

(19)日本国特許庁(JP)

(12)公表特許公報(A)

(11)公表番号

特表2022-525941

(P2022-525941A)

(43)公表日 令和4年5月20日(2022.5.20)

(51)国際特許分類	F I	テーマコード(参考)
C 1 2 N 15/09 (2006.01)	C 1 2 N 15/09 1 1 0	4 B 0 2 9
C 1 2 Q 1/6869(2018.01)	C 1 2 Q 1/6869 Z	4 B 0 6 3
C 1 2 M 1/00 (2006.01)	C 1 2 M 1/00 A	

審査請求 未請求 予備審査請求 未請求 (全62頁)

(21)出願番号	特願2021-556554(P2021-556554)	(71)出願人	597160510 リジェネロン・ファーマシューティカルズ・インコーポレイテッド REGENERON PHARMACEUTICALS, INC. アメリカ合衆国10591-6707 ニューヨーク州タリータウン、オールド・ソー・ミル・リバー・ロード777番
(86)(22)出願日	令和2年3月17日(2020.3.17)	(74)代理人	100105957 弁理士 恩田 誠
(85)翻訳文提出日	令和3年11月12日(2021.11.12)	(74)代理人	100068755 弁理士 恩田 博宣
(86)国際出願番号	PCT/US2020/023152	(74)代理人	100142907 弁理士 本田 淳
(87)国際公開番号	WO2020/190944	(74)代理人	100152489
(87)国際公開日	令和2年9月24日(2020.9.24)		
(31)優先権主張番号	62/820,106		
(32)優先日	平成31年3月18日(2019.3.18)		
(33)優先権主張国・地域又は機関	米国(US)		
(81)指定国・地域	AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA,RW,SD,SL,ST,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,RU,TJ,TM),EP(AL,A T,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC,		

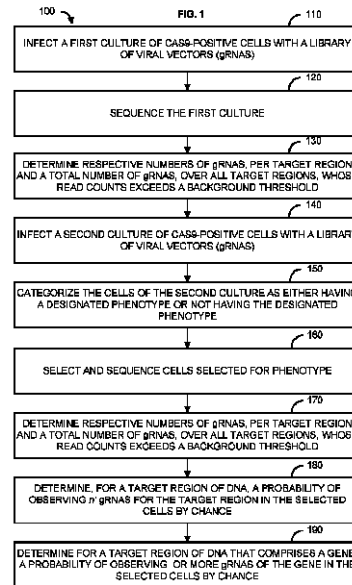
最終頁に続く

最終頁に続く

(54)【発明の名称】 CRISPR選択のための方法およびシステム

(57)【要約】

CRISPR陽性選択のための方法およびシステムが記載されている。CRISPR陽性選択は、DNA配列決定を使用して、CRISPRガイドRNAによるその振動が表現型と相関している遺伝子を識別する。



【特許請求の範囲】

【請求項 1】

(A) cas9 陽性細胞の第一の培養物をウイルスベクターのライブラリーで感染させることであって、前記ライブラリーが、前記細胞のゲノム内の DNA の標的領域を切断するための少なくとも 3 つのガイド RNA (gRNA) を含む、感染させること、前記細胞を配列決定して、前記 gRNA の各々のリード数を取得すること、DNA の標的領域当たりで、そのリード数が背景閾値を超える、gRNA のそれぞれの数を合計すること () であって、式中 $n = n$ である、合計すること、および、すべての標的領域にわたって、そのリード数が前記背景閾値を超える、gRNA の総数を合計すること () であって、式中 $N = N$ である、合計することと；

10

(B) 前記ウイルスベクターのライブラリーで、cas9 陽性細胞の第二の培養物を感染させること、前記第二の培養物の細胞を、指定された表現型を有する、または前記指定された表現型を有しないのいずれかとしてカテゴリ化すること、前記指定された表現型を有する細胞を選択し、前記選択された細胞を配列決定して前記 gRNA の各々の選択後リード数を取得すること、その選択後リード数が前記背景閾値を超える前記選択された細胞において、DNA の標的領域当たりで、gRNA のそれぞれの数を合計すること () であって、式中 $n' = n'$ である、合計すること、および、

20

すべての標的領域にわたって、そのリード数が前記閾値を超える前記選択された細胞の、gRNA の総数を合計すること () であって、式中 $N' = N'$ である、合計することと；

式

【数 1】

$$\frac{\binom{n'}{N'} \cdot \binom{N'-n'}{N-n'}}{\binom{N'}{N}}$$

に従って、前記選択された細胞内の前記標的領域について n' gRNA を偶然に観察する確率を計算することであって、式中

30

【数 2】

$$\binom{x}{y} = \frac{y!}{x!(y-x)!}$$

は y オブジェクトから x オブジェクトを選択する方法の数を算出する、計算すること、および、

遺伝子を含む DNA の標的領域について、式

【数 3】

$$\sum_{i=n'}^n \frac{\binom{i}{N'} \cdot \binom{N'-i}{N-i}}{\binom{N'}{N}}$$

40

に従って、前記選択された細胞内の遺伝子の n' 以上の gRNA を偶然に観察する確率を計算することと、を含む、方法。

【請求項 2】

(A) cas9 陽性細胞の第一の培養物をウイルスベクターのライブラリーで感染させることであって、前記ライブラリーが、前記細胞のゲノム内の DNA の標的領域の転写を強化するための少なくとも 3 つのガイド RNA (gRNA) を含む、感染させること、

50

前記細胞を配列決定して、前記 g R N A の各々のリード数を取得すること、
D N A の標的領域当たりで、そのリード数が背景閾値を超える、g R N A のそれぞれの数を合計すること () であって、式中 $n = n$ である、合計すること、および、
すべての標的領域にわたって、そのリード数が前記背景閾値を超える、g R N A の総数を合計すること () であって、式中 $N = N$ である、合計することと；

(B) 前記ウイルスベクターのライブラリーで、c a s 9 陽性細胞の第二の培養物を感染させること、

前記第二の培養物の前記細胞を、指定された表現型を有する、または前記指定された表現型を有しないのいずれかとしてカテゴリズすること、

前記指定された表現型を有する細胞を選択し、前記選択された細胞を配列決定して前記 g R N A の各々の選択後リード数を取得すること、

その選択後リード数が前記背景閾値を超える前記選択された細胞において、D N A の標的領域当たりで、g R N A のそれぞれの数を合計すること () であって、式中 $n = n'$ である、合計すること、および、

すべての標的領域にわたって、そのリード数が前記閾値を超える前記選択された細胞の、g R N A の総数を合計すること () であって、式中 $N = N'$ である、合計することと；

(C) D N A の標的領域について、

前記式

【数 4】

$$\frac{\binom{n'}{N'} \cdot \binom{N'-n'}{N-n'}}{\binom{N'}{N}}$$

20

に従って、前記選択された細胞内の前記標的領域について n' g R N A を偶然に観察する確率を計算することであって、式中

【数 5】

$$\binom{x}{y} = \frac{y!}{x!(y-x)!}$$

30

は、 y オブジェクトから x オブジェクトを選択する方法の数を算出する、計算すること、および、

遺伝子を含む D N A の標的領域について、前記式

【数 6】

$$\sum_{i=n'}^n \frac{\binom{i}{N'} \cdot \binom{N'-i}{N-i}}{\binom{N'}{N}}$$

40

に従って、前記選択された細胞内の遺伝子の n' 以上の g R N A を偶然に観察する確率を計算することと、を含む、方法。

【請求項 3】

前記第一の培養物を、C R I S P R 技術に従って感染させる、請求項 1 ~ 2 のいずれか一項に記載の方法。

【請求項 4】

前記第一の培養物の前記 c a s - 9 陽性細胞が、一つ以上の選択マーカ含有するように改変される、請求項 1 ~ 3 のいずれか一項に記載の方法。

【請求項 5】

前記指定された表現型が蛍光である、請求項 1 ~ 4 のいずれか一項に記載の方法。

50

【請求項 6】

前記指定された表現型が細胞生存である、請求項 1 ~ 5 のいずれか一項に記載の方法。

【請求項 7】

前記一つ以上の選択マーカーが、蛍光マーカーを含む、請求項 4 に記載の方法。

【請求項 8】

前記蛍光マーカーが、FRETバイオセンサーの一部である、請求項 7 に記載の方法。

【請求項 9】

前記一つ以上の選択マーカーが、検出可能な酵素である、請求項 4 に記載の方法。

【請求項 10】

前記検出可能な酵素が、 β -ガラクトシダーゼである、請求項 9 に記載の方法。

10

【請求項 11】

前記検出可能な酵素が、ルシフェラーゼである、請求項 9 に記載の方法。

【請求項 12】

標的領域が、遺伝子を含む、請求項 1 ~ 11 のいずれか一項に記載の方法。

【請求項 13】

前記第二の培養物の細胞を、前記指定された表現型を有する、または前記指定された表現型を有しないのいずれかとしてカテゴライズすることが、選択機構を前記第二の細胞培養物に適用することを含む、請求項 1 ~ 12 のいずれか一項に記載の方法。

【請求項 14】

前記選択機構が、前記第二の細胞集団を薬剤に曝露することか、または前記第二の細胞集団をタンパク質活性または発現レベルを特定する物質に曝露することのうちの一つ以上を含む、請求項 13 に記載の方法。

20

【請求項 15】

前記指定された表現型を有する細胞を選択することが、前記一つ以上の選択マーカーに基づいて前記細胞をソートすることを含む、請求項 4 に記載の方法。

【請求項 16】

前記選択された細胞内の前記遺伝子の n' 以上の gRNA を偶然に観察する前記確率に基づいて、前記標的領域が陽的に選択されると識別することをさらに含む、請求項 1 ~ 15 のいずれか一項に記載の方法。

【請求項 17】

前記標的領域が陽的に選択されると識別することが、前記選択された細胞における対象の配列の n' 以上の gRNA を偶然に観察する前記確率が閾値を満たすと決定することを含む、請求項 16 に記載の方法。

30

【請求項 18】

各 gRNA について、濃縮スコアを決定することをさらに含む、請求項 1 ~ 17 のいずれか一項に記載の方法。

【請求項 19】

各 gRNA について、前記濃縮スコアを決定することが、 N/N' を評価することを含む、請求項 18 に記載の方法。

【請求項 20】

前記 cas-9 陽性細胞が、不活性化 cas-9 を含む、請求項 2 ~ 19 のいずれか一項に記載の方法。

40

【請求項 21】

前記不活性化 cas-9 が、少なくとも一つの転写活性化ドメインに融合される、請求項 20 に記載の方法。

【請求項 22】

前記 DNA の標的領域が、下流遺伝子またはタンパク質を調節する、請求項 1 ~ 21 のいずれか一項に記載の方法。

【請求項 23】

下流遺伝子またはタンパク質を調節することが、前記下流遺伝子またはタンパク質の活性

50

化または阻害を含む、請求項 22 に記載の方法。

【請求項 24】

前記選択された細胞内の前記遺伝子の n' 以上の gRNA を偶然に観察する前記確率に基づいて、前記標的領域を第二の遺伝子の修飾因子として識別することをさらに含む、請求項 1 ~ 23 のいずれか一項に記載の方法。

【請求項 25】

前記選択された細胞内の前記遺伝子の n' 以上の gRNA を偶然に観察する前記確率に基づいて、前記標的領域を治療標的として識別することをさらに含む、請求項 1 ~ 24 のいずれか一項に記載の方法。

【請求項 26】

前記選択された細胞内の前記遺伝子の n' 以上の gRNA を偶然に観察する前記確率に基づいて、前記標的領域が前記指定された表現型と相関すると識別することをさらに含む、請求項 1 ~ 25 のいずれか一項に記載の方法。

10

【請求項 27】

前記選択された細胞内の前記遺伝子の n' 以上の gRNA を偶然に観察する前記確率に基づいて、前記標的領域が保護効果を示すと識別することをさらに含む、請求項 1 ~ 26 のいずれか一項に記載の方法。

【請求項 28】

DNA の複数の標的領域の各々に対する少なくとも 3 つのガイド RNA (gRNA) のライブラリーを含むベクターを用いた感染後の第一の細胞集団を配列決定することに基づいて、そのリード数が背景閾値を超える、前記 DNA の複数の標的領域の各々について存在する gRNA のそれぞれの数 (n) を決定することと、

20

前記 DNA の複数の標的領域の各々に対する前記 gRNA のそれぞれの数に基づいて、そのリード数が前記背景閾値を超える、前記 DNA の複数の標的領域の全ての標的領域にわたる、前記第一の細胞集団に存在する gRNA の総数 (N) を決定することと、

前記 DNA の複数の標的領域の各々に対する前記少なくとも 3 つの gRNA のライブラリーを含むベクターを用いた感染後の第二の細胞集団を配列決定することに基づいて、そのリード数が前記背景閾値を超える、前記 DNA の複数の標的領域の各々について存在する gRNA のそれぞれの数 (n') を決定することと、

前記 DNA の複数の標的領域の各々に対する前記 gRNA のそれぞれの数に基づいて、そのリード数が前記背景閾値を超える前記 DNA の複数の標的領域の全ての標的領域にわたる前記第二の細胞集団に存在する gRNA の総数 (N') を決定することと、

30

前記複数の標的領域の各標的領域について、 n 、 N 、 n' 、および N' に基づいて、前記選択された細胞内の前記標的領域について n' gRNA を偶然観察する確率を決定することと、

対象の配列を含む標的領域について、前記選択された細胞内の前記標的領域について n' gRNA を偶然に観察する前記確率に基づいて、前記選択された細胞における前記対象の配列の n' 以上の gRNA を偶然に観察する確率を決定することと、

前記選択された細胞における前記対象の配列の n' 以上の gRNA を偶然に観察する前記確率に基づいて、前記対象の配列が陽的に選択されると識別することと、を含む、方法

40

【請求項 29】

前記第一の細胞集団および前記第二の集団が、cas-9 陽性細胞を含む、請求項 28 に記載の方法。

【請求項 30】

DNA の複数の標的領域の各々に対する少なくとも 3 つの gRNA のライブラリーを含むベクターを用いた感染後の第一の細胞集団を配列決定することに基づいて、そのリード数が背景閾値を超える、前記 DNA の複数の標的領域の各々について存在する gRNA のそれぞれの数 (n) を決定することが、

前記第一の細胞集団に、ウイルスベクターの前記ライブラリーを感染させることと、

50

前記第一の細胞集団の細胞を配列決定して、前記 g R N A の各々のリード数を取得することと、

そのリード数が前記背景閾値を超える各 g R N A を、前記 D N A の複数の標的領域の各々について、数えることと、を含む、請求項 29 に記載の方法。

【請求項 31】

前記 D N A の複数の標的領域の各々に対する前記 g R N A のそれぞれの数に基づいて、そのリード数が前記背景閾値を超える、前記 D N A の複数の標的領域の全ての標的領域にわたる、前記第一の細胞集団に存在する g R N A の総数 (N) を決定することが、そのリード数が前記背景閾値を超える前記第一の細胞集団に存在する各 g R N A を数えることを含む、請求項 30 に記載の方法。

10

【請求項 32】

前記 D N A の複数の標的領域の各々に対する前記少なくとも 3 つの g R N A のライブラリーを含むベクターを用いた感染後の第二の細胞集団の配列決定に基づいて、そのリード数が前記背景閾値を超える、前記 D N A の複数の標的領域の各々について存在する g R N A (n ') のそれぞれの数を決定することが、

前記第二の細胞集団の細胞を、指定された表現型を有する、または前記指定された表現型を有しないのいずれかとしてカテゴライズすることと、

前記指定された表現型を有する細胞を選択することと、

前記選択された細胞を配列決定して、前記 g R N A の各々のリード数を取得することと、そのリード数が前記背景閾値を超える各 g R N A を、前記 D N A の複数の標的領域の各々について、数えることと、を含む、請求項 31 に記載の方法。

20

【請求項 33】

前記 D N A の複数の標的領域の各々に対する前記 g R N A のそれぞれの数に基づいて、そのリード数が前記背景閾値を超える前記 D N A の複数の標的領域の全ての標的領域にわたる前記第二の細胞集団に存在する g R N A の総数 (N ') を決定することが、そのリード数が前記背景閾値を超える前記第二の細胞集団の選択された細胞に存在する各 g R N A を数えることを含む、請求項 28 ~ 32 のいずれか一項に記載の方法。

【請求項 34】

前記複数の標的領域の各標的領域について、 n 、 N 、 n ' 、および N ' に基づいて、前記選択された細胞内の前記標的領域について n ' g R N A を偶然に観察する前記確率を決定することが、

30

【数 7】

$$\frac{\binom{n'}{N'} \cdot \binom{N'-n'}{N-n'}}{\binom{N'}{N}}$$

を評価することを含み、式中

【数 8】

$$\binom{x}{y} = \frac{y!}{x!(y-x)!}$$

40

は、 y オブジェクトから x オブジェクトを選択する方法の数を算出する、請求項 28 ~ 33 のいずれか一項に記載の方法。

【請求項 35】

対象の配列を含む標的領域について、前記選択された細胞内の前記標的領域について n ' g R N A を偶然に観察する前記確率に基づいて、前記選択された細胞における前記対象の配列の n ' 以上の g R N A を偶然に観察する確率を決定することは、

【数 9】

50

$$\sum_{i=n'}^n \frac{\binom{i}{N'} \cdot \binom{N'-i}{N-i}}{\binom{N'}{N}}$$

を評価することを含む、請求項 28 ~ 34 のいずれか一項に記載の方法。

【請求項 36】

前記選択された細胞における前記対象の配列の n' 以上の gRNA を偶然に観察する前記確率に基づいて、前記対象の配列が陽的に選択されると識別することが、前記選択された細胞における前記対象の配列の n' 以上の gRNA を偶然に観察する前記確率が閾値を

10

【請求項 37】

各 gRNA について、濃縮スコアを決定することをさらに含む、請求項 28 ~ 36 のいずれか一項に記載の方法。

【請求項 38】

各 gRNA について、濃縮スコアを決定することが、 N/N' を評価することを含む、請求項 37 に記載の方法。

【請求項 39】

前記 cas-9 陽性細胞が、不活性化 cas-9 を含む、請求項 28 ~ 38 のいずれか一項に記載の方法。

20

【請求項 40】

前記不活性化 cas-9 が、少なくとも一つの転写活性化ドメインに融合される、請求項 39 に記載の方法。

【請求項 41】

装置であって、
一つ以上のプロセッサと、
プロセッサ実行可能命令であって、前記一つ以上のプロセッサによって実行されると、前記装置に、

(A) 第一のリード数データであって、
cas9 陽性細胞の第一の培養物をウイルスベクターのライブラリーで感染させることであって、前記ライブラリーが、前記細胞のゲノム内の DNA の標的領域を切断するための少なくとも 3 つのガイド RNA (gRNA) を含む、感染させることと、
前記細胞を配列決定して、前記 gRNA の各々のリード数を取得することと、によって生成される、第一のリード数データを受信させる、
前記第一のリード数データに基づく、DNA の標的領域当たりの、そのリード数が背景閾値を超える、gRNA のそれぞれの数の合計 () であって、式中 $\sum = n$ である、合計をさせる、および、
前記第一のリード数データに基づく、すべての標的領域にわたる、そのリード数が前記背景閾値を超える、gRNA の総数の合計 () であって、式中 $\sum = N$ である、合計をさせる；

30

40

(B) 第二のリード数データであって、
前記ウイルスベクターのライブラリーで、cas9 陽性細胞の第二の培養物を感染させることと、

前記第二の培養物の前記細胞を、指定された表現型を有する、または前記指定された表現型を有しないのいずれかとしてカテゴライズすることと、

前記指定された表現型を有する細胞を選択し、前記選択された細胞を配列決定して前記 gRNA の各々の選択後リード数を取得することと、によって生成される、第二のリード数データを受信させる、

前記第二のリード数データに基づく、その選択後リード数が前記背景閾値を超える前記選択された細胞における、DNA の標的領域当たりの、gRNA のそれぞれの数の合計 (

50

)であって、式中 $n' = n'$ である、合計をさせる、および、前記第二のリード数データに基づく、すべての標的領域にわたる、そのリード数が前記閾値を超える前記選択された細胞の、gRNAの総数の合計()であって、式中 $n' = N'$ である、合計をさせる；

(C) DNAの標的領域について、前記式

【数10】

$$\frac{\binom{n'}{N'} \cdot \binom{N'-n'}{N-n'}}{\binom{N'}{N}}$$

10

に従った、前記選択された細胞内の前記標的領域について n' gRNAを偶然に観察する確率の計算であって、式中

【数11】

$$\binom{x}{y} = \frac{y!}{x!(y-x)!}$$

は、 y オブジェクトから x オブジェクトを選択する方法の数を算出する、計算をさせる、および、

20

遺伝子を含むDNAの標的領域について、式

【数12】

$$\sum_{i=n'}^n \frac{\binom{i}{N'} \cdot \binom{N'-i}{N-i}}{\binom{N'}{N}}$$

に従った、前記選択された細胞内の遺伝子の n' 以上のgRNAを偶然に観察する確率の計算をさせる、プロセッサ実行可能命令、を格納するメモリと、を備える装置。

30

【請求項42】

前記第一の培養物を、CRISPR技術に従って感染させる、請求項41に記載の装置。

【請求項43】

前記第一の培養物の前記cas-9陽性細胞が、一つ以上の選択マーカ含有するように改変される、請求項41～42のいずれか一項に記載の装置。

【請求項44】

標的領域が遺伝子を含む、請求項41～43のいずれか一項に記載の装置。

【請求項45】

前記第二の培養物の細胞を、指定された表現型を有する、または前記指定された表現型を有しないのいずれかとしてカテゴライズすることが、選択機構を前記第二の細胞培養物に適用することを含む、請求項41～44のいずれか一項に記載の装置。

40

【請求項46】

前記選択機構が、前記第二の細胞集団を薬剤に曝露することか、または前記第二の細胞集団をタンパク質活性または発現レベルを特定する物質に曝露することのうちの一つ以上を含む、請求項45に記載の装置。

【請求項47】

前記指定された表現型を有する細胞を選択することが、前記一つ以上の選択マーカに基づいて前記細胞をソートすることを含む、請求項41～46のいずれか一項に記載の装置。

【請求項48】

50

前記選択された細胞内の前記遺伝子の n' 以上の $gRNA$ を偶然に観察する前記確率に基づいて、前記標的領域が陽的に選択されると識別することをさらに含む、請求項 41 ~ 47 のいずれか一項に記載の装置。

【請求項 49】

前記標的領域が陽的に選択されると識別することが、前記選択された細胞における前記対象の配列の n' 以上の $gRNA$ を偶然に観察する前記確率が閾値を満たすと決定することを含む、請求項 48 に記載の装置。

【請求項 50】

各 $gRNA$ について、濃縮スコアを決定することをさらに含む、請求項 41 ~ 49 のいずれか一項に記載の装置。

10

【請求項 51】

各 $gRNA$ について、濃縮スコアを決定することが、 N/N' を評価することを含む、請求項 50 に記載の装置。

【請求項 52】

前記 $cas-9$ 陽性細胞が、不活性化 $cas-9$ を含む、請求項 41 ~ 51 のいずれか一項に記載の装置。

【請求項 53】

前記不活性化 $cas-9$ が、少なくとも一つの転写活性化ドメインに融合される、請求項 52 に記載の装置。

【請求項 54】

一つ以上のガイド RNA ($gRNA$) を偶然に観察する確率を決定するための非一時的コンピュータ可読媒体であって、プロセッサ実行可能命令であって、一つ以上のプロセッサによって実行されると、前記一つ以上のプロセッサに、

20

(A) 第一のリード数データであって、

$cas9$ 陽性細胞の第一の培養物をウイルスベクターのライブラリーで感染させることであって、前記ライブラリーが、前記細胞のゲノム内の DNA の標的領域を切断するための少なくとも3つのガイド RNA ($gRNA$) を含む、感染させることと、

前記細胞を配列決定して、前記 $gRNA$ の各々のリード数を取得することと、によって生成される、第一のリード数データを受信させる、

前記第一のリード数データに基づく、 DNA の標的領域当たりの、そのリード数が背景閾値を超える、 $gRNA$ のそれぞれの数の合計 () であって、式中 $= n$ である、合計をさせる、および、

