

【公報種別】特許法第17条の2の規定による補正の掲載

【部門区分】第1部門第1区分

【発行日】平成17年12月22日(2005.12.22)

【公表番号】特表2004-533262(P2004-533262A)

【公表日】平成16年11月4日(2004.11.4)

【年通号数】公開・登録公報2004-043

【出願番号】特願2003-510461(P2003-510461)

【国際特許分類第7版】

C 1 2 N 15/09

C 1 2 Q 1/68

【F I】

C 1 2 N 15/00 Z N A A

C 1 2 Q 1/68 A

【誤訳訂正書】

【提出日】平成17年2月14日(2005.2.14)

【誤訳訂正1】

【訂正対象書類名】特許請求の範囲

【訂正対象項目名】全文

【訂正方法】変更

【訂正の内容】

【特許請求の範囲】

【請求項1】

遺伝子転写調節特性を有する単離された及び／又は組換えられたDNA配列であって、前記配列は、(a)図26のSEQ ID:1-119から成る群から選択される配列と；(b)図26のSEQ ID:1-119から成る群から選択される配列の機能的フラグメント、誘導体又は相同物とから成る群から選択される配列を含み、前記配列は、

- 250マイクロg/ml ゼオシン及び0.1ng/ml ドキシサイクリンの存在下で、5週間細胞を成長すること；

- 5週間後になお成長するコロニーを選択すること；

- 前記コロニーから単離されたプラスミド中の遺伝子転写調節抑制を有する前記DNA配列の存在を検証すること；

のステップを含む工程によって決定されるような遺伝子転写調節特性を有し、

ここで前記細胞は、LexA DNA結合性ドメイン及びHP1又はHPC2のコード領域を含むLexA-リプレッサー融合タンパク質をTet-Off転写的調節システムの制御下で発現するU-2 OSヒト骨肉腫細胞株からのものであり、及び

ここで前記細胞は、4つのLexAオペレーター部位及びゼオシン耐性遺伝子を制御するSV40プロモーターの間に置かれた、前記プラスミド内のポリリンカー配列内にクローニングされた転写調節特性を有する前記DNAフラグメントを含むプラスミドを含み、及び

ここでさもなければ前記ゼオシン耐性遺伝子は、もし転写調節特性を有する前記DNAフラグメントの存在がなければ、前記LexA-リプレッサー融合タンパク質の発現故に抑制されるであろう、

前記DNA配列。

【請求項2】

約50～5000塩基対のテスト核酸配列内でSTAR配列の存在を予測するための方法であって

a)図26のSEQ ID:1-65及びSEQ ID:85-119から成る群から選択されるSTAR配列を含むSTAR配列のセットに対してパターン頻度分析を実行して、ランダム配列と比較してSTAR配列内の有意に過剰に表されている配列パターンのセットを識別すること、

b) ステップa)で得られた前記セットに対してステップワイズ線形判別式分析によりSTAR配列の予測のためのモデルをトレーニングして、STAR配列とランダム配列との間の最大の判別力を有するパターンのサブセットを得ること、

c) ステップb)で得られた前記サブセットを使用して、前記テスト核酸配列内のSTAR配列の存在を予測すること、

を含む、前記方法。

【請求項3】

前記ランダム配列が、

i) 前記テスト配列と同じ種(species)から得られ；及び／又は

ii) 前記テスト配列と類似のAT/CG含量を有する、

請求項2に記載の方法。

【請求項4】

ステップa)の前記パターン頻度分析は、

a) 少なくとも5塩基、及び／又は

b) 少なくとも6塩基、及び／又は

c) 少なくとも8塩基、

を含む配列パターンを使用して行われる、請求項2又は請求項3に記載の方法。

【請求項5】

約50～5000塩基対のテスト核酸配列内でSTAR配列の存在を予測するための方法であって

a) 表9及び／又は表10にリストされたパターンを含む配列のセットに対してステップワイズ線形判別式分析によりSTAR配列の予測のためのモデルをトレーニングして、STAR配列とランダム配列との間の最大の判別力を有するパターンのサブセットを得ること、

b) ステップa)で得られた前記サブセットを使用して、前記テスト核酸配列内のSTAR配列の存在を予測すること、

を含む、前記方法。

【請求項6】

約50～5000塩基対のテスト核酸配列内でSTAR配列の存在を予測するための方法であって

配列パターンのセットを使用して、前記テスト核酸配列内でSTAR配列の存在を予測することを含み、ここで配列パターンの前記セットは、GGACCC、CCCTGC、AAGCCC、CCCCCA、AGCACC、CCCN{16}AGC、GGCN{9}GAC、CACN{13}AGG、CTGN{4}GCCから成る群から選択される配列パターンを含む、前記方法。

【請求項7】

約50～5000塩基対のテスト核酸配列内でSTAR配列の存在を予測するための方法であって

配列パターンGGACCC、CCCTGC、AAGCCC、CCCCCA、AGCACC、CCCN{16}AGC、GGCN{9}GAC、CACN{13}AGG、及びCTGN{4}GCCを使用して、前記テスト核酸配列内でSTAR配列の存在を予測することを含む、前記方法。

【請求項8】

前記方法が、

d) 前記予測されたSTAR配列が遺伝子転写システムにおいて遺伝子転写調節特性を有するかどうかを決定することをさらに含む、請求項2～7のいずれか一項に記載の方法を含むSTAR配列を識別するための方法。

【請求項9】

予測されたSTAR配列が遺伝子転写システムにおいて遺伝子転写調節特性を有するかどうかを決定することが、

- 250マイクロg/ml ゼオシン及び0.1ng/ml ドキシサイクリンの存在下で、5週間細胞を成長すること；

- 5週間後になお成長するコロニーを選択すること；

- 前記コロニーから単離されたプラスミド中のSTAR配列の存在を検証すること；
を含み、

ここで前記細胞は、LexA DNA結合性ドメイン及びHP1又はHPC2のコード領域を含むLexA-リプレッサー融合タンパク質をTet-Off転写的調節システムの制御下で発現するU-2 OSヒト骨肉腫細胞株からのものであり、及び

ここで前記細胞は、4つのLexAオペレーター部位及びゼオシン耐性遺伝子を制御するSV40プロモーターの間に置かれた、前記プラスミド内のポリリンカー配列内にクローニングされたSTAR配列を含むプラスミドを含み、及びここでさもなければ前記ゼオシン耐性遺伝子は、もし転写調節特性を有する前記STAR配列の存在がなければ、前記LexA-リプレッサー融合タンパク質の発現故に抑制されるであろう。

請求項8記載の方法。

【請求項10】

遺伝子転写調節活性を有する単離された及び／又は組み換えられた配列であって、前記配列は請求項9記載の方法によって得られる、前記配列。

【請求項11】

パターン頻度分析でランダム配列と比較して少なくとも1つの有意に過剰に表されている配列パターンを含む単離された及び／又は組換えられた配列であって、ここで前記有意に過剰に表されている配列パターンは、表9又は表10にリストされた配列パターンであり、前記単離された及び／又は組換えられた配列は、

- 250マイクロg/ml ゼオシン及び0.1ng/ml ドキシサイクリンの存在下で、5週間細胞を成長すること；

- 5週間後になお成長するコロニーを選択すること；

- 前記コロニーから単離されたプラスミド中の遺伝子転写調節抑制を有する前記配列の存在を検証すること；

のステップを含むプロセスによって決定されるような遺伝子転写調節特性を有し、

ここで前記細胞は、LexA DNA結合性ドメイン及びHP1又はHPC2のコード領域を含むLexA-リプレッサー融合タンパク質をTet-Off転写的調節システムの制御下で発現するU-2 OSヒト骨肉腫細胞株からのものであり、及び

ここで前記細胞は、4つのLexAオペレーター部位及びゼオシン耐性遺伝子を制御するSV40プロモーターの間に置かれた、前記プラスミド内のポリリンカー配列内にクローニングされた転写調節特性を有する前記配列を含むプラスミドを含み、及び

ここでさもなければ前記ゼオシン耐性遺伝子は、もし転写調節特性を有する前記配列の存在がなければ、前記LexA-リプレッサー融合タンパク質の発現故に抑制されるであろう

、

前記配列。

【請求項12】

前記少なくとも1つの有意に過剰に表されている配列パターンは、配列パターンGGACCC、CCCTGC、AAGCCC、CCCCCA、AGCACCC、CCCN{16}AGC、GGCN{9}GAC、CACN{13}AGG、又はCTGN{4}GCCを含む、請求項11に記載の単離された及び／又は組換えられた配列。

【請求項13】

遺伝子転写調節特性を有する単離された及び／又は組換えられたDNAフラグメントであって、そのDNAフラグメントは、

a) 250マイクロg/ml ゼオシン及び0.1ng/ml ドキシサイクリンの存在下で、5週間細胞を成長すること；

b) 5週間後になお成長するコロニーを選択すること；

c) 遺伝子転写調節特性を有する前記DNA配列を単離し及び／又は分析すること；

d) ステップa)～c)を繰り返すことによって前記単離されたDNAフラグメントの前記遺伝子転写調節特性を確認すること

を含む方法によって得られることができ、

ここで前記細胞は、LexA DNA結合性ドメイン及びHP1又はHPC2のコード領域を含むLexA-

リプレッサー融合タンパク質をTet-Off転写的調節システムの制御下で発現するU-2 OSヒト骨肉腫細胞株からのものであり、及び

ここで前記細胞は、4つのLexAオペレーター部位及びゼオシン耐性遺伝子を制御するSV40プロモーターの間に置かれた、前記プラスミド内のポリリンカー配列内にクローニングされた転写調節特性についてスクリーニングされるべき前記DNAフラグメントを含むプラスミドを含み、及び

ここでさもなければ前記ゼオシン耐性遺伝子は、もし転写調節特性を有する前記DNAフラグメントの存在がなければ、前記LexA-リプレッサー融合タンパク質の発現故に抑制されるであろう。

前記DNAフラグメント。

【請求項14】

前記DNAフラグメントが植物若しくは脊椎動物、又はそれらの派生体から得られ、或いは合成DNAフラグメントすなわち遺伝子工学によって構築されたDNAフラグメントである、請求項13に記載の遺伝子転写調節特性を有する単離された及び／又は組換えられたDNAフラグメント。

【請求項15】

レポーター遺伝子が、ショウジョウバエSCSエレメントによって隣接されている又は遺伝子転写調節特性を有するDNAフラグメントによって隣接されていない場合と比べて、U-2 OS細胞のゲノム内に統合される前記レポーター遺伝子を隣接する場合に前記DNAフラグメントが、レポーター過剰発現クローンのより高い割合を提供する、請求項13又は請求項14に記載の遺伝子転写調節特性を有する単離された及び／又は組換えられたDNAフラグメント。

【請求項16】

前記DNAフラグメントは、同一の核酸単独と比較して転写ユニットをコードする動作可能にリンクされた核酸に、(i)転写のより高い予測性、(ii)より高い転写、及び／又は(iii)長期にわたる転写のより高い安定性を与えることが可能である、請求項13～15のいずれか一項に記載の遺伝子転写調節特性を有する単離された及び／又は組換えられたDNAフラグメント。

【請求項17】

請求項13～16のいずれか一項に記載の単離された及び／又は組換えられたDNAフラグメントであって、ヒトから得られたものである、前記DNAフラグメント。

【請求項18】

請求項13～16のいずれか一項に記載の単離された及び／又は組換えられたDNAフラグメントであって、植物から得られたものである、前記DNAフラグメント。

【請求項19】

遺伝子転写調節特性を有する配列の存在を検出するための方法であって、種(species)の細胞の染色体の一部分における請求項1又は10～18のいずれか一項に記載の遺伝子転写を助長する特性を含む配列を識別すること、及び前記配列と、異なる種の染色体の配列との間の有意な相同性を検出すること、及び異なる種の染色体の前記配列が、転写システムにおける遺伝子転写調節特性を有するかどうかを決定することを含む、前記方法。

【請求項20】

脊椎動物種又は植物種の約50～5000塩基対の核酸配列内の遺伝子転写調節特性を有する配列の存在を検出する方法であって、請求項1又は10～18のいずれか一項に記載の遺伝子転写を助長する特性を含む配列を識別すること、及び前記核酸配列の隣接(fanking)配列が少なくとも1つの他の種内で保存されているかどうかを識別すること、及び前記少なくとも1つの他の種内の前記保存された隣接配列に隣接する前記配列が転写システムにおいて遺伝子転写調節特性を有するかどうかを決定することを含む、前記方法。

【請求項21】

前記配列が転写システムにおいて遺伝子転写調節特性を有するかどうかを決定することが、

- 250マイクロg/ml ゼオシン及び0.1ng/ml ドキシサイクリンの存在下で、5週間細胞を成長すること；

- 5週間後になお成長するコロニーを選択すること；

- 前記コロニーから単離されたプラスミド中の転写調節特性を有する前記配列の存在を検証すること；

を含み、

ここで前記細胞は、LexA DNA結合性ドメイン及びHP1又はHPC2のコード領域を含むLexA-リプレッサー融合タンパク質をTet-Off転写的調節システムの制御下で発現するU-2 OSヒト骨肉腫細胞株からのものであり、及び

ここで前記細胞は、4つのLexAオペレーター部位及びゼオシン耐性遺伝子を制御するSV40プロモーターの間に置かれた、前記プラスミド内のポリリンカー配列内にクローニングされた遺伝子転写特性を有する前記配列を含むプラスミドを含み、及び

ここでさもなければ前記ゼオシン耐性遺伝子は、もし転写調節特性を有する前記配列の存在がなければ、前記LexA-リプレッサー融合タンパク質の発現故に抑制されるであろう

、
請求項19又は請求項20に記載の方法。

【請求項22】

請求項1又は10～18のいずれか一項に記載の核酸配列又は修飾によってそれから得られる核酸配列を備えられたDNA構築物。

【請求項23】

目的の核酸に動作可能にリンクされたプロモーターを含む、請求項22に記載のDNA構築物。

【請求項24】

目的の核酸を有する前記プロモーターが請求項1若しくは10～18のいずれか一項に記載の核酸配列又は修飾によってそれから得られる配列によって隣接されている、請求項23に記載のDNA構築物。

【請求項25】

目的の核酸の転写を調節するために、請求項1若しくは10～18のいずれか一項に記載のDNA配列又はそれから得られる配列を使用する方法。

【請求項26】

細胞内で遺伝子生成物を生産するための方法であって、i) 目的の遺伝子及びii) 請求項1若しくは10～18のいずれか一項に記載のDNA配列又はそれから得られる配列を含む発現カセットを用意すること、及び細胞内で前記発現カセットの転写をさせることを含む、前記方法。

【請求項27】

目的の前記遺伝子が請求項1若しくは10～18のいずれか一項に記載のDNA配列又はそれから得られる配列によって隣接されている、請求項26に記載の方法。

【請求項28】

請求項23又は請求項24に記載のDNA構築物を含む宿主細胞。

【誤訳訂正2】

【訂正対象書類名】明細書

【訂正対象項目名】全文

【訂正方法】変更

【訂正の内容】

【発明の詳細な説明】

【発明の名称】遺伝子転写を調節する特性を含むDNA配列及びそのようなDNA配列を検出し且つ使用するための方法

【技術分野】

【0001】

本発明は、医薬及び細胞生物学の分野に関連する。本発明は、詳細には遺伝子転写の調

節の手段及び方法に関連する。本発明は、DNA配列が遺伝子転写を調節する特性及び／又は遺伝子転写を抑制する特性を含むかどうかを決定するための方法にさらに関連する。

【背景技術】

【0002】

様々なゲノムプロジェクトの発達に伴い、完全な生物体ゲノムの配列が利用可能になってきている。データの洪水は、多くの研究員の目的(interest)を引き起こしている。より注目され得る発見の1つは、ヒトゲノムはショウジョウバエ(fruit fly)のような簡単な生物体のゲノムよりも極めて多くの遺伝子をコードしていないという所見であった。多くの研究員の焦点は、遺伝子の同定から遺伝子発現及び遺伝子機能の決定に今シフトしている。そのような技術の例が、DNAマイクロアレイ、機能的ゲノミクス(genomics)利用及びプロテオミクス(proteomics)である。これらの技術は、それらがコードする配列の機能及び発現の周りに集中されていることで共通点を有する。しかしながら、我々の遺伝子の知識は劇的に増加する一方、遺伝子の発現がどのように制御されるかの理解は、この迅速に増加する知識に適用する能力を制限している。これは例えば、遺伝子組み換え植物及び動物の生成、並びにヒト遺伝子治療の場合である。これらの利用において、外来の核酸は、コードする配列の発現を得るために細胞内に典型的に導入される。細胞のゲノム内への外来の核酸のしばしばの組み込みは、導入された配列の長期にわたる機能のために必要とされる。しかしながら、ゲノム内への配列の組み込みは、発現の予測不可能性をもたらす。なぜならば、周りのDNAが、組み込まれた配列の転写に影響を及ぼすためである。この予測不可能性は、導入された配列が周りのDNAの効果に影響を及ぼす転写から組み込まれた配列を機能的に単離するために十分なゲノム情報をいまだ提供されることでできないという事実に一部拠るものである。他の部分では、これは、周りのDNAの効果に影響を与える転写について十分に知られていないという事実に拠る。

【発明の開示】

【発明が解決しようとする課題】

【0003】

本発明は、インシスで(in cis)遺伝子の転写に影響を与える能力を含むDNA配列に関連がある。典型的には、必要ではないけれども、調査された配列は機能的タンパク質をそれ自身でコードしていない。インシスで遺伝子転写に影響を与える能力を有する種々の配列エレメントが、同定されてきている。これらのエレメントは、プロモーター、エンハンサー及びサイレンサーから境界エレメント(boundary element)及びマトリックス結合領域(matrix attachment regions)に及んでいる。

【0004】

非常に多くの種々の型の調節塩基配列(regulatory sequences)が発見されてきているという事実は、有効な発現力セットを設計することが非常に簡単であるという印象を与える。しかしながら、全く逆が真実である。発現力セットの設計は、試行錯誤によって今なおしばしばおこなわれている、ターゲット細胞又はその後代において外来遺伝子のある種の発現を得ることを可能にすることはかなり頻繁に可能である。しかしながら、非常にしばしば、発現力セットがターゲット細胞において表わすことができる発現のレベル又は発現の持続性を、何らかの正確性で予測することは困難である。

【課題を解決するための手段】

【0005】

本発明はなかんずく、新規の転写調節エレメントを検出し且つ単離するための手段及び方法を提供する。遺伝子転写を調節する特性を有するDNA配列を検出し、任意的に選択するための方法であって、転写システムに様々なフラグメント含有ベクターを備えることを含み、前記ベクターは、i) 遺伝子転写を抑制する特性を有するエレメント、及びii) レポーター遺伝子の転写を指示するプロモーターを含み、前記方法は前記遺伝子転写を調節する特性を有する前記DNA配列を識別するために前記転写システムにおいて選択ステップを実行することをさらに含む方法が提供される。好ましい実施態様では、前記フラグメントは、i) 遺伝子転写を抑制する特性を有する前記エレメント、及びii) 前記レポー

ター遺伝子の転写を指示する前記プロモーターの間に配置される。RNAポリメラーゼは、どこでRNA合成が開始すべきかを合図するプロモーターと呼ばれる特定の配列に結合後、転写工程を開始する。調節する特性は、所定の細胞型及び/又は所定のプロモーターにおいてインシスで前記プロモーターからの転写を助長することができる。同一のDNA配列は、1つの型の細胞において又は1つの型のプロモーターとともに、助長する特性を含むことができる。一方、それは他の細胞において又は他の型のプロモーターとともに、遺伝子転写を調節する他の特性を含むことができるか又は含まない。転写は、特定のプロモーターの転写に対する調節エレメント(又はそれに結合するタンパク質)の直接の影響を通じて影響されうる。しかしながら、転写は、間接的な影響によってもまた影響されうる。例えば調節エレメントが、1つ又はそれ以上の他の調節エレメントの機能に影響を与えるからである。遺伝子転写を調節する特性は、安定な(stable)遺伝子転写特性をもまた含みうる。安定とは、観察された転写レベルが少なくとも30細胞分割の間、有意には変化されないことを意味する。安定な特性は、発現特徴が多くの細胞分割にわたって予測可能であるべきところの状況において有益である。典型的な例は、外来遺伝子でトランسفエクションされた細胞株(cell line)である。他の例は、形質転換動物及び植物並びに遺伝子治療である。導入された発現カセットが細胞分割若しくは植物又は動物世代の数の増加後に夫々に非常にしばしば違うように機能する。好ましい実施態様では、安定な特性は、形質転換植物又は動物の後の世代において遺伝子転写を維持する能力を含む。もちろん、発現が誘発可能である場合、前記特性は形質転換植物又は動物の後の世代において発現の誘発性を維持する特性を含む。しばしば、発現レベルは、細胞分割の数の増加に伴い劇的に落ちる。本発明の方法では、細胞分割の数の増加につれて転写レベルの劇的な落ち込みを少なくとも部分的に防ぐことが可能であるDNA配列を検出し、任意的に選択する、ことが可能である。従って、好ましい実施態様では、前記遺伝子転写を調節する特性は、安定な遺伝子転写特性を含む。驚くべきことには、前記方法は転写の長期的な安定性を必ずしも測定しないという事実にもかかわらず、前記安定な遺伝子転写特性を有するDNA配列を含むフラグメントは、本発明の方法で検出され且つ任意的に選択され得る。本発明の好ましい実施態様では、前記遺伝子転写を調節する特性は、遺伝子転写を助長する安定な特性を含む。目的の遺伝子を有する発現ベクターへの遺伝子転写を調節する特性を有するDNA配列の組み込みは、細胞のゲノムへの発現ベクターの組み込みに応答して目的の前記遺伝子のより高いレベルの転写を生じること、且つその上、前記より高い遺伝子発現レベルはまた、遺伝子転写を調節する特性を有する前記DNA配列の不存在におけるよりもより安定であるということが観察されている。

【0006】

細胞のゲノム内に目的の遺伝子を導入するために且つ目的の前記遺伝子の発現を取得するために設計された実験において、次のことが観察されている。目的の前記遺伝子とともに遺伝子転写を調節する特性を有するDNAがまた導入された場合、前記DNA配列が目的の前記遺伝子とともに導入されないときよりも、目的の前記遺伝子の遺伝子生成物を或る量よりも多く発現したクローンが、より多く検出された。すなわち、本発明は、前記細胞のゲノムに目的の前記遺伝子を備えると目的の遺伝子の遺伝子生成物を有するレベルよりも多く発現する細胞の数を増加する方法をまた提供し、該方法は、目的の前記遺伝子とともに遺伝子転写を調節する特性を含むDNA配列を前記細胞に備えることを含む。

【0007】

遺伝子転写を調節する特性を有するフラグメントを検出することの機会は、フラグメントがそこから得られるところの起源で変わる。典型的には、前記特性を有するフラグメントの存在又は不存在についての先行する知識がない。それらの状況において、多くのフラグメントは、遺伝子転写を調節する特性を有するDNA配列を含まない。これらの状況において、前記特性を有するDNA配列のための正式な選択ステップが導入される。これは、前記レポーター遺伝子の生産物の特性に基づき前記配列を含むセレクションベクターによって行われ、これはそれに有利にも不利にも選択されうる。例えば、前記遺伝子生成物は、蛍光若しくは色沈着(例えば、緑色蛍光タンパク質及びその誘導体、ルシフェラーゼ、又

はアルカリフォスファターゼ)を誘発し、又は抗生物質耐性を与え、若しくはアポトーシス及び細胞死を誘発しうる。

【0008】

本発明の方法は、遺伝子転写を助長する特性を含むDNA配列を検出し、且つ任意的に選択する、ために特に適している。目的の遺伝子を含む発現ベクターに組み込まれると、ベクターが遺伝子転写を抑制する特性を有するエレメントを含まなくとも、選択されたDNA配列の少なくとも幾つかは宿主細胞中で目的の前記遺伝子の遺伝子転写を劇的に増加することができる。この遺伝子転写を助長する特性は、外来遺伝子でトランسفェクションされた細胞株において又は形質転換動物及び植物において非常に有用である。

【0009】

前記転写システムは、細胞フリーのイン ビトロ (*in vitro*) 転写システムでありうる。自動化された現在の専門的技術では、そのような細胞フリーシステムは、正確且つ迅速でありうる。しかしながら、本発明のために、前記転写システムは宿主細胞を好ましくは含む。宿主細胞を使用することは、フラグメントが細胞内で活性により検出され且つ任意的に選択されることを保証する。

【0010】

本発明の方法において、遺伝子転写を抑制する特性を有するエレメントは、使用される転写システムにおいてプロモーターからの転写を抑制する。前記抑制は、検出不可能な発現レベルにまで進む必要はない。重要なことは、抑制の不存在又は存在において発現レベルの相違が、検出可能であり且つ任意的に選択可能であるということである。好ましい実施態様では、前記ベクターにおける遺伝子転写抑制は、遺伝子転写を抑制するクロマチンをもたらす。この好ましい実施態様では、遺伝子転写を抑制するクロマチンの形成を少なくとも部分的に妨げることが可能であるDNA配列が検出され且つ任意的に選択されうる。1つの観点では、遺伝子転写を抑制するクロマチンの形成を少なくとも部分的に妨げることが可能であるDNA配列は、安定な遺伝子転写特性を含む。好ましい実施態様では、遺伝子転写抑制に関するDNA配列が、タンパク質複合体によって認識されるDNA配列であり、及び前記転写システムが前記複合体を含む。好ましくは、前記複合体は、HP1を含むヘテロクロマチン結合性タンパク質 (heterochromatin-binding protein)、ポリコーム - グループ (Pc-G) タンパク質、ヒストン脱アセチル化酵素活性又はMeCP2 (メチル - CpG - 結合性タンパク質) を含む。多くの生物体が、1つ又はそれ以上のこれらのタンパク質を含む。これらのタンパク質はしばしば、他の種においても活性を示す。従って、前記複合体は、2つ又はそれ以上の種からのタンパク質をまた含むことができる。既知のクロマチンに関連付けられたタンパク質複合物の前記されたセットは、多くの塩基対にわたる長い範囲の抑制をもたらすことができる。前記複合体は、細胞分割すると娘細胞に遺伝子の抑制された状態を安定に移すことによつて関係する。この方法で選択された配列は、多くの塩基対にわたる長い範囲の抗抑制をもたらすことができる (ファン デル フラッグら、2000年)。

【0011】

使用されるベクターは、DNAクローニングに適切であり且つ転写システムで使用されうる任意のベクターであり得る。宿主細胞が使用される場合、該ベクターが、エピソーム的に複製するベクターであることが好ましい。このようにして、ベクターの組込みの異なる部位による効果は避けられる。組込みの部位でベクターに隣接する (flanking) DNAエレメントは、プロモーターの転写のレベルに対して影響を与える且つそれによって、遺伝子転写を調節する特性を有するDNA配列を含むフラグメントの影響をまねくことができる。好ましい実施態様では、前記ベクターが、イプシュタイン - バール (Epstein-Barr) ウィルス (EBV)、OriP、及び核抗原 (EBNA-1) からの複製起源を含む。そのようなベクターは、適切な条件下で真核生物の細胞の多くの型において複製することができ且つクロマチン内に組入れる (assemble) ことが可能である。

【0012】

他の観点では、本発明は、i) 植物若しくは脊椎動物から単離されたDNA配列又はそれ

らの誘導体、又は i i) 合成DNA配列すなわち遺伝子エンジニアリングによって構築されたDNA配列を含むDNA配列であって、本発明に従う方法によって検出され、選択され及び任意的にクローニングされ得る、抑制を阻害する配列であるところの前記DNA配列を提供する。他の観点では、本発明は、 i) 植物若しくは脊椎動物から単離されたDNA配列又はそれらの誘導体、又は i i) 合成DNA配列すなわち遺伝子エンジニアリングによって構築されたDNA配列を含むDNA配列であって、本発明の方法によって検出され、選択され及び任意的にクローニングされるところの前記DNA配列を提供する。好ましくは、前記DNA配列は、表4Aに記載されている配列又はそれらの機能的相同物を含む。表4に記載されている配列の機能的相同物は、表4(表4A又は表4Bである)に与えられた情報で導出された配列である。例えば、表4に示された配列に又は配列から塩基を削除し、修飾し及び/又は挿入することによって表4における配列から導出されうる配列であり、ここで前記導出された配列は表4に記載されている配列の種類と同一の活性(しかし、必ずしも同じ量ではない)を含む。機能的相同物は、表4に記載されている配列の2又はそれ以上からの部分を含む配列である。合成DNA配列は、生物体に存在する配列から直接的に又は間接的に導出されない配列である。例えばショウジョウバエscs又はscs'配列を含む配列は、該scs又はscs'配列が人工的に生成されたときであっても合成配列ではない。

【0013】

1つの観点において、本発明はより高次の遺伝子調節の知識を増加すること並びにこの知識を利用するための手段及び方法に関連がある。シングル遺伝子の転写を指示し且つ調節するエレメント例えば古典的なプロモーター及びエンハンサーは特性が調べられているが、一方、染色体領域全体の遺伝子転写能力を支配するより高位の調節エレメントはまだ殆ど注意が払われていない。そのような高位エレメントに関連する我々の知識の殆どは、胚形成の研究からきている。胚形成の間、細胞は種々の発達経路に委ねられる。委ねられた後には、多くの細胞分割の後でさえも細胞はそれらの運命をめったに変化しない。

【0014】

細胞型特異的遺伝子転写パターンの安定な伝達はプロモーターの文脈に依存しないが、その代わりにDNAの構造における変化に及びクロマチンと称される関係するタンパク質によって調整されることがだんだんに判ってきた。染色体レベルでの遺伝子調節は、DNAの修飾(例えば、メチル化)、ヒストンの修飾(例えばアセチル化及び又はメチル化)、及び離れた染色体のエレメント間での長い範囲の相互作用を包含する。

【0015】

クロマチンテンプレートは、DNA、ヒストン及び非ヒストンタンパク質の高度に凝縮された複合体であり、それは核内にゲノム全体をパッケージすることができ且つ同時に特定の遺伝子の適切な転写を可能にする。真核細胞染色体は、遺伝子転写の活性のための均一のテンプレートではない。クロマチン及びクロマチン領域の種々の型は区別され得、それらは遺伝子転写に区別をつけて影響を与える。いわゆるヘテロクロマチン(heterochromatin)領域は「閉ざされた(closed)」クロマチン構造を識別し、一方、真正クロマチン(euchromatin)はより拡散し且つ「開いた(open)」クロマチン構造に関連する。真正クロマチン領域は、構造的な変化に付されることができ、条件的ヘテロクロマチン(facultative heterochromatin)及び条件的真正クロマチン(facultative euchromatin)と言われる多かれ少なかれ凝縮された構造を生じる。条件的真正クロマチン又は条件的ヘテロクロマチンの形成は、クロマチンが仲介する遺伝子調節の潜在する機構を表わすと信じられており、遺伝子を細胞型特異的な様式で活性な又は抑制された状態に保持する。

【0016】

全ての真核生物において、幾つかのクロマチンに関連付けられたタンパク質複合体が識別されており、それは細胞型特異性の維持に關係し、それらの1つはポリコーム・グループ(Pc-G)複合体である。Pc-G複合体は、遺伝子の安定な抑制に關係し、そこではクロマチン構造における変化は重要な役割を果たすと信じられている。同様に、トリソラックス(trithorax)グループ(TrG)と名付けられたタンパク質の第2のクラスが識別されており、それはPcGタンパク質の活性を妨げる。TrGタンパク質は、遺伝子の維持に關係する。

それ故に、それらの個々の活性のモードに基づき、PcG及びTrGタンパク質は、遺伝子転写パターンの遺伝性の伝達にとって重要である細胞のメモリーシステムを表わす。

【0017】

PcG及びTrG複合体がそれらの目標遺伝子にどのように関連付けられているかは、いまだ不明である。遺伝子研究は、遺伝子の転写的に不活性な状態を維持するシス(cis)活動調節配列(regulatory sequence)の特徴を調べた。これらのシス活動調節配列によって仲介されるサイレンシング(silencing)は、機能的PcGタンパク質の存在に依存し、それ故にこれらの配列はPcG応答エレメント(PcG response elements; PRE)と名付けられている。クロマチンのPcG仲介された抑制に関する配列は識別されている。しかしながら(脊椎動物又は植物のいずれにおいても)クロマチンの抑制を仲介するために要求される全ての配列情報を含む完全なPREはまだ見つかってはいない。

【0018】

加えて、コヒーレントな様式(coherent way)で長い範囲の抑制能力を有する配列を研究することはまだ可能ではない。これは、そのような長い範囲の活動配列を体系的にスクリーンすることが不可能であることに大部分拠る。1つの観点では、本発明はDNAにおけるそのような配列を体系的に検出するための手段及び方法を提供する。1つの実施態様では、本発明は、遺伝子転写抑制特性を有するDNA配列を識別するための方法であって、

テスト核酸のコレクションを用意すること、

プロモーターの転写制御下にあるテスト核酸及び第1のレポーター遺伝子を含む発現ベクターのコレクションを生成すること、

細胞に発現ベクターの前記コレクションを備えること、

それらの後代を含む細胞又はベクターを選択すること、ここで前記第1のレポーター遺伝子の転写は抑制されている、及び

前記細胞中の前記テスト核酸を識別すること、

を含む。前記識別されたテスト核酸は、前記プロモーター機能を抑制する能力を有し、従って遺伝子転写を抑制する特性を有する。好ましくは、前記識別されたテスト核酸はまた回収され且つクローニングされる。前記特性を有するDNA配列の不存在下におけるレベルと比較して、前記プロモーターに物理的にリンクされた場合、前記特性は前記プロモーターからの転写のレベルを減少する能力を少なくとも一部分含む。好ましい実施態様では、前記遺伝子転写抑制特性は、遺伝子転写を抑制するクロマチン特性を有する。すなわち、転写のレベルの前記減少は、遺伝子転写を抑制する構成を有するクロマチンの結果である。この構成は好ましくは、前記プロモーターを包含する。しかしながら、前記構成がエンハンサーなどを包含することもまた可能であり、それによって前記プロモーター上で前記エンハンサーの転写を助長する効果を少なくとも一部分不活性にする。特に好ましい実施態様では、遺伝子転写を抑制するクロマチン特性を有する前記DNA配列はポリコーム・グループ様応答エレメントを含む。

【0019】

上記に記載された方法を使用して、プロモーターからの転写のレベルを減少する能力を含む幾つかの核酸配列及び従って遺伝子転写を抑制する特性を含む核酸配列を検索することが可能である。遺伝子転写を抑制する特性を有するエレメント例えばポリコーム・グループ様応答エレメントの1又はそれ以上のコンセンサス配列が推論されうるような配列類似性について、類似の機能を有する配列が互いに比較されうる。その上、生命体のゲノムの全体の配列が知られ且つ簡単に理解されることを考えると、これらのゲノム又はそれらの一部をスクリーンすること、及びゲノムにおけるこれらの配列の存在を予測することは可能である。ゲノム中の遺伝子転写を調節する特性及び/又は遺伝子転写を抑制する特性を含むDNA配列の存在及び配置の知識は、ゲノムにおける遺伝子転写のより高位の調節の我々の知識を大いに増加する。

【0020】

ポリコーム・グループ応答エレメントは、前記エレメントを有する1つ又はそれ以上のポリコーム・グループタンパク質の直接的な及び/又は間接的な相互作用に応答してプロモ

ーターからの転写を抑制することができるエレメントである。ポリコーム・グループ様応答エレメントは、ポリコーム・グループ応答エレメント又は代替的には前記エレメントを有する1つ又はそれ以上のタンパク質の直接的な及び／又は間接的な相互作用に応答してプロモーターの転写を抑制することができるエレメントである。ここで、前記1つ又はそれ以上のタンパク質は、ポリコームグループに属しないが、前記相互作用の結果、遺伝子転写を抑制するクロマチンが形成される。そのようなタンパク質の例は、クロマチンに関連付けられたタンパク質例えばヘテロクロマチンタンパク質1(HP1)(アイゼンベルグら、1990年)である。遺伝子活性を抑制するクロマチンに関連付けられた他のタンパク質は、メチル-CpG結合性タンパク質、MeCP2(ナンら、1997年)である。好ましい実施態様では、本発明のポリコーム・グループ様応答エレメントは、長い距離好ましくは2000塩基対よりも多くに亘ってプロモーターの転写を抑制する能力を有する(ファン・デル・フラッグル、2000年)。

【0021】

テスト核酸のコレクションは、多くの方法で生成され得る。テスト配列として人工的な配列を使用し、遺伝子転写を抑制する特性のためのコンセンサス配列を得ることが可能である。異なる特性は異なるコンセンサス配列を含むことができる。好ましくは、前記コレクションは染色体のDNAから生成される。このようにして、染色体において自然に生じる配列を含む遺伝子転写を抑制する特性が見つけられる。これは、染色体におけるこれらの特性の位置が、前記位置におけるより高位の遺伝子転写に対するそれらの影響が消滅するとすぐに決定されうるという有利点を有する。

【0022】

レポーター遺伝子は、その存在が細胞において直接的に又は間接的に検出されうるところの発現生成物をコードする遺伝子である。遺伝子転写を抑制する特性を検出するための方法では、細胞内への発現ベクターの転移(transfer)は、前記レポーター遺伝子の発現を導出する。しかしながら、テスト核酸が遺伝子転写を抑制する特性例えばポリコーム・グループ様応答エレメントを含む場合、発現は前記細胞において抑制され、それによって前記レポーター遺伝子の少なくとも部分的に減少した発現を導出する。従って、前記プロモーターの転写を抑制することが可能な核酸の存在又は非存在は、前記細胞における前記発現生成物を検出することによって検出され得、それによって減少した検出又は非検出が遺伝子転写を抑制する特性の存在を示す。レポーター遺伝子は、蛍光レポータータンパク質をコードし得る。すると、減少した発現は蛍光手段によって例えばフローサイトメーター(flow cytometer)において検出され得る。蛍光を示さない又は低い蛍光を示す細胞は、蛍光活性化細胞ソーター(a fluorescence activated cell sorter)及び発現ベクター及び／又は例えば增幅反応によって単離されたテスト核酸を使用しソートされ得る。好ましくは、前記第1のレポーター遺伝子は選択可能なレポーター遺伝子を含み、その発現は前記細胞に前記第1のレポーター遺伝子の低いレベルを発現しない又は低いレベルで発現する細胞全体にわたって少なくとも成長不利点を直接的に又は間接的に備える。遺伝子転写を抑制する特性を有するDNA配列についてスクリーンされる場合、前記第1レポーター遺伝子の発現は好ましくは直接的に又は間接的に前記細胞に対して毒性である。そのような毒性発現生成物の非限定的な例は、リシン(ricin)又はそれらの毒性変異体である。他の例では、前記第1のレポーター遺伝子は、アポトーシスを誘発する遺伝子生産物をコードする。好ましくは、前記アポトーシスを誘発する遺伝子生産物は、アデノウィルス(adenovirus)13S E1A又はそれらの機能的等価体である(ブレッケンリッジ及びショール、2000年)。他の実施態様では、前記アポトーシスを誘発する遺伝子生産物は、アポプチン(apoptin)又はそれらの機能的等価体を含む(ピーターセン及びノートボーン、2000年)。

【0023】

他の例はいわゆる自殺遺伝子生産物例えば単純ヘルペスウィルスチミジンキナーゼ(HSV-tk)をコードする遺伝子である。HSV-tkを発現する細胞の培地へのガンシクロビル(gancyclovir)の添加は、これらの細胞において毒性物質の形成を生じ、それ故にこれらの

細胞を殺す。特に好ましい実施態様では、前記自殺遺伝子はシトシンデアミナーゼ (cytosine deaminase) を含む。シトシンデアミナーゼはシトシンをウラシルに転化する。この酵素活性は、原核生物及び低度の真核生物において見つけられ、しかし高度の真核生物においては存在しない。該遺伝子は、プロドラッグ 5 - フルオロシトシン (5-FC) と組み合わせて代謝性自殺遺伝子として使用される。シトシンデアミナーゼは、非毒性 5 - FC を 5 - フルオロウラシルに転化することができ、それはDNA合成を妨害することによって細胞を殺し、それによってアポトーシスを引き起こす (ミューレンら、1992年；ウェイ及びフーバー、1996年)。

【 0 0 2 4 】

前記第1のレポーター遺伝子の転写を制御するプロモーターは、活性である又は前記細胞内で活性化され得る任意のプロモーターであり得る。特定のプロモーターを選択することによって、遺伝子転写を抑制する特性例えば前記特定のプロモーターの転写を抑制することができるポリコーム - グループ様応答エレメントを選択することが可能である。このようにして、前記特定のプロモーターが属するところのプロモーターのクラスを特異的に抑制する特性を選択することが可能である。好ましい実施態様では、前記プロモーターは、これを含む細胞にシグナルを備えると、前記プロモーターの活性が誘発されうるところのプロモーターを含む。そのような誘発性プロモーターは好ましくは、テトラサイクリン応答性プロモーターを含む。該シグナルはここでは、テトラサイクリン、ドキシサイクリン及び等価な化合物である。そのようなプロモーターは、真核生物細胞においてテトラサイクリン応答性にまた適合されうる (インら、1996年)。テトラサイクリン又はそれらの等価体を添加すると遺伝子の発現を誘発し又は抑制するプロモーター及び処理する分子は入手可能である。

