

(19)日本国特許庁(JP)

## (12)特許公報(B2)

(11)特許番号

特許第7064439号

(P7064439)

(45)発行日 令和4年5月10日(2022.5.10)

(24)登録日 令和4年4月26日(2022.4.26)

(51)国際特許分類

F I

C 1 2 Q 1/6806(2018.01)

C 1 2 Q 1/6806

Z Z N A

C 1 2 Q 1/6874(2018.01)

C 1 2 Q 1/6874

Z

C 1 2 N 15/11 (2006.01)

C 1 2 N 15/11

Z

C 1 2 N 15/10 (2006.01)

C 1 2 N 15/10

1 1 0 Z

請求項の数 27 (全45頁)

(21)出願番号 特願2018-543062(P2018-543062)

(86)(22)出願日 平成28年11月4日(2016.11.4)

(65)公表番号 特表2018-538006(P2018-538006  
A)

(43)公表日 平成30年12月27日(2018.12.27)

(86)国際出願番号 PCT/US2016/060576

(87)国際公開番号 WO2017/079593

(87)国際公開日 平成29年5月11日(2017.5.11)

審査請求日 令和1年11月1日(2019.11.1)

(31)優先権主張番号 62/250,900

(32)優先日 平成27年11月4日(2015.11.4)

(33)優先権主張国・地域又は機関  
米国(US)

(73)特許権者 521359911

アトレカ インコーポレイテッド

アメリカ合衆国 9 4 0 7 0 カリフォル

ニア州 サン カルロス インダストリアル

ロード 8 3 5 スイート 4 0 0

(74)代理人 100102978

弁理士 清水 初志

(74)代理人 100102118

弁理士 春名 雅夫

(74)代理人 100160923

弁理士 山口 裕孝

(74)代理人 100119507

弁理士 刑部 俊

(74)代理人 100142929

弁理士 井上 隆一

最終頁に続く

(54)【発明の名称】 単一細胞に関連する核酸の解析のための、核酸バーコードの組み合わせセット

## (57)【特許請求の範囲】

## 【請求項 1】

核酸試料の細胞起源を特定する方法であって、

(a) 単一細胞に由来する核酸分子を含む核酸試料を含み、かつ独特の5'バーコードを各構築物が含む1つ以上の5'核酸構築物および独特の3'バーコードを各構築物が含む1つ以上の3'核酸構築物を含むバーコードのセットも含むことができる、複数の反応混合物を形成する工程であって、前記1つ以上の独特の5'バーコードおよび3'バーコードが、前記複数の反応混合物の30%、40%、50%、60%、70%、80%、90%または100%に存在し、かつ各5'核酸構築物が、固体表面に結合された複数コピーで提供され、かつ各3'核酸構築物が、固体表面に結合された複数コピーで提供される工程と、

(b) 前記試料の前記核酸分子にバーコードの前記セットを組み込む工程と、

(c) 前記単一細胞の前記核酸分子に組み込まれたバーコードの前記セットを特定し、それによって前記核酸試料の細胞起源を特定する工程であって、前記核酸試料における前記1つ以上の独特の5'バーコードと3'バーコードとの組み合わせが、前記核酸試料の細胞起源を特定する、工程

とを含む、前記方法。

## 【請求項 2】

前記反応混合物を形成する工程が、バーコードの前記セットに前記単一細胞を接触させることと、前記単一細胞に由来する前記核酸試料を提供するために、前記細胞を溶解することとを含む、または

前記反応混合物を形成する工程が、懸濁緩衝液中の細胞を提供することと、前記単一細胞とバーコードの前記セットとを含む前記反応混合物を形成するための条件下で、バーコードの前記セットを含むバーコーディング緩衝液に、前記細胞を含む前記懸濁緩衝液を接触させることとを含む、請求項 1 に記載の方法。

【請求項 3】

前記構築物が、独特の分子識別子 (UMI) 配列をさらに含む、請求項 1 または 2 に記載の方法。

【請求項 4】

前記固体表面がビーズである、請求項 1 または 2 に記載の方法。

10

【請求項 5】

各 5' 核酸構築物及び 3' 核酸構築物が、光切断可能なリンカーまたは制限酵素部位をさらに含み、前記反応混合物を制限酵素に接触させる工程および/または前記反応混合物を UV 光に曝露して、前記固体表面から前記 5' 核酸構築物及び前記 3' 核酸構築物を遊離させる工程をさらに含む、請求項 1 または 2 に記載の方法。

【請求項 6】

前記 5' 核酸構築物が、プロモータを含む二本鎖核酸構築物であり、且つ前記 5' 核酸構築物が、複数コピーの前記 5' 核酸構築物を作り出すための鋳型として役立つ、請求項 1 または 2 に記載の方法。

【請求項 7】

20

バーコードの前記セットが、1、2、3、4、5、6、7、または 8 個の 5' 核酸構築物と、1、2、3、4、5、6、7、または 8 個の 3' 核酸構築物とを含む、請求項 1 または 2 に記載の方法。

【請求項 8】

前記 3' 核酸構築物が結合配列をさらに含む、請求項 1 または 2 に記載の方法。

【請求項 9】

前記結合配列がポリ - T 配列を含む、請求項 8 に記載の方法。

【請求項 10】

前記ポリ - T 配列が、前記試料の前記核酸分子におけるポリ - A 配列に結合する、請求項 9 に記載の方法。

30

【請求項 11】

前記 3' 核酸構築物における前記バーコードが、逆転写によって組み込まれる、請求項 1 または 2 に記載の方法。

【請求項 12】

前記逆転写により、前記試料の前記核酸分子の末端に 2 ~ 5 個のシトシン残基が付加される、請求項 11 に記載の方法。

【請求項 13】

前記 2 ~ 5 個のシトシン残基が、前記 5' 核酸構築物における結合配列に相補性の結合配列を提供する、請求項 12 に記載の方法。

【請求項 14】

40

前記 5' 核酸構築物における前記結合配列が、ポリ - G 配列を含む、請求項 13 に記載の方法。

【請求項 15】

前記 5' 核酸構築物における前記結合配列が、前記試料の前記核酸分子におけるポリ - C 配列に結合する、請求項 14 に記載の方法。

【請求項 16】

前記 5' 核酸構築物が、逆転写酵素を用いて前記核酸試料に組み込まれる、請求項 1 または 2 に記載の方法。

【請求項 17】

前記逆転写酵素が M - M L V 酵素である、請求項 16 に記載の方法。

50

## 【請求項 18】

前記逆転写酵素が M M L V H - 逆転写酵素である、請求項 17 に記載の方法。

## 【請求項 19】

5' バーコード及び 3' バーコードを各配列が含む複数の核酸配列の複数コピーを作り出すために、前記核酸試料がさらに増幅される、請求項 1 または 2 に記載の方法。

## 【請求項 20】

前記核酸分子が m R N A を含む、請求項 1 ~ 19 のいずれか 1 項に記載の方法。

## 【請求項 21】

前記核酸試料における前記バーコードが、配列決定によって特定される、請求項 1 ~ 20 のいずれか 1 項に記載の方法。

10

## 【請求項 22】

前記核酸分子の 1 つ以上が、2 つ以上の独特のバーコードを含む、請求項 1 に記載の方法。

## 【請求項 23】

前記独特のバーコードが、前記核酸分子の 5' 末端、3' 末端、または 5' 末端と 3' 末端の双方に位置する、請求項 22 に記載の方法。

## 【請求項 24】

複数の反応混合物を形成する工程を含む、請求項 1 または 2 に記載の方法。

## 【請求項 25】

2 つ以上の独特のバーコードが、前記反応混合物の 30%、40%、50%、60%、70%、80%、90% または 100% に存在する、請求項 24 に記載の方法。

20

## 【請求項 26】

前記複数の反応混合物を形成する工程が、0.5、0.6、0.7、0.8、0.9、1.0、2.0、3.0、4.0 または 5.0 より大きいラムダで複数の反応混合物にバーコードの前記セットを加えることを含む、請求項 24 に記載の方法。

## 【請求項 27】

前記反応混合物が液滴である、請求項 1 に記載の方法。

## 【発明の詳細な説明】

## 【技術分野】

## 【0001】

関連出願への相互参照

30

本出願は、その全体が参照によって本明細書に組み入れられる 2015 年 11 月 4 日に出願された米国仮特許出願番号 62, 250, 900 の優先権を主張する。

## 【背景技術】

## 【0002】

発明の背景

細胞集団のゲノム及びトランスクリプトームのプロファイリングは、バイオテクノロジー及び医学の分野におけるかなりの影響と共に強く高い関心のある領域である。核酸の配列決定技術における最近の進歩によって従来のアプローチの実践的な処理能力をはるかに上回るペースでの規模でプロファイリング研究が可能になっている。単一細胞のゲノムまたはトランスクリプトームのプロファイリングに向けた初期の試みには、ゲノムまたはトランスクリプトームの最適な解析方法を適用することができる FACS 選別またはさらにマルチウェルマイクロタイタープレートのウェルのような別個の反応バイアルへの個々の細胞の手選が含まれた。この形式では、別個の反応は個々に操作され、制御され得る。たとえば、各ウェルが単一細胞のみを含有することを確実にするのは比較的簡単である。複数細胞の同時隔離はこれらの細胞のゲノムまたはトランスクリプトームのプロファイリングが互いに巻き込まれ合ってしまう、個々のプロファイリングを解きほぐすことが難しいまたは不可能であるので、このことは重要である。これらの形式の短所は、それらは時間がかかり、試薬がかかり、人手がかかり、これらの理由で相対的に低い処理能力のプロセスしかできないことである。さらに最近のアプローチは細胞の隔離に向かう傾向があり、マイクロウェルまたはナノウェルまたは液滴のようなナノリットルまたはピコリットルの大

40

50

きさの容器で化学試薬及び分子試薬を必要としている。これらの形式は時間、コスト及び処理能力という点で大幅な改善を可能にするが、細胞及び他の反応成分を容器に正確に負荷するのを制御するのがかなり難題であるという短所を伴う。成分は無作為に容器に割り当てられることが多く、その場合、多数の容器の間での成分の分配はポアソン分布に従う。細胞の場合、たとえば、システムが、細胞の少なくとも95%が単独で隔離される（言い換えれば、細胞を伴う容器のたった5%が2つ以上の細胞を含有するのを容認される）ことを必要とするのであれば、条件は各容器が平均で0.1の細胞を受け取るように設定されなければならない。ポアソン用語では、これは0.1の期待値、または「ラムダ」である。この場合、容器の90.48%は細胞を受け取らず、9.05%の容器は単一細胞を受け取り、0.45%は2個の細胞を受け取り、0.02%は3個の細胞を受け取り、さらに低い比率は4、5、6個等の細胞を受け取るであろう。この場合、細胞を受け取る容器のうち、95.1%は正確に1個の細胞を含有する。反応容器すべての90%を超えるものが細胞を欠いており非生産的なので、そのようなシステムの非効率性は明らかである。これらのポアソン分布の2つ以上の交点が含まれる場合、非効率性は度合を増す。ゲノム及びトランスクリプトームのプロファイリングに対する一般的なアプローチには、細胞を起源とする遺伝子及び/または転写産物にその後付加される核酸バーコードと共にその細胞を隔離することが含まれる。その後、バーコードを用いて各遺伝子または転写産物を起源の細胞に明瞭に割り当て、結果としてその細胞のゲノム及びトランスクリプトームのプロファイルを含む。そのようなスキームの曖昧性解消は、単独で隔離される細胞だけでなく、単独のバーコードのみを受け取る同一容器にも左右される。バーコードは単一の核酸分子であってもよいが、さらに一般的には、その表面が同一核酸バーコードの多数のコピーで官能化されるポリスチレンビーズのような粒子を介して届けられる。いずれの場合も、独特のバーコードの反応容器への無作為な割り当ては、前と同じ負荷統計、すなわち、期待値（ラムダ）によって決定されるポアソン分布に従うであろう。正確に1個の細胞が正確に1つのバーコードと共に同時隔離される容器のみが明白なゲノムまたはトランスクリプトームのプロファイルを生じるであろう。特定のバーコードとの多重細胞隔離は、上流データの解析の際、同時隔離した細胞の個々のプロファイルの結合体である単一のゲノムまたはトランスクリプトームのプロファイルのように見えるであろう。複数細胞の複数バーコードとの同時隔離の場合は一層さらに複雑な曖昧性を生じる。この曖昧性は単一細胞の遺伝子またはトランスクリプトームのプロファイリングを含むあらゆる適用に対して有害な意味合いを有する。たとえば、適用にトランスクリプトームのプロファイルによって特定される特定のクローンの定量が関与するのであれば、複数のバーコードがそのクローン型の細胞と共に同時隔離される場合は下流の解析ではそのクローン型の複数細胞の見掛けを与えるので、試料における細胞の集団全体のその見掛けの比率を人為的に誇張する。細胞とバーコードを受け取る容器の少なくとも95%が正確に1個の細胞と正確に1つのバーコードを受け取ることを要求するこの設計のプロファイリングシステムについては、所望の結果を与える細胞とバーコードのラムダの組み合わせの連続した繋がりがあ（容器がそれぞれの少なくとも一方を受け取れなければ、それは非生産性であり、考慮から除外することに留意のこと）。代表的な組み合わせの1つは細胞及びバーコードの双方について0.05のラムダである。この場合、容器は平均で0.05の細胞と0.05のバーコードを受け取るであろう。それぞれ細胞とバーコードのうち、容器の95.12%は何も受け取らず、4.76%は1つ受け取り、0.12%は2つ受け取り、さらに低い比率は3、4、5個等を受け取るであろう。少なくとも1個の細胞を受け取る容器のうち、97.52%が正確に1個の細胞を受け取り、バーコードについても同様である。これら2つの比率を重ね合わせると、 $97.52\% \times 97.52\% = 95.1\%$ の正確に1個の細胞と正確に1つのバーコードとしての細胞・バーコード同時隔離の事象を生じる。しかしながら、この基準を達成するために、99.76%を超える容器が細胞を受け取らない、バーコードを受け取らない、または両方とも受け取らないので非生産的である。言い換えれば、0.24%未満の容器が細胞プロファイルを作り出す機会を有するであろう。加えて、5%未満の細胞がバーコードと共に同時隔離されるので、95%を超える細胞

10

20

30

40

50

が解析から失われる。これはシステムの処理能力を大きく低下させ、これは解析からの細胞の95%の喪失を生じるので、解析する対象とする細胞が稀な細胞である場合、特に有害である。これらのピコリットル及びナノリットル規模の方法は、バーコードを受け取る容器の比率をバーコーディングスキームの忠実性を犠牲にすることなく増やせば、処理能力（すなわち、処理される細胞の数）及び試料の適用範囲（すなわち、処理される原試料における細胞の総数の比率）という点で意味のある改善を実現してもよい。

#### 【0003】

核酸バーコードの単クローン性の「パケット」または「セット」、たとえば、バーコード核酸で官能化したビーズ、核酸バーコードを負荷したヒドロゲルまたは類似のバーコード送達法を使用するゲノム及びトランスクリプトームのプロファイリングのアプローチについては、核酸バーコードのパケットはほぼ常にある程度の多クローン性を示す。多クローン性の程度はバーコードのパケットを作り出すのに使用される方法に左右される。たとえば、方法の1つは、汎用プライミング配列を持つDNAオリゴで官能化され、その後PCRによってビーズの表面で増幅する個々のバーコード分子で乳化されるビーズを使用する。このスキームでは、エマルジョンにおける各液滴はすべてが同一のバーコードを受け取る幾つかのビーズを含有してもよいが、理想的な場合、所与の液滴はたった1つのバーコード核酸のみを受け取ることになる。2つ以上のバーコード核酸が単一の液滴に入るのであれば、その液滴で生じるビーズは多クローン性であろう。バーコード核酸分子はバーコードの濃度を変えることによって調整することができるポアソン分布に従って液滴に割り当てられるであろう。ここで再び二律背反がある。低い濃度のバーコードは高い程度の単クローン性を有するビーズを作り出すが、全体的に少ないバーコード付きのビーズが作り出されるというコストのマイナス面がある。さらに高い濃度は、多くの比率のビーズが高い多クローン性を示すであろうバーコード付きビーズの収量を増やすであろう。多クローン性はこの方法の設計に組み込まれている。核酸バーコードをパッケージングする別の方法には組み合わせ合成が含まれる。384ウェルプレートで実施され、各ウェルは多数の粒子。この場合ヒドロゲルを含有し、それは汎用DNAオリゴプライマーで官能化される。各ウェルはPCRを介してそのウェルの各ヒドロゲルに組み込まれる違った最初のバーコード配列も含有する。次いでヒドロゲルは一緒にプールされ、別の384ウェルプレートに再分配され、その各ウェルは、再びPCRによって各ヒドロゲルに組み込まれる違った二次バーコード配列を含有する。この方法によって得られるヒドロゲルのライブラリは約150Kのバーコードの多様性を有し、理論的には高い程度の単クローン性を示すはずである。しかしながら、実際のところ、方法は無視できない程度（約8%）の多クローン性を生じることが報告されている（Klein, A. Cell, 161 (2015) (非特許文献1)）。この多クローン性は、いずれかのPCR工程でのPCRプレートのウェル間におけるヒドロゲルもしくはバーコード核酸の混入、または元のバーコード核酸ライブラリ。当初のまたは二次的なまたは双方のバーコード配列の混入のいずれかから生じてもよい。設計によってまたは混入によって、バーコードパケットのライブラリに対する多クローン性は、単一バーコードを各細胞に割り当てるように設計されているゲノム及びトランスクリプトームのプロファイリングのシステムを混乱させるであろう。バーコーディングスキームは、バーコードのパケットのライブラリにおけるある程度の多クローン性に対処する能力から利益を得る。

#### 【0004】

核酸バーコードのポアソン系分布から生じる非効率性を克服するまたは回避するのに利用されているアプローチの幾つかの例がある。個々のバーコードが充填され、細胞（0.1～0.2のラムダを持つポアソン分布に従って無作為に負荷される）と共に液滴に封入される直前に液滴装置の水性の流れに導入されるヒドロゲル粒子を使用するトランスクリプトームのプロファイリング法が報告されている。ヒドロゲルは高度に従順なので、それらは狭い流入口の単一の縦列に詰めることができ、液滴形成の速度に同期化することができる高度に秩序のある行列に導入することができる。その結果、液滴はほぼすべて独特のバーコードの多数のコピーと共に1個のヒドロゲル粒子（Abate, R. Lab on a

10

20

30

40

50

Chip, 9, (2009) (非特許文献2))を受け取る。しかしながら、そのように慎重に調整された実験が高度に最適化された環境で且つ専門家の手によって実行できる一方で、このアプローチが広汎な使用のために十分に拡張可能且つ着実であるかどうかは今後の課題である。加えて、スキームは多クローン性のバーコードパケットについては堅固ではない。他のアプローチには、重複伸長PCRの場合のように、複数のゲノムまたはトランスクリプトームの対象とする標的を一緒に単一の連続アンプリコンに組み込むことが関与している。これらのアプローチは、各標的にバーコードを付加することではなく、対象とする標的を物理的に連結するための逃れた汎用プライマーのセットを頼る。プライマーのセットは汎用であり、多数のコピーが容器ごとに導入されるので、それによって標的が起源のその細胞に割り当てられるメカニズムはポアソン系分布の短所に直面しない。しかしながら、これらのアプローチは各細胞から標的とされることができる潜在的な遺伝子または転写産物の数で厳しく限定される。全部ではないが既存の方法のほとんどは現在2つに限定されている (DeKosky, BJ, Nat. Biotechnol. 31 (2) (2013) (非特許文献3))。