30

前記第一のリード数データに基づく、すべての標的領域にわたる、そのリード数が前記背景閾値を超える、 $gRNA$ の総数の合計 () であって、式中 $= N$ である、合計をさせる；

(B) 第二のリード数データであって、

$cas9$ 陽性細胞の第二の培養物を、前記ウイルスベクターのライブラリーに感染させることと、

前記第二の培養物の前記細胞を、指定された表現型を有する、または前記指定された表現型を有しないのいずれかとしてカテゴライズすることと、

40

前記指定された表現型を有する細胞を選択し、前記選択された細胞を配列決定して前記 $gRNA$ の各々の選択後リード数を取得することと、によって生成される、第二のリード数データを受信させる、

前記第二のリード数データに基づく、その選択後リード数が前記背景閾値を超える前記選択された細胞における、 DNA の標的領域当たりの、 $gRNA$ のそれぞれの数の合計 () であって、式中 $= n'$ である、合計をさせる、および、

前記第二のリード数データに基づく、すべての標的領域にわたる、そのリード数が前記閾値を超える前記選択された細胞の、 $gRNA$ の総数の合計 () であって、式中 $= N'$ である、合計をさせる；

(C) DNA の標的領域について、

50

前記式

【数 1 3】

$$\frac{\binom{n'}{N'} \cdot \binom{N'-n'}{N-n'}}{\binom{N'}{N}}$$

に従った、前記選択された細胞内の前記標的領域について n' gRNA を偶然に観察する確率の計算であって、式中

【数 1 4】

$$\binom{x}{y} = \frac{y!}{x!(y-x)!}$$

10

は、 y オブジェクトから x オブジェクトを選択する方法の数を算出する、計算をさせる、および、

遺伝子を含む DNA の標的領域について、前記式

【数 1 5】

$$\sum_{i=n'}^n \frac{\binom{i}{N'} \cdot \binom{N'-i}{N-i}}{\binom{N'}{N}}$$

20

に従った、前記選択された細胞内の遺伝子の n' 以上の gRNA を偶然に観察する確率の計算をさせる、プロセッサ実行可能命令、を格納する非一時的コンピュータ可読媒体。

【請求項 5 5】

前記第一の培養物が、CRISPR 技術に従って感染される、請求項 5 4 に記載の非一時的コンピュータ可読媒体。

【請求項 5 6】

前記第一の培養物の前記 cas-9 陽性細胞が、一つ以上の選択マーカ含有するように改変される、請求項 5 4 ~ 5 5 のいずれか一項に記載の非一時的コンピュータ可読媒体。

30

【請求項 5 7】

標的領域が遺伝子を含む、請求項 5 4 ~ 5 6 のいずれか一項に記載の非一時的コンピュータ可読媒体。

【請求項 5 8】

前記第二の培養物の細胞を、前記指定された表現型を有する、または前記指定された表現型を有しないのいずれかとしてカテゴライズすることが、選択機構を前記第二の細胞培養物に適用することを含む、請求項 5 4 ~ 5 7 のいずれか一項に記載の非一時的コンピュータ可読媒体。

【請求項 5 9】

前記選択機構が、前記第二の細胞集団を薬剤に曝露することか、または前記第二の細胞集団をタンパク質活性または発現レベルを特定する物質に曝露することのうちの一つ以上を含む、請求項 5 8 に記載の非一時的コンピュータ可読媒体。

40

【請求項 6 0】

前記指定された表現型を有する細胞を選択することが、前記一つ以上の選択マーカに基づいて前記細胞をソートすることを含む、請求項 5 4 ~ 5 9 のいずれか一項に記載の非一時的コンピュータ可読媒体。

【請求項 6 1】

前記選択された細胞内の前記遺伝子の n' 以上の gRNA を偶然に観察する前記確率に基づいて、前記標的領域が陽的に選択されると識別することをさらに含む、請求項 5 4 ~

50

60のいずれか一項に記載の非一時的コンピュータ可読媒体。

【請求項62】

前記標的領域が陽的に選択されると識別することが、前記選択された細胞における前記対象の配列の n' 以上のgRNAを偶然に観察する前記確率が閾値を満たすと決定することを含む、請求項61に記載の非一時的コンピュータ可読媒体。

【請求項63】

各gRNAについて、濃縮スコアを決定することをさらに含む、請求項54～62のいずれか一項に記載の非一時的コンピュータ可読媒体。

【請求項64】

各gRNAについて、前記濃縮スコアを決定することが、 N/N' を評価することを含む、請求項63に記載の非一時的コンピュータ可読媒体。 10

【請求項65】

前記cas-9陽性細胞が、不活性化cas-9を含む、請求項54～64のいずれか一項に記載の非一時的コンピュータ可読媒体。

【請求項66】

前記不活性化cas-9が、少なくとも一つの転写活性化ドメインに融合される、請求項65に記載の非一時的コンピュータ可読媒体。

【請求項67】

装置であって、
一つ以上のプロセッサと、 20

プロセッサ実行可能命令であって、前記一つ以上のプロセッサによって実行されると、前記装置に、

DNAの複数の標的領域の各々に対する少なくとも3つのガイドRNA(gRNA)のライブラリーを含むベクターを用いた感染後の第一の細胞集団を配列決定することに基づいて、そのリード数が背景閾値を超える、前記DNAの複数の標的領域の各々について存在するgRNAのそれぞれの数(n)を決定させる、

前記DNAの複数の標的領域の各々に対する前記gRNAのそれぞれの数に基づいて、そのリード数が前記背景閾値を超える、前記DNAの複数の標的領域の全ての標的領域にわたる、前記第一の細胞集団に存在するgRNAの総数(N)を決定させる、

前記DNAの複数の標的領域の各々に対する前記少なくとも3つのgRNAのライブラリーを含むベクターを用いた感染後の第二の細胞集団を配列決定することに基づいて、そのリード数が前記背景閾値を超える、前記DNAの複数の標的領域の各々について存在するgRNAのそれぞれの数(n')を決定させる、 30

前記DNAの複数の標的領域の各々に対する前記gRNAのそれぞれの数に基づいて、そのリード数が前記背景閾値を超える前記DNAの複数の標的領域の全ての標的領域にわたる前記第二の細胞集団に存在するgRNAの総数(N')を決定させる、

前記複数の標的領域の各標的領域について、 n 、 N 、 n' 、および N' に基づいて、前記選択された細胞内の前記標的領域について n' gRNAを偶然に観察する確率を決定させる、

対象の配列を含む標的領域について、前記選択された細胞内の前記標的領域について n' gRNAを偶然に観察する前記確率に基づいて、前記選択された細胞における前記対象の配列の n' 以上のgRNAを偶然に観察する確率を決定させる、および、 40

前記選択された細胞における前記対象の配列の n' 以上のgRNAを偶然に観察する前記確率に基づいて、前記対象の配列が陽的に選択されることを識別させる、プロセッサ実行可能命令、を格納するメモリと、を備える装置。

【請求項68】

前記第一の細胞集団および前記第二の集団が、cas-9陽性細胞を含む、請求項67に記載の装置。

【請求項69】

前記プロセッサ実行可能命令であって、前記一つ以上のプロセッサによって実行されると、前記装置に、DNAの複数の標的領域の各々に対して少なくとも3つのgRNAのライ 50

ブラリーを含むベクターを用いた感染後の第一の細胞集団を配列決定することに基づいて、そのリード数が背景閾値を超える、前記DNAの複数の標的領域の各々について存在するgRNAのそれぞれの数(n)を決定させる、プロセッサ実行可能命令が、前記装置に

、
前記第一の細胞集団に前記ウイルスベクターのライブラリーを感染させることと、
前記第一の細胞集団の細胞を配列決定して、前記gRNAの各々のリード数を取得することと、
によって生成される、第一のリード数データを受信させる、および、
前記第一のリード数データに基づいて、前記DNAの複数の標的領域の各々について、そのリード数が前記背景閾値を超える各gRNAを数えさせる、請求項67~68のいずれか一項に記載の装置。

10

【請求項70】

前記プロセッサ実行可能命令であって、前記一つ以上のプロセッサによって実行されると、前記装置に、前記DNAの複数の標的領域の各々に対する前記gRNAのそれぞれの数に基づいて、そのリード数が前記背景閾値を超える、前記DNAの複数の標的領域の全ての標的領域にわたる、前記第一の細胞集団に存在するgRNAの総数(N)を決定させる、プロセッサ実行可能命令が、前記装置に、そのリード数が前記背景閾値を超える前記第一の細胞集団に存在する各gRNAを数えさせる、請求項69に記載の装置。

【請求項71】

前記プロセッサ実行可能命令であって、前記一つ以上のプロセッサによって実行されると、前記装置に、前記DNAの複数の標的領域の各々に対する前記少なくとも3つのgRNAのライブラリーを含むベクターを用いた感染後の第二の細胞集団を配列決定することに基づいて、そのリード数が前記背景閾値を超える、前記DNAの複数の標的領域の各々について存在するgRNAのそれぞれの数(n')を決定させる、プロセッサ実行可能命令が、前記装置に、

20

前記第二の細胞集団の細胞を、指定された表現型を有する、または前記指定された表現型を有しないのいずれかとしてカテゴリズすることと、

前記指定された表現型を有する細胞を選択することと、

前記選択された細胞を配列決定して、前記gRNAの各々のリード数を取得することと、
によって生成される、第二のリード数データを受信させる、および、

前記第二のリード数データに基づいて、前記DNAの複数の標的領域の各々について、そのリード数が前記背景閾値を超える各gRNAを数えさせる、請求項67~70のいずれか一項に記載の装置。

30

【請求項72】

前記プロセッサ実行可能命令であって、前記一つ以上のプロセッサによって実行されると、前記装置に、前記DNAの複数の標的領域の各々に対する前記gRNAのそれぞれの数に基づいて、そのリード数が前記背景閾値を超える前記DNAの複数の標的領域の全ての標的領域にわたる前記第二の細胞集団に存在するgRNAの総数(N')を決定させる、プロセッサ実行可能命令が、前記装置に、そのリード数が前記背景閾値を超える前記第二の細胞集団の前記選択された細胞に存在する各gRNAを数えさせる、請求項67~71のいずれか一項に記載の装置。

40

【請求項73】

前記プロセッサ実行可能命令であって、前記一つ以上のプロセッサによって実行されると、前記装置に、前記複数の標的領域の各標的領域について、n、N、n'、およびN'に基づいて、前記選択された細胞内の標的領域についてn'gRNAを偶然に観察する前記確率を決定させる、プロセッサ実行可能命令が、前記装置に

【数16】

$$\frac{\binom{n'}{N'} \cdot \binom{N'-n'}{N-n'}}{\binom{N'}{N}}$$

50

を評価させ、式中

【数 1 7】

$$\binom{x}{y} = \frac{y!}{x!(y-x)!}$$

は、y オブジェクトから x オブジェクトを選択する方法の数を算出する、請求項 6 7 ~ 7 2 のいずれか一項に記載の装置。

【請求項 7 4】

前記プロセッサ実行可能命令であって、前記一つ以上のプロセッサによって実行されると、前記装置に、前記選択された細胞内の標的領域について n' gRNA を偶然に観察する前記確率に基づいて、対象の配列を含む標的領域について、前記選択された細胞における前記対象の配列の n' 以上の gRNA を偶然に観察する確率を決定させる、プロセッサ実行可能命令が、前記装置に

10

【数 1 8】

$$\sum_{i=n'}^n \frac{\binom{i}{N'} \cdot \binom{N'-i}{N-i}}{\binom{N'}{N}}$$

20

を評価させる、請求項 6 7 ~ 7 3 のいずれか一項に記載の装置。

【請求項 7 5】

前記プロセッサ実行可能命令であって、前記一つ以上のプロセッサによって実行されると、前記装置に、前記選択された細胞における前記対象の配列の n' 以上の gRNA を偶然に観察する前記確率に基づいて、前記対象の配列が陽的に選択されることを識別させる、プロセッサ実行可能命令が、前記装置に、前記選択された細胞における前記対象の配列の n' 以上の gRNA を偶然に観察する前記確率が閾値を満たすと評価させる決定させる、請求項 6 7 ~ 7 4 のいずれか一項に記載の装置。

【請求項 7 6】

前記一つ以上のプロセッサによって実行されると、前記装置に、各 gRNA について濃縮スコアを決定させる、プロセッサ実行可能命令をさらに含む、請求項 6 7 ~ 7 6 のいずれか一項に記載の装置。

30

【請求項 7 7】

前記プロセッサ実行可能命令であって、前記一つ以上のプロセッサによって実行されると、前記装置に、各 gRNA について濃縮スコアを決定させる、プロセッサ実行可能命令が、前記装置に N / N' を評価させる、請求項 7 6 に記載の装置。

【請求項 7 8】

前記 cas - 9 陽性細胞が、不活性化 cas - 9 を含む、請求項 6 7 ~ 7 7 のいずれか一項に記載の装置。

【請求項 7 9】

前記不活性化 cas - 9 が、少なくとも一つの転写活性化ドメインに融合される、請求項 7 8 に記載の装置。

40

【請求項 8 0】

一つ以上のガイド RNA (gRNA) を偶然に観察する確率を決定するための非一時的コンピュータ可読媒体であって、プロセッサ実行可能命令であって、一つ以上のプロセッサによって実行されると、前記一つ以上のプロセッサに、DNA の複数の標的領域の各々に対する少なくとも 3 つのガイド RNA (gRNA) のライブラリーを含むベクターを用いた感染後の第一の細胞集団を配列決定することに基づいて、そのリード数が背景閾値を超える、DNA の複数の標的領域の各々について存在する gRNA のそれぞれの数 (n) を決定させる、

50

前記 DNA の複数の標的領域の各々に対する前記 gRNA のそれぞれの数に基づいて、そのリード数が前記背景閾値を超える、前記 DNA の複数の標的領域の全ての標的領域にわたる、前記第一の細胞集団に存在する gRNA の総数 (N) を決定させる、
 前記 DNA の複数の標的領域の各々に対する前記少なくとも 3 つの gRNA のライブラリーを含むベクターを用いた感染後の第二の細胞集団を配列決定することに基づいて、そのリード数が前記背景閾値を超える、前記 DNA の複数の標的領域の各々について存在する gRNA のそれぞれの数 (n') を決定させる、
 前記 DNA の複数の標的領域の各々に対する前記 gRNA のそれぞれの数に基づいて、そのリード数が前記背景閾値を超える前記 DNA の複数の標的領域の全ての標的領域にわたる前記第二の細胞集団に存在する gRNA の総数 (N') を決定させる、
 前記複数の標的領域の各標的領域について、n、N、n'、および N' に基づいて、前記選択された細胞内の前記標的領域について n' gRNA を偶然に観察する確率を決定させる、対象の配列を含む標的領域について、前記選択された細胞内の前記標的領域について n' gRNA を偶然に観察する前記確率に基づいて、前記選択された細胞における前記対象の配列の n' 以上の gRNA を偶然に観察する確率を決定させる、および、
 前記選択された細胞における前記対象の配列の n' 以上の gRNA を偶然に観察する前記確率に基づいて、前記対象の配列が陽性的に選択されることを識別させる、プロセッサ実行可能命令、を格納する、非一時的コンピュータ可読媒体。

10

【請求項 81】

前記第一の細胞集団および前記第二の集団が、cas-9 陽性細胞を含む、請求項 80 に記載の非一時的コンピュータ可読媒体。

20

【請求項 82】

前記プロセッサ実行可能命令であって、前記一つ以上のプロセッサによって実行されると、前記装置に、DNA の複数の標的領域の各々に対して少なくとも 3 つの gRNA のライブラリーを含むベクターを用いた感染後の第一の細胞集団を配列決定することに基づいて、そのリード数が背景閾値を超える、前記 DNA の複数の標的領域の各々について存在する gRNA のそれぞれの数 (n) を決定させる、プロセッサ実行可能命令が、前記一つ以上のプロセッサに、
 前記第一の細胞集団に前記ウイルスベクターのライブラリーを感染させることと、
 前記第一の細胞集団の細胞を配列決定して、前記 gRNA の各々のリード数を取得することと、によって生成される、第一のリード数データを受信させる、および、
 前記第一のリード数データに基づいて、前記 DNA の複数の標的領域の各々について、そのリード数が前記背景閾値を超える各 gRNA を、数えさせる、請求項 80 ~ 81 のいずれか一項に記載の非一時的コンピュータ可読媒体。

30

【請求項 83】

前記プロセッサ実行可能命令であって、前記一つ以上のプロセッサによって実行されると、前記装置に、前記 DNA の複数の標的領域の各々に対する前記 gRNA のそれぞれの数に基づいて、そのリード数が前記背景閾値を超える、前記 DNA の複数の標的領域の全ての標的領域にわたる、前記第一の細胞集団に存在する gRNA の総数 (N) を決定させる、プロセッサ実行可能命令が、前記一つ以上のプロセッサに、そのリード数が前記背景閾値を超える前記第一の細胞集団に存在する各 gRNA を数えさせる、請求項 80 ~ 82 のいずれか一項に記載の非一時的コンピュータ可読媒体。

40

【請求項 84】

前記プロセッサ実行可能命令であって、前記一つ以上のプロセッサによって実行されると、前記装置に、前記 DNA の複数の標的領域の各々に対する前記少なくとも 3 つの gRNA のライブラリーを含むベクターを用いた感染後の第二の細胞集団を配列決定することに基づいて、そのリード数が前記背景閾値を超える、前記 DNA の複数の標的領域の各々について存在する gRNA のそれぞれの数 (n') を決定させる、プロセッサ実行可能命令が、前記一つ以上のプロセッサに、
 前記第二の細胞集団の細胞を、指定された表現型を有する、または前記指定された表現型

50

を有しないのいずれかとしてカテゴリズすることと、
 前記指定された表現型を有する細胞を選択することと、
 前記選択された細胞を配列決定して、前記 gRNA の各々のリード数を取得することと、
 によって生成される、第二のリード数データを受信させる、および、
 前記第二のリード数データに基づいて、前記 DNA の複数の標的領域の各々について、その
 リード数が前記背景閾値を超える各 gRNA を数えさせる、請求項 80 ~ 83 のいずれ
 か一項に記載の非一時的コンピュータ可読媒体。

【請求項 85】

前記プロセッサ実行可能命令であって、前記一つ以上のプロセッサによって実行されると
 、前記装置に、前記 DNA の複数の標的領域の各々に対する前記 gRNA のそれぞれの数 10
 に基づいて、そのリード数が前記背景閾値を超える前記 DNA の複数の標的領域の全ての
 標的領域にわたる前記第二の細胞集団に存在する gRNA の総数 (N') を決定させる、
 プロセッサ実行可能命令が、前記一つ以上のプロセッサに、そのリード数が前記背景閾値
 を超える前記第二の細胞集団の前記選択された細胞に存在する各 gRNA を数えさせる、
 請求項 80 ~ 84 のいずれか一項に記載の非一時的コンピュータ可読媒体。

【請求項 86】

前記プロセッサ実行可能命令であって、前記一つ以上のプロセッサによって実行されると
 、前記装置に、前記複数の標的領域の各標的領域について、n、N、n'、および N' に基
 づいて、前記選択された細胞内の前記標的領域について n' gRNA を偶然に観察する前
 記確率を決定させる、プロセッサ実行可能命令が、前記一つ以上のプロセッサに 20

【数 19】

$$\frac{\binom{n'}{N'} \cdot \binom{N'-n'}{N-n'}}{\binom{N'}{N}}$$

を評価させ、式中

【数 20】

$$\binom{x}{y} = \frac{y!}{x!(y-x)!}$$

は、y オブジェクトから x オブジェクトを選択する方法の数を算出する、請求項 80 ~ 8
 5 のいずれか一項に記載の非一時的コンピュータ可読媒体。

【請求項 87】

前記プロセッサ実行可能命令であって、前記一つ以上のプロセッサによって実行されると
 、前記装置に、前記選択された細胞内の前記標的領域について n' gRNA を偶然に観察
 する前記確率に基づいて、対象の配列を含む標的領域について、前記選択された細胞にお
 ける前記対象の配列の n' 以上の gRNA を偶然に観察する確率を決定させる、プロセッ
 サ実行可能命令が、前記一つ以上のプロセッサに 40

【数 21】

$$\sum_{i=n'}^n \frac{\binom{i}{N'} \cdot \binom{N'-i}{N-i}}{\binom{N'}{N}}$$

を評価させる、請求項 80 ~ 86 のいずれか一項に記載の非一時的コンピュータ可読媒体
 。

【請求項 88】

前記プロセッサ実行可能命令であって、前記一つ以上のプロセッサによって実行されると 50

、前記装置に、前記選択された細胞における前記対象の配列の n' 以上の $gRNA$ を偶然に観察する前記確率に基づいて、前記対象の配列が陽的に選択されることを識別させる、プロセッサ実行可能命令が、前記一つ以上のプロセッサに、前記選択された細胞における前記対象の配列の n' 以上の $gRNA$ を偶然に観察する前記確率が閾値を満たすと評価させる決定させる、請求項 80 ~ 87 のいずれか一項に記載の非一時的コンピュータ可読媒体。

【請求項 89】

前記一つ以上のプロセッサによって実行されると、前記一つ以上のプロセッサに、各 $gRNA$ について、濃縮スコアを決定させるプロセッサ実行可能命令を、さらに含む、請求項 80 ~ 88 のいずれか一項に記載の非一時的コンピュータ可読媒体。

10

【請求項 90】

前記プロセッサ実行可能命令であって、前記一つ以上のプロセッサによって実行されると、前記装置に、各 $gRNA$ について、濃縮スコアを決定させる、プロセッサ実行可能命令が、前記一つ以上のプロセッサに、 N/N' を評価させることを含む、請求項 89 に記載の非一時的コンピュータ可読媒体。

【請求項 91】

前記 $cas-9$ 陽性細胞が、不活性化 $cas-9$ を含む、請求項 80 ~ 90 のいずれか一項に記載の非一時的コンピュータ可読媒体。

【請求項 92】

前記不活性化 $cas-9$ が、少なくとも一つの転写活性化ドメインに融合される、請求項 91 に記載の非一時的コンピュータ可読媒体。

20

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、CRISPR 選択のための方法およびシステムに関する。

関連出願の相互参照

本出願は、2019年3月18日に提出された米国仮特許出願第 62/820,106 号の利益を主張するものであり、その全体が参照により本明細書に援用される。

【背景技術】

【0002】

クラスター化して規則的な配置の短い回文配列リピート (CRISPR) - Cas9 技術は、ゲノム工学に大革命を起こした。このシステムでは、ガイド RNA ($gRNA$) が、Cas9ヌクレアーゼに標的化ゲノム領域での二重鎖切断を誘導させる。 $gRNA$ の 5' 末端は、標的化領域に相補的な約 20ヌクレオチドのヌクレオチド配列を含む。二本鎖切断が非相同末端結合 (NHEJ) によって修復される場合、挿入および欠失は高頻度で発生し、したがって標的化ゲノム遺伝子座を効率的にロックアウトする。レンチウイルス送達方法の開発により、ゲノムスケール CRISPR/Cas9 ロックアウトのライブラリーの作成が可能となった。これらのライブラリーは、哺乳類細胞株上で陰性および陽性選択スクリーニングの両方が行われることを可能にする。CRISPR/Cas9 ロックアウトスクリーンでは、各遺伝子はいくつかの $gRNA$ によって標的化され、異なる遺伝子ロックアウトを運ぶ変異体プールは、高スループット配列決定によって決定され得る。CRISPR 活性化 (CRISPRa) はまた、活性化された遺伝子が高スループット配列決定によって決定され得る $gRNA$ ライブラリーとともに使用され得る。

30

40

【0003】

ゲノムワイド CRISPR/Cas9 ロックアウトまたは遺伝子活性化技術は、有効な遺伝子振動スクリーニング技術である。目的は、表現型に関連する $gRNA$ 、従ってそれに対応する罹患遺伝子を識別することである。しかしながら、これらのスクリーンによって生成されるデータは、コンピュータ解析にいくつかの課題を提起する。CRISPR 研究は、多くの場合、複数の複製を用いて実施される。CRISPR は、各実験がスクリーニングライブラリー内で同じ $gRNA$ ウィルス力価を使用しない場合があり、レンチウイル

