

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号
特許第4293634号
(P4293634)

(45) 発行日 平成21年7月8日 (2009.7.8)

(24) 登録日 平成21年4月17日 (2009.4.17)

(51) Int. Cl.

F I

C 1 2 Q 1/68 (2006.01)

C 1 2 N 15/09 (2006.01)

C 1 2 Q 1/68 A

C 1 2 Q 1/68 Z

C 1 2 N 15/00 Z N A A

請求項の数 10 (全 50 頁)

(21) 出願番号	特願平9-501818	(73) 特許権者	598114930
(86) (22) 出願日	平成8年6月6日 (1996.6.6)		ソレクサ・インコーポレイテッド
(65) 公表番号	特表平11-507528		S o l e x a , I n c .
(43) 公表日	平成11年7月6日 (1999.7.6)		アメリカ合衆国 カリフォルニア 945
(86) 国際出願番号	PCT/US1996/009513		45, ヘイワード, インダストリアル プ
(87) 国際公開番号	W01996/041011		ールバード 25861
(87) 国際公開日	平成8年12月19日 (1996.12.19)		3832 Bay Center Pla
審査請求日	平成12年3月3日 (2000.3.3)		ce, Hayward, Califo
審査番号	不服2005-17398 (P2005-17398/J1)		rn i a 94545 U. S. A.
審査請求日	平成17年9月9日 (2005.9.9)	(74) 代理人	100094318
(31) 優先権主張番号	08/478, 238		弁理士 山田 行一
(32) 優先日	平成7年6月7日 (1995.6.7)	(74) 代理人	100123995
(33) 優先権主張国	米国 (US)		弁理士 野田 雅一
(31) 優先権主張番号	PCT/US95/12791	(74) 代理人	100107456
(32) 優先日	平成7年10月12日 (1995.10.12)		弁理士 池田 成人
(33) 優先権主張国	カナダ (CA)		最終頁に続く

(54) 【発明の名称】 分類および同定のためのオリゴヌクレオチドタグ

(57) 【特許請求の範囲】

【請求項 1】

D N A がロードされた微粒子と称する、微粒子に結合したポリヌクレオチドのライブラリー
の平行配列決定法であって、

(i) 各ポリヌクレオチドにオリゴヌクレオチドタグを結合し、タグーポリヌクレオチド
結合体のサンプルを提供する工程、

(i i) 前記タグーポリヌクレオチド結合体に捕捉部分を結合させる工程、

(i i i) 前記タグーポリヌクレオチド結合体を微粒子にロードさせる工程であって、
各微粒子の表面には 1 種類のタグ相補物が結合され、微粒子上に結合されたタグ相補物と
前記タグーポリヌクレオチド結合体とを相互作用させることにより、D N A がロードされ
た微粒子が形成され、

前記 D N A がロードされた微粒子は、D N A がロードされた微粒子及びロードされてい
ない微粒子の双方を含むより多数の微粒子の中に含まれるように、異なる前記タグ相補物
の数は、タグーポリヌクレオチド結合体の数の少なくとも 10 倍である工程、

(i v) 前記捕捉部分が捕捉表面に結合する選別工程に前記より多数の微粒子を供し、そ
れによって、前記 D N A がロードされた微粒子を前記ロードされていない微粒子から分離
する工程、

(v) 前記捕捉表面から前記 D N A がロードされた微粒子を離す工程、

(v i) 工程 (i) - (v) を新しいタグーポリヌクレオチド結合体のサンプルとタグ相
補物を含む微粒子を用いて繰り返す工程、

(v i i) それぞれの工程 (i) - (v) において得られたDNAがロードされた微粒子を保存する工程、及び、

(v i i i) 前記保存したDNAがロードされた微粒子上のポリヌクレオチドを平行して同時に配列決定される工程、

を含む方法。

【請求項2】

前記捕捉部分がビオチンである、請求項1記載の方法。

【請求項3】

前記捕捉表面がアビジン化ビーズである、請求項2記載の方法。

【請求項4】

前記アビジン化ビーズが磁気化されている、請求項3記載の方法。

【請求項5】

増幅時にビオチン化プライマーによって前記ビオチンが前記ポリヌクレオチドに結合する、請求項2記載の方法。

【請求項6】

各DNAがロードされた微粒子が一種類のみのポリヌクレオチドに結合している、請求項1記載の方法。

【請求項7】

前記工程 (v i i) の後にDNAがロードされた微粒子が平面上に広げられる、請求項1記載の方法。

【請求項8】

前記平行配列決定が、処理及び検出の連続サイクルを通じて段階的なヌクレオチドの同定によって行われる、請求項1記載の方法。

【請求項9】

前記処理が標識したオリゴヌクレオチドプローブとのテンプレートの伸長を含む、請求項8記載の方法。

【請求項10】

前記オリゴヌクレオチドプローブが蛍光標識されている、請求項9記載の方法。

【発明の詳細な説明】

発明の分野

本発明は、一般に、オリゴヌクレオチドタグを用いて、分子（特に、ポリヌクレオチド）を同定、分類、および／または探知するための方法に関し、より詳細には、タグのその相補物への特異的なハイブリダイゼーションによりそのようなタグ化ポリヌクレオチドを分類および分析する方法に関する。

背景

オリゴヌクレオチドおよびそのアナログの特異的なハイブリダイゼーションは、広く多様な研究的適用、医学的適用および産業的適用で用いられている基礎的なプロセスである。これは、診断アッセイにおける疾患関連のポリヌクレオチドの同定、新規の標的ポリヌクレオチドのクローニングのためのスクリーニング、ポリヌクレオチド混合物のプロットにおける特異的ポリヌクレオチドの同定、特異的な標的ポリヌクレオチドの増幅、不適切に発現された遺伝子の治療的ブロック、DNA配列決定などを含む（例えば、Sambrookら、Molecular Cloning: A Laboratory Manual, 第2版 (Cold Spring Harbor Laboratory, New York, 1989) ; KellerおよびManak、DNA Probes, 第2版 (Stockton Press, New York, 1993) ; Milliganら、J. Med. Chem. , 36:1923-1937 (1993) ; Drmanacら、Science, 260:1649-1652 (1993) ; Bains, J. DNA Sequencing and Mapping, 4:143-150 (1993)) 。

特異的ハイブリダイゼーションはまた、オリゴヌクレオチドタグで標識された化合物を探知、回収、および同定する方法として提案されている。例えば、複数のDNA配列決定において、オリゴヌクレオチドタグを用いて同じ配列決定反応において生じるDNAフラグメントからなる、ゲル上の電気泳動的に分離されたバンドを同定する。このように、多くの配列決定反応由来のDNAフラグメントがゲルの同じレーンで分離され、これは、次いで異なる

10

20

30

40

50

る固相物質でプロットされる。そこでは、異なる配列決定反応由来のフラグメントバンドが、相補的なタグに特異的にハイブリダイズするオリゴヌクレオチドプローブで可視化される (Churchら、Science, 240:185-188 (1988))。オリゴヌクレオチドタグの同様の使用はまた、爆発物、潜在的な汚染物質 (例えば、原油)、および模造品の防止および検出のために通貨を同定するために提供されてきた (例えば、Dollinger、265-274頁、Mullisら編、The Polymerase Chain Reaction (Birkhauser, Boston, 1994) に概説された)。さらに最近、オリゴヌクレオチドタグを用いるシステムはまた、複雑なコンビナトリアル (combinatorial) ケミカルライブラリー中の個々の分子を操作および同定する方法として (例えば、このような薬物候補物のライブラリーのスクリーニングを助けるために) 提案されている (BrennerおよびLerner. Proc.Natl.Acad.Sci.、89:5381-5383 (1992) : Alper. Science、264:1399-1401 (1994) : およびNeedelsら、Proc.Natl.Acad.Sci.、90:10700-10704 (1993))。

このようなタグ化スキームをうまく実施することは、タグとその相補的プローブとの間での特異的なハイブリダイゼーションの達成における成功に大部分依存する。すなわち、オリゴヌクレオチドタグが物質をうまく同定するためには、擬陽性および擬陰性のシグナルの数を最小限にしなければならない。不幸にも、このような偽のシグナルは希ではない。なぜなら塩基対形成および塩基の積み重ねの自由エネルギーは、二重鎖または三重鎖の構造におけるヌクレオチド間で広く変化するためである。例えば、デオキシアデニン (A) およびその相補物に結合したチミジン (T) の繰り返し配列からなる二重鎖は、ミスマッチを含む部分的に相補的な標的に結合したデオキシグアノシン (G) およびデオキシシチジン (C) の繰り返し配列からなる同じ長さの二重鎖より低い安定性を有し得る。従って、大きなコンビナトリアルケミカルライブラリー由来の所望の化合物が前者のオリゴヌクレオチドを用いてタグ化された場合、完全にマッチしたATに富む二重鎖を検出するように設計されたハイブリダイゼーション条件下では、GCに富むオリゴヌクレオチドを用いて標識された所望されない化合物 (ミスマッチした二重鎖においてさえ) が、ATに富むタグからなる完全にマッチした二重鎖とともに検出されるという重大な可能性が存在する。Brennerら (上記) により提案された分子タグ化システムでは、密接に関連するタグのミスハイブリダイゼーションの関連する問題は、いわゆる「無限 (comma-less)」コードを用いることにより処理され、これはその相補的タグについて登録をはずれた (またはフレームシフトした) プローブが、その5またはそれ以上の3塩基ワード、すなわち各「コドン」について1つ以上のミスマッチを有する二重鎖を生じることを保証する。

たとえば塩化テトラメチルアンモニウムのような試薬がオリゴヌクレオチド二重鎖の塩基特異的な安定性の相違を無効にするために利用可能であるとしても、このような試薬の効果はしばしば限定され、そしてその存在は、選択された化合物のさらなる操作 (例えば、ポリメラーゼ連鎖反応 (PCR) などによる増幅) を不可能にし得るか、またはより困難にし得る。

このような問題は、例えば多数回のPCR、リバースドットブロッキングなどを介する多数または複雑な遺伝子座の分析における多数のハイブリダイゼーションプローブの同時使用を非常に困難にする。その結果、HLA遺伝子のような特定の座の直接的な配列決定は、遺伝子型の同定のための特異的なハイブリダイゼーションを用いる間接的な方法に対する信頼し得る代替法として奨励される (例えば、Gyllensteinら、Proc.Natl.Acad.Sci.、85:7652-7656 (1988))。

異なる固相支持体上で、クローン化されかつ同一にタグ化されたDNAを分類する能力は、特に、並行して多くのサンプルに同時に適用し得る非ゲルベースの配列決定方法論と組み合わせられる場合、このような配列決定を容易にする。

上記を考慮すると、タグの多くのレパートリーを提供するが、また天然の塩基対形成および塩基の積み重ねの自由エネルギーの差を変化させるための特別な試薬を用いる必要はなく、擬陽性および擬陰性のシグナルの発生を最小限にするオリゴヌクレオチドベースのタグ化システムが利用可能である場合、これは有用である。このようなタグ化システムにより、コンビナトリアルケミカルライブラリーの構築および使用、DNAの大規模なマッピング

10

20

30

40

50

グおよび配列決定、遺伝子同定、医学診断などを含む多くの領域での適用が見い出される。

発明の要旨

本発明の目的は、化合物の探知、回収、および同定のための分子タグ化システムを提供することである。

本発明の別の目的は、オリゴヌクレオチドタグおよびその相補物の特異的ハイブリダイゼーションにより固相物質の表面上の同一の分子、または分子のサブクラス（特にポリヌクレオチド）を分類するための方法を提供することである。

本発明のさらなる目的は、疾患または正常な組織での遺伝子の発現パターンを分析するための方法を提供することである。

本発明のなおさらなる目的は、同時分析および／または配列決定用の標的ポリヌクレオチドの数千ものフラグメント（特にランダムに重複するフラグメント）をタグ化および分類するためのシステムを提供することである。

本発明の別の目的は、数百塩基対から数万塩基対の範囲の長さを有する標的ポリヌクレオチドを配列決定するための迅速かつ信頼できる方法を提供することである。

本発明のさらなる目的は、従来のサンガー法に基づく配列決定技術を用いる大規模配列決定プロジェクトにおいて必要とされる多くの異なるテンプレート調製工程を軽減するための方法を提供することである。

本発明は、オリゴヌクレオチドタグの使用により分子の種類または亜集団を探知、同定、および／または分類するための方法および物質を提供することにより、これらの目的および他の目的を達成する。本発明の重要な特徴は、オリゴヌクレオチドタグが最少に交差ハイブリダイズするセット（minimally cross-hybridizing set）のオリゴヌクレオチドのメンバーであることである。このようなセットのオリゴヌクレオチドの配列は、同じセットの他のどのメンバーの配列と少なくとも2ヌクレオチドだけ異なる。従って、このようなセットの各メンバーは、2つ未満のミスマッチを有する他のいずれのメンバーの相補物とも二重鎖（または三重鎖）を形成し得ない。本発明のオリゴヌクレオチドタグの相補物（本明細書中では、「タグ相補物」という）は、天然のヌクレオチドまたは非天然のヌクレオチドアナログを含み得る。好ましくは、タグ相補物は、固相支持体に結合される。このようなオリゴヌクレオチドタグは、その対応するタグ相補物と一緒に使用される場合、分子（特に、ポリヌクレオチド）を分類、探知、または標識するためのハイブリダイゼーションの特異性を増強する手段を提供する。

最少に交差ハイブリダイズするセットのオリゴヌクレオチドタグおよびタグ相補物は、コンビナトリアル的または個々のいずれかで合成され得、それは、所望されるセットのサイズ、および交差ハイブリダイゼーションを最少化すると考えられる程度（言い換えれば、特異性が増強されると考えられる程度）に依存する。例えば、最少に交差ハイブリダイズするセットは、少なくとも4ヌクレオチドだけ互いに異なる個々に合成された10マーの配列のセットからなり得る。このようなセットは、332の最大のサイズを有する（3種類のヌクレオチドからなり、そして付録Icに開示されるようなコンピュータプログラムを用いて計数した場合）。あるいは、最少に交差ハイブリダイズするセットのオリゴヌクレオチドタグはまた、最少に交差ハイブリダイズするセットからそれら自身が選択されるサブユニットからコンビナトリアル的に組み立てられ得る。例えば、少なくとも3ヌクレオチドだけ互いに異なる最少に交差ハイブリダイズする12マーのセットは、各々が3ヌクレオチドだけ互いに異なる最少に交差ハイブリダイズする4マーのセットから選択される3つのサブユニットを組み立てることにより合成され得る。このような実施態様では、最大限、 9^3 、すなわち729の12マーのサイズのセットが得られる。数字9は、付録Iaのコンピュータプログラムに列記されたオリゴヌクレオチドの数である。これは、10マーとして、4つの異なるタイプのヌクレオチドの3つのみを使用されると仮定する。セットは「最大」として記載される。なぜなら、付録Ia～Icのコンピュータプログラムは、任意の入力（例えば、メンバー間のヌクレオチドの数の長さ、組成、差異）に対して最大のセットを提供する。最少に交差ハイブリダイズするさらなるセットは、このように計算されたセッ

10

20

30

40

50

トのサブセットから形成され得る。

オリゴヌクレオチドタグは一本鎖であり得、そして二重鎖形成による一本鎖タグ相補物への特異的にハイブリダイゼーションのために、または三重鎖形成により二本鎖タグ相補物への特異的なハイブリダイゼーションのために設計され得る。オリゴヌクレオチドタグはまた、二本差であり得、そして三重鎖形成による一本鎖タグ相補物への特異的なハイブリダイゼーションのために設計され得る。

コンビナトリアル的に合成される場合、本発明のオリゴヌクレオチドタグは、好ましくは、複数のサブユニットからなる。各サブユニットは、長さが3～9ヌクレオチドのオリゴヌクレオチドからなる。ここで、各サブユニットは、最少に交差ハイブリダイズする同じセットから選択される。このような実施態様では、利用可能なオリゴヌクレオチドタグの数は、タグ当たりのサブユニットの数、およびそのサブユニットの長さに依存する。数は、一般に、タグの長さのすべての可能な配列の数よりはるかに少ない。タグについて、 n ヌクレオチドの長さは、 4^n である。

本発明の1つの局面では、固相支持体に結合したオリゴヌクレオチドタグの相補物が、各々がタグを含有するポリヌクレオチドの混合物からポリヌクレオチドを分類するために使用される。この実施態様では、オリゴヌクレオチドタグの相補物は、固相支持体（例えば、微小ビーズ、または単一の支持体上の合成位置の配置（array）の特定の位置）の表面上で合成される。その結果、同一配列の集団が、特定の領域で得られる。すなわち、各支持体（ビーズの場合）または各領域（配置の場合）の表面は、特定の配列を有する相補物の1つのタイプのみにより誘導体化される。このようなビーズまたは領域の集団は、異なる配列の相補物のレパートリーを含有する。オリゴヌクレオチドタグおよびタグ相補物に関連して本明細書中で使用される用語「レパートリー」は、特定の実施態様でのタグまたは対応するタグ相補物のセットを組み立てるオリゴヌクレオチドタグの、最少に交差ハイブリダイズするセットを意味する。

分類すべき各ポリヌクレオチドは、異なるポリヌクレオチドが異なるタグを有するように、結合したオリゴヌクレオチドタグを有する。下記により十分に説明するように、この条件は、ポリヌクレオチドの集団より実質的に大きいタグのレパートリーを用いることにより、そしてタグ化されたポリヌクレオチドの完全な集まり（ensemble）からタグ化されたポリヌクレオチドの十分に小さいサンプルを取ることにより達成される。このようなサンプリングの後、支持体およびポリヌクレオチドの集団が、オリゴヌクレオチドタグとその個々の相補物とを特異的にハイブリダイゼーションを可能にする条件下で混合される場合、同一のポリヌクレオチドは、特定のビーズまたは領域で分類される。ポリヌクレオチドの分類された集団は、次いで、微量生化学技術により固相支持体上で操作され得る。

一般的に、本発明の方法は以下の工程を包含する：（a）分子集団中の各分子にタグのレパートリー由来のオリゴヌクレオチドタグを結合させる工程であって、（i）集団中の実質的にすべての異なる分子または分子の異なる亜集団が異なる結合オリゴヌクレオチドタグを有し、そして（ii）レパートリー由来の各オリゴヌクレオチドタグが最少に交差ハイブリダイズするセットから選択される、工程；および（b）オリゴヌクレオチドタグを、1つ以上の固相支持体に結合させたその相補物のそれぞれと特異的にハイブリダイズさせることにより、そのような支持体上の集団の分子を分類する、工程。

本発明の重要な局面は、並行する配列決定のために、ポリヌクレオチドを分類するためのオリゴヌクレオチドタグの使用である。好ましくは、このような配列決定は以下の工程により行われる：（a）標的ポリヌクレオチドをカバーする複数のフラグメントを標的ポリヌクレオチドから形成する工程；（b）タグのレパートリー由来のオリゴヌクレオチドタグをその複数の各フラグメントに結合させる工程であって、（i）実質的にすべての異なるフラグメントが異なる結合オリゴヌクレオチドタグを有し、そして（ii）そのレパートリー由来の各オリゴヌクレオチドタグが最少に交差ハイブリダイズするセットから選択される、工程；（c）1つ以上の固相支持体上のフラグメントを、オリゴヌクレオチドタグと固相支持体に結合したその相補物のそれぞれとの特異的なハイブリダイゼーションにより分類する工程；（d）下記のように、好ましくは単一塩基配列決定方法論により、複数

10

20

30

40

50

の各フラグメントの一部分のヌクレオチド配列を決定する工程；ならびに（e）標的ポリヌクレオチドのヌクレオチド配列を、フラグメントの配列を対照することにより決定する工程。