【 0 0 2 5 】

本発明の発現ベクターでトランスフェクションされた細胞は、典型的に低い頻度で且つ遺伝子転写を抑制する特性を有するDNA配列の存在に関連付けられない理由のために、前記第1のレポーター遺伝子の発現生成物の検出可能な量を発現しない。これは例えば、前記第1のレポーター遺伝子のコード配列を破壊する組み換え事象に拠ることができる。本発明の好ましい実施態様では、発現ベクターの前記コレクションは、第2のレポーター遺伝子をさらに含む。前記第2のレポーター遺伝子の発現は好ましくは、第2のプロモーターの制御下である。前記第2のレポーター遺伝子の発現生成物の発現の検出のための方法は、前記テスト核酸の発現を抑制する活性を確認するために使用され得、それによって前記第1のレポーター遺伝子を誤って表わさない細胞の数の少なくとも一部を減少する。好ましい実施態様では、前記第2のレポーター遺伝子は発現力セットを含む細胞を選択するために使用される。このようにして、前記発現力セットを含まない細胞は、容易に無視されうる。この目的を達成するために、前記第2のレポーター遺伝子の発現生成物は好ましくは、ポジティブ優性選択可能なレポーター遺伝子を含む。好ましくは、前記ポジティブ優性選択可能なレポーター遺伝子は、さもなくば毒性化合物に対する耐性を付与することが可能な発現生成物をコードする。非限定的な例は、G418耐性及びハイグロマイシン耐性である。

【 0 0 2 6 】

遺伝子転写を抑制する特性は転写を抑制することができることを考慮すると、この実施態様では、発現ベクターは、遺伝子転写を調節する特性を有して、遺伝子転写を抑制する特性を有するDNA配列の転写抑制効果を妨害することを出来る少なくとも1つのDNA配列をさらに含むことが好ましい。発現ベクターにおける前記転写を妨害するエレメントの配置は好ましくは、それが前記第2のレポーター遺伝子の転写のレベルに対する前記遺伝子転写を抑制する特性の低減する効果を効率的に邪魔するようなものである。好ましい実施態様では、遺伝子転写を調節する特性を有する前記DNA配列は、前記第1のレポーター遺伝子及び前記第2のレポーター遺伝子を含む発現力セットを機能的に分離する。好ましくは、前記第2のレポーター遺伝子 (及び前記第2のレポーター遺伝子の転写を制御するプロモーター) は、遺伝子転写を調節する特性を有するDNA配列によって隣接される。遺伝子

転写を調節する特性を有するDNA配列の例は、表1及び表2にリストされたいわゆるSTARエレメントである。

【0027】

本発明の方法は、遺伝子転写を調節する特性及び／又は遺伝子転写を抑制する特性を有する多くのエレメントのクローニング及び識別をもたらす。そのようなエレメントは、前記特性を実行する際に役立たない例えは遺伝子転写を抑制するクロマチンの形成に関係しない無関係な核酸を含んでもよい。そのようなエレメントにおける機能的配列は、当業者に知られている種々の方法によって描写され得る。1つの実施態様では、削除及び／又は置換は、遺伝子転写を調節する特性又は遺伝子転写を抑制する特性を有するDNA配列においてなされる。そのようにして修飾されるDNAは、本発明の方法で活性の試験をされる。これは、シングル修飾された核酸を使用し又は前記修飾された核酸を含むテスト核酸のコレクションを生成することによって行われうる。本発明のDNA配列内の機能的配列の説明は、遺伝子転写を調節する特性及び／又は遺伝子転写を抑制する特性を有するエレメントのためのコンセンサス配列の説明を可能にする。それぞれが種々の機能性及び発現パターンを含む幾つかのポリコーム・グループ様複合体があることを考慮すると、コンセンサス配列の1より多い型が本発明の方法で見つけられることが予想される。同じように、コンセンサス配列の1より多い型が遺伝子転写を調節する特性を含むエレメントのために見つけられることが予想される。従って、本発明は、遺伝子転写を調節する特性及び／又は遺伝子転写を抑制する特性を有する単離された及び／又は組み換えられた核酸例えはポリコーム・グループ様応答エレメントのライブラリーをさらに提供する。1つの実施態様では、前記ライブラリーは、同一のコンセンサス配列を含む単離された及び又は組み換えられた核酸を含む。好ましい実施態様では、前記ライブラリーは、コンセンサス配列の1より多い型を含む。前記ライブラリーは例えは、所与のDNA分子がDNAを調節する特性を有するかどうかを決定するために使用されうる。好ましい実施態様では、前記ライブラリーは、遺伝子転写を助長する機能を有する全てのエレメントを本質的に含み、該エレメントは、染色体の安定な遺伝子転写特性及び／又は遺伝子転写抑制特性を有するエレメント例えはポリコーム・グループ様応答エレメントを含む。染色体上のこれらのエレメントの配置に対する知識とともに、このことは前記染色体上に自然に存在する遺伝子の遺伝子発現のより高位の調節のための及び組み換え手段によって前記染色体中に導入された遺伝子（外来核酸）のための予測を行うことを当業者に可能にする。そのような予測は例えは、外来DNAの挿入のために前記染色体上に安定な候補位置を選択するために使用されうる。適切な位置は、或る細胞、細胞型及び／又は組織において特異的に発現されることを予想される位置であり得る。好ましくは、前記染色体は、21番染色体又は22番染色体を含む。特に好ましい実施態様では、細胞において遺伝子転写を調節する特性又は遺伝子転写を抑制する特性を有する全てのDNA配列が、ライブラリーにある。この実施態様では、ゲノム全体が、適切な候補位置の予測のために使用されうる。1つの実施態様では、前記ライブラリーは、植物からヒトに亘る種々の異なる細胞株において生成されている。種々の細胞株及び／又は種(species)において、遺伝子転写を抑制する特性を有するDNA配列と相互作用することが可能な種々のタンパク質（又はタンパク質複合体）が発現され、遺伝子転写を抑制する特性を有する種々のDNAエレメントをもたらす。同様に、遺伝子転写を調節する特性を有するDNA配列と直接的に又は間接的に相互作用する種々のタンパク質が発現されるであろう。それ故に、ライブラリーの作成は、細胞型依存性であり且つ関連したタンパク質の存在に依存する。これはまた、ポリコーム・グループ様応答エレメントの場合もある。HP1が細胞型1(cell type one)において発現される場合、HP1に依存するエレメントは本発明の方法によって検出されるであろう。HP1が細胞型2(cell type two)において発現されない場合、本発明の方法は、細胞型1から回収されたエレメントを検出しないであろう。

【0028】

本発明の1つの観点では、前記ライブラリーは、遺伝子転写を抑制するクロマチンの形成を少なくとも一部を妨げることが可能な少なくとも一つのエレメントを含む。染色体又

はゲノム上の遺伝子転写を抑制する特性を有するDNA配列の位置の知識とともに、そのように妨げるエレメントの位置の知識は、前記染色体又はゲノムにおいて（挿入された）遺伝子の遺伝子転写のより高位の調節のより正確な予測を可能にする。好ましくは、前記ライブラリーは他の転写調節エレメント例えばエンハンサー及びサイレンサーをさらに含む。そのような配列はより高位の遺伝子調節に限定的影響を有しているけれども、そのような他の配列の位置についての情報は、その中に導入された外来配列の発現のためのゲノムにおける適切な位置についての予測の正確性をさらに増加する。好ましくは、前記ライブラリーは、遺伝子転写を調節する特性を有する全てのDNA配列及び／又は染色体の全ての他の調節配列を本質的に含む。

【0029】

染色体がすでに何千万の塩基から典型的に成ることを考慮すれば、ライブラリーが高位の遺伝子調節を与えるところの情報は少なくとも部分的に自動化されたシステム内に組み入れられることが好ましい。

【0030】

本発明のライブラリーの他の使用は、染色体上の配列の目的の修飾、たとえば、「より高位の」調節配列が変異された後の遺伝子の転写に対する予測の生成である。例えば、本発明の1つ又はそれ以上のポリコーム・グループ様応答エレメント、及び／又は前記染色体上の他の調節エレメントが変異されうる。このことは、ポリコーム・グループ様応答エレメント及び／又は他の発現を調節するエレメントの近傍にある遺伝子の転写レベルを変化させると予期される。

【0031】

本発明のライブラリー又はシステムのまた別の使用は、ゲノムにおける変異から生じる遺伝子発現の予測である。変異が変更された遺伝子転写を生じる場合には、そのような変更された遺伝子転写の検出は、前記自然に生じる変異の存在を示すことができる。このアプローチは、例えば診断用アッセイにおいてテストされるべき配列又はタンパク質の数を制限する際に使用される。このことは、マイクロアレイアプローチにおいて特に重要である。なぜならば、これらのアプローチにおいて、試験されるべき発現された配列の数はアレイが最大に保存できる配列の数によって制限されるからである。本発明の手段及び方法を用いて、マイクロアレイアプローチにおいてテストされるべき配列の数を制限することが可能である。

【0032】

本発明のシステム又はライブラリーのまた別の使用は、薬のターゲットの発見である。調整エレメント（それらが「より高位の」エレメントか又はそうでなくとも）は、それらに結合できる蛋白質（複合体）の故に機能する。本発明のシステムは、特定のタンパク質（複合体）の結合を又は機能を妨げないための薬のターゲティングが特定の遺伝子の代替の発現の見込みを有するかどうかを決定するために使用されうる。

【0033】

本発明は、本発明の方法によって取得可能な遺伝子転写を抑制する特性を有するDNA配列をさらに提供する。好ましい実施態様では、遺伝子転写を抑制する特性を有する前記DNA配列は、脊椎動物又は植物から得られる。より好ましくは、遺伝子転写を抑制する特性を有する前記DNA配列は、表4Bに従う配列又はそれらの機能的相同体を含む。DNA構築物に本発明のDNA配列を備えること又はそのようなDNA配列を修飾することもまた可能である。好ましい実施態様では、目的の核酸に動作可能にリンクされたプロモーターを含むDNA構築物が用意される。好ましくは、遺伝子転写を調節する及び／又は抑制する特性を有する前記DNA配列の活性の特性の量が、前記プロモーターと比較して前記構築物内での前記DNA配列の方向に依存する。好ましくは、前記遺伝子転写を調節する及び／又は抑制する特性がシグナルの存在に依存する。好ましくは、前記シグナルがDNA結合性タンパク質（DNA-binding protein）を含む。好ましくは、前記シグナルがヒト免疫不全ウィルスTATタンパク質（a human immuno-deficiency virus TAT protein）を含む。

【0034】

遺伝子転写を調節する特性又は遺伝子転写を抑制する特性を有するDNA配列の使用の一つは勿論、目的の遺伝子の転写の調節である。目的の遺伝子の転写は、前記特性を有するDNA配列が備えられ又は除去されるように前記遺伝子の近傍において配列を変更することによって変更されうる。特定の発現特徴は、遺伝子転写を調節する特性及び／又は遺伝子転写を抑制する特性を有するDNA配列（の一部）を結合することによって設計されうる。例えば、発現ベクターにおける安定な遺伝子転写特性を有する配列の複製は、ターゲット細胞における前記ベクターの導入に応答して、前記ターゲット細胞又は後代における発現の改善された安定性を導出する。遺伝子転写を調節する特性を有するDNA配列を結合することによって、変更された遺伝子転写調節特性は、種類若しくは量又は両方で生成されうる。

【0035】

望ましい遺伝子転写を調節する特性及び／又は遺伝子転写を抑制する特性を有するDNA配列を設計することもまた可能である。他のタンパク質及びDNA配列とともにDNA結合性タンパク質は、DNA配列の特性を決定する。1つ又はそれ以上の他のタンパク質結合DNA配列を、ある特性を有するDNA配列内に挿入することが可能である。結合性タンパク質の結合を可能にすることによって、該特性を妨げる又は指示することが可能であり、その結果設計者の特性を有するDNA配列の生成を可能にする。特定の遺伝子転写を調節する特性及び／又は遺伝子転写を抑制する特性を有するDNA配からタンパク質結合部位を除去することも勿論また可能であり、それによって生じたDNA配列の特性を変更する。付加と除去の組み合わせが、また可能である。特定の遺伝子転写を調節する特性及び／又は遺伝子転写を抑制する特性は、本発明において記載された検出方法を調整することによって選択可能であり得る。例えば、誘発可能な遺伝子転写調節特性及び／又は遺伝子転写抑制特性を有するDNA配列を合成することが可能である。例えば遺伝子転写を抑制する特性を有するDNA配列にTAT結合性エレメントを含めることによって、TATを含む細胞において遺伝子転写を抑制する特性を少なくとも一部不活性にすることが可能である。同様に、シグナルの不存在下又は存在下においてそれらのターゲット配列にのみ結合する利用可能なDNA結合性タンパク質がある。そのようなタンパク質の非限定的な例は、TETリプレッサー及びそれらの種々の変異体、lac-リプレッサー、ステロイドホルモン受容体、レチノイン酸受容体及び誘導体である。細胞型特異的な遺伝子転写調節特性及び／又は遺伝子転写抑制特性を有するDNA配列を設計することが例え可能である。例えば、上記したTAT例の場合である。言及されるDNA配列は、TATを発現するHIV感染した細胞に特異的にされうる。あるいは、DNA配列は細胞型特異的な様式で発現されるタンパク質複合体に特異的に作成されうる。

【0036】

遺伝子転写を調節する特性及び／又は遺伝子転写を抑制する特性を有するDNA配列を含む発現構築物は、前記発現構築物の1より上のコピーを含む細胞内で前記構築物から発現を得るために適切である。発現構築物が前記細胞のゲノムにおいて存在する場合及びまた発現力セットが前記細胞中において1より上のコピーで存在する場合もまたである。その上、それらは、1つより上のコピーにおいて同一の位置内に統合された場合にさえ働く。

【0037】

本発明の好ましい実施態様では、遺伝子転写を調節する特性を有する前記DNA配列は、いわゆるSTAR (Stabilizing Anti-Repression : 安定化抗抑制) 配列を含む。本明細書中で使用されるSTAR配列という語は、記載された遺伝子転写を調節する特性の1又はそれ以上を有するDNA配列を言う。

【0038】

いくつかの方法が、ある共通の特性を共有するDNA配列のファミリーから配列識別子を抽出するために当業者にとって利用可能である。そのような配列識別子は次に、1つ又はそれ以上の識別子を共有する配列を識別するために使用され得る。そのような1つ又はそれ以上の識別子を共有する配列は、同じファミリーの配列のメンバーである可能性が高い、すなわち、ファミリーの共通の特徴を共有する可能性が高いようである。本明細書では、STAR活性を含む非常に多くの配列（いわゆるSTAR配列）が、STAR活性を含む配列に特徴

的である配列識別子（パターン）を得るために使用された。これらのパターンは、テスト配列がSTAR活性を含んでいる可能性があるかどうかを決定するために使用されうる。従って、1つの観点から、本発明は、約50～5000塩基対の核酸配列内のSTAR配列の存在を検出するための方法であって、前記配列における少なくとも1つの配列パターンの出現の頻度を決定し、そして出現の頻度がSTAR配列を含む少なくとも1つの配列における前記少なくとも1つの配列パターンの出現の頻度を表わすことを決定することを含む、前記方法を提供する。原則的に、配列パターンがSTAR配列を表わすかどうかを決定するための任意の方法が適している。多くの種々の方法が、当業者にとって利用可能である。本発明の好ましい実施態様では、出現の頻度が、STAR配列を含む少なくとも1つの配列における前記少なくとも1つの配列パターンの出現の頻度を表わすことを決定する前記ステップが、前記少なくとも1つの配列パターンの出現の頻度が、前記少なくとも1つのSTAR配列及び少なくとも1つの対照配列の間で有意に異なることを決定することを含む。原則的に、任意の有意な相違が、STAR配列の存在のために識別力がある。しかしながら、特に好ましい実施態様では、前記少なくとも1つの配列パターンの出現の頻度が、前記少なくとも1つの対照配列と比較して、STAR配列を含む前記少なくとも1つの配列において有意に高い。STAR配列を含む相当多くの配列が、本発明で識別されている。パターンが対照配列とSTAR配列を含む配列とを区別する際にいかに有効であるかをテストするためにこれらの配列を使用することが可能である。いわゆる判別式分析（discriminant analysis）を使用して、一つの種（species）におけるSTAR配列の任意のセットに基づき、最適な判別配列（discriminative sequence）パターン又はそれらの組み合わせを決定することが可能である。従って、好ましくは前記パターンの少なくとも1つが、STAR配列を含む前記少なくとも1つの配列と対照配列との間の所望の及び好ましくは最適の判別に基づいて選択される。所望の判別は、バイオインフォマティックスを通じて前記パターンに関連づけられた或る重みファクターであり得る。

【0039】

好ましい実施態様では、テスト核酸における配列パターンの出現の頻度は、STAR配列を含むことが知られている配列における出現の頻度と比較される。この場合において出現の頻度が同様である場合、パターンはSTAR配列を含む配列を表わすと考えられる。好ましい実施態様では、他の基準が使用される。STR配列を含む配列におけるパターンの出現の頻度が、対照配列における前記パターンの出現の頻度と比較される。該2つの頻度を比較することによって、このように分析された各パターンについて、STAR配列を含む配列における前記頻度が対照配列における頻度と有意に異なるかどうかを決定できる。この実施態様では、もしSTAR配列を含む少なくとも1つの配列におけるパターンの出現の頻度が対照配列における同じパターンの出現の頻度と有意に異なる場合、配列パターンは、STAR配列を含む配列を表わすと考えられる。STAR配列を含む配列の多くの数を使用することによって、統計的な相違が確立されうるところのパターンの数が増加し、その結果出現の頻度がSTAR配列を含む配列を表わすところのパターンの数を増大する。好ましくは、出現の頻度はSTAR配列を含む少なくとも2配列における、より好ましくはSTAR配列を含む少なくとも5配列における前記少なくとも1つの配列パターンの出現の頻度を表わす。より好ましくは、STAR配列を含む少なくとも10配列におけるものである。より好ましくは、出現の頻度は、STAR配列を含む少なくとも20配列における少なくとも1つの配列の出現の頻度を表わす。特に好ましい実施態様では、出現の頻度はSTAR配列を含む少なくとも50配列における前記少なくとも1つの配列パターンの出現の頻度を表わす。STAR配列を含む配列を表示するところのパターンは使用される対照核酸の型にまた依存する。使用される対照配列の型は好ましくは、STAR配列の存在が検出されるべきであるところの配列に基づき選択される。好ましい実施態様では、前記対照配列が、STAR配列を含む前記少なくとも1つの配列と類似のAT/CG含量を有するランダム配列を含む。他の好ましい実施態様では、対照配列は、前記STAR配列を含む前記配列と同じ種から得られる。例えば、テスト配列が、植物細胞において活性であるSTAR配列の存在を検査される場合、対照配列は好ましくは植物細胞から取得される。同様に、ヒト細胞におけるSTAR活性をテストするために、対照核酸が好ましく

は、ヒトゲノムからまた取得される。好ましい実施態様では、対照配列は、STAR配列を含む前記少なくとも1つの配列の塩基の50%~150%を含む。特に好ましい実施態様では、前記対照配列は、STAR配列を含む前記少なくとも1つの配列の塩基の90%~110%を含む。より好ましくは、95%~105%である。

【0040】

パターンは、2よりも多い任意の数の塩基を含むことができる。好ましくは、少なくとも1つの配列パターンは少なくとも5、より好ましくは少なくとも6の塩基を含む。他の実施態様では、少なくとも1つの配列パターンは少なくとも8塩基を含む。好ましい実施態様では、前記少なくとも1つの配列パターンは、表9及び/又は表10にリストされたパターンを含む。パターンは、塩基の連続したリストから成ってもよい。しかしながら、該パターンは、判別できない又は部分的にのみ判別できる多くの塩基によって1つ又はそれ以上の回数で中断される塩基をもまた含んでもよい。部分的に判別的な塩基は例えば、プリン(purine)として示される。

【0041】

好ましくは、STAR活性の存在は機能的アッセイを使用して検証される。幾つかの方法は配列がSTAR活性を含むかどうかを決定するために明細書中に示される。STAR活性は、前記配列が次の機能の少なくとも1つを実行することができる場合に確認される：(i)本発明の遺伝子転写を抑制するエレメントを含む配列の効果を少なくとも部分的に阻害する、(ii)クロマチンに関連づけられた抑制を少なくとも部分的にロックする、(iii)エンハンサーの活性を少なくとも部分的にロックする、(iv)同一の核酸のみと比較して、転写ユニットをコードする動作可能にリンクされた核酸に、(iv-a)転写のより高い予測可能性、(iv-b)より高い転写、及び/又は(iv-c)長期にわたる転写のより高い安定性を与える。

【0042】

本発明において識別されるSTAR活性を含む非常に多くの配列は、種類において同じであり量において必ずしも同じでない活性を含む配列を生成し且つ識別する広い種々の可能性を開く。例えば、本発明において識別された配列を変更し、且つ変更された配列のSTAR活性を試験することは当業者の到着できる範囲内に十分にある。それ故に、そのような変更された配列はまた本発明の一部である。変更は、配列における1つ又はそれ以上の塩基の削除、挿入及び変異を含みうる。

【0043】

STAR活性を含む配列は、400塩基の広がりで識別された。しかしながら、それら400塩基の全てがSTAR活性を保存するために必要とされるのではないことが予測される。400~500塩基のフラグメントに或る特性を与える配列の境界を定める方法は、よく知られている。STAR活性を含むフラグメントの最小配列長は、約50塩基であると見積もられる。

【0044】

表9及び表10は、STAR活性を含む核酸分子において過剰に表わされていると判った6塩基のパターンをリストする。この過剰の表現は、STAR配列の特徴であると考えられる。該表は、65個のSTAR配列のファミリーのために生成された。同様の表が異なるセットのSTAR配列から又はより小さい若しくはより大きいセットのSTAR配列から開始して生成される。パターンがSTARエレメントを含まない配列と比較して前記STAR配列において過剰に表わされているならば、パターンはSTAR配列を表わす。これは、ランダム配列でありうる。しかしながら、無関係なバイアスを排除するために、STAR配列を含む配列は好ましくは、ゲノム又はその有意な部分と比較される。好ましくは脊椎動物又は植物のゲノム、より好ましくはヒトゲノムである。ゲノムの有意な部分は例えば、一つの染色体である。好ましくはSTAR配列及び前記対照配列を含む配列は、同一の種の核酸から取得される。

【0045】

より多くのSTAR配列が、配列パターンの出現の頻度の決定のために使用されると、過剰又は低度に表わされているパターンがSTARをよりよく代表する。核酸によって発現されうる多くの機能性特性がそれに結合するプロテイン性分子によって達成されることを考慮す

ると、代表パターンはSTAR配列において過剰に表わされることが好ましい。そのように過剰に表わされるパターンは、そのようなタンパク質性分子のための結合部位の一部分であり得る。好ましくは出現の頻度は、STAR配列を含む少なくとも2配列における、より好ましくはSTAR配列を含む少なくとも5配列における前記少なくとも1つの配列パターンの出現の頻度を表わす。より好ましくはSTAR配列を含む少なくとも10配列である。より好ましくは出現の頻度は、STAR配列を含む少なくとも20配列における前記少なくとも1つの配列パターンの出現の頻度を表わす。特に好ましい実施態様では、出現の頻度はSTAR配列を含む少なくとも50配列における前記少なくとも1つの配列パターンの出現の頻度を表わす。好ましくは、STAR配列を含む前記配列は、図26に記載されている配列の少なくとも1つを含む。

【0046】

STAR活性は、図26にリストされた配列によって共有される特性である。しかしながら、これは、それらが同一の識別子配列を全て共有しなければならないことを意味しない。異なる識別子が存在することは極めて十分に可能である。識別子は、それを含むフラグメントに共通の特性を与えうるが、しかしながらこれはそうである必要はない。

【0047】

1又は複数の配列パターンの出現の頻度を決定するためにSTAR活性を含むより多くの配列を使用することによって、そのようなSTAR配列において他よりもよりしばしば存在する又は存在しないパターンを選択することが可能である。このようにして、STAR配列において非常にしばしば過剰に表わされている又は低度に表わされているパターンを見つけることが可能である。しばしば過剰に又は低度に表わされているパターンは、テストセットにおける候補STAR配列を識別する蓋然性がより高い。過剰に又は低度に表わされているパターンのセットを用いる他の方法は、どのパターン又はパターンの組み合わせが配列におけるSTARを識別するために最も適されているかを決定することである。いわゆる判別統計を使用して、我々はSTARエレメントを含む配列を識別する際に最も良く実行するパターンのセットを識別している。好ましい実施態様では、STAR配列を検出するための前記配列パターンの少なくとも1つが、配列パターンGGACCC、CCCTGC、AAGCCC、CCCCCA及び/又はAGCACCを含む。他の実施態様では、STAR配列を決定するための前記配列パターンの少なくとも1つが、配列パターンCCCN{16}AGC、GGCN{9}GAC、CACN{13}AGG及び/又はCTGN{4}GCCを含む。

【0048】

STAR配列のリストは、それらの中の1つ又はそれ以上のコンセンサス配列を検出するためにまた使用されうる。それ故に、本発明はSTARエレメントのためのコンセンサス配列をまた提供する。このコンセンサス配列は勿論、テスト配列における候補STARエレメントを識別するために使用されうる。

【0049】

その上、一旦STARエレメントを含む配列が脊椎動物において識別されると、それは脊椎動物に属する他の種におけるSTARエレメントを含む配列を識別するために配列相同性によって使用されうる。好ましくは、哺乳動物のSTAR配列が、他の哺乳動物種におけるSTAR配列をスクリーンするために使用される。同様に一旦、STAR配列が植物種で識別されると、それは他の植物種における同様の機能を有する相同配列をスクリーンするために使用されうる。1つの観点における本発明は、本発明に従う方法によって取得可能なSTAR配列を提供する。さらにSTAR配列のコレクションが提供される。好ましくは前記STAR配列は、脊椎動物又は植物STAR配列である。より好ましくは、前記STAR配列は、哺乳動物STAR配列又は被子植物（单子葉植物例えは稻、又は双子葉植物例えはシロイヌナズナ（Arabidopsis））である。より好ましくは、前記STAR配列は、靈長類及び/又はヒトSTAR配列である。

【0050】

STAR活性を含む配列のリストは、テスト配列がSTARエレメントを含むかどうかを決定するために使用されうる。上記したように、この目的のためのそのようなリストを使用するための多くの種々の方法がある。好ましい実施態様では、本発明は約50~5000塩基対の核

酸配列がSTAR配列を含むかどうかを決定するための方法であって、本発明のSTAR配列のコレクションにおける配列パターンの出現の頻度を含む配列パターンの第1のテーブルを生成すること、少なくとも1つの参照(reference)配列における配列パターンの出現の頻度を含む前記パターンの第2のテーブルを生成すること、出現の頻度が前記2つのテーブル間で異なるところの少なくとも1つのパターンを選択すること、前記選択されたパターンの少なくとも1つの出現の頻度を約50～5000塩基対の核酸配列内において決定すること、及び前記テスト核酸における出現の頻度がSTAR配列の前記コレクションにおける前記選択されたパターンの出現を表わすかどうかを決定することを含む、前記方法を提供する。

【0051】

代替的に、前記決定することは、前記テスト核酸における出現の頻度がSTAR配列の前記コレクションにおける前記選択されたパターンの出現の頻度を表わすかどうかを決定することを含む。好ましくは、前記方法は、前記候補STARが遺伝子転写を調節する特性を有するかどうかを、本発明の方法を使用し決定することをさらに含む。好ましくは、STARの前記コレクションは、図26に記載されているような配列を含む。

【0052】

他の観点において、本発明は、本発明の方法によって取得されるSTAR配列を含む単離された及び／又は組換えられた核酸配列を提供する。

【0053】

上記したように、STAR配列は指向的な様式で、すなわち、それを含むフラグメントの1つの側に向って、他の側に向ってよりも多く、その活性を發揮することができる。STAR活性は、STARエレメントの数を複製することによって多量に増幅されうる。後者は、STARエレメントがSTAR活性を含む1つ又はそれ以上のエレメントを含んでもよいことを示唆する。それを含むフラグメント上にSTAR活性を与えることが可能な配列を識別する他の方法は、脊椎動物又は植物配列、STAR活性を含む配列からの選択を含み、そして選択された配列に隣接する配列が他の種において保存されるかどうかを識別することを含む。そのような保存された隣接配列は、機能的配列(functional sequences)である可能性が高い。1つの観点から、それ故に本発明は、STARエレメントを含む脊椎動物又は植物種からの約50～5000塩基対の配列を選択すること、及び前記種において前記選択された配列に隣接する配列が少なくとの1つの他の種において保存されるかどうかを識別することを含む、STAR配列を含む配列を識別する方法を提供する。それ故に本発明は、約50～5000塩基対の核酸配列内のSTAR配列の存在を検出する方法であって、一つの種(species)の細胞の染色体の一部分におけるSTAR配列を含む配列を識別すること、そして前記配列及び異なる種の染色体の配列との間の有意な相同意性を検出することを含む方法を提供する。好ましくは、前記種は、植物又は脊椎動物種を含み、好ましくは哺乳動物種である。本発明は脊椎動物種又は植物種の約50～5000塩基対の核酸配列内のSTARエレメントの存在を検出するための方法であって、前記核酸配列の隣接配列が少なくとも1つの他の種に保存されるかどうかを識別することを含む、前記方法をまた提供する。

【0054】

バイオインフォマティクス的な情報を使用し、STAR配列を含む配列の存在を検出するための本発明の方法はその性質として反復可能であることに注目することは重要である。STAR配列を含むより多くの配列が、本発明の方法で識別されると、より多くのパターンが、STAR配列を含む配列と対照配列との間で判別可能であることが見い出される。これらの新規に見つかった判別可能なパターンを使用して、STAR配列を含む多くの配列が識別され、その結果それは判別可能であるパターンのセットを増加し、等々である。この反復性の観点は、本発明において提供される方法の重要な観点である。

【0055】

配列に関連する用語「特性」は、前記配列の活性を云う。明細書中で使用される用語「STAR」、「STAR配列」又は「STARエレメント」は、上記した遺伝子転写を調節する特性の1つ以上を有するDNA配列を云う。用語「SINC」又は「SINCエレメント」(下記にリストされている)は、上記した遺伝子転写を抑制する特性の1つ又はそれ以上を含むDNA配

列を云う。本明細書中で使用される用語「DNA配列」は、別記しない限りは、塩基の特定の順序のリストを云うのではなく、DNAの物理的な片を云う。DNA配列に関連して転写特性は、前記DNA配列が目的の遺伝子の転写について有する効果を云う。本明細書中で使用される「特性」(quality)は、転写システムにおける核酸又はタンパク質の検出可能な特性又は属性を云う。

(実施例)

【実施例1】

【0056】

STARエレメント及びSINCエレメントを単離する方法

材料及び方法：

プラスミド及び菌株

STARエレメントのための選択ベクター、pSelect-SV40-zeo(「pSelect」、図1)は、次のようにして構築される：pREP4ベクター(Invitrogen V004-50)が、プラスミドバックボーンとして使用される。それは、靈長類の細胞株における高いコピーのエピソーマル複製のために、複製のイプシュタイン(Epstein)バール(Barr)oriPオリジン及びEBNA-1核性抗原を提供し、；哺乳動物細胞における選択のために、チミジンキナーゼプロモーター及びポリアデニレーション部位を有するハイグロマイシン(hygromycin)耐性遺伝子を提供し；及び大腸菌における維持のために、アンピシリン耐性遺伝子及び複製のcolE1オリジンを提供する。ベクターは、XbaI及びNheI制限部位の間に4つの連続したLexAオペレーター部位を含む(ブンケル及びキングストン、1994年)。LexAオペレーターとNheI部位の間に、次の制限部位から成るポリリンカーが埋め込まれる；HindIII-Ascl-BamHI-AscI-HindIII。NheI部位とSalI部位の間に、SV40プロモーター及びポリアデニル化部位を有するゼオシン(zeocin)耐性遺伝子があり、それはpSV40/Zeo(Invitrogen V502-20)から導出される；これは、STARスクリーンのための選択可能なマーカーである。

【0057】

pSDHベクター(図2)は次のようにして構築される。pGL3-コントロール由来のルシフェラーゼレポーター遺伝子(Promega E1741)は、PCRによって増幅され、そしてSacII/BamHI-消化されたpUHD10-3(ゴッセン及びビュヤード、1992年)内に挿入される。これはTet-Offプロモーターの制御下に及びSV40ポリアデニル化シグナルの上流にルシフェラーゼを配置する。マルチクローニング部位はPCRによって、Tet-Offプロモーター(MCS1,Xhol-NotI-EcoRI-SalI)の上流かつポリアデニル化シグナル(MSCII,NheI-BgIII-EcoRV-HindIII)の下流に導入される。遺伝子ライプラリーは、プラセンタ(placenta)(Clontech 6550-1)から精製された又はバクテリア/P1(BAC/PAC)人工染色体内で運ばれたヒトゲノムDNAのSau3AI消化によって構築される。BAC/PACクローンは、1q12細胞発生の領域(クローンRP1154H19及びRP3328E19)からの又は異種再生の遺伝子のHOXクラスター(クローンRP1167F23、RP1170019及びRP11387A1)からのゲノムDNAを含む。該DNAはサイズで分画され、及び0.5~2 kbサイズフラクションは、標準の技術(サムブルックら、1989年)によって、BamHI-消化されたpSelectベクター内にライゲーションされる。

【0058】

宿主株の構築は、記載されている(ファン デル フラッグら、2000年)。手短に言えば、それらはU-2 OSヒト骨肉腫細胞株(American Type Culture Collection HTB-96)に基づく。U-2 OSは、Tet-リプレッサーDNA結合性ドメイン及びVP16転写促進ドメインから成るタンパク質キメラをコードするpTet-Offプラスミド(Clontech K1620-A)で安定にトランスフェクションされる。該細胞株は次に、LexA DNA結合性ドメイン、及びHP1又はHPC2(DNAにつながれた場合に遺伝子発現を抑制する2つのショウジョウバエポリコーム-グループタンパク質)のいずれかのコード領域を含む融合タンパク質遺伝子で引き続き安定的にトランスフェクションされる。LexA-リプレッサー遺伝子は、Tet-Off転写的調整システムの制御下にある(ゴッセン及びビュヤード、1992年)。

【0059】

ライプラリースクリーニング及びSTARエレメント特性

pSelectにおける遺伝子ライブラリーは、トランスフェクション試薬の提供者（Life Technologies）によって推奨されているように、リン酸カルシウム沈殿（グラハム及びファン デル エブ、1973年；ウィグラーら、1978年）によってU-2 OS/Tet-Off/LexA-リプレッサー細胞株内にトランスフェクションされる。トランスフェクションされた細胞は、ハイグロマイシン選択（25マイクロg/ml）及びテトラサイクリン抑制（ドキシサイクリン（doxycycline）、10ナノg/ml）下で、1週間（50% コンフルエンス（confluence））培養される。その後ドキシサイクリン濃度は、LexA-リプレッサー遺伝子を誘発するために0.1ナノg/mlにまで減少され、そして2日後にゼオシンが250マイクロg/mlまで添加された。細胞はさらに、（エンプティpSelectでトランスフェクションされた）コントロール培養がゼオシンによって殺されるまで、更に4～5週間培養される。

【0060】

ライブラリートransfekションからのゼオシン耐性コロニーは増殖され、且つプラスミドDNAは、単離され且つ標準技術（サムブルックら、1989年）によって大腸菌内にレスキューされる。レスキューされたDNA内における候補STARエレメントは、U-2 OS/Tet-Off/LexA-リプレッサーへの再トランスフェクション及びドキシサイクリン濃度の低下後に、制限エンドヌクレアーゼマッピング（サムブルックら、1989年）、DNA配列分析（サンガーら、1977年）によって且つSTAR活性（ゼオシン耐性）について解析される。

【0061】

ヒトゲノムにおける公知の配列に対応するDNA配列を有する候補STARエレメントは、ヒトゲノムデータベース（<http://www.ncbi.nlm.nih.gov/genome/seq/HsBlast.html>、2001年6月20日）のBLAST調査（アルツシュルら、1990年）によって識別される。エレメントの染色体の位置が、反復DNAの比率及び近傍の遺伝子の識別とともに記録される。再トランスフェクション後に、STAR活性を示すそれらの候補は、pSDHプラスミド内にSTARフラグメントをサブクローニングすること及びU-2 OS染色体のDNA内への安定な組み込みによって更に特徴付けられる。pSDHプラスミドは、U-2 OS細胞内にpBABE-puro（モルゲンスターん及びランド、1990年）でコトランスフェクションされ、且つピューロマイシン耐性について選択される。STARエレメント毎に、およそ30個の個々のクローンの集団が分離され且つ培養される。クローンは、製造者の指示書（ロッシュ 1669893）に従いルシフェラーゼ活性を周期的にアッセイされる。

【0062】

結果：

STARエレメント機能的特徴付け。ヒトゲノムDNA並びにHOX及び1q12座のスクリーンは、17個の正真正銘のSTARエレメントを生じた。基準は、（1）エレメントが、pSelectに基づくクローンを宿主U-2 OSヒト骨肉腫細胞株内に再トランスフェクションした後にSTAR活性を示した（初期スクリーンで発現される抗リプレッサー活性はプラスミド特異的であり、且つ宿主細胞内における人工的な変化によるものではないことを示す）；（2）エレメントが、ヒトゲノム配列データベースにおける配列と一致するDNA配列を含む（クローンは例えばバクテリア起源又はベクター起源からの汚染するDNA配列を含まないことを示す）。

【0063】

STARエレメントは、pSDHプラスミド内にサブクローニングされ、且つ宿主細胞ゲノム内に組み込まれる。レポーター遺伝子の発現は、ゲノム内にランダムに組み入れられた後にレポーター遺伝子をサイレンシングから防ぐためのSTARエレメントの能力を示すために、安定なトランスフェクションの集団においてアッセイされる。これは、（1）高発現を示すクローンの割合及び（2）STARエレメントによって導出された過剰発現の程度に対する情報を提供する。

【0064】

クローンによるルシフェラーゼレポーター遺伝子の発現は、それがSTARエレメントを含まないプラスミドの平均レベル（参照レベル）よりも2倍上である場合、有意であると考えられる。全てのプラスミドにとって、発現レベルにおける分散が、クローンの間で観

察される：非発現から参照レベルより有意に上にまで発現する及び殆ど過剰発現しないから多く過剰発現するまで。優れたSTAR活性は、幾つかの高く過剰発現するクローンを含む多くの過剰発現するクローンをもたらすプラスミドによって明示される。

【0065】

代表的な実験からの結果は、表1及び図3～5に示される。

【0066】

結果は、テストされるヒトSTARエレメントが、保護されていないレポーター遺伝子又はショウジョウバエSCSエレメント（ケルム及びシェドル、1992年）によって保護されたレポーター遺伝子よりもより高い割合の過剰発現クローンを生じることを示す。その上、これらのプラスミドによる過剰発現の程度は、保護されていない又はSCS保護されたレポーターよりもSTAR保護されたレポーター遺伝子からのものが非常に大きい。

【0067】

STARエレメント配列及びゲノム配置データ

表2は、17個のSTARエレメントの夫々の染色体位置、ならびにすぐ隣の遺伝子の識別及びエレメントの反復DNA含量をリストする。STARエレメントは多くの染色体中に分布される。それらは、それらの真のDNA配列及び反復DNA含量において多様であり、且つ近隣の遺伝子との様々な程度の関連を示す。

【0068】

SINCエレメントスクリーン

材料及び方法：

SINCスクリーンのためのプラスミド、pSINC-Select（「pSS」、図6）は、次のようにして構築される：pREP4ベクター（Invitrogen V004-50）が、プラスミドバックボーンとして使用される。それは、複製のイプシュタイン（Epstein）バール（Barr）oriPオリジン及び靈長類細胞株における高いコピーのエピソーマル複製のためのEBNA-1核性抗原を提供し；哺乳動物細胞における選択のために、チミジンキナーゼプロモーター及びポリアデニレーション部位を有するハイグロマイシン（hygromycin）耐性遺伝子を提供し；及びアンピシリン耐性遺伝子及び大腸菌における維持のための複製のcolE1オリジンを提供する。ベクターはTet-Offプロモーターを含み、それはTet-Off転写的調節システムによる調節のために、プラスミドpUDH10-3（ゴッセン及びビュヤード、1992年）からのタンデムテット（Tet）応答エレメント（TRE）から成る。TREは、融合タンパク質（シトシンデアミナーゼ／ウラシルフォスホリボシルトランスフェラーゼ；Invivogen porf-codaupp）をコードするcodA::upp遺伝子の発現を調節する。これは、いわゆる「自殺遺伝子」である；codA::upp酵素の活性は、プロドラッグ 5-フルオロシトシン（5-FC）を毒性薬物 5-フルオロウラシル（5-FU）に転化し、それによってアポトーシス及び細胞死をもたらす（ミューレンら、1992年；ティラビーら、1998年；ウェイ及びフーバー、1996年）。Tet-Offプロモーターの上流に、スクリーニングのためのSau3AI-消化されたゲノムDNAをクローニングするためのBgIII制限部位がある。pREP4 DNAは、クローニングされたSINCエレメントによってpREP4コンポーネントにおける必須のプラスミドエレメントのサイレンシングを防止するためにSTAR配列によってゲノムDNA及び自殺遺伝子から分離される。

【0069】

ヒト22番染色体（Invitrogen/Research Genetics 96010-22）を含むBACクローンのライブライリーからのゲノムDNAは、Sau3AIで部分的に消化され、そしてBgIII-消化されたpSS内にライゲーションされる（サムブルックら、1989年）。組み換えプラスミドのライブライリーは、トランスフェクション試薬の提供者（Life Technologies）によって推奨されているように、リン酸カルシウム沈殿（グラハム及びファン デル エブ、1973年；ウィグラーら、1978年）によってU-2 OS/Tet-Off細胞株内にトランスフェクションされる。トランスフェクションされた細胞は、ハイグロマイシン選択（25マイクロg/ml）及びテトラサイクリン抑制（ドキシサイクリン（doxycycline）、10ナノg/ml）下で、3週間培養される。その後5-FCが1マイクロg/mlの濃度まで添加され、そして細胞はSINCエレメントを選択するために3週間の間更に培養される。

