF
TTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQ
IMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSC
HYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDR
HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLF
IQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDL
AEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQ
DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILE
DIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVI
EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIM
NFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKG
YKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIG
NKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV

NBE3（rAPOBEC1-XTEN-Cas9n-UGI-NLS）（配列番号157）

【化１１２】

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKF
TTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQ
IMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSC
HYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDR
HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLF
IQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDL
AEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQ
DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILE
DIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVI
EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIM
NFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKG
YKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIG
NKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV

pmCDA1-XTEN-dCas9-UGI（細菌）（配列番号159）

MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQS
GTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTL
KIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWL
EKTLKRAEKRRSELSIMIQVKILHTTKSPAVSGSETPGTSESATPESDKKYSIGLAIGTNSV
GWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK
NRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE
NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAE
DAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASM
IKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK
MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKI
LTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLP
NEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQ
LKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL
FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVD
ELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQ
LQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNR
GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVE
TRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHH
AHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIM
NFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKEL
LGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADAN
LDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATL
IHQSITGLYETRIDLSQLGGDSGGSMTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKP
ESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML

pmCDA1-XTEN-nCas9-UGI-NLS（哺乳類構築物）（配列番号160）

MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQS
GTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTL
KIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWL
EKTLKRAEKRRSELSIMIQVKILHTTKSPAVSGSETPGTSESATPESDKKYSIGLAIGTNSV
GWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK
NRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH
LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE
NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAE
DAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASM
IKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK
MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKI
LTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLP
NEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQ
LKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL
FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVD
ELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQ
LQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNR
GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVE
TRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHH
AHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIM
NFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKEL
LGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADAN
LDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATL
IHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPES

DILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV

huAPOBEC3G-XTEN-dCas9-UGI（細菌）（配列番号161）

MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFL
EGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTAR
IYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLS
GRLRAILQSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFH
RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH
MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRR
LENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI
GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL
PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT
FDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWM
TRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR
FNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKED
IQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAREN
QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQ
LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE
NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE
SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETN
GETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDW
DPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGY
KEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS
PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI
IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGG
SMTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTS
DAPEYKPWALVIQDSNGENKIKML

huAPOBEC3G-XTEN-nCas9-UGI-NLS（哺乳類構築物）（配列番号162）

MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFL
EGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTAR
IYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLS
GRLRAILQSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFH
RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH
MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRR
LENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI
GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL
PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT
FDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWM
TRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR
FNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKED
IQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAREN
QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQ
LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE

NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE
SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETN
GETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDW
DPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGY
KEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS
PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI
IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGG
STNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDA
PEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV

huAPOBEC3G(D316R_D317R)-XTEN-nCas9-UGI-NLS（哺乳類構築物）（配列番号163）

MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFL
EGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTAR
IYRRQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLS
GRLRAILQSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFH
RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH
MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRR
LENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI
GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL
PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT
FDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWM
TRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR
FNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKED
IQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAREN
QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQ
LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE
NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE
SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETN
GETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDW
DPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGY
KEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS
PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI
IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGG
STNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDA
PEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV

【０４４９】

ベースコールMatlabスクリプト

【化１１３】

```
WTnuc='GCGGACATGGAGGACGTGCGCGGCCGCCTGGTGCAGTACCGCGGCGAGGTGCAGGCCATGCTCG
GCCAGA
GCACCGAGGAGCTGCGGGTGCGCCTCGCCTCCCACCTGCGCAAGCTGCGTAAGCGGCTCCTCCGCGATGC
CGATGAC
CTGCAGAAGCGCCTGGCAGTGTACCAGGCCGGGGCCCGCGAGGGCGCCGAGCGCGGCCTCAGCGCCATC
CGCGAGCG CCTGGGGCCCCTGGTGGAACAG'(SEQ ID NO: 164);
%cycle through fastq files for different samples  files=dir('*.fastq');
for d=1:20
filename=files(d).name;
%read fastq file
[header,seqs,qscore]=fastqread(filename);
seqsLength=length(seqs);                                % number of sequences  seqsFile=
strrep(filename,'.fastq','');                           % trims off .fastq
%create a directory with the same name as fastq file  ifexist(seqsFile,'dir');
error('Directory already exists. Please rename or move it before  moving on.');
end
mkdir(seqsFile);                                        % make directory
wtLength=length(WTnuc);                                     % length of wildtype sequence
%% aligning back to the wildtype nucleotide sequence
%
```

【化１１４】

```
% AlN is a matrix of the nucleotide alignment  window=1:wtLength;
sBLength=length(seqs);                                    % number of sequences
% counts number of skips  nSkips = 0;
ALN=repmat('',[sBLengthwtLength]);
% iterate through each sequencing read  for i = 1:sBLength
%If you only have forward read fastq files leave as is
%If you have R1 foward and R2 is reverse fastq files uncomment the
%next four lines of code and the subsequent end statement
%                ifmod(d,2)==0;
%                        reverse=seqrcomplement(seqs{i});
%                        [score,alignment,start]=
swalign(reverse,WTnuc,'Alphabet','NT');
%                else

[score,alignment,start]=swalign(seqs{i},WTnuc,'Alphabet','NT');
%                end

% length of the sequencing read  len=
length(alignment(3,:));
% if there is a gap in the alignment , skip = 1 and we will
% throw away the entire read  skip = 0;
for j = 1:len
if (alignment(3,j) == '-' || alignment(1,j) == '-')  skip = 1;
                        break;
end
%in addition if the qscore for any given base in the read is
%below 31 the nucleotide is turned into an N (fastq qscores that  are not letters)
ifisletter(qscore{i}(start(1)+j-1)) else
alignment(1,j) = 'N';
end

end
if skip == 0 && len>10
ALN(i,   start(2):(start(2)+length(alignment)-1))=alignment(1,:);
```

【化１１５】

```
                end
    end
```

【化１１６】

```
% with the alignment matrices we can simply tally up the occurrences of
% each nucleotide at each column in the alignment these
% tallies ignore bases annotated as N
% due to low qscores
TallyNTD=zeros(5,wtLength); fori=1:wtLength

TallyNTD(:,i)=[sum(ALN(:,i)=='A'),sum(ALN(:,i)=='C'),sum(ALN(:,i)=='G'),sum(A
LN(:,i)=='T'),sum(ALN(:,i)=='N')];
end
% we then save these tally matrices in the respective folder for
% further processing

save(strcat(seqsFile,'/TallyNTD'),'TallyNTD'); dlmwrite(strcat(seqsFile,'/TallyNTD.txt'),TallyNTD,'precision',
'%.3f','newline','pc'); end
```

【０４５０】

インデル検出Ｍａｔｌａｂスクリプト

【化１１７】

```
WTnuc='GCGGACATGGAGGACGTGCGCGGCCGCCTGGTGCAGTACCGCGGCGAGGTGCAGGCCATGCTCGGC
CAGA
GCACCGAGGAGCTGCGGGTGCGCCTCGCCTCCCACCTGCGCAAGCTGCGTAAGCGGCTCCTCCGCGATGCCG
ATGAC
CTGCAGAAGCGCCTGGCAGTGTACCAGGCCGGGGCCCGCGAGGGCGCCGAGCGCGGCCTCAGCGCCATCCGC
GAGCG CCTGGGGCCCCTGGTGGAACAG'(SEQ ID NO: 164);
%cycle through fastq files for different samples files=dir('*.fastq');
%specify start and width of indel window as well as length of each flank indelstart=154;
width=30; flank=10;

for d=1:3
filename=files(d).name;
%read fastq file
[header,seqs,qscore]=fastqread(filename);
seqsLength=length(seqs);                                    % number of sequences seqsFile
=strcat(strrep(filename,'.fastq',''),'_INDELS');
%create a directory with the same name as fastq file+_INDELS ifexist(seqsFile,'dir');
error('Directory already exists. Please rename or move it before moving on.');
end
mkdir(seqsFile);                                            % make directory
wtLength=length(WTnuc);                                     % length of wildtype sequence sBLength=
length(seqs);                                               % number of sequences

% initialize counters and cell arrays
nSkips = 0; notINDEL=0;
ins={};
dels={}; NumIns=0;
NumDels=0;
% iterate through each sequencing read for i = 1:sBLength
%search for 10BP sequences that should flank both sides of the "INDEL WINDOW"
windowstart=strfind(seqs{i},WTnuc(indelstart-flank:indelstart));

                windowend=strfind(seqs{i},WTnuc(indelstart+width:indelstart+width+flank
));
%if the flanks are found proceed
```

【化１１８】

```
iflength(windowstart)==1&&length(windowend)==1
%if the sequence length matches the INDEL window length save as
%not INDEL
if   windowend-windowstart==width+flank notINDEL=notINDEL+1;
%if the sequence is two or more bases longer than the INDEL
%window length save as an Insertion
elseif   windowend-windowstart>=width+flank+2  NumIns=NumIns+1;
ins{NumIns}=seqs{i};
%if the sequence is two or more bases shorter than the INDEL
%window length save as a Deletion
elseif   windowend-windowstart<=width+flank-2   NumDels=NumDels+1;
dels{NumDels}=seqs{i};
%keep track of skipped sequences that are either one base
%shorter or longer than the INDEL window width  else
nSkips=nSkips+1;
end
%keep track of skipped sequences that do not possess matching flank
%sequences else
nSkips=nSkips+1;

            end
end
```

【化１１９】

```
fid=fopen(strcat(seqsFile,'/summary.txt'),'wt');
fprintf(fid, 'Skipped reads %i\n not INDEL %i\n Insertions %i\n Deletions
%i\n', [nSkips, notINDEL, NumIns, NumDels]);  fclose(fid);
save(strcat(seqsFile,'/nSkips'),'nSkips');  save(strcat(seqsFile,'/notINDEL'),'notINDEL');
save(strcat(seqsFile,'/NumIns'),'NumIns');  save(strcat(seqsFile,'/NumDels'),'NumDels');
save(strcat(seqsFile,'/dels'),'dels');
C = dels;
fid = fopen(strcat(seqsFile, '/dels.txt'), 'wt');  fprintf(fid,'"%s"\n',C{:});
fclose(fid);
save(strcat(seqsFile,'/ins'),'ins');  C = ins;
fid = fopen(strcat(seqsFile, '/ins.txt'), 'wt');  fprintf(fid,'"%s"\n',C{:});
fclose(fid);

end
```

【０４５１】

例５：Cas9バリアント配列

本開示は、（例えば、dCas9またはCas9ニッカーゼを創出するための）1つ以上の変異を含み得るCas9バリアント、例えば1つ以上の生物からのCas9蛋白質を提供する。いくつかの態様においては、下でアステリスクによって同定されているCas9蛋白質のアミノ酸残基の1つ以上が変異させられ得る。いくつかの態様において、配列番号6によって提供されるアミノ酸配列のD10および／もしくはH840残基、またはいずれかのCas9蛋白質、例え

ば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異が変異させられる。いくつかの態様において、配列番号6によって提供されるアミノ酸配列のD10残基、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異が、Dを除くいずれかのアミノ酸残基に変異させられる。いくつかの態様において、配列番号6によって提供されるアミノ酸配列のD10残基、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異がAに変異させられる。いくつかの態様において、配列番号6によって提供されるアミノ酸配列のH840残基、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する残基はHである。いくつかの態様において、配列番号6によって提供されるアミノ酸配列のH840残基、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異が、Hを除くいずれかのアミノ酸残基に変異させられる。いくつかの態様において、配列番号6によって提供されるアミノ酸配列のH840残基、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する変異がAに変異させられる。いくつかの態様において、配列番号6によって提供されるアミノ酸配列のD10残基、またはいずれかのCas9蛋白質、例えば本明細書において提供されるCas9アミノ酸配列のいずれか1つの対応する残基はDである。

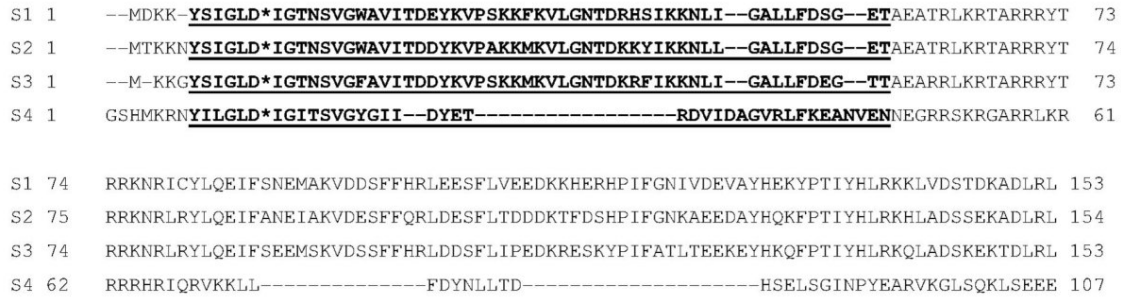
【０４５２】

種々の種からのいくつものCas9配列をアライメントして、配列番号6または配列番号567のD10およびH840の対応する相同アミノ酸残基が他のCas9蛋白質において同定され得、相同アミノ酸残基の対応する変異を有するCas9バリアントの創出を許すかどうかを決定した。アライメントは、NCBIの制約に基づくマルチプルアライメントツール（COBALT（st-va.ncbi.nlm.nih.gov/tools/cobaltにおいてアクセス可能）を用いて、次のパラメータによって実施した。アライメントパラメータ：Gap penalties -11,-1；End-Gap penalties -5,-1. CDD Parameters: Use RPS BLAST on；Blast E-value 0.003；Find Conserved Columns and Recompute on. Query Clustering Parameters: Use query clusters on；Word Size 4；Max cluster distance 0.8；Alphabet Regular。

【０４５３】

4つのCas9配列の例示的なアライメントが下で提供されている。アライメント中のCas9配列は：配列1（S1）：配列番号567 ¦ WP_010922251 ¦ gi 499224711 ¦ II型CRISPR RNAによってガイドされるエンドヌクレアーゼCas9 [Streptococcus pyogenes]；配列2（S2）：配列番号568 ¦ WP_039695303 ¦ gi 746743737 ¦ II型CRISPR RNAによってガイドされるエンドヌクレアーゼCas9 [Streptococcus gallolyticus]；配列3（S3）：配列番号569 ¦ WP_045635197 ¦ gi 782887988 ¦ II型CRISPR RNAによってガイドされるエンドヌクレアーゼCas9 [Streptococcus mitis]；配列4（S4）：配列番号570 ¦ 5AXW_A ¦ gi 924443546 ¦ Staphylococcus aureus Cas9である。HNHドメイン（太字かつ下線付き）およびRuvCドメイン（枠囲み）が4つの配列のそれぞれにおいて同定される。S1のアミノ酸残基10および840ならびにアライメントされた配列の相同アミノ酸はそれぞれのアミノ酸残基に後続するアステリスクによって同定されている。

【化１２０】


S1 1 --MDKK-YSIGLD*IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLI--GALLFDSG--ETAEATRLKRTARRYT 73
S2 1 --MTKKNYSIGLD*IGTNSVGWAVITDDYKVPAKKMKVLGNTDKKYIKKNLL--GALLFDSG--ETAEATRLKRTARRYT 74
S3 1 --M-KKGYSIGLD*IGTNSVGFAVITDDYKVPSKKMKVLGNTDKRFIKKNLI--GALLFDEG--TTAEARRLKRTARRYT 73
S4 1 GSHMKRNYILGLD*IGITSVGYGII--DYET----------RDVIDAGVRLFKEANVENNEGRRSKRGARRLKR 61
S1 74 RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRL 153
S2 75 RRKNRLRYLQEIFANEIAKVDESFFQRLDESFLTDDDKTFDSHPIFGNKAEEDAYHQKFPTIYHLRKHLADSSEKADLRL 154
S3 74 RRKNRLRYLQEIFSEEMSKVDSSFFHRLDDSFLIPEDKRESKYPIFATLTEEKEYHKQFPTIYHLRKQLADSKEKTDLRL 153
S4 62 RRRHRIQRVKKLL----------FDYNLLTD----------HSELSGINPYEARVKGLSQKLSEEE 107

【化１２１】


S1 154 IYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEK 233
S2 155 VYLALAHMIKFRGHFLIEGELNAENTDVQKIFADFVGVYNRTFDDSHLSEITVDVASILTEKISKSRRLENLIKYYPTEK 234
S3 154 IYLALAHMIKYRGHFLYEEAFDIKNNDIQKIFNEFISIYDNTFEGSSLSGQNAQVEAIFTDKISKSAKRERVLKLFPDEK 233
S4 108 FSAALLHLAKRRG----------VHNVNEVEEDT---------- 131
S1 234 KNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEIT 313
S2 235 KNTLFGNLIALALGLQPNFKTNFKLSEDAKLQFSKDTYEEDLEELLGKIGDDYADLFTSAKNLYDAILLSGILTVDDNST 314
S3 234 STGLFSEFLKLIVGNQADFKKHFDLEDKAPLQFSKDTYDEDLENLLGQIGDDFTDLFVSAKKLYDAILLSGILTVTDPST 313
S4 132 -----GNELS----------TKEQISRN---------- 144
S1 314 KAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKM--DGTEELLV 391
S2 315 KAPLSASMIKRYVEHHEDLEKLKEFIKANKSELYHDIFKDKNKNGYAGYIENGVKQDEFYKYLKNILSKIKIDGSDYFLD 394
S3 314 KAPLSASMIERYENHQNDLAALKQFIKNNLPEKYDEVFSDQSKDGYAGYIDGKTTQETFYKYIKNLLSKF--EGTDYFLD 391
S4 145 ----SKALEEKYVAELQ---------- 165
S1 392 KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEE 471
S2 395 KIEREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGDYYPFLKEKQDRIEKILTFRIPYYVGPLVRKDSRFAWAEYRSDE 474
S3 392 KIEREDFLRKQRTFDNGSIPHQIHLQEMNAILRRQGEYYPFLKDNKEKIEKILTFRIPYYVGPLARGNRDFAWLTRNSDE 471
S4 166 --EVRGSINRFKTSD--------YVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGP--GEGSPFGW------K 227
S1 472 TITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL 551
S2 475 KITPWNFDKVIDKEKSAEKFITRMTLNDLYLPEEKVLPKHSHVYETYAVYNELTKIKYVNEQGKE-SFFDSNMKQEIFDH 553
S3 472 AIRPWNFEEIVDKASSAEDFINKMTNYDLYLPEEKVLPKHSLLYETFAVYNELTKVKFIAEGLRDYQFLDSGQKKQIVNQ 551
S4 228 DIKEW----------YEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEK---LEYYEKFQIIEN 289
S1 552 LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR---FNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFED 628
S2 554 VFKENRKVTKEKLLNYLNKEFPEYRIKDLIGLDKENKSFNASLGTYHDLKKIL-DKAFLDDKVNEEVIEDIIKTLTLFED 632
S3 552 LFKENRKVTEKDIIHYLHN-VDGYDGIELKGIEKQ---FNASLSTYHDLLKIIKDKEFMDDAKNEAILENIVHTLTIFED 627
S4 290 VFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEF---TNLKVYHDIKDITARKEII---ENAELLDQIAKILTIYQS 363
S1 629 REMIEERLKTYAHLFDDKVMKQLKR-RRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKED 707
S2 633 KDMIHERLQKYSDIFTANQLKKLER-RHYTGWGRLSYKLINGIRNKENNKTILDYLIDDGSANRNFMQLINDDTLPFKQI 711
S3 628 REMIKQRLAQYDSLFDEKVIKALTR-RHYTGWGKLSAKLINGICDKQTGNTILDYLIDDGKINRNFMQLINDDGLSFKEI 706
S4 364 SEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDE------LWHTNDNQIAIFNRLKLVP---------- 428
S1 708 IQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT------QKGQKNSRERM 781
S2 712 IQKSQVVGDVDDIEAVVHDLPGSPAIKKGILQSVKIVDELVKVMG-GNPDNIVIEMARENQTT------NRGRSQSQQRL 784
S3 707 IQKAQVIGKTDDVKQVVQELSGSPAIKKGILQSIKIVDELVKVMG-HAPESIVIEMARENQTT------ARGKKNSQQRY 779
S4 429 -KKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYG--LPNDIIIELAREKNSKDAQKMINEMQKRNRQTN 505
S1 782 KRIEEGIKELGSQIL-------KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSD----YDVDH*IVPQSFLKDD 850
S2 785 KKLQNSLKELGSNILNEEKPSYIEDKVENSHLQNDQLFLYYIQNGKDMYTGDELDIDHLSD----YDIDH*IIPQAFIKDD 860
S3 780 KRIEDSLKILASGL---DSNILKENPTDNNQLQNDRLFLYYLQNGKDMYTGEALDINQLSS----YDIDH*IIPQAFIKDD 852
S4 506 ERIEEIIRTTGK----------ENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDH*IIPRSVSFDN 570
S1 851 SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDN-LTKAERGGL-SELD------KAGFIKRQLV 922

【化１２２】

```
S2 861  SIDNRVLTSSAKNRGKSDDVPSLDIVRARKAEWVRLYKSGLISKRKFDN-LTKAERGGL-TEAD------KAGFIKRQLV 932
S3 853  SLDNRVLTSSKDNRGKSDNVPSIEVVQKRKAFWQQLLDSKLISERKFNN-LTKAERGGL-DERD------KVGFIKRQLV 924
S4 571  SFNNKVLVKQEEASKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLV 650

S1 923  ETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYP 1002
S2 933  ETRQITKHVAQILDARFNTEHDENDKVIRDVKVITLKSNLVSQFRKDFEFYKVREINDYHHAHDAYLNAVVGTALLKKYP 1012
S3 925  ETRQITKHVAQILDARYNTEVNEKDKKNRTVKIITLKSNLVSNFRKEFRLYKVREINDYHHAHDAYLNAVVAKAILKKYP 1004
S4 651  DTRYATRGLMNLLRSYFRVN-------NLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIA---------- 712

S1 1003 KLESEFVYGDYKVYDVRKMIAKSEQ--EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG--- 1077
S2 1013 KLASEFVYGEYKKYDIRKFITNSSD-----KATAKYFFYSNLMNFFKTKVKYADGTVFERPIIETNAD-GEIAWNKQ--- 1083
S3 1005 KLEPEFVYGEYQKYDLKRYISRSKDPKEVEKATEKYFFYSNLLNFFKEEVHYADGTIVKRENIEYSKDTGEIAWNKE--- 1081
S4 713  --NADFIFKEWKKLDKAKKVMENQM-------------------------FEEKQAESMPEIETEQEYKEIFITPHQIK 764

S1 1078 -----RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD---WDPKKYGGFDSPTVAYSVLVVAKV 1149
S2 1084 -----IDFEKVRKVLSYPQVNIVKKVETQTGGFSKESILPKGDSDKLIPRKTKKVYWDTKKYGGFDSPTVAYSVFVVADV 1158
S3 1082 -----KDFAIIKKVLSLPQVNIVKKREVQTGGFSKESILPKGNSDKLIPRKTKDILLDTTKYGGFDSPVIAYSILLIADI 1156
S4 765  HIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKL----KKLIN-KSP----EKLLMYHH 835

S1 1150 EKGKSKKLKSVKELLGITIMERSSFEKNPI-DFLEAKG-----YKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKG 1223
S2 1159 EKGKAKKLKTVKELVGISIMERSFFEENPV-EFLENKG-----YHNIREDKLIKLPKYSLFEFEGGRRRLLASASELQKG 1232
S3 1157 EKGKAKKLKTVKTLVGITIMEKAAFEENPI-TFLENKG-----YHNVRKENILCLPKYSLFELENGRRRLLASAKELQKG 1230
S4 836  DPQTYQKLK--------LIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKV 907

S1 1224 NELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH------ 1297
S2 1233 NEMVLPGYLVELLYHAHRADNF-----NSTEYLNYVSEHKKEFEKVLSCVEDFANLYVDVEKNLSKIRAVADSM------ 1301
S3 1231 NEIVLPVYLTTLLYHSKNVHKL-----DEPGHLEYIQKHRNEFKDLLNLVSEFSQKYVLADANLEKIKSLYADN------ 1299
S4 908  VKLSLKPYRFD-VYLDNGVYKFV-----TVKNLDVIK--KENYYEVNSKAYEEAKKLKKISNQAEFIASFYNNDLIKING 979

S1 1298 RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSIT--------GLYETRI----DLSQL 1365
S2 1302 DNFSIEEISNSFINLLTLTALGAPADFNFLGEKIPRKRYTSTKECLNATLIHQSIT--------GLYETRI----DLSKL 1369
S3 1300 EQADIEILANSFINLLTFTALGAPAAFKFFGKDIDRKRYTTVSEILNATLIHQSIT--------GLYETWI----DLSKL 1367
S4 980  ELYRVIGVNNDLLNRIEVNMIDITYR-EYLENMNDKRPPRIIKTIASKT---QSIKKYSTDILGNLYEVKSKKHPQIIKK 1055

S1 1366 GGD  1368
S2 1370 GEE  1372
S3 1368 GED  1370
S4 1056 G--  1056
```

【０４５４】

アライメントは、当分野において公知のアライメントプログラムおよびアルゴリズムを用いて参照配列または参照残基とアライメントするアミノ酸配列または残基を同定することによって、参照Cas9アミノ酸配列またはアミノ酸残基に対して相同であるアミノ酸配列およびアミノ酸残基が、異なる種からのCas9配列を包含するがこれらに限定されないCas9配列バリアントにおいて同定され得ることを実証する。本開示は、配列番号567～570（例えば、それぞれS1、S2、S3、およびS4）においてアステリスクによって同定されているアミノ酸残基の1つ以上が本明細書に記載される通り変異しているCas9バリアントを提供する。配列番号567～570においてアステリスクによって同定されている残基に対応する配列番号6のCas9の残基D10およびH840は、本明細書においては「相同な」または「対応する」残基と言われる。かかる相同残基は、例えば上に記載されている通り配列アライメントによって、および参照配列または残基とアライメントする配列または残基を同定することによって、同定され得る。類似に、本明細書の配列番号6によって同定される

変異、例えば配列番号6の残基10および840の変異に対応するCas9配列の変異は、本明細書においては「相同な」または「対応する」変異と言われる。例えば、上の4つのアライメントされた配列について配列番号6またはS1（配列番号567）のD10A変異に対応する変異は、S2ではD11A、S3ではD10A、S4ではD13Aであり；配列番号6またはS1（配列番号567）のH840Aの対応する変異はS2ではH850A、S3ではH842A、S4ではH560Aである。

【０４５５】

例6：次世代のC→T編集因子

塩基編集因子3（BE3）構築物の代替としてのシチジンデアミナーゼの他のファミリーを調べた。標的化可能な基質を拡大するための狭いまたは異なる編集ウィンドウ、替わりの配列特異性を有し、より高い活性を有するように、異なるC→T編集因子を開発した。

【０４５６】

例4に記載されている方法を用いて、HeK-3部位におけるpmCDA1（Petromyzon marinusからのシチジンデアミナーゼ1）活性を評価している（図42）。pmCDA1-nCas9-UGI-NLS（nCas9は本明細書に記載されるCas9ニッカーゼを指示している）構築物は、rAPOBEC1（BE3）によってアクセス可能でないいくつかの部位（例えば、位置9、5、4、および3における相補的鎖上のC塩基）に対して活性である。

【０４５７】

HeK-2部位におけるpmCDA1活性が図43によって与えられている。pmCDA1-XTEN-nCas9-UGI-NLS構築物は「G」に隣接する部位において活性であるが、rAPOBEC1アナログ（BE3構築物）は「G」に隣接している「C」、例えば相補鎖の位置11におけるC塩基に対して低い活性を有する。

【０４５８】

Tに変換された標的C（図44）、Aに変換されたC（図45）、およびGに変換されたC（図46）を有する合計のシーケンシングリードのパーセントがCDAおよびAPOBEC1（BE3構築物）について示されている。

【０４５９】

HeK-2部位におけるhuAPOBEC3G活性が図47に示される。2つの構築物を用いた：huAPOBEC3G-XTEN-nCas9-UGI-NLSおよびhuAPOBEC3G*(D316R_D317R)-XTEN-nCas9-UGI-NLS。図47に示される通り、huAPOBEC3G-XTEN-nCas9-UGI-NLS構築物はrAPOBEC1（BE3）とは異なる配列特異性を有し、APOBEC1と比較して位置4におけるAPOBEC3Gの減少した活性によって指示される通り、編集ウィンドウは狭く見える。huAPOBEC3Gになされた変異（D316RおよびD317R）はssDNA結合を増大させ、編集される部位を拡大することへの観察可能な効果をもたらした（図47のAPOBEC3GをAPOBEC3G_RRと比較せよ）。変異はAPOBEC3G結晶構造に基づいて選んだ：Holden et al., Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implication. Nature. (2008); 121-4を見よ。これらの内容全体は参照によって本明細書に組み込まれる。

【０４６０】

例7：pmCDA1/huAPOBEC3G/rAPOBEC1はE. coliにおいて働く

Fプラスミド上にコードされたlacZを有する細菌株を用いて、A→I変換のためのLacZセレクション最適化を行った。シチジンデアミナーゼによるG→AがlacZ活性を回復させるように、決定的なグルタミン酸残基を変異させた（例えば、GAG→GGG、Glu→Gly変異）（図48）。セレクションアッセイには株CC102を選択した。APOBEC1およびCDA構築物をセレクションアッセイに用いて、G→A変換を最適化した。

【０４６１】

lacZ復帰頻度へのデアミナーゼ構築物をコードするプラスミドのコピー数の効果を評価するために、CDAおよびAPOBEC1デアミナーゼを、異なる複製起点（ゆえに異なるコピー数）SC101、CloDF3、RSF1030、およびPUC（コピー数：PUC＞RSF1030＞CloDF3＞SC101）を有する4つのプラスミドにクローニングし、誘導性プロモーター下に置いた

。変異したLacZ遺伝子を含有するFプラスミドを抱えるE. coli細胞に、プラスミドを個々に形質転換した。デアミナーゼの発現を誘導し、LacZ活性を各構築物について検出した（図49）。図49に示されている通り、デアミナーゼがクローニングされたプラスミドコピー数にかかわらず、CDAは全ての場合にAPOBEC1よりも有意に高い活性を見せた。さらに、コピー数の観点からは、デアミナーゼ活性はそれらがクローニングされたプラスミドのコピー数と正に相関し、すなわちPUC＞CloDF3＞SC101であった。

【０４６２】

lacZ座位におけるゲノムDNAのシーケンシングによって、LacZ復帰を確認した。修正されたLacZ遺伝子を含有するゲノムDNAを得るために、細胞をX-galを含有する培地で成長させた。ここでは、LacZ活性を有する細胞は青いコロニーを形成する。青いコロニーを選択し、ラクトースを含有する最少培地によって成長させた。細胞をスピンダウンし、洗浄し、最少培地プレート（ラクトース）に再度プレーティングした。それから、最も高い希釈における青いコロニーを選択し、そのゲノムDNAをlacZ座位においてシーケンシングした（図50）。

【０４６３】

異なるシチジンデアミナーゼ（例えばCDAおよびAPOBEC1）の活性を試験するために、クロラムフェニコール復帰アッセイを設計した。クロラムフェニコール耐性を細菌に授ける変異体CAT1遺伝子を抱えるプラスミドがRSF1030を複製起点として構築される。変異体CAT1遺伝子はH195R（CAC→CGC）変異を有するCAT1蛋白質をコードし、蛋白質を不活性にする（図51）。CGCコドンのG塩基と塩基対形成したCの脱アミノ化はコドンをCACコドンに再び変換し、蛋白質の活性を回復させるであろう。図52に示されている通り、CDAは、E. coliにおいて、クロラムフェニコール耐性遺伝子の活性を回復させることにおいてrAPOBECに優る。セレクションプラスミド（pNMG_ch_5）を有するS1030におけるchlorの最小阻害濃度（MIC）はおよそ1μg/mLであった。rAPOBEC-XTEN-dCas9-UGIおよびCDA-XTEN-dCas9-UGI両方はセレクションプラスミドのDNA修正を誘導した（図53）。

【０４６４】

次に、huAPOBEC3G-XTEN-dCas9-UGI蛋白質を同じアッセイによって試験した。興味深いことに、huAPOBEC3G-XTEN-dCas9-UGIはrAPOBEC1-XTEN-dCas9-UGI融合蛋白質とは異なる配列特異性を見せた。rAPOBEC11-XTEN-dCas9-UGI融合体（位置3、6、および8が編集された）と比較して、APOBEC3G-XTEN-dCas9-UGI融合体では、位置8のみが編集された（図54）。

【０４６５】

例8：より少ないオフターゲット編集を有するC→T塩基編集因子

現行の塩基編集テクノロジーは、ゲノムDNA中のC:G塩基対からT:A塩基対への配列特異的変換を許す。これはシチジンデアミナーゼ酵素によるシトシンからウラシルへの触媒的な直接変換によってされ、それゆえに、従来のゲノム編集テクノロジーとは違って、第1ステップとしてDNAに二本鎖DNA切断（DSB）を導入しない。Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016), "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage." Nature 533, 420-424を見よ；これの内容全体は参照によって本明細書に組み込まれる。代わりに、触媒的にデッドなSpCas9（dCas9）またはSpCas9ニッカーゼ（dCas9(A840H)）がrAPOBEC1、pmCDA1、またはhAPOBEC3Gなどのシチジンデアミナーゼ酵素にテザリングされる。当該ゲノム座位はsgRNAによってコードされ、DNA結合および局所的変性は融合体のdCas9部分によって容易化される。しかしながら、単にwtdCas9およびwtCas9がオフターゲットDNA結合および切断を見せる通り、現行の塩基編集因子もまたCas9オフターゲット座位においてC→T編集を見せ、これはそれらの治療上の有用性を限定する。

【０４６６】

蛋白質とその標的DNAの糖-リン酸バックボーンとの間の非特異的静電相互作用を中和するSpCas9への単に3つから4つの変異の導入が、SpCas9のDNA結合特異性を増大させる

ことが報告されている。Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016) "High-fidelity CRISPR－Cas9 nucleases with no detectable genome-wide off-target effects." Nature 529, 490-495；およびSlaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2015)"Rationally engineered Cas9 nucleases with improved specificity. Science 351, 84-88を見よ；それぞれの内容全体はここに参照によって本明細書に組み込まれる。よって、4つの報告された中和変異を最初に報告された塩基編集因子BE3（配列番号48）に組み込み、この酵素のオフターゲットC→T編集もまた劇的に縮減され（図55）、オンターゲット編集の減少を有さない（図56）ことを見出した。

【０４６７】

図55に示される通り、Lipofectamine 2000を用いて、BE3またはHF-BE3およびEMX1配列にマッチするsgRNAを発現するプラスミドによって、HEK293T細胞をトランスフェクションした。トランスフェクションの3日後に、ゲノムDNAを抽出し、PCRによって増幅し、GUIDE-seq法を用いてJoungらによって先に決定されたEMX1 sgRNAのオンターゲット座位プラス上位10個の公知のCas9オフターゲット座位における高スループットDNAシーケンシングによって分析した。Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P., et al. (2015) "GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases." Nat Biotech 33, 187-197を見よ；これらの内容全体は参照によって本明細書に組み込まれる。EMX1オフターゲット5座位は増幅せず、示されていない。オンターゲットおよびオフターゲットプロトスペーサーおよびプロトスペーサー隣接モチーフ（PAM）の配列が表示されている（図55）。プロトスペーサー内の元々のCの各位置にTを有する合計のDNAシーケンシングリードのパーセンテージとして定められる細胞性のC→T変換パーセンテージが、BE3およびHF-BE3について示されている。

【０４６８】

図56では、Lipofectamine 2000を用いて、BE3またはHF-BE3および指示されているゲノム座位にマッチするsgRNAを発現するプラスミドによって、HEK293T細胞をトランスフェクションした。トランスフェクションの3日後に、ゲノムDNAを抽出し、PCRによって増幅し、オンターゲット座位における高スループットDNAシーケンシングによって分析した。各プロトスペーサー内の標的Cに全ての4つの塩基を有する合計のDNAシーケンシングリードのパーセンテージが、BE3またはHF-BE3による処置について示されている（図56）。インデル形成の頻度もまた示されている。

【０４６９】

HF-BE3の一次蛋白質配列（配列番号48）：

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLE
LYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESATP
ESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGN
IVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV
DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIA
LSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYID
GGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRR
QEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKG
ASAQSFIERMTAFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQK
KAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGALSRKLI
NGIRDKQSGKTILDFLKSDGFANRNFMALIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL

AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEG
IKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG
GLSELDKAGFIKRQLVETRAITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESIL
MLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKM
LSGGSPKKKRKV

【０４７０】

例9：Cas9バリアントを用いる塩基編集因子の開発と、塩基編集テクノロジーの標的範囲および正確性を増大させるための塩基編集因子処理能力の調節

従来のゲノム編集プラットフォームとは違って、塩基編集テクノロジーは、二本鎖切断（DSB）を誘導することなしにDNAの正確な1ヌクレオチド変化を許す。Komor, A. C. et al. Nature 533, 420-424（2016）を見よ。現行の世代の塩基編集因子はNGG PAMを排他的に用いる。これは、ゲノム内の所望の塩基を編集するその能力を限定する。なぜなら、塩基編集因子が、標的塩基がPAMのおよそ15塩基上流の4塩基領域（「脱アミノ化ウィンドウ」）内に置かれている正確な場所に置かれることを必要とするからである。Komor, A. C. et al. Nature 533, 420-424（2016）を見よ。その上、シチジンデアミナーゼの高い処理能力が原因で、塩基編集因子はその脱アミノ化ウィンドウ内の全てのシチジンをチミジンに変換し得、これは研究者による所望の1つ以外のアミノ酸変化を誘導し得る。Komor, A. C. et al. Nature 533, 420-424（2016）を見よ。

【０４７１】

Cas9バリアントを有する塩基編集因子の開発によって塩基編集の範囲を拡大する

異なるPAM特異性を有するCas9ホモログおよび他のRNAによってガイドされるDNA結合因子が、塩基編集因子アーキテクチャに組み込まれている。Kleinstiver, B. P. et al. Nature 523, 481-485 (2015)；Kleinstiver, B. P. et al. Nature Biotechnology 33, 1293-1298 (2015)；およびZetsche, B. et al. Cell 163, 759-771 (2015)を見よ；それぞれの内容全体は参照によって本明細書に組み込まれる。さらにその上、種々のCas9蛋白質のPAM特異性を広げたイノベーションもまた組み込まれて、塩基編集因子の標的範囲をより多くさえも拡大している。Kleinstiver, B. P. et al. Nature 523, 481-485 (2015)；およびKleinstiver, B. P. et al. Nature Biotechnology 33, 1293-1298 (2015)を見よ。塩基編集因子の現行のパレットが表4にまとめられている。

表4．Cas9バリアントから作られた新たな塩基編集因子

【表１】

種	PAM	塩基編集因子名称	Cas9バリアントの参考文献
S. pyogenes	...NGG	BE3	野生型
	...NGA	VQR BE3または EQR BE3	Kleinstiver, B. P. *et al.*
	...NGCG	VRER BE3	Kleinstiver, B. P. *et al.*
S. aureus	...NNGRRT	SaBE3	Wild-type
	...NNNRRT	SaKKH BE3	Kleinstiver, B. P. *et al.*
L. bacterium	TTTN...	dCpf1 BE2	Zetsche, B. *et al.*

【０４７２】
rAPOBEC1の部位特異的変異導入によって塩基編集因子の処理能力を調節する
塩基編集因子の処理能力はデアミナーゼ酵素に点変異を作ることによって調節され得ることが推論された。塩基編集因子が平均で1ラウンドのシチジン脱アミノ化をなお触媒し得るが、該当するタイムスケール内においては別の脱アミノ化にアクセスし触媒することが非蓋然的である、デアミナーゼの触媒活性を軽度に縮減する変異の組み込みを追求した。実際には、もたらされる塩基編集因子はより狭い脱アミノ化ウィンドウを有するであろう。
【０４７３】
この研究において探られたrAPOBEC1変異は表5にリスト化されている。変異のいくつかはrAPOBEC1触媒作用の軽度の見かけ上の棄損をもたらした。これは、複数のシチジンが脱アミノ化ウィンドウ内に見出されるときには、別のものと比べて1つのシチジンの選好的な編集として現出した。これらの変異のいくつかを組み合わせることは相加的な効果を有し、塩基編集因子が基質シチジンをより高い厳格さで差別することを許した。二重変異体および三重変異体のいくつかは、互いにすぐ隣である複数のシチジンのうちの1つのシチジンの選択的編集を許した（図57）。
表5．検討されたrAPOBEC1点変異
【表２】

本研究において研究されたrAPOBEC1変異	APOBEC3Gの対応する変異	参考文献
W121R / W122R	D315R / D316R	Holden, L. G. et al.
R126A	R320A	Chen, K-M. et al.
R126E	R320E	Chen, K-M. et al.
R118A	R313A	Chen, K-M. et al.
W90A	W285A	Chen, K-M. et al.
W90Y	W285Y	
R132E	R326E	

【０４７４】
塩基編集因子PAM拡大および処理能力調節
他のRNAによってガイドされるDNA結合因子を用いることによってゲノム中の編集可能なシチジンを拡大するために、次世代の塩基編集因子を設計した（図58）。NGG PAMを用いることは「ウィンドウ」内の1つの標的のみを許すが、複数の異なるPAMの使用は、Cas9が選択的脱アミノ化を成し遂げるためにどこにでも位置することを許す。種々の新たな塩基編集因子をCas9バリアントから作出した（図59および表4）。異なるPAM部位（NGA、図60；NGCG、図61；NNGRRT、図62；およびNNHRRT、図63）を探求した。シチジンデアミナーゼ点変異導入の速度論的調節によって、選択的脱アミノ化が首尾良く達成された（図65および表5）。
【０４７５】
それから、脱アミノ化ウィンドウへの種々の変異の効果を、複数のシチジンを有するスペーサーを用いて細胞培養によって検討した（図66および67）。
【０４７６】
さらに、限定された数のシチジンを有する異なるゲノム部位への種々の変異の効果を調べた（図68～71）。およそ1つのシチジンがスペーサーの脱アミノ化ウィンドウ（windown）内において編集されるであろうが、シチジンの残りは無傷で残されるであろうことが見出された。総体的に、編集の選好性は次の通りである：$C_6 > C_5 >> C_7 \quad C_4$。

【０４７７】
Cpf1を用いる塩基編集
Cas9ホモログCpf1はAsCpf1、LbCpf1として、またはいずれかの他の種から得られ得る。BE2およびBE3等価物を包含する融合体構築物の模式図が図73に示されている。BE2等価物はCas9の代わりに触媒不活性型Cpf2酵素（dCpf1）を用い、BE3等価物はCpf1変異体を包含し、これは標的鎖にニックを入れる。下の模式図は、それの上に例示されている2つのイノベーションを組み合わせるための異なる融合アーキテクチャを図示している（図73）。Cpf1 BE2を用いるHEK293T細胞TTTN PAM部位の塩基編集結果を異なるスペーサーによって調べた（図64A～64C）。いくつかの態様において、Cpf1は本明細書において提供される塩基編集因子のいずれかにおいてCas9ドメインの代わりに用いられ得る。いくつかの態様において、Cpf1は、配列番号9と少なくとも（at lesst）50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、98%、99%、または99.5%同一である蛋白質である。
Cpf1の全長蛋白質配列（配列番号9）：
MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
EEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFN
QNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHEN
RKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDID
YKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYS
QQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKS
IKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLD
NPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIA
QNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKAN
ILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKE
PDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFF
SAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKD
FGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSK
GRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNP
KKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRG
ERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKE
MKEGYLSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLV
FKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKY
ESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKN
HNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSK
TGTELDYLISPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGK
KLNLVIKNEEYFEFVQNRNN
【０４７８】
例10：塩基編集の増大したフィデリティ
BE3およびHF-BE3のプラスミド送達間の違いを調べ、2つはオンターゲット座位を同等の効率で編集することが見出された（図74および75）。しかしながら、HF-BE3はBE3よりもずっと少ないオフターゲット座位を編集しており、HF-BE3がBE3よりもずっと高いDNA特異性を有することを意味した（図76）。HEK細胞へのデアミナーゼ蛋白質リポフェクションは、BE3の蛋白質送達が、同等のオンターゲット活性だが、BE3のプラスミドDNA送達よりもずっと良好な特異性をもたらすことを実証している。改善されたトランスフェクション手続きと良好なプラスミドとを用いて（n=2）、実験は次の条件を用いた：蛋白質送達は125nM Cas9:sgRNA複合体であり、プラスミド送達は750ng BE3/HF-BE3プラスミド+250ng sgRNAプラスミドであり、リポフェクションはウェルあたり1.5μLのLipofectamine 2000によってであった。EMX-1オフターゲット部位2およびFANCFオフターゲット部位1は、アッセイされたオフターゲットの全てと比較してBE3による最も多いオフターゲット編集を示したが（図77および78）、HEK-3は送達方法のいずれか

についてオフターゲットの有意な編集を示さなかった（図79）。HEK-4はオンターゲット部位においていくらかのC→T編集を示し、そのオフターゲット部位1、3、および4は全てのアッセイされた部位のうち最も多いオフターゲット編集を示した（図80）。

【０４７９】

ゼブラフィッシュへのマイクロインジェクションによるBE3蛋白質の送達

TYRガイドRNAをsgRNA活性についてin vitroアッセイによって試験した（図81および82）。%HTSリードは、精製されたBE3蛋白質との2hインキュベーションおよびもたらされた生成物のPCRの間に、いかに多くのC残基がT残基に変換されたかを示している。実験は、60bpのそのゲノムコンテキストに標的脱アミノ化部位を有する80mer合成DNA基質（substate）を用いた。これは%の編集されたDNA鎖と同じではない。なぜなら、1つの鎖のみにニックが入ったからであり、そのため、生成物はPCRによって増幅されない。編集されたHTSリードの割合はx/(2-x)に等しく、ここでxは編集されたTHSリードの実際の割合である。60%の編集では、編集された塩基の実際の割合は75%である。「オフターゲット」は、同じDNA基質とインキュベーションされたがオフターゲットsgRNAに結合されたBE3に相当する。sgRNA sgRH_13、sgHR_17、および可能性としてsgHR_16はin vivoインジェクション実験の有望な標的であるように見えることが見出された。

【０４８０】

BE3蛋白質の送達をゼブラフィッシュにおいてin vivoで試験した。ゼブラフィッシュ胚（n=16～24）に、スクランブルsgRNA、sgHR_13、sgHR_16、またはsgHR_17どちらかおよび精製されたBE3をインジェクションした。各条件からの3つの胚を独立して分析し（単一の胚）、各条件について、インジェクションされた胚の全てをプールし、プールとしてシーケンシングした。結果は図83～85に示されている。

【０４８１】

例11：疾患を処置するための塩基編集因子の使用

本明細書において提供される塩基編集因子または複合体（例えばBE3）は核酸を改変するために用いられ得る。例えば、塩基編集因子は核酸（例えばDNA）中のシトシンをチミンに変化させるために用いられ得る。かかる変化は、とりわけ蛋白質のアミノ酸配列を変改するため、開始コドンを壊すかもしくは作出するため、終止コドンを作出するため、スプライシングドナーを破壊するため、スプライシングアクセプターを破壊するため、または制御配列を編集するためになされ得る。可能なヌクレオチド変化の例は図86に示されている。

【０４８２】

本明細書において提供される塩基編集因子または複合体（例えばBE3）は、対象におけるアポリポ蛋白質Eのアイソフォームを編集するために用いられ得る。例えば、アポリポ蛋白質Eアイソフォームは、アルツハイマー病を発生するより低いリスクに関連するアイソフォームを生むように編集され得る。アポリポ蛋白質Eはアミノ酸112および158において異なる4つのアイソフォームを有する。APOE4は遅発型アルツハイマー病の最も大きくかつ最も一般的な遺伝学的リスク因子である。核酸配列CGCによってコードされるAPOE4のアルギニン残基158は、CGC核酸配列を残基158にシステインをコードするTGCに変化させるために塩基編集因子（例えばBE3）を用いることによって、システインに変化させられ得る。この変化はAPOE3rアイソフォームを生み、これはより低いアルツハイマー病リスクに関連する。図87を見よ。

【０４８３】

塩基編集因子BE3がマウスアストロサイトにおいてAPOE4をAPOE3rへと編集するために用いられ得るかどうかを試験した（図88）。APOE4のアルギニン158をコードする核酸を標的化して、Cas9+鋳型またはBE3によって、APOE4マウスアストロサイトをヌクレオフェクションした。Cas9+鋳型は0.3%のみの編集を生み、26%のインデルを有したが、BE3は75%の編集を生み、5%のインデルを有した。2つの追加の塩基編集されたシトシンはサイレントであり、アミノ酸配列への変化を生まない（図88）。

【０４８４】

本明細書において提供される塩基編集因子または複合体は、例えばPRNP遺伝子に変異を導入することによって、プリオン蛋白質疾患、例えばクロイツフェルト・ヤコブ病および致死性家族性不眠症を処置するために用いられ得る。PRNP変異を復帰させることは治療上の結果を生まずにあり得、PRNP中のインデルは病原性であり得る。従って、PRNP遺伝子に未成熟終止コドンを導入するように塩基編集因子（例えばBE3）を用いてPRNPを変異させ得るかどうかを試験した。そのガイドRNAと結び付いたBE3がHEK細胞または膠芽腫細胞に導入され、PRNP遺伝子を編集して残基37のコードされるアルギニンを終止コドンに変化させることができた。BE3は41%の編集を生んだ（図89）。

【０４８５】

編集され得る追加の遺伝子は次を包含する：増大したアルツハイマーのリスクを処置するためのArg112およびArg158のAPOE編集；アルツハイマーのリスクを減少させるためのAla673のAPP編集；致死性家族性不眠症および他のプリオン蛋白質疾患を処置するためのArg37のPRNP編集；デュシェンヌ型筋ジストロフィーを処置するためのエキソン23および51スプライシング部位のDMD編集；肥満リスクを処置するためのイントロン1のFTO編集；ペンドレッド症候群（遺伝子性難聴）を処置するためのエキソン8のPDS編集；先天性難聴を処置するためのエキソン8のTMC1編集；慢性肉芽腫症を処置するための種々の患者に該当する変異のCYBB編集。本明細書において提供される塩基編集因子を用いて処置され得る追加の疾患は下の表6に示されている。

【０４８６】

UGIもまた鍵となる役割を演ずる。UDG（これをUGIが阻害する）をノックアウトすることは、C→T塩基編集のクリーンさおよび効率を劇的に改善することが示された（図90）。さらにその上、ニッカーゼを有しかつUGIなしの塩基編集因子は成果の混合物を生成することが示され、非常に高いインデル率を有した（図91）。

【０４８７】

例12：塩基編集の標的化範囲を拡大する

塩基編集はゲノム編集の新たなアプローチであり、これは、触媒的に欠陥がある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有する融合蛋白質を用いて、二本鎖DNA切断を創出することなしに、ドナーDNA鋳型を要求することなしに、かつ過剰な確率的な挿入および欠失を誘導することなしに、プログラム可能な1ヌクレオチドC→T（またはG→A）変化をDNA中に誘導する[1]。5つの新たなC→T（またはG→A）塩基編集因子の開発が本明細書に記載され、これらは異なるプロトスペーサー隣接モチーフ（PAM）特異性を有する天然および人工のCas9バリアントを用いて、塩基編集によって標的化され得る部位の数を2.5倍拡大する。加えて、見かけ上の編集ウィンドウの幅をおよそ5ヌクレオチドから1または2ヌクレオチドに狭める変異したシチジンデアミナーゼドメインを含有する新たな塩基編集因子を操作し、先には同等の効率で編集されたであろう近隣のCヌクレオチド同士の差別を可能にした。一緒になって、これらの開発は塩基編集の標的化範囲を実質的に増大させる。

【０４８８】

CRISPR-Cas9ヌクレアーゼは標的化されたゲノム編集を媒介するために幅広く用いられてきた[2]。ほとんどのゲノム編集用途において、Cas9はシングルガイドRNA（sgRNA）との複合体を形成し、sgRNA配列によって指定される標的部位における二本鎖DNA切断（DSB）を誘導する。細胞は主として非相同末端結合（NHEJ）修復経路によってこのDSBに応答し、これは、遺伝子を破壊するフレームシフト変異を引き起こし得る確率的な挿入または欠失（インデル）をもたらす。DSBをフランキングする配列に対する高い程度の相同性を有するドナーDNA鋳型の存在下においては、相同組み換え修復（HDR）として公知の代替的な経路によって、遺伝子修正が達成され得る[3,4]。残念ながら、ほとんどの非摂動的な条件においては、HDRは非効率的であり、細胞状態および細胞型に依存し、より大きい頻度のインデルによって支配される[3,4]。ヒト疾患に関連する公知の遺伝子多様性のほとんどは点変異であるので[5]、より効率的にかつクリーンに正確な点変異を作り得る方法が必要とされる。

【０４８９】
DSBを誘導することなしにプログラム可能な様式でT:A塩基対によるC:G塩基対の標的化された置き換えを可能にする塩基編集が[1]、最近記載された。塩基編集は、Streptococcus pyogenes Cas9（SpCas9）の触媒不活性型（dCas9）またはニッカーゼ形態、APOBEC1などのシチジンデアミナーゼ、およびウラシルグリコシラーゼ阻害因子（UGI）などの塩基除去修復の阻害因子の間の融合蛋白質を用いて、sgRNAによって指定される5ヌクレオチドウィンドウ内のシチジンをウリジンに変換する[1]。第3世代塩基編集因子BE3は、種々の細胞株において、疾患に該当する点変異を包含するC:G塩基対をT:A塩基対に変換し、他のゲノム編集方法を用いて達成され得るものよりも高い効率および低いインデル頻度を有する[1]。爾後の研究はデアミナーゼ-dCas9融合アプローチを種々の設定において検証した[6,7]。

【０４９０】
BE3による効率的な編集は、プロトスペーサーのPAMから遠位の末端に近い5ヌクレオチドウィンドウ（位置4～8。PAMを位置21～23としてカウントする）内に標的Cを置くNGG PAMの存在を要求する[1]。多くの当該部位は標的Cの13から17ヌクレオチド下流のNGGを欠くので、このPAM要件は、BE3によって効率的に標的化され得るヒトゲノム中の部位の数を実質的に限定する。その上、BE3の高い活性および処理能力は編集ウィンドウ内の全てのCからTへの変換をもたらし、これは見込みとして標的座位に所望ではない変化を導入し得る。本明細書においては、これらの限定の両方に対処する新たなC:G→T:A塩基編集因子が記載される。

【０４９１】
DNAに結合し、一本鎖DNAバブルを含有する「Rループ」複合体[8]を形成するいずれかのCas9ホモログは、原理的に塩基編集因子へと変換され得ることが考えられた。これらの新たな塩基編集因子は、非NGG PAM部位が編集されることを許すことによって、標的化可能座位の数を拡大するであろう。Staphylococcus aureusからのCas9ホモログ（SaCas9）はSpCas9よりもかなり小さく（1053vs.1368残基）、哺乳類細胞における効率的なゲノム編集を媒介し得、NNGRRT PAMを要求する[9]。BE3においてSpCas9をSaCas9によって置き換えてSaBE3を創出し、SaBE3および6つのヒトゲノム座位を標的化するsgRNAをコードするプラスミドによって、HEK293T細胞をトランスフェクションした（図92Aおよび92B）。3d後に、ゲノム座位を高スループットDNAシーケンシング（HTS）に付して、塩基編集効率を定量した。SaBE3は、ヒト細胞において種々のゲノム部位の標的CのC→T塩基編集を可能にし、非常に高い変換効率（トランスフェクションされた細胞の濃縮なしに、合計のDNA配列のおよそ50～75%がCからTに変換された）が、位置6～11のCを標的化することから生じた。一般的に、NNGRRT含有標的部位におけるSaBE3の効率はNGG含有標的部位におけるBE3のものを超えた[1]。おそらくはそのより高い平均効率が原因で、SaBE3は古典的BE3活性ウィンドウの外の位置の標的Cにおける検出可能な塩基編集をもまたもたらし得る（図92C）。比較して、公知のSpCas9 PAM選好性に従って、BE3は同じ条件において有意に縮減された編集（0～11%）を示した（図106A）[10]。これらのデータは、SaBE3が、BE3にはアクセス可能でない部位における非常に効率的な塩基編集を容易化し得ることを示している。

【０４９２】
PAM特異性を拡大または変改する最近操作されたCas9バリアントを適用することによって、塩基編集因子の標的化範囲をさらに拡大した。最近、Joungらは、NGA（VQR-Cas9）、NGAG（EQR-Cas9）、またはNGCG（VRER-Cas9）PAM配列を受け入れる3つのSpCas9変異体を報告した[11]。加えて、Joungらは、そのPAM要件をNNNRRTへと緩和する3つの変異を含有するSaCas9バリアント（SaKKH-Cas9）を操作した[12]。BE3のSpCas9部分をこれらの4つのCas9バリアントによって置き換えて、VQR-BE3、EQR-BE3、VRER-BE3、およびSaKKH-BE3を生成した。これらはそれぞれNNNRRT、NGA、NGAG、およびNGCG PAMを標的化する。各新たな塩基編集因子について、これらの構築物および6つのゲノム座位を標的化するsgRNAをコードするプラスミドによって、HEK293T細胞を

トランスフェクションし、HTSを用いてC→T塩基変換を測定した。

【０４９３】

SaKKH-BE3は、処置された非濃縮細胞の62%までの効率で、NNNRRT PAMを有する部位を編集した（図92D）。予想される通り、SaBE3はNNNHRRT（ここではH=A、C、またはT）であるPAMを含有する標的を効率的に編集する能力があった（図92D）。VQR-BE3、EQR-BE3、およびVRER-BE3は、BE3のものに類似の編集ウィンドウの予想されるPAM要件を有するゲノム座位において、処置された非濃縮細胞の50%までという、より控えめだがなお実質的な塩基編集効率を見せた（図92Eおよび92F）。一般的に、VQR-BE3、EQR-BE3、およびVRER-BE3の塩基編集効率は、対応するCas9ヌクレアーゼの報告されたPAM要件と密に符合した；例えば、EQR-BE3はNGAH PAM配列を含有する標的を効率的に編集する能力がなかった（図92F）。対照的に、蓋然的にはそのPAM制限が原因で、BE3はNGAまたはNGCG PAMを有する部位を効率的に編集する能力がなかった（0～3%）（図106B）。

【０４９４】

総合すると、SaBE3、SaKKH-BE3、VQR-BE3、EQR-BE3、およびVRER-BE3の特性は、塩基編集因子が、Cas9ホモログおよび人工のバリアントを活用するそれらの能力を容易化するモジュール性を見せることを確立する。

【０４９５】

次に、変改された活性ウィンドウ幅を有する塩基編集因子を開発した。BE3の活性ウィンドウ内の全てのCは効率的にTに変換され得る[1]。このウィンドウの幅を調節する能力は、BE3活性ウィンドウ内に存在するCのある部分集合のみを編集することが重要であるケースにおいて有用であろう。

【０４９６】

APOBEC1とdCas9との間のリンカーの長さは、in vitroでAPOBEC1によってアクセス可能である塩基の数を調節することが先に観察されている[1]。しかしながら、HEK293T細胞では、リンカー長さを変えることは編集ウィンドウの幅を有意には調節せず、複雑な細胞内環境においては、dCas9およびシチジンデアミナーゼの相対的な向きおよび柔軟性はリンカー長さによって強くは決定されないことを示唆した（図96）。次に、sgRNAの5'末端を短縮することは、RNA-DNAヘテロ二重鎖の形成によってデアミナーゼにとってアクセス可能な一本鎖DNAの長さを縮減することによって、塩基編集ウィンドウを狭め得ることが考えられた。BE3および編集ウィンドウ内に複数のCを有する座位を標的化する異なるスペーサー長さのsgRNAをコードするプラスミドによって、HEK293T細胞をコトランスフェクションした。17から19塩基のスペーサーを有する短縮されたsgRNAを用いるときに、塩基編集の幅の整合的な変化は観察されなかった（図95A～95C）。sgRNAスペーサーを17塩基よりも少数まで短縮することは活性の大きい損失をもたらした（図95A）。

【０４９７】

代替的なアプローチとして、デアミナーゼドメインの変異は複数の可能なメカニズムによって編集ウィンドウの幅を狭め得ることが考えられた。第1に、いくつかの変異は、基質結合、結合したDNAの立体配座、または活性部位への基質アクセス可能性を、デアミナーゼ活性部位へのCの非最適な提示に対する寛容性を縮減するやり方で変改し得る。第2に、APOBEC1の高い活性は蓋然的にDNA結合イベントあたり複数のCの脱アミノ化に寄与するので[1,13,14]、塩基編集因子のデアミナーゼドメインの触媒効率を縮減する変異は、DNAから解離する前に、それが逐次的なラウンドの脱アミノ化を触媒することを防止し得る。ひとたびいずれかのC:G→T:A編集イベントが起こると、sgRNAはもはや完璧には標的DNA配列にマッチせず、標的座位への塩基編集因子の再結合はより好都合でないであろう。元々の編集ウィンドウ内の複数のシチジンを見分ける新たな塩基編集因子を発見するための作業において、両方の戦略を試験した。

【０４９８】

入手可能なAPOBEC1構造の非存在に鑑みて、APOBEC1のものに対してその活性部位含有ドメインの42%の配列類似性を共有する同じファミリーからのシチジンデアミナーゼAP

OBEC3Gの触媒活性を調節することが先に報告されているいくつかの変異を同定した[15]。対応するAPOBEC1変異をBE3に組み込み、HEK293T細胞における塩基編集効率および編集ウィンドウ幅へのそれらの効果を、位置3、4、5、6、8、9、10、12、13、および14にCを含有する（部位A）；または位置5、6、7、8、9、10、11、および13にCを含有する（部位B）2つのCリッチなゲノム部位において評価した。

【０４９９】

APOBEC1変異R118AおよびW90Aはそれぞれが塩基編集効率の劇的な損失に至った（図97C）。R132Eは編集効率の一般的な減少に至ったが、編集ウィンドウの形状を変化させるかまたは実質的に狭めるかしなかった（図97C）。対照的に、実質的な編集効率を維持しながら編集ウィンドウの幅を狭めるいくつかの変異が見出された（図93Aおよび97C）。「編集ウィンドウ幅」は、編集効率がその標的について半最大値を超える人為的に算出されたウィンドウ幅に相当するように定められた。試験された2つのCリッチなゲノム部位のBE3の編集ウィンドウ幅は5.0（部位A）および6.1（部位B）ヌクレオチドであった。

【０５００】

APOBEC1のR126はssDNAのリン酸バックボーンと相互作用することが予測される[13]。先の研究は、対応する変異をAPOBEC3Gに導入することが触媒作用を少なくとも5倍減少させたことを示している[14]。興味深いことに、BE3のAPOBEC1に導入されたときに、R126AおよびR126Eは、最も強く編集された位置（C5、C6、およびC7）においてはBE3に対して相対的に活性を増大させるかまたは維持したが、他の位置においては編集活性を減少させた（図93Aおよび97C）。よって、これらの2つの変異のそれぞれは、部位Aおよび部位Bにおける編集ウィンドウの幅をそれぞれ4.4および3.4ヌクレオチド（R126A）または4.2および3.1ヌクレオチド（R126E）まで狭めた（図93Aおよび97C）。

