

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2007-526776
(P2007-526776A)

(43) 公表日 平成19年9月20日(2007.9.20)

(51) Int. Cl.	F I	テーマコード (参考)
C12Q 1/68 (2006.01)	C12Q 1/68 Z N A A	2 G O 4 5
GO1N 33/15 (2006.01)	GO1N 33/15 Z	4 B O 2 4
GO1N 33/50 (2006.01)	GO1N 33/50 Z	4 B O 6 3
C12Q 1/02 (2006.01)	C12Q 1/02	
C12N 15/09 (2006.01)	C12N 15/00 A	

審査請求 未請求 予備審査請求 未請求 (全 77 頁)

(21) 出願番号 特願2007-502063 (P2007-502063)
 (86) (22) 出願日 平成17年3月3日 (2005.3.3)
 (85) 翻訳文提出日 平成18年11月6日 (2006.11.6)
 (86) 国際出願番号 PCT/US2005/007249
 (87) 国際公開番号 W02005/088306
 (87) 国際公開日 平成17年9月22日 (2005.9.22)
 (31) 優先権主張番号 60/550,074
 (32) 優先日 平成16年3月4日 (2004.3.4)
 (33) 優先権主張国 米国 (US)
 (31) 優先権主張番号 60/604,470
 (32) 優先日 平成16年8月25日 (2004.8.25)
 (33) 優先権主張国 米国 (US)

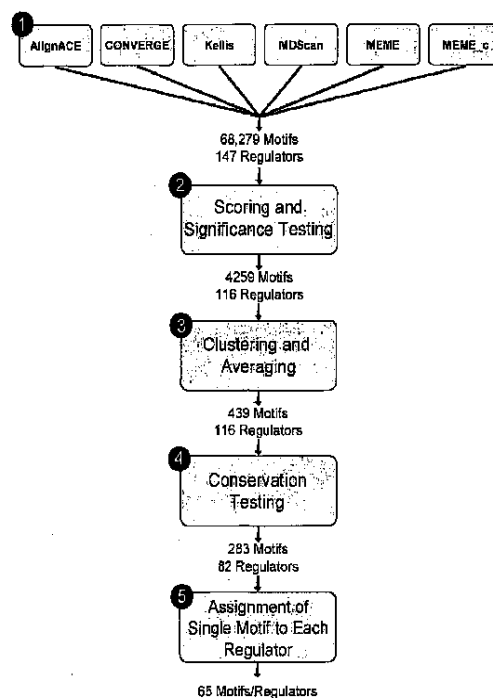
(71) 出願人 502168404
 ホワイトヘッド・インスティテュート・フ
 ォー・バイオメディカル・リサーチ
 Whitehead Institute
 for Biomedical Res
 earch
 アメリカ合衆国02142-1479マサ
 チューセッツ州ケンブリッジ、ナイン・ケ
 ンブリッジ・センター
 (74) 代理人 100099623
 弁理士 奥山 尚一
 (74) 代理人 100096769
 弁理士 有原 幸一
 (74) 代理人 100107319
 弁理士 松島 鉄男

最終頁に続く

(54) 【発明の名称】 生物活性DNA結合部位及び関連する方法

(57) 【要約】

本発明は、細胞中の目的のタンパク質が結合する生物活性DNA結合部位の同定に関する。本発明はまた、タンパク質が結合する生物活性DNA結合部位を変化させる薬剤及び条件の同定に関する。本発明の一態様はまた、転写調節因子によって調節される経路の同定方法及び経路の活性の調整方法を提供する。



【特許請求の範囲】**【請求項 1】**

細胞ゲノム中の目的のタンパク質の生物活性 DNA 結合部位セットを同定する方法であって、

(i) 該細胞中の該目的のタンパク質が結合するゲノム DNA 領域セットを同定するステップと、

(i i) 同定したゲノム DNA の該領域中の候補 DNA 結合部位を同定するステップと、ここで、候補 DNA 結合部位は、前記目的のタンパク質の DNA 配列モチーフに対応する配列を含み、

(i i i) 前記候補 DNA 結合部位が、前記細胞が得られる種と異なる 1 つ又は複数の種中の同等なゲノム領域中に保存されているかどうかを決定するステップと、ここで、少なくとも 1 つの前記異なる種の中に保存された候補 DNA 結合部位が生物活性 DNA 結合部位である、を含む、方法。 10

【請求項 2】

ステップ (i) は、ゲノム DNA 領域の前記セットから前記タンパク質の DNA 配列モチーフを同定することをさらに含む、請求項 1 に記載の方法。

【請求項 3】

前記 DNA 配列モチーフは、適切なコントロールと比較して統計的に有意な量のゲノム DNA 領域の前記セットによって富化される、請求項 2 に記載の方法。 20

【請求項 4】

前記適切なコントロールは、前記細胞中の前記目的のタンパク質によって結合されないゲノム領域のセットを含む、請求項 3 に記載の方法。

【請求項 5】

前記適切なコントロールは、無作為に選択したゲノム領域のセットを含む、請求項 3 に記載の方法。

【請求項 6】

前記適切なコントロールは、無作為に生成した配列のセットを含む、請求項 3 に記載の方法。

【請求項 7】

前記適切なコントロールは、前記細胞中の前記目的のタンパク質の変異形態によって結合されるゲノム領域のセットを含む、請求項 3 に記載の方法。 30

【請求項 8】

前記ゲノム DNA 領域は、プロモーター領域を含む、請求項 1 に記載の方法。

【請求項 9】

前記ゲノム DNA 領域は、約 50 b p ~ 約 10 k b の長さである、請求項 1 に記載の方法。

【請求項 10】

ステップ (i) は、前記目的のタンパク質のゲノム規模の位置分析 (G W L A) を行うことを含む、請求項 1 に記載の方法。 40

【請求項 11】

G W L A は、クロマチン免疫沈降 (C h I P) 及びその後の DNA マイクロアレイでの分析 (C h I P - c h i p) を含む、請求項 10 に記載の方法。

【請求項 12】

前記少なくとも 1 つの異なる種中の同等なゲノム領域が、前記目的のタンパク質の前記 DNA 配列モチーフに適合する核酸配列を含む場合、候補 DNA 結合部位は保存されている、請求項 2 に記載の方法。

【請求項 13】

少なくとも 1 つのアルゴリズムを使用して、前記 DNA 配列モチーフを同定する、請求項 2 に記載の方法。 50

【請求項14】

前記アルゴリズムは、AlignACE、MEME、MDscan、Kellis Method、Mogul、Verbunculus、YMF、BioProspector、Motif Sampler、及びSUPERPOSITIONから成る群から選択される、請求項13に記載の方法。

【請求項15】

前記DNA配列モチーフを、アルゴリズムの組み合わせを使用して同定する、請求項2に記載の方法。

【請求項16】

前記候補DNA結合部位は、20bp長未満である、請求項1に記載の方法。

10

【請求項17】

前記DNA配列モチーフは、少なくとも1つの位置で縮重される、請求項1に記載の方法。

【請求項18】

前記1つ又は複数の異なる種が、前記細胞と同一の属に分類される、請求項1に記載の方法。

【請求項19】

前記ステップ(iii)は、前記候補DNA結合部位が2つ又はそれ以上の異なる種中の同等なゲノム領域に保存されるかどうかを決定することを含む、請求項1に記載の方法。

20

【請求項20】

前記目的のタンパク質は、転写調節因子である、請求項1に記載の方法。

【請求項21】

前記目的のタンパク質は、DNA結合ドメインを含む、請求項1に記載の方法。

【請求項22】

前記目的のタンパク質は、DNA結合ドメインを含まない、請求項1に記載の方法。

【請求項23】

前記DNA結合ドメインは、ジンクフィンガー、ウイングドヘリックス、ロイシンジッパー、ホメオドメイン、及びヘリックス-ループ-ヘリックス(HLH)から成る群から選択される、請求項21又は請求項22に記載の方法。

30

【請求項24】

前記生物活性DNA結合部位のセットは、1つ又は複数の生物活性DNA結合部位を含む、請求項1に記載の方法。

【請求項25】

前記生物活性DNA結合部位セットは、10個又はそれ以上の生物活性結合部位を含む、請求項1に記載の方法。

【請求項26】

共に少なくとも1つのオルソログ遺伝子配列を含む場合、2つのゲノムDNA領域は等価である、請求項1に記載の方法。

【請求項27】

それぞれが各ゲノム中に第1及び第2の読み取り枠(ORF)に隣接する遺伝子間領域を含む2つのゲノムDNA領域は、(i)該2つの領域中の前記第1のORFがオルソログORFである場合、及び(ii)該2つの領域中の前記第2のORFがオルソログORFである場合に等価であると考えられる、請求項1に記載の方法。

40

【請求項28】

前記細胞は真核細胞である、請求項1に記載の方法。

【請求項29】

前記細胞は幹細胞である、請求項28に記載の方法。

【請求項30】

前記細胞は哺乳類の細胞である、請求項28に記載の方法。

50

- 【請求項 3 1】
前記細胞はヒト細胞である、請求項 3 0 に記載の方法。
- 【請求項 3 2】
前記細胞は初代細胞である、請求項 1 に記載の方法。
- 【請求項 3 3】
前記細胞は、組織生検に由来する、請求項 3 1 に記載の方法。
- 【請求項 3 4】
前記組織生検を、障害を罹患した被験体から単離する、請求項 3 3 に記載の方法。
- 【請求項 3 5】
前記細胞は、単細胞生物である、請求項 1 に記載の方法。 10
- 【請求項 3 6】
細胞ゲノム中の目的のタンパク質の生物活性 DNA 結合部位セットを変化させる薬剤の同定方法であって、
 (i) 実験細胞を候補薬剤と接触させるステップと、
 (i i) 請求項 2 に記載の方法にしたがって、ステップ (i) の前記細胞ゲノム中の目的のタンパク質の生物活性 DNA 結合部位セットを同定し、それにより、生物活性 DNA 結合部位の実験セットを生成するステップと、
 (i i i)
 (1) 前記生物活性 DNA 結合部位の実験セットと、
 (2) 前記目的のタンパク質の生物活性 DNA 結合部位のコントロールセットと 20
 を比較するステップと
 を含み、前記実験セット及び該コントロールセットが異なる場合に候補薬剤が同定される、方法。
- 【請求項 3 7】
前記コントロールセットは、前記候補薬剤と接触していないコントロール細胞に由来する、請求項 3 6 に記載の方法。
- 【請求項 3 8】
細胞中の目的のタンパク質によって転写的に調節される経路を同定する方法であって、
 (i) 請求項 2 に記載の方法にしたがって前記細胞ゲノム中の目的のタンパク質の生物活性 DNA 結合部位セットを同定するステップと、 30
 (i i) ステップ (i) で同定された該生物活性 DNA 結合部位セットへの前記目的のタンパク質の結合によって調節される可能性が高い少なくとも 2 つの候補遺伝子を同定するステップと
 を含み、少なくとも 2 つの候補遺伝子が同一経路のメンバーである場合に、前記目的のタンパク質によって転写的に調節される経路が同定される、方法。
- 【請求項 3 9】
前記目的のタンパク質についての生物活性 DNA 結合部位セットを変化させる薬剤又は条件への細胞の曝露によって、前記目的のタンパク質によって転写的に調節される経路を調整するステップをさらに含む、請求項 3 8 に記載の方法。
- 【請求項 4 0】 40
前記経路は生化学的経路である、請求項 3 8 に記載の方法。
- 【請求項 4 1】
前記経路は遺伝子発現経路である、請求項 3 8 に記載の方法。
- 【請求項 4 2】
前記経路は調節的経路である、請求項 3 8 に記載の方法。
- 【請求項 4 3】
候補遺伝子についてのプロモーターが少なくとも 1 つの生物活性 DNA 結合部位を含む場合に、前記候補遺伝子は目的のタンパク質によって調節される可能性が高い、請求項 3 8 に記載の方法。
- 【請求項 4 4】 50

候補遺伝子のプロモーター領域は、約 3 k b の 5 ' ~ 約 1 k b の 3 ' の転写開始部位を含む、請求項 4 3 に記載の方法。

【請求項 4 5】

目的のタンパク質が細胞ゲノムに異なって結合する 2 つの条件セットを同定する方法であって、

(i) 請求項 1 に記載の方法にしたがって細胞ゲノム中の前記目的のタンパク質についての第 1 の生物活性 D N A 結合部位セットを同定するステップと、ここで、該細胞は第 1 の条件セットに曝露され、

(i i) 請求項 1 に記載の方法にしたがって細胞ゲノム中の前記目的のタンパク質についての第 2 の生物活性 D N A 結合部位セットを同定するステップと、ここで、該細胞は第 2 の条件セットに曝露され、

(i i i) 前記第 1 の生物活性 D N A 結合部位セットと該第 2 の生物活性 D N A 結合部位セットとを比較し、該 2 つのセットが異なるかどうかを決定するステップとを含む、目的のタンパク質が細胞ゲノムに異なって結合する 2 つの条件セットを同定する方法。

10

【請求項 4 6】

細胞ゲノムに対する、目的の遺伝子によってコードされるポリペプチドの結合活性と相関する該目的の遺伝子の遺伝子産物の性質を同定する方法であって、

(i) 請求項 4 4 に記載の方法にしたがって、目的のタンパク質が前記細胞ゲノムに異なって結合する 2 つの条件セットを同定するステップと、

(i i) (a) 前記第 1 の条件セットに曝露された細胞中、及び (b) 前記第 2 の条件セットに曝露された細胞中の前記目的の遺伝子の遺伝子産物の性質を決定するステップと、

20

(i i i) 該遺伝子産物の少なくとも 1 つの性質がステップ (i i) の前記 2 つの細胞で異なるかどうかを決定するステップとを含む、それにより、細胞ゲノムへの目的の遺伝子の前記結合活性に相関する性質を同定する、方法。

【請求項 4 7】

細胞ゲノムへの目的の遺伝子によってコードされるポリペプチドの結合活性と相関する前記目的の遺伝子の遺伝子産物の性質を同定する方法であって、

(i) 請求項 3 6 に記載の方法にしたがって、細胞ゲノム中の目的のタンパク質の生物活性 D N A 結合部位セットを変化させる薬剤を同定するステップと、

(i i) (a) 該薬剤と接触される細胞中、及び (b) 該薬剤と接触されない細胞中の前記目的の遺伝子の遺伝子産物の性質を決定するステップと、

30

(i i i) 該遺伝子産物の少なくとも 1 つの性質がステップ (i i) の前記 2 つの細胞で異なるかどうかを決定するステップとを含む、それにより、細胞ゲノムへの目的の遺伝子の結合活性に相関する性質を同定する、方法。

【請求項 4 8】

前記性質は、タンパク質修飾、発現レベル、酵素活性、及び細胞内局在化から成る群から選択される、請求項 4 6 又は請求項 4 7 に記載の方法。

40

【請求項 4 9】

前記発現産物は m R N A である、請求項 4 6 又は請求項 4 7 に記載の方法。

【請求項 5 0】

前記発現産物はポリペプチドである、請求項 4 6 又は請求項 4 7 に記載の方法。

【請求項 5 1】

前記性質は、前記遺伝子産物の発現レベルを含む、請求項 4 6 又は請求項 4 7 に記載の方法。

【請求項 5 2】

前記性質は、前記遺伝子産物の前記細胞内局在を含む、請求項 4 6 又は請求項 4 7 に記

50

載の方法。

【請求項 5 3】

前記性質は、遺伝子産物のリン酸化状態を含む、請求項 4 6 又は請求項 4 7 に記載の方法。

【請求項 5 4】

前記性質は、前記遺伝子産物の分子量を含む、請求項 4 6 又は請求項 4 7 に記載の方法。

【請求項 5 5】

前記性質は、前記遺伝子産物の等電点を含む、請求項 4 6 又は請求項 4 7 に記載の方法。

10

【請求項 5 6】

前記性質は、前記遺伝子産物の核酸配列又はアミノ酸配列を含む、請求項 4 6 又は請求項 4 7 に記載の方法。

【請求項 5 7】

前記性質は、前記目的のタンパク質の別のポリペプチドとの物理的会合を含む、請求項 4 6 又は請求項 4 7 に記載の方法。

【請求項 5 8】

前記性質は、ポリペプチド遺伝子産物の酵素活性を含む、請求項 4 6 又は請求項 4 7 に記載の方法。

【請求項 5 9】

前記性質は、ポリペプチド遺伝子産物のオリゴマー状態を含む、請求項 4 6 又は請求項 4 7 に記載の方法。

20

【請求項 6 0】

目的のタンパク質が細胞ゲノムに異なって結合する 2 つの細胞の遺伝子型を同定する方法であって、

(i) 第 1 の遺伝子型の細胞ゲノム中の前記目的のタンパク質の第 1 の生物活性 DNA 結合部位セットを同定するステップと、

(i i) 第 2 の遺伝子型の細胞ゲノム中の前記目的のタンパク質の第 2 の生物活性 DNA 結合部位セットを同定するステップと、

(i i i) 前記第 1 の生物活性 DNA 結合部位セットと前記第 2 の生物活性 DNA 結合部位セットとを比較し、該 2 つのセットが異なるかどうかを決定するステップとを含む、方法。

30

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

[関連出願の相互参照]

本出願は、2004年8月25日に出願され、発明の名称が「生物活性 DNA 結合部位及び関連する方法」である米国特許出願番号 60 / 604 , 470 号及び 2004年3月4日に出願され、発明の名称が「真核生物ゲノムの転写調節コード及びその方法」である米国特許出願番号 60 / 550 , 074 号の出願日の利益を主張する。引用された出願の全ての教示は、参照することにより本明細書に組み込まれるものとする。

40

【0 0 0 2】

[連邦政府による資金提供を受けた研究開発の記載]

本明細書中に記載の発明は、その全体又は一部が、国立衛生研究所助成金番号 H G 0 0 2 6 6 8 によって支援された。米国政府は、本発明に一定の権利を有する。

【背景技術】

【0 0 0 3】

[発明の背景]

ゲノム配列は、遺伝子発現プログラムを調節し、タンパク質及び他の遺伝子産物を特定するために必要な情報を含む。DNA 結合転写調節因子は、特定の配列への結合によって

50

ゲノムの調節コードを解釈し、遺伝子発現を誘導又は抑制する (Jacob et al. J Mol Biol 3, 318-56 (1961), Kellis et al. Nature 423,241-54 (2003), Cliften et al. Science 301,71-6 (2003))。ゲノム配列の実質的部分は、調節されると考えられているが (Pritsker et al Genome Res 14, 99-108 (2004); Wang, et al. Bioinformatics 19,2369-80 (2003); Blanchette et al. Nucleic Acids Res 31,3840-2 (2003); Iyer et al. Nature 409,533- 8. (2001); Ren et al. Science 290,2306-9. (2000))、実質的に調節コードに寄与する DNA 配列は明確に定義されていない。対照的に、ヌクレオチド配列をタンパク質分子に翻訳するために使用されるトリプレットコードは既知である (Lee et al. Science 298,799-804. (2002), Lieb et al. Nat Genet 28,327-34 (2001), Roth et al. Nat Biotechnol 16, 939-45. (1998))。ゲノムの転写調節コードの知識により、全体的 (global) 遺伝子調節を支配する原理に対する新規の洞察を得ることができる。 10

【発明の開示】

【発明が解決しようとする課題】

【0004】

最近、酵母ゲノム中の機能的配列エレメントの同定に比較ゲノミクスが使用されている (Pritsker et al. Genome Res 14,99-108 (2004), Wang, et al. Bioinformatics 19,2369-80 (2003), Liu et al. Nat Biotechnol 20,835-9 (2002), Bailey et al. Proc Int Conf Intell Syst Mol Biol 3,21-9 (1995))。複数の酵母種のゲノム配列の比較分析により、系統発生的に保存された配列が明らかとなり、これらの配列を使用して、遺伝子及び推定調節エレメントの同定が容易になった。しかし、保存配列情報のみでは、転写調節因子 20 によって結合した配列のサブセット、結合調節因子の同一性、又は調節因子がその結合部位を占める条件は明らかとならない。

【0005】

したがって、 *in vivo* で転写調節因子によって結合した生物活性 DNA 結合部位の同定のための新規の方法及びアルゴリズムを開発する必要がある。

【課題を解決するための手段】

【0006】

[発明の概要]

本発明は、目的のタンパク質の生物活性 DNA 結合部位に関連する新規の方法を提供する。本発明の一態様は、生細胞などの細胞のゲノム中の目的のタンパク質の生物活性 DNA 結合部位を同定する方法を提供する。本発明の実施形態では、目的のタンパク質は、転写調節因子、DNA 組換えを媒介するタンパク質、DNA 修復を媒介するタンパク質、DNA 修飾を媒介するタンパク質、又は DNA 複製を媒介するタンパク質である。本発明の一態様はまた、目的のタンパク質の DNA 配列モチーフを同定する方法を提供する。 30

【0007】

本発明はまた、細胞中の目的のタンパク質が結合する生物活性 DNA 結合部位セットを変化させるアンチセンス薬、抗体、ポリペプチド、又は小分子などの薬剤を同定する方法を提供する。このような薬剤を、特に、転写調節因子などのタンパク質の細胞ゲノムへの結合又は結合の欠如により疾患を発症する場合、治療的に使用することができる。本発明は、さらに、目的のタンパク質が細胞ゲノムに異なって結合する条件又は細胞遺伝子型を 40 同定する方法を提供する。

【0008】

本発明は、さらに、目的のタンパク質によって調節される細胞経路 (生化学的経路、調節経路、遺伝子発現経路、及びシグナル伝達経路が含まれる) を同定する方法を提供する。本発明は、さらに、細胞ゲノムへの目的のタンパク質の差分結合に相関する転写調節因子などの目的のタンパク質の性質を同定する方法を提供する。

【発明を実施するための最良の形態】

【0009】

[発明の詳細な説明]

I . 概要

本発明は、部分的に、生物活性DNA結合部位に関連する新規の方法を提供する。本発明の一態様は、細胞ゲノム中の目的のタンパク質の生物活性DNA結合部位セットを同定する方法を提供し、この方法は、(i)細胞中の目的のタンパク質が結合するゲノムDNA領域セットを同定するステップと、(ii)同定したゲノムDNA領域中の候補DNA結合部位を同定するステップと、ここで、候補DNA結合部位は、目的のタンパク質のDNA配列モチーフに対応する配列を含み、(iii)候補DNA結合部位が、その細胞の異なる1つ又は複数の種中の同等なゲノム領域中に保存されているかどうかを決定するステップとを含み、ここで、少なくとも1つの種中に保存される候補DNA結合部位が生物活性DNA結合部位と考えられる。一実施形態では、ステップ(i)は、ゲノムDNA領域セットから目的のタンパク質のDNA配列モチーフを同定するステップをさらに含む。10

特定の実施形態では、DNA配列モチーフは、適切なコントロールと比較して統計的に有意な量のゲノムDNA領域セットによって富化される。本明細書中で使用される、用語「統計的に有意な」は、得られた結果が特定の可能性レベルで変動する可能性がないことを意味する。2つの最も一般的に特定される有意レベルは、0.05 ($p = 0.05$)及び0.01 ($p = 0.01$)である。0.05及び0.01に等しい有意レベルは、エラー確率がそれぞれ100のうち5及び100のうち1であることを意味する。しかし、より高い有意レベル(0.005、0.001、0.0001、0.00001、又はこれらの値の間など)を本明細書中に記載の方法と組み合わせて使用することができる。

【0010】

一実施形態では、適切なコントロールは、細胞中の目的のタンパク質によって結合しないゲノム領域セットを含む。別の実施形態では、適切なコントロールは、無作為に選択されたゲノム領域セットを含む。別の実施形態では、適切なコントロールは、無作為に生成された配列のセットを含む。別の実施形態では、適切なコントロールは、細胞中の目的のタンパク質変異形態によって結合されるゲノム領域セットを含む。20

【0011】

一実施形態では、少なくとも1つの異なる種中の同等なゲノム領域が目的のタンパク質のDNA配列モチーフに適合する核酸配列を含む場合、候補DNA結合部位は保存されている。別の実施形態では、DNA配列モチーフを少なくとも1つのアルゴリズムを使用して同定する。別の実施形態では、DNA配列モチーフを、アルゴリズムの組み合わせを使用して同定する。一実施形態では、アルゴリズムは、AlignACE、MEME、MDscan、Kellis Method、Mogul、Verbunculus、YMF、BioProspector、Motif Sampler、及びSUPERPOSITIONから成る群から選択される。30

【0012】

本明細書中に記載の生物活性DNA結合部位セットの同定方法の一実施形態では、ゲノムDNA領域は、プロモーター領域を含む。別の実施形態では、ゲノムDNA領域は、約50bp~約10kbの長さである。別の実施形態では、ステップ(i)は、目的のタンパク質のゲノム規模の位置分析(GWLA)を行うことを含む。特定の実施形態では、GWLAはChIP-chipを含むが、目的のタンパク質の結合部位の任意の同定方法を使用することができる。別の実施形態では、候補DNA結合部位は、20bp長未満(less than)である。40

【0013】

一実施形態では、DNA配列モチーフは、少なくとも1つの位置で生成される。別の実施形態では、1つ又は複数の異なる種が、細胞と同一の属に分類される。別の実施形態では、ステップ(iii)は、候補DNA結合部位が2つ又はそれ以上の異なる種中の同等なゲノム領域中に保存されるかどうかを決定することを含む。別の実施形態では、目的のタンパク質は、転写調節因子である。特定の実施形態では、目的のタンパク質は、DNA結合ドメインを含む。別の実施形態では、目的のタンパク質は、DNA結合ドメインを含まない。特定の実施形態では、DNA結合ドメインは、ジンクフィンガー、ウイングド(winged)ヘリックス、ロイシンジッパー、ホメオドメイン、及びヘリックス-ループ-ヘリ50

ックス (H L H) から成る群から選択される。別の実施形態では、目的のタンパク質は、DNA組換えを媒介するタンパク質、DNA修復を媒介するタンパク質、DNA修飾を媒介するタンパク質、又はDNA複製を媒介するタンパク質である。

【 0 0 1 4 】

一実施形態では、生物活性DNA結合部位セットは、1つ又は複数の生物活性DNA結合部位を含む。別の実施形態では、生物活性DNA結合部位セットは、10又はそれ以上の生物活性DNA結合部位を含む。別の実施形態では、共に少なくとも1つのオルソロガス遺伝子配列を含む場合、2つのゲノムDNA領域は等価である。別の実施形態では、それぞれが各ゲノム中に第1及び第2の読み取り枠 (O R F) に隣接する遺伝子間領域を含む2つのゲノムDNA領域は、(i) 2つの領域中の第1のORFがオルソロガスORFである場合、及び(i i) 2つの領域中の第2のORFがオルソロガスORFである場合に等価であると考えられる。

10

【 0 0 1 5 】

一実施形態では、細胞は、哺乳動物細胞などのような真核細胞であり、より好ましくはヒト細胞である。別の実施形態では、細胞は、組織生検などに由来する初代細胞である。特定の実施形態では、組織生検を、障害を罹患した被験体から単離する。さらに別の実施形態では、細胞は、酵母細胞などの単細胞生物である。一実施形態では、細胞は幹細胞である。本明細書中で使用される、用語「幹細胞」は、細胞系列を生じ、且つ分裂の際に異なる娘細胞 (供給源又は部分的に分化した幹細胞に置換される細胞) を産生する細胞として特徴づけることができる細胞をいう。幹細胞には、胚幹細胞、臍帯血幹細胞、及び成体 / 末梢血幹細胞が含まれる。

20

【 0 0 1 6 】

本発明の別の態様は、細胞ゲノム中の目的のタンパク質の生物活性DNA結合部位セットを変化させる薬剤の同定方法を提供し、この方法は、(i) 実験細胞を候補薬剤と接触させるステップと、(i i) 本明細書中に記載の方法のいずれかを使用するステップの実験細胞ゲノム中の目的のタンパク質についての生物活性DNA結合部位セットを同定し、それにより、生物活性DNA結合部位の実験セットを生成するステップと、(i i i) (1) 生物活性DNA結合部位の実験セットと、(2) 目的のタンパク質の生物活性DNA結合部位のコントロールセットとを比較するステップとを含み、この実験セット及びコントロールセットが異なる場合に上記候補薬剤が同定される。一実施形態では、コントロールセットは、候補薬剤と接触していないコントロール細胞に由来する。

30

【 0 0 1 7 】

本発明のさらに別の態様は、細胞中の目的のタンパク質によって転写的に調節される経路を同定する方法を提供し、この方法は、(i) 本明細書中に記載の方法のいずれかにしたがって細胞ゲノム中の目的のタンパク質の生物活性DNA結合部位セットを同定するステップと、(i i) (i) で同定された生物活性DNA結合部位セットへの目的のタンパク質の結合によって調節される可能性が高い少なくとも2つの候補遺伝子を同定するステップとを含み、この少なくとも2つの候補遺伝子が同一経路のメンバーである場合に目的のタンパク質によって転写的に調節される経路が同定される。特定の実施形態では、少なくとも2、3、4、又は5つの候補遺伝子が同一経路のメンバーである場合、目的のタンパク質によって転写的に調節される経路が同定される。一実施形態では、経路は生化学的経路である。別の実施形態では、経路は遺伝子発現経路である。別の実施形態では、経路は、調節経路である。別の実施形態では、候補遺伝子のプロモーターが少なくとも1つの生物活性DNA結合部位を含む場合に、候補遺伝子は目的のタンパク質によって調節される可能性が高い。一実施形態では、候補遺伝子のプロモーター領域は、3 kbの5' ~ 1 kbの3'の転写開始部位を含む。さらに別の実施形態では、以下でさらに考察するように、本発明は、目的のタンパク質が結合する結合部位を変化させる薬剤又は条件への細胞の曝露による、目的のタンパク質によって転写的に調節される経路の調整方法を提供する。

40

【 0 0 1 8 】

50

本発明の別の態様は、目的のタンパク質が細胞ゲノムに異なって結合する2つの条件セットを同定する方法を提供し、この方法は、(i)本明細書中に記載の方法のいずれか1つにしたがって細胞ゲノム中の目的のタンパク質の第1の生物活性DNA結合部位セットを同定するステップと、ここで、細胞は第1の条件セットに曝露される、(ii)生物活性DNA結合部位セットの同定のための本明細書中に記載の方法のいずれか1つにしたがって細胞ゲノム中の目的のタンパク質の第2の生物活性DNA結合部位セットを同定するステップと、ここで、細胞は第2の条件セットに曝露される、(iii)上記第1の生物活性DNA結合部位セットと上記第2の生物活性DNA結合部位セットとを比較し、この2つのセットが異なるかどうかを決定するステップとを含む。

【0019】

10

本発明の別の態様は、細胞ゲノムへの目的の遺伝子によってコードされるポリペプチドの結合活性と相関する目的の遺伝子の遺伝子産物の性質を同定する方法を提供し、(i)上記目的のタンパク質が細胞ゲノムに異なって結合する2つの条件セットを同定すること、(ii)(a)第1の条件セットに曝露された細胞中、及び(b)第2の条件セットに曝露された細胞中の目的の遺伝子の遺伝子産物の性質を決定すること、及び(iii)上記遺伝子産物の少なくとも1つの性質がステップ(ii)の2つの細胞で異なるかどうかを決定することを含み、それにより、細胞ゲノムへの目的の遺伝子の結合活性に相関する性質を同定することを含む。