【 0 0 7 0 】

候補SINCを含むコロニーは収穫され、そしてプライマーPCR1及びPCR2(図6)とともにポリメラーゼ連鎖反応において使用される；PCR生成物は、慣用技術(サムブルックら、1989年)によって、HindIII及びXbaI制限エンドヌクレアーゼで消化され且つpBluescript II SK(+) (Stratagene 212207)内にクローニングされる。候補SINCエレメントのDNA配列が決定され(サンガーら、1977年)及びヒトゲノムにおける対応する配列は、ヒトゲノムデータベース(<http://www.ncbi.nlm.nih.gov/genome/seq/HsBlast.html>、2001年6月20日)のBLAST調査(アルツシュルら、1990年)によって識別される。エレメントの染色体の位置が、反復DNAの比率及び近傍の遺伝子の識別とともに記録される。

【 0 0 7 1 】

結果：

選択期間の終わりでは、コロニーはコントロール培養(エンプティpSS)において明らかでなく、且つ多くのコロニーがゲノムDNAを有するpSSを含む培養内で明らかである。これらの生き残ったクローニーは、候補SINCエレメントを含む。これらの生き残ったクローニーは、候補SINCエレメントを含む。エレメントはPCRによって回収され且つ標準クローニングベクター、pBluescript内にサブクローニングされる。エレメントのDNA配列が決定され、そしてヒトゲノム配列と比較される(表3)。全ての場合において、シーケンスされたエレメントは、予想されたように、22番染色体上に見つけられる。

【 実施例 2 】**【 0 0 7 2 】**

STAR、SINC又は組み合わされたSTAR/SINCによる導入遺伝子の発現特性

背景：

部位特異的組み換えは、異種のDNAをそれらの染色体の位置から正確に除去するために使用される。これは、2つのシステムの1つによって日常的に実行されている：バクテリオファージP1のcreリコンビナーゼ及びloxPターゲット(フェングル、1999年)、又は酵母のFLPリコンビナーゼ及びFRT(FLPリコンビナーゼターゲット)(ウィグレイら、1994年)。これらのシステムでは、(レポーター遺伝子及び/又は選択可能なマーカーを通常含む)DNA領域は、loxP又はFRTターゲットによって染色体中で隣接される。その後、リコンビナーゼの活性は、染色体から該DNA領域の正確な切除を触媒する。リコンビナーゼは、その2つの認識配列を単一部位に変え(resolve)、それらの間の配列を削除する。従って、DNAの全長は、リコンビナーゼの導入又は活性化が起きると、インビボ(invivo)で引き続き削除されるべきターゲット部位により隣接されなければならない(シュヴェンクら、1995年；ディメッキ、1996年)。Cre及びFIPリコンビナーゼは、最小の6個(loxP)又は8個(FRT)の塩基対をもつスペーサーによって分離された、2個の13塩基対の反転された反復の間での組み替えを触媒する(セネコフら、1985年)。loxP配列はATAACTTCGTATAであり、且つFRT配列はGAAGTTCCCTATACである。

【 0 0 7 3 】

プロトコル：

慣用的なDNAクローニング(サムブルックら、1989年)を使用し、(レポータータンパク質例えば緑色蛍光タンパク質(GFP)をコードする)レポーター遺伝子(ビリリューゼンら、1997年)又はルシフェラーゼ(ヒメス及びシャノン、2000年)が構築され、それはSTARエレメントの対によって、SINCエレメントの対によって又はSTAR/SINC組み合わせエレメントの対によってプラスミドにおいて隣接される。各場合において、エレメントはリコンビナーゼターゲット部位によってそれら自身隣接される。1つのエレメントはloxP部位の対によって隣接され、及び他はFRT部位の対によって隣接される(図1)。転写すると、プラスミドは細胞の小さいパーセントで宿主染色体内に組み込まれ、そして組み込まれたものは、抗生物質耐性によって選択される。同様の構築物は3つのテストエレメント(STAR、SINC、STAR/SINC)の夫々のために作成される。

【 0 0 7 4 】

慣用の技術(「SuperFect Transfection Reagent Handbook」、Qiagen、1997年11月)

、を使用し、これらのプラスミドは、U-2 OSヒト骨肉腫細胞株内にトランスフェクトされ、且つハイグロマイシン耐性について選択される。ハイグロマイシン耐性単離体は、細胞株のゲノム内に安定に組み込まれたプラスミドを有する。夫々の単離体は、細胞培養培地内で増殖され、且つ組み換えレポーター遺伝子の発現は例えばフローサイトメトリー（ステュルラ、2000年）によってアッセイされる。

【0075】

その後慣用技術（トランスフェクション、又はホルモン刺激）を使用して、前記からの安定な単離体は、リコンビナーゼ活性を導入し又は活性化するように処理される。これは、例えばcreリコンビナーゼ活性はSTAR1の切除を触媒し、そして引き続きFLPリコンビナーゼ活性はSTAR2の切除を触媒するように、順次行われる。これらの細胞におけるレポーター遺伝子の発現のレベルがアッセイされ、値が、親の、STARを含む単離物の値を基準として比較される。

【実施例3】

【0076】

STARの配列分析；エレメント機能のための最小必須配列の決定；エレメント間の配列の保存；及びタンデム及び複数のエレメントの特性

背景：STAR又はSINCエレメントを含むDNAフラグメントが、夫々pSelect（図1）又はpSS（図6）プラスミドを使用して遺伝子の選択によって単離される。このセクションはSTAR又はSINC活性を有するそれらのフラグメント内のDNA配列を特徴付けるためのアプローチを記載する。

【0077】

プロトコル：

DNA配列：オリゴヌクレオチドが、DNAフラグメントをシーケンスするためにpSelect及びpSS選択プラスミドの配列に基づき設計される。フラグメントはジデオキシ連鎖終止技術（サンガーら、1977年）を使用してシーケンスされる。その後、DNA配列は、公のヒトゲノム配列データベース（http://www.ncbi.nlm.nih.gov:80/cgi-bin/Entrez/hum_srch?chr=hum_chr.inf&query）を使用して染色体位置に配置される。フラグメント配列の近接における遺伝子及び遺伝子密度は、ゲノム配列アノテーションから記録される。それらの遺伝子の転写的活性は、公のDNAマイクロアレイのデータベース（<http://arrays.rockefeller.edu/xenopus/links.html>）及びSAGE（Serial Analysis of Gene Expression;<http://bioinfo.amc.uva.nl/HTM-bin/index.cgi>）データから決定される。いったんSTAR及びSINC配列に対する位置の情報が編集されると、データは基礎となるコンセンサス配列の観点から分析される。コンセンサス配列又はトレンド（これは、特定のヌクレオチド組み合わせに豊む例えはC及びG塩基に豊むローカルエリアであると理解される）が、類似性探索アルゴリズム例えはclustalw（ヒギンズら、1996年）及びプロサム類似性スコアリング（blobum similarity scoring）（アルツシュル及びギッシュ、1996年）を使用し検出される。その後、任意の基礎となるコンセンサス又はトレンドは、BLAST調査（アルツシュルら、1990年）を実行することによってゲノムスケール上の他の潜在的なSTARを識別するために使用される。従来の研究は、既知のインシユレーター及び境界エレメントに結合する転写調節タンパク質を識別した（ガスツナーら、1999年；ゲラシモヴァ及びコルセス、1998年）。記載されている実施例では、タンパク質結合部位は、インシユレーター又は境界機能に必須であるところのDNase I過敏部位と一致する。STARエレメントが既知の調節タンパク質によって結合されるという仮定は、STARエレメントにおいて生じる配列モチーフのための転写ファクターのTRANSFACデータベース（<http://transfac.gdf.de/TRANSFAC/>）を検索することによって試験される。STAR又はSINCコレクションのメンバー間で共通である配列モチーフは、対応する転写ファクターがそのエレメントに結合するところの指標である。

【0078】

最小の必須配列：この配列知識を使用して、STAR（又はSINC）エレメントは、機能性の切り詰められ、そして機能について試験される。これは、STAR-又はSINC-含有フラグメン

トのサブフラグメントを標準技術（サムブルックら、1989年）によってpSelect又はpSS内にクローニングするために、ポリメラーゼ連鎖反応（PCR）を使用して行われる。サブフラグメントを含むプラスミドは、U-2 OS細胞内にトランスフェクションされ、且つ抗生物質耐性（STARエレメント）又はプロドラッグ耐性（SINCエレメント）をアッセイすることによって機能について試験される。

【0079】

方向性：STAR及びSINCエレメントは、夫々pSelect及びpSSプラスミドを使用してそれらの方向性をテストされる。例えば、pSelectスクリーンによって単離されたSTARエレメントの方向性は、5' 3' 配位と言われる。エレメントの配向は、慣用の組み換えDNA技術（サムブルックら、1989年）によって、反転される。生じたプラスミドは、U-2 OS細胞株内にトランスフェクションされ、且つレポーター遺伝子の発現がアッセイされる（ビリューゼンら、1997年、ヒメス及びシャノン、2000年）。反転配位エレメントを有するプラスミドからの発現のレベルは、5' 3' 配位のそれと比較される。反転配位プラスミドが類似の発現レベルを有する場合、STARエレメントは方向性を示さない。

【0080】

エレメントの組み合わせ及び倍数：STARエレメントが混合された塩基対において機能することができるかどうかを決定するために、種々のエレメントが組み合わされ且つ試験される。分析は、組み換えDNA技術（サムブルックら、1989年）によって、MCSIに1つのSTARエレメントを及びMCSIIに異なるSTARを挿入することによってpSDHプラスミド内で実行される。生じたプラスミドはトランスフェクションされ、及びレポーター遺伝子の発現がアッセイされる（ビリューゼンら、1997年；ヒメス及びシャノン、2000年）；結果は、MCSI及びMCSIIで同一のエレメントを含むプラスミドからの発現と比較される；もし該発現が、プラスミドの2つの型と同様であれば、種々のSTARエレメントはお互いに邪魔をしないということが結論づけられる。

【0081】

単一のSTAR又は複数のSINCエレメントの強度は、エレメントのタンデム反復と比較される。これは、DNAリガーゼによる目的のSTAR又はSINCエレメントのコンカタメリゼーション（concatamerization）及び組み換えDNA技術（サムブルックら、1989年）によるpSDH又はpSSプラスミド内へのライゲーション生成物の挿入によって行われる。生じたプラスミドは、U-2OS細胞内にトランスフェクションされ、且つレポーター遺伝子の発現がアッセイされる（ビリューゼンら、1997年、ヒメス及びシャノン、2000年）；結果は、単一のSTAR又は複数のSINCエレメントを含むプラスミドからの発現と比較される。

【実施例4】

【0082】

STAR、SINC又はそれらの組み合わせが機能するところの距離の決定

背景：

STARエレメントは、単一及び複数の導入遺伝子の発現を最適化するために使用される。STARエレメントの単一対が大きい又は複数の導入遺伝子をサイレンシングから防ぐことができるかを決定するために、STARエレメントが活動するところの範囲を決定することが必要である。同様の情報が、SINCエレメント及びSTAR/SINC組み合わせのために決定される。

【0083】

プロトコル：

STAR及びSINCエレメントは、次のようにpSelect又はpSS夫々に基づく誘導プラスミドを使用して、距離を超えるそれらの機能を試験される。500bpから10kbのランダムDNAフラグメントのライブラリーは、標準DNAクローニング技術（サムブルックら、1989年）によって組み立てられる。STAR又はSINC活性を有しないフラグメントは、上記したようにpSelect及びpSSで試験することによってこのライブラリーから選択される。STARエレメント及びSTAR/SINC組み合わせのために、これらのフラグメントは適切なpSelectプラスミド（図1）におけるクローニング部位とレポーター遺伝子のプロモーターとの間に挿入される。ブ

ラスミドのこのセットは、U-2 OS細胞株内にトランスフェクションされ、そして発現が上記したように測定される。レポーター遺伝子発現の強度は、プロモーターからのSTARエレメントを分離するランダムフラグメントの長さに関連させられる。SINCエレメントは、類似の様式で評価される：ランダムDNAフラグメントは適切なpSSプラスミドのSINCエレメントとプロモーターとの間に挿入され、そしてレポーター遺伝子の抑制の程度は、ランダムDNAフラグメントの長さに関連させられる。

【実施例5】

【0084】

実施例5（a）

STARエレメントのための遺伝的選択における自然に生じるSINCエレメントの使用

背景：

STARエレメントのための現在のスクリーンは、選択プラスミドにおいて選択可能なマークターの抑制を提供するためにキメラlexA-PcGタンパク質を使用する。自然に生じるSINCエレメントを使用し選択を繰り返すことによって、これらの自然に生じるSINCエレメントによる抑制活性に対して特異的であるSTARエレメントが識別される。

【0085】

SINCエレメントスクリーンは、「Tet-Off」プロモーターをサイレンシングし且つcodA:upp自殺遺伝子の抑制をロックすることができるゲノムDNAのランダムに生成されたフラグメントを識別するための遺伝子選択の能力に基づく。この選択から回収したSINCエレメントは、ゲノムのサイレンシングするエレメントのランダムサンプリングを表し、且つ種々のクラスのエレメントが回収される。このプロトコルのために、これらの多様なSINCエレメントは、前記したlexA-PcGを基にした選択において回収されたそれらと異なるクラスのSTARエレメントを回収するために使用される。

【0086】

プロトコル：

現在の選択からのSINCエレメントは特徴付けられ、機能的特性及びDNA配列特性に基づきクラスに分類される（機能的特性は抑制の強度を含む；配列特性は識別可能な保存されたモチーフを含む；実施例3参照）。各クラスからの代表的なエレメントは、標準DNAクローニング技術（サムブルックら、1989年）を介してpSelectプラスミドにおけるlexA結合部位を置換するために使用される。遺伝子バンクは、これらの新しいプラスミドの各自で作成され、そして記載されたように（ファン デル フラッグら、2000年）新しい、SINC特異的なSTARエレメントを識別するために使用される。これは、全体のゲノムDNAを用い且つ使用されるSINCエレメントをまた含むBACクローンからのDNAを用いて行われる。

【0087】

実施例5（b）

STARエレメント及びSINCエレメントの最大長の決定

背景：

STARエレメントは、pSelectプラスミドを使用して回収されたDNAのフラグメントとしてクローニングされ、それは2kbよりも少ないゲノムDNAフラグメントでおこなわれる。しかしながら、これらは、より拡張されたSTARエレメントの部分であってもよい。拡張されたSTAR活性は、次の実験によって試験される。

【0088】

プロトコル：

pSelectにおいてクローニングされたSTARエレメントは、ヒトゲノム配列にマッピングされる。それらがより拡張されたSTARエレメントの部分であるかを決定するために、クローンを包含する4kbの領域が、PCRによって増幅され且つ標準の組み換えDNA技術（サムブルックら、1989年）によってpSelect及び/又はpSDHプラスミド内にクローニングされる。生じたプラスミドは、U-2 OS細胞内にトランスフェクションされ且つ上記したようにレポーター遺伝子発現についてアッセイされる；オリジナルの2kb STARエレメントを含むプラスミドは、対照として含まれる。3つの可能な結果が予想されうる：（1）、対照及び

拡張されたSTAR単離物による同様の発現、STARエレメントがオリジナルの2kbフラグメントに限られることを示す；(2)拡張されたSTAR単離物による低い発現、STARエレメントが2kbフラグメント内に含まれ且つ長距離で効果的に活動しないこと又は拡張されたフラグメントがSINCエレメントを含むことを示唆する；(3)拡張されたSTAR単離物によるより高い発現、拡張された領域がより完全なSTARエレメントを含むことを示唆する。結果(3)の場合、実行は6kbのより大きいPCRフラグメントで繰り返される。

【0089】

STARエレメントは、様々なタンパク質が結合しているところの部位の複合物(composite)であってもまたよい。それ故に、STAR活性を有する大きなDNAフラグメントは、STAR活性を有するより小さいフラグメントに分割しうる(実施例3を参照)。2kbよりも大きいエレメントは、それらが2kb(内部的欠失を含む)未満へのトランケーション(truncation)後にSTAR活性をまだ示す場合、STARエレメントとして認識される。

【実施例6】

【0090】

STARエレメント、SINCエレメント、又はそれらの組み合わせ及び近傍の導入遺伝子のメチル化及びヒストンアセチル化状態

背景：

STARエレメント及びSINCエレメントの調節特性は、ローカルなクロマチン構造に関連付けられ、それはDNAそれ自身によって及びDNAに関連付けられたタンパク質によって決定される。遺伝子発現における変化に関連付けられるクロマチン構造の変化は、巨大分子の2次的修飾、特にDNAのメチル化又はヒストンタンパク質のアセチル化によってしばしば生成される。STARエレメント及びSINCエレメントで且つ近傍の導入遺伝子で生じる2次的変性を識別することは、これらのエレメントの特質(hallmarks)を提供する。

【0091】

プロトコル：

DNAメチル化：STARエレメント又はSINCエレメント又はそれらの組み合わせは、標準技術(サムブルックら、1989年)によってpSelectプラスミド内にクローニングされる。U-2 OS細胞は、レポーター遺伝子での基本的DNAメチル化を決定するために、これらのプラスミドで且つ対照としてSTARエレメント又はSINCエレメントを欠くpSelectで安定にトランスフェクションされる。細胞は収穫され且つクロマチンは標準手法(トーマス、1998年)によって精製される。該DNAは、別の反応(サムブルックら、1989年)においてHpaII及びMspI制限エンドヌクレアーゼで消化される。これらの制限酵素の両方は、非メチル化配列CCGGを切断することができる。外側のCがメチル化される場合、MspI及びHpaIIの両方が開裂できない。しかしながら、内部のCがメチル化される場合、HpaIIと違って、MspIは配列を開裂できる。該DNAがサンプロット法に付され、そして該プロットが間接的なエンドラベリング(パジン及びカドナガ、1998年)によって解析される。対照として、裸の対応するpSelectプラスミド、メチル化されていないDNAは記載された酵素でまた切断され、そしてサンプロット法に付される。

【0092】

種々のサイズのDNAフラグメントの比較は、DNAがインビボ(in vivo)でメチル化されるかどうかを明らかにする。

【0093】

ヒストンアセチル化：DNAメチル化分析のために使用される同一のトランスフェクションされた細胞株が、これらの実験のために使用される。下記に記載された方法は、STARエレメント及びSINCエレメント及びレポーター遺伝子上のヒストンアセチル化パターンの高い解決マップを生成する(リットラ、2001年)。核のミクロコッカル(micrococcal)ヌクレアーゼ消化は、ショ糖グラジエントで画分化され、且つ精製されたヌクレオソームモノマー及びダイマーは、抗アセチルヒストン抗体での免疫沈降によって、アセチル化されたヒストンに豊むようにされる。ヌクレオソーム画分及び免疫沈降物は分析に付され、例えばプライマー及びタックマン(Taqman)プローブ(それはレポーター遺伝子若しくはSTAR

又はSINCエレメントにアニールして、0.1kbのムービングウィンドウを有する0.2kb生成物を与える)を使用し、リアルタイムPCR(ジュングラ、2000年)によって分析される。次に、PCRの間のタックマンプロープ蛍光シグナルの増加の割合(それはサンプルにおけるテンプレートDNAの数に比例している)が測定される。又クレオソーム画分及び免疫沈降物におけるテンプレートDNAの数の割合は、レポーター遺伝子及びSTARエレメント又はSINCエレメント上の(又はエレメントの不存在におけるレポーター遺伝子上の)各0.1kbのためのヒストンアセチル化のパターンの詳細マップを提供する。

【実施例7】

【0094】

インビボ(*in vivo*)又クレオソーム配置及びDNase I過敏性部位

背景:

クロマチンは、DNA、ヒストン及び非ヒストンタンパク質から成る。該ヒストンは、又クレオソームを作成するためにDNAの~150bpによってラップされているコア粒子を形成する。又クレオソームは、リンクーDNAの50~75bpによって分離される。染色体のDNA上に安定に配置された又クレオソームは遺伝子発現を抑制し、又クレオソームを排除する又はさもなくばクロマチンを改造するファクターは、この抑制を克服することができる。染色体の領域における又クレオソームの配置は、ミクロコッカル(micrococcal)又クレアーゼ(MNase)アッセイによって分析される; MNaseは、リンクーDNAにおいて優先的にクロマチンを切断する。同様に、DNAのある領域は、構成的に非ヒストンタンパク質にさらされ、且つこれらはしばしば調節領域すなわちシス・行動(acting)調節ファクターが結合するところの部位である。実験的に、これらの部位は酵素DNaseIによる消化に対して過敏である。

【0095】

プロトコル:

レポーター遺伝子上の及びSTARエレメント又はSINCエレメントのいずれか上の又クレオソームの位置を決定するために、MNaseが使用される(サルツ及びジョスト、1993年)。核は、培養されたU-2 OS細胞から精製され且つ上記したように(ヒストンアセチル化)MNaseで消化される。STARエレメント及びSINCエレメント又はレポーター遺伝子におけるDNase I過敏部位を探すために、精製された核が上記したように(ウォールラスら、1998年)適切な濃度(例えば、100マイクロg/mlゲノムDNA及び20~100 U/ml DNase I)でDNaseIで処理される。裸の(naked)DNAは、対照としてDNase Iで消化される。両方の技術にとって、レポーター遺伝子及びSTARエレメント又はSINCエレメントが、(タナカら、1996年; ファンデルフラッギーら、2000年)に、記載されているように、プライマーの拡張又は間接的なエンドラベリング及びサザンプロット法を使用し詳細マップされる。MNaseアッセイは、STARエレメント若しくはSINCエレメント又はレポーター遺伝子上の又クレオソームの位置に対応するオートラジオグラム上の離れたバンドの位置(ladder)を明らかにする。DNase I過敏部位は、裸のDNA対照において存在しない又は比較的顕著でないところの生じたオートラジオグラムにおける離れたバンドとして現れる。

【実施例8】

【0096】

STARエレメント及びSINCエレメントの細胞型、組織依存性及びプロモーター依存性

背景:

あるインシュレーター又は境界エレメントが組織特異性を表示しうることが報告されている(タカダら、2000年)。STARエレメントは、インシュレーター及び境界エレメントと共通の多くの特性を有する。無差別の及び組織特異的STARエレメント及びSINCエレメントは導入遺伝子的用途においてバイオテクノロジー的価値を有する。下記に記載されたアッセイは、細胞型依存性を評価するために実行される。エレメントの細胞及び組織特異性は、DNAマイクロアレイの公知のデータベース(<http://arrays.rochester.edu/xenopus/links.html>)及びSAGE(Serial Analysis of Gene Expression;<http://bioinfo.amc.uva.nl/HTM-bin/index.cgi>)データを使用しヒトゲノムにおけるエレメントの近傍の遺伝子の

発現を検査することによってさらに検査される。

【0097】

プロトコル：

STARエレメントはpSDHプラスミド内で試験され、且つSINCエレメントはpSSプラスミド内で試験される。3つの細胞株が、標準のプロトコルを使用しトランスフェクションされる：ヒト U-2 OS骨肉腫細胞株（ヘルディンら、1986年）、アフリカサバンナモンキー（African green monkey）の腎臓由来のヴェロ（Vero）細胞株（シミズら、1967年）、及びチャイニーズハムスター卵巣由来のCHO細胞株（カオ及びブック、1968年）。すべての3つの細胞型において機能することができるエレメントは、無差別として分類される。該細胞株の1つ又は2つにおいてのみ活性を示すものは、それらの細胞型機能性において制限されていると分類される。

【0098】

プロモーター特異性：

STARエレメント及びSINCエレメントは、2つのプロモーター、完全なサイトメガロウィルス（CMV）プロモーター又はテトラサイクリン応答エレメント（Tetracycline Response Element）及び最小のCMVプロモーター（tTA転写的活性剤との組み合わせで）を用いて機能との関連で、現在選択され且つ試験される。プロモーター特異性を評価するために、STAR及びSINC機能が、他の一般に使用されるウイルスプロモーターすなわち猿のウイルス型40（SV40）初期及び後期プロモーター、アデノウイルスE1A及び主な後期プロモーター、並びにラウス肉腫ウイルス（Rous sarcoma virus; RSV）長末端反復（ドールら、1996年；スミスら、2000年；ウイバー及びカダン、2000年；クスウラ、1995年）で試験される。これらのプロモーターの夫々は、STARエレメント又はSINCエレメントとともに標準の技術（サムブルックら、1989年）によってpSelect及びpSSプラスミド内に夫々別個にクローニングされる。生じたプラスミドは、上記したように、U-2 OS細胞株内にトランスフェクションされ、且つレポーター遺伝子発現についてアッセイされる。これらのプロモーターをサイレンシングさせるSINCエレメントの能力又はサイレンシングに対して保護するためのSTARエレメントの能力は、STARエレメント又はSINCエレメントを欠くプラスミドとの比較によって決定される。

【実施例9】

【0099】

STARエレメント及びSINCエレメントの改善の方法

背景：

改善されたSTARエレメント及びSINCエレメントが開発される。改善は抗抑制又は抑制活性の増加した強度並びに誘発可能な及び組織特異的な活性を有するエレメントをもたらすこれらの改善は組み合わせの技術によって作成される。

【0100】

プロトコル

強制された進化：エラープローンPCR（Error Prone PCR）（チェリーら、1999年；ヘンケ及びボルンショイヤー、1999年）は、エレメント当たり1つ又は2つのポイント変異の平均を導出するために使用される。突然変異したエレメントは、例えば蛍光活性化されたセルソーティング及び抗生物質耐性（ベネットら、1998年）によってレポーター選択可能なマーカー融合タンパク質を含むpSelect（又はpSS）プラスミドを使用しスクリーンされる。エラープローンPCR及び選択の次の段階は、活性において更なる改善を有するエレメントを導出するために実行される。

【0101】

タンデム及び異種の組み合わせ：上記したように、エレメントのタンデム及び異種の組み合わせは、シングルエレメント（実施例3）と比較して活性について試験される。

【0102】

STARエレメント及びSINCエレメントの相対的な優位が、ケースバイケースに基づいて試験される。それは、エレメントの強度を試験するために使用される；例えば新しいSTAR工

レメントが、既知の強いSINCエレメントエレメントに対して優位である場合、該STARは非常に強いとして分類される。STAR及びSINCの間の優位関係は細胞型 - 、組織 - 又はプロモーター - 特異的であるところの可能性が、また考えられる（実施例8）。優位性試験は、標準組み換えDNA技術（サムブルックら、1989年）によって個々のSTARエレメントの上流に配置された個々のSINCエレメントを有するpSelectプラスミドを利用する。該プラスミドは、U-2 OS細胞にトランスフェクションされ、且つレポーター遺伝子発現がアッセイされる。SINC優位性は、STARエレメントのみを有するプラスミドよりもより低い発現によって表わされ、一方STAR優位性は、SINCエレメントのみを有するプラスミドよりもより高い発現によって表わされる。

【0103】

新規な特徴（例えば、誘発性、組織特異性）を付加するためにSTARエレメント及びSINCエレメントへの他のDNA結合性タンパク質のための結合部位の導入

【0104】

背景：

調節可能なSTARエレメント及びSINCエレメントは、それらをシングル依存性DNA結合性タンパク質のための結合部位と結合することによって生成される。1つの例では、これはSTAR若しくはSINC又はSTAR/SINC組み合わせ及びグルココルチコイド応答エレメント（GRE：glucocorticoid response element）の並置を含む。グルココルチコイド刺激の不存在下、STARエレメント又はSINCエレメントは、記載されているように機能する。刺激に応答して、自然に生じるグルココルチコイドレセプターはGREに結合し且つSTAR又はSINC機能を妨害する。

【0105】

プロトコル：

慣用のDNAクローニング（サムブルックら、1989年）を使用し、GREは、STARエレメント又はSINCエレメントに隣接したpSelect又はpSSベクター内に夫々導入される。該プラスミドは、上記したようにU-2 OS細胞内にトランスフェクションされる。細胞は、2つの培地に分割される；1つは、グルココルチコイド（10マイクロM）で処理される。レポーター遺伝子の発現は測定され且つ2つの培地間で比較される。発現における相違は、シングル依存性DNA結合性タンパク質の活性によってSTAR及びSINC機能を調節する能力を示す。

【0106】

無差別のSTAR及びSINCエレメント：

これらの特性を試験し又は助長することは、種々の細胞株における培養、抗生物質的選択なしの長期間培養を含む（実施例8及び10）。

【実施例10】

【0107】

STARエレメント及びSINCエレメントは導入遺伝子の維持のための連続的な選択の必要性を防ぐ。

背景：

導入遺伝子学において、選択マーカーに頼ることは、2つの欠点がある：選択剤は通常高価であり且つ細胞に代謝コスト（metabolic cost）を与え、且つ特に導入遺伝子それ自身が製品（例えば作物、遺伝子治療ベクター）内にある場合、選択可能マーカーを導入遺伝子用途に含めることに対する法律的且つ倫理的な異議がある。STARエレメント及びSINCエレメントは、遺伝子組み換え単離物を確立した後に選択を維持することの必要性を減少し又は除去する。従って、耐性遺伝子は、導入遺伝子発現の減少した損失を有する部位特異的組み替えによって形質転換されたゲノムから除去されうる。

【0108】

プロトコル：

レポーター遺伝子に隣接する染色体的に統合されたSTARエレメントを含む安定的にトランスフェクションされたU-2 OS細胞株は、上記したようにトランス - 行動（trans-acting）抗生物質耐性プラスミドとpSDHプラスミドの同時トランスフェクションによって生成さ

れる。実験は、選択の不存在下に長期に渡る（3～6ヶ月）培養の間にこれらの細胞株におけるレポーター遺伝子発現レベルの安定性を試験することを含む。これは、pSDHプラスミドにおけるルシフェラーゼ又はGFPレポーター遺伝子に隣接するSTARエレメントで試験される。抗生物質耐性遺伝子は、抗生物質選択マーカーがリコンビナーゼターゲット部位によって隣接されるところの（pSDHに基づく）発現プラスミドを構築することによって除去される。選択可能なマーカーは次に、上記したように（実施例2）、リコンビナーゼ活性によって切り取られる。

【実施例11】

【0109】

予測可能性及び収率が発現システムにおけるSTARエレメントの適用によって改善される。

STARエレメントは、導入遺伝子発現ユニットに対する転写抑制機能の効果をブロックするために機能する。これらの抑制影響は、ヘテロクロマチン（「位置影響」、（ボイヴィン及びデュラ、1998年））又は導入遺伝子の近傍のコピー（「反復導入された遺伝子サイレンシング（repeat-induced gene silencing）」、（ガーリックら、1998年））によりうる。異種タンパク質生成のためのSTARエレメントの2つの利点は、高発現性プライマリー組み換え宿主細胞を見い出すことの増加した予測可能性、及び生成サイクルの間の増加した収率である。これら利益は本実施例において示される。

【0110】

材料及び方法：

pSDHベクター及びSTAR含有誘導体の構築：pSDH-Tetベクターは、プライマーC67及びC68を使用しプラスミドpREP4-HSF-Luc（ファン・デル・フラッグら、2000年）からのルシフェラーゼオープンリーディングフレームのポリメラーゼ連鎖反応増幅（PCR）（PCRプライマー及び突然変異オリゴヌクレオチドのすべてが表5にリストされている）、及びSacII/BamHIフラグメントをSacII/BamHI-消化されたpUHD10-3へ挿入すること（ゴッセン及びビュヤード、1992年）によって構築された。ルシフェラーゼ発現ユニットはプライマーC65及びC66で再増幅され且つそれを2つのマルチクローニング部位（MCSI及びMSCII）に隣接するためにpUHD10-3内に再挿入された。次に、Ascl部位は、EcoRIでの消化及び（アニールされたオリゴヌクレオチドD93及びD94を含む）リンカーの挿入によってMCSI内に導入された。CMVプロモーターは、ベクター pSDH-CMVを生成するために、プライマーD90及びD91を用いてプラスミドpCMV-Bsd（Invitrogen K510-01）から増幅され及びSalI/SacII消化且つライゲーションによりpSDH-Tet内のTet-Offプロモーターを置換するために使用された。このベクターにおけるルシフェラーゼオープンリーディングフレームは、次のようにしてSEAP（Secreted Alkaline Phosphatase）によって置換された：ベクター pSDH-CMVは、SacII及びBamHIで消化されそして平滑（blunt）にされた：SEAPオープンリーディングフレームは、ベクター pSDH-CSを生成するために、EcoRI/SAI1消化によってpSEAP-ベーシック（Clontech 6037-1）から分離され、平滑にされ且つpSDH-CMV内にライゲーションされた。SV40プロモーターの制御下のピューロマイシン耐性遺伝子は、プライマーC81及びC82を使用し、PCRによってプラスミドpBabe-Puro（モルゲンスター及びランド、1990年）から単離された。これは、pGL3-puroを生成するために、Ncol/XbaIで消化されたベクターpGL3-コントロール（BamHI部位が除去された）（Promega E1741）内にライゲーションされた。pGL3-puroは、SV-40-ピューロ耐性遺伝子を分離するためにBgIII/SalIで消化されて、それは平滑にされ且つ、NheI消化された、平滑末端（blunt-ended）pSDH-CS内にライゲーションされた。生じたベクター pSDH-CSPは、図7に示される。全てのクローニングステップは、従来技術で知られている方法（サムブルックら、1989年）に従い、試薬の製造者によって提供される指示書に従い実行された。

【0111】

STARエレメントは、適切な制限酵素でSTARエレメントおよびpSDH-CSPベクターを消化し、引き続きライゲーションすることによる2段階で、MCSI及びMSCII内に挿入された。組み換えられたpSDHベクターにおけるSTARエレメントの配位は、制限マッピングによって決定された。挿入物の識別及び配位は、DNA配列解析によって検証された。配列決定は、

製造者の指示書に従って、ベックマン (Beckman) CEQ2000自動化DNAシーケンサーを使用しジデオキシ方法 (サンガーラ、1977年) によって実行された。簡単にいようと、DNAは、QIAprep Spin Miniprep及びプラスミド ミディ (Midi) キット (夫々QIAGEN 27106及び12145) を使用し大腸菌から精製された。サイクル配列決定が、染料ターミネーター (CEQ Dye Terminator Cycle Sequencing Kit、Beckman 608000) の存在下でカスタムオリゴヌクレオチドC85、E25及びE42 (表5) を使用して、実行された。

【0112】

pSDHプラスミドを有するCHO細胞のトランスフェクション及び培養：

チャイニーズハムスター卵巣細胞株CHO-K1 (ATCCCL-61) が、HAMS-F12培地 + 2mMグルタミン、100 U/mlペニシリン及び100マイクログラム/mlストレプトマイシンを含む10%ウシ胎仔血清 (Fetal Calf Serum) 内で、37 /5%CO₂下で培養された。細胞は、製造者によって記載されたようにSuperFect (QIAGEN) を使用し、pSDH-CSPベクター及び、MCSI及びMSCIIにおいてSTAR6又はSTAR49を含むその誘導体でトランスフェクションされた。簡単に言えば、細胞は、培養容器に接種され且つ70~90%コンフルエンスにまで一晩成長された。SuperFect試薬は、6マイクロリットル/マイクログラムの割合 (例えば10cmペトリ皿に、20マイクログラムDNA及び120マイクロリットルSuperFect) で (PvuIでの消化によって本実施例において線形化された) プラスミドDNAと一緒にされ、且つ細胞に添加された。一晩のインキュベーション後、トランスフェクション混合物は、新鮮な培地で置き換えられ、及びトランスフェクションされた細胞はさらにインキュベートされた。一晩の培養後、5マイクログラム/ml ピューロマイシンが添加された。ピューロマイシン選択は2週間に完了し、その後個々のピューロマイシン耐性CHO/pSDH-CSPクローンがランダムに分離されそしてさらに培養された。

【0113】

SEAP (Secreted Alkaline Phosphatase) アッセイ：

CHO/pSDH-CSPクローンの培養培地におけるSEAP活性 (ベルゲルら、1988年；ヘンソルンら、1988年；カイン、1997年；ヤングら、1997年) は、製造者 (Clontech Great EscAPE kit #K2041) によって記載されるように決定された。簡単に言えば、培地の分取は、65度熱不活性化され、その後アッセイバッファー及びCSPD化学発光法基質と一緒にされ、そして室温で10分間インキュベートされた。その後、基質変換の速度は、照度計 (luminometer) (Turner 20/20TD) において決定された。細胞密度は、Coulter ACT10 細胞カウンターにおいてトリプシン化された細胞を算えることによって決定された。

【0114】

pSDHプラスミドを有するU-2 OS細胞のトランスフェクション及び培養：

ヒト骨肉腫U-2 OS細胞株 (ATCC #HTB-96) がダルベッコMEM培地 (Dulbecco's Modified Eagle Medium) +グルタミン、ペニシリン及びストレプトマイシン (上述) を含む10%ウシ胎仔血清内で、37 /5%CO₂下で培養された。細胞は、SuperFect (上述) を使用し、(プラスミドpBabe-Puroとともに) pSDH-CMVベクター及び、MCSI及びMSCIIにおいてSTAR6又はSTAR8を含むその誘導体で同時トランスフェクションされた。ピューロマイシン選択は2週間に完了し、その後個々のピューロマイシン耐性U-2OS/pSDH-CMVクローンがランダムに分離されそしてさらに培養された。

【0115】

ルシフェラーゼアッセイ：

ルシフェラーゼ活性 (ヒメス及びシャノン、2000年) が、照度計 (Turner 20/20TD) を使用しアッセイキット製造者の指示書 (Roche 1669893) に従い、再懸濁された細胞においてアッセイされた。合計細胞タンパク質濃度は、製造者の指示書 (Sigma B-9643) に従いビシンコニン酸 (bicinchoninic acid) 方法によって決定され、且フルシフェラーゼデータを正規化するために使用された。

【0116】

結果：

pSDH-CSPベクター又はSTAR6若しくはSTAR49を含むpSDH-CSPプラスミド (表6) を含む

組み換えCHO細胞クローンは、3週間培養された。その後、培養上清におけるSEAP活性が決定され、且つ細胞数に基づき表わされる(図8)。見られうるように、発現ユニットにおいてSTARエレメントを有するクローンは分離され、それはSTARエレメントを含まない発現ユニットのクローンよりも2~3倍高いSEAP活性を発現する。その上、STARのないクローンの最大の活性の又はそれより上のSEAP活性を発現するSTAR含有クローンの数は、かなり高い: STARクローン集団の25%~40%がpSDH-CSPクローンの最高SEAP発現を超える。

【0117】

pSDH-CMVベクター又は、STAR6若しくはSTAR8(表6)を含むpSDH-CMVプラスミドを含む組み換えU-2 OS細胞クローンは、3週間培養された。その後、宿主細胞内のルシフェラーゼ活性が決定され且つ総細胞タンパク質に正規化された相対ルシフェラーゼ単位(図9)として表現される。発現ユニットに隣接するSTARエレメントを有する組み換えU-2 OSクローンは、STARのないクローンよりもより高い収率を有した: STAR8クローンから観察された最高の発現は、STARのないクローンからの発現よりも2~3倍高かった。STAR6クローンは、STARのないクローンよりも5倍高い最大の発現レベルを有した。STARエレメントは同様により高い予測可能性を与えた: 両方のSTARエレメントについて、クローンの15~20%が、最高の発現レベルを有するSTARのないクローンと類似の又はそれよりも大きいレベルでルシフェラーゼ発現を示した。

【0118】

これらの結果は、強いCMVプロモーターとともに使用される場合、STARエレメントは異種のタンパク質(ルシフェラーゼ及びSEAP)の収率を増加することを示す。本実施例において導入された3つの全てのSTARエレメントは、高められた収率を提供する。STARエレメントによって与えられた増加した予測可能性は、STARなしのクローンによって示された最高の収率と等しいか又はそれよりも大きい収率を有するクローンの大きい割合によって明らかにされた。

【実施例12】

【0119】

STARエレメントは導入遺伝子発現の安定性を改善する

組み換えられた宿主細胞の培養の間、抗生物質の選択を維持することは通常のプラクティスである。これは、導入遺伝子の転写的サイレンシング又は組み換えのようなプロセスによってゲノムからの導入遺伝子の損失を防ぐことを目的とされる。しかしながら、それは異種のタンパク質の生産にとって多くの理由のために望ましいものではない。第1に、使用される抗生物質はとても高価であり、生産物の単位コストに重大な寄与をする。第2に、生物薬剤的な使用の点から、タンパク質は、明らかに純粋でなければならず、生成物に抗生物質の痕跡があつてはならない。異種タンパク質生産のためのSTARエレメントの一つの利点は、抗生物質選択の不存在下であつても、それらは、長期の培養の間、導入遺伝子に対して安定な発現を与えるということである: この特性は、本実施例で示される。

【0120】

材料及び方法:

U-2 OS細胞株は、プラスミドpSDH-Tet-STAR6でトランスフェクションされ、且つ実施例11に記載されているように培養された。個々のピューロマイシン耐性クローンが分離され、かつドキシサイクリンの存在なしにさらに培養された。週間隔で、1:20の希釈率で細胞は新しい培地容器に移された。ルシフェラーゼ活性は、実施例11に記載されているように定期的な間隔で測定された。15週間後、培養物は2つの複製に分けられた: 一つの複製はピューロマイシンを受け続け、一方他の複製は実験の残りにおいて抗生物質を受けなかった(25週間合計)。

【0121】

結果:

表7は、抗生物質あり又はなしで、長期にわたる成長の間、STAR6に隣接された発現ユニットによるルシフェラーゼ発現についてのデータを表わす。見られうるように、レポーター導入遺伝子、ルシフェラーゼの発現は、実験の持続の間、U-2 OS宿主細胞において安

定に保持される。該培養物が2つの処理（抗生物質をプラスと抗生物質なし）に分けられた後、ルシフェラーゼの発現は、抗生物質選択の不存在下で基本的に安定であった。これは、長期に渡る培養の間、サイレンシング又は損失から導入遺伝子を防ぐためのSTARエレメントの能力を示す。それは、この特性は抗生物質選択から独立であることをまた示す。それ故に、異種のタンパク質の生産は、抗生物質の又は難しい下流の処理の費用を生じることなしに可能である。

