

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6633524号
(P6633524)

(45) 発行日 令和2年1月22日 (2020.1.22)

(24) 登録日 令和1年12月20日 (2019.12.20)

(51) Int. Cl.

F I

C O 7 K	14/00	(2006.01)	C O 7 K	14/00	Z N A
C 1 2 N	9/22	(2006.01)	C 1 2 N	9/22	
C 1 2 N	15/09	(2006.01)	C 1 2 N	15/09	1 0 0
C 1 2 Q	1/44	(2006.01)	C 1 2 Q	1/44	
C 1 2 Q	1/68	(2018.01)	C 1 2 Q	1/68	

請求項の数 22 (全 95 頁) 最終頁に続く

(21) 出願番号 特願2016-536473 (P2016-536473)
 (86) (22) 出願日 平成26年8月22日 (2014.8.22)
 (65) 公表番号 特表2016-534116 (P2016-534116A)
 (43) 公表日 平成28年11月4日 (2016.11.4)
 (86) 国際出願番号 PCT/US2014/052231
 (87) 国際公開番号 W02015/027134
 (87) 国際公開日 平成27年2月26日 (2015.2.26)
 審査請求日 平成29年8月22日 (2017.8.22)
 (31) 優先権主張番号 61/868,846
 (32) 優先日 平成25年8月22日 (2013.8.22)
 (33) 優先権主張国・地域又は機関
米国 (US)
 (31) 優先権主張番号 14/320,519
 (32) 優先日 平成26年6月30日 (2014.6.30)
 (33) 優先権主張国・地域又は機関
米国 (US)

(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 02
 138, 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
 (72) 発明者 リウ, デービッド, アール.
 アメリカ合衆国 マサチューセッツ州 0
 2420, レキシントン, ウィットマン
 サークル 3

最終頁に続く

(54) 【発明の名称】 操作された転写活性化因子様エフェクター (T A L E) ドメインおよびその使用

(57) 【特許請求の範囲】

【請求項 1】

単離された転写活性化因子様エフェクター (T A L E) ドメインであって、ここで、
 (i) 単離された T A L E ドメインが、配列番号 2、配列番号 3 または配列番号 4 のアミノ酸配列を含む N 末端 T A L E ドメインであり、ここで、単離された N 末端ドメインの正味電荷が、生理的 p H での標準 N 末端ドメイン (配列番号 1) の正味電荷より少ない；または

(i i) 単離された T A L E ドメインが、下記アミノ酸置換：K 4 8 Q、R 5 2 Q、および R 6 1 Q、ならびに任意に 1 または 2 以上の下記アミノ酸置換：K 3 7 Q、K 3 8 Q、R 4 9 Q、および R 5 3 Q を含む点で、配列番号 2 の標準 C 末端ドメインとは異なるアミノ酸配列を含む C 末端 T A L E ドメインであり、ここで、C 末端ドメインの正味電荷が、生理的 p H での標準 C 末端ドメイン (配列番号 2) の正味電荷より少ない；
 前記単離された T A L E ドメイン。

【請求項 2】

C 末端ドメインの正味電荷が、+ 6 以下、+ 5 以下、+ 4 以下、+ 3 以下、+ 2 以下、+ 1 以下、+ 0 以下、- 1 以下、- 2 以下、- 3 以下、- 4 以下、または - 5 以下である、請求項 1 に記載の単離された T A L E ドメイン。

【請求項 3】

単離された T A L E ドメインの、少なくとも 1 個の、少なくとも 2 個の、少なくとも 3 個の、少なくとも 4 個の、少なくとも 5 個の、少なくとも 6 個の、少なくとも 7 個の、少

なくとも8個の、少なくとも9個の、少なくとも10個の、少なくとも11個の、少なくとも12個の、少なくとも13個の、少なくとも14個の、または少なくとも15個のカチオン性アミノ酸が、生理的pHで電荷を示さないかまたは負電荷を示すアミノ酸残基で置き換えられている、請求項1または2に記載の単離されたT A L Eドメイン。

【請求項4】

N末端ドメインが、配列番号1のK57、K78、R84、R97、K110、K113およびR114に対応する位置における1または2以上のアミノ酸置換を含む、請求項1～3のいずれか一項に記載の単離されたT A L Eドメイン。

【請求項5】

単離されたT A L Eドメインが、次の構造：

[N末端ドメイン] - [T A L E反復アレイ] - [C末端ドメイン] - [エフェクタードメイン]

または

[エフェクタードメイン] - [N末端ドメイン] - [T A L E反復アレイ] - [C末端ドメイン]

を含むT A L E分子中に含まれる、請求項1～4のいずれか一項に記載の単離されたT A L Eドメイン。

【請求項6】

エフェクタードメインが、ヌクレアーゼドメイン、転写活性化因子もしくはリプレッサードメイン、リコンビナーゼドメイン、またはエピゲノム修飾酵素ドメインを含む、請求項5に記載の単離されたT A L Eドメイン。

【請求項7】

T A L E分子が、疾患または障害に関連することが知られている遺伝子内の標的配列に結合する、請求項5または6に記載の単離されたT A L Eドメイン。

【請求項8】

単離されたT A L Eドメインが、Q3バリエーション配列（配列番号22においてK48Q、R52Q、R61Q）またはQ7バリエーション配列（配列番号22においてK37Q、K38Q、K48Q、R49Q、R52Q、R53Q、R61Q）を含む点で、配列番号22の標準C末端ドメインとは異なるアミノ酸配列を含む、C末端T A L Eドメインである、請求項1～7のいずれか一項に記載の単離されたT A L Eドメイン。

【請求項9】

単離されたT A L Eドメインが、配列番号22のR8、R30およびR57に対応する位置における1または2以上のアミノ酸置換をさらに含む、請求項1～8のいずれか一項に記載の単離されたT A L Eドメイン。

【請求項10】

以下を含む、転写活性化因子様エフェクターヌクレアーゼ（T A L E N）であって、

（a）ヌクレアーゼ切断ドメイン；

（b）ヌクレアーゼ切断ドメインにコンジュゲートした、請求項1～4のいずれか一項に定義されるC末端ドメイン；

（c）C末端ドメインにコンジュゲートしたT A L E反復アレイ；および

（d）T A L E反復アレイにコンジュゲートした、請求項1～4のいずれか一項に定義されるN末端ドメイン、

前記T A L E N。

【請求項11】

ヌクレアーゼ切断ドメインが、F o k Iヌクレアーゼドメインであり、任意にF o k Iヌクレアーゼドメインが、配列番号26～30で提供される配列を含む、請求項10に記載のT A L E N。

【請求項12】

T A L E Nが単量体であり、任意にT A L E N単量体が、別のT A L E N単量体と二量体化してT A L E N二量体を形成することができ、任意に二量体が、ヘテロ二量体である

10

20

30

40

50

、請求項 1 0 または 1 1 に記載の T A L E N。

【請求項 1 3】

T A L E N が、疾患または障害に関連することが知られている遺伝子内の標的配列に結合し、任意に T A L E N が、二量体化の際に標的配列を切断する、請求項 1 0 ~ 1 2 のいずれか一項に記載の T A L E N。

【請求項 1 4】

疾患が、H I V / A I D S または増殖性疾患である、請求項 1 0 ~ 1 3 のいずれか一項に記載の T A L E N。

【請求項 1 5】

T A L E N が、C C R 5 標的配列、A T M 標的配列または V E G F A 標的配列に結合する、請求項 1 0 ~ 1 4 のいずれか一項に記載の T A L E N。

10

【請求項 1 6】

請求項 1 0 ~ 1 5 のいずれか一項に記載の T A L E N と、該 T A L E N と共にヘテロ二量体を形成することができる異なる T A L E N とを含む組成物であって、ここで該二量体がヌクレアーゼ活性を示す、前記組成物。

【請求項 1 7】

請求項 1 0 ~ 1 5 のいずれか一項に記載の T A L E N または請求項 1 6 に記載の組成物、および薬学的に許容し得る賦形剤を含む、医薬組成物。

【請求項 1 8】

(a) 医薬組成物が、対象への投与のために処方されるか、
(b) 医薬組成物が、対象の細胞内の標的配列を切断するための有効量の T A L E N を含むか、および / または
(c) T A L E N が疾患または障害と関連することが知られている遺伝子内の標的配列に結合し、およびここで前記組成物が疾患または障害に関連する症状を軽減するための有効量の T A L E N を含む、
請求項 1 7 に記載の医薬組成物。

20

【請求項 1 9】

標的配列を含む核酸分子を、T A L E N が標的配列に結合して標的配列を切断するのに好適な条件下で、標的配列に結合する T A L E N と接触させることを含む、核酸分子中の標的配列の切断における使用のための、請求項 1 0 ~ 1 5 のいずれか一項に記載の T A L E N。

30

【請求項 2 0】

標的配列が細胞に含まれている、請求項 1 9 に記載の T A L E N。

【請求項 2 1】

請求項 1 ~ 9 のいずれか一項に定義される単離された転写活性化因子様エフェクター (T A L E) ドメインを含む操作された T A L E N を製造する方法であって、方法が、標準 N 末端 T A L E N ドメインおよび / または標準 C 末端 T A L E N ドメインの少なくとも 1 つのアミノ酸を、生理的 p H で電荷を有さないかまたは負電荷を有するアミノ酸で置き換えること ;

したがって、減少した正味電荷の N 末端ドメインおよび / または C 末端ドメインを有する、操作された T A L E N を生成すること、を含む、前記方法。

40

【請求項 2 2】

請求項 1 0 ~ 1 5 のいずれか一項に記載の T A L E N 、請求項 1 6 に記載の組成物、または請求項 1 7 または 1 8 に記載の医薬組成物を含み、任意に T A L E N を賦形剤と接触させて、核酸を T A L E N と接触させるのに好適な組成物を生成するための、賦形剤および指示書を含む、キット。

【発明の詳細な説明】

【技術分野】

50

【 0 0 0 1 】

関連出願

本出願は、35 U.S.C. § 365(c)の下で、2014年6月30日に出願されたU.S.S.N. 14/320,519の優先権を主張し、また、35 U.S.C. § 119(e)の下で、2013年8月22日に出願された米国仮特許出願U.S.S.N. 61/868,846の優先権を主張する；これらの出願の各々は、本明細書に参照により組み込まれる。

【 0 0 0 2 】

政府支援

本発明は、米国政府の支援を受け、Defense Advanced Research Projects Agencyによって授与されたグラントHR0011-11-2-0003およびN66001-12-C-4207；National Institute of General Medical Sciencesによって授与されたグラントT32GM007753；およびNational Institutes of Healthによって授与されたグラントDP1 GM105378のもとでなされた。米国政府は本発明において一定の権利を有する。

【 背景技術 】

【 0 0 0 3 】

発明の背景

転写活性化因子様エフェクターヌクレアーゼ (T A L E N) は、F o k I制限エンドヌクレアーゼ切断ドメインとDNA結合転写活性化因子様エフェクター (T A L E) 反復アレイとの融合体である。T A L E Nは、所望の標的DNA配列に特異的に結合して切断するように操作することができ、これはin vitroおよびin vivoでの核酸分子、遺伝子およびゲノムの操作のために有用である。操作されたT A L E Nは、限定することなく基礎研究および治療的用途を含む多くの用途の場面において有用である。例えば操作されたT A L E Nは、遺伝子ロックアウトまたはロックインの作製の場面において、ゲノムの操作のために使用することができ、これはそれぞれ、非相同末端結合 (N H E J) を介した、標的遺伝子のロックアウトのための標的ゲノム部位でのDNA切断の誘導によるか、または、相同性配向型修復 (H D R) を介した、外来DNA鋳型を使用する標的ゲノム配列の置換による。T A L E Nはしたがって、遺伝子操作された細胞、組織、および生物の生成に有用である。

【 0 0 0 4 】

T A L E Nは、天然および合成の配列を含む、任意の所望の標的DNA配列を切断するように設計することができる。しかし、密接に関連するオフターゲット配列から標的配列を区別するT A L E Nの能力は、深くは研究されていない。この能力およびこれに影響を与えるパラメータを理解することは、所望の特異性レベルを有するT A L E Nの設計に、およびまた、切断すべきユニークな標的配列を選択するため、例えば望ましくないオフターゲット切断の可能性を最小限にするために重要である。

【 発明の概要 】

【 0 0 0 5 】

T A L E Nは、in vitroおよびin vivoでの遺伝子およびゲノムの操作のための用途の広いツールであり、その理由は、T A L E Nが、核酸分子内の実質的に全ての標的配列に結合して切断するように設計することができるからである。例えば、T A L E Nは、DNA切断の誘導を介した、細胞のゲノム内のDNA配列の標的化された除去のために使用することができ、該DNA切断は、次に、非相同末端結合 (N H E J) を介して、細胞のDNA修復機構により修復される。T A L E Nはまた、ゲノム配列に挿入される配列を含む核酸の存在下で、相同性配向型修復 (H D R) を介した、標的化された配列置換のためにも使用することができる。T A L E Nは生細胞のゲノムを操作するために使用可能であるため、得られた遺伝的に修飾された細胞は、形質転換された細胞または組織培養物および生物を生成するために、使用することができる。

【 0 0 0 6 】

複雑な試料の場面で、例えばゲノムの場面において、T A L E NがDNA配列の標的切断のために使用されるシナリオでは、T A L E Nが特定の標的配列のみに結合して切断し

、オフターゲット切断活性はないか、または最小であることが、多くの場合に望ましい（例えば、PCT出願公開WO2013/066438 A2を参照、この全内容は本明細書に参照により組み込まれる）。いくつかの態様において、理想的なTALENは、その意図する標的配列にのみ特異的に結合し、全くオフターゲット活性を有さず、したがって単一の配列の標的切断、例えば、全ゲノムの場面においては目的の遺伝子の単一のアレルのみの標的切断を、可能とする。

【0007】

本開示のいくつかの側面は、TALENの、オフターゲット配列を切断するという傾向およびオフターゲットTALEN活性の性質に影響するパラメータについては、あまり理解されていないとの認識に基づく。ここで紹介する研究は、TALENのオフターゲット活性をもたらし構造パラメータの、より良い理解を提供する。オフターゲット活性を有さないか、最小限にしか有さない、操作されたTALENを生成するための方法およびシステムが、本明細書に提供され、また、増加したオンターゲット切断効率および最小限のオフターゲット活性を有する、操作されたTALENも提供される。TALENによる、非特異的またはオフターゲットDNA結合を低減するための、本明細書に提供される戦略、方法、および試薬は、他のDNA結合タンパク質にも適用可能であることが、当業者によって理解されるであろう。特に、DNAへの非特異的結合を低減するために、DNA結合タンパク質のアミノ酸配列を修飾するための戦略であって、カチオン性アミノ酸残基を、カチオン性ではないか、非荷電であるか、または生理的pHでアニオン性ではないアミノ酸残基で置換することによる、前記戦略を使用して、例えば、他のTALENエフェクタータンパク質、操作されたジンクフィンガータンパク質（ジンクフィンガーヌクレアーゼを含む）、およびCas9タンパク質などの特異性を減少させることができる。

【0008】

本開示のいくつかの側面は、操作され、単離された転写活性化因子様エフェクター（TALE）ドメインを提供する。いくつかの態様において、単離されたTALEドメインはN末端TALEドメインであり、単離されたN末端ドメインの正味電荷は、生理的pHでの標準N末端ドメイン（配列番号1）の正味電荷より少ない。いくつかの態様において、単離されたTALEドメインはC末端TALEドメインであり、C末端ドメインの正味電荷は、生理的pHでの標準C末端ドメイン（配列番号22）の正味電荷より少ない。いくつかの態様において、単離されたTALEドメインはN末端TALEドメインであり、標的核酸分子に対するN末端ドメインの結合エネルギーは、標準N末端ドメイン（配列番号1）の結合エネルギーより少ない。いくつかの態様において、単離されたTALEドメインはC末端TALEドメインであり、標的核酸分子に対するC末端ドメインの結合エネルギーは、標準C末端ドメイン（配列番号22）の結合エネルギーより少ない。いくつかの態様において、C末端ドメインの正味電荷は、+6以下、+5以下、+4以下、+3以下、+2以下、+1以下、+0以下、-1以下、-2以下、-3以下、-4以下、または-5以下である。いくつかの態様において、C末端ドメインはアミノ酸配列を含み、該アミノ酸配列は、標準C末端ドメイン配列とは次の点で、すなわち標準C末端ドメイン配列の少なくとも1つのカチオン性アミノ酸残基が、生理的pHで電荷を示さないかまたは負電荷を示すアミノ酸残基で置き換えられている点で異なっている。いくつかの態様において、N末端ドメインはアミノ酸配列を含み、該アミノ酸配列は、標準N末端ドメイン配列とは次の点で、すなわち標準N末端ドメイン配列の少なくとも1つのカチオン性アミノ酸残基が、生理的pHで電荷を示さないかまたは負電荷を示すアミノ酸残基で置き換えられている点で異なっている。いくつかの態様において、単離されたTALEドメインの、少なくとも1個の、少なくとも2個の、少なくとも3個の、少なくとも4個の、少なくとも5個の、少なくとも6個の、少なくとも7個の、少なくとも8個の、少なくとも9個の、少なくとも10個の、少なくとも11個の、少なくとも12個の、少なくとも13個の、少なくとも14個の、または少なくとも15個のカチオン性アミノ酸残基は、生理的pHで電荷を示さないかまたは負電荷を示すアミノ酸残基で置き換えられている。いくつかの態様において、少なくとも1つのカチオン性アミノ酸残基は、アルギニン（R）またはリジ

10

20

30

40

50

ン（K）である。いくつかの態様において、生理的 pH で電荷を示さないかまたは負電荷を示すアミノ酸残基は、グルタミン（Q）またはグリシン（G）である。いくつかの態様において、少なくとも 1 つのリジンまたはアルギニン残基は、グルタミン残基で置き換えられている。いくつかの態様において、C 末端ドメインは、1 または 2 以上の次のアミノ酸置換：K 7 7 7 Q、K 7 7 8 Q、K 7 8 8 Q、R 7 8 9 Q、R 7 9 2 Q、R 7 9 3 Q、R 8 0 1 Q、を含む。いくつかの態様において、C 末端ドメインは、Q 3 バリエーション配列（K 7 8 8 Q、R 7 9 2 Q、K 8 0 1 Q）を含む。いくつかの態様において、C 末端ドメインは、Q 7 バリエーション配列（K 7 7 7 Q、K 7 7 8 Q、K 7 8 8 Q、R 7 8 9 Q、R 7 9 2 Q、R 7 9 3 Q、R 8 0 1 Q）を含む。いくつかの態様において、N 末端ドメインは、標準 N 末端ドメインのランケット型である。いくつかの態様において、C 末端ドメインは、標準 C 末端ドメインのランケット型である。いくつかの態様において、ランケットされたドメインは、標準ドメインの残基の 90 % 未満、80 % 未満、70 % 未満、60 % 未満、50 % 未満、40 % 未満、30 % 未満、または 25 % 未満を含む。いくつかの態様において、ランケットされた C 末端ドメインは、60 個未満、50 個未満、40 個未満、30 個未満、29 個未満、28 個未満、27 個未満、26 個未満、25 個未満、24 個未満、23 個未満、22 個未満、21 個未満、または 20 個未満のアミノ酸残基を含む。いくつかの態様において、ランケットされた C 末端ドメインは、60、59、58、57、56、55、54、53、52、51、50、49、48、47、46、45、44、43、42、41、40、39、38、37、36、35、34、33、32、31、30、29、28、27、26、25、24、23、22、21、20、19、18、17、16、15、14、13、12、11、または 10 個の残基を含む。いくつかの態様において、単離された TALE ドメインは、次の構造：[N 末端ドメイン] - [TALE 反復アレイ] - [C 末端ドメイン] - [エフェクタードメイン]；または [エフェクタードメイン] - [N 末端ドメイン] - [TALE 反復アレイ] - [C 末端ドメイン]、を含む TALE 分子中に含まれる。いくつかの態様において、エフェクタードメインは、ヌクレアーゼドメイン、転写活性化因子もしくはリプレッサードメイン、リコンビナーゼドメイン、またはエピゲノム修飾酵素ドメインを含む。いくつかの態様において、TALE 分子は、疾患または障害に関連することが知られている遺伝子内の標的配列に結合する。

【0009】

本開示のいくつかの側面は、標準 TALE N と比較して、修飾された正味電荷および / またはそれらの標的核酸配列に結合する修飾された結合エネルギーを有する、転写活性化因子様エフェクターヌクレアーゼ（TALE N）を提供する。典型的には、本発明の TALE N は、（a）ヌクレアーゼ切断ドメイン；（b）ヌクレアーゼ切断ドメインにコンジュゲートした C 末端ドメイン；（c）C 末端ドメインにコンジュゲートした TALE 反復アレイ；および（d）TALE 反復アレイにコンジュゲートした N 末端ドメイン、を含む。いくつかの態様において、（i）生理的 pH での N 末端ドメインの正味電荷は、生理的 pH での標準 N 末端ドメイン（配列番号 1）の正味電荷より少なく；および / または（ii）生理的 pH での C 末端ドメインの正味電荷は、生理的 pH での標準 C 末端ドメイン（配列番号 22）の正味電荷より少ない。いくつかの態様において、（i）標的核酸分子に対する N 末端ドメインの結合エネルギーは、標準 N 末端ドメイン（配列番号 1）の結合エネルギーより少なく；および / または（ii）標的核酸分子に対する C 末端ドメインの結合エネルギーは、標準 C 末端ドメイン（配列番号 22）の結合エネルギーより少ない。いくつかの態様において、（i）標的核酸分子に対する N 末端ドメインの結合エネルギーは、標準 N 末端ドメイン（配列番号 1）の結合エネルギーより少ない；および / または（ii）標的核酸分子に対する C 末端ドメインの結合エネルギーは、標準 C 末端ドメイン（配列番号 22）の結合エネルギーより少ない。いくつかの態様において、生理的 pH での C 末端ドメインの正味電荷は、+ 6 以下、+ 5 以下、+ 4 以下、+ 3 以下、+ 2 以下、+ 1 以下、+ 0 以下、- 1 以下、- 2 以下、- 3 以下、- 4 以下、または - 5 以下である。いくつかの態様において、N 末端ドメインはアミノ酸配列を含み、該アミノ酸配列は、標準

N末端ドメイン配列の少なくとも1つのカチオン性アミノ酸残基が、カチオン性電荷を有さないか、電荷を全く有さないか、またはアニオン性電荷を有するアミノ酸残基で置き換えられている点で、標準N末端ドメイン配列とは異なっている。いくつかの態様において、C末端ドメインはアミノ酸配列を含み、該アミノ酸配列は、標準C末端ドメイン配列の少なくとも1つのカチオン性アミノ酸残基が、カチオン性電荷を有さないか、電荷を全く有さないか、またはアニオン性電荷を有するアミノ酸残基で置き換えられている点で、標準C末端ドメイン配列とは異なっている。いくつかの態様において、N末端ドメインおよび/またはC末端ドメインにおいて、少なくとも1個の、少なくとも2個の、少なくとも3個の、少なくとも4個の、少なくとも5個の、少なくとも6個の、少なくとも7個の、少なくとも8個の、少なくとも9個の、少なくとも10個の、少なくとも11個の、少なくとも12個の、少なくとも13個の、少なくとも14個の、または少なくとも15個のカチオン性アミノ酸は、カチオン性電荷を有さないか、電荷を全く有さないか、またはアニオン性電荷を有するアミノ酸残基で置き換えられている。いくつかの態様において、少なくとも1つのカチオン性アミノ酸残基は、アルギニン(R)またはリジン(K)である。いくつかの態様において、カチオン性アミノ酸を置き換えるアミノ酸残基は、グルタミン(Q)またはグリシン(G)である。本開示の側面に従って置換され得るC末端ドメインの正に荷電した残基としては、限定はされないが、アルギニン(R)残基およびリジン(K)残基、例えば、C末端ドメインのR747、R770、K777、K778、K788、R789、R792、R793、R797、およびR801である(例えば配列番号22を参照、番号付けは完全長TALENタンパク質のそれぞれの残基の位置を示し、配列番号22に提供されるC末端ドメインの同等の位置は、R8、R30、K37、K38、K48、R49、R52、R53、R57、R61である)。本開示の側面に従って置換され得るN末端ドメインの正に荷電した残基としては、限定はされないが、アルギニン(R)残基およびリジン(K)残基、例えば、K57、K78、R84、R97、K110、K113、およびR114(例えば、配列番号1を参照)が挙げられる。いくつかの態様において、少なくとも1つのリジンまたはアルギニン残基は、グルタミン残基で置き換えられている。いくつかの態様において、C末端ドメインは、1または2以上の次のアミノ酸置換：K777Q、K778Q、K788Q、R789Q、R792Q、R793Q、R801Q、を含む。いくつかの態様において、C末端ドメインは、Q3バリエーション配列(K788Q、R792Q、K801Q)を含む。いくつかの態様において、C末端ドメインは、Q7バリエーション配列(K777Q、K778Q、K788Q、R789Q、R792Q、R793Q、R801Q)を含む。いくつかの態様において、N末端ドメインは、標準N末端ドメインのランケット型である。いくつかの態様において、C末端ドメインは、標準C末端ドメインのランケット型である。いくつかの態様において、ランケットされたドメインは、標準ドメインの残基の90%未満、80%未満、70%未満、60%未満、50%未満、40%未満、30%未満、または25%未満を含む。いくつかの態様において、ランケットされたC末端ドメインは、60個未満、50個未満、40個未満、30個未満、29個未満、28個未満、27個未満、26個未満、25個未満、24個未満、23個未満、22個未満、21個未満、または20個未満のアミノ酸残基を含む。いくつかの態様において、ランケットされたC末端ドメインは、60、59、58、57、56、55、54、53、52、51、50、49、48、47、46、45、44、43、42、41、40、39、38、37、36、35、34、33、32、31、30、29、28、27、26、25、24、23、22、21、20、19、18、17、16、15、14、13、12、11、または10個の残基を含む。いくつかの態様において、ヌクレアーゼ切断ドメインは、FokIヌクレアーゼドメインである。いくつかの態様において、FokIヌクレアーゼドメインは、配列番号26~30で提供される配列を含む。いくつかの態様において、TALENは単量体である。いくつかの態様において、TALEN単量体は、別のTALEN単量体と二量化してTALEN二量体を形成する。いくつかの態様において、二量体は、ヘテロ二量体である。いくつかの態様において、TA

10

20

30

40

50

LENは、疾患または障害に関連することが知られている遺伝子内の標的配列に結合する。いくつかの態様において、TALENは、二量体化の際に標的配列を切断する。いくつかの態様において、処置または予防される疾患は、HIV感染またはAIDS、または増殖性疾患である。いくつかの態様において、TALENは、HIV感染またはAIDSの処置または予防において、CCR5 (C-Cケモカイン受容体5型) 標的配列に結合する。いくつかの態様において、TALENは、ATM (毛細血管拡張性運動失調症変異) 標的配列に結合する。いくつかの態様において、TALENは、VEGFA (血管内皮増殖因子A) 標的配列に結合する。

【0010】

本開示のいくつかの側面は、本明細書に記載のTALENを、例えばTALEN単量体を含む、組成物を提供する。いくつかの態様において、組成物は、本発明のTALEN単量体および、ヘテロ二量体を形成する別のTALEN単量体を含み、ここで二量体は、ヌクレアーゼ活性を示す。いくつかの態様において、組成物は医薬組成物である。

【0011】

本開示のいくつかの側面は、本明細書に提供されるTALENを含む組成物を提供する。いくつかの態様において、組成物は、*in vitro*で細胞または組織と接触させるのに好適になるように処方される。いくつかの態様において、医薬組成物は、例えば、*in vitro*または*ex vivo*での細胞または組織内の標的配列を切断するための、TALENの有効量を含む。いくつかの態様において、TALENは、目的の遺伝子内の標的配列、例えば疾患または障害と関連することが知られている遺伝子内の標的配列に結合し、および組成物は、疾患または障害に関連する徴候および/または症状を軽減するための、TALENの有効量を含む。本開示のいくつかの側面は、本明細書に提供されるTALENおよび薬学的に許容し得る賦形剤を含む、医薬組成物を提供する。いくつかの態様において、医薬組成物は、対象への投与のために処方される。いくつかの態様において、医薬組成物は、対象における細胞の標的配列を切断するためのTALENの有効量を含む。いくつかの態様において、TALENは、疾患または障害と関連することが知られている遺伝子内の標的配列に結合し、組成物は、疾患または障害に関連する徴候および/または症状を緩和するための、TALENの有効量を含む。

【0012】

本開示のいくつかの側面は、本明細書に提供されるTALENを用いて、核酸分子内の標的配列を切断する方法を提供する。いくつかの態様において、方法は、標的配列を含む核酸分子を、標的配列に結合する本発明のTALENと、TALENが標的配列に結合して切断するのに好適な条件下で、接触させることを含む。いくつかの態様において、TALENは、単量体として提供される。いくつかの態様において、本発明のTALEN単量体は、本発明のTALEN単量体と二量体化してヌクレアーゼ活性を有するヘテロ二量体を形成することができる異なるTALEN単量体を含む、組成物中で提供される。いくつかの態様において、本発明のTALENは、医薬組成物中で提供される。いくつかの態様において、標的配列は、細胞のゲノム内にある。いくつかの態様において、標的配列は、対象の中にある。いくつかの態様において、方法は、TALENを含む例えば医薬組成物などの組成物を、対象に対して、TALENが標的部位に結合して切断するのに十分な量で、投与することを含む。

【0013】

本開示のいくつかの側面は、操作されたTALENを製造する方法を提供する。いくつかの態様において、方法は、標準N末端TALENDメインおよび/または標準C末端TALENDメインの少なくとも1個のアミノ酸を、置き換えられるアミノ酸と比較して生理的pHで電荷を有さないかまたは負電荷を有するアミノ酸で置き換えること；および/または、N末端TALENDメインおよび/またはC末端TALENDメインをトランケートして、正に荷電した断片を除去すること；したがって、生理的pHで減少した正味電荷のN末端ドメインおよび/またはC末端ドメインを有する、操作されたTALENを生成すること、を含む。いくつかの態様において、置換されている少なくとも1つのアミノ

10

20

30

40

50

酸は、カチオン性アミノ酸または生理的 pH で正電荷を有するアミノ酸を含む。本開示の側面に従って置換され得る C 末端ドメインの正に荷電した残基としては、限定はされないが、アルギニン (R) 残基およびリジン (K) 残基、例えば、C 末端ドメインの R 7 4 7、R 7 7 0、K 7 7 7、K 7 7 8、K 7 8 8、R 7 8 9、R 7 9 2、R 7 9 3、R 7 9 7、および R 8 0 1 が挙げられる。本開示の側面に従って置換され得る N 末端ドメインの正に荷電した残基としては、限定はされないが、アルギニン (R) 残基およびリジン (K) 残基、例えば、K 5 7、K 7 8、R 8 4、R 9 7、K 1 1 0、K 1 1 3、および R 1 1 4 が挙げられる。いくつかの態様において、少なくとも 1 つのアミノ酸を置き換えるアミノ酸は、カチオン性アミノ酸または中性のアミノ酸である。いくつかの態様において、トランケートされた N 末端 T A L E N ドメインおよび / またはトランケートされた C 末端 T A L E N ドメインは、それぞれの標準ドメインの残基の 9 0 % 未満、8 0 % 未満、7 0 % 未満、6 0 % 未満、5 0 % 未満、4 0 % 未満、3 0 % 未満、または 2 5 % 未満を含む。いくつかの態様において、トランケートされた C 末端ドメインは、6 0 個未満、5 0 個未満、4 0 個未満、3 0 個未満、2 9 個未満、2 8 個未満、2 7 個未満、2 6 個未満、2 5 個未満、2 4 個未満、2 3 個未満、2 2 個未満、2 1 個未満、または 2 0 個未満のアミノ酸残基を含む。いくつかの態様において、トランケートされた C 末端ドメインは、6 0、5 9、5 8、5 7、5 6、5 5、5 4、5 3、5 2、5 1、5 0、4 9、4 8、4 7、4 6、4 5、4 4、4 3、4 2、4 1、4 0、3 9、3 8、3 7、3 6、3 5、3 4、3 3、3 2、3 1、3 0、2 9、2 8、2 7、2 6、2 5、2 4、2 3、2 2、2 1、2 0、1 9、1 8、1 7、1 6、1 5、1 4、1 3、1 2、1 1、または 1 0 個のアミノ酸残基を含む。いくつかの態様において、方法は、標準 N 末端 T A L E N ドメインおよび / または標準 C 末端 T A L E N ドメインの少なくとも 2 個、少なくとも 3 個、少なくとも 4 個、少なくとも 5 個、少なくとも 6 個、少なくとも 7 個、少なくとも 8 個、少なくとも 9 個、少なくとも 1 0 個、少なくとも 1 1 個、少なくとも 1 2 個、少なくとも 1 3 個、少なくとも 1 4 個、または少なくとも 1 5 個のアミノ酸を、生理的 pH で電荷を有さないかまたは負電荷を有するアミノ酸で置き換えることを含む。いくつかの態様において、置き換えられるアミノ酸は、アルギニン (R) またはリジン (K) である。いくつかの態様において、生理的 pH で電荷を有さないかまたは負電荷を有するアミノ酸は、グルタミン (Q) またはグリシン (G) である。いくつかの態様において、方法は、少なくとも 1 つのリジンまたはアルギニン残基を、グルタミン残基で置き換えることを含む。

【 0 0 1 4 】

本開示のいくつかの側面は、本明細書に提供される操作された T A L E N、またはかかる T A L E N を含む組成物 (例えば、医薬組成物) を含む、キットを提供する。いくつかの態様において、キットは、T A L E N を賦形剤と接触させて、核酸を T A L E N と接触させるのに好適な組成物を生成するための、賦形剤および指示書を含む。いくつかの態様において、賦形剤は薬学的に許容可能な賦形剤である。

【 0 0 1 5 】

上記の概要は、本明細書に開示された技術のいくつかの態様、利点、特徴、および使用について、非限定的に説明することを意図する。本明細書に開示された技術の他の態様、利点、特徴、および使用は、詳細な説明、図面、例、および特許請求の範囲から明らかである。

【図面の簡単な説明】

【 0 0 1 6 】

【図 1】T A L E N 構築物 (architecture) および選択スキーム。(A) T A L E N の構築物。T A L E N 単量体は、N 末端ドメインに続いて T A L E 反復のアレイ、C 末端ドメイン、および F o k I ヌクレアーゼ切断ドメインを含む。各 T A L E 反復の 1 2 番目と 1 3 番目のアミノ酸 (R V D、赤) は、特定の DNA 塩基対を認識する。2 つの異なる T A L E N は、それらに対応する半部位に結合し、F o k I の二量体化および DNA 切断を可能にする。本研究で用いた C 末端ドメインバリエーションが示されている。(B) 部分的にラ

ンダム化された左半部位 (L)、スペーサー (S)、右半部位 (R) および定常領域 (太い黒線) を含む DNA オリゴヌクレオチドの一本鎖ライブラリーを環状化し、ローリングサークル増幅によりコンカテマー化した。得られた DNA ライブラリーを、目的の *in vitro* 翻訳した TALEN と共にインキュベートした。切断されたライブラリーメンバーを平滑化し、アダプター # 1 にライゲートした。ライゲーション産物は、アダプター # 1 からなる 1 つのプライマーと、アダプター # 2 - 定常配列からなる、定常領域にアニーリングする別のプライマーを使用して、PCR により増幅した。長さが

【数 1】

1½

10

の標的配列カセットのアンプリコンをゲル精製によって単離し、ハイスループット DNA シーケンシングおよびコンピューター分析に供した。

【0017】

【図 2】*in vitro* 選択の結果。選択を耐えた配列 (灰色) および選択前配列 (黒) の割合を、両方の半部位における変異数の関数として、CCR5A TALEN (A) および ATM TALEN (B) について示す。 (C) 標的半部位における全ての位置、および単一の隣接位置における、L18 + R18 CCR5A TALEN についての特異性スコア。色は、最大特異性スコアの 1.0 ~ 白 (特異性なし、スコア 0) ~ 最大の負のスコアの -1.0 の範囲である。枠内の塩基は、意図された標的塩基を表す。 (D) L18 + R18 ATM TALEN について、 (C) と同様。 (E) 16 の変異体 DNA 配列 (変異は赤色) について、L13 + R13 CCR5B TALEN の選択からの、オンターゲット DNA (OnB) と比較した濃縮値 (enrichment value)。 (F) オンターゲット切断 (= 1) に対して正規化された、 (E) にリストされた配列についての、*in vitro* の個別の TALEN 切断効率 (切断された DNA の、全 DNA に対する比率) と、オンターゲット濃縮値 (= 1) に対して正規化された、選択におけるそれらの濃縮値の間の対応。 (G) (F) で使用されたオンターゲットおよびオフターゲット配列の、PAGE によって分析された個別アッセイ。

20

【0018】

【図 3】オンターゲットおよび予測されたオフターゲットゲノム部位での、TALEN により誘導された細胞修飾。 (A) TALEN なし、またはヘテロ二量体 EL / K K、ヘテロ二量体 EL D / K K R、もしくはホモ二量体 (Homo) FokI バリエーションを含む CCR5A TALEN のいずれかで処置した細胞について、細胞修飾率を示したものであり、該細胞修飾率は、オンターゲット部位 (On) および予測オフターゲット部位 (Off) について、配列の総数に相対的な、TALEN 切断に整合して観察された挿入または欠失 (インデル) の数のパーセンテージとして示される。 (B) ATM TALEN について、 (A) と同様。 (C) EL D / K K R FokI ドメインを含む CCR5A TALEN で処置した細胞について、オンターゲット部位およびオフターゲット部位での修飾された配列の例。示された各例について、未修飾のゲノム部位が最初の配列であり、続いて欠失を含むトップ 3 の配列が示される。括弧内の数字はシーケンシングの数を表し、半部位は下線太字で示す。

30

40

【0019】

【図 4】ヒトゲノムにおける、TALEN 特異性およびオフターゲット部位の存在量の両方を考慮した、TALEN 長の関数としての予測されたオフターゲットゲノム切断。 (A) オンターゲット (ゼロ変異) および 1 ~ 6 個の変異を含むオフターゲット配列の濃縮値を、様々な TALE 反復アレイ長さの CCR5B TALEN について示す。TALEN は、長さが 32 bp (L16 + R16)、29 bp (L16 + R13 または L13 + R16)、26 bp (L16 + R10 または L13 + R13 または L10 + R16)、23 bp (L13 + R10 または L10 + R13) または 20 bp (L10 + R10) の DNA 部位を標的とした。 (B) 9 個の CCR5B オンターゲット配列 (L10、L13、または L16 を、R10、R13、または R16 と組み合わせたもの) の各々に関連するヒト

50

ゲノム中の部位の数、これにより、2つの半部位の間に12～25bpのスペーサー長が可能となる。(C)全9個のCCR5BTALENについて、全体的なゲノムのオフターゲット切断頻度を、一定数の変異を含むヒトゲノム中の部位の数に、(A)に示された同数の変異を含むオフターゲット配列の濃縮値を掛け合わせることでより予測した。おそらくは選択感度の限界により、濃縮値は高い変異数において横ばいとなるため、高い変異濃縮値は、濃縮値を変異数の関数として近似することにより、外挿する必要があった(表9)。全体的な予測のゲノム切断を、ヒトゲノムで1回より多く発生することが観測された部位の変異数についてのみ、算出した。

【0020】

【図5】標準または操作されたC末端ドメインを含むTALENの、*in vitro*特異性および個別切断効率。(AおよびB)標準、Q3、Q7、または28-aaC末端ドメインを含む(A)CCR5ATALENまたは(B)ATMTALENの選択についての、オンターゲット濃縮値。(C)CCR5Aオンターゲット配列(OnC)と小文字で示す変異を有する二重変異配列。(D)ATMオンターゲット配列(OnC)と小文字で示す変異を有する単一変異配列。(E)標準または操作されたQ7C末端ドメインのいずれかを含むCCR5ATALENを用いた、(C)にリストされたDNA配列の、個別の*in vitro*切断効率。(F)ATMTALENについて、(E)と同様。

【0021】

【図6】ヒト細胞における操作されたTALENの特異性。標準および操作されたTALENの細胞修飾効率を、全配列からのTALEN誘導性修飾に整合するインデルのパーセンテージとして表し、これを、オンターゲットCCR5A配列および、最も高度に切断された試験されたオフターゲット基質である、CCR5Aオフターゲット部位#5について示す。オンターゲットのオフターゲット修飾に対する比率として定義される細胞特異性を、各バーの対の下に示す。

【0022】

【図7】ヒトCCR5およびATM遺伝子における標的DNA配列。本研究で使用するTALENのための標的DNA配列を、黒で示す。各半部位標的の5'Tを認識するN末端TALEN末端に注目し(5')、TALENは、標的とされた塩基対の数に応じて命名される。示されたCCR5L18とR18を標的とするTALENを、CCR5ATALENと呼び、一方、示されたL10、L13、L16、R10、R13またはR16半部位を標的とするTALENを、CCR5BTALENと呼ぶ。

