オシドの１つ以上の可能な正体が制限される。
【0213】
ヌクレオシドＮは、好ましくは、独立して選択され、Ａ、Ｇ、ＣまたはＴ（あるいは、任意選択で縮重低減性ヌクレオシド）であり得る。好ましくは、ヌクレオシドＮの配列は、１つ以上のＮが縮重低減性ヌクレオシドであり得る以外は、長さｋの可能な全ての配列を表し、したがって、プローブは、２つの部分を含有し、そのうち、ヌクレオシドＮからなる部分は非拘束部分とよばれ、ヌクレオシドＸからなる部分は、拘束部分とよばれる。上記のように、該部分は、連続的なヌクレオシドである必要はない。拘束部分および非拘束部分を含有するプローブは、本明細書において部分拘束プローブという。好ましくは、拘束部分内の１個以上のヌクレオシドは、プローブの近位端、すなわち、伸長可能なプローブ末端にライゲートされるヌクレオシドを含有する末端（これは、種々の本発明の実施形態において、オリゴヌクレオチドプローブ５’または３’末端のいずれかであり得る）に存在する。
【0214】
任意のオリゴヌクレオチドプローブの拘束部分は、ある特定の配列のみを有するものであり得るため、プローブの拘束部分内のヌクレオシドの１つ以上の正体がわかることは、他方のヌクレオシドの１つ以上に関する情報を提供する。該情報は、他方のヌクレオシドの１つ以上を正確に同定するのに充分である場合もそうでない場合もあるが、拘束部分の他方のヌクレオシドの１つ以上の正体の１つ以上の可能性を排除するには充分である。配列決定法ＡＢのある特定の好ましい実施形態において、プローブの拘束部分の１つのヌクレオシドの正体がわかることは、拘束部分の他方のヌクレオシドの各々を正確に同定する、すなわち、拘束部分を含むヌクレオシドの正体および順序を決定するのに充分である。
【0215】
上記の配列決定法のように、鋳型に相補的な伸長プローブ内の最も近位ヌクレオシドは、初期オリゴヌクレオチドの伸長可能な末端にライゲートされ（伸長、ライゲーションおよび検出の第１のサイクル）、伸長、ライゲーションおよび検出の後続のサイクルでは、伸長されたオリゴヌクレオチドプローブの伸長可能な末端にライゲートされる。検出により、新たにライゲートされたプローブが属するプローブファミリー名が決定される。プローブの拘束部分内の各位置は、少なくとも２重に縮重しているため、プローブファミリー名は、それ自体では、拘束部分内の任意のヌクレオチドが同定されない。しかしながら、拘束部分の配列は、長さｊの可能な全ての配列の部分集団の１つであるため、プローブファミリーが同定されることにより、拘束部分の配列のある特定の可能性が排除される。プローブの拘束部分は、その配列決定部分を構成する。したがって、属するプローブファミリーが同定されることによってプローブの拘束部分の１個以上のヌクレオシドの
正体の１つ以上の可能性が排除されることにより、伸長プローブハイブリダイズする鋳型内のヌクレオチドの正体１つ以上の可能性が排除される。本発明の好ましい実施形態において、部分拘束プローブは、切れやすい連結を任意の２つのヌクレオシド間に含む。
【0216】
ある特定の実施形態において、部分拘束プローブは、一般構造$(X)_j(N)_k$を有し、式中、Ｘは、ヌクレオシドを表し、$(X)_j$は各位置で少なくとも２重に縮重しており、そのため、Ｘは、異なる塩基対合特異性を有する任意の少なくとも２ヌクレオシドであり得、Ｎは任意のヌクレオシドを表し、ｊは少なくとも２であり、ｋは１～１００であり、少なくとも１つのＮまたはプローブ末端のＸ以外のＸは検出可能な部を含む。好ましくは、$(N)_k$独立して、各位置で４重に縮重しており、そのため、各プローブにおいて、$(N)_k$は、$(N)_k$の１つ以上の位置が縮重低減性ヌクレオチドで占められ得る以外は、長さｋの可能な全ての配列を表す。$(X)_j$内のヌクレオシドは同一であっても異なっていてもよいが、独立して選択されない。換言すると、各プローブにおいて、$(X)_j$は、長さｊの可能な全ての配列の部分集団のみを表し得る。したがって、$(X)_j$の１つ以上のヌクレオチドの正体により、他方のヌクレオシドの１つ以上の可能な正体が制限される。したがって、プローブは２つの部分を含有し、その$(N)_k$は非拘束部分であり、（

Ｘ）$_{j}$は拘束部分である。

【0217】

本発明のある特定の好ましい実施形態において、部分拘束プローブは、構造５’－（Ｘ）$_{j}$（Ｎ）$_{k}$ N_B＊－３’または３’－（Ｘ）$_{j}$（Ｎ）$_{k}$ N_B＊－５’（式中、Ｎは任意のヌクレオシド表し、N_Bはリガーゼによって伸長可能でない部分を表し、＊は検出可能な部分を表し、（Ｘ）$_{j}$は、各位置で少なくとも２重に縮重しているプローブの拘束部分であり、（Ｘ）$_{j}$内のヌクレオシドは同一であっても異なっていてもよいが、独立して選択されない、少なくとも１つのヌクレオシド間結合は切れやすい連結であり、ｊは少なくとも２であり、ｋは１～１００である、ただし、検出可能な部分は、任意のヌクレオシドＮまたはプローブ末端のＸ以外のＸ上に、N_Bの代わりまたはこれに加えて存在し得るものとする）を有する。切れやすい連結は、（Ｘ）$_{j}$内の２つのヌクレオシド間、（Ｘ）$_{j}$内の最も遠位ヌクレオチドと（Ｎ）ｋの最も近位ヌクレオシド間、（Ｎ）$_{k}$内のヌクレオシド間、または（Ｎ）ｋ内の末端ヌクレオシドとN_B間に存在し得る。好ましくは、切れやすい連結はホスホロチオレート結合である。

【0218】

別のより好ましい本発明の実施形態では、プローブは、構造５’－（ＸＹ）（Ｎ）$_{k}$ N_B＊－３’または３’－（ＸＹ）（Ｎ）$_{k}$ N_B＊－５’（式中、Ｎは任意のヌクレオシド表し、N_Bはリガーゼによって伸長可能でない部分を表し、＊は検出可能な部分を表し、ＸＹはプローブの拘束部分であり、ここで、ＸおよびＹは、同一または異なるが独立して選択されないヌクレオシドを表し、ＸおよびＹは少なくとも２重に縮重しており、少なくとも１つのヌクレオシド間結合は切れやすい連結であり、ｋは１～１００（両端を含む）である、ただし、検出可能な部分は、任意のヌクレオシドＮまたはプローブ末端のＸ以外のＸ上に、N_Bの代わりまたはこれに加えて存在し得るものとする）を有する。好ましくは、切れやすい連結はホスホロチオレート結合である。構造５’－（ＸＹ）（Ｎ）$_{k}$ N_B＊－３’を有するプローブは、５’→３’方向の配列決定に有用である。構造３’－（ＸＹ）（Ｎ）$_{k}$ N_B＊－５’　を有するプローブは、３’→５’方向の配列決定に有用である。

【0219】

ある特定の好ましいプローブの構造を、以下により詳細に示す。５’→３’方向の配列決定には、構造５’－Ｏ－Ｐ－Ｏ－（Ｘ）$_{j}$（Ｎ）$_{k}$－Ｏ－Ｐ－Ｓ－（Ｎ）$_{i}$ N_B＊－３’（式中、Ｎは任意のヌクレオシドを表し、N_Bはリガーゼによって伸長可能でない部分を表し、＊は検出可能な部分を表し、（Ｘ）$_{j}$は、各位置で少なくとも２重に縮重しているプローブの拘束部分であり、（Ｘ）$_{j}$内のヌクレオシドは同一であっても異なっていてもよいが、独立して選択されない、ｊは少なくとも２であり、（ｋ＋ｉ）は１～１００であり、ｋは１～１００であり、ｉは０～９９である、ただし、検出可能な部分は、（Ｎ）$_{i}$の任意のヌクレオシド上にN_Bの代わりまたはこれに加えて存在し得るものとする）を有する部分拘束プローブ。本発明のある特定の実施形態において、（Ｘ）$_{j}$は（ＸＹ）であり、ここで、ＸおよびＹは、少なくとも２重に縮重しており、同一または異なるが独立して選択されないヌクレオチドを表す。本発明のある特定の実施形態において、ｉは０である。

【0220】

５’→３’方向の配列決定のための他の好ましいプローブは、構造５’－Ｏ－Ｐ－Ｏ－（Ｘ）$_{j}$－Ｏ－Ｐ－Ｓ－（Ｎ）$_{i}$ N_B＊－３’を有し、式中、Ｎは任意のヌクレオシドを表し、N_Bはリガーゼによって伸長可能でない部分を表し、＊は検出可能な部分を表し、（Ｘ）$_{j}$は、各位置で少なくとも２重に縮重しているプローブの拘束部分であり、（Ｘ）$_{j}$内のヌクレオチドは同一であっても異なっていてもよいが、独立して選択されない、ｊは少なくとも２であり、ｉは１～１００である、ただし、（Ｎ）$_{i}$の任意のヌクレオシド上に、N_Bの代わりまたはこれに加えて存在し得るものとする。本発明のある特定の実施形態において、（Ｘ）$_{j}$は（ＸＹ）であり、ここで、位置ＸおよびＹは少なくとも２重に縮重しており、ＸおよびＹは、同一または異なるが独立して選択されないヌクレオシドを

表す。５’→３’方向の配列決定のためのまた他の好ましいプローブは、構造５’－Ｏ－Ｐ－Ｏ－（Ｘ）$_{j}$－Ｏ－Ｐ－Ｓ－（Ｘ）$_{k}$（Ｎ）$_{i}$ N_B＊－３’を有し、式中、Ｎは任意のヌクレオシドを表し、N_Bはリガーゼによって伸長可能でない部分を表し、＊は検出可能な部分を表し、（Ｘ）$_{j}$－Ｏ－Ｐ－Ｓ－（Ｘ）$_{k}$は、各位置で少なくとも２重に縮重しているプローブの拘束部分であり、（Ｘ）$_{j}$－Ｏ－Ｐ－Ｓ－（Ｘ）$_{k}$内の位置は少なくとも２重に縮重しており、同一であっても異なっていてもよいが、独立して選択されない、ｊおよびｋはともに少なくとも１であり、（ｊ＋ｋ）は少なくとも２（例えば、２、３、４または５）であり、ｉは１～１００である、ただし、検出可能な部分は、（Ｎ）$_{i}$の任意のヌクレオシド上に、N_Bの代わりまたはこれに加えて存在し得るものとする。

【0221】

本発明のある特定の実施形態において、ｊおよびｋはともに１である。

【0222】

３’→５’方向の配列決定では、構造５’－N_B＊（Ｎ）$_{i}$－Ｓ－Ｐ－Ｏ－（Ｎ）$_{k}$－Ｏ－Ｐ－Ｏ－（Ｘ）$_{j}$－３’（式中、Ｎは任意のヌクレオシドを表し、N_Bはリガーゼによって伸長可能でない部分を表し、＊は検出可能な部分を表し、（Ｘ）$_{j}$は、各位置で少なくとも２重に縮重しているプローブの拘束部分であり、（Ｘ）$_{j}$内のヌクレオシドは同一であっても異なっていてもよいが、独立して選択されない、ｊは少なくとも２であり、（ｋ＋ｉ）は１～１００であり、ｋは１～１００であり、ｉは０～９９である、ただし、検出可能な部分は、（Ｎ）$_{i}$の任意のヌクレオシド上に、N_Bの代わりまたはこれに加えて存在し得るものとする）を有する部分拘束プローブ。本発明のある特定の実施形態において、（Ｘ）$_{j}$は（ＸＹ）であり、ここで、ＸおよびＹ少なくとも２重に縮重しており、同一または異なるが独立して選択されないヌクレオシドを表す。本発明のある特定の実施形態において、ｉは０である。

【0223】

３’→５’方向の配列決定のための他の好ましいプローブは、構造５’－N_B＊（Ｎ）$_{i}$－Ｓ－Ｐ－Ｏ－（Ｘ）$_{j}$－３’を有し、式中、Ｎは任意のヌクレオシドを表し、N_Bはリガーゼによって伸長可能でない部分を表し、＊は検出可能な部分を表し、（Ｘ）$_{j}$は、各位置で少なくとも２重に縮重しているプローブの拘束部分であり、（Ｘ）$_{j}$内のヌクレオシドは同一であっても異なっていてもよいが、独立して選択されない、ｊは少なくとも２であり、ｉは１～１００である、ただし、検出可能な部分は、（Ｎ）$_{i}$の任意のヌクレオシド上に、N_Bの代わりまたはこれに加えて存在し得るものとする。本発明のある特定の実施形態において、（Ｘ）$_{j}$は（ＸＹ）であり、ここで、ＸおよびＹ少なくとも２重に縮重しており、同一または異なるが独立して選択されないヌクレオシドを表す。本発明のある特定の実施形態において、ｊは２～５であり、例えば、任意の部分拘束プローブにおいて２、３、４または５である。

【0224】

３’→５’方向の配列決定のためのまた他の好ましいプローブは、構造５’－N_B＊（Ｎ）$_{i}$－Ｓ－Ｐ－Ｏ－（Ｘ）$_{k}$－Ｏ－Ｐ－Ｏ－（Ｘ）$_{j}$－３’（式中、Ｎは任意のヌクレオシドを表し、N_Bはリガーゼによって伸長可能でない部分を表し、＊は検出可能な部分を表し、－（Ｘ）$_{k}$－Ｏ－Ｐ－Ｏ－（Ｘ）$_{j}$は、各位置で少なくとも２重に縮重しているプローブの拘束部分であり、－（Ｘ）$_{k}$－Ｏ－Ｐ－Ｏ－（Ｘ）$_{j}$内のヌクレオシドは同一であっても異なっていてもよいが、独立して選択されない、ｊおよびｋはともに少なくとも１であり、（ｊ＋ｋ）は少なくとも２（例えば、２、３、４または５）であり、ｉは１～１００である、ただし、検出可能な部分は、（Ｎ）$_{i}$の任意のヌクレオシド上に、N_Bの代わりまたはこれに加えて存在し得るものとする）を有する。ある特定の実施形態において、ｊ＝１およびｋ＝１である。

【0225】

切れやすい連結が（Ｘ）$_{j}$の最も近位ヌクレオシドと（Ｘ）$_{j}$の次に最も近位ヌクレオシドとの間に位置する本発明の実施形態では、プローブファミリーの序列リスト名は、伸長されたオリゴヌクレオチドプローブが、各サイクルで１つのヌクレオチドによって伸長

されるため、単一の初期オリゴヌクレオチドから開始される伸長、ライゲーション、検出および切断の連続的サイクルによって得られ得る。切れやすい連結が、その他の２つのヌクレオシド間に位置する本発明の実施形態では、プローブファミリーの序列リスト名は、配列決定法Ａで記載のような、結合反応において異なる位置にハイブリダイズする初期オリゴヌクレオチドが使用される複数の配列決定反応から得られる結果から合成される。

【0226】

上記のもの以外の任意の多数の構造を有するプローブが、配列決定法ＡＢに使用され得ることは理解されよう。例えば、プローブは、ＸＮＹ（Ｎ）$_k$（式中、拘束されたヌクレオシドＸおよびＹは隣接してない）、またはＸＩＹ（Ｎ）$_k$（式中、Ｉは普遍塩基である）などの構造を有するものであり得る。（Ｎ）$_k$Ｘ（Ｎ）$_l$、（Ｎ）$_i$Ｘ（Ｎ）$_j$Ｙ（Ｎ）$_k$Ｚ（Ｎ）$_l$、（Ｎ）$_i$Ｘ（Ｎ）$_j$ＹＩＺ（Ｎ）$_l$および（Ｎ）$_i$Ｘ（Ｎ）$_j$Ｙ（Ｎ）$_k$Ｚ（Ｉ）$_l$は、さらなる可能性を表す。上記のプローブのように、これらのプローブは、切れやすい連結、検出可能な部分、および一方の末端にリガーゼによって伸長可能でない部分を含む。好ましくは、プローブは、リガーゼによって伸長可能でない部分と反対のプローブ末端のヌクレオチドに結合された検出可能な部分含まない。任意のこれらの構造または他のものを有するプローブを含むプローブファミリーは、各プローブファミリーが異なる配列の複数の標識されたプローブを含み、配列内の各位置において、プローブファミリーが、該位置に異なる塩基を有する少なくとも２つのプローブを含むという基準を満たす。好ましくは、各プローブ内のヌクレオシドの総数は１００以下、例えば３０以下である。

【0227】

オリゴヌクレオチド伸長プローブファミリーのコード化。本発明の配列決定法は、コード化されたプローブファミリーを利用するものである。「コード化」は、ある特定の標識を、規定の１組の配列の１つを有する部分を含むプローブと会合させ、その結果、該規定の組の配列のメンバーである配列を有する部分を含むプローブが該標識で標識されるスキームをいう。一般に、コード化では、複数の識別可能な標識の各々を１つ以上プローブと会合させ、その結果、各識別可能な標識が異なる群のプローブと会合され、各プローブがたった１つの標識（これは、検出可能な部分の組合せを含み得る）によって標識される。好ましくは、各プローブ群のプローブは、各々、同じ規定の組の配列のメンバーである配列を有する部分を含む。該部分は、単一のヌクレオシドであってもよく、多くのヌクレオシド長、例えば、２、３、４、５またはそれ以上のヌクレオシド長であってもよい。該部分の長さは、完全長のプローブ小画分のみを構成するものであってもよく、プローブ全体を構成するものであってもよい。規定の組の配列は、単一の配列のみを含むものであってもよく、該部分の長さに応じて、任意の数の異なる配列を含むものであってもよい。例えば、該部分が単一のヌクレオシドである場合、規定の組の配列は、最大４つの要素（Ａ、Ｇ、Ｃ、Ｔ）　を有し得る。部分が２つのヌクレオシド長である場合、規定の組の配列は、１６個までの要素（ＡＡ、ＡＧ、ＡＣ、ＡＴ、ＧＡ、ＧＧ、ＧＣ、ＧＴ、ＣＡ、ＣＧ、ＣＣ、ＣＴ、ＴＡ、ＴＧ、ＴＣ、ＴＴ）　を有し得る。一般に、規定の組の配列は、可能な配列の総数よりも少数の要素を含み、コード化では、１つより多くの規定の組の配列を用いる。

【0228】

本明細書に記載した配列決定法Ａでは、一般的に、近位ヌクレオシドプローブ内の（すなわち、伸長可能なプローブ末端にライゲートされたヌクレオシド）と標識の正体との間に直接的な対応がある単一のコード化を有する１組のプローブを利用する。近位ヌクレオシドは、鋳型内でハイブリダイズするヌクレオチドに相補的であり、そのため、新たにライゲートされたプローブの近位ヌクレオシドの正体により、伸長された二本鎖の位置と反対側に位置する鋳型内のヌクレオチドの正体が決定される。一般的な意味において、本明細書に記載の他方の配列決定法に有用なプローブは、構造Ｘ（Ｎ）$_k$を有し、式中、Ｘは近位ヌクレオシドであり、各ヌクレオシドＮは４重に縮重しており、そのため、長さｋの可能な全ての配列が、該プローブを構成するオリゴヌクレオチドプローブ分子のプールに

提示される。したがって、例えば、一部のオリゴヌクレオチドプローブ分子がＡを位置ｋ＝１に含有する場合、あるものはＧを位置ｋ＝１に含有し、あるものはＣを位置ｋ＝１に含有し、あるものはＴを位置ｋ＝１に含有し、他の位置ｋについても同様であり、（Ｎ）$_k$ 内のＸに隣接するヌクレオシドは位置ｋ＝１を占めるとみなされる；（Ｎ）$_k$ 内の次のヌクレオシドは位置ｋ＝２を占めるとみなされるなどである。しかしながら、任意の所与のオリゴヌクレオチドプローブにおいて、Ｘは、たった１つの塩基対合特異性を表し、これは、典型的には、ある特定のヌクレオシドの正体、例えば、Ａ、Ｇ、ＣまたはＴに対応する。したがって、Ｘは、典型的には、ある特定のプローブを構成するプローブ分子のプール内で均一にＡ、Ｇ、ＣまたはＴである。図２は、構造Ｘ（Ｎ）$_k$ を有するプローブの適当なコード化を示す。このコード化によれば、Ｘ＝Ｃを有するプローブは標識「赤色」に割り当てられる；Ｘ＝Ａを有するプローブは標識「黄色」に割り当てられる；Ｘ＝Ｇを有するプローブは標識「緑色」に割り当てられる；およびＸ＝Ｔを有するプローブは標識「青色」に割り当てられる。したがって、プローブの配列決定部分とその標識間に１対１対応が存在する。

【0229】

新たにライゲートされた伸長プローブの標識の正体が、伸長プローブの最も近位ヌクレオシドの正体に対応する上記のアプローチは、標識の正体が伸長プローブ最も近位ヌクレオシドのみの正体に対応するのではなく、伸長プローブ内の最も近位の２つ以上のヌクレオシドの配列に対応し、その結果、鋳型内の多数のヌクレオチドの正体が、伸長、ライゲーションおよび検出の１サイクルで決定され得る（典型的には、その後、切断を行なう）コード化を包含するように拡張され得ることは認識されよう。しかしながら、かかるコード化は、依然として標識をオリゴヌクレオチド伸長プローブ内の単一の配列に会合させるものであり得、そのため、対向する位置にある鋳型内の相補的なヌクレオチドの正体が同定され得る。例えば、上記のように、１サイクルで２つのヌクレオチドを同定するためには、各々が対応する標識を有する（すなわち、１６種類の識別可能な標識）１６種類の異なるオリゴヌクレオチドプローブが必要とされ得る。

【0230】

配列決定法ＡＢでは、標識をプローブと会合させる代替的なアプローチを用いる。標識の正体とプローブの配列決定部分の配列間の１対１対応ではなく、同じ標識を、異なる配列決定部分を有する多数のプローブに割り当てる。プローブは部分的に拘束されており、プローブの拘束部分がその配列決定部分である。したがって、同じ標識が、各々が異なる配列拘束部分を有し、該配列が規定の１組の配列の１つである複数の異なるプローブに割り当てられる。上記のように、同じ標識を含むプローブは、「プローブファミリー」を構成する。該方法では、各々が異なる配列の拘束部分を有し、該配列が規定の１組の配列の１つである複数のプローブを含む複数のかかるプローブファミリーを用いる。

【0231】

複数のプローブファミリーをプローブファミリーの「コレクション」という。プローブファミリーのコレクション内の各プローブファミリー内のプローブは、該コレクション内の他のプローブファミリーを標識するために使用される標識と識別可能な標識で標識されている。各プローブファミリーは、好ましくは、その独自の規定の組の配列を有する。好ましくは、各プローブファミリー内のプローブの拘束部分は同じ長さであり、好ましくは、プローブファミリーのコレクション内のプローブファミリーの拘束部分は同じ長さである。好ましくは、プローブファミリーのコレクション内のプローブファミリーの規定配列の組の組合せは、拘束部分の長さの可能な全ての配列を含む。好ましくは、プローブファミリーのコレクションは、４つの識別可能に標識されたプローブファミリーを含むか、または外ファミリーからなる。好ましくは、プローブの拘束部分は２ヌクレオシド長である。

【0232】

広範なさまざまな異なるようにコード化された識別可能に標識されたプローブファミリーのコレクションが上記の基準を満足し、本発明の方法を実施するために使用され得る。

しかしながら、ある特定のプローブファミリーのコレクションが好ましい。部分拘束プローブを含む４つの識別可能に標識されたプローブファミリーの好ましいコレクションの例示的なコード化を図２５Ａに示す。図２５Ａに示すように、拘束部分は、プローブ内の２つの最も３’側のヌクレオシドからなる。プローブファミリーは、「赤色」、「黄色」、「緑色」および「青色」に標識される。各プローブファミリー内のプローブは、その配列が規定の１組の配列の１つであり、該規定の組は各プローブファミリーで異なる拘束部分を含む。例えば、プローブの近位端であるみなされる各配列の３’末端からから始めると、「赤色」プローブファミリーの規定の組の配列は｛ＣＴ、ＡＧ、ＧＡ、ＴＣ｝である；「黄色」のプローブファミリーの規定の組の配列は｛ＣＣ、ＡＴ、ＧＧ、ＴＡ｝である；「緑色」のプローブファミリーの規定の組の配列は｛ＣＡ、ＡＣ、ＧＴ、ＴＧ｝である；「青色」のプローブファミリーの規定の組の配列は｛ＣＧ、ＡＡ、ＧＣ、ＴＴ｝である。各規定の組は、その他の組の１つに存在する任意のメンバーを含まず、好ましい特徴である。また、プローブファミリーのコレクション内のプローブファミリーの規定配列の組の組合せは、長さ２の可能な全ての配列、すなわち、可能な全てのジヌクレオシドを含む。このプローブファミリーのコレクションの別の特徴（好ましいが必要ではない）は、プローブの拘束部分内の各位置が、４重に縮重している、すなわちＡ、Ｇ、ＣまたはＴのいずれかで占められ得ることである。このプローブファミリーのコレクションの別の特徴（好ましいが必要ではない）は、規定配列の組の各々において、たった１つの配列が、任意の特異的ヌクレオシドを任意の位置に、例えば、最も近位の位置または任意の他の位置に有することである。規定配列の組の各々において、たった１つの配列が、任意の特異的ヌクレオシドを拘束部分内の２位以降に（最も近位ヌクレオシドを１位とみなす）を有することは特に好ましいが必要ではない。例えば、赤色プローブファミリーの規定の組の配列において、たった１つの配列がＴを２位に有する；たった１つの配列がＧを２位に有する；たった１つの配列がＡを２位に有する；たった１つの配列がＣを２位に有する。

【0233】

図２５Ａに示したものなどの任意の特定のコード化が与えられると、プローブファミリーの１つのプローブの拘束部分内の１個以上のヌクレオシドの正体がわかることにより、該プローブの拘束部分内の他方のヌクレオチドに関する情報が提供される。最も一般的な意味において、プローブファミリー内のプローブの拘束部分内の１個以上のヌクレオシドの正体がわかることにより、該プローブファミリーの規定の組の配列は、該正体を該位置を有するヌクレオシドを有する配列を含まないため、他方の位置の１つにおけるヌクレオシドの１つ以上の可能な正体を排除するのに充分な情報が提供される。典型的には、プローブファミリー内のプローブの拘束部分内の１個以上のヌクレオシドの正体がわかることにより、複数のヌクレオシド、例えば、その他のヌクレオシドの各々の１つ以上の可能な正体を排除するのに充分な情報が提供される。好ましいコード化では、プローブファミリー内のプローブの含有された部分の１個以上のヌクレオシドの正体がわかることにより、プローブ内のその他のヌクレオシドの各々の１つ以外すべての可能性が排除される。例えば、図２５Ａに示すコード化されたプローブファミリーの場合、プローブが赤色ファミリーのメンバーであることがわかっている場合、また、最も近位ヌクレオシドがＣであることもわかっている場合、隣接するヌクレオシドはＴのはずである。同様に、プローブが緑色ファミリーのメンバーであることがわかっている場合、また、最も近位ヌクレオシドがＧであることもわかっている場合、隣接するヌクレオシドはＴのはずである。したがって、拘束部分の１つのヌクレオシドの正体がわかることは、他方のヌクレオシドの１つ以外すべての可能性を排除するのに充分であり、そのため、他方のヌクレオシドの正体が完全に特定化される。しかし、プローブの拘束部分の少なくとも１つのヌクレオシドの正体がわからないと、拘束部分の各位置のヌクレオシドがＡ、Ｇ、ＣまたはＴであり得るため、属する　プローブファミリー名がわかったことに基づいて、プローブ内の特異的ヌクレオシドの正体に関する情報を取得することはまったくできない。図２５Ｂは、プローブファミリーの好ましいコレクション（上側パネル）および配列決定法ＡＢを用いたライゲーション、検出および切断のサイクル（下側パネル）を示す。

【0234】
本発明者は、図２５Ａに示すような、２ヌクレオシド長であって、プローブファミリーのコレクションの好都合な特徴を有する拘束部分を含む２４個のプローブファミリーのコレクションを設計した。これらのプローブファミリーは、プローブが属するプローブファミリー名がわかること、およびプローブ内の１つのヌクレオシドの正体がわかることが、拘束部分の他方のヌクレオシドを正確に同定するのに充分であるという点で最大の情報を与える。これは、すべてのプローブおよび各拘束部分内のすべてのヌクレオシドについてそうである。プローブファミリーの２４個の好ましいコレクションの各々のコード化スキームを表１に示す。表１では、１～２４の範囲のコード化ＩＤを各プローブファミリーのコレクションに割り当てている。各コード化により、配列決定法ＡＢにおける使用のための一般構造（ＸＹ）N_kの好ましいプローブファミリーのコレクションの拘束部分が規定され、それによりコレクションそれ自体が規定される。表１において、コード化ＩＤの下の欄の値１は、該コード化に従って、それぞれ第１および第２の欄に示されたヌクレオシドＸおよびＹを含むプローブが、第１のプローブファミリーに割り当てられることを示す；（ｉｉ）コード化ＩＤの下の欄の値２は、該コード化に従って、それぞれ第１および第２の欄に示されたヌクレオシドＸおよびＹを含むプローブが、第２のプローブファミリーに割り当てられることを示す；（ｉｉｉ）コード化ＩＤの下の欄の値３は、該コード化に従って、それぞれ第１および第２の欄に示されたヌクレオシドＸおよびＹを含むプローブが、第３のプローブファミリーに割り当てられることを示す；ならびに（ｉｖ）コード化ＩＤの下の欄の値４は、該コード化に従って、それぞれ第１および第２の欄に示されたヌクレオシドＸおよびＹを含むプローブが、第４のプローブファミリーに割り当てられることを示す。値１、２、３および４は各々、標識を表す。例えば、コード化９により図２５Ａに示すプローブファミリーのコレクションが規定され、この場合、１は青色を表し、２は緑色を表し、３は赤色を表し、４は黄色を表す。標識に対する値の割り当ては随意であり、例えば、１は、充分等しく緑色、赤色または黄色を表し得ることは認識されよう。値１、２、３および４と標識との関連を変更することによって各プローブファミリー内のプローブの組は変更され得ず、単に、異なる標識が各プローブファミリーと関連することになり得るだけである。
【0235】
表１：オリゴヌクレオチドプローブファミリーコード化
【0236】

【表1】

		コード化ＩＤ																						
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
X	Y																								
A	A	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
C	A	2	4	3	2	2	4	3	2	2	3	4	3	2	3	3	4	2	3	4	4	2	4	3	4
G	A	4	3	2	3	3	2	4	4	3	2	2	4	4	2	4	3	4	2	3	2	3	2	4	3
T	A	3	2	4	4	4	3	2	3	4	4	3	2	3	4	2	2	3	4	2	3	4	3	2	2
A	C	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
C	C	1	1	1	1	1	1	1	1	4	4	3	4	4	4	4	3	3	4	3	3	3	3	4	3
G	C	3	4	4	4	4	3	3	3	1	1	1	1	3	3	3	4	1	1	1	1	4	4	3	4
T	C	4	3	3	3	3	4	4	4	3	3	4	3	1	1	1	1	4	3	4	4	1	1	1	1
A	G	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
C	G	4	2	4	4	4	2	4	4	1	1	1	1	1	1	1	1	4	2	2	2	4	2	2	2
G	G	1	1	1	1	2	4	2	2	4	4	4	2	2	4	2	2	2	4	4	4	1	1	1	1
T	G	2	4	2	2	1	1	1	1	2	2	2	4	4	2	4	4	1	1	1	1	2	4	4	4
A	T	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
C	T	3	3	2	3	3	3	2	3	3	2	2	2	3	2	2	2	1	1	1	1	1	1	1	1
G	T	2	2	3	2	1	1	1	1	2	3	3	3	1	1	3	1	3	3	2	3	2	3	2	2
T	T	1	1	1	1	2	2	3	2	1	1	1	1	2	3	1	3	2	2	3	2	3	2	3	3

好ましいプローブファミリーのコレクションを規定する表１の使用をさらに説明するため、コード化１７を考察する。このコード化によれば、拘束部分ＡＡ、ＧＣ、ＴＧおよびＣＴを有するプローブが標識１（例えば、赤色）に割り当てられ、拘束部分ＣＡ、ＡＣ、ＧＧおよびＴＴを有するプローブが標識２（例えば、黄色）に割り当てられ、拘束部分ＴＡ、ＣＣ、ＡＧおよびＧＴを有するプローブが標識３（例えば、緑色）に割り当てられ、拘束部分ＧＡ、ＴＣ、ＣＧおよびＡＴを有するプローブが標識４（例えば、青色）に割り当てられている。得られるプローブファミリーのコレクションを図２６に示す。

【0237】

図２７Ａ～２７Ｃは、プローブファミリーの２４個の好ましいコレクションを図式的に規定する代替法を表す。該方法では、図２７Ａのものなどのダイアグラムを利用する。かかるダイアグラム内の第１の欄は第１の塩基を表す。各標識は、４種類の異なる塩基配列に結合されている。その各々は、第１の欄の塩基を選択された標識の欄の塩基と並置させて示されている。例えば、項目「第１の塩基」の欄にＡがある場合、配列ＡＡを有する拘束部分を有するプローブは、プローブファミリー１（標識１）に割り当てられ、配列ＡＣを有する拘束部分を有するプローブは、プローブファミリー２（標識２）に割り当てられ、配列ＡＧを有する拘束部分を有するプローブは、プローブファミリー３（標識３）　に割り当てられ、配列ＡＴを有する拘束部分を有するプローブは、プローブファミリー４（標識４）に割り当てられる。プローブファミリーに対する割り当ては、Ｃ、ＧまたはＴで始まる拘束部分を有するプローブと同様にして行なわれる。したがって、図２７Ａに示す

ように塩基で空欄を埋めたダイアグラムは、図２７Ｂに示すようなコード化に翻訳され、ここでは、｛ＡＡ、ＣＣ、ＧＧ、ＴＴ｝の組の拘束部分を有するプローブはプローブファミリー１に割り当てられ、｛ＡＣ、ＣＡ、ＧＴ、ＴＧ｝の組の拘束部分を有するプローブはプローブファミリー２に割り当てられ、｛ＡＧ、ＣＴ、ＧＣ、ＴＡ｝の組の拘束部分を有するプローブはプローブファミリー３に割り当てられ、｛ＡＴ、ＣＧ、ＧＡ、ＴＣ｝の組の拘束部分を有するプローブはプローブファミリー４に割り当てられる。図２７Ｃは、プローブファミリーの２４個の好ましいコレクションの各々を作製するために、図２７Ａダイアグラムの陰影部分の代わりに挿入され得るダイアグラムを示す。配列決定法ＡＢにおいて好ましいプローブファミリーのコレクションを使用する方法は、以下にさらに記載する。

【0238】

表１に規定されるコード化されたプローブファミリーの２４個のコレクションは、配列決定法ＡＢにおける使用のためのプローブファミリーのコレクションの好ましい実施形態のみを表す。プローブファミリー名がわかることとともに、拘束部分内の１個以上のヌクレオシドの正体がわかることによって、１つ以上他のヌクレオシドに関する情報が提供される同じ基本原理を用いる広範なさまざまな他のコード化スキーム、プローブファミリーおよびプローブ構造が使用され得る。プローブファミリーの好ましいコレクションと比較すると、プローブファミリーのやや好ましいコレクションは、一般的に、（ｉ）少なくとも一部プローブに関して、プローブファミリー名がわかることによってもたらされる情報の量およびヌクレオシドの正体が少ない；または（ｉｉ）少なくとも一部のプローブに関して、プローブファミリー名がわかることによってもたらされる情報の量がより多いため、やや好ましい。

【0239】

一般に、プローブファミリーのやや好ましいコレクションは、配列決定法ＡＢを、好ましいプローブファミリーのコレクションが使用される様式と同様にして行うために使用され得る。しかしながら、デコード化に必要とされる該工程が異なり得る。例えば、状況によっては、候補配列を互いに比較することが、配列の少なくとも一部分を決定するのに充分であり得る。

【0240】

プローブが２ヌクレオシド長である拘束部分を含むプローブファミリーのやや好ましいコレクションの一例を図２８に示す。このコード化によれば、｛ＡＡ、ＡＣ、ＧＡ、ＧＣ｝の組の拘束部分を有するプローブはプローブファミリー１に割り当てられ、｛ＣＡ、ＣＣ、ＴＡ、ＴＣ｝の組の拘束部分を有するプローブはプローブファミリー２に割り当てられ、｛ＡＧ、ＡＴ、ＧＧ、ＧＴ｝の組の拘束部分を有するプローブはプローブファミリー３に割り当てられ、｛ＣＧ、ＣＴ、ＴＧ、ＴＴ｝の組の拘束部分を有するプローブはプローブファミリー４に割り当てられる。このプローブファミリーのコレクションでは、プローブファミリー名がわかることにより、プローブファミリー名を決定するためにその標識が検出された新たにライゲートされた伸長プローブ内の近位ヌクレオシドと反対の位置に位置する鋳型内のヌクレオチドの正体についてのある特定の可能性が排除される。例えば、プローブファミリー名が１である場合、新たにライゲートされた伸長プローブ内の近位ヌクレオシドはＡまたはＧのはずであり、そのため、相補的鋳型内のヌクレオチドはＴまたはＣのはずである。拘束部分内の各位置において少なくとも２つの可能性があるため、ヌクレオチドは正確には同定され得ないが、好ましいプローブファミリーのコレクションが使用される場合とは対照的に、一部の可能性を排除するのには充分な情報が１サイクルから得られる。

【0241】

本発明のある特定の実施形態において、拘束部分が３ヌクレオシド長である部分拘束プローブが使用される。その拘束部分が長さ３可能な全ての配列を含む（これは、好ましい）プローブを含めるため、プローブファミリーのコレクションは、$4^3=64$種の異なるプローブを含むものであるのがよい。図２９Ａは、拘束部分３ヌクレオシド長（トリヌク

レオシド）を有するプローブを含むプローブファミリーのコレクション拘束部分を作製するために使用され得るダイアグラムを示す。この図は、Ａ、Ｇ、ＣおよびＴで表示された４組の列、ならびにプローブファミリー名１、２、３および４の４つの欄を示す。各組みの４列は、内部にヌクレオシドの正体を有するボックスと反対である。トリヌクレオシドのプローブファミリーを決定するため、トリヌクレオシドの最後のヌクレオシドを含むボックスをまず選択する。該ボックスに隣接する４つの列内で、トリヌクレオシド内の第１のヌクレオシドを特定する文字が標示された列を選択する。該列内で、トリヌクレオシドの第２のヌクレオシドを含む欄を選択する。該トリヌクレオシドは、該欄の上部に標示されたプローブファミリーに割り当てられる。例えば、以下の手順に従って、トリヌクレオシド「ＴＣＧ」をプローブファミリーに割り当てる。最後のヌクレオシドが「Ｇ」であるため、「Ｇ」を含むボックスと対向する位置に位置する４列の組、すなわち、第３の組の列に注目する。第１のヌクレオシド「Ｔ」であるため、考慮事項を４つの組の最後の列にさらに限定する。プローブファミリー割り当ては、真ん中のヌクレオシドを含む欄の表題によって決定される。真ん中のヌクレオシドは「Ｃ」であるため、トリヌクレオシドはプローブファミリー１に割り当てられる。類似したプロセスにより、プローブファミリー割り当て：ＡＡＡ＝１；ＡＴＡ＝２；ＡＧＡ＝３；ＧＴＡ＝４；ＧＡＧ＝１；ＴＧＧ＝２などがもたらされる。このプロセスは、可能な全てのトリヌクレオシドがプローブファミリーに割り当てられるまで継続される。

【0242】

図２９Ｂは、３ヌクレオシド長の拘束部分を有するプローブを含むプローブファミリーのコレクションさらなる拘束部分を構築するための手順を示す。該手順は、かかるコレクションを拘束部分が２ヌクレオシド長でありコレクションが４つのプローブファミリーを含む上記のプローブファミリーの２４個の好ましいコレクションの各々から構築するために使用される。プローブファミリーの好ましいコレクションを示す例示的なダイアグラムを図の上側部分に示す。このダイアグラムの欄を、上側のダイアグラムの各欄に割り当てられた色に従って、図の下側部分の欄に直接マッピングしている。したがって、上側のダイアグラムの欄は、左から右に向かって青色、緑色、黄色および赤色である。下側のダイアグラムの欄１のエントリーは、上から下に向かって青色、緑色、黄色および赤色であり、４ヌクレオシドの各組は、上側のダイアグラム欄に対応する。下側ダイアグラムの第２、３および４欄は、欄１の４ヌクレオシドの各組を下方に進行的に移動させることにより作製される。

【0243】

「プローブファミリー」は、各々が同じ標識を有する複数の異なるプローブを含む単一の「スーパープローブ」であるとみなされ得ることは認識されよう。この場合、プローブを構成するプローブ分子は、一般的に、該プローブの任意の部分において実質的に同一の分子の実質的に同一の分子集団ではない。用語「プローブファミリー」の使用は、任意の限定的な効果を有することを意図するものではなく、かかる「スーパープローブ」を構成し得るプローブの特徴の記載の便宜上のために用いる。

【0244】

デコード化

上記のように、少なくとも２つの識別可能に標識されたプローブファミリーを含むプローブファミリーのコレクションを用いる伸長、ライゲーション、検出および切断の連続的サイクルにより、プローブファミリーの序列リスト名が、単一の配列決定反応、または序列リスト内の鋳型内の異なる部位から開始される多重配列決定反応で決定されたプローブファミリー名の合成のいずれかからもたらされる。実施されるサイクル数は、所望の配列の長さおよそ同等であるはずである。序列リストは、相当な量の情報を含むが、直接的に目的の配列が得られる形態ではない。おそらく目的の配列を表す配列を得るため、さらなる工程（１つまたは複数）（その少なくとも１つは、さらなる配列に関する情報の少なくとも１つの項目の収集を伴う）を行なわなければならない。おそらく目的の配列を表す配列は、本明細書において「正確な」配列といい、正確な配列をプローブファミリーの序列

リストから抽出するプロセスは、を「デコード化」という。上記の「序列リスト」内の要素は、リストの作製中またはその後のいずれかで再編成され得ることは認識されよう。ただし、リスト内の要素と鋳型内のヌクレオチド間の対応を含む情報の内容は保持されるものとし、また、再編成、再分化および／または置換（ｐｅｒｍｕｔａｔｉｏｎ）は、デコード化プロセス中に適切に考慮されるものとする（後述）。したがって、用語「序列リスト」は、上記のようにして作製された再編成、再分化および／または置換された変形例の序列リストを包含することが意図される（ただし、かかる再分化および／または置換された変形例は実質的に同じ情報内容を含むものとする）。

【0245】

序列リストは、さまざまなアプローチを用いてデコード化され得る。一部のこれらのアプローチは、プローブファミリーの序列リスト名からの１組の少なくとも１つの候補配列の作製を伴う。候補配列の組により、目的を達成するのに充分な情報が提供され得る。好ましい実施形態において、おそらく目的の配列を表す配列を候補配列の中から、または候補配列と比較される１組の配列から選択するために、１つ以上のさらなる工程が行なわれる。例えば、アプローチの一例において、少なくとも１つの候補配列の少なくとも一部分を、少なくとも１つの他の配列と比較する。正確な配列は、比較に基づいて選択される。本発明のある特定の実施形態において、デコード化は、該方法を反復すること、および元のプローブファミリーのコレクションと異なるようにコード化されたプローブファミリーのコレクションを用いて第２のプローブファミリーの序列リスト名を得ることを伴う。第２のプローブファミリーの序列リストからの情報は、正確な配列を決定するために使用される。ある一部の実施形態において、別のコード化されたプローブファミリーのコレクションを用いて伸長、ライゲーションおよび検出の１サイクルという少ないサイクルから得られる情報は、正確な配列の選択の可能にするのに充分である。換言すると、別のコード化されたプローブファミリーを用いて同定された第１のプローブファミリーにより、どの候補配列が正確であるかを決定するのに充分な情報が提供される。

【0246】

他のデコード化アプローチは、任意の利用可能な配列決定法、例えば、１サイクルの配列決定法Ａによって鋳型内の少なくとも１つのヌクレオチドを特異的に同定することを伴う。１つ以上のヌクレオチド（１つまたは複数）に関する情報が、プローブファミリーの序列リスト名をデコード化するための「鍵」として用いられる。あるいはまた、鋳型の配列決定される部分は、その配列が未知である領域に加えて、既知配列の領域を含むものであり得る。配列決定法ＡＢが、未知配列と既知配列の少なくとも１つのヌクレオチドとの両方を含む鋳型の部分に適用される場合、既知配列が、プローブファミリーの序列リスト名をデコード化するための「鍵」として用いられ得る。以下のセクションに候補配列を作製するプロセスを記載する。その後のセクションに、既知配列との比較、第２の組の候補配列との比較、および既知ヌクレオチド正体の利用によって正確な配列を選択するための候補配列の使用を記載する。

【0247】

候補配列の作製

鋳型の配列決定される領域は、伸長、ライゲーションおよび切断の連続的サイクルによって生成される伸長された二本鎖に相補的であることは認識されよう。したがって、伸長された二本鎖の候補配列を作製することは、鋳型の配列決定される領域の候補配列を作製することに等価である。実際には、鋳型の配列決定される領域の候補配列を作製し得るか、または伸長された二本鎖の候補配列を作製し、その相補配列を得て鋳型の配列決定される領域の候補配列を決定し得る。後者のアプローチをここに記載する。候補配列をプローブファミリー名のリストから作製するため、プローブファミリーのリストの第１のメンバーを考慮する。該プローブファミリーと関連する拘束部分の組により、配列内の最初のヌクレオチドの可能性が、拘束部分の長さに相当する長さに限定される。例えば、拘束部分がジヌクレオチドである場合、伸長された二本鎖内の最初のジヌクレオチドの可能な配列は、該プローブファミリーに含まれるプローブ内に存在する拘束部分に限定される（した

がって、鋳型の配列決定される領域内の第１のジヌクレオチドの可能な配列は、該プローブファミリーに含まれるプローブ内に存在する拘束部分に相補的な組合せに限定される）。最初のジヌクレオチドの可能性は、典型的にはコンピュータによって記録される。同様に、伸長された二本鎖内の第２のジヌクレオチド（すなわち、第１のジヌクレオチドから１ヌクレオチドずれたジヌクレオチド）の可能な配列が、第２のプローブファミリーに含まれるプローブ内に存在する拘束部分に限定される（したがって、鋳型内の第２のジヌクレオチド、すなわち、第１のジヌクレオチドから１ヌクレオチドずれたジヌクレオチドの可能な配列が、第２のプローブファミリーに含まれるプローブ内に存在する拘束部分に相補的な組合せに限定される）。第２のジヌクレオチドの可能な配列もまた記録される。続くジヌクレオチドの可能な配列（ｐｏｓｓｉｂｉｌｉｔｙ）も同様に、決定対象の配列の所望の長さに相当するまで、またはリスト内のプローブファミリーがなくなるまでジヌクレオチドの可能な配列が記録される。

【0248】

可能な配列の記録プロセスの代表例を図３０に示す。図において、プローブファミリー名のリストは、図２５Ａに示すプローブファミリーコレクションを用いて作製されたものと想定している。図３０の左端の欄はプローブファミリーのリストを示し、上から下の順に、黄色、緑色、赤色、青色である。リスト内の各プローブファミリーに対応するジヌクレオチドの配列可能性を図の右側に示す。ヌクレオチド位置を配列可能性の上部に示す。配列は１位から始まり、そのため、第１のジヌクレオチドは１位と２位を占める；第２のジヌクレオチドは２位と３位を占めるなどである。黄色プローブファミリーでは、該可能性は、図３０に示すようにＣＣ、ＡＴ、ＧＧおよびＴＡである。緑色プローブファミリーでは、該可能性は、ＣＡ、ＡＣ、ＧＴおよびＴＧであるなどである。各ジヌクレオチドの可能な配列の記録プロセスは、所望の配列長さに達するまで継続される。

【0249】

可能性の組を作製した後、第１の仮定を、候補配列内の第１のヌクレオチドの正体において行ない、これは、図３０において１位と表示する配列の５’位置であると仮定するものである。第１の仮定は、該ヌクレオチドがＡである、該ヌクレオチドがＧである、該ヌクレオチドがＣであるまたは該ヌクレオチドがＴであるとするものであり得る。

【0250】

各ジヌクレオチドの可能な配列は、隣接するジヌクレオチドが重複する、すなわち、第１のジヌクレオチドの第２のヌクレオチドはまた、第２のジヌクレオチドの第１のヌクレオチドであるため、隣接するジヌクレオチドの可能な配列によって限定されることが観察される。例えば、第１のヌクレオチドをＣと仮定した場合、第１のジヌクレオチドはＣＣとなるはずである。第１のジヌクレオチドがＣＣであるならば、第２のジヌクレオチドは、Ｃをその第１の位置に有するはずである。Ｃをその第１の位置に有する第２のジヌクレオチドの唯一可能な配列はＣＡであるため、第２のジヌクレオチドがＣＡのはずであることは明白である。したがって、最初のの３つのヌクレオチドの配列は、ＣＣＡのはずである。同様に、第３のジヌクレオチドの可能な配列は、第２のジヌクレオチドの可能な配列によって限定される。第２のジヌクレオチドがＣＡであるならば、第３のジヌクレオチドは、その第１の位置にＡを有することが唯一の可能性であるため、ＡＧのはずである。したがって、最初の４つのヌクレオチドの配列はＣＣＡＧのはずである。このプロセスを継続すると、最初の５つヌクレオチドに対して５’－ＣＣＡＧＣ－３’の配列がもたらされる。したがって、ＣＣＡＧＣは第１の候補配列である。

【0251】

第２の候補配列は、第１のヌクレオチドがＡであると仮定することによりもたらされる。この仮定により、ＡＴが第１のジヌクレオチドとなる。ＴＧは、ＡＴが第１のジヌクレオチドの配列と整合する第２のジヌクレオチドの唯一の可能な配列である。ＧＡは、ＴＧが第２のジヌクレオチドの配列と整合する第３のジヌクレオチドの唯一の可能な配列である。ＡＡは、ＧＡが第３のジヌクレオチドの配列と整合する第４のジヌクレオチドの唯一の可能な配列である。これらのジヌクレオチドを完全長の候補配列に合成すると、ＡＴＧ

ＡＡが得られる。同様に、第１のヌクレオチドがＧであるという仮定により、候補配列ＧＧＴＣＧが得られ、第１のヌクレオチドがＴであるという仮定により、候補配列ＴＡＣＴＴが得られる。したがって、各々、仮定された配列の第１のヌクレオチドの異なるヌクレオチドで始まる４つの候補配列がもたらされる。

【0252】

仮定は、その他のヌクレオチドの１つではなく第１のヌクレオチドで行なわなければならないとする要件はない。例えば、仮定は、充分等しく第４のヌクレオチドの正体において行なわれたものであり得、この場合、候補配列は、鋳型に沿って「逆行方向」（すなわち、３’→５’方向）に移動させることによりもたらされ得る。例えば、第４のヌクレオチドがＴであると仮定することは、第４のジヌクレオチドはＴＴのはずであり、第３のジヌクレオチドはＣＴのはずであり、第２のジヌクレオチドはＡＣのはずであり、第１のジヌクレオチドはＣＣのはずであることを意味する（ヌクレオチドは、５’→３の向きに記載されるが、その正体は、配列において３’→５’に移動させることによりもたらされる）。あるいはまた、仮定は、配列の中央の任意のヌクレオチドにおいて行なわれ得、ジヌクレオチドの正体は、５’→３’および３’→５の両方向に移動させることによりもたらされる。ヌクレオチドの１つに対する仮定の非存在下では、各位置がＡ、Ｇ、ＣまたはＴで占められ得るため、各ヌクレオチドの正体が完全に決定されない状態になることは認識されよう。

【0253】

好ましいプローブファミリーのコレクションを用いる場合、任意の単一のヌクレオチド（例えば、第１のヌクレオチド）の正体を仮定すると、唯一の候補配列が生成される。しかしながら、プローブファミリーのやや好ましいコレクションを用いる場合、１つより多くのヌクレオチドの正体を仮定することが必要であり得る。すなわち、第１のヌクレオチドの正体を仮定することでは、残りの配列は完全には特定されない。例えば、プローブファミリーのやや好ましいコレクションは、その規定配列がＡＡおよびＡＣであるメンバーファミリーを含むものであり得る。かかる場合では、第１のヌクレオチドがＡであると仮定することにより、第２のヌクレオチドに対して２つの可能性が残る。プローブファミリーのやや好ましいコレクションを用いる配列決定を以下にさらに論考する。拘束部分が非連続的なヌクレオチドからなる場合、上記のアプローチは、若干の修正を伴ってやはり使用され得ることは認識されよう。

【0254】

候補配列と既知配列との比較による配列同定

一般的に、上記のようにして、伸長された二本鎖の候補配列が決定された場合、鋳型の配列決定される領域に対応する候補配列は、その相補配列を得ることにより得られる。場合によっては、候補配列自体が、目的を達成するのに充分な情報を提供する。例えば、配列決定の目的が、単にある特定の配列可能性を排除することである場合、候補配列とその可能な配列とを比較することで充分であり得る。例えば、図３０に示す候補配列は、配列決定対象の領域はポリＡテイルの一部でないという決定を可能にし得るものである。より長い配列により、配列決定対象の領域がベクターの一部でないことが確認され得る。

【0255】

多くの場合、正確な配列を明確に決定することが望ましい。好ましい本発明の実施形態によれば、正確な配列は、鋳型の配列決定される領域の候補配列を１組の既知配列と比較することにより同定される。既知配列の組は、例えば、ある特定の目的の生物体の１組の配列であり得る。例えば、ヒトＤＮＡを配列決定する場合、候補配列を、Ｈｕｍａｎ　Ｄｒａｆｔ　Ｇｅｎｏｍｅ　Ｓｅｑｕｅｎｃｅと比較し得る。公的に入手可能なヒトゲノム配列情報源の案内については、ＵＲＬ　ｗｗｗ．ｎｃｂｉ．ｎｉｈ．ｇｏｖ／ｇｅｎｏｍｅ／ｇｕｉｄｅ／ｈｕｍａｎ／のウェブサイトを参照のこと。別の例として、感染因子（例えば、被験体から単離された細菌またはウイルス）由来の核酸が配列決定される場合、該細菌またはウイルスのバリアント株の配列を含むデータベースが検索され得る。完全または部分配列のいずれかを含む多くのかかる生物体指定データベースが、当該技術分野で

知られており、配列決定作業が加速されると、より多くが利用可能となる。一部の代表的な例としては、マウス（例えば、ウェブサイトＵＲＬ　ｗｗｗ．ｎｃｂｉ．ｎｌｍ．ｎｉｈ．ｇｏｖ／ｇｅｎｏｍｅ／ｓｅｑ／ＭｍＨｏｍｅ．ｈｔｍｌを参照）、ヒト免疫不全ウイルス（例えば、ウェブサイトＵＲＬ　ｈｉｖ－ｗｅｂ．ｌａｎｌ．ｇｏｖ／ｃｏｎｔｅｎｔ／ｈｉｖ－ｄｂ／ｍａｉｎｐａｇｅ．ｈｔｍｌを参照）、マラリア種Ｐｌａｓｍｏｄｉｕｍ　ｆａｌｃｉｐａｒｕｍ（例えば、ウェブサイトＵＲＬ　ｗｗｗ．ｔｉｇｒ．ｏｒｇ／ｔｄｂ／ｅｄｂ２／ｐｆａｌ／ｈｔｍｌｓ／ｉｎｄｅｘ．ｓｈｔｍｌを参照）などのデータベースが挙げられる。もちろん、生物体指定の組の配列を用いる必要はない。広範なさまざまな生物体およびウイルス由来の配列を含むＧｅｎＢａｎｋ（ウェブサイトＵＲＬ　ｗｗｗ．ｎｃｂｉ．ｎｌｍ．ｎｉｈ．ｇｏｖ／Ｇｅｎｂａｎｋ／）などのデータベースを検索してもよい。データベースは、鋳型を誘導した生物体またはウイルス由来の任意の配列を含むものである必要はない。一般に、配列はゲノム配列、ｃＤＮＡ配列、ＥＳＴｓなどであり得る。多数の配列が検索され得る。

【0256】

単純に検索を行なうことは、目的を達するのに充分であり得る。例えば、ウイルス系核酸が患者から単離された場合、候補配列を該ウイルスの１組の既知配列と比較することにより、該ウイルス系核酸が、該ウイルス由来の配列を含むか含まないかいずれであるかが、配列のマッチングを調べられない場合であっても、決定され得る。マッチングの存在により、患者が該ウイルスに感染していることが確認され得、一方、マッチングの欠如は、患者が該ウイルスに感染していないことを示し得る。

【0257】

ある特定の実施形態において、既知配列の組は、より狭い範囲の配列を含み、これは、配列決定が行なわれる目的に特別に調整されたものであり得る。したがって、配列決定される核酸に関する情報は、既知配列の組を選択するために使用され得る。例えば、鋳型がある特定の遺伝子の配列を示すものであることがわかっている場合、既知配列は、所与の目的の遺伝子座の遺伝子、変異型および野生型配列の異なる対立遺伝子を表すものなどであり得る。候補配列のどれが正確であるかを決定するためには、候補配列を単一の既知配列と比較することだけが必要であり得る。例えば、本発明のある特定の実施形態において、鋳型は、目的の領域を含むＤＮＡ増幅することにより得られる（例えば、目的の領域に隣接するプライマーを用いて）。目的の領域は、変異または多型（例えば、ある特定の疾患と関連する変異または多型）が存在し得る部位を包含する。鋳型が、ある特定の目的の領域由来の配列を示すものであることがわかっている場合、候補配列を該領域の単一の参照配列と、例えば、野生型または変異型形態の配列と比較するだけでよい。換言すると、鋳型の配列の一部または全部がわかっている場合は、複数の既知配列との比較を行なうことは必要とされ得ない。それどころか、既知配列の一部または全部を含む候補配列が適宜選択される。例えば、ＢＲＣＡ１およびＢＲＣＡ２遺伝子内の変異は、乳癌リスクの増大と関連していることがわかっており、被験体がかかる変異を保有している否かの決定において重要な関心となる。鋳型がＢＲＣＡ１遺伝子由来の配列を含むことがわかっている場合、例えば、該遺伝子の一部分を包含する目的の領域と隣接するプライマーが鋳型のクローン性の集団を作製するために使用された場合、正確な配列を決定するためには、候補配列を野生型または変異型ＢＲＣＡ１配列と比較するだけでよい。