10

【先行技術文献】

【非特許文献】

【0005】

【文献】Klein, A. Cell, 161 (2015)

Abate, R. Lab on a Chip, 9, (2009)

DeKosky, BJ, Nat. Biotechnol. 31 (2) (2013)

20

【発明の概要】

【0006】

本明細書で提供されるのは、核酸試料の起源を特定する方法である。該方法は、単一細胞に由来する核酸分子を含む核酸試料とバーコードのセットとを含む反応混合物を形成する工程と、バーコードのセットを試料の核酸分子に組み込む工程と、単一細胞の核酸分子に組み込まれたバーコードのセットを特定し、それによって核酸試料の起源を特定する工程とを含む。提供されるのはまた、各構築物がそれぞれ独特の5'バーコード配列または3'バーコード配列を含む5'核酸構築物及び3'核酸構築物のセットを提供する工程と、単一細胞に由来する核酸分子を含む核酸試料を提供する工程と、核酸試料を5'核酸構築物及び3'核酸構築物のセットに接触させる工程と、それぞれ5'核酸構築物及び3'核酸構築物に由来する5'バーコード配列及び3'バーコード配列を核酸分子に組み込む工程と、5'バーコード配列及び3'のバーコード配列のセットを特定し、それによって単一細胞に由来する核酸試料の起源を特定する工程とを含む、核酸試料の起源を特定する方法である。

30

[本発明1001]

核酸試料の起源を特定する方法であって、

(a) 単一細胞に由来する核酸分子を含む核酸試料とバーコードのセットとを含む反応混合物を形成する工程と、

(b) 前記試料の前記核酸分子にバーコードの前記セットを組み込む工程と、

(c) 前記単一細胞の前記核酸分子に組み込まれたバーコードの前記セットを特定し、それによって前記核酸試料の起源を特定する工程と

40

を含む、前記方法。

[本発明1002]

前記反応混合物を形成する工程が、

バーコードの前記セットに前記単一細胞を接触させることと、

前記単一細胞に由来する前記核酸試料を提供するために、前記細胞を溶解することとを含む、本発明1001の方法。

[本発明1003]

前記反応混合物を形成する工程が、

懸濁緩衝液中の細胞を提供することと、

前記単一細胞とバーコードの前記セットとを含む前記反応混合物を形成するための条件下

50

で、バーコードの前記セットを含むバーコーディング緩衝液に、前記細胞を含む前記懸濁緩衝液を接触させることとを含む、本発明1001の方法。

[本発明1004]

バーコードの前記セットが、独特の5'バーコードを各構築物が含む1つ以上の5'核酸構築物と、独特の3'バーコードを各構築物が含む1つ以上の3'核酸構築物とを含む、本発明1001~1003のいずれかの方法。

[本発明1005]

前記反応混合物を形成する工程が、

懸濁緩衝液中の細胞を提供することと、

前記単一細胞と少なくとも1つの5'核酸構築物と少なくとも1つの3'核酸構築物とを含む前記反応混合物を形成するための条件下で、前記5'核酸構築物と前記3'核酸構築物とを含むバーコーディング緩衝液に、前記懸濁緩衝液を接触させることと

を含む、本発明1004の方法。

[本発明1006]

前記反応混合物を形成する工程が、

懸濁緩衝液中の細胞を提供することと、

前記単一細胞と少なくとも1つの5'核酸構築物と少なくとも1つの3'核酸構築物とを含む前記反応混合物を形成するための条件下で、前記5'核酸構築物を含む第1のバーコーディング緩衝液及び前記3'核酸構築物を含む第2のバーコーディング緩衝液に、前記懸濁緩衝液を接触させることと

を含む、本発明1004の方法。

[本発明1007]

前記構築物が、独特の分子識別子(UMI)配列をさらに含む、本発明1004~1006のいずれかの方法。

[本発明1008]

各5'核酸構築物が、固体表面に結合された複数コピーで提供される、本発明1004~1007のいずれかの方法。

[本発明1009]

各3'核酸構築物が、固体表面に結合された複数コピーで提供される、本発明1004~1008のいずれかの方法。

[本発明1010]

前記固体表面が、ビーズ、マイクロビーズ、微粒子、ミクロスフェア、ナノ粒子、ナノビーズまたはヒドロゲルである、本発明1008または1009の方法。

[本発明1011]

前記反応混合物が、前記単一細胞に由来する前記核酸試料と、複数コピーの前記5'核酸構築物を含む少なくとも1つのビーズと、複数コピーの前記3'核酸構築物を含む少なくとも1つのビーズとを含む、本発明1004~1010のいずれかの方法。

[本発明1012]

各5'核酸構築物及び3'核酸構築物が、光切断可能なリンカーまたは制限酵素部位をさらに含む、本発明1004~1011のいずれかの方法。

[本発明1013]

各5'核酸構築物が光切断可能なリンカーをさらに含み、且つ各3'核酸構築物が制限酵素部位をさらに含む、本発明1004~1011のいずれかの方法。

[本発明1014]

各5'核酸構築物が制限酵素部位をさらに含み、且つ各3'核酸構築物が光切断可能なリンカーをさらに含む、本発明1004~1011のいずれかの方法。

[本発明1015]

前記反応混合物を制限酵素に接触させる工程と、

前記固体表面から前記5'核酸構築物及び前記3'核酸構築物を遊離させるために、前記反

10

20

30

40

50

応混合物をUV光に曝露する工程と  
をさらに含む、本発明1013または1014の方法。

[本発明1016]

前記5'核酸構築物及び前記3'核酸構築物が制限酵素部位をさらに含み、  
前記方法が、

前記固体表面から前記5'核酸構築物及び前記3'核酸構築物を遊離させるために、前記反  
応混合物を制限酵素に接触させる工程  
をさらに含む、本発明1012の方法。

[本発明1017]

前記5'核酸構築物及び前記3'核酸構築物が光切断可能なリンカーをさらに含み、  
前記方法が、

前記固体表面から前記5'核酸構築物及び前記3'核酸構築物を遊離させるために、前記反  
応混合物をUV光に曝露する工程  
をさらに含む、本発明1012の方法。

[本発明1018]

前記5'核酸構築物が、プロモータを含む二本鎖核酸構築物であり、且つ前記5'核酸構築  
物が、複数コピーの前記5'核酸構築物を作り出すための鋳型として役立つ、本発明1004  
~1017のいずれかの方法。

[本発明1019]

バーコードの前記セットが、1、2、3、4、5、6、7、または8個の5'核酸構築物と、1  
、2、3、4、5、6、7、または8個の3'核酸構築物とを含む、本発明1001~1018のい  
ずれかの方法。

[本発明1020]

前記核酸試料における5'バーコードと3'バーコードとの組み合わせが、前記核酸試料の  
起源を特定する、本発明1001~1019のいずれかの方法。

[本発明1021]

前記3'核酸構築物が結合配列をさらに含む、本発明1001~1020のいずれかの方法。

[本発明1022]

前記結合配列がポリ-T配列を含む、本発明1021の方法。

[本発明1023]

前記ポリ-T配列が、前記試料の前記核酸分子におけるポリ-A配列に結合する、本発  
明1022の方法。

[本発明1024]

前記3'核酸構築物における前記バーコードが、逆転写によって組み込まれる、本発明10  
01~1023のいずれかの方法。

[本発明1025]

前記逆転写が、前記3'バーコード配列を含む核酸分子を作り出す、本発明1024の方法。

[本発明1026]

前記逆転写酵素が、前記試料の前記核酸分子の末端に2~5個のシステイン残基を付加す  
る、本発明1025の方法。

[本発明1027]

前記2~5個のシステイン残基が、前記5'核酸構築物における結合配列に相補性の結合配  
列を提供する、本発明1026の方法。

[本発明1028]

前記5'核酸構築物における前記結合配列が、ポリ-G配列を含む、本発明1027の方法。

[本発明1029]

前記5'核酸構築物における前記結合配列が、前記試料の前記核酸分子におけるポリ-C  
配列に結合する、本発明1028の方法。

[本発明1030]

前記5'核酸構築物が、逆転写酵素を用いて前記核酸試料に組み込まれる、本発明1001~

10

20

30

40

50



1029のいずれかの方法。

[本発明1031]

前記逆転写酵素がM - M L V 酵素である、本発明1024 ~ 1030のいずれかの方法。

[本発明1032]

前記逆転写酵素がM M L V H - 逆転写酵素である、本発明1031の方法。

[本発明1033]

5' バーコード及び3' バーコードを各配列が含む複数の核酸配列の複数コピーを作り出すために、前記核酸試料がさらに増幅される、本発明1001 ~ 1032のいずれかの方法。

[本発明1034]

前記核酸分子がmRNAを含む、本発明1001 ~ 1033のいずれかの方法。

10

[本発明1035]

前記mRNAが、少なくとも1、3、10、30、100、300、1,000、3,000、10,000、30,000、100,000、300,000、または1,000,000個のmRNAを含む、本発明1034の方法。

[本発明1036]

前記核酸試料における前記バーコードが、配列決定によって特定される、本発明1001 ~ 1035のいずれかの方法。

[本発明1037]

バーコードの前記セットが、独特の5' バーコードを各構築物が含む少なくとも2つの5' 核酸構築物を含む、本発明1001 ~ 1003のいずれかの方法。

20

[本発明1038]

バーコードの前記セットが、独特の3' バーコードを各構築物が含む少なくとも2つの3' 核酸構築物を含む、本発明1001 ~ 1003のいずれかの方法。

[本発明1039]

前記核酸分子の1つ以上が、2つ以上の独特のバーコードを含む、本発明1001の方法。

[本発明1040]

前記独特のバーコードが、前記核酸分子の5' 末端、3' 末端、または5' 末端と3' 末端の双方に位置する、本発明1039の方法。

[本発明1041]

複数の反応混合物を形成する工程を含む、本発明1001 ~ 1040のいずれかの方法。

30

[本発明1042]

2つ以上の独特のバーコードが、前記反応混合物の30%、40%、50%、60%、70%、80%、90%または100%に存在する、本発明1041の方法。

[本発明1043]

前記複数の反応混合物を形成する工程が、0.5、0.6、0.7、0.8、0.9、1.0、2.0、3.0、4.0または5.0より大きいラムダで複数の反応混合物にバーコードの前記セットを加えることを含む、本発明1041の方法。

[本発明1044]

前記反応混合物が液滴である、本発明1001 ~ 1040のいずれかの方法。

[本発明1045]

核酸試料の起源を特定する方法であって、

40

(a) 独特の3' バーコード配列と結合配列とバーコード関連配列とを各構築物が含む3' 核酸構築物のセットを提供する工程と、

(b) 単一細胞に由来する核酸分子を含む核酸試料を提供する工程と、

(c) 前記核酸試料を3' 核酸構築物の前記セットに接触させる工程と、

(d) 前記バーコード関連配列から前記3' バーコード配列及び結合配列を分離する工程と、

(e) 前記3' 核酸構築物に由来する前記3' バーコード配列を前記核酸分子に組み込む工程と、

(f) 前記バーコード関連配列を用いて3' バーコード配列の前記セットを特定し、それによって前記単一細胞に由来する前記核酸試料の起源を特定する工程と

50

を含む、前記方法。

[本発明1046]

3'核酸構築物の前記セットが、2、3、4、5、6、7または8個の3'核酸構築物を含む、本発明1045の方法。

[本発明1047]

前記3'核酸構築物が、独特の分子識別子(UMI)配列をさらに含む、本発明1045または1046の方法。

[本発明1048]

前記3'核酸構築物が、ヘアピンを形成することができる、本発明1045~1047のいずれかの方法。

[本発明1049]

前記3'バーコード配列及び結合配列が、制限酵素配列または光切断可能なリンカーによって前記バーコード関連セグメントから分離される、本発明1045~1048のいずれかの方法。

[本発明1050]

前記3'核酸構築物の複数コピーが、固体表面に結合される、本発明1045~1049のいずれかの方法。

[本発明1051]

前記固体表面が、ビーズ、マイクロビーズ、微粒子、ミクロスフェア、ナノ粒子、ナノビーズまたはヒドロゲルである、本発明1050の方法。

[本発明1052]

制限酵素が、前記3'バーコード配列を前記固体表面から遊離させ、且つ前記結合配列を前記バーコード関連セグメントから分離する、本発明1051の方法。

[本発明1053]

UV光が、前記3'バーコード配列を前記固体表面から遊離させ、且つ前記結合配列を前記バーコード関連セグメントから分離する、本発明1051の方法。

[本発明1054]

前記3'核酸構築物における前記結合配列が、ポリ-T配列を含む、本発明1045~1053のいずれかの方法。

[本発明1055]

前記3'核酸構築物の前記結合配列が、前記試料の前記核酸分子におけるポリ-A配列に結合する、本発明1054の方法。

[本発明1056]

前記3'核酸構築物における前記バーコード配列が、逆転写によって組み込まれる、本発明1045~1055のいずれかの方法。

[本発明1057]

前記逆転写が、前記3'バーコード配列を含む核酸分子を作り出す、本発明1056の方法。

[本発明1058]

前記バーコード関連セグメントがプライミング配列を含む、本発明1045~1057のいずれかの方法。

[本発明1059]

前記プライミング配列が回文配列である、本発明1058の方法。

[本発明1060]

一方のバーコード関連セグメントの前記プライミング配列が、前記核酸試料における他の任意のバーコード関連セグメントのプライミング配列と結合することができる、本発明1058の方法。

[本発明1061]

双方のバーコード関連セグメントを単一核酸に組み込むために、互いに接触している前記バーコード関連セグメントが、DNAポリメラーゼによって伸長される、本発明1059の方法。

10

20

30

40

50

[本発明1062]

前記核酸分子に組み込まれた前記バーコード配列を特定する工程をさらに含む、本発明1045～1061のいずれかの方法。

[本発明1063]

前記核酸試料における前記バーコード配列が、配列決定によって特定される、本発明1062の方法。

[本発明1064]

前記バーコード関連セグメントが、配列決定によって特定される、本発明1062の方法。

【図面の簡単な説明】

【0007】

【図1】液滴形成について流れ集束の構成で油流入流路とその後の水性流入流路があるマイクロ流体液滴チップ：(1)懸濁緩衝液中の細胞及び(2)バーコーディングミックス中のバーコード付きビーズの模式図である。懸濁緩衝液及びバーコーディングミックスは、合流して液滴中で所望される最終的なミックスを構成する比で成分を融合させる、異なるマイクロ流体流路に位置する。バーコード付きビーズ及び細胞はポアソン分布によって記載されたように水性液滴に負荷される。液滴当たりのビーズ及び液滴当たりの細胞の平均値(ラムダ)はそれらの流入の流れにおけるその成分の濃度の関数である。液滴はその中でバーコーディング反応が行われる反応容器である。

【図2】2つのビーズのセット(3'及び5')の例となる組成を示す模式図である。各3'ビーズにおけるバーコード核酸は、ビーズからバーコード配列を切断するのに使用される制限部位(RS)と、下流の解析で使用される独特の分子識別子配列(UMI)と、バーコード配列(BC3)と、ポリ-T配列(TTT)とを含む。各5'ビーズにおけるバーコード核酸は、制限部位(RS)と、バーコード配列(BC5)と、3つのグアニン残基とを含む。

【図3】液滴の中で、細胞が溶解され、mRNA転写産物が遊離され、バーコーディング反応にアクセス可能であることを示す模式図である。制限酵素(RE)が制限部位でビーズから3'及び5'双方のバーコードを切断する。3'バーコードのポリ-T配列がmRNA転写産物の3'末端でポリ-A尾部にアニーリングすると逆転写が開始される。この3'バーコードのプライマーから出発して、逆転写(RT)酵素が5'から3'の方向に逆転写して3'バーコードとmRNA配列の双方を含む単一のcDNAアンプリコンを作り出す。転写産物の末端に達すると、RT酵素(M-MLV酵素)は、5'バーコードの3つのグアニン残基のためにアニーリング部位を提供する幾つかのシトシン残基を付加する。5'バーコードがいったんアニーリングされると、RT酵素は「鋳型切り替え」を行い、5'バーコード配列を同じcDNAに組み込む。

【図4】2つの3'バーコード付きビーズ及び3つの5'バーコード付きビーズが細胞と共に同時封入されることを示す模式図である。各3'バーコードは各5'バーコードを対合し、逆もまた同様で、この細胞を起源とするような転写産物すべてを特定するまとまりのある独特のバーコードのセットを形成する。

【図5】逆転写が時期尚早に終結すると、cDNAは5'バーコードを組み込めないことを示す模式図である。3'バーコードのライブラリは、1つのバーコードが複数の細胞と共に同時封入されることはまずあり得ないような方法(バーコードの総数という点で)で設計することができる。この場合、非完全長のバーコード付きcDNAは程度の高い信頼性で適正なプロファイルに割り当てることができる。

【図6】同じ2つの3'バーコード配列が2つの異なる液滴に同時封入されるが、それにもかかわらず2つのバーコードのセットは一方の液滴における余分な3'バーコードのせいで区別可能であることを示す模式図である。この第3の3'BCは他の液滴の2つの5'BCとは対合しないので、2つのバーコードのセットを分離するラインは明白に引くことができる。

【図7】例となるバーコードのセットの使用を示す模式図である。ここで、5'バーコードのビーズのセットは二本鎖DNAオリゴで官能化され、そのハイブリッド形成した(非共

10

20

30

40

50

有結合)鎖はT7プロモータ配列(「T7」と、バーコード配列と、3つのシトシン残基とを含む。制限酵素によってビーズ表面から切断されるのではなく、T7プロモータは、T7 RNAポリメラーゼがビーズ表面のDNAオリゴ鋳型に由来するRNAバーコードのコピーを結合し、転写できるようにする。これらのRNAバーコードのコピーは、完全長の3'バーコード付きcDNAの3'末端にそれらがアニーリングするのを可能にする3つのグアニン残基によって3'末端で終結するであろう。RT酵素は5'バーコードRNAオリゴ鋳型に対して鋳型切り替えを行い、5'バーコードをcDNAに組み込む。

【図8】各ビーズはその表面が一本鎖バーコードオリゴヌクレオチドで官能化されることを示す模式図である。オリゴは、二本鎖セグメントが制限部位を含有するヘアピン構造を形成する。制限酵素は双方の鎖を切断し、転写産物・バーコーディングセグメントとループ状セグメントを遊離する。ループ状セグメントは、転写産物・バーコーディングセグメント上のバーコードと同一であるまたは単にそれに関連するバーコードと「バーコード関連セグメント」(BAS)を含む。BASは、単一のアンプリコンにて2つのループ状セグメントの2つのバーコードと一緒に物理的に連結するDNAポリメラーゼ反応にて複数のループ状セグメントが互いにプライムオフできるようにする。これを達成するには、BASは回文状であってもよいし、またはその双方が各ビーズに組み込まれる違うセット(たとえば、2つの相補性の配列)で構成され得る。

【図9】mRNAのポリ-A尾部に選択的にアニーリングするポリ-T尾部を持つ転写産物・バーコーディングセグメントを示す模式図である。転写産物の配列はバーコードも運ぶ単一のcDNAに逆転写される。