【0023】

【図8】ヒートマップにおける、すべてのCCR5ATALENの選択からの特異性プロファイル。CCR5ATALENの選択内のすべての標的化塩基対に対する特異性スコアが示される。標的半部位における全ての位置、および単一の隣接位置における、L18+R18CCR5ATALENについての特異性スコア。色は、最大特異性スコアの1.0～白(スコア0、特異性なし)～最大の負のスコアの-1.0の範囲である。枠内の塩基は、意図された標的塩基を表す。右側のタイトルは、選択に使用されるTALENが、標準TALEN構築物(これは標準C末端ドメイン、野生型N末端ドメイン、およびEL/KK FokIパリアントを含む)と異なるかどうかを示す。選択は、表2にリストされた条件に対応する。(A)標準、Q3、Q7、28-aa、32nM標準、8nM標準、4nM標準、32nMQ7、および8nMQ7のCCR5ATALEN選択の、特異性プロファイル。(B)4nMQ7、N1、N2、N3、標準ELD/KKR、Q3ELD/KKR、Q7ELD/KKRおよびN2ELD/KKRのCCR5ATALEN選択の、特異性プロファイル。指定されていない場合、TALEN濃度は16nMである。

【0024】

【図9】すべてのCCR5ATALEN選択からの、棒グラフで表した特異性プロファイル。CCR5ATALENの選択内のすべての標的化された塩基対に対する特異性スコアが示される。正の特異性スコアは、特異性スコア1.0の完全な特異性までは、その

10

20

30

40

50

位置での他の可能性に対する当該塩基対の濃縮を表す。負の特異性スコアは、 -1.0 の完全な非特異性スコアまでは、その塩基対に対する濃縮を表す。指定された位置は、X軸上に積み重ね棒グラフとしてプロットし（同一位置の複数の指定された塩基対は、短い棒を前側に互いの上にプロットし、端から端には並べない）、一方、指定されない塩基対は、細いグループ化された棒としてプロットした。右側のタイトルは、選択に使用されるT A L E Nが、標準T A L E N構築物（これは標準C末端ドメイン、野生型N末端ドメイン、およびE L / K K F o k Iバリエーションを含む）と異なるかどうかを示す。選択は、表2にリストされた条件に対応する。（A）標準、Q3、Q7、28 - a a、32 nM標準、および8 nM標準のC C R 5 A T A L E N選択の、特異性プロファイル。（B）4 nM標準、32 nM Q7、8 nM Q7、4 nM Q7、N1、およびN2のC C R 5 A T A L E N選択の、特異性プロファイル。（C）N3、標準E L D / K K R、Q3 E L D / K K R、Q7 E L D / K K RおよびN2 E L D / K K RのC C R 5 A T A L E N選択の、特異性プロファイル。指定されていない場合、T A L E N濃度は16 nMである。

【0025】

【図10】ヒートマップにおける、すべてのA T M T A L E Nの選択からの特異性プロファイル。A T M T A L E Nの選択内のすべての標的化塩基対に対する特異性スコアが示される。標的半部位における全ての位置、および単一の隣接位置における、L18 + R18 A T M T A L E Nについての特異性スコア。色は、最大特異性スコアの $1.0 \sim 白$ （スコア0、特異性なし）～最大の負のスコアの -1.0 の範囲である。枠内の塩基は、意図された標的塩基を表す。右側のタイトルは、選択に使用されるT A L E Nが、標準T A L E N構築物（これは標準C末端ドメイン、野生型N末端ドメイン、およびE L / K K F o k Iバリエーションを含む）と異なるかどうかを示す。選択は、表2にリストされた条件に対応する。（A）（12 nM）標準、Q3、（12 nM）Q7、24 nM標準、6 nM標準、3 nM標準、24 nM Q7、および6 nM Q7のA T M T A L E N選択の、特異性プロファイル。（B）N1、N2、N3、標準E L D / K K R、Q3 E L D / K K R、Q7 E L D / K K RおよびN2 E L D / K K RのA T M T A L E N選択の、特異性プロファイル。指定されていない場合、T A L E N濃度は12 nMである。

【0026】

【図11】すべてのA T M T A L E N選択からの、棒グラフで表した特異性プロファイル。A T M T A L E Nの選択内のすべての標的化された塩基対に対する特異性スコアが示される。正の特異性スコアは、特異性スコア 1.0 の完全な特異性までは、その位置での他の可能性に対する当該塩基対の濃縮を表す。負の特異性スコアは、 -1.0 の完全な非特異性スコアまでは、その塩基対に対する濃縮を表す。指定された位置は、X軸上に積み重ね棒グラフとしてプロットし（同一位置の複数の指定された塩基対は、短い棒を前側に互いの上にプロットし、端から端には並べない）、一方、指定されない塩基対は、細いグループ化された棒としてプロットした。右側のタイトルは、選択に使用されるT A L E Nが、標準T A L E N構築物（これは標準C末端ドメイン、野生型N末端ドメイン、およびE L / K K F o k Iバリエーションを含む）と異なるかどうかを示す。選択は、表2にリストされた条件に対応する。（A）標準、Q3、Q7、28 - a a、32 nM標準、および8 nM標準のA T M T A L E N選択の、特異性プロファイル。（B）3 nM標準、24 nM Q7、6 nM Q7、N1、N2およびN3のA T M T A L E N選択の、特異性プロファイル。（C）標準E L D / K K R、Q3 E L D / K K R、Q7 E L D / K K RおよびN2 E L D / K K RのA T M T A L E N選択の、特異性プロファイル。指定されていない場合、T A L E N濃度は12 nMである。

【0027】

【図12】ヒートマップにおける、すべてのC C R 5 B T A L E Nの選択からの特異性プロファイル。C C R 5 B T A L E Nの選択内のすべての標的化塩基対に対する特異性スコアが示される。標的半部位における全ての位置、および単一の隣接位置における、左（L10、L13、L16）および右（R10、R13、R16）半部位のすべての可能

10

20

30

40

50

な組み合わせを標的とする C C R 5 B T A L E N についての特異性スコア。色は、最大特異性スコアの 1 . 0 ~ 白 (スコア 0、特異性なし) ~ 最大の負のスコアの - 1 . 0 の範囲である。枠内の塩基は、意図された標的塩基を表す。右側のタイトルは、選択に使用される C C R 5 B T A L E N についての標的化された左 (L) および右 (R) 標的半部位を示す。選択は、表 2 にリストされた条件に対応する。

【 0 0 2 8 】

【図 1 3】すべての C C R 5 B T A L E N 選択からの、棒グラフで表した特異性プロファイル。C C R 5 B T A L E N の選択内のすべての標的化された塩基対に対する特異性スコアが示される。正の特異性スコアは、特異性スコア 1 . 0 の完全な特異性までは、その位置での他の可能性に対する当該塩基対の濃縮を表す。負の特異性スコアは、 - 1 . 0 の完全な非特異性スコアまでは、その塩基対に対する濃縮を表す。指定された位置は、X 軸上に積み重ね棒グラフとしてプロットし (同一位置の複数の指定された塩基対は、短い棒を前側にして互いの上にプロットし、端から端には並べない)、一方、指定されない塩基対は、細いグループ化された棒としてプロットした。右側のタイトルは、選択に使用される C C R 5 B T A L E N についての標的化された左 (L) および右 (R) 標的半部位を示す。選択は、表 2 にリストされた条件に対応する。

【 0 0 2 9 】

【図 1 4】二重変異体配列の予測された濃縮値に対する観測された濃縮値。(A) L 1 3 + R 1 3 C C R 5 A T A L E N の選択について、個別の配列の観測された二重変異体濃縮値 (選択後配列の存在量 ÷ 選択前配列の存在量) を、オンターゲット濃縮値 (定義により = 1 . 0) について正規化し、オンターゲット濃縮値について正規化された成分の単一変異体の濃縮値を乗じることにより算出した予測二重変異体濃縮値に対して、プロットした。したがって予測二重変異体濃縮値は、各単一変異から二重変異体濃縮値に対して、独立した寄与を果たす。(B) 観測された二重変異体配列濃縮を、予測二重変異体配列濃縮で割り算し、2 つの変異の間の距離 (塩基対の単位) の関数としてプロットしたもの。同一の半部位に 2 つの変異を有する配列のみを検討した。

【 0 0 3 0 】

【図 1 5】操作された T A L E N ドメインおよび T A L E N 濃度の、特異性に対する効果。(A) C C R 5 A 部位の各位置での標的塩基対の特異性スコアを、標準、Q 3、Q 7、または 2 8 - a a C 末端ドメインを含有する C C R 5 A T A L E N について算出した。Q 3、Q 7、または 2 8 - a a C 末端ドメイン T A L E N の特異性スコアから、標準 C 末端ドメインを有する T A L E N の特異性スコアを引き算したものを示す。(B) 操作された N 末端ドメイン N 1、N 2、または N 3 を含有する C C R 5 A T A L E N について、(A) と同様。(C) (A) と同様だが、1 6 n M、8 n M または 4 n M にてアッセイされた標準 C C R 5 A T A L E N の特異性スコアから、3 2 n M にてアッセイされた標準 C C R 5 A T A L E N の特異性スコアを引き算したものの差を比較。(D ~ F) A T M T A L E N について、(A ~ C) と同様。選択は、表 2 に記載された条件に対応する。

【 0 0 3 1 】

【図 1 6】T A L E N のスペーサー長の選好。(A) C C R 5 A T A L E N であって、次の様々な組み合わせを含有するもの：標準、Q 3、Q 7、または 2 8 - a a C 末端ドメイン；N 1、N 2、または N 3 の N 末端変異；および、4、8、1 6 または 3 2 n M での E L / K K または E L D / K K R の F o k I バリエーション、による各選択に対して、DNA スペーサー長の濃縮値を、選択後配列中の DNA スペーサー長の存在量を選択前ライブラリー配列の DNA スペーサー長の存在量で割り算することにより、算出した。(B) A T M T A L E N について、(A) と同様。

【 0 0 3 2 】

【図 1 7】T A L E N の DNA 切断部位の選好。(A) C C R 5 A T A L E N であって、次の様々な組み合わせを含有するもの：標準の、Q 3、Q 7、または 2 8 - a a C 末端ドメイン；N 1、N 2、または N 3 の N 末端変異；および、4、8、1 6 または 3 2 n M での E L / K K または E L D / K K R の F o k I バリエーション、の各選択に対して、各可能

なDNAスペーサー長についての右半部位に先行するスペーサーDNA塩基対の数のヒストグラムを、選択全体の総配列カウントに対して正規化して示す。(B)ATMTALENについて、(A)と同様。

【0033】

【図18】異なるアミノ酸置換を有するN末端ドメインを含むTALENの、DNA切断部位の選好。

【0034】

【図19】例示のTALENプラスミドコンストラクト。

【0035】

定義

10

本明細書および特許請求の範囲で使用される場合、単数形「a」、「an」および「the」は、文脈が明らかに他を示さない限り、単数形と複数形の言及を含む。したがって例えば、「薬剤」への言及は、1つの薬剤および複数の薬剤を含む。

【0036】

本明細書で使用する用語「標準配列」は、DNA、RNA、またはアミノ酸の配列であって、その種類の知られている分子の中で各位置における塩基またはアミノ酸の最も一般的な選択を反映している前記配列を指す。例えば、タンパク質ドメインの標準アミノ酸配列は、その種類の知られている全てのドメイン、またはその種類の知られている主要なドメインの各位置に存在するアミノ酸の、最も一般的な選択を反映し得る。いくつかの態様において、標準配列はコンセンサス配列である。

20

【0037】

核酸配列の文脈において本明細書中で使用される用語「コンセンサス配列」および「コンセンサス部位」は、複数の類似の配列の各位置で見出される最も頻繁なヌクレオチド残基を表す、算出された配列を指す。典型的には、コンセンサス配列は、類似の配列が互いに比較され、類似の配列モチーフが算出される、配列アラインメントによって決定される。ヌクレアーゼ標的部位配列の文脈において、ヌクレアーゼ標的部位のコンセンサス配列は、いくつかの態様において、所与のヌクレアーゼによって最も頻繁に結合された、最も高い親和性で結合された、および/または最高の効率で切断された、配列であってよい。

【0038】

用語「コンジュゲートする」、「コンジュゲートされた」および「コンジュゲーション」とは、2つの実体の関連を指し、2つの実体とは例えば、2つのタンパク質などの2つの分子、2つのドメイン(例えば、結合ドメインと切断ドメイン)、またはタンパク質と薬剤(例えば、タンパク質結合ドメインと小分子)である。関連とは、例えば、直接的もしくは間接的な(例えば、リンカーを介した)共有結合を介するものか、または非共有結合性相互作用を介するものであり得る。いくつかの態様において、関連は共有結合性である。いくつかの態様において、2つの分子は、両方の分子を連結するリンカーを介してコンジュゲートされている。例えば、2つのタンパク質がお互いにコンジュゲートされて(例えば操作されたヌクレアーゼの結合ドメインと切断ドメインなど)タンパク質融合体を形成するいくつかの態様においては、2つのタンパク質は、ポリペプチドリナーを介して、例えば、1つのタンパク質のC末端を別のタンパク質のN末端に連結するアミノ酸配列を介して、コンジュゲートされてもよい。

30

40

【0039】

本明細書で使用する用語「有効量」は、所望の生物学的応答を誘発するのに十分な、生物学的に活性な薬剤の量を意味する。例えば、いくつかの態様において、TALENヌクレアーゼの有効量は、例えば、無細胞アッセイにおいて、または標的細胞、組織、もしくは生物体において、ヌクレアーゼにより特異的に結合され切断される標的部位の切断を誘導するのに十分な、ヌクレアーゼの量を指してもよい。当業者によって理解されるように、薬剤、例えば、ヌクレアーゼ、ハイブリッドタンパク質、またはポリヌクレオチドの有効量は、種々の要因に応じて、例えば、所望の生物学的応答、特定のアレル、ゲノム、標的部位、細胞、または標的とされる組織、および使用される薬剤に応じて、変化し得る。

50

【0040】

本明細書で使用する用語「操作された」とは、ヒトによって設計され、生産され、製造され、合成され、および／または製作される、分子、複合体、物質、または実体を指す。したがって、操作された生成物は、自然界に存在しない生成物である。いくつかの態様において、操作された分子または複合体、例えば、操作されたT A L E N単量体、二量体、または多量体は、特定の要件を満たすために、または特定の所望の特徴を有するように設計されたT A L E Nであり、例えば、最小限のオフターゲット結合で目的の標的配列に結合するため、特定の最小または最大の切断活性を有するため、および／または特定の安定性を有するために、設計されたT A L E Nである。

【0041】

本明細書中で使用される用語「単離された」とは、分子、複合体、物質、または実体であって、(1)(天然であれ、実験設定においてであれ)最初に生成された時にそれが関連していた成分の少なくともいくつかから分離されたもの、および／または(2)ヒトによって生産、製造、合成、および／または製作されたもの、を指す。単離された物質および／または実体は、最初に関連していた他の成分の、少なくとも約10%、約20%、約30%、約40%、約50%、約60%、約70%、約80%、約90%、またはそれ以上から、分離することができる。いくつかの態様において、単離された薬剤は、約80%より高く、約85%、約90%、約91%、約92%、約93%、約94%、約95%、約96%、約97%、約98%、約99%、または約99%より高く、純粋である。本明細書で使用される場合、物質は、それが他の成分を実質的に含まない場合に「純粋」である。

【0042】

核酸またはタンパク質の文脈において本明細書で使用する用語「ライブラリー」はそれぞれ、2または3以上の異なる核酸またはタンパク質の集団を指す。例えば、ヌクレアーゼ標的部位のライブラリーは、異なるヌクレアーゼ標的部位を含む少なくとも2つの核酸分子を含む。いくつかの態様において、ライブラリーは、少なくとも 10^1 、少なくとも 10^2 、少なくとも 10^3 、少なくとも 10^4 、少なくとも 10^5 、少なくとも 10^6 、少なくとも 10^7 、少なくとも 10^8 、少なくとも 10^9 、少なくとも 10^{10} 、少なくとも 10^{11} 、少なくとも 10^{12} 、少なくとも 10^{13} 、少なくとも 10^{14} 、または少なくとも 10^{15} の、異なる核酸またはタンパク質を含む。いくつかの態様において、ライブラリーのメンバーはランダム化配列、例えば、完全にまたは部分的にランダム化された配列を含んでよい。いくつかの態様において、ライブラリーは、互いに関連しない核酸分子を、例えば完全なランダム化配列を含む核酸を含む。他の態様において、ライブラリーの少なくともいくつかのメンバーは関連することができ、例えば、コンセンサス標的部位配列などの特定の配列のバリエーションまたは誘導体であり得る。

【0043】

本明細書中で使用される用語「リンカー」は、2つの分子または部分、例えばヌクレアーゼの結合ドメインおよび切断ドメインを連結する、化学基または分子を指す。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、またはそれらに隣接して、それぞれを共有結合を介して連結し、したがって両者を結び付ける。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、ペプチドまたはタンパク質)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。

【0044】

用語「ヌクレアーゼ」は、本明細書において、例えばタンパク質または小分子などの薬剤であって、核酸分子中のヌクレオチド残基を連結するホスホジエステル結合を切断することができるものを指す。いくつかの態様において、ヌクレアーゼはタンパク質であり、例えば、核酸分子に結合でき、核酸分子内のヌクレオチド残基を連結するリン酸ジエステル結合を切断することができる、酵素である。ヌクレアーゼは、ポリヌクレオチド鎖内のホスホジエステル結合を切断するエンドヌクレアーゼ、またはポリヌクレオチド鎖の末端

でホスホジエステル結合を切断するエキソヌクレアーゼであってよい。いくつかの態様において、ヌクレアーゼは、特定のヌクレオチド配列内の特定のホスホジエステル結合を結合および/または切断する、部位特異的ヌクレアーゼであり、これは本明細書において、「認識配列」、「ヌクレアーゼ標的部位」または「標的部位」とも呼ばれる。いくつかの態様において、ヌクレアーゼは一本鎖標的部位を認識し、一方他の態様において、ヌクレアーゼは、例えば二本鎖DNA標的部位などの二本鎖標的部位を認識する。多くの天然に存在するヌクレアーゼの標的部位、例えば、多くの天然に存在するDNA制限ヌクレアーゼは、当業者に知られている。多くの場合、EcoRI、HindIII、またはBamHIなどのDNAヌクレアーゼは、パリンドロームの長さ4～10塩基対の二本鎖DNAの標的部位を認識し、2つのDNA鎖の各々を標的部位内の特定の位置で切断する。いくつかのエンドヌクレアーゼは、二本鎖核酸の標的部位を対称的に切断し、すなわち、両方の鎖を同じ位置で切断して、末端が塩基対ヌクレオチドを含むようにし、これは本明細書で平滑末端とも称される。他のエンドヌクレアーゼは、二本鎖核酸の標的部位を非対称的に切断し、すなわち、両方の鎖を異なる位置で切断して、末端が不對ヌクレオチドを含むようにする。二本鎖DNA分子の末端の不對ヌクレオチドはまた、「オーバーハング」とも呼ばれ、例えば、不對ヌクレオチド(単数または複数)がそれぞれのDNA鎖の5'または5'末端を形成するかどうかに依存して、「5'-オーバーハング」または「3'-オーバーハング」と呼ばれる。不對ヌクレオチドで終了する二本鎖DNA分子は、粘着(スティッキー)末端とも呼ばれ、何故なればそれらが、相補的な不對ヌクレオチド(単数または複数)を含む他の二本鎖DNA分子末端に粘着するからである。ヌクレアーゼタンパク質は、典型的には、タンパク質と核酸基質との相互作用を媒介する「結合ドメイン」、および核酸骨格内のホスホジエステル結合の切断を触媒する「切断ドメイン」を含む。いくつかの態様において、ヌクレアーゼタンパク質は、単量体の形態で、核酸分子に結合し切断することができ、他の態様において、ヌクレアーゼタンパク質は、標的核酸分子を切断するために、二量体化または多量体化しなければならない。天然に存在するヌクレアーゼの結合ドメインおよび切断ドメイン、ならびに特定の標的部位に結合するヌクレアーゼを生成するために組み合わせることができる、モジュラー結合ドメインおよび切断ドメインは、当業者に知られている。例えば、転写活性化因子様要素を、所望の標的部位に特異的に結合する結合ドメインとして使用し、切断ドメイン、例えばFokIの切断ドメインに融合またはコンジュゲートして、所望の標的部位を切断する操作されたヌクレアーゼを生成することができる。

【0045】

本明細書中で使用される用語「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指す。典型的には、ポリマー核酸、例えば3または4以上のヌクレオチドを含む核酸分子は、隣接するヌクレオチドがホスホジエステル結合を介して互いに連結されている直鎖状分子である。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)を指す。いくつかの態様において、「核酸」は、3または4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指す。本明細書で使用する場合、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つのヌクレオチドのストリング)を指すために互換的に使用することができる。いくつかの態様において、「核酸」は、RNAならびに一本鎖および/または二本鎖DNAを包含する。核酸は、天然に存在するものであってよく、例えばゲノムの文脈において、転写産物、mRNA、tRNA、rRNA、siRNA、snRNA、プラスミド、コスミド、染色体、染色分体、または他の天然に存在する核酸分子である。一方、核酸分子は非天然分子であってもよく、例えば組換えDNAまたはRNA、人工染色体、操作されたゲノム、またはそれらの断片、または合成DNA、RNA、DNA/RNAハイブリッド、または非天然のヌクレオチドまたはヌクレオシドを含む。さらに、用語「核酸」、「DNA」、「RNA」および/または類似の用語は、核酸アナログ、すなわちホスホジエステル骨格以外を有する類似体を含む。核酸は、天然源から精

10

20

30

40

50

製される、組換え発現系を用いて生産されて任意に精製される、化学的に合成される、などが可能である。適切な場合、例えば化学的に合成された分子の場合、核酸は、化学的に修飾された塩基または糖、および骨格修飾を有する類似体などの、ヌクレオシド類似体を含むことができる。特に断らない限り、核酸配列は、5'から3'方向に表される。いくつかの態様において、核酸は、以下であるか、またはこれを含む：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）；ヌクレオシド類似体（例えば、2-アミノアデノシン、2-チオチミジン、イノシン、ピロロピリミジン、3-メチルアデノシン、5-メチルシチジン、2-アミノアデノシン、C5-プロモウリジン、C5-フルオロウリジン、C5-ヨードウリジン、C5-プロピニルウリジン、C5-プロピニル-シチジン、C5-メチルシチジン、2-アミノアデノシン、7-デアザアデノシン、7-デアザグアノシン、8-オキソアデノシン、8-オキソグアノシン、O(6)-メチルグアニン、および2-チオシチジン）；化学的修飾塩基；生物学的修飾塩基（例えば、メチル化塩基）；挿入（intercalated）塩基；修飾糖（例えば、2'-フルオロリボース、リボース、2'-デオキシリボース、アラビノース、およびヘキソース）；および/または修飾リン酸基（例えば、ホスホロチオエートおよび5'-N-ホスホロアミダイト結合）。

10

【0046】

本明細書で使用する用語「医薬組成物」は、疾患または障害の処置の文脈において、対象に投与することができる組成物を指す。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、および薬学的に許容し得る賦形剤を含む。

20

用語「予防」または「予防する」とは、疾患、障害、または症状を発症するリスクのある対象（例えば、対照対象または対象の対照群と比較してリスクが上昇したか、年齢を適合および/または性別を適合させた対象の平均的なリスクと比較して、リスクが高い）に対する予防的処置であって、対象が疾患、障害、または症状を発症する確率の低下（予防なしの確率と比較して）をもたらすものを指すか、および/または、既に確立された障害のさらなる進行の抑制を指す。

【0047】

本明細書で使用する用語「増殖性疾患」は、細胞または組織の恒常性が妨げられて、細胞または細胞集団が異常に高い増殖速度を示す、任意の疾患を指す。増殖性疾患は、前新生物過形成状態および新生物疾患などの、過剰増殖性疾患を含む。新生物疾患は、細胞の異常増殖を特徴とし、良性および悪性両方の新生物を含む。悪性新生物はまた、がんとも呼ぶ。

30

【0048】

用語「タンパク質」、「ペプチド」および「ポリペプチド」は、本明細書中で交換可能に使用され、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のポリマーを指す。この用語は、任意のサイズ、構造、または機能のタンパク質、ペプチド、またはポリペプチドを指す。典型的には、タンパク質、ペプチド、またはポリペプチドは、少なくとも3個のアミノ酸長である。タンパク質、ペプチド、またはポリペプチドは、個々のタンパク質またはタンパク質の集合を指してもよい。タンパク質、ペプチド、またはポリペプチド中の1または2以上のアミノ酸は、例えば炭水化物基、ヒドロキシル基、リン酸基、ファルネシル基、イソファルネシル基、脂肪酸基、およびコンジュゲーション用、官能化用、もしくは他の修飾用のリンカーなどの、化学的実体を添加することによって、修飾されてもよい。タンパク質、ペプチド、またはポリペプチドはまた、単一の分子であってもよく、または多分子複合体であってもよい。タンパク質、ペプチド、またはポリペプチドは、天然に存在するタンパク質またはペプチドのほんの断片であってもよい。タンパク質、ペプチド、またはポリペプチドは、天然の、組換えの、もしくは合成のもの、またはそれらの任意の組み合わせであってもよい。タンパク質は、例えば、核酸結合ドメインおよび核酸切断ドメインなどの、異なるドメインを含んでもよい。いくつかの態様におい

40

50

て、タンパク質は、タンパク質性部分、例えば核酸結合ドメインを構成するアミノ酸配列、および有機化合物、例えば核酸切断剤として作用し得る化合物、を含む。

【0049】

核酸配列の文脈において本明細書で使用する用語「ランダム化」は、遊離ヌクレオチドの混合物を組み込むために、例えば、4つすべてのヌクレオチドA、T、GおよびCの混合物を組み込むために合成された配列内の、配列または残基を指す。ランダム化残基は、典型的には、ヌクレオチド配列内の文字Nによって表される。いくつかの態様において、ランダム化配列または残基は完全にランダム化され、この場合、ランダム化残基を、それぞれの配列残基の合成ステップの間に同量の組み込まれるべきヌクレオチドを添加することによって（例えば、25%のT、25%のA、25%のG、および25%のC）、合成する。いくつかの態様において、ランダム化配列または残基は部分的にランダム化され、この場合、ランダム化残基を、それぞれの配列残基の合成ステップの間に異なる量の組み込まれるべきヌクレオチドを添加することによって（例えば、79%のT、7%のA、7%のG、および7%のC）、合成する。部分的なランダム化は、与えられた配列を鋳型とするが、所望の頻度で変異が組み込まれる配列の生成を可能にする。例えば、周知のヌクレアーゼ標的部位を合成の鋳型として使用する場合、各ステップにおいてそれぞれの残基で表されるヌクレオチドを79%で合成に添加し、残りの3つのヌクレオチドをそれぞれ7%ずつ添加する部分的ランダム化は、合成され部分的にランダム化された標的部位の混合物をもたらす、これは依然として元の標的部位のコンセンサス配列を表すが、元の標的部位とは、各残基において、合成された各残基の統計的頻度が21%で異なっている（二項分布）。いくつかの態様において、部分的ランダム化配列は、平均して5%以上、10%以上、15%以上、20%以上、25%以上、または30%以上の二項分布により、コンセンサス配列とは異なる。いくつかの態様において、部分的ランダム化配列は、平均して10%未満、15%未満、20%未満、25%未満、30%未満、40%未満、または50%未満の二項分布により、コンセンサス配列とは異なる。

【0050】

本明細書において用語「対象」とは、例えば個々の哺乳類などの、個々の生物を指す。いくつかの態様において、対象は、いずれかの性別で発達の任意の段階におけるヒトである。いくつかの態様において、対象は、非ヒト哺乳動物である。いくつかの態様において、対象は、非ヒト霊長類である。いくつかの態様において、対象は、齧歯類である。いくつかの態様において、対象は、ヒツジ、ヤギ、ウシ、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、爬虫類、魚類、昆虫、ハエ、または線虫である。

【0051】

ヌクレアーゼの文脈において本明細書中で使用される用語「標的核酸」および「標的ゲノム」は、それぞれ、与えられたヌクレアーゼの少なくとも1つの標的部位を含む核酸分子またはゲノムを指す。

【0052】

本明細書中で交換可能に使用される「標的部位」および「ヌクレアーゼ標的部位」は、ヌクレアーゼによって結合され切断される、核酸分子内の配列を指す。標的部位は、一本鎖または二本鎖であってもよい。二量体化するヌクレアーゼ、例えばFokI DNA切断ドメインを含むヌクレアーゼの文脈において、標的部位は、典型的には、左半部位（ヌクレアーゼの1つの単量体によって結合される）、右半部位（ヌクレアーゼの第2の単量体によって結合される）、および切断が行われる半分部位の間のスペーサー配列を含む。この構造（[左半部位] - [スペーサー配列] - [右半部位]）は本明細書において、LSR構造と呼ぶ。いくつかの態様において、左半部位および/または右半部位は、10~18のヌクレオチド長である。いくつかの態様において、一方または両方の半部位は、より短いまたはより長い。いくつかの態様において、左右の半部位は、異なる核酸配列を含む。

【0053】

本明細書で使用する用語「転写活性化因子様エフェクター (T A L E)」とは、高度に変な 2 アミノ酸モチーフ (反復可変二残基、R V F) を含む高度に保存された 3 3 ~ 3 4 のアミノ酸配列を含む D N A 結合ドメインを含む、タンパク質を指す。R V D モチーフは、核酸配列に対する結合特異性を決定し、当業者に周知の方法に従って、所望の D N A 配列に特異的に結合するよう、操作することができる。(例えば、以下を参照: Miller, Jeffrey; et.al. (February 2011). "A TALE nuclease architecture for efficient genome editing". Nature Biotechnology 29 (2): 143-8; Zhang, Feng; et.al. (February 2011). "Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription". Nature Biotechnology 29 (2): 149-53; Geiler, R.; Scholze, H.; Hahn, S.; Streubel, J.; Bonas, U.; Behrens, S. E.; Boch, J. (2011), Shiu, Shin-Han. ed. "Transcriptional Activators of Human Genes with Programmable DNA-Specificity". PLoS ONE 6 (5): e19509; Boch, Jens (February 2011). "TALEs of genome targeting". Nature Biotechnology 29 (2): 135-6; Boch, Jens; et.al. (December 2009). "Breaking the Code of DNA Binding Specificity of TAL-Type II Effectors". Science 326 (5959): 1509-12; および Moscou, Matthew J.; Adam J. Bogdanove (December 2009). "A Simple Cipher Governs DNA Recognition by TAL Effectors". Science 326 (5959): 1501; これらの各々の全内容は、参照により本明細書に組み込まれる)。アミノ酸配列と D N A 認識との間の単純な関係は、適切な R V D を含有する反復セグメントの組み合わせを選択することにより、特定の D N A 結合ドメインの操作を可能にした。

【 0 0 5 4 】

本明細書で使用する用語「転写活性化因子様要素ヌクレアーゼ」(T A L E N) は、例えば F o k I ドメインなどの D N A 切断ドメインに対する、転写活性化因子様エフェクター D N A 結合ドメインを含む、人工的ヌクレアーゼを指す。操作された T A L E コンストラクトを生成するための、多くのモジュールアセンブリスキームが報告されている (Zhang, Feng; et.al. (February 2011). "Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription". Nature Biotechnology 29 (2): 149-53; Geiler, R.; Scholze, H.; Hahn, S.; Streubel, J.; Bonas, U.; Behrens, S. E.; Boch, J. (2011), Shiu, Shin-Han. ed. "Transcriptional Activators of Human Genes with Programmable DNA-Specificity". PLoS ONE 6 (5): e19509; Cermak, T.; Doyle, E. L.; Christian, M.; Wang, L.; Zhang, Y.; Schmidt, C.; Baller, J. A.; Somia, N. V. et al. (2011). "Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting". Nucleic Acids Research; Morbitzer, R.; Elsaesser, J.; Hausner, J.; Lahaye, T. (2011). "Assembly of custom TALE-type DNA binding domains by modular cloning". Nucleic Acids Research; Li, T.; Huang, S.; Zhao, X.; Wright, D. A.; Carpenter, S.; Spalding, M. H.; Weeks, D. P.; Yang, B. (2011). "Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes". Nucleic Acids Research; Weber, E.; Gruetzner, R.; Werner, S.; Engler, C.; Marillonnet, S. (2011). Bendahmane, Mohammed. ed. "Assembly of Designer TAL Effectors by Golden Gate Cloning". PLoS ONE 6 (5): e19722; これらの各々の全内容は、参照により本明細書に組み込まれる)。

【 0 0 5 5 】

用語「処置」、「処置する」および「処置すること」とは、本明細書に記載のように、疾患または障害、またはその 1 または 2 以上の症状の、発症を逆転、緩和または遅延させる、またはその進行を阻害するための、臨床的介入を指す。本明細書において、用語「処置」、「処置する」および「処置すること」とは、本明細書に記載のように、疾患または障害、またはその 1 または 2 以上の症状の、発症を逆転、緩和または遅延させる、またはその進行を阻害するための、臨床的介入を指す。いくつかの態様において、処置は、1 または 2 以上の症状が発症した後、および / または疾患が診断された後に行ってもよい。他

の態様において、処置は、症状の不在下で、例えば、症状の発症を予防または遅延させるため、または疾患の発症または進行を阻害するために行ってもよい。例えば、処置は、感受性の高い個人に対して、症状の発症前に行ってもよい（例えば、症状の履歴に照らして、および/または遺伝的または他の感受性因子に照らして）。処置はまた、症状が解消した後に、例えばその再発を予防または遅延させるために、継続してもよい。

【0056】

本発明の特定の態様の詳細な説明

転写活性化因子様エフェクターヌクレアーゼ（T A L E N）は、F o k I制限エンドヌクレアーゼの切断ドメインと、D N A結合転写活性化因子様エフェクター（T A L E）反復アレイとの融合体である。T A L E Nは、オフターゲット切断活性を低減し、したがって標的D N A配列に特異的に結合するように操作することができ、例えばゲノム内、in v i t r oまたはin v i v oで標的D N A配列を切断するために使用することができる。かかる操作されたT A L E Nは、in v i v oまたはin v i t r oでゲノムを操作するために使用することができ、例えば標的ゲノム部位でのD N A切断の誘導を介した遺伝子ノックアウトまたはノックインのために、非相同末端結合（N H E J）を介した標的遺伝子ノックアウトのため、または外来D N A鋳型を使用した相同性配向型修復（H D R）を介した標的ゲノム配列置換のために、使用できる。

【0057】

T A L E Nは、天然および合成配列を含む任意の所望の標的D N A配列を切断するように、設計することができる。しかし、密接に関連するオフターゲット配列から標的配列を区別するT A L E Nの能力は、深く研究されていなかった。この能力およびそれに影響を与えるパラメータを理解することは、その治療的使用のために所望の特異性レベルを有するT A L E Nの設計に、およびまた、オフターゲット切断の可能性を最小化する目的で切断すべきユニークな標的配列を選択するために、重要である。

【0058】

本開示のいくつかの側面は、 10^{1-2} の潜在的なオフターゲット部位でのin v i t r o選択およびハイスループットシーケンシングによる、41のT A L E Nのプロファイリングから得られた切断特異性データに基づく。選択結果のコンピューター分析は、ヒトゲノムにおけるオフターゲット基質を予測し、そのうち13は、ヒト細胞においてT A L E Nにより修飾された。本開示のいくつかの側面は、次の驚くべき知見に基づく：（i）T A L E N反復は、比較的独立してD N Aに結合する；（i i）より長いT A L E Nは、ミスマッチにより寛容であるが、ゲノムの場面でより特異的である；および（i i i）過剰なD N A結合エネルギーは、T A L E N特異性の低減につながり得る。これらの知見に基づき、最適化されたT A L E Nを、非特異的D N A結合を減少させるために設計された変異を用いて操作した。これらの操作されたT A L E Nのいくつかは、ヒト細胞において一般的に使用されるT A L E Nと比較して、例えば34 ~ > 116倍高い改善された特異性を示す。

【0059】

ゲノム中に部位特異的な変化を操作する能力は、重要な治療的意味を持つ強力な研究能力を表す。T A L E Nは、F o k I制限エンドヌクレアーゼ切断ドメインとD N A結合T A L E反復アレイの融合体である（図1A）。これらのアレイは、各々が位置12と13のアミノ酸である反復可変二残基（R V D）を使用して単一のD N Aヌクレオチドを認識する、複数の34アミノ酸T A L E反復配列からなる^{1, 2}。4つのD N A塩基対の各々を認識可能なR V Dの例が知られており、事実上任意のD N A配列に結合することができるT A L E反復のアレイの構築を可能とする。T A L E Nは、真正のヘテロ二量体F o k Iバリエーションの使用を介して、ヘテロ二量体としてのみ活性であるように操作可能である^{3, 4}。この構成において、2つの異なるT A L E N単量体はそれぞれが、1つの標的半部位に結合して2つの半部位の間のD N Aスパーサー配列内で切断するように、設計される。

【0060】

例えば哺乳動物細胞などの細胞において、T A L E Nによって誘導される二重鎖切断は、非相同末端結合(N H E J)を介した標的遺伝子ノックアウト⁵、または、外来DNA鋳型を使用し相同性配向型修復(H D R)を介した標的ゲノム配列置換^{6, 7}をもたらすことができる。T A L E Nは、種々の生物^{8 - 11}および細胞株^{7, 12, 13}においてゲノムを操作するために、首尾良く使用されている。

【0061】

オフターゲット部位でのT A L E N媒介性のDNA切断は、ゲノム座位での意図しない変異をもたらし得る。S E L E X実験は、単量体T A L Eタンパク質のDNA結合特異性を特徴付けたが^{5, 7}、活性な二量体ヌクレアーゼのDNA切断特異性は、それらの成分の単量体DNA結合ドメインの特異性とは異なる可能性がある¹⁴。T A L E Nで処置した4つの酵母株¹⁵および、T A L E Nで処置した細胞由来の2つのヒト細胞株¹⁶の全ゲノムシーケンシングは、T A L E誘発性のゲノムオフターゲット変異の証拠がないことを明らかにし、これは、アフリカツメガエル(Xenopus)¹⁷およびヒト細胞株¹⁸におけるオフターゲットゲノム修飾が観察されないとの他の報告と整合する。対照的に、T A L E Nは、ゼブラフィッシュ^{13, 19}、ラット⁹、ヒト初代線維芽細胞²⁰および胚性幹細胞⁷でのin vivoのオンターゲット配列と比較して、2~11個の変異を含むオフターゲット部位を切断することが観察された。関連する変異標的部位の大規模なセットでのT A L E N切断の測定値から生成される、T A L E Nの特異性の系統のおよび包括的プロファイルは、これまで記載されていない。かかる広い特異性プロファイルは、研究ツールおよび治療剤としてのT A L E Nの可能性を理解し、改善するための、基本である。

【0062】

本明細書に記載される研究の一部は、DNA切断特異性のために以前に記載されたin vitro選択¹⁴の修正版を使用した、41のT A L E N対の能力であって、それらの標的配列の各々の10¹²のオフターゲットバリエーションを切断する前記能力をプロファイルするために行った実験に関する。これらの実験からのこれらの結果は、T A L E N切断特異性の総合的なプロファイルを提供する。in vitroでの選択結果を使用して、ヒトゲノムにおけるオフターゲット基質をコンピューターにより予測し、これらのうち13個が、ヒト細胞においてT A L E Nによって切断されることが確認された。

【0063】

驚くべきことに、塩基対当たりより少なく特異的であるにも関わらず、より長い標的部位を切断するように設計されたT A L E Nは、ヒトゲノムにおける潜在的なオフターゲット部位の数を考慮した場合、より短い部位を標的とするものよりも一般に全体として高い特異性を示すことが判明した。選択結果はまた、過剰な非特異的T A L E N結合エネルギーが、オンターゲットの切断に比べてより大きいオフターゲット切断を生じるというモデルを示唆する。このモデルに基づき、我々は、現在使用されているT A L E Nコンストラクトよりも、in vitroで実質的に改善されたDNA切断特異性を有し、ヒト細胞において30~>150倍高い特異性を有するT A L E Nを操作した。

【0064】

本開示のいくつかの側面は、本明細書の別の箇所により詳細に説明するように、3つの異なる配列の1つを標的とするように設計された41のヘテロ二量体T A L E Nの特異性のプロファイリングから得られたデータに基づく。プロファイリングは、in vitro選択法¹⁴の改善版(P C T出願公開W02013/066438 A2にも記載されており、この全内容は参照により本明細書に組み込まれる)を、選択のスループットと感度を高める修飾を加えて実施した(図1B)。