50

ス感染速度が実験間で異なる場合があり、gRNAが実験間で同じ効率の遺伝子を標的としない場合があるという変動性に影響を受けやすい。したがって、観察されたgRNA存在量は、同じ表現型の細胞を用いても、実験間で非常に変動的である。既存の技術は、表現型に関連するgRNAを識別するためにリード数に依存しており、具体的には、正規化されたgRNAリード数の平均および分散を使用して、表現型を有するまたは有しない細胞間でgRNAの存在量が有意に異なるかどうかを試験する。

【0004】

しかしながら、こうした技術は、上述の実験間の変動性の問題には対処せず、むしろ、選択前実験と選択後実験との間で高度な均質性を想定している。これらの技術は、単一のCRISPR実験内および/またはCRISPR実験間の変動性には対処しない。

10

【発明の概要】

【発明が解決しようとする課題】

【0005】

したがって、陽性および陰性選択スクリーンを介して遺伝子を識別する際に、CRISPRの変動性の問題に対処するコンピューティングテクノロジーにおける技術的改善に対するニーズがある。

【課題を解決するための手段】

【0006】

以下の概説および以下の発明を実施するための形態は両方とも、あくまで例示的かつ説明的なものであって、限定的なものではないことを理解されたい。

20

一実施形態では、方法は、(A) ウイルスペクターのライブラリーで、cas9陽性細胞の第一の培養物を感染させることを含み、ライブラリーは、細胞のゲノム内のDNAの標的領域を切断するための少なくとも3つのガイドRNA (gRNA) を含み、細胞を配列決定してgRNAの各々のリード数を取得し、DNAの標的領域当たりで、そのリード数が背景閾値 (background threshold) を超える、gRNAのそれぞれの数を合計し()、式中 $n = n$ で、かつgRNAの総数を合計し()、式中 $N = N$ である。方法は、(B) ウイルスペクターのライブラリーで、cas9陽性細胞の第二の培養物を感染させることと、第二の培養物の細胞を、指定された表現型を有する、または指定された表現型を有しないのいずれかとしてカテゴライズすることと、指定された表現型を有する細胞を選択し、選択された細胞を配列決定して、gRNAの各々の選択後リード数を取得することと、DNAの標的領域当たりで、選択後リード数が背景閾値を超える選択された細胞において、gRNAのそれぞれの数を合計すること() であって、式中 $n' = n'$ である、合計することと、標的領域全体にわたって、そのリード数が閾値を超える選択された細胞の、gRNAの総数を合計すること() であって、式中 $N' = N'$ である、合計することと、を含む。方法は、(C) DNAの標的領域について、式

30

【0007】

【数1】

$$\frac{\binom{n'}{n} \cdot \binom{N' - n'}{N - n}}{\binom{N'}{N}}$$

40

【0008】

に従って、選択された細胞内の標的領域について n' gRNAを偶然に観察する確率を計算すること、式中

【0009】

【数2】

$$\binom{x}{y} = \frac{y!}{x!(y-x)!}$$

50

【 0 0 1 0 】

は y オブジェクトから x オブジェクトを選択する方法の数を算出し、および遺伝子を含む DNA の標的領域について、式

【 0 0 1 1 】

【 数 3 】

$$\sum_{i=n'}^n \frac{\binom{i}{n} \cdot \binom{N'-i}{N-n}}{\binom{N'}{N}}$$

10

【 0 0 1 2 】

に従って、選択された細胞内の遺伝子の n' 以上の gRNA を偶然に観察する確率を計算すること、を含む。

一実施形態では、方法は、DNA の複数の標的領域の各々に対する少なくとも 3 つのガイド RNA (gRNA) のライブラリーを含むベクターを用いた感染後の、第一の細胞集団の配列決定に基づいて、そのリード数が背景閾値を超える、DNA の複数の標的領域の各々について存在する gRNA のそれぞれの数 (n) を決定すること、DNA の複数の標的領域の各々に対する gRNA のそれぞれの数に基づいて、そのリード数が背景閾値を超える、DNA の複数の標的領域の全ての標的領域にわたる、第一の細胞集団に存在する gRNA の総数 (N) を決定すること、DNA の複数の標的領域の各々に対する少なくとも 3 つの gRNA のライブラリーを含むベクターを用いた感染後の第二の細胞集団の配列決定に基づいて、そのリード数が背景閾値を超える、DNA の複数の標的領域の各々について存在する gRNA のそれぞれの数 (n') を決定すること、DNA の複数の標的領域の各々に対する gRNA のそれぞれの数に基づいて、そのリード数が背景閾値を超える DNA の複数の標的領域の全ての標的領域にわたる第二の細胞集団に存在する gRNA の総数 (N') を決定すること、n、N、n'、および N' に基づき、複数の標的領域の各標的領域に対して、選択された細胞内の標的領域について n' gRNA を偶然に観察する確率を決定すること、対象の配列を含む標的領域について、選択された細胞内の標的領域について n' gRNA を偶然に観察する確率に基づいて、選択された細胞における対象の配列の n' 以上の gRNA を偶然に観察する確率を決定すること、および選択された細胞における対象の配列の n' 以上の gRNA を偶然に観察する確率に基づいて、対象の配列が陽的に選択されると識別すること、を含む。

20

30

【 0 0 1 3 】

追加の利点は、その一部が下記説明に記載されているか、または実践によって知ることができるであろう。これらの利点は、添付の特許請求の範囲において特に指摘されている要素および組み合わせによって実現され、達成されるであろう。

【 図面の簡単な説明 】

【 0 0 1 4 】

【 図 1 】 図 1 は、CRISPR 陽性選択のための例示的な方法である。

【 図 2 】 図 2 は、別の例示的な方法を示す。

40

【 図 3 】 図 3 は、例示的な操作環境を示す。

【 図 4 】 図 4 は、遺伝子当たり三つ (A ~ C または D ~ F) または四つ (G ~ J または K ~ N) の gRNA のいずれかを有する、四つの仮説遺伝子に対する人工リード数を格納する例示的なデータ構造を示し、それによって各 gRNA の選択前および選択後リード数がとられ、それによって閾値 30 を超える選択前 n および選択後 n' リード数を識別し合計する (N および N')、また、確率は示される式に従って決定される。

【 図 5 】 図 5 は、N 個の gRNA を最初に使用し、10 日後の細胞培養で N' 個の gRNA が存在する、開示された方法の人工データおよび人工結果を格納する例示的なデータ構造を示す。示されるように、標的領域 1 または標的領域 2 は、n 個の gRNA を有し、10 日間の培養後の細胞中に存在する n' 個の gRNA の可能性が提示される (p 値として

50

提示される確率)。

【図6】図6は、約21,000個の遺伝子(G#)が関与する実験の結果を示し、四つの選択前および三つの選択後実験からの各gRNA(g)のリード数が示され、リード数を合計し(合計)、閾値30を超えるリード数を識別し(存在の合計(sum of presence))、合計する(合計)。

【図7】図7は、三日間(d03)、六日間(d06)、および十日間(d10)にわたる細胞培養のいくつかの平行実験(xp#)で使用されたサンプルgRNAライブラリー(Gecko AおよびGecko B)を示し、その後、FRET蛍光によって測定されるタウ凝集が決定された。Gecko AおよびBライブラリーの構成は、特定の遺伝子標的化gRNA、特定のマイクロRNA標的化gRNA、および非標的化gRNA、および標的当たりのgRNAのおよその数を含むとして示されている。

10

【図8】図8は、3日間、6日間、および10日間にわたる各実験(xp#)で使用されたGecko Aライブラリーからの各gRNAの正規化前のリード数を示し、それによって、gRNAを使用して標的の不活化を方向付けた。

【図9】図9は、図8に示す実験からのGecko Aライブラリーのリード数の正規化を示す。正規化は、リード数の中央値に基づいており、10日目は選択後である。

【図10】図10は、標的が、異なる数のgRNAを有してもよく、また一部のサンプルが、10日目に他のものよりも多くの「存在する」gRNAを有するため、「存在」の頻度を集計する別の方法は、遺伝子が「存在する」確率を計算することである、ということを示す。確率は、Gecko AライブラリーgRNAを使用して、10日目(選択後)にスクリーンされた五つの異なる遺伝子について計算された。

20

【発明を実施するための形態】

【0015】

本明細書に組み込まれ、かつ本明細書の一部をなす添付の図面は、実施形態を例証し、この説明とともに、本方法およびシステムの原理を説明する役割を果たすものである。本方法およびシステムに関する開示および説明に先立って、本方法およびシステムが特定の方法、特定のコンポーネントまたは特定の実装形態に限定されないことを理解すべきである。本明細書中で使用されている用語は、もっぱら特定の実施形態の説明を目的としたものであって、限定することを意図するものではないこともまた、理解すべきである。

【0016】

本明細書および添付の特許請求の範囲で使用される場合、単数形「a」、「an」、および「the」は、文脈から他の意味に解釈されることが明白な場合を除き、複数の指示対象を含む。

30

【0017】

本明細書で使用される場合、用語「プローブ」および「ガイドRNA(gRNA)」および「ガイド」は互換的に使用される。一実施形態では、gRNAはまた、gRNAをコードするDNAの形態で提供されてもよい。

【0018】

本明細書で使用される場合、「Casタンパク質」は、野生型タンパク質(すなわち、自然界で生じるもの)、改変Casタンパク質(すなわち、Casタンパク質バリエーション)、または野生型もしくは改変Casタンパク質の断片であり得る。Casタンパク質はまた、野生型または改変Casタンパク質の触媒活性に関して、活性バリエーションまたは断片であってもよい。

40

【0019】

第一の態様では、本開示は、例えば、他の遺伝子または遺伝子産物の発現を調節する遺伝子または遺伝子産物を識別する方法を特徴とする。本方法は、例えば、CRISPRガイドを用いた摂動後の陽性選択を明示するのに有用である。

【0020】

一実施形態では(図1に示す)、方法は、cas9陽性細胞の第一の培養物をウイルスベクターのライブラリーで感染させる工程110であって、ライブラリーは、細胞のゲノム

50

内のDNAの標的領域を切断するための少なくとも3つのガイドRNAを含む、感染させる工程、細胞を配列決定してgRNAの各々のリード数を取得する工程120、DNAの標的領域当たりで、そのリード数が背景閾値を超える、gRNAのそれぞれの数を合計し() (= n)、標的領域全体にわたり、そのリード数が背景閾値を超える、gRNAの総数を合計する() (= N)工程130、cas9陽性細胞の第二の培養物を、ウイルスベクターのライブラリーで感染させる工程140、第二の培養物の細胞を、指定された表現型を有する、または指定された表現型を有しないのいずれかとしてカテゴライズする工程150、指定された表現型を有する細胞を選択し、選択された細胞を配列決定してgRNAの各々の選択後リード数を取得する工程160、DNAの標的領域当たりで、その選択後リード数が背景閾値を超える選択された細胞において、gRNAのそれぞれの数を合計し() (= n')、標的領域全体にわたりそのリード数が閾値を超える選択された細胞のgRNAの総数を合計する() (= N')工程170、DNAの標的領域について、式

10

【0021】

【数4】

$$\frac{\binom{n'}{n} \cdot \binom{N'-n'}{N-n}}{\binom{N'}{N}}$$

20

【0022】

に従って、

選択された細胞内の標的領域についてn' gRNAを偶然に観察する確率を計算する工程180であって、式中

【0023】

【数5】

$$\binom{x}{y} = \frac{y!}{x!(y-x)!}$$

30

【0024】

は、yオブジェクトからxオブジェクトを選択する方法の数を算出する、計算する工程、および遺伝子を含むDNAの標的領域について、式

【0025】

【数6】

$$\sum_{i=n'}^n \frac{\binom{i}{n} \cdot \binom{N'-i}{N-n}}{\binom{N'}{N}}$$

40

【0026】

に従って選択された細胞内の遺伝子のn'以上のgRNAを偶然に観察する確率を計算する工程190、を含む。

一実施形態では(これもまた図1に示す)、方法は、cas9陽性細胞の第一の培養物をウイルスベクターのライブラリーで感染させる工程110であって、ライブラリーは、細胞のゲノム内のDNAの標的領域の転写を強化するための少なくとも3つのガイドRNA(gRNA)を含む、感染させる工程、細胞を配列決定してgRNAの各々のリード数を取得する工程120、DNAの標的領域当たりで、そのリード数が背景閾値を超える、gRNAのそれぞれの数を合計し() (= n)、標的領域全体にわたりそのリード数が背景閾値を超える、gRNAの総数を合計する() (= N)工程130、cas9陽

50

性細胞の第二の培養物を、ウイルスベクターのライブラリーで感染させる工程 140、第二の培養物の細胞を、指定された表現型を有する、または指定された表現型を有しないのいずれかとしてカテゴリ化する工程 150、指定された表現型を有する細胞を選択し、選択された細胞を配列決定して gRNA の各々の選択後リード数を取得する工程 160、DNA の標的領域当たりで、その選択後リード数が背景閾値を超える選択された細胞において、gRNA のそれぞれの数を合計し () (= n ')、標的領域全体にわたり、そのリード数が閾値を超える選択された細胞の、gRNA の総数を合計する () (= N ') 工程 170、DNA の標的領域について、式

【0027】

【数7】

$$\frac{\binom{n'}{n} \cdot \binom{N'-n'}{N-n}}{\binom{N'}{N}}$$

10

【0028】

に従って、

選択された細胞内の標的領域について n ' gRNA を偶然に観察する確率を計算する工程 180 であって、式中

【0029】

【数8】

$$\binom{x}{y} = \frac{y!}{x!(y-x)!}$$

20

【0030】

は、y オブジェクトから x オブジェクトを選択する方法の数を算出する、計算する工程、および遺伝子を含む DNA の標的領域について、式

【0031】

【数9】

$$\sum_{i=n'}^n \frac{\binom{i}{n} \cdot \binom{N'-i}{N-n}}{\binom{N'}{N}}$$

30

【0032】

に従って選択された細胞内の遺伝子の n ' 以上の gRNA を偶然に観察する確率を計算する工程 190、を含む。

一実施形態では、第一の培養物は、CRISPR 技術に従って感染させることができる。

一実施形態では、cas9 陽性細胞の第一の培養物をウイルスベクターのライブラリーで感染させること (例えば、図 2 ; 201)、および cas9 陽性細胞の第二の培養物をウイルスベクターのライブラリーで感染させること (例えば、図 2 ; 205) は、CRISPR gRNA ライブラリーを使用することを含む。CRISPR gRNA ライブラリーは、例えば、ゲノム中の各遺伝子を標的とする一つ以上の gRNA (例えば、sgRNA) を含むゲノムワイド gRNA ノックアウトライブラリーを含むノックアウトライブラリーであってもよく、ここでゲノムは任意のタイプのゲノムであってもよい。一部の実施形態では、gRNA ライブラリーは、プールされたライブラリーを含んでもよい。プールされたライブラリーの非限定的な例は、ゲノムスケール CRISPR ノックアウト (GeCKO) ライブラリーを含む。例えば、Shalem O et al. (2014) Science 343 : 84 - 7 および Sanjana NE et al. (2014)

40

50

Nat. Methods 11:783-4を参照のこと。ライブラリー中のgRNAは、DNA中の任意の数の標的領域（例えば、遺伝子）を標的とすることができる。例えば、gRNAは、約50個以上の遺伝子、約100個以上の遺伝子、約200個以上の遺伝子、約300個以上の遺伝子、約400個以上の遺伝子、約500個以上の遺伝子、約1000個以上の遺伝子、約2000個以上の遺伝子、約3000個以上の遺伝子、約4000個以上の遺伝子、約5000個以上の遺伝子、約10000個以上の遺伝子、または約20000個以上の遺伝子を標的とすることができる。一部のライブラリーでは、gRNAは、特定のシグナル伝達経路の遺伝子を標的とするように選択することができる。gRNAライブラリーは、広範な感染多重度（MOI）を使用して投与することができる。一部の態様では、より低いMOIを使用して、細胞当たり一つのgRNAをもたらす感染を促進することができる。

10

【0033】

一実施形態では、cas陽性細胞は、DNAの標的領域を切断するためのcasタンパク質、または転写を調節する（例えば、転写を増強または抑制する）ためのcasタンパク質を含み得る。DNAの標的領域を切断するためのcasタンパク質は、RNA結合ドメインおよびヌクレアーゼドメインを含み得る。転写を調節するためのcasタンパク質は、ヌクレアーゼ活性をもはや有しないように、不活性化される。不活性化Cas（例えば、dCas-9）は、転写活性化因子または転写リプレッサーに融合され得る。したがって、一実施形態では、野生型活性を有するcas-9、または不活性化cas-9を含むcas-9陽性細胞が開示される。不活性化cas-9は、例えば、少なくとも一つの転写活性化ドメインに融合され得る。一つ以上のgRNAは、対象の遺伝子の転写開始部位の上流の標的配列に結合することができ、DNAを切断する代わりに、dCas-9および転写調節因子は、対象の遺伝子の転写を活性化するか、または転写を抑制するかのいずれかで役割を果たすことができる。dCas-9が一つ以上の転写活性化因子に融合される場合、一つ以上の転写活性化因子は、対象の遺伝子の転写開始部位に転写因子を動員し、それにより転写を活性化または上方制御する。dCas-9が転写リプレッサーに融合される場合、転写は阻害、下方制御、または抑制される。

20

【0034】

一実施形態では、DNAの標的領域は遺伝子であってもよい。一実施形態では、DNAの標的領域は、遺伝子のプロモーター領域または調節因子領域であってもよい。一実施形態では、DNAの標的領域は、下流遺伝子またはタンパク質を調節する。一実施形態では、下流遺伝子またはタンパク質を調節することは、下流遺伝子またはタンパク質の活性化または阻害を含む。

30

【0035】

一部の実施形態では、細胞は、選択マーカー系を含む。例えば、本明細書の方法が開示されるように、第一の培養物のcas-9陽性細胞は、一つ以上の選択マーカーを含有するように改変され得る。選択マーカー系は、調節されているタンパク質でのみ活性化される選択マーカーに融合または連結されたマーカータンパク質に関与することができる。「マーカータンパク質」は、調節され得る任意のタンパク質であってもよく、調節されるとは、マーカータンパク質が形状を変えることか、一つ以上の他のタンパク質または核酸に結合することか、活性の変化を有することか、または発現レベルの変化を有することを意味する。マーカータンパク質は、一つ以上のgRNAによって標的化される遺伝子によって調節されてもよく、その結果、遺伝子がCas9によって切断される場合、マーカータンパク質は調節されず、選択マーカーが活性化されない。したがって、活性化された選択マーカーを有しない細胞を、細胞内の選択マーカーに融合または連結されたマーカータンパク質を調節する遺伝子を含む細胞として選択することができる。例えば、タウは、CFPと融合または連結することができる。別のタウタンパク質をYFPに融合または連結することができる。タウタンパク質の凝集に伴い、CFPに当たった光が青色光として放射され、YFPが励起され、黄色光が放射される。タウタンパク質の凝集がない場合、CFPから放射される青色光は、他のタウタンパク質のYFPを励起することができないため、

40

50

黄色光は放射されない。したがって、タウを調節する遺伝子が一つ以上の gRNA によって標的化され、したがって切断される場合、黄色光は放射されない。最終的に、この遺伝子は、タウを調節する（例えば、タウ凝集を引き起こす）遺伝子として識別され得る。CRISPRaにおいて、gRNAが標的領域に結合する場合、下流遺伝子が活性化され、したがって、過剰発現または過剰量の選択マーカーを有する細胞が選択され得る。例えば、上述のタウタンパク質を使用して、gRNAを有しない細胞と比較して増加した量の黄色光が、タウを調節する遺伝子を示すことができる。既知の経路、具体的には疾患経路に存在する任意の数のタンパク質をマーカータンパク質として使用して、そのマーカータンパク質を調節し、最終的に特定の疾患に関与することが見出され得る遺伝子を識別することができる。

10

【0036】

一実施形態では、細胞のセットは、（例えば、表現型によって）カテゴライズされ、選択されてもよく（例えば、図2；206）、初期感染後のある程度の時間、例えば、感染後、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15日など、実施され得る。実施形態では、細胞のセット中に含有される選択マーカーを利用する、多くのタイプのスクリーン/選択機構が使用され得る。一実施形態では、選択機構は、第二の細胞集団を薬剤に曝露するか、または第二の細胞集団をタンパク質活性または発現レベルを特定する物質に曝露することのうちの一つ以上を含む。一実施形態では、指定された表現型を有する細胞を選択することは、一つ以上の選択マーカーに基づいて細胞をソートすることを含む。一実施形態では、細胞の生存率を、選択に使用することができる。

20

【0037】

一実施形態では、第二の培養物の細胞を、指定された表現型を有する、または指定された表現型を有しないのいずれかとしてカテゴライズすることは、細胞における指定された表現型の存在または不在を特定することを含む。第二の培養物の細胞を、指定された表現型を有する、または指定された表現型を有しないのいずれかとしてカテゴライズすることは、選択機構を第二の細胞培養物に適用することを含み得る。一実施形態では、表現型が特定されると、指定された表現型を有する（または指定された表現型を有しない）細胞を選択することが実施され得る。表現型は、細胞成長、増殖、形態、酵素機能、シグナル伝達、発現パターン、下流発現パターン、レポーター遺伝子活性化、ホルモン放出、成長因子放出、神経伝達物質放出、リガンド結合、アポトーシス、および生成物形成における変化などのアッセイで測定できる任意の観察可能な特徴または機能的効果である。一実施形態では、指定された表現型は蛍光または細胞生存であり得る。

30

【0038】

細胞は、例えば、マーカーのゲノム統合によって、またはゲノム内に統合されていない細胞内マーカーの存在によって、直接選択することができる表現型を伝達するために改変され得る。本明細書で使用される場合、「マーカー」は最も一般的には、細胞中に存在する（例えば、発現される）場合に、そのマーカーを含有するとしてその細胞を視覚化するか、または識別する属性または表現型をもたらす生物学的特徴または形質を指す。様々なマーカータイプが一般的に使用され、そして例えば、可視マーカー、例えば発色、例えばlacZ相補（ β -ガラクトシダーゼ）、または蛍光、例えば緑色蛍光タンパク質（GFP）またはGFP融合タンパク質、RF P、BF P、ルシフェラーゼ、 β -ガラクトシダーゼ、強化緑色蛍光タンパク質（eGFP）、シアン蛍光タンパク質（CFP）、黄色蛍光タンパク質（YFP）、強化黄色蛍光タンパク質（eYFP）、青色蛍光タンパク質（BF P）、強化青色蛍光タンパク質（eBF P）、DsRed、ZsGreen、MmGFP、mPlum、mCherry、tdTomato、mイチゴ、J-赤、mオレンジ、mKO、mシトリン、Venus、YPet、エメラルド、CyPet、セルリアン、T-サファイア、およびアルカリホスファターゼの発現など、表現型マーカー（成長率、細胞形態、コロニーの色またはコロニーの形態、温度感受性）、栄養要求性マーカー（成長要件）、抗生物質感受性および耐性、抗原感受性により区別できる生体分子などの分子マーカー（例えば、血液型抗原および組織適合性マーカー）、細胞表面マーカー（例えば、H2K

40

50

K)、酵素マーカー、ならびに核酸マーカー、例えば制限断片長多型(RFLP)、一塩基多型(SNP)、および他の様々な増幅可能な遺伝子多型、であり得る。したがって、例えば、一つ以上の選択マーカーは、 β -ガラクトシダーゼまたはルシフェラーゼなどの、検出可能な酵素であり得る。