【0258】

より一般的な場合では、候補配列を既知配列の組と比較することにより、任意の候補配列類似する任意の既知配列が同定される。ただし、候補配列は充分な長さのものであり、データベースが１つより多くの候補配列と同一のであるか非常に類似する配列を含む尤度は非常に小さいものとする。換言すると、候補配列が充分に長い場合、その１つより多くが既知配列の組に提示されることは考えにくい。候補配列は、「マッチング」しているとみなされる任意の配列と比較される。典型的には、マッチングが存在することが確立される必要とされる同一性の程度の閾値を設定することが望ましい。例えば、既知配列は、候補配列と既知配列が少なくとも５０％、少なくとも６０％、少なくとも７０％、少なくと

も８０％、少なくとも９０％、少なくとも９５％、少なくとも９９％またはまさに１００％同一である場合、マッチングしているとみなされ得る。典型的には、同一性割合は、少なくとも１０ヌクレオチド長、例えば１０～１５ヌクレオチド、１５～２０ヌクレオチド、２０～２５ヌクレオチド、２５～３０ヌクレオチドなどの領域で評価する。該領域の長さは、さまざまな異なる基準、例えば、限定されないが、複数の既知配列内の配列の数、複数の既知配列の正体または供給源などに応じて選択され得る。例えば、候補配列をＧｅｎＢａｎｋなどの大きなデータベースの配列と比較する場合、より少ない配列を含むデータベースが利用される場合は、長鎖の使用が望ましいことがあり得る。本発明のある特定の実施形態では、配列を、必ずしも互いに隣接していない複数の異なる領域において比較する。好ましくは、合わせた領域の長さは、少なくとも１０ヌクレオチド長、例えば１０～１５ヌクレオチド、１５～２０ヌクレオチド、２０～２５ヌクレオチド、２５～３０ヌクレオチドなどである。場合によっては、既知配列の組内の多数の配列がマッチングし得る。配列は、例えば、鋳型を誘導したものと同じ生物体に見られる相同遺伝子、異なる生物体由来の相同遺伝子、、偽遺伝子、ｃＤＮＡおよびゲノム配列などを示すものであり得る。

【0259】

一般に、既知配列の組内の配列にもっとも類似する候補配列が適宜選択される。あるいはまた、例えば、配列決定法が高エラー率に供されたかもしれないと思われる理由がある場合、適宜データベースから対応する配列を選択することが好ましいことがある。例えば、エラー率が所定の閾値より高いことがわかっている場合、正確な配列としてデータベースから配列を選択することが好ましいことがある。

【0260】

マッチングの尤度が多数の候補配列に見られることを確実にするために必要とされる長さは、さまざまな考慮事項、例えば、限定されないが、既知配列の具体的な組、許容されるマッチングの閾値などに依存する。一般に、典型的な生物体のゲノムにおいて、１回で提示される配列の長さは約２５～２６ヌクレオチドにすぎない。したがって、およそこの長さの候補配列の作製が、正確な配列を同定するのに充分である。一般に、候補配列は、少なくとも１０ヌクレオチド長、好ましくは少なくとも１５、少なくとも２０ヌクレオチド長、例えば、２０～２５、２５～３０、３０～３５、３５～４０、４５～５０またはさらに長鎖であるのがよい。

【0261】

第１の組の候補配列と第２の組の候補配列との比較による配列同定

本発明のある特定の実施形態において、デコード化は、第１のコード化に従ってコード化されたプローブファミリーの第１のコレクションを用いて第１のプローブファミリーの序列リスト作製し、これから第１の組の候補配列を作製し、次いで、第２のプローブファミリーの序列リストを同じ鋳型から、第２のコード化に従ってコード化されたプローブファミリーの第２のコレクションを用いて作製し、これから第２の組の候補配列を作製することにより行なわれる。新たに合成されたＤＮＡ鎖を２つの配列決定反応の間で鋳型から取り出すか、または同一の配列の鋳型を、プローブファミリーの第２のコレクションを用いて配列決定する。候補配列の組を比較する。どのプローブファミリーのコレクションが使用されるかとは無関係に、候補配列の１つが正確な配列であり、その他は正確でない（または、よくても部分的に正確である）ことは認識されよう。したがって、どの候補配列の組も正確な配列を含むが、ほとんどの場合、候補配列の任意の所与の組のその他の候補配列は、候補配列の別の組に見られるものと異なる。したがって、候補配列の２つの組を単純に比較することにより、正確な配列が決定され得る。２種類の異なるようにコード化されたプローブファミリーのコレクションを用いて等しい長さの候補配列を作製する必要はない。本発明の好ましい実施形態において、プローブファミリーの第２のコレクションを用いて作製された候補配列は２ヌクレオチドという短いものであり得、または同等に、プローブファミリーの第２のコレクションを用いて作製されたプローブファミリーの序列リストは、１要素（すなわち、ライゲーションおよび検出の１サイクル）という短いもの

であり得る。

【0262】

図３１Ａ～３１Ｃは、２種類の識別可能に標識された好ましいプローブファミリーを用いた候補配列の作製およびデコード化の一例を示す。図３１　Ａは、第１のコード化に従ってコード化されたプローブファミリーの好ましいコレクションを示す。図３１Ｂは、プローブファミリー黄色、緑色、赤色、青色（これは、「２３１４」と表示され得、ここで、赤色＝１、黄色＝２、グリーン＝３および青色＝４である）の序列リストからの４種類の候補配列の作製を示し、その正確な配列をＣＡＧＧＣ（太字で示す）と仮定している。図３１Ｃは、第２のコード化に従ってコード化されたプローブファミリーの好ましいコレクションを示す。鋳型内の第１のジヌクレオチドはＣＡであるため、黄色プローブファミリーの最上のプローブが伸長の第１のサイクルにおいて伸長可能な末端にライゲートされる。これにより、第１のジヌクレオチドに対して下記の組の候補配列：ＣＡ、ＴＣ、ＧＧ、ＡＴがもたらされる。第１のプローブファミリーのコレクションを用いて作製された候補配列のうち、配列ＣＡＧＧＣだけが、任意のこれらのジヌクレオチドで始まる。したがって、これは正確な配列のはずである。一般に、第１および第２のプローブファミリーのコレクションは、下記の基準：第１と第２のプローブファミリーのコレクションを比較する場合、（ｉ）第１のコレクション内の各プローブファミリーの４種類のプローブのうち３種類が第２のコレクションの新たなプローブファミリーに割り当てられるべきである；および（ｉｉ）３種の再割り当されたプローブの各々が第２のコレクションの異なるプローブファミリーに割り当てられるべきであるを満たすべきであることが好ましい。

【0263】

既知ヌクレオチド正体を用いたプローブファミリーの序列リストのデコード化

上記のように、候補配列は、伸長された二本鎖または鋳型内の単一のヌクレオチドの正体を仮定することにより作製され得る。使用される具体的なプローブファミリーコレクションに応じて、一般的に、少なくとも４種類の候補配列を作製することが必要である。しかしながら、多数の候補配列の作製は、鋳型内の少なくとも１つのヌクレオチドの正体がわかっている場合は（したがって、伸長された二本鎖の場合も）回避され得る。その場合は、単一の候補配列を作製するだけでよい。候補配列を作製するための方法は、上記のものと同一である。鋳型内の少なくとも１つのヌクレオチドの正体は、任意の配列決定法、例えば、限定されないが、配列決定法Ａ、１組の識別可能に標識されたヌクレオチドおよびポリメラーゼを用いる初期オリゴヌクレオチドからのプライマー伸長などを用いて決定され得る。鋳型内の１つ以上のヌクレオチドが、まず、配列決定法ＡＢ以外の配列決定法を用いて配列決定され得、次いで、初期オリゴヌクレオチドおよび任意の伸長産物が除去され得、同じ鋳型が配列決定法ＡＢを用いる配列決定に供され得ることは認識されよう（または逆も同様）。

【0264】

別のアプローチは、その配列が決定される部分に加えて既知の正体の１つ以上の既知ヌクレオチドを含む鋳型を単純に配列決定することである。例えば、鋳型の初期オリゴヌクレオチドが結合する領域と未知配列が始まる領域との間の部分に、１つ以上の既知の正体のヌクレオチドが含まれ得る。鋳型のこの部分を配列決定法ＡＢに供することにより、配列内の１つ以上のヌクレオチドの正体が予め決定され、したがって、単一の候補配列を作製するために使用され得、これは正確な配列である。

【0265】

したがって、上記の方法は、（ｉ）どの正体が既知ヌクレオチドの正体と一致するか、およびその近位ヌクレオチドが、既知の正体のヌクレオチドに隣接するヌクレオチドに対向してライゲートされたプローブの拘束部分の可能な配列を調べることにより、正体を、既知の正体のヌクレオチドに隣接する鋳型内のヌクレオチドに割り当てる工程；（ｉｉ）

どの正体が、その近位ヌクレオチドが後続のヌクレオチドに対向してライゲートされたプローブの拘束部分の可能な配列と一致するかを調べることにより、正体を後続のヌクレオチドに割り当てる工程；ならびに（ｉｉｉ）　配列が決定されるまで工程（ｉｉ）を反

復する工程を含む。伸長された二本鎖と鋳型の配列決定される領域との間に正確な対応があるため、これらの工程は、同じ工程を伸長された二本鎖において行なうことと等価であることを理解されたい。

【0266】

やや好ましいプローブファミリーを用いる配列決定

やや好ましいプローブファミリーのコレクションは、配列決定法ＡＢを、好ましいプローブファミリーのコレクションが使用される様式と同様にして行なうために使用され得る。しかしながら、結果はいくつかの点で異なり得る。例えば、配列のある特定部分は候補配列から、さらなる情報を必要せずに完全に同定され得る。図３２は、図２８に示すようにしてコード化されたプローブファミリーのやや好ましいコレクションを用いる配列決定の一例を示す。配列の決定は、一般的に、好ましいプローブファミリーのコレクションで記載したようにして行なう。目的の鋳型は配列「ＧＣＡＴＧＡ」を有し、これは、プローブファミリーの序列リストでは「１２３４１」となる。１位のヌクレオチドをＡと仮定すると、「ＡＣＡＴＧＡ」が候補配列として得られる。しかしながら、好ましいプローブファミリーのコレクションの場合とは異なり、標識「１」は、第１のヌクレオチドとしてＡを有する２つの異なるジヌクレオチド、すなわち「ＡＡ」および「ＡＧ」と関連するため、第２のヌクレオチドに対して２つの可能性がある。したがって、１位のヌクレオチドがＡであると仮定すること、「ＡＣＡＴＧＣ」が第２の候補配列として得られる。１位のヌクレオチドがＧであると仮定すると、「ＧＣＡＴＧＡ」が候補配列として得られ、また、「ＧＣＡＴＧＣ」が候補配列として得られる。標識「１」は、ＣまたはＴを１位に有するジヌクレオチドとはなんら関連しないため、「Ｃ」または「Ｔ」で始まる候補配列は得られない。図３２は、互いにアライメントさせた４つの候補配列を示す。すべての候補配列の中央の４つのヌクレオチドがＣＡＴＧであることが観察される。したがって、正確な配列は、ＣＡＴＧを２～５位に含むものであるはずである。もし、これらのヌクレオチドが目的のものでない場合は、さらなるデコード工程を行なう必要はない。

【0267】

上記のように、プローブファミリーのコレクションは、４種類の異なるプローブファミリーからなるものである必要はく、２つよりより多く4^N（ここで、Ｎは拘束部分の長さである）までの任意の数からなるものであり得る。しかしながら、４つより少ないファミリーを用いる場合、４つより多くの候補配列作製することが必要であり得、一方、４つより多くのプローブファミリーを用いる場合は、さらなる標識が必要とされる。これらおよび他の理由により、４つのプローブファミリーからなるコレクションが好ましい。

【0268】

候補配列を互いに比較することによる配列同定

本発明のある特定の実施形態において、目的の配列の一部または全部は、候補配列を互いに比較することにより決定され得る。一般に、かかる比較は、その全長において候補配列のどれが正確であるかを決定するのに充分でないことがあり得る。しかしながら、候補配列の２つ以上が該配列の一部分において同一であるか充分類似している場合、この情報は、鋳型の上記の部分内のヌクレオチドの配列を明確に同定するのに充分であり得る。

【0269】

所望の場合、鋳型は、同定された配列を有するさらなる部分を得るため、さらに１回以上、別のコード化されたプローブファミリーを用いて配列決定され得る。これらの部分を組み合わせ、所望の長さの配列が合成され得る。

【0270】

プローブファミリーを用いたエラー補正。多くの場合、同じＤＮＡ　配列の一部または全部を提示する多数の鋳型を配列決定し、該配列をアライメントすることが望ましい。鋳型が目的の領域の一部しか含まない場合、重複断片を合成することによって、より長い配列を得る。例えば、生物体のゲノムを配列決定する場合、典型的には、ＤＮＡを断片化し、充分な断片を、各ＤＮＡ鎖がいくつかの（例えば、４～１２個の）異なる断片内に提示されるように配列決定する。重複配列を合成してより長い配列にするためのコンピュータ

ソフトウエアが当業者に知られている。

【0271】

慣用的な配列決定法が用いられる場合、ほとんどの場合、ある領域において、断片の１つ（異常断片という）がその他のものと該領域内の単一の位置で異なる場合以外は、多数の断片が完璧にアライメントされる。孤立性の差が配列決定エラーを表すか否か、または真性の差（例えば、一塩基多型）が該位置に存在するか否かを調べることは、問題となり得る。

【0272】

本発明は、配列決定法ＡＢを用いてエラーチェックを行なうこと新規な方法を提供する。該方法によれば、同じＤＮＡ鎖を示す断片を含む鋳型が、上記のような識別可能に標識されたプローブファミリーのコレクションを用いて配列決定され、各鋳型のプローブファミリーの序列リストがもたらされる。プローブファミリーの序列リストをアライメントする。いくつかのリストが、単一の位置でその他の断片異なる１つのリスト以外、リスト内の所定の長さ、例えば、１０、１５、２０または２５それ以上の要素において完璧にアライメントされたら、その差は、配列決定エラーに起因する。実際に多型が存在する場合、異常断片から得られた序列プローブリストは、２つ以上の隣接する位置で、その他の断片から得られた序列プローブリストと異なる。

【0273】

例えば、表１のコード化４を用いたプローブファミリーの好ましいコレクションを使用する配列決定法ＡＢを、配列５’－ＣＡＧＡＣＧＡＣＡＡＧＴＡＴＡＡＴＧ－３’を含む鋳型に適用すると、下記のプローブファミリーの序列リスト：「２３３２４３２２１３２４４４１４２」が得られ、これを以下に示す。

【0274】

【化1】

```
 2332432213244414 2
CAGACGACAAGTATAATG
```

実際にＳＮＰが存在する場合（例えば、ＣＡＧＡＣＧＡＧＡＡＧＴＡＴＡＡＴＧ、式中、下線のヌクレオチが多型性部位を表す）、リスト内の２つの連続する要素に変化がもたらされる：２３３２４３３３１３２４４４１４２式中、下線は、ＳＮＰの結果として生じた変化を示す。プローブファミリーの序列リストとＳＮＰを含む配列間の対応を以下に示す。

【0275】

【化2】

```
 2332433313244414 2

CAGACGAGAAGTATAATG
```

しかしながら、ライゲートされた伸長プローブと会合した標識の同定におけるエラーは、プローブファミリーの序列リストにおける単一のエラーおよびその点の前方（ｆｏｒｗａｒｄ）から得られる候補配列における変化をもたらす。例えば、７回目にライゲートされ伸長プローブ２３３２４３３２１３２４４４１４２（式中、下線の数字は誤同定された標識を表す）と会合した標識の決定におけるエラーによって、得られる候補配列はＣＡＧＡＣＧＡＧＴＴＣＡＴＡＴＴＡＣに変化する（式中、下線部分は、配列決定エラーの結果として生じた変化を示す）。プローブファミリーの序列リストと配列間の対応を以下に示す。

【0276】

【化3】

233243321324441 42

CAGACGAGTTCATATTAC

３塩基、４標識スキームを用いる場合、ＳＮＰを含む断片により、異常断片のプローブファミリーの序列リストにおいて３連続の差がもたらされるが、配列決定エラーでは、１つの差しかもたらされない。例えば、図２９に示すようにしてコード化されたプローブファミリーのコレクションを用いる場合、配列ＣＡＧＡＣＧＡＣＡＡＧＴＡＴＡＡＴＧのプローブファミリー正体の序列リストを以下に示す。

【0277】

【化4】

2322224132412244

CAGACGACAAGTATAATG

ＳＮＰを含む異常断片、例えば、ＣＡＧＡＣＧＡＧＡＡＧＴＡＴＡＡＴＧは、ＳＮＰを含まない断片から得られる序列リストと比べて３つの連続する位置が異なるプローブファミリーの序列リストをもたらし得、これを以下に示す。

【0278】

【化5】

2322213332412244

CAGACGAGAAGTATAATG

配列決定エラーは、たった１つの差をプローブファミリーの序列リスト内にもたらし得、エラー点の前方と完全に異なる候補配列が作製され得る。

【0279】

したがって、断片（異常断片）から作製されたプローブファミリーの序列リストを、同じＤＮＡ鎖を示すが、他方の序列リストと単一の孤立性の位置で異なる他の断片から作製されたプローブファミリーの序列リストとアライメントさせる場合、差を含む序列リストが配列決定エラーを表すことがあり得る（プローブファミリーの誤同定）。断片（異常断片）から作製されたプローブファミリーの序列リストを、同じＤＮＡ鎖を示すが、他方の序列リストと２つ以上の連続する位置で異なる他の断片から作製されたプローブファミリーの序列リストとアライメントさせる場合、異常断片がＳＮＰを含むものであることがあり得る。好ましくは、プローブファミリーの序列リストのアライメントされる部分は、少なくとも３または４要素の長さ、好ましくは少なくとも６、８またはそれ以上の要素の長さである。好ましくは、アライメントされる部分は、少なくとも６６％同一、少なくとも７０％同一、少なくとも８０％同一、少なくとも９０％同一、またはそれ以上、例えば１００％同一である。

【0280】

同様に、断片の候補配列を、配列の第１の部分において同じＤＮＡ鎖を示すが、他の断片の候補配列と配列の第２の部分において実質的に異なる他の断片の候補配列とアライメントさせる場合、配列決定エラーが起こることがあり得る。ある断片の候補配列を、配列の２つの部分において同じＤＮＡ鎖を示すが単一の位置で異なる他の断片の候補配列とアライメントさせる場合、異常断片がＳＮＰを含むものであることがあり得る。好ましくは、候補配列のアライメントされる部分は、少なくとも４ヌクレオチド長である。好ましくは、アライメントされる部分は、少なくとも６６％同一、少なくとも７０％同一、少なくとも８０％同一、少なくとも９０％同一、またはそれ以上、例えば１００％同一である。

【0281】

したがって、本発明は、（ａ）配列決定法ＡＢを用いて複数の鋳型を配列決定する工程、ここで、鋳型は、単一の核酸配列の重複断片を表す；（ｂ）工程（ａ）で得られた配列アライメントする工程；および（ｃ）配列が第１の部分と全体的に実質的に同一であり、第２の部分と実質的に異なる場合（各部分は少なくとも３ヌクレオチド長を有する）、配

列間の差が配列決定エラーを表すと決定する工程を含む、一塩基多型を配列決定エラーと識別する方法を提供する。本発明は、さらに、（ａ）単一の核酸配列の重複断片を表す複数の鋳型を用いて配列決定法ＡＢを行なうことにより、プローブファミリーの複数の序列リスト得る工程；（ｂ）工程（ａ）で得られたプローブファミリーの序列リストをアライメントし、該リストが少なくとも９０％同一であるアライメント領域を得る工程；および（ｃ）該リストがアライメント領域内のたった１つの位置で異なる場合、プローブファミリーの序列リスト間の差は、配列決定エラーを表すと決定する工程；または（ｄ）該リストがアライメント領域内の２つ以上の隣接する位置で異なる場合、プローブファミリーの序列リスト間の差は一塩基多型を表すと決定する工程を含む、一塩基多型を配列決定エラーと識別する方法を提供する。

【0282】

非局在化情報収集

当該技術分野でよく知られているように、「ビット」（２進数）は、基数を２とする、換言すると、１または０のいずれかである１桁をいい、デジタルデータの最小単位を表す。ヌクレオチドは任意の４種類の異なる正体を有し得るため、ヌクレオチドの正体の指定には２ビットが必要であることは認識されよう。例えば、Ａ、Ｇ、ＣおよびＴは、それぞれ００、０１、１０および１１で表され得る。識別可能に標識されたプローブファミリーの好ましいコレクションにおけるプローブファミリー名の指定では、４つの識別可能に標識されたプローブファミリーがあるため、２ビットが必要である。

【0283】

最も慣用的な配列決定の形態および配列決定法Ａでは、各ヌクレオチドは、別個の単位として同定され、１つのヌクレオチドに対応する情報が一度に収集される。各検出工程では、２ビットの情報が単一のヌクレオチドから取得される。対照的に、配列決定法ＡＢでは、各検出工程において、２ビット未満の情報が複数のヌクレオチドの各々から取得されるが、プローブファミリーの好ましいコレクションが使用される場合は、やはり１回の検出工程で２ビットの情報が取得される。プローブファミリーの序列リスト内の各プローブファミリー名は、鋳型内の少なくとも２つのヌクレオチドの正体を表し、その正確な数は、プローブの配列決定部分の長さによって決定される。例えば、配列５’－ＣＡＧＡＣＧＡＣＡＡＧＴＡＴＡＡＴＧ－Ｓ’から、表１のコード化４に従ってコード化されたプローブファミリーのコレクションを用いて得られるプローブファミリーの序列リストを考察されたい。

【0284】

【化6】

2332432213244414 2
CAGACGACAAGTATAATG

プローブファミリー２は、ジヌクレオチドＣＡが、プローブファミリー２のプローブ内に存在する指定される部分の１つであるため、リスト内の第１のプローブファミリーである。プローブファミリー３は、ジヌクレオチドＡＧが、プローブファミリー３のプローブ内に存在する指定される部分の１つであるため、リスト内の第２のプローブファミリーである。上記のように、４つのプローブファミリーがあるため、各プローブファミリー正体は、２ビットの情報で表される。したがって、各検出工程では、２つのヌクレオチドに関する２ビットの情報が収集され、平均１ビットの情報が各ヌクレオチドからもたらされる。

【0285】

したがって、本発明は、多数の伸長、ライゲーションおよび検出のサイクルを含む、配列決定方法であって、検出工程が、平均２ビットの情報を同時に、任意の個々のヌクレオチドの２ビットの情報を取得することなく、鋳型内の少なくとも２つのヌクレオチドの各々から取得することを含む配列決定方法を提供する。本発明は、さらに、（ａ）伸長、ライゲーション、検出および切断の連続的サイクルを行なう工程、ここで、検出工程が、平

均２ビットの情報を同時に、任意の個々のヌクレオチドの２ビットの情報を取得することなく、鋳型内の少なくとも２つのヌクレオチドの各々から取得することを含む；ならびに（ｂ）工程（ａ）で得られた情報を少なくとも１ビットのさらなる情報と組み合わせて配列を決定する工程を含む、オリゴヌクレオチドプローブファミリーの第１のコレクションを用いて、鋳型ポリヌクレオチド内のヌクレオチドの配列を決定する方法を提供する。種々の本発明の実施形態において、少なくとも１ビットのさらなる情報は、鋳型内のヌクレオチドの正体、候補配列を少なくとも１つの既知配列と比較することにより得られる情報；およびオリゴヌクレオチドプローブファミリーの第２のコレクションを用いて該方法を反復することにより得られる情報からなる群より選択される項目を含む。

【0286】

したがって、該方法では、個々のヌクレオチドから２ビットの情報は取得されないが、平均２ビットの情報が鋳型から各サイクルで収集され、プローブファミリーの好ましいコレクションが使用される場合は、非局在化様式で収集される。２または３つのプローブファミリーのコレクションが使用される場合は、２ビット未満の情報が各サイクル中で収集される。

【0287】

非局在化情報の収集はいくつかの利点を有し、例えば、上記のものなどのエラーチェック法の適用が可能になる。また、各鋳型内のヌクレオチドは、好ましい実施形態では、非局在化情報の収集によって、１回より多くのインテロゲーションが行なわれる（ｉｎｔｅｒｒｏｇａｔｅ）ため、特定のヌクレオチドと会合したフルオロフォアの検出における系統的偏重（ｓｙｓｔｅｍａｔｉｃ　ｂｉａｓ）の回避が補助され得る。

【0288】

本明細書に記載のプローブファミリーおよびプローブファミリーのコレクションは、プローブの伸長、ライゲーションおよび切断の連続的サイクルを伴う方法に加え、さまざまな配列決定法に使用され得る。本発明はまた、上記のような配列および構造を有し、プローブが任意選択で切れやすい連結を含まないプローブファミリーおよびプローブファミリーのコレクションを提供する。例えば、プローブは、ホスホジエステル主鎖連結のみを含むものであり得、および／またはトリガー残基含まないものであり得る。本発明のある一部の実施形態において、プローブファミリーは、伸長およびライゲーションの連続的サイクルを用いるが、各サイクル中で切断を伴わない配列決定を行うために使用される。例えば、プローブファミリーは、ＷＯ２００５０２１７８６および当該技術分野において他に記載されたものなどのライゲーション系の方法に使用され得る。プローブファミリーをかかる方法で使用するためには、プローブ上の標識は、例えば、ＷＯ２００５０２１７８６に開示のように、核酸の切れやすい連結を切断することなく除去され得るように、切断可能なリンカーによって結合されているのがよい。かかる方法は、例えば、ＷＯ２００５０２１７８６に記載されたライゲーションカセットではなくプローブファミリーを用い、多数の反応を並行して、または連続的に行ない、次いで、プローブファミリーのリストを合成することにより、プローブファミリーの序列リスト作製するために使用され得る。リストは、上記のようにしてデコード化される。

【0289】

Ｉ．キット

本発明の種々の実施形態を行なためのさまざまなキットが提供され得る。キットのある特定のものは、ホスホロチオレート結合を含む伸長オリゴヌクレオチドプローブを含む。キットは、１種類以上の初期オリゴヌクレオチドをさらに含むものであり得る。キットは、ホスホロチオレート連結の切断に適した切断試薬（例えば、$AgNO_3$）および該切断を行なうための適切なバッファーを含むものであり得る。キットのある特定のものは、トリガー残基、例えば、損傷塩基または無塩基残基を含有するヌクレオシドを含む伸長オリゴヌクレオチドプローブを含む。キットは、１種類以上の初期オリゴヌクレオチドをさらに含むものであり得る。キットは、ヌクレオシドと隣接する無塩基残基間の連結の切断に適した切断試薬および／または損傷塩基をポリヌクレオチドから除去するのに適した試薬

（例えば、ＤＮＡグリコシラーゼ）を含むものであり得る。ある特定のキットは、二糖ヌクレオチドを含むオリゴヌクレオチドプローブを含み、切断試薬として過ヨウ素酸塩を含む。ある特定の実施形態において、キットは、識別可能に標識されたオリゴヌクレオチドプローブファミリーのコレクションを含む。

【0290】

キットは、ライゲーション試薬（例えば、リガーゼ、バッファーなど）および具体的な本発明の実施形態を行なうための取扱い説明書を含むものであり得る。使用され得るその他の酵素、例えば、ホスファターゼ、ポリメラーゼに適切なバッファーを含めてもよい。場合によっては、これらのバッファーは同一であり得る。キットはまた、鋳型を固定するための支持体、例えば、磁気ビーズを含むものであり得る。ビーズは、ＰＣＲ増幅を行なうためにプライマーにより官能性付与されていてもよい。他の任意選択的な成分としては、洗浄溶液；ＰＣＲ増幅のための鋳型の挿入用ベクター；ＰＣＲ試薬、例えば、増幅プライマー、熱安定性ポリメラーゼ、ヌクレオチドなど；エマルジョン調製のための試薬；ゲル調製のための試薬などが挙げられる。

【0291】

ある特定の好ましいキットには、ホスホロチオレート結合を含む蛍光標識されたオリゴヌクレオチドプローブが、該プローブの異なる末端ヌクレオチドに対応するプローブが異なるスペクトルにより分離可能な蛍光染料を担持するように提供される。より好ましくは、４種類のかかるプローブが提供され、これは、４種類のスペクトルにより分離可能な蛍光染料の各々と、プローブの４種類の可能な末端ヌクレオチドとの間の１対１対応を可能にする。

【0292】

キットは、対合末端または断片ライブラリーを生成するのに好適なオリゴヌクレオチドおよび／またはベクターを含有する場合がある。キットは、ライブラリーのメンバーである鋳型分子の相補的な共通部分である１つ以上のブロッキングオリゴヌクレオチドを含有する場合がある。

【0293】

識別物、例えば、バーコード、高周波ＩＤタグなどをキット内またはキット上に存在させてもよい。識別物は、例えば、品質管理、在庫管理、追跡、ワークステーション間の移動などの目的でキットを固有に識別するために使用され得る。

【0294】

キットは、一般的に、ある種の個々の試薬が個別に収容され得るように、１つ以上の容器（ｖｅｓｓｅｌ）または収容器（ｃｏｎｔａｉｎｅｒ）を含む。キットはまた、個々の収容器を、市販用に比較的コンパクトに（ｃｌｏｓｅ　ｃｏｎｆｉｎｅｍｅｎｔ）封入するための手段、例えば、取扱い説明書、発泡スチレンなどのパッケージング材料などが封入され得るプラスチックボックスを含むものであり得る。

【0295】

Ｊ．Ｐａｒａｌｌｅｌ配列決定および自動配列決定システム

Ｍａｃｅｖｉｃｚは、特定の配列を有する単一鋳型種の配列決定を開示しているが、種々の配列を有する複数の鋳型の同時配列決定と並行してこの方法を実施する可能性については述べていない。本発明者等は、高いスループットで配列決定を効率的に実施するためには、上述のように、複数の担体（例えば、ビーズ）を、それぞれの担体に特定の配列の鋳型が付着するように調製し、それぞれの担体に付着した鋳型に対して、本明細書に記載の方法を同時に実施することが望ましいことを認識した。本手法のある実施形態において、複数のこのような担体は、スライドなどの平面の基材の中または上に配列される。ある実施形態において、担体は、ゲルなどの半固形培地の中または上に配列される。担体は、無作為に配列される場合があり、すなわち、基材上の各担体の位置は既定されていない。担体は、規則的な間隔を置いて位置する必要もなければ、列や行などの規則正しい配置に配置する必要もない。好ましくは、担体の多くまたはほとんどから個々のシグナルを検出することが可能な密度で、担体が配列される。ある好ましい実施形態において、担体は、

主として単一の焦点面に分布する。同じ配列の鋳型が付着した複数の担体が、例えば品質管理の目的で含まれる場合もある。配列決定反応は、担体のそれぞれに付着した鋳型において、並行して実施される。

【0296】

シグナルは、種々のイメージング様式を含めた種々の手段のいずれかを使用して収集される場合がある。好ましくは、検出前に基材（例えば、基材上に配置された半固形担体に埋め込まれたビーズ）上に配列される微粒子に対して配列決定が実施される実施形態の場合、画像化装置の分解能は１μｍ以下である。例えば、ＣＣＤカメラを取り付けた走査型顕微鏡、または十分な分解能を有するマイクロアレイスキャナーを使用できると考えられる。あるいは、蛍光検出のために装備される顕微鏡に取り付けたフローセルまたは流体工学ワークステーションにビーズを通過させることもできる。その他のシグナル収集方法には光ファイバー束が含まれる。適切な画像収集処理ソフトウェアが使用される場合もある。

【0297】

本発明のある実施形態において、配列決定はマイクロ流体デバイスで行われる。例えば、鋳型が付着したビーズをデバイスに充填し、その中に試薬を流す場合がある。例えばＰＣＲを使用した鋳型合成もデバイスで行うことができる。米国特許第６，６３２，６５５号では、好適なマイクロ流体デバイスの例が記載されている。

【0298】

本発明は、配列情報を複数の鋳型から並行して、すなわち、実質的に同時に収集するために使用され得るさまざまな自動配列決定システムを提供する。好ましくは、鋳型は、実質的に平面状の基材上でアレイ状である。図２１は、本発明のシステムの一例の写真を示す。上側の写真に示すように、本発明のシステムは、ＣＣＤ　カメラ、蛍光顕微鏡、可動ステージ、Ｐｅｌｔｉｅｒフローセル、温度制御装置、液体操作装置および専用のコンピュータを備える。これらの構成要素の種々の置き換えがなされ得ることは認識されよう。例えば、代替的に画像捕捉装置が使用され得る。このシステムのさらなる詳細は実施例９に示す。

【0299】

本発明の自動配列決定システムならびに関連する画像処理方法およびソフトウエアは、さまざまな配列決定法、例えば、本明細書に記載のライゲーション系の方法ならびに他の方法、例えば、限定されないが、合成方法による配列決定、例えば、合成による蛍光インサイチュで配列決定（ＦＩＳＳＥＱ）など（例えば、を参照　Ｍｉｔｒａ　ＲＤら．，Ａｎａｌ　Ｂｉｏｃｈｅｍ．，３２０（１）：５５－６５，２００３）の両方を行うために使用され得ることは認識されよう。本明細書に記載のライゲーション系の配列決定法の場合のように、ＦＩＳＳＥＱは、半固相支持体内または該支持体上に直接固定化された鋳型上、半固相支持体内または該支持体上の微粒子上に固定化された鋳型上、基材に直接結合された鋳型上などで行なわれ得る。

【0300】

本発明のシステムの重要な態様の１つは、フローセルである。一般に、フローセルは、内部を流体が流動し得る流入ポートおよび流出ポートを有するチャンバを含む。種々のフローセルならびにその製造のための材料および方法の論考については、例えば、米国特許第６，４０６，８４８号および同第６，６５４，５０５号ならびにＰＣＴ公開公報ＷＯ９８０５３３００を参照のこと。流体の流動により、種々の試薬をフローセル内に位置する実体（例えば、鋳型、微粒子、解析物など）に添加し、これらから除去することが可能になる。

【0301】

好ましくは、本発明の配列決定システムにおける使用に適したフローセルは、流体がその表面を流動するような基材、例えば、スライドなどの実質的に平面状の基材が取り付けられる位置、および照光、励起、シグナル取得などを可能にする窓を備える。本発明の方法によれば、微粒子などの実体は、典型的にはフローセル内に配置する前に、基材上にア

レイ状にする。

【0302】

本発明のある特定の実施形態において、フローセルは垂直に配向させ、これにより、気泡をフローセルの上面から散逸させることが可能になる。フローセルは、流体経路がフローセルの下部から上に流れるように、例えば、流入ポートがセルの下部に存在し、流出ポートがセルの上に存在するように配設する。導入され得る気泡は浮揚性であるため、これは、照光窓を障害することなく速やかに流出ポートに浮揚する。気泡を液体の表面まで、その密度が該液体のものより低いことによって上昇させるこのアプローチを、本明細書では「重量測定気泡移動」という。したがって、本発明は、重量測定気泡移動が可能になるように配向させたフローセルを含む配列決定システムを含むを提供する。好ましくは、自身に直接もしくは間接的に結合された微粒子（例えば、共有結合もしくは非共有結合により基材に結合）、または基材に接着もしくは固定された半固相支持体内または該支持体上に固定化された微粒子を有する基材をフローセル内に垂直になるように、すなわち、基材の最大の平面状の表面が、設置面に対して垂直になるように取り付ける。好ましい実施形態では、微粒子は、支持体または基材内またはその上面に固定化されているため、互いに対して実質的に固定された位置に維持され、これにより逐次的な画像取得と画像レジストレーションが容易になる。

【0303】

図２４Ａ～Ｊは、種々の向きの本発明のフローセルまたはその一部分の概略図を示す。本発明のフローセルは、任意のさまざまな目的、例えば、限定されないが、解析方法（例えば、配列決定、ハイブリダイゼーションアッセイなどの核酸解析法；タンパク質解析法、結合アッセイ、スクリーニングアッセイなどに使用され得る。フローセルはまた、合成を行うため、例えば、コンビナトリアルライブラリーを作製するためなどに使用され得る。

【0304】

図２２は、別の本発明の自動配列決定システムの概略図を示す。フローセルは、自動温度制御ステージ（実施例９に記載のものと類似）上に取り付けられ、液体操作システム（例えば、マルチポートバルブを有するシリンジポンプなど）に接続されている。該ステージは、他の工程、例えば、伸長、ライゲーションおよび切断などが別のフローセルで行なわれている間に、１つのフローセルが画像化されるのを可能にするために、多数のフローセルを収容する。このアプローチは、高価な光学システム　の利用を最大限にするとともに、処理量を増大させる。

【0305】

流体ラインには、気泡を検出ため、および試薬の使用をモニターするための光学および／または伝導性センサーが設置されている。流体工学システム内の温度制御およびセンサーにより、試薬が、長時間の安定に適切な温度に維持されるが、フローセル内に侵入するとき作業温度またはで上昇せず、アニーリング、ライゲーションおよび切断工程中での温度変動が回避されることが確保される。試薬は、好ましくは、誤負荷を回避するためキット内に予め装填しておく。

【0306】

光学装置としては、４つのカメラが含まれ、各々が、４つのフィルターセットの１つから１つの画像を撮影する。光退色性効果を低減させるため、照光光学装置は、画像化される領域だけに照光されるように、視野周辺の多重照光が回避されるように設計されたものであり得る。画像化光学装置は、標準的な無限遠補正顕微鏡対物レンズおよび標準的なビームスプリッタおよびフィルターから構成されたものであり得る。標準的な２，０００×２，０００ピクセルＣＣＤカメラは、画像を取得するために使用され得る。該システムには、光学装置のための適切な機械的支持体が組み込まれれている。照光強度は、好ましくは、後の解析ソフトウエアによる使用のためにモニターされ、記録される。

【0307】

複数の画像（例えば、代表的な実施形態において、およそ１８００以上の非重複画像フ

ィールド）を速やかに取得するため、該システムには、好ましくは、高速オートフォーカスシステムを用いる。画像自体の解析に基づくオートフォーカスシステムは、当該技術分野でよく知られている。これは、一般的に、１回のフォーカシング事象に少なくとも５フレームを必要とする。これは、フォーカシング画像を取得するために余分な照光が必要とされるという点で、低速および高価の両方である（光退色性が増大する）。本発明のある特定の実施形態において、択一的なオートフォーカスシステム、例えば、該機械的システムが応答し得るのと同等に高速にフォーカシングし得る独立した光学装置に基づくシステムが使用される。かかるシステムは当該技術分野で知られており、例えば、コンシューマＣＤプレイヤーに用いられるフォーカシングシステム（これは、ＣＤ回転時にリアルタイムでサブミクロンフォーカシングを維持する）が挙げられる。

【0308】

本発明のある特定の実施形態において、該システムは、遠隔操作される。特定のプロトコルを実行するためのスクリプトは、セントラルデータベース内に保存し、各配列決定実行のためにダウンロードされ得る。試料は、バーコードを付けて、試料追跡の完全性および試料と最終データとの関連付けが維持され得る。リアルタイムのセントラルモニタリングにより、プロセスエラーの迅速な解決が可能になる。ある特定の実施形態において、機器によって収集された画像は、直ちにセントラルマルチテラバイト保存システムおよび１つ以上のプロセッサバンクにアップロードされる。セントラルデータベースからの追跡データを用い、プロセッサ（１つまたは複数）により画像が解析され、配列データが作製され、任意選択で、例えば、機器の性能を調節するために、バックグラウンド蛍光レベルおよびビーズ密度などのメトリクスが処理される。

【0309】

制御ソフトウエアは、ポンプ、ステージ、カメラ、フィルター、温度制御を適正にシーケンス処理するため、ならびに画像データの注記付けおよび保存のために使用される。ユーザーインターフェースは、例えば、作業者が機器をセットアップし、維持するのを補助するために設けられ、好ましくは、スライドの負荷／取り出しおよび流体ラインの準備（ｐｒｉｍｉｎｇ）のためのステージを位置合わせする機能を含むものである。例えば、作業者に、種々の実行パラメータ、例えば、温度、ステージ位置、光学フィルター構成の現状、実行プロトコルの状態などを示すためのディスプレイ機能を含め得る。好ましくは、試薬ロットおよび試料ＩＤなどの追跡データを記録するデータベースへのインターフェースを含める。

【0310】

Ｋ．画像およびデータ処理方法

本発明は、少なくとも一部は、コンピュータ可読媒体上に保存されたコンピュータコード（すなわち、ソフトウエア）として行なわれ得るさまざまな画像およびデータ処理方法を提供する。さらなる詳細は、実施例９および１０に示す。また、一般に、配列決定法ＡおよびＢでは、ともに、一般的に、例えば、多重配列決定反応で収集されたデータの追跡管理、かかるデータの合成、候補配列の作製、配列比較の実施などを伴う該処理工程を行なうための適切なコンピュータソフトウエアが使用される。

【0311】

Ｌ．配列情報を保存するコンピュータ可読媒体

また、本発明は、本発明の配列決定法を適用することにより得られた情報を保存するコンピュータ可読媒体を提供する。情報としては、生データ（すなわち、さらなる処理または解析がされていないデータ）、処理または解析されデータなどが挙げられる。データには、画像、数値も含まれる。情報は、典型的には、検索が容易なように配設された、例えば、コンピュータメモリ内に保存されたデータベース、すなわち、情報（例えば、データ）のコレクションに保存され得る。情報としては、例えば、配列および配列に関する任意の情報（例えば、部分配列）、配列と参照配列との比較、配列解析の結果、ゲノム情報、例えば、多型情報など（例えば、ある特定の鋳型が多型を含むか否か）または変異情報など、連鎖情報（すなわち、例えば、染色体内での核酸配列の別の核酸配列に対する物理的

な位置に関する情報）、疾患関連情報（すなわち、疾患の存在または疾患に対する感受性を、被験体の身体的形質、例えば、被験体の対立遺伝子と相関させる情報）などが挙げられる。情報は、試料ＩＤ、被験体ＩＤなどと関連したものであり得る。試料、被験体などに関するさらなる情報は、例えば、限定されないが、試料の供給源、試料において行なわれた処理工程、情報の解釈、試料または被験体の供給源などに含まれ得る。本発明はまた、前述の任意の情報をコンピュータ可読形式に受信する、例えば、コンピュータ可読媒体に保存することを含む方法を含む。該方法は、該情報にに基づく診断的、予後的もしくは予測的情報を提供する工程、または第３者に、好ましくはコンピュータ可読媒体に保存された情報を単に提供する工程をさらに含み得る。

【0312】

以下の実施例は例示の目的で提供され、本発明を限定することは意図されない。

【0313】

（実施例１：ホスホロチオレート含有オリゴヌクレオチドの効率的な切断およびライゲーション）

この実施例では、３’－Ｓホスホロチオレート結合を含有する伸長オリゴヌクレオチドの効率的なライゲーションおよび切断を示す実験を記載する。

【0314】

材料および方法

ライゲーション配列決定プロトコル

鋳型調製：サイクル的オリゴヌクレオチドライゲーションおよび切断による配列決定の可能性の評価を示すため、および該方法のある特定の態様の変形の効果を検討するため、２組のモデルビーズ系鋳型集団を調製した。実施例に記載したような好ましい実施では、サイクル的オリゴヌクレオチドライゲーションおよび切断により、鎖は３’→５’方向に伸長される。したがって、ライゲーション効率を評価するため、モデル鋳型を、ビーズに５’末端において結合させ、同じ結合領域を３’末端に有するように設計した。１組は、二重ビオチン部分によってストレプトアビジンコート磁気ビーズ（１ミクロン）に結合させた短鎖（７０ｂｐ）オリゴヌクレオチドで構成されたものであった。これらの短鎖鋳型集団の各々は、同一のプライマー結合領域（４０ｂｐ）および特有の配列領域（３０ｂｐ）を３’末端に有するように設計した。この短鎖オリゴヌクレオチド鋳型集団をライゲーション配列決定鋳型１～７（ＬＳＴ１～７）と命名した。

【0315】

第２の組のビーズ系鋳型集団は、１８３ｂｐのスペーサー配列（ヒトｐ５３エキソン由来）を各鋳型集団内に挿入することにより誘導した長鎖のＰＣＲ生成ＤＮＡ断片（２３２ｂｐ）から設計した。鋳型は、二重ビオチン含有フォワードプライマー、および短鎖鋳型集団と同じ３０塩基の特有３’末端配列を含有するリバースプライマーを用いて増幅させた。鋳型は、水酸化ナトリウム含有バッファーを用いて一方の鎖を融解して離すことにより単鎖に作製した。これらの長鎖鋳型集団は、係属中の特許出願に記載された短鎖断片対末端ライブラリーから生成される種を模倣するように設計し、長鎖ＬＳＴ１～７と命名した。

【0316】

プライマーハイブリダイゼーション：２．５μＬの１００μＭ　ＦＡＭ標識プライマーを、１００μＬの１×クレノウバッファーと予備混合した。この溶液を、バッファーの除去後、鋳型が結合された磁気ビーズ（１０^{6}／μＬ）のアリコート３０μＬに添加し、得られた溶液を充分混合した。鋳型／プライマーハイブリダイゼーションを行なわせた後（ハイブリダイゼーション反応は６５℃で２分間、４０℃で２分間および氷上で２分間行なった）、プライマー／バッファーを除去し、ビーズを、３×洗浄ＩＥバッファーを用いて洗浄し、次いで、３００μＬ（１０^{6}／ｍＬ）をＴＥＮＴバッファー（１０ｍＭ　Ｔｒｉｓ，２ｍＭ　ＥＤＴＡ，３０ｍＭ　ＮａＯＡｃおよび０．０１％Ｔｒｉｔｏｎ　Ｘ－１００含有）に再懸濁した。

【0317】

ライゲーション１：次いで、ＬｉｇＳｅｑ－ＦＡＭがハイブリダイズした２．５×１０^{6}のＬＳＴ７ビーズを、３０分間３７℃で、１μＬの１００μＭ　ＬＳＴ７－１九量体，４μＬの５×Ｔ４リガーゼバッファー（Ｉｎｖｉｔｒｏｇｅｎ），１４μＬのH_2Oおよび１μＬのＴ４リガーゼ（１ｕ／μＬ，Ｉｎｖｉｔｒｏｇｅｎ）を含有する混合物中でインキュベートした。

【0318】

切断１：次いで、ビーズを１００μＬのＬＳＷａｓｈ１（１×ＴＥ，３０ｍＭ酢酸ナトリウム，０．０１％Ｔｒｉｔｏｎ　Ｘ１００含有）で３回洗浄し；この溶液の１０μＬアリコートを取り出し、解析のために保存した。次いで、ビーズ（ＩＸ）を１００μＬの３０ｍＭ酢酸ナトリウム中で洗浄した。５０μＬの５０ｍＭ　$AgNO_3$をこの溶液に添加し、得られた混合物を３７℃で２０分間インキュベートした。$AgNO_3$を除去し、ビーズを１回、１００μＬの３０ｍＭ酢酸ナトリウム中で洗浄した。次いで、ビーズを１００μＬのＬＳＷａｓｈ１で３回洗浄し、９０μＬの洗浄。（ＴＥＮＴバッファー）に再懸濁し、この溶液の１０μＬアリコートを取り出し、解析のために保存した。

【0319】

ライゲーション２：ＴＥＮＴバッファーの除去後、ビーズを１４μＬのH_2Oに再懸濁し、３７℃で３０分間、１μＬの１００μＭ　ＬＳＴ７－５九量体，４μＬの５×Ｔ４リガーゼバッファー（Ｉｎｖｉｔｒｏｇｅｎ）および１μＬのＴ４リガーゼ（１ｕ／μＬ，Ｉｎｖｉｔｒｏｇｅｎ）を含有する混合物とともにインキュベートした。

【0320】

切断２：ビーズを、１００μＬのＬＳＷａｓｈ１（１×ＴＥ，３０ｍＭ酢酸ナトリウム，０．０１％Ｔｒｉｔｏｎ　Ｘ１００）中で３回洗浄し、４５μＬのＷａｓｈ１Ｅに再懸濁した。この混合物の１５μＬアリコートを取り出して解析のために保存した。次いで、ビーズを１００μＬの３０ｍＭ酢酸ナトリウムで１回洗浄し、５μＬの２０ｍＭ酢酸ナトリウムに再懸濁した。５０μＬの５０ｍＭ　$AgNO_3$をビーズに添加し、混合物を３７℃で２０分間インキュベートした。$AgNO_3$の除去後、ビーズを、１００μＬの３０ｍＭ酢酸ナトリウムで１回洗浄した。次いで、ビーズを、１００μＬのＬＳＷａｓｈ１中で３回洗浄し、３０μＬのＷａｓｈ１Ｅに再懸濁した。この混合物　２０μＬアリコートを取り出して解析のために保存した。

【0321】

結果

この実験は、図８を参照すると、より良好に理解されよう。図８の上側のセクションは、実験手順の全体的な概要を示す。初期オリゴヌクレオチド（プライマー）を、ビオチン結合によってビーズ結合された鋳型（ＬＳＴ７で示す）にハイブリダイズさせた。初期オリゴヌクレオチドには５’リン酸基を含め、その３’末端をＦＡＭで蛍光標識した。２種類の９量体（九量体）オリゴヌクレオチドプローブ（第１切断可能なオリゴおよび第２切断可能なオリゴ）を、内部にホスホロチオレート含有チミジン塩基（ｓＴ）（下線）を含むように合成した。第１の切断可能なプローブをプライマーの伸長可能な末端に、Ｔ４ＤＮＡリガーゼを用いてライゲートし、次いで、硝酸銀を用いて切断した。切断により、伸長プローブの末端の５ヌクレオチドが除去され、伸長可能な末端が、プローブのプライマーにライゲートされたままの部分上に生成した。次いで、第２の切断可能なプローブを伸長可能な末端にライゲートさせ、次いで、同様に切断させた。

【0322】

蛍光キャピラリー電気泳動ゲルシフトアッセイを、ライゲーションおよび切断の工程をモニタリングするために使用した。このアッセイでは、プライマーを鋳型鎖に、５’リン酸基が、接近する（ｉｎｃｏｍｉｎｇ）オリゴヌクレオチドプローブのライゲーション基質としての機能を果たし得る（フルオロフォアは、移動度型キャピラリーゲル電気泳動のレポーターとしての機能を果たす）ようにハイブリダイズさせる。各工程後、ビーズのアリコートを解析のために取り出した。オリゴヌクレオチドプローブのライゲーション後、磁気ビーズを磁石を用いて回収し、ライゲートされたプライマーおよびプローブ

（１つまたは複数）からなるライゲート種を、鋳型ビーズから熱変性によって放出させ、自動ＤＮＡ配列決定機器（ＡＢＩ　３７３０）を用いる蛍光キャピラリー電気泳動に供した。標識されたサイズ標準（リサミンラダー；　１５～１２０ヌクレオチドのサイズ範囲；クロマトグラムにおいて１組の橙色ピークとして出現、図８参照）。典型的なゲルシフトにおいて、潜在ピークとしては、ｉ）プライマーピーク（無伸長またはプライマー伸長の不足による）、ｉｉ）アデニル化ピーク（ＤＮＡリガーゼの作用による非生産的ライゲーション接合部の５’末端アデノシン残基の結合による－図８Ｆの機構を参照、また、Ｌｅｈｍａｎ，Ｉ．Ｒ．，Ｓｃｉｅｎｃｅ，１８６：７９０－７９７，１９７４も参照）、およびｉｉｉ）終了ピーク（オリゴプローブの結合による）が挙げられる。ライゲーション効率を評価するためのゲルシフトアッセイの使用の利点の１つは、ピーク下面積が、各種の濃度と直接相関することである。

【0323】

図８Ａは、Ｔ４　ＤＮＡリガーゼおよびホスホジエステル結合のみを含有する正確にマッチングしたプローブ（図８Ａの左に示す）を用いて行なわれる対照ライゲーションを示す。橙色ピークはサイズマーカーを表す。左の青色ピークは、ライゲーションの非存在下のプライマー位置を示す。正確にマッチングしたプローブライゲーションは、左へのシフトをもたらす（矢印）。図８Ｂは、同じ条件下で、内部に　チオレート含有Ｔ塩基を含有するプローブ（図８Ｂの左に示す）を用いて行なわれるライゲーションを示す。対照プローブで観察されるものと同一のシフトが見られた（矢印）。ライゲートされたホスホロチオレート含有プローブを含有するビーズ連結鋳型集団を、次いで、硝酸銀とともにインキュベートし、プローブ切断を誘導した。ゲルシフト解析により、左にシフトされた４ｂｐ切断産物（図８Ｃ）の標示によって効率的な切断を確認した。予測される切断産物を図８Ｃの左に示す。切断されたビーズ系鋳型集団を、次いで、第２の回目のライゲーションに曝露し、右にシフトされた１３ｂｐ伸長産物（図８Ｄ）の出現によって生産的ライゲーションが示された。予測される切断産物を図８Ｄの左に示す。２回目の切断により、予測された左にシフトされた８ｂｐ切断産物（図８Ｅ）によって示されるように、効率的な多数の切断工程がなされ得ることが確認された。これらの結果は、成功裏のライゲーションおよびホスホロチオレート結合を含有するプローブの切断を示す。

【0324】

これらの実験で、ライゲーションが１００％完了まで進行しなかったことは明白であるが、より高い完了度が、Ｔ４　ＤＮＡリガーゼを用いた他の実験において観察された（下記参照）。確かに、ライゲーションが完了まで進行することは望ましいが、必須要件ではない。例えば、上記のようなライゲーション工程後、任意の非ライゲート５’末端を５’－ホスファターゼで処理することにより有効に「キャッピングする」ことが可能である。しかしながら、その場合、ライゲーション可能な分子の自然減により、行なわれ得る連続ライゲーションの回数が制限され得る。所与の回数の連続ライゲーションについて、読出し長さは、各ライゲーション／切断サイクル後に残ったプローブの長さ、および
配列決定反応の回数（各々の後には、プライマーの除去およびプライマー結合部位の異なる部分に結合するプライマーのハイブリダイゼーションが起こり、これらは、所与の鋳型において行なわれ得、「リセット」回数ともよばれる）に依存する。これは、プローブの５’末端側に位置する切断可能な連結を有するより長鎖のプローブの使用の場合に論じられる。本発明者らの実験では、六量体プローブにより、八量体およびより長鎖のプローブよりも多量の非ライゲーション可能アデニル化産物をもたらす。したがって、八量体およびより長鎖プローブは、実質的に完了までライゲートされる（下記参照）。また、六量体プローブの５’末端への蛍光部分の付加により、ライゲーション効率が低下するようであるが、八量体プローブへの蛍光部分の付加は、影響はほとんどまたは全くない。これらの理由のため、八量体またはより長鎖のプローブは好ましいとみなされる。

【0325】

さらなる実験（後述）により、ホスホロチオレート結合および縮重低減性ヌクレオチドを含有するプローブのライゲーションおよび切断；ライゲートされた伸長プローブの３’

末端特異性および選択性；ゲル内ライゲーションおよび切断；最小シグナル損失でのプライマーハイブリダイゼーションおよび除去の連続的サイクル；３’→５’伸長でのＴ４またはＴａｑリガーゼの１００％忠実度；ならびにライゲートされた伸長プローブの４色スペクトル分離可能性を示した。該方法を行なうための自動システムを構築した。

【0326】

（実施例２：縮重低減性ヌクレオチドを含有するホスホロチオレート含有オリゴヌクレオチドの効率的な切断およびライゲーション）

しかしながら、プローブ長さと競合する考慮事項は、伸長オリゴヌクレオチドの忠実度およびその後続のライゲーション効率に対する影響である。Ｔ４　ＤＮＡリガーゼの忠実度は、接合後、第５番目の塩基以降で急速に低下することが示されている（Ｌｕｏら．，Ｎｕｃｌｅｉｃ　Ａｃｉｄ　Ｒｅｓ．，２４：３０７１－３０７８および３０７９－３０８５，１９９６）。ミスマッチを新たなライゲーション接合部の５’側に導入した場合、ライゲーション効率は自然減により低下し得るが、バックグラウンドシグナルにおける位相の散逸または増大は生じない（合成方法によるポリメラーゼ系配列決定に見られる大きな障害）。

【0327】

プローブの組は、キャラクタライズされていないＤＮＡのデノボ（ｄｅ　ｎｏｖｏ）配列決定を可能にするため、好ましくは、任意のＤＮＡ配列にハイブリダイズできるものであるのがよい。しかしながら、標識されたプローブの組の複雑度は、４重縮重塩基の長さおよび数にともなって指数関数的に増大する。また、複合プローブの組は、あらゆるプローブ種のほぼ等しい提示を維持したまま、合成するのはより困難であり、生成することよりも困難である。また、これは、各種を一定濃度に維持するのに、より高濃度のプローブ混合物を必要とする。この複雑度に対処するための方法の一例は、４重縮重塩基の代わりに普遍塩基、例えば、デオキシイノシンなどがある特定の位置に組み込まれたヌクレオチドの使用である。

【0328】

４重縮重塩基（Ｎ；等モル量のＡ、Ｃ、Ｇ、Ｔ）および普遍塩基イノシン（Ｉ）を、八量体内の種々の位置に有する１２種類のオクタヌクレオチドプローブを設計した（イノシンは、Ｂ－ＤＮＡ内の任意の４つの標準塩基と二座に配位した水素結合をすることができる；イノシン塩基対の安定性の順は、Ｉ：Ｃ＞Ｉ：Ａ＞Ｉ：Ｔ≒Ｉ：Ｇである）。これらのプローブ設計の評価の目的の１つは、依然としてイノシン塩基の存在下での効率的なライゲーションを支持したまま、どれだけ低い八量体複雑度が達成され得るかを調べることであった。

【0329】

最初の研究では、いくつかのオリゴヌクレオチドプローブを、ビーズ系鋳型（長鎖－ＬＳＴ１）にＴ４　ＤＮＡリガーゼを用いてライゲートさせた。ライゲーションされると、フルオロフォア標識プライマー（３’ＦＡＭプライマー）は、ライゲートされたオリゴヌクレオチドプローブの量に比例して右にシフトする。設計プローブＮ１８－９は最高レベル完了を示し、プローブの効率的なライゲーションにより、＞９９％のプライマー集団が右にシフトした（図９参照）。これらの反応は２５℃で行なった；反応温度を３７℃に上げると、ライゲーションは、いくぶん効率が下がり、完了率はより変動的であった。

【0330】

データをより詳細に調べると、接合部の３’側の最初の５つのヌクレオチド内により少数のイノシン塩基を有するプローブ（下線）は、より高いライゲーション効率を示すことが示された。ライゲーション効率に対する配列状況の潜在的効果をさらに検討し、評価するため、ライゲーション接合部の３’側の最初の５つの塩基内にたった１つのイノシン残基を有する４種類の設計オリゴヌクレオチドプローブを、すべての鋳型においてスクリーニングした。図１０は、Ｔ４　ＤＮＡリガーゼを使用し、多数の鋳型上の選択したプローブ組成物でのゲルシフトアッセイを用いて評価したときのライゲーション完了を示す。これらの最初の実験のデータにより、ライゲーション効率、したがって完了は、イノシン残