【0065】

簡単に述べると、T A L E Nは、>10¹²のDNA配列のライブラリーに対してプロファイリングされ、切断産物を捕捉し分析して、各T A L E Nの特異性およびオフターゲット活性を決定した。選択データは、in vitroでのオフターゲットT A L E N切断の効率を正確に予測し、またT A L E Nが、標的配列全体を通して全体的に高度に特異的であることを示したが、従来のT A L E Nではオフターゲット切断のいくつかのレベルが生じ、

これは、T A L E N使用のいくつかのシナリオにおいては望ましくない場合がある。本明細書に記載した実験の結果、驚くべきことに、T A L E N反復がそれらのそれぞれのDNA塩基対に、隣接するミスマッチに対するわずかに増加した許容範囲を超えて、独立して結合することが見出され、これは、塩基対当たりのT A L E N特異性が、標的部位の長さとは無関係であるという認識を知らしめた。本明細書の別の箇所により詳細に説明するように、短いT A L E Nは長いT A L E Nよりも、標的塩基対当たり大きい特異性を有するが、しかし、長いT A L E Nは短いT A L E Nよりも、20 ~ 32 bpの部位を標的とする試験されたT A L E N長については、全ゲノムの場面における潜在的な切断部位のセットに対してより特異的であることが、実験的に検証された。

【0066】

10

本開示のいくつかの側面は、長いT A L E N中の過剰な結合エネルギーが、対応するオンターゲット切断効率の効率増加なしでオフターゲット配列の切断を可能にすることにより、特異性を減少させるという、驚くべき発見に基づく。本開示のいくつかの側面は、T A L E Nが、オンターゲット切断効率を損なうことなく、オフターゲット結合エネルギー低減することによって、それらの標的配列をより特異的に切断するように操作することができるという、驚くべき発見に基づく。T A L E N特異性が、効率的なオンターゲット切断を可能とするために必要とされるものを越えて、非特異的DNA結合エネルギーを低減することにより、改善可能であるとの認識は、改善された標的部位特異性を有する操作されたT A L E Nを生成するための基礎となる。

【0067】

20

典型的には、T A L E N単量体、例えば本明細書で提供されるT A L E N単量体は、次の構造を含むか、または次の構造のものである：

[N末端ドメイン] - [T A L E 反復アレイ] - [C末端ドメイン] - [ヌクレアーゼドメイン]

式中、各「 - 」は個別に、共有結合または非共有結合的なコンジュゲーションを示し、および式中、コンジュゲーションは、例えば直接結合を介して直接的であっても、または例えばリンカードメインを介して間接的であってもよい。図1も参照されたい。

【0068】

本開示のいくつかの側面は、以前に使用されたT A L E Nに比べて、強化された特異性を有するT A L E Nを提供する。一般に、T A L E Nの配列特異性は、特定のヌクレオチド配列に結合するT A L E 反復アレイによって付与される。T A L E 反復アレイは複数の34アミノ酸T A L E 反復配列からなり、その各々は、位置12と13のアミノ酸である反復可変二残基(RVD)を使用して、1つのDNAヌクレオチドを認識する。本開示のいくつかの側面は、T A L E 反復アレイの特異的結合が、二量体化および核酸切断のために充分であること、および、非特異的核酸結合活性が、T A L E NのN末端および/またはC末端ドメインに起因することを、提供する。

30

【0069】

この認識に基づき、改善されたT A L E Nが本明細書に提供されるように操作される。N末端ドメインを介した非特異的結合が、生理的pHで正に荷電した(カチオン性)アミノ酸残基によって付与される過剰な結合エネルギーを介して、発生し得ることが発見されたので、本明細書で提供される改善されたT A L E Nの一部は、標準T A L E Nと比較して、減少した正味電荷および/またはそれらの標的核酸配列に結合するための減少した結合エネルギーを有する。この電荷の減少は、修飾N末端およびC末端ドメインを介した、オフターゲット結合の減少につながる。したがって標的認識および結合の一部は、T A L E 反復アレイの特異的認識および結合活性により狭く限定される。得られたT A L E Nは、したがって、非修飾ドメインを使用するT A L E Nと比較して、結合の特異性の増加および、次に、改善されたT A L E Nによる標的部位の切断の特異性の増加を示す。

40

【0070】

いくつかの態様において、N末端ドメインの正味電荷が標準N末端ドメイン(配列番号1)の正味電荷より少ないT A L E N；および/またはC末端ドメインの正味電荷が、標

50

標準C末端ドメイン（配列番号22）の正味電荷より少ないT A L E Nが、提供される。いくつかの態様において、N末端ドメインの標的核酸分子への結合エネルギーが、標準N末端ドメイン（配列番号1）の結合エネルギーより少ないT A L E N、および/またはC末端ドメインの標的核酸分子への結合エネルギーが、標準C末端ドメイン（配列番号22）の結合エネルギーより少ないT A L E Nが、提供される。いくつかの態様において、修飾T A L E NのN末端ドメインであって、T A L E N標的核酸分子へのその結合エネルギーが、標準N末端ドメイン（配列番号1）の結合エネルギーより少ないものが、提供される。いくつかの態様において、修飾T A L E NのC末端ドメインであって、T A L E N標的核酸分子へのその結合エネルギーが、標準C末端ドメイン（配列番号22）の結合エネルギーより少ないものが、提供される。いくつかの態様において、提供されるT A L E NにおけるN末端および/またはC末端ドメインの結合エネルギーは、少なくとも5%、少なくとも10%、少なくとも15%、少なくとも20%、少なくとも25%、少なくとも30%、少なくとも35%、少なくとも40%、少なくとも45%、少なくとも50%、少なくとも55%、少なくとも60%、少なくとも65%、少なくとも70%、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98%、または少なくとも99%、低減されている。

【0071】

いくつかの態様において、標準N末端ドメインおよび/または標準C末端ドメインは、生理的pHで正に荷電したアミノ酸残基を、荷電していないかまたは負に荷電したアミノ酸残基で置き換えるように、修飾される。いくつかの態様において、修飾は、正に荷電した残基の、負に荷電した残基による置き換えを含む。いくつかの態様において、修飾は、正に荷電した残基の、中性（非荷電）残基による置き換えを含む。いくつかの態様において、修飾は、正に荷電した残基の、電荷を有さないかまたは負電荷を有する残基による置換を含む。いくつかの態様において、修飾N末端ドメインおよび/または修飾C末端ドメインの正味電荷は、+10以下、+9以下、+8以下、+7以下、+6以下、+5以下、+4以下、+3以下、+2以下、+1以下、0以下、-1以下、-2以下、-3以下、-4以下、-5以下、または-10以下である。いくつかの態様において、修飾N末端ドメインおよび/または修飾C末端ドメインの正味電荷は、+5~-5の間、+2~-7の間、0~-5の間、0~-10の間、-1~-10の間、または-2~-15の間である。いくつかの態様において、修飾N末端ドメインおよび/または修飾C末端ドメインの正味電荷は、負である。いくつかの態様において、修飾N末端ドメインおよび修飾C末端ドメインの正味電荷は、共に負である。いくつかの態様において、修飾N末端ドメインおよび修飾C末端ドメインの正味電荷は、中性またはわずかに正である（例えば、+2未満または+1未満）。いくつかの態様において、修飾N末端ドメインおよび修飾C末端ドメインの正味電荷は、共に、中性またはわずかに正である（例えば、+2未満または+1未満）。

【0072】

いくつかの態様において、修飾N末端ドメインおよび/または修飾C末端ドメインは、標準ドメイン配列の少なくとも1つのカチオン性アミノ酸残基が、生理的pHで電荷を示さないかまたは負電荷を示すアミノ酸残基で置き換えられている点で、それぞれの標準ドメイン配列とは異なっているアミノ酸配列を含む。いくつかの態様において、修飾N末端ドメインおよび/または修飾C末端ドメインにおいて、少なくとも1個の、少なくとも2個の、少なくとも3個の、少なくとも4個の、少なくとも5個の、少なくとも6個の、少なくとも7個の、少なくとも8個の、少なくとも9個の、少なくとも10個の、少なくとも11個の、少なくとも12個の、少なくとも13個の、少なくとも14個の、または少なくとも15個のカチオン性アミノ酸は、生理的pHで電荷を示さないかまたは負電荷を示すアミノ酸残基で置き換えられている。いくつかの態様において、修飾N末端ドメインおよび/または修飾C末端ドメインにおいて、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15個のカチオン性アミノ酸は、生理的pHで電荷を示さないかまたは負電荷を示すアミノ酸残基で置き換えられている。

【0073】

いくつかの態様において、カチオン性アミノ酸残基は、アルギニン（R）、リジン（K）、またはヒスチジン（H）である。いくつかの態様において、カチオン性アミノ酸残基は、RまたはHである。いくつかの態様において、生理的pHで電荷を示さないかまたは負電荷を示すアミノ酸残基は、グルタミン（Q）、グリシン（G）、アスパラギン（N）、スレオニン（T）、セリン（S）、アスパラギン酸（D）またはグルタミン酸（E）である。いくつかの態様において、生理的pHで電荷を示さないかまたは負電荷を示すアミノ酸残基は、Qである。いくつかの態様において、修飾N末端ドメインおよび/または修飾C末端ドメインにおいて、少なくとも1つのリジンまたはアルギニン残基は、グルタミン残基で置き換えられている。

10

【0074】

いくつかの態様において、C末端ドメインは、次のアミノ酸置換の1または2以上を含む：K777Q、K778Q、K788Q、R789Q、R792Q、R793Q、R801Q。いくつかの態様において、C末端ドメインは、次のアミノ酸置換の2または3以上を含む：K777Q、K778Q、K788Q、R789Q、R792Q、R793Q、R801Q。いくつかの態様において、C末端ドメインは、次のアミノ酸置換の3または4以上を含む：K777Q、K778Q、K788Q、R789Q、R792Q、R793Q、R801Q。いくつかの態様において、C末端ドメインは、次のアミノ酸置換の4または5以上を含む：K777Q、K778Q、K788Q、R789Q、R792Q、R793Q、R801Q。いくつかの態様において、C末端ドメインは、次のアミノ酸置換の5または6以上を含む：K777Q、K778Q、K788Q、R789Q、R792Q、R793Q、R801Q。いくつかの態様において、C末端ドメインは、次のアミノ酸置換の6または7以上を含む：K777Q、K778Q、K788Q、R789Q、R792Q、R793Q、R801Q。いくつかの態様において、C末端ドメインは、次のアミノ酸置換の全7つを含む：K777Q、K778Q、K788Q、R789Q、R792Q、R793Q、R801Q。いくつかの態様において、C末端ドメインは、Q3バリエーション配列（K788Q、R792Q、R801Q、配列番号23参照）を含む。いくつかの態様において、C末端ドメインは、Q7バリエーション配列（K777Q、K778Q、K788Q、R789Q、R792Q、R793Q、R801Q、配列番号24参照）を含む。

20

30

【0075】

いくつかの態様において、N末端ドメインは、標準N末端ドメインのトランケート型である。いくつかの態様において、C末端ドメインは、標準C末端ドメインのトランケート型である。いくつかの態様において、トランケートされたN末端ドメインおよび/またはトランケートされたC末端ドメインは、標準ドメインの残基の90%未満、80%未満、70%未満、60%未満、50%未満、40%未満、30%未満、または25%未満を含む。いくつかの態様において、トランケートされたC末端ドメインは、60個未満、50個未満、40個未満、30個未満、29個未満、28個未満、27個未満、26個未満、25個未満、24個未満、23個未満、22個未満、21個未満、または20個のアミノ酸残基を含む。いくつかの態様において、トランケートされたC末端ドメインは、60、59、58、57、56、55、54、53、52、51、50、49、48、47、46、45、44、43、42、41、40、39、38、37、36、35、34、33、32、31、30、29、28、27、26、25、24、23、22、21、20、19、18、17、16、15、14、13、12、11、または10個の残基を含む。いくつかの態様において、修飾N末端ドメインおよび/または修飾C末端ドメインはトランケートされており、1または2以上のアミノ酸置換を含む。当業者には、いくつかの態様においては、トランケートを収容するために、例えばトランケートされたC末端ドメインなどのトランケートされたドメインを使用してTALLENにおけるDNAスペーサー長を調整することが望ましいことが明らかであろう。

40

50

【 0 0 7 6 】

いくつかの態様において、核酸切断ドメインと呼ばれることもあるヌクレアーゼドメインは、非特異的な切断ドメイン、例えば、F o k Iヌクレアーゼドメインである。いくつかの態様において、ヌクレアーゼドメインは単量体であり、核酸を切断するために、二量体化または多量体化しなければならない。T A L E N単量体の、ホモまたはヘテロ二量体化または多量体化は、典型的には、単量体の、二量体化を可能にするのに充分近接している結合配列への、例えば、同一の核酸分子（例えば同一の二本鎖核酸分子）上で互いに近接している配列への結合を介して、生じる。

【 0 0 7 7 】

最も一般的に使用されるドメイン、例えば、最も広く使用されるN末端およびC末端ドメインは、本明細書において標準ドメインと呼ぶ。標準N末端ドメインの例示的な配列（配列番号1）および標準C末端ドメインの例示的な配列（配列番号22）が、本明細書中に提供される。F o k Iヌクレアーゼドメインの典型的な配列も、本明細書中に提供される。さらに、C C R 5結合T A L E反復アレイを形成する、T A L E反復の例示的な配列が提供される。以下に提供される配列は例示的であり、本開示に包含されるいくつかの態様を説明する目的のために提供されることが、理解されるであろう。これらは限定的であることを意味せず、本開示の側面に従って有用である追加の配列は、この開示に基づいて当業者には明らかであろう。

【 0 0 7 8 】

標準N末端ドメイン：

【 数 2 】

VDLRTLGYSSQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMI
AALPEATHEAIVGVGKQWSGARRALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVH
AWRNALTGAPLN（配列番号1）

【 0 0 7 9 】

修飾N末端ドメイン：N 1

【 数 3 】

VDLRTLGYSSQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMI
AALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPPLQLDTGQLLQIAKRGGVTAVEAVH
AWRNALTGAPLN（配列番号2）

【 0 0 8 0 】

修飾N末端ドメイン：N 2

【 数 4 】

VDLRTLGYSSQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMI
AALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPPLQLDTGQLLQIAQRGVTAVEAVH
AWRNALTGAPLN（配列番号3）

【 0 0 8 1 】

修飾N末端ドメイン：N 3

【 数 5 】

VDLRTLGYSSQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMI
AALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPPLQLDTGQLLQIAQQGGVTAVEAVH
AWRNALTGAPLN（配列番号4）

【 0 0 8 2 】

T A L E反復アレイ：L 1 8 C C R 5 A

10

20

30

40

50

【数 6】

MTPDQVVAIASNGGGKQALETVQRLLPVLCQDH (配列番号 5)
 GLTPEQVVAIASHDGGKQALETVQRLLPVLCQAH (配列番号 6)
 GLTPDQVVAIASNIGGKQALETVQRLLPVLCQAH (配列番号 7)
 GLTPAQVVAIASNGGGKQALETVQRLLPVLCQDH (配列番号 8)
 GLTPDQVVAIASNGGGKQALETVQRLLPVLCQDH (配列番号 9)
 GLTPEQVVAIASNIGGKQALETVQRLLPVLCQAH (配列番号 10)
 GLTPDQVVAIASHDGGKQALETVQRLLPVLCQAH (配列番号 11)
 GLTPAQVVAIASNIGGKQALETVQRLLPVLCQDH (配列番号 12)
 GLTPDQVVAIASHDGGKQALETVQRLLPVLCQDH (配列番号 13)
 GLTPEQVVAIASHDGGKQALETVQRLLPVLCQAH (配列番号 14)
 GLTPDQVVAIASNGGGKQALETVQRLLPVLCQAH (配列番号 15)
 GLTPAQVVAIASNNGGKQALETVQRLLPVLCQDH (配列番号 16)
 GLTPDQVVAIASHDGGKQALETVQRLLPVLCQDH (配列番号 17)
 GLTPEQVVAIASNIGGKQALETVQRLLPVLCQAH (配列番号 18)
 GLTPDQVVAIASNNGGKQALETVQRLLPVLCQAH (配列番号 19)
 GLTPAQVVAIASHDGGKQALETVQRLLPVLCQDH (配列番号 20)
 GLTPEQVVAIASNGGGRPALE (配列番号 21)

10

20

【0083】

標準 C 末端ドメイン：

【数 7】

SIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRV
 A (配列番号 22)

30

【0084】

修飾 C 末端ドメイン：Q 3

【数 8】

SIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIQRTNQORIPERTSHQV
 A (配列番号 23)

40

【0085】

修飾 C 末端ドメイン：Q 7

【数 9】

SIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVQQGLPHAPALIQQTNQQIPERTSHQV
 A (配列番号 24)

【0086】

修飾 C 末端ドメイン：28 - a a

【数 10】

SIVAQLSRPDPALAALTNDHLVALACLG (配列番号 25)

50

【 0 0 8 7 】

F o k I : ホモ二量体

【 数 1 1 】

GSQLVKSELEEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHL
 GGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRNKHINPNEWW
 KVYPSSVTEFKFLFVSGHFHFGNYKAQLTRLNHIITNCNGAVLSVEELLIGGEMIKAGTLTLEE
 VRRKFNNGEINF* (配列番号 26)

【 0 0 8 8 】

F o k I : E L

【 数 1 2 】

GSQLVKSELEEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHL
 GGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMERYVEENQTRNKHLNPNEWW
 KVYPSSVTEFKFLFVSGHFHFGNYKAQLTRLNHIITNCNGAVLSVEELLIGGEMIKAGTLTLEE
 VRRKFNNGEINF* (配列番号 27)

10

【 0 0 8 9 】

F o k I : K K

【 数 1 3 】

GSQLVKSELEEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHL
 GGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVKENQTRNKHINPNEWW
 KVYPSSVTEFKFLFVSGHFHFGNYKAQLTRLNHKTNCNGAVLSVEELLIGGEMIKAGTLTLEE
 VRRKFNNGEINF* (配列番号 28)

20

【 0 0 9 0 】

F o k I : E L D

【 数 1 4 】

GSQLVKSELEEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHL
 GGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMERYVEENQTRDKHLNPNEWW
 KVYPSSVTEFKFLFVSGHFHFGNYKAQLTRLNHIITNCNGAVLSVEELLIGGEMIKAGTLTLEE
 VRRKFNNGEINF* (配列番号 29)

30

【 0 0 9 1 】

F o k I : K K R

【 数 1 5 】

GSQLVKSELEEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHL
 GGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVKENQTRNKHINPNEWW
 KVYPSSVTEFKFLFVSGHFHFGNYKAQLTRLNRKTNCNGAVLSVEELLIGGEMIKAGTLTLEE
 VRRKFNNGEINF* (配列番号 30)

40

【 0 0 9 2 】

いくつかの態様において、標準N末端ドメイン、T A L E 反復アレイ、修飾C末端ドメイン、およびヌクレアーゼドメインを含むT A L E Nが、本明細書に提供される。いくつかの態様において、修飾N末端ドメイン、T A L E 反復アレイ、標準C末端ドメイン、およびヌクレアーゼドメインを含むT A L E Nが、本明細書に提供される。いくつかの態様

50

において、修飾N末端ドメイン、T A L E 反復アレイ、修飾C末端ドメイン、およびヌクレアーゼドメインを含むT A L E Nが、本明細書に提供される。いくつかの態様において、ヌクレアーゼドメインは、F o k Iヌクレアーゼドメインである。いくつかの態様において、F o k Iヌクレアーゼドメインは、ホモ二量体F o k Iドメイン、またはF o k I - E L、F o k I - K K、F o k I - E L D、もしくはF o k I - K K Rドメインである。

【 0 0 9 3 】

本明細書に提供される標準および修飾ドメインの特定の配列の全ての可能な組み合わせは、本開示によって包含され、以下を含む。

【 表 1 - 1 】

TALEN	N 末端ドメイン	TALE 反復アレイ	C 末端ドメイン	ヌクレアーゼドメイン
1	標準	配列特異的	Q3	ホモ二量体
2	標準	配列特異的	Q3	EL
3	標準	配列特異的	Q3	KK
4	標準	配列特異的	Q3	ELD
5	標準	配列特異的	Q3	KKR
6	標準	配列特異的	Q7	ホモ二量体
7	標準	配列特異的	Q7	EL
8	標準	配列特異的	Q7	KK
9	標準	配列特異的	Q7	ELD
10	標準	配列特異的	Q7	KKR
11	標準	配列特異的	トランケート(28aa)	ホモ二量体
12	標準	配列特異的	トランケート(28aa)	EL
13	標準	配列特異的	トランケート(28aa)	KK
14	標準	配列特異的	トランケート(28aa)	ELD
15	標準	配列特異的	トランケート(28aa)	KKR
16	N1	配列特異的	標準	ホモ二量体
17	N1	配列特異的	標準	EL
18	N1	配列特異的	標準	KK
19	N1	配列特異的	標準	ELD
20	N1	配列特異的	標準	KKR
21	N1	配列特異的	Q3	ホモ二量体
22	N1	配列特異的	Q3	EL
23	N1	配列特異的	Q3	KK
24	N1	配列特異的	Q3	ELD
25	N1	配列特異的	Q3	KKR
26	N1	配列特異的	Q7	ホモ二量体
27	N1	配列特異的	Q7	EL

10

20

30

40

【表 1 - 2】

TALEN	N 末端ドメイン	TALE 反復アレイ	C 末端ドメイン	ヌクレアーゼドメイン
28	N1	配列特異的	Q7	KK
29	N1	配列特異的	Q7	ELD
30	N1	配列特異的	Q7	KKR
31	N1	配列特異的	トランケート(28aa)	ホモ二量体
32	N1	配列特異的	トランケート(28aa)	EL
33	N1	配列特異的	トランケート(28aa)	KK
34	N1	配列特異的	トランケート(28aa)	ELD
35	N1	配列特異的	トランケート(28aa)	KKR
36	N2	配列特異的	標準	ホモ二量体
37	N2	配列特異的	標準	EL
38	N2	配列特異的	標準	KK
39	N2	配列特異的	標準	ELD
40	N2	配列特異的	標準	KKR
41	N2	配列特異的	Q3	ホモ二量体
42	N2	配列特異的	Q3	EL
43	N2	配列特異的	Q3	KK
44	N2	配列特異的	Q3	ELD
45	N2	配列特異的	Q3	KKR
46	N2	配列特異的	Q7	ホモ二量体
47	N2	配列特異的	Q7	EL
48	N2	配列特異的	Q7	KK
49	N2	配列特異的	Q7	ELD
50	N2	配列特異的	Q7	KKR
51	N2	配列特異的	トランケート(28aa)	ホモ二量体
52	N2	配列特異的	トランケート(28aa)	EL
53	N2	配列特異的	トランケート(28aa)	KK
54	N2	配列特異的	トランケート(28aa)	ELD
55	N2	配列特異的	トランケート(28aa)	KKR
56	N3	配列特異的	標準	ホモ二量体
57	N3	配列特異的	標準	EL

10

20

30

40

【表 1 - 3】

TALEN	N 末端ドメイン	TALE 反復アレイ	C 末端ドメイン	ヌクレアーゼドメイン
58	N3	配列特異的	標準	KK
59	N3	配列特異的	標準	ELD
60	N3	配列特異的	標準	KKR
61	N3	配列特異的	Q3	ホモ二量体
62	N3	配列特異的	Q3	EL
63	N3	配列特異的	Q3	KK
64	N3	配列特異的	Q3	ELD
65	N3	配列特異的	Q3	KKR
66	N3	配列特異的	Q7	ホモ二量体
67	N3	配列特異的	Q7	EL
68	N3	配列特異的	Q7	KK
69	N3	配列特異的	Q7	ELD
70	N3	配列特異的	Q7	KKR
71	N3	配列特異的	トランケート(28aa)	ホモ二量体
72	N3	配列特異的	トランケート(28aa)	EL
73	N3	配列特異的	トランケート(28aa)	KK
74	N3	配列特異的	トランケート(28aa)	ELD
75	N3	配列特異的	トランケート(28aa)	KKR
76	標準	配列特異的	標準	EL
77	標準	配列特異的	標準	KK
78	標準	配列特異的	標準	ELD
79	標準	配列特異的	標準	KKR
80	標準	配列特異的	トランケート(28aa)	ホモ二量体
81	標準	配列特異的	トランケート(28aa)	EL
82	標準	配列特異的	トランケート(28aa)	KK
83	標準	配列特異的	トランケート(28aa)	ELD
84	標準	配列特異的	トランケート(28aa)	KKR

表 1：本開示に包含される例示の TALEN。使用されるそれぞれの TALE 反復アレイは、具体的な標的配列に依存する。当業者は、かかる配列特異的な TALE 反復アレイを、本開示および当該技術分野の知識に基づいて設計することができるであろう。異なる N 末端、C 末端、およびヌクレアーゼドメインに対する配列は、上に提供されている（配列番号 1～4 および 22～30 参照）。

【0094】

当業者は、本明細書で提供される例示的な配列は、単に説明目的のためであり、本開示の範囲の限定を意図しないことを理解するであろう。本開示はまた、本発明の TALEN ドメイン、例えば本明細書に記載の修飾された N 末端ドメイン、C 末端ドメイン、およびヌクレアーゼドメインの、他の TALEN 配列の場面における、例えば他の修飾または非修飾 TALEN 構造における使用も包含する。本開示の側面に従って有用な、記載の原理およびパラメータを満足する追加の配列は、当業者には明らかであろう。

【0095】

いくつかの態様において、提供されるT A L E Nは、単量体である。いくつかの態様において、T A L E N単量体は、別のT A L E N単量体と二量体化してT A L E N二量体を形成することができる。いくつかの態様において形成された二量体は、ホモ二量体である。いくつかの態様において、二量体はヘテロ二量体である。

【0096】

いくつかの態様において、本明細書で提供されるT A L E Nは、その標的部位を高い特異性で切断する。例えば、いくつかの態様において、ゲノム内の所望の標的部位を切断し、かつ一方で、ヌクレアーゼがその意図される標的部位を切断するのに有効な濃度において、1未満、2未満、3未満、4未満、5未満、6未満、7未満、8未満、9未満、または10未満のオフターゲット部位に結合および/または切断するように操作された、改善されたT A L E Nが提供される。いくつかの態様において、ゲノム内の任意の他の部位とは、少なくとも3個、少なくとも4個、少なくとも5個、少なくとも6個、少なくとも7個、少なくとも8個、少なくとも9個、または少なくとも10個のヌクレオチド残基で異なるように選択された所望のユニークな標的部位を切断するように操作されたT A L E Nが、提供される。

【0097】

本開示のいくつかの側面は、本明細書に提供されるT A L E Nをコードする核酸を提供する。例えば、表1に記載のT A L E Nをコードする核酸が、本明細書で提供される。いくつかの態様において、T A L E Nをコードする核酸は、異種プロモーターの制御下にある。いくつかの態様において、コードする核酸は、発現コンストラクト、例えば、プラスミド、ウイルスベクターまたは直鎖発現コンストラクトに含まれている。いくつかの態様において、核酸または発現コンストラクトは、細胞、組織、または生物中に存在する。

【0098】

本明細書に提供されるT A L E Nをコードする例示的な核酸のマップを、図19に示す。かかる核酸の例示の配列を下に提供する。当業者は、本明細書に提供されるマップおよび配列は例示であり、本開示の範囲を限定するものではないことを理解するであろう。

【0099】

本明細書の別の箇所では説明したように、本開示により提供される改善されたT A L E Nを含むT A L E Nは、実質的に任意の核酸配列に結合する（および切断する）ように、使用される配列特異的T A L E 反復アレイに基づいて、操作することができる。いくつかの態様において、本明細書に提供される改善されたT A L E Nは、疾患または障害と関連することが知られている遺伝子内の標的配列に結合する。いくつかの態様において、本明細書に提供されるT A L E Nは、治療目的のために使用することができる。たとえば、いくつかの態様において、本明細書に提供されるT A L E Nは、以下の1または2以上を含むがこれらに限定されない様々な疾患、障害、および/または状態のいずれかの処置のために使用され得る：自己免疫疾患（例えば、糖尿病、狼瘡、多発性硬化症、乾癬、関節リウマチ）；炎症性疾患（例えば、関節炎、骨盤内炎症性疾患）；感染症（例えば、ウイルス感染（例えば、H I V、H C V、R S V）、細菌感染、真菌感染、敗血症）；神経学的障害（例えば、アルツハイマー病、ハンチントン病；自閉症；デュシェンヌ型筋ジストロフィー）；心臓血管疾患（例えば、アテローム性動脈硬化症、高コレステロール血症、血栓症、凝固障害、黄斑変性症などの血管新生障害）；増殖性疾患（例えばがん、良性腫瘍）；呼吸器疾患（例えば、慢性閉塞性肺疾患）；消化器疾患（例えば、炎症性腸疾患、潰瘍）；筋骨格障害（例えば、線維筋痛、関節炎）；内分泌、代謝、および栄養障害（例えば、糖尿病、骨粗しょう症）；泌尿器疾患（例えば、腎疾患）；精神障害（例えば、うつ病、統合失調症）；皮膚疾患（例えば創傷、湿疹）；血液およびリンパ管障害（例えば、貧血、血友病）など。いくつかの態様において、T A L E Nは、二量体化の際に標的配列を切断する。いくつかの態様において、本明細書に提供されるT A L E Nは、疾患または障害に関連するアレル内の標的部位を切断する。いくつかの態様において、T A L E Nは標的部位を切断し、これが、疾患または障害の処置または予防をもたらす。いくつかの態様

において、疾患は、H I V / A I D Sである。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、T A L E Nは、C C R 5 標的配列（例えば、H I Vに関連するC C R 5 配列）に結合する。いくつかの態様において、T A L E Nは、A T M 標的配列（例えば、毛細血管拡張性運動失調症に関連するA T M 標的配列）に結合する。いくつかの態様において、T A L E Nは、V E G F A 標的配列（例えば、増殖性疾患に関連するV E G F A 配列）に結合する。いくつかの態様において、T A L E Nは、C F T R 標的配列（例えば、嚢胞性線維症に関連するC F T R 配列）に結合する。いくつかの態様において、T A L E Nは、ジストロフィン標的配列（例えば、デュシェンヌ型筋ジストロフィーに関連するジストロフィン遺伝子配列）に結合する。いくつかの態様において、T A L E Nは、ヘモクロマトーシス、血友病、シャルコー・マリー・トゥース病、神経線維腫症、フェニルケトン尿症、多発性嚢胞腎疾患、鎌状赤血球病、またはテイ・サックス病に関連する標的配列に結合する。好適な標的遺伝子、例えば、リストされている疾患の原因となる遺伝子は、当業者に知られている。疾患または障害に関連する追加の遺伝子および遺伝子配列は、当業者には明らかであろう。

【 0 1 0 0 】

本開示のいくつかの側面では、以前に使用されたT A L E エフェクタードメインと比較して、非特異的核酸結合活性が減少したN末端およびC末端T A L E エフェクタードメインなどの、単離されたT A L E エフェクタードメインを提供する。本明細書で提供される単離されたT A L E エフェクタードメインは、好適なT A L E エフェクター分子の場面において、例えばT A L E ヌクレアーゼ、T A L E 転写活性化因子、T A L E 転写リプレッサー、T A L E リコンビナーゼ、およびT A L E エピゲノム修飾酵素などで、使用することができる。その場面において単離されたT A L E ドメインが使用できる追加の好適なT A L E エフェクターは、本開示に基づいて当業者には明らかであろう。一般に、本明細書で提供される単離されたN末端およびC末端ドメインは、生理的p Hで正に荷電した（カチオン性）アミノ酸残基によって付与される過剰な結合エネルギーを最適化する、例えば最小化するように、操作される。本明細書で提供される改善されたN末端またはC末端T A L E ドメインのいくつかは、それぞれの標準T A L E ドメインと比較して、標的核酸配列に結合するための減少した正味電荷および/または減少した結合エネルギーを有している。T A L E エフェクター分子の一部として使用する場合、例えば、T A L E ヌクレアーゼ、T A L E 転写活性化因子、T A L E 転写リプレッサー、T A L E リコンビナーゼ、またはT A L E エピゲノム修飾酵素の一部として使用する場合は、荷電におけるこの減少は、修飾N末端およびC末端ドメイン（単数または複数）を介した、オフターゲット結合の減少につながる。したがって、標的認識および結合の部分は、本明細書の他の箇所により詳細に説明されるように、より狭く、T A L E 反復アレイの特異的認識および結合活性に限定される。得られたT A L E エフェクター分子は、したがって、非修飾ドメインを使用するT A L E エフェクター分子と比較して、結合の特異性の増加、ひいてはT A L E エフェクターのそれぞれの効果の特異性の増加を示す（例えば、T A L E ヌクレアーゼによる標的部位の切断、T A L E 転写活性化因子による標的遺伝子の活性化、T A L E 転写リプレッサーによる標的遺伝子の発現の抑制、T A L E リコンビナーゼによる標的配列の組換え、またはT A L E エピゲノム修飾酵素による標的配列のエピジェネティックな修飾）。

【 0 1 0 1 】

いくつかの態様において、正味電荷が標準N末端ドメイン（配列番号1）の正味電荷より小さい、単離されたN末端T A L E ドメインが提供される。いくつかの態様において、正味電荷が標準C末端ドメイン（配列番号22）の正味電荷より小さい、単離されたC末端T A L E ドメインが提供される。いくつかの態様において、標的核酸分子への結合エネルギーが、標準N末端ドメイン（配列番号1）の結合エネルギーより小さい、単離されたN末端T A L E ドメインが提供される。いくつかの態様において、標的核酸分子への結合エネルギーが、標準C末端ドメイン（配列番号22）の結合エネルギーより小さい、単離されたC末端T A L E ドメインが提供される。いくつかの態様において、本明細書で提供される単離されたN末端および/または単離されたC末端T A L E ドメインの結合エネル

10

20

30

40

50

ギーは、少なくとも5%、少なくとも10%、少なくとも15%、少なくとも20%、少なくとも25%、少なくとも30%、少なくとも35%、少なくとも40%、少なくとも45%、少なくとも50%、少なくとも55%、少なくとも60%、少なくとも65%、少なくとも70%、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98%、または少なくとも99%、低減されている。

【0102】

いくつかの態様において、標準N末端ドメインおよび/または標準C末端ドメインは、生理的pHで正に荷電したアミノ酸残基を、荷電していないかまたは負に荷電したアミノ酸残基で置き換えるよう修飾されて、本明細書で提供される単離されたN末端および/またはC末端ドメインが達成される。いくつかの態様において、修飾は、正に荷電した残基の、負に荷電した残基による置換を含む。いくつかの態様において、修飾は、正に荷電した残基の、中性（非荷電）残基による置換を含む。いくつかの態様において、修飾は、正に荷電した残基の、電荷を有さないかまたは負電荷を有する残基による置換を含む。いくつかの態様において、本明細書で提供される単離N末端ドメインおよび/または単離C末端ドメインの正味電荷は、生理的pHにおいて+10以下、+9以下、+8以下、+7以下、+6以下、+5以下、+4以下、+3以下、+2以下、+1以下、0以下、-1以下、-2以下、-3以下、-4以下、-5以下、または-10以下である。いくつかの態様において、単離N末端TALドメインおよび/または単離C末端TALドメインの正味電荷は、生理的pHにおいて+5~-5の間、+2~-7の間、0~-5の間、0~-10の間、-1~-10の間、または-2~-15の間である。いくつかの態様において、単離N末端ドメインおよび/または単離C末端ドメインの正味電荷は、負である。いくつかの態様において、単離N末端TALドメインおよび/または単離C末端TALドメインが提供され、単離N末端TALドメインおよび単離C末端TALドメインの正味電荷は、共に負である。いくつかの態様において、単離N末端TALドメインおよび単離C末端TALドメインの正味電荷は、中性またはわずかに正である（例えば、生理的pHで+2未満または+1未満）。いくつかの態様において、単離N末端TALドメインおよび/または単離C末端TALドメインが提供され、単離N末端TALドメインおよび単離C末端TALドメインの正味電荷は、共に、中性またはわずかに正である（例えば、生理的pHで+2未満または+1未満）。

【0103】

いくつかの態様において、本明細書に提供される単離N末端ドメインおよび/または単離C末端ドメインは、標準ドメイン配列の少なくとも1つのカチオン性アミノ酸残基が、生理的pHで電荷を示さないかまたは負電荷を示すアミノ酸残基で置き換えられている点で、それぞれの標準ドメイン配列とは異なっているアミノ酸配列を含む。いくつかの態様において、提供される単離N末端ドメインおよび/または単離C末端ドメインの少なくとも1個の、少なくとも2個の、少なくとも3個の、少なくとも4個の、少なくとも5個の、少なくとも6個の、少なくとも7個の、少なくとも8個の、少なくとも9個の、少なくとも10個の、少なくとも11個の、少なくとも12個の、少なくとも13個の、少なくとも14個の、または少なくとも15個のカチオン性アミノ酸は、生理的pHで電荷を示さないかまたは負電荷を示すアミノ酸残基で置き換えられている。いくつかの態様において、単離N末端ドメインおよび/または単離C末端ドメインの1、2、3、4、5、6、7、8、9、10、11、12、13、14、15個のカチオン性アミノ酸は、生理的pHで電荷を示さないかまたは負電荷を示すアミノ酸残基で置き換えられている。

【0104】

いくつかの態様において、カチオン性アミノ酸残基は、アルギニン（R）、リジン（K）、またはヒスチジン（H）である。いくつかの態様において、カチオン性アミノ酸残基は、RまたはHである。いくつかの態様において、生理的pHで電荷を示さないかまたは負電荷を示すアミノ酸残基は、グルタミン（Q）、グリシン（G）、アスパラギン（N）、スレオニン（T）、セリン（S）、アスパラギン酸（D）またはグルタミン酸（E）で

ある。いくつかの態様において、生理的 pH で電荷を示さないかまたは負電荷を示すアミノ酸残基は、Q である。いくつかの態様において、単離 N 末端ドメインおよび/または修飾 C 末端ドメインの少なくとも 1 つのリジンまたはアルギニン残基は、グルタミン残基で置き換えられている。

【0105】

いくつかの態様において、次のアミノ酸置換の 1 または 2 以上を含む単離 C 末端 T A L E ドメインが、本明細書で提供される：K 7 7 7 Q、K 7 7 8 Q、K 7 8 8 Q、R 7 8 9 Q、R 7 9 2 Q、R 7 9 3 Q、R 8 0 1 Q。いくつかの態様において、単離 C 末端ドメインは、次のアミノ酸置換の 2 または 3 以上を含む：K 7 7 7 Q、K 7 7 8 Q、K 7 8 8 Q、R 7 8 9 Q、R 7 9 2 Q、R 7 9 3 Q、R 8 0 1 Q。いくつかの態様において、単離 C 末端ドメインは、次のアミノ酸置換の 3 または 4 以上を含む：K 7 7 7 Q、K 7 7 8 Q、K 7 8 8 Q、R 7 8 9 Q、R 7 9 2 Q、R 7 9 3 Q、R 8 0 1 Q。いくつかの態様において、単離 C 末端ドメインは、次のアミノ酸置換の 4 または 5 以上を含む：K 7 7 7 Q、K 7 7 8 Q、K 7 8 8 Q、R 7 8 9 Q、R 7 9 2 Q、R 7 9 3 Q、R 8 0 1 Q。いくつかの態様において、単離 C 末端ドメインは、次のアミノ酸置換の 5 または 6 以上を含む：K 7 7 7 Q、K 7 7 8 Q、K 7 8 8 Q、R 7 8 9 Q、R 7 9 2 Q、R 7 9 3 Q、R 8 0 1 Q。いくつかの態様において、単離 C 末端ドメインは、次のアミノ酸置換の 6 または 7 以上を含む：K 7 7 7 Q、K 7 7 8 Q、K 7 8 8 Q、R 7 8 9 Q、R 7 9 2 Q、R 7 9 3 Q、R 8 0 1 Q。いくつかの態様において、単離 C 末端ドメインは、次のアミノ酸置換の全 7 つを含む：K 7 7 7 Q、K 7 7 8 Q、K 7 8 8 Q、R 7 8 9 Q、R 7 9 2 Q、R 7 9 3 Q、R 8 0 1 Q。いくつかの態様において、単離 C 末端ドメインは、Q 3 バリエーション配列 (K 7 8 8 Q、R 7 9 2 Q、R 8 0 1 Q、配列番号 2 3 参照) を含む。いくつかの態様において、単離 C 末端ドメインは、Q 7 バリエーション配列 (K 7 7 7 Q、K 7 7 8 Q、K 7 8 8 Q、R 7 8 9 Q、R 7 9 2 Q、R 7 9 3 Q、R 8 0 1 Q、配列番号 2 4 参照) を含む。