【0020】

20

本発明の関連する態様は、細胞ゲノムへの目的の遺伝子によってコードされるポリペプチドの結合活性と相関する目的の遺伝子の遺伝子産物の性質を同定する方法を提供し、この方法は、(i)細胞ゲノム中の目的のタンパク質の生物活性DNA結合部位セットを変化させる薬剤を同定すること、(ii)(a)上記薬剤と接触した細胞中、及び(b)上記薬剤と接触していない細胞中の目的の遺伝子の遺伝子産物の性質を決定すること、及び(iii)上記遺伝子産物の少なくとも1つの性質がステップ(ii)の2つの細胞で異なるかどうかを決定することを含み、それにより、細胞ゲノムへの目的の遺伝子の結合活性に相関する性質を同定することを含む。

【0021】

一実施形態では、性質は、タンパク質修飾、発現レベル、酵素活性、及び細胞内局在化から成る群から選択される。一実施形態では、発現産物はmRNA又はポリペプチドである。別の実施形態では、性質は、遺伝子産物の発現レベルを含む。

30

【0022】

別の実施形態では、性質は、遺伝子産物の細胞内局在化を含む。別の実施形態では、性質は、遺伝子産物のリン酸化状態を含む。別の実施形態では、性質は、遺伝子産物の分子量を含む。別の実施形態では、性質は、遺伝子産物の等電点を含む。別の実施形態では、性質は、遺伝子産物の核酸配列又はアミノ酸配列を含む。別の実施形態では、性質は、目的のタンパク質の別のポリペプチドとの物理的結合を含む。

【0023】

本発明の別の態様は、目的のタンパク質が細胞ゲノムに異なって結合する2つの細胞遺伝子型を同定する方法を提供し、この方法は、(i)第1の遺伝子型の細胞ゲノム中の目的のタンパク質の第1の生物活性DNA結合部位セットを同定するステップと、(ii)第2の遺伝子型の細胞ゲノム中の目的のタンパク質の第2の生物活性DNA結合部位セットを同定するステップと、(iii)上記第1の生物活性DNA結合部位セットと上記第2の生物活性DNA結合部位セットとを比較し、この2つのセットが異なるかどうかを決定するステップとを含む。

40

【0024】**II. 定義**

便宜上、明細書、実施例、及び添付の特許請求の範囲中で使用した一定の用語をここに集める。他で定義しない限り、本明細書中で使用される全ての技術用語及び科学用語は、本発明に属する当業者によって一般的に理解されている意味と同義である。

50

【0025】

冠詞「a」及び「an」は、1つ又は1つを超える（すなわち、少なくとも1つの）文法上の冠詞の対象物に本明細書中で使用される。例として、「an element」は、1つの要素又は1つを超える要素を意味する。

【0026】

用語「含む」は、句「含むが、これらに限定されない」を意味するために本明細書中で使用され、この句と交換に使用することが可能である。

【0027】

用語「又は」は、別途明確に示されない限り、用語「及び/又は」を意味するために本明細書中で使用され、この用語と交換に使用することが可能である。

10

【0028】

用語「等」は、句「等であるが、これらに限定されない」を意味するために本明細書中で使用され、この句と交換に使用することが可能である。

【0029】

本発明の方法によって治療を受けるべき「患者」又は「被験体」は、ヒト又は非ヒト動物のいずれか、好ましくは哺乳動物を意味することができる。

【0030】

用語「コードする」は、DNA分子の転写に起因するRNA産物、RNA分子の翻訳に起因するタンパク質、又はDNA分子の転写及びその後のRNA産物の翻訳に起因するタンパク質を含む。

20

【0031】

用語「プロモーター」は、遺伝子の転写を開始させるDNA配列を意味するために本明細書中で使用される。プロモーターは、典型的には、遺伝子の5'側で見出され、開始コドンの近位に存在する。プロモーターが誘導性の場合、転写速度はインデューサーに反応して増加する。プロモーターは、転写調節因子の結合部位としての機能を果たすDNA結合エレメントに作動可能に連結することができる。用語「哺乳動物プロモーター」は、哺乳動物細胞で活性なプロモーターを意味するために本明細書中で使用される。同様に、「原核生物プロモーター」は、原核細胞で活性なプロモーターをいう。

【0032】

用語「発現」は、DNAからポリペプチドが産生される過程を意味するために本明細書中で使用される。この過程は、遺伝子のmRNAへの転写及びこのmRNAのポリペプチドへの翻訳を含む。使用される文脈によって、「発現」は、RNA、タンパク質、又はその両方の産生をいうことができる。

30

【0033】

用語「組換え」は、天然で隣接しない配列を含む任意の核酸を意味するために本明細書中で使用される。組換え核酸を、例えば、分子生物学的方法を使用して*in vitro*で生成するか、例えば、相同組換え又は非同組換えによる新規の染色体位置への核酸の挿入によって*in vivo*で生成することができる。

【0034】

用語「転写調節因子」は、一定の環境条件下でプロモーター駆動DNA配列の転写を防止又は阻害するように作用する生化学的エレメント（例えば、リプレッサー又は核阻害タンパク質）、一定の環境条件下でプロモーター駆動DNA配列の転写を受容又は刺激するように作用する生化学的エレメント（例えば、インデューサー又はエンハンサー）をいう。

40

【0035】

用語「マイクロアレイ」は、紙、ナイロン、若しくは他の膜型、フィルター、チップ、ガラスライド、又は任意の他の適切な固体支持体などの基板上で合成された異なるポリヌクレオチド又はオリゴヌクレオチドのアレイをいう。

【0036】

用語「障害」及び「疾患」は、包含的に使用され、身体の任意の部位、器官、又は系（

50

又はこれらの任意の組み合わせ)の正常な構造又は機能からの任意の逸脱をいう。特定の疾患は、特徴的な症状及び徴候(生物学的変化、化学的变化、及び物理的变化が含まれる)によって現れ、しばしば、種々の他の要因(人口統計学的要因、環境要因、雇用要因、遺伝的要因、及び病歴要因が含まれるが、これらに限定されない)に関連する。一定の特徴的な徴候、症状、及び関連する要因を種々の方法によって定量し、それにより重要な診断情報を得ることができる。

【0037】

用語「調整」は、応答のアップレギュレート(すなわち、活性化又は刺激)、ダウンレギュレート(すなわち、阻害又は抑制)、又はこれら2つが組み合わせて起こること若しくは個別に起こることをいう。「調整因子(modulator)」は、調整する化合物又は分子であり、例えば、アゴニスト、アンタゴニスト、活性化因子、刺激因子、抑制因子、又は阻害剤であり得る。

10

【0038】

用語「アゴニスト」は、タンパク質(例えば、ポリペプチドX)の生物活性を模倣するかアップレギュレートする(例えば、増強するか補完する)薬剤をいう。アゴニストは、野生型タンパク質又は野生型タンパク質の少なくとも1つの生物活性を有する誘導体であり得る。アゴニストはまた、遺伝子発現をアップレギュレートするか少なくとも1つのタンパク質の生物活性を増加させる化合物であり得る。アゴニストはまた、ポリペプチドの別の分子(例えば、標的ペプチド又は核酸)との相互作用を増加させる化合物であり得る。

20

【0039】

用語「アンタゴニスト」は、少なくとも1つのタンパク質の生物活性をダウンレギュレートする(例えば、抑制するか阻害する)薬剤をいう。アンタゴニストは、タンパク質と別の分子(例えば、標的ペプチド又は酵素基質)との間の相互作用を阻害するか減少する化合物であり得る。アンタゴニストはまた、遺伝子発現をダウンレギュレートするか、発現タンパク質の存在量を減少させる化合物であり得る。

【0040】

本明細書中で使用される、用語「遺伝子間領域」は、隣接読み取り枠(ORF)の間に存在するゲノムDNA配列をいう。遺伝子間領域は、プロモーター、オペレーター、又はエンハンサーなどの調節エレメントを含み得るが、調節配列をコード領域中に配置することもできる。

30

【0041】

III. 生物活性DNA結合部位の同定方法

本発明の一態様は、細胞ゲノム中の目的のタンパク質の生物活性DNA結合部位セットを同定する方法を提供する。1つの特定の態様では、方法は、(i)細胞中の目的のタンパク質が結合するゲノムDNA領域セットを同定するステップと、(ii)同定したゲノムDNA領域中の候補DNA結合部位を同定するステップと、ここで、候補DNA結合部位は、目的のタンパク質のDNA配列モチーフに対応する配列を含み、(iii)候補DNA結合部位が、その細胞の種と異なる1つ又は複数の種中の同等なゲノム領域中に保存されているかどうかを決定するステップとを含み、少なくとも1つの種中に保存される候補DNA結合部位は生物活性DNA結合部位である。

40

【0042】

本発明の方法は、任意の特定の細胞型に制限されない。細胞は、例えば、原核細胞又は真核細胞であり得る。1つの好ましい実施形態では、細胞は哺乳動物細胞である。他の好ましい細胞には、霊長類細胞及びげっ歯類細胞(ヒト細胞及びマウス細胞など)並びに酵母、ゼブラフィッシュ、線虫(Celegans)、又はショウジョウバエなどのモデル生物由来の細胞が含まれる。本発明の方法で使用される細胞は、癌細胞株などの不死化細胞株を含む、in vitroで長期にわたって継代した細胞であってもよい。本発明で使用することができる細胞株には、胚幹細胞及び成体幹細胞などの幹細胞も含まれる。

【0043】

50

他の実施形態では、細胞は、幾らかでもあるならば (if any)、*in vitro* で最小に培養された初代細胞である。1つの例示的实施形態では、細胞は、新たに単離された組織に由来する (組織生検などに由来する)。このような細胞は、このような細胞中での遺伝子発現パターンが *in vivo* 状態と最も密接に類似すると予想されるので、いくつかの実施形態で好ましい。いくつかの実施形態では、細胞は、障害を罹患した被験体に由来する。このような細胞によって疾患を洞察し、障害の治療又は予防のための治療薬の同定の一助となり得る。

【0044】

いくつかの実施形態では、細胞は、単一の細胞型に由来する一方で、他の実施形態は、1つを超える細胞型の細胞を含み得る。例えば、ある実施形態では肝細胞のみを使用することができ、他の実施形態では、肝細胞、ニューロン、及び膵臓細胞を使用することができる。他の実施形態では、細胞には、生物全体又は生物の組み合わせ由来の細胞が含まれ得る。

10

【0045】

本明細書中に記載の方法における目的のタンパク質は、DNAと直接又は間接的に会合することができる。いくつかの実施形態では、目的のタンパク質は、DNA分子に直接結合することができるDNA結合ドメインを含む。例示的なDNA結合ドメインには、ジンクフィンガー、ウイングドヘリックス、ロイシンジッパー、ホメオドメイン、又はヘリックス-ループ-ヘリックス (HLH) が含まれる。別の実施形態では、目的のタンパク質は、DNA結合ドメインを欠く。本発明のいくつかの実施形態では、目的のタンパク質は、特定の転写因子、活性化補助因子、補抑制物質、又はその複合体を含む。転写因子は、プロモーター、エンハンサー、及びサイレンサーエレメントなどの特定の同族DNAエレメントに結合し、遺伝子発現の調節を担う。転写因子は、細胞の状況に依存して、転写活性化因子、転写抑制因子、又はその両方であり得る。一実施形態では、目的のタンパク質は、TRANSFACデータベース (E. Wingender, X. Chen, R. Hehl, H. Karas, I. Liebig, V. Matys, T. Meinhardt, M. Pr, I. Reuter及びF. Schacherer. TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res.*, 28: 316-319, 2000を参照のこと) に列挙された転写調節因子のうちのいずれか1つである。

20

【0046】

転写因子は、既知又は同定された転写因子の任意のクラス又は型に属し得る。既知のファミリー又は構造関連転写因子の例には、ヘリックス-ループ-ヘリックス、ロイシンジッパー、ジンクフィンガー、リングフィンガー、及びホルモン受容体が含まれる。転写因子を、疾患とのその既知の関連又は1つ又は複数の遺伝子の調節に基づいて選択することもできる。例えば、*c-myc*、*Rel/Nf-kB*、*neuroD*、*c-fos*、*c-jun*、及びE2Fなどの転写因子を標的化することができる。任意の転写活性化補助因子又は補抑制物質に対する抗体も本発明に従って使用することができる。特定の活性化補助因子の例には、CBP、CTIIA、及びSRAが含まれ、補抑制物質の特定の例には、mSin3タンパク質、MITR、及びLEUNIGが含まれる。さらに、ヒストンアセチラーゼ (HAT) 及びヒストンデアセチラーゼ (HDAC) などの転写複合体に関連するタンパク質によって調節される遺伝子を、本明細書中に記載の方法を使用して決定することもできる。

30

40

【0047】

本明細書中に記載の方法の他の実施形態では、目的のタンパク質は基本転写調節因子又は基本転写機構の構成要素である。特定の実施形態では、基本転写機構の構成要素は、RNAポリメラーゼ (*pol I*、*pol II*、*pol III*、TBP、NTF-1、及びSp1が含まれる) 及びTFIIDの任意の他の構成要素 (例えば、TAF (例えば、TAF250、TAF150、TAF135、TAF95、TAF80、TAF55、TAF31、TAF28、及びTAF20) が含まれる)、又はポリメラーゼホロ酵素の任意の他の成分を含む。

【0048】

50

本明細書中に記載の方法のいくつかの実施形態では、目的のタンパク質は細胞に固有である。「固有である」は、目的のタンパク質が細胞中に天然に存在することを意味する。他の実施形態では、目的のタンパク質は、細胞に由来せず、その代わりに、細胞操作（タンパク質の微量注入、タンパク質のリポソーム媒介送達（Weiner他, (1994) Immunomethods ; 4 (3): 201-9）、又はT a tポリペプチドへの融合などの細胞に侵入可能なポリペプチドへの融合（Becker-Hapak他, (2001) Methods. 24 (3): 247-56）が含まれる）によって導入される。より好ましい実施形態では、非天然タンパク質を、標準的な遺伝子操作技術及び組換えDNA技術によって細胞に導入する。

【0049】

いくつかの実施形態では、目的のタンパク質は、本方法で使用される細胞種と異なる種の細胞に固有である。いくつかの実施形態では、目的のタンパク質は、ウイルスタンパク質である。このような実施形態では、細胞を、細胞がウイルスに感染される条件下でウイルスと接触させて細胞中でウイルスタンパク質を発現させることができる。いくつかの実施形態では、目的のタンパク質をコードする遺伝子は、ミスセンス変異、ナンセンス変異、欠失、挿入、又は逆位を保有し、その結果、変異タンパク質を細胞中で発現することができる。いくつかの実施形態では、目的のタンパク質は融合タンパク質である。1つの特定の実施形態では、融合タンパク質は、タンパク質の視覚化又は精製が可能な「タグ」に融合した転写調節因子を含む。好ましいタグには、GFPポリメラーゼ、GSTドメイン、mycタグ、Hisタグ、又はクロマチンと複合体化した場合に目的のタンパク質を免疫沈降させる任意の他のタグが含まれる。他の実施形態では、目的のタンパク質は、人工転写調節因子を含む（WO 02 / 3 1 1 6 6号を参照のこと）。

【0050】

本明細書中に記載の方法のいくつかの実施形態では、目的のタンパク質は、疾患や障害の原因として関わっている。疾患や障害を引き起こし得る転写調節因子の例は、Medical Genetics、L. V. Jorde他、Elsevier Science 2003、Principles of Internal Medicine、第15版、Braunwald他編、McGraw-Hill、2001、American Medical Association Complete Medical Encyclopedia (Random House、Incorporated、2003)、及びThe Mosby Medical Encyclopedia、Glanze編（Plume、1991）等の科学文献や医学文献で見出すことができる。いくつかの実施形態において、障害は、以下の器官又は組織のうちの少なくとも1つの機能の損傷により特徴づけられる：脳、脊髄、心臓、動脈、食道、胃、小腸、大腸、肝臓、膵臓、肺、腎臓、尿道、卵巣、胸部、子宮、睾丸、陰茎、結腸、前立腺、骨、頭皮、筋肉、軟骨、甲状腺、副腎、下垂体、骨髄、血液、胸腺、脾臓、リンパ節、肌、目、耳、鼻、歯又は舌。

【0051】

生物活性DNA結合部位には、目的のタンパク質が結合する細胞のゲノムDNA上の部位が含まれる。当業者は、ゲノム中の多数の配列が目的のタンパク質のDNA配列モチーフと適合し得る一方で、これらの部位の多くは*in vivo*で目的のタンパク質によって結合されず、それにより、生物活性DNA結合部位と見なさないと認識するであろう。

【0052】

一実施形態では、生物活性DNA結合部位は、DNAへの目的のタンパク質の特異的結合に必要な最小のDNA配列を含む。別の実施形態では、生物活性DNA結合部位は、目的のタンパク質が天然又は生理学的条件下（生細胞中など）でDNAと結合する場合、目的のタンパク質によって物理的に接触されるDNA配列ストレッチを含む。別の実施形態では、生物活性DNA結合部位は、プロモーター領域又は目的のタンパク質が細胞中の天然の条件下で結合する任意の他の遺伝子の機能的に定義された調節領域を含む。

【0053】

本明細書中に記載の方法のいくつかの実施形態では、生物活性DNA結合部位セットは、1つの生物活性DNA結合部位を含む。別の実施形態では、生物活性DNA結合部位セットは、少なくとも2、3、4、5、6、7、8、9、10、12、15、20、25、30、40又は50の生物活性DNA結合部位を含有する。いくつかの実施形態では、細

胞が1つのセットの条件下で増殖する場合、セットは目的のタンパク質が結合する生物活性DNA結合部位を含み、一方、他の実施形態では、セットは1つより多くのセットの条件を含む。従って、セットは目的のタンパク質が結合する部位の「スナップショット」であることがあり、又はそれぞれが異なるセットの条件で取られた、様々なスナップショットの累積したセットであることがある。セットは1つの細胞型又は1つより多い細胞型において目的のタンパク質が結合する部位を含むことができる。例えば、セットは肝細胞のE2Fが結合する部位を含むことができ、又は肝細胞、脂肪細胞、若しくは神経細胞中の結合される部位を含むことができる。いくつかの実施形態において、セットは生物における目的のタンパク質の結合部位を含むことができる。例えば、セットは線虫における転写因子の部位を含むことができ、これは例えば、丸ごとの線虫(whole worm)が出発物質として使われる場合に得ることのできるセットである。別の実施形態では、セットは細胞のゲノム中の遺伝子の少なくとも5%、10%、20%、30%、50%、70%、80%、90%、95%、98%又は99%を有するプロモーター領域に位置する結合部位を含む。

【0054】

目的のタンパク質が細胞中で結合するゲノムDNA領域セットの同定は、当該技術分野において公知の任意の技術を用いて達成してもよい。好ましい一実施形態では、ゲノム領域はゲノムワイド位置解析(GWLA)を用いて同定される。GWLAはPCT国際出願番号WO01/16378号及びWO02/059371号、並びに米国特許番号第6,410,243号に記載されている。好ましい実施形態において、細胞中の目的のタンパク質が結合するゲノムDNA領域セットの同定は、クロマチン免疫沈降(ChIP)及びそれに続くDNAマイクロアレイ上の分析(ChIP-chip分析)(Buck及びLieb、(2004) Genomics 83:349-360、Ren他(2004) Methods Enzymol. 376:304-315、Urnov他(2003) Journal of Cellular Biochemistry 88:684-694、及びOrlando、Valerio(2000) TIBS 25:99-103を参照のこと)を含む。

【0055】

好ましい一実施形態において、目的のタンパク質が結合するゲノムDNA領域(すなわちクロマチンフラグメント)はクロマチン免疫沈降(ChIP)を用いて単離される。簡潔にいうと、この技術では、特殊な抗体を用いて、一致する抗原(すなわち転写調節因子)を含むクロマチン混合物を免疫沈降させ、免疫沈降物に存在するヌクレオチド配列を調べることを含む。抗体による特定の配列の免疫沈降は、この配列との抗原の相互作用を示す。例えばO'Neill他in Methods in Enzymology、第274巻、Academic Press、San Diego、1999、pp. 189-197、Kuo他(1999) Method 19:425-433、及び上記Ausubel他、第21章を参照のこと。従って、ある実施形態において、目的のタンパク質が結合するDNAフラグメントは目的のタンパク質と結合する抗体を用いて、同定される。

【0056】

本明細書中に記載の一実施形態では、目的のタンパク質をクロマチンに架橋する(Solomon, M. J. 及びVarshavsky, A., Proc. Natl. Sci. USA 82: 6470-6474; Orlando, V., TIBS, 25: 99-104)。細胞ゲノムへの細胞のDNA結合タンパク質の結合に使用することができる種々の方法が存在する。例えば、UV光を使用することができる。特定の実施形態では、ホルムアルデヒドを使用して、細胞のゲノムDNAにDNA結合タンパク質を架橋する。次いで、クロマチンを超音波処理などによって剪断して、目的のタンパク質に結合したクロマチンのフラグメントを生成することができる。

【0057】

一実施形態では、クロマチン免疫沈降技術を以下のように適用する。天然の転写調節因子又は組換え転写調節因子等の目的のタンパク質を発現する細胞を、この転写調節因子が安定に結合する場合、クロマチンに転写調節因子を架橋する薬剤で処理する。転写調節因子を、例えば、ホルムアルデヒド処理又は紫外線照射によってクロマチンに架橋することができる。架橋後、細胞核酸を単離し、断片化し、転写調節因子に指向する抗体の存在下でインキュベートする。抗体-抗原複合体を沈殿させ、架橋を戻す(例えば、ホルムアル

デヒド誘導性DNA-タンパク質架橋を加熱によって戻すことができる)、免疫沈降DNAの配列内容物を、特定の配列(例えば、プロモーター領域)の存在について試験する。抗体は、転写調節因子上のエピトープに直接結合することができるか、抗Myc抗体と共に使用する場合、mycタグなどの調節因子上のタグに結合することができる(Santa Cruz Biotechnology, sc-764)。さらに別の実施形態では、転写調節因子又は転写調節因子に対して使用されるタグに親和性を有する非抗体薬剤を抗体の代わりに使用する。例えば、転写調節因子が6ヒスチジンタグなどの親和性タグを含む場合、複合体を、ニッケル含有セファロースによるアフィニティクロマトグラフィによって単離することができる。ChIP法のさらなるバリエーションを、Kurdistani他Methods. 2003 31(1): 90-5; O'Neill他Methods. 2003, 31(1): 76-82; Spencer他, Methods. 2003; 31(1): 67-75; 及びOrlando他Methods 11: 205-214 (1997)に見出すことができる。 10

【0058】

本明細書中に記載の方法の一実施形態では、コントロール免疫沈降反応由来のDNAフラグメントを、コントロールとして単離クロマチンの代わりに使用する。例えば、試験される転写因子と反応しない抗体をクロマチンIP手順で使用してコントロールクロマチンを単離し、その後、転写調節因子と反応する抗体を使用して単離したクロマチンと比較することができる。好ましい実施形態では、試験される転写因子と反応しない抗体はまた、他の転写調節因子又はDNA結合タンパク質と反応しない。

【0059】

一実施形態では、標識プローブを、クロマチンフラグメント、任意選択的に、コントロールクロマチンフラグメントから生成する。標識プローブを、ライゲーション媒介ポリメラーゼ連鎖反応(LM-PCR)を使用してテンプレートDNAから生成することができる(例えば、Current Protocols in Molecular Biology, Ausubel, F. M. 他編1991及び米国出願第2003/0143599号(その教示全体が本明細書中で参照することにより組み込まれる)を参照のこと)。特定の実施形態では、LM-PCRは、LM-PCR反応への蛍光タグ化ヌクレオチドの含有による蛍光標識増幅DNAを含む。一般に、全ゲノムベースの増幅方法(例えば、実質的に無作為の増幅方法)は、多置換増幅(Multiple Displacement Amplification)(MDA)(Hosono他, Genome Res. 2003; 13(5): 954-64)、DOP(Telenius, Genomics 1992; 13: 718-725)、プライマー伸長予備増幅(すなわちPEP)(Zhang他, Proc. Natl. Acad. Sci. 1992; 89: 5847-5851)、標識プライマー及び/又はヌクレオチドと組み合わせたAlu内(inter-Alu)PCRなどを用いることができる。しかし、さらに他の態様では、事前の増幅ステップを使用せずにプローブを標識する。 20 30

【0060】

記載の方法の一実施形態では、コントロールDNAフラグメント由来の標識プローブ及び標識プローブを、全ゲノム又はゲノムのサブセット(例えば、単数の染色体又は複数の染色体)を示す実験スポット又はフィーチャーを含むDNAアレイとハイブリッド形成させる。増幅コントロールクロマチンと比較した増幅クロマチンフラグメント由来のマイクロアレイ上の各実験スポットの蛍光強度は、目的のタンパク質が特定のスポットに存在するDNA領域に結合するかどうかを示す。したがって、本明細書中に記載の方法を、全ゲノムにわたるタンパク質-DNA相互作用の検出に適用することができる。 40

【0061】

本明細書中に記載の方法のいくつかの実施形態では、標識プローブをDNAマイクロアレイとハイブリッド形成させて、目的のタンパク質によって結合したゲノムDNA領域を同定する。「バイオチップ」又は「アレイ」とも呼ばれるマイクロアレイは、典型的には、化学反応及び生化学反応を行うための直径が μm から mm の範囲の小型のデバイスであり、特に、本発明の実施形態に適切である。アレイを、本質的に半導体産業、及び/又は生化学産業で公知、且つ利用可能な任意の、及び全ての技術を使用したマイクロ電子工学及び/又はマイクロ加工によって構築することができるが、このような技術がポリヌクレオチド配列の沈殿及びスクリーニングを受け入れることが可能であり、適合可能である場合に限られる。マイクロアレイは、マイクロアレイのサンプル処理が速く、且つプロフィール及 50

び他のデータの収集 (generating cost) コストが低い点で特に望ましい。

【 0 0 6 2 】

DNA マイクロアレイ及びマイクロアレイ由来のデータの分析方法は当該技術分野で十分に説明されており、DNA Microarrays: A Molecular Cloning Manual, Ed by Bowtel 及び Sambrook 編 (Cold Spring Harbor Laboratory Press, 2002); Kohana による Microarrays for an Integrative Genomics (MIT Press, 2002); Knudsen による A Biologist's Guide to Analysis of DNA Microarray Data (Wiley, John & Sons, Incorporated, 2002); 及び Schema による DNA Microarrays: A Practical Approach, 第 205 巻 (Oxford University Press, 1999); 及び Methods of Microarray Data Analysis II, Lin 他 編 (Kluwer Academic Publishers, 2002) (その全体が本明細書中で参照することにより組み込まれる) が含まれる。

【 0 0 6 3 】

いくつかの実施形態では、本発明で使用される DNA マイクロアレイを、プロモーター配列と共に核酸を含むスポットを使用して構築することができる。本明細書中に記載の方法のいくつかの実施形態では、遺伝子のプロモーター領域は、遺伝子の転写開始部位の少なくとも 700 bp 上流から少なくとも 200 bp 下流までを含む。いくつかの実施形態では、プロモーター領域は、少なくとも約 30、40、50、又は 60 ヌクレオチド長を含む。特定の実施形態では、マイクロアレイのスポット上に見出される遺伝子のプロモーター領域は、少なくとも 30 ヌクレオチドの配列を含み、この配列は遺伝子の転写開始部位の 3 kb 上流から 1 kb 下流までに及ぶ領域と同一である。マイクロアレイ上にスポットティングしたより小さなプローブ (例えば、約 30 ~ 200 ヌクレオチド) は、目的のタンパク質が結合するより小さな対応するゲノム領域を同定することができるという利点を有するが、より大きなプローブと同一範囲を得るためにマイクロアレイ上により多数のプローブをスポットティングしなければならないという欠点を有する。

【 0 0 6 4 】

いくつかの実施形態では、目的のタンパク質が結合するゲノム DNA 領域の配列は、転写因子が結合することが見出されたマイクロアレイ上のスポット上の核酸配列に対応する。マイクロアレイ上のスポットが重複配列を有するいくつかの実施形態では、目的のタンパク質が結合するゲノム領域の配列は、スポットティングした各核酸の長さよりも短いと推定することができる。例えば、目的のタンパク質が配列 C - D - E - F 及び E - F - G - H を有するプローブに結合し、各文字が、例えば、約 20 ~ 500 bp のストレッチを示し、目的のタンパク質も配列 A - B - C - D 及び G - H - I - J を有するプローブに結合できない場合、目的のタンパク質が結合する領域は E - F であると推定することができる。したがって、好ましい実施形態では、アレイはスポットティングした核酸を含み、その配列は貼り付けられている (tiled) (すなわち、重複領域を共有している)。

【 0 0 6 5 】

本明細書中に記載の方法の一実施形態では、同定したゲノム領域中の候補 DNA 結合部位の同定は、同定したゲノム領域の配列をスキャンして目的のタンパク質についての DNA 配列モチーフに対応する配列を同定することを含む。具体例では、3つのゲノム領域 (それぞれ約 1 kb) が同定され、目的のタンパク質の DNA 配列モチーフが GCANTGC である場合、3つのそれぞれの 1 kb の配列を、このモチーフと適合する配列 (すなわち、サブシーケンス GCAATGC、GCAGTGC、GCACTGC、及び GCATTGC のうちのいずれか) についてスキャンする。このようなサブシーケンスがゲノム領域中で同定された場合、このサブシーケンスを候補 DNA 結合部位と決定する。したがって、候補 DNA 結合部位は、DNA 配列モチーフと同一の配列を有し (すなわち、DNA 配列モチーフのサブシーケンスの 1 つと同一の配列を有する)、同定されたゲノム領域中に存在する。同定したゲノム領域は、候補結合部位を含まないかもしれないか、1つを含むかも知れないか、1より多く含むかも知れない。

【 0 0 6 6 】

目的のタンパク質が結合するゲノム領域中の DNA 配列モチーフによって定義された配

列を有する候補結合部位の同定を、ゲノム領域の手作業による調査によって行うことができる。しかし、より好ましくは、当該技術分野で既知のソフトウェアプログラム及びアルゴリズムを使用して、プロセスを部分的又は完全に自動化する。本発明で使用することができるプログラムには、A h b プログラム (Rajewsky 他 (2002) BMC Bioinformatics, 3:30)、C l o v e r プログラム (Frith 他 (2004) Nucleic Acids Res., 32,1372-1381 を参照のこと)、及び M o t i f S c a n n e r プログラム (Thijs 他 (2001), Proceedings Recomb' 2001,305-312) が含まれるが、これらに限定されない。別の実施形態では、F u z z n u c プログラムを使用して、DNA 配列モチーフに対応するゲノム領域セット中の候補結合部位を同定する。F f u z z n u c は、H G M P, 英国から流通されている生物学的ソフトウェアツールの E M B O S S スーツの一部である一プログラムである (Rice 他 (2000) E M B O S S : The European Molecular Biology Open Software Suite. Trends in Genetics, 第16巻, No 6. pp. 276-277 を参照のこと)。さらに別の実施形態では、R O V E R (Relative OVER-abundance of cis-elements) プログラムを使用して、候補結合部位を同定する (Haverty 他 (2004) Nucleic Acids Res., 32,179-188 を参照のこと)。文献中に記載されているさらなるアプローチを使用することもできる (Sharan 他 (2003). Bioinformatics, 19 (Suppl 1), 1283-1291.; 及び Elkonet 他 (2003) Genome Res., 13,773-780. を参照のこと)。