【０５０１】

APOBEC1のW90（APOBEC3GのW285に対応する）は、APOBEC3G活性部位に疎水性ポケットを形成し、基質結合を援助することが予測される[13]。この残基をAlaに変異させることはAPOBEC3Gの触媒活性を無効化した[13]。BE3では、W90Aはほぼ完全に塩基編集効率を無効化した（図97C）。対照的に、W90Yは、部位Aおよび部位Bにおける編集ウィンドウ幅をそれぞれ3.8および4.9ヌクレオチドまで狭めながら、塩基編集活性を控えめにのみ減少させることが見出された（図93A）。これらの結果は、シチジンデアミナーゼドメインへの変異が対応する塩基編集因子の活性ウィンドウ幅を狭め得ることを実証している。

【０５０２】

塩基編集活性を劇的に縮減することなしに編集ウィンドウを狭めた3つの変異W90Y、R126E、およびR132Eを、二重および三重変異した塩基編集因子へと組み合わせた。二重変異体W90Y+R126Eは、BE3様の最大編集効率だが、実質的に狭められた編集ウィンドウ幅を有する塩基編集因子（YE1-BE3）をもたらした（部位Aおよび部位Bにおける幅=それぞれ2.9および3.0ヌクレオチド（図93A）。W90Y+R132E塩基編集因子（YE2-BE3）は、控えめにより低い編集効率（BE3と比較して試験された5つの部位において平均1.4倍低い最大編集収率である）、および実質的に狭められた編集ウィンドウ幅（部位Aおよび部位Bの幅=それぞれ2.7および2.8ヌクレオチド）をもまた見せた（図97C）。R126E+R132E二重変異体（EE-BE3）はYE2-BE3に類似の最大編集効率および編集ウィンドウ幅を示した（図97C）。三重変異体W90Y+R126E+R132E（YEE-BE3）は、2.0倍低い平均の最大編集収率だが、C_6位置のほかでは非常に少ない編集、ならびに部位Aおよび部位Bについてはそれぞれ2.1および1.4ヌクレオチドという編集ウィンドウ幅を見せた（図97C）。一緒にすると、これらのデータは、シチジンデアミナーゼドメインの変異が編集ウィンドウ幅に強く影響し得、いくつかのケースでは編集効率への最小限のまたは控えめな効果のみを有することを指示している。

【０５０３】

BE3活性ウィンドウ内に複数のCを含有する4つの良く研究されたヒトゲノム部位を標的化するBE3、YE1-BE3、YE2-BE3、EE-BE3、およびYEE-BE3の塩基編集成果をHEK29

3T細胞においてさらに比較した[1]。これらの標的座位は標的Cを位置4および5（HEK部位3）、位置4および6（HEK部位2）、位置5および6（EMX1）、または位置6、7、8、および11（FANCF）に含有した。BE3は、位置4～8活性ウィンドウ内のいずれかのCを編集することについて少ない（＜1.2倍）選好性を見せた。対照的に、YE1-BE3は、C_4に対してC_5を編集することについて1.3倍の選好性（HEK部位3）、C_4に対してC_6について2.6倍の選好性（HEK部位2）、C_6に対してC_5について2.0倍の選好性（EMX1）、C_7に対してC_6について1.5倍の選好性（FANCF）を見せた（図93B）。YE2-BE3およびEE-BE3はYE1-BE3よりもいくぶん多大な位置特異性（より狭い活性ウィンドウ）を見せ、C_4に対してC_5を編集することについて平均2.4倍の選好性（HEK部位3）、C_4に対してC_6について9.5倍の選好性（HEK部位2）、C_6に対してC_5について2.9倍の選好性（EMX1）、C_6に対してC_7について2.6倍の選好性（FANCF）であった（図93B）。YEE-BE3は最も多大な位置選択性を示し、C_4に対してC_5を編集することについて2.9倍の選好性（HEK部位3）、C_4に対してC_6について29.7倍の選好性（HEK部位2）、C_6に対してC_5について7.9倍の選好性（EMX1）、C_6に対してC_7について7.9倍の選好性（FANCF）を有した（図93B）。知見は、両方のヌクレオチドがBE3編集ウィンドウ内にあるときでさえも、変異体塩基編集因子が隣接するC同士を差別し得ることを確立している。

【０５０４】

これらの4つの変異体およびBE3の生成物分布をHTSによってさらに分析し、それらの見かけ上の処理能力を評価した。BE3は、処置されたHEK293T細胞のT_4-T_5（HEK部位3）、T_4-T_6（HEK部位2）、およびT_5-T_6（EMX1）生成物を圧倒的に創出し、1つのTを含有する生成物よりも平均で7.4倍多くの2つのTを含有する生成物をもたらした。対照的に、YE1-BE3、YE2-BE3、EE-BE3、およびYEE-BE3は一重編集されたC_4-T_5、C_4-T_6、およびT_5-C_6生成物について実質的により高い選好性を示した（図93C）。YE1-BE3は1.4という平均の1T対2T生成物比を有する生成物を生んだ。YE2-BE3およびEE-BE3はそれぞれ4.3および5.1という平均の1T対2T生成物比を有する生成物を生んだ（図93C）。上の結果と整合して、YEE-BE3三重変異体は、3つのゲノム座位において1Tの生成物を平均で14.3倍選好した（図93C）。1つのCのみが標的ウィンドウ内である標的部位では（HEK部位4、位置C_5）、全ての4つの変異体はBE3と同等の編集効率を見せた（図98）。これらの知見は、これらのBE3変異体が減少した見かけ上の処理能力を有し、BE3編集ウィンドウ内に複数のCを含有する標的部位において1つのCのみの変換を選好し得ることを指示している。これらのデータはこれらの変異体塩基編集因子についてC5＞C6＞C7　C4という位置選好性をもまた示唆しているが、この選好性は標的配列に依存して異なり得る。

【０５０５】

APOBEC1のウィンドウ調節変異をVQR-BE3に適用し、NGA PAMによって標的化される部位における基質の選択的な塩基編集を許した（図107A）。しかしながら、これらの変異をSaKKH-BE3に適用したときには、塩基編集効率の線形的な減少が基質選択性の改善なしに観察され、BE3およびそのバリアントのものとは異なるこの塩基編集因子の速度論的平衡および基質アクセス可能性を示唆した（図107B）。

【０５０６】

この研究において記載されている変改されたPAM特異性を有する5つの塩基編集因子は、一緒になって、原理的に塩基編集によって修正され得るClinVarデータベース中の疾患関連変異の数を2.5倍増大させる（図94Aおよび94B）。類似に、狭められた編集ウィンドウを有する塩基編集因子の開発は、非標的Cの同等の改変なしに塩基編集によって修正され得るしかるべく位置したNGG PAMを有するClinVarエントリーの画分をおよそ倍化させる（BE3の31%からYEE-BE3の59%まで）（図94Aおよび94B）。

【０５０７】

まとめると、異なるPAM特異性を有するCas9バリアントを用いる塩基編集因子を開発することによって、および様々な編集ウィンドウ幅を有するデアミナーゼ変異体のコレクションを開発することによって、塩基編集の標的化範囲が実質的に拡大された。理論上は、塩基編集は、一本鎖特異的なヌクレオチドデアミナーゼ酵素の基質としての用をなし得

る一本鎖DNAのバブルを作出する他のプログラム可能なDNA結合蛋白質（例えばCpf1[16]）を用いて可能であるはずである。

【０５０８】

材料および方法

クローニング。

PCRはQ5ホットスタート高フィデリティDNAポリメラーゼ（New England Biolabs）を用いて行った。BEおよびsgRNAのプラスミドはUSERクローニング（New England Biolabs）を用いて構築し、先に報告されているプラスミドから得た[1]。DNAベクター増幅はNEB10ベータコンピテントセル（New England Biolabs）を用いて実施した。

【０５０９】

細胞培養。

HEK293T（ATCC CRL-3216）は10%（v/v）ウシ胎児血清（FBS）を添加したダルベッコ改変イーグル培地プラスGlutaMax（Thermo Fisher）によって37℃において5%CO_2で培養した。APOE遺伝子のApoE4アイソフォームを含有する不死化ラットアストロサイト（Taconic Biosciences）は、10%（v/v）ウシ胎児血清（FBS）および200μg/mLジェネティシン（Thermo Fisher Scientific）を添加したダルベッコ改変イーグル培地プラスGlutaMax（Thermo Fisher Scientific）によって維持した。

【０５１０】

トランスフェクション。

HEK293T細胞を48ウェルコラーゲンコートBioCoatプレート（Corning）に接種し、およそ85%コンフルエンシーでトランスフェクションした。製造者のプロトコールに従ってウェルあたり1.5μlのLipofectamine 2000（Thermo Fisher Scientific）を用いて、750ngのBEおよび250ngのsgRNA発現プラスミドをトランスフェクションした。

【０５１１】

ゲノムDNAサンプルの高スループットDNAシーケンシング。

トランスフェクションされた細胞を3d後に収穫し、ゲノムDNAを製造者の説明書に従ってAgencourt DNAdvanceゲノムDNA単離キット（Beckman Coulter）を用いて単離した。当該ゲノム領域を、補足的な配列にリスト化されているフランキングHTSプライマーペアによるPCRによって増幅した。PCR増幅は製造者の説明書に従ってPhusionホットスタートII DNAポリメラーゼ（Thermo Fisher）によって実施した。PCR生成物をRapidTips（Diffinity Genomics）を用いて精製した。二次PCRを行ってシーケンシングアダプターを取り付けた。生成物をゲル精製し、KAPAライブラリー定量キット-Illumina（KAPA Biosystems）を用いて定量した。サンプルは先に記載されている通りIllumina MiSeqによってシーケンシングした[1]。

【０５１２】

データ分析。

ヌクレオチド頻度は先に記載されているMATLABスクリプト[1]を用いて査定した。簡潔には、リードをSmith-Watermanアルゴリズムによって参照配列に対してアライメントした。30よりも下のQスコアを有するベースコールはプレースホルダーヌクレオチド（N）によって置き換えた。このクオリティ閾値は1000分の1という理論上の予想エラー率でヌクレオチド頻度をもたらす。

【０５１３】

塩基編集処理能力の分析はカスタムのpythonスクリプトを用いて行った。このプログラムは、標的部位をフランキングするべき7ヌクレオチド配列の完璧なマッチによって決定される20ヌクレオチドプロトスペーサー配列まで、シーケンシングリードをトリミングする。それから、これらの標的をコンソリデーションし、豊富さによってソーティングして塩基編集生成物の頻度を査定した。

【０５１４】

ヒト疾患関連変異のClinVarデータベースのバイオインフォマティクス分析は、先に記載されているものに類似だが小さい調整を有する様式で行なった[1]。これらの調整は、カス

タマイズ可能な長さおよび配列のPAMを有する標的の同定を可能にする。加えて、この改善されたスクリプトは標的C位置の優先度順位を包含し（C_5＞C_6＞C_7＞C_8　C_4）、それゆえに、オンターゲットCがウィンドウ内の唯一のシトシンであるかまたは編集ウィンドウ内のいずれかのオフターゲットCよりも高い予測される編集効率を有する位置に置かれているかどちらかの、標的部位の同定を可能にする。

例12の参考文献

【表３】

1 Komor, A. C. *et al.* Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420-424 (2016).
2 Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature biotechnology* **32**, 347-355 (2014).
3 Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823 (2013).
4 Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protocols* **8**, 2281-2308 (2013).
5 Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D862–D868 (2015).
6 Nishida, K. *et al.* Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729-1-8 (2016).
7 Ma, Y. *et al.* Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* doi:10.1038/nmeth.4027 (2016).
8 Jiang, F. *et al.* Structures of a CRISPR–Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867-71 (2016).

【表４】

9 Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186-191 (2015).
10 Zhang, Y. *et al.* Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. *Sci. Rep.* **4**, (2014).
11 Kleinstiver, B. P. *et. al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481-485 (2015).
12 Kleinstiver, B. P. *et. al.* Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nat. Biotechnol.* **33**, 1293-1298 (2015).
13 Holden, L. G. et al. Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature* **452**, 121-124 (2008).
14 Chen, K.-M. et al. Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature* **452**, 116-119 (2008).
15 Harris, R. S., Petersen-Mahrt, S. K. & Neuberger, M. S. RNA Editing Enzyme APOBEC1 and Some of Its Homologs Can Act as DNA Mutators. *Molecular Cell* **10**, 1247-1253 (2002).
16 Zetsche, B. *et al.* Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163**, 759–771 (2015).

【０５１５】

例13

改善されたトランスフェクション手続きとより良好なプラスミドとを用いて、生物学的レプリケート（n=3）を用いて4つのHF変異をBE3のCas9部分にインストールした。変異はプラスミド送達によるオンターゲット編集を有意には左右しない（図99）。試験された濃度では、BE3蛋白質送達は働く；しかしながら、オンターゲット編集はプラスミド送達よりも低い（図100）。インストールされたHF変異を有するBE3の蛋白質送達は、オンターゲット編集効率を縮減するが、なおいくつかの編集された細胞を生む（図101）。

【０５１６】
リポフェクションとHF変異をインストールすることとは、両方、オフターゲット脱アミノ化イベントを減少させることが示された。図102に示される4つの部位では、最も高いGUIDE-Seqリードおよび脱アミノ化イベントを有するオフターゲット部位（OT）をアッセイした（Komor et al., Nature, 2016）。最も近接した対応するCにおけるオンターゲット編集によってオフターゲット編集を割ることによって、特異性比を算出した。オフターゲット編集が検出可能ではないケースでは、比を100に設定した。それゆえに、より高い特異性比はより特異的な構築物を指示している。BE3プラスミド送達は、BE3の蛋白質送達、HF-BE3のプラスミド送達、またはHF-BE3の蛋白質送達よりもずっと高いオフターゲット／オンターゲット編集を示した（図102および105）。

【０５１７】
精製された蛋白質HF-BE3およびBE3を、最も緩いモチーフを有するスペーサー内の異なる位置においてCをT残基に変換するそれらの性能についてin vitroで分析した。BE3およびHF-BE3蛋白質両方は塩基編集の同じ「ウィンドウ」を有することが見出された（図103および104）。

【０５１８】
疾患標的のリストが表8に与えられている。表8の編集されるべき塩基は太字によって指示され、下線付きである。

表8．塩基編集因子疾患標的

【表５】

遺伝子	疾患	スペーサー	配列番号	PAM	編集因子	欠陥	細胞
RB1	網膜芽細胞腫	AATCTAGTAAATAAATTGATGT	571	AAAAGT	SAKKH-BE3	スプライシング棄損	J82
PTEN	癌	GACCAACGGCTAAGTGAAGA	572	TGA	VQR-BE3	W111R	MC116
PIK3CA	癌	TCCTTTCTTCACGGTTGCCT	573	ACTGGT	SAKKH-BE3	K111R	CRL-5853
PIK3CA	癌	CTCCTGCTCAGTGATTTCAG	574	AGA	VQR-BE3	Q546R	CRL-2505
TP53	癌	TGTCACACATGTAGTTGTAG	575	TGG	YEE-BE3	N239D	SNU475
HRAS	癌	CCTCCCGGCCGGCGGTATCC	576	AGG	YEE-BE3	Q61R	MC/CAR

表6．塩基編集因子を用いて処置され得る例示的な疾患．プロトスペーサーおよびPAM配列（配列番号577～589）がsgRNA（PAM）カラムに示されている。PAM配列は括弧で示されており、編集されるべき塩基は下線付きによって指示されている。

【表６】

疾患標的	遺伝子記号	変化した塩基	sgRNA (PAM)	塩基編集因子
プリオン病	PRNP	R37*	GGCAGCCGATACCCGGGGCA(GGG)	BE3
			GGGCAGCCGATACCCGGGGC(AGG)	
ベンドレッド症候群	Slc26a4	c.919-2A>G	TTATTGTCCGAAATAAAAGA(AGA)	BE3
			ATTGTCCGAAATAAAAGAAG(AGG)	(VQR
			TTGTCCGAAATAAAAGAAGA(GGA)	SaCas9)
			GTCCGAAATAAAAGAAGAGGAAAA(AAT)	
			GTCCGAAATAAAAGAAGAGGAAAAA(ATT)	
先天性難聴	Tmc1	c.545A>G	CAGGAAGCACGAGGCCACTG(AGG)	BE3
			AACAGGAAGCACGAGGCCAC(TGA)	YE-BE3
			AGGAAGCACGAGGCCACTGA(GGA)	YEE-BE3
後天性難聴	SNHL	S33F	TTGGATTCTGGAATCCATTC(TGG)	BE3
アルツハイマー病	APP	A673T	TCTGCATCCATCTTCACTTC(AGA)	BE3 VQR
ニーマン・ピック病C型	NPC1	I1061T	CTTACAGCCAGTAATGTCAC(CGA)	BE3 VQR

【０５１９】

例14：塩基編集構築物を試験する

BE3、BE4-pmCDA1、BE4-hAID、BE4-3G、BE4-N、BE4-SSB、BE4-$(GGS)_3$、BE4-XTEN、BE4-32aa、BE4-2xUGI、およびBE4を包含するいくつかの塩基編集構築物を、異なる標的配列（すなわち、EMX1、FANCF、HEK2、HEK3、HEK4、およびRNF2）内のシトシン（C）残基を編集するそれらの能力について試験した。例えば、これらの構築物がC→T変異を生成することができるかどうかを試験した。塩基編集構築物の模式図が図109に示されている。試験された標的配列は図109にもまた示されており、標的シトシンは付番され、赤色で指示されている。

【０５２０】

次のアミノ酸配列をこの例の塩基編集構築物に用いた：

UGI：

TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML（配列番号736）

rAPOBEC1：

SSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK（配列番号737）

pmCDA1：

MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAV（配列番号81）

hAID：

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL（配列番号49）

hAPOBEC3G：

MELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLT

IFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPW
NNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWV
LLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCA
QEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVD
HQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN（配列番号60）
SSB（一本鎖DNA結合蛋白質）：
ASRGVNKVILVGNLGQDPEVRYMPNGGAVANITLATSESWRDKATGEMKEQTEWHRVV
LFGKLAEVASEYLRKGSQVYIEGQLRTRKWTDQSGQDRYTTEVVVNVGGTMQMLGGRQ
GGGAPAGGNIGGGQPQGGWGQPQQPQGGNQFSGGAQSRPQQSAPAAPSNEPPMDFDDD
IPF（配列番号590）
リンカー配列：
XTEN:
SGSETPGTSESATPES（配列番号604）
32aa：
SGGSSGGSSGSETPGTSESATPESSGGSSGGS（配列番号605）
SGGS：
SGGS（配列番号606）
$(GGS)_3$：
GGSGGSGGS（配列番号610）
【０５２１】
図109に示されている構築物のアミノ酸配列は下で規定されている：
【０５２２】
BE3：
MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLE
LYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESATP
ESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGN
IVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV
DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIA
LSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYID
GGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRR
QEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKG
ASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQK
KAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLI
NGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL
AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEG
IKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG
GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESIL

MLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKM
LSGGSPKKKRKV（配列番号174）
【０５２３】
BE4-pmCDA1：
[00523] MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWG
YAVNKPQSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQE
LRGNGHTLKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQ
LNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAVSGSETPGTSESATPESDKKYSIGL
AIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEK
YPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKS
NFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKA
PLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFI
KPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNR
EKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNF
DKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT
VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTR
SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN
NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY
SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSV
KELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQK
GNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILA
DANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNK
PESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV
（配列番号175）
【０５２４】
BE4-hAID：
MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVEL
LFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDR
KAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRR
ILLPLYEVDDLRDAFRTLGLSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKV
PSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSN
EMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK
ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAK
AILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY
DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDL
TLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKL
NREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLL
YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIEC
FDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQ

LIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHK
PENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY
LQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQI
LDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEK
NPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLY
LASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDEST
DENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV（配列番号176）

【０５２５】

BE4-3G：

MELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLT
IFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPW
NNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWV
LLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCA
QEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVD
HQGCPFQPWDGLDEHSQDLSGRLRAILQNQENSGSETPGTSESATPESDKKYSIGLAIGT
NSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPT
IYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQ
LFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPL
SASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREK
IEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT
VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLT
LTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK
VVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVE
NTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSD
KNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQ
LVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNY
HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSN
IMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVK
ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKG
NELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA
TLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKP
ESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV（
配列番号177）

【０５２６】

BE4-N：

MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSD
APEYKPWALVIQDSNGENKIKMLGGSSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKET

CLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGEC
SRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFV
NYSPSNEAHWPRYPHLWVRLYYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQR
LPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKV
LGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDD
SFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYL
ALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLS
KSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNL
LAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALV
RQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVY
NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLF
DDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLT
FKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEM
ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDM
YVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKN
YWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY
PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLI
ETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARK
KDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA
KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKL
KGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQA
ENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD
SGGSPKKKRKV（配列番号178）

【０５２７】

BE4-SSB：

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLE
LYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESATP
ESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGN
IVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV
DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIA
LSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYID
GGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRR
QEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKG
ASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQK
KAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLI
NGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL
AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEG
IKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG
GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE

QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSGGSGGSASRGVNKVILVGNL
GQDPEVRYMPNGGAVANITLATSESWRDKATGEMKEQTEWHRVVLFGKLAEVASEYLR
KGSQVYIEGQLRTRKWTDQSGQDRYTTEVVVNVGGTMQMLGGRQGGGAPAGGNIGGGQ
PQGGWGQPQQPQGGNQFSGGAQSRPQQSAPAAPSNEPPMDFDDDIPFSGGSPKKKRKV
（配列番号179）

【０５２８】

BE4-$(GGS)_3$：

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLE
LYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESATP
ESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGN
IVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV
DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIA
LSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYID
GGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRR
QEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKG
ASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQK
KAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLI
NGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL
AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEG
IKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG
GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSGGSGGSTNLSDIIEKETGKQL
VIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSN
GENKIKMLSGGSPKKKRKV（配列番号180）

【０５２９】

BE4-XTEN：

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLE
LYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESATP
ESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGN
IVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV

DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIA
LSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYID
GGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRR
QEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKG
ASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQK
KAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLI
NGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL
AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEG
IKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG
GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGSETPGTSESATPESTNLSDIIEKE
TGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVI
QDSNGENKIKMLSGGSPKKKRKV（配列番号181）

【０５３０】

BE4-32aa：

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLE
LYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGGSSGGSSGSETP
GTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIK
KNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESF
LVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRG
HFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIA
QLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK
EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGS
IPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSE
ETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYV
TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL
GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLK
RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQK
GQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKL
ITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGD
YKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV
WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG
GFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKD
LIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQ
KQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT

NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDI
IEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPW
ALVIQDSNGENKIKMLSGGSPKKKRKV（配列番号182）

【０５３１】

BE4-2XUGI：

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLE
LYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESATP
ESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGN
IVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV
DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIA
LSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYID
GGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRR
QEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKG
ASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQK
KAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLI
NGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL
AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEG
IKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG
GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESIL
MLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKM
LSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLL
TSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV（配列番号183）

【０５３２】

BE4：

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLE
LYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGGSSGGSSGSETP
GTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIK
KNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESF
LVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRG
HFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIA
QLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK
EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGS
IPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSE
ETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYV

TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL
GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLK
RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQK
GQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKL
ITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGD
YKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV
WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG
GFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKD
LIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQ
KQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT
NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSGGSGGS
TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAP
EYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEE
VIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKK
KRKV（配列番号184）

【０５３３】

EMX1、FANCF、HEK2、HEK3、HEK4、およびRNF2の標的シトシン残基を変異させる図109の塩基編集構築物の能力が図110～115に示されている。編集された標的シトシンのパーセンテージ（T、A、またはGに変異したC残基の割合を包含する）ならびに創出されたインデルの%が図110～115に示されている。編集された標的シトシンのパーセンテージは次式を用いて算出した：100-[Cを有するシーケンシングリードの%]。図110～115の円グラフは、指示されている種々の塩基を有するリードの分布を示しており、塩基編集されたリードの全て（指示されている塩基にC以外のヌクレオチドを有するリード）を眺めて、それらのいかなるパーセンテージがA、G、またはTを有するかを意味している。下の表9～14は図110～115の円グラフによって指示されている変異パーセンテージの値を示している。

【０５３４】

観察されるC→非T編集は蓋然的にはUDG（ウラシルDNAグリコシラーゼ）が原因である。例えば、ひとたびCがウラシル中間体に変換されると、UDGはそれを脱塩基部位へと変換し得る。それから、この脱塩基部位は他の内在性酵素によってプロセシングされ、究極的には、インデルまたはCを置き換える他の塩基（例えばGもしくはA）に至る。我々は、UDGノックアウト細胞株が増大したC→T編集を示し、インデルは少しから全くないことを示した。

表9 EMX1

【表７－１】

EMX1		C_5
BE3	A	1.9%
	C	51.9%
	G	6.0%
	T	40.3%
pmCDA1	A	0.2%
	C	88.9%
	G	0.8%
	T	10.0%
hAID	A	0.3%
	C	83.6%
	G	0.6%
	T	15.4%
hAPOBEC3G	A	0.0%
	C	98.9%

【表７－２】

	G	0.2%
	T	1.0%
BE4-N	A	1.4%
	C	85.6%
	G	3.5%
	T	9.5%
EMX1		
BE4-SSB	A	0.3%
	C	96.7%
	G	0.7%
	T	2.3%
BE4-$(GGS)_3$	A	1.7%
	C	36.8%
	G	4.1%
	T	57.4%
BE4-XTEN	A	2.1%
	C	45.7%
	G	5.1%
	T	47.1%
BE4-32aa	A	1.6%
	C	46.3%
	G	4.6%
	T	47.5%
BE4-2XUGI	A	0.6%
	C	60.2%
	G	1.9%
	T	37.3%

表10 FANCF

【表８－１】

FANCF		C_8
BE3	A	1.6%
	C	74.4%
	G	0.7%
	T	23.3%
pmCDA1	A	0.5%
	C	88.8%
	G	0.3%
	T	10.3%

【表８－２】

hAID	A	0.3%
	C	89.6%
	G	0.2%
	T	9.9%
hAPOBEC3G	A	2.1%
	C	75.4%
	G	10.4%
	T	12.1%
BE4-N	A	1.8%
	C	83.5%
	G	0.6%
	T	14.2%
BE4-SSB	A	0.3%
	C	97.1%
	G	0.1%
	T	2.6%
BE4-$(GGS)_3$	A	2.3%
	C	45.7%
	G	1.4%
	T	50.6%
BE4-XTEN	A	1.1%
	C	56.3%
	G	0.4%
	T	42.1%
BE4-32aa	A	1.2%
	C	70.0%
	G	0.6%
	T	28.2%
BE4-2XUGI	A	0.9%
	C	57.8%
	G	1.1%
	T	40.2%

表11 HEK2

【表９－１】

HEK2		C_6
BE3	A	0.9%

【表９－２】

	C	28.2%
	G	52.9%
	T	18.1%
pmCDA1	A	1.8%
	C	73.5%
	G	3.5%
	T	21.2%
hAID	A	2.1%
	C	56.9%
	G	7.7%
	T	33.3%
hAPOBEC3G	A	0.1%
	C	86.5%
	G	9.9%
	T	3.5%
BE4-N	A	1.0%
	C	37.6%
	G	57.0%
	T	4.4%
BE4-SSB	A	0.2%
	C	78.4%
	G	20.0%
	T	1.4%
BE4-$(GGS)_3$	A	0.6%
	C	11.1%
	G	40.6%
	T	47.7%
BE4-XTEN	A	1.2%
	C	24.8%
	G	44.6%
	T	29.4%
BE4-32aa	A	1.1%
	C	26.3%
	G	41.8%
	T	30.7%
BE4-2XUGI	A	0.8%

【表９－３】

	C	37.0%
	G	21.6%
	T	40.6%

表12 HEK3

【表１０－１】

HEK3		C_5
BE3	A	2.23%
	C	38.06%
	G	12.77%
	T	46.95%
pmCDA1	A	0.21%
	C	76.57%
	G	0.12%
	T	23.09%
hAID	A	0.28%
	C	60.23%
	G	1.03%
	T	38.45%
hAPOBEC3G	A	3.11%
	C	33.89%
	G	28.59%
	T	34.41%
BE4-N	A	2.6%
	C	64.1%
	G	13.5%
	T	19.8%
BE4-SSB	A	0.4%
	C	92.9%
	G	2.8%
	T	3.9%
BE4-$(GGS)_3$	A	1.3%
	C	9.9%
	G	7.9%
	T	80.8%
BE4-XTEN	A	2.3%
	C	15.9%
	G	12.2%

【表１０－２】

	T	69.6%
BE4-32aa	A	1.3%
	C	14.9%
	G	9.9%
	T	73.9%
BE4-2XUGI	A	0.6%
	C	23.4%
	G	3.8%
	T	72.2%

表13 HEK4

【表１１－１】

HEK4		C_5
BE3	A	8.40%
	C	41.89%
	G	24.54%
	T	25.17%
pmCDA1	A	0.50%
	C	87.53%
	G	0.01%
	T	11.95%
hAID	A	0.93%
	C	71.32%
	G	0.69%
	T	27.06%
hAPOBEC3G	A	0.12%
	C	99.37%
	G	0.35%
	T	0.16%
BE4-N	A	7.3%
	C	56.6%
	G	25.7%
	T	10.3%
BE4-SSB	A	2.1%
	C	86.8%
	G	5.8%
	T	5.2%

【表１１－２】

BE4-$(GGS)_3$	A	6.7%
	C	13.0%
	G	19.8%
	T	60.5%
BE4-XTEN	A	7.5%
	C	19.7%
	G	25.4%
	T	47.4%
BE4-32aa	A	7.9%
	C	21.8%
	G	25.1%
	T	45.3%
BE4-2XUGI	A	3.4%
	C	22.2%
	G	12.4%
	T	62.0%

表14 RNF2

【表１２－１】

RNF2		C_6
BE3	A	2.46%
	C	46.65%
	G	19.87%
	T	31.03%
pmCDA1	A	0.60%
	C	83.52%
	G	1.33%
	T	14.55%
hAID	A	0.36%
	C	75.03%
	G	3.20%
	T	21.40%
hAPOBEC3G	A	0.10%
	C	86.60%
	G	3.70%
	T	9.59%
BE4-N	A	5.1%
	C	50.0%

【表１２－２】

	G	28.8%
	T	16.2%
BE4-SSB	A	1.1%
	C	89.9%
	G	4.9%
	T	4.1%
BE4-$(GGS)_3$	A	2.0%
	C	23.0%
	G	14.0%
	T	61.0%
BE4-XTEN	A	2.6%
	C	32.4%
	G	16.0%
	T	49.0%
BE4-32aa	A	2.2%
	C	29.2%
	G	18.5%
	T	50.0%
BE4-2XUGI	A	0.7%
	C	45.0%
	G	6.5%
	T	47.8%

【０５３５】

例15：標的鎖を切断するCpf1ニッカーゼを含む塩基編集因子.

上で論じられる通り、本明細書において提供される融合蛋白質のいずれかの核酸によってプログラム可能なDNA結合蛋白質（napDNAbp）はCpf1蛋白質であり得る。いくつかの態様において、Cpf1蛋白質はCpf1ニッカーゼ（nCpf1）である。Cpf1ニッカーゼ、例えば、非標的鎖を切断するCpf1ニッカーゼ（AsCpf1のR1225A；およびLbCpf1のR1138A）はYamano et al.,"Crystal structure of Cpf1 in complex with guide RNA and target DNA." Cell (165) 2016, p.949-962に記載されており；これの内容全体は参照によってここに組み込まれる。しかしながら、標的鎖を切断するニッカーゼ（例えば、塩

基編集因子蛋白質のCpf1ニッカーゼ）は塩基編集効率を改善することが予想される。

【０５３６】

蛍光標識DNAを用いて、非標的鎖よりもむしろ標的鎖に選好的にニックを入れるCpf1変異体を同定した（図116を見よ）。図116では、DNA構築物1～3の上の鎖は非標的鎖であり、下の鎖は標的鎖である。野生型LbCpf1、R836A（LbCpf1）、およびR1138A（LbCpf1）を用いてin vitroアッセイが実施される。R836A（LbCpf1）は「不具化された」ニッカーゼであるように見え、それが標的鎖を非標的鎖よりも効率的に切ることを意味する。図117に示されている通り、非標的鎖は切られていず、蛍光性の350切片は観察されない。2時間後には、両方の鎖が切られている。異なる強度は、非標的鎖よりも多くの標的鎖が切られていることを示唆している。

【０５３７】

AsCp1-BE3による塩基編集ウィンドウを確立する

napDNAbpとしてLbCpf1（R836A）またはAsCpf1（R912A）を有する塩基編集蛋白質（例えばBE3（SpCas9-BE3））は低い効率で塩基を編集することが示された（0.1%～0.4%）。AsCpf1（R912A）napDNAbpを有する塩基編集因子は、EMX1、FANCF、HEK3、およびHEK4部位の標的Cをより効率的に変異させた。試験した構築物の編集ウィンドウは第7の塩基から第11の塩基までであるように見える。数は、最も高い数を有するBE3およびより低いものを有する自滅型のBE（すなわち、非標的鎖を切断するAPOBEC-AsCpf1(R1225A)-UGI）による傾向と整合している。EMX1に対するCas9-BE3による図118陽性対照：5～6%を見よ。AsCpf1のインデル値：＞20%。AsCpf1のR912はCpf1ファミリーの多くのメンバーにおいて保存されている。LbCpf1の対応する残基はR836であり、これはRがAに変異しているときには「不具化された」ニッカーゼであると信じられる。

【０５３８】

Cpf1-BEの最適化（リンカー）

インデルデータはCpf1がDNA標的部位にアクセスし得ることを示唆している。それゆえに、Cpf1塩基編集蛋白質の最適化は具体的なAPOBEC蛋白質、リンカー、および／またはUGIドメインにフォーカスした。LbCpf1(R836A)およびAsCpf1(R912A)両方を用いて、様々なリンカーによって、図119に示されている構築物を試験した。短く言うと、APOBECおよびCpf1ドメイン（例えばAsCpf1またはLbCpf1）間の異なるリンカー配列（すなわち、XTEN、GGS、$(GGS)_3$（配列番号610）、$(GGS)_5$（配列番号610）、および$(GGS)_7$（配列番号610））を試験した。図120を見よ。構築物を、TGCTTCTC$_8$CAGCCCTGGCCTGG（配列番号592）であるHEK3部位のC$_8$残基を変異させるそれらの能力について試験した。AsCpf1を有する塩基編集蛋白質の編集レベルは1%超に達し、LbCpf1を有する塩基編集蛋白質は塩基編集効率の比較的な縮減を示した。図121に示されている通り、Centre of Integrative Bioinformatics VUによって維持されているデータベースからのリンカーは、AsCpf1-BE3についてGGS型リンカーほどの有意な改善を示さなかった。図121に示されているリンカーを下に示す：

【０５３９】

PDBコード	長さ（aa）	配列

【０５４０】

1au7A_1　10　KRRTTISIAA（配列番号593）

【０５４１】

1c1kA_1　19　ALVFYREYIGRLKQIKFKF（配列番号594）

【０５４２】

1c20A_1　14　LPIMAKSVLDLYEL（配列番号595）

【０５４３】

1ee8A_1　5　LLRLG（配列番号596）

【０５４４】

1f1zA_1　15　TDKEINPVVKENIEW（配列番号597）

【０５４５】

1ignA_1 8 PPSIKRKF（配列番号598）

【０５４６】

1jmcA_1 9 LPTVQFDFT（配列番号599）

【０５４７】

1sfe_1 14 LPLDIRGTAFQQQV（配列番号600）

【０５４８】

2ezx_1 8 AYVVLGQF（配列番号601）

【０５４９】

2reb_1 8 INFYGELV（配列番号602）

【０５５０】

Cpf1-BEの最適化（向き）

Cas9はCおよびN末端の間にアミノ酸のストレッチを有するが（図123の赤色の四角形を見よ）、AsCpf1は有さない（図122を見よ）。その上、AsCpf1はより短い距離をNおよびC末端間に有する。これらの観察はAPOBEC（N末端上）とUGI（C末端上）との間の見込まれる干渉を指示しており、これによってUGIは非標的鎖へのAPOBECアクセスを障害し得る。1つの解決策は、APOBECおよびUGIを同じ末端、NまたはCどちらかに動かすことである。従って、アーキテクチャNLS-UGI-APOBEC-XTEN-AsCpf1；UGI-APOBEC-XTEN-AsCpf1-NLS；およびAsCpf1-XTEN-APOBEC-NLSを有する構築物が試験されるであろう。

【０５５１】

Cpf1-BEの最適化（内部短縮）

非標的鎖が分離されているCpf1の公知の結晶構造はない（図124シアンを見よ）。図124に示される通り、編集ウィンドウは赤色の円内に在るはずであることが信じられる。APOBECを妨害しようとし得るヘリックス状領域（図124の四角形を見よ）がある。この領域はアミノ酸配列K(661)KTGDQK(667)（配列番号603）を含む。

【０５５２】

2つ、4つ、または6つの残基の取り除きが塩基編集効率を改善するかどうかを試験するために、標的部位としてHEK3を用いて、AsCpf1(R912A) napDNAbpを有する塩基編集因子によって実験を執り行った。T663およびD665が欠失しているときに、編集レベルはおよそ2.6%まで増大する-対照レベルからの6倍の増大（下の表7を見よ）。この実験に用いた構築物はAPOBEC-XTEN-AsCpf1(R912A)-SGGS-UGIであった。

表7

【表１３】

欠失	C8の編集	C9の編集
T663, D665	2.59%	1.29%
K662, T663, D665, Q666	0.15%	0.15%
K661, K662, T663, D665, Q666, K667	0.22%	0.21%

参考文献

【表１４】

1. Humbert O, Davis L, Maizels N. Targeted gene therapies: tools, applications, optimization. *Crit Rev Biochem Mol.***2012;** 47(3):264-81. PMID: 22530743.
2. Perez-Pinera P, Ousterout DG, Gersbach CA. Advances in targeted genome editing. *Curr Opin Chem Biol.* **2012;** 16(3-4):268-77. PMID: 22819644.
3. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet.* **2010;** 11(9):636-46. PMID: 20717154.
4. Joung JK, Sander JD. TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol.* **2013;** 14(1):49-55. PMID: 23169466.
5. Charpentier E, Doudna JA. Biotechnology: Rewriting a genome. *Nature.* **2013;** 495, (7439):50-1. PMID: 23467164.
6. Pan Y, Xia L, Li AS, Zhang X, Sirois P, Zhang J, Li K. Biological and biomedical applications of engineered nucleases. *Mol Biotechnol.* **2013;** 55(1):54-62. PMID: 23089945.
7. De Souza, N. Primer: genome editing with engineered nucleases. *Nat Methods.* **2012;** 9(1):27. PMID: 22312638.
8. Santiago Y, Chan E, Liu PQ, Orlando S, Zhang L, Urnov FD, Holmes MC, Guschin D, Waite A, Miller JC, Rebar EJ, Gregory PD, Klug A, Collingwood TN. Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proc Natl Acad Sci U S A.* **2008;** 105(15):5809-14. PMID: 18359850.
9. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley

【表１５】

GQ, Lander ES. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet.* **1999;** 22(3):231-8. PMID: 10391209.
10. Jansen R, van Embden JD, Gaastra W, Schouls LM. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol.* **2002;** 43(6):1565-75. PMID: 11952905.
11. Mali P, Esvelt KM, Church GM. Cas9 as a versatile tool for engineering biology. *Nat Methods.* **2013;** 10(10):957-63. PMID: 24076990.
12. Jore MM, Lundgren M, van Duijin E, Bultema JB, Westra ER, Waghmare SP, Wiedenheft B, Pul U, Wurm R, Wagner R, Beijer MR, Barendregt A, Shou K, Snijders AP, Dickman MJ, Doudna JA, Boekema EJ, Heck AJ, van der Oost J, Brouns SJ. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol.* **2011;** 18(5):529-36. PMID: 21460843.
13. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. *Science.* **2010;** 327(5962):167-70. PMID: 20056882.
14. Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. *Nature.* **2012;** 482(7385):331-8. PMID: 22337052.
15. Gasiunas G, Siksnys V. RNA-dependent DNA endonuclease Cas9 of the CRISPR system: Holy Grail of genome editing? *Trends Microbiol.* **2013;** 21(11):562-7. PMID: 24095303.
16. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell.* **2013;** 152(5):1173-83. PMID: 23452860.
17. Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, Thakore PI, Glass KA, Ousterout DG, Leong KW, Guilak F, Crawford GE, Reddy TE, Gersbach CA. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods.* **2013;** 10(10):973-6. PMID: 23892895.
18. Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, Yang L, Church GM. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol.* **2013;** 31(9):833-8. PMID: 23907171.
19. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS, Qi LS. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell.* **2013;** 154(2):442-51. PMID: 23849981.
20. Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc.* **2013;** 8(11):2180-96. PMID: 24136345.
21. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. RNA-guided human genome engineering via Cas9. *Science.* **2013;** 339(6121):823-6. PMID: 23287722.

【表１６】

22. Cole-Strauss A, Yoon K, Xiang Y, Byrne BC, Rice MC, Gryn J, Holloman WK, Kmiec EB. Correction of the mutation responsible for sickle cell anemia by an RNA-DNA oligonucleotide. *Science.* **1996;** 273(5280):1386-9. PMID: 8703073.
23. Tagalakis AD, Owen JS, Simons JP. Lack of RNA-DNA oligonucleotide (chimeraplast) mutagenic activity in mouse embryos. *Mol Reprod Dev.* **2005;** 71(2):140-4. PMID: 15791601.
24. Ray A, Langer M. Homologous recombination: ends as the means. *Trends Plant Sci.* **2002;** 7(10):435-40. PMID 12399177.
25. Britt AB, May GD. Re-engineering plant gene targeting. *Trends Plant Sci.* **2003;** 8(2):90-5. PMID: 12597876.
26. Vagner V, Ehrlich SD. Efficiency of homologous DNA recombination varies along the Bacillus subtilis chromosome. *J Bacteriol.* **1988;** 170(9):3978-82. PMID: 3137211.
27. Saleh-Gohari N, Helleday T. Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human cells. *Nucleic Acids Res.* **2004;** 32(12):3683-8. PMID: 15252152.
28. Lombardo A, Genovese P, Beausejour CM, Colleoni S, Lee YL, Kim KA, Ando D, Urnov FD, Galli C, Gregory PD, Holmes MC, Naldini L. Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol.* **2007;** 25(11):1298-306. PMID: 17965707.
29. Conticello SG. The AID/APOBEC family of nucleic acid mutators. *Genome Biol.* **2008;** 9(6):229. PMID: 18598372.
30. Reynaud CA, Aoufouchi S, Faili A, Weill JC. What role for AID: mutator, or assembler of the immunoglobulin mutasome? *Nat Immunol.* **2003;** 4(7):631-8.
31. Bhagwat AS. DNA-cytosine deaminases: from antibody maturation to antiviral defense. *DNA Repair (Amst).* **2004;** 3(1):85-9. PMID: 14697763.
32. Navaratnam N, Sarwar R. An overview of cytidine deaminases. *Int J Hematol.* **2006;** 83(3):195-200. PMID: 16720547.
33. Holden LG, Prochnow C, Chang YP, Bransteitter R, Chelico L, Sen U, Stevens RC, Goodman MF, Chen XS. Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature.* **2008;** 456(7218):121-4. PMID: 18849968.
34. Chelico L, Pham P, Petruska J, Goodman MF. Biochemical basis of immunological and retroviral responses to DNA-targeted cytosine deamination by activation-induced cytidine deaminase and APOBEC3G. *J Biol Chem.* **2009;** 284(41). 27761-5. PMID: 19684020.
35. Pham P, Bransteitter R, Goodman MF. Reward versus risk: DNA cytidine deaminases triggering immunity and disease. *Biochemistry.* **2005;** 44(8):2703-15. PMID 15723516.
36. Chen X, Zaro JL, Shen WC. Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev.* **2013;** 65(10):1357-69. PMID: 23026637.
37. Lee JW, Soung YH, Kim SY, Lee HW, Park WS, Nam SW, Kim SH, Lee JY, Yoo NJ, Lee SH. PIK3CA gene is frequently mutated in breast carcinomas and hepatocellular carcinomas. *Oncogene.* **2005;** 24(8):1477-80. PMID: 15608678.

【表１７】

38. Ikediobi ON, Davies H, Bignell G, Edkins S, Stevens C, O'Meara S, Santarius T, Avis T, Barthorpe S, Brackenbury L, Buck G, Butler A, Clements J, Cole J, Dicks E, Forbes S, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Hunter C, Jenkinson A, Jones D, Kosmidou V, Lugg R, Menzies A, Mironenko T, Parker A, Perry J, Raine K, Richardson D, Shepherd R, Small A, Smith R, Solomon H, Stephens P, Teaque J, Tofts C, Varian J, Webb T, West S, Widaa S, Yates A, Reinhold W, Weinstein JN, Stratton MR, Futreal PA, Wooster R. Mutation analysis of 24 known cancer genes in the NCI-60 cell line set. *Mol Cancer Ther.* **2006;** 5(11):2606-12. PMID: 17088437.

39. Cox, D. B., Platt, R. J. & Zhang, F. Therapeutic genome editing: prospects and challenges. *Nature medicine* **21**, 121-131, doi:10.1038/nm.3793 (2015).

40. Hilton, I. B. & Gersbach, C. A. Enabling functional genomics with genome engineering. *Genome research* **25**, 1442-1455, doi:10.1101/gr.190124.115 (2015).

41. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature biotechnology* **32**, 347-355, doi:10.1038/nbt.2842 (2014).

42. Maruyama, T. *et al.* Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nature biotechnology* **33**, 538-542, doi:10.1038/nbt.3190 (2015).

43. Chu, V. T. *et al.* Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nature biotechnology* **33**, 543-548, doi:10.1038/nbt.3198 (2015).

44. Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* **3**, e04766, doi:10.7554/eLife.04766 (2014).

45. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823, doi:10.1126/science.1231143 (2013).

46. Rong, Z., Zhu, S., Xu, Y. & Fu, X. Homologous recombination in human embryonic stem cells using CRISPR/Cas9 nickase and a long DNA donor template. *Protein & cell* **5**, 258-260, doi:10.1007/s13238-014-0032-5 (2014).

47. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).

48. Harris, R. S., Petersen-Mahrt, S. K. & Neuberger, M. S. RNA Editing Enzyme APOBEC1 and Some of Its Homologs Can Act as DNA Mutators. *Molecular Cell* **10**, 1247-1253 (2002).

49. Jinek, M. *et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997, doi:10.1126/science.1247997 (2014).

50. Schellenberger, V. *et al.* A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. *Nature biotechnology* **27**, 1186-1190, doi:10.1038/nbt.1588 (2009).

【表１８】

51. Saraconi, G., Severi, F., Sala, C., Mattiuz, G. & Conticello, S. G. The RNA editing enzyme APOBEC1 induces somatic mutations and a compatible mutational signature is present in esophageal adenocarcinomas. *Genome biology* **15**, 417- (2014).
52. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569-573, doi:10.1038/nature13579 (2014).
53. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823 (2013).
54. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR- Cas nucleases. *Nature biotechnology* **33**, 187-197, doi:10.1038/nbt.3117 (2015).
55. Kunz, C., Saito, Y. & Schar, P. DNA Repair in mammalian cells: Mismatched repair: variations on a theme. *Cellular and molecular life sciences : CMLS* **66**, 1021-1038, doi:10.1007/s00018-009- 8739-9 (2009).
56. D., M. C. *et al.* Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. *Cell* **82**, 701-708 (1995).
57. Caldecott, K. W. Single-strand break repair and genetic disease. *Nature reviews. Genetics* **9**, 619-631, doi:10.1038/nrg2380 (2008).
58. Fukui, K. DNA mismatch repair in eukaryotes and bacteria. *Journal of nucleic acids* **2010**, doi:10.4061/2010/260512 (2010).
59. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences* **109**, E2579–E2586, doi:10.1073/pnas.1208507109 (2012).
60. Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186-191, doi:10.1038/nature14299 (2015).
61. Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nature biotechnology* **32**, 677-683, doi:10.1038/nbt.2916 (2014).
62. Wu, X. *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature biotechnology* **32**, 670-676, doi:10.1038/nbt.2889 (2014).
63. Beale, R. C. L. *et al.* Comparison of the Differential Context-dependence of DNA Deamination by APOBEC Enzymes: Correlation with Mutation Spectra in Vivo. *Journal of Molecular Biology* **337**, 585-596, doi:10.1016/j.jmb.2004.01.046 (2004).
64. Kim, J., Basak, J. M. & Holtzman, D. M. The role of apolipoprotein E in Alzheimer's disease. *Neuron* **63**, 287-303, doi:10.1016/j.neuron.2009.06.026 (2009).
65. Liu, C. C., Kanekiyo, T., Xu, H. & Bu, G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nature reviews. Neurology* **9**, 106-118, doi:10.1038/nrneurol.2012.263 (2013).
66. Sjöblom, T. *et al.* The Consensus Coding Sequences of Human Breast and Colorectal Cancers. *Science* **314**, 268-274, doi:10.1126/science.1133427 (2006).

【表１９】

67. Stephens, P. J. *et al.* The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486**, 400-404, doi:10.1038/nature11017 (2012).
68. Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Research*, doi:10.1093/nar/gkv1222 (2015).
69. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science*, doi:10.1126/science.aad5227 (2015).
70. Davis, K. M., Pattanayak, V., Thompson, D. B., Zuris, J. A. & Liu, D. R. Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nature chemical biology* **11**, 316-318, doi:10.1038/nchembio.1793 (2015).
71. Zuris, J. A. *et al.* Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nature biotechnology* **33**, 73-80, doi:10.1038/nbt.3081 (2015).
72. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481-485, doi:10.1038/nature14592 (2015).
73. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA- programmed Cas9 nuclease specificity. *Nature Biotechnology* **31**, 839-843, doi:10.1038/nbt.2673 (2013).
74. Shcherbakova, D. M. & Verkhusha, V. V. Near-infrared fluorescent proteins for multicolor in vivo imaging. *Nature Methods* **10**, 751-754, doi:10.1038/nmeth.2521 (2013).
75. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protocols* **8**, 2281-2308, doi:10.1038/nprot.2013.143 (2013).
76. Jiang, F. *et al.* Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science*, doi:10.1126/science.aad8282 (2016).
77. Tsai, S. Q. *et al.* Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotech* **32**, 569-576, doi:10.1038/nbt.2908 (2014).
78. Lieber, M. R., Ma, Y., Pannicke, U. & Schwarz, K. Mechanism and regulation of human non-homologous DNA end-joining. *Nat Rev Mol Cell Biol* **4**, 712-720 (2003).
79. Heller, R. C. & Marians, K. J. Replisome assembly and the direct restart of stalled replication forks. *Nat Rev Mol Cell Biol* **7**, 932-943 (2006).
80. Pluciennik, A. *et al.* PCNA function in the activation and strand direction of MutLα endonuclease in mismatch repair. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 16066-16071, doi:10.1073/pnas.1010662107 (2010).
81. Seripa, D. *et al.* The missing ApoE allele. *Annals of human genetics* **71**, 496-500, doi:10.1111/j.1469-1809.2006.00344.x (2007).
82. Kleinstiver, B. P. *et al.* High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490-495, doi:10.1038/nature16526 (2016).
83. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotech* **34**, 339-344, doi:10.1038/nbt.3481 (2016).

【表２０】

84. Simonelli, V., Narciso, L., Dogliotti, E. & Fortini, P. Base excision repair intermediates are mutagenic in mammalian cells. *Nucleic acids research* **33**, 4404-4411, doi:10.1093/nar/gki749 (2005).
85. Barnes, D. E. & Lindahl, T. Repair and Genetic Consequences of Endogenous DNA Base Damage in Mammalian Cells. *Annual Review of Genetics* **38**, 445-476, doi:doi:10.1146/annurev.genet.38.072902.092448 (2004).

【０５５３】
例16：蛋白質工学および蛋白質送達によって塩基編集のDNA特異性および適用性を改善する
二本鎖DNA切断、過剰な確率的な挿入および欠失、または相同組み換え修復への依存性なしに、1つの塩基対から別のものへのプログラム可能な変換を可能にするゲノム編集アプローチ、塩基編集が開発された。塩基編集の用途はオフターゲット活性と細胞内DNA送達への依拠とによって限定される。ここでは、これらの限定に対処する2つの進歩が記載される。第1に、第3世代塩基編集因子（BE3）に変異をインストールすることによってオフターゲット塩基編集を縮減して、高フィデリティ塩基編集因子（HF-BE3）を創出した。次に、BE3およびHF-BE3を精製し、リボヌクレオ蛋白質（RNP）複合体として哺乳類細胞に送達し、DNA不使用の塩基編集を確立する。BE3のRNP送達は、同等のオンターゲット編集レベルを維持しながら、HF-BE3のプラスミドトランスフェクションよりも高い特異性さえも授ける。最後に、これらの進歩を適用して、BE3 RNPをゼブラフィッシュ胚および生体マウスの内耳両方に送達して、in vivoの特異的なDNA不使用の塩基編集を達成する。
【０５５４】
導入
従来のゲノム編集薬剤はゲノム編集の第1ステップとして二本鎖DNA切断（DSB）を導入する[1～4]。細胞はDSBに対して主として非相同末端結合（NHEJ）によって応答し、確率的な挿入または欠失（インデル）を切断部位にもたらす[1,5]。より正確な変化をゲノムDNAに創出するためには、相同組み換え修復（HDR）を用いて、外来的に供給されるDNAドナー鋳型のものによって、切断部位の周辺のゲノムDNAを置き換え得る[6～8]。残念ながら、典型的には、HDRは競合するNHEJからもたらされる過剰なインデルを伴い、主として有糸分裂細胞に限定される。加えて、ほとんどのゲノム編集方法は、哺乳類細胞への外来性プラスミドまたはウイルスDNAの送達、次に薬剤の細胞内発現に依拠する[9～12]。これらの送達方法は、オンターゲット座位が編集された後でさえも連続的なコントロールされないCas9およびsgRNA発現をもたらし、オフターゲット座位におけるゲノム編集の機会を増大させる[1,13]。
【０５５５】
C:G塩基対からT:A塩基対への直接的なプログラム可能な標的化された変換を可能にするゲノム編集の異なるアプローチ、塩基編集が最近記載された[3,14]。第3世代塩基編集因子BE3は、1つの蛋白質中に（i）ガイドRNAによって指定される座位において小さい一本鎖DNAバブルを開裂させる触媒的に棄損されたCas9、（ii）一本鎖DNAバブル中のおよそ5ヌクレオチドのウィンドウ内のCをUに変換するテザリングされた一本鎖特異的シチジンデアミナーゼ、（iii）塩基除去修復を阻害し、それによって塩基編集の効率および生成物選択性を改善するウラシルグリコシラーゼ阻害因子（UGI）、ならびに（iv）G含有DNA鎖を置き換えるように細胞性のミスマッチ修復を操るためのニッカーゼ活性、を含有する。これらのコンポーネントの組み合わせは哺乳類細胞における効率的かつ永続的なC→T（またはG→A）変換を可能にし、最小限のインデル形成を有する。先に（prefiously）報告されてから[14]、他の研究者は、哺乳類細胞[15～17]および植物[18]におけるCas9によって導かれるC→T変換を容易化するこの戦略および関係するアプローチの能力を確認している。
【０５５６】
ここでは、塩基編集のDNA特異性を多大に改善し、かつホストゲノムとの組換えのリスクおよび細胞傷害性と関連付けられている外来性DNAを供給することなしにin vitroおよびin vivoの塩基編集を許す2つの進歩が記載される[18,19]。第1に、オフターゲット編集イベント縮減し、オンターゲット編集活性の控えめな減少のみを有する、Cas9のDNA親和性を減少させることが公知の変異を組み込むBE3の変異体形態[20]を操作する。次に、ガイドRNAと複合体化した塩基編集因子蛋白質の脂質によって媒介される送達は、プラスミドDNA送達と比較して、オンターゲット塩基編集の明らかな縮減なしにより大きい特異性向上さえももたらすことを明らかにする。典型的には、RNPとしての塩基編集因子の送達は、

培養ヒトおよびマウス細胞において、高度に反復的なゲノムDNA配列を標的化する悪評あるプロミスキャスなガイドRNAについてさえも、測定可能なレベルよりも下までオフターゲット編集を縮減する。これらの進歩は、マウスおよびゼブラフィッシュにおける高度に特異的なDNA不使用のin vivo塩基編集が実証されることを可能にする。

【０５５７】

結果

高フィデリティ塩基編集因子を操作する

Cas9ヌクレアーゼおよびそれらの関連融合体構築物はオフターゲットゲノム座位のDNAに結合し切断することが示されている[21～24]。Joungらは4つの点変異（N497A、R661A、Q695A、Q926A）を含有する高フィデリティSpCas9バリアントHF-Cas9を開発した。これらは、Cas9とDNA標的鎖のリン酸バックボーンとの間の非特異的相互作用を消去するように設計されており（図125A）[20]、それらのDNA切断特異性を多大に増大させるTALENにおける非特異的DNA相互作用の先の無効化と整合している[25]。塩基編集因子はCas9によって作出される一本鎖DNAバブル内の非標的鎖に対して働くので[14]、「HF-BE3」を創出するためにHF-Cas9からのこれらの4つの点変異をBE3に導入することは、その塩基変換性能を変改することなしにオフターゲット塩基編集を縮減し得るという仮説が立てられ得る（図125Aおよび125C）。

【０５５８】

BE3およびHF-BE3をHis6タグ付き蛋白質としてコードするプラスミドをE. coliによって過剰発現し、第1にニッケルアフィニティークロマトグラフィーによって、それからカチオン交換クロマトグラフィーによって精製した（図126A～126B）。発現および精製条件の大がかりな最適化後に、BE3およびHF-BE3蛋白質は培養培地のリットルあたり～2mgの収率で普通に生成され得る（図126A～126C）。

【０５５９】

精製された塩基編集因子蛋白質を用いて、HF-BE3およびBE3の塩基編集効率および編集ウィンドウの幅を生化学的に比較した。TC反復を含むプロトスペーサーを有する合成dsDNA 79merにおいて、in vitroC→U変換効率を測定した。標的dsDNA（250nM）をBE3:sgRNAまたはHF-BE3:sgRNA（２μＭ）と30minに渡って37℃でインキュベーションした。インキュベーション後に、編集されたDNAをウラシルに対して寛容なポリメラーゼを用いて増幅し、高スループットDNAシーケンシング（HTS）によってシーケンシングした。同等の編集効率および活性ウィンドウ幅がin vitroのHF-BE3およびBE3で観察された（図125B）。これらの知見は、BE3への高フィデリティ変異の導入が、固有のオンターゲット塩基編集効率を損なうことまたはin vitroのもたらされるHF-BE3蛋白質の編集ウィンドウの幅を変化させることをしないことを指示している。

【０５６０】

HF-BE3はDNAトランスフェクション後の編集特異性を向上させる

次に、BE3およびHF-BE3の塩基編集効率、特異性、および編集ウィンドウ幅をプラスミドDNAトランスフェクション後の哺乳類細胞において比較した。4つの良く研究された内在性のゲノム座位（HEK293部位3、FANCF、EMX1、およびVEGFA部位2）を、哺乳類細胞におけるオンおよびオフターゲット塩基編集を取り調べるために選んだ[14,24]。VEGFA部位2は高度に反復的であり、20個のプロトスペーサーヌクレオチドのうち14個のCを含有し、公知のオフターゲットゲノム編集の悪評ある高い率に関連している[20,22,24,26]。この部位は、それが恐るべき特異性の難題を突き付けるので、包含されるように選ばれた。ほとんどのヌクレアーゼに基づくゲノム編集用途とは対照的に、塩基編集は、標的ヌクレオチドを編集ウィンドウ内に置くためにプロトスペーサーの正確な場所に依拠し、通常は、ガイドRNAの選択肢の柔軟性は少し利用可能であるかまたは利用可能ではない。よって、高度に反復的なプロミスキャスなsgRNA標的についてさえも向上した特異性を有する塩基編集因子の開発は重大である[3,14]。

【０５６１】

4つの塩基編集因子:sgRNA対のそれぞれによるプラスミドトランスフェクション[24]後に

、オンターゲット座位および公知のオフターゲット座位をPCRによって増幅し、HTSによって分析した。これらの4つの内在性のゲノム座位のHEK293T細胞におけるオンターゲット編集はHF変異の導入によって軽度に縮減され；編集はBE3では平均29±5%、HF-BE3では21±3%であった（n=3つの生物学的レプリケートの平均±s.e.m.）（図127A～127D、128A）。

【０５６２】

3つの標準的な非反復的な標的部位（HEK293部位3、FANCF、およびEMX1）のそれぞれについて、Cas9および同じガイドRNAによる処置から改変されることが先に報告されたオフターゲット座位から、編集ウィンドウ内にCを含有する3つの最も高頻度に改変されるオフターゲット座位を調べた（表15）[24]。BE3プラスミドによって細胞をトランスフェクションしたときには、HEK293部位3、FANCF、およびEMX1の9つの最も高頻度に改変されるCas9オフターゲット座位におけるC→T変換は平均1.1±0.3%であった（図127A～C；n=3つの生物学的レプリケートの平均±s.d.）。HF変異のインストールは、平均オフターゲット編集の絶対レベルを0.03±0.005%まで37倍縮減し、測定可能なオフターゲットC→T変換の1つの場合のみを有した（図127A；EMX1 オフターゲット1のC_5）。

【表２１－１】

部位	配列	配列番号	GUIDE-Seq カウント
EMX1	GAGTC$_5$C$_6$GAGCAGAAGAAGAA**GGG**	480	4,521
EMX1 オフターゲット 1	GAGTC$_5$TAAGCAGAAGAAGAA**GAG**	481	1,445
EMX1 オフターゲット 2	GAGGC$_5$C$_6$GAGCAGAAGAAAGA**CGG**	482	700
EMX1 オフターゲット 3	GAGTC$_5$C$_6$TAGCAGGAGAAGAA**GAG**	483	390
HEK293 部位 3 オンターゲット	GGCC$_4$C$_5$AGACTGAGCACGTGA**TGG**	484	2,074
HEK293 部位 3 オフターゲット 1	CACC$_4$C$_5$AGACTGAGCACGTGC**TGG**	485	327