【実施例13】

【0122】

STARエレメントの最小必須配列

STARエレメントは実施例1に記載された遺伝子スクリーンから分離される。該スクリーンは、ほぼ0.5~2キロ塩基（上述）にサイズ分画されたヒトゲノムDNAで構築されたライブラリーを使用する。該STARエレメントは、500から2361塩基対の範囲である（表6）。分離されている多くのSTARエレメントにとって、STAR活性は当初分離されたクローンよりも小さいDNAフラグメントによって与えられるようである。STAR活性にとって必須であるこれらの最小のフラグメントサイズを決定することは、2つの理由で有用である。第1に、より小さい機能的STARエレメントは、コンパクトな発現ベクターの設計に有利であり、なぜならばより小さいベクターが、より高い効力で宿主細胞をトランスフェクションするからである。第2に、最小必須のSTAR配列を決定することは、助長された機能性のそれらの配列の変性を許容する。2つのSTAR配列は、それらの最小必須配列を決定するためにファインマッピングされる。

【0123】

材料及び方法：

STAR10（1167塩基対）及びSTAR27（1520塩基対）が、詳細マッピングされた。それらはPCRによって増幅され、ほぼ等しい長さ（図10説明）のサブフラグメントを与えた。初期のテストのために、これらは、BamHI部位でpSelectベクター内にクローニングされ、且つ実施例1に記載されているようにU-2 OS/Tet-Off/LexA-HP1細胞内にトランスフェクションされた。ハイグロマイシン耐性について選択の後、LexA-HP1がドキシサイクリン濃度を低くすることによって導出された。その後、トランスフェクションされた細胞は、LexA-HP1結合による抑制からSV40-Zeo発現ユニットを保護するためのSTARフラグメントの能力をテストするために、ゼオシンとともにインキュベーションされた。

【0124】

結果：

この実験では、STAR10及びSTAR27は、予想されたように遺伝子サイレンシングに対して良い保護を与える（図10）。これは、ゼオシンの存在下における強力な成長によって表わされる。

【0125】

3個のSTAR10サブフラグメントのうち、10A（～400塩基対）は、トランスフェクションされた細胞に対してゼオシンの存在下における活発な成長を与え、これは完全長STARエレメントのそれを超える。他の2個のサブフラグメントを含むpSelect構築物でトランスフェクションされた細胞は、ゼオシンの存在下で成長しない。これらの結果は、STAR10の抗リプレッサー活性に責任のあるDNA配列を包含するとして、～400塩基対10、Aフラグメントを識別する。

【0126】

STAR27は、本実験においてトランスフェクションされた細胞に対して、ゼオシン中での中庸の成長を与える。このSTARのサブフラグメントの一つ、27B（～500塩基対）は、ゼオシン含有培地において宿主細胞の弱い成長を許容する。これは、このSTARの抗リプレッサー活性がサブフラグメント27B上に部分的に局地化されるが、完全な活性は同様に27A及び/又は27C（夫々～500塩基対）からの配列を要求することを示す。

【実施例14】

【0127】

培養された哺乳動物細胞の多様な株におけるSTARエレメント機能

異種のタンパク質発現のための宿主細胞株の選択は、タンパク質の質、収率及び単位コストのために重要なパラメータである。考慮例えばポスト翻訳変性、分泌経路能力及び細胞株不死は、特定の生物薬剤的な生成システムのために適切な細胞株を指示する。この理由のために、収率、予測可能性及び安定性の観点においてSTARエレメントによって与えられる利点は、多様の細胞株において取得可能であるべきである。これは、それが元々クローニングされているところのヒトU-2 OS細胞株、及びバイオテクノロジーにおいて広く適用されているCHO細胞株におけるSTAR6の機能を比較することによって試験された。

【0128】

材料及び方法：

実施例11の実験が参照される。

【0129】

結果：

CHO細胞内のSEAPレポーター遺伝子の発現は、図8に表わされている；U-2 OS細胞におけるルシフェラーゼレポーター遺伝子の発現は、図9に示される。これらの2つの実験の結果の比較によって、STAR6エレメントは、両方の細胞株において機能することは明らかである：レポーター遺伝子発現はそれらの両方においてより予測可能であり、及び該レポーター遺伝子がSTAR6による位置効果から保護される場合、各細胞株のクローンはより高い収率を示した。これらの2つの細胞株は、異なる種（ヒト及びハムスター）並びに異なる組織型（骨及び卵巣）から得られ、このSTARエレメントが異種タンパク質発現を改善する際にその中で利用されうるところの宿主細胞の広い範囲を反映する。

【実施例15】

【0130】

種々の転写プロモーターとの関係におけるSTARエレメント機能

導入遺伝子転写は、導入遺伝子のオープンリーディングフレームを外来のプロモーターの制御下に置くことによって達成される。プロモーターの選択は、異種のタンパク質の性質及び生成システムによって影響される。ほとんどの場合、強い構造的プロモーター（*constitutive promoter*）は、それらが提供できる高い収率の故に好まれる。幾つかのウィルスプロモーターは、これらの特性を有する；CMV-IEG（cytomegalovirus immediate early gene）のプロモーター／エンハンサー（「CMVプロモーター」）は、通常の生物化学的な使用において最強のプロモーターであると一般にみなされている（ボシャルトら、1985年；ドールら、1996年；フェッキング及びホフステッテル、1986年）。シミアンウィルスSV40プロモーターもまた、中庸に強力であり（ボシャルトら、1985年；フェッキング及びホフステッテル、1986年）、且つ哺乳動物細胞ベクターにおける異所性の発現のためにしばしば使用される。Tet-Offプロモーターは導出可能である：該プロモーターは、tTAプラスミド（Clontech K1620-A）を発現する細胞株においてテトラサイクリン又は関連する抗生物質（ドキシサイクリンが通常使用される）の存在下で抑制され、抗生物質の除去は転写的誘発を生じる（ドイシュレら、1995年；ゴッセン及びビュヤード、1992年；イズミ及びギルバート、1999年；ウマナラ、1999年）。

【0131】

材料及び方法：

pSDH-Tet及びpSDH-CMVベクターの構築が、実施例11に記載されている。pSDH-SV40は、プラスミド pSelect-SV40-Zeo（実施例1）からのSV-40プロモーター（プライマー D41及びD42）のPCR増幅によって構築され、引き続きSaclI及びSalIでPCR生成物を消化した。pSDH-CMVベクターは、CMVプロモーターを除去するためにSaclI及びSalIで消化され、且つベクター及びSV40フラグメントはpSDH-SV40を生成するためにともにライゲーションされた。STAR6は実施例11で記載されたようにMCSI及びMCSII内にクローニングされた。プラスミド pSDH-Tet、pSDH-Tet-STAR6、pSDH-Tet-START7、pSDH-SV40及びpSDH-SV40-STAR6は、製造者によって記載されたようにSuperFectを使用し、U-2 OS内にpBabe-Puroで同時にトランスフェクションされた。細胞培養、ピューロマイシン選択及びルシフェラーゼアッ

セイは、実施例 1 1 に記載されているように実施された。

【 0 1 3 2 】

結果：

図 9、1 1 及び 1 2 は、下記の3つの異なるプロモーターからのルシフェラーゼレポーター遺伝子の発現を比較する：2つの強く且つ構造的なウィルスプロモーター（CMV及びSV40）、及び導出可能なTet-Offプロモーター。全ての3つのプロモーターは、U-2 OS 細胞においてSTAR6エレメントとの関連においてテストされた。結果は、全ての3つのプロモーターからの収率及び予測可能性は、STAR6によって増加されることを示す。実施例 1 1 及び 1 4 に記載されているように、STAR6はCMVプロモーターの関連において有益である（図 9）。同様の改善が、SV-40プロモーターとの関連において見られる（図 1 1）：最高発現するSTAR6クローンからの収率は、最良のpSDH-SV40クローンよりも2～3倍良く、且つ6個のSTARクローン（集団の20%）は最良のSTARのないクローンよりも高い収率を有する。導出する（低いドキシサイクリン）濃度下でTet-Offプロモーターの関連において、STAR6は導入遺伝子発現の収率及び予測可能性をまた改善する（図 1 2）：最高発現するSTAR6クローンは、最良のpSDH-Tetクローンよりも20倍高い収率を有し、及び9個のSTAR6クローン（集団の35%）は、最良のSTARのないクローンよりも高い収率を有する。このSTARエレメントはその導入遺伝子を保護する特性において多様的であり、なぜならばそれは、転写の様々なバイオテクノロジー的に有益なプロモーターとの関連において機能するからであると結論される。

【 実施例 1 6 】

【 0 1 3 3 】

STARエレメント機能は方向性でありうる

短い核酸配列が対称（例えば回文）でありうるのに対して、より長い自然に生じる配列は典型的には非対称である。結果として、核酸配列の情報内容は方向性であり、且つ配列自体はそれらの5' 及び3' 末端に関して記載されうる。核酸配列情報の方向性は、組み換えDNA分子が当業者に知られている標準クローニング技術（サムブルックら、1989年）を使用して組み立てられるところの準備に影響を与える。STARエレメントは、長く、非対称的なDNA配列であり、且つそれらがpSelectベクターにおいて元々クローニングされた配向に基づく方向性を有する。上記された実施例では、pSDHベクターにおける2つのSTARエレメントを使用し、この方向的に保持された。この配向は、ゼオシン耐性遺伝子に相対的に、自然の又は5' -3' 配向として記載される（図 1 3 参照）。この実施例では、STAR機能のための方向性の重要性が、pSDH-Tetベクターにおいてテストされる。pSDHベクターにおけるレポーター遺伝子は目的のSTARエレメントのコピーによって両方の側で隣接される故に、各STARコピーの配向が考慮されなければならない。この実施例は、自然の配向を逆の配向と比較する（図 1 3）。

【 0 1 3 4 】

材料及び方法：

STAR66エレメントは、実施例 1 1 に記載されているようにpSDH-Tet内にクローニングされた。U-2 OS細胞は、プラスミド pSDH-Tet-STAR66-native及びpSDH-Tet-STAR66-oppositeで同時トランسفエクションされた。夫々のクローンは分離され、そして培養された；ルシフェラーゼ発現のレベルは、記載されたように（上述）決定された。

【 0 1 3 5 】

結果：

自然の配向及び逆の配向におけるSTAR66活性の比較の結果が、図 1 4 に示される。STAR66が逆の配向にある場合、一つのクローンのみの収率がかなり高い（60ルシフェラーゼ単位）。その一方、STAR66が自然の配向にある場合の最高発現クローンの収率はかなり高く（100ルシフェラーゼ単位）、且つ予測可能性は同様にはるかに高い：自然配向の集団の7個（30%）のクローンが、逆配向の集団からの最高発現クローンのレベルより上のルシフェラーゼを発現し、且つ自然配向の集団の15個（60%）のクローンが、10相対的ルシフェラーゼ単位よりも上のルシフェラーゼを発現する。それ故に、STAR66機能は方向性である

ことが示される。

【実施例 17】

【0136】

STARエレメントとの関連における導入遺伝子の発現はコピー数依存性である

異質タンパク質発現のための導入遺伝子発現ユニットは、細胞分割の間の安定な保存を保証するために宿主細胞のゲノム内に一般的に組み込まれる。組み込みは、ゲノム内に挿入された発現ユニットの一つ又は複数のコピーを生じうる：複数のコピーはタンデムアレイとして存在するか又は存在しない。STARエレメントによって保護された導入遺伝子のために示された増加した収率（上述）は、STARエレメントは導入遺伝子発現ユニットがゲノム内の組み込みの部位に関連付けられた転写への影響に独立的に機能することを許容する（位置効果からの独立性（ボイヴィン及びデュラ、1998年））ことを示す。それは、STARエレメントは各発現ユニットが、それらがタンデムアレイとして組み込まれた場合に、発現ユニットの近傍するコピーに独立的に機能することを許容することをさらに示唆する（反復導出された遺伝子サイレンシングからの独立性（*independence from repeat-induced gene silencing*）（ガーリックら、1998年））。コピー数依存性は、下記の実施例で記載されるように導入遺伝子発現レベル及びコピー数の間の関係から決定される。

【0137】

材料及び方法：

U-2 OS細胞は、pSDH-Tet-STAR10で同時トランスフェクションされ、そして記載されたように（上述）ピューロマイシン選択下で培養された。8個の個々のクローンが分離され、そしてさらに培養された。その後、細胞は収穫され、そして一つの部分は、記載したように（上述）ルシフェラーゼ活性を試験された。残った細胞は溶解され、そしてゲノムDNAはDNeasy Tissue Kit (QIAGEN 69504) を使用し製造者によって記載されたように精製された。DNAサンプルは、UV分光測定法によって数値化された。3マイクログラムの各ゲノムDNAサンプルが、製造者によって記載されたように（New England Bioabs）一晩PvuII及びXbaIで消化され、そしてアガロースゲル電気泳動法によって分離された。DNAフラグメントは、記載されたように（サムブルックら、1989年）ナイロン膜に移され、そして（BamH I/SacII-消化されたpSDH-Tetから分離された）ルシフェラーゼ遺伝子に対する放射活性的にラベル化されたプローブとハイブリダイズされた。プロット（blot）は、記載されたように（サムブルックら、1989年）洗われ、そしてフォスファーイメージャスクリーン（Personal F/X, BioRad）にさらされた。生じたオートラジオグラム（図15）はルシフェラーゼDNAバンドの相対強度を決定するために、デンシトメトリーによって解析され、それは導入遺伝子コピー数を表わす。

【0138】

結果：

pSDH-Tet-STAR10クローン集団からのコロニーにおけるルシフェラーゼの酵素活性及びコピー数（DNAバンド強度）が、図16に示される。導入遺伝子コピー数は、これらのpSDH-Tet-STAR10クローンにおけるルシフェラーゼ発現のレベルに高く相関している（ $r=0.86$ ）。このことは、STAR10が導入遺伝子発現ユニットに対してコピー数增加依存性を与え、導入遺伝子発現をタンデムアレイにおける他の導入遺伝子コピーに独立に及び組み込みの部位における遺伝子サイレンシング影響に独立にする。

【実施例 18】

【0139】

STARエレメントはエンハンサープロッカーとして機能し、しかしエンハンサーとして機能しない

遺伝子プロモーターは、転写を開始するためのそれらの能力に対してポジティブな及びネガティブな影響の両方に付される。ポジティブな影響を働くエレメントの重要な分類は、エンハンサーである。エンハンサーは、それらが遠くに配置された場合（多くのキロ塩基対）にでさえ、プロモーターに影響を特徴的に与えることができる。ヘテロクロマチン形成（例えばポリコーム グループタンパク質）によって作用するネガティブな影響

は上記に記載されており、且つそれらはSTAR活性のターゲットである。エンハンサー機能のための及びヘテロクロマチン形成のための生化学的な基礎は、基本的に類似しており、なぜならばそれらはいずれも、DNAに対するタンパク質の結合に関係するからである。それ故に、STARエレメントがネガティブ影響ならびにポジティブ影響をブロックすることができるか、換言すれば組み込みの部位の近傍におけるゲノム的エンハンサーからの導入遺伝子を保護することができるかどうかを決定することが重要である。エンハンサー活性からの導入遺伝子を保護する能力は、バイオテクノロジーの用途において導入遺伝子の安定な且つ予測可能な能力を保証する。この実施例は、エンハンサーブロッキングアッセイにおけるSTARエレメントの能力を検査する。

【0140】

これらの機能に対して重要なところのSTAR活性の他の特徴は、それらが導入遺伝子に対して与える増加した収率である（実施例11）。ヘテロクロマチン形成タンパク質が候補STARエレメントの近傍に結合される場合、複数のSTARがゼオシン発現の高いレベルを維持するそれらの能力に基づき分離される。複数のSTARはゼオシン発現ユニット内へのヘテロクロマチンの拡散をブロックすることが予測される故に、高い発現が生じると予測される。しかしながら、第2のシナリオは、ゼオシン耐性クローニングにおけるDNAフラグメントがエンハンサーを含むことである。エンハンサーは、ポリコーム・グループタンパク質例えばSTARスクリーンの方法において使用されるそれらの抑制効果を克服する能力を有することが示されている（チングル、1995年）。この現象によって分離されたエンハンサーは、偽のポジティブとみなされ、なぜならばエンハンサーはSTARのために本明細書中に請求された特性を有しないからである。STARエレメントがエンハンサーでないことを示すために、それらはエンハンサーブロッキングアッセイにおいて試験された。

【0141】

エンハンサーブロッキングアッセイ及びエンハンサーアッセイは、方法論的に及び概念的に類似である。これらアッセイは、図17において図式的に示される。エンハンサーをブロックするSTARエレメントの能力は、E47/E-boxエンハンサーシステムを使用し実行される。E47タンパク質は、それがプロモーターの近傍に配置されたE-box DNA配列に結合されている場合、該プロモーターによる転写を活性化することができる（クオングラ、2002年）。E47は、B及びTリンパ球分化の制御に通常関係する（クオングラ、2002年）が、しかしそれは異所的に発現した場合、多様な細胞型において機能することが可能である（ピーターソンら、2002年）。E-boxは回文的DNA配列、CANNTGである（クノフラーら、2002年）。エンハンサーブロッキングアッセイにおいて、E-boxは、発現ベクターにおいて（最小プロモーターを含む）ルシフェラーゼレポーター遺伝子の上流に配置される。STARエレメントのためのクローニング部位は、E-boxと該プロモーターとの間に配置される。E47タンパク質は、第2のプラスミド上にコードされる。アッセイは、E47プラスミド及びルシフェラーゼ発現ベクターの両方を細胞内にトランスフェクションすることによって実行される：E47タンパク質は、発現され且つE-boxに結合され、及びE47/E-box複合体はエンハンサーとして作用することができる。ルシフェラーゼ発現ベクターがSTARエレメントを含まない場合、E47/E-box複合体はルシフェラーゼ発現を助長する（図17A、状況1）。STARエレメントがE-boxとプロモーターとの間に挿入された場合、エンハンサーをブロックするそれらの能力が、ルシフェラーゼ活性の低減した発現によって示される（図17A、状況2）；STARがエンハンサーをブロックできない場合、ルシフェラーゼ発現は活性化される（図17A、状況3）。

【0142】

エンハンサーとして作用するSTARエレメントの能力は、同一のルシフェラーゼ発現ベクターを利用する。E47の不存在において、E-boxそれ自身は転写に影響を与えない。代わりに、STARエレメントによるエンハンサー行動は、ルシフェラーゼ転写の活性化を結果する。アッセイは、E47プラスミドなしにルシフェラーゼ発現ベクターをトランスフェクションすることによって実行される。発現ベクターがSTARエレメントを含まない場合、ルシフェラーゼ発現は低い（図17B、状況1）。STARエレメントがエンハンサー特性を有しな

い場合において、STARエレメントがベクター内に存在するときルシフェラーゼ発現は低い（図17B、状況2）。STARエレメントがエンハンサー特性を有する場合、ルシフェラーゼ発現はSTAR含有ベクターにおいて活性化される（図17B、状況3）。

【0143】

材料及び方法：

ルシフェラーゼ発現ベクターは、pGL3-E-boxルシフェラーゼ（W.Romanowのギフト）を生成するために、E-boxを及び、プラスミドmu-E5+E2x6-cat(x)（ルエジンスキーラ、1991年）からのヒトアルカリ性フォスファターゼ最小プロモーターを、プラスミドpGL3-basic（Promega E1751）においてルシフェラーゼ遺伝子の上流に挿入することによって構築される。E47発現プラスミドは、pHBApr-1-neoプラスミドにおいてペータ-アクチンプロモーターの制御下でE47オープンリーディングフレームを含む；E47は構造的にこのプラスミド（W.Romanowのギフト）から発現される。

【0144】

STARエレメント1、2、3、6、10、11、18及び27は、ルシフェラーゼ発現ベクター内にクローニングされている。ショウジョウバエscsエレメントおよびニワトリベータ-グロビンHS4-6xコア（「HS4」）エレメントを含むクローンは、ポジティブ対照（それらは、エンハンサーをブロックし且つ固有のエンハンサー特性を有しないとして知られている（チュングラ、1993年；ケルム及びシェドル、1992年））として含まれ、及びエンプティルシフェラーゼ発現ベクターは、ネガティブ対照として含まれている。全てのアッセイは、U-2 OS細胞株を使用し実行された。エンハンサー-ブロッキングアッセイにおいて、E47プラスミドは、ルシフェラーゼ発現ベクター（エンプティベクター、又はSTAR若しくはポジティブ対照エレメントを含む）で同時トランスフェクションされた。エンハンサー-アッセイにおいて、E47プラスミドは、エンハンサー活性のためのポジティブ対照としてSTARのないルシフェラーゼ発現ベクターで同時トランスフェクションされた；全ての他のサンプルは同時トランスフェクションの間にモック（mock）プラスミドを受けた。この一時的にトランスフェクションされた細胞は、プラスミドトランスフェクションの48時間後に、ルシフェラーゼ活性をアッセイされた（上述）。E-boxを含まない又はSTAR/コントロールエレメントを含むプラスミドから発現したルシフェラーゼ活性は差し引かれ、及びルシフェラーゼ活性は、記載されたように（上述）タンパク質含量に対して正規化された。

【0145】

結果：

図18は、エンハンサー-ブロッキングアッセイの結果を示す。STARエレメント（又は既知のエンハンサー-ブロッキングエレメントscs及びHS4）の不存在下で、E-47/E-boxエンハンサー複合体は、ルシフェラーゼ（「ベクター」）の発現を活性化する。；発現のこの助長されたレベルは、100に正規化された。エンハンサー活性は、テストされた全てのSTARエレメントによってブロックされる。エンハンサー活性は、予想されたように（ベルラ、2001年；ゲラシモワ及びコルセス、2001年）、HS4及びscsエレメントによってまたブロックされる。これらの結果は、転写的サイレンシングの拡散をブロックするそれらの能力（ネガティブ影響）に加えて、STARエレメントはエンハンサーの活性をブロックすることができる（ポジティブ影響）ことを示す。

【0146】

図19は、エンハンサー-アッセイの結果を示す。E47/E-box複合体による増強によるルシフェラーゼ発現のレベルは、100にセットされた（「E47」）。比較によると、STARエレメントのいずれも、ルシフェラーゼ発現の有意な活性化をもたらさない。予想されるように、scs及びHS4エレメントはレポーター遺伝子の活性をまた生じない。それ故に、少なくともテストされたSTARエレメントはエンハンサー特性を保有しないと結論付けられる。

【実施例19】

【0147】

サイレンシングを導出するクロマチン（SINC；Silencing Inducing Chromatin）エレメントの特徴付け

材料及び方法：

SINCスクリーンの一般的な特性は実施例1に記載されており、その幾つかの観点はここに要約される。ゲノムDNAにおけるSINCエレメントをスクリーニングするために使用されるpSSベクターの一つのバージョンは、pSS-codA::uppである(図20)。それは、2つのSTAR6エレメントによって隣接された自殺遺伝子発現ユニットから成る。Tet-Offプロモーターの制御下のcodA::upp自殺遺伝子から成る発現ユニットは、BglIII制限サイトの下流にある。pSSベクターの第二のバージョン、pSS-hrGFP(図21)は、一つのSTAR6エレメントをSTAR8で置換し及び自殺遺伝子を緑色蛍光タンパク質(Stratagene 240059)をコードするhrGFP遺伝子で置換することによって生成された。22番染色体(Research Genetics 96010-22)からのヒトゲノムDNAは、Sau3AIで部分的に消化され、かつサイズフラクション化された。0.5から10キロ塩基対フラクションは、pSS-codA::uppのBglIII部位内にライゲーションされた。このライブラリーは、1.2キロ塩基対の平均挿入サイズを有する~20,000独立クローンを表わした。該ライブラリーは、大腸菌内で増幅された。増幅されたライブラリーからの精製されたDNAは、標準技術(リン酸カルシウム;Life Technologies 18306-019)によってU-2 OS/Tet-Off細胞(ファンデルフラッグル、2000年)内にトランسفェクションされた。対照トランسفェクションは、エンプティpSS-codA::uppベクターDNAを使用して実行され、2400のハイグロマイシン耐性コロニーを生じた。トランسفェクションされた細胞は、高ドシサイクリン(10ng/ml)で3週間にわたりハイグロマイシン耐性(25mg/ml)のために選択され、そして1800のハイグロマイシン耐性コロニーは、ライブラリートransフェクションから回収された。その後、これらのコロニーは、プロドラッグ5-フルオロシトシン(5-FC)を4日間で5mg/mlの増加になるように1mg/mlで、10ng/mlのドキシサイクリン濃度でインキュベートされた。3週間後、全ての(3個の弱く成長するものを除き)(エンプティpSS-codA::uppでトランسفェクションされた)対照コロニーは、死んだ;ライブラリートransフェクションされたコロニーの58が生き残った。これらのコロニーは、プロドラッグ処理から回収されることを可能にし、そしてさらに培養された。5-FC耐性単離物が回収され、該細胞は溶解され、そしてSINCエレメントを回収するために、DNAの一部分がプライマーD30及びD51を使用しPCR増幅に付された。6つの5-FC耐性コロニーからのPCR生成物は、慣用の方法(サムブルックら、1989年)によって、pBluescript II SK(+)プラスミド(Stratagene 212207)のHindIII及びXbaI部位の間でクローニングされた。候補SINCエレメントのDNA配列は、pBluescriptベクターのための商業的に入手可能なプライマー(Stratagene 300301及び300302)を使用し、記載されたように決定された(上述)。これらのSINCエレメントの配列は、表4Bにあらわされる。

【0148】

6個の候補SINCエレメントが、それらの自然の配向においてプラスミドpSS-hrGFP内にクローニングされ、及び生じたプラスミドは、U-2OS/Tet-Off細胞内にトランسفェクションされた。ハイグロマイシン耐性のための選択後、pSS-hrGFP-SINC形質転換体の集団は、高いドキシサイクリン濃度(10ng/ml)でさらに培養された。合計の細胞性RNAは、製造者によって記載されたように、RNeasy Mini Kit(QIAGEN 74104)を使用して抽出された。これらの集団におけるGFP mRNA豊富さのノーザンプロット法分析は、標準技術(サムブルックら、1989年)を使用して評価された。GFPプローブは、phrGFP-1における塩基対690~1419を含むBamHI-EcoRIフラグメントであった。プロットは、pSS-hrGFP導出されたプラスミドコピー数のための対照としてハイグロマイシンmRNAについて且つ遺伝子的にコードされたmRNA量のための対照としてベータアクチンについてまたプローブされた。ハイグロマイシンプローブは、pREP4(Invitrogen)における8219~10144から伸張するSful-Sal Iフラグメントであり、及びベータ-アクチンプローブは、Clontech, #9800-1からであった。ハイブリダイゼーション及び洗浄後、該プロットはフォスファーイメージャースクリーンにさらされ、及び放射活性シグナルが、BioRad Personal F/X phosphorimagerを使用して可視化され且つ数値化された。

【0149】

結果：

GFPレポーター遺伝に近接するクローニングされたSINCエレメントは、レポーター遺伝子転写のサイレンシングを導出するであろうが、他の遺伝子の転写に影響を与えないであろう。SINC活性の正確な測定は、絶対的なGFP発現を単に測定するのではなくて、2つの参照遺伝子（reference gene）の発現に相対的なGFPの発現を決定することにより、この事実を利用する。一つの参照遺伝子は、pSS-hrGFPプラスミド（STARエレメントによって定義されるドメインの外側、図21）上のハイグロマイシン耐性遺伝子であり、及び他はゲノムベータ-アクチン遺伝子である。SINC活性は、ハイグロマイシン及びベータ-アクチンシグナルに対するGFPシグナルの割合における減少としてRNAプロット分析によって数値化される。特徴付けられている候補SINCエレメントのうち、あるものは、GFP転写において有意な相対的な減少を示し、これらのDNAはサイレントクロマチンの形成を導出することができることを示す。SINC35エレメント（表4BにおいてラベルされたPSINCS35）は、これらの候補の最強の活性を有する。それは、GFP/ハイグロマイシン割合において69%減少及びGFP/ベータ-アクチシグナルにおける75%減少をもたらす。元の明細書中に記載された他の5つの候補及びその明細書の提出後に分離され且つ特徴付けられた多数の他の候補SINCエレメントにおけるSINC活性の強度は、より弱い。それ故に、SINC35は、多様なバイオテクノロジー的用途において、サイレントクロマチンの導出のための有力な遺伝的エレメントとして優れた能力を有する。

【実施例20】

【0150】

STARエレメントはマウスとヒトの間で保存される

ヒトゲノムデータベース（<http://genome.ucsc.edu/cgi-bin/hgGateway>）に対するSTAR DNA配列のBLAT分析は、これらの配列の幾つかがヒトゲノムの他の領域と高い配列保存を有することを明らかにした。これらの複製された領域は、候補STARエレメントである；それらがSTAR活性を示さない場合、それらはクローニングされたSTARのパラログ（paralogs）と考えられる（2個の遺伝子又は遺伝子的エレメントは、それらが複製イベントから得られる場合、パラロガス（paralogous）であると言われる（リー、1997年））。

【0151】

マウスゲノム（http://www.ensembl.org/Mus_musculus/blastview）に対するヒトSTARのBLAST分析は、マウスとヒトとの間に高い配列保存の領域をまた表わす。この配列保存は、65個のヒトSTARエレメントのうちの15個のフラグメントに見られている。保存は64%～89%の範囲であり、141個の塩基ないし909塩基対の長さにわたる（表8）。配列保存のこれらの程度は、注目すべきであり、これらのDNA配列がマウスゲノム内で同様にSTAR活性を与えうることを示唆する。表8のマウス及びヒトゲノムからの配列の幾つかは、オルトロログ（ortholog）として厳密に定義される（2個の遺伝子又は遺伝子的エレメントは、それらが種形成イベント（リー、1997年）から得られる場合、オルトロガス（orthologous）であると言われる。）。例えば、STAR6は、ヒト及びマウスゲノムの両方においてSLC8A1及びHAAO遺伝子の間にある。他の場合、クローニングされたヒトSTARは、ヒトゲノム内にパラログを有し、そのオルトロログはマウスゲノム内で識別されている。例えば、STAR3aは、ヒト15番染色体の15q11.2領域のフラグメントである。この領域は、ヒト5番染色体上の5q33.3でのDNAフラグメントと96.9%同一性があり（パラロガス）、それはIL12B インターロイキン遺伝子に近い。これらのヒトDNAは、マウス11番染色体上の11B2領域のフラグメントとほぼ80%同一性を共有する。該11B2フラグメントは、（マウス）IL12Bインターロイキン遺伝子にも近い。それ故にSTAR3a及びマウス11B2フラグメントは、パラログとして厳密に定義されうる。マウス及びヒトゲノムにおける高い配列保存の領域の間でSTAR活性が共有されるという仮説を試験するために、マウスにおける保存された配列を有するヒトSTARの一つ、STAR18がより詳細に解析された。オリジナルのSTAR18クローンで検出されたマウスゲノムにおける配列保存は、約500塩基対のためにヒト2番染色体上で左側に伸びる（図22；左及び右は、2番染色体のアームの標準記述に関連する）。この実施例では、我々は、配列保存の領域が、オリジナルのクローンよりも長さにおいてより伸張する「

ヒト」における「自然に生じる（naturally occurring）」STARエレメントを定義するかどうかを試験する。我々は、このSTARエレメントのSTAR機能がマウス及びヒトの間に保存されるかどうかをまた試験する。

【0152】

材料及び方法：

STAR18の周りのマウス/ヒト配列保存の領域は、PCR增幅によってヒトBACクローンRP11-387A1から3つのフラグメントとして回収された：完全な領域（プライマーE93及びE94）、左方向の半分（プライマーE93及びE92）、及び右方向の半分（プライマーE57及びE94）。相同なマウス領域からの対応するフラグメントが、同一の様式でBACクローン RP23-400H17から回収された（夫々、プライマーE95及びE98、E95及びE96、並びにE97及びE98）。全てのフラグメントは、pSelectベクター内にクローニングされ、且つU-2 OS/Tet-Off/LexA-HP1細胞株（上述）内にトランスフェクションされた。トランスフェクション後に、ハイグロマイシン選択は。トランスフェクションされた細胞を選択するために実行された。LexA-HP1タンパク質は、ドキシサイクリン濃度を低くすることによって導出され、且つ抗生物質ゼオシンに耐える（STAR活性の尺度）ためのトランスフェクションされた細胞の能力が、細胞成長を監視することによって評価された。

【0153】

結果：

オリジナルSTAR18クローンは、ゼオシン耐性遺伝子のサイレンシングを防止するその能力に基づき、pSelectベクター内にライゲーションされた、Sau3AI消化されたヒトDNAから分離された。マウスゲノムを有するヒトSTAR18クローン（497塩基対）の配列は、オルトロガスヒト及びマウスSTAR18領域の間に高い配列類似性（72%）を示した。それは、クローニングされた領域の左端を定義するSau3AI部位の直接左側に488塩基対伸張する領域において高い類似性（73%）をまた明らかにした。これらの領域の外で、ヒトとマウスDNAの間の配列類似性は、60%以下に落ちる。

【0154】

図22に示されているように、ヒト及びマウスSTAR18エレメントの両方が、LexA-HP1リプレッサータンパク質を発現する宿主細胞にゼオシンに対する生存を与える。オリジナルの497個塩基対STAR18クローン及びそのマウスオルトログはいずれも、成長する能力を与える（図22、a及びd）。両方のゲノムからの高い類似性の近接する488個塩基対領域はまた、成長する能力を与え、及び事実それらの成長表現型はオリジナルのSTAR18クローンのそれよりもより強健である（図22、b及びe）。配列類似性の全領域が試験された場合、マウスとヒトの両方からのこれらのDNAは成長を与え、及び成長表現型は、2つのサブ-フラグメントよりもより強健である（図22、c及びf）。これらの結果は、ヒトSTAR18のSTAR活性はマウスからのそのオルトログ内に保存されることを示している。これらオルトロガス領域間の高い配列保存は特に注目に値し、なぜならばそれらはタンパク質をコードする配列ではないためであり、それらは変異を通じるそれらの進化的多様性を防いでいるある調節機能を有するという結論を導く。

【0155】

この分析は、オリジナルのスクリーニング・プログラムによって識別されたクローニングされたSTARエレメントが、ある場合において部分的なSTARエレメントを表すかも知れず、且つそれらがその中に埋め込まれるところのゲノムDNAの分析はより強力なSTAR活性を有する配列を識別することができるということを示す。

【実施例21】

【0156】

STARエレメントは特徴的なDNA配列モチーフを含む

STARエレメントは、導入遺伝子発現の観点においてそれらの抗抑制表現型に基づき分離される。この抗抑制表現タイプは、STARエレメントに関連付けられるクロマチン形成を調節する基礎的な生化学的プロセスを反映する。これらのプロセスは、典型的に配列特異的であり且つタンパク質結合又はDNA構造から生じる。これは、STARエレメントがDNA配列類

似性を共有することを示唆する。STARエレメント間の配列類似性の識別は、機能的スクリーン及びテストによってすでに識別されているエレメントの特徴である配列モチーフを提供する。該配列モチーフは、その機能が本特許の請求項に一致するところの新しいSTARエレメントを認識し且つ請求するためにまた有益である。該機能は、真核宿主細胞内で発現される導入遺伝子の改善された収率及び安定性を含む。

【0157】

STARエレメントを特徴付ける配列モチーフを識別する他の利点は、：（1）ゲノムデータベースにおける新しいSTARエレメントの予測及び識別のための探索モチーフの提供、（2）エレメントの修飾のための理論的根拠の提供及び（3）STAR活性の機能的分析のための情報の提供を含む。バイオインフォマティックスを使用し、STARエレメント間の配列類似性が識別された；該結果は、本実施例で表わされる。

【0158】

バイオインフォマティックス及び統計的背景。

調節DNAエレメントは、配列特異的DNA結合性タンパク質との相互作用を介して典型的に機能する。DNAエレメント例えばその調節特性が識別されているがその相互作用タンパク質が知られていないところのSTARエレメントのバイオインフォマティックス分析は、配列モチーフの識別のための統計学的アプローチを要求する。これは、参照配列（例えば完全なヒトゲノム）と比較して、調節DNAエレメント（例えばSTARエレメント）のセットにおいて過剰に現れている短いDNA配列パターンを検出する方法によって達成されうる。該方法は、各調節エレメントにおけるパターンの観察された、及び予想された事象の数を決定する。予想された事象の数は、参照配列における各パターンの観察された事象の数から計算される。

【0159】

DNA配列パターンは、所定の長さ例えば6塩基対のオリゴヌクレオチドでありうる。最も簡単な分析において、4つのヌクレオチド（A、C、G及びT）から成る6塩基対オリゴヌクレオチド（六量体）にとって、 $4^6 = 4096$ の区別されるオリゴヌクレオチド（AAAAAA～TTTTTTまでの全ての組み合わせ）がある。調節配列及び参照配列が完全にランダムであり且つA、C、G及びTヌクレオチドの等しい割合を有した場合、各六量体の予想された頻度は、1/4096 (~ 0.00024) である。しかしながら、参照配列における各六量体の実際の頻度は、G : C塩基対などの含量におけるかたよりの故にこれよりも典型的には異なる。それ故に、参照配列における各オリゴヌクレオチドの頻度は、パターンのための「頻度テーブル」を生成するために、数えることによって経験的に決定される。

【0160】

その後、参照配列の該パターン頻度テーブルは、調節エレメントセットにおける各パターンの事象の予想された頻度を計算するために使用される。予想された頻度は、パターンの事象の観察された頻度と比較される。セットにおける「過剰に表わされている（over-represented）」パターンが識別される；例えば、六量体ACGTGAが、20キロ塩基対の配列において5回生じることを予想されるが、15回生じることを観察されるなら、それは3倍過剰に表わされている。エレメントがゲノム全体と同一の六量体組成を有していた場合、その六量体配列パターンの該15事象の10回が調節エレメント内に予想されない。いったん過剰に表わされているパターンが識別されたなら、統計テストはそれらの過剰表現が有意であるか又は偶然によるものかどうかを決定するために適用される。このテストのために、有意性インデックス、「sig」が各パターンのために計算される。有意性インデックスは、各パターンの事象の蓋然性から導出され、それは2項分布によって見積もられる。蓋然性は可能なパターン（六量体について4096）の数を考慮する。最も高いsig値は、最も過剰に表わされているオリゴヌクレオチドに対応する（ファン ヘルデンら、1998年）。実務において、 $sig >= 0$ であるオリゴヌクレオチドは、過剰に表わされていると考えられる。 $sig >= 0$ であるパターンは、調節エレメント配列のセットにおいて一回の機会（= 10^0 ）により過剰に表れることがありうる。しかしながら、 $sig >= 1$ でパターンが、 10 （= 10^1 ）配列セットにおいて一回、 $sig >= 2$ で 100 （= 10^2 ）配列セットにおいて一回などと過剰

に表わされることが予想される。調節エレメントセットにおいて有意に過剰に表わされているパターンは、分類のためのモデル及び調節エレメント配列予測を展開するために使用される。これは、判別式分析 (Discriminant Analysis)、いわゆる当業者に知られた統計的分類（ヒュベルティ、1994年）の「管理された（supervised）」方法を使用する。判別式分析において、既知の又は分類された項目（例えば、STARエレメント）のセットは、特定の変異体（例えば配列パターン例えば六量体）に基づきそれらの項目を認識するためのモデルを「トレーニングする（train）」ために使用される。次に、トレーニングされたモデルは、他の項目が既知の項目のセットに属するように分類されるべきである（例えば、DNA配列がSTARエレメントである）かどうかを予測するために使用される。この実施例では、トレーニングセットにおける既知の項目が、STARエレメント（ポジティブなトレーニングセット）である。それらは、STARエレメントと同じ長さを有するゲノム（ネガティブなトレーニングセット）からランダムに選択される配列と対照される。判別式分析は、ポジティブを区別する変異体のセットに基づきネガティブからポジティブを区別するための基準を確立する；この実施例では、変異体は、有意に過剰に表わされているパターン（例えば、六量体）である。

【0161】

過剰に表わされているパターンの数がトレーニングセットのサイズと比較して高い場合に、該モデルは過剰なトレーニングによりバイアスをかけられうる。過剰トレーニングは、変異体のフォワードステップワイズ選択を適用することによって回避される（ヒュベルティ、1994年）。ステップワイズ判別式分析の目標は、ポジティブとネガティブとの間で最大の区別を提供する変異体の最小数を選択することである。該モデルは、ポジティブな及びネガティブなトレーニングセットにおける項目を適切に分類するそれらの能力について変異体を一つずつ評価することによりトレーニングされる。これは、モデルへの新しい変異体の追加がモデルの予言的な力を有意に増加することのないまで（すなわち、分類エラー割合が最小化されるまで）行われる。次に、この最適化されたモデルは、「新しい」項目がポジティブか又はネガティブかを予測するための試験をするために使用される（ヒュベルティ、1994年）。

【0162】

複雑な項目例えばDNA配列にとって、ポジティブトレーニングセットのあるエレメントがネガティブ（偽のネガティブ）として分類され、及びネガティブトレーニングセットのあるメンバーはポジティブ（偽のポジティブ）として分類されるであろうことは、分類統計において固有である。トレーニングされたモデルが、新しい項目をテストするために適用される場合、同じタイプの誤分類が生じることが予想される。明細書中に記載したバイオインフォマティクス方法において、第1のステップ、パターン頻度分析は、配列パターンの大きなセット（例えば、合計4096六量体）を、有意に過剰に表わされたパターンの最小セット（例えば100六量体）にまで減少する；第2のステップ、ステップワイズ判別式分析は、過剰に表わされたパターンのセットを、最大の判別力を有するそれらのパターンのサブセット（例えば、5~10六量体）にまで減少する。それ故に、このアプローチは、調節DNAエレメント例えばSTARエレメントを識別するための、簡単且つ強力な基準を提供する。

【0163】

DNA結合性タンパク質は、それらが占領する結合部位のタイプに基づき区別されうる。あるものは隣接している配列を認識する；このタイプのタンパク質にとって、長さ6塩基対（六量体）のオリゴヌクレオチドであるパターンは、バイオインフォマティクス分析にとって実りのあるものである（ファン ヘルデンら、1998年）。他のタンパク質は、二分染色体を配列決定するために結合する：接触は、固定された幅の非保存領域によって分離された高く保存されたトリヌクレオチドの対の間に作成される（ファン ヘルデンら、2000年）。二分染色体結合性タンパク質によって結合されているかも知れないSTARエレメントにおける配列を識別するために、頻度分析がこのタイプのパターンのために行われ、ここで二つのトリヌクレオチド間のスペースは0~20で変化される（すなわち、XXXN{0-20}X