基がライゲーション接合部３’側の最初の５つの位置（下線）に配置された場合、変動的であり配列依存性であることが示された。八量体の効率的なライゲーションは、一貫して観察されたが、設計オリゴヌクレオチドプローブＮＩ８－９では、本明細書に示すように、試験したすべての鋳型において＞９９％完了であった。

【0331】

なんら理論に拘束されることを望まないが、このデータ（アデニル化中間体の存在を含む）は、Ｔ４　ＤＮＡリガーゼのためのコアＤＮＡ結合部位内の好ましくないイノシン塩基対により、ＤＮＡタンパク質複合体が不安定化され、酵素結合および後続のライゲーションが充分に低減されるという結論を裏付ける。しかしながら、興味深い疑問は、かかるイノシン塩基対の不安定化が、ライゲートされるオリゴヌクレオチドプローブの忠実度に影響し得るのかどうかということであった。

【0332】

（実施例３：プローブライゲーションの忠実度）

Ｔａｑ　ＤＮＡリガーゼなどの細菌ＮＡＤ依存性リガーゼは、ライゲーション接合部において高い配列忠実度を有し、３’側のミスマッチは本質的にニック閉鎖（ｎｉｃｋ－ｃｌｏｓｕｒｅ）活性を有しないが、５’側のミスマッチはある程度許容性（ｔｏｌｅｒａｔｅｄ）であることが報告されている（Ｌｕｏら．，Ｎｕｃｌｅｉｃ　Ａｃｉｄ　Ｒｅｓ．，２４：３０７１－３０７８および３０７９－３０８５，１９９６）。他方、Ｔ４　ＤＮＡリガーゼは、いくぶんストリンジェンシーが低く、ミスマッチが接合部の３’および５’の両側で許容されることが報告されている。したがって、Ｔ４　ＤＮＡリガーゼを用いてプローブライゲーションの忠実度を、本発明者らのシステムの状況におけるＴａｑ　ＤＮＡリガーゼとの比較において評価することは、興味深かった。

【0333】

本発明者らは、標準的なＡＢＩ配列決定手法を用いてライゲートされるオリゴヌクレオチドの配列忠実度を評価する２つの方法を開発した。第１の方法は、ライゲーション産物をクローニングし、配列決定するように設計した。この方法では、ライゲーション伸長産物をアダプター配列に結合し、クローニングし、細菌内で形質転換した。個々のクローンを採取し、配列決定して、ライゲーション接合部における各位置でのミスマッチ頻度の定量的評価を得た。第２の方法は、ライゲーション産物を直接配列決定するように設計した。該アプローチでは、単鎖ライゲーション産物をビーズ系鋳型から変性させ、相補プライマーを用いて直接配列決定した。精度の低い位置は配列図（ｔｒａｃｅ）において多数の重複ピークを示し、該位置における配列忠実度を示す定性的評価が得られる。

【0334】

第１の方法を用いて、Ｔ４およびＴａｑ　ＤＮＡリガーゼによるプローブライゲーションの相対忠実度を評価した。単一のビーズ系鋳型集団（ＬＳＴ１）を普遍配列決定プライマーにハイブリダイズさせ、これを初期オリゴヌクレオチドとして用いた。次いで、溶液系ライゲーション反応を、縮重オリゴヌクレオチドプローブ（Ｎ７Ａ、３ＡＮＮＮＮＮＮＮ５’、２０００ｐｍｏｌ）の存在下で３７℃、３０分間、Ｔ４　ＤＮＡリガーゼ（１５Ｕ／１×10^6　ビーズ）またはＴａｑ　ＤＮＡリガーゼ（６０Ｕ／１×10^6　ビーズ）のいずれかを用いて行なった（図１１、パネルＡ）。ライゲーション産物をクローニングし、配列決定して、そのライゲーション接合部の３’側（１～８位）の各ＤＮＡリガーゼの位置ごとの忠実度を評価した（図１１、パネルＢおよびＣ）。その結果、Ｔ４　ＤＮＡリガーゼは、最初の５つ位置ではＴａｑ　ＤＮＡリガーゼと本質的に同じレベルの忠実度を有するが、６～８位では忠実度が低いことが示された。これらの結果は、３種類の縮重イノシン含有設計プローブ（３’－ＮＮＮＮＮＩＩＩ－５’、３’－ＮＮＮＮＮＩＮＩ－５’および３’－ＮＮＮＩＮＮＮＩ－５’）での、７種類すべての鋳型（ＬＳＴ１～７）のライゲーション接合部におけるＤＮＡ配列を評価した後続のクローニング実験によってさらに支持された。これらの研究により、Ｔ４　ＤＮＡリガーゼはライゲーション接合部において６～８位では低い配列忠実度を有することが確認されたが、最初の５つ位置では試験したすべての鋳型において高い忠実度が示される（データ示さず）。

【0335】
直接配列決定法を使用し、縮重イノシン含有プローブを用いてＴ４　ＤＮＡリガーゼの忠実度を評価した。オリゴヌクレオチドプローブは、Ｔ４　ＤＮＡリガーゼおよびビーズ系鋳型を含むライゲーション反応物において、２５℃および３７℃で評価した。オリゴヌクレオチドプローブライゲーション効率は、ゲルシフトアッセイを用いて評価した（図１２、パネルＡ）。オリゴヌクレオチドプローブライゲーションにおけるＴ４　ＤＮＡリガーゼの忠実度を評価するため、ＡＢＩ３７３０ｘｌ　ＤＮＡ解析装置を用いたライゲーション反応の直接配列決定を行なった（図１２、パネルＢ）。正確にマッチングしたオリゴプローブと２種類の代表的な縮重イノシン含有オリゴプローブ（ＮＩ８－９およびＮＩ８－１１）のライゲーションにより、＞９９％完了および非常に低頻度のミスマッチ（配列図における多重ピークの非存在）が得られた。このデータにより、プローブが効率的にライゲートされることによっても高い配列忠実度がもたらされることが示される。
【0336】
さらなる実験において、単一のビーズ系鋳型集団（ＬＳＴ１）を、５’リン酸基を含む普遍配列決定プライマーにハイブリダイズさせ、これを初期オリゴヌクレオチドとして用いた。溶液系ライゲーション反応を３７℃で３０分間、Ｔ４　ＤＮＡリガーゼ（１Ｕ／２５０，０００ビーズ）を用い、縮重イノシン含有オリゴヌクレオチドプローブ（３’ＮＮＮＮＮｉｉｉ５’、３’ＮＮＮＮＮｉＮｉ５’または３’ＮＮＮｉＮＮＮｉ５’、６００ｐｍｏｌ）の存在下で行なった。ライゲーション産物をクローニングし、コロニーを採取し、配列決定した。配列忠実度は、ライゲーション接合部における各位置に提示されたクローンの数を計算することにより決定した。結果を図１２のパネルＣ～Ｆに表で示す。これらの研究は、Ｔ４　ＤＮＡリガーゼを用いた縮重イノシン含有プローブの３’→５’ライゲーションが、最初の１～５位で高レベルの忠実度を有することを示す。
【0337】
（実施例４：ゲル内ライゲーションおよび切断）
サイクル的オリゴヌクレオチドライゲーション方法を検討、開発および最適化するための最初の実験は、上記のように、溶液中でビーズ系鋳型を用いて行なった。第２の組の実験では、ライゲーションおよび切断を、スライド上のポリアクリルアミドゲル内に包埋したビーズ系鋳型において行なった。
【0338】
スライドは、各々が自身に結合された単鎖ＤＮＡ鋳型のクローン性の集団を有する数百万個のビーズを５％ポリアクリルアミドと混合し、ガラススライド上で重合を行なわせることにより調製した。テフロン（登録商標）マスクを用いてビーズ含有ポリアクリルアミド溶液を封入した。図１４（上）は、Ｃｙ３標識プライマーがハイブリダイズされた鋳型が結合されたビーズがポリアクリルアミドゲル内に固定化されたスライドの一部分の蛍光画像を示す（このスライドは、異なる実験に使用したが、ここで使用したスライドの典型である）。図１４（下）は、ポリアクリルアミド溶液を封入するためのテフロン（登録商標）マスクを設けたスライドの概略図を示す。
【0339】
反応体をスライド内に、適切な溶液中への手作業によるスライドの浸漬、または自動層流型フローセル内へのスライドの配置のいずれかによって導入した。最初の研究により、効率的なゲル内ライゲーションが実際に、かかるスライド上のポリアクルルアミドマトリックス内に固定化されたビーズに結合された鋳型において行なわれ得ることが確認された。図１５に示す実験では、単鎖ＤＮＡ鋳型ビーズをアクリルアミドおよびＤＡＴＤを含むスライド上に固定化した。重合後、３’フルオロフォア標識された普遍５’リン酸化プライマー（Ｓｅｑプライマー）をゲル内に拡散させ、ハイブリダイズさせた（パネルＡ）。スライドを洗浄して非結合ｓｅｑプライマーを除去し、Ｔ４　ＤＮＡリガーゼ（１０Ｕ）およびオリゴヌクレオチドプローブを含むライゲーションカクテルを重層し、３７℃で３０分間インキュベートした。次いで、スライドを、過ヨウ素酸ナトリウム（０．１Ｍ）を含有するバッファー中でインキュベートし、アクリルアミドポリマーを消化させ、ビーズ

系鋳型集団を放出させた。ライゲーション産物を熱により鋳型鎖から変性させ、回収し、上記のゲルシフトアッセイを用いて解析した。ゲル内ライゲーション反応をＴ４　ＤＮＡリガーゼの非存在下で行ない、非ライゲート配列決定プライマーに代表的な単一のピークが示された（パネルＢ）。ライゲーション反応は、八量体プローブを用い、Ｔ４　ＤＮＡリガーゼの存在下で行ない、効率的なゲル内オリゴヌクレオチドライゲーションが示され、＞９９％のビーズ系鋳型集団が効率的にライゲートされた（パネルＣ）。

【0340】

（実施例５：４色検出）

検出効率を最大にするには、各可能な塩基付加産物に対応する異なる標識を有する１組のオリゴヌクレオチドプローブを用いることが望ましい。これを、図１５に概略を示し、適切な励起および発光フィルターを備えた本発明者らの自動配列決定機器においてモデル化した。プローブの特異性および選択性の課題に取り組むため、３組の八量体プローブを設計した。第１の組は、４種類の特有の鋳型集団に相補的で、異なる３’塩基および５’色素標識を有する４種類の八量体を含むものとした。第２の組は、特有の３’塩基および５’色素を有する７種類の特有の八量体を含むものとした。第３の組は、各々が異なる５’色素標識で同定される特有の３’末端塩基を有する４種類の縮重イノシン含有八量体設計プローブに対応させた。

【0341】

４色スペクトル正体を確認するため、プローブ組番号１を用いて、４種類の特有の鋳型集団（図１６参照）を検出した。ポリアクリルアミド内に包埋されたビーズに結合された４種類の特有の単鎖鋳型集団を含むスライドを調製した（パネルＡ）。各ビーズは、自身に結合された鋳型のクローン性の集団を有した。５’リン酸基を含有する普遍配列決定プライマーをインサイチュでハイブリダイズさせ、ライゲーション反応を、４種類の特有のフルオロフォアプローブ（Ｃｙ５、ＣＡＬ　６１０、ＣＡＬ　５６０、ＦＡＭ；各１００ｐｍｏｌ）およびＴ４　ＤＮＡリガーゼ（１０Ｕ／スライド）を含むオリゴヌクレオチドプローブ混合物を用いて行なった。スライドを３７℃で３０分間インキュベートし、洗浄して非結合プローブを除去した。スライドを明るい光で画像化し、白色光塩基画像（パネルＢ）を得た。蛍光励起には、４種類の帯域通過フィルター（ＦＩＴＣ、Ｃｙ３、ＴｘＲｅｄおよびＣｙ５）を用いた。蛍光画像捕捉をライゲーションの前後で行なった。個々の集団を擬似カラー表示し（パネルＣ）、画像値のスペクトル正体をプロットし、最小シグナル重複を確認した（パネルＤ）。

【0342】

（実施例６：ゲル内でのライゲーション特異性および選択性の実証）

３’末端特異性を確認するため、プローブ組番号２を用いて、単一の鋳型集団のインテロゲーションを行なった（図１７参照）。ポリアクリルアミドゲル内に包埋された自身に結合され単一の鋳型集団（ＬＳＴ１．Ｔ）を有するビーズを有するスライドを調製し、普遍配列決定プライマーとインサイチュでハイブリダイズさせた（パネルＡ）。ゲル内ライゲーション反応を、Ｔ４　ＤＮＡリガーゼ（１０Ｕ／スライド）を使用し、単一の３’塩基だけが異なる４種類の５’末端－標識されたプローブで構成されたオリゴヌクレオチドプローブ混合物を用いて行なった。スライドを３７℃で３０分間インキュベートし、洗浄して非結合プローブ集団を除去した。スライドを白色光で画像化し、塩基画像を得（パネルＢ）、蛍光励起には、４種類の帯域通過フィルター（ＦＩＴＣ、Ｃｙ３、ＴｘＲｅｄおよびＣｙ５）を用いた。蛍光画像捕捉をライゲーションの前後で行ない、単一のＦＡＭ系プローブ集団（青色スポット）がＴ４　ＤＮＡリガーゼを用いたゲル内ライゲーション後に存在し、スペクトル重複がないことを確認した（パネルＣ、Ｄ）。このデータは、Ｔ４　ＤＮＡリガーゼを用いたプローブ特異性がストリンジェントであり、ライゲーション接合部の最初の３’末端塩基によって決定されることを示す。

【0343】

３’末端特異性および選択性をさらに実証するため、プローブ組番号２を用いて、単一の塩基差を含み異なる量で存在するビーズ系鋳型集団の混合物を同定した。図１８のパネ

ルＡに示すような、各々が自身に結合された４つの鋳型集団（各々は一塩基多型を有する（ＬＳＴ１；Ａ、Ｇ、ＣまたはＴ））の１つを有するビーズの混合物を有するスライドを調製した。ビーズを、スライド上のポリアクリルアミドゲル内に包埋した。ビーズ系鋳型集団は、パネルＤに概略を示すように、種々の異なる頻度で用いた。スライドをインサイチュで普遍配列決定プライマーとハイブリダイズさせた。ゲル内ライゲーション反応を、Ｔ４　ＤＮＡリガーゼ（１０Ｕ／スライド）および等モル量（各１００ｐｍｏｌ）の４種類の５’末端標識プローブ（単一の３’塩基だけ異なる）を含有するオリゴヌクレオチドプローブ混合物を用いて行なった。スライドを３７℃で３０分間インキュベートし、洗浄して非結合プローブ集団を除去した。スライドを白色光で画像化し、塩基画像を得（パネルＢ）、蛍光励起には、４種類の帯域通過フィルター（ＦＩＴＣ、Ｃｙ３、ＴｘＲｅｄおよびＣｙ５）を用いた。個々のプローブ画像を重ね撮りし、擬似カラー表示した（パネルＣ）。蛍光画像を、ビーズ呼び出し（ｃａｌｌｉｎｇ）ソフトウエアを用いて一覧にした。結果をパネルＤに示す。観察されたライゲーション頻度（Ｏｂｓ）は、予測した頻度（Ｅｘｐ）と相関したことが確認される。このデータは、ライゲーション後、多数の鋳型の存在下での高いプローブ特異性およびプローブ選択性を示し、ライゲーションによって、一塩基多型（ＳＮＰ）、すなわち、集団の異なる個体のゲノムＤＮＡ鎖内の単一のヌクレオチド塩基内に生じた改変を検出できる能力を示す。

【0344】

（実施例７：４色縮重イノシン含有伸長プローブを用いたゲル内でのライゲーション特異性および選択性の実証）

別の組の実験をプローブ組番号３を用いて行ない、プローブライゲーションの特異性および選択性を、４色縮重イノシン含有オリゴヌクレオチドプローブプールを用いてを評価した。結果を図１９に示す。ビーズ系スライドを上記のようにして調製したが、４種類の特有の単鎖鋳型集団をビーズ上に異なる量で存在せ、これを、次いで、普遍配列決定プライマーとインサイチュでハイブリダイズさせた（パネルＡ）。ゲル内ライゲーション反応を、Ｔ４　ＤＮＡリガーゼ（１０Ｕ／スライド）の存在下、５種類の縮重塩基（Ｎ；複雑度4^5＝１０２４）、２種類の普遍塩基（Ｉ、イノシン）および単一の既知ヌクレオチドを、特異的５’フルオロフォア（Ｇ－Ｃｙ５、Ａ－ＣＡＬ　６１０、Ｔ－ＣＡＬ５６０、Ａ－ＦＡＭ；各６００ｐｍｏｌ）に対応する３’末端に有するように設計された八量体からなるプローブプールを用いて行なった。スライドを３７℃で３０分間インキュベートし、洗浄して非結合プローブ集団を除去した。スライドを白色光で画像化し、塩基画像を得（パネルＢ）、蛍光励起には、４種類の帯域通過フィルター（ＦＩＴＣ、Ｃｙ３、ＴｘＲｅｄおよびＣｙ５）を用いた。個々のプローブ画像を重ね撮りし、擬似カラー表示した（パネルＣ）。蛍光画像を一覧にし、各ライゲーション産物の頻度を、ビーズ呼び出しソフトウエアを用いて表で示した（パネルＤ）；ビーズシグナル値の上部９０％表す未処理生データおよびフィルター処理データのスペクトル散布図をパネルＥに示す。このデータは、観察されたライゲーション頻度（Ｏｂｓ）が、各鋳型の既知濃度に基づいて予測した頻度（Ｅｘｐ）と相関したことを示す。これにより、縮重した普遍塩基含有プローブプールをＴ４　ＤＮＡリガーゼとともに使用すると、特異的および選択的なゲル内ライゲーションがもたらされ得ることが確認される。

【0345】

（実施例８：　ゲル内での初期オリゴヌクレオチドのハイブリダイゼーションおよび除去の反復サイクルの実証）

自動フローセル内に取り付けた顕微鏡スライド上のゲル内に固定化された鋳型において実験を行ない（下記参照）、初期オリゴヌクレオチドのアニーリングおよびストリッピングの多数のサイクルが、スライド上のゲル内に包埋されたビーズに結合された鋳型に適用され得、シグナル損失は最少であることを確認した。４４塩基蛍光標識初期オリゴヌクレオチドを使用した。図２０に示すように、１０サイクルで生じたシグナル損失は最少であった。初期オリゴヌクレオチドを、図２０ではプライマーという。上記のように、合成手順によるポリメラーゼ系配列決定の大きな欠点の１つは正（ｐｏｓｉｔｉｖｅ）および負

（ｎｅｇａｔｉｖｅ）の両方の位相の散逸が個々の鋳型鎖において起こる傾向である。正の位相の散逸は、ヌクレオチドが成長鎖内に誤って組み込まれ、したがって、該特定の鎖の塩基配列が残りの鋳型から得られる配列より前で実行（ｒｕｎ　ａｈｅａｄ）され、ｎ＋１塩基呼び出し分位相ずれが引き起こされた場合に起こる。負の位相の散逸は、より一般的であり、鎖が充分に伸長されず、バックグラウンド塩基呼び出しがもたらされ、成長鎖より後ろ（ｎ－１）で実行された場合に起こる。伸長産物効率的にストリッピングできること、および差次的に位置する初期オリゴヌクレオチドをハイブリダイズすることにより鋳型を「リセット」できることで、シグナル自然減がほとんどまたは全くなしで、非常に長い読出し長さが可能になる。

【0346】

（実施例９：自動配列決定システム）

この実施例では、１つ以上の鋳型から配列情報を収集するために使用され得る代表的な本発明の自動配列決定システムを記載する。好ましくは、鋳型を、実質的に平面状の基材、例えば、ガラス顕微鏡スライド上に配置させる。例えば、鋳型は、基材上でアレイ状であるビーズに結合されたものであり得る。該システムの写真を図２１に示す。該システムは、Ｏｌｙｍｐｕｓエピ蛍光顕微鏡本体（横向きに設置）を主部とし、自動オートフォーカシングステージおよびＣＣＤカメラを備える。回転式ホルダ内の４つのフィルタキューブにより、さまざまな励起および発光波長での４色検出が可能である。ペルチェ（ｐｅｌｔｉｅｒ）温度制御を伴うフローセルは、スライドなどの基材を収容するように開放および閉鎖が可能であり（ゲルなどの半固相支持体を含む領域の縁部周囲を密閉するためのガスケットを有する）、ステージ上に設置されている。フローセルの垂直配向は、本発明のシステムの重要な態様であり、気泡がフローセルの上部から散逸するのを可能にする。セルは、各洗浄工程前にすべての試薬が噴出されるように、空気が充分充填されたものであり得る。フローセルは、４種類の差次的に標識されたプローブ混合物、切断試薬、任意の他の所望の試薬、酵素並行化バッファー、洗浄バッファーおよび単一のポートを介するフローセルへの空気の送達を可能にする２つの９－ポートＣａｖｒｏシリンジポンプを有する流体取り扱い部と接続されている。システムの操作は完全に自動化されており、多数のＩ／Ｏポートを有する専用のコンピュータを用いて、制御ソフトウエアによりプログラミング可能である。Ｃｏｏｋｅ　Ｓｅｎｓｉｃａｍカメラには１．３メガピクセル冷却ＣＣＤが組み込まれているが、より低いまたはより高い感度を有するカメラもまた使用され得る（例えば、４メガピクセル、８メガピクセルなども使用され得る）。フローセルには、０．２５ミクロンステージが用いられており、形状は１ミクロンである。

【0347】

（実施例１０：画像取得および処理方法）

この実施例では、自身に結合された標識された核酸を有するビーズのアレイからの画像の取得および処理のための代表的な方法を記載する。正確な素性の同定およびアライメントは、各取得画像の信頼できる解析に重要である。素性は、まず、各ビーズについて最も強いピクセル以外は排除するにより同定される。所与の画像のピクセル値をヒストグラムにプロットする；バックグラウンドに対応するピクセルを排除し、残りのピクセル値を分取する。均一な画像では、すべてのビーズがほぼ同じ強度であり、アルゴリズムにより、ピクセル値の下部８０～９０％を除外する。次いで、上部１０～２０％値を有するピクセルをスキャンし、４ピクセル半径内で極大のものを同定する。次いで、該領域内の平均強度および周辺の平均強度を記録する。これらの値により正規分布を形成し、次いで、その値が該分布から外れるピクセルを除く。最初に無視したピクセルの割合、該円形領域の大きさ、および正規分布内の可能なビーズを排除した切り捨て値を、すべてパラメータ化し、必要であれば調整し得る。該組の各画像について素性行列を作製することによりアライメントを行なう。次いで、得られた行列を、最も高頻度のｘ、ｙ座標オフセットについて検索し、最適なアライメントを同定する。

【0348】

ビーズ画像を、伸長プローブ添加の前に、Ｃｙ５チャネル（配列決定プライマーに対応

）に収集する。これらの画像を用いて、各ビーズについて位置的座標および蛍光単位（ＲＦＵ値）としての生シグナル強度の両方をマーキングする（ｍａｒｋｉｎｇ）素性マップを作製する。各後続の二本鎖伸長後、画像の組を、Ｃｙ３－標識ヌクレオチドを添加する前と後の両方で取得する。これらの画像を元のＣｙ５　画像に対してアライメントし、次いで、ＲＦＵ値をビーズの各々に割り当て、記録する。各塩基付加の非標識（伸長前）および標識（蛍光付加）画像間の強度を差を差し引くことによりベースライン補正を適用する。これらのベースラインを差し引いた値を、次いで、各素性でＣｙ５画像において見られた強度によって標準化し、ビーズが伸長された、またはされていないとみなす（すなわち、ビーズは、該ビーズに結合された二本鎖が伸長された場合、伸長されたとみなす）基準を形成する。これらの方法を用い、約１，３００画像／スライドで１つの画像あたり数千の素性が解析され、１回の実験で５～１０００００００鋳型種の解析がもたらされ得る。アルゴリズムは、さらなる効率増強のために、後日でも容易にＭＡＴＬＡＢからＣ＋にポート移動（ｐｏｒｔ）され得るように設計した。

【0349】

（実施例１１：ビーズアライメントおよび追跡および配列でコード化）

この実施例では、自身に結合された標識され核酸を有するビーズのアレイからの画像処理および取得されデータからの配列決定のための代表的な方法を記載する。

【0350】

画像解析は、ビーズサイズに適合した直径を有するゼロインテグラルサーキュラートップハットカーネル（ｚｅｒｏ－ｉｎｔｅｇｒａｌ　ｃｉｒｃｕｌａｒ　ｔｏｐ－ｈａｔ　ｋｅｒｎｅｌ）を用いて画像を回転させることにより開始する。これは、自動的にバックグラウンドをゼロに標準化すると同時に、個々のビーズの中心を極大によりを同定する。最大値を位置し、他の極大から孤立性のものが、アライメント点として使用される。これらのアライメント点は、各画像について時系列でコンピュータ計算される。各対の画像について、アライメント点を比較し、変位ベクトルを、すべての共通アライメント点の平均変位に基づいてコンピュータ計算する。これにより、サブピクセル分解によるペアワイズ画像変位が提供される。

【0351】

Ｎ個の画像では、Ｎ＊（Ｎ－１）／２のペアワイズ変位があるが、これらのうちＮ－１だけは、残りが非依存的な組から計算され得るため、非依存的である。例えば、画像１と２間および画像１と３間の変位を測定すると、画像２と３間の変位が示唆される。画像２と３間の測定された変位が示唆された変位と同じでない場合、測定値は不一致である。この不一致の大きさは、どれだけ良好にアライメントアルゴリズムが機能しているかを測るための測定基準として使用され得る。本発明者らの最初の試験は、各寸法において、一般的に、０．１ピクセル未満である不一致を示す（図２３参照）。

【0352】

いったん、時系列の画像がアライメントされたら、個々のビーズを追跡するのに２つの方法がある。ビーズ密度が、別のビーズと接触していないビーズの大部分で低い場合、個々の各ビーズの光学質量中心が同定され得、ビーズ周囲の領域が積分され、ビーズ強度がコンピュータ計算される。ビーズ密度が非常に高いため、ビーズの大部分が接触している場合は、その周囲にある暗いバックグラウンドバンドによって個々のビーズを同定することは可能でない。しかしながら、サブピクセル分解のためアライメントされたすべての画像で、隣接するピクセルの相関性を経時的にコンピュータ計算することにより、同じビーズに属するピクセルを同定することが可能である。高度に相関するピクセル対は、確信的に同じビーズに割り当てられ得る。同様の手法がＤＮＡ配列決定ゲル内でのレーン追跡に適用され、良好な結果が得られている（Ｂｌａｎｃｈａｒｄ，Ａ．Ｐ．Ｓｅｑｕｅｎｃｅ－ｓｐｅｃｉｆｉｃ　ｅｆｆｅｃｔｓ　ｏｎ　ｔｈｅ　ｉｎｃｏｅｐｏｒａｔｉｏｎ　ｏｆ　ｄｉｄｅｏｘｙｎｕｃｌｅｏｔｉｄｅｓ　ｂｙ　ａ　ｍｏｄｉｆｉｅｄ　Ｔ７　ｐｏｌｙｍｅｒａｓｅ，Ｃａｌｉｆｏｒｎｉａ　Ｉｎｓｔｉｔｕｔｅ　ｏｆ　Ｔｅｃｈｎｏｌｏｇｙ，１９９３）。いったんビーズが全４色時系列により追跡されたら、配列は、どの

色がプローブオリゴヌクレオチドのどの最も３’側の塩基に対応するかがわかることによりデコード化される。

【0353】

（実施例１１：処理量計算）

一般に、配列決定システムの処理量は、主に、装置が１日あたりにもたらすことができる画像の数および１つの画像あたりの配列データのヌクレオチド（塩基）の数によって規定される。装置は、好ましくは、カメラが常時作動中で維持されるように設計され、計算は、１００％カメラ利用に基づいて行なう。１つの塩基の正体を決定するのに各ビーズが４色で画像化される実施において、１つのカメラで４画像、２つのカメラで２画像、または４つのカメラで１画像のいずれかが使用され得る。４つのカメラでの画像化により、他の選択肢よりも劇的に高い処理量が可能になり、好ましいシステムでは、該アプローチが用いられる。

【0354】

本発明者らの最初の試験は、５０ピクセル／ビーズのピクセル密度（５．４平方ミクロンを表す）は、標準的な画像解析に適度な（ｃｏｍｆｏｒｔａｂｌｅ）密度を提供することを示す。４メガピクセルＣＣＤ　カメラ（現在では、一般的）を用いることにより、単一のＣＣＤフレームで約８０，０００個のビーズの画像化ができる（本発明者らの現在の画像データに基づく）。別個のカメラでの４種類の画像の捕捉およびフローセル上の次のフィールドへの移動に１．５秒より多くかからない。７５％のビーズにより有用な情報が得られる場合、本発明者らは、およそ８０，０００ビーズ* ０．７５／１．５＝４０，０００塩基／秒の生配列データのデータを収集することができる。

【0355】

１００％カメラ利用の維持における大きな問題の１つは、ライゲーション／切断化学反応の１サイクルを行なうのにかかる時間全フローセルの画像化に必要とされる時間に適合させることである。伸長、切断およびライゲーションの１サイクルにかかる時間の適度な推定は１．５時間（５，４００秒間）である。５，４００秒間は１，８００画像フィールドまたは約１５ｍｍ×４５ｍｍの面積に適応し、これは、フローセルに適度なサイズである。４種類のカメラを用いるシステムの処理量の伸長な推定は、１５ｍｍ×４５ｍｍフローセルで４０，０００塩基／秒である。これは、２８実行／日で約６５０塩基読出し長さ（２０塩基／秒）の処理量に基づくと、およそ２，０００　ＡＢＩ３７３０ｘ１配列決定装置に相当し、本発明者らは、これらの装置を用いて達成した。ビーズ密度を２００，０００／画像に２．５倍増大させると、１００，０００塩基／秒への処理量の全体的な増加が可能になり、これは、ほぼ５，０００　ＡＢＩ３７３０ｘ１装置に相当する。この処理量レベルでの全出力／日は、約８．６Ｇｂ／日であり、１２×ヒトゲノム配列決定を完了するのに必要とされる時間は、約４．２日となり得る。

【0356】

本発明の配列決定本明細書に記載の方法は、さまざまな異なる配列決定システム、画像捕捉および処理方法などを用いて実施され得ることに注意されたい。例えば、米国特許第６，４０６，８４８号および同第６，６５４，５０５号ならびにＰＣＴ公開公報ＷＯ９８０５３３００の論考を参照のこと。

【0357】

（実施例１２：その上面で鋳型合成を微粒子の調製方法）

この実施例では、鋳型分子のクローン性の集団が各微粒子に結合されるように鋳型が増幅され得る（例えば、ＰＣＲにより）ように自身に結合された増幅プライマーを有する微粒子（この実施例では磁気ビーズ）のプロトコル調製を記載する。一般に、増幅ビーズは、クローンＰＣＲ反応に必要とされる１つのプライマーが自身に結合されている。このプライマーは、共有結合によりカップリングさせてもよく、または例えば、ビオチン標識し、ビーズ表面上のストレプトアビジンに結合させてもよい。ビーズは、標準的なＰＣＲ反応（例えば、マイクロタイタープレートのウェル、チューブ内など）、実施例１３に記載のエマルジョンＰＣＲ反応などに使用され、自身に結合された鋳型分子のクローン性の集

団を有するビーズが得られ得る。
【0358】
材料
１×ＴＥ：１０ｍＭ　Ｔｒｉｓ（ｐＨ８）１ｍＭ　ＥＤＴＡ
１×ＰＣＲバッファー：（ＴｈｅｒｍｏＰｏｌバッファー，ＮＥＢ）
２０ｍＭ　Ｔｒｉｓ－ＨＣｌ（ｐＨ８．８）
１０ｍＭ　ＫＣｌ
１０ｍＭ$(NH_4)_2SO_4$
２ｍＭ　$MgSO_4$
０．１％Ｔｒｉｔｏｎ　Ｘ－１００
１Ｍ　ベタイン（　１×ＰＣＲ－Ｂバッファーの場合のみ添加）
１×結合および洗浄バッファー
５ｍＭ　Ｔｒｉｓ　ＨＣｌ（ｐＨ７．５）
０．５ｍＭ　ＥＤＴＡ
１　ＭＮａＣｌ
ＤＮＡ捕捉プライマー（２０量体，５００μＭストック）
二重ビオチン－（ＨＥＧ）５－Ｐ１：５’－二重ビオチン－（ＨＥＧ）５－ＣＴＡ　ＡＧＧ　ＴＡＧ　ＣＧＡ　ＣＴＧ　ＴＣＣ　ＴＡ－３’
（ＨＥＧ）５＝ヘキサエチレングリコールリンカー、１８炭素含有スペーサー、使用され得るいくつかの異なるスペーサー部分の１種類。例えば、ビーズの表面から離れるオリゴのＰ１プライマー部分を増大させるため、スペーサーを含めることが有用である。本明細書に記載の任意のプライマーには、かかるスペーサー部分が組み込まれ得る。
Ｄｙｎａｌストック磁気ビーズ（１μｍ直径）＝１０ｍｇ／ｍｌ（７～１２×10^6ビーズ／μｌ）。
【0359】
方法
１．５０μｌビーズ（約４５０×10^6　ビーズ）を除去。
２．２００μｌ　１×ＴＥバッファーを添加、充分混合。磁石で分離。
３．１×ｗｉｔｈ　２００μｌ　１×ＴＥバッファー洗浄。磁石で分離。
４．１００μｌ　Ｂ／Ｗバッファー中に再懸濁。
５．３μｌのＰ１オリゴ（５００μＭストック＝１５００ｐｍｏｌ）を添加。
６．室温で＞３０分間回転。
７．２００μｌ　１×ＴＥバッファーで３回洗浄。
８．５０μｌ（初期容量）１×ＴＥバッファー中に再懸濁。
９．使用前にＤＮＡ捕捉ビーズを４℃保存または氷上に配置。ビーズは、１週間以内に使用するのがよい（ビーズは、＞１週間の保存期間で凝集する傾向がある）。
【0360】
（実施例１３：エマルジョン中の微粒子上でＰＣＲを行なうための方法）
この実施例では、エマルジョン中の微粒子上でＰＣＲを行ない、自身に結合されたクローン鋳型を有する微粒子を生成させるために使用され得る方法を記載する。微粒子（以下、ＤＮＡビーズをいう名称を用いる）を、まず、第１のプライマー（Ｐ１）により官能性付与させる。第２のプライマー（Ｐ２）を水相中に存在させ、そこでＰＣＲ反応を行なわせる。所望の場合は、低濃度をＰ１を水相中に含め得る（例えば、２０倍低い）。そうすることにより、鋳型の水相中での迅速な構築が可能になり、これは，さらなる増幅のための基材となる。Ｐ１が溶液中ですると、反応は、微粒子に結合されたＰ１の利用に向かって誘導される。Ｐ１＿Ｐ２　ｄｅｇｅｎ１０は、Ｐ１およびＰ２にハイブリダイズしてＰＣＲによる増幅をもたらす配列およびオリゴヌクレオチド集団に4^{10}の複雑度を与える１０縮重塩基鎖（オリゴヌクレオチド合成中に組み込まれる）を有するオリゴヌクレオチド鋳型（１００ｂｐ）である。
Ｉ．エマルジョンプロトコル（１μｍビーズ）

１．油相の調製：
Ｓｐａｎ　８０（７％）
Ｔｗｅｅｎ　８０（０．４％）
低分子鉱油中で調製
新たに作製した油相のみ使用
全油相＝４５０μｌ
２．水相の調製：（２×１０^{9}滴が作製されたと概算，１１５μＬ／滴）
【0361】
【化7】

試薬（ストック）	（μｌ）反応液あたり	最終
dH_2O	156.0	---
$MgCl_2$ バッファー（10X）	32.0	1X
dNTP (100 mM ea)	11.3	各3.5 mM
$MgCl_2$ (1M)	7.3	23mM
ベタイン（5 M)	32.0	0.5M
P1 (プライマー1)(10 μM)	1.6	11.25pmol
P2 (プライマー2)(200 μM)	40.0	5625 pmol
P1_P2 degen10 (100 pM)	6.6	5.9x10^7/ul
DNA ビーズ (8M/μl)	25.0	150M/エマルジョン
白金 Taq (5 U/μl)	9.0	0.28U/ul

全水性容量＝３２０μｌ
最終反応＝２５５μｌ　水相　：４５０μｌ　油相
３．エマルジョンの添加まで水相チューブを氷に移動。
４．４５０μｌ　油相を２ｍｌクリオバイアルに添加。
５．クリオバイアルＵＰＲＩＧＨＴをＩＫＡ　ボルテックスに取り付けた発泡アダプター内に配置。ボルテックスを２５００ｒｐｍに設定。
６．アリコート水相（３　アリコート，各８５μｌ＝２５５μｌ）を油相に入れて振とうする。単分散性水相を、先端をチューブ内に配置し、ゆっくりと水相を先端から振とう油相内に分配することにより撹拌２ｍｌクリオバイアルに添加。残留水相に２回添加を反復。
７．エマルジョンの振とうを２４分間　２５００ｒｐｍで継続、
８．エマルジョンの約１００μｌ　アリコートを９６ウェルプレート内に移す（合計＝４ウェル）。また、残留水相のアリコート（６５μｌ）を、溶液系ＰＣＲ対照反応のために別のウェルに。プレートを密封し、次のセクションに概略を示すようにサイクルさせる。
ＩＩ．エマルジョン増幅（１μｍ　ビーズ）
１．１μｍ　ビーズエマルジョン（プライマー使用Ｔｍｓ＝６２Ｃ）のＰＣＲサイクリングパラメータ：
プログラム：ＤＴＢ－ＰＣＲ
９４Ｃ，２分　ｎ＝１
９４Ｃ，１５秒
５７Ｃ，３０ｓ　ｎ＝１００
７０Ｃ，６０秒
５５Ｃ，５分　ｎ＝１
１０Ｃ，随意時間
２．サイクリング時間は約６時間である。
３．サイクリング後のエマルジョンを観察。良好なエマルジョンは均一な琥珀色に見え、分離水相は観察されない。「破断」したエマルジョン（溶液から分離）は、チューブの底部に異なる水相を有する。ビーズのこの集団はクローン性でないため、この段階での回収は回避。
４．明るい視野での顕微鏡検査を用いてサイクル後エマルジョンを評価。サイクル的エマ

ルジョンの２μｌアリコートを取り出し、ガラススライドに１滴落とす。エマルジョンアンプルに２２×６０ｍｍガラスカバースリップを重層。
５．２０×対物レンズを用いてエマルジョンを観察。ビーズは、好ましくは、単分散性であり、液滴の大部分が単一のビーズを含む。
注：エマルジョンアンプルが多数の多ビーズ液滴を含む場合、プールエマルジョン反応を単一の１．５ｍｌのエッペンドルフチューブに入れ、６０００ｒｐｍで１５秒間回転させる。チューブの底部に蓄積したビーズ懸濁液を取り出す。この集団は、遊離ビーズおよび単一ビーズ液滴より重い多ビーズ液滴の両方で構成され、したがって、短時間のスピン後、チューブの底部に沈降する、このビーズ集団はクローンでない、したがって、後続の処理前に回避するのがよい。工程４および５を反復することによりエマルジョンを再評価し、
エマルジョンアンプル中での単一のビーズ含有液滴の完全性を確認する。
６．次のセクションに概要を示すプロトコルを用いたエマルジョンの破壊（崩壊）
ＩＩＩ．エマルジョン破壊および融解（１μｍビーズ’）
ビーズ破壊洗浄（ＢＢＷ）バッファー
２％Ｔｒｉｔｏｎ　Ｘ－１００　２％Ｔｗｅｅｎ　２０；１０ｍＭ　ＥＤＴＡ
融解溶液　１００ｍＭ　ＮａＯＨ
１×ＴＥ：１０ｍＭ　Ｔｒｉｓ（ｐＨ８）１ｍＭ　ＥＤＴＡ
１×結合および洗浄（Ｂ／Ｗ）バッファー
５ｍＭ　Ｔｒｉｓ－ＨＣｌ（ｐＨ７．５）
０．５ｍＭ　ＥＤＴＡ
１Ｍ　ＮａＣｌ
１．各エマルジョンセット（４　アリコート）を単一の１．５ｍｌエッペンドルフチューブ中にプール。
２．８００μｌ　ＢＢＷバッファーを添加。反応チューブの１０秒間のボルテックスによりエマルジョンの破壊。
３．８０００ｒｐｍで２分間スピン。
４．上部８００μｌ（主に油相）の除去。チューブの底部でＤＮＡビーズをペレット化。
５．８００μｌ　ＢＢＷを添加、ボルテックスおよび８０００ｒｐｍで２分間スピン、上部６００μｌ除去。
６．
さらに２回６００μｌ　１×ＴＥで洗浄、磁石を用いて各洗浄液を交換。
８．５０μｌ　融解溶液をビーズペレットに添加、激しいペッティイングにより再懸濁試料。融解溶液中で５分間室温でビーズのインキュベート。チューブ断続的に打つ。
９．チューブを磁石に配置して融解溶液を除去。１×１００μｌ　融解溶液で洗浄し、第２鎖の完全な除去を確保。
１０．ビーズペレットを２回１×ＴＥで洗浄、２０μｌ　ＴＥバッファー中に再懸濁して４Ｃで保存または２０μｌ　１×Ｂ／Ｗバッファー（次の工程が富化の場合）。ビーズが凝集しているようであれば、１×ＰＣＲ－　Ｂバッファーに交換。
１１．富化プロトコルを継続（任意選択的）。
【0362】
（実施例１４：自身に結合されたクローン鋳型集団を有する微粒子の富化方法）
　この実施例では、上面での鋳型増幅が例えばＰＣＲエマルジョンで成功裏に行なわれた微粒子富化方法を記載する。該方法は、自身に結合された捕捉オリゴヌクレオチドを有する大きな微粒子を利用する。捕捉オリゴヌクレオチドは鋳型内に存在するヌクレオチド領域に相補的なヌクレオチド領域を含む。
Ｉ　．エマルジョン富化（１ｕｍ）
Ａ．富化ビーズ（捕捉実体）の調製
富化ビーズ：
Ｓｐｈｅｒｏｔｅｃｈストレプトアビジンコートされたポリスチレンビーズ（約６．５ｕ

ｍ）
ビーズストック（０．５％ｗ／ｖ）：３３，１２５　ビーズ／μｌ
Ｐｅｒ　プロトコル：（３３，１２５　ビーズ／μｌ）（８００μｌ）＝２６．５×１０^{6}　ビーズ
使用：
１１９０００００００ビーズ／エマルジョン－エマルジョンクローン性の概算（２％）：約３Ｍ　鋳型－陽性ビーズ／エマルジョン。２～３富化ビーズ／概算鋳型陽性エマルジョンビーズ＝１００００００００　富化ビーズ／エマルジョン反応を添加。
富化オリゴヌクレオチド（捕捉剤）：
Ｐ２－富化（３５量体，Ｔｍ＝７３Ｃ）
５’－二重ビオチン－１８－炭素スペーサー－ｔｔａｇｇａｃｃｇｔｔａｔａｇｔｔａｇｇｔｇａｔｇｃａｔｔａｃｃｃｔｇ　３’
（または）
Ｐ２－富化（例えば、３５量体まで，Ｔｍ＝５２Ｃ）
５’－二重ビオチン－１８－炭素スペーサー－ｇｇｔｇａｔｇｃａｔｔａｃｃｃｔｇ　３’
グリセロール溶液　－　６０％（ｖ／ｖ）
６ｍｌ　グリセロール
４ｍｌ　ヌクレアーゼ無含有H_2O
１．８００μｌのビーズを除去、１３，０００ｒｐｍで１分間の遠心分離によりＢ／Ｗバッファーに交換。１×の５００μｌ　Ｂ／Ｗバッファーでの洗浄、および１００μｌ　Ｂ／Ｗバッファーへの再懸濁。
２．２０μｌ　富化オリゴを添加。（５００μＭストック＝１０，０００ｐｍｏｌ／ｒｘｎ）。
３．ビーズ反応物を室温で１時間回転。
４．ビーズを３回、５００μｌ　１×ＴＥバッファーを用いて洗浄。１３，０００ｒｐｍで１分間遠心分離により洗浄間でビーズをペレット化。
５．ビーズを２５μｌ　Ｂ／Ｗバッファーに再懸濁。濃度＝１Ｍ　富化ビーズ／μｌ。注：４種類の富化エマルジョン集団を２０～３０μｌ　１×Ｂ／Ｗバッファーにプールし、約４０Ｍ　鋳型陽性ビーズを得る。次いで、多数のスライドに行ない得る。
Ｂ．富化手順
１．２０μｌの富化ビーズを、エマルジョン誘導型ビーズ（２０μｌ）を含有するチューブに添加。
穏和なピペッティングによりビーズ混合物を再懸濁（または概算鋳型陽性エマルジョンビーズごとに２～３　富化ビーズ増大を与える比を使用）。
２．ビオチン化Ｐ２－富化プライマーでコーティングされた富化ビーズを用いる場合、ビーズ混合物を６５Ｃで２分間インキュベート。チューブを氷に１０分間取り出す。
注：最初の実験は、１００－サイクルＰＣＲ（例えば、Ｐ２ＰＣＲ）に使用されるプライマー配列を含有する富化ビーズの使用は、富化すると、プライマーを含有するビーズを富化能力により効率的が低くなり得ることを示した。液滴中の２量体種誘導ビーズは鋳型がなかった。上記のＰ２－富化プライマーが負荷された富化ビーズを用いる場合、この短鎖プライマーのＴｍの低下によりビーズ混合物を５０Ｃ　２分間インキュベートする。
３．ビーズ混合物を、３００μｌ　６０％グリセロール溶液を含む１．５ｍｌエッペンドルフチューブに重層。
４．１３，０００ｒｐｍで１分間遠心分離
５．スピン後、陰性ビーズは、チューブの底部にペレット化する。結合された鋳型ビーズを含む富化ビーズは、グリセロール相上部に浮遊する。上相ビーズ集団を改修し、きれいな１．５ｍｌエッペンドルフチューブに移す。
注：チューブの底部ペレット化したビーズ（鋳型なしのビーズ）は洗浄し、鋳型陽性ビーズで概略を示した同じ洗浄レジメン後、磁石を用いて解析され得る。

６．ビーズを相から引き出すため、１ｍｌ　ヌクレアーゼ無含有Ｈ２Ｏを添加し、グリセロール濃度を希釈する。穏和なピペッティングを用いてビーズ混合物を再懸濁する。１３，０００ｒｐｍで１分間スピンする。
７．スピン後、上清みを除去し、２回１００μｌ　ＴＥを用いて洗浄。
８．１００μｌ　融解溶液を添加し、ビーズペレットを洗浄。チューブを５分間室温で回転。
９．さらに１００μｌ　融解溶液を添加、鋳型ビーズを磁石を用いて単離。
１０．２回１００μｌ　ＴＥ洗浄により非磁石性富化ビーズを除去し、磁石でＤＮＡビーズを富化ビーズから離す。
１１．鋳型ビーズを１０～２０μｌ　１×ＴＥに再懸濁。ビーズ凝集しているようであれば、１×ＰＣＲ－Ｂバッファー中に含むで希釈。
１２．鋳型含有ビーズは、他の富化集団でプールされ得、次の実施例で記載のようにしてスライド上に負荷し得る。
【0363】
（実施例１５：半固相支持体に固定化された微粒子アレイ調製方法）
この実施例では、スライド上に位置する半固相支持体内に自身に結合された鋳型微粒子が固定化され（例えば、包埋され）スライドの調製を記載する。かかるスライドは、ポロニルスライドとよばれ得る。この実施例に使用される半固相支持体は、ポリアクリルアミドである。プロトコルの一例では、ポリメラーゼ分子を鋳型付近に捕捉して増幅を増強する方法を用いる。
【0364】
スライドの調製
Ａ．ガラススライド：結合シラン処理
結合シラン処理により、ガラススライド表面へのアクリルアミドゲル結合が容易になる。スライドは、使用前に結合シラン処理で前処理するのがよい。
注：
＊＊　結合シラン処理溶液を化学的結合で保存。
＊＊　結合シラン処理は刺激性である。溶液調製時は化学的に作業する。
＊＊　　ストック結合シラン処理溶液がきれないように確保する。
＊＊　ラックへの出し入れの移動時は、スライドの表面に触れないようにする。
【0365】
結合シラン処理溶液の調製：
１．１－Ｌ　プラスチック容器内に添加：
１　Ｌ　ｄＨ２０，１　撹拌バー
２２０ｕｌ濃酢酸を添加（ｐＨ３．５にする）。４ｍｌ結合シラン試薬混合溶液を＞１５分間、撹拌プレートを用いて添加。
スライド処理：
２．スライド（同じ方向に向ける）を裏返したプラスチック　３８４ウェルプレートに負荷。
３．ｄH_2Ｏでリンス処理によりスライド洗浄、ウェルをドレイン。
４，１００％エタノールでリンス。ウェルをドレイン。
５．再度ｄH_2Ｏでリンス、ウェルをドレイン、ベンチをを有する組織培養フード内に配置、ＵＶ光照射。スライドを洗浄して乾燥させる（約３０分）。
６．プレートをプラスチック容器内に配置し、スライドを結合シラン処理溶液で覆う。
７．溶液およびスライドを１時間反応させる。容器を断続的に撹拌し、ガラスへの結合シラン処理の均一なコーティングを確保。
８．インキュベーション後、スライドを３回ｄＨ２Ｏでリンス処理。
９．１回１００％エタノールでリンス処理。ウェルをドレイン。
１０．使用前にスライドを充分乾燥させる。
１１．結合シラン処理スライドをデシケータ内で乾燥。

【0366】
Ｂ．アクリルアミド系スライド（小さいマスク）
・　非捕捉プロトコル
１．すべての試薬氷上に配置。以下の冷却試薬を１．５ｍｌエッペンドルフチューブに添加：
【0367】
【化8】

試薬	amt (μl)	
	2 スライド	1 スライド
1x TE	13	6.5
ビーズ (1-3M, １×ＴＥ中で希釈)	10	5
Rhinohide	1	0.5
40% アクリルアミド:Bis (19:1, F/S)	5	2.5
TEMED (5%, １×ＴＥ中)	2	1
APS (0.5%, 新たに作製)	3	1.5
合計	**34 μl**	**17 μl**

混合物を激しくピペッティングしてビーズをかき回す。ガラスカバースリップ下で１７μｌ／スライド負荷。
上下に重合。室温で６０分間.
きれいなかみそり刃でカバースリップを除く。
スライドを浸漬し。２回ＩＥバッファー中で１５分間洗浄（非結合ビーズを除去）。
包埋ビーズを有するスライドを４ＣでＩＥ保存。
２．フルオロフォアー標識配列決定プライマーを包埋ビーズ集団にハイブリダイズさせる。
洗浄１Ｅから１×ＰＣＲ－Ｂバッファーに、１×ＰＣＲ－Ｂバッファーを含有するＣｏｐｌｉｎ　ｊａｒ短時間浸漬させることによりスライドを平衡化させる。
３．１．５ｍｌエッペンドルフチューブ内で、１～６μｌ（１００μＭストック）プライマーを９９μｌ　１×ＰＣＲバッファーに添加する。アクリルアミドマトリックス上に、１００μｌ　プライマー溶液を滴下し、ガラスカバースリップを重層するか、ガスケットを封止する。
４．＜ＤＥＶＩＮ＞プログラムを用いてスライドを加熱することにより、プライマーを包埋されたビーズにハイブリダイズさせる。（６５Ｃで２分間、３０Ｃまで低速アニーリング）。スライド２回２分間洗浄ＩＥ中で洗浄。スライドは、ライゲーション系配列決定に供する準備ができている。
【0368】
・捕捉プロトコル
１．ｓｓＤＮＡ鋳型ビーズは、１Ｍ／μｌで調製する。［ポロニルスライド、４－５Ｍビーズ／スライド］。
２．ビーズ混合物を３０μｌ　１×ＰＣＲバッファー中に再懸濁。
３．１ｕｌの配列決定プライマー（１００μＭストック）を添加；充分混合
４．６５Ｃまで２分間加熱
５．氷上に５分間取り出す
６．８０μｌで１×ＴＥ洗浄
７．すべての固形分を磁石を用いて除去
８．以下に概要を示すようにして試薬を添加：
【0369】

【化9】

試薬	amt(μl) 2スライド
1x バッファー	1.5
10x バッファー	2.0
高濃度 (HC) 酵素	16.0
40% アクリルアミド:Bis (19:1, F/S)	14.4
Rhinohide	2.0
TEMED (5%, １×ＴＥ中)	2.0
APS (0.5%, 新たに作製)	1.5
合計	39.4 μl

混合物を激しくピペッティングしてビーズをかき回す。ガラスカバースリップ下で１７μｌ／スライド負荷。

９．重合、　好ましくは、上下にする（例えば、ＭＪ　Ｒｅｓｅａｒｃｈ　Ｔｅｔｒａｄ　ＰＣＲ装置で＜Ｐｏｌ－１＞サイクリングプロフィールを使用）

１０．かみそり刃でカバースリップを除くスライドを２回１Ｅバッファー中に１０分間浸漬および洗浄（非結合ビーズを除去するため）。

１１．ポロニルスライドはライゲーション系配列決定に供する準備ができている。

１２．包埋されたビーズを有するポロニルスライドは、ガスケット内に４Ｃで洗浄ＩＥ中に保存され得る。

【0370】

（実施例１６：固相支持体に結合された微粒子アレイ調製方法）

この実施例では、その上面に自身に結合された鋳型を有する微粒子が固相支持体に結合されたスライド調製を記載する。

【0371】

１．反応性ＮＨＳを有するポリマーテザーを用いて調製されたガラススライドを－２０Ｃで保存する（スライドH，製造番号１０７０９３６；Ｓｃｈｏｔｔ　Ｎｅｘｔｅｒｉｏｎ；Ｓｃｈｏｔｔ　Ｎｏｒｔｈ　Ａｍｅｒｉｃａ，Ｉｎｃ．，Ｅｌｍｓｆｏｒｄ，ＮＹ）

２．乾燥剤の存在下、使用前にスライドを室温に平衡化させる。

【0372】

３．５０ｍｌ　１×ＰＢＳ（３００ｍＭリン酸ナトリウム，ｐＨ８．７）中で５分間スライドを洗浄する。洗浄を２回繰り返す。

【0373】

４．スライドを溶液から取り出し、接着性ガスケットで覆う（試料負荷を可能にするため）。

【0374】

５．別のチューブで、アリコート１００～４０００００００タンパク質コートまたはＤＮＡコートビーズを１×ＰＢＳ，ｐＨ８．７に。ＤＮＡは、例えば、配列決定用のＤＮＡ鋳型であり得る。ＤＮＡは、例えば、ＮＨＳとの反応用のアミンリンカーを含むものであり得る。

【0375】

６．バッファー交換によってビーズ試料を３回１×ＰＢＳ，ｐＨ８．７で洗浄する。

【0376】

７．ビーズを１２５ｍｌ　１×ＰＢＳ，ｐＨ８．７中に再懸濁する。

【0377】

８．ビーズ溶液をスライドガスケット内に負荷し、スライド表面を均一にコートする。

【0378】

９．スライドを暗室チャンバ内に封鎖し、インキュベーションのために１～２時間室温で反応させる。

【0379】
１０．インキュベーション後、非結合ビーズ溶液を除去し、スライドを５０ｍｌ　１×ＴＥ（１０ｍＭ　Ｔｒｉｓ，１ｍＭ　ＥＤＴＡ，ｐＨ８）に移す。
【0380】
１１．１５分間／洗浄で定常的撹拌により、５０ｍｌ　１×ＴＥを用いて５回スライドを洗浄する。
【0381】
１２．スライドは、１×ＴＥ中に４Ｃで数週間保存され得る。
【0382】
１３．所望の場合は、ビーズ集団は、白色光（ＷＬ）を用いる明るい視野での画像解析により、または蛍光により、フルオロフォア系色素に結合された相補ＤＮＡオリゴヌクレオチドを用いて評価され得る。ＤＮＡ鋳型は、例えば、ライゲーション系配列決定を用いて配列決定され得る。
【0383】
図３３Ａは、自身にビーズが結合されたスライドの概略図を示す。ほんの少量の割合のＤＮＡ鋳型分子がスライドに結合されていることに注意されたい。１ミクロンビーズ（ＤｙｎａビーズＭｙＯｎｅ　ストレプトアビジンビーズ；Ｄｙｎａｌ　Ｂｉｏｔｅｃｈ，Ｉｎｃ．，製造番号６５０．０１）を使用した。しかしながら、広範なさまざまなビーズた使用され得る。
【0384】
図３３Ｂは、スライドに結合されたビーズ集団を示す。下側パネルは、白色光下でのスライドの同じ領域（左）および蛍光顕微鏡検査を示す。上側パネルは、ビーズ密度の範囲を示す。
【0385】
（実施例１６）　ゲル非含有ビーズに基づくアレイを使用した、オリゴヌクレオチド伸長およびライゲーションによる配列決定
本実施例では、ビオチンとストレプトアビジンの相互作用を介して基材（ガラススライド）に付着した微粒子のアレイの調製について記載し、ライゲーション、開裂および検出のサイクルによる配列決定の成功を実証する。ビオチン化鋳型が付着した微粒子は、エマルジョンＰＣＲを使用して調製し、後述の通り、半固形培地の非存在下で、ＰＥＧを含む連結を介して、ストレプトアビジンで官能基化した基材に付着させた。本方法では、増幅前にビオチン化プライマーを付着させたストレプトアビジンコートビーズを使用する。生産的な鋳型増幅が行われた粒子の増幅と濃縮の後、鋳型をビオチン化した。次いで、ビオチン化した鋳型が付着した微粒子を、ストレプトアビジンコートスライドとともにインキュベートした。したがって、本方法では、ビオチンとストレプトアビジンの連結を２回使用した。その他の手法では、プライマーを微粒子に連結させるか、あるいは増幅させた鋳型を基材に連結させる代替手段が使用されると考えられる。
【0386】
材料および方法：
ＢＡＣ　Ｅｃｏ　ｖ２．１ビーズの調製
ＭｙＯｎｅストレプトアビジンビーズ（１ミクロン）をビオチン化Ｐ１プライマーでコーティングし（図を参照）、エマルジョンＰＣＲにおいて、本発明者等のＢＡＣ－Ｅｃｏ（ｖ２．１）ライブラリーの鋳型が付着したビーズ群を作り出すために使用した。エマルジョンを破壊し、ビーズを標準的な方法でエクソヌクレアーゼにより精製および処理した。完全に伸長したＰＣＲ産物を有するビーズを、Ｐ２濃縮オリゴで被覆した濃縮ビーズに結合させることにより濃縮した（図を参照）。濃縮したビーズの溶液中の挙動を改善するために、これらのビーズをビオチン化Ｐ１オリゴとともにインキュベートして、ストレプトアビジンコーティングを露出したビーズの部分を被覆した。
【0387】
スライドへのＢＡＣ　Ｅｃｏ　ｖ２．１ビーズの沈着