【0039】

「選択マーカー」または「スクリーニングマーカー」または「陽性選択マーカー」は、細胞中に存在するとき、選択マーカー形質を発現しない他の細胞からのそれらの細胞の選択または分離を可能にする属性または表現型をもたらすマーカーを指す。様々な遺伝子が、選択マーカーとして使用され得る、例えば、薬剤耐性または栄養要求性救済をコードする遺伝子は広く知られている。例えば、カナマイシン(ネオマイシン)耐性は、細菌性カナマイシン耐性をコードする遺伝子を担持するプラスミドを取り込んだ細菌を選択するための形質として使用することができる(例えば、酵素ネオマイシンホスホトランスフェラーゼII)。非トランスフェクト細胞は、培養物がネオマイシンまたは類似の抗生物質で処理されると最終的に死滅する。細胞のセットは、薬物スクリーンで使用され、薬剤耐性を付与する遺伝子を識別することができる。細胞は、対象の薬剤で処理されてもよく、富化されているgRNAは、変異時に薬剤耐性を付与する遺伝子と関連付けられる。ウイルスまたは細菌病原体に対する耐性のスクリーンを使用して、感染または病原体の複製を防ぐ遺伝子を識別することができる。薬剤耐性スクリーンにおいてと同様に、病原体曝露後の生存は、強力な選択を提供する。癌では、陰性選択CRISPRスクリーンは、分子標的化療法の基礎を提供することができる特定の癌サブタイプにおける「癌遺伝子依存性」を識別し得る。発生研究については、ヒトおよびマウスの多能性細胞におけるスクリーニングは、多能性または別個の細胞型への分化に必要な遺伝子を正確に示し得る。

10

20

【0040】

一実施形態では、細胞は、選択マーカー系を含む。選択マーカー系は、蛍光タンパク質FREYバイオセンサーを含み得る。例えば、タウは、CFPと融合または連結することができる。別のタウタンパク質をYFPに融合または連結することができる。タウタンパク質の凝集に伴い、CFPに当たった光が青色光として放射され、YFPが励起され、黄色光が放射される。タウタンパク質の凝集がない場合、CFPから放射される青色光は、他のタウタンパク質のYFPを励起することができないため、黄色光は放射されない。したがって、タウを調節する遺伝子が一つ以上のgRNAによって標的化され、したがって切断される場合、黄色光は放射されない。最終的に、この遺伝子は、タウを調節する(例えば、タウ凝集を引き起こす)遺伝子として識別され得る。CRISPRaにおいて、gRNAが標的領域に結合する場合、下流遺伝子が活性化され、したがって、過剰発現または過剰量の選択マーカーを有する細胞が選択され得る。例えば、上述のタウタンパク質FREYバイオセンサーを使用して、gRNAを有しない細胞と比較して増加した量の黄色光が、タウを調節する遺伝子に結合されたgRNAを示すことができる。既知の経路、具体的には疾患経路に存在する任意の数のタンパク質をマーカータンパク質として使用して、そのマーカータンパク質を調節し、最終的に特定の疾患経路に関与することが見出される遺伝子を識別することができる。

30

【0041】

一実施形態では、第一の培養物の細胞および第二の培養物の選択された細胞は配列決定されてもよい(例えば、図2; 202および207)。第一の培養物の細胞および第二の培養物の選択された細胞は、感染後、異なる時点で配列決定されてもよい。第一の培養物の細胞および第二の培養物の選択された細胞は、選択前のリード数(第一の培養物の細胞)を生成し、選択後のリード数(第二の培養物の選択された細胞)を生成するために、感染後の異なる時点で配列決定されてもよい。例えば、第一の培養物の細胞は、感染の3日後に配列決定されてもよく、第二の培養物の選択された細胞は、感染の10日後に配列決定されてもよい。細胞は、NGSなどの任意の利用可能な配列決定技術を使用して配列決定することができる。細胞の核酸を配列決定して、配列データを生成することができる。配列データは、リード数を含むことができる。配列データは、ライブラリーの一つ以上のg

40

50

R N A についてのリード数を含むことができる。細胞を配列決定することは、データ構造に格納され得る配列データを生成することができる。データ構造は、一つ以上の核酸配列および/またはサンプル識別子を含むことができる。

【 0 0 4 2 】

配列決定から生じるリード数は、複製の頻繁な不在、g R N A ノックアウト効率の変動性、およびリード数分布の変動性を含む、C R I S P R スクリーン解析に存在する従来のバイアスを被る。これらのバイアスは、陰性二項式アプローチ、 \log_2 比アプローチ (\log_2 - r a t i o a p p r o a c h e)、および対の t 検定アプローチに従ってリード数を解析する際に不良な結果をもたらす。こうしたアプローチは、同じ遺伝子についての g R N A 間、ならびにリピート間のある一定の均質性/一致を必要とする。これらの既存のアプローチは、例えば、異なる感染効率、異なる遺伝子編集効率、スクリーニングライブラリー内の初期ウイルス数、および同じ表現型を有する他のガイドの存在から生じ得る、同じ遺伝子についての g R N A 間およびリピート間の大きな変動を処理することができない。これらのバイアスは、C R I S P R 実験内、および複数の C R I S P R 実験の間に存在し得る。本明細書に記述された、標的領域当たりの g R N A の合計数と、すべての標的領域にわたる g R N A の総数とを決定する工程は、リード数の大きな変動に対して堅牢であるリード数処理を表す。本開示の方法の一実施形態は、各ガイドの正確なリード数の代わりに、個々の実験における遺伝子当たりのガイドの陽性発生に基づくため、既存のアプローチに対する利点を提供する。

10

【 0 0 4 3 】

リード数を正規化することができる。例えば、異なるサンプルからのリード数は、ライブラリーサイズおよびリード数分布の影響を調整するために、中央値正規化され得る。実施形態では、所与の N C R I S P R / C a s 9 ノックアウトスクリーニング実験は、M g R N A のセットで行われ、実験 j 中の g R N A のリード数 i は、 x_{ij} 、 $1 \leq i \leq M$ 、 $1 \leq j \leq N$ である。配列決定深度 (またはライブラリーサイズ) は実験間で異なる場合があるため、リード数は、全実験に比率中央値法を適用することによって調整することができる。実施形態では、調整されたリード数 x'_{ij} は、方程式：

20

【 0 0 4 4 】

【 数 1 0 】

$$x'_{ij} = \frac{x_{ij}}{\sum_{i=1}^N x_{ij}} * S$$

30

【 0 0 4 5 】

に従って計算されてもよく、式中、S は、 $j = 1 \sim M$ について

【 0 0 4 6 】

【 数 1 1 】

$$\left(\sum_{i=1}^N x_{i1}, \sum_{i=1}^N x_{i2}, \dots, \sum_{i=1}^N x_{iM} \right)$$

40

【 0 0 4 7 】

の中央値である。別の実施形態では、調整されたリード数 x'_{ij} は、 x_{ij} / S_j の丸めた値として計算されてもよく、式中、 s_j は、実験 j における寸法係数であり、個々の s g R N A リード数から計算された全ての寸法係数の中央値として算出される：

【 0 0 4 8 】

【 数 1 2 】

50

$$s_j = \text{median}_i \left\{ \frac{x_{ij}}{\hat{x}_i} \right\}$$

【 0 0 4 9 】

式中、

【 0 0 5 0 】

【 数 1 3 】

$$\hat{x}_i$$

10

【 0 0 5 1 】

は、gRNAのリード数*i*の幾何平均である。

【 0 0 5 2 】

【 数 1 4 】

$$\hat{x}_i = \left(\prod_{k=1}^N x_{ik} \right)^{1/N}$$

【 0 0 5 3 】

あるいは、gRNA当たりのリード数および遺伝子当たりのリード数は、100万当たりのカウント、総数、または寸法係数正規化のいずれかを使用して正規化されてもよい。Anders, S. および Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol.*, 11, R106 (その全体が参照により本明細書に組み込まれる)を参照されたい。

20

【 0 0 5 4 】

一実施形態では、本開示の方法は、配列データを使用して、DNAの標的領域当たりで、その選択後リード数が背景閾値を超える選択された細胞における、gRNAのそれぞれの数(*n'*)の合計()を決定する(例えば、図2; 203、208)。本方法はまた、全ての標的領域にわたって、そのリード数が背景閾値を超える選択された細胞の、gRNAの総数(*N'*)を決定することも含み得る(例えば、図2; 204、209)。

30

【 0 0 5 5 】

したがって、開示された方法は、選択された細胞の配列におけるライブラリーからgRNAの陽性発生を識別することができる。リード数を含む配列データを解析して、DNAの各標的領域(例えば、遺伝子)について、選択後「存在の合計」または*n'*を決定することができる。各gRNAの個々のリード数を背景閾値と比較することによって、「存在の合計」、またはDNAの標的領域当たりで存在するgRNAのそれぞれの数を決定することができる。DNAの標的領域当たりで、そのリード数が背景閾値を超える、gRNAのそれぞれの数を決定することは、コンピューティングデバイスによって実施され得る。背景閾値は、配列データ中の背景ノイズを低減するのに十分な任意の値とすることができる。例えば、30を背景閾値として使用してもよい。したがって、リード数によって示される存在するgRNAの量と対比して、「存在の合計」は存在するgRNAの数(そのリード数が背景閾値を超える)を示す。

40

【 0 0 5 6 】

一実施形態における、cas9陽性細胞の第二の培養物をウイルスベクターのライブラリーで感染させる工程と、第二の培養物の細胞を、指定された表現型を有する、または指定された表現型を有しないのいずれかとしてカテゴリ化する工程と、指定された表現型を有する細胞を選択し、選択された細胞を配列決定して、gRNAの各々の選択後リード数を取得する工程と、DNAの標的領域当たりで、その選択後リード数が背景閾値を超える選択された細胞において、gRNAのそれぞれの数を合計する()工程であって、式中

50

= n' である、合計する工程と、標的領域全体にわたり、そのリード数が閾値を超える選択された細胞の、gRNAの総数を合計する工程()であって、式中 = N' が任意の回数リピートされ得る、合計する工程。

【0057】

一実施形態では、本開示の方法はさらに、選択された細胞内の遺伝子の n' 以上の gRNA を偶然に観察する確率に基づいて、標的領域が陽的に選択されると識別することを含む(例えば、図2; 211)。一部の実施形態では、標的領域が陽的に選択されると識別することは、選択された細胞における対象の配列の n' 以上の gRNA を偶然に観察する確率が閾値を満たすと決定することを含む。

【0058】

一部の実施形態では、本開示の方法はさらに、選択された細胞内の遺伝子の n' 以上の gRNA を偶然に観察する確率に基づいて、標的領域を第二の遺伝子の修飾因子(modifier)として識別することを含む。一部の実施形態では、本開示の方法はさらに、選択された細胞内の遺伝子の n' 以上の gRNA を偶然に観察する確率に基づいて、標的領域を治療標的として識別することを含む。一部の実施形態では、本開示の方法はさらに、選択された細胞内の遺伝子の n' 以上の gRNA を偶然に観察する確率に基づいて、標的領域が指定された表現型と相関すると識別することを含む。一部の実施形態では、本開示の方法はさらに、選択された細胞内の遺伝子の n' 以上の gRNA を偶然に観察する確率に基づいて、標的領域が保護効果を示すと識別することを含む。したがって、開示された方法は、疾患経路に關与する標的領域(例えば、遺伝子)、一つ以上の他の遺伝子/タンパク質の調節に關与する標的領域(例えば、遺伝子)、および/または表現型に關連する標的領域(例えば、遺伝子)の識別を助けることができる。候補遺伝子の遺伝子発現の調製が、選択された表現型に変化を生じさせる場合、DNAの領域(例えば、候補遺伝子)は、選択された表現型と「關連」してもよい。

【0059】

一実施形態では、開示された方法は、各 gRNA について、濃縮スコアを決定することをさらに含む。一部の実施形態では、各 gRNA について、濃縮スコアを決定することは、N/N' を評価することを含む。濃縮スコアは、閾値を超えるように決定されてもよい。閾値は、他の標的領域の他の濃縮スコアに関するものであってもよい。高濃縮スコアを有し、偶然に gRNA が存在する確率が低い標的領域を使用して、標的領域を表現型と關連すると識別してもよい。

【0060】

例示的な実施形態において、方法およびシステムは、図3に示され、以下に記載されるように、コンピュータ301上で実装され得る。同様に、方法およびシステムは、一つ以上のコンピュータを利用して、一つ以上の場所で一つ以上の機能を実施できる。図3は、方法を実施するための例示的な操作環境を示すブロック図である。この例示的な操作環境は、あくまで操作環境の一例にすぎず、操作環境アーキテクチャの使用または機能の範囲に關する何らかの制限を示唆することを意図したものではない。また、いかなる操作環境も、例示的な操作環境において図示されるコンポーネントのいずれか一つもしくは組み合わせに關連する何らかの依存性または要件を有するものとして解釈すべきではない。

【0061】

本方法およびシステムは、多数の他の汎用もしくは特殊用途向けコンピューティングシステム環境または構成で動作可能であり得る。システムおよび方法と共に使用するのに好適であり得るコンピューティングシステム、環境、および/または構成の例には、パーソナルコンピュータ、サーバーコンピュータ、ラップトップデバイス、およびマルチプロセッサシステムが含まれるが、これらに限定されるものではない。追加的な例には、セットトップボックス、プログラマブル大衆消費電子製品、ネットワークPC、ミニコンピュータ、メインフレームコンピュータ、上記のシステムまたはデバイスのいずれかを含む分散コンピューティング環境が含まれる。

【0062】

10

20

30

40

50

方法およびシステムの処理は、ソフトウェアコンポーネントを介して実施できる。システムおよび方法は、一つ以上のコンピュータまたは他のデバイスを介して実行されるプログラムモジュールなどの、コンピュータ実行可能命令の一般的なコンテキストで記述できる。概して、プログラムモジュールは、コンピュータコード、ルーチン、プログラム、オブジェクト、コンポーネント、データ構造などを含み、それらによって特定のタスクが実行されるかまたは特定の抽象データ型が実施される。また、方法は、通信ネットワーク経由でリンクされたりリモートプロセッシングデバイスを介してタスクが実施されるグリッドベースおよび分散コンピューティング環境においても実践することができる。分散コンピューティング環境において、プログラムモジュールは、メモリ記憶デバイスを含むローカルおよびリモートコンピュータストレージ媒体の両方に配置できる。

10

【0063】

さらに、システムおよび方法は、コンピュータ301の形態のコンピューティングデバイスを介して実装され得る。コンピュータ301のコンポーネントには、限定されるものではないが、一つ以上のプロセッサ303と、システムメモリ312と、一つ以上のプロセッサ303を含む様々なシステムコンポーネントをシステムメモリ312に連結するシステムバス313と、が含まれ得る。システムは並列計算を利用できる。

【0064】

システムバス313は、多様なバスアーキテクチャのいずれかを使用した、メモリバスもしくはメモリコントローラ、周辺機器用バス、アクセラレーテッドグラフィックスポート、またはローカルバスを含む、いくつかの可能なタイプのバス構造のうちの一つ以上を表す。バス313、および本記述中に指定されている全てのバスはまた、有線または無線ネットワーク接続経由で実装することもでき、一つ以上のプロセッサ303を含む各サブシステム、大容量ストレージデバイス304、オペレーティングシステム305、ソフトウェア306、データ307、ネットワークアダプタ308、システムメモリ312、入出力インターフェース310、ディスプレイアダプタ309、ディスプレイデバイス311、およびヒトマシンインターフェース302は、この形態のバスを介して接続された物理的に別個の位置にある一つ以上のリモートコンピューティングデバイス314a、b、c内に収容され、事実上完全に分散されたシステムを実装し得る。

20

【0065】

コンピュータ301は、典型的には、様々なコンピュータ可読媒体を含む。例示的な可読媒体は、コンピュータ301によりアクセスできる任意の利用可能な媒体であってよく、例えば、揮発性および不揮発性媒体、リムーバブルおよび非リムーバブル媒体の両方が挙げられるが、これらに限定されるものではない。システムメモリ312は、ランダムアクセスメモリ(RAM)などの揮発性メモリ、および/またはリードオンリメモリ(ROM)などの不揮発性メモリの形態のコンピュータ可読媒体を含む。システムメモリ312は、典型的には、データ307のようなデータ、および/または一つ以上のプロセッサ303によって直ちにアクセス可能であり、かつ/または現在操作されているオペレーティングシステム305およびソフトウェア306などのプログラムモジュールを含む。

30

【0066】

別の実施形態では、コンピュータ301はまた、他のリムーバブル/非リムーバブルな、揮発性/不揮発性コンピュータストレージ媒体を含むこともできる。一例として、図3は、コンピュータ301用のコンピュータコード、コンピュータ可読命令、データ構造、プログラムモジュール、および他のデータの揮発性ストレージを提供することができる、大容量ストレージデバイス304が示されている。例えば、限定されるものではないが、大容量記憶デバイス304は、ハードディスク、リムーバブル磁気ディスク、リムーバブル光学式ディスク、磁気カセットまたは他の磁気ストレージデバイス、フラッシュメモリカード、CD-ROM、デジタル多用途ディスク(DVD)または他の光学式ストレージ、ランダムアクセスメモリ(RAM)、リードオンリメモリ(ROM)、および/または電氣的消去可能プログラマブルリードオンリメモリ(EEPROM)であり得る。

40

【0067】

50

任意選択的に、任意の数のプログラムモジュールが、オペレーティングシステム 305 およびソフトウェア 306 を例として含む大容量ストレージデバイス 304 に格納され得る。オペレーティングシステム 305 およびソフトウェア 306 (またはそれらのいくつかの組み合わせ) の各々には、プログラミングおよびソフトウェア 306 の要素を含めることができる。データ 307 はまた、大容量ストレージデバイス 304 に格納できる。データ 307 は、一つ以上のデータベースのいずれかに格納できる。そのようなデータベースの例には、DB 2 (登録商標)、MICROSOFT (登録商標) Access、MICROSOFT (登録商標) SQL Server、ORACLE (登録商標)、および/または MYSQL (登録商標)、POSTGRESQL (登録商標) が含まれる。データベースは、集中型とすることができ、または複数のシステムにわたって分散することができる。データ 307 は、配列決定データを含んでもよい。配列決定データは、リードデータ (例えば、リード数) を配列決定することを含んでもよい。コンピュータ 301 は、例えば、図 1 および図 2 のそれぞれ工程 120 および 202 で生成される第一のリード数データを受信してもよい。コンピュータ 301 は、例えば、図 1 および図 2 のそれぞれ工程 160 および 207 で生成される第二のリード数データを受信してもよい。

10

【0068】

別の実施形態では、ユーザは、入力デバイス (図示せず) を介して、コンピュータ 301 内にコマンドおよび情報を入力することができる。そのような入力デバイスの例としては、以下に限定されないが、キーボード、指示デバイス (例えば、「マウス」)、マイクロホン、ジョイスティック、スキャナー、グローブなどの触覚入力デバイス、および/または他の身体被覆物が挙げられる。これらのおよび他の入力デバイスは、システムバス 313 に連結されているヒト機械インターフェース 302 を介して、一つ以上のプロセッサ 303 に接続することができるが、パラレル・ポート、ゲーム・ポート、IEEE 1394 ポート (Firewire (登録商標) ポートとも称される)、シリアル・ポートやユニバーサル・シリアル・バス (USB) などの他のインターフェースおよびバス構造によって接続することができる。

20

【0069】

さらに別の実施形態では、ディスプレイデバイス 311 はまた、ディスプレイアダプタ 309 などのインターフェースを介してシステムバス 313 に接続できる。コンピュータ 301 に二つ以上のディスプレイアダプタ 309 を設けることができ、コンピュータ 301 に二つ以上のディスプレイデバイス 311 を設けることもできることが予期される。例えば、ディスプレイデバイスは、モニター、液晶ディスプレイ (LCD)、またはプロジェクターとすることができ、ディスプレイデバイス 311 に加えて、他の出力周辺デバイスには、入出力インターフェース 310 を介してコンピュータ 301 に接続できるスピーカ (図示せず) およびプリンタ (図示せず) などのコンポーネントを含めることができる。本方法の任意の工程および/または結果は、任意のフォームで出力デバイスに出力できる。そのような出力は、テキスト、グラフィカル、アニメーション、オーディオ、および/または触覚 (tactile) を含むが、これらに限定されない任意のフォームの視覚的表象であり得る。ディスプレイ 311 およびコンピュータ 301 は、一つのデバイスの一部である場合もあれば、別々のデバイスである場合もある。

30

40

【0070】

コンピュータ 301 は、一つ以上のリモートコンピューティングデバイス 314 a、b、c への論理的接続を使用してネットワーク環境で動作することができる。一例として、リモートコンピューティングデバイスは、パーソナルコンピュータ、ポータブルコンピュータ、スマートフォン、サーバー、ルーター、ネットワークコンピュータ、ピアデバイスまたは他の共通ネットワークノードなどであり得る。コンピュータ 301 とリモートコンピューティングデバイス 314 a、b、c との間の論理的接続は、ローカルエリアネットワーク (LAN) および/または一般的なワイドエリアネットワーク (WAN) などのネットワーク 315 を介して行うことができる。そのようなネットワーク接続は、ネットワークアダプタ 308 経由であり得る。ネットワークアダプタ 308 は、有線および無線の両

50

方の環境で実装できる。一実施形態では、システムメモリ 312 は、ネットワーク 315 経由で、一つ以上のリモートコンピューティングデバイス 314 a、b、c にアクセス可能にされた一つ以上のオブジェクトを格納できる。したがって、コンピュータ 301 は、クラウドベースのオブジェクトストレージとして機能できる。別の実施形態では、一つ以上のリモートコンピューティングデバイス 314 a、b、c のうちの一つ以上には、コンピュータ 301 へのアクセスを許可されたオブジェクトを一つ以上、および/または一つ以上のリモートコンピューティングデバイス 314 a、b、c のうち他方へのアクセスを許可されたオブジェクトを一つ以上格納できる。したがって、一つ以上のリモートコンピューティングデバイス 314 a、b、c はまた、クラウドベースのオブジェクトストレージとして機能できる。

10

【0071】

そのようなプログラムおよびコンポーネントは、コンピューティングデバイス 301 の異なるストレージコンポーネント内に様々な時間に存在し、コンピュータの一つ以上のプロセッサ 303 を介して実行されることが認識されるが、例証の便宜上、本明細書においてアプリケーションプログラムおよびオペレーティングシステム 305 などの他の実行可能プログラムコンポーネントは、離散的ブロックとして示されている。一実施形態では、ソフトウェア 306 および/またはデータ 307 の少なくとも一部分は、コンピューティングデバイス 301、リモートコンピューティングデバイス 314 a、b、c、および/またはそれらの組み合わせのうちの一つ以上の上で格納および/または実行され得る。したがって、ソフトウェア 306 および/またはデータ 307 は、ソフトウェア 306 および/またはデータ 307 へのアクセスを、ネットワーク 315 (例えば、インターネット) を介して実施できるクラウドコンピューティング環境内で動作できる。さらに、一実施形態では、コンピューティングデバイス 301、リモートコンピューティングデバイス 314 a、b、c、および/またはそれらの組み合わせのうちの一つ以上にわたって、データ 307 を同期化できる。

20

【0072】

ソフトウェア 306 の実装形態は、何らかの形態のコンピュータ可読媒体上に格納される場合もあれば、またはそのコンピュータ可読媒体を介して伝送される場合もある。方法のいずれも、コンピュータ可読媒体上に具現化されたコンピュータ可読命令によって実施することができる。コンピュータ可読媒体は、コンピュータによってアクセス可能な任意の利用可能媒体とすることができる。例として、かつ限定を意図するものではないが、コンピュータ可読媒体は、「コンピュータストレージ媒体」および「通信媒体」を含み得る。「コンピュータストレージ媒体」は、コンピュータ可読命令、データ構造、プログラムモジュールもしくは他のデータなどの情報を記憶するための任意の方法または技術で実装される揮発性および不揮発性のリムーバブル媒体および非リムーバブル媒体を備える。例示的なコンピュータストレージ媒体は、限定されるものではないが、RAM、ROM、EEPROM、フラッシュメモリもしくは他のメモリ技術、CD-ROM、デジタル多用途ディスク(DVD)、または他の光学式ストレージ、磁気カセット、磁気テープ、磁気ディスクストレージデバイスもしくは他の磁気ストレージデバイス、または、所望の情報を格納する目的に使用でき、かつコンピュータがアクセスできる任意の他の媒体を具備する。