XX、ここでXはトリヌクレオチドを構成する特定のヌクレオチド、及びNは長さにおいて0~20塩基対のランダムヌクレオチドである）。二分染色体頻度分析の結果は、上記したように線形判別式分析のためにまた使用される。

【0164】

材料及び方法：

元の特許出願に記載されている遺伝子的スクリーンを使用し、66個のSTARエレメントがヒトゲノムDNAから最初に分離され、且つ詳細に（表6）特徴付けられた。該スクリーンは、胎座（Clontech 6550-1）から精製された又はバクテリアの/P1染色体（BAC/PAC）人工染色体内に運ばれている、ヒトゲノムDNAのSau3AI消化によって構築された遺伝子ライブラリー上で実行された。BAC/PACクローンは、1番染色体の領域（クローン RP1154H19及びRP3328E19）からの、相同異質遺伝子（クローン RP1167F23、RP1170019及びRP11387A1）のHOXクラスターからの又はヒト22番染色体（Research Genetics 96010-22）からのゲノムDNAを含む。該DNAはサイズフラクションされ、そして0.5~2kbサイズフラクションは、標準技術（サムブルックら、1989年）によって、BamHI-消化されたpSelectベクター内にライゲーションされた。低いドキシサイクリン濃度でゼオシンに対する耐性を与えたヒトゲノムDNAを含むpSelectプラスミドが分離され且つ大腸菌内で増殖された。表6のSTARエレメントを生じたスクリーンは、ヒトゲノムのほぼ1~2%をアッセイした。

【0165】

これら66個のプラスミド内のヒトゲノムDNAは挿入物は、Beckman CEQ2000 automated DNA sequencerを使用し製造者の指示書を用いて、ジデオキシ法（サンガーら、1977年）によってシーケンスされた。簡単に言えば、DNAは、QIAprep Spin Miniprep及びプラスミド Midi Kits（夫々、QIAGEN 27106及び12145）を使用し、大腸菌から精製された。サイクルシーケンシング（cycle sequencing）が、染料ターミネーター（CEQ Dye Terminator Cycle Sequencing Kit, Beckmann 608000）の存在において、pSelectベクター（プライマーD89及びD95、表5）に対応するカスタムオリゴヌクレオチドを使用して実行された。組み立てられたSTARDNA配列は、BLAT（Basic Local Alignment Tool（ケント、2002年））；<http://genome.ucsc.edu/cgi-bin/hgGateway>（表6）を使用し、ヒトゲノム（データベースは、2001年8月に作成された）内に配置された。全体として、一緒にされたSTAR配列は、平均長1.3キロ塩基対とともに、85.6キロ塩基対を含む。

【0166】

ヒトゲノムDNA内のSTARエレメントを区別する配列モチーフは、2段階の工程を使用してバイオインフォマティクス分析によって、次のように（図式的な図の図23を参照）識別された。分析は2つの入力データセットを有する：（1）STARエレメント（STAR1 - STAR65が使用された；表6）のDNA配列、及び（2）ヒトゲノムのDNA配列（1番染色体を除き、それはその大きいサイズにより含むことが実現不可能であった；二分染色体分析にとって、ヒトゲノムDNA配列のランダムサブセット（~27Mb）が使用された）。

【0167】

パターン頻度分析：

分析における第1のステップは、RSA-Toolsソフトウェア（調節配列分析ツール；<http://www.uclm.ulb.ac.be/bioinformatics/rsa-tools/>；参照（ファンヘルデンら、1998年；ファンヘルデンら、2000年、ファンヘルデンら、2000年）を使用し、下記の情報を決定する：（1）ヒトゲノム内の全ての二分染色体及び六量体オリゴヌクレオチドの頻度；（2）65個のSTARエレメントにおけるオリゴヌクレオチド及び二分染色体の頻度；及び（3）ゲノムと比較してSTARエレメント内の過剰に表わされたこれらのオリゴヌクレオチド及び二分染色体の有意性インデックス。対照分析は、表6のSTARエレメントの長さに一致するヒトゲノム（すなわち、 2689×10^3 キロ塩基対）からランダムに選択された65個の配列で行われた。

【0168】

判別式分析：

過剰に表わされたオリゴヌクレオチド及び二分染色体は、線形判別式分析（ヒューベル

ティ、1994年)によるSTARエレメントの予測のためのモデルをトレーニングするために使用された。変異体の予備選択は、頻度分析の過剰に表わされたオリゴヌクレオチド(oligos)又は二分染色体から最高の夫々の判別力を有する50個のパターンを選択することによって実行された。次に、これらの予備選択された変異体は、変異体の最も判別できる組み合わせを選択するために(ヒュベルティ、1994年)、ステップワイズ線形判別式分析におけるモデルトレーニングのために使用された。変異体選択は、分類エラー割合(誤ったネガティブ分類のパーセント)を最小化することに基づいていた。さらに、予想されるエラー割合は、(誤ったポジティブ分類のパーセントを最小化する)ランダム配列の対照セットに対する同一の判別アプローチを適用することによって見積もられた。

【0169】

判別式分析のトレーニング段階からの予測モデルは、2つの方法でテストされた。第1に、モデル(トレーニングセット)を生成するために使用されたSTARエレメント及びランダム配列が分類された。第2に、19個の候補STARエレメント(上記したように、ゼオシン選択によって最近クローニングされた)のコレクションにおける配列が分類された。これら候補STARエレメントは、表11(配列ID:67-84)にリストされる。

【0170】

結果:

パターン頻度分析は、参照配列としてヒトゲノムを使用し、65個のSTARエレメントに対してRSA-Toolsで実行された。166個の六量体のオリゴヌクレオチドがゲノム全体(表9)と比較してSTARエレメントのセット内で過剰に表わされる($\text{sig} >= 0$)ことが見つかった。最も大きく過剰に表わされるオリゴヌクレオチドCCCCACは、65個のSTARエレメント間で107回生じ、しかし49回のみ生じることが予想される。それは8.76の有意性係数を有する;言い換えれば、その過剰に表わされたことがランダム偶然によるものであることの蓋然性は、 $1/10^{8.76}$ 、すなわち5億回に1回よりも少ない。

【0171】

オリゴヌクレオチドの95個は、1よりも大きい有意性係数を有し、及びそれ故にSTARエレメントにおいて高く過剰に表わされている。過剰に表わされたオリゴヌクレオチドのうち、それらの観察された、及び予想された事象は、夫々6及び1(オリゴ163について、CGCGAA、 $\text{sig}=0.02$)から133及び95(オリゴ120について、CCCAGG、 $\text{sig}=0.49$)までの範囲である。予想された事象における相違は、ファクター例えばヒトゲノムのG:C含量を反映する。それ故に、それらの事象の数におけるオリゴヌクレオチド間の相違は、それらの過剰に表わされていることよりも重要ではない;例えば、オリゴ2(CAGCGG)は $36/9 = 4$ 倍過剰に表わされており、それは5千万回に1回のランダムチャンス($\text{sig}=7.75$)の蓋然性を有する。

【0172】

表9は、夫々の過剰に表わされたオリゴヌクレオチドが見つかったところのSTARエレメントの数をまた表わす。例えば、最も重要なオリゴヌクレオチド、オリゴ1(CCCAC)は107回生じ、しかし51個のSTARのみにおいて見つかっている。すなわち、それは平均してSTARあたり2コピーとして生じる。最も少なく豊富なオリゴヌクレオチド、番号166(AATCGG)は、平均してSTARあたり単一のコピーとして生じる(11個のSTAR上で13個の事象);単一のコピーオリゴヌクレオチドはしばしば、特に比較的低い豊富さのオリゴを生じる。他の極端な場合、オリゴ4(CAGCCC)はそれが見つけられるところのSTAR(37個のSTAR)において平均3回生じる。最も広まっているオリゴヌクレオチドは番号120(CCCAGG)であり、それは58個のSTAR上で生じ(STARあたり平均2回)、もっとも少なく広まっているオリゴヌクレオチドは番号114(CGTCGC)であり、それは6個のSTARのみで(STARあたり平均1回のみ)生じる。

【0173】

二分染色体頻度分析の結果が、表10に与えられる。730個の二分染色体が、参照配列と比較してSTARエレメントのセットにおいて過剰に表わされていた($\text{sig} >= 0$)。最も著しく過剰に表わされた二分染色体、CCCN{2}CGGは、65個のSTARエレメントの間で36回生じ

、しかし7回のみ生じると予測される。それは、9.31の有意性係数を有する；言い換れば、その過剰に表わされてさいることは偶然によるものであるとの蓋然性は、 $1/10^{9.3}$ ¹、すなわち20億回に1回よりも少ない。

【0174】

397個の二分染色体は、1よりも多い有意性係数を有し、且つそれ故にSTARエレメント内で高く過剰に表わされている。該過剰に表わされている二分染色体の間で、観察された、及び予想されたそれらの事象は、夫々9及び1（5個の二分染色体（番号380、435、493、640及び665）について）から118及び63（番号30（AGGN{2}GGG）について、 $\text{sig} > = 4.44$ ）までの範囲である。

【0175】

パターン頻度分析によってSTARエレメントにおいて過剰に表わされていることが見つかったオリゴヌクレオチド及び二分染色体は、線形判別式分析によってそれらの判別の能力について試験された。判別モデルは、50個の最も判別できるオリゴヌクレオチド（表9）又は二分染色体（表10）パターンの間で最良の組み合わせのステップワイズ選択によってトレーニングされた。該モデルは、4個の（二分染色体）又は5個の変異体の組み込み後に最適なエラー割合で達成された。オリゴ分析から判別可能な変異体は、番号11、30、94、122及び160（表9）であり；二分染色体解析からのそれらは、番号73、194、419及び497（表10）である。

【0176】

次に、判別モデルは、トレーニングセットおよびそれらの関連付けられたランダム配列において65個のSTARエレメントを分類するために使用された。オリゴヌクレオチド変異体を使用する該モデルは、STARエレメント（真のポジティブ）として65個のSTARエレメントのうち46個を分類し；二分染色体モデルは、真のポジティブとしてSTARエレメントのうち49個を分類する。組み合わせで、該モデルは、STARエレメントとして65個のSTARエレメントのうち59個を分類する（91%、図24）。誤ったポジティブ割合（STARとして分類されるランダム配列）は、二分染色体モデルにつき7個、オリゴヌクレオチドモデルにつき8個及び2つのモデルの結合された予測について13個（20%）であった。LDAによってSTARとして分類されなかった表6のSTARエレメントは、STAR7、22、35、44、46及び65である。これらのエレメントは、機能的アッセイにおいて抗リプレッサー活性の安定化を示し、それ故にLDAによってそれらはSTARとして分類されないという事実はそれらがSTARエレメントの他のクラス（又は複数のクラス）を表していることを示唆する。

【0177】

次に、該モデルは、表11にリストされたテストセットにおいて19個の候補STARエレメントを分類するために使用された。二分染色体モデルは、STARエレメントとしてこれらの候補STARの12個を分類し、オリゴヌクレオチドモデルは、STARとして14個を分類する。STARエレメントとして分類される候補の組合わされた数は、15(79%)個である。これは、65個のSTARのトレーニングセットで得られるよりも低い割合の分類であり；これは、2つの理由から予想される。第一に、判別モデルは表6の65個のSTARでトレーニングされ、このトレーニングセットに基づく判別可能な変異体は、テストセットにおいて比較的よく表わされていないかも知れない。第二に、テストセットにおける候補STAR配列は、インビボ（in vivo）機能の観点から十分にまだ特徴付けられていず、弱い抗抑制特性のみを有するエレメントを含んでいるかも知れない。

【0178】

この分析は、STARエレメントのバイオインフォマティックス分類に対する統計的アプローチの能力を示す。STAR配列は、全体としてヒトゲノムと比較して有意に過剰に表わされている多くの二分染色体及び六量体オリゴヌクレオチドパターンを含む。これらのパターンは、STAR活性を与えるタンパク質の結合部位を表してもよく；いずれにしても、それらはSTARエレメント配列を認識するために使用されるうる配列モチーフのセットを形成する。

【0179】

判別式分析によってSTARエレメントを認識するためにこれらのパターンを使用し、本発

明の遺伝子的スクリーンによって取得されるエレメントの高い割合が、STARとして事実上分類される。これは、これらのエレメントの間の基礎的な配列及び機能的な類似性を反映する。本明細書に記載されている方法（パターン頻度分析及び引き続く判別式分析）の重要な観点は、それが反復可能であることである；例えば一つのトレーニングセット内において表6の66個のSTARエレメントを有する表11の19個の候補STARエレメントを含むことによって、改善された判別モデルがトレーニングされうる。次に、この改善されたモデルは、STARとして他の候補調節エレメントを分類するために使用されうる。本発明の方法を使用し、バイオインフォマティックス分析の反復と組み合わせた、ゲノム配列の大規模イン

ビボ(*in vivo*)スクリーニングは、ゲノムがその全体でスクリーンされる場合にエレメントの100%認識及び予測に漸近的に近づく、STARエレメントを判別する手段を提供する。STAR機能のこれら説得力のある且つ総合的な予測は、全てのヒトSTARエレメントが認識され且つ導入遺伝子発現を改善する際に使用するために利用可能であるということを保証する。

【実施例22】

【0180】

シロイヌナズナ(*Arabidopsis thaliana*)からのSTARエレメントのクローニング及び特徴付け

導入遺伝子サイレンシングは、転写的及び転写後のレベルのいずれにおいても形質転換植物において生じる(メイヤー、2000年；ヴァンス及びヴァウチエルト、2001年)。いずれの場合にも、導入遺伝子発現の望ましい結果は、サイレンシングによって危うくされうる；導入遺伝子の低い発現及び不安定性は望ましい形質(例えば害虫耐性)の貧しい発現又は組み換えタンパク質の低い収率を生じる。それは、貧弱な予測可能性をまた生じる：バイオテクノロジー的に有益なレベルで導入遺伝子を発現する形質転換植物の割合は低く、それは、有益な発現特性を有するもののために形質転換された個体の骨の折れ且つ高価なスクリーニングを必要とする。本実施例は、形質転換植物における転写的な導入遺伝子サイレンシングを防ぐ際に使用するための、双子葉植物シロイヌナズナ(*Arabidopsis thaliana*)のゲノムからのSTARエレメントの分離を記載する。シロイヌナズナは、それが十分に研究されたモデル生物体である故に、本実施例のために選択された：それはコンパクトなゲノムを有し、それは遺伝子的及び組み換えDNA操作に対して敏感に反応し、かつそのゲノムは配列決定されている(ベヴァンら、2001年、イニチアティブ、2000年、マイケル、1998年)。

【0181】

材料及び方法：

ゲノムDNAは、記載されているように(スタムら、1998年)、*Arabidopsis thaliana* ecotype Columbiaから分離され、且つMboIで部分的に消化された。該消化されたDNAは、アガロースゲル電気泳動及び該ゲルからの精製(QIAquick Gel Extraction Kit, QIAGEN 28706)によって0.5~2キロ塩基対にサイズフラクションされ、引き続きpSelectベクター(上述)内にライゲーションされた。U-2 OS/Tet-Off/LexA-HP1細胞株内へのトランスフェクション及び低いドキシサイクリン濃度でのゼオシン耐性の選択は、前記した様に(上述)実行された。プラスミドは、ゼオシン耐性コロニーから分離され且つU-2 OS/Tet Off/LexA-HP1細胞株内にトランスフェクションされた。

【0182】

再トランスフェクションに応答してゼオシン耐性を与えられたシロイヌナズナゲノムDNAフラグメントの配列決定は、前記した様に(上述)実行された。DNA配列は、BLAST分析((アルツシュルら、1990年)；URL <http://www.ncbi.nlm.nih.gov/blast/Blast>)によってシロイヌナズナゲノムの配列と比較された。

【0183】

STAR活性は、逆転写PCR(RT-PCR)によって組み換え宿主細胞におけるハイグロマイシン-及びゼオシン-耐性遺伝子のmRNAレベルを測定することによってさらにテストされた。U-2 OS/Tet-Off/LexA-HP1細胞株の細胞は、シロイヌナズナSTARエレメント、又はショウ

ジョウバエscsエレメントを含むpSelectプラスミド又は挿入物を含まないもの（上述）でトランスフェクションされた。これらは、高いドキシサイクリン濃度で2週間の間、ハイグロマイシンで培養され、その後ドキシサイクリン濃度は、lexA-HP1リプレッサータンパク質を導出するために0.1ng/mlに少なくされた。10日後、総RNAは、RNeasy mini kit (QI AGEN 74104)によって製造者によって記載されてように単離された。第一ストランドcDNA合成は、oligo(dT)18プライマーを使用するRevert Aid First Strand cDNA Synthesis kit (MBI Fermentas 1622)を使用して製造者によって記載されているように実行された。cDNAの分けた一部分 (aliquot) は、(ゼオシンマーカーについて) プライマーD58及びD80、及び(ハイグロマイシンマーカーについて) D70及びD71を使用し、並びにTaq DNAポリメラーゼ (Promega M2661)を使用しPCR反応におけるテンプレートとして使用された。反応条件は、1分間の間94、1分間の間54、及び90秒の間72 の15~20サイクルであった。これらの条件は、投入したRNAとPCR生成物DNAの間に直線関係を生じる。PCR生成物はアガロースゲル電気泳動法によって分離され、及びゼオシン及びハイグロマイシンバンドは、テンプレートとして、精製されたpSelectプラスミドを有する前記の様に生成されたPCR生成物を使用し、前記したように（サムブルックら、1989年）サザンプロット法によって検出された。ゼオシンとハイグロマイシンシグナルの割合は、ゼオシン遺伝子の正規化された発現レベルに対応する。

【0184】

結果：

pSelectベクターにおけるシロイヌナズナゲノムDNAのライプラリーは、大腸菌内に69,000個のプライマリークローンを含み、その80%が挿入物を持っていた。平均の挿入物サイズは、ほぼ1000塩基対であった；それ故にライプラリーは、シロイヌナズナゲノムのほぼ40%を表す。

【0185】

(シロイヌナズナゲノムのほぼ16%を表している)このライプラリーの部分は、U-2 OS/Tet-Off/LexA-HP1細胞株内にトランスフェクションされた。ハイグロマイシン選択は、形質転換体を分離するために行われ、それは27,000個の生き残ったクローンをもたらした。その後、これらは、低いドキシサイクリン濃度でゼオシン選択に付された。56個のゼオシン耐性コロニーからの想定されるSTAR含有プラスミドが大腸菌内にレスキューされ且つU-2 OS/Tet-Off/LexA-HP1細胞内に再トランスフェクションされた。これらのプラスミドの44個（テストされたプラスミドの79%）は、低いドキシサイクリン濃度で宿主細胞に対してゼオシン耐性を与え、プラスミドがSTARエレメントを有することを示している。これは、ヒトU-2 OS細胞におけるpSelectスクリーンは、植物ゲノムDNAからのSTARエレメントの検出で高く効率的であることを示す。

【0186】

これら44個の候補STARエレメントのDNA配列が決定された。それらのうち35個は、シロイヌナズナ核ゲノム配列のデータベースにおける単一座であるとして識別された（表12；配列ID:85~配列ID:119）。他の4個は、葉緑体ゲノム由来であるとして識別され、4個は2つの座からのDNAフラグメントのキメラであり、及び1個はシロイヌナズナゲノムデータベースにおいて見つからなかった。

【0187】

クローニングされたシロイヌナズナSTARエレメントの強度は、RT-PCRアッセイを使用し、ゼオシン耐性遺伝子の転写抑制を防ぐそれらの能力を評価することによってテストされた。それらのサンプルの間で投入されたRNAの対照として、各STARトランスフェクションのためのハイグロマイシン耐性遺伝子の転写レベルもまた評価された。この分析は、シロイヌナズナSTARエレメントの12個について実行されている。該結果（図25）は、シロイヌナズナSTARエレメントが、転写抑制からゼオシン耐性遺伝子を保護するそれらの能力においてショウジョウバエscsエレメント（ポジティブ対照）及びエンブティベクター（「SV40」；ネガティブ対照）よりも優れていることを示す。特に、STAR-A28及びSTAR-A30は、lexA-HP1リプレッサーが発現される場合、（ハイグロマイシン耐性遺伝子mRNAの内部対

照に正規化された) scsエレメントよりも2倍高いゼオシン耐性遺伝子発現のレベルを可能にする。

【0188】

これらの結果は、本願発明の方法が、ヒト以外の他の種のゲノムからSTARエレメントを回収するために成功裡に適用されうることを示す。植物ゲノムからのSTARエレメントに対するその成功的な適用は特に重要である。何故ならば、それは本発明の方法が適用可能であるところの広い分類学的範囲を示すからであり且つ植物はバイオテクノロジー的発展の重要なターゲットであるからである。

【0189】

参考文献

(アルツシュル及びギッシュ、1996年) Altschul, S. F. and Gish, W. Local alignment statistics. *Methods Enzymol.*, 266巻, 460-480頁, (1996年).

(アルツシュルら、1990年) Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. Basic local alignment search tool. *J Mol Biol.*, 215巻, 403-410頁, (1990年).

(ベルら、2001年) Bell, AC, West, AG, and Felsenfeld, G. Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. *Science*, 291巻, 447-50頁, (2001年).

(ベネットら、1998年) Bennett, R. P., Cox, C. A. and Hoeffler, J. P. Fusion of green fluorescent protein with the Zeocin-resistance marker allows visual screening and drug selection of transfected eukaryotic cells. *Biotechniques*, 24巻, 478-482頁, (1998年).

(ベルゲルら、1988年) Berger, J, Hauber, J, Hauber, R, Geiger, R, and Cullen, BR. Secreted placental alkaline phosphatase:a powerful new quantitative indicator of gene expression in eukaryotic cells. *Gene*, 66巻, 1-10頁, (1988年).

(ベヴァンら、2001年) Bevan, M, Mayer, K, White, O, Eisen, JA, Preuss, D, Bureau, T, Salzberg, SL, and Mewes, HW. Sequence and analysis of the *Arabidopsis* genome. *Curr Opin Plant Biol.*, 4巻, 105-10頁, (2001年).

(ビリューゼンら、1997年) Bierhuizen, M.F., Westerman, Y., Visser, T. P., Wognum, A. W. and Wagemaker, G. Green fluorescent protein variants as markers of retroviral-mediated gene transfer in primary hematopoietic cells and cell lines. *Biochem Biophys Res Commun*, 234巻, 371-375頁, (1997年).

(ボイヴィン及びデュラ、1998年) Boivin, A, and Dura, JM. In vivo chromatin accessibility correlates with gene silencing in *Drosophila*. *Genetics*, 150巻, 1539-49頁, (1998年).

(ボシャルトら、1985年) Boshart, M, Weber, F, Jahn, G, Dorsch-Hasler, K, Fleckenstein, B, and Schaffner, W. A verystrong enhancer is located upstream of an immediate early gene of humancytomegalovirus. *Cell*, 41巻, 521-30頁, (1985年).

(ブレッケンリッジ及びショール、2000年) Breckenridge, D. G. and Shore, G. C. Regulation of apoptosis by E1A and Myc oncoproteins. *Crit Rev Eukaryot Gene Expr*, 10巻, 273-280頁, (2000年).

(ブンケル及びキングストン、1994年) Bunker, C.A. and Kingston, R. E. Transcriptional repression by Drosophila and mammalian Polycomb group proteins in transfected mammalian cells. *Mol Cell Biol*, 14巻, 1721-1732頁, (1994年).

(チェリーら、1999年) Cherry, J. R., Lamsa, M. H., Schneider, P., Vind, J. , Svedsen, A. , Jones, A. and Pedersen,A. H. Directed evolution of a fungal peroxidase. *Nat Biotechnol*, 17巻, 379-384頁, (1999年).

(チュングら、1993年) Chung, JH, Whiteley, M, and Felsenfeld, G. A 5'element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in *Drosophila* Cell. 74巻, 505-14頁,(1993年).

(ドイシュレら、1995年) Deuschle, U,Meyer, WK, and Thiesen, HJ. Tetracycline-reversible silencing of eukaryotic promoters. *Mol Cell Biol*, 15巻, 1907-14頁, (1995年).

(ドールら、1996年) Doll, R. F., Crandall, J. E. , Dyer, C. A. , Aucoin, J. M. and Smith, F. I. Comparison of promoter strengths on gene delivery into mammalian brain cells using AAV vectors. *Gene Ther*, 3巻, 437-447頁, (1996年).

(アイゼンベルグら、1990年) Eissenberg, J. C., James T. C., Foster-Hartnett D. M ., Hartnett T., Ngan V., and Elgin S. C. R. Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in *Drosophila melanogaster*. *Proc Natl Acad Sci (USA)*, 87巻, 9923-9927頁, (1990年).

(フェングら、1999年) Feng, Y. Q., Seibler, J., Alami, R. , Eisen, A. , Westerman, K. A. , Leboulch, P., Fiering,S. and Bouhassira, E. E. Site-specific chromosomal integration in mammalian cells: highly efficient CRE recombinase mediated cassette exchange. *J Mol Biol*, 292巻, 779-785頁, (1999年).

(フェッキング及びホフステッテル、1986年) Foecking,MK, and Hofstetter, H. Powerful and versatile enhancer promoter unit for mammalian expression vectors. *Gene*, 45巻, 101-5頁, (1986年).

(ガーリックら、1998年) Garrick, D,Fiering, S, Martin, DI, and Whitelaw, E. Repeat-induced gene silencing in mammals. *Nat Genet*, 18巻, 56-9頁,(1998年).

(ガスツナーら、1999年) Gaszner, M., Vazquez, J. and Schedl, P. The Zw5 protein, a component of the scs chromatin domain boundary, is able to block enhancer promoter interaction. *Genes Dev*, 13巻, 2098-2107頁, (1999年).

(ゲラシモヴァ及びコルセス、1998年) Gerasimova,T. I. and Corces, V. G. Polycomb and trithorax group proteins mediate the function of a chromatin insulator. *Cell*, 92巻, 511-521頁, (1998年).

(ゲラシモワ及びコルセス、2001年) Gerasimova, TI, and Corces, VG. Chromatin insulators and boundaries: effects on transcription and nuclear organization. *Annu Rev Genet*, 35巻, 193-208頁, (2001年).

(ゴッセン及びビュヤード、1992年) Gossen, M. and Bujard, H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc Natl Acad Sci U S A, 89巻, 5547-5551頁, (1992年).

(グラハム及びファン デル イービー、1973年) Graham, F. L. and van der Eb, A. J. Transformation of rat cells by DNA of human adenovirus 5. Virology, 54巻, 536-539頁, (1973年).

(ヘルディンら、1986年) Heldin, C. H., Johnsson, A., Wennergren, S., Wernstedt, C., Betsholtz, C. and estermark, B. A human osteosarcoma cell line secretes a growth factor structurally related to a homodimer of PDGF A-chains. Nature, 319巻, 511-514頁, (1986年).

(ヘンケ及びボルンショイマー、1999年) Henke, E. and Bornscheuer, U. T. Directed evolution of an esterase from *Pseudomonas fluorescens*. Random mutagenesis by error-prone PCR or a mutator strain and identification of mutants showing enhanced enantio selectivity by a resorufin-based fluorescence assay. Biol Chem, 380巻, 1029-1033頁, (1999年).

(ヘンソルンら、1988年) Henthorn, P., Zervos, P., Raducha, M., Harris, H., and Kadlec, T. Expression of a human placental alkaline phosphatase gene in transfected cells: use as a reporter for studies of gene expression. Proc Natl Acad Sci U S A, 85巻, 6342-6頁, (1988年).

(ヒギンスら、1996年) Higgins, D. G., Thompson, J. D. and Gibson, T. J. Using CLUSTAL for multiple sequence alignments. Methods Enzymol, 266巻, 383-402頁, (1996年).

(ヒメス及びシャノン、2000年) Himes, S. R. and Shannon, M. F. Assays for transcriptional activity based on the luciferase reporter gene. Methods Mol Biol, 130巻, 165-174頁, (2000年).

(ヒュベルティ、1994年) Huberty, CJ Applied discriminant analysis, Wiley and Sons, New York, (1994年).

(イニチアティブ、2000年) Initiative, AG. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature, 408巻, 796-815頁, (2000年).

(イズミ及びギルバート、1999年) Izumi, M. and Gilbert, DM. Homogeneous tetracycline-regulatable gene expression in mammalian fibroblasts. J Cell Biochem, 76巻, 280-9頁, (1999年).

(ジュングラ、2000年) Jung, R., Soondrum, K. and Neumaier, M. Quantitative PCR. Clin Chem Lab Med, 38巻, 833-836頁, (2000年).

(カイン、1997年) Kain, SR. Use of secreted alkaline phosphatase as a reporter of gene expression in mammalian cells. Methods Mol Biol, 63巻, 49-60頁, (1997年).

(カオ及びプック、1968年) Kao, F. T. and Puck, T. T. Genetics of somatic mammali

an cells, VII. Induction and isolation of nutritional mutants in Chinese hamster cells. Proc Natl Acad Sci U S A, 60巻, 1275-81頁, (1968年).

(ケルム及びシェドル、1992年) Kellum, R. and Schedl, P. A group of scs elements function as domain boundaries in an enhancer-blocking assay. Mol Cell Biol, 12巻, 2424-2431頁, (1992年).

(ケント、2002年) Kent, W.J. BLAT--theBLAST-like alignment tool. Genome Res, 12巻, 656-64頁, (2002年).

(クノフラーら、2002年) Knofler, M., Meinhardt, G., Bauer, S., Loregger, T., Vasicek, R., Bloor, DJ., Kimber, SJ., and Husslein, P. Human Hand1 basic helix-loop helix (bHLH) protein: extra-embryonic expression pattern, interaction partners and identification of its transcriptional repressor domains. Biochem J, 361巻, 641-51頁, (2002年).

(リー、1997年) Li, W-H Molecular Evolution, Sinauer Associates, Sunderland MA, (1997年).

(マインケら、1998年) Meinke, DW, Cherry, JM, Dean, C, Rounsley, SD, and Koornneef, M. Arabidopsis thaliana: a model plant for genome analysis. Science, 282巻, 662巻, 679-82頁, (1998年).

(リットら、2001年) Litt, M. D., Simpson, M., Recillas-Targa, F., Prioleau, M. N. and Felsenfeld, G. Transitions in histone acetylation reveal boundaries of three separately regulated neighboring loci. EMBO J, 20巻, 2224-2235頁, (2001年).

(メイヤー、2000年) Meyer, P. Transcriptional transgene silencing and chromatin components. Plant Mol Biol, 43巻, 221-34頁, (2000年).

(モルゲンステルン及びランド、1990年) Morgenstern, J. P. and Land, H. Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. Nucleic Acids Res, 18巻, 3587-3596頁, (1990年).

(ミューレンら、1992年) Mullen, C. A., Kilstrup, M. and Blaese, R. M. Transfer of the bacterial gene for cytosine deaminase to mammalian cells confers lethal sensitivity to 5-fluorocytosine: a negative selection system. Proc Natl Acad Sci U S A, 89巻, 33-37頁, (1992年).

(ナンら、1997年) Nan, X., Javier Campoy, F., and Bird A. MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. Cell, 88巻, 471-481頁, (1997年).

(ピーターソンら、2002年) Petersson, K., Ivars, F., and Sigvardsson, M. The pT alpha promoter and enhancer are direct targets for transactivation by E box-binding proteins. Eur J Immunol, 32巻, 911-20頁, (2002年).

(パジン及びカドナガ、1998年) Pazin, M. J. and Kadonaga, J. T. Transcriptional and structural analysis of chromatin assembled in vitro. In Gould, H. (ed.) Chroma

tin: A Practical Approach. Oxford University Press, Oxford, 172-194頁, (1998年).

(ピーターセン及びノートボーン、2000年) Pietersen, A. and H. M. Noteborn. Apoptin. *Adv Exp Med Biol*, 465巻, 153-161頁, (2000年).

(クォングラ、2002年) Quong, MW, Romanow, WJ, and Murre, C. E protein function in lymphocyte development. *Annu Rev Immunol*, 20巻, 301-22頁, (2002年).

(ルエジンスキーら、1991年) Ruezinsky, D, Beckmann, H, and Kadesch, T. Modulation of the IgH enhancer's cell type specificity through a genetic switch. *Genes Dev*, 5巻, 29-37頁, (1991年).

(サルツ及びジョスト、1993年) Saluz, H. P. and Jost, J. P. Approaches to characterize protein-DNA interactions in vivo. *Crit Rev Eukaryot Gene Expr*, 3巻, 1-29頁, (1993年).

(サムブルックら、1989年) Sambrook, J., Frisch, E. F. and Maniatis, T. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Plainview NY, (1989年).

(サンガーら、1977年) Sanger, F., Nicken, S. and Coulson, A. R. DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci U S A*, 74巻, 5463-5467頁, (1977年).

(シミズら、1967年) Simizu, B., Rhim, J. S. and Wiebenga, N. H. Characterization of the Tacaribe group of arboviruses. I. Propagation and plaque assay of Tacaribe virus in a line of African green monkey kidney cells (Vero). *Proc Soc Exp Biol Med*, 125巻, 119-123頁, (1967年).

(スミスら、2000年) Smith, R. L., Traul, D. L., Schaack, J., Clayton, G. H., Staley, K. J. and Wilcox, C. L. Characterization of promoter function and cell-type-specific expression from viral vectors in the nervous system. *J Virol*, 74巻, 11254-11261頁, (2000年).

(スタムら、1998年) Stam, M, Viterbo, A, Mol, JN, and Kooter, JM. Position-dependent methylation and transcriptional silencing of transgenes in inverted T-DNA repeats: implications for posttranscriptional silencing of homologous hostgenes in plants. *Mol Cell Biol*, 18巻, 6165-77頁, (1998年).

(ステュルら、2000年) Stull, R. A., Hyun, W. C. and Pallavicini, M. G. Simultaneous flow cytometric analyses of enhanced green and yellow fluorescent proteins and cell surface antigens in doubly transduced immature hematopoietic cell populations. *Cytometry*, 40巻, 126-134頁, (2000年).

(タカダら、2000年) Takada, T., Iida, K., Akasaka, K., Yasue, H., Torii, R., Tsujimoto, G., Taira, M. and Kimura, H. Evaluation of heterologous insulator function with regard to chromosomal position effect in the mouse blastocyst and fetus. *Mol Reprod Dev*, 57巻, 232-237頁, (2000年).

(タナカラ、1996年) Tanaka, S., Livingstone-Zatchej, M. and Thoma, F. Chromatin

structure of the yeast URA3 gene at high resolution provides insight into structure and positioning of nucleosomes in the chromosomal context. *J Mol Biol*, 257巻, 919-934頁, (1996年).

(トーマス、1998年) Thomas, J. O. Isolation and fractionation of chromatin and linker histones. In Gould, H. (ed.) Chromatin: A Practical Approach. Oxford University Press, Oxford, 1-34頁, (1998年).

(ティラビーラ、1998年) Tiraby, M., Cazaux, C., Baron, M., Drocourt, D., Reynes, J. P. and Tiraby, G. Concomitant expression of *E. coli* cytosine deaminase and uracil phosphoribosyltransferase improves the cytotoxicity of 5-fluorocytosine. *FEMS Microbiol Lett*, 167巻, 41-49頁, (1998年).

(ウマナラ、1999年) Umana, P., Jean-Mairet, J., and Bailey, J.E. Tetracycline-regulated overexpression of glycosyltransferases in Chinese hamster ovary cells. *Biotechnol Bioeng*, 65巻, 542-9頁, (1999年).

(ファン デル フラッグら、2000年) van der Vlag, J., den Blaauwen, J. L., Sewalit, R. G., van Driel, R. and Otte, A. P. Transcriptional repression mediated by polycomb group proteins and otherchromatin-associated repressors is selectively blocked by insulators. *J Biol Chem*, 275巻, 697-704頁, (2000年).

(ファン ヘルデンら、1998年) van Helden, J., Andre, B., and Collado-Vides, J. Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. *J Mol Biol*, 281巻, 827-42頁, (1998年).

(ファン ヘルデンら、2000年) van Helden, J., Andre, B., and Collado-Vides, J. A web site for the computational analysis of yeast regulatory sequences. *Yeast*, 16巻, 177-87頁, (2000年).

(ファン ヘルデン、2000年) van Helden, J., Rios, AF., and Collado-Vides, J. Discovering regulatory elements in non-coding sequences by analysis of spaced dyads. *Nucleic Acids Res*, 28巻, 1808-18頁, (2000年).

(ヴァンス及びヴァウチェルト、2001年) Vance, V. and Vaucheret, H. RNA silencing in plants--defense and counterdefense. *Science*, 292巻, 2277-80頁, (2001年).

(ウォールラスら、1998年) Wallrath, L. L., Swede, M. J. and Elgin, S. C. R. Mapping chromatin structure in *Drosophila*. In Gould, H. (ed.) Chromatin: A Practical Approach. Oxford University Press, Oxford, 59-77頁, (1998年).

(ウィバー及びカダン、2000年) Weaver, L. S. and Kadan, M. J. Evaluation of adenoviral vectors by flow cytometry. *Methods*, 21巻, 297-312頁, (2000年).

(ウェイ及びフーバー、1996年) Wei, K. and Huber, B. E. Cytosine deaminase gene as a positive selection marker. *J BiolChem*, 271巻, 3812-3816頁, (1996年).

(ウィグラーら、1978年) Wigler, M., Pellicer, A., Silverstein, S. and Axel, R. (Biochemical transfer of single-copy eukaryotic genes using total cellular DNA as donor. *Cell*, 14巻, 725-731頁, (1978年).

(ウィグレイら、1994年) Wigley, P., Becker, C. , Beltrame, J., Blake, T. , Crocker, L., Harrison, S. , Lyons, I. , McKenzie, Z., Tearle, R. , Crawford, R. and et al. Site-specific transgene insertion:an approach. Reprod Fertil Dev, 6巻, 585-588頁, (1994年).

(クスウら、1995年) Xu, Z. Z., Krougliak, V. , Prevec, L. , Graham, F. L. and Both, G. W. Investigation of promoter function in human and animal cells infected with human recombinant adenovirus expressing rotavirus antigen VP7sc. J Gen Virol, 76巻, 1971-1980頁, (1995年).

(ヤングら、1997年) Yang, TT, Sinai, P, Kitts, PA, and Kain, SR. Quantification of gene expression with a secreted alkaline phosphatase reporter system. Biotechniques, 23巻, 1110-4頁, (1997年).

(インら、1996年) Yin, D. X., Zhu, L. , and Schimke R. T. Tetracyclin-controlled gene expression system achieves high-level and quantitative control of gene expression. Analyt Biochem, 235巻, 195-201頁, (1996年).

(チンク及びパロ、1995年) Zink, D. , and Paro, R. Drosophila Polycomb-group regulated chromatin inhibits the accessibility of a trans-activator to its target. DNA Embo J, 14巻, 5660-71頁, (1995年).