【0067】

本明細書中に記載の方法のいくつかの実施形態では、目的のタンパク質の DNA 配列モチーフは知られていないか、定義が不十分である。したがって、いくつかの実施形態では、本方法は、同定されたゲノム DNA 配列から目的のタンパク質の DNA 配列モチーフを同定することをさらに含む。一実施形態では、DNA 配列モチーフの同定は、一般に、適切なコントロールと比較して目的のタンパク質が結合するゲノム DNA 領域中に富化される 1 つ又は複数の DNA 配列を同定することを含む。

【0068】

特定の実施形態では、適切なコントロールは、コントロールゲノム領域セットを含む。コントロールセットは、細胞中の目的のタンパク質によって結合されないゲノム領域セットを含み得る。他の実施形態では、コントロールセットは、無作為に選択されたゲノム領域セットを含む。さらに別の実施形態では、コントロールセットは、無作為化 DNA 配列を含む。好ましい実施形態では、無作為化 DNA 配列の塩基組成は、(i) 細胞ゲノム、(i i) 目的のタンパク質が結合するゲノム領域、又は (i i i) プロモーター又は他の調節領域の平均 (average) のいずれかの塩基組成と実質的に同一である。例えば、目的のタンパク質によって結合されるゲノム領域中の 60% 及び 40% の塩基対がそれぞれ A - T 及び G - C 塩基対である場合、無作為化は、これと同一の塩基対比率を含み得る。あるいは、適切なコントロールは、頻度予測を含むことができ、この予測は、実際に無作為化配列を生成する必要がなくモチーフ中の塩基対の頻度に基づいて所与のモチーフを予想する。

【0069】

別の実施形態では、コントロールセットは、その DNA 結合ドメイン中に欠失又は置換を有する変異形態等の目的のタンパク質の変異形態によって結合されるゲノム領域セットを含み得るのに対して、他の実施形態では、コントロールセットは、転写因子、基本転写機構の成分、又はヒストンなどの第 2 のタンパク質によって結合されるゲノム領域セットを含む。

【0070】

DNA 配列モチーフを同定するための当業者に既知の任意のアルゴリズム又はソフトウェアプログラムを用いて、同定されたゲノム領域セットから DNA 配列モチーフの統計的に富化されたセットを同定することができる。いくつかの実施形態において、DNA 配列モチーフは 1 つ又は複数の以下のプログラム又はアルゴリズムを用いて同定される。G i b b s S a m p l e r (Rajewsky 他 (2002) BMC Bioinformatics, 3:30, Lawrence 他 (1993) Science, 262, 208-214)、R ' M E S プログラム (Schbath S, (1997) J. Comp. Biol., 4,

189-192)、V e r b u m c u l u s プログラム (Apostolico他、(2000)Journal of Computational Biology、第7巻、第1/2号、Apostolico他(2004)Journal of Computer and Science Technology、第19巻、第1号、pp. 22-41)、Y M F プログラム (Sinha他(2002)Nucleic Acids Research、第30巻、第24号、5549-5560、Sinha他(2000)Eighth International Conference on Intelligent Systems for Molecular Biology, San Diego, CA, 344-354)、A l i g n A C E (Aligns Nucleic Acid Conserved Elements) (Hughes他(2000), Journal of Molecular Biology, 296(5):1205-14, Roth他、(1998)Nature Biotechnology, 16(10):939-45, 1998)、B i o P r o s p e c t o r プログラム (Liu X他(2001)Pac. Symp. Biocomput., 127-38)、M E M E プログラム (Bailey他(1994)Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, 28-36, AAAI Press)、M o t i f S a m p l e r (Thijs G(2001)他 Bioinformatics, 17(12), 1113-1122, Thijs G他 Journal of Computational Biology (special issue Recomb'2001), 9(2), 447-464, 2002)、及びS U P E R P O S I T I O N (Shinozaki D他、(2003)Bioinformatics, 19 Suppl 2:11206-11214)。

10

20

30

40

50

【0071】

特定の実施形態において、M o t i f D i s c o v e r y スキャン (M D s c a n) プログラムを用いて、同定されたゲノム領域セットからDNA配列モチーフを同定する (Liu XS他、(2002)Nat. Biotechnol., 20(8):835-9)。別の特定の実施形態において、M o g u l プログラム (多重アルゴリズムを包含する) を用いてDNA配列モチーフを同定する (Rust他 Int. Conf. on Systems Biology (ISMB2003)、2003)。さらなるアルゴリズムもまた記載されている (Bailey他(1994)Proc. Int. Conf. Intell. Syst. Mol. Biol., 2、28-36を参照のこと)。いくつかの実施形態では、1つより多くのアルゴリズムを用いてDNA配列モチーフを同定する。

【0072】

別の実施形態において、同定されるDNA配列モチーフの有効性について、ワンハイブリッドシステム又は *in vitro* でのDNA結合部位選択法などのアプローチを用いて実験的に試験した (Blackwell TK(1995)Methods Enzymol., 254:604-18、Blackwell他(1990)Science. 250(4984):1149-51、Blackwell他(1990)Science. 250(4984):1104-10を参照のこと)。さらなる方法については、米国特許出願番号第2004/0115794号に開示されており、その全体が本明細書中に参照することにより組み込まれるものとする。

【0073】

別の実施形態において、以下のモチーフ見出プログラムの少なくとも1つが使用される：A l i g n A C E (Roth、F. P.、Hughes、J. D.、Estep、P. W. & Church、G. M. Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. Nat Biotechnol 16、939-45(1998))、M E M E (Bailey、T. L. & Elkan、The value of prior knowledge in discovering motifs with MEME. Proc Int Conf Intell Syst Mol Biol 3、21-9(1995))、M D s c a n (Liu、X. S.、Brutlag、D. L. & Liu、J. S. An algorithm for finding protein-DNA-binding sites with applications to chromatin-immunoprecipitation microarray experiments. Nat Biotechnol 20、835-9(2002))、Kellis他によって記載されている保存に基づく方法 (conservation-based method) (Kellis他、Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature 423、241-54(2003))、本明細書中では「K e l l i s 方法」として参照され、新規の保存に基づく方法はC O N V E R G E と呼ばれている。

【0074】

目的のタンパク質の生物活性DNA結合部位セットを同定するための本明細書中に記載の方法の一実施形態は、候補DNA結合部位が細胞の種とは異なった1つ又は複数の種中で保存されているかどうかを決定することを含む。特定の実施形態において、上記方法は候補DNA結合部位が1つ又は複数の種中の同等なゲノム領域において保存されているか

どうかを決定することを含む。ある実施形態では、異なる種が細胞と同じ系統学的階級の下に分類される。別の実施形態では、異なる種は細胞と同じ順序で分類される。好ましい実施形態では、異なる種は細胞と同じ属で分類される。別の実施形態において、細胞のリボソームRNAと異なる種のリボソームRNAとの間の配列同一性の割合(percent sequence identity)は少なくとも、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、97.5%、98%、99%、99.5%、99.8%又は99.0%である。別の実施形態において、細胞のmtDNAと異なる種のmtDNAとの間の配列同一性の割合は、少なくとも、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、97.5%、98%、99%、99.5%、99.8%又は99.0%である。別の実施形態では、細胞のコード領域と異なる種のコード領域との間の平均配列同一性の割合は、少なくとも、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、97.5%、98%、99%、99.5%、99.8%又は99.0%である。

【0075】

一実施形態では、候補DNA結合部位は、DNA結合モチーフに適合する配列が同等なゲノム領域中に存在する場合、生物活性DNA結合部位と考えられる。特定の実施形態では、候補DNA結合部位は、DNA結合モチーフに適合する配列が少なくとも1つの異なる種中の同等なゲノム領域中に存在する場合、生物活性DNA結合部位と考えられる。一実施形態では、2つのゲノムDNA領域は、これらがオルソログ遺伝子を含む領域などのオルソログ配列である場合、等価であると考えられる。用語「オルソログ遺伝子」は、そのヌクレオチド配列中で互いに十分に類似しており、これらの遺伝子座が共通の先祖の遺伝子を起源とすると示唆される、異なる種中のゲノム遺伝子座をいう。オルソログ遺伝子は、ゲノム内で遺伝子が重複される場合よりもむしろ1つの系列が2つの種に分かれる場合に生じる。オルソログタンパク質は、2つの異なる種の遺伝子によってコードされ、この遺伝子はオルソログと考えられる。関連する実施形態では、2つの領域は、これらがオルソログ遺伝子配列を含む場合、等価であると考えられる。

【0076】

一実施形態では、等価又はオルソログなゲノム領域の同定を、同定されたゲノム領域と別の種由来のゲノム配列とを比較するためのDNAアラインメントプログラムを使用することができる。一実施形態では、候補結合部位に隣接する少なくとも50、100、200、400、600、800、1,000、1,500、3,000、5,000、又は10,000塩基の配列を使用して、他の種のゲノム配列中の同等なゲノム領域を検索する。DNA配列のアラインメント及び比較のためのプログラム及びアルゴリズムは、当該技術分野で既知である。一実施形態では、DNA Block Aligner (DBA) プログラム (GeneWiseソフトウェアセットの一部) を使用して、同定されたゲノム配列と別の種由来のゲノムDNAとを比較する (Ewan他(2004) GeneWise and Genomewise. Genome Research 14: 988-995を参照のこと)。進化的に保存された配列の同定方法は、米国特許出願第2003/0017474号及び同第2003/0119015号 (その内容全体が本明細書中で参照することにより組み込まれる) に記載されている。使用することができる他のプログラムには、Mauveプログラム (Darling他(2004) Genome Res; 14 (7):1394-403)、OrthoMCLプログラム (Li他(2003) Genome Res; 13 (9): 2178-89) が含まれる。Thomas他(2003) Nature 14; 424 (6950): 788-93も参照のこと。

【0077】

ゲノム領域間の核酸配列の相同性を、当該技術分野で既知の種々の配列比較アルゴリズム及びプログラムのいずれかを使用して評価することができる。このようなアルゴリズム及びプログラムには、TB LASTN、BLASTP、FASTA、TFASTA、及びCLUSALW (Pearson及びLipman, 1988, Proc. Natl. Acad. Sci# USA 85 (8): 2444-2448; Altschul他, 1990, J. Mol. Biol. 215 (3): 403-410; Thompson他, 1994, Nucl

Acids Res. 22 (2): 4673-4680; Higgins他, 1996, Methods Enzymol. 266: 383-402; Altschul他, 1990, J. Mol. Biol. 215 (3): 403-410; Altschul他, 1993, Nature Genetics 3: 266-272)が含まれるが、これらに決して制限されない。特定の実施形態では、核酸配列の相同性を、当該技術分野で既知のBasic Local Alignment Search Tool (「BLAST」)を使用して評価する(例えば、Karlin及びAltschul, 1990, Proc. Natl. Acad. Sci. USA 87: 2267-2268; Altschul他, 1990, J. Mol. Biol. 215: 403-410; Altschul他, 1993, Nature Genetics 3: 266-272; Altschul他, 1997, Nuc. Acids Res. 25: 3389-3402を参照のこと)。

【0078】

BLASTプログラムは、クエリーアミノ酸配列又は核酸配列と、好ましくは、タンパク質配列又は核酸配列のデータベースから得た試験配列との間の類似のセグメント(本明細書中で「高スコアリングセグメント対」と呼ばれる)の同定によって相同配列を同定する。高スコアリングセグメント対を、好ましくは、スコアリング行列によって同定し(すなわち、アラインメントする)、スコアリング行列は、その多くが当該技術分野で既知である。使用されるスコアリング行列は、BLOSUM62行列であり得る(Gonnet他, 1992, Science 256: 1443-1445; Henikoff 及びHenikoff, 1993, Proteins 17:49-61)。PAM行列又はPAM250行列も使用することができる(例えば、Schwartz及びDayhoff編, 1978, Matrices for Detecting Distance Relationships: Atlas of Protein Sequence and Structure, Washington: National Biomedical Research Foundationを参照のこと)。BLASTプログラムは、同定された全ての高スコアリングセグメント対の統計的有意性を評価し、好ましくは、ユーザー指定相同率などのユーザー指定の有意性の閾値を満たすセグメントを選択する。一実施形態では、高スコアリングセグメント対の統計的有意性を、Karlinの統計的有意性の式を使用して評価する(例えば、Karlin及びAltschul, 1990, Proc. Natl. Acad. Sci. USA 87: 2267-2268を参照のこと)。上記アルゴリズムと共に使用したパラメーターを、研究した配列の長さ及び相同性の程度に依存して適合することができる。いくつかの実施形態では、パラメーターは、ユーザーからの指示のないアルゴリズムによって使用されたデフォルトパラメーターであり得る。

【0079】

2配列間の配列の比較並びに同一率及び類似率の決定を、数学アルゴリズムを使用して行うことができる。(Computational Molecular Biology, Lesk, A. M. 編, Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D. W. 編, Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part 1, Griffin, A. M. 及びGriffin, H. G. 編, Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; 及びSequence Analysis Primer, Gribskov, M. 及びDevereux, J. 編, M Stockton Press, New York, 1991)。1つの好ましい実施形態では、2つのアミノ酸配列間の同一率を、Blossom62行列又はPAM250行列のいずれかを使用したGCGソフトウェアパッケージ(<http://www.gcg.com>で利用可能)中のGAPプログラムに組み込まれたNeedleman and Wusch(J Mol. Biol. (48): 444-453 (1970))を使用して決定する。さらに別の実施形態では、2つのヌクレオチド配列間の同一率を、NWsgapdna-CMP行列を使用したGCGソフトウェアパッケージ(Devereux, J. 他, Nuc. Acids Res. 12(1):387 (1984))(<http://www.gcg.com>で利用可能)中のGAPプログラムを使用して決定する。別の実施形態では、2つのアミノ酸配列又はヌクレオチド配列の間の同一率を、ALIGNプログラム(バージョン2.0)に組み込まれたE. Myers及びW. Millerのアルゴリズム(CABIOS, 4: 11-17 (1989))を使用して決定する。

【0080】

特定の実施形態では、候補DNA結合部位は、目的のタンパク質のDNA配列モチーフに対応するDNA結合部位が、候補DNA結合部位に隣接する2kb、1.5kb、1kb、800bp、600bp、500bp、400bp、300bp、200bp、150bp、100bp、80bp、60bp、又は50bp未満のウィンドウ(window)に同

10

20

30

40

50

等な別の種中のゲノム領域内に存在する場合、保存されたDNA結合部位である。例えば、配列GGGACCCの候補DNA結合部位が細胞中のX遺伝子の転写開始部位の550bp上流に存在する例示的な制限されない例では、ウィンドウが300bpであり、且つ目的のタンパク質のDNA配列モチーフがGGGNCCCである場合、配列GGGACCC、GGGTCCC、GGGGCCC、又はGGGCCCCのいずれかがX遺伝子の転写開始部位から700~400bpに同等な他の種の領域中で見出された場合、候補DNA結合部位は保存されており、これは、これらの配列が全てDNA配列モチーフに適合するからである。ウィンドウ及び転写開始部位に先立つ領域中の2つの種間の配列同一性に依存して、細胞から700から400bpのウィンドウに同等な第2の種中の領域を、転写活性部位に関連してシフトすることができ、それにより、例えば、第2の種が転写開始部位の直前の10bpが欠失した場合、同等な位置は、690~390bpであり得る。当業者は、2領域間の配列アラインメントに基づいて、任意のこのようなシフトを確認することができるであろう。

10

【0081】

特定の実施形態において、別の種のオーソログ遺伝子のプロモーター中に目的のタンパク質のDNA配列モチーフに対応するDNA結合部位が存在する場合、候補DNA結合部位は保存されているDNA結合部位であると考えられる。ある特定の実施形態において、目的のタンパク質のDNA配列モチーフに対応するDNA結合部位が別の種のオーソログ遺伝子のプロモーター中に存在する場合、遺伝子のプロモーター中の候補DNA結合部位は、保存されているDNA結合部位であると考えられ、プロモーター領域は、転写開始部位に対して約5kb、4kb、3kb、2.5kb、2kb、1.5kb、1kb、900bp、800bp又は700bp5'から、転写開始部位に対して約50bp、100bp、150bp、200bp、250bp、300bp、400bp、500bp、600bp、800bp又は1,000bp3'までを含む。別の実施形態において、プロモーター領域は、転写開始部位に対して約5kb、4kb、3kb、2.5kb、2kb、1.5kb、1kb、900bp、800bp又は700bp5'から、転写開始部位に対して約50bp、100bp、150bp、200bp、250bp、300bp、400bp、500bp、600bp、800bp又は1,000bp3'までから成る。別の実施形態において、いくつか又は全てのプロモーター領域が実験的に定義され、結果として異なる遺伝子間のプロモーター領域の大きさは異なり得る。

20

【0082】

IV. さらなる方法

本発明の別の態様は、細胞ゲノム中の目的のタンパク質の生物活性DNA結合部位セットを変化させる薬剤の同定方法を提供する。1つの特定の態様は、細胞ゲノム中の目的のタンパク質の生物活性DNA結合部位セットを変化させる薬剤の同定方法を提供し、この方法は、(i)実験細胞を候補薬剤と接触させるステップと、(ii)本明細書中に記載の方法のいずれかにしたがって、ステップ(i)の細胞ゲノム中の目的のタンパク質の生物活性DNA結合部位セットを同定し、それにより、生物活性DNA結合部位の実験セットを生成するステップと、(iii)(1)生物活性DNA結合部位の実験セットと、(2)目的のタンパク質の生物活性DNA結合部位のコントロールセットとを比較するステップとを含み、実験セット及びコントロールセットが異なる場合に候補薬剤が同定される。

30

40

【0083】

薬剤の同定方法の一実施形態では、生物活性DNA結合部位の実験セットが薬物と接触していないコントロール細胞中の目的のタンパク質の生物活性DNA結合部位セットである、生物活性DNA結合部位のコントロールセットを提供する。一実施形態では、生物活性DNA結合部位のコントロールセットは、(a)異なる投薬量の薬剤と接触したか、(b)異なる薬剤と接触したか、(c)その組み合わせであるコントロール細胞中の目的のタンパク質の生物活性DNA結合部位セットである。あるいは、コントロールセットは、薬剤との接触前の細胞又は細胞群に由来し得る。いくつかの実施形態において、この方法は、単細胞よりもむしろ細胞集団と薬剤とを接触させること、その後、細胞集団又は小集

50

団を使用するステップを含む。

【0084】

好ましい実施形態では、実験細胞と薬剤との接触以外の実験条件（細胞の遺伝子型、細胞の成長条件、クロマチンの単離条件、免疫沈降条件など）は、実験細胞とコントロール細胞との間で実質的に同一である。一実施形態では、実験細胞を *in vitro* で薬剤と接触させるのに対して、他の実施形態では、細胞を *in vivo* で接触させる。細胞を、例えば、細胞を含む生物への薬剤の投与によって薬剤と *in vivo* で接触させることができる。

【0085】

いくつかの実施形態では、細胞を、細胞中での薬剤の遺伝的発現（RNA 又はポリペプチド薬等の薬剤をコードする実験細胞への導入遺伝子の導入など）によって薬剤と接触させる。導入遺伝子は、任意のタンパク質（転写調節因子又はキナーゼ及びホスファターゼなどの転写調節因子の活性を調節するタンパク質など）をコードし得る。導入遺伝子はまた、ヘアピン RNA などの阻害 RNA をコードすることができ、その結果、ヘアピン RNA が指示される遺伝子の機能をロックダウンすることができる。いくつかの実施形態では、導入遺伝子は、病状に関連する。

【0086】

いくつかの実施形態では、実験セット及びコントロールセットは、少なくとも1つの生物活性 DNA 結合部位が異なり、一方のセットで存在するが他方のセットで存在しない場合、候補薬剤が同定される。他の実施形態では、少なくとも2、3、4、5、10、20、又は30の生物活性 DNA 結合部位が他方のセットと比較して一方のセットで異なって存在する場合、候補薬剤が同定される。別の実施形態では、少なくとも0.5%、1%、2%、3%、4%、5%、又は10%の生物活性 DNA 結合部位が他方のセットと比較して一方のセットで異なって存在する場合、候補薬剤が同定される。

【0087】

本発明の関連する態様は、目的のタンパク質が細胞ゲノムに異なって結合する条件を同定する方法を提供する。本発明の1つの特定の態様は、目的のタンパク質が細胞ゲノムに異なって結合する2つの条件セットを同定する方法を提供し、この方法は、(i) 本明細書中に記載の方法のいずれかにしたがって細胞ゲノム中の目的のタンパク質の第1の生物活性 DNA 結合部位セットを同定するステップと、ここで細胞は第1の条件セットに曝露される、(ii) 本明細書中に記載の方法のいずれかにしたがって細胞ゲノム中の目的のタンパク質の第2の生物活性 DNA 結合部位セットを同定するステップと、ここで細胞は第2の条件セットに曝露される、(iii) 第1の生物活性 DNA 結合部位セットと第2の生物活性 DNA 結合部位セットとを比較し、2つのセットが異なるかどうかを決定するステップとを含む。

【0088】

一実施形態では、条件セットは、環境条件セットを含む。条件は、細胞が存在する物理的環境条件、化学的環境条件、及び/又はその部位の生物学的条件であり得る。任意の適切な時間、曝露することができる。曝露は、連続的、一過性、定期的、散發的などであり得る。物理的条件には、細胞が存在するサンプルの任意の物理的状态が含まれる。物理的状态は、サンプルの温度若しくは圧力又はその部位の光の量若しくは質（電磁放射）であり得る。あるいは又はさらに、物理的状态は、特に、その部位の電場、磁場、及び/又は粒子放射などに関し得る。化学的条件には、サンプル集団が処理される流動物の任意の化学的態様が含まれる。化学的態様は、特に、pH、イオン強度、及び/又は流動物の組成（ガス濃度）などに関し得る。細胞が生物中に存在する場合、特に、条件セットは、食事、身体的活動、性的行為、ストレスレベル、又は意識状態（起きている状態対眠っている状態など）、及び病状を含み得る。

【0089】

別の実施形態では、条件セットは、生物学的条件セットを含む。生物学的条件には、細胞が処理される共通の流動物体積 (fluid volume) の任意の生物学的態様（栄養素など）が

含まれる。生物学的条件には、成長因子、ケモカイン、又はサイトカインの有無も含まれ得る。生物学的態様には、さらなる細胞の存在、非存在、濃度、活性、又は型が含まれ得る。

【0090】

本発明の関連する態様はまた、目的のタンパク質が細胞ゲノムに異なって結合する2つの細胞遺伝子型を同定する方法を提供し、この方法は、(i)第1の遺伝子型の細胞ゲノム中の目的のタンパク質の第1の生物活性DNA結合部位セットを同定するステップと、(ii)第2の遺伝子型の細胞ゲノム中の目的のタンパク質の第2の生物活性DNA結合部位セットを同定するステップと、(iii)第1の生物活性DNA結合部位セットと第2の生物活性DNA結合部位セットとを比較し、2つのセットが異なるかどうかを決定するステップとを含む。

10

【0091】

一実施形態では、細胞の遺伝子型は、目的のタンパク質をコードする遺伝子の遺伝子型が異なる。他の実施形態では、細胞は複数の遺伝子の遺伝子型が異なり、そのうちの1つは目的のタンパク質であり得る。

【0092】

本発明の別の態様は、細胞中の目的のタンパク質によって調節される少なくとも1つの経路を同定する方法を提供する。本発明の1つの特定の態様は、細胞中の目的のタンパク質によって転写的に調節される経路を同定する方法を提供し、この方法は、(i)本明細書中に記載の方法のいずれかにしたがって細胞ゲノム中の目的のタンパク質の生物活性DNA結合部位セットを同定するステップと、(ii)ステップ(i)で同定された生物活性DNA結合部位セットへの目的のタンパク質の結合によって調節される可能性が高い少なくとも2つの候補遺伝子を同定するステップとを含み、少なくとも2つの候補遺伝子が同一経路のメンバーである場合に上記経路が同定される。

20

【0093】

本明細書中で使用される、用語「経路」は、生成物又は活性が生成される2つ又はそれ以上の連続的分子相互作用に参与する系の構成要素セットを意味することを意図する。経路は、種々の生成物又は活性を生成し、これらには、例えば、分子間相互作用、核酸又はポリペプチドの発現の変化、2つ又はそれ以上の分子の複合体の形成又は解離、代謝産物の蓄積又は破壊、酵素又は結合活性の活性化又は不活化が含まれ得る。したがって、用語「経路」には、種々の経路型(例えば、生化学的経路、遺伝子発現経路、及び調節経路など)が含まれる。同様に、経路には、これらの例示的経路型の組み合わせが含まれ得る。

30

【0094】

一実施形態では、経路は、生化学的経路である。生化学的経路には、例えば、代謝などにおいてある化合物が別の化合物に変換される酵素経路、並びに酵素活性、ポリペプチド構造、及びポリペプチド機能活性が変化するシグナル伝達経路が含まれ得る。生化学的経路の特定の例には、ガラクトースがグルコース-6-リン酸に変換される経路及び光受容体ロドプシンによって受容される光の粒子によってサイクリックAMPが生成される経路が含まれる。多数の他の生化学的経路が存在し、当業者に既知である。

【0095】

いくつかの実施形態では、生化学的経路は炭水化物代謝経路であり、特定の実施形態では、解糖/糖新生、クエン酸回路(TCA回路)、ペントースリン酸経路、ペントース及びグルクロン酸の相互交換、フルクトース及びマンノース代謝、ガラクトース代謝、アスコルビン酸及びアルダレート(aldarate)代謝、デンプン及びグルコース代謝、アミノ糖代謝、ヌクレオチド糖代謝、ピルビン酸代謝、グリオキシル酸及びジカルボン酸代謝、プロピオン酸代謝、ブタン酸代謝、C₅分岐二塩基酸代謝、イノシトール代謝、及びイノシトールリン酸代謝から成る群から選択される。

40

【0096】

いくつかの実施形態では、生化学的経路はエネルギー代謝経路であり、特定の実施形態では、酸化的リン酸化、ATP合成、光合成、炭素固定、還元的カルボン酸サイクル(C

50

O₂固定)、メタン代謝、窒素代謝、及び硫黄代謝から成る群から選択される。

【0097】

いくつかの実施形態では、生化学的経路は脂質代謝経路であり、特定の実施形態では、脂肪酸生合成(経路1)、脂肪酸生合成(経路2)、脂肪酸代謝、ケトン体の合成及び分解、ステロイドの生合成、胆汁酸生合成、C21-ステロイドホルモン代謝、アンドロゲン及びエストロゲン代謝、グリセロ脂質代謝、リン脂質分解、プロスタグランジン及びロイコトリエン代謝から成る群から選択される。

【0098】

いくつかの実施形態では、生化学的経路は、ヌクレオチド代謝経路であり、特定の実施形態では、プリン代謝及びピリミジン代謝から成る群から選択される。

10

【0099】

いくつかの実施形態では、生化学的経路はアミノ酸代謝経路であり、特定の実施形態では、グルタミン酸代謝、アラニン及びアスパラギン酸代謝、グリシン、セリン及びスレオニン代謝、メチオニン代謝、システイン代謝、バリン、ロイシン及びイソロイシン分解、バリン、ロイシン及びイソロイシン生合成、リジン生合成、リジン分解、アルギニン及びプロリン代謝、ヒスチジン代謝、チロシン代謝、フェニルアラニン代謝、トリプトファン代謝、フェニルアラニン、チロシン及びトリプトファン生合成、尿素サイクル、 α -アラニン代謝、タウリン及びヒポタウリン代謝、アミノホスホネート代謝、セレノアミノ酸代謝、シアノアミノ酸代謝、D-グルタミン及びD-グルタミン酸代謝、D-アルギニン及びD-オルニチン代謝、D-アラニン代謝及びグルタチオン代謝から成る群から選択される。

20

【0100】

いくつかの実施形態では、生化学的経路はグリカン生合成及び代謝経路であり、特定の実施形態では、N-グリカン生合成、N-グリカン分解、O-グリカン生合成、コンドロイチン/ヘパラン硫酸生合成、ケラタン硫酸生合成、グリコサミノグリカン分解、リポ多糖生合成、シルコシルホスファチジルイノシトール(clycosylphosphatidylinositol)(GPI)-アンカー生合成、ペプチドグリカン生合成、グリコスフィンゴリピド代謝、血液型グリコリピド生合成-ラクトシリース、血液型グルコリピド生合成-新ラクトシリース、グロボシド代謝及びガングリオシド生合成から成る群から選択される。

【0101】

いくつかの実施形態では、生化学的経路はポリケチド及び非リボソームペプチド経路の生合成であり、特定の実施形態では、I型ポリペプチド構造、12員環、14員環、16員環マクロライドの生合成、アンサマイシン生合成、ポリケチド糖単位生合成、非リボソームペプチド構造及びシデロホアグループ非リボソームペプチド生合成から成る群から選択される。

30

【0102】

いくつかの実施形態では、生化学的経路は補因子及びビタミン経路の代謝であり、特定の実施形態では、チアミン代謝、リボフラビン代謝、ビタミンB6代謝、ニコチン酸及びニコチンアミド代謝、パントテン酸及びCoA生合成、ピオチン生合成、葉酸生合成、葉酸による1つの炭素プール(One carbon pool by folate)、レチノール代謝、ポルフィリン及びクロロフィル代謝並びにユビキノン生合成から成る群から選択される。

40

【0103】

いくつかの実施形態では、生化学的経路は第二の代謝経路における生合成であり、特定の実施形態では、テルペノイド生合成、ジテルペノイド生合成、モノテルペノイド生合成、リモネン及びピネン分解、インドール及びトコンアルカロイド生合成、フラボノイド、スチルベン及びリグニン生合成、アルカロイド生合成I、アルカロイド生合成II、ペニシリン及びセファロsporin生合成、 β -ラクタム耐性、ストレプトミシン生合成、テトラサイクリン生合成、クラブラン酸生合成及びピューロマイシン生合成から成る群から選択される。

【0104】

50

一実施形態では、経路は、遺伝子発現経路である。遺伝子発現経路には、例えば、特定の遺伝子の発現を誘導、増強、又は抑制する分子が含まれ得る。したがって、遺伝子発現経路には、1つ又は複数の調節された遺伝子のプロモーター又は他の調節領域中の特定のDNA配列に結合するリプレッサー及び転写因子として機能するポリペプチドが含まれ得る。遺伝子発現経路の例は、成長刺激に応答する細胞周期遺伝子発現の誘導である。

【0105】

一実施形態では、経路は、調節経路である。調節経路には、例えば、特定の条件下で細胞機能を調節する経路が含まれ得る。調節経路は、例えば、系の構成要素の活性又は生化学的遺伝子発現若しくは他の経路型の活性の変化によって細胞機能を調節する。活性の変化には、例えば、特定の条件下での経路構成要素の発現、活性、又は物理的相互作用の変化の誘導が含まれる。調節経路の特定の例には、生化学系の環境刺激に反応して細胞機能を活性化する経路（細胞成長シグナルの存在に反応した細胞分化の阻害並びにガラクトースの存在及び糖抑制の非存在に反応したガラクトースの輸送及び触媒の活性化など）が含まれる。用語「構成要素」は、ネットワーク又は経路に関して使用する場合、生化学系、ネットワーク、又は経路の分子構成成分（例えば、ポリペプチド、核酸、他の高分子、又は他の生体分子など）を意味することを意図する。