ｓｓＤＮＡを含有する濃縮したＢＡＣ－Ｅｃｏ　ｖ２．１ビーズを、ストレプトアビジンでコーティングしたＯｐｔｉ－Ｃｈｅｍスライド（Ａｃｃｅｌ８　Ｔｅｃｈｎｏｌｏｇｙ　Ｃｏｒｐｏｒａｔｉｏｎ）上に沈着した。本プロセスの準備として、ビーズをターミナルトランスフェラーゼ（Ｎｅｗ　Ｅｎｇｌａｎｄ　Ｂｉｏｌａｂｓ）およびビオチン－１１－ｄｄＡＴＰ（Ｐｅｒｋｉｎ　Ｅｌｍｅｒ）とともにインキュベートし、ビオチン部分をＤＮＡ鋳型分子の３’端に共有結合的に付着させた。ビーズを等しい数のＭｙＯｎｅ　Ｃａｒｂｏｘｙｌｉｃ　Ａｃｉｄビーズ（Ｄｙｎａｌ）と混合し、５ｍＭのトリス塩酸塩（ｐＨ８．０）、５ｍＭのＥＤＴＡ、０．０００５％のＴｒｉｔｏｎ　Ｘ－１００、および１０％のＰＥＧ８０００（Ａｍｅｒｉｃａｎ　Ｂｉｏａｎａｌｙｔｉｃａｌ）を含有する沈着緩衝液に入れた。懸濁液をＣｏｖａｒｉｓ　Ｓ２ソニケーターで短時間超音波処理し、ストレプトアビジンでコーティングしたＯｐｔｉ－Ｃｈｅｍスライド（Ａｃｃｅｌ８　Ｔｅｃｈｎｏｌｏｇｙ　Ｃｏｒｐｏｒａｔｉｏｎ）上に沈着させた。スライドをＴＥ緩衝液で３回洗浄し、使用前に圧縮空気で乾燥させた。懸濁液をＬｉｆｔｅｒＳｌｉｐ（Ｅｒｉｅ　Ｓｃｉｅｎｔｉｆｉｃ　Ｃｏｍｐａｎｙ）で被覆し、スライド上に平坦な水層を生成し、蒸発を減少させた。これらのスライドを高湿度チャンバーに入れて室温にて４５分間インキュベートし、端部の蒸発を減少させつつ、ビーズを表面に定着させ、結合させた。ＴＥ緩衝液を充填したトレーにスライドを上下逆にして浸漬することにより、カバースリップを除去した。約１分間軽く攪拌すると、ほとんどのカルボン酸ビーズが除去された（別の実験で示された通り）。これらのスライドを即座にアセトンに浸漬し、圧縮空気を使用して乾燥させた。

【0388】

ゲル非含有スライド上でのサイクル式ライゲーション配列決定で使用した試薬は、Ｒｅｓｅｔ緩衝液を除けばアクリルアミド系ゲルの場合と同じであった。非ゲルアレイでは、１０ｍＭのＮａＯＨおよび０．１％のドデシルスルホン酸ナトリウム（Ｆｌｕｋａ）を含有するアルカリ系Ｒｅｓｅｔ緩衝液を使用した。図３８および図３９に示す通り、３００パネルゲル非含有アレイ（約１８×１８ｍｍ）に濃縮ＢＡＣ－Ｅｃｏライブラリービーズを播種し、自動小型フローセル器に入れ、アルカリ性Ｒｅｓｅｔに５０回暴露させて、ゲル非含有環境におけるビーズの安定性を検証した。５０サイクルの流動プログラムの後、ゲル非含有アレイはパネル１枚当たり２６，０００個を超えるビーズを含有した（４メガピクセルカメラ）。次いで、サイクル式ライゲーションおよび開裂を使用して、ゲル非含有アレイの配列決定を行った。各蛍光チャネルの高いＲＦＵ値によって証明されているように、サイクル１のデータの評価によって、本発明者等の２塩基４色プローブセットの効率的なライゲーションが裏付けられた。その後、ビーズ群をベースコールし、スペクトル純度プロットにプロットし、Ｓａｔａｙ分析および密度プロット評価によって優れた配列決定性能を示した。

【0389】

均等物および範囲

当該技術分野業者には、単に常套的な実験手法を用いることで、本明細書に記載した本発明の具体的な実施形態に対する多くの均等物が認識され、または確認できよう。本発明の範囲は、上記の記載に限定されることは意図されるのではなく、添付の特許請求の範囲に示されるとおりである。以下の特許請求の範囲において、「ａ」、「ａｎ」および「ｔｈｅ」などの冠詞は、そうでないと特に記載のない限り、あるいは文脈から明白にそうでない場合以外は、１つではなく、１つ以上を意味し得る。１つの群内の１つ以上のメンバー間の「または」を含む請求項または詳細説明は、そうでないと特に記載のない限り、あるいは文脈から明白にそうでない場合以外は、１つ、１つより多く、またはその群のメンバーすべてが、所与の生成物またはプロセスに存在し、使用され、あるいは、関連する場合が満たされるとみなされる。請求項における「場合により」の使用は、任意の特徴が存在する実施形態と、任意の特徴が存在しない実施形態とを本発明が包含することを示している。

【0390】

さらにまた、本発明は、記載の１つ以上の請求項からの１つ以上の限定、要素、原因、記述用語などが別の請求項に導入されたあらゆる変形例、組合せおよび置換を包含することを理解されたい。特に、別の請求項に従属する任意の請求項は、ベースとなる同じ請求項に従属する任意の他の請求項に見られる１つ以上の限定を含むように変形され得る。

【0391】

また、任意の１つ以上の実施形態は、具体的な除外が本明細書に明白に示されていない場合であっても、特許請求の範囲から明確に除外されることがあり得ることを理解されたい。また、本明細書および／または特許請求の範囲に配列決定に有用な試薬（例えば、鋳型、ミクロスフィア、プローブ、プローブファミリーなど）が開示されている場合、かかる開示はまた、該試薬を、当業者が別なふうに理解し得る場合、または本発明に特に記載のない限り、本明細書に開示された具体的な方法、または当該技術分野で知られた他の方法のいずれかに従って用いる配列決定方法も包含することを理解されたい。また、本明細書および／または特許請求の範囲に配列決定法が開示されている場合、当業者が別なふうに理解し得る場合、または本明細書においてかかる方法におけるその試薬の使用が明白に除外されている場合を除き、本明細書に開示された任意の１種類以上の試薬が該方法に使用され得る。さらに、配列決定に有用な特定の成分が本明細書および特許請求の範囲に開示されている場合、本発明は、その試薬の作製方法もまた包含することを理解されたい。用語「成分」は、配列決定に用いられる任意の要素、例えば、鋳型、自身に結合された鋳型を有する微粒子、ライブラリーなどをいうために広義で使用される。さらにまた、図面は、本明細書に組み込まれた一部分であり、本発明は、図面に示された構造、例えば、自身に結合された鋳型を有する微粒子、および図面に開示された方法を含む。

【0392】

本明細書において範囲を示す場合、両端を含む。さらにまた、ことを理解されたい。特に記載のない限り、あるいは文脈から明白にそうでない場合以外、および当業者に理解される場合、範囲として標示される値は、本発明の異なる実施形態において記載された範囲内の任意の具体的な値または下位範囲が、特に文脈においてそうでないと明白に記載されていない限り、該範囲の下限の単位の１０分の１まで想定され得る。

【図面の簡単な説明】

【0393】

【図1A】図１Ａは、初期化後、２サイクルの伸長、ライゲーションおよび同定を概略的に示す。

【図1B】図１Ｂは、伸長が鋳型の自由末端から支持体に向かって内方に進行する実施形態における初期化後、２サイクルの伸長、ライゲーションおよび同定を概略的に示す。

【図2】図２は、オリゴヌクレオチドプローブに対する色の割り当てのスキームを示し、この場合、プローブの３’塩基の正体が、フルオロフォアの色を同定することにより決定される。

【図3A】図３Ａは、鋳型の結合領域内の異なる位置における初期オリゴヌクレオチドのハイブリダイゼーション後、伸長プローブのライゲーションにより得られる伸長された二本鎖を概略的に示す。

【図3B】図３Ｂは、鋳型分子を６塩基ごとに読み出されるように設計された伸長プローブを用いて伸長、ライゲーションおよび切断方法を使用することにより、連続配列（配列番号７として配列表に記載される）の合成を概略的に示す。

【図4A】図４Ａは、５’－Ｓ－ホスホロチオレート結合（３’－Ｏ－Ｐ－Ｓ－５’）　を示す。

【図4B】図４Ｂは、３’－Ｓ－ホスホロチオレート結合（３’－Ｓ－Ｐ－Ｏ－５’）　を示す。

【図5A】図５Ａは、３’－Ｏ－Ｐ－Ｓ－５’ホスホロチオレート結合を有する伸長プローブを用いた５’→３’方向における配列決定のための１サイクルの伸長、ライゲーションおよび切断を概略的に示す。

【図5B】図５Ｂは、３’－Ｓ－Ｐ－Ｏ－５’ホスホロチオレート結合を有する伸長プロー

ブを用いた３’→５’方向における配列決定のための１サイクルの伸長、ライゲーションおよび切断を概略的に示す。
【図6】図６Ａ～６Ｆは、単一の鋳型において行なわれた数個の配列決定反応のより詳細な概略図である。この反応では、鋳型の異なる部分に結合する初期オリゴヌクレオチドを用いている。図６Ｆに示される配列は、配列番号８として記載される。
【図7】図７は、ｄＡおよびｄＧの３’－ホスホロアミデートの合成スキームを示す概略である。
【図8-1】図８Ａ～８Ｅは、ホスホロチオレート結合を含有する伸長プローブの成功裏のライゲーションおよび切断の２サイクルを示すゲルシフトアッセイの結果を示す。示される配列は、配列番号１０、６２および８３として記載される。
【図8-2】図８Ｆは、ＤＮＡリガーゼによるライゲーションの機構の概略図を示す。
【図9】図９は、縮重イノシン含有オリゴヌクレオチドプローブのライゲーション効率を示すゲルシフトアッセイの結果を示す。示される配列は、配列番号１４および６２として記載される。
【図10】図１０は、多数の鋳型における縮重イノシン含有オリゴヌクレオチドプローブのライゲーション効率を示すゲルシフトアッセイの結果を示す。
【図11】図１１は、３’→５’伸長での２つのＤＮＡリガーゼ（Ｔ４　ＤＮＡリガーゼおよびＴａｑＤＮＡリガーゼ）の各々の忠実度を評価するために行なった解析の結果を示す。示される配列は、配列番号１６および６２として記載される。
【図12】図１２は、（Ａ）　縮重イノシン含有オリゴヌクレオチドプローブおよびライゲーション反応の直接配列決定解析のライゲーション効率を示す（Ｂ）　オリゴヌクレオチドプローブライゲーションにおけるＴ４　ＤＮＡリガーゼの忠実度を評価するために行なったゲルシフトアッセイの結果を示す。結果をパネルＣ～Ｆに表で示す。図１２Ｂに示される配列は、配列番号５６～６２および８４として記載される。
【図13】図１３Ａ～１３Ｃは、ビーズ系鋳型をスライド上のポリアクリルアミドゲル内に包埋した場合のゲル内ライゲーションを示す実験の結果を示す。図１３Ａは、ライゲーション反応の概略を示す。ゲル内ライゲーション反応は、Ｔ４　ＤＮＡリガーゼの非存在下（Ｂ）および存在下（Ｃ）で行なった。示される配列は、配列番号６２および６６として記載される。
【図14A】図１４Ａは、第１の増幅プライマーが結合されたビーズ上で、蛍光標識された第２の増幅プライマーおよび過剰の鋳型を用いて行なったエマルジョンＰＣＲ反応の画像を示す。
【図14B】図１４Ｂ（上）は、Ｃｙ３－標識オリゴヌクレオチドがハイブリダイズする鋳型結合された結合されたビーズがポリアクリルアミドゲル固定化されたスライドの一部分の蛍光画像を示す（このスライドは、異なる実験に使用したが、ここで使用したスライドの典型である）。図１４Ｂ（下）は、ポリアクリルアミド溶液を封入するためのテフロン（登録商標）マスクを設けたスライドの概略図を示す。
【図15】図１５は、プローブの特異性および選択性の課題に取り組むために設計され３組の標識されたオリゴヌクレオチドプローブを示し、また、１組の４つのスペクトルにより分離可能な標識の励起および発光の値を示す。
【図16】図１６は、オリゴヌクレオチドプローブの４色スペクトルの正体を確認する実験の結果を示す。４つの特有の単鎖鋳型集団を含むスライド（Ａ）をハイブリダイゼーションに供し、４種類の特有のフルオロフォアプローブを含有するオリゴヌクレオチドプローブ混合物を用いたライゲーション反応物を、明るい光の下で、および（Ｂ）　４種類の帯域通過フィルターを用いた蛍光励起によりライゲーションの前と後で画像化した。個々の集団は擬似カラー表示した（Ｃ）。スペクトルの正体（これは、最小シグナル重複を示した）を（Ｄ）にプロットする。示される配列は、配列番号６２～７０として記載される。
【図17】図１７は、オリゴヌクレオチド伸長プローブのライゲーション特異性を確認する実験からの結果を示す。図１７（Ａ）は、ライゲーションの概略的概要を示す。図１７（Ｂ）は、明るい光での画像であり、図１７（Ｃ）は、ライゲーション後のポリアクリルア

ミドゲル内に包埋されたビーズ集団の対応する蛍光画像である。図１７（Ｄ）は、ライゲーション前（ｐｒｅ）または後（ｐｏｓｔ）の各標識から検出された蛍光を示す。示される配列は、配列番号６３および７８として記載される。

【図18】図１８は、オリゴヌクレオチド伸長プローブのライゲーションの特異性および選択性を確認する別の実験からの結果を示す。図１８（Ａ）は、ライゲーションの概略的概要を示す。図１８（Ｂ）は、明るい光での画像であり、図１８（Ｃ）は、ライゲーション後のポリアクリルアミドゲル内に包埋されたビーズ集団の対応する蛍光画像である。図１８（Ｄ）は、ライゲーション頻度の予測対観察結果を示し、集団内の特定の伸長プローブの割合に基づいて予測した頻度と、観察された頻度との高い相関性を示す。図１８（Ａ）に示される配列は、配列番号６２、６６および７１～７８として記載される。

【図19】図１９は、オリゴヌクレオチド伸長プローブプールを含有する縮重した普遍塩基を用いて、特異的および選択的なゲル内ライゲーションがもたらされ得ることを確認する実験からの結果を示す。図１９（Ａ）は、ライゲーションの概略的概要実験を示し、ライゲーション後の４種類の差次的に標識された縮重イノシン含有プローブプールを示す。図１９（Ｂ）明るい光での画像であり、図１９（Ｃ）は、ライゲーション後のポリアクリルアミドゲル内に包埋されたビーズ集団の対応する蛍光画像である。図１９（Ｄ）は、ライゲーション頻度の予測対観察結果を示し、集団内の特定の伸長プローブの割合に基づいて予測した頻度と、観察された頻度との高い相関性を示す。図１９（Ｅ）は、生の未処理のデータおよびビーズシグナル値の上部９０％を表すフィルター処理データの散布図を示す。図１９（Ａ）に示される配列は、配列番号６２、６６、６７、６８、７０および７９～８２として記載される。

【図20】図２０は、鋳型に対する初期オリゴヌクレオチド（プライマー）のハイブリダイゼーションとストリッピングの連続的サイクルにおいて検出されたシグナルを示すチャートである。図に示されるように、１０サイクルで生じたシグナル損失は最少であった。

【図21】図２１は、配列情報を、例えば、実質的に平面状の支持体内または該支持体上のアレイ状の鋳型から収集するために使用され得る自動配列決定システムの写真である。また、該システムの種々の構成部品の動作を制御するため、収集した画像データを処理および保存するため、ユーザーインターフェースを提供するためなどの専用のコンピュータを示す。図の下側部分に、重量測定気泡移動（ｇｒａｖｉｍｅｔｒｉｃ　ｂｕｂｂｌｅ　ｄｉｓｐｌａｃｅｍｅｎｔ）が達成されるように配設されたフローセルの拡大図を示す。

【図22】図２２は、実質的に平面状の支持体内または該支持体上のアレイ状の鋳型の配列を決定するために使用され得るハイスループット自動配列決定機器の概略図を示す。

【図23】図２３は、アライメントの不一致の散布図を示し、３０フレームで最少の不一致を示す。

【図24】図２４Ａ～Ｉは、さまざまな異なる方向から見た本発明フローセルまたはその一部分の概略図を示す。

【図25A】図２５Ａは、２ヌクレオチド長である拘束部分を含む部分拘束プローブを含むプローブファミリーの好ましいコレクションの例示的なコード化を示す。

【図25B】図２５Ｂは、プローブファミリーの好ましいコレクション（上側パネル）およびライゲーション、検出および切断のサイクル（下側パネル）　を示す。

【図26】図２６は、２ヌクレオチド長である拘束部分を含む拘束プローブを含むプローブファミリーの別の好ましいコレクション例示的なコード化を示す。

【図27A】図２７Ａは、表１に示すプローブファミリーの２４個の好ましいコレクションを図式的に規定するための代替法を表す。

【図27B】図２７Ｂは、表１に示すプローブファミリーの２４個の好ましいコレクションを図式的に規定するための代替法を表す。

【図27C】図２７Ｃは、表１に示すプローブファミリーの２４個の好ましいコレクションを図式的に規定するための代替法を表す。

【図28】図２８は、プローブが、２ヌクレオチド長である拘束部分を含むプローブファミリーのやや好ましいコレクションを示す。

【図29A】図２９Ａは、拘束部分３ヌクレオチド長を有するプローブを含むプローブファミリーのコレクションのための拘束部分を作製するために使用され得るダイアグラムを示す。
【図29B】図２９Ｂは、拘束部分３ヌクレオチド長を有するプローブを含むプローブファミリーのコレクションのための拘束部分を、プローブファミリーの２４個の好ましいコレクションから作製するために使用され得るマッピングスキームのダイアグラムを示す。
【図30】図３０は、配列決定がプローブファミリーのコレクションを用いて行なわれる方法を示す。好ましい組のプローブファミリーを用いた実施形態を示す。
【図31A】図３１Ａ～３１Ｃは、配列決定が、プローブファミリーの第１のコレクションを用いて候補配列を作製し、プローブファミリーの第２のコレクションを用いてデコード化することで行なわれる方法を示す。
【図31B】図３１Ｂは、配列決定が、プローブファミリーの第１のコレクションを用いて候補配列を作製し、プローブファミリーの第２のコレクションを用いてデコード化することで行なわれる方法を示す。
【図31C】図３１Ｃは、配列決定が、プローブファミリーの第１のコレクションを用いて候補配列を作製し、プローブファミリーの第２のコレクションを用いてデコード化することで行なわれる方法を示す。
【図32】図３２は、配列決定が、プローブファミリーのやや好ましいコレクションを用いて行なわれる方法を示す。
【図33】図３３Ａは、自身にビーズが結合されたスライドの概略図を示す。ＤＮＡ鋳型がビーズに結合されている。図３３Ｂは、スライドに結合されたビーズ集団を示す。下側パネルは、白色光下（左）および蛍光顕微鏡検査でのスライドの同じ領域を示す。上側パネルは、ビーズ密度の範囲を示す。
【図34】図３４Ａ～３４Ｃは、核酸の個々の集団としての核酸断片（鋳型）内に存在する１対のタグの両方のタグを増幅し、これを、増幅プロセスにより微粒子に捕捉するためのスキームを示す。
【図35-1】図３５Ａおよび３５Ｂは、図３５のスキームのプライマー設計および増幅の詳細を示す。明確にするために核酸断片（鋳型）の両方の鎖を示す。同じ配列を有するプライマーおよびプライマー結合領域を同じ色で示す。例えば、Ｐ１を紺色で表し、微粒子上および溶液中に存在するプライマーＰ１が、鋳型の標示した鎖と対応させて着色した部分と同じ配列を有することを示す。Ｐ１と標示した鋳型の紺色領域は、対応するプライマー（Ｐ１）が、実際には他方の鎖の相補的部分に結合し、プライマーＰ１と同じ配列を有するものであっても、プライマー結合領域とよばれ得る。
【図35-2】図３５Ｃおよび３５Ｄは、それぞれ、図３５Ａおよび３５Ｂの方法によって生成された微粒子に結合された第１および第２のタグの配列決定を示す。
【図36】図３６Ａは、対合末端ライブラリーの鋳型分子を示し、ライブラリーのメンバーに共通する、鋳型の前方アダプター部分、後方アダプター部分、および内部アダプター部分にハイブリダイズするブロッキングオリゴヌクレオチドを示している。図の下部には、アダプターとブロッキングオリゴヌクレオチドの例示的な配列が示されている。図３６Ａ～図３６Ｃの「ｄｄＢａｓｅ」とは、ジデオキシヌクレオシドを示す。「固有のＤＮＡ配列」とは、配列決定する標的領域を表す。図３６Ｂは、断片ライブラリーの鋳型分子を示し、ライブラリーのメンバーに共通する、鋳型分子の前方アダプター、後方アダプター、および内部アダプター部分にハイブリダイズするブロッカーオリゴヌクレオチドを示している。図の下部には、アダプターおよび相補性ブロッキングオリゴヌクレオチドの例示的な配列が示されている。図３６Ｃは、鋳型分子にローリングサークル増幅（ＲＣＡ）が行われたライブラリーの分子を示す。ＲＣＡは、鋳型分子の固有部分（２）、ならびにアダプター領域（１）とパドロック領域（３）の複数の複製を作成する。この図は、ライブラリーのメンバーに共通する、鋳型のアダプター部分およびパドロック部分にハイブリダイズするブロッキングオリゴヌクレオチドを示している。
【図37】いくつかのパドロックプローブ配列と、ＲＣＡを使用した鋳型分子の合成後にパ

ドロック領域をブロックするとされるオリゴヌクレオチドの例示的な配列を示す。
【図38】半固形培地を使用せずに基材上に作成される微粒子のアレイ（ゲル非含有微粒子アレイ）を示す。
【図39】ゲル非含有微粒子アレイを使用した、ライゲーションに基づく配列決定の結果を示す。
【図40】表面上に位置する微粒子の概略図を示し、鋳型の伸長によって生じるとされる接触パッチと核酸コロニーの予想されるサイズを例示する。

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2007/066931

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, EMBASE, BIOSIS, MEDLINE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO 2006/084132 A (AGENCOURT BIOSCIENCE CORP [US]; MCKERNAN KEVIN [US]; BLANCHARD ALAN [U) 10 August 2006 (2006-08-10) the whole document	1-12
X	WO 2005/040425 A (ISIS INNOVATION [GB]; SHCHEPINOV MIKHAIL S [GB]; MIR KALIM [GB]) 6 May 2005 (2005-05-06) page 30 - page 72; claims 27-57	1-3,6-12
X	WO 98/31831 A (BRAX GENOMICS LTD [GB]; SCHMIDT GUNTER [GB]; THOMPSON ANDREW HUGIN [GB) 23 July 1998 (1998-07-23) the whole document	1,6-12
X	US 6 306 597 B1 (MACEVICZ STEPHEN C [US]) 23 October 2001 (2001-10-23) the whole document	1-3,6-12

-/--

[X] Further documents are listed in the continuation of Box C.　　[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
3 April 2008	01/08/2008

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Botz, Jürgen

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

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2007/066931

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 96/33205 A (SPECTRAGEN INC [US]) 24 October 1996 (1996-10-24) claims 1-17; figures 1-4	1-3,6-12
X	WO 97/46704 A (LYNX THERAPEUTICS INC [US]) 11 December 1997 (1997-12-11) the whole document	1-12
Y	US 5 714 330 A (BRENNER SYDNEY [GB] ET AL) 3 February 1998 (1998-02-03) the whole document	1-12
Y	METELEV VALERI G ET AL: "New chemically reactive dsDNAs containing single internucleotide monophosphoryldithio links: Reactivity of 5'-mercapto-oligodeoxyribon ucleotides" NUCLEIC ACIDS RESEARCH, vol. 29, no. 19, 1 October 2001 (2001-10-01), pages 4062-4069, XP002474826 ISSN: 0305-1048 the whole document	1-12

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

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2007/066931

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

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

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

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

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

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

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

see additional sheet

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

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

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

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

1-12 (completely)

Remark on Protest

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

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

☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet (2)) (April 2005)

International Application No. PCT/US2007 /066931

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-12 (completely)

A method for identifying a sequence of nucleotides in a template polynucleotide, the method comprising the steps of: (a.) extending an initializing oligonucleotide along the template polynucleotide by ligating an oligonucleotide probe thereto to form an extended duplex, wherein the oligonucleotide probe is attached to a microparticle, wherein the micorparticle is attached to a substrate, and wherein the microparticle is not immobilized in a semi-solid support;
(b.) identifying one or more nucleotides of the polynucleotide; and repeating steps (a.) and (b.) until the sequence of nucleotides is determined.

---

2. claims: 13 - 30 (completely)

A method for determining a sequence of nucleotides in a template polynucleotide the method comprising the steps of: (a) providing a probe-template duplex comprising a probe hybridized to a template polynucleotide, the probe having an extendable terminus; (b) ligating an extension oligonucleotide probe to said extendable terminus to form an extended duplex containing an extended oligonucleotide probe, wherein the oligonucleotide probe is attached to a microparticle, wherein the microparticle is attached to a substrate, and wherein the microparticle is not immobilized in a semi-solid support; (c) identifying, in the extended duplex, at least one nucleotide in the template polynucleotide that is either (1) complementary to the just-ligated extension probe or (2) a nucleotide residue in the template polynucleotide which is immediately downstream of the extended oligonucleotide probe; (d) generating an extendable terminus on the extended oligonucleotide probe, if an extendable terminus is not already present, such that the terminus generated is different from the terminus to which the last extension probe was ligated; and (e) repeating steps (b), (c) and (d) until a sequence of nucleotides in the template polynucleotide is determined.

---

3. claims: 31 - 37 (completely)

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

A method for identifying a sequence of nucleotides in a template polynucleotide, the method comprising the steps of: (a) providing a template polynucleotide attached to a microparticle that is immobilized in or on a semi-solid support or is attached to a substantially planar, rigid substrate; (b) contacting the template polynucleotide with a blocking oligonucleotide; (c) extending an initializing oliognucleotide along the template polynucleotide by ligating an oligonucleotide probe thereto to form an extendable duplex, wherein the oligonucleotide probe optionally comprises a scissile linkage; (d) identifying one or more nucleotides of the polynucleotide; and (e) repeating steps (c) and (d) until the sequence of nucleotides is determined.

---

4. claims: 38-44 (completely)

A method for determining a sequence of nucleotides in a template polynucleotide, the method comprising the steps of: (a) providing a probe-template duplex comprising a probe hybridized to a template polynucleotide, the probe having an extendable terminus and the template having a blocking oligonucleotide hybridized thereto, the probe-template duplex being attached to a microparticle that is embedded in or on a semi-solid support or attached to a substrate; (b) ligating an extension oligonucleotide probe to said extendable terminus, to form an extended duplex containing an extended oligonucleotide probe wherein the extension probe comprises a phosphorothiolate linkage; (c) identifying, in the extended duplex, at least one nucleotide in the template polynucleotide that is either (1) complementary to the just-ligated extension probe or (2) a nucleotide residue in the template polynucleotide which is immediately downstream of the extended oligonucleotide probe; (d) generating an extendable terminus on the extended oligonucleotide probe, if an extendable terminus is not already present, such that the terminus generated is different from the terminus to which the last extension probe was ligated; and (e) repeating steps (b), (c) and (d) until a sequence of nucleotides in the template polynucleotide is determined;

---

5. claims: 45 - 50 (completely)

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

A method for determining a sequence of nucleotides in a template polynucleotide, the method comprising the steps of: (a) amplifying a template polynucleotide molecule in a compartment of an emulsion in the presence of a microparticle so that a microparticle having a clonal population of template polynucleotides attached thereto is produced; (b) recovering the microparticle from the emulsion; (c) embedding the microparticle in or on a semi-solid support or attaching the microparticle to a substrate; (d) extending an initializing oligonucleotide along the template polynucleotide by ligating an oligonucleotide probe thereto to form an extended duplex, wherein the oligonucleotide probe comprises a scissile linkage; (c) identifying one or more nucleotides of the polynucleotide; and (f) repeating steps (d) and (e) until the sequence of nucleotides is determined.

---

6. claims: 51 - 84 (completely)

A method for determining information about a sequence of nucleotides in a template polynucleotide using a first collection of at least 2 distinguishably labeled oligonucleotide probe families, the method comprising the steps of: (a) extending an initializing oligonucleotide along the template polynucleotide by ligating an oligonucleotide probe thereto to form an extended duplex wherein the oligonucleotide probe is a member of the collection of distinguishably labeled oligonucleotide probe families and has a blocking oligonucleotide hybridized thereto; (c) detecting a label associated with the oligonucleotide probe; and (d) repeating steps (b) and (c) until an ordered list of probe family names is obtained; and (d) using the ordered list of probe family names to climinate one of more possibilities for the sequence of nucleotides.

---

7. claims: 85 - 87 (completely)

A method for determining information about a sequence of nucleotides in a template polynucleotide using a first collection of at least 2 distinguishably labeled oligonucleotide probe families, the method comprising steps (a) - (e) of claim 85 as filed;

---

8. claims: 88, 89 (completely)

A method for determining information about a sequence of nucleotides in a template polynucleotide using a first collection of oligonucleotide probe families, the method comprising steps (a) and (b) of claim 88 as filed;

International Application No. PCT/US2007 /066931

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

9. claims: 90 - 94 (completely)

A method of preparing a plurality of template polynucleotides comprising steps (a) and (b) of claim 90 as filed;

---

10. claims: 95 - 98 (completely)

A collection of components for preparing a population of microparticles, the collection comprising (a), (b) and (c) as specified within claim 95 as filed;

---

11. claims: 99 - 115 (completely)

A template comprising a population of substantially identical template molecules as specified within claim 99 as filed as well as a support or substrate having the population of claim 99 attached thereto, as well as a collection of templates of claim 99 wherein the templates comprise different segments of interest, as well as an array comprising a population of microparticles of claim 104, wherein the microparticles have templates comprising different segments of interest attached thereto, as well as a microparticle attached to the substrate, wherein the microparticle has a template attached thereto, as well as a population of microparticles as detailed within claim 115;

---

12. claims: 116 - 118 (completely)

A method for preparing an array as detailed within claim 116;

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2007/066931

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 2006084132	A	10-08-2006	AU	2006210553 A1	10-08-2006
			CA	2596496 A1	10-08-2006
			EP	1844162 A2	17-10-2007
			KR	20070112785 A	27-11-2007
			US	2008003571 A1	03-01-2008
WO 2005040425	A	06-05-2005	EP	1689881 A2	16-08-2006
			US	2007219367 A1	20-09-2007
WO 9831831	A	23-07-1998	AU	728805 B2	18-01-2001
			AU	5570298 A	07-08-1998
			CA	2277520 A1	23-07-1998
			EP	0960215 A1	01-12-1999
			IL	130886 A	19-02-2004
			JP	2001508308 T	26-06-2001
			NZ	336770 A	26-01-2001
			US	7094531 B1	22-08-2006
US 6306597	B1	23-10-2001	NONE		
WO 9633205	A	24-10-1996	AU	718937 B2	04-05-2000
			AU	5549096 A	07-11-1996
			CA	2218017 A1	24-10-1996
			EP	0871646 A1	21-10-1998
			JP	11503908 T	06-04-1999
			JP	2007282639 A	01-11-2007
			JP	2008086328 A	17-04-2008
			US	5750341 A	12-05-1998
			US	5969119 A	19-10-1999
WO 9746704	A	11-12-1997	AT	356221 T	15-03-2007
			AU	733782 B2	24-05-2001
			AU	3374097 A	05-01-1998
			CA	2256700 A1	11-12-1997
			CN	1230226 A	29-09-1999
			CZ	9803979 A3	14-07-1999
			DE	69737450 T2	29-11-2007
			EP	0923650 A1	23-06-1999
			HK	1021206 A1	15-06-2007
			HU	0003944 A2	28-03-2001
			JP	2000515006 T	14-11-2000
			NO	985698 A	08-02-1999
			PL	331513 A1	19-07-1999
US 5714330	A	03-02-1998	US	5856093 A	05-01-1999

Form PCT/ISA/210 (patent family annex) (April 2005)

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

(72)発明者　マッカーナン，　ケビン
アメリカ合衆国　マサチューセッツ　０１９４５，　マーブルヘッド，　ペクォート　ロード　１０
(72)発明者　ブランチャード，　アラン
アメリカ合衆国　マサチューセッツ　０１９４９，　ミドルトン，　キャンベル　ロード　４
(72)発明者　コスタ，　ジーナ
アメリカ合衆国　マサチューセッツ　０１９２９，　エセックス，　シェイ　コート　４

Ｆターム(参考)　4B024 AA11 CA01 CA04 CA05 CA09 CA10 CA11 CA20 HA11 HA12
4B029 AA07 CC02 CC03 CC08 FA15
4B063 QA01 QA13 QA18 QA19 QQ42 QQ52 QR32 QR35 QR41 QR55
QR59 QR62 QR82 QR83 QS25 QS32 QX01 QX02

Espacenet

Bibliographic data: JP2009538123 (A) — 2009-11-05

REAGENTS, METHODS, AND LIBRARIES FOR GEL-FREE BEAD-BASED SEQUENCING

Inventor(s):

Applicant(s):

Classification: - international: ***C12M1/00; C12N15/09; C12Q1/68***
- cooperative: **C12Q1/6837 (EP); C12Q1/6869 (EP); C12Q1/6874 (EP)** more

Application number: JP20090506766 20070419
Global Dossier

Priority number(s): US20060793702P 20060419 ; WO2007US66931 20070419

Also published as: AU2007237909 (A1) CA2649725 (A1) CN101495654 (A) EP2007907 (A2) US2009062129 (A1) more

Abstract not available for JP2009538123 (A)
Abstract of corresponding document: WO2007121489 (A2)

The present invention provides methods for determining a nucleic acid sequence by performing successive cycles of duplex extension along a single stranded template. The cycles comprise steps of extension, ligation, and, preferably, cleavage. In certain embodiments the methods make use of extension probes containing phosphorothiolate linkages and employ agents appropriate to cleave such linkages. In certain embodiments the methods make use of extension probes containing an abasic residue or a damaged base and employ agents appropriate to cleave linkages between a nucleoside and an abasic residue and/or agents appropriate to remove a damaged base from a nucleic acid. The invention provides methods of

determining information about a sequence using at least two distinguishably labeled probe families. In certain embodiments the methods acquire less than 2 bits of information from each of a plurality of nucleotides in the template in each cycle. In certain embodiments the sequencing reactions are performed on templates attached to microparticles, which are immobilized in or on a semi-solid support or attached to a substrate. The invention further provides sets of labeled extension probes containing phosphorothiolate linkages or trigger residues that are suitable for use in the method. In addition, the invention includes performing multiple sequencing reactions on a single template by removing initializing oligonucleotides and extended strands and performing subsequent reactions using different initializing oligonucleotides. The invention further provides efficient methods for preparing templates, particularly for performing sequencing multiple different templates in parallel. The invention also provides methods for performing ligation and cleavage. The invention also provides new libraries of nucleic acid fragments containing paired tags, and methods of preparing microparticles having multiple different templates (e.g., containing paired tags) attached thereto and of sequencing the templates individually. The invention also provides automated sequencing systems, flow cells, image processing methods, and computer-readable media that store computer-executable instructions (e.g., to perform the image-processing methods) and/or sequence information. In certain embodiments the sequence information is stored in a database. The invention further provides blocking oligonucleotides and methods of use thereof to facilitate sequencing. Further provided are arrays comprising microparticles having templates attached thereto and attached to a substrate in the absence of a semi-solid medium and methods of sequencing the templates. The invention also provides arrays of nucleic acid colonies produced using microparticles to "imprint" templates on a semi-solid medium or substrate, and methods of sequencing thereof.

Espacenet

Description: JP2009538123 (A) — 2009-11-05

REAGENTS, METHODS, AND LIBRARIES FOR GEL-FREE BEAD-BASED SEQUENCING

Description not available for JP2009538123 (A)
Description of corresponding document: WO2007121489 (A2)

A high quality text as facsimile in your desired language may be available amongst the following family members:

CA2649725 (A1) CN101495654 (A) US2009062129 (A1) WO2007121489 (A2)

The EPO does not accept any responsibility for the accuracy of data and information originating from other authorities than the EPO; in particular, the EPO does not guarantee that they are complete, up-to-date or fit for specific purposes.

REAGENTS, METHODS, AND LIBRARIES FOR GEL-FREE BEAD-BASED SEQUENCING

GOVERNMENT SUPPORT

This invention was made with government support under Grant Number ROI-HG- 003570, awarded by NIH. The government has certain rights in the invention.

Cross-Reference to Related Applications

The present application claims the benefit of and priority to copending United States provisional application number 60/793,702 filed April 19, 2006, the entire contents of which are herein incorporated by reference. This application is related to provisional applications USSN 60/649,294, filed Feb. 1, 2005; USSN 60/656,599, filed Feb. 25, 2005; USSN 60/673,749, filed April 21, 2005, USSN 60/699,541, filed July 15, 2005, USSN 60/722,526, filed September 30, 2005, and USSN 11/345,979, all of which are herein incorporated by reference.

Background of the Invention

Nucleic acid sequencing techniques are of major importance in a wide variety of fields ranging from basic research to clinical diagnosis. The results available from such technologies can include information of varying degrees of specificity . For example, useful information can consist of determining whether a particular polynucleotide differs in sequence from a reference polynucleotide, confirming the presence of a particular polynucleotide sequence in a sample, determining partial sequence information such as the identity of one or more nucleotides within a polynucleotide, determining the identity and order of nucleotides within a polynucleotide, etc.

DNA strands are typically polymers composed of four types of subunits, namely deoxyribonucleotides containing the bases adenine (A), cytosine (C), guanine (G), and thymidine (T). These subunits are attached to one another by covalent phosphodiester bonds that link the 5' carbon of one deoxyribose group to the 3' carbon of the following group. Most naturally occurring DNA consists of two such strands, which are aligned in an antiparallel orientation and are held together by hydrogen bonds formed between complementary bases, i.e., between A and T and between G and C. [0005] DNA sequencing first became possible on a large scale with the development of the chain termination or dideoxynucleotide method (Sanger, et al., Proc. Natl. Acad. Sci. 74:5463-5467, 1977) and the chemical degradation method (Maxam & Gilbert, Proc. Natl. Acad. Sci. 74:560-564, 1977), of which the former has been most extensively employed, improved upon, and automated. In particular, the use of fluorescently labeled chain terminators was of key importance in the development of automatic DNA sequencers. Common to both of the above approaches is the production of one or more collections of labeled DNA fragments of differing sizes, which must then be separated on the basis of length to determine the identity of the nucleotide at the 3' end of the fragment (in the chain termination method) or the identity of the nucleotide that was most recently removed from the fragment (in the case of the chemical degradation method). [0006] Although currently available sequencing technologies have allowed the achievement of major landmarks such as the sequencing of a number of complete genomes, these techniques have a number of disadvantages, and

considerable need for improvement remains in a number of areas. Separation of labeled DNA fragments has typically been achieved using polyacrylamide gel electrophoresis. However, this step has proven to be a major bottleneck limiting both the speed and accuracy of sequencing in many contexts. While capillary electrophoresis (CAE) proved to be the breakthrough that allowed the completion of the Human Genome Project (Venter, et al., Science, 291 :1304-1351, 2001; Lander, et al., Nature, 409:860-921, 2001), significant shortcomings remain. For example, CAE still requires a time-consuming separation step and still involves discrimination based on size, which can be inaccurate.

A variety of alternatives to the chain termination method have been proposed. In one approach, often referred to as "sequencing by synthesis", an oligonucleotide primer is first hybridized to a target template. The primer is then extended by successive cycles of polymerase-catalyzed addition of differently labeled nucleotides, whose incorporation into the growing strand is detected. The identity of the label serves to identify the complementary nucleotide in the template. Alternately, multiple reactions can be performed in parallel using each of the nucleotides, and incorporation of a labeled nucleotide in the reaction that uses a particular nucleotide identifies the complementary nucleotide in the template. (See, e.g., Melamede, U.S. Pat. No. 4,863,849; Cheeseman, U.S. Pat. No. 5,302,509, Tsien et al, International application WO 91/06678; Rosenthal et al, International application WO 93/21340; Canard et al, Gene, 148: 1-6 (1994); Metzker et al, Nucleic Acids Research, 22: 4259-4267 (1994)). [0008] To efficiently sequence polynucleotides of any significant length, it is desirable that the polymerase incorporates exactly one nucleotide in each cycle. Therefore it is generally necessary to use nucleotides that act as chain terminators, i.e., their incorporation prevents further extension by the polymerase. The incorporated nucleotide must then be modified, either enzymatically or chemically, to allow the polymerase to incorporate the next nucleotide. A variety of nucleotide analogs that can serve as chain terminators but can be modified after their incorporation such that they can be extended in a subsequent step have been proposed. Such "reversible terminators" have been described, for example, in U.S. Pat. Nos. U.S. 5,302,509; 6,255,475; 6,309, 836; 6,613,513. However, it has proven difficult to identify reversible terminators that can be incorporated by polymerase with

high efficiency, probably due to the fact that given the small size of a nucleotide, modifications that affect the ability of the nucleotide to act as a terminator also affect its incorporation into a growing polynucleotide strand.

Other sequencing approaches include pyrosequencing, which is based on the detection of the pyrophosphate (PPi) that is released during DNA polymerization (see, e.g., U.S. Pat. Nos. 6,210,891 and 6,258,568. While avoiding the need for electrophoretic separation, pyrosequencing suffers from a large number of drawbacks that have as yet limited its widespread applicability (Franca, et al, Quarterly Reviews of Biophysics, 35(2): 169-200, 2002). Sequencing by hybridization has also been proposed as an alternative (U.S. Pat. No. 5,202,231; WO 99/60170; WO 00/56937; Drmanac, et al., Advances in Biochemical Engineering/Biotechnology, 77:76-101, 2002) but has a number of disadvantages including the potential for error in discriminating between highly similar sequences. Single-molecule sequencing by exonuclease, which involves labeling every base in one strand and then detecting sequentially cleaved 3 ' terminal nucleotides in a sample stream is theoretically a very powerful method for rapidly determining the sequence of a long DNA molecule (Stephan, et al., J. Biotechnol, 86:255-267, 2001). However, various technical hurdles remain to be overcome before realization of this potential (Stephan, et al., 2001).
[0010] Diagnostic tests based upon particular sequence variations are already in use for a variety of different diseases. The sequencing of the human genome is widely thought to herald an era of personalized medicine in which therapies, including preventive therapies, will be tailored to the particular genetic make-up of the patient or will be selected based upon the identification of particular alleles or mutations. There is an increasing need for rapid and accurate determination of sequence variants of pathogenic agents such as HIV. Thus it is evident that the demand for accurate and rapid sequence determination will expand greatly in the immediate future. Improved methods for sequence determination of all types are therefore needed.

Summary of the Invention

The present invention provides new and improved sequencing methods that avoid the necessity for performing fragment separation and also in certain

embodiments avoid the need to use polymerase enzymes. An alternative to the methods discussed in the Background is described in U.S. Pat. Nos. 5,740,341 and 6,306,597, to Macevicz. The methods are based on repeated cycles of duplex extension along a single-stranded template. In preferred embodiments of these methods a nucleotide is identified in each cycle. The present invention provides improvements to these methods. The improvements allow efficient implementation of the methods and are particularly suited for high throughput sequencing. In addition, the invention provides methods for sequence determination that involve repeated cycles of duplex extension along a single-stranded template but do not involve identification of any individual nucleotide during each cycle.

In one aspect, the invention provides improved methods for sequencing based on successive cycles of duplex extension along a single-stranded template, ligation of labeled extension probes, and detection of the label. In general, extension starts from a duplex formed by an initializing oligonucleotide and a template. The initializing oligonucleotide is extended by ligating an oligonucleotide probe to its end to form an extended duplex, which is then repeatedly extended by successive cycles of ligation. During each cycle, the identity of one or more nucleotides in the template is determined by identifying a label on or associated with a successfully ligated oligonucleotide probe. The label of the newly added probe can also be detected prior to ligation, instead of, or in addition to, after ligation. Generally it is preferred to detect the label after ligation.

In preferred embodiments the probe has a non-extendable moiety in a terminal position (at the opposite end of the probe from the nucleotide that is ligated to the growing nucleic acid strand of the duplex) so that only a single extension of the extended duplex takes place in a single cycle. By "non-extendable" is meant that the moiety does not serve as a substrate for ligase without modification. For example, the moiety may be a nucleotide residue that lacks a 5' phosphate or 3' hydroxyl group. The moiety may be a nucleotide with a blocking group attached thereto that prevents ligation. In preferred embodiments of the invention the non-extendable moiety is removed after ligation to regenerate an extendable terminus so that the duplex can be further extended in subsequent cycles. [0014] To allow removal of the non-extendable moiety, in certain

embodiments of the invention the probe contains at least one internucleoside linkage that can be cleaved under conditions that will not substantially cleave phosphodiester bonds. Such linkages are referred to herein as "scissile internucleosidic linkages" or "scissile linkages". Cleavage of the scissile internucleosidic linkage removes the non-extendable moiety and either regenerates an extendable probe terminus or leaves a terminal residue that can be modified to form an extendable probe terminus. The scissile internucleosidic linkage may be located between any two nucleosides in the probe. Preferably the scissile linkage is located at least several nucleotides away from (i.e., distal to) the newly formed bond. The nucleotides in the extension probe between the terminal nucleotide that is ligated to the extendable terminus and the scissile linkage need not hybridize perfectly to the template. These nucleotides may serve as a "spacer" and allow identification of nucleotides located at intervals along the template without performing a cycle for each nucleotide within the interval.

The scissile internucleosidic linkage and the label are preferably located such that cleavage of the scissile internucleosidic linkage separates the extension probe into a labeled portion and a portion that remains part of the growing nucleic acid strand, allowing the labeled portion to diffuse away (e.g., upon raising the temperature). For example, the label may be attached to the terminal nucleotide of the extension probe, at the opposite end from the nucleotide that is ligated. Alternately, the label may be removed using any of a number of approaches.

The present inventors have discovered that phosphorothiolate linkages, in which one of the bridging oxygen atoms in the phosphodiester bond is replaced by a sulfur atom, are particularly advantageous scissile internucleosidic linkages. The sulfur atom in the phosphorothiolate linkage may be attached to either the 3' carbon of one nucleoside or the 5' carbon of the adjacent nucleoside.

In certain embodiments of the methods described above a plurality of sequencing reactions is performed. The reactions use initializing oligonucleotides that hybridize to different sequences of the template such that the terminus at which the first ligation occurs is located at different positions with respect to the template. For

example, the locations at which the first ligation occurs may be shifted, or "out of phase", relative to one another by 1 nucleotide increments. Thus after each cycle of extension with oligonucleotide probes of the same length, the same relative phase exists between the ends of the initializing oligonucleotides on the different templates. The reactions can be performed in parallel, in separate compartments each containing copies of the same template, or in series, i.e., by removing the extended duplex from the template after obtaining sequence information using a first initializing oligonucleotide and then performing additional reaction (s) using initializing oligonucleotides that hybridize to different sequences of the template. [0018] In another aspect, the invention provides solutions that are of use for a variety of nucleic acid manipulations. In one embodiment, the invention provides a solution containing or consisting essentially of 1.0-3.0% SDS, 100-300 mM NaCl, and 5-15 mM sodium bisulfate (NaHSO4) in water. The solution may contain or consist essentially of about 2% SDS, about 200mM NaCl, and about 10mM sodium bisulfate (NaHSO4) in water. For example, in one embodiment the solution contains 2% SDS, 200mM NaCl, and 10mM sodium bisulfate (NaHSO4) in water. In another embodiment the solution consists essentially of 2% SDS, 200mM NaCl, and 10mM sodium bisulfate (NaHSO4) in water. In certain embodiments the solution has a pH between 2.0 and 3.0, e.g., 2.5. The solutions are useful to separate double-stranded nucleic acids, e.g., double-stranded DNA, into individual strands, i.e., to denature (melt) double-stranded nucleic acids. In certain embodiments both strands are DNA. In other embodiments both strands are RNA. In other embodiments one strand is DNA and the other strand is RNA. In other embodiments one or both strands contains both RNA and DNA. In other embodiments one or both of the strands contains at least one nucleotide other than A, G, C, or T. In some embodiments one or both of the strands contains a non-naturally occurring nucleotide. In yet other embodiments one or more of the residues is a trigger residue, e.g., an abasic residue or damaged base. In some embodiments one or more residues contains a universal base. In some embodiments one or both of the strands contains a scissile linkage.

The double-stranded nucleic acids may be fully or partially double-stranded. They may be free in solution or one or both strands may be physically associated

with (e.g., covalently or noncovalently attached to) a solid or semi-solid support or substrate. Of particular note, double-stranded nucleic acids incubated in these solutions are effectively separated into single strands in the absence of heat or harsh denaturants that could cause gel delamination (e.g., when the nucleic acids are located in or attached to a semi-solid support such as a polyacrylamide gel) or could disrupt noncovalent associations such as streptavidin (SA)-biotin association (e.g., when the nucleic acids are attached to a support or substrate via a SA-biotin association). In one embodiment the solutions are used to separate double-stranded nucleic acids wherein one of the nucleic acids is attached to a bead via a SA-biotin association.

The invention also provides a method of separating strands of a double-stranded nucleic acid comprising the step of: contacting the double stranded nucleic acid with any of the afore-mentioned solutions, e.g., an aqueous solution containing about 1.0-3.0% SDS, about 100-300 mM NaCl, and about 5-15 mM sodium bisulfate (NaHSO4), e.g., containing 1.0-3.0% SDS, 100-300 mM NaCl, and 5-15 mM sodium bisulfate (NaHSO4). In one embodiment the solution contains about 2% SDS, 20OmM NaCl, and 1OmM sodium bisulfate (NaHSO4), e.g., 2% SDS, 20OmM NaCl, and 1OmM sodium bisulfate (NaHSO4). In another embodiment the solution consists essentially of 2% SDS, 20OmM NaCl, and 1OmM sodium bisulfate (NaHSO4) in water. In certain embodiments the solution has a pH between 2.0 and 3.0, e.g., 2.5. In some embodiments the double-stranded nucleic acid is incubated in the solution. In other embodiments the double-stranded nucleic acid (preferably attached to a support or substrate) is washed with the solution. In some embodiments the double-stranded nucleic acid is contacted with the solution for a time sufficient to separate at least 10% of the double-stranded nucleic acid molecules into single strands. In some embodiments the double-stranded nucleic acid is contacted with the solution for a time sufficient to separate at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99% or more of the double- stranded nucleic acids into single strands. In an exemplary embodiment the double-stranded nucleic acid is contacted with the solution for between 15 seconds and 3 hours. In another embodiment the double-stranded nucleic acid is contacted with the solution for between 1 minute and 1 hour. In certain embodiments the double-stranded nucleic acid is

contacted with the solution for about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, or 60 minutes. The methods may comprise a further step of removing the solution or removing some or all of the nucleic acids from the solution following a period of incubation. [0021] The solutions find use in one or more steps of a number of the sequencing methods described herein and may be employed in any of these methods. For example, the solutions may be used to separate an extended duplex from a template. The solutions may be used following cleavage of a scissile linkage to remove the portion of an extension probe that is no longer attached to the extended duplex. The solutions are also of use in separating strands of a triple-stranded nucleic acids or in separating double-stranded regions of a single nucleic acid strand that contains self-complementary portions that have hybridized to one another. [0022] In another aspect, the invention provides methods for obtaining information about a sequence using a collection of at least two distinguishably labeled oligonucleotide probe families. The probes in the probe families contain an unconstrained portion and a constrained portion. As in the methods described above, extension starts from a duplex formed by an initializing oligonucleotide and a template. The initializing oligonucleotide is extended by ligating an oligonucleotide probe to its end to form an extended duplex, which is then repeatedly extended by successive cycles of ligation. The probe has a non-extendable moiety in a terminal position (at the opposite end of the probe from the nucleotide that is ligated to the growing nucleic acid strand of the duplex) so that only a single extension of the extended duplex takes place in a single cycle. During each cycle, a label on or associated with a successfully ligated probe is detected, and the non-extendable moiety is removed or modified to generate an extendable terminus. The label corresponds to the probe family to which the probe belongs.

Successive cycles of extension, ligation, and detection produce an ordered list of probe families to which successive successfully ligated probes belong. The ordered list of probe families is used to obtain information about the sequence. However, knowing to which probe family a newly ligated probe belongs is not by itself sufficient to determine the identity of a nucleotide in the template. Instead, knowing to which probe family the newly ligated probe belongs eliminates

certain sequences as possibilities for the sequence of the constrained portion of the probe but leaves at least two possibilities for the identity of the nucleotide at each position. Thus there are at least two possibilities for the identity of the nucleotides in the template that are located at opposite positions to the nucleotides in the constrained portion of the newly ligated probe (i.e., the nucleotides that are complementary to the nucleotides in the constrained portion of the probe).

In certain embodiments, after performing a desired number of cycles, a set of candidate sequences is generated using the ordered series of probe family identities. The set of candidate sequences may provide sufficient information to achieve an objective. In preferred embodiments of the invention one or more additional steps are performed to select the correct sequence from among the candidate sequences. For example, the sequences can be compared with a database of known sequences, and the candidate sequence closest to one of the sequences in the database is selected as the correct sequence. In other embodiments the template is subjected to another round of sequencing by successive cycles of extension, ligation, detection, and cleavage, using a differently encoded set of probe families, and the information obtained in the second round is used to select the correct sequence. In other embodiments at least one item of information is combined with the information obtained from ordered list of probe family identities to determine the sequence. [0025] The invention also provides methods of performing error checking when templates are sequenced using probe families. Certain of the methods distinguish between single nucleotide polymorphisms (SNPs) and sequencing errors. [0026] The invention also provides nucleic acid fragments (e.g., DNA fragments) containing at least two segments of interest (e.g., at least two tags) and at least three primer binding regions (PBRs), such that at least two distinct templates, each corresponding to a segment of interest, can be amplified from each fragment. A "primer binding region" is a portion of a nucleic acid to which an oligonucleotide can hybridize such that the oligonucleotide can serve as an amplification primer, sequencing primer, initializing oligonucleotide, etc. Thus the primer binding region should have a known sequence in order to allow selection of a suitable complementary olignucleotide. As used herein and in the figures, a portion of a nucleic acid strand used in a

method of the invention may be referred to as a primer binding region regardless of whether, in the practice of the method, the primer actually binds to the region or binds to the corresponding portion of a complementary strand of the nucleic acid strand. Thus a portion of a nucleic acid may be referred to as a primer binding region regardless of whether, when used in a method of the invention, a primer actually binds to that region (in which case the sequence of the primer is complementary or substantially complementary to that of the region) or binds to the complement of the region (in which case the sequence of the primer is identical to or substantially identical to the sequence of the primer binding region) A segment of interest is any segment of nucleic acid for which sequence information is desired. For example, a sequence of interest may be a tag, and for purposes of the present disclosure it will be assumed that the segment of interest is a tag (also referred to herein and elsewhere as an "end tag"). However, it is to be understood that the invention is not limited to segments of interest that are tags. In certain embodiments the at least two tags are a paired tag. The nucleic acid fragments can contain one or more pairs of tags, e.g., one or more paired tags, e.g., 2, 3, 4, 5, or more pairs of paired tags. The invention further provides libraries containing such nucleic acid fragments, and methods for making the templates and libraries.

The invention further provides a microparticle, e.g., a bead, having at least two distinct populations of nucleic acids attached thereto, wherein each of the at least two populations consists of a plurality of substantially identical nucleic acids, and wherein the populations were produced by amplification (e.g., PCR amplification) from a single nucleic acid fragment. In some embodiments the single nucleic acid fragment contains a 5 ' tag and 3' tag, wherein the 5' and 3' tags are a paired tag. In some embodiments in which the single nucleic acid fragment contains a 5' tag and a 3' tag of a pair, one of the populations of nucleic acids attached to the microparticle comprises at least a portion of the 5' tag and one of the populations of nucleic acids attached to the microparticle comprises at least a portion of the 3' tag. In preferred embodiments one of the populations comprises a complete 5' tag and one of the populations comprises a complete 3' tag.

The nucleic acid fragment contains multiple PBRs, at least one of which is located between the tags and at

least two of which flank a portion of the nucleic acid fragment that contains the tags, so that a region comprising at least a portion of the 5' tag can be amplified, and a region comprising at least a portion of the 3' tag can be amplified, to produce two distinct populations of nucleic acids. In preferred embodiments the entire 5' tag and the entire 3' tag can be amplified. For example, the nucleic acid fragment can contain first and second primer binding sites flanking the 5 ' tag and also third and fourth primer binding sites flanking the 3' tag. A PCR amplification using primers that bind to the first and second primer binding sites amplifies the 5' tag. A PCR amplification using primers that bind to the third and fourth primer binding sites amplifies the 3' tag. It will be appreciated that the primers should be selected so that extension from each primer proceeds towards the region of the DNA fragment containing the tag to be amplified. Alternately, a first primer binding site can be located upstream of one of the tags, and a second primer binding site can be located downstream of the other tag, and a third primer binding site can be located between the two tags. The third primer binding site serves as a binding site for a forward primer for a PCR amplification that amplifies one of the tags and serves as a binding site for a reverse primer for a PCR amplification that amplifies the other tag. Thus in one embodiment the invention provides a microparticle, e.g., a bead, having at least two distinct populations of nucleic acids attached thereto, wherein each of the at least two populations consists of a plurality of substantially identical nucleic acids, and wherein a first distinct population comprises a 5' tag and a second distinct population comprises a 3' tag.