【表２１－２】

HEK293 部位3オフターゲット 2	GACAC$_5$AGACTGGGCACGTGA**GGG**	486	306
HEK293 部位 3 オフターゲット 3	AGCTC$_5$AGACTGAGCAAGTGA**GGG**	487	136
VEGFA 部位 2 オンターゲット	GAC$_3$C$_4$C$_5$C$_6$C$_7$TC$_9$C$_{10}$ACCCCGCCTCC**GG**	488	540
VEGFA 部位 2オフターゲット 1	CTAC$_4$C$_5$C$_6$C$_7$TC$_9$C$_{10}$ACCCCGCCTC**CG G**	489	1,925
VEGFA 部位 2 オフターゲット 2	ATTC$_4$C$_5$C$_6$C$_7$C$_8$C$_9$C$_{10}$ACCCCGCCTC**AG G**	490	1,549
VEGFA 部位2 オフターゲット 3	ACAC$_4$C$_5$C$_6$C$_7$C$_8$C$_9$C$_{10}$ACCCCGCCTCA**GG**	491	1,178
VEGFA 部位 2 オフターゲット4	TGC$_3$C$_4$C$_5$C$_6$C$_7$C$_8$C$_9$C$_{10}$ACCCCACCTCT**GG**	492	1,107
FANCF オンターゲット	GGAATC$_6$C$_7$C$_8$TTC$_{11}$TGCAGCACC**TGG**	493	4,816
FANCF オフターゲット 1	GGAAC$_5$C$_6$C$_7$C$_8$GTC$_{11}$TGCAGCACC**AG G**	494	2,099
FANCF オフターゲット 2	GGAGTC$_6$C$_7$C$_8$TC$_{10}$C$_{11}$TACAGCACC**AG G**	495	524
FANCF オフターゲット 3	AGAGGC$_6$C$_7$C$_8$C$_9$TC$_{11}$TGCAGCACC**AG G**	496	150

表15：本研究において研究されたオンおよびオフターゲットヒトゲノム座位のプロトスペーサーおよびPAM配列。オフターゲット部位はそれらのGUIDE-Seqリードカウントに基づいて選んだ[45]。特定のsgRNAの編集ウィンドウ内のシトシンが付番されている。PAM配列は太字で示されている。それらのそれぞれのオンターゲット座位とは異なるオフターゲット座位のプロトスペーサー塩基は下線付きにしている。マウスサンプルにおいて取り調べたゲノム配列については図132Eを見よ。

【０５６３】

極度に難題である部位におけるHF-BE3特異性を特徴付けるために、高度に反復的なVEGFA部位2座位を標的化するときのBE3およびHF-BE3オフターゲット活性を比較した。BE3処置は、このsgRNAに関連する4つの試験されたオフターゲット部位の活性ウィンドウ内に配置されたシトシンの平均で15+5%の編集に至る（このパラグラフにおいて引用される全ての平均値はn=3つの生物学的レプリケートの平均±s.d.に相当する）。対照的に、HF-BE3は同じオフターゲット部位における絶対的なオフターゲット編集の3倍縮減（5.0±2.3%）に至る（図127D）。BE3のトランスフェクションと比較したときに、HF-BE3は、オフターゲット座位に配置された57個のシトシンの27個においてオフターゲット編集を有意に（p＜0.05，スチューデントの両側t検定）縮減したが（表16）、HF-BE3処置は、16個の取り調べられたオンターゲットシトシン残基の3つのみにおいてオンターゲット編集の有意な縮減（p＜0.05，スチューデントの両側t検定）に至る。

【表２２】

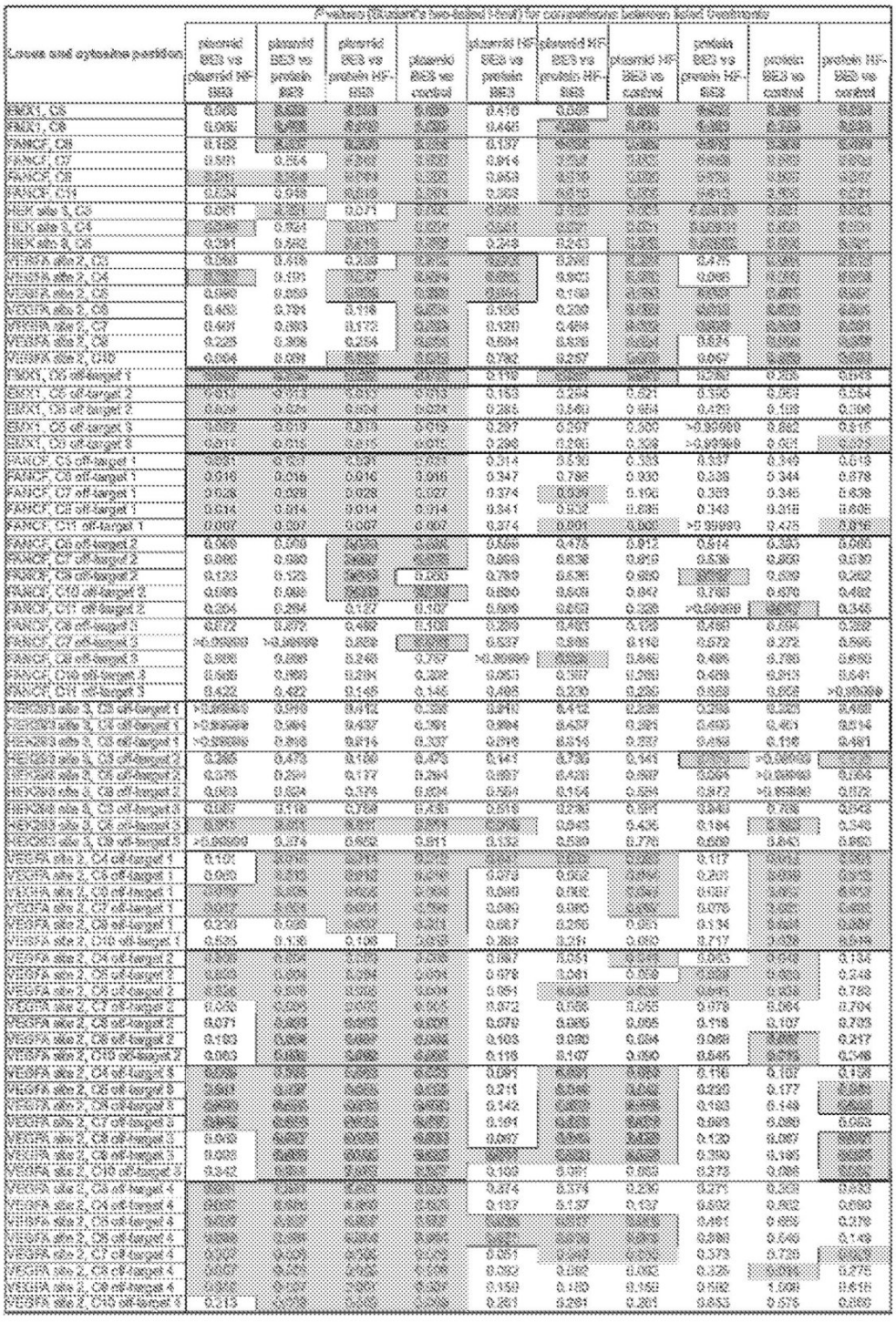


表16：本研究において評価された全ての座位の異なる処置条件における塩基編集の違いのp値。p値は材料と方法に記載されている通りスチューデントの両側t検定を用いて算出した。p値が有意な違い（$p<0.05$）を指示したときには、対応するエントリーを強調した。

【０５６４】

オフターゲット座位における絶対的な編集間の違いを考えることに加えて、観察されたオンターゲット効率をオフターゲット効率によって割ることによって、オンターゲット：オフターゲット編集特異性比をもまた算出した（図129A～129B）。このメトリックは、HF変異のインストールに関連するオンターゲット編集のいずれかの縮減を考慮しており、塩基編集の効率および特異性両方に対して敏感な用途にとって有用である。HF-BE3によ

るオフターゲット編集は、いくつかのオフターゲット座位では高スループットシーケンシングの検出限界よりも下であった。これらのケースでは、検出上限に等しい控え目なオフターゲット編集効率を想定した（0.025%C→T変換；方法を見よ）。この分析に基づくと、2つの構築物のプラスミド送達を行なったときには、本明細書において調べられた全ての34個のシトシンにおけるHF変異のインストールによる特異性比の平均の改善は19倍であった。総合すると、これらの結果は、非反復的な部位（図129A）および高度に反復的な部位（図129B）では、HF-BE3は、BE3と比較して、実質的に向上した塩基編集特異性をもたらし、オンターゲット編集効率の控えめな縮減のみを有することを確立している。

【０５６５】

BE3のRNP送達はDNA不使用の塩基編集を可能にする

次に、培養ヒト細胞に直接的に送達されたときの塩基編集を媒介するDNA不含のRNP形態のBE3の能力を研究した。カチオン性脂質試薬はリボヌクレオ蛋白質（RNP）複合体を包含する負荷電蛋白質または蛋白質:核酸複合体を哺乳類細胞に強力に送達し得ることと、RNP送達はオフターゲットゲノム編集を実質的に縮減し得ることとが最近確立された[27～29]。

【０５６６】

商業的に入手可能なカチオン性脂質Lipofectamine 2000を、EMX1、HEK293部位3、FANCF、またはVEGFA部位2座位を標的化するガイドRNAと予め複合体化させられた精製されたBE3蛋白質またはHF-BE3蛋白質どちらかと組み合わせ、もたらされた脂質:RNP複合体をHEK293T細胞とインキュベーションした。72h後に、ゲノムDNAを収穫し、オンターゲットおよびオフターゲット塩基編集を高スループットDNAシーケンシングによって分析した。全てのCas9に基づくテクノロジーと同じく、異なるゲノム座位における編集効率には実質的な変動が観察された（図127および130）。異なる処置間のオンターゲット編集効率に関連する傾向を表示するために、4つの試験された座位において平均のオンターゲット塩基編集効率を算出した（図128A）。BE3の蛋白質送達（200nM）はプラスミドトランスフェクションで観察されたものと同等のオンターゲット編集効率に至る（それぞれ26±4%vs.29±5%；n=3つの生物学的レプリケートの平均±s.e.m.；図128A）。

【０５６７】

対照的に、HF-BE3の蛋白質送達は、研究された4つのゲノム座位においてBE3の蛋白質送達と比較してオンターゲット編集を縮減した（それぞれ13±3%vs.26±4%の平均編集効率；n=3つの生物学的レプリケートの平均±s.e.m.；図128A）。HF-BE3およびBE3は試験管内では同等の編集効率を有し（図125B）、HF-BE3がHEK293T細胞においてプラスミドから発現されるときには編集は軽度にのみ縮減されるので（図127A～D）、HF-BE3蛋白質送達からの編集の減少した効率は哺乳類細胞における減少したHF-BE3安定性の結果であり得るということを推測することは魅力的である。より低い安定性はプラスミドからの連続的な発現によって埋め合わせられ得るが、1回の蛋白質送達後では埋め合わせられ得ない。この観察は、$CD34^+$造血幹および前駆細胞にヌクレオフェクションされたときに、精製されたCas9と比較して精製されたHF-Cas9による縮減されたオンターゲットインデル形成の最近の報告と整合している[30]。この研究が査読中である間に、Kim et al.はエレクトロポレーションを用いてマウス胚へのBE3のRNP送達を実証した[31]。本発明者の知識の及ぶ限り、本アプローチは、限定されたin vivoの治療上の妥当性を有するエレクトロポレーションなしに哺乳類細胞の個々のヌクレオチドに正確な変化を創出することができる始めのDNA不使用の技術である。

【０５６８】

塩基編集因子のRNP送達はDNA特異性を多大に向上させる

重要なことに、BE3およびHF-BE3のRNP送達は実質的なオンターゲット塩基編集に至ったが、EMX1、FANCF、およびHEK293部位3に関連する9つの試験されたオフターゲット座位のいずれかにおいて、測定可能な塩基編集（＜0.025%）の場合は観察されなかった（図130A～130C）。対照的に、BE3のプラスミド送達は、塩基編集活性ウィンドウ内の全てのシーケンシングされたシトシンにおける平均で1.1±0.3%（n=3つの生物学的

レプリケートの平均±s.d.）のオフ編集と、これらの9つのオフターゲット座位に配置された16個のオフターゲットシトシンの11個における検出可能なオフターゲット編集とに至る（図127A～127D）。試験された3つの非反復的な座位のオフターゲット座位において、BE3蛋白質送達はプラスミド送達のものよりも26倍高い平均の特異性比に至る（図127A）。これらの結果は、塩基編集因子のRNP送達が塩基編集のDNA特異性を劇的に増大させるということを明らかにしている。

【０５６９】

BE3またはHF-BE3どちらかのプラスミド送達と比較して、BE3またはHF-BE3どちらかの蛋白質送達は、高度にプロミスキャスなVEGFA部位2座位における多大に改善された塩基編集特異性をもまたもたらした（図127および130を比較せよ；表16を見よ）。この部位に関連するオフターゲット座位における塩基編集の絶対頻度は、蛋白質送達によって、BE3（プラスミド送達：15±4%オフターゲット編集；蛋白質送達：1.3±0.4%オフターゲット編集；このパラグラフの全ての値はn=3つの生物学的レプリケートの平均±s.d.に相当する）およびHF-BE3（プラスミド送達：5±2%オフターゲット編集；蛋白質送達：0.5±0.1%オフターゲット編集）両方について少なくとも10倍縮減された。全ての4つの研究された座位において、オンターゲット：オフターゲット編集の塩基編集特異性比はBE3のプラスミド送達と比較してBE3の蛋白質送達では平均で66倍増大した（図129）。総合すると、これらの結果は、反復的および非反復的標的部位両方について、RNP対DNA送達が、高フィデリティCas9変異の存在または非存在よりも塩基編集特異性の強い決定要因であることを明らかにしている。

【０５７０】

HF変異の導入も送達方法もどちらも、塩基編集に関連する低いインデル率を実質的には変改しなかった。この研究の全ての処置条件における全てのオンターゲット座位のインデル頻度は低いままに留まり（典型的には≦5%；図131A）、編集：インデル比は、試験された全てのケースにおいて、最適化されたHDRプロトコールを用いる先の研究よりも高いままに留まった（典型的には≧10倍；図131B）[30,32,33]。非反復的なsgRNAsでは、非常に少数のインデルがオフターゲット座位において観察されたが（図131C）、BE3のプラスミド送達はVEGFA部位2に関連するオフターゲット座位において5%までのインデルを創出したことが注意される（図131C）。

【０５７１】

一緒にすると、これらの結果は、塩基編集因子の蛋白質送達がオンターゲット塩基編集効率を維持し、プラスミドDNAの送達に対して相対的に編集特異性を多大に向上させることを確立している。

【０５７２】

RNP送達はオンおよびオフターゲット編集を切り離す

BE3の蛋白質送達に関連する塩基編集特異性の際立った向上に鑑みると、この改善が細胞に送達される活性なゲノム編集薬剤の合計の数量の縮減の結果であるかどうかを検討した。EMX1を標的化するsgRNAを用いて、プラスミド（図128B）および蛋白質送達（図128C）の用量反応研究を行った。処置条件間のトランスフェクション効率を最大化するために、Lipofectamine 2000の体積は全ての試験で1.5μLであり、塩基編集因子蛋白質:sgRNAモル比は蛋白質送達では1:1.1に維持した。プラスミド送達では、1:3というsgRNAプラスミド：BE3プラスミドの質量比（モル比～1:1）および1.5μLのLipofectamine 2000を用いた。プラスミド送達の試験された全ての条件においてオフターゲット塩基編集が観察されたが（図128B）、試験された全ての蛋白質送達条件においてオフターゲット編集は事実上観察されなかった（図128C）。

【０５７３】

線形回帰分析を行なって、プラスミドおよび蛋白質送達についてオンおよびオフターゲット編集間の関係性を査定した。プラスミド送達では、オフターゲット編集はオンターゲット編集率と密に関連していたが（R^2=0.95，非ゼロスロープのp=0.0012，F検定）、蛋白質送達を用いるオフターゲットおよびオンターゲット編集間には有意な関連はなかっ

た（R^2=0.078，非ゼロスロープのp=0.59，F検定）。

【０５７４】

これらのデータは、塩基編集因子の蛋白質送達が、用量とは独立している固有の特異性の利点をオファーすることを指示している。先の観察と一緒になって[29,34]、これらの知見は、DNA送達と比較して蛋白質送達からの塩基編集のより高いDNA特異性が、塩基編集因子に対するゲノムの延長された暴露を回避し、それによって、オンターゲット座位がすでに改変された後に、塩基編集因子がオフターゲット座位をプロセシングする機会を最小化する蛋白質送達の能力から生ずるモデルを支持している。

【０５７５】

ゼブラフィッシュおよびマウスにおけるDNA不使用の塩基編集

上の観察は、検出可能なオフターゲット塩基編集を消去しながらオンターゲット塩基編集を維持するためのBE3の蛋白質送達の有望さを示唆している。よって、BE3の蛋白質送達がゼブラフィッシュにおいて特異的な点変異を創出するために用いられ得るかどうかを、チロシナーゼ座位を標的化するBE3:sgRNA複合体をゼブラフィッシュ受精胚にインジェクションすることによって試験した。インジェクションの4日後にもたらされたゼブラフィッシュ幼生からゲノムDNAを収穫し、高スループットシーケンシングによって塩基編集およびインデル頻度を測定した（図132A）。試験された3つのBE3:sgRNA複合体の2つはin vivoで実質的な点変異を誘導した（TYR1：C_3→T_3 5.3±1.8%、TYR2：C_4→T_4 4.3±2.1%；n=3つのインジェクションされた胚の平均±s.d.；図132A）。ゼブラフィッシュ座位の配列は表17にリスト化されている。

【表２３】

HEK

サンプルの記載	1	2	3	アンプリコン
Protein, BE3	2372566	2956091	1593911	VEGFA On Target
Protein, HF-BE3	3834801	3898741	3834671	VEGFA On Target
Plasmid, BE3	3152131	2808911	3356681	VEGFA On Target
Plasmid, HF-BE3	1963231	2519651	3692011	VEGFA On Target
Control	390748	395523	353614	VEGFA On Target
Protein, BE3	192801	264721	247991	FANCF Off Target Site #1
Protein, HF-BE3	363831	300071	391931	FANCF Off Target Site #1
Plasmid, BE3	355801	295571	222431	FANCF Off Target Site #1
Plasmid, HF-BE3	403711	271871	282481	FANCF Off Target Site #1
Control	351061	372741	349391	FANCF Off Target Site #1
Protein, BE3	829781	1246891	838401	VEGFA Off Target Site #1
Protein, HF-BE3	1424041	1404821	1422201	VEGFA Off Target Site #1
Plasmid, BE3	1120271	1170711	1008941	VEGFA Off Target Site #1
Plasmid, HF-BE3	1141871	988761	1225531	VEGFA Off Target Site #1
Control	768541	905471	892711	VEGFA Off Target Site #1
Protein, BE3	145141	255151	193251	VEGFA Off Target Site #2
Protein, HF-BE3	246781	243631	253121	VEGFA Off Target Site #2
Plasmid, BE3	169431	199181	102251	VEGFA Off Target Site #2
Plasmid, HF-BE3	122001	147691	177971	VEGFA Off Target Site #2
Control	87391	136481	78181	VEGFA Off Target Site #2
Protein, BE3	1792421	1116931	1739091	VEGFA Off Target Site #3
Protein, HF-BE3	2438991	3005031	2761191	VEGFA Off Target Site #3
Plasmid, BE3	2084761	2913701	1554401	VEGFA Off Target Site #3
Plasmid, HF-BE3	1171741	1540331	1991521	VEGFA Off Target Site #3
Control	1192631	1704361	1216861	VEGFA Off Target Site #3
Protein, BE3	2377991	2629471	1853711	VEGFA Off Target Site #4
Protein, HF-BE3	3132531	2336991	2449221	VEGFA Off Target Site #4
Plasmid, BE3	2430941	2303161	2344211	VEGFA Off Target Site #4
Plasmid, HF-BE3	1709581	1600911	1406931	VEGFA Off Target Site #4
Control	1586911	1487201	1372701	VEGFA Off Target Site #4
Protein, BE3	286841	282371	433151	HEK3 On Target
Protein, HF-BE3	493001	425761	576901	HEK3 On Target
Plasmid, BE3	550081	558131	543101	HEK3 On Target
Plasmid, HF-BE3	551991	113841	76591	HEK3 On Target
Control	637411	428781	485241	HEK3 On Target
Protein, BE3	1048271	1817921	1610901	HEK3 Off Target Site # 1
Protein, HF-BE3	2045801	1755611	1773031	HEK3 Off Target Site # 1
Plasmid, BE3	1785841	1522641	2068631	HEK3 Off Target Site # 1
Plasmid, HF-BE3	1912971	1384251	1607891	HEK3 Off Target Site # 1
Control	1903031	1900611	1905161	HEK3 Off Target Site # 1
Protein, BE3	1460891	951131	1350151	HEK3 Off Target Site # 2
Protein, HF-BE3	1559471	1365411	1579911	HEK3 Off Target Site # 2
Plasmid, BE3	1500361	1284381	1589051	HEK3 Off Target Site # 2
Plasmid, HF-BE3	3710771	1236421	1425621	HEK3 Off Target Site # 2
Control	1303221	1345451	1418331	HEK3 Off Target Site # 2
Protein, BE3	1450581	1753381	1618371	HEK3 Off Target Site # 3
Protein, HF-BE3	2123371	1789931	1798871	HEK3 Off Target Site # 3
Plasmid, BE3	1864521	1665001	804411	HEK3 Off Target Site # 3
Plasmid, HF-BE3	1637321	1184531	1347191	HEK3 Off Target Site # 3
Control	1314611	1344701	1556081	HEK3 Off Target Site # 3
Protein, BE3	419861	616781	678901	FANCF On Target
Protein, HF-BE3	410571	558501	864111	FANCF On Target
Plasmid, BE3	391141	485751	700741	FANCF On Target
Plasmid, HF-BE3	416171	556381	757181	FANCF On Target
Control	688521	594221	812651	FANCF On Target
Protein, BE3	1134621	805291	1913441	FANCF Off Target Site # 1
Protein, HF-BE3	2026621	2339811	2030241	FANCF Off Target Site # 1
Plasmid, BE3	2089121	2020441	1072341	FANCF Off Target Site # 1
Plasmid, HF-BE3	864941	1139891	868071	FANCF Off Target Site # 1
Control	922551	723861	566611	FANCF Off Target Site # 1
Protein, BE3	962711	1174421	843741	FANCF Off Target Site # 2
Protein, HF-BE3	1056241	1023121	1053431	FANCF Off Target Site # 2
Plasmid, BE3	1010021	987471	700521	FANCF Off Target Site # 2
Plasmid, HF-BE3	3089661	697871	831841	FANCF Off Target Site # 2
Control	999861	1003441	1006591	FANCF Off Target Site # 2
Protein, BE3	255241	1824511	653881	FANCF Off Target Site # 3
Protein, HF-BE3	718581	755531	717851	FANCF Off Target Site # 3
Plasmid, BE3	609801	573601	781691	FANCF Off Target Site # 3
Plasmid, HF-BE3	683161	346591	857181	FANCF Off Target Site # 3
Control	496851	573881	604181	FANCF Off Target Site # 3
Protein, BE3	469811	907931	794391	EMX1 On Target
Protein, HF-BE3	586291	711861	615751	EMX1 On Target
Plasmid, BE3	708171	827361	757061	EMX1 On Target
Plasmid, HF-BE3	770381	711231	785111	EMX1 On Target
Control	621831	485741	684391	EMX1 On Target
Protein, BE3	1659051	2575631	1428881	EMX 1 Off Target Site # 1
Protein, HF-BE3	1483391	1513001	1307121	EMX 1 Off Target Site # 1
Plasmid, BE3	1019501	1032261	2030041	EMX 1 Off Target Site # 1
Plasmid, HF-BE3	1679691	1751931	970101	EMX 1 Off Target Site # 1
Control	1014761	1504351	1023271	EMX 1 Off Target Site # 1
Protein, BE3	1367381	2134381	1187111	EMX 1 Off Target Site # 2
Protein, HF-BE3	1234131	1261141	1093751	EMX 1 Off Target Site # 2
Plasmid, BE3	855761	866001	1695921	EMX 1 Off Target Site # 2
Plasmid, HF-BE3	1403171	1457381	1370501	EMX 1 Off Target Site # 2
Control	848181	1251391	854541	EMX 1 Off Target Site # 2
Protein, BE3	119401	365931	249461	EMX1 Off Target Site # 3
Protein, HF-BE3	267621	315661	363771	EMX1 Off Target Site # 3
Plasmid, BE3	324201	215471	146591	EMX1 Off Target Site # 3
Plasmid, HF-BE3	314771	166921	173851	EMX1 Off Target Site # 3
Control	173851	281281	327171	EMX1 Off Target Site # 3

NIH 3T3細胞処置からのマウスサンプル

サンプルの記載	レプリケート 1	2	3	アンプリコン
Plasmid, BE3	166411	1022161	4636131	On Target VEGFA (Mus)
Plasmid, HF-BE3	893301	1265451	1009931	On Target VEGFA (Mus)
Protein, BE3	889981	816971	5112431	On Target VEGFA (Mus)
Protein, HF-BE3	1282181	291931	1315151	On Target VEGFA (Mus)
Control	187671	388661	589851	On Target VEGFA (Mus)
Plasmid, BE3	1747821	1675041	1825651	CFD Off Target 1
Plasmid, HF-BE3	1671201	1825201	1923891	CFD Off Target 1
Protein, BE3	2305691	2126051	1381441	CFD Off Target 1
Protein, HF-BE3	2286681	2114571	1833701	CFD Off Target 1
Control	1717381	1911171	2087931	CFD Off Target 1
Plasmid, BE3	2064751	2273321	2060891	CFD Off Target 2
Plasmid, HF-BE3	2138091	2030281	1990781	CFD Off Target 2
Protein, BE3	2159951	2757541	2499691	CFD Off Target 2
Protein, HF-BE3	2509181	2720631	2410591	CFD Off Target 2
Control	1937601	1759591	2469631	CFD Off Target 2
Plasmid, BE3	603881	1262781	732811	CFD Off Target 3
Plasmid, HF-BE3	890451	1285081	517881	CFD Off Target 3
Protein, BE3	1671951	3300461	1116331	CFD Off Target 3
Protein, HF-BE3	821201	3093521	1039331	CFD Off Target 3
Control	832041	1769391	566131	CFD Off Target 3
Plasmid, BE3	1928461	1137091	1710781	CFD Off Target 4
Plasmid, HF-BE3	2056011	1514341	1889431	CFD Off Target 4
Protein, BE3	2181941	1819931	2063981	CFD Off Target 4
Protein, HF-BE3	2119661	1489761	1868381	CFD Off Target 4
Control	1839331	1303181	1974761	CFD Off Target 4

マウス蝸牛サンプル

サンプルの記載	レプリケート 1	2	3	アンプリコン
Stria vascularies	378891	2057061	6209151	On Target (VEGFA)
Organ of Corti	1484471	1750041	2907531	On Target (VEGFA)
Modiolus	1828061	1813821	6126931	On Target (VEGFA)
Uninjected control	2282221	2419791	2727591	On Target (VEGFA)
Stria vascularies	444571	2444871	2446461	CFD Off Target 1
Organ of Corti	1363351	1183181	347471	CFD Off Target 1
Modiolus	671761	2095431	686991	CFD Off Target 1
Uninjected control	3431001	3427171	3790151	CFD Off Target 1
Stria vascularies	72962	3198831	2657931	CFD Off Target 2
Organ of Corti	1584561	1314301	605301	CFD Off Target 2
Modiolus	920141	2515091	814131	CFD Off Target 2
Uninjected control	3991381	3459651	4839201	CFD Off Target 2
Stria vascularies	83251	803221	1425561	CFD Off Target 3
Organ of Corti	810141	459761	181051	CFD Off Target 3
Modiolus	99281	755551	1134151	CFD Off Target 3
Uninjected control	3991381	3459651	4839201	CFD Off Target 3
Stria vascularies	2321941	3977701	5540541	CFD Off Target 4
Organ of Corti	3134721	2853021	1768721	CFD Off Target 4
Modiolus	2301051	3713991	2581421	CFD Off Target 4
Uninjected control	5245031	6379461	6247091	CFD Off Target 4

ゼブラフィッシュサンプル

	1	2	3	アンプリコン
Treated zebrafish	723551	494981	8106131	TYR 1
Scrambled sgRNA	1079191	985021	924291	TYR 1
Treated zebrafish	514341	480141	415471	TYR 2
Scrambled sgRNA	614661	623741	667651	TYR 2
Treated zebrafish	64871	572471	7588331	TYR 3
Scrambled sgRNA	645961	712341	756241	TYR 3

マウス蝸牛サンプル　非関連sgRNAによって処置

サンプル	Stria	Cortil	Modiolus	アンプリコン
	537459	249767	389274	On Target (VEGFA)

表17：参照配列とアライメントし、材料と方法に記載されているクオリティフィルターをパスするHTSリードの数。

【０５７６】

最後に、マウスにおいてDNA不使用の高特異性の塩基編集を達成するために、これらの開発を適用した。in vivoのオンおよびオフターゲット塩基編集を観察する蓋然性を最大化するために、VEGFA部位2を標的化する高度に反復的なsgRNAを用いた；便利なことに、マウスおよびヒトゲノムはこの標的部位において同一である。

【０５７７】

培養マウスNIH/3T3細胞を用いて、BE3蛋白質送達がこの座位における効率的なオンタ

ーゲット塩基編集34±11%を生むことが確認された（図133A；このパラグラフの全ての編集パーセンテージはn=3つの生物学的レプリケートの平均±s.d.に相当する）。Cutting Frequency Determinant（CFD）アルゴリズム[29,34]を用いて、VEGFA部位2 sgRNAに関連するマウスゲノム中のオフターゲット座位を予測した（表18）。培養NIH/3T3細胞を用いて、上位4つの予測されるオフターゲット座位の2つが培養マウス細胞においてBE3のプラスミド送達によって本当に改変されることが確認された（CFDオフターゲット座位1，9±5%編集；およびCFDオフターゲット座位4，3±2%編集；図133B～133E）。ヒト細胞からの結果と整合して、BE3の蛋白質送達は陰性対照のものに類似のレベルまでオフターゲット編集を縮減した（図133Cおよび133E）。CFDオフターゲット座位1および4の平均の塩基編集特異性比はBE3のプラスミド送達の28±13からBE3の蛋白質送達の≧780±300まで増大した（値は平均±s.e.m.に相当する；n=3つの生物学的レプリケート）。

【表２４】

部位	配列	配列番号	CFD スコア	座位の記載
オンターゲット	GACCCCCTCCACCCCGCCTCCGG	497		VEGFA 部位 2
オフターゲット 1	TCCCCCCTCCACCCCACCTCCGG	498	0.7857	遺伝子間 :mmu-mir-21c-Nrp1/Mir1903
オフターゲット 2	TGCCCACCTCACCCCGCCTCTGG	499	0.65	イントロン:Vipr1
オフターゲット 3	GCCCCTCCCAACCCCACCTCTGG	500	0.6323	イントロンNos1ap
オフターゲット 4	CACCCCCCTCACCCCGCCTCAGG	501	0.625	遺伝子間 :Unc5b-mmu-mir-6408

表18：VEGFA部位2 sgRNAに関連するマウスゲノム中の予測されるオフターゲット座位のプロトスペーサーおよびPAM配列。CFDスコア[46]はCRISPOR[47]を用いて算出した。オンターゲット配列とは異なるオフターゲットプロトスペーサー内の位置は下線付きである。

【０５７８】

マウスのDNA不使用の塩基編集を確立するために、BE3:sgRNA複合体をLipofectamine 2000と組み合わせ（図132B）、P1-P2の仔マウスへの蝸牛内注射を行った。注射された蝸牛組織を注射の3～4日後に収穫し、蝸牛領域あたり5～7サンプルにマイクロダイセクションした。注射されていないマウスからの対照蝸牛を同時に収穫した。ゲノムDNAを収穫された組織から抽出し、qPCRによって指数期後期まで増幅し、高スループットDNAシーケンシングに付してC→T変換を測定した。塩基編集因子蛋白質に暴露された細胞から排他的にDNAを回収することは可能ではないので、処置された細胞の塩基編集効率を定量化することは不可能であるが、蝸牛の3つの領域の組織から、一義的な塩基編集を観察した：コルチ器の基底末端、血管条、およびモダイオラス（図132C～132D）。処置された組織サンプルにおいて、有意なインデル形成は検出されなかった（＜0.1%のインデル；図134B）。

【０５７９】

標的C→T変換を含有する蝸牛細胞のパーセンテージ（図132C）は、培養の処置されたNIH/3T3細胞で観察されたもの（図133A）よりも有意に低く、脂質に基づく蛋白質送達の高度に局在した性質、および塩基編集因子に暴露された細胞から排他的にDNAを単離する不能性と整合した。それでもやはり、局所送達は、いずれの細胞型が編集されるかのコントロールならびに調製および投与の容易さを包含する、アクセス可能な用途にとって鍵となる利点をオファーする。

【０５８０】

最後に、BE3:sgRNA:脂質複合体の蝸牛内注射後のオフターゲット編集を分析した。確認されたオフターゲット部位CFD座位1およびCFD座位4を包含する、BE3:VEGFA部位2 sgRNA:脂質複合体を注射されたマウスの蝸牛組織からのゲノムDNA中の全ての4つの予測されるオフターゲット座位の分析は、試験されたオフターゲット座位のいずれかについて

、非処置の対照サンプルで観察されたものよりも上の検出可能なC→T変換またはインデル形成を明らかにしなかった（図134A）。

【０５８１】

一緒になって、これらのin vivo塩基編集結果は、高いDNA配列特異性を有する動物のゲノムDNA中の個々のヌクレオチドの正確な変換のための、ウイルス不使用のDNA不使用の戦略を確立している。

【０５８２】

考察

この研究において開発および実装された戦略は、オフターゲット塩基編集を取り除くかまたは縮減することおよびin vivo塩基編集を支持するDNA不使用の送達方法を確立することによって、塩基編集の実用性および適用性を拡大する。蛋白質送達は、同じ構築物のプラスミド送達と比較してヒトおよびマウス細胞における塩基編集特異性を改善し（図137、130、および133）、ゼブラフィッシュおよびマウス蝸牛における特異的な塩基編集を可能にする（図132）。

【０５８３】

Cas9のDNA特異性を向上させることが公知のBE3変異をインストールすることによって、高フィデリティ塩基編集因子を創出した[20]。Cas9へのこれらの変異のインストールは、ここで取り調べられたEMX1座位を包含する非反復的なsgRNAに関連するオフターゲット座位において、検出不可能なインデル形成をもたらすことが報告されている（図127A）[20]。HF-BE3で観察された特異性向上は、実質的だが、より控えめであった；HF-BE3は、哺乳類細胞にプラスミドDNAとして送達されたときには、反復的および非反復的な座位両方において検出可能なオフターゲット塩基編集を見せた（図127A、127D、133C、および133E）。塩基編集は、Cas9によって媒介されるインデル形成とは違ってDNA切断を要求せず、DNA結合およびRループ形成のみを要し[14]、HF-Cas9の向上した特異性のいくつかは、すでに結合されたオフターゲット座位における棄損されたDNA切断から生じ得る、という事実から、この特異性向上の違いが生じ得ることを推測することは魅力的である。

【０５８４】

オフターゲット塩基編集を縮減するための第2の試みにおいては、塩基編集因子のRNP送達がオンおよびオフターゲット編集の切り離しに至ることが実証された（図128B～128C）。RNP送達は、プラスミド送達と同等のオンターゲット編集を維持しながら非反復的部位におけるオフターゲット編集を静め（図130A～130Cおよび128A）、高度に反復的なVEGFA部位2においてさえもオフターゲット編集を多大に縮減した（図130D）。塩基編集因子のRNP送達は、細胞の用量が典型的にはコントロールまたは特徴付けることが難しいin vivoの編集用途にとって特に有用であり得る。

【０５８５】

ドナーDNA鋳型の送達とカップリングされたCas9のRNP送達は、哺乳類細胞においてHDRに基づくゲノム編集を行うために先に用いられている。しかしながら、これらのアプローチは、低い効率、細胞状態依存性、および典型的には特に点変異修正について所望のHDR成果のものを超えるインデル形成効率によって限定されたままに留まっている[29,30,32,35]。対照的に、DNA不使用の塩基編集は、in vivoおよび細胞両方において確率的なインデルに対して相対的に実質的過剰の編集された生成物を創出する（図132A、134A、および134B）。本発明者の知識の及ぶ限り、塩基編集因子のRNP送達は、外来性DNAを要求することなしにゲノムDNAに特異的かつ正確な改変を創出するための第1の戦略に相当する。

【０５８６】

方法

プラスミドのクローニング

この研究のプラスミドはUSERクローニングによって創出した。Phusion Uホットスタートポリメラーゼ（Thermo Fisher）を用いて点変異をインストールし、先に報告された構築物から蛋白質発現プラスミドを構築した[36]。蛋白質配列は補足情報にリスト化されて

おり、BE3およびHF-BE3の発現のためのプラスミドはAddgeneから入手可能である。

【０５８７】

BE3およびHF-BE3の発現および精製

His6N末端精製タグを有する細菌コドン最適化された塩基編集因子をコードするプラスミドによって、BL21 Star（DE3）E. coliコンピテントセルを形質転換した。シングルコロニーを50μg mL^{-1}カナマイシンを含有するルリア・ベルターニ（LB）ブロスによって37℃で一晩成長させた。細胞を2Lの同じ培地によって1:200希釈し、OD_{600}=0.70～0.75まで37℃で成長させた。培養物を氷上で60minに渡ってインキュベーションし、蛋白質発現を0.5mMイソプロピル-β-D-1-チオガラクトピラノシド（IPTG，GoldBio）によって誘導した。発現は18℃で振盪しながら14～16hに渡って持続した。爾後の精製ステップは4℃で実施した。細胞を20minの6,000gにおける遠心によって収集し、細胞収集緩衝液（100mMトリス（ヒドロキシメチル）-アミノメタン（Tris）-HCl，pH8.0，1M NaCl，20%グリセロール，5mMトリス（2-カルボキシエチル）ホスフィン（TCEP；GoldBio），0.4mMフッ化フェニルメタンスルホニル（PMSF；Sigma Aldrich），および用いられた50mL緩衝液あたり1個のcOmplete EDTA不含プロテアーゼ阻害剤ペレット（Roche））中に再懸濁した。細胞をソニケーションによって溶解し（合計6min，3sオン，3sオフ）、溶解物を25,000gにおける遠心によって清澄化した（20min）。

【０５８８】

清澄化された溶解物をHis-Purニッケルニトリロ酢酸（ニッケル-NTA）樹脂（培養のリットルあたり1mL樹脂，Thermo Fisher）と4℃で回転しながら60～90minに渡ってインキュベーションした。結合した蛋白質を溶出緩衝液（（100mMトリス（ヒドロキシメチル）-アミノメタン（Tris）-HCl，pH8.0，0.5M NaCl，20%グリセロール，5mMトリス（2-カルボキシエチル）ホスフィン（TCEP；GoldBio），200mMイミダゾール）によって溶出する前に、樹脂を20カラム体積の細胞収集緩衝液によって洗浄した。もたらされた蛋白質画分は、5mL Hi-Trap HP SP（GE Healthcare）カチオン交換カラムによって、Akta Pure FPLCを用いてさらに精製した。蛋白質含有画分を3,000gで遠心される100,000kDaカットオフ（Millipore）を有するカラムを用いて濃縮し、濃縮された溶液は.22μm PVDF膜（Millipore）を通して無菌濾過した。

【０５８９】

無菌濾過後に、蛋白質を還元剤適合性のビシンコニン酸（BCA）アッセイ（Pierce Biotechnology）によって定量し、液体窒素中で瞬間凍結し、-80℃においてアリコートで保存した。発現された蛋白質の配列は補足的な注2にリスト化されている。

【０５９０】

sgRNAのin vitro転写

所望の20bp sgRNAプロトスペーサーおよびsgRNAバックボーンの上流にT7 RNAポリメラーゼプロモーター配列を含有する線状DNA断片を、補足情報にリスト化されているプライマーを用いてPCR（Q5ホットスタートマスターミックス，New England Biolabs）によって創出し、mineluteカラム（Qiagen）によって濃縮した。sgRNAを、20μL反応あたり1μgの線状鋳型で14～16hに渡って16℃においてHiScribe T7高収率RNA合成キット（New England Biolabs）によって転写した。sgRNAは製造者の説明書に従ってMEGAClear転写クリーンアップキット（Thermo Fisher）を用いて精製した。精製されたsgRNAは-80℃においてアリコートで保存した。

【０５９１】

in vitro脱アミノ化アッセイ

in vitro脱アミノ化アッセイの鋳型として用いられたDNAオリゴヌクレオチドの配列は補足的な注3に示されている。全てのオリゴヌクレオチドはIDTから購入した。相補的な配列を有する合成された一本鎖オリゴヌクレオチド同士をTris緩衝液pH8.0中で組み合わせ（5μLの100μM溶液）、5minに渡る95℃への加熱、次に0.1℃$秒^{-1}$の速度における37℃への徐々の冷却によってアニーリングして、79塩基対（bp）dsDNA基質を創出した。フレッシュに融かされた塩基編集因子蛋白質（10μL反応体積中に2μM終濃度）を、指示

されているsgRNA（2.2μM終濃度）と反応緩衝液（20mM HEPES pH7.5，150mM KCl，0.5mM DTT，0.1mM EDTA，10mM $MgCl_2$）[37]中において5分に渡って室温で複合体化した。それから、アニーリングされたdsDNA基質を250nMの終濃度まで追加した。蛋白質変性が99℃における5minの加熱によって行われる前に、反応は30minに渡って37℃で進んだ。PB緩衝液（Qiagen，100μL）およびイソプロパノール（25μL）の追加は、蛋白質が基質DNAから解離させられることを保証した。DNAをMineluteカラム（Qiagen）によって精製し、もたらされた生成物を、Uに対して寛容なポリメラーゼ（Phusion Uホットスタート，Thermo Fisher）および補足情報にリスト化されているプライマーを用いて、15サイクルのqPCR（12ngインプットDNA，50μL反応体積）によって直線範囲の一番上まで増幅した。下に記載されている通りIllumina MiSeqによるシーケンシングのために調製される前に、増幅されたDNAをRapidTip2（Diffinity Genomics）を用いて精製し、第2ラウンドのPCR（8サイクル，5ngインプット）によってバーコード化した。

【０５９２】

ゲノムDNAの精製およびシーケンシング

ゲノムDNAを製造者の説明書に従ってAgencourt DNAdvanceゲノムDNA単離キット（Beckman Coulter）を用いて単離した。第1のPCRでは、製造者の説明書に従ってだが3%DMSOおよびSYBR Gold核酸染色（Thermo Fisher）の追加によってQ5ホットスタートDNAポリメラーゼ（NEB）を用いて、DNAを直線範囲の一番上まで増幅した。全てのアンプリコンについて、用いられたPCRプロトコールは、98℃における2minの最初の加熱ステップ、次に、最適化された数の増幅サイクル（98℃で12s，61℃で25s，72℃で30s）であった。ゼブラフィッシュおよびトランスフェクションされた細胞サンプルでは30ngのインプットDNAを50μL反応に用い、蝸牛サンプルでは20ngを25μL反応に用いた。qPCRを行って、各アンプリコンの最適なサイクル数を決定した。増幅されたDNAをRapid Tip2（Diffinity Genomics）を用いて精製し、さらなるPCR（8サイクル，5ngインプット）によってバーコード化した。第1ラウンドPCRに用いられたユニークなフォワードおよびリバースプライマーはアニーリング領域の5'に定常領域を含有し（フォワード：5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNN-3'（配列番号502）、リバース：5'-TGGAGTTCAGACGTGTGCTCTTCCGATCT-3'（配列番号503））、これは第2ラウンドPCRのための増幅されたDNAへのバーコード化プライマーの結合を容易化した。

【０５９３】

第2ラウンドPCRは3つの領域を有するプライマーを用いた：アンプリコンがIlluminaフローセルに結合することを許す5'定常領域（斜体化）、8塩基バーコード化領域（X）、およびバーコード化プライマーが第1ラウンドPCRアンプリコンに結合することを許す3'定常領域（太字）。プライマー配列の例は：

フォワード：

【化１２３】

5’-

*AATGATACGGCGACCACCGAGATCTACAC*XXXXXXXX**ACACTCTTTCCCTACACGAC**

-3’

（配列番号504）

リバース：

【化１２４】

5’-

CAAGCAGAAGACGGCATACGAGATXXXXXXX**GTGACTGGAGTTCAGACGTGTGC**

TCTTC-3’

（配列番号505）
であった。
【０５９４】
シーケンシングアダプターおよび二重バーコード化配列はTruSeqインデックスアダプター（Illumina）に基づく。バーコード化されたサンプルをプールし、ゲル抽出（Qiagen）によって精製し、それから、製造者の説明書に従ってQubit dsDNA HSキット（Thermo Fisher）およびqPCR（KAPA Biosystems）を用いる定量の前に、Ampureビーズ（Beckman Coulter）を用いて精製した。製造者の説明書に従ってMiSeq（Illumina）によって、180～250塩基（アンプリコンサイズに依存する）からのシングルエンドリードを用いて、プールされたサンプルのシーケンシングを行なった。
【０５９５】
PCR増幅に用いられたオリゴヌクレオチドの配列は補足的な注3に示されている。全てのオリゴヌクレオチドはIDTから得た。この研究の各アンプリコンについて最適化された数のPCRサイクルは次の通りである：VEGFA部位2ヒトゲノムDNA（アニーリング温度は全ての伸長ステップについて25秒に渡って61℃であった）：オンターゲット：29サイクル、オフターゲット#1：32サイクル、オフターゲット#2：28サイクル、オフターゲット#3：27サイクル、オフターゲット#4：27サイクル、VEGFA部位2マウスゲノムDNA：オンターゲット：31サイクル、オフターゲット#1、#2、#3、および#4：31サイクル。HEK293部位3：オフターゲット#1：29サイクル、オフターゲット#2：28サイクル、オフターゲット#3：28サイクル。FANCFオフターゲット#1：29サイクル、オフターゲット#2：28サイクル、オフターゲット3：28サイクル。EMX1オフターゲット#1、#2、および#3：28サイクル。ゼブラフィッシュDNAの増幅のためのTYR1、TYR2、およびTYR3 sgRNA：32サイクル。先に記載されている通り、EMX1、FANCF、およびHEK293部位3座位のオンターゲット増幅のための最適化されたプロトコールを踏襲した[14]。
【０５９６】
ゲノムDNAシーケンシングリードの分析およびアライメント
シーケンシングリードは先に記載されている通り分析した[14]。簡潔には、シーケンシングリードを、MiSeqレポーター（Illumina）を用いてデマルチプレックスし、先に報告されているカスタムのMatlabスクリプトによって個々のFASTQファイルを分析した[14]。Smith-Watermanアルゴリズムを用いて、リードを参照配列に対してアライメントした。30よりも下のQスコアを有するベースコールはプレースホルダーヌクレオチド（N）によって置き換えた。このクオリティ閾値は1,000分の1という予想エラー率を有するヌクレオチド頻度をもたらす。インデル頻度を先に公開されているカスタムのMatlabスクリプトによって定量した。これは、nCas9切断部位の周りの30塩基ウィンドウにおいて生起し、最小で2塩基挿入または欠失であるインデルをカウントする[14]。処置されたサンプルおよび非処置の対照におけるインデル形成間に有意な違いがある場合には（スチューデントの両側t検定，$p<0.05$）、インデルは検出可能であると定められた。
【０５９７】
シーケンシングされたアンプリコンの1つ、マウスゲノムのVEGFA部位2 sgRNAに関連するCFDオフターゲット#3については、インデル形成を精確に測定することは可能ではなかった。この座位のプロトスペーサーは12個のグアニン塩基によって直接的に先行されており、これは、この部位のPCRおよび高スループットシーケンシングをランダムな挿入または欠失をしがちにする；複数の独立した非処置の対照サンプルでは、シーケンシングリードの20%ほども高い欠失率が観察された。このオフターゲット座位ではいずれかの処置条件において有意な塩基編集は検出されなかったので（図132および133）、この座位ではインデル形成もまた無視できると思われる。
【０５９８】
phred.II Q30スコアはベースコールの推定される99.9%の正確度に対応する[38]。ベースコールエラーがランダムに4塩基に分布することを想定する場合には、所与の位置における正しくないベースコールの0.1%確率は、0.1/4=0.025%というベースコールの下限

に対応する。この閾値よりも下に落ちるC→T編集パーセンテージは検出不可能として分類した。偶発的な脱アミノ化[39]またはPCRの間のポリメラーゼエラーもまたアーティファクトのC→T編集を導入し得る。塩基編集因子によって誘導されるC→T編集をアーティファクトのC→T編集率から見分けるために、非処置の対照細胞を各アンプリコンについてシーケンシングし、$p<0.05$を閾値とするスチューデントの両側t検定を用いて、特定の条件におけるC→T編集が統計的に有意であるかどうかを算出した。統計的に有意な編集率＞0.025%を有するオフターゲット部位は測定可能と考えられた。各サンプルのアライメントおよびクオリティフィルター処理されたリードの数は表17に包含されている。

【０５９９】

ゲノムDNA配列アライメントの統計的分析

別様に注意されない限り、本文において挙げられる平均値は$n \geqq 3$の独立した生物学的レプリケートを表しており、平均±標準偏差を記した。

【０６００】

図2および3に表示されている高スループットシーケンシングデータの統計的分析を、処置されたサンプルのオンおよびオフターゲット編集パーセンテージを陰性対照サンプル（非処置）で測定されたいずれかの編集と比較することによって行なった。スチューデントの両側t検定が用いられており、個々のp値は表16に示されている。*$p \leqq 0.05$、**$p \leqq 0.01$、および***$p \leqq 0.001$。編集が検出限界（0.025%）よりも下であるときには、有意性は算出しなかった；全ての非処置の対照サンプルは検出不可能な編集を示した。

【０６０１】

図128Aでは、塩基編集活性ウィンドウ内のシトシン（HEK293部位3およびEMX1ではC_4～C_8、FANCFおよびVEGFA部位2ではC_4～C_9）の編集を平均することによって、平均のオンターゲット塩基編集を算出した。

【０６０２】

塩基編集活性のsgRNA依存的な違いを考慮するために、塩基編集：インデル比を算出した（図130B）。この比は、インデルを含有するHTSリードのパーセンテージによって、C→T変換（各部位について塩基編集ウィンドウにおいて平均される）を有するHTSリードのパーセンテージを割ることによって創出された。上に記載されている通り、特定の座位のオフターゲット編集が検出限界よりも下である場合には、特異性比を算出する目的のために、それは検出法の推定される上側限度（0.025%）と控え目に想定された。

【０６０３】

in vitro編集されたDNAのデータ分析

シーケンシングリードはMiSeqレポーター（Illumina.）を用いて自動デマルチプレックスした。オンラインパッケージusegalaxy.orgを用いて、クオリティフィルター処理を行った[40]。30よりも少ないかまたはそれに等しいIlluminaクオリティスコアを有する個々の塩基はプレースホルダーヌクレオチド「N」に変換し、FASTQ Groomer、次にFASTA Maskerを用いた[41]。爾後に、もたらされたクオリティフィルター処理されたFASTQファイルを補足的な注1によって提供されるカスタムのpythonスクリプトによって分析した。標的DNA配列の両側をフランキングする2つの14塩基配列に対する厳密なマッチについて、シーケンシングリードをスキャンした。厳密なマッチが見出されない場合には、リードは分析から除外した。両方の14塩基配列が見つかり、かつそれらの間の配列の長さが予想されるプロトスペーサー長さ（20塩基）に等しい場合には、フランキング領域間に見出されたプロトスペーサー配列をセーブし、プロトスペーサー内の各部位において高スループットシーケンシングによってコールされた塩基を計算した。

【０６０４】

細胞培養

HEK293T（ATCC CRL-3216）およびNIH/3T3（ATCC CRL-1658）両方は10%（v/v）ウシ胎児血清（FBS）を添加したダルベッコ改変イーグル培地プラスGlutaMax（Thermo Fisher）によって37℃において5%CO_2で維持した。細胞はATCCから得、購入時にはATCCによってマイコプラズマ不含であることが検認された。

【０６０５】
HEK293T細胞への塩基編集因子のプラスミドトランスフェクション
HEK293T細胞を抗生物質不含培地の48ウェルコラーゲンコートBioCoatプレート（Corning）に接種し、およそ70%コンフルエンシーでトランスフェクションした。別様に注意されない限り、製造者のプロトコールに従ってウェルあたり1.5μlのLipofectamine 2000（Thermo Fisher）を用いて、750ngのBEおよび250ngのsgRNA発現プラスミドをトランスフェクションした。
【０６０６】
HEK293T細胞への塩基編集因子の蛋白質トランスフェクション
HEK293T細胞を250μL抗生物質不含培地の48ウェルコラーゲンコートBioCoatプレート（Corning）に接種し、およそ70%コンフルエンシーでトランスフェクションした。塩基編集因子蛋白質を1.1×モル過剰の必要なsgRNAと室温で5minに渡ってインキュベーションした。それから、複合体を1.5μLのLipofectamine 2000（Thermo Fisher）とインキュベーションし、プラスミド送達のための製造者のプロトコールに従ってトランスフェクションした。別様に注意されない限り、BE蛋白質を200nMの終濃度まで追加した（275μLという合計のウェル体積に基づく）。
【０６０７】
NIH/3T3細胞への塩基編集因子のプラスミドトランスフェクション
NIH/3T3細胞を抗生物質不含DMEM培地の48ウェルコラーゲンコートBioCoatプレート（Corning）に接種し、およそ75%コンフルエンシーでトランスフェクションした。別様に注意されない限り、製造者のプロトコールに従ってウェルあたり1μLのP3000試薬（Thermo Fisher）によって1.4μLのLipofectamine 3000を用いて、600ngのBEおよび200ngのsgRNA発現プラスミドをトランスフェクションした。
【０６０８】
NIH/3T3細胞への塩基編集因子の蛋白質トランスフェクション
NIH/3T3細胞を抗生物質不含DMEM培地の48ウェルコラーゲンコートBioCoatプレート（Corning）に接種し、およそ75%コンフルエンシーでトランスフェクションした。塩基編集因子蛋白質を1.1倍モル過剰の指示されているsgRNAと25℃で5minに渡ってインキュベーションした。それから、複合体を1.4μLのLipofectamine 3000（Thermo Fisher）とインキュベーションし、プラスミド送達のための製造者のプロトコールに従ってトランスフェクションした。その追加は蛋白質沈殿および塩基編集効率の縮減に至るので、P3000試薬は用いなかった。別様に注意されない限り、BE蛋白質は400nMの終濃度で追加した（275μLという合計ウェル体積に基づく）。
【０６０９】
カチオン性脂質によって封入されたBE3蛋白質:ガイドRNAの蝸牛内送達
全ての動物実験はマサチューセッツ眼科耳鼻科病院の施設内動物実験委員会によって認可された。先に記載されている通り混合遺伝学的バックグラウンドのP1-P2マウスにおいて蝸牛内送達を行なった[42]。外科手術前に、体温を下げることによってマウスに麻酔した。右耳の近くへ、耳後部に切り込みを入れ、耳骨胞を持ち上げて蝸牛を露出した。BE3蛋白質（57.7μM）を1:1.1モル比でsgRNA（100μM）と予め複合体化させ、それから1:1体積比でLipofectamine 2000（Thermo Fisher）と混合した。もたらされた溶液（1.2～1.5μL）を、250nL min^{-1}の速度で、ナノリットルマイクロポンプ（WPI，UMP3+Micro4+NanoFil）に取り付けられた蝸牛基底回転の中央階に蝸牛骨包からガラスピペット（末端直径5μm）によって注射した。注射後に、切開を閉じ、マウスを加熱パッド上に持って行って回復させた。3～4日後に、マウスの蝸牛をコルチ器、血管条、およびモダイオラスへと解剖した。各組織を5つおよび7つの間の別個の切片へとさらにマイクロダイセクションし、各サンプルについて別個にDNA抽出、次に高スループットシーケンシングを、上に記載されている通り行なった。図132および図134に提示されているデータは、各蝸牛領域について1つのマイクロダイセクションサンプルの抽出からもたらされたシーケンシングデータを示している。

【０６１０】
ゼブラフィッシュ胚へのBE3蛋白質:ガイドRNAのマイクロインジェクション
ゼブラフィッシュ（チュービンゲン系統）はボストン小児病院の内部規制審査を遵守して標準的な条件において維持した。適切なsgRNAまたは非関連sgRNA対照と1:1モル比で予め複合体化させられたおよそ2nLのBE3蛋白質を、1細胞ステージゼブラフィッシュ胚にインジェクションした（4.5μM終濃度）。受精の4日後に、50mM NaOH中で30分に渡って95℃において先に記載されている通り幼生からDNAを抽出し[43]、もたらされた溶液をTris-HClによって中和した。ゲノムDNAを定量し、PCRによって増幅し、上に記載される通りにシーケンシングした。
【０６１１】
蛋白質ゲル分析
示されている全ての蛋白質ゲルはプレキャスト4～12%ポリアクリルアミドBis-Tris Plus（Thermo Fisher）であった。それらはMOPS緩衝液（Thermo Fisher）によって180Vで50minに渡って流した。ローディングのためには、サンプルは、10minに渡る100mM DTTおよび変性のための1×ドデシル硫酸リチウム（LDS）サンプル緩衝液（Thermo Fisher）中における99℃への加熱によって調製した。製造者の説明書に従ってInstant Blue蛋白質染色（Expedion）を用いてゲルを染色した。
【０６１２】
細胞溶解物分析では、製造者の説明書に従う100μL B-PER（Thermo Fisher）による溶解の前に、誘導後一晩培養物の2mLを15,000gでペレット化した。
【０６１３】
データ入手可能性
この研究の知見を支持する高スループットシーケンシングデータは、NCBI配列リードアーカイブデータベースにアクセッション番号SRP097884で寄託した。蛋白質発現のためのHF-BE3およびBE3ならびに哺乳類発現のためのHF-BE3をコードするプラスミドは、Addgeneからアクセッション番号87439（pCMV-HF-BE3）、87438（pET42b-HF-BE3）、87437（pET42b-BE3）によって入手可能である。
参考文献

【表２５】

1 Cox, D. B., Platt, R. J. & Zhang, F. Therapeutic genome editing: prospects and challenges. *Nature medicine* **21**, 121-131, doi:10.1038/nm.3793 (2015).
2 Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096, doi:10.1126/science.1258096 (2014).
3 Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell* **168**, 20-36, doi:10.1016/j.cell.2016.10.044 (2017).
4 Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature biotechnology* **32**, 347-355, doi:10.1038/nbt.2842 (2014).
5 Hilton, I. B. & Gersbach, C. A. Enabling functional genomics with genome engineering. *Genome Res* **25**, 1442-1455, doi:10.1101/gr.190124.115 (2015).
6 Wyman, C. & Kanaar, R. DNA double-strand break repair: all's well that ends well. *Annual review of genetics* **40**, 363-383, doi:10.1146/annurev.genet.40.110405.090451 (2006).
7 Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281-2308, doi:10.1038/nprot.2013.143 (2013).
8 Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823, doi:10.1126/science.1231143 (2013).
9 Bodles-Brakhop, A. M., Heller, R. & Draghia-Akli, R. Electroporation for the Delivery of DNA-based Vaccines and Immunotherapeutics: Current Clinical Developments. *Mol Ther* **17**, 585-592, doi:10.1038/mt.2009.5 (2009).
10 Kay, M. A., Glorioso, J. C. & Naldini, L. Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics. *Nature medicine* **7**, 33-40, doi:Doi 10.1038/83324 (2001).
11 Midoux, P., Pichon, C., Yaouanc, J. J. & Jaffres, P. A. Chemical vectors for gene delivery: a current review on polymers, peptides and lipids containing histidine or imidazole as nucleic acids carriers. *Brit J Pharmacol* **157**, 166-178, doi:10.1111/j.1476-5381.2009.00288.x (2009).
12 Yin, H. *et al.* Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nature biotechnology* **32**, 551-553, doi:10.1038/nbt0914-952d (2014).

【表２６】

13 Hu, J. H., Davis, K. M. & Liu, D. R. Chemical Biology Approaches to Genome Editing: Understanding, Controlling, and Delivering Programmable Nucleases. *Cell chemical biology* **23**, 57-73, doi:10.1016/j.chembiol.2015.12.009 (2016).
14 Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420-424, doi:10.1038/nature17946 (2016).
15 Hess, G. T. *et al.* Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nature methods* **13**, 1036-1042, doi:10.1038/nmeth.4038 (2016).
16 Ma, Y. *et al.* Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nature methods* **13**, 1029-1035, doi:10.1038/nmeth.4027 (2016).
17 Nishida, K. *et al.* Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729, doi:10.1126/science.aaf8729 (2016).
18 Li, J., Sun, Y., Du, J., Zhao, Y. & Xia, L. Generation of targeted point mutations in rice by a modified CRISPR/Cas9 system. *Mol Plant* **In Press**, doi:10.1016/j.molp.2016.12.001 (2016).
19 Koyama, S., Ishii, K. J., Coban, C. & Akira, S. Innate immune response to viral infection. *Cytokine* **43**, 336-341, doi:10.1016/j.cyto.2008.07.009 (2008).
20 Kleinstiver, B. P. *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490-495, doi:10.1038/nature16526 (2016).
21 Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology* **31**, 827-832, doi:10.1038/nbt.2647 (2013).
22 Fu, Y. F. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology* **31**, 822-826, doi:10.1038/nbt.2623 (2013).
23 Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature biotechnology* **31**, 839-843, doi:10.1038/nbt.2673 (2013).
24 Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature biotechnology* **33**, 187-197, doi:10.1038/nbt.3117 (2015).
25 Guilinger, J. P. *et al.* Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. *Nature methods* **11**, 429-435, doi:10.1038/nmeth.2845 (2014).
26 Singh, R., Kuscu, C., Quinlan, A., Qi, Y. J. & Adli, M. Cas9-chromatin binding information enables more accurate CRISPR off-target prediction. *Nucleic acids research* **43**, 677-683, doi:ARTN e118
10.1093/nar/gkv575 (2015).
27 Liang, X. Q. *et al.* Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J Biotechnol* **208**, 44-53, doi:10.1016/j.jbiotec.2015.04.024 (2015).
28 Wang, M. *et al.* Efficient delivery of genome-editing proteins using bioreducible lipid nanoparticles. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 2868-2873, doi:10.1073/pnas.1520244113 (2016).
29 Zuris, J. A. *et al.* Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nature biotechnology* **33**, 73-80, doi:10.1038/nbt.3081 (2015).