【0106】

いくつかの態様において、標準 N 末端ドメインのランケット型である、単離 N 末端 T A L E ドメインが提供される。いくつかの態様において、標準 C 末端ドメインのランケット型である、単離 C 末端 T A L E ドメインが提供される。いくつかの態様において、ランケットされた N 末端ドメインおよび/またはランケットされた C 末端ドメインは、標準ドメインの残基の 90% 未満、80% 未満、70% 未満、60% 未満、50% 未満、40% 未満、30% 未満、または 25% 未満を含む。いくつかの態様において、ランケットされた C 末端ドメインは、60 個未満、50 個未満、40 個未満、30 個未満、29 個未満、28 個未満、27 個未満、26 個未満、25 個未満、24 個未満、23 個未満、22 個未満、21 個未満、または 20 個のアミノ酸残基を含む。いくつかの態様において、ランケットされた C 末端ドメインは、60、59、58、57、56、55、54、53、52、51、50、49、48、47、46、45、44、43、42、41、40、39、38、37、36、35、34、33、32、31、30、39、38、37、36、35、34、33、32、31、30、29、28、27、26、25、24、23、22、21、20、19、18、17、16、15、14、13、12、11、または 10 個の残基を含む。いくつかの態様において、ランケットされ、1 または 2 以上のアミノ酸置換を含む、単離 N 末端 T A L E ドメインおよび/または単離 C 末端 T A L E ドメインが提供される。いくつかの態様において、単離 N 末端 T A L E ドメインは、配列番号 2 ~ 5 のいずれかに提供されるアミノ酸配列を含む。いくつかの態様において、単離 C 末端 T A L E ドメインは、配列番号 2 3 ~ 2 5 のいずれかに提供されるアミノ酸配列を含む。

【0107】

本明細書で提供される単離された C および N 末端 T A L E エドメインは、任意の T A L E エフェクター分子の場面で、例えば T A L E ヌクレアーゼ、T A L E 転写活性化因子、T A L E 転写リプレッサー、T A L E リコンビナーゼ、T A L E エピゲノム修飾酵素、および任意のその他の好適な T A L E エフェクター分子の一部として使用できること

が、当業者には明らかであろう。いくつかの態様において、本明細書で提供される T A L E ドメインは、次の構造を含むかまたは本質的にこれからなる T A L E 分子の場面において、使用される。

[N 末端ドメイン] - [T A L E 反復アレイ] - [C 末端ドメイン] - [エフェクタードメイン]

または

[エフェクタードメイン] - [N 末端ドメイン] - [T A L E 反復アレイ] - [C 末端ドメイン]

この式中、エフェクタードメインは、いくつかの態様において、ヌクレアーゼドメイン、転写活性化因子もしくはリプレッサードメイン、リコンビナーゼドメイン、またはエピゲノム修飾酵素ドメインであってよい。

【 0 1 0 8 】

当業者には明らかなように、いくつかの態様においては、ランケットを収容するために、例えば本明細書で提供されるようなランケットされた C 末端ドメインなどのランケットされたドメインを使用する場合に、スパーサーを含む T A L E エフェクタードメインにおけるかかる D N A スパーサー長を調整することが好ましい。

【 0 1 0 9 】

本開示のいくつかの側面は、例えば、T A L E N 単量体などの T A L E N を含む組成物を提供する。いくつかの態様において、組成物は、T A L E N 単量体と、該 T A L E N とヘテロ二量体を形成することができる別の T A L E N 単量体とを含む。

【 0 1 1 0 】

いくつかの態様において、T A L E N は、例えばヒト対象などの対象への投与用に処方された組成物中に提供される。例えば、いくつかの態様において、T A L E N および薬学的に許容し得る賦形剤を含む、医薬組成物が提供される。いくつかの態様において、医薬組成物は、対象への投与のために処方される。いくつかの態様において、医薬組成物は、対象における細胞内の標的配列を切断するための T A L E N の有効量を含む。いくつかの態様において、T A L E N は、疾患または障害に関連することが知られている遺伝子内の標的配列に結合し、ここで組成物は、疾患または障害に関連する症状を軽減するための T A L E N の有効量を含む。

【 0 1 1 1 】

例えば、いくつかの態様は、本明細書で提供される T A L E N、またはかかるヌクレアーゼをコードする核酸、および薬学的に許容し得る賦形剤を含む、医薬組成物を提供する。医薬組成物は、任意に 1 または 2 以上の追加の治療的に活性な物質を含んでもよい。

【 0 1 1 2 】

本明細書に記載の医薬組成物の製剤は、薬理学の分野で既知のまたは今後開発される任意の方法によって、製造することができる。一般にかかる製造方法は、活性成分を、賦形剤および / または 1 もしくは 2 以上の他の補助成分と会合させるステップ、次いで必要に応じて、および / または望ましい場合には、生成物を所望の単一または複数用量単位に成形および / または包装すること、を含む。

【 0 1 1 3 】

医薬製剤は、薬学的に許容し得る賦形剤をさらに含んでもよく、これは本明細書中で使用する場合、所望の特定の剤形に適するように、任意の全ての溶媒、分散媒質、希釈剤、または他の液体ビヒクル、分散もしくは懸濁助剤、表面活性剤、等張剤、増粘剤または乳化剤、防腐剤、固体結合剤、潤滑剤等を含む。RemingtonのThe Science and Practice of Pharmacy, 21st Edition, A. R. Gennaro (Lippincott, Williams & Wilkins, Baltimore, MD, 2006; 参照により本明細書に組み込まれる) には、医薬組成物の処方に使用される種々の賦形剤および、その製造のための知られている技術が開示されている。任意の従来の賦形剤媒体が、例えば任意の望ましくない生物学的効果を生じるか、あるいは医薬組成物の任意の他の成分 (単数または複数) と有害な様式で相互作用することによって物質またはその誘導体と不適合である限りを除き、その使用は、本発明の範囲内にあること

10

20

30

40

50

【 0 1 1 4 】

10

【 0 1 1 5 】

50

chnology 29 (2): 143-8. doi:10.1038/nbt.1755. PMID 21179091 ; Hockemeyer, D.; Wang, H.; Kiani, S.; Lai, C. S.; Gao, Q.; Cassady, J. P.; Cost, G. J.; Zhang, L. et al. (2011). "Genetic engineering of human pluripotent cells using TALE nucleases". Nature Biotechnology 29 (8). doi:10.1038/nbt.1927 ; Wood, A. J.; Lo, T. -W.; Zeitler, B.; Pickle, C. S.; Ralston, E. J.; Lee, A. H.; Amora, R.; Miller, J. C. et al. (2011). "Targeted Genome Editing Across Species Using ZFNs and TALENs". Science 333 (6040): 307. doi:10.1126/science.1207773. PMC 3489282. PMID 21700836 ; Tesson, L.; Usal, C.; Menoret, S. V.; Leung, E.; Niles, B. J.; Remy, S. V.; Santiago, Y.; Vincent, A. I. et al. (2011). "Knockout rats generated by embryo microinjection of TALENs". Nature Biotechnology 29 (8): 695. doi:10.1038/nbt.1940 ; Huang, P.; Xiao, A.; Zhou, M.; Zhu, Z.; Lin, S.; Zhang, B. (2011). "Heritable gene targeting in zebrafish using customized TALENs". Nature Biotechnology 29 (8): 699. doi:10.1038/nbt.1939 ; Doyon, Y.; Vo, T. D.; Mendel, M. C.; Greenberg, S. G.; Wang, J.; Xia, D. F.; Miller, J. C.; Urnov, F. D. et al. (2010). "Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures". Nature Methods 8 (1): 74-79. doi:10.1038/nmeth.1539. PMID 21131970 ; Szczeppek, M.; Brondani, V.; Buchel, J.; Serrano, L.; Segal, D. J.; Cathomen, T. (2007). "Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases". Nature Biotechnology 25 (7): 786. doi:10.1038/nbt1317. PMID 17603476 ; Guo, J.; Gaj, T.; Barbas Iii, C. F. (2010). "Directed Evolution of an Enhanced and Highly Efficient FokI Cleavage Domain for Zinc Finger Nucleases". Journal of Molecular Biology 400 (1): 96. doi:10.1016/j.jmb.2010.04.060. PMC 2885538. PMID 20447404 ; Mussolino, C.; Morbitzer, R.; Lutge, F.; Dannemann, N.; Lahaye, T.; Cathomen, T. (2011). "A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity". Nucleic Acids Research. doi:10.1093/nar/gkr597 ; Zhang, Feng; et.al. (February 2011). "Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription". Nature Biotechnology 29 (2): 149-53. doi:10.1038/nbt.1775. PMC 3084533. PMID 21248753 ; Morbitzer, R.; Elsaesser, J.; Hausner, J.; Lahaye, T. (2011). "Assembly of custom TALE-type DNA binding domains by modular cloning". Nucleic Acids Research. doi:10.1093/nar/gkr151 ; Li, T.; Huang, S.; Zhao, X.; Wright, D. A.; Carpenter, S.; Spalding, M. H.; Weeks, D. P.; Yang, B. (2011). "Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes". Nucleic Acids Research. doi:10.1093/nar/gkr188 ; Geiler, R.; Scholze, H.; Hahn, S.; Streubel, J.; Bonas, U.; Behrens, S. E.; Boch, J. (2011). "Transcriptional Activators of Human Genes with Programmable DNA-Specificity". In Shiu, Shin-Han. PLoS ONE 6 (5): e19509. doi:10.1371/journal.pone.0019509 ; Weber, E.; Gruetzner, R.; Werner, S.; Engler, C.; Marillonnet, S. (2011). "Assembly of Designer TAL Effectors by Golden Gate Cloning". In Benda hmane, Mohammed. PLoS ONE 6 (5): e19722. doi:10.1371/journal.pone.0019722 ; Sander et al. "Targeted gene disruption in somatic zebrafish cells using engineered TALENs". Nature Biotechnology Vol 29:697-98 (5 August 2011) Sander, J. D.; Cade, L.; Khayter, C.; Reyon, D.; Peterson, R. T.; Joung, J. K.; Yeh, J. R. J. (2011). "Targeted gene disruption in somatic zebrafish cells using engineered TALENs". Nature Biotechnology 29 (8): 697. doi:10.1038/nbt.1934 ; これらの全内容は、参照により本明細書に組み込まれる。

【 0 1 1 6 】

いくつかの態様において、本明細書に記載の T A L E N、T A L E Nドメイン、T A L E Nをコードするもしくは T A L E Nドメインをコードする核酸、組成物および試薬は、単離されている。いくつかの態様において、本明細書に記載の T A L E N、T A L E Nド

10

20

30

40

50

メイン、T A L E NをコードするもしくはT A L E Nドメインをコードする核酸、組成物および試薬は、精製されており、例えば、少なくとも60%、少なくとも70%、少なくとも80%、少なくとも90%、または少なくとも95%純粋である。

【0117】

本開示のいくつかの側面は、本明細書に記載の本発明のT A L E Nを用いて、核酸分子中の標的配列を切断する方法を提供する。いくつかの態様において、方法は、標的配列を含む核酸分子を、T A L E Nが標的配列に結合して切断するために好適な条件下で、標的配列に結合するT A L E Nと接触させることを含む。いくつかの態様において、T A L E Nは、単量体として提供される。いくつかの態様において、本発明のT A L E N単量体は、本発明の第1のT A L E N単量体と二量化してヌクレアーゼ活性を有するヘテロ二量体形成することができる別のT A L E N単量体を含む、組成物中に提供される。いくつかの態様において、本発明のT A L E Nは、医薬組成物中に提供される。いくつかの態様において、標的配列は、細胞内にある。いくつかの態様において、標的配列は、細胞のゲノム内にある。いくつかの態様において、標的配列は、対象中にある。いくつかの態様において、この方法は、T A L E Nを含有する例えば医薬組成物などの組成物を、対象に対して、T A L E Nが標的部位に結合してトランケートするのに十分な量で投与することを含む。

【0118】

本開示のいくつかの側面は、操作されたT A L E Nを製造する方法を提供する。いくつかの態様において、方法は、標準N末端T A L E Nドメインおよび/または標準C末端T A L E Nドメインの少なくとも1個のアミノ酸を、生理的pHで電荷を有さないかまたは負電荷を有するアミノ酸で置き換えること；および/または、N末端T A L E Nドメインおよび/またはC末端T A L E Nドメインをトランケートして、正に荷電した断片を除去すること；したがって、減少した正味電荷のN末端ドメインおよび/またはC末端ドメインを有する、操作されたT A L E Nを生成すること、を含む。いくつかの態様において、置換される少なくとも1つのアミノ酸は、カチオン性アミノ酸または生理的pHで正電荷を有するアミノ酸を含む。いくつかの態様において、少なくとも1つのアミノ酸を置き換えるアミノ酸は、カチオン性アミノ酸または中性のアミノ酸である。いくつかの態様において、トランケートされたN末端T A L E Nドメインおよび/またはトランケートされたC末端T A L E Nドメインは、それぞれの標準ドメインの残基の90%未満、80%未満、70%未満、60%未満、50%未満、40%未満、30%未満、または25%未満を含む。いくつかの態様において、トランケートされたC末端ドメインは、60個未満、50個未満、40個未満、30個未満、29個未満、28個未満、27個未満、26個未満、25個未満、24個未満、23個未満、22個未満、21個未満、または20個未満のアミノ酸残基を含む。

【0119】

いくつかの態様において、トランケートされたC末端ドメインは、60、59、58、57、56、55、54、53、52、51、50、49、48、47、46、45、44、43、42、41、40、39、38、37、36、35、34、33、32、31、30、29、28、27、26、25、24、23、22、21、20、19、18、17、16、15、14、13、12、11、または10個のアミノ酸残基を含む。いくつかの態様において、方法は、標準N末端T A L E Nドメインおよび/または標準C末端T A L E Nドメインの少なくとも2個、少なくとも3個、少なくとも4個、少なくとも5個、少なくとも6個、少なくとも7個、少なくとも8個、少なくとも9個、少なくとも10個、少なくとも11個、少なくとも12個、少なくとも13個、少なくとも14個、または少なくとも15個のアミノ酸を、生理的pHで電荷を有さないかまたは負電荷を有するアミノ酸で置き換えることを含む。いくつかの態様において、置き換えられるアミノ酸は、アルギニン(R)またはリジン(K)である。いくつかの態様において、生理的pHで電荷を有さないかまたは負電荷を有するアミノ酸は、グルタミン(Q)またはグリシン(G)である。いくつ

かの態様において、方法は、少なくとも1つのリジンまたはアルギニン残基を、グルタミン残基で置き換えることを含む。

【0120】

いくつかの態様において、本明細書で提供される改善されたT A L E Nは、組換え技術によって設計され、および/または生成される。いくつかの態様において、設計すること、および/または生成することは、所望の標的配列またはそれらの半部位に特異的に結合する、T A L E 反復アレイを設計することを含む。

【0121】

本開示のいくつかの側面は、本明細書に提供される操作されたT A L E NまたはかかるT A L E Nを含む組成物（例えば、医薬組成物）を含む、キットを提供する。いくつかの態様において、キットは、T A L E Nを賦形剤と接触させて、核酸をT A L E Nと接触させるのに好適な組成物を生成するための、賦形剤および指示書を含む。いくつかの態様において、賦形剤は、薬学的に許容し得る賦形剤である。

【0122】

典型的には、キットは、キットの構成要素を格納する容器、および、キットの構成要素を如何にして貯蔵し使用すべきかを述べた書面による指示書を含む。

【0123】

本発明のこれらおよび他の態様の機能および利点は、以下の例からより完全に理解されるであろう。以下の例は、本発明の利点を説明し、および特定の態様を記述することを意図しているが、本発明の完全な範囲を例示することは意図していない。したがって、例は本発明の範囲を限定するものではないことが理解されるであろう。

【実施例】

【0124】

例1

材料および方法

オリゴヌクレオチド、PCRおよびDNA精製

すべてのオリゴヌクレオチドは、Integrated DNA Technologies (IDT) から購入した。オリゴヌクレオチド配列は、表10に記載されている。PCRは、50 μ Lの1 \times HF緩衝液中の0.4 μ Lの2 U / μ L Phusion Hot Start II DNAポリメラーゼ (Thermo-Fisher)、0.2 mMのdNTPミックス (0.2 mMのdATP、0.2 mMのdCTP、0.2 mMのdGTP、0.2 mMのdTTP) (NEB)、0.5 μ M ~ 1 μ Mの各プライマーを用い、次のプログラムで実施した：特に断りのない限り、98、1分；[98、15秒；62、15秒；72、1分]を35サイクル。DNA反応物の多くは、以下ではQカラム精製と呼ばれるQIAquick PCR Purification Kit (Qiagen)、または以下ではMカラム精製と呼ばれるMinElute PCR Purification Kit (Qiagen) を用いて精製した。

【0125】

T A L E Nの構築

標準T A L E Nプラスミドは、FLASH法^{1 2}により、各T A L E Nが10 ~ 18塩基対を標的として構築した。N末端変異は、PCRによりQ5 Hot Start Master Mix (NEB) [98、22秒；62、15秒；72、7分]を用いて、リン酸化TAL-N1fwd (N1に対し)、リン酸化TAL-N2fwd (N2に対し)、またはリン酸化TAL-N3fwd (N3に対し)、およびリン酸化TALNrevをプライマーとして使用して行った。1 μ LのDpnI (NEB) を添加し、反応物を37で30分間インキュベートし、次いでMカラム精製した。溶出した~25 ngのDNAを、10 μ Lの2 \times Quick Ligase Buffer、1 μ LのQuick Ligase (NEB) 中、20 μ Lの総容量で、室温 (~21) で15分、分子内で平滑末端ライゲーションした。このライゲーション反応物1 μ Lを、トップ10のケミカルコンピテント細胞 (Invitrogen) に形質転換した。C末端ドメインの変異は、PCRによりTAL-CifwdとTAL-Cirevプライマーを用いてクローニングした後、Qカラム精製した。この溶出DNA ~ 1 ngを鋳型として、TALCifwdおよびTAL-Q3 (Q3に対し) またはTAL-Q7 (Q7に対し) の

何れかをプライマーとするPCRに使用し、次いでQカラム精製した。このDNA断片～1 μgを、1×NEBuffer 4中のHpaIおよびBamHIで消化し、HpaIおよびBamHIで前もって消化された約2 μgの所望のTALENプラスミド中にクローニングした。

【0126】

in vitroでのTALENの発現

全てが3×FLAGタグを含むTALENタンパク質を、in vitro転写/翻訳により発現させた。TALENをコードするプラスミドまたはプラスミドなし(「空の溶解物」対照)800 ngを、TNT(登録商標)Quick Coupled Transcription/Translation System、T7バリエーション(Promega)を最終容量20 μLで30 にて1.5時間用いて、in vitro転写/翻訳反応物中に添加した。ウェスタンブロットを使用し、抗FLAG M2モノクローナル抗体(Sigma-Aldrich)を用いてタンパク質を可視化した。TALEN濃度は、1 ng～16 ngのN末端FLAGタグ付き細菌アルカリホスファターゼ(Sigma-Aldrich)の標準曲線と比較することにより、算出した。

【0127】

DNA切断のためのin vitro選択

選択前ライブラリーを、部分的にランダム化された標的半部位配列(CCR5A、ATM、またはCCR5B)および完全にランダム化された10～24 bpスペーサー配列を含有する、10 pmolのオリゴライブラリーを用いて製造した(表10)。オリゴヌクレオチドライブラリーを、2.5 mMのMnCl₂を補足した1×CircLigase II反応緩衝液(33 mMのトリス酢酸、66 mMの酢酸カリウム、0.5 mMのジチオスレイトール、pH 7.5)の全20 μL中の、100単位のCircLigase II ssDNA Ligase(Epicentre)で60 にて16時間インキュベートすることにより、個別に環状化し、ついで80 で10分インキュベートした。2.5 μLの各環状化反応物を、Illustra TempliPhi 100 Amplification Kit(GE Healthcare)を使用するローリングサークル増幅のための基質として、30 で16時間、50 μLの反応物中で用いた。得られたコンカテマーライブラリーを、Quant-iT(商標)PicoGreen(登録商標)dsDNA Kit(Invitrogen)で定量し、異なるスペーサー長のライブラリーを、等モル比で混合した。

【0128】

CCR5B配列ライブラリーでの選択については、500 ngの選択前ライブラリーを、1×NEBuffer 3中、in vitro転写/翻訳TALEN+空の溶解物(合計30 μL)で37 にて2時間消化した。すべてのCCR5B TALENについて、in vitro転写/翻訳TALEN濃度をウェスタンブロットにより定量し(プロット中、TALENを4 で16時間保存)、次いでTALENを、単量体当たり40 nMの最終濃度で添加した。CCR5AおよびATM配列ライブラリーでの選択については、合わせた選択前ライブラリーを、300,000 MWCOスピンカラム(Sartorius)を用い、1×NEBuffer 3での3回の500 μL洗浄により、さらに精製した。125 ngの選択前ライブラリーを、1×NEBuffer 3中、合計24 μLの新鮮なin vitro転写/翻訳TALEN+空の溶解物で37 で30分消化した。すべてのCCR5AおよびATM TALENについて、6 μLのin vitro転写/翻訳左TALENおよび6 μLの右TALENを使用し、これらはそれぞれ、CCR5AおよびATM TALENについて、切断反応物の最終濃度16 nM±2 nMまたは12 nM±1.5 nMに対応する。これらのTALEN濃度は、消化と並行して行ったウェスタンブロットにより定量した。

【0129】

すべての選択について、TALEN消化ライブラリーを、1 μLの100 μg/μLのRNase A(Qiagen)で2分間インキュベートした後、Qカラム精製した。50 μLの精製DNAを、3 μLの10 mM dNTPミックス(10 mMのdATP、10 mMのdCTP、10 mMのdGTP、10 mMのdTTP)(NEB)、6 μLの10×NEBuffer 2、および1 μLの5 U/μL Klenow Fragment DNA Polymerase(NEB)で、室温で30分インキュベートした後、Qカラム精製した。50 μLの溶出DNAを、各試料に対応するバーコードを含む加熱および冷却#1アダプターでライゲートした(異なるTALEN濃

度またはコンストラクトによる選択) (表10A)。ライゲーションは、1×T4 DNA リガーゼ緩衝液(50 mMのTris-HCl、10 mMのMgCl₂、1 mMのATP、10 mMのDTT、pH 7.5)中で1 µLの400 U/µL T4 DNA リガーゼ(NEB)により60 µLの総容量で、室温にて16時間実施し、次いでQカラム精製した。

【0130】

6 µLの溶出DNAのPCRによる増幅を、150 µLの総反応容量(3×50 µLの反応物に分割)で、表10Aの#2Aのアダプタープライマーを用いて14~22サイクル実施した。PCR産物を、Qカラムで精製した。各DNA試料は、Quant-iT(商標)PicoGreen(登録商標)dsDNA Kit(Invitrogen)で定量化した後、等モル混合物にプールした。500 ngのプールされたDNAについて、5%TBEの18ウェルCriterion PAGEゲル(BioRad)を200 Vで30分実行し、~230塩基対の長さのDNA(1.5標的部反復+アダプター配列に対応する)を単離し、Qカラム精製した。溶出DNA~2 ngをPCRで5~8サイクル、#2Bアダプタープライマー(表10A)で増幅し、Mカラム精製した。

【0131】

10 µLの溶出DNAを、12 µLのAMPure XPビーズ(Agencourt)を用いて精製し、Illumina/Universal Library Quantification Kit(Kapa Biosystems)を用いて定量した。DNAは、Illuminaの指示に従ってハイスループットDNAシーケンシングのために製造し、MiSeq DNA Sequencer(Illumina)を用い、12 pMの最終溶液および156 bpのペアエンドリードを使用してシーケンシングした。シーケンシングのための選択前ライブラリーを製造するために、選択前ライブラリーを、1 µL~4 µLの適切な制限酵素で(CCR5A=Ts p45 I、ATM=Acc 65 I、CCR5B=Av a I(NEB))37で1時間消化した後、上記と同様2 pmo lの加熱および冷却#1ライブラリーアダプター(表10A)でライゲートした。選択前ライブラリーDNAを上記と同様に、#2Aライブラリーアダプタープライマーおよび#2Bライブラリーアダプタープライマーを、それぞれ#2Aアダプタープライマーおよび#2Bアダプタープライマー(表10A)の代わりに使用して製造した。得られた選択前ライブラリーDNAを、TALEN消化試料と共にシーケンシングした。

【0132】

個別のin vitro TALEN切断アッセイ

TALEN消化のための個別のDNA基質の構築を、表9Bに特定されたオリゴヌクレオチドのペアを制限クローニング14でpUC19(NEB)中に組み合わせることにより、実施した。対応するクローニングされたプラスミドの増幅を、PCRにより(59で15秒アニーリング)24サイクル、pUC190fwdおよびpUC190revプライマー(表10B)を用いて実施し、Qカラム精製した。50 ngの増幅したDNAを、1×NEBuffer 3中、それぞれ3 µLのin vitro転写/翻訳TALEN左および右単量体(~16 nMから~12 nMの最終TALEN濃度に相当)および6 µLの空の溶解物で、総反応容量120 µLで消化した。消化反応物を37で30分間インキュベートし、次いで1 µLの100 µg/µL RNase A(Qiagen)で2分間インキュベートし、Mカラム精製した。全体で10 µLの溶出DNAに、グリセロールを15%まで添加したものを、5%TBE 18ウェルCriterion PAGEゲル上で(Bio-Rad)、200 Vで45分間分析し、次に1×SYBR Gold(Invitrogen)で10分間染色した。バンドを、AlphaImager HP(Alpha Innotech)で可視化し、定量した。

【0133】

細胞のTALEN切断アッセイ

TALENを、哺乳動物発現ベクター12にクローニングし、得られたTALENベクターを前に記載のようにしてU2OS-EGFP細胞にトランスフェクトした^{1 2}。ゲノムDNAを、前に記載のようにして2日後に単離した^{1 2}。各アッセイについて、50 ngの単離したゲノムDNAの増幅を、PCRにより[98で15秒; 67.5で15秒; 72で22秒]、表10Cに指定のように4%DMSOありまたはなしのプライマー対を用

いて35サイクル実施した。各ゲノム部位のPCR反応物の相対的DNA含量は、Quant-iT (商標) PicoGreen (登録商標) dsDNA Kit (Invitrogen) で定量し、次いで等モル混合物中にプールし、TALENなしの試料と全てのTALEN処置試料とを分離した。150 ~ 350 bpに相当するDNAを、上述のようにPAGEによって精製した。

【0134】

44 µLの溶出DNAを、5 µLの1×T4 DNAリガーゼ緩衝液および1 µLの10 U/µLポリヌクレオチドキナーゼ (NEB) で、37 °Cで30分インキュベートし、Cカラム精製した。43 µLの溶出DNAを、1 µLの10 mMのdATP (NEB)、5 µLの10×NEBuffer 2、および1 µLの5 U/µLのDNAクレノウ断片 (3' 5' エキソ-) (NEB) で、37 °Cで30分インキュベートし、Mカラム精製した。10 µLの溶出DNAを、上記と同様に10 pmolの加熱および冷却G (ゲノム) アダプター (表10A) でライゲートした。8 µLの溶出DNAを、PCRにより、各試料に対応するバーコードを含むG-Bプライマーで6~8サイクル増幅した。各試料のDNAを、Quant-iT (商標) PicoGreen (登録商標) dsDNA Kit (Invitrogen) で定量し、次いで等モル混合物中にプールした。合わせたDNAを、上記のようにMiSeqを用いてハイスループットシーケンシングに供した。

【0135】

データ分析

Illuminaシーケンシングリードをフィルタリングし、アルゴリズムのセクションで概説されるようにUnix Bashで記述されたスクリプトで解析した。ソースコードは、リクエストにより入手可能である。特異性スコアは、前に記載のようにして算出した¹⁴。表3の様々なTALEN選択中の変異数の分布に関する統計解析は、前に記載のようにして実施した¹⁴。表7の修飾部位の統計解析は、前に記載のようにして実施した¹⁴。

【0136】

アルゴリズム

すべてのスクリプトはbashまたはMATLABで記述された。

【0137】

選択前配列および選択された配列のコンピューターによるフィルタリング

選択前配列について

1) 最初の4塩基 (ランダムな塩基) のリードの直後の、16 bpの定常配列 (CCR5 A = CGTCACGCTCACCCT、CCR5 B = CCTCGGGACTCCACGCT、ATM = GGTACCCCACTCCGCGT) を探索し、1つの変異を可能にする16 bp定常配列を有する配列のみを承認する。

2) 定常配列から、少なくとも最小可能限度の完全部位長から最大の完全部位長までの距離で離れている位置における、9 bpの最終配列を探索して、完全部位の存在を確認し、この9 bpの最終配列を有する配列のみを承認する。(最終配列: CCR5 A = CGTCACGCT、CCR5 B = CCTCGGGAC、ATM = GGTACGTGC)。

3) 完全部位内の各半部位の最良の例を探索し、左の後に右という、正しい左および右半部位の順序のすべての配列を承認する。

4) 2つの半部位と、左半部位の左の単一隣接ヌクレオチドおよび右半部位の右の単一隣接ヌクレオチドの間の、DNAスペーサー配列 (半部位と定常配列の間の配列) を決定する。

5) シーケンシングリード品質スコアによりフィルタリングして、半部位の位置の4分の3に渡って品質スコアA以上の配列のみを承認する。

【0138】

選択された配列について

1) 異なる選択条件に対応する正しい5 bpのバーコードから開始される全ての配列の、全配列のリードおよび位置の品質スコアを、別々のファイルに出力する。

2) 初期配列から少なくとも最小可能限度の完全部位長離れ、かつ初期配列から最大の完全部位長離れている位置で反復される、5 bpのバーコードの直後の初期16 bp配列を探索して、反復配列を有する完全部位の存在を確認し、1つの変異を可能にする16 bp

の反復を有する配列のみを承認する。

3) 完全部位内の16bpの定常配列を探索し、1つの変異を可能にする定常配列を有する配列のみを承認する。配列の解析を、定常配列+5'配列から開始して2番目の反復配列まで、次にバーコードの後の初期配列から定常配列まで実施して、1つの完全部位の均等物を挟む複数定常配列を得る。

CONSTANT - LFLANK - LHS - SPACER - RHS - RFLANK - CONSTANT

LFLANK = 左隣接配列 (単一のランダム塩基として設計)

LHS = 左半分部位配列

RHS = 右半分部位配列

RFLANK = 右隣接配列 (単一のランダム塩基として設計)

10

CONSTANT = 定常配列 (CCR5A = CGTCACGCTCACCCT、CCR5B = CCTCGGGACTCCACGCT、ATM = GGTACCCCACTCCGCGT)

4) 完全部位内の各半部位の最良の例を探索し、左の後に右という、正しい左および右半部位の順序のすべての配列を承認する。

5) 半部位の位置で、対応するスペーサー (2つの半部位間の配列)、左隣接および右隣接配列 (半部位と定常配列の間の配列) を決定する。

6) 5bpのバーコード配列の後のリードの開始から、定常配列の始まりまでの配列を得ることにより、配列末端を決定する。

SEQUENCESTART - RHS - RFLANK - CONSTANT

7) シーケンシングリード品質スコアによりフィルタリングして、半部位の位置の4分の3に渡って品質スコアA以上の配列のみを承認する。

20

8) 選択された配列を、スペーサー側からの各半部位中に0~5個の塩基対の末端を有する配列の数の平均として計算された配列末端の背景の量よりも2.5倍豊富なスペーサー中に配列末端を有する配列のみを承認することによって (配列末端の背景数は、比較のための配列末端の背景として利用した配列末端に最も近い半部位を有する両方の半部位について計算した)、配列末端によってフィルタリングした。

【0139】

CCR5B標的部位に関連するゲノムオフターゲット部位のコンピューター探索

1) Patmatchプログラム³⁹を使用して、パターン配列について次のようにヒトゲノム (GRCh37/hg19ビルド) を探索した: CCR5B左半部位配列 (L16、L13またはL10) NNNNNNNN... CCR5B右部位配列 (R16、R13またはR10) [M, 0, 0]、ここでNの数は12~25で変化し、およびM (可能な変異を示す) は0~14で変化する。

30

2) プログラムがX個以下の変異を有するすべての配列を出力するため、出力されるオフターゲット部位の数は脱累積されて、標的部位から離れた変異の特定数である、ヒトゲノム中のオフターゲット部位の数がもたらされるからである。

【0140】

ゲノム部位の配列におけるインデルの同定

1) 各配列について、プライマー配列を使用してゲノム部位を同定した。

2) 標的部位の左の8bpに対応する参照ゲノム配列、および完全標的部位の右の8bp (またはシーケンシングリードの最末端のゲノム部位については6bp) の参照ゲノム配列を含む配列は、標的部位配列と考えられた。

40

3) 参照ゲノム部位と同じ大きさに対応する任意の標的部位配列は、非修飾と考えられ、参照サイズではない任意の配列は、ClustalW⁴⁰で参照ゲノム部位と整列させた。

4) 2つ半部位配列の間のDNAスペーサー配列における2つより多くの挿入または2つより多くの欠失を有する整列された配列は、インデルと考えられた。

【0141】

結果

CCR5およびATMを標的とするTALENの特異性プロファイリング

我々は、種々の長さの左および右半部位を標的とするTALEN、および異なるドメイ

50

ンバリエントを有するT A L E Nを含む、合計41のヘテロ二量体T A L E N対（以下T A L E Nと呼ぶ）の特異性をプロファイリングした。41のT A L E Nのそれぞれは、2つの異なるヒト遺伝子C C R 5およびA T Mにおける、3つの異なる配列（C C R 5 A、C C R 5 B、またはA T Mと呼ぶ）の1つを標的とするように設計した（図7）。我々は、前述のin vitro選択法^{1 4}の改良版に、選択のスループットおよび感度を高める変更を加えて使用した（図1 B）。

【0142】

簡単に述べると、 $> 10^{12}$ のDNA配列の選択前ライブラリーにおいて、それぞれを、3 nM ~ 40 nMのin vitroで翻訳されたT A L E Nで消化した。これらの濃度は、ヒト細胞核当たり~ 20から~ 200の二量体T A L E N分子に相当し^{2 1}、これは比較的10
低レベルの細胞タンパク質発現である^{2 2, 2 3}。切断されたライブラリーメンバーは、アダプターライゲーションによって捕捉してゲル精製により単離した、遊離の5'ーリン酸を含んでいた（図1 B）。対照試料において、選択前ライブラリーのすべてのメンバーは、定常配列において制限エンドヌクレアーゼにより切断されており、これらがアダプターライゲーションによって捕捉されゲル精製によって単離されることが可能であった。この選択プロセスを耐えたT A L E N処置試料または対照試料のハイスループットシーケンシングとコンピューター解析により、T A L E Nで切断されたすべての配列の存在量だけでなく、選択前の対応する配列の存在量が明らかにされた。選択を耐えた各ライブラリーメンバーの濃縮値の算出は、その選択後配列存在量を、その選択前存在量で割り算することにより行った。選択前DNAライブラリーは、その各々が、理論的には、オンターゲット20
配列に対して6個以下の変異を有する全ての可能なDNA配列の少なくとも10のコピーを含有するだけ、十分に大きかった。

【0143】

試験した全41のT A L E Nについて、選択を耐えたDNAは、選択前ライブラリーに存在したよりも有意に少ない平均変異を標的半部位に含んでいた（表3および4）。例えば、18 bpの左および右半部位を標的とするT A L E Nで処置後の選択を耐えたDNA配列における変異の平均数は、C C R 5 A配列について4.06、A T M配列について3.18であり、比較して、対応する選択前ライブラリーではそれぞれ7.54と6.82の変異であった（図2 Aおよび2 B）。すべての選択について、オンターゲット配列は8 ~ 640倍に濃縮された（表5）。我々の選択結果をin vitroで検証するために、13 bpの左および右半部位（L13 + R13）を標的とするC C R 5 B T A L E Nの、16の多様なオフターゲット基質のそれぞれを切断する能力をアッセイした（図2 Eおよび2 F）。得られた個別のin vitro切断効率は、観測された濃縮値と良好に相関した（図2 G）。

【0144】

全4つの可能な塩基対に対する、T A L E N標的部位における各位置での特異性を決定するために、特異性スコアを選択前と選択後の塩基対頻度の差として算出し、これを、完全な特異性（1.0と定義）から完全な反特異性（anti-specificity）（-1.0と定義）までの選択前頻度の最大可能な変化について、正規化した。試験したすべてのT A L E Nについて、両方の半部位内の全ての位置で標的とされた塩基対が好ましく、ただし唯一の例外は、右半部位でのいくつかのA T M T A L E Nについてのスパーサーに最も近い塩基対であった（図2 C、2 Dおよび図8 ~ 13）。N末端ドメインによって認識される5' Tヌクレオチドが高度に特異的であり、5' DNA末端（N末端T A L E N末端）は、一般に、3' DNA末端よりも高い特異性を示す；両方の観測は、以前の報告と整合する^{2 4, 2 5}。まとめると、これらの結果は、選択データが正確にin vitroでのオフターゲットT A L E N切断の効率を予測し、T A L E Nが、標的配列全体にわたって高度に特異的であることを示す。

【0145】

細胞におけるT A L E Nのオフターゲット切断

選択によって報告されたオフターゲット切断活性が、細胞でのオフターゲット切断に関

10

20

30

40

50

連しているかどうかを試験するために、我々はin vitroでの選択結果を用いて、機械学習アルゴリズムを、ヒトゲノムで潜在的なT A L E Nオフターゲット部位を生成するように訓練した^{2 6}。このコンピューターのステップが必要であるのは、選択前ライブラリーは6個以下の変異を有する全ての配列をカバーするが、一方C C R 5配列およびA T M配列については、ヒトゲノムのほぼすべての潜在的なオフターゲット部位が、標的配列に対して6より多くの位置で異なるからである。このアルゴリズムは、切断が観察されなかった標的ライブラリーからの配列に対する、T A L E Nにより切断された配列で生じる、標的の各位置の各ヌクレオチドの事後確率を算出する^{2 7}。次にこれらの事後確率を用いて、アルゴリズムの訓練に使用したT A L E Nが、10 ~ 30 b pの単量体スペーシングのヒトゲノムにおける、全ての潜在的標的配列を切断する尤度(likelihood)をスコア付けした。機械学習アルゴリズムを使用して、我々は、7 ~ 14の位置においてオンターゲット配列と異なる、36のC C R 5 Aおよび36のA T M T A L E Nオフターゲット部位を同定した(表6)。

【0146】

C C R 5 AおよびA T M T A L E Nについての72のベストスコアのゲノムオフターゲット部位を、C C R 5 AまたはA T MいずれかのT A L E Nを発現するヒトU2OS-EGFP細胞12から精製したゲノムDNAから増幅した³。潜在的なゲノムオフターゲット部位のDNAスペーサー内に3または4以上の塩基対の挿入または欠失を含む配列であって、未処置の対照試料と比べてT A L E N処置試料中に有意に高い数値で存在するものは、T A L E N誘導性の修飾と考えられた。我々が首尾良く増幅した35のC C R 5 Aオフターゲット部位から、T A L E N誘導性の修飾を有する6つのオフターゲット部位を同定した；同様に、我々が首尾良く増幅した31のA T Mオフターゲット部位から、T A L E N誘導性の修飾を有する7つのオフターゲット部位を観測した(図3および表7)。修飾されたオンターゲットおよびオフターゲット部位の検査により、3から数十に及ぶ塩基対の範囲の欠失の保有率が得られ(図3)、これは、T A L E N誘導性のゲノム修飾の、以前に説明された特徴と整合していた^{2 8}。

【0147】

これらの結果をまとめると、機械学習アルゴリズムによって処理されたin vitro選択データが、ヒト細胞においてT A L E N誘導性修飾を受ける真正のオフターゲット基質を予測できることを示す。T A L E N反復は、相対的独立性で塩基対に生産的に結合する。選択データにおいて定量的に特徴付けられた多数のオフターゲット基質は、標的配列内の1つの位置での変異が、他の位置に生産的に結合するT A L E N反復の能力に影響を与えるかどうか評価することを可能にした。我々は、L 13 + R 13 C C R 5 B T A L E Nに対するあらゆる可能な二重変異体配列の予測濃縮値を、対応する2つの単一変異濃縮からの独立した寄与を仮定して生成した。一般的には、予測濃縮値は、各二重変異配列について実際に観測された濃縮値に非常に類似しており(図14A)、成分単一変異は、二重変異配列の全体のトランケート性に独立して寄与することを示唆する。観測と予測の二重変異濃縮値の間の差は、2つの変異の間の距離とは比較的無関係であったが、ただし、2つの隣接するミスマッチは、予想されるよりもわずかに良好に耐容されていた(図14B)。

