

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号

特許第7095031号

(P7095031)

(45)発行日 令和4年7月4日(2022.7.4)

(24)登録日 令和4年6月24日(2022.6.24)

(51)国際特許分類

F I

C 1 2 N 15/09 (2006.01)

C 1 2 N 15/09

Z Z N A

C 1 2 Q 1/6855(2018.01)

C 1 2 Q 1/6855

Z

C 1 2 Q 1/686(2018.01)

C 1 2 Q 1/686

Z

請求項の数 25 外国語出願 (全67頁)

(21)出願番号 特願2020-126325(P2020-126325)
 (22)出願日 令和2年7月27日(2020.7.27)
 (62)分割の表示 特願2016-575174(P2016-575174)
)の分割
 原出願日 平成27年6月23日(2015.6.23)
 (65)公開番号 特開2020-191885(P2020-191885)
 A)
 (43)公開日 令和2年12月3日(2020.12.3)
 審査請求日 令和2年8月19日(2020.8.19)
 (31)優先権主張番号 62/015,911
 (32)優先日 平成26年6月23日(2014.6.23)
 (33)優先権主張国・地域又は機関
 米国(US)
 (31)優先権主張番号 62/077,844
 (32)優先日 平成26年11月10日(2014.11.10)
 最終頁に続く

(73)特許権者 592017633
 ザ ジェネラル ホスピタル コーポレイ
 ション
 アメリカ合衆国 マサチューセッツ州 ボ
 ストン フルーツ ストリート 5 5
 (74)代理人 100092783
 弁理士 小林 浩
 (74)代理人 100120134
 弁理士 大森 規雄
 (74)代理人 100196966
 弁理士 植田 渉
 (74)代理人 100104282
 弁理士 鈴木 康仁
 (72)発明者 ジョン, ジェイ. キース
 アメリカ合衆国 マサチューセッツ州,
 最終頁に続く

(54)【発明の名称】 シークエンシングによって評価されるゲノムワイドでバイアスのないDSBの同定 (G U
 I D E - S e q)

(57)【特許請求の範囲】

【請求項1】

細胞のゲノムDNA (g DNA) 内の二本鎖切断 (D S B) を検出するための方法であっ
 て、
 前記細胞を平滑末端の二本鎖オリゴデオキシヌクレオチド (d s O D N) と接触させるス
 テップであって、前記 d s O D N の両鎖が前記細胞のゲノムにオルソゴナルであり、前記
 d s O D N の 5 ' 末端がリン酸化されており、さらに (a) ホスホロチオエート結合が両方
 の 3 ' 末端に存在するか、または (b) ホスホロチオエート結合が両方の 3 ' 末端および両
 方の 5 ' 末端に存在する、ステップ、
 前記細胞が D S B を修復するのに十分な条件で前記細胞を維持するステップ、1つまたは
 複数の D S B において d s O D N を組み込むステップ、
 組み込まれた d s O D N を含む g DNA の一部を増幅するステップ、並びに
 前記 g DNA の増幅された部分をシークエンスするステップ
 を含み、それによって前記細胞の前記 g DNA 内の D S B を検出する、方法。

【請求項2】

g DNA の一部を増幅するステップが、
 前記 DNA を断片化すること、
 前記細胞からの断片化されたゲノム DNA の末端をユニバーサルアダプターでライゲーシ
 ョンすること、および
 ライゲーションされた DNA に対してポリメラーゼ連鎖反応 (P C R) を行うこと、

を含む、請求項 1 に記載の方法。

【請求項 3】

前記細胞が哺乳動物細胞である、請求項 1 又は 2 に記載の方法。

【請求項 4】

前記 dsODN が 30 ~ 35 nt の長さである、請求項 1 ~ 3 のいずれか一項に記載の方法。

【請求項 5】

前記 dsODN が、5'末端でリン酸化されており、3'末端でホスホロチオエート化されている、請求項 1 ~ 4 のいずれか一項に記載の方法。

【請求項 6】

前記 dsODN が、無作為化された DNA バーコードを含む、請求項 1 ~ 5 のいずれか一項に記載の方法。

【請求項 7】

請求項 1 ~ 6 のいずれか一項に記載の方法であって、
前記 gDNA を断片に切断するステップ、および
前記断片を、末端修復、a テーリング、および単一テールのシーケンシングアダプターのライゲーションによってシーケンシング用に調製するステップ
を含む方法。

【請求項 8】

前記 dsODN が 15 ~ 50 nt の間の長さである、請求項 1 ~ 3 および 5 ~ 7 のいずれか一項に記載の方法。

【請求項 9】

細胞の gDNA 内の二本鎖切断 (DSB) を検出するための方法であって、
前記細胞を平滑末端の二本鎖オリゴデオキシヌクレオチド (dsODN) と接触させるステップであって、前記 dsODN の両鎖が前記細胞のゲノムにオルソゴナルであり、前記 dsODN の 5'末端がリン酸化されており、さらに (a) ホスホロチオエート結合が両方の 3'末端に存在するか、または (b) ホスホロチオエート結合が両方の 3'末端および両方の 5'末端に存在する、ステップ、
DSB がゲノム DNA 内で生じるのに十分な条件、かつ前記細胞が DSB を修復するのに十分な条件で前記細胞を維持するステップ、1 つまたは複数の DSB において dsODN を組み込むステップ、
組み込まれた dsODN を含む gDNA の一部を増幅するステップ、並びに
前記 gDNA の増幅された部分をシーケンスするステップ
を含み、それによって前記細胞の前記 gDNA 内の DSB を検出する、方法。

【請求項 10】

gDNA の一部を増幅するステップが、
前記 DNA を断片化すること、
前記細胞からの断片化されたゲノム DNA の末端をユニバーサルアダプターでライゲーションすること、および
ライゲーションされた DNA に対してポリメラーゼ連鎖反応 (PCR) を行うこと、
を含む、請求項 9 に記載の方法。

【請求項 11】

前記細胞が哺乳動物細胞である、請求項 9 又は 10 に記載の方法。

【請求項 12】

前記 dsODN が 30 ~ 35 nt の長さである、請求項 9 ~ 11 のいずれか一項に記載の方法。

【請求項 13】

前記 dsODN が 15 ~ 50 nt の間の長さである、請求項 9 ~ 11 のいずれか一項に記載の方法。

【請求項 14】

10

20

30

40

50

前記 dsODN が、5'末端でリン酸化されており、3'末端でホスホロチオエート化されている、請求項 9 ~ 12 のいずれか一項に記載の方法。

【請求項 15】

前記 dsODN が、無作為化された DNA バーコードを含む、請求項 9 ~ 14 のいずれか一項に記載の方法。

【請求項 16】

請求項 9 ~ 15 のいずれか一項に記載の方法であって、

前記 gDNA を断片に切断するステップ、および

前記断片を、末端修復、a テーリング、および単一テールのシーケンシングアダプターのライゲーションによってシーケンシング用に調製するステップ

を含む方法。

10

【請求項 17】

細胞のゲノム DNA (gDNA) 内の二本鎖切断 (DSB) に目的の二本鎖オリゴデオキシヌクレオチド (dsODN) を効率良く組み込むための方法であって、

前記細胞を平滑末端の dsODN と接触させるステップであって、前記 dsODN の両鎖が前記細胞のゲノムにオルソゴナルであり、前記 dsODN の 5'末端がリン酸化されており、さらに (a) ホスホロチオエート結合が両方の 3'末端に存在するか、または (b) 2

つのホスホロチオエート結合が両方の 3'末端および両方の 5'末端に存在する、ステップ、

前記細胞が DSB を修復するのに十分な条件で前記細胞を維持するステップ、および 1 つ

または複数の DSB において dsODN を組み込むステップ、

を含む、方法。

20

【請求項 18】

組み込まれた dsODN を含む gDNA の一部を増幅するステップ、および

前記 gDNA の増幅された部分をシーケンスするステップ

を含み、それによって前記細胞の前記 gDNA 内の DSB を検出する、請求項 17 に記載の方法

【請求項 19】

gDNA の一部を増幅するステップが、

前記 DNA を断片化すること、

前記細胞からの断片化されたゲノム DNA の末端をユニバーサルアダプターでライゲーションすること、および

ライゲーションされた DNA に対してポリメラーゼ連鎖反応 (PCR) を行うこと、

を含む、請求項 18 に記載の方法。

30

【請求項 20】

前記細胞が哺乳動物細胞である、請求項 17 ~ 19 のいずれか一項に記載の方法。

【請求項 21】

前記 dsODN が 30 ~ 35 nt の長さである、請求項 17 ~ 20 のいずれか一項に記載の方法。

【請求項 22】

前記 dsODN が 15 ~ 50 nt の間の長さである、請求項 17 ~ 20 のいずれか一項に記載の方法。

40

【請求項 23】

前記 dsODN が、5'末端でリン酸化されており、3'末端でホスホロチオエート化されている、請求項 17 ~ 22 のいずれか一項に記載の方法。

【請求項 24】

前記 dsODN が、無作為化された DNA バーコードを含む、請求項 17 ~ 23 のいずれか一項に記載の方法。

【請求項 25】

請求項 17 ~ 24 のいずれか一項に記載の方法であって、

前記 gDNA を断片に切断するステップ、および

50

前記断片を、末端修復、aテーリング、および単一テールのシークエンシングアダプターのライゲーションによってシークエンシング用に調製するステップを含む方法。

【発明の詳細な説明】

【技術分野】

【0001】

優先権の主張

本出願は、2014/06/23に出願された、米国仮特許出願第62/015,911号明細書；2014/11/10に出願された、第62/077,844号明細書；2014/11/12に出願された、第62/078,923号明細書；および2014/12/5に出願された、第62/088,223号明細書の利益を主張する。前述のものの全内容は、参照により本明細書によって組み込まれている。

10

【0002】

連邦政府による資金提供を受けた研究または開発

本発明は、National Institutes of Healthによって与えられた助成金第DP1GM105378号の下で、政府の支援を受けて行われた。政府は、本発明において一定の権利を有する。

【0003】

技術分野

生細胞において改変ヌクレアーゼの切断部位の位置を同定するための、高感度なバイアスのないゲノムワイドな方法が提供される。

20

【背景技術】

【0004】

ヒトの薬の長い間保持されてきた目標は、遺伝性の遺伝疾患を治療することである。ゲノム編集は、疾患を治癒させるまたは予防するために内在性遺伝子内の変異を直接に修正するという強力な概念を包含する。このアプローチの新しい例は、HIVのコレセプターである、CCR5を破壊するように改変されたジンクフィンガーヌクレアーゼ(ZFN)治療薬の臨床試験である(1)。このex vivo自己細胞療法アプローチは、CCR5におけるホモ変異を有する個体からの骨髓細胞を移植された「ベルリンの患者」、Timothy BrownにおけるHIVの治療の成功を再現することを試みるものである。もう1つの最近の例は、6ヶ月齢の対象に由来する造血幹細胞におけるZFNでの遺伝子ターゲティングによるX連鎖重症複合免疫不全症の修正である(2)。

30

【0005】

以下の4つの主なクラスの改変ヌクレアーゼがある：1)メガヌクレアーゼ、2)ジンクフィンガーヌクレアーゼ、3)転写活性化因子エフェクター様ヌクレアーゼ(TALEN)、および4)クラスター化された規則的な間隔の短鎖反復回文配列(CRISPR) Cas RNA誘導型ヌクレアーゼ(RGN)。

【発明の概要】

【0006】

しかし、これらの新しい治療および研究のツールの採用は、それらの特異性の実証に依存しうる。ヒトおよび他の真核細胞におけるオフターゲット作用を理解および同定することは、これらのヌクレアーゼが研究および治療的適用のために広範に使用される場合には決定的に不可欠であろう。

40

【0007】

GUIDE-Seqは、改変ヌクレアーゼによって誘発される変異、例えば、オフターゲット変異を検出するための、バイアスのないゲノムワイドかつ高感度な方法を可能にする。したがって、該方法は、哺乳動物生細胞においてゲノムワイドな規模で変異を評価するための最も包括的でバイアスのない方法を可能にする。該方法は、ヌクレアーゼで誘発されたDSBの中にdsODNが効率良く捕捉されうるいかなる細胞型において利用されてもよい。

50

【 0 0 0 8 】

したがって、一態様では、本発明は、細胞のゲノムDNA内の、二本鎖切断(DSB)、例えば、外来性の改変ヌクレアーゼによって誘発された、例えば、オフターゲットDSBを検出するための方法を提供する。該方法は、細胞を二本鎖オリゴデオキシヌクレオチド(dsODN)と接触させるステップであって、好ましくは、dsODNが15および75ntの間の長さ、例えば、15~50nt、50~75nt、30~35nt、60~65nt、または50~65ntの長さであり、dsODNの両鎖が細胞のゲノムにオルソログスであり、好ましくは、dsODNの5'末端がリン酸化されており、さらにまた好ましくは、ホスホロチオエート結合が両方の3'末端に存在する、または2つのホスホロチオエート結合が両方の3'末端および両方の5'末端に存在する、ステップ、細胞内で外来性の改変ヌクレアーゼを、ヌクレアーゼが細胞のゲノムDNA内にDSBを誘発するのに十分な時間、かつ細胞がDSBを修復して、1つまたは複数のDSBにおいてdsODNを組み込むのに十分な時間、発現または活性化させるステップ、組み込まれたdsODNを含むゲノムDNAの一部を増幅するステップ、およびゲノムDNAを増幅された部分をシークエンスするステップを含み、それによって細胞のゲノムDNA内のDSBを検出する。

10

【 0 0 0 9 】

一部の実施形態では、ゲノムDNAの一部を増幅するステップは、例えば切断によって、DNAを断片化すること、細胞からの断片化されたゲノムDNAの末端をユニバーサルアダプターでライゲーションすること、組み込まれたdsODNと相補的なプライマー(プライマーA)およびユニバーサルアダプターと相補的なプライマー(プライマーB)を用いて、ライゲーションされたDNAに対して1回目のポリメラーゼ連鎖反応(PCR)を行うこと、次いで、プライマーAと相補的な3'ネステッドプライマー(プライマーC)、プライマーBと相補的な3'ネステッドプライマー(プライマーD)、およびプライマーDと相補的なプライマー(プライマーE)を使用して2回目のPCRを行うことを含む。一部の実施形態では、プライマーEは、精製または結合配列、例えば、フローセル結合配列、および同定配列、例えば、バーコードまたは無作為の分子インデックスのうちの一つまたは複数を含む。

20

30

【 0 0 1 0 】

一部の実施形態では、改変ヌクレアーゼは、メガヌクレアーゼ、ジンクフィンガーヌクレアーゼ、転写活性化因子エフェクター様ヌクレアーゼ(TALEN)、およびクラスター化された規則的な間隔の短鎖反復回文配列(CRISPR)/CasRNA誘導型ヌクレアーゼ(CRISPR/CasRGN)からなる群から選択される。

【 0 0 1 1 】

別の態様では、本発明は、複数のガイドRNAのうちの一つが最も特異的であるか、すなわち、最も少ないオフターゲットDSBを誘発するかを決定するための方法を提供する。該方法は、第1の細胞の集団を第1のガイドRNAおよび二本鎖オリゴデオキシヌクレオチド(dsODN)と接触させるステップであって、好ましくは、dsODNが15および75ntの間の長さ、例えば、15~50nt、50~75nt、60~65nt、30~35ntまたは50~65ntの長さであり、dsODNの両鎖が細胞のゲノムにオルソログスであり、好ましくは、dsODNの5'末端がリン酸化されており、さらにまた好ましくは、ホスホロチオエート結合が両方の3'末端に存在する、または2つのホスホロチオエート結合が両方の3'末端および両方の5'末端に存在する、ステップ、第1の細胞の集団内で外来性のCas9改変ヌクレアーゼを、ヌクレアーゼが細胞のゲノムDNA内にDSBを誘発するのに十分な時間、かつ細胞がDSBを修復して、1つまたは複数のDSBにおいてdsODNを組み込むのに十分な時間、発現または活性化させるステップ、

40

50

組み込まれた dsODNを含む、第1の細胞の集団からのゲノムDNAの一部を増幅するステップ、および

第1の細胞の集団からのゲノムDNAを増幅された部分をシーケンスするステップ、dsODNが第1の細胞の集団のゲノムDNA内に組み込まれた部位の数を決定するステップ、

第2の細胞の集団を第2のガイドRNAおよび二本鎖オリゴデオキシヌクレオチド(dsODN)と接触させるステップであって、好ましくは、dsODNが15および75ntの間の長さ、例えば、15~50nt、50~75nt、30~35nt、60~65nt、または50~65ntの長さであり、dsODNの両鎖が細胞のゲノムにオルソログスであり、好ましくは、dsODNの5'末端がリン酸化されており、さらにまた好ましくは、2つのホスホロチオエート結合が両方の3'末端および両方の5'末端に存在する、ステップ、

第2の細胞の集団内で外来性のCas9改変ヌクレアーゼを、ヌクレアーゼが第2の細胞の集団のゲノムDNA内にDSBを誘発するのに十分な時間、細胞がDSBを修復して、1つまたは複数のDSBにおいてdsODNを組み込むのに十分な時間、発現または活性化させるステップ、

組み込まれたdsODNを含む、第2の細胞の集団からのゲノムDNAの一部を増幅するステップ、および

第2の細胞の集団からのゲノムDNAを増幅された部分をシーケンスするステップ、dsODNが第2の細胞の集団のゲノムDNA内に組み込まれた部位の数を決定するステップ、

dsODNが第1の細胞の集団のゲノムDNA内に組み込まれた部位の数を、dsODNが第2の細胞の集団のゲノムDNA内に組み込まれた部位の数と比較するステップを含み、より少ない(オフターゲット)部位で組み込まれたdsODNがより特異的である。該方法は、第3、第4、第5、第6、またはそれより多種の細胞の集団について繰り返されてもよい。「より少ない」オフターゲット部位とは、より少ない数のDSB部位および/または(1つまたは複数の)個別の部位におけるDSBの発生の頻度の低下の両方を含みうる。

【0012】

本明細書においてさらに、本明細書に記載の末端保護されたdsODNの使用によって目的の短鎖dsDNAをDSBの部位の中に効率良く組み込むための方法が提供される。

【0013】

一部の実施形態では、細胞は、哺乳動物細胞である。

【0014】

一部の実施形態では、改変ヌクレアーゼはCas9ヌクレアーゼであり、方法は、Cas9ヌクレアーゼをゲノム内の標的配列に導くガイドRNA、例えば、単一のガイドまたはtracrRNA/crRNA対を細胞内で発現させるステップも含む。

【0015】

一部の実施形態では、dsODNはビオチン化され、例えば、共有結合でdsODNに付加されたビオチンを含む、かつ/または無作為化されたDNAバーコードまたはCreもしくはLox部位を含む。dsODNがビオチン化される、上記の請求項のいずれかの方法。

【0016】

一部の実施形態では、本明細書に記載の方法は、ゲノムgDNAを断片に切断するステップ、およびdsODNを含む断片をビオチンへの結合によって単離するステップを含む。

【0017】

一部の実施形態では、dsODNは、平滑末端であるまたは5'末端に1、2、3、もしくは4ntのオーバーハンギングを有する、5'末端でリン酸化されている、および/または3'末端でホスホロチオエート化されている。

10

20

30

40

50

【 0 0 1 8 】

一部の実施形態では、dsODNは、平滑末端であり、5'末端でリン酸化されており、3'末端でホスホロチオエート化されている。

【 0 0 1 9 】

一部の実施形態では、dsODNは、無作為化されたDNAバーコード、Lox認識部位、制限酵素認識部位、および/またはタグ配列を含む。

【 0 0 2 0 】

一部の実施形態では、方法は、ゲノムgDNAを断片に切断するステップ、および断片を、末端修復/aテーリング/シークエンシングアダプター、例えば、単一テールのシークエンシングアダプターのライゲーションによってシークエンシング、例えば、ハイスループットシークエンシング用に調製するステップを含む。

10

【 0 0 2 1 】

一部の実施形態では、DSBは、(例えば、脆弱な部位での)バックグラウンドのゲノムのDSBまたは鍵となる細胞タンパク質の小分子阻害因子によって引き起こされるDSBである。

【 0 0 2 2 】

他に定めがない限り、本明細書中で使用されるすべての技術的および科学的な用語は、本発明が属する当技術分野の当業者によって一般的に理解されるのと同じ意味を有する。方法および材料は、本発明における使用のために本明細書に記載されている；他の、当技術分野において知られている好適な方法および材料も使用されうる。材料、方法、および例は、単に例示的なものであって限定することを意図するものではない。本明細書中で言及されているすべての刊行物、特許出願、特許、配列、データベースエントリー、および他の参照は、参照によりその全体が組み込まれている。矛盾する場合には、定義を含めて、本明細書が優先されることになる。

20

【 0 0 2 3 】

本発明の他の特徴および利点は、以下の発明を実施するための形態および図面から、ならびに特許請求の範囲から明らかであろう。

【 図面の簡単な説明 】

【 0 0 2 4 】

【 図 1 】 図 1 A - B . C R I S P R - C a s ヌクレアーゼを介した dsODN 捕捉の最適化を示した図である。(a) 使用される短鎖オリゴヌクレオチドタグの配列が示されている。使用されるすべてのオリゴヌクレオチドは、5'リン酸化される。タグオリゴヌクレオチドは、RFLPによる組み込み頻度の評価を可能にする診断用のNdeI制限部位も含む。(b) 下のグラフは、RFLPによる短鎖dsODNの組み込み(%)を示す。5'および3'ホスホロチオエート結合の両方を有するdsODN(各セットにおける左側のバー)についての組み込み率が、5'ホスホロチオエート結合のみを有するdsODN(各セットにおける中央のバー)およびdsODNなしの対照(各セットにおける右側のバー)と比較される。

30

【 図 2 - 1 】 図 2 A . V E G F 部位 1 についての組み込みの特徴付けを示した図である。(a) RFLPアッセイは、VEGF部位1について示されており、Qiaxcelキャピラリー電気泳動機器上で解析されるように、NdeI制限部位を有するdsODNの好結果の組み込みを実証している。

40

【 図 2 - 2 】 図 2 B . V E G F 部位 1 についての組み込みの特徴付けを示した図である。(b) サンガーシークエンシングデータは、目的のVEGF部位1標的部位でのdsODN組み込みについて示されている。dsODN配列は、灰色で強調されている。VEGF A 部位 1 を標的としたガイドRNA/Cas9複合体によって認識される部位は太字で、隣接したプロトスペーサー隣接モチーフ(PAM)配列は下線で強調されている。この部位でCas9によって誘発される、予想される二本鎖切断の位置は、小さな黒色の矢印で示されている。

【 図 3 】 例示的なGUIDE-seq法の概要を示した図である。

50

【図4-1】図4A. 1GUIDE-Seq法によって発見されたCRISPR-Cas オフターゲット切断部位を示した図である。データは、4つの部位、VEGF部位1~3、およびEMX1について示されている。標的部位配列に対するミスマッチが強調されている。小さな中実の黒色の矢印は、目的のオンターゲット部位を示すために使用されているのに対して、小さな破線の矢印は、以前の研究(Fu et al., 2013)において検出されていた既知のオフターゲット部位に印をつけるために使用されている。

【図4-2】図4B. GUIDE-Seq法によって発見されたCRISPR-Cas オフターゲット切断部位を示した図である。データは、4つの部位、VEGF部位1~3、およびEMX1について示されている。標的部位配列に対するミスマッチが強調されている。小さな中実の黒色の矢印は、目的のオンターゲット部位を示すために使用されているのに対して、小さな破線の矢印は、以前の研究(Fu et al., 2013)において検出されていた既知のオフターゲット部位に印をつけるために使用されている。

10

【図4-3】図4C. GUIDE-Seq法によって発見されたCRISPR-Cas オフターゲット切断部位を示した図である。データは、4つの部位、VEGF部位1~3、およびEMX1について示されている。標的部位配列に対するミスマッチが強調されている。小さな中実の黒色の矢印は、目的のオンターゲット部位を示すために使用されているのに対して、小さな破線の矢印は、以前の研究(Fu et al., 2013)において検出されていた既知のオフターゲット部位に印をつけるために使用されている。

【図4-4】図4D. GUIDE-Seq法によって発見されたCRISPR-Cas オフターゲット切断部位を示した図である。データは、4つの部位、VEGF部位1~3、およびEMX1について示されている。標的部位配列に対するミスマッチが強調されている。小さな中実の黒色の矢印は、目的のオンターゲット部位を示すために使用されているのに対して、小さな破線の矢印は、以前の研究(Fu et al., 2013)において検出されていた既知のオフターゲット部位に印をつけるために使用されている。

20

【図4-5】図4E. GUIDE-Seq法によって発見されたCRISPR-Cas オフターゲット切断部位を示した図である。データは、4つの部位、VEGF部位1~3、およびEMX1について示されている。標的部位配列に対するミスマッチが強調されている。小さな中実の黒色の矢印は、目的のオンターゲット部位を示すために使用されているのに対して、小さな破線の矢印は、以前の研究(Fu et al., 2013)において検出されていた既知のオフターゲット部位に印をつけるために使用されている。

30

【図5-1】図5A-B. 例示的なGUIDE-Seq法の設計、最適化および適用を示した図である。(a)例示的なGUIDE-Seq法の図式的概要。(b)ヒト細胞におけるRGNで誘発されたDSBの中へのdsODN組み込みの最適化。RFLPアッセイによって測定されるような、異なる修飾オリゴヌクレオチドについての組み込みの割合が示されている。対照反応物は、RGNをコードしているプラスミドのみでトランスフェクトされた(すなわち、dsODNなし)。

【図5-2】図5C. 例示的なGUIDE-Seq法の設計、最適化および適用を示した図である。(c)ゲノム配列の読みのマッピングによりDSB部位の同定が可能とされる方法の図式的例示。二方向のマッピングの読みまたは同じ方向へのマッピングであるが異なるプライマーによって増幅された読みは、GUIDE-seqアッセイにおけるDSBのサインである。図1Aも参照されたい。

40

【図5-3】図5D-1. 例示的なGUIDE-Seq法の設計、最適化および適用を示した図である。(d1~d3)GUIDE-Seqに基づいた、RGNで誘発されたDSBの同定。ゲノムにマップされたGUIDE-Seqの読みの開始部位により、数塩基対以内までのDSBのマッピングが可能とされる。本発明者らがGUIDE-Seqによって評価した10種のRGNのオンターゲット部位についてのマップされた読みが示されている。すべての場合において、標的部位配列は、20bpのプロトスペーサー配列がx軸上の左側に、かつPAM配列が右側に示されている。すべての場合において、最も高いピークが、RGN切断事象の予想される部位である、NGG PAM配列の5'端の3から4bp以内に入る方法に留意されたい。

50

【図5-4】図5D-2. 例示的なGUIDE-Seq法の設計、最適化および適用を示した図である。(d1~d3)GUIDE-Seqに基づいた、RGNで誘発されたDSBの同定。ゲノムにマップされたGUIDE-Seqの読みの開始部位により、数塩基対以内までのDSBのマッピングが可能とされる。本発明者らがGUIDE-Seqによって評価した10種のRGNのオンターゲット部位についてのマップされた読みが示されている。すべての場合において、標的部配列は、20bpのプロトスペーサー配列がx軸上の左側に、かつPAM配列が右側に示されている。すべての場合において、最も高いピークが、RGN切断事象の予想される部位である、NGG PAM配列の5'端の3から4bp以内に入る方法に留意されたい。

【図5-5】図5D-3. 例示的なGUIDE-Seq法の設計、最適化および適用を示した図である。(d1~d3)GUIDE-Seqに基づいた、RGNで誘発されたDSBの同定。ゲノムにマップされたGUIDE-Seqの読みの開始部位により、数塩基対以内までのDSBのマッピングが可能とされる。本発明者らがGUIDE-Seqによって評価した10種のRGNのオンターゲット部位についてのマップされた読みが示されている。すべての場合において、標的部配列は、20bpのプロトスペーサー配列がx軸上の左側に、かつPAM配列が右側に示されている。すべての場合において、最も高いピークが、RGN切断事象の予想される部位である、NGG PAM配列の5'端の3から4bp以内に入る方法に留意されたい。