10

【0106】

一実施形態では、経路は、シグナル伝達経路である。シグナル伝達経路には、MAPKシグナル伝達経路、Wntシグナル伝達経路、TGF-シグナル伝達経路、トール様受容体シグナル伝達経路、Jak-STATシグナル伝達経路、二次メッセンジャーシグナル伝達経路、及びホスファチジルイノシトールシグナル伝達経路が含まれる。

20

【0107】

経路同定のための本明細書中に記載の方法の一実施形態は、目的のタンパク質によって調節される可能性が高い候補遺伝子を同定することを含む。一実施形態では、候補遺伝子は、候補遺伝子のプロモーターが少なくとも1つの目的のタンパク質の同定された生物活性DNA結合部位を含む場合、目的のタンパク質によって調節される可能性が高い。特定の実施形態では、候補遺伝子のプロモーター領域は、転写開始部位の約3kb5'から1kb3'までを含む。別の特定の実施形態では、候補遺伝子のプロモーター領域は、転写開始部位の約2kb5'から0.5kb3'までを含む。別の実施形態では、候補遺伝子は、候補遺伝子の調節エレメントが少なくとも1つの同定された生物活性DNA結合部位を含む場合、目的のタンパク質によって調節される可能性が高い。一実施形態では、調節エレメントは、プロモーター領域及びエンハンサー領域から成る群から選択される。

30

【0108】

一実施形態では、候補遺伝子は、少なくとも1つの同定された生物活性DNA結合部位が遺伝子の転写開始部位の約3kb5'～1kb3'内に存在する場合、目的のタンパク質によって調節される可能性が高い。遺伝子が1つを超える転写開始部位を有する場合、一実施形態では、候補遺伝子は、少なくとも1つの同定された生物活性DNA結合部位が少なくとも1つの遺伝子の転写開始部位の約3kb5'～1kb3'内に存在する場合、目的のタンパク質によって調節される可能性が高い。特定の実施形態では、少なくとも1つの同定された生物活性DNA結合部位は、少なくとも1つの遺伝子の転写開始部位の約2kb5'～0.5kb3'内に存在する。

40

【0109】

別の実施形態では、候補遺伝子は、少なくとも1つの同定された生物活性DNA結合部位が目的の遺伝子の開始メチオニンコドンの約4kb5'～1kb3'内に存在する場合、又は開始メチオニンの使用が異なる複数の遺伝子産物が存在する場合には、少なくとも1つの同定された生物活性DNA結合部位が約4kb5'～1kb3'又は1つの開始メチオニン内に存在する場合、目的のタンパク質によって調節される可能性が高い。

【0110】

当業者は、候補遺伝子を、例えば、遺伝子の機能及び経路へのその分類を記載するいくつかのデータベースのうちのいずれかの調査並びに/又は文献の調査によって経路に割り

50

当てることができる (Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology. Gerhard Michal (編集者) Wiley, John & Sons, Incorporated, (1998); Biochemistry of Signal Transduction and Regulation, Gerhard Krauss, Wiley, John & Sons, Incorporated, (2003); Signal Transduction. Bastien D. Gomperts, Academic Press, Incorporated (2003)も参照のこと)。使用することができるデータベースには、<http://www.genome.jp/kegg/cegg4.html>; Pubmed、OMIM、及びEntrez (<http://www.ncbi.nih.gov>); Swiss-Protデータベース (<http://www.expasy.org/>) が含まれるが、これらに限定されない。

【0111】

1つの好ましい実施形態では、遺伝子が割り当てられた経路を、Biomolecular Interaction Network Database (BIND) (<http://www.blueprint.org/bind/>、より好ましくは<http://www.blueprint.org/bind/search/bindsearch.html>) (Bader GD, Betel D, Hogue CW. (2003) BIND: the Biomolecular Interaction Network Database. Nucleic Acids Res. 31(1): 248-50;及びBader GD, Hogue CW. (2003) An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics. 4(1)も参照のこと)を使用して同定する。BIMDデータベースの1つの特徴は、クエリ-遺伝子を割り当てる経路を列挙し、それにより、遺伝子が割り当てられる経路の同定を可能にする。さらに、米国特許出願番号2003/0100996号は、経路の同定及び遺伝子の経路への分類を容易にするために使用することができる経路データベースの確立方法及び経路検索の実施方法を記載している。

10

20

【0112】

本発明の別の態様は、遺伝子調節の基礎をなす機構の同定方法を提供する。本発明は、細胞ゲノムへのその差分結合に相関する目的のタンパク質の生化学的変化の同定方法を提供する。本発明の1つの特定の態様は、細胞ゲノムへの目的の遺伝子によってコードされるポリペプチドの結合活性と相関する目的の遺伝子の遺伝子産物の性質を同定する方法を提供し、この方法は、(i)本明細書中に記載の方法のいずれかにしたがって、目的のタンパク質が細胞ゲノムに異なって結合する2つの条件セットを同定すること、(ii)(a)第1の条件セットに曝露された細胞中、及び(b)第2の条件セットに曝露された細胞中の目的の遺伝子の遺伝子産物の性質を決定すること、及び(iii)遺伝子産物の少なくとも1つの性質がステップ(ii)の2つの細胞で異なるかどうかを決定することを含む、それにより、細胞ゲノムへの目的の遺伝子の結合活性に相関する性質を同定することを含む。

30

【0113】

本発明の別の態様は、細胞ゲノムへの目的の遺伝子によってコードされるポリペプチドの結合活性と相関する目的の遺伝子の遺伝子産物の性質を同定する方法を提供し、この方法は、(i)本明細書中に記載の方法のいずれかにしたがって細胞ゲノム中の目的のタンパク質の生物活性DNA結合部位セットを変化させる薬剤を同定すること、(ii)(a)薬剤と接触した細胞中、及び(b)薬剤と接触していない細胞中の目的の遺伝子の遺伝子産物の性質を決定すること、及び(iii)遺伝子産物の少なくとも1つの性質がステップ(ii)の2つの細胞で異なるかどうかを決定することを含む、それにより、細胞ゲノムへの目的の遺伝子の結合活性に相関する性質を同定することを含む。

40

【0114】

本発明の関連する態様は、細胞ゲノムへの目的の遺伝子によってコードされるポリペプチドの結合活性と相関する目的の遺伝子の遺伝子産物の性質を同定する方法を提供し、この方法は、(i)本明細書中に記載の方法のいずれかにしたがって細胞ゲノム中の目的のタンパク質の生物活性DNA結合部位セットを変化させる薬剤を同定すること、(ii)(a)ある投薬量の薬剤と接触した細胞中、及び(b)異なる投薬量の薬剤と接触した細胞中の目的の遺伝子の遺伝子産物の性質を決定すること、及び(iii)遺伝子産物の少なくとも1つの性質がステップ(ii)の2つの細胞で異なるかどうかを決定することを含む、それにより、細胞ゲノムへの目的の遺伝子の結合活性に相関する性質を同定すること

50

を含む。異なる投薬量の薬剤は、例えば、異なる濃度の薬剤を異なる持続時間で細胞に曝露することを含み得る。

【0115】

本明細書中に記載の遺伝子産物の性質を同定する方法の一実施形態では、発現産物は、mRNAである。別の実施形態では、発現産物はポリペプチドである。

【0116】

本明細書中に記載される遺伝子産物の特性を同定する方法の一実施形態において、遺伝子産物の性質は、遺伝子産物の翻訳後修飾、遺伝子産物の発現レベル、遺伝子産物の酵素活性、遺伝子産物のオリゴマー状態及び細胞内局在化から成る群から選択される。本明細書中で使用される場合、「翻訳後修飾」は、タンパク質のペプチド鎖の伸長とは考えられないタンパク質へのペプチド部分又は非ペプチド部分の添加を表す。翻訳後タンパク質修飾の例としては、リン酸化、アセチル化、メチル化、ADP-リボシル化、ユビキチン化、グルコシル化、カルボニル化、スモ化(sumoylation)、ビオチン化又はポリペプチド側鎖若しくは疎水性基の添加が挙げられる。1つの特定の実施形態において、評価されるリン酸化は、チロシン、セリン、スレオニン、又はヒスチジン残基のリン酸化である。別の特定の実施形態において、評価されるポリペプチド側鎖の添加はユビキチンの添加である。ユビキチン化では、全ての真核細胞に豊富に存在する進化的に高く保存された76アミノ酸ポリペプチドであるユビキチンが、標的タンパク質の1つ又は複数のリジン側鎖の1つのアミノ基へ共有結合する。さらに別の特定の実施形態において、評価される疎水性基の添加は、脂肪酸(例えば、ミリスチン酸又はパルミチン酸)の添加、イソプレノイド(例えば、ファルネシル又はゲンラニルゲンラニル(genranylgentranyl))の添加、又はグリコシル-ホスファチジルイノシトールアンカー(例えば、グリコシルを含む炭水化物群)の添加である。

10

20

【0117】

リン酸化には、チロシン、セリン、トレオニン、又はヒスチジンのリン酸化が含まれ得る。これらの修飾に使用することができる抗体には、例えば、ホスホチロシン特異的抗体、ホスホセリン特異的抗体、ホスホセリン特異的抗体、ホスホ-トレオニン-プロリン抗体が含まれ得る。これらの修飾の検出に使用することができる抗体には、Ser73でのリン酸化c-Junなどのタンパク質のリン酸化残基に特異的な抗体が含まれる。アセチル化を、アセチル化リジン抗体の使用によって検出することができる。メチル化特異的抗体を使用して、1つ又は複数のアミノ酸がメチル化したタンパク質を検出することができる。ADPリボシル化特異的抗体を使用して、ADPリボシル化修飾を有するタンパク質を検出することができる。ポリペプチド鎖付加の例は、ユビキチン化である。例えば、ユビキチン特異的抗体又はポリユビキチン特異的抗体を使用して、標的タンパク質のユビキチン化を検出することができる。

30

【0118】

当業者に既知の任意の一般的方法を適用して、遺伝子産物の発現レベルを決定することができる(mRNAレベル又はタンパク質レベルなど)。mRNAレベルを、例えば、ノーザンブロット、逆転写ポリメラーゼ連鎖反応(RT-PCR)、RNA保護アッセイ、又はmRNAの少なくとも一部とハイブリッド形成することができるプローブを含むDNAマイクロアレイなどの技術を使用して決定することができる。同様に、タンパク質レベルを、ウェスタンブロットティング、免疫サンドイッチアッセイ、ELISAアッセイ、又は任意の他の免疫学的技術などの当該技術分野で既知の技術を使用して定量することができる。核酸及びタンパク質の定量技術は、例えば、Molecular Cloning: A Laboratory Manual, 第3版, Sambrook及びRussell編(Cold Spring Harbor Laboratory Press: 2001); 及びCurrent Protocols in Cell Biology, Bonifacino, Dasso, Lippincott-Schwartz, Harford, 及びYamada編, John Wiley and Sons, Inc., New York, 1999(その全体が本明細書中で参照することによって組み込まれる)に見出すことができる。

40

【0119】

タンパク質の細胞内局在化を、免疫細胞化学によって決定することができる。あるいは

50

、細胞由来のオルガネラを精製し、異なるオルガネラ調製物中のポリペプチドの存在を決定することができる。あるいは、目的のタンパク質を蛍光タンパク質に融合して、*in vivo*でタンパク質の細胞内局在化をライブイメージングすることができる。例示的蛍光タンパク質には、緑色蛍光タンパク質 (GFP)、DsRed、zFP538、mRFP1、BFP、CFP、YFP、これらの変異体、又はこれらの機能活性フラグメントが含まれる。GFPは米国特許第5,491,084号に記載されており、zFP538はZagranichny他Biochemistry. 2004; 43 (16): 4764-72に記載されている。細胞内区画には、核、小胞体 (ER)、ゴルジ装置、被覆小窩、ミトコンドリア、エンドソーム、及び細胞質が含まれるが、これらに限定されない。mRNA分子の細胞内局在化を、固定細胞サンプルに対する*in-situ*ハイブリッド形成技術を使用して決定することもできる。

10

【0120】

遺伝子産物が酵素である一実施形態では、遺伝子産物の酵素活性は、遺伝子産物の性質である。遺伝子産物の酵素活性を、特定の酵素型についての標準的アッセイを使用して決定することができる (例えば、Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis, Robert Allen Copeland著; Wiley, John & Sons, Incorporated, 2000;及びEnzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems; Irwin H. Segel著; Wiley, John & Sons, Incorporated (1994)を参照のこと)。

【0121】

本明細書中に記載の遺伝子産物の性質の同定方法の一実施形態では、遺伝子産物の性質は、ポリペプチド遺伝子産物のオリゴマー状態である。オリゴマー状態は、目的のタンパク質が、単量体、二量体、三量体、六量体、又は他の多量体型として存在するかどうかをいう。目的のタンパク質のオリゴマー状態を、例えば、未変性ポリアクリルアミドゲル電気泳動 (PAGE) 又はゲル濾過クロマトグラフィを使用して決定し、目的のタンパク質を含む複合体のサイズを評価することができる。

20

【0122】

本明細書中に記載の方法のいくつかの実施形態では、2つの細胞中で少なくとも1つの遺伝子産物の性質が異なるかどうかの決定は、定量的測定によって遺伝子の性質が異なるかどうかを決定することを含む。他の実施形態では、これは、定量的測定によって、少なくとも10%、20%、40%、50%、75%、100%、又は200%異なるかどうかを決定することを含む。

30

【0123】

いくつかの実施形態では、タンパク質遺伝子産物の性質は、その分子量、その等電点、そのアミノ酸組成、そのアミノ酸配列、その二量体化状態 (例えば、単量体、二量体、三量体など)、又は複合体を形成するための他のポリペプチドとの会合などの物理的性質を含む。当該技術分野で既知の任意の技術 (質量分析、SDS-PAGE、等電点電気泳動、ペプチド配列決定、ゲル濾過、及び免疫沈降) を使用して、これらの性質を測定することができる。

【0124】

現行の方法のいくつかの実施形態では、遺伝子産物の性質の増加は、目的のタンパク質によって結合された生物活性DNA結合部位数の増加と相関する一方で、他の実施形態では、目的のタンパク質によって結合された生物活性DNA結合部位数の減少に相関する。他の実施形態では、遺伝子産物の性質の変化は、部位の総数を変化させない生物活性DNA結合部位の変化に相関する。

40

【0125】

いくつかの実施形態では、細胞を2つを超える条件又は2つを超える候補薬剤で試験し、その結果、相関関係をさらなる条件又は候補薬剤に拡大することができる。

【0126】

V. 候補薬剤

50

本明細書中に記載の方法のいくつかの実施形態では、候補薬剤は、小分子薬物、アンチセンス核酸、抗体、ペプチド、リガンド、脂肪酸、ホルモン、又は代謝産物を含む。

【0127】

候補薬剤（例えば、単一の化合物、2つ又はそれ以上の化合物の組み合わせ、化合物のライブラリー）として使用することができる例示的化合物には、核酸、ペプチド、ポリペプチド、ペプチド模倣物、抗体、アンチセンスオリゴヌクレオチド、RNAi構築物（siRNAが含まれる）、リボザイム、化学化合物、及び有機小分子が含まれる。化合物を、個別、化合物の組み合わせ、又は化合物のライブラリーとしてスクリーニングすることができる。理論に拘束されないが、本発明は、細胞表現型の調整が、増幅、生存、又は特定の系列への分化を調整し、それにより、細胞表現型が調整される特定の遺伝子及びシグナル伝達経路の活性化又は阻害を含み得ることを意図する。

【0128】

候補薬剤を、個別、1つ又は複数の他の化合物との組み合わせ、又は化合物のライブラリーとしてスクリーニングすることができる。化合物には、核酸、ペプチド、ペプチド模倣物、RNAi構築物、アンチセンスオリゴヌクレオチド、リボザイム、抗体、及び小分子が含まれる。

【0129】

アンチセンスオリゴヌクレオチドは、特定のタンパク質をコードするmRNAのコード鎖（センス鎖）に相補的（又はアンチセンス）である比較的短い核酸である。アンチセンスオリゴヌクレオチドは典型的にはRNAベースであるが、DNAベースでもあり得る。さらに、アンチセンスオリゴヌクレオチドをしばしば修飾してその安定性を増加させる。

【0130】

理論に拘束されないが、これらの比較的短いオリゴヌクレオチドのmRNAへの結合は、二本鎖RNAのストレッチを誘導して内因性RNAアーゼによるメッセージの分解を誘発すると考えられる。さらに、時折、オリゴヌクレオチドをメッセージのプロモーター付近に結合するように特異的にデザインし、これらの環境下で、アンチセンスオリゴヌクレオチドは、メッセージの翻訳をさらに妨害することができる。アンチセンスオリゴヌクレオチドが機能する特定の機構と無関係に、細胞又は組織へのその投与によって特定のタンパク質をコードするmRNAが分解される。したがって、アンチセンスオリゴヌクレオチドは、特定のタンパク質の発現及び/又は活性を減少させる。

【0131】

オリゴヌクレオチドは、DNA、RNA、それらのキメラ混合物、誘導體、又は修飾型(version)であり、一本鎖又は二本鎖であり得る。オリゴヌクレオチドは、例えば、分子の安定性の改善、ハイブリダイゼーション等のために、塩基部分、糖部分又はリン酸骨格で修飾され得る。オリゴヌクレオチドは、ペプチドのような別に添加される群（例えば、宿主細胞受容体を標的とするため）、細胞膜を介した輸送を促進する化合物（例えば、Letsinger他、1989、Proc. Natl. Acad. Sci. U.S.A. 86:6553-6556、Lemaitre他、1987、Proc. Natl. Acad. Sci. 84:648-652、1988年12月15日刊行のPCT公開番号WO88/09810を参照のこと）又は血液脳関門（例えば、1988年4月25日刊行のPCT公開番号WO89/10134を参照のこと）、ハイブリダイゼーション誘導分裂剤（例えば、Krol他、1988、Bio Techniques 6:958-976を参照のこと）又は挿入剤（例えば、Zon、1988、Pharm. Res. 5:539-549を参照のこと）を含み得る。このために、オリゴヌクレオチドは、他の分子と結合し得る。

【0132】

アンチセンスオリゴヌクレオチドは、少なくとも1つの修飾された塩基部分を含み得る。少なくとも1つの修飾された塩基部分は、5-フルオロウラシル、5-プロモウラシル、5-クロロウラシル、5-ヨードウラシル、ヒポキサンチン、キサンチン、4-アセチルシトシン、5-(カルボキシヒドロキシトリエチル)ウラシル、5-カルボキシメチルアミノメチル-2-チオウリジン、5-カルボキシメチルアミノメチルウラシル、ジヒドロウラシル、-D-ガラクトシルケオシン、イノシン、N6-イソペンテニルアデニン

、 1 - メチルグアニン、 1 - メチルイノシン、 2 , 2 - ジメチルグアニン、 2 - メチルアデニン、 2 - メチルグアニン、 3 - メチルシトシン、 5 - メチルシトシン、 N 6 - アデニン、 7 - メチルグアニン、 5 - メチルアミノメチルウラシル、 5 - メトキシアミノメチル - 2 - チオウラシル、 - D - マンノシルケオシン、 5 ' - メトキシカルボキシメチルウラシル、 5 - メトキシウラシル、 2 - メチルチオ - N 6 - イソペンテニルアデニン、 ウラシル - 5 - オキシ酢酸 (v)、ワイプトキソシン、プソイドウラシル、ケオシン、 2 - チオシトシン、 5 - メチル - 2 - チオウラシル、 2 - チオウラシル、 4 - チオウラシル、 5 - メチルウラシル、ウラシル - 5 - オキシ酢酸メチルエステル、ウラシル - 5 - オキシ酢酸 (v)、 5 - メチル - 2 - チオウラシル、 3 - (3 - アミノ - 3 - N - 2 - カルボキシプロピル) ウラシル、 (a c p 3) w 及び 2 , 6 - ジアミノプリンを含む群から選択されるが、これに限定されない。

10

【 0 1 3 3 】

アンチセンスオリゴヌクレオチドはまた、アラビノース、 2 - フルオロアラビノース、キシロース、及びヘキソースが含まれるが、これらに限定されない群から選択される少なくとも 1 つの修飾糖部分を含み得る。アンチセンスオリゴヌクレオチドは、中性ペプチド様骨格も含み得る。このような分子は、ペプチド核酸 (P N A) - オリゴマーと呼ばれ、例えば、Perry-O'Keefe他 (1996) Proc. Natl. Acad. Sci. U.S.A. 93: 14670及びEglom他 (1993) Nature 365: 566に記載されている。P N A オリゴマーの 1 つの利点は、D N A の中性骨格のために培地のイオン強度から本質的に独立して相補 D N A と結合する能力である。さらに別の実施形態では、アンチセンスオリゴヌクレオチドは、ホスホロチオエート、ホスホロジチオエート、ホスホラミドチオエート、ホスホラミデート、ホスホロジアミデート、メチルホスホネート、アルキルホスホトリエステル、及びホルムアセタール、又はこれらのアナログから成る群から選択される少なくとも 1 つの修飾リン酸骨格を含む。

20

【 0 1 3 4 】

なおさらなる実施形態では、アンチセンスオリゴヌクレオチドは、アノマーオリゴヌクレオチドである。アノマーオリゴヌクレオチドは、相補 R N A と特異的な二本鎖ハイブリッドを形成し、通常単位と対照的に、鎖は互いに並行している (Gautier他, 1987, Nucl. Acids Res. 15: 6625-6641)。オリゴヌクレオチドは、 2 ' - O - メチルリボヌクレオチド (Inoue他, 1987, Nucl. Acids Res. 15: 6131-6148)又はキメラ R N A - D N A アナログ (Inoue他, 1987, FEBS Lett. 215: 327-330)である。

30

【 0 1 3 5 】

本発明のオリゴヌクレオチドを、当該技術分野で既知の標準的方法 (例えば、自動化 D N A 合成機 (Biosearch, Applied Biosystemsなどから市販されているものなど)の使用)によって合成することができる。例として、ホスホロチオエートオリゴヌクレオチドを、Stein他 (1988, Nucl. Acids Res. 16: 3209)によって合成することができ、メチルホスホネートオリゴヌクレオチドを、調節された細孔ガラスポリマー支持体等の使用によって調製することができる (Sarin他, 1988, Proc. Natl. Acad. Sci. U.S.A. 85: 7448-7451)。

【 0 1 3 6 】

当業者は、適切なオリゴヌクレオチドを容易に選択することができる。特定のタンパク質をコードする核酸配列を仮定すると、当業者は、このタンパク質に結合するアンチセンスオリゴヌクレオチドをデザインし、*in vitro*又は*in vivo*系でのこれらのオリゴヌクレオチドを試験して、これらが特定のタンパク質をコードする m R N A に結合して分解を媒介することを確認することができる。特定のタンパク質に特定に結合して分解を媒介するアンチセンスオリゴヌクレオチドをデザインするために、オリゴヌクレオチドによって認識される配列は、特定のタンパク質に固有であるか実質的に固有であることが重要である。例えば、タンパク質のいたる所で頻繁に反復する配列は、特定のメッセージを特異的に認識して分解するオリゴヌクレオチドのデザインの理想的な選択ではないかもしれない。当業者は、オリゴヌクレオ

40

50

チドをデザインし、このオリゴヌクレオチドと公的に利用可能なデータベースに登録されている核酸配列とを比較して、配列が特定のタンパク質に特異的又は実質的に特異的であることを確認することができる。

【0137】

別の例では、1つを超えるメッセージに結合し分解を媒介するアンチセンスオリゴヌクレオチドをデザインすることが望ましいかもしれない。1つの例では、メッセージは、イソ型又は機能的に不必要なタンパク質などの関連タンパク質をコードし得る。このような場合、当業者は、これらの関連タンパク質をコードする核酸配列をアラインメントし、両方のメッセージを認識するオリゴヌクレオチドをデザインすることができる。

【0138】

アンチセンスDNA又はRNAを細胞に送達するための多数の方法が開発されており、例えば、アンチセンス分子を、組織部位に直接注射することができるか、所望の細胞を標的化するようにデザインされた修飾アンチセンス分子(例えば、標的細胞表面上に発現した受容体又は抗原に特異的に結合するペプチド又は抗体に結合するアンチセンス)を、全身投与することができる。

【0139】

しかし、一定の例では、内因性mRNAの翻訳の抑制に十分なアンチセンスの細胞内濃度を達成することは困難であり得る。したがって、別のアプローチは、アンチセンスオリゴヌクレオチドが強力なpol IIIプロモーター又はpol IIプロモーターの調節下におかれた組換えDNA構築物を使用する。例えば、ベクターを、細胞によって取り込まれてアンチセンスRNAの転写を指示するように*in vivo*で導入することができる。このようなベクターは、転写されて所望のアンチセンスRNAを産生することができる限り、エピソームを保持するか染色体に組み込まれるようになり得る。このようなベクターを、当該技術分野で標準的な組換えDNAテクノロジー法によって構築することができる。ベクターは、哺乳動物細胞中での複製及び発現のために使用されるプラスミド、ウイルス、又は当該技術分野で既知の他のベクターであり得る。アンチセンスRNAをコードする配列を、哺乳動物、好ましくはヒトの細胞中で作用するための当該技術分野で既知の任意のプロモーターによって発現させることができる。このようなプロモーターは、誘導性又は構成性であり得る。このようなプロモーターには、以下が含まれるが、これらに限定されない: SV40初期プロモーター領域(Bernoist及びChambon, 1981, Nature 290:304-310)、ラウス肉腫ウイルスの3'長末端反復中に含まれるプロモーター(Yamamoto他, 1980, Cell 22: 787-797)、ヘルペスチミジンキナーゼプロモーター(Wagner他, 1981, Proc. Natl. Acad. Sci. U.S.A. 78: 1441-1445)、メタロチオネイン遺伝子の調節配列(Brinster他, 1982, Nature 296: 39-42)など。プラスミド、コスミド、YAC、又はウイルスベクターの任意の型を使用して、組織部位に直接導入することができる組換えDNA構築物を調製することができる。あるいは、所望の組織選択的に感染するウイルスベクターを使用することができ、この場合、別経路で投与することができる(例えば、全身投与)。

【0140】

RNAi構築物は、標的遺伝子の発現を特異的に遮断することができる二本鎖RNAを含む。「RNA干渉」又は「RNAi」は、植物及び蠕虫で認められる現象に最初に適用された用語であり、二本鎖RNA(dsRNA)が特異的且つ転写後様式で遺伝子発現を遮断する。理論に拘束されないが、RNAiは、mRNA分解に関与するようであるが、現在のところ、生化学的機構が活発に研究されている。作用機構に関するいくつかの謎にもかかわらず、RNAiは、有用な*in vitro*又は*in vivo*での遺伝子発現の阻害方法を提供する。

【0141】

本明細書中で使用される、用語「dsRNA」は、siRNA分子又は二本鎖の特徴を有し、且つヘアピンRNA部分などの細胞中でsiRNAにプロセシングされることのできる他のRNAをいう。

【0142】

10

20

30

40

50

用語「機能喪失」は、本発明のRNAi法によって阻害される遺伝子をいう場合、RNAi構築物の非存在下でのレベルと比較した場合の遺伝子発現レベルの減少をいう。

【0143】

本明細書中で使用される、句「RNAiを媒介する」は、RNAのRNAiプロセスによる分解を区別する能力をいう（示す）（例えば、配列独立性dsRNA応答（例えば、PKR応答）よりもむしろ配列特異的様式で分解する）。

【0144】

本明細書中で使用される、用語「RNAi構築物」は、明細書を通して、小干渉RNA（siRNA）、ヘアピンRNA、及びin vivoで分解されてsiRNAを形成することができる他のRNA種を含むように使用される一般名である。本明細書中のRNAi構築物はまた、細胞中でdsRNA又はヘアピンRNAを形成する転写物及び/又はin vivoでsiRNAを産生することができる転写物を生じることができる発現ベクター（RNAi発現ベクターともいう）を含む。

10

【0145】

「RNAi発現ベクター」（本明細書中で、「dsRNAコードプラスミド」ともいう）は、RNAを発現（転写）して細胞中でsiRNA部分を産生し、構築物を発現するために使用される複製可能な核酸構築物をいう。このようなベクターには、（2）転写されて二本鎖RNA（細胞中でアニリングされてsiRNAを形成する2つのRNA部分又はsiRNAにプロセッシングすることができる単一のヘアピンRNA）を産生する「コード」配列及び（3）適切な転写開始配列及び転写終結配列に作動可能に連結された（1）

遺伝子発現において調節の役割を果たす遺伝子エレメント（単数又は複数）（例えば、プロモーター、オペレーター、又はエンハンサー）のアセンブリを含む転写単位が含まれる。プロモーター及び他の調節エレメントの選択は、一般に、意図する宿主細胞によって変化する。一般に、組換えDNA技術で有用な発現ベクターは、しばしば、「プラスミド」の形態であり、これは、環状二本鎖DNAループをいい、そのベクター形態は染色体に結合しない。本明細中で、「プラスミド」及び「ベクター」は、プラスミドが最も一般的に使用されるベクター形態であるので、交換可能に使用される。しかし、本発明は、等価の機能を果たし、且つ当該技術分野でその後既知となるような他の発現ベクター形態を含むことを意図する。

20

【0146】

RNAi構築物は、細胞の生理学的条件下で阻害すべき遺伝子（すなわち、「標的」遺伝子）のmRNA転写物の少なくとも一部のヌクレオチド配列とハイブリッド形成するヌクレオチド配列を含む。二本鎖RNAは、RNAiを媒介する能力を有する天然RNAに十分に類似していることのみが必要である。したがって、本発明は、遺伝子の変異、株の多型、又は進化的相違によって予想され得る配列の変動に耐え得るという利点を有する。標的配列とRNAi構築物の配列との間の耐え得るヌクレオチドのミスマッチ数は、5塩基対中1塩基対、10塩基対中1塩基対、20塩基対中1塩基対、又は50塩基対中1塩基対に過ぎない。siRNA二重鎖の中心のミスマッチは最も重要であり、本質的に、標的RNAを切断できない。対照的に、標的RNAに相補的なsiRNAの3'末端のヌクレオチドは、標的認識の特異性に有意に寄与しない。

30

40

【0147】

配列同一性を、当該技術分野で既知の配列比較及びアラインメントアルゴリズム（Gribkov及びDevereux, Sequence Analysis Primer, Stockton Press, 1991及びその引用文献を参照のこと）及び、例えば、デフォルトパラメーターを使用してBESTFITソフトウェアプログラムにて実施するSmith-Watermanアルゴリズム（例えば、University of Wisconsin Genetic Computing Group）によるヌクレオチド配列間の相違率の計算によって最適化することができる。阻害RNAと標的遺伝子の一部との間の配列同一性が90%を超えるか、さらに100%であることが好ましい。あるいは、RNAの二重鎖領域を、標的遺伝子転写物の一部とハイブリッド形成することができる（例えば、400mM NaCl、40mM PIPES（pH6.4）、1mM EDTA、50又は70で12～