The invention further provides a population of microparticles, e.g., beads, wherein individual microparticles having at least two distinct populations of nucleic acids attached thereto, wherein each of the at least two populations consists of a plurality of substantially identical nucleic acids, and wherein the populations were produced by amplification (e.g., PCR amplification) from a single DNA fragment. The substantially identical populations can be, e.g,. a 5' tag and a 3' tag. The invention further provides arrays of such microparticlesand methods of sequencing that involve sequencing the populations of substantially identical nucleic acids. For example, in one embodiment, each of the two populations of substantially identical nucleic acids attached to an

individual microparticle comprises a different primer binding region (PBR), so that by using different sequencing primers, one of the populations can be sequenced without interference from the other population. If more than two substantially identical populations of substantially identical nucleic acids are attached to a single microparticle, each of the populations can have a unique (i.e., distinct) PBR, such that a primer that binds to a given PBR does not bind to a PBR present in the other substantially identical populations of nucleic acids attached to the microparticle. Thus the methods of the invention allow for producing microparticles having at least two different substantially identical populations of nucleic acids attached thereto (e.g., a multiple copies of template containing a 5' tag and multiple copies of template containing a 3' tag), wherein the tags are paired tags. In accordance with the inventive methods, the templates contain different PBRs, which provide binding sites for sequencing primers. Therefore, by selecting a sequencing primer complementary to the PBR in the template that contains the 5 ' tag, sequence information can be obtained from the 5 ' tag without interference from the template containing the 3 ' tag, even though the template containing the 3 ' tag is also present on the same microparticle. By selecting a sequencing primer complementary to the PBR in the template that contains the 3 ' tag, sequence information can be obtained from the 3 ' tag without interferene from the template containing the 5 ' tag, even though the template containing the 5' tag is also present on the same microparticle. The fact that both of the paired tags are present on the same microparticle means that the sequence of the 5' and 3' paired tags can be associated with one another, just as would be the case if they were present within a single template.

Also provided are arrays of microparticles attached to a substrate. In one embodiment microparticles are tethered to a substrate via a single-stranded template, that is attached to the microparticle at one terminus and attached to the substrate at the other terminus. The means of attachment at either or both ends may be covalent or noncovalent. In certain embodiments either or both means of attachment comprises a biotin-binding moiety and biotin.

Also provided are arrays comprising nucleic acid colonies generated by copying templates attached to

microparticles and, optionally, amplifying the copied templates. Also provided are blocking oligonucleotides and methods of use thereof as well as compositions comprising blocking oligonucleotides. [0032] The invention also provides automated sequencing systems that may be used, e.g., to sequence templates arrayed in or on a substantially planar support. The invention further provides image processing methods, which may be stored on a computer-readable medium such as a hard disc, CD, zip disk, flash memory, or the like. In certain preferred embodiments the system achieves 40,000 nucleotide identifications per second, or more. In certain preferred embodiments the system generates 8.6 gigabytes (Gb) of sequence data per day (24 hours), or more. In certain embodiments the system produces 48 Gb of sequence information (nucleotide identifications) per day, or more.

The invention also provides a computer-readable medium that stores information generated by applying the inventive sequencing methods. The information may be stored in a database.

This application refers to various patents, patent applications, journal articles, and other publications, all of which are incorporated herein by reference. In addition, the following standard reference works are incorporated herein by reference: Current Protocols in Molecular Biology, John Wiley & Sons, N.Y., edition as of July 2002; Sambrook, Russell, and Sambrook, Molecular Cloning: A Laboratory Manual, 3<rd> ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2001. In the event of a conflict between the instant specification and any document incorporated by reference, the specification shall control, it being understood that the determination of whether a conflict or inconsistency exists is within the discretion of the inventors and can be made at any time.

Brief Description of the Drawing

Please note that many of Figures 1-35 are in color in USSN 11/345,979 and that these figures are incorporated herein by reference and can be used in place of those provided herewith. [0035] Fig. IA diagramatically illustrates initialization followed by two cycles of extension, ligation, and identification.

Fig. IB diagramatically illustrates initialization followed

by two cycles of extension, ligation, and identification in an embodiment in which extension proceeds inwards from the free end of the template towards a support.

Fig. 2 shows a scheme for assigning colors to oligonucleotide probes in which the identity of the 3' base of the probe is determined by identifying the color of a fluorophore. [0038] Fig. 3A diagramatically shows extended duplexes resulting from hybridization of initializing oligonucleotides at different positions in the binding region of a template followed by ligation of extension probes. [0039] Fig. 3B diagramatically shows assembly of a continuous sequence by using the extension, ligation, and cleavage method with extension probes designed to read every 6th base of the template molecule.

Fig. 4A illustrates a 5 '-S- phosphorothiolate linkage (3 '-O-P-S-5 ').

Fig. 4B illustrates a 3'-S-phosphorothiolate linkage (3'-S-P-O-5').

Fig. 5 A diagramatically illustrates a single cycle of extension, ligation, and cleavage for sequencing in the 5 '->3' direction using extension probes having 3'-O-P-S-5' phosphorothiolate linkages.

Fig. 5B diagramatically illustrates a single cycle of extension, ligation, and cleavage for sequencing in the 3' -^ 5' direction using extension probes having 3'-S-P-O-5' phosphorothiolate linkages.

Fig. 6A-6F is a more detailed diagrammatic illustration of several sequencing reactions performed on a single template. The reactions utilize initializing oligonucleotides that bind to different portions of the template.

Fig. 7 is a schematic showing a synthesis scheme for 3'-phosphoroamidites of dA and dG.

Figs. 8A-8E shows results of a gel shift assay demonstrating two cycles of successful ligation and cleavage of extension probes containing phosphorothiolate linkages.

Fig. 8F shows a schematic diagram of the mechanism of ligation by DNA ligases.

Fig. 9 results of a gel shift assay demonstrating the ligation efficiency of degenerate inosine-containing oligonucleotide probes.

Fig. 10 shows results of a gel shift assay demonstrating the ligation efficiency of degenerate inosine-containing oligonucleotide probes on multiple templates.

Fig. 11 shows results of an analysis conducted to assess the fidelity of each of two DNA ligases (T4 DNA ligase and Tag DNA ligase) for 3 '-^5' extensions.

Fig. 12 shows results of a gel shift assay (A) demonstrating the ligation efficiency of degenerate inosine-containing oligonucleotide probes and of a direct sequencing analysis of the ligation reactions (B) conducted to assess the fidelity of T4 DNA ligase in oligonucleotide probe ligation. Results are tabulated in panels C-F.

Fig. 13A-13C shows results of an experiment that demonstrates in-gel ligation when bead-based templates are embedded in polyacrylamide gels on slides. Fig. 13A shows a schematic of the ligation reaction. In gel ligation reactions were performed in the absence (B) and in the presence (C) of T4 DNA ligase.

Fig. 14A shows an image of an emulsion PCR reaction performed on beads having attached first amplifcation primers, using a fluorescently labeled second amplification primer and an excess of template.

Fig. 14B (top) shows a fluorescence image of a portion of a slide on which beads with an attached template, to which a Cy3 -labeled oligonucleotide was hybridized, were immobilized within a polyacrylamide gel. (This slide was used in a different experiment, but is representative of the slides used here.) Fig. 14B (bottom) shows a schematic diagram of a slide equipped with a Teflon mask to enclose the polyacrylamide solution.

Fig. 15 illustrates three sets of labeled oligonucleotide probes designed to address issues of probe specificity and selectivity and also shows excitation and emission values for a set of four spectrally resolvable labels.

Fig. 16 shows results of an experiment confirming

4-color spectral identity of oligonucleotide probes. Slides containing four unique single-stranded template populations (A) were subjected to hybridization and ligations reactions using an oligonucleotide probe mixture that contained four unique fluorophore probes, were imaged under bright light (B) and with fluorescence excitation using four bandpass filters before and after ligation. Individual populations were pseudocolored (C). The spectral identity, which showed minimal signal overlap, is plotted in (D).

Fig. 17 shows an experiment confirming ligation specificity of oligonucleotide extension probes. Fig. 17 (A) shows a schematic outline of the ligation. Fig. 17(B) is a bright light image, and Fig. 17(C) is a corresponding fluorescence image of a population of beads embedded in a polyacrylamide gel after ligation. Fig. 17(D) shows fluorescence detected from each label before (pre) or after (post) ligation.

Fig. 18 shows another experiment confirming ligation specificity and selectivity of oligonucleotide extension probes. Fig. 18(A) shows a schematic outline of the ligation. Fig. 17(B) is a bright light image, and Fig. 18(C) is a corresponding fluorescence image of a population of beads embedded in a polyacrylamide gel after ligation. Fig. 18(D) shows expected versus observed ligation frequencies, showing a high correlation between frequencies expected based on the proportion of particular extension probes in a population and frequencies observed. [0059] Fig. 19 shows an experiment confirming that degenerate and universal base containing oligonucleotide extension probe pools can be used to afford specific and selective in-gel ligation. Fig. 19(A) shows a schematic outline of the ligation experiment, illustrating four differentially labeled degenerate inosine-containing probe pools following ligation. Fig. 19(B) is a bright light image, and Fig. 19(C) is a corresponding fluorescence image of a population of beads embedded in a polyacrylamide gel after ligation. Fig. 19(D) shows expected versus observed ligation frequencies, showing a high correlation between frequencies expected based on the proportion of particular extension probes in a population and frequencies observed. Fig. 19(E) shows a scatter plot of the raw unprocessed data and filtered data reprsenting the top 90% of bead signal values.

Fig. 20 is a chart showing the signal detected in

sequential cycles of hybridization and stripping of an initializing oligonucleotide (primer) to a template. As shown in the figure, minimal signal loss occurred over 10 cycles.

Fig. 21 is a photograph of an automated sequencing system that may be used to gather sequence information, e.g., from templates arrayed in or on a substantially planar support. Also shown is a dedicated computer for controlling operation of various components of the system, processing and storing collected image data, providing a user interface, etc. The lower portion of the figure shows an enlarged view of a flow cell oriented to achieve gravimetric bubble displacement.

Fig. 22 shows a schematic diagram of a high throughput automated sequencing instrument that may be used to sequence templates arrayed in or on a substantially planar support.

Fig. 23 shows a scatter plot of alignment inconsistency, illustrating minimal inconsistency over 30 frames.

Figs. 24A-I shows schematic diagrams of inventive flow cells or portions thereof in a variety of different views.

Fig. 25 A shows an exemplary encoding for a preferred collection of probe families comprising partially constrained probes comprising constrained portions that are 2 nucleotides in length.

Fig. 25B shows a preferred collection of probe families (upper panel) and a cycle of ligation, detection, and cleavage (lower panel). [0067] Fig. 26 shows an exemplary encoding for another preferred collection of probe families comprising partially constrained probes comprising constrained portions that are 2 nucleotides in length.

Figs. 27A-27C represent an alternate method to schematically define the 24 preferred collections of probe families that are defined in Table 1.

Fig. 28 shows a less preferred collection of probe families in which the probes comprise constrained portions that are 2 nucleotides in length.

Fig. 29A shows a diagram that can be used to generate

constrained portions for a collection of probe families that comprises probes with a constrained portion 3 nucleotides long.

Fig. 29B shows a diagram a mapping scheme that can be used to generate constrained portions for a collection of probe families that comprises probes with a constrained portion 3 nucleotides long from the 24 preferred collections of probe families.

Fig. 30 shows a method in which sequence determination is performed using a collection of probe families. An embodiment using a preferred set of probe families is depicted.

Figs. 3 IA - 31C show a method in which sequence determination is performed using a first collection of probe families to generate candidate sequences and a second collection of probe families to decode.

Fig. 32 shows a method in which sequence determination is performed using a less preferred collection of probe families.

Fig. 33 A shows a schematic diagram of a slide with beads attached thereto. DNA templates are attached to the beads.

Figure 33B shows a population of beads attached to a slide. The lower panels show the same region of the slide under white light (left) and fluorescence microscopy. The upper panel shows a range of bead densities.

Figures 34A - 34C show a scheme for amplifying both tags of a paired tag present in a nucleic acid fragment (template) as individual populations of nucleic acids and capturing them to a microparticle via the amplification process.

Figures 35A and 35B show details of primer design and amplification for the scheme of Figure 35. Both strands of a nucleic acid fragment (template) are shown for clarity. Primers and primer binding regions having the same sequence are presented in the same color. For example, PI is represented in dark blue, indicating that primer PI, which is present on the microparticle and in solution, has the same sequence as the correspondingly colored portion of the indicated strand of the template.

The dark blue region of the template, labeled PI, may be referred to as a primer binding region even though the corresponding primer (PI) in fact binds to the complementary portion of the other strand and has the same sequence as primer PI .

Figures 35C and 35D show sequencing of the first and second tags, respectively, attached to a microparticle produced by the method of Figures 35 A and 35B.

Figure 36A depicts a template molecule from a paired-end library showing blocking oligonucleotides hybridized to the forward adapter, reverse adapter, and internal adapter portions of the template, which are common to members of the library. The lower portion of the figure shows exemplary sequences for the adapters and blocking oligonucleotides. "ddBase" in Figures 36A-36C indicates a dideoxy nucleoside. "Unique

DNA sequence" represents a target region to be sequenced.

Figure 36B depicts a template molecule from a fragment library showing blocker oligonucleotides hybridized to the forward adapter, reverse adapter, and internal adapter portions of the template molecule, which are common to members of the library. The lower portion of the figure shows exemplary sequences for the adapters and the complementary blocking oligonucleotides.

Figure 36C depicts a molecule from a library in which the template molecules have undergone rolling circle amplification (RCA). RCA creates multiple copies of the unique portion of the template molecule (2) as well as the adapter regions (1) and padlock region (3). The figure shows blocking oligonucleotides hybridized to the adapter and padlock portions of the template, which are common to members of the library.

Figure 37 shows several padlock probe sequences and exemplary sequences for oligonucleotides that would block the padlock region following synthesis of a template molecule using RCA.

Figure 38 shows an array of microparticles generated on a substrate without use of a semi-solid medium (gel-free microparticle array).

Figure 39 shows results of ligation-based sequencing

using a gel-free microparticle array.

] Figure 40 shows a schematic diagram of a microparticle located on a surface and illustrates the expected size of the contact patch and nucleic acid colony that would result from template extension.

Definitions

To facilitate understanding of the description of the invention, the following definitions are provided. It is to be understood that, in general, terms not otherwise defined are to be given their meaning or meanings as generally accepted in the art. [0088] As used herein, an "abasic residue" is a residue that has the structure of the portion of a nucleoside or nucleotide that remains after removal of the nitrogenous base or removal of a sufficient portion of the nitrogenous base such that the resulting molecule no longer participates in hydrogen bonds characteristic of a nucleoside or nucleotide. An abasic residue may be generated by removing a nitrogenous base from a nucleoside or nucleotide. However, the term "abasic" is used to refer to the structural features of the residue and is independent of the manner in which the residue is produced. The terms "abasic residue" and "abasic site" are used herein to refer to a residue within a nucleic acid that lacks a purine or pyrimidine base.

An "apurinic/apyrimidinic (AP) endonuclease", as used herein, refers to an enzyme that cleaves a bond on either the 5 ' side, the 3 ' side, or both the 5 ' and 3 ' sides of an abasic residue in a polynucleotide. In certain embodiments of the invention the AP endonuclease is an AP lyase. Exampes of AP endonucleases include, but are not limited to, E. coli endonuclease VIII and homo logs thereof and E. coli endonuclease III and homo logs thereof. It is to be understood that references to specific enzymes, e.g., endonucleases such as E. coli Endo VIII, Endo V, etc., are intended to encompass homo logs from other species that are recognized in the art as being homologs and as possessing similar biochemical activity with respect to removal of damaged bases and/or cleavage of DNA containing abasic residues or other trigger residues.

As used herein, the term "array" refers to a collection of entities that is distributed over or in a support matrix; preferably, individual entities are spaced at a distance

from one another sufficient to permit the identification of discrete features of the array by any of a variety of techniques. The entities may be, for example, nucleic acid molecules, clonal populations of nucleic acid molecules, microparticles (optionally having clonal populations of nucleic acid molecules attached thereto), etc. When used as a verb, the term "array" and variations thereof refers to any process for forming an array, e.g., distributing entities over or in a support matrix.

A "damaged base" is a purine or pyrimidine base that differs from an A, G, C, or T in such a manner as to render it a substrate for removal from DNA by a DNA glycosylase. Uracil is considered a damaged base for purposes of the present invention. In some embodiments of the invention the damaged base is hypoxanthine. [0092] "Degenerate", with respect to a position in a polynucleotide that is one of a population of polynucleotides, means that the identity of the base that forms part of the nucleoside occupying that position varies among different members of the population. Thus the population contains individual members whose sequence differs at the degenerate position. The term "position" refers to a numerical value that is assigned to each nucleoside in a polynucleotide, generally with respect to the 5' or 3' end. For example, the nucleoside at the 3' end of an extension probe may be assigned position 1. Thus in a pool of extension probes of structure 3'-XXXNXXXX-5', the N is at position 4. Position 4 is considered degenerate if, in different members of the pool, the identity of N can vary. The pool of extension probes is also said to be degenerate at position N. A position is said to be k-fold degenerate if it can be occupied by nucleosides having any of k different identities. For example, a position that can be occupied by nucleosides comprising either of 2 different bases is 2-fold degenerate.

"Determining information about a sequence" encompasses "sequence determination" and also encompasses other levels of information such as eliminating one or more possibilities for the sequence. It is noted that performing sequence determination on a polynucleotide typically yields equivalent information regarding the sequence of a perfectly complementary (100% complementary) polynucleotide and thus is equivalent to sequence determination performed directly on a perfectly complementary polynucleotide. [0094]

"Independent", with respect to a plurality of elements, e.g., nucleosides in an oligonucleotide probe molecule or portion thereof, means that the identity of each element does not limit and is not limited by the identity of any of the other elements, e.g., the identity of each element is selected without regard for the identity of any of the other element(s). Thus knowing the identity of one or more of the elements does not provide any information regarding the identity of any of the other elements. For example, the nucleosides in the sequence NNNN are independent if the identity of each N can be A, G, C, or T, regardless of the identity of any other N. [0095] "Ligation" means to form a covalent bond or linkage between the termini of two or more nucleic acids, e.g. oligonucleotides and/or polynucleotides, in a template-driven reaction. The nature of the bond or linkage may vary widely and the ligation may be carried out enzymatically or chemically.

The term "microparticle" is used herein to refer to particles having a smallest cross-sectional dimension of 50 microns or less, preferably 10 microns or less. In certain embodiments the smallest cross-sectional dimension is approximately 3 microns or less, approximately 1 micron or less, approximately 0.5 microns or less, e.g., approximately 0.1, 0.2, 0.3, or 0.4 microns. Microparticles may be made of a variety of inorganic or organic materials including, but not limited to, glass (e.g., controlled pore glass), silica, zirconia, cross-linked polystyrene, polyacrylate, polymehtymethacrylate, titanium dioxide, latex, polystyrene, etc. See, e.g., U.S. Pat. No. 6,406,848, for various suitable materials and other considerations. Dyna beads, available from Dynal, Oslo, Norway, are an example of commercially available microparticles of use in the present invention. Magnetically responsive microparticles can be used. The magnetic responsiveness of certain preferred microparticles permits facile collection and concentration of the microparticle-attached templates after amplification, and facilitates additional steps (e.g., washes, reagent removal, etc.). In certain embodiments of the invention a population of microparticles having different shapes (e.g., some spherical and others nonspherical) is employed.

The term "microsphere" or "bead" is used herein to refer to substantially spherical microparticles having a diameter of 50 microns or less, preferably 10 microns or

less. In certain embodiments the diameter is approximately 3 microns or less, approximately 1 micron or less, approximately 0.5 microns or less, e.g., approximately 0.1, 0.2, 0.3, or 0.4 microns. In certain embodiments of the invention a population of monodisperse microspheres is used, i.e., the microspheres are of substantially uniform size. For example, the diameters of the microparticles may have a coefficient of variation of less than 5%, e.g., 2% of less, 1% or less, etc. However, in other embodiments the coefficient of variation of a population of microparticles is 5% or greater, e.g., 5%, between 5% and 10% (inclusive), between 10% and 25%, inclusive, etc. In certain embodiments a mixed population of microparticles is used. For example, a mixture of two populations, each of which has a coefficient of variation of less than 5%, may be used, resulting in a mixed population that is not monodisperse. As an example, a mixture of microspheres having diameters of 1 micron and 3 microns can be employed. In certain embodiments of the invention additional information is provided by the size of the microsphere when sequencing is performed using templates attached to microspheres of a population that is not monodisperse. For example, different libraries of templates may be attached to differently sized microspheres. Also, since fewer template molecules may be attached to smaller particles, the intensity of the signals may vary, which may facilitate multiplex sequencing.

The term "nucleic acid sequence" as used herein can refer to the nucleic acid material itself and is not restricted to the sequence information (i.e. the succession of letters chosen among the five base letters A, G, C, T, or U) that biochemically characterizes a specific nucleic acid, e.g., a DNA or RNA molecule. Nucleic acids shown herein are presented in a 5' -^ 3' orientation unless otherwise indicated.

A "nucleoside" comprises a nitrogenous base linked to a sugar molecule. As used herein, the term includes natural nucleosides in their 2'-deoxy and 2'-hydroxyl forms as described in Kornberg and Baker, DNA Replication, 2nd Ed. (Freeman, San Francisco, 1992) and nucleoside analogs. For example, natural nucleosides include adenosine, thymidine, guanosine, cytidine, uridine, deoxyadenosine, deoxythymidine, deoxyguanosine, and deoxycytidine. Nucleoside "analogs" refers to synthetic nucleosides having

modified base moieties and/or modified sugar moieties, e.g. described generally by Scheit, Nucleotide Analogs (John Wiley, New York, 1980). Such analogs include synthetic nucleosides designed to enhance binding properties, reduce degeneracy, increase specificity, and the like. Nucleoside analogs include 2-aminoadenosine, 2-thiothymidine, pyrrolo-pyrimidine, 3- methyl adenosine, C5-propynylcytidine, C5-propynyluridine, C5-bromouridine, C5- fluorouridine, C5-iodouridine, C5-methylcytidine, 7-deazaadenosine, 7-deazaguanosine, 8- oxoadenosine, 8-oxoguanosine, O(6)-methylguanine, 2-thiocytidine, etc. Nucleoside analogs may comprise any of the universal bases mentioned herein.

[00100] The term "organism" is used herein to indicate any living or nonliving entity that comprises nucleic acid that is capable of being replicated and is of interest for sequence determination. It includes plasmids; viruses; prokaryotic, archaebacterial and eukaryotic cells, cell lines, fungi, protozoa, plants, animals, etc.

[00101] "Perfectly matched duplex" in reference to the protruding strands of probes and template polynucleotides means that the protruding strand from one forms a double stranded structure with the other such that each nucleoside in the double stranded structure undergoes Watson-Crick basepairing with a nucleoside on the opposite strand. The term also comprehends the pairing of nucleoside analogs, such as deoxyinosine, nucleosides with 2- aminopurine bases, and the like, that may be employed to reduce the degeneracy of the probes, whether or not such pairing involves formation of hydrogen bonds. [00102] The term "plurality" means more than one.

[00103] The term "polymorphism" is given its ordinary meaning in the art and refers to a difference in genome sequence among individuals of the same species. A "single nucleotide polymorphism" (SNP) refers to a polymorphism at a single position. [00104] "Polynucleotide", "nucleic acid", or "oligonucleotide" refers to a linear polymer of nucleosides (including deoxyribonucleosides, ribonucleosides, or analogs thereof) joined by internucleosidic linkages. Typically, a polynucleotide comprises at least three nucleosides. In certain embodiments of the invention one or more nucleosides in an extension probe comprises a universal base. Usually oligonucleotides range in size from a few monomeric units, e.g. 3-4, to several

hundreds of monomeric units. Whenever a polynucleotide such as an oligonucleotide is represented by a sequence of letters, such as "ATGCCTG," it will be understood that the nucleotides are in 5'->3' order from left to right and that "A" denotes deoxyadenosine, "C" denotes deoxycytidine, "G" denotes deoxyguanosine, and "T" denotes thymidine, unless otherwise noted. The letters A, C, G, and T may be used to refer to the bases themselves, to nucleosides, or to nucleotides comprising the bases, as is standard in the art.

[00105] In naturally occuring polynucleotides, the internucleoside linkage is typically a phosphodiester bond, and the subunits are referred to as "nucleotides". However, oligonucleotide probes comprising other internucleoside linkages, such as phosphorothiolate linkages, are used in certain embodiments of the invention. It will be appreciated that one or more of the subunits that make up such an oligonucleotide probe with a non-phosphodiester linkage may not comprise a phosphate group. Such analogs of nucleotides are considered to fall within the scope of the term "nucleotide" as used herein, and nucleic acids comprising one or more internucleoside linkages that are not phosphodiester linkages are still referred to as "polynucleotides", "oligonucleotides", etc. In other embodiments, a polynucleotide such as an oligonucleotide probe comprises a linkage that contains an AP endonuclease sensitive site. For example, the oligonucleotide probe may contain an abasic residue, a residue containing a damaged base that is a substrate for removal by a DNA glycosylase, or another residue or linkage that is a substrate for cleavage by an AP endonuclease. In another embodiment an oligonucleotide probe contains a disaccharide nucleoside. [00106] The term "primer" refers to a short polynucleotide, typically between about 10- 100 nucleotides in length, that binds to a target polynucleotide or "template" by hybridizing with the target. The primer preferably provides a point of initiation for template-directed synthesis of a polynucleotide complementary to the target, which can take place in the presence of appropriate enzyme(s), cofactors, substrates such as nucleotides, oligonucleotides, etc. The primer typically provides a terminus from which extension can occur. In the case of primers for synthesis catalyzed by a polymerase enzyme such as a DNA polymerase (e.g., in

"sequencing by synthesis", polymerase chain reaction (PCR) amplifcation, etc.), the primer typically has, or can be modified to have, a free 3' OH group. Typically a PCR reaction employs a pair of primers (first and second amplification primers) including an "upstream" (or "forward") primer and a "downstream" (or "reverse") primer, which delimit a region to be amplified. In the case of primers for synthesis by successive cycles of extension, ligation (and optionally cleavage), the primer typically has, or can be modified to have, a free 5 ' phosphate group or 3 ' OH group that serves as a substrate for DNA ligase.

[00107] As used herein, a "probe family" refers to a group of probes, each of which comprises the same label.

[00108] As used herein "sequence determination", "determining a nucleotide sequence", "sequencing", and like terms, in reference to polynucleotides includes determination of partial as well as full sequence information of the polynucleotide. That is, the term includes sequence comparisons, fingerprinting, and like levels of information about a target polynucleotide, as well as the express identification and ordering of each nucleoside of the target polynucleotide within a region of interest. In certain embodiments of the invention "sequence determination" comprises identifying a single nucleotide, while in other embodiments more than one nucleotide is identified. In certain embodiments of the invention, sequence information that is insufficient by itself to identify any nucleotide in a single cycle is gathered. Identification of nucleosides, nucleotides, and/or bases are considered equivalent herein. It is noted that performing sequence determination on a polynucleotide typically yields equivalent information regarding the sequence of a perfectly complementary (100% complementary) polynucleotide and thus is equivalent to sequence determination performed directly on a perfectly complementary polynucleotide. [00109] "Sequencing reaction" as used herein refers to a set of cycles of extension, ligation, and detection. When an extended duplex is removed from a template and a second set of cycles is performed on the template, each set of cycles is considered a separate sequencing reaction though the resulting sequence information may be combined to generate a single sequence.

[00110] "Semi-solid", as used herein, refers to a

compressible matrix with both a solid and a liquid component, wherein the liquid occupies pores, spaces or other interstices between the solid matrix elements. Exemplary semi-solid matrices include matrices made of polyacrylamide, cellulose, polyamide (nylon), and cross-linked agarose, dextran and polyethylene glycol. A semi-solid support may be provided on a second support, e.g., a substantially planar, rigid support, also referred to as a substrate, which supports the semisolid support.

[00111] "Support", as used herein, refers to a matrix on or in which nucleic acid molecules, microparticles, and the like may be immobilized, i.e., to which they may be covalently or noncovalently attached or, in or on which they may be partially or completely embedded so that they are largely or entirely prevented from diffusing freely or moving with respect to one another.

[00112] A "trigger residue" is a residue that, when present in a nucleic acid, renders the nucleic acid more susceptible to cleavage (e.g., cleavage of the nucleic acid backbone) by a cleavage agent (e.g., an enzyme, silver nitrate, etc.) or combination of agents than would be an otherwise identical nucleic acid not including the trigger residue, and/or is susceptible to modification to generate a residue that renders the nucleic acid more susceptible to such cleavage. Thus presence of a trigger residue in a nucleic acid can result in presence of a scissile linkage in the nucleic acid. For example, an abasic residue is a trigger residue since the presence of an abasic residue in a nucleic acid renders the nucleic acid susceptible to cleavage by an enzyme such as an AP endonuclease. A nucleoside containing a damaged base is a trigger residue since the presence of a nucleoside comprising a damaged base in a nucleic acid also renders the nucleic acid more susceptible to cleavage by an enzyme such as an AP endonuclease, e.g., after removal of the damaged base by a DNA glycosylase. The cleavage site may be at a bond between the trigger residue and an adjacent residue or may be at a bond that is one or more residues removed from the trigger residue. For example, deoxyinosine is a trigger residue since the presence of a deoxyinosine in a nucleic acid renders the nucleic acid more susceptible to cleavage by E. coli Endonuclease V and homologs thereof. Such enzymes cleave the second phosphodiester bond 3' to deoxyinosine. Any of the probes disclosed herein may contain one or more

trigger residues. The trigger residue may, but need not, comprise a ribose or deoxyribose moiety. Preferably the cleavage agent is one that does not substantially cleave a nucleic acid in the absence of a trigger residue but exhibits significant cleavage activity against a nucleic acid that contains the trigger residue under the same conditions, which conditions may include the presence of agents that modify the nucleic acid to render it sensitive to the cleavage agent. For example, preferably if the cleavage agent is present in a composition containing nucleic acids that are identical in length and composition except that one of them contains the trigger residue and the other of them does not contain the trigger residue, the likelihood that the nucleic acid containing the trigger residue will be cleaved is at least: 10; 25; 50; 100; 250; 500; 1000; 2500; 5000; 10,000; 25,000; 50,000; 100,000; 250,000; 500,000; 1,000,000 or more, as great as the likelihood that the nucleic acid not containing the trigger residue will be cleaved, e.g., the ratio of the likelihood of cleavage of a nucleic acid containing a trigger residue to the likelihood of cleavage of a nucleic acid not containing the trigger residue but otherwise identical is between 10 and 10<6>, or any integral subrange thereof. It will be appreciated that the ratio may differ depending upon the particular nucleic acid and location and nucleotide context of the trigger residue.

[00113] Preferably if the nucleic acid containing the trigger residue needs to be modified in order to render the nucleic acid susceptible to cleavage by a cleavage agent, such modification occurs readily in the presence of suitable modifying agent(s), e.g., the modification occurs in reasonable yield and in a reasonable period of time. For example, in certain embodiments of the invention at least 50%, at least 60%, at least 70%, preferably at least 80%, at least 90% or more preferably at least 95% of the nucleic acids containing the trigger residue are modified within, e.g., 24 hours, preferably within 12 hours, more preferably within less than 1 minute to 4 hours.

[00114] A variety of suitable trigger residues and corresponding cleavage reagents are exemplified herein. Any trigger residue and cleavage reagent having similar activity to those described herein may be used. One of ordinary skill in the art will be able to determine whether a particular trigger residue and cleavage reagent combination is suitable for use in the present

invention, e.g., whether the cleavage efficiency and speed, the selectivity of the cleavage agent for nucleic acids containing a trigger residue, etc, are suitable for use in the methods of the invention. Note that a "trigger residue" is distinguished from a nucleotide that simply forms part of a restriction enzyme site in that the ability of the trigger residue to confer increased susceptibility to cleavage does not, in general, depend significantly on the particular sequence context in which the trigger residue is found although, as noted above, the context can have some influence on the susceptibility to modification and/or cleavage. Of course depending on the surrounding nucleotides, a trigger residue may form part of a restriction site. Thus, in most cases, the cleavage agent is not a restriction enzyme, though use of an enzyme that is both a restriction enzyme and has non-sequence specific cleavage ability is not excluded.
[00115] A "universal base", as used herein, is a base that can "pair" with more than one of the bases typically found in naturally occurring nucleic acids and can thus substitute for such naturally occurring bases in a duplex. The base need not be capable of pairing with each of the naturally occurring bases. For example, certain bases pair only or selectively with purines, or only or selectively with pyrimidines. Certain preferred universal bases (fully universal bases) can pair with any of the bases typically found in naturally occurring nucleic acids and can thus substitute for any of these bases in duplex. The base need not be equally capable of pairing with each of the naturally occurring bases. If a probe mix contains probes that comprise (at one or more positions) a univeral base that does not pair with all of the naturally occurring nucleotides, it may be desirable to utilize two or more universal bases at that position in the particular probe so that at least one of the universal bases pairs with A, at least one of the universal bases pairs with G, at least one of the universal bases pairs with C, and at least one of the universal bases pairs with T.

[00116] A number of universal bases are known in the art including, but not limited to, hypoxanthine, 3-nitropyrrole, 4- nitroindole, 5-nitroindole, 4-nitrobenzimidazole, 5- nitroindazole, 8-aza-7-deazaadenine, 6H,8H-3,4-dihydropyrimido[4,5-c][1,2] oxazin-7-one (P. Kong Thoo Lin. and D.M. Brown, Nucleic Acids Res., 1989, 17, 10373-10383), 2-amino-6-methoxyaminopurine (D.M. Brown and P. Kong Thoo Lin, Carbohydrate Research, 1991, 216, 129-139), etc.

Hypoxanthine is one preferred fully univeral base. Nucleosides comprising hypoxanthine include, but are not limited to, inosine, isoinosine, 2'-deoxyinosine, and 7-deaza-2'-deoxyinosine, 2-aza-2'deoxyinosine.

[00117] Additional universal bases are known in the art as described, for example, in relevant portions of Loakes, D. and Brown, D. M., Nucl. Acids Res. 22:4039-4043, 1994; Ohtsuka, E. et al, J. Biol. Chem. 260 (5):2605-2608, 1985; Lin, P.K.T. and Brown, D. M., Nucleic Acids Res. 20(19):5149-5152, 1992; Nichols, R. et al., Nature 369(6480): 492-493, 1994; Rahmon, M. S. and Humayun, N. Z., Mutation Research 377 (2): 263-8, 1997; Berger, M., et al., Nucleic Acids Research, 28 (15):2911-2914, 2000; Amosova, O., et al., Nucleic Acids Res. 25 (10): 1930-1934, 1997; and Loakes, D., Nucleic Acids Res. 29(12):2437-47, 2001. The universal base may, but need not, form hydrogen bonds with an oppositely located base. The universal base may form hydrogen bonds via Watson-Crick or non- Watson-Crick interactions (e.g., Hoogsteen interactions).

[00118] In certain embodiments of the invention rather than using an oligoucleotide probe comprising a universal base, an oligonucleotide probe comprising an abasic residue is used. The abasic residue can occupy a position opposite any of the four naturally occurring nucleotides and can thus serve the same function as a nucleotide comprising a universal base. In some embodiments of the invention the linkage adjacent to an abasic residue is cleaved by an AP endonuclease, but abasic residues are also of use as described here (i.e., to serve the function of a universal base) in embodiments in which other scissile linkages (e.g., phosphorothiolates) are present and other cleavage reagents are used.

Detailed Description of Certain Preferred Embodiments of the Invention

[00119] A. Sequencing by Successive Cycles of Extension, Ligation, and Cleavage [00120] The overall scheme of one aspect of the invention is shown diagrammatically in Fig. IA, and generally resembles a method taught in U.S. Pat. Nos. 5,740,341 and 6,306,597, both issued to Macevicz. For purposes of convenience, these patents will be referred to collectively as "Macevicz" herein. In particular, Macevicz teaches a method for identifying a sequence of

nucleotides in a polynucleotide, the method comprising the steps of: (a) extending an initializing oligonucleotide along the polynucleotide by ligating an oligonucleotide probe thereto to form an extended duplex; (b) identifying one or more nucleotides of the polynucleotide; and (c) repeating steps (a) and (b) until the sequence of nucleotides is determined.

[00121] Macevicz further teaches a method for determining a sequence of nucleotides in a template polynucleotide, the method comprising the steps of: (a) providing a probe-template duplex comprising an initializing oligonucleotide probe hybridized to a template polynucleotide, said probe having an extendable probe terminus; (b) ligating an extension oligonucleotide probe to said extendable probe terminus, to form an extended duplex containing an extended oligonucleotide probe; (c) identifying, in the extended duplex, at least one nucleotide in the template polynucleotide that is either (1) complementary to the just- ligated extension probe or (2) a nucleotide residue in the template polynucleotide which is immediately downstream of the extended oligonucleotide probe; (d) generating an extendable probe terminus on the extended probe, if an extendable probe terminus is not already present, such that the terminus generated is different from the terminus to which the last extension probe was ligated; and (e) repeating steps (b), (c) and (d) until a sequence of nucleotides in the target polynucleotide is determined. In certain embodiments of these methods each extension probe has a chain-terminating moiety at a terminus distal to the initializing oligonucleotide probe. In certain embodiments the step of regenerating includes cleaving a chemically scissile internucleosidic linkage in the extended oligonucleotide probe. [00122] Referring to Fig. IA, polynucleotide template 20 comprising a polynucleotide region 50 of unknown sequence and binding region 40 is attached to support 10. Nucleotide 41, at the distal end of binding region 40, and nucleotide 51, at the proximal end of polynucleotide region 50, are adjacent to one another. An initializing oligonucleotide 30 is provided that hybridizes with binding region 40 to form a duplex at a location in binding region 40. Initializing oligonucleotide 30 is also referred to as a "primer" herein, and binding region 40 may be referred to as a "primer binding region". The duplex may, but need not be, a perfectly matched duplex. The initializing oligonucleotide has an extendable terminus 31. In Fig. IA, the initializing

oligonucleotide binds to the binding region such that extendable terminus 31 is located opposite nucleotide 41. However, the initializing oligonucleotide could bind elsewhere in the binding region, as discussed further below. An extension oligonucleotide probe 60 of length N is hybridized to the template adjacent to the initializing oligonucleotide. Terminal nucleotide 61 of the extension oligonucleotide probe is ligated to extendable terminus 31.

[00123] Terminal nucleotide 61 is complementary to the first unknown nucleotide in polynucleotide region 50. Therefore, the identity of terminal nucleotide 61 specifies the identity of nucleotide 51. Preferably nucleotide 51 is identified by detecting a label (not shown) associated with an extension probe known to have A, G, C, or T, as terminal nucleotide 61. The label is removed following detection. Fig. 2 shows a scheme for assigning different labels, e.g., fluorophores of different colors, to extension probes having different 3' terminal nucleotides.

[00124] Following ligation and detection, an extendable probe terminus is generated on extension probe 60 if probe 60 does not already have such a terminus. A second extension probe 70, preferably also of length N, is annealed to the template adjacent to extension probe 60 and is ligated to the extendable terminus of probe 60. The identity of terminal nucleotide 71 of extension probe 70 specifies the identity of oppositely located nucleotide 52 in polynucleotide 50. Terminal nucleotide 71 therefore constitutes the "sequence determining portion" of the extension probe, by which is meant the portion of the probe whose hybridization specificity is used as a basis from which to determine the identity of one or more nucleotides in the template. It will be appreciated that typically additional nucleotides in the extension probe will hybridize with the template, but only those nucleotides in the probe whose identity is associated with a particular label are used to identify nucleotides in the template. [00125] In preferred embodiments of the invention, generation of the extendable terminus involves cleavage of an internucleoside linkage as described further below. Preferably cleavage also removes the label. Cleavage removes a number of nucleotides M from the extension probe (not shown). Therefore, the duplex is extended by N-M nucleotides in each cycle, and nucleotides located at intervals of N-M in the template are identified. It is to be understood that

multiple copies of a given template will typically be attached to a single support, and the sequencing reaction will be performed simultaneously on these templates. [00126] Macevicz teaches that the oligonucleotide probes should generally be capable of being ligated to an initializing oligonucleotide or extended duplex to generate the extended duplex of the next extension cycle; the ligation should be template-driven in that the probe should form a duplex with the template prior to ligation; the probe should possess a blocking moiety to prevent multiple probe ligations on the same template in a single extension cycle; the probe should be capable of being treated or modified to regenerate an extendable end after ligation; and the probe should possess a signaling moiety (i.e., a detectable moiety) that permits the acquisition of sequence information relating to the template after a successful ligation.

[00127] Macevicz teaches characteristics of certain suitable initializing oligonucleotides, extension oligonucleotide probes, templates, binding sites, and various methods for synthesizing, designing, producing, or obtaining such components. Macevicz further teaches certain suitable ligases, ligation conditions, and a variety of suitable labels. Macevicz also teaches an alternative method for identification using polymerase extension to add a labeled chain-terminating nucleotide to a newly ligated extension probe. The identity of the added nucleotide identifies the nucleotide located oppositely in the template. [00128] As will be appreciated by one of ordinary skill in the art, references to templates, initializing oligonucleotides, extension probes, primers, etc., generally mean populations or pools of nucleic acid molecules that are substantially identical within a relevant region rather than single molecules. Thus, for example, a "template" generally means a plurality of substantially identical template molecules; a "probe" generally means a plurality of substantially identical probe molecules, etc. In the case of probes that are degenerate at one or more positions, it will be appreciated that the sequence of the probe molecules that comprise a particular probe will differ at the degenerate positions, i.e., the sequences of the probe molecules that constitute a particular probe may be substantially identical only at the nondegenerate position(s). For purposes of description the singular form is to be understood to include single molecules and populations of substantially identical molecules. Where

it is intended to refer to a single nucleic acid molecule (i.e., one molecule), the terms "template molecule", "probe molecule", "primer molecule", etc., will be used. In certain instances the plural nature of a population of substantially identical nucleic acid molecules will be explicitly indicated.

[00129] A population of substantially identical nucleic acid molecules may be obtained or produced using any of a variety of known methods including chemical synthesis, biological synthesis in cells, enzymatic amplification in vitro from one or more starting nucleic acid molecules, etc. For example, using methods well known in the art, a nucleic acid of interest can be cloned by inserting it into a suitable expression vector, e.g., a DNA or RNA plasmid, which is then introduced into cells, e.g., bacterial cells, in which it replicates. Plasmid DNA or RNA containing copies of the nucleic acid of interest is then isolated from the cells. Genomic DNA isolated from viruses, cells, etc., or cDNA produced by reverse transcription of mRNA) can also be a source of a population of substantially identical nucleic acid molecules (e.g., template polynucleotides whose sequence is to be determined) without an intermediate step of cloning or in vitro amplification, though generally it is preferred to perform such an intermediate step.

[00130] It will be understood that members of a population need not be 100% identical, e.g., a certain number of "errors" may occur during the course of synthesis. Preferably at least 50% of the members of a population are at least 90%, or more preferably at least 95% identical to a reference nucleic acid molecule (i.e., a molecule of defined sequence used as a basis for a sequence comparison). More preferably at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 99%, or more of the members of a population are at least 90%, or more preferably at least 95% identical, or yet more preferably at least 99% identical to the reference nucleic acid molecule. Preferably the percent identity of at least 95% or more preferably at least 99% of the members of the population to a reference nucleic acid molecule is at least 98%, 99%, 99.9% or greater. Percent identity may be computed by comparing two optimally aligned sequences, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions, and

multiplying the result by 100 to yield the percentage of sequence identity. It will be appreciated that in certain instances a nucleic acid molecule such as a template, probe, primer, etc., may be a portion of a larger nucleic acid molecule that also contains a portion that does not serve a template, probe, or primer function. In that case individual members of a population need not be substantially identical with respect to that portion.
[00131] Macevicz teaches methods in which a template is attached to a support such as a bead and extension proceeds towards the end of the template that is located distal to the support, as shown in Fig. IA. Thus the binding region is located closer to the support than the unknown sequence, and the extended duplex grows in the direction away from the support. However, the inventors have unexpectedly discovered that the method can advantageously be practiced using an alternative approach in which the binding region is located at the end of the template that is distal to the support, and extension proceeds inwards toward the support. This embodiment is depicted in Fig. IB, in which the various elements are numbered as in Fig. IA. The inventors have determined that sequencing "inwards" from the distal end of the template towards the support provides superior results. In particular, sequencing from the distal end of the template towards a support such as a bead results in higher ligation efficiencies than sequening outwards from the support.
[00132] As further taught by Macevicz, preferably the oligonucleotide probes are applied to templates as mixtures comprising oligonucleotides of all possible sequences of a predetermined length. For example, a mixture of probes containing all possible sequences of 6 nucleotides in length (hexamers) of structure NNNNNN (which may also be represented as (N)k, where k = 6) would contain 4<6> (4096) probe species. Generally the probes are of structure X(N)kN*, where N represents any nucleotide, and k is between 1 and 100, * represents a label, and X represents a nucleotide whose identity corresponds to the label. In certain embodiments k is between 1 and 100, between 1 and 50, between 1 and 30, between 1 and 20, e.g., between 4 and 10. One or more of the nucleotides may comprise a universal base. Generally the probe is 4-fold degenerate at positions represented by N or comprises a degeneracy-reducing nucleotide at one or more positions represented by N. If desired, the mixture can be divided into subsets of probes ("stringency classes) whose perfectly matched duplexes with complementary

sequences have similar stability or free energy of binding. The subsets may be used in separate hybridization reactions as taught by Macevicz. [00133] The complexity (i.e., the number of different sequences) of probe mixtures can be reduced by a number of methods, including using so-called degeneracy-reducing nucleotides or nucleotide analogs. For example, a library of probes containing all possible sequences of 8 nucleotides would contain 4<8> probes. The number of probes can be reduced to 4<6> while retaining various desirable features of an octamer library, such as the length, by using universal bases at two of the positions. The present invention comprehends the use of any of the universal bases mentioned above or described in the references cited above.[00134] Depending on the embodiment, the extended duplex or initializing oligonucleotide may be extended in either the 5' -^ 3' direction or the 3' -^ 5' direction by oligonucleotide probes, as described further below. Generally, the oligonucleotide probe need not form a perfectly matched duplex with the template, although such binding may be preferred. In embodiments in which a single nucleotide in the template is identified in each extension cycle, perfect base pairing is only required for identifying that particular nucleotide. For example, in embodiments where the oligonucleotide probe is enzymatically ligated to an extended duplex, perfect base pairing, i.e. proper Watson-Crick base pairing, is required between the terminal nucleotide of the probe which is ligated and its complement in the template. Generally, in such embodiments, the rest of the nucleotides of the probe serve as "spacers" that ensure the next ligation will take place at a predetermined site, or number of bases, along the template. That is, their pairing, or lack thereof, does not provide further sequence information. Likewise, in embodiments that rely on polymerase extension for base identification, the probe primarily serves as a spacer, so specific hybridization to the template is not critical.

[00135] The methods described above allow partial determination of a sequence, i.e., the identification of individual nucleotides spaced apart from one another in a template. In preferred embodiments of the invention, in order to gather more complete information, a plurality of reactions is performed in which each reaction utilizes a different initializing oligonucleotide i. The initializing oligonucleotides i bind to different portions of the binding region. Preferably the initializing oligonucleotides bind at

positions such the extendable termini of the different initializing oligonucleotides are offset by 1 nucleotide from each other when hybridized to the binding region. For example, as shown in Fig. 3, seqeuncing reactions 1...N are performed. Initializing oligonucleotides ii... in have the same length and bind such that their terminal nucleotides 31, 32, 33, etc., hybridize to successive adjacent positions 41, 42, 43, etc., in binding region 40. Extension probes ei...en thus bind at successive adjacent regions of the template and are ligated to the extendable termini of the initializing oligonucleotides. Terminal nucleotide 61 of probe en ligated to in is complementary to nucleotide 55 of polynucleotide region 50, i.e., the first unknown polynucleotide in the template. In the second cycle of extension, ligation, and detection, terminal nucleotide 71 of probe ei2 is complementary to nucleotide 56 of polynucleotide region 50, i.e., the second nucleotide of unknown sequence. Likewise, terminal nucleotides of extension probes ligated to duplexes initialized with initializing oligonucleotides i2, i3, i4, and so on, will be complementary to the third, fourth, and fifth nucleotides of unknown sequence 50. It will be appreciated that the initializing oligonucleotides may bind to regions progressively further away from polynucleotide region 50 rather than progressively closer to it.

[00136] The spacer function of the non-terminal nucleotides of the extension probes allows the acquisition of sequence information at positions in the template that are considerably removed from the position at which the initializing oligonucleotide binds without requiring a correspondingly large number of cycles to be performed on any given template. For example, by successive cycles of ligation of probes of length N, followed by cleavage to remove a single terminal nucleotide from the extension probe, nucleotides at intervals of N-I nucleotides can be identified in successive rounds. For example, nucleotides at positions 1, N, 2N- 1, 3N-2, 4N-3, and 5N-4 in the template can be identified in 6 cycles where the nucleotide at position 1 in the template is the nucleotide opposite the nucleotide that is ligated to the extendable probe terminus in the duplex formed by the binding of the initializing oligonucleotide to the template. Similarly, if cleavage removes two nucleotides from the extension probes of length N, then nucleotides at positions separated from each other by N-2 nucleotides can be identified in successive rounds. For example,

nucleotides at positions 1, N-l, 2N-3, 3N-5, 4N-7 in the template can be identified in 6 cycles. Thus if the probes are 8 nucleotides in length and 2 nucleotides are removed in each cycle, nucleotides at positions 1, 7, 13, 19, and 25 are identified. Thus the number of cycles needed to identify a nucleotide at a distance X from the first nucleotide in the template is on the order of X/M, where M is the length of the extension probe that remains following cleavage, rather than on the order of X.

[00137] For example, the schematic depicted in Figure 3B shows the net result of using the extension, ligation, and cleavage method with extension probes designed to read every 6th base of the template. By serially stripping and sequencing the template using 6 initializing nucleotides that bind to positions that are offset within the binding region and combining the results, all template bases are elucidated over a defined length. For instance, if 10 serial ligations are performed for each of the 6 reactions, the resulting read length will be 60 sequential base pairs, whereas if 15 serial ligations are performed for each reaction the resultant read length will be 90 sequential base pairs.

[00138] While not wishing to be bound by any theory, the inventors suggest that in contrast to this approach, most serial sequencing by synthesis methods struggle with error accumulation that ultimately limits the potential for long read lengths. An advantageous feature of certain of the methods described herein is that they allow the identification of every n<[Lambda]> base (depending on the position of the cleavable moiety in the probe), such that after a given number of cycles (y), one reaches the n*y-(n-l)th base (e.g., the 71<st> base in the foregoing example after 15 cycles, or the 115th base after 20 cycles using a probe with 6 bases on the 3' side of the cleavage site). The ability to "reset" the initializing oligonucleotide at the n-1, n-2, etc., positions greatly minimizes serial error accumulation (via dephasing or attrition) for a given read length since the process of stripping the extended strands from the template and hybridizing a new initializing oligonucleotide effectively resets background signals to zero. For example, comparing the polymerase based sequencing by synthesis and the ligation based approaches described herein, if the signal to noise ratio at each extension cycle is 99:1, the ratio after 100 cycles for the polymerase based approach will be 37:63

and for the ligase based method, 85:15. The net result for the ligase based method is a large increase in read length over polymerase based methods.

[00139] The ability to identify nucleotides using fewer cycles than would be required if it was necessary to perform a cycle for each preceding nucleotide in the template is important for a number of reasons. In particular, it is unlikely that each step in the method will occur with 100% efficiency. For example, some templates may not be successfully ligated to an extension probe; some extension probes may not be cleaved, etc. Thus in each cycle the reactions occurring on different copies of the template become progressively dephased, and the number of templates from which useful and accurate information can be acquired is reduced. It is thus particularly desirable to minimize the number of cycles required to read nucleotides located more than a few positions away from the extendable terminus of the initializing oligonucleotide. However, increasing the length of the extension probe potentially results in greater complexity of the probe mixture, which decreases the effective concentration of each individual probe sequence. As described herein, degeneracy-reducing nucleotides can be used to reduce the complexity but may result in decreased hybridization strength and/or decreased ligation efficiency. The inventors have recognized the need to balance these competing factors in order to optimize results. Thus in a preferred embodiment of the invention extension probes 8 nucleotides in length are used, with degeneracy-reducing nucleotides at selected positions. In addition, the inventors have recognized the importance of selecting appropriate scissile linkages and cleavage conditions and times to optimize the efficiency of the cleavage step (i.e., the percentage of linkages that is successfully cleaved in each cleavage step) and its specificity for the appropriate linkage.
[00140] B. Oligonucleotide Extension Probe Design
[00141] While Macevicz mentions that degeneracy-reducing nucleoside analogs may be used in the oligonucleotide extension probes, he does not teach specific positions at which it is particularly desirable to include a residue comprising such residues in the extension probes and does not teach particular probe structures (i.e., sequences) that incorporate degeneracy- reducing nucleosides. The present inventors have recognized that it may be particularly advantageous to utilize degeneracy-reducing

nucleosides (e.g., nucleosides that comprise a universal base) at particular positions and in particular numbers in the oligonucleotide extension probes. For example, in certain embodiments of the invention most or all of the nucleotides at position 6 or greater (counting from X), comprise a universal base. For example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 100% of the nucleotides at position 6 or greater may comprise a univeral base. The nucleotides need not all comprise the same universal base. In certain embodiments of the invention hypoxanthine and/or a nitro-indole is used as a universal base. For example, nucleosides such as inosine can be used.

[00142] The inventors have recognized that superior results may be achieved using extension probes that are greater than 6 nucleotides in length, and in which one or more of the nucleotides at position 6 or greater from the proximal terminus of the probe, counting from the nucleotide to be ligated to the extendable probe terminus, is a degeneracy-reducing nucleotide, e.g., comprises a universal base (i.e., if the most proximal nucleotide is considered position 1, one or more of the nucleotides at position 6 or greater comprises a universal base), e.g., 1, 2, or 3 of the nucleotides at position 6 or greater in the case of octamer probes comprises a universal base. For example, for sequencing in the 3 '-^5' direction, probes having the structure 3'-XNNNNsINI-5' can be used, where X and N represent any nucleotide, "s" represents a scissile linkage, such that cleavage occurs between the fifth and sixth residues counting from the 3' end, and at least one of the residues between the scissile linkage and the 5' end preferably has a label that corresponds to the identity of X. Another design is 3'-XNNNNsNII-5'. Yet another probe design is 3'-XNNNNsIII-5'. This design yields a probe mixture with a modest complexity of 1024 different species, is long enough to prevent formation of significant adenylation products (see Example 1), and has the advantage that the resulting extension product remaining after cleavage would consist of unmodified DNA. One drawback is that this probe extends the primer by only 5 bases at a time. Since the read length is a function of the extension length times the number of cycles, each additional base on the extension length has the potential to increase the read length by the lx the cycle number (e.g. 20 bases if 20 cycles are used). Another probe design leaves one or more inosines (or other universal base) at the end of the extension probe

following cleavage to create a 6 base, or longer, extended duplex. For example, with the probe 3'-XNNNNIsII-5', the duplex would be extended by 6 bases at a time, leaving a 5' inosine at the junction. In each of these designs, at least one of the residues between the scissile linkage and the 5' end preferably has a label that corresponds to the identity of X. In certain embodiments of the invention the third nucleotide from the distal terminus of the probe, counting from the end opposite the nucleotide to be ligated to the extendable probe terminus, comprises a universal base, (i.e., if the distal terminus is considered position K, the nucleotide at position K-2 comprises a universal base).

[00143] In certain embodiments of the invention locked nucleic acid (LNA) bases are used at one or more positions in an initializing oligonucleotide probe, extension probe, or both. Locked nucleic acids are described, for example, in U.S. Pat. No. 6,268,490; Koshkin, AA, et al, Tetrahedron, 54:3607-3630, 1998; Singh, SK, et al, Chem. Comm., 4:455-456, 1998. LNA can be synthesized by automatic DNA synthesizers using standard phosphoramidite chemistry and can be incorporated into oligonucleotides that also contain naturally occurring nucleotides and/or nucleotide analogues. They can also be synthesized with labels such as those described below.

[00144] C. Templates, Libraries,Supports, Blockers, and Methods for their Preparation and Use

[00145] The invention provides a variety of methods for preparing nucleic acid templates and supports. The invention also provides libraries for use in ligation-based sequencing or for other purposes. The invention also provides blocker oligonucleotides and methods of using them in the context of sequencing by successive cycles of oligonucleotide ligation, detection, and cleavage of for other purposes.

[00146] Macevicz teaches a process in which a template comprising a plurality of substantially identical template molecules is first synthesized, e.g., by amplification in a tube or other vessel as in conventional polymerase chain reaction (PCR) methods. Macevicz teaches that the amplified template molecules are preferably attached to supports such as magnetic microparticles (e.g., beads) after synthesis.

[00147] The inventors have recognized that templates to be sequenced may desirably be synthesized on or in a support itself, e.g., by using supports such as microparticles or various semi-solid support materials such as gel matrices to which one of a pair of amplification primers is attached prior to performing the PCR reaction. This approach avoids the need for a separate step of attaching the template molecules to the support after synthesis. Thus a plurality of template species of differing sequence can be conveniently amplified in parallel. For example, according to the methods described below, synthesis on microparticles results in a population of individual microparticles, each with multiple copies of a particular template molecule (or its complement) attached thereto, wherein the template molecules attached to each microparticle differ in sequence from the template molecules attached to other microparticles. Each of the supports thus has a clonal population of templates attached thereto, e.g., support A will have multiple copies of template X attached thereto; support B will have multiple copies of template Y attached thereto; support C will have multiple copies of template Z attached thereto, etc. By "clonal population of templates", "clonal population of nucleic acids", etc., is meant a population of substantially identical template molecules, preferably generated by successive rounds of amplification that start from a single template molecule of interest (starting template). The substantially identical template molecules may be substantially identical to the starting template or to its complement. [00148] Amplification is typically performed using PCR, but other amplification methods may also be used (see below). It will be understood that members of a clonal population need not be 100% identical, e.g., a certain number of "errors" may occur during the course of synthesis, e.g., during amplification. Preferably at least 50% of the members of a clonal population are at least 90%, or more preferably at least 95% identical to a starting template molecule (or to its complement). More preferably at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 99%, or more of the members of a population are at least 90%, or more preferably at least 95% identical, or yet more preferably at least 99% identical to the starting template molecule (or to its complement). Preferably the percent identity of at least 95% or more preferably at least 99% of the members of the population to a starting template molecule (or to its complement) is at least 98%, 99%, 99.9% or greater. [00149] Amplification primers may be

attached to supports using any of a variety of techniques. For example, one end of the primer (the 5' end) of the primer may be functionalized with one member of a binding pair (e.g., biotin), and the support functionalized with the other member of the binding pair (e.g., streptavidin). Any similar binding pair may be used. For example, nucleic acid tags of defined sequence may be attached to the support and primers having complementary nucleic acid tags can be hybridized to the the nucleic acid tags attached to the support. Various linkers and crosslinkers can also be used.

[00150] Methods for performing PCR are well known in the art and are described, for example, in U.S. Pat. Nos. 4,683,195, 4,683,202, and 4,965,188, and in Dieffenbach, C. and Dveksler, GS, PCR Primer: A Laboratory Manual, 2<nd> ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2003. Methods for amplifying nucleic acids on microparticles are well known in the art and are described, for example, standard PCR can be performed in wells of a microtiter dish or in tubes on beads with primers attached thereto (e.g., beads prepared as in Example 12. While PCR is a convenient amplification method, any of numerous other methods known in the art can also be used. For example, multiple strand displacement amplification, helicase displacement amplification (HDA), nick translation, Q beta replicase amplification, rolling circle amplification, and other isothermal amplification methods etc., can be used.