30

40

【0073】

ソフトウェア 306 は、本明細書に開示される方法の一部または全ての工程を実施するように構成されてもよい。一実施形態では、ソフトウェア 306 は、DNA の複数の標的領域の各々に対する少なくとも 3 つのガイド RNA (gRNA) のライブラリーを含むベクターを用いた感染後の、第一の細胞集団の配列決定に基づいて、そのリード数が背景閾値を超える、DNA の複数の標的領域の各々について存在する gRNA のそれぞれの数 (n) を決定するように、DNA の複数の標的領域の各々に対する gRNA のそれぞれの数に基づいて、そのリード数が背景閾値を超える、DNA の複数の標的領域の全ての標的領域にわたる、第一の細胞集団に存在する gRNA の総数 (N) を決定するように、DNA の複数の標的領域の各々に対する少なくとも 3 つの gRNA のライブラリーを含むベクター

50

を用いた感染後の第二の細胞集団の配列決定に基づいて、そのリード数が背景閾値を超える、DNAの複数の標的領域の各々について存在するgRNAのそれぞれの数(n')を決定するように、DNAの複数の標的領域の各々に対するgRNAのそれぞれの数に基づいて、そのリード数が背景閾値を超えるDNAの複数の標的領域の全ての標的領域にわたって、第二の細胞集団に存在するgRNAの総数(N')を決定するように、n、N、n'、およびN'に基づき、複数の標的領域の各標的領域に対して、選択された細胞内の標的領域についてn' gRNAを偶然に観察する確率を決定するように、対象の配列を含む標的領域について、選択された細胞内の標的領域についてn' gRNAを偶然に観察する確率に基づいて、選択された細胞における対象の配列のn'以上のgRNAを偶然に観察する確率を決定するように、および選択された細胞における対象の配列のn'以上のgRNAを偶然に観察する確率に基づいて、対象の配列が陽的に選択されることを識別するように、構成されてもよい。

10

【0074】

DNAの複数の標的領域の各々に対するgRNAのそれぞれの数に基づいて、そのリード数が背景閾値を超える、DNAの複数の標的領域の全ての標的領域にわたる、第一の細胞集団に存在するgRNAの総数(N)を決定することは、そのリード数が背景閾値を超える第一の細胞集団中に存在する各gRNAを数えることを含み得る。

【0075】

DNAの複数の標的領域の各々に対するgRNAのそれぞれの数に基づいて、そのリード数が背景閾値を超えるDNAの複数の標的領域の全ての標的領域にわたる、第二の細胞集団中に存在するgRNAの総数(N')を決定することは、そのリード数が背景閾値を超える第二の細胞集団の選択された細胞中に存在する各gRNAを数えることを含み得る。

20

【0076】

複数の標的領域の各標的領域について、n、N、n'、およびN'に基づいて、選択された細胞内の標的領域についてn'ガイドを偶然に観察する確率を決定することは、

【0077】

【数15】

$$\frac{\binom{n'}{n} \cdot \binom{N'-n'}{N-n}}{\binom{N'}{N}}$$

30

【0078】

を評価することを含むことができ、式中

【0079】

【数16】

$$\binom{x}{y} = \frac{y!}{x!(y-x)!}$$

40

【0080】

は、yオブジェクトからxオブジェクトを選択する方法の数を算出する。対象の配列を含む標的領域について、選択された細胞内の標的領域についてn'ガイドを偶然に観察する確率に基づいて、選択された細胞における対象の配列のn'以上のgRNAを偶然に観察する確率を決定することは、

【0081】

【数17】

50

$$\sum_{i=n'}^n \frac{\binom{i}{n} \cdot \binom{N'-i}{N-n}}{\binom{N'}{N}}$$

【 0 0 8 2 】

を評価することを含み得る。

選択された細胞における対象の配列の n' 以上の $gRNA$ を偶然に観察できる確率に基づいて、対象の配列が陽的に選択されると識別することは、選択された細胞における対象の配列の n' 以上の $gRNA$ を偶然に観察する確率が閾値を満たすことを決定することを含み得る。

10

【 0 0 8 3 】

ソフトウェア 306 はさらに、各 $gRNA$ について、濃縮スコアを決定するように構成され得る。各 $gRNA$ について、濃縮スコアを決定することは、 N/N' を評価することを含み得る。

【 0 0 8 4 】

ソフトウェア 306 は、データ 307 として存在する $gRNA$ のリード数および数を格納するように構成されてもよい。図 4 は、データ 307 の実施形態を表す例示的なデータ構造 410 を示す。データ構造 410 は、一つ以上の表、アレイ等を備えてもよい。データ構造 410 は、複数の行および複数の列を含み得る。列「遺伝子#」は、DNA の固有の遺伝子または標的領域の識別子を含んでもよい。列「遺伝子当たりの $gRNA$ 」は、DNA の遺伝子または標的領域に結合し得る固有の $gRNA$ の識別子を含む。列「 $gRNA$ リード数 (選択前)」には、リード数の値が含まれる。列「 n 」は、列「 $gRNA$ リード数 (選択前)」に由来する DNA の標的領域 (例えば遺伝子) 当たりで存在する $gRNA$ の数を含む。列「 $gRNA$ リード数 (選択後)」には、リード数の値が含まれる。列「 n' 」は、列「 $gRNA$ リード数 (選択後)」に由来する DNA の標的領域 (例えば遺伝子) 当たりで存在する $gRNA$ の数を含む。

20

【 0 0 8 5 】

図 4 のデータ構造 410 は、図示の目的でデータ 307 として格納される人工値を含んで示されている。遺伝子 1 については、 $gRNA$ A、B、および C はライブラリーに存在し、遺伝子 1 に結合してもよい。配列決定後、 $gRNA$ A、B、および C は、それぞれ 100、200、および 10 のリード数で細胞内に存在した。背景閾値 30 は、リード数に適用されてもよく、またライブラリーからの $gRNA$ の陽性発生数は、 $gRNA$ A および B のみが 30 を超えるリード数を有するため、2 ($n_1 = 3$) であると、ソフトウェア 306 によって決定されてもよい。遺伝子 2 については、 $gRNA$ D、E、および F がライブラリーに存在し、遺伝子 2 に結合してもよい。配列決定後、 $gRNA$ D、E、および F は、それぞれ 500、300、および 200 のリード数で細胞内に存在した。背景閾値 30 は、リード数に適用されてもよく、またライブラリーからの $gRNA$ の陽性発生数は、 $gRNA$ D、E、および F が 30 を超えるリード数を有するため、3 ($n_2 = 3$) であると、ソフトウェア 306 によって決定されてもよい。遺伝子 3 については、 $gRNA$ G、H、I、および J がライブラリーに存在し、遺伝子 3 に結合してもよい。配列決定後、 $gRNA$ G、H、I、および J は、それぞれ 200、250、10、および 300 のリード数で細胞内に存在した。背景閾値 30 は、リード数に適用されてもよく、またライブラリーからの $gRNA$ の陽性発生数は、 $gRNA$ G、H、および J のみが 30 を超えるリード数を有するため、3 ($n_3 = 3$) であると、ソフトウェア 306 によって決定されてもよい。遺伝子 4 については、 $gRNA$ K、L、M、および N がライブラリーに存在し、遺伝子 4 に結合してもよい。配列決定後、 $gRNA$ K、L、M、および N は、それぞれ 100、200、200、および 100 のリード数で細胞中に存在した。背景閾値 30 は、リード数に適用されてもよく、ライブラリーからの $gRNA$ の陽性発生数は、 $gRNA$ K、L、M、および N が 30 を超えるリード数を有するため、4 (n_4

30

40

50

= 4) であるとソフトウェア 306 によって決定されてもよい。

【0086】

図4のデータ構造410に示すように、gRNAの総数またはNは12である。Nは、各遺伝子についてのnの値からの $2 + 3 + 3 + 4$ に由来する($n_1 + n_2 + n_3 + n_4$)。したがって、細胞集団を感染させるために使用された14個の異なるgRNA (gRNA A ~ N) のオリジナルライブラリーのうち、背景閾値を超える量で存在したのは12個の異なるgRNAのみであった。nおよびNの値は、データ構造に格納されてもよい。nおよびNの値に加えて、データ構造は、一つ以上の核酸配列(例えば、標的領域および/またはgRNA配列)および/または一つ以上のgRNA識別子を含み得る。

【0087】

列「遺伝子#」は、DNAの固有の遺伝子または標的領域の識別子を含む。列「遺伝子当たりのgRNA」は、DNAの遺伝子または標的領域に結合し得る固有のgRNAの識別子を含む。列「gRNAリード数(選択後)」には、リード数の値が含まれる。列「n'」は、列「gRNAリード数(選択後)」に由来するDNAの標的領域(例えば遺伝子)ごとに存在するgRNAの数を含む。図4のデータ構造410に示すように、遺伝子1については、gRNA A、B、およびCがライブラリーに存在し、遺伝子1に結合してもよい。選択後、配列決定後、gRNA A、B、およびCは、それぞれ0、50、および0のリード数で細胞内に存在した。背景閾値30は、リード数に適用されてもよく、またライブラリーからのgRNAの陽性発生数は、gRNA Bのみが30を超えるリード数を有するため、1($n'_1 = 1$)であると、ソフトウェア306によって決定されてもよい。遺伝子2については、gRNA D、E、およびFがライブラリーに存在し、遺伝子2に結合してもよい。選択後、配列決定後、gRNA D、E、およびFは、それぞれ100、100、および100のリード数で細胞内に存在した。背景閾値30は、リード数に適用されてもよく、またライブラリーからのgRNAの陽性発生数は、gRNA D、E、およびFが30を超えるリード数を有するため、3($n'_2 = 3$)であると、ソフトウェア306によって決定されてもよい。遺伝子3については、gRNA G、H、I、およびJがライブラリーに存在し、遺伝子3に結合してもよい。選択後、配列決定後、gRNA G、H、I、およびJは、それぞれ0、0、0、および20のリード数で細胞内に存在した。背景閾値30は、リード数に適用されてもよく、またライブラリーからのgRNAの陽性発生数は、いずれも30を超えるリード数を有しないため、0($n'_3 = 0$)であると、ソフトウェア306によって決定されてもよい。遺伝子4については、gRNA K、L、M、およびNがライブラリーに存在し、遺伝子4に結合してもよい。選択後、配列決定後、gRNA K、L、M、およびNは、それぞれ0、0、0、および20のリード数で細胞内に存在した。背景閾値30は、リード数に適用されてもよく、またライブラリーからのgRNAの陽性発生数は、いずれのgRNAも30を超えるリード数を有しないため、0($n'_4 = 0$)であると、ソフトウェア306によって決定されてもよい。

【0088】

そのリード数が背景閾値を超える選択された細胞内に存在する、全標的領域にわたる、gRNAの総数またはN'は、ソフトウェア306によって決定され得る。図4のデータ構造410に示すように、gRNAの総数またはN'は4である。N'は、各遺伝子についてのnの値からの $1 + 3 + 0 + 0$ に由来する($n'_1 + n'_2 + n'_3 + n'_4$)。したがって、細胞集団に感染させるために使用された14個の異なるgRNA (gRNA A ~ N) のオリジナルライブラリーのうち、背景閾値を超える量で選択された細胞内に存在したのは4個の異なるgRNAのみであった。n'およびN'の値は、データ構造に格納されてもよい。n'およびN'の値に加えて、データ構造は、一つ以上の核酸配列(例えば、標的領域および/またはgRNA配列)および/または一つ以上のgRNA識別子を含み得る。

【0089】

実施形態では、データ307はまた、ソフトウェア306の一つ以上の結果を格納するように構成されてもよい。図5は、例えば、データ構造410を入力として使用して、ソフ

10

20

30

40

50

トウェア 306 によって生成される、結果データ構造 510 の例を示す。データ構造 510 は、一つ以上の表、アレイ等を備えてもよい。図 5 は、例示の目的での人工データおよび人工結果を示す。正式な統計 p 値は、ライブラリーサイズ、遺伝子当たりのガイドの数、および各実験における陽性ガイドの総数を考慮して実験リポートにわたるガイドの数を陽性的に観察するために計算されてもよい。データ構造 510 は、63, 950 gRNA の初期ライブラリー (N) を示す。データ構造 510 は、選択後、4, 946 の gRNA (N') (gRNA の数量ではなく固有の gRNA の数) が実験 1 の細胞集団に、および 13, 606 の gRNA (N') (gRNA の量ではなく固有の gRNA の数) が実験 2 の細胞集団に残っていたことを示す。データ構造 510 に示されるように、標的領域 1 は、標的領域 1 の少なくとも一部に結合することができる三つの (3) gRNA を有する。実験 1 では、選択後細胞集団において、3 つの gRNA のうち 3 つが偶然に存在する確率は、選択後に 4, 946 の gRNA が細胞集団に残っていることを考慮すると、0.000462378 である。実験 2 では、選択後細胞集団において、3 つの gRNA のうち 3 つが偶然に存在する確率は、選択後に 13, 606 の gRNA が細胞集団に残っていることを考慮すると、0.009629 である。データ構造 510 は、3 つ中 2 つ、3 つ中 1 つ、および 3 つ中 0 の gRNA が、選択後細胞集団において偶然に存在する確率の結果を示す。

10

【0090】

データ構造 510 に示されるように、標的領域 2 は、標的領域 2 の少なくとも一部に結合することができる四つの (4) gRNA を有する。データ構造 510 は、4 つ中 4 つ、4 つ中 3 つ、4 つ中 2 つ、4 つ中 1 つ、および 4 つ中 0 の gRNA が、選択後細胞集団において偶然に存在する確率の結果を示す。データ構造 510 に示される結果から、確率は閾値を下回っている (例えば、十分に小さい) とソフトウェア 306 によって決定され得る。例えば、標的領域 2、実験 1 については、4 つの gRNA のうち 4 つが偶然に存在する確率は $3.57411E-05$ であり、4 つの gRNA のうち 4 つが、単に偶然に存在したわけではない可能性を示している。

20

【実施例】

【0091】

A. 実施例 1: タウ凝集の遺伝子修飾因子を識別するためのゲノムワイド CRISPR / Cas9 スクリーニングプラットフォームの開発

異常なタウタンパク質凝集のプロセスを変更する遺伝子および経路を特定するため、CRISPRヌクレアーゼ (CRISPRn) sgRNA ライブラリーを用いたゲノムワイドスクリーンを実施するためのプラットフォームを開発し、タウ疾患関連タンパク質凝集体 (すなわち、破壊された場合に、タウ線維化タンパク質の供給源に曝露されたときのタウ凝集体形成に対して細胞の感受性をより高くする遺伝子) によって細胞が「播種される」可能性を調節する遺伝子を識別した。スクリーンは、CFP または YFP のいずれかに融合される、P301S 病原性変異を有するタウ微小管結合ドメイン (MBD) を含むタウ 4 リピートドメイン、tau_4RD を安定的に発現する HEK293T 細胞からなるタウバイオセンサーヒト細胞株を採用した。すなわち、HEK293T 細胞株は、蛍光タンパク質 CFP または蛍光タンパク質 YFP: タウ 4RD - CFP / タウ 4RD - YFP (TCY) に融合された疾患関連タンパク質パリアントを安定的に発現する二つの導入遺伝子を含有し、タウリピートドメイン (4RD) は P301S 病原性変異を含む。

30

40

【0092】

これらのバイオセンサーシステムでは、タウ - CFP / タウ - YFP タンパク質凝集は、FRET シグナルを産生し、これはドナー CFP からアクセプター YFP への蛍光エネルギーの伝達の結果である。タウ凝集体を含有する FRET 陽性細胞は、フローサイトメトリーによりソーティングおよび単離することができる。ベースラインでは、刺激されていない細胞は、最小限の FRET シグナルで、安定した可溶性状態のレポーターを発現する。刺激 (例えば、種粒子のリボソームトランスフェクション) 時に、レポータータンパク質は凝集体を形成し、FRET シグナルを産生する。凝集体含有細胞は、FACS によって単

50

離することができる。安定的に増殖する凝集体含有細胞株、A g g [+] は、A g g [-] 細胞株のクローン性段階希釈 (c l o n a l s e r i a l d i l u t i o n) によって単離することができる。

【 0 0 9 3 】

遺伝子スクリーニングに有用となるように、このタウバイオセンサー細胞株にいくつかの改変を行った。まず、これらのタウバイオセンサー細胞を、レンチウイルスベクターを介して C a s 9 発現導入遺伝子 (S p C a s 9) を導入することによって改変した。C a s 9 を発現するクローン性トランスジェニック細胞株を、ブラストサイジンで選択し、クローン性段階希釈により単離して、単細胞誘導クローン (s i n g l e - c e l l - d e r i v e d c l o n e s) を得た。クローンを、q R T - P C R による C a s 9 発現のレベル、およびデジタル P C R による D N A 切断活性について評価した。 10

【 0 0 9 4 】

具体的には、C a s 9 変異効率を、二つの選択された標的遺伝子に対して、g R N A をコードするレンチウイルスの形質導入の3日後および7日後にデジタル P C R により評価した。切断効率は、より低発現のクローンで C a s 9 レベルによって制限された。十分なレベルの C a s 9 発現を有するクローンは、最大活性を達成するために必要であった。より低い C a s 9 発現を有するいくつかの誘導クローンは、標的配列を効率的に切断できなかったが、より高い発現を有するクローン (スクリーニングに使用されるものを含む) は、培養物中での3日後、約 8 0 % の効率で、遺伝子 P E R K および S N C A の標的配列で変異を生成することができた。効率的な切断は、g R N A 形質導入後3日ですでに観察されており、7日後ではわずかな改善のみであった。クローン 7 B 1 0 - C 3 を、後続のライブラリースクリーンに使用する高性能クローンとして選択した。 20

【 0 0 9 5 】

第二に、細胞をタウ播種活性に対して感受性にするための試薬および方法を開発した。タウ細胞間増殖は、凝集体含有細胞によって分泌されるタウ凝集活性から生じる場合がある。タウ凝集の細胞増殖を試験するために、Y F P に融合される、P 3 0 1 S 病原性変異を有するタウ微小管結合ドメイン (M B D) を含む、タウリピートドメイン、タウ__ 4 R D を安定的に発現する H E K 2 9 3 T 細胞からなるタウ Y F P 細胞株のサブクローンを取得した。

【 0 0 9 6 】

凝集状態でタウ - Y F P タンパク質が安定的に存在する細胞 (A g g [+]) を、リポフェクタミン試薬と混合された組み換え線維化タウで、これらのタウ - Y F P 細胞を処理して、これらの細胞によって安定的に発現されるタウ - Y F P タンパク質の凝集を播種することによって取得した。次いで、単細胞誘導クローンを得るために、「播種される」細胞を段階希釈した。次いで、これらのクローンを増殖させて、経時的な増殖継代および複数継代培養で、タウ Y F P 凝集体がすべての細胞で安定的に持続するクローン性細胞株を識別した。これらのタウ - Y F P __ A g g [+] クローンのうちの一つを使用して、タウ - Y F P __ A g g [+] 細胞が四日間コンフルエントであった培地を収集して、ならし培地を産生した。次いで、ならし培地 (C M) を、3 : 1 の C M : 新鮮培地の比率でナイーブバイオセンサータウ - C F P / タウ - Y F P 細胞に塗布し、タウ凝集をこれらのレシピエント細胞のうち少数の割合で誘導するようにした。リポフェクタミンは使用しなかった。リポフェクタミンを使用してレシピエント細胞をトリックしてタウ凝集を強制 / 増加させることなく、可能な限り生理学的なアッセイを有するために、リポフェクタミンは使用されなかった。凝集の尺度として F R E T シグナルを産生する細胞の割合を評価するためにフローサイトメトリーを使用することによって測定されるように、ならし培地は、細胞の約 0 . 1 % で一貫して F R E T を誘導した。 40

【 0 0 9 7 】

B . 実施例 2 : タウ凝集の遺伝子的修飾因子を識別するためのゲノムワイド C R I S P R / C a s 9 スクリーニング
F R E T (+) 細胞における濃縮 s g R N A としてのタウ凝集の変更遺伝子を明らかにす 50

るため、凝集体 (Agg [-]) なしの Cas9 発現 タウ - CFP / タウ - YFP バイオセンサー細胞を、二つのヒトゲノムワイド CRISPR sgRNA ライブラリー (GeCKO A および GeCKO B) (図 7) で、レンチウイルス送達アプローチを用いて形質導入し、各標的遺伝子でノックアウト変異を導入した。各 CRISPR sgRNA ライブラリーは、1 遺伝子当たり約 3 つの sgRNA の平均カバレッジ (二つのライブラリーを合わせて、1 遺伝子当たり合計 6 つの sgRNA) を有する機能的ノックアウト用の 5' 構成エクソンを標的とする。リード数分布 (すなわち、ライブラリー中の各 sgRNA の表現) は正常であり、各ライブラリーで類似していた。sgRNA は、標的外ゲノム配列とのミスマッチが二つ以下の sgRNA を避けることによって、標的外効果を回避するように設計した。このライブラリーは、19,050 のヒト遺伝子および 1864 miRNA を、1000 の非標的対照 sgRNA でカバーする。ライブラリーを、sgRNA 当たり > 300 個の細胞のカバレッジで、複数の感染 (MOI) < 0.3 で形質導入した。タウバイオセンサー細胞を、ピューロマイシン選択下で増殖させて、細胞ごとに固有の sgRNA の統合および発現を有する細胞を選択した。ピューロマイシン選択は、1 μ g / mL で形質導入の 24 時間後に開始した。五つの独立したスクリーニング複製を、一次スクリーンで使用した。

【0098】

形質導入後の 3 日目および 6 日目の細胞継代時に、完全形質導入細胞集団のサンプルを採取した。6 日目の継代後、細胞をならし培地中で増殖させ、それらを播種活性に対して感受性にさせた。10 日目に、蛍光補助細胞ソーティング (FACS) を使用して、FRET [+] 細胞の亜集団を特異的に単離した。スクリーニングは、五つの複製された実験から構成された。統合された sgRNA 構築物の DNA 単離および PCR 増幅により、各時点での sgRNA レポートリーの次世代配列決定 (NGS) による特徴解析が可能となった。

【0099】

NGS データの統計解析により、五つの実験の 10 日目の FRET [+] 亜集団において濃縮された sgRNA を、3 日目および 6 日目の早期の時点での sgRNA レポートリーと比較して、識別することが可能となった。潜在的なタウ修飾因子を識別する最初の戦略は、DNA 配列決定を使用して、DESeq アルゴリズムを使用した各サンプルにおいて、sgRNA リード数を生成し、10 日目対 3 日目で、または 10 日目対 6 日目でより豊富であるが、6 日目対 3 日目ではそうではない sgRNA を見出すことであった (倍率変化 (fc) 1.5 および陰性二項検定 $p < 0.01$)。Fc 1.5 は、(10 日目カウントの平均) / (3 日目または 6 日目カウントの平均) の比率 1.5 を意味する。P < 0.01 は、10 日目カウントと 3 日目カウントとの間の統計的な差異がない、または 6 日目カウント < 0.01 である可能性を意味する。DESeq アルゴリズムは、「配列カウントデータの差次的発現の解析」のために広く使用されているアルゴリズムである。例えば、参照により本明細書に組み込まれる Anders 他、(2010)、Genome Biology、11:R106 を参照されたい。

【0100】

具体的には、二つの比較が各ライブラリーで使用され、有意な sgRNA が識別された：10 日目対 3 日目、および 10 日目対 6 日目 これら四つの比較のそれぞれについて、DESeq アルゴリズムを使用し、有意であるとみなされるカットオフ閾値は、陰性二項検定 $p < 0.01$ に加えて、倍率変化 1.5 であった。各ライブラリーに対するこれらの比較のそれぞれにおいて有意なガイドが識別されると、遺伝子は、以下の二つの基準のうちの一つを満たす場合、有意であるとみなされた：(1) その遺伝子に対応する少なくとも二つの sgRNA は、一つの比較において有意であるとみなされた (10 日目対 3 日目、または 10 日目対 6 日目のいずれか)。および (2) 両方の比較において少なくとも一つの sgRNA が有意であった (10 日目対 3 日目、および 10 日目対 6 日目)。このアルゴリズムを使用して、五つの遺伝子を、第一のライブラリーから有意であると識別し、四つの遺伝子を、第二のライブラリーから識別した。表 1 を参照されたい。