50

16時間のハイブリッド形成、その後の洗浄)ヌクレオチド配列として機能的に定義することができる。

【0148】

化学合成法又は組換え核酸技術によって、RNAi構築物を産生することができる。処理細胞の内因性RNAポリメラーゼは、*in vivo*で転写を媒介することができるか、クローン化RNAポリメラーゼを*in vitro*での転写に使用することができる。RNAi構築物は、例えば、細胞ヌクレアーゼに対する感受性を減少させ、生物学的利用能を改良し、処方物の特徴を改良し、及び/又は他の薬物動態学的性質を変化させるためのリン酸-糖骨格又はヌクレオシドのいずれかに対する修飾を含み得る。例えば、天然RNAのホスホジエステル結合を、少なくとも1つの窒素又は硫黄ヘテロ原子を含むように修飾することができる。RNA構造を修飾して、特定の遺伝子を阻害する一方で、dsRNAに対する一般的応答を回避することができる。同様に、塩基を修飾して、アデノシンデアミナーゼ活性を遮断することができる。RNAi構築物を、酵素又は部分的/全有機合成によって産生することができ、任意の修飾リボヌクレオチドを、*in vitro*での酵素又は有機合成によって導入することができる。

10

【0149】

RNA分子の化学修飾方法を、RNAi構築物の修飾に適用することができる(例えば、Heidenreich他(1997) *Nucleic Acids Res*, 25: 776-780; Wilson他(1994) *J Mol Recog* 7: 89-98; Chen他(1995) *Nucleic Acids Res* 23: 2661-2668; Hirschbein他(1997) *Antisense Nucleic Acid Drug Dev* 7: 55-61を参照のこと)。例示に過ぎないが、RNAi構築物の骨格を、ホスホチオエート、ホスホラミデート、ホスホロジチオエート、キメラメチルホスホネート-ホスホジエステル、ペプチド核酸、5-プロピニル-ピリジン含有オリゴマー、又は糖修飾物(例えば、2'置換リボヌクレオチド、a配座)で修飾することができる。

20

【0150】

二本鎖構造を、単一の自己相補RNA鎖又は2つの相補RNA鎖によって形成することができる。RNA二重鎖形成を、細胞の内側又は外側のいずれかで開始することができる。RNAを、1細胞あたり少なくとも1コピーが送達される量で導入することができる。より高い用量(例えば、1細胞あたり少なくとも5、10、100、500、又は1,000コピー)の二本鎖材料によってより有効に阻害することができる一方で、より低用量でも特定の適用に有用であり得る。阻害は、RNAの二重鎖領域に対応するヌクレオチド配列が遺伝子阻害のために標的化されるという点で、配列特異的である。

30

【0151】

一定の実施形態では、目的のRNAi構築物は、「小干渉RNA」*m*又は「*siRNA*」である。これらの核酸は、約19~30ヌクレオチド長、さらにより好ましくは21~23ヌクレオチド長(例えば、より長い二本鎖RNA「を刻む」ヌクレアーゼによって生成されたフラグメントの長さに対応する)である。*siRNA*は、ヌクレアーゼ複合体を補充し、特定の配列への対合によって複合体を標的*mRNA*に誘導することが理解される。結果として、標的*mRNA*は、タンパク質複合体中のヌクレアーゼによって分解される。一定の実施形態では、21~23ヌクレオチドの*siRNA*分子は、3'ヒドロキシル基を含む。

40

【0152】

本発明の*siRNA*分子を、当業者に既知の多数の技術を使用して得ることができる。例えば、*siRNA*を、当該技術分野で既知の方法を使用して、化学合成するか組換えによって産生することができる。例えば、短いセンス及びアンチセンスRNAオリゴマーを合成し、アニーリングして、各末端に2-ヌクレオチドオーバーハングを有する二本鎖RNA構造を形成することができる(Caplen他(2001) *Proc Natl Acad Sci USA*, 98: 9742-9747; Elbashir他(2001) *EMBO J*, 20: 6877-88)。次いで、下記などのように、これらの二本鎖*siRNA*構造を、受動的取り込み又は最適な送達系のいずれかによって細胞に直接導入することができる。

50

【0153】

一定の実施形態では、*siRNA* 構築物を、例えば、酵素ダイサーの存在下でのより長い二本鎖 *RNA* のプロセッシングによって生成することができる。一実施形態では、ショウジョウバエ *in vitro* 系を使用する。この実施形態では、*dsRNA* を、ショウジョウバエ胚由来の可溶性抽出物と組み合わせ、それにより、組み合わせを生成する。組み合わせを、*dsRNA* を約 21 ~ 約 23 ヌクレオチドの *RNA* 分子にプロセッシングする条件下に維持する。

【0154】

siRNA 分子を、当業者に既知の多数の技術を使用して精製することができる。例えば、ゲル電気泳動を使用して、*siRNA* を精製することができる。あるいは、未変性方法（未変性カラムクロマトグラフィなど）を使用して、*siRNA* を精製することができる。さらに、クロマトグラフィ（例えば、サイズ排除クロマトグラフィ）、グリセロール勾配遠心分離、抗体を使用したアフィニティ精製を使用して *siRNA* を精製することができる。

10

【0155】

一定の好ましい実施形態では、*siRNA* 分子の少なくとも1つの鎖は、約1 ~ 約6 ヌクレオチド長の3'オーバーハングを有するが、2 ~ 4 ヌクレオチド長でもよい。より好ましくは、3'オーバーハングは、1 ~ 3 ヌクレオチド長である。一定の実施形態では、一方の鎖が3'オーバーハングを有し、他方の鎖が平滑末端であるかオーバーハングも有する。オーバーハングの長さは、各鎖で同一であっても異なってもよい。*siRNA* の安定性をさらに増強するために、3'オーバーハングを分解に対して安定化することができる。一実施形態では、アデノシン又はグアノシンヌクレオチドなどのプリンヌクレオチドの含有によって *RNA* を安定化する。あるいは、ピリミジンヌクレオチドの修飾アナログによる置換（例えば、ウリジンヌクレオチド3'オーバーハングの2'-デオキシチミジンによる置換）が許容され、*RNAi* の有効性に影響を与えない。2'ヒドロキシルの非存在により、組織培養培地中でのオーバーハングのヌクレアーゼ耐性が増強され、*in vivo* で有利であり得る。

20

【0156】

他の実施形態では、*RNAi* 構築物は、長い二本鎖 *RNA* の形態である。一定の実施形態では、*RNAi* 構築物は、少なくとも25、50、100、200、300、又は400塩基である。一定の実施形態では、*RNAi* 構築物は、400 ~ 800塩基対である。二本鎖 *RNA* を細胞内で消化して、例えば、細胞中に *siRNA* 配列を産生する。しかし、*in vivo* での長い二本鎖 *RNA* の使用は必ずしも実用的とは限らず、これは、おそらく配列独立性 *dsRNA* 応答によって生じ得る有害な影響による。このような実施形態では、局所送達系及び/又はインターフェロン若しくはPKRの効果を減少させる薬剤の使用が好ましい。

30

【0157】

一定の実施形態では、*RNAi* 構築物は、ヘアピン構造（ヘアピン *RNA* と呼ばれる）の形態である。ヘアピン *RNA* を外因的に合成することができるか、*in vivo* で *RNA* ポリメラーゼIIIプロモーターからの転写によって形成することができる。哺乳動物細胞における遺伝子スプライシングのためのこのようなヘアピン *RNA* の作製及び使用の例は、例えば、Paddison他、Genes Dev, 2002, 16: 948-58; McCaffrey他、Nature, 2002, 418: 38-9; McManus他、RNA, 2002, 8: 842-50; Yu他、Proc Natl Acad Sci USA, 2002, 99: 6047-52に記載されている。好ましくは、このようなヘアピン *RNA* を細胞中又は動物中で操作して、所望の遺伝子の連続的且つ安定な抑制を確実にする。*siRNA* を細胞中でのヘアピン *RNA* のプロセッシングによって産生することができることが当該技術分野で既知である。

40

【0158】

さらに他の実施形態では、プラスミドを使用して、例えば、転写産物として二本酸 *RNA* を送達させる。このような実施形態では、プラスミドを、*RNAi* 構築物の各センス鎖

50

及びアンチセンス鎖の「コード配列」を含むようにデザインする。コード配列は、例えば、逆プロモーターに隣接した同一の配列であり得るか、それぞれ個別のプロモーターの転写調節下の2つの個別の配列であり得る。コード配列の転写後、相補RNA転写物が塩基対合して、二本鎖RNAを形成する。

【0159】

PCT出願W001/77350号は、導入遺伝子を二方向で転写して真核細胞中に同一導入遺伝子のセンス及びアンチセンスRNA転写物を生成するための例示的ベクターを記載している。したがって、一定の実施形態では、本発明は、以下の固有の特徴を有する組換えベクターを提供する：組換えベクターは、反対方向に配置された2つの重複転写単位を有し、且つ目的のRNAi構築物の導入遺伝子に隣接しているウイルスレプリコンを含み、2つの重複転写単位は宿主細胞中の同一の導入遺伝子からセンス及びアンチセンスのRNA転写物を生成する。

10

【0160】

RNAi構築物は、標的核酸配列と同一若しくは実質的に同一の二本鎖RNAの長いストレッチ又は標的核酸配列領域のみと同一若しくは実質的に同一の二本鎖RNAの短いストレッチを含み得る。長RNAi構築物又は短RNAi構築物のいずれかの例示的な作製及び送達方法を、例えば、W001/68836号及びW001/75164号に見出すことができる。

【0161】

mRNA転写物を触媒的に切断するようにデザインしたリボザイム分子を使用して、mRNAの翻訳を防止することもできる（例えば、1990年10月4日公開のPCT国際公開W090/11364号；Sarver他，1990，Science 247：1222-1225及び米国特許第5,093,246号を参照のこと）。部位特異的認識配列でmRNAを切断するリボザイムを使用して特定のmRNAを破壊することができる一方で、ハンマーヘッドリボザイムの使用が好ましい。ハンマーヘッドリボザイムは、隣接領域によって指示される位置でmRNAを切断して、標的mRNAとの相補的塩基対合物を形成する。標的mRNAが以下の2塩基の配列を有することのみが必要である：5'-UG-3'。ハンマーヘッドリボザイムの構築及び産生は当該技術分野で既知であり、Haseloff及びGerlach，1988，Nature，334：585-591により完全に記載されている。

20

【0162】

本発明のリボザイムはまた、RNAエンドリボヌクレアーゼ（以後、「Cech型リボザイム」）（テトラヒメナ中に天然に存在するもの（IVS又はL-19 IVS RNAとして既知）及びThomas Cech及び共同研究者によって広く記載されているもの（Zaug他，1984，Science，224：574-578；Zaug及びCech，1986，Science，231：470-475；Zaug他，1986，Nature，324：429-433；University Patents Inc.の公開国際特許番号W088/04300号；Been及びCech，1986，Cell，47：207-216）など）を含む。Cech型リボザイムは、標的RNA配列とハイブリッド形成し、その後標的RNAが切断される8塩基対の活性部位を有する。本発明は、8塩基対活性部位配列を標的化するCech型リボザイムを含む。

30

【0163】

アンチセンスアプローチなどの場合、リボザイムは、修飾オリゴヌクレオチドから構成され（例えば、安定性の改善、標的化などのため）、*in vitro*又は*in vivo*で細胞に送達することができる。好ましい送達方法は、強力構成性polIII又はpolIIプロモーターの調節下でリボザイムを「コードする」DNA構築物を使用し、その結果、トランスフェクトされた細胞が標的化メッセージを破壊して翻訳を阻害するのに十分な量のリボザイムが産生されることを含む。アンチセンス分子と異なるリボザイムは触媒性であるので、より低い細胞内濃度で有効である。

40

【0164】

抗体を、特定のタンパク質の活性の阻害剤として使用することができる。抗体は、特定のエピトープに対して並外れた親和性及び特異性を有し得る。タンパク質上のエピトープ

50

への抗体の結合がこのタンパク質の機能を妨害し得るような方法で、特定のタンパク質に結合する抗体。例えば、抗体は、適切なタンパク質 - タンパク質相互作用を立体的に妨害して活性部位を塞ぐことによってタンパク質機能を阻害することができる。あるいは、特定のタンパク質上のエピトープへの抗体の結合により、このタンパク質の高次構造が変化し、それにより、もはや適切に機能することができなくなる。

【0165】

モノクローナル抗体又はポリクローナル抗体を、標準的なプロトコールを使用して作製することができる(例えば、Antibodies: A Laboratory Manual Harlow及びLame編(Cold Spring Harbor Press: 1988)を参照のこと)。マウス、ハムスター、ラット、ヤギ、又はウサギなどの哺乳動物を、ペプチドの免疫原性形態で免疫性を与えることができる。タン

10

【0166】

ポリペプチドの抗原性調製物での動物の免疫付与後、抗血清を得ることができ、所望ならば、血清からポリクローナル抗体を単離することができる。モノクローナル抗体を産生するために、抗体産生細胞(リンパ球)を、免疫付与動物から採取し、標準的な体細胞融合手順によって骨髓腫細胞などの不死化細胞と融合してハイブリドーマ細胞を得ることができる。このような技術は、当該技術分野で既知であり、例えば、ハイブリドーマ技術(K

20

【0167】

本明細書中で使用される、用語「抗体」は、特定のポリペプチドとも特異的に反応するそのフラグメントを含むことを意図する。従来技術を使用して抗体を断片化し、フラグメントを、全抗体について上記と同一の様式で有用性についてスクリーニングすることができる。例えば、F(a b)₂フラグメントを、ペプシンでの抗体の処理によって生成す

30

【0168】

特定のポリペプチドに指向するモノクローナル抗体及びポリクローナル抗体(A b)並びに抗体フラグメント(F a b、F(a b)₂、F v、及びs c F vなど)を使用して、特定のタンパク質の作用を遮断することができる。このような抗体を、生物学的過程における特定のタンパク質の役割をさらに理解するための実験の場又は治療の場いずれかで使用することができる。

40

【0169】

ペプチド、ポリペプチド、変異ポリペプチド、及びペプチドフラグメントが候補薬剤となり得る。ポリペプチドの例としては、特定のポリペプチドに、少なくとも60%、70%、75%、80%、85%、90%、95%、98%、99%、又は100%相同であるアミノ酸配列が挙げられる。フラグメントの例としては、全長ポリペプチドの、少なくとも5、6、7、8、9、10、15、20、25、50、75、100、125、150、200、250又は250よりも大きいアミノ酸残基のフラグメントが挙げられる。ペプチド及びポリペプチドは、特定のタンパク質の機能を作動(agonize)、又は拮抗し、それにより細胞表現型を調整することができる。

【0170】

50

有機小分子は、特定のタンパク質の発現及び/又は活性を作動又は拮抗し、それにより細胞表現型を調整することができる。有機小分子は、分子量が2,500 amu未満、より好ましくは1,500 amu未満、さらにより好ましくは750 amu未満の炭素含有分子を意味する。本発明の文脈では、このような有機小分子は、特定の分化細胞型への細胞の分化を促進することができるであろう。

【0171】

小分子を、有機分子及び/又は化合物のライブラリーをスクリーニングし、それにより所望の機能を有する化合物を同定することによって容易に同定することができる。理論に拘束されないが、有機小分子は、多数の方法のいずれかで細胞表現型に影響を与え得る。例として、小分子は、細胞表面受容体に影響を与えるように細胞表面で作用することができる。さらなる例として、小分子は、細胞内で、特定のシグナル伝達経路に沿った細胞内シグナル伝達に影響を与えるように作用することができる。本発明の方法は偏りがなく、その作用機構と無関係に細胞表現型を調整する小分子化合物を同定可能である。

10

【0172】

ペプチド又はポリペプチドである化合物に加えて、本発明は、ペプチド及びポリペプチドをコードするヌクレオチド配列を含む核酸を意図する。本明細書中で使用される、用語「核酸」は、等価物を含むことを意図する。用語「等価物」は、特定のヌクレオチド配列と機能的に同等なヌクレオチド配列が含まれると理解される。同等なヌクレオチド配列には、1つ又は複数のヌクレオチドの置換、付加、又は欠失によって異なる配列(対立遺伝子変異型及び遺伝コードの縮重によるパリエーションなど)が含まれる。同等な配列には、ストリンジェントな条件下で(すなわち、約1Mの塩中で形成されたDNA二重鎖の融点(T_m)より約20~27低いものに等価)所与のヌクレオチド配列とハイブリッド形成するヌクレオチド配列も含まれ得る。ストリンジェントなハイブリッド形成条件のさらなる例には、65の0.2×SSCでの洗浄ステップが含まれる。

20

【0173】

遺伝コードの縮重のために特定のペプチド又はポリペプチド候補薬剤をコードするヌクレオチド配列と異なる配列を有する核酸も本発明の範囲内に含まれる。このような核酸は、機能的に同等なペプチドをコードするが、遺伝コードの縮重のために当該技術分野で既知の野生型配列由来の配列と異なる。例えば、多数のアミノ酸は、1つを超えるトリプレットによって示される。同一アミノ酸を特定するコドン又は同義語(例えば、CAU及びCACはそれぞれヒスチジンをコードする)により、アミノ酸配列に影響を与えない「サイレント」変異を得ることができる。しかし、アミノ酸配列を変化させるDNA配列多型も存在すると予想される。

30

【0174】

生物学的条件には、細胞集団が配置される共通の流動物の任意の生物学的態様が含まれる。生物学的態様には、特に、細胞、ウイルス、小胞、オルガネラ、生体抽出物、及び/又は生体混合物の存在、非存在、濃度、活性、又は型が含まれる。本明細書中に記載のアッセイは、細胞集団セットに対する各ライブラリーメンバーの活性を試験するための条件のライブラリーをスクリーニングすることができる。ライブラリーは、一般に、2つ又はそれ以上の異なるメンバーの収集物(collection)を含む。これらのメンバーは、特に、任意の適切な又は所望の共通の特性に関連する、分子、リガンド、化合物、トランスフェクション材料、受容体、抗体、及び/又は細胞(ファージ、ウイルス、ホールセル、組織、及び/又は細胞抽出物)の形態の化学モジュレーター(又は候補モジュレーター)であり得る。この共通の特徴は、「型」であり得る。したがって、ライブラリーは、特に、2つ又はそれ以上の化合物、2つ又はそれ以上の異なる細胞、2つ又はそれ以上の異なる抗体、2つ又はそれ以上の異なる核酸、2つ又はそれ以上の異なるリガンド、2つ又はそれ以上の異なる受容体、又は2つ又はそれ以上の異なるファージの収集物、又は異なるタンパク質の発現によって区別される細胞集団を含み得る。この共通の特徴はまた、「機能」であり得る。したがって、ライブラリーは、特に、型と無関係である2つ又はそれ以上の結合パートナー(例えば、リガンド及び/又は受容体)、アゴニスト、又はアンタゴニス

40

50

トの収集物を含み得る。

【0175】

ライブラリーメンバーを、任意の適切な機構 (*in vitro*での化学合成、*in vitro*での酵素合成、及び/又は細胞若しくは生物における生合成が含まれる)によって産生及び/又は精製若しくは回収することができる。化学的及び/又は酵素的に合成されたライブラリーには、化合物のライブラリー(合成オリゴヌクレオチド(DNA、RNA、ペプチド核酸、及び/又はこれらの混合物若しくは修飾された誘導体)、小分子(約100Da~10kDa)、ペプチド、炭水化物、脂質など)が含まれ得る。このような化学的及び/又は酵素的に合成されたライブラリーを、各ライブラリーメンバーの方向性のある(*directed*)合成、ライブラリーメンバーセットの組み合わせ合成、及び/又は無作為合成アプローチによって形成することができる。生合成によって産生されたライブラリーメンバーには、特に、プラスミド、相補DNA、ゲノムDNA、RNA、ウイルス、ファージ、細胞、タンパク質、ペプチド、炭水化物、脂質、細胞外基質、細胞溶解物、細胞混合物、及び/又は細胞から分泌された物質などのライブラリーが含まれ得る。ライブラリーメンバーは、単独又は2つ又はそれ以上のメンバーの群/プールとしての細胞集団の接触アレイ(*contact arrays*)であり得る。

10

【0176】

実験

本発明を、ここに一般的に記載し、以下の実施例を参照してより容易に理解されるであろうが、実施例は、本発明の一定の態様及び実施形態の例示のみを目的とし、本発明を制限することを意図せず、当業者は、上記の教示及び以下の実施例から、特許請求の範囲に記載の発明の範囲を逸脱することなく、他のDNAマイクロアレイ、転写調節因子、細胞型、抗体、ChIP条件、又はデータ分析法(全てこれらに制限されない)を使用することができることと認識するであろう。本発明の実施には、適切な場合且つ他で示さない限り、細胞生物学、細胞培養、分子生物学、トランスジェニック生物学、微生物学、ウイルス学、組換えDNA、及び免疫学の従来技術を使用し、これらは当業者の範囲内である。このような技術は、文献に記載されている。例えば、*Molecular Cloning: A Laboratory Manual*, 第3版, Sambrook及びRussell編(Cold Spring Harbor Laboratory Press:2001); *the treatise, Methods In Enzymology* (Academic Press, Inc., N. Y.); *Using Antibodies*, Second Edition by Harlow and Lane, Cold Spring Harbor Press, New York, 1999; *Current Protocols in Cell Biology*, Bonifacino, Dasso, Lippincott-Schwartz, Harford,及びYamada編, John Wiley and Sons, Inc., New York, 1999;及び*PCR Protocols*, Bartlett他編, Humana Press, 2003を参照のこと。

20

30

【0177】

種々の刊行物、特許、及び特許出願が本願を通して引用されており、その内容全体が本明細書中で参照することにより組み込まれる。

【0178】

実験手順

以下の手順は、実験例に従った。

【0179】

遺伝子試薬

203転写調節因子を、既知及び推定転写因子並びに核酸結合タンパク質についてのYPD及びMIPSデータベースの検索によって同定した(Mewes, H. W., Albermann, K., Heumann, K., Liebl, S. & Pfeiffer, F. MIPS: a database for protein sequences, homology data and yeast genome information. *Nucleic Acids Res* 25, 28-30 (1997); Hodges, P. E., McKee, A. H., Davis, B. P., Payne, W. E. & Garrels, J. I. The Yeast Proteome Database (YPD): a model for the organization and presentation of genome-wide functional data. *Nucleic Acids Res* 27,69-73 (1999); Costanzo, M. C. 他. YPD, PombePD and WormPD: model organism volumes of the BioKnowledge library, an integrated resource for protein information. *Nucleic Acids Res* 29,75-9 (2

40

50

001))。反復した Myc エピトープコード配列が調節因子をコードする内因性遺伝子に組み込まれた 203 個の調節因子についてそれぞれ酵母染色を行った。Myc エピトープコード配列及び標的化遺伝子の 5' 末端又は 3' 末端のいずれかと相同な領域に隣接する選択マーカーを含む PCR 構築物を、W303 酵母株 Z1256 に形質転換した。エピトープタグ化タンパク質のゲノム組み込み及び発現を、それぞれ PCR 及びウェスタンブロッティングによって確認した。

【0180】

成長条件

調節因子が特定の環境において成長に不可欠である場合又は特定の環境下での遺伝子発現の調節に調節因子が関連する他の証拠が存在する場合、特定の環境におけるプロファイリングについて調節因子を選択した。 10

【0181】

使用した環境条件の簡単な説明を以下に示す。

【0182】

富化培地。細胞を、YPD (1% 酵母抽出物 / 2% ペプトン / 2% グルコース) 中で OD 600 が約 0.8 になるまで成長させた。

【0183】

高い高酸素。細胞を、YPD 中で OD 600 が約 0.5 になるまで成長させ、その後過酸化水素 (最終濃度 4 mM) で 30 分間処理した。

【0184】

中程度の高酸素。細胞を、YPD 中で OD 600 が約 0.5 になるまで成長させ、その後過酸化水素 (最終濃度 0.4 mM) で 20 分間処理した。 20

【0185】

アミノ酸枯渇。細胞を、合成完全培地で OD 600 が約 0.6 になるまで成長させ、その後アミノ酸生合成の阻害剤であるスルホメツロンメチル (最終濃度 0.2 µg/ml) で 2 時間処理した。

【0186】

栄養素欠乏。細胞を、YPD 中で OD 600 が約 0.8 になるまで成長させ、その後ラパマイシン (最終濃度 100 nM) で 20 分間処理した。

【0187】

線維化誘導。細胞を、1% ブタノールを含む YPD 中で 90 分間又は 14 時間 (約 0.8 の OD 600 に相当する) 成長させた。 30

【0188】

交配誘導。細胞を、YPD 中で OD 600 が約 0.8 になるまで成長させ、その後因子フェロモン (5 µg/ml) で 30 分間処理した。

【0189】

高温。細胞を、30 の YPD 中で OD 600 が約 0.5 になるまで成長させ、その後 37 に温度をシフトして 45 分間成長させた。

【0190】

ガラクトース培地。細胞を、ガラクトース (2%) を補足した YEP 培地中で OD 600 が約 0.8 になるまで成長させた。 40

【0191】

ラフィノース培地。細胞を、ラフィノース (2%) を補足した YEP 培地中で OD 600 が約 0.8 になるまで成長させた。

【0192】

酸性培地。細胞を、YPD 中で OD 600 が約 0.5 になるまで成長させ、その後コハク酸 (最終濃度 0.05 M) で 30 分間処理し、pH 4.0 に到達させた。

【0193】

リン酸欠乏培地。細胞を、リン酸を欠く合成完全培地中で最終 OD 600 が約 0.8 になるまで成長させた。 50

【0194】

ビタミン欠乏培地。細胞を、チアミンを欠く合成完全培地中で最終OD600が約0.8になるまで成長させた。

【0195】

株情報

203の各調節因子について、反復したMycエピトープコード配列が調節因子をコードする内因性遺伝子に組み込まれた株を生成した。Mycエピトープコード配列及び標的化遺伝子の5'末端又は3'末端のいずれかと相同な領域に隣接する選択可能なマーカを含むポリメラーゼ連鎖反応(PCR)構築物を、W303酵母株Z1256に形質転換した。エピトープタグ化タンパク質のゲノム組み込み及び発現を、それぞれPCR及びウエスタンブロッティングによって確認した。

【0196】

ゲノム規模の位置分析

ゲノム規模の位置分析を以前に記載のように行った(Orlando, V. Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. Trends Biochem Sci 25,99-104 (2000); Tessier, D. 他 A DNA Microarrays Fabrication Strategy for Research Laboratories. (Rehm, H. & Reed, G. 編) (Wiley-VCH, Weinheim, Germany, 2002))。結合したタンパク質を、in vivoでDNAにホルムアルデヒド架橋し、その後、細胞溶解し、超音波処理してDNAを切断した。架橋材料を、抗myc抗体で免疫沈降し、その後、架橋を逆反応させてタンパク質からDNAを分離した。免疫沈降したDNA及び非富化サンプル由来のDNAを増幅し、ライゲーション媒介性PCRによって差分的に蛍光標識した。これらのサンプルを、出芽酵母ゲノムの遺伝子間領域を示すスポッティングしたPCR産物から成るマイクロアレイとハイブリッド形成させた。スポットの相対強度を、結合相互作用に確率スコア(P値)を割り当てるエラーモデルの基本として使用した。全マイクロアレイデータは、ArrayExpress(アクセッション番号: E-WMIT-10)及び筆者のウェブサイトから利用可能である。

【0197】

成長環境

出願人は、富化培地中の203個全ての調節因子をプロファイリングした。さらに、出願人は、少なくとも1つの他の環境条件下で84個の調節因子をプロファイリングした。調節因子のリストを、表1に示す。

【0198】

マイクロアレイのデザイン

酵母ゲノム間領域プライマーセット(Research Genetics)を使用して、出願人は、約6,000個のDNAフラグメントを増幅してプリントし、これらは、酵母ゲノム中の本質的に全ての既知の遺伝子間領域を示す(Tessier, D. 他 A DNA Microarrays Fabrication Strategy for Research Laboratories. (Rehm, H. & Reed, G. 編) (Wiley-VCH, Weinheim, Germany, 2002))。スポッティングしたPCR産物の平均サイズは480bpであり、サイズは、60bp~1,500bpの範囲であった。

【0199】

生データの分析

Axon200Bスキャナを使用してマイクロアレイをスキャンし、GenePix5.0を使用して画像を分析した。バックグラウンドに対応するカラムを強度から差し引き、さらなる分析のためにバックグラウンドの標準偏差を抽出した。免疫沈降(試験)及び非富化(コントロール)サンプルを示す2つのチャンネルの強度を、各チャンネルの中央値を使用して正規化因子を計算し、全データセットを単一の中央値の強度に正規化した。試験チャンネルとコントロールチャンネルとの強度の対数比を計算した。免疫沈降反応の偏りを説明するために、これらの対数比を、全アレイにわたる各スポットの平均ログ比を引くことによって各スポットについて正規化した。次いで、試験チャンネルの強度を調整してこの正規化比を得た。最後に、エラーモデル(Hughes他(2000)Cell 102, 109-26)を使用して、

各チップにおける富化有意性を計算し、複製物についてのデータと組み合わせて、各遺伝子間領域についての富化の最終平均比及び有意性を得た。筆者のウェブサイトに記載のように、調節された可能性が最も高い各遺伝子間領域を遺伝子に割り当てた。

【0200】

出願人は、Lee他 Science 298,799-804. (2002)で使用した分析と比較して本発明者らの分析では新規の改良点を含んでいた。明白に、出願人は、分析から人為的スポットを排除し、正規化のためのより信頼できるプローブを選択し、各アレイに質の評価指標を割り当てて低品質の実験を同定した。

【0201】

エラーの評価

出願人は、以前に、 $P = 0.001$ の閾値を満たすゲノム規模の結合データに6~10%の偽陽性を推定していた。本研究は、両方に結合し、結合部位特異性に対して保存された適合を含むDNA領域に注目する($P = 0.001$)。エラー比を決定するためにLee et al. Science 298,799-804. (2002)によって使用し、本発明者らの結合部位基準を満たす47部位のうち、45部位が独立した遺伝子特異的ChIP実験によって確認された。したがって、このデータセット中の偽陽性の頻度は、約4%のようである。

【0202】

偽陰性比は評価がより困難であるが、本ゲノム位置データセットでは約24%のようである。この評価は、 $P = 0.001$ のゲノム規模の位置データで同定されず、且つ保存結合部位(12/50)に関連する細胞周期調節因子について文献に報告されている結合相互作用数の決定に由来する。この調節因子群及びその標的は広範に研究されているので、出願人は、分析のために細胞周期の文献を選択した。

【0203】

調節因子結合特異性

調節因子の推定特異性を、一連のモチーフ発見プログラムを結合データによって同定された遺伝子間配列に適用することによって同定した。得られた特異性の予測を、一定の評価指標を使用して有意性についてフィルタリングし、その後クラスター化して代表的モチーフを得た(図6)。出願人は、以下の6つの方法を使用して、調節因子によって結合した特定の配列を同定した: AlignACE11、MEME13、MDscan12、Kellis 他の方法、及び保存データ、MEMEプログラムを変化させることなく使用するMEME_c及びCONVERGE.MEME_cを組み込む2つのさらなる新規の方法。しかし、上記方法は、狭義の酵母種に保存されない塩基を文字「N」で置換した修飾配列のセットに適用する。CONVERGEは、複数のゲノム由来の配列情報を使用した特異性の発見のための新規の期待値最大化(EM)ベースのアルゴリズムである。狭義の種にわたって同一の部位の検索よりもむしろ、MEME_cの場合のように、CONVERGEは、全てのアラインメントした配列が同一の特異性モデルと一致する遺伝子座を検索する。