[00151] Template molecules can be obtained from any of a variety of sources. For example, DNA may be isolated from a sample, which may be obtained or derived from a subject. The word "sample" is used in a broad sense to denote any source of a template on which sequence determination is to be performed. The phrase "derived from" is used to indicate that a sample and/or nucleic acids in a sample obtained directly from a subject may be further processed to obtain template molecules. The source of a sample may be of any viral, prokaryotic, archaebacterial, or eukaryotic species. In certain embodiments of the invention the source is a human. The sample may be, e.g., blood or another body fluid containing cells; sperm; a biopsy sample, etc. Genomic or mitochondrial DNA from any organism of interest may be sequenced. cDNA may be sequenced. RNA may also be sequenced, e.g., by first reverse

transcribing to yield cDNA, using methods known in the art such as RT-PCR. Mixtures of DNA from different samples and/or subjects may be combined. Samples may be processed in any of a variety of ways. Nucleic acids may be isolated, purified, and/or amplified from a sample using known methods. Of course entirely artificial, synthetic nucleic acids, recombinant nucleic acids not derived from an organism can also be sequenced.

[00152] Templates can be provided in double or single-stranded form. Typically when a template is initially provided in double-stranded form the two strands will subsequently be separated (e.g., the DNA will be denatured), and only one of the two strands will be amplified to produce a localized clonal population of template molecules, e.g., attached to a microparticle, immobilized in or on a semi-solid support, etc.

[00153] Templates may be selected or processed in a variety of additional ways. For example, templates obtained from DNA that has been subjected to treatment to with a methyl-sensitive restriction enzyme (e.g., MspI) can be used. Such treatment, which results in DNA fragments, can be performed prior to amplification. Fragments containing methylated bases do not amplify. Sequence information obtained from the hypomethylated templates may be compared with sequence information obtained from templates derived from the same source, which were not subjected to selection for hypomethylation. [00154] Templates may be inserted into, provided in, or derived from a library. For example, hypomethylated libraries are known in the art. Inserting templates into libraries can allow for the convenient concatenation of additional nucleotide sequences to the ends of templates, e.g., tags, binding sites for primers or initializing oligonucleotides, etc. For example, certain strategies allow the addition of tags having a plurality of binding sites, e.g., a binding site for an amplification primer, a binding site for an initializing oligonucleotide, a binding site for a capture agent, etc.

[00155] A variety of suitable libraries are known in the art. For example, libraries of particular interest, and methods for their construction, are described in USSN 10/978,224, PCT publications WO2005042781 and WO2005082098, and Shendure, J., et al, Science, 309 (5741):1728-32, 2005, Sciencexpress, 4 Aug. 2005 (www.sciencexpress.org). Of course it will be

understood that other methods of generating such libraries could also be used. Certain libraries of particular interest contain a plurality of nucleic acid fragments (typically DNA), each of which contain two nucleic acid segments of interest, separated by sequences that are complementary to amplification and/or sequencing primers that are used in sequencing steps, i.e., these sequences serve as primer binding regions (PBRs). In embodiments of particular interest, the nucleic acid segments are portions of a contiguous piece of naturally occurring DNA. For example, the segments may be from the 5 ' and 3 ' end of a contiguous piece of genomic DNA as desribed in the afore-mentioned references. Such nucleic acid segments are referred to herein in a manner consistent with the afore-mentioned references, as "tags" or "end tags". Two tags derived from a single contiguous nucleic acid, e.g., from the 5' and 3' ends thereof, are referred to as "a paired tag", "paired tags", or "a ditag". It will be appreciated that a "paired tag" comprises two tags, even if used in the singular. By selecting the contiguous pieces of DNA from which the tags of a paired tag are derived to be within a predefined size limit, the distance separating the two tags is constrained. [00156] In addition to being separated by sequences that are complementary to sequencing and/or amplification primers, the nucleic acid fragments of the libraries typically also contain sequences complementary to sequencing and/or amplification primers flanking the tags, i.e., a first such sequence may be located 5' to the tag that is closer to the 5' end of the fragment, and a second such sequence may be located 3 ' to the tag that is located closer to the 3 ' end of the fragment. It is noted that the position of the two tags as present in the contiguous nucleic acid from which the tags are derived may, but need not, correspond with the position of the tag in the DNA fragment of the library in various embodiments. [00157] The nucleic acid fragments and the tags can have a range of different sizes. Typically the nucleic acid fragments may be, for example, between 80 and 300 nucleotides in length, e.g., between 100-200, 100-150, approximately 150 nucleotides in length, approximately 200 nucleotides in length, etc. The tags can be, e.g., between 15-25 nucleotides in length, e.g., approximately 17-18 nucleotides in length, etc. It is noted that these lengths are exemplary and are not intended to be limiting. Shorter or longer fragments and/or tags could be used.

[00158] It should also be noted that while obtaining the paired tags from a single contiguous nucleic acid affords a convenient method for library construction, the important aspect of the paired tags is the fact that they are separated from one another by a distance ("separation distance") in the nucleic acid from which they were originally derived, wherein the separation distance falls within a predetermined range of distances. The fact that the tags are separated by a separation distance that falls within a predetermined range allows the sequence of the tags to be aligned against a reference sequence (e.g., a reference genome sequence). Without wishing to be bound by any theory, this can be advantageous in certain applications such as genome resequencing, wherein it allows the use of shorter read lengths while still allowing accurate placement of the sequences with respect to the reference genome. The 5' and 3' tags of a paired tag represent (i.e., they have the sequence of) segments of a larger piece of nucleic acid, e.g., genomic DNA, which segments are located within a predefined distance from one another in a naturally occurring piece of DNA, e.g., within a piece of genomic DNA. For example, in certain embodiments of the invention the 5' and 3' tags of a paired tag represent segments of DNA located within up to 500 nucleotides of each other, within up to 1 kB of each other, within up to 2 kB of each other, within up to 5 kB of each other, within up to 10 kB of each other, within up to 20 kB of each other, in a naturally occurring piece of DNA. In certain embodiments the 5' and 3' tags of a paired tag are located between 500 nucleotides and 2 kB apart, e.g, between 700 nucleotides and 1.2 kB apart, approximately 1 kB apart, etc., in a naturally occurring piece of DNA. It is noted that the exact distance separating the two tags of a paired tag is not of major importance and is typically not known. In addition, while the tags are originally obtained from a larger piece of nucleic acid, the word "tag" applies to any nucleic acid segment that has the sequence of the tag, whether present in its original sequence context or in a library fragment, amplification product from a library fragment, template to be sequenced, etc. [00159] A nucleic acid fragment (e.g., a library molecule) may have the following structure:

[00160] Linker 1 - Tag 1 - Linker 3 - Tag 1 - Linker 2

[00161] Tag 1 and Tag 2 can be 5' and 3' tags of a paired tag. Either of the tags can be the 5 ' tag or the 3 '

tag. Linker 1 and Linker 2 contain primer binding regions for one or more primers. In certain embodiments Linkers 1 and 2 each contain a PBR for an amplification primer and a PBR for a sequencing primer. The primers in each linker can be nested, such that the sequencing primer PBR is located internal to the amplification primer PBR. Linker 3 may contain PBRs for one or more sequencing primers to allow for sequencing of Tag 1 and Tag 2. The term "linker" as used in reference to a library of nucleic acid fragments refers to a nucleic acid sequence that is present in multiple nucleic acid fragments of a library, e.g., in substantially all fragments of the library. A linker may or may not actually have served a linking function during construction of the library and can simply be considered to be a defined sequence that is common to most or all members of a given library. Such a sequence is also referred to as a "universal sequence". Thus a nucleic acid complementary to the linker or a portion thereof would hybridize to multiple members of the library and could be used as an amplification primer or sequencing primer for most or all molecules in the library. [00162] In certain embodiments of the present invention, a nucleic acid fragment has the following structure:

[00163] Linker 1 - Tag 1 - Internal Adaptor - Tag 2 - Linker 2

[00164] Tag 1 and Tag 2 and Linker 1 and Linker 2 contain PBRs as described above. Internal Adaptor contains two primer binding regions, which may be referred to as IA and IB, as discussed further below. These PBRs are of use to produce microparticles having two distinct substantially identical populations of nucleic acids attached thereto, wherein nucleic acids of one of the populations comprise Tag 1 and nucleic acids of the other population comprise Tag 2. The two distinct populations of nucleic acids have at least partially different sequences, e.g., they differ in the sequence of the tag regions. The Internal adaptor can contain a spacer region between the two primer binding regions. The spacer region may contain abasic residues, which will prevent a polymerase from extending through the spacer. Of course spacer regions containing any other blocking group that would prevent polymerase extension through the spacer could be used.

[00165] In other embodiments, a nucleic acid fragment includes one or more additional tags (e.g, 2, 4, 6, etc.)

and one or more additional internal adaptors. For example, a nucleic acid fragment can have the following structure:

[00166] Linker 1 - Tag 1 - Internal Adaptor 1 - Tag 2 - Linker 2 - Tag 3 - Internal Adaptor 2 - Tag 4 - Linker 3

[00167] It is noted that the inventive nucleic acid fragments and libraries of such fragments, microparticles containing two or more substantially identical populations of nucleic acids, and arrays of such microparticles can be used in a wide variety of sequencing methods other than the ligation-based sequencing methods described herein. For example, sequencing methods such as FISSEQ, pyrosequencing, etc., can be used. See, e.g., WO2005082098. Of course the ligation-based methods can also advantageously be employed. It will be appreciated that in the context of the ligation-based methods described herein, the term "sequencing primer" may be understood to mean "initializing oligonucleotide".

[00168] In certain embodiments of the invention the templates to be sequenced are synthesized by PCR in individual aqueous compartments (also called "reactors") of an emulsion. Preferably the compartments each contain a particulate support such as a bead having a suitable first amplification primer attached thereto, a first copy of the template, a second amplification primer, and components needed for the PCR reaction (e.g., nucleotides, polymerase, cofactors, etc.). Methods for preparing emulsions are described for example, in U.S. Pat. Nos. 6,489,103 (Griffiths); 5,830,663 (Embleton); and in U.S. Pub. No. 20040253731 (Ghadessy). Methods for performing PCR within individual compartments of an emulsion to produce clonal populations of templates attached to microparticles ("emulsion PCR") are described, e.g., in Dressman, D., et al, Proc. Natl. Acad. ScL, 100 (15):8817-8822, 2003, and in PCT publication WO2005010145.

[00169] Methods described in the afore -mentioned references, or modifications thereof, may be used to produce clonal populations of templates attached to microparticles for sequencing. In a preferred and non-limiting embodiment, short (<500 nucleotide) templates suitable for PCR are created by attaching (e.g., by ligation) a universal adaptor sequence to each end of a

population of different target sequences (templates). (Universal in this context means that the same adaptor sequence is attached to each template, to create "adapted" templates that can be amplified using a single pair of PCR amplification primers.) A bulk PCR reaction is prepared with the adapted templates, one free amplification primer, microparticles with a second amplification primer attached thereto, and other PCR reagents (e.g., polymerase, cofactors, nucleotides, etc.). The aqueous PCR reaction is mixed with an oil phase (containing light mineral oil and surfactants) in a 1 :2 ratio. This mixture is vortexed to create a water-in-oil emulsion. One milliliter of mixture is sufficient to create more than 4x10<9> aqueous compartments within the emulsion, each a potential PCR reactor. Aliquots of the emulsion sample are dispensed into the wells of a microtiter plate (e.g., 96 well plate, 384 well plate, etc.) and thermally cycled to achieve solid-phase PCR amplification on the microparticles. To ensure clonality, the microparticle and template concentrations are carefully controlled so that the reactors rarely contain more than one bead or template molecule. For example, in certain embodiments of the invention at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or more of the reactors contain a single bead and a single template. Members of each clonal populations of templates are thus spatially localized in proximity to one another as a result of their attachment to the microparticle. In general, the points of attachment of the templates may be substantially uniformly distributed on the surface of the particle. Microparticles that have a clonal population of templates attached thereto (typically many thousands to millions of copies of the templates) following an amplification procedure are referred to as having undergone template amplification. [00170] It is of particular interest to use PCR emulsion methods to produce populations of microparticles in which individual microparticles have distinct populations of amplified nucleic acid fragments that contain a 5' tag and a 3' tag of a paired tag attached thereto. In other words, it is of particular interest to produce populations of microparticles in which individual particles have different nucleic acid fragments from a library such as those described above amplified and attached thereto.

[00171] Methods known in the art for amplifying DNA in emulsions (e.g., described in the references mentioned above), are limited in terms of their ability to achieve amplification of large nucleic acid molecules and

attachment of these molecules to microparticles. For example, it has been demonstrated that the PCR efficiency decays exponentially with longer amplicons. This decrease in PCR efficiency reduces the efficiency with which nucleic acid fragments containing paired tags and primer binding sites, such as those described above, can be amplified in PCR emulsions and attached to microparticles via such amplification. Thus methods in which a single population of substantially identical nucleic acid fragments containing first and second tags of a paired tag are amplified in a PCR emulsion and attached to beads via such amplification suffer from a number of limitations. [00172] The present invention provides an approach that allows the use of smaller amplicons while still preserving the paired tag information that arises when a single nucleic acid fragment containing 5' and 3' tags of a paired tags is attached via amplification to a microparticle. The invention provides a microparticle, e.g., a bead, having at least two distinct populations of nucleic acids attached thereto, wherein each of the at least two populations consists of a plurality of substantially identical nucleic acids, and wherein a first population of substantially identical nucleic acids comprises a first nucleic acid segment of interest, e.g., 5' tag, and a second population of nucleic acids comprises a second nucleic acid segment of interest, e.g., 3' tag. The first and second populations of nucleic acids are amplified from a single larger nucleic acid fragement that contains the two tags and also contains appropriately positioned primer binding sites flanking and separating the tags, so that two amplification reactions can be performed either sequentially or, preferably, simultaneously, in a single reactor of a PCR emulsion in the presence of a microparticle and amplification reagents. The microparticle has attached thereto two different populations of primers, one of which corresponds in sequence with a primer binding region external to one of the tags in the nucleic acid fragment, and the other of which corresponds in sequence with a primer binding region external to the other tag in the nucleic acid fragment, i.e., the primer binding regions flank the two tags.

[00173] Also provided are primers that bind to primer binding regions located between the two tags, so that two separate PCR reactions can be performed, each amplifying a portion of the nucleic acid fragment containing one of the tags. The amplified nucleic acid

segments contain additional primer binding regions, which are different from one another. These additional primer binding regions are present in the nucleic acid fragment and are located internal to the PBRs for the amplification primers, i.e., they are nested. These additional PBRs serve as binding regions for two different sequencing primers. Thus by applying one or the other of the two different sequencing primers to a microparticle having the two populations of substantially identical nucleic acid segments attached thereto, either one or the other of the two nucleic acid segments can be sequenced without interference due to the presence of the other nucleic acid segment. Each of the nucleic acid segments is significantly shorter than the nucleic acid fragment from which it was amplified, thus improving the efficiency with which emulsion-based PCR can be performed using libraries of fragments containing paired tags, while still preserving the association between the tags of a paired tag. [00174] The methods described above may be better understood by reference to the various panels of Figures 34 and 35 in which portions of nucleic acids having the same sequence are assigned the same color. The description above is to be interpreted consistently with Figures 34 and 35. Figures 34A and 35 A show the same steps, with figure 35 A providing additional details. As shown in Figures 34A and 35A, paired-end library fragments containing two tags (Tag 1 and Tag 2) are constructed with an internal adapter cassette (IA- IB) and unique flanking linker sequences (PI and P2), i.e., PI and P2 are distinct from one another. Both the internal adapter cassette and the flanking linker sequences contain nucleotide sequences that afford both PCR amplification and DNA sequencing. PCR primer regions are designed as to allow the use of nested DNA sequencing primers. DNA capture microparticles (beads) are generated by attaching two oligonucleotide sequences that are identical to the unique flanking linker sequences. For PCR amplification, DNA capture microparticles bound with oligonucleotides having PI and P2 sequences, are seeded into reactions containing a single di-tag library fragment (i.e., a library fragment containing a 5' tag and 3' tag of a paired tag) and solution-based PCR primers.

[00175] Solution-based flanking linker primers (PI and P2) are added in limiting amounts in comparison to the internal adapter primers (IA and IB) and will serve to promote efficient drive-to-bead amplification of PCR-

generated tag products (i.e., [P1"IB], [P2"IA]). If desired, controlling the amount of primers appropriately can also ensure that the populations of nucleic acids contain substantially the same number of nucleic acids, e.g., approximately half the nucleic acids on an individual microparticle belong to the first population and approximately half the nucleic acids on an individual microparticle belong to the second population. Thus a form of asymmetric PCR can be employed, if desired, in order to control the ratio of the different populations.

[00176] During amplification, as shown in Figures 34B and 35B (where Figure 35B again provides additional details relative to Figure 34B), the single paired-end library fragment, in the presence of the four oligonucleotide primers (PI, P2, IA and IB), will generate two unique PCR products. One population contains Tag 1 flanked by PI and IA, and a second population contains Tag 2 flanked by P2 and IB.

[00177] Following amplification microparticles will be loaded with two unique PCR populations corresponding to Tag 1 and Tag 2 generated from the initial library fragment. Each tag thus contains a unique set of priming regions to allow serial sequencing of each tag as shown in Figures 34C, 35C, and 35D. Figures 35C and 35D show sequential sequencing of tags 1 and 2, using different sequencing primers. Any of a variety of sequencing methods can be used.

[00178] The above methods can be used to generate microparticles having more than two distinct populations of nucleic acid sequences attached thereto, e.g., 4, 6, 8, 12, 16, 20, populations, e.g., wherein the populations comprise 2, 3, 4, 6, 8, 10 paired tags. Each population can be individually sequenced by providing a unique primer binding region in each sequence, as described above in the case of two tags.

[00179] The invention encompasses nucleic acid fragments having the structures shown in Figures 34 and 35 and described above, libraries of such fragments, microparticles having nucleic acid segments from such fragments attached thereto, populations of such microparticles wherein the individual microparticles have populations of nucleic acids attached thereto that differ in sequence from those of other microparticles, arrays of microparticles, amplification primers for amplifying nucleic acid segments (tags) from the nucleic

acid fragments, sequencing primers for sequencing nucleic acid segments attached to microparticles, methods for making the fragments, libraries and microparticles, and methods of sequencing the nucleic acids attached to the microparticles. The invention encompasses kits containing any combination of the afore-mentioned components, optionally also containing one or more enzymes, buffers, or other reagents useful in amplification, sequencing, etc.

[00180] If desired, a variety of methods may be used to enrich for microparticles that have templates attached thereto. For example, a hybridization-based method can be used in which an oligonucleotide (capture agent) complementary to a portion of an amplification product (template) attached to the microparticles is attached to a capture entity such as another (preferably larger) microparticle, microtiter well, or other surface. The portion of the amplification product may be referred to as a target region. The target region may be incorporated into templates during amplification, e.g., at one end of the portion of the template having unknown sequence. For example, the target region may be present in the amplification primers that is not attached to the microparticle, so that a complementary portion is present in the amplified template. Thus multiple different templates can include the same target region, so that a single capture agent will hybridize to multiple different templates, allowing the capture of multiple microparticles using only a single oligonucleotide sequence as the capture agent. Microparticles that have been subjected to amplification are exposed to the capture agent under conditions in which hybridization can occur. As a result, microparticles having amplified templates attached thereto are attached to the capture entity via the capture agent. Unattached microparticles are then removed, and the retained microparticles released (e.g., by raising the temperature). In certain embodiments in which a particulate capture entity is used, aggregates consisting of the capture entity with microparticles attached thereto after hybridization are separated from particulate capture entities lacking attached microparticles and from microparticles that are not attached to a capture entity, e.g., by centrifugation in a viscous solution such as glycerol. Other methods of separation based on size, density, etc., can also be used. Hybridization is but one of a number of methods that can be used for enrichment. For example, capture agents having an affinity for any of a number of different

ligands that can be incorporated into a template (e.g., during synthesis) may be used. Multiple rounds of enrichment can be used. [00181] Figure 14A shows an image of compartments of a water-in-oil emulsion, in which PCR reactions were performed on beads having first amplification primers attached thereto, using a fluorescently labeled second amplification primer and an excess of template. Aqueous reactors fluoresce weakly from diffuse free primer whereas beads strongly fluoresce from primers accumulating on the bead as a result of solid-phase amplification (i.e., fluorescent primers are incorporated into the amplified templates that are attached to the beads via the first amplification primer). Bead signal is uniform in the different sized reactors.

[00182] Following amplification, microparticles are collected (e.g., by use of a magnet in the case of magnetic particles) and used for sequencing by repeated cycles of extension, ligation, and cleavage as described herein. In certain embodiments of the invention the microparticles are arrayed in or on a semi-solid support prior to sequencing, as described below. Examples 12, 13, 14, and 15 provide additional details of representative and nonlimiting methods that may be used to (i) prepare microparticles having an amplification primer attached thereto, for synthesis of templates on the microparticles (Example 12); (ii) preparation of an emulsion comprising a plurality of reactors for performing PCR (Example 13); (iii) PCR amplification in compartments of an emulsion (Example 13); (iv) breaking the emulsion and recovering microparticles (Example 13); (v) enriching for microparticles having clonal template populations attached thereto (Example 14); (vi) preparation of glass slides to serve as substrates for a semi-solid polyacrylamide support (Example 15); and (vii) mixing microparticles with unpolymerized acrylamide, forming an array of microparticles having templates attached thereto, embedded in acrylamide on a substrate (Example 15). Example 15 also describes a protocol for polymerase trapping, which is used in certain of the methods when performing PCR in a semi-solid support. One of ordinary skill in the art will recognize that numerous variations on these methods may be used.

[00183] In other embodiments of the invention, the templates are amplified by PCR in a semi-solid support such as a gel having suitable amplification primers

immobilized therein. Templates, additional amplification primers, and reagents needed for the PCR reaction are present within the semi-solid support. One or both of a pair of amplification primers is attached to the semi-solid support via a suitable linking moiety, e.g., an acrydite group. Attachment may occur during polymerization. Additional reagents (e.g., templates, second amplification primer, polymerase, nucleotides, cofactors, etc.) may be present in prior to formation of the semi-solid support (e.g., in a liquid prior to gel formation), or one or more of the reagents may be diffused into the semi-solid support after its formation. The pore size of the semi-solid support is selected to allow such diffusion. As is well known in the art, in the case of a polyacrylamide gel, pore size is determined mainly by the concentration of acrylamide monomer and to a lesser extent by the crosslinking agent. Similar considerations apply in the case of other semi-solid support materials. Appropriate cross-linkers and concentrations to achieve a desired pore size can be selected. In certain embodiments of the invention an additive such as a cationic lipid, polyamine, polycation, etc., is included in the solution prior to polymerization, which forms in-gel micelles or aggregates surrounding the microparticles. Methods disclosed in U.S. Pat. Nos. 5,705,628, 5,898,071, and 6,534,262 may also be used. For example, various "crowding reagents" can be used to crowd DNA near beads for clonal PCR. SPRI (R) magnetic bead technology and/or conditions can also be employed. See, e.g., U.S. Pat. No. 5,665,572, demonstrating effective PCR amplification in the presence of 10% polyethylene glycol (PEG). In certain embodiments of the inventive methods amplification (e.g., PCR), ligation, or both, are performed in the presence of a reagent such as betaine, polyethylene glycol, PVP -40, or the like. These reagents may be added to a solution, present in an emulsion, and/or diffused into a semi-solid support. [00184] The semi-solid support may be located or assembled on a substantially planar rigid substrate. In certain preferred embodiments the substrate is transparent to radiation of the excitation and emission wavelengths used for excitation and detection of typical labels (e.g., fluorescent labels, quantum dots, plasmon resonant particles, nanoclusters), e.g., between approximately 400-900 nm. Materials such as glass, plastic, quartz, etc., are suitable. The semi-solid support may adhere to the substrate and may optionally be affixed to the substrate using any of a variety of methods. The

substrate may or may not be coated with a substance that enhances adherence or bonding, e.g., silane, polylysine, etc. U.S. Pat. No. 6,511,803 describes methods for synthesizing clonal populations of templates using PCR in semi-solid supports, methods for preparing semi-solid supports on substantially planar substrates, etc. Similar methods may be used in the present invention. The substrate may have a well or depression to contain the liquid prior to formation of the semi-solid substrate. Alternately, a raised barrier or mask may be used for this purpose.

[00185] The above approach provides an alternative to the use of reactors in emulsions to generate spatially localized populations of clonal templates. The clonal populations are present at discrete locations in the semi-solid support, such that a signal can be acquired from each population during sequencing for purposes of detecting a newly ligated extension probe, e.g., by imaging. In some embodiments of the invention, two or more distinct clonal populations are amplified from a single nucleic acid fragment and are present as a mixture at a discrete location in the semi-solid support. Each of the clonal populations in the mixture may comprise a tag, e.g., so that the discrete location contains fragments containing a 5' tag and fragments containing a 3' tag. The clonal templates comprising the 5' tag and the 3' tag contain different sequencing primers, so that they can be sequenced independently of one another. This approach is identical to the approach described above for producing multiple populations of substantially identical nucleic acids on a microparticle and obtaining sequencing information for both members of a paired tag from a single microparticle. [00186] In general, a semi-solid support for use in any of the inventive methods forms a layer of about 100 microns or less in thickness, e.g., about 50 microns thick or less, e.g., between about 20 and 40 microns thick, inclusive. A cover slip or other similar object having a substantially planar sufrace can be placed atop the semi-solid support material, preferably prior to polymerization, to help produce a uniform gel layer, e.g. to form a gel layer that is substantially planar and/or substantially uniform in thickness.

[00187] In yet other embodiments of the invention, modifications to the above methods are used, in which templates are synthesized by PCR on microparticles having a suitable amplification primer attached thereto,

wherein the microparticles are immobilized in or on a semi-solid support prior to template synthesis, i.e., they are fully or partially embedded in the semi-solid support. Generally the microparticles are completely surrounded by the semi-solid support material, though they may rest on an underlying substrate. The microparticles thus remain at substantially fixed positions with respect to one another unless the semi-solid support is disrupted. This approach provides another alternative to the use of emulsions to generate spatially localized populations of clonal templates. Microparticles may be mixed with liquid prior to formation of the semi-solid support. Alternatively, microparticles may be arrayed on a substantially planar substrate, and liquid added to the microparticle array prior to polymerization, crosslinking, etc. The microparticles have a first amplification primer attached thereto. The second amplification primer may, but need not be, be attached to the semi-solid support. Additional reagents (e.g., template, second amplification primer, polymerase, nucleotides, co factors, etc.) may be present prior to formation of the semi-solid support (e.g., in a liquid prior to gel formation), or one or more of these reagents may be diffused into the semi-solid support after gel formation. The semi-solid substrate is generally formed as described above, e.g., on a glass slide.

[00188] In certain embodiments of the invention the gel can be solubilized (e.g., digested or depolymerized or dissolved) so that microparticles with attached clonal template populations can be conveniently recovered (e.g., by use of a magnet in the case of magnetic particles) following template synthesis. Gels that can be solubilized, digested, depolymerized, dissolved, etc., are referred to herein as "reversible". Conventional polyacrylamide polymerization involves the use of N-N' methylenebisacrylamide (BIS) as a crosslinking agent together with a suitable catalyst to initiate polymerization (e.g., N,N,N',N'- tetramethylethylenediamine (TEMED)). To produce a reversible gel an alternative cross- linking agent such as N-N' diallyltartardiamide (DATD) may be used. This compound is structurally similar to BIS but possesses cis-diol groups that can be cleaved by periodic acid, e.g., in a solution containing sodium periodate (Anker, H. S.: F.E.B.S. Lett., 7: 293, 1970). Thus DATD gels can be readily solubilized. Gels made using DATD as the crosslinker are highly transparent and bind well to glass Another crosslinking agent with DATD-like properties of forming reversible gels is

ethylene diacrylate (Choules, G. L. and Zimm, B. S.: Anal. Biochem., 13: 336-339, 1965). N,N'-bisacrylylcystamine (BAC) is another crosslinker that can be used to form a reversible polyacrylamide gel. Another crosslinking agent that can be used to form gels that dissolve in periodate is N,N'-(1,2-Dihydroxyethylene)bis- acrylamide (DHEBA). Any of a variety of other materials that form reversible semi-solid supports can also be used. For example, thermo-reversible polymers such as Pluronics (available from BASF) can be used. Pluronics are a family of poly (ethylene oxide)- poly(propylene oxide)-poly(ethylene oxide) (PEO-PPO-PEO) triblock copolymers (Nace, V. M., et al, Nonionic Surfactants, Marcel-Dekker, NY, 1996). These materials become semi-solid (gel) at elevated temperatues (e.g., temperatures greater than room temperature) and liquify upon cooling. Various methods can be used to chemically derivatize Pluronics, e.g., to facilitate attachment of primers thereto (see, e.g., Neff, J.A. et al., J. Biomed. Mater. Res., 40:511, 1998; Prud'homme, RK, et al., Langmuir, 12:4651, 1996). [00189] After solubilization, the microparticles can be collected and subjected to sequencing using repeated cycles of extension, ligation, and cleavage. Prior to sequencing, the microparticles may be arrayed in or on a second semi-solid support, e.g., at a higher density than that at which they were present in or on the first semi-solid support. The semisolid support is typically itself supported by a substantially planar and rigid substrate, e.g., a glass slide.

[00190] Thus two general approaches may be used to produce semi-solid supports having an array of microparticles bearing clonal template populations embedded in or on the semisolid support. The first approach involves performing amplification on microparticles that are not present in the semi-solid support (e.g., by emulsion-based PCR) and then immobilizing the microparticles in or on a semi-solid support. The second general approach involves immobilizing microparticles in or on a semi-solid support and then performing amplication. In either case, it may be desirable to employ procedures to reduce clumping of the microparticles and/or to align the microparticles substantially in a single focal plane. For example, when immobilizing particles in a polyacrylamide gel, the concentrations of monomer and crosslinker are selected so that the particles will sink to the bottom of the solution prior to complete polymerization, so that they

settle on an underlying planar substrate and are thus arranged in a single plane. In certain embodiments of the invention an object having a substantially planar surface, such as a cover slip, is placed on top of the liquid acrylamide (or other material capable of forming a semi-solid support) containing microparticles so that the acrylamide is trapped between two layers of a "sandwich" structure. The sandwich is then turned over, so that by the action of gravity the microparticles sink down and rest on the cover slip (or other object having a substantially planar surface). After polymerization, the cover slip is removed. The microparticles are thus embedded in substantially a single plane, close to the surface of the semi-solid support, (e.g., tangent to the surface).

[00191] Rather than immobilizing supports such as microparticles in a semi-solid matrix as described above, in certain embodiments of the invention microparticles are either covalently or noncovalently attached to a substantially planar, rigid substrate without use of a semi-solid support to immobilize them resulting in a "gel-free" or "gel-less" microparticle array. A variety of methods for attaching microparticles to substrates such as glass, plastic, quartz, silicon, etc., are known in the art. The substrate may or may not be coated (e.g., spin- coated) or functionalized with a material (e.g., any of a variety of polymers) or agent that facilitates attachment. The coating may be a thin film, self-assembled monolayer, etc. Eitherthe microparticles, a moiety attached to the microparticles, or oligonucleotides attached to the microparticles (e.g., the templates) can be attached to the substrate. In certain embodiments of the invention the substrate is not treated with a silanizing agent or if so treated, the treatment does not result in effective silanization, e.g., the silanization is not effective to permit formation of an array of microparticles immobilized by a polyacrylamide layer on a flat glass surface in a manner that is stable to subsequent manipulation and/or contact with fluids such as that which takes place during multiple cycles of ligation-based sequencing described herein, where "stable" in this context means that the gel typically remains affixed to the substrate during the manipulation and/or contact with fluids and does not significantly buckle, detach, or delaminate. The inventors have recognized that avoiding the use of a semisolid medium such as a gel to make the microparticle array may afford a number of advantages. For example, (i) diffusion of

reagents is more rapid, and removal of unwanted species such as unligated probes, enzymes, etc., is faster in the absence of the semi-solid medium; (ii) gels such as acrylamide may not remain stably affixed to the substrate in the absence of effective silanization; (iii) polymerization is sensitive to environmental features such as oxygen; thus eliminating the polymerization step removes a potential source of inconsistency in the array production process; (iv) absence of the semi-solid medium facilitates getting more of the microparticles into a single focal plane; (v) microparticles are more stably affixed in position when attached to the substrate than when embedded in a semisolid medium, particularly one in which polymerization is compromised. [00192] In general, any of a wide variety of methods known in the art can be used to modify nucleic acids such as oligonucleotide primers, probes, templates, etc., to facilitate the attachment of such nucleic acids to microparticles or to other supports or substrates. In addition, any of a wide variety of methods known in the art can be used to modify microparticles or others supports to facilitate the attachment of nucleic acids thereto, to facilitate the attachment of microparticles to supports or substrates, etc. Microspheres are available that have surface chemistries that facilitate the attachment of a desired functionality. Some examples of these surface chemistries include, but are not limited to, amino groups including aliphatic and aromatic amines, carboxylic acids, aldehydes, amides, chloromethyl groups, hydrazide, hydroxyl groups, sulfonates and sulfates. These groups may react with groups present in nucleic acids, or nucleic acids may be modified by attachment of a reactive group. In addition, a large number of stable bifunctional groups are well known in the art, including homobifunctional and heterobifunctional linkers. See, e.g., Pierce Chemical Technical Library, available at the Web site having URL www.piercenet.com (originally published in the 1994-95 Pierce Catalog) and G. T. Hermanson, Bioconjugate Techniques, Academic Press, Inc., 1996. See also U.S. Pat. No. 6,632,655.

[00193] In general, any pair of molecules that exhibit affinity for one another such that they form a binding pair may be used to attach microparticles or templates to a substrate. The first member of the binding pair is attached covalently or noncovalently to the substrate, and the second member of the binding pair is attached covalently or noncovalently to the microparticles or

templates. For purposes of description, the first member of the binding pair, i.e., the binding partner attached to the substrate, is referred to herein as BPI, and the second member of the binding pair, i.e., the binding partner attached to the microparticles or templates, is referred to as (BP2). The first binding partner (BPI) may be attached to the substrate via a linker. The second binding partner (BP2) may be attached to the microparticles or templates via a linker. For example, according to one approach, a slide or other suitable substrate is modified with an amine-reactive group (e.g., using a PEG linker containing an amine-reactive group). The amine-reactive group reacts under aqueous conditions (e.g. at pH 8.0) with an amine, e.g., a lysine in any protein, for example, streptavidin. Microparticles functionalized with a moiety bearing an amine will therefore become immobilized on the substrate. The moiety bearing an amine can be a protein or a suitably functionalized nucleic acid, e.g., a DNA template. Multiple moieties can be attached to a bead. For example, a bead may have proteins attached thereto that react with the NHS ester to attach the bead to the substrate and may also have DNA templates attached thereto, which can be sequenced after the bead is attached to the substrate. Suitably coated slides bearing a polymer tether having an amine-reactive NHS moiety on one end are commercially available, e.g., from Schott Nexterion, S chott North America, Inc., Elmsford, NY 10523). Alternately, coated slides (e.g., biotin-coated slides) are available from Accelr8 Technology Corporation, Denver, CO. Their OptiChem(TM) technology represents but one method for attaching microparticles to a substrate. See, e.g., U.S. Pat. No. 6,844,028. Alternately, microparticles may be attached to a substrate by functionalising polynucleotides on the bead with biotin by, e.g., the use of terminal transferase with biotin-dideoxyATP and/or biotin- deoxyATP, and then contacting them with a substrate such as a streptavidin-coated slide (available from, e.g., Accelr8 Technology Corporation, Denver, CO) (see U.S. Pat. No. 6,844,028) under conditions which promote formation of a biotin-streptavidin bond. In one embodiment, the streptavidin is attached to the substrate using a PEG linker. In one embodiment, the microparticle-bound polynucleotides are functionalized with biotin after their synthesis. In another embodiment biotin is incorporated into polynucleotides during synthesis by using biotinylated primers during amplification, e.g., when performing emulsion PCR. For

example, a first primer P1 is covalently or noncovalently attached to the microparticles. The second primer, P2, which is not bound to the microparticles, comprises a biotin moiety so that the resulting PCR product comprises biotin.

[00194] The invention therefore provides methods of capturing microparticles having nucleic acid templates attached thereto, and tethering them to the surface of a substrate, e.g., a substantially planar, rigid substrate such as a glass slide or the like. In an embodiment of particular interest, a population of microparticles having different clonal populations of templates attached thereto is produced (e.g., using emulsion PCR), wherein the templates comprise a biotin moiety. Biotin may be attached to the templates using standard methods following amplification. The microparticles are then contacted with a substantially planar, rigid substrate such as a glass slide having a biotin-binding moiety, e.g., a biotin-binding protein such as streptavidin attached thereto. The biotin on the template molecules binds to the biotin-binding moiety, thus attaching the microparticles to the substrate via a linkage comprising biotin and a biotin-binding protein. The attachment of the microparticles to the substrate may thus be indirect, wherein the template serves as a tether. In one embodiment, one end of the template molecules is attached to a biotin-binding moiety attached to the beads and the other end of the template molecules is attached to a biotin-binding moiety attached to the substrate.

[00195] In certain embodiments one terminus of a single-stranded template is attached to a microparticle and the other terminus of the single-stranded template is attached to the substrate. Thus in one embodiment both the 3' and 5' termini of a single-stranded template participate in linkages that serve to attach the microparticle to the substrate, wherein a first linkage is between the microparticle and the template and a second linkage is between the template and the substrate. The resulting structure is stable to heat and to other conditions that would tend to cause hybridized nucleic acids to dissociate.

[00196] As described in Example 16, it has been discovered that templates attached to streptavidin-coated microparticles can be biotinylated after their synthesis during emulsion PCR and that the resulting

biotinylated templates efficiently and robustly bind to streptavidin- coated substrates. In one embodiment, a biotin-streptavidin linkage is used at two stages in the method: (i) biotinylated primers are attached to streptavidin-coated microparticles prior to template amplification (e.g., prior to emulsion PCR) and (ii) after amplification, microparticle-bound templates biotinylated at their free end (i.e., the end not attached to the microparticle) are attached to a strepatividin-coated substrate, thereby anchoring the microparticles to the substrate as well. Optionally, following step (i), a population of microparticles that have been subjected to emulsion PCR (or other amplification method) can be enriched for microparticles that have undergone amplification. Prior to step (ii), and optionally following enrichment, the microparticles can be incubated with a biotinylated oligonucleotide in order to cover any part of the microparticle surface that has exposed streptavidin. These methods result in an array of microparticles stably attached to the surface of a substrate without the need for a semi-solid medium. In an embodiment of particular interest the substrate is a substantially planar, rigid substrate such as a glass slide or the like. While the biotin/streptavidin interaction is exemplified herein, it will be appreciated that streptavidin is only one of a number of proteins that bind to biotin, any of which could be used in the present invention. For example, avidin is an egg white protein that, like bacterial streptavidin, binds to biotin with high affinity and selectivity. NeutrAvidin is a derivative of avidin that has been processed to remove its carbohydrates. CaptAvidin is an avidin derivative that has reduced affinity for biotinylated molecules above pH 9. Consequently, biotinylated molecules can be allowed to bind at neutral pH and released at pH ~10. Neutravidin and CaptAvidin are described in The Handbook of Fluorescent Probes and Research Products, online edition

(http://probes.invitrogen.com/handbook/sections/0706.html; visited April 17, 2006) and are available from Invitrogen, Carlsbad, CA. Furthermore, the invention encompasses the use of any pair of molecules that display a specific and high affinity interaction. For example, the members of a specific binding pair could be an antibody and an antigen, a receptor and a ligand of the receptor (e.g., a small molecule or peptide), a metal and a metal binding agent (e.g., Ni+ and a 6X His tag), etc. The invention provides microparticles attached to substrates using any

of the methods described above and further provides arrays comprising microparticles attached to substrates, wherein the microparticles have different templates attached thereto.

[00197] In certain embodiments of the invention, formation of a gel-free microparticle array serves to separate microparticles that have multiple copies of a template attached thereto (e.g., at least thousands and typically millions of copies of a template attached thereto) from microparticles that do not that have multiple copies of a template attached thereto. In one embodiment the substrate has a first binding partner (BP1) attached thereto, wherein the template molecules attached to the microparticles comprise a second binding partner (BP2), and wherein BP1 and BP2 specifically bind to one another, i.e., they are members of a specific binding pair. When the gel-free microparticle array is formed as described above, only those microparticles that have templates comprising BP2 attached thereto will become attached to the substrate. In another embodiment the substrate has a first reactive moiety (R1) attached thereto, wherein the template molecules attached to the microparticles comprise a second reactive moiety (R2), and wherein R1 and R2 react with each other to form a covalent bond. When the gel-free microparticle array is formed as described above, only those microparticles that have templates comprising BP2 or R2 attached thereto will become attached to the substrate. After allowing binding or reaction to occur, the unattached microparticles can be removed, e.g., by gentle agitation and/or washing. The method is typically applied to a population of microparticles that includes microparticles having different clonal populations of templates attached thereto and also includes some microparticles that do not have multiple copies of a template attached thereto. For example, the method may be used to separate microparticles that have undergone template amplification (e.g., during emulsion PCR) from microparticles that have not undergone substantial template amplication. In one embodiment the method comprises steps of: (i) providing a substrate having a first member of a specific binding pair or a reactive moiety attached thereto; (ii) contacting the substrate with a population of microparticles at least some of which have multiple copies of a template comprising a second member of the specific binding pair or a reactive moiety attached thereto under conditions suitable for

binding to occur (either between the members of the binding pair or between the reactive moieties); and (iii) removing unbound microparticles. Specific binding partners that form strong non- covalent linkages (e.g., strepatividin and biotin) are of particular interest for achieving enrichment. In another embodiment, hybridization between complementary oligonucleotides is used. For example, in one embodiment an oligonucleotide selected to be complementary to a portion of the free PCR primer that is incorporated into a template during emulsion PCR (the free PCR primer being the one that is not attached to the microparticle) is attached to the substrate. Since the free PCR primer is only present on the microparticle if amplification was successful, only those microparticles that underwent successful template amplification become attached to the substrate. A ligase may be used to quality check the hybridization event and covalently link a biotinylated splint or primer to the 3' end of the templates on the beads. For example, the following sequence of steps can be performed, where "bead" represents a microparticle, P2 represents at least a portion of an amplification primer sequence, "ds" means "double-stranded" , "array" refers to the substrate to which the microparticles that have underone successful amplifcation can become attached via biotin. A microparticle having a double-stranded template attached thereto is provided. In the first step, the unbound template is removed, e.g., by raising the temperature. In the second step a double-stranded nucleic acid having a single-stranded extension is hybridized to the template. The double-stranded nucleic acid serves as a bridge or splint by which biotin can be stably linked to the template. The strand of the double-stranded nucleic acid not having the single- stranded extension has a biotin moiety attached at the opposite terminus to the single- stranded extension. In the third step, ligase is present. The double-stranded nucleic acid comprising biotin will be ligated to the template if successful hybridization has occurred, thus stably linking biotin to the template. In the fourth step, the strand of the splint that was not ligated to the template is released, e.g., by raising the temperature. Interaction of biotin with streptavidin bound to a substrate or support results in creation of an array of microparticles.

Bead -template -[P2]-ACGTTGGGGCTA

J, (1) melt unbound template from bead

Bead template -[P2]-ACGTTGGGGCTA

j, (2) hybridize a dsDNA-biotin splint to the template

P2

Bead -template -[P2]
-ACGTTGGGGCTApAAAAAAAAAAAAAAAAAAA
Biotin- array

TGCAACCCCGAT TTTTTTTTTTTTTTTTTTTTTTT

I (3) ligate dsDNA-biotin to P2

Bead -template -[P2]
-ACGTTGGGGCTAAAAAAAAAAAAAAAAAAA Biotin-
array

TGCAACCCCGAT TTTTTTTTTTTTTTTTTTTTTTT

j, (4) melt off non-bead bound DNA

Bead template -[P2]
-ACGTTGGGGCTAAAAAAAAAAAAAAAAAAA Biotin-
array [00198]

[00199] The method can be used to separate microparticles that have multiple templates attached thereto from microparticles that do not have multiple templates attached thereto or have substantially fewer templates attached thereto, wherein the templates are attached to the microparticles after amplification or synthesis. The microparticles to be separated may have been subjected to any type of condition in which amplification or synthesis of a microparticle-bound template occurs, or in which multiple copies of an amplified template may become attached to the microparticles. The amplification method may be PCR amplification, rolling circle amplification, or any other type of nucleic acid amplification. The method can be combined with and/or used in conjunction with any of the other methods and compositions of the invention. The contacting step typically occurs in a liquid medium. In certain embodiments of the invention during the contacting step liquid containing microparticles is allowed to flow across a substrate that has a specific binding pair or reactive moiety attached thereto. The substrate may, for example, be placed in a chamber

such as a flow cell having a fluid inlet and a fluid outlet. Microparticles may be flowed over the substrate until a desired density or number of microparticles attached to the substrate is reached. The change in density or number may be monitored over time (e.g., by imaging). In an embodiment of particular interest the method is used to separate microparticles that have undergone amplification during emulsion PCR from microparticles that have not undergone substantial template amplification during emulsion PCR. The method enriches for microparticles that undergone template amplification. The templates attached to the microparticles bound to the substrate can be subjected to a variety of further reactions or manipulationss. For example, they can be sequencing, e.g., using ligation-based sequencing as described herein, or using other sequencing methods such as FISSEQ, pyrosequencing, etc. For example, any of the inventive sequencing methods described herein can be performed on templates attached to microparticles that are attached to a substrate without using and/or in the absence of a semi-solid medium.

[00200] In any of the embodiments of the invention in which microparticles are attached to a substrate or semi-solid medium, the microparticles can subsequently be released and, optionally, removed (e.g., by washing). The appropriate method to release the microparticles will depend on the particular covalent or noncovalent linkage by which they are attached to the substrate or semi-solid medium. Any suitable method can be used provided it does not significantly damage the DNA template or result in its release from the substrate or semisolid medium. For example, in one embodiment the microparticles are attached to the substrate or semi-solid medium by a cleavable linker, e.g., one that contains a disulfide or ester linkage.

[00201] In certain embodiments of the invention microparticles are used to generate an array of clonal populations of templates that are stably attached to a semi-solid medium. In this method, microparticles having one or more template molecules attached thereto are incubated in the presence of a semi-solid medium located on a substrate, e.g., a polyacrylamide gel located on a substantially planar, rigid substrate, and the templates are hybrized to primers immobilized in and/or attached to the semi-solid medium. The primers are then extended (e.g., using a DNA polymerase),

resulting in synthesis of a complementary template attached to or immobilized in the semi-solid medium. The microparticles are released, e.g., by raising the stringency of the incubation (e.g., by raising the temperature) so that the two complementary template strands become separated. Alternate methods of releasing the microparticles, e.g., by cleaving the template attached thereto or otherwise detaching the microparticle from the template could also be used.

[00202] The process transfers a copy or "imprint" of the microparticle-bound template to the semi-solid medium. The efficiency of this process may be defined as the number of template molecules that are copied from a microparticle to the semi-solid medium divided by the number of template molecules attached to the microparticle. Based on geometrical and physical considerations, and without limiting the invention in any way, a microparticle of lum in diameter with about 150,000 template molecules 200 bp in size attached thereto would have a contact patch of about 500nm in diameter, as shown in Figure 40. The contact patch refers to the region of the semi-solid medium or substrate that would be in close enough proximity to a microparticle located on the surface of the medium or partically embedded therein so that templates complementary to those attached to the microparticle could be synthesized by extending primers located in or on the semi-solid medium or substrate. Specifically, 1 micron diameter beads have an area of 3.1x 10<6> nm<2>, so that 150,000 DNA molecules on a bead gives an average area of 20.9 nm<2> or average distance of 4.57 nm. The diameter of B-DNA is about 1.9 nm, and 200 bp B-DNA is 68 nm long. Therefore the contact patch of a 1 micron bead out to a separation of 68 nm is 252 nm in radius or 199,000 nm<2> in area. At 20.9 nm<2> per DNA molecule, the patch would be expected to contain as many as 9500 molecules, or about 13% of the number of molecules on the bottom half of the bead. [00203] Optionally, one or more rounds of amplification of the template that remains associated with the semi-solid medium is performed. In one embodiment, the amplification is rolling circle amplification (RCA; U.S. Pat. No. 5,854,033; 6,143,495). Prior to performing RCA, steps including (i) hybridization of a circularizable probe ("padlock probe") to two non-adjacent regions of the template, (ii) filling of the resulting gap using polymerase, and (iii) ligation of the ends, may be performed. It will be appreciated that

template molecules for use in RCA should include regions complementary to the circularizable probe in addition to a portion to be sequenced. [00204] Primer extension and optional amplification results in an array of "spots", or nucleic acid "colonies", attached to or immobilized in the semi-solid medium. The colonies are located at position corresponding to the locations at which the microparticles were deposited. Many or most of the colonies consist of a single clonal population of templates or, in certain embodiments of the invention, two or up to several clonal populations of templates (if the microparticle had two or more different templates attached thereto). A similar approach could be used to generate arrays of nucleic acid colonies directly on a substrate such as a glass slide without use of a semi-solid medium by attaching the primers to the substrate itself rather than to a semi-solid medium located on the substrate. [00205] Without wishing to be bound by any theory, forming an array of nucleic acid colonies using microparticles as described above provides a number of advantages. The microparticles can be subjected to template amplification and, optionally, enrichment, prior to their use to form the array, so that each nucleic acid spot arises from amplification of multiple copies of a template derived from a single microparticle rather than from amplification of a single template. Furthermore, the use of microparticles, which can be arranged on the surface of a semi-solid medium in close proximity to one another, provides for an efficient use of the surface of the semi-solid medium yet results in discrete spots that can be readily distinguished from one another during detection. The spots will typically be smaller in size than the microparticles, allowing them to be more clearly distinguished from one another. For example, if the DNA on a 1 micron diameter particle located within 250 nm of the contact point between the particle and a flat surface becomes attached to the flat surface and is copied, then after releasing the particle, the result would be a patch of DNA on the surface 500 nm in diameter. If two 1 micron beads are touching, then the centers of the DNA patches they leave behind will be 1 micron apart, leaving 500 nm spaces between the closest edges of the patches. With the capacity to pack millions of microparticles on the surface of a small substrate such as a glass slide, this process provides an efficient way to achieve high density arrays of template colonies that are readily imaged without interference from neighboring colonies and that contain a sufficient number of template molecules to enable easy and

reliable detection over multiple sequencing cycles.

[00206] The templates attached to the microparticles bound to the substrate can be subjected to a variety of further reactions or manipulationss. They can be sequencing, e.g., using ligation-based sequencing as described herein, or using other sequencing methods such as FISSEQ, pyrosequencing, etc. For example, any of the inventive sequencing methods described herein can be performed on templates that are present in nucleic acid colonies in a semi-solid medium, wherein the colonies are formed using a microparticle as described above.

[00207] Arrays of microparticles or nucleic acid colonies formed according to the methods described herein may be generally random. As used herein, the terms "randomly-patterned" or "random" refer to a non-ordered, non-Cartesian distribution (in other words, not arranged at pre-determined points or locations along the x- and y axes of a grid or at defined "clock positions", degrees or radii from the center of a radial pattern) of entities (features) over a support, that is not achieved through an intentional design (or program by which such a design may be achieved) or by placement of individual entities. Such a "randomly-patterned" or "random" array of entities may be achieved by dropping, spraying, plating, spreading, distributing, etc., a solution, emulsion, aerosol, vapor or dry preparation comprising a pool of entities onto or into a support and allowing them to settle onto or into the support without intervention in any manner to direct them to specific sites in or on the support. For example, entities may be suspended in a solution that contains precursors to a semi-solid support (e.g., acrylamide monomers). The solution is then distributed on a second support and the semisolid support forms on the second support. Entities are embedded in or on the semi-solid support. Of course non-random arrays can also be used. Close packing of microparticles may result in a regular grid-like array of microparticles or nucleic acid colonies synthesized therefrom. Generally the methods for forming arrays used herein are distinct from methods in which, for example, synthesis of a polynucleotide occurs by sequential application of individual nucleotide subunits at predefined locations on a substrate.

[00208] Figure 14B (top) shows a fluorescence image of a slide (1 inch by 3 inch) having a polyacrylamide gel

thereon. Beads (1 micron diameter) with a fluorescently labeled oligonucleotide hybridized to templates attached to the beads are immobilized in the gel. The image shows a bead surface density (i.e., number of beads per unit area of the substrate, within the region where the beads are located) sufficient to image approximately 280 million beads per slide. The surface density and imageable area are sufficient to image at least 500 million beads on a single slide. For example, Figure 14B (bottom) shows a schematic diagram of a slide with a Teflon(R) mask surrounding a clear area in which beads are to be embedded in a semi-solid support layer such as a polyacrylamide gel. The area of this mask is 864 mm<2>. With 500 million beads, the surface density is 578,000 beads per mm<2>. A close-packed hexagonal array of 1 micron beads gives 1,155,000 beads per mm<2>, so this embodiment results in an array having 52% of the theoretical maximum density. It will be appreciated that smaller and larger numbers of beads, and greater or lesser bead surface densities, can be used than in this particular embodiment.

[00209] Microparticles may be arrayed in or on a substantially planar semi-solid support, or on another support or substrate, at a variety of densities, which can be defined in a number of ways. For example, the density may be expressed in terms of the number of microparticles (e.g., spherical microparticles) per unit area of a substantially planar array. In certain embodiments of the invention the number of microparticles per unit area of a substantially planar array is at least 80% of the number of microparticles in a hexagonal array (by "hexagonal array" is meant a substantially planar array of microparticles in which every microparticle in the array contacts at least six other adjacent microparticles of equal area as described in U.S. Pat. No. 6,406,848). However, in other embodiments of the invention the microparticle density is lower, e.g., the number of microparticles per unit area of a substantially planar array is less than 80%, less than 70%, less than 60%, or less than 50% of the number of microparticles in a hexagonal array. Without wishing to be bound by any theory, it may be preferable to utilize lower densities such as these in order, for example, to allow adequate diffusion of reagents such as enzymes, primers, cofactors, etc., and to avoid a reagent partitioning effect that may occur if certain reagents have differential affinity for microparticles or become entrapped therein. Such an effect may result in

different reaction conditions at different positions on the array and may even prevent access to certain locations on the array by these reagents. These problems may be exacerbated when reactions are performed in a flow cell since the reagents move through the flow cell in a directional manner. In certain embodiments of the invention a mixing device, e.g., devices that achieve fluid mixing by mechanical or acoustical means, is included within the chamber of a flow cell. A number of suitable mixing devices are known in the art.

[00210] The inventive sequencing methods can be practiced using templates arranged in array formats of all types, including both random and nonrandom arrays, which can be arrays of microparticles or arrays of templates themselves. For example, supports with templates arrayed thereon are described in U.S. Pat. No. 5,641,658 and PCT Pub. No. WO0018957. Arrays may be located on a wide variety of substrates such as filters, membranes (e.g., nylon), metal surfaces, etc. Additional examples of array formats on which sequencing by repeated cycles of extension, ligation, and cleavage can be performed are arrays of beads located in wells at the terminal or distal end of individual optical fibers in a fiber optic bundle. See, e.g., bead arrays and "arrays of arrays" described in US publications and patents, e.g., 6,023,540; 6,429,027, 20040185483, 2002187515, PCT applications US98/05025, and PCT US98/09163, and PCT publication WO0039587. Beads with templates attached thereto can be arrayed as described therein. Amplification is preferably performed prior to formation of the array. Arrays formed on such substrates need not necessarily be substantially planar.

[00211] In other embodiments, PCR is performed on arrays that comprise oligonucleotides attached to a substrate or support, (see, e.g., U.S. Pat. Nos. 5,744,305; 5,800,992; 6,646,243 and related patents (Affymetrix); PCT publications WO2004029586; WO03065038; WO03040410 (Nimblegen)). In general, such oligonucleotides have a free 3' or 5' end. If desired, the end can be modified, e.g., by adding a phosphate group or an OH group to a 3' end if one is not already present. Template molecules comprising a region complementary to the oligonucleotide attached to the support or substrate are hybridized to the oligonucleotide, and PCR is performed in situ on the array, resulting in a clonal template population at each

location on the array. The oligonucleotide attached to the array may serve as one of the amplification primers. The templates are then sequenced using the ligation-based methods described herein. Sequencing can also be performed on templates in arrays such as those described in U.S. Pub. No. 20030068629.

[00212] Yet other methods for preparation of DNA arrays on surfaces can be used. For example, alkanethiols modified with terminal aldehyde groups can used to prepare a self- assembled monolayer (SAM) on a gold surface. The aldehyde groups of the monolayer may be reacted with amine-modified oligonucleotides or other amine -bearing biomolecules to form a Schiff base, which may then be reduced to a stable secondary amine by treatment with sodium cyanoborohydride (Peelen & Smith, Langmuir, 21(1):266-71, 2005). PCR amplification of templates can then be performed. Alternately, microparticles having clonal populations of templates attached thereto may be attached to surfaces by reacting an amine group on the microparticle or on templates or oligonucleotides attached to the particle, with such surfaces.

[00213] Still another method of obtaining microparticles with clonal template populations attached thereto is the "solid phase cloning" approach described in U.S. Pat. No. 5,604,097, which makes use of oligonucleotide tags for sorting polynucleotides onto microparticles such that only polynucleotides of the same sequence will be attached to any particular microparticle.

[00214] In certain embodiments of the invention sequencing by repeated cycles of extension, ligation, and cleavage is performed by diffusing sequencing reagents (e.g., extension probes, ligase, phosphatase, etc.) into a semi-solid support such as a gel having clonal populations of templates immobilized in or on the support such that each clonal population is localized to a spatially distinct region of the support. In certain embodiments the templates are attached directly to the semi-solid support as described above. However, in other embodiments the templates are immobilized on a second support such as a microparticle that is in turn immobilized in or on the semi-solid support, as also described above.

[00215] As described in Example 1, the inventors have shown that robust ligation and cleavage can be

performed on templates attached to beads that are immobilized in polyacrylamide gels. The invention thus provides a method of ligating a first polynucleotide to a second polynucleotide comprising steps of: (a) providing a first polynucleotide immobilized in or on a semi-solid support; (b) contacting the first polynucleotide with a second polynucleotide and a ligase; and (c) maintaining the first and second polynucleotides in the presence of ligase under suitable conditions for ligation. Suitable conditions include the provision of appropriate buffers, co factors, temperature, times, etc., for the particular ligase being used. In a preferred embodiment the semi-solid support is a gel such as an acrylamide gel. In a further preferred embodiment the first polynucleotide is immobilized in or on the semi-solid support as a result of attachment to a support such as a bead, which is itself immobilized in or on the semi-solid support, e.g., by being partly or completely embedded in the support matrix. Alternately, the first polynucleotide may be attached directly to the semi-solid support via a linkage such as an acrydite moiety. The linkage may be covalent or noncovalent (e.g., via a biotin-avidin interaction). U.S. Pat. No. 6,511,803 describes a variety of methods that may be used to a attach a nucleic acid molecule to a preferred support of the invention, i.e., a polyacrylamide gel.