【表２７】

30 DeWitt, M. A. *et al.* Selection-free genome editing of the sickle mutation in human adult hematopoietic stem/progenitor cells. *Sci Transl Med* **8**, 360ra134, doi:10.1126/scitranslmed.aaf9336 (2016).
31 Kim, K. *et al.* Highly efficient RNA-guided base editing in mouse embryos. *Nature biotechnology*, doi:10.1038/nbt.3816 (2017).
32 Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nature biotechnology* **34**, 339-344, doi:10.1038/nbt.3481 (2016).
33 Lin, Y. C. *et al.* Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nature communications* **5**, 4767, doi:10.1038/ncomms5767 (2014).
34 Davis, K. M., Pattanayak, V., Thompson, D. B., Zuris, J. A. & Liu, D. R. Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat Chem Biol* **11**, 316-318, doi:10.1038/Nchembio.1793 (2015).
35 Lin, S., Staahl, B., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife*, e04766, doi:10.7554/eLife.04766 (2014).
36 Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature biotechnology* **34**, 184-191, doi:10.1038/nbt.3437 (2016).
37 Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).
38 Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8**, 186-194 (1998).
39 Liu, M. & Schatz, D. G. Balancing AID and DNA repair during somatic hypermutation. *Trends in immunology* **30**, 173-181, doi:10.1016/j.it.2009.01.007 (2009).
40 Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic acids research* **44**, W3-W10, doi:10.1093/nar/gkw343 (2016).
41 Blankenberg, D. *et al.* Manipulation of FASTQ data with Galaxy. *Bioinformatics* **26**, 1783-1785, doi:10.1093/bioinformatics/btq281 (2010).
42 Hu, L. X. *et al.* Diphtheria Toxin-Induced Cell Death Triggers Wnt-Dependent Hair Cell Regeneration in Neonatal Mice. *J Neurosci* **36**, 9479-9489, doi:10.1523/Jneurosci.2447-15.2016 (2016).
43 Meeker, N. D., Hutchinson, S. A., Ho, L. & Trede, N. S. Method for isolation of PCR-ready genomic DNA from zebrafish tissues. *Biotechniques* **43**, 610, 612, 614 (2007).
44 Zuris, J.A. et al. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. Nature biotechnology 33, 73-80 (2015).
45 Tsai, S.Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nature biotechnology 33, 187-197 (2015).
46 Doench, J.G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nature biotechnology 34, 184-191 (2016).
47 Haeussler, M. et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol 17, 148 (2016).

補足情報

補足的な注1：クオリティフィルター処理されたin vitro編集されたDNAを分析するために用いられたPythonスクリプト。

【数１】

```
from __future__ import print_function
from __future__ import division

import Bio #This will import the BioPython suite
from Bio import SeqIO #Necessary to read/write sequence handles
from Bio.Seq import Seq
import os
import collections
import csv

inputfile = "please_specify_your_input_file_here_containing_filtered_reads" #specify th
e filenames that contain sequences
filenames = []

for file in os.listdir(inputfile):
    if file.endswith(".fastqsanger"):
        filenames.append(file)

spacer = []
list_of_filenames = []

for file in filenames:
    site = {}
    output = open(file + ".txt", "w")
    list_of_filenames.append(file + ".txt") #allows calling of the txt files that come
from fastq files later
    for rec in SeqIO.parse(file, "fastq"):
        split1=rec.seq.tostring().split("GTTCGCGGCGATCG") #14-
base pair constant_region_before_protospacer
        if len(split1)>=2:
            split2=split1[1].split("TGGATCGCCTGGCA") #14-
base pairc constant_region_after_protospacer
            site=split2[0]
            if len(site)==20:
                output.write(site + "\n")

BASES = 'ATGCN'
UNRECOGNIZED = 'X'
BASE_SEPERATOR = dict(zip(BASES, ',,,,\n'))
a_index = 0
t_index = 1
g_index = 2
c_index = 3
n_index = 4

def get_counts_by_column(base, count, library):
    current_count = library[count]
    if base == 'A':
        current_count[a_index] += 1
    elif base == 'T':
        current_count[t_index] += 1
    elif base == 'G':
        current_count[g_index] += 1
    elif base == 'C':
        current_count[c_index] += 1
    elif base == 'N':
        current_count[n_index] += 1

```

【数２】

```
    def dna_counts(list_of_sequences, sample):
        first_oligo = list_of_sequences[0]
        for i in range (len(first_oligo)):
            sample.append([0,0,0,0,0])
        for j in range(len(first_oligo)):
            for i in range(len(list_of_sequences)):
                get_counts_by_column(list_of_sequences[i][j], j, libname)


    for file in list_of_filenames:
        spacer_list = open(file).read().splitlines()
        output2=[]
        dna_counts(spacer_list, output2)
        with open(file  + ".csv", "wb") as f:
            writer = csv.writer(f)
            writer.writerows(output2)
```

補足的な注2：本研究に用いられた蛋白質の配列

発現されたBE3の蛋白質配列

MGSSHHHHHHSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIW
RHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTL
FIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPH
LWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSE
TPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK
KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEK
KNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA
KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIH
LGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMR
KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHD
LLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD
VDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVIT
LKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYD
VRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGR
DFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT
VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA
AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETG
KQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQ
DSNGENKIKMLSGGSPKKKRKV（配列番号185）

発現されたHF-BE3の蛋白質配列

MGSSHHHHHHSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIW
RHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTL
FIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPH

LWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSE
TPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK
KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEK
KNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA
KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIH
LGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMTAFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMR
KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHD
LLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
GWGALSRKLINGIRDKQSGKTILDFLKSDGFANRNFMALIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD
VDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRAITKHVAQILDSRMNTKYDENDKLIREVKVIT
LKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYD
VRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGR
DFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT
VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA
AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETG
KQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQ
DSNGENKIKMLSGGSPKKKRKV（配列番号186）

補足的な注3：本研究に用いられたオリゴヌクレオチドの配列

ヒト細胞のHTSのためにオフターゲットゲノムDNAを増幅するために用いられた未公開のプライマー（配列番号506～513）

【化１２５】

fwd_VEGFA_site2_off_target_1_human	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCTACAAGTAACAGTCCAAGAA
rev_VEGFA_site2_off_target_1_human	TGGAGTTCAGACGTGTGCTCTTCCGATCTTTCTGCAACTTAACTTACGTGAAA
fwd_VEGFA_site2_off_target_2_human	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACCAAGCCCATTTGTCCAGG
rev_VEGFA_site2_off_target_2_human	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCTTCTTTTTGAGCTTTGGGC
fwd_VEGFA_site2_off_target_3_human	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCATACCAGCAGCAGTTCC
rev_VEGFA_site2_off_target_3_human	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCACCTCAGCTCCTGCAC
fwd_VEGFA_site2_off_target_4_human	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCCACTGATTCTACACCATGGT
rev_VEGFA_site2_off_target_4_human	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGAGTTCCCAACCTTTTTGACA

他のプライマー（HEK_3, EMX1, FANCFに関連するオフターゲット部位のための）は先に公開された。

マウス細胞のHTSのためにオフターゲットゲノムDNAを増幅するために用いられたプライマー（配列番号514～521）

【化１２６】

fwd_VEGFA_site2_off_target_1_murine	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTGGCTGGAGATTCAGAGACAC
rev_VEGFA_site2_off_target_1_murine	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGCCCCTTCTGACACACATAC
fwd_VEGFA_site2_off_target_2_murine	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACCCCTCAAGGCTTGACATTTC
rev_VEGFA_site2_off_target_2_murine	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGAAAAGTTGGGAGAGGGGATG
fwd_VEGFA_site2_off_target_3_murine	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTTGTACCCCAGTCCCCTCATC
rev_VEGFA_site2_off_target_3_murine	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGAAGTTACGGGGATGTCACTTG
fwd_VEGFA_site2_off_target_4_murine	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTAACATCCAGTCTCCCAAACACA
rev_VEGFA_site2_off_target_4_murine	TGGAGTTCAGACGTGTGCTCTTCCGATCTACACACACACACTACTAGGACA

マウス細胞のHTSのためにオンターゲットゲノムDNAを増幅するために用いられたプライマー（配列番号522～523）

【化１２７】

fwd_VEGFA_site2_on_target_murine	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCGCTACTACGGAGCGAGAAG
rev_VEGFA_site2_on_target_murine	TGGAGTTCAGACGTGTGCTCTTCCGATCTACAGGGGCAAAGTGAGTGAC

sgRNAのT7転写の基質としての用をなすためのPCR生成物を創出するために用いられたプライマー（配列番号524～529）

【化１２８】

rev_sgRNA_T7: used in all cases	AAAAAAAGCACCGACTCGGTGCCAC
fwd_sgRNA_T7_EMX1	TAATACGACTCACTATAGGGAGTCCGAGCAGAAGAAGAAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_FANCF	TAATACGACTCACTATAGGGGAATCCCTTCTGCAGCACCGTTTTAGAGCTAGAAATAGCA

【化１２９】

fwd_sgRNA_T7_HEK_site_3	TAATACGACTCACTATAGGGGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_VEGFA_site_2	TAATACGACTCACTATAG GACCCCCTCCACCCCGCCTCGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_TC_repeat_in_vitro	TAATACGACTCACTATAGGTCTCTCTCTCTCTCTCTCTCTCGTTTTAGAGCTAGAAATAGCA

sgRNAトランスフェクションプラスミドを創出するために用いられたプライマー（配列番号530～531）

先に記載される通り、pFYF1320プラスミドを鋳型として用いた（Komor et al.）。他のsgRNAプラスミドの配列は先に報告されている。

【化１３０】

rev_sgRNA_plasmid	GGTGTTTCGTCCTTTCCACAAG
fwd_VEGFA_site_2	GACCCCCTCCACCCCGCCTCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC

in vitroデアミナーゼアッセイに用いられたssDNA基質の配列（配列番号532-533）

【化１３１】

fwd_TC_repeat_substrate	ACGTAAACGGCCACAAGTTCGCGGCGATCGTCTCTCTCTCTCTCTCTCTCTGGATCGCCTGGCATCTTC TTCAAGGACG
rev_TC_repeat_substrate	CGTCCTTGAAGAAGATGCCAGGCGATCCAGAGAGAGAGAGAGAGAGAGAGACGATCGCCGCGAACTTGT GGCCGTTTACGT

ヒト細胞のHTSのためにオフターゲットゲノムDNAを増幅するために用いられた先に公開されているプライマー（配列番号534～557）

【化１３２】

fwd_EMX1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCAGCTCAGCCTGAGTGTTGA
rev_EMX1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCGTGGGTTTGTGGTTGC
fwd_FANCF_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCATTGCAGAGAGGCGTATCA
rev_FANCF_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGTCCCAGGTGCTGAC
fwd_HEK293_site3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATGTGGGCTGCCTAGAAAGG
rev_HEK293_site3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCCAGCCAAACTTGTCAACC
fwd_EMX1_off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGTAGCCTCTTTCTCAATGTGC
rev_EMX1_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCTTTCACAAGGATGCAGTCT
fwd_EMX1_off2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGAGCTAGACTCCGAGGGGA
rev_EMX1_off2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCTCGTCCTGCTCTCACTT
fwd_EMX1_off3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGAGGCTGAAGAGGAAGACCA
rev_EMX1_off3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGCCCAGCTGTGCATTCTAT
fwd_FANCF_off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAACCCACTGAAGAAGCAGGG
rev_FANCF_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGTGCTTAATCCGGCTCCAT
fwd_FANCF_off2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCAGTGTTTCCATCCCGAA
rev_FANCF_off2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCTGACCTCCACAACTCT

【化１３３】

fwd_FANCF_off3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTGGGTACAGTTCTGCGTGT
rev_FANCF_off3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCACTCTGAGCATCGCCAAG
fwd_HEK293_site3_off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCCCTGTTGACCTGGAGAA
rev_HEK293_site3_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACTGTACTTGCCCTGACCA
fwd_HEK293_site3_off2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTGGTGTTGACAGGGAGCAA
rev_HEK293_site3_off2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGAGATGTGGGCAGAAGGG
fwd_HEK293_site3_off3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGAGAGGGAACAGAAGGGCT
rev_HEK293_site3_off3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCCAAAGGCCCAAGAACCT

ゼブラフィッシュのHTSのためにオンターゲットゲノムDNAを増幅するために用いられたプライマー（配列番号558～563）

【化１３４】

fwd_TYR1_zebrafish	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTTCCCCCGAGTCTGCACCT
rev_TYR1_zebrafish	TGGAGTTCAGACGTGTGCTCTTCCGATCTCGAACTTGCATTCGCCGCAA
fwd_TYR2_zebrafish	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTCTGCCTTGGCATCGGGTG
rev_TYR2_zebrafish	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACCATACCGCCCCTAGAACTAACATTC
fwd_TYR3_zebrafish	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACAACTGCTTTCCATGGTGTGT
rev_TYR3_zebrafish	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCCAGGGCTTTCGTGGAGA

【表２８】

部位	配列	配列番号
TYR1	$GTC_3AGGTC_8GAGGGTTCTGTCAGG$	564
TYR2	$CTTC_4C_5AGGATGAGAACACAGAGG$	565
TYR3	$CAAC_4C_5AC_7TGCTCAAAGATGCTGG$	566

補足的な表1．この研究において研究されたゼブラフィッシュゲノム座位のプロトスペーサーおよびPAM配列。

【０６１４】

例17：より高い効率および生成物純度を有するC:G→T:A塩基編集因子

塩基編集は、二本鎖DNA切断を誘導することまたは相同組み換え修復を要求することなしの、Cas9バリアントとシチジンデアミナーゼとの人工の融合体を用いる標的C:G塩基対からT:A塩基対へのプログラム可能な変換である（1）。第3世代塩基編集因子（BE3）および関係するテクノロジーは多くの研究者によって幅広い範囲の生物に首尾良く用いられてきた（2～13）。塩基編集の生成物分布－標的C:G塩基対が、所望のT:A生成物と併せて、所望ではない副生成物の混合物に変換される頻度－は、標的部位依存的な様式で変わる（2,3,6～8）。ここで、我々はヒト細胞における塩基編集成果の決定要因を特徴付け、所望ではない生成物の形成がウラシルN-グリコシラーゼ（UNG）に依存し、塩基編集活性ウィンドウ内に1つのCのみを含有する標的部位において生起することがより蓋然的であることを確立する。いくつかの配列について塩基編集効率を増大させるシチジンデアミナーゼホモログを用いる構築物CDA1-BE3およびAID-BE3を操作した。加えて、塩基編集された中間体へのUNGのアクセスをより効率的にブロックする第4世代S. pyogenes Cas9由来塩基編集因子（BE4）をもまた操作した。BE3と比較して、BE4はC:G→T:A塩基編集の効率をおよそ50%増大させながら、所望ではない副生成物の頻度を半減させる。これらの改善は、S. aureus Cas9由来BE4（SaBE4）を生むためにもまた適用した。これはBE4よりも実質的に小さく、代替的な標的化範囲を有する。

【０６１５】

導入

従来のゲノム編集方法はゲノム標的座位に二本鎖DNA切断（DSB）を導入する（14）。DSB損傷に対する細胞応答は主として非相同末端結合（NHEJ）および関係するプロセスによって進む（15）。NHEJは通常はDSBをフランキングする2つの末端を再度つなぎ合わせるが、典型的なゲノム編集条件においては、DSBが連続的に再導入され、ついには、DSBの部位における挿入および欠失（インデル）または転座の蓄積ならびに対応するゲノム座位の破壊をもたらす（16）。活発に分裂している細胞もまたDSBの周辺の領域に対する相同性を含有するドナーDNA鋳型の存在下において相同組み換え修復（HDR）を開始す

ることによってDSBに応答し得、これは、研究者がNHEJによって可能であるよりも正確にかつ予測可能にゲノムを操ることを許す（17）。HDR依存的なゲノム編集は、NHEJ成果との競合および有糸分裂へのHDRの依存性から生ずる低い効率によって限定される（18）。

【０６１６】

塩基編集の開発が報告されており、これは、HDRまたはDSBの導入を要求することなしに、プログラム可能な様式で、C:G塩基対からT:A塩基対への不可逆的な直接的変換を可能にする（1）。塩基編集因子は、触媒的に棄損されたCas9蛋白質および塩基除去修復阻害因子にテザリングされた一本鎖DNA特異的シチジンデアミナーゼ酵素からなる（1,4,9,10）。対応するガイドRNAによってプログラムされ、Cas9バリアントは当該ゲノム座位に結合する。蛋白質:RNA:DNAの3者「Rループ」複合体（19）の形成は、テザリングされたシチジンデアミナーゼ酵素の基質としての用をなす一本鎖DNAの小さい（～5nt）ウィンドウを露出する。このウィンドウ内のいずれかのシチジンは加水分解的にウラシルへと脱アミノ化され、G:U中間体をもたらす。

【０６１７】

塩基除去修復（BER）はG:Uミスマッチに対する細胞の一次応答であり、ウラシルN-グリコシラーゼ（UNG）によるウラシルの切り出しによって開始される（20）。編集されたG:U中間体をUNGによる切り出しから保護するための作業において、83アミノ酸のウラシルグリコシラーゼ阻害因子（UGI）を触媒的にデッドなCas9（dCas9）のC末端に直接的に融合させた（1）。G:UミスマッチのGをAによって選好的に置き換えるように細胞性のDNAミスマッチ修復システムを操るために、dCas9のAla840アミノ酸をHisに復帰させ、Cas9蛋白質が新たに形成されたウラシルに対向するDNA鎖にニックを入れることを可能にし、G:U中間体から所望のA:UおよびA:T生成物へのずっと効率的な変換をもたらした（1）。これらの2つの工学作業を組み合わせることは、UGIに4アミノ酸リンカーを介して共有結合的に連結されるS. pyogenes dCas9（A840H）に16アミノ酸リンカーを介してテザリングされたAPOBEC1シチジンデアミナーゼ酵素の3部融合体からなる単一の蛋白質BE3をもたらした（1）。今や、BE3および関係する塩基編集因子は、植物ゲノム編集、in vivoの哺乳類ゲノム編集、標的化された変異導入、およびノックアウト研究を包含する幅広い種々の用途に採用されている（2～13）。最近、変改されたPAM要件（4）、狭められた編集ウィンドウ（4）、縮減されたオフターゲット編集（10）、および小分子依存性（21）を有するBE3バリアントを報告することによって、塩基編集の範囲は拡大されてきた。

【０６１８】

いくつかの座位では、BE3などの塩基編集因子は、標的C:G塩基対が所望のT:A生成物よりもむしろG:CまたはA:T塩基対に変換されている所望ではない副生成物を生ぜしめる（2,3,6～8）。ここで、我々は塩基編集生成物純度の決定要因を照らし出し、UNG活性が所望ではない副生成物の形成に要求されることを確立する。所望ではない生成物を最小化するためには、ウラシル中間体へのUNGアクセスをブロックすることは、編集ウィンドウ内に単一のCがある標的座位にとって特に重大であることが決定された。BE3を包含する先に記載された塩基編集因子と比較してより高い効率および多大に改善された生成物純度を有する塩基編集を行う、第4世代塩基編集因子BE4およびSaBE4を操作した。

【０６１９】

結果

副生成物形成にはUNG活性が要求される

ウラシル含有DNA鎖内における脱塩基部位の形成およびエラープローンな解消が原因で、所望ではない塩基編集副生成物が塩基除去修復の間に生じ得る。ウラシルN-グリコシラーゼ（UNG）を欠く細胞における塩基編集の生成物純度が改善するかどうかを決定するために、BE3およびEMX1、FANCF、HEK2、HEK3、HEK4、またはRNF2座位を標的化するsgRNAをコードするプラスミドによって、HAP1細胞（ハプロイドヒト細胞株）およびHAP1 UNG^{-}細胞をヌクレオフェクションした（標的配列については図135Bを見よ）。ヌ

クレオフェクションの3日後に、ゲノムDNAを抽出し、標的座位をPCRによって増幅し、高スループットDNAシーケンシング（HTS）によって分析した。塩基編集生成物純度は、標的CがTへと編集されている編集されたシーケンシングリード（標的CがA、G、またはTに変換されたリード）のパーセントとして定められる。BE3によって処置されたHAP1細胞の塩基編集生成物純度は6つの座位の12個の標的Cにおいて平均68±6%であった（n=3つの生物学的レプリケートの平均±S.D.）。HAP1 UNG^{-}細胞では、試験された全ての12個の標的Cは生成物純度＞98%で塩基編集された（図135A）。加えて、全ての6つの試験された座位におけるインデル頻度はUNGノックアウトによって7～100倍減少した（図135C）。これらのデータは、UNG活性が塩基編集の間の所望ではない生成物の形成にとって必要であると強くほのめかしており、脱塩基部位形成およびエラープローンポリメラーゼによる爾後の塩基除去修復が標的ヌクレオチドのランダム化とインデルをもたらす時たまの鎖切断とに至るモデルと整合している。

【０６２０】

複数の編集可能なCを有する標的はより高い生成物純度を見せる

BE3による塩基編集効率はGの直ちに下流である（全てではないが）いくつかの標的Cについてはより低くあり得（1）、APOBEC1の公知の配列選好性と整合している（22）（図136A）。かかる標的を効率的に編集するための作業において、異なる配列選好性を有する3つの一本鎖DNA特異的シチジンデアミナーゼ酵素CDA1（CDA1-BE3を創出するため）、AID（AID-BE3を創出するため）、またはAPOBEC3G（APOBEC3G-BE3を創出するため）によってAPOBEC1デアミナーゼを置き換えたBE3バリアントを創出した（23）。これらのBE3バリアントおよびEMX1、FANCF、HEK2、HEK3、HEK4、またはRNF2座位を標的化するsgRNAをコードするプラスミドによって、HEK293T細胞をトランスフェクションした。トランスフェクションの3日後に、ゲノムDNAを抽出し、標的座位をPCRによって増幅し、塩基編集についてHTSを用いて査定した。直ちにGに後続する標的Cのより効率的な編集がBE3と比較してCDA1-BE3およびAID-BE3によって観察された（図136A、図140A～D、および図146～151）。一般的に、CDA1-BE3およびAID-BE3は、Gに後続しない標的CにおいてBE3よりも低い編集効率を見せた（図140A～D）。対照的に、APOBEC3G-BE3は予測不可能な配列選好性を見せ、BE3と比較してC→T編集の総体的な低い収率を有した。これらの知見は、CDA1-BE3およびAID-BE3がいくつかの標的5'-GC-3'配列についてBE3と比べてより高い編集効率をオファーし得ることを示唆する。

【０６２１】

これらのデータを分析する間に、CDA1-BE3およびAID-BE3がBE3よりも多くのCを編集する部位においては、CDA1-BE3およびAID-BE3の生成物純度は典型的にはBE3のものよりも高いことが注意された（図136A～D）。例えば、HEK4座位においては、BE3は1つのCのみを効率的に編集するが（Gによって先行されていないC）、CDA1-BE3およびAID-BE3両方は3つのCを編集する（図140A～C）。この座位におけるBE3の生成物純度は50±7%であり（n=3つの生物学的レプリケートの平均±S.D.）、CDA1-BE3およびAID-BE3の生成物純度はそれぞれ97±2%および93±2%である。その上、それぞれ84±3%および91±2%の生成物純度でBE3によって編集されるEMX1およびFANCFは、同等の効率で編集される複数のCを含有し（図S2）、それぞれ28±3%および64±3%というずっと低い生成物純度でBE3によって編集されるHEK2およびRNF2は、不等な効率で編集される複数のCを含有する（図141A～C）。HEK2座位内の両方のCを同等の効率で編集するCDA1-BE3およびAID-BE3は、この座位におけるずっと高い生成物純度（それぞれ85±5%および81±4%）を見せる（図136A～Dおよび図140C）。HEK2およびRNF2部位において、複数のCがBE3によって同等の効率で最初にUに変換され、それからDNA修復システムによって異なる効率でプロセシングされるという可能性は排除された。これに鑑みると、これらの部位がCDA1-BE3またはAID-BE3に対してBE3によって処置されるときには、観察された異なる生成物分布よりもむしろ、類似の生成物分布が予想されるであろう（図136Bおよび図146～151）。代わりに、単離されたG:UはG:U損傷のクラスターよりも容易くUNGによってプロセシングされ得る。BE3のシチジンデアミナーゼドメインの処理能力（1,2

4）は、複数の編集可能なCを含有する座位におけるBE3の所在時間を増大させ得、それによって、1つの編集可能なCを含有する座位よりも有効にUNGによるアクセスをブロックすることが可能である。

【０６２２】

生成物純度、個々のシーケンシングリードの編集されたCの数、およびUNG活性の間の関係性をさらに分析した。明示的なUNG阻害の非存在下における塩基編集されたDNAの運命を明らかにするために、BE3のUGIコンポーネントを取り除いてBE3Bを創出した。BE3またはBE3BおよびEMX1、FANCF、HEK2、HEK3、HEK4、またはRNF2座位を標的化するsgRNAをコードするプラスミドによって、HEK293T細胞をトランスフェクションした。予想される通り、上で確立された塩基編集成果を多様化することにおけるUNGの役割に鑑みると、全ての標的Cにおける生成物純度は、BE3によって処置されたDNAと比較してBE3Bによって処置されたDNAでは多大に減少し、非Tを含有する編集生成物の画分は平均で1.8±0.4倍増大した（図142B）。

【０６２３】

複数C部位HEK2、HEK3、およびRNF2を標的化するsgRNAおよびBE3またはBE3Bどちらかによって処置されたHEK293T細胞からの個々のDNAシーケンシングリードを分析した。各部位について、一次標的Cは最も効率的に改変されたヌクレオチドに指名される。全ての3つの部位において、一次標的Cの所望ではないC→非T編集を含有するシーケンシングリードの平均で80±10%は、その1塩基編集イベントのみを見せた（図142A～Dおよび図143）。対照的に、同じ3つの複数C部位において、一次標的CのクリーンなC→T編集を含有するシーケンシングリードのずっと低い平均で32±4%が、そのクリーンな1塩基編集イベントのみを見せた（図142A～Dおよび図143）。加えて、BE3Bによって処置された、編集ウィンドウ内に1つのCのみを含有する部位HEK4 DNAの生成物の分布は、A:G:Tでは1:3:1の比をほぼ踏襲した（図143D）。総合すると、これらの観察は、所与の標的中の単一のシチジンがUGIの非存在下においてUに変換されるときには、それはUNGによって開始されるBERによって効率的にプロセシングされて生成物の混合物を与えることを指示している。

【０６２４】

これらのデータは、クラスターしたG:Uミスマッチが単離されたG:Uミスマッチとは異なって処理され、クリーンなC→T編集物を生成することがより蓋然的であるモデルと整合している。単一のC→T編集イベントのみが所望であるときに、上の観察は、所望ではない副生成物を最小化するためにはUNG阻害が決定的であることを示唆している。しかしながら、dCas9-デアミナーゼ融合体、例えばTAM（8）およびCRISPR-X（2）を用いて標的化されたランダムな変異導入を行うときには、上の観察は、1つの編集可能なCのみを有する標的部位は生成物混合物を最大化するであろうということを示唆している。

【０６２５】

改善された生成物純度のためのBE3アーキテクチャの最適化

BE3のUGIコンポーネントを一本鎖DNA結合（SSB）蛋白質によって置き換えて、SSBが、UNGによってアクセスされることからRループのウラシル含有ssDNA部分をブロックし得るようなSSB-BE3を生んだ。SSB-BE3による塩基編集効率の大きい減少が観察され、4つの部位における全ての7つのCは平均で1.9±0.5%のみのC→T変換を見せた（図137C）。

【０６２６】

塩基編集成果を決定するステップの間のAPOBEC、UGI、およびUNGの相対的な位置どりは公知ではないので、UNG阻害を改善するための作業において、UGIをBE3のN末端に再配置した（N-UGI-BE3）。UGIをBE3のN末端に動かすことは、BE3と比較して2.3±0.6倍の全ての7つの試験された標的CにおけるC→T編集パーセンテージの平均の減少（図137C）、およびBE3と比較して平均2.2±0.5倍である全ての4つの部位における総体的な生成物純度の減少（図137B）をもたらした。

【０６２７】

対照的に、UGIの追加のコピーをBE3のC末端に付加することは（BE3-2xUGI）、BE3に対して相対的に生成物純度の大きい増大とBE3のものと同等のC→T編集パーセンテージとをもたらした。非T編集生成物は試験された4つの座位において平均で2.2±0.8倍減少した（図137B）。これらの観察は、UGIの第2のコピーの追加がG:U塩基編集中間体へのUNGのアクセスを実質的に減少させ、それによって生成物純度を多大に改善することを示唆している。

【０６２８】

上の実験はコンポーネントのアーキテクチャに対する塩基編集成果の敏感さをもまた明らかにしているので、次に、我々はBE3コンポーネント間のリンカーを最適化して、生成物純度および編集効率をさらに増大させた。我々は、rAPOBEC1-dCas9(A840H)リンカーを16アミノ酸（BE3）から32アミノ酸（BE3C）に、dCas9(A840H)-UGIリンカーを4（BE3）から9（BE3D）から16アミノ酸（BE3E、図138A）に変えた。dCas9(A840H)-UGIリンカーが4アミノ酸（BE3）の代わりに長さ9アミノ酸残基（BE3D）であるときには、非T生成物形成は平均で1.3±0.1倍減少し（図138D）、C→T編集効率の明らかな違いを有さなかった（図138C）。rAPOBEC1-dCas9(A840H)リンカーを16アミノ酸（BE3）から32アミノ酸（BE3C）に増大させることは、HEK2座位においてC→T編集効率を平均で1.2±0.1倍上昇させた（図138C）。この座位は、先に、最も不均一に編集された試験された複数C部位であり（図141A～C）、このリンカーを延長することは、2.6±0.2倍から1.8±0.1倍への、C_4と比べてC_6の選好的な編集（C_6において編集されているシーケンシングリードのパーセンテージ対C_4のものの比）の縮減に至った。我々は、このより長いリンカーは、Rループ中のssDNAへのより良好なアクセスをデアミナーゼに許し得、複数の標的Cが塩基編集ウィンドウ内に存在するときにはより一様な脱アミノ化をもたらし得ることを推論した。BE3Cは試験された他の座位においてもまた同等のまたは改善された塩基編集効率および生成物純度を見せた（図138C～D）。

【０６２９】

向上した効率および生成物純度を有するC:G→T:A塩基編集因子BE4

全ての3つの改善－rAPOBEC1-dCas9リンカーを32アミノ酸に延長すること、dCas9-UGIリンカーを9アミノ酸に延長すること、およびUGIの第2のコピーを別の9アミノ酸リンカーによって構築物のC末端に付加することを組み合わせることによって、塩基編集因子構築物BE4を操作した。Nishida et. al.（9）によって報告された代替的な塩基編集因子構築物Target-AIDをもまたBE4と同じプラスミドバックボーンにクローニングした。BE3、BE4、またはTarget-AIDおよびEMX1、FANCF、HEK2、HEK3、HEK4、またはRNF2座位を標的化するsgRNAをコードするプラスミドによって、HEK293T細胞をトランスフェクションした。トランスフェクションの3日後に、ゲノムDNAを抽出し、標的座位をPCRによって増幅し、HTSによって分析した。BE3に対して相対的に、BE4について全ての12個の編集されたCにおける1.5±0.3倍のC→T編集効率の平均の増大が観察された（図139C）。分析された同じ位置におけるTarget-AIDのC→T編集の平均の効率はBE3のものよりも1.5±0.5倍低く、BE4のものよりも2.1±0.5倍低かったが、CDA1デアミナーゼを用いるTarget-AIDはBE3およびBE4に対して相対的にシフトした編集ウィンドウを有するように見え、最適な編集を位置C_3およびC_4の周りに有することに注意することが重要である（図139C）。このシフトされた編集ウィンドウは、Target-AIDとBE3またはBE4との間における効率および生成物純度の比較を難しくする。なぜなら、同じガイドRNAを用いるときでさえも、所与の標的Cは異なる編集ウィンドウ内のより最適なまたはより最適でない位置に在り得るからである。

【０６３０】

より多大なC→T編集効率に加えて、BE4は試験された全てのゲノム座位においてBE3に対して相対的に実質的に改善された生成物純度をもまた見せ、2.3±0.3倍の非T生成物形成の平均の減少を有した（図139D）。インデルに至り得る塩基除去修復をさらに妨げることから予想される通り（25）、BE3と比較してBE4処置後の全ての6つの座位において平均2.3±1.1倍であるインデル率の減少もまた観察された（図144A～C）。一緒にする

と、これらの結果は、試験された全ての座位において、BE4がC→T編集の高い効率、高い生成物純度、および低いインデル形成率をオファーすることを指示している。

【０６３１】

BE4の改善をS. aureus Cas9（26）とインテグレーションしてSaBE4を創出した。これは、S. pyogenes dCas9(A840H)をより小さいS. aureus dCas9(A580N)によって置き換え、その代替的なPAM要件が原因で異なる標的にアクセスし得る。SaBE3（4）またはSaBE4およびFANCF、HEK3、またはHEK4座位を標的化するsgRNAをコードするプラスミドによって、HEK293T細胞をトランスフェクションした。BE4およびBE3を比較する結果と整合して、我々は、全ての10個の編集されたCにおいて、SaBE4ではSaBE3に対して相対的に1.4±0.2倍というC→T編集効率の平均の増大を（図145A）、所望ではない非T編集生成物の1.8±0.5倍の平均の減少（図145B）と共に観察した。これらの結果は、BE4の向上から生ずる塩基編集効率および生成物純度のゲインが他のCas9ホモログに由来する塩基編集因子にもまた当てはまることを指示している。

【０６３２】

材料と方法

プラスミドのクローニング

この研究における全てのプラスミドはPhusion Uホットスタートポリメラーゼ（Thermo Fisher）を用いてUSERクローニングによって創出した。デアミナーゼおよびSSB遺伝子はgBlocks遺伝子断片として合成し（Integrated DNA Technologies）、Target-AIDはAddgeneから得た（プラスミド#79620）。蛋白質配列は補足的な注にリスト化されている。

【０６３３】

細胞培養

HEK293T（ATCC CRL-3216）細胞は10%（v/v）ウシ胎児血清（FBS）を添加したダルベッコ改変イーグル培地プラスGlutaMax（Thermo Fisher）によって37℃において5%CO_2で維持した。HAP1（Horizon Discovery C631）およびHAP1 UNG$^-$（Horizon Discovery HZGHC001531c012）は10%（v/v）ウシ胎児血清（FBS）を添加したイスコフ改変ダルベッコ培地プラスGlutaMax（Thermo Fisher Scientific）によって37℃において5%CO_2で維持した。

【０６３４】

トランスフェクション

HEK293T細胞を48ウェルコラーゲンコートBioCoatプレート（Corning）に接種し、およそ75%コンフルエンシーにおいてトランスフェクションした。簡潔には、製造者のプロトコールに従ってウェルあたり1.5μLのLipofectamine 2000（Thermo Fisher Scientific）を用いて、750ngのBEおよび250ngのsgRNA発現プラスミドをトランスフェクションした。

【０６３５】

製造者のプロトコールに従ってSE Cell Line 4D Nucleofector（商標）XキットS（Lonza）を用いて、HAP1およびHAP1 UNG$^-$細胞をヌクレオフェクションした。簡潔には、4D Nucleofector（商標）プログラムDZ-113を用いて、300ngのBEおよび100ngのsgRNA発現プラスミドによって4×10^5細胞をヌクレオフェクションした。

【０６３６】

ゲノムDNAサンプルの高スループットDNAシーケンシング

トランスフェクションされた細胞を3日後に収穫し、細胞を溶解緩衝液（10mM Tris-HCl pH8.0，0.05% SDS，25μg/mLプロテイナーゼK）中において37oCで1hr、次に80℃で30minに渡ってインキュベーションすることによって、ゲノムDNAを単離した。先に記載されている通り、フランキングHTSプライマーペアによるPCRによって、当該ゲノム領域を増幅した（6,1）。製造者の説明書に従って、かつ先に記載されている通り、Phusion高フィデリティDNAポリメラーゼ（Thermo Fisher）によってPCR増幅を実施した。シーケンシングアダプターを含有するプライマーによるPCRによって、精製されたDNAを

増幅した。生成物をゲル精製し、QuantiT（商標）PicoGreen dsDNAアッセイキット（Thermo Fisher）およびKAPAライブラリー定量キットIllumina（KAPA Biosystems）を用いて定量した。サンプルは先に記載されている通りIllumina MiSeqによってシーケンシングした。

【０６３７】

データ分析

シーケンシングリードはMiSeqレポーター（Illumina）を用いて自動デマルチプレックスし、個々のFASTQファイルを先に記載されている通りカスタムのMatlabスクリプトを用いて分析した（1）。Smith-Watermanアルゴリズムを用いて、各リードを適切な参照配列に対してペアワイズアライメントした。31よりも下のQスコアを有するベースコールはNによって置き換え、それゆえにヌクレオチド頻度を算出することからは除外した。この処置はおよそ1,000分の1という予想MiSeqベースコールエラー率を生む。リードおよび参照配列がギャップを含有しないアライメントされた配列をアライメント表に保存し、これから塩基頻度を各座位について集計し得た。

【０６３８】

先に記載されているMatlabスクリプトを用いてインデル頻度を定量した（5,6,1）。簡潔には、インデルが生起し得るウィンドウの両側をフランキングする2つの10bp配列との厳密なマッチについて、シーケンシングリードをスキャンした。厳密なマッチが見つからない場合には、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厳密にマッチする場合には、リードはインデルを含有しないと分類した。インデルウィンドウが参照配列よりも2塩基以上長いかまたは短い場合には、シーケンシングリードをそれぞれ挿入または欠失として分類した。

【０６３９】

編集ウィンドウ内の複数の標的シチジンにおける塩基編集成果間の相互依存性（連鎖不平衡）を評価するために、BEによって処置された細胞からの標的部位配列をカスタムのPythonスクリプトによって分析した（補足的な注1）。簡潔には、プロトスペーサーの各側をフランキングする2つの7bp配列に対する厳密なマッチについて、シーケンシングリードをスキャンした。介在する領域が厳密に20bpではない場合には、それはさらなる分析から除外した。最も多い非T編集成果を有する位置におけるヌクレオチドのアイデンティティ（一次標的C）に基づいて、プロトスペーサー配列を4つの群へとさらにフィルタリングした。これらの4つの群のそれぞれおよびプール全体について、プロトスペーサー内の20個の位置のそれぞれにおけるヌクレオチドの豊富さを計算した。

【０６４０】

例18：LbCpf1（ヌクレアーゼデッド、ヌクレアーゼ活性型、およびニッカーゼ）を含む塩基編集因子。

上で論じられている通り、本明細書において提供される融合蛋白質のいずれかの核酸によってプログラム可能なDNA結合蛋白質（napDNAbp）はLbCpf1蛋白質であり得る。いくつかの態様において、LbCpf1蛋白質はヌクレアーゼ不活性型、ヌクレアーゼ活性型、またはLbCpf1ニッカーゼである。LbCpf1の形態を含む融合蛋白質のいくつかの構築物を、C→T編集をなすそれらの能力について異なる標的配列において試験した。試験された構築物の模式図が図152に示されている。構築物10は[Apobec]-[LbCpf1]-[UGI]-[UGI]のドメイン配置を有し；構築物11は[Apobec]-[LbCpf1]-[UGI]のドメイン配置を有し；構築物12は[UGI]-[Apobec]-[LbCpf1]のドメイン配置を有し；構築物13は[Apobec]-[UGI]-[LbCpf1]のドメイン配置を有し；構築物14は[LbCpf1]-[UGI]-[Apobec]のドメイン配置を有し；構築物15は[LbCpf1]-[Apobec]-[UGI]のドメイン配置を有する。各構築物について、3つの異なるLbCpf1蛋白質を用いた（D/N/A。これはヌクレアーゼデッドLbCpf1（D）；LbCpf1ニッカーゼ（N）、およびヌクレアーゼ活性型LbCpf1（A）を言う）。これらの構築物のそれぞれについて、ドメインを連結するリンカーは下に示されており、XTENは配列SGSETPGTSESATPES（配列番号604）を有するXTENリンカーを言い、BPNLSは配列KRTADGSEFEPKKKRKV（配列番号740）を有する核局在配列を言う。構築物

はN末端（左）からC末端（右）へと示されている。

構築物10：Apobec-SGGSSGGSXTENSGGSSGGS-LbCpf1-SGGSGGSGGS-UGI-SGGSGGSGGS-UGI-SGGS-BPNLS

構築物11：Apobec-SGGSSGGSXTENSGGSSGGS-LbCpf1-SGGSGGSGGS-UGI-SGGS-BPNLS

構築物12：UGI-SGGSGGSGGS-Apobec-SGGSSGGSXTENSGGSSGGS-LbCpf1-SGGS-BPNLS

構築物13：Apobec-SGGSGGSGGS-UGI-SGGSSGGSXTENSGGSSGGS-LbCpf1-SGGS-BPNLS

構築物14：LbCpf1-SGGSGGSGGS-UGI-SGGSGGSGGS-Apobec-SGGS-BPNLS

構築物15：LbCpf1-SGGSGGSGGS-Apobec-SGGSGGSGGS-UGI-SGGS-BPNLS

【０６４１】

実験に用いられた共通のガイドバックボーン（sgRNA）はGTAATTTCTACTAAGTGTAGAT（配列番号741）[ガイド配列]TTTTTTTであり、配列番号741のTのそれぞれはウラシル（U）である。ここでは、構築物を特異的なヌクレオチド配列へと標的化するガイド配列は角括弧間に示されている。いくつかの態様において、本明細書において提供される構築物のいずれかは、GTAATTTCTACTAAGTGTAGAT（配列番号741）のバックボーン配列を含むsgRNAと複合体化させられ、配列番号741のTのそれぞれはウラシル（U）である。いくつかの態様において、本明細書において提供されるガイドRNAのいずれかは、ある配列、例えば標的DNA配列に対して完璧に相補的である14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、または30ヌクレオチドを含むガイド配列を含む。行われた実験において、試験されたガイド配列を下に示す：

EMX23：TACTTTGTCCTCCGGTTCTGGAA（配列番号742）

EMX20：TACTTTGTCCTCCGGTTCTG（配列番号743）

EMX19：TACTTTGTCCTCCGGTTCT（配列番号744）

EMX18：TACTTTGTCCTCCGGTTC（配列番号745）

EMX17：TACTTTGTCCTCCGGTT（配列番号746）

Hek2_23：CAGCCCGCTGGCCCTGTAAAGGA（配列番号747）

Hek2_20：CAGCCCGCTGGCCCTGTAAA（配列番号748）

Hek2_19：CAGCCCGCTGGCCCTGTAA（配列番号749）

Hek2_18：CAGCCCGCTGGCCCTGTA（配列番号750）

Hek2_17：CAGCCCGCTGGCCCTGT（配列番号751）

【０６４２】

種々の構築物および標的配列を用いるC→T塩基対編集パーセンテージを実証するデータは図153～159に示されており、実験の編集パーセンテージ値（インデルカウントに基づく調整後）およびインデルのパーセンテージは図160～166に示されている。

【０６４３】

補足的な配列

CDA1-BE3、AID-BE3、BE4、およびSaBE4融合蛋白質のアミノ酸配列。

【化１３５】

CDA1-**XTEN**-dCas9-**UGI**-*NLS* －次配列 (CDA1-BE3):

MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTE
RGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACK
LYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKR
*RSELSIMIQVKILHTTKSPAV***SGSETPGTSESATPES**DKKYSIGLAIGTNSVGWAVITDEYK

【化１３６】

VPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIF
SNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGV
DAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQL
SKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD
EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGT
EELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK
VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLF
EDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVEN
TQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSD
KNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN
NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKK
TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSK
KLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF
SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRY
TSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGS**TNLSDIIEKETGKQLVIQESILMLP
EEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIK
ML**SGGS*PKKKRKV*

（配列番号165）

【化１３７】

AID-**XTEN**-dCas9-**UGI**-*NLS*　　一次配列　(AID-BE3):

*MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFL
RYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPE
GLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVD
DLRDAFRTLGL***SGSETPGTSESATPES***DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKV
LGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD
DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLI
YLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSA
RLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDD
LDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTL
LKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN
REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLA
RGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY
EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE
CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER
LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY

【化１３８】

LYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNV
PSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITK
HVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAY
LNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFK
TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSK
ESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL
ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADAN
LDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGDSGGS**TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGN**
KPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGS*PKKK*
RKV

（配列番号１６６）

【化１３９】

rAPOBEC1-リンカー-dCas9-**UGI**-***UGI***-*NLS* 一次配列 (BE4):

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVE
VNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNR
QGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPC
*LNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK***SGGSSGGSSGSETPGTSESATPESS**
GGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF
DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE
KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFL
AAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH
QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI
TPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL
GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQL
KRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKA
QVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT
QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE
NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKL
ESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK
DWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE
AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY
EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGDSGGSGGSGGS**TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTA**
YDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGS***TNLSDIIEK***
ETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWAL
VIQDSNGENKIKMLSGGS*PKKKRK*

（配列番号１６７）

【化１４０】

rAPOBEC1-リンカー-SaCas9d-**UGI**-***UGI***-NLS　一次配列　(SaBE4):

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVE
VNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNR
QGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPC
*LNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK***SGGSSGGSSGSETPGTSESATPESS**
GGSSGGSGKRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGA
RRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLA
KRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFK
TSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYE
MLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQ
KKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIA
KILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQ
IAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIEL
AREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYS
LEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSK
ISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMN
LLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKE
WKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVD
KKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDP
QTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLD
ITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEA
KKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMND
KRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKGGSPKKKRKVSSDYKDHDG
DYKDHDIDYKDDDDKSGGSGGSGGS**TNLSDIIEKETGKQLVIQESILMLPEEVEEVIG**
NKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGG
SGGS***TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLL***
TSDAPEYKPWALVIQDSNGENKIKMLSGGS*PKKKRKV*

（配列番号１６８）

例１７の参考文献

【表２９】

1. A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533, 420-424 (2016).

2. G. T. Hess et al., Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. Nat Methods 13, 1036-1042 (2016).

3. K. Kim et al., Highly efficient RNA-guided base editing in mouse embryos. Nat Biotechnol 35, 435-437 (2017).

4. Y. B. Kim et al., Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. Nat Biotechnol 35, 371-376 (2017).

【表３０】

5. C. Kuscu et al., CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. Nat Meth advance online publication, (2017).

6. J. Li, Y. Sun, J. Du, Y. Zhao, L. Xia, Generation of Targeted Point Mutations in Rice by a Modified CRISPR/Cas9 System. Mol Plant 10, 526-529 (2017).

7. Y. Lu, J. K. Zhu, Precise Editing of a Target Base in the Rice Genome Using a Modified CRISPR/Cas9 System. Mol Plant 10, 523-525 (2017).

8. Y. Ma et al., Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. Nat Methods 13, 1029-1035 (2016).

9. K. Nishida et al., Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science 353, (2016).

10. H. A. Rees et al., Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. Nat Commun 8, 15790 (2017).

11. L. Yang et al., Engineering and optimising deaminase fusions for genome editing. Nat Commun 7, 13330 (2016).

12. Y. Zong et al., Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. Nat Biotechnol 35, 438-440 (2017).

13. Z. Shimatani et al., Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. Nat Biotech 35, 441-443 (2017).

14. A. C. Komor, A. H. Badran, D. R. Liu, CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. Cell 168, 20-36 (2017).

15. A. J. Davis, D. J. Chen, DNA double strand break repair via non-homologous end-joining. Translational cancer research 2, 130-143 (2013).

16. M. M. Vilenchik, A. G. Knudson, Endogenous DNA double-strand breaks: Production, fidelity of repair, and induction of cancer. Proceedings of the National Academy of Sciences 100, 12871-12876 (2003).

17. F. Liang, M. Han, P. J. Romanienko, M. Jasin, Homology-directed repair is a major doublestrand break repair pathway in mammalian cells. Proceedings of the National Academy of Sciences 95, 5172-5177 (1998).

18. Y. Miyaoka et al., Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. Scientific Reports 6, 23549 (2016).

【表３１】

19. M. M. Jore et al., Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat Struct Mol Biol 18, 529-536 (2011).

20. L. H. Pearl, Structure and function in the uracil-DNA glycosylase superfamily. Mutation Research/DNA Repair 460, 165-181 (2000).

21. W. Tang, J. H. Hu, D. R. Liu, Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation. 8, 15939 (2017).

22. G. Saraconi, F. Severi, C. Sala, G. Mattiuz, S. G. Conticello, The RNA editing enzyme APOBEC1 induces somatic mutations and a compatible mutational signature is present in esophageal adenocarcinomas. Genome Biology 15, 417 (2014).

23. R. M. Kohli et al., Local Sequence Targeting in the AID/APOBEC Family Differentially Impacts Retroviral Restriction and Antibody Diversification. Journal of Biological Chemistry 285, 40956-40964 (2010).

24. L. Chelico, P. Pham, M. F. Goodman, Stochastic properties of processive cytidine DNA deaminases AID and APOBEC3G. Philosophical Transactions of the Royal Society B: Biological Sciences 364, 583-593 (2009).

25. E. A. Kouzminova, A. Kuzminov, Patterns of chromosomal fragmentation due to uracil-DNA incorporation reveal a novel mechanism of replication-dependent double-stranded breaks. Molecular Microbiology 68, 202-215 (2008).

26. F. A. Ran et al., In vivo genome editing using Staphylococcus aureus Cas9. Nature 520, 186-191 (2015).

27. F. d. A. di Fagagna, G. R. Weller, A. J. Doherty, S. P. Jackson, The Gam protein of bacteriophage Mu is an orthologue of eukaryotic Ku. EMBO Reports 4, 47-52 (2003).

28. C. Shee et al., Engineered proteins detect spontaneous DNA breakage in human and bacterial cells. eLife 2, e01222 (2013).

【０６４４】

等価物および範囲

当業者は、本明細書に記載される態様の多くの等価物を認識するか、または普通に過ぎない実験手順を用いて確かめる能力があるであろう。本開示の範囲は上の記載に限定されることを意図されず、むしろ添付の請求項において規定される通りである。

【０６４５】

反対に指示されないかまたはコンテキストから別様に明白でない限り、「a」、「an」、および「the」などの冠詞は1つまたは1つよりも多くを意味し得る。反対に指示されないかまたはコンテキストから別様に明白でない限り、「または」を群の2つ以上のメンバー間に包含する請求項または記載は、群のメンバーの1つ、1つよりも多く、または全てが存在する場合には満足していると考えられる。「または」を2つ以上の群メンバー間に包含する群の開示は、群の厳密に1つのメンバーが存在する態様、群の1つよりも多くのメンバーが存在する態様、および群のメンバーの全てが存在する態様を提供する。簡潔の目的のために、それらの態様は本明細書に個々に明記しなかったが、これらの態様のそれぞれは本明細書によって提供され、具体的に請求または放棄され得ることは理解されるであろう。

【０６４６】
本発明が、請求項の1つ以上からのまたは明細書の1つ以上の該当する部分からの1つ以上の限定、要素、節、または記述用語が別の請求項に導入される全ての変形、組み合わせ、および順列を包摂することは理解されるであろう。例えば、別の請求項に従属するある請求項は、同じ基本請求項に従属するいずれかの他の請求項に見出される限定の1つ以上を包含するように改変され得る。さらにその上、請求項が組成物を記載しているところでは、別様に指示されないかまたは矛盾もしくは不整合が生ずるであろうことが当業者にとって明白ではない限り、もしあれば、本明細書において開示される作るかもしくは用いる方法のいずれかに従ってまたは当分野において公知の方法に従って組成物を作るかまたは用いる方法が包含されることは理解されるであろう。

【０６４７】
要素がリストとして、例えばマーカッシュ群フォーマットで提示されているところでは、要素のあらゆる可能な部分群もまた開示されていること、およびいずれかの要素または要素の部分群が群から取り除かれ得ることは理解されるであろう。用語「含む」は開放的であることを意図され、追加の要素またはステップの包含を許可することもまた注意される。一般的に、ある態様、生成物、または方法が特定の要素、特質、またはステップを含むと言われるところでは、かかる要素、特質、またはステップからなるかまたは本質的になる態様、生成物、または方法もまた提供されていることは理解されるべきである。簡潔の目的のために、それらの態様は本明細書において個々に明記しなかったが、これらの態様のそれぞれは本明細書によって提供され、具体的に請求または放棄され得ることは理解されるであろう。

【０６４８】
範囲が与えられているところでは、エンドポイントは包含される。さらにその上、別様に指示されないかまたはコンテキストおよび／もしくは当業者の理解から別様に明白でない限り、範囲として表現されている値は、コンテキストが明瞭に別様に述べていない限り、範囲の下限の単位の十分の一まで、いくつかの態様において記されている範囲内のいずれかの具体的な値をとり得ることは理解されるであろう。簡潔の目的のために、各範囲内の値は本明細書において個々に明記しなかったが、これらの値のそれぞれは本明細書によって提供され、具体的に請求または放棄され得ることは理解されるであろう。別様に指示されていないかまたはコンテキストおよび／もしくは当業者の理解から別様に明白でない限り、範囲として表現されている値は所与の範囲内のいずれかの部分範囲をとり得、部分範囲のエンドポイントは範囲の下限の単位の十分の一と同じ程度の正確度で表現されることもまた理解されるであろう。加えて、本発明のいずれかの特定の態様は請求項のいずれか1つ以上から明確に除外され得ることは理解されるべきである。範囲が与えられているところでは、範囲内のいずれかの値は請求項のいずれか1つ以上から明確に除外され得る。本発明の組成物および／または方法のいずれかの態様、要素、特質、用途、または側面はいずれかの1つ以上の請求項から除外され得る。簡潔の目的のために、1つ以上の要素、特質、目的、または側面が除外されている態様の全てが本明細書において明確に規定されているわけではない。

【図面】

【図１】

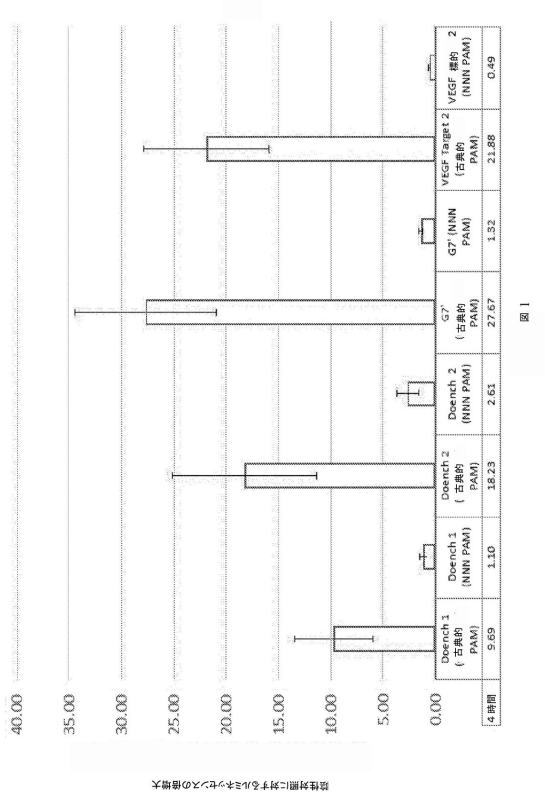


図 1

【図２】

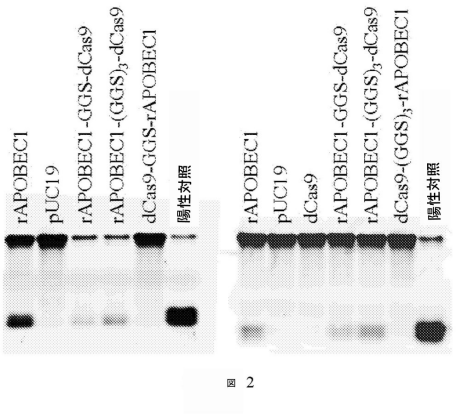


図 2

【図３】

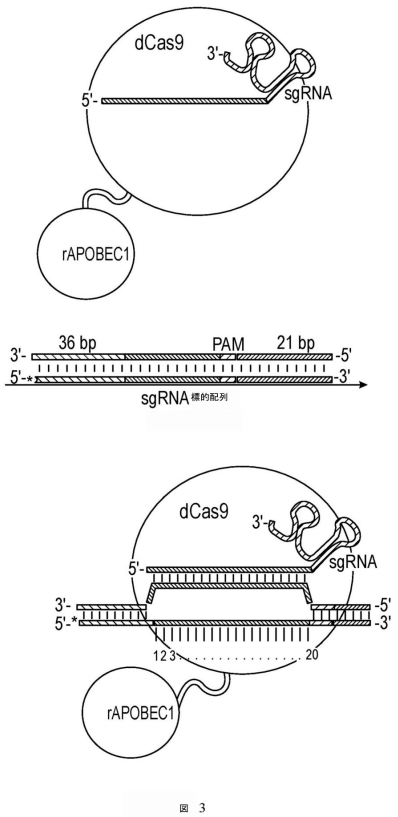


図 3

【図４】

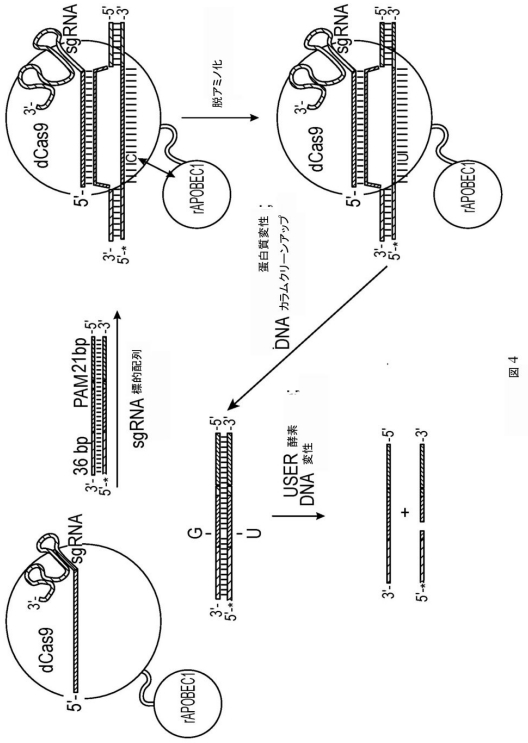


図 4

【図５】

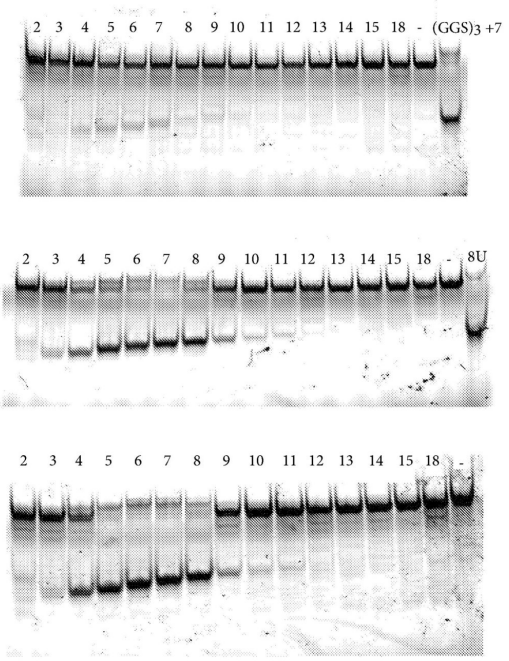


図 5

【図６】

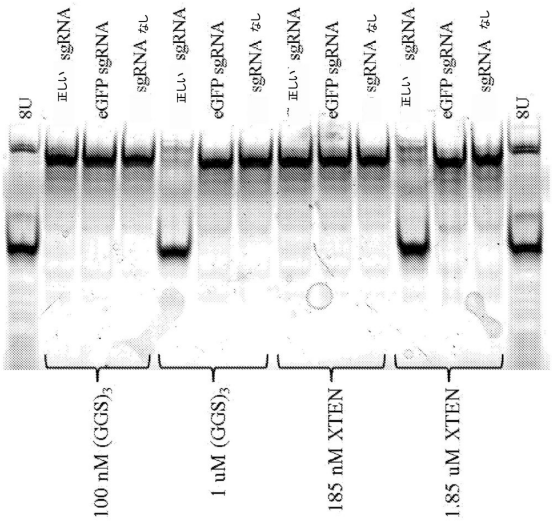


図 6

【図７】

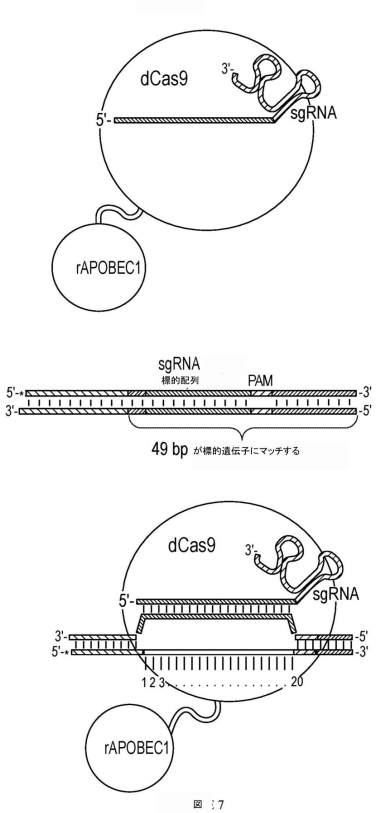


図 7

【図８】

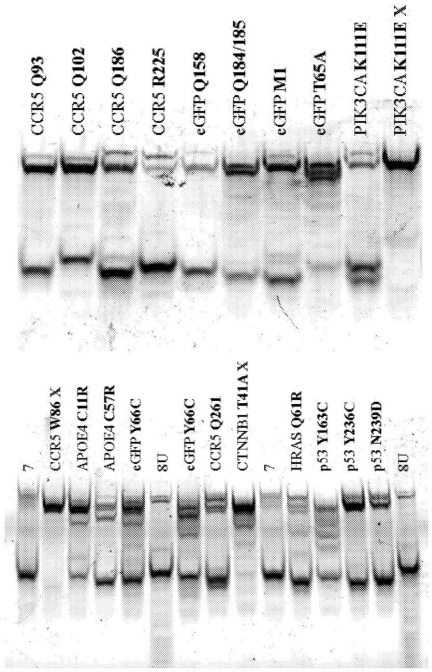


図 8

【図９】

EMX1		C_5	C_6	C_{10}
0 uM XTEN	A	0.1%	0.1%	0.1%
	C	99.8%	99.8%	99.8%
	G	0.0%	0.0%	0.1%
	T	0.0%	0.1%	0.0%
1.85 uM XTEN	A	0.1%	0.0%	0.1%
	C	60.4%	61.0%	99.1%
	G	0.0%	0.0%	0.1%
	T	39.5%	39.0%	0.7%

FANCF		C_6	C_7	C_8	C_{11}
0 uM XTEN	A	0.1%	0.1%	0.1%	0.1%
	C	99.8%	99.8%	99.9%	99.9%
	G	0.0%	0.1%	0.0%	0.0%
	T	0.0%	0.0%	0.0%	0.0%
1.85 uM XTEN	A	0.1%	0.1%	0.1%	0.1%
	C	63.9%	64.7%	65.0%	72.6%
	G	0.0%	0.0%	0.0%	0.0%
	T	36.0%	35.1%	34.9%	27.3%