本発明の別の重要な局面は、所定の組織または細胞タイプ中で発現している遺伝子のプロフィール、またはその頻度分布を決定するである。ここで、このような各遺伝子は、その配列の一部分により同定される。好ましくは、そのような頻度分布は、以下の工程により決定される：（a）mRNA分子の集団からのcDNAライブラリーの形成（cDNAライブラリー中の各cDNA分子は、結合したオリゴヌクレオチドタグを有する）を行う工程であって、（i）実質的にすべての異なるcDNA分子が異なる結合オリゴヌクレオチドタグを有し、そして（ii）そのレパートリー由来の各オリゴヌクレオチドタグが最少に交差ハイブリダイズするセットから選択される、工程；（b）cDNA分子を、オリゴヌクレオチドタグと1つ以上の固相支持体に結合したその相補物のそれぞれとの特異的なハイブリダイゼーションにより分類する工程；（c）分類されたcDNA分子のそれぞれの一部分のヌクレオチド配列を決定する工程；および（d）分類されたcDNA分子の一部分のヌクレオチド配列に由来するmRNA分子の頻度分布を形成する工程。

本発明は、オリゴヌクレオチドで分子をタグ化または標識する現在の方法の大きな欠点を克服する。本発明に従ってタグの配列をコードすることにより、タグと別のタグに対する相補物との間でミスマッチした任意の二重鎖または三重鎖の安定性は、タグとそれ自身の相補物との間で完全にマッチした任意の二重鎖の安定性よりはるかに低い。従って、GCに富むタグのミスマッチした二重鎖が完全にマッチしたATに富むタグよりも安定であるために生じる不正確な分類の問題は、除去される。

微細ビーズのような固相支持体と組み合わせて用いた場合、本発明は、大規模のDNA配列決定のような大規模の並行操作において特に有用なポリヌクレオチドを操作および分類するための容易な自動化システムを提供する。ここでは、多くの標的ポリヌクレオチドまたは単一標的ポリヌクレオチドの多くのセグメントが同時に配列決定および／または分析される。

【図面の簡単な説明】

図1は、最少に交差ハイブリダイズするセットを生じるための一般的なアルゴリズムを例示するフローチャートである。

図2は、本発明に従ってポリヌクレオチド配列決定のような並行操作を行うための装置を図で例示する。

図3は、固相支持体上に連結したプローブの分類による遺伝子型決定のための実施態様を例示する。

定義

オリゴヌクレオチドタグに関して本明細書中に用いられる「相補物」または「タグ相補物」は、オリゴヌクレオチドタグが特異的にハイブリダイズして完全にマッチした二重鎖または三重鎖を形成するオリゴヌクレオチドをいう。特異的なハイブリダイゼーションが三重鎖を生じる実施態様では、オリゴヌクレオチドタグは二本鎖または一本鎖のいずれかであることが選択され得る。従って、三重鎖が形成される場合、用語「相補物」は、一本鎖オリゴヌクレオチドタグの二本鎖相補物または二本鎖オリゴヌクレオチドタグの一本鎖相補物のいずれかを包含することを意味する。

本明細書で用いられる用語「オリゴヌクレオチド」は、モノマー対モノマーの相互作用の通常のパターン（例えば、ワトソン・クリック（Watson-Crick）型の塩基対形成、塩基の積み重ね、フーグスティーン（Hoogsteen）または逆フーグスティーン型の塩基対形成など）により標的ポリヌクレオチドに特異的に結合し得る、天然のまたは改変されたモノマーまたは連結物（デオキシリボヌクレオシド、リボヌクレオシド、それらのアノマー型、ペプチド核酸（PNA）などを含む）の直鎖オリゴマーを含む。通常、モノマーは、ホスホジエステル結合またはそのアナログにより連結され、少数のモノマーユニット（例えば3～4）から数十のモノマーユニットのサイズの範囲のオリゴヌクレオチドを形成する。オリゴヌクレオチドが一連の文字（例えば、「ATGCTG」）により示される場合はいつも、

他に記載されない限り、ヌクレオチドは左から右に5'→3'の順であり、「A」はデオキシアデノシンを示し、「C」はデオキシシチジンを示し、「G」はデオキシグアノシンを示し、および「T」はチミジンを示すことが理解される。ホスホジエステル結合のアナログは、ホスホロチオエート、ホスホロジチオエート、ホスホルアニリデート、ホスホルアミデートなどを含む。通常、本発明のオリゴヌクレオチドは4つの天然のヌクレオチドを包含する；しかし、それらはまた、非天然のヌクレオチドアナログを含有し得る。天然または非天然のヌクレオチドを有するオリゴヌクレオチドが用いられ得るとき、例えば、酵素によるプロセッシングが必要とされる場合、通常、天然のヌクレオチドからなるオリゴヌクレオチドが必要であることは当業者には明らかである。

二重鎖に関して「完全にマッチした」は、各鎖中の全てのヌクレオチドが他鎖中のヌクレオチドとワトソン-クリック塩基対形成を受けるように、二重鎖を作るポリヌクレオチド鎖またはオリゴヌクレオチド鎖が、互いに二本鎖構造を形成することを意味する。この用語はまた、用いられ得るヌクレオシドアナログ（例えば、デオキシイノシン、2-アミノプリン塩基を有するヌクレオシドなど）の対形成を含む。三重鎖に関しては、この用語は、三重鎖が完全にマッチした二重鎖および第3鎖からなり、ここで、全てのヌクレオチドが完全にマッチした二重鎖の塩基対を伴うフーグスティーンまたは逆フーグスティーン会合を受けることを意味する。逆に、二重鎖におけるタグとオリゴヌクレオチドとの間の「ミスマッチ」は、二重鎖または三重鎖におけるヌクレオチドの対またはトリプレット (triplet) がワトソン-クリックおよび／またはフーグスティーンおよび／または逆フーグスティーン結合を受け損なうことを意味する。

本明細書中に用いられる「ヌクレオシド」は、2'-デオキシ型および2'-ヒドロキシル型（例えば、KornbergおよびBaker、DNA Replication、第2版 (Freeman, San Francisco, 1992) に記載されたように）を含む天然ヌクレオシドを含む。ヌクレオシドに関して「アナログ」は、改変された塩基部分および／または改変された糖部分を有する合成ヌクレオシド（例えば、Scheit, Nucleotide Analogs (John Wiley, New York, 1980) ; UhlmanおよびPeyman, Chemical Reviews, 90:543-584 (1990) などに記載された）を含む。ただし、それらは特異的にハイブリダイズし得る。このようなアナログは、結合特性を増強し、複雑度 (complexity) を低減し、特異性を増すなどのように設計された合成ヌクレオシドを包含する。

ポリヌクレオチドに関して、本明細書中に用いられる「配列決定」または「ヌクレオチド配列を決定する」は、ポリヌクレオチドの部分的配列情報ならびに全配列情報の決定を含む。すなわち、本用語は、配列の比較、フィンガープリンティング、および標的ポリヌクレオチドに関する類似するレベルの情報、ならびに標的ポリヌクレオチドにおけるヌクレオシド、通常各ヌクレオシドの発現同一性および順序を包含する。本用語はまた、標的ポリヌクレオチド内の4つのタイプのヌクレオチドのうち、1、2、または3つのヌクレオチドの同一性、順序、および位置の決定を包含する。例えば、いくつかの実施態様において、配列決定は、標的ポリヌクレオチド「CATCGC...」内の1つのタイプのヌクレオチド（例えば、シトシン）の順序および位置を同定することによって、行われ得る。その結果、その配列はバイナリーコード（例えば、「C-（非C）-（非C）-C-（非C）-C...」に対して「100101...」）などによって表される。

ポリヌクレオチドの集団に関して、本明細書中に用いられる用語「複雑度」は、集団内に存在する異なる種の分子の数を意味する。

発明の詳細な説明

本発明は、分子（特にポリヌクレオチド）をオリゴヌクレオチドタグを用いることにより標識および分類する方法を提供する。本発明のオリゴヌクレオチドタグは、オリゴヌクレオチドの最少に交差ハイブリダイズするセットに属する。従って、レパートリーの任意の2つのオリゴヌクレオチドタグの配列は、2ヌクレオチドだけ異なり、「より厳密 (closer)」ではない。特定の実施態様では、レパートリーの任意の2つのオリゴヌクレオチドタグの配列は、例えばオリゴヌクレオチドが3つより少なくミスマッチしたヌクレオチドを有する同じセットの別のメンバー相補物と二重鎖または三重鎖を形成し得ないように最

少に交差ハイブリダイズするセットを設計することなどにより、「さらに」離れることさえあり得る。このような実施態様では、より大きい特異性が達成されるが、タグの総レパートリーはより小さい。従って、特定の長さのタグについて、所望の特異性の程度と所望のレパートリーのサイズとの間で妥協しなければならない。本発明は、配列決定、フィンガープリンティング、または他のタイプの分析のような並行操作のためにポリヌクレオチドを標識および分類するのに特に有用である。

オリゴヌクレオチドタグおよびタグ相補物

最少に交差ハイブリダイズするセットのオリゴヌクレオチドのヌクレオチド配列は、図3に図示した一般的なアルゴリズムに従い、そしてソースコードが付録IaおよびIbに挙げられるプログラムにより例示されたような簡単なコンピュータプログラムにより簡便に計数される。付録Iaのプログラムminhxは、3種のヌクレオチドからなる4マーのサブユニットを有する全ての最少に交差ハイブリダイズするセットを算定する。付録IbのプログラムtagNは、最少に交差ハイブリダイズするセットのより長いオリゴヌクレオチドを計数する。同様のアルゴリズムおよびコンピュータプログラムが、本発明の任意の実施態様について最少に交差ハイブリダイズするセットのオリゴヌクレオチドを列挙するために、容易に記載される。以下の表Iは、オリゴヌクレオチドの差の示される長さおよび数について、最少に交差ハイブリダイズするオリゴヌクレオチドのセットについての指針を提供する。数をもたらすために、上記のコンピュータプログラムを用いた。

表I

オリゴヌクレオチド のワード長	最少に交差 ハイブリダイズする セットの オリゴヌクレオチド間の ヌクレオチド差	最少に交差 ハイブリダイズ するセットの 最大サイズ	4ワードによる レパートリーの サイズ	5ワードによる レパートリーの サイズ
4	3	9	6561	5.90×10^4
6	3	27	5.3×10^3	1.43×10^4
7	4	27	5.3×10^3	1.43×10^4
7	5	8	4096	3.28×10^4
8	3	190	1.30×10^3	2.48×10^{11}
8	4	62	1.48×10^4	9.16×10^6
8	5	18	1.05×10^3	1.89×10^9
9	5	39	2.31×10^8	9.02×10^7
10	5	332	1.21×10^{10}	
10	6	28	6.15×10^3	1.72×10^7
11	5	187		
18	6	≈ 25000		
18	12	24		

本発明のいくつかの実施態様について、きわめて大きなレパートリーのタグが必要とされない場合、最少に交差ハイブリダイズするセットのオリゴヌクレオチドタグが個別に合成され得る。数百～数千または数万のオリゴヌクレオチドを含むセットが、種々の並行合成アプローチ（例えば、Frankら、米国特許第4,689,405号；Frankら、Nucleic Acids Research, 11:4365-4377 (1983)；Matsonら、Anal. Biochem., 224:110-116 (1995)；Fodorら、国際出願PCT US93/04145；Peaseら、Proc. Natl. Acad. Sci., 91:5022-5026 (1994)；Southernら、J. Biotechnology, 35:217-227 (1994)、Brennan、国際出願PCT/US94/05896；Lashkariら、Proc. Natl. Acad. Sci., 92:7912-7915 (1995)；などに開示されるように）によって直接合成され得る。

好ましくは、本発明のオリゴヌクレオチドタグは、3ヌクレオチドと6ヌクレオチドとの間の長さのサブユニットからコンビナトリアル的に合成され、そして最少に交差ハイブリダイズする同じセットから選択される。この範囲のオリゴヌクレオチドについて、そのようなセットの数は、図3のアルゴリズムに基づくコンピュータプログラムによって計数され得る。

図3のアルゴリズムは、最少に交差ハイブリダイズするセットのサブユニットの特性、す

なわち長さ、メンバー間の塩基の違いの数、および組成（例えば、2、3、または4種の塩基からなる）を最初に定義することにより実施される。所定の長さおよび組成の全ての可能性のある配列からなるテーブル M_n ($n=1$) が作製される (100)。最初のサブユニット S_1 を選択し、そしてテーブルの最後まで $i=n+1$ である連続するサブユニット S_i と比較する

(120)。連続するサブユニットが最少に交差ハイブリダイズするセットのメンバーであるために必要なミスマッチの数を有する場合はいつも、新たなテーブル M_{n+1} (125) に保存され、これは前述の工程120までの経路で以前に選択されたサブユニットをも含む。例えば、比較の第一セットでは、 M_2 は S_1 を含む；比較の第二セットでは、 M_3 は S_1 および S_2 を含む；比較の第三セットでは、 M_4 は S_1 、 S_2 、および S_3 を含む；など。同様に、テーブル M_j における比較は、 S_j と M_j における全ての連続サブユニットとの間である。各連続するテーブル M_{n+1} は、サブユニットが工程130までの連続する経路において除去されるので、その前の物より小さいことに注意する。テーブル M_n の全てのサブユニットが比較された (140) 後に、古いテーブルは新たなテーブル M_{n+1} に置き換わり、そして次のラウンドの比較が開始される。選択されたサブユニット S_i と比較するための連続するサブユニットを含まないテーブル M_n (すなわち $M_n = M_{n+1}$) に達する場合、処理は停止する (160)。

好ましくは、最少に交差ハイブリダイズするセットは、セットの他の全てのサブユニットのように二重鎖の安定性にほぼ等しい寄与をするサブユニットを含有する。この場合、全てのサブユニットとその相補物との間で完全にマッチした二重鎖の安定性はほぼ等しい。このようなセットを選択するための指針は、至適なPCRプライマーの選択および二重鎖安定性の計算についての公開された技術により提供される、例えばRychlikら、Nucleic Acids Research, 17:8543-8551 (1989) および18:6409-6412 (1990) ; Breslauerら、Proc. Natl. Acad. Sci., 83:3746-3750 (1986) ; Wetmur, Crit. Rev. Biochem. Mol. Biol., 26:227-259 (1991) など。より短いタグ（例えば、約30ヌクレオチドまたはそれより少ない）について、RychlikおよびWetmurにより記載されたアルゴリズムが好ましく、そしてより長いタグ（例えば、約30~35ヌクレオチドまたはそれより大きい）については、Suggsら、Brown編のICN-UCLA Symp. Dev. Biol., 第23巻 (Academic Press, New York, 1981) の683-693頁に開示されたアルゴリズムが簡便に用いられ得る。明らかに、本発明の範囲内で最少に交差ハイブリダイズするサブユニットのセットを設計するために当業者に利用可能な多くのアプローチがある。例えば、サブユニットを組み立てる場合、末端ヌクレオチドの異なる塩基の積み重ねエネルギーの影響を最小にするために、サブユニットは、同じ末端ヌクレオチドを有するように提供され得る。このように、サブユニットを連結する場合、全ての隣接する末端ヌクレオチドの塩基の積み重ねエネルギーの総量は同じであり、これによりタグ融解温度の多様性が低減または除去される。

下記のイタリック体で示される末端ヌクレオチドの「ワード」はまた、それと他のいずれかのタグ相補物の類似末端「ワード」との間で完全なマッチが常に形成されるようにタグの各末端に付加され得る。そのようなタグの増大は、以下の形態を有する：

<i>W</i>	<i>W</i> ₁	<i>W</i> ₂	...	<i>W</i> _{<i>k</i>-1}	<i>W</i> _{<i>k</i>}	<i>W</i>
<i>W</i> '	<i>W</i> ₁ '	<i>W</i> ₂ '	...	<i>W</i> _{<i>k</i>-1} '	<i>W</i> _{<i>k</i>} '	<i>W</i> '

ここで、プライムされた W' は、相補物を示す。常に完全にマッチした二重鎖を形成するタグの末端を伴う場合、全てのミスマッチしたワードは、内部のミスマッチであり、それによってタグ相補物二重鎖の安定性が減少し、その二重鎖は、そうでなければ、それらの末端でミスマッチしたワードを有する。内部ミスマッチを有する二重鎖が、末端で同じミスマッチを有する二重鎖よりも有意に安定ではないことは、周知である。

最少に交差ハイブリダイズするセットの好ましい実施態様は、そのサブユニットが4つの天然ヌクレオチドのうちの3つからなるセットである。下記でより詳細に述べるように、オリゴヌクレオチドタグ中の1つのタイプのヌクレオチドの欠損は、DNAポリメラーゼの5'→3' エクソヌクレアーゼ活性の使用により標的ポリヌクレオチドが固相支持体上にロードされることを可能にする。以下は、A、G、およびTからなる群から選択される4つのヌクレオチドのおおのを含有するサブユニットの最少に交差ハイブリダイズするセッ

トの例である：

表II

ワード：	w_1	w_2	w_3	w_4
配列：	GATT	TGAT	TAGA	TTTG

ワード：	w_5	w_6	w_7	w_8
配列：	GTAA	AGTA	ATGT	AAAG

10

このセットでは、各メンバーは、全ての他のメンバーの相補物と3つのミスマッチした塩基を有する二重鎖を形成する。

さらなる典型的な最少に交差ハイブリダイズするセットが下記の表IIIに挙げられる。明らかに、さらなるセットは異なるグループのヌクレオチドを置換することによるか、または公知の最少に交差ハイブリダイズするセットのサブユニットを用いることにより作製され得る。

表III

4 マーサブユニットの典型的な最少に交差ハイブリダイズするセット

20

セット1	セット2	セット3	セット4	セット5	セット6
CATT	ACCC	AAAC	AAAG	AACA	AACG
CTAA	AGGG	ACCA	ACCA	ACAC	ACAA
TCAT	CACG	AGGG	AGGC	AGGG	AGGC
ACTA	CCGA	CACG	CACC	CAAG	CAAC
TACA	CGAC	CCGC	CCGG	CCGC	CCGG
TTTC	GAGC	CGAA	CGAA	CGCA	CGCA
ATCT	GCAG	GAGA	GAGA	GAGA	GAGA
AAAC	GGCA	GCAG	GCAC	GCCG	GCCC
	AAAA	GGCC	GGCG	GGAC	GGAG
セット7	セット8	セット9	セット10	セット11	セット12
AAGA	AAGC	AAGG	ACAG	ACCG	ACGA
ACAC	ACAA	ACAA	AACA	AAAA	AAAC
AGCG	AGCG	AGCC	AGGC	AGGC	AGCG
CAAG	CAAG	CAAC	CAAC	CACC	CACA
CCCA	CCCC	CCCG	CCGA	CCGA	CCAG
CGGC	CGGA	CGGA	CGCG	CGAG	CGGC
GACC	GACA	GACA	GAGG	GAGG	GAGG
GCGG	GCGG	GCGC	GCCC	GCAC	GCCC
GGAA	GGAC	GGAG	GGAA	GGCA	GGAA

30

40

本発明のオリゴヌクレオチドタグおよびそれらの相補物は、自動化DNA合成機（例えば、Applied Biosystems, Inc. (Foster City, California) モデル392または394型DNA/RNA合成機）で、ホスホルアミダイトケミストリーのような標準的なケミストリー（例えば、以下の参考文献で開示される）を用いて簡便に合成される：BeaucageおよびIyer, Tetrahedron. 48:2223-2311 (1992)；Molkoら、米国特許第4,980,460号；Kosterら、米国特許第4,725,677号；Caruthersら、米国特許第4,415,732号；同第4,458,066号；および同第4,973,679号；など。得られたオリゴヌクレオチドが特異的にハイブリダイズし得るという条件で、別のケミストリー（例えば、ホスホロチオエート、ホスホルアミデートなどのような非天然骨格基（backbone group）をもたらす）も用いられ得る。いくつかの実施態様では、タグは、酵素により処理または操作され得る天然に存在するヌクレオチドを含有し