【図5-6】図5E-F. 例示的なGUIDE-Seq法の設計、最適化および適用を示した図である。(e)本研究において解析された10種のRGNについてのGUIDE-Seqによって同定された既知および新規のオフターゲット切断部位の数。4種のRGNについてのすべての既知のオフターゲット切断は、GUIDE-Seqによって同定された。(f)本報告の10種のRGNについてのGUIDE-Seqによって検出されたオフターゲット部位の総数に対するヒトゲノム(y軸)に直交性のオンターゲット部位の散乱プロット。直交性は、オンターゲット部位に対して相対的に1から6つのミスマッチを有するヒトゲノム内の部位の総数として算出された。

【図5-7】図5G. 例示的なGUIDE-Seq法の設計、最適化および適用を示した図である。(g)本報告の10種のRGNについてのGUIDE-Seqによって検出されるオフターゲット部位の総数に対するオンターゲット部位GC含量(y軸)の散乱プロット。

【図5-8】図5H-I. 例示的なGUIDE-Seq法の設計、最適化および適用を示した図である。(h)EMX1を標的とするRGNについてのCRISPR/Cas9オンターゲットおよびオフターゲット部位の染色体イデオグラム。残りのRGNについてのさらなるイデオグラムは、図13の中で見出すことができる。(i)本研究において調査された10種のRGNについてのGUIDE-Seqによって同定されたオフターゲット切断部位のゲノム位置。

【図6-1】図6A. 10種のRGNについてのGUIDE-Seqによって同定されたオフターゲット部位の配列を示した図である。各RGNについて、目的の標的配列が一番上の行に示され、切断された部位が下に示され、オンターゲット部位に対するミスマッチが示されかつ色で強調されている。GUIDE-Seqシーケンシングの読み数は、それぞれの部位の右側に示されている。オンターゲット部位は四角形でかつ既知のオフターゲット部位はひし形で印がつけられている。データは、以下の部位を標的とするRGNについて示されている:(a)VEGFA部位1。オフターゲット部位は、RNF2部位を標的とするRGNでは見出されなかった。

【図6-2】図6B. 10種のRGNについてのGUIDE-Seqによって同定されたオフターゲット部位の配列を示した図である。各RGNについて、目的の標的配列が一番上の行に示され、切断された部位が下に示され、オンターゲット部位に対するミスマッチが示されかつ色で強調されている。GUIDE-Seqシーケンシングの読み数は、それぞれの部位の右側に示されている。オンターゲット部位は四角形でかつ既知のオフターゲット部位はひし形で印がつけられている。データは、以下の部位を標的とするRGNに

10

20

30

40

50

ついて示されている：(b) VEGFA 部位 2。オフターゲット部位は、RNF2 部位を標的とする RGN では見出されなかった。

【図 6 - 3】図 6 C - E . 10 種の RGN についての GUIDE - Seq によって同定されたオフターゲット部位の配列を示した図である。各 RGN について、目的の標的配列が一番上の行に示され、切断された部位が下に示され、オンターゲット部位に対するミスマッチが示されかつ色で強調されている。GUIDE - Seq シークエンシングの読み数は、それぞれの部位の右側に示されている。オンターゲット部位は四角形でかつ既知のオフターゲット部位はひし形で印がつけられている。データは、以下の部位を標的とする RGN について示されている：(c) VEGFA 部位 3、(d) EMX1、(e) FANCF。オフターゲット部位は、RNF2 部位を標的とする RGN では見出されなかった。

10

【図 6 - 4】図 6 F - J . 10 種の RGN についての GUIDE - Seq によって同定されたオフターゲット部位の配列を示した図である。各 RGN について、目的の標的配列が一番上の行に示され、切断された部位が下に示され、オンターゲット部位に対するミスマッチが示されかつ色で強調されている。GUIDE - Seq シークエンシングの読み数は、それぞれの部位の右側に示されている。オンターゲット部位は四角形でかつ既知のオフターゲット部位はひし形で印がつけられている。データは、以下の部位を標的とする RGN について示されている：(f) HEK293 部位 1、(g) HEK293 部位 2、(h) HEK293 部位 3、(i) HEK293 部位 4、(j) RNF2。オフターゲット部位は、RNF2 部位を標的とする RGN では見出されなかった。

【図 7 - 1】図 7 A . GUIDE - Seq 切断部位が真正の RGN オフターゲット変異部位であることを示した図である。(a) GUIDE - Seq 切断部位におけるインデル変異を確認するために使用される、AMP に基づいたシークエンシング法の図式的概要が、図の上半分に示されている。マップされたインデル変異のヒストグラムプロットは、3 種の RGN オフターゲット部位について示されている。欠失は X 軸の上に示されているのに対して、挿入は下に示されている。標的部位全体(すなわち、プロトSpacer および PAM 配列)の境界は点線で示され、かつプロトSpacer と PAM 配列との間の境界は他の 2 本の間の点線として示されている。RGN 切断は、プロトSpacer の 5' 端から 3 から 4 bp に生じると予測される。

20

【図 7 - 2】図 7 B - F . GUIDE - Seq 切断部位が真正の RGN オフターゲット変異部位であることを示した図である。(b) ~ (f) 以下を標的とした RGN についての GUIDE - Seq によって同定された切断部位についてのインデル頻度(x 軸)および GUIDE - Seq シークエンシングの読み数(y 軸)の散乱プロット：VEGFA 部位 1、VEGFA 部位 2、VEGFA 部位 3、EMX1、および FANCF。

30

【図 8】図 8 A - E . RGN で誘発されるオフターゲット配列特性の解析を示した図である。(a) (GUIDE - Seq によって検出されるように)切断される特定の数のミスマッチを有する潜在的な RGN オフターゲット部位の画分。(b) 特定の数のミスマッチを有する RGN オフターゲット切断部位についての GUIDE - Seq の読み数のプロット(対数目盛)(c) RGN オフターゲット部位についての GUIDE - Seq の読み数に対する、プロトSpacer 内のミスマッチ部位の影響。塩基は、1 から 20 の番号が付けられ、20 は PAM に隣接した塩基である。(d) 線形回帰解析によって推定される、ゆらぎトランジション、非ゆらぎトランジション、およびトランスバージョンのミスマッチの影響。(e) ミスマッチ数、ミスマッチ型、ミスマッチ部位、PAM 密度、発現レベル、およびゲノム位置(遺伝子間/エクソン/イントロン)の影響についての個別の一変量解析によって説明される GUIDE - Seq の読み数相違の割合。

40

【図 9 - 1】図 9 A . RGN オフターゲット部位を同定するための GUIDE - Seq とコンピュータ予測または CHIP - Seq 法との比較を示した図である。(a) 9 種の RGN についての MIT CRISPR Design Tool および GUIDE - Seq によって予測されるオフターゲット部位間の重複を例示しているベン図。

【図 9 - 2】図 9 B . RGN オフターゲット部位を同定するための GUIDE - Seq とコンピュータ予測または CHIP - Seq 法との比較を示した図である。(b) 9 種の

50

RGNについてのE-CRISPコンピュータ予測プログラムおよびGUIDE-Seqによって予測されるオフターゲット部位間の重複を例示しているベン図。

【図9-3】図9C-D. RGNオフターゲット部位を同定するためのGUIDE-Seqとコンピュータ予測またはChIP-Seq法との比較を示した図である。(c) MIT CRISPR Design Toolによって予測される、予測されない、および考慮されない、GUIDE-Seqによって同定される真正のRGNオフターゲット部位の数を示しているヒストグラム。MIT CRISPR Design Toolによって予測される部位は、プログラムによって提供されるスコアに基づく五分位群に分けられる。各バーは、オフターゲット部位に対してミスマッチの数に基づいて細分類された部位を有する。バルジ部位は、スキップされた塩基部位をgRNA-プロトスペーサーDNA境界面に有するものである。(d) E-CRISPコンピュータ予測ツールによって予測される、予測されない、および考慮されない、GUIDE-Seqによって同定される真正のRGNオフターゲット部位の数を示しているヒストグラム。部位は、(c)に記載のように再分割される。

10

【図9-4】図9E-F. RGNオフターゲット部位を同定するためのGUIDE-Seqとコンピュータ予測またはChIP-Seq法との比較を示した図である。(e) ChIP-Seqによって同定されたdCas9結合部位とGUIDE-Seqによって同定されたRGNオフターゲット切断部位との間の重複を例示しているベン図。(f) 目的のオフターゲット部位に対して配列におけるミスマッチの数によって分類された、GUIDE-Seqによって同定されたRGNオフターゲット部位およびChIP-Seqによって同定されたdCas9結合部位のヒストグラムプロット。GUIDE-SeqおよびChIP-Seqミスマッチのカーネル密度評価が示されている。点線は、各クラスの部位についてのミスマッチの平均数を示す。

20

【図10-1】図10A. RGNによって誘発される大規模な構造変化を示した図である。(a) 転座を検出するためのAMP戦略の図式的概要。さらなる詳細は方法に。

【図10-2】図10B. RGNによって誘発される大規模な構造変化を示した図である。(b) RGNによって誘発される構造的バリエーションのCircosプロット。5種のRGNおよび細胞の対照についてのデータが示されている。染色体は円に配置されており、転座は2つの染色体位置間の弧として示されている。長さが1kbを超える欠失または逆位は、直線として示される。オフターゲット、オフターゲット、または切断点ホットスポット以外の部位は、「その他」として分類されている。

30

【図10-3】図10C. RGNによって誘発される大規模な構造変化を示した図である。(c) 第6染色体上のVEGFA部位1オフターゲット部位と第17染色体上のオフターゲット部位との間で検出される転座の例。4つすべての可能な相互転座が、AMPを使用して検出された。

【図10-4】図10C-D. RGNによって誘発される大規模な構造変化を示した図である。(d) AMPによって検出されるVEGFA部位2における2つのオフターゲット部位間の大きな欠失および逆位の例。(e) 5種のRGNで観察される、RGN誘発性およびRGN非依存性の異なる構造的バリエーションのまとめの表。Cas9のみ、dsODNオリゴのみ、および細胞のみでの対照も示されている。

40

【図10-5】図10F. RGNによって誘発される大規模な構造変化を示した図である。(f) U2OSおよびHEK293細胞における切断点ホットスポットの位置を例示している染色体イデオグラム。2つのホットスポットは染色体1および10のセントロメア領域において重複する。

【図11-1】図11A-B. tru-gRNAによって指向されるRGNのGUIDE-Seqプロファイルを示した図である。(a) 適合された全長gRNAおよびトランケートされたgRNAによってVEGFA部位1、VEGFA部位3、およびEMX1標的部位に指向されるRGNについて同定された既知および新規のオフターゲット切断部位の数。全長gRNAによって指向されるRGNについてのデータは、図1eに示されているものと同じであり、比較の容易性のためにここにまた示されていることに留意されたい

50

。(b) 適合された全長 gRNA およびトランケートされた gRNA によって VEGFA 部位 1、VEGFA 部位 3、および EMX1 標的部位に指向される RGN についてのオンターゲットおよびオフターゲット部位を示している染色体イデオグラム。全長 gRNA によって指向される RGN についてのイデオグラムは、図 1 h および図 1 3 A ~ B に示されているものと同じであり、比較の容易性のためにここにまた示されていることに留意されたい。

【図 1 1 - 2】図 1 1 C - D . tru - gRNA によって指向される RGN の GUIDE - Seq プロファイルを示した図である。(c) ~ (d) 適合された全長 gRNA およびトランケートされた gRNA によって VEGFA 部位 1、VEGFA 部位 3、および EMX1 標的部位に指向される RGN についてのオンターゲットおよびオフターゲット部位を示している染色体イデオグラム。全長 gRNA によって指向される RGN についてのイデオグラムは、図 1 h および図 1 3 A ~ B に示されているものと同じであり、比較の容易性のためにここにまた示されていることに留意されたい。

10

【図 1 1 - 3】図 1 1 E . tru - gRNA によって指向される RGN の GUIDE - Seq プロファイルを示した図である。(e) tru - gRNA によって指向される RGN によって誘発される DSB の GUIDE - Seq に基づいた同定。本発明者らが GUIDE - Seq によって評価した tru - gRNA によって指向される 3 種の RGN のオンターゲット部位についてのマップされた読みが示されている。すべての場合において、標的領域配列は、20 bp のプロトスペーサー配列が x 軸上の左側に、かつ PAM 配列が右側に示されている。RGN が全長 gRNA によって指向されるので、最も高いピークが、RGN 切断事象の予想される部位である、NGG PAM 配列の 5' 端の 3 から 4 bp 以内に入る方法に留意されたい。

20

【図 1 1 - 4】図 1 1 F - H . tru - gRNA によって指向される RGN の GUIDE - Seq プロファイルを示した図である。(f) ~ (h) tru - gRNA によって指向される RGN についての GUIDE - Seq によって同定されたオフターゲット部位の配列。各 RGN について、目的の標的配列が一番上の行に示され、切断された部位が下に示され、オンターゲット部位に対するミスマッチが示されかつ色で強調されている。GUIDE - Seq シークエンシングの読み数は、それぞれの部位の右側に示されている。目的のオンターゲット部位は四角形で印がつけられており、全長 gRNA および tru - gRNA の両方によって指向される RGN の既知のオフターゲット部位は濃い灰色のひし形で印がつけられており、かつ tru - gRNA によって指向される RGN のみで見出される既知のオフターゲット部位は薄い灰色のひし形で印がつけられている。既知のオフターゲット部位は、以前の報告 FU et al., Nat Biotechnol 32, 279-284 (2014)) において 0.1% 以上の変異生成頻度を有することが示されていたものであった。データは、tru - gRNA によって (f) VEGFA 部位 1、(g) VEGFA 部位 3、および (h) EMX1 標的部位に指向される RGN について示されている。

30

【図 1 2】dsODN 挿入およびインデル変異の検証のための GUIDE - Seq および AMP に基づいたシークエンシングの詳細な図式的概要を示した図である。両方のプロトコールについての詳細は、方法の中で見出すことができる。

【図 1 3 - 1】図 A - C . GUIDE - Seq によって評価された 10 種すべての RGN についての CRISPR / Cas9 オンターゲットおよびオフターゲット部位の染色体イデオグラムを示した図である。

40

【図 1 3 - 2】図 D - F . GUIDE - Seq によって評価された 10 種すべての RGN についての CRISPR / Cas9 オンターゲットおよびオフターゲット部位の染色体イデオグラムを示した図である。

【図 1 3 - 3】図 G - I . GUIDE - Seq によって評価された 10 種すべての RGN についての CRISPR / Cas9 オンターゲットおよびオフターゲット部位の染色体イデオグラムを示した図である。

【図 1 3 - 4】図 J . GUIDE - Seq によって評価された 10 種すべての RGN についての CRISPR / Cas9 オンターゲットおよびオフターゲット部位の染色体イデオ

50

グラムを示した図である。

【図14】GUIDE-Seqの読み数に対する因子の独立した影響を示すための多重因子線形回帰モデルを示した図である。

【図15-1】図15A．オフターゲット切断部位として以前に特徴付けされた7つのChIP-Seq結合部位についてのマップされたインデル変異のヒストグラムプロットである。実験試料および対照試料が各部位について並べて示されている。

【図15-2】図15B．オフターゲット切断部位として以前に特徴付けされた7つのChIP-Seq結合部位についてのマップされたインデル変異のヒストグラムプロットである。実験試料および対照試料が各部位について並べて示されている。

【図15-3】図15C．オフターゲット切断部位として以前に特徴付けされた7つのChIP-Seq結合部位についてのマップされたインデル変異のヒストグラムプロットである。実験試料および対照試料が各部位について並べて示されている。

10

【図15-4】図15D．オフターゲット切断部位として以前に特徴付けされた7つのChIP-Seq結合部位についてのマップされたインデル変異のヒストグラムプロットである。実験試料および対照試料が各部位について並べて示されている。

【図16-1】図16A．EGFPに対して標的化されたTALEN、ZFN、およびRFNを使用した、3つの型のdsODNの組み込み頻度を示しているグラフである。dsODNのすべてが5'リン酸化されていた。dsODNは、表示されている通り、無作為化された5'もしくは3'の4bpのオーバーハングを有していたまたは平滑末端であった。

【図16-2】図16B-C．U2OS細胞における2つの内在性標的部位、CCR5およびAPCでTALENによって誘発される二本鎖切断(DSB)の中への、平滑末端の、5'リン酸化された、34bpの二本鎖オリゴデオキシヌクレオチド(dsODN)(oSQT685/686)の効率的な組み込みを示しているグラフである。(16B)RFPLP解析により、2つの内在性部位、CCR5およびAPCでTALENによって誘発されるDSBの中へのdsODNタグoSQT685/686の組み込み(%)が示されている。(16C)累積の変異生成頻度は、これらの2つの内在性標的部位におけるT7E1アッセイによって測定される。

20

【図17】図17A-B．異なるdsODN末端保護の比較を示している棒グラフである；この実験において使用されたdsODNは、リン酸化されかつ平滑末端であって、かつ5'および3'ホスホロチオエート修飾の両方、または3'ホスホロチオエート修飾のみのいずれかを有していた。17A、ヒトU2OS細胞におけるRFN；17B、マウスES細胞におけるCas9。

30

【図18-1】図18A-B．マウスES細胞における3'ホスホロチオエート修飾されたオリゴの異なる濃度での実験を示しているグラフである。18A、Nanog sgRNA/Cas9；18B、Phc1 sgRNA/Cas9。dsODNは、リン酸化されかつ平滑末端であって、かつ5'および3'ホスホロチオエート修飾の両方、または3'ホスホロチオエート修飾のみのいずれかを有していた。実験は、ヒトU2OS細胞においてダイマーRNA誘導型FokIヌクレアーゼで(図18A)、またはマウスES細胞において標準Cas9で(図18B)行われた。

【図18-2】図18C．マウスES細胞における3'ホスホロチオエート修飾されたオリゴの存在下での破壊の割合のT7E1解析を示しているグラフである。

40

【図19】図19A-B．U2OS細胞における3つの内在性標的部位、VEGFA3、EMX1、およびFANCF1でCas9によって誘発される二本鎖切断(DSB)の中へのビオチン化されたdsODNタグの効率的な組み込みを示す図である。(19A)RFPLP解析により、U2OS細胞における3つの内在性標的部位、VEGFA3、EMX1、およびFANCF1でCas9によって誘発されるDSBの中への標準dsODN(oSQT685/686)と比較した、ビオチン化されたdsODN(oSQT1261/1262)の組み込み率(%)が示される。(19B)T7EIにより、U2OS細胞における3つの内在性標的部位、VEGFA3、EMX1、およびFANCF1での標準dsODN(oSQT685/686)と比較した、ビオチン化されたdsODN(oS

50

QT1261/1262)での推定される変異生成頻度(%)が示される。

【図20】図20A-B.より長いdsODNタグは、CRISPR-Cas9で誘発されるDSBの部位において効率良く組み込むために最適化されうること示す図である。(20A)RFLP解析により、75、50、または25pmolでトランスフェクトされるとき60bpのdsODN(oSQT1255/1256、oSQT1257/1258、およびoSQT1259/1260)の組み込み率(%)が示される。U2OS細胞において2つの内在性部位、EMX1およびFANCF1で試験された。(20B)T7E1により、75、50、または25pmolでトランスフェクトされるとき60bpのdsODN(oSQT1255/1256、oSQT1257/1258、およびoSQT1259/1260)の推定NHEJ率(%)が示される。U2OS細胞において2つの内在性部位、EMX1およびFANCF1で試験された。

10

【図21】異なるsgRNAを使用した改変VQRおよびVREER SpCas9バリエーションについてのGUIDE-seqによって同定されたオフターゲット切断部位の数を示しているグラフである。

【図22】オフターゲット部位での野生型およびD1135E SpCas9バリエーション間の特異性における、GUIDE-seqで検出される変化をまとめているグラフである。D1135Eについての読み数のない部位での特異性における推定される増加(倍)は、プロットされていない。

【図23】図23A-B.以下を示しているグラフである。(23A)制限断片長多型解析によって推定される、オンターゲット部位でのGUIDE-seqオリゴタグ組み込みの平均頻度。エラーバーは、平均値の標準誤差、n=4を表す；(23B)GUIDE-seq実験のためにT7E1によって検出されるオンターゲット部位における平均変異生成頻度。エラーバーは、平均値の標準誤差、n=4を表す。

20

【発明を実施するための形態】

【0025】

本明細書に記載のシーケンシングによって評価されるゲノムワイドでバイアスのないDSBの同定(GUIDE-Seq)法は、生細胞、例えば、非相同末端結合(NHEJ)修復経路が有効な細胞において改変ヌクレアーゼ切断部位の位置を同定するための高感度な、バイアスのない、かつゲノムワイドな方法を可能にする。一部の実施形態では、該方法は、ヌクレアーゼで誘発された切断の中への短い二本鎖オリゴデオキシヌクレオチド(dsODN)の捕捉(NHEJ経路によって媒介されると推定されるプロセス)および次いでゲノムの挿入の部位を同定するための挿入されたdsODN配列の使用、例えば、挿入されたdsODN配列を使用してハイスループットシーケンシングのためのゲノム挿入の部位を選択的に増幅する、PCRに基づいたディープシーケンシングアプローチを使用すること、または、例えば、溶液ハイブリッド捕捉を使用して、付加されたタグ、例えば、ピオチンなどを使用して、挿入されたdsODNを含むゲノム断片を選択的にプルダウンすること、に依存する。本明細書に記載されているのは、培養ヒト細胞におけるGUIDE-Seq法の開発および検証である；本明細書に記載されている一般的なアプローチは、すべての哺乳動物細胞においておよびNHEJ経路が活性であるまたは活性であると推定されるあらゆる細胞型または生物において有効なはずである。

30

40

【0026】

この最初のシーケンシングプロセスによって同定される潜在的なオフターゲット部位は、ヌクレアーゼ成分だけが発現される細胞におけるNHEJ修復のインデル変異特性について解析される可能性もある。これらの実験は、増幅およびその後のディープシーケンシングを使用して行われてもよく、それぞれのヌクレアーゼによって誘発されるオフターゲット変異の頻度のさらなる確認および定量を可能にするはずである。

【0027】

二本鎖オリゴデオキシヌクレオチド(dsODN)

本明細書に記載の方法において、天然に存在しないdsODNは、細胞内で発現される。本方法において、dsODNの両鎖は、細胞のゲノムにオルソログスである(すなわち

50

、細胞のゲノム内に存在しないまたは細胞のゲノム内に存在する配列に相補的である、すなわち、細胞のゲノム内に存在する配列に10%、20%、30%、40%、または50%以下の同一性を有する)。dsODNは、好ましくは15および75ntの間の長さ、例えば、15~50nt、50~75nt、30~35nt、60~65nt、もしくは50~65ntの長さ、または15および50ntの間の長さ、例えば、20~40もしくは30~35、例えば、32~34ntの長さでありうる。dsODNの各鎖は、固有のPCRプライミング配列を含んでいなければならない(すなわち、dsODNは、それぞれの鎖上に1つずつ、2つのPCRプライマー結合部位を含む)。一部の実施形態では、dsODNは、制限酵素認識部位、好ましくは、細胞のゲノムにおいて相対的にまれな部位を含む。

10

【0028】

dsODNは、好ましくは修飾される；好ましくは、dsODNの5'末端は、リン酸化される；さらにまた好ましくは、2つのホスホロチオエート結合が、両方の3'末端および両方の5'末端上に存在する。好ましい実施形態では、dsODNは、平滑末端である。一部の実施形態では、dsODNは、5'または3'末端上に無作為な多様な1、2、3、4つまたはそれより多くのヌクレオチドオーバーハングを含む。

【0029】

dsODNは、1つまたは複数のさらなる修飾、例えば、当技術分野において知られているまたはPCT/US2011/060493に記載されているようなものも含んでいてもよい。例えば、一部の実施形態では、dsODNは、ビオチン化される。GUIDE-seq dsODNタグのビオチン化バージョンは、ゲノムのDSBの部位の中への組み込みのための基質として使用される。ビオチンは、dsODNに対して内側のどこにあってもよいが(例えば、修飾されたチミジン残基(ビオチン-dT)、またはビオチンアジドを使用する)、5'または3'末端上にあってはならない。実施例4に示されているように、こうしたオリゴを効率良く組み込むことが可能である。これは、GUIDE-seq dsODNタグを含む断片を回収する代替の方法を可能にする。一方、一部の実施形態では、これらの配列は、ネステッドPCRによって検索および同定され、このアプローチでは、これらは、ビオチンを使用することによって、例えば、ストレプトアビジンコーティング付き磁気ビーズに結合させることによって、または溶液ハイブリッド捕捉を使用して、物理的にプルダウンされる；例えば、Gnirke et al., *Nature Biotechnology* 27, 182 - 189 (2009)を参照されたい。主な利点は、両方のランキング配列の検索であり、これは、オフターゲット切断部位を同定するために参照ゲノムに配列をマップすることへの依存性を低下させる。

20

30

【0030】

改変ヌクレアーゼ

以下の4つの主なクラスの改変ヌクレアーゼがある：1)メガヌクレアーゼ、2)ジンクフィンガーヌクレアーゼ、3)転写活性化因子エフェクター様ヌクレアーゼ(TALEN)、および4)クラスター化された規則的な間隔の短鎖反復回文配列(CRISPR) Cas RNA誘導型ヌクレアーゼ(RGN)。例えば、Gaj et al., *Trends Biotechnology* 2013 Jul;31(7):397-405を参照されたい。ヌクレアーゼは、当技術分野において知られている方法を使用して、細胞内で一過的にまたは安定に発現されてもよい；典型的には、発現を得るために、タンパク質をコードしている配列は、直接的な転写のためのプロモーターを含む発現ベクター中にサブクロニングされる。好適な真核発現系は、当技術分野においてよく知られかつ、例えば、Sambrook et al., *Molecular Cloning, A Laboratory Manual* (4th ed. 2013); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (2006); および *Current Protocols in Molecular Biology* (Ausubel et al., eds., 2010)に、記載されている。真核細胞および原核細胞の形質転換は、標準的な技術(例えば、上記の参考文献およびMorrison, 1977, *J. Bacteriol.* 132:349-351; Clark-Curti

40

50

ss & Curtiss, *Methods in Enzymology* 101:347-362 (Wu et al., eds, 1983)を参照されたい) によって行われる。

【 0 0 3 1 】

ホーミングメガヌクレアーゼ

メガヌクレアーゼは、多様な生物、例えば、細菌、酵母、藻類および植物の細胞小器官などに由来している配列特異的エンドヌクレアーゼである。内在性メガヌクレアーゼは、12から30塩基対の認識部位を有する；18bpおよび24bpの長さのメガヌクレアーゼ認識部位を有するカスタマイズされたDNA結合部位が記載されており、いずれかが本方法およびコンストラクトにおいて使用されてもよい。例えば、Silva, G., et al., *Current Gene Therapy*, 11:11-27, (2011) ; Arnould et al., *Journal of Molecular Biology*, 355:443-58 (2006) ; Arnould et al., *Protein Engineering Design & Selection*, 24:27-31 (2011) ; および Stoddard, *Q. Rev. Biophys.* 38, 49 (2005) ; Grizot et al., *Nucleic Acids Research*, 38:2006-18 (2010)を参照されたい。