【図10】2つのバーコード付きビーズが細胞と共に同時封入される場合、その細胞に由来する各転写産物は2つのバーコード(BC1aまたはBC2a)のいずれかによってバーコード付けされてもよいことを示す模式図である。同じ液滴は、それら2つの転写産物のバーコードのそれぞれに相当するループ状セグメントの多数のコピーも含有するであろう。

【図11】ループ状セグメントは、BASを介したループ状セグメント間の相互アニーリングが、ループ状構造で自己アニーリングされたままの単一のループ状セグメントに対比して好まれるように設計することができることを示す模式図である。DNAポリメラーゼを含んで各ループ状セグメントを伸長し、他のバーコードの相補体を組み込むことができる。

【図12】試料に関連付けたバーコードを示す模式図である。5'バーコード(5'BC)または3'バーコード(3'BC)は、いずれかの試料に関連付けることができる：(a)試料に関連付けた核酸の各末端で一方が、または(b)核酸の5'末端で双方が、または(c)核酸の3'末端で双方が、または(d~f)核酸の中で、または試料に関連付けられた2つの核酸が共有結合する場合、内部的に関連付けられ得る。バーコードは、ライゲーション技術またはポリメラーゼ鎖反応(PCR)のような増幅技術を含むが、これらに限定されない既知の方法を介して試料に付加することができる。

【図13】試料に関連付けたバーコードのセットの解析を示す模式図である。バーコード付きの核酸は、丸い頂点(5'バーコード)と四角の頂点(3'バーコード)が辺(核酸を表す)によって接続される二部グラフとして表すことができる。反応容器では、5'バーコードと3'バーコードの組み合わせすべてを核酸に関連付けることができるので、丸い頂点はすべて辺によって四角の頂点に接続されるはずであり、最大クリークを形成する。最大クリークを使用することは試料の中でバーコードと核酸を正しく関連付けるであろう。

【図14】試料に関連付けられた複数のバーコードのセットがバーコードの「不一致」を解決することを示す模式図である。(上)単独のバーコードのセットが使用され、異なる反応容器が偶然同一のバーコードのビーズを含有すれば、異なる試料に由来する核酸の誤った関連付けが生じるであろう。(真ん中)複数のバーコードのセットが使用されるとしたら、双方の反応容器における同一のバーコード(5'バーコード)は他のバーコードのセット(ここでは3'バーコード)との関連によって明白にすることができる。共有される5'バーコードは双方のクリークに属するであろうが、5'バーコードの頂点からの辺(核酸

10

20

30

40

50

を表す)はたった1つのクリークに独特に属するであろう。(下)2つの反応容器にて同一の5'バーコード及び3'バーコードがあるならば、辺は双方のクリークに属するであろうし、核酸がどの試料に関連付けられるのかは不明である。これは検出することができ、解析からは捨てられ得る(点線によって示される)。

【図15】観察された回数(×軸、カウント)に対比した配列読み取り対で観察されたバーコード対の頻度(y軸、割合)を示す模式図である。半分を超えるバーコード対が正確に1回観察されたが、それは直接除去されてもよい実験的ノイズ(たとえば、配列決定の誤差)を表すものと矛盾しない。わずかなバーコード対が10回を超えて観察された。50でのピンに50回以上観察されたバーコードの対すべてが含まれる。

【発明を実施するための形態】

【0008】

発明の詳細な説明

反応容器に単一の独特のバーコードを導入するように試みるのではなく、本出願は、生産反応のほとんどの場合またはすべての場合、その容器にてゲノムまたはトランスクリプトーム標的すべてについて起源の細胞を特定する区別可能なバーコードのセットを形成する複数のバーコードの計画的導入を指向する。これは幾つかの異なる方法のいずれかで実施することができ、決定的な共通する特徴は、対象とするゲノムまたはトランスクリプトーム標的にバーコードを付けることに加えて、単一の容器に複数のバーコードを導入する場合、容器におけるバーコードすべてを直接または間接的にその容器における他のバーコードすべてに関連付けることができるメカニズムも存在することである。単一バーコードではなく複数バーコードが個々の容器に割り当てられると、容器の大半が空なので非生産的であるという要件なしでバーコードの無作為な分配を適用することができる。提供される方法はまた、2つ以上のポアソン分布の交点を適用することの非効率性も低下させる。

【0009】

処理能力及び適用範囲という点でそのようなバーコーディングスキームの利益を実証するために、細胞とバーコードを受け取る容器の少なくとも95%が正確に1つの細胞と1つのバーコードを受け取る場合を考察する。多重バーコードのスキームでは、相当する要件は、受け取られるバーコードの数が今や無関係の実践的検討のためのものなので、細胞を受け取る容器の95%を超えるものが正確に1個の細胞を受け取ることである。容器当たり0.1個の細胞の平均値でこの評点に達し、その場合、少なくとも1個の細胞を受け取る容器の95.1%は正確に1個の細胞を受け取るであろう。容器当たりのバーコードの平均数は、バーコードの無駄のなさ、下流の解析の複雑さ及び他の因子の間での試料の適用範囲の間での所望の均衡に基づいて選択することができる。容器当たり3つのバーコードの平均値については、相当するポアソン分布は容器の95.02%が少なくとも1つのバーコードを受け取るであろうことを決定づけている。単一バーコードのアプローチでは、容器の0.24%未満が少なくとも1個の細胞と1つのバーコードを受け取るのでゲノムまたはトランスクリプトームのプロファイリングシステムに生産的に加わる機会を有するの対して、この類似の多重バーコードのアプローチでは、容器の9.04%が少なくとも1個の細胞と1つのバーコードを受け取る。これは潜在的に生産的な容器の数でほぼ40倍の増加であり、得られる処理能力(プロファイルされる細胞の数という点で)は従って高まると期待され得る。重要なことに、このことはプロファイリングされる細胞の数だけでなく比率も反映する。単一バーコードのアプローチでは容器の4.88%のみが少なくとも1つのバーコードを受け取るにすぎないので、それに応じて低い比率の細胞がプロファイリングされるのに対して、容器当たり3つのバーコードの平均値を伴う多重ピーズのアプローチでは、容器の95.02%が少なくとも1つのバーコードを受け取る。これは試料の適用範囲でほぼ20倍の増加を可能にし、容器当たりのバーコードのさらに大きな平均数が選択されれば一層さらに大きな前進を実現することができる。試料適用範囲のこの改善は極めて有益であり、相対的に稀な細胞クローンのプロファイルが求められる適用について特にそうである。

【0010】

10

20

30

40

50

多重バーコードのアプローチの別の利益は、単クローン性のバーコードのセットを作り出すことを目指す方法のほとんどまたはすべての共通する結果である多クローン性に対してスキームが安定していることである。単一バーコードのアプローチでは、2つ以上の違うバーコードを含有する単一のセットがそのバーコードのセットが割り当てられる細胞のプロファイルを混乱させるのに対して、所与の容器におけるバーコードすべてを一緒に包括的なバーコードのセットの中に関連付けるプロファイリングシステムは多クローン性のバーコードのセットに対処することができる。それはまた、ビーズのバーコーディング過程に対する収率及び／またはコストという点で有意な改善を可能にすることもできる。例として本明細書に記載されているエマルションPCRに基づく方法を用いて、容認できる閾値がバーコード付きビーズの少なくとも95%は単クローン性でなければならないということであれば、ビーズの10%未満はバーコードを受け取り、ビーズの残りは無駄であるという結果を伴って、乳化過程の間でのバーコードコピーのラムダは0.1に設定されなければならない。システムが多クローン性のビーズのさらに大きな割合が認容できる。たとえば、ビーズの99%が単コード化、二重コード化、または三重コード化されなければならない（この場合、約70%が単コード化される）のであれば、バーコードコピーのラムダは0.7に設定することができ、ビーズの>50%は少なくとも1つのバーコードを受け取るという結果を伴う。これは生産収量で5倍を超える増加をもたらす。

#### 【0011】

（表1）ラムダ = 0.1の場合に対比させたバーコード付きビーズの収量の対応する倍率増加を伴った、emPCRビーズバーコーディング反応における液滴当たりのバーコードの平均コピー数（ラムダ）の関数としてのビーズのクローン性の百分率

b.c. ラムダ	クローン性				0.1ラムダに対比させた処理能力増加倍率
	単一	二重	三重	四重	
0.1	95.1%	4.8%	0.2%	0.0%	NA
0.2	90.3%	9.0%	0.6%	0.0%	1.9
0.3	85.7%	12.9%	1.3%	0.1%	2.72
0.4	81.3%	16.3%	2.2%	0.2%	3.46
0.5	77.1%	19.3%	3.2%	0.4%	4.13
0.6	73.0%	21.9%	4.4%	0.7%	4.74
0.7	69.1%	24.2%	5.6%	1.0%	5.29

#### 【0012】

単一バーコードによってではなく、バーコードの区別できるセットを用いて個々の遺伝子またはトランスクリプトームのプロファイルを特定することによって、スキームはバーコードの任意の1つが複数の反応容器に入る例を特定することができる。単一バーコード配列の2つの他の違うセットとの関連付けは、そのような場合が特定され、別の解析メカニズムによって分類されるまたは廃棄されるようにする。

#### 【0013】

バーコードと試料に関連付けられた核酸との正確な割り当ては、グラフ理論を用いて、特に二部グラフを用いて及び最大クリークを探索して行うことができる。二部グラフは、頂点（またはノード）を、各辺が第1の集合における頂点を第2の集合における頂点に接続する2つの互いに素な集合に分割することができるグラフである。我々の目的では、頂点の2つの集合は2つのバーコードのセット（5'バーコード及び3'バーコードのセット）

を表し、頂点を接続する辺はその中に組み込まれたその5'バーコードとその特定の3'バーコードを有する核酸を表す(図13)。反応容器では、5'バーコードと3'バーコードの組み合わせすべてを核酸に関連付けることができるので、丸い頂点はすべて辺によって四角の頂点すべてに接続されて最大クリークを形成するはずである。クリークは、クリークにおける2つの違う頂点がどれも皆隣接するような無向グラフの頂点のサブセットである。最大クリークは、1つ以上の隣接する頂点を含めることによって伸長できない、それがさらに大きなクリークのサブセットではないことを意味するクリークである。コンピュータプログラムを用いて最大クリークを情報学的に解き明かすことは試料内でバーコードと核酸を正しく関連付けるであろう(図13及び14を参照のこと)。図14は複数バーコードのセットを使用することの追加の利点を説明している。いずれか1つのバーコードが複数の反応容器に入る例(またはバーコード「不一致」)を検出し、解決することができる。これはたった1つのバーコードを使用する場合可能ではない。

10

#### 【0014】

機能的なセット内での複数バーコードの組み合わせはさらに大きな多様性をコード空間に本質的に加える。たとえば、各cDNAにコピーが各末端(5'及び3')でバーコード1つを受け取り、各セット(5'及び3')でのコード空間が100である本発明の考えられる一実施形態では、システムのコード空間全体は $100 \times 100 = 10,000$ の可能なバーコードの組み合わせである。さらに大きなバーコーディングの忠実性で多数の細胞にバーコードを付けるために、これは作り出されるバーコード付きビーズのさらに小さなライブラリを可能にすることができる。

20

#### 【0015】

便宜上、バーコードのセットは5'バーコード及び3'バーコードと呼ばれる。複数のバーコードのセットを用いて核酸を適当な反応容器に割り当てるので核酸試料の起源を特定することができる。バーコードのセットを5'末端及び3'末端のいずれかで、または5'末端で双方を、または3'末端で双方を、または核酸試料内で内部的に核酸試料に関連付けることができる(図12を参照のこと)。たとえば、増幅(たとえば、ポリメラーゼ鎖反応(PCR)、ライゲーション反応及び転位反応(Green, et al., Molecular Cloning, A Laboratory Manual, 4th Edition, Cold Spring Harbor Press (2012)を参照してください)のような、しかしこれらに限定されない、当該技術で既知の方法を用いてバーコードを核酸に付加することができる。少なくとも2つ以上のバーコードセットが核酸試料に関連付けられる限り、バーコード付きビーズ入りの負荷反応容器におけるポアソン統計を克服すること及びいずれか1つのバーコードが複数の反応容器に入るバーコード「不一致」を解決することを含む本明細書で言及される利点すべてを伴う適当な反応容器(及び従って試料)に核酸を割り当てることができ、その際、任意の1つのバーコードが複数の反応容器に入る。

30

#### 【0016】

複数のバーコード付きビーズを単一の液滴に含めることは、それを介して遺伝子または転写産物のバーコーディング反応が進む分子生物学から利益が得られる可能性もある。バーコード核酸分子はバーコーディング反応の反応物質であり、反応の効率はその反応物質の供給に直接関係するので、当然、各反応で複数のバーコードを含むことはバーコード付き産物の高い収量をもたらすであろう。

40

#### 【0017】

本明細書で提供されているのは、核酸試料の起源を特定する方法である。該方法は、単一細胞に由来する核酸分子を含む核酸試料とバーコードのセットとを含む反応混合物を形成する工程と、試料の核酸分子にバーコードのセットを組み込む工程と、単一細胞の核酸分子に組み込まれたバーコードのセットを特定し、それによって核酸試料の起源を特定する工程とを含む。任意で、バーコードのセットは各構築物が独特の5'バーコードを含む少なくとも2つの5'核酸構築物を含む。任意で、バーコードのセットは、各構築物が独特の3'バーコードを含む少なくとも2つの3'核酸構築物を含む。任意で、核酸分子の1つ以上

50

は、2つ以上の独特のバーコードを含む。任意で、独特のバーコードは核酸分子の5'末端、3'末端、または5'末端と3'末端の双方に位置する。任意で、方法は複数の反応混合物を形成する工程を含む。任意で、2つ以上の独特のバーコードは、反応混合物の30%、40%、50%、60%、70%、80%、90%または100%に存在する。任意で、複数の反応混合物を形成する工程は、0.5、0.6、0.7、0.8、0.9、1.0、2.0、3.0、4.0または5.0を超えるラムダで複数の反応混合物にバーコードのセットを加えることを含む。

#### 【0018】

提供されるのはまた、各構築物が独特の3'バーコード配列を含む3'核酸構築物のセットを提供する工程と、単一細胞に由来する核酸分子を含む核酸試料を提供する工程と、3'核酸構築物のセットに核酸試料を接触させる工程と、3'核酸構築物に由来する3'バーコード配列を核酸分子に組み込む工程と、3'バーコード配列のセットを特定し、それによって単一細胞に由来する核酸試料の起源を特定する工程とを含む、核酸試料の起源を特定する方法である。

#### 【0019】

用語が本明細書で使用されるとき、配列をポリヌクレオチドに「組み込むこと」は、たとえば、ホスホジエステル結合によってポリヌクレオチドの3'または5'の末端にてポリヌクレオチドの残りに一連のヌクレオチドを共有結合することを指し、その際、ヌクレオチドは配列によって指示される順序で連結される。ポリヌクレオチドが配列またはその相補体を含むのであれば、その配列はそのポリヌクレオチドに「組み込まれて」おり、または同等にそのポリヌクレオチドはその配列を「組み込む」。配列のポリヌクレオチドへの組込みは酵素的に（たとえば、ライゲーションまたは重合によって）または化学合成（たとえば、ホスホロアミダイト化学反応によって）を用いて生じることができる。

#### 【0020】

用語「プライマー」は、標的核酸に選択的にアニーリングし、またはハイブリッド形成することができるポリヌクレオチド（オリゴヌクレオチド）またはその類似体を指す。プライマーは、たとえば、適当な酵素（複数可）、補因子、基質、たとえば、ヌクレオチド（dNTP）等の存在下で好適な条件下にてDNA合成のための開始プライマーとして役立つ。プライマーは、対応するポリヌクレオチド鋳型、隣接配列またはプライマーの3'末端からの増幅産物に対して相補性の配列の合成を可能にする。通常、プライマーは約10～100ヌクレオチドの間の長さであることができる。プライマーの結合部位または汎用プライミング配列はプライマーが結合して増幅を開始することができる核酸配列を指す。

#### 【0021】

本明細書で使用されるとき、用語「増幅する」及び「増幅」は、配列またはその相補体も含有するさらに多くのポリヌクレオチドを生成するように全体としてまたは部分的にポリヌクレオチドの配列を酵素的にコピーすることを指す。コピーされる配列は鋳型配列と呼ばれる。増幅の例には、RNAポリメラーゼによるDNAが鋳型のRNA合成、逆転写酵素によるRNAが鋳型の第1の鎖のcDNAの合成、及び熱安定性のDNAポリメラーゼを用いたDNAが鋳型のPCR増幅が挙げられる。増幅にはプライマーによる伸長反応すべてが含まれる。

#### 【0022】

本明細書で使用されるとき、用語「等温の」は、一定の温度または温度の範囲で行われる、酵素反応などの反応を指す。

#### 【0023】

用語「関連した」は、試料と、その試料を起源とするまたはそれに由来するDNA分子、RNA分子または他のポリヌクレオチドとの間の関係を指すのに本明細書で使用される。その用語は、1つのバーコード配列と別のバーコード配列との間、または1つのバーコード配列と他のバーコードのセットとの間の関係を指すのに使用することができる。たとえば、バーコードがバーコードの同じセットにあるのであれば、そのバーコードは別のバーコードに関連する。或いは、バーコード配列がバーコード配列のセットにおけるバーコー



ドの1つであるならば、そのバーコードはバーコードのセットに関連することができる。ポリヌクレオチドは、それが内在性ポリヌクレオチドならば、すなわち、試料が選択される時点でそれが試料に存在するならば、または内在性ポリヌクレオチドに由来するならば、その試料に関連する。たとえば、細胞にとって内在性のmRNAはその細胞に関連する。これらのmRNAの逆転写から生じるcDNA及びcDNAのPCR増幅から生じるDNAアンプリコンはそのmRNAの配列を含有し、その細胞にも関連する。試料に関連するポリヌクレオチドは試料に位置する必要はなくまたは試料で合成される必要はなく、試料が破壊された後（たとえば、細胞が溶解された後）でさえ、試料に関連すると見なされる。分子バーコーディングの技法または他の技法を用いて混合物におけるどのポリヌクレオチドが特定の試料に関連するのかを判定することができる。

10

#### 【0024】

用語が本明細書で使用されるとき、「反応容量」（または同等に「容器」または「区画」）は、液体、たとえば、水溶液の容量が保持され、液体または周りの媒体の他のそのような容量から隔離された（たとえば、単離された）ままであることができる空間である。反応容量とその周囲の間の隔離は反応容量の周りの固体バリアからまたは位相分離から生じることができる。たとえば、疎水性のキャリア流体に浮遊された水性のマイクロ流体液滴は、水がキャリア流体と非混和性であるので、反応容量を構成することができる。従って、キャリア流体にて互いに分離されている2つの液滴は隔離されたままであり、一方の液滴に溶解された核酸または他の親水性の種は液滴から出ることはできず、または別の液滴に移行することはできない。反応容量はまた、たとえば、フラスコ、ビーカー、遠心管、及びマルチウェルプレートのウェルによっても定義することができる。