HEK293部位 2		C_4	C_6	C_{11}
0 uM XTEN	A	0.1%	0.1%	0.1%
	C	99.9%	99.9%	99.9%
	G	0.0%	0.0%	0.0%
	T	0.0%	0.0%	0.1%
1.85 uM XTEN	A	0.1%	0.1%	0.1%
	C	80.6%	76.9%	99.6%
	G	0.0%	0.0%	0.0%
	T	19.3%	22.9%	0.3%

HEK293部位 3		C_3	C_4	C_5	C_9
0 uM XTEN	A	0.1%	0.1%	0.0%	0.1%
	C	99.8%	99.9%	99.9%	99.9%
	G	0.0%	0.0%	0.0%	0.0%
	T	0.1%	0.0%	0.0%	0.0%
1.85 uM XTEN	A	0.1%	0.1%	0.0%	0.1%
	C	92.2%	74.8%	71.5%	96.6%
	G	0.0%	0.0%	0.0%	0.0%
	T	7.7%	25.1%	28.5%	3.3%

HEK293部位 4		C_3	C_5	C_8	C_{11}
0 uM XTEN	A	0.1%	0.0%	0.1%	0.0%
	C	99.8%	99.9%	99.8%	99.9%
	G	0.0%	0.0%	0.0%	0.0%
	T	0.0%	0.0%	0.1%	0.0%
1.85 uM XTEN	A	0.1%	0.1%	0.1%	0.1%
	C	98.8%	60.1%	97.0%	99.4%
	G	0.0%	0.0%	0.0%	0.0%
	T	1.1%	39.8%	2.9%	0.5%

RNF2		C_3	C_6
0 uM XTEN	A	0.1%	0.0%
	C	99.9%	99.9%
	G	0.0%	0.0%
	T	0.0%	0.0%
1.85 uM XTEN	A	0.1%	0.0%
	C	59.1%	57.8%
	G	0.0%	0.0%
	T	40.8%	42.1%

図 9

【図１０】

EMX1		C_5	C_6	C_{10}
非処置	A	0.0%	0.0%	0.0%
	C	99.5%	99.7%	100.0%
	G	0.0%	0.1%	0.0%
	T	0.5%	0.2%	0.0%
XTEN	A	0.7%	0.5%	0.0%
	C	93.5%	95.8%	100.0%
	G	2.1%	0.3%	0.0%
	T	3.6%	3.3%	0.0%
XTEN-UGI	A	0.2%	0.0%	0.0%
	C	81.8%	82.5%	100.0%
	G	0.6%	0.3%	0.0%
	T	17.4%	17.1%	0.0%

FANCF		C_6	C_7	C_8	C_{11}
非処置	A	0.0%	0.0%	0.2%	0.1%
	C	99.9%	99.8%	99.8%	99.9%
	G	0.0%	0.0%	0.0%	0.0%
	T	0.1%	0.1%	0.0%	0.0%
XTEN	A	0.3%	0.1%	0.0%	0.0%
	C	98.1%	99.2%	99.0%	99.8%
	G	0.4%	0.0%	0.0%	0.0%
	T	1.2%	0.7%	1.0%	0.2%
XTEN-UGI	A	0.0%	0.0%	0.1%	0.0%
	C	93.2%	93.5%	93.4%	98.2%
	G	0.0%	0.0%	0.0%	0.0%
	T	6.7%	6.5%	6.5%	1.8%

HEK293部位2		C_4	C_6	C_{11}
非処置	A	0.3%	0.2%	0.2%
	C	99.7%	99.7%	99.7%
	G	0.0%	0.0%	0.0%
	T	0.0%	0.0%	0.0%
XTEN	A	0.3%	0.3%	0.3%
	C	99.7%	99.4%	99.7%
	G	0.0%	0.3%	0.0%
	T	0.0%	0.0%	0.0%
XTEN-UGI	A	0.3%	0.2%	0.2%
	C	98.8%	98.2%	99.8%
	G	0.0%	0.3%	0.0%
	T	0.9%	1.3%	0.0%

HEK293部位2		C_3	C_4	C_5	C_9
非処置	A	0.0%	0.0%	0.0%	0.0%
	C	100.0%	100.0%	100.0%	99.9%
	G	0.0%	0.0%	0.0%	0.0%
	T	0.0%	0.0%	0.0%	0.1%
XTEN	A	0.0%	0.6%	0.3%	0.1%
	C	100.0%	95.8%	95.8%	99.2%
	G	0.0%	0.2%	0.7%	0.4%
	T	0.0%	3.4%	3.2%	0.3%
XTEN-UGI	A	0.0%	0.3%	0.3%	0.0%
	C	96.8%	83.0%	79.2%	98.5%
	G	0.0%	0.0%	1.1%	0.2%
	T	3.2%	16.8%	19.4%	1.3%

HEK293部位4		C_3	C_5	C_8	C_{11}
非処置	A	0.0%	0.4%	0.0%	0.0%
	C	99.8%	97.6%	99.9%	100.0%
	G	0.0%	1.0%	0.0%	0.0%
	T	0.2%	1.0%	0.0%	0.0%
XTEN	A	0.0%	1.1%	0.0%	0.0%
	C	99.6%	92.2%	99.9%	100.0%
	G	0.0%	2.2%	0.0%	0.0%
	T	0.4%	4.5%	0.0%	0.0%
XTEN-UGI	A	0.0%	0.5%	0.0%	0.0%
	C	99.4%	86.7%	99.1%	100.0%
	G	0.0%	1.8%	0.0%	0.0%
	T	0.6%	11.0%	0.9%	0.0%

RNF2		C_3	C_6
非処置	A	0.0%	0.0%
	C	99.9%	99.5%
	G	0.0%	0.2%
	T	0.0%	0.3%
XTEN	A	0.0%	0.0%
	C	99.8%	99.3%
	G	0.0%	0.2%
	T	0.2%	0.5%
XTEN-UGI	A	0.0%	0.0%
	C	99.6%	99.1%
	G	0.0%	0.4%
	T	0.4%	0.5%

図 10

【図１１Ａ】

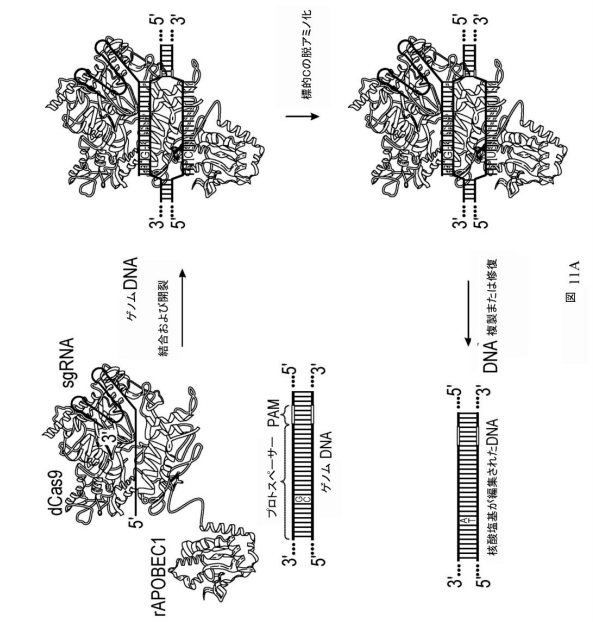


図 11A

【図１１Ｂ－Ｃ】

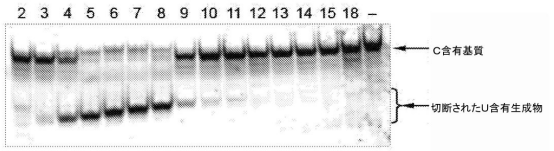


図 11B

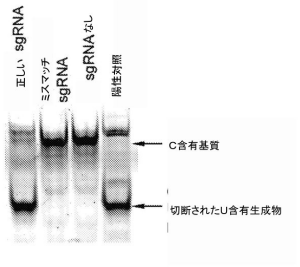


図 11C

【図１２Ａ】

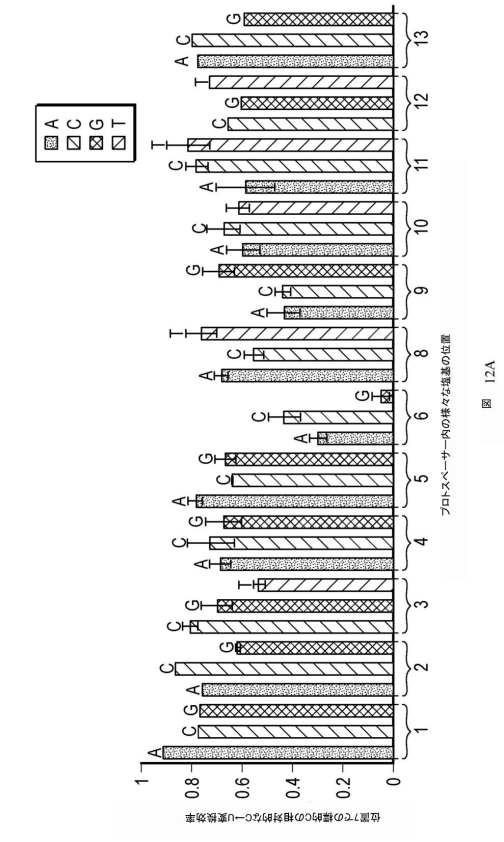


図 12A

【図１２Ｂ】

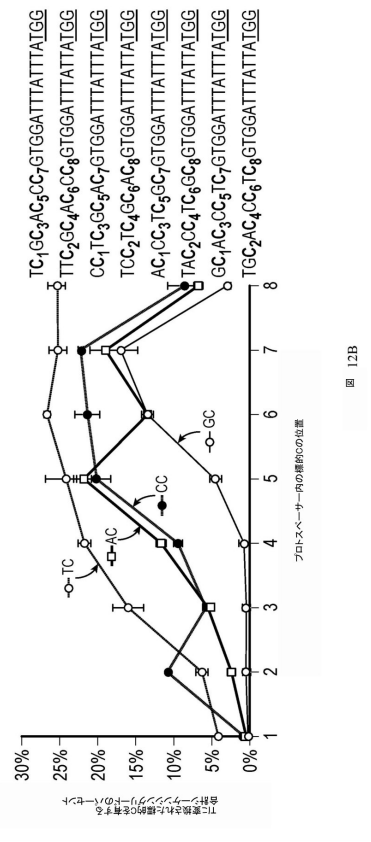


図 12B

【図１３Ａ】

EMX1:	GAGTC$_5$C$_6$GAGCAGAAGAAGAAGGGG
FANCF:	GGAATC$_6$C$_7$C$_8$TTC$_{11}$TGCAGCACCTGG
HEK293 部位 2:	GAAC$_4$AC$_6$AAAGCATAGACTGCGGG
HEK293 部位 3:	GGCC$_4$C$_5$AGACTGAGCACGTGATGG
HEK293 部位 4:	GGCAC$_5$TGCGGCTGGAGGTCCGGG
RNF2:	GTC$_3$ATC$_6$TTAGTC$_{12}$ATTACCTGAGG

図 13A

【図１３Ｂ】

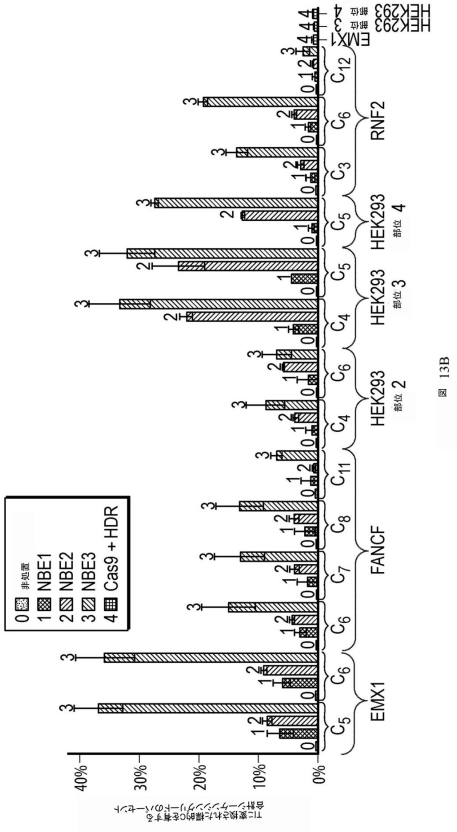


図 13B

【図１３Ｃ】

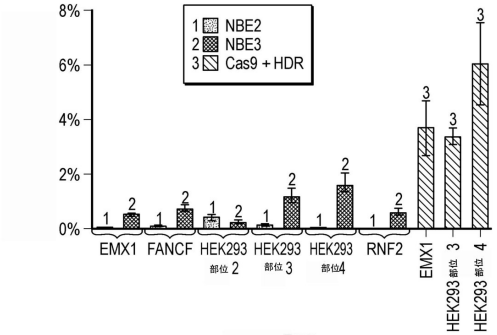


図13C

【図１４】

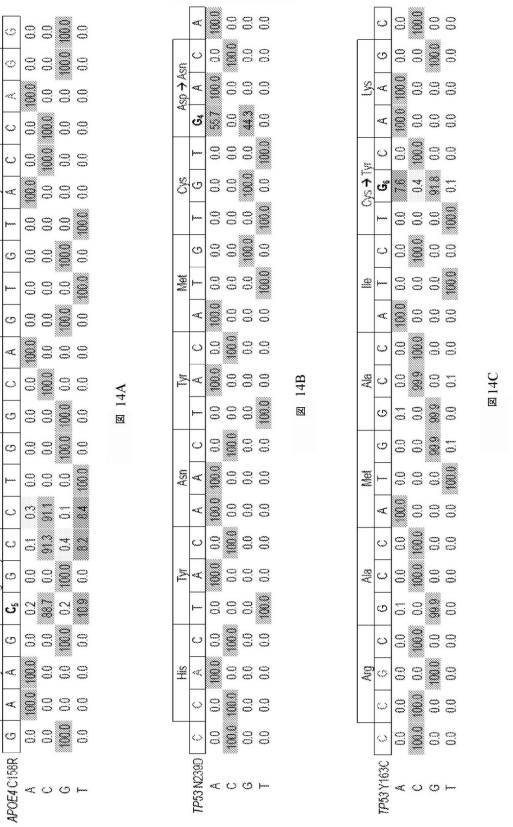


図 14A　図 14B　図 14C

【図１５】

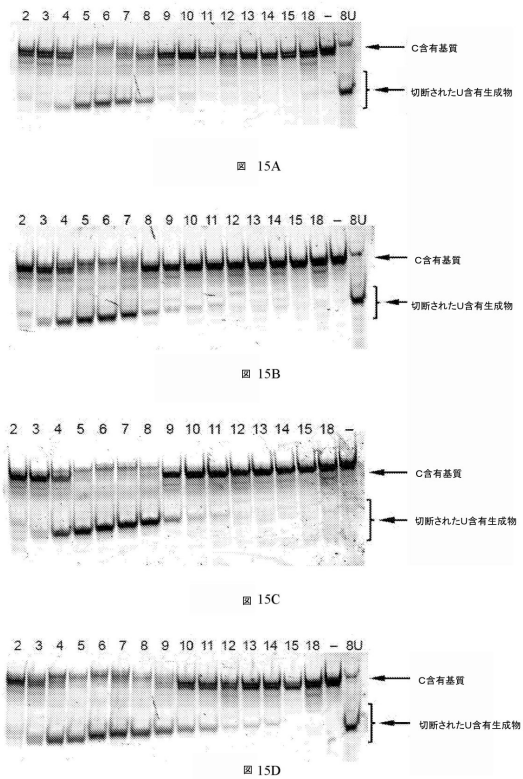


図 15A　図 15B　図 15C　図 15D

【図１６】

APOE4 Cys112Arg: 5'-GGAGGACGTGC$_{11}$GCGGCCGCCTGG
APOE4 Cys158Arg: 5'-GAAGC$_{8}$GCCTGGCAGTGTACCAGG
CTNNB1 Thr41Ala: 5'-CTGTGGC$_{7}$AGTGGCACCAGAATGG
HRAS Gln61Arg: 5'-CCTCCC$_{8}$GGCCGGCGGTATCCAGG
p53 Tyr163Cys: 5'-GCTTGC$_{8}$AGATGGCCATGGCGCGG
p53 Tyr236Cys: 5'-ACACATGC$_{8}$AGTTGTAGTGGATGG
p53 Asn239Asp: 5'-TGTC$_{4}$ACACATGTAGTTGTAGTGG

図 16A

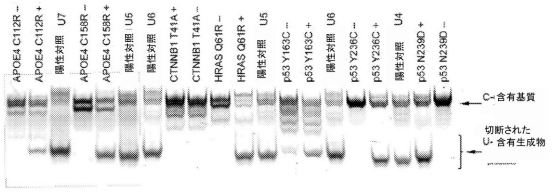


図 16B

【図１７】

プロトスペーサーおよびPAM配列 : 5'-TTCCCCCCCCGATTTATTTATGG-3'

配列	合計リードの%
CCCCCCCC	62.4
TTTTTTCC	18.2
TTTTTTTC	13.4
TTTTTTTT	3.3
TCCCCCCC	0.8
CCCCTTCC	0.3
CCCTTTCC	0.3
TTTTTCCC	0.3
CCCCTCCC	0.3

図 17

【図１８Ａ－Ｄ】

EMX1: GAGTC$_5$C$_6$GAGCAGAAGAAGAAGGG
FANCF: GGAATC$_6$C$_7$C$_8$TTC$_{11}$TGCAGCACCTGG
HEK293 部位 2: GAAC$_4$AC$_6$AAAGCATAGACTGCGGG
HEK293 部位 3: GGCC$_4$C$_5$AGACTGAGCACGTGATGG
HEK293 部位 4: GGCAC$_5$TGCGGCTGGAGGTCCGGG
RNF2: GTC$_3$ATC$_6$TTAGTCATTACCTGAGG

図 18A

EMX1	C$_5$	C$_6$
NBE1	6.2%	6.5%
NBE1 + UGI	9.7%	10.1%
NBE2	8.0%	8.7%

図 18B

FANCF	C$_6$	C$_7$	C$_8$	C$_{10}$
NBE1	3.7%	3.2%	3.4%	2.4%
NBE1 + UGI	7.5%	7.6%	7.5%	1.6%
NBE2	4.7%	4.6%	4.6%	0.8%

図 18C

HEK293 部位 2	C$_4$	C$_6$
NBE1	0.4%	0.4%
NBE1 + UGI	1.6%	2.6%
NBE2	3.4%	5.9%

図 18D

【図１８Ｅ－Ｈ】

HEK293 部位 3	C$_4$	C$_5$
NBE1	2.0%	1.9%
NBE1 + UGI	6.5%	6.7%
NBE2	10.0%	12.5%

図 18E

HEK293 部位 4	C$_5$
NBE1	1.4%
NBE1 + UGI	5.4%
NBE2	8.2%

図 18F

RNF2 部位	C$_3$	C$_6$
NBE1	0.7%	1.4%
NBE1 + UGI	3.4%	3.9%
NBE2	2.5%	3.7%

図 18G

非プロトスペーサー	Cs	C	T
非処置		99.93%	0.03%
NBE1		99.95%	0.03%
NBE1 + UGI		99.91%	0.06%
NBE2		99.92%	0.04%

図 18H

【図１９】

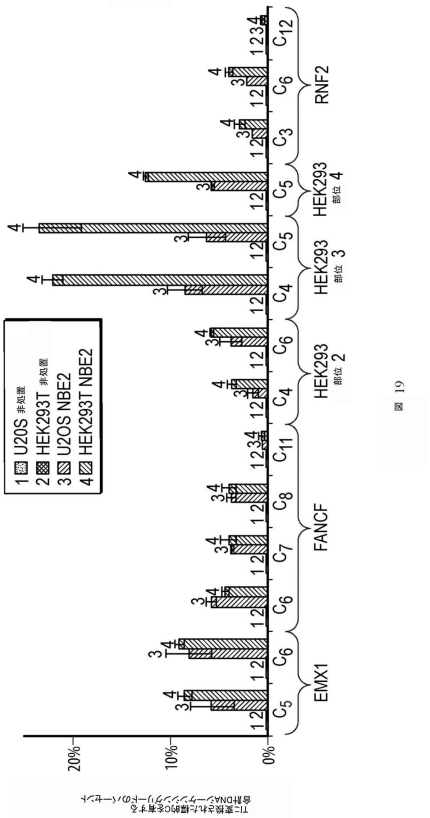


図 19

【図２０】

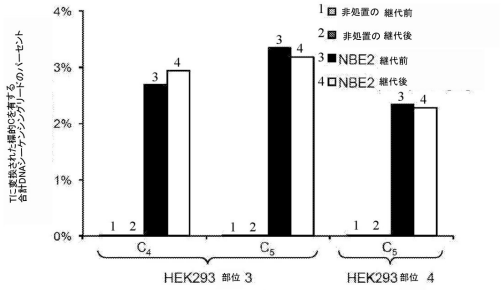


図 20

【図２１】

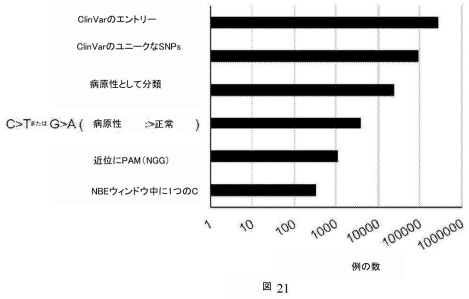


図 21

【図２２－１】

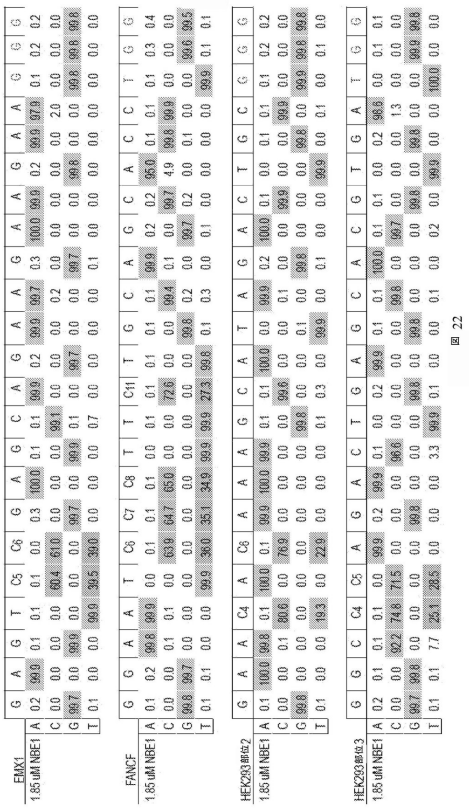


図 22

【図２２－２】

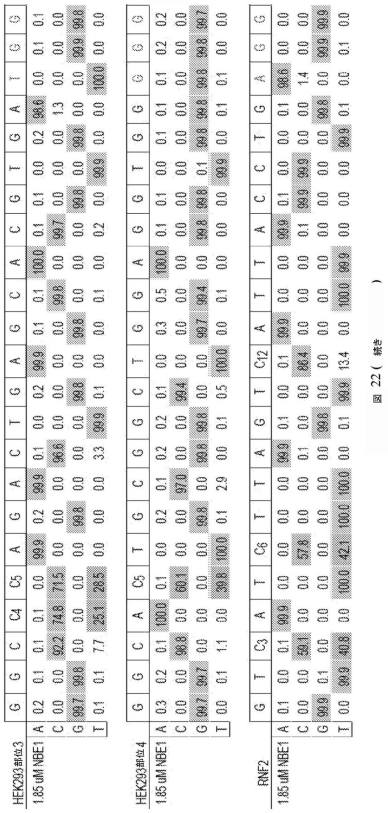


図 22（続き）

【図２３－１】

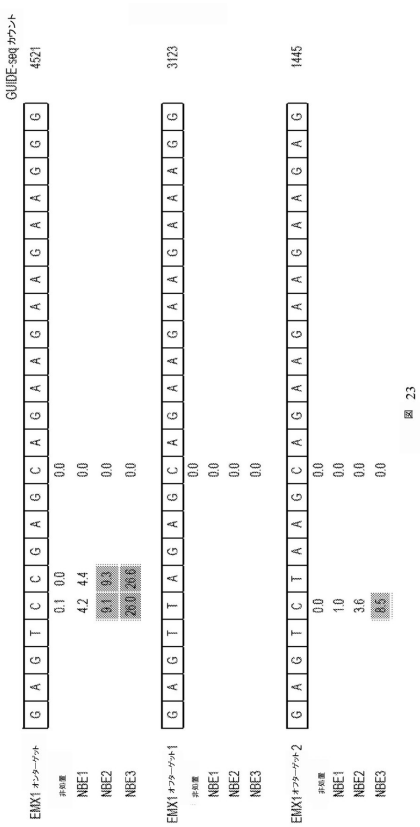


図 23

【図２３－２】

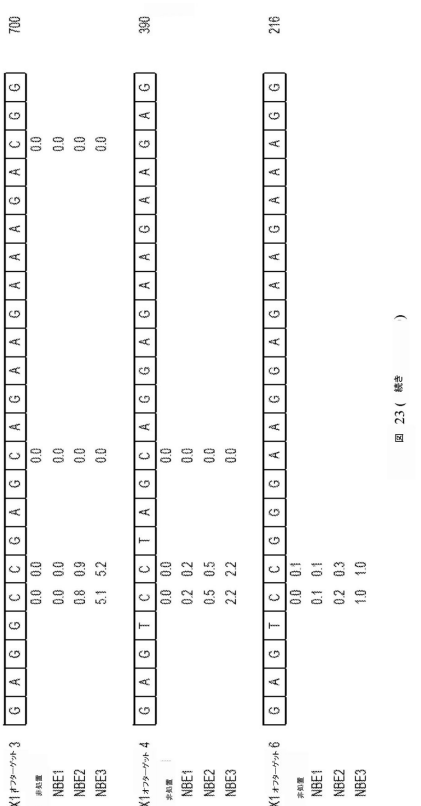


図 23（続き）

【図２３－３】

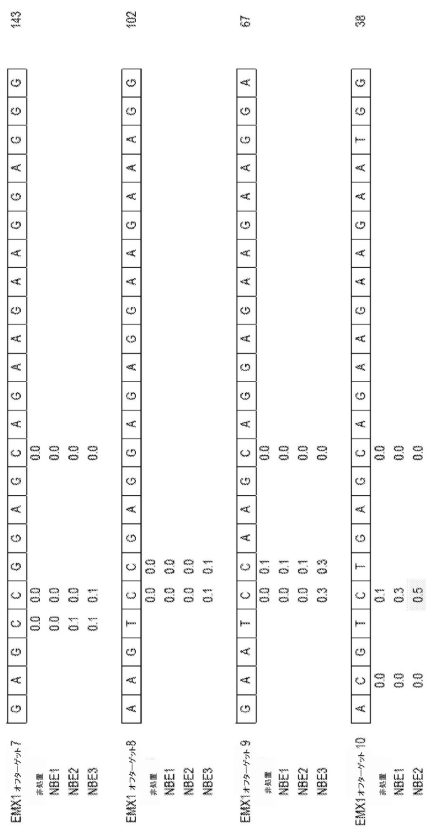


図 23（続き）

【図２４－１】

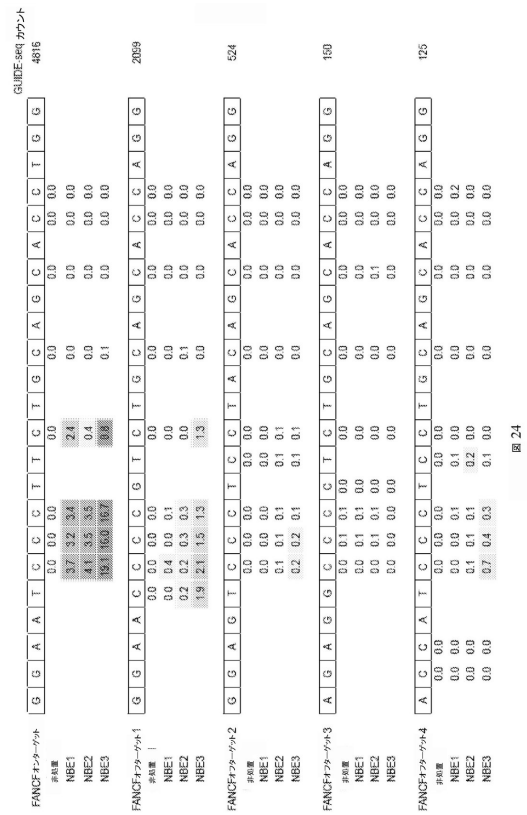


図 24

【図２４－２】

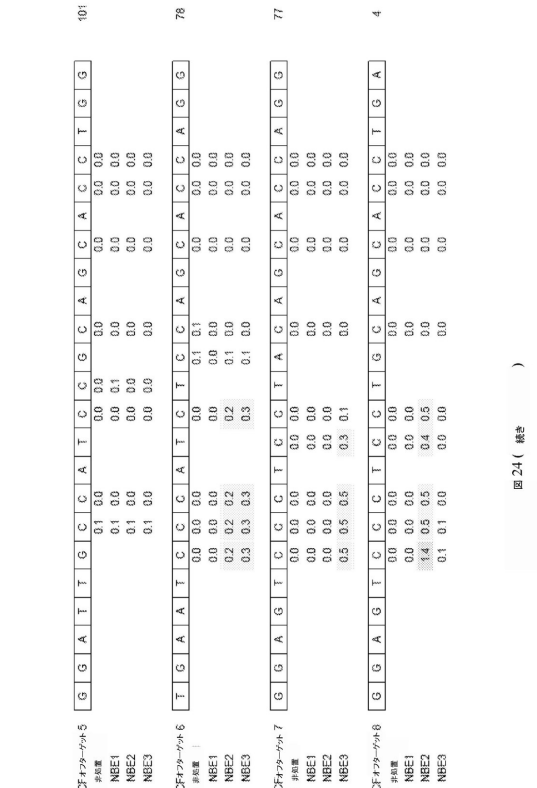


図 24（続き）

【図２５】

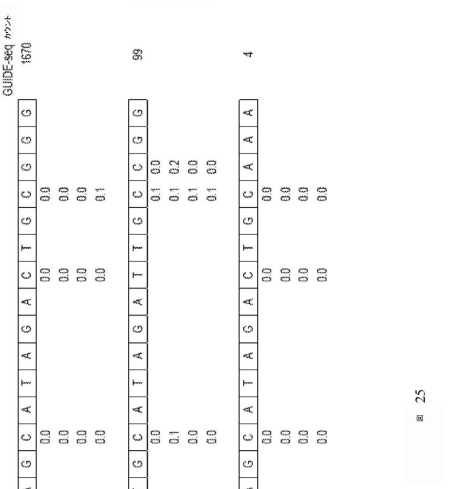


図 25

【図２６】

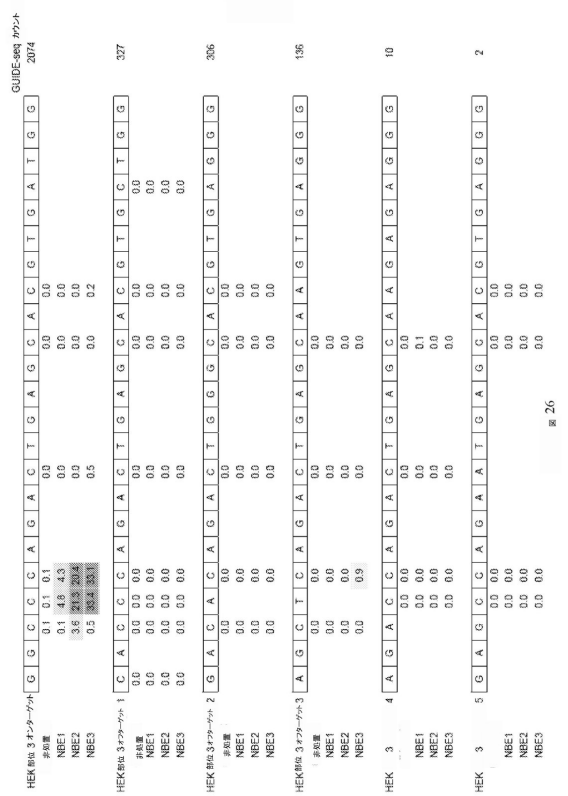


図 26

【図２７－１】

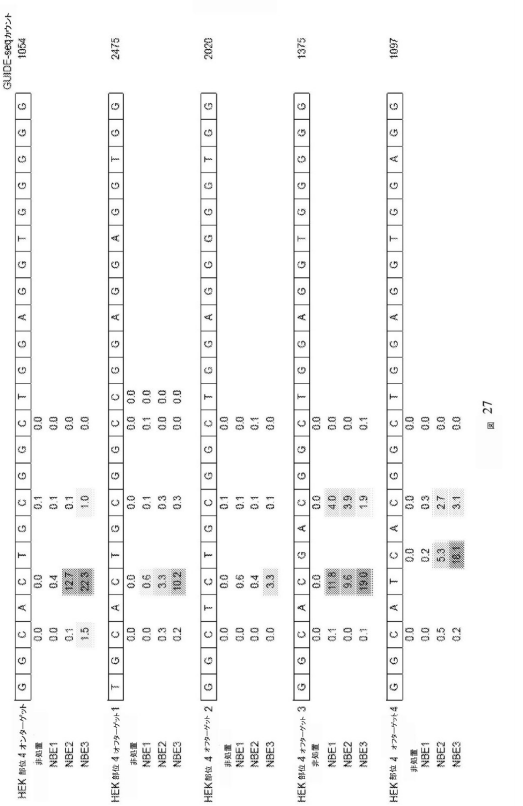


図 27

【図２７－２】

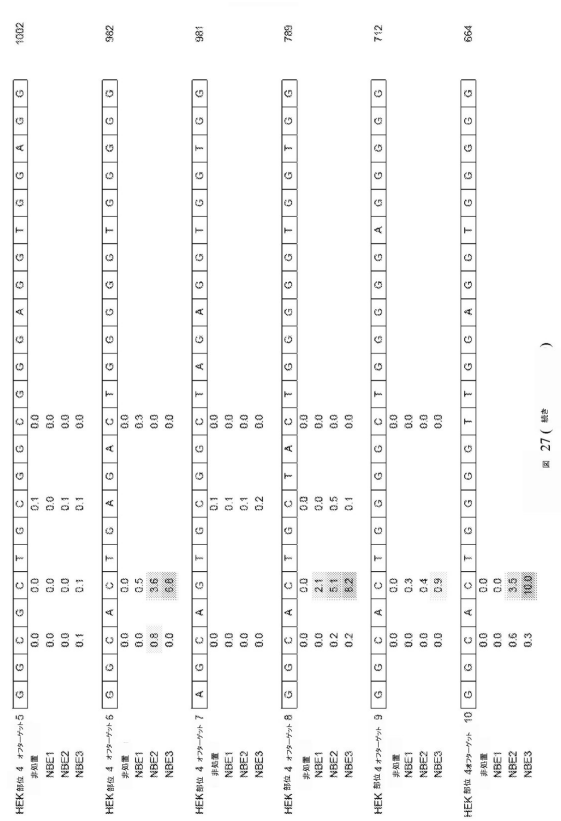


図 27（続き）

【図２８】

非プロトスペーサー	Cs	C (%)	T (%)
非処置		99.94	0.04
NBE1		99.92	0.05
NBE2		99.92	0.05
NBE3		99.94	0.03

図 28

【図２９Ａ】

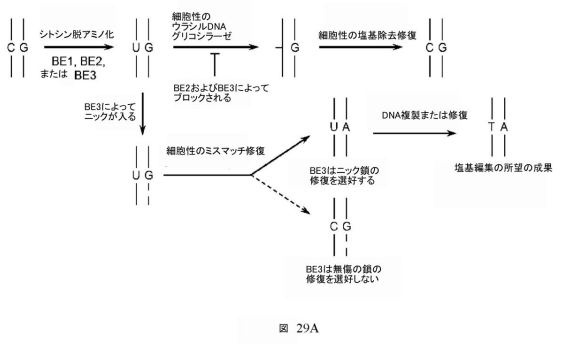


図 29A

【図２９Ｂ】

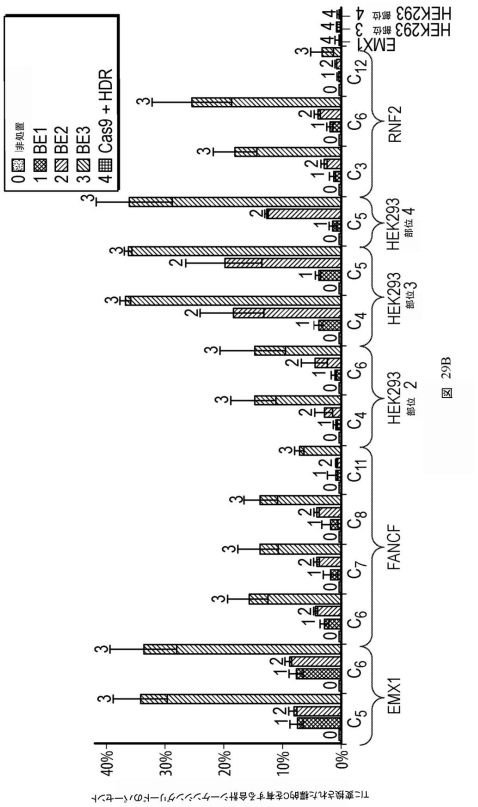


図 29B

【図２９Ｃ】

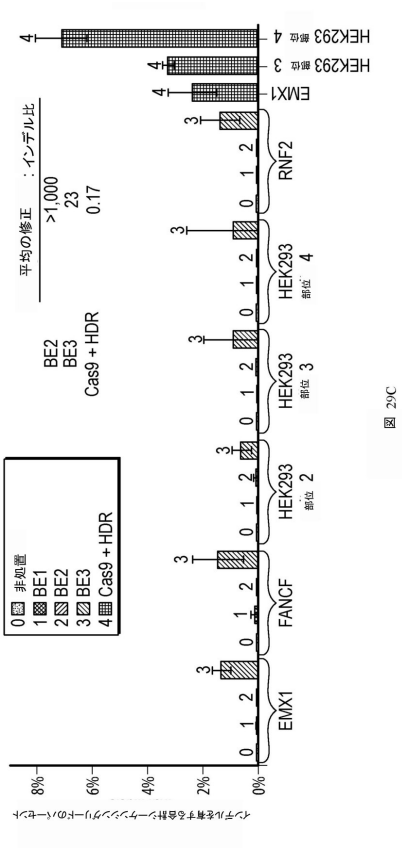


図 29C

【図３０Ａ】

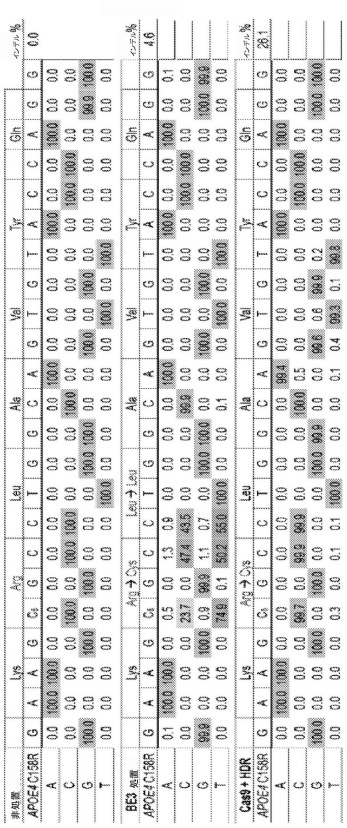


図 30A

【図３０Ｂ】

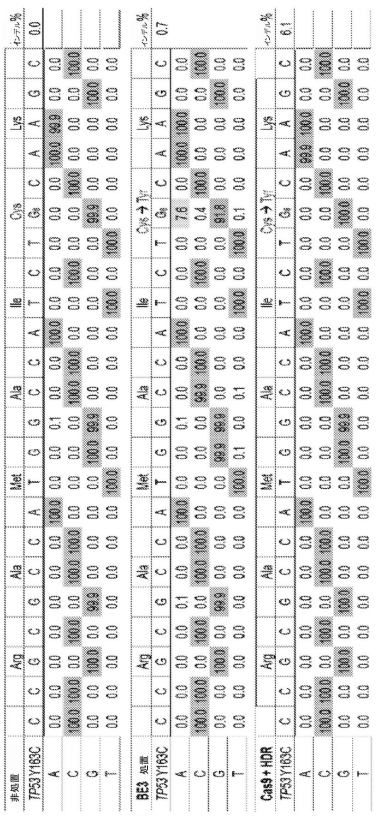


図 30B

【図３１－１】

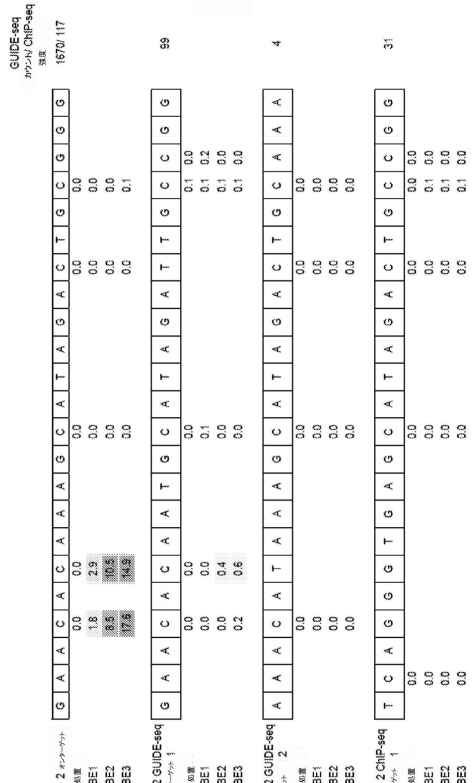


図 31

【図３１－２】

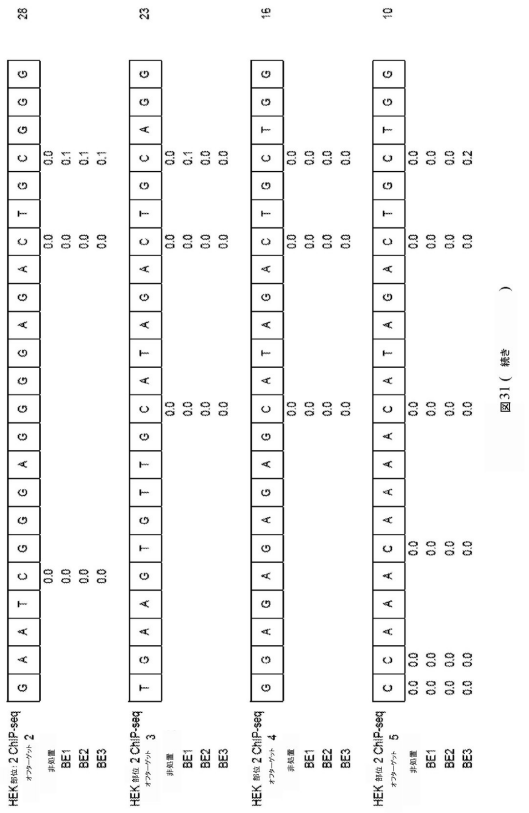


図31（続き）

【図３２－１】

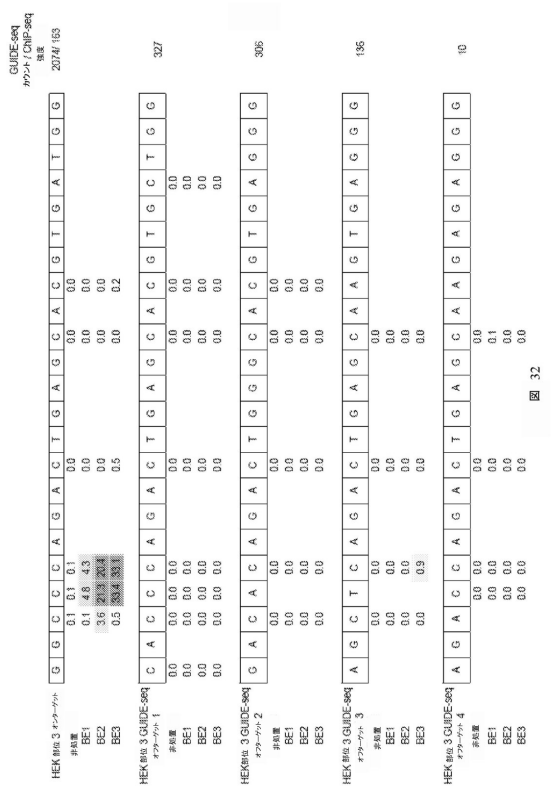


図32

【図３２－２】

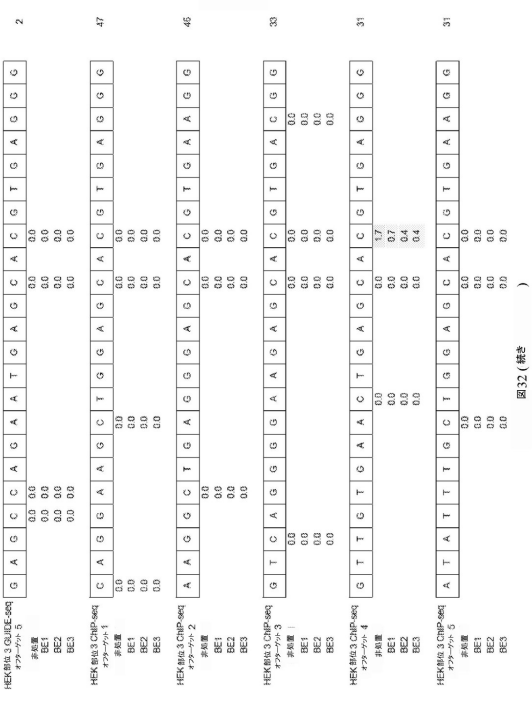


図32（続き）

【図３３－１】

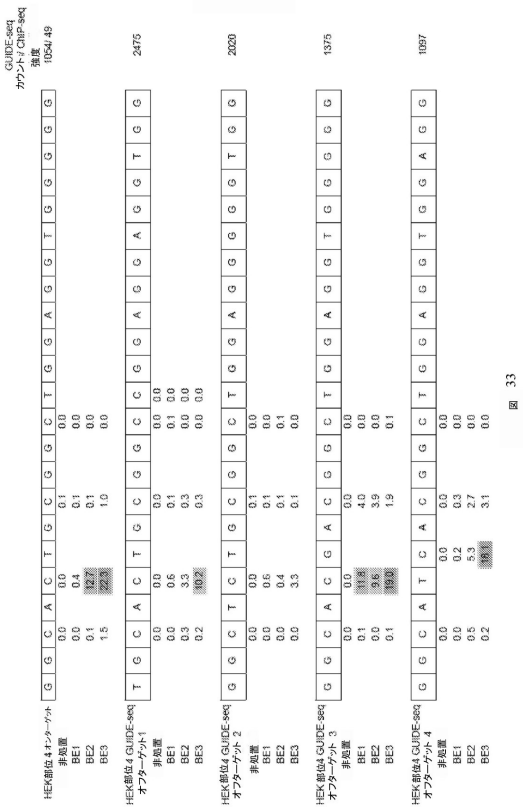
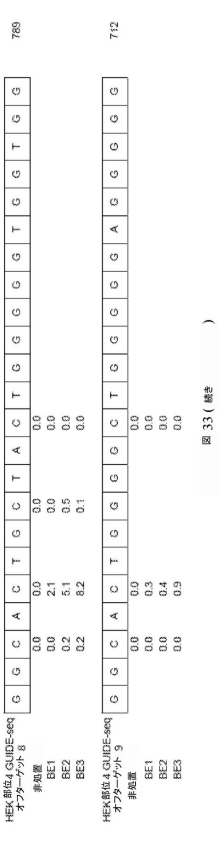
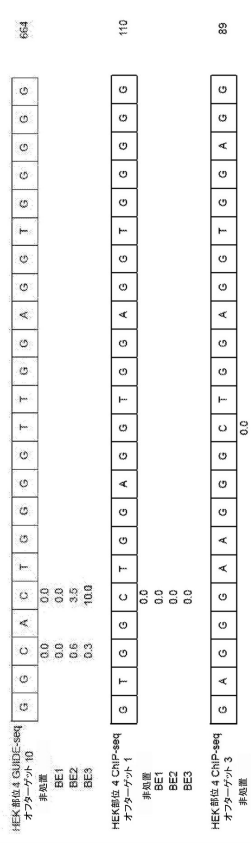


図33

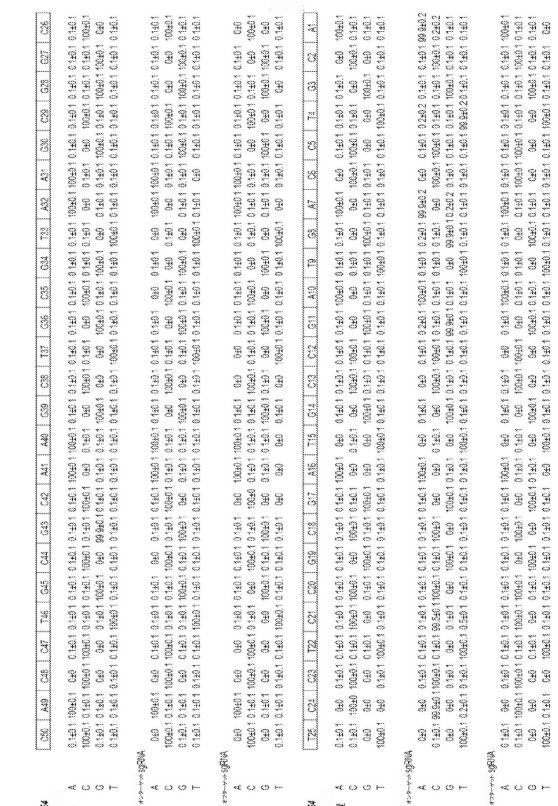
【図３３－２】



【図３３－３】

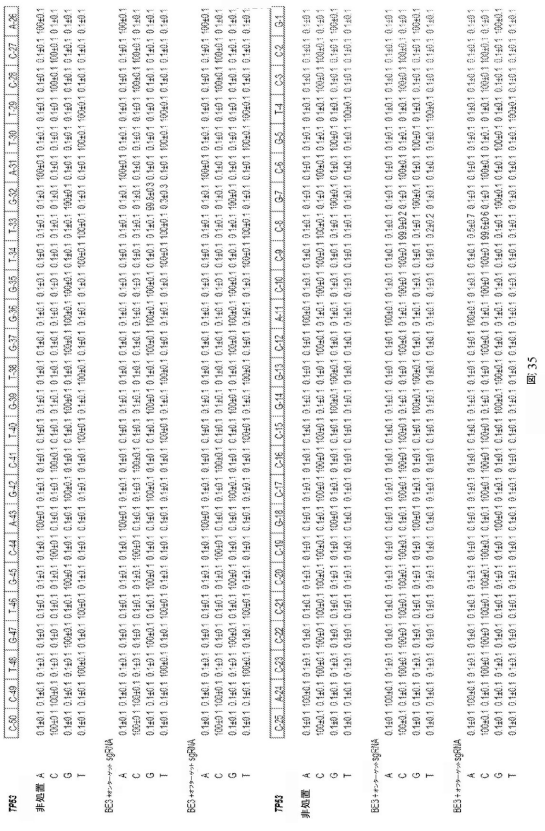


【図３４－１】

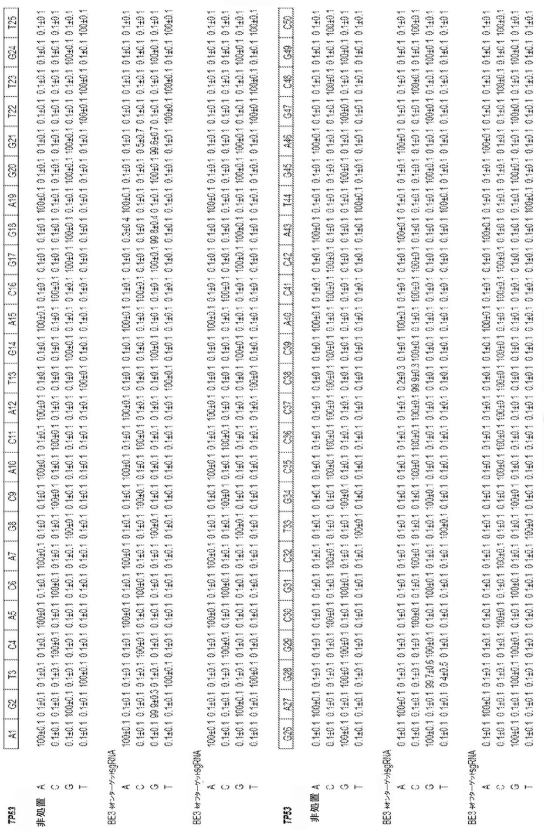


【図３４－２】

【図３５－１】



【図３５－２】



【図３６Ａ－Ｃ】

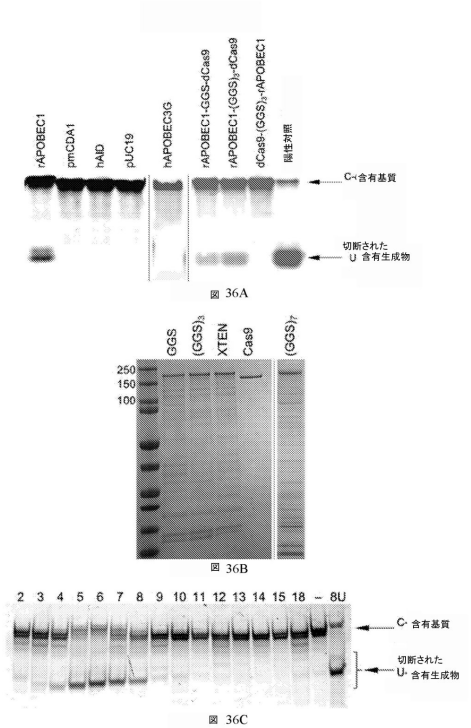


図 36A

図 36B

図 36C

【図３６Ｄ－Ｆ】

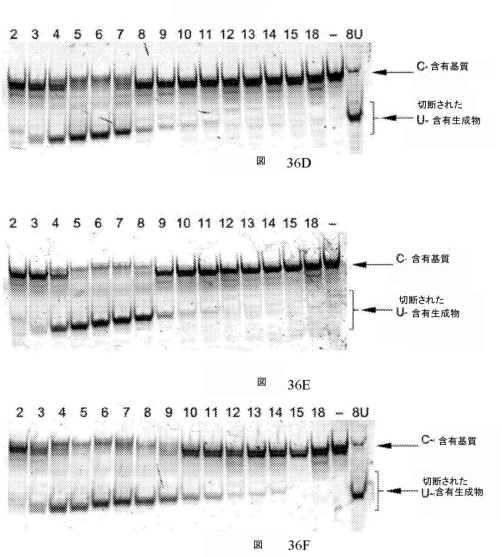


図 36D

図 36E

図 36F

【図３７Ａ】

EMX1: GAGTC$_5$C$_6$GAGCAGAAGAAGAAGGG

FANCF: GGAATC$_6$C$_7$C$_8$TTC$_{11}$TGCAGCACCTGG

HEK293 部位 2: GAAC$_4$AC$_6$AAAGCATAGACTGCGGG

HEK293 部位 3: GGCC$_4$C$_5$AGACTGAGCACGTGATGG

HEK293 部位 4: GGCAC$_5$TGCGGCTGGAGGTCCGGG

RNF2: GTC$_3$ATC$_6$TTAGTC$_{12}$ATTACCTGAGG

図 37A

【図３７Ｂ】

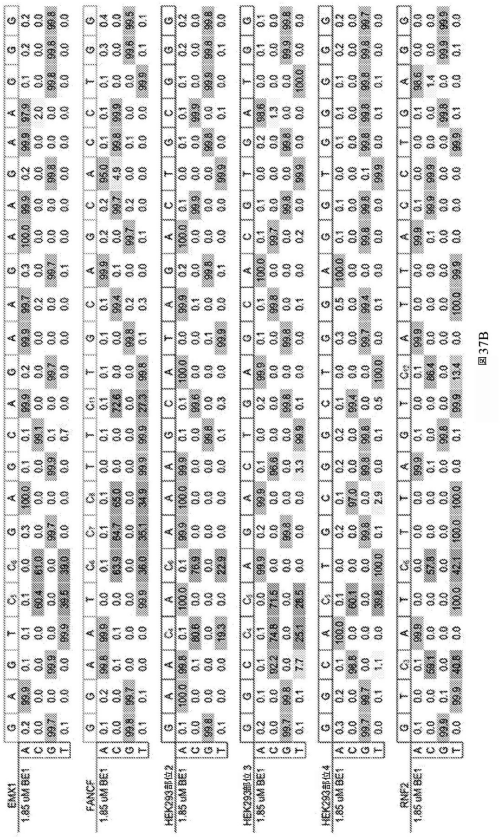


図37B

【図３７Ｃ】

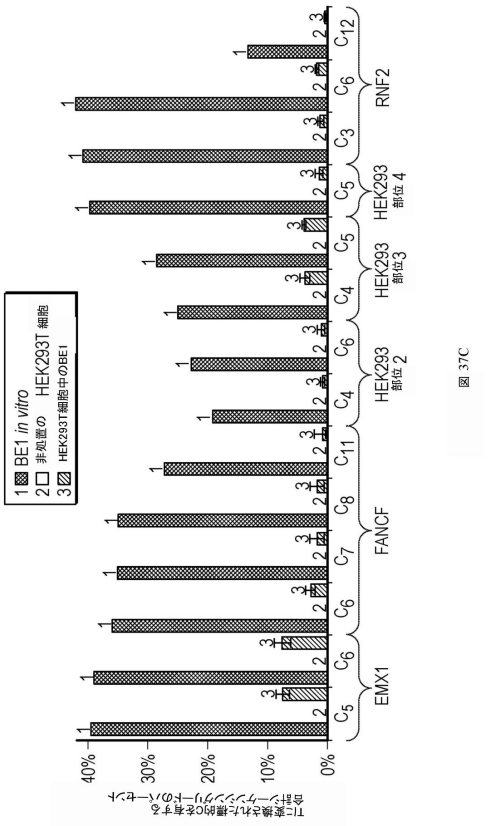


図 37C

【図３８】

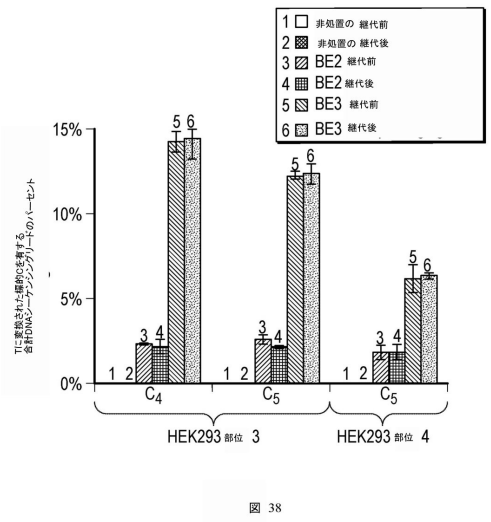


図 38

【図３９Ａ】

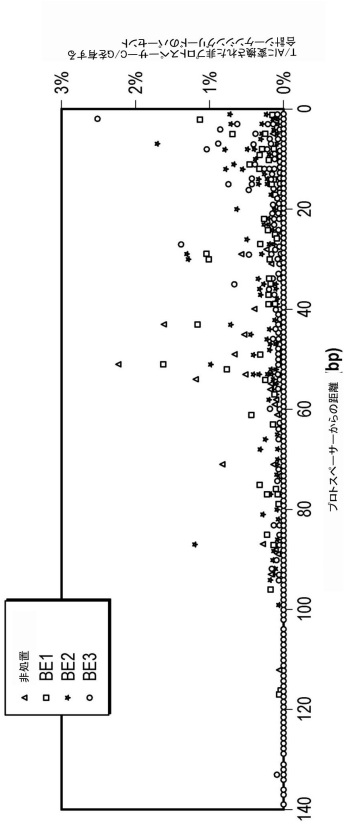


図 39A

【図３９Ｂ】

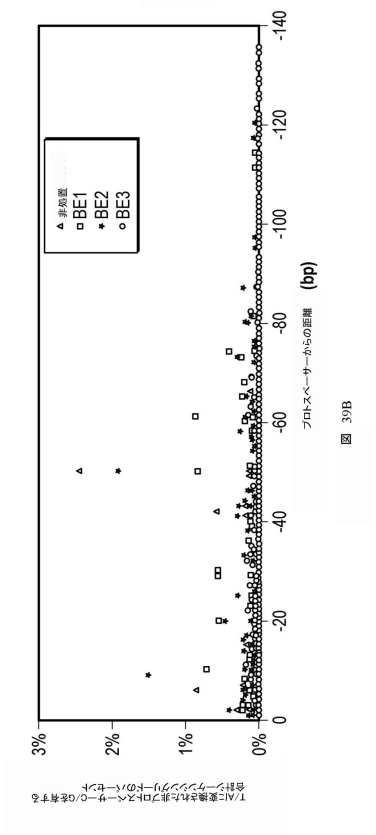


図 39B

【図３９Ｃ】

非プロトスペーサー C/G	非処置 C/G (%)	平均 T/A (%)	最も低い T/A (%)	最も高い T/A (%)
BE1	99.95 ± 0.14	0.02 ± 0.02	0.00	2.44
BE2	99.95 ± 0.24	0.03 ± 0.03	0.00	1.64
BE3	99.95 ± 0.13	0.03 ± 0.03	0.00	1.92
	99.97 ± 0.09	0.02 ± 0.02	0.00	2.52