10

【 0 0 3 2 】

CRISPR - Casヌクレアーゼ

近年の研究は、クラスター化された、規則的な間隔の、短鎖反復回文配列 (CRISPR) / CRISPR関連 (Cas) システム (Wiedenheft et al., *Nature* 482, 331-338 (2012) ; Horvath et al., *Science* 327, 167-170 (2010) ; Terns et al., *Curr Opin Microbiol* 14, 321-327 (2011)) が、細菌、酵母およびヒト細胞において、ならびに全生物、例えば、ショウジョウバエ、ゼブラフィッシュおよびマウスなどにおいて *in vivo* で、ゲノム編集を行うための簡単かつきわめて効率的な方法の基礎として役に立ちうることを実証している (Wang et al., *Cell* 153, 910-918 (2013) ; Shen et al., *Cell Res* (2013) ; Dicarlo et al., *Nucleic Acids Res* (2013) ; Jiang et al., *Nat Biotechnol* 31, 233-239 (2013) ; Jinek et al., *Elife* 2, e00471 (2013) ; Hwang et al., *Nat Biotechnol* 31, 227-229 (2013) ; Cong et al., *Science* 339, 819-823 (2013) ; Mali et al., *Science* 339, 823-826 (2013c) ; Cho et al., *Nat Biotechnol* 31, 230-232 (2013) ; G

20

ratz et al., *Genetics* 194(4):1029-35 (2013))。化膿レンサ球菌 (*S. pyogenes*) からの Cas9ヌクレアーゼ (以下、簡単に Cas9) は、17~20ヌクレオチドの改変ガイドRNA (gRNA) (例えば、単一ガイドRNAまたは crRNA / tracrRNA対) とプロトスペーサー隣接モチーフ (PAM) (例えば、PAMは配列NGGまたはNAGに合致している) の隣にある目的の標的ゲノムDNA配列の相補鎖との間の単純な塩基対相補性を介して誘導されうる (Shen et al., *Cell Res* (2013) ; Dicarlo et al., *Nucleic Acids Res* (2013) ; Jiang et al., *Nat Biotechnol* 31, 233-239 (2013) ; Jinek et al., *Elife* 2, e00471 (2013) ; Hwang et al., *Nat Biotechnol* 31, 227-229 (2013) ; Cong et al., *Science* 339, 819-823 (2013) ; Mali et al., *Science* 339, 823-826 (2013c) ; Cho et al., *Nat Biotechnol* 31, 230-232 (2013) ; Jinek et al., *Science* 337, 816-821 (2012))。

30

【 0 0 3 3 】

一部の実施形態では、本システムは、細菌においてコードされるようなまたは哺乳動物細胞における発現のためにコドン最適化されたいずれかの、化膿レンサ球菌 (*S. pyogenes*) または黄色ブドウ球菌 (*Staphylococcus aureus*) からの野生型またはバリエーションの Cas9タンパク質を利用する。ガイドRNAは、細胞内で Cas9と一緒に発現される。ガイドRNAまたはヌクレアーゼのいずれか、または両方は、細胞内で一過的にまたは安定に発現されうる。

40

ガイドRNAまたはヌクレアーゼのいずれか、または両方は、細胞内で一過的にまたは安定に発現されうる。

【 0 0 3 4 】

TALエフェクター反復配列

キサントモナス属 (*Xanthomonas*) の中の植物病原性細菌のTALエフェクターは、宿主DNAを結合することおよびエフェクター特異的宿主遺伝子を活性化することによって

50

、疾患において重要な役割を果たす、または防御を始動させる。特異性は、エフェクター可変数の、不完全な、典型的には約33～35アミノ酸の反復に依存する。多型は、主にリピート部位12および13に存在し、これは、本明細書において反復可変二残基(RVD)と呼ばれる。TALエフェクターのRVDは、それらの標的部位においてヌクレオチドに、直接的、直線的な様式で、1つのRVDが1つのヌクレオチドに、一部は縮重で、かつ見かけ上の前後関係の依存性なく、対応する。一部の実施形態では、ヌクレオチド特異性を付与する多型領域は、三残基またはトリプレットとして発現されうる。

【0035】

それぞれのDNA結合反復は、標的DNA配列において塩基対の認識を決定するRVDを含みうる。ここで、それぞれのDNA結合反復は、標的DNA配列内の1つの塩基対を認識する役割を果たす。一部の実施形態では、RVDは、以下のうちの1つまたは複数：Cを認識するためのHA；Cを認識するためのND；Cを認識するためのHI；Gを認識するためのHN；Gを認識するためのNA；GまたはAを認識するためのSN；Tを認識するためのYG；およびGを認識するためのNK、ならびに以下のうちの1つまたは複数を含みうる：Cを認識するためのHD；Tを認識するためのNG；Aを認識するためのNI；GまたはAを認識するためのNN；AまたはCまたはGまたはTを認識するためのNS；CまたはTを認識するためのN*、ここで、*は、RVDの第2の部位内のギャップを表す；Tを認識するためのHG；Tを認識するためのH*、ここで、*は、RVDの第2の部位内のギャップを表す；およびTを認識するためのIG。

【0036】

TALEタンパク質は、ゲノム工学において(例えば、植物において生物燃料または生物再生可能材料に有用な形質を付加または促進するために)相同組換えを促進することができる標的化されたキメラヌクレアーゼとして研究およびバイオテクノロジーにおいて有用でありうる。これらのタンパク質は、例えば、転写因子としても有用でありえ、かつ特に、非限定的な例として病原体(例えば、ウイルス)に対する療法のように、きわめて高いレベルの特異性を必要とする治療的適用のためにも有用でありうる。

【0037】

改変TALE配列を作製するための方法は、当技術分野において知られており、例えば、米国特許出願公開第61/610,212号明細書、およびReyon et al., *Nature Biotechnology* 30,460-465 (2012)に記載されている高速のライゲーションに基づいた自動化

可能な固相ハイスループット(FLASH)システム；ならびにBogdanove & Voytas, *Science* 333, 1843-1846 (2011)；Bogdanove et al., *Curr Opin Plant Biol* 13, 394-401

(2010)；Scholze & Boch, *J. Curr Opin Microbiol* (2011)；Boch et al., *Science* 326, 1509-1512 (2009)；Moscou & Bogdanove, *Science* 326, 1501 (2009)；Miller et al.,

Nat Biotechnol 29, 143-148 (2011)；Morbiter et al., *T. Proc Natl Acad Sci U S A* 107, 21617-21622 (2010)；Morbiter et al., *Nucleic Acids Res* 39, 5790-5799 (2011)；

Zhang et al., *Nat Biotechnol* 29, 149-153 (2011)；Geissler et al., *PLoS ONE* 6, e19509 (2011)；Weber et al., *PLoS ONE* 6, e19722 (2011)；Christian et al., *Genetics* 186, 757-761 (2010)；Li et al., *Nucleic Acids Res* 39, 359-372 (2011)；Mahfo

uz et al., *Proc Natl Acad Sci U S A* 108, 2623-2628 (2011)；Mussolino et al., *Nucleic Acids Res* (2011)；Li et al., *Nucleic Acids Res* 39, 6315-6325 (2011)；Cermak et al., *Nucleic Acids Res* 39, e82 (2011)；Wood et al., *Science* 333, 307 (2011)；Hockemeyer et al. *Nat Biotechnol* 29, 731-734 (2011)；Tesson et al., *Nat Biotechnol* 29, 695-696 (2011)；Sander et al., *Nat Biotechnol* 29, 697-698 (2011)；

Huang

10

20

30

40

50

et al., Nat Biotechnol 29, 699-700 (2011) ; および Zhang et al., Nat Biotechnol 29, 149-153 (2011) に記載されている方法を参照されたい ; これらのすべては、参照により本明細書にその全体が組み込まれている。

【 0 0 3 8 】

ジンクフィンガー

ジンクフィンガータンパク質は、独立に折りたたまれた亜鉛を含むミニドメインである、1つまたは複数のジンクフィンガーを含むDNA結合タンパク質であり、その構造は、当技術分野においてよく知られており、かつ、例えば、Miller et al., 1985, EMBO J., 4:1609 ; Berg, 1988, Proc. Natl. Acad. Sci. USA, 85:99 ; Lee et al., 1989, Science . 245:635 ; および Klug, 1993, Gene, 135:83 に定義されている。DNA に結合したジンクフィンガータンパク質 Z i f 2 6 8 およびそのバリエーションの結晶構造は、半分保存された相互作用のパターンを示し、そこでは、典型的にはジンクフィンガーのアルファヘリックスからの3つのアミノ酸が、DNA内の3つの隣接した塩基対すなわち「サブサイト」と接触する (Pavletich et al., 1991, Science, 252:809 ; Elrod-Erickson et al., 1998, Structure, 6:451) 。したがって、Z i f 2 6 8 の結晶構造により、ジンクフィンガー-DNA結合ドメインが、ジンクフィンガーとDNA配列内の3塩基対「サブサイト」との間の1対1の相互作用を有するモジュールの様式で機能する可能性があることが示唆される。天然に存在するジンクフィンガー転写因子において、複数のジンクフィンガーは、典型的には、直列の配列内に一緒に結合されて、連続したDNA配列の配列特異的認識を達成する (Klug, 1993, Gene 135:83) 。

【 0 0 3 9 】

複数の研究は、DNA結合に関与するアルファヘリックスの部位においてアミノ酸を無作為化することおよびファージディスプレイのような選択方法を使用して目的のDNA標的部位に結合する能力のある望ましいバリエーションを同定することによって個々のジンクフィンガーのDNA結合特性を人為的に改変することが可能であることを示している (Rebar et al., 1994, Science, 263:671 ; Choo et al., 1994 Proc. Natl. Acad. Sci. USA, 91:11163 ; Jamieson et al., 1994, Biochemistry 33:5689 ; Wu et al., 1995 Proc. Natl. Acad. Sci. USA, 92: 344) 。

こうした組換えジンクフィンガータンパク質は、機能ドメイン、例えば、転写活性化因子、転写抑制因子、メチル化ドメイン、およびヌクレアーゼと融合させて、遺伝子発現を調節し、DNAメチル化を変化させ、モデル生物、植物、およびヒト細胞のゲノム内に標的とする変化を導入することができる (Carroll, 2008, Gene Ther., 15:1463-68 ; Cathomen, 2008, Mol. Ther., 16:1200-07 ; Wu et al., 2007, Cell. Mol. Life Sci., 64:2933-44) 。

【 0 0 4 0 】

「モジュール集合体」として知られている、ジンクフィンガー配列を改変するための1つの既存の方法は、予め選択されたジンクフィンガーモジュールを配列内に単純に一緒に連結することを提唱している (Segal et al., 2003, Biochemistry, 42:2137-48 ; Beerli et al., 2002, Nat. Biotechnol., 20:135-141 ; Mandell et al., 2006, Nucleic Acids Res., 34:W516-523 ; Carroll et al., 2006, Nat. Protoc. 1:1329-41 ; Liu et al., 2002, J. Biol. Chem., 277:3850-56 ; Bae et al., 2003, Nat. Biotechnol., 21:275-280 ; Wright et al., 2006, Nat. Protoc., 1:1637-52) 。

いかなる研究者によって実施されるのにも十分に簡単であるにもかかわらず、近年の報告は、この方法のために、特に、ジンクフィンガーヌクレアーゼの文脈において、高い失敗率 (Ramirez et al., 2008, Nat. Methods, 5:374-375 ; Kim et al., 2009, Genome Res. 19:1279-88) 、いずれの所与の

標的遺伝子のためにもきわめて多数のジンクフィンガータンパク質の構築および細胞に基づいた試験を典型的には必要とするという制限 (Kim et al., 2009, Genome Res. 19:1279-88) を示している。

10

20

30

40

50

【0041】

無作為化されたライブラリーからジンクフィンガー配列を同定する、組合せの選択に基づいた方法は、モジュール集合体よりも高い成功率を有することが示されている (Maeder et al., 2008, Mol. Cell, 31:294-301; Joung et al., 2010, Nat. Methods, 7:91-92; Isalan et al., 2001, Nat. Biotechnol., 19:656-660)。好ましい実施形態では、ジンクフィンガー配列は、国際公開第2011/017293号パンフレットおよび国際公開第2004/099366号パンフレットに記載されている、またはこれらに記載されているように作製される。さらなる好適なジンクフィンガーDBDは、米国特許第6,511,808号明細書、第6,013,453号明細書、第6,007,988号明細書、および第6,503,717号明細書ならびに米国特許出願公開第2002/0160940号明細書に記載されている。

10

【0042】

細胞

本明細書に記載の方法は、ゲノムDNAにおいてDSBを修復する能力のあるいかなる細胞においても使用されうる。真核細胞における2つの主要なDSB修復経路は、相同組換え(HR)および非相同末端結合(NHEJ)である。好ましくは、該方法は、NHEJの能力のある細胞において行われる。NHEJ活性を検出するための方法は、当技術分野において知られている; NHEJの標準経路および副経路の概要については、Liu et al., Nucleic Acids Res. Jun 1, 2014; 42(10):6106-6127を参照されたい。

20

【0043】

シーケンシング

本明細書中で使用される場合、「シーケンシング」は、核酸の配列を決定するいかなる方法をも含む。鎖終結(サンガー)シーケンシングおよび色素終結シーケンシングを含む、シーケンシングのいかなる方法も本方法において使用されうる。好ましい実施形態では、何千または何百万ものシーケンシング反応を並行に行うハイスループットシーケンシング技術である、次世代シーケンシング(NGS)が使用される。異なるNGSプラットフォームは異なるアッセイ化学を使用するが、これらはすべて、多数の鋳型上で同時に行われる多数のシーケンシング反応から配列データを生成する。典型的には、配列データは、スキャナーを使用して収集され、次いで、バイオインフォマティクスの集結および解析される。したがって、シーケンシング反応は、並行に、実施、読み取り、集結、および解析される; 例えば、米国特許第20140162897号明細書、ならびにVoelkerding et al., Clinical Chem., 55: 641-658, 2009; およびMacLean et al., Nature Rev. Microbiol., 7: 287-296 (2009)を参照されたい。NGS法には、鋳型増幅を必要とするものもあり、必要としないものもある。増幅を必要とする方法は、ピロシーケンシング(例えば、米国特許第6,210,899号明細書および第6,258,568号明細書を参照されたい; Rocheによって商品化されている); Solexa/Illuminaプラットフォーム(例えば、米国特許第6,833,246号明細書、第7,115,400号明細書、および第6,969,488号明細書を参照されたい); ならびに補助されたオリゴヌクレオチドライゲーションおよび検出(SOLID)プラットフォーム(Applied Biosystems; 例えば、米国特許第5,912,148号明細書および第6,130,073号明細書を参照されたい)を含む。増幅を必要としない方法(例えば、一分子シーケンシング法)は、ナノポアシーケンシング、Heliscope(米国特許第7,169,560号明細書; 第7,282,337号明細書; 第7,482,120号明細書; 第7,501,245号明細書; 第6,818,395号明細書; 第6,911,345号明細書; および第7,501,245号明細書); 合成によるリアルタイムシーケンシング(例えば、米国特許第7,329,492号明細書を参照されたい); ゼロモード導波路(ZMW)を使用した一分子リアルタイム(SMRT)DNAシーケンシング法; ならびに、米国特許第7,170,050号明細書; 第7,302,146号明細書; 第7,313,308号明細書; および第7,476,503号明細書に記載されているものを含む)、他の方法。例えば、米国特

30

40

50

許出願公開第2013/0274147号明細書；米国特許出願公開第2014/0038831号明細書；Metzker, Nat Rev Genet 11(1): 31-46 (2010)を参照されたい。

【0044】

代替的に、ハイブリダイゼーションに基づいたシーケンス法または他のハイスループットな方法、例えば、マイクロアレイ解析、NANOSTRING、ILLUMINA、または他のシーケンシングプラットフォームも使用されてもよい。

実施例

【0045】

本発明は、以下の実施例においてさらに記載されるが、これは、特許請求の範囲に記載の本発明の範囲を限定するものではない。

【実施例1】

【0046】

最初の実験では、ヌクレアーゼで誘発された二本鎖切断(DSB)の中にdsODNカセットを組み込むプロセスを最適化した。以前に公表された実験は、5'末端に2つのホスホロチオエート結合修飾を有しているdsODNを、哺乳動物細胞においてジンクフィンガーヌクレアーゼ(ZFN)で誘発されたDSBの中に捕捉できたことを実証している(Orlando et al., Nucleic Acids Res. 2010 Aug;38(15):e152)。しかし、こうしたssODNの捕捉を使用してきた低頻度のDSBでも同定するために、こうした切断の中へのその捕捉の割合を改良するためにdsODNの特性を最適化した。当初の努力は、化膿レンサ球菌(Streptococcus pyogenes)からのクラスター化された規則的な間隔の短

鎖反復回文配列(CRISPR)RNA誘導型ヌクレアーゼCas9によって誘発されたDSB中にdsODNを捕捉することに焦点を当てた。Cas9は、平滑末端を有するDSBを誘発することが報告されており、したがって、dsODNバリエーションは、平滑末端化されるように設計した。最適化実験は、両方の5'末端のリン酸化および(5'末端上のものに加えて)両方の3'末端上の2つのホスホロチオエート結合の導入により、Cas9で誘発されたDSBの中へのdsODNの捕捉の割合の大幅な上昇が導かれたことを示した(図1A~B)。サンガーシーケンシングより、この特定のDSB中にdsODNを成功裏に捕捉したことが検証された(図2A~B)。

【0047】

Cas9で誘発されたDSB中にdsODNが効率良く組み込まれることが確立されたので、次の実験では、次世代ディープシーケンシング法を哺乳動物細胞のゲノム内のdsODN組み込み部位の捕捉、増幅および同定に使用し得るかどうかを決定することが求められた。これを行うために、(それぞれの鎖上に1つの)2つのPCRプライマー結合部位を含む34bpのdsODNを利用した；これらの配列が選択された理由は、それらがヒトゲノムにそれぞれオルソログであるためである。

【0048】

使用しdsODNの配列を、表1に示す：

【0049】

【表1】

表1

鎖	配列(5'から3')	配列番号
FWD	/5Phos/G*T*TTAATTGAGTTGTCATATGTTAATAACGGT*A*T	1
REV	/5Phos/A*T*ACCGTTATTAACATATGACAACCTCAATTAA*A*C	2

/5Phos/は、5'リン酸化を表す。

*は、隣接ヌクレオチド間のホスホロチオエート結合を表す。

【0050】

この dsODN を、Cas9 をコードするプラスミド、および異なる内在性ヒト遺伝子配列 (EMX1 および VEGFA 部位 1、2、および 3) にそれぞれ標的化された、4 種の異なる標的的特異的 gRNA のうちの 1 種をコードするプラスミドと共にヒト U2OS 細胞内にトランスフェクトした。これら 4 種の特定 gRNA は、これらのそれぞれについて真正のオフターゲット部位が以前に同定されていた (Fu et al., Nat Biotechnol. 2013; Table 1) という理由で選択した。トランスフェクションは、以下のように行った: dsODN を、STE (100 mM TrisHCl, 500 mM NaCl, 10 mM EDTA) 中でそれぞれ 100 μ M の濃度でアニールさせる。U2OS 細胞では、2E5 細胞を溶液 SE およびプログラム DN-100 でヌクレオフェクトするために、500 ng の Cas9 発現プラスミド、250 ng の gRNA 発現プラスミド、および 100 pmol の dsODN を使用した。

10

【0051】

トランスフェクション 3 日後にゲノム DNA を採取し (Agencourt Ampure XP)、dsODN にコードされた制限部位の存在に基づいて、dsODN がこれらの細胞内のオンターゲット部位の中に効率良く組み込まれていることを検証するために、PCR ベースの制限断片長多型 (RFLP) アッセイを使用した。

【0052】

トランスフェクトされた細胞のゲノム内の dsODN 組み込みの位置を包括的に同定するために、これらの挿入部位を選択的に増幅してさらにそれらが次世代シーケンシング技術を使用してシーケンスされるのを可能にする PCR ベースの方法を使用した。この戦略の一般的な概要を図 3 に示す。ゲノム DNA は、Covaris Adaptive Focused Acoustic (AFA) 焦点式超音波発生装置で 500 bp の平均長に切断した。切断された gDNA を末端修復し (Enzymatics)、A テール付加して (Enzymatics)、切断された DNA の末端に半機能的シーケンシングアダプター (米国特許第 20130303461 号明細書) を連結した (Enzymatics)。これらの酵素的ステップのそれぞれを精製するために、固相可逆固定法 (SPRI) 磁気ビーズ精製を使用した (Agencourt XP)。

20

【0053】

dsODN 配列を有する DNA 断片は、次いで、dsODN に特異的なプライマーを、シーケンシングアダプターにアニールするプライマーと一緒に使用して増幅した。(上記のようにそれぞれの鎖上に 1 つの) 2 つの潜在的なプライミング部位が dsODN 内にあるので、望ましい配列を選択的に増幅するために以下のように 2 つの独立した PCR 反応を行った。

30

【0054】

標的化されたシーケンシングライブラリーを作製するために、2 ラウンドのネステッド PCR を行った。第一ラウンドの PCR は、組み込み dsODN と相補的なプライマー (プライマー A) およびユニバーサルアダプターと相補的なプライマー (プライマー B) を使用して行った。第二ラウンドの PCR は、プライマー A と相補的な 3' ネステッドプライマー (プライマー C)、プライマー B と相補的な 3' ネステッドプライマー (プライマー D)、ならびに、シーケンシングのために用意のできた「完全な」分子を作製するためにフローセル結合配列および無作為な分子インデックスを付加した、プライマー D と相補的なプライマー (プライマー E) を使用して行った。各ラウンドの PCR を精製するために、SPRI 磁気ビーズを使用した。(Agencourt Ampure XP)

40

【0055】

このアプローチによる dsODN 含有ゲノム配列の増幅は、ゲノム DNA の無作為な切断によって誘発された切断にシーケンシングアダプターがライゲーションされるので、挿入地点に隣接したフランキング配列に依存することもそれによってバイアスされることもない。dsODN に最も近い末端上に次世代シーケンシングアダプター配列およびインデクシングバーコードを付加するためにさらなるラウンドの PCR を行い、結果として、次世代シーケンシング用に準備できた断片のライブラリーを得た。この一般的な方法

50

は、本明細書において、シークエンシングによって評価されるゲノムワイドでバイアスのないDSBの同定 (Genomewide Unbiased Identification of DSBs Evaluated by Sequencing) として、GUIDE-Seqと呼ぶ。

【0056】

GUIDE-Seqを使用して構築されたライブラリーのディープシークエンシングにより、共発現された4種のgRNA/Cas9ヌクレアーゼのそれぞれの存在下においてdsODNが挿入されるようになった広範囲のゲノム座位が明らかにされた。生のディープシークエンシングデータの解析において、挿入の真正の部位が、両方向での少なくとも1つの読みによってカバーされたゲノム座位として同定されうることは理にかなっていた。dsODNがいずれの方向においても挿入できたためおよびdsODN配列内の一方または他方のいずれかの鎖に特異的なプライマーを使用して増幅が行われたためという両方の理由により、両方向での読みが可能であった。調査した4種のgRNAについて、この基準を満たした合計465のゲノム座位を同定した。これらの465座位のうちの36%では、使用したgRNAのオンターゲット部位と同様であり、かつ、オンターゲット部位に対して相対的に6つものミスマッチを有する、挿入地点の25bp以内の配列も同定された(図4A~E)。この方法は、ここで調査された4種すべてのgRNAについて、すべての既知の真正オフターゲット部位(図4に示されている既知のオフターゲット部位のすべては、Fu et al., Nat Biotechnol. 2013からの表1にも示されている)ならびにこれまでに知られていない多くのさらなるオフターゲット部位の発見にも成功した。

【実施例2】

【0057】

カスタマイズ可能なCRISPR-Cas RNA誘導型ヌクレアーゼ(RGN)は、広範囲の研究および潜在的な臨床の適用を有する強力な、カスタマイズ可能なゲノム編集試薬である^{1, 2}; しかし、ヒトにおけるRGNの治療的使用には、有害な結末のリスクを最小限にするためにこれらのオフターゲット作用についての完全な知識が必要とされることになる。化膿レンサ球菌(*S. pyogenes*) Cas9ヌクレアーゼによるDNA切断は、プログラム可能な約100ntのガイドRNA(gRNA)によって指向される³。ターゲティングは、gRNAの5'末端にある17~20ntによって介され、これは、5'-NGG型のプロトスペーサー隣接モチーフ(PAM)の隣にある「プロトスペーサー」DNA部位と相補的である。非相同末端結合(NHEJ)による、プロトスペーサー内のCas9で誘発されたDNA二本鎖切断(DSB)の修復は、可変長の挿入/欠失変異(インデル)を誘発しうる。本発明者らのグループおよび他は、RGNで誘発された意図されていないインデルが、プロトスペーサー内の5つもの部位によって異なるオフターゲット切断部位で生じうること、または代替のPAM配列を有することを以前に示している^{4~7}。染色体転座は、RGNで誘発されたオンターゲットおよびオフターゲットの切断事象の連結によって生じうる^{8~11}。RGNを使用するex vivoおよびin vivoでの治療戦略には、きわめて大きな細胞集団の修飾が必要とされることが予想されるので、臨床的適用では、低頻度であっても変化の同定は、決定的に重要であろう。こうした変化は好ましくない臨床的結末をもたらすので、細胞クローンのたとえまれなサブセットにおいても癌化の誘導(例えば、腫瘍抑制遺伝子の変異を不活性化することまたは腫瘍形成性染色体転座の形成)が特に懸念される。

【0058】

ゲノム内のどこにでも生じうるインデルまたは高次ゲノム再配列の包括的な同定は、簡単には対処されない課題であり、残念なことに、生細胞におけるバイアスのない、ゲノムワイドな、RGNで誘発されたオフターゲット変異の同定のための高感度の方法はまだ記載されていない^{12, 13}。全ゲノム再シークエンシングは、編集された単一細胞クローンにおけるRGNオフターゲット変化を同定するための試みに使用されてきた^{14, 15}。しかし、きわめて多数のゲノムをシークエンスする高いコストにより、この方法は、細胞集団における低頻度の事象を見つけるのには非実用的とされる¹²。本発明者らおよび

他者は、オンターゲット部位に対する配列類似性⁴、⁵または部分的に変性した結合部位ライブラリーからの *in vitro* 選択⁶ のいずれかによって同定される潜在的なオフターゲット部位におけるインデル変異を同定するために、焦点式ディープシーケンシングを使用してきた。しかし、これらのアプローチは、オフターゲット配列の性質についての仮定を作るものであり、したがって、ゲノム内の他の場所の他の変異部位を見落とす可能性がある。ChIP-Seqも、触媒的に不活性の (dead) Cas9 (dCas9) と複合体形成した gRNA についてのオフターゲット結合部位を同定するために使用されてきたが、公表された研究の大多数は、これらの部位のうち、活性 Cas9 ヌクレアーゼによる切断のオフターゲット部位であるものは、たとえあるとしても、きわめて少ないことを示唆している¹⁶ ~ ¹⁹。

10

【0059】

ここで、本発明者らは、ヒト生細胞における異なる10種のRGNについての最初の包括的な特異性ランドスケープの作製を可能にする、シーケンシングによって評価されるゲノムワイドでバイアスのないDSBの同定 (GUIDE-Seq) のための新規な方法の開発を記載する。これらのプロファイルにより、オフターゲットDSBの総数は、個々のRGNについて広範に変化したことが明らかにされ、化膿レンサ球菌 (*S. pyogenes*) または他の種からのRGNの特異性についての広い結論は、大規模な調査に基づいていなければならないが、少数のgRNAだけにに基づいてはならないことが示唆された。本発明者らの発見は、オフターゲット作用がそこで生じうる配列の範囲および性質を拡大もさせた。直接的な比較により、GUIDE-Seqは、RGNオフターゲット部位を同定するための2つの広範に使用されるコンピュータのアプローチおよびChIP-Seq法よりも実質的に優れていたことが実証された。予想外なことに、GUIDE-Seqは、RGNで誘発されたDSBと一緒に、転座などの高次ゲノム変化に關与しうるRGN非依存性のDNA切断点ホットスポットも同定した。最後に、本発明者らは、直接的な比較において、gRNAの相補性領域をトランケートすることにより、それらのゲノムワイドなオフターゲットDSBプロファイルが大きく改良されることを示しており、RGNの特異性を改良するために設計された向上を評価するためのGUIDE-Seqの有用性を実証している。ここに概説される実験は、RGNの特異性を評価するためにこれまでに記載された中で、ならびに、治療的使用のために考慮されうる、プラットフォームに対するあらゆる改良の中で、最も精密な戦略を提供する。