20

#### 【0025】

たとえば、単一細胞に由来する試料に関連したRNAにバーコードのセットを「付加すること」には、バーコードと単一細胞または単一細胞に由来するRNAとを含有する反応混合物を形成することが含まれるので、RNA及びバーコードがバーコーディング反応に加わることができる。いったん付加されると、バーコードは、たとえば、RNAとのハイブリッド形成によって1つ以上のRNAと直接反応することができ、または重合反応もしくはRNA分子が鋳型として役立つ一連の反応（たとえば、逆転写またはRT-PCR）に加わることができる。

#### 【0026】

用語「ポリヌクレオチド（複数可）」は、DNA分子及びRNA分子及びそれらの類似体（たとえば、ヌクレオチド類似体を用いて、または核酸化学を用いて生成されるDNAまたはRNA）のような核酸を指す。所望のように、ポリヌクレオチドは、たとえば、当該技術で承認されている核酸化学を用いて合成で、またはたとえば、ポリメラーゼを用いて酵素的に作られてもよく、所望であれば修飾することができる。典型的な修飾には、メチル化、ピオチン化、及び他の当該技術で知られる修飾が挙げられる。加えて、ポリヌクレオチドは一本鎖または二本鎖であることができ、所望であれば、検出可能な部分に連結することができる。一部の態様では、ポリヌクレオチドは、たとえば、DNAとRNAを含むハイブリッド分子を含むことができる。

30

#### 【0027】

「G」、「C」、「A」、「T」及び「U」はそれぞれ一般にそれぞれ塩基としてグアニン、シトシン、アデニン、チミジン及びウラシルを含有するヌクレオチドを表す。しかしながら、用語「リボヌクレオチド」または「ヌクレオチド」は、修飾されたヌクレオチドまたは代理置換部分も指すことができることが理解されるであろう。当業者は、そのような置換部分を持つヌクレオチドを含むオリゴヌクレオチドの塩基対合特性を実質的に変えることなく、グアニン、シトシン、アデニン及びウラシルが他の部分によって置き換えられてもよいことを十分に承知している。たとえば、限定しないで、塩基としてイノシンを含むヌクレオチドはアデニン、シトシンまたはウラシルを含有するヌクレオチドと塩基対合してもよい。従って、ウラシル、グアニンまたはアデニンを含有するヌクレオチドは、たとえば、イノシンを含有するヌクレオチドによってヌクレオチド配列にて置き換えられ

40

50

てもよい。別の例では、オリゴヌクレオチドにおけるどこかでのアデニン及びシトシンはそれぞれグアニン及びウラシルで置き換えて標的 mRNA との G - U のゆらぎ塩基対合を形成することができる。そのような置換部分を含有する配列は本明細書に記載されている組成物及び方法に好適である。

#### 【0028】

本明細書で使用されるとき、且つ特に指示されない限り、用語「相補性の」は、第2のヌクレオチド配列に関連して第1のヌクレオチド配列を記載するのに使用される場合、当業者によって理解されるように、第1のヌクレオチド配列を含むポリヌクレオチドの特定の条件下で第2のヌクレオチド配列を含むポリヌクレオチドとハイブリッド形成し、二本鎖構造を形成する能力を指す。そのような条件は、たとえば、ストリンジェントな条件であることができ、ストリンジェントな条件には 400 mM の NaCl、40 mM の PIPES、pH 6.4、1 mM の EDTA で 50 または 70 にて 12 ~ 16 時間にその後の洗浄が含まれてもよい。生物の内部で遭遇してもよいような生理的に関連する条件のような他の条件を適用することができる。当業者は、ハイブリッド形成したヌクレオチドの最終的な適用に従って2つの配列の相補性を調べるのに最も適する一揃いの条件を決定することができるであろう。

10

#### 【0029】

相補性の配列は、一方または双方のヌクレオチド配列の長さまたは長さの一部にわたる第2のヌクレオチド配列を含むポリヌクレオチドの領域に対する第1のヌクレオチド配列を含むポリヌクレオチドの領域の塩基対合を含む。そのような配列は本明細書では互いに関連して「相補性である」と呼ぶことができる。しかしながら、本明細書で第1の配列が第2の配列に関して「実質的に相補性である」と言われる場合、2つの配列は相補性であることができ、またはそれらは、塩基が対合する領域内で1つ以上の、しかし一般に約5、4、3または2つを超えないミスマッチの塩基対を含んでもよい。ミスマッチの塩基対を伴う2つの配列については、2つのヌクレオチド配列が塩基の対合を介して互いに結合する限り、配列は「実質的に相補性である」と見なされるであろう。

20

#### 【0030】

「相補性の」配列は本明細書で使用されるとき、ハイブリッド形成するその能力に関する上記の実施形態が満たされる限り、非ワトソン・クリック塩基対及び/または非天然のヌクレオチドや修飾されたヌクレオチドから形成される塩基対を含んでもよく、またはそれらから全体的に形成されてもよい。そのような非ワトソン・クリック塩基対には、G : U ゆらぎ塩基対合またはフーグスティーン型塩基対合が挙げられるが、これらに限定されない。

30

#### 【0031】

2つ以上の核酸またはポリペプチドの配列の文脈で用語、パーセント「同一性」は、以下に記載されている配列比較アルゴリズム（たとえば、BLASTP及びBLASTNまたは当業者に利用可能な他のアルゴリズム）の1つを用いて、または目視検査によって測定されるような、最大一致について比較し、並べたとき、特定の比率の同一であるヌクレオチドまたはアミノ酸の残基を有する2つ以上の配列または部分配列を指す。適用に応じて、パーセント「同一性」は比較される配列の領域、たとえば、機能的ドメインにわたって存在することができ、または代わりに比較される2つの配列の完全長にわたって存在することができる。

40

#### 【0032】

配列比較については通常、一方の配列は試験配列が比較される参照配列として役割を果たす。配列比較アルゴリズムを使用する場合、試験配列及び参照配列をコンピュータに入力し、必要に応じて部分配列座標を指定し、配列アルゴリズムのプログラムパラメータを指定する。次いで配列比較アルゴリズムは、指定されたプログラムパラメータに基づいて参照配列と比べた試験配列（複数可）のパーセント配列同一性を算出する。

#### 【0033】

比較のための配列の最適な整列化は、たとえば、Smith及びWaterman, Ad

50

v. Appl. Math. 2: 482 (1981) の局所相同性アルゴリズムによって、Needleman 及び Wunsch, J. Mol. Biol. 48: 443 (1970) の相同性整列化アルゴリズムによって、Pearson 及び Lipman, Proc. Nat'l. Acad. Sci. USA, 85: 2444 (1988) の類似性法についての検索によって、これらのアルゴリズムのコンピュータ化された実施 (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.) によって、または目視検査 (一般に、Ausubel, et al., 上記を参照のこと) によって行うことができる。

10

#### 【0034】

パーセント配列同一性及び配列類似性を決定するのに好適であるアルゴリズムの一例は、Altschul, et al., J. Mol. Biol. 215: 403-410 (1990) に記載されている BLAST アルゴリズムである。BLAST 解析を実施するためのソフトウェアは全米バイオテクノロジー情報センターのウェブサイトを通じて公的に利用できる。BLAST アルゴリズムのパラメータ、W、T 及び X は整列化の感度と速度を決定する。BLASTN のプログラム (ヌクレオチド配列用) は、11 の語長 (W)、期待値 (E) または 10、M = 5、N = -4 及び双方の鎖の比較を初期設定として使用する。

#### 【0035】

20

同一の配列は、一方または双方のヌクレオチド配列の全長にわたって第2のヌクレオチド配列を含むポリヌクレオチドに対する第1のヌクレオチド配列を含むポリヌクレオチドの100%同一性を含む。そのような配列は本明細書では互いに関して「完全に同一である」ということができる。しかしながら、一部の態様では、第1の配列が本明細書では第2の配列に関して「実質的に同一である」と言われる場合、2つの配列は完全に相補性であることができ、またはそれらは整列化の際、1つ以上の、しかし一般に約5、4、3または2つを超えないミスマッチのヌクレオチドを有してもよい。一部の態様では、第1の配列が本明細書では第2の配列に関して「実質的に同一である」と言われる場合、2つの配列は完全に相補性であることができ、またはそれらは、互いに少なくとも約50、60、70、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、または99%同一であってもよい。本明細書に記載されている2つのヌクレオチド配列のパーセント同一性を決定するには、上述の BLASTN の初期設定の設定を使用することができる。

30

#### 【0036】

第1の配列が本明細書では第2の配列の同一性に関して「違う」と言われる場合、2つの配列は整列化の際、少なくとも1つ以上のミスマッチのヌクレオチドを有する。一部の態様では、違う配列は整列化の際、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50個以上のミスマッチのヌクレオチドを有することができる。一部の態様では、違う配列は互いに約1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、60、70、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99または100%未満同一であることができる。一部の態様では、第1の配列が本明細書では第2の配列に関して「違う」と言われる場合、2つの配列は実質的にまたは完全に同一の配列を有することができるが、代わりに配列内の修飾の異なるパターンに基づいて互いに異なることができる。そのような修飾、たとえば、メチル化は当該技術で一般に

40

50

知られている。

【0037】

一部の態様では、ポリヌクレオチドはポリヌクレオチドのライブラリに存在することができる。一部の態様では、ポリヌクレオチドのライブラリは複数のポリヌクレオチドを含むことができる。一部の態様では、複数のポリヌクレオチドにおける各ポリヌクレオチドは単一試料に由来することができる。一部の態様では、単一試料は単一細胞である。

【0038】

ヌクレオチド配列を記載するのに本明細書では従来の表記法を使用する：一本鎖ヌクレオチド配列の左側末端は5'末端であり；二本鎖ヌクレオチド配列の左側方向は5'方向と呼ばれる。新生RNA転写産物への5'から3'へのヌクレオチドの付加の方向は転写方向と呼ばれる。mRNAと同じ配列を有するDNA鎖は「コーディング鎖」と呼ばれる。

10

【0039】

用語「メッセンジャーRNA」または「mRNA」はイントロンがなく、ポリペプチドに翻訳することができるRNAを指す。

【0040】

用語「cDNA」は一本鎖または二本鎖の形態でmRNAに対して相補性であるまたは同一であるDNAを指す。

【0041】

用語「アンプリコン」は核酸増幅反応、たとえば、RT-PCRの増幅産物を指す。

【0042】

用語「ハイブリッド形成する」は相補性の核酸との配列に特異的な非共有結合の相互作用を指す。ハイブリッド形成は核酸配列の全部または一部に生じてよい。当業者は核酸二本鎖またはハイブリッドの安定性はT<sub>m</sub>によって決定され得ることを認識するであろう。ハイブリッド形成の条件に関する追加の指針はCurrent Protocols in Molecular Biology, John Wiley & Sons, N.Y., 1989, 6.3.1-6.3.6及びSambrook, et al., Molecular Cloning, a Laboratory Manual, Cold Spring Harbor Laboratory Press, 1989, Vol. 3にて見いだされてもよい。

20

【0043】

本明細書で使用されるとき、「領域」はポリヌクレオチドのヌクレオチド配列の連続する部分を指す。領域の例は本明細書に記載されており、それらには同定領域、試料同定領域、バーコード領域、結合領域、cDNA領域、等が挙げられる。一部の態様では、ポリヌクレオチドは1つ以上の領域を含むことができる。一部の態様では、ポリヌクレオチドは、2つ未満、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50個、またはそれ以上の領域を含むことができる。一部の態様では、領域は結合することができる。一部の態様では、領域は操作可能に結合することができる。一部の態様では、領域は物理的に結合することができる。

30

【0044】

本明細書で使用されるとき、「バーコード」または「バーコード配列」は、たとえば、少なくとも1つのヌクレオチド配列を後で特定するためにその少なくとも1つのヌクレオチド配列に結合することができる独特の配列標識を指す。

【0045】

本明細書で使用されるとき、「バーコードセット」または「バーコードのセット」は試料に由来するヌクレオチド配列に結合することができる配列の独特のセットを指し、その際、たとえば、ヌクレオチド配列を後で特定するためにそのヌクレオチド配列はセットにおける1つのバーコード配列に結合する。

40

【0046】

50

用語「バーコード構築物」、「バーコード付き構築物」及び「バーコード構築物分子」は独特のバーコード配列を含むオリゴヌクレオチドを指すのに本明細書では相互交換可能に使用される。

【0047】

用語「3'核酸構築物」または「3'核酸バーコード構築物」は、3'から「5'核酸バーコード構築物」であるバーコード配列を含むオリゴヌクレオチドを指す。3'核酸構築物は核酸分子、たとえば、mRNAの3'末端に組み込むことができる。用語「3'バーコード」または「3'バーコード配列」は3'核酸構築物または3'核酸バーコード構築物に位置するバーコードを指す。通常、3'核酸構築物は核酸分子の3'末端で配列に相補性の配列を含有し、たとえば、3'核酸構築物はmRNAのポリ-A配列に相補性のポリ-T配列を含む。10  
通常、3'構築物はmRNA配列の3'末端に結合し、逆転写されてその5'末端で3'核酸構築物を含むcDNA分子を生成する。たとえば、図2及び3を参照のこと。それはまた、5'末端で、3'末端で、または内部的に（それは核酸の3'末端で及び別の核酸の5'末端で組み込まれて1つのさらに長い核酸を形成する）核酸に組み込むこともできる。たとえば、図12を参照のこと。3'核酸バーコード構築物は、バーコード配列に加えて、汎用プライミング配列、または汎用プライミング領域、及び結合部位を含むことができる。

【0048】

用語「5'核酸構築物」または「5'核酸バーコード構築物」は5'から「3'核酸バーコード構築物」であるバーコード配列を含むオリゴヌクレオチドを指す。5'核酸構築物は核酸分子、たとえば、mRNAの5'末端またはcDNAの3'末端に組み込むことができる。用20  
語「5'バーコード」または「5'バーコード配列」は5'核酸構築物または5'核酸バーコード構築物に位置するバーコードを指す。通常、5'核酸構築物は核酸分子の3'末端で配列に相補性の配列を含有し、たとえば、5'核酸構築物はcDNA配列の3'末端に組み込まれるのポリ-C配列に相補性のポリ-G配列を含む。5'核酸構築物は、3'核酸構築物を含むcDNA分子の3'末端に組み込むことができる。たとえば、図2及び3を参照のこと。それはまた、5'末端で、3'末端で、または内部的に（それは核酸の3'末端で及び別の核酸の5'末端で組み込まれて1つのさらに長い核酸を形成する）核酸に組み込むこともできる。たとえば、図12を参照のこと。5'核酸バーコード構築物は、バーコード配列に加えて、汎用プライミング配列、または汎用プライミング領域、及び結合部位を含むことができる。30

【0049】

本明細書で使用されるとき、「バーコード構築物ビーズ」は1つ以上のバーコード構築物に結合されるビーズを指す。

【0050】

本明細書で使用されるとき、バーコーディングミックスは1つ以上のバーコード構築物を含む組成物を指す。ミックスは、細胞を溶解すること、逆転写、制限酵素消化及び/またはバーコーディング反応を行うことのための追加の試薬を含むことができる。

【0051】

本明細書で使用されるとき、「バーコーディング」または「バーコーディング反応」はバーコード配列またはバーコード配列の相補体を核酸に連結する反応を指す。バーコード構築物は必ずしも核酸と共有結合する必要はないが、バーコード配列の情報自体が核酸に結び付けられるまたは組み込まれる。「バーコーディング核酸」、「バーコーディング細胞」、40  
「細胞に由来するバーコーディング核酸」、「反応容器に由来するバーコーディング核酸」、及び「バーコーディング反応容器」は相互交換可能に使用される。

【0052】

本明細書で使用されるとき、「同定領域」は、たとえば、少なくとも1つのヌクレオチド配列の後での同定のために、その少なくとも1つのヌクレオチド配列に結合することができるヌクレオチド配列の標識（たとえば、独特のバーコード配列）を指す。一部の態様では、バーコードのセットは試料の同定領域として使用される。

【0053】

本明細書で使用されるとき、「アダプター領域」または「アダプター分子」は第1のヌクレオチド配列を第2のヌクレオチド配列に結合するリンカーを指す。一部の態様では、アダプター領域はリンカーとして作用するヌクレオチド配列の連続する部分を含むことができる。

【0054】

用語「試料」は、対象（たとえば、哺乳類対象、動物対象、ヒト対象、または非ヒト動物対象）から採取されるRNA、DNA、単一細胞または複数細胞、または細胞の断片、または体液のアリコートを含むことができる。試料は、遠心分離、静脈穿刺、採血、切除、ぬぐい取り、射精、マッサージ、生検、針吸引、洗浄試料、剥離、外科切除、レーザー捕捉顕微解剖、勾配分離、または当該技術で既知の介入または他の手段を含む、今や既知のまたは最近発見された任意の手段を用いて当業者によって選択され得る。試料はまた、対象とする試料に関連することが知られている1つ以上のマーカーを用いて当業者によって選択され得る。試料はまた、細胞選別及びFACSのような当該技術で既知の方法を用いて選択され得る。

10

【0055】

本発明の方法は所望の試料で実践することができる。一部の実施形態では、各試料は細胞を含み、たとえば、単一細胞であることができる。細胞はマイクロ流液体滴のような反応容量の中に入れることができ、所望であれば、溶解して反応容量の中にRNA分子を遊離することができる。この目的で、細胞は都合のよいときに溶解緩衝液に接触させることができる。

20

【0056】

本明細書で提供されているのは、核酸試料の起源を特定する方法である。任意で、方法は、単一細胞に由来する核酸分子を含む核酸試料とバーコードのセットとを含む反応混合物を形成する工程と、試料の核酸分子にバーコードのセットを組み込む工程と、単一細胞の核酸分子に組み込まれたバーコードのセットを特定し、それによって核酸試料の起源を特定する工程とを含む。任意で、反応混合物を形成する工程は、バーコードのセットに単一細胞を接触させることと、細胞を溶解して単一細胞に由来する核酸試料を提供することとを含む。任意で、反応混合物を形成する工程は、単一細胞とバーコードのセットとを含む反応混合物を形成するための条件下で、懸濁緩衝液中の細胞を提供することと、細胞を含む懸濁緩衝液を、バーコードのセットを含むバーコーディング緩衝液に接触させることとを含む。任意で、反応混合物は液滴である。

30

【0057】

本明細書で提供されるとき、バーコードのセットは少なくとも2つの独特のバーコードの1つ以上のコピーを含む。従って、たとえば、バーコードのセットはそれぞれ独特の5'バーコードを含む5'核酸構築物の1つ以上と、それぞれ独特の3'バーコードを含む3'核酸構築物の1つ以上とを含むことができる。任意で、バーコードのセットは各構築物が独特の3'バーコードを含む少なくとも2つの3'核酸構築物を含む。任意で、バーコードのセットは1、2、3、4、5、6、7、または8個の5'核酸構築物と1、2、3、4、5、6、7、または8個の3'核酸構築物とを含む。1、2、3、4、5、6、7、または8種の核酸構築物のそれぞれは独特のバーコード配列を有し、複数のコピーで提供され得る。任意で、バーコードのセットは、2つの異なる5'バーコードを含む2つの異なる5'核酸構築物と、2つの異なる3'バーコードを含む2つの異なる3'核酸構築物とを含む。任意で、バーコードのセットは3つの異なる5'バーコードを含む3つの異なる5'核酸構築物と3つの異なる3'バーコードを含む3つの異なる3'核酸構築物とを含む。任意で、バーコードのセットは、2つの異なる3'バーコードを含む2つの異なる3'核酸構築物と、3つの異なる5'バーコードを含む3つの異なる5'核酸構築物とを含む。各種類の核酸構築物の複数のコピーには同じバーコードを有する同じ種類の複数のコピーのそれぞれが提供され得る。従って、たとえば、バーコードのセットは、異なる種類のそれぞれが独特の5'または3'のバーコード配列を有する、1、2、3、4、5、6、7、または8つの異なる種類の5'核酸または3'核酸の構築物の複数のコピーを含むことができる。言い方を変えれば