【0204】

出願人が使用した各プログラムは、1つ又は複数の統計スコアを使用してその結果の有意性を測定することを試みる。しかし、これらのプログラムの報告が遺伝子間領域の無作為な選択に適用した場合でさえも高スコアでの結果が報告されることが出願人によって認められた。真のモチーフを区別するために、出願人は、実験手順に記載の統計測定セットを選択し、これらのスコアを経験的確率に変換し、類似のスコアを有するモチーフを同一のプログラムによって無作為に選択した配列中に見出すことができた。これらのP値を評価するために、出願人は、無作為に選択した種々のサイズの配列のセットに対して各プログラムを50回実行した。出願人は、これらのスコアによって有意である($P = 0.001$)と判断したモチーフのみを許可した。

【0205】

全プログラム由来の有意なモチーフを互いにプールし、k-medoidsアルゴリズムを使用してクラスター化した。各クラスター内のアラインメントしたモチーフを共に平

10

20

30

40

50

均化してコンセンサスモチーフを産生し、その保存にしたがってフィルタリングした。この手順により、典型的には、各調整因子についていくつかの異なるコンセンサスモチーフが産生された。各調節因子についての1つの特異性を選択するために、出願人は、結果をTRANSFAC 27、YPD 28、及びSCPD 29データベース中の情報と比較した。以前の情報を利用できない場合、出願人は、最も有意な統計スコアを使用して特異性を選択する。

【0206】

モチーフ発見の概要

以下に詳述し、図6にまとめた5ステップのプロセスで結合モチーフを同定した。第1に、一連のモチーフ発見プログラムを結合データによって同定された遺伝子間配列に適用することによってモチーフを発見した。得られた特異性の予測を、一定の評価指標を使用して有意性についてフィルタリングし、クラスター化して代表的モチーフを得た。保存ベースの評価指標を使用して、これらのモチーフの最も信頼性の高いサブセットを同定した。因子について複数の有意な結合モチーフが見出される場合について、出願人は、Transfac (Matys, V. 他 TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* 31,374-8 (2003))、YPD (Csank, C. 他 Three yeast proteome databases: YPD, PombePD, and CalPD (MycoPathPD). *Methods Enzymol* 350, 347-73 (2002))、及びSCPD (Zhu, J. & Zhang, M. Q. SCPD: a promoter database of the yeast *Saccharomyces cerevisiae*. *Bioinformatics* 15,607-11 (1999))データベース由来の統計的スコア又は情報を使用して各調節因子の単一モチーフを選択した。配列入力ファイル、中間体モチーフ発見出力、及び完成したモチーフの行列表示は、筆者のウェブサイトで利用可能である。

【0207】

ステップ1：初期モチーフ発見

モチーフ発見プログラムは、発見特異性に関する強度が異なる。分析をできるだけ総括的にするために、出願人は、結合データに対して異なる5つのモチーフ発見プログラムを提供した：AlignACE (Roth, F. P. , Hughes, J. D. , Estep, P. W. & Church, G. M. Finding DNA regulatory motifs within aligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nat Biotechnol* 16, 939-45 (1998))、MEME (Bailey, T. L. & Elkan, C. The value of prior knowledge in discovering motifs with MEME. *Proc Int Conf Intell Syst Mol Biol* 3, 21-9 (1995))、MDscan (Liu, X. S. , Brutlag, D. L. & Liu, J. S. An algorithm for finding protein-DNA-binding sites with applications to chromatin-immunoprecipitation microarray experiments. *Nat Biotechnol* 20,835-9 (2002))、Kellis他に記載の保存ベースの方法 (Kellis, M. , Patterson, N. , Endrizzi, M. , Birren, B. & Lander, E. S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 423,241-54 (2003))、CONVERGEと呼ばれる新規の保存ベースの方法 (下記)。MEMEプログラムを使用して、保存情報が組み込まれた修正入力も分析した (「プローブ配列」を参照のこと)。

【0208】

より完全に検索するために、これらの各プログラムを、異なるパラメータを使用して複数回実施した。デフォルト設定を使用し、異なる乱数シードで、AlignACEを10回実行し、サンプリングされるモチーフスペースを増大させた。AlignACEの実施によって得られた結果を、分析のためにグループ分けした。次数5のMarkovバックグラウンドモデル、「ZOOFS」モチーフモデル、及び「-minsites20-DNA-revcomp」オプションを使用して、MEMEを実行した。7~11及び12~18のモチーフ幅範囲を使用して、MEMEを繰り返し実行した。MDscanを実行するために、結合のP値に従って配列を順位づけし、「-s30 -r5 -t10」オプションを使用してプログラムを実行した。MDscanが固定された幅のモチーフのみを検索するという事実を補うために、1回の各幅範囲が8~15塩基でプログラムを繰り返す

10

20

30

40

50

返し実行した。記載のように、Kellis 他の方法をデータに適用した (Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 423,241-54 (2003))。8 及び 15 のモチーフ幅を使用して、CONVERGE を 2 回実行した。

【0209】

MEME_c

出願人は、狭義の酵母種中の各塩基保存を伝達するために入力配列を修飾することによって、ALIGNACE、MEME、及び MDscan の性能を改良することができるかどうか試験した。狭義の種 (Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 423, 241-54 (2003)) についての ClustalW (Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22,4673-80 (1994)) アラインメントを使用して、他のゲノムの 2/3 又は 3/4 が保存されていない場合、出願人は、酵母属中の遺伝子を文字「N」に置換した。試験したプログラムのうち、MEME のみが修飾配列を使用することができた。

【0210】

CONVERGE

出願人は、入力配列のセット中に過剰表示され、且つ複数のゲノムにわたって保存されているモチーフを同定するために CONVERGE をデザインした。CONVERGE 入力配列は、一次ゲノムに対応する非ギャップ挿入 DNA 配列及びギャップを含み得る 1 つ又は複数の選択的にアラインメントされた配列から成る。アルゴリズムは、MEME の ZOOFS モデルに基づき、次数 5 の Markov バックグラウンドモデルを使用する。しかし、MEME が入力配列のセットにわたってモチーフモデルとの適合を検索するのに対して、CONVERGE は複数の配列アラインメントにわたって各配列を検索する。詳細には、CONVERGE は、アラインメントした各配列中の同一部位で生じたモチーフの確率の産物として、アラインメント中の部位で生じるモチーフの確率を処理する。したがって、CONVERGE は、モチーフの発見による柔軟な様式で保存された部位を定義する。全詳細を、本明細書中に示す。

【0211】

プローブ配列

モチーフ発見プログラムを、 $P = 0.001$ の P 値で結合したプローブの配列に適用した。出願人は、いくつかの遺伝子間領域がその全長にわたって高度に相同であることを見出し、その結果、全サブシーケンスが過剰表示されるので、モチーフ発見の結果が非対称になる。この偏りを除去するために、BLAST (Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* 215, 403-10 (1990)) を使用して、その長さの 50% を超えて配列が高度に類似するプローブ対を同定した。各対について、より短い遺伝子間領域を、モチーフ発見計算から省く。このプロセスにより、いくつかの実験では 9 つまでの領域が除去されたが、平均して 1 つ未満であった。

【0212】

マイクロアレイ上に存在する配列を決定するために、出願人は、アレイを構築するために使用した PCR の予想産物を計算した。Research Genetics のプライマー配列を、<http://www.resgen.com/products/YeIRP.php3> から入手し、2002 年 3 月改訂の酵母ゲノムを、SGD から入手した (Dwight, S. S. et al. *Nucleic Acids Res* 30, 69-72 (2002))。2 つを超える異なるゲノム配列を増幅すると予想されるプローブを、計算から省いた。反復非転写フィーチャーに隣接する 25 個のプローブ配列 (例えば、テロメア反復、X エlement、及び Y' エlement) も省略した。

【0213】

P S S M 表示

その後の分析のために、全プログラム由来のモチーフを、標準的な位置特異的スコア行列 (P S S M) に変換した。A l i g n A C E 及び M D s c a n によって結合部位をアラインメントし、これらを最初にアラインメントの各位置の各塩基 (A、C、G、T) の頻度を示す行列に変換した。K e l l i s 他の方法は、不確定 (ambiguity) コードを含む文字列としてモチーフを表示し、これも頻度の行列に変換した。(例えば、モチーフが特定の位置に文字「S」を含む場合、値 0.5 を「C」及び「G」の両方に割り当てる。) 塩基頻度の行列を確率に変換し、次数 0 のバックグラウンド確率に比例して 0.001 の疑似計数を使用して調整した (A 及び T で 3.1×10^{-4} の疑似計数、G 及び C で 1.9×10^{-4} の疑似計数)。対数 - 尤度スコアを、評価した確率を各文字についてのバックグラウンド確率で割り、二進法で計算した。C O N V E R G E 及び M E M E の両方により、確率行列が得られ、これを直接使用した。

【0214】

ステップ 2 : モチーフスコアリング及び有意性試験

出願人は、結合プローブ及び非結合プローブで見出される頻度の比較によって、各モチーフの有意性を試験した。モチーフ過剰表示の計算に対する異なるアプローチをカプセル化するために、出願人は、以下の 3 つの異なる評価指標を使用した: 富化、R O C A U C、及び K e l l i s 他に記載の方法 (「C C 4 スコア」) によって発見されたモチーフ。富化スコアは、可能な全遺伝子標的と比較した結合プローブの間のモチーフの発生の直接測定であるが、各遺伝子間領域内のモチーフ発生数を区別しない。R O C A U C 評価指標は、モチーフ発生数が区別される因子である場合により感度が高い。最後に、C C 4 評価指標により、結合プローブ間のモチーフ保存の重要性を説明する方法が得られる。これらのスコアを、以下の「有意性閾値」に記載の遺伝子間領域の無作為な選択の計算から得た有意性閾値と比較した。

【0215】

富化スコア

富化スコアを得るために、超幾何学的分布を使用して、結合プローブ中のモチーフの頻度を、遺伝子間領域をゲノムから無作為に選択した場合に予想される頻度と比較した。配列は、配列が最大可能行列スコアの少なくとも 70% とスコアリングされた少なくとも 1 つ又は複数の部位を含む場合、モチーフを含むと見なした。富化についての P 値を、式:

【0216】

【数 1】

$$p = \sum_{i=b}^{\min(B,g)} \frac{\binom{B}{i} \binom{G-B}{g-i}}{\binom{G}{g}}$$

【0217】

(式中、B は結合した遺伝子間領域数であり、G はマイクロアレイ (又はゲノム) 上に示された遺伝子間領域の総数である) にしたがって計算した。数量 b 及び g は、モチーフに適合した B 及び G の遺伝子間領域数を示す。数量 $-\log_{10}(p)$ を、富化スコアという。

【0218】

R O C A U C (R e c e i v e r O p e r a t i n g C h a r a c t e r i s t i c A r e a U n d e r C u r v e)

R O C A U C は、含まれるモチーフ適合数にしたがって結合プローブセット及び非結合プローブセットの順序付け及び互いに対する部分的順序付けによって、アセンブリされた受信者動作特性曲線下領域をいう。出願人は、Clarke and Granek (Clarke, N. D. & Granek, J. A. Rank order metrics for quantifying the association of sequence feat

ures with gene regulation. *Bioinformatics* 19,212-8 (2003))に記載の方法及びコードを使用した。

【0219】

保存 C C 4

Kellis他の方法(Kellis, M. , Patterson, N. , Endrizzi, M. , Birren, B. & Lander, E. S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 423,241-54 (2003))を使用して発見したモチーフを、C C 4 評価指標にしたがって判断し、このC C 4 評価指標は、結合プローブ間の保存モチーフの発生を、同一の結合プローブセットの間の全 3 - g a p - 3 モチーフで認められる予想率と比較する。認められた比の 2 項確率を計算し、同等な z スコアに関して報告する。

10

【0220】

有意性閾値

遺伝子間領域の無作為な選択に適用した場合でさえも、モチーフ発見プログラムは、高度な過剰表示評価指標(「富化」及び「R O C A U C」など)を使用してモチーフを産生することを出願人は認めた。真のモチーフを同定するために、出願人は、各評価指標由来のスコアを、経験的確率に変換し、類似のスコアを有するモチーフを同一のプログラムによって無作為に選択した配列中に見出すことができた。出願人は、0 . 0 0 1 以下の P 値を有するモチーフのみを許可した。経験的に、既知の特異性を有する多数の調節因子の正確なモチーフが同定されることが認められた。これらの閾値を評価するために、出願人は、1 0、2 0、3 0、4 0、5 0、6 0、7 0、8 0、1 0 0、1 2 0、1 4 0、及び 1 6 0 プローブセットに対して無作為に選択した配列について各プログラムを 5 0 回実行した。

20

【0221】

これらの無作為な実行由来の認められたスコアを、正規分布によってパラメーター化した。各プログラム及び各評価指標についての 0 . 0 0 1 の P 値と同等な臨界値を、表 8 に示す。経験的分布が正規分布でなかった場合(シャピロ-ウルクス検定による)、対応する評価指標を使用しないで、類似の結合プローブ数を使用した調整因子の関連プログラムによって生成されたモチーフを評価した。

【0222】

特定の実験のために、出願人は、結合プローブ配列数に最も近いサイズを有する無作為セット由来の閾値を使用した。例えば、3 2 の遺伝子間配列に対する 1 0 回の A l i g n A C E の実行によって見出されたモチーフを仮定すると、富化スコアは 2 5 であった。3 0 の遺伝子間配列のセットをそれぞれ無作為に選択した 5 0 セットに対する 1 0 回の A l i g n A C E の実行によって関連するスコア分布が得られた。得られた富化スコアの分布の平均は 1 4 . 1 であり、標準偏差は 2 . 1 であり、したがって、P 0 . 0 0 1 の有意性に対応する富化は 2 0 . 4 3 である。候補モチーフのスコアがより高いので、有意と見なされる。

30

【0223】

ステップ 3 : モチーフクラスター化及び平均化

K - m e d o i d s クラスター化

40

次いで、各実験の有意なモチーフセットを、下記の距離評価指標を使用した k - m e d o i d s クラスター化(Hastie他 *The elements of Statistical Learning ; Data mining, inference and prediction* (Springer-Verlag, New York, 2001))によってクラスター化した。k - m e d o i d s アルゴリズムを 5 0 0 回実施して、クラスター間距離の最小和を有するクラスター化を見出した。最適クラスター数を見出すために、このプロセスを、1 0 クラスターを使用して最初に実施し、その後、クラスターのメンバーと他のクラスターの m e d i o i d との間全平均距離が十分に長くなるまで(0 . 1 8 以上)、クラスターを徐々に減少させて繰り返した。

【0224】

モチーフ間距離

50

出願人は、モチーフの比較を補助するための距離評価指標を構築した。2つのアラインメントしたモチーフ「a」及び「b」の間の距離Dを、

【0225】

【数2】

$$D(a,b) = \frac{1}{w} \sum_{i=1}^w \frac{1}{\sqrt{2}} \sum_{L \in \{ACGT\}} (a_{i,L} - b_{i,L})^2$$

【0226】

(式中、wはモチーフ幅であり、 $a_{i,L}$ 及び $b_{i,L}$ はそれぞれ、モチーフa及びbの位置iで認められる塩基Lの評価された確率である)と定義する。w及び2による正規化により、部分距離としての解釈が容易になる。例えば、0.20の距離は、2つのモチーフが約20%異なることを示す。

【0227】

特に、モチーフの最適なアラインメントは知られていない。したがって、出願人は、全アラインメントの中のモチーフ間の最短距離を使用し、アラインメントは、モチーフの少なくとも7つの塩基が重複しているか、モチーフがより短い場合、最も短いモチーフの長さよりも2塩基少ない。モチーフの逆相補物とのアラインメントを含む。

【0228】

モチーフ平均化

1つのクラスターを示す各モチーフを、クラスターを含むアラインメントしたモチーフの各行列の位置での確率の平均化によって計算した。平均化モチーフの側面の情報に乏しい位置(low-information positions)を除去した。

【0229】

ステップ4：平均化モチーフの保存試験

出願人は、平均化モチーフの保存を試験し、その後の分析では以下の2つの保存基準を満たすモチーフに注目した。第1に、モチーフの全てのインスタンス(instance)と比較したモチーフの保存されたインスタンスの頻度が、少なくとも全遺伝子間領域と同様に結合した遺伝子間領域内の高さであることが必要であった。第2に、モチーフは少なくとも3つの結合された保存インスタンスを有することが必要であった。

【0230】

出願人は、最大モチーフの少なくとも60%のスコアを有する場合、配列がモチーフに適合すると見なした。出願人は、「保存インスタンス」は、少なくとも2つの他の狭義の種のアラインメントした配列もモチーフと適合することを意味すると定義した。1つ以下のアラインメントされた配列が利用可能な場合、部位を、「保存されていない」として処理した。

【0231】

ステップ5：1つのモチーフの各調節因子への割り当て

しばしば、モチーフ発見プロセスにより、いくつかの有意な異なる平均モチーフ(3つの平均)が得られた。これらのモチーフは、タンパク質の所望の結合特異性を示すことができるか、これらは、結合パートナーの特異性から生じ得るか、他の生物学的有意性を有し得る。プロファイリングした転写因子の結合特異性を示すモチーフを同定するために、出願人は、利用可能ならば、クラスター化のために使用した同一のモチーフ間距離評価指標を使用して、Transfac(Matys, V. 他 TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic Acids Res 31,374-8 (2003)), YPD(Csank, C. 他 Three yeast proteome databases: YPD, PombePD, and CalPD (MycopathPD). Methods Enzymol 350, 347-73 (2002)), 及びSCP(D Zhu, J. & Zhang, M. Q. SCPD: a promoter database of the yeast Saccharomyces cerevisiae. Bioinformatics 15,607-11 (1999))データベース中の結合データを特異性と比較した(上記)。このようなデータを利用できない調節因子が21個存在した。これらの場合、出願人は、最良の富化スコアを有

するモチーフを選択する。

【0232】

これらのデータベース由来の特異性データは、生配列、不確定性コード、及び行列の形態で時々利用可能である。行列を使用しない調節因子について、出願人は、実験的に決定した特異性情報の正文 (body) を示すために1つのコンセンサス配列をアセンブリし、これを上記のようにPSSMに変換した。データベースからアセンブリしたモチーフの質を独立して評価する方法が存在しないので、出願人は、許容閾値 (permissive threshold) を使用して、発見したモチーフとデータベースのモチーフとの間の類似性を検出した。0.24未満にスコアリングされたモチーフを適合として許可する一方で、0.35未満のスコアを有するモチーフを手動で試験した。調節コードマップ (Regulatory Code Map) 中で使用されたモチーフのスコアを、表2に示す。

10

【0233】

文献由来のモチーフ

出願人は、(1)有効なモチーフ発見のために結合した遺伝子間領域が非常に少なかったか(10未満)、(2)文献に類似の発見されたモチーフがステップ4の保存によって消失したか、又は(3)発見されたモチーフがステップ5中の文献と適合しなかったかのいずれかである、残存調節因子についてのデータベース由来のモチーフを使用した。これらのモチーフが少なくとも1つの結合した保存インスタンスを有する場合のみ、これらのモチーフをインキュベートした。得られた102モチーフの概要(表3)を、その後の全ての分析で使用した。

20

【0234】

調節コードマップ

102個の調節因子の結合モチーフ(表3)を、位置分析データ及び保存データと融合して、遺伝子間領域中の活性結合部位のマップを作成した。全マップは、<http://web.wi.mit.edu/fraenkel/regulatory#map/>で利用可能である。対応する因子によって結合した遺伝子間領域内の各モチーフの保存された全ての発生を見出すことによってマップを構築した。

【0235】

出願人は、P 0.001の結合P値の閾値及び上記「保存試験」の項中に記載の保存の定義を使用した。異なる結合閾値及び保存閾値を使用して構築したマップの異形もオンラインで利用可能である。

30

【0236】

読み取り枠の開始コドン(ATG)から隣接する上流領域中の結合部位までの距離の分布は、上記データに由来した。これらを、10,000個の「無作為化」ゲノムに対して計算した分布と比較し、このゲノムは、各遺伝子間領域中の結合部位を隣接遺伝子間の無作為に独立して再分布していた。100~500の領域(図2Cの灰色の領域)は、予想よりも多数の結合部位を含む。

【0237】

プロモーターの分類

プロモーターを、全実験で集まった結合データに基づいて分類した。プロモーターを、1つを超える調節因子が集合データで結合した場合、任意の特定の条件下で結合した調節因子数と無関係に、複数の調節因子構造を有すると定義した。同様に、プロモーターが集合データの1つの調節因子によって正確に結合した場合、プロモーターを1つの調節因子構造に割り当てた。

40

【0238】

反復モチーフ構造を使用する傾向のある調節因子を、カイ二乗分析によって同定した。各調節因子について、出願人は、1つの部位を含むプロモーター数及び複数の部位を含むプロモーター数を計算した。次いで、これらの値を、全因子の平均に基づいた期待値と比較した。

【0239】

50

同時出現調節モチーフを、超幾何学的分布に基づいた2つの調節因子の結合が独立している帰無仮説下で両調節因子によって結合した認められた遺伝子間領域数(又はそれ以上)を発見する確率を示すP値に基づいて決定した。

【0240】

調節因子の挙動分類

各調節因子の結合を、ペアワイズ様式で、調節因子が研究された各環境条件下について比較した。P = 0.001で結合し、且つ対応するモチーフに対する適合が保存された領域のみを、本分析に含めた。いくつかの調節因子を、正確に比較した条件によって複数のカテゴリーに分類する。「不変条件」カテゴリーについて、調節因子に対する結合プローブの重複比は0.66を超え、結合プローブ数の比は0.66と1.5との間であった。「可能な条件」カテゴリーについて、ある環境下で調節因子はプローブに結合しなかった。「拡大条件」カテゴリーは、調節因子に対する結合プローブの重複比0.66を超え、結合プローブ数の比は0.66未満であるか1.5を超えた。「変化する条件」カテゴリーは、両環境下で調節因子は少なくとも1つのプローブに結合し、結合プローブの重複比は0.66未満であった。

10

【0241】

予想特異性の実験による確認

出願人は、自動化方法を使用して発見したモチーフと文献中のモチーフとを比較し、矛盾が最も大きい調節因子(Cin5)を選択した(表2)。発見したモチーフ、TTAcrtAAは、以前に報告された部位と比較して1つ塩基が挿入されている(Fernandes, L., Rodrigues Pousada, C. & Struhl, K. Yap, a novel family of eight bZIP proteins in *Saccharomyces cerevisiae* with distinct biological functions. *Mol Cell Biol* 17, 6982-93 (1997)) (TTACTAA)。以前に既知の部位は、Cin5によって結合されたプローブ中にあまり豊富でない一方で(P = 0.02)、発見したモチーフは、非常に豊富である(P = 10^{-38.4})。

20

【0242】

出願人は、*in vivo*データから推測したCin5の特異性もこの調節因子の*in vitro*特性を示すかどうかを試験するためにゲルシフトアッセイを使用した(図7)。Cin5のDNA結合ドメインを、チオレドキシン及びポリヒスチジンペプチドに融合したpET-32ベクター(Novagen)の誘導体にクローン化し、大腸菌中で発現させ、アフィニティークロマトグラフィによって精製した。タンパク質を、配列gcgacaTTACCTAAaggcを含むCy5標識オリゴヌクレオチドとインキュベートし、同一の配列又は以前に公開された結合部位(gcgacaTTACTAAaggc)(Fernandes他 *Mol Cell Biol* 17, 6982-93(1997))のいずれかを含む非標識競合物で攻撃誘発した。反応物を、0.5 × TBEで泳動した10%アクリルアミドゲルで分析した。TTACGTAAのコア配列を含むプローブについて類似の結果が得られた。

30

【0243】

調節コード

潜在的な結合部位は、これらが2つの基準を満たす場合、調節コードマップ中に含まれた。第1に、遺伝子座は、出芽酵母ゲノム及び少なくとも2つの他の狭義の酵母ゲノム中の調節因子について最大可能スコアの60%で、特異性モデルと適合しなければならなかった。第2に、遺伝子座は、任意の条件下で対応する調節因子によって結合するプローブも含むゲノム間領域中に存在しなければならなかった(P = 0.001)。プロモーター構造及び環境特異的結合の全分析は、このマップに基づき、以下にさらに記載している。以下の実施例に記載の全方法に関するより詳細な情報を、<http://web.wi.mit.edu/young/regulatory#code>に見出すことができる。

40

【実施例1】

【0244】

実施例1：生物活性DNA結合部位の同定

出願人は、ゲノム規模の位置分析(Iyer et al. *Nature* 409,533-8. (2001), Ren et

50

al. Science 290,2306-9. (2000), Lee et al. Science 298,799-804. (2002), Lieb et al. Nat Genet 28, 327-34 (2001)) を使用して、富化培地条件における 203 個の DNA 結合転写調節因子のゲノム占有及び他の 12 の環境条件の少なくとも 1 つにおけるこれらの調節因子のうち 84 個のゲノム占有を決定した (表 1、図 5、<http://web.wi.mit.edu/young/regulatory#code>)。これらの 203 個のタンパク質は、酵母ゲノム中にコードされるほぼ全ての DNA 結合転写調節因子を含む可能性が高い。調節因子がさらなる環境での成長に不可欠である場合、又はさらなる環境下での遺伝子発現の調節に関与する他の証拠が存在する場合、さらなる環境中でのプロファイリングのために調節因子を選択した。ゲノム規模の位置データにより、調節因子とプロモーター領域との間で 11,000 個の固有の相互作用が高い信頼性で同定された ($P = 0.001$)。 10

【0245】

転写調節因子の認識部位として機能する可能性が高いシス調節配列を同定するために、出願人は、ゲノム規模の位置データ由来の情報、系統発生的に保存された配列、及び先行の知識を融合した (図 1A)。出願人は、6 つのモチーフ発見方法 11~13 を使用して、10 個を超えるプローブに結合する 147 個の調節因子についての 68,279 個の DNA 配列モチーフを発見した (実験手順を参照のこと; 図 6)。これらのモチーフから、出願人は、クラスター化及びストリンジェントな統計的検定によって各調節因子の最も可能性が高い特異性を誘導した。このモチーフ発見プロセスにより、116 個の各調節因子の有意性の高い ($P = 0.001$) モチーフが同定された。出願人は、4 種の関連酵母種のうちの 3 種にわたって保存されている必要があるさらなる基準を使用してこれらの調節因子の 65 個について 1 つの信頼性の高いモチーフを決定した。新規及び「再発見した」モチーフの例を、図 1B に示し、発見したモチーフと以前に記載のモチーフとの比較を、表 2 に示す。発見したモチーフにより、以前に利用可能であった情報よりも有意により多くの情報が得られる、21 個の調節因子については、文献中に先行の特異性情報は存在せず、出願人がモチーフを報告した 17 個の調節因子のみについての詳細な確率行列は、以前に決定されていた (Knuppel et al. J Comput Biol 1, 191-8 (1994))。計算由来のモチーフ (TTACRTAA) と以前に報告した部位 (TTACTAA、表 2) との間で最も大きな相違を示した *Cin5* の場合、出願人は、出願人が報告したモチーフも好ましい *in vitro* 標的であることを見出した (図 7)。出願人は、発見したモチーフに保存試験も合格した文献由来のさらなるモチーフを補足し、出願人は、その後の全ての分析中の 102 個の調節因子についての配列モチーフの概要を使用した。 20

【実施例 2】

【0246】

実施例 2：転写調節コードの構築

出願人は、酵母ゲノム配列上にモチーフをマッピングすることによって転写調節コードの最初のバージョンを構築し、このモチーフは、調節因子によって高い信頼性 ($P = 0.001$) で結合し、且つ狭義の出芽酵母種の間で保存されている (図 2、<http://web.wi.mit.edu/fraenkel/regulatory#map>)。このマップは、1,296 個のプロモーター領域内に 3,353 の相互作用を含む。より信頼性の低い情報を使用して構築した多数のプロモーターを含む調節部位のマップを、筆者のウェブサイトで閲覧することもできる。マップを構築するために使用される情報が複数の成長環境由来の結合データを含み得るので、マップは、ゲノム内の転写調節可能性を記載している。出願人が以下により詳細に記載するように、任意の一環境下での成長の際、マップ中で同定された結合部位のサブセットのみが転写調節因子を占める。 40

【0247】

特定の転写調節因子の機能が以前に確立されている場合、調節マップ中に結合する遺伝子の機能は、この先行の情報と高度に一致する。例えば、アミノ酸生合成調節因子である *Gcn4* 及び *Leu3* は、*BAP2* (染色体 II) のプロモーター中の部位に結合し、アミノ酸輸送体をコードする (図 2A)。6 つの十分に研究されている細胞周期転写調節因子は、*YHP1* (染色体 IV) のプロモーターに結合し、細胞周期の G1 期の調節に関連 50

している。呼吸 H a p 5 の調節因子は、C O X 4 (染色体 V I I) の上流に結合し、呼吸電子輸送鎖の成分をコードする。確立された機能を有する調節因子が未知の機能の遺伝子に結合する場合、これらの標的遺伝子は、このような機能的プロセスに新規に関与する。

【 0 2 4 8 】

調節因子結合データと配列保存データとの組み合わせの有用性を、図 2 B に示す。遺伝子 B A P 2 の 8 8 4 塩基対遺伝子間領域上流内ので起こる本研究に記載の調節因子 D N A 結合特異性に適合する全配列 (表 2) を、上のパネルに示す。したがって、複数の酵母種で保存されているこれらの配列のサブセットは、調節因子相互作用の候補である可能性が高く、中央のパネルに示す。これらの保存調節部位の存在は、この配列による調節の可能性を示すが、部位がいくつかの成長条件下で調節因子によって実際に結合されるかどうかを示さない。結合情報の組み込み (下のパネル) により、試験した条件下で成長した細胞中の調節因子によって利用される保存配列を同定する。

10

【 0 2 4 9 】

転写調節因子の結合部位の分布により、酵母プロモーター中のこれらの部位が組織化されることが明らかとなる (図 2 C)。結合部位は、プロモーター領域にわたり均一に分布していないが、むしろ、鋭く尖った分布を示す。タンパク質コード配列の 1 0 0 塩基対 (b p) 上流の領域中に結合部位はほとんど存在しない。この領域は、典型的には、転写開始部位を含み、転写開始装置によって結合される。転写調節因子結合部位の大部分 (7 4 %) は、タンパク質コード配列の 1 0 0 b p 上流と 5 0 0 b p 上流との間に存在し、無作為で予想されるよりもはるかに多い (5 3 %)。5 0 0 b p より離れた領域は、無作為で