20

30

【0060】

方法

以下の材料および方法がこの実施例において使用された。

【0061】

ヒト細胞培養およびトランスフェクション

U2OSおよびHEK293細胞は、10% FBS、2mM GlutaMax (Life Technologies)、およびペニシリン/ストレプトマイシンを補ったAdvanced DMEM (Life Technologies) 中で5%のCO₂と共に37°Cで培養した。U2OS細胞 (プログラムDN-100) およびHEK293細胞 (プログラムCM-137) は、Lonza Nucleofector 4-D上で製造業者の説明書に従って20µlの溶液SE中でトランスフェクトした。dsODN組み込み率は、NdeIを使用した制限断片長多型 (RFLP) アッセイによって評価した。切断産物は、以前に記載されているように (Tsai et al., Nat. Biotechnol 32, 569-576 (2014))、Qiaxcelキャピラリー電気泳動機器 (Qiagen) によって泳動および定量化した。

40

【0062】

GUIDE-SeqのためのゲノムDNAの単離および調製

ゲノムDNAは、固相可逆固定法磁気ビーズ (Agencourt DNAdvance) を使用して単離し、Covaris S200超音波発生装置で500bpの平均長に切断し、末端修復し、Aテール付加し、8ntの無作為な分子インデックスを組み込ん

50

でいる半機能的アダプターにライゲーションした。標的濃縮のために、オリゴタグと相補的なプライマーでの、2ラウンドのネステッドアンカーDPCRを使用した。例示的なGUIDE-Seqプロトコルの全詳細は、本明細書中で見出すことができる。

【0063】

シーケンシングの読みの加工および統合

配列の最初の6つの同じ塩基ならびに同一の8ntの分子インデックスを共有する読みは、それらが同じ起源のプレPCR鋳型断片に由来するとみなされるので、一緒にビンギンされる。これらの読みは、各部位において大多数の塩基を選択することによって単一のコンセンサスの読みに統合した。ノーコール(N)塩基は、10%より高い不一致の読みを有する状況で割り当てた。塩基クオリティスコアは、予め統合された読みの間で最も高くなるようにした。統合された読みは、BWA-MEM(Li and Durbin, *Bioinformatics* 26, 589-595 (2010))を使用してヒトゲノム参照(GrCh37)にマップした。

10

【0064】

オフターゲット切断部位の同定

マッピングのクオリティ 50を有する読みの開始マッピング部位を表にし、10bpのスライド式ウィンドウを使用して近くに開始マッピング部位を有する領域をグループ化した。組み込まれたdsODNを有するゲノムウィンドウは、以下の判定基準のうちの一つによって同定した：1)参照配列内の逆鎖に位置する2つ以上の固有の分子インデックス付加された読み、または2)フォワードおよびリバースプライマーによって増幅された2つ以上の固有の分子インデックス付加された読み。推定された切断点の両側の側方の25bpの参照配列を目的の標的部位に対してアライメントし、目的の標的配列から8つ以下のミスマッチを有するRGNオフターゲット部位を判定した。SNPおよびインデルは、これらの部位において、分子インデックスおよびSAMtoolsに基づいたカスタムピンコンセンサスバリエーション判定アルゴリズムによって判定し、参照配列と異なったオフターゲット配列は、対応する細胞特異的な配列と置き換えた。

20

【0065】

AMPに基づいたシーケンシング

GUIDE-Seqで検出されたDSBのAMP検証のために、プライマーは、以前に記載されているように(Zheng, Z. et al. *Anchored multiplex PCR for targeted next-generation sequencing*. *Nat Med* 2014 Nov 10. doi: 10.1038/nm.3729 (2014))、推定

30

された二本鎖切断点の側方の領域に設計し、8ntの分子インデックスを付加した。可能な場合には、本発明者らは、各DSBの側方に位置するように2つのプライマーを設計した。

【0066】

AMP検証データの解析

平均クオリティスコア > 30を有する読みは、GUIDE-Seqで推定されたDSB部位とオーバーラップした挿入、欠失、および組み込みについてPythonを使用して解析した。PCRまたはシーケンシングの誤りからのノイズが入るのを最小限にするために、1bpのインデルは、それらが予測されたDSB部位の1bp以内であった場合のみに含めた。組み込みおよびインデルの頻度は、統合された分子インデックス付加された読みに基づいて算出した。

40

【0067】

構造的バリエーション

転座、大きな欠失、および逆位は、スプリットBWA-MEMアライメントに基づいたカスタムアルゴリズムを使用して同定した。同一染色体上の50塩基以内の候補の融合切断点は、Cas9切断部位の周囲の潜在的な切断を含むようにグループ化した。融合事象は、独自にマップされた少なくとも3つのスプリットの読みで判定し、パラメーターもsegemehlツールにより使用した(Hoffmann, *Genome biology*, 2014))。マップ

50

グストランデッドネスは、2つの関与しているDSBの間の相互融合の同定のため、および欠失または逆位を決定するために維持した。1 kbの染色体部位内にDSBを含む融合は、単一のCas9切断によって引き起こされる大きなインデルが考えられるために無効とした。残りの融合DSBは、以下の4つのカテゴリーに分類した：GUIDE-seqに基づく「オンターゲット」、「オフターゲット」もしくは「バックグラウンド」、または他に「その他」。

【0068】

GUIDE-SeqおよびChIP-Seqによって検出される部位ならびにコンピュータ上での予測の比較

本発明者らは、MIT CRISPR Design Toolを使用して、10種すべてのRGNについての潜在的なオフターゲット部位を同定した。このツールは、それぞれの潜在的なオフターゲット部位に、対応する百分位数を定める。本発明者らは、次いで、これらの百分位数を可視化目的で五分位群にグループ化した。E-CRISPRツールはオフターゲットをランク付けしないので、本発明者らは、単純に、E-CRISPRによって正確に予測されたGUIDE-seqオフターゲットを見つけた。これらのGUIDE-Seq対コンピュータ上での予測の両方について、本発明者らは、さらに、コンピュータ上での方法によって予測されなかったGUIDE-Seq結果を、MITツール（最大4つ）およびE-CRISPR（最大3つ）の範囲内のミスマッチ数を有するオフターゲット、ならびにこれらの予測ツールの閾値を超えるミスマッチ数を有するものに分割した。ChIP-Seq予測とのGUIDE-Seqオフターゲットの比較において、同じ技術を使用し、ChIP-Seqによって正確に予測されたGUIDE-Seqオフターゲットを見つけた。これらの比較のそれぞれについて、なされたすべてのグループ化は、オフターゲットミスマッチ数によって再分割し、正しくおよび誤って予測されたRGNオフターゲットの特性をより良好に特徴付けした。

【0069】

オフターゲット切断率に対する、ミスマッチ、DNAアクセシビリティおよび局所的PAM密度の影響の解析

本発明者らは、推定される切断率に適合した線形回帰モデルを使用して、4つ以下のミスマッチを有する潜在的なオフターゲット部位において特異性に対する、ミスマッチ部位、ミスマッチ型およびDNAアクセシビリティの影響を評価した。ミスマッチ部位の共変動は、PAMの上流のオーバーラップしていない5つの4 bpのウィンドウのそれぞれの中のミスマッチされた塩基の数として定義した。ミスマッチ型の共変動は、i) (標的TがCによって置き換えられ、標的GがAによって置き換えられた) ゆらぎ対形成を結果として生じるミスマッチの数、ii) (標的CがTによって置き換えられ、標的AがGによって置き換えられた) 非ゆらぎプリン-ピリミジン塩基対形成を結果として生じるミスマッチの数、およびiii) プリン-プリンまたはピリミジン-ピリミジン対形成を結果として生じるミスマッチとしての数と定義した。

【0070】

3つの因子のそれぞれは、別々のモデルにおいて、 $\log_2(1 + \text{GUIDE-Seq 読み数})$ によって推定される、相対切断率の予測因子として使用した。作用の大きさの推定は、標的部位間変動性のために調整した。各因子によって説明されるサイト内切断率変動性の比率は、回帰平方和(SS)に基づく偏イータ二乗統計： $F = \text{SS}_{\text{factor}} / (\text{SS}_{\text{factor}} + \text{SS}_{\text{error}})$ によって評価した。一因子モデルに加えて、本発明者らは、3つすべての因子、発現レベル、および1 kbのウィンドウ内のPAM密度を含む組み合わされた線形回帰モデルも適合させて、オフターゲット切断可能性に対するそれらの独立した寄与を評価した。

【0071】

Guide-seqライブラリー調製のための例示的な試薬および装置

室温で保存

品目 販売会社

10

20

30

40

50

Covaris S220マイクロチューブ、Covaris
 エタノール、200プルーフ(100%) Sigma Aldrich
 Micro Amp Optical 96ウェルプレート Applied Bios
 systems
 ヌクレアーゼ非含有H₂O Promega
 Qubitアッセイチューブ、500チューブ/パック Invitrogen
 Qubit dsDNA BRキット-500アッセイ Invitrogen
 TMACバッファー、5M Sigma Aldrich
 - 塩化テトラメチルアンモニウム
 1×TEバッファー/10mM トリス-HCl、pH8.0 Invitrogen 10
 UltraPure 0.5M EDTA、pH8.0(Gibco)(4×100m
 L) Life Technologies
 【0072】
 4 で保存
 品目 販売会社
 Agencourt AMPure XPビーズ - 60mL Beckman C
 oulter
 【0073】
 - 20 で保存
 品目 カタログ# 20
 25mM dNTP溶液ミックス Enzymatics, Inc.
 Slowライゲーションバッファー Enzymatics, Inc.
 末端修復ミックス(低濃度) Enzymatics, Inc.
 T4 DNAリガーゼ Enzymatics, Inc.
 - 10× T4 DNAリガーゼバッファー(Slowライゲーションバッファ
 ー)
 Platinum(登録商標)Taq DNAポリメラーゼ Life Techno
 logies
 - 10× PCRバッファー(MgCl₂なし)
 - 50mM MgCl₂ 30
 qPCR Illuminaライブラリー定量化キット KAPA Biosyste
 ms, Inc.
 【0074】
 機器
 96ウェルプレート磁気スタンド Invitrogen
 Qubit蛍光光度計2.0 Life Technologies
 Covaris S-2 Focused Ultra-sonicator(商標)
 機器 Covaris
 卓上遠心分離機 Thermo Scientific
 卓上ボルテクサー Thermo Scientific 40
 サーマサイクラー Eppendorf
 Miseq Illumina
 【0075】
 GUIDE-seqライブラリー調製のための例示的なプロトコール
 Yアダプター調製
 Yアダプターは、Miseq共通オリゴを試料バーコードアダプター(A01からA1
 6、表3を参照されたい)のそれぞれとアニールすることによって作製される。アダプタ
 ーは、8merのNNWNNWNN(N=A、C、T、またはG;W=AまたはT)分子
 インデックスも含む。
 1×TEバッファー 80.0μL 50

A # # (1 0 0 μ M) 1 0 . 0 μ L

M i S e q C o m m o n A d a p t e r _ M I (1 0 0 μ M) 1 0 . 0 μ L

合計 1 0 0 . 0 μ L

アニーリングプログラム：9 5 1 秒間；6 0 1 秒間；4 まで緩徐な傾斜の低下（約 - 2 / 分）；4 で維持する。 - 2 0 で保存する。

【 0 0 7 6 】

インプット定量化および切断

1 . d s D N A は、Q u b i t に よ っ て 定 量 化 さ れ、1 × T E バ ッ フ ェ ー を 使 用 し て 4 0 0 n g が 1 2 0 u l の 最 終 容 量 に さ れ る。

2 . 各 試 料 は、C o v a r i s S 2 の た め の 標 準 操 作 プ ロ ト コ ール に 従 っ て 5 0 0 b p の 平 均 長 に 切 断 さ れ る。

3 . 1 2 0 u l の A M P u r e X P S P R I ビ ーズ (1 × 比 率) で の 精 製 は、製 造 業 者 の プ ロ ト コ ール に 従 っ て 行 い、1 5 u l の 1 × T E バ ッ フ ェ ー 中 に 溶 出 さ れ る。

【 0 0 7 7 】

末端修復、A - テーリングおよびライゲーション

末端修復

4 . 2 0 0 μ L の P C R チ ュ ー ブ ま た は 9 6 プ レ ー ト 中 の ウ ェ ル に、以 下 を (反 応 ご と に) 添 加 す る :

ヌクレアーゼ非含有 H₂O 0 . 5 μ L

d N T P ミ ッ ク ス、5 m M 1 . 0 μ L

S L O W ラ イ ゲ ー シ ョ ン バ ッ フ ェ ー、1 0 × 2 . 5 μ L

末端修復ミックス (低濃度) 2 . 0 μ L

T a q ポ リ メ ラ ー ゼ 用 の バ ッ フ ェ ー、1 0 × (M g²⁺ な し) 2 . 0 μ L

T a q ポ リ メ ラ ー ゼ (非 ホ ッ ト ス タ ー ト) 0 . 5 μ L

合計 8 . 5 μ L

+ (前 の ス テ ッ プ か ら の) D N A 試 料 1 4 . 0 μ L

合計 2 2 . 5 μ L

末端修復サーモサイクラープログラム：1 2 1 5 分 間、3 7 1 5 分 間；7 2 1 5 分 間；4 で 維 持 す る

【 0 0 7 8 】

アダプターライゲーション

5 . 試 料 反 応 チ ュ ー ブ ま た は ウ ェ ル に、以 下 の 試 薬 を 順 番 に 添 加 す る (ピ ペ ッ テ ィ ン グ に よ っ て 混 合 す る) :

アニールされた Y a d a p t e r _ M I (1 0 μ M) 1 . 0 μ L

T 4 D N A リ ガ ー ゼ 2 . 0 μ L

+ (前 の ス テ ッ プ か ら の) D N A 試 料 2 2 . 5 μ L

合計 2 5 . 5 μ L

アダプターライゲーションサーモサイクラープログラム：1 6 3 0 分 間、2 2 3 0 分 間、4 で 維 持 す る

6 . 0 . 9 × S P R I c l e a n (2 2 . 9 5 u l A m p u r e X P ビ ーズ)、1 2 u L の 1 × T E バ ッ フ ェ ー 中 に 溶 出

【 0 0 7 9 】

P C R

P C R 1 (オ リ ゴ タ グ プ ラ イ マ ー [発 見] ま た は 大 規 模 な プ ラ イ マ ー プ ール [デ ィ ー プ シ ー ク エ ン シ ン グ 検 証])

7 . 以 下 の マ ス タ ー ミ ッ ク ス を 調 製 す る :

ヌクレアーゼ非含有 H₂O 1 1 . 9 μ L

T a q ポ リ メ ラ ー ゼ 用 の バ ッ フ ェ ー、1 0 × (M g C l₂ な し) 3 . 0 μ L

d N T P ミ ッ ク ス、1 0 m M 0 . 6 μ L

M g C l₂、5 0 m M 1 . 2 μ L

10

20

30

40

50

Platinum Taqポリメラーゼ、5 U / μ l 0.3 μ L
 GSP1プライマー(10 μ M) / プライマープール(*) 1.0 μ L *
 TMAC(0.5 M) 1.5 μ L
 P5__1、10 μ M 0.5 μ L
 合計 20.0 μ L

+ (ステップ6からの) DNA 試料 10.0 μ L

合計 30.0 μ L

* 発見用には、+ / (センス) および - / (アンチセンス) 反応について別々のマスターミックスを作製して、別々のPCR反応を行う。

* ディープシーケンシング検証用には、1種のマスターミックスが作製される。プライマープールは、30 μ lの反応物中に30 pmolの総量に標準化されるべきである。

10

発見サーモサイクラープログラム(タッチダウン) :

95 5分間、

15サイクルの[95 30秒間、70 (-1 / サイクル) 2分間、72 30秒間]、

10サイクルの[95 30秒間、55 1分間、72 30秒間]、

72 5分間、

4 維持

検証サーモサイクラープログラム :

20

95 5分間、

14サイクルの[95 30秒間、20%の傾斜で65 まで下げる、65 5分間]、

72 5分間、

4 維持

8. 1.2 x SPRI clean (36.0 μ L)、15 μ lの1 x TEバッファー中に溶出

【0080】

PCR2(オリゴタグプライマー[発見]または大規模なプライマープール[ディープシーケンシング検証])

30

9. 以下のマスターミックスを調製する :

ヌクレアーゼ非含有H₂O 5.4 μ L

Taqポリメラーゼ用のバッファー、10 x (Mg²⁺ + なし) 3.0 μ L

dNTPミックス、10 mM 0.6 μ L

MgCl₂、50 mM 1.2 μ L

Platinum Taqポリメラーゼ、5 U / μ L 0.3 μ L

GSP2プライマー(10 μ M) / プライマープール(*) 1.0 μ L

TMAC(0.5 M) 1.5 μ L

P5__2、10 μ M 0.5 μ L

合計 13.5 μ L

40

+ P7__# (10 μ M) * 1.5 μ L

+ (ステップ8からの) ビーズ付きのDNA 試料 15.0 μ L

合計 30.0 μ L

プライマー濃度は、PCR1に記載の詳細に従うべきである

* P7__#では、Illuminaシーケンサー上での良好な画像記録のために、1つのシーケンシング泳動において少なくとも4種は使用されるべきである(例えば、P701~P704またはP705~P708)

発見サーモサイクラープログラム(タッチダウン) :

PCR1用と同じ

検証サーモサイクラープログラム :

50

PCR1用と同じ

10. 0.7 x SPRI clean (21.0 uL)、30 uLの1 x TEバッファ中に溶出

【0081】

qPCRによるライブラリー定量化およびシーケンシング

qPCR定量化

11. 製造業者の説明書に従って、Illuminaライブラリー定量化キット用のKapa Biosystemsキットを使用してライブラリーを定量化する。

【0082】

標準化およびシーケンシング

12. 各試料についてqPCR操作によって得られる1 uLあたりの分子の数の平均量推定値を使用して、全セットのライブラリーを 1.2×10^{10} 分子に標準化することを行ない、シーケンシング用に一緒にプールされるライブラリーの数によって分割する。これは、各試料についての分子によるインプット、およびさらに、各試料についての容量によるインプットを与えることになる。

プーリング後、Vacufugeでライブラリーをシーケンシング用に10 uLの最終容量までSpeedvac(乾燥減量)する。

ライブラリーを変性させ、Illumina Miseq試薬キットV2-300サイクル(2 x 150 bp対末端)でのシーケンシングのために、以下を除き、Illumina標準プロトコールに従ってMiseq上にロードする：

1) 3 uLの100 μMのカスタムシーケンシングプライマーIndex 1をMiseq試薬カートリッジ部位13に添加する(Indexプライマーミックス)。3 uLの100 μMのカスタムシーケンシングプライマーRead 2をMiseq試薬カートリッジ部位14に添加する(Read 2プライマーミックス)。

2) 対末端Nexteraシーケンシングプロトコールで以下の数のサイクル「151 | 8 | 16 | 151」でシーケンスする。

下流のバイオインフォマティクス解析のために関連する管路に対してbclまたはfastqフォーマットのいずれかでシーケンシングデータを出す。

【0083】

10

20

30

40

50

【表 2 - 1】

表3 GUIDE-Seqに必要とされる共通のプライマー

P7 アダプター	配列(5' → 3')	配列番号
P701	CAAGCAGAAGACGGCATAACGAGATTCGCCTTAGTGACTGGAGTCCCTCTCTATGG GCAGTCGGTGA	3
P702	CAAGCAGAAGACGGCATAACGAGATCTAGTACGGTGACTGGAGTCCCTCTCTATG GGCAGTCGGTGA	4
P703	CAAGCAGAAGACGGCATAACGAGATTTCTGCCTGTGACTGGAGTCCCTCTCTATGG GCAGTCGGTGA	5
P704	CAAGCAGAAGACGGCATAACGAGATGCTCAGGAGTGACTGGAGTCCCTCTCTATG GGCAGTCGGTGA	6
P705	CAAGCAGAAGACGGCATAACGAGATAGGAGTCCGTGACTGGAGTCCCTCTCTATG GGCAGTCGGTGA	7
P706	CAAGCAGAAGACGGCATAACGAGATCATGCCCTAGTGACTGGAGTCCCTCTCTATGG GCAGTCGGTGA	8
P707	CAAGCAGAAGACGGCATAACGAGATGTAGAGAGGTGACTGGAGTCCCTCTCTATG GGCAGTCGGTGA	9
P708	CAAGCAGAAGACGGCATAACGAGATCCTCTCTGGTGACTGGAGTCCCTCTCTATGG GCAGTCGGTGA	10
P5 アダプター	配列(5' → 3')	
P5_1	AATGATACGGCGACCACCGAGATCTA	11
P5_2	AATGATACGGCGACCACCGAGATCTACAC	12
カスタムシー クエンシング プライマー	配列(5' → 3')	
インデックス1	ATCACCGACTGCCCATAGAGAGGACTCCAGTCAC	13
リード2	GTGACTGGAGTCCCTCTCTATGGGCAGTCGGTGAT	14
Illumina Y アダ プター 1~16(分 子インデック スタグ NNWNNWNN 付き)	配列(5' → 3')	
Miseq 共通 アダプター	[Phos]GATCGGAAGAGC*C*A	15
A01	AATGATACGGCGACCACCGAGATCTACACTAGATCGCANNWNNWNNACACTCT TTCCCTACACGACGCTCTCCGATC	16
A02	AATGATACGGCGACCACCGAGATCTACACTCTCTATNNWNNWNNACACTCTT TCCCTACACGACGCTCTCCGATC*T	17
A03	AATGATACGGCGACCACCGAGATCTACACTATCCTCTNNWNNWNNACACTCTT TCCCTACACGACGCTCTCCGATC*T	18
A04	AATGATACGGCGACCACCGAGATCTACACAGAGTAGANNWNNWNNACACTCT TTCCCTACACGACGCTCTCCGATC*T	19
A05	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGNNWNNWNNACACTCT TTCCCTACACGACGCTCTCCGATC*T	20
A06	AATGATACGGCGACCACCGAGATCTACACTGCATANNWNNWNNACACTCT TTCCCTACACGACGCTCTCCGATC*T	21
A07	AATGATACGGCGACCACCGAGATCTACACAAGGAGTANNWNNWNNACACTCT TTCCCTACACGACGCTCTCCGATC*T	22
A08	AATGATACGGCGACCACCGAGATCTACACTAAGCCTNNWNNWNNACACTCT TTCCCTACACGACGCTCTCCGATC*T	23
A09	AATGATACGGCGACCACCGAGATCTACACGACATTGTNNWNNWNNACACTCT TTCCCTACACGACGCTCTCCGATC*T	24

10

20

30

40

【 0 0 8 4 】

50

【表 2 - 2】

A10	AATGATACGGCGACCACCGAGATCTACACACTGATGGNNWNNWNNACACTCT TTCCCTACACGACGCTCTCCGATC*T	25
A11	AATGATACGGCGACCACCGAGATCTACACGTACCTAGNNWNNWNNACACTCT TTCCCTACACGACGCTCTCCGATC*T	26
A12	AATGATACGGCGACCACCGAGATCTACACCAGAGCTANNWNNWNNACACTCT TTCCCTACACGACGCTCTCCGATC*T	27
A13	AATGATACGGCGACCACCGAGATCTACACCATAGTGANNWNNWNNACACTCT TTCCCTACACGACGCTCTCCGATC*T	28
A14	AATGATACGGCGACCACCGAGATCTACACTACCTAGTNNWNNWNNACACTCT TTCCCTACACGACGCTCTCCGATC*T	29
A15	AATGATACGGCGACCACCGAGATCTACACCAGATATNNWNNWNNACACTCT TTCCCTACACGACGCTCTCCGATC*T	30
A16	AATGATACGGCGACCACCGAGATCTACACTGGATTGTNNWNNWNNACACTCT TTCCCTACACGACGCTCTCCGATC*T	31
プライマー名	配列(5' → 3')	鎖/方向
Nuclease_off_+_GSP1	GGATCTCGACGCTCTCCCTATACCGTTATTAACATATGACA	+
Nuclease_off_-_GSP1	GGATCTCGACGCTCTCCCTGTTAATTGAGTTGTCATATGTTAATA AC	-
Nuclease_off_+_GSP2	CCTCTCTATGGGCAGTCGGTGATACATATGACAACCAATTA AAC	+
Nuclease_off_-_GSP2	CCTCTCTATGGGCAGTCGGTGATTGAGTTGTCATATGTTAATA AAC GGTA	-

* は、ホスホロチオエート結合修飾を示す

【 0 0 8 5 】

結果

例示的な G U I D E - S e q 法の概要

一部の実施形態では、G U I D E - S e q は、2つのステージからなる（図 5 B）：ステージ I では、ヒト生細胞のゲノム内の D S B が、これらの切断における平滑末端二本鎖オリゴデオキシヌクレオチド（d s O D N）の組み込みによってタグ付けされる。ステージ II では、バイアスのない増幅および次世代シーケンシングを使用して、ゲノム D N A 内の d s O D N 組み込み部位がヌクレオチドレベルで正確にマップされる。

【 0 0 8 6 】

ステージ I のために、本発明者らは、ヒト細胞において R G N で誘発された D S B の中に平滑末端の、5'リン酸化された d s O D N を組み込む条件を最適化した。最初の実験では、本発明者らは、こうした d s O D N が R G N で誘発された D S B 中に組み込まれるのを観察するのに失敗した。細胞内でオリゴを安定させるために設計された両方の D N A 鎖の 5'末端に2つのホスホチオレート結合を有する d s O D N²0 を使用して、本発明者らは、適度な検出可能な組み込み頻度だけを観察した（図 5 B）。しかし、両鎖の 3'末端でのホスホチオレート結合の付加は、強力な組み込み効率をもたらした（図 5 B）。これらの組み込み率は、これらの部位において R G N のみによって（すなわち、d s O D N の非存在下で）誘発されるインデルの頻度よりも2分の1から3分の1低いだけであった。

【 0 0 8 7 】

ステージ II のために、本発明者らは、組み込まれた d s O D N を有する断片のみを、バイアスのない様式で、選択的に増幅およびシーケンスすることを可能にする新規戦略を開発した（図 5 A）。本発明者らは、d s O D N、および R G N 成分をコードしているプラスミドがトランスフェクトされた細胞からの無作為に切断されたゲノム D N A に最初に「単一テール」次世代シーケンシングアダプターをライゲーションすることによって、これを達成した。本発明者らは、次いで、d s O D N に特異的にアニールする1つのプ

10

20

30

40

50

ライマーおよびシーケンシングアダプターにアニールするもう1つのプライマーによって開始される一連のPCR反応を行った(図5Aおよび図12)。シーケンシングアダプターが単一テールのみ付加されているので、これは、他の方法、例えば直線的増幅を介する(LAM)-PCR²¹、²²などに特有のバイアスなしで、dsODNに隣接した配列の特異的な一方向の増幅を可能にする。本発明者らは、我々の戦略を単一テールアダプター/タグ(single-tail adapter/tag:STAT)-PCR法と称する。dsODN鎖のそれぞれにアニールするプライマーを使用してSTAT-PCR反応を行うことにより、本発明者らは、それぞれの組み込まれたタグの両側上の隣接ゲノム配列の読みを得ることができた(図5C)。増幅プロセス中の無作為な8bpの分子バーコードの組み込み(図12)により、PCRバイアスの修正が可能となり、それによって、ハイスループットシーケンシングから得られた固有のシーケンシングの読みの正確な定量化が可能となる。