40

50

、バーコードのセットは、第 1、第 2、第 3、第 4、第 5、第 6、第 7、または第 8 の種類のうちの 1 つ以上の 5 ' 核酸構築物の 1 つ以上のコピーと、第 1、第 2、第 3、第 4、第 5、第 6、第 7、または第 8 の種類のうちの 1 つ以上の 3 ' 核酸構築物の 1 つ以上のコピーとを含むことができる。任意で、バーコードのセットは、第 1、第 2、第 3、第 4、第 5、第 6、第 7、または第 8 の種類のうちの 1 つ以上の 3 ' 核酸構築物の 1 つ以上のコピーを含む。構築物は RNA または DNA の分子、または RNA-DNA のハイブリッドであることができる。たとえば、構築物は共通のオリゴヌクレオチド鎖にて DNA ヌクレオチドに共有結合した RNA ヌクレオチドを含むことができる。構築物は一本鎖または二本鎖であることもできる。二本鎖であれば、構築物は 1 つ以上の平滑末端または一本鎖オーバーハングを伴う末端を有することができる。

10

**【 0 0 5 8 】**

本明細書で提供されているのはまた、3 ' 核酸構築物のセットを用いて核酸試料の起源を特定する方法である。方法は、各構築物が独特の 3 ' バーコード配列、結合配列及びバーコード関連配列を含む 3 ' 核酸構築物のセットを提供する工程と、単一細胞に由来する核酸分子を含む核酸試料を提供する工程と、核酸試料を 3 ' 核酸構築物のセットに接触させる工程と、3 ' バーコード配列及び結合配列をバーコード関連配列から分離する工程と、3 ' 核酸構築物に由来する 3 ' バーコード配列を核酸分子に組み込む工程と、バーコード関連配列を用いて 3 ' バーコード配列のセットを特定し、それによって単一細胞に由来する核酸試料の起源を特定する工程とを含む。任意で、3 ' 核酸構築物のセットは 2、3、4、5、6、7、または 8 個の 3 ' 核酸構築物を含む。任意で、3 ' 核酸構築物の複数のコピーは固体表面、たとえば、ビーズ、マイクロビーズ、微粒子、ミクロスフェア、ナノビーズまたはヒドロゲルに結合される。任意で、各ビーズは同じ核酸構築物を含み、複数のビーズがバーコードのセットにて提供される。従って、任意で、各固体表面、たとえば、ビーズは同じ種類の 3 ' 核酸構築物の複数のコピーを含有する。従って、たとえば、バーコードのセットは、独特の 3 ' バーコードを含む第 1 の 3 ' 核酸構築物の複数のコピーを含む第 1 の表面と、第 1 の構築物上のバーコードとは異なる独特の 3 ' バーコードを含む第 2 の 3 ' 核酸構築物の複数のコピーを含む第 2 の表面とを含むことができる。

20

**【 0 0 5 9 】**

任意で、反応混合物を形成する工程は、懸濁緩衝液中の細胞を提供することと、単一細胞を含む反応混合物を形成するための条件下で 5 ' 核酸構築物及び / または 3 ' 核酸構築物を含むバーコード緩衝液に懸濁緩衝液を接触させることとを含む。反応混合物は各構築物が独特のバーコード有する少なくとも 2 つの構築物と単一細胞とを含む。任意で、反応混合物は少なくとも 1 つの 5 ' 核酸構築物と少なくとも 1 つの 3 ' 核酸構築物とを含む。任意で、反応混合物は少なくとも 2 つの 3 ' 核酸構築物を含む。任意で、反応混合物を形成する工程は、懸濁緩衝液中の細胞を提供することと、単一細胞及び少なくとも 1 つの 5 ' 核酸構築物及び少なくとも 1 つの 3 ' 核酸構築物を含む反応混合物を形成するための条件下で 5 ' 核酸構築物を含む第 1 のバーコーディング緩衝液及び 3 ' 核酸構築物を含む第 2 のバーコーディング緩衝液に懸濁緩衝液を接触させることとを含む。任意で、反応混合物は、単一細胞に由来する核酸試料と、5 ' 核酸構築物の複数コピーを含む少なくとも 1 つのビーズと、3 ' 核酸構築物の複数コピーを含む少なくとも 1 つのビーズとを含む。

30

40

**【 0 0 6 0 】**

溶解に先立って、浸透圧保護剤を含む細胞懸濁緩衝液に細胞を有利に浮遊させることができる。浸透圧保護剤は、細胞を浸透圧ストレスから保護することができ、バーコーディングに先立って、必ず、細胞生理が安定で、落ち着いたままであるようにすることができる。一部の実施形態では、細胞はバーコード構築物と共に細胞懸濁緩衝液に浮遊させる。一部の実施形態では、細胞は、逆転写、PCR 及び / または溶解のための試薬と接触させる前に細胞懸濁緩衝液に浮遊させる。細胞懸濁緩衝液はどんな反応容量にも含めることができ、水性反応容量を形成し、組み合わせるための本明細書に記載されている方法と適合性である。

**【 0 0 6 1 】**

50

細胞は、逆転写、PCR及び/または溶解のための試薬と接触させる前に細胞懸濁緩衝液に浮遊させる。細胞懸濁緩衝液はどんな反応容量にも含めることができ、水性反応容量を形成し、組み合わせるための本明細書に記載されている方法と適合性である。

#### 【0062】

任意で、各5'核酸構築物は固体表面に結合された複数コピーで提供される。任意で、各3'核酸構築物は固体表面に結合された複数コピーで提供される。本方法及び組成物で使用するすることができる固体表面を有する固形支持体の例には、ビーズ、クロマトグラフィの樹脂、マルチウェルプレート、微量遠心分離管、または固体表面を有する他の物体が挙げられる。任意で、固体表面はビーズ、マイクロビーズ、微粒子、ミクロスフェア、ナノ粒子、ナノビーズまたはヒドロゲルである。従って、所望のメカニズムまたは捕捉化学、たとえば、ピオチン・アビジン、ピオチン・ストレプトアビジン、または金・ジオールの相互作用を用いてバーコード構築物を固形支持体に結合することができる。任意で、バーコード構築物が結合される固体表面は水溶液に接触させられ、バーコード構築物はこの溶液中に遊離される。水溶液は、たとえば、バーコード構築物が付加される、単一細胞に由来する試料に関連するRNA分子と同じ反応容量であることができる。或いは、バーコード構築物用の固体表面に接触する水溶液は、標的のRNAまたは細胞とは異なる反応容量で保持することができ、バーコード構築物は、2つの反応容量を合わせて、バーコードのセットと単一細胞または単一細胞に由来するRNA試料とを含む反応混合物を生成する際に、これらのRNAまたは細胞に加えることができる。

#### 【0063】

提供されているのは、ポリヌクレオチドを固形支持体に結合する方法であり、該ポリヌクレオチドはバーコード配列を含有する。ポリヌクレオチドはバーコード構築物であることができる。任意で、方法には、逆エマルジョンの親水性区画（すなわち、水性液滴）を生成することが含まれる。区画は、所望のように、たとえば、疎水性キャリア流体に水溶液を混合すること、及び任意で混合物を攪拌することによって生成することができる。水溶液は、固形支持体、オリゴヌクレオチド、及びその中に懸濁される試薬を有することができるので、各区画は、区画が形成されるとポリヌクレオチドを固形支持体に結合するために必要な成分すべてを含有する。これらの実施形態では、区画に固形支持体を加えることに先立って、捕捉部分を介してオリゴヌクレオチドを固形支持体の表面に結合する。このオリゴヌクレオチドは本明細書では「結合されたオリゴヌクレオチド」と呼ばれ、バーコードオリゴヌクレオチドの3'配列に対して相補性の3'配列を含有する。従って、ポリヌクレオチドは、結合されたオリゴヌクレオチドとバーコードオリゴヌクレオチドが関与するポリメラーゼ伸長反応を介して固形支持体上で形成され、この反応は区画の中で行われる。

#### 【0064】

任意で、親水性区画が形成される場合、バーコードオリゴヌクレオチドは低い濃度または限界濃度（たとえば、区画当たり1モル）で存在する。無作為化配列を有するバーコードオリゴヌクレオチドのライブラリを用いて複数のバーコード構築物ビーズを調製する場合、この濃度は好都合である。バーコードオリゴヌクレオチドがどれも皆、異なるバーコード配列を有すると想定され、各区画の固形支持体がたった1つのバーコードを有することが所望であるならば、そのときは、1つのバーコードオリゴヌクレオチド（多くてもまたは平均で）が区画ごとに存在することができる。いったんこの条件が満たされると、複数の固形支持体（たとえば、複数のビーズ）が区画に存在することができ、または結合されたオリゴヌクレオチドの複数コピーが各固形支持体に結合することができるが、区画におけるポリメラーゼ伸長反応から生じるオリゴヌクレオチドすべてが同じバーコード配列を含有するであろう。

#### 【0065】

本方法で使用するための固形支持体には、ビーズ、たとえば、約0.1~10マイクロメートルの範囲での直径を有する金属及び/またはポリマー材料で作られる球形ビーズが含まれる。他の特徴を有するビーズを代わりにまたは追加で使用するすることができる。固形支

10

20

30

40

50



持体は捕捉部分で官能化されてオリゴヌクレオチドを表面に結合することができる。補足部分の例には、アビジン、ストレプトアビジン、ビオチン、カルボキシル基、エポキシ基、ヒドロキシル基、チオール基及び金が挙げられる。一部の補足部分はそれらが特異的に且つ非共有結合で結合する結合相手を有する。たとえば、ストレプトアビジンは結合相手としてビオチンを利用する。そのような捕捉部分は固形支持体に直接（たとえば、共有結合で）結合することができ、結合相手は結合されるオリゴヌクレオチドに結合することができ、逆もまた同様なので、結合されるオリゴヌクレオチドは非共有結合の相互作用を介して固形支持体に結合する。他の補足部分は結合されるオリゴヌクレオチドと固形支持体との間で直接的な共有結合を提供する。

#### 【 0 0 6 6 】

結合されるオリゴヌクレオチドは好ましくは、その 5' 末端で固形支持体に結合する一本鎖 DNA 分子である。従って、結合されるオリゴヌクレオチドの 3' 末端は溶液中で遊離であり、バーコードオリゴヌクレオチドとハイブリッド形成するとき、DNA ポリメラーゼのような酵素によって伸長することができる。伸長反応はバーコードオリゴヌクレオチドを用いて鋳型化されるので、バーコード配列はビーズに結合した DNA 鎖に組み込まれる。所望であれば、結合されるオリゴヌクレオチド及び / またはバーコードオリゴヌクレオチドは分子内二次構造を出来るだけ抑えるように設計された配列を有することができる。

#### 【 0 0 6 7 】

ポリヌクレオチドを固形支持体に結合するための本明細書に記載されている方法のいずれかを用いて、試料、細胞または RNA にバーコードを付けるのに使用するために 1 つ以上の固形支持体を調製することができる。各固形支持体に結合されるポリヌクレオチド（複数可）はバーコード配列を含み、バーコード構築物として役立つことができる。本方法を使用してバーコードライブラリを調製することもでき、それは、それぞれバーコード配列に関連する複数の固形支持体を含む。任意の 2 つの固形支持体（たとえば、ビーズ）は、全体的にまたは部分的に互いに異なるバーコード配列を有することができる。任意で、バーコードライブラリにおける固形支持体はどれも皆、異なるバーコード配列に関連付けられる。

#### 【 0 0 6 8 】

本方法に従って調製されるバーコード構築物ビーズはバーコード構築物に結合したビーズを含む。ビーズはバーコード構築物の複数のコピー、たとえば、少なくとも 10、30、100、300、1,000、3,000、10,000、30,000、100,000、300,000 または 1,000,000 のコピーに結合することができる。任意で、1 つのビーズに結合したバーコード構築物の各コピーは同じバーコード配列を含む。本方法は複数のバーコード構築物ビーズを含むビーズ状のバーコードライブラリの調製も可能にする。ライブラリにおけるビーズはどれも皆、異なるバーコード配列に関連付けることができ、各ビーズにおけるバーコード構築物のコピーは同じバーコード配列を含むことができる。

#### 【 0 0 6 9 】

任意で、本方法を用いて、バーコード構築物ビーズ上で 1 つ以上の試料（たとえば、細胞）から調製されたまたは得られた cDNA を物理的に捕捉することによってポリヌクレオチドのライブラリを調製することができる。各ビーズは 3' 末端にて cDNA 結合部位を伴ったバーコード構築物を含む。ビーズは酵素に接触させて結合部位を一本鎖にすることができる（たとえば、溶液中で遊離の鋳型分子の末端で 3' オーバーハングを脱離すること）。次いで、cDNA が結合部位を介して鋳型分子のコピーに結合するように試料に由来する 1 つ以上の cDNA にビーズを接触させる。好まれる実施形態では、結合部位は 1 つ以上の G ヌクレオチド、たとえば、ポリ - G トラクトを含み、逆転写酵素によって cDNA の末端に付加される鋳型化されないポリ - C トラクトに対して相補性である。

#### 【 0 0 7 0 】

ポリヌクレオチドのライブラリにおけるビーズを所望に応じて用いて、たとえば、複数の試料に由来する cDNA の配列を決定し、または様々な試料から cDNA を分離すること

10

20

30

40

50

ができる。後者の場合、様々な試料に対応するビーズを遠心分離または磁気作用を用いて沈殿させ、次いで常法を用いて再浮遊させ、分離することができる。所望であれば、ビーズ上の鋳型分子へのcDNAの結合に続いて、鋳型分子を酵素的に伸長させ、それによってビーズに結合したDNA二本鎖にcDNA配列を組み込み、これらの配列をバーコード配列に関連付けることができる。試料に由来するcDNA分子のコピーの数がビーズ上のバーコード構築物のコピーの数に匹敵するのであれば、その時は、これらのcDNA分子を少数のビーズ上で(たとえば、試料当たり多くても約1、3、10、30、100、300、または1000個のビーズ)捕捉することができる。常法を用いて、または上記で議論されたように、試料に由来するRNAを逆転写してcDNAを生成することができる。

#### 【0071】

核酸構築物が固体表面に結合されてバーコードを含む構築物を核酸試料の核酸に組み込ませるようにする場合、種々の方法を用いて5'核酸構築物及び3'核酸構築物を遊離することができる。たとえば、5'核酸構築物及び3'核酸構築物はそれぞれ光切断可能なリンカーまたは制限酵素部位を含むことができる。任意で、各5'核酸構築物はさらに光切断可能なリンカーを含み、各3'核酸構築物はさらに制限酵素部位を含む。任意で、各5'核酸構築物はさらに制限酵素部位を含み、各3'核酸構築物はさらに光切断可能なリンカーを含む。任意で、方法はさらに、反応混合物を制限酵素に接触させ、且つ反応混合物をUV光に曝露して固体表面から5'核酸構築物及び3'核酸構築物を遊離させることを含む。任意で、5'核酸構築物及び3'核酸構築物はさらに制限酵素部位を含み、方法はさらに、反応混合物を制限酵素に接触させて固体表面から5'核酸構築物及び3'核酸構築物を遊離させることを含む。任意で、5'核酸構築物及び3'核酸構築物はさらに光切断可能なリンカーを含み、方法はさらに、反応混合物をUV光に曝露して固体表面から5'核酸構築物及び3'核酸構築物を遊離させることを含む。

#### 【0072】

上記で言及されたように、任意で、3'核酸構築物は独特の3'バーコードと結合配列とバーコード関連配列とを含む。任意で、3'核酸構築物はヘアピンを形成することができる。任意で、3'バーコード配列及び結合配列は制限酵素配列または光切断可能なリンカーによってバーコード関連セグメントから分離される。任意で、制限酵素は固体表面から3'バーコード配列を遊離させ、バーコード関連セグメントから結合配列を分離する。任意で、UV光は固体表面から3'バーコード配列を遊離させ、バーコード関連セグメントから結合配列を分離する。任意で、3'核酸構築物上の結合配列はポリ-T配列を含む。任意で、3'核酸構築物上の結合配列は試料の核酸分子上のポリ-A配列に結合する。任意で、3'核酸構築物上のバーコード配列は逆転写によって組み込まれる。任意で、逆転写は3'バーコード配列を含む核酸分子を作り出す。任意で、バーコード関連セグメントはプライミング配列を含む。任意で、プライミング配列は回文配列である。任意で、一方のバーコード関連セグメントのプライミング配列は、核酸試料における他の任意のバーコード関連セグメントのプライミング配列と結合することができる。任意で、互いに接触しているバーコード関連セグメントはDNAポリメラーゼによって伸長されて双方のバーコード関連セグメントを単一の核酸に組み込む。

#### 【0073】

制限酵素及び制限酵素部位は既知であり、酵素は市販されている。制限酵素消化は、核酸分子における特定の位置のみで機能する酵素による核酸分子の触媒的切断を指す。そのような酵素の使用は当業者によって理解されている。従って、5'核酸構築物及び3'核酸構築物は、適当な酵素(たとえば、制限エンドヌクレアーゼ)への曝露の際、構築物の切断を促すエンドヌクレアーゼ制限部位を含有することができる。任意で、制限酵素は、N b . B s r D I、N b . B t s I、N t . B b v C I、N t . B s p Q I、N t . B s m A I、N t . B s t N B I、N t . A l w I、及びN t . B s m A Iから成る群から選択される。任意で、固体表面に結合された5'核酸構築物及び3'核酸構築物は二本鎖であり、構築物の一方の鎖のみが固体表面から遊離される。従って、任意で、5'核酸構築物及び/または3'核酸構築物はニッキングエンドヌクレアーゼ制限部位を含むことができ、1つ以

10

20

30

40

50

上のニッキングエンドヌクレアーゼ酵素を用いて鎖を遊離することができる。ニッキングエンドヌクレアーゼ制限部位は、Nb . BsrDI、Nb . BtsI、Nt . BbvCI、Nt . BspQI、Nt . BsmAI、Nt . BstNBI、Nt . AlwI、及びNt . BsmAIから成る群から選択される酵素に特異的な部位であることができる。任意で、鎖を置き換えるDNAポリメラーゼは制限酵素と共に含められ得る。鎖を置き換えるポリメラーゼはニッキングエンドヌクレアーゼ酵素の前に、と同時にまたはその後で添加することができる。鎖を置き換えるDNAポリメラーゼは、クレノウエキソ - 、Bst大断片及びBst大断片の操作された変異体から成る群から選択することができる。