【 0 1 9 0 】

【表1】

| 表1 STARエレメントは導入遺伝子発現を改善する | | | |
|---------------------------|-----------------|----------------|--------|
| プラスミド | 過剰発現 クローン, % | 過剰発現倍数 (範囲) | クローンの数 |
| エンプティ | 12 | 3-11 | 25 |
| SCS(ポジティブ コントロール) | 24 | 3-160 | 21 |
| STAR-6 | 62 | 2-200 | 26 |
| STAR-3 | 39 | 5-820 | 28 |
| STAR-8 | 63 | 7-315 | 19 |
| STAR-4 | 31 | 25-1500 | 18 |
| STAR-1 | 57 | 5-80 | 28 |

ルシフェラーゼレポーター遺伝子の発現は、STARエレメントなしの(「エンプティ(empty)」、ネガティブ対照)又はSTARエレメントを含む(ポジティブ対照エレメント、ショウジョウバエからのSCSを含む、)、組み込まれたpSDHプラスミドを含む細胞株において測定される。ネガティブ対照の平均発現レベルは、参照レベルとして定義され、且つクローンの発現レベルが参照レベルの2倍よりも大きい場合、過剰発現であると考えられる

。各プラスミドの過剰発現するクローンのパーセント、及び過剰発現倍数が、各プラスミドのために解析されたクローンの数とともに報告される。

【0191】

【表2】

表2 クローニングされたSTARエレメント

| クローン | 染色体の位置 ¹ | 近傍の遺伝子 ² | 反復配列 |
|---------|--------------------------|--|--|
| STAR-1 | N.d. | | |
| STAR-2 | N.d. | | |
| STAR-3 | 正の 5q33.3 逆の 10q22.2 | ヒストン中のChr部分 アセチルトランスフェラーゼ 遺伝子 | |
| STAR-4 | 正の 1p31.1 逆の 14q24.1 | 10kb内に遺伝子がない G-タンパク質シグナル の調節のイントロン | 83% 反復 LINE2 及び LTR ERV_ClassI |
| STAR-5 | 正の 3q13.1 逆の 10q22.1* | | |
| STAR-6 | 2p21 | L5 kb 未知の推定のキナーゼ タンパク質に関連付けられた R 20 kb 微小管 | 19% SINE (MIR) 29% LINE |
| STAR-7 | 1q32.2 | | 12% Alu 4% MIR (SINE) LINE1 2.5% L31CR1 11.5% MER1 7% 低複雑性 2% |
| STAR-8 | 9q32 | 亜鉛フィンガータンパク質 を含む ZFP KRAB ボックス | 35% ERV_ClassI (LTR) 2% 単純反復 |
| STAR-9 | STAR4 参照 | | |
| STAR-10 | N.d. | | |
| STAR-11 | 2p25.1 | R 15 kb 未知のDNA結合性 タンパク質阻害剤 (Myc型) | 12% Alu (SINE) 26% MaLRs (LINE) |
| STAR-12 | 5q35.3 | R 15 kb 未知のADAM TS2 ファミリーメタロ プロテイナーゼ | 3% 低複雑性 |
| STAR-13 | STAR4 及び 9 参照 | | |
| STAR-14 | 正の N.d. 逆の 20q13.33 | | |
| STAR-15 | 1p36.36 | L 6 kb 電位差門のついた K チャンネルサブユニット R 4 kb 未知 | 14% LTR (MaLRs) |
| STAR-16 | 正の 8p23.1 逆の 8p22 など | | シーケンスされた 部分上の反復なし |
| STAR-17 | 2q31.1 | L 6 kb BTEB1 転写ファクター R 40 kb HNRNP | 10% 単純及び 低複雑性 |

¹ 染色体中の位置は、ヒトゲノムデータベースに対するSTARクローンからのDNA配列データのBLAST調査によって決定される。該位置は、各染色体の細胞遺伝学的イデオグラムと言われる標準命名法に従い与えられる；例えば1p2.3は1番染色体の短いアーム(arm)の第2の細胞遺伝学的バンドの第3の細胞遺伝学的サブバンドである(<http://www.ncbi.nlm.nih.gov>)

[h.gov/Class/MLACourse/Genetics/chrombanding.html](http://www.ncbi.nlm.nih.gov/Class/MLACourse/Genetics/chrombanding.html))。F, 正の(Forward)配列反応結果; R, 逆の(Reverse)配列反応結果; N.d., 決められていない(not yet determined)。

²ヒトゲノムマップビュービルド(Human Genome Map View Build)22に基づく(http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/hum_srch?chr=hum_chr.inf&query、2001年4月)。

L, 左; R, 右。

*位置があいまい、幾つかのヒット

【0192】

【表3】

| SINC | 長さ(nt) | 染色体の位置 ¹ | 備考 |
|-----------|--------|---------------------|--|
| psinks 9 | 700 | 22q11.21 | LTRを含む; 最も近い遺伝子ZNF 74, RNA結合性タンパク質。LTR非常に反復的。 |
| psinks 12 | 750 | 22q12.3 | ガン形成に関係しているアセチルグルコサミニルトランスフェラーゼ様タンパク質(664 kb)のイントロン内に位置している。 |
| psinks 19 | 600 | 22q13.1 | カルシウムチャネルのイントロン内に位置している、脳においてほとんど独占的に発現している。 |
| psinks 28 | 950 | 22q13.31 | 未知の機能の腎臓のイントロン内に位置している、SINEエレメントを含む |
| psinks 30 | 700 | 22q13.33 | SINEの一部を含む |
| psinks 35 | 650 | 22q11.21 | 溶質キャリヤーのためのエクソンをカバーする。(ミトコンドリアのための核遺伝子) |

¹染色体中の位置は、ヒトゲノムデータベースに対するSTARクローンからのDNA配列データのBLAST調査によって決定される。該位置は、各染色体の細胞遺伝学的イデオグラムと言われる標準命名法に従い与えられる; 例えば1p2.3は1番染色体の短いアーム(arm)の第2の細胞遺伝学的バンドの第3の細胞遺伝学的サブバンドである(<http://www.ncbi.nlm.nih.gov/Class/MLACourse/Genetics/chrombanding.html>)。

【0193】

【表4A】

表4A: 1つのストランド(正の; forward)又は逆のストランド
(逆の; reverse)における種々のSTARエレメントの配列

正のSTAR3

ACGTNCTAAGNAAACCATTATTATCATGACATTAAACCTATAAAAATAGGC
GTATCACGAGGCCCTTCGTCTTCACTCGAGCGGCCAGCTGGATCTGA
GTACTGAAATAGGAGTAAATCTGAAGAGCAAATAAGATGAGCCAGAAAAC
CATGAAAAGAACAGGGACTACCAGTTGATTCCACAAGGACATTCCAAGG
TGAGAAGGCCATACCTCCACTACCTGAACCAATTCTGTATGCAGATT
TAGCAAGGTTATAAGGTAGCAAAGATTAGACCCAAGAAAATAGAGAACT
TCCAATCCAGTAAAAATCATAGCAAATTATTGATGATAACAATTGTCTCC
AAAGGAACCAGGCAGAGTCGTGCTAGCAGAGGAAGCACGTGAGCTGAAA
ACAGCCAAATCTGCTTGTTCATGACACAGGAGCATAAGTACACACCA
CCAAC TGACCTATTAAGGCTGTGGTAAACCGATTGATAGAGAGAGGTTCT
AAATACATGGTCCCTCATAGGCAAACCGCAGTTCACTCCGAACGTAGTC
CCTGGAAATTGATGTCCAGNATAGAAAAGCANAGCAGNCNNNNNTAT
ANATNNNGNTGANCCANATGNTNNCTGNNC

逆のSTAR3

GAGCTAGCGGCGGCCAAGCTGGATCCCGCCCCGCCCTCCGCCCTCG
AGCCCCGCCCTGCCCTAGAGGCCCTGCCGAGGGGCGGGCCTGTCCC
TCCTCCCCCTTCCCCGCCCTACCGTCACGCTCAGGGCAGCCTGACC
CCGAGCGGCCCGCGGTGACCTCGCGCAGAGGCCTGTGGAGGGCGT
CGCAAGCCCTGAATCCCCCCCCGTCTGTTCCCCCTCCGCCAGTCTC
CTCCCCCTGGAACCGCGGGGTGGGTGACAGACCTGGCTGCGGCCAC
CGCCACCGCGCCTGCCGGGGCGCTGCCGCTGCCAGAAGAACTGCGGCT
GCCGCCTGGAGGAGGTGCCGCTGCCACCGCTGCCGCCGCC
AGGGGTAGGAGCTAAGCCGCCATTGTGTCCTCTGTTGTCGT
TGACATGAATCCGACATGACACTGATTACAGCCAATGGAGTCTCATTAA
ACCCGAGTCGCGGTCCGCCCGCGCTGCCATTGGAGGAGACCAAAG

ACACTTAAGGCCACCGTTGGCCTACGGGTCTGTCACCCACTCACT
AACCACTCTGCAGCCCATTGGGGCAGGTTCTGCCGGTCATNTCGCTTCC
AATAAACACACCCCTCGACCCCATTCCCCCTCGGGAACCACCC
CCGGGGAGGGTCCACTGGNCAATACCAATTNAANAGAACCGCTNGGG
TCCGCCTNTTNCGGCNCCTATTGGGTT

正の STAR4

GGGGAGGATTCTTGGCTGCTGAGTTGAGATTAGGTTGAGGGTAGTGAA
GGTAAAGGCAGTGAGACCACGTAGGGTCATTGCAGTAATCCAGGCTGG
AGATGATGGTGGTCAGTGGAAATAGCAGTCATGTGCTGTAACAACCTC
AGCTGGGAAGCAGTATATGTGGCGTTATGACCTCAGCTGGAACAGCAATG
CATGTGGTGGTGTAAATGACCCCCAGCTGGTAGGGTGCATGTGATGGAACA
ACCTCAGCTGGTAGCAGTGTACTTGATAAAATGTTGGCATACTCTACATT
TGTTATGAGGGTAGTGCCATTAAATTCTCCACAAATTGGTTGTCACGTAT
GAGTGAAAAGAGGAAGTGATGGAAGACTTCAGTGCTTTGGCCTGAATAA
ATAGAAGACGTCACTTCAGTAATGGAGACAGGGAAAGACTAANGNAGGGT
GGATTCACTAGAGCAGGTGTTCACTTTGAATATGATGAACTCTGAGAGA
GGAAAAACTTTCTACCTCTTAGTTTGNGNCTGGACTTAANATTAAAG
GACATANGACNGAGANCAGACCAAATNTCGANGTTTATATTACTT
GCNGAGGGATTNCAAGAAAAAGAAGACCCAANANCCATTGGTCAAAA
CTATNTGCCTTTAANAAAAAGANAATTACAATGGGANANANAAGTGTGN
CTNGGCAAAATTGGG

逆の STAR4

GGATTNGAGCTAGCGGCGCGCCAAGCTTGGATCTTAGAAGGGACAGAGTG
GGGCATGGAAATGCACCAACCAGGGCAGTGCAGCTGGTCACTGCCAGCTC
CNCTCATGGCAGAGGGCTGGCCTTGCAGCCGACCAGGCACTGAGCG
CCATCCCAGGGCCCTGCCAGCCCTCAGCAGGGCCAGGACACACAAGCCT
TTGACTTCCTCCTGTCAGTGCTGCCATTCTGTTGTGGTCATCACT
CCTTCCCTGTCCTCAGACTGCCAGCACTCAAGGATGTCCTGTGGTGGCA

TCAGACCATATGCCCTGAANAGGAGTGAGTTGGTGTGCGCGCC
 CANAGAGCTGCTGCCCTGAAAGATGCAAGTGGGAATGATGATGNTCAC
 CATCNTCTGACACCAAGCCCTTGGATAGAGGCCAACAGTGAGGATGG
 GGCTGCAGTGCATTGCCAAGGCAACTCTGTNNNTGACTGCTACANGACANT
 CCCAGGACCTGNGAAGNNCTATANATNTGATGCNAGGCACCT

正のSTAR6

CCACCACAGACATCCCCTCTGGCCTCCTGAGTGGTTCTTCAGCACAGCTT
 CCAGAGCCAATTAAACGTTCACTCTATGCTATAGACAAAAAGGGTTTG
 ACTAAACTCTGTGTTTAGAGAGGGAGTTAAATGCTGTTAACTTTAGGG
 GTGGGCGAGAGGAATGACAAATAACAACCTGTCTGAATGTTTACATTTC
 TCCCCACTGCCTCAAGAAGGTCACAACGAGGTCACTGATAAGGAGT
 AAGACCTCCCAGCCGGACTGTCCCTCGGCCCCAGAGGACACTCCACAGA
 GATATGCTAACTGGACTTGGAGACTGGCTCACACTCCAGAGAAAAGCATG
 GAGCACGAGCGCACAGAGCANGGCCAAGGTCCAGGGACNGAATGTCT
 AGGAGGGAGATTGGGGTGAGGGTANTCTGATGCAATTACTGNGCAGCTC
 AACATTCAAGGGAGGGGAAGAAAGAACNGTCCCTGTAAGTAAGTTGTNC
 ANCAGAGATGGTAAGCTCAAATTNAACTTGGCTGCTGGAAAGTTNN
 GGGCCNANANAANAAACANAAANATTGAGGTTANACCCACTAACCN
 TATNANTANTTATTAAACCCCTAATTANACCTGGATANCCTAAAATAT
 CNTNTNAAACGGAACCCTCNTTCCNTTNAAATNNNAAGGCCATTN
 NGNNNCAGTAAAATCTNNNTAAGNNNTGGGCCNAACAAACNTNTTC
 CNAGACACNTTTNTCCNGGNATTNTAATTATTCTAANCC

逆のSTAR6

ATCGTGTCTTCCAGGGACATGGATGAAGCTGGAAGCCATCATCCTCAG
 CAAACTAACACAGGAACAGAAAACCAAATACCACATGTTCTCACTCATAAG
 TGGGAGCTGAACAGTGAGAACACATGGACACAGGGAGGGAACATCACA
 CACCAAGGCCTGTCTGGTGTGGGAGGGAGGGAGGCATCAGGACAAA

TAGCTAATGCATGTGGGGCTTAAACCTAGATGACGGGTTGATAGGTGCAG
CAATCCACTATGGACACATATACTATGTAACAACCCNACCTNTTGACAT
GTATCCCAGAACTTAAAGGAAAATAAAAATTAAAAAAATTNCCTGGAA
TAAAAAAAGAGTGTGGACTTGGTGAGATN

正の STAR8

GGATCACCTCGAAGAGAGTCTAACGTCCGTAGGAACGCTCTCGGGTTCAC
AAGGATTGACCGAACCCCAGGATACGTCGCTCTCCATCTGAGGCTTGNTC
CAAATGGCCCTCCACTATTCCAGGCACGTGGGTGTCTCCCTAACTCTCC
CTGCTCTCCTGAGCCCAGTGCCTATCACCCATCGGTGCAGGTCTTCT
GAANAGCTCGGGTGGATTCTCTCCATCCCACTTCCCTTCCAAGAAAGAA
GCCACCGTTCCAAGACACCCAATGGGACATTCCNTTCCACCTCCTTNTC
NAAAGTTNGCCCAGGTGTTCTAACAGGTTAGGGAGAGAANCCCCAGG
TTTNAGTTNCAAGGCATAGGACGCTGGCTGAACACACACACACNCTC

逆の STAR8

GGATCCCAGCTCTGCACCGCAAACCTCTACGGGCCCTGCAGGACGGCGGC
CTCCTGCCGCTTGGACGCCAGNCAGGAGCTCCCCGGCAGCAGCAGAGCA
GAAAGAAGGATGGCCCCGCCCCACTTCGCCTCCGGCGGTCTCCCTCCCG
CCGGCTACGGACATAGATGGCTGCCTAGCTCCGGAAGCCTAGCTCTTGT
TCCGGGCATCCTAACCGAAGACACGGTTTCTCCCTCCGGGCCTCACCACA
TCTGGGACTTGACGACTCGGACCTCTCCATTGAATGGTTGCGCGTTC
TCTGGGAAAG

正の STAR18

TGGATCCTGCCGCTCGCGTCTTAGTGTTCCTCCCTAACAGACTTCCCTTG
TTTGTTGTCTTGTGCAGTATTTACAGCCCTCTGTGTTTCTTATT
CTCGTACACACACCGCAGTTAACGGGTGATGTGTGTATAATTAAAAGGAC
CCTTGGCCCATACTTCTAACCTTAAAGGGACTGGGATTGGGTTGACTG
AAATATGTTTGGTGGGGATGGGACGGTGGACTTCCATTCTCCCTAAACT

GGAGTTTGGTCGGAATCAAAACTAAAAGAAACCTCTGGGAGACTGGAA
ACCTGATTGGAGCACTGAGGAACAAGGAAATGAAAAGGCAGACTCTCTGA
ACGTTGATGAAATGGACTCTGTGAAAATTAACAGTGAATATTCACTGTT
GCACTGTACGAAGTCTCTGAAATGTAATTAAAAGTTTATTGAGCCCCCG
AGCTTGCGCTTGC CGT ATT TT CCG GT CG CG AC AT CCC ACC G CG C AGA
GCCTCGCCTCCCCGCTGNCCCTCAGCTCCGATGACTCCCCGCCCGCCC
TGCTCGGTGACAGACGTTCTACTGCTCCAATCGGAGGCACCCTCGCGG

逆の STAR18

TGGATCCTGCCGCTCGCGTCTAGTGTTCCTCCCTCAAGACTTTCTTCTG
TTTGTTGTCTTGTGCAGTATTTACAGCCCTCTGTGTTTCTTATT
CTCGTACACACACCGCAGTTAAGGGTATGTGTGTATAATTAAAAGGAC
CCTTGGCCCATACTTCTAATTCTTAGGGACTGGGATTGGGTTGACTG
AAATATGTTTGGGGATGGGACGGTGGACTCCATTCTCCCTAAACT
GGAGTTTGGTCGGAATCAAAACTAAAAGAAACCTCTGGGAGACTGGAA
ACCTGATTGGAGCACTGAGGAACAAGGAAATGAAAAGGCAGACTCTCTGA
ACGTTGATGAAATGGACTCTGTGAAAATTAACAGTGAATATTCACTGTT
GCACTGTACGAAGTCTCTGAAATGTAATTAAAAGTTTATTGAGCCCCCG
AGCTTGCGC

--
【 0 1 9 4 】

【表4B】

表4B: 種々のSINCエレメントの配列

PSINKS 9

GATCAGGA TAATAAGTAC GCTGGGAAGA CAACAAAATG
 ATTAAATCT TAGACAAGTC ATTCTAGGTG TCTCCACTGT
 TTCAGTTCTT GCATTCAATT TTGTGGTATC TTTTCCCTTT
 TACCAATAAA AAAGCTCCCT GACATCACAT TGTGGCAGTC
 CCCATGGTTT GCCGCAGTTA CTGCGGGACT GAACGAAGGA
 GGACGAATGA AGAAATGAAA ACCAAGGAAA AAAGGAGCTG
 TTTAAAGAAG GGTCCAGGGA AGAAGAAGAG GGCTCCCAGC
 TTCTAGTGAG CAAGGGCAGC AGCCCTGAGC TTCTACAGCC
 CTTCATATT ATTGAGTAGA AAGAGCAGGG AGCAGGGAGGT
 AATGATTGGT CAGCTCTCA ATTGATCACA GGTCACATT
 ATTGCTAAC A GATTACAT GTGCCTAAC TCAAGAACG
 CCGCGCCTGG GGCATGACTG CCCTCAGCAT TCCCTCTGGG
 TGGCAGACGC AGTTGCCAA CATTCTGCAT TCATGAGAAC
 AGTTTACTGT TTACTCATAT AACCTCCAGT GGTACACCGA
 GTTGATC

PSINKS 12

GATCTAA TTTCTCTGTA TTTAATTCCC ATGTCTATT
 TGTCTATTTT CAAGATTGAT TTACATTGCA GGTTCCGATG
 CAACCACTGA CTTACATTGC AGGTTCTAAT GTAACCCTG
 TCCTTAACGA GTACATAGAT TTGTTCCCTT CTCTCCAGGA
 GCATGAGATT TGTTGCCTCC AGGAAAGGCA ACAAACTAC
 TATTCCCTTA AGGACAGTGG TTCTCAAAGG ATTGTCCCTGG
 GAACAGCAGC ATCACCTACA CAGTAGTTAG AAATGGCACAT
 TCTGAGGCCT CCCAAGACCT GCTAACTCAG ACACCTGGGG

AGAAGAAGGG GTTCCAACAA GCCTTCTAGG TCATTCTGAT
GCATGCTGGA GTTGAGAAT CGATGCTCTA GGAAAAACAC
CAGTACTAT CTACCATCAA CTTGACCACT CAAGTGTAC
CATTCACTGA AGTTTAACTA CAATGTCCAG AGAATTAATT
GTGTACCAGG CACTATGCCG AAGGCTGAAT GCTGCCTCAC
AATCCANAGT GGTATGTGTG TAAATGACTA AATAAAATGC
AAAATGGGAT GACATG

PSINKS 19

G ATCCTCCATC TGCTCCACCC ACTTCCATGT AAGTGATCCT
GGGCTGATCA CTTCCTCTCT CTAGACTTCG TTTCTTTTTT
TTCTTTTTA GACCGAGTCT CACTCTGTCA CCCAGGCTGG
AGTGCAGTGG TGAGAGATCTT GCTCACTGCA ACCTCCACCT
CCTGGGTTCA AGCAATTCTC CTGTCTCAGC CTCCTGAGTA
GATAGGACTA TAGGTGCACA CCACCATAACC TAGCTAATT
TTGTTTTTT AGTAGAGATG GGGTTTCACC ATATTGGTCA
GCCTGTTCTC AAACCTCCTGA CCTCAGGTGA TCCACCCACC
TCAGCCTCCC AAAGTGCTGG GATTACAGGT GTGAGCCACC
GAGCCGGGCT GCCCTTCTCT GGACTTTGAT TTCCTCATCT
ATAAAACAGA CAACAATCCC TACTATGACC ATCCAGAAGG
GTTAATCTAT GCTTCATTGC AATCCTAATC AAAAATCCCA
ACATTTGGC CGTGGAGCCT GCCCAGATGG TTCTAGGATT
TATTTGGATG GGAAAATAGT CAAGACAAGC TT

PSINKS 28

GAT CATGGAGGGA GAGAACAAACC AACCAACACAC
TGACTGGTCA CCCCTGAAGT TCACAGCCAC TACCCTGTAG
AGGCCCGAG GTGCCGGCA AGCCCAGTAT ACTTCCATCT

AAACTCCCCT TGCACCTGCT CCTCCTGTT CAGACAATGA
GCTGTAACAC GCACATCCAC ACCACACATC ACCCACAGCA
GGGGCAGGAG GCAGCTAAGC ATGGGCTTCA GAGTCCTCCC
ACCAGCAGCG CCTACCAGCT ACAAGCCTGA CGTCTCTGTG
TGTGTGTGTA AATTCACTA AATATTCTT CCTTGTTTT
TTAAAAATT ACATGAAATG CACATTTTG CTGTGACAGA
AGCATGTAAC TGTGATCCTA ACACACCTAC TCCTCCGCCT
TTTACTGCCG TCTGCTTCCC TCTCTCTCC ACGCCCACTC
GAUTGCAGTA TCGATGCCAA CAACATGATG TGTGTCCTTC
CATGTTCCC TGCTCATGCA TTCGCATGTA AGCCACCGCA
CATGTCACTG TATGTACACA CAGGGGATTG TGAGGCCAAT
GTTTACAAG GATTACGTTA TACACCCTT TCTGCAGTGA
GTTTTCCA GGCAACCTCC CAGGCCCAT GGTGTAGCTC
TGGGTCAATC CTTTTTTTTT TTTTGGAGA CAGAGTCTCA
CTCTGTCGCC CAGGCTGGAG TGCAGTGGTG CAATTGGGC
TCACTGCAAC CTCCGCCTCC CGGGTTCAAG CGATTCTCCT
GCCTCAGCCT CCTGAGTAGC TGACATTACA AGCGCGCACT
ACCACACCCG GCTAATTTT GTATTTTAG TAGAGATA
GTTTCACCAT GTGGTCAGG CTGGTCTTGA ACTCCTGAGC
TCGTGATC

PSINKS 30

GATCC ACCCGCCTCG GACTCCAAA GTGCTGGGAT
TACAGGTGTG AGCCACTGTG CTTGGCCCGA ATCAGGAATA
ATTCTGATGG CTAAGGAAGA CAGCTTCCGA GAGAGTAGGA
GAAAGGGCAC AGGATTCCAG GCAGAAGGCC CATCTAGGGC
AAAGGCGAAG GTGTGGCTCA GCCTGCCTCC TTTGGGAAAT
GGCGAGTGTG TTCTGGCTC AGGGTTCTTGT GTAAGGGACA
GAGAAGACTC GGGAAAGATC AGTTGAGCTG GAATGTGCAG

GCTCTTGAGT ACCCTGCTCA GGAGCTGGAG GTGGGCTACC
CTGCAAACTC CAGGCCATGA AGCCCAGGAA GATGTCAGGC
TGGTCTTCCC ATGCCCTTGT GTATCTGAGA CCAACTGTCA
CTAAATGTTT CCTTTACGCC CTGGACACAC AGCTAGACTC
TACTTCTCAG ATTCTCTTGA AATACAAGTC TTTAGCCAGA
GGGTGTGGAG GGAAATGCTG TGTATCACTT TGAGGTTGAG
GCCATCAAAG CCTCCCACAG GTGGCCCCCT CTTTCTCTCC
CCACGTACTT ATGATGTTGA TGCCAAGGC AGCTTGAGTA
CTACCTGCTG AAGGCAGGGC CTCTGTCACC ATAGATC

PSINKS 35

GATCCAC CTGCCTCGGC CTCCCAAAGT GCTGGGATTA
CAGGCATGAG CCACCATGCC TGGCCAAAAA CTTCTACCTG
CTTGGAAAGT TGACTGGTCA CACAGCCTAG CAAATGAGGT
TGGGATGTGG GATGTGCCTG GTTCCAATCC CAGCCCTTA
CTGTTCCCCT AGGAGGTGGG GACAGGCCTC ACCCAGGCGT
CCAGCATCCT GCAGCTGAAT CTTGAGCATT TCCATGGGAC
AGGTCAACCAC GACCTGGCAC ATCCCAGCCC CACACCCGGC
AAGCATCTCC ATCTTCAGGT TCCGCTGCAT CCTATGGGAA
CAGGCGTCAG GCTCCTTCAG CCGCAGGCCA CAGGCCTGCC
CTGGTGCAGC TGCCCTTTG TGAGAGGGGG ACTTTCCCTG
GATGGCACCC GTGGCTGCCA CTCACCCAGC TGGTCAAGTC
ATCAGCTAGC CCTTAGGTGT GGTCTCTGTA CGGACAGGGG
ACTAAAGTTA AAACAAAGCC TGCTAGGGAG GTAGCACCGC
ATGGAAGCTG AAACAGTGAC AGAGAAAAGT ACCCAGACCA
GGCGTTGTCC TTGATC

【 0 1 9 5 】

【表5】

表5. ポリメラーゼ連鎖反応 (PCR) 又はDNA突然変異のために
使用されるオリゴヌクレオチド

| 番号 | 配列 |
|-----|---|
| C65 | AACAAAGCTTGATATCAGATCTGCTAGCTTGGTCGAGCTGATAC TTCCC |
| C66 | AAACTCGAGCGGCCGCGAATTCGTCGACTTACCACTCCCTAT CAGTGATAGAG |
| C67 | AAACCGCGGCATGGAAGACGCCAAAACATAAAGAAAGG |
| C68 | TATGGATCCTAGAATTACACGGCGATCTTCC |
| C81 | AAACCATGGCCGAGTACAAGCCCACGGTGC GCC |
| C82 | AAATCTAGATCAGGCACCGGGCTGCGGGTCATGC |
| C85 | CATTTCCCCGAAAGTGCCACC |
| D30 | TCACTGCTAGCGAGTGGTAAACTC |
| D41 | GAAGTCGACGAGGCAGGCAGAAGTATGC |
| D42 | GAGCCGCGGTTAGTCCTCACCTTGT CG |
| D51 | TCTGGAAGCTTGCTGAAGAAC |
| D89 | GGGCAAGATGTCGTAGTCAGG |
| D90 | AGGCCCATGGTCACCTCCATCGCTACTGTG |
| D91 | CTAATCACTCACTGTGTAAT |
| D93 | AATTACAGGCGCGCC |
| D94 | AATTGGCGCGCCTGT |
| D95 | TGCTTGCATACTCTGCCTGCCTC |
| E12 | TAGGGGGATCCAAATGTT |
| E13 | CCTAAAAGAAGATCTTAGC |
| E14 | AAGTGTGGATCCACTTGG |
| E15 | TTTGAAGATCTACCAAATGG |
| E16 | GTTCGGGATCCACCTGGCCG |
| E17 | TAGGCAAGATCTGGCCCTC |
| E18 | CCTCTCTAGGGATCCGACCC |
| E19 | CTAGAGAGATCTCCAGTAT |
| E20 | AGAGTTCCGGATCCGCCTGG |
| E21 | CCAGGCAGACTCGGAAC TCT |
| E22 | TGGTGAAACCGGATCCCTAC |
| E23 | AGGTCAGGAGATCTAGACCA |
| E25 | CCATTTCGCTTCCTTAGCTCC |

| | |
|-----|------------------------------------|
| E42 | CGATGTAACCCACTCGTGCACC |
| E57 | AGAGATCTAGGATAATTTCG |
| E92 | AGGCCTAGCACCGTTCTACTCTTTCTACTCTG |
| E93 | GATCAAGCTTACCGTCTAAAGGCATTATATAG |
| E94 | AGGCCTAGCACCGTTAGAGTTAGTGATCCAGG |
| E95 | GATCAAGCTTACCGTCAGTAAAGGTTCTGTATGG |
| E96 | AGGCCTAGCACCGTTCTACTCTTTCTACTCTG |
| E97 | CGAGGAAGCTGGAGAAGGAGAAGCTG |
| E98 | CAAGGGCCGCAGCTTACACATGTTC |
| D58 | CCAAGTTGACCAGTGCC |
| D80 | GTTCGTGGACACGACCTCCG |
| D70 | TACAAGCCAACCACGGCCT |
| D71 | CGGAAGTGCTTGACATTGGG |

【 0 1 9 6 】

【表6】

表 6. ゲノム位置及び長さを含む、本発明の STAR エレメント

| STAR | 位置 ¹ | 長さ ² |
|-----------------|---------------------|-----------------|
| 1 | 2q31.1 | 750 |
| 2 | 7p15.2 | 916 |
| 3 ³ | 15q11.2 及び 10q22.2 | 2132 |
| 4 | 1p31.1 及び 14q24.1 | 1625 |
| 5 ⁴ | 20q13.32 | 1571 |
| 6 | 2p21 | 1173 |
| 7 | 1q34 | 2101 |
| 8 | 9q32 | 1839 |
| 9 ⁴ | 10p15.3 | 1936 |
| 10 | Xp11.3 | 1167 |
| 11 | 2p25.1 | 1377 |
| 12 | 5q35.3 | 1051 |
| 13 ⁴ | 9q34.3 | 1291 |
| 14 ⁴ | 22q11.22 | 732 |
| 15 | 1p36.31 | 1881 |
| 16 | 1p21.2 | 1282 |
| 17 | 2q31.1 | 793 |
| 18 | 2q31.3 | 497 |
| 19 | 6p22.1 | 1840 |
| 20 | 8p13.3 | 780 |
| 21 | 6q24.2 | 620 |
| 22 | 2q12.2 | 1380 |
| 23 | 6p22.1 | 1246 |
| 24 | 1q21.2 | 948 |
| 25 ⁵ | 1q21.3 | 1067 |
| 26 | 1q21.1 | 540 |
| 27 | 1q23.1 | 1520 |
| 28 | 22q11.23 | 961 |
| 29 | 2q13.31 | 2253 |
| 30 | 22q12.3 | 1851 |
| 31 | 9q34.11 及び 22q11.21 | 1165 |
| 32 | 21q22.2 | 771 |

| STAR | 位置 ¹ | 長さ ² |
|------|--------------------------------|-----------------|
| 33 | 21q22.2 | 1368 |
| 34 | 9q34.14 | 755 |
| 35 | 7q22.3 | 1211 |
| 36 | 21q22.2 | 1712 |
| 37 | 22q11.23 | 1331 |
| 38 | 22q11.1 及び 22q11.1 | ~1000 |
| 39 | 22q12.3 | 2331 |
| 40 | 22q11.21 | 1071 |
| 41 | 22q11.21 | 1144 |
| 42 | 22q11.1 | 735 |
| 43 | 14q24.3 | 1231 |
| 44 | 22q11.1 | 1591 |
| 45 | 22q11.21 | 1991 |
| 46 | 22q11.23 | 1871 |
| 47 | 22q11.21 | 1082 |
| 48 | 22q11.22 | 1242 |
| 49 | Chr 12 ランダムクローン, 及び 3q26.32 | 1015 |
| 50 | 6p21.31 | 2361 |
| 51 | 5q21.3 | 2289 |
| 52 | 7p15.2 | 1200 |
| 53 | Xp11.3 | 1431 |
| 54 | 4q21.1 | 981 |
| 55 | 15q13.1 | 501 |
| 56 | 3p25.3 を含む | 741 |
| 57 | 4q35.2 | 1371 |
| 58 | 21q11.2 | 1401 |
| 59 | 17 ランダムクローン | 872 |
| 60 | 4p16.1 及び 6q27 | 2068 |
| 61 | 7p14.3 及び 11q25 | 1482 |
| 62 | 14q24.3 | 1011 |
| 63 | 22q13.3 | 1421 |
| 64 | 17q11.2 | 1414 |
| 65 | 7q21.11=28.4 | 1310 |
| 66 | 20q13.33 及び 6q14.1 | ~2800 |

¹ 染色体中の位置は、ヒトゲノムデータベースに対するSTARエレメントからのDNA配列データのBLAST調査によって決定される。該位置は、各染色体の細胞遺伝学的イデオグラムと言及される標準命名法に従い与えられる。例えば1p2.3は1番染色体の短いアーム(arm)の第2の細胞遺伝学的バンドの第3の細胞遺伝学的サブバンドである(<http://www.ncbi.nlm.nih.gov/Class/MLACourse/Genetics/chrombanding.html>)。正の(forward)及び逆の

(reverse) シーケンシング反応が種々のゲノム座からのDNAを識別した場合、両方の座が示される。

²正確な長さは、DNA配列分析によって決定され、およその長さが制限マッピングによって決定される。

³STAR3の配列及び位置が表2及び表4の集合故に洗練されている。

⁴表2及び表4におけるこれら数を有するSTARは無視されており（以下、「oldSTAR5」などとして言及される）及びSTARエレメントに割り当てられたそれら数がDNA配列別表に示される。oldSTAR5、oldSTAR14及びoldSTAR16、クローニングされたDNAは、2よりも多い染色体位置からのキメラであった；oldSTAR9及びoldSTAR13、クローニングされたDNAは、STAR4に同一であった。

⁵表4「STAR18」に同一

【0197】

【表7】

表7. STARエレメントは、導入遺伝子発現¹に対して
長期間にわたり安定性をもたらす

| | 細胞分割 ² | ルシフェラーゼ発現 ³ |
|--|-------------------|------------------------|
| STAR6に ピューロマイシン をプラス | 42 | 18,000 |
| | 60 | 23,000 |
| | 84 | 20,000 |
| | 108 | 16,000 |
| STAR6に ピューロマイシン をなし ⁴ | 84 | 12,000 |
| | 108 | 15,000 |
| | 144 | 12,000 |

¹プラスミドpSDH-Tet-STAR6は、U-2 OS細胞内にトランスフェクションされ、且つクローニングは実施例1に記載されたようにドキシサイクリンのない培地内で分離されかつ培養された。細胞は、1:20の希釈率で、新鮮な培養容器に移された。

²細胞分割の数は、1週間で培養が細胞コンフルエンス（それは～6細胞分割を表わす）に達するとの推定に基いている。

³ルシフェラーゼは、実施例1に記載されているようにアッセイされた。

⁴60細胞分割後に、細胞は2つの培養容器に移された；一つは、第1の60細胞分割についてはピューロマイシンを含む培養培地を供給され、且つ第2は抗生素質を欠く培養培地を供給された。

【0198】

【表8】

表8. ヒトSTARエレメント及びそれらの推定のマウス
オルトログ (ortholog) 及びパラログ (paralog)

| 配列:ID | STAR | ヒト ¹ | マウス ² | 類似性 ³ |
|-------|------|-----------------|------------------|--|
| 1 | 1 | 2q31.1 | 2D | 600 bp 69% |
| 2 | 2 | 7p15.2 | 6B3 | 909 bp 89% |
| 3 | 3a | 5q33.3 | 11B2 | 248 bp 83% |
| 4 | 3b | 10q22.2 | 14B | 1. 363 bp 89% 2. 163 bp 86% |
| 5 | 6 | 2p21 | 17E4 | 437 bp 78% |
| 6 | 12 | 5q35.3 | 11b1.3 | 796 bp 66% |
| 7 | 13 | 9q34.3 | 2A3 | 753 bp 77% |
| 8 | 18 | 2q31.3 | 2E1 | 497 bp 72% |
| 9 | 36 | 21q22.2 | 16C4 | 166 bp 79% |
| 10 | 40 | 22q11.1 | 6F1 | 1. 270 bp 75% 2. 309 bp 70% |
| 11 | 50 | 6p21.31 | 17B1 | 1. 451 bp 72% 2. 188 bp 80% 3. 142 bp 64% |
| 12 | 52 | 7p15.2 | 6B3 | 1. 846 bp 74% 2. 195 bp 71% |
| 13 | 53 | Xp11.3 | XA2 | 364 bp 64% |
| 14 | 54 | 4q21.1 | 5E3 | 1. 174 bp 80% 2. 240 bp 73% 3. 141 bp 67% 4. 144 bp 68% |
| 15 | 61a | 7p14.3 | 6B3 | 188 bp 68% |

¹ヒトゲノムにおけるSTARエレメントの細胞遺伝位置。

²マウスゲノムにおけるSTARエレメント オルトログの細胞遺伝位置。

³高い配列類似性を示す領域の長さ、及び類似性のパーセント。ある場合では、一つより上の高い類似性のブロックが生じる；それらの場合、各ブロックは、別々に記載される。類似性 < 60%は、有意であると考えられない。