10

【0088】

ヒト細胞におけるCRISPR RGNのゲノムワイドなオフターゲット切断プロファイル

本発明者らは、U2OSまたはHEK293ヒト細胞系のいずれかにおいて、Cas9、および多様な内在性ヒト遺伝子を標的とする異なる10種のgRNAでGUIDE-Seqを行った(表1)。dsODN組み込み部位を解析すること(方法)によって、本発明者らは、ヌクレオチドレベルまでマップされた、10種のRGNのそれぞれによって誘発されたDSBの正確なゲノム部位を同定することができた(図5D)。これらのゲノムウィンドウのうち、>80%については、本発明者らは、オフターゲット部位であるまたはそれに関連するいずれかの重複標的配列を同定することができた(方法)。興味深いことに、各RGNについて本発明者らが同定したオフターゲット部位の総数は広範に変化し、ゼロから>150までの範囲に及び(図5E)、いずれの特定のRGNについても望ましくない切断のゲノムワイドな程度が極端に多くまたは少なくなりうることが示された。本発明者らは、(1から6つのミスマッチを有するゲノム部位の総数によって測定されるような)ヒトゲノムに対するgRNAプロトスペーサー配列の直交性と、GUIDE-Seqによって本発明者らが観察したオフターゲット部位の総数との間にいかなる明らかな相関も観察しなかった(図5F)。オフターゲット配列は、ゲノムの全体にわたって分散して見出され(図5Ggおよび図13A~J)、エクソン、イントロン、および非コード遺伝子間領域において減少する(図5H)。本発明者らが同定したオフターゲット配列の中に含まれていたのは、すべてRGNのうち4種についての既知の真正のオフターゲット部位4、5であった(図6A~J)。さらに重要なことには、GUIDE-Seqは、ヒトゲノムの全体にわたって位置する、多数の新しい、これまでに知られていないオフターゲット部位を同定した(図5E、5G、6A~Jおよび13A~J)。

20

30

【0089】

40

50

【表 3】

表 1			
標的部位名	細胞	配列	配列番号
EMX1	U2OS	GAGTCCGAGCAGAAGAAGAANGG	36
VEGFA 部位 1	U2OS	GGGTGGGGGGGAGTTTGCTCCNGG	37
VEGFA 部位 2	U2OS	GACCCCTCCACCCCGCCTCNGG	38
VEGFA 部位 3	U2OS	GGTGAGTGAGTGTGTGCGTGNGG	39
RNF2	U2OS	GTCATCTTAGTCATTACCTGNGG	40
FANCF	U2OS	GGAATCCCTTCTGCAGCACCNGG	41
HEK293 部位 1	293	GGGAAAGACCCAGCATCCGTNGG	42
HEK293 部位 2	293	GAACACAAAGCATAGACTGCNGG	43
HEK293 部位 3	293	GGCCCAGACTGAGCACGTGANGG	44
HEK293 部位 4	293	GGCACTGCGGCTGGAGGTGGNGG	45
トランケートされた VEGFA 部位 1	U2OS	GTGGGGGGGAGTTTGCTCCNGG	
トランケートされた VEGFA 部位 3	U2OS	GAGTGAGTGTGTGCGTGNGG	
トランケートされた EMX1	U2OS	GTCCGAGCAGAAGAAGAANGG	

10

20

【0090】

本発明者らは、次に、(図6A~Jに示される)GUIDE-Seqによって同定された各オフターゲット部位についてのシーケンシングの読みの数が、RGNのみによって(すなわち、dsODNの非存在下で)誘発されるはずであったインデルの相対頻度の代わりになっているかどうかを試験した。ヌクレアーゼ成分が発現されたヒトU2OS細胞における5種のRGNについての、アンカードマルチプレックスPCR(AMP)に基づいた次世代シーケンシング(方法)によるこれらの部位の試験より、>80%(132からの106)がRGN切断に特徴的な可変長のインデルを有していたことが示され、GUIDE-Seqが真正のRGNオフターゲット部位を同定するという本発明者らの結論がさらに支持される(図7A)。検出されたインデル頻度の範囲は、0.03%から60.1%までの範囲に及んだ。重要なことに、本発明者らは、5つすべてのRGNオフターゲット部位についてGUIDE-Seqの読み数とインデル変異頻度との間に陽性の直線的相関を認めた(図7A~F)。したがって、本発明者らは、所与の部位についてのGUIDE-Seqの読み数が、RGNによるその配列の切断効率の定量的尺度を表すと結論する。

30

【0091】

RGNで誘発されるオフターゲット配列特性の解析

10種すべてのRGNについてのGUIDE-Seqによって本発明者らが同定したオフターゲット部位の視覚的検査により、RGNがそこで切断可能なバリエーション配列の多様性が強調される。これらの部位は、プロトスペーサー配列内の6つものミスマッチ(最大7つのミスマッチを有する部位の*in vitro*の切断を示している以前の報告⁶と一致する)、非標準PAM(以前に記載されているNAGおよびNGA配列⁵、23だけでなく、新規なNAA、NGT、NGC、およびNCG配列も)、およびgRNA/プロトスペーサー境界面での1bpの「バルジ」型ミスマッチ²⁴を有しうる(図6A~J)。プロトスペーサーミスマッチは、標的部位の5'末端に生じる傾向にあるが、特定の3'末端部位でも見出されることがあり、部位に基づいてミスマッチ作用を予測するための簡単な法則はないという考えが支持される⁴。興味深いことに、一部のオフターゲット部位は、実際に、それらの適合したオンターゲット部位よりも高いシーケンシング読み数を

40

50

有し(図6A~D、6J)、オフターゲット変異頻度は、特定の場合において、目的のオンターゲット部位でのものよりも高くなりうるという本発明者らの以前の知見4と一致している。特に、RGNのうちの4種についての既知のオフターゲット部位の多くは、高い読み数を有し(図6A~D)、以前の解析は、最も効率良く切断される部位を主に同定していたことが示唆される。

【0092】

10種すべてのRGNにおける本発明者らのGUIDE-Seqデータの定量的解析は、本発明者らが異なる可変のもの、例えば、オフターゲット部位切断におけるミスマッチ数、位置、および型などの寄与および影響を定量化するのを可能にした。本発明者らは、RGNによって切断される特定の数のプロトスペーサーミスマッチを有する全ゲノム部位の割合は、ミスマッチの数の増大と共に低下することを見出した(図8A)。さらに、シーケンシングの読み数は、ミスマッチの数の増大と共に全体的な下降傾向を示す(図8B)。一般に、標的部位の5'末端のより近くに位置するプロトスペーサーミスマッチは、3'末端のより近いものよりもGUIDE-Seqの読み数におけるより小さい減少と関連する傾向にあるが、PAMから1から4bp離れて位置するミスマッチは、驚くべきことに、5から8bp離れて位置するものよりもいくらか良好な耐容性を示す(図8C)。興味深いことに、ミスマッチの性質も、GUIDE-Seqの読み数に対する影響に関連する。ゆらぎミスマッチは、オフターゲット部位において頻繁に生じ、本発明者らの解析は、それらが、他の非ゆらぎミスマッチよりもより小さい、GUIDE-Seqの読み数に対する影響に関連することを示唆している(図8D)。これらの結果と一致して、本発明者らは、一変量回帰分析においてオフターゲット切断における最も程度の大きいバリエーションを説明する単一因子が、ミスマッチ数、部位、および型であることを見出している。対照的に、近位PAM配列の密度、遺伝子発現レベル、またはゲノム位置(遺伝子間/イントロン/エクソン)などの他の因子は、GUIDE-Seq切断読み数においてきわめてより小さな割合の相違を説明する(図8E)。ミスマッチ部位、ミスマッチ型、遺伝子発現レベル、および近位PAM配列の密度を含む複数の因子を考慮した組み合わせられた線形回帰モデルは、一変量分析と一致した結果を得た(図14)。この分析により、本発明者らは、それぞれのさらなるゆらぎミスマッチがオフターゲット切断率を、平均において、かつそれらの部位に依存して、約2分の1~3分の1だけ低下させるのに対して、さらなる非ゆらぎミスマッチは切断率を約3分の1だけ低下させることを独立的に推定することも可能となった(図14)。

【0093】

既存のオフターゲット予測法でのGUIDE-Seqの比較

GUIDE-Seqの効力が確立されたので、本発明者らは、次に、オフターゲット変異部位を予測するための以下の2つの普及している既存のコンピュータの方法との本発明者らの新しい方法の直接的な比較を行った: MIT CRISPR Design Tool²⁵ (crispr.mit.edu) および E-CRISP プログラム²⁶ (www.e-crisp.org/E-CRISP/)。これらの両方のプログラムは、ミスマッチ数および部位についての特定の「法則」に基づいて潜在的なオフターゲット部位を同定することを試みるものであり、これまでの刊行物においてオフターゲット部位を同定するために使用されてきた。GUIDE-Seqによって本発明者らが特徴付けした10種のRGNを使用した本発明者らの比較において、本発明者らは、両方のプログラムが、実験的に検証されたオフターゲット部位のきわめて大多数を同定できなかったことを見出した(図9A~B)。これらの部位の多くは、E-CRISPおよびMITプログラムが、単に、それぞれ3つおよび4つを超えるミスマッチを有するオフターゲットを考慮しないので見落とされた(図9C~D)。考慮された配列の中でも、これらのプログラムは、依然として真正のオフターゲット部位の大多数を同定することができず(図9C~D)、切断が生じることになるまたは生じないことになるか否かを決定する因子を説明するためのこれらの現在限られた能力を強調している。MITプログラムによって割り当てられたランク付けスコアは、それが正確に同定する部位の中でいくらかの予測的能力を有するが

、特に、見落とされた部位が、わずか1つのミスマッチを有するものを含むことは留意すべきことである(図9C~D)。最後に、両方のプログラムが、GUIDE-Seqによって同定されない多くの「偽陽性」部位を出すことに留意することは重要である(図9A~B)。本発明者らは、MITおよびE-CRISPの両方が、真正のRGNオフターゲット部位の同定において本発明者らのGUIDE-Seq法よりも実質的に低い有効性で機能すると結論する。

【0094】

dCas9結合部位の決定についてのChIP-Seq法とのGUIDE-Seqの比較本発明者らは、RGNオフターゲット部位の同定について、GUIDE-Seqを、以前に記載されているChIP-Seq法と直接に比較することも探求した。本発明者らがGUIDE-Seqによって評価したRGNのうちの4つは、オフターゲット結合部位の大規模なセットの同定につながった、触媒的に不活性なCas9(dCas9)でのChIP-Seq実験において以前に特徴付けされていたgRNA¹⁸を使用した。直接的な比較により、GUIDE-Seqによって同定されたCas9オフターゲット切断部位と、ChIP-Seqによって同定されたdCas9オフターゲット結合部位との間できわめて少ない重複しか示されない；4種のgRNAについて本発明者らが同定した、RGNで誘発された149個のオフターゲット切断部位の中で3つだけが、以前に公表された同じgRNAを使用したCas9-ChIP-Seq実験によって以前に同定されているものであった(図9E)。この重複の欠如は、dCas9オフターゲット結合部位がCas9オフターゲット切断部位とは基本的に異なるからである可能性が高く、本発明者らのデータによって支持される仮説は、GUIDE-Seqによって同定されたこれらの4種のgRNAについてのCas9オフターゲット切断部位が、ChIP-Seq(図9F)および、きわめて少ないdCas9結合部位が活性Cas9の存在下でインデルの証拠を示すことを示している、以前の研究の結果^{16~19}によって同定されたそれらの結合部位よりも平均ではるかに少ないミスマッチを有することを示している。GUIDE-Seqは、ChIP-Seqによって以前に同定された4つのオフターゲット部位を同定することができず、かつ次いで、Cas9による変異生成の標的であることが示されたが、本発明者らは、これは、それらの部位がその以前の研究¹⁸において真正のオフターゲット切断部位として誤って同定されたからであると考えている。その研究からのシーケンシングデータの注意深い解析により、それらの部位において見出されたインデル変異のきわめて大多数は、RGN切断活性によってではなく、PCRまたはシーケンシングの誤りによって代わりに生じた可能性が高いことが示唆される(図15A~D)。まとめると、これらの発見により、GUIDE-Seqが、真正のオフターゲット切断部位の同定についてChIP-Seqよりも実質的に優れていることが実証され、ChIP-Seqによって発見された(たとえあるとしても)きわめて少ないdCas9オフターゲット結合部位が実際のCas9オフターゲット切断部位を表すという考えについての実験的支持が得られる。

【0095】

GUIDE-Seqによるヒト細胞におけるRGN非依存性DSBホットスポットの同定本発明者らのGUIDE-Seq実験は、本発明者らの研究のために使用されたU2OSおよびHEK293細胞における合計30個の固有のRGN非依存性DSBホットスポットの存在も予想外に明らかにした(表2)。本発明者らは、RGNをコードしているプラスミドなしでdsODNのみをトランスフェクトしたU2OSおよびHEK293細胞での対照実験からのゲノムDNAを解析するときには、これらの部位を明らかにした(方法)。特異的な塩基対部位に正確に位置するRGNで誘発されたDSBとは対照的に、RGN非依存性DSBは、それらが生じるそれぞれの座位においてより広く分散したdsODN組み込みパターンを有する(方法)。これらの30個の切断点ホットスポットは、多くの染色体にわたって分布していて、セントロメアもしくはテロメア領域にまたはその近くに存在するように見えた(図10F)。興味深いことに、少数のこれらのDSB(2つ)のみが両方の細胞系に共通であって、大多数は細胞系特異的であるようであった(U2

10

20

30

40

50

OSにおける25つおよびHEK293細胞における7つ；図10Fおよび表2）。知る限りでは、GUIDE-Seqは、ヒト生細胞において、直接的でかつバイアスのない、切断点ホットスポットの同定を、それらの存在を明らかにするために潜在的に有毒な薬剤（例えば、アフィジコリンなどのDNA複製阻害剤）を必要とすることなく、可能にする最初の方法である。

【0096】

【表4】

表2 ヒトU2OSおよびHEK293細胞におけるRGN非依存性の切断点ホットスポットのまとめ

細胞	染色体	開始	終了	間隔(bp)
U2OS	chr1	121484547	121485429	882
U2OS	chr1	236260170	236260754	584
U2OS	chr3	197900267	197900348	81
U2OS	chr4	191044096	191044100	4
U2OS	chr5	10020	10477	457
U2OS	chr7	16437577	16439376	1799
U2OS	chr7	158129486	158129491	5
U2OS	chr9	140249964	140249977	13
U2OS	chr9	140610510	140610516	6
U2OS	chr10	42599569	42599575	6
U2OS	chr11	129573467	129573469	2
U2OS	chr11	134946499	134946506	7
U2OS	chr12	95427	95683	256
U2OS	chr12	29944278	29946544	2266
U2OS	chr16	83984266	83984271	5
U2OS	chr17	63965908	63967122	1214
U2OS	chr18	63765	63769	4
U2OS	chr18	37381409	37381971	562
U2OS	chr2	9877829	9877857	28
U2OS	chr2	182140586	182140587	1
U2OS	chr2	209041635	209041637	2
U2OS	chr2	242838677	242838859	182
U2OS	chr22	49779897	49782342	2445
U2OS	chr22	49780337	49780338	1
U2OS	chrX	155260204	155260352	148
HEK293	chr1	121484526	121485404	878
HEK293	chr6	58778207	58779300	1093
HEK293	chr7	61968971	61969378	407
HEK293	chr10	42385171	42385189	18
HEK293	chr10	42400389	42400394	5
HEK293	chr10	42597212	42599582	2370
HEK293	chr19	27731978	27731991	13

【0097】

大規模なゲノム再編成におけるRGNで誘発されたDSBおよびRGN非依存性DSBの両方の関与

RGNで誘発されたDSBおよびRGN非依存性DSBにおけるインデルを同定するために設計された本発明者らの次世代シーケンシング実験の結果を解析する過程において、本発明者らは、これらの切断の一部が、転座、逆位および大きな欠失に関与していることも発見した。使用されたAMP法により、本発明者らがこれらの大規模なゲノムの変化を観察することが可能となった。これは、調査される各DSB部位について、この方法では

10

20

30

40

50

、1対の隣接座位特異的なプライマーではなく、ただ1つの固定された末端にアンカーされる、座位特異的なネステッドプライマーのみが使用されるからである(図10A)。したがって、AMPに基づいたシーケンシングは、DSBにおいてインデル変異が生じているかどうかを同定するだけでなく、DSBが別の配列に結合されているかどうかを検出することができる。

【0098】

本発明者らが調査した5種のRGNについて、AMPシーケンシングにより、RGNで誘発されるオンターゲットおよびオフターゲットのDSBが多様な転座に関与していることが明らかにされた(図10B)。少なくとも1つの場合において、本発明者らは、1対のDSBから結果として生じる4種すべての起こりうる転座事象を観察することができた(図10C)。2つのDSBが同じ染色体上に存在した場合に、本発明者らは、大きな欠失および逆位も観察した(図10B)。少なくとも1つの場合では、本発明者らは、RGNで誘発された2つの切断間の大きな欠失ならびに同じ介在性配列の逆位の両方を観察した(図10D)。重要なことに、本発明者らの結果は、RGNで誘発されたDSBとRGN非依存性DSBとの間の転座(および欠失または逆位)も明らかにし(図10B)、細胞ゲノムに対するRGNのオフターゲット作用を評価するときに、これらの2つの型の切断の間の相互作用が考慮される必要があることが示唆された。本発明者らのデータは、これらの大規模なゲノム再編成の頻度がきわめて低くなる可能性が高いことを示唆しているが、正確な定量化は、本発明者らの既存のデータセットのシーケンシングの深さではできなかった。シーケンシングの読みの数を増やすことにより、検出の感度が高められ、これらの重要なゲノムの変化のより良好な定量化が可能となるはずである。

【0099】

トランケートされたgRNAによって指向されるRGNのGUIDE-Seqプロファイル

本発明者らのグループによる以前の研究は、全長gRNAs 27によって指向されるRGNの既知のオフターゲット部位におけるトランケートされた17または18ntの相補性領域を有するgRNAの使用により、変異頻度が低下されうること示している27。しかし、この解析は少数の既知のオフターゲット部位に限定されていたので、これらのトランケートされたgRNA(tru-gRNA)のゲノムワイドな特異性は、本発明者らの以前の実験では不明確なままであった。本発明者らは、3種のtru-gRNAによって指向されるRGNのゲノムワイドなDSBプロファイルを得るためにGUIDE-Seqを使用した。そのそれぞれは、本発明者らが上記でアッセイした10種の全長gRNAのうちの1つより短いバージョンである。

【0100】

本発明者らの結果は、3つのすべての場合において、GUIDESeqによって同定されるオフターゲット部位の総数は、tru-gRNAの使用で実質的に減少されたことを示している(図11A~D)。GUIDE-Seqの読みのマッピングにより、本発明者らが、オンターゲット(図11E)およびオフターゲット(図示せず)部位の切断位置を正確に同定することが可能となった。予想されたように、かつ本発明者らが全長gRNAで観察したように、オフターゲット部位のリストに含まれていたのは、3種のtru-gRNAによって指向されるRGNについての12個の既知のオフターゲット部位のうち10個であった(図11F~H)。本発明者らが同定したオフターゲット部位の配列は、プロトスペーサー内に1つまたは2つのミスマッチを主に有していたが、一部の部位は、4つも有していた(図11F~H)。さらに、一部の部位は、NAG、NGA、およびNTGの形態の代替のPAM配列を有していた(図11F~H)。これらのデータは、gRNAのトランケーションがRGNのオフターゲット作用を実質的に低減できることのゲノムワイドな規模での確認を可能にし、RGNプラットフォームについての特異性の向上を評価するためにGUIDESeqが使用されうる方法を示す。

【0101】

考察

10

20

30

40

50

GUIDE - Seqは、RGNで誘発されたDSBを検出するための、バイアスのない、高感度、かつゲノムワイドな方法を可能にする。該方法は、オフターゲット部位の性質についての仮説をたてること（例えば、オフターゲット部位が配列においてオンターゲット部位に密接に関連すると推定すること）なくDSBを検出するのでバイアスがない。GUIDE - Seqは、エクソン、イントロン、および遺伝子間領域内を含む、オフターゲット部位をゲノムワイドに同定し、最大6つのプロトスペーサーミスマッチおよび/または、以前の研究⁵、²³に記載されている代替のNAGおよびNGA配列以外の、新しいミスマッチのPAM部位を有していた。この実施例において本発明者らが調査したRGNでは、GUIDE - Seqにより、すべての既知のオフターゲット部位の同定が成功しただけでなく、何百もの新しい部位も同様に明らかにされた。

10

【0102】

ヒト細胞においてすべてのRGNオフターゲット部位を包括的に同定するための実際的な黄金標準方法が現在ないことにより、本発明者らはGUIDE - Seqの確実性を有する感度を知ることができなくなっているが、本発明者らは、以下の理由で、それが低い偽陰性率を有する可能性がきわめて高いと考えている：第一に、RGNで誘発された平滑末端化されたすべてのDSBは、平滑末端化されたdsODNをNHEJによって取り込むはずであり、仮説は、GUIDE - Seqの読み数（これはdsODNの取り込みを測定する）とRGNの存在下におけるインデル頻度（これはDSB形成およびそれらのうちの変異誘発性修復の率を測定する）との間で本発明者らが観察する強い相関によって支持された（図7B～F）。本発明者らは、これらの相関が、広範囲のインデル変異生成頻度を示す130を超える部位を含むことを注記する。第二に、以前に同定されたオフターゲット部位をベンチマークとして使用して（これは、現在、成功を判断する唯一の方法である）、GUIDE - Seqは、0.12%の低さまで広がっている一範囲の変異生成頻度を示す40個のこれらの部位のうち38個を検出することができた。この方法は、4種の全長gRNAについての28個すべての既知のオフターゲット部位および3種のtrugRNAについての12個の既知のオフターゲット部位のうち10個を検出した。検出されなかった2つのオフターゲット部位のうち1つは、本発明者らの生のデータにおいて捕捉の証拠を示したが、シーケンシングの読みが一方向だけであってかつただ1つのプライマーに由来していたので本発明者らの読みの判定アルゴリズムによって除かれた（方法）。（この部位についての二方向のマッピングの読みがないのは、読みを正確にマップすることを難しくさせている片側のオフターゲット部位上の反復領域が原因である可能性があった。）他の検出されないオフターゲット部位は以前に記載されている。

20

30

【0103】

重要なことに、本発明者らが評価したRGNのうち1つは、（GUIDE - Seq法の現在の検出限界で）いかなる検出可能なオフターゲット作用も生じず、一部のgRNAがきわめて少ない望ましくない変異を誘導する、またはおそらく全く誘導しない、という興味深い可能性を高めた。

【0104】

本発明者らの検証実験は、GUIDE - Seqが、0.1%という低い頻度でRGNによって変異誘発されるオフターゲット部位を感度良く検出できることを示しているが、その検出能力は、いくらかの簡単な変更でさらに改良されうる。インデルを検出するために次世代シーケンシングを使用する戦略は、プラットフォームの誤り率（典型的には約0.1%）によって制限される。対照的に、GUIDE - Seqは、インデルではなくdsODN挿入部位を同定するためにシーケンシングを使用し、したがって、誤り率によって制限されるのではなく、シーケンシングの深さによって制限される。例えば、本発明者らは、本発明者らのシーケンシング検証実験において本発明者らがインデルを見出さなかった本発明者らのGUIDE - Seq実験において検出される少数の部位が、0.1%未満のインデル変異頻度を有する可能性の高い部位を実際に表すと考えている。これと一致して、本発明者らは、これらの26部位のうち3つ以外のすべては、100未満のGUIDE - Seq読み数を有したことを注記する。まとめると、これらの知見により、本

40

50

発明者らが、単にシーケンシングの読みの数を増加させることによって（および増幅のために鋳型として使用されるゲノムの数を増大させることによって）GUIDE-Seqの感度を増大させることができることが示唆される。例えば、1000倍多くの読みが得られるシーケンシングプラットフォームの使用により、検出が可能となるはずである。

【0105】

本発明者らのGUIDE-Seq実験によって可能となった直接的な比較により、RGNオフターゲット部位を予測するための2つの既存のコンピュータプログラムの限界が示される。これらのプログラムは、GUIDE-Seqによって見出される真正のオフターゲット部位を同定するのに失敗していただけでなく、切断を示さない多くの部位を過判定もしていた。これは、これらのプログラムによって使用されるパラメーターが、本発明者らのGUIDE-Seq実験によって同定されたより多数のプロトスペーサーミスマッチおよび代替のPAM配列を説明しないというオフターゲット部位の性質についてのより限定的な仮定に基づいていたのであれば全く驚くべきことではない。より良好な予測的プログラムが将来開発されうる可能性があるが、そうするためには、より多数のRGNについての実験的に決定されたゲノムワイドなオフターゲット部位が必要とされることになる。こうしたプログラムが開発されうるまで、オフターゲット部位の同定は、GUIDE-Seqなどの実験法によって最も効果的に対処されることになる。

【0106】

本発明者らの実験結果は、dCas9のオフターゲット結合部位とCas9のオフターゲット切断部位との間の明らかな相違を詳しく説明する。4種の異なるgRNAについてのdCas9のChIP-SeqおよびCas9のGUIDE-Seqのデータの比較により、2セットの部位間でごくわずかな直接的な重複があることおよび2クラスの部位におけるミスマッチの平均数が実際に実質的に異なることが示される。さらに、本発明者らは、Cas9によって変異誘発されると以前に報告されている少数のdCas9結合部位でさえも、RGNで誘発される真正の切断部位ではない可能性がきわめて高いことを示している。まとめると、本発明者らの結果は、ChIP-Seqで捕捉されているDNA部位に対するdCas9の結合が、Cas9ヌクレアーゼによるDNA部位の切断というよりも異なる生物学的プロセスを表していることを示し、プロトスペーサーとのgRNAの5'末端の結合が効率的な切断に必要とされることを示している近年の研究の結果⁹と一致している。ChIP-Seqアッセイは、dCas9融合タンパク質のゲノムワイドな結合の特徴付けにおける役割を疑いなく有するであろうが、その方法は、触媒的に活性化RGNのゲノムワイドなオフターゲット切断部位を決定するには明らかに有効でない。

【0107】

GUIDE-Seqは、細胞内のDSB部位を同定するための以前に記載されている他のゲノムワイドな方法を超えるいくつかの重要な利点を有する。最近記載されたBLESS (breaks labeling, enrichment on streptavidin and next-generation sequencing):

切断標識、ストレプトアビジン上での濃縮および次世代シーケンシング)オリゴヌクレオチドタギング法は、固定され、浸透性にされた細胞上で*in situ*で行われる²⁷。細胞固定に伴う人為的結果を生じやすいことに加えて、BLESSは、一瞬存在する切断物だけを捕捉することになる。対照的に、GUIDE-Seqは、生細胞上で行われ、より長期間の時間(日)にわたって生じるDSBを捕捉し、それによって感度がより高くかつ包括的なアッセイとなっている。DSBの近くの領域内への組み込み欠損レンチウイルス(IDLV)DNAの捕捉およびLAM-PCRによるそれらの座位の同定は、ヒト細胞において改変ジंकフィンガーヌクレアーゼ(ZFN)²²および転写活性化因子様エフェクターヌクレアーゼ(TALEN)²⁸についての少数のオフターゲット部位を同定するために使用されてきた。しかし、IDLV組み込み事象は、一般に数が少なく、かつ実際のオフターゲットDSBから500bpも遠くの距離にわたって広範に分散してお