図 39C

【図４０Ａ】

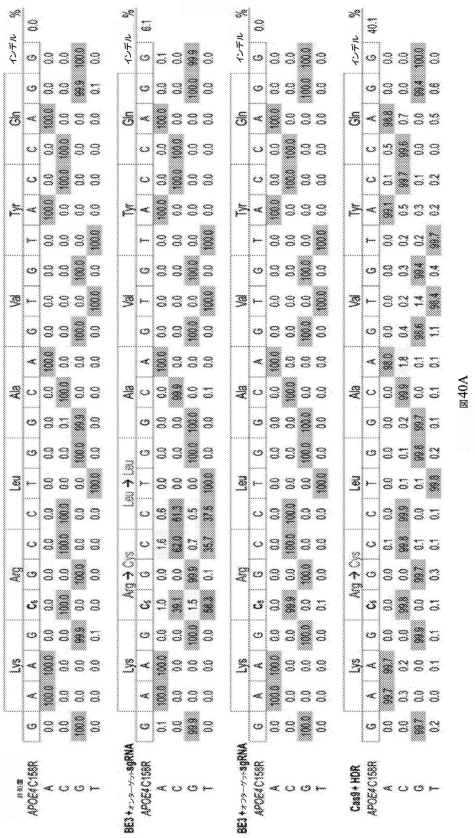


図 40A

【図４０Ｂ】

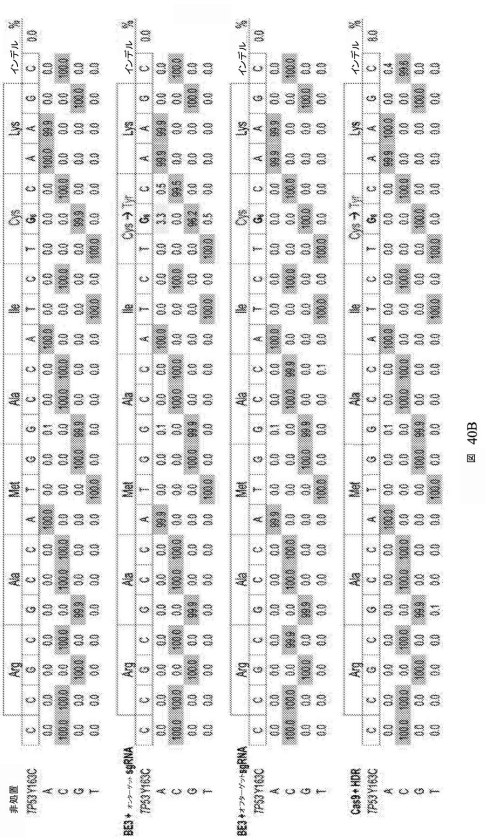


図 40B

【図４１】

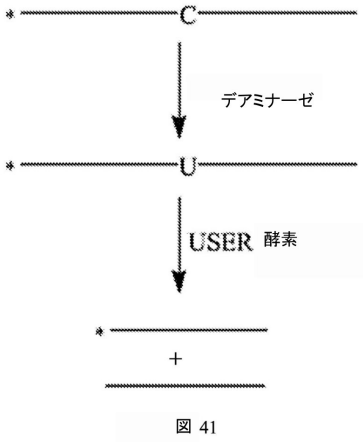


図 41

【図４２】

図 42

【図４３】

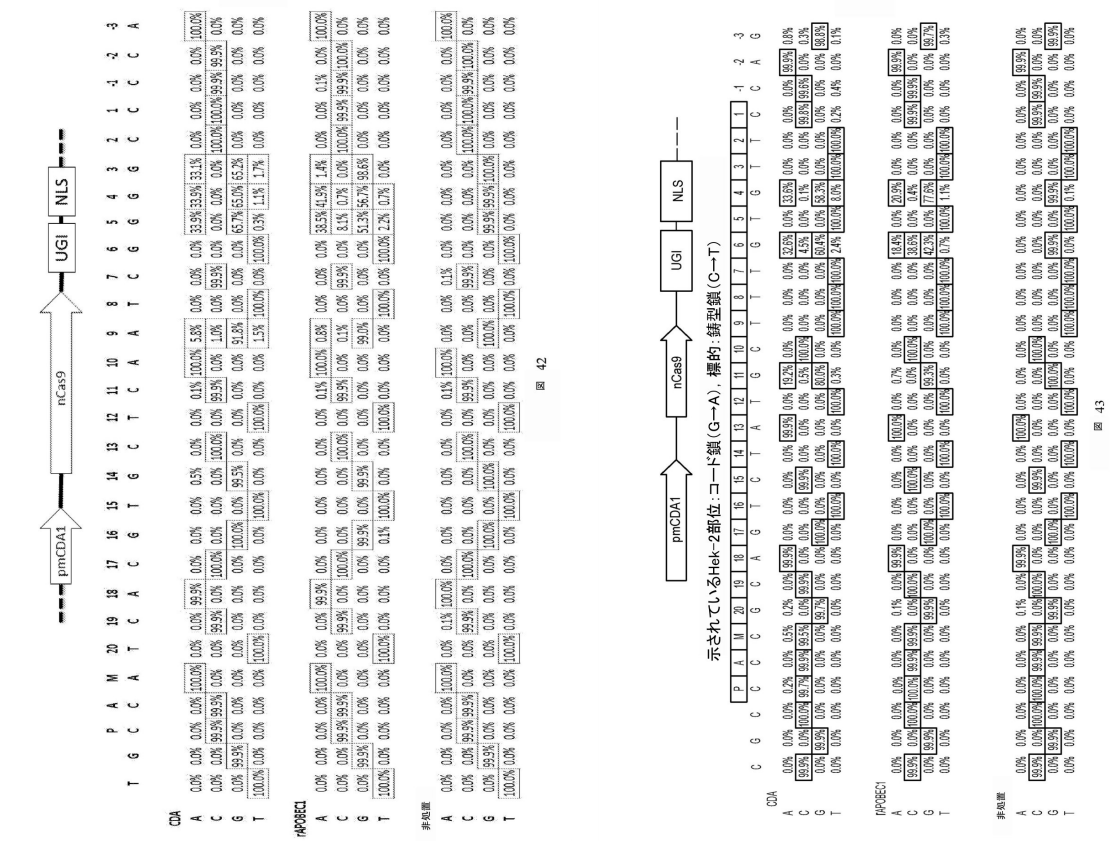


図 43

【図４４】

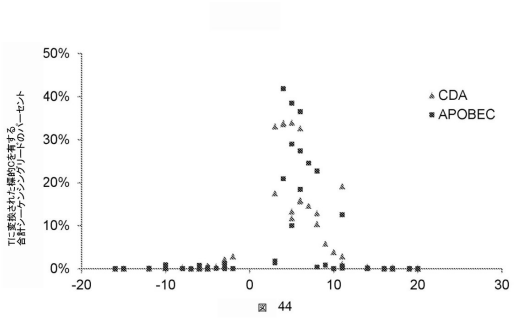


図 44

【図４５】

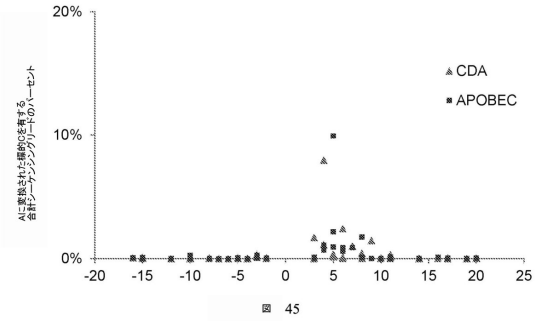


図 45

【図４６】

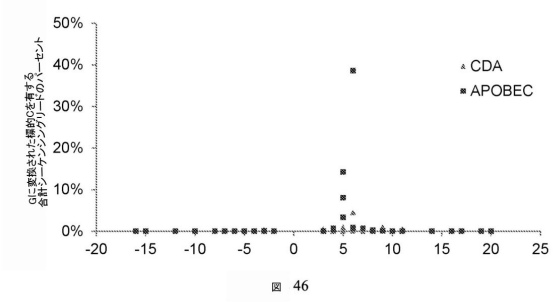


図 46

【図４７】

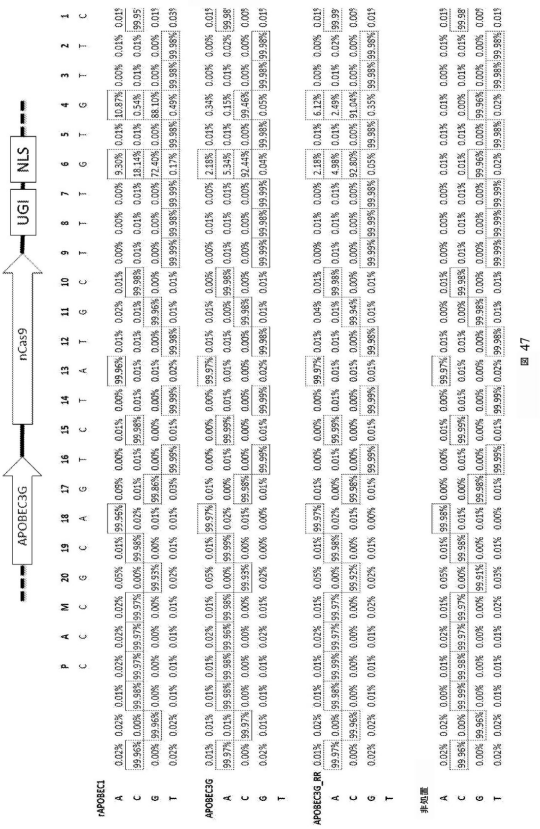


図 47

【図４８】

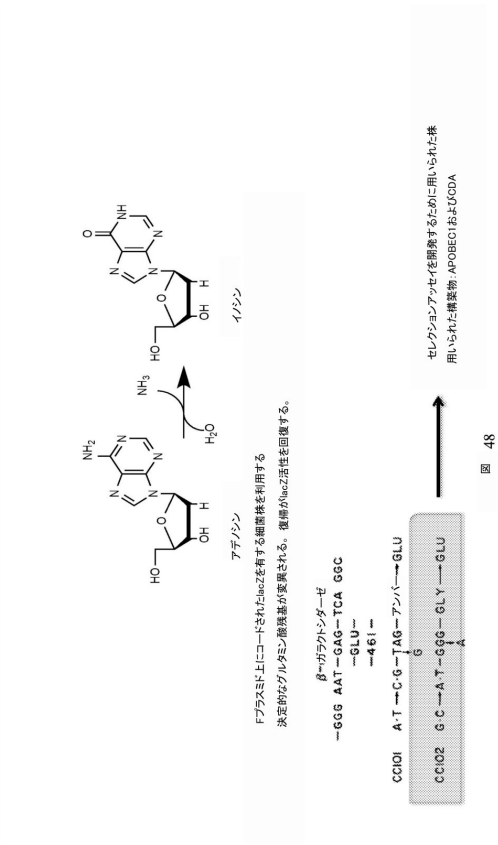


図 48

【図４９－１】

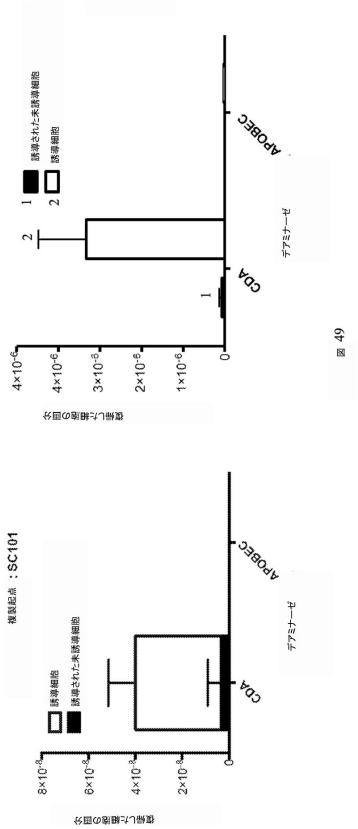


図 49

【図４９－２】

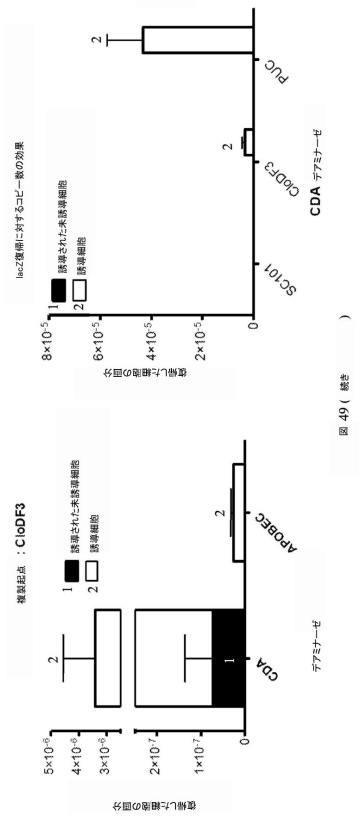


図 49（続き）

【図５０】

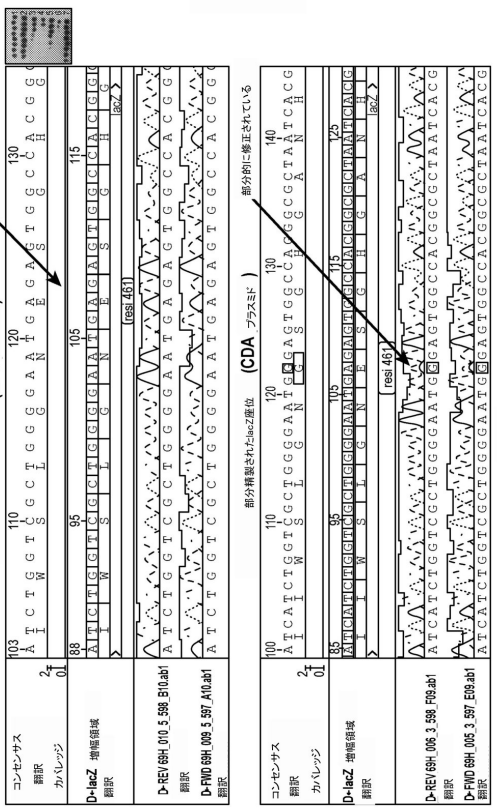


図 50

【図５１】

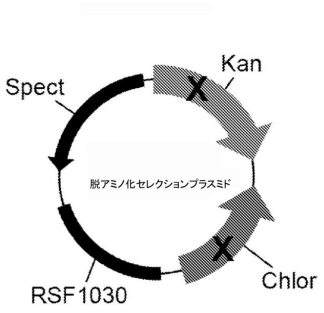
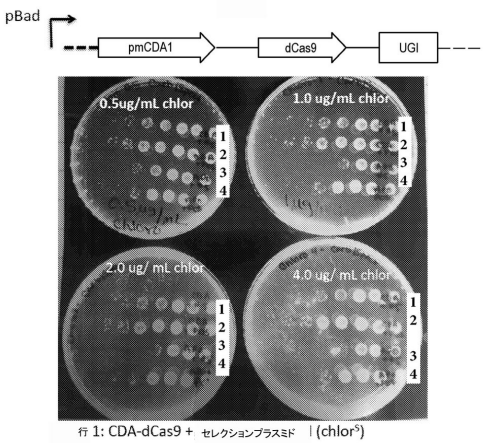


図 51

【図５２】



行 1: CDA-dCas9 + セレクションプラスミド (chlor^S)

行 2: CDA-dCas9 + 陽性対照セレクション (chlor^R)

行 3: rAPOBEC-dCas9 + セレクションプラスミド (chlor^S)

行 4: rAPOBEC-dCas9 + 陽性対照セレクション (chlor^R)

図 52

【図５３Ａ】

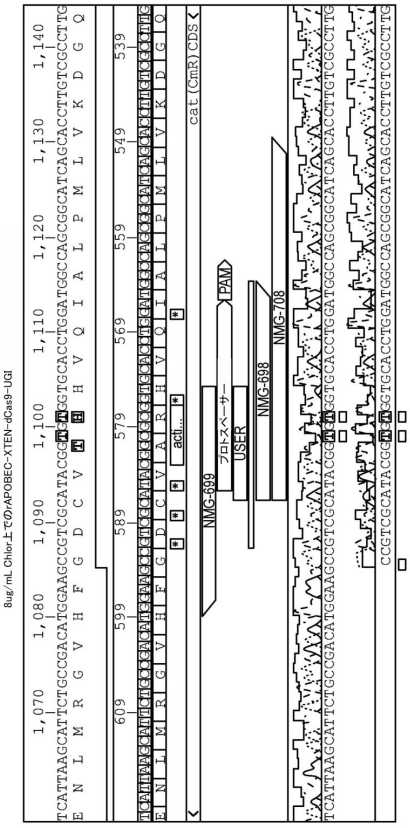


図 53A

【図５３Ｂ】

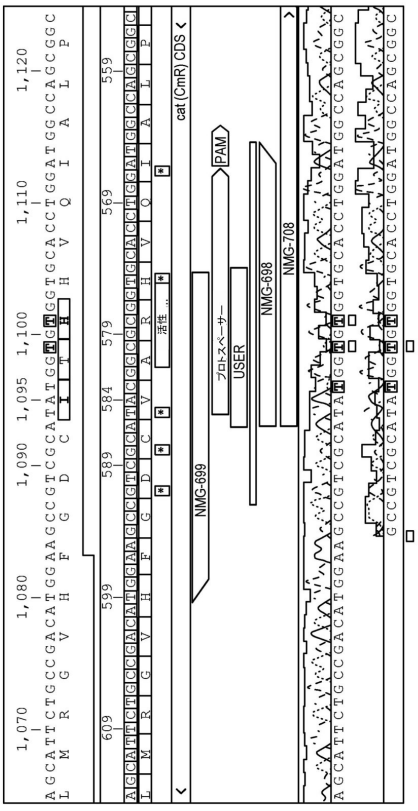


図 53B

【図５４】

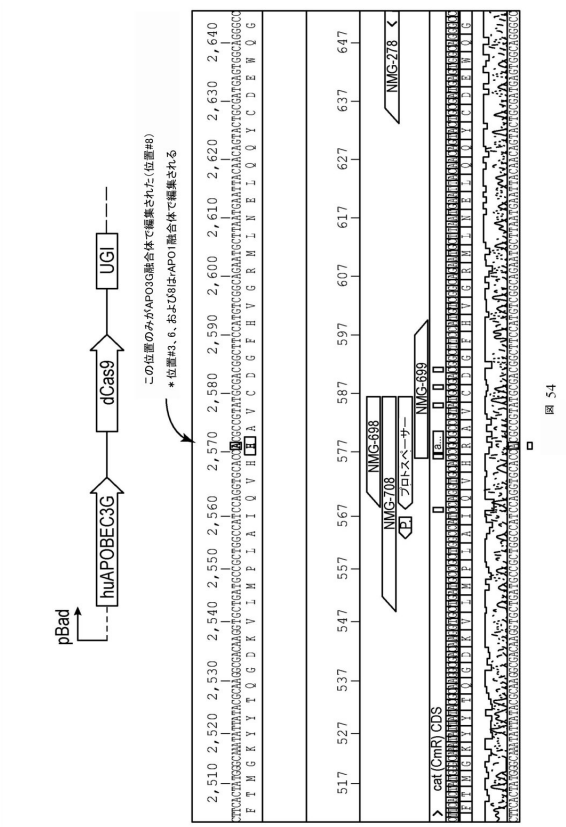


図 54

【図５５－１】

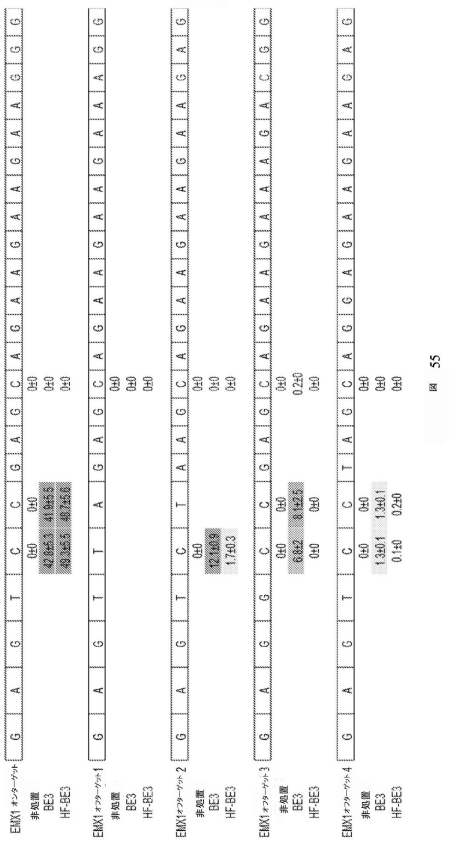


図 55

【図５５－２】

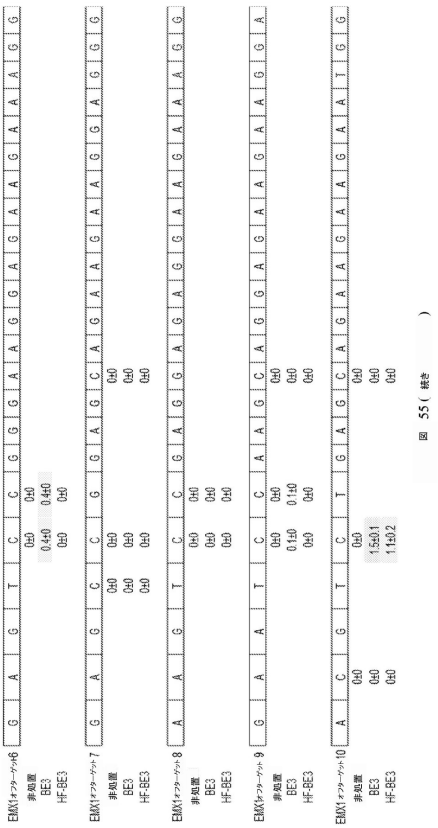


図 55（続き）

【図５６】

EMX1		C_2	C_6	C_{12}	インデル %
非処置	A	0±0	0±0	0±0	0±0
	C	99.9±0	99.9±0	99.9±0	
	G	0±0	0±0	0±0	
	T	0±0	0±0	0±0	
BE3	A	1.3±0.1	0.7±0.2	0±0	2.6±0.3
	C	51.7±1	55.2±0.1	99.9±0	
	G	4±0.7	1.9±0.4	0±0	
	T	42.8±0.2	41.9±0.4	0±0	
HF BE3	A	1.5±0.7	0.5±0.1	0±0	1±0.1
	C	44.8±8.5	49.9±6.6	99.9±0	
	G	4.2±1.2	0.6±0.2	0±0	
	T	49.3±6.5	48.7±6.3	0±0	

RNF2		C_3	C_6	C_{12}	インデル %
非処置	A	0±0	0±0	0±0	0±0
	C	99.9±0	99.9±0	99.9±0	
	G	0±0	0±0	0±0	
	T	0±0	0±0	0±0	
BE3	A	0.4±0	1.5±0.2	0.1±0	2.3±0.3
	C	70.9±2.7	45±0	95±0.2	
	G	0±0	12.6±2.4	0.1±0	
	T	28.5±2.7	40.7±2.7	4.6±0.2	
HF BE3	A	0±0	0.2±0	0±0	0.5±0.1
	C	76.8±1.8	64.5±3.3	95.4±0.1	
	G	0±0	2.4±0.5	0±0	
	T	23±1.8	32.7±2.6	4.5±0.1	

FANCF		C_6	C_7	C_8	C_{11}	インデル %
非処置	A	0±0	0±0	0±0	0±0	0±0
	C	99.9±0	99.9±0	99.9±0	99.9±0	
	G	0±0	0±0	0±0	0±0	
	T	0±0	0±0	0±0	0±0	
BE3	A	0.8±0.1	1.2±0.3	1.2±0.1	0.2±0.1	5.8±0.9
	C	64.7±5.9	67.9±5.1	68.9±5.2	85.4±2.7	
	G	0.5±0	0.7±0	0.7±0.1	0.1±0	
	T	33.8±5.7	29.9±4.8	29±5	14.1±2.6	
HF BE3	A	0.9±0.2	2±0.4	0.9±0.2	0.4±0	5.9±0.7
	C	52.5±8.4	57.7±7.6	61.9±6.8	84.7±2.4	
	G	1.4±0.2	0.6±0	0.5±0.1	0.3±0	
	T	45±8	39.5±7.2	36.5±6.5	14.4±2.2	

HEK3		C_6	C_7	C_8	C_{11}	インデル %
非処置	A	0±0	0±0	0±0	0±0	0±0
	C	99.9±0	99.9±0	99.9±0	99.9±0	
	G	0±0	0±0	0±0	0±0	
	T	0±0	0±0	0±0	0±0	
BE3	A	0±0	1.6±0.2	1.5±0.3	0.2±0	2.7±0.4
	C	98.7±0.2	48.6±5.4	40.8±6.7	98.6±0.1	
	G	0±0	1.4±0.2	10.5±1.2	0.3±0	
	T	1.1±0.2	48.2±4.9	47±5.2	0.7±0.1	
HF BE3	A	0±0	0.8±0.2	1.1±0.3	0.4±0	4.1±0.7
	C	97.8±0.4	65±3.8	40.4±6.3	95.9±0.5	
	G	0±0	0.9±0.2	6.1±1.1	0.7±0.2	
	T	2±0.4	33±3.4	52.2±5.2	2.7±0.2	

HEK4		C_6	C_7	C_8	C_{11}	インデル %
非処置	A	0±0	0±0	0±0	0±0	0±0
	C	99.9±0	99.9±0	99.9±0	99.9±0	
	G	0±0	0±0	0±0	0±0	
	T	0±0	0±0	0±0	0±0	
BE3	A	0±0	5.4±1.3	0±0	0±0	3±0.7
	C	98.7±0	44.5±8.2	98.9±0	99.8±0	
	G	0±0	18.1±2.3	0±0	0±0	
	T	1.1±0.1	31.9±4.8	0.9±0.1	0.1±0	
HF BE3	A	0±0	5.2±0.8	0±0	0±0	1.3±0
	C	99.6±0	36.6±0	99.5±0.1	99.9±0	
	G	0±0	16±2.8	0±0	0±0	
	T	0.2±0	42.1±5.3	1.3±0.1	0±0	

図 56

【図５７】

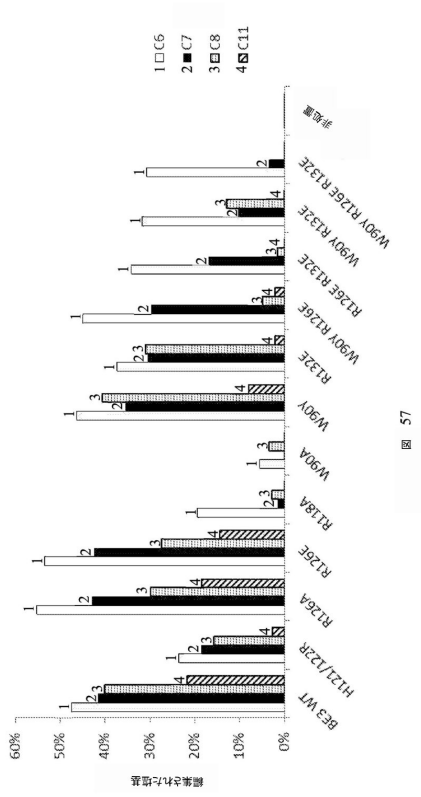


図 57

【図５８】

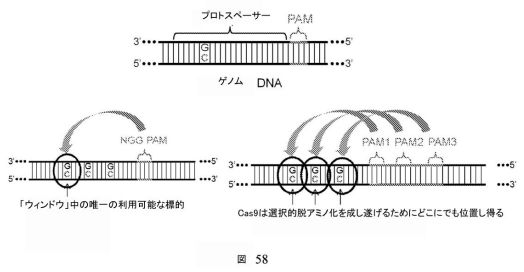


図 58

【図５９】

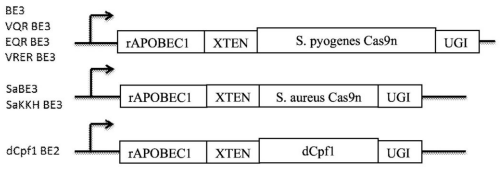


図 59

【図６０】

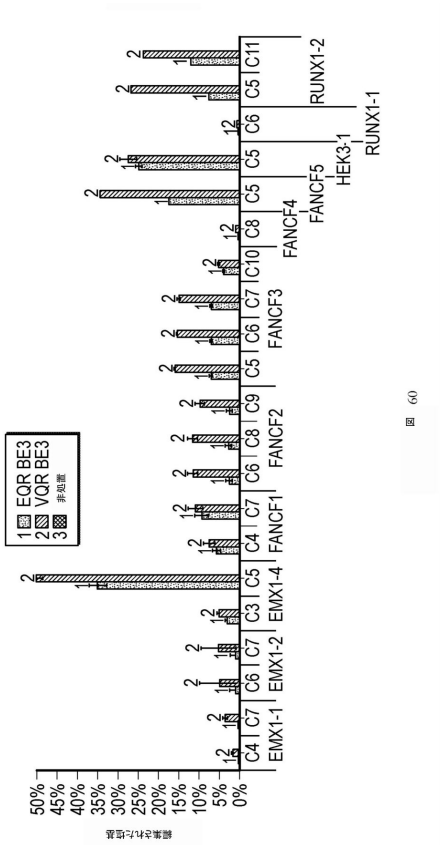


図 60

【図６１】

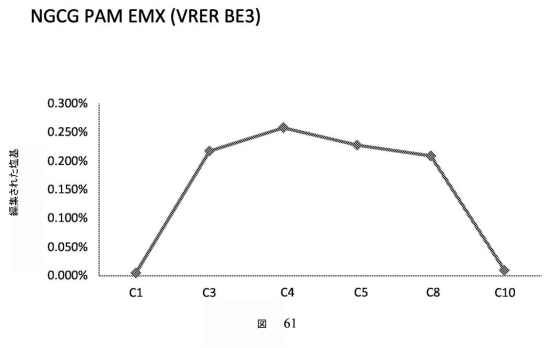


図 61

【図６２】

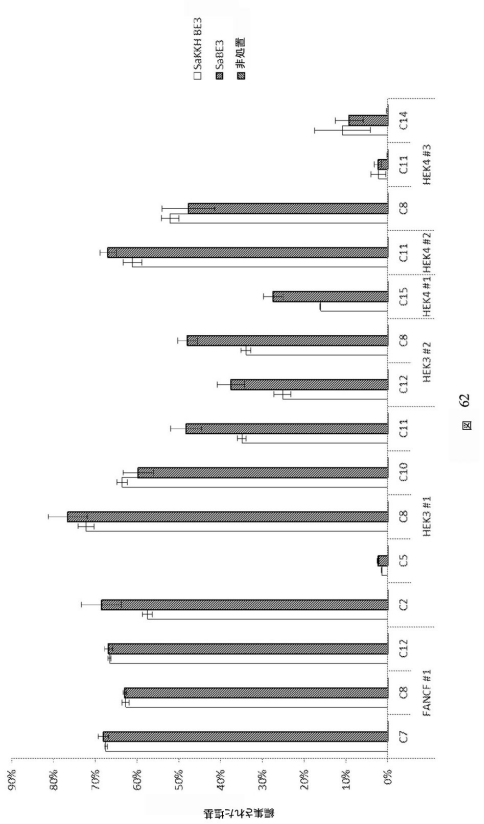


図 62

【図６３】

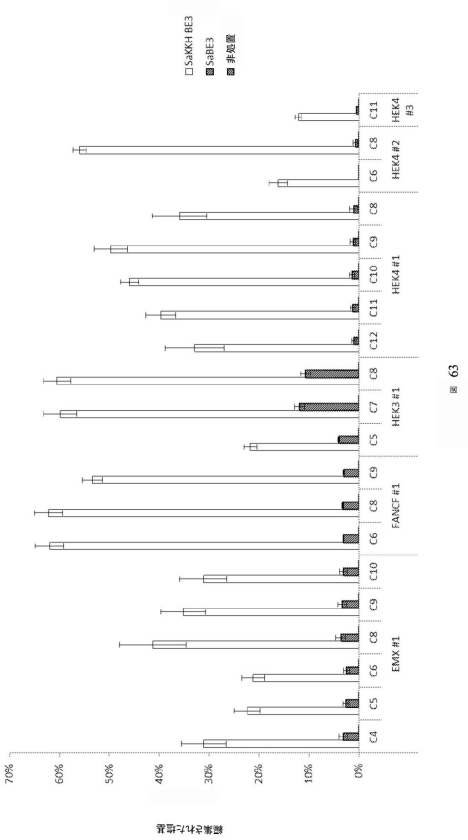


図 63

【図６４】

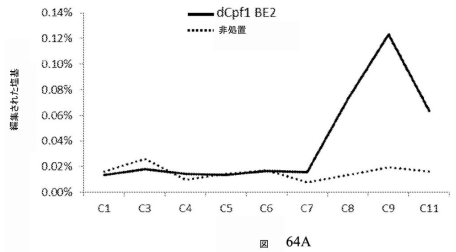


図 64A

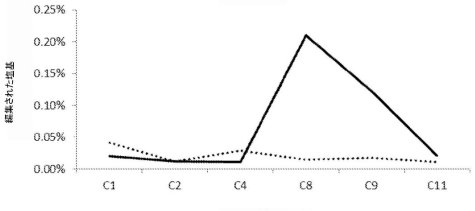


図 64B

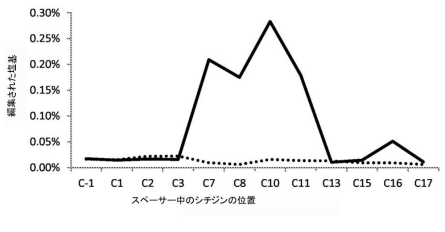


図 64C

【図６５】

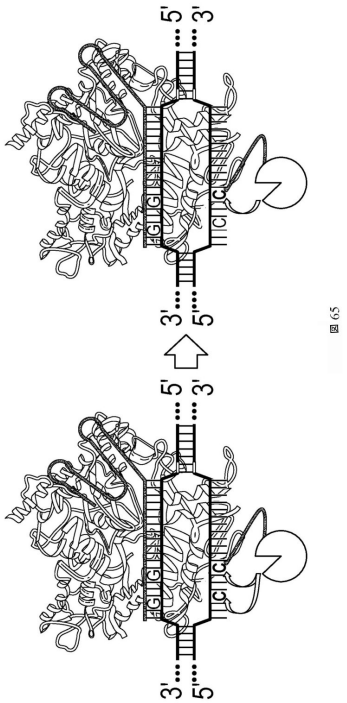


図 65

【図６６】

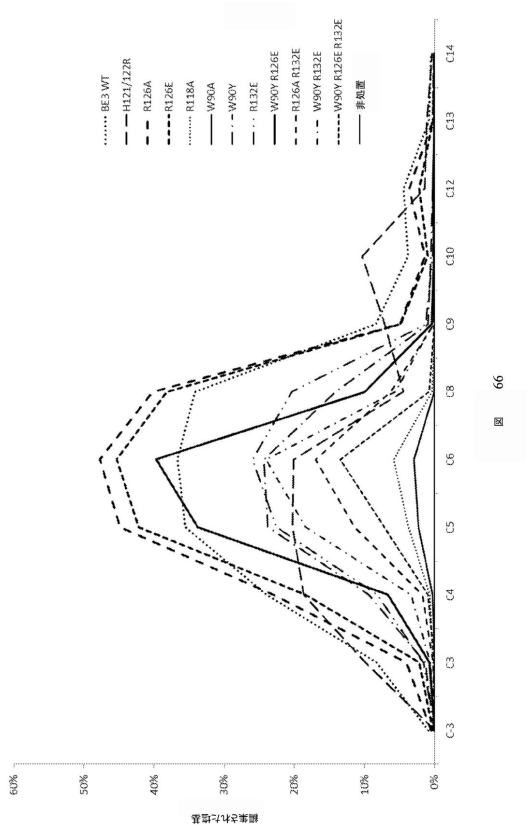


図 66

【図６７】

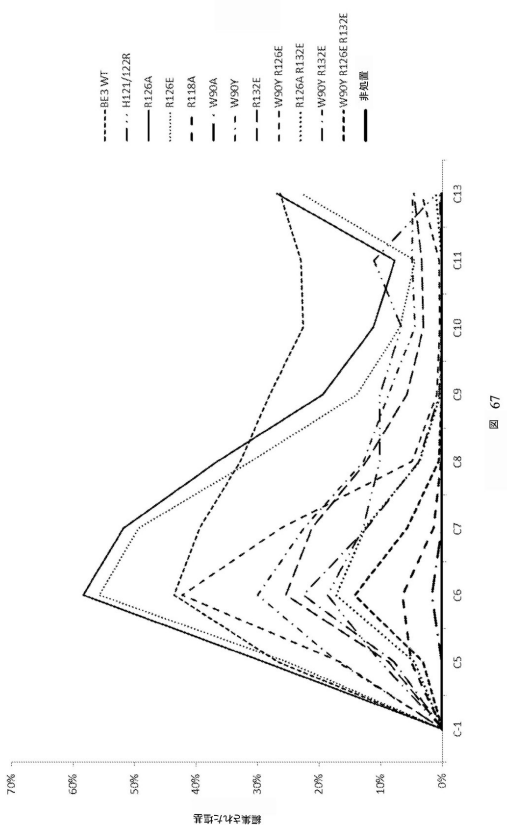


図 67

【図６８】

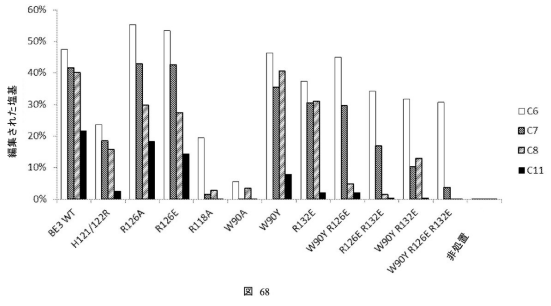


図 68

【図６９】

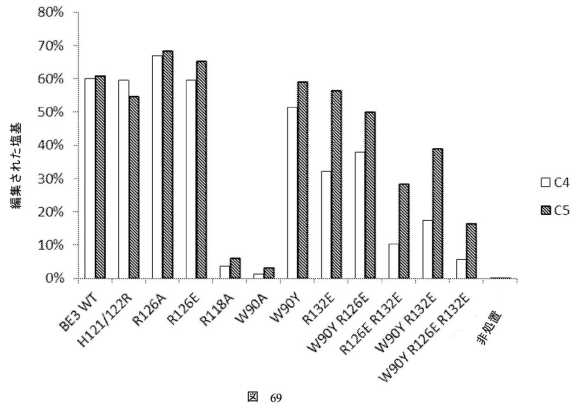


図 69

【図７０】

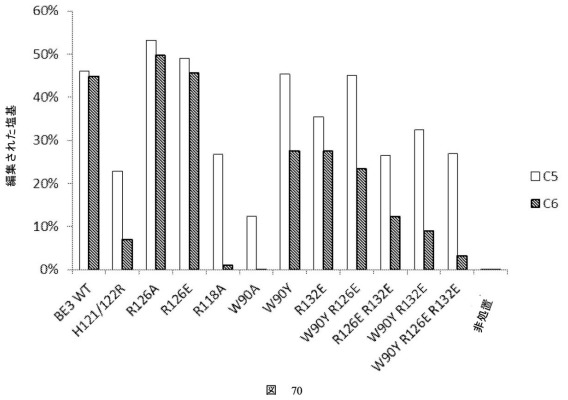


図 70

【図７１】

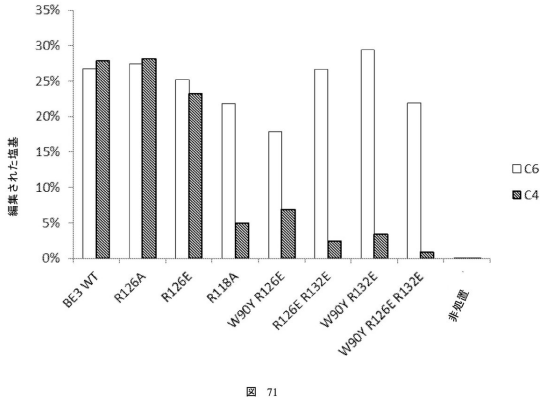


図 71

【図７２】

BE3

塩基	Lys				Arg→Cys				Leu→Leu				Ala				Tyr				Gln			
	A	A	G	G	C	G	C	C	T	T	G	G	C	C	A	A	T	T	G	G	A	A	C	C
A	0.2	99.9	0.0	0.0	1.0	0.1	1.8	1.2	0.0	0.0	0.0	0.1	0.1	99.8	0.0	0.0	0.0	0.0	0.1	99.8	0.1	99.9	0.1	0.0
C	0.0	0.0	0.1	0.0	38.1	0.0	49.8	52.3	0.0	0.0	0.0	0.0	99.7	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	99.9	99.9
G	99.8	0.0	0.0	99.9	1.8	99.8	1.3	0.7	0.1	99.9	99.8	0.0	0.1	99.9	0.1	99.9	0.1	99.9	0.1	0.0	0.0	0.0	0.0	0.0
T	0.0	0.0	0.1	0.1	59.2	0.1	47.0	45.8	99.9	0.1	0.1	0.2	0.0	0.0	0.0	99.8	0.0	0.0	99.8	0.0	0.0	0.0	0.1	0.1

BE3 W90Y R132E

塩基	Lys				Arg→Cys				Leu→Leu				Ala				Tyr				Gln			
	A	A	G	G	C	G	C	C	T	T	G	G	C	C	A	A	T	T	G	G	A	A	C	C
A	0.0	99.9	99.9	0.0	0.5	0.1	0.3	0.1	0.0	0.0	0.0	0.0	0.1	99.8	0.0	0.0	0.0	0.0	0.1	99.9	0.1	99.9	0.1	0.0
C	0.0	0.0	0.0	0.1	78.0	0.0	94.8	98.9	0.0	0.0	0.0	0.0	99.8	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	99.9
G	100.0	0.0	0.0	99.9	0.5	99.9	0.1	0.1	0.0	99.9	99.9	99.9	0.0	0.1	0.1	100.0	0.1	99.9	0.0	0.0	0.0	0.0	0.0	0.0
T	0.0	0.0	0.0	0.0	21.0	0.0	4.8	0.9	99.9	0.0	0.0	0.1	0.0	0.0	0.0	99.8	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0

図 72

【図７３】

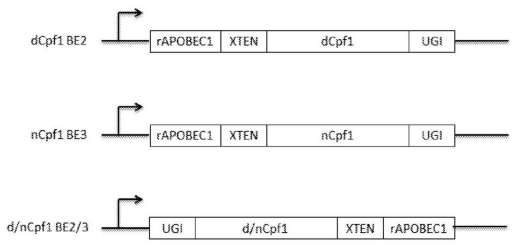


図 73

【図７４】

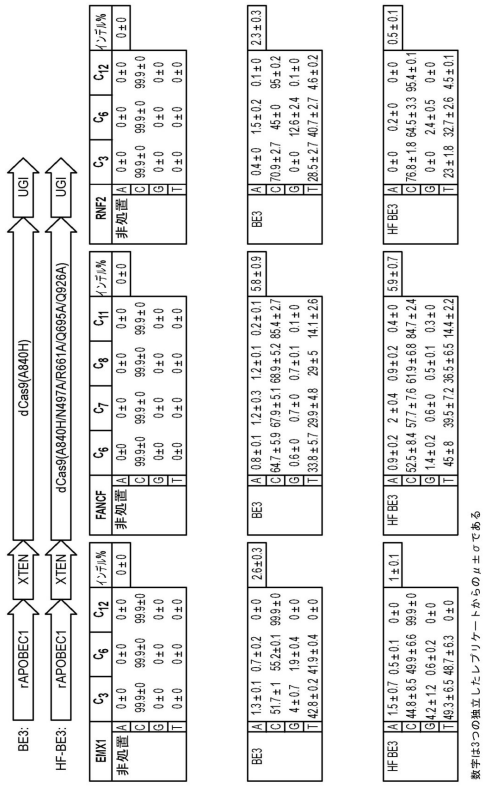


図 74

【図７５】

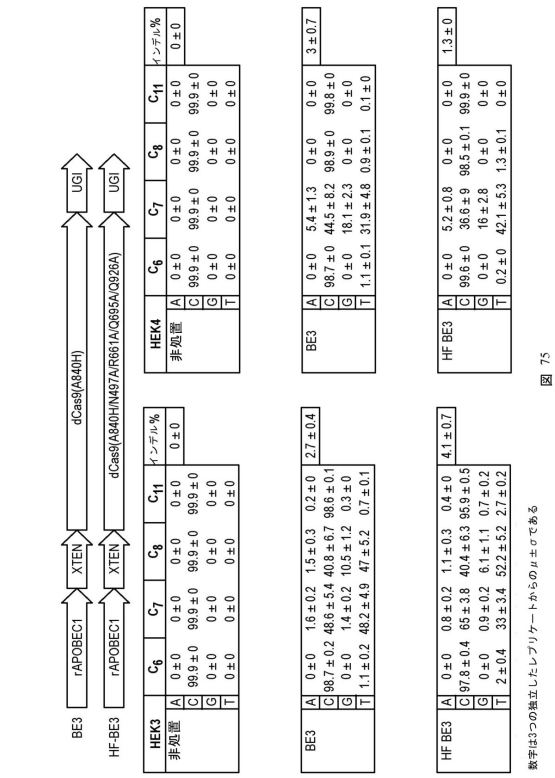


図 75

【図７６－１】

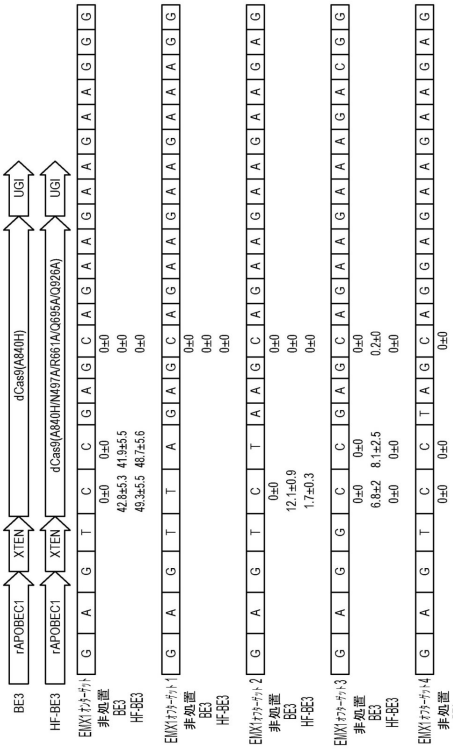
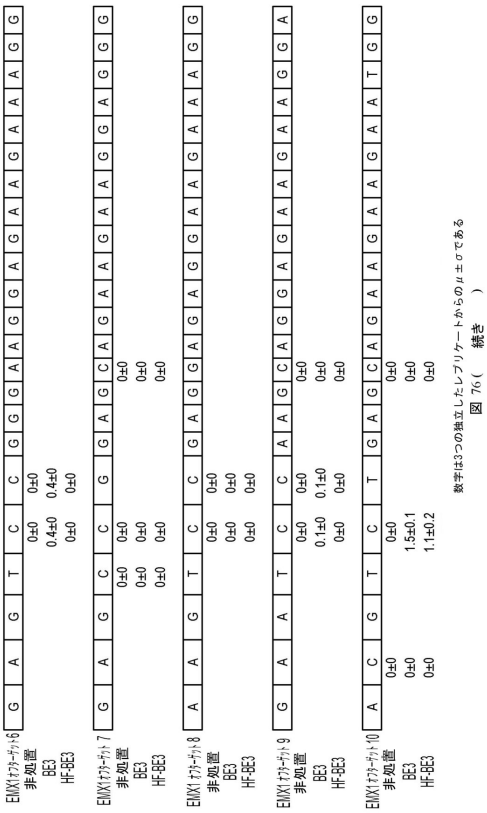


図 76

【図７６－２】



数字は3つの独立したレプリケートからのμ±σである

図 76（　続き　）

【図７７】

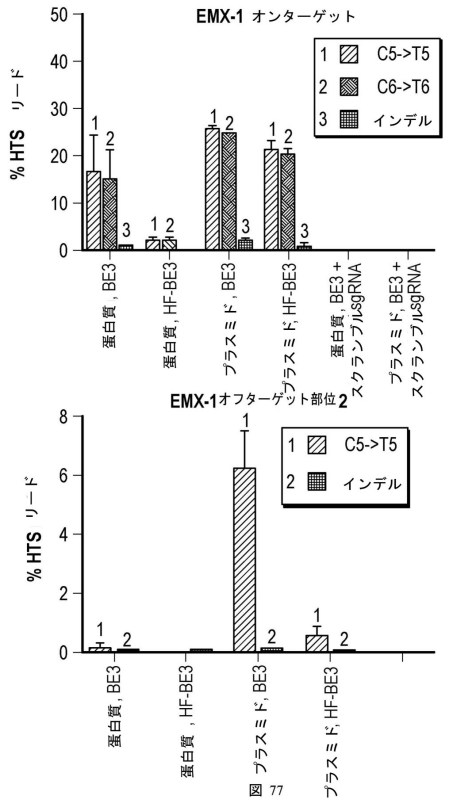


図 77

【図７８】

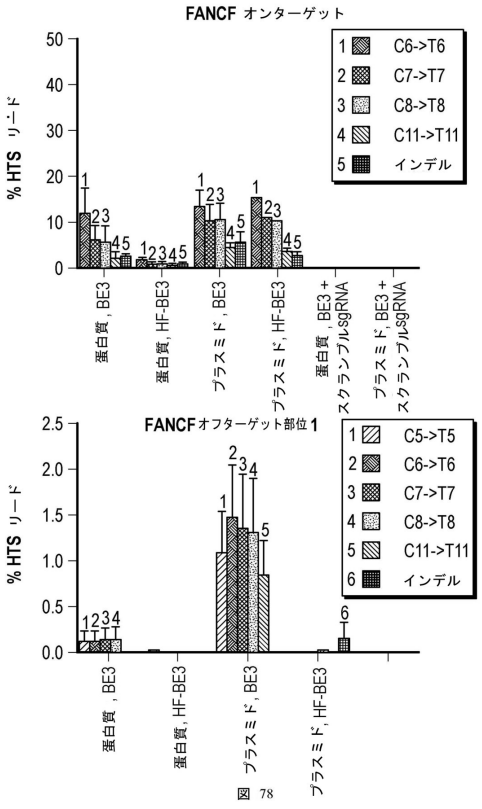


図 78

【図７９】

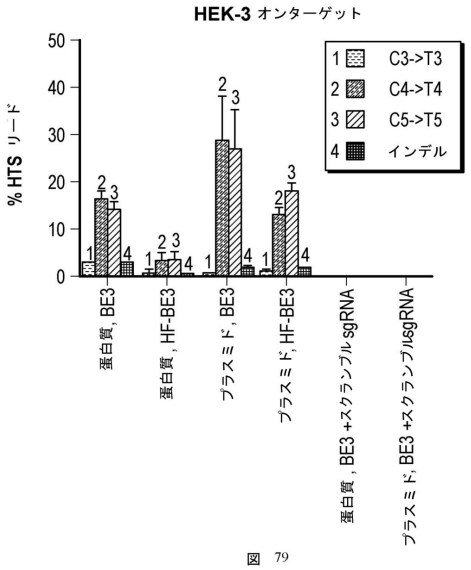


図 79

【図８０－１】

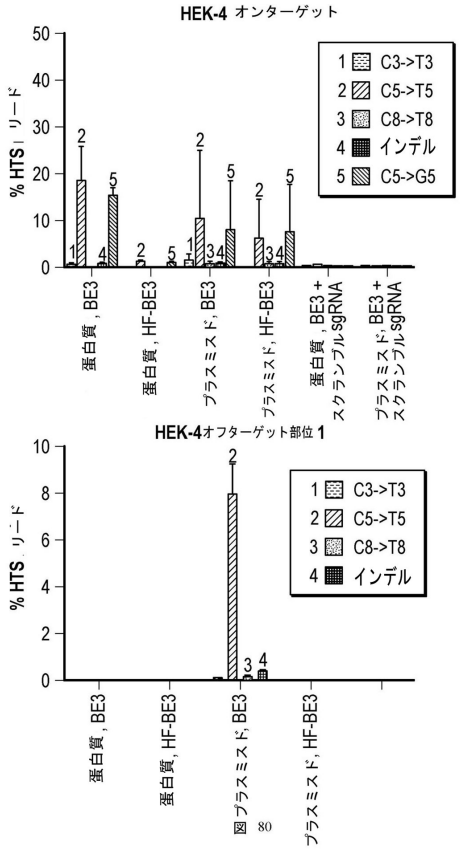


図 80

【図８０－２】

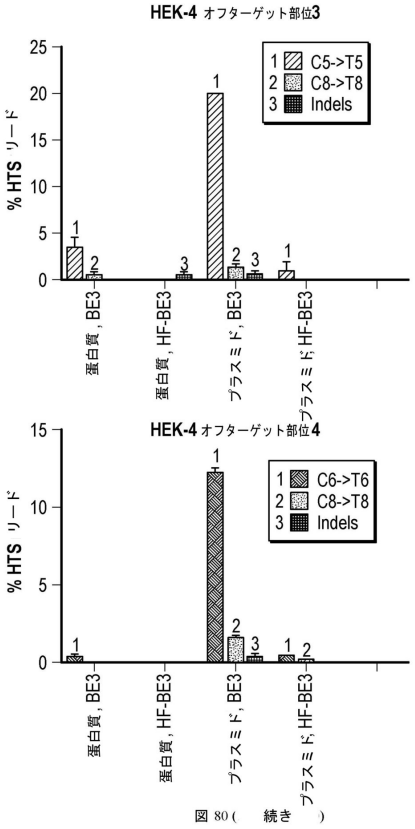


図 80 (続き)

【図８１】

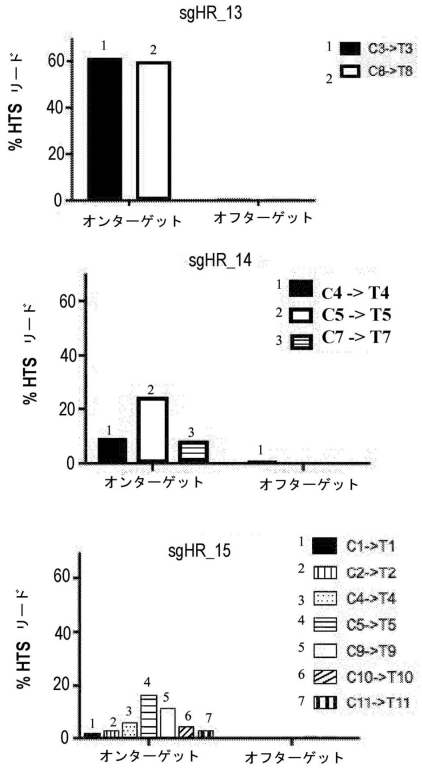


図 81

【図８２】

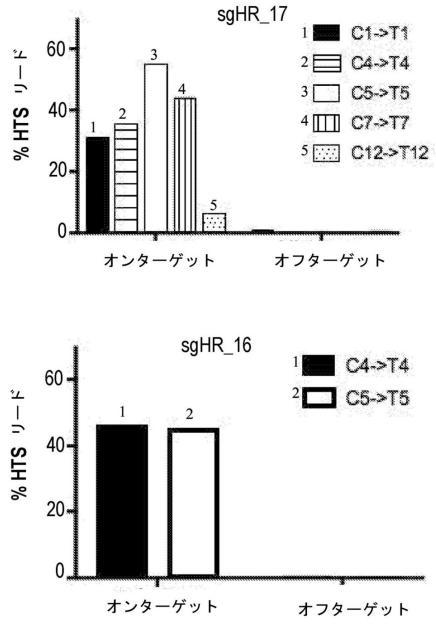


図 82

【図８３】

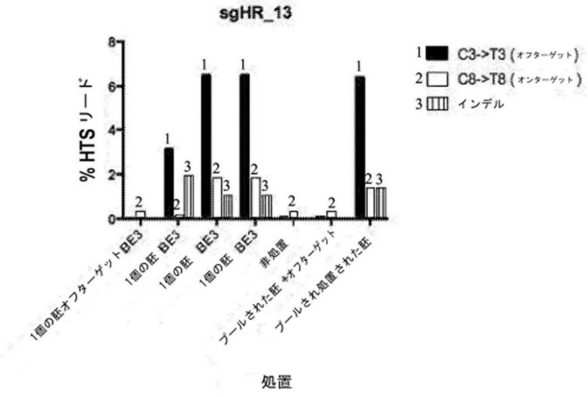


図 83

【図８４】

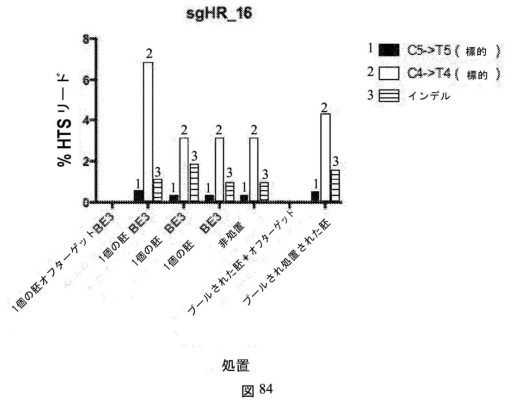


図 84

【図８５】

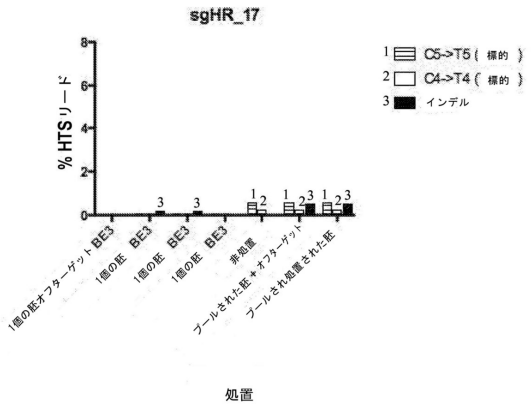


図 85

【図８６】

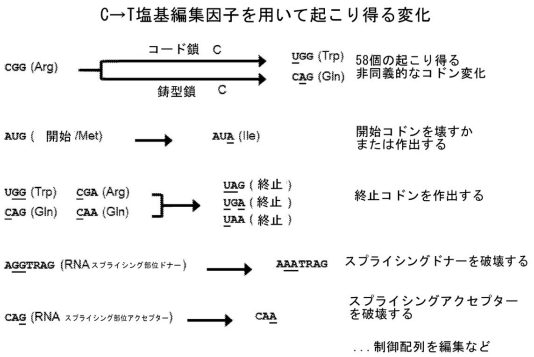
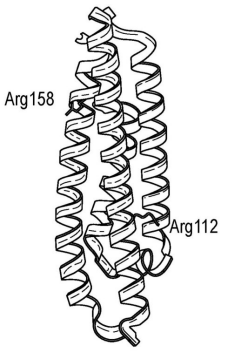


図 86

【図87】

APOE4→APOE3rの塩基編集



- アポリポ蛋白質Eはアミノ酸112および158が異なる4つのアイソフォームを有する
- APO4Eは遅発型アルツハイマー病にとっての、最も大きくかつ最も一般的な遺伝学的リスク因子である

アレル	残基 112	残基 158	AD リスク（公知または推定）
APOE2	Cys TGC	Cys TGC	低
APOE3	Cys TGC	Arg CGC	中
APOE3r	Arg CGC	Cys TGC	中
APOE4	Arg CGC	BE3 → Arg CGC	高

図 87

【図88】

非処置		Lys			Arg 158			Leu			Ala			Val			Tyr			Gln			インデル%
APOE4 C158R		G	A	A	G	C_5	G	C	C	T	G	G	C	A	G	T	G	T	A	C	C	A	G	G	0.0
	A	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	
	C	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	
	G	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	100.0	
	T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

Cas9 + HDR		Lys			Arg 158 → Cys			Leu			Ala			Val			Tyr			Gln			インデル%
APOE4 C158R		G	A	A	G	C_5	G	C	C	T	G	G	C	A	G	T	G	T	A	C	C	A	G	G	26.1
	A	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.4	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	
	C	0.0	0.0	0.0	0.0	99.7	0.0	99.9	99.9	0.0	0.0	0.0	100.0	0.5	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	
	G	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	99.9	0.0	0.0	99.6	0.6	99.9	0.2	0.0	0.0	0.0	0.0	100.0	100.0	
	T	0.0	0.0	0.0	0.0	**0.3**	0.0	0.1	0.1	100.0	0.0	0.0	0.0	0.1	0.4	99.3	0.1	99.8	0.0	0.0	0.0	0.0	0.0	0.0	

BE3 処置		Lys			Arg 158 → Cys			Leu → Leu			Ala			Val			Tyr			Gln			インデル%
APOE4 C158R		G	A	A	G	C_5	G	C_7	C_8	T	G	G	C	A	G	T	G	T	A	C	C	A	G	G	4.6
	A	0.1	100.0	100.0	0.0	0.5	0.0	1.3	0.9	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.1	
	C	0.0	0.0	0.0	0.0	23.7	0.0	47.4	43.5	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	
	G	99.9	0.0	0.0	100.0	0.9	99.9	1.1	0.7	0.0	100.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	99.9	
	T	0.0	0.0	0.0	0.0	**74.9**	0.1	**50.2**	**55.0**	100.0	0.0	0.0	0.1	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	

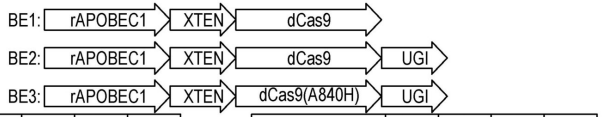
図 88

【図89】

BE3 処置		Gly			Ser → Ser			Arg 37 → 終止			Tyr			Pro			Gly			Glu				
PRNP R37X		G	G	C	A	G	C_7	C_8	G	A	T	A	C	C	C	G	G	G	G	C	A	G	G	G
	A	0.0	0.0	0.1	100.0	0.0	1.1	0.3	0.1	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
	C	0.0	0.0	98.9	0.0	0.0	54.9	57.8	0.0	0.0	0.0	0.0	99.8	99.8	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
	G	100.0	100.0	0.0	0.0	100.0	2.1	0.6	99.9	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	100.0	100.0	100.0
	T	0.0	0.0	1.1	0.0	0.0	41.9	41.2	0.0	0.0	100.0	0.0	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

図 89

【図90－1】



EMX1・親		C_5	C_6	C_{10}	インデル%
非処置	A	0.0%	0.0%	0.0%	0.01%
	C	100.0%	100.0%	100.0%	
	G	0.0%	0.0%	0.0%	
	T	0.0%	0.0%	0.0%	
BE1	A	0.4%	0.1%	0.0%	0.18%
	C	95.8%	97.4%	100.0%	
	G	0.9%	0.2%	0.0%	
	T	2.9%	2.3%	0.0%	
BE2	A	0.3%	0.1%	0.0%	0.15%
	C	94.9%	95.7%	100.0%	
	G	1.0%	0.1%	0.0%	
	T	3.8%	4.0%	0.0%	
BE3	A	1.4%	0.8%	0.0%	2.54%
	C	78.5%	84.7%	100.0%	
	G	3.0%	1.4%	0.0%	
	T	17.1%	13.0%	0.0%	

EMX1-UDG KO		C_5	C_6	C_{10}	インデル%
非処置	A	0.0%	0.0%	0.0%	0.01%
	C	99.9%	99.9%	100.0%	
	G	0.0%	0.0%	0.0%	
	T	0.1%	0.1%	0.0%	
BE1	A	0.0%	0.0%	0.0%	0.04%
	C	76.5%	76.6%	100.0%	
	G	0.0%	0.0%	0.0%	
	T	23.5%	23.4%	0.0%	
BE2	A	0.0%	0.0%	0.0%	0.02%
	C	96.3%	96.3%	100.0%	
	G	0.0%	0.0%	0.0%	
	T	3.7%	3.7%	0.0%	
BE3	A	0.0%	0.0%	0.0%	0.04%
	C	84.5%	86.8%	100.0%	
	G	0.0%	0.0%	0.0%	
	T	15.5%	13.2%	0.0%	