#### 【0074】

光切断可能なリンカー及びその使用方法も既知である。本明細書で使用されるとき、用語「光切断可能なリンカー」は2つの分子、たとえば、2つの核酸または核酸と固体表面を結合するまたは操作可能に連結する任意の化学基を指す。任意で、リンカーは2 - ニトロベンジル部分を含む。たとえば、その全体が参照によって本明細書に組み入れられる米国特許第5,643,722号を参照のこと。光切断可能なリンカーには、アルファ - 置換した2 - ニトロベンジル部分[たとえば、1 - (2 - ニトロフェニル)エチル部分]、3,5 - ジメトキシベンジル部分、チオヒドロキサム酸、7 - ニトロインドリン部分、9 - フェニルキサンチル部分、ベンゾイン部分、ヒドロキシフェナシル部分、及びNHS - ASA部分が挙げられるが、これらに限定されない。リンカーは、適当な波長の光、たとえば、紫外線への曝露によって切断され、基質から核酸分子を回収することができる(J. Olejnik及びK. Rothschild, Methods Enzymol, 291: 135 - 154, 1998を参照のこと)。

#### 【0075】

任意で、提供される構築物は独特の分子識別子(UMI)配列を含むことができる。任意で、UMI配列は無作為化したヌクレオチドを含有し、バーコード配列とは無関係にバーコード構築物に組み込まれる。従って、同一バーコード配列を含有するバーコード構築物のセットは異なるUMI配列を含有することができる。同一バーコード配列を含有するが、異なるUMI配列を含有するバーコード構築物のセットが1つの試料に関連するRNAに加えられる場合、RNA配列はどれも皆バーコーディングの間に、異なるUMI配列に連結することができる。各ビーズ上の構築物が同一バーコード配列と異なるUMI配列のライブラリとを含有する場合、UMI配列を伴ったバーコード構築物ビーズを生成することができる。

#### 【0076】

バーコード構築物は、バーコード配列に加えて、汎用プライミング配列または汎用プライミング領域、及び結合部位を含むことができる。従って、任意で、3'核酸構築物及び5'核酸構築物は通常結合配列を含む。任意で、プライマー結合部位または汎用プライミング配列を含む5'核酸構築物及び3'核酸構築物。任意で、3'核酸構築物はポリ - T配列を含む結合配列を有する。ポリ - T配列は試料の核酸分子のポリ - A配列に結合する。3'核酸構築物の核酸分子への結合の際、そのとき、3'核酸構築物におけるバーコードは逆転写によって組み込むことができる。逆転写は3'バーコード配列を含む核酸分子を作り出す。従って、3'核酸構築物は逆転写のためのプライマーとして役立つことができる。ここで、構築物の結合部位はRNA結合部位(たとえば、mRNA結合部位)であり、1つ以上のRNAにおける配列領域に相補性の配列領域を含有する。任意で、結合部位は、バーコード構築物が加えられる試料におけるRNAすべてに共通する配列領域に対して相補性である。たとえば、結合部位はポリ - Tトラクトであることができ、それは真核細胞mRNAのポリ - A尾部に対して相補性である(図3)。代わりにまたは加えて、結合部位は無作為配列トラクトを含むことができる。試料に関連するRNAに3'構築物を加える際、逆転写が生じ、cDNAの第1の鎖が合成され得るので、3'バーコード配列はcDNAの第1の鎖に組み込まれる。逆転写は、適当な条件、たとえば、適当な緩衝液や逆転写酵素の存在、及びバーコード構築物のRNAへのアニーリングに適する温度及び酵素の活性を必要とすることが認識されるであろう。プライマーの3'末端が鋳型に相補性であり、鋳型に直接

10

20

30

40

50

アニーリングすることができる場合、DNAプライマーとRNA鋳型が関与する逆転写が最も効率的であることも認識されるであろう。従って、3'核酸構築物は結合部位がRNAの3'末端に存在するように設計することができる。

#### 【0077】

上記で議論したように、任意で、バーコードと結合配列とバーコード関連配列とを含む3'核酸構築物のセットが使用される。任意で、プライマー結合部位または汎用プライミング配列を含む3'核酸構築物。各構築物はヘアピン構造を形成する一本鎖DNAオリゴであることができる。任意で、構築物は線状である。ヘアピン用の構築物がヘアピンの二本鎖セグメントであるならば、双方の鎖の制限を指示する制限部位を組み込むことができる。これは、バーコードと同様に「バーコード関連配列」(BAS)も含有してもよい別個のループ状のセグメントからバーコード及び結合領域を切断し、分離する。たとえば、図8を参照のこと。線状のオリゴは、ビーズから切断される2つ以上のセグメントも生じる複数の制限部位と共に使用することができる。バーコード領域には、バーコードと、ポリ-Tセグメントと、任意でUMIとが含まれる。このセグメントは、液滴内の細胞によって遊離されるmRNA転写産物のポリ-A尾部にアニーリングし、そのmRNAの逆転写に刺激を与えるので、作り出されるcDNAにバーコードを組み込む。たとえば、図9を参照のこと。これとは別に、ループ状のセグメントはそれが切断されるバーコードと同一であるまたはそれに相当するそれ自体のバーコードを含む。複数のビーズが細胞と共に同時封入される場合、2つの転写産物バーコーディングセグメント(BC1a及びBC2a、図10)は双方ともRTバーコーディング反応のためのプライマーとして役立つであろうし、その細胞に由来するトランスクリプトームは双方のバーコードを組み込むであろう。これとは別に、そのそれぞれのバーコード(BC1b及びBC2b、図10)を運ぶ2つの対応するループ状セグメントも存在するであろう。BAS上の結合領域はBC1b及びBC2bループ状セグメントが交差アニーリングするのを可能にするであろうし、DNAポリメラーゼは反応ミックスに含まれて各セグメントを伸長し、他のバーコードの相補体を組み込むことができる(図11)。このようにBAS'は、その細胞に由来するトランスクリプトームを一体化するためのセットとしてバーコードを関連付けるために、その後下流で配列決定することができる単一アンプリコンにて物理的に連結される。BAS上の結合配列は回文配列であることができ、各BASは実施例でさらに記載されるように別のBAS上の配列に相補性である配列を含むことができる。

#### 【0078】

5'核酸構築物及び3'核酸構築物を使用する場合、いったん3'核酸構築物がcDNAの第1の鎖に組み込まれると、任意で、逆転写酵素は試料における核酸分子の末端に2~5個のシステイン残基を付加する。2~5個のシステイン残基は5'核酸構築物上の結合配列に相補性である結合配列を提供する。任意で、5'核酸構築物上の結合配列はポリ-G配列を含む。5'核酸構築物上の結合配列は次いで試料の核酸分子におけるポリ-C配列に結合することができる。5'核酸構築物は逆転写酵素を用いて核酸試料に組み込まれる。たとえば、図3を参照のこと。この過程は、その全体が参照によって本明細書に組み入れられるWO2012/148497に少なくとも記載されている。手短には、H-MMLV逆転写酵素は3'dCテイリング活性を有し、非鋳型化dCを第1の鎖cDNAに付加する。少なくとも1Gで終了するバーコード付き構築物、たとえば、5'核酸構築物も存在するのであれば、構築物は第1の鎖cDNAの3'dCと塩基対合することができ、逆転写酵素は鋳型切り替えを受け、転写を継続する。従って、逆転写酵素はホスホジエステル結合を介して第1の鎖cDNAの3'末端にバーコード配列を共有結合で付加する。得られる構築物は、たとえば、図7で示すように、5'から3'で5'バーコード cDNA配列 3'バーコードの配列を一般に含む核酸分子である。任意で、核酸試料はさらに増幅されて複数の核酸配列の複数のコピーを作り出し、各配列は5'バーコード及び3'バーコードを含む。任意で、5'核酸構築物及び3'核酸構築物は5'バーコード及び3'バーコードを含む核酸配列を増幅するのに使用される汎用プライミング部位を含む。

#### 【0079】

10

20

30

40

50

任意で、逆転写酵素はM - M L V 酵素である。任意で逆転写酵素はM M L V H - 逆転写酵素である。

#### 【 0 0 8 0 】

上記で言及されたように、溶解試薬が液滴に存在し、細胞からRNAを遊離すれば、且つ逆転写酵素とプライマーと他の適当な試薬が存在すれば、逆転写は起こり得る。溶解及び逆転写を促す酵素及び試薬は、たとえば、酵素と試薬を含有する液滴をバーコードと細胞を含有する液滴と合わせることによって一度にいっせいに加えることができ、または段階的に加えることができる。

#### 【 0 0 8 1 】

任意で、5' 核酸構築物は、プロモータを含む二本鎖核酸構築物であり、5' 核酸構築物は5' 核酸構築物の複数コピーを作り出す鋳型として役立つ。任意で、プロモータはRNAポリメラーゼのプロモータである。たとえば、図7を参照のこと。従って、任意で、5' 核酸は以下の配列：5' - T7プロモータ 汎用プライミング配列 バーコード配列 結合配列 - 3' を有する二本鎖DNA (dsDNA) 鋳型分子である。T7プロモータ配列はT7 RNAポリメラーゼによる鋳型からの5' 核酸構築物の合成を可能にする。汎用プライミング配列は下流で使用されるPCRプライマーに対する相補性のために使用される。結合配列は1つ以上のグアニン塩基 (G) から成り、第1の鎖cDNAの3' 末端へのバーコード付き構築物の相補性の塩基対合を可能にする。それぞれT3及びSP6のRNAポリメラーゼによるRNAバーコード付きの構築物の合成を可能にするT3及びSP6のプロモータ配列のような、しかし、これらに限定されない他のプロモータ配列を使用することができる。完全長またはほぼ完全長の5' 核酸構築物が大部分の場合に合成される限り、特異的なプロモータ配列を有さない他のRNAポリメラーゼも使用されてもよい。

#### 【 0 0 8 2 】

提供される核酸試料は単一細胞に由来し、試料に関連するRNAはmRNAを含むことができる。試料は、たとえば、少なくとも1、3、10、30、100、300、1,000、3,000、10,000、30,000、100,000、300,000、または1,000,000個のmRNA分子を含むことができ、それは遺伝子、対立遺伝子、リーディングフレームまたは違う配列の数を表すことができる。任意で、試料に関連するRNAは、試料に由来するmRNAすべて、細胞の完全なまたは部分的なトランスクリプトーム、または細胞に由来する全RNAを含む。

#### 【 0 0 8 3 】

任意の理論に束縛されないで、本方法は試料当たりバーコードを付けることができるRNAの数に制限を設けない。従って、試料当たり作り出される対象とするポリヌクレオチドの数は少なくとも10、30、100、300、1,000、3,000、10,000、30,000、100,000、300,000、または1,000,000個であることができる。対象とする各ポリヌクレオチドは複数コピーで存在することができる。さらに、方法の1回の実行にてバーコードを付けることができる細胞または試料の数は、バーコード配列の独特の組み合わせでバーコードのセットを調製する難題 (上記で議論された) によってのみ限定される。一部の実施形態では、1つ以上の試料は少なくとも10、30、100、300、1,000、3,000、10,000、30,000、100,000、300,000、または1,000,000個の細胞を含む。試料 (たとえば、それぞれ単一細胞である) は同一対象または異なる対象から得ることができる。たとえば、少なくとも2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90または100個の異なる対象が試料を提供することができる。

#### 【 0 0 8 4 】

核酸試料における5' バーコード及び3' バーコードの組み合わせは単一細胞に由来する核酸試料の起源を特定する。任意で、核酸試料におけるバーコードは配列決定によって特定される。各バーコード配列は、バーコード配列が独特のバーコードセットで再利用されるので、異なる試料に由来する集合コンティグに関連付けられるように期待される。用語コンティグは本明細書で使用されるとき、重複する配列データ (読み取り) を指す。同一の

コンティグがバーコード配列 a、b、c を使用しているのが観察されてもよい。そしてバーコード配列 a、b 及び d は別のコンティグに関連するのが観察されてもよい。これから、次いで、a、b 及び c はバーコードセット 1 を構成し、a、b 及び d はバーコードセット 2 を構成すると結論付けることができる。バーコードセット 1 は 1 つの細胞に由来する核酸試料に関連する一方でバーコードセット 2 は第 2 の異なる細胞に由来する核酸試料に関連する。セットにおける 5' バーコードすべてがセットにおける 3' バーコードすべてに関連付けられるので、それらは最大二部クリークとしてグラフ理論で表すことができる（図 13 を参照のこと）。最大二部クリークはバーコードセット及び核酸を単一細胞に関連付けるであろう。

【0085】

提供されるのはまた、本明細書で開示されている方法を用いて生成される組成物である。従って、本出願はその生成のための RNA 及び DNA の構築物の組成物を提供する。提供されるのはまた、とりわけ、バーコード構築物ビーズのライブラリ、RNA バーコード構築物と共に負荷されるエマルション液滴のライブラリ、細胞と共にバーコード構築物のライブラリを含有するエマルション、バーコード付き cDNA のライブラリ、及びマイクロ流体液滴を生成する装置である。

【0086】

提供されるのは、バーコード付きビーズの集団を含む組成物である。組成物は複数のビーズを含み、各ビーズはその表面に複数コピーの 5' 核酸構築物または 3' 核酸構築物を結合している。従って、提供されるのは、各ビーズがその表面に複数コピーの 5' 核酸構築物を結合している、複数のビーズを有する組成物である。提供されるのはまた、各ビーズがその表面に複数コピーの 3' 核酸構築物を結合している、複数のビーズを含む組成物である。任意で、組成物は、細胞を含む懸濁緩衝液に適宜加えると、5' 核酸構築物を含む少なくとも 1 つのビーズと 3' 核酸構築物を含む少なくとも 1 つのビーズと単一細胞とを含む反応混合物を生じるバーコーディングミックスである。好ましくは、反応混合物は少なくとも 3 つの異なる核酸構築物、たとえば、2 つの 5' 核酸構築物と 1 つの 3' 核酸構築物、または 2 つの 3' 核酸構築物と 1 つの 5' 核酸構築物を含む。

【0087】

本明細書でさらに開示されるのは、本明細書に記載されている提供されている構築物を含むキットである。キットは本明細書に記載されているバーコード構築物に結合された複数の固形支持体を含むことができる。任意で、キットは 1 つ以上のバーコード構築物のライブラリを含み、各ライブラリは同一バーコードを持つ複数のバーコード構築物を含む。たとえば、提供されるのは、1、2、3、4、5、6、7、8 以上の種類の 5' 核酸構築物及び/または 3' 核酸構築物を含む、5' 核酸構築物及び 3' 核酸構築物を含むバーコードビーズのライブラリを含むキットである。任意で、キットは本明細書に記載されているバーコーディングミックス及び/または懸濁緩衝液を含む。キットは使用のための指示書を含むことができる。キットは、好適な容器にて、本明細書で開示されているようなビーズ、構築物、またはライブラリ、1 つ以上の対照、及び当該技術で周知の種々の緩衝液、試薬、酵素及び他の標準成分を含んでもよい。従って、キットはさらに、細胞溶解、制限酵素消化及びバーコーディング反応を行うための酵素または他の試薬を含むことができる。

【0088】

容器は 1 つ以上のウェルを含むプレートにて少なくとも 1 つのウェルを含むことができる。容器には、その中で提供される試薬、核酸構築物、ビーズ、ライブラリまたは細胞が入れられてもよく、一部の例では好適に等分されてもよい少なくとも 1 つのバイアル、試験管、フラスコ、ビン、注射器、液滴または他の容器手段を挙げることができる。追加の成分が提供される場合、キットはこの成分が入れられてもよい追加の容器を含有することができる。キットはまた、試薬、核酸構築物、ビーズ、ライブラリまたは細胞を含有するための手段及び市販のために厳重な管理下にある他の試薬容器も含むことができる。そのような容器は、所望のバイアルが保持される射出成形または中空成形のプラスチック容器を含んでもよい。容器は使用及び/または警告のための指示書を伴ったラベル付けを含むこ

10

20

30

40

50

とができる。

【0089】

提供されるのはまた、その全体が参照によって本明細書に組み入れられるU.S.S.N. 14/586, 857に記載されたもののような反応容量を生成し、輸送するための装置である。これらの容量はマイクロ流体規模で存在することができ、キャリア流体から位相分離することができる。装置によって取り扱うことができる反応容量の例には逆エマルション（すなわち、水/油エマルション）における水性液滴が挙げられる。装置は、バーコード構築物、試料（たとえば、細胞）、及び/またはこれらの試料から得られるRNAが別々にまたは一緒に液滴に封入されるようにする。装置はまた試薬も液滴に導入されるようにするので、バーコード構築物分子を酵素的に生成することができ、個々の試料に由来するRNAにバーコードを付けることができる。

10

【0090】

装置は一般に3つのマイクロ流体経路を含み、それぞれ圧力源及び流れセンサーに繋がれる。マイクロ流体経路のための圧力源は経路を介して流体を運び、圧力源の下流に存在する流れセンサーを用いて経路を介する流れの速度を測定することができる。一部の実施形態では、第1の経路と第2の経路が第1の接合部で合流して複合経路を形成する。たとえば、図1を参照のこと。

【0091】

本明細書に記載されているような装置は、IDEX Corporation (Lake Forest, Illinois, U.S.A.) から入手可能な配管及び流体工学成分から、及びDolomite Microfluidics (Charlestown, Massachusetts, U.S.A.) から入手可能なマイクロ流体液滴チップを用いて組み立てることができる。マイクロ流体液滴チップの一部の特徴は、参照によって本明細書に組み入れられる米国特許第7,268,167号、同第7,375,140号、同第7,717,615号、同第7,772,287号、同第8,741,192号及び同第8,883,864号に記載されている。好適な圧力源には注射器ポンプ及び圧力ポンプが挙げられる。圧力ポンプはDolomite Microfluidicsから入手可能である。圧力源は独立して制御することができる。

20

【0092】

任意で、第1と第2のマイクロ流体経路が水溶液を輸送する。各経路は注入ポートと弁（たとえば、四方弁）を含んで、経路でインラインの注入ポートに導入される溶液を移動させる。任意で、水性のキャリア流体を保持するリザーバを各四方弁の上流に配置する。キャリア流体は下流に運ばれるので、または第1の接合部に向かって水溶液のプラグを下流に押すので、水性のキャリア流体は四方弁で水溶液と混ざり合うことができる。任意で、流れ抵抗器が各マイクロ流体経路に配置される。

30

【0093】

いったん水溶液が第1または第2のマイクロ流体経路に導入されると、それは、第1の接合部に向かう溶液の流れを測る試料ループを通過することができる。測ることは、所望に応じて、たとえば、試料ループに沿って配置される流体抵抗または弁を用いて達成することができる。任意で、1つの試料ループが第1及び第2のマイクロ流体経路のそれぞれに関連付けられ、試料ループは熱冷却ユニットに接触する。熱冷却ユニットが含まれて水溶液における酵素、核酸または他の生体成分の熱変性を防ぎ、または酵素反応のための最適な温度を定めることができる。第1及び第2のマイクロ流体経路のための試料ループに接触した熱冷却ユニットの部分は独立してまたは一緒に制御することができる。それが試料ループを通過する水溶液の温度を常温から逸脱させることができるという条件で、物質または装置を熱冷却ユニットとして使用することができる。好適な熱冷却装置の例はPeltier装置及びアイスピンである。