10

20

30

40

50

り22、28、切断事象の位置を正確にマップすることおよび実際のオフターゲット部位の配列を推定することの両方を難しくさせている。さらに、LAM-PCRは、配列のバイアスおよび/またはシーケンシングの読みの低い効率を欠点としている。まとめると、これらの制約により、IDL V捕捉によるより低頻度のZFNオフターゲット切断部位を検出するのは明らかに不可能であることも説明されうる²⁹。対照的に、dsODNは、GUIDE-SeqでDSB内にきわめて効率良くかつ正確に組み込まれ、単一ヌクレオチド分解能での切断のマッピングおよびヌクレアーゼオフターゲット切断部位の簡単に直接的な同定を可能にする。さらに、LAM-PCRとは対照的に、本発明者らのSTAT-PCR法は、効率的な、バイアスのない増幅、およびdsODNが組み込まれているゲノムDNA断片のシーケンシングを可能にする。本発明者らは、STAT-PCRが、GUIDE-Seqにおけるその使用を超えたより一般的な有用性を有しうることを注記する；例えば、ゲノムワイドな規模でウイルスの組み込み部位をマップしようとすることは研究に有用でありうる。

10

【0108】

GUIDE-Seqは高感度であるが、その検出能力は、いくらかの簡単な変更でさらに改良しうる。インデルを検出するために次世代シーケンシングを使用する戦略は、プラットフォームの誤り率(典型的には約0.1%)によって制限される。対照的に、GUIDE-Seqは、インデルではなくdsODN挿入部位を同定するためにシーケンシングを使用し、したがって、誤り率によって制限されるのではなく、シーケンシングの深さによって制限される。例えば、本発明者らは、本発明者らのシーケンシング検証実験において本発明者らがインデルを見出さなかった本発明者らのGUIDE-Seq実験において検出される少数の部位が、0.1%未満の変異頻度を有する可能性の高い部位を実際に表すと考えている。これと一致して、本発明者らは、これらの26部位のうち3つ以外のすべては、100未満のGUIDE-Seq読み数を有したことを注記する。まとめると、これらの知見により、本発明者らが、単にシーケンシングの読みの数を増加させることによって(および増幅のために鋳型として使用されるゲノムの数を増大させることによって)GUIDE-Seqの感度を増大させることができうることを示唆される。例えば、1000倍多くの読みが得られるシーケンシングプラットフォームの使用により、位の3桁低い(すなわち、0.0001%)変異生成頻度を有する部位の検出が可能となるはずであり、本発明者らは、技術における継続した改良でさらなる増大が生じるのを期待する。

20

30

【0109】

本発明者らの実験の予想外の結果は、GUIDE-Seqが、RGNの非存在下でも細胞内で生じる切断点ホットスポットも同定することができたという認識であった。本発明者らのAMPに基づいたシーケンシング実験は、dsODNの捕捉だけでなくこれらの部位におけるインデルの形成も検証したので、本発明者らは、これらのDSBはGUIDE-Seqのただの人為的結果ではないと考えている。重要なことに、多くのホットスポットは本発明者らの研究において調査された2種の細胞系のそれぞれに固有であるが、いくつかは両方に共通しているようである。将来の研究において、1つの細胞型においていくつかの部位が切断点ホットスポットであるが別の細胞型ではそうではない理由を律するパラメーターを定義することは興味深いであろう。さらに、本発明者らの結果は、これらの切断点ホットスポットが転座に関与しうることを示しているので、細胞型特異的な切断点ホットスポットの存在は、特定のゲノム再編成が特定の細胞型のみが生じて他では生じない理由を説明するための一助となる可能性がある。知る限りでは、GUIDE-Seqは、ヒト生細胞において、DNA複製を阻害する薬剤の添加を必要とすることなく、切断点ホットスポットを同定できることが記載される最初の方法である²⁷。したがって、本発明者らは、それが、これらの切断を同定および試験するための有用なツールを提供することになると予想する。

40

【0110】

本発明者らの研究は、RGNによって誘発される転座を同定するためのこれまでに記載

50

された最も包括的な定性的アプローチを確立する。GUIDE-Seqによって発見されたRGN誘発性およびRGN非依存性のDSB部位のAMPに基づいた標的化されたシーケンシングは、両クラスの部位に関与する転座、欠失、および逆位を含む大規模なゲノム再編成を見出すことができ、大規模なゲノム再編成を同定するとき両クラスの切断を考慮することの重要性を強調している。さらに、おそらく必ずしもすべてのRGN誘発性またはRGN非依存性のDSBが大規模な変化に関与するわけではないであろうし、いくつかの部位がそれらの再編成に寄与して他の部位はしない理由を理解することは、さらなる研究のために重要な領域となるであろう。

【0111】

GUIDE-Seqは、ゲノムワイドな規模におけるRGNプラットフォームに対する特異性の改良を評価する重要な手段も提供することになる。この報告において、本発明者らは、トランケートされたgRNAの導入によってどのようにゲノム規模でオフターゲット作用を低減させることができるかを示すためにGUIDE-Seqを使用し、このアプローチが、適合された全長gRNAの既知のオフターゲット部位における変異を低減させうるといふ本発明者らのグループによる以前の結果³⁰を拡張した。これは、他の細菌もしくは古細菌からの代替のCas9ヌクレアーゼの、または5'オーバーハングを生成する二量体ZFN、TALEN、およびCRISPR RNA誘導型FokIヌクレアーゼ³¹、³²または5'もしくは3'オーバーハングを生成する対になったCas9ニッカーゼ³³、³⁴などのヌクレアーゼの、ゲノムワイドな特異性を評価するように適合されてもよい；しかし、GUIDE-Seqを拡張してこれらの他の型のDSBを検出するためには、こうした切断の中へのその効率的な捕捉を確実にするためのdsODNのさらなる修飾および最適化が疑いなく必要となるであろう。この方法は、他の細菌または古細菌からの代替のCas9ヌクレアーゼ³⁵の特異性を評価するためにも使用されう。本発明者らは、本発明者が評価した10種のgRNAについてのオフターゲット部位の数におけるきわめて広い変動性を見出したので、1つの重要な注意は、任意の新しいCas9プラットフォームの特異性についての結論を広く出す前に多数のgRNAを調査する必要性である。

【0112】

GUIDE-SeqおよびAMPに基づいたシーケンシングを使用する本発明者らの例示的なアプローチは、RGNによって誘発されるオフターゲット変異およびゲノム再編成の評価のための新しい黄金標準を確立する。本発明者らは、GUIDE-Seqが、NHUが活性でありかつ必要とされる成分が効率良く導入されうあらゆる細胞における使用のために拡張されうことを期待する；例えば、本発明者らは、ヒトK562およびマウス胚性幹細胞において効率的なdsODN組み込みをすでに達成している（データは示されていない）。最も重要なことに、ここに概説されている戦略は、厳しい前臨床経路の一部として、治療的使用のために提唱されるあらゆるRGNの潜在的なオフターゲット作用を客観的に評価して、それによって診療所におけるこれらの試薬の使用のための見込みを実質的に改善するために使用しう。

【実施例3】

【0113】

本方法の一部の実施形態において使用されうdsODNについての必要条件を探索するために、さらなる実験を行った。

【0114】

以下のdsODNsを、実施例3の実験において使用した：

【0115】

10

20

30

40

50

【表 5】

dsODN 型	配列	配列番号
リン酸化された、5'オーバーハングの、5'末端保護された F	/5Phos/N*N*NNGTTTAATTGAGTTGTCATATGTTAATAACGGT*A*T	47
リン酸化された、5'オーバーハングの、5'末端保護された R	/5Phos/N*N*N*NNATACCGTTATTAA CATATGACAACCTCAATTAA*A*C	48
リン酸化された、3'オーバーハングの、3'末端保護された F	/5Phos/G*T*T*TTAATTGAGTTGTCATATGTTAATAACGGTATNN*N*N	49
リン酸化された、3'オーバーハングの、3'末端保護された R	/5Phos/A*T*T*ACCGTTATTAACATATGACAACCTCAATTAAACNN*N*N	50
リン酸化された、平滑末端の、5'および3'末端保護された F	/5Phos/G*T*T*TTAATTGAGTTGTCATATGTTAATAACGGT*A*T	51
リン酸化された、平滑末端の、5'および3'末端保護された R	/5Phos/A*T*T*ACCGTTATTAACATATGACAACCTCAATTAA*A*C	52
リン酸化された、平滑末端の、3'末端保護された F	/5Phos/GTTTAATTGAGTTGTCATATGTTAATAACGGT*A*T	1
リン酸化された、平滑末端の、3'末端保護された R	/5Phos/ATACCGTTATTAACATATGACAACCTCAATTAA*A*C	2

/5Phos/は、5'リン酸化を示す

*は、ホスホロチオエート結合を示す

すべてのオリゴは、STE 中でアニールされた。

【0116】

最初に、EGFPに対して標的化されたTALEN、ZFN、およびRFNを使用して、3つの型のdsODNの組み込み頻度を評価した。2E5 U2OS-EGFP細胞は、500 ngの各TALENモノマー（合計1 ug）、500 ngの各ZFNモノマー（合計1 ug）、または325 ngのマルチプレックスgRNAプラスミドおよび975 ngのFokI-dCas9発現プラスミドならびに100 pmolのdsODNでヌクレオフェクトした。使用された3種のdsODNは、5'ホスホロチオエート結合を有する4 bpの5'オーバーハング、3'ホスホロチオエート結合を有する4 bpの3'オーバーハングのいずれかを有していた、または5'および3'ホスホロチオエート結合を有する平滑末端であった。すべてのdsODNは、5'リン酸化されていた。組み込み頻度は、NdeI制限断片長多型(RFLP)アッセイで推定し、キャピラリー電気泳動を使用して定量化した；簡単に説明すると、標的部位は、単離されたゲノムDNAからPCRによって増幅した。PCRは、NdeI制限酵素(20U)で37℃で3時間消化し、1.8x Ampure XPで精製した。精製された切断産物は、Qiaxcelキャピラリー電気泳動機器(Qiagen)によって泳動および定量化した。図16Aは、5'リン酸化され、かつ3'ホスホロチオエート化された、平滑末端化したdsODNが最も高い組み込み率を有していたことを示している。

【0117】

上記で使用された同じオリゴ(配列番号1および2)を、Lonza Nucleofector 4-D上で製造業者の説明書に従って20 μlの溶液SE(Lonza)中でU2OS細胞内にトランスフェクトした(プログラムDN-100)。500 ngの各TALENモノマー(CCR5についてはTAL1252/TAL1301かつAPCについてはTAL2294/2295)および100 pmolのdsODNをトランスフェ

クトした。図16B~Cは、(上記の)NdeI制限断片長多型(RFLP)解析またはT7E1アッセイによって決定されるような、U2OS細胞における2つの内在性標的部位、CCR5およびAPCでTALENによって誘発される二本鎖切断(DSB)の中への、平滑末端の、5'リン酸化された、34bpの二本鎖オリゴデオキシヌクレオチド(dsODN)(oSQT685/686)の効率的な組み込みの証拠を示している(簡単に説明すると、標的部位は、単離されたゲノムDNAからPCRによって増幅した。PCRは、1.8x Ampure XPで精製した。精製されたPCR産物(200ng)は、以下のプロトコールに従ってハイブリダイズした: 95 5分間、-2 /秒で95~85、-1 /10秒で85~25; 10 で維持する。T7エンドヌクレアーゼI(10U)を反応物に添加し、これを37 で15分間インキュベートした。反応物は、EDTA(25mM)の添加によって停止し、1.8x Ampure XPで精製した。精製された切断産物は、Qiaxcelキャピラリー電気泳動機器(Qiagen)によって泳動および定量化した)。

10

【0118】

さらなる実験を行い、2E5 U2OS-EGFP細胞を、325ngのマルチプレックスgRNAプラスミドおよび975ngのFokI-dCas9発現プラスミドならびに100pmolのdsODNでヌクレオフェクトした。さらに、3E5マウスES細胞を、200ngの単一gRNAプラスミドおよび600ngのCas9発現プラスミド、ならびに100pmolのdsODNでヌクレオフェクトした。2種のdsODNについて以下を比較した: 1)平滑末端、リン酸化、5'および3'ホスホロチオエート修飾および2)平滑末端、リン酸化、3'ホスホロチオエート修飾のみ。組み込み頻度は、NdeI制限断片長多型(RFLP)アッセイで推定し、キャピラリー電気泳動を使用して定量化した。

20

【0119】

実験は、ヒトU2OS細胞においてダイマーRNA誘導型FokIヌクレアーゼで(図17A)、またはマウスES細胞において標準Cas9で(図17B)を行い、3'ホスホロチオエート修飾のみを有するdsODNが最も高い割合の組み込みを有していたことが示された。

【0120】

マウスES細胞において異なる濃度の3'ホスホロチオエート修飾されたオリゴを試験するために、さらなる実験を行った。3E5マウスES細胞は、200ngの単一gRNAプラスミドおよび600ngのCas9発現プラスミド、ならびに下記のような異なる量のdsODNでヌクレオフェクトした。平滑末端の、リン酸化された、3'のみホスホロチオエート修飾されたdsODNをこの実験において使用した。精製dsODNと未精製dsODNとの間の比較では、アニールされたオリゴは、Sephadex G-25カラムを使用して精製した。dsODNは、1、2、5、10、25、50、および100pmolの濃度で試験した。組み込み頻度は、NdeI制限断片長多型(RFLP)アッセイで推定し、キャピラリー電気泳動を使用して定量化した。結果は、図18Aおよび18Bに示されており、50pmolまたは100pmolが最高の活性をもたらしたことを示した。Sephadex G-25カラムを通してのオリゴの精製は、割合を有意に高めなかった(図18Aおよび18Bを参照されたい)。変異生成頻度はT7E1アッセイによって推定し、これにより、3'修飾されたdsODNの存在下であっても、破壊の一般的な割合が高かったことが示された。

30

40

【0121】

dsODNの長さも評価した。図20A~Bは、より長い(例えば、60bp)dsODNタグが、CRISPR-Cas9で誘発されたDSBの部位で効率良く組み込まれたことを示している。これらのより長いdsODNは、PCR増幅の人為的結果のバイオフィンフォマティクスの選別を可能にすることによってGUIDE-seqの精度を向上させるために使用されうる。これらの配列は、より長いタグ内に存在する配列が含まれていなかったすべてのものとして認識されうる。

50

【 0 1 2 2 】

【表 6】

ssODN	配列	配列番号
oSQT1255	/5Phos/C*C*GCTTGCAGAGGGTATATTTGGTTAT CATATG GGACGAGTAGACTGAGATGAAGGTT*T*A	53
oSQT1256	/5Phos/T*A* AACCTTCATCTCAGTCTACTCGTCC CATATG ATAACCAAATATACCCTCTGCAAGC*G*G	54
oSQT1257	/5Phos/A*G*GACTGCATTCTTGATACTTAGACT CATATG TTCCTCTGGTACCGCGTAGATGTTT*A*C	55
oSQT1258	/5Phos/G*T* AAACATCTACGCGGTACCAGAGGAA CATATG AGTCTAAGTATACAAGAATGCAGTC*C*T	56
oSQT1259	/5Phos/A*C* CAATCAGTCACGAGCCTAGGAGATT CATATG GGTAAGAGAGTACATAATGCTTCC*G*G	57
oSQT1260	/5Phos/C*C*GGAAGCATTATGTGACTCTCTTACC CATATG AATCTCCTAGGCTCGTGACTGATTG*G*T	58

* は、ホスホロチオエート結合を示す

【 0 1 2 3 】

これらの実験により、dsODNタグ取り込みの効率は、5'および3'末端の両方よりもむしろ3'上のみで修飾されている、より長い、オリゴを使用することによって高められうること、および形質転換したがん細胞系由来ではない細胞（例えば、マウスES細胞）を含む多様な細胞系においてdsODNタグの効率的な捕捉が生じることが示される。

【実施例 4】

【 0 1 2 4 】

この実施例では、GUIDE-seq dsODNタグのビオチン化バージョンを、ゲノムのDSBの部位の中への組み込みのための基質として使用した。実施例4に示されているように、こうしたオリゴを効率良く組み込むことが可能であった。実験は、IDT DNAから得られた、ビオチン化dsODNを使用して、上記のように行った。

【 0 1 2 5 】

【表 7】

dsODN	配列	配列番号
oSQT1261	/5Phos/G*T*TTAATTGAG/iBiodT/TGTCATATG TTAATAACGGT*A*T	59
oSQT1262	/5Phos/A*T*ACCGTTA/iBiodT/TAA CATATG ACAACCTCAATTAA*A*C	60

iBiodT - ビオチンdTタグ

*は、ホスホロチオエート結合を示す

【 0 1 2 6 】

図19A~Bは、U2OS細胞における3つの内在性標的部位、VEGFA3、EMX1、およびFANCF1でCas9によって誘発される二本鎖切断(DSB)の中へのビオチン化されたdsODNタグの効率的な組み込みについての証拠を提供する。この向上により、ビオチンとストレプトアビジンの強固な結合親和性を利用することによるタグ付き断片の直接的な物理的捕捉を可能にしよう。(A)RFLP解析により、U2OS細胞における3つの内在性標的部位、VEGFA3、EMX1、およびFANCF1でCas9によって誘発されるDSBの中への標準dsODN(oSQT685/686)と比較

10

20

30

40

50

した、ビオチン化された dsODN (o S Q T 1 2 6 1 / 1 2 6 2) の組み込み率 (%) が示される。(b) T 7 E I により、U 2 O S 細胞における3つの内在性標的部位、V E G F A 3、E M X 1、および F A N C F 1 での標準 dsODN (o S Q T 6 8 5 / 6 8 6) と比較した、ビオチン化された dsODN (o S Q T 1 2 6 1 / 1 2 6 2) での推定変異生成頻度 (%) が示される。

【 0 1 2 7 】

ビオチン化が細胞内で維持されるとすると、これを使用して、ビオチン化された ssODN を含む DNA 断片を物理的にプルダウンすること、ならびに捕捉された断片をシーケンスおよびマップすることができる。

【 実施例 5 】

【 0 1 2 8 】

本実施例では、例示的な G U I D E - S e q 法を、バリエント Cas9 タンパク質と共に使用する。

【 0 1 2 9 】

バリエントの化膿レンサ球菌 (Streptococcus pyogenes) Cas9 (S p C a s 9) および黄色ブドウ球菌 (Staphylococcus aureus) Cas9 (S a C a s 9) タンパク質は、参照により本明細書に組み込まれている、米国特許出願公開第 6 1 / 1 2 7 , 6 3 4 号明細書および第 6 2 / 1 6 5 , 5 1 7 号明細書、ならびに Kleinstiver et al., "Engineered CRISPR-Cas9 nucleases with altered PAM specificities." Nature (2015) doi:10.1038/nature14592 に記載のように作製した。オフターゲット作用は、上記のように評価した。

【 0 1 3 0 】

図 2 1 は、E M X 1、F A N C F、R U N X 1、V E G F A、または Z N F 6 2 9 を標的とする sgRNA (配列については表 4 を参照されたい) を使用した、D 1 1 3 5 V / R 1 3 3 5 Q / T 1 3 3 7 R (V Q R バリエント) または D 1 1 3 5 V / G 1 2 1 8 R / R 1 3 3 5 E / T 1 3 3 7 R (V R E R バリエント) における変異を含む改変 SpCas9 バリエントについての G U I D E - s e q によって同定されたオフターゲット切断部位の数を示している。これは、G U I D E - s e q が、Cas9 の改変バージョンのゲノムワイドな特異性をプロファイルするのにも使用されうることを実証している。G U I D E - s e q は、ヒト細胞において N G A または N G C G P A M を含む内在性部位を標的とすることによって V Q R および V R E R SpCas9 バリエントの特異性プロファイルを決定するためにも使用した。

【 0 1 3 1 】

10

20

30

40

50

【表 8】

表 4					
名称	スペーサー長(nt)	スペーサー配列	配列番号	延長された PAM 付きの配列	配列番号
EMX1 NGA 4-20	20	GCCACGAAGCAGGCCAATGG	61	GCCACGAAGCAGGCCAATGGGGAG	62
FANCF NGA 1-20	20	GAATCCCTTCTGCAGCACCT	63	GAATCCCTTCTGCAGCACCTGGAT	64
FANCF NGA 3-20	20	GCGGCGGCTGCACAACCAGT	65	GCGGCGGCTGCACAACCAGTGGAG	66
FANCF NGA 4-20	20	GGTTGTGCAGCCGCCGCTCC	67	GGTTGTGCAGCCGCCGCTCCAGAG	68
RUNX1 NGA 1-20	20	GGTGCATTTTCAGGAGGAAG	69	GGTGCATTTTCAGGAGGAAGCGAT	70
RUNX1 NGA 3-20	20	GAGATGTAGGGCTAGAGGGG	71	GAGATGTAGGGCTAGAGGGGTGAG	72
VEGFA NGA 1-20	20	GCGAGCAGCGTCTTCGAGAG	73	GCGAGCAGCGTCTTCGAGAGTGAG	74
ZNF629 NGA 1-20	20	GTGCGGCAAGAGCTTCAGCC	75	GTGCGGCAAGAGCTTCAGCCAGAG	76
FANCF NGCG 3-20	20	GCAGAAGGGATTCCATGAGG	77	GCAGAAGGGATTCCATGAGGTGCG	78
FANCF NGCG 4-19	19	GAAGGGATTCCATGAGGTG	79	GAAGGGATTCCATGAGGTGCGCG	80
RUNX1 NGCG 1-19	19	GGGTGCATTTTCAGGAGGA	81	GGGTGCATTTTCAGGAGGAAGCG	82
VEGFA NGCG 1-20	20	GCAGACGGCAGTCACTAGGG	83	GCAGACGGCAGTCACTAGGGGCG	84
VEGFA NGCG 2-20	20	GCTGGGTGAATGGAGCGAGC	85	GCTGGGTGAATGGAGCGAGCAGCG	86

【 0 1 3 2 】

図 2 2 は、本明細書に記載の例示的な G U I D E - s e q 法を使用して検出されたオフターゲット部位での野生型と D 1 1 3 5 E S p C a s 9 バリエーションとの間の特異性における変化を示している。G U I D E - s e q は、3つの内在性のヒト細胞の部位での野生型 S p C a s 9 と D 1 1 3 5 E との間の読み数の相違を決定するためにも使用した。

【 0 1 3 3 】

G U I D E - s e q d s O D N タグ組み込みも、野生型および改変 C a s 9 D 1 1 3 5 E バリエーションで3つの遺伝子において行った。結果は、図 2 3 A ~ B に示されているが、G U I D E - s e q が改変 C a s 9 バリエーションのプロファイルに使用されうるといふさらなる証拠を示している。

【 0 1 3 4 】

参考文献

1. Sander, J.D. & Joung, J.K. CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol 32, 347-355 (2014).
2. Hsu, P.D., Lander, E.S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. Cell 157, 1262-1278 (2014).
3. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816-821 (2012).
4. Fu, Y. et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol 31, 822-826 (2013).
5. Hsu, P.D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol 31, 827-832 (2013).
6. Pattanayak, V. et al. High-throughput profiling of off-target DNA cleavage r

- reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* 31, 839-843 (2013).
7. Cradick, T.J., Fine, E.J., Antico, C.J. & Bao, G. CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res* 41, 9584-9592 (2013).
 8. Cho, S.W. et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res* 24, 132-141 (2014).
 9. Ghezraoui, H. et al. Chromosomal translocations in human cells are generated by canonical nonhomologous end-joining. *Mol Cell* 55, 829-842 (2014). 10
 10. Choi, P.S. & Meyerson, M. Targeted genomic rearrangements using CRISPR/Cas technology. *Nat Commun* 5, 3728 (2014).
 11. Gostissa, M. et al. IgH class switching exploits a general property of two DNA breaks to be joined in cis over long chromosomal distances. *Proc Natl Acad Sci U S A* 111, 2644-2649 (2014).
 12. Tsai, S.Q. & Joung, J.K. What's changed with genome editing? *Cell Stem Cell* 15, 3-4 (2014).
 13. Marx, V. Gene editing: how to stay on-target with CRISPR. *Nat Methods* 11, 1021-1026 (2014). 20
 14. Veres, A. et al. Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. *Cell Stem Cell* 15, 27-30 (2014).
 15. Smith, C. et al. Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell Stem Cell* 15, 12-13 (2014).
 16. Duan, J. et al. Genome-wide identification of CRISPR/Cas9 off-targets in human genome. *Cell Res* 24, 1009-1012 (2014). 30
 17. Wu, X. et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol* 32, 670-676 (2014).
 18. Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat Biotechnol* 32, 677-683 (2014).
 19. Cencic, R. et al. Protospacer Adjacent Motif (PAM)-Distal Sequences Engage CRISPR Cas9 DNA Target Cleavage. *PLoS One* 9, e109213 (2014).
 20. Orlando, S.J. et al. Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology. *Nucleic Acids Res* 38, e152 (2010). 40
 21. Schmidt, M. et al. High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR). *Nat Methods* 4, 1051-1057 (2007).
 22. Gabriel, R. et al. An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat Biotechnol* 29, 816-823 (2011).
 23. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L.A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* 31, 233-239 (2013). 50

24. Lin, Y. et al. CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Res* 42, 7473-7485 (2014).
25. Ran, F.A. et al. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* 8, 2281-2308 (2013).
26. Heigwer, F., Kerr, G. & Boutros, M. E-CRISP: fast CRISPR target site identification. *Nat Methods* 11, 122-123 (2014).
27. Crosetto, N. et al. Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat Methods* 10, 361-365 (2013).
28. Osborn, M.J. et al. TALEN-based gene correction for epidermolysis bullosa. *Mol Ther* 21, 1151-1159 (2013). 10
29. Sander, J.D. et al. In silico abstraction of zinc finger nuclease cleavage profiles reveals an expanded landscape of off-target sites. *Nucleic Acids Res* (2013).
30. Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M. & Joung, J.K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* 32, 279-284 (2014).
31. Tsai, S.Q. et al. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol* 32, 569-576 (2014).
32. Guilinger, J.P., Thompson, D.B. & Liu, D.R. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat Biotechnol* 32, 577-582 (2014). 20
33. Mali, P. et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* 31, 833-838 (2013).
34. Ran, F.A. et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154, 1380-1389 (2013).
35. Fonfara, I. et al. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Res* 42, 2577-2590 (2014). 30

【 0 1 3 5 】

他の実施形態

本発明は、その詳細な説明と併せて記載されているが、上記の記載は、例示することが意図されるものであり、添付の特許請求の範囲の範囲によって定義される、本発明の範囲を制限するものではないことは理解されるべきである。他の態様、利点および改変は、添付の特許請求の範囲の範囲内である。

本発明の様々な実施形態を以下に示す。

1. 細胞のゲノムDNA内の二本鎖切断(DSB)を検出するための方法であって、前記細胞を二本鎖オリゴデオキシヌクレオチド(dsODN)と接触させるステップであって、前記dsODNが好ましくは15および50ntの間の長さであり、前記dsODNの両鎖が前記細胞のゲノムにオルソロガスであり、好ましくは、前記dsODNの5'末端がリン酸化されており、さらにまた好ましくは、ホスホロチオエート結合が両方の3'末端に存在する、または2つのホスホロチオエート結合が両方の3'末端および両方の5'末端に存在する、ステップ、 40

前記細胞内で外来性の改変ヌクレアーゼを、前記ヌクレアーゼが前記細胞の前記ゲノムDNA内にDSBを誘発するのに十分な時間、かつ前記細胞が前記DSBを修復して、1つまたは複数のDSBにおいてdsODNを組み込むのに十分な時間、発現または活性化させるステップ、 50

組み込まれた dsODNを含むゲノムDNAの一部を増幅するステップ、および前記ゲノムDNAを増幅された部分をシーケンスするステップ

を含み、それによって前記細胞の前記ゲノムDNA内のDSBを検出する、方法。

2. 細胞のゲノムDNA内の二本鎖切断(DSB)を検出するための方法であって、前記細胞を二本鎖オリゴデオキシヌクレオチド(dsODN)と接触させるステップであって、前記dsODNが好ましくは50および75 ntの間の長さであり、前記dsODNの両鎖が前記細胞のゲノムにオルソログスであり、好ましくは、前記dsODNの5'末端がリン酸化されており、さらにまた好ましくは、ホスホロチオエート結合が両方の3'末端に存在する、または2つのホスホロチオエート結合が両方の3'末端および両方の5'末端に存在する、ステップ、