【 0 1 0 1 】

【 表 1 】

表 1 戦略# 1 を使用して識別された遺伝子。

ライブラリー#1					
10 日目対 3 日目		10 日目対 6 日目		6 日目対 3 日目	
遺伝子	有意な gRNA	遺伝子	有意な gRNA	遺伝子	有意な gRNA
標的遺伝子 1	1	標的遺伝子 1	1	標的遺伝子 1	0
標的遺伝子 2	3	標的遺伝子 2	1	標的遺伝子 2	0
標的遺伝子 15	1	標的遺伝子 15	1	標的遺伝子 15	0
標的遺伝子 16	1	標的遺伝子 16	1	標的遺伝子 16	0
標的遺伝子 17	2	標的遺伝子 17	0	標的遺伝子 17	0
ライブラリー#2					
10 日目対 3 日目		10 日目対 6 日目		6 日目対 3 日目	
遺伝子	有意な gRNA	遺伝子	有意な gRNA	遺伝子	有意な gRNA
標的遺伝子 2	1	標的遺伝子 2	1	標的遺伝子 2	0
標的遺伝子 18	1	標的遺伝子 18	1	標的遺伝子 18	0
標的遺伝子 19	1	標的遺伝子 19	1	標的遺伝子 19	0
標的遺伝子 20	1	標的遺伝子 20	1	標的遺伝子 20	0

10

20

【 0 1 0 2 】

しかしながら、この第一の戦略は、各実験群内のリード数の均質性の一定のレベルが極めて厳しいことを必要とする。同じ sgRNA について、スクリーニングライブラリー中の初期ウイルス数、感染または遺伝子編集効率、および遺伝子編集後の相対成長率など、多くの因子が各実験群（3 日目、6 日目、または 10 日目サンプル）内のサンプル間でリード数の変動性を生じ得る。したがって、第二の戦略も、正確なリード数の代わりに、10 日目（選択後）の各サンプルにおける遺伝子当たりのガイドの陽性発生（リード数 > 30）に基づいて使用した。

【 0 1 0 3 】

選択前 CRISPR 実験を四回繰り返した。図 6 に示すように、選択前実験「Exp 1」の遺伝子「G 1」については、sgRNA「g 1」、「g 2」、および「g 3」は、それぞれ 121、1000、および 302 のリード数で細胞集団中に存在した。遺伝子「G 2」については、sgRNA「g 4」、「g 5」、「g 6」、および「g 7」は、それぞれ 443、2012、534、および 150 のリード数で細胞集団中に存在した。かかるリード数データは、遺伝子「G 1」から遺伝子「G 21, 000」について生成される。各遺伝子について、「存在の合計」または n を決定した。各 sgRNA の個々のリード数を背景閾値と比較することによって、DNA の標的領域当たり存在する sgRNA のそれぞれの数である「存在の合計」を決定した。この場合、30 を背景閾値として使用した。したがって、リード数によって示される存在する sgRNA の量と対比して、「存在の合計」は質的に存在している sgRNA の数を示す。sgRNA g 1、g 2、および g 3 のリード数が各々 30 の背景閾値を超えるため、遺伝子 G 1 に対応する sgRNA の存在の合計は 3 である。sgRNA g 4、g 5、g 6、および g 7 のリード数が各々 30 の背景閾値を超えるため、遺伝子 G 2 に対応する sgRNA の存在の合計は 4 である。リード数が背景閾値を超える細胞集団中に存在する、全標的領域にわたる、sgRNA の総数または N は、59,010 として示される。したがって、細胞集団に感染するために使用された約 64,000 の異なる sgRNA のオリジナルライブラリーのうち、背景閾値を超えた量で存在したのは約 59,000 の異なる sgRNA のみであった。

30

40

【 0 1 0 4 】

表現型選択を伴う CRISPR 実験を四回繰り返した。しかしながら、配列決定の前に、

50

細胞集団中の細胞を、蛍光技術（例えば、FRET 蛍光）を使用して表現型に従ってソートした。Cas9 / CRISPR が、細胞の標的領域（例えば、遺伝子）を切断し、細胞が蛍光を発しなかった場合、遺伝子は首尾よくロックアウトされた。細胞が蛍光を発する場合、遺伝子はロックアウトされなかった。次いで蛍光細胞を配列決定することができ、非蛍光細胞は配列決定されなかった。選択された細胞は、特定の表現型 / マーカーを示す細胞を表す。

【0105】

図6に示すように、選択後実験「Exp1」の遺伝子「G1」については、gRNA「g1」、「g2」、および「g3」は、それぞれ0、8、および12のリード数で細胞集団中に存在した（または存在しなかった）。遺伝子「G2」については、gRNA「g4」、「g5」、「g6」、および「g7」は、それぞれ4、25、4、および150のリード数で細胞集団中に存在した（または存在しなかった）。かかるリード数データは、遺伝子「G1」から遺伝子「G21,000」について生成される。各遺伝子について、「存在の合計」または n' を決定した。各 gRNA の個々のリード数を背景閾値と比較することによって、DNA の標的領域当たり存在する gRNA のそれぞれの数である「存在の合計」を決定した。この場合、30 を背景閾値として使用した。したがって、リード数によって示される存在する gRNA の量と対比して、「存在の合計」は存在する gRNA の数を示す。gRNA g1、g2、および g3 のリード数が各々30の背景閾値を超えないため、遺伝子 G1 に対応する gRNA の存在の合計は0である。gRNA g7 のリード数のみが30の背景閾値を超えるため、遺伝子 G2 に対応する gRNA の存在の合計は1である。リード数が背景閾値を超えた細胞集団中に存在する、全標的領域にわたる、gRNA の総数または N' は、4,320 として示される。したがって、細胞集団に感染するために使用された約 64,000 の異なる gRNA のオリジナルライブラリーのうち、背景閾値を超える量で存在したのは約 4,320 の異なる gRNA のみであった。

10

20

【0106】

正式な統計 p 値を、ライブラリーサイズ、遺伝子当たりのガイドの数、および選択後サンプル中のポジティブガイドの総数を考慮して、選択後サンプル中のガイドの数を陽的に観察するために計算した。リード数が考慮されると、選択された細胞（選択後）内の標的領域（例えば、遺伝子）に対する n' ガイドを偶然に観察する確率を、式

【0107】

【数18】

$$\frac{\binom{n'}{n} \cdot \binom{N'-n'}{N-n}}{\binom{N'}{N}}$$

30

【0108】

に従って決定した。説明として、

【0109】

【数19】

$$\binom{x}{y} = \frac{y!}{x!(y-x)!}$$

40

【0110】

これは y オブジェクトから x オブジェクトを選択する方法の数を決定する。

選択された細胞（選択後）内の標的領域（例えば、遺伝子）について n' ガイドを偶然観察する確率を決定した後、選択された細胞（選択後）内の遺伝子の n' 以上の gRNA を偶然に観察する確率を、式

【0111】

【数20】

50

$$\sum_{i=n'}^n \frac{\binom{i}{n} \cdot \binom{N'-i}{N-n}}{\binom{N'}{N}}$$

【 0 1 1 2 】

に従って決定した。

選択された細胞（選択後）内の標的領域の n' 以上の $gRNA$ を偶然に観察する確率が決定されると、平均濃縮 $gRNA$ を標的領域レベルで決定した。選択前と比較した、選択後遺伝子のリード数の全体的な濃縮を、陽性遺伝子を識別するために追加パラメータとして使用した。平均濃縮は、濃縮スコアによって表された。濃縮スコアは、 N/N' を評価することによって決定された。図 6 から引用すると、濃縮スコアは $59010/4320$ 、または 13.66 である。

10

【 0 1 1 3 】

選択された細胞（選択後）内の標的領域の n' 以上の $gRNA$ を偶然に観察する確率は、標的領域が陽的に選択されるかどうかを評価するために使用され得る。濃縮スコアは、標的領域が陽的に選択されるかどうかを評価するためにさらに使用され得る。選択された細胞（選択後）内の標的領域の n' 以上の $gRNA$ を偶然に観察する確率より有意に低い、選択された細胞における標的領域の n' 以上の $gRNA$ を観察する確率を有する標的領域は、陽的に選択された標的領域として識別してもよい。さらに、閾値を超える濃縮スコアを有する標的領域は、陽的に選択された標的領域を示し得る。

20

【 0 1 1 4 】

したがって、この第二の戦略は、CRISPR 陽性選択のための新しい、より感度の高い解析方法を表す。CRISPR 陽性選択の目的は、DNA 配列決定を使用して、 $sgRNA$ による摂動が表現型と相関する遺伝子を識別することである。ノイズ背景を低減するために、同じ遺伝子について複数の $sgRNA$ を、実験の複製と一緒に、これらの実験で通常使用する。しかしながら、現在、同一遺伝子に対する $sgRNA$ 間、ならびに技術的反復の間にある程度の均質性/一致を必要とする、一般的に使用されている統計解析法は、うまく機能しない。これは、これらの方法は、多くの可能性のある理由（例えば、異なる感染または遺伝子編集効率、スクリーニングライブラリー内の初期ウイルス数、および同じ表現型を有する他の $sgRNA$ の存在）のために、 $sgRNA$ 間の大きな変動および同じ遺伝子に対するリピートを処理できないからである。対照的に、本実施例に示される方法は、大きな変動に対して堅牢である。これは、各 $sgRNA$ の正確なリード数ではなく、個々の実験における遺伝子当たりのガイドの陽性発生に基づく。正式な統計 p 値は、ライブラリーサイズ、遺伝子当たりの $sgRNA$ 数、および各実験における陽性 $sgRNA$ の総数を考慮して、実験リピートにわたる $sgRNA$ 数を陽的に観察するために計算される。表現型選択の前後の相対的 $sgRNA$ 配列リード濃縮もパラメータとして使用される。この方法は、DESeq、MAGCK、およびその他のものを含む、広く使用されている最新の方法よりも良好に機能する。具体的には、この方法には、以下の工程が含まれる。

30

40

【 0 1 1 5 】

(1) 各実験について、陽性表現型を有する細胞内の任意の存在するガイドを識別すること。

(2) 遺伝子レベルで、各実験においてガイドが存在するランダムな可能性（別名 p 値）を計算すること。複数の実験にわたって存在する全体的な可能性は、フィッシャーの複合確率検定 (Fisher's combined probability test) (参照: Fisher, R. A. Fisher, R. A. (1948). 「質問と回答 # 14」。The American Statistician) によって計算される。すなわち、まず、複数の実験からの p 値を使用して検定統計量を計算する：

【 0 1 1 6 】

50

【数 2 1】

$$\phi = -2 \sum_{k=1}^K p_k,$$

【0 1 1 7】

式中、 p_k は、 k 番目の実験について計算された p 値であり、 K は、実験の総数である。次いで、 K 実験にわたる p 値の組み合わせは、 $2 * K$ の自由度で、カイ二乗分布下での値を観察する確率と等しい。

【0 1 1 8】

(3) 遺伝子レベルでガイドの平均濃縮を計算すること：濃縮スコア = 選択後の相対的存在量 / 選択前の相対的存在量。相対存在量 = 遺伝子についてのガイドのリード数 / 全ガイドのリード数。 10

【0 1 1 9】

(4) 存在しているランダムな可能性よりも有意に低く、ならびに特定の濃縮スコアを上回る遺伝子を選択すること。

C. 実施例 3

CRISPR / Cas9 活性化および不活性化変異誘発 (inactivating mutagenesis) を使用して、機能的ノックアウトのためのコードエクソンを標的化する CRISPR / Cas9 gRNA ライブラリー (hGeCKO-A および hGeCKO-B) を使用する、タウおよび τ -シヌクレイン原線維化および増殖の遺伝子修飾因子をスクリーニングした。図 7 は、サンプル識別子、実験番号、配列決定までの時間を示し、どのライブラリー (Gecko A または Gecko B) が実験に使用されたかを識別する。 20

【0 1 2 0】

Gecko A ライブラリーは、約 63,950 の gRNA から構成された。63,950 の gRNA のうち、56,116 の gRNA が 18,874 の遺伝子を標的とし、遺伝子の大多数が 3 つの gRNA によって標的とされた。63,950 の gRNA のうち、6,834 の gRNA が 1,795 のマイクロ RNA (microRNA) を標的とし、マイクロ RNA (microRNA) の大多数が 4 つの gRNA によって標的とされた。

【0 1 2 1】

Gecko B ライブラリーは、約 56,869 の gRNA から構成された。56,869 の gRNA のうち、55,869 の gRNA が 18,834 の遺伝子を標的とし、遺伝子の大多数が 3 つの gRNA によって標的とされた。56,869 個の gRNA のうち、マイクロ RNA を標的とした gRNA はなかった。1,000 個の gRNA のうち、標的は無かった。 30

【0 1 2 2】

3 日目および 6 日目には、表現型は示されなかった。10 日目、サンプルは表現型陽性であった。

図 8 は、Gecko A ライブラリーに感染したサンプルからの DNA リード数を示す。各バーは、ウイルスに感染したサンプルを表す。各サンプルは、3 日目 (d03)、6 日目 (d06)、または 10 日目 (d10) で配列決定され、各サンプルは、全ゲノムではなく gRNA の配列決定リードアウトを表す。 40

【0 1 2 3】

図 9 は、Gecko A ライブラリーに感染したサンプルから中央値までのリード数の正規化を示す。正規化は、gRNA リード数を各サンプル中のリード数の合計で割ることによって行われ、全サンプルにわたるリード数の合計の中央値で割って掛けた。下の棒グラフは定性的で、30 の閾値を超えるリード数を示し、10 日目は選択後である。同様のリード数の合計を考慮すると、3 日目および 6 日目のサンプルは、多種多様な gRNA を有し、10 日目のサンプルは、はるかに少ない種類の gRNA を有した。

【0 1 2 4】

図 10 は、ライブラリーサイズ、遺伝子当たりの gRNA 数、および各実験における陽性 gRNA の総数を考慮して、実験レポートにわたる gRNA 数を陽的に観察するために計算された正式な統計 p 値を示す。p 値は、Gecko A ライブラリーに感染し、10 日目に配列決定された五つのサンプルについて示される。

【0125】

実施形態

実施形態 1。

(A) cas9 陽性細胞の第一の培養物をウイルスベクターのライブラリーで感染させることであって、ライブラリーが、細胞のゲノム内の DNA の標的領域を切断するための少なくとも 3 つのガイド RNA (gRNA) を含む、感染させること、細胞を配列決定して gRNA の各々のリード数を取得すること、

DNA の標的領域当たりで、そのリード数が背景閾値を超える、gRNA のそれぞれの数を合計すること () であって、式中 $n = n$ である、合計すること、および全ての標的領域にわたって、そのリード数が背景閾値を超える、gRNA の総数を合計すること () であって、式中 $N = N$ である、合計すること；

(B) ウイルスベクターのライブラリーで、cas9 陽性細胞の第二の培養物を感染させること、

第二の培養物の細胞を、指定された表現型を有する、または指定された表現型を有しないのいずれかとしてカテゴライズすること、指定された表現型を有する細胞を選択し、選択された細胞を配列決定して gRNA の各々の選択後リード数を取得すること、

その選択後リード数が背景閾値を超える選択された細胞において、DNA の標的領域当たりで、gRNA のそれぞれの数を合計すること () であって、式中 $n' = n'$ である、合計すること、および、

全ての標的領域にわたって、そのリード数が閾値を超える選択された細胞の、gRNA の総数を合計すること () であって、式中 $N' = N'$ である、合計すること；(C) DNA の標的領域について、式

【0126】

【数 2 2】

$$\frac{\binom{n'}{n} \cdot \binom{N'-n'}{N-n}}{\binom{N'}{N}}$$

30

【0127】

に従って、選択された細胞内の標的領域について n' gRNA を偶然に観察する確率を計算することであって、式中

【0128】

【数 2 3】

$$\binom{x}{y} = \frac{y!}{x! (y-x)!}$$

40

【0129】

は y オブジェクトから x オブジェクトを選択する方法の数を算出する、計算すること、および遺伝子を含む DNA の標的領域について、式

【0130】

【数 2 4】

50

$$\sum_{i=n'}^n \frac{\binom{i}{n} \cdot \binom{N'-i}{N-n}}{\binom{N'}{N}}$$

【0131】

に従って、選択された細胞内の遺伝子の n' 以上の gRNA を偶然に観察する確率を計算すること、を含む、方法。

実施形態 2

(A) cas9 陽性細胞の第一の培養物をウイルスベクターのライブラリーで感染させることであって、ライブラリーが、細胞のゲノム内の DNA の標的領域の転写を強化するための少なくとも 3 つのガイド RNA (gRNA) を含む、感染させること、

細胞を配列決定して、gRNA の各々のリード数を取得すること、DNA の標的領域当たりで、そのリード数が背景閾値を超える、gRNA のそれぞれの数を合計 () することであって、式中 $n = n'$ である、合計すること、および、すべての標的領域にわたって、そのリード数が背景閾値を超える、gRNA の総数を合計すること () であって、式中 $N = N'$ である、合計すること；

(B) ウイルスベクターのライブラリーで、cas9 陽性細胞の第二の培養物を感染させること、

第二の培養物の細胞を、指定された表現型を有する、または指定された表現型を有しないのいずれかとしてカテゴリ化すること、

指定された表現型を有する細胞を選択し、選択された細胞を配列決定して gRNA の各々の選択後リード数を取得すること、

DNA の標的領域当たりで、その選択後リード数が背景閾値を超える選択された細胞において、gRNA のそれぞれの数を合計すること () であって、式中 $n = n'$ である、合計すること、および、

すべての標的領域にわたって、そのリード数が閾値を超える選択された細胞の、gRNA の総数を合計すること () であって、式中 $N = N'$ である、合計すること；

(C) DNA の標的領域について、式

【0132】

【数 25】

$$\frac{\binom{n'}{n} \cdot \binom{N'-n'}{N-n}}{\binom{N'}{N}}$$

【0133】

に従って、選択された細胞内の標的領域について n' gRNA を偶然に観察する確率を計算することであって、式中

【0134】

【数 26】

$$\binom{x}{y} = \frac{y!}{x!(y-x)!}$$

【0135】

は y オブジェクトから x オブジェクトを選択する方法の数を算出する、計算すること、および、

遺伝子を含む DNA の標的領域について、式

【0136】

10

20

30

40

50

【数 27】

$$\sum_{i=n'}^n \frac{\binom{i}{n} \cdot \binom{N'-i}{N-n}}{\binom{N'}{N}}$$

【0137】

に従って、選択された細胞内の遺伝子の n' 以上の gRNA を偶然に観察する確率を計算すること、を含む、方法。

実施形態 3 第一の培養物を、CRISPR 技術に従って感染させる、実施形態 1 ~ 2 のいずれかに記載の方法。 10

【0138】

実施形態 4 第一の培養物の cas-9 陽性細胞が、一つ以上の選択マーカ-を含有するように改変される、実施形態 1 ~ 3 のいずれかに記載の方法。

実施形態 5 指定された表現型が蛍光である、実施形態 1 ~ 4 のいずれかに記載の方法。

【0139】

実施形態 6 指定された表現型が細胞生存である、実施形態 1 ~ 5 のいずれかに記載の方法。

実施形態 7 一つ以上の選択マーカ-が、蛍光マーカ-を含む、実施形態 4 に記載の方法。

【0140】

実施形態 8 蛍光マーカ-が、FRET バイオセンサーの一部である、実施形態 7 に記載の方法。 20

実施形態 9 一つ以上の選択マーカ-が、検出可能な酵素である、実施形態 4 に記載の方法。

【0141】

実施形態 10 検出可能な酵素が、 β -ガラクトシダーゼである、実施形態 9 に記載の方法。

実施形態 11 検出可能な酵素が、ルシフェラーゼである、実施形態 9 に記載の方法。

【0142】

実施形態 12 標的領域が、遺伝子を含む、実施形態 1 ~ 11 のいずれかに記載の方法。 30

実施形態 13 第二の培養物の細胞を、指定された表現型を有する、または指定された表現型を有しないのいずれかとしてカテゴライズすることが、選択機構を第二の細胞培養物に適用することを含む、実施形態 1 ~ 12 のいずれかに記載の方法。

【0143】

実施形態 14 選択機構は、第二の細胞集団を薬剤に曝露することか、または第二の細胞集団をタンパク質活性または発現レベルを特定する物質に曝露することのうちの一つ以上を含む、実施形態 13 に記載の方法。

【0144】

実施形態 15 指定された表現型を有する細胞を選択することが、一つ以上の選択マーカ-に基づいて細胞をソートすることを含む、実施形態 4 に記載の方法。 40

実施形態 16 選択された細胞内の遺伝子の n' 以上の gRNA を偶然に観察する確率に基づいて、標的領域が陽的に選択されると識別することをさらに含む、実施形態 1 ~ 15 のいずれかに記載の方法。

【0145】

実施形態 17 標的領域が陽的に選択されると識別することが、選択された細胞における対象の配列の n' 以上の gRNA を偶然に観察する確率が閾値を満たすと決定することを含む、実施形態 16 に記載の方法。

【0146】

実施形態 18 各 gRNA について、濃縮スコアを決定することをさらに含む、実施形態 1 ~ 17 のいずれかに記載の方法。 50

実施形態 19 各 gRNA について、濃縮スコアを決定することが、N/N' を評価することを含む、実施形態 18 に記載の方法。

【0147】

実施形態 20 cas-9 陽性細胞が、不活性化 cas-9 を含む、実施形態 2 ~ 19 のいずれかに記載の方法。

実施形態 21 不活性化 cas-9 が、少なくとも一つの転写活性化ドメインに融合される、実施形態 20 に記載の方法。

【0148】

実施形態 22 DNA の標的領域が、下流遺伝子またはタンパク質を調節する、実施形態 1 ~ 21 のいずれかに記載の方法。

10

実施形態 23 下流遺伝子またはタンパク質を調節することが、下流遺伝子またはタンパク質の活性化または阻害を含む、実施形態 22 に記載の方法。

【0149】

実施形態 24 選択された細胞内の遺伝子の n' 以上の gRNA を偶然に観察する確率に基づいて、標的領域を第二の遺伝子の修飾因子として識別することをさらに含む、実施形態 1 ~ 23 のいずれかに記載の方法。

【0150】

実施形態 25 選択された細胞内の遺伝子の n' 以上の gRNA を偶然に観察する確率に基づいて、標的領域を治療標的として識別することをさらに含む、実施形態 1 ~ 24 のいずれかに記載の方法。

20

【0151】

実施形態 26 選択された細胞内の遺伝子の n' 以上の gRNA を偶然に観察する確率に基づいて、標的領域が指定された表現型と相関すると識別することをさらに含む、実施形態 1 ~ 25 のいずれかに記載の方法。

【0152】

実施形態 27 選択された細胞内の遺伝子の n' 以上の gRNA を偶然に観察する確率に基づいて、標的領域が保護効果を示すと識別することをさらに含む、実施形態 1 ~ 26 のいずれかに記載の方法。

【0153】

実施形態 28 方法であって、DNA の複数の標的領域の各々に対する少なくとも 3 つのガイド RNA (gRNA) のライブラリーを含むベクターを用いた感染後の第一の細胞集団の配列決定に基づいて、そのリード数が背景閾値を超える、DNA の複数の標的領域の各々について存在する gRNA のそれぞれの数 (n) を決定すること；DNA の複数の標的領域の各々に対する gRNA のそれぞれの数に基づいて、そのリード数が背景閾値を超える、DNA の複数の標的領域の全ての標的領域にわたる、第一の細胞集団に存在する gRNA の総数 (N) を決定すること；DNA の複数の標的領域の各々に対する少なくとも 3 つの gRNA のライブラリーを含むベクターを用いた感染後の第二の細胞集団の配列決定に基づいて、そのリード数が背景閾値を超える、DNA の複数の標的領域の各々について存在する gRNA のそれぞれの数 (n') を決定すること；