50

得る、一方、対応するタグ相補物は、分類する間、より安定な二重鎖の形成を促進するペプチド核酸または同様の化合物のような非天然ヌクレオチドアナログを含有し得る。

微粒子が支持体として用いられる場合、オリゴヌクレオチドタグおよびタグ相補物のレパートリーは、例えば、Shortleら、国際特許出願PCT/US93/03418またはLyttleら、Biotechniques, 19:274-280 (1995) に開示されるような「分裂および混合 (split and mix)」技術を介するサブユニット様式 (subunit-wise) 合成により作製され得る。簡単に述べると、合成の基本ユニットは、オリゴヌクレオチドタグのサブユニットである。好ましくは、ホスホルアミダイトケミストリーを用い、3' ホスホルアミダイトオリゴヌクレオチドが最少に交差ハイブリダイズするセット (例えば、上記に最初に挙げたセットについて、8つの4マーの3'-ホスホルアミダイトが存在する) における各サブユニットについて調製される。合成は、Shortleらにより開示されたように、またはヌクレオシドモノマーを用いる種々のオリゴヌクレオチドライブラリーを生じるために用いられる技術 (例えば、Teleniusら、Genomics, 13:718-725 (1992) ; Welshら、Nucleic Acids Research, 19:5275-5279 (1991) ; Grothuesら、Nucleic Acids Research, 21:1321-1322 (1993) ; Hartley、ヨーロッパ特許出願第90304496.4号 ; Lamら、Nature, 354:82-84 (1991) ; Zuckermanら、Int. J. Pept. Protein Research, 40:498-507 (1992) ; などに開示された) に直接類似して進められる。一般的に、これらの技術はカップリング工程の間、成長するオリゴヌクレオチドへの活性化されたモノマーの混合物の適用を単に必要とするのみである。好ましくは、オリゴヌクレオチドタグおよびタグ相補物を、タグの構築において用いられる異なる種類のワードの数以上の多数の合成チャンバを有するDNA合成機上で合成する。すなわち、好ましくは、各タイプのワードに対応する合成チャンバが存在する。この実施態様において、ワードが5つのヌクレオチドからなるのであれば、各合成チャンバにおいて5つのモノマーカップリングが存在するように、ワードはヌクレオチド毎に付加される。ワードが完全に合成された後、合成支持体をチャンバから取り除き、混合し、そして次のサイクルのワード付加のためのチャンバに再配分する。この後者の実施態様は、例えば、ホスホルアミデートケミストリーにおいて、モノマー付加の高いカップリング収率を利用する。

二本鎖型のタグは、相補物鎖を別個に合成し、続いて二重鎖形成を可能にする条件下で混合することにより作製され得る。あるいは、二本鎖タグは、プライマー結合部位として働く公知のオリゴヌクレオチド配列に連結した一本鎖レパートリーを最初に合成することにより形成され得る。次いで、第2の鎖は、一本鎖レパートリーとプライマーとを一緒にし、次いでポリメラーゼで伸長することにより合成される。この後者のアプローチは、Oliphantら、Gene, 44:177-183 (1986) に記載される。次いで、このような二重鎖タグは、本発明に従って、標的ポリヌクレオチドの分類および操作用の標的ポリヌクレオチドと一緒にクローニングベクターに挿入され得る。

PNAまたはオリゴヌクレオチドN3'→P5' ホスホルアミデートのような、増強された結合特性を有する核酸から作製されたタグ相補物を用いる場合、DNAポリメラーゼの3'→5' エキソヌクレアーゼ活性を用いて、tag一本鎖を与える「ストリッピング」反応の代わりとして、天然のヌクレオチドとそれらのPNAまたはホスホルアミデート相補物との間のDループの形成を介して分類が実施され得る。

本発明のオリゴヌクレオチドタグは、長さが12~60のヌクレオチドまたは塩基対の範囲であり得る。好ましくは、オリゴヌクレオチドタグは、長さが18~40のヌクレオチドまたは塩基対の範囲である。さらに好ましくは、オリゴヌクレオチドタグは、長さが25~40のヌクレオチドまたは塩基対の範囲である。サブユニットの好ましい数およびさらに好ましい数に関して、それらの範囲は、以下のように表され得る：

表 IV

好ましい実施態様におけるタグ中のサブユニットの数

サブユニット中 のモノマー	オリゴヌクレオチドタグ中のヌクレオチド		
	(12-60)	(18-40)	(25-40)
3	4-20 サブユニット	6-13 サブユニット	8-13 サブユニット
4	3-15 サブユニット	4-10 サブユニット	6-10 サブユニット
5	2-12 サブユニット	3-8 サブユニット	5-8 サブユニット
6	2-10 サブユニット	3-6 サブユニット	4-6 サブユニット

最も好ましくは、オリゴヌクレオチドタグが一本鎖であり、そして特異的なハイブリダイゼーションが、タグ相補物とのワトソン・クリック対形成により起こる。

好ましくは、本発明の一本鎖オリゴヌクレオチドタグのレパートリーは、少なくとも100のメンバーを含む；より好ましくは、そのようなタグのレパートリーは少なくとも、1,000のメンバーを含む；そして最も好ましくは、そのようなタグのレパートリーは、少なくとも10,000のメンバーを含む。

三重鎖タグ

特異的なハイブリダイゼーションが三重鎖形成を介して生じる実施態様では、タグ配列のコード化は、二重鎖形成タグについての原理と同じ原理に従う。しかし、サブユニット配列の選択において、さらなる制約が存在する。一般に、フーグスティーン型の結合を介する第3鎖の会合は、二本鎖標的においてホモピリミジン-ホモプリントラックに沿って最も安定である。通常、塩基トリプレットは、T-A^{*}TまたはC-G^{*}Cモチーフ（ここで、「-」はワトソン・クリック対形成を示し、そして「^{*}」はフーグスティーン型の結合を示す）を形成する。しかし、他のモチーフもまた可能である。例えば、フーグスティーン塩基対形成は、鎖の状態および組成に依存して、第3鎖（フーグスティーン鎖）と第3鎖が結合する二重鎖のプリンに富む鎖との間に平行および逆平行（antiparallel）の配向を可能にする。特定の実施態様で所望されるように、三重鎖の安定性を最大にするか、またはそうでなければ調節するための、適切な配列、配向、条件、ヌクレオシド型（例えば、リボースヌクレオシドまたはデオキシリボースヌクレオシドのどちらかが用いられる）、塩基修飾（例えば、メチル化シトシンなど）の選択についての広範囲にわたる指針が文献中に見られる。例えば、Robertsら、Proc.Natl.Acad.Sci.、88:9397-9401（1991）；Robertsら、Science、258:1463-1466（1992）；Robertsら、Proc.Natl.Acad.Sci.、93:4320-4325（1996）；Distefanoら、Proc.Natl.Acad.Sci.、90:1179-1183（1993）；Mergnyら、Biochemistry、30:9791-9798（1991）；Chengら、J.Am.Chem.Soc.、114:4465-4474（1992）；BealおよびDervan、Nucleic Acids Research、20:2773-2776（1992）；BealおよびDervan、J.Am.Chem.Soc.、114:4976-4982（1992）；Giovannangeliら、Proc.Natl.Acad.Sci.、89:8631-8635（1992）；MoserおよびDervan、Science、238:645-650（1987）；McShanら、J.Biol.Chem.、267:5712-5721（1992）；Yoonら、Proc.Natl.Acad.Sci.、89:3840-3844（1992）；Blumeら、Nucleic Acids Research、20:1777-1784（1992）；ThuongおよびHelene、Angew.Chem.Int.Ed.Engl.、32:666-690（1993）；Escudeら、Proc.Natl.Acad.Sci.、93:4365-4369（1996）など。一本鎖タグまたは二重鎖タグをそれらの一本鎖相補物または二重鎖相補物にアニールするための条件は、周知である（例えば、Jiら、Anal.Chem.、65:1323-1328（1993）；Cantorら、米国特許第5,482,836号など）。三重鎖タグの使用は、アニーリングのためのタグをその相補物に曝すために、ポリメラーゼによる「ストリッピング」反応を必要としないという利点を有する。

好ましくは、三重鎖ハイブリダイゼーションを用いる本発明のオリゴヌクレオチドタグは、二本鎖DNAであり、そして対応するタグ相補物は、一本鎖である。より好ましくは、タグとその相補物との間で形成される三重鎖のpH安定性の範囲を広げるために、タグ相補物において、5-メチルシトシンがシトシンの代わりに用いられる。三重鎖を形成するための好ましい条件は、上記の参考文献において十分に開示されている。簡単に説明すると、ハイブリダイゼーションは、濃縮された塩溶液（例えば、1.0M NaCl、1.0M酢酸カリウムなど、pH5.5未満（または、5-メチルシトシンが用いられる場合、6.5）で生じる。ハイブリダイゼーションの温度は、タグの長さおよび組成に依存する；しかし、18~20マー以上のタグの場合、室温でのハイブリダイゼーションが適切である。洗浄は、より低い濃度の塩溶液（例えば、10mM酢酸ナトリウム、100mM MgCl₂、pH5.8、室温）で行われ得る。タグは、同様の塩溶液（pH9.0）におけるインキュベーションによってそれらのタグ相補物から溶出され得る。

三重鎖を形成するオリゴヌクレオチドの、最少に交差ハイブリダイズするセットは、付録Icのコンピュータプログラムまたは類似のプログラムによって作製され得る。二本鎖の8マーワードのセットの例を大文字で、対応する相補物を小文字で以下に列挙する。そのような各ワードは、セット中の他のワードのそれぞれと3塩基対だけ異なる。

表 V

二本鎖8マータグの最少に交差ハイブリダイズするセットの例

5'-AAGGAGAG	5'-AAAGGGGA	5'-AGAGAAGA	5'-AGGGGGGG
3'-TTCCTCTC	3'-TTTCCCCT	3'-TCTCTTCT	3'-TCCCCCCC
3'-ttccctctc	3'-tttccccct	3'-tctcttctt	3'-tcccccccc
5'-AAAAA	5'-AAGAGAGA	5'-AGGAAAAG	5'-GAAGGAG
3'-TTTTTTT	3'-TTCTCTCT	3'-TCCTTTTC	3'-CTTCTCTC
3'-ttttttttt	3'-ttctctctt	3'-tcctttttc	3'-cttctcttc
5'-AAAAAGGG	5'-AGAAGAGG	5'-AGGAAGGA	5'-GAAGAAGG
3'-TTTTTCCC	3'-TCTTCTCC	3'-TCCTTCCT	3'-CTTCTTCC
3'-ttttttccc	3'-tcttctctc	3'-tccttcctt	3'-cttctcttc
5'-AAAGGAAG	5'-AGAAGGAA	5'-AGGGGAAA	5'-GAAGAGAA
3'-TTTCCTTC	3'-TCTTCCTT	3'-TCCCCTTT	3'-CTTCTCTT
3'-tttcccttc	3'-tcttccctt	3'-tcccccttt	3'-cttctcttt

表 VI

タグ相補物と三重鎖を形成する種々の二本鎖タグのレパートリーサイズ

オリゴヌクレオチド のワード長	最少に交差 ハイブリダイズする セットの オリゴヌクレオチド間の ヌクレオチド差	最少に交差 ハイブリダイズ するセットの 最大サイズ	4ワードによる レパートリーの サイズ	5ワードによる レパートリーの サイズ
4	2	8	4096	3.2×10^4
6	3	8	4096	3.2×10^4
8	3	16	6.5×10^4	1.05×10^5
10	5	8	4096	
15	5	92		
20	6	765		
20	8	92		
20	10	22		

好ましくは、本発明の二本鎖オリゴヌクレオチドタグのレパートリーは、少なくとも10のメンバーを含む；より好ましくは、そのようなタグのレパートリーは、少なくとも100のメンバーを含む。好ましくは、ワードは、コンビナトリアル的に合成される二本鎖オリゴヌクレオチドタグについて、4ヌクレオチドと8ヌクレオチドとの間の長さであり、そしてオリゴヌクレオチドタグは、12塩基対と60塩基対との間の長さである。より好ましくは

、そのようなタグは、18塩基対と40塩基対との間の長さである。

固相支持体

本発明で用いられる固相支持体は、広い種類の形態を有し得、例えば、微粒子、ビーズ、および膜、スライド、プレート、マイクロマシーンドチップ (micromachined chip) などを含む。同様に、本発明の固相支持体は、広い種類の組成物を含み得、例えば、ガラス、プラスチック、シリコン、アルカンチオレート誘導体化金 (alkanethiolate-derivatized gold)、セルロース、低架橋ポリスチレンおよび高架橋ポリスチレン、シリカゲル、ポリアミドなどが挙げられる。好ましくは、分離した粒子の集団は、各々が同じタグ（および他はなし）の相補的配列の均一のコーティング（または集団）を有するように用いられるか、あるいは、単一支持体または数個の支持体は、各々が同じタグ（および他はなし）への相補的配列の均一のコーティング（または集団）を含む空間的に離れた領域で用いられるかのいずれかである。後者の実施態様では、領域の面積は、特定の適用に従って変化し得、通常、領域は、数 μm^2 （例えば、3~5）から数百 μm^2 （例えば、100~500）の面積の範囲である。好ましくは、このような領域は、空間的に分離され、そのため隣接領域での事象（例えば、蛍光発光）により発生されるシグナルが、用いられる検出システムにより識別され得る。いくつかの適用では、1つより多いタグ相補物（例えば、同時配列分析、または別々にタグ化された分子を極めて近接させるための）の均一コーティングの領域を有することが、望ましくあり得る。

タグ相補物は、その上でタグ相補物が合成される固相支持体と共に用いられ得、あるいは、別々に合成され得、そして使用のために固相支持体に結合され得る。例えば、このことは以下に開示される：Lundら、Nucleic Acids Research、16:10861-10880 (1988)；Albr etsenら、Anal.Biochem.、189:40-50 (1990)；Wolfら、Nucleic Acids Research、15:29 11-2926 (1987)；またはGhoshら、Nucleic Acids Research、15:5353-5372 (1987)。好ましくは、タグ相補物は、同じ固相支持体上で合成され、そして同じ固相支持体と共に用いられ得、それは、種々の形態を含み、そして種々の結合部分を含み得る。このような支持体は、タグ相補物の均一集団が合成される領域の微粒子、またはアレイ (array)、またはマトリックスを含み得る。広い種類の微粒子支持体が、本発明で用いられ得、制御された孔のガラス (CPG)、高架橋ポリスチレン、アクリルコポリマー、セルロース、ナイロン、デキストラン、ラテックス、ポリアクロレインなどにより作製される微粒子が挙げられ、以下の典型的な参考文献中に開示されている：Meth.Enzymol.、Section A、11-147 頁、第44巻 (Academic Press、New York、1976)；米国特許第4,678,814号；同第4,413,0 70号；および同第4,046,720号；ならびにPon、19章、Agrawal (編)、Methods in Molecu lar Biology、第20巻 (Humana Press、Totowa、NJ、1993)。微粒子支持体としては、市販のヌクレオシドー誘導体化CPGおよびポリスチレンビーズ（例えば、Applied Biosystem s、Foster City、CAから入手可能）；誘導体化磁気ビーズ；ポリエチレングリコールでグラフト化されたポリスチレン（例えば、TentaGelTM、Rapp Polymere、Tubingen Germany）などがさらに挙げられる。支持体の特性（例えば、材料、多孔度、サイズ、形状など）の選択、および用いられる結合部分のタイプは、タグを用いる際の条件に依存する。例えば、酵素による連続的な処理を含む適用において、酵素の立体障害を最小にし、そして基質への接近を容易にする支持体およびリンカーが好ましい。最も適切な微粒子支持体を選択する際に、考慮されるべき他の重要な要因としては、サイズの均一性、合成支持体としての効率、既知の表面積の度合い、および光学特性（例えば、以下により十分に説明するように、表面で多くのビーズを扱う場合、透明で平滑なビーズは、計測的な (instrument ational) 利点を提供する）が挙げられる。

微粒子表面上でタグを結合させるおよび／または合成するための典型的な結合部分は、以下に開示されている：Ponら、Biotechniques、6:768-775 (1988)；Webb、米国特許第4,6 59,774号；Baranyら、国際特許出願PCT/US91/06103；Brownら、J.Chem.Soc.Comm.、198 9:891-893；Damhaら、Nucleic Acids Research、18:3813-3821 (1990)；Beattieら、Cli nical Chemistry、39:719-722 (1993)；MaskosおよびSouthern、Nucleic Acids Researc h、20:1679-1684 (1992) など。

上記のように、タグ相補物はまた、単一の（または数個の）固相支持体上で合成され得、タグ相補物で均一にコーティングされた領域のアレイを形成する。すなわち、このようなアレイ中の各領域内で、同じタグ相補物が合成される。このようなアレイを合成するための技術は、以下に開示される：McGallら、国際出願PCT/US93/03767；Peaseら、Proc.Natl.Acad.Sci.、91:5022-5026（1994）；SouthernおよびMaskos、国際出願PCT/GB89/01114；MaskosおよびSouthern（上記）；Southernら、Genomics、13:1008-1017（1992）；およびMaskosおよびSouthern、Nucleic Acids Research、21:4663-4669（1993）。

好ましくは、本発明は、同じタグ配列の相補物で均一にコーティングされた微粒子またはビーズを用いて実行される。微粒子支持体、およびそれらの表面とオリゴヌクレオチドとを共有結合または非共有結合する方法は、周知であり、以下の参考文献により例示される：BeaucageおよびIyer（上記）；Gait（編）Oligonucleotide Synthesis；A Practical Approach（IRL Press、Oxford、1984）；および上記の参考文献。一般に、微粒子のサイズおよび形態は重要ではない；しかし、最少の試薬およびサンプルの使用によりオリゴヌクレオチドタグの多くのレパートリーの構築および操作を容易にするので、直径数 μm （例えば、1~2 μm ）から数百 μm （例えば、200~1000 μm ）の範囲のサイズの微粒子が好ましい。

いくつかの好ましい適用では、制御された孔の多孔性ガラス（CPG）またはポリスチレン支持体が、本発明における固相支持体として用いられる。塩基不安定性のリンカーおよび結合された最初のヌクレオチドを有するこのような支持体は、入手可能である（例えば、Applied Biosystems（Foster City、CA））。好ましくは、500オングストローム~1000オングストロームの間の孔サイズを有する微粒子が用いられる。

他の好ましい適用では、孔のない微粒子が、その光学特性のために用いられ、この微粒子は、平面支持体（例えば、顕微鏡スライド）の上で非常に多くの微粒子を感知する場合、有利に用いられ得る。特に、好ましい孔のない微粒子は、Bangs Laboratories（Carmel、IN）から入手できるグリシダルメタクリレート（glycidial methacrylate）（GMA）ビーズである。このような微粒子は、種々のサイズで有用であり、そしてタグまたはタグ相補物を合成するための種々の結合基で誘導体化される。好ましくは、タグ化された微粒子の多量（massively）の平行操作のために、直径5 μm のGMAビーズが用いられる。

固相支持体上での分類のための、ポリヌクレオチドへのタグの結合

本発明の重要な局面は、例えば、cDNAライブラリー由来のポリヌクレオチドの集団を分類し、これを、微粒子、または、固相支持体上の別々の領域に結合させることである。それによって、微粒子または領域のそれぞれに、実質的にただ一種類だけのポリヌクレオチドが結合する。この目的は、実質的にすべての異なるポリヌクレオチドが、異なる結合タグを有することを確実にすることによって実施される。この条件は、次に、タグーポリヌクレオチド結合体の完全な集合体からサンプルをとって分析することによって達成される（同一ポリヌクレオチドが異なるタグを有することは許される。なぜなら、それは、単に、同じポリヌクレオチドが、2つの異なる位置で作用するか、または、2回分析されるということをもたらすにすぎないからである）。このようなサンプリングは、以下のいずれを用いてもよい：タグをポリヌクレオチドに結合させた後で、明らかに（例えば、大きな混合体から小容量を取り出すことにより）行ない得るし、ポリヌクレオチドとタグとを処理するのに用いられる技術の二次的効果として固有的に実施され得るし、または明らかにおよび処理工程の固有部分としての両用で実施され得る。