【0199】

【表9】

**表9. STARエレメント内で過剰に表れたオリゴヌクレオチドパターン
(6塩基対)**

該パターンは、有意性係数に従いランク付けされる。これらは、参照としてヒトゲノムの配列を有するRSAツールを使用して決定された。線形判別式分析におけるほとんどの判別式変数を含むパターンがアストリスクで示される。

| 番号 | オリゴ ヌクレオチド 配列 | 観察された 出現 | 期待された 出現 | 有意性係数 | 一致する STARの数 |
|----|---------------------|-------------|-------------|-------|----------------|
| 1 | CCCCAC | 107 | 49 | 8.76 | 51 |
| 2 | CAGCGG | 36 | 9 | 7.75 | 23 |
| 3 | GGCCCC | 74 | 31 | 7.21 | 34 |
| 4 | CAGCCC | 103 | 50 | 7.18 | 37 |
| 5 | GCCCCC | 70 | 29 | 6.97 | 34 |
| 6 | CGGGGC | 40 | 12 | 6.95 | 18 |
| 7 | CCCCGC | 43 | 13 | 6.79 | 22 |
| 8 | CGGCAG | 35 | 9 | 6.64 | 18 |
| 9 | AGCCCC | 83 | 38 | 6.54 | 40 |
| 10 | CCAGGG | 107 | 54 | 6.52 | 43 |
| 11 | GGACCC* | 58 | 23 | 6.04 | 35 |
| 12 | GCGGAC | 20 | 3 | 5.94 | 14 |
| 13 | CCAGCG | 34 | 10 | 5.9 | 24 |
| 14 | GCAGCC | 92 | 45 | 5.84 | 43 |
| 15 | CCGGCA | 28 | 7 | 5.61 | 16 |
| 16 | AGCGGC | 27 | 7 | 5.45 | 17 |
| 17 | CAGGGG | 86 | 43 | 5.09 | 43 |
| 18 | CCGCC | 43 | 15 | 5.02 | 18 |
| 19 | CCCCCG | 35 | 11 | 4.91 | 20 |
| 20 | GCCGCC | 34 | 10 | 4.88 | 18 |
| 21 | GCCGGC | 22 | 5 | 4.7 | 16 |
| 22 | CGGACC | 19 | 4 | 4.68 | 14 |
| 23 | CGCCCC | 35 | 11 | 4.64 | 19 |
| 24 | CGCCAG | 28 | 8 | 4.31 | 19 |
| 25 | CGCAGC | 29 | 8 | 4.29 | 20 |
| 26 | CAGCCG | 32 | 10 | 4 | 24 |
| 27 | CCCACG | 33 | 11 | 3.97 | 26 |
| 28 | GCTGCC | 78 | 40 | 3.9 | 43 |
| 29 | CCCTCC | 106 | 60 | 3.87 | 48 |
| 30 | CCCTGC* | 92 | 50 | 3.83 | 42 |
| 31 | CACCCC | 77 | 40 | 3.75 | 40 |
| 32 | GCGCCA | 30 | 10 | 3.58 | 23 |
| 33 | AGGGGC | 70 | 35 | 3.55 | 34 |

| | | | | | |
|----|--------|-----|----|------|----|
| 34 | GAGGGC | 66 | 32 | 3.5 | 40 |
| 35 | GCGAAC | 14 | 2 | 3.37 | 13 |
| 36 | CCGGCG | 17 | 4 | 3.33 | 12 |
| 37 | AGCCGG | 34 | 12 | 3.29 | 25 |
| 38 | GGAGCC | 67 | 34 | 3.27 | 40 |
| 39 | CCCCAG | 103 | 60 | 3.23 | 51 |
| 40 | CCGCTC | 24 | 7 | 3.19 | 19 |
| 41 | CCCCTC | 81 | 44 | 3.19 | 43 |
| 42 | CACCGC | 33 | 12 | 3.14 | 22 |
| 43 | CTGCC | 96 | 55 | 3.01 | 42 |
| 44 | GGGCCA | 68 | 35 | 2.99 | 39 |
| 45 | CGCTGC | 28 | 9 | 2.88 | 22 |
| 46 | CAGCGC | 25 | 8 | 2.77 | 19 |
| 47 | CGGCC | 28 | 10 | 2.73 | 19 |
| 48 | CCGCCG | 19 | 5 | 2.56 | 9 |
| 49 | CCCCGG | 30 | 11 | 2.41 | 17 |
| 50 | AGCCGC | 23 | 7 | 2.34 | 17 |
| 51 | GCACCC | 55 | 27 | 2.31 | 38 |
| 52 | AGGACC | 54 | 27 | 2.22 | 33 |
| 53 | AGGGCG | 24 | 8 | 2.2 | 18 |
| 54 | CAGGGC | 81 | 47 | 2.18 | 42 |
| 55 | CCCGCC | 45 | 21 | 2.15 | 20 |
| 56 | GCCAGC | 66 | 36 | 2.09 | 39 |
| 57 | AGCGCC | 21 | 6 | 2.09 | 18 |
| 58 | AGGCC | 64 | 34 | 2.08 | 32 |
| 59 | CCCACC | 101 | 62 | 2.05 | 54 |
| 60 | CGCTCA | 21 | 6 | 2.03 | 17 |
| 61 | AACGCG | 9 | 1 | 1.96 | 9 |
| 62 | GCGGCA | 21 | 7 | 1.92 | 14 |
| 63 | AGGTCC | 49 | 24 | 1.87 | 36 |
| 64 | CCGTCA | 19 | 6 | 1.78 | 14 |
| 65 | CAGAGG | 107 | 68 | 1.77 | 47 |
| 66 | CCCGAG | 33 | 14 | 1.77 | 22 |
| 67 | CCGAGG | 36 | 16 | 1.76 | 25 |
| 68 | CGCGGA | 11 | 2 | 1.75 | 8 |
| 69 | CCACCC | 87 | 53 | 1.71 | 45 |
| 70 | CCTCGC | 23 | 8 | 1.71 | 20 |
| 71 | CAAGCC | 59 | 32 | 1.69 | 40 |
| 72 | TCCGCA | 18 | 5 | 1.68 | 17 |
| 73 | CGCCGC | 18 | 5 | 1.67 | 9 |
| 74 | GGGAAC | 55 | 29 | 1.63 | 39 |
| 75 | CCAGAG | 93 | 58 | 1.57 | 49 |

| | | | | | |
|-----|----------|----|----|------|----|
| 76 | CGTTCC | 19 | 6 | 1.53 | 16 |
| 77 | CGAGGA | 23 | 8 | 1.5 | 19 |
| 78 | GGGACC | 48 | 24 | 1.48 | 31 |
| 79 | CCGCGA | 10 | 2 | 1.48 | 8 |
| 80 | CCTGCG | 24 | 9 | 1.45 | 17 |
| 81 | CTGCGC | 23 | 8 | 1.32 | 14 |
| 82 | GACCCC | 47 | 24 | 1.31 | 33 |
| 83 | GCTCCA | 66 | 38 | 1.25 | 39 |
| 84 | CGCCAC | 33 | 15 | 1.19 | 21 |
| 85 | GCGGGA | 23 | 9 | 1.17 | 18 |
| 86 | CTGCGA | 18 | 6 | 1.15 | 15 |
| 87 | CTGCTC | 80 | 49 | 1.14 | 50 |
| 88 | CAGACG | 23 | 9 | 1.13 | 19 |
| 89 | CGAGAG | 21 | 8 | 1.09 | 17 |
| 90 | CGGTGC | 18 | 6 | 1.06 | 16 |
| 91 | CTCCCC | 84 | 53 | 1.05 | 47 |
| 92 | GC GGCC | 22 | 8 | 1.04 | 14 |
| 93 | CGGCGC | 14 | 4 | 1.04 | 13 |
| 94 | AAGCCC * | 60 | 34 | 1.03 | 42 |
| 95 | CCGCAG | 24 | 9 | 1.03 | 17 |
| 96 | GCCCAC | 59 | 34 | 0.95 | 35 |
| 97 | CACCCA | 92 | 60 | 0.93 | 49 |
| 98 | GCGCCC | 27 | 11 | 0.93 | 18 |
| 99 | ACCGGC | 15 | 4 | 0.92 | 13 |
| 100 | CTCGCA | 16 | 5 | 0.89 | 14 |
| 101 | ACGCTC | 16 | 5 | 0.88 | 12 |
| 102 | CTGGAC | 58 | 33 | 0.88 | 32 |
| 103 | GCCCCA | 67 | 40 | 0.87 | 38 |
| 104 | ACCGTC | 15 | 4 | 0.86 | 11 |
| 105 | CCCTCG | 21 | 8 | 0.8 | 18 |
| 106 | AGCCCG | 22 | 8 | 0.79 | 14 |
| 107 | ACCCGA | 16 | 5 | 0.78 | 13 |
| 108 | AGCAGC | 79 | 50 | 0.75 | 41 |
| 109 | ACCGCG | 14 | 4 | 0.69 | 7 |
| 110 | CGAGGC | 29 | 13 | 0.69 | 24 |
| 111 | AGCTGC | 70 | 43 | 0.64 | 36 |
| 112 | GGGGAC | 49 | 27 | 0.64 | 34 |
| 113 | CCGCAA | 16 | 5 | 0.64 | 12 |
| 114 | CGTCGC | 8 | 1 | 0.62 | 6 |
| 115 | CGTGAC | 17 | 6 | 0.57 | 15 |
| 116 | CGCCCA | 33 | 16 | 0.56 | 22 |
| 117 | CTCTGC | 97 | 65 | 0.54 | 47 |

| | | | | | |
|-----|----------|-----|----|------|----|
| 118 | AGCGGG | 21 | 8 | 0.52 | 17 |
| 119 | ACCGCT | 15 | 5 | 0.5 | 11 |
| 120 | CCCAGG | 133 | 95 | 0.49 | 58 |
| 121 | CCCTCA | 71 | 45 | 0.49 | 39 |
| 122 | CCCCCA * | 77 | 49 | 0.49 | 42 |
| 123 | GGCGAA | 16 | 5 | 0.48 | 14 |
| 124 | CGGCTC | 29 | 13 | 0.47 | 19 |
| 125 | CTCGCC | 20 | 8 | 0.46 | 17 |
| 126 | CGGAGA | 20 | 8 | 0.45 | 14 |
| 127 | TCCCCA | 95 | 64 | 0.43 | 52 |
| 128 | GACACC | 44 | 24 | 0.42 | 33 |
| 129 | CTCCGA | 17 | 6 | 0.42 | 13 |
| 130 | CTCGTC | 17 | 6 | 0.42 | 14 |
| 131 | CGACCA | 13 | 4 | 0.39 | 11 |
| 132 | ATGACG | 17 | 6 | 0.37 | 12 |
| 133 | CCATCG | 17 | 6 | 0.37 | 13 |
| 134 | AGGGGA | 78 | 51 | 0.36 | 44 |
| 135 | GCTGCA | 77 | 50 | 0.35 | 43 |
| 136 | ACCCCA | 76 | 49 | 0.33 | 40 |
| 137 | CGGAGC | 21 | 9 | 0.33 | 16 |
| 138 | CCTCCG | 28 | 13 | 0.32 | 19 |
| 139 | CGGGAC | 16 | 6 | 0.3 | 10 |
| 140 | CCTGGA | 88 | 59 | 0.3 | 45 |
| 141 | AGGCGA | 18 | 7 | 0.29 | 17 |
| 142 | ACCCCT | 54 | 32 | 0.28 | 36 |
| 143 | GCTCCC | 56 | 34 | 0.27 | 36 |
| 144 | CGTCAC | 16 | 6 | 0.27 | 15 |
| 145 | AGCGCA | 16 | 6 | 0.26 | 11 |
| 146 | GAAGCC | 62 | 38 | 0.25 | 39 |
| 147 | GAGGCC | 79 | 52 | 0.22 | 42 |
| 148 | ACCCTC | 54 | 32 | 0.22 | 33 |
| 149 | CCCGGC | 37 | 20 | 0.21 | 21 |
| 150 | CGAGAA | 20 | 8 | 0.2 | 17 |
| 151 | CCACCG | 29 | 14 | 0.18 | 20 |
| 152 | ACTTCG | 16 | 6 | 0.17 | 14 |
| 153 | GATGAC | 48 | 28 | 0.17 | 35 |
| 154 | ACGAGG | 23 | 10 | 0.16 | 18 |
| 155 | CCGGAG | 20 | 8 | 0.15 | 18 |
| 156 | ACCCAC | 60 | 37 | 0.12 | 41 |
| 157 | CTGGGC | 105 | 74 | 0.11 | 50 |
| 158 | CCACGG | 23 | 10 | 0.09 | 19 |
| 159 | CGGTCC | 13 | 4 | 0.09 | 12 |

| | | | | | |
|-----|----------|----|----|------|----|
| 160 | AGCACC * | 54 | 33 | 0.09 | 40 |
| 161 | ACACCC | 53 | 32 | 0.08 | 38 |
| 162 | AGGGCC | 54 | 33 | 0.08 | 30 |
| 163 | CGCGAA | 6 | 1 | 0.02 | 6 |
| 164 | GAGCCC | 58 | 36 | 0.02 | 36 |
| 165 | CTGAGC | 71 | 46 | 0.02 | 45 |
| 166 | AATCGG | 13 | 4 | 0.02 | 11 |

【 0 2 0 0 】

【 表 1 0 】

表 10. STAR エレメントにおいて過剰に表れた二分染色体パターン

該パターンは、有意性係数に従いランク付けされる。これらは、参照としてヒトゲノムからのランダム配列を有するRSA-ツールを使用して決定された。線形判別式分析におけるほとんどの判別式変数を含むパターンがアスタリスクで示される。

| 番号 | 二分染色体配列 | 観察された出現 | 期待された出現 | 有意性係数 |
|----|-------------|---------|---------|-------|
| 1 | CCCN{2}CGG | 36 | 7 | 9.31 |
| 2 | CCGN{6}CCC | 40 | 10 | 7.3 |
| 3 | CAGN{0}CGG | 36 | 8 | 7.13 |
| 4 | CGCN{15}CCC | 34 | 8 | 6.88 |
| 5 | CGGN{9}GCC | 33 | 7 | 6.82 |
| 6 | CCCN{9}CGC | 35 | 8 | 6.72 |
| 7 | CCCN{1}GCG | 34 | 8 | 6.64 |
| 8 | CCCN{0}CAC | 103 | 48 | 6.61 |
| 9 | AGCN{16}CCG | 29 | 6 | 5.96 |
| 10 | CCCN{4}CGC | 34 | 8 | 5.8 |
| 11 | CGCN{13}GGA | 26 | 5 | 5.77 |
| 12 | GCGN{16}CCC | 30 | 7 | 5.74 |
| 13 | CGCN{5}GCA | 25 | 5 | 5.49 |
| 14 | CCCN{14}CCC | 101 | 49 | 5.43 |
| 15 | CTGN{4}CGC | 34 | 9 | 5.41 |
| 16 | CCAN{12}GCG | 28 | 6 | 5.37 |
| 17 | CGGN{11}CAG | 36 | 10 | 5.25 |
| 18 | CCCN{5}GCC | 75 | 33 | 4.87 |
| 19 | GCCN{0}CCC | 64 | 26 | 4.81 |
| 20 | CGCN{4}GAC | 19 | 3 | 4.78 |
| 21 | CGGN{0}CAG | 33 | 9 | 4.76 |
| 22 | CCCN{3}CGC | 32 | 8 | 4.67 |
| 23 | CGCN{1}GAC | 20 | 3 | 4.58 |

| | | | | |
|----|-------------|-----|----|------|
| 24 | GCGN{2}GCC | 29 | 7 | 4.54 |
| 25 | CCCN{4}GCC | 76 | 34 | 4.53 |
| 26 | CCCN{1}CCC | 103 | 52 | 4.53 |
| 27 | CCGN{13}CAG | 33 | 9 | 4.5 |
| 28 | GCCN{4}GGA | 64 | 27 | 4.48 |
| 29 | CCGN{3}GGA | 26 | 6 | 4.46 |
| 30 | AGGN{2}GGG | 118 | 63 | 4.44 |
| 31 | CACN{5}GCG | 22 | 4 | 4.42 |
| 32 | CGCN{17}CCA | 27 | 6 | 4.39 |
| 33 | CCCN{9}GGC | 69 | 30 | 4.38 |
| 34 | CCTN{5}GCG | 28 | 7 | 4.37 |
| 35 | GCGN{0}GAC | 19 | 3 | 4.32 |
| 36 | GCCN{0}GGC | 40 | 7 | 4.28 |
| 37 | GCGN{2}CCC | 26 | 6 | 4.27 |
| 38 | CCGN{11}CCC | 32 | 9 | 4.17 |
| 39 | CCCN{8}TCG | 23 | 5 | 4.12 |
| 40 | CCGN{17}GCC | 30 | 8 | 4.12 |
| 41 | GGGN{5}GGA | 101 | 52 | 4.11 |
| 42 | GGCN{6}GGA | 71 | 32 | 4.1 |
| 43 | CCAN{4}CCC | 96 | 48 | 4.1 |
| 44 | CCTN{14}CCG | 32 | 9 | 4.09 |
| 45 | GACN{12}GGC | 45 | 16 | 4.07 |
| 46 | CGCN{13}CCC | 30 | 8 | 4.04 |
| 47 | CAGN{16}CCC | 92 | 46 | 4.02 |
| 48 | AGCN{10}GGG | 75 | 35 | 3.94 |
| 49 | CGGN{13}GGC | 30 | 8 | 3.93 |
| 50 | CGGN{1}GCC | 30 | 8 | 3.92 |
| 51 | AGCN{0}GGC | 26 | 6 | 3.9 |
| 52 | CCCN{16}GGC | 64 | 28 | 3.89 |
| 53 | GCTN{19}CCC | 67 | 29 | 3.87 |
| 54 | CCCN{16}GGG | 88 | 31 | 3.81 |
| 55 | CCCN{9}CGG | 30 | 8 | 3.77 |
| 56 | CCCN{10}CGG | 30 | 8 | 3.76 |
| 57 | CCAN{0}GCG | 32 | 9 | 3.75 |
| 58 | GCCN{17}CGC | 26 | 6 | 3.74 |
| 59 | CCTN{6}CGC | 27 | 7 | 3.73 |
| 60 | GGAN{1}CCC | 63 | 27 | 3.71 |
| 61 | CGCN{18}CAC | 24 | 5 | 3.7 |
| 62 | CGCN{20}CCG | 21 | 4 | 3.69 |
| 63 | CCGN{0}GCA | 26 | 6 | 3.69 |
| 64 | CGCN{20}CCC | 28 | 7 | 3.69 |
| 65 | AGCN{15}CCC | 67 | 30 | 3.65 |

| | | | | |
|-----|---------------|----|----|------|
| 66 | CCTN{7}GGC | 69 | 31 | 3.63 |
| 67 | GCCN{5}CGC | 32 | 9 | 3.61 |
| 68 | GCCN{14}CGC | 28 | 7 | 3.59 |
| 69 | CAGN{11}CCC | 89 | 45 | 3.58 |
| 70 | GGGN{16}GAC | 53 | 21 | 3.57 |
| 71 | CCCN{15}GCG | 25 | 6 | 3.57 |
| 72 | CCCN{0}CGC | 37 | 12 | 3.54 |
| 73 | CCCN{16}AGC * | 67 | 30 | 3.54 |
| 74 | AGGN{9}GGG | 96 | 50 | 3.52 |
| 75 | CGCN{12}CTC | 28 | 7 | 3.46 |
| 76 | CACN{8}CGC | 23 | 5 | 3.43 |
| 77 | CCAN{7}CCG | 31 | 9 | 3.42 |
| 78 | CGGN{1}GCA | 25 | 6 | 3.41 |
| 79 | CGCN{14}CCC | 29 | 8 | 3.4 |
| 80 | AGCN{0}CCC | 76 | 36 | 3.4 |
| 81 | CGCN{13}GTC | 18 | 3 | 3.37 |
| 82 | CGGN{3}GCA | 26 | 7 | 3.35 |
| 83 | CGGN{0}GGC | 34 | 11 | 3.35 |
| 84 | GCCN{14}CCC | 68 | 31 | 3.33 |
| 85 | ACCN{7}CGC | 21 | 4 | 3.32 |
| 86 | AGGN{7}CGG | 33 | 10 | 3.31 |
| 87 | CCCN{16}CGA | 22 | 5 | 3.3 |
| 88 | CGCN{6}CAG | 31 | 9 | 3.29 |
| 89 | CAGN{11}GCG | 29 | 8 | 3.29 |
| 90 | CCGN{12}CCG | 19 | 4 | 3.26 |
| 91 | CGCN{18}CAG | 27 | 7 | 3.24 |
| 92 | CAGN{1}GGG | 80 | 39 | 3.21 |
| 93 | CGCN{0}CCC | 32 | 10 | 3.2 |
| 94 | CGGN{18}GCC | 26 | 7 | 3.18 |
| 95 | CGGN{15}GGC | 27 | 7 | 3.15 |
| 96 | CCCN{15}AGG | 72 | 34 | 3.14 |
| 97 | AGGN{20}GCG | 26 | 7 | 3.14 |
| 98 | CGGN{5}CTC | 26 | 7 | 3.13 |
| 99 | TCCN{17}CGA | 23 | 5 | 3.12 |
| 100 | GCNN{4}CCC | 30 | 9 | 3.08 |
| 101 | CCCN{2}CGC | 30 | 9 | 3.07 |
| 102 | CGTN{3}CAG | 28 | 8 | 3.06 |
| 103 | CCGN{13}GAG | 27 | 7 | 3.05 |
| 104 | CTCN{6}CGC | 28 | 8 | 3.04 |
| 105 | CGCN{4}GAG | 21 | 5 | 3.03 |
| 106 | GCNN{5}GGA | 24 | 6 | 3.03 |
| 107 | CCGN{1}CAG | 27 | 7 | 3.01 |

| | | | | |
|-----|-------------|----|----|------|
| 108 | CGCN{11}CCG | 18 | 3 | 2.99 |
| 109 | GCGN{19}CCC | 26 | 7 | 2.98 |
| 110 | CGCN{18}GAA | 21 | 5 | 2.98 |
| 111 | GGGN{19}GGA | 78 | 39 | 2.95 |
| 112 | CCAN{1}CGG | 24 | 6 | 2.94 |
| 113 | CCCN{7}GCG | 25 | 6 | 2.94 |
| 114 | AGGN{10}CCC | 84 | 43 | 2.92 |
| 115 | CCAN{0}GGG | 97 | 52 | 2.88 |
| 116 | CAGN{10}CCC | 82 | 41 | 2.87 |
| 117 | CCGN{18}CCG | 19 | 4 | 2.86 |
| 118 | CCGN{18}GGC | 26 | 7 | 2.85 |
| 119 | CCCN{2}GCG | 24 | 6 | 2.84 |
| 120 | CGCN{1}GGC | 25 | 7 | 2.83 |
| 121 | CCGN{5}GAC | 19 | 4 | 2.81 |
| 122 | GGAN{0}CCC | 52 | 22 | 2.8 |
| 123 | CCCN{1}CCG | 29 | 9 | 2.78 |
| 124 | CCCN{15}ACG | 23 | 6 | 2.75 |
| 125 | AGCN{8}CCC | 66 | 31 | 2.73 |
| 126 | CCCN{3}GGC | 60 | 27 | 2.71 |
| 127 | AGGN{9}CGG | 31 | 10 | 2.7 |
| 128 | CCCN{14}CGC | 27 | 8 | 2.7 |
| 129 | CCGN{0}CCG | 19 | 4 | 2.7 |
| 130 | CGCN{8}AGC | 23 | 6 | 2.69 |
| 131 | CGCN{19}ACC | 21 | 5 | 2.68 |
| 132 | GCGN{17}GAC | 17 | 3 | 2.66 |
| 133 | AGCN{1}GCG | 24 | 6 | 2.63 |
| 134 | CCGN{11}GGC | 31 | 10 | 2.63 |
| 135 | CGGN{4}AGA | 26 | 7 | 2.63 |
| 136 | CGCN{14}CCG | 17 | 3 | 2.62 |
| 137 | CCTN{20}GCG | 24 | 6 | 2.62 |
| 138 | CCAN{10}CGC | 26 | 7 | 2.61 |
| 139 | CCCN{20}CAC | 69 | 33 | 2.6 |
| 140 | CCGN{11}GCC | 27 | 8 | 2.6 |
| 141 | CGCN{18}CCC | 26 | 7 | 2.59 |
| 142 | CGGN{15}CGC | 16 | 3 | 2.57 |
| 143 | CGCN{16}GCC | 24 | 6 | 2.55 |
| 144 | CGCN{20}GGC | 23 | 6 | 2.54 |
| 145 | CGCN{19}CCG | 18 | 4 | 2.52 |
| 146 | CGGN{10}CCA | 28 | 8 | 2.51 |
| 147 | CGCN{17}CCC | 26 | 7 | 2.51 |
| 148 | CGCN{11}ACA | 23 | 6 | 2.51 |
| 149 | CGGN{0}ACC | 17 | 3 | 2.5 |

| | | | | |
|-----|-------------|----|----|------|
| 150 | GCGN{10}GCC | 24 | 6 | 2.49 |
| 151 | GCGN{8}GAC | 17 | 3 | 2.49 |
| 152 | CCCN{15}GGG | 84 | 32 | 2.44 |
| 153 | CGGN{16}GGC | 27 | 8 | 2.44 |
| 154 | CGCN{16}CCA | 23 | 6 | 2.42 |
| 155 | GCCN{3}CCC | 73 | 36 | 2.4 |
| 156 | CAGN{4}GGG | 94 | 51 | 2.4 |
| 157 | CCCN{6}GCG | 23 | 6 | 2.38 |
| 158 | CCGN{16}CGC | 17 | 3 | 2.38 |
| 159 | CCCN{17}GCA | 61 | 28 | 2.37 |
| 160 | CGCN{13}TCC | 24 | 6 | 2.37 |
| 161 | GCCN{1}CGC | 29 | 9 | 2.36 |
| 162 | CCGN{19}GAG | 26 | 7 | 2.35 |
| 163 | GGGN{10}GGA | 89 | 48 | 2.35 |
| 164 | CAGN{5}CCG | 32 | 11 | 2.35 |
| 165 | CGCN{3}AGA | 19 | 4 | 2.32 |
| 166 | GCCN{0}GCC | 29 | 9 | 2.32 |
| 167 | CCCN{8}GGC | 61 | 28 | 2.31 |
| 168 | CCTN{6}GCG | 22 | 6 | 2.29 |
| 169 | GACN{6}CCC | 48 | 20 | 2.29 |
| 170 | CGGN{1}CCC | 26 | 8 | 2.27 |
| 171 | CCCN{15}CCG | 30 | 10 | 2.27 |
| 172 | CAGN{9}CCC | 84 | 44 | 2.26 |
| 173 | CGGN{10}GGC | 27 | 8 | 2.26 |
| 174 | CGAN{10}ACG | 10 | 1 | 2.26 |
| 175 | GCGN{3}TCC | 21 | 5 | 2.26 |
| 176 | CCCN{3}GCC | 75 | 38 | 2.24 |
| 177 | GCGN{1}ACC | 17 | 3 | 2.24 |
| 178 | CCGN{9}AGG | 27 | 8 | 2.23 |
| 179 | CGCN{16}CAG | 26 | 8 | 2.23 |
| 180 | GGCN{0}CCC | 62 | 29 | 2.22 |
| 181 | AGGN{12}CCG | 26 | 8 | 2.19 |
| 182 | CCGN{0}GCG | 16 | 3 | 2.19 |
| 183 | CCGN{2}GCC | 30 | 10 | 2.18 |
| 184 | CCGN{11}GTC | 19 | 4 | 2.17 |
| 185 | CAGN{0}CCC | 88 | 47 | 2.17 |
| 186 | CCCN{5}CCG | 32 | 11 | 2.17 |
| 187 | GCCN{20}CCC | 66 | 32 | 2.15 |
| 188 | GACN{2}CGC | 18 | 4 | 2.14 |
| 189 | CGCN{6}CAC | 23 | 6 | 2.13 |
| 190 | AGGN{14}GCG | 25 | 7 | 2.1 |
| 191 | GACN{5}CGC | 17 | 3 | 2.1 |

| | | | | |
|-----|--------------|----|----|------|
| 192 | CCTN{19}CCG | 29 | 9 | 2.1 |
| 193 | CCGN{12}GGA | 24 | 7 | 2.08 |
| 194 | GGCN{9}GAC * | 44 | 18 | 2.08 |
| 195 | AGGN{10}GGG | 94 | 52 | 2.07 |
| 196 | CCGN{10}GAG | 25 | 7 | 2.07 |
| 197 | CGCN{6}GGA | 20 | 5 | 2.06 |
| 198 | CGCN{7}AGC | 23 | 6 | 2.04 |
| 199 | CCAN{13}CGG | 26 | 8 | 2.03 |
| 200 | CGGN{6}GGA | 25 | 7 | 2.03 |
| 201 | CGCN{19}GCC | 24 | 7 | 2.03 |
| 202 | CCAN{12}CGC | 24 | 7 | 2.02 |
| 203 | CGGN{1}GGC | 41 | 16 | 2.02 |
| 204 | GCGN{3}CCA | 25 | 7 | 2.01 |
| 205 | AGGN{1}CGC | 21 | 5 | 2 |
| 206 | CTCN{5}CGC | 24 | 7 | 1.98 |
| 207 | CCCN{0}ACG | 30 | 10 | 1.97 |
| 208 | CAGN{17}CCG | 29 | 9 | 1.96 |
| 209 | GGCN{4}CCC | 62 | 30 | 1.96 |
| 210 | AGGN{8}GCG | 26 | 8 | 1.96 |
| 211 | CTGN{1}CCC | 88 | 48 | 1.94 |
| 212 | CCCN{16}CAG | 85 | 46 | 1.94 |
| 213 | CGCN{9}GAC | 16 | 3 | 1.93 |
| 214 | CAGN{6}CCG | 29 | 9 | 1.92 |
| 215 | CGTN{12}CGC | 11 | 1 | 1.92 |
| 216 | CTCN{7}GCC | 69 | 35 | 1.92 |
| 217 | CGCN{19}TCC | 22 | 6 | 1.92 |
| 218 | CCCN{7}GCC | 67 | 33 | 1.91 |
| 219 | CAGN{13}CGG | 30 | 10 | 1.9 |
| 220 | CGCN{1}GCC | 27 | 8 | 1.9 |
| 221 | CGCN{17}CCG | 17 | 4 | 1.89 |
| 222 | AGGN{4}CCC | 63 | 31 | 1.89 |
| 223 | AGCN{10}CGC | 21 | 5 | 1.89 |
| 224 | CCCN{11}CGG | 30 | 10 | 1.88 |
| 225 | CCCN{8}GCC | 75 | 39 | 1.86 |
| 226 | CCGN{1}CGG | 22 | 3 | 1.86 |
| 227 | CCCN{1}ACC | 71 | 36 | 1.85 |
| 228 | CGCN{0}CAG | 25 | 7 | 1.85 |
| 229 | CCGN{19}TGC | 23 | 6 | 1.82 |
| 230 | GCGN{4}CGA | 12 | 2 | 1.82 |
| 231 | CCGN{19}GCC | 30 | 10 | 1.82 |
| 232 | CCAN{10}CCC | 85 | 46 | 1.81 |
| 233 | CAGN{13}GGG | 91 | 51 | 1.81 |

| | | | | |
|-----|-------------|----|----|------|
| 234 | AGCN{18}CGG | 23 | 6 | 1.81 |
| 235 | CGAN{8}CGC | 11 | 1 | 1.81 |
| 236 | AGCN{4}CCC | 63 | 31 | 1.8 |
| 237 | GGAN{6}CCC | 61 | 30 | 1.8 |
| 238 | CGGN{13}AAG | 23 | 6 | 1.8 |
| 239 | ACCN{11}CGC | 19 | 5 | 1.79 |
| 240 | CCGN{12}CAG | 28 | 9 | 1.78 |
| 241 | CCCN{12}GGG | 76 | 29 | 1.77 |
| 242 | CACN{17}ACG | 22 | 6 | 1.76 |
| 243 | CAGN{18}CCC | 82 | 44 | 1.76 |
| 244 | CGTN{10}GTC | 19 | 5 | 1.75 |
| 245 | CCCN{13}GCG | 23 | 6 | 1.75 |
| 246 | GCAN{1}CGC | 20 | 5 | 1.73 |
| 247 | AGAN{4}CCG | 24 | 7 | 1.73 |
| 248 | GCGN{10}AGC | 22 | 6 | 1.72 |
| 249 | CGCN{0}GGA | 12 | 2 | 1.72 |
| 250 | CGGN{4}GAC | 17 | 4 | 1.69 |
| 251 | CCCN{12}CGC | 26 | 8 | 1.68 |
| 252 | GCCN{15}CCC | 65 | 33 | 1.68 |
| 253 | GCGN{6}TCC | 20 | 5 | 1.66 |
| 254 | CGGN{3}CAG | 33 | 12 | 1.65 |
| 255 | CCCN{3}CCA | 88 | 49 | 1.65 |
| 256 | AGCN{3}CCC | 59 | 28 | 1.65 |
| 257 | GGGN{16}GCA | 65 | 33 | 1.65 |
| 258 | AGGN{8}CCG | 28 | 9 | 1.64 |
| 259 | CCCN{0}CCG | 29 | 10 | 1.64 |
| 260 | GCGN{5}GAC | 16 | 3 | 1.64 |
| 261 | CCCN{9}ACC | 60 | 29 | 1.64 |
| 262 | CTGN{5}CGC | 25 | 8 | 1.64 |
| 263 | CGCN{14}CTC | 23 | 7 | 1.64 |
| 264 | CGGN{14}GCA | 23 | 7 | 1.63 |
| 265 | CCGN{8}GCC | 26 | 8 | 1.62 |
| 266 | CCGN{7}CAC | 23 | 7 | 1.62 |
| 267 | AGCN{8}GCG | 21 | 6 | 1.61 |
| 268 | CGGN{16}GGA | 29 | 10 | 1.61 |
| 269 | CCAN{12}CCG | 26 | 8 | 1.61 |
| 270 | CGGN{2}CCC | 26 | 8 | 1.6 |
| 271 | CCAN{13}GGG | 71 | 37 | 1.6 |
| 272 | CGGN{15}GCA | 21 | 6 | 1.6 |
| 273 | CGCN{9}GCA | 20 | 5 | 1.58 |
| 274 | CGGN{19}CCA | 26 | 8 | 1.58 |
| 275 | GGGN{15}CGA | 20 | 5 | 1.57 |

| | | | | |
|-----|-------------|----|----|------|
| 276 | CCCN{10}CGC | 26 | 8 | 1.57 |
| 277 | CTCN{14}CGC | 26 | 8 | 1.55 |
| 278 | CACN{11}GCG | 20 | 5 | 1.55 |
| 279 | CCGN{2}GGC | 24 | 7 | 1.55 |
| 280 | CTGN{18}CCC | 85 | 47 | 1.54 |
| 281 | GGGN{13}CAC | 58 | 28 | 1.54 |
| 282 | CCTN{15}GGC | 62 | 31 | 1.54 |
| 283 | CCCN{20}CGA | 20 | 5 | 1.54 |
| 284 | CCCN{8}CGA | 20 | 5 | 1.53 |
| 285 | GAGN{7}CCC | 61 | 30 | 1.53 |
| 286 | CGCN{2}CCG | 22 | 6 | 1.53 |
| 287 | CCCN{0}TCC | 98 | 57 | 1.52 |
| 288 | AGCN{0}GCC | 21 | 6 | 1.52 |
| 289 | CCCN{2}TCC | 82 | 45 | 1.52 |
| 290 | CCGN{5}CCC | 30 | 10 | 1.52 |
| 291 | CGCN{13}CGC | 16 | 3 | 1.51 |
| 292 | CCCN{1}CGC | 28 | 9 | 1.51 |
| 293 | GCCN{16}GCA | 53 | 25 | 1.51 |
| 294 | CCCN{16}CCA | 84 | 46 | 1.5 |
| 295 | CCGN{13}CGC | 19 | 5 | 1.5 |
| 296 | CCGN{17}CAG | 28 | 9 | 1.49 |
| 297 | CGGN{18}GGC | 26 | 8 | 1.49 |
| 298 | CCGN{14}AGG | 28 | 7 | 1.49 |
| 299 | CCCN{5}CGG | 26 | 8 | 1.49 |
| 300 | CCCN{6}GGA | 58 | 28 | 1.49 |
| 301 | ACGN{2}CCC | 20 | 5 | 1.49 |
| 302 | CCAN{9}CCG | 27 | 9 | 1.48 |
| 303 | CCCN{19}CCA | 78 | 42 | 1.48 |
| 304 | CAGN{0}GGG | 77 | 41 | 1.48 |
| 305 | AGCN{1}CCC | 58 | 28 | 1.47 |
| 306 | GCGN{7}TCC | 27 | 9 | 1.46 |
| 307 | ACGN{18}CCA | 25 | 8 | 1.46 |
| 308 | GCTN{14}CCC | 61 | 30 | 1.46 |
| 309 | GCGN{14}CCC | 23 | 7 | 1.46 |
| 310 | GCGN{19}AGC | 20 | 5 | 1.45 |
| 311 | CCGN{8}CAG | 29 | 10 | 1.45 |
| 312 | GCGN{6}GCC | 22 | 6 | 1.45 |
| 313 | GCGN{10}GCA | 20 | 5 | 1.44 |
| 314 | CCTN{7}GCC | 69 | 36 | 1.44 |
| 315 | GCCN{13}GCC | 54 | 26 | 1.42 |
| 316 | CCCN{14}GCC | 63 | 32 | 1.42 |
| 317 | CCCN{15}CGG | 26 | 8 | 1.42 |

| | | | | |
|-----|-------------|----|----|------|
| 318 | CCAN{13}CGC | 23 | 7 | 1.42 |
| 319 | AGCN{11}GGG | 67 | 35 | 1.41 |
| 320 | GGAN{0}GCC | 64 | 32 | 1.4 |
| 321 | GCCN{3}TCC | 61 | 30 | 1.4 |
| 322 | CCTN{5}GCC | 69 | 36 | 1.39 |
| 323 | CGGN{18}CCC | 25 | 8 | 1.39 |
| 324 | CCTN{3}GGC | 59 | 29 | 1.38 |
| 325 | CCGN{0}CTC | 22 | 6 | 1.38 |
| 326 | AGCN{17}GCG | 19 | 5 | 1.37 |
| 327 | ACGN{14}GGG | 20 | 5 | 1.37 |
| 328 | CGAN{12}GGC | 19 | 5 | 1.37 |
| 329 | CCCN{20}CGC | 24 | 7 | 1.37 |
| 330 | ACGN{12}CTG | 24 | 7 | 1.36 |
| 331 | CCGN{0}CCC | 36 | 14 | 1.36 |
| 332 | CCGN{10}GGA | 23 | 7 | 1.36 |
| 333 | CCCN{3}GCG | 21 | 6 | 1.36 |
| 334 | GCGN{14}CGC | 22 | 3 | 1.35 |
| 335 | CCGN{8}CGC | 16 | 4 | 1.35 |
| 336 | CGCN{10}ACA | 22 | 6 | 1.34 |
| 337 | CCCN{19}CCG | 28 | 10 | 1.33 |
| 338 | CACN{14}CGC | 20 | 5 | 1.32 |
| 339 | GACN{3}GGC | 46 | 21 | 1.32 |
| 340 | GAAN{7}CGC | 19 | 5 | 1.32 |
| 341 | CGCN{16}GGC | 21 | 6 | 1.31 |
| 342 | GGCN{9}CCC | 64 | 33 | 1.31 |
| 343 | CCCN{9}GCC | 64 | 33 | 1.31 |
| 344 | CGCN{0}TGC | 26 | 9 | 1.3 |
| 345 | CCTN{8}GGC | 67 | 35 | 1.3 |
| 346 | CCAN{8}CCC | 82 | 46 | 1.29 |
| 347 | GACN{2}CCC | 42 | 18 | 1.28 |
| 348 | GGCN{1}CCC | 54 | 26 | 1.27 |
| 349 | CGCN{0}AGC | 24 | 7 | 1.26 |
| 350 | AGGN{4}GCG | 28 | 10 | 1.26 |
| 351 | CGGN{6}TCC | 22 | 6 | 1.25 |
| 352 | ACGN{19}GGC | 20 | 5 | 1.25 |
| 353 | CCCN{8}ACG | 21 | 6 | 1.24 |
| 354 | CCCN{18}GCC | 62 | 31 | 1.24 |
| 355 | GCCN{2}CGA | 19 | 5 | 1.24 |
| 356 | CCCN{8}GCG | 28 | 10 | 1.23 |
| 357 | CCCN{0}CTC | 76 | 41 | 1.23 |
| 358 | GCCN{11}CGC | 27 | 9 | 1.22 |
| 359 | AGCN{9}CCC | 59 | 29 | 1.22 |

| | | | | |
|-----|-------------|-----|----|------|
| 360 | GCTN{0}GCC | 71 | 38 | 1.21 |
| 361 | CGCN{3}CCC | 26 | 9 | 1.21 |
| 362 | CCCN{2}CCC | 117 | 72 | 1.19 |
| 363 | GCCN{9}CGC | 23 | 7 | 1.19 |
| 364 | GCAN{19}CGC | 19 | 5 | 1.19 |
| 365 | CAGN{4}CGG | 32 | 12 | 1.18 |
| 366 | CAGN{2}GGG | 80 | 44 | 1.17 |
| 367 | GCCN{16}CCC | 67 | 35 | 1.16 |
| 368 | GAGN{5}CCC | 60 | 30 | 1.16 |
| 369 | CCTN{16}TCG | 20 | 6 | 1.16 |
| 370 | CCCN{2}GGC | 62 | 32 | 1.15 |
| 371 | GCGN{13}GGA | 24 | 8 | 1.15 |
| 372 | GCCN{17}GGC | 66 | 25 | 1.15 |
| 373 | CCCN{14}GGC | 58 | 29 | 1.14 |
| 374 | AGGN{8}CCG | 31 | 12 | 1.14 |
| 375 | CACN{0}CGC | 32 | 12 | 1.14 |
| 376 | CGGN{18}CAG | 28 | 10 | 1.14 |
| 377 | AGCN{1}GCC | 57 | 28 | 1.13 |
| 378 | CGCN{18}GGC | 23 | 7 | 1.13 |
| 379 | CCCN{5}AGG | 64 | 33 | 1.11 |
| 380 | AACN{0}GCG | 9 | 1 | 1.11 |
| 381 | CCCN{10}CCA | 88 | 50 | 1.09 |
| 382 | CGCN{13}GAG | 20 | 6 | 1.09 |
| 383 | CGCN{7}GCC | 25 | 8 | 1.08 |
| 384 | CCCN{9}CCG | 28 | 10 | 1.07 |
| 385 | CGCN{16}CCC | 24 | 8 | 1.05 |
| 386 | GAAN{13}CGC | 18 | 5 | 1.05 |
| 387 | GGCN{3}CCC | 49 | 23 | 1.03 |
| 388 | TCCN{11}CCA | 87 | 50 | 1.03 |
| 389 | CACN{0}CCC | 70 | 38 | 1.02 |
| 390 | CGCN{16}CCG | 15 | 3 | 1.02 |
| 391 | CGGN{15}AGC | 21 | 6 | 1.02 |
| 392 | CCCN{12}GCG | 21 | 6 | 1.02 |
| 393 | CCCN{9}GAG | 59 | 30 | 1.01 |
| 394 | CCGN{20}TCC | 24 | 8 | 1.01 |
| 395 | CGCN{0}CGC | 17 | 4 | 1.01 |
| 396 | ATGN{7}CGG | 20 | 6 | 1 |
| 397 | GGGN{20}GCA | 59 | 30 | 1 |
| 398 | CGGN{4}GGC | 26 | 9 | 0.99 |
| 399 | CGGN{16}AGC | 22 | 7 | 0.99 |
| 400 | CGGN{5}GGC | 25 | 8 | 0.99 |
| 401 | GCGN{0}GGA | 25 | 8 | 0.98 |

| | | | | |
|-----|---------------|----|----|------|
| 402 | GGCN{20}CAC | 52 | 25 | 0.98 |
| 403 | CCCN{9}CCC | 97 | 58 | 0.97 |
| 404 | ACCN{17}GGC | 44 | 20 | 0.97 |
| 405 | CCCN{6}CGA | 18 | 5 | 0.96 |
| 406 | AAGN{10}CGG | 26 | 9 | 0.96 |
| 407 | CGCN{17}CAC | 21 | 6 | 0.95 |
| 408 | CCCN{16}CGG | 25 | 8 | 0.94 |
| 409 | GACN{18}GGC | 39 | 17 | 0.94 |
| 410 | GGGN{15}GAC | 47 | 22 | 0.92 |
| 411 | GCCN{4}TCC | 66 | 35 | 0.92 |
| 412 | GGCN{15}CCC | 56 | 28 | 0.92 |
| 413 | CAGN{12}CGC | 24 | 8 | 0.92 |
| 414 | CCAN{3}GCG | 22 | 7 | 0.91 |
| 415 | CCGN{16}GAG | 22 | 7 | 0.9 |
| 416 | AGCN{2}CGC | 24 | 8 | 0.89 |
| 417 | GAGN{4}CCC | 54 | 27 | 0.89 |
| 418 | AGGN{3}CGC | 23 | 7 | 0.88 |
| 419 | CACN{13}AGG * | 67 | 36 | 0.88 |
| 420 | CCCN{4}CAG | 88 | 51 | 0.88 |
| 421 | CCCN{2}GAA | 63 | 33 | 0.87 |
| 422 | CGCN{19}GAG | 21 | 6 | 0.87 |
| 423 | ACGN{18}GGG | 21 | 6 | 0.87 |
| 424 | CCCN{4}GGC | 62 | 32 | 0.87 |
| 425 | CGGN{9}GAG | 28 | 10 | 0.86 |
| 426 | CCCN{3}GGG | 66 | 26 | 0.86 |
| 427 | GAGN{4}GGC | 66 | 35 | 0.85 |
| 428 | CGCN{5}GAG | 18 | 5 | 0.84 |
| 429 | CCGN{20}AGG | 24 | 8 | 0.84 |
| 430 | CCCN{15}CCC | 88 | 51 | 0.83 |
| 431 | AGGN{17}CCG | 25 | 8 | 0.82 |
| 432 | AGGN{6}GGG | 89 | 52 | 0.82 |
| 433 | GGCN{20}CCC | 57 | 29 | 0.82 |
| 434 | GCAN{17}CGC | 19 | 5 | 0.82 |
| 435 | CGAN{11}ACG | 9 | 1 | 0.81 |
| 436 | CGCN{2}GGA | 19 | 5 | 0.81 |
| 437 | CTGN{5}CCC | 79 | 45 | 0.8 |
| 438 | TCCN{20}CCA | 77 | 43 | 0.8 |
| 439 | CCAN{2}GGG | 59 | 30 | 0.8 |
| 440 | CCGN{15}GCG | 14 | 3 | 0.8 |
| 441 | CCAN{5}GGG | 69 | 38 | 0.79 |
| 442 | CGGN{1}TGC | 24 | 8 | 0.79 |
| 443 | CCCN{14}GCG | 21 | 6 | 0.79 |