[00216] The invention further provides a method of cleaving a polynucleotide comprising steps of: (a) providing a polynucleotide immobilized in or on a semi-solid support, wherein the polynucleotide comprises a scissile linkage; (b) contacting the polynucleotide with a cleavage agent; and (c) maintaining the polynucleotide in the presence of the cleavage agent under conditions suitable for cleavage. Suitable conditions include the provision of appropriate buffers, temperatures, times, etc., for the particular cleavage agent. In a preferred embodiment the semi-solid support is a gel such as an acrylamide gel. In a further preferred embodiment the polynucleotide is immobilized in the semi-solid support as a result of attachment to a support such as a bead, which is itself immobilized in the semi-solid support. Alternately, the polynucleotide may be attached directly to the semi-solid support via a linkage such as an acrydite moiety. The linkage may be covalent or noncovalent (e.g., via a biotin-avidin interaction).As will be appreciated, DNA templates prepared according to many of the methods described herein typically contain a region to be sequenced and also contain conserved

priming regions on either or both the 3' and 5' ends (PBRs). "Conserved" or "common" regions refers to sequences that are common to a plurality of templates that contain different regions to be sequenced, i.e., the templates, though differing in part of their sequence, also contain portions that are identical. Templates may also contain one or more conserved internal adapter sequence. Additionally, rolling circle amplification (RCA) of DNA templates not only generates additional copies of these conserved sequences but also introduces copies of yet another region of conserved sequence from the RCA probe. As a result, the portions of the library molecules to be sequenced (referred to as "target regions", "segment of interest", etc.) may represent less than half of the actual template nucleic acid. The invention encompasses the recognition that when single stranded, these known/common non-target regions can sequester sequencing probes and are potential sites for mispriming of the sequencing primers (e.g., the initializing oligonucleotides). The invention provides blocking oligonucleotides that are complementary to non-target sequences present in polynucleotide templates. As used herein, a "blocking oligonucleotide" is an oligonucleotide that stably hybridizes to a non-target sequence in a template, wherein the non-target sequence is common to a plurality of templates that comprise different target regions under conditions suitable for sequencing. The non-target region is distinct from the region to which an initializing oligonucleotide would bind. The invention further provides polynucleotide templates that have one or more blocking oligonucleotides hybridized thereto. [00217] In certain embodiments of the invention the templates are synthesized using emulsion PCR.

[00218] In embodiments of particular interest the DNA templates are members of a fragment library and contain forward and reverse adapters as shown in Figure 36B. A first blocking oligonucleotide is complementary to the forward adapter, and a second blocking oligonucleotide is complementary to the reverse adapter. In other embodiments the DNA templates are members of a paired-end library and contain forward and reverse adapters and also an internal adaptor, as shown in Figure 36A. A first blocking oligonucleotide is complementary to the forward adapter, a second blocking oligonucleotide is complementary to the reverse adapter, and a third blocking oligonucleotide is complementary to the

internal adapter. In other embodiments the templates are amplified using RCA and contain adapter regions and padlock regions as shown in Figure 36C and 37. Blocking oligonucleotides are complementary to the adapter and padlock regions as present in the templates. It will be appreciated that in RCA, the padlock probe is copied by polymerase to produce its complement. Therefore, to block the RCA complement present in the template, the same sequence as the padlock probe is to be used as a blocking oligonucleotide. The specific oligonucleotides shown in Figures 36 and 37, and their complements, are aspects of this invention, but it will be recognized that the sequence of the blocking oligonucleotides is selected to be complementary to the particular conserved sequences present in the template, which can vary. Additionally included in the invention are oligonucleotides that differ in sequence by not more than 1, 2, 3, 4, or 5 nucleotides from those depicted in Figure 36 or 37. [00219] Without limiting the invention in any way, the inventors the blocking oligonucleotides may be used to counteract the afore-mentioned problems or others that may arise due to the presence of many copies of these common sequences, e.g., by acting as a template complexity reduction tool, eliminating potential mispriming sites, and/or facilitating access of the extension oligonucleotides to the target region of the template. In certain embodiments of the invention the blocking oligonucleotides provide increased sequencing efficiency, e.g., a higher signal to noise ratio.

[00220] The blocking oligonucleotides are typically hybridized to the single-stranded template DNA prior to annealing of the sequencing primer, preventing these regions from subsequent hybridization with either the sequencing primer (e.g., the initializing oligonucleotide in ligation-based sequencing) or probes (e.g., extension probes in ligation- based sequencing). They would typically remain present during successive cycles of ligation, detection, (and cleavage, in those embodiments of the invention in which the extension oligonucleotide is cleaved). In certain embodiments of the invention the blocking oligonucleotides are not substrates for polymerases or ligases, e.g., they are not enzymatically extendable by typical polymerase or ligase enzymes. In one embodiment, the blocking oligonucleotides lack 3' hydroxyl groups and 5' phosphates. These groups may be absent or may be removed following synthesis, or

the 3' and/or 5' end of the oligonucleotide may be capped or blocked with a moiety that is not a substrate for extension or ligation. In certain embodiments of the invention a blocking oligonucleotide comprises a 3' terminal dideoxynucleoside. In certain embodiments of the invention a blocking oligonucleotide comprises a terminal 3' end dideoxycytosine (3'ddC).

In certain embodiments of the invention padlock probes for use with a paired tag library are designed to allow RCA of single tags individually (Tag #1 only, Tag #2 only) or across both tags (Tag #1 -internal-Tag #2) (Figure 37). [00221] The blocking oligonucleotides can be shorter than the conserved regions, i.e., they may be complementary to only a portion of a conserved region. The blocking oligonucleotides need not be perfectly complementary to the conserved regions, although this may be preferred. Typically they will be at least 80%, preferably at least 90% complementary to all or part of the conserved region. The size of a blocking oligonucleotide may vary depending on the length of the common sequences to be blocked. Typical lengths are between 10 and 50 nucleotides. Two or more blocking oligonucleotides, each complementary to a portion of a conserved region to be blocked, may be used instead of a single longer oligonucleotide.

[00222] The blocking oligonucleotides may find particular use in ligation-based sequencing as described herein. Thus any of the methods described herein may include a step of contacting a template polynucleotide with one or more blocking oligonucleotides prior to contacting the template with an initializing oligonucleotide, prior to forming or providing a probe-template duplex, and/or prior to forming an extended duplex. However, the blocking oligonucleotides may also be used when performing other sequencing methods such as FISSEQ, pyrosequencing, etc.

[00223] D. Sequencing With Re-initialization Using Different Initializing Oligonucleotides

[00224] In a preferred embodiment of the instant invention, the extended strand generated by extending a first initializing oligonucletide is removed from the template following a sufficient number of cycles and a second initializing oligonucleotide is annealed to the binding region, followed by cycles of extension, ligation, and detection. The process is repeated with any number

of different initializing oligonucleotides. In embodiments in which the extension probes are cleaved, preferably the number of different initializing oligonucleotides used (and thus the number of reactions) equals the length of the portion of the extension probe that remains hybrized to the template following release of the distal portion of the probe. Thus according to this embodiment sequence information (e.g., the order and identity of each nucleotide) can be obtained from the templates that are attached to a single support while still reading deep into the sequence using substantially fewer cycles than would be required if successive nucleotides were identified in each cycle. [00225] Embodiments in which the initializing oligonucleotides are bound sequentially to the same template have certain advantages over an approach that requires dividing the template into multiple aliquots, such as the methods taught by Macevicz. For example, applying the initializing oligonucleotides to the same template avoids the need to keep track of, and later, combine data acquired from multiple aliquots. In embodiments in which the supports are arrayed in a random fashion such that the position of individual supports is not predetermined, it would be difficult or impossible to reliably combine partial sequence information from multiple supports each of which had templates of the same sequence attached thereto.

[00226] E. Identification of Multiple Nucleotides in Each Cycle on a Single Template [00227] Macevicz teaches identification of single nucleotides in the template in each cycle of extension, ligation, and detection. However, the inventors have recognized that the methods may be modified to allow identification of multiple nucleotides in the template in each cycle. In this case the extension probes are labeled so that the identily of two or more, preferably contiguous, nucleotides abutting the extended duplex can be determined from the label. In other words, the sequence determining portion of the extension probes is more than a single nucleotide and typically comprises the proximal nucleotide, the immediately adjacent nucleotide, and possibly one or more additional, preferably contiguous nucleotides, all of which hybridize specifically to the template. For example, rather than using 4 labels to identify the bases A, G, C, and T, 16 distinguishably labeled probes or probe combinations are used to identify the 16 possible dinucleotides AA, AG, AC, AT, GA, GG, GC, GT, CA, CG, CC, CT, TA,

TG, TC, and TT. The sequence determining portion of each distinguishably labeled extension probe is complementary to one of these dinucleotides. Similar methods utilizing more labels allow identification of longer nucleotide sequences in each cycle. [00228] F. Labels

[00229] The term "label" is used herein in a broad sense to denote any detectable moiety or plurality of detectable moieties attached to or associated with a probe, by which probes of different species (e.g., probes with different terminal nucleotides) may be distinguished from one another. Thus there need not be a one to one correspondence between a label and a specific detectable moiety. For example, multiple detectable moieties can be attached to a single probe, resulting in a combined signal that allows the probe to be distinguished from probes having a different detectable moiety or set of detectable moieties attached thereto. For example, combinations of detectable moieties can be used in accordance with a labeling scheme referred to as "Combinatorial Multicolor Coding", which is described in U.S. Pat. No. 6,632,609 and in Speicher, et al, Nature Genetics, 12:368-375, 1996. [00230] The probes of the invention can be labeled in a variety of ways, including the direct or indirect attachment of fluorescent or chemiluminescent moieties, colorimetric moieties, enzymatic moieties that generate a detectable signal when contacted with a substrate, and the like. Macevicz teaches that the probes may be labeled with fluorescent dyes, e.g. as disclosed by Menchen et al, U.S. Pat. No. 5,188,934; Begot et al PCT application PCT/US90105565. The terms "fluorescent dye", and "fluorophore" as used herein refer to moieties that absorb light energy at a defined excitation wavelength and emit light energy at a different wavelength. Preferably the labels selected for use with a given mixture of probes are spectrally resolvable. As used herein, "spectrally resolvable" means that the labels may be distinguished on the basis of their spectral characteristics, particularly fluorescence emission wavelength, under conditions of operation. For example, the identity of the one or more terminal nucleotides may be correlated to a distinct wavelength of maximum light emission intensity, or perhaps a ratio of intensities at different wavelengths. The spectral characteristic(s) of a label that is/are used to detect and identify a label is referred to as a "color" herein. It will be appreciated that a label is frequently identified on the basis of a specific

spectral characteristic, e.g., the frequency of maximum emission intensity in the case of labels that consist of a single detectable moiety, or the frequencies of emission peaks in the case of labels that consist of multiple detectable moieties. [00231] Preferably, four probes are provided that allow a one-to-one correspondence between each of four spectrally resolvable fluorescent dyes and the four possible terminal nucleotides of the probes. Sets of spectrally resolvable dyes are disclosed in U.S. Pat. Nos. 4,855,225 and 5,188,934; International application PCT/US90/05565; and Lee et al, Nucleic Acids Researchs, 20: 2471-2483 (1992). In certain embodiments a set consisting of FITC, HEX(TM), Texas Red, and Cy5 is preferred. Numerous suitable fluorescent dyes are commercially available, e.g., from Molecular Probes, Inc., Eugene OR. Specific examples of fluorescent dyes include, but are not limited to: Alexa Fluor dyes (Alexa Fluor 350, Alexa Fluor 488, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, Alexa Fluor 633, Alexa Fluor 660 and Alexa Fluor 680), AMCA, AMCA-S, BODIPY dyes (BODIPY FL, BODIPY R6G, BODIPY TMR, BODIPY TR, BODIPY 530/550, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY 630/650, BODIPY 650/665), CAL dyes, Carboxyrhodamine 6G, carboxy-X-rhodamine (ROX), Cascade Blue, Cascade Yellow, Cyanine dyes (Cy3, Cy5, Cy3.5, Cy5.5), Dansyl, Dapoxyl, Dialkylaminocoumarin, 4',5'-Dichloro-2',7'-dimethoxy-fluorescein, DM-NERF, Eosin, Erythrosin, Fluorescein, FAM, Hydroxycoumarin, IRDyes (IRD40, IRD 700, IRD 800), JOE, Lissamine rhodamine B, Marina Blue, Methoxycoumarin, Naphtho fluorescein, Oregon Green 488, Oregon Green 500, Oregon Green 514, Oyster dyes, Pacific Blue, PyMPO, Pyrene, Rhodamine 6G, Rhodamine Green, Rhodamine Red, Rhodol Green, 2',4',5',7'-Tetra- bromosulfone-fluorescein, Tetramethyl-rhodamine (TMR), Carboxytetramethylrhodamine (TAMRA), Texas Red, Texas Red-X. See The Handbook of Fluorescent Probes and Research Products, 9th ed., Molecular Probes, Inc., for further description. [00232] Rather than being directly detectable themselves, some fluorescent groups transfer energy to another group in the process of nonradiative fluorescent resonance energy transfer (FRET), and the second group produces the detected signal. The use of quenchers, i.e., is also within the scope of the invention. The term "quencher" refers to a moiety that is capable of absorbing the energy of an excited fluorescent label when located in close proximity

and of dissipating that energy without the emission of visible light. Examples of quenchers include, but are not limited to DABCYL (4-(4'-dimethylaminophenylazo) benzoic acid) succinimidyl ester, diarylrhodamine carboxylic acid, succinimidyl ester (QSY-7), and 4', 5'-dinitrofluorescein carboxylic acid, succinimidyl ester (QSY-33) (all available from Molecular Probes), quencherl (Ql; available from Epoch), or "Black hole quenchers" BHQ-I, BHQ-2, and BHQ-3 (available form BioSearch, Inc.).

[00233] In addition to the various detectable moieties mentioned above, the present invention also comprehends use of spectrally resolvable quantum dots, metal nanoparticles or nanoclusters, etc., which may either be directly attached to an oligonucleotide probe or may be embedded in or associated with a polymeric matrix which is then attached to the probe. As mentioned above, detectable moieties need not themselves be directly detectable. For example, they may act on a substrate which is detected, or they may require modification to become detectable.

[00234] As described above, in certain embodiments of the invention a label consists of a plurality of detectable moieties. The combined signal from these detectable moieties produces a color that is used to identify the probe. For example, a "purple" probe of a particular sequence could be constructed by attaching "blue" and "red" detectable moieties thereto. Alternatively, a distinct color can be generated by combining two species of probe having the same sequence but labeled with different detectable moieties to produce a mixed probe. Thus a "purple" probe of a particular sequence can be produced by constructing two species of probe having that sequence. "Red" detectable moieties are attached to the first species, and "blue" detectable moieties are attached to the second species. Aliquots of these two species are mixed. Various shades of purple can be produced by mixing aliqots in different ratios. This approach offers a number of advantages. Firstly, it allows the production of multiple distinguishable probes using a smaller number of detectable moieties. Secondly, using a mixed probe can provide a degree of redundancy that may help reduce bias that may result from interactions between particular detectable moieties and particular nucleotides.

[00235] In certain embodiments of the invention a

detectable moiety is attached to a nucleotide in an oligonucleotide extension probe by a cleavable linkage, which allows removal of the detectable moiety following ligation and detection. Any of a variety of different cleavable linkages may be used. As used herein, in the context of a detectable moiety and a nucleotide in an oligonucleotide probe, the term "cleavable linkage" refers to a chemical moiety that joins a detectable moiety to a nucleotide, and that can be cleaved to remove the detectable moiety from the nucleotide when desired, essentially without altering the nucleotide or the nucleic acid molecule it is attached to. Cleavage may be accomplished, for example, by acid or base treatment, or by oxidation or reduction of the linkage, or by light treatment (photocleavage), depending upon the nature of the linkage. Examples of cleavable linkages and cleavage agents are described in Shimkus et al., 1985, Proc. Natl. Acad. Sci. USA 82: 2593-2597; Soukup et al., 1995, Bioconjug. Chem. 6: 135-138; Shimikus et al., 1986, DNA 5: 247-255; and Herman and Fenn, 1990, Meth. Enzymol. 184: 584-588. More generally, "cleavable linkage" refers to a moiety that can be used to link two molecules or entities together and can be readily cleaved, thereby allowing separation of the molecules or entities, without substantially altering their structure, e.g, under conditions consistent with stability of the molecules or entities.

[00236] For example, as described in U.S. Pat. No. 6,511,803, a disulfide linkage can be reduced and thereby cleaved using thiol compound reducing agents such as dithiothreitol (DTT). Fluorophores are available with a sulfhydryl (SH) group available for conjugation (e.g., Cyanine 5 or Cyanine 3 fluorophores with SH groups; New England Nuclear -DuPont), as are nucleotides with a reactive aryl amino group (e.g., dCTP). A reactive pyridyldithiol will react with a sulflhydryl group to give a sulfhydryl bond that is cleavable with reducing agents such as dithiothreitol. An NHS-ester heterobifunctional crosslinker (Pierce) can be used to link a deoxynucleotide comprising a reactive aryl amino group to a pyridyldithiol group, which is in turn reactive with the SH on a fluorophore, to yield a disulfide bonded, cleavable nucleotide-fluorophore complex useful in the methods of the invention. Alternatively, a cis-glycol linkage between a nucleotide and a fluorophore can be cleaved by periodate. A variety of cleavable linkages are described in US Pat. Nos. 6,664,079, and 6,632,655, US Published

Application 20030104437, WO 04/18497 and WO 03/48387.[00237] In other embodiments of the invention a detectable moiety that can be rendered nondetectable by exposure to electromagnetic energy such as light (photobleaching) is used. [00238] In those embodiments of the invention that employ extension probes having a label that is attached to the probe by a cleavable linkage, or having a label that can be photobleached, a the sequencing methods will typically include a step of cleavage or photobleaching in one or more cycles after ligation and label detection have been performed. As mentioned above, cleavage of the scissile linkage present in the oligonucleotide extension probes may not proceed to completion (i.e., less than 100% of the newly ligated probes may be cleaved in the cycle in which they were ligated). Since such probes generally comprise a non-extendable terminus, or are capped, they will not contribute to successive cycles. However, failure to cleave the probe means that the label remains associated with the template molecule to which the probe ligated, which contributes background signal (i.e., background fluorescence) that can increase the noise in subsequent cycles. Incorporating a step of cleavage or photobleaching to remove the label or render it undetectable reduces this background and improves the signal to noise ratio. Cleavage or photobleaching can be performed as often as every cycle, or less frequently, such as every other, every third, or every fifth or more cycles. In certain embodiments of the invention it is not necessary to actually add any additional steps to achieve cleavage of the cleavable linker. For example, a cleavage agent such as DTT may already be present in a wash buffer that may be used to remove unligated extension probes. [00239] G. Preferred Scissile Linkages

[00240] The inventors have discovered that extension probes having at least one phosphorothiolate linkage are particularly useful in the practice of methods for sequencing by successive cycles of extension, ligation, detection, and cleavage. In such linkages one of the bridging oxygen atoms of a phosphodiester bond is replaced by a sulfur atom. The phosphorothiolate linkage can be either a 5'-S- phosphorothiolate linkage (3'-O-P-S-5') as shown in Fig. 4A or a 3'-S-phosphorothiolate linkage (3'-S-P-O-5') as shown in Fig. 4B. It is to be understood that the phosphorus atom in linkages represented as 3'-O-P-S-5' or 3'-S- P-O-5' may be attached to two non-bridging oxygen atoms as

shown in Figs. 4A and 4B (as in typical phosphodiester bonds). Alternately, the phosphorus atom could be attached to any of a variety of other atoms or groups, e.g., S, CH3, BH3, etc. Thus one aspect of the invention is labeled olignucleotide probes comprising phosphorothiolate linkages. While the probes find particular use in the sequencing methods described herein, they may also be used for a variety of other purposes. In particular, the invention provides (i) an oligonucleotide of the form 5'-O-P-O-X-O-P-S-(N) kNB*-3'; and (ii) an oligonucleotide of the form 5'-NB*(N) k -S- P-O-X-3'. In each of these probes N represents any nucleotide, NB represents a moiety that is not extendable by ligase, * represents a detectable moiety, X represents a nucleotide, and k is between 1 and 100. In certain embodiments k is between 1 and 50, between 1 and 30, between 1 and 20, e.g., between 4 and 10, with the proviso that a detectable moiety may be present on any nucleotide of (N)k instead of, or in addition to, NB. The terminal nucleotides in any of these probes may or may not include a phosphate group or a hydroxyl group. Furthemore, it will be appreciated that the phosphorus atoms will generally be attached to two additional (non-bridging) oxygen atoms in preferred embodiments. [00241] Methods for synthesizing oligonucleotides containing 5'-S- phosphorothiolate or 3'-S-phosphorothiolate linkages are known in the art, and certain of these methods are amenable to automated solid phase oligonucleotide synthesis. Synthesis procedures are described, for example, in Cook, AF, J. Am. Chem. Soc, 92:190-195, 1970; Chladek, S. et al, J. Am. Chem. Soc, 94:2079-2084, 1972; Rybakov, VN, et al, Nucleic Acids Res., 9:189-201, 1981; Cosstick, R. and Vyle, JS, J. Chem. Soc. CHem. Commun., 992-992, 1988; Mag, M., et al., Nucleic Acids Res., 19(7); 1437-1441, 1991; Xu, Y, and Kool, ET, Nucleic Acids Res., 26(13): 3159-3164, 1998; Cosstick, R. and Vyle, JS, Tetrahedron Lett., 30:4693-4696, 1989; Cosstick, R. and Vyle, JS, Nucleic Acids Res., 18:829-835, 1990; Sun, SG and Piccirilli, JA, Nucl. NucL, 16:1543-1545, 1997; Sun SG, et al., RNA, 3:1352-1363, 1997; Vyle, JS, et al., Tetrahedron Lett., 33:3017-3020, 1992; Li, X., et al., J. Chem. Soc. Perkin Trans., 1 :2123-22129, 1994; Liu, XH and Reese, CB, Tetrahedron Lett., 37: 925-928, 1996; Weinstein, LB, et al., J. Am. Chem. Soc, 118:10341-10350, 1996; and Sabbagh, G., et al., Nucleic Acids Res., 32(2):495-501, 2004. In addition, the present inventors have developed new synthesis methods. For example, Figure 7 shows a

synthesis scheme for a 3'- phosphoroamidite of dA. A similar scheme may be used for synthesis of a 3'-phosphoroamidite of dG. These phosphoroamidites may be used to synthesize oligonucleotides containing 3'-S-phosphorothiolate linkages associated with purine nucleosides, e.g., using an automated DNA synthesizer.

[00242] Phosphorothiolate linkages can be cleaved using a variety of metal-containing agents. The metal can be, for example, Ag, Hg, Cu, Mn, Zn or Cd. Preferably the agent is a water-soiuble salt that provides Ag<+>, Hg<++>, Cu <+>, Mn<++>, Zn<+> or Cd<+> anions (salts that provide ions of other oxidation states can also be used). I2 can also be used. Silver-containing salts such as silver nitrate (AgNO3), or other salts that provide Ag<+> ions, are particularly preferred. Suitable conditions include, for example, 50 mM AgNO3 at about 22-37[deg.]C for 10 minutes or more, e.g., 30 minutes. Preferably the pH is between 4.0 and 10.0, more preferably between 5.0 and 9.0, e.g., between about 6.0 and 8.0, e.g., about 7.0. See, e.g., Mag, M., et al, Nucleic Acids Res., 19(7):1437-1441, 1991. An exemplary protocol is provided in Example 1.

[00243] Sequencing in the 5 ' ->3 ' direction may be performed using extension probes containing a 3'-O-P-S-5' linkage. Fig. 5 A shows a single cycle of hybridization, ligation, and cleavage using an extension probe of the form 5 '-O-P-O-X-O-P-S-NNNNNB*-3 ' where N represents any nucleotide, NB represents a moiety that is not extendable by ligase (e.g., NB is a nucleotide that lacks a 3 ' hydroxyl group or has an attached blocking moiety), * represents a detectable moiety, and X represents a nucleotide whose identity corresponds to the detectable moiety. Alternately, any of a large number of blocking moieties can be attached to the 3' terminal nucleotide to prevent multiple ligations. For example, attaching a bulky group to the sugar portion of the nucleotide, e.g., at the 2' or 3' position, will prevent ligation. A fluorescent label may serve as an appropriate bulky group.

[00244] A template containing binding region 40 and polynucleotide region 50 of unknown sequence is attached to a support, e.g., a bead. In a preferred embodiment, as shown in Fig. 5 A, the binding region is located at the opposite end of the template from the point of attachment to the support. An initializing oligonucleotide 30 with an extendable terminus (in this

case a free 3' OH group) is annealed to binding region 40. Extension probe 60 is hybridized to the template in polynucleotide region 50. Nucleotide X forms a complementary base pair with unknown nucleotide Y in the template. Extension probe 60 is ligated to the initializing oligonucleotide (e.g., using T4 ligase). Following ligation, the label attached to extension probe 60 is detected (not shown). The label corresponds to the identity of nucleotide X. Thus nucleotide Y is identified as the nucleotide complementary to nucleotide X. Extension probe 60 is then cleaved at the phosphorothiolate linkage (e.g., using AgNO3 or another salt that provides Ag<+> ions), resulting in an extended duplex. Cleavage leaves a phosphate group at the 3' end of the extended duplex. Phosphatase treatment is used to generate an extendable probe terminus on the extended duplex. The process is repeated for a desired number of cycles.

[00245] In a preferred embodiment sequencing is performed in the 3 '-^5' direction using extension probes containing a 3'-S-P-O-5' linkage. Fig. 5B shows a single cycle of hybridization, ligation, and cleavage using an extension probe of the form 5 '-NB*-NNNN-S- P-O-X-3' where N represents any nucleotide, NB represents a moiety that is not extendable by ligase (e.g., NB is a nucleotide that lacks a 5' phosphate group or has an attached blocking moiety), * represents a detectable moiety, and X represents a nucleotide whose identity corresponds to the detectable moiety.

[00246] A template containing binding region 40 and polynucleotide region 50 of unknown sequence is attached to a support, e.g., a bead. In a preferred embodiment, as shown in Fig. 5B, the binding region is located at the opposite end of the template from the point of attachment to the support. An initializing oligonucleotide 30 with an extendable terminus (in this case a free 5' phosphate group) is annealed to binding region 40. Extension probe 60 is hybridized to the template in polynucleotide region 50. Nucleotide X forms a complementary base pair with unknown nucleotide Y in the template. Extension probe 60 is ligated to the initializing oligonucleotide (e.g., using T4 ligase). Following ligation, the label attached to extension probe 60 is detected (not shown). The label corresponds to the identity of nucleotide X. Thus nucleotide Y is identified as the nucleotide complementary to nucleotide X. Extension probe 60 is

then cleaved at the phosphorothiolate linkage (e.g., using AgNO3 or another salt that provides Ag<+> ions), resulting in an extended duplex. Cleavage leaves an extendable monophosphate group at the 5' terminus of the extended duplex and it is therefore unnecessary to perform an additional step to generate an extendable terminus. The process is repeated for a desired number of cycles.

[00247] It will be appreciated that a number of variations of this scheme can be used. For example, the probe may be shorter or longer than 6 nucleotides; the label need not be on the 3' terminal nucleotide; the P-S linkage can be located between any two adjacent nucleotides, etc. In the embodiments described above, successive cycles of extension, ligation, detection, and cleavage, result in identification of adjacently located nucleotides. However, by placing the P-S linkage closer to the distal end of the extension probe (i.e., the end opposite to that at which ligation occurs), the nucleotides that are sequentially identified will be spaced at intervals along the template, as described above and shown in Figs. 1 and 6. [00248] Fig. 6A-6F is a more detailed diagrammatic illustration of several sequencing reactions performed sequentially on a single template. Sequencing is performed in the 3 ' ->5 ' direction using extension probes containing 3'-S-P-O-5' linkages. Each sequencing reaction comprises multiple cycles of extension, ligation, detection, and cleavage. The reactions utilize initializing oligonucleotides that bind to different portions of the template. The extension probes are 8 nucleotides in length and contain phosphorothiolate linkages located between the 6th and 7th nucleotides counting from the 3' end of the probe. Nucleotides 2-6 serve as a spacer such that each reaction allows the identification of a plurality of nucleotides spaced at intervals along the template. By performing multiple reactions in series and appropriately combining the partial sequence information obtained from each reaction, the complete sequence of a portion of the template is determined.

[00249] Fig. 6A shows initialization using a first initializing oligonucleotide (referred to as a primer in Figs. 6A-6F) that is hybridized to an adapter sequence (referred to above as a binding region) in the template to provide an extendable duplex. Figs. 6B-6D show several cycles of nucleotide identification in which every 6th base of the template is read. In Fig. 6B, a first

extension probe having a 3 ' terminal nucleotide complementary to the first unknown nucleotide in the template sequence binds to the template and is ligated to the extendable terminus of the primer. The label attached to the extension probe identifies the probe as having an A as the 3 ' terminal nucleotide and thus identifies the first unknown nucleotide in the template sequence as A. Fig. 6C shows cleavage of the extension oligonucleotide at the phosphorothiolate linkage with AgNO3 and release of a portion of the extension probe to which a label is attached. Fig. 6D shows additional cycles of extension, ligation, and cleavage. Since the probes contain a spacer 5 nucleotides in length, the sequencing reaction identifies every 6th nucleotide in the template.

[00250] Following a desired number of cycles the extended strand, including the first initializing oligonoucleotide, is removed and a second initializing oligonucleotide that binds to a different portion of the binding region from that at which the first initializing oligonucleotide bound, is hybridized to the template. Fig. 6E shows a second sequencing reaction in which initialization is performed with a second initializing oligonucleotide, followed by several cycles of nucleotide identification. Fig. 6F shows initialization using a third initializing oligonucleotide followed by several cycles of nucleotide identification. Extension from the second initializing oligonucleotide allows identification of every 6th base in a different "frame" from the nucleotides identified in the first sequencing reaction. [00251] Although extension probes comprising phosphorothiolate linkages are preferred in certain embodiments of the invention, a variety of other scissile linkages may be advantageously employed. For example, a large number of variations on the O-P-0 linkage found in naturally occuring nucleic acids are known (see, e.g., Micklefield, J. Curr. Med. Chem., 8: 1157-1179, 2001). Any structures described therein that contain a P-O bond can be modified to contain a scissile P-S bond. For example, an NH-P-O bond can be changed to an NH-P-S bond.

[00252] In some embodiments of the invention the extension probes comprise a trigger residue that renders the nucleic acid susceptible to cleavage by a cleavage agent or combination thereof, optionally following modification of the trigger residue by a modifying agent. In particular, the inventors have

discovered that enzymes involved in DNA repair areadvantageous cleavage reagents for use in the practice of methods for sequencing by successive cycles of extension, ligation, detection, and cleavage. In general, the presence of a trigger residue such as a damaged base or abasic residue in an extension probe may render the probe susceptible to cleavage by one or more DNA repair enzymes, optionally following modification by a DNA glycosylase. Thus extension probes comprising linkages that are substrates for cleavage by enzymes involved in DNA repair such as AP endonucleases are of use in the invention. Extension probes containing residues that are substrates for modification by enzymes involved in DNA repair, such as DNA glycosylases, wherein the modification renders the probe susceptible to cleavage by an AP endonuclease, are also of particular use in the invention. In some embodiments the extension probe comprises an abasic residue, i.e., it lacks a purine or pyrimidine base. The linkage between the abasic residue and an adjacent nucleoside is susceptible to cleavage by an AP endonuclease and is therefore a scissile linkage. In certain embodiments of the invention the abasic residue comprises 2' deoxyribose. In some embodiments the extension probe comprises a damaged base. The damaged base is a substrate for an enzyme that removes damaged bases, such as a DNA glycosylase. Following removal of the damaged base, the linkage between the resulting abasic residue and an adjacent nucleoside is susceptible to cleavage by an AP endonuclease and is therefore considered a scissile linkage in accordance with the invention. [00253] Many different AP endonucleases are of use as cleavage reagents in the present invention. Two major classes of AP endonuclease have been distinguished on the basis of the mechanism by which they cleave linkages adjacent to abasic residues. Class I AP endonucleases, such as endonuclease III (Endo III) and endonuclease VIII (Endo VIII) of E. coli and the human homo logs hNTHI, NEIL1, NEIL2, and NEIL3, are AP lyases that cleave DNA on the 3' side of the AP residue, resulting in a 5' portion that has a 3' terminal phosphate and a 3' portion that bears a 5' terminal phosphate. Class II AP endonucleases such as endonuclease IV (Endo IV) and exonuclease III (Exo III) of E. coli cleave the DNA 5' of the AP site, which produces a 3' OH and 5' deoxyribose phosphate moiety at the termini of the resulting fragments. See, e.g., Doublie, S., et al, Proc. Natl. Acad. Sci. 101(28), 10284-10289, 2004; Haltiwanger, B.M., et

al, Biochem J., 345, 85-89, 2000; Levin, J. and Demple, B., Nucl. Acids. Res, 18(17), 1990, and references in all of the foregoing for further discussion of various Class I and Class II AP endonucleases and conditions under which they remove damaged bases from DNA and/or cleave DNA containing an abasic residue. One of ordinary skill in the art will appreciate that a variety of homologs of these enzymes exist in other organisms (e.g., yeast) and are of use in the present invention.
[00254] Certain enzymes are bifunctional in that they possess both glycosylase activity that removes a damaged base to generate an AP residue and also display a lyase activity that cleaves the phosphodiester backbone 3' to the AP site generated by the glycosylase activity. Thus these dual activity enzymes are both AP endonucleases and DNA glycosylases. For example, Endo VIII acts as both an N-glycosylase and an AP-lyase. The N-glycosylase activity releases damaged pyrimidines from double-stranded DNA, generating an apurinic (AP site). The AP-lyase activity cleaves 3 ' and 5 ' to the AP site leaving a 5 ' phosphate and a 3 ' phosphate. Damaged bases recognized and removed by Endonuclease VIII include urea, 5, 6-dihydroxythymine, thymine glycol, 5-hydroxy-5-methylhydanton, uracil glycol, 6- hydroxy-5, 6-dihydrothymine and methyltartronylurea. See, e.g., Dizdaroglu, M., et al, Biochemistry, 32,12105-12111, 1993 and Hatahet, Z. et al., J. Biol. Ozem.,269,18814-18820, 1994; Jiang, D., et al., J. Biol. Chem., 272(51), 32220-32229, 1997; Jiang, D., et al., J. Bact, 179(11), 3773-3782, 1997.

[00255] Fpg (formamidopyrimidine [fapy]-DNA glycosylase) (also known as 8- oxoguanine DNA glycosylase) also acts both as a N-glycosylase and an AP-lyase. The N- glycosylase activity releases damaged purines from double stranded DNA, generating an apurinic (AP site). The AP-lyase activity cleaves both 3 ' and 5 ' to the AP site thereby removing the AP site and leaving a 1 base gap. Some of the damaged bases recognized and removed by Fpg include 7, 8-dihydro-8-oxoguanine (8-oxoguanine), 8-oxoadenine, fapy-guanine, methy-fapy-guanine, fapy-adenine, aflatoxin Bl-fapy-guanine, 5-hydroxy-cytosine and 5-hydroxy-uracil. See, e.g., Tchou, J. et al. J. Biol. Chem., 269, 15318-15324, 1994; Hatahet, Z. et al. J. Biol. Chem., 269, 18814-18820, 1994; Boiteux, S., et al, EMBOJ., 5, 3177-3183, 1987; Jiang, D., et al., J. Biol. Chem., 272 (51), 32220-32229, 1997; Jiang, D., et al., J. Bact., 179

(11), 3773-3782, 1997.

[00256] A number of DNA glyscosylases and AP endonucleases are commercially available, e.g., from New England Biolabs, Ipswich, MA.

[00257] In some embodiments of the invention extension probes comprising a site that is a substrate for cleavage by an AP endonuclease are used in the sequencing method as described above for extension probes containing a phosphorothiolate linkage or in sequencing methods AB (see below). In any of these methods, following ligation of an extension probe to a growing nucleic acid strand, the extension probe is cleaved using an AP endonuclease to remove the portion of the probe that comprises a label. [00258] Depending on the particular AP endonuclease, and depending on whether sequencing is performed in the 3 '-^5' or the 5 '->3' direction, it may be necessary or desirable to treat the extended duplex with a polynucleotide kinase or a phosphatase following cleavage in order to generate an extendable probe terminus on the extended duplex (see Figs. 5 A and 5B for depiction of extendable probe termini). Thus in certain methods of the invention an extendable terminus is generated by treatment with a polynucleotide kinase or phosphatase. One of ordinary skill in the art will appreciate that appropriate buffers will be employed for the various enzymes, and additional steps of washing may be included to remove enzymes and provide appropriate conditions for subsequent steps in the methods. [00259] In other embodiments the extension probe comprises a damaged base that is a substrate for removal by a DNA glycosylase. A wide range of cytotoxic and mutagenic DNA bases are removed by different DNA glycosylases, which initiate the base excision repair pathway following damage to DNA (Krokan, H.E., et al, Biochem J, 325 (Pt 1):1-16, 1997). DNA glycosylases cleave the N-glycosydic bond between the damaged base and deoxyribose, thus releasing a free base and leaving an apurinic/apyrimidinic (AP) site. In some embodiments the extension probe comprises a uracil residue, which is removed by a uracil-DNA glycosylase (UDG). UDGs are found in all living organisms studied to date, and a large number of these enzymes are known in the art and are of use in this invention (Frederica, et al, Biochemistry, 29, 2353-2537, 1990; Krokan, supra). For example, mammalian cells contain at least 4 types of UDG:

mitochondrial UNGI and nuclear UNG2, SMUGI, TDG, and MBD4 (Krokan, et al., Oncogene, 21, 8935-8948, 2002). UNGI and UNG2 belong to a highly conserved family typified by E. coli Ung.

In embodiments in which the extension probe comprises a damaged base, following ligation of the extension probe to an extendable probe terminus, the extended duplex is contacted with a glycosylase that removes the damaged base, thereby producing an abasic residue. An extension probe that comprises a damaged base that is subject to removal by a glycosylase is considered to be "readily modifiable to comprise a scissile linkage". The extended duplex is then contacted with an AP endonuclease, which cleaves a linkage between the abasic residue and an adjacent nucleoside, as described above. In certain embodiments of the invention a dual activity enzyme that is both a DNA glycosylase and an AP endonuclease is used to perform both of these reactions. In some embodiments the extended duplex containing a damaged base is contacted with a DNA glycosylase and an AP endonuclease. The enzymes can be used in combination or sequentially (i.e., glycosylase followed by endonuclease) in various embodiments of the invention. [00260] In some embodiments of the invention an extension probe comprises a trigger residue which is deoxyinosine. As noted above, E.coli Endonuclease V (Endo V), also called deoxyinosine 3 ' endonuclease, and homo logs thereof cleave a nucleic acid containing deoxyinosine at the second phoshodiester bond 3 ' to the deoxyinosine residue, leaving a 3 ' OH and 5' phosphate termini. Thus this bond serves as a scissile linkage in the extension probe. Endo V and its cleavage properties are known in the art (Yao, M. and Kow Y.W., J. Biol. Chem., Ill, 30672-30673 (1996); Yao, M. and Kow Y.W., J. Biol. Chem., 270, 28609- 28616 (1995); He, B, et al, Mutat. Res., 459, 109-114 (2000). In addition to deoxyinosine, Endo V also recognizes deoxyuridine, deoxyxanthosine, and deoxyoxanosine (Hitchcock, T. et al., Nuc. Acids Res., 32(13), 32(13) (2004). Mammalian homologs such as mEndo V also exhibit cleavage activity (Moe, A., et al., Nuc. Acids Res., 31(14), 3893-3900 (2004). While Endo V is a preferred cleavage agent for probes comprising deoxyinosine, other cleavage reagents may also be used to cleave probes comprising deoxyinosine. For example, as a damaged base, hypoxanthine may be subject to removal by an appropriate DNA glycosylase,

and the resulting extension probe containing an abasic residue is then subject to cleavage by an endonuclease.

[00261] It will be appreciated that if deoxyinosine is used as a trigger residue, it may be desirable to avoid using deoxyinosine elsewhere in the probe, particularly at positions between the terminus that will be ligated to the extendable probe terminus and the trigger residue. Thus if the probe comprises one or more universal bases, a nucleoside other than deoxyinosine may be used. It will also be appreciated that where a trigger residue that renders a nucleic acid containing the trigger residue susceptible to cleavage by a particular cleavage agent is used in an extension probe, it may be desirable to avoid including other residues in the probe (or in other probes that would be used in a sequencing reaction together with that extension probe) that would trigger cleavage by the same cleavage agent. [00262] The present invention encompasses the use of any enzyme that cleaves a nucleic acid that comprises a trigger residue. Additional enzymes may be identified by perusing the catalog of enzyme suppliers such as New England Biolabs(R), Inc. The New England Biolabs Catalog , 2005 edition (New England Biolabs, Ipswich, MA 01938-2723) is incorporated herein by reference, and the present invention contemplates use of any enzyme disclosed therein that cleaves a nucleic acid containing a trigger residue, or a homo log of such an enzyme. Other enzymes of use include, e.g., hOGGI and homologs thereof (Radicella, JP, et al., Proc Natl Acad Sci USA., 94(15):8010-5, 1997). [00263] Methods for synthesizing oligonucleotides containing a trigger residue such as a damaged base, abasic residue, etc. are known in the art. Methods for synthesizing oligonucleotides containing site that is a substrate for an AP endonuclease, e.g., oligonucleotides containing an abasic residue are known in the art and are generally amenable to automated solid phase oligonucleotide synthesis. In some embodiments an oligonucleotide containing uridine at the desired location of the abasic residue is synthesized. The oligonucleotide is then treated with an enzyme such as a UDG, which removes uracil, thereby producing an abasic residue wherever uridine was present in the oligonucleotide. [00264] In some embodiments of the invention the oligonucleotide probe comprises a disaccharide nucleoside as described in Nauwelaerts, K., et al, Nuc. Acids. Res., 31 (23), 2003. Following ligation, the extended duplex is cleaved using periodate (NaIO4), followed by treatment

with base (e.g., NaOH) to remove the label, resulting in a free 3' OH and P5- OPO3H2 group. Depending on whether sequencing is performed in the 3' - >5' or 5 '-^3', it may be necessary or desirable to treat the extended duplex with a polynucleotide kinase or phosphatase to generate an extendable terminus. Thus in certain methods of the invention an extendable terminus is generated by treatment with a polynucleotide kinase or phosphatase. [00265] A polynucleotide comprising a disaccharide nucleoside is considered to comprise an abasic residue. For example, a polynucleotide containing a ribose residue inserted between the 3 'OH of one nucleotide and the 5' phosphate group of the next nucleotide is considered to comprise an abasic residue. [00266] Capping

[00267] In some cases, fewer than all probes with extendable termini participate in a successful ligation reaction in each cycle of extension, ligation, and cleavage. It will be appreciated that if such probes participated in succeeding cycles, the accuracy of each nucleotide identification step would progressively decline. Although the inventors have shown that use of extension probes containing phophorothiolate linkages allows ligation with high efficiency, in certain embodiments of the invention a capping step is included to prevent those extendable termini that do not undergo ligation from participating in future cycles. When sequencing in the 5 '->3' direction using extension probes containing a 3'-O-P-S-5' phosphorothiolate linkage, capping may be performed by extending the unligated extendable termini with a DNA polymerase and a non-extendable moiety, e.g., a chain-terminating nucleotide such as a dideoxynucleotide or a nucleotide with a blocking moiety attached, e.g., following the ligation or detection step. When sequencing in the 3 '-^5' direction using extension probes containing a 3'-S-P-O-5' phosphorothiolate linkage, capping may be performed, e.g., by treating the template with a phosphatase, e.g., following ligation or detection. Other capping methods may also be used. [00268] H. Sequencing Using Oligonucleotide Probe Families

[00269] In the sequencing methods described above, referred to collectively as "Methods A", there is a direct and known correspondence between the label attached to any particular extension probe and the identity of one or more nucleotides at the proximal terminus of the probe (i.e., the terminus that is ligated to the extendable

probe terminus of the extended duplex. Therefore, identifying the label of a newly ligated extension probe is sufficient to identify one or more nucleotides in the template. The invention provides additional sequencing methods, referred to collectively as "Methods AB", and also involving successive cycles of extension, ligation, and, preferably, cleavage, that adopt a different approach to nucleotide identification.

[00270] The invention provide sequencing methods AB that use a collection of at least two distinguishably labeled oligonucleotide probe families. Each probe family is assigned a name based on the label, e.g., "red", "blue", "yellow", "green". As in the methods described above, extension starts from a duplex formed by an initializing oligonucleotide and a template. The initializing oligonucleotide is extended by ligating an oligonucleotide probe to its end to form an extended duplex, which is then repeatedly extended by successive cycles of ligation. The probe has a non-extendable moiety in a terminal position (at the opposite end of the probe from the nucleotide that is ligated to the growing nucleic acid strand of the duplex) so that only a single extension of the extended duplex takes place in a single cycle. During each cycle, a label on or associated with a successfully ligated probe is detected, and the non- extendable moiety is removed or modified to generate an extendable terminus. Detection of the label identifies the name of the probe family to which the probe belongs. [00271] Successive cycles of extension, ligation, and detection produce an ordered series of label names. The labels correspond to the probe families to which successfully ligated probes that hybridize to the template at successive positions belong. The probes have proximal termini that are located opposite different nucleotides in the template following ligation. Thus there is a correspondence between the order of probe family names and the order of nucleotides in the template.

[00272] In embodiments of the invention in which the scissile linkage is located between the proximal nucleoside in the extension probe and the adjacent nucleoside, the ordered list of probe family names may be obtained by successive cycles of extension, ligation, detection, and cleavage that begin from a single initializing oligonucleotide since the extended oligonucleotide probe is extended by one nucleotide in each cycle. If the scissile linkage is located between two

of the other nucleosides, the ordered list of probe family names is assembled from results obtained from a plurality of sequencing reactions in which initializing oligonucleotides that hybridize to different positions within the binding reaction are used, as described for sequencing methods A.

[00273] Knowing which probe family a newly ligated probe belongs to is not by itself sufficient to determine the identity of a nucleotide in the template. Instead, determining the probe family name eliminates certain combinations of nucleotides as possibilities for the sequence of at least a portion of the probe but leaves at least two possibilities for the identity of each nucleotide. Thus knowledge of the probe family name, in the absence of additional information, leaves open least two possibilities for the identity of the nucleotides in the template that are located at opposite positions to the nucleotides in the newly ligated probe. Therefore any single cycle of extension, ligation, detection (and, optionally, cleavage) does not itself identify any nucleotide in the template. However, it does allow elimination of one or more possible sequences for the template and thereby provides information about the sequence. In certain embodiments of the invention, with appropriate design of the probes and probe families as described below, the sequence of the template can still be determined. In certain embodiments of the invention sequencing methods AB thus comprise two phases: a first phase in which an ordered list of probe family names is obtained, and a second phase in which the ordered list is decoded to determine the sequence of the template. [00274] Unless otherwise indicated, sequencing methods A and AB generally employ similar methods for synthesizing probes, preparing templates, and performing the steps of extension, ligation, cleavage, and detection.

[00275] Features of Oligonucleotide Extension Probes and Probe Families for Sequencing Methods AB

[00276] Probe families for use in sequencing methods AB are characterized in that each probe family comprises a plurality of labeled oligonucleotide probes of different sequence and, at each position in the sequence, a probe family comprises at least 2 probes having different bases at that position. Probes in each probe family comprise the same label. Preferably the probes comprise a scissile internucleoside linkage. The

scissile linkage can be located anywhere in the probe. Preferably the probes have a moiety that is not extendable by ligase at one terminus. Preferably the probes are labeled at a position between the scissile linkage and the moiety that is not extendable by ligase, such that cleavage of the scissile linkage following ligation of a probe to an extendable probe terminus results in an unlabeled portion that is ligated to the extendable probe terminus and a labeled portion that is no longer attached to the unlabeled portion.

[00277] The probes in each probe family preferably comprise at least j nucleosides X, wherein j is at least 2, and wherein each X is at least 2-fold degenerate among the probes in the probe family. Probes in each probe family further comprise at least k nucleosides N, wherein k is at least 2, and wherein N represents any nucleoside. In general, j + k is equal to or less than 100, typically less than or equal to 30. Nucleosides X can be located anywhere in the probe. Nucleosides X need not be located at contiguous positions. Similarly nucleosides N need not be located at contiguous positions. In other words, nucleosides X and N can be interspersed. Nevertheless, nucleosides X can be considered to have a 5'->3' sequence, with the understanding that the nucleosides need not be contiguous. For example, nucleosides X in a probe of structure XANXGNNXCN would be considered to have the sequence AGC. Similarly, nucleosides N can be considered to have a sequence.

[00278] Nucleosides X can be identical or different but are not independently selected, i.e., the identity of each X is constrained by the identity of one or more other nucleosides X in the probe. Thus in general only certain combinations of nucleosides X are present in any particular probe and within the probes in any particular probe family. In other words, in each probe, the sequences of nucleosides X can only represent a subset of all possible sequences of length j. Thus the identity of one or more nucleotides in X limits the possible identities for one or more of the other nucleosides.

[00279] Nucleosides N are preferably independently selected and can be A, G, C, or T (or, optionally, a degeneracy-reducing nucleoside). Preferably the sequence of nucleosides N represents all possible sequences of length k, except that one or more N may

be a degeneracy-reducing nucleoside. The probes thus contain two portions, of which the portion consisting of nucleosides N is referred to as the unconstrained portion and the portion consisting of nucleosides X is referred to as the constrained portion. As described above, the portions need not be contiguous nucleosides. Probes that contain a constrained portion and an unconstrained portion are referred to herein as partially constrained probes. Preferably one or more nucleosides in the constrained portion is at the proximal end of the probes, i.e., at the end that contains the nucleoside that will be ligated to the extendable probe terminus, which can be either the 5' or 3' end of the oligonucleotide probe in different embodiments of the invention.

[00280] Since the constrained portion of any oligonucleotide probe can only have certain sequences, knowing the identity of one or more of the nucleosides in the constrained portion of a probe provides information about one or more of the other nucleosides. The information may or may not be sufficient to precisely identify one or more of the other nucleosides, but it will be sufficient to eliminate one or more possibilities for the identity of one or more of the other nucleosides in the constrained portion. In certain preferred embodiments of sequencing methods AB, knowing the identity of one nucleoside in the constrained portion of a probe is sufficient to precisely identify each of the other nucleosides in the constrained portion, i.e., to determine the identity and order of the nucleosides that comprise the constrained portion. [00281] As in the sequencing methods described above, the most proximal nucleoside in an extension probe that is complementary to the template is ligated to an extendable terminus of an initializing oligonucleotide (in the first cycle of extension, ligation, and detection) and to an extendable terminus of an extended oligonucleotide probe in subsequent cycles of extension, ligation, and detection. Detection determines the name of the probe family to which the newly ligated probe belongs. Since each position in the constrained portion of the probe is at least 2-fold degenerate, the name of the probe family does not in itself identify any nucleotide in the constrained portion. However, since the sequence of the constrained portion is one of a subset of all possible sequences of length j, identifying the probe family does eliminate certain possibilities for the sequence of the constrained portion. The constrained portion of the probe constitutes its

sequence determining portion. Therefore, eliminating one or more possibilities for the identity of one or more nucleosides in the constrained portion of the probe by identifying the probe family to which it belongs eliminates one or more possibilities for the identity of a nucleotide in the template to which the extension probe hybridizes. In preferred embodiments of the invention the partially constrained probes comprise a scissile linkage between any two nucleosides.

[00282] In certain embodiments the partially constrained probes have the general structure (X)j(N)k, in which X represents a nucleoside, (X), is at least 2-fold degenerate at each position such that X can be any of at least 2 nucleosides having different base-pairing specificities, N represents any nucleoside, j is at least 2, k is between 1 and 100, and at least one N or X other than the X at the probe terminus comprises a detectable moiety. Preferably (N)k is independently 4-fold degenerate at each position so that, in each probe, (N)k represents all possible sequences of length k, except that one or more positions in (N)k may be occupied by a degeneracy-reducing nucleotide. Nucleosides in (X), can be identical or different but are not independently selected. In other words, in each probe, (X), can only represent a subset of all possible sequences of length j. Thus the identity of one or more nucleotides in (X), limits the possible identities for one or more of the other nucleosides. The probes thus contain two portions, of which (N)k is the unconstrained portion and (X), is the constrained portion. [00283] In certain preferred embodiments of the invention the partially constrained probes have the structure 5'-(X)j(N)kNB*-3' or 3'-(X)j(N) kNB*-5', wherein N represents any nucleoside, NB represents a moiety that is not extendable by ligase, * represents a detectable moiety, (X), is a constrained portion of the probe that is at least 2-fold degenerate at each position, nucleosides in (X), can be identical or different but are not independently selected, at least one internucleoside linkage is a scissile linkage, j is at least 2, and k is between 1 and 100, with the proviso that a detectable moiety may be present on any nucleoside N or X other than the X at the probe terminus instead of, or in addition to, NB. The scissile linkage can be between two nucleosides in (X),, between the most distal nucleotide in (X), and the most proximal nucleoside in (N)k, between nucleosides within (N)k, or between the terminal nucleoside in (N)k and NB. Preferably the scissile linkage is a phosphorothiolate

linkage. [00284] In yet more preferred embodiments of the invention the probes have the structure 5'-(XY)(N) kNB*-3' or 3'-(XY)(N)kNB*-5', wherein N represents any nucleoside, NB represents a moiety that is not extendable by ligase, * represents a detectable moiety, XY is a constrained portion of the probe in which X and Y represent nucleosides that are identical or different but are not independently selected, X and Y are at least 2-fold degenerate, at least one internucleoside linkage is a scissile linkage, and k is between 1 and 100, inclusive, with the proviso that a detectable moiety may be present on any nucleotide N or X other than the X at the probe terminus instead of, or in addition to, NB. Preferably the scissile linkage is a phosphorothiolate linkage. Probes having the structure 5'-(XY)(N)kN[beta] *-3' are of use for sequencing in the 5 '->3' direction. Probes having the structure 3'-(XY)(N)kN[beta]*-5' are of use for sequencing in the 3 '-^5' direction.

[00285] The structure of certain preferred probes is represented in more detail as follows. For sequencing in the 5 '->3' direction, partially constrained probes having the structure 5'- O-P-O-(X)J(N)k-O-P-S-(N)1NB*-3' where N represents any nucleoside, NB represents a moiety that is not extendable by ligase, * represents a detectable moiety, (X), is a constrained portion of the probe that is at least 2-fold degenerate at each position, nucleosides in (X), can be identical or different but are not independently selected, j is at least 2, (k+i) is between 1 and 100, k is between 1 and 100, and i is between 0 and 99, with the proviso that a detectable moiety may be present on any nucleoside Of (N)1 instead of, or in addition to, NB. In certain embodiments of the invention (X), is (XY) in which X and Y are at least 2-fold degenerate and represent nucleotides that are identical or different but are not independently selected. In certain embodiments of the invention i is 0.

[00286] Other preferred probes for sequencing in the 5 ' ->3 ' direction have the structure 5 '-0-P-O-(X)J-O-P-S-(N)1NB^-S ' in which N represents any nucleoside, NB represents a moiety that is not extendable by ligase, * represents a detectable moiety, (X), is a constrained portion of the probe that is at least 2-fold degenerate at each position, nucleotides in (X), can be identical or different but are not independently selected, j is at least 2, and i is between 1 and 100, with the proviso that a detectable moiety may be present on any nucleoside Of (N)1 instead of, or in addition to, NB. In certain

embodiments of the invention (X), is (XY), in which positions X and Y are at least 2-fold degenerate and X and Y represent nucleosides that are identical or different but are not independently selected. Yet other preferred probes for sequencing in the 5'->3' direction have the structure 5 '-0-P-O-(X)J-O-P-S-(XX(N)1NB*- 3' in which N represents any nucleoside, NB represents a moiety that is not extendable by ligase, * represents a detectable moiety, (X)j-O-P-S-(X)k is a constrained portion of the probe that is at least 2-fold degenerate at each position, positions in (X)j-O-P-S-(X)k are at least 2-fold degenerate and can be identical or different but are not independently selected, j and k are both at least 1 and (j+k) is at least 2 (e.g., 2, 3, 4, or 5), and i is between 1 and 100, with the proviso that a detectable moiety may be present on any nucleoside of (N)1 instead of, or in addition to, NB. In certain embodiments of the invention j and k are both 1. [00287] For sequencing in the 3 '-^5' direction, partially constrained probes having the structure 5'-NB*(N)rS-P-O-(N)k-O-P-O-(X)j-3' where N represents any nucleoside, NB represents a moiety that is not extendable by ligase, * represents a detectable moiety, (X), is a constrained portion of the probe that is at least 2-fold degenerate at each position, nucleosides in (X)j can be identical or different but are not independently selected, j is at least 2, (k+i) is between 1 and 100, k is between 1 and 100, and i is between 0 and 99, with the proviso that a detectable moiety may be present on any nucleoside Of (N)1 instead of, or in addition to, NB. In certain embodiments of the invention (X), is (XY) in which X and Y are at least 2-fold degenerate and represent nucleosides that are identical or different but are not independently selected. In certain embodiments of the invention i is 0.