実質的にすべての異なるcDNAが異なるタグを有するcDNAライブラリーを構築するに当たって、その複雑度、または異なるタグの数は、細胞または組織サンプルから抽出されたmRNAの全数をはるかに越えているタグレパートリーの使用が好ましい。タグレパートリーの複雑度は、ポリヌクレオチド集団の複雑度の少なくとも10倍であることが好ましい。より好ましくは、タグレパートリーの複雑度は、ポリヌクレオチド集団の複雑度の少なくとも100倍である。以下に、典型的に、9ワードのタグのすべてのレパートリーを含むプライマー混合物を用いたcDNAライブラリー構築のためのプロトコルを開示する。タグ含有プライマーのそのような混合体は、 8^9 または約 1.34×10^8 の複雑度を有する。Winslowら、Nuclei

c Acids Research、19:3251-3253 (1991) により示されるように、ライブラリー構築用の mRNA は、10~100個というごく少数の哺乳動物細胞から抽出することができる。単一の哺乳動物細胞は、約 3.4×10^4 の異なる種類の mRNA 分子の約 5×10^5 コピーを含んでいるので、標準的な技術を用いると、mRNA を約 100 個の細胞、または（理論的に）約 5×10^7 の mRNA 分子から単離され得る。この数をプライマー混合体の複雑度と比べると、それ以上の工程を有することなく、さらに、mRNA が完全な効率（1% 以下の効率がさらに正確であるが）で cDNA に変換されると仮定すると、この cDNA ライブラリー構築プロトコルにより、異なるタグの全数の多くても 37% を含む集団が得られることが判明する。すなわち、まったく明らかな（overt）サンプリングをしなくても、このプロトコルは、タグレパートリーの 37% 以下を含むサンプルを固有的に生成する。これらの条件下で重複を得る確率は約 5% であり、この数字は好ましい範囲の中にある。10 個の細胞からの mRNA を用いてサンプリングしたタグレパートリーの割合は、仮に処理工程がすべて 100% の効率で行なわれたとしても、僅か 3.7% に低下する。実際には、cDNA ライブラリーを構築するための処理工程の効率は極めて低い。「経験則」では、良いライブラリーならば、 10^6 個の哺乳動物細胞から抽出される mRNA に由来する約 10^8 個の cDNA クローンを含んでいるはずである。

上記のプロトコルにおいて、さらに大量の mRNA の使用、あるいは一般に、さらに大量のポリヌクレオチドのための使用（ここで、このような分子の数はタグレパートリーの複雑度を越える）、その際には、タグーポリヌクレオチド結合体の混合物は、タグと mRNA またはポリヌクレオチドとのすべての可能な対形成を含む可能性がある。このような場合、明らかなサンプリングは、タグーポリヌクレオチド結合体の出発混合物を連続希釈した後、サンプル容量を取り出して実施することができる。必要な希釈量は、出発材料の量および処理工程の効率に依存する。これらの量は容易に見積もられる。

mRNA を 10^6 個の細胞（これは、poly (A) ⁺ RNA の約 $0.5 \mu\text{g}$ に相当する）から抽出し、そしてプライマーが、通常のプロトコル（例えば、Sambrook ら、Molecular Cloning、第 2 版、8.61 頁 [1mg/ml での 1.8kb mRNA $10 \mu\text{L}$ は約 1.68×10^{-11} モルに等しく、1mg/ml での 18mer プライマーの $10 \mu\text{L}$ は約 1.68×10^{-9} モルに等しい]）で要求される濃度の約 10~100 倍の濃度過剰で存在する場合、cDNA ライブラリー中のタグーポリヌクレオチド結合体の全数は、単純に、mRNA の出発数以下、すなわちタグーポリヌクレオチド結合体を含む約 5×10^{11} 個のベクター以下となる：ここでも、cDNA 構築の各工程（第 1 鎖の合成、第 2 鎖の合成、ベクターへの連結等）が完全な効率で実施されたと仮定する。これは、きわめて控えめの推定である。実際の数はいくらかに少ない。

n 個のタグーポリヌクレオチド結合体のサンプルを、反応混合物から無作為に抽出する（これは、あるサンプル容量を取り出すことで実行される）と、同じタグを有する結合体を引く確率はポアソン分布 $P(r) = e^{-\lambda} (\lambda)^r / r!$ で表わされる。ここで、r は同じタグを有する結合体の数であり、 $\lambda = np$ であり、ここで、p は所定のタグが選られられる確率である。もしも、 $n = 10^6$ であり、かつ、 $p = 1 / (1.34 \times 10^8)$ と仮定すると、 $\lambda = 0.00746$ であり、 $P(2) = 2.76 \times 10^{-5}$ となる。従って、百万個の分子から成るサンプルであっても、十分好ましい範囲の中に予想される重複の数を生じる。このようなサンプルは、以下のようにして容易に得られる： 5×10^{11} 個の mRNA が、挿入体としてタグー cDNA 結合体を含む 5×10^{11} 個のベクターに完全に変換され、かつこの 5×10^{11} 個のベクターが、 $100 \mu\text{l}$ の容量を有する反応液中に存在すると仮定する。この原液からの $10 \mu\text{l}$ を、TE のような適切な緩衝液の $90 \mu\text{l}$ を含む容器に移すことによって、4 つの 10 倍連続希釈を実施することができる。この処理をさらに 3 回希釈繰り返すと、 $1 \mu\text{l}$ 当たり 5×10^5 個のベクター分子を含む $100 \mu\text{l}$ 溶液が得られ得る。この溶液から得られる $2 \mu\text{l}$ アリコートは、挿入体としてのタグ cDNA 結合体を含む 10^6 個のベクターを生成する。次に、このサンプルを、コンピテントな宿主細胞を直ちに形質変換により増幅し、その後培養する。

もちろん、上述したように、上記の処理のいずれの工程も完全な効率で進行しない。特に、あるサンプルのタグーポリヌクレオチド結合体を増幅するのにベクターを用いる場合、宿主を形質変換する工程の効率がきわめて悪い。通常、ベクターのわずか 1% が、宿主によって取り込まれ、そして複製される。従って、このような増幅法の場合、 10^6 個の結合

体のサンプルを得るためには、さらに少ない数の希釈が必要とされる。

オリゴヌクレオチドタグのレパートリーは、多くの方法でポリヌクレオチド集団に結合し得る。この中には、直接的な酵素による連結、増幅、例えばタグ配列を含むプライマーを用いたPCRなどが含まれる。最初の連結工程では、タグーポリヌクレオチド結合体のきわめて大規模な集団が生成されるので、一般に、単一のタグが、多くの異なるポリヌクレオチドに結合される。しかし、前記したように、結合体の十分に小さいサンプルを用いることによって、「重複」、すなわち、同じタグが2つの異なるポリヌクレオチドに結合した例を得る確率は、無視できるほど小さくすることができる。一般に、サンプルが大きければ大きいほど、重複を得る確率は大きくなる。従って、設計上の妥協は、大きなサンプルのタグーポリヌクレオチド結合体を選ぶこと（これによって、例えば、ショットガン配列決定操作における標的ポリヌクレオチドを十分にカバーできる、または急速に変化するmRNAプールを十分に代表させることができる）と、小さなサンプルを選ぶこと（これによって、重複の数を最小にすることができる）との間にある。たいていの実施態様では、重複の存在は、単に、さらなるノイズ源を付加するか、または配列決定の場合には、走査および信号処理を僅かに複雑化させるに過ぎない。なぜなら、多数の蛍光信号を与える微粒子は単純に無視することができるからである。

タグを分子、特にポリヌクレオチドに結合させるのに関して本明細書中で用いる「実質的にすべて」という用語は、本質的に重複のない、タグー分子結合体の集団を得るのに用いられるサンプリング手順の統計的性質を反映することを意味する。タグー分子結合体の実際の割合に関して、実質的にすべての意味は、タグをどのように使用するかによって異なる。核酸配列決定のためには、好ましくは、実質的にすべては、そのポリヌクレオチドの少なくとも80%がユニーク（一意的）なタグを結合することを意味する。より好ましくは、実質的にすべてが、そのポリヌクレオチドの少なくとも90%がユニークなタグを結合することを意味する。さらにより好ましくは、そのポリヌクレオチドの少なくとも95%がユニークなタグを結合することを意味する。そして、最も好ましくは、そのポリヌクレオチドの少なくとも99%がユニークなタグを結合することを意味する。

ポリヌクレオチドの集団がメッセンジャーRNA（mRNA）から成る場合、好ましくは、オリゴヌクレオチドタグは、タグ配列の相補物を含む一組のプライマーを用いて、mRNAを逆転写することによって結合され得る。典型的な一組のこのようなプライマーは、下記の配列を有し得る。

5'-mRNA-[A]_n-3'

[T]₁₉GG[W, W, W, C]₉ACCAGCTGATC-5'-ビオチン

ここで、「[W, W, W, C]₉」は、それぞれ4個のヌクレオチドから成る9個のサブユニットのオリゴヌクレオチドタグ配列を表わし、かつ、「[W, W, W, C]」は、前記のサブユニット配列を表わす。すなわち、「W」は、TまたはAを表わす。下線部の配列は、任意の制限エンドヌクレアーゼ部位を示し、この部位は、そのような部位が用いられる場合、ビオチンを介する固相支持体への結合からポリヌクレオチドを遊離するのに使用することができる。上記のプライマーに関しては、微粒子に結合される相補物は、下記の形態を取り得る。

5'-[G, W, W, W]₉TGG-リンカー-微粒子

逆転写後、mRNAは、例えば、RNase H消化によって除去され、そしてcDNAの第2の鎖は、例えば、下記の形態のプライマーを用いて合成される。

5'-NRRGATCYNNN-3'

ここで、Nは、A、T、G、またはCのいずれか一つであり；Rは、プリンを含むヌクレオチドであり、そしてYは、ピリミジンを含むヌクレオチドである。この特定のプライマーは、得られた二重鎖DNAの中にBstY1制限部位を形成する。これは、SalI部位と一緒に、例えばBamHIおよびXhoI部位を有するベクターへのクローン化を容易にする。BstY1およびSalIによる消化の後、典型的な結合体は、下記の形態を取る。

5'-RCGACCA[C, W, W, W]₉GG[T]₁₉-cDNA-NNNR

GGT[G, W, W, W]₉CC[A]₁₉-rDNA-NNNYCTAG-5'

次に、このポリヌクレオチドータグ結合体は、分子生物学の標準的な技術を用いて操作され得る。例えば、前記の結合体（実際には、混合体である）を、市販のクローニングベクター、例えばStratagene Cloning System (La Jolla, CA) に挿入してもよい；宿主（例えば、市販の宿主細菌）にトランスフェクトし；次いでそれを培養し、結合体の数を増やす。次に、このクローニングベクターを、標準的な技術（例えば、Sambrookら、Molecular Cloning、第2版（Cold Spring Harbor Laboratory, New York, 1989））を用いて単離してもよい。あるいは、適切なアダプターとプライマーとを用いて、この結合体集団をPCRによって増やしてもよい。

10

好ましくは、リガーゼに基づく配列決定法を用いる場合には、BstYIおよびSalIによる消化フラグメントを、下記の単一コピー制限部位を有するBamHI-/XhoI-消化ベクターに挿入してクローン化する：

5'-GAGGATGCCTTTATGGATCCACTCGAGATCCCAATCCA-3'

Fok I BamHI XhoI

これは、FokI部位を付加している。これによって、下記に、より十分に記述される配列決定プロセスの開始を可能にする。

タグは、標準的なクローニング法による既存のライブラリーのcDNAに結合され得る。既存のベクターからcDNAを取り出し、単離し、次に、タグのレパートリーを含むベクターに連結する。好ましくは、このタグ含有ベクターは、2つの制限酵素で切断することにより線状とし、その結果、切り出されたcDNAが、所定の方向に連結され得る。この線状タグ含有ベクターの濃度は、cDNA挿入物の濃度を実質的に越えるものであって、従って、連結がタグの固有的なサンプリングを与える。

20

増幅後に一本鎖タグを暴露する一般的な方法は、T4 DNAポリメラーゼ、または同様の酵素の5'→3'エキソヌクレアーゼ活性によって、標的ポリヌクレオチド含有結合体を消化することを含む。単一のデオキシヌクレオシドトリホスフェートの存在下で用いると、このようなポリメラーゼは、テンプレート鎖上で、この単一デオキシヌクレオチドトリホスフェートの相補物が到達するまで、二本鎖フラグメントの非テンプレート鎖に存在する3'の縮退末端からヌクレオチドを切断する。このようなヌクレオチドに達すると、5'→3'の消化は実効的に停止する。なぜなら、除去活性がヌクレオチドを取り去るよりも、ポリメラーゼによる伸長活性の方が、大きい速度でヌクレオチドを付加するからである。その結果、3ヌクレオチドで構築された一本鎖タグが、固相支持体にロードされたされるための容易に調製される。

30

この技術はまた、標的ポリヌクレオチド内のFokI部位を優先的にメチル化し、一方、ポリヌクレオチドの末端部の単一FokI部位をメチル化せずに残しておくために使用され得る。最初に、末端のFokI部位を、デオキシシチジントリホスフェートとともにポリメラーゼを用いて一本鎖とする。次に、このフラグメントの二本鎖部分をメチル化し、その後、一本鎖の末端部を、4つのヌクレオシドトリホスフェートのすべての存在下でDNAポリメラーゼを用いて充填し、それによって、FokI部位を再生する。この手順は、FokI以外のエンドヌクレアーゼに一般化できることは明らかである。

40

オリゴヌクレオチドタグを、例えば、前述のように一本鎖にし、特定のハイブリダイゼーション用に調製した後、ポリヌクレオチドを、このタグの相補配列を含む微粒子と、タグおよびその相補物の間で完全にマッチした二重鎖の形成を促すような条件下で混合する。このような条件の作製に関して、文献には広範な指針がある。そのような指針を供給する典型的な文献としては、Wetmur、Critical Reviews in Biochemistry and Molecular Biology、26:227-259（1991）、Sambrookら、Molecular Cloning: A Laboratory Manual、第2版（Cold Spring Harbor Laboratory, New York, 1989）などがある。好ましくは、ハイブリダイゼーションの条件は、十分にストリンジェントであって、その結果、完全にマ

50

ッチした配列のみが安定な二重鎖を形成する。そのような条件下では、それらのタグを通じて特定のハイブリダイズするポリヌクレオチドが、微粒子に結合した相補的な配列に連結され得る。最後に、この微粒子を洗浄し、未連結および／またはミスマッチしたタグを有するポリヌクレオチドを除去する。

合成支持体として通常使用されるCPG微粒子を用いる場合、微粒子表面におけるタグ相補物の密度は、典型的には、いくつかの配列決定操作に必要とされるものよりも大きい。すなわち、結合したポリヌクレオチドを、種々の酵素によって連続的に処理することが要求される配列決定法の場合、ぎっしりと詰めこまれたポリヌクレオチドは、比較的嵩張る酵素がポリヌクレオチドに接触するのを妨げる傾向を有し得る。そのような場合、ポリヌクレオチドは、好ましくは、タグ相補物がポリヌクレオチドよりも十分に過剰になるように、例えば、10:1～100:1またはそれ以上になるように微粒子と混合される。これによって、微粒子表面におけるポリヌクレオチド密度が、酵素の接触を妨げるほどにならないことを確実にする。好ましくは、微粒子表面におけるポリヌクレオチド間の平均間隔は、30～100nmの程度である。標準的なCPG支持体およびBallotiniビーズ（固体ガラス支持体の一種）の選択比の指針は、MaskosおよびSouthern, Nucleic Acids Research, 20:1679-1684 (1992)に見ることができる。好ましくは、配列決定の適用に関しては、直径が20～50 μm の範囲にある標準的なCPGビーズは、約 10^5 個のポリヌクレオチドでロードされ、そして直径が5～10 μm 範囲にあるGMAビーズは、数万個（例えば、 $4 \times 10^4 \sim 6 \times 10^4$ 個）のポリヌクレオチドでロードされる。

好ましい実施態様においては、タグ相補物は、微粒子上において、コンビナトリアル的に合成される。従って、合成の終了時に、微粒子の複雑な混合物が得られる。これから、タグ化されたポリヌクレオチドをロードするために取り出される。微粒子サンプルのサイズは、いくつかの因子に依存する。その因子としては、タグ相補物レパートリーのサイズ、ロードされた微粒子の観察に用いられる装置の性質（例えば、その能力、同じタグ相補物による微粒子の複数コピーに対する許容度（すなわち「ビーズ重複」））などである。下の表は、微粒子サンプルのサイズ、微粒子直径、および様々な直径を有する微粒子の充填配列の近似的な物理ディメンジョンに関する指針を提供する。

微粒子直径	5 μm	10 μm	20 μm	40 μm
10 ⁵ 平方オングストロー				
μ当たり1個の微粒子		3×10^5	1.26×10^6	5×10^6
にロードされるポリヌ				
クレオチドの最大数				
10 ⁶ 個の微粒子単層の平	0.45×0.45cm	1×1cm	2×2cm	4×4cm
均面積				

微粒子のサンプルが、ある所定のタグ相補物を含むか、または所定のタグ相補物が複数のコピー中に存在する確率は、下記の表に示すように、ポアソン分布によって表わされる。

表VII

サンプル中の微粒子の数 (レパートリーサイズの 割合として表わす)	サンプル中のタグ相補物 の、レパートリーの割合	ユニークなタグ相補物を 結合させた、サンプル中 の微粒子の割合	サンプル中のもう一つ別 の微粒子として同じタグ 相補物を有するサンプル 中の微粒子の割合（「ビ ーズ重複」）
m	$1-e^{-m}$	$m(e^{-m})/2$	$m^2(e^{-m})/2$
1.000	0.63	0.37	0.18
0.693	0.50	0.35	0.12
0.405	0.33	0.27	0.05
0.285	0.25	0.21	0.03
0.223	0.20	0.18	0.02
0.105	0.10	0.09	0.005
0.010	0.01	0.01	

特異性の高い分類および選別

分類の動力学は、タグ相補物に対するオリゴヌクレオチドタグのハイブリダイゼーション速度に依存し、これは、ハイブリダイゼーション反応におけるタグの複雑度に依存する。従って、妥協点は、分類速度の増加がハイブリダイゼーション反応に関わるタグの複雑度を下げるといふ犠牲により達成され得るように、分類速度とタグの複雑度との間に存在する。以下に説明するように、この妥協の作用は、「選別」によって緩和され得る。

ハイブリダイゼーションの特異性は、サンプルを十分に小さく取ることによって高めることができる。そうすることによって、サンプル中の高比率のタグがユニークとなり、そしてサンプル中の実質的にすべてのタグの最も隣接するタグが、少なくとも2ワードしか変わらないことになる。この後の方の条件は、サンプルの数が、使用するレパートリーの大きさの約0.1%以下である多数のタグーヌクレオチド結合体を含むサンプルを採取することによって満たされる。例えば、タグが表IIから選択される8ワードで構築される場合、 8^8 の集合体、すなわち約 1.67×10^7 個のタグおよびタグ相補物が生成される。上述のタグーcDNA結合体のライブラリーにおいて、0.1%サンプルは、約16,700個の異なるタグが存在することを意味する。これを、レパートリーに等価な微粒子、すなわちこの例では、 1.67×10^7 個の微粒子サンプルに直接ロードされた場合、サンプリングされた微粒子のごくわずかな小部分しかロードされない。ロードされた微粒子の密度は、「選別」工程を設けることによって、例えば、さらに効率的な配列決定のために、高めることができる。この「選別」工程では、サンプリングされたタグーcDNA結合体を用いて、ロードされない微粒子から、ロードされた微粒子を分離する。従って、上の例では、「0.1%」サンプルは、わずか16,700個のcDNAしか含んでいなくとも、サンプリングおよび選別工程を、所望される程の多くのロードされた微粒子が蓄積するまで、繰り返すことができる。

選別工程は、タグーcDNA結合体のサンプルを供給することで実施され得る。その各々は、オリゴヌクレオチドタグの反対側の終端または遠位の終端に捕捉部分を含む。好ましくは、捕捉部分は、タグcDNA結合体から遊離され得る種類のものである。それによって、そのタグーcDNA結合体は、単一塩基配列決定法によって配列決定され得るからである。このような部分は、ビオチン、ジゴキシゲニン、またはリガンドのような三重鎖結合領域などを