20

【 0 2 5 0 】

出願人は、D N A 結合部位配列の特定の配置はプロモーター内で起こり、これらのプロモーターの構造により調節機構を示す手がかりが得られることが示唆されることに留意する (図 3)。例えば、単一の調節因子についての D N A 結合部位の存在は、最も単純なプロモーター構造であり、予想されるように、出願人は、この特徴内の遺伝子セットはしばしば共通の生物機能に関与することを見出した (表 4)。第 2 のプロモーター構造型は、特定の結合部位配列の反復から成る。反復結合部位は、調節因子 D a l 8 0 による安定な結合に必要であることが示されている (Cunningham他 J Bacteriol 175, 5851-61 (1993))。この反復プロモーター構造により、H I S 4 遺伝子について認められたように、段階的転写応答も可能である (Donahue他 Cell 32, 89-98 (1983))。多数の調節因子 (D i g 1、M b p 1、及び S w i 6 が含まれる) は、反復モチーフについての実質的に有意な優先を示す (表 5)。第 3 のプロモータークラスは、複数の異なる調節因子の結合部位を含む。このプロモーター配置は、遺伝子を組み合わせ調節に供することができることを意味し、出願人は、多くの場合、種々の調節因子を使用して変化する成長条件に対する異なる応答を実行することができるかと予想する。実際、出願人は、このカテゴリー中の多くの遺伝子は複数の代謝経路に必要であり、且つ環境特異的様式で調節される産物をコードすることに留意する。第 4 のプロモーター構造型では、出願人は、ここで、特定の調節因子対の結合部位が偶然によって同一のプロモーター領域内で予想されるよりも頻繁に起こると考察する (表 6)。この「同時発生」モチーフ構造は、2 つの調節因子が物理的に相互作用するか、複数のゲノムで機能を共有することを意味する。

30

40

【 実施例 3 】

【 0 2 5 1 】

実施例 3：生物活性 D N A 結合部位の分布を変化させる条件の同定

複数の細胞成長条件下でのいくつかの調節因子についてのゲノム規模の結合実験の実施により、出願人は、調節配列のサブセットへの調節因子の結合が、細胞の環境条件に高く依存することを確認する (図 8)。出願人は、4 つの調節因子結合挙動の共通パターンを認めた (図 4、表 7)。それぞれの 4 つの群中の十分に研究された調節因子によって使用される調節機構に関する事前の情報により、他の調節因子の環境依存性結合挙動を説明す

50

るための仮説が示唆される。

【0252】

「不変条件」調節因子は、2つの異なる成長環境中で本質的に同一のプロモーターセット（ノイズの限度内）に結合する（図4）。Leu3（アミノ酸生合成に關与する遺伝子を調節することが既知である）は、この群で最も研究されている調節因子である。in vivoでのLeu3の結合は必要であることが示されているが、Leu3調節遺伝子の活性化には十分ではない（Kirkpatrick 他 Mol Cell Biol 15, 4021-30 (1995)）。むしろ、これらの遺伝子の調節的制御には、ロイシン代謝前駆体がLeu3と会合して、負の調節因子から正の調節因子に変換することが必要である。出願人は、「不変条件」挙動を示す他の垂鉛クラスター型調節因子が類似の様式で調節されることが既知であることを留意する（Axelrod 他 Mol Cell Biol 11, 564-7(1991), Ma 他 Cell 50, 137-42 (1987)）。したがって、このクラスの他のいくつかの調節因子の活性化機能又は抑制機能がDNA結合から独立していることを提案することが妥当である。

10

【0253】

「可能な条件」の調節因子は、ある条件下で検出可能な遺伝子を結合しないが、環境が変化すると相当な数のプロモーターに結合する。Msn2は、このクラスで最も研究されている調節因子であり、Msn2依存性転写に關与する機構により、どのようにしてこのクラスの他の調節因子が結合するのかについての手がかりが得られる。Msn2は、ストレスの非存在下で細胞を成長させた場合に核から排除されるが、細胞がストレスに供されると核に迅速に蓄積されることが知られている（Beck 他 Nature 402, 689-92 (1999), Chi et al. Genes Dev 15, 1078-92. (2001)）。この可能な条件の挙動はまた、チアミン生合成調節因子Thi2、窒素調節因子Gat1、及び発生調節因子Rim101でも認められた。出願人は、多数のこれらの転写調整因子が核排除又は条件特異的結合の極端なバージョンを生じる別の機構によって調節されると示唆した。

20

【0254】

「拡大条件」の調節因子は、ある条件下で標的プロモーターのコアセットに結合するが、別の条件下では拡大したプロモーターセットに結合する。Gcn4は、この「拡大」クラスに分類される最も研究された調節因子である。Gcn4レベルは、制限された栄養を含む培地に酵母を導入した場合に6倍になると報告されており（Albrecht 他 J Biol Chem 273, 12696-702. (1998)），due largely to increased nuclear protein stability（Chi 他 Genes Dev 15, 1078-92. (2001), Kornitzer 他 EMBO J 13, 6021-30. (1994)）、この条件下で、出願人は、Gcn4が拡大した遺伝子セットに結合することを見出した。興味深いことに、Gcn4レベルが低い場合に結合するプローブは、より高濃度のタンパク質で排他的に結合するプローブよりも既知のGcn4結合部位により適合し、このことは、固有のタンパク質親和性及びタンパク質濃度に基づいた特性についての簡潔なモデルと一致する（図9）。このクラスの多数の調節因子による結合部位の拡大は、DNA結合に利用可能な調節因子レベルの増加を反映し得る。

30

【0255】

「変化する条件」の調節因子は、2つの異なる条件下で結合したプロモーターセットの基準を変化させる。Ste12は、その挙動がこの「変化する」クラスに分類される最も研究された調節因子である。他の調節因子との相互作用に依存して、Ste12の特異性が変化して、その細胞機能が変化し得る（Zeitlinger, et al. Cell 113, 395-404 (2003)）。例えば、繊維状成長条件下で、Ste12はTec1と相互作用して、Ste12自体がDNA結合特異性を有するようになる（Baur et al. Mol Cell Biol 17, 4330-7 (1997)）。この条件変更挙動は、転写調節因子Aft2、Skn7、及びUme6でも認められた。出願人は、多数の転写調節因子の結合特異性を他の調節因子との相互作用又は環境依存性の修飾（例えば、キメラ）によって変化させることができることを提案する。

40

【0256】

真核生物ゲノム配列の実質的部分は、調節性を示すと考えられるが（Kellis et al. Nature 423, 241-54 (2003), Cliften et al. Science 301, 71-6 (2003), Waterston et al.

50

Nature 420,520-62 (2002))、ゲノム発現の調節に実際に寄与するDNA配列は明確に定義されていない。種々の環境下での特異的調節因子によって結合されたDNA配列のマッピングにより、ゲノム中に組み込まれた調節の可能性が同定され、全体的遺伝子発現に寄与する機構のモデル化のための骨組みが得られる。出願人は、酵母中の調節配列のマッピングのためにここで使用したアプローチを使用して、高等真核生物中でのゲノム発現を調節する配列をマッピングすることもできると認識する。

【0257】

【表1A】

表1—試験した調節因子及び環境条件のリスト

A1	Dat1	Hap3	Met18	Pho4 ¹¹	Sig1 ¹	Swi4	YDR266C
Abf1	Dig1 ^{5,6}	Hap4 ^{2,3}	Met28 ³	Pip2	Sip3	Swi5	YDR520C
Abt1	Dot6	Hap5 ³	Met31 ³	Ppr1	Sip4 ³	Swi6	YER051W
Aca1	Ecm22	Hir1	Met32 ³	Put3 ^{2,3}	Skn7 ^{1,2,7}	Tbs1	YER130C
Ace2	Eds1	Hir2	Met4 ³	Rap1 ³	Skol	Tec1 ^{5,6}	YER184C
Adr1 ^{3,7}	Fap7	Hir3	Mga1 ¹	Rco1	Smk1	Thi2 ¹²	YFL044C
Aft2 ^{1,2}	Fhl1 ^{1,3,4}	Hms1	Mig1 ⁸	Rcs1 ^{1,2,3}	Smp1	Tos8	YFL052W
Arg80 ³	Fkh1	Hms2	Mig2 ¹	Rdr1	Snf1	Tye7	YGR067C
Arg81 ³	Fkh2 ^{1,2}	Hog1	Mig3	Rds1 ¹	Snt2	Uga3 ^{3,4}	Yhp1
Aro80 ³	Fzf1	Hsf1 ^{1,2,7}	Mot3 ^{1,2,3}	Reb1 ^{1,2}	Sok2 ⁵	Ume6 ¹	YJL206C ^{1,2}
Arr1 ¹	Gal3	Ifh1	Msn1	Rfx1	Spt10	Upc2	YKL222C
Ash1 ⁵	Gal4 ^{8,9}	Ime1 ¹	Msn2 ^{1,2,4,7,10}	Rgm1	Spt2	Usv1	YKR064W
Ask10	Gal80	Ime4 ¹	Msn4 ^{1,2,4,10}	Rgt1 ⁸	Spt23	War1	YLR278C
Azf1	Gat1 ^{3,4,7}	Ino2	Mss11 ⁵	Rim101 ^{1,2}	Srd1	Wtm1	YML081W
Bas1 ³	Gat3	Ino4	Mth1 ⁸	Rlm1 ⁵	Stb1	Wtm2	YNR063W
Bye1	Gcn4 ^{3,4}	Ixr1	Ndd1	Rlr1	Stb2	Xbp1 ^{2,7}	Yox1
Cad1 ^{1,3}	Gcr1	Kre33	Ndt80	Rme1	Stb4	Yap1 ^{1,2,7}	YPR022C
Cbf1 ³	Gcr2 ³	Kss1 ^{5,6}	Nnf2	Rox1 ^{1,2}	Stb5	Yap3 ¹	YPR196W
Cha4 ³	Gln3 ^{3,4}	Leu3 ³	Nrg1 ^{1,2}	Rph1 ^{1,2,3}	Stb6	Yap5 ¹	Yrr1
Cin5 ^{1,2}	Gts1	Mac1 ¹	Oaf1	Rpil	Ste12 ^{5,6}	Yap6 ^{1,2}	Zap1
Crz1	Gzf3 ^{1,4}	Mal13	Opi1	Rpn4 ^{1,2}	Stp1 ³	Yap7 ^{1,2}	Zms1
Cst6	Haa1	Mal33 ^{1,2}	Pdc2	Rtg1 ^{3,4}	Stp2	YBL054W	
Cup9	Hac1	Mbfl	Pdr1 ²	Rtg3 ^{1,2,3,4}	Stp4	YBR239C	
Dal80 ⁴	Hal9	Mbp1 ^{1,2}	Pdr3	Rts2	Sum1	YBR267W	
Dal81 ^{3,4}	Hap1	Mcm1 ^{5,6}	Phd1 ⁵	Sfl1	Sut1	YDR026C	
Dal82 ^{3,4}	Hap2 ⁴	Mds3	Pho2 ^{1,2,3,11}	Sfp1 ^{1,2,3}	Sut2	YDR049W	

1大い到高酸素
2中程度の高酸素
3アミノ酸飢餓

4栄養素喪失
5線維化
6交配

7高温
8ガラクトース
9ラフィノース

10酸性
11リン酸喪失
12ビタミン喪失

【表1B】

*全調節因子を、富化培地でプロファイリングした。これらのサブセットを、示されるように、少なくとも1つの他の環境条件でプロファイリングした。条件の完全な説明を、筆者のウェブサイトで見出すことができる。

10

20

30

40

50

【 0 2 5 8 】

【 表 2 A 】

表 2 - 発見した特異性と文献との類似性

調節因子	距離1	発見されたもの	文献	
Abf1	0.143	rTCAytnnnnAcg	rTCAyTnnnnACGw	
Ace2	0.18	tGCTGGT	GCTGGT	
			ATCTTCAAAAAGTGCA-	
Aft2	0.15	rCACCC	CCCATTTCAGGTGC	10
Azf1	0.203	YwTTkcKkTyycgkykky	TTTTCTT	
Bas1	0.045	TGACTC	TGACTC	
Cad1	0.089	mTTAsTmAkC	TACTAA	
Cbf1	0.105	tCACGTG	rTCACrTGA	
Cin5	0.324	TTAcrTAA	TACTAA	
Fkh1	0.123	gtAAAacAA	GGTAAACAA	
Fkh2	0.212	GTAAACA	GGTAAACAA	
Gal4	0.11	CGGnnnnnnnnnnCg	CGGnnnnnnnnnnCCG	20
Gat1	0.004	aGATAAG	GATAA	
Gcn4	0.123	TGAsTCa	ArTGACTCw	
Gln3	0.148	GATAAGa	GATAAGATAAG	
Hap1	0.191	GGnnaTAnCGs	CGGnnnTAnCGG	
Hap4	0.146	gnCcAAtcA	YCnNCCAATNANM	
Hsf1	0.198	TTCynmmnnnTTC	TTCTAGAAmTTCT	
Ino2	0.236	CAcaTGc	ATTTCACATC	
Ino4	0.163	CATGTGaa	CATGTGAAAT	
Leu3	0.131	cCGgtacCGG	yGCCGGTACCGGyk	30
Mbp1	0.073	ACGCGt	ACGCGT	
Mcm1	0.181	CCnrAtnngg	wTTCCyAAwnnGGTAA	
Msn2	0.308	mAGGGGsgg	mAGGGG	

【表 2 B】

Nrg1	0.042	GGaCCCT	CCCT	
Pdr1	0.301	ccGCCgRAwr	CCGCGG	
Pho4	0.096	CACGTGs	cacgtnkg	
Rap1	0.181	cayCCrtrCa	wrmACCCATACayy	
Rcs1	0.184	ggGTGcant	AAmTGGGTGCAkT	
Reb1	0.055	TTACCCG	TTACCCGG	
Rpn4	0.049	GGTGGCAAA	GGTGGCAAA	10
Sip4	0.184	CGGnynAATGGrr	yCGGAyrrAwGG	
Skn7	0.228	GnCnnGsCs	ATTTGGCyGGsCC	
Stb5	0.058	CGGnstTAta	CGG	
Ste12	0.087	tgAAAC	ATGAAAC	
Sum1	0.221	gyGwCAswaaw	AGyGwCACAAAAk	
Sut1	0.295	gcsGsgnnsG	CGCG	
Swi4	0.122	CgCsAAA	CnCGAAA	
Swi6	0.214	CGCgaaa	CnCGAAA	20
Tec1	0.064	CATTcyy	CATTcy	
Tye7	0.193	tCACGTGa	CAnnTG	
Ume6	0.16	taGCCGCCsa	wGCCGCCGw	
Yap1	0.124	TTaGTmAGc	TTAsTmA	
Yap7	0.15	mTkAsTmA	TTACTAA	
Zap1	0.085	ACCCTmAAGGTyrT	ACCCTAAAGGT	

1既知の特異性からの距離を、実験手順に記載の距離評価指標を使用して計算した。

30

【 0 2 5 9 】

【表 3 A】

表 3 - 調節因子の特異性

調節因子	発見された特異性1	既知の特異性1, 2	プログラム4	
Abf1	rTCAytnmnnAcg	rTCAyTnnmACGw	A, C, D, K, M, N	
Ace2	tGCTGGT	GCTGGT	K	
Adr1		GGrGk		
Af2	rCACCC	...AAAGTGCACCCATT...	A, C, D, M, N	40

【表 3 B】

調節因子	発見された特異性1	既知の特異性1, 2	プログラム4	
Arr1		TTACTAA		
Ash1		yTGA CT		
Azf1	YwTTkcKkTyycgykky	TTTTICTT	N	
Bas1	TGACTC	TGACTC	A, K, M, N	
Cad1	mTTAsTmAkC	TTACTAA	A, C, D, M, N	
Cbf1	tCACGTG	rTCACrTGA	A, C, D, K, M, N	10
Cin5	TTAcrTAA	TTACTAA	A, C, D	
Dal80		GATAA		
Dal81		AAAAGCCGCGGGCGGGATT		
Dal82	GATAAG		D, K	
Dig1	TgAAAca		A, C, D, K, M, N	
Fhl1	TGTayGGrtg		A, C, D, K, M, N	
Fkh1	gtAAAcAA	GGTAAACAA	A, C, D, K, M, N	
Fkh2	GTAAACA	GGTAAACAA	A, C, D, K, M, N	
Gal4	CGGmnnnnnnnnncCg	CGGmnnnnnnnnnCCG	A, K	20
Gal80		CGGmnnnnnnnnnCCG		
Gat1	aGATAAG	GATAA	K	
Gcn4	TGAsTCa	ArTGACTCw	A, C, D, K, M, N	
Gcr1		GGCTTCCwC		
Gln3	GATAAGa	GATAAGATAAG	C, D, K	
Gzf3		GATAAG		
Hac1		kGmCAGCGTGTC		
Hap1	GGnnaTAnCGs	CGGnnnTAnCGG	C, M	30
Hap2		CCAAT		
Hap3		CCAAT		
Hap4	gnCcAAtcA	YCNNCCAATNANM	A, C, D, M, N	
Hap5		CCAAT		
Hsf1	TTCynnnnnnTTC	TTCTAGAAAnnTTCT	A, C, D, K, M, N	
Ime1	AAkGAAAnkWA		A	
Ino2	CAcaTGc	ATTTACATC	C, D, M, N	
Ino4	CATGTGaa	CATGTGAAAT	A, C, D, K, M, N	40

【表 3 C】

調節因子	発見された特異性1	既知の特異性1, 2	プログラム4	
Leu3	cCGgtacCGG	yGCCGGTACCGGyk	A, D, K, M	
Mac1		GAGCAAA		
Mbp1	ACGCGt	ACGCGT	A, C, D, K, M, N	
Mcm1	CCnrAtnngg	wTTCCyAAwnnGGTAA	A, C, D, M, N	
Met31		AAACTGTGG		
Met32		AAACTGTGG		10
Met4	RMmAwsTGKSgyGsc		C	
Mot3		yAGGyA		
Msn2	mAGGGGsgg	mAGGGG	M	
Msn4		mAGGGG		
Ndd1	CCnrAwnnGG		A, D	
Nrg1	GGaCCCT	CCCT	A, C, D, M, N	
Opi1		TCGAAyC		
Pdr1	ccGCCgRAwr	CCGCGG	M	
Pdr3		TCCGCGGA		20
Phd1	scnGCngg		A, D, N	
Pho2	SGTGCGsygyG		N	
Pho4	CACGTGs	cacgtkng	D, K, N	
Put3		CGGnnnnnnnnnnCCG		
Rap1	cayCCrrCa	wrmACCCATACAy	A, C, D, M, N	
Rcs1	ggGTGcant	AAmTGGGTGCAkT	C, D, M, N	
Rds1	kCGGCCGa		D, N	
Reb1	TTACCCG	TTACCCGG	A, C, D, K, M, N	30
Rfx1	TTgccATggCAAC		D	
Rgt1		CGGAnnA		
Rim101		TGCCAAG		
Rlm1		CTAwwwTAG		
Rlr1	ATTTTCnnCwTt		N	
Rox1		ysyATTGTT		
Rph1		CCCCTTAAGG		
Rpn4	GGTGGCAAA	GGTGGCAAA	A, C, D, K, M, N	40

【表 3 D】

調節因子	発見された特異性1	既知の特異性1, 2	プログラム4	
Rtg3		GGTCAC		
Sfp1	ayCcrACay		A, C, D, M, N	
Sig1	ArGmAwCrAmAA		M	
Sip4	CGGnynAATGGrr	yCGGAyrrAwGG	D	
Skn7	GnCnnGsCs	ATTTGGCyGGsCC	A, C, D, M, N	
Sko1		ACGTCA		10
Smp1		ACTACTAwwwTAG		
Snt2	yGGCGCTAyca		A, C, D, M, N	
Sok2	tGCAnna		A	
Spt2	ymtGTmTytAw		M	
Spt23	rAAATsaA		C	
Stb1	rracGCsAa		C, D, K, M, N	
Stb4	TCGgmnCGA		K	
Stb5	CGGnstTAta	CGG	D, N	
Ste12	tgAAAC	ATGAAAC	A, C, D, K, M, N	20
Stp1		rCGGCnnnrCGGC		
Sum1	gyGwCAswaaw	AGyGwCACAAAk	A, C, D, M, N	
Sut1	gcsGsgnnsG	CGCG	A, D, M	
Swi4	CgCsAAA	CnCGAAA	A, C, D, K, M, N	
Swi5		kGCTGr		
Swi6	CGCgaaa	CnCGAAA	A, C, D, M, N	
Tec1	CATTCyy	CATTCy	C	
Thi2	gmAAcyntwAgA		C, D	30
Tye7	tCACGTGa	CAnnTG	A, C, D, M	
Uga3		CCGnnnnCGG		
Ume6	taGCCGCCsa	wGCCGCCGw	A, C, D, K, M, N	
Xbp1		CTTCGAG		
Yap1	TTaGTmAGc	TTAsTmA	A, C, D, M	
Yap3		TACTAA		
Yap5		TACTAA		
Yap6		TACTAA		40

【表 3 E】

調節因子	発見された特異性 ¹	既知の特異性 ^{1, 2}	プログラム ⁴
Yap7	mTkAsTmA	TTACTAA	A, C, D, M, N
YDR026C	tTACCCGGm		C, D, M, N
Yhp1		TAATTG	
Yox1		YAATA	
Zap1	ACCCTmAAGGTyrT	ACCCTAAAGGT	N

10

¹ 確率行列のテキスト表示。小文字は、より弱い優先性を示す（確率行列の位置での情報量が少ない）。不確定性コード：S=C又はG、W=A又はT、R=A又はG、Y=C又はT、K=G又はT、M=A又はC、n=A、C、G、又はT。

² 既知の特異性は、YPD、SCP D、及びTRANSFACデータベースから得た。

³ プログラムコード：A=A l i g n A C E、C=CONVERGE、D=MD s c a n、K=K e l l i s 他、M=MEME、N=MEME__c。

【 0 2 6 0 】

【表 4 A】

表 4 - 単一の調節因子構造の結合標的の間で過剰表示されたMIPSカテゴリー

20

調節因子

r	P値 ¹	MIPSカテゴリー ²
Bas1	6.10e-09	ヌクレオチド機構*
Fhl1	1.73e-15	リボゾーム生合成
Gal4	2.18e-04	C-化合物及び炭水化物代謝*
Gat1	4.92e-05	窒素及び硫黄の代謝*
Gat1	2.63e-02	mRNA転写*
Gat1	4.38e-02	アミノ酸代謝
Gcn4	8.72e-12	アミノ酸代謝*
Gzf3	2.21e-02	輸送機構
Hap3	6.03e-03	脂質、脂肪酸、及びイソプレノイドの代謝
Hap3	1.61e-02	アラントイン及びアラントエートの輸送体
Hap3	2.50e-02	他のエネルギー発生活性
Hap4	3.33e-10	呼吸

30

40

【表 4 B】

Hap4	1.78e-05	ミトコンドリア輸送	
Hap4	1.03e-02	輸送機構	
Hap4	2.12e-02	タンパク質複合体のアセンブリ	
Hsf1	6.58e-06	ストレス応答*	
Ino4	5.31e-03	脂質、脂肪酸、及びイソプレノイドの代謝*	
Mbp1	1.04e-04	DNA プロセッシング	
Met32	1.13e-04	アミノ酸代謝*	10
Met32	1.21e-03	窒素及び硫黄の代謝*	
Met32	4.64e-02	アミノ酸輸送体	
Mot3	3.89e-02	DNA プロセッシング	
Msn2	4.40e-02	エネルギー貯蔵代謝 (グリコーゲン、トレハロース)	
Put3	3.45e-02	他の輸送促進物	
Reb1	2.09e-05	小胞輸送 (ゴルジネットワーク等)	
Rfx1	3.57e-02	他のタンパク質合成活性	
Rox1	3.43e-02	細胞死	
Rpn4	2.49e-13	タンパク質分解*	20
Rtg3	8.50e-03	他の転写活性	
Sig1	2.97e-02	細胞周期	
Sip4	2.69e-03	グリオキシル酸回路	
Sip4	1.57e-02	解糖及び糖新生	
Stb4	4.02e-02	電子輸送及び膜結合エネルギー 窒素及び硫黄の代謝*	
Stb5	2.42e-02	保存	
Ste12	5.56e-03	細胞分化*	30
Sut1	5.37e-03	グリオキシル酸回路	
Swi6	7.96e-03	窒素及び硫黄の代謝	
Thi2	1.15e-02	mRNA 転写*	
Thi2	2.45e-02	ビタミン、補因子、及び補欠分子団の代謝	

1P 値は、超幾何学的分布に基づいた、遺伝子が無作為に選択される帰無仮説下で特定された M I P S レベル 2 カテゴリーで観察された遺伝子数 (又はそれ以上) が認められる確率を示す。この値は、ボンフェローニの補正を使用して複数のカテゴリーの試験のために補正されている。

40

【表 4 C】

2アスタリスク (*) は、カテゴリーが調節因子自体にも関連することを示す。

【 0 2 6 1 】

【表 5】

表 5 - 反復モチーフに優先度を有する調節因子

調節因子	P 値 ¹	非反復	反復
Dig1	1.43e-08	O: 25 E: 45	O: 38 E: 17
Mbp1	2.99e-08	O: 34 E: 56	O: 44 E: 21
Swi6	7.36e-06	O: 34 E: 50	O: 37 E: 20
Sok2	1.34e-05	O: 13 E: 24	O: 21 E: 9
Bas1	2.84e-04	O: 6 E: 12	O: 12 E: 5
Ste12	5.57e-04	O: 48 E: 62	O: 39 E: 24
Swi4	7.29e-04	O: 27 E: 38	O: 26 E: 14
Phd1	7.89e-03	O: 15 E: 21	O: 15 E: 8
Aft2	9.73e-03	O: 22 E: 29	O: 19 E: 11
Swi5	1.05e-02	O: 11 E: 16	O: 12 E: 6
Sfp1	3.03e-02	O: 7 E: 10	O: 8 E: 4
Ino2	4.77e-02	O: 11 E: 15	O: 10 E: 5

10

¹P 値は、カイ二乗分布に基づいた、各調節因子の分布が全調節因子と同一の平均分布である帰無仮説下で非反復及び反復モチーフ構造のプロモーター数が認められる片側検定を行った確率を示す。O = 認められた発生数、E = 予想される発生数。

20

【 0 2 6 2 】

【表 6 A】

表 6 - 同時発生調節因子対1

Ace2, Fkh2	Dig1, Swi4	Mbp1, Stb1	Rlm1, Sko1
Ace2, Swi5	Dig1, Swi6	Mbp1, Swi4	Rox1, Sut1
Aft2, Rcs1	Dig1, Tec1	Mbp1, Swi6	Sip4, Stp1
Arr1, Yap3	Fhl1, Rap1	Mcm1, Ndd1	Skn7, Sok2

30

【表 6 B】

Azf1, Gzf3	Fhl1, Sfp1	Mcm1, Ste12	Skn7, Sut1	
Bas1, Met4	Fkh1, Fkh2	Mcm1, Swi4	Skn7, Swi6	
Cad1, Yap1	Fkh2, Mcm1	Mcm1, Swi6	Skn7, Xbp1	
Cad1, Yap7	Fkh2, Ndd1	Mcm1, Tec1	Sko1, Sok2	
Cbfl, Met31	Fkh2, Swi6	Met31, Met32	Sok2, Sut1	
Cbfl, Met32	Gat1, Spt23	Met31, Met4	Sok2, Swi6	
Cbfl, Met4	Gcn4, Gln3	Met32, Met4	Spt23, Yox1	
Cbfl, Pho4	Gcn4, Leu3	Mot3, Rox1	Stb1, Swi4	10
Cbfl, Tye7	Gcr1, Tye7	Mot3, Skn7	Stb1, Swi6	
Cin5, Phd1	Gln3, Hap2	Msn2, Msn4	Stb1, Tec1	
Cin5, Skn7	Gzf3, Pdr1	Msn4, Nrg1	Ste12, Swi4	
Cin5, Sok2	Hap2, Hap3	Nrg1, Rlm1	Ste12, Swi6	
Cin5, Sut1	Hap2, Hap4	Nrg1, Skn7	Ste12, Tec1	
Cin5, Xbp1	Hap2, Hap5	Phd1, Rox1	Swi4, Swi6	
Cin5, Yap6	Hap3, Hap5	Phd1, Skn7	Swi4, Tec1	
Dal82, Gat1	Hap4, Hap5	Phd1, Sok2	Swi6, Tec1	20
Dal82, Gln3	Hsf1, Msn4	Phd1, Sut1	Yap1, Yap7	
Dal82, Hap2	Ino2, Ino4	Phd1, Swi6	Yap6, Yap7	
Dig1, Mcm1	Ino4, Sko1	Rap1, Sfp1		
Dig1, Ste12	Mac1, Rcs1	Rim101, Yox1		

1同時発生調節因子対を示す ($P \leq 0.005$)。P値は、超幾何学的分布に基づいた、2つの調節因子の結合が独立している帰無仮説下で両方の調節因子によって結合された遺伝子間領域の数 (又はそれ以上) が認められる確率を示す。

30

【0 2 6 3】

【表 7 A】

図7—調節因子の挙動分類1

不変2	可能3	拡大4	変化5
Fhl1	Adr1	Bas1	Adr1
Gal4	Arr1	Cad1	Aft2
Gcn4	Ash1	Cbfl	Cad1
Hsf1	Dal81	Cin5	Cin5
Leu3	Fhl1	Dal82	Dal80
Put3	Gat1	Fkh2	Dal82
Ste12	Hap4	Gal4	Dig1

40

【表 7 B】

Ume1	Hsf1	Gcn4	Fkh2	
Yap7	Mot3	Gln3	Gat1	
	Msn2	Hap2	Gln3	
	Pdr1	Mac1	Gzf3	
	Phd1	Mbp1	Hap4	
	Pho2	Mcm1	Hap5	
	Put3	Met31	Mbp1	
	Rap1	Met32	Mot3	
	Rgt1	Met4	Msn2	
	Rim101	Nrg1	Msn4	
	Rlm1	Rcs1	Phd1	10
	Rph1	Rds1	Pho4	
	Rpn4	Reb1	Reb1	
	Rtg3	Rox1	Rox1	
	Sfp1	Rpn4	Rtg3	
	Sig1	Rtg3	Skn7	
	Sip4	Skn7	Ste12	
	Sok2	Ste12	Tec1	
	Stp1		Ume6	
	Thi2		Yap1	
	Uga3		Yap6	
	Xbp1			
	Yap1			20
	Yap7			

1各調節因子の結合を、調節因子がプロファイリングされたそれぞれの環境条件下で2つ1組の様式で比較した。いくつかの調節因子を、比較された条件に厳密に依存して複数のカテゴリーに分類する。

2調節因子 ($P \leq 0.001$) に対する結合プローブの重複比は0.66を超え、結合プローブ数の比は0.66と1.5との間であった。

3調節因子は、ある環境下でプローブに結合しない。

4調節因子に対する結合プローブの重複比は0.66を超え、結合プローブ数の比は0.66と1.5との間であった。

5調節因子は両方の環境下で少なくとも1つのプローブに結合し、結合プローブの重複比は0.66を超えた。

それ以上) が認められる確率を示す。

30

【0264】

【表 8 A】

表 8 - モチーフスコアの有意なカットオフ ($P \leq 0.001$)

富化スコア1

配列数 Converge AlignACE MDscan MEME MEME_c

40

【表 8 B】

10	12.70	20.32	11.78	13.54	n/a
20	11.96	21.14	12.95	12.89	9.81
30	11.43	20.43	13.30	12.57	n/a
40	11.34	20.62	14.04	11.64	7.53
50	10.74	19.94	12.23	12.81	7.43
60	10.50	19.71	10.95	12.37	n/a
70	10.34	18.30	13.25	11.34	n/a
80	10.20	19.40	12.84	11.93	n/a
100	9.36	20.31	11.56	10.58	2.91
120	n/a	18.59	13.14	10.94	n/a
140	8.14	18.52	11.26	10.87	n/a
160	n/a	20.04	11.38	9.77	n/a