30

DNA の複数の標的領域の各々に対する gRNA のそれぞれの数に基づいて、そのリード数が背景閾値を超える DNA の複数の標的領域の全ての標的領域にわたる第二の細胞集団中に存在する gRNA の総数 (N') を決定すること；複数の標的領域の各標的領域について、n、N、n'、および N' に基づいて、選択された細胞内の標的領域について n' gRNA を偶然に観察する確率を決定すること；対象の配列を含む標的領域について、選択された細胞内の標的領域について n' gRNA を偶然に観察する確率に基づいて、選択された細胞における対象の配列の n' 以上の gRNA を偶然に観察する確率を決定すること；および、選択された細胞における対象の配列の n' 以上の gRNA を観察する確率に基づいて、対象の配列が陽的に選択されると識別すること、を含む、方法。

40

【0154】

実施形態 29 第一の細胞集団および第二の集団が、cas-9 陽性細胞を含む、実施形

50

態 28 に記載の方法。

実施形態 30 DNA の複数の標的領域の各々に対する少なくとも 3 つの gRNA のライブラリーを含むベクターを用いた感染後の第一の細胞集団の配列決定に基づいて、そのリード数が背景閾値を超える、DNA の複数の標的領域の各々について存在する gRNA のそれぞれの数 (n) を決定することが、ウイルスベクターのライブラリーで第一の細胞集団を感染させること、第一の細胞集団の細胞を配列決定して、gRNA の各々のリード数を取得すること、および DNA の複数の標的領域の各々に対して、そのリード数が背景閾値を超える各 gRNA を数えること、を含む、実施形態 29 に記載の方法。

【0155】

実施形態 31 DNA の複数の標的領域の各々に対する gRNA のそれぞれの数に基づいて、そのリード数が背景閾値を超える、DNA の複数の標的領域の全ての標的領域にわたる、第一の細胞集団に存在する gRNA の総数 (N) を決定することは、そのリード数が背景閾値を超える第一の細胞集団に存在する各 gRNA を数えることを含む、実施形態 30 に記載の方法。

10

【0156】

実施形態 32 DNA の複数の標的領域の各々に対する少なくとも 3 つの gRNA のライブラリーを含むベクターを用いた感染後の第二の細胞集団の配列決定に基づいて、そのリード数が背景閾値を超える、DNA の複数の標的領域の各々について存在する gRNA のそれぞれの数 (n') を決定することが、第二の細胞集団の細胞を、指定された表現型を有する、または指定された表現型を有しないのいずれかとしてカテゴライズすること、指定された表現型を有する細胞を選択すること、選択された細胞を配列決定して gRNA の各々のリード数を取得すること、および DNA の複数の標的領域の各々について、そのリード数が背景閾値を超える各 gRNA を数えること、を含む、実施形態 31 に記載の方法。

20

【0157】

実施形態 33 DNA の複数の標的領域の各々に対する gRNA のそれぞれの数に基づいて、そのリード数が背景閾値を超える DNA の複数の標的領域の全ての標的領域にわたる第二の細胞集団に存在する gRNA の総数 (N') を決定することが、そのリード数が背景閾値を超える第二の細胞集団の選択された細胞内に存在する各 gRNA を数えることを含む、実施形態 28 ~ 32 のいずれか一つに記載の方法。

30

【0158】

実施形態 34 複数の標的領域の各標的領域について、n、N、n'、および N' に基づいて、選択された細胞内の標的領域について n' gRNA を偶然に観察する確率を決定することが、

【0159】

【数 28】

$$\frac{\binom{n'}{n} \cdot \binom{N'-n'}{N-n}}{\binom{N'}{N}}$$

40

【0160】

を評価することを含み、式中

【0161】

【数 29】

$$\binom{x}{y} = \frac{y!}{x!(y-x)!}$$

【0162】

50

は、y オブジェクトから x オブジェクトを選択する方法の数を算出する、実施形態 28 ~ 33 のいずれか一つに記載の方法。

実施形態 35 対象の配列を含む標的領域について、選択された細胞内の標的領域について n' gRNA を偶然に観察する確率に基づいて、選択された細胞における対象の配列の n' 以上の gRNA を偶然に観察する確率を決定することは、

【0163】

【数30】

$$\sum_{i=n'}^n \frac{\binom{i}{n} \cdot \binom{N'-i}{N-n}}{\binom{N'}{N}}$$

10

【0164】

を評価することを含む、実施形態 28 ~ 34 のいずれか一つに記載の方法。

実施形態 36 選択された細胞における対象の配列の n' 以上の gRNA を偶然に観察できる確率に基づいて、対象の配列が陽的に選択されると識別することが、選択された細胞における対象の配列の n' 以上の gRNA を偶然に観察する確率が閾値を満たすと決定することを含む、実施形態 28 ~ 35 のいずれか一つに記載の方法。

【0165】

実施形態 37 各 gRNA について、濃縮スコアを決定することをさらに含む、実施形態 28 ~ 36 のいずれか一つに記載の方法。

20

実施形態 38 各 gRNA について、濃縮スコアを決定することが、 N/N' を評価することを含む、実施形態 37 に記載の方法。

【0166】

実施形態 39 cas-9 陽性細胞が、不活性化 cas-9 を含む、実施形態 28 ~ 38 のいずれか一つに記載の方法。

実施形態 40 不活性化 cas-9 が、少なくとも一つの転写活性化ドメインに融合される、実施形態 39 に記載の方法。

【0167】

実施形態 41 装置であって、一つ以上のプロセッサと、プロセッサ実行可能命令であって、一つ以上のプロセッサによって実行されると、装置に、(A) 第一のリード数データを受信させ、ここで第一のリード数データは、cas9 陽性細胞の第一の培養物を、ウイルスベクターのライブラリーに感染させることであって、ライブラリーが、細胞のゲノム内の DNA の標的領域を切断するための少なくとも 3 つのガイド RNA (gRNA) を含む、感染させること、および細胞を配列決定して gRNA の各々のリード数を取得することによって生成され；第一のリード数データに基づいて、DNA の標的領域当たりで、そのリード数が背景閾値を超える、gRNA のそれぞれの数を合計させ ()、式中 n であり、および第一のリード数データに基づいて、標的領域全体にわたり、そのリード数が背景閾値を超える、gRNA の総数を合計させ ()、式中 N であり；(B) 第二のリード数データを受信させ、ここで第二のリード数データは、cas9 陽性細胞の第二の培養物を、ウイルスベクターのライブラリーに感染させること、第二の培養物の細胞を、指定された表現型を有する、または指定された表現型を有しないのいずれかとしてカテゴライズすること、および指定された表現型を有する細胞を選択し、選択された細胞を配列決定して gRNA の各々の選択後リード数を取得することによって生成され；第二のリード数データに基づいて、DNA の標的領域当たりで、その選択後リード数が背景閾値を超える第二の細胞における、gRNA のそれぞれの数を合計させ ()、式中 n' であり、および第二のリード数データに基づいて、標的領域全体にわたり、そのリード数が閾値を超える選択された細胞の、gRNA の総数を合計させ ()、式中 N' であり；(C) DNA の標的領域について、式

30

40

【0168】

50

【数 3 1】

$$\frac{\binom{n'}{n} \cdot \binom{N' - n'}{N - n}}{\binom{N'}{N}}$$

【0 1 6 9】

に従って、選択された細胞内の標的領域について n' gRNA を偶然に観察する確率を計算させ、式中

【0 1 7 0】

10

【数 3 2】

$$\binom{x}{y} = \frac{y!}{x! (y-x)!}$$

【0 1 7 1】

は y オブジェクトから x オブジェクトを選択する方法の数を算出し、および遺伝子を含む DNA の標的領域について、式

【0 1 7 2】

20

【数 3 3】

$$\sum_{i=n'}^n \frac{\binom{i}{n} \cdot \binom{N' - i}{N - n}}{\binom{N'}{N}}$$

【0 1 7 3】

に従って、選択された細胞内の遺伝子の n' 以上の gRNA を偶然に観察する確率を計算させる、プロセッサ実行可能命令、を格納するメモリと、を備える装置。

実施形態 4 2 第一の培養物を、CRISPR 技術に従って感染させる、実施形態 4 1 に記載の装置。

30

【0 1 7 4】

実施形態 4 3 第一の培養物の cas-9 陽性細胞が、一つ以上の選択マーカ-を含有するように改変される、実施形態 4 1 ~ 4 2 のいずれか一つに記載の装置。

実施形態 4 4 標的領域が遺伝子を含む、実施形態 4 1 ~ 4 3 のいずれか一つに記載の装置。

【0 1 7 5】

実施形態 4 5 第二の培養物の細胞を、指定された表現型を有する、または指定された表現型を有しないのいずれかとしてカテゴリ-化することが、選択機構を第二の細胞培養物に適用することを含む、実施形態 4 1 ~ 4 4 のいずれか一つに記載の装置。

【0 1 7 6】

40

実施形態 4 6 選択機構が、第二の細胞集団を薬剤に曝露することか、または第二の細胞集団をタンパク質活性または発現レベルを特定する物質に曝露することのうちの一つ以上を含む、実施形態 4 5 に記載の装置。

【0 1 7 7】

実施形態 4 7 指定された表現型を有する細胞を選択することが、一つ以上の選択マーカ-に基づいて細胞をソ-トすることを含む、実施形態 4 1 ~ 4 6 のいずれか一つに記載の装置。

【0 1 7 8】

実施形態 4 8 選択された細胞内の遺伝子の n' 以上の gRNA を偶然に観察する確率に基づいて、標的領域が陽的に選択されると識別することをさらに含む、実施形態 4 1 ~

50

47のいずれか一つに記載の装置。

【0179】

実施形態49 標的領域が陽的に選択されると識別することが、選択された細胞における対象の配列の n' 以上の $gRNA$ を偶然に観察する確率が閾値を満たすと決定することを含む、実施形態48に記載の装置。

【0180】

実施形態50 各 $gRNA$ について、濃縮スコアを決定することをさらに含む、実施形態41～49のいずれか一つに記載の装置。

実施形態51 各 $gRNA$ について、濃縮スコアを決定することが、 N/N' を評価することを含む、実施形態50に記載の装置。

【0181】

実施形態52 $cas-9$ 陽性細胞が、不活性化 $cas-9$ を含む、実施形態41～51のいずれか一つに記載の装置。

実施形態53 不活性化 $cas-9$ が、少なくとも一つの転写活性化ドメインに融合される、実施形態52に記載の装置。

【0182】

実施形態54 一つ以上のガイド RNA ($gRNA$) を偶然に観察する確率を決定するための非一時的コンピュータ可読媒体であって、プロセッサ実行可能命令であって、一つ以上のプロセッサによって実行されると、一つ以上のプロセッサに、(A) 第一のリード数データを受信させ、ここで第一のリード数データは、 $cas-9$ 陽性細胞の第一の培養物を、ウイルスベクターのライブラリーに感染させることであって、ライブラリーが、細胞のゲノム内の DNA の標的領域を切断するための少なくとも3つのガイド RNA ($gRNA$) を含む、感染させること、および細胞を配列決定して $gRNA$ の各々のリード数を取得することによって生成され；第一のリード数データに基づいて、 DNA の標的領域当たりで、そのリード数が背景閾値を超える、 $gRNA$ のそれぞれの数を合計させ()、式中 $= n$ であり、および第一のリード数データに基づいて、標的領域全体にわたり、そのリード数が背景閾値を超える、 $gRNA$ の総数を合計させ()、式中 $= N$ であり；(B) 第二のリード数データを受信させ、ここで第二のリード数データは、 $cas-9$ 陽性細胞の第二の培養物を、ウイルスベクターのライブラリーに感染させること、第二の培養物の細胞を、指定された表現型を有する、または指定された表現型を有しないのいずれかとしてカテゴライズすること、および指定された表現型を有する細胞を選択し、選択された細胞を配列決定して $gRNA$ の各々の選択後リード数を取得することによって生成され；第二のリード数データに基づいて、 DNA の標的領域当たりで、その選択後リード数が背景閾値を超える第二の細胞における、 $gRNA$ のそれぞれの数を合計させ()、式中 $= n'$ であり、および第二のリード数データに基づいて、標的領域全体にわたり、そのリード数が閾値を超える選択された細胞の、 $gRNA$ の総数を合計させ()、式中 $= N'$ であり；(C) DNA の標的領域について、式

【0183】

【数34】

$$\frac{\binom{n'}{n} \cdot \binom{N'-n'}{N-n}}{\binom{N'}{N}}$$

【0184】

に従って、選択された細胞内の標的領域について n' $gRNA$ を偶然に観察する確率を計算させ、式中

【0185】

【数35】

10

20

30

40

50

$$\binom{x}{y} = \frac{y!}{x!(y-x)!}$$

【0186】

は y オブジェクトから x オブジェクトを選択する方法の数を算出し、および遺伝子を含む DNA の標的領域について、式 n' に従って、選択された細胞内の遺伝子の

【0187】

【数36】

$$\sum_{i=n'}^n \frac{\binom{i}{n} \cdot \binom{N'-i}{N-n}}{\binom{N'}{N}}$$

10

【0188】

以上の gRNA を偶然に観察する確率を計算させる、プロセッサ実行可能命令、を格納する、非一時的コンピュータ可読媒体。

実施形態 55 第一の培養物は、CRISPR 技術に従って感染される、実施形態 54 に記載の非一時的コンピュータ可読媒体。

【0189】

実施形態 56 第一の培養物の cas-9 陽性細胞が、一つ以上の選択マーカ含有するように改変される、実施形態 54 ~ 55 のいずれか一つに記載の非一時的コンピュータ可読媒体。

【0190】

実施形態 57 標的領域が遺伝子を含む、実施形態 54 ~ 56 のいずれか一つに記載の非一時的コンピュータ可読媒体。

実施形態 58 第二の培養物の細胞を、指定された表現型を有する、または指定された表現型を有しないのいずれかとしてカテゴライズすることが、選択機構を第二の細胞培養物に適用することを含む、実施形態 54 ~ 57 のいずれか一つに記載の非一時的コンピュータ可読媒体。

30

【0191】

実施形態 59 選択機構が、第二の細胞集団を薬剤に曝露することか、または第二の細胞集団をタンパク質活性または発現レベルを特定する物質に曝露することのうちの一つ以上を含む、実施形態 58 に記載の非一時的コンピュータ可読媒体。

【0192】

実施形態 60 指定された表現型を有する細胞を選択することが、一つ以上の選択マーカに基づいて細胞をソートすることを含む、実施形態 54 ~ 59 のいずれか一つに記載の非一時的コンピュータ可読媒体。

【0193】

実施形態 61 選択された細胞内の遺伝子の n' 以上の gRNA を偶然に観察する確率に基づいて、標的領域が陽的に選択されると識別することをさらに含む、実施形態 54 ~ 60 のいずれか一つに記載の非一時的コンピュータ可読媒体。

40

【0194】

実施形態 62 標的領域が陽的に選択されると識別することが、選択された細胞における対象の配列の n' 以上の gRNA を偶然に観察する確率が閾値を満たすと決定することを含む、実施形態 61 に記載の非一時的コンピュータ可読媒体。

【0195】

実施形態 63 各 gRNA について、濃縮スコアを決定することをさらに含む、実施形態 54 ~ 62 のいずれか一つに記載の非一時的コンピュータ可読媒体。

実施形態 64 各 gRNA について、濃縮スコアを決定することが、N / N' を評価する

50

ことを含む、実施形態 63 に記載の非一時的コンピュータ可読媒体。

【0196】

実施形態 65 cas-9 陽性細胞が、不活性化 cas-9 を含む、実施形態 54 ~ 64 のいずれか一つに記載の非一時的コンピュータ可読媒体。

実施形態 66 不活性化 cas-9 が、少なくとも一つの転写活性化ドメインに融合される、実施形態 65 に記載の非一時的コンピュータ可読媒体。

【0197】

実施形態 67 装置であって、一つ以上のプロセッサと、プロセッサ実行可能命令であって、一つ以上のプロセッサによって実行されると、装置に、DNA の複数の標的領域の各々に対する少なくとも 3 つのガイド RNA (gRNA) のライブラリーを含むベクターを用いた感染後の第一の細胞集団の配列決定に基づいて、そのリード数が背景閾値を超える、DNA の複数の標的領域の各々について存在する gRNA のそれぞれの数 (n) を決定させ、DNA の複数の標的領域の各々に対する gRNA のそれぞれの数に基づいて、そのリード数が背景閾値を超える、DNA の複数の標的領域の全ての標的領域にわたる、第一の細胞集団に存在する gRNA の総数 (N) を決定させ、DNA の複数の標的領域の各々に対する少なくとも 3 つの gRNA のライブラリーを含むベクターを用いた感染後の第二の細胞集団の配列決定に基づいて、そのリード数が背景閾値を超える、DNA の複数の標的領域の各々について存在する gRNA のそれぞれの数 (n') を決定させ、DNA の複数の標的領域の各々に対する gRNA のそれぞれの数に基づいて、そのリード数が背景閾値を超える DNA の複数の標的領域の全ての標的領域にわたる第二の細胞集団に存在する gRNA の総数 (N') を決定させ、複数の標的領域の各標的領域について、n、N、n'、および N' に基づき、選択された細胞内の標的領域について n' gRNA を偶然に観察する確率を決定させ、対象の配列を含む標的領域について、選択された細胞内の標的領域について n' gRNA を偶然に観察する確率に基づいて、選択された細胞における対象の配列の n' 以上の gRNA を偶然に観察する確率を決定させ、および選択された細胞における対象の配列の n' 以上の gRNA を偶然に観察する確率に基づいて、対象の配列が陽的に選択されることを識別させる、プロセッサ実行可能命令、を格納するメモリと、を備える装置。

【0198】

実施形態 68 第一の細胞集団および第二の集団が、cas-9 陽性細胞を含む、実施形態 67 に記載の装置。

実施形態 69 プロセッサ実行可能命令であって、一つ以上のプロセッサによって実行されると、装置に、DNA の複数の標的領域の各々に対する少なくとも 3 つの gRNA のライブラリーを含むベクターを用いた感染後の第一の細胞集団の配列決定に基づいて、そのリード数が背景閾値を超える、DNA の複数の標的領域の各々について存在する gRNA のそれぞれの数 (n) を決定させる、プロセッサ実行可能命令が、装置に、ウイルスベクターのライブラリーで第一の細胞集団を感染させることと、第一の細胞集団の細胞を配列決定して、gRNA の各々のリード数を取得することと、によって生成される第一のリード数データを受信させ、および第一のリード数データに基づいて、DNA の複数の標的領域の各々について、そのリード数が背景閾値を超える各 gRNA を数えさせる、実施形態 67 ~ 68 のいずれか一つに記載の装置。

【0199】

実施形態 70 プロセッサ実行可能命令であって、一つ以上のプロセッサによって実行されると、装置に、DNA の複数の標的領域の各々に対する gRNA のそれぞれの数に基づいて、そのリード数が背景閾値を超える、DNA の複数の標的領域の全ての標的領域にわたる、第一の細胞集団に存在する gRNA の総数 (N) を決定させる、プロセッサ実行可能命令が、装置に、そのリード数が背景閾値を超える第一の細胞集団に存在する各 gRNA を数えさせる、実施形態 69 に記載の装置。

【0200】

実施形態 71 プロセッサ実行可能命令であって、一つ以上のプロセッサによって実行さ

れると、装置に、DNAの複数の標的領域の各々に対する少なくとも3つのgRNAのライブラリーを含むベクターを用いた感染後の第一の細胞集団の配列決定に基づいて、そのリード数が背景閾値を超える、DNAの複数の標的領域の各々について存在するgRNAのそれぞれの数(n')を決定させる、プロセッサ実行可能命令が、装置に、第二の細胞集団の細胞を、指定された表現型を有する、または指定された表現型を有しないのいずれかとしてカテゴライズすることと、指定された表現型を有する細胞を選択することと、選択された細胞を配列決定してgRNAの各々のリード数を取得することと、によって生成される第二のリード数データを受信させ、および第二のリード数データに基づいて、DNAの複数の標的領域の各々について、そのリード数が背景閾値を超える各gRNAを数えさせる、実施形態67~70のいずれか一つに記載の装置。

10

【0201】

実施形態72 プロセッサ実行可能命令であって、一つ以上のプロセッサによって実行されると、装置に、DNAの複数の標的領域の各々に対するgRNAのそれぞれの数に基づいて、そのリード数が背景閾値を超えるDNAの複数の標的領域の全ての標的領域にわたる第二の細胞集団に存在するgRNAの総数(N')を決定させる、プロセッサ実行可能命令が、装置に、そのリード数が背景閾値を超える第二の細胞集団の選択された細胞に存在する各gRNAを数えさせる、実施形態67~71のいずれか一つに記載の装置。

【0202】

実施形態73 プロセッサ実行可能命令であって、一つ以上のプロセッサによって実行されると、装置に、複数の標的領域の各標的領域について、 n 、 N 、 n' 、および N' に基づいて、選択された細胞内の標的領域について n' gRNAを偶然に観察する確率を決定させる、プロセッサ実行可能命令が、装置に、

20

【0203】

【数37】

$$\frac{\binom{n'}{n} \cdot \binom{N'-n'}{N-n}}{\binom{N'}{N}}$$

【0204】

を評価させ、式中

30

【0205】

【数38】

$$\binom{x}{y} = \frac{y!}{x!(y-x)!}$$

【0206】

は、 y オブジェクトから x オブジェクトを選択する方法の数を算出する、実施形態67~72のいずれか一つに記載の装置。

40

実施形態74 プロセッサ実行可能命令であって、一つ以上のプロセッサによって実行されると、装置に、選択された細胞内の標的領域について n' gRNAを偶然に観察する確率に基づいて、対象の配列を含む標的領域について、選択された細胞における対象の配列の n' 以上のgRNAを偶然に観察する確率を決定させる、プロセッサ実行可能命令が、装置に

【0207】

【数39】

50

$$\sum_{i=n'}^n \frac{\binom{i}{n} \cdot \binom{N'-i}{N-n}}{\binom{N'}{N}}$$

【 0 2 0 8 】

を評価させる、実施形態 6 7 ~ 7 3 のいずれか一つに記載の装置。

実施形態 7 5 プロセッサ実行可能命令であって、一つ以上のプロセッサによって実行されると、装置に、選択された細胞における対象の配列の n ' 以上の g R N A を偶然に観察する確率に基づいて、対象の配列が陽性的に選択されると識別させる、プロセッサ実行可能命令が、装置に、選択された細胞における対象の配列の n ' 以上の g R N A を偶然に観察する確率が閾値を満たすと評価させる決定させる、実施形態 6 7 ~ 7 4 のいずれか一つに記載の装置。

10

【 0 2 0 9 】

実施形態 7 6 一つ以上のプロセッサによって実行されると、装置に、各 g R N A について、濃縮スコアを決定させる、プロセッサ実行可能命令をさらに含む、実施形態 6 7 ~ 7 6 のいずれか一つに記載の装置。

【 0 2 1 0 】

実施形態 7 7 プロセッサ実行可能命令であって、一つ以上のプロセッサによって実行されると、装置に、各 g R N A について、濃縮スコアを決定させる、プロセッサ実行可能命令が、装置に N / N ' を評価させる、実施形態 7 6 に記載の装置。

20

【 0 2 1 1 】

実施形態 7 8 c a s - 9 陽性細胞が、不活性化 c a s - 9 を含む、実施形態 6 7 ~ 7 7 のいずれか一つに記載の装置。

実施形態 7 9 不活性化 c a s - 9 が、少なくとも一つの転写活性化ドメインに融合される、実施形態 7 8 に記載の装置。

【 0 2 1 2 】

実施形態 8 0 一つ以上のガイド R N A (g R N A) を偶然に観察する確率を決定するための非一時的コンピュータ可読媒体であって、プロセッサ実行可能命令であって、一つ以上のプロセッサによって実行されると、一つ以上のプロセッサに、D N A の複数の標的領域の各々に対する少なくとも 3 つのガイド R N A (g R N A) のライブラリーを含むベクターを用いた感染後の第一の細胞集団の配列決定に基づいて、そのリード数が背景閾値を超える、D N A の複数の標的領域の各々について存在する g R N A のそれぞれの数 (n) を決定させ ; D N A の複数の標的領域の各々に対する g R N A のそれぞれの数に基づいて、そのリード数が背景閾値を超える、D N A の複数の標的領域の全ての標的領域にわたる、第一の細胞集団に存在する g R N A の総数 (N) を決定させ ; D N A の複数の標的領域の各々に対する少なくとも 3 つの g R N A のライブラリーを含むベクターを用いた感染後の第二の細胞集団の配列決定に基づいて、そのリード数が背景閾値を超える、D N A の複数の標的領域の各々について存在する g R N A のそれぞれの数 (n ') を決定させ ; D N A の複数の標的領域の各々に対する g R N A のそれぞれの数に基づいて、そのリード数が背景閾値を超える D N A の複数の標的領域の全ての標的領域にわたる第二の細胞集団に存在する g R N A の総数 (N ') を決定させ ; 複数の標的領域の各標的領域について、n、N、n'、および N' に基づき、選択された細胞内の標的領域について n' g R N A を偶然に観察する確率を決定させ ; 対象の配列を含む標的領域について、選択された細胞内の標的領域について n' g R N A を偶然に観察する確率に基づいて、選択された細胞における対象の配列の n' 以上の g R N A を偶然に観察する確率を決定させ、および、選択された細胞における対象の配列の n' 以上の g R N A を偶然に観察する確率に基づいて、対象の配列が陽性的に選択されると識別させる、プロセッサ実行可能命令、を格納する、非一時的コンピュータ可読媒体。