40

【0094】

任意で、第1のマイクロ流体経路を通して輸送される水溶液は細胞及びバーコード構築物ビーズを含有する。一部の実施形態では、第2のマイクロ流体経路を通して輸送される水

50

溶液は細胞溶解のための試薬及び対象とするポリヌクレオチドを作り出すための試薬（たとえば、バーコード構築物分子を生成するための酵素）を含有する。各マイクロ流体経路に関連付けられた注入ポート、弁及び／または試料ループはその経路を通過する水溶液の内容物を受け入れるように構成するまたは特注することができる。たとえば、第1のマイクロ流体経路に関連する試料ループは細胞及びビーズを受け入れるために大きな内径を有することができる。第1と第2のマイクロ流体経路の間で細胞、ビーズ及び試薬を割り当てるために多数の他の選択肢が存在するので、これらの成分のすべてが第1の接合部で混ぜ合わせられることが認識されるであろう。たとえば、第1のマイクロ流体経路を通過して細胞を輸送することができ、第2のマイクロ流体経路を通過してビーズを輸送することができる。各経路は所望のように、それが運ぶ水溶液の内容物を考慮して構成することができる。

10

#### 【0095】

組み合わせた経路の合流から生じる、液滴を含有する流体経路は試料経路を構成する。試料経路は、第2の接合部の下流に存在する試料回収容器に届けられる。試料回収容器では、液滴を熱循環に供することができる。液滴を壊して開け、バーコード付きの核酸を回収することができる。

#### 【0096】

操作中、装置を用いてバーコード構築物ビーズ及び細胞を水性のマイクロ流体液滴に封入するので、各液滴は平均でほぼ2つのビーズと1つの細胞を含有する。各液滴におけるビーズ及び細胞の数は、たとえば、装置に負荷される溶液におけるビーズまたは細胞の濃度を調整することによって、または3つのマイクロ流体経路における流速を調整することによって所望のように調整することができる。各液滴に含まれる試薬はバーコード構築物分子が液滴にてビーズ1つから酵素的に生成されるようにする。これらの試薬は細胞1つが溶解され、細胞に由来するRNAがバーコーディング反応を受けることができるようにする。従って、細胞に由来するRNAは液滴内でバーコードを付けることができ、これらのRNAに由来する（及びバーコード配列を含有する）核酸は、複数の細胞に由来する核酸が混合される場合、後で1つの細胞に辿り着くことができる。

20

#### 【0097】

開示されているのは、開示されている方法及び組成物で 사용할ことができる、それらと併せて使用することができる、それらを調製することで使用することができる材料、組成物及び成分、またはそれらの生成物である。これらの及び他の材料は本明細書で開示され、これらの材料の組み合わせ、サブセット、相互作用、群等が開示される場合、これらの化合物のそれぞれ種々の個々の及び総体の組み合わせ及び順列の特定の参照は明白に開示されなくてもよい一方で、それぞれは具体的に熟考され、本明細書に記載されていることが理解される。たとえば、方法が開示され、議論され、方法を含む多数の分子に対して行うことができる多数の修飾が議論されるのであれば、方法の組み合わせと順列のそれぞれ及びどれも皆、ならびに可能である修飾は、特に反対に指示されない限り、具体的に熟考される。同様に、これらの任意のサブセットまたは組み合わせも具体的に熟考され、開示される。この概念は、開示されている組成物を用いた方法における工程を含むが、これらに限定されない本開示の態様すべてに適用される。従って、実施することができる種々の追加の工程があるならば、これらの追加の工程のそれぞれは開示されている方法の具体的な方法工程または方法工程の組み合わせで実施することができ、そのような組み合わせまたは組み合わせのサブセットのそれぞれは具体的に熟考されており、開示されると見なされるべきであることが理解される。

30

40

#### 【0098】

本明細書で引用されている出版物及びそれらが引用されている材料はその全体が参照によって本明細書に具体的に組み入れられる。

#### 【0099】

多数の実施形態が記載されている。それにもかかわらず、種々の改変が行われてもよいことが理解されるであろう。従って、他の実施形態は以下のクレームの範囲の中にある。

50



## 【実施例】

## 【0100】

実施例1．単一細胞に関連する核酸の解析のための核酸バーコードのセット

バーコード付きビーズライブラリの2つの違うセット（5'セット及び3'セット）を用いて各細胞のトランスクリプトームにバーコードを付ける。図2を参照のこと。各3'バーコードは制限部位と独特の分子識別子とバーコード配列とmRNA転写産物のポリ-A尾部を標的とするポリ-T配列（従って各真核細胞にて全RNAのかなり大半を構成するリボソームRNA及びtRNAを回避する）とから成る。各5'バーコードは3'ビーズセットの制限部位と同じであってもよい制限部位（RE）と最終的なバーコード配列の相補体と3つのグアニン残基とを含む。対象とする細胞と共にこれら2つのビーズのセットを液滴に封入する。図1にて示すもののような液滴装置を使用することができる。

10

## 【0101】

所望の比率の液滴が各セットのビーズを少なくとも1つ受け取るように各ビーズのセット（3'及び5'）の濃度を調整する。たとえば、液滴当たり各種の3'のビーズの平均値については、90%を超える液滴が各種のビーズを少なくとも1つ受け取るであろう。処理能力と単一細胞の封入との間の最適な均衡について同様に細胞濃度を調整する。この実施例の目的で、平均値は液滴当たり0.1個の細胞であり、その際、細胞を受け取る液滴の95%を超える事例が正確に1個の細胞を受け取るであろう。生産的な液滴は1個の細胞と5'及び3'のセットのそれぞれから少なくとも1つのビーズを受け取るであろう。成功した封入事象ののち、反応は図3で説明されたように進むことができる。

20

## 【0102】

mRNAは完全に逆転写され、cDNAの各末端でのバーコード配列の組込みが生じる。各種のビーズ正確に1個が1つの細胞と共に同時封入される場合、完全に逆転写されているmRNA転写産物は、その細胞にとって独特であるバーコードのセットを含む5'バーコード及び3'バーコードの同一対を示すであろう。一方または双方のビーズの複数が1つの細胞と共に同時封入される場合、5'/3'の組み合わせのどれも皆、生産されるcDNAライブラリにて示されると予想され、5'/3'の組み合わせのこのセットがその細胞に独特であるバーコードのセットである。cDNAのどれも皆少なくとも3'バーコードを運ぶであろう。逆転写が不完全である場合、5'バーコードはcDNAに組み込まれなくてもよい。3'バーコードのみが組み込まれるような場合でさえ、バーコードがその液滴にて十分な数の5'バーコードと関連してそのバーコードのセットの一部としてそれを構築している限り、不完全なcDNAは下流の解析ではその適正なトランスクリプトームに依然として潜在的に割り当てられ得る。

30

## 【0103】

図4は、各バーコードセットの複数のバーコード付きビーズが1つの細胞と共に単一の液滴に同時封入される場合を説明している。各細胞内での多数のmRNA転写産物のせいで、上述のバーコーディングのスキームによって、3'バーコードはすべて液滴にて5'バーコードすべてと一致する機会を有するであろう。これらのバーコードは細胞を起源とする転写産物すべてを特定する独特のバーコードセットを形成する。

## 【0104】

多くの場合、コーディング反応は完了に達することができなくてもよく、転写産物配列の一部のみが転写されるであろう。不完全ではあるけれども、これはとにかく潜在的に有用な情報である。図5は、5'バーコードを組み込むことができない非完全長のcDNAの場合でさえ、これらのcDNAは依然として適正なバーコードセットに割り当てられ得ることを説明している。従って、1つのバーコードしか持たない核酸を高度の信頼で正しいトランスクリプトームに割り当てることができる（逆もまた同様に同じことが言える）。これについての要件は、3'バーコードのコード空間（独特の3'バーコードの総数）が十分に大きいので任意の1つの3'バーコードが複数の細胞と共に同時封入される可能性が十分に低いことである。それが生じる場合でさえ、この重複は検出することができ、その特定の共有された3'バーコードを持つ非完全長のcDNAを安全に捨てることができる。対照

40

50

的に、単一のバーコードシステムの場合、同一バーコードを受け取っている2つの細胞のプロファイルは解析の間に統合されることになる。

【0105】

図6は同じ個々のバーコードの若干の重複がある場合でさえ、バーコードのセットは独特のバーコード1つの包含によって区別することができることを明らかにしている。

【0106】

5'及び3'のセットのそれぞれについての必要とされるコード空間(独特のバーコードの数)が決定されなければならない。バーコードは十分に独特なので配列決定誤差の存在でそれらを区別することができなければならない、すなわち、少数の呼び間違えた塩基の事象にて1つのバーコードが別のバーコードに変換されるべきではない。バーコードのセットの確立は、配列決定誤差がバーコード間の曖昧さを作り出し得る事例を解決するのに役立つであろう。

10

【0107】

代替の一実施形態は、それには鋳型切り替えが関与するという点で上述のバーコーディングのスキームに類似するが、この場合、5'バーコードビーズセットは図7で示されるように異なる組成を有する。このことは、mRNA転写産物から完全に逆転写されているcDNAの各末端にバーコードを組み込む「鋳型切り替え」のバーコーディングスキームには複数のアプローチがあることを実証している。

【0108】

実施例2. 単一細胞に関連する核酸の解析のための核酸バーコードのセット 非鋳型切り替え

20

この実施例は非鋳型切り替え法を記載している。バーコード付きビーズの単一セットしかない。各ビーズはヘアピン構造を形成する一本鎖DNAオリゴを有する。ヘアピンの二本鎖セグメントでは、双方の鎖の制限を指示する制限部位が組み込まれる。これは、転写産物のバーコーディングセグメントと、バーコードと同様に「バーコード関連配列」(BAS)も含有する別のループ状セグメントとの双方を切断する。図8を参照のこと。

【0109】

転写産物のバーコードセグメントにはUMIとバーコードとポリ-Tセグメントとが含まれる。このセグメントは、液滴内で細胞によって遊離されるmRNA転写産物のポリ-A尾部にアニーリングし、そのmRNA配列の逆転写に刺激を与えるので、作り出されるcDNAにバーコードを組み込む(図9)。

30

【0110】

それとは別に、ループ状セグメントはそれが切断されるバーコードと同一であるまたはそれに相当するそれ自体のバーコードを含む。それはまたプライミング配列、回文配列であるBASも含む。

【0111】

複数のビーズが細胞と共に同時封入される事象では、2つの転写産物バーコーディングセグメント(BC1a及びBC2a、図10)が双方ともRTバーコーディング反応のプライマーとして役立ち、その細胞に由来するトランスクリプトームは双方のバーコードを組み込むであろう。それとは別に、各バーコード(BC1b及びBC2b、図10)を運ぶ2つの対応するループ状セグメントも存在するであろう。

40

【0112】

回文状のBASはBC1b及びBC2bループ状セグメントが交差アニーリングするのを可能にし、DNAポリメラーゼが反応ミックスに含まれて各セグメントを伸長し、他のバーコードの相補体を組み込む(図11)。このように、BASは、バーコードをセットとして関連付けてその細胞に由来するトランスクリプトームを一体化するために、その後下流で配列決定することができる単一のアンプリコンに物理的に連結される。

【0113】

回文配列が望ましくなければ、BASが2つの相補性の配列を含み、各ビーズが大まかに同等の比率で双方のBAS配列を運ぶシステムも設計されてもよい。「ループ状」セグメ

50

ントとして記載されているものはループ状ではなく線状であってもよい。これは使用される制限酵素に左右される。同様に、ヘアピンを使用するのではなく、ビーズから切断される2つ以上のセグメントも生じる複数の制限部位を持つ線状オリゴを使用することができる。ここで再び、「ループ状」セグメントはループ状でなくてもよいが、代わりに線状である。

#### 【 0 1 1 4 】

実施例3．バーコードがどの容器に位置するのかという演繹的知識を持たずに複数のバーコードを用いてどの核酸が同一試料に関連するのかを特定すること

細胞にバーコードを付けるためのP B M Cの調製。凍結保存したP B M Cを解凍し、リン酸緩衝化生理食塩水（C o r n i n g , M a n a s s a s , V A）で細胞ペレットを3回洗淨し、96ウェルプレートに一回F A C S選別した、または25 mMのN a C lと153 mMのベタインを含む等張細胞懸濁緩衝液（C S B）に再浮遊させた。次いで細胞を数え、逆転写のためにウェル当たり平均1個の細胞にC S Bで希釈した。

#### 【 0 1 1 5 】

オリゴプレートの調製。バーコード付きオリゴD Tオリゴ（3'バーコード、以後、「o D T B C」と呼ぶ）及びバーコード付き鋳型切り替えオリゴ（5'バーコード。以後「w I D B C」と呼ぶ）はI n t e g r a t e d D N A T e c h n o l o g i e s（C o r a l v i l l e , I A）から取り寄せた。到着の際、o D T B C及びw I D B Cは双方とも10 μ Mに希釈し、使用まで-20℃で保った。それらは以下の配列を有する。

#### 【 0 1 1 6 】

（表2）オリゴの配列

名称	配列
oDT BC	GGC TCG GAG ATG TGT ATA AGA GAC AG [7量体のバーコード] NNN NNN NNT TTT TTT TTT TTT TTT V (SEQ ID NO:1)
wID BC	GGA AGA TAG GGA TAA CAG GGT AAT G [7量体のバーコード] ACG GG (SEQ ID NO:2)
pltwll_LibPCR_Fw	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CGT CGG CAG CGG AAG ATA GGG ATA ACA GGG TAA TG (SEQ ID NO:3)
pltwll_LibPCR_RV	CAA GCA GAA GAC GGC ATA CGA GAT [インデックス配列] GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG A (SEQ ID NO:4)
pltwll_PCR2_Fw	AATGATACGGCGACCAC (SEQ ID NO:5)
pltwll_PCR2_Rv	CAAGCAGAAGACGGCATAC (SEQ ID NO:6)

#### 【 0 1 1 7 】

バーコードは、それらが互いから少なくとも2編集距離離れるように、すなわち、1つのバーコードをもう1つのバーコードに変換するのに少なくとも2つの変異（置換、挿入または欠失）を要するように設計された。割り当ての誤りのせいで別のバーコードとして間違っ

#### 【0118】

逆転写及びバーコーディング。2のラムダを持つ無作為ポアソン分布を介して各ウェルに分配されたoDT BC及びwID BCを含有する96ウェルプレート

10

#### 【0119】

試料をプールすること、ライブラリの調製及び配列決定。試料をプールするために凍結したプレートを氷上で解凍した。各プレートのRT産物を2等分（96ウェルプレートの左半分と右半分）してプールし、製造元の指示書に従ってSelect-a-Size DNA Clean & Concentrator (Zymo Research, Irvine, CA)を用いて濃縮し、300bpより短い物質を除いた。手短には、500µLのselect-a-size DNA結合緩衝液を100µLのcDNAに加え、穏やかにピペッティングすることによって混合し、その後、10,000gで30秒間遠心分離した。cDNAを含有しているカラムをDNA洗浄緩衝液で2回洗浄し、DNA溶出緩衝液を用いて15µLで溶出した。Q5 DNAポリメラーゼ (New England Biosciences) と0.2µMのプライマーpltwll\_LibPCR\_Fw及びpltwll\_LibPCR\_RVとを用いたPCRに1µLの各試料を使用した。Eppendorf Mastercycler Nexus熱循環器 (Hauppauge, NY)を用い、以下のPCR熱循環条件：95で5分間の変性；98で30秒間、60で30秒間、及び68で1分間を10サイクル；ならびに68で5分間の最終伸長を用いた。以前記載されたようにselect-a-Size DNA Clean & Concentrator (Zymo Research, Irvine, CA)を用いてPCR産物を再び濃縮して300bpより短い産物を取り除いた。次いで0.2µMのプライマーpltwll\_PCR2\_Fw及びpltwll\_PCR2\_RVを用いて2回目のPCRを行った。Pippin Prep DNAサイズ選択システム (Beverly, MA)を用いて300bp~1200bpの産物をサイズ選択し、Illumina MiSeq system (San Diego, CA)によって配列決定した。

20

30

40

#### 【0120】

解析及び結果。ライブラリのMiSeq配列決定に続いて、48ウェルのセットから得られた読み取り対をPythonスクリプトを用いて解析した。ポアソン分布に従ったバーコードの無作為分布の結果、これらのウェルのうち39個が、少なくとも1つのwIDと少なくとも1つのwIDとを含有した。9個のウェルは、wIDバーコードもoDTバーコードも含有しなかった。読み取りの既知の構造（5'-wID-ACGGG-mRNA-BdA16-N8-oDT）を用いて、oDT BCを逆行読み取りから特定し、wID BCを順行読み取りから特定した。後に続く解析でバーコードの既知の配列を使用しなかったということは、この方法が、バーコードがどの試料に関連するかという以前の知識を持たないで核酸の起源の試料を決定することができ、データ解析のやり方がシークエンサ

50

ーによって導入される誤差に対して安定であることを実証している。

#### 【0121】

「二重バーコード付き転写産物の読み取り」のセットはoDT BC及びwID BCの双方が特定される読み取り対から生成された。セットは、6,964個の違うバーコード対を含む合計107,795個の読み取り対から成った。二重バーコード付き転写産物から観察されたバーコード対の頻度のヒストグラムを図15に示す。

#### 【0122】

低レベルの交差混入がオリゴの合成から予想された (IDT DNA Technologies, Reduced Barcode Contamination Using Oligonucleotides With TruGrade (商標) Processing, Decoded, 2(4), 2012を参照のこと)。バーコードのオリゴはセットと一緒に合成した。一緒に合成されるN個のバーコードと共に、供給される各オリゴが他のN-1個のオリゴのそれぞれの等量混合物を含有するのであれば、各所望の産物オリゴが作り出されるオリゴの画分1-pのみを含むように、混入から生じる対に対比させた予想されるバーコード対を有するであろう読み取りの画分を推定することができる。予想される対を伴う読み取りの画分は $(1-p)^* (1-p)$ として推定することができる。混入である画分は $(1 - (1-p)^* (1-p))$ として推定することができる。混入のせいで(N-1)個の異なるオリゴバーコードがあるならば、合計 $N * N - 1$ 個の混入物対がある。次いで混入物対の比率は $f = (1 - p * p) / \{ p * p * (N * N - 1) \}$ として推定することができ、読み取りの所与の数Rについての混入物当たりの読み取り対の数は $f * R$ である。

#### 【0123】

24のプレートで合計288個のバーコードを一度に整えた。モルレベルで10%までのオリゴの交差混入を控えめに可能にして、40未満の違う読み取り対のバーコード対が混入であるらしいことが推定された。従って、我々の解析でのさらなる検討のために最低50の読み取り対でバーコードの所与の対が観察されることを要する閾値が設定された。

#### 【0124】

データのウェル及び細胞への割り当て。バーコードが各ウェルに分配された以前の知識を使用しないで出力読み取りを単一の細胞及びウェルに割り当てた。この解析の出発点は、少なくとも50の読み取り対によってデータにて支持されたwID BCとoDT BCの183対のセットだった。BCのこれら183対は合計94,712の読み取り対の合計によって支持された。これらの対の1つ1つは、その頂点が観察されたwID BC及びoDT BCであり、少なくとも50の読み取りで辺に割り当てられたグラフにて辺(wID BC、oDT BC)と見なされた。我々のデータセットを表すこのグラフは、ノードの一方の群を構成するwID BCとノードの他方の群を構成するoDT BCを伴った二部であった。1つ以上のwID BC及び1つ以上のoDT BCを有する細胞/ウェルはデータにおける辺すべてから成るグラフの完全な二部部分グラフを形成する辺のセットを生成すべきであることが特徴とされた(図13及び14も参照のこと)。従って、データセットすべてのグラフを完全な二部グラフに分解するアルゴリズムを開発することによってグラフの辺を細胞/ウェルに割り当てた。