10

ROC a.u.c.¹

20

配列数	Converge	AlignACE	MDscan	MEME	MEME_c
10	n/a	n/a	n/a	n/a	n/a
20	0.812	0.842	0.857	0.925	n/a
30	0.758	0.773	0.793	0.831	0.785
40	0.720	0.713	0.758	0.764	0.737
50	0.687	0.674	0.719	0.737	0.711
60	0.670	0.662	0.688	0.706	0.654
70	0.663	0.641	0.686	0.684	0.664
80	0.643	0.626	0.670	0.675	0.648
100	0.634	0.615	0.664	0.633	0.606
120	0.624	0.604	0.629	0.624	0.602
140	0.608	n/a	0.634	n/a	0.590
160	0.594	0.580	0.613	0.593	0.588

30

1方法に記載の遺伝子間領域の無作為な選択の計算から得られた「富化」及び「ROC a. u. c.」特異性評価指標についてのモチーフスコア有意性 $P \leq 0.001$ 閾値。「n/a」を含むエントリーは、経験的分布が正規分布でなかったことを示す。CC4評価指標の閾値(4.95)は、配列数に依存しない。

40

【図面の簡単な説明】

【0265】

【図1A】酵母転写調節因子の結合部位特異性を発見するための一般的計画を示す図である。実施例に記載のように、転写調節因子の認識部位として機能する可能性が高いシス調節配列を、ゲノム規模の位置データ、系統発生的に保存された配列、及び以前に公表されたデータ由来の情報の組み合わせによって同定した。調節配列モチーフの概要を、表3に見出すことができる。

【図1B】酵母転写調節因子の結合部位特異性を発見するための一般的計画を示す図である。「再発見された」選択配列特異性及び新規に発見された選択配列特異性を示す。カラ

50

ムの高さは、位置情報量に比例し、各文字は、その頻度及び情報量の産物に比例する高さを有する (Schneider et al. Nucleic Acids Res 18,6097-100 (1990))。

【図 2 A】酵母転写調節マップの設計図を示す。in vivo で転写調節因子によって結合した遺伝子の位置 (大きな長方形) 及び保存 DNA 配列の位置 (小ボックス) を示す染色体部分。

【図 2 B】酵母転写調節マップの設計図を示す。結合データと配列保存データとの組み合わせ。図は、大要由来のモチーフに適合する全配列 (上)、全てのこのような保存配列 (中央)、及び調節因子に結合した全てのこのような保存配列 (下) を示す。

【図 2 C】酵母転写調節マップの設計図を示す。調節因子結合部位の分布。「actual」と表示した線は、読み取り枠の開始コドンから隣接する上流領域中の結合部位までの距離の分布を示す。「randomized」と表示した線は、無作為な分布を示す。 10

【図 3】酵母プロモーターの構造を示す図である。単一の調節因子の構造：単一の調節因子の結合部位配列の 1 つ又は複数のコピーを含むプロモーター領域。反復モチーフの構造：調節因子の結合部位配列の複数のコピーを含むプロモーター領域。多調節因子の構造：1 つを超える調節因子の結合部位配列の 1 つ又は複数のコピーを含むプロモーター領域。同時発生 (co-occurring) 調節因子の構造：調節因子の再発 (recurrent) 対の結合部位配列を含むプロモーター。例示目的のために、全ての部位を示しているわけではなく、尺度もおおよそである。さらなる情報を、表 4 ~ 6 に見出すことができる。

【図 4】転写調節コードの環境特異的活用を示す図である。4 パターンのゲノム規模の結合挙動を、左側の図に示し、転写調節因子を円で示し、標的遺伝子/プロモーターセットの上下に配置する。調節因子と標的遺伝子/プロモーターとの間の線は、結合事象を示す。環境依存性挙動の特定の例を、右側に示す。円は調節因子を示し、ボックスは特定のプロモーター領域内の DNA 結合配列を示す。出願人は、調節因子は異なる条件対を比較した場合に異なる挙動を示し得ることに留意している。 20

【図 5】「actual」と表示した線として 1 調節因子あたりに結合したプロモーター領域数の分布を示す図である。複数の条件下でプロファイリングした調節因子について、全条件下で結合したプロモーター領域の集合 (union) を報告する。調節因子とプロモーター領域の間で無作為に割り当てた P 値の同一のセットの平均無作為化分布を、「randomized」と表示した線として示す。 30

【図 6】モチーフの発見及び割り当ての概要を示す図である。一連のモチーフ発見プログラムを結合データによって同定された遺伝子間配列に適用することによってモチーフを同定した。得られた特異性の予測を、有意性についてフィルタリングし、クラスター化して代表的モチーフを得た。保存ベースの評価指標を使用して、これらのモチーフの最も信頼性の高いサブセットを同定した。因子について複数の有意な結合モチーフが見出される場合について、出願人は、特異性データベース由来の統計的スコア又は情報を使用して各調節因子に対して 1 つのモチーフを選択した。方法の完全な記載を、実験手順で見出すことができる。

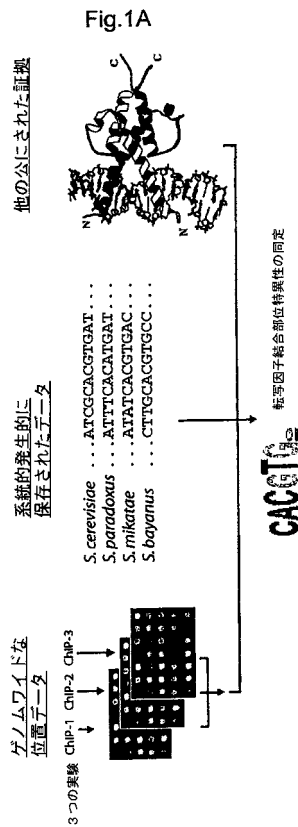
【図 7】2 つの配列の C i n 5 結合の比較を示す図である。組換え C i n 5 を細菌から精製し、配列 (gcgacaTTACCTAAgggc) を含む C y 5 標識オリゴヌクレオチドとインキュベートし、以下の 2 つの非標識競合物のうち 1 つで攻撃誘発した：同一の配列 (レーン 2 ~ 8) 又は以前に公開された結合部位 (gcgacaTTACTAAgggc、レーン 9 ~ 15)。各競合物の濃度は、3 倍で変化した。本発明者らが発見したモチーフに基づくプローブは、以前に公開された特異性に基づいたプローブと比較して、シフトしたバンドの競合において約 2.7 倍良好であった。TTACGTAA のコア配列を含むプローブについて類似の結果が得られた。 40

【図 8】25 の調節因子についての 2 つの異なる条件下で結合したプロモーター領域数の 2 つ 1 組の比較を示す図である ($P < 0.001$ のゲノム規模の位置データのみに基づく)。暗色のバーは、富化培地で結合したプロモーター領域数を示し、明色のバーは、アミノ酸枯渇培地中での成長下で結合したプロモーター領域数を示す。

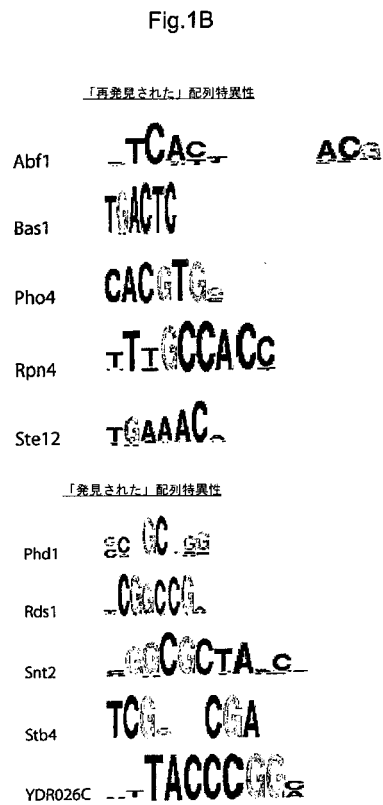
【図 9】異なる条件下で結合した遺伝子間領域の間の G c n 4 結合部位の質を示す図であ 50

る。各遺伝子間領域を、Gcn4結合特異性に最良に適合するサブシーケンス(TGASTCA)の質に基づいてスコアリングした。富化培地条件では、68%の遺伝子間領域は、Gcn4特異性に高品質で適合する。枯渇条件下で、Gcn4タンパク質レベルは上昇し、結合した遺伝子間領域セットは拡大する。新規に結合した領域のうち、27%のみが高品質に適合する。対照的に、全遺伝子間領域のうちの3%のみがこの質で適合する。

【 図 1 A 】

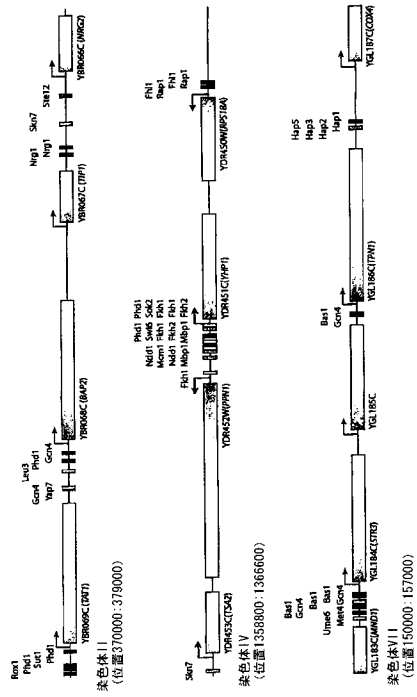


【 図 1 B 】



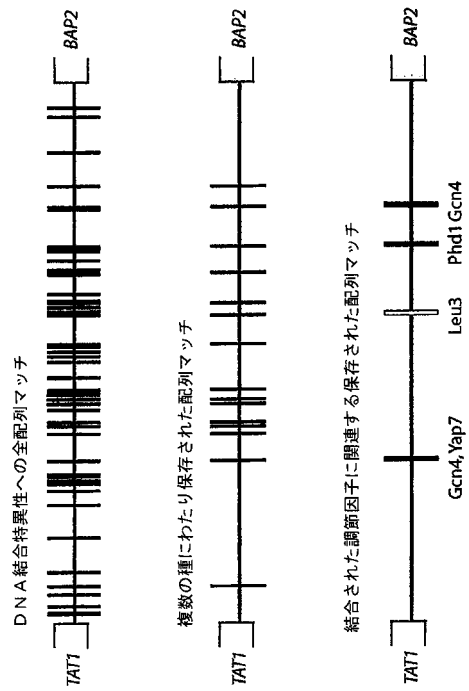
【 図 2 A 】

Fig.2A



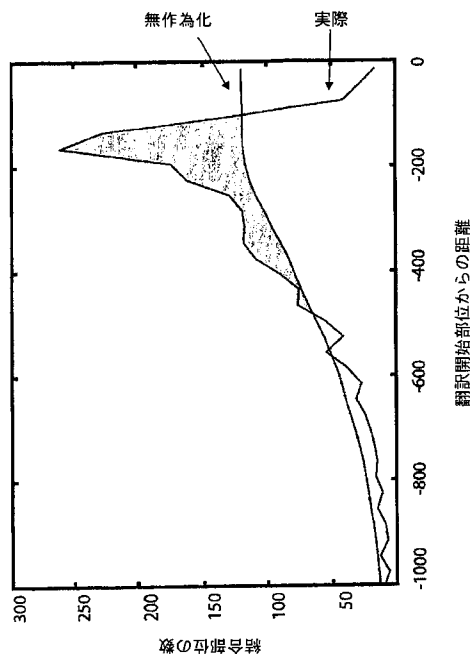
【 図 2 B 】

Fig.2B



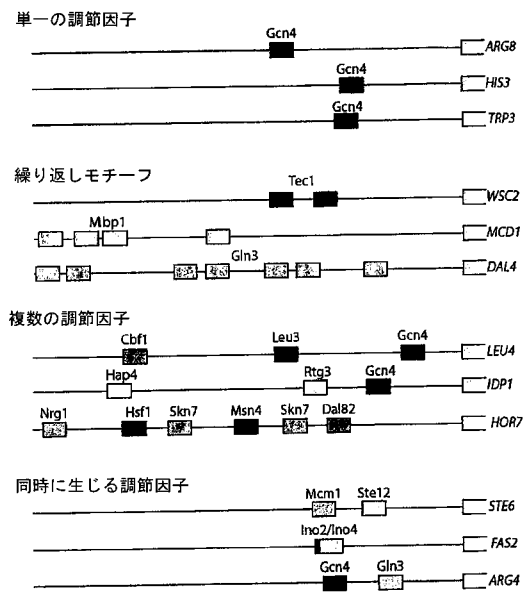
【 図 2 C 】

Fig.2C

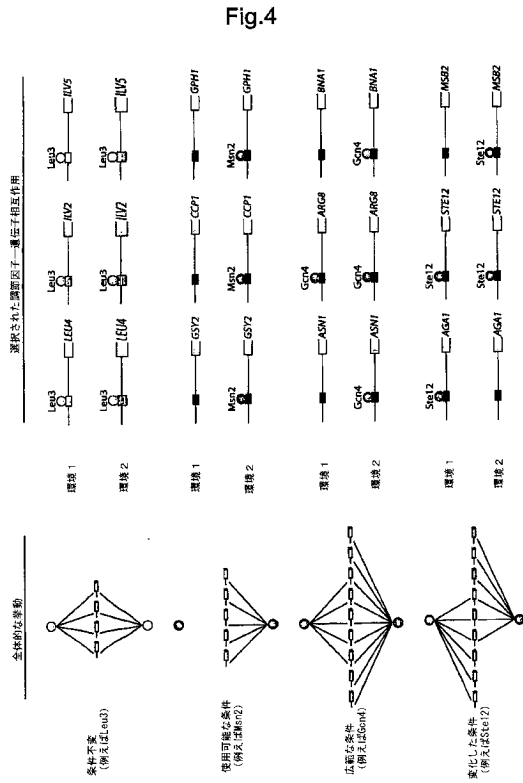


【 図 3 】

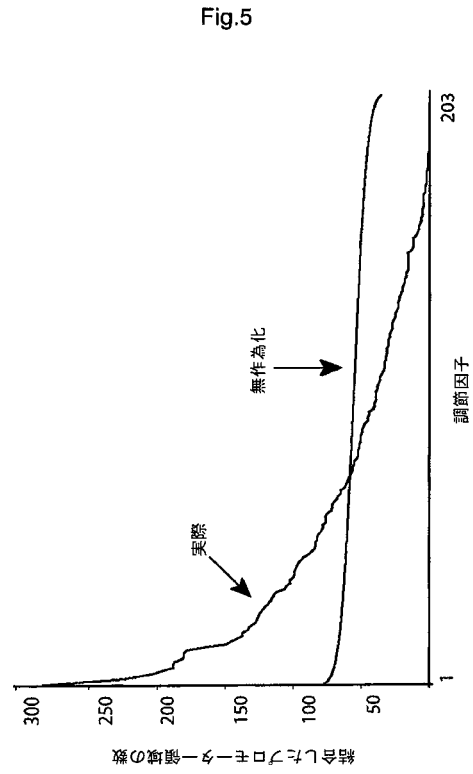
Fig.3



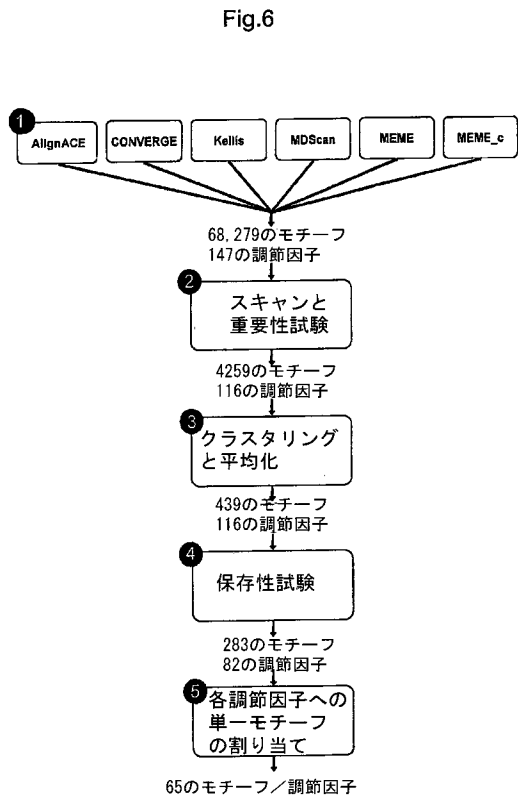
【 図 4 】



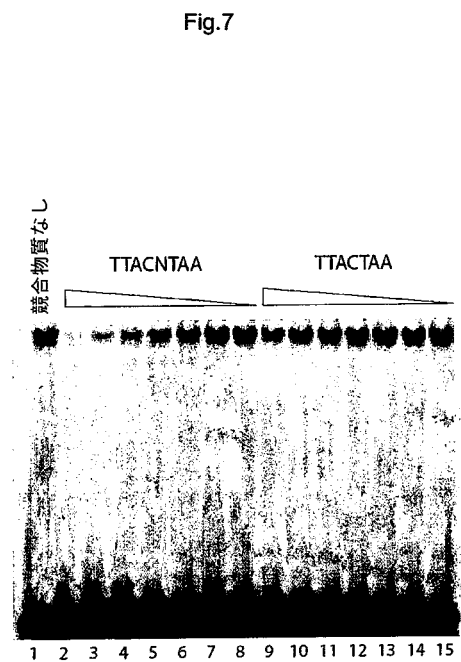
【 図 5 】



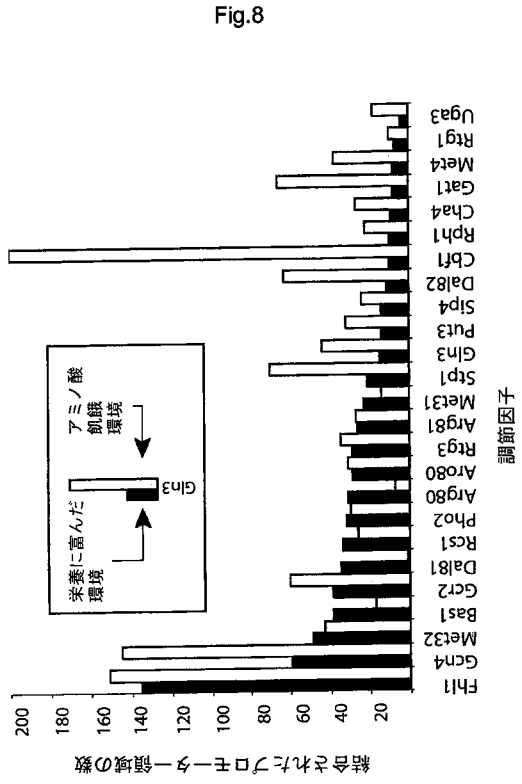
【 図 6 】



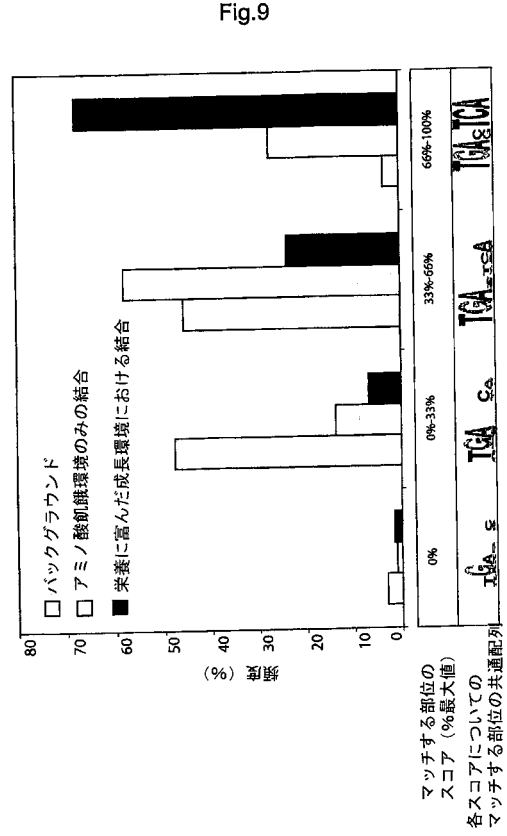
【 図 7 】



【 図 8 】



【 図 9 】



【 国際調査報告 】

INTERNATIONAL SEARCH REPORT

 International Application No
 PCT/US2005/007249

A. CLASSIFICATION OF SUBJECT MATTER IPC 7 C12Q1/68 G01N33/53		
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) IPC 7 G01N C12Q		
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 practical, search terms used) EPO-Internal, WPI Data, PAJ, BIOSIS, EMBASE		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
E	WO 2005/054461 A (WHITEHEAD INSTITUTE FOR BIOMEDICAL RESEARCH; ODOM, DUNCAN, T; YOUNG, R) 16 June 2005 (2005-06-16) claims page 11, line 4 - page 12, line 21 page 28, line 16 - page 29, line 22 page 32, line 5 - line 17	1-60
P,X	WO 2004/097577 A (NEW YORK UNIVERSITY; CHEREPINSKY, VERA; FENG, JIA-WU; REJALI, MARC; MI) 11 November 2004 (2004-11-11) page 24, line 3 - page 45, line 12 claims 73-94	1-35
P,X	WO 2004/087965 A2 (REGULOME CORPORATION; STAMATOYANNOPOULOS, JOHN, A; MCARTHUR, MICHAEL;) 14 October 2004 (2004-10-14) claims	1-60
-/-		
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex.		
* Special categories of cited documents : *A* document defining the general state of the art which is not considered to be of particular relevance *E* earlier document but published on or after the international filing date *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) *O* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed *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 *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 *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. *&* document member of the same patent family		
Date of the actual completion of the international search 29 August 2005		Date of mailing of the international search report 19/09/2005
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Authorized officer Routledge, B

2

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US2005/007249

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO 2004/053106 A (REGULOME CORPORATION; STAMATOYANNOPOULOS, JOHN, A; MCARTHUR, MICHAEL;) 24 June 2004 (2004-06-24) claims page 16, line 25 - line 32 page 17, line 15 - page 22, line 17 page 38, line 7 - page 48, line 7	1-60
P,X	MUKHERJEE ET AL: "Rapid analysis of the DNA binding specificities of transcription factors with DNA microarrays." NATURE GENETICS, vol. 36, no. 12, December 2004 (2004-12), pages 1331-1339, XP002342489 the whole document	1-35
Y	WO 02/14550 A (TRANSGENETICS INCORPORATED; BURGESS, ROBERT, M., JR; LUNYAK, VICTORIA;) 21 February 2002 (2002-02-21) claims page 43, paragraph 2	1-60
Y	WO 01/16378 A (WHITEHEAD INSTITUTE FOR BIOMEDICAL RESEARCH) 8 March 2001 (2001-03-08) claims	1-60
Y	US 6 109 776 A (HAAS ET AL) 29 August 2000 (2000-08-29) claims column 4, line 24 - column 5, line 55 figures 9-15	1-60
X	ODOM D T ET AL: "CONTROL OF PANCREAS AND LIVER GENE EXPRESSION BY HNF TRANSCRIPTION FACTORS" SCIENCE, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE,, US, vol. 303, no. 5662, 27 February 2004 (2004-02-27), pages 1378-1381, XP001205726 ISSN: 0036-8075	1-35
Y	page 1378, right-hand column, paragraph 2 page 1379, left-hand column, paragraph 4 page 1381, left-hand column, paragraph 2 - middle column, paragraph 1	1-60
X	BAR JOSEPH ET AL: "Computational discovery of gene modules and regulatory networks" NATURE BIOTECHNOLOGY, vol. 21, no. 11, November 2003 (2003-11), pages 1337-1342, XP002342488 the whole document	1-60
	-/--	

2

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US2005/007249

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	LEE T I ET AL: "Transcriptional regulatory networks in Saccharomyces cerevisiae" SCIENCE, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE,, US, vol. 298, no. 5594, 25 October 2002 (2002-10-25), pages 799-804, XP002327643 ISSN: 0036-8075 page 799, right-hand column, paragraph 2 - paragraph 3 page 804, middle column, paragraph 4	1-60
X	WYRICK ET AL: "Deciphering gene expression regulatory networks." CURRENT OPINION IN GENETICS AND DEVELOPMENT, vol. 12, Apr11 2002 (2002-04), pages 130-136, XP002342490 the whole document	1-60
X	WEINMANN A S ET AL: "Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis" GENES AND DEVELOPMENT, COLD SPRING HARBOR LABORATORY PRESS, NEW YORK, US, vol. 16, no. 2, 15 January 2002 (2002-01-15), pages 235-244, XP002327644 ISSN: 0890-9369 page 236, left-hand column, line 3 - line 11 page 241, right-hand column, paragraph 5 - page 242, left-hand column, paragraph 1	1-35
Y	HARTEMINK A J ET AL: "COMBINING LOCATION AND EXPRESSION DATA FOR PRINCIPLED DISCOVERY OF GENETIC REGULATORY NETWORK MODELS" PROCEEDINGS OF THE PACIFIC SYMPOSIUM ON BIOCOMPUTING, 2002, pages 437-449, XP009046817 the whole document	1-60
Y		1-60

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2005/007249

Box No. 1 Nucleotide and/or amino acid sequence(s) (Continuation of item 1.b of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application and necessary to the claimed invention, the international search was carried out on the basis of:
- a. type of material
- a sequence listing
- table(s) related to the sequence listing
- b. format of material
- In written format
- In computer readable form
- c. time of filing/furnishing
- contained in the international application as filed
- filed together with the international application in computer readable form
- furnished subsequently to this Authority for the purpose of search
2. In addition, in the case that more than one version or copy of a sequence listing and/or table relating thereto has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2005/007249**Box 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:
2. Claims Nos.: 1-60 (partly)
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).

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

Remark on Protest

- The additional search fees were accompanied by the applicant's protest.
- No protest accompanied the payment of additional search fees.

International Application No. PCT/US2005/007249

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box II.2

Claims Nos.: 1-60 (partly)

Present claims 1-60 relate to methods which recite an extremely large number of possible proteins of interest. Support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT is to be found, however, for only a very small proportion of the methods claimed. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Consequently, the search has been carried out for those parts of the claims which appear to be supported and disclosed, namely those parts relating to methods involving transcription factors/regulators.

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 Guideline C-VI, 8.5), should the problems which led to the Article 17(2) declaration be overcome.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US2005/007249

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2005054461 A	16-06-2005	WO 2005054461 A2	16-06-2005
WO 2004097577 A	11-11-2004	WO 2004097577 A2	11-11-2004
WO 2004087965 A2	14-10-2004	US 2004191779 A1 JP 2004298178 A	30-09-2004 28-10-2004
WO 2004053106 A	24-06-2004	AU 2003302777 A1 WO 2004053106 A2	30-06-2004 24-06-2004
WO 0214550 A	21-02-2002	AU 8318001 A CA 2419479 A1 EP 1425579 A2 WO 0214550 A2	25-02-2002 21-02-2002 09-06-2004 21-02-2002
WO 0116378 A	08-03-2001	CA 2383928 A1 EP 1220949 A2 JP 2003508066 T WO 0116378 A2 US 6410243 B1	08-03-2001 10-07-2002 04-03-2003 08-03-2001 25-06-2002
US 6109776 A	29-08-2000	NONE	

フロントページの続き

(81) 指定国 AP(BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), EA(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), EP(AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW

(72) 発明者 ハービソン, クリストファー
アメリカ合衆国マサチューセッツ州 0 2 1 6 9 , クインシー , クインシー・アヴェニュー 1 4 0
 , # 2 3

(72) 発明者 ヤング, リチャード・エイ
アメリカ合衆国マサチューセッツ州 0 2 1 9 3 , ウェストン , ハイランド・ストリート 2 1 6

(72) 発明者 ゴードン, デイヴィッド・ビー
アメリカ合衆国マサチューセッツ州 0 2 1 4 3 , ソマーヴィル, コンコード・アヴェニュー 1 3
0 , アpartment 4 ディー

(72) 発明者 フレンケル, アーネスト
アメリカ合衆国マサチューセッツ州 0 2 1 3 8 - 6 0 1 8 , ケンブリッジ, サリー・ストリート
2 8 , # 1

F ターム(参考) 2G045 AA40 CB01 DA12 DA13 DA14 DA36
4B024 AA11 CA04 CA05 CA06 CA09 CA10 CA12 DA12 EA04 FA02
FA10 GA18 GA19 HA08 HA12 HA14
4B063 QA01 QA13 QA18 QQ06 QQ07 QQ08 QQ20 QQ42 QQ53 QQ79
QR08 QR32 QR36 QR42 QR48 QR55 QR66 QR69 QR75 QR76
QR77 QR82 QS03 QS12 QS16 QS17 QS25 QS34 QS36 QX02

专利名称(译)	生物活性DNA结合位点和相关方法		
公开(公告)号	JP2007526776A	公开(公告)日	2007-09-20
申请号	JP2007502063	申请日	2005-03-03
申请(专利权)人(译)	怀特黑德生物医学研究所		
[标]发明人	ハービソンクリストファー ヤングリチャードエイ ゴードンデイヴィッドビー フレンケルアーネスト		
发明人	ハービソン,クリストファー ヤング,リチャード・エイ ゴードン,デイヴィッド・ビー フレンケル,アーネスト		
IPC分类号	C12Q1/68 G01N33/15 G01N33/50 C12Q1/02 C12N15/09 G01N33/53		
CPC分类号	G01N33/5308 C12Q2522/00		
FI分类号	C12Q1/68.ZNA.A G01N33/15.Z G01N33/50.Z C12Q1/02 C12N15/00.A		
F-TERM分类号	2G045/AA40 2G045/CB01 2G045/DA12 2G045/DA13 2G045/DA14 2G045/DA36 4B024/AA11 4B024/CA04 4B024/CA05 4B024/CA06 4B024/CA09 4B024/CA10 4B024/CA12 4B024/DA12 4B024/EA04 4B024/FA02 4B024/FA10 4B024/GA18 4B024/GA19 4B024/HA08 4B024/HA12 4B024/HA14 4B063/QA01 4B063/QA13 4B063/QA18 4B063/QQ06 4B063/QQ07 4B063/QQ08 4B063/QQ20 4B063/QQ42 4B063/QQ53 4B063/QQ79 4B063/QR08 4B063/QR32 4B063/QR36 4B063/QR42 4B063/QR48 4B063/QR55 4B063/QR66 4B063/QR69 4B063/QR75 4B063/QR76 4B063/QR77 4B063/QR82 4B063/QS03 4B063/QS12 4B063/QS16 4B063/QS17 4B063/QS25 4B063/QS34 4B063/QS36 4B063/QX02		
优先权	60/550074 2004-03-04 US 60/604470 2004-08-25 US		
外部链接	Espacenet		

摘要(译)

本发明涉及细胞中目标蛋白质结合的生物活性DNA结合位点的鉴定。本发明还涉及改变蛋白质结合的生物活性DNA结合位点的试剂和条件的鉴定。本发明的一个实施方案还提供了鉴定由转录调节子调节的途径的方法和调节途径活性的方法。