図 90

【図９０－２】

FANCF- 親		C_6	C_7	C_8	C_{11}	インデル%
非処置	A	0.0%	0.0%	0.0%	0.0%	0.03%
	C	99.9%	99.9%	99.9%	99.9%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	0.1%	0.1%	0.1%	0.0%	
BE1	A	0.5%	0.5%	0.3%	0.0%	0.13%
	C	94.2%	97.8%	97.9%	99.8%	
	G	0.7%	0.0%	0.0%	0.0%	
	T	4.7%	1.6%	1.8%	0.2%	
BE2	A	0.3%	0.4%	0.2%	0.0%	0.25%
	C	95.3%	97.3%	97.6%	99.8%	
	G	0.4%	0.1%	0.0%	0.0%	
	T	4.0%	2.3%	2.1%	0.1%	
BE3	A	2.4%	3.2%	2.2%	0.9%	18.88%
	C	60.3%	72.6%	73.8%	86.6%	
	G	1.2%	0.6%	0.4%	0.3%	
	T	36.2%	23.6%	23.5%	12.3%	

FANCF-UDG KO		C_6	C_7	C_8	C_{11}	インデル%
非処置	A	0.0%	0.0%	0.0%	0.0%	0.02%
	C	99.9%	99.9%	100.0%	100.0%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	0.1%	0.0%	0.0%	0.0%	
BE1	A	0.0%	0.1%	0.0%	0.0%	0.09%
	C	60.9%	61.1%	61.3%	94.7%	
	G	0.1%	0.1%	0.1%	0.0%	
	T	39.0%	38.8%	38.8%	5.2%	
BE2	A	0.0%	0.0%	0.0%	0.0%	0.02%
	C	88.6%	88.6%	88.7%	98.9%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	11.4%	11.4%	11.3%	1.1%	
BE3	A	0.0%	0.1%	0.1%	0.0%	0.35%
	C	47.9%	48.5%	48.8%	55.8%	
	G	0.0%	0.1%	0.3%	0.2%	
	T	52.1%	51.3%	50.8%	44.0%	

図 90（ 続き ）

【図９０－３】

HEK3- 親		C_3	C_4	C_5	C_9	インデル%
非処置	A	0.0%	0.0%	0.0%	0.0%	0.00%
	C	100.0%	99.9%	99.9%	100.0%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	0.0%	0.0%	0.0%	0.0%	
BE1	A	0.0%	0.4%	0.3%	0.1%	0.07%
	C	99.9%	96.3%	94.4%	99.8%	
	G	0.0%	0.1%	1.9%	0.1%	
	T	0.1%	3.2%	3.4%	0.1%	
BE2	A	0.0%	0.2%	0.3%	0.1%	0.05%
	C	100.0%	97.7%	96.1%	99.8%	
	G	0.0%	0.0%	1.2%	0.1%	
	T	0.0%	2.1%	2.4%	0.1%	
BE3	A	0.1%	2.9%	2.4%	0.3%	3.27%
	C	99.4%	62.6%	55.9%	98.8%	
	G	0.0%	1.5%	10.4%	0.3%	
	T	0.6%	33.0%	31.3%	0.6%	

HEK3-UDG KO		C_3	C_4	C_5	C_9	インデル%
非処置	A	0.0%	0.0%	0.0%	0.0%	0.00%
	C	100.0%	100.0%	100.0%	100.0%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	0.0%	0.0%	0.0%	0.0%	
BE1	A	0.0%	0.0%	0.0%	0.0%	0.03%
	C	96.0%	50.3%	41.1%	96.3%	
	G	0.0%	0.0%	0.1%	0.1%	
	T	4.0%	49.6%	58.8%	3.6%	
BE2	A	0.0%	0.0%	0.0%	0.0%	0.00%
	C	99.0%	80.2%	73.2%	98.8%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	1.0%	19.7%	26.8%	1.1%	
BE3	A	0.0%	0.0%	0.0%	0.0%	0.09%
	C	98.5%	55.6%	40.3%	99.4%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	1.4%	44.3%	59.7%	0.6%	

図 90（ 続き ）

【図９０－４】

HEK4- 親		C_3	C_5	C_8	C_{11}	インデル%
非処置	A	0.0%	0.0%	0.0%	0.0%	0.00%
	C	100.0%	100.0%	99.9%	100.0%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	0.0%	0.0%	0.1%	0.0%	
BE1	A	0.0%	1.4%	0.0%	0.0%	0.08%
	C	100.0%	95.5%	99.9%	100.0%	
	G	0.0%	2.2%	0.0%	0.0%	
	T	0.0%	0.9%	0.1%	0.0%	
BE2	A	0.0%	0.4%	0.0%	0.0%	0.07%
	C	100.0%	97.7%	99.9%	100.0%	
	G	0.0%	1.1%	0.0%	0.0%	
	T	0.0%	0.7%	0.1%	0.0%	
BE3	A	0.0%	7.1%	0.0%	0.0%	6.05%
	C	99.9%	58.5%	99.7%	100.0%	
	G	0.0%	17.4%	0.0%	0.0%	
	T	0.1%	17.0%	0.2%	0.0%	

HEK4-UDG KO		C_3	C_5	C_8	C_{11}	インデル%
非処置	A	0.0%	0.0%	0.0%	0.0%	0.00%
	C	100.0%	100.0%	99.9%	100.0%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	0.0%	0.0%	0.1%	0.0%	
BE1	A	0.0%	0.0%	0.0%	0.0%	0.00%
	C	99.9%	67.5%	99.0%	100.0%	
	G	0.0%	0.0%	0.1%	0.0%	
	T	0.1%	32.5%	1.0%	0.0%	
BE2	A	0.0%	0.0%	0.0%	0.0%	0.00%
	C	100.0%	91.9%	99.7%	100.0%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	0.0%	8.1%	0.2%	0.0%	
BE3	A	0.0%	0.0%	0.0%	0.0%	0.03%
	C	99.9%	65.9%	99.7%	100.0%	
	G	0.0%	0.3%	0.0%	0.0%	
	T	0.1%	33.8%	0.3%	0.0%	

図 90（ 続き ）

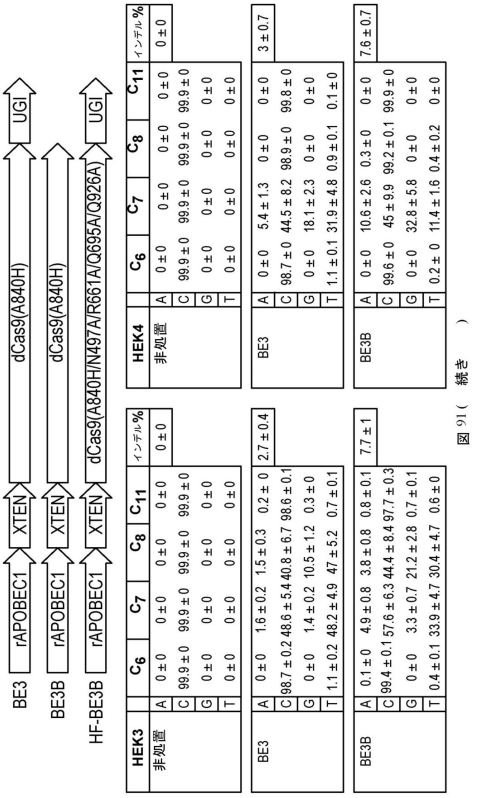
【図９１－１】

EMX1		C_3	C_6	C_{12}	インデル%
非処置	A	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C	99.9 ± 0	99.9 ± 0	99.9 ± 0	
	G	0 ± 0	0 ± 0	0 ± 0	
	T	0 ± 0	0 ± 0	0 ± 0	
BE3	A	1.3 ± 0.1	0.7 ± 0.2	0 ± 0	2.6 ± 0.3
	C	51.7 ± 1	55.2 ± 0.1	99.9 ± 0	
	G	4 ± 0.7	1.9 ± 0.4	0 ± 0	
	T	42.8 ± 0.2	41.9 ± 0.4	0 ± 0	
BE3B	A	4.9 ± 1.4	3.9 ± 1.1	0 ± 0	11.6 ± 2.6
	C	52.6 ± 12.1	61.9 ± 10.5	99.9 ± 0	
	G	12.3 ± 3.3	11.3 ± 3.1	0 ± 0	
	T	30.1 ± 7.3	22.7 ± 6.1	0 ± 0	

FANCF		C_6	C_7	C_8	C_{11}	インデル%
非処置	A	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C	99.9 ± 0	99.9 ± 0	99.9 ± 0	99.9 ± 0	
	G	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
	T	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
BE3	A	0.8 ± 0.1	1.2 ± 0.3	1.2 ± 0.1	0.2 ± 0.1	5.8 ± 0.9
	C	64.7 ± 5.9	67.9 ± 5.1	68.9 ± 5.2	85.4 ± 2.7	
	G	0.6 ± 0	0.7 ± 0	0.7 ± 0.1	0.1 ± 0	
	T	33.8 ± 5.7	29.9 ± 4.8	29 ± 5	14.1 ± 2.6	
BE3B	A	2.9 ± 0.9	4.1 ± 1.4	2.8 ± 0.9	1.1 ± 0	20.5 ± 3.8
	C	67.3 ± 8.8	79.3 ± 5.8	79.5 ± 5.2	92 ± 1.9	
	G	2 ± 0.5	0.7 ± 0.3	0.9 ± 0.2	0.5 ± 0.1	
	T	27.7 ± 7.3	15.6 ± 4	16.5 ± 4	6.1 ± 1.3	

図 91

【図９１－２】



HEK3		C_6	C_7	C_8	C_{11}	インデル%
非処置	A	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C	99.9 ± 0	99.9 ± 0	99.9 ± 0	99.9 ± 0	
	G	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
	T	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
BE3	A	0 ± 0	1.6 ± 0.2	1.5 ± 0.3	0.2 ± 0	2.7 ± 0.4
	C	98.7 ± 0.2	48.6 ± 5.4	40.8 ± 6.7	98.6 ± 0.1	
	G	0 ± 0	1.4 ± 0.2	10.5 ± 1.2	0.3 ± 0	
	T	1.1 ± 0.2	48.2 ± 4.9	47 ± 5.2	0.7 ± 0.1	
BE3B	A	0.1 ± 0	4.9 ± 0.8	3.8 ± 0.8	0.8 ± 0.1	7.7 ± 1
	C	99.4 ± 0.1	57.6 ± 6.3	44.4 ± 8.4	97.7 ± 0.3	
	G	0 ± 0	3.3 ± 0.7	21.2 ± 2.8	0.7 ± 0.1	
	T	0.4 ± 0.1	33.9 ± 4.7	30.4 ± 4.7	0.6 ± 0	

HEK4		C_6	C_7	C_8	C_{11}	インデル%
非処置	A	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C	99.9 ± 0	99.9 ± 0	99.9 ± 0	99.9 ± 0	
	G	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
	T	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
BE3	A	0 ± 0	5.4 ± 1.3	0 ± 0	0 ± 0	3 ± 0.7
	C	98.7 ± 0	44.5 ± 8.2	98.9 ± 0	99.8 ± 0	
	G	0 ± 0	18.1 ± 2.3	0 ± 0	0 ± 0	
	T	1.1 ± 0.1	31.9 ± 4.8	0.9 ± 0.1	0.1 ± 0	
BE3B	A	0 ± 0	10.6 ± 2.6	0.3 ± 0	0 ± 0	7.6 ± 0.7
	C	99.6 ± 0	45 ± 9.9	99.2 ± 0.1	99.9 ± 0	
	G	0 ± 0	32.8 ± 5.8	0 ± 0	0 ± 0	
	T	0.2 ± 0	11.4 ± 1.6	0.4 ± 0.2	0 ± 0	

図 91（続き）

【図９２Ａ－Ｂ】

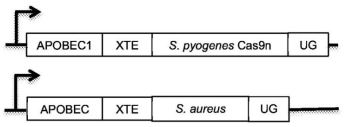


図 92A

種	PAM	塩基編集因子	参照
S. pyogenes	NGG	BE3	野生型
	NGA	VQR, EQR BE3	Ref #7
	NGCG	VRER BE3	Ref #7
S. aureus	NNGRRT	SaBE3	野生型
	NNNRRT	SaKKHBE3	Ref #8

図 92B

【図９２Ｃ】

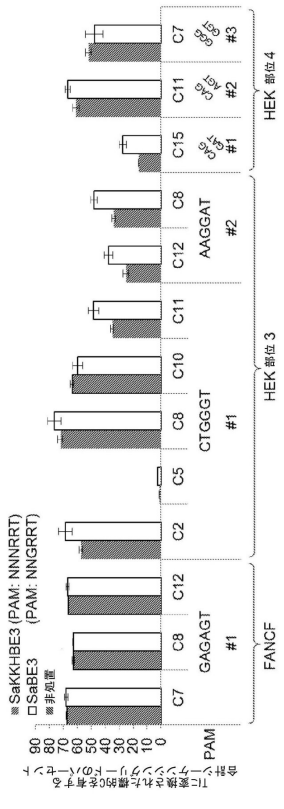


図 92C

【図９２Ｄ】

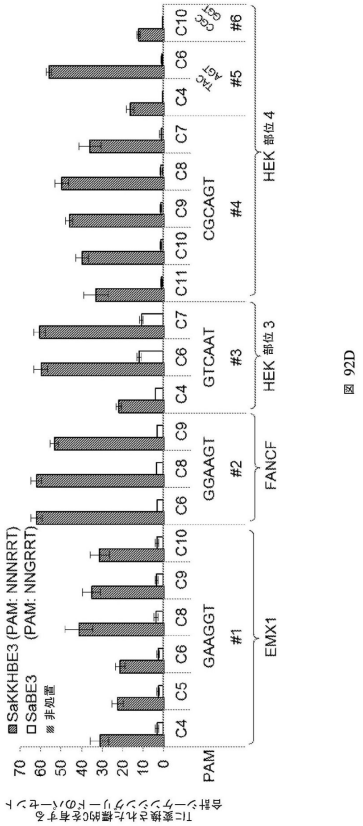


図 92D

【図９２Ｅ－Ｆ】

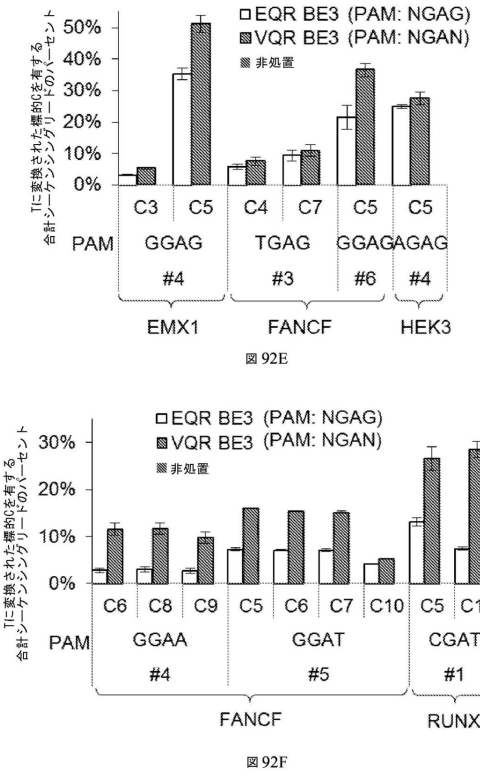


図 92E

図 92F

【図９２Ｇ】

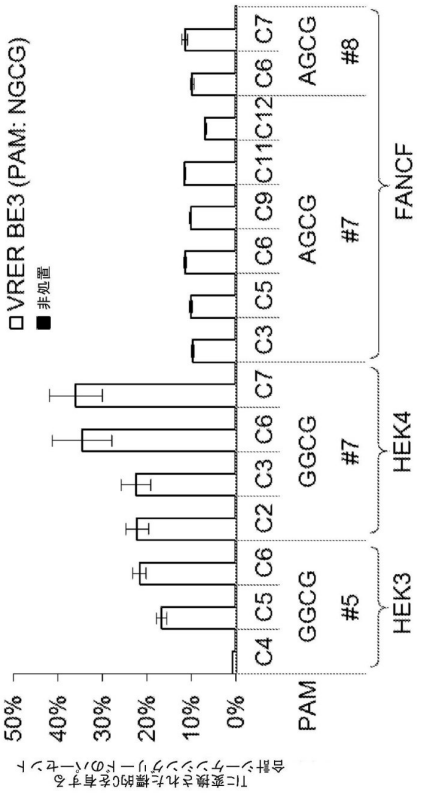


図 92G

【図９３Ａ】

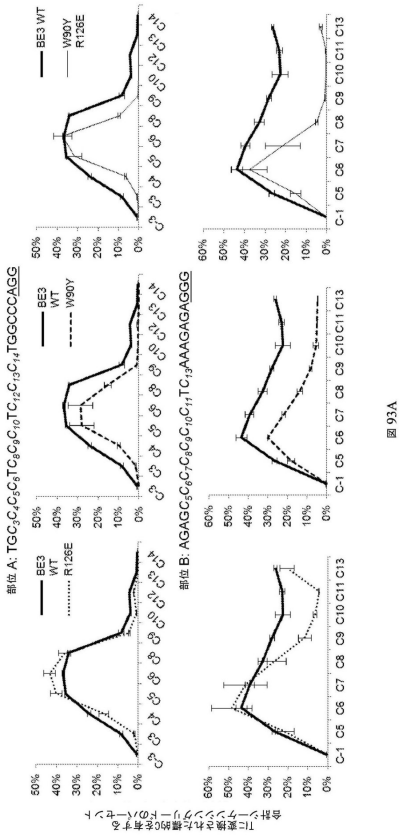


図 93A

【図９３Ｂ】

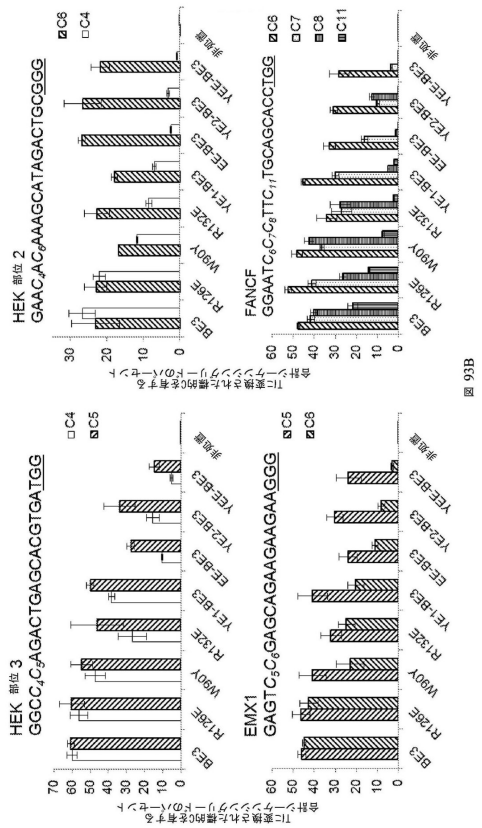


図 93B

【図９３Ｃ】

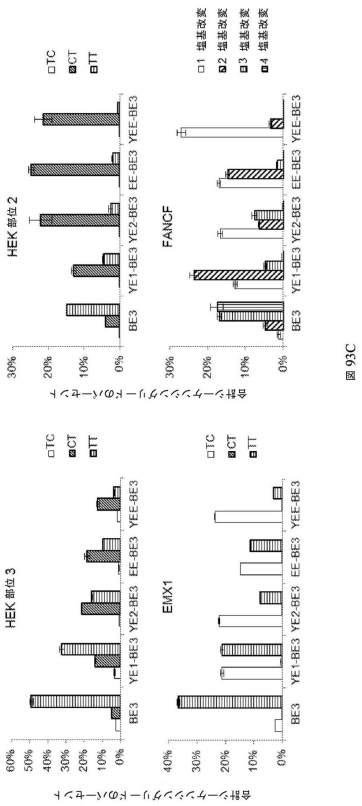


図 93C

【図９４】

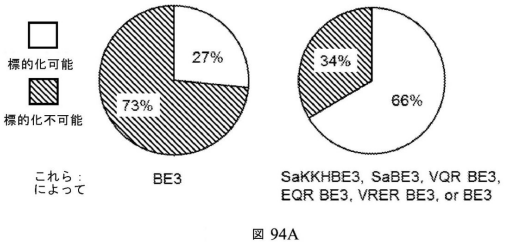


図 94A

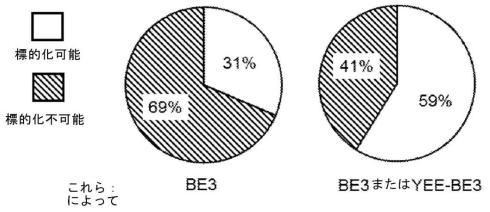


図 94B

【図９５Ａ】

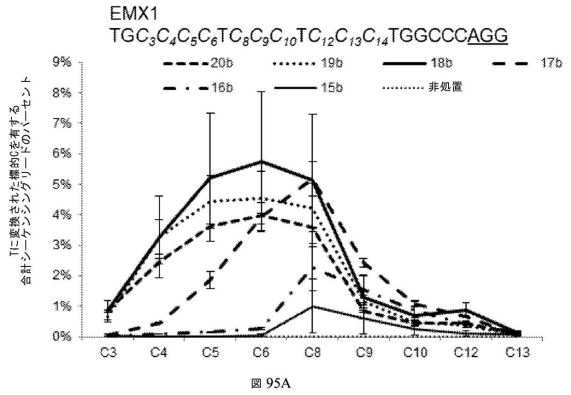


図 95A

【図９５Ｂ】

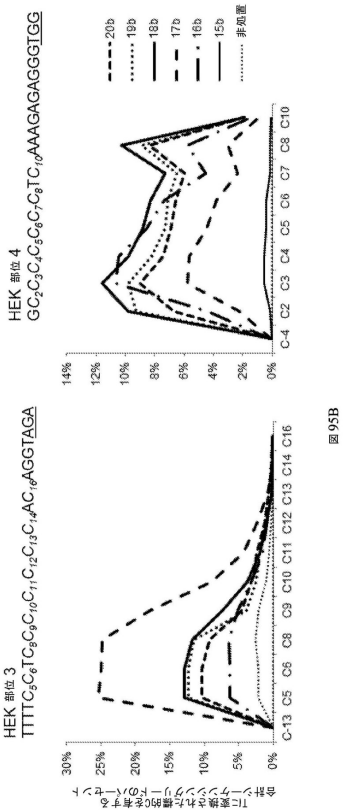


図 95B

【図９６】

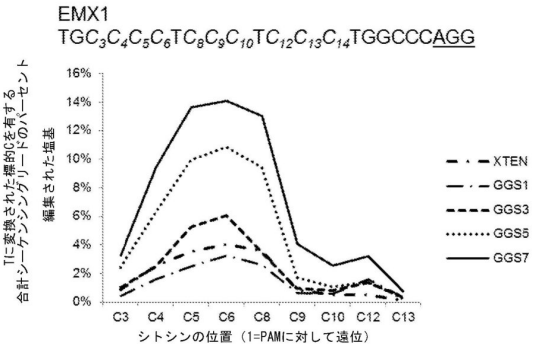


図96

【図９７Ａ－Ｂ】

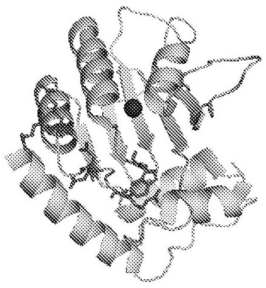


図 97A

APOBEC1 変異	APOBEC3G 変異	参照
R126A	R320A	#9,10
R126E	R320E	#9,10
W90A	W285A	#9,10
W90Y	W285Y	本研究
R132E	R326E	本研究

図 97B

【図９７Ｃ－１】

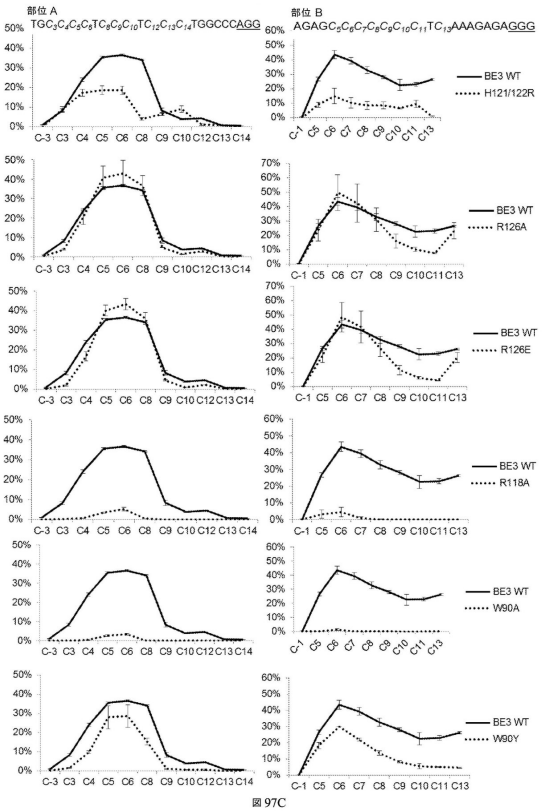
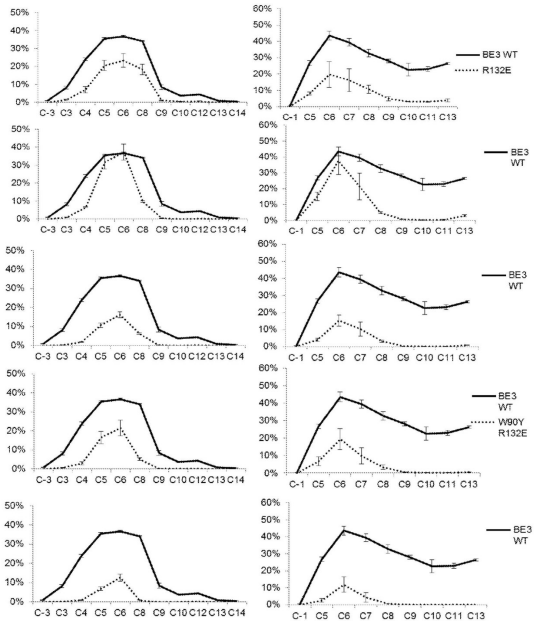


図 97C

【図９７Ｃ－２】



編集ウィンドウ幅

	部位 A	部位 B		部位 A	部位 B
BE3 WT	5.0	6.1	W90Y	3.8	4.9
H121/122R	4.2	7.4	R132E	4.0	3.0
R126A	4.4	3.4	W90Y R126E	2.9	3.0
R126E	4.2	3.1	R126E R132E	2.9	3.0
R118A	2.4	3.6	W90Y R132E	2.7	2.8
W90A	2.5	1.1	W90Y R126E R132E	2.1	1.4

図 97C（続き）

【図９８】

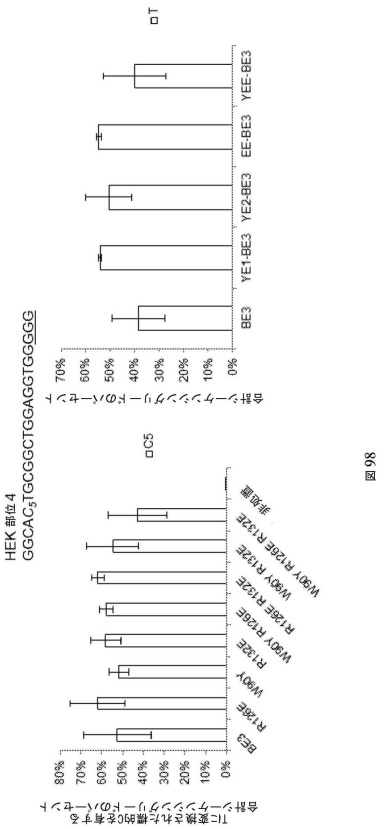


図 98

【図９９】

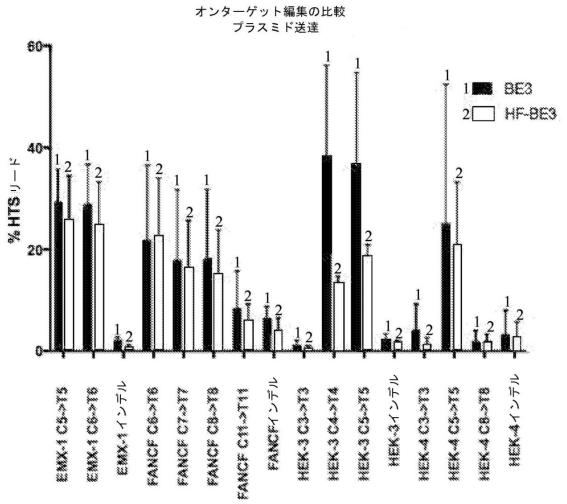


図 99

【図１００】

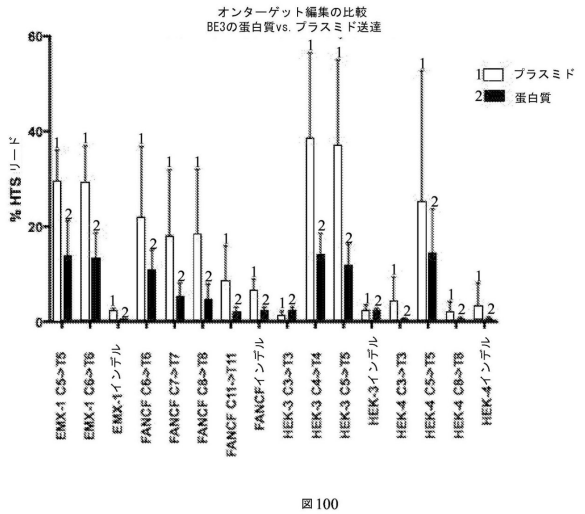


図 100

【図１０１】

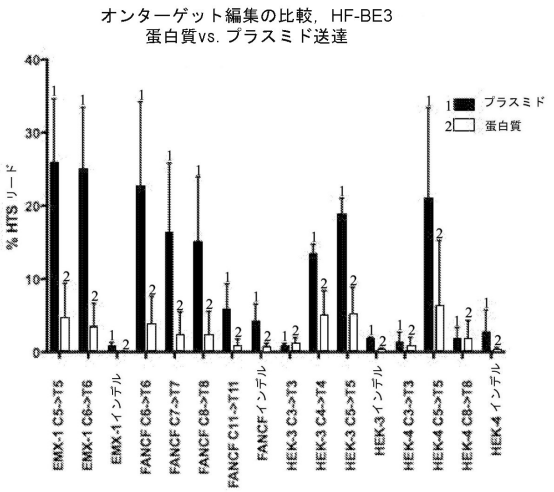


図 101

【図１０２】

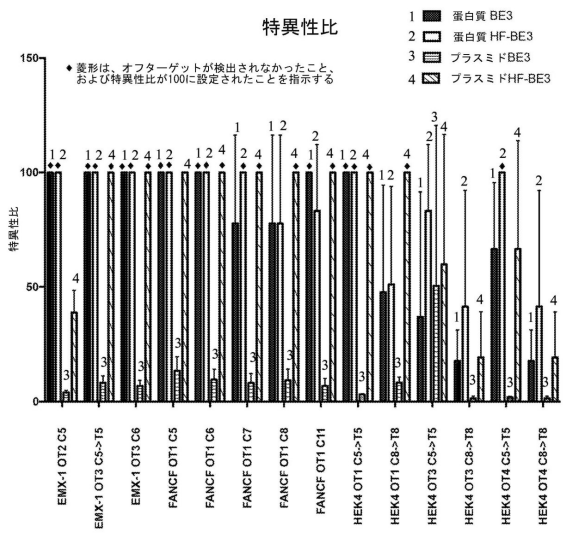


図 102

【図１０３】

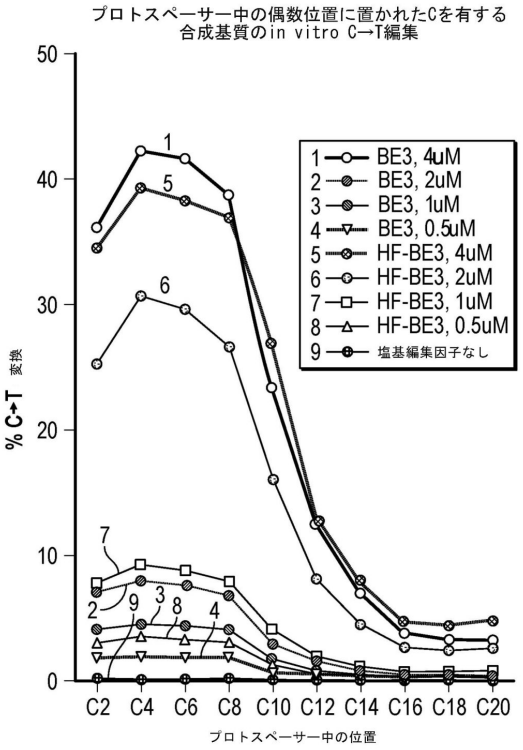


図 103

【図１０４】

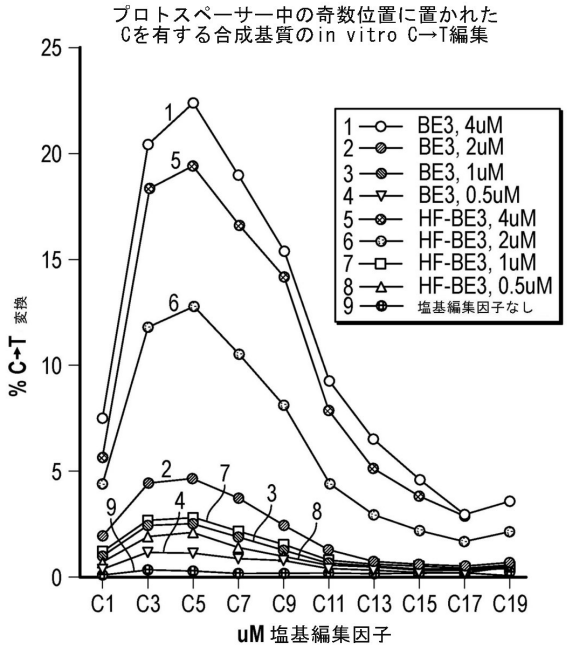


図 104

【図１０５】

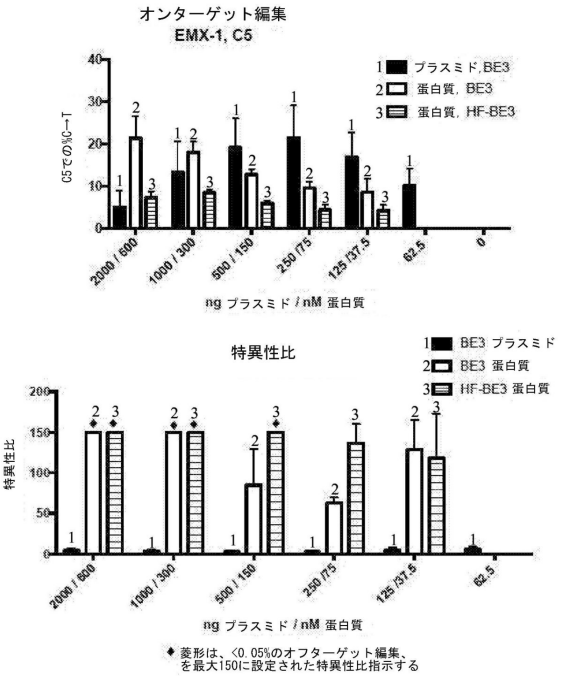


図 105

【図１０６Ａ】

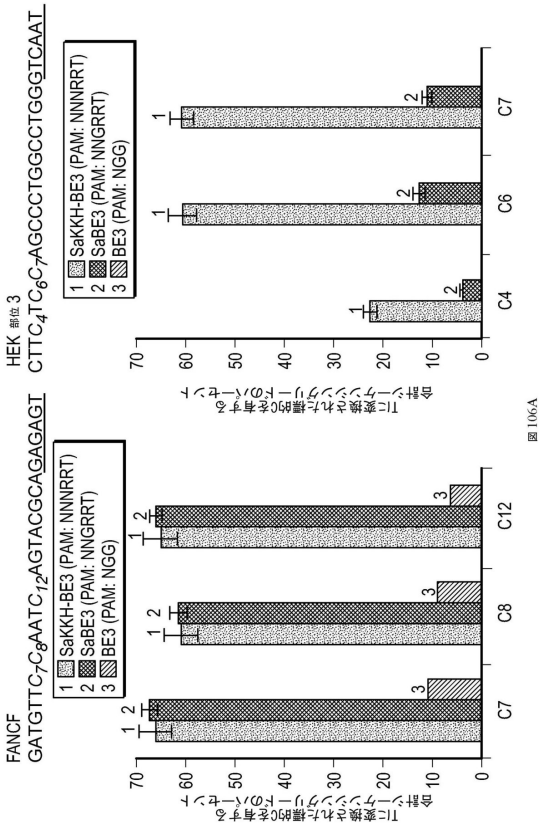


図 106A

【図１０６Ｂ】

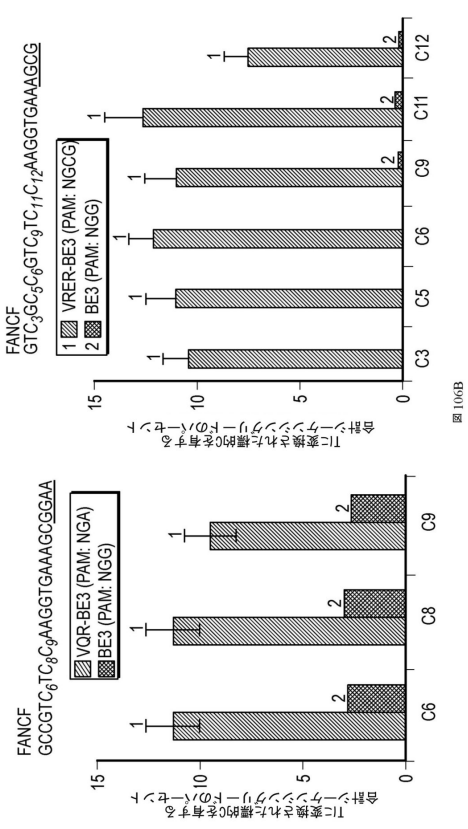


図 106B

【図１０７Ａ】

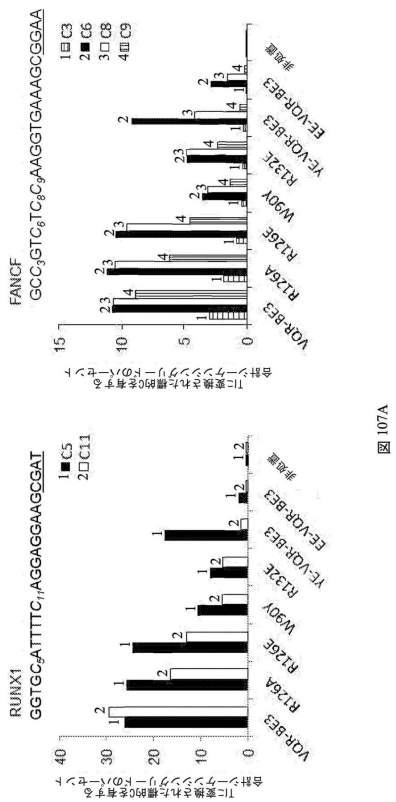


図 107A

【図１０７Ｂ】

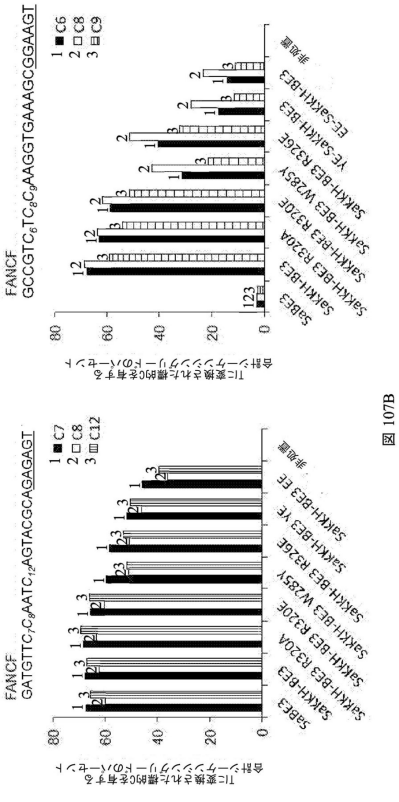


図 107B

【図１０８】

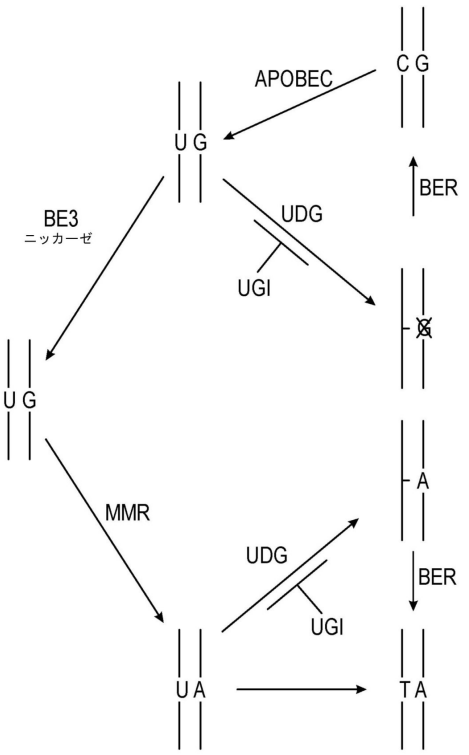


図 108

【図１０９】

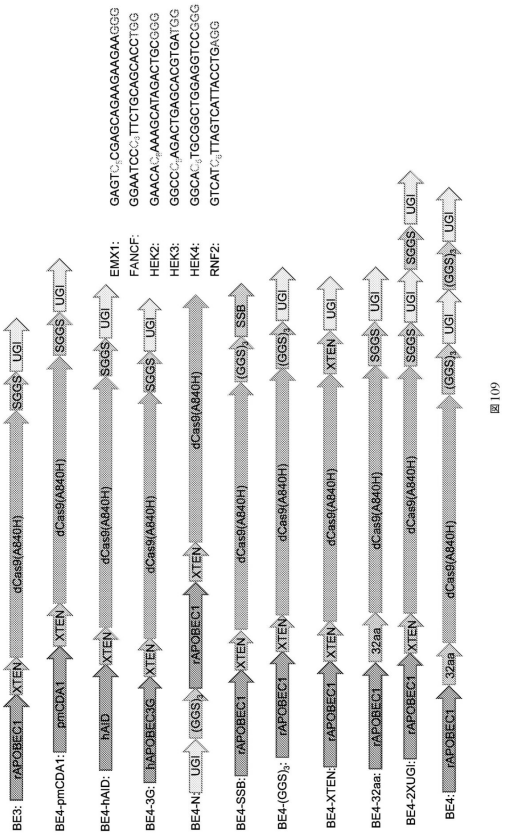


図 109

【図１１０】

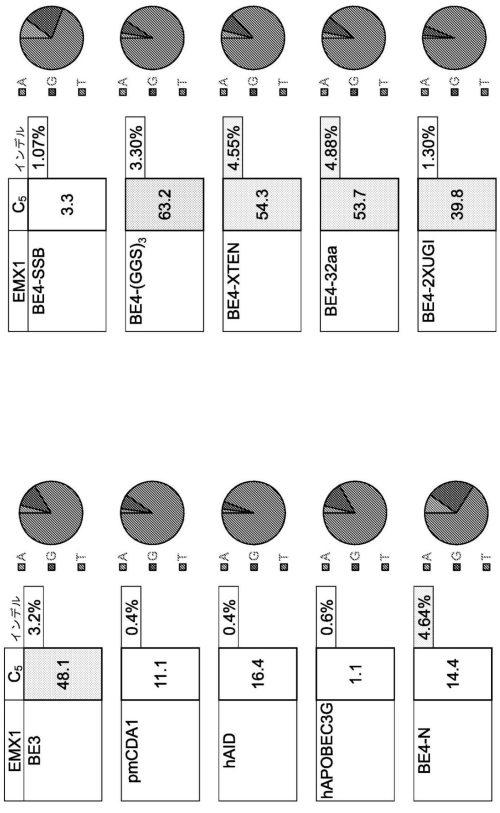


図 110

【図１１１】

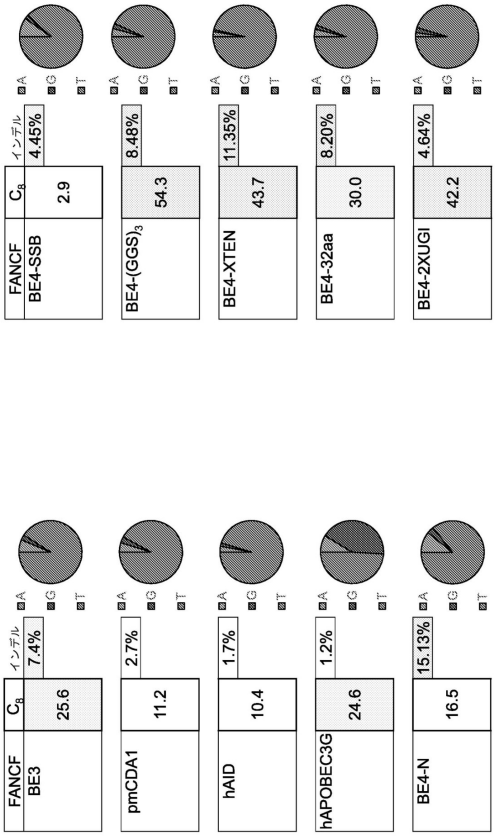


図 111

【図１１２】

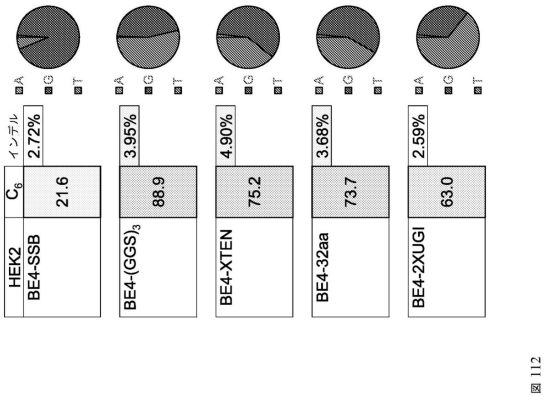


図 112

【図１１３】

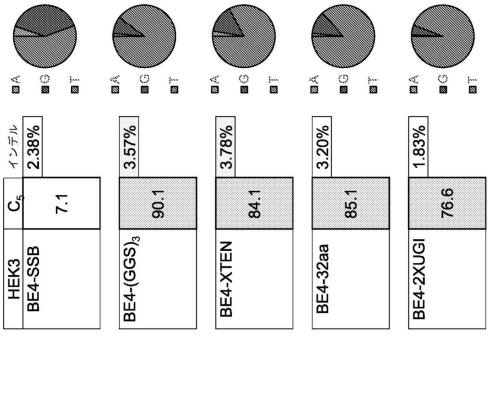
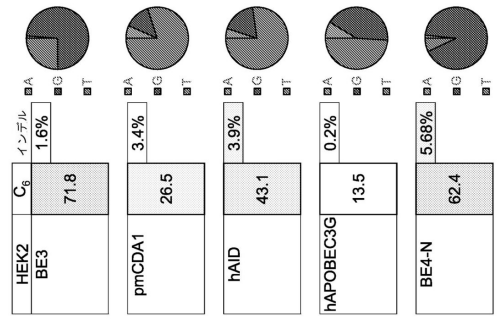
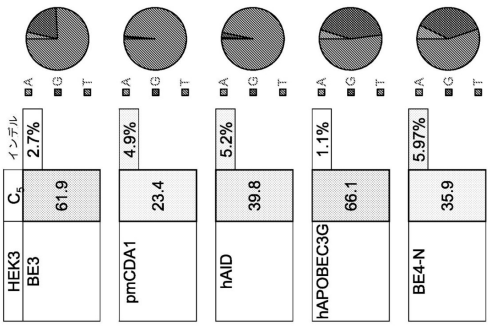


図 113





【図１１４】

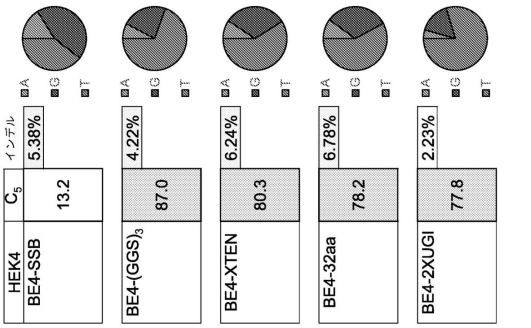


図 114

【図１１５】

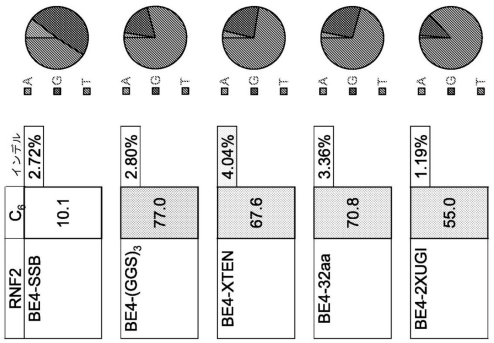
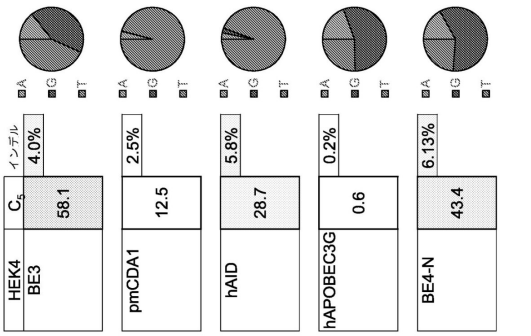
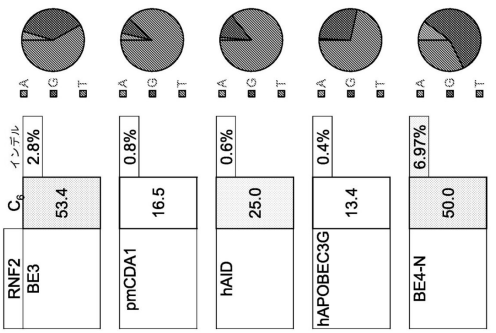
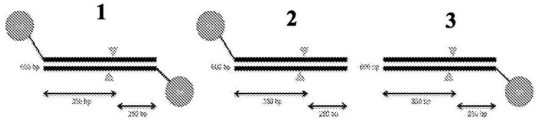


図 115





【図１１６】



上の鎖：非標的鎖；下の鎖：標的鎖

図 116

【図１１７】

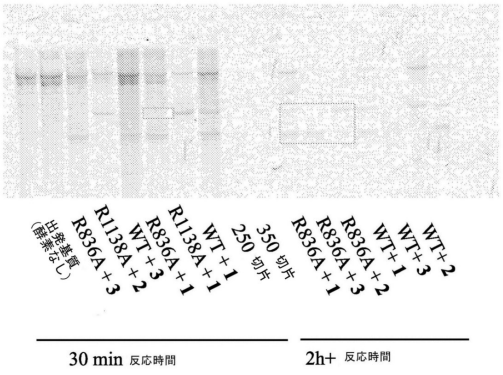


図 117

【図１１８】

PAM

FANCF3	TTC	G1	C2	G3	G4	G7	C10	C11	C15	G17	C20	C22										
BG		0.01005	0.03522	0.01383	0.00933	0.00502	0.01872	0.02614	0.03740	0.10504	0.01318	0.02099										
AsCpf1		0.00895	0.02769	0.07439	0.00672	0.06495	0.02395	0.01795	0.02691	0.01570	0.02237	0.02456										
APOBEC-AsCpf1(R912A)-UGI		0.00883	0.02606	0.01430	0.00539	0.00490	0.29513	0.28158	0.04373	0.01425	0.02795	0.01631										
APOBEC-AsCpf1(R1225A)-UGI		0.00796	0.03754	0.01131	0.00749	0.01124	0.02769	0.02908	0.03846	0.02016	0.01732	0.01651										
FANCF2	TTC	C1	G2	G3	C4	C7	G9	G10	C11	G12	G13	C15	C16	G17	C18							
BG		0.02628	0.01963	0.02346	0.00318	0.00738	0.00544	0.00850	0.08758	0.01268	0.00637	0.00499	0.07694	0.01182	0.01805							
AsCpf1		0.03489	0.03325	0.05248	0.00392	0.00742	0.02484	0.00751	0.06139	0.16805	0.01851	0.02139	0.10988	0.00927	0.00992							
APOBEC-AsCpf1(R912A)-UGI		0.02440	0.02677	0.00699	0.00481	0.19389	0.01115	0.00534	0.11463	0.01660	0.00427	0.00722	0.19175	0.01119	0.01065							
APOBEC-AsCpf1(R1225A)-UGI		0.02855	0.04216	0.00322	0.00479	0.03587	0.01004	0.00585	0.06624	0.00386	0.00638	0.00444	0.07656	0.00797	0.01061							
HEK3	TTC	C3	C5	C6	C9	C12	C13	C14	C18	C19												
BG		0.00907	0.00726	0.01446	0.01538	0.01084	0.01281	0.00729	0.01540	0.10834												
AsCpf1		0.01118	0.00670	0.01594	0.00892	0.01334	0.01342	0.01797	0.01118	0.03951												
APOBEC-AsCpf1(R912A)-UGI		0.00647	0.02464	0.44917	0.30191	0.03754	0.12192	0.01533	0.01251	0.00754												
APOBEC-AsCpf1(R1225A)-UGI		0.00770	0.00768	0.10243	0.07236	0.01202	0.00770	0.00551	0.01965	0.00549												
HEK4		G21	G20	G18	G16	C15	C13	G12	G10	G7	G6	C5	G3	GAAA								
BG		0.00638	0.01001	0.01459	0.01938	0.01008	0.01285	0.01284	0.00548	0.01458	0.02557	0.01089	0.02636									
AsCpf1		0.00988	0.00862	0.02466	0.01984	0.03343	0.01600	0.00865	0.01600	0.01718	0.02216	0.01252	0.00488									
APOBEC-AsCpf1(R912A)-UGI		0.00323	0.01185	0.03012	0.00759	0.01735	0.01734	0.04550	0.15878	0.00862	0.02266	0.00542	0.00644									
APOBEC-AsCpf1(R1225A)-UGI		0.01142	0.01045	0.03044	0.02435	0.01626	0.02104	0.01051	0.01238	0.01045	0.03428	0.01810	0.00473									

BG～バックグラウンド；R912A－BE3；R1225A－自滅型 BE（非標的鎖にニックを入れる ）

図 118

【図１１９－１２１】

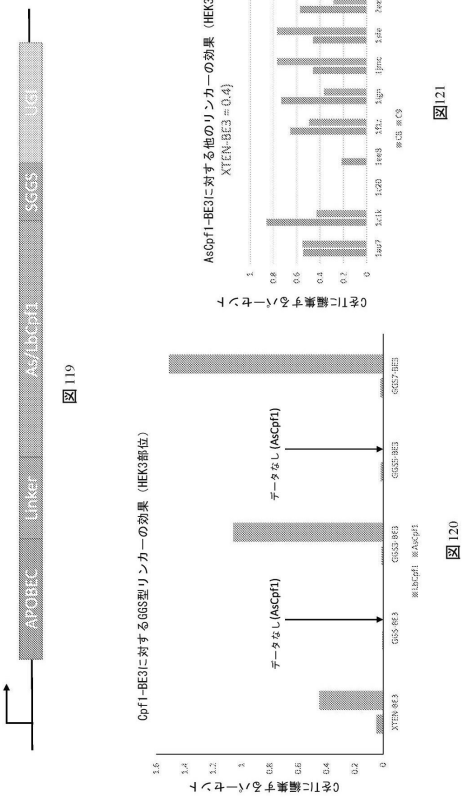


図 119

図 120

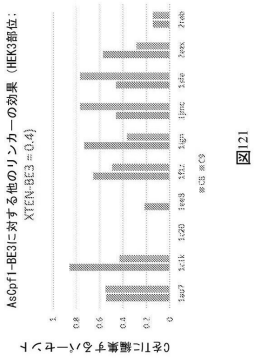


図 121

【図１２２】

図 122

【図１２３】

図 123

【図１２２－１２３】

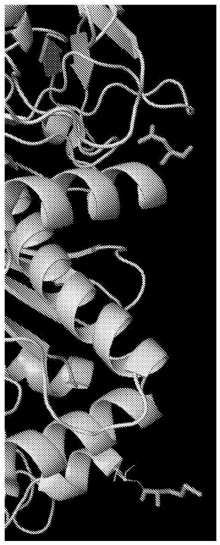


図122



図123

【図１２４】



図124

【図１２５】

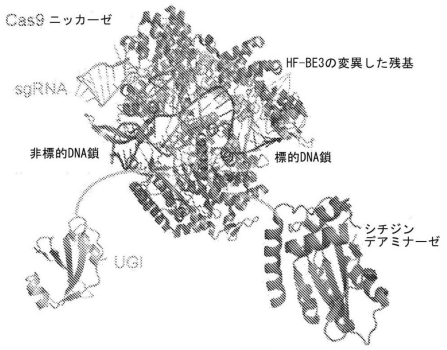


図125A

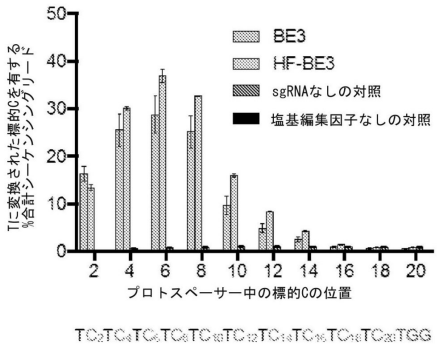
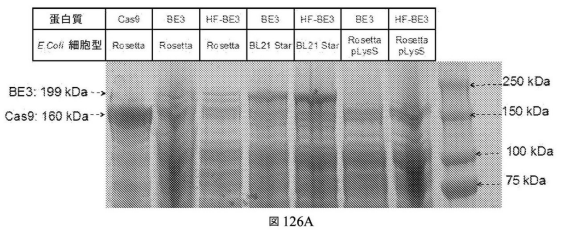


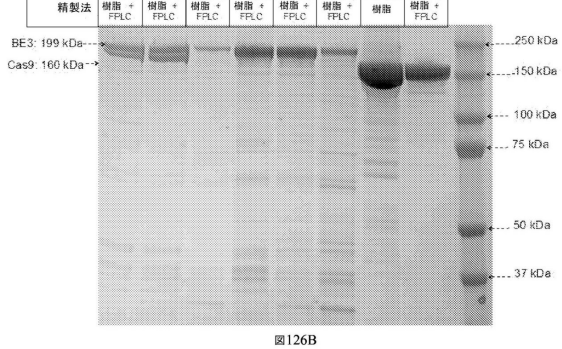
図125B

【図１２６Ａ－Ｂ】



蛋白質	Cas9	BE3	HF-BE3	BE3	HF-BE3	BE3	HF-BE3
E.Coli 細胞型	Rosetta	Rosetta	Rosetta	BL21 Star	BL21 Star	Rosetta pLysS	Rosetta pLysS

図126A



蛋白質	HF-BE3	BE3	HF-BE3	BE3	HF-BE3	BE3	Cas9	Cas9
His_6タグ	N 末	N 末	C 末	C 末	C 末	C 末	N 末	N
誘導 OD_{600}	0.7	0.7	0.4	0.7	0.7	0.4	0.4	0.4
精製法	ニッケル樹脂 + FPLC	ニッケル樹脂 + FPLC	ニッケル樹脂 + FPLC	ニッケル樹脂 + FPLC	ニッケル樹脂 + FPLC	ニッケル樹脂 + FPLC	ニッケル樹脂	ニッケル樹脂 + FPLC

図126B

【図１２６Ｃ】

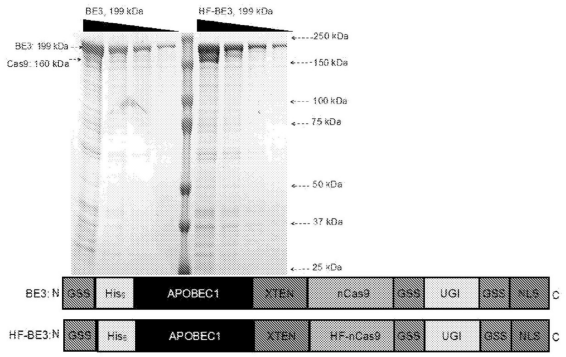
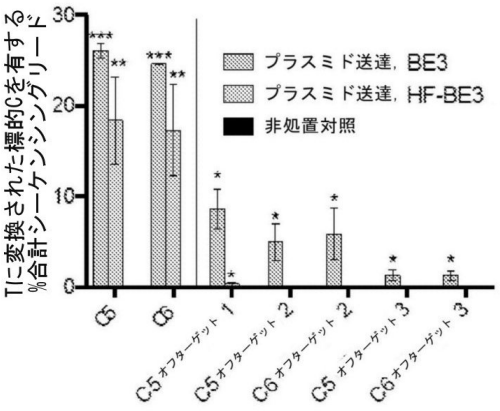


図 126C

【図１２７Ａ】

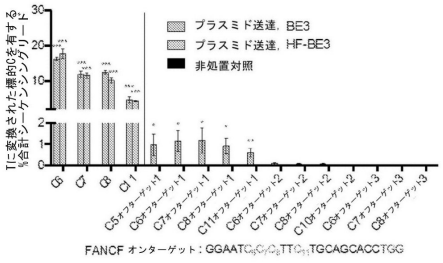


EMX1 オンターゲット：GAGTCCGAGCAGAAGAAGAAGGG

EMX1 オフターゲット1: GAGTCCTAAGCAGAAGAAGAAGAG
EMX1 オフターゲット2: GAGGCCGAGCAGAAGAAAGACGG
EMX1 オフターゲット3: GAGTCCTAGCAGGAGAAGAAGAG

図 127A

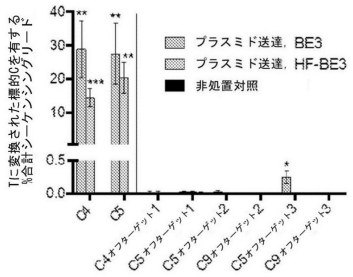
【図１２７Ｂ－Ｃ】



FANCF オンターゲット：GGAATCCCTTCGTCAGCACCTGG

FANCF オフターゲット1: GGAACCCCTTCGTCTGCAGCACCAGG
FANCF オフターゲット2: GGAGTCCCTTCGTCTACAGCACCAGG
FANCF オフターゲット3: AGAGGCCCTTCGTCTGCAGCACCAGG

図 127B

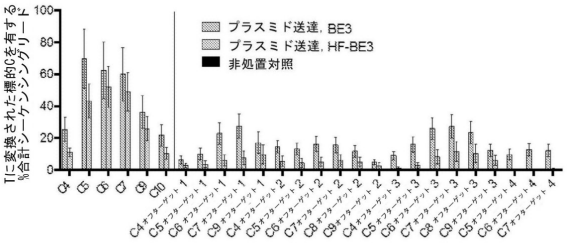


HEK293 部位 3 オンターゲット：GCCCCAGACTGAGCACGTGATGG

HEK293 部位 3 オフターゲット1: CACCCAGACTGAGCACGTGCTGG
HEK293 部位 3 オフターゲット2: GACACCAGACTGGGCACGTGAGGG
HEK293 部位 3 オフターゲット3: AGCTCCAGACTGAGCAAGTGAGGG