含んでいてもよい。好ましくは、このような捕捉部分は、ビオチン成分を含む。ビオチンは、多くの標準的な技術を用いて、タグ-cDNA結合体に結合させることができる。PCRプライマー結合部位を含む適切なアダプターが、タグ-cDNA結合体に結合している場合、ビオチンは、サンプリング後の増幅時に、ビオチン化プライマーを用いて結合させることができる。あるいは、タグ-cDNA結合体がクローニングベクターの挿入体である場合、ビオチンは、適切な制限酵素で消化してタグ-cDNA結合体を切り出し、次に単離し、そしてタグの遠位部の突出鎖を、ビオチン化ウリジントリホスフェートの存在下で、DNAポリメラーゼによって充填してもよい。

タグ-cDNA結合体が捕捉された後、それは、多くの方法でビオチン部分から遊離させることができる。例えば、還元によって切断される（例えば、Hermanら、Anal. Biochem.、156:48-55（1986））、または光化学的に切断される（例えば、Olejnikら、Nucleic Acids Research、24:361-366（1996））、またはPCRプライマーに制限部位を導入することによって酵素的に切断される化学連結による。後の実施態様は、前述のタグ-ポリヌクレオチド結合体のライブラリーを考察することによって例示することができる。

5'-RCGACCA[C, W, W, W]₉GG[T]₁₉-cDNA-NNNR

GGT[G, W, W, W]₉CC[A]₁₉-rDNA-NNNYCTAG-5'

下記のアダプターを、これらのフラグメントの末端に連結して、PCRによる増幅を可能にし得る。

5'-XXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXYGAT

右アダプター

GATCZZACTAGTZZZZZZZZZZZZ-3'

ZZTGATCAZZZZZZZZZZZZ

左アダプター

ZZTGATCAZZZZZZZZZZZZ-5'-ビオチン

左プライマー

ここで、「ACTAGT」はSpeIの認識部位（これは、単一塩基配列決定に好都合な互い違いの切断部位を残す）であり、XおよびZは、それぞれのプライマーのアニーリングおよび解離の温度がほぼ同じでなるように選ばれたヌクレオチドである。アダプターの連結およびビオチン化プライマーを用いるPCR増幅の後に、結合体のタグを、T4 DNAポリメラーゼのエキソヌクレアーゼ活性により一本鎖とし、結合体を、タグ相補物を結合させた微粒子（例えば、等価なレパートリー）を有する微粒子と一緒にする。ストリンジェントな条件下で（タグの誤った結合を最少にするために）アニーリングした後、結合体を、好ましくは、タグ相補物に連結し、そしてロードされた微粒子を、アビジン化磁気ビーズによる捕捉、または同様の捕捉技術によって、ロードされない微粒子から分離する。

実施例に戻ると、この処理によって、異なるタグを有する約10,500（=16,700×0.63）個のロードされた微粒子の蓄積が得られる。これらは、SpeIによる切断によって、磁気ビーズから遊離させることができる。この処理を、微粒子とタグ-cDNA結合体の新しいサンプルについて40～50回繰り返すことにより、4～5×10⁵個のcDNAが、遊離した微粒子をプールすることによって蓄積し得る。次に、このプールした微粒子について、単一塩基配列決定技術によって同時に配列決定し得る。

サンプリングと選別工程とを何回繰り返すかを決定すること、すなわち、もっと一般的に

言えば、どのぐらいの数のcDNAを分析するかを決定することは、目的によって異なる。目的が、大量の、比較的ありふれた配列の中での変化、例えば、ある集団の中の5%以上となる変化をモニターすることである場合、比較的小さなサンプル、すなわち全集団のサイズの小さな割合のものが、相対量に関して統計的に有意な推定を可能にし得る。一方、珍しい配列、例えば、集団の0.1%以下となるものをモニターしたい場合、大きなサンプルが要求される。一般に、サンプルのサイズと、そのサンプルに基づく相対量の推定の信頼性との間には、直接的な関係がある。信頼できる統計的推定を行なうために、適切なサンプルのサイズに関しては、文献には広範な指針がある。例えば、Kollerら、*Nucleic Acids Research*, 23:185-191 (1994)、Good, *Biometrika*, 40:16-264 (1953) ; Bungeら、*J. Am. Stat. Assoc.*、88:364-373 (1993) などがある。好ましくは、 $3.0 \sim 3.5 \times 10^4$ 個の異なる配列を有する、 $10^5 \sim 10^8$ 個の独立クローンを含む一連のcDNAライブラリー分析に基づいて遺伝子発現に起こる変化をモニターする場合、各ライブラリーの分析に、少なくとも 10^4 個の配列のサンプルが蓄積される。より好ましくは、各ライブラリーの分析に、少なくとも 10^5 個の配列のサンプルが蓄積される。最も好ましくは、各ライブラリーの分析に、少なくとも 5×10^5 個の配列を有するサンプルが蓄積される。あるいは、サンプリングされる配列の数は、95%信頼限界が、集団サイズの0.1%を越えない場合、0.1%～5%の範囲内にある頻度で存在する配列の相対量を推定するに十分であることが好ましい。

単一塩基DNA配列決定

本発明は、DNA配列決定の従来の方法、例えば、Hultmanら、*Nucleic Acids Research*, 17:4937-4946 (1989) によって開示された方法とともに用いられ得る。しかし、複数のポリヌクレオチドの並行または同時の配列決定に関しては、接近したサイズのDNAフラグメントの電気泳動での分離またはペプチド配列決定におけるような分離分析手順による切断したヌクレオチドの分析のいずれも必要でないDNA配列決定方法論が好ましい。好ましくは、方法論は、処理および検出の連続的なサイクルによって、通常一度に1つ、配列中のヌクレオチドの段階的な同定を可能にする。このような方法論を、本明細書中で「単一塩基」配列決定方法という。単一塩基アプローチは、以下の参考文献で開示されている：Cheeseman、米国特許第5,302,509号；Tsienら、国際出願第W0 91/06678号；Rosenthalら、国際出願第W0 93/21340号；Canardら、*Gene*, 148:1-6 (1994) ；およびMetzkerら、*Nucleic Acids Research*, 22:4259-4267 (1994) 。

本発明で用いられるのに適切であり、DNAフラグメントの電気泳動での分離を必要としないDNA配列決定の「単一塩基」方法が、国際出願第PCT/US95/03678号に記載されている。簡単には、この方法は、以下の工程を包含する：(a) 突出鎖を有するポリヌクレオチドの末端にプローブを連結して連結複合体を形成する工程であって、このプローブは、ポリヌクレオチドの鎖に対して相補的な突出鎖を有し、かつ、このプローブは、ヌクレアーゼ認識部位を有する、工程；(b) この連結複合体から非連結プローブを除去する工程；(c) 連結プローブの同一性によりポリヌクレオチドの突出鎖における1つまたはそれ以上のヌクレオチドを同定する工程；(d) ヌクレアーゼで連結複合体を切断する工程；および(e) ポリヌクレオチド、またはその一部のヌクレオチド配列が決定されるまで(a)から(d)までの工程を繰り返す工程。

単一蛍光色素のような単一シグナル発生部分が、並行配列決定操作において、異なる空間的にアドレス可能 (addressable) な固相支持体 (例えば、固定化微粒子) に結合されたいくつかの異なる標的ポリヌクレオチドを配列決定する場合に用いられ得る。これは、4セットのプローブを提供することにより完成され得る。これらのプローブは異なる微粒子上の複数の標的ポリヌクレオチドに対して連続的に適用される。このようなプローブの例示的なセットを以下に示す：

セット1	セット2	セット3	セット4
ANNNN...NN N...NNTT...T*	dANNNN...NN d N...NNTT...T	dANNNN...NN N...NNTT...T	dANNNN...NN N...NNTT...T
dCNNNN...NN N...NNTT...T	CNNNN...NN N...NNTT...T*	dCNNNN...NN N...NNTT...T	dCNNNN...NN N...NNTT...T
dGNNNN...NN N...NNTT...T	dGNNNN...NN N...NNTT...T	GNNNN...NN N...NNTT...T*	dGNNNN...NN N...NNTT...T*
dTNNNN...NN N...NNTT...T	dTNNNN...NN N...NNTT...T	dTNNNN...NN N...NNTT...T	TNNNN...NN N...NNTT...T*

10

ここで、列記した各プローブは、上段の鎖の3'末端ヌクレオチドの同一性を固定させ、そして突出鎖の他の位置が各3マー順列のヌクレオチドで充填されるような $4^3=64$ オリゴヌクレオチドの混合物を表すか、または複雑度の少ないアナログを表す。列記したプローブはまた、「T*」で示される末端チミジンに結合されたシグナル発生部分を有する1本鎖ポリTテイルとともに示される。非標識プローブ上の「d」は、連結ブロッキング部分または3'ヒドロキシルの非存在を示し、これは非標識プローブが連結されるのを防止する。好ましくは、このような3'末端ヌクレオチドはジデオキシヌクレオチドである。本実施態様において、セット1のプローブが複数の標的ポリヌクレオチドに最初に適用され、そしてリガーゼで処理されて、標識プローブの3'末端アデノシンに相補的なチミジンを有する標的ポリヌクレオチドが連結される。非標識プローブが、同時に適用されて不適切な連結を最少化する。「A」で終了するプローブとの連結複合体を形成する標的ポリヌクレオチドの位置が、プローブ上に有する標識により生じるシグナルによって同定される。洗浄および切断後、セット2のプローブが適用される。この場合、「C」で終了するプローブとの連結複合体を形成する標的ポリヌクレオチドが位置により同定される。同様に、セット3および4のプローブが適用され、そして陽性シグナルの位置が同定される。4セットのプローブを連続して適用する本願のプロセスは、所望の数のヌクレオチドが標的ポリヌクレオチド上で同定されるまで継続される。明らかに、当業者は類似のセットのプローブを構築し得る。これらのプローブは、異なる長さの突出鎖、非標識プローブの連結をブロックするための異なる部分、プローブを標識するための異なる手段などを有するような、多くの変化を有し得る。

20

30

微粒子表面での酵素的処理および/または結合事象の観察装置

本発明の目的は、タグおよびその相補物の特異的ハイブリダイゼーションによって、微粒子の表面上の同一分子、特に、ポリヌクレオチド进行分类することである。一旦、このような分類が行われると、分子の存在またはそれらに対して行われる操作は、タグ化分子の性質、微粒子が、分離して、または「バッチ」で検出されるかどうか、繰り返される測定が所望であるかなどに応じて、多くの方法で検出され得る。代表的に、分類された分子は、例えば、薬物開発において、結合のためにリガンドに曝されるか、または、例えば、ポリヌクレオチド配列決定において化学的または酵素的処理に供される。これらの使用の両方において、多くの微粒子上でのこのような事象または処理に対応するシグナルを同時に観察することがしばしば望まれる。分類された分子を担持する微粒子（本明細書において「ロードされた」微粒子と称する）が、このような大規模な並行操作に役立つ（例えば、Lamら（上記）に示されている）。

40

好ましくは、発光シグナル（例えば、化学ルミネセンス、蛍光など）が、事象または処理を検出するために用いられるときは常に、ロードされた微粒子平面基板（例えば、ガラススライド）上に、（例えば、国際特許出願第PCT/US91/09217号、同第PCT/NL90/00081号および同第PCT/US95/01886号に記載されているような）走査システムでの検査のために広がっている。走査システムは、再現性よく基板を走査し、かつ、座標系によって、所定の領域に各微粒子の位置を規定し得るべきである。ポリヌクレオチド配列決定適用において、微粒子の位置同定は、連続的な走査工程において繰り返し可能であることが重要である。

50

このような走査システムは、市販の構成要素（例えば、1つまたはそれ以上の光電子増倍管、あるいは、CCDアレイならびに、例えば、蛍光シグナルを励起、収集および分類するための適切な光学機器を備える検出システムとともに用いられるデジタルコンピュータによって制御されるx-y移動テーブル）から構築され得る。いくつかの実施態様において、共焦光学システムが所望であり得る。4色配列決定での使用に適切な走査システムの一例が、図5に模式的に示されている。基板300（例えば、固定化微粒子を有する顕微鏡スライド）が、x-y移動テーブル302上に置かれ、これが、種々の市販のパーソナルコンピュータ（例えば、486-ベースの機械またはApple Computer（Cupertino, CA）から入手可能なPowerPCモデル7100または8100）のうちのいずれかであり得る適切にプログラムされたデジタルコンピュータ304に接続され、そしてそれにより制御される。テーブル移動およびデータ収集機能のためのコンピュータソフトウェアは、National Instrumentsから入手可能なLab Windowsのような市販の実験ソフトウェアによって提供され得る。

10

基板300およびテーブル302は、光を集め、基板300に固定した微粒子へ伝達し得る1つまたはそれ以上の対物レンズ308を有する顕微鏡306と動作が連動する。好適には、レーザーである光源312からの励起光線310が、光線スプリッター314（例えば、二色性ミラー）へ向けられ、これが、光線を顕微鏡306および対物レンズ308を通して再度向けられ、続いて、光線が基板300上に集中される。レンズ308は、微粒子から発光された蛍光316を集め、それを光線スプリッター314を通してシグナル分配光学機器318へ向け、それが、続いて、蛍光をいくつかの蛍光特性（例えば、強度、寿命など）を電気シグナルに変換するための1つまたはそれ以上の適切な光学電気デバイスへ向ける。シグナル分配光学機器318は、帯域通過フィルター、ファイバー光学機器、回転ミラー、固定化位置ミラーおよびレンズ、回折格子などのような当該分野において標準的な種々の構成要素を備える得る。図5に示すように、シグナル分配光学機器318は蛍光316を4つの別々の光電子増倍管330、332、334および336へ向け、次いで、これらの出力は、プリアンプおよびフォトンカウンター350、352、354および356に向けられる。フォトンカウンターの出力は、コンピュータ304によって収集され、そこで、蓄積され、分析され、ビデオ360において観察され得る。あるいは、シグナル分配光学機器318は、蛍光シグナル318をCCDアレイ上に向ける回折格子であり得る。

20

走査の際の位置付けの安定性および再現性が、間隔が密接な微粒子を分離するための解像度を、かなりの程度まで、決定する。好適には、走査システムは、間隔が密接な（例えば、1粒子直径またはそれ以下分だけ分離されている）微粒子を解像し得るべきである。従って、ほとんどの適用に対して（例えば、CPG微粒子を用いて）、走査システムは、少なくとも10~100mのオーダーで対象物を解像する能力を有するべきである。いくつかの実施態様において、さらに高い解像度が所望され得るが、解像度が増すにつれて、基板を完全に走査するのに必要な時間が増す。従って、いくつかの実施態様において、速度と解像度との間で妥協が必要であり得る。走査時間の増加は、微粒子が配置されていることが知られている（例えば、最初のフル走査から）位置のみを走査するシステムによって達成され得る。好適には、微粒子サイズおよび走査システム解像度が、1 cm²当たり約1万~10万の間の微粒子の密度で平面上にランダムに配置される蛍光標識された微粒子の解像度を可能にするように選択される。

30

40

配列決定適用において、ロードされた微粒子は、種々の方法で基板の表面に固定化され得る。固定は、微粒子が、顕著な損失なく、試薬への曝露および洗浄の連続的なサイクルを行うのに十分に強力であるべきである。基板がガラスであるとき、その表面は、市販の試薬（例えば、Pierce Chemical）を用いて、アルキルアミノリンカーで誘導体化され得、これは次に再び従来のケミストリーを用いて、アビジンに架橋され、アビジン化された表面を形成し得る。ビオチン部分は、多くの方法でロードされた微粒子に導入され得る。例えば、ポリヌクレオチドにタグを結合するために用いられるクローニングベクターの画分（例えば、10~15パーセント）が、タグの反対側のポリヌクレオチドの末端でポリヌクレオチド挿入物の隣接する（消化において粘着末端を提供する）唯一の制限部位を含有する

50

ように操作される。この部位は、微粒子にロードするためのポリヌクレオチドおよびタグで切り出される。ロード後、ロードされたポリヌクレオチドの約10～15パーセントが、微粒子表面から遠位の唯一の制限部位を有する。関連の制限エンドヌクレアーゼで消化後、ビオチン部分を含有している適切な二本鎖アダプターが、粘着末端に連結される。次いで、得られた微粒子が、アビジン化されたガラス表面に広げられ、そこで、ビオチン-アビジン結合を介して固定される。

あるいは、そして好適には、連結による配列決定が用いられる場合、最初の連結工程において、プローブの混合物が、ロードされた微粒子に適用される：プローブの一面分は、配列決定方法に必要なようなII型制限認識部位を含み、そしてプローブの一面分は、このような認識部位を有さないが、その代わりに、その非連結末端においてビオチン部分を

10

含む。好適には、混合物は、約10～15パーセントのビオチン化プローブを含む。なお別の実施態様において、DNAがロードされた微粒子が、ガラス基板に適用されると、DNAは、数時間（例えば、24時間）のインキュベーションで、ガラス表面に非特異的に吸着され、微粒子の顕著な損失なく、反復される試薬への露出および洗浄を可能とするために十分に強力な結合を形成し得る。好適には、このようなガラス基板は、フローセル（flow cell）であり、これは、ガラススライドにエッチングされた流路を含み得る。好適には、このような流路は、流体がそこを通過して吸い上げられ得るように閉じられ、かつ、微粒子の単層が規定された観察領域内に閉じ込められるように、微粒子の直径と十分に近い深さを有する。

並行配列決定

20

本発明のタグ化システムは、数千ベースまでの長さのポリヌクレオチドを配列決定するための単一塩基配列決定方法とともに用いられ得る。タグ化システムは、標的ポリヌクレオチドの何千ものフラグメントが、1つまたはそれ以上の固相支持体上に分類され、同時に配列決定され得ることを可能にする。この方法の好適な実施によると、各分類されたフラグメントの一部分が、上記のような走査システムまたは画像分析システムに関連する、顕微鏡スライドのような共通の基板に固定された何千ものロードされた微粒子のそれぞれにおいて、段階的に配列決定される。配列決定されたフラグメントの部分のサイズは、いくつかの要因、例えば、生成および分類されたフラグメントの数、標的ポリヌクレオチドの長さ、用いられた単一塩基方法の速度および正確さ、同時にモニターされ得る微粒子および/または別個の領域の数などに依存する。好適には、12～50塩基が、各微粒子または領域で同定され、さらに好適には、18～30塩基が、各微粒子または領域で同定される。この情報により、標的ポリヌクレオチドの配列が、それらの重複する領域を介して12～50塩基のフラグメントを対照することによって決定される（例えば、米国特許第5,002,867号に記載されている）。以下の引例は、所定の長さの標的ポリヌクレオチドを首尾良く再構築するために配列決定されなければならないフラグメントの部分を決定する際のさらなる指針を提供する：LanderおよびWaterman、Genomics、2:231-239（1988）：Drmanacら、Genomics、4:114-128（1989）：Bains、DNA Sequencing and Mapping、4: 143-150（1993）：Bains、Genomics、11:294-301（1991）：Drmanacら、J.Biomolecular Structure and Dynamics、8: 1085-1102（1991）：およびPevzner、J.Biomolecular Structure and Dynamics、7: 63-73（1989）。好適には、標的ポリヌクレオチドの長さは、1キロベース～50キロベースの間である。より好適には、長さは、10キロベース～40キロベースの間である。LanderおよびWaterman（上記）は、配列決定されるフラグメントの数（すなわち、サンプルサイズ）、各フラグメントから得られた配列情報の量、および標的ポリヌクレオチドが、ギャップがない部分的な配列、すなわち、「アイランド」から再構築され得る標的ポリヌクレオチドの可能性の間の相互関係に関する指針を提供する。本発明において、所定のサンプルサイズおよびフラグメント配列のサイズで得られ得る最大のポリヌクレオチドサイズを以下に示す。

30

40

サンプルのサイズ

最大標的ポリヌクレオチドのおおよその長さ

	30塩基／フラグメント	50塩基／フラグメント
1,000	3キロベース	4キロベース
10,000	22キロベース	32キロベース
20,000	40キロベース	65キロベース
30,000	60キロベース	85キロベース
100,000	180キロベース	300キロベース