【0148】

2つ以上の変異の潜在的な相互依存関係を決定するために、我々は、L 13 + R 13 C C R 5 B T A L E Nについての選択後の標的における、選択濃縮値と変異数の関係を評価した(図4A、黒線)。0 ~ 5の変異について、濃縮値は、変異の平均数(m)の単純な指数関数に密接に従った(表8)。この関係は、連続した各変異が一定量(G)ずつ結合エネルギーを低減し、T A L E N結合($K_{eq}(m)$)が、 $K_{eq}(m) \sim e^{-G \cdot m}$ という指数関数的減少をもたらすというモデルと整合する。したがって、観測された指数関数的な関係は、典型的なミスマッチからの結合エネルギーの平均の減少が、T A L E N : DNA相互作用中に既に存在するミスマッチの数とは無関係であることを示唆する。まとめると、これらの結果は、T A L E N反復がそれぞれのDNA塩基対に、隣接のミスマッチに対するわずかに増加した耐性を超えて、独立して結合することを示す。

【0149】

長いT A L E Nは認識塩基対当たりの特異性が低い

T A L E N反復の独立した結合は、塩基対当たりのT A L E N特異性が、標的部位の長さとは無関係であることを単純化して予測する。T A L E Nアレイの長さとおフターゲット切断の間の関係を実験的に特徴づけるために、我々は、10、13、または16bp(5' Tを含む)を標的とするT A L E Nを、左(L10、L13、L16)および右(R10、R13、R16)の両方の半部位について構築した。左および右のC C R 5 B T A L E Nの9つすべての可能な組み合わせを表すT A L E Nを、in vitro選択に供した。結果は、より短いT A L E Nが、より長いT A L E Nよりも、標的化塩基対当たりより大きな特異性を有することが明らかになった(表3)。例えば、L10 + R10 T A L E Nにより切断された配列は、認識塩基対当たり平均して0.032の変異を含んでいるのに対し、L16 + R16 T A L E Nにより切断されたものは、認識塩基対当たり平均して0.067の変異を含んでいた。16 + 16塩基対を標的とする最長のC C R 5 B T A L E N、または18 + 18bpを標的とするC C R 5 AおよびA T M T A L E Nによる選択の場合、平均の選択濃縮値は、変異数の関数としての単純な指数関数的減少には従わない(図4Aおよび表8)。

10

【0150】

我々は次のように仮定した：長いT A L E Nにおけるより多数のT A L E N反復からの過剰な結合エネルギーは、より多くの変異を有する配列の切断を可能にすることにより、より少ない変異を有する配列において対応する切断の増加なしで、特異性を低下させるが、これは、後者の配列が既にほぼ完全に切断されているからである。実際、これらのより長いT A L E Nについての個々のDNA配列のin vitro切断効率は、標的部位における少数の変異の存在とは無関係であり(図5C - 5F)、わずかな変異を含有する配列のほぼ完全な結合および切断を示唆する。同様に、より高いT A L E N濃度はまた、わずかな変異を有する配列の濃縮値の減少をもたらし、一方で多くの変異を有する配列の濃縮値を増加させる(表5)。これらの結果は共に、長いT A L E Nアレイまたは高いT A L E N濃度のいずれかから生じた、過剰なT A L E N結合が、各認識塩基対で観測されたT A L E N DNA切断特異性を低下させるというモデルを支持する。

20

【0151】

長いT A L E Nはゲノムの場面において少ないおフターゲット切断を誘導する

長いT A L E Nは短いT A L E Nより、ミスマッチ配列(図4A)に対してより耐容であるが、ヒトゲノムには、短い標的部位よりも長い標的部位について、はるかに少ない数の密接に関連するおフターゲット部位が存在する(図4B)。おフターゲット部位の存在量と切断効率は両方とも、ゲノムの場面におけるおフターゲット切断イベントの数に寄与するため、我々は、全体的なゲノムの切断特異性をT A L E N長の関数として、与えられた長さの変異配列の外挿平均濃縮度値に、ヒトゲノムにおける対応する変異配列の数を乗じて算出した。より長い標的部位の長さに起因する潜在的なおフターゲット部位の存在量の減少は、より長いT A L E Nについて観測された認識塩基対当たりの特異性の減少を上回るに十分な大きさである(図4C)。その結果、長いT A L E Nは、短いT A L E Nよりも、20 ~ 32bpの部位を標的とした試験されたT A L E N長さに関して、ヒトゲノムにおける潜在的な切断部位のセットに対しより特異的であることが、予測されている。

30

40

【0152】

改善された特異性を有するT A L E Nの操作

上記の知見は、T A L E N特異性を、効率的なおフターゲット切断を可能にするために必要とされるものを超えた非特異的DNA結合エネルギーを低減することによって、改善可能であることを示唆する。最も広く使用されている、T A L E N反復アレイとF o k Iヌクレアーゼドメインの間の63 - a a C末端ドメインは、10個のカチオン性残基を含む。我々は、標準T A L E N C末端ドメインのカチオン性電荷を減少させることは、非特異的DNA結合を減少させ²⁹、T A L E N特異性を改善するであろうと推測した。

【0153】

50

我々は、2つのC末端ドメインバリエーションを構築し、ここで、標準63 - aa C末端ドメインの3個(「Q3」、K788Q、R792Q、R801Q)または7個(「Q7」、K777Q、K778Q、K788Q、R789Q、R792Q、R793Q、R801Q)のカチオン性ArgまたはLysは、Glnに変異されている。我々は、標準、操作Q3、および操作Q7のC末端ドメイン、ならびに、Q7 C末端ドメインと同一の理論正味電荷(-1)を有する、以前に報告された28 - aaトランケートされたC末端ドメイン⁵、を含有するCCR5AおよびATM TALEN上でin vitroでの選択を行った。

【0154】

CCR5AとATMの選択のためのオンターゲット配列濃縮値は、C末端ドメインの正味電荷が減少するにつれて、実質的に増加した(図5Aおよび5B)。例えば、ATMの選択は、Q7、Q3、および標準63 - aaのC末端バリエーションそれぞれについて、510、50、および20のオンターゲット濃縮値をもたらした。これらの結果は、C末端ドメイン中のカチオン性残基が部分的に中性残基によって置換または完全に除去されたTALENバリエーションが、標準63 - aa C末端ドメインを含むTALENよりも、in vitroで実質的により特異的であることを示唆する。同様に、TALEN N末端における1つ、2つ、または3つのカチオン性の残基をGlnに変異させることもまた、切断特異性を増加させた(表5、および図8 - 11)。

【0155】

Q7の、in vitroの標準63 - aa C末端ドメインを超える、高いDNA切断特異性を確認するために、16のオフターゲットDNA基質の代表的なコレクションを、in vitroで標準または操作いずれかのQ7 C末端ドメインを含むTALENで消化した。標準63 - aa C末端ドメインTALENを有するATMおよびCCR5A TALENは、0、1、または2つの変異を有する標的部位に、同等のin vitro切断活性を実証する(図5C - 5F)。これとは対照的に、試験された16のオフターゲット基質のうち11については、操作されたQ7 TALEN変異体は、標準63 - aa C末端ドメインTALENよりも、1または2つの変異を有するオフターゲットDNA基質に対して実質的に高い(〜4倍以上)識別を示し、ただし、Q7 TALENはそれぞれのオンターゲット配列を、標準63 - aa C末端ドメインTALENと同等かそれ以上の効率で切断した(図5C - 5F)。全体的に、個別の切断アッセイは選択結果と一致しており、操作されたQ7 C末端ドメインを有するTALENが標準63 - aa C末端ドメインのTALENよりも、in vitroで実質的により特異的であることを示す。

【0156】

ヒト細胞における操作されたTALENの改善された特異性

in vitroで観測された、操作されたTALENの特異性の増加が、ヒト細胞にも適用されるかどうかを決定するために、オンターゲット部位およびトップ36の予測オフターゲット部位のTALEN誘導性の修飾率を、標準63 - aa、Q3、またはQ7のC末端ドメインおよびELD/KKまたはELD/KKRのFokIドメインの全6つの可能な組み合わせを含むCCR5AおよびATM TALEN(合計12のTALEN)について測定した。

【0157】

両方のFokIバリエーションについて、Q3 C末端ドメインのTALENは、8%〜24%の修飾の範囲の顕著なオンターゲット活性を実証し、これは、標準63 - aa C末端ドメインのTALENの活性に匹敵する。標準63 - aaまたはQ3 C末端ドメインおよびELD/KKR FokIドメインを有するTALENは両方とも、細胞内のCCR5AおよびATMオンターゲット部位を修飾するのに、Q7 C末端ドメインの対応するTALENよりも、〜5倍、より活性である(図6および表7)。

【0158】

in vitroで観測された改善された特異性と一致して、操作されたQ7 TALENは、Q3バリエーションよりも特異的であり、これは次に標準63 - aa C末端ドメインTALEN

ENよりも特異的である。標準63 - aa C末端ドメインと比較して、Q3 C末端ドメインのTALENは、ELD/KKRのFokIドメインを有するCCR5AおよびATM部位についてそれぞれ、13倍および9倍以上も、細胞特異性の平均増加（オンターゲット部位のオフターゲット部位に対する細胞修飾率の比として定義される）を示す（表7）。これらの平均の改善は、多くのオフターゲット配列に対しては操作されたTALENによる観測された切断イベントがないかまたはほとんどないために、下限値としてのみ表すことができる。最も多く切断されたオフターゲット部位について（CCR5Aオフターゲット部位#5）、Q3 C末端ドメインは、標準63 - aa C末端ドメインより34倍さらに特異的であり（図6）、Q7 C末端ドメインは>116倍さらに特異的である。

【0159】

10

これらの結果は共に、CCR5およびATM配列を標的とするために、標準63 - aa C末端ドメインを操作されたQ3 C末端ドメインで置き換えることが、細胞内でのオンターゲット部位に対して匹敵する活性、最も容易に切断されるオフターゲット部位に対して細胞における特異性の34倍の改善、およびその他のオフターゲット部位に対する特異性の一貫した増加をもたらすことを明らかにする。低い活性が必要な場合は、操作されたQ7 C末端ドメインは、特異性の付加的な増加を提供する。

【0160】

改善されたTALEN DNA切断特異性のためのN末端ドメインの操作

本明細書に記載のTALEN結合および特異性のモデルは、過剰なTALEN結合エネルギーを低減することが、TALENのDNA切断特異性を増加させることを予測する。この予測をさらに試験し、TALEN特異性を潜在的にさらに強化するために、我々は、CCR5AおよびATMを標的とするTALENのN末端ドメインにおいて、1個（「N1」、K150Q）、2個（「N2」、K150QおよびK153Q）、または3個（「N3」、K150Q、K153Q、およびR154Q）のLysまたはArg残基を、Glnに変異させた。これらのN末端残基は、以前の研究において、DNAに非特異的に結合することが示されており、これらの特定の残基における、カチオン性電荷を中和するための変異は、非特異的DNA結合エネルギーを減少させる³³。我々は、これらのN末端変異からの非特異的結合エネルギーの減少は、過剰なTALEN結合エネルギーを減少させ、特異性の増加をもたらすとの仮説をたてた。これら3つのTALENバリエーションでのin vitro選択は、より少ないカチオン性のN末端TALENが実際に、オンターゲット切断のより大きな濃縮値を示すことを明らかにした（表5）。

20

30

【0161】

N末端およびC末端ドメインならびにTALEN濃度の特異性に対する効果

試験したすべてのTALENコンストラクトは、両方の半部位にわたって意図した塩基対を特異的に認識するが（図8～13）、ただし、ATM TALENのいくつかは、スパーサーに隣接する塩基対とは特異的に相互作用しない（ほとんどのC末端TALE反復により標的化）（図10および11）。標準TALENと、操作されたC末端またはN末端ドメインを含有するものの、広い特異性プロファイルを比較するために、標準、Q3、またはQ7のC末端ドメインおよびN1、N2、またはN3のN末端ドメインを有するCCR5AおよびATM TALENを使用した選択からの各標的塩基対の特異性スコアから、標準TALEN（標準63 - aa C末端ドメイン、野生型N末端ドメイン）上の選択からの対応する特異性スコアを差し引いた。

40

【0162】

結果を図15に示す。特異性を増加させるC末端ドメインにおける変異は、各半部位の中央およびC末端において最も強く特異性を増加させた。同様に、N末端における特異性増大変異は、TALENのN末端（5' DNA末端）付近の位置で最も強く特異性を増加させる傾向にあったが、ただし、右半部位を標的とするATM TALENのN末端における変異は、特異性を顕著に変化させなかった。これらの結果は、いずれかの末端での弱い結合が、この末端近くのTALE反復における特異性の増加を要求するとの、局所的結合補償モデルと整合する。TALEN濃度の特異性に対する効果を特徴付けるために、3

50

nM ~ 16 nM の範囲の3つの異なる濃度で行ったATMおよびCCR5A TALENの選択からの特異性スコアから、アッセイした最高TALEN濃度、すなわちATMについて24 nM、またはCCR5Aについて32 nMで行った対応する選択からの特異性スコアを差し引いた。結果(図15)は、特異性スコアが、TALENの濃度が減少するにつれて、半部位にわたってかなり均一に増加することを示す。

【0163】

DNAスパーサー長と切断部位の選好

様々なTALEN構築物(C末端変異、N末端変異、およびFokIバリエーション)のスパーサー長および様々なTALEN濃度の選好を評価するために、10 ~ 24塩基対のスパーサー長を有するライブラリーメンバーの濃縮値の算出を、標準、Q3、Q7、または28 - aa C末端ドメイン; N1、N2、またはN3 N末端変異; およびEL/KKまたはELD/KKR FokIバリエーションの様々な組み合わせを有するCCR5AおよびATM TALENによる選択のそれぞれについて、4 nM ~ 32 nMのCCR5AおよびATM TALENにて実施した(図16)。試験した全濃度において、N末端バリエーション、C末端バリエーション、およびのFokIバリエーションは、3つの顕著な例外を除いて、14 ~ 24塩基対の範囲の広いDNAスパーサー長の選好を示した。第1に、以前の報告と整合して³⁴⁻³⁶、CCR5A 28 - aa C末端ドメインは、標準C末端ドメインのより広範なDNAスパーサー長の選好よりはるかに狭い、DNAスパーサー長の選好を示した。第2に、Q7 C末端ドメインを含有するCCR5A TALENは、標準C末端ドメインバリエーションと比較して、12塩基のスパーサーに対する耐性の増加を示した(図16)。このわずかに広がったスパーサー長の選好は、おそらくはTALEN: DNA境界にそったより少数の非特異的タンパク質: DNA相互作用から生じる、Q7 C末端ドメインのより大きな立体配座的柔軟性を反映している可能性がある。第3に、Q7 C末端ドメインのATM TALENおよびN3変異体のN末端ドメインのATM TALENは、狭いスパーサーの選好を示した。

【0164】

これらの、低いDNA結合親和性を有するより特異的なTALEN(表5)は、非最適なDNAスパーサー長の切断速度と競争力のある、より速いオフ速度(off-rate)を有することができ、観測されたスパーサー長の選好を変化させる。以前の報告は、TALEN C末端ドメインの長さをDNAスパーサー長の選好の主要な決定因子として焦点を当てているが、これらの結果は、C末端ドメインの正味の電荷ならびに全体的なDNA結合親和性もまた、TALENのスパーサー長の選好に影響を与える可能性があることを示唆する。

【0165】

我々はまた、スパーサー内のTALENのDNA切断の位置を特徴付けた。我々は、各配列の右半部位に先だって観測されるスパーサーDNA塩基の数を報告するヒストグラムを、標準、Q3、Q7、または28 - aa C末端ドメイン; N1、N2、またはN3 N末端変異; およびEL/KKまたはELD/KKR FokIバリエーションの様々な組み合わせを有するCCR5AおよびATM TALENによる選択からの配列の各々において作製した(図17)。ヒストグラムのピークは、スパーサー内のDNA切断の最も可能性の高い場所を表すと解釈された。切断位置は、TALEN C末端ドメインとDNAスパーサー長によって課される立体配座的制約から予想されるように、TALEN結合半部位の間のDNAスパーサー長に依存する。

【0166】

考察

10¹²の密接に関連するオフターゲット配列でチャレンジした41のTALENのin vitro選択およびその後の分析は、次の4つの重要な知見を通して、TALEN特異性についての我々の理解を報告する: (i) TALENは、その意図する標的塩基対に対して、全ての位置において高度に特異的であり、各TALE反復アレイのN末端TALEN末端近くで(結合したDNAの5'末端に相当)特異性が増加する; (ii) 長いTALE

Nは、ゲノムの場面においてより特異的であり、一方短いT A L E Nは、ヌクレオチド当たりのより高い特異性を有する；(i i i) T A L E N反復はそれぞれ、対応する塩基対に比較的独立的に結合する；および(i v) 過剰なD N A結合親和性は、オフターゲット部位に対してT A L E N活性を増加させ、したがって特異性の減少をもたらす。

【 0 1 6 7 】

C末端ドメインまたはN末端におけるより多くのT A L E N反復またはより多くのカチオン性残基を有するT A L E Nについての特異性の、観測された減少は、過剰なT A L E N結合親和性が混乱の増加につながるというモデルと整合する。過剰な結合エネルギーはまた、長いC末端ドメインのT A L E Nの5'末端Tで以前に報告された混乱を説明し³⁰、また高いT A L E Nタンパク質濃度がin vivoでより多くのオフターゲット部位の切断をもたらすとの報告とも整合する⁹。細胞内でのT A L E Nタンパク質発現の低減は、理論的には、オフターゲット切断を減らすことができるが、いくつかのT A L E Nコンストラクトのそれらの標的D N A配列に対するK d値は既に、ヒト細胞の核内での理論的な最小のタンパク質濃度である~ 0.2 nMと、同等かまたはそれ以下である可能性がある。

【 0 1 6 8 】

かかるT A L E Nの特異性を、それらの発現レベルを低下させることによって改善することの困難さは、所望のレベルのオンターゲット切断を達成するのに十分なT A L E N濃度を維持する必要性と相まって、本研究で記載されているような、より高い固有の特異性を有するT A L E Nを操作することの価値を強調する。我々の知見は、低下した非特異的D N A結合を有する変異体C末端ドメインを使用して、T A L E NのD N A結合親和性を微調整し、オンターゲット配列が効率的にしかし最小限の過剰な結合エネルギーで切断され、オンターゲット部位とオフターゲット部位の間のより良好な識別をもたらし得ることを示唆する。46の総塩基対までを標的とするT A L E Nは、細胞において活性であることが示されているため¹⁵、ここで提示された結果は、特異性が、変異N末端およびC末端ドメインの組み合わせを有する操作されたT A L E Nによって、さらに改善することができるという考えと整合し、ここで該T A L E Nは、低減した非特異的D N A結合および、追加のオンターゲットD N A結合に寄与するより多数のT A L E N反復、および、Gを認識するより特異的な(しかし低親和性の)N K R V Dを、付与するものである^{25, 31}。

【 0 1 6 9 】

本研究では、S E L E Xまたはインテグラーゼ欠損レンチウイルスベクター(I D L V)などの方法³²を使用した他の研究よりもより多くの、真正なT A L E Nゲノムのオフターゲット部位を同定した。我々のモデルおよび得られた改善T A L E Nは、ゲノム中に存在する目的の標的配列に密接に関連する少ない数の配列により本質的に制限されている、細胞オフターゲット切断方法からは、または、D N A切断活性を測定せず、したがって活性な二量体T A L E Nを特徴づけない、単量体T A L E N反復アレイによるS E L E X実験⁵からは、得ることが困難であろう。対照的に、本研究の各T A L E Nは、哺乳動物ゲノムにおける異なる配列の数よりも数桁大きいライブラリーサイズの、そのオンターゲット配列の10¹²個の密接なバリエーションのいずれかを切断する能力について評価した。オフターゲット配列空間のこの密な有効範囲により、D N A切断の特異性と、標的塩基対の位置、T A L E N反復長、T A L E N濃度、ミスマッチの位置、および操作されたT A L E Nドメイン組成物との間の詳細な関係の解明が可能となった。

【 0 1 7 0 】

例 2

標準N末端ドメイン配列の少なくとも1つのカチオン性アミノ酸残基が、生理学pHで電荷を示さないかまたは負電荷を示すアミノ酸残基で置き換えられている、多数のT A L E Nを生成した。T A L E Nは、それらのN末端ドメインにおいて、グリシン(G)および/またはグルタミン(Q)の置換を含んでいた(図18参照)。操作されたT A L E Nの切断の選好の評価は、グリシン(G)に対する変異が、グルタミン(Q)に対するものと同等であることを実証した。T A L E NのN末端ドメイン(K 1 5 0 Q、K 1 5 3 Q、

および R 1 5 4 Q) における正に荷電したアミノ酸の変異は、Q または G のいずれかへの変異に対する結合親和性およびオフターゲット切断に、類似の減少をもたらした。例えば、M 3 および M 4 の N 末端を含む T A L E N、これはそれぞれ、Q または G のいずれかに変異した同一アミノ酸 (R 1 5 4) を含むが、これらはほぼ同量の切断を示した。同様に、Q または G 置換が位置 K 1 5 0 および R 1 5 4 で導入されている点でのみ異なる、M 6 および M 8 の N 末端を含む T A L E N、および、Q または G 置換が位置 K 1 5 0、K 1 5 3、および R 1 5 4 で導入されている点でのみ異なる、M 9 および M 1 0 の N 末端を含む T A L E N は、類似の切断活性を示した。

【 0 1 7 1 】

例 3

本明細書において提供される操作された T A L E N の、クローニングおよび発現のためのプラスミドを生成した。プラスミドのマップを、図 1 9 に示す。プラスミドは、例えば本明細書で提供される操作されたドメインなどの N 末端および C 末端ドメイン、および T A L E 反復の、モジュラークローニングを可能にし、所望の操作された T A L E N をコードする組換え核酸を生成する。プラスミドはまたアミノ酸タグをコードし、該タグは例えば、N 末端 F L A G タグおよび C 末端 V 5 タグであり、これらは必要に応じてコードされた T A L E N の精製または検出のために利用することができる。これらのタグの使用は任意であり、当業者は、タグ付けされた T A L E N タンパク質をコードするために、T A L E N をコードする配列は、タグをコードする配列とインフレームでクローニングされなければならないことを理解するであろう。

【 0 1 7 2 】

図 1 9 に示すようなクローニングベクターの例示的な配列を、以下に提供する。当業者は、以下の配列は例示的な態様の説明であり、本開示を限定するものではないことを理解するであろう。

10

20

【数 16 - 1】

>pExpCCR5A-L18_(63aa)

GACGGATCGGGGAGATCTCCCGATCCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCC
AGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAG
GCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGAT
ATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATA
TGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGT
CAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGT
AAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAAT
GGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCA
TCGCTATTACCATGGTGATGCGGTTTTTGGCAGTACATCAATGGGCGTGATAGCGGTTTGA CTACGCGGGATTTC
CAAGTCTCCACCCCAATTGACGTCAATGGGAGTTTGT TTTGGCACCAAAATCAACGGGACTTTC AAAATGTCGTA
ACA ACTCCGCCCCATTGACGCAAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGC
TAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGC
ACCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATG
GCCCCAAGAAGAAGAGGAAGGTGGGCATTACCGCGGGGTACCTATGGTGGACTTGAGGACACTCGGTTATTCTG
CAACAGCAACAGGAGAAAAATCAAGCCTAAGGTCAGGAGCACCGTCGCGCAACACCACGAGGCGCTTGTGGGGCAT
GGCTTCACTCATGCGCATATTGTGCGCTTTCACAGCACCCCTGCGGCGCTTGGGACGGTGGCTGTCAAATACCAA
GATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACAGTGGTCGGGAGCGCGA
GCAC TTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACACCGGGCAGCTGCTG
AAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGCGCTCACCGGGGCCCCC
TTGAACCTGACCCAGACCAGGTAGTCGCAATCGCGTCAAACGGAGGGGAAAGCAAGCCCTGGAAACCGTGCAA
AGGTTGTGCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCAITCCACGACGGT
GGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCACGGGCTGACTCCCGATCAA
GTTGTAGCGATTGCGTCGAACATTGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTGTGT

10

20

【数 1 6 - 2】

CAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCGAATGGCGGCGGTAAGCAGGCGCTGGAAACA
GTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCAGACCAGGTAGTCGCAATCGCGTCAAAC
GGAGGGGGAAAGCAAGCCCTGGAAACCGTGCAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCG
GAGCAAGTCGTGGCCATTGCAAGCAACATCGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCAGTT
CTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGTCGCATGACGGAGGGAAACAAGCATTG
GAGACTGTCCAACGGCTCCTTCCCGTGTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCC
TCCAATATTGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTG
ACCCAGACCAGGTAGTCGCAATCGCGTCACATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAGGTTGTTG
CCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCCACGACGGTGGCAAACAG
GCTCTTGAGACGGTTCAGAGACTTCTCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCG
ATTGCGTCCAACGGTGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTGTGTCAAGCCCAC
GGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGC
CTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCAGACCAGGTAGTCGCAATCGCGTCACATGACGGGGGA
AAGCAAGCCCTGGAAACCGTGCAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTC
GTGGCCATTGCAAGCAACATCGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCAGTTCTCTGTCAA
GCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAATAACAATGGAGGGAAACAAGCATTGGAGACTGTC
CAACGGCTCCTTCCCGTGTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAGCCATGAT
GGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACACCCGAA
CAGGTGGTCGCCATTGCTTCTAATGGGGGAGGACGGCCAGCCTTGGAGTCCATCGTAGCCCAATTGTCCAGGCCC
GATCCCGCGTTGGCTGCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGACCCGCGCTCGAT
GCAGTCAAAAAGGGTCTGCCTCATGCTCCCGCATTGATCAAAAAGAACCAACCGGCGGATTCCCGAGAGAACTTCC
CATCGAGTCGCGGGATCCCACTAGTCAAAAGTGAAGTGGAGGAGAAAGAAATCTGAAGTTCGTACATAAATTGAAA
TATGTGCCTCATGAATATATTGAATTAATTGAAATGGCAGAAATCCACTCAGGATAGAATTCTTGAAATGAAG
GTAATGGAATTTTTTATGAAAGTTTATGGATATAGAGGTAAACATTTGGGTGGATCAAGGAAACCGGACGGAGCA
ATTTATACTGTGCGAICTCCTATTGATTACGGTGTGATCGTGGATACTAAAGCTTATAGCGGAGGTTATAATCTG
CCAATTGGCCAAGCAGATGAAATGGAGCGATATGTCGAAGAAATCAAACACGAAACAAACATATCAACCCTAAT
GAATGGTGGAAAGTCTATCCATCTTCTGTAACGGAATTTAAGTTTTTATTTGTGAGTGGTCACTTTAAAGGAAAC
TACAAAGCTCAGCTTACACGATTAAATCATATCACTAATTGTAATGGAGCTGTTCTTAGTGTAGAAGAGCTTTTA
ATTGGTGGAGAAATGATTAAAGCCGGCACATTAACCTTAGAGGAAGTGGAGACGGAATTTAATAACGGCGAGATA
AACTTTTAAGGGCCCTTCGAAGGTAAGCCTATCCCTAACCCCTCTCCTCGGTCTCGATTCTACGCGTACCGGTCT
CATCACCATCACCATTGAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTT
TGCCCCCTCCCCCGTGCCTTCTTGACCCGGAAGGTGCCACTCCCACTGTCTTTCTTAATAAAATGAGGAAATT
GCATCGCATTGTCTGAGTAGGTGTCTATTCTATTTCTGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGG
GAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCT
AGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCT
ACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTGCTTTCTTCCCTTCTCTCGCCACGTTCCGCGGCTTTCCC
CGTCAAGCTCTAAATCGGGGCATCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAACTT
GATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACG
TTCTTTAATAGTGGACTCTTGTTCCAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAA
GGGATTTTGGGGATTTTCGGCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGT
GGAATGTGTGTGAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGGCAGGCAGAGTATGCAAGCAATGCACTC

10

20

30

40

【数 16 - 3】

AATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAAT
TAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCCATCCCGCCCTAACTCCCGCCAGTTCGCCCATTTCTCCG
CCCCATGGCTGACTAATTTTTTTTATTTATGTCAGAGGCCGAGGCCGCTCTGCCCTCTGAGCTATTCCAGAAGTAG
TGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGAT
CAGCACGIGTTGACAATTAATCATCGGCATAGTATATCGGCATAGTATAATACGACAAGGTGAGGAACTAAACCA
TGCCCAAGCCTTTGTCTCAAGAAGAATCCACCCTCATTGAAAGAGCAACGGCTACAATCAACAGCATCCCCATCT
CTGAAGACTACAGCGTCGCCAGCGCAGCTCTCTCTAGCGACGGCCGCATCTTCACTGGTGTCAATGTATATCAT
TTACTGGGGGACCTTGTGCAGAACTCGTGGTGTGGGCACTGCTGCTGCTGCGGCAGCTGGCAACCTGACTTGT
TCGTCGCGATCGGAAATGAGAACAGGGGCATCTTGAGCCCTGCGGACGGTGTGACAGGTGCTTCTCGATCTGC
ATCTGGGATCAAGCGATAGTGAAGGACAGTGATGGACAGCCGACGGCAGTTGGGATTCTGTAATTGCTGCCCT
CTGGTTATGTGTGGGAGGGCTAAGCACTTCGTGGCCGAGGAGCAGGACTGACACGTGCTACGAGATTTTCGATTCC
ACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGG
GATCTCATGCTGGAGTTCTTCGCCACCCCAACTTGTATTGTCAGCTTATAATGGTTACAAATAAAGCAATAGC
ATCACAATTTTACAAATAAAGCATTTTTTTTCACTGCATTTCTAGTTGTGTTTGTCCAACTCATCAATGTATCT
TATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCTGTGTGAAAT
TGTATCCGCTCACAATTCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTG
AGCTAACTCACATTAATTGCGTTGCGCTCACTGCCGCTTTCCAGTCGGGAAACCTGTGTCGTCAGCTGCATTAA
TGAATCGGCCAACGCGCGGGGAGAGCGGTTTTCGCTATTGGGCGCTCTTCCGCTTCTCTGCTCACTGACTCGCTG
CGCTCGGTCGTTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGG
GATAACGCAGGAAAGACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCTTGCTGGCG
TTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACA
GGACTATAAAGATACAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCCTCTCTGTTCCGACCCTGCCGCTTACC
GGATACCTGTCGCCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTG
GTGTAGGTCGTTGCTCCAAGCTGGGCTGTGTGACGAACCCCCGTTTCAGCCGACCGCTGCGCCTTATCCGGT
AACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGC
AGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCTAACTACGGCTACACTAGAAGGACAGTA
TTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGAAAAAGAGTTGGTAGCTCTTGATCCGGCAACAAACC
ACCGCTGGTAGCGGTGGTTTTTTTGTGTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCT
TTGATCTTTTCTACGGGCTCTGACGCTCAGTGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCA
AAAAGGATCTTACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAACT
TGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTTATCCATAGTT
GCCGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCACTGCCCCAGTGTGCAATGATACCG
CGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGT
CCTGCAACTTTATCCGCTCCATCCAGTCTATTAATTGTGCGGGAAGCTAGAGTAAGTAGTTGCCAGTTAAT
AGTTTGGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGGTATGGCTTCATTACG
TCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCT
CCGATCGTTGTGAGAAGTAAGTTGGCCGAGTGTTATCACTCATGTTATGGCAGCACTGCATAATTCTCTTACT
GTATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGG
CGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATC
ATTGAAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTTCGATGTAACCCACT
CGTGCACCCAACGTATCTTCAGCATCTTTTACTTTTACCAGCGTTTCTGGGTGAGCAAAAAACGGAAGGCAAAAT
GCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCTTTTTTCAATATTATTGAAGC
ATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAACAAATAGGGGTTCCG
CGCACATTTCCCCGAAAAGTGCCACCTGACGTC (配列番号 42)

10

20

30

40

【表 2 - 1】

1. Moscou, M.J. & Bogdanove, A.J. A simple cipher governs DNA recognition by TAL effectors. *Science* 326, 1501 (2009).
2. Boch, J. *et al.* Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326, 1509-1512 (2009).
3. Doyon, Y. *et al.* Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat Methods* 8, 74-79 (2011).
4. Cade, L. *et al.* Highly efficient generation of heritable zebrafish gene mutations using homo- and heterodimeric TALENs. *Nucleic Acids Res* 40, 8001-8010 (2012).
5. Miller, J.C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* 29, 143-148 (2011).
6. Bedell, V.M. *et al.* In vivo genome editing using a high-efficiency TALEN system. *Nature* 491, 114-118 (2012).
7. Hockemeyer, D. *et al.* Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol* 29, 731-734 (2011).
8. Cermak, T. *et al.* Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* 39, e82 (2011).
9. Tesson, L. *et al.* Knockout rats generated by embryo microinjection of TALENs. *Nat Biotechnol* 29, 695-696 (2011).
10. Moore, F.E. *et al.* Improved somatic mutagenesis in zebrafish using transcription activator-like effector nucleases (TALENs). *PLoS One* 7, e37877 (2012).
11. Wood, A.J. *et al.* Targeted genome editing across species using ZFNs and TALENs. *Science* 333, 307 (2011).
12. Reyon, D. *et al.* FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol* 30, 460-465 (2012).
13. Mussolino, C. *et al.* A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res* 39, 9283-9293 (2011).
14. Pattanayak, V., Ramirez, C.L., Joung, J.K. & Liu, D.R. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat Methods* 8, 765-770 (2011).

【表 2 - 2】

15. Li, T. *et al.* Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res* 39, 6315-6325 (2011).
16. Ding, Q. *et al.* A TALEN Genome-Editing System for Generating Human Stem Cell-Based Disease Models. *Cell Stem Cell* (2012).
17. Lei, Y. *et al.* Efficient targeted gene disruption in *Xenopus* embryos using engineered transcription activator-like effector nucleases (TALENs). *Proc Natl Acad Sci U S A* 109, 17484-17489 (2012). 10
18. Kim, Y. *et al.* A library of TAL effector nucleases spanning the human genome. *Nat Biotechnol* 31, 251-258 (2013).
19. Dahlem, T.J. *et al.* Simple methods for generating and detecting locus-specific mutations induced with TALENs in the zebrafish genome. *PLoS Genet* 8, e1002861 (2012).
20. Osborn, M.J. *et al.* TALEN-based Gene Correction for Epidermolysis Bullosa. *Molecular Therapy* (2013). 20
21. Maul, G.G. & Deaven, L. Quantitative determination of nuclear pore complexes in cycling cells with differing DNA content. *J Cell Biol* 73, 748-760 (1977).
22. Huang, B. *et al.* Counting low-copy number proteins in a single cell. *Science* 315, 81-84 (2007).
23. Beck, M. *et al.* The quantitative proteome of a human cell line. *Mol Syst Biol* 7, 549 (2011).
24. Meckler, J.F. *et al.* Quantitative analysis of TALE-DNA interactions suggests polarity effects. *Nucleic Acids Res* (2013). 30
25. Christian, M.L. *et al.* Targeting G with TAL effectors: a comparison of activities of TALENs constructed with NN and NK repeat variable di-residues. *PLoS One* 7, e45383 (2012).
26. Sander, J.D. *et al.* Abstraction of zinc finger nuclease cleavage profiles reveals an expanded landscape of off-target mutations. Submitted (2013).
27. Witten, I.H. & Frank, E. *Data mining: practical machine learning tools and techniques*, Edn. 2nd. (Morgan Kaufman, San Francisco; 2005). 40
28. Kim, Y., Kweon, J. & Kim, J.S. TALENs and ZFNs are associated with different mutation signatures. *Nat Methods* 10, 185 (2013).
29. McNaughton, B.R., Cronican, J.J., Thompson, D.B. & Liu, D.R. Mammalian cell penetration, siRNA transfection, and DNA transfection by supercharged proteins. *Proc Natl Acad Sci U S A* 106, 6111-6116 (2009).

【表 2 - 3】

30. Sun, N., Liang, J., Abil, Z. & Zhao, H. Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease. *Mol Biosyst* 8, 1255-1263 (2012).
31. Cong, L., Zhou, R., Kuo, Y.C., Cunniff, M. & Zhang, F. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat Commun* 3, 968 (2012).
32. Gabriel, R. *et al.* An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat Biotechnol* 29, 816-823 (2011). 10
33. Gao, H., Wu, X., Chai, J. & Han, Z. Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. *Cell Res* 22, 1716-1720 (2012).
34. Li, T. *et al.* Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res* 39, 6315-6325 (2011).
35. Miller, J.C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* 29, 143-148 (2011). 20
36. Mahfouz, M.M. *et al.* De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proc Natl Acad Sci U S A* 108, 2623-2628 (2011).
37. Pattanayak, V., Ramirez, C.L., Joung, J.K. & Liu, D.R. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat Methods* 8, 765-770 (2011).
38. Sander, J.D. *et al.* Abstraction of zinc finger nuclease cleavage profiles reveals an expanded landscape of off-target mutations. Submitted (2013). 30
39. Yan, T. *et al.* PatMatch: a program for finding patterns in peptide and nucleotide sequences. *Nucleic Acids Res* 33, W262-266 (2005).
40. Larkin, M.A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947-2948 (2007).