アルゴリズムの工程は以下のとおりだった。

1. グラフの連結成分すべてをコンピュータで計算する。
2. 完全な二部グラフ(2クリーク)である連結成分をそれぞれウェル/細胞に割り当てた。
3. 他の連結成分を以下のように2クリークの和集合に分解した。
  - a. 成分の部分グラフである2クリークすべてを決定した(以下で説明されるアルゴリズム)
  - b. 3. a.におけるグラフの設定から、最大である(他の2クリークに含まれない)2クリークを選択した。
  - c. 「切り詰め(pruning)」工程を適用してデータに存在しそうにない2クリー

10

20

30

40

50

クを除いた。

上記アルゴリズムの工程 3 a . は以下のように行った。

1 . 以下のように集合 O 及び I が生成された。

a . 連結成分における各  $w I D \quad B C$  については、辺によってそれに接続され、それを退出 ( $w I D \quad B C$ ) と呼ぶ  $o D T \quad B C$  の集合を決定し、次いでそれを O 退出 ( $w I D \quad B C$ ) とそのサブセットのすべての双方に加えた。

b . a . と同様の方法で、集合 I を以下のように生成した：連結成分における各  $o D T \quad B C$  については、辺によってそれに接続され、それを進入 ( $o D T \quad B C$ ) と呼ぶ  $w I D \quad B C$  の集合を決定し、次いでそれを I 進入 ( $o D T \quad B C$ ) とそのサブセットのすべての双方に加えた。

10

2 . I における各集合 i 及び O における各集合 o について今や、i における  $w I D \quad B C$  と o における  $o D T \quad B C$  との間の辺すべてがデータに存在するかどうか、及びこれは、2 クリークがその成分の部分グラフである 2 クリークすべての集合に加えられる場合であるかどうかを決定した。

上記アルゴリズムの工程 3 c . は以下のように行った。

1 .  $= 2$  のみで細胞の割合 0 . 0 1 6 6 は所与の種類の 5 を超えるバーコードを有するであろう。従って、5 を超える  $w I D \quad B C$  または  $o D T \quad B C$  を有するアルゴリズムによって構築された 2 クリークは有効な構築物として拒絶され、フィルターで除去された。

2 . 単一の  $w I D \quad B C$  または 4 以上の  $o D T \quad B C$  を有するクリーク (または逆もまた同様) はさらに小さい 2 クリークで構成される曖昧な構築物であることの高い尤度を有した。

20

この曖昧さを考慮して、これらの 2 クリークは同様にフィルターで除去された。  
工程 3 b . 及び 3 c . の経験則は、各成分をそれが構成する最もありそうな二部クリークに分解する最大尤度法によって置き換えられてもよく、これは結果をさらに改善すると予想される。尤度は、我々の経験則 3 c . で議論したようなパラメータ に依存し、且つ成分構成する種々の 2 クリーク間での  $w I D \quad B C$  及び  $o D T \quad B C$  における重複の量にも依存する。

【 0 1 2 5 】

( 表 3 ) 解析の結果

30

40

50

成分	成分の サイズ	w B C の数	o d t B C の数	切り詰め 最大 2クリーク の数	完全に 再構築 された 2クリーク の数	部分的に 再構築 された 2クリーク の数	1ウェル を交差する 2クリーク の数	2ウェル を交差する 2クリーク の数
1	48	25	23	12	6	3	0	3
2	20	12	8	3	3	0	0	0
3	15	8	7	4	2	1	0	1
4	11	4	7	2	2	0	0	0
5	10	2	8	2	1	0	0	1
6	8	2	6	1	1	0	0	0
7	6	2	4	1	1	0	0	0
8	5	2	3	1	1	0	0	0
9	4	3	1	1	0	0	1	0
10	4	2	2	1	1	0	0	0
11	4	2	2	2	1	0	1	0
12	4	3	1	1	1	0	0	0
13	2	1	1	1	1	0	0	0
14	2	1	1	1	1	0	0	0
15	2	1	1	1	1	0	0	0

## 【 0 1 2 6 】

表 2 は、このデータセットに寄与した 3 9 ウェルに由来するデータが 1 5 の成分を生成したことを示す（ 4 8 ウェルのうち 3 9 だけが w I D B C 及び o D T B C 双方を有したことに留意のこと。 9 ウェルは w I D パーコードのみまたは o D T パーコードのみを有したのでパーコード対を得ることができなかった）。 w I D B C 及び o D T B C の数ならびに分解アルゴリズムから生じた切り詰め最大 2 クリークの数を表示する。加えて、表 2 は、予想クリークに正確に一致した、予想された 2 クリークの適正なサブセットだった、または 1 もしくは 2 の予想されたグラフと交差した 2 クリークの数を表示する。

## 【 0 1 2 7 】

10

20

30

40

50

15の連結成分から34の完全な二部グラフが生成された。これら34のうち、23は予想されたグラフと完全な一致だった。さらに4つのグラフは予想されたグラフの適正なサブセットだった。従って、得られたクリークの $23/34 = 67\%$ は完全に生成され、 $27/34 = 79\%$ は少なくとも部分的に再構築され、それらに割り当てられた間違っただけを有さなかった。全体で、実験に含まれたウェルのうち $27/48 = 56\%$ は上手く割り当てられた関連バーコードとmRNAだった。バーコードオリゴの製造の間での、データ処理にてフィルターで除去され得る交差混入のせいで誤った読み取り対のデータセットが存在するにもかかわらず、これらの思い通りの再構築が可能だった。

#### 【0128】

この多重バーコーディングスキームは、目標が単一バーコードを各細胞に割り当てることである単一バーコーディングスキームよりも大きく優れている。ポアソン負荷による単一バーコーディングスキームでは、 $\lambda = 0.1$ 以下を使用することが自然の選択となる( $0.06$ のラムダを使用するMacosko, et al., Cell, 161:5(2015)を参照のこと)。 $\lambda = 0.1$ では、細胞のおよそ $90.4\%$ がバーコードを受け取らず、およそ $9.0\%$ が正確に1つのバーコードを受け取り、およそ $0.5\%$ が1を超えるバーコードを受け取ることになる。ここで記載されている実験では、 $5.8$ 倍に等しい $56\%$ ( $100 \sim 90.4\%$ )もの多くの細胞が単一バーコーディングスキームを用いて上手く解析されている。

10

20

30

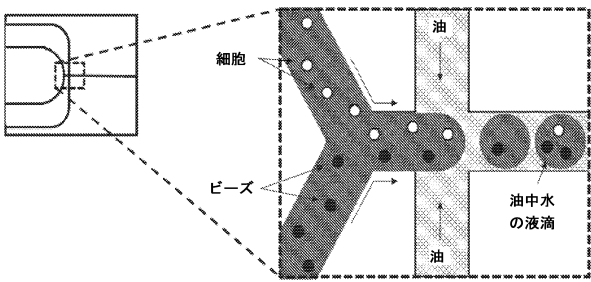
40

50

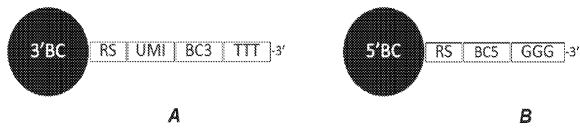


【図面】

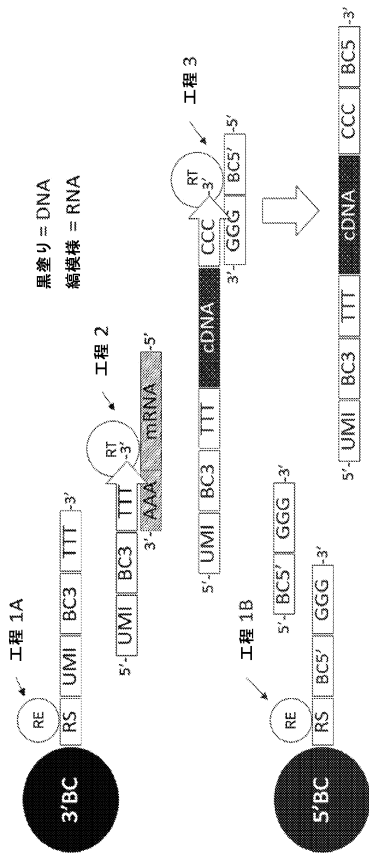
【図 1】



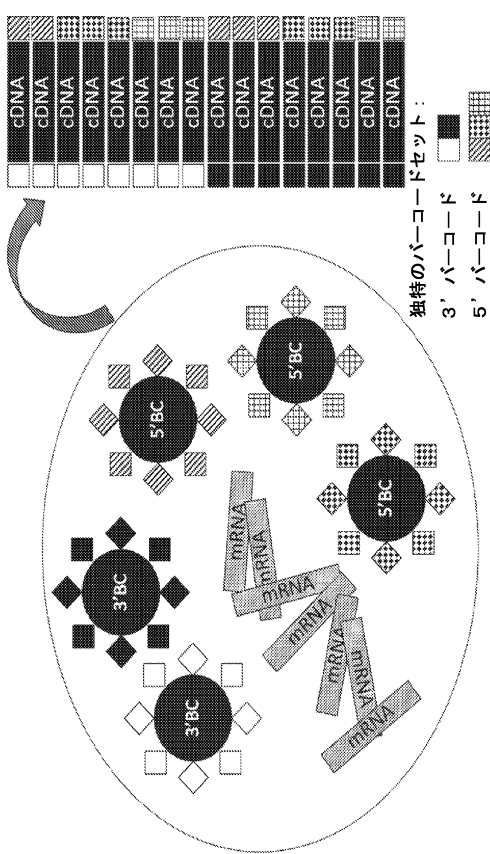
【図 2】



【図 3】



【図 4】



10

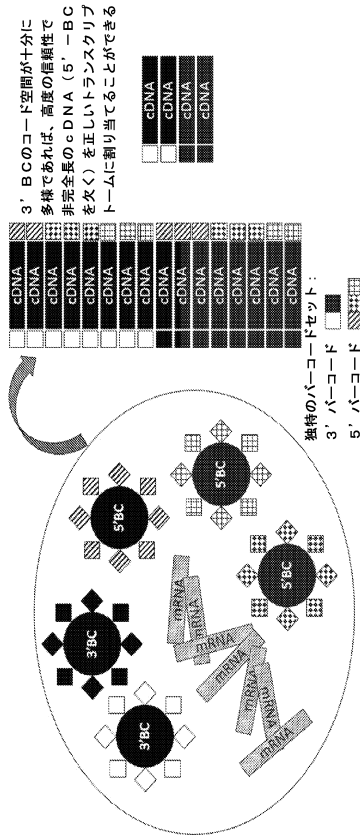
20

30

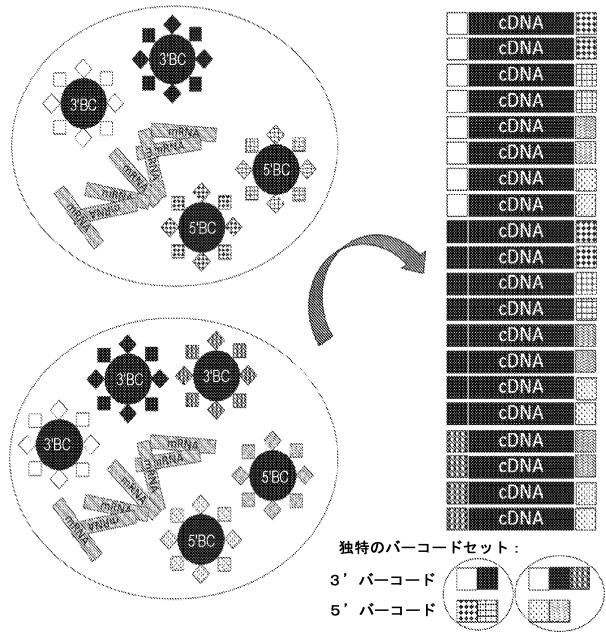
40

50

【図 5】



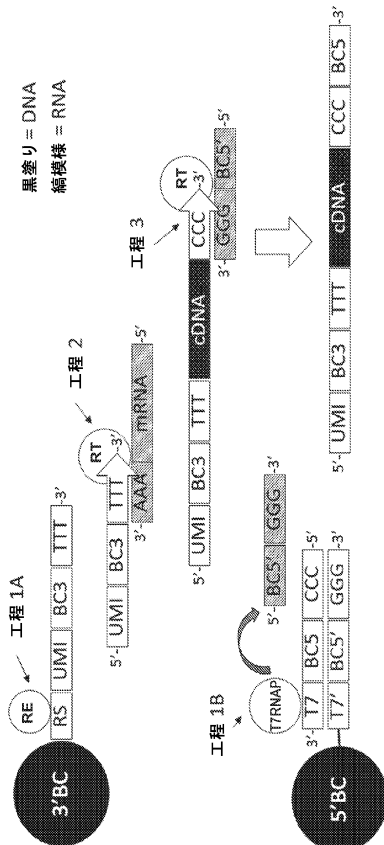
【図 6】



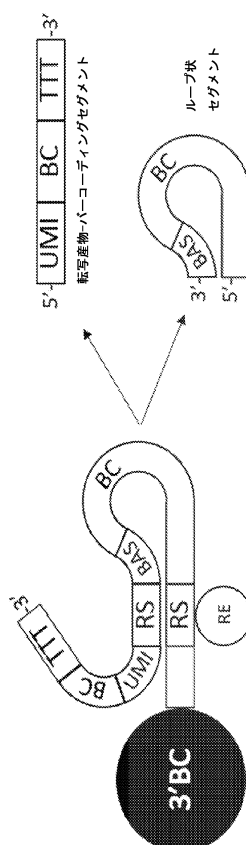
10

20

【図 7】



【図 8】



30

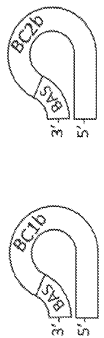
40

50

【 図 9 】

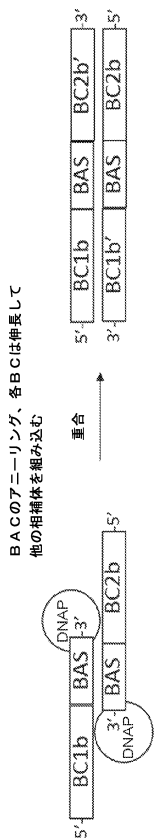


【 図 10 】

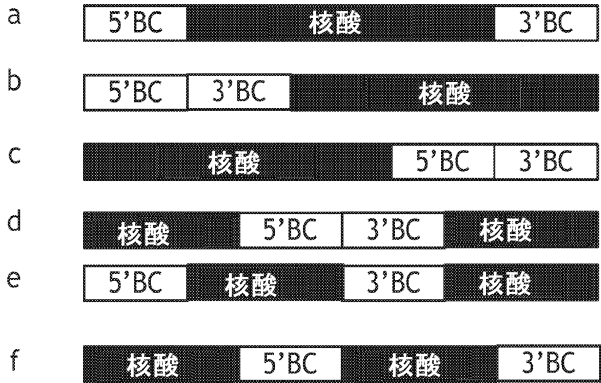


10

【 図 11 】



【 図 12 】

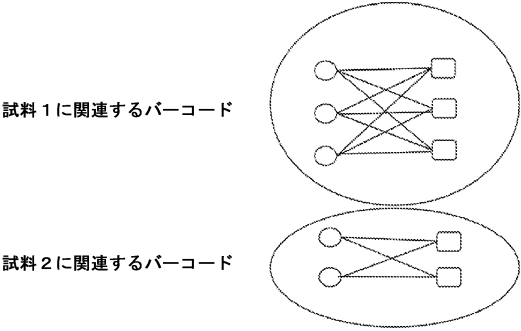


30

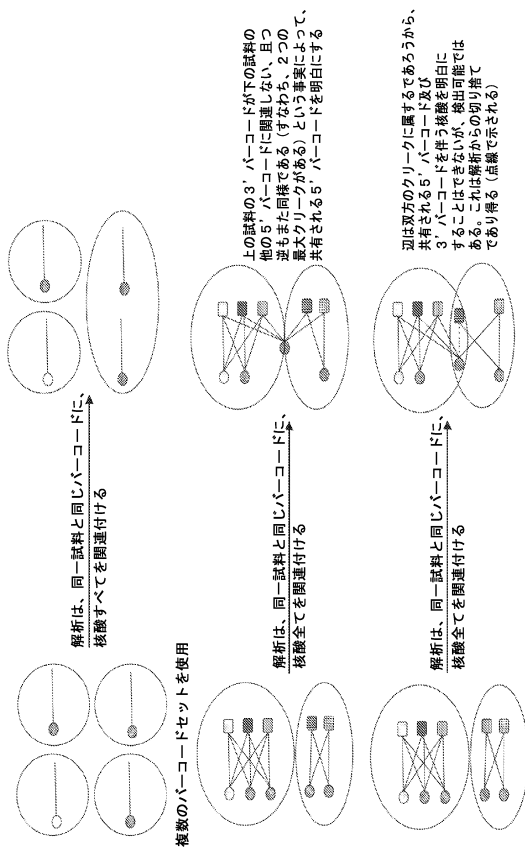
40

50

【図 13】



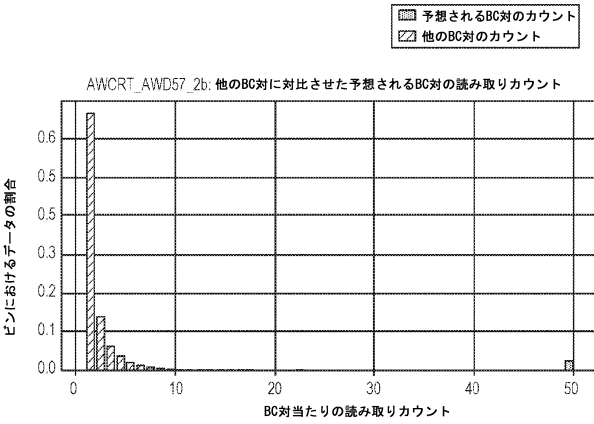
【図 14】



10

20

【図 15】



30

【配列表】

0007064439000001.app

40

50

## フロントページの続き

- (74)代理人 100148699  
弁理士 佐藤 利光
- (74)代理人 100128048  
弁理士 新見 浩一
- (74)代理人 100129506  
弁理士 小林 智彦
- (74)代理人 100205707  
弁理士 小寺 秀紀
- (74)代理人 100114340  
弁理士 大関 雅人
- (74)代理人 100121072  
弁理士 川本 和弥
- (72)発明者 ウィジー ギャリー  
アメリカ合衆国 9 4 1 1 0 カリフォルニア州 サンフランシスコ レキシントン ストリート 1 3 5
- 審査官 池上 京子
- (56)参考文献 国際公開第 2 0 1 6 / 1 3 8 4 9 0 ( W O , A 1 )  
国際公開第 2 0 1 5 / 1 0 3 3 3 9 ( W O , A 1 )
- (58)調査した分野 (Int.Cl. , D B 名)  
C 1 2 Q 1 / 0 0 - 3 / 1 0  
J S T P l u s / J M E D P l u s / J S T 7 5 8 0 ( J D r e a m I I I )  
C A p l u s / M E D L I N E / E M B A S E / B I O S I S ( S T N )