10

前記細胞内で外来性の改変ヌクレアーゼを、前記ヌクレアーゼが前記細胞の前記ゲノムDNA内にDSBを誘発するのに十分な時間、かつ前記細胞が前記DSBを修復して、1つまたは複数のDSBにおいてdsODNを組み込むのに十分な時間、発現または活性化させるステップ、

組み込まれたdsODNを含むゲノムDNAの一部を増幅するステップ、および前記ゲノムDNAを増幅された部分をシーケンスするステップ

を含み、それによって前記細胞の前記ゲノムDNA内のDSBを検出する、方法。

3. ゲノムDNAの一部を増幅するステップが、

前記DNAを断片化すること、

前記細胞からの断片化されたゲノムDNAの末端をユニバーサルアダプターでライゲーションすること、

20

前記組み込まれたdsODNと相補的なプライマー(プライマーA)および前記ユニバーサルアダプターと相補的なプライマー(プライマーB)を用いて、ライゲーションされたDNAに対して1回目のポリメラーゼ連鎖反応(PCR)を行うこと、

次いで、プライマーAと相補的な3'ネステッドプライマー(プライマーC)、プライマーBと相補的な3'ネステッドプライマー(プライマーD)、およびプライマーDと相補的なプライマー(プライマーE)を使用して2回目のPCRを行うこと

を含む、上記1または2に記載の方法。

4. プライマーEが、

精製もしくは結合配列、および/または同定配列

30

のうちの1つまたは複数を含む、上記3に記載の方法。

5. 前記改変ヌクレアーゼが、メガヌクレアーゼ、ジンクフィンガーヌクレアーゼ、転写活性化因子エフェクター様ヌクレアーゼ(TALEN)、およびクラスター化された規則的な間隔の短鎖反復回文配列(CRISPR)/Cas RNA誘導型ヌクレアーゼ(CRISPR/Cas RGN)からなる群から選択される、上記1または2に記載の方法。

6. 前記DSBがオフターゲットDSBである、上記1~5に記載の方法。

7. 前記DSBが、外来性の改変ヌクレアーゼによって誘発された、上記1~5に記載の方法。

8. 複数のガイドRNAのうちのいずれが最も特異的であるか、すなわち、最も少ないオフターゲットDSBを誘発するかを決定する方法であって、

40

第1の細胞の集団を第1のガイドRNAおよび二本鎖オリゴデオキシヌクレオチド(dsODN)と接触させるステップであって、前記dsODNが好ましくは15および50 ntの間の長さであり、前記dsODNの両鎖が前記細胞のゲノムにオルソログスであり、好ましくは、前記dsODNの5'末端がリン酸化されており、さらにまた好ましくは、ホスホロチオエート結合が両方の3'末端に存在する、または2つのホスホロチオエート結合が両方の3'末端および両方の5'末端に存在する、ステップ、

前記第1の細胞の集団内で外来性のCas9改変ヌクレアーゼを、前記ヌクレアーゼが前記細胞のゲノムDNA内にDSBを誘発するのに十分な時間、かつ前記細胞が前記DSBを修復して、1つまたは複数のDSBにおいてdsODNを組み込むのに十分な時間、発

50

現または活性化させるステップ、

組み込まれた dsODN を含む、前記第 1 の細胞の集団からのゲノム DNA の一部を増幅するステップ、および

前記第 1 の細胞の集団からの前記ゲノム DNA の増幅された部分をシーケンスするステップ、

前記 dsODN が前記第 1 の細胞の集団の前記ゲノム DNA 内に組み込まれた部位の数を決定するステップ、

第 2 の細胞の集団を第 2 のガイド RNA および二本鎖オリゴデオキシヌクレオチド (dsODN) と接触させるステップであって、前記 dsODN が好ましくは 15 および 50 nt の間の長さであり、前記 dsODN の両鎖が前記細胞のゲノムにオルソロガスであり、好ましくは、前記 dsODN の 5' 末端がリン酸化されており、さらにまた好ましくは、ホスホロチオエート結合が両方の 3' 末端に存在する、または 2 つのホスホロチオエート結合が両方の 3' 末端および両方の 5' 末端に存在する、ステップ、

前記第 2 の細胞の集団内で外来性の Cas9 改変ヌクレアーゼを、前記ヌクレアーゼが前記第 2 の細胞の集団のゲノム DNA 内に DSB を誘発するのに十分な時間、かつ前記細胞が前記 DSB を修復して、1 つまたは複数の DSB において dsODN を組み込むのに十分な時間、発現または活性化させるステップ、

組み込まれた dsODN を含む、前記第 2 の細胞の集団からのゲノム DNA の一部を増幅するステップ、および

前記第 2 の細胞の集団からの前記ゲノム DNA の増幅された部分をシーケンスするステップ、

前記 dsODN が前記第 2 の細胞の集団の前記ゲノム DNA 内に組み込まれた部位の数を決定するステップ、

前記 dsODN が前記第 1 の細胞の集団の前記ゲノム DNA 内に組み込まれた部位の数を、前記 dsODN が前記第 2 の細胞の集団の前記ゲノム DNA 内に組み込まれた部位の数と比較するステップ

を含み、より少ない (オフターゲット) 部位で組み込まれた前記 dsODN がより特異的である、方法。

9 . 複数のガイド RNA のうちのいずれが最も特異的であるか、すなわち、最も少ないオフターゲット DSB を誘発するかを決定する方法であって、

第 1 の細胞の集団を第 1 のガイド RNA および二本鎖オリゴデオキシヌクレオチド (dsODN) と接触させるステップであって、前記 dsODN が好ましくは 50 および 75 nt の間の長さであり、前記 dsODN の両鎖が前記細胞のゲノムにオルソロガスであり、好ましくは、前記 dsODN の 5' 末端がリン酸化されており、さらにまた好ましくは、ホスホロチオエート結合が両方の 3' 末端に存在する、または 2 つのホスホロチオエート結合が両方の 3' 末端および両方の 5' 末端に存在する、ステップ、

前記第 1 の細胞の集団内で外来性の Cas9 改変ヌクレアーゼを、前記ヌクレアーゼが前記細胞のゲノム DNA 内に DSB を誘発するのに十分な時間、かつ前記細胞が前記 DSB を修復して、1 つまたは複数の DSB において dsODN を組み込むのに十分な時間、発現または活性化させるステップ、

組み込まれた dsODN を含む、前記第 1 の細胞の集団からのゲノム DNA の一部を増幅するステップ、および

前記第 1 の細胞の集団からの前記ゲノム DNA の増幅された部分をシーケンスするステップ、

前記 dsODN が前記第 1 の細胞の集団の前記ゲノム DNA 内に組み込まれた部位の数を決定するステップ、

第 2 の細胞の集団を第 2 のガイド RNA および二本鎖オリゴデオキシヌクレオチド (dsODN) と接触させるステップであって、前記 dsODN が好ましくは 50 および 75 nt の間の長さであり、前記 dsODN の両鎖が前記細胞のゲノムにオルソロガスであり、好ましくは、前記 dsODN の 5' 末端がリン酸化されており、さらにまた好ましくは、

10

20

30

40

50

ホスホロチオエート結合が両方の3'末端に存在する、または2つのホスホロチオエート結合が両方の3'末端および両方の5'末端に存在する、ステップ、

前記第2の細胞の集団内で外来性のCas9改変ヌクレアーゼを、前記ヌクレアーゼが前記第2の細胞の集団のゲノムDNA内にDSBを誘発するのに十分な時間、かつ前記細胞が前記DSBを修復して、1つまたは複数のDSBにおいてdsODNを組み込むのに十分な時間、発現または活性化させるステップ、

組み込まれたdsODNを含む、前記第2の細胞の集団からのゲノムDNAの一部を増幅するステップ、および

前記第2の細胞の集団からの前記ゲノムDNAを増幅された部分をシーケンシングするステップ、

前記dsODNが前記第2の細胞の集団の前記ゲノムDNA内に組み込まれた部位の数を決定するステップ、

前記dsODNが前記第1の細胞の集団の前記ゲノムDNA内に組み込まれた部位の数を、前記dsODNが前記第2の細胞の集団の前記ゲノムDNA内に組み込まれた部位の数と比較するステップ

を含み、より少ない(オフターゲット)部位で組み込まれた前記dsODNがより特異的である、前記方法。

10. 前記細胞が哺乳動物細胞である、上記1~9のいずれかに記載の方法。

11. 前記改変ヌクレアーゼがCas9ヌクレアーゼであり、前記方法が、前記Cas9ヌクレアーゼを前記ゲノム内の標的配列に導くガイドRNAを前記細胞内で発現させるステップも含む、上記1~10のいずれかに記載の方法。

12. 前記dsODNが30~35ntの長さまたは60~65ntの長さである、上記のいずれかに記載の方法。

13. 前記dsODNがピオチン化される、上記のいずれかに記載の方法。

14. 前記ゲノムgDNAを断片に切断するステップ、およびdsODNを含む断片を前記ピオチンへの結合によって単離するステップを含む、上記13に記載の方法。

15. 前記dsODNが平滑末端である、上記のいずれかに記載の方法。

16. 前記dsODNが前記5'末端に1、2、3、または4ntのオーバーハンギングを有する、上記のいずれかに記載の方法。

17. 前記dsODNが、5'末端でリン酸化されており、3'末端でホスホロチオエート化されている、上記のいずれかに記載の方法。

18. 前記dsODNが、無作為化されたDNAバーコードを含む、上記のいずれかに記載の方法。

19. 上記のいずれかに記載の方法であって、前記ゲノムgDNAを断片に切断するステップ、および前記断片を、末端修復/αテーリング/単一テールのシーケンシングアダプターのライゲーションによってシーケンシング用に調製するステップを含む方法。

10

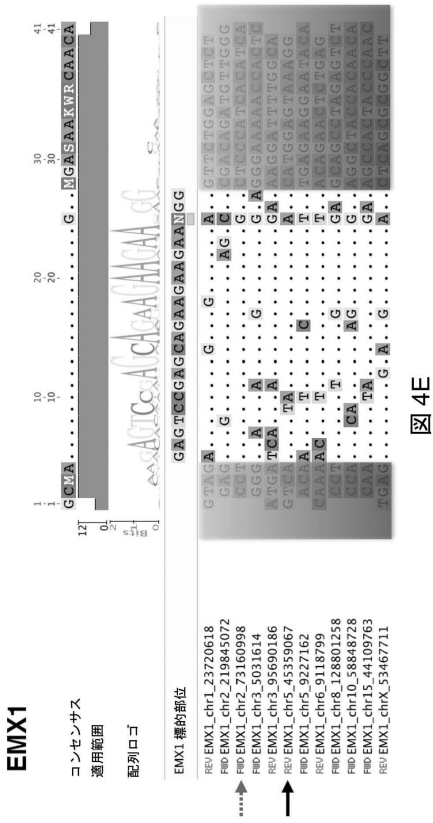
20

30

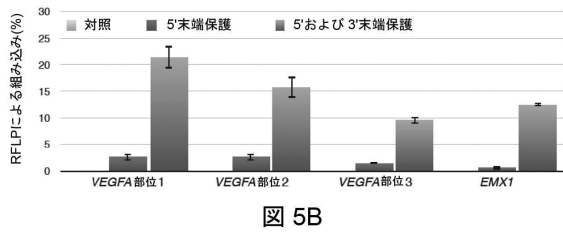
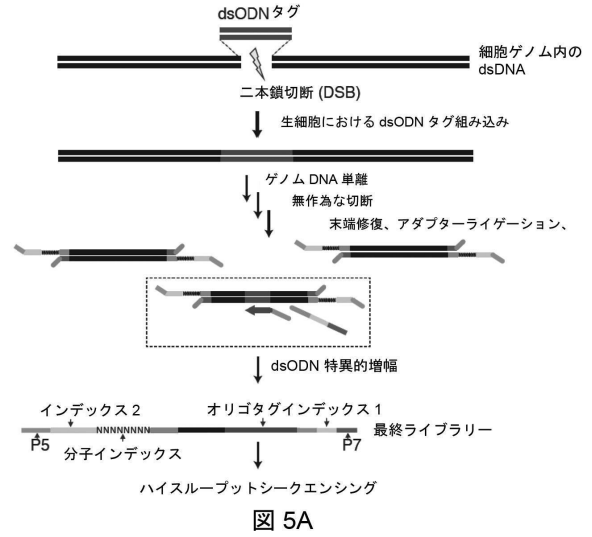
40

50

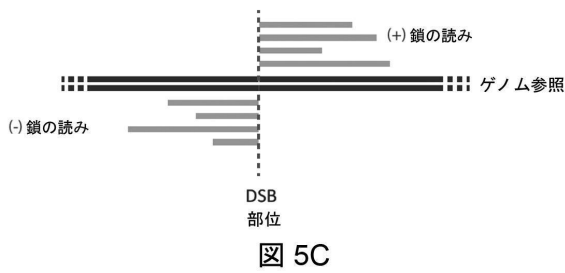
【 図 4 - 5 】



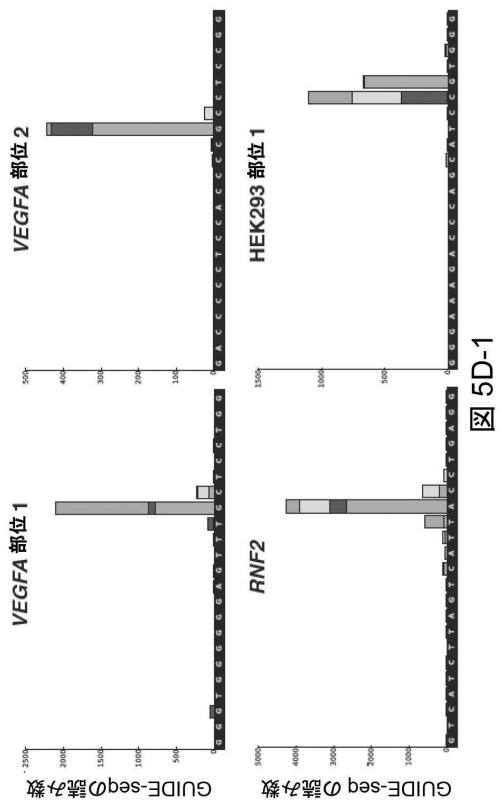
【 図 5 - 1 】



【 図 5 - 2 】



【 図 5 - 3 】



10

20

30

40

50

【 図 5 - 4 】

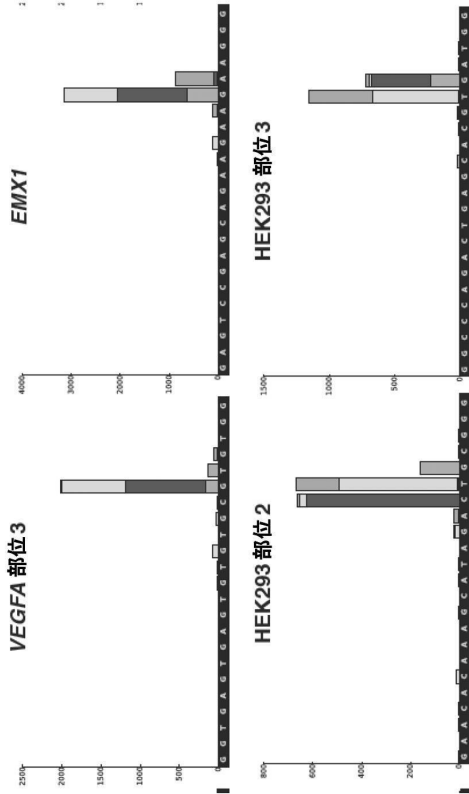


図 5D-2

【 図 5 - 5 】

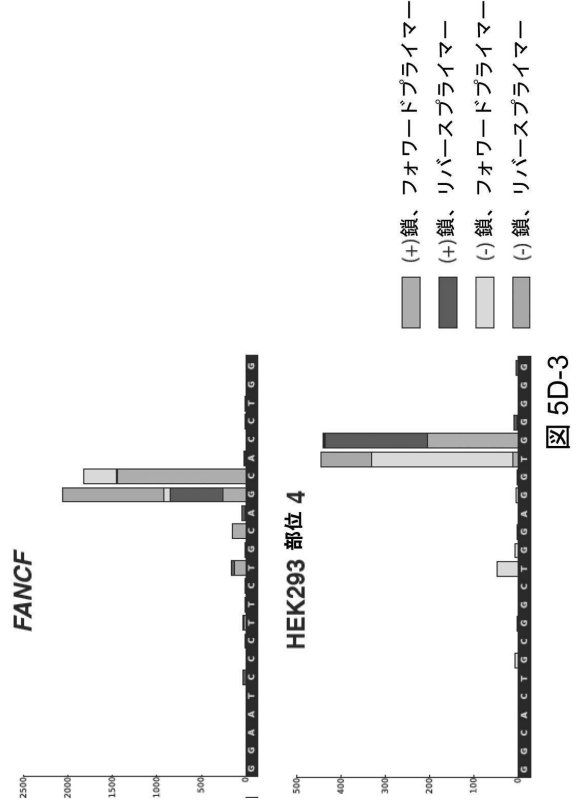


図 5D-3

【 図 5 - 6 】

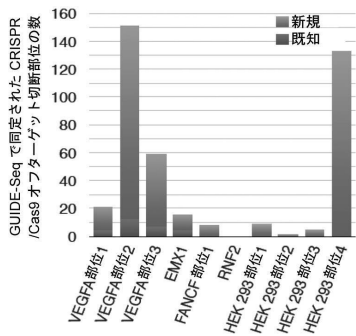


図 5E

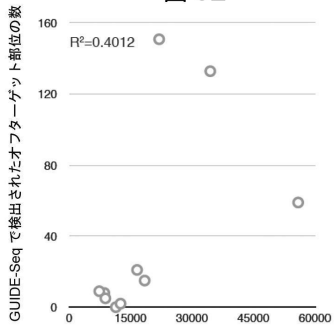


図 5F

【 図 5 - 7 】

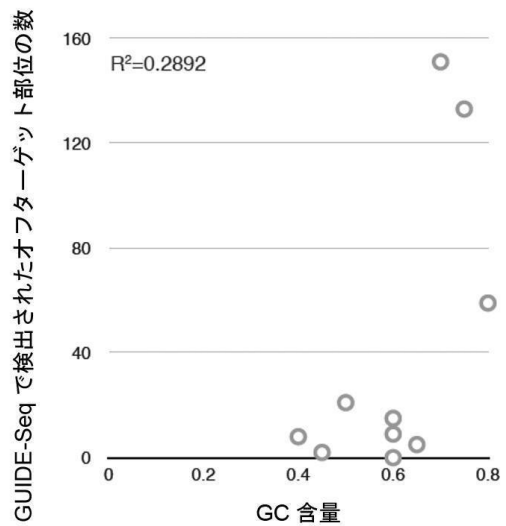


図 5G

10

20

30

40

50

【 図 5 - 8 】

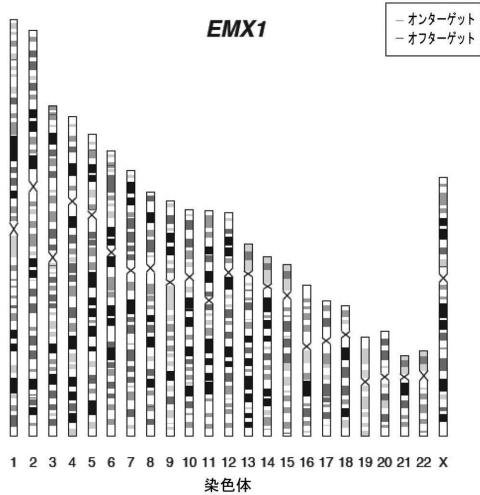


図 5H

標的	遺伝子間	エクソン	イントロン
VEGFA 部位 1	8	1	12
VEGFA 部位 2	39	24	88
VEGFA 部位 3	19	6	34
EMX1	9	1	5
FANCF	3	3	2
RNF2	0	0	0
HEK 293 部位 1	6	1	2
HEK 293 部位 2	1	0	1
HEK 293 部位 3	2	1	2
HEK 293 部位 4	39	20	74

図 5I

【 図 6 - 1 】

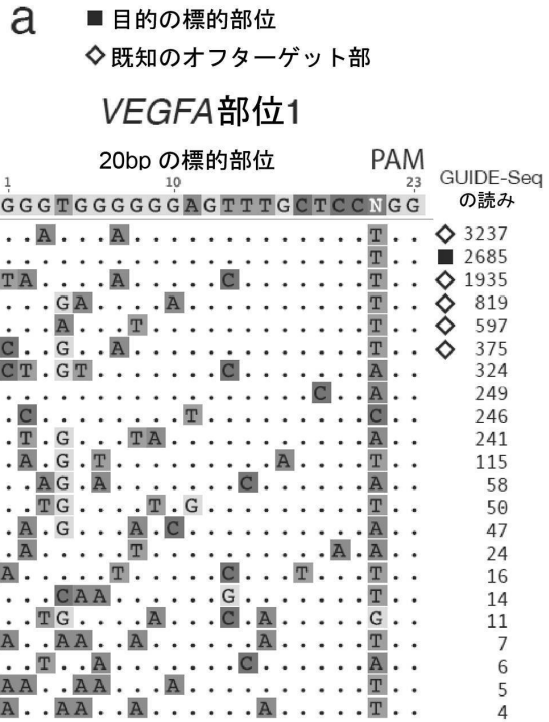


図 6A

【 図 6 - 2 】

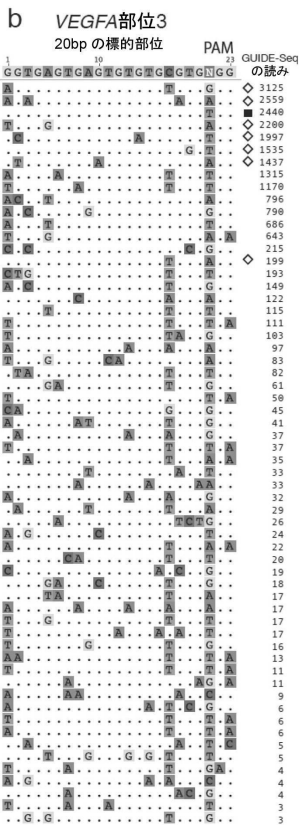


図 6B

【 図 6 - 3 】

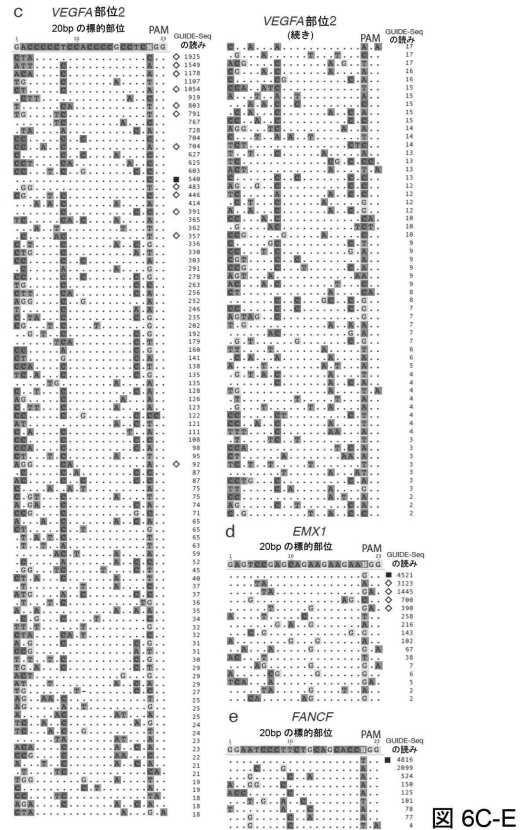


図 6C-E

10

20

30

40

50

【 図 9 - 1 】

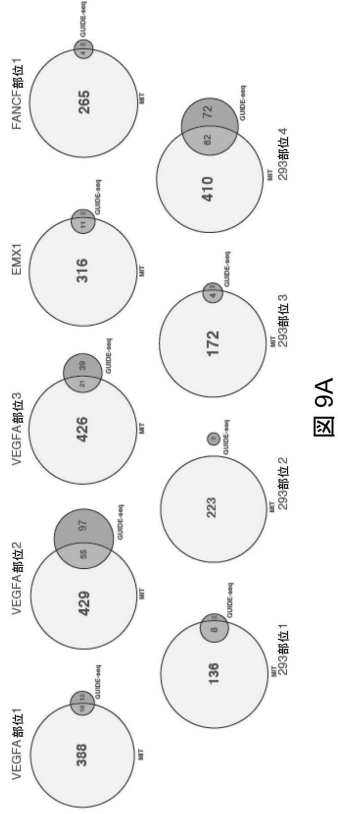


図 9A

【 図 9 - 2 】

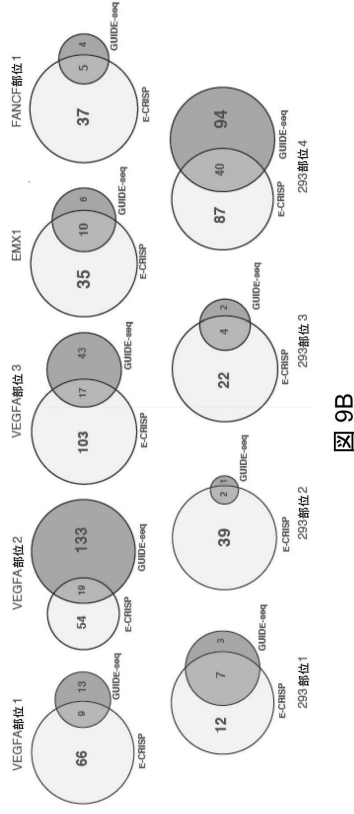


図 9B

【 図 9 - 3 】

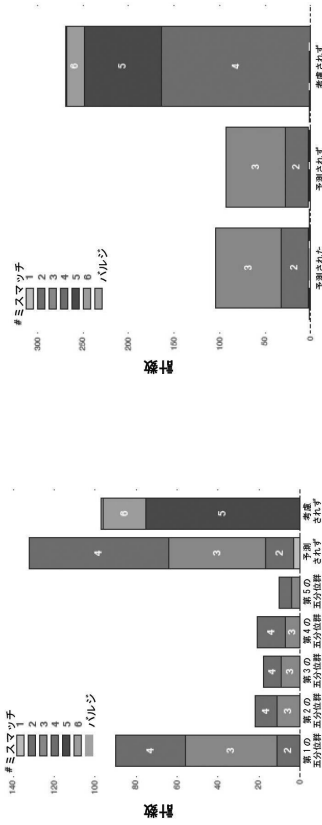


図 9C

図 9D

【 図 9 - 4 】

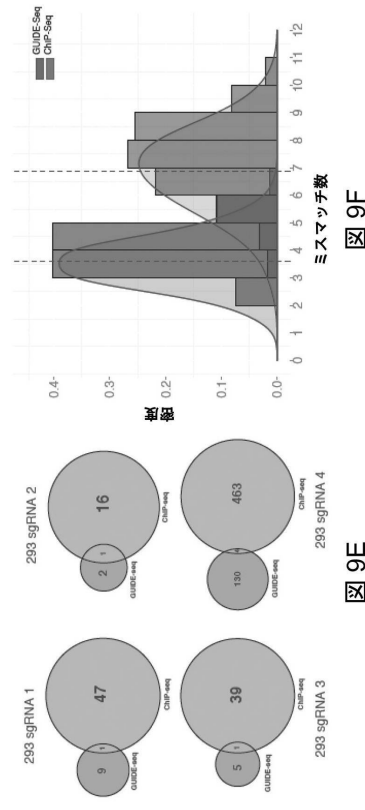
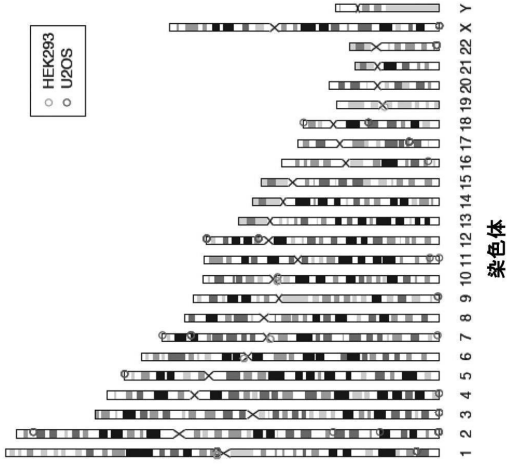