| | | | | |
|-----|-------------|----|----|------|
| 444 | CAGN{0}CCG | 27 | 10 | 0.79 |
| 445 | GCCN{9}TCC | 60 | 31 | 0.78 |
| 446 | AGGN{20}CGC | 22 | 7 | 0.78 |
| 447 | CCCN{6}GAC | 42 | 19 | 0.77 |
| 448 | CGGN{11}CCA | 23 | 7 | 0.76 |
| 449 | GGGN{14}CAC | 57 | 29 | 0.75 |
| 450 | GCAN{15}CGC | 19 | 5 | 0.74 |
| 451 | CGCN{2}ACA | 20 | 6 | 0.74 |
| 452 | ACCN{9}CCC | 57 | 29 | 0.73 |
| 453 | GCGN{9}CGC | 20 | 3 | 0.73 |
| 454 | CAGN{15}GCG | 23 | 7 | 0.73 |
| 455 | CCCN{18}GTC | 45 | 21 | 0.72 |
| 456 | GCGN{3}CCC | 24 | 8 | 0.72 |
| 457 | CGGN{11}GCC | 23 | 8 | 0.72 |
| 458 | CCCN{1}CGG | 24 | 8 | 0.71 |
| 459 | GCCN{4}CCA | 70 | 38 | 0.71 |
| 460 | CCCN{4}CCG | 30 | 12 | 0.7 |
| 461 | CGTN{2}GCA | 21 | 6 | 0.7 |
| 462 | AGCN{7}TCG | 18 | 5 | 0.69 |
| 463 | CCGN{15}GAA | 20 | 6 | 0.69 |
| 464 | ACCN{5}CCC | 62 | 33 | 0.69 |
| 465 | CGCN{14}GAG | 19 | 5 | 0.68 |
| 466 | CCCN{7}CGC | 30 | 12 | 0.68 |
| 467 | GAGN{12}CGC | 21 | 6 | 0.68 |
| 468 | GGCN{17}CCC | 58 | 30 | 0.67 |
| 469 | ACGN{11}CTC | 21 | 7 | 0.65 |
| 470 | ACAN{9}CGG | 24 | 8 | 0.65 |
| 471 | CTGN{7}CCC | 82 | 47 | 0.65 |
| 472 | CCCN{2}GCC | 72 | 40 | 0.65 |
| 473 | CGGN{2}GCA | 24 | 8 | 0.64 |
| 474 | CCCN{0}TGC | 83 | 48 | 0.64 |
| 475 | CGCN{7}ACC | 18 | 5 | 0.63 |
| 476 | GCAN{2}GCC | 54 | 27 | 0.63 |
| 477 | GCGN{8}CCA | 20 | 6 | 0.63 |
| 478 | AGCN{0}CGC | 22 | 7 | 0.63 |
| 479 | GCGN{2}GCA | 18 | 5 | 0.63 |
| 480 | CCGN{2}GTC | 18 | 5 | 0.62 |
| 481 | CCGN{3}ACA | 21 | 7 | 0.62 |
| 482 | ACGN{13}TGG | 21 | 7 | 0.62 |
| 483 | CCAN{8}CGC | 23 | 8 | 0.62 |
| 484 | CCGN{9}GGC | 23 | 8 | 0.61 |
| 485 | CCAN{5}CCG | 25 | 9 | 0.61 |

| | | | | |
|-----|--------------|----|----|------|
| 486 | AGGN{3}GGG | 97 | 59 | 0.61 |
| 487 | CAGN{2}GGC | 78 | 45 | 0.61 |
| 488 | CCCN{8}CAG | 81 | 47 | 0.61 |
| 489 | AGCN{5}CAG | 80 | 46 | 0.6 |
| 490 | CGGN{16}GCC | 22 | 7 | 0.6 |
| 491 | GCGN{15}CCC | 23 | 8 | 0.6 |
| 492 | CCCN{11}GCC | 59 | 31 | 0.59 |
| 493 | CGAN{2}ACG | 9 | 1 | 0.59 |
| 494 | CGGN{4}GCC | 22 | 7 | 0.59 |
| 495 | CACN{6}CGC | 19 | 6 | 0.59 |
| 496 | CGGN{5}ACG | 11 | 2 | 0.59 |
| 497 | CTGN{4}GCC * | 66 | 36 | 0.59 |
| 498 | GGGN{18}CGA | 18 | 5 | 0.59 |
| 499 | CCTN{8}CGC | 22 | 7 | 0.59 |
| 500 | GCCN{4}CCC | 67 | 37 | 0.58 |
| 501 | CGGN{10}GCC | 22 | 7 | 0.58 |
| 502 | GCCN{5}GGA | 54 | 27 | 0.57 |
| 503 | ACCN{7}GCG | 15 | 4 | 0.57 |
| 504 | CCCN{8}CGC | 24 | 8 | 0.57 |
| 505 | CAGN{5}CCC | 77 | 44 | 0.56 |
| 506 | CACN{14}GGA | 63 | 34 | 0.56 |
| 507 | CCCN{1}GCC | 94 | 57 | 0.55 |
| 508 | CCCN{5}AGC | 67 | 37 | 0.55 |
| 509 | GGCN{5}GGA | 59 | 31 | 0.55 |
| 510 | CGAN{17}GAG | 19 | 6 | 0.55 |
| 511 | CGCN{7}ACA | 18 | 5 | 0.54 |
| 512 | CCAN{13}CCC | 87 | 52 | 0.54 |
| 513 | CGGN{20}GGC | 24 | 8 | 0.54 |
| 514 | CCCN{17}GCC | 58 | 30 | 0.53 |
| 515 | CCTN{10}CCG | 30 | 12 | 0.53 |
| 516 | CCCN{8}CCG | 27 | 10 | 0.53 |
| 517 | CGCN{3}GAG | 18 | 5 | 0.52 |
| 518 | CGCN{7}AAG | 17 | 5 | 0.51 |
| 519 | CGGN{11}GGA | 23 | 8 | 0.51 |
| 520 | CCGN{15}CCG | 15 | 4 | 0.51 |
| 521 | CCCN{3}GCA | 57 | 30 | 0.51 |
| 522 | CGGN{2}CAG | 24 | 8 | 0.5 |
| 523 | AGGN{2}CCG | 24 | 8 | 0.5 |
| 524 | CCCN{4}CAC | 69 | 38 | 0.5 |
| 525 | GGAN{19}CCC | 56 | 29 | 0.49 |
| 526 | CCCN{8}CAC | 68 | 38 | 0.49 |
| 527 | ACCN{6}CCG | 18 | 5 | 0.49 |

| | | | | |
|-----|-------------|----|----|------|
| 528 | CCCN{6}GGC | 54 | 28 | 0.49 |
| 529 | CCCN{6}CCG | 29 | 11 | 0.48 |
| 530 | CGCN{14}GCC | 26 | 9 | 0.47 |
| 531 | CCGN{5}TCC | 25 | 9 | 0.46 |
| 532 | GCCN{6}GCC | 55 | 28 | 0.46 |
| 533 | CGGN{7}GGA | 24 | 8 | 0.45 |
| 534 | GGGN{6}GGA | 87 | 52 | 0.44 |
| 535 | GCCN{12}TCC | 60 | 32 | 0.44 |
| 536 | AGTN{16}CCG | 17 | 5 | 0.44 |
| 537 | GGCN{19}GCC | 68 | 29 | 0.44 |
| 538 | CCGN{3}CCG | 22 | 7 | 0.44 |
| 539 | CCCN{8}ACC | 58 | 31 | 0.44 |
| 540 | CAGN{15}GCC | 77 | 44 | 0.44 |
| 541 | CCCN{17}CGG | 24 | 8 | 0.44 |
| 542 | GCGN{1}CCA | 22 | 7 | 0.44 |
| 543 | CCCN{14}CAG | 79 | 46 | 0.44 |
| 544 | CCCN{8}CCC | 89 | 53 | 0.44 |
| 545 | ACAN{12}GCG | 23 | 8 | 0.43 |
| 546 | AGGN{4}CCG | 23 | 8 | 0.43 |
| 547 | CGCN{13}GCC | 23 | 8 | 0.43 |
| 548 | GAGN{2}CGC | 23 | 8 | 0.42 |
| 549 | CCCN{9}GCG | 21 | 7 | 0.42 |
| 550 | CGCN{17}ACA | 17 | 5 | 0.42 |
| 551 | GCGN{17}CCA | 23 | 8 | 0.42 |
| 552 | AAGN{18}CCG | 20 | 6 | 0.42 |
| 553 | CGCN{1}GGA | 18 | 5 | 0.41 |
| 554 | CCAN{1}CCC | 90 | 54 | 0.41 |
| 555 | CGTN{18}TGC | 20 | 6 | 0.41 |
| 556 | TCCN{14}CGA | 17 | 5 | 0.41 |
| 557 | CACN{5}GGG | 56 | 29 | 0.4 |
| 558 | CCGN{12}GCA | 21 | 7 | 0.4 |
| 559 | CTGN{6}CCC | 77 | 44 | 0.4 |
| 560 | CGGN{8}GGC | 32 | 13 | 0.4 |
| 561 | CCAN{11}GGG | 68 | 38 | 0.4 |
| 562 | ACGN{19}CAA | 21 | 7 | 0.39 |
| 563 | GGGN{20}CCC | 72 | 31 | 0.39 |
| 564 | CGCN{3}CAG | 23 | 8 | 0.39 |
| 565 | AGCN{17}GGG | 58 | 31 | 0.37 |
| 566 | CACN{20}CCG | 21 | 7 | 0.37 |
| 567 | ACGN{17}CAG | 24 | 8 | 0.37 |
| 568 | AGGN{1}CCC | 60 | 32 | 0.37 |
| 569 | CGTN{12}CAC | 20 | 6 | 0.37 |

| | | | | |
|-----|-------------|----|----|------|
| 570 | CGGN{9}GGC | 23 | 8 | 0.37 |
| 571 | CGCN{10}GCG | 18 | 3 | 0.37 |
| 572 | CCCN{6}CTC | 80 | 47 | 0.36 |
| 573 | CCGN{10}AGG | 23 | 8 | 0.36 |
| 574 | CCCN{18}CAG | 79 | 46 | 0.36 |
| 575 | AGCN{17}CCG | 21 | 7 | 0.36 |
| 576 | AGCN{9}GCG | 18 | 5 | 0.36 |
| 577 | CCAN{3}GGC | 62 | 34 | 0.36 |
| 578 | CCCN{11}GGC | 57 | 30 | 0.35 |
| 579 | ACGN{5}GCA | 23 | 8 | 0.35 |
| 580 | CCCN{14}CGG | 23 | 8 | 0.35 |
| 581 | CCCN{5}CCA | 91 | 55 | 0.35 |
| 582 | CCGN{1}AGG | 22 | 7 | 0.34 |
| 583 | GGGN{10}GAC | 45 | 22 | 0.34 |
| 584 | CGCN{15}CCA | 20 | 6 | 0.34 |
| 585 | CCTN{19}CGC | 22 | 7 | 0.34 |
| 586 | CGTN{3}CGC | 10 | 2 | 0.33 |
| 587 | AGCN{14}CCG | 21 | 7 | 0.33 |
| 588 | GGCN{2}CGA | 17 | 5 | 0.33 |
| 589 | CAGN{8}CCC | 79 | 46 | 0.33 |
| 590 | CCGN{2}GAC | 16 | 4 | 0.33 |
| 591 | AGCN{19}AGG | 70 | 40 | 0.32 |
| 592 | CCTN{4}GGC | 64 | 35 | 0.32 |
| 593 | CCGN{11}AGC | 22 | 7 | 0.32 |
| 594 | CACN{4}CGC | 18 | 5 | 0.32 |
| 595 | CCGN{1}CCC | 30 | 12 | 0.31 |
| 596 | CTGN{13}GGC | 73 | 42 | 0.31 |
| 597 | CGCN{16}ACC | 15 | 4 | 0.31 |
| 598 | CACN{18}CAG | 79 | 46 | 0.31 |
| 599 | GGCN{8}GCC | 68 | 29 | 0.29 |
| 600 | GGGN{15}GGA | 78 | 46 | 0.29 |
| 601 | CCGN{16}GCC | 22 | 7 | 0.29 |
| 602 | CCGN{20}ACC | 18 | 5 | 0.29 |
| 603 | CGAN{7}CCC | 17 | 5 | 0.28 |
| 604 | CCGN{6}CTC | 23 | 8 | 0.28 |
| 605 | CGGN{10}CTC | 22 | 7 | 0.28 |
| 606 | CAGN{16}CGC | 23 | 8 | 0.28 |
| 607 | CCAN{3}AGG | 77 | 45 | 0.27 |
| 608 | GCCN{18}GCC | 52 | 27 | 0.27 |
| 609 | CGCN{18}GGA | 19 | 6 | 0.26 |
| 610 | CCGN{20}GGC | 22 | 7 | 0.26 |
| 611 | ACAN{10}GCG | 17 | 5 | 0.26 |

| | | | | |
|-----|-------------|----|----|------|
| 612 | CGGN{5}CCC | 25 | 9 | 0.25 |
| 613 | CCCN{7}TCC | 75 | 43 | 0.25 |
| 614 | ACGN{10}CGC | 10 | 2 | 0.25 |
| 615 | CCCN{3}TCC | 81 | 48 | 0.25 |
| 616 | CCGN{8}CGG | 20 | 3 | 0.24 |
| 617 | CCAN{15}CGG | 22 | 7 | 0.24 |
| 618 | CCGN{6}CCG | 17 | 5 | 0.24 |
| 619 | CAGN{3}GCG | 25 | 9 | 0.24 |
| 620 | GAGN{1}CCC | 62 | 34 | 0.24 |
| 621 | CCGN{18}TGC | 22 | 7 | 0.23 |
| 622 | CCCN{7}CCA | 85 | 51 | 0.23 |
| 623 | CGGN{3}CCA | 24 | 9 | 0.23 |
| 624 | ACGN{1}CCC | 18 | 5 | 0.23 |
| 625 | CGGN{13}TGA | 21 | 7 | 0.22 |
| 626 | CTCN{6}GGC | 53 | 28 | 0.22 |
| 627 | GCGN{2}GAC | 15 | 4 | 0.22 |
| 628 | GGGN{11}ACC | 49 | 25 | 0.22 |
| 629 | CGCN{4}GGA | 17 | 5 | 0.22 |
| 630 | CCCN{11}CCG | 27 | 10 | 0.22 |
| 631 | CCGN{19}GCA | 20 | 6 | 0.22 |
| 632 | GCGN{0}GCA | 20 | 6 | 0.21 |
| 633 | AGAN{7}CCC | 61 | 33 | 0.21 |
| 634 | CGGN{2}CCA | 21 | 7 | 0.21 |
| 635 | CCCN{7}CCC | 89 | 54 | 0.21 |
| 636 | ACCN{4}GCG | 15 | 4 | 0.2 |
| 637 | CCTN{15}CGC | 20 | 6 | 0.2 |
| 638 | AGCN{9}GTC | 44 | 21 | 0.2 |
| 639 | CCCN{18}CTC | 74 | 43 | 0.2 |
| 640 | CGCN{18}CGA | 9 | 1 | 0.19 |
| 641 | CCCN{15}GCC | 62 | 34 | 0.18 |
| 642 | ACCN{11}GGC | 45 | 22 | 0.18 |
| 643 | AGGN{15}CGC | 29 | 12 | 0.18 |
| 644 | GCGN{0}CCA | 27 | 10 | 0.18 |
| 645 | GCGN{9}AGC | 18 | 5 | 0.17 |
| 646 | GGGN{18}GCA | 59 | 32 | 0.17 |
| 647 | CCCN{17}CAG | 77 | 45 | 0.17 |
| 648 | CCAN{8}CGG | 22 | 8 | 0.16 |
| 649 | CCGN{10}GGC | 21 | 7 | 0.16 |
| 650 | GCAN{0}GCC | 76 | 44 | 0.16 |
| 651 | CAGN{2}CGC | 20 | 6 | 0.16 |
| 652 | CGCN{8}GGC | 19 | 6 | 0.16 |
| 653 | CTGN{17}GGC | 65 | 36 | 0.16 |

| | | | | |
|-----|-------------|----|----|------|
| 654 | GGGN{14}ACC | 46 | 23 | 0.16 |
| 655 | CCGN{1}TGC | 20 | 6 | 0.16 |
| 656 | CAGN{8}CGC | 22 | 8 | 0.15 |
| 657 | AAGN{11}CGC | 17 | 5 | 0.15 |
| 658 | CCGN{6}TCC | 22 | 8 | 0.14 |
| 659 | CCAN{18}CCC | 72 | 42 | 0.14 |
| 660 | CCAN{0}CCC | 84 | 51 | 0.14 |
| 661 | GAGN{6}CCC | 53 | 28 | 0.14 |
| 662 | AGCN{20}GGC | 52 | 27 | 0.14 |
| 663 | CAGN{0}CGC | 21 | 7 | 0.14 |
| 664 | CCGN{12}CTC | 22 | 8 | 0.14 |
| 665 | CGCN{15}ACG | 9 | 1 | 0.13 |
| 666 | GGCN{17}CGA | 15 | 4 | 0.13 |
| 667 | CCGN{16}AAG | 19 | 6 | 0.13 |
| 668 | CGCN{14}TCC | 19 | 6 | 0.12 |
| 669 | AGGN{7}CGC | 20 | 7 | 0.12 |
| 670 | CGGN{7}CCC | 22 | 8 | 0.12 |
| 671 | CGCN{4}GCC | 34 | 15 | 0.12 |
| 672 | CGAN{6}CCC | 17 | 5 | 0.12 |
| 673 | CCCN{19}GGA | 60 | 33 | 0.11 |
| 674 | CCCN{16}GCG | 28 | 11 | 0.11 |
| 675 | CCAN{7}CGC | 20 | 7 | 0.11 |
| 676 | CCCN{6}GCC | 80 | 48 | 0.11 |
| 677 | GCCN{14}TCC | 55 | 29 | 0.11 |
| 678 | AGGN{14}GCC | 64 | 36 | 0.1 |
| 679 | CGCN{11}GCC | 20 | 7 | 0.1 |
| 680 | TCCN{0}GCA | 17 | 5 | 0.09 |
| 681 | GCGN{8}CCC | 27 | 11 | 0.09 |
| 682 | CCAN{11}GCG | 19 | 6 | 0.09 |
| 683 | CACN{4}GGG | 51 | 26 | 0.09 |
| 684 | CGGN{7}TCC | 20 | 7 | 0.09 |
| 685 | GCGN{5}GCC | 20 | 7 | 0.09 |
| 686 | ACGN{12}CAG | 26 | 10 | 0.09 |
| 687 | CCGN{19}CGC | 14 | 4 | 0.08 |
| 688 | CGGN{8}TGC | 18 | 5 | 0.08 |
| 689 | CCCN{1}GAG | 65 | 37 | 0.07 |
| 690 | GCGN{19}TGA | 18 | 6 | 0.07 |
| 691 | GGCN{15}GCC | 70 | 31 | 0.07 |
| 692 | CCGN{7}CCC | 27 | 11 | 0.07 |
| 693 | ACAN{19}CCC | 63 | 35 | 0.07 |
| 694 | ACCN{16}GGG | 47 | 24 | 0.07 |
| 695 | AGAN{1}GGC | 64 | 36 | 0.07 |

| | | | | |
|-----|-------------|----|----|------|
| 696 | GGGN{17}TGA | 64 | 36 | 0.06 |
| 697 | CAGN{5}GGG | 83 | 50 | 0.06 |
| 698 | GCCN{13}CGC | 22 | 8 | 0.06 |
| 699 | GCGN{7}GGA | 19 | 6 | 0.06 |
| 700 | CAGN{14}CCA | 94 | 58 | 0.06 |
| 701 | CCGN{4}GTC | 16 | 4 | 0.06 |
| 702 | CCCN{13}CGC | 22 | 8 | 0.06 |
| 703 | GCGN{14}ACC | 15 | 4 | 0.05 |
| 704 | CAGN{20}GGG | 81 | 49 | 0.05 |
| 705 | CCGN{4}CCC | 27 | 11 | 0.05 |
| 706 | CGCN{5}GGC | 18 | 6 | 0.05 |
| 707 | CCTN{6}GGC | 57 | 31 | 0.05 |
| 708 | AGGN{3}GGC | 67 | 38 | 0.05 |
| 709 | CGGN{11}CGC | 14 | 4 | 0.05 |
| 710 | CTGN{18}GGA | 77 | 46 | 0.04 |
| 711 | CACN{17}CCA | 74 | 43 | 0.04 |
| 712 | CGGN{3}GAG | 22 | 8 | 0.04 |
| 713 | CCCN{9}CCA | 82 | 49 | 0.03 |
| 714 | CCCN{1}ACG | 18 | 6 | 0.03 |
| 715 | CAGN{1}GCC | 72 | 42 | 0.03 |
| 716 | AGGN{6}CCG | 23 | 8 | 0.03 |
| 717 | AGCN{9}GGG | 57 | 31 | 0.03 |
| 718 | CCCN{7}GGC | 54 | 29 | 0.02 |
| 719 | CCTN{13}CCC | 88 | 54 | 0.02 |
| 720 | CCGN{19}TTC | 20 | 7 | 0.02 |
| 721 | CCCN{7}CCG | 27 | 11 | 0.02 |
| 722 | CGAN{6}GGC | 17 | 5 | 0.01 |
| 723 | CGGN{4}CTC | 21 | 7 | 0.01 |
| 724 | CGGN{0}CGC | 13 | 3 | 0.01 |
| 725 | CCTN{13}ACG | 19 | 6 | 0.01 |
| 726 | GGGN{6}CAC | 53 | 28 | 0.01 |
| 727 | CCCN{16}CGC | 21 | 7 | 0.01 |
| 728 | CCCN{10}CTC | 76 | 45 | 0 |
| 729 | CCCN{0}CAG | 92 | 57 | 0 |
| 730 | GCCN{5}CCC | 65 | 37 | 0 |

【 0 2 0 1 】

【表11】

表11. 線形判別式分析によって試験された候補 STAR エレメント

| 候補 STAR | 位置 ¹ | 長さ |
|---------|-----------------|-----------------|
| T2 F | 20q13.33 | ~2800 |
| T2 R | 6q14.1 | ~2800 |
| T3 F | 15q12 | ~2900 |
| T3 R | 7q31.2 | ~2900 |
| T5 F | 9q34.13 | ND ² |
| T5 R | 9q34.13 | ND |
| T7 | 22q12.3 | ~1200 |
| T9 F | 21q22.2 | ~1600 |
| T9 R | 22q11.22 | ~1600 |
| T10 F | 7q22.2 | ~1300 |
| T10 R | 6q14.1 | ~1300 |
| T11 F | 17q23.3 | ~2000 |
| T11 R | 16q23.1 | ~2000 |
| T12 | 4p15.1 | ~2100 |
| T13 F | 20p13 | ~1700 |
| T13 R | 1p13.3 | ~1700 |
| T14 R | 11q25 | ~1500 |
| T17 | 2q31.3 | ND |
| T18 | 2q31.1 | ND |

¹ 染色体の位置は、ヒトゲノムデータベースに対するSTARエレメントからのDNA配列データのBLAST調査によって決定される。該位置は、各染色体の細胞遺伝学的イデオグラムと言及される標準命名法に従い与えられる。例えば1p2.3は1番染色体の短いアーム(arm)の第2の細胞遺伝学的バンドの第3の細胞遺伝学的サブバンドである(<http://www.ncbi.nlm.nih.gov/Class/MLACourse/Genetics/chrombanding.html>)。正の(forward)及び逆の(reverse)シーケンシング反応が種々のゲノム座からのDNAを識別した場合、両方の座が示される。

F, 正の(Forward)配列反応結果; R, 逆の(Reverse)配列反応結果。

正の及び逆のシーケンス結果は、種々のゲノム位置にマップされ、各配列は、ヒトゲノムデータベースからの配列情報に基づき(制限マッピングによって決定されたように)オリジナルのクローニングの全長にまで拡張された。

² ND: 決められていない(not determined)。

【0202】

【表12】

表12. 染色体位置及び長さを含む、本発明のアラビドプシス
STARエレメント

| STAR | 染色体 | 長さ, kb |
|------|-----|--------|
| A1 | I | 1.2 |
| A2 | I | 0.9 |
| A3 | I | 0.9 |
| A4 | I | 0.8 |
| A5 | I | 1.3 |
| A6 | I | 1.4 |
| A7 | II | 1.2 |
| A8 | II | 0.8 |
| A9 | II | 0.9 |
| A10 | II | 1.7 |
| A11 | II | 1.9 |
| A12 | II | 1.4 |
| A13 | II | 1.2 |
| A14 | II | 2.1 |
| A15 | II | 1.4 |
| A16 | II | 0.7 |
| A17 | II | 1.5 |
| A18 | III | 1.5 |
| A19 | III | 0.7 |
| A20 | III | 2.0 |
| A21 | IV | 1.8 |
| A22 | IV | 0.8 |
| A23 | IV | 0.6 |
| A24 | IV | 0.5 |
| A25 | V | 0.9 |
| A26 | V | 1.9 |
| A27 | V | 1.1 |
| A28 | V | 1.6 |
| A29 | V | 0.9 |
| A30 | V | 2.0 |
| A31 | V | 2.0 |
| A32 | V | 1.3 |

| | | |
|-----|----|-----|
| A33 | V | 0.9 |
| A34 | I | 0.9 |
| A35 | II | 1.1 |

【図面の簡単な説明】

【0203】

【図1】STARエレメントを選択し且つ特徴付けるためのプラスミドのpSelectファミリー。ショウジョウバエSV40プロモーターの制御下の耐性マーカー（ゼオシン又はピューロマイシン）又はレポーター遺伝子（GFP又はルシフェラーゼ）がAscl及びHindIII部位によって隣接されるBamHIクローニング部位に近接している。クローニング部位の上流にlexAオペレーターがあり、lexAタンパク質がそれに結合できる。オペレーターへのキメラlexA-ポリコームグループタンパク質の結合は、マーカー又はレポーター遺伝子の抑制をもたらす。抑制をブロックするクローニング部位に挿入されたDNAフラグメントは、マーカー又はレポーター遺伝子の永続的な発現によって識別される。このプラスミドは、培養された哺乳動物細胞内でoriP配列によりエピソーム的に複製する。

【図2】STARエレメントをテストするためのプラスミドのpSDHファミリー。2つのマルチクローニング部位（MCSI及びMCSII）は、発現が上流プロモーター（CMV、Tet-Off、またはSV40）によって駆動されるところのレポーター遺伝子（GFP又はルシフェラーゼ）に隣接する。テストされるべきSTARエレメントは、MCSI及びMCSIIで挿入される。これらは、特有の制限部位（MCSI: Xhol、NotI、EcoRI及びSalI；MCSII:HindIII、EcoRV、BglII及びNheI）を含む。該プラスミドは哺乳動物細胞のゲノムにおいてランダムに組み込まれた後に複製する。

【図3】ルシフェラーゼを過剰発現するクローンの割合。U-2 OSヒト骨肉種細胞は、（Tet-Offプロモーターの制御下でルシフェラーゼプロモーター遺伝子を含む）pSDHプラスミドで安定にトランスフェクションされ、及び個々のトランスフェクションされたクローンは単離され且つ培養された。ルシフェラーゼ発現は、酵素的に測定された。STARなし（STAR less）pSDHを含むクローンによる平均ルシフェラーゼ発現（「参照レベル」）が決定された。全てのプラスミドのためのセットからのクローンは、それらのルシフェラーゼ活性が前記参照レベルより2倍より高い場合に、「過剰発現（over-expressing）」としてスコアされる。各プラスミドにおける過剰発現するクローンのパーセントがプロットされる。

【図4】過剰発現するクローンによる過剰発現倍数。ゲノムDNA内に組み込まれたSTAR含有pSDHプラスミドにおける過剰発現の範囲は、参照レベルによって各クローンのルシフェラーゼ活性を割算することによって決定された。有意な発現（上記参照レベルの2倍より多い）を示すそれらのために、真の倍数増加が記載された；これらのデータの最小値及び中央値が各プラスミドのためにプロットされる。

【図5】過剰発現するクローンによる過剰発現倍数。ゲノムDNA内に組み込まれたSTAR含有pSDHプラスミドにおける過剰発現の範囲は、参照レベルによって各クローンのルシフェラーゼ活性を割算することによって決定された。有意な発現（上記参照レベルの2倍より多い）を示すそれらのために、真の倍数増加が記載された；これらのデータの最小値及び中央値が各プラスミドのためにプロットされる。

【図6】SINCエレメントを選択し且つ特徴付けるためのpSS（SINC-Select）プラスミド。codA::upp自殺遺伝子は、プロドラッグ5'-フルオロシトシンを毒性薬物5'-フルオロウラシルに転化するタンパク質をコードする。低くされたテトラサイクリン濃度による導出に応答し、宿主細胞はプロドラッグに感受性になる。サイレンシング活性を有するクローニング部位（BglII及びXhol）で挿入されたゲノムDNAフラグメントは、自殺遺伝子の発現を防止し且つプロドラッグ耐性コロニーの形成を可能にする。STARエレメントは、ブ

プラスミドの機能的コンポーネントへのサイレンシングされたクロマチンの拡散を防ぐために選択コンポーネントに隣接する。このプラスミドは、培養された哺乳動物細胞内でoriP配列によりエピソーム的に複製する。

【図7】STAR活性を試験するために使用されるpSDH-CSPプラスミド。SEAP (Secreted Alkaline Phosphatase) レポーター遺伝子は、CMVプロモーターの制御下にあり、且つピュロマイシン耐性選択可能なマークター (puro) は、SV40プロモーターの制御下にある。STARエレメントがその中でクローニングされうるマルチクローニング部位が、これら2つの遺伝子に隣接している。該プラスミドは、大腸菌内の増殖のために複製のオリジン (ori) 及びアンピシリン耐性遺伝子 (ampR) をまた有する。

【図8】STAR6及びSTAR49は、予測可能性及び導入遺伝子発現の収率を改善する。pSDH-CSP、pSDH-CSP-STAR6、又はpSDH-CSP-STAR49でトランスフェクションされたCHO細胞によるCMVプロモーターからのSEAPの発現が決定された。STAR含有構築物は、pSDH-CSP構築物のみと比較してより大きい予測可能性及び高い収率を与える。

【図9】STAR6及びSTAR8は、予測可能性及び導入遺伝子発現の収率を改善する。pSDH-CMV、pSDH-CMV-STAR6、又はpSDH-CMV-STAR8でトランスフェクションされたU-2 OS細胞によるCMVプロモーターからのルシフェラーゼの発現が決定された。STAR含有構築物は、pSDH-CMV構築物のみと比較してより大きい予測可能性及び高い収率を与える。

【図10】STAR10及びSTAR27の最小必須配列。STARエレメントの部分は、PCRによって増幅された：STAR10は、フラグメント10Aを与えるためにプライマーE23及びE12で、フラグメント10Bを与えるためにプライマーE13及びE14で、及びフラグメント10Cを与えるためにプライマーE15及びE16で増殖された。STAR27は、フラグメント27Aを与えるためにプライマーE17及びE18で、フラグメント27Bを与えるためにプライマーE19及びE20で、及びフラグメント27Cを与えるためにプライマーE21及びE22で増殖された。これらサブフラグメントは、pSelectベクター内にクローニングされた。U-2 OS/Tet-Off/LexA-HP1細胞内にトランスフェクションした後、ゼオシン存在下における培養の成長が観察された。成長速度は、活発な (+++) から悪い (+/-) と様々であり、一方ある培養は、試験されたDNAフラグメントにおけるSTAR活性の不存在によりゼオシン処理で生存することに失敗した (-)。

【図11】SV40プロモーターとの関連におけるSTARエレメント機能。pSDH-SV40及びpSDH-SV40-STAR6は、ヒト骨肉腫内U-2 OS細胞株にトランスフェクションされ、且フルシフェラーゼの発現はピュロマイシン耐性クローンにおいてSTAR6による遺伝子サイレンシングからの保護あり又はなしでアッセイされた。

【図12】Tet-Offプロモーターとの関連におけるSTARエレメント機能。pSDH-Tet及びpSDH-Tet-STAR6は、ヒト骨肉腫内U-2 OS細胞株にトランスフェクションされ、且フルシフェラーゼの発現はピュロマイシン耐性クローンにおいてSTAR6による遺伝子サイレンシングからの保護あり又はなしでアッセイされた。

【図13】STARエレメントの配向の説明図であって、それらがpSelectベクター内にクローニングされる場合 (パネルA)、それらがそれらの自然の配向を保存するためにpSDHベクター内にクローニングされる場合 (パネルB) 及び、それらが逆の配向においてpSDHベクター内にクローニングされる場合 (パネルC)。

【図14】STAR66機能の方向性。STAR66エレメントは、自然 (STAR66 native) 又は反対の配向 (STAR66 opposite) のいずれかでpSDH-Tet内にクローニングされ、且つU-2 OS細胞内にトランスフェクションされた。ルシフェラーゼ活性は、ピュロマイシン耐性クローンでアッセイされた。

【図15】STAR機能のコピー数依存性。U-2 OSゲノムDNA内に組み込まれた、pSDH-Tet-STAR10内のルシフェラーゼ発現ユニットのサザンプロット法。放射活性ルシフェラーゼDNAプローブは、各クローンのゲノムにおける導入遺伝子DNAの量を検出するために使用され、その後それはフォスファーイメージヤー (phosphorimager) で数値化された。

【図16】STAR機能のコピー数依存性。各クローンにおけるpSDH-Tet-STAR10発現ユニットのコピー数は、フォスファーイメージヤーによって決定され、且つ各クローンによって

発現されたルシフェラーゼレポーター酵素の活性と比較される。

【図17】エンハンサープロッキング及びエンハンサー・アッセイ。エンハンサー・プロッキング及びエンハンサー活性のためのSTARをテストするために使用されるルシフェラーゼ発現ベクターが、概略的に示される。E47エンハンサータンパク質のためのE-Box結合部位は、STARエレメントのためのクローニング部位の上流にある。STARクローニング部位の下流に、ヒトアルカリリコット・プロモーター(mp)の制御下のルシフェラーゼ遺伝子がある。ヒストグラムは、3つの可能性のある実験状況(本文参照)のための予想された結果を示している。パネルA: エンハンサー・プロッキングアッセイ。パネルB: エンハンサー・アッセイ。

【図18】エンハンサー・プロッキングアッセイ。最小プロモーターからのルシフェラーゼ発現は、エンプティベクター(ベクター)におけるE47/E-boxエンハンサーによって活性化される。エンハンサー・プロッカー(SCS, HS4)又はSTARエレメント(STARエレメント1, 2, 3, 6, 10, 11, 18及び27)の挿入は、E47/E-boxエンハンサーによるルシフェラーゼ活性をブロックする。

【図19】エンハンサー・アッセイ。最小プロモーターからのルシフェラーゼ発現は、エンプティベクター(E47)におけるE47/E-boxエンハンサーによって活性化される。SCS及びHS4エレメント又は種々のSTARエレメント(STARエレメント1, 2, 3, 6, 10, 11, 18及び27)の挿入は、レポーター遺伝子の転写を活性化しない。

【図20】SINCエレメントの分離のために使用されるpSS-codA::uppベクター。codA::upp自殺遺伝子は、プロドラッグ5-フルオロシトシンを毒性薬物5-フルオロウラシルに転化するタンパク質をコードする。低くされたドキシサイクリン濃度による導出に応答して、宿主細胞はプロドラッグに感受性になる。サイレンシング活性を有するBgIIIクローニング部位で挿入されたゲノムDNAフラグメントは、自殺遺伝子の発現を防止し且つプロドラッグ耐性コロニーの形成を可能にする。STARエレメントは、プラスミドの機能的コンポーネントへのサイレンシングされたクロマチンの拡散を防止するために選択コンポーネントに隣接する。プラスミドは、ハイグロマイシン耐性遺伝子を有する哺乳動物細胞内にトランسفエクション後、及びアンピシリン耐性遺伝子を有する大腸菌細胞内にトランسفエクション後に選択される。それは、培養された哺乳動物細胞内でoriP及びEBNA-1配列により、及び大腸菌内でori配列によりエピソーム的に複製する。

【図21】pSS-hrGFPプラスミドは、自殺遺伝子を(緑色蛍光タンパク質をコードする)hrGFPでの置換及びSTAR6をGFPレポーター遺伝子の下流でSTAR8での置換を除いてpSS-codA::Auppプラスミドに同一である。

【図22】マウスとヒトとの間のSTAR18配列保存。497個の塩基対STAR18を含むヒトゲノムの領域が示される(黒いボックス);該エレメントは、ヒト2番染色体上のHOXD8及びHOXD4ホメオボックス遺伝子の間に生じる。それは、72%配列同一性を共有するマウス2番染色体における領域と並べられる。STAR18のすぐ左のヒト2番染色体の領域は、マウス2番染色体でまた高く保存されている(73%同一性;グレー・ボックス);これらの領域を超えると、該同一性は60%以下に落ちる。別々に又は組み合わせてのいずれかでゼオシンでの成長を与えるヒト及びマウス由来のこれらの領域の能力が示される;- 成長しない; + 普通の成長; ++ 活発な成長、+++ 迅速な成長。

【図23】バイオインフォマティクス分析ワークフローの説明図。詳細には、本文を参照。

【図24】65個のSTARエレメントのトレーニングセットの分類に対する判別式分析の結果。ステップワイズ線形判別式分析(Linear Discriminant Analysis; LDA)によってSTARとして正確に分類されるSTARエレメントは、ベン(Venn)ダイアグラム内に示される。LD Aの変異体は、6量体のオリゴヌクレオチド(「オリゴ; oligos」)のための及び二分染色体のための頻度分析結果から選択された。該ダイアグラムは、STARを正確に分類する2つのセットの変異体の一一致を示す。

【図25】U-2 OS/Tet-Off/IlexA-HP1細胞が、候補シロイヌナズナSTARエレメントでトランسفエクションされかつ低いドキシサイクリン濃度で培養された。総RNAは単離され且

つRT-PCRに付された；ゼオシン及びハイグロマイシン耐性mRNAに対応するバンドは、サザンプロット法によって検出され、且つフォスファーイメージヤーで数値化された。ハイグロマイシンシグナルに対するゼオシンシグナルの割合は、12の種々のシロイヌナズナSTARエレメント、ショウジョウバエscsエレメントで隣接された又は隣接エレメントを有しないゼオシン発現ユニットを含む形質転換体(transfectants)のために示される。

【図26】STAR1 - STAR65を含む配列(配列ID(SEQ ID):1-65)STAR66及びテストセットを含む配列(配列ID:66-84)シロイヌナズナSTAR A1-A35を含む配列(配列ID:85-119)

)
【誤訛訂正3】

【訂正対象書類名】図面

【訂正対象項目名】図13

【訂正方法】変更

【訂正の内容】

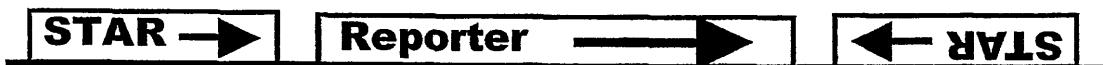
【図13】

FIG 13 STAR エレメント配向

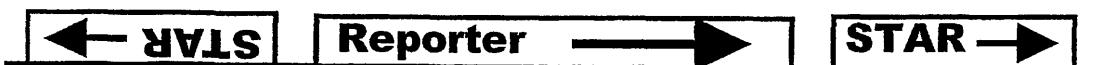
A. クローニングされたSTARエレメントを有するpSelectベクター：



B. pSDHベクター、自然の配向における複数のSTAR：



C. pSDHベクター、逆の配向における複数のSTAR：



【誤訳訂正4】

【訂正対象書類名】図面

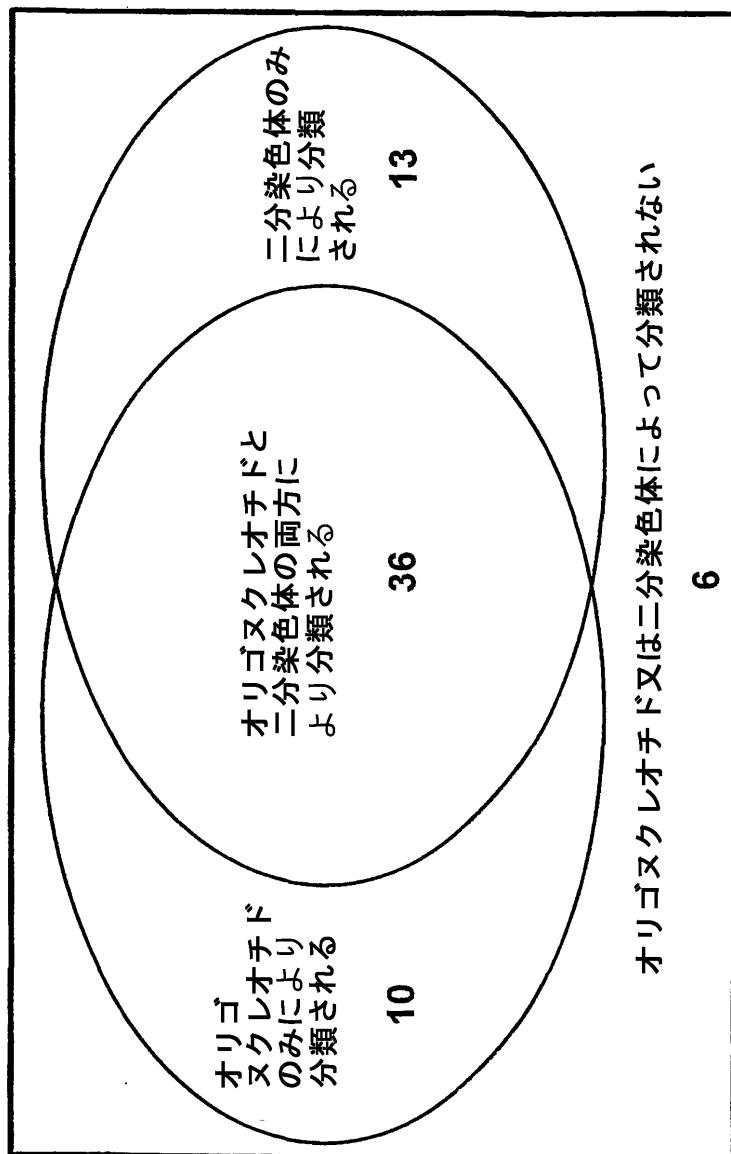
【訂正対象項目名】図24

【訂正方法】変更

【訂正の内容】

【図24】

Fig 24
オリゴスクレオチド及び二分染色体モデルでの
判別式分析によるSTARの分類



65個のSTARのトレーニングセット

【誤訳訂正5】

【訂正対象書類名】図面

【訂正対象項目名】図25

【訂正方法】変更

【訂正の内容】

【図25】

FIG 25
シロイヌナズナSTAR強度のRT-PCRアッセイ