10

フラグメントは、最少の重複で標的ポリヌクレオチドをカバーするフラグメントのセットを生成しようとする、いわゆる「直接的」なアプローチ、およびランダムに重複するフラグメントが生成される、いわゆる「ショットガン」アプローチを含む、種々の方法で標的ポリヌクレオチドから生成され得る。好適には、フラグメント生成についての「ショットガン」アプローチが、その簡便性および固有の縮重のために用いられる。例えば、標的ポリヌクレオチドをカバーする、ランダムに重複するフラグメントは、以下の従来の「ショットガン」配列決定プロトコルで生成される（例えば、Sambrookら（上記）に開示されている）。本明細書に用いられるような、この文脈での「カバーする（cover）」は、標的ポリヌクレオチド配列の各部分が、生成されたフラグメントの各サイズ範囲（例えば、全

20

てのフラグメントは長さが、100塩基対～200塩基対の範囲である）で表されることを意味する。簡潔には、適切なクローニングベクター（例えば、ファージ）における挿入物としての標的ポリヌクレオチドから始まって、ベクターは拡大され、精製され、そして適切な制限酵素で消化され、約10～15 μgの精製された挿入物を生じる。代表的に、プロトコルは、1マイクログラムの出発DNA当たり約500～1000のサブクローンをもたらす。挿入物は、調製ゲル電気泳動によってベクターフラグメントから分離され、従来の方法によってゲルから取り出され、そしてTE（Tris-EDTA）のような標準的な緩衝液に再懸濁される。ベクターから挿入物を切り出すために選択される制限酵素は、好適には、挿入物に適合性の粘着末端を残し、これにより、挿入物が、ランダムに重複するフラグメントを生成するために、調製において自己連結され得る。Sambrookら（上記）に説明されるように、環状化

30

DNAは、以下に用いられる断片化方法における直鎖状DNAに比べて、フラグメントのよりよいランダム分配を与える。挿入物を例えば、T4リガーゼで従来のプロトコルを用いて、自己連結した後、精製された連結挿入物が、標準的なプロトコル（例えば、超音波処理またはMn⁺⁺の存在下におけるDNaseI消化）によって断片化される。断片化後、フラグメントの末端が、例えば、Sambrookら（上記）に記載されるように修復され、そして修復フラグメントは、ゲル電気泳動を用いて、サイズによって分離される。300～500塩基対の範囲のフラグメントが選択され、従来の手段によってゲルから溶出され、上記のように、タグを有するベクター中に連結され、タグフラグメント結合体のライブラリーを形成する。

30

上記のように、数千のタグフラグメント結合体を含有するサンプルが、ライブラリーから採取され、そして拡大され、その後、タグフラグメント挿入物がベクターから切り出され、そして上記のように、微粒子上のタグ相補物への特異的なハイブリダイゼーションのために調製される。標的ポリヌクレオチドのサイズに応じて、複数のサンプルが、タグフラグメントライブラリーから採取され、別々に拡大され、微粒子上にロードされ、そして配列決定され得る。上記のように、選択された重複の数は、サンプル中に表されたタグレパートリーの画分に依存する（トリプル（同じタグを有する3つの異なるポリヌクレオチド）または上記を得る可能性は、安全に無視し得る）。上記のように、サンプル中の重複の可能性は、ポアソン分布 $p(\text{重複}) = m^2 e^{-m} / 2$ （ここで、 m は、サンプル中のタグレパートリー画分である）から評価され得る。以下の表VIは、所定のタグサイズ、サンプルサイズおよびレパートリーの多様性についてサンプル中で重複を得る可能性を挙げている。

40

50

表VIII

8ワードセットから のタグのワード数	タグレパートリーの サイズ	サンプルのサイズ	サンプリングされた レパートリーの画分	重複の可能性
7	2.1×10^6	3000	1.43×10^{-3}	10^{-6}
8	1.68×10^7	3×10^4	1.78×10^{-3}	1.6×10^{-6}
		3000	1.78×10^{-4}	1.6×10^{-8}
9	1.34×10^8	3×10^5	2.24×10^{-3}	2.5×10^{-6}
		3×10^4	2.24×10^{-4}	2.5×10^{-8}
10	1.07×10^9	3×10^6	2.8×10^{-3}	3.9×10^{-6}
		3×10^5	2.8×10^{-4}	3.9×10^{-8}

いずれの場合においても、次いで、ロードされた微粒子が、好適には、アビジン-ビオチン結合を介して、ガラス顕微鏡スライド上に分散および固定される。好適には、それぞれのランダムフラグメントの少なくとも15~20のヌクレオチドが、単一塩基方法で同時に配列決定される。次いで、標的ポリヌクレオチドの配列が、アセンブリングコンティグ (assembling contig) に用いられるアルゴリズムと類似の、または上記の文献に開示されたようなハイブリダイゼーションによる配列決定のために開発されたようなアルゴリズムを用いて、それらの重複する部分によって、ランダムなフラグメントの部分的な配列を照合することによって再構築される。

本発明の方法を実施するためのキット

本発明は、本発明の種々の実施態様を実施するためのキットを含む。好適には、本発明のキットは、固相支持体に結合されたタグ相補物のレパートリーを含む。さらに、本発明のキットは、対応するタグのレパートリー（例えば、分類されるべきポリヌクレオチドを増幅するためのプライマーとして、または分類されるべきポリヌクレオチドを増幅するためにもまた用いられ得るクローニングベクターの要素として）を含み得る。好適には、タグ相補物のレパートリーは、微粒子に結合される。キットはまた、酵素的プロセッシング、検出ケミストリー（例えば、蛍光タグまたは化学ルミネセンスタグなど）のための適切な緩衝液、使用説明書、プロセッシング酵素（例えば、リガーゼ、ポリメラーゼ、トランスフェラーゼなど）を含み得る。配列決定のための重要な実施態様において、キットはまた、プロセッシングのためにロードされた微粒子を固定するための、アビジン化された顕微鏡スライドのような基板を含み得る。

cDNAライブラリーにおける新規なポリヌクレオチドの同定

cDNAライブラリーにおける新規なポリヌクレオチドは、上記のように微粒子に結合されたcDNA分子のライブラリーを構築することによって、同定され得る。次いで、ライブラリーの大きな画分、またはライブラリー全体でさえも、並行して部分的に配列決定され得る。mRNAの単離、およびおそらく、Soaresら、Proc. Natl. Acad. Sci.、91:9228-9232 (1994)、または同様の文献によって教示されるように、その集団の正規化の後、以下のプライマーが、従来のプロトコルを用いて、逆転写酵素で第1鎖の合成のためのポリAテールにハイブリダイズされ得る：

5'-mRNA- [A]_n-3'
[T]₁₉-[プライマー部位]-GG[W,W,W,C]₉ACCAGCTGATC-5'

ここで、[W,W,W,C]₉は、上記のようなタグを表し、「ACCAGCTGATC」は、二本鎖形態で制限部位を形成する任意の配列であり、そして「プライマー部位」は、PCRによって目的のポリヌクレオチドを増幅するためのプライマー結合部位として後に用いられる、ライブラリーのすべてのメンバーに共通の配列である。

従来の技術による逆転写および第2の鎖の合成の後、二本鎖フラグメントが、上記のようにクローニングベクターに挿入され、そして増幅される。次いで、増幅されたライブラリーがサンプリングされ、そしてサンプルが増幅される。増幅されたサンプルからのクローニングベクターが単離され、そしてタグ化されたcDNAフラグメントが切り出され、そして精製される。上記のようなポリメラーゼでタグを一本鎖にした後、フラグメントはメチル化され、そして本発明によって微粒子上に分類される。好適には、上記のように、クローニングベクターは、タグ化されたcDNAが、分類および微粒子への連結後に好適な単一塩基方法による即座の配列決定を可能にする、FokIのようなエンドヌクレアーゼで切り出され得るように構築される。

次いで、段階的な配列決定が、本発明に従って、ライブラリー全体またはライブラリーの1つまたはそれ以上の大きな画分で同時に行われ、ついには、十分な数のヌクレオチドが、ライブラリーが由来する生物のゲノムに特有の表示のために各cDNAについて同定される。例えば、ライブラリーが哺乳動物のmRNAに由来するなら、14~15ヌクレオチドの長さのランダムに選択された配列が、代表的な哺乳動物のゲノムの2000~3000メガ塩基の間で特有の表示を有すると予想される。もちろん、細菌または他の下等生物に由来するライブラリーに特有の表示には、ずっと少ないヌクレオチドの同定で十分である。好適には、少なくとも20~30ヌクレオチドが、特有の表示を確実にし、かつ、以下に記載するように、適切なプライマーの構築を可能にするために同定される。次いで、表にされた配列は、特有のcDNAを同定するために、既知の配列と比較され得る。

次いで、特有のcDNAは、従来の技術（例えば、プライム部位およびその配列が決定されたcDNAの部分に関するプライマーで製造されたPCRアンプリコンからのプローブを構築する技術）によって単離される。次いで、プローブは、従来のスクリーニングプロトコルを用いてライブラリー中のcDNAを同定するために用いられ得る。

新しいcDNAを同定するための上記方法はまた、確立された測定においてかまたは動的に変化する集団に関してのいずれかで、mRNA集団をフィンガープリントするために用いられ得る。部分的な配列情報は、上記方法で記載された別々の微粒子に結合されたcDNAの大きなサンプル（例えば、10,000~100,000またはそれ以上）から同時に得られる。部分配列の度数分布は、異なる細胞または組織型由来のmRNA集団、ならびにガンのような疾患組織由来のmRNA集団を同定し得る。このようなmRNAフィンガープリントは、疾患状態をモニターおよび診断する（例えば、国際特許出願PCT/US95/21944、これは同じ目的のための発現配列タグ（EST）の使用を記載している）ために有用である。

分類されたポリヌクレオチドでロードされた微粒子でのサイクル配列決定

本発明に従い、標識DNAフラグメントの生成と分離とを必要とする従来の配列決定法を用いて、平行的配列決定を実施し得る。特に、均一なテンプレート集団でロードされた単離微粒子を用いて、サイクル配列決定法により、標識された伸長産物を生成することができる。サイクル配列決定法とは、DNA配列決定に対する基本的なSanger法の周知の変法であって、下記の文献に十分に記載される：Craxton, Methods, 第2巻（1991年2月）；Wozny, 欧州特許公報0409 078 A2（1993年1月23日）；Fuller, 国際特許出願PCT/US92/07303；およびFuller, 国際特許出願PCT/US94/03264。簡単に言うと、標準的な配列決定反応混合物において、熱的に安定なポリメラーゼを用い、それによって、同じテンプレート上で、反復する伸長反応を実施され得る。これによって、少量のテンプレートから、電気泳動による分離後の検出のために十分な量の伸長産物を得ることができる。典型的には、サイクル配列決定法は、次の工程を包含する：（a）配列決定反応混合物に、テンプレート、プライマー、ヌクレオシドトリホスフェート、鎖停止ヌクレオシドトリホスフェート、および熱的に安定なDNAポリメラーゼを供給する工程、（b）テンプレートを変性させる工程、（c）プライマーを変性テンプレートにアニーリングさせる工程、（d）プライマーを伸長して、伸長産物を形成する工程、および（e）工程（b）~（d）を、分離時に検出され得るように、十分量の伸長産物が蓄積するまで繰り返す工程。繰り返すサイクル数は、数多くの要因に依存する。そのような要因として、出発テンプレートの量と質、用いる検出装置、用いる分離装置などである。従来実施されているところでは、伸長サイクルは、典

型的には、10～100回繰り返される。テンプレート量は、数十フェムトモルというごく少量から、数十ピコモルの範囲である。変性工程は、反応混合物を、92～95℃の範囲内の温度で加熱して行なわれ、アニーリング工程は、35～75℃の範囲内の温度で行なわれ、伸長工程は、65～85℃の範囲内の温度で、熱的に安定なDNAポリメラーゼ、例えばTaqまたはVent（それぞれ、Perkin-Elmer Corp. , Norwalk, CT、およびNew England Biolabsから入手可能）によって行なわれる。

タグ相補物は、Albretsenら、Anal. Biochem.、189:40-50（1990）により記載されているように、磁気微粒子上に調製してもよい。そうすることによって、数フェムトモルのタグ相補物は、4.5μm直径の磁気ビーズにロードされ得る。タグ相補物は、5'末端または3'末端のいずれかで、微粒子に結合させてもよい。5'末端で結合させた場合、テンプレートは、3'末端でのタグの特異的ハイブリダイゼーションを通じて分類され得る。本実施態様においては、テンプレートは、下記に示すように、5'末端にプライマー相補物を有する。

3' - [オリゴヌクレオチドタグ] - [テンプレート] - [プライマー相補物] - 5'

次に、タグ相補物を、テンプレートの長さに伸長し、それによって、微粒子に共有結合的に結合されるテンプレートの相補物が得られるようにする。テンプレートを加熱によって除き、そして微粒子を洗浄する。微粒子を、例えば、フロー分類（flow sorting）によって分離した後、プライマーのアニーリング、伸長、および変性の工程の繰り返しサイクルを実施する。

タグ相補物は、その3'末端で微粒子に結合し、それによって微粒子に対して直接的な共有従来の合成が可能である場合、オリゴヌクレオチドタグおよびプライマー相補物の順序は、下記に示すように、逆になる。

5' - [オリゴヌクレオチドタグ] - [テンプレート] - [プライマー相補物] - 3'

タグ相補物の5'末端はまた、例えば、市販の試薬を用いてリン酸化される。オリゴヌクレオチドタグによる特異的ハイブリダイゼーションの後、プライマーを、テンプレートの3'末端においてプライマー相補物にアニーリングし、3'→5'エクソヌクレアーゼ活性を持たないDNAポリメラーゼにより伸長させる。次に、伸長反応により残されたニックを連結し、元のテンプレートを加熱によって除く。微粒子を分離した後、サイクル配列決定を上記のように実施することができる。

ロードされた微粒子の分離は、フロー分類によって実行することができる。この方法では、懸濁微粒子を、一列でノズルを通過して、ジェット流となって流出するように仕組む。このジェット流は、規則的な一連の荷電した小滴に分散され、基板上の所定の標的容器、ウェル、またはその他の反応位置に向けられる。微粒子は、ジェット流において、光散乱によって都合よく検出され、そして散乱度を用いて、小滴が、微粒子を全然含んでいないか、1個含んでいるか、複数個含んでいるかが決定される。このようなフロー分類、および配列決定試薬の送達に特に有用な装置は、Brennan、国際特許出願PCT/US94/05896に開示されている。個々のロードされた微粒子が、適切な配列決定試薬を含む複数の反応部位またはウェルに配られたならば、反応の集合を熱的に一緒にサイクルさせて、伸長産物を生成することができる。サイクルが完了した後、伸長産物を電気泳動で分離する。好ましくは、電気泳動分離は、ゲルを含まない分離媒体による毛細管電気泳動を用いて実施される。この方法によれば、伸長産物の好都合なロードおよび急速な分離を可能にする。実質的に同時に、多数のサンプルを4色蛍光によって検出することのできる装置もまた、市販されている。例えば、MathiesおよびHuang、Nature、359:167-169（1992）；Huangら、Anal. Chem.、64:2149-2154（1992）；Huangら、Anal. Chem.、64:967-972（1992）などに開示されているタイプである。好ましくは、同時に数千サイクルの配列反応が実行されることが好ましい。さらに好ましくは、テンプレートの混合体が、1000～10,000の間の異なるタイプのオリゴヌクレオチドタグレパートリーを有する微粒子の集団において分類される。

遺伝子型分析用多座プローブの分類

多くの疾患状態および／または病気に対する感受性が、複雑な遺伝的形質および／または

10

20

30

40

50

変異パターン（例えば、HLAタイプ、多くの癌におけるp53遺伝子の変異パターン、嚢胞性線維症遺伝子、レッシューナイアン（Lesch-Nyhan）症候群、ダッケン（Duchenne）筋ジストロフィーなど）と関連する。Landerら、Science、265:2037-2048（1994）；Collins、Science、256:774-779（1992）；Tsuiら、国際特許出願PCT/CA90/00267；Hedrumら、Biotechniques、17:118-129（1994）；Santamariaら、国際特許出願PCT/US92/01675；Chamberlainら、Nucleic Acids Research、16:11141-11156（1988）など。従来、このような複雑な遺伝的形質のための好都合なアッセイを構築する一つの方法は、Chamberlainら（上に引用）およびGrossmanら、国際特許出願PCT/US93/03229に記載されているような、いわゆる多重PCRまたは多重連結アッセイを用いることであった。通常、このような方法は、同じ反応混合物中の多数の遺伝子配列を同時に増幅し、その後、目的の配列の特異的検出を要求する。本発明のオリゴヌクレオチドタグは、このようなアッセイにおいて増幅される遺伝子配列を同定するための、簡便で好都合な手段を提供し得る。その最も簡便な型式において、本発明のこの実施態様は、多重PCRで使用されるPCRプライマーにオリゴヌクレオチドタグを結合させることによって実施され得る。前述したように、一対の片方のプライマーは、オリゴヌクレオチドタグを有し、同じ一対の他方のプライマーは、捕捉部分を有することによって、うまく増幅された配列を単離し、次に遊離することができる。遊離後、配列は、あらかじめ定められた空間アドレスに結合させた一組のタグ相補物を有する固相支持体に与えられる。次に、タグの特異的ハイブリダイゼーションのパターンを検出して、サンプルの遺伝子型を同定する。

好ましい実施態様では、PCRを用いて、多数の標的部位、すなわち変異または疾患関連配列が存在する多数の部位を含む、目的の遺伝子配列を増幅する。好ましくは、わずか2個またはきわめて少数対のプライマーを用いて、標的配列を増幅し、標的長さを釣り合わせることで、プライマーのアニーリング温度などのような、多重PCRに関わる困難を避ける。増幅後、特定の遺伝子型が、Grossmanら（上に引用）およびGrossmanら、米国特許第5,514,543号に記載されているのと同様の様式で検出される。上記文献は、PCRおよび連結反応条件、連結プローブのサイズなどの選択についての指針を提供する。上記文献では、標的配列が同様に増幅され、その後、連結プローブの集団を、DNAリガーゼの存在下に与える。連結プローブは、分析対象であるサンプル中に存在し得る標的に対して共に相補的な二つの異なる配列から成る：1つは電気泳動的移動度の改変剤に結合され、他方は蛍光標識に結合される。この二つのプローブが、サンプル中の標的配列と完全な二重鎖を形成する場合、それらのプローブは、移動度改変部分が、標的に相補的な結合配列を通じて、蛍光標識に結合されるように連結される。次に、この混合物の成分を電気泳動的に分離し、それによって、ゲル上の蛍光バンドのパターンが、サンプル中に存在する標的の遺伝子型を呈示する。図3に示したように、本発明のオリゴヌクレオチドタグを、電気泳動移動度改変剤の代わりに用いてもよい。この場合、空間的分離は、連結された配列を固相支持体の特定位置に分類することによって実施され得る。図3に戻ると、標的配列（200）は、PCRで増幅することが好ましく、その後、連結プローブの集団（206～216）を、変性したアンプリコンに与える（204）。本実施態様において、連結プローブは、オリゴヌクレオチドタグ（206）、標的配列に対して相補的な第1の配列（208）、標的配列に対して相補的であり、かつ第1の配列と連続する第2の配列（210）（従って、両方が完全に標的配列に対して相補的である場合、それらは連結され得るような）、および信号発生手段（214）を備える尾部（212）を含む。信号発生手段（214）は、蛍光標識であることが好ましい。好ましくは、連結プローブの第1と第2の配列は、DNAリガーゼによって連結される。従って、断端配列（216）の5'末端は、例えばリン酸化試薬によって、リン酸化されなければならない。リン酸化試薬は、例えば、Urdeaら、米国特許第5,332,845号に記載される。連結プローブとリガーゼとを投与した後、標的配列と完全に適合する二重鎖を形成するプローブは、共有結合的に連結される（218および220）。次に、プローブ標的二重鎖を変性し、そして固相支持体に与える（222）。支持体は、タグ t_1 から t_k までの一つ一つに対して、十分に規定された空間位置に結合したタグ相補物を有している。非特異的に結合した配列を洗浄除去した後、蛍光標識に連結された、オリゴヌクレオチドタグ t_i および t_j のタ