図 127C

【図１２７Ｄ】



VEGFA部位2 オンターゲット：GACCCCCTCCTCAGCCCGCCTCCGG

VEGFA部位2 オフターゲット 1: CTACCCCTCCTCTCCACCCCGCCTCCGG
VEGFA部位2 オフターゲット 2: ATTCCCCTCCTCCCCACCCCGCCTCAGG
VEGFA部位2 オフターゲット 3: ACACCCCTCCTCCCCACCCCGCCTCAGG
VEGFA部位2 オフターゲット 4: TGCCCCTCCTCCCCCACCCCACCTCTGG

図 127D

【図１２８Ａ】

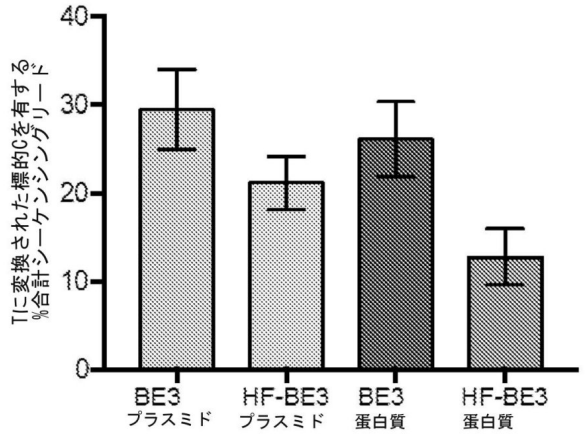
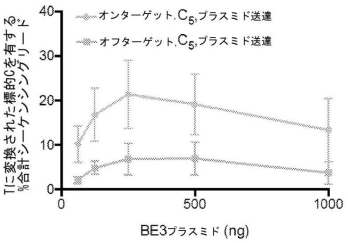


図 128A

【図１２８Ｂ－Ｃ】



EMX1: GAGTC5C6GAGCAGAAGAAGAAGGG

EMX1 オフターゲット 1: GAGTC5TAAGCAGAAGAAGAAGAG

図 128B

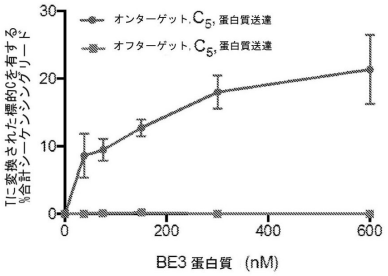


図 128C

【図１２９】

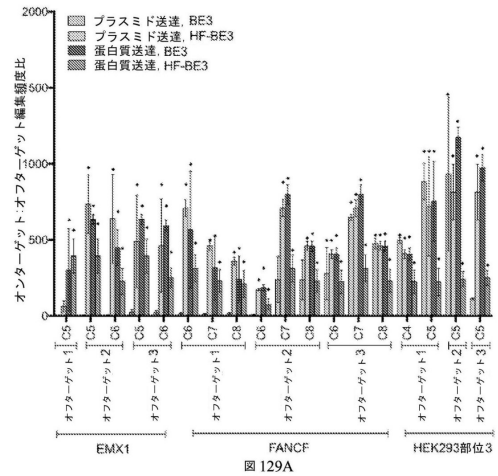


図 129A

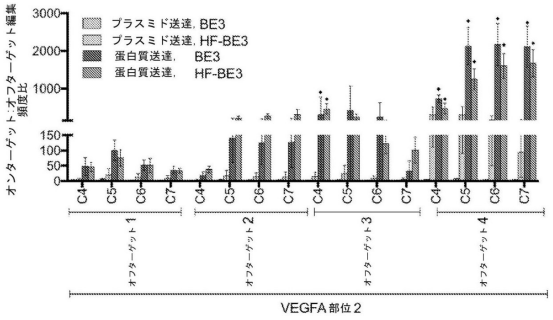
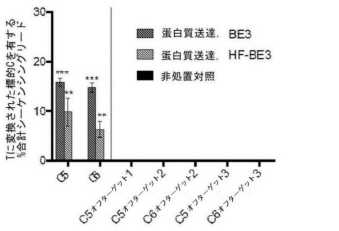


図 129B

【図１３０Ａ－Ｂ】



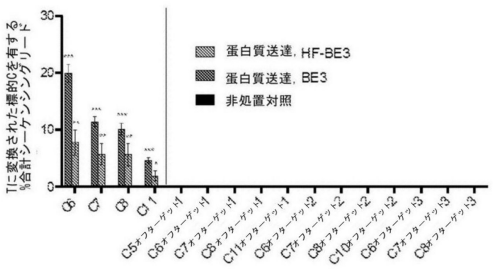
EMX1オンターゲット: GAGTC5C6GAGCAGAAGAAGAAGGG

EMX1オフターゲット 1: GAGTC5TAAGCAGAAGAAGAAGAG

EMX1オフターゲット 2: GAGGC5C6GAGCAGAAGAAGAACGG

EMX1オフターゲット 3: GAGTC5C6TAGCAGGAGAAGAAGAG

図 130A



FANCFオンターゲット: GGAATC5C6C7C8TTC11TGCAGCACCTGG

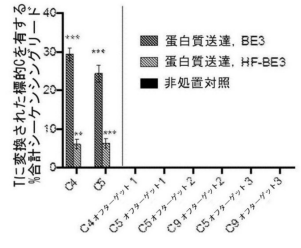
FANCFオフターゲット1: GGAAC5C6C7C8GTC11TGCAGCACCAGG

FANCFオフターゲット2: GGAGTC5C6C7C8TC10C11TACAGCACCAGG

FANCFオフターゲット3: AGAGGC5C6C7C8C9TC11TGCAGCACCAGG

図 130B

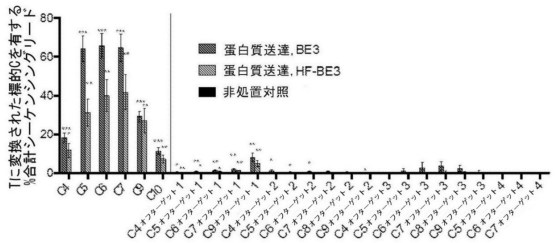
【図１３０Ｃ－Ｄ】



HEK293部位３ オンターゲット: GCCCCCCAGACTGAGCACGTGATGG

HEK293 部位３ オフターゲット１: CACCCCAGACTGAGCACGTGCTGG
HEK293 部位３ オフターゲット２: GACACCAGACTGGGCACGTGAGGG
HEK293部位３ オフターゲット３: AGCTCCAGACTGAGCAAGTGAGGG

図130C

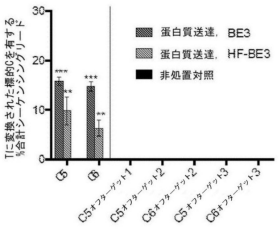


VEGFA部位２ オンターゲット : GACCCCCTCCCCTCCCCGCCTCCGG

VEGFA部位２ オフターゲット １: CTACCCCTCCCCTCCCCGCCTCCGG
VEGFA部位２ オフターゲット ２: ATTCCCCCCCCCCCCCCCCCCTCAGG
VEGFA部位２ オフターゲット ３: ACACCCCCCCCCCCCCCCCCGCCTCAGG
VEGFA部位２ オフターゲット ４: TGCCCCCCCCCCCCCCCCCCACCCTCTGG

図130D

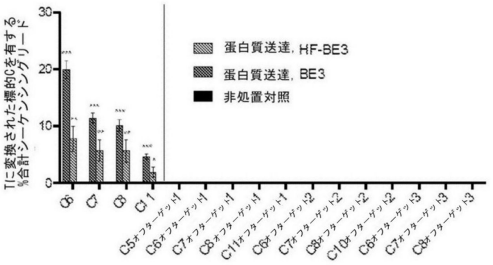
【図１３１Ａ－Ｂ】



EMX1オンターゲット: GAGTCCGAGCAGAAGAAGAAGGG

EMX1オフターゲット 1: GAGTCTAAGCAGAAGAAGAAGAG
EMX1オフターゲット 2: GAGGCCGAGCAGAAGAAAGACGG
EMX1オフターゲット 3: GAGTCCTAGCAGGAGAAGAAGAG

図 130A



FANCFオンターゲット: GGAATCCCCTTCTGCAGCACCTGG

FANCFオフターゲット1: GGAACCCCGTCTGCAGCACCAGG
FANCFオフターゲット2: GGAGTCCCCTCCTACAGCACCAGG
FANCFオフターゲット3: AGAGGCCCCCCTCTGCAGCACCAGG

図 130B

【図　　】

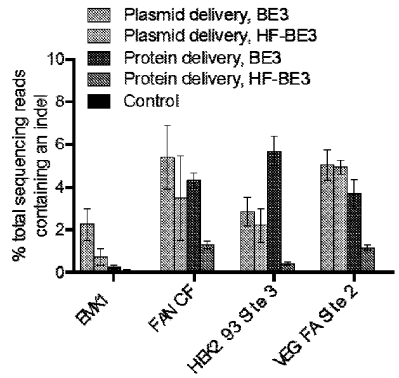


FIGURE 131A

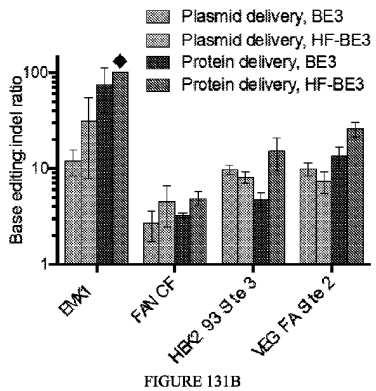


FIGURE 131B

【図１３１Ｃ】

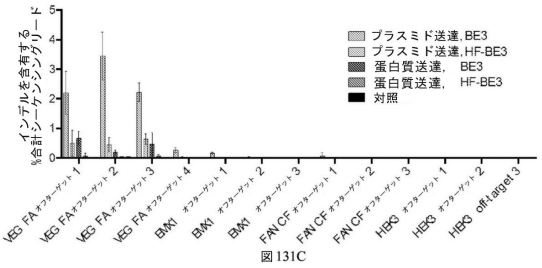


図131C

【図１３２Ａ】

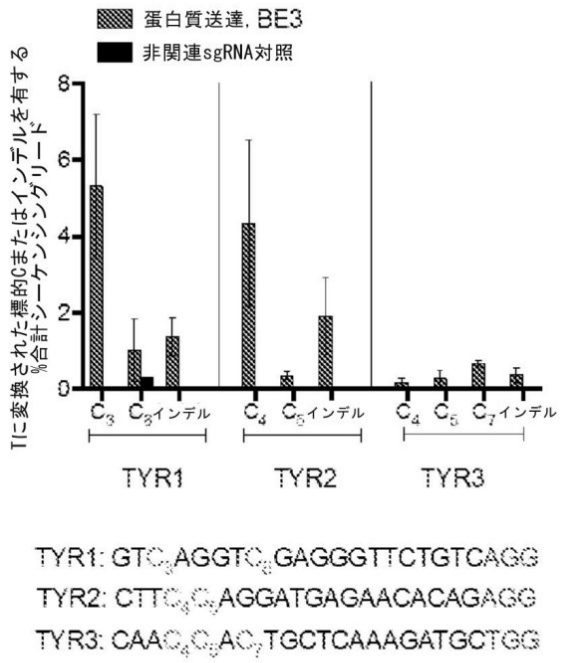


図 132A

【図１３２Ｂ－Ｃ】

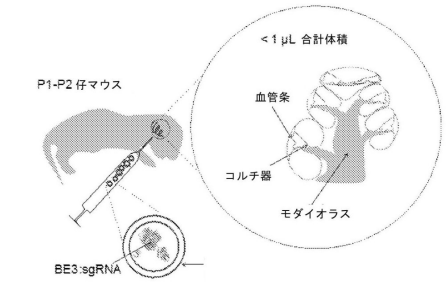
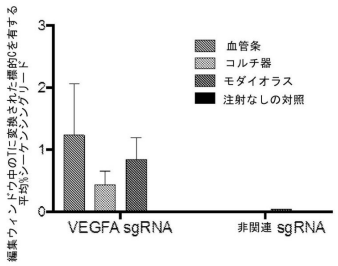


図 132B



VEGFA 部位 2: GACC4C5C6C7TC9C10ACCCCGCCTCCGG

図 132C

【図１３２Ｄ】

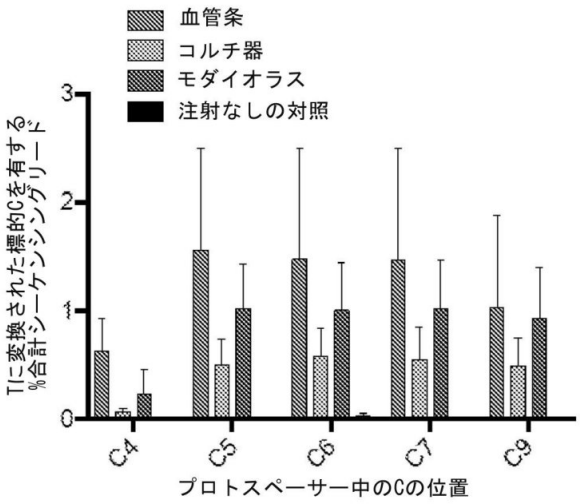


図 132D

【図１３３Ａ】

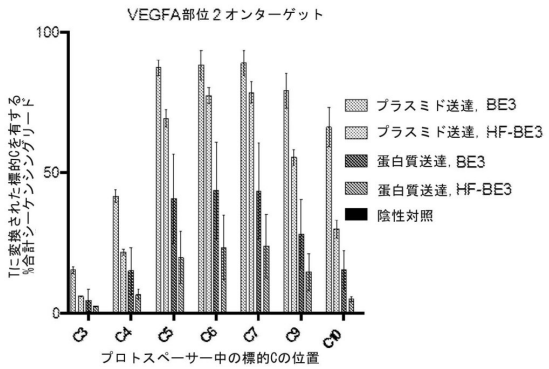


図 133A

【図１３３Ｂ－Ｃ】

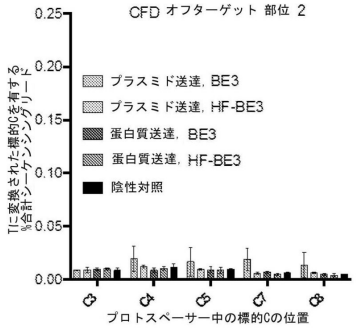


図 133B

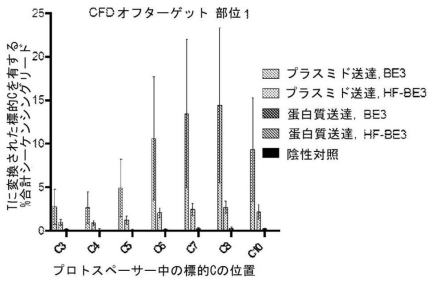


図 133C

【図１３３Ｄ－Ｅ】

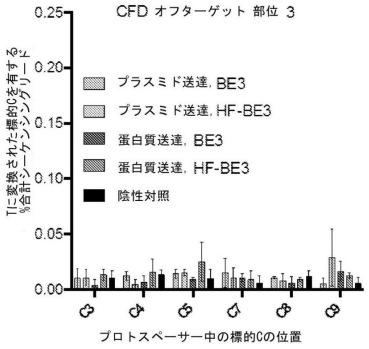


図 133D

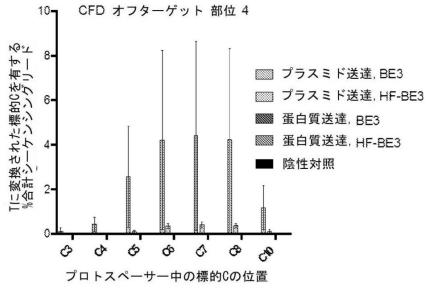


図 133E

【図１３４】

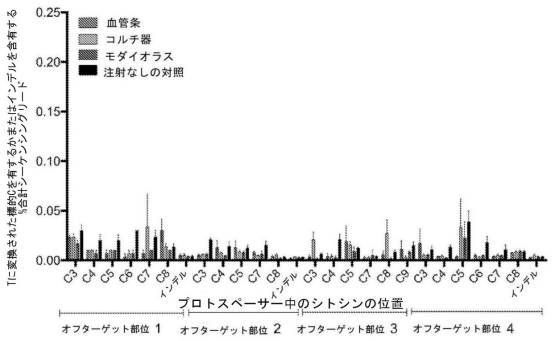


図 134A

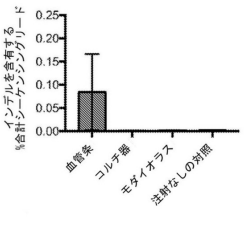


図 134B

【図１３５】

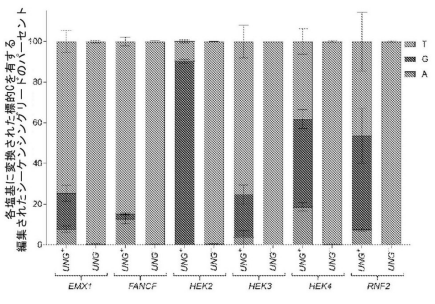


図135A

EMX1: GAGTC$_5$CGAGCAGAAGAAGAAGGG
FANCF: GGAATCC$_7$CTTCTGCAGCACCTGG
HEK2: GAACAC$_8$AAAGCATAGACTGCGGG
HEK3: GGCCC$_8$AGACTGAGCACGTGATGG
HEK4: GGCAC$_5$TGCGGCTGGAGGTCCGGG
RNF2: GTCATC$_8$TTAGTCATTACCTGAGG

図 135B

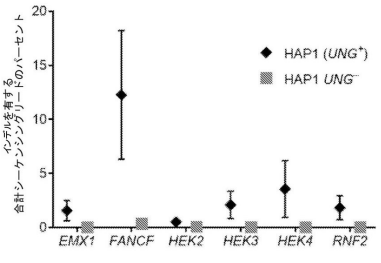


図 135C

【図１３６Ａ】

HEK4		G	G	C	A	C	T	G	G	C	G	G	A	G	T	G	G	G
非処置	A	0.0	0.1	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	100.0	0.0	100.0	0.0	0.0	99.9	0.0	0.0	100.0	0.1	0.0	0.2	0.0	0.0	0.0
	G	99.9	99.9	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	100.0	100.0	99.9	99.9	99.9
	T	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.5	0.1	0.0	0.0
BE3	A	0.1	0.0	0.1	99.9	8.4	0.1	0.0	0.1	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	99.0	0.0	41.9	0.0	0.0	99.3	0.0	0.0	99.8	0.1	0.0	0.0	0.0	0.0	0.0
	G	99.9	99.9	0.0	0.0	24.5	0.0	100.0	0.0	100.0	100.0	0.0	0.1	100.0	100.0	100.0	100.0	100.0
	T	0.0	0.0	0.9	0.0	25.2	99.9	0.0	0.5	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0
AID-BE3	A	0.3	0.4	0.3	100.0	0.9	0.0	0.0	0.9	0.0	0.0	0.1	0.0	0.0	99.8	0.0	0.0	0.0
	C	0.0	0.0	81.0	0.0	71.3	0.0	0.0	80.7	0.0	0.0	97.9	0.0	0.0	0.2	0.0	0.0	0.0
	G	99.7	99.6	0.1	0.0	0.7	0.0	100.0	0.1	100.0	100.0	0.0	0.0	99.3	100.0	100.0	100.0	99.9
	T	0.0	0.0	18.0	0.0	27.1	100.0	0.0	18.3	0.0	0.0	2.0	0.0	0.0	99.8	0.0	0.0	0.1

図136A

【図１３６Ｂ－Ｄ】

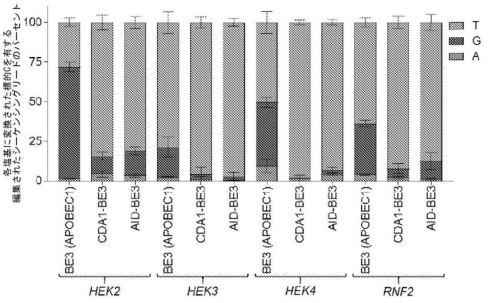


図136B

HEK2: GAACAC$_6$AAAGCATAGACTGCGGG
HEK3: GGCCC$_5$AGACTGAGCACGTGATGG
HEK4: GGCAC$_5$TGCGGCTGGAGGTCCGGG
RNF2: GTCATC$_6$TTAGTCATTACCTGAGG

図136C

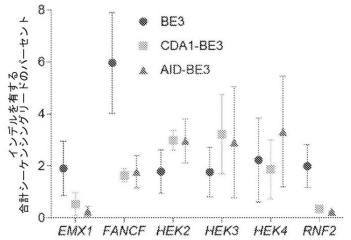


図136D

【図１３７Ａ－Ｂ】

HEK2: GAAC$_4$AC$_6$AAAGCATAGACTGCGGG
HEK3: GGCC$_4$C$_5$AGACTGAGCACGTGATGG
HEK4: GGCAC$_5$TGCGGCTGGAGGTCCGGG
RNF2: GTC$_3$ATC$_6$TTAGTCATTACCTGAGG

図137A

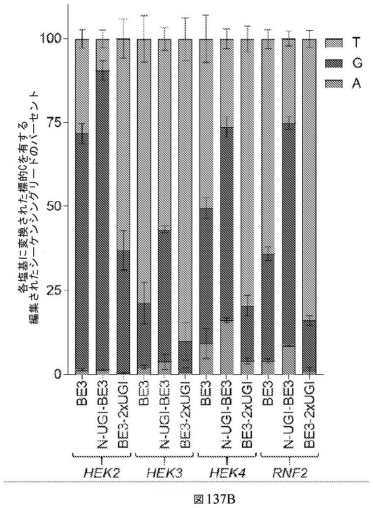


図137B

【図１３７Ｃ】

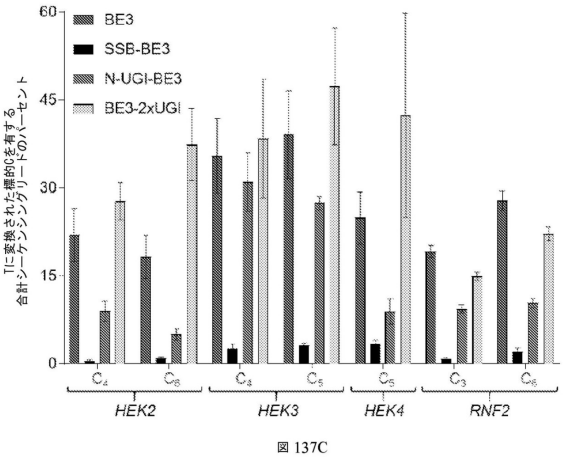


図137C

【図１３８Ａ－Ｃ】

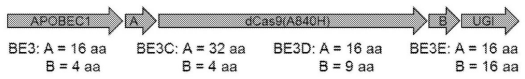


図 138A



図 138B

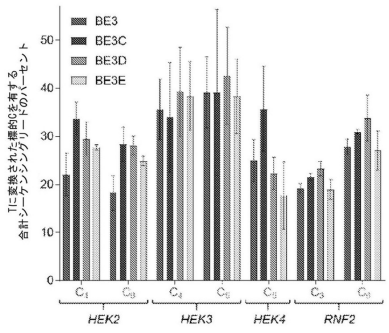


図 138C

【図１３８Ｄ】

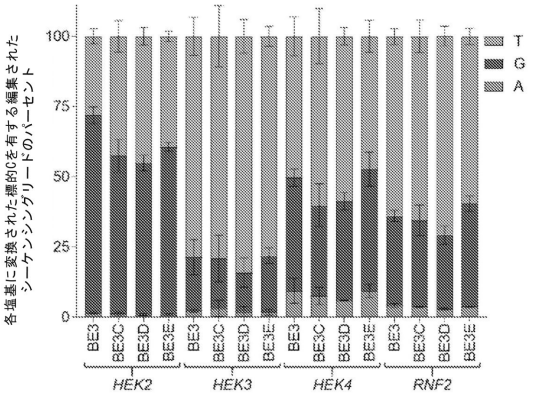


図１３８Ｄ

【図１３９Ａ－Ｃ】

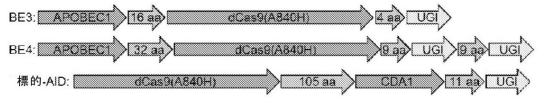


図１３９Ａ



図１３９Ｂ

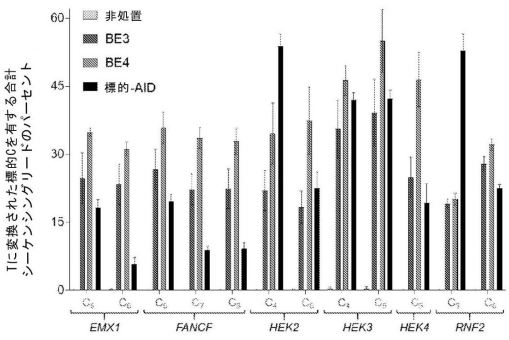


図１３９Ｃ

【図１３９Ｄ】

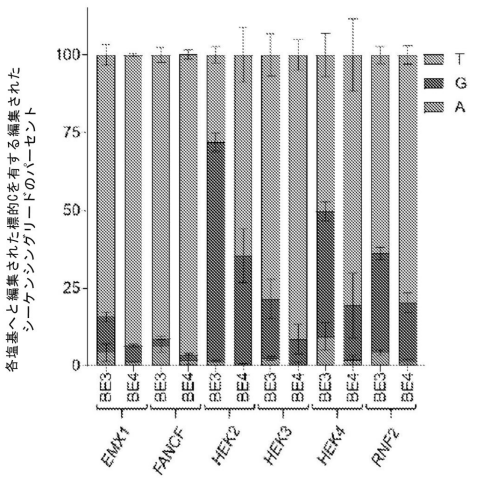


図１３９Ｄ

【図１４０】

EMX1: GAGTC$_5$C$_6$GAGCAGAAGAAGAAGGG

FANCF: GGAATC$_6$C$_7$C$_8$TTCTGCAGCACCTGG

HEK2: GAAC$_3$AC$_6$AAAGCATAGACTGCGGG

HEK3: GGC$_3$C$_4$C$_5$AGACTGAGCACGTGATGG

HEK4: GGC$_3$AC$_5$TGC$_8$GGCTGGAGGTCCGGG

RNF2: GTC$_3$ATC$_6$TTAGTCATTACCTGAGG

図１４０Ａ

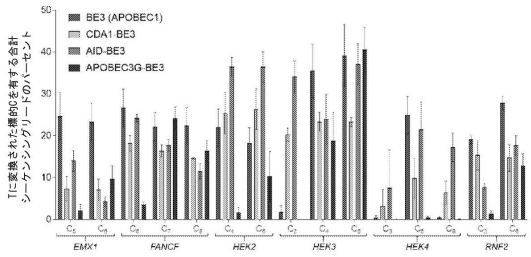


図１４０Ｂ

所与の配列を有するクリーンに編集されたリードのパーセント

HEK2プロトスペーサー配列	BE3	CDA1-BE3	AID-BE3
GAAT$_A$AT$_7$AAAGCATAGACTGCGGG	77 ± 1	88 ± 2	92 ± 2
GAAC$_3$AT$_7$AAAGCATAGACTGCGGG	23 ± 1	5 ± 1	4 ± 1
GAAT$_A$AC$_6$AAAGCATAGACTGCGGG	0 ± 0	7 ± 2	4 ± 2

図１４０Ｃ

【図１４１Ａ－Ｂ】

EMX1: GAGTC$_5$C$_6$GAGCAGAAGAAGAAGGG

FANCF: GGAATC$_6$C$_7$C$_8$TTCTGCAGCACCTGG

HEK2: GAAC$_3$AC$_6$AAAGCATAGACTGCGGG

HEK3: GGCC$_4$C$_5$AGACTGAGCACGTGATGG

HEK4: GGCAC$_5$TGCGGCTGGAGGTCCGGG

RNF2: GTC$_3$ATC$_6$TTAGTCATTACCTGAGG

図１４１Ａ

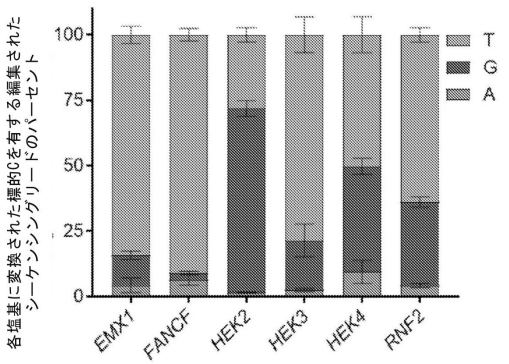


図１４１Ｂ

【図１４１Ｃ】

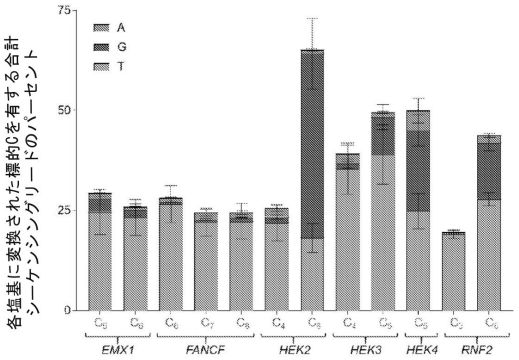


図１４１Ｃ

【図１４２Ａ－Ｂ】

HEK2: GAAC$_4$AC$_6$AAAGCATAGACTGCGGG

HEK3: GGCC$_4$C$_5$AGACTGAGCACGTGATGG

HEK4: GGCAC$_5$TGCGGCTGGAGGTCCGGG

RNF2: GTC$_3$ATC$_6$TTAGTCATTACCTGAGG

図１４２Ａ

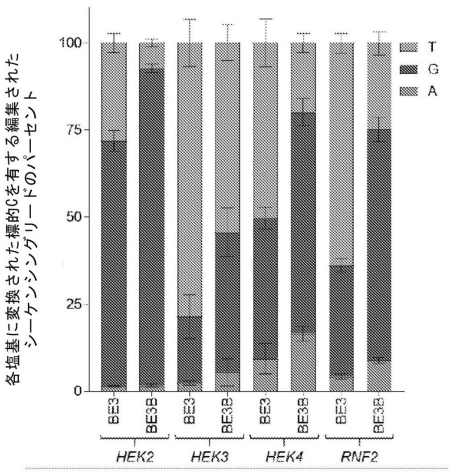


図１４２Ｂ

【図１４２Ｃ－Ｄ】

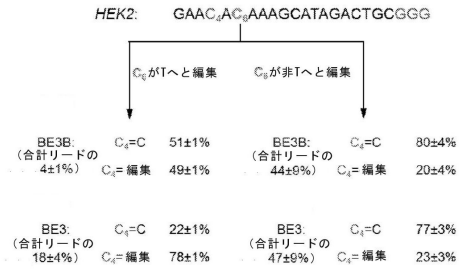


図１４２Ｃ

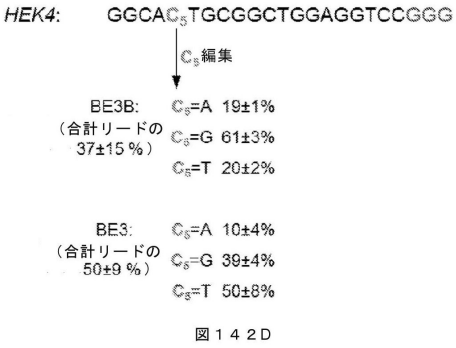


図１４２Ｄ

【図１４３】

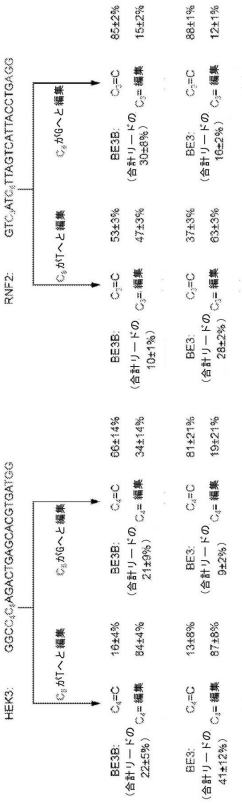


図１４３

【図１４４】

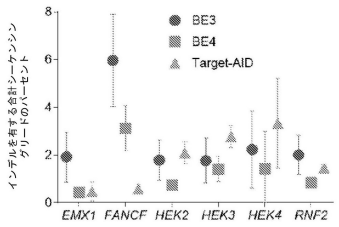


図１４４Ａ

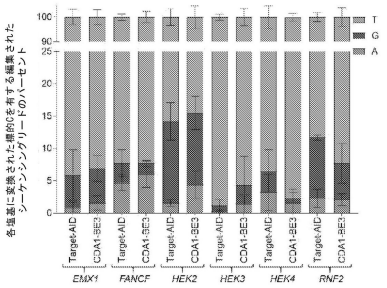


図１４４Ｂ

EMX1: GAGTC5CGAGCAGAAGAAGAAGGG
FANCF: GGAATCC7CTTCTGCAGCACCTGG
HEK2: GAACAC6AAAGCATAGACTGCGGG
HEK3: GGCCC5AGACTGAGCACGTGATGG
HEK4: GGCAC5TGCGGCTGGAGGTCCGGG
RNF2: GTCATC6TTAGTCATTACCTGAGG

図１４４Ｃ

【図１４５】

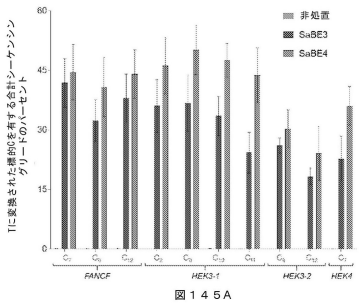


図１４５Ａ

FANCF: GATGTTC7C8AATC12AGTACGCAGAGAGT
HEK3-1: TC2TGCTTC8TC10C11AGCCCTGGCCTGGGT
HEK3-2: ACGTGCTC8AGTC12TGGGCCCCAAGGAT
HEK4: GTGGCAC7TGCGGCTGGAGGTGGGGGT

図１４５Ｂ

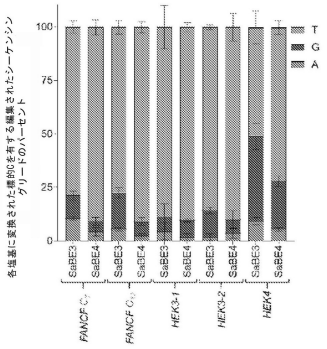


図１４５Ｃ

【図１４６】

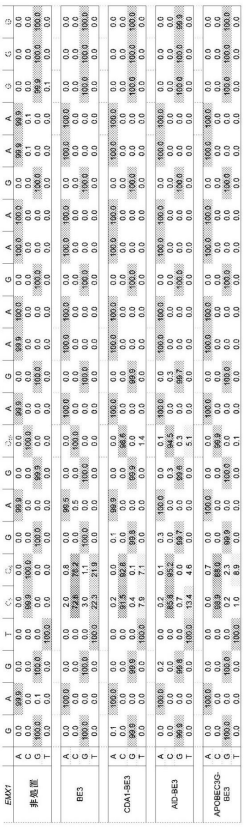


図１４６

【図１４７】

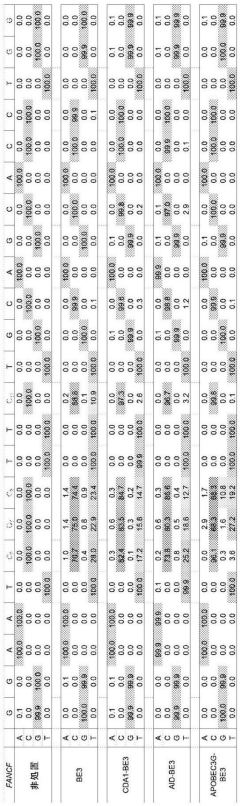


図１４７

【図１４８】

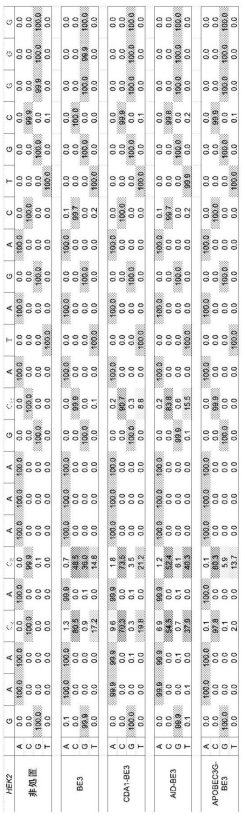


図１４８

【図１４９】

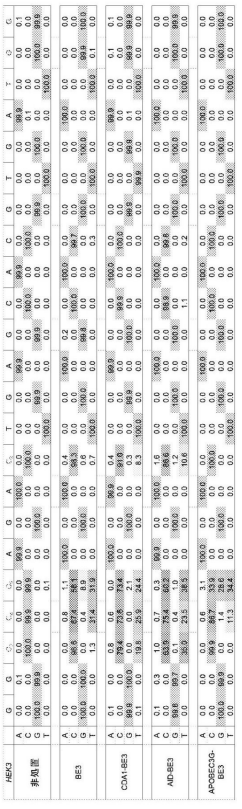


図１４９

【図１５０】

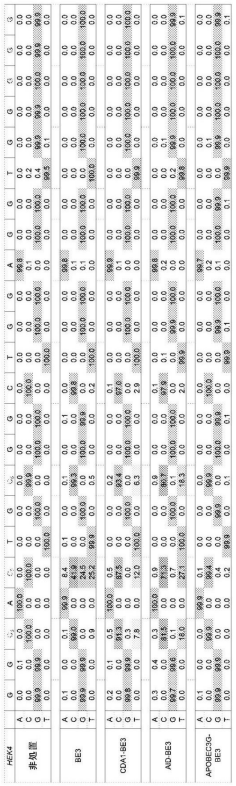


図１５０

【図１５１】

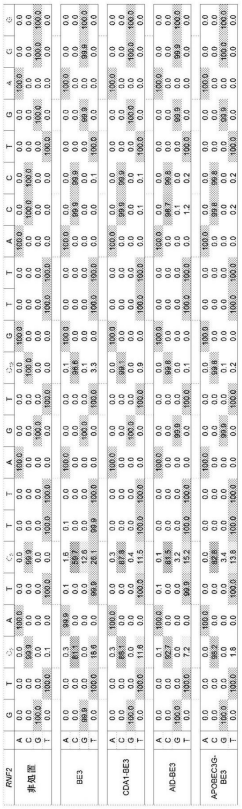


図１５１

【図１５２】

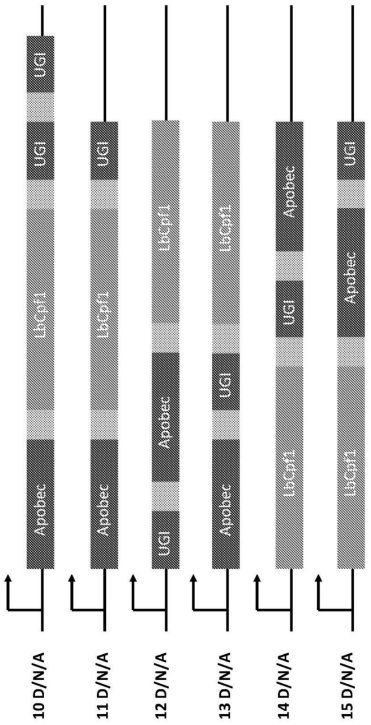


図１５２

【図１５３】

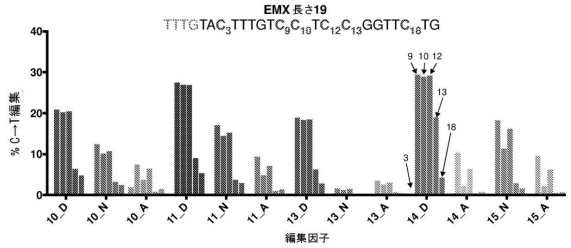


図１５３

【図１５４】

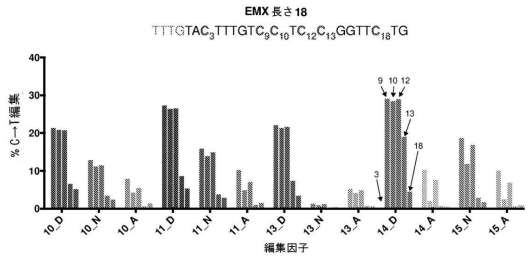


図１５４

【図１５５】

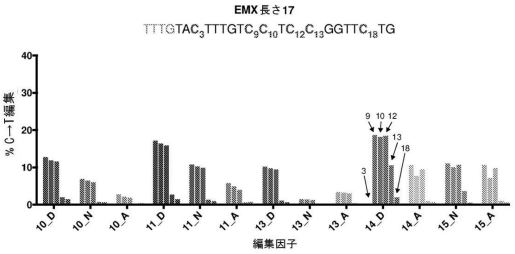


図１５５

【図１５６】

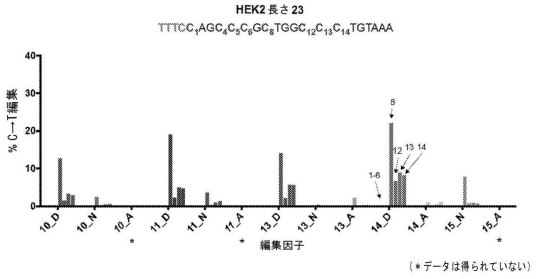


図１５６

【図１５７】

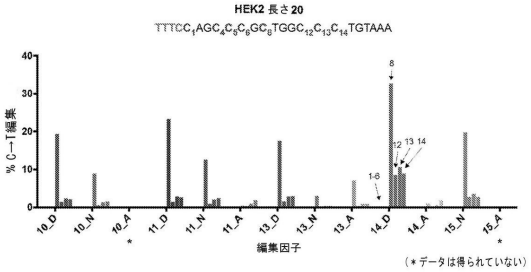


図１５７

【図１５８】

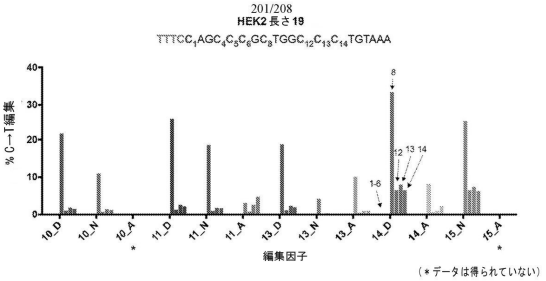


図１５８

【図１５９】

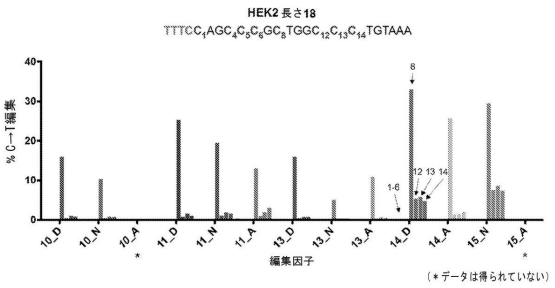


図１５９

【図１６０】

EMX 19

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C3	0.025	0.039	2.662	0.012	0.013	0.041	0.02	0.013	0.012	0.033	0.012	0.02	0.021
C9	20.899	13.15	10.456	27.535	18.921	13.072	19.043	1.591	3.638	29.39	14.271	20.95	13.924
C10	20.226	10.784	5.206	26.909	16.026	6.748	18.445	1.266	2.658	28.933	2.934	13.039	3.051
C12	20.453	11.385	9.095	26.849	16.893	9.983	18.589	1.437	3.127	29.165	8.695	18.597	9.07
C13	6.39	3.396	1.085	8.946	4.126	1.3	6.317	0.257	0.659	18.973	0.547	3.247	0.8
C18	4.757	2.603	2.013	5.304	3.233	1.914	2.805	0.235	0.367	4.302	0.992	1.848	0.994
インデル(%)	0.028979	6.0862	28.9183	0.038623	9.9557	28.6068	0.94296	1.4573	3.78	0.024623	27.6872	13.0741	31.3295
C3	0.024992276	0.036626538	1.892194851	0.011995537	0.011718756	0.029227121	0.010981141	0.012810055	0.011546641	0.032995187	0.008677541	0.011738518	0.011442381
C9	20.89234371	12.3496647	7.423202555	27.5243652	17.0562083	9.33251911	18.86343211	1.567814836	3.500483611	29.38276333	10.31917977	18.21097611	9.561680842
C10	20.22013871	10.12769442	3.700051133	26.89860669	14.44652355	4.817613141	18.2721071	1.247550588	2.557527619	28.92587258	2.121657555	11.33402681	2.095138696
C12	20.44707291	10.69208861	6.464686062	26.83863091	15.22607665	7.127183616	18.41371332	1.41003686	3.008779413	29.15781817	6.287597986	16.16560696	6.228414353
C13	6.388148424	3.185912651	0.772139445	8.942544779	3.731553822	0.92811316	6.257413222	0.252525474	0.634089858	18.96832831	0.395553102	2.822468397	0.549364
C18	4.755621471	2.444577621	1.438071462	5.301951444	2.914365722	1.366446585	2.778549971	0.231575555	0.353112741	4.309048072	0.717342988	1.606339063	0.682584477

図１６０

【図１６１】

EMX 18

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C3	0.023	0.008	0.032	0.044	0.035	0.016	0.03	0.016	0.016	0.018	0.02	0.017	0.01
C9	21.333	13.606	10.333	27.294	17.216	14.646	22.315	1.297	5.485	29.094	14.566	21.847	13.66
C10	20.865	11.831	5.528	26.409	15.041	6.931	21.548	0.932	4.309	28.456	2.996	13.762	3.337
C12	20.773	12.165	7.147	26.572	16.178	10.157	21.839	1.202	5.059	28.937	10.88	19.625	9.359
C13	6.538	3.591	0.848	8.562	4.096	1.387	7.426	0.129	0.729	18.993	0.945	3.353	0.809
C18	5.174	2.54	1.788	5.313	3.146	2.196	3.443	0.243	0.565	4.517	0.647	2.03	1.154
インデル(%)	0.039639	5.7886	23.1911	0.035862	7.9573	30.2672	1.1014	1.1121	4.6798	0.040521	29.8821	14.2217	26.5552
C3	0.022990	0.007536	0.024578	0.043984	0.032214	0.011157	0.029669	0.015826	0.015251	0.017992	0.014023	0.014582	0.007444
C9	21.325438	12.818403	7.936663	27.284211	15.846071	10.213080	22.069226	1.282576	5.228312	29.082210	10.213373	18.739985	10.03255
C10	20.856729	11.146150	4.245999	26.399529	13.844142	4.833187	21.310670	0.921635	4.107347	28.444469	2.100732	11.804809	2.450852
C12	20.764765	11.460816	5.489532	26.562470	14.890668	7.082770	21.598465	1.188632	4.822248	28.925274	7.628827	16.833991	6.873698
C13	6.535408	3.388313	0.651339	8.558929	3.770068	0.967193	7.344221	0.127565	0.694884	18.985303	0.662614	2.876146	0.594168
C18	5.171949	2.392969	1.373434	5.311094	2.895663	1.531334	3.405078	0.240297	0.538559	4.515169	0.453662	1.741299	0.847552

図１６１

【図１６２】

EMX 17

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C3	0.016	0.014	0.044	0.012	0.018	0.012	0.013	0.016	0.03	0.014	0.011	0.016	0.005
C9	12.742	6.895	2.862	17.142	10.77	5.941	10.177	1.471	3.395	18.719	11.13	11.103	11.493
C10	11.901	6.427	2.199	16.385	10.277	5.064	9.741	1.393	3.216	18.206	8.1	10.086	7.747
C12	11.548	6.014	1.916	15.933	9.872	4.041	9.468	1.287	2.986	18.492	9.847	10.751	10.523
C13	1.983	0.768	0.252	2.696	1.357	0.599	1.155	0.083	0.285	10.532	0.971	3.674	1.082
C18	1.457	0.613	0.361	1.471	0.987	0.773	0.667	0.043	0.245	1.978	0.729	0.545	0.621
インデル(%)	0.017245	0.052423	1.7787	0.025203	0.11499	2.7876	0.14938	0.019738	0.041079	0.04713	4.2923	0.37931	6.7995
C3	0.015997240	0.013992660	0.043217371	0.011996098	0.0179793	0.011665490	0.012980581	0.015996840	0.029987680	0.0139934	0.010552785	0.015939310	0.004660030
C9	12.739802810	6.891385430	2.811095610	17.137679700	10.757651560	5.775388680	10.161797600	1.470709650	3.393600570	18.710177000	10.652267000	11.060388520	10.711533500
C10	11.898894770	6.423630770	2.159886390	16.380870500	10.265118250	4.922835940	9.726448890	1.392725050	3.214678900	18.197419500	7.752323700	10.047742800	7.220242740
C12	11.546000500	6.010847280	1.881920110	15.928898440	9.860648190	3.932835080	9.453857670	1.286745970	2.984773380	18.483284700	9.424337220	10.710220040	9.807488620
C13	1.992656310	0.767597390	0.247517680	2.695320530	1.355439550	0.582302280	1.153274660	0.082983620	0.284882920	10.527039630	0.929321770	3.660066150	1.080429410
C18	1.456748740	0.612678650	0.354578880	1.470629200	0.985885050	0.775145185	0.666003640	0.042991510	0.244899360	1.977067770	0.697709130	0.542932760	0.578775110

図１６２

【図１６３】

Hek2 23

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C1	0.042	0.021	0	0.039	0.084	0	0.045	0.07	0.04	0.045	0.056	0.057	0
C4	0.016	0.018	0	0.023	0.013	0	0.027	0.022	0.016	0.035	0.008	0.016	0
C5	0.031	0.018	0	0.037	0.052	0	0.048	0.021	0.044	0.047	0.157	0.017	0
C6	0.064	0.045	0	0.064	0.02	0	0.077	0.068	0.051	0.078	0.032	0.08	0
C8	12.758	2.444	0	19.198	3.66	0	14.134	0.423	2.22	22.158	0.977	7.763	0
C12	1.418	0.291	0	2.281	0.446	0	2.163	0.101	0.21	6.745	0.235	0.809	0
C13	0.282	0.603	0	5.021	1.121	0	5.713	0.042	0.308	8.933	0.599	0.921	0
C14	2.898	0.641	0	4.719	1.38	0	5.614	0.096	0.38	8.268	1.066	0.746	0
インデル(%)	0.052343	11.0664		0.041898			2.2549	6.3138	13.9681	0.085777	59.2508	11.3052	
C1	0.041978020	0.018676060	0	0.038983660	0.084	0	0.043985300	0.065580340	0.034413760	0.044961400	0.022261955	0.056559604	0
C4	0.015991630	0.016008050	0	0.022990360	0.013	0	0.026539118	0.020610960	0.013776510	0.034969980	0.008325994	0.016419117	0
C5	0.030898377	0.016008050	0	0.036898450	0.052	0	0.046911765	0.019674100	0.037854404	0.044959960	0.066976264	0.016907812	0
C6	0.063956650	0.040020120	0	0.063973193	0.02	0	0.075263973	0.066270662	0.049387827	0.077933809	0.031803974	0.079953564	0
C8	12.757721	2.435759718	0	19.189956440	3.66	0	13.812352100	0.394625263	2.169098318	22.039253500	0.938196988	6.881547232	0
C12	1.417137271	0.287937218	0	2.280964564	0.446	0	2.144139251	0.096766988	0.210455595	6.745000000	0.234762564	0.809000000	0
C13	3.280282	0.596269821	0	5.018895863	1.121	0	5.586417756	0.039134982	0.264979235	8.925337754	0.244087711	0.811687911	0
C14	2.896548311	0.570086438	0	4.717702283	1.38	0	5.487405911	0.089938375	0.326921221	8.266900795	0.434386471	0.661566321	0

図１６３

【図１６４】

Hek2 20

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C1	0.094	0.051		0.036	0.116	0.126	0.035	0.022	0.042	0.032	0.02	0.133	
C4	0.064	0.023		0.026	0.011	0.01	0.019	0.012	0.011	0.025	0.05	0.016	0
C5	0.033	0.017	0	0.051	0.054	0.048	0.046	0.02	0.029	0.035	0.075	0.043	0
C6	0.096	0.083	0	0.058	0.032	0.022	0.055	0.016	0.047	0.139	0.166	0.083	0
C8	19.339	9.265	0	23.342	13.199	1.173	17.974	3.225	7.353	32.768	2.835	20.891	0
C12	1.402	0.551	0	1.491	1.036	0.713	1.68	0.188	0.471	8.549	0.3	2.967	0
C13	2.377	1.409	0	2.974	2.174	2.814	3.043	0.356	0.895	10.654	1.81	3.81	0
C14	2.129	1.545	0	2.604	2.523	5.582	3.076	0.316	0.955	8.964	4.97	3.086	0
インデル(%)	0.10237	3.2815	24.3902	0.10375	4.3674	66.4779	1.8097	1.6132	3.0719	0.093678	63.4082	5.4543	
C1	0.02397543	0.04932644	0	0.03596265	0.11093382	0.04223785	0.03436661	0.0216451	0.0407098	0.03197002	0.00731836	0.12574578	0
C4	0.05394472	0.02224526	0	0.02597303	0.01051959	0.00335221	0.01865616	0.01869349	0.01066209	0.02497658	0.00585469	0.01512731	0
C5	0.03296622	0.01644215	0	0.05094709	0.05164161	0.01609061	0.04516754	0.01967736	0.02810915	0.03496721	0.02744385	0.04065465	0
C6	0.09590172	0.08027636	0	0.05793983	0.03060243	0.00737486	0.05400467	0.01574189	0.04555621	0.13886979	0.06074239	0.07847293	0
C8	19.31920227	8.96096903	0	23.3177827	12.6225469	0.39321423	17.6487245	3.1729743	7.127212319	32.7373036	1.037377531	19.7515422	0
C12	1.400564777	0.53291894	0	1.489453097	0.99075374	0.23901257	1.64959704	0.18496718	0.456533135	8.540991477	0.109775412	2.805170923	0
C13	2.374566667	1.362763867	0	2.970914488	2.07905272	0.943311889	2.98793083	0.35025701	0.867506505	10.64019505	0.662311582	3.602191177	0
C14	2.126820254	1.494300983	0	2.601298835	2.4128105	1.872120362	3.020333363	0.310902295	0.925663362	8.955660271	1.818612461	2.917680312	0

図１６４

【図１６５】

Hek2 19

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C1	0.072	0.031	0	0.039	0.062	0.406	0.023	0.019	0.025	0.06	0.224	0.04	0
C4	0.034	0.019	0	0.021	0.06	0.018	0.039	0.033	0.025	0.012	0.009	0.008	0
C5	0.03	0.037	0	0.044	0.045	0.027	0.018	0.031	0.033	0.024	0.021	0.012	0
C6	0.053	0.066	0	0.042	0.071	0.164	0.033	0.031	0.046	0.098	0.117	0.034	0
C8	21.611	11.001	0	25.556	18.67	6.38	18.944	4.272	10.143	33.441	15.227	25.017	0
C12	1.049	0.704	0	1.345	1.022	1.548	1.154	0.195	0.489	6.54	1.177	6.539	0
C13	1.914	1.39	0	2.629	1.819	5.23	2.301	0.253	0.881	8.065	1.823	7.452	0
C14	1.545	1.212	0	2.238	1.761	9.623	1.98	0.151	0.975	6.517	4.174	6.365	0
インデル(%)	0.056704	0.17936		0.04541	0.43984	50.2032	0.61981	0.18691	0.22381	0.14694	45.816	0.52493	
C1	0.07195917	0.0309444	0	0.03898229	0.0617273	0.20217501	0.02285744	0.01896449	0.02494405	0.05991184	0.12137216	0.03979003	0
C4	0.03398072	0.01896592	0	0.02099046	0.0597361	0.00896342	0.03875827	0.03293832	0.02494405	0.01198237	0.00487656	0.00795801	0
C5	0.02998299	0.03693364	0	0.04398002	0.04480207	0.01344514	0.01788843	0.03094206	0.03252614	0.02396473	0.01137864	0.01193701	0
C6	0.05296995	0.06588162	0	0.04198093	0.07068771	0.08166675	0.03279546	0.03094206	0.04589705	0.097856	0.06339528	0.03382152	0
C8	21.5987457	10.98126886	0	25.544395	18.58788819	3.177935841	18.8265832	4.2640152	10.120299	33.39186818	8.250597688	24.8856783	0
C12	1.04840518	0.70273731	0	1.34438924	1.01750484	0.77085446	1.14684739	0.19463553	0.48790557	6.530390012	0.637745688	6.50467483	0
C13	1.91291469	1.3875069	0	2.627806617	1.81099931	2.604372647	2.28673817	0.25252712	0.87902823	8.053149291	0.987774321	7.41288222	0
C14	1.54412392	1.20982616	0	2.236983721	1.75325442	4.791946061	1.967727761	0.150717771	0.972817851	6.507423921	2.261640161	6.331584211	0

図１６５

【図１６６】

Hek2 18

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C1	0.097	0.027	0	0.035	0.041	0.315	0.099	0.034	0.034	0.028	0.125	0.074	0
C4	0.054	0.016	0	0.027	0.018	0.017	0.018	0.038	0.019	0.014	0.021	0.035	0
C5	0.052	0.018	0	0.051	0.034	0.045	0.083	0.039	0.022	0.014	0.088	0.035	0
C6	0.096	0.035	0	0.088	0.08	0.047	0.092	0.051	0.027	0.061	0.288	0.107	0
C8	16.596	10.427	0	25.032	19.569	14.498	15.982	5.043	10.317	38.484	29.476	29.702	0
C12	1.517	0.503	0	1.602	1.161	1.138	1.513	0.208	0.709	5.445	1.741	7.751	0
C13	0.877	0.803	0	1.627	1.669	2.289	0.803	0.315	0.709	5.922	2.141	9.732	0
C14		0.819	0	1.132	1.609	3.469	0.751	0.375	0.636	4.806	2.316	7.414	0
インデル(%)	0.076858	0.17892		0.048785	0.13947	11.5698	0.14868	0.1177	0.076621	0.23744	9.8678	0.50485	
C1	0.036927156	0.026951169	0	0.034982893	0.04094282	0.27858348	0.03894201	0.03395598	0.03897395	0.02793352	0.11266525	0.07362641	0
C4	0.013989234	0.01597137	0	0.006956659	0.01179149	0.01650466	0.01179324	0.0179527	0.01898544	0.01396076	0.02799276	0.0348233	0
C5	0.016196003	0.01796779	0	0.05097312	0.03955838	0.03977954	0.01429094	0.038541	0.01248314	0.01396076	0.07931854	0.0348233	0
C6	0.059356496	0.03493738	0	0.031984839	0.07988042	0.04154664	0.08187808	0.05093997	0.06954637	0.08080707	0.18747498	0.10645981	0
C8	16.02675138	10.4048944	0	25.28553944	19.4617498	13.0960767	15.5672246	5.59706459	10.9176284	32.3495428	25.5498415	29.7520495	0
C12	0.570561114	0.50210003	0	0.801686074	1.15938075	1.0964581	0.5122372	0.2277164	0.31675711	5.347357642	1.32584666	7.5725158	0
C13	1.1913919	0.90138455	0	1.606421093	1.86639331	2.03932172	0.8018961	0.31462925	0.70845676	5.587986429	1.5492016	8.7078153	0
C14	0.879162596	0.81753465	0	1.131144775	1.60675593	3.06707146	0.74986941	0.31455863	0.63551269	4.794458865	2.087461751	7.378570942	0

図１６６

【配列表】
0007191388000001.app

フロントページの続き

(51)国際特許分類		FI		
Ｃ１２Ｎ	*1/15 (2006.01)*	Ｃ１２Ｎ	1/15	
Ｃ１２Ｎ	*1/19 (2006.01)*	Ｃ１２Ｎ	1/19	
Ｃ１２Ｎ	*1/21 (2006.01)*	Ｃ１２Ｎ	1/21	
Ｃ１２Ｎ	*5/10 (2006.01)*	Ｃ１２Ｎ	5/10	
Ｃ１２Ｎ	*9/16 (2006.01)*	Ｃ１２Ｎ	9/16	Ｚ
Ａ６１Ｋ	*38/50 (2006.01)*	Ａ６１Ｋ	38/50	
Ａ６１Ｋ	*38/47 (2006.01)*	Ａ６１Ｋ	38/47	
Ａ６１Ｋ	*47/65 (2017.01)*	Ａ６１Ｋ	47/65	
Ａ６１Ｋ	*47/59 (2017.01)*	Ａ６１Ｋ	47/59	
Ａ６１Ｋ	*48/00 (2006.01)*	Ａ６１Ｋ	48/00	
Ａ６１Ｋ	*31/7105(2006.01)*	Ａ６１Ｋ	31/7105	
Ａ６１Ｋ	*31/713(2006.01)*	Ａ６１Ｋ	31/713	
Ａ６１Ｋ	*31/711(2006.01)*	Ａ６１Ｋ	31/711	
Ｃ１２Ｎ	*15/62 (2006.01)*	Ｃ１２Ｎ	15/62	Ｚ ＺＮＡ

(33)優先権主張国・地域又は機関
米国(US)
(31)優先権主張番号 62/475,830
(32)優先日 平成29年3月23日(2017.3.23)
(33)優先権主張国・地域又は機関
米国(US)
(31)優先権主張番号 62/551,951
(32)優先日 平成29年8月30日(2017.8.30)
(33)優先権主張国・地域又は機関
米国(US)
アメリカ合衆国 カリフォルニア州 ９２１２２、サンディエゴ、ジュディシャル ドライブ ９１３５、＃３４１９
(72)発明者 チェン，リ－ウェイ
アメリカ合衆国 マサチューセッツ州 ０２１４３、サマービル、エイヴォン ストリート ８６、アパートメント ２
(72)発明者 レス，ホリー，エー．
アメリカ合衆国 マサチューセッツ州 ０２１４４、サマービル、ホウソーン ストリート ３４
審査官 松田 芳子
(56)参考文献 国際公開第２０１７／０７０６３２（WO，Ａ２）
国際公開第２０１８／０３９４３８（WO，Ａ１）
Nature, 2016,vol.533, no.7603,p.420-424
Mol.Cell, 2015, vol.60, no.3,p.385-397
Nature, 2016, vol.542,no.7640, p.237-241
PNAS USA, 2016, vol.113,no.15, p.4057-4062
(58)調査した分野 (Int.Cl.，ＤＢ名)
Ｃ１２Ｎ １５／０９
ＪＳＴＰｌｕｓ／ＪＭＥＤＰｌｕｓ／ＪＳＴ７５８０（ＪＤｒｅａｍＩＩＩ）
ＣＡｐｌｕｓ／ＭＥＤＬＩＮＥ／ＥＭＢＡＳＥ／ＢＩＯＳＩＳ（ＳＴＮ）