【 0 1 7 3 】

本明細書において、例えば背景、要約、詳細な説明、例、および／または参考文献のセクションで言及された全ての刊行物、特許、特許出願、刊行物、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、その全体が、個々の刊行物、特許、特許出願、刊行物、およびデータベースエントリーが具体的かつ個別に参照により本明細書に組み込まれたかのようにして、本明細書に参照により組み込まれる。矛盾する場合、本明細書中の任意の定義を含む本出願が支配する。

【 0 1 7 4 】

均等物および範囲

当業者は、日常的な実験を超えるものを使用することなく、本明細書に記載の本発明の特定の態様に対する多くの均等物を認識し、または確認することができるであろう。本発明の範囲は上記の説明に限定されることを意図せず、むしろ添付の特許請求の範囲に記載されている通りである。

【 0 1 7 5 】

特許請求の範囲において、「a」、「an」、および「the」などの冠詞は、逆が示されるかまたは文脈から明らかでない限り、1または1より多くを意味することができる。群の1または2以上のメンバーの間に「または」を含む特許請求の範囲または記載は、逆が示されるかまたは文脈から明らかでない限り、群のメンバーの1つ、1つより多く、または全てが、所与の製品またはプロセスに存在するか、用いられるか、またはその他に関連している場合に、満たされていると考えられる。本発明は、群の正確に1つのメンバーが、所与の製品またはプロセスに存在するか、用いられるか、またはその他に関連している態様も包含する。本発明はまた、群のメンバーの1つより多く、または全てが、所与の製品またはプロセスに存在するか、用いられるか、またはその他に関連している態様も包含する。

10

【0176】

さらに、本発明は、すべての変更、組み合わせ、および、1または2以上の特許請求の範囲からのまたは説明の関連する部分からの1または2以上の限定、要素、節、記述用語などが、別の特許請求の範囲に導入されるところの置換を、包含することが理解されるべきである。例えば、別の請求項に従属する任意の請求項は、同じ基本請求項に従属する任意の他の請求項に見出される1または2以上の限定を含むように、修飾することができる。さらに、請求項が組成物を述べる場合、本明細書に開示される目的のいずれかのために組成物を使用する方法、および、本明細書に開示されるかまたは当該技術分野において知られている他の方法のいずれかによって組成物を作製する方法も、別の記載がない限り、または矛盾もしくは不一致生じるであろうことが当業者に明らかでない限り、含まれることが理解されるべきである。

20

【0177】

要素がリストとして、例えばマーカッシュグループ形式で表される場合、要素の各サブグループも開示されており、任意の要素（単数または複数）がグループから削除され得ることを理解すべきである。また、用語「含む」は、オープンであることが意図され、追加の要素または工程の包含を許容することに留意されたい。一般に、本発明、または本発明の側面が、特定の要素、特徴、工程等を含むと言及される場合、本発明の特定の態様または本発明の側面は、かかる要素、特徴、工程等からなるか、または本質的にこれらからなることが、理解されるべきである。単純化のために、かかる態様は、本明細書においてこのような言葉で具体的に記載されていない。したがって、1または2以上の要素、特徴、工程等を含む本発明の各態様について、本発明はまた、これらの要素、特徴、工程等からなるか、本質的にこれらからなる態様も提供する。

30

【0178】

範囲が与えられている場合、終点も含まれる。さらに、他に指示がない限り、または文脈および/または当業者の理解から別が明らかでない限り、範囲として表される値は、本発明の異なる態様において記載された範囲内の任意の特定の値を、文脈が明確に別を指示しない限り、範囲の下限の単位の10分の1までの値でとることができることが理解されるべきである。また、他に指示がない限り、または文脈および/または当業者の理解から別が明らかでない限り、範囲で表された値は、所与の範囲内の任意の下位範囲をとることができ、ここで下位範囲の終点は、範囲の下限の単位の10分の1までと同程度の精度で表される。

40

【0179】

さらに、本発明の任意の特定の態様は、任意の1または2以上の特許請求の範囲から明示的に除外できることが理解されるべきである。範囲が与えられる場合、範囲内の任意の値を、任意の1または2以上の特許請求の範囲から明示的に除外してよい。本発明の組成物および/または方法の任意の態様、要素、特徴、用途、または側面は、任意の1または2以上の特許請求の範囲から除外することができる。簡潔にするため、1または2以上の要素、特徴、目的、または側面が除外される態様のすべてが、本明細書に明示的に記載されているわけではない。

【0180】

50

【表 3 - 1】

A

選択名	標的部位	左+右 半部位	部位長	N末端 ドメイン	C末端 ドメイン	FokI ドメイン	TALEN conc. (nM)	
CCR5A 32 nM 標準	CCR5A	L18+R18	36	標準	標準	EL/KK	32	
CCR5A 16 nM 標準 (or CCR5A 32 標準)	CCR5A	L18+R18	36	標準	標準	EL/KK	16	
CCR5A 8 nM 標準	CCR5A	L18+R18	36	標準	標準	EL/KK	8	10
CCR5A 4 nM 標準	CCR5A	L18+R18	36	標準	標準	EL/KK	4	
CCR5A Q3	CCR5A	L18+R18	36	標準	Q3	EL/KK	16	
CCR5A 32 nM Q7	CCR5A	L18+R18	36	標準	Q7	EL/KK	32	
CCR5A 16 nM Q7 (or CCR5A Q7)	CCR5A	L18+R18	36	標準	Q7	EL/KK	16	
CCR5A 8 nM Q7	CCR5A	L18+R18	36	標準	Q7	EL/KK	8	
CCR5A 4 nM Q7	CCR5A	L18+R18	36	標準	Q7	EL/KK	4	
CCR5A 28-aa	CCR5A	L18+R18	36	標準	28-aa	EL/KK	16	
CCR5A N1	CCR5A	L18+R18	36	N1	標準	EL/KK	16	
CCR5A N2	CCR5A	L18+R18	36	N2	標準	EL/KK	16	
CCR5A N3	CCR5A	L18+R18	36	N3	標準	EL/KK	16	
CCR5A 標準 ELD/KKR	CCR5A	L18+R18	36	標準	標準	ELD/KKR	16	20
CCR5A Q3 ELD/KKR	CCR5A	L18+R18	36	標準	Q3	ELD/KKR	16	
CCR5A Q7 ELD/KKR	CCR5A	L18+R18	36	標準	Q7	ELD/KKR	16	
CCR5A N2 ELD/KKR	CCR5A	L18+R18	36	N2	標準	ELD/KKR	16	

【 0 1 8 1 】

【表 3 - 2】

B

選択名	標的部位	左+右 半部位	部位長	N末端 ドメイン	C末端 ドメイン	FokI ドメイン	TALEN conc. (nM)	
ATM 32 nM 標準	ATM	L18+R18	36	標準	標準	EL/KK	24	
ATM 16 nM 標準 (or ATM 標準)	ATM	L18+R18	36	標準	標準	EL/KK	12	30
ATM 8 nM 標準	ATM	L18+R18	36	標準	標準	EL/KK	6	
ATM 4 nM 標準	ATM	L18+R18	36	標準	標準	EL/KK	3	
ATM Q3	ATM	L18+R18	36	標準	Q3	EL/KK	12	
ATM 32 nM Q7	ATM	L18+R18	36	標準	Q7	EL/KK	24	
ATM 16 nM Q7 (or ATM Q7)	ATM	L18+R18	36	標準	Q7	EL/KK	12	
ATM 8 nM Q7	ATM	L18+R18	36	標準	Q7	EL/KK	6	
ATM 4 nM Q7	ATM	L18+R18	36	標準	Q7	EL/KK	3	
ATM 28-aa	ATM	L18+R18	36	標準	28aa	EL/KK	12	
ATM N1	ATM	L18+R18	36	N1	標準	EL/KK	12	
ATM N2	ATM	L18+R18	36	N2	標準	EL/KK	12	
ATM N3	ATM	L18+R18	36	N3	標準	EL/KK	12	40
ATM 標準 ELD/KKR	ATM	L18+R18	36	標準	標準	ELD/KKR	12	
ATM Q3 ELD/KKR	ATM	L18+R18	36	標準	Q3	ELD/KKR	12	
ATM Q7 ELD/KKR	ATM	L18+R18	36	標準	Q7	ELD/KKR	12	
ATM N2 ELD/KKR	ATM	L18+R18	36	N2	標準	ELD/KKR	12	

【 0 1 8 2 】

【表 3 - 3】

C

選択名	標的部位	左+右 半部位	部位長	N末端 ドメイン	C末端 ドメイン	FokI ドメイン	TALEN conc. (nM)
L16+R16 CCR5B	CCR5 B	L16+R16	32	標準	標準	EL/KK	10
L16+R13 CCR5B	CCR5 B	L16+R13	29	標準	標準	EL/KK	10
L16+R10 CCR5B	CCR5 B	L16+R10	26	標準	標準	EL/KK	10
L13+R16 CCR5B	CCR5 B	L13+R16	29	標準	標準	EL/KK	10
L13+R13 CCR5B	CCR5 B	L13+R13	26	標準	標準	EL/KK	10
L13+R10 CCR5B	CCR5 B	L13+R10	23	標準	標準	EL/KK	10
L10+R16 CCR5B	CCR5 B	L10+R16	26	標準	標準	EL/KK	10
L10+R13 CCR5B	CCR5 B	L10+R13	23	標準	標準	EL/KK	10
L10+R10 CCR5B	CCR5 B	L10+R10	20	標準	標準	EL/KK	10

10

表 2 . 選択に使用した T A L E N コンストラクトおよび濃度。CCR5 A 標的配列 (A) 、 A T M 標的配列 (B) および C C R 5 B 標的配列 (C) を標的とする T A L E N を使用した各選択について、選択名、標的 DNA 部位、T A L E N N 末端ドメイン、T A L E N C 末端ドメイン、T A L E N F o k I ドメイン、および T A L E N 濃度 (conc.) が示されている。

20

【 0 1 8 3 】

【表 4 - 1】

A

選択名	Seq. 数	平均 mut.	Stdev mut.	Mut/bp	P値 vs. ライブラリー	P値 vs. 他のTALEN vs. CCR5A 標準 vs. CCR5A Q3
CCR5A 32 nM 標準	53883	4.327	1.483	0.120	3.3E-10	ELD/KKR = 0.260
CCR5A 16 nM 標準	28940	4.061	1.438	0.113	5.4E-10	ELD/KKR = 0.028
CCR5A 8 nM 標準	29568	3.751	1.394	0.104	3.3E-10	
CCR5A 4 nM 標準	34355	3.347	1.355	0.093	1.5E-10	
CCR5A Q3	51694	3.841	1.380	0.107	1.7E-10	
CCR5A 32 nM Q7	48473	2.718	1.197	0.076	4.4E-11	
CCR5A 16 nM Q7	56593	2.559	1.154	0.071	3.1E-11	
CCR5A 8 nM Q7	43895	2.303	1.157	0.064	3.0E-11	
CCR5A 4 nM Q7	43737	2.018	1.234	0.056	2.1E-11	
CCR5A 28-aa	47395	2.614	1.203	0.073	4.0E-11	
CCR5A N1	64257	3.721	1.379	0.103	1.1E-10	vs. CCR5A 8 nM 標準 = 0.039
CCR5A N2	45467	3.148	1.306	0.087	8.2E-11	
CCR5A N3	24064	2.474	1.493	0.069	8.1E-11	
CCR5A 標準 ELD/KKR	46998	4.336	1.491	0.120	4.0E-10	
CCR5A Q3 ELD/KKR	56978	4.098	1.415	0.114	2.2E-10	
CCR5A Q7 ELD/KKR	54903	3.234	1.330	0.090	7.3E-11	
CCR5A N2 ELD/KKR	79632	3.286	1.341	0.091	5.2E-11	

30

40

【 0 1 8 4 】

【表 4 - 2】

B

選択名	Seq. 数	平均 mut.	Stdev mut.	Mut./bp	P値 vs. ライブラリー	P値 vs. 他のTALEN vs. ATM 標準 ELD/KKR =0.012
ATM 24 nM 標準	89571	3.262	1.360	0.091	6.54E-11	
ATM 12 nM 標準 (or ATM 標準)	96703	3.181	1.307	0.088	5.36E-11	
ATM 6 nM 標準	78852	2.736	1.259	0.076	3.63E-11	
ATM 3 nM 標準	82527	2.552	1.258	0.071	2.71E-11	
ATM Q3	96582	2.551	1.248	0.071	2.31E-11	vs. ATM 4 nM 標準 =0.222
ATM 24 nM Q7	10166	1.885	2.125	0.052	2.06E-10	
ATM 12 nM Q7 (or ATM Q7)	4662	1.626	2.083	0.045	5.31E-10	
ATM 6 nM Q7	1290	1.700	2.376	0.047	7.16E-09	vs. ATM 16 nM Q7 =0.035
ATM N1	84402	2.627	1.318	0.073	2.92E-11	
ATM N2	62470	2.317	1.516	0.064	2.69E-11	
ATM N3	1605	2.720	2.363	0.076	2.69E-08	
ATM 標準 ELD/KKR	107970	3.279	1.329	0.091	5.48E-11	
ATM Q3 ELD/KKR	104099	2.846	1.244	0.079	3.15E-11	
ATM Q7 ELD/KKR	21108	1.444	1.56	0.040	3.02E-11	
ATM N2 ELD/KKR	70185	2.45	1.444	0.06805	2.82E-11	

10

20

【 0 1 8 5 】

【表 4 - 3】

C

選択名	Seq. 数	平均 mut.	Stdev mut.	Mut./bp	P値 vs. ライブラリー	P値 vs. 他のTALEN
L16+R16 CCR5B	34904	2.134	1.168	0.067	4.7E-11	
L16+R13 CCR5B	38229	1.581	1.142	0.055	2.7E-11	
L16+R10 CCR5B	37801	1.187	0.949	0.046	2.2E-11	
L13+R16 CCR5B	46608	1.505	1.090	0.052	1.7E-11	
L13+R13 CCR5B	53973	0.996	1.025	0.038	8.8E-12	
L13+R10 CCR5B	60550	0.737	0.884	0.032	7.4E-12	
L10+R16 CCR5B	36927	1.387	0.971	0.053	3.0E-11	
L10+R13 CCR5B	58170	0.839	0.882	0.036	9.1E-12	
L10+R10 CCR5B	57331	0.646	0.779	0.032	1.0E-11	

30

表 3 . T A L E N 消化により選択された配列の統計値。統計値は、C C R 5 A 標的配列 (A)、A T M 標的配列 (B) および C C R 5 B 標的配列 (C) に対するそれぞれの T A L E N 選択について示される。Seq. 数：ハイスループットシーケンシングされコンピューターによりフィルタリングされた選択配列の総カウント数。平均mut.：選択された配列内の平均変異数。Stdev. mut.：選択された配列における変異の標準偏差。Mut./bp：標的部位長 (b p) について正規化された平均変異数。P 値vs.ライブラリー：T A L E N 選択配列分布の、対応する選択前ライブラリー配列分布 (表 5) に対する P 値は、以前の報告のようにして決定した⁵。P 値vs.他の T A L E N：すべての T A L E N 消化間のすべてのペアワイズ比較を算出し、0 . 0 1 ~ 0 . 5 の間の P 値を示す。3 n M の Q 7 A T M および 2 8 - a a A T M 選択について、これらの選択が実施されたにも関わらず、解釈に十分な配列が得られなかったことに注意されたい。

40

50

【 0 1 8 6 】

【 表 5 】

ライブラリー名	標的部位	左+右 半部位	部位長	Seq. 数	平均 mut.	Stdev mut.	Mut./bp
CCR5A ライブラリー	CCR5A	L18+R18	36	158643	7.539	2.475	0.209
ATM ライブラリー	ATM	L18+R18	36	212661	6.820	2.327	0.189
CCR5B ライブラリー	CCR5B	L16+R16	32	280223	6.500	2.441	0.203
CCR5B ライブラリー	CCR5B	L16+R13	29	280223	5.914	2.336	0.204
CCR5B ライブラリー	CCR5B	L16+R10	26	280223	5.273	2.218	0.203
CCR5B ライブラリー	CCR5B	L13+R16	29	280223	5.969	2.340	0.206
CCR5B ライブラリー	CCR5B	L13+R13	26	280223	5.383	2.230	0.207
CCR5B ライブラリー	CCR5B	L13+R10	23	280223	4.742	2.106	0.206
CCR5B ライブラリー	CCR5B	L10+R16	26	280223	5.396	2.217	0.208
CCR5B ライブラリー	CCR5B	L10+R13	23	280223	4.810	2.100	0.209
CCR5B ライブラリー	CCR5B	L10+R10	20	280223	4.169	1.971	0.208

10

表 4 . 選択前ライブラリーからの配列の統計値。各選択前ライブラリーは、CCR5A 標的配列、ATM 標的配列およびCCR5B 標的配列の変異配列の分布を含む。Seq. 数：ハイスループットシーケンシングされコンピューターによりフィルタリングされた選択配列の総カウント数。平均mut.：配列の平均変異数。Stdev. mut.：配列の標準偏差。Mut./bp：標的部位長 (b p) について正規化した平均変異数。

20

【 0 1 8 7 】

【 表 6 - 1 】

A

選択	濃縮値								
	0 Mut.	1 Mut.	2 Mut.	3 Mut.	4 Mut.	5 Mut.	6 Mut.	7 Mut.	8 Mut.
CCR5A 32 nM 標準	9.879	9.191	8.335	6.149	4.205	2.269	1.005	0.325	0.085
CCR5A 16 nM 標準	12.182	13.200	10.322	7.195	4.442	2.127	0.748	0.216	0.052
CCR5A 8 nM 標準	19.673	17.935	13.731	8.505	4.512	1.756	0.531	0.116	0.028
CCR5A 4 nM 標準	36.737	29.407	19.224	9.958	4.047	1.242	0.302	0.058	0.014
CCR5A Q3	18.550	16.466	12.024	8.070	4.632	1.938	0.572	0.126	0.026
CCR5A 32 nM Q7	60.583	54.117	31.082	11.031	2.640	0.469	0.073	0.013	0.006
CCR5A 16 nM Q7	62.294	64.689	35.036	10.538	2.183	0.322	0.046	0.010	0.006
CCR5A 8 nM Q7	97.020	91.633	38.634	8.974	1.485	0.189	0.029	0.010	0.007
CCR5A 4 nM Q7	197.239	130.497	38.361	6.535	0.896	0.120	0.025	0.019	0.017
CCR5A 28-aa	70.441	62.213	33.481	10.488	2.317	0.402	0.064	0.012	0.006
CCR5A N1	19.038	18.052	13.858	8.788	4.546	1.697	0.499	0.115	0.025
CCR5A N2	41.715	35.752	22.638	10.424	3.777	0.989	0.194	0.038	0.007
CCR5A N3	173.897	88.392	31.503	8.770	1.853	0.350	0.089	0.036	0.027
CCR5A 標準 ELD/KKR	8.101	10.012	8.220	6.147	4.119	2.291	1.019	0.330	0.083
CCR5A Q3 ELD/KKR	14.664	12.975	9.409	6.819	4.544	2.235	0.797	0.198	0.041
CCR5A Q7 ELD/KKR	37.435	32.922	21.033	10.397	3.867	1.087	0.238	0.046	0.010
CCR5A N2 ELD/KKR	35.860	31.469	20.135	10.189	3.983	1.155	0.260	0.050	0.013

30

40

【 0 1 8 8 】

【表 6 - 2】

B

選択	濃縮値								
	0 Mut.	1 Mut.	2 Mut.	3 Mut.	4 Mut.	5 Mut.	6 Mut.	7 Mut.	8 Mut.
ATM 24 nM 標準	19.900	16.881	12.162	6.318	2.629	0.884	0.228	0.057	0.015
ATM 12 nM 標準	20.472	17.645	12.724	6.549	2.606	0.803	0.189	0.039	0.007
ATM 6 nM 標準	41.141	29.522	17.153	6.551	1.872	0.431	0.082	0.017	0.006
ATM 3 nM 標準	56.152	37.152	18.530	6.196	1.562	0.308	0.058	0.015	0.008
ATM Q3	50.403	36.687	19.031	6.245	1.513	0.294	0.057	0.016	0.010
ATM 24 nM Q7	353.148	90.350	13.475	1.531	0.186	0.128	0.116	0.118	0.103
ATM 12 nM Q7	513.385	89.962	11.310	0.860	0.190	0.093	0.115	0.092	0.111
ATM 6 nM Q7	644.427	82.074	7.650	0.677	0.170	0.205	0.163	0.164	0.071
ATM N1	57.218	35.388	17.808	6.124	1.644	0.383	0.076	0.023	0.011
ATM N2	119.240	53.618	18.977	4.742	0.992	0.233	0.076	0.044	0.037
ATM N3	201.158	55.468	15.244	3.187	0.764	0.307	0.154	0.173	0.287
ATM 標準 ELD/KKR	19.356	15.692	11.855	6.403	2.706	0.899	0.224	0.054	0.011
ATM Q3 ELD/KKR	32.816	25.151	16.172	6.727	2.095	0.506	0.095	0.018	0.004
ATM Q7 ELD/KKR	447.509	93.166	13.505	1.543	0.170	0.053	0.049	0.045	0.045
ATM N2 ELD/KKR	90.625	45.525	18.683	5.369	1.267	0.274	0.076	0.035	0.027

【 0 1 8 9 】

【表 6 - 3】

C

選択	濃縮値								
	0 Mut.	1 Mut.	2 Mut.	3 Mut.	4 Mut.	5 Mut.	6 Mut.	7 Mut.	8 Mut.
L16+R16 CCR5B	59.422	35.499	13.719	3.770	0.737	0.132	0.024	0.011	0.008
L16+R13 CCR5B	80.852	31.434	7.754	1.380	0.218	0.040	0.022	0.016	0.017
L16+R10 CCR5B	64.944	20.056	3.867	0.515	0.056	0.010	0.006	0.006	0.007
L13+R16 CCR5B	101.929	34.255	8.131	1.299	0.167	0.033	0.016	0.011	0.014
L13+R13 CCR5B	113.102	22.582	3.037	0.315	0.044	0.022	0.017	0.017	0.016
L13+R10 CCR5B	74.085	11.483	1.270	0.121	0.022	0.013	0.011	0.013	0.008
L10+R16 CCR5B	60.186	22.393	5.286	0.777	0.084	0.012	0.006	0.006	0.008
L10+R13 CCR5B	74.204	13.696	1.673	0.152	0.021	0.011	0.010	0.009	0.010
L10+R10 CCR5B	43.983	7.018	0.740	0.061	0.013	0.007	0.007	0.008	0.005

表 5 . 変異数の関数としての配列の濃縮値。CCR5A 標的配列 (A)、ATM 標的配列 (B) および CCR5B 標的配列 (C) に対するそれぞれの TALEN 選択について、濃縮値は、TALEN 消化からの選択後配列の部分的存在量 (fractional abundance) を、選択前配列の部分的存在量で割り算し、半部位における総変異数 (Mut.) の関数として算出した。

【 0 1 9 0 】

【表 7 - 1】

A

CCR5A 部位	スコア	Mut.	左半部位	スペーサー長	右半部位	遺伝子	
OnCCR5 A	0.008	0	TTCAATTACACCTGCAGCT	18	AGTATCAATTCTGGAAGA	CCR5	
OffC-1	0.747	9	TaCATcACAtaTGCAaaT	29	tGTATCAAtTTCTGGgAGA	ARL17A & LRR37A	
OffC-2	0.747	9	TaCATcACAtaTGCAaaT	29	tGTATCAAtTTCTGGgAGA	ARL17A & LRR37A	
OffC-3	0.747	9	TaCATcACAtaTGCAaaT	29	tGTATCAAtTTCTGGgAGA	ARL17A & LRR37A	10
OffC-4	0.747	11	TcCATaACACaTcttctCT	10	tGcATCAAtTcCTGGAAGA	ZSCAN5A	
OffC-5	0.804	11	TcCAaTACctCTGCcaCa	14	AGgAgCAAcTCTGGgAGA		
OffC-6	0.818	10	TTCAgTcCATcCTGaAaac	16	gGTATCAAtTTCTGGAgGA	KL	
OffC-7	0.834	14	TaCAaaACcCtTGCCaaa	27	taTATCAATTtgGGgAGA		
OffC-8	0.837	12	TcCAagACACCTGCttac	26	tcTATCAATTtgGGgAGA		
OffC-9	0.874	10	TTCAATaACAtCTtaAaaT	27	AaTAcCAAcTCTGGAtGA	ZEB2	
OffC-10	0.89	12	TcCAaaACAtCTGaAaaT	25	tGgATCAAAAttgGGAAGA		
OffC-11	0.896	12	TTCAgaACACaTGactac	21	tGTATCAgTTaTGGAtGA	GABPA	
OffC-12	0.904	13	TcCATaAtAtCTtCctCT	28	gGgATtAATTtgGGAAGGA		
OffC-13	0.905	11	TgCAaTATACCTGttGaT	16	ctcATCAATTCTGGgtGA		
OffC-14	0.906	12	TTCAATaACACtccacctT	16	gGTATCAAACTCTGGggGA	SYN3	20
OffC-15	0.906	12	TcCATgACACaaaagaCT	26	gGTATCtAtTcCTGGAAtA	SPOCK3	
OffC-16	0.906	9	TTCcTTcCACCaGtgtCc	28	AGcATCAATcCTGGAAGA		
OffC-17	0.907	10	TTaATaACAtCTcCAaCT	24	gGcAcCAAACTCTGGAtGA	ATP13A5	
OffC-18	0.909	13	TcCATcACcCCTcCctCc	10	gGTgcCAgcTCTGGAgGA	TBC1D7	
OffC-19	0.909	8	TTCAATTACTCCTcCtctCT	30	ctTATCACTTtTGGAAGA		
OffC-20	0.912	10	TgCATTTACACaTtatGtg	17	AGcAgCACTTCTGGAAGA		
OffC-21	0.913	11	TTCAaaACACaTaCatCT	28	AacAaCatTcCTGtAAGA	PRKAG2	
OffC-22	0.913	10	TcCATTTACcaCTGCAGaT	25	gacATCAgTTaTGGAtGA		
OffC-23	0.925	13	TTCcagACcCCTtCctCa	13	gacATCAAACTCTGGgAGA		
OffC-24	0.927	12	TTCcAAACACCCcGCttCc	26	taTATCctTTCTGGAAtA		
OffC-25	0.93	12	TgaAaTACACCTGCctaT	13	gGccTCAAggCTGGAtGA	IL15	30
OffC-26	0.93	12	TgCcaaACctCTGtcaCc	22	AGgATCAcTTCTGGAAGA		
OffC-27	0.931	12	TgCcaaACctCTGtcaCc	22	AGgATCAcTTCTGGAAGA		
OffC-28	0.931	8	TTtATTTACACtTcCAGaT	19	gaTATCctTTCTGGAAGA	ADIPOR2	
OffC-29	0.932	13	TaCAaaAaAcTtTctGag	27	tGTATCAATTtgGGgAGA	FBXL17	
OffC-30	0.932	11	TcCAaaACACCCaCAGac	19	gGTATagATTgTGGAAGA	ZNF365	
OffC-31	0.934	13	TTCAATTcCACaTcCccac	25	gtTATCAAcAtgGGAAGA	MYO18B	
OffC-32	0.934	11	TTCAaTAtgCCaaCAGCT	11	AGctTCAATctgGGAAGA		
OffC-33	0.934	12	TTCAaTACACtTGtctaT	12	tGTgTCAAtTTCTGGgttA		
OffC-34	0.935	11	TTCAacACACCTtCAaaa	12	tGTgTCAAtTaaTGGAAGA		
OffC-35	0.935	10	TTCAaaACAtCTGacatT	10	AaTAgAAATCTGGAAGA		
OffC-36	0.935	11	CTCcTaAtACCTGCaaaT	21	gaTATtAtTTCTGGAgGA		40

【 0 1 9 1 】

【表 7 - 2】

B

ATM 部位	スコア	Mut.	左半部位	スパーサー長	右半部位	遺伝子
OnATM	0.000	0	TGAATTGGGATGCTGTTT	18	TTTATTTTACTGCTTTTA	ATM
OHA-1	0.595	7	TGAATaGGAaataTaTTT	20	TTTATTTTACTGCTTTTA	
OHA-2	0.697	9	TGgATTcaGATaCTcTTT	10	TTTATTTTtttTaTtTTTA	
OHA-3	0.697	9	TGgATTcaGATaCTcTTT	10	TTTATTTTtttTaTtTTTA	
OHA-4	0.697	9	TGgATTcaGATaCTcTTT	10	TTTATTTTtttTaTtTTTA	
OHA-5	0.697	9	TGgATTcaGATaCTcTTT	10	TTTATTTTtttTaTtTTTA	10
OHA-6	0.697	9	TGgATTcaGATaCTcTTT	10	TTTATTTTtttTaTtTTTA	
OHA-7	0.697	9	TGgATTcaGATaCTcTTT	10	TTTATTTTtttTaTtTTTA	
OHA-8	0.7	8	TGcATaGGAATGCTaaTT	10	TTTATTTTACTaTtTaTa	MGAT4C
OHA-9	0.708	10	TGAATTaaaATcCTGcTT	19	gTTATaTgACTaTtTTTA	BRCA2
OHA-10	0.711	10	TccATTaaaATACTaTTT	18	TTTATTTTAtTaTtTTTA	CPNE4
OHA-11	0.715	10	TGAATTGaGagaagcaTT	16	TTTATTTTAtTaTtTTTA	
OHA-12	0.725	10	TGAAGTGGGATaCTGTTa	29	ggTATaTTAaaTtTTTA	
OHA-13	0.729	9	TGAATTatGAaGCTacTT	17	TTTATTgTAaTaTtTTTA	NAAALADL2
OHA-14	0.731	9	TGAATaaGGATGCTaTTa	25	TTTATTTAttTaTtTTTA	
OHA-15	0.744	10	TGAATgGGGAcCaGccca	29	TTTATTTTAtTaTtTTTA	
OHA-16	0.752	9	TaAATgGaaATGCTGTTC	24	aTTATTTTAtTGTtTTTt	20
OHA-17	0.761	9	gGAAaTGGGATaCTGagT	15	TTTATgTTACTaTtTcTa	
OHA-18	0.781	11	TGgATcGaagTGaTtaTT	23	TTTATTTTAtTaTtTTTA	CIDEC
OHA-19	0.792	11	TGAATTGaGATtCacagc	23	TTTATTTTtttTaTtTTTA	
OHA-20	0.803	8	TGAATTaGGAatCTGaTT	10	TTTATTTTAtTaTtTaTTA	THSD7B
OHA-21	0.807	12	TaAATTaaaATACTccag	23	aTTATTTTAAaTGTtTTTA	ARID1B
OHA-22	0.811	10	TGAATaGGAATatTcTTT	12	TTTATTTAttTaTtTTTA	
OHA-23	0.811	9	TagATTGaaATGCTGTTT	15	TTTtTaTTATaTtTTTA	KLHL4
OHA-24	0.816	10	TGAcTaGaaATGaTgaTT	25	TTTATTTTctTaTtTTTA	
OHA-25	0.817	12	TGAATTTtaaAaaaTGTcc	13	aTTATTTTAtTaTtTTTA	
OHA-26	0.817	12	TGAATTTtaaAaaaTGTcc	13	aTTATTTTAtTaTtTTTA	
OHA-27	0.817	10	TGgATccaGATaCTcTTT	10	TTTATTTTtttTaTtTTTA	30
OHA-28	0.819	7	TGgAgTGaGATcCTGTTT	21	TTTATTTTAtTGTtaTTA	
OHA-29	0.824	8	TGAACtTGGATGaTaTaT	24	TTTATTTgAtTaTCTTTA	
OHA-30	0.832	9	TGtATTGGGATaCcaTTT	26	TcTATTTTAtTaTtTTTt	
OHA-31	0.833	9	TcAATTGGGATGaTcaTa	23	TTTATTCtAtTtTtTTTA	
OHA-32	0.835	9	TGAAagGGAaAGtTGgaT	23	TTTATTTTACTaTtTTTA	
OHA-33	0.841	9	TGgtTTGGGATcCTGTgT	27	TTTATgTTttTaTtTTTA	PICHD2
OHA-34	0.841	9	TGAAaTGGGATGagcTTg	28	TTTATTTTAtTaTtTTaA	
OHA-35	0.844	10	TGAATTGGGATaCTGTag	29	cTTAAaTaAaTaTtTTTA	ST6GALNAC3
OHA-36	0.844	10	TGAATTGtGgTatTGccT	18	TTTATggtttTGTCTTTA	

表 6 . ヒトゲノム中の予測オフターゲット部位。(A) in vitro の C C R 5 A T A L E N 選択の出力に対して訓練された機械学習「classifier」アルゴリズムを使用し、10 ~ 30 塩基対のスパーサー長を可能にする、標的部位の 6 変異配列をスコア付けした。C C R 5 A T A L E N についてのベストスコアを有する得られた 36 の予測オフターゲット部位を、classifier スコア、変異数、左および右半部位配列 (オンターゲットからの変異を小文字で示す)、塩基対における半部位の間のスパーサー長、および、それが遺伝子内にある場合は予測されたオフターゲット部位が生じる遺伝子 (イントロンを含む) と共に示す。(B) A T M T A L E N について、(A) と同様。配列は配列番号 43 ~ XX に関連する。

【0192】

【表 8 - 1】

A

C末端ドメイン FokI ドメイン CCR5A 部位	TALENなし TALENなし	Q7 EL/KK	Q7 ELD/KKR	Q3 EL/KK	Q3 ELD/KKR	標準 EL/KK	標準 ELD/KKR	標準 Homo
OnC								
インデル	5	147	705	1430	3731	841	2004	3943
総数	23644	7192	12667	16843	15381	8546	7267	8422
修飾%	0.021%	2.044%	5.566%	8.490%	24.257%	9.841%	27.577%	46.818%
P値		1.3E-33	2.5E-160	<1.0E-200	<1.0E-200	5.9E-200	<1.0E-200	<1.0E-200
特異性								
OffC-1								
インデル	0	0	1	1	0	0	1	1
総数	51248	38975	79858	35491	77804	34227	87497	42498
修飾%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P値								
特異性								
OffC-2								
インデル	0	0	0	0	0	0	10	0
総数	124356	96280	157387	93337	159817	85603	163332	114663
修飾	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	0.006%	<0.006%
P値							1.6E-03	
特異性		>307	>835	>1274	>3639	>1476	>4137	>7023
OffC-3								
インデル	5	0	4	1	0	0	6	3
総数	99085	75958	130027	72919	131132	67192	136796	90039
修飾	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P値								
特異性								
OffC-4								
インデル	0	1	0	0	0	0	0	0
総数	45377	44674	52876	35133	53909	26034	42284	40452
修飾	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P値								
特異性								
OffC-5								
インデル	0	0	0	3	22	134	385	395
総数	27009	28172	26036	22432	25800	25273	17045	17977
修飾	<0.006%	<0.006%	<0.006%	0.013%	0.085%	0.527%	2.209%	2.261%
P値					2.7E-06	4.5E-31	4.9E-87	2.8E-89

【 0 1 9 3 】

【表 8 - 2】

特異性		>576	>1450	635	285	19	12	21
OffC-6								
インデル	0	0	0	0	0	0	0	0
総数	10766	12309	10886	9240	10558	10500	5943	6560
修飾	<0.009%	<0.008%	<0.009%	<0.011%	<0.009%	<0.010%	<0.017%	<0.015%
P値								
特異性								
OffC-7								
総数	0	0	0	0	0	0	0	0
修飾	15626	28825	22138	31742	19577	11902	33200	15400
P値	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.008%	<0.006%	<0.006%
特異性								
OffC-9								
インデル	0	0	0	1	0	0	0	0
総数	40603	39765	47974	51595	44002	34520	25211	30771
修飾	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P値								
特異性								
OffC-10								
インデル	0	0	0	0	0	0	0	0
総数	4142	9591	5187	1413	7975	4378	2215	3779
修飾	<0.024%	<0.010%	<0.019%	<0.071%	<0.013%	<0.023%	<0.045%	<0.026%
P値								
特異性								
OffC-11								
インデル	0	0	0	0	0	0	0	0
総数	71180	55455	65015	44847	70907	50967	65257	60191
修飾	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P値								
特異性								
OffC-12								
インデル	0	0	0	0	0	0	0	0
総数	3242	1784	30274	14006	4897	19830	9747	12910
修飾	<0.031%	<0.056%	<0.006%	<0.007%	<0.020%	<0.006%	<0.010%	<0.008%
P値								
特異性								
OffC-13								
インデル	0	0	0	0	0	0	0	0

【 0 1 9 4 】

【表 8 - 3】

総数	65518	52459	53413	38156	61600	47922	57211	78546
修飾	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P値								
特異性								

OffC-14

インデル	0	0	0	0	0	0	2	0
総数	34607	7217	26301	8339	29845	1081	9471	19026
修飾	<0.006%	<0.014%	<0.006%	<0.012%	<0.006%	<0.093%	0.021%	<0.006%
P値								
特異性								

10

OffC-15

インデル	0	0	0	0	0	0	16	2
総数	4989	4880	6026	9370	9156	7371	6967	4662
修飾	<0.020%	<0.020%	<0.017%	<0.011%	<0.011%	<0.014%	0.230%	0.043%
P値							6.3E-05	
特異性		>100	>335	>796	>2221	>725	120	1091

OffC-16

インデル	0	1	1	1	14	1	12	0
総数	36228	34728	34403	34866	44362	38384	38536	32636
修飾	<0.006%	<0.006%	<0.006%	<0.006%	0.032%	<0.006%	0.031%	<0.006%
P値					1.8E-04		5.3E-04	
特異性		>307	>835	>1274	769	>1476	886	>7023

20

OffC-17

インデル	0	0	0	0	0	0	0	0
総数	32112	23901	31273	33968	27437	29670	27133	31299
修飾	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P値								
特異性								

30

OffC-18

インデル	0	0	0	0	0	0	0	0
総数	9437	9661	13505	14900	13848	12720	6624	12804
修飾	<0.011%	<0.010%	<0.007%	<0.007%	<0.007%	<0.008%	<0.015%	<0.008%
P値								
特異性								

OffC-19

インデル	1	1	1	2	2	2	1	0
総数	22868	11479	22702	15258	20733	17449	14638	28478
修飾	<0.006%	0.009%	<0.006%	0.013%	0.010%	0.011%	0.007%	<0.006%
P値								

40

【 0 1 9 5 】

【表 8 - 4】

特異性

OffC-20

インデル	0	0	0	0	0	1	0	0
総数	23335	26164	30782	15261	20231	21184	14144	18972
修飾	<0.006%	<0.006%	<0.006%	<0.007%	<0.006%	<0.006%	<0.007%	<0.006%
P値								
特異性								

OffC-21

インデル	0	0	0	0	0	0	0	0
総数	34302	27573	31694	24451	25826	27192	18110	21161
修飾	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P値								
特異性								

OffC-22

インデル	1	0	0	0	0	0	0	0
総数	81037	86687	74274	79004	93477	92089	75359	104857
修飾	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P値								
特異性								

OffC-23

インデル	0	0	0	0	0	0	0	0
総数	18812	19337	23034	25603	25023	28615	17172	21033
修飾	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P値								
特異性								

OffC-24

インデル	0	0	1	0	0	0	0	1
総数	23538	21673	24594	27687	18343	29113	21709	26610
修飾	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P値								
特異性								

OffC-25

インデル	0	0	0	0	0	0	0	0
総数	28941	25326	25871	10641	21422	20171	18946	18711
修飾	<0.006%	<0.006%	<0.006%	<0.009%	<0.006%	<0.006%	<0.006%	<0.006%
P値								
特異性								

OffC-26

--	--	--	--	--	--	--	--	--

【 0 1 9 6 】

【表 8 - 5】

インデル	0	0	1	0	0	0	0	0
総数	71831	48494	62650	45801	60175	65137	28795	64632
修飾	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P値								
特異性								
OffC-27								
インデル	0	0	0	0	0	0	0	0
総数	12181	2423	11258	7188	5126	4003	2116	4603
修飾%	<0.008%	<0.041%	<0.009%	<0.014%	<0.020%	<0.025%	<0.047%	<0.022%
P値								
特異性								
OffC-28								
インデル	0	0	0	0	6	1	12	5
総数	10651	6410	16179	13980	13022	7232	7379	8998
修飾%	<0.009%	<0.016%	<0.006%	<0.007%	0.046%	0.014%	0.163%	0.056%
P値					1.4E-02		5.3E-04	
特異性		>131	>835	>1187	526	712	170	843
OffC-29								
インデル	0	0	0	0	0	0	0	0
総数	4262	3766	4228	6960	3234	1516	2466	1810
修飾%	<0.023%	<0.027%	<0.024%	<0.014%	<0.031%	<0.066%	<0.041%	<0.055%
P値								
特異性								
OffC-30								
インデル	0	0	0	0	0	0	0	0
総数	11840	12257	9617	34097	20507	5029	22248	6285
修飾%	<0.008%	<0.008%	<0.010%	<0.006%	<0.006%	<0.020%	<0.006%	<0.016%
P値								
特異性								
OffC-31								
インデル	0	0	0	0	0	0	0	0
総数	64522	67791	50085	50056	56241	48287	72230	100410
修飾%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P値								
特異性								
OffC-32								
インデル	0	0	0	0	0	0	0	0
総数	1944	6888	9330	3207	4591	6699	13607	19115
修飾%	<0.051%	<0.015%	<0.011%	<0.031%	<0.022%	<0.015%	<0.007%	<0.006%

【 0 1 9 7 】

【表 8 - 6】

P値

特異性

OffC-33

インデル	0	0	0	0	0	0	0	0
総数	34475	27039	18547	33467	15745	17075	4	18844
修飾%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<25.000%	<0.006%
P値								
特異性								

10

OffC-34

インデル	0	0	0	0	0	0	0	0
総数	9052	18858	13647	11796	6945	6114	4979	9072
修飾%	<0.011%	<0.006%	<0.007%	<0.008%	<0.014%	<0.016%	<0.020%	<0.011%
P値								
特異性								

OffC-35

インデル	0	0	0	0	0	0	0	0
総数	23839	22290	25133	24190	10	10459	22554	11897
修飾%	<0.006%	<0.006%	<0.006%	<0.006%	<10.000%	<0.010%	<0.006%	<0.008%
P値								
特異性								

20

OffC-36

インデル	1	0	0	1	2	1	19	5
総数	23412	24394	23427	24132	19723	28369	12461	18052
修飾	<0.006%	<0.006%	<0.006%	<0.006%	0.010%	<0.006%	0.152%	0.028%
P値							2.6E-05	
特異性		>307	>835	>1274	2392	>1476	181	1690

B

30

C末端ドメイン	TALENなし	Q7	Q7	Q3	Q3	標準	標準	標準
Forkドメイン	TALENなし	EL/KK	ELD/KKR	EL/KK	ELD/KKR	EL/KK	ELD/KKR	Homo
ATM 部位								

On-A

インデル	3	0	46	104	309	1289	410	909
総数	8888	1869	2520	1198	1808	19025	2533	5003
修飾	0.03%	0.00%	1.83%	8.68%	17.09%	6.78%	16.19%	18.17%
P値	0		2.2E-11	3.2E-26	4.9E-81	6.4E-276	4.5E-105	1.5E-228
特異性								

40

OffA-1

インデル	0	0	1	0	1	0	13	34
------	---	---	---	---	---	---	----	----

【 0 1 9 8 】

【表 8 - 7】

総数	52490	45383	34195	32325	47589	39704	50349	44056
修飾	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	0.026%	0.077%
P値							3.1E-04	5.5E-09
特異性		>0	>274	>1302	>2564	>1016	627	235

OffA-2

インデル	0	0	0	0	0	0	0	0
総数	8777	11846	11362	12273	20704	3776	5650	5025
修飾	<0.011%	<0.008%	<0.009%	<0.008%	<0.006%	<0.026%	<0.018%	<0.020%
P値								
特異性								

10

OffA-3

インデル	0	0	0	0	1	0	0	0
総数	47338	14352	21253	17777	26512	19483	43728	29469
修飾	<0.006%	<0.007%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P値								
特異性								

OffA-4

インデル	0	0	0	0	0	0	0	0
総数	12292	532	1383	2597	861	2598	1356	3573
修飾	<0.008%	<0.188%	<0.072%	<0.039%	<0.116%	<0.038%	<0.074%	<0.028%
P値								
特異性								

20

OffA-5

インデル	0	0	0	0	0	0	0	0
総数	60859	22846	25573	19054	25315	31754	66622	60925
修飾	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P値								
特異性								

30

OffA-6

インデル	0	0	0	0	0	0	0	0
総数	60859	22846	25573	19054	25315	31754	66622	60925
修飾	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P値								
特異性								

OffA-7

インデル	0	0	0	0	0	0	0	0
総数	60859	22846	25573	19054	25315	31754	66622	60925
修飾	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P値								

40

【 0 1 9 9 】

【表 8 - 8】

特異性

OffA-8

インデル	0	0	0	0	0	0	0	0
総数	9170	1614	5934	3215	2450	12750	10120	13003
修飾	<0.011%	<0.062%	<0.017%	<0.031%	<0.041%	<0.008%	<0.010%	<0.008%
P値								
特異性								