図 9E

図 9F

【図 10 - 1】



染色体
図 10A

【図 10 - 2】

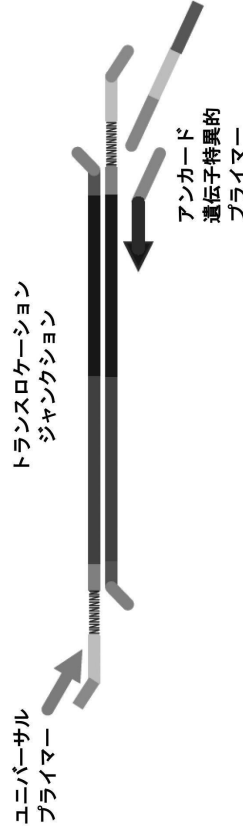


図 10B

【図 10 - 3】

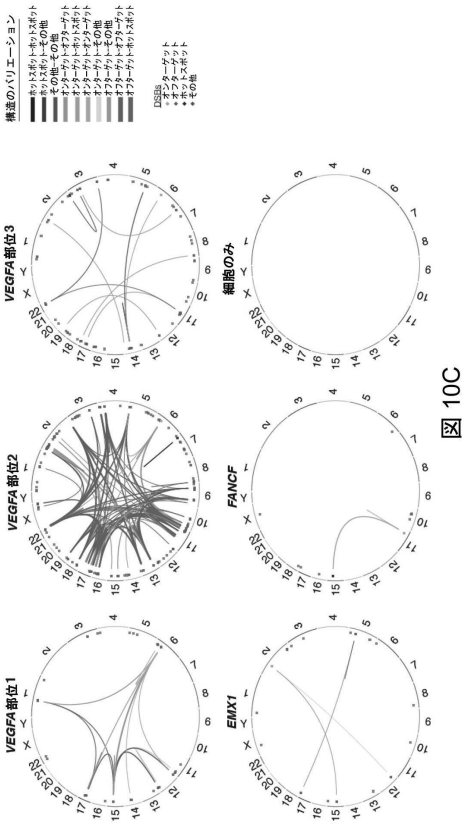


図 10C

【図 10 - 4】

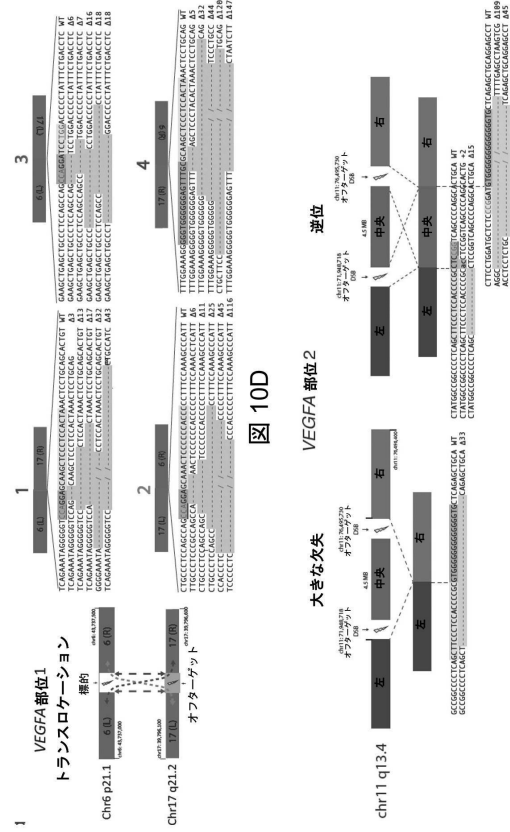


図 10D

図 10E

10

20

30

40

50

【 図 1 0 - 5 】

加減	スクレアーゼに阻害された構造要素(SV)		スクレアーゼ非依存性の構造要素(SV)		SVあたりのスプリットの数の平均値(最小, 最大)	
	152120-2122 X	発生	152120-2122 X	発生	平均値(最小, 最大)	平均値(最小, 最大)
Cas9のみ	0	0	0	0	3 (3, 3)	1 (1, 1)
細胞のみ	0	0	0	0	17 (4, 80)	2 (1, 3)
オリゴのみ	0	0	0	0	17 (4, 80)	2 (1, 3)
Cas9およびgRNA	0	1	0	0	17 (4, 80)	2 (1, 3)
Cas9およびtru-gRNA	0	0	0	0	17 (4, 80)	2 (1, 3)
Cas9およびgRNA	11	6	5	0	47.2 (3, 1282)	1.9 (1, 4)
Cas9およびtru-gRNA	86	24	35	1	47.2 (3, 1282)	1.9 (1, 4)
Cas9およびgRNA	9	2	1	0	7.1 (3, 24)	1.5 (1, 4)
Cas9およびtru-gRNA	9	2	1	0	7.1 (3, 24)	1.5 (1, 4)

図 10F

【 図 1 1 - 1 】

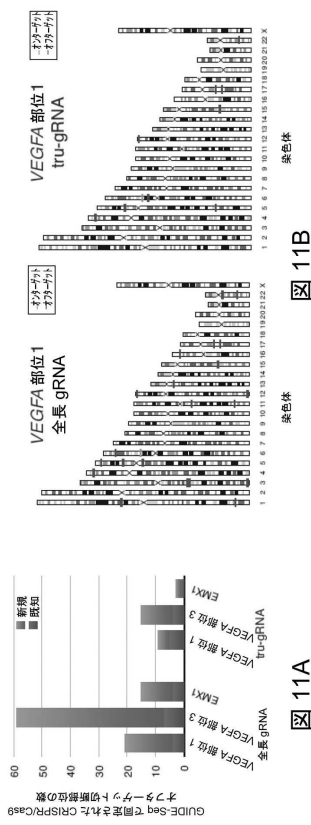


図 11A

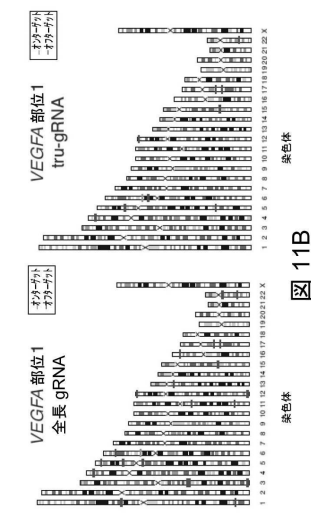


図 11B

【 図 1 1 - 2 】

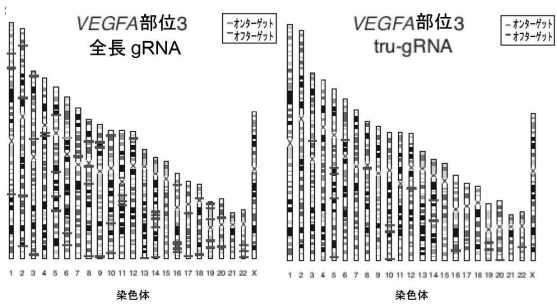


図 11C

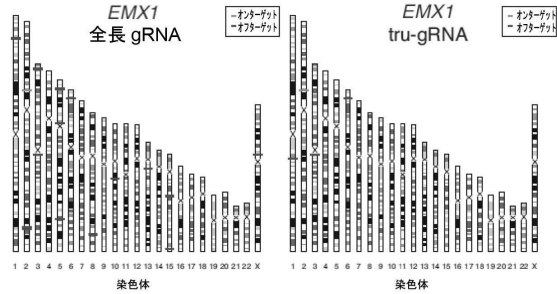


図 11D

【 図 1 1 - 3 】

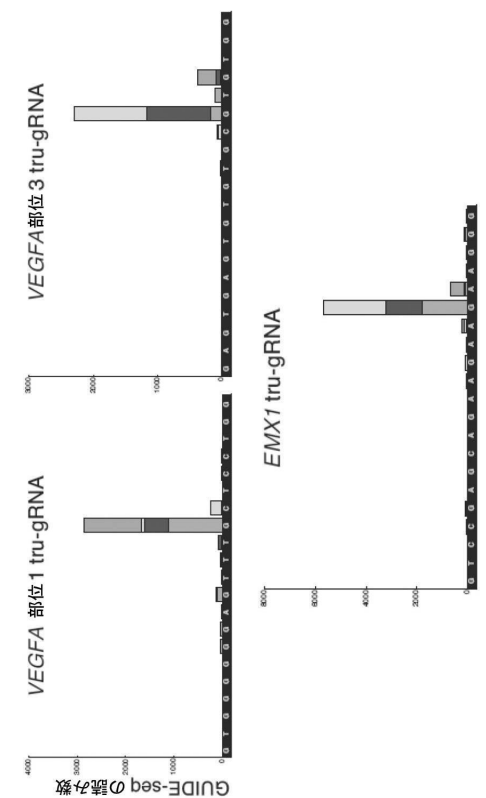


図 11E

10

20

30

40

50

【 図 13 - 3 】

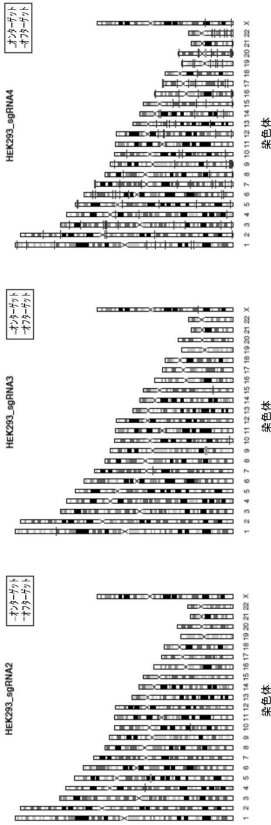


図 13I

図 13H

図 13G

【 図 13 - 4 】

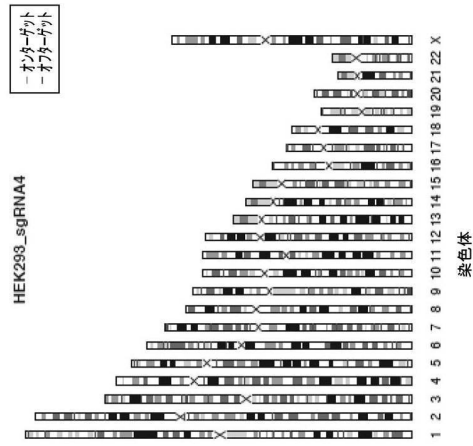


図 13J

10

20

【 図 14 】

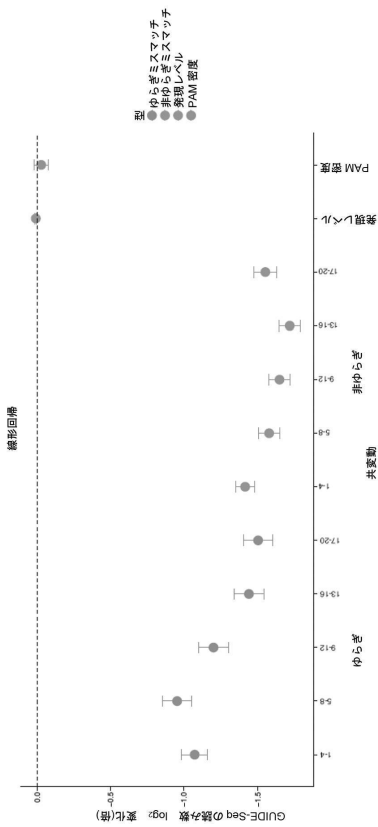
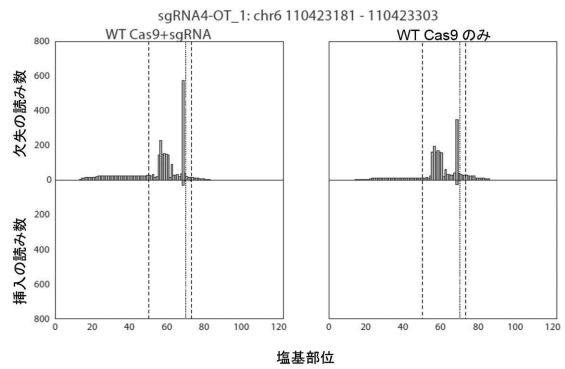
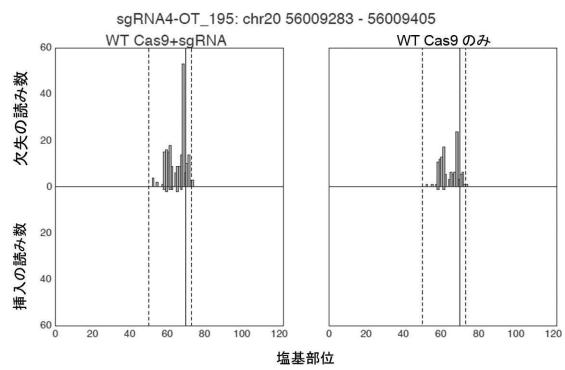


図 14

【 図 15 - 1 】



30

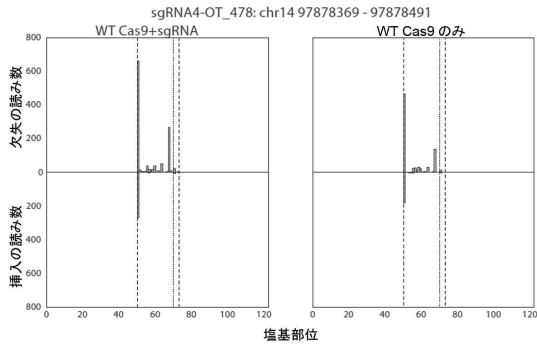


40

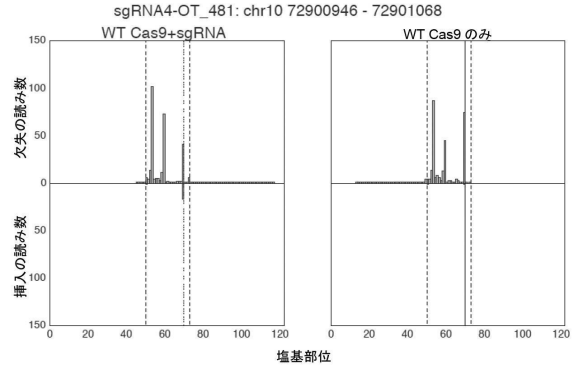
図 15A

50

【 図 15 - 2 】



【 図 15 - 3 】



10

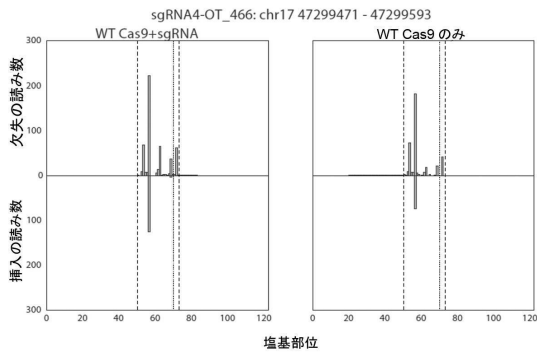


図 15B

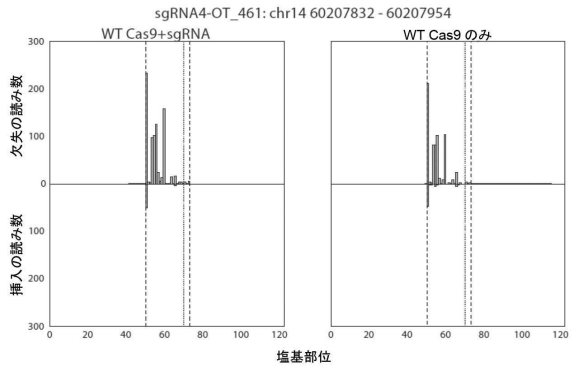


図 15C

20

【 図 15 - 4 】

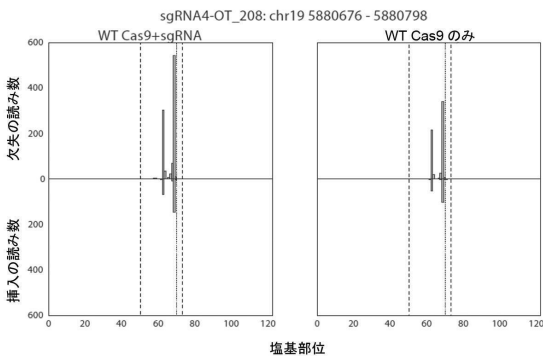


図 15D

【 図 16 - 1 】

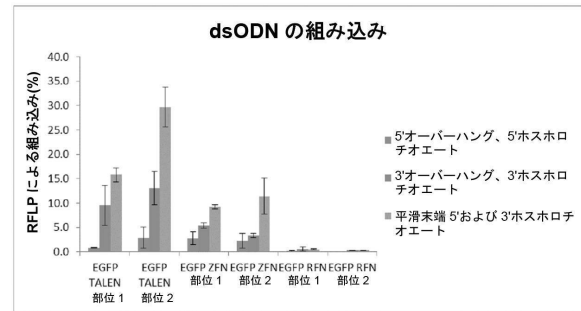


図 16A

30

40

50

【 図 16 - 2 】

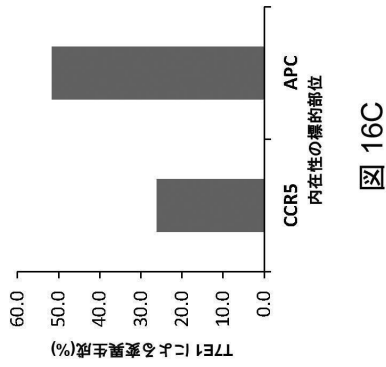


図 16C

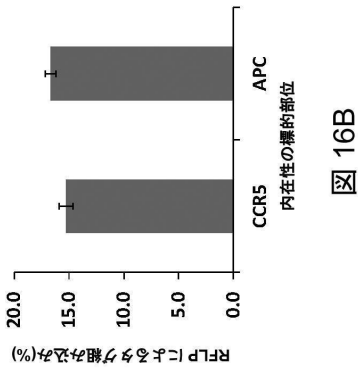


図 16B

【 図 17 】

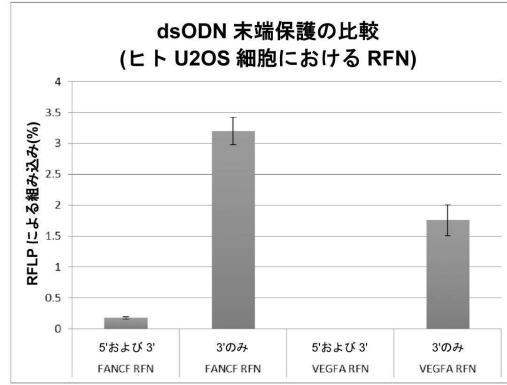


図 17A

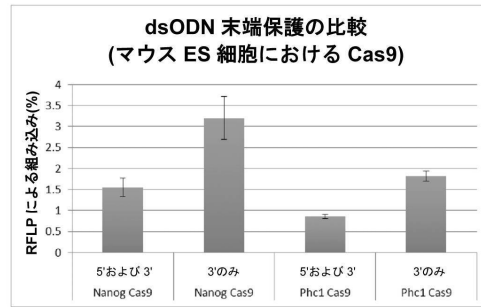


図 17B

【 図 18 - 1 】

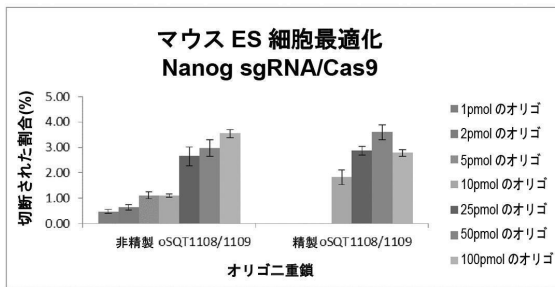


図 18A

【 図 18 - 2 】

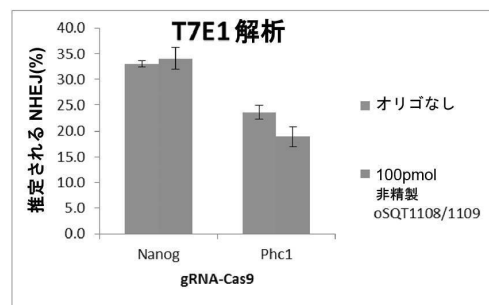


図 18C

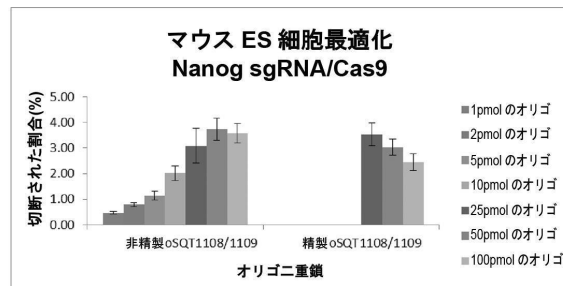


図 18B

10

20

30

40

50

【 図 19 】

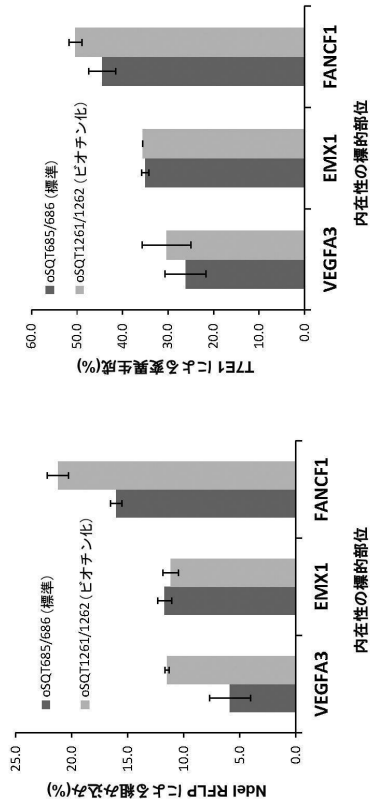


図 19B

図 19A

【 図 20 】

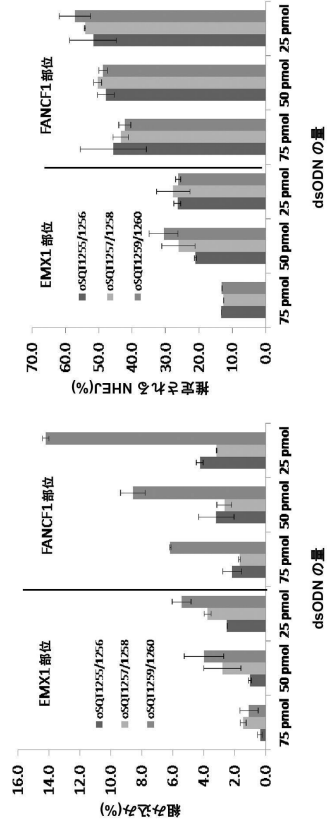


図 20B

図 20A

【 図 21 】

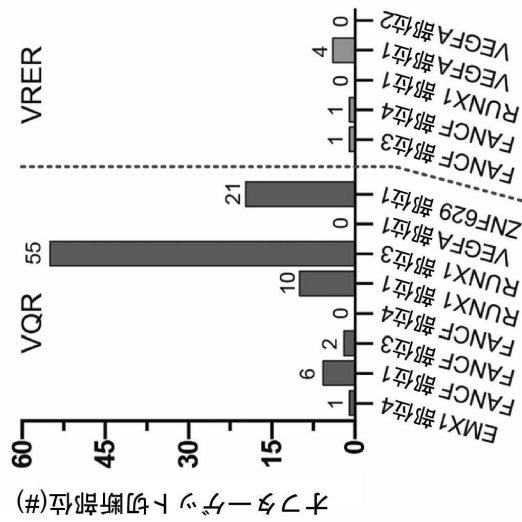


図 21

【 図 22 】

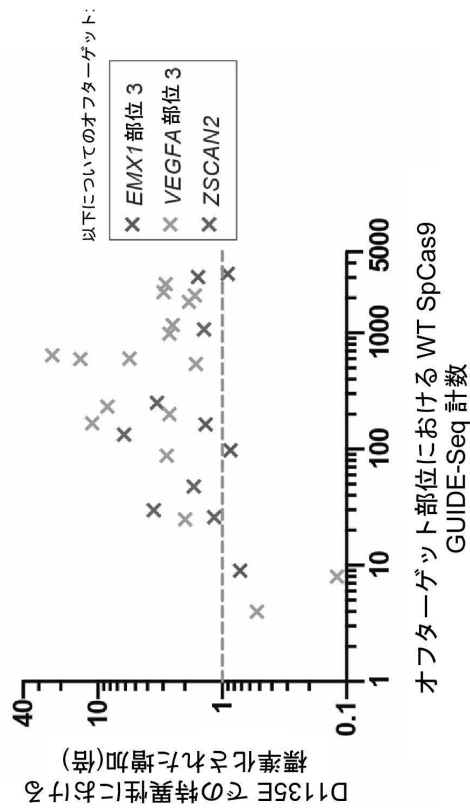


図 22

【図 23】

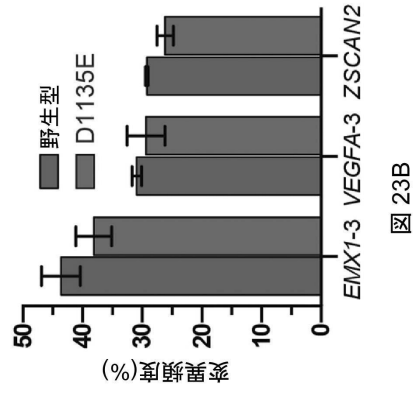


図 23B

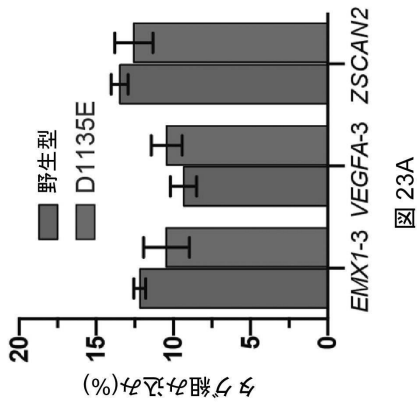


図 23A

【配列表】

0007095031000001.app

10

20

30

40

50

フロントページの続き

(33)優先権主張国・地域又は機関

米国(US)

(31)優先権主張番号 62/078,923

(32)優先日 平成26年11月12日(2014.11.12)

(33)優先権主張国・地域又は機関

米国(US)

(31)優先権主張番号 62/088,223

(32)優先日 平成26年12月5日(2014.12.5)

(33)優先権主張国・地域又は機関

米国(US)

ウィンチェスター, マグノリア ウェイ 1

(72)発明者 ツァイ, シェンダー

アメリカ合衆国 マサチューセッツ州 02129, チャールズタウン, アpartment 1, ブ
ロスペクト ストリート 32

審査官 田中 晴絵

(56)参考文献 米国特許出願公開第2013/0143204(US, A1)

特表2013-518602(JP, A)

特表2009-502202(JP, A)

米国特許出願公開第2005/0221341(US, A1)

米国特許出願公開第2005/0202490(US, A1)

米国特許出願公開第2012/0208705(US, A1)

特表2013-544498(JP, A)

特表2014-506788(JP, A)

国際公開第2014/008447(WO, A1)

nature protocols, 9(6), 2014年05月08日, 1255-81

(58)調査した分野 (Int.Cl., DB名)

C12N 15/00 - 15/90

C12Q 1/68 - 1/6897

CAplus/REGISTRY/MEDLINE/EMBASE/BIOSIS(ST
N)