10

20

30

40

50

一般に、本発明のこの実施態様は、標的ポリヌクレオチド中の、複数の選択された標的配列の存在または非存在を検出するのに下記の工程を用い得る：（１）標的ポリヌクレオチドに、複数の連結プローブを付加する工程、ここで、各連結プローブは、標的オリゴヌクレオチド中の標的配列の選ばれた一つに隣接する部分に対して、配列が相補的である第１のオリゴヌクレオチドと第２のオリゴヌクレオチドとを含み、第１のオリゴヌクレオチドは、結合したオリゴヌクレオチドタグを有し、各オリゴヌクレオチドタグは、最少に交差ハイブリダイズする同じセットから選ばれ、そして各連結プローブは、異なるオリゴヌクレオチドタグを有する；（２）連結プローブを、標的ポリヌクレオチドとハイブリッドさせる工程；（３）ハイブリダイズした第１および第２のオリゴヌクレオチドを、第１および第２のオリゴヌクレオチドが隣接標的配列と完全にマッチした二重鎖を形成する場合には必ず、第１および第２のオリゴヌクレオチドを連結するのに有効な条件下で処理する工程；（４）連結した第１および第２のオリゴヌクレオチドを、未連結の第１および第２のオリゴヌクレオチドから分離する工程；（５）連結した第１および第２のオリゴヌクレオチドを、そのオリゴヌクレオチドタグと、それぞれの相補物とを特異的にハイブリダイズさせることによって分類する工程、ここで、このそれぞれの相補物は、１個以上の固相支持体の空間的に離れた領域に、実質的に同一のオリゴヌクレオチドの均一な集団として結合されている；および（６）１個以上の固相支持体の上に連結された第１および第２のオリゴヌクレオチドの存在または非存在により、選択した標的配列の有無を検出する工程。

pUC19由来の複数の標的ポリヌクレオチドの分類

5'-pTCGACC (w₁) (w₂) (w₃) (w₄) (w₅) (w₆) (w₇) (w₈) (w₁)A
GG(**) (**) (**) (**) (**) (**) (**) (**) (**) (**) TTCGAp-5'

5'-PTCGACC (w6) (w7) (w8) (w1) (w2) (w6) (w4) (w2) (w1) A
GG (**) (**) (**) (**) (**) (**) (**) (**) (**) (**) TTCGAp-5'

5'-pTCGACC (w₃) (w₂) (w₁) (w₁) (w₅) (w₈) (w₈) (w₄) (w₄)A
GG(**) (**) (**) (**) (**) (**) (**) (**) (**) TTCGAp-5'

ここで、「p」はモノホスフェートを示し、 w_i は表IIで定義されるサブユニットを表し、そして用語「(*)」はそれぞれの相補物を表す。pUC19を、SalIおよびHindIIIで消化し、大きなフラグメントを精製し、そしてタグ1、タグ2およびタグ3で別々に連結して、それぞれ、pUC19-1、pUC19-2、pUC19-3を形成する。この3つの組換え体を、別々に増幅し単離し、その後、pUC19-1を、HindIIIおよびAatIで消化し、pUC19-2をHindIIIおよび

SspIで消化し、pUC19-3をHindIIIおよびXmnIで消化する。小さなフラグメントを従来のプロトコルを用いて単離して、それぞれ、約250、375および575塩基対の長さの3つの二本鎖フラグメントを得る。それぞれは、タグに隣接する陥没した3'鎖および反対側の末端には平滑または3'突出鎖を有する。約12nmolの各フラグメントを、製造者が推薦する、33M デオキシシトシントリホスフェートを含有する反応緩衝液中で、5ユニットのT4 DNAポリメラーゼと混合する。反応混合物を37℃で30分間インキュベートし、その後、氷の上に置くことによって反応を止める。次いで、フラグメントを、従来の手段で精製する。

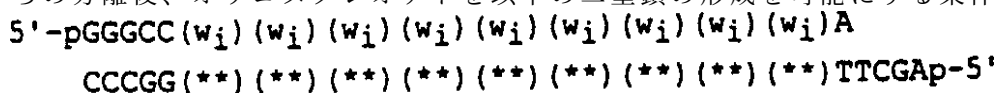
CPG微粒子（粒子サイズ37~74mm、孔サイズ500オングストローム、Pierce Chemical）を、MaskosおよびSouthern、Nucleic Acids Research、20:1679-1684（1992）によって開示されたリンカーで誘導体化する。3つのアリコートに分離後、タグ1、タグ2およびタグ3の相補物を、従来の自動化DNA合成機（例えば、model 392 DNA合成機（Applied Biosystems、Foster City、CA））を用いて、微粒子上で合成する。異なって誘導体化された微粒子のそれぞれ約1mgを、別個の容器に入れる。

pUC19-1、pUC19-2、およびpUC19-3から切り出され、T4 DNAポリメラーゼ処理したフラグメントを、製造者推薦のTaq DNAリガーゼ（New England Biolabs）の用の緩衝液50Lに再懸濁する。次いで、混合物を、誘導体化CPG微粒子をそれぞれ1mg含有している3つの容器に等分する。5ユニットのTaq DNAリガーゼを各容器に添加し、その後、それらを55℃で15分間インキュベートする。氷の上に置くことによって反応を止め、そして微粒子を、遠心分離およびTEへの再懸濁を繰り返すことによって数回洗浄する。最後に、微粒子をNdeI反応緩衝液（New England Biolabs）で再懸濁し、結合したポリヌクレオチドを消化する。微粒子からの分離後、NdeI消化によって放出されたポリヌクレオチドフラグメントを、Sequenase DNAポリメラーゼおよびフルオレセイン標識チミジントリホスフェート（Applied Biosystems、Foster City、CA）でインキュベートすることにより、蛍光標識する。次いで、そのフラグメントを、Applied Biosystems model 373 DNA配列決定機を用いて、非変性ポリアクリルアミドゲル上で別々に分析する。

実施例 2

SV40フラグメントの並行配列決定

表IIから選択される9つの4-ヌクレオチドサブユニットからなる36マーのタグのレパートリーを、上記のような、分裂または混合手段によって、タグおよびタグ相補物を別々に合成することによって調製する。SmaI/HindIII消化のM13mp19に連結可能なように、レパートリーを合成する。従って、実施例Iと同様に、1セットのオリゴヌクレオチドはAの添加で開始し、9回の分裂および混合の合成が続く。このオリゴヌクレオチドを、表IIのサブユニットに対応する3'-ホスホルアミダイト誘導体化4マーによりサブユニット様式で伸張する。次いで、SmaI認識部位の半分（GGG）、2つのCおよび5'-モノホスフェートを、例えば、Clontech Laboratories（Palo Alto、CA）から入手可能なPhosphate-ON試薬を用いて、ヌクレオチド毎に添加することで合成を完了する。他方のセットのオリゴヌクレオチドは、3つのC（SmaI認識部位の一部）および2つのGの添加で開始し、9回の分裂および混合合成が続く。このオリゴヌクレオチドを、表IIのサブユニットの相補物に対応する3'-ホスホルアミダイト誘導体化4マーにより伸張する。HindIII認識部位および5'-モノホスフェートのヌクレオチド毎の添加により合成を完了する。合成支持体からの分離後、オリゴヌクレオチドを以下の二重鎖の形成を可能にする条件下で混合する：



次いで、二重鎖の混合物をSmaI/HindIIIの消化M13mp19に連結する。タグ相補物のレパートリーを上記のようにCPG微粒子上で合成する。

次に、FokI部位ならびにEcoRIおよびSmaI部位の一部を含有する、以下のアダプターを調製する：

5'-pAATTCGGATGATGCATGCATCGACCC
GCCTACTACGTACGTAGCTGGGp-5'

Eco RI Fok I

Sma I

このアダプターを、上記のように、EcoRI/SmaI消化のM13に連結する。

別に、Sambrookら（上記）に記載のプロトコルに従って超音波処理により、SV40 DNAを断片化する。得られたフラグメントを標準的なプロトコルを用いて修復し、そしてサイズによって分離する。300～500塩基対の範囲のフラグメントを選択し、上記のSmaI消化のM13に連結してフラグメント-タグ結合体のライブラリーを形成し、次いで、増幅する。数千の異なるフラグメント-タグ結合体を含むサンプルをライブラリーから取り出し、さらに増幅し、そしてフラグメント-タグ挿入物を、EcoRIおよびHindIIIで消化することにより切り出す。切り出されたフラグメント-タグ結合体を、実施例 I に記載のように、デオキシシチジントリホスフェートの存在下でT4 DNAポリメラーゼで処理して、特異的なハイブリダイゼーションのためのオリゴヌクレオチドタグをCPG微粒子に曝露する。

ハイブリダイゼーションおよび連結後、実施例 I に記載のように、ロードされた微粒子をFokIで処理して、所定の配列の4-ヌクレオチド突出鎖を生成する。以下のプローブの10:1の混合物（プローブ 1：プローブ 2）を微粒子上のポリヌクレオチドに連結する。

プローブ 1 **FAM- ATCGGATGAC**
 TAGCCTACTGAGCT

プローブ 2 **ビオチン - ATCCAATGAC**
 TAGGTTACTGAGCT

FAMは、Applied Biosystemsから入手可能なアミノホスフェートリンカー（Aminolinker）を介してプローブ 1 の上段の鎖の 5' -ヒドロキシルに結合されるフルオレセイン色素を表す。ビオチンはまた、Aminolinker部分を介して結合され得、そして必要に応じて、ポリエチレンオキシドリンカー（例えば、Jaschkeら（上記））を介してさらに伸張され得る。

次いで、ロードされた微粒子をアビジン化ガラススライドの表面上に堆積する。試薬および洗浄溶液は表面上へ送達され、そして表面上から除去され得る。結合した微粒子を有するアビジン化スライドを、走査型蛍光顕微鏡（例えば、Newport Model PM500-Cモーションコントローラー、488nmの励起光を発生するSpectra-Physics Model 2020アルゴンイオンレーザー、および520nm波長域発光フィルターなどの装置を備えたZeiss Axioskop）で検査する。励起光および蛍光発光を、同じ対物レンズを通して、それぞれ、送達および収集する。二色性ミラーによって励起光および収集蛍光を分離する。二色性ミラーは、収集された蛍光を一連の帯域フィルターを介して、モニターされている蛍光物質に対応するフォトン計数デバイス（例えば、Hamamatsu model 9403-02光電子増倍管、Stanford Research Systems model SR445増幅器およびmodel SR430多重チャンネルスケーラーおよびデジタルコンピューター（例えば、486ベースのコンピューター）を備える）へ向ける。コンピューターは、微粒子の位置を記録するスライドの二次元マップを生成する。

最初のプローブを除去するためのFokIでの切断後、結合微粒子上のポリヌクレオチドを、以下に記載の好適な単一塩基配列決定方法に従い、プローブ連結、洗浄、検出、切断および洗浄工程を20サイクル行う。各検出工程において、走査システムが、各微粒子で同定された塩基に対応する蛍光発光を記録する。一般に、以下の反応および洗浄を、特に記載されない限り、製造者（New England Biolabs）の推薦する使用酵素用の緩衝液を用いて実施する。標準的な緩衝液もまた、Sambrookら（上記）に記載される。

以下の 4 組の混合プローブを、標的ポリヌクレオチドへの添加用に提供する：

**TAMRA- ATCGGATGACATCAAC
TAGCCTACTGTAGTTGANNN**

**FAM- ATCGGATGACATCAAC
TAGCCTACTGTAGTTGCNNN**

**ROX- ATCGGATGACATCAAC
TAGCCTACTGTAGTTGGNNN**

**JOE- ATCGGATGACATCAAC
TAGCCTACTGTAGTTGTNNN**

10

ここで、TAMRA、FAM、ROX、およびJOEは、Aminolinker IIによって結合された分光解像可能な蛍光標識である（すべて、Applied Biosystems, Inc.、Forster City, California から入手可能である）；太字のヌクレオチドは、FokIエンドヌクレアーゼの認識部位であり、「N」は、4つのヌクレオチド、A、C、G、Tのいずれか1つを表す。TAMRA（テトラメチルローダミン）、FAM（フルオレセイン）、ROX（ローダミンX）、およびJOE（2',7'-ジメトキシ-4',5'-ジクロロフルオレセイン）およびそれらのオリゴヌクレオチドへの結合はまた、Fungら米国特許第4,855,225号に記載される。

上記のプローブを、約5モル過剰量の標的ポリヌクレオチド末端中で、以下のようにインキュベートする：プローブを、200ユニットのT4 DNAリガーゼおよびアンカー化標的ポリヌクレオチドとともに60分間16℃でT4 DNAリガーゼ緩衝液中でインキュベートする：次いで、洗浄後、標的ポリヌクレオチドを、100ユニットのT4ポリヌクレオチドキナーゼとともに30分間37℃で製造者の推薦する緩衝液中でインキュベートし、洗浄し、そして再度、200ユニットのT4 DNAリガーゼおよびアンカー化標的ポリヌクレオチドとともに30分間16℃でT4 DNAリガーゼ緩衝液中でインキュベートする。洗浄は、スライド上に洗浄緩衝液（例えば、TE）を大量に連続的に流すことによって達成される（Sambrookら（上記）に開示される）。連結ーリン酸化ー連結のサイクルおよび最後の洗浄後、結合微粒子が、蛍光標識の存在について走査され、その位置および特徴が走査システムによって記録される。次いで、標識された標的ポリヌクレオチド、即ち、連結複合体を、製造者の推薦する緩衝液中で30分間37℃で、10ユニットのFokIとともにインキュベートし、その後、TE中で洗浄する。その結果、標的ポリヌクレオチドは各鎖上で1つのヌクレオチド分短くされ、次の連結および切断サイクルのために準備される。20のヌクレオチドが同定されるまでこのプロセスを続ける。

20

30

実施例 3

タグライブラリーの構築

例示的なタグライブラリーを以下のように構築して、式、

3'-TGGC-[⁴(A,G,T)₉]-CCCCp

により規定される、ヌクレオチドA、G、およびTからなる化学合成9ワードタグを形成する。ここで、「⁴(A,G,T)₉」は、各タグがA、G、およびTの9個の4マーワードからなるタグ混合物を示し；そして「p」は5'ホスフェートを示す。この混合物を、以下の右および左のプライマーの結合領域に連結する：

40

**5'- AGTGGCTGGGCATCGGACCG
TCACCGACCCGTAGCCp**

**5'- GGGGCCAGTCAGCGTCGAT
GGGTCAGTCGCAGCTA**

左

右

右および左のプライマーの結合領域を上記タグ混合物と連結した後、連結構造の1本鎖部分をDNAポリメラーゼで充填し、次いで、以下に示す右および左のプライマーと混合し、そして増幅してタグライブラリーを得る。

左プライマー

5'- AGTGGCTGGGCATCGGACCG

5'- AGTGGCTGGGCATCGGACCG- $\{^4\{ (A, G, T) 9 \}$ -GGGGCCCAGTCAGCGTCGAT
 TCACCGACCCGTAGCCTGGC- $\{^4\{ (A, G, T) 9 \}$ -CCCCGGGTCAGTCGCAGCTA

CCCCGGGTCAGTCGCAGCTA-5'

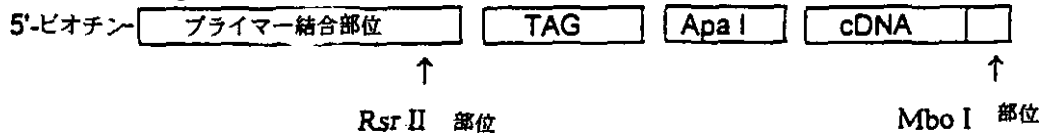
右プライマー

左のプライマー結合領域の下線部分はRsrII認識部位を示す。右のプライマー結合領域の最も左の下線領域は、Bsp120I、ApaI、およびEco0109Iの認識部位、ならびにHgaIの切断部位を示す。右のプライマー結合領域の最も右の下線領域は、HgaIの認識部位を示す。必要に応じて、右または左のプライマーは、（従来の試薬、例えば、Clontech Laboratories, Palo Alto, CAから入手可能な試薬を用いて）ビオチンを結合させて合成され、増幅および／または切断後の精製を容易にし得る。

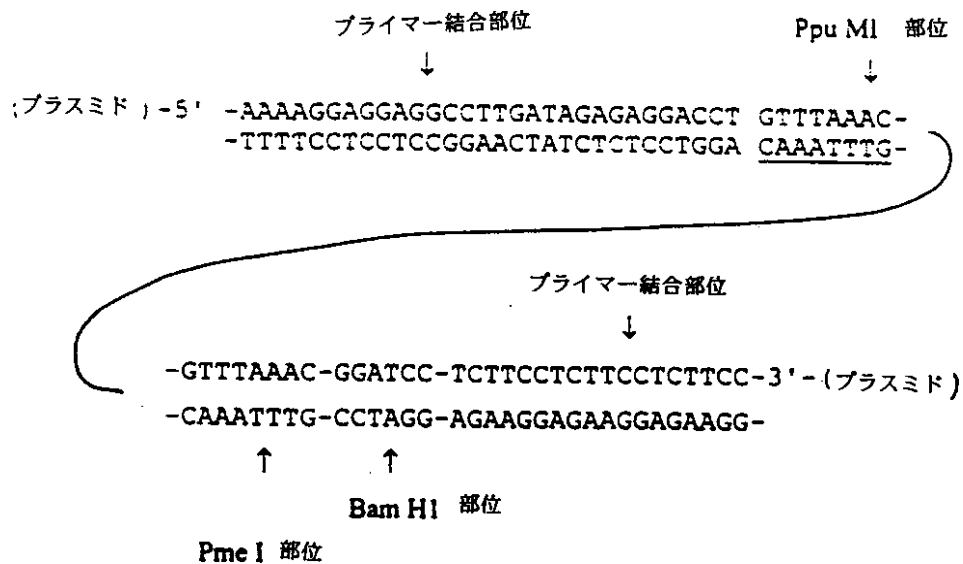
実施例 4

cDNA「サイン (signature)」配列決定のためのタグ-ポリヌクレオチド結合体のプラスミドライブラリーの構築

cDNAを、mRNAのポリ A 領域の境界で結合したpGGCCCT₁₅ (AまたはGまたはC) を第 1 の鎖の合成のためのプライマーとして、N₈ (AまたはT) GATCを第 2 の鎖の合成のためのプライマーとして用いて、従来のプロトコルによりmRNAサンプルから生成する。すなわち、両方は、第 2 の鎖のプライマーが 2 つの形態で存在し、そして第 1 の鎖のプライマーが 3 つの形態で存在するような縮重プライマーである。第 2 の鎖のプライマー中のGATC配列はMboIの認識部位に対応し；他の 4 塩基認識部位は同様に使用され得る（例えば、BamHI、SphI、EcoRIなどの認識部位）。第 2 の鎖のプライマーの認識部位に隣接する A および T の存在は、ストリッピングおよび交換反応を次の工程で使用して、「GGCCC」の 5 塩基 5' 突出を生じさせ得ることを確実にする。第 1 の鎖のプライマーをmRNAサンプルにアニールし、そして逆転写酵素で伸長し、その後、RNA鎖を逆転写酵素のRNase H活性により分解して、1 本鎖cDNAを残す。第 2 の鎖のプライマーを、従来のプロトコルを用いて、アニールし、そしてDNAポリメラーゼにより伸長する。第 2 の鎖の合成後、得られたcDNAを、製造者のプロトコルを用いて、CpGメチラーゼ (New England Biolabs, Beverly, MA) によりメチル化する。次いで、cDNAの 3' 鎖を、dATPおよびdTTPの存在下でT4 DNAポリメラーゼを用いて、上記のストリッピングおよび交換反応によりカットバック (cut back) し、その後、cDNAを、先にHgaIで切断した実施例 3 のタグライブラリーに連結して以下の構築物を得る：



以下のクローニングベクターを、例えば、Bluescriptファージミド (Stratagene, La Jolla, CA) のような市販されているプラスミドから出発して、別々に構築する。