OffA-9

インデル	0	0	0	0	0	0	0	3
総数	8753	12766	9504	10114	11086	10676	9013	11110
修飾	<0.011%	<0.008%	<0.011%	<0.010%	<0.009%	<0.009%	<0.011%	0.027%
P値								
特異性								

10

OffA-10

インデル	1	0	0	2	2	3	5	7
総数	8151	16888	8804	7061	8891	32138	14889	40120
修飾	0.012%	<0.006%	<0.011%	0.028%	0.022%	0.009%	0.034%	0.017%
P値								
特異性								

20

OffA-11

インデル	0	0	1	0	0	0	9	76
総数	41343	32352	28834	28709	26188	32519	24894	19586
修飾	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	0.036%	0.388%
P値							2.7E-03	2.5E-18
特異性		>0	>274	>1302	>2564	>1016	448	47

OffA-12

インデル	0	0	0	0	0	0	0	0
総数	13186	2326	13981	12911	21134	9220	7792	8068
修飾	<0.008%	<0.043%	<0.007%	<0.008%	<0.006%	<0.011%	<0.013%	<0.012%
P値								
特異性								

30

OffA-13

インデル	0	0	0	0	0	2	9	0
総数	32704	32015	12312	23645	26315	24078	36111	22364
修飾	<0.006%	<0.006%	<0.008%	<0.006%	<0.006%	0.008%	0.025%	<0.006%
P値							2.7E-03	
特異性		>0	>225	>1302	>2564	816	649	>2725

40

OffA-15

【 0 2 0 0 】

【表 8 - 9】

インデル	0	0	0	0	1	0	0	0
総数	14654	15934	12313	6581	13053	18996	10916	21519
修飾	<0.007%	<0.006%	<0.008%	<0.015%	0.008%	<0.006%	<0.009%	<0.006%
P値								
特異性								
OffA-16								
インデル	1	0	0	0	0	0	0	12
総数	65190	35639	37252	30378	31489	22590	13594	20922
修飾	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.007%	0.057%
P値								7.9E-04
特異性		>0	>274	>1302	>2564	>1016	>2200	317
OffA-17								
インデル	0	0	0	0	0	0	0	6
総数	1972	606	1439	2113	2862	728	597	636
修飾	<0.051%	<0.165%	<0.069%	<0.047%	<0.035%	<0.137%	<0.168%	0.943%
P値								1.4E-02
特異性		>0	>26	>183	>489	>49	>97	19
OffA-18								
インデル	0	0	0	0	0	0	0	0
総数	5425	995	1453	1831	3132	1934	1534	5816
修飾	<0.018%	<0.101%	<0.069%	<0.055%	<0.032%	<0.052%	<0.065%	<0.017%
P値								
特異性								
OffA-19								
インデル	1	2	0	1	1	1	1	3
総数	31094	41252	33213	29518	32337	25904	27575	38711
修飾	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	0.008%
P値								
特異性								
OffA-21								
インデル	0	0	0	0	0	0	0	0
総数	15297	9710	16719	12119	15483	21692	16558	16418
修飾	<0.007%	<0.010%	<0.006%	<0.008%	<0.006%	<0.006%	<0.006%	<0.006%
P値								
特異性								
OffA-22								
インデル	27	41	38	46	32	50	55	57
総数	9406	11150	11516	10269	13814	14057	11685	14291
修飾	0.287%	0.368%	0.330%	0.448%	0.232%	0.356%	0.471%	0.399%

【 0 2 0 1 】

【表 8 - 1 0】

P値

特異性

OffA-23

インデル	1	0	0	0	0	0	10	20
総数	5671	9363	2203	7011	7078	12068	3484	8619
修飾	0.018%	<0.011%	<0.045%	<0.014%	<0.014%	<0.008%	0.287%	0.232%
P値							3.5E-03	9.1E-05
特異性		>0	>40	>609	>1210	>818	56	78

OffA-24

インデル	4	0	0	1	0	1	0	2
総数	17288	7909	14261	29936	6943	6333	14973	19953
修飾	0.023%	<0.013%	<0.007%	<0.006%	<0.014%	0.016%	<0.007%	0.010%
P値								
特異性								

OffA-25

インデル	0	0	0	0	0	0	0	0
総数	20089	45320	50758	108581	11574	20948	123827	74151
修飾	<0.006%	<0.006%	<0.006%	<0.006%	<0.009%	<0.006%	<0.006%	<0.006%
P値								
特異性								

OffA-27

インデル	0	0	0	0	1	0	0	0
総数	47338	14352	21253	17777	26512	19483	43728	29469
修飾	<0.006%	<0.007%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P値								
特異性								

OffA-29

インデル	0	0	0	0	0	0	0	0
総数	5174	12618	36909	18063	16486	17934	9999	35072
修飾	<0.019%	<0.008%	<0.006%	<0.006%	<0.006%	<0.006%	<0.010%	<0.006%
P値								
特異性								

OffA-30

インデル	4	4	0	7	4	4	0	3
総数	45082	56631	36333	88651	69652	20362	29180	21350
修飾	0.009%	0.007%	<0.006%	0.008%	<0.006%	0.020%	<0.006%	0.014%
P値								
特異性								

【 0 2 0 2】

10

20

30

40

【表 8 - 1 1】

OffA-32								
インデル	0	0	0	0	0	0	0	0
総数	13405	6721	14013	7513	14136	22376	6407	13720
修飾	<0.007%	<0.015%	<0.007%	<0.013%	<0.007%	<0.006%	<0.016%	<0.007%
P値								
特異性								
OffA-33								
インデル	0	0	0	0	1	1	0	4
総数	108222	46866	157329	48611	92559	152094	201408	225805
修飾	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P値								
特異性								
OffA-34								
インデル	0	0	0	0	0	0	0	2
総数	3889	3158	2903	2235	2112	3022	2322	2481
修飾	<0.026%	<0.032%	<0.034%	<0.045%	<0.047%	<0.033%	<0.043%	0.081%
P値								
特異性								
OffA-35								
インデル	0	0	0	1	0	0	0	33
総数	48482	37431	38043	31033	44803	37257	41073	47273
修飾	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	0.070%
P値								9.2E-09
特異性		>0	>274	>1302	>2564	>1016	>2428	260
OffA-36								
インデル	0	0	2	0	0	0	0	0
総数	27115	17075	45425	35059	22298	19610	12620	27170
修飾	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.008%	<0.006%
P値								
特異性								

表 7 . オンターゲットおよび予測オフターゲットゲノム部位において T A L E N により誘導される細胞修飾。(A) T A L E N なしで、または標準、Q 3 もしくは Q 7 C 末端ドメイン、ならびに E L / K K ヘテロ二量体、E L D / K K R ヘテロ二量体、もしくはホモ二量体 (H o m o) F o k I ドメインのいずれかを含有する T A L E N で処置したヒト細胞から単離されたゲノム D N A から増幅された、C C R 5 A オンターゲットおよび各予測されたゲノムオフターゲット部位のシーケンシングの結果。インデル：T A L E N 誘導性の切断と整合する、挿入または欠失を含む観測された配列の数。総数：配列カウントの総数。修飾：インデルの数を、配列の総数で割ったパーセンテージ。潜在的な修飾の上限値は、観測されたインデルなしの部位について 1 未満のインデルを仮定し、配列カウントの総数で割り算して修飾率の上限に到達するか、または検出の理論的限界値 (1 / 1 6 , 4 0 0) のうち、より保守的な (大きい) 値を取る。P 値：各 T A L E N 処置試料と未処置の対照試料との間で前に報告されたように 5 計算する。0 . 0 5 未満の P 値が示される。特異性：各部位について、オンターゲットゲノム修飾頻度のオフターゲットゲノム修飾頻度に対する比率。(B) A T M 標的部位について、(A) と同様。

【 0 2 0 3 】

【表 9】

TALEN 選択	a	b	R ²
L13+R10 CCR5B	1.00	-1.88	0.999937
L10+R10 CCR5B	1.00	-1.85	0.999901
L10+R13 CCR5B	1.00	-1.71	0.999822
L13+R13 CCR5B	1.00	-1.64	0.999771
L13+R16 CCR5B	1.00	-1.15	0.998286
L16+R10 CCR5B	1.00	-1.24	0.998252
L10+R16 CCR5B	1.01	-1.08	0.996343
L16+R13 CCR5B	1.01	-1.04	0.995844
L16+R16 CCR5B	1.03	-0.70	0.977880
L18+R18 ATM	1.08	-0.36	0.913087
L18+R18 CCR5A	1.13	-0.21	0.798923

表 8 . 変異数の関数としての濃縮値の指数フィッティング。変異の関数としての選択後配列の濃縮値を、オンターゲット濃縮（定義により = 1 . 0）に対して正規化した。0 ~ 4 個の変異を有する配列の正規化濃縮値を、指数関数 $a * e^b$ にフィットさせ、ここで R² は、非線形最小二乗法を用いて報告された。

【 0 2 0 4 】

【表 1 0】

TALEN 選択	範囲	a	b	R ²
L16+R16 CCR5B	3-5	1.00	-1.638	0.99998
L16+R13 CCR5B	2-4	1.00	-1.733	0.99998
L16+R10 CCR5B	2-4	1.00	-2.023	0.99999
L13+R16 CCR5B	2-4	1.00	-1.844	0.99997
L13+R13 CCR5B	1-3	1.00	-2.014	0.99998
L13+R10 CCR5B	1-3	1.00	-2.205	0.99999
L10+R16 CCR5B	2-4	1.00	-1.929	0.99995
L10+R13 CCR5B	1-3	1.00	-2.110	0.99998
L10+R10 CCR5B	1-3	1.00	-2.254	0.99999

表 9 . 変異数の関数としての濃縮値の指数フィッティングおよび外挿。変異数の関数としての、全 9 個の C C R 5 B 選択からの全ての配列の濃縮値を、示した範囲内の最小変異数（定義により = 1 . 0）を有する配列の濃縮値に対して正規化した。指定された変異の範囲からの配列の正規化された濃縮値は、指数関数 $a * e^b$ にフィットさせ、ここで R² は、非線形最小二乗法を用いて報告された。これらの指数関数的減少 b を用いて、すべての平均濃縮値を 5 つの変異を超えて外挿した。

【 0 2 0 5 】

【表 1 1 - 2】

	NNNNNNNNNNNNNNNT%T%T%A%T%T%T%A%T%T%T%A%T%T%T%A%GGTACCCC A
#1 アダプター-fwd**1	AATGATACGGCGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTACTGT
#1 アダプター-rev**1	ACAGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 アダプター-fwd**2	AATGATACGGCGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTCTGAA
#1 アダプター-rev**2	TTGAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 アダプター-fwd**3	AATGATACGGCGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTTGCAA
#1 アダプター-rev**3	TTGCAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 アダプター-fwd**4	AATGATACGGCGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTTGACT
#1 アダプター-rev**4	AGTCAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 アダプター-fwd**5	AATGATACGGCGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTGCATT
#1 アダプター-rev**5	AATGCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 アダプター-fwd**6	AATGATACGGCGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTCATGA
#1 アダプター-rev**6	TCATGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 アダプター-fwd**7	AATGATACGGCGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTATGCT
#1 アダプター-rev**7	AGCATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 アダプター-fwd**8	AATGATACGGCGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTCTAGT
#1 アダプター-rev**8	ACTAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 アダプター-fwd**9	AATGATACGGCGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTGCTAA
#1 アダプター-rev**10	TTAGCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 アダプター-fwd**10	AATGATACGGCGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTCAGTA
#1 アダプター-rev**11	TACTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 アダプター-fwd**11	AATGATACGGCGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTGTACT
#1 アダプター-rev**12	AGTACAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 アダプター-fwd**12	AATGATACGGCGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTACTGT
#1 アダプター-rev**13	ACAGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 アダプター-fwd**13	AATGATACGGCGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTGCTAA
#1 アダプター-rev**14	TTAGCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 アダプター-fwd**14	AATGATACGGCGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTCAGTA
#1 アダプター-rev**14	TACTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 アダプター-fwd**15	AATGATACGGCGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTGTACT
#1 アダプター-rev**15	AGTACAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 アダプター-fwd**16	AATGATACGGCGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTACTGT
#1 アダプター-rev**16	ACAGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#2A プライマー-fwd	AATGATACGGCGACCAAC
#2A プライマー-rev*CCR5A	GTTCAGACGTGTGCTCTTCCGATCTNNNNAGTGGTGAGCGTGACG
#2A プライマー-rev*ATM	GTTCAGACGTGTGCTCTTCCGATCTNNNNACGCGGAGTGGGGTACC
#2A プライマー-rev*CCR5B	CAGACGTGTGCTCTTCCGATCTNNNNAGCGGTGGAGTCCCGAGG
#2B プライマー-fwd	AATGATACGGCGACCAAC
#2B プライマー-rev**1	CAAGCAGAAGACGGCATACGAGATGTTGACTGTGACTGGAGTTCAGACGTGTGCTCTTC
#2B プライマー-rev**2	CAAGCAGAAGACGGCATACGAGATACGGAAGTGTGACTGGAGTTCAGACGTGTGCTCTTC
#2B プライマー-rev**3	CAAGCAGAAGACGGCATACGAGATCTAATCATGTGACTGGAGTTCAGACGTGTGCTCTTC
#2B プライマー-rev**4	CAAGCAGAAGACGGCATACGAGATCGGGACGGGTGACTGGAGTTCAGACGTGTGCTCTTC
#1 Lib. アダプター-fwd*CCR5A	GTACCCAGATCGGAAGAGCACACGTCTGAACTCCAGTCACACAGTATCTCGATGCCGCTT CTGCTTG
#1 Lib. アダプター-rev*CCR5A	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTG
#1 Lib. アダプター-fwd*ATM	GTACGATCGATCGGAAGAGCACACGTCTGAACTCCAGTCACCTAGGCATCTCGATGCCGTC TTCTGCTTG
#1 Lib. アダプター-rev*ATM	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCGCATC
#1 Lib. アダプター-fwd*CCR5B	TCGGGAACGTGATCGGAAGAGCACACGTCTGAACTCCAGTCACCGTCTAATCTCGATGCCGT CTTCTGCTTG
#1 Lib. アダプター-rev*CCR5B	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCACGTT
#2A Lib. プライマー-rev	CAAGCAGAAGACGGCATACGA
#2A Lib. プライマー-fwd*CCR5A	AATGATACGGCGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTNNNNCG TCACGCTCACCACT

【 0 2 0 7 】

【表 11 - 3】

#2A Lib. プライマー - fwd*ATM	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNG GTACCCCACTCCGCGT
#2A Lib. プライマー - fwd*CCR5B	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCC TCGGGACTCCACGCT
#2B Lib. プライマー - rev	CAAGCAGAAGACGGCATACGA
#2B Lib. プライマー - fwd	AATGATACGGCGACCAC
G アダプター - fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
G アダプター - rev	/5Phos/GATCGGAAGAGCACACGCTCTGAACCTCCA
G-B プライマー - fwd	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC
G-B プライマー - rev**1	CAAGCAGAAGACGGCATACGAGATGTGCGGACGTGACTGGAGTTCAGACGTGTGCT
G-B プライマー - rev**2	CAAGCAGAAGACGGCATACGAGATCGTTTCACGTGACTGGAGTTCAGACGTGTGCT
G-B プライマー - rev**3	CAAGCAGAAGACGGCATACGAGATAAGGCCACGTGACTGGAGTTCAGACGTGTGCT
G-B プライマー - rev**4	CAAGCAGAAGACGGCATACGAGATTCCGAAACGTGACTGGAGTTCAGACGTGTGCT
G-B プライマー - rev**5	CAAGCAGAAGACGGCATACGAGATTACGTACGGTGAAGTTCAGACGTGTGCT
G-B プライマー - rev**6	CAAGCAGAAGACGGCATACGAGATATCCACTCGTGAAGTTCAGACGTGTGCT
G-B プライマー - rev**7	CAAGCAGAAGACGGCATACGAGATAAAGGAATGTGACTGGAGTTCAGACGTGTGCT
G-B プライマー - rev**8	CAAGCAGAAGACGGCATACGAGATATATCAGTGTGACTGGAGTTCAGACGTGTGCT
CCR5AonCfwd	CGACGGTCTAGAGTCTTCAATACACCTGCAGCTCTCATTTTCCATACAGT
CCR5Amut1fwd	CGACGGTCTAGAGTCTTCAATACACCTGCAGCTCTCATTTTCCATACAGT
CCR5Amut2fwd	CGACGGTCTAGAGTCTTCAATACACCTGCAGCTCTCATTTTCCATACAGT
CCR5Amut3fwd	CGACGGTCTAGAGTCTTCAATACACCTGCAGCTCTCATTTTCCATACAGT
CCR5Amut4fwd	CGACGGTCTAGAGTCTTCAATGACCTGCAGCTCTCATTTTCCATACAGT
CCR5AonCrev	CCGACGAAGCTTTTCTTCCAGAATTGATACTGACTGTATGGAAAATGA
CCR5Amut1rev	CCGACGAAGCTTTTCTTCCAGAATTGATACTGACTGTATGGAAAATGA
CCR5Amut2rev	CCGACGAAGCTTTTCTTCCAGAATTGATACTGACTGTATGGAAAATGA
CCR5Amut3rev	CCGACGAAGCTTTTCTTCCAGAATTGATACTGACTGTATGGAAAATGA
CCR5Amut4rev	CCGACGAAGCTTTTCTTCCAGAATTGATACTGACTGTATGGAAAATGA
ATMonAfwd	CGACGGTCTAGATTGAAATGGGATGCTGTTTTAGGTATTCTATTCAAAT
ATMmut1fwd	CGACGGTCTAGATTGAAATGGGATGCTGTTTTAGGTATTCTATTCAAAT
ATMmut2fwd	CGACGGTCTAGATTGAAATGGGATGCTGTTTTAGGTATTCTATTCAAAT
ATMmut3fwd	CGACGGTCTAGATTGAAATGGGATGCTGTTTTAGGTATTCTATTCAAAT
ATMmut4fwd	CGACGGTCTAGATTGAAATGGGATGCTGTTTTAGGTATTCTATTCAAAT
ATMonArev	CCGACGAAGCTTAATAAAGACAGTAAATAAATTTGAATAGAATACCTAAAA
ATMmut1rev	CCGACGAAGCTTAATAAAGACAGTAAATAAATTTGAATAGAATACCTAAAA
ATMmut2rev	CCGACGAAGCTTAATAAAGACAGTAAATAAATTTGAATAGAATACCTAAAA
ATMmut3rev	CCGACGAAGCTTAATAAAGACAGTAAATAAATTTGAATAGAATACCTAAAA
ATMmut4rev	CCGACGAAGCTTAATAAAGACAGTAAATAAATTTGAATAGAATACCTAAAA
CCR5BonBfwd	CGACGGTCTAGAAAGGTCTTCAATACACCTGCAGCTCTCATTTTCCATACAGTCA
CCR5Bmut1fwd	CGACGGTCTAGAGTCTTCAATACACCTGCAGCTCTCATTTTC
CCR5Bmut2fwd	CGACGGTCTAGAGTCTTCAATACACCTGCAGCTCTCATTTTC
CCR5Bmut3fwd	CGACGGTCTAGAGTCTTCAATACACCTGCAGCTCTCATTTTC
CCR5Bmut4fwd	CGACGGTCTAGAGTCTTCAATACACCTGCAGCTCTCATTTTC
CCR5Bmut5fwd	CGACGGTCTAGAGTCTTCAATACACCTGCAGCTCTCATTTTC
CCR5Bmut6fwd	CGACGGTCTAGAGTCTTCAATGACCTGCAGCTCTCATTTTC
CCR5BonBrev	CCGACGAAGCTTTCTTCCAGAATTGATACTGACTGTATGGAAAATGAGAGCT
CCR5Bmut1rev	CCGACGAAGCTTTCTTCCAGAATTGATACTGACTGTATGGAAAATGAGAGCT
CCR5Bmut2rev	CCGACGAAGCTTTCTTCCAGAATTGATACTGACTGTATGGAAAATGAGAGCT
CCR5Bmut3rev	CCGACGAAGCTTTCTTCCAGAATTGATACTGACTGTATGGAAAATGAGAGCT
CCR5Bmut4rev	CCGACGAAGCTTTCTTCCAGAATTGATACTGACTGTATGGAAAATGAGAGCT
CCR5Bmut5rev	CCGACGAAGCTTTCTTCCAGAATTGATACTGACTGTATGGAAAATGAGAGCT
CCR5Bmut6rev	CCGACGAAGCTTTCTTCCAGAATTGATACTGACTGTATGGAAAATGAGAGCT
CCR5Bmut7rev	CCGACGAAGCTTTCTTCCAGAATTGATACTGACTGTATGGAAAATGAGAGCT
pUC19Ofwd	GCGACACGGAAATGTTGAATACTCAT
pUC19Orev	CAGCGAGTCAGTGAGCGA

【 0 2 0 8 】

10

20

30

40

【表 11 - 4】

B

DNA 基質名	オリゴヌクレオチドの組み合わせ
A1	ATMmut1fwd + ATMonArev
A2	ATMmut2fwd + ATMonArev
A3	ATMonAfwd + ATMmut1rev
A4	ATMonAfwd + ATMmut2rev
A5	ATMmut2fwd + ATMmut2rev
A6	ATMmut3fwd + ATMmut3rev
A7	ATMmut1fwd + ATMmut1rev
A8	ATMmut4fwd + ATMmut4rev
C1	CCR5Amut1fwd + CCR5AonCrev
C2	CCR5Amut2fwd + CCR5AonCrev
C3	CCR5Amut3fwd + CCR5AonCrev
C4	CCR5Amut4fwd + CCR5AonCrev
C5	CCR5AonAfwd + CCR5Amut1rev
C6	CCR5AonAfwd + CCR5Amut2rev
C7	CCR5AonAfwd + CCR5Amut3rev
C8	CCR5AonAfwd + CCR5Amut4rev
B1	CCR5Bmut1fwd + CCR5BonBrev
B2	CCR5Bmut2fwd + CCR5BonBrev
B3	CCR5Bmut3fwd + CCR5BonBrev
B4	CCR5BonBfwd + CCR5Bmut1rev
B5	CCR5BonBfwd + CCR5Bmut2rev
B6	CCR5BonBfwd + CCR5Bmut3rev
B7	CCR5BonBfwd + CCR5Bmut4rev
B8	CCR5Bmut4fwd + CCR5BonBrev
B9	CCR5Bmut5fwd + CCR5BonBrev
B10	CCR5Bmut6fwd + CCR5BonBrev
B11	CCR5BonBfwd + CCR5Bmut5rev
B12	CCR5BonBfwd + CCR5Bmut6rev
B13	CCR5BonBfwd + CCR5Bmut7rev
B14	CCR5Bmut1fwd + CCR5Bmut1rev
B15	CCR5Bmut2fwd + CCR5Bmut2rev
B16	CCR5Bmut1fwd + CCR5Bmut3rev

【 0 2 0 9 】

【表 1 1 - 5】

C

部位	Fwd プライマー	Rev プライマー	PCR
OnCCR5A	TCACTTGGGTGGTGGCTGTG	GACCATGACAAGCAGCGGCA	
OffC-1	AGTCCAAGACCAGCCTGGGG	AAGAACCTGTTGTCTAATCCAGCA	
OffC-2	GAACCTGTTGTCTAATCCAGCGTC	CTGCAAGAAGGCCAGGCA	
OffC-3	AGTCCAAGACCAGCCTGGGG	AAGAACCTGTTGTCTAATCCAGCA	
OffC-4	TGACCTGTTTGTTCAGGTCTTCC	CCATATGGTCCCTGTCGCAA	
OffC-5	TCCAGTTGCTGTCCCTTCAGA	ACAGGGAGAGCCACCAATGC	
OffC-6	GCCCGGCCTGTCCTGTATTT	CACCCACACATGCACTTCCC	
OffC-7	TGGCTATTCTAGTTCTTTTGCAT	CCATGCCCTAGGGATTTGTGGA	
OffC-8	CGCTGAAGGCTGTCACCCCTAA	TGGACCTAAGAGTCCTGCCCAT	
OffC-9	CCACCACCACACAACCTTCACA	CAGCTGGCGAGAAGTGCAAA	ND
OffC-10	TTCCAGGTCCCTTGCACAAATA	GCAAGGTCGTTGGATAGAAGTTGA	
OffC-11	CACCGAAAGCAACCCATTCC	TGATCTGCCCACCCCAGACT	
OffC-12	TTCATTCTCACCATCTGGAATTGG	TCTGGCTGGACTGCTCTGGTT	
OffC-13	TGGCATGTGGATCAGTACCCA	TAGAACATGCCCCGGAACAG	
OffC-14	CTGACGTCCATGTCAACGGG	TTTGAATTCCCCCTCCCCAT	
OffC-15	GCTCCTTTCTGAGAAGCACCCAT	GGCAGATGGTGGCAGGTCTT	
OffC-16	ATGAGGGGCTTGGATTGGCTG	CCACCTCCCCCACTGCAATA	

10

20

【 0 2 1 0 】

【表 11 - 6】

OffC-17	GGAGGCCCTTCATTGTGTCACG	AACTCCACCTGGGTGCCCTA	
OffC-18	CGTGGTCCCCCAGAAATCAC	GGAGCAGGAGTTGGTGGCAT	
OffC-19	GATTGCATAGGTTAGCATTGCC	GCCCCTGTTGGTTGACTCCC	
OffC-20	TTCCAGCGAATGGAAAGTGCT	AAGCCCAGGAATAAGGGCCA	
OffC-21	AAGCATGCTCACACTGTGGTGTA	TTGCTTGAGGCGGAAGTTGC	
OffC-22	TGACCCTCCAGCAAAGGTGA	CCCCAGGACTGAGCATGAG	
OffC-23	GCTTTGCTTGCACTGTGCCTT	GGGGACAGACTGTGAGGGCT	
OffC-24	TCAAAAGGATGTGATCTGCCACA	GGCCTCTTTGAGGGCCAGTT	
OffC-25	CCAGGGCTCAATTCTTAGACCG	AAAAGAGCAGGGCTGCCATC	
OffC-26	TGTTTCATGCCTGCACAGTGG	TGGATGTGCCCTCTACCACA	
OffC-27	TTTGGCAAGGAATTCACAGTTC	TCATGCCTGCACAGTGGTTG	
OffC-28	GGAGGATGTCTTTGTGGTAGGGG	CGCTGCCAAGCAAACCTCAA	
OffC-29	TCCCCCAACTTCACTGTTTTT	GCAATGAGCATGTGGACACCA	
OffC-30	TTCTCTGTTTCCAGTGATTTTCA	GTGCAAAACAGCCAGTTGC	+DMSO
OffC-31	TGGCTTGGTTAATGGACAATGG	CCTGCAAGGAGCAAGGCTTC	+DMSO
OffC-32	TGGGCTTCGTTGACTTAAAGAG	GGACAAGAGGGCCAGGGTTT	
OffC-33	TCTTAAACATGTGGAACCCAGTCAT	TGAAAACCCACAGAGTGGGAGA	
OffC-34	GCAGATTCATTAGCGTTTGTGGC	TGCATGGGTGTAAATGTAGCAGAAA	
OffC-35	CCAAGGATCAATACCTTTGGAGGA	GCCCTCCCTTGAATCAGGCT	
OffC-36	TTCCCCAACCAGGGGCGAGT	GTGGTGAGTGGGTGTGGCAG	+DMSO
OnATM	AGCGCCTGATTCGAGATCCT	AGCGCCTGATTCGAGATCCT	
OffA-1	CCTGCCATTGAATTCAGCCT	TGTCTGCCTTTCTGTCCCC	
OffA-2	GACTGCCACTGCACTCCAC	GGATACCCTTGCCTCCAC	
OffA-3	CCTCCCATTTTCTTCTCCA	CTGGGAGACACAGGTGGCAG	
OffA-4	TCCTCCAATTTCTTCTCCA	CTGGGAGACACAGGTGGCAG	
OffA-5	CTGGGAGACACAGGTGGCAG	AGGACCAATGGGGCCAATCT	
OffA-6	CTGGGAGACACAGGTGGCAG	AGGACCAATGGGGCCAATCT	
OffA-7	CTGGGAGACACAGGTGGCAG	AGGACCAATGGGGCCAATCT	
OffA-8	GCATGCCAAAGAAATTGTAGGC	TTCCCCCTGTCATGGTCTTCA	
OffA-9	GCATCTCTGCATTCCTCAGAAAGTG	AGAAACTGAGCAAGCCTCAGTCAA	
OffA-10	GGGATACCAAAGAGCTTTTGTGTT	CAGAGGCTGCATGATGCCTAATA	
OffA-11	TGCAGCTACGGATGAAAACCAT	TCAGAATACCTCCCCGCCAG	
OffA-12	GCATAAAGCACAGGATGGGAGA	TCCCTCTTAAACGGTTATGTTGGC	
OffA-13	TGGGTAAAGTAATTCGAAAGGAGAA	ATGTGCCCCACACATTGCC	+DMSO
OffA-14	GAGTGAGCCACTGCACCCAG	CGTGTGGTGGTGGCACAAG	ND
OffA-15	CCTCCCTCTGGCTCCCTCCC	ACCAGGGCCTGTTGGGGGT	
OffA-16	TGCTCCCTGACCTTCTGAGA	CCATTGGAATGAGAACCTTCTGG	
OffA-17	GGTGAACAATCCACCTGTATTAGC	GAATGTGACACCACCACCGC	
OffA-18	GGCTTTGCAACATAAACTCA	CCTTCTGAGCAGCTGGGACAA	
OffA-19	CACTGGAACCCAGGAGGTGG	CCTCCCATTTGAGCCTTGGT	
OffA-20	CAGCCTGCCTGGGTGACAG	CATCTGAGCTCAAACTGTCTGC	+DMSO
OffA-21	GCCACTGCATTGCATTTCC	TGAGGGCAGGTCTGTTTCCCTG	ND
OffA-22	GGGAGGATCTCTCGAGTCCAGG	CCTTGCCTGACTTGCCCTGT	
OffA-23	TGTTTAGTAATTAAGACCCTGGCTTC	GCGACAGGTACAAAGCAGTCCAT	
OffA-24	GCCCTTTGATTTCACTGTTTCCC	CATTGCTGCCATTGCACTCC	
OffA-25	AACTGGCACATGTACTCCT	ACATGATTTGATTTTTCATGTGTTT	
OffA-26	GGGTGGAAGGTGAGAGGAGATT	CGCAGATGGGCATGTTATTG	ND
OffA-27	CCTCCCATTTTCTTCTCCA	GACTGCCACTGCACTCCAC	
OffA-28	AGCCAAGATTGCACCATTCG	GTCCCTGACGGAGGCTGAGA	ND
OffA-29	TGGTTGGAATTTTGGCTCTGTCAC	TGCAATATCAATACCCTGCTTTCCTC	

【 0 2 1 1 】

【図 2 - 1】

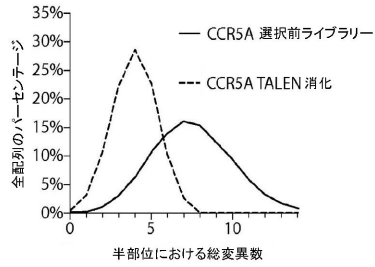


図 2A

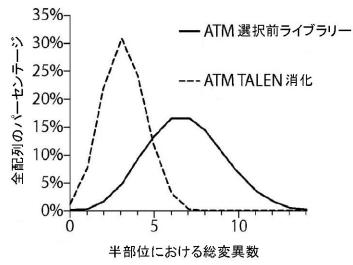


図 2B

【図 2 - 2】

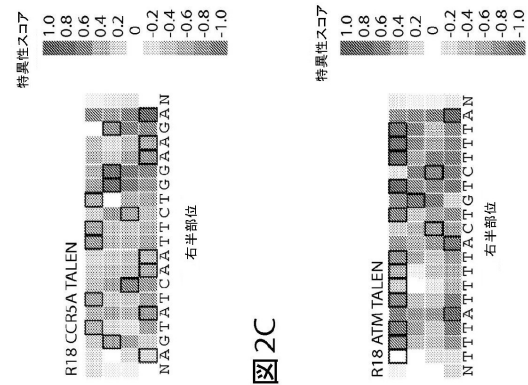


図 2C

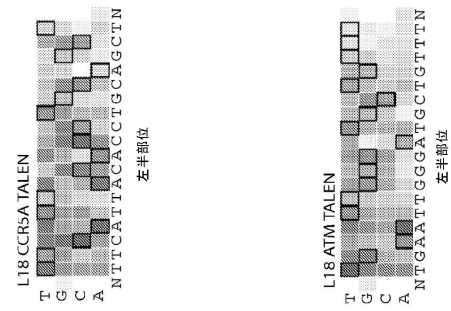


図 2D

【図 2 - 3】

DNA	濃縮	左半部位	右半部位
OnB	1.0	TCATTACACCTGC	CATACAGTCAGTA
B1	0.47	TCATTACACCTGt	CATACAGTCAGTA
B2	0.46	TCATaACACCTGC	CATACAGTCAGTA
B3	0.13	TCATTACACCeGC	CATACAGTCAGTA
B4	1.1	TCATTACACCTGC	CATACAGTaaGTA
B5	0.84	TCATTACACCTGC	gATACAGTCAGTA
B6	0.21	TCATTACACCTGC	CATtCAGTCAGTA
B7	0.03	TCATTACACCTGC	CATACAGTCgGTA
B8	0.25	TCATaACACCTGt	CATACAGTCAGTA
B9	0.25	TCATTAtACCTaC	CATACAGTCAGTA
B10	0.00	TCATTgCACCeGC	CATACAGTCAGTA
B11	0.55	TCATTACACCTGC	gATACAGTaaGTA
B12	0.04	TCATTACACCTGC	CACAcAtTCAGTA
B13	0.03	TCATTACACCTGC	CATACtAtTCAGTA
B14	0.30	TCATTACACCTGt	CATACAGTaaGTA
B15	0.18	TCATaACACCTGC	gATACAGTCAGTA
B16	0.04	TCATTACACCTGt	CATtCAGTCAGTA

図 2E

【図 2 - 4】

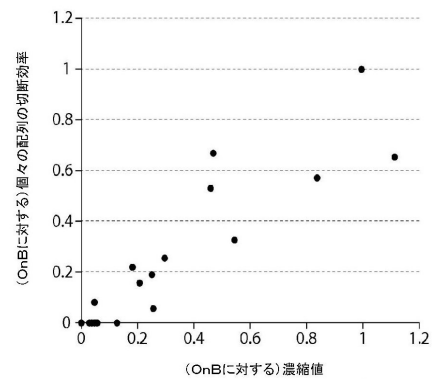


図 2F

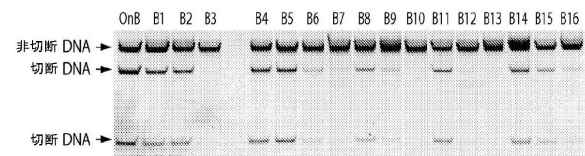


図 2G

【図3-1】

部位	TALENなし	CCR5A EL/KK FokI	CCR5A ELD/KKR FokI	CCR5A Homo FokI
OnCCR5A	<0.006%	9.84%	27.6%	46.8%
OffC-2	<0.006%	<0.006%	0.006%	<0.006%
OffC-5	<0.006%	0.45%	1.96%	2.78%
OffC-15	<0.020%	<0.014%	0.230%	0.043%
OffC-16	<0.006%	<0.006%	0.031%	<0.006%
OffC-28	<0.009%	0.014%	0.163%	0.056%
OffC-36	<0.006%	<0.006%	0.152%	0.028%

図3A

部位	TALENなし	ATM EL/KK FokI	ATM ELD/KKR FokI	ATM Homo FokI
OnATM	0.006%	6.78%	16.2%	18.2%
OffA-1	<0.006%	<0.006%	0.03%	0.08%
OffA-11	<0.006%	<0.006%	0.04%	0.39%
OffA-13	<0.006%	0.01%	0.02%	<0.006%
OffA-16	<0.006%	<0.006%	<0.006%	0.06%
OffA-17	<0.05%	<0.14%	<0.17%	0.94%
OffA-23	0.02%	<0.006%	0.29%	0.23%
OffA-35	<0.006%	<0.006%	<0.006%	0.07%

図3B

【図3-2】

OnCCR5A
 TTCATTACACCTGCAGCTCTCATTTTCCATACAGTCAGTATCAATTCGTGGAAGA (7267)
 TTCATTACACCTGCAGCTCTCAT-----ACAGTCAGTATCAATTCGTGGAAGA (76)
 TTCATTACACCTGCAG-----TCAGTATCAATTCGTGGAAGA (63)
 TTCATTACACCTG-----GAAGA (61)

OffC-2
 TACATCACATATGCAAAITGACTCAAAATGGATCATAGACCTAAATGCTATCATTTCTGGGAGA (163332)
 TACATCACATATGCAAAITGACTCAAAATGGATCA---ACCTAAATGCTATCATTTCTGGGAGA (6)
 TACATCACATATGCAAAITGACTCAAAATG-----GACCTAAATGCTATCATTTCTGGGAGA (4)

OffC-5
 TCCAATACCTCTGCCACACCCAGGCAITGGCCAGGAGCAACTCTGGGAGA (17045)
 TCCAATACCTCTGCCACAC-----CCAGGAGCAACTCTGGGAGA (28)
 TCCAATACCTCTG-----GCAITGGCCAGGAGCAACTCTGGGAGA (12)
 TCCAATAC-----CTCTGGGAGA (10)

OffC-15
 TCCATGACACAAAAGACTTCCCTGATTTCTCTTAAGGCATCACTGATCTATCTCTGGAATA (6967)
 TCCATGACACAAAAGACTTCCCTGATTTCTCTTAAGG-----CTGATCTATCTCTGGAATA (6)

OffC-16
 TTCCTCCACCACTGTGCCACAGCTTTCACACTGATCACCATAATCCAGCATCAATCCTGGGAAGA (38536)
 TTCCTCCACCACTGTGCCACAGTC-----CACCATAATCCAGCATCAATCCTGGGAAGA (4)

OffC-28
 TTTATTACACTTCCAGATCTTTTATTTTAAAGTTACAGATATCCTTTCTGGGAAGA (7379)
 TTTATTACACTT-----CCAGATATCCTTTCTGGGAAGA (3)
 TTTATTACACTTCCAGATCTTT-----ATATCCTTTCTGGGAAGA (2)
 TTTATTACACTTCCAGATCTTT-----TATCCTTTCTGGGAAGA (2)

OffC-36
 CTCCTTAATACCTGCAAAITATAAGGACACTATTTGACTTGATATTATTTCTGGGAGGA (12461)
 CTCCTTAATACCTGCAAAITATAAGGACACT---GACTTGATATTATTTCTGGGAGGA (11)

図3C

【図4-1】

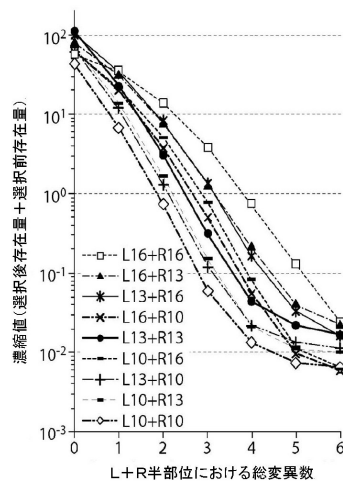


図4A

【図4-2】

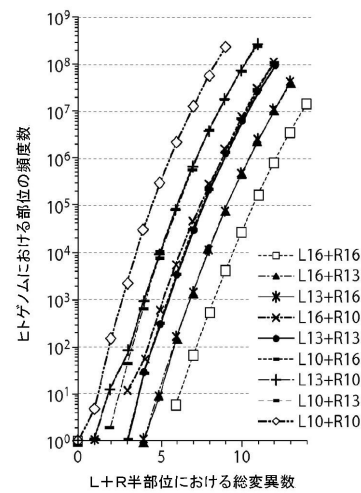


図4B

【図 4 - 3】

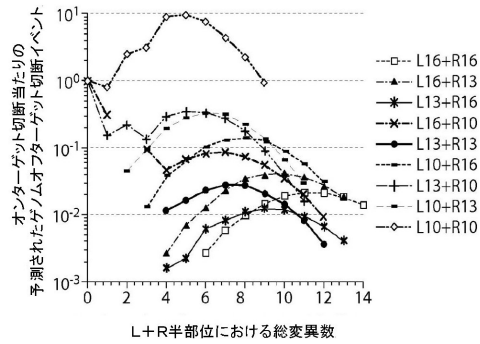


図 4C

【図 5 - 1】

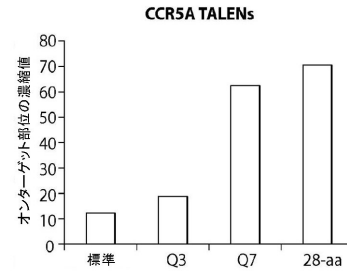


図 5A

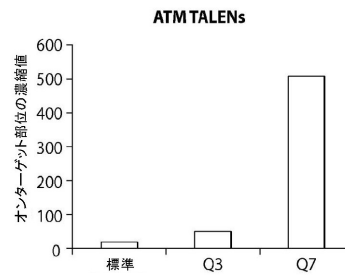


図 5B

【図 5 - 2】

DNA	左半部位	右半部位
OnC	TTCATTACACCTGCAGCT	AGTATCAATTCTGGAAGA
C1	TTCATTACaTCTGCAcCT	AGTATCAATTCTGGAAGA
C2	TTCaATACACCTGTAGCT	AGTATCAATTCTGGAAGA
C3	TTCATTACACCCcGCAGCa	AGTATCAATTCTGGAAGA
C4	TTaATTgCACCTGCAGCT	AGTATCAATTCTGGAAGA
C5	TTCATTACACCTGCAGCT	AGTATgAATTCTGtAAGA
C6	TTCATTACACCTGCAGCT	AGTATCAAcTCTGGAgGA
C7	TTCATTACACCTGCAGCT	AtTATCAATTCaGGAAGA
C8	TTCATTACACCTGCAGCT	AGTAaCRAItgCTGGAAGA