30

40

【 0 2 1 3 】

50

実施形態 8 1 第一の細胞集団および第二の集団が、c a s - 9 陽性細胞を含む、実施形態 8 0 に記載の非一時的コンピュータ可読媒体。

実施形態 8 2 プロセッサ実行可能命令であって、一つ以上のプロセッサによって実行されると、装置に、DNAの複数の標的領域の各々に対する少なくとも3つのgRNAのライブラリーを含むベクターを用いた感染後の第一の細胞集団の配列決定に基づいて、そのリード数が背景閾値を超える、DNAの複数の標的領域の各々について存在するgRNAのそれぞれの数(n)を決定させる、プロセッサ実行可能命令が、一つ以上のプロセッサに、ウイルスベクターのライブラリーで第一の細胞集団を感染させることと、第一の細胞集団の細胞を配列決定して、gRNAの各々のリード数を取得することと、によって生成される第一のリード数データを受信させ、および第一のリード数データに基づいて、DNAの複数の標的領域の各々について、そのリード数が背景閾値を超える各gRNAを数えさせる、実施形態 8 0 ~ 8 1 のいずれか一つに記載の非一時的コンピュータ可読媒体。

10

【0214】

実施形態 8 3 プロセッサ実行可能命令であって、一つ以上のプロセッサによって実行されると、装置に、DNAの複数の標的領域の各々に対するgRNAのそれぞれの数に基づいて、そのリード数が背景閾値を超える、DNAの複数の標的領域の全ての標的領域にわたる、第一の細胞集団に存在するgRNAの総数(N)を決定させる、プロセッサ実行可能命令が、一つ以上のプロセッサに、そのリード数が背景閾値を超える第一の細胞集団に存在する各gRNAを数えさせる、実施形態 8 0 ~ 8 2 のいずれか一つに記載の非一時的コンピュータ可読媒体。

20

【0215】

実施形態 8 4 プロセッサ実行可能命令であって、一つ以上のプロセッサによって実行されると、装置に、DNAの複数の標的領域の各々に対する少なくとも3つのgRNAのライブラリーを含むベクターを用いた感染後の第二の細胞集団の配列決定に基づいて、そのリード数が背景閾値を超える、DNAの複数の標的領域の各々について存在するgRNAのそれぞれの数(n')を決定させる、プロセッサ実行可能命令が、一つ以上のプロセッサに、第二の細胞集団の細胞を、指定された表現型を有する、または指定された表現型を有しないのいずれかとしてカテゴライズすることと、指定された表現型を有する細胞を選択することと、選択された細胞を配列決定してgRNAの各々のリード数を取得することと、によって生成される第二のリード数データを受信させ、および、第二のリード数データに基づいて、DNAの複数の標的領域の各々について、そのリード数が背景閾値を超える各gRNAを数えさせる、実施形態 8 0 ~ 8 3 のいずれか一つに記載の非一時的コンピュータ可読媒体。

30

【0216】

実施形態 8 5 プロセッサ実行可能命令であって、一つ以上のプロセッサによって実行されると、装置に、DNAの複数の標的領域の各々に対するgRNAのそれぞれの数に基づいて、そのリード数が背景閾値を超えるDNAの複数の標的領域の全ての標的領域にわたる第二の細胞集団に存在するgRNAの総数(N')を決定させる、プロセッサ実行可能命令が、一つ以上のプロセッサに、そのリード数が背景閾値を超える第二の細胞集団の選択された細胞に存在する各gRNAを数えさせる、実施形態 8 0 ~ 8 4 のいずれか一つに記載の非一時的コンピュータ可読媒体。

40

【0217】

実施形態 8 6 プロセッサ実行可能命令であって、一つ以上のプロセッサによって実行されると、装置に、複数の標的領域の各標的領域について、n、N、n'、およびN'に基づいて、選択された細胞内の標的領域についてn'gRNAを偶然に観察する確率を決定させる、プロセッサ実行可能命令が、一つ以上のプロセッサに

【0218】

【数40】

50

$$\frac{\binom{n'}{n} \cdot \binom{N'-n'}{N-n}}{\binom{N'}{N}}$$

【 0 2 1 9 】

を評価させ、式中

【 0 2 2 0 】

【 数 4 1 】

$$\binom{x}{y} = \frac{y!}{x! (y-x)!}$$

10

【 0 2 2 1 】

は、 y オブジェクトから x オブジェクトを選択する方法の数を算出する、実施形態 8 0 ~ 8 5 のいずれか一つに記載の非一時的コンピュータ可読媒体。

実施形態 8 7 プロセッサ実行可能命令であって、一つ以上のプロセッサによって実行されると、装置に、選択された細胞内の標的領域について n' gRNA を偶然に観察する確率に基づいて、対象の配列を含む標的領域について、選択された細胞における対象の配列の n' 以上の gRNA を偶然に観察する確率を決定させる、プロセッサ実行可能命令が、一つ以上のプロセッサに

20

【 0 2 2 2 】

【 数 4 2 】

$$\sum_{i=n'}^n \frac{\binom{i}{n} \cdot \binom{N'-i}{N-n}}{\binom{N'}{N}}$$

【 0 2 2 3 】

を評価させる、実施形態 8 0 ~ 8 6 のいずれか一つに記載の非一時的コンピュータ可読媒体。

30

実施形態 8 8 プロセッサ実行可能命令であって、一つ以上のプロセッサによって実行されると、装置に、選択された細胞における対象の配列の n 以上の gRNA を偶然に観察する確率に基づいて、対象の配列が陽的に選択されると識別させる、プロセッサ実行可能命令が、一つ以上のプロセッサに、選択された細胞における対象の配列の n' 以上の gRNA を偶然に観察する確率が閾値を満たすと評価させる決定させる、実施形態 8 0 ~ 8 7 のいずれか一つに記載の非一時的コンピュータ可読媒体。

【 0 2 2 4 】

実施形態 8 9 一つ以上のプロセッサによって実行されると、一つ以上のプロセッサに、各 gRNA について、濃縮スコアを決定させるプロセッサ実行可能命令を、さらに含む、実施形態 8 0 ~ 8 8 のいずれか一つに記載の非一時的コンピュータ可読媒体。

40

【 0 2 2 5 】

実施形態 9 0 一つ以上のプロセッサによって実行されると、装置に、各 gRNA について、濃縮スコアを決定させる、プロセッサ実行可能命令が、一つ以上のプロセッサに、 N/N' を評価させることを含む、実施形態 8 9 に記載の非一時的コンピュータ可読媒体。

【 0 2 2 6 】

実施形態 9 1 cas - 9 陽性細胞が、不活性化 cas - 9 を含む、実施形態 8 0 ~ 9 0 のいずれか一つに記載の非一時的コンピュータ可読媒体。

実施形態 9 2 不活性化 cas - 9 が、少なくとも一つの転写活性化ドメインに融合される、実施形態 9 1 に記載の非一時的コンピュータ可読媒体。

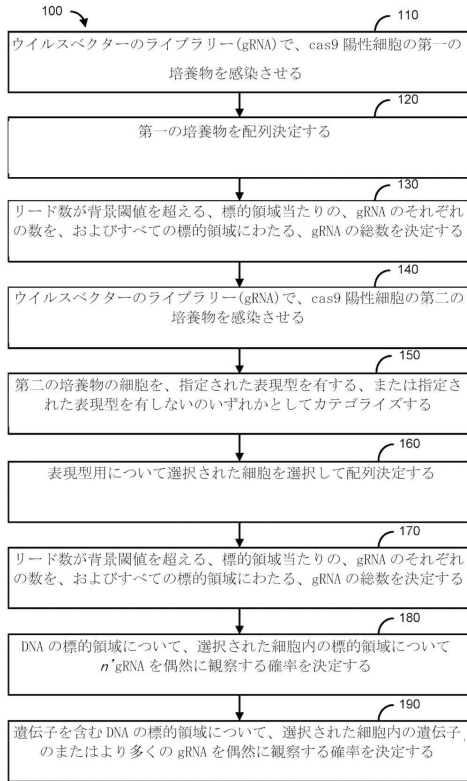
50

【 0 2 2 7 】

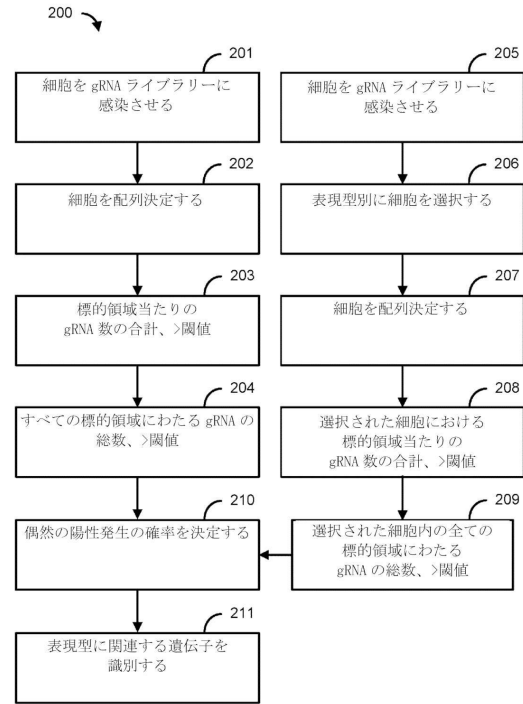
当業者は、通常の実験だけを用いることで、本明細書に記載の方法および組成物の特定の実施形態の多数の同等物を認識し、または確認できる。かかる同等物は、以下の特許請求の範囲に包含されることが意図される。

【 図 面 】

【 図 1 】



【 図 2 】

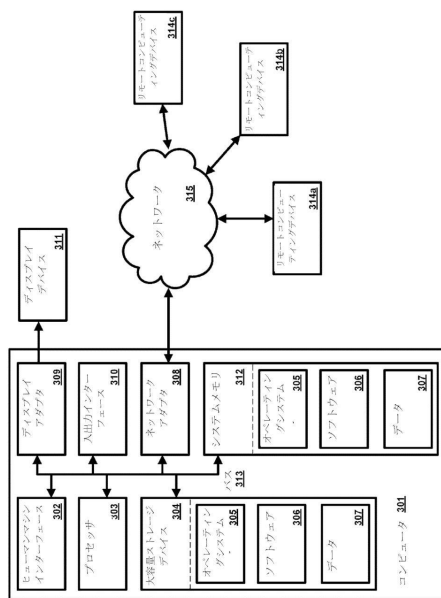


10

20

30

【 図 3 】



【 図 4 】

遺伝子#	遺伝子ごとの gRNA	gRNA リード数 (選択前)	n 閾値=30	gRNA リード数 (選択後)	n' 閾値=30
1	A B C	A = 100 B = 200 C = 10	2	A = 0 B = 50 C = 0	1
2	D E F	D = 500 E = 300 F = 200	3	D = 100 E = 100 F = 100	3
3	G H I J	G = 200 H = 250 I = 10 J = 300	3	G = 0 H = 0 I = 0 J = 20	0
4	K L M N	K = 100 L = 200 M = 200 N = 100	4	K = 0 L = 0 M = 0 N = 20	0
			N=12		N'=4

40

50

【 図 5 】

		N=63950		実験 1 N=4946		実験 2 N=13606	
		n'	n	確率		確率	
標的領域 1		3	中 3	0.000462378		0.009629	
		2	中 3	0.017017099		0.116534	
		1	中 3	0.214545542		0.512117	
		0	中 3	1		1	
標的領域 2		4	中 4	3.57411E-05		0.002048	
		3	中 4	0.001742291		0.032372	
		2	中 4	0.032291907		0.200696	
		1	中 4	0.275296754		0.615924	
		0	中 4	1		1	

510

【 図 6 】

遺伝子	アミノ酸	適応値					適応値						
		Exp1	Exp2	Exp3	Exp4	Exp5	Exp1	Exp2	Exp3	Exp4	Exp5		
G1	81	121	141	335	2012	0	0	0	0	0	0	0	5
G1	62	1000	342	345	534	3	8	0	0	0	0	0	6
G1	63	302	300	123	150	12	12	5	0	0	0	0	7
G2	64	373	68	56	42	5	12	3	0	0	0	0	12
G2	66	534	335	141	121	4	4	1	1	0	0	0	0
G3	67	150	342	345	1000	150	150	0	0	0	0	0	6
G3	68	150	123	300	302	15	222	2	432	3	29	2	1501
G3	69	432	456	598	443	3	432	2	432	3	29	2	453
G3	610	123	789	615	789	3	123	1	501	3	29	2	453
...
G63998	231	1050	907	899	2089	3	3	0	678	3	0	0	44
G71000	445	967	3	899	2089	3	3	0	325	3	0	0	36
G84900	885	122	1050	907	2089	3	3	0	17	3	0	0	44
G84900	886	122	1050	907	2089	3	3	0	17	3	0	0	44
合計	63950	20,000,000	55,000	20,000,000	60,123	20,000,000	60,123	20,000,000	59,950	20,000,000	6,320	20,000,000	11,024

10

20

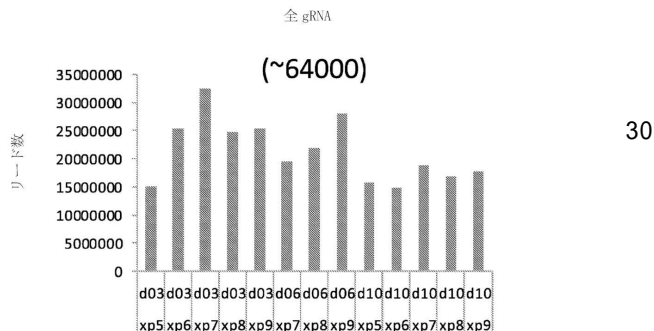
【 図 7 】

3 日目および6 日目：凝集体なし
10 日目：FRET+

ゲート A		ゲート B	
ゲート	遺伝子	ゲート	遺伝子
ゲート A	grNA	ゲート B	grNA
ゲート A	56,116	ゲート B	56,869
ゲート A	5,854	ゲート B	0
ゲート A	1,000	ゲート B	1,000
ゲート A	63,950	ゲート B	56,869

ゲート A		ゲート B	
ゲート	遺伝子	ゲート	遺伝子
ゲート A	標的	ゲート B	標的
ゲート A	18,871 遺伝子	ゲート B	18,834 遺伝子
ゲート A	1,795 遺伝子 RNA	ゲート B	遺伝子 RNA
ゲート A	なし	ゲート B	なし
ゲート A	63,950	ゲート B	1

【 図 8 】

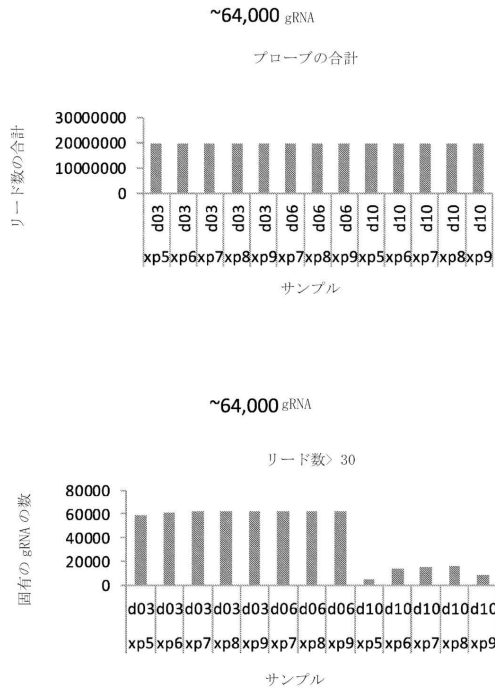


30

40

50

【 図 9 】



【 図 10 】

N=63950	実験 1 N=4946	実験 2 N=13606	実験 3 N=14742	実験 4 N=15468	実験 5 N=7767
n	n	確率	確率	確率	確率
3 中 3	0.000462378	0.009629	0.012248	0.014149	0.001791
2 中 3	0.017017099	0.116534	0.134919	0.147207	0.040666
1 中 3	0.214545542	0.512117	0.544405	0.564274	0.321905
0 中 3	1	1	1	1	1
4 中 4	3.57411E-05	0.002048	0.002823	0.003422	0.000217
3 中 4	0.001742291	0.032372	0.040524	0.04633	0.006512
2 中 4	0.032291907	0.200696	0.229313	0.248084	0.074821
1 中 4	0.275296754	0.615924	0.649435	0.669671	0.404267
0 中 4	1	1	1	1	1

10

20

30

40

50

【 国際調査報告 】

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2020/023152

A. CLASSIFICATION OF SUBJECT MATTER INV. G16B35/20 ADD. C12N15/09 According to International Patent Classification (IPC) or to both national classification and IPC												
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) G16B C40B C12N Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EPO-Internal												
C. DOCUMENTS CONSIDERED TO BE RELEVANT												
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.										
X	SWEE HOE ONG ET AL: "Optimised metrics for CRISPR-KO screens with second-generation gRNA libraries", SCIENTIFIC REPORTS, vol. 7, no. 1, 7 August 2017 (2017-08-07), pages 1-10, XP055644757, DOI: 10.1038/s41598-017-07827-z A whole document, in particular abstract, p. 1, 2, 4,6, 8, table 1, fig. 2 ----- -/--	1-18, 20-27, 41-50, 52-63, 65,66 19,51,64										
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.												
* Special categories of cited documents : <table border="0"> <tr> <td>*A* document defining the general state of the art which is not considered to be of particular relevance</td> <td>*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</td> </tr> <tr> <td>*E* earlier application or patent but published on or after the international filing date</td> <td>*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</td> </tr> <tr> <td>*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</td> <td>*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</td> </tr> <tr> <td>*O* document referring to an oral disclosure, use, exhibition or other means</td> <td>*Z* document member of the same patent family</td> </tr> <tr> <td>*P* document published prior to the international filing date but later than the priority date claimed</td> <td></td> </tr> </table>			*A* document defining the general state of the art which is not considered to be of particular relevance	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention	*E* earlier application or patent but published on or after the international filing date	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone	*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art	*O* document referring to an oral disclosure, use, exhibition or other means	*Z* document member of the same patent family	*P* document published prior to the international filing date but later than the priority date claimed	
A document defining the general state of the art which is not considered to be of particular relevance	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention											
E earlier application or patent but published on or after the international filing date	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone											
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art											
O document referring to an oral disclosure, use, exhibition or other means	*Z* document member of the same patent family											
P document published prior to the international filing date but later than the priority date claimed												
Date of the actual completion of the international search		Date of mailing of the international search report										
9 June 2020		18/06/2020										
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016		Authorized officer Lüdemann, Susanna										

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No PCT/US2020/023152

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WEI LI ET AL: "MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens", GENOME BIOLOGY, BIOMED CENTRAL LTD, vol. 15, no. 12, 5 December 2014 (2014-12-05), page 554, XP021208793, ISSN: 1465-6906, DOI: 10.1186/S13059-014-0554-4 the whole document	1-27, 41-66
A	----- ANOB M. CHAKRABARTI ET AL: "Target-Specific Precision of CRISPR-Mediated Genome Editing", MOLECULAR CELL, vol. 73, no. 4, 1 February 2019 (2019-02-01), pages 699-713.e6, XP055701322, AMSTERDAM, NL ISSN: 1097-2765, DOI: 10.1016/j.molcel.2018.11.031 the whole document	1-27, 41-66
A	----- Anonymous: "CRISPR 101: A Desktop Resource", CRISPR 101: A Desktop Resource, 1 May 2017 (2017-05-01), pages 1-195, XP055689797, Retrieved from the Internet: URL:https://cdn2.hubspot.net/hubfs/306096/ CRISPR 101 ebook/2nd%20Edition/CRISPR%2010 1%202nd%20Ed%20Final%20May%202018_2.pdf?hs CtaTracking=bf6f5328-0652-4ef5-ab78-dfa031 255da0]1a1481f7-5c09-4f4c-9ebb-94fd110ce62 7 [retrieved on 2020-04-28] the whole document	1-27, 41-66
A	----- LINDE A. MILES ET AL: "Design, execution, and analysis of pooled in?vitro CRISPR/Cas9 screens", FEBS JOURNAL, vol. 283, no. 17, 16 June 2016 (2016-06-16), pages 3170-3180, XP055428636, GB ISSN: 1742-464X, DOI: 10.1111/febs.13770 the whole document	1-27, 41-66

1

Form PCT/ISA/210 (continuation of second sheet) (April 2005)

page 2 of 2

50

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2020/023152

Box No. II Observations where certain claims were found unsearchable (Continuation of Item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

10

2. Claims Nos.: 28-40, 67-92
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210

3. Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

20

Box No. III Observations where unity of invention is lacking (Continuation of Item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

30

2. As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

40

Remark on Protest

The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box II.2

10

Claims Nos.: 28-40, 67-92

The present application contains 90 claims, of which 7 are independent. There is no clear distinction between the independent claims because of overlapping scope. There are so many claims, and they are drafted in such a way that the claims as a whole are not in compliance with the provisions of clarity and conciseness of Article 6 PCT, as it is particularly burdensome for a skilled person to establish the subject-matter for which protection is sought.

The non-compliance with the substantive provisions is to such an extent that a meaningful search of the whole claimed subject-matter could not be carried out (Article 17(2) PCT and PCT Guidelines 9.30).

20

The applicant's attention is drawn to the fact that claims relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure. If the application proceeds into the regional phase before the EPO, the applicant is reminded that a search may be carried out during examination before the EPO (see EPO Guidelines C-IV, 7.2), should the problems which led to the Article 17(2) declaration be overcome.

30

40

50

フロントページの続き

MK,MT,NL,NO,PL,PT,RO,RS,SE,SI,SK,SM,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,KM,ML,MR,N
E,SN,TD,TG),AE,AG,AL,AM,AO,AT,AU,AZ,BA,BB,BG,BH,BN,BR,BW,BY,BZ,CA,CH,CL,CN,CO,CR,CU,
CZ,DE,DJ,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IR,IS,JO,JP,KE,K
G,KH,KN,KP,KR,KW,KZ,LA,LC,LK,LR,LS,LU,LY,MA,MD,ME,MG,MK,MN,MW,MX,MY,MZ,NA,NG,N
I,NO,NZ,OM,PA,PE,PG,PH,PL,PT,QA,RO,RS,RU,RW,SA,SC,SD,SE,SG,SK,SL,ST,SV,SY,TH,TJ,TM,TN,
TR,TT,TZ,UA,UG,US,UZ,VC,VN,WS,ZA,ZM,ZW

弁理士 中村 美樹

(72)発明者 フューリー、ウェン

アメリカ合衆国 1 0 5 9 1 ニューヨーク州 タリータウン オールド ソー ミル リバー ロード
7 7 7 リジェネロン・ファーマシューティカルズ・インコーポレイテッド内

(72)発明者 パイ、ユー

アメリカ合衆国 1 0 5 9 1 ニューヨーク州 タリータウン オールド ソー ミル リバー ロード
7 7 7 リジェネロン・ファーマシューティカルズ・インコーポレイテッド内

F ターム (参考) 4B029 AA07 AA23 FA15

4B063 QA11 QQ42 QR10 QS34 QX02