10

プラスミドをPpuMIおよびPmeIを用いて切断し（挿入物が配向されるようにRsrII適合性末端および平滑末端を得るため）、次いでDAMメチラーゼでメチル化する。タグ含有構築物をRsrIIで切断し、次いで開環プラスミドに連結し、その後、結合体をMboIおよびBamHIで切断してプラスミドを連結および閉環させる。次いで、プラスミドを、増幅し、単離し、そして本発明に従って使用する。

20

付録 Ia

最少に交差ハイブリダイズするセットを生成するためのコンピュータプログラムの例
 （一本鎖タグ／一本鎖タグ相補物）

```

integer*2 sub1(6),mset1(1000,6),mset2(1000,6)
dimension nbase(6)

write(*,*)'ENTER SUBUNIT LENGTH'
read(*,100)nsup
format(i1)
open(1,file='sub4.dat',form='formatted',status='new')

nset=0
do 7000 m1=1,3
  do 7000 m2=1,3
    do 7000 m3=1,3
      do 7000 m4=1,3
        sub1(1)=m1
        sub1(2)=m2
        sub1(3)=m3
        sub1(4)=m4

      ndiff=3
      sab1と少なくともndiffヌクレオチドだけ異なる
      サブユニットのセットを生成する
      mset1に保存する

      jj=1
      do 900 j=1,nsup
        mset1(1,j)=sub1(j)

      do 1000 k1=1,3
        do 1000 k2=1,3
          do 1000 k3=1,3
            do 1000 k4=1,3

              nbase(1)=k1
              nbase(2)=k2
              nbase(3)=k3
              nbase(4)=k4

```

30


```

c
      do 1600 m=npass-2,jj
        n=0
        do 1600 j=1,nsup
          if(mset1(npass+1,j).eq.1.and.mset1(m,j).ne.1.or.
            mset1(npass+1,j).eq.2.and.mset1(m,j).ne.2.or.
            mset1(npass+1,j).eq.3.and.mset1(m,j).ne.3) then
            n=n+1
          endif
          continue
          if(n.ge.ndiff) then
            kk=kk+1
            do 1625 i=1,nsup
              mset2(kk,i)=mset1(m,i)
            endif
            continue
          endif
        enddo
      enddo
      c
      c
      c
      c
      c
      c
      c
      c
      do 2000 k=1,kk
        do 2000 m=1,nsup
          mset1(k,m)=mset2(k,m)
        enddo
        if(kk.lt.jj) then
          jj=kk
          goto 1700
        endif
      enddo
      c
      c
      nset=nset+1
      write(1,7009)
      format(/)
      do 7008 k=1,kk
        write(1,7010) (mset1(k,m),m=1,nsup)
      enddo
      format(4i1)
      write(*,*)
      write(*,120) kk,nset
      format(1x,'Subunits in set=',i5,2x,'Set No=',i5)
      120
      7000
      continue
      close(1)
      c
      c
      end
      c
      c
      *****
      *****

```

付録 1b

最少に交差ハイブリダイズするセットを生成するためのコンピュータプログラムの例
(一本鎖タグ／一本鎖タグ相補物)

10

20

30

40

C
C
C
C
C
C
C

C
C

10
C
C

11
C
C

C
C
C
C
C
C

8
C
C
C
C
C
C
C

```
do 1000 k1=1,3
```

30

```

do 1000 k2=1,3
  do 1000 k3=1,3
    do 1000 k4=1,3
      do 1000 k5=1,3
        do 1000 k6=1,3
          do 1000 k7=1,3
            do 1000 k8=1,3
              do 1000 k9=1,3
                do 1000 k10=1,3
do 1000 k11=1,3
  do 1000 k12=1,3
    do 1000 k13=1,3
      do 1000 k14=1,3
        do 1000 k15=1,3
          do 1000 k16=1,3
            do 1000 k17=1,3
              do 1000 k18=1,3
                do 1000 k19=1,3
                  do 1000 k20=1,3
c
c
nbase(1)=k1
nbase(2)=k2
nbase(3)=k3
nbase(4)=k4
nbase(5)=k5
nbase(6)=k6
nbase(7)=k7
nbase(8)=k8
nbase(9)=k9
nbase(10)=k10
nbase(11)=k11
nbase(12)=k12
nbase(13)=k13
nbase(14)=k14
nbase(15)=k15
nbase(16)=k16
nbase(17)=k17
nbase(18)=k18
nbase(19)=k19
nbase(20)=k20
c
c
c
do 1250 nn=1,jj
  n=0
  do 1200 j=1,nsup
    if(mset(nn,j).eq.1 .and. nbase(j).ne.1 .or.
1      mset(nn,j).eq.2 .and. nbase(j).ne.2 .or.
2      mset(nn,j).eq.3 .and. nbase(j).ne.3 .or.
3      mset(nn,j).eq.4 .and. nbase(j).ne.4) then
      n=n+1
    endif
    continue
1200
c
c
if(n.lt.ndiff) then
  goto 1000
endif
1250
c
c
continue

  jj=jj+1
  write(*,130)(nbase(i),i=1,nsup),jj
  do 1100 i=1,nsup

```

10

20

30

40

```

                mset(jj,i)=nbase(i)
            continue
        c
        c
    1000    continue
    c
    c
        write(*,*)
    130    format(10x,20(1x,i1),5x,i5)
        write(*,*)
        write(*,120) jj
    120    format(1x,'Number of words=',i5)
    c
    c
        end
    c
    c
    c
        *****
        *****

```

10

付録 Ic

最少に交差ハイブリダイズするセットを生成するためのコンピュータプログラムの例
 (二本鎖タグ／一本鎖タグ相補物)

プログラム 3tagN

```

c
c
c
c プログラム3tagNは i)N-サブユニット長および ii)初期ホモプリン配列を与える
c 二重鎖サブユニットの最小に交差ハイブリダイズするセットを生成する。
c
c
c      character*1 sub1(20)
c      integer*2 mset(10000,20), nbase(20)
c
c
c      write(*,*)'ENTER SUBUNIT LENGTH'
c      read(*,100)nsup
100      format(i2)
c
c
c      write(*,*)'ENTER SUBUNIT SEQUENCE a & g only'
c      read(*,110)(sub1(k),k=1,nsup)
110      format(20a1)
c
c
c      ndiff=10
c
c
c      Let a=1 and g=2
c
c
c
c      do 800 kk=1,nsup
c      if(sub1(kk).eq.'a') then
c          mset(1,kk)=1
c      endif
c          if(sub1(kk).eq.'g') then
c              mset(1,kk)=2
c          endif
800      continue
c
c
c
c      jj=1
c
c
c      do 1000 k1=1,3
c          do 1000 k2=1,3
c              do 1000 k3=1,3
c                  do 1000 k4=1,3
c                      do 1000 k5=1,3
c                          do 1000 k6=1,3
c                              do 1000 k7=1,3
c                                  do 1000 k8=1,3
c                                      do 1000 k9=1,3
c                                          do 1000 k10=1,3
c                                  do 1000 k11=1,3
c                                      do 1000 k12=1,3
c                                          do 1000 k13=1,3

```

10

20

30

```

do 1000 k14=1,3
do 1000 k15=1,3
do 1000 k16=1,3
do 1000 k17=1,3
do 1000 k18=1,3
do 1000 k19=1,3
do 1000 k20=1,3
c
c
nbase(1)=k1
nbase(2)=k2
nbase(3)=k3
nbase(4)=k4
nbase(5)=k5
nbase(6)=k6
nbase(7)=k7
nbase(8)=k8
nbase(9)=k9
nbase(10)=k10
nbase(11)=k11
nbase(12)=k12
nbase(13)=k13
nbase(14)=k14
nbase(15)=k15
nbase(16)=k16
nbase(17)=k17
nbase(18)=k18
nbase(19)=k19
nbase(20)=k20
c
c
do 1250 nn=1,jj
n=0
do 1200 j=1,nsub
1 if(mset(nn,j).eq.1.and. nbase(j).ne.1 .or.
2 mset(nn,j).eq.2 .and. nbase(j).ne.2 .or.
3 mset(nn,j).eq.3 .and. nbase(j).ne.3 .or.
mset(nn,j).eq.4 .and. nbase(j).ne.4) then
n=n+1
endif
1200 continue
c
c
if(n.lt.ndiff) then
goto 1000
endif
1250 continue
c
jj=jj+1
write(*,130)(nbase(i),i=1,nsub),jj
do 1100 i=1,nsub
mset(jj,i)=nbase(i)
1100 continue
c
1000 continue
c
write(*,*)
130 format(10x,20(1x,i1),5x,i5)
write(*,*)
write(*,120) jj
120 format(1x,'Number of words=',i5)
c
c
end

```

配列表

(1) 一般的情報：

(i) 出願人：シドニー プレナー

(ii) 発明の名称：分類および同定のためのオリゴヌクレオチドタグ

(iii) 配列数：16

(iv) 連絡住所：

10

20

30

40

50

- (A) 名称：ステファン シー．マセビッツ，スペクトラジェン，インコーポレイテッド
- (B) 番地：3832 ベイ センター プレイス
- (C) 市：ヘイワード
- (D) 州：カリフォルニア
- (E) 国：アメリカ合衆国
- (F) 郵便番号：94545

(v) コンピューター読み出し形態：

- (A) 媒体型：3.5 インチ ディスケット
- (B) コンピューター：IBM 互換機
- (C) OS：ウィンドウズ 3.1
- (D) ソフトウェア：マイクロソフト ワード 5.1

10

(vi) 現在の出願データ：

- (A) 出願番号：
- (B) 出願日：
- (C) 分類：

(vii) 先願の出願データ：

- (A) 出願番号：PCT/US95/12791
- (B) 出願日：1995年10月12日

(vii) 先願の出願データ：

- (A) 出願番号：08/478,238
- (B) 出願日：1995年 6 月 7 日

20

(vii) 先願の出願データ：

- (A) 出願番号：08/485,105
- (B) 出願日：1995年 6 月 7 日

(vii) 先願の出願データ：

- (A) 出願番号：PCT/US95/12791
- (B) 出願日：1995年10月12日

(viii) 代理人／事務所情報：

- (A) 氏名：ステファン シー．マセビッツ
- (B) 登録番号：30,285
- (C) 照会／記録番号：cbd4wo

30

(ix) 電話回線情報：

- (A) 電話：(510) 670-9365
- (B) テレファックス：(510) 670-9302

(2) 配列番号 1 の情報：

(i) 配列の特色：

- (A) 長さ：38ヌクレオチド
- (B) 型：核酸
- (C) 鎖の数：一本鎖
- (D) トポロジー：直鎖状

40

(xi) 配列：配列番号 1：

GAGGATGCCT TTATGGATCC ACTCGAGATC CCAATCCA

38

(2) 配列番号 2 の情報：

(i) 配列の特色：

- (A) 長さ：26ヌクレオチド
- (B) 型：核酸
- (C) 鎖の数：一本鎖
- (D) トポロジー：直鎖状

(xi) 配列：配列番号 2：

AATTCGGATG ATGCATGCAT CGACCC

26

50

(2) 配列番号 3 の情報：		
(i) 配列の特色：		
(A) 長さ：14ヌクレオチド		
(B) 型：核酸		
(C) 鎖の数：一本鎖		
(D) トポロジー：直鎖状		
(xi) 配列：配列番号 3：		
TAGCCTACTG AGCT	14	
(2) 配列番号 4 の情報：		
(i) 配列の特色：		10
(A) 長さ：16ヌクレオチド		
(B) 型：核酸		
(C) 鎖の数：一本鎖		
(D) トポロジー：直鎖状		
(xi) 配列：配列番号 4：		
ATCGGATGAC ATCAAC	16	
(2) 配列番号 5 の情報：		
(i) 配列の特色：		
(A) 長さ：11ヌクレオチド		
(B) 型：核酸		20
(C) 鎖の数：二本鎖		
(D) トポロジー：直鎖状		
(xi) 配列：配列番号 5：		
ACCAGCTGAT C	11	
(2) 配列番号 5 の情報：		
(i) 配列の特色：		
(A) 長さ：11ヌクレオチド		
(B) 型：核酸		
(C) 鎖の数：一本鎖		30
(D) トポロジー：直鎖状		
(xi) 配列：配列番号 5：		
CTAGTCGACC A	11	
(2) 配列番号 6 の情報：		
(i) 配列の特色：		
(A) 長さ：11ヌクレオチド		
(B) 型：核酸		
(C) 鎖の数：一本鎖		
(D) トポロジー：直鎖状		40
(xi) 配列：配列番号 6：		
NRRGATCYNN N	11	
(2) 配列番号 7 の情報：		
(i) 配列の特色：		
(A) 長さ：22ヌクレオチド		
(B) 型：核酸		
(C) 鎖の数：一本鎖		
(D) トポロジー：直鎖状		
(xi) 配列：配列番号 7：		50

GGGTCGATGC ATGCATCATC CG**22**

- (2) 配列番号 8 の情報：
(i) 配列の特色：
(A) 長さ：10ヌクレオチド
(B) 型：核酸
(C) 鎖の数：一本鎖
(D) トポロジー：直鎖状
(xi) 配列：配列番号 8：

ATCGGATGAC**10**

10

- (2) 配列番号 9 の情報：
(i) 配列の特色：
(A) 長さ：10ヌクレオチド
(B) 型：核酸
(C) 鎖の数：一本鎖
(D) トポロジー：直鎖状
(xi) 配列：配列番号 9：

ATCNNNNNAC**10**

- (2) 配列番号10の情報：
(i) 配列の特色：
(A) 長さ：14ヌクレオチド
(B) 型：核酸
(C) 鎖の数：一本鎖
(D) トポロジー：直鎖状
(xi) 配列：配列番号10：

20

TCGAGTNNNN NGAT**14**

- (2) 配列番号11の情報：
(i) 配列の特色：
(A) 長さ：16ヌクレオチド
(B) 型：核酸
(C) 鎖の数：一本鎖
(D) トポロジー：直鎖状
(xi) 配列：配列番号11：

30

ATCGGATGAC ATCAAC**16**

- (2) 配列番号12の情報：
(i) 配列の特色：
(A) 長さ：20ヌクレオチド
(B) 型：核酸
(C) 鎖の数：一本鎖
(D) トポロジー：直鎖状
(xi) 配列：配列番号12：

40

NNNAGTTGAT GTCATCCGAT**20**

- (2) 配列番号13の情報：
(i) 配列の特色：
(A) 長さ：20ヌクレオチド
(B) 型：核酸
(C) 鎖の数：一本鎖
(D) トポロジー：直鎖状

50

(xi) 配列：配列番号13：

NNNCGTTGAT GTCATCCGAT

20

(2) 配列番号14の情報：

(i) 配列の特色：

(A) 長さ：20ヌクレオチド

(B) 型：核酸

(C) 鎖の数：一本鎖

(D) トポロジー：直鎖状

(xi) 配列：配列番号14：

NNNGGTTGAT GTCATCCGAT

20

10

(2) 配列番号15の情報：

(i) 配列の特色：

(A) 長さ：20ヌクレオチド

(B) 型：核酸

(C) 鎖の数：一本鎖

(D) トポロジー：直鎖状

(xi) 配列：配列番号15：

NNNTGTTGAT GTCATCCGAT

20

20

(2) 配列番号16の情報：

(i) 配列の特色：

(A) 長さ：37ヌクレオチド

(B) 型：核酸

(C) 鎖の数：二本鎖

(D) トポロジー：直鎖状

(xi) 配列：配列番号16：

NNNNNGGATG NNNNNNNNNN NNNTNNNNNNN NNNNNNNN

37--.

【図 1】

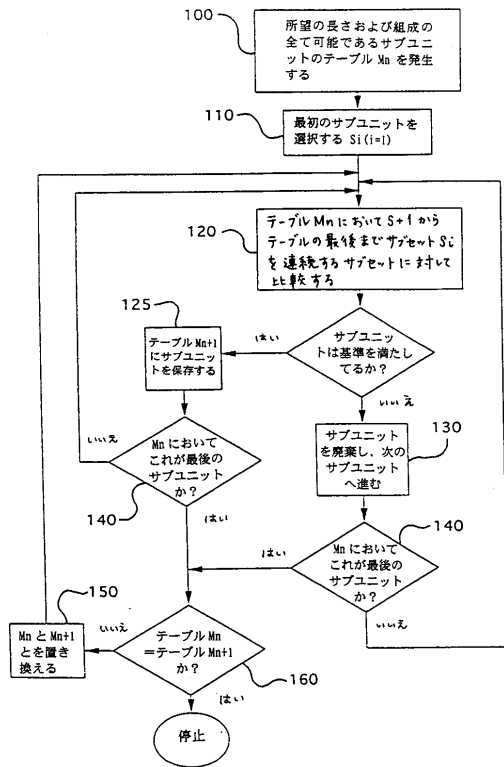


Fig. 1

【図 2】

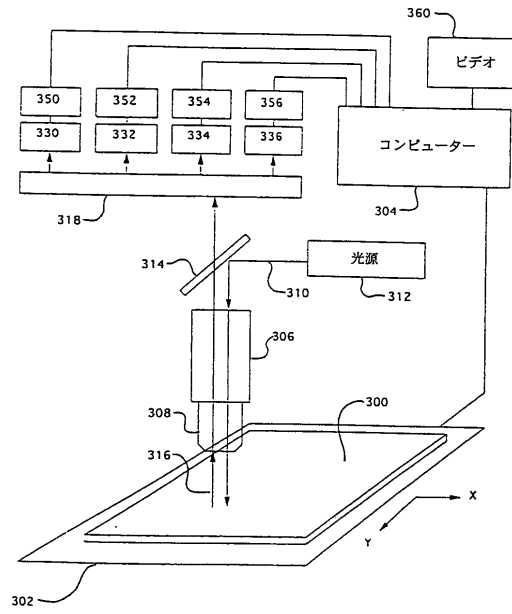


Fig. 2

【図 3】

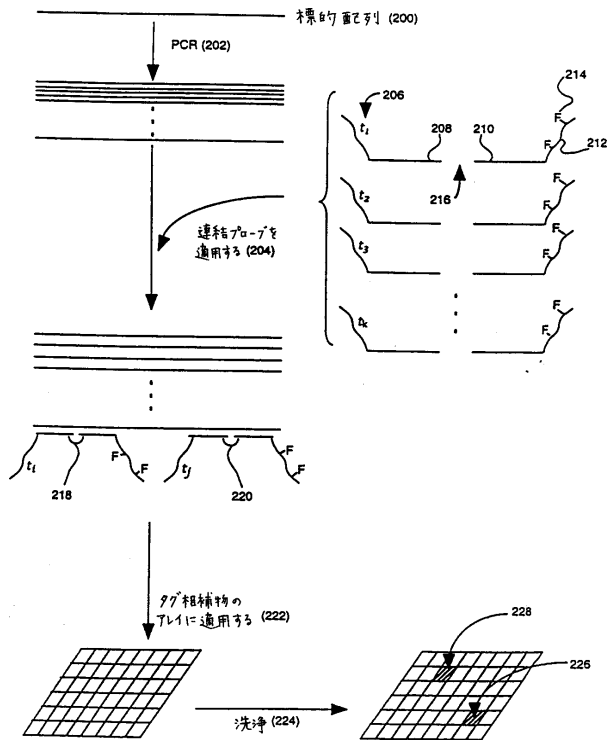


Figure 3.

フロントページの続き

(74)代理人 100126653

弁理士 木元 克輔

(72)発明者 ブレナー, シドニー

イギリス国 シービー2 3ピージェイ ケンブリッジ, セント エドワーズ パッセージ 17
ビー

(72)発明者 アルブレヒト, グレン

アメリカ合衆国 カリフォルニア 94061, レッドウッド シティ, オリバー ストリート
1113

合議体

審判長 平田 和男

審判官 鵜飼 健

審判官 小暮 道明

(56)参考文献 特表平8-504082 (JP, A)

特表平10-507357 (JP, A)

特開平1-137982 (JP, A)

特開平5-7490 (JP, A)

特表平5-504095 (JP, A)

特開昭63-188399 (JP, A)

国際公開第96/12014 (WO, A1)

(58)調査した分野(Int.Cl., DB名)

C12Q1/68

C12N15/00

CA/BIOSIS/WPIDS/MEDLINE(STN)