図 5C

【図 5 - 3】

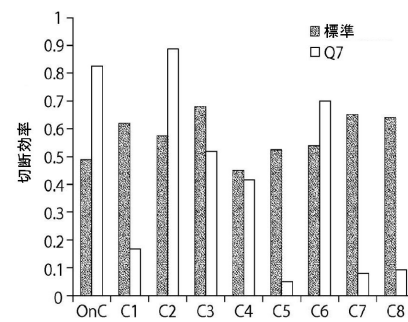


図 5E

DNA	左半部位	右半部位
OnA	TGAATTGGGATGCTGTTT	TTTATTTTACTGTCCTTTA
A1	TGAATTGGGtTGCTGTTT	TTTATTTTACTGTCCTTTA
A2	TGAATTGcGATGCTGTTT	TTTATTTTACTGTCCTTTA
A3	TGAATTGGGATGCTGTTT	TTTATTTcACTGTCCTTTA
A4	TGAATTGGGATGCTGTTT	TTTATTTTACTaTCTTTA
A5	TGAATTGcGATGCTGTTT	TTTATTTTACTaTCTTTA
A6	TGAgtTGGGATGCTGTTT	TTTATgtTACTGTCCTTTA
A7	TGAATTGGGtTGCTGTTT	TTTATTTgACTGTCCTTTA
A8	TGAATTGGGATGCTGaTT	TTTATTTTACTGTCcTTA

図 5D

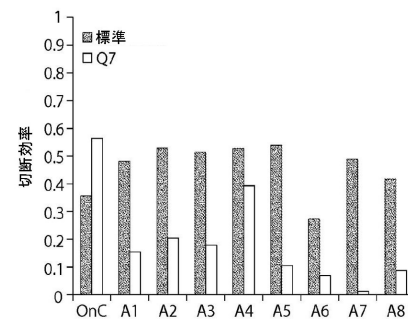


図 5F

【 図 6 】

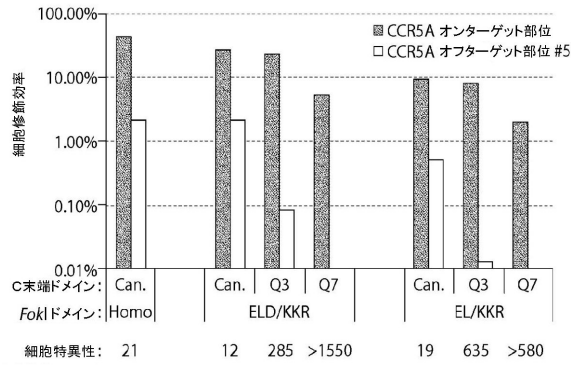


図 6

【圖 7】

CCR5 標的部位

TALEN 单量体

CCR5A L18 5'-TTCATTACACCTGCAGCT
CCR5B L16 5'-TCTTCATTACACCTGC
CCR5B L13 5'-TCATTACACCTGC
CCR5B L10 5'-TTACACCTGC

TCTTCATTACACCTGCAGCTCTCATTTTCCATACAGTCAGTATCAATTCTGGAAGA
AGAAGTAATGTGGACGTCGAGAGTAAAAGGTATGTTCAGTCATAGTTAAGACCTTCT

CCR5A R18	TCATAGTTAAGACCTTCT-5'
CCR5B R16	GTATGTCAGTCATAGT-5'
CCR5B R13	GTATGTCAGTCAT-5'
CCR5B R10	GTATGTCAGT-5'

图 7A

ATM 標的部位

TALEN 单量体

ATM L18 5'-TGAATTGGGATGCTGTT

TGAATTGGGATGCTGTTTTAGGTATTCTATTCAAATTTATTTTACTGTCTTTA
ACTTAACCTACGACAAAAATCCATAAGATAAGTTTAAATAAAATGACAGAAAT

ATM R18 AAATAAAATGACAGAAAT-5'

図7B

【 図 8 - 1 】

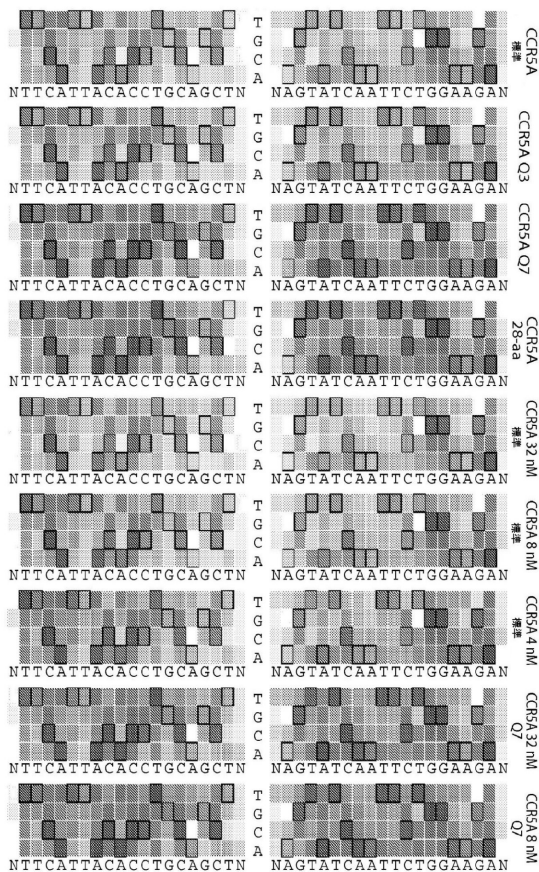


図 8A

【 図 8 - 2 】

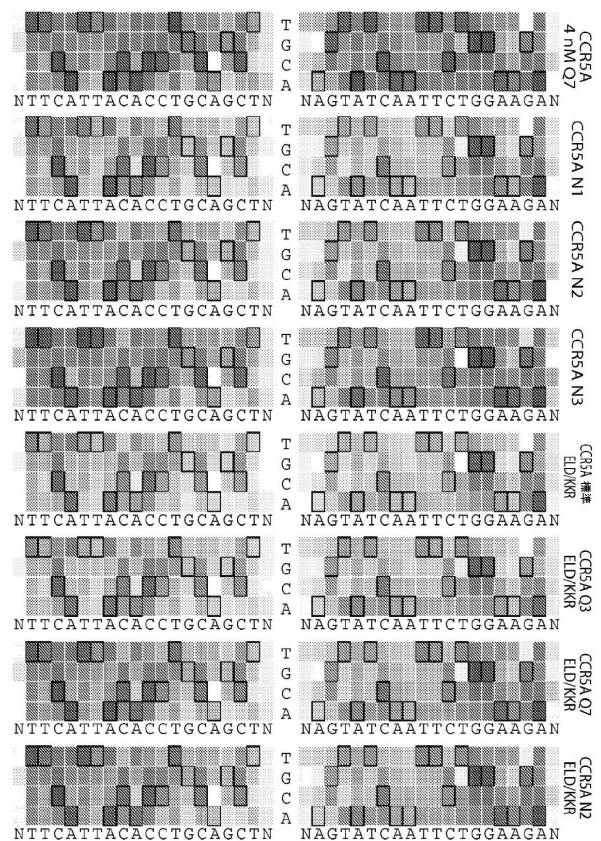


图 8B

【図9-1】

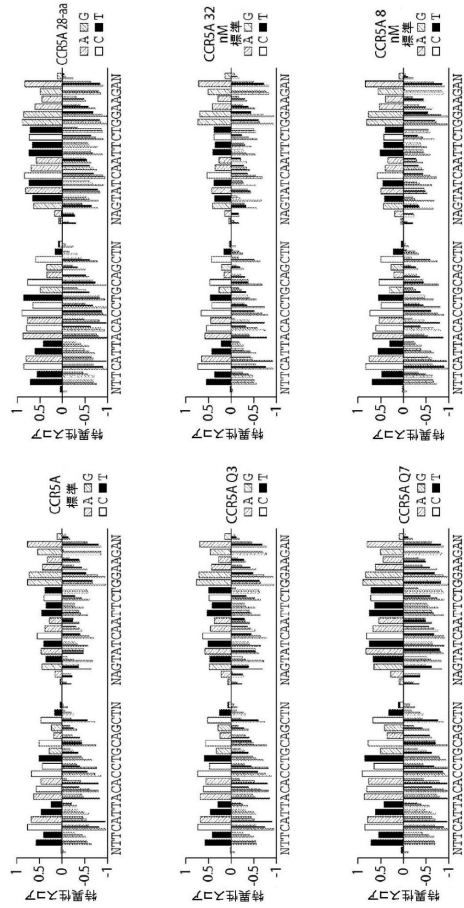


図9A

【図9-2】

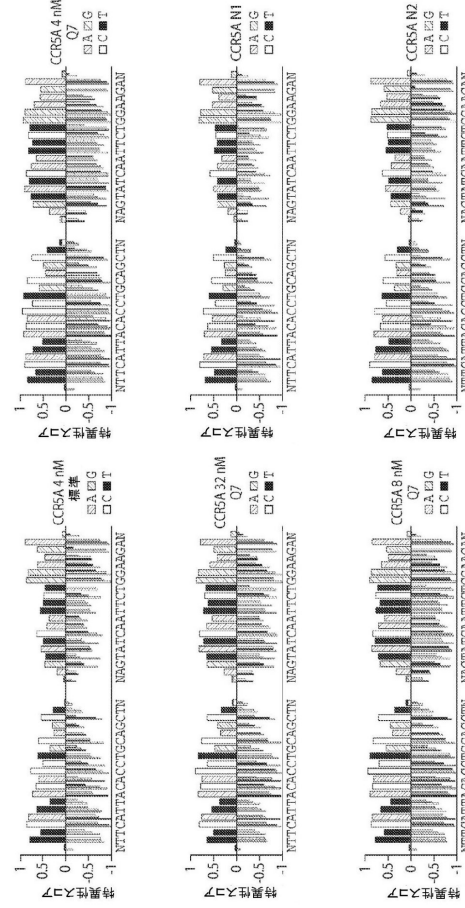


図9B

【図9-3】

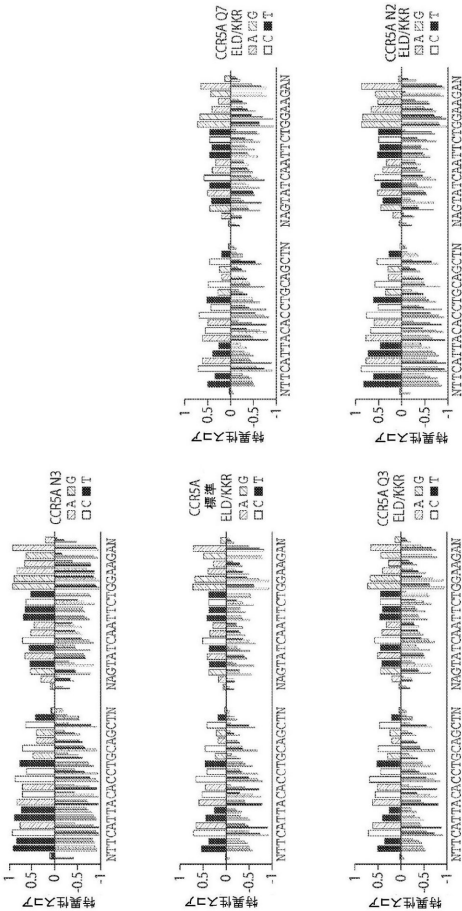


図9C

【図10-1】

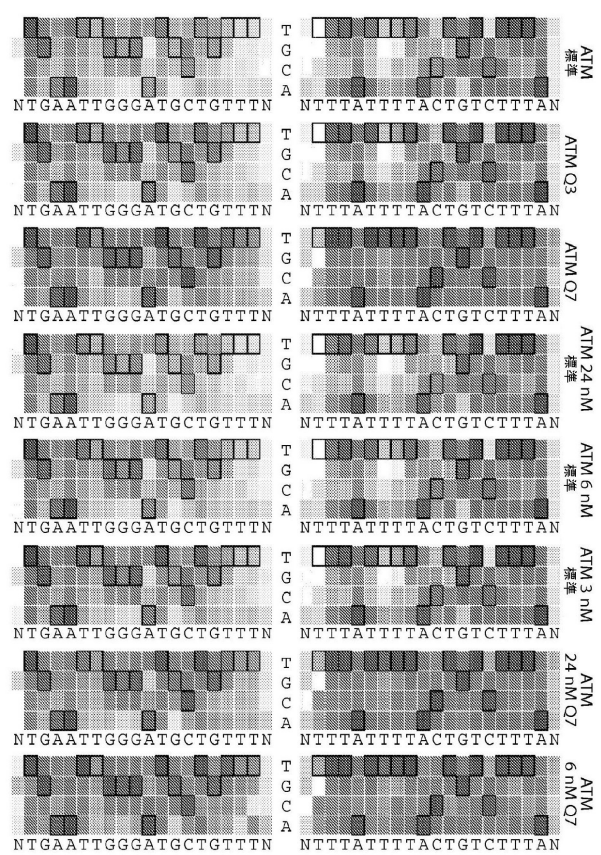


図10A

【図 10 - 2】

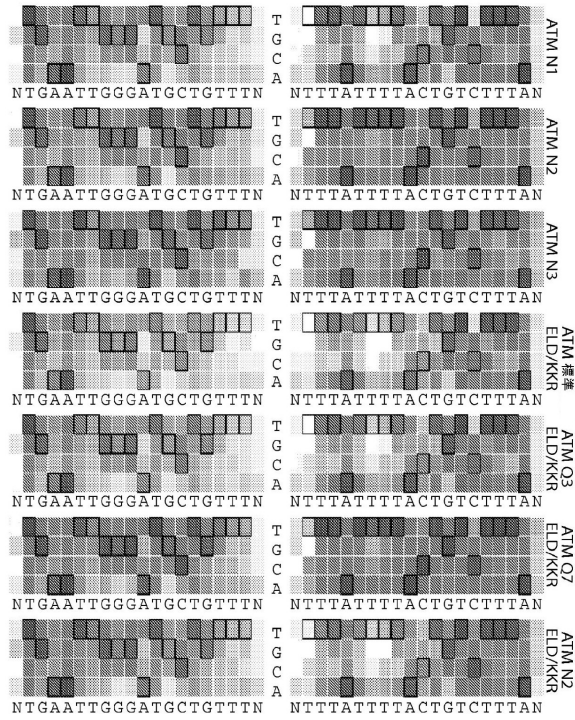


図10B

【図 11 - 1】

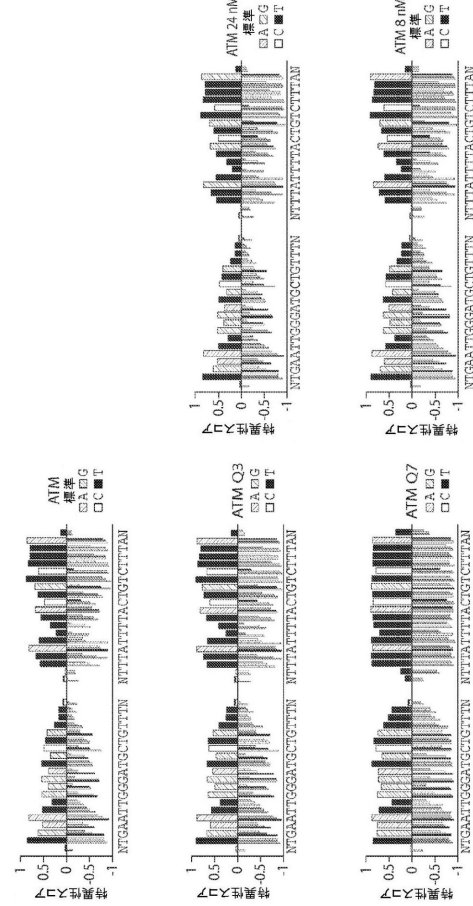


図11A

【図 11 - 2】

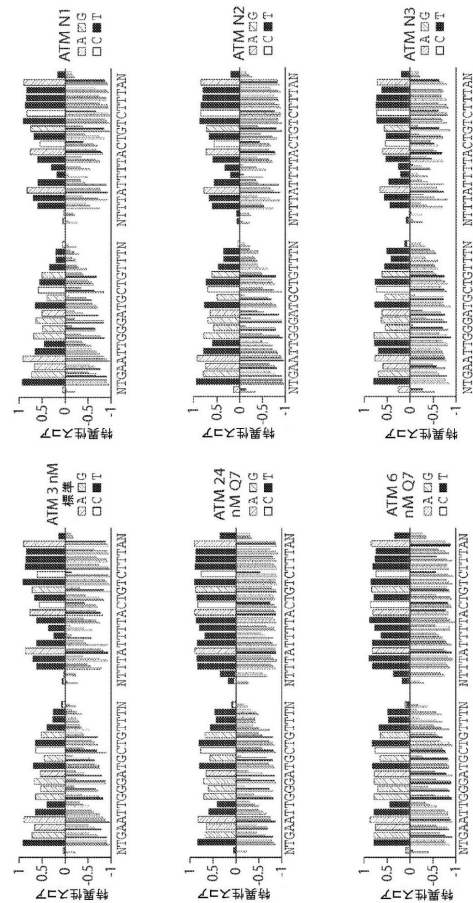


図11B

【図 11 - 3】

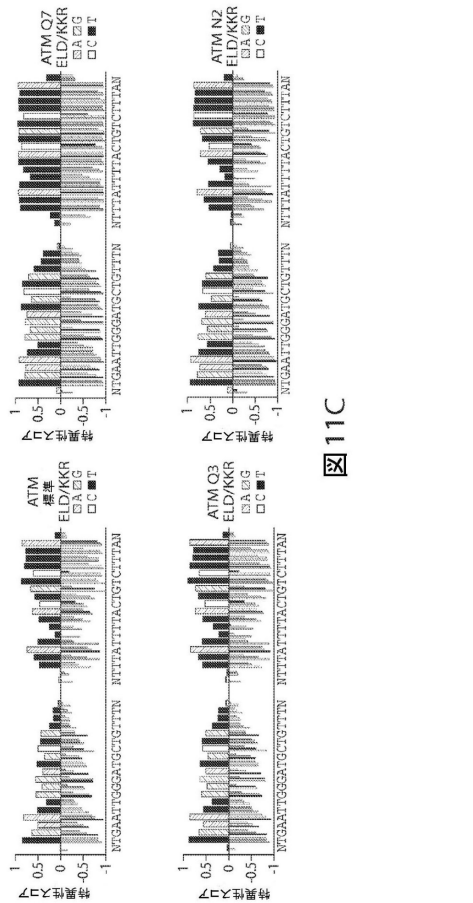


図11C

【図 12】

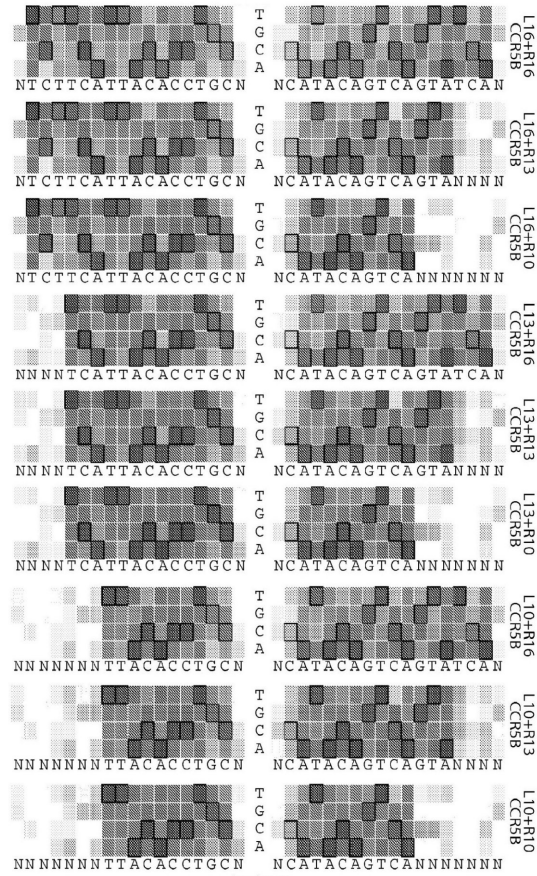


図 12

【図 13 - 2】

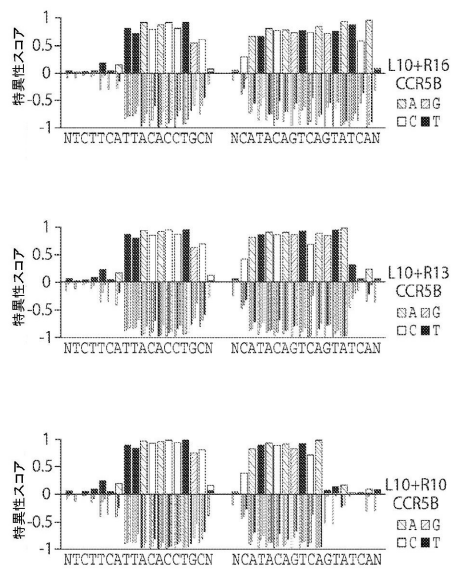


図 13B

【図 13 - 1】

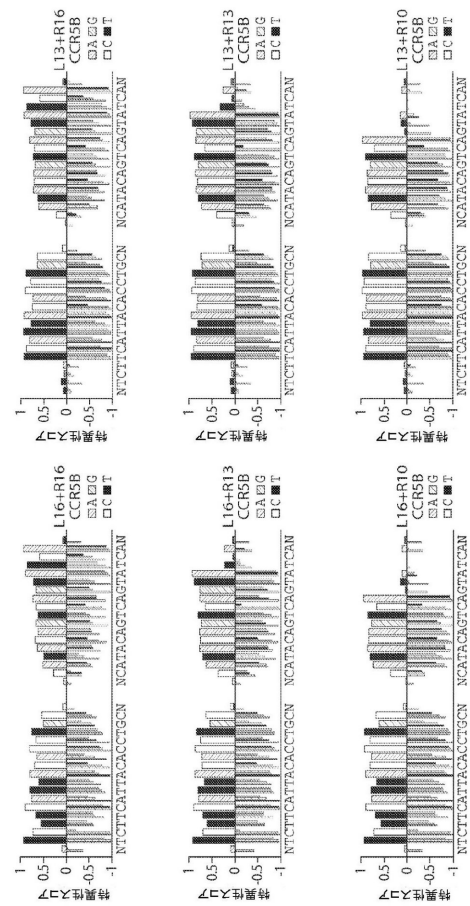


図 13A

【図 14】

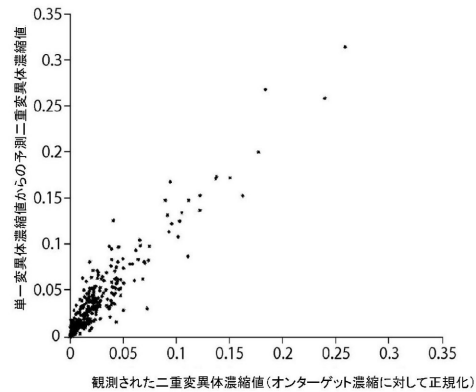


図 14A

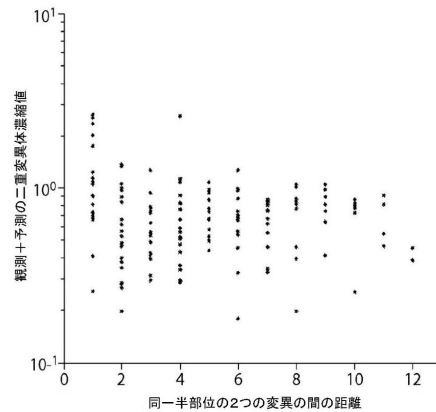
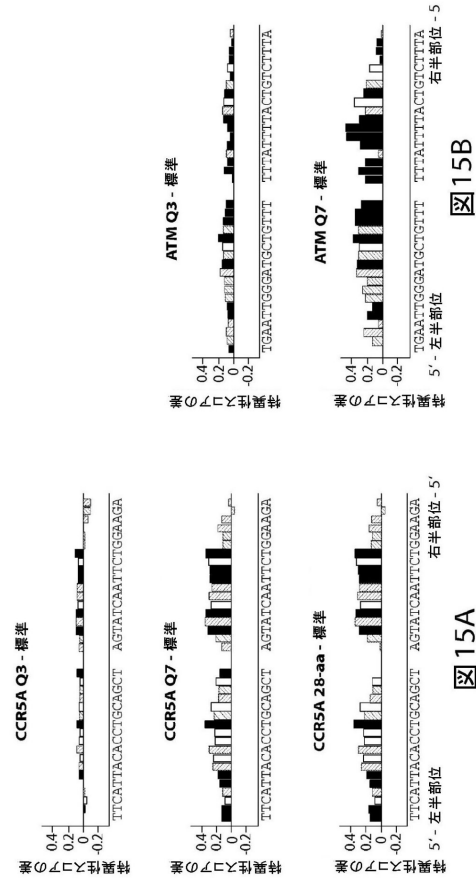
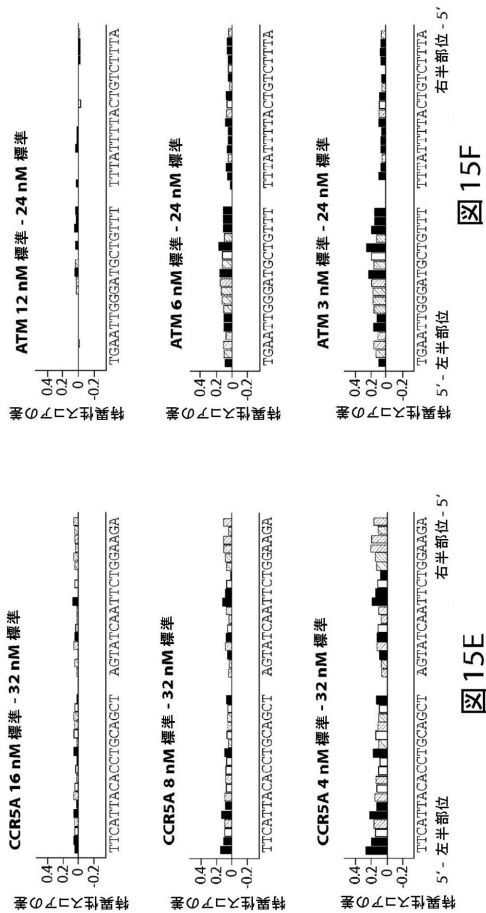


図 14B

【 図 1 5 - 1 】



【 図 1 5 - 3 】



15B

15A

图 15F

图 15E

【 図 1 5 - 2 】

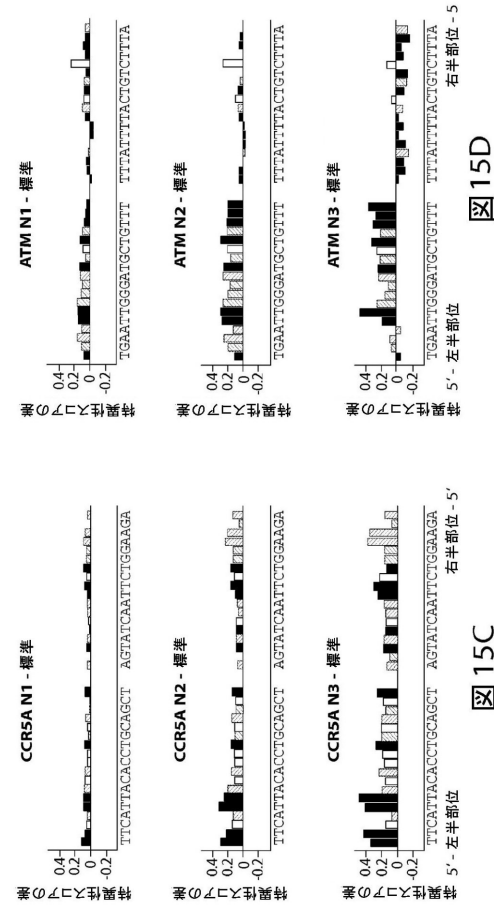


图 15D

Figure 15C

【 図 1 6 - 1 】

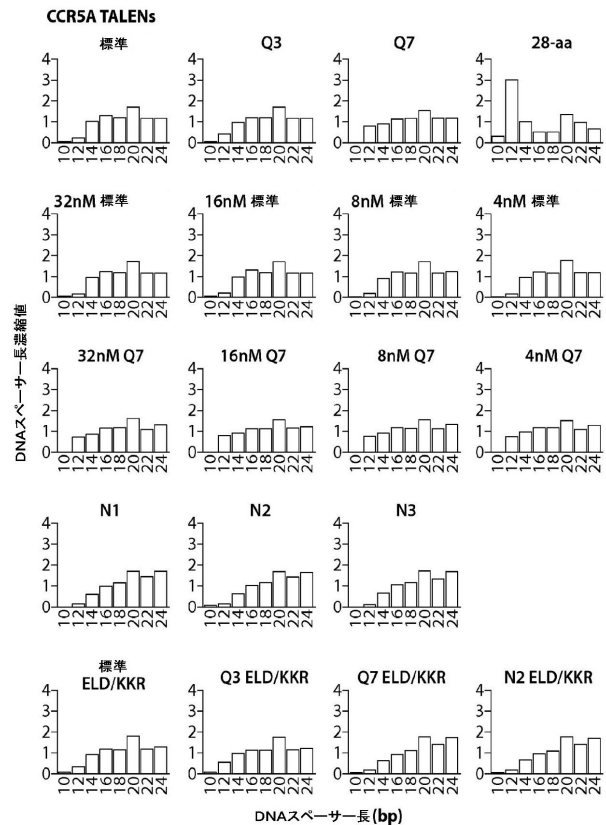


図 16A

【図16-2】

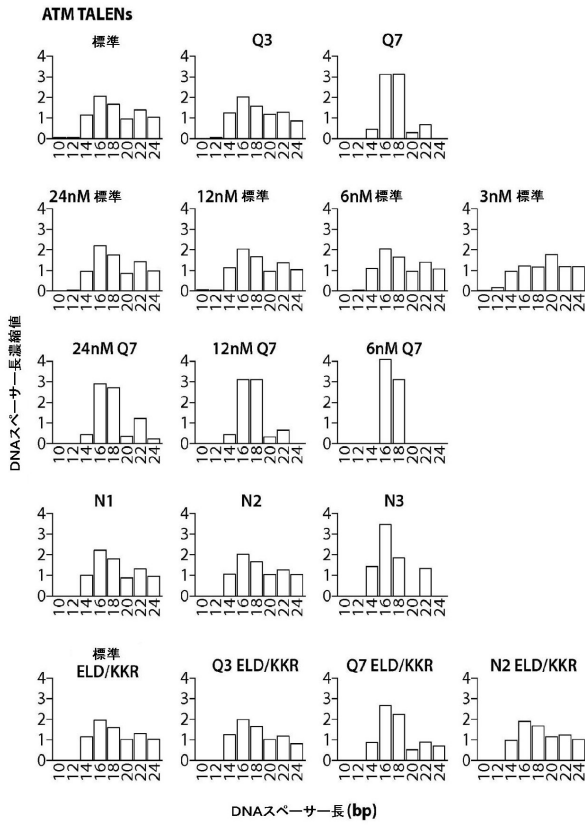


図16B

【図17-1】

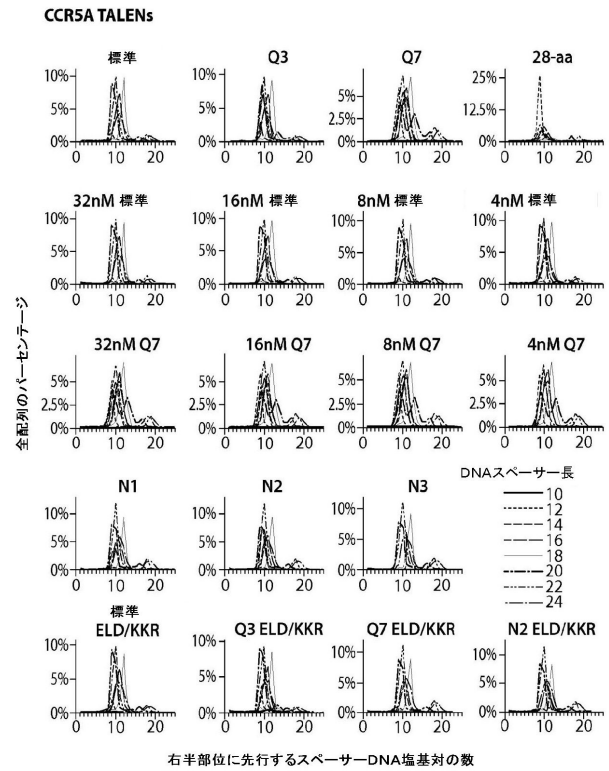


図17A

【図17-2】

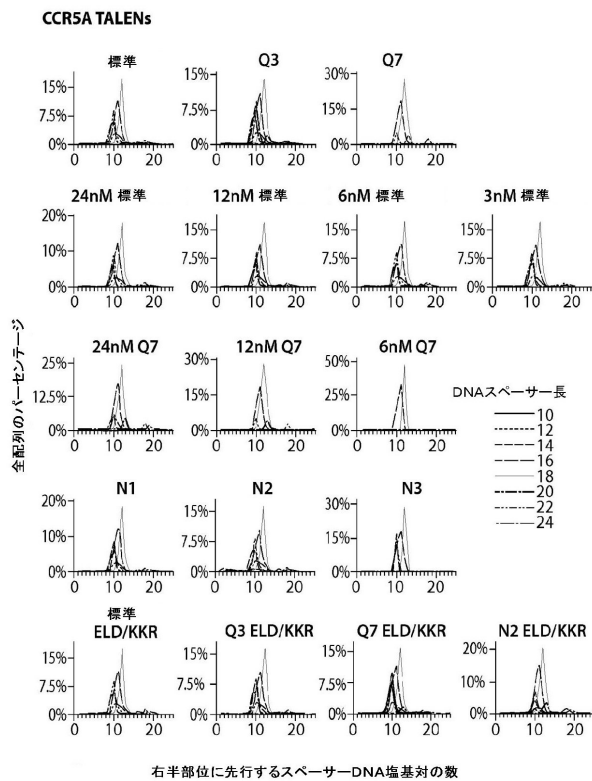


図17B

【図18-1】

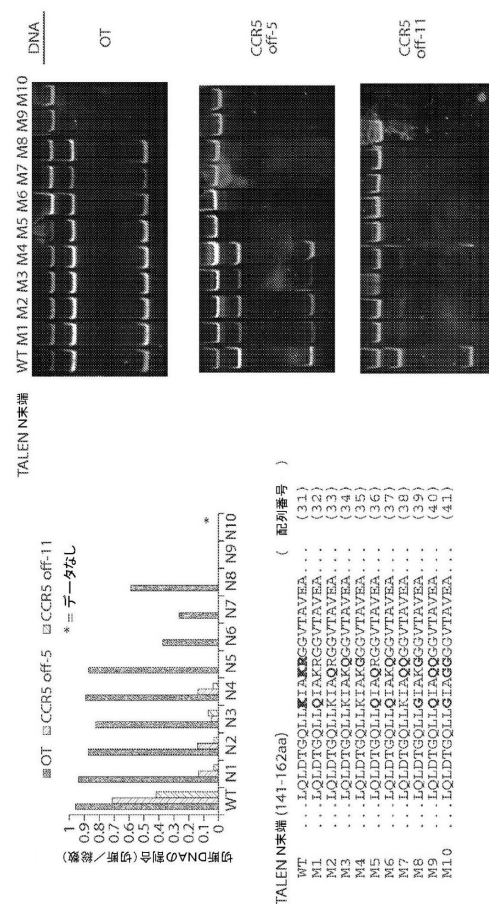


図18-1

【図 18 - 2】

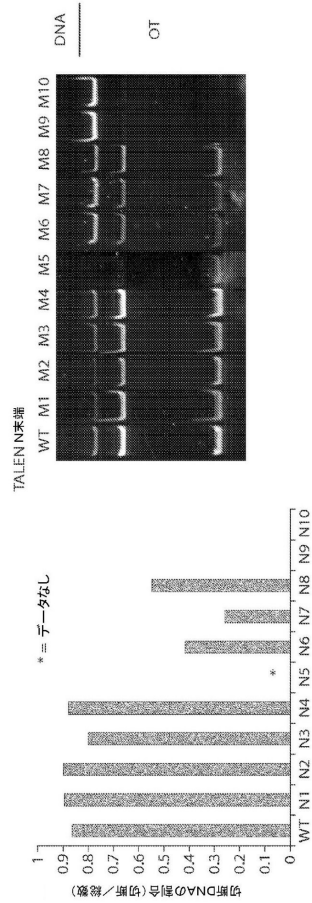


図 18-2

【図 19】

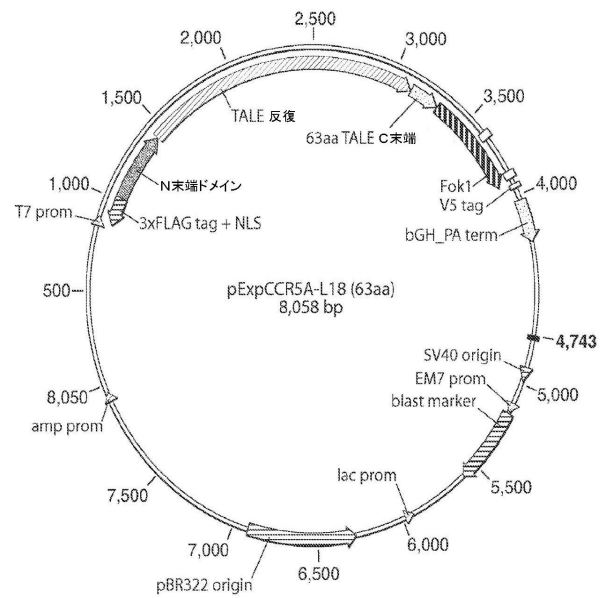


図 19

【配列表】

0006633524000001.app

フロントページの続き

(51)Int.Cl.		F I	
C 1 2 P 21/02 (2006.01)		C 1 2 P 21/02	C
A 6 1 K 38/16 (2006.01)		A 6 1 K 38/16	
A 6 1 P 35/00 (2006.01)		A 6 1 P 35/00	
A 6 1 P 31/18 (2006.01)		A 6 1 P 31/18	
A 6 1 P 43/00 (2006.01)		A 6 1 P 43/00	1 0 5

前置審査

(72)発明者 ギリンジャー, ジョン, ポール
 アメリカ合衆国 コロラド州 8 1 4 3 2、リッジウェイ、ベアー カブ ドライブ 3 4 0

(72)発明者 パッタナヤク, ヴィクラム
 アメリカ合衆国 マサチューセッツ州 0 2 1 3 8、ケンブリッジ、ホールデン グリーン 1 2
 3

審査官 植原 克典

(56)参考文献 国際公開第 2 0 1 2 / 1 3 8 9 2 7 (W O , A 1)
 国際公開第 2 0 1 3 / 0 4 7 8 4 4 (W O , A 1)
 特開 2 0 1 1 - 0 8 1 0 1 1 (J P , A)
 Nature Biotechnology, 2 0 1 1 年, Vol. 29, No. 2, pp. 143-148, 1-20
 ウイルス, 2 0 0 7 年, Vol. 57, No. 2, pp. 133-140
 PNAS, 2 0 0 9 年, Vol. 106, No. 15, pp. 6111-6116

(58)調査した分野(Int.Cl., D B 名)
 C 1 2 N 1 5 / 0 0 - 1 5 / 9 0
 C A p l u s / M E D L I N E / E M B A S E / B I O S I S (S T N)
 U n i P r o t / G e n e S e q