[00288] Other preferred probes for sequencing in the 3 ' ->5 ' direction have the structure 5 '-NB*(N)I-S-P-O-(X) J-3 ' in which N represents any nucleoside, NB represents a moiety that is not extendable by ligase, * represents a detectable moiety, (X), is a constrained portion of the probe that is at least 2-fold degenerate at each position, nucleosides in (X), can be identical or different but are not independently selected, j is at least 2, and i is between 1 and 100, with the proviso that a detectable moiety may be present on any nucleoside Of (N)1 instead of, or in addition to, NB. In certain embodiments of the invention (X), is (XY) in which X and Y are at least 2-fold degenerate and represent

nucleosides that are identical or different but are not independently selected. In certain embodiments of the invention j is between 2 and 5, e.g., 2, 3, 4, or 5, in any of the partially constrained probes. [00289] Yet other preferred probes for sequencing in the 3 ' ->5 ' direction have the structure 5 '-NB^(N)1-S-P-O-(XVO-P-O-(X)J-S ' where N represents any nucleoside, NB represents a moiety that is not extendable by ligase, * represents a detectable moiety, -(X)k- O-P-O-(X)j is a constrained portion of the probe that is at least 2-fold degenerate at each position, nucleosides in -(X)k-O-P-O-(X), can be identical or different but are not independently selected, j and k are both at least 1 and (j+k) is at least 2 (e.g., 2, 3, 4, or 5), i is between 1 and 100, with the proviso that a detectable moiety may be present on any nucleoside of (N)1 instead of, or in addition to, NB. In certain embodiments j = 1 and k =1. [00290] In embodiments of the invention in which the scissile linkage is located between the most proximal nucleoside in (X), and the next most proximal nucleoside in (X),, the ordered list of probe family names may be obtained by successive cycles of extension, ligation, detection, and cleavage that begin from a single initializing oligonucleotide since the extended oligonucleotide probe is extended by one nucleotide in each cycle. In embodiments of the invention in which the scissile linkage is located between two of the other nucleosides, the ordered list of probe family names is assembled from results obtained from a plurality of sequencing reactions in which initializing oligonucleotides that hybridize to different positions within the binding reaction are used, as described for sequencing methods A. [00291] It will be understood that probes having any of a large number of structures other than those described above can be employed in sequencing methods AB. For example, probes can have structures such as XNY(N)k in which the constrained nucleosides X and Y are not adjacent, or XIY(N)k where I is a universal base. (N)kX(N)i, (N),X (N)jY(N)kZ(N)i, (N)1X(N)JYIZ(N)I, and (N),X(N)jY(N)kZ (I)i represent additional possibilities. As in the probes described above, these probes comprise a scissile linkage, a detectable moiety, and a moiety at one terminus that is not extendable by ligase. Preferably the probes do not comprise a detectable moiety attached to the nucleotide at the opposite end of the probe from the moiety that is not extendable by ligase. Probe families comprising probes having any of these structures, or others, satisfy the criterion that each probe family comprises a plurality of labeled probes of different

sequence and, at each position in the sequence, a probe family comprises at least 2 probes having different bases at that position. Preferably the total number of nucleosides in each probe is 100 or less, e.g., 30 or less.

[00292] Encoding Oligonucleotide Extension Probe Families. The inventive sequencing method makes use of encoded probe families. An "encoding" refers to a scheme that associates a particular label with a probe comprising a portion that has one of a defined set of sequences, such that probes comprising a portion that has a sequence that is a member of the defined set of sequences are labeled with the label. In general, an encoding associates each of a plurality of distinguishable labels with one or more probes, such that each distinguishable label is associated with a different group of probes, and each probe is labeled by only a single label (which can comprise a combination of detectable moieties). Preferably the probes in each group of probes each comprise a portion that has a sequence that is a member of the same defined set of sequences. The portion may be a single nucleoside or may be multiple nucleosides in length, e.g., 2, 3, 4, 5, or more nucleosides in length. The length of the portion may constitute only a small fraction of the entire length of the probe or may constitute up to the entire probe. The defined set of sequences may contain only a single sequence or may contain any number of different sequences, depending on the length of the portion. For example, if the portion is a single nucleoside, the defined set of sequences could have at most 4 elements (A, G, C, T). If the portion is two nucleosides in length, the defined set of sequences could have up to 16 elements (AA, AG, AC, AT, GA, GG, GC, GT, CA, CG, CC, CT, TA, TG, TC, TT). In general, the defined set of sequences will contain fewer elements than the total number of possible sequences, and an encoding will employ more than one defined set of sequences.

[00293] Sequencing methods A described herein generally make use of a set of probes having a simple encoding in which there is a direct correspondence between the proximal nucleoside in the probe (i.e., the nucleoside that is ligated to the extendable probe terminus) and the identity of the label. The proximal nucleoside is complementary to the nucleotide with which it hybridizes in the template, so the identity of the proximal nucleoside in a newly ligated probe determines

the identity of the nucleotide in the template that is located at the opposite position in the extended duplex. In a general sense, probes of use in the other sequencing methods described herein have the structure X(N)k, in which X is the proximal nucleoside, and each nucleoside N is 4-fold degenerate, such that all possible sequences of length k are represented in the pool of oligonucleotide probe molecules that constitutes the probe. Thus, for example, some oligonucleotide probe molecules will contain A at position k = 1 , others will contain G at position k = 1 , others will contain C at position k = 1 , others will contain T at position k = 1, and similarly for other positions k, where the nucleoside adjacent to X in (N)k is considered to occupy position k = 1 ; the next nucleoside in (N)k is considered to occupy position k = 2, etc. However, in any given oligonucleotide probe, X represents only a single base pairing specificity, which typically corresponds to a particular nucleoside identity, e.g., A, G, C, or T. Thus X is typically uniformly A, G, C, or T in the pool of probe molecules that constitute a particular probe. Figure 2 shows a suitable encoding for probes having the structure X(N)k. According to this encoding, probes having X = C are assigned the label "red"; probes having X = A are assigned the label "yellow"; probes having X = G are assigned the label "green"; and probes having X = T are assigned the label "blue". Thus there is a one-to-one corresondence between the sequence determining portion of the probe and its label.

[00294] It will be appreciated that the above approach in which the identity of the label of a newly ligated extension probe corresponds to the identity of the most proximal nucleoside in the extension probe may be broadened to encompass encodings in which the identity of the label corresponds not to the identity of only the most proximal nucleoside in the extension probe but rather to the sequence of the most proximal 2 or more nucleosides in the extension probe, so that the identity of multiple nucleotides in the template can be determined in a single cycle of extension, ligation, and detection (typically followed by cleavage). However, such encodings would still associate a label with a single sequence in the oligonucleotide extension probe so that the identity of the oppositely located complementary nucleotides in the template could be identified. For example, as described above, in order to identify two nucleotides in a single cycle, 16 different

oligonucleotide probes, each with a corresponding label (i.e., 16 distinguishable labels) would be needed.

[00295] Sequencing method AB employs an alternative approach to associating labels with probes. Rather than a one-to-one correspondence between the identity of the label and the sequence of the sequence determining portion of the probe, the same label is assigned to multiple probes having different sequence determining portions. The probes are partially constrained, and the constrained portion of the probe is its sequence determining portion. Thus the same label is assigned to a plurality of different probes, each having a constrained portion with a different sequence, wherein the sequence is one of a defined set of sequences. As mentioned above, probes comprising the same label constitute a "probe family". The method employs a plurality of such probe families, each comprising a plurality of probes having a constrained portion with a different sequence, wherein the sequence is one of a defined set of sequences.[00296] A plurality of probe families is referred to as a "collection" of probe families. Probes in each probe family in a collection of probe families are labeled with a label that is distinguishable from labels used to label other probe families in the collection. Each probe family preferably has its own defined set of sequences. Preferably the constrained portions of the probes in each probe family are the same length, and preferably the constrained portions of probe families in a collection of probe families are of the same length. Preferably the combination of sets of defined sequences for probe families in a collection of probe families includes all possible sequences of the length of the constrained portion. Preferably a collection of probe families comprises or consists of 4 distinguishably labeled probe families. Preferably the constrained portion of the probes is 2 nucleosides in length. [00297] A wide variety of differently encoded collections of distinguishably labeled probe families will satisfy the above criteria and may be used to practice the inventive method. However, certain collections of probe families are preferred. An exemplary encoding for a preferred collection of 4 distinguishably labeled probe families comprising partially constrained probes is shown in Figure 25 A. As depicted in Figure 25 A, the constrained portion consists of the 2 most 3' nucleosides in the probe. The probe families are labeled "red", "yellow", "green", and "blue". Probes in each probe family comprise a constrained portion whose sequence is one

of a defined set of sequences, the defined set being different for each probe family. For example, beginning at the 3 ' end of each sequence, which is considered to be the proximal end of the probe, the defined set of sequences for the "red" probe family is {CT, AG, GA, TC}; the defined set of sequences for the "yellow" probe family is {CC, AT, GG, TA}; the defined set of sequences for the "green" probe family is {CA, AC, GT, TG}; the defined set of sequences for the "blue" probe family is {CG, AA, GC, TT} . Each defined set does not contain any member that is present in one of the other sets, a characteristic that is preferred. In addition, the combination of sets of defined sequences for probe families in a collection of probe families includes all possible sequences of length 2, i.e., all possible dinucleosides. Another characteristic of this collection of probe families, which is preferred but not required, is that each position in the constrained portion of the probes is 4-fold degenerate, i.e., it can be occupied by either A, G, C, or T. Another characteristic of this collection of probe families, which is preferred but not required, is that within each set of defined sequences only a single sequence has any specific nucleoside at any position, e.g., at the most proximal position or at any of other positions. It is particularly preferred, but not required, that within each set of defined sequences only a single sequence has any specific nucleoside at position 2 or higher within the constrained portion, considering the most proximal nucleoside to be at position 1. For example, in the defined set of sequences for the Red probe family, only one sequence has T at position 2; only one sequence has G at position 2; only one sequence has A at position 2; only one sequence has C at position 2.

[00298] Given any particular encoding such as that depicted in Figure 25 A, knowing the identity of one or more nucleosides in the constrained portion of a probe in one of the probe families provides information about the other nucleotides in the constrained portion of that probe. In the most general sense, knowing the identity of one or more nucleosides in the constrained portion of a probe in a probe family provides sufficient information to eliminate one or more possible identities for a nucleoside at one of the other positions, because the defined set of sequences for that probe family will not contain a sequence having a nucleoside with that identity at that position. Typically knowing the identity of one or more nucleosides in the constrained portion of a

probe in a probe family provides sufficient information to eliminate one or more possible identities for a plurality of nucleosides, e.g., each of the other nucleosides. For preferred encodings, knowing the identity of one or more nucleosides in the contrained portion of a probe in the probe family eliminates all but one possibility for each of the other nucleosides in the probe. For example, in the case of the encoded probe families shown in Figure 25 A, if it is known that a probe is a member of the red family, and if it is also known that the most proximal nucleoside is C, then the adjacent nucleoside must be T. Similarly, if it is known that a probe is a member of the green family, and if it is also known that the most proximal nucleoside is G, then the adjacent nucleoside must be T. Thus knowing the identity of one nucleoside in the constrained portion is sufficient to eliminate all possibilities for the other nucleoside except one, so the identity of the other nucleoside is completely specified. Yet without knowing the identity of at least one nucleoside in the constrained portion of a probe it is not possible to gain any information at all about the identity of any specific nucleoside in the probe based only on knowing the name of the probe family to which it belongs since the nucleoside at each position of the constrained portion could be A, G, C, or T. Fig. 25B shows a preferred collection of probe families (upper panel) and a cycle of ligation, detection, and cleavage (lower panel) using sequencing methods AB.

[00299] The inventors have designed 24 collections of probe families containing constrained portions that are 2 nucleosides in length and that have the advantageous features of the collection of probe families depicted in Figure 25 A. These probe families are maximally informative in that knowing the name of the probe family to which a probe belongs, and knowing the identity of one nucleoside in the probe, is sufficient to precisely identify the other nucleoside in the constrained portion. This is the case for all probes, and for all nucleosides in each constrained portion. The encoding schemes for each of the 24 preferred collections of probe families are shown in Table 1. Table 1 assigns an encoding ID ranging from 1 to 24 to each collection of probe families. Each encoding defines the constrained portions of a collection of preferred probe families of general structure (XY)Nk for use in sequencing methods AB, and thereby defines the collection itself. In Table 1 a value of 1 in the column under an encoding ID indicates

that, according to that encoding, a probe comprising nucleosides X and Y as indicated in the first and second columns, respectively, is assigned to the first probe family; (ii) a value of 2 in the column under an encoding ID indicates that, according to that encoding, a probe comprising nucleosides X and Y as indicated in the first and second columns, respectively, is assigned to the second probe family; (iii) a value of 3 in the column under an encoding ID indicates that, according to that encoding, a probe comprising nucleosides X and Y as indicated in the first and second columns, respectively, is assigned to the third probe family; and (iv) a value of 4 in the column under an encoding ID indicates that, according to that encoding, a probe comprising nucleosides X and Y as indicated in the first and second columns, respectively, is assigned to the fourth probe family. The values 1, 2, 3, and 4, each represent a label. For example, encoding 9 defines the collection of probe families depicted in Figure 25 A, in which 1 represents blue, 2 represents green, 3 represents red, and 4 represents yellow. It will be appreciated that the assignment of values to labels is arbitrary, e.g., 1 could equally well represent green, red, or yellow. Changing the association between values 1, 2, 3, and 4, and the labels would not change the set of probes in each probe families but would merely associate a different label with each probe family.

[00300]

Table 1 : Oligonucleotide Probe Family Encodings

Image available on "Original document"

[00301] To further illustrate the use of Table 1 to define the collections of preferred probe families, consider encoding 17. According to this encoding, probes having constrained portions AA, GC, TG, and CT are assigned to label 1 (e.g., red); probes having constrained portions CA, AC, GG, and TT are assigned to label 2 (e.g., yellow); probes having constrained portions TA, CC, AG, and GT are assigned to label 3 (e.g., green); and probes having constrained portions GA, TC, CG, and AT are assigned to label 4 (e.g., blue). The resulting collection of probe families is depicted in Figure 26.

[00302] Figures 27A-27C represent an alternate method to schematically define the 24 preferred collections of

probe families. The method makes use of diagrams such as that in Figure 27 A. The first column in such a diagram represents the first base. Each label is attached to four different base sequences, each of which is given by juxtaposing the base from the first column with the base from the chosen label's column. For example, if there is an A in the column with the heading "First base", then a probe with constrained portion having sequence AA is assigned to probe family 1 (label 1); a probe with constrained portion having sequence AC is assigned to probe family 2 (label 2); a probe with constrained portion having sequence AG is assigned to probe family 3 (label 3); and a probe with constrained portion having sequence AT is assigned to probe family 4 (label 4). Assignments to probe families are made in a similar manner for probes with constrained portions beginning with C, G, or T. Thus a diagram filled in with bases as shown in Figure 27 A translates to the encoding shown in Figure 27B, in which probes having constrained portions in the set {AA, CC, GG, TT} are assigned to probe family 1; probes having constrained portions in the set {AC, CA, GT, TG} are assigned to probe family 2; probes having constrained portions in the set {AG, CT, GC, TA} are assigned to probe family 3; and probes having constrained portions in the set {AT, CG, GA, TC} are assigned to probe family 4. Figure 27C shows diagrams that may be inserted in place of the shaded portion of the diagram in Figure 27 A in order to generate each of the 24 preferred collections of probe families. Methods of using the preferred collections of probe families in sequencing methods AB are described further below.

[00303] The 24 collections of encoded probe families defined by Table 1 represent only the preferred embodiments of collections of probe families for use in sequencing methods AB. A wide variety of other encoding schemes, probe families, and probe structures can be used that employ the same basic principle, in which knowing a probe family name, together with knowledge of the identity of one or more nucleosides in a constrained portion, provides information about one or more other nucleosides. As compared with a preferred collection of probe families, the less preferred collections of probe families are generally less preferred because: (i) at least with respect to some probes, the amount of information afforded by knowing a probe family name and a nucleoside identity is less; or (ii) at least with respect to some probes, the amount of

information afforded by knowing a probe family name is more. [00304] In general, less preferred collections of probe families may be used to perform sequencing methods AB in a similar manner to the way in which preferred collections of probe families are used. However, the steps needed for decoding may differ. For example, in some situations comparing candidate sequences with each other may be sufficient to determine at least a portion of a sequence.

[00305] An example of a less preferred collection of probe families in which the probes comprise constrained portions that are 2 nucleosides in length is shown in Figure 28. According to this encoding, probes having constrained portions in the set {AA, AC, GA, GC} are assigned to probe family 1; probes having constrained portions in the set {CA, CC, TA, TC} are assigned to probe family 2; probes having constrained portions in the set {AG, AT, GG, GT} are assigned to probe family 3; and probes having constrained portions in the set {CG, CT, TG, TT} are assigned to probe family 4. In this collection of probe families, knowing the name of a probe family eliminates certain possibilities for the identity of a nucleotide in the template that is located opposite the proximal nucleoside in a newly ligated extension probe whose label was detected to determine the name of the probe family. For example, if the probe family name is 1, then the proximal nucleoside in a newly ligated extension probe must be A or G, so the complementary nucleotide in the template must be T or C. Since there are at least two possibilities at each position in the constrained portion, the nucleotide cannot be precisely identified, but information sufficient to rule out some possibilities is obtained from the single cycle, in contrast to the situation when preferred collections of probe families are employed.

[00306] In certain embodiments of the invention partially constrained probes in which the constrained portion is 3 nucleosides in length are used. In order to contain probes whose constrained portions include all possible sequences of length 3, as is preferred, the collection of probe families should comprise 4<3> = 64 different probes. Figure 29 A shows a diagram that can be used to generate constrained portions for a collection of probe families that comprises probes with a constrained portion 3 nucleosides long (trinucleosides). The figure shows 4 sets of rows indicated A, G, C, and T, and 4 columns with probe family names 1, 2, 3, and 4. Each

set of 4 rows is opposite a box with a nucleoside identity inside. To determine a probe family for a trinucleoside, the box containing the last nucleoside in the trinucleoside is first selected. Within the four rows adjacent to that box, the row labeled with the letter identifying the first nucleoside in the trinucleoside is selected. Within that row, the column containing the second nucleoside of the trinucleoside is selected. The trinucleoside is assigned to the probe family indicated at the top of the column. For example, the following procedure is followed to assign the trinucleoside "TCG" to a probe family: Since the last nucleoside is a "G", attention is confined to the set of 4 rows located opposite the box containing "G", i.e., the third set of rows. Since the first nucleoside is "T", consideration is further limited to the last row in the set of 4. The probe family assignment is determined by the heading of the column that contains middle nucleoside. Since the middle nucleoside is "C", the trinucleoside is assigned to probe family 1. A similar process yields the following probe family assignments: AAA = 1; ATA = 2; AGA = 3; GTA = 4; GAG = 1; TGG = 2, etc. The process continues until all possible trinucleosides have been assigned to a probe family. [00307] Figure 29B shows a procedure for constructing additional constrained portions for a collection of probe families that comprises probes with a constrained portion 3 nucleosides long. The procedure is used to construct such a collection from each of the 24 preferred collections of probe families described above, in which constrained portions are 2 nucleosides in length and the collection contains 4 probe families. An exemplary diagram representing a preferred collection of probe families is shown in the upper portion of the figure. The columns of this diagram map directly into the columns of the lower portion of the figure in accordance with the color assigned to each column in the upper diagram. Thus the columns in the upper diagram are blue, green, yellow, and red, moving from left to right. The entries under column 1 in the lower diagram are blue, green, yellow, and red, moving from top to bottom, with each set of 4 nucleosides corresponding to a column in the upper diagram. Columns 2, 3, and 4 in the lower diagram are generated by progressively moving each set of 4 nucleosides in column 1 downwards.

[00308] It will be appreciated that a "probe family" can be considered to be a single "super-probe" comprising a plurality of different probes, each with the same label. In

this case, the probe molecules that constitute the probe will generally not be a population of substantially identical molecules across any portion of the probe. Use of the term "probe family" is not intended to have any limiting effect but is used for convenience to describe the characteristics of probes that would constitute such a "super-probe". [00309] Decoding

[00310] As described above, successive cycles of extension, ligation, detection, and cleavage using a collection of probe families comprising at least two distinguishably labeled probe families yields an ordered list of probe family names either from a single sequencing reaction or from assembling probe family names determined in multiple sequencing reactions that initiate from different sites in the template into an ordered list. The number of cycles performed should be approximately equivalent to the length of sequence desired. The ordered list contains a substantial amount of information but not in a form that will immediately yield the sequence of interest. Further step(s), at least one of which involves gathering at least one item of additional information about the sequence, must be performed in order to obtain a sequence that is most likely to represent the sequence of interest. The sequence that is most likely to represent the sequence of interest is referred to herein as the "correct" sequence, and the process of extracting the correct sequence from the ordered list of probe families is referred to as "decoding". It will be appreciated that elements in an "ordered list" as described above could be rearranged either during generation of the list or thereafter, provided that the information content, including the correspondence between elements in the list and nucleotides in the template, is retained, and provided that the rearrangement, fragmentation, and/or permutation is appropriately taken into consideration during the decoding process (discussed below). The term "ordered list" is thus intended to encompass rearranged, fragmented, and/or permuted versions of an ordered list generated as described above, provided that such rearranged, fragmented, and/or permuted versions include substantially the same information content.

[00311] The ordered list can be decoded using a variety of approaches. Some of these approaches involve generating a set of at least one candidate sequence from the ordered list of probe family names. The set of

candidate sequences may provide sufficient information to achieve an objective. In preferred embodiments one or more additional steps are performed to select the sequence that is most likely to represent the sequence of interest from among the candidate sequences or from a set of sequences with which the candidate sequence is compared. For example, in one approach at least a portion of at least one candidate sequence is compared with at least one other sequence. The correct sequence is selected based on the comparison. In certain embodiments of the invention, decoding involves repeating the method and obtaining a second ordered list of probe family names using a collection of probe families that is encoded differently from the original collection of probe families. Information from the second ordered list of probe families is used to determine the correct sequence. In some embodiments information obtained from as little as one cycle of extension, ligation, and detection using the alternately encoded collection of probe families is sufficient to allow selection of the correct sequence. In other words, the first probe family identified using the alternately encoded probe family provides sufficient information to determine which candidate sequence is correct.

[00312] Other decoding approaches involve specifically identifying at least one nucleotide in the template by any available sequencing method, e.g., a single cycle of sequencing method A. Information about the one or more nucleotide(s) is used as a "key" to decode the ordered list of probe family names. Alternately, the portion of the template that is sequenced may comprise a region of known sequence in addition to a region whose sequence is unknown. If sequencing methods AB are applied to a portion of the template that includes both unknown sequence and at least one nucleotide of known sequence, the known sequence can be used as a "key" to decode the ordered list of probe family names. The following section describes the process of generating candidate sequences. Subsequent sections describe using the candidate sequences to select the correct sequence by comparing with known sequences, by comparing with a second set of candidate sequences, and by utilizing a known nucleotide identity.
[00313] Generating Candidate Sequences

[00314] It will be appreciated that the region of the template to be sequenced is complementary to the extended duplex that is produced by successive cycles

of extension, ligation, and cleavage. Therefore, generating a candidate sequence for the extended duplex is equivalent to generating a candidate sequence for the region of the template to be sequenced. In practice, one could generate candidate sequences for the region of the template to be sequenced, or one could generate candidate sequences for the extended duplex and take their complement to determine candidate sequences for the region of the template to be sequenced. The latter approach is described here. To generate a candidate sequence from a list of probe family names, the first member of the list of probe families is considered. The set of constrained portions associated with that probe family limits the possibilities for the initial nucleotides in the sequence, out to a length equivalent to the length of the constrained portion. For example, if the constrained portion is a dinucleotide, then the possible sequences for the first dinucleotide in the extended duplex are limited to those constrained portions that occur in probes that fall within that probe family (and thus the possible sequences for the first dinucleotide in the region of the template to be sequenced are limited to those combinations that are complementary to the constrained portions that occur in probes that fall within that probe family). The possibilities for the first dinucleotide are recorded, typically by a computer. Similarly, the possible sequences for the second dinucleotide in the extended duplex (i.e., the dinucleotide that is one nucleotide offset from the first dinucleotide) are limited to those constrained portions that occur in probes that fall within the second probe family (and therefore, the possible sequences for the second dinucleotide in the template, i.e., the dinucleotide that is one nucleotide offset from the first dinucleotide are limited to those combinations that are complementary to the constrained portions that occur in probes that fall within the second probe family). The possible sequences for the second dinucleotide are also recorded. Possibilities for succeeding dinucleotides are likewise recorded until possibilities have been recorded for dinucleotides that correspond to the desired length of the sequence to be determined or there are no more probe families in the list.

[00315] A representative example of the process of recording possibilities is depicted in Figure 30, in which it is assumed that a list of probe family names has been generated using the probe family collection shown in Figure 25 A. The leftmost column of Figure 30 shows

the list of probe families in order from top to bottom: Yellow, Green, Red, Blue. The sequence possibilities for the dinucleotide corresponding to each probe family in the list are shown on the right side of the figure. Nucleotide positions are indicated above the sequence possibilities. The sequence begins at position 1, so the first dinucleotide occupies positions 1 and 2; the second dinucleotide occupies positions 2 and 3, etc. For the Yellow probe family, the possibilities are CC, AT, GG, and TA, as shown in Figure 30. For the Green probe family, the possibilities are CA, AC, GT, and TG, etc. The process of recording the possible sequences of each dinucleotide is continued until a desired sequence length has been reached. [00316] After the sets of possibilities are generated, a first assumption is made about the identity of the first nucleotide in the candidate sequence, which is assumed to be at the 5' position of the sequence, indicated as position 1 in Figure 30. The first assumption can be that the nucleotide is A, that the nucleotide is G, that the nucleotide is C, or that the nucleotide is T.

[00317] It will be observed that the possible sequences for each dinucleotide are limited by the possible sequences of the adjacent dinucleotides, since adjacent dinucleotides overlap, i.e., the second nucleotide of the first dinucleotide is also the first nucleotide of the second dinucleotide. For example, if the first nucleotide is assumed to be C, then the first dinucleotide must be CC. If the first dinucleotide is CC, then the second dinucleotide must have a C at its first position. Since the only possible sequence for the second dinucleotide that has a C at its first position is CA, it is evident that the second dinucleotide must be CA. Therefore the sequence of the first 3 nucleotides must be CCA. Similarly, the possible sequences for the third dinucleotide are limited by the possible sequences of the second dinucleotide. If the second dinucleotide is CA, then the third dinucleotide must be AG since that is the only possibility that has A at its first position. Thus the sequence of the first 4 nucleotides must be CCAG. Continuing this process results in a sequence of 5'-CCAGC-3' for the first 5 nucleotides. CCAGC is thus the first candidate sequence. [00318] A second candidate sequence is generated by assuming that the first nucleotide is A. This assumption yields AT for the first dinucleotide. TG is the only possible sequence for the second dinucleotide that is consistent with a sequence of AT for the first dinucleotide. GA is the only possible

sequence for the third dinucleotide that is consistent with a sequence of TG for the second dinucleotide. AA is the only possible sequence for the fourth dinucleotide that is consistent with a sequence of GA for the third dinucleotide. Assembling these dinucleotides into a full length candidate sequence yields ATGAA. Similarly, an assumption that the first nucleotide is a G yields the candidate sequence GGTCG, and an assumption that the first nucleotide is a T yields the candidate sequence TACTT. Thus 4 candidate sequences are generated, each beginning with a different nucleotide assumed to be the first nucleotide in the sequence.

[00319] There is no requirement that the assumption must be made about the first nucleotide rather than one of the other nucleotides. For example, an assumption could equally well have been made about the identity of the fourth nucleotide, in which case the candidate sequences would have been generated by moving "backwards" along the template (i.e., in a 3' -^ 5' direction). For example, assuming that the fourth nucleotide is T means that the fourth dinucleotide must be TT; the third dinucleotide must be CT; the second dinucleotide must be AC; and the first dinucleotide must be CC. (Nucleotides are written in the 5 '->3' orientation although their identities are generated by moving from 3' ->5' in the sequence.) Alternately, an assumption can be made about any nucleotide in the middle of the sequence, and dinucleotide identities generated by moving both in the 5 '->3' and the 3 '-^5 directions. It will be appreciated that in the absence of an assumption about one of the nucleotides, the identity of each nucleotide remains completely undetermined since each position could be occupied by A, G, C, or T.

[00320] When using preferred collections of probe families, assuming the identity of any single nucleotide (e.g., the first nucleotide) generates one and only one candidate sequence. However, when less preferred collections of probe families are used it may be necessary to assume an identity for more than one nucleotide, i.e., assuming an identity for a first nucleotide does not entirely specify the rest of the sequence. For example, a less preferred collection of probe families may include a family with members whose defined sequences are AA and AC. In such a case, assuming that the first nucleotide is A leaves two possibilities for the second nucleotide. Sequencing using less preferred collections of probe families is

discussed further below. It will be appreciated that if the constrained portions consist of noncontiguous nucleotides, the approach described above can still be used with minor modifications.

[00321] Sequence Identification by Comparing Candidate Sequences with Known Sequences

[00322] Generally if the candidate sequences of the extended duplexes were determined, as described above, corresponding candidate sequences for the region of the template to be sequenced are obtained by taking their complements. In some instances, the candidate sequences themselves will provide enough information to achieve an objective. For example, if the purpose of sequencing is simply to rule out certain sequence possibilities, then comparing the candidate sequences with those possibilities would be sufficient. The candidate sequences shown in Figure 30 would allow a determination that the region being sequenced was not part of a poly A tail, for example. A longer sequence could confirm that the region being sequenced was not part of a vector.

[00323] In many instances it will be desirable to explicitly determine the correct sequence. According to a preferred embodiment of the invention the correct sequence is identified by comparing the candidate sequences for the region of the template to be sequenced with a set of known sequences. The set of known sequences may, for example, be a set of sequences for a particular organism of interest. For example, if human DNA is being sequenced, then the candidate sequences can be compared with the Human Draft Genome Sequence. See the web site having URL www.ncbi.nih.gov/genome/guide/human/for a guide to publicly available human genome sequence resources As another example, if nucleic acid derived from an infectious agent (e.g., a bacterium or virus isolated from a subject) is being sequenced, a database containing sequences of variant strains of that bacterium or virus can be searched. Many such organism-specific databases, containing either complete or partial sequences, are known in the art, and more will become available as sequencing efforts accelerate. Some representative examples include databases for the mouse (see, e.g., the web site having URL www.ncbi.nlm.nih.gov/genome/seq/MmHome.html), human immunodeficiency virus (see, e.g., the web site

having URL hiv-web.lanl.gov/content/hiv-db/mainpage.html), malaria species Plasmodium falciparum (see, e.g., the web site having URL www.tigr.org/tdb/edb2/pfal/htmls/index.shtml), etc. Of course it is not necessary to use an organism-specific set of sequences. A database such as GenBank (web site having URL www.ncbi.nlm.nih.gov/Genbank/), which contains sequences from a wide variety of organisms and viruses, can be searched. The database need not even contain any sequences from the organism or virus from which the template was derived. In general, the sequences can be genomic sequences, cDNA sequences, ESTs, etc. Multiple sequences can be searched.

[00324] Simply performing the search may be sufficient to achieve an objective. For example, if viral nucleic acid is isolated from a patient, comparing the candidate sequences with a set of known sequences of that virus can determine that the viral nucleic acid either does or does not contain sequences from that virus, even if the matching sequence is never examined. The existence of a match would confirm that the patient is infected with the virus, while lack of a match would indicate that the patient is not infected with the virus. [00325] In certain embodiments the set of known sequences contains a narrower range of sequences, which may be specifically tailored to the purpose for which the sequencing is performed. Thus information about the nucleic acid being sequenced may be used to select the set of known sequences. For example, if it is known that the template represents sequence of a particular gene, the known sequences may represent different alleles of a gene, mutant and wild type sequences at a given locus of interest, etc. It may only be necessary to compare the candidate sequences with a single known sequence to determine which of the candidate sequences is correct. For example, in certain embodiments of the invention the template is obtained by amplifying DNA that contains a region of interest (e.g., using primers that flank the region of interest). The region of interest may encompass a site at which mutations or polymorphisms may exist, e.g., mutations or polymorphisms that are associated with a particular disease. If it is known that the template represents a sequence from a particular region of interest, then the candidate sequences need only be compared with a single reference sequence for that region, e.g., a wild type or mutant form of the sequence. In other words, if part or all of the sequence

of the template is known, it may not be necessary to perform a comparison with a plurality of known sequences. Instead, a candidate sequence that comprises all or part of the known sequence is selected as correct. For example, mutations in the BRCAl and BRCA2 genes are known to be associated with an increased risk of breast cancer, and there is significant interest in determining whether subjects carry such mutations. If it is known that the template comprises sequence from the BRCAl gene, e.g., if primers flanking a region of interest that encompasses a portion of the gene were used to produce a clonal population of templates, then the candidate sequences need only be compared against the wild type or mutant BRCAl sequence to determine the correct sequence. [00326] In the more general case, comparing the candidate sequences with the set of known sequences will identify any known sequences that are similar to any of the candidate sequences. Provided that the candidate sequences are of sufficient length, the likelihood that a database will contain sequences that is identical to or closely resemble more than one of the candidate sequences are very small. In other words, if the candidate sequences are long enough, it is unlikely that more than one of them will be represented in the set of known sequences. The candidate sequences are compared with any sequences that are considered to be a "match". It will typically be desirable to set a threshhold for the degree of identity required to establish that a match exists. For example, a known sequence may be considered to be a match if a candidate sequence and the known sequence are at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 99%, or even 100% identical. Typically the percent identity will be evaluated over a window of at least 10 nucleotides in length, e.g, 10-15 nucleotides, 15-20 nucleotides, 20-25 nucleotides, 25-30 nuclotides, etc. The length of the window may be selected according to a variety of different criteria including, but not limited to, the number of sequences in the plurality of known sequences, the identity or source of the plurality of known sequences, etc. For example, if a candidate sequence is being compared against sequences in a large database such as GenBank, it may be desirable to use a longer length than if a database containing fewer sequences is used. In certain embodiments of the invention sequences are compared across a plurality of different windows, not necessarily adjacent to one another. Preferably the

combined length of the windows is at least 10 nucleotides in length, e.g, 10-15 nucleotides, 15-20 nucleotides, 20-25 nucleotides, 25-30 nuclotides, etc. In some instances multiple sequences in the set of known sequences may match. The sequences may, for example, represent homologous genes found in the same organism as that from which the template was derived, homologous genes from different organisms, pseudogenes, cDNA and genomic sequences, etc. [00327] In general, the candidate sequence that most closely resembles a sequence in the set of known sequences is selected as correct. Alternately, e.g., if there is reason to believe that the sequencing method may have been subject to a high error rate it may be preferable to select the corresponding sequence from the database as correct. For example, if the error rate is known to be above a predetermined threshhold it may be preferable to select a sequence from the database as the correct sequence.

[00328] The length required in order to ensure that the likelihood of matches being found for multiple candidate sequences will depend on a variety of considerations including, but not limited to, the particular set of known sequences, the threshhold for accepting matches, etc. In general, a sequence of length -25-26 nucleotides would only be represented once in the genome of a typical organism. Therefore generating candidate sequences of approximately this length is sufficient to identify the correct sequence. In general, the candidate sequence should be at least 10 nucleotides in length, preferably at least 15, at least 20 nucleotides in length, e.g., between 20-25, 25-30, 30-35, 35-40, 45-50, or even longer. [00329] Sequence Identification by Comparing a First Set of Candidate Sequences with a Second Set of Candidate Sequences

[00330] In certain embodiments of the invention decoding is performed by generating a first ordered list of probe families using a first collection of probe families encoded according to a first encoding, generating a first set of candidate sequences therefrom and then generating a second ordered list of probe families from the same template using a second collection of probe families encoded according to a second encoding and generating a second set of candidate sequences therefrom. The newly synthesized DNA strand is removed from the template between the two sequencing reactions, or a template of identical sequence is

sequenced using the second collection of probe families. The sets of candidate sequences are compared. It will be appreciated that regardless of which collection of probe families is used, one of the candidate sequences will be the correct sequence while the others are not correct (or are at best partially correct). Thus every set of candidate sequences will contain the correct sequence, but in most cases the other candidate sequences in any given set candidate sequences will differ from those found in another set of candidate sequences. Therefore, by simply comparing the two sets of candidate sequences, the correct sequence can be determined. It is not necessary to generate candidate sequences of equal length using the two differently encoded collections of probe families. In preferred embodiments of the invention the candidate sequences generated using the second collection of probe families can be as short as 2 nucleotides or, equivalently, the ordered list of probe families generated using the second collection of probe families can be as short as 1 element (i.e., a single cycle of ligation and detection).

[00331] Figures 31A - 31C show an example of candidate sequence generation and decoding using two distinguishably labeled preferred probe families. Figure 31A shows a preferred collection of probe families encoded according to a first encoding. Figure 3 IB shows generation of 4 candidate sequences from the ordered list of probe families Yellow, Green, Red, Blue (which could be represented as "2314" in which Red = 1, Yellow = 2, Green = 3, and Blue = 4), of which the correct sequence is assumed to be CAGGC (shown in bold). Figure 31C shows a preferred collection of probe families encoded according to a second encoding. Since the first dinucleotide in the template is CA, the uppermost probe in the Yellow probe family will ligate to the extendable terminus in the first cycle of extension. This results in the following set of candidate sequences for the first dinucleotide: CA, TC, GG, AT. Among the candidate sequences generated using the first collection of probe families, only the sequence CAGGC begins with any of these dinucleotides. Therefore it must be the correct sequence. In general, it is preferred that the first and second collections of probe families should fulfill the following criteria: When the first and second collections of probe families are compared, (i) 3 of the 4 probes in each of the probe families in the first collection should be assigned to a new probe family in the second collection; and (ii) each of the 3 reassigned probes

should be assigned to a different probe family in the second collection.

[00332] Using a Known Nucleotide Identity to Decode an Ordered List of Probe Families [00333] As described above, candidate sequences can be generated by assuming an identity for a single nucleotide in the extended duplex or template. Depending on the specific probe family collection used, it will generally be necessary to generate at least 4 candidate sequences. However, generation of multiple candidate sequences can be avoided if the identity of at least one nucleotide in the template (and therefore also in the extended duplex) is known. In that case, it will only be necessary to generate a single candidate sequence. The method for generating the candidate sequence is identical to that described above. The identity of the at least one nucleotide in the template may be determined using any sequencing method including, but not limited to sequencing methods A, primer extension from an initializing oligonucleotide using a set of distinguishably labeled nucleotides and a polymerase, etc. It will be appreciated that one or more nucleotides in the template can first be sequenced using a sequencing method other than sequencing method AB, and the initializing oligonucleotide and any extension products can then be removed, and the same template subjected to sequencing using sequencing methods AB (or vice versa). [00334] Another approach is to simply sequence a template that contains one or more known nucleotides of known identity in addition to a portion whose sequence is to be determined. For example, the portion of the template between the region to which the initializing oligonucleotide binds and at which the unknown sequence begins can include one or more nucleotides of known identity. By subjecting this portion of the template to sequencing methods AB, the identity of one or more nucleotides in the sequence will be predetermined and can thus be used to generate a single candidate sequence, which will be the correct sequence. [00335] The methods described above therefore comprise steps of (i) assigning an identity to a nucleotide in the template adjacent to a nucleotide of known identity by determining which identity is consistent with the identity of the known nucleotide and the possible sequences of the constrained portion of the probe whose proximal nucleotide ligated opposite the nucleotide adjacent to the nucleotide of known identity; (ii) assigning an identity to a succeeding nucleotide by

determining which identity is consistent with possible sequences of the constrained portion of the probe whose proximal nucleotide ligated opposite the succeeding nucleotide; and (iii) repeating step (ii) until the sequence is determined. It is to be understood that these steps are equivalent to performing the same steps on the extended duplex since there is a precise correspondence between the extended duplex and the region of the template to be sequenced.

[00336] Sequencing With Less Preferred Probe Families

[00337] Less preferred collections of probe families may be used to perform sequencing methods AB in a similar manner to the way in which preferred collections of probe families are used. However, the results may differ in a number of respects. For example, certain portions of the sequence may be fully identified from the candidate sequences without the need for additional information. Figure 32 shows an example of sequence determination using a less preferred collection of probe families encoded as shown in Figure 28. Sequence determination generally proceeds as described for preferred collections of probe families. The template of interest has the sequence "GCATGA", which results in "12341" as the ordered list of probe families. Assuming that the nucleotide at position 1 is A yields "ACATGA" as a candidate sequence. However, unlike the case with the preferred collections of probe families, there are two possibilities for the second nucleotide since the label "1" is associated with two different dicnucleotides that have A as the first nucleotide, i.e., "AA" and "AG". Thus assuming that the nucleotide at position 1 is A yields "ACATGC" as a second candidate sequence. Assuming that the nucleotide at position 1 is G yields "GCATGA" as a candidate sequence and also yields "GCATGC" as a candidate sequence. Since the label "1" is not associated with any dinucleotides that have C or T at position 1, no candidate sequences beginning with "C" or "T" are generated. Figure 32 shows the 4 candidate sequences aligned with each other. It will be observed that the middle 4 nucleotides of all the candidate sequences are CATG. Therefore, the correct sequence must include CATG at positions 2 - 5. If only these nucleotides are of interest, there is no need to perform further decoding steps. [00338] As mentioned above, collections of probe families need not consist of four different probe families but can consist of any number greater than 2, up to 4<N>, where N is the length of the

constrained portion. However, if fewer than 4 families are used it may be necessary to generate more than 4 candidate sequences, while if more than 4 probe families are used additional labels will be required. For these and other reasons collections consisting of 4 probe families are preferred.

[00339] Sequence Identification by Comparing Candidate Sequences with Each Other [00340] In certain embodiments of the invention part or all of a sequence of interest may be determined by comparing candidate sequences with each other. In general, such a comparison may not be sufficient to determine which of the candidate sequences is correct across its entire length. However, if two or more of the candidate sequences are identical or sufficiently similar over a portion of the sequences, this information may be sufficient to explicitly identify the sequence of nucleotides in the template within that portion as described above.

[00341] If desired, the template can be sequenced one or more additional times using alternatively encoded probe families to yield additional portions with an identifed sequence. These portions can be combined to assemble a sequence of a desired length. [00342] Error Correction Using Probe Families. It is often desirable to sequence multiple templates that represent all or part of the same DNA sequence and to align the sequences. If the templates contain only part of a region of interest, a longer sequence is then obtained by assembling overlapping fragments. For example, when sequencing the genome of an organism , typically the DNA is fragmented, and enough fragments are sequenced so that each stretch of DNA is represented in several (e.g., 4-12) different fragments. Computer software for assembling overlapping sequences into a longer sequence is known to one of skill in the art.

[00343] When conventional sequencing methods are used, it is frequently the case that multiple fragments align perfectly over a region except that one of the fragments (referred to as an anomalous fragment) differs from the others at a single position within the region. Determining whether the isolated difference represents a sequencing error or whether a genuine difference (e.g., a single nucleotide polymorphism) exists at the position the can be problematic.

[00344] The invention provides novel methods of performing error checking using sequencing methods AB. According to the method, templates comprising fragments that represent the same stretch of DNA are sequenced using a collection of distinguishably labeled probe families as described above, resulting in an ordered list of probe families for each template. The ordered lists of probe families are aligned. If several lists align perfectly over a predetermined length, e.g, 10, 15, 20, or 25 or more elements in the lists, except for one list that differs at a single position from the other fragments, the difference is ascribed to a sequencing error. If an actual polymorphism exists, the ordered probe list generated from the anomalous fragment will differ at two or more adjacent positions from the ordered probe lists generated from the other fragments.

[00345] For example, applying sequencing methods AB using a preferred collection of probe families that uses encoding 4 in Table 1 to a template comprising the sequence 5'- C AGACGAC AAGT AT AATG-3' yields the following ordered list of probe families: "23324322132444142", as shown below:

2332432213244414 2 CAGACGACAAGTATAATG

[00346] If there is an actual SNP (e.g., CAGACGAGAAGTATAATG, in which the underlined nucleotide represents the polymorphic site), it results in changes in two consecutive elements in the list: 23324333.132444142, in which underlining indicates the change that occurs as a result of the SNP. The correspondence between the ordered list of probe families and sequence containing a SNP is shown below: 233243^3132444 142

CAGACGAGAAGTATAATG

[00347] However, an error in identifying the label associated with a ligated extension probe results in a single error in the ordered list of probe families and a change in the resulting candidate sequence from that point forward. For example, an error in determining the label associated with the 7th ligated extension probe 2332433.2132444142 (in which the underlined number represents the misidentif[iota]ed label) changes the resulting candidate sequence to CAGACGAGTTC AT ATTAC, in which the underlined portion indicates the change that occurs as a result of the sequencing error.

The correspondence between the ordered list of probe families and the sequence is shown below: 233243^2 132444 142

CAGACGAGTTCATATTAC

[00348] When using a 3 base, 4 label scheme, a fragement that contains a SNP results in 3 consecutive differences in the ordered list of probe families for the anomalous fragment, while a sequencing error results in only 1 difference. For example, when the collection of probe families encoded as shown in Figure 29 is used, an ordered list of probe family identities for the sequence CAGACGACAAGTATAATG is shown below:

2322224132412244 CAGACGACAAGTATAATG

[00349] An anomalous fragment containing a SNP, e.g., CAGACGAGAAGTATAATG, would result in an ordered list of probe families that differs at 3 consecutive positions relative to ordered lists generated from fragments that do not contain the SNP, as shown below:

2322213332412244 CAGACGAGAAGTATAATG

[00350] A sequencing error would result in only a single difference in the ordered list of probe families and would result in a completely different generated candidate sequence from the point of the error forward.

[00351] Thus when an ordered list of probe families generated from a fragment (an anomalous fragment) aligns with ordered lists of probe families generated from other fragments that represent the same stretch of DNA but differs from the other ordered lists at a single isolated position, it is likely that the ordered list containing the difference represents a sequencing error (misidentification of a probe family). When an ordered list of probe families generated from a fragment (an anomalous fragment) aligns with ordered lists of probe families generated from other fragments that represent the same stretch of DNA but differs from the other ordered lists at 2 or more consecutive positions, it is likely that the anomalous fragment contains a SNP. Preferably the aligned portions of the ordered lists of probe families are at least 3 or 4 elements in length, preferably at least 6, 8, or more elements in length. Preferably the aligned portions are at least 66% identical, at least 70% identical, at least 80% identical,

at least 90% identical, or more, e.g., 100% identical. [00352] Similarly, when a candidate sequence for a fragment aligns with candidate sequences for other fragments that represent the same stretch of DNA over a first portion of the sequence but differs substantially from candidate sequences for other fragments over a second portion of the sequence, is it likely that a sequencing error occurred. When a candidate sequence for a fragment aligns with candidate sequences for other fragments that represent the same stretch of DNA over two portions of the sequence but differ at a single position, it is likely that the anomalous fragment contains a SNP. Preferably the aligned portions of the candidate sequences are at least 4 nucleotides in length. Preferably the aligned portions are at least 66% identical, at least 70% identical, at least 80% identical, at least 90% identical, or more, e.g., 100% identical.

[00353] The invention therefore provides a method of distinguishing a single nucleotide polymorphism from a sequencing error comprising steps of: (a) sequencing a plurality of templates using sequencing methods AB, wherein the templates represent overlapping fragments of a single nucleic acid sequence; (b) aligning the sequences obtained in step (a); and (c) determining that a difference between the sequences represents a sequencing error if the sequences are substantially identical across a first portion and substantially different across a second portion, each portion having a length of at least 3 nucleotides. The invention further provides a method of distinguishing a single nucleotide polymorphism from a sequencing error comprising steps of: (a) obtaining a plurality of ordered lists of probe families by performing sequencing methods AB using a plurality of templates that represent overlapping fragments of a single nucleic acid sequence; (b) aligning the ordered lists of probe families obtained in step (a) to obtain an aligned region within which the lists are at least 90% identical; and (c) determining that a difference between the ordered lists of probe families represents a sequencing error if the lists differ at only one position within the aligned region; or (d) determining that a difference between the ordered lists of probe families represents a single nucleotide polymorphism if the lists differ at two or more consecutive positions within the aligned region. [00354] Delocalized Information Collection

[00355] As is well known in the art, a "bit" (binary digit)

refers to a single digit number in base 2, in other words, either a 1 or a zero, and represent the smallest unit of digital data. Since a nucleotide can have any of 4 different identities, it will be appreciated that specifying the identity of a nucleotide requires 2 bits. For example, A, G, C, and T could be represented as 00, 01, 10, and 11, respectively. Specifying the name of a probe family in a preferred collection of distinguishably labeled probe families requires 2 bits since there are four distinguishably labeled probe families.

[00356] In most conventional forms of sequencing, and in sequencing methods A, each nucleotide is identified as a discrete unit, and information corresponding to one nucleotide at a time is gathered. Each detection step acquires two bits of information from a single nucleotide. In contrast, sequencing methods AB acquire less than two bits of information from each of a plurality of nucleotides in each detection step while still acquiring 2 bits of information per detection step when a preferred collection of probe families is used. Each probe family name in an ordered list of probe families represents the identity of at least 2 nucleotides in the template, with the exact number being determined by the length of the sequence determining portion of the probes. For example, consider the ordered list of probe families obtained from the sequence 5'- C AGACGAC AAGTAT AATG-3 ' using a collection of probe families encoded according to encoding 4 in Table 1 :

2332432213244 4142 CAGACGACAAGTATAATG

[00357] Probe family 2 is the first probe family in the list since the dinucleotide CA is one of the specified portions present in probes of probe family 2. Probe family 3 is the second probe family in the list since the dinucleotide AG is one of the specified portions present in probes of probe family 3. As mentioned above, since there are 4 probe families, each probe family identity represents 2 bits of information. Thus each detection step gathers 2 bits of information about 2 nucleotides, resulting in an average of 1 bit of information from each nucleotide.

[00358] The invention therefore provides a method for determining a sequence, wherein the method comprises multiple cycles of extension, ligation, and detection, and wherein the detecting step comprises simultaneously acquiring an average of two bits of information from

each of at least two nucleotides in the template without acquiring two bits of information from any individual nucleotide. The invention further provides a method for determining a sequence of nucleotides in a template polynucleotide using a first collection of oligonucleotide probe families, the method comprising the steps of: (a) performing sequential cycles of extension, ligation, detection, and cleavage, wherein an average of two bits of information are simultaneously acquired from each of at least two nucleotides in the template during each cycle without acquiring two bits of information from any individual nucleotide; and (b) combining the information obtained in step (a) with at least one bit of additional information to determine the sequence. In various embodiments of the invention the at least one bit of additional information comprises an item selected from the group consisting of: the identity of a nucleotide in the template, information obtained by comparing a candidate sequence with at least one known sequence; and information obtained by repeating the method using a second collection of oligonucleotide probe families. [00359] Thus while the methods do not acquire 2 bits of information from individual nucleotides, an average of 2 bits of information is gathered from the template in each cycle, but in a delocalized manner when preferred collections of probe families are used. When using collections of 2 or 3 probe families, less than 2 bits of information are gathered during each cycle.

[00360] Delocalized information collection has a number of advantages including allowing the application of error checking methods such as those described above. In addition, since each nucleotide in the template is interrogated more than once in preferred embodiments, delocalized information collection can help avoid systematic biases in detecting fluorophores associated with particular nucleotides.

[00361] The probe families and collections of probe families described herein can be used in a variety of sequencing methods in addition to methods that involve successive cycles of extension, ligation, and cleavage of the probe. The invention also provides probe families and collections of probe families having the sequences and structures as described above, wherein the probes optionally do not contain a scissile linkage. For example, the probes can contain only phosphodiester backbone linkages and/or may not contain a trigger residue. In some embodiments of the invention the probe families

are used to perform sequecing using successive cycles of extension and ligation, but not involving cleavage during each cycle. For example, the probe families can be used in a ligation-based method such as that described in WO2005021786 and elsewhere in the art. To use the probe families in such a method, the label on the probe should be attached by a cleavable linker, e.g,. as disclosed in in WO2005021786, such that it can be removed without cleaving a scissile linkage of the nucleic acid. Such a method can be used to generate an ordered list of probe families, e.g., by performing multiple reactions in parallel or sequentially, using the probe families rather than the ligation casssettes described in WO2005021786, and then assembling the list of probe families. The list is decoded as described above. [00362] I. Kits

[00363] A variety of kits may be provided for carrying out different embodiments of the invention. Certain of the kits include extension oligonucleotide probes comprising a phosphorothiolate linkage. The kits may further include one or more initializing oligonucleotides. The kits may contain a cleavage reagent suitable for cleaving phosphororothiolate linkages, e.g., AgNO3 and appropriate buffers in which to perform the cleavage. Certain of the kits include extension oligonucleotide probes comprising a trigger residue such as a nucleoside containing a damaged base or an abasic residue. The kits may further include one or more initializing oligonucleotides. The kits may contain a cleavage reagent suitable for cleaving a linkage between a nucleoside and an adjacent abasic residue and/or a reagent suitable for removing a damaged base from a polynucleotide, e.g., a DNA glycosylase. Certain kits contain oligonucleotide probes that comprise a disaccharide nucleotide and contain periodate as a cleavage reagent. In certain embodiments the kits contain a collection of distinguishably labeled oligonucleotide probe families.

[00364] Kits may further include ligation reagents (e.g., ligase, buffers, etc.) and instructions for practicing the particular embodiment of the invention. Appropriate buffers for the other enzymes that may be used, e.g., phosphatase, polymerases, may be included. In some cases, these buffers may be identical. Kits may also include a support, e.g. magnetic beads, for anchoring templates. The beads may be functionalized with a primer for performing PCR amplification. Other optional

components include washing solutions; vectors for inserting templates for PCR amplification; PCR reagents such as amplification primers, padlock probes, thermostable polymerase, nucleotides; reagents for preparing an emulsion; reagents for preparing a gel, etc.

[00365] In certain preferred kits, fluorescently labeled oligonucleotide probes comprising phosphorothiolate linkages are provided such that probes corresponding to different terminal nucleotides of the probe carry distinct spectrally resolvable fluorescent dyes. More preferably, four such probes are provided that allow a one-to-one correspondence between each of four spectrally resolvable fluorescent dyes and the four possible terminal nucleotides of a probe.

[00366] The kits may contain oligonucleotides and/or vectors suitable for producing a paired-end or fragment library. The kits may contain one or more blocking oligonucleotides that are complementary common portions of template molecules that are members of the library.

[00367] An identifier, e.g., a bar code, radio frequency ID tag, etc., may be present in or on the kit. The identifier can be used, e.g., to uniquely identify the kit for purposes of quality control, inventory control, tracking, movement between workstations, etc.

[00368] Kits will generally include one or more vessels or containers so that certain of the individual reagents may be separately housed. The kits may also include a means for enclosing the individual containers in relatively close confinement for commercial sale, e.g., a plastic box, in which instructions, packaging materials such as styrofoam, etc., may be enclosed.

[00369] J. Parallel Sequencing and Automated Sequencing Systems

[00370] Macevicz discloses sequencing a single template species having a particular sequence. He does not discuss the possibility of performing his method in parallel to simultaneously sequence a plurality of templates having different sequences. The inventors have recognized that in order to efficiently perform sequencing in a high throughput manner, it is desirable to prepare a plurality of supports (e.g., beads), as described above, such that each support has templates

of a particular sequence attached thereto, and to perform the methods described herein simultaneously on templates attached to each support. In certain embodiments of this approach, a plurality of such supports are arrayed in or on on a planar substrate such as a slide. In certain embodiments the supports are arrayed in or on a semisolid medium such as a gel. The supports may be arrayed in a random fashion, i.e., the location of each support on the substrate is not predetermined. The supports need not be located at regularly spaced intervals or positioned in an ordered arrangement of rows and columns, etc. Preferably the supports are arrayed at a density such that it is possible to detect an individual signal from many or most of the supports. In certain preferred embodiments the supports are primarily distributed in a single focal plane. Multiple supports having templates of the same sequence attached thereto may be included, e.g., for purposes of quality control. Sequencing reactions are performed in parallel on templates attached to each of the supports.

[00371] Signals may be collected using any of a variety of means, including various imaging modalities. Preferably, for embodiments in which sequencing is performed on microparticles that are arrayed on a substrate (e.g., beads embedded in a semi-solid support positioned on a substrate) prior to detection, the imaging device has a resolution of 1 [mu]m or less. For example, a scanning microscope fitted with a CCD camera, or a microarray scanner with sufficient resolution could be used. Alternately, beads can be passed through a flow cell or fluidics workstation attached to a microscope equipped for fluorscence detection. Other methods for collecting signal include fiber optic bundles. Appropriate image acquisition and processing software may be used.

[00372] In certain embodiments of the invention sequencing is performed in a microfluidic device. For example, beads with attached templates may be loaded into the device and reagents flowed therethrough. Template synthesis, e.g., using PCR, can also be performed in the device. U.S. Pat. No. 6,632,655 describes an example of a suitable microfluidic device.

[00373] The invention provides a variety of automated sequencing systems that can be used to gather sequence information from a plurality of templates in parallel, i.e., substantially simultaneously. Preferably the templates are arrayed on a substantially planar substrate. Fig. 21 shows a photograph of one of the

inventive systems. As shown in the upper part of the photograph, the inventive system comprises a CCD camera, a fluorescence microscope, a movable stage, a Peltier flow cell, a temperature controller, a fluid handling device, and a dedicated computer. It will be appreciated that various substitutions of these components can be made. For example, alternative image capture devices can be used. Further details of this system are provided in Example 9.

[00374] It will be appreciated that the inventive automated sequencing system and associated image processing methods and software can be used to practice a variety of sequencing methods including both the ligation-based methods described herein and other methods including, but not limited to, sequencing by synthesis methods such as fluorescence in situ sequencing by synthesis (FISSEQ) (see, e.g., Mitra RD, et al., Anal Biochem., 320(I):55-65, 2003). As is the case for the ligation-based sequencing methods described herein, FISSEQ may be practiced on templates immobilized directly in or on a semi-solid support, templates immobilized on microparticles in or on a semi-solid support, templates attached directly to a substrate, etc.

[00375] One important aspect of the inventive system is a flow cell. In general, a flow cell comprises a chamber that has input and output ports through which fluid can flow. See, e.g., U.S. Pat. Nos. 6,406,848 and 6,654,505 and PCT Pub. No. WO98053300 for discussion of various flow cells and materials and methods for their manufacture. The flow of fluid allows various reagents to be added and removed from entities (e.g., templates, microparticles, analytes, etc.) located in the flow cell.

[00376] Preferably a suitable flow cell for use in the inventive sequencing system comprises a location at which a substrate, e.g. a substantially planar substrate such as a slide, can be mounted so that fluid flows over the surface of the substrate, and a window to allow illumination, excitation, signal acquisition, etc. In accordance with the inventive methods, entities such as microparticles are typically arrayed on the substrate before it is placed within the flow cell.

[00377] In certain embodiments of the invention the flow cell is vertically oriented, which allows air bubbles to escape from the top of the flow cell. The flow cell is

arranged such that the fluid path runs from bottom to top of the flow cell, e.g., the input port is at the bottom of the cell and the output port is at the top of the cell. Since any bubbles that may be introduced are buoyant, they rapidly float to the output port without obscuring the illumination window. This approach, in which gas bubbles are allowed to rise to the surface of a liquid by virtue of their lower density relative to that of the liquid is referred to herein as "gravimetric bubble displacement". Thus the invention provides a sequencing system comprising a flow cell oriented so as to allow gravimetric bubble displacement. Preferably the substrate having microparticles directly or indirectly attached thereto (e.g., covalently or noncovalently linked to the substrate) or immobilized in or on a semi-solid support that is adherent to or affixed to the substrate is mounted vertically within the flow cell, i.e., the largest planar surface of the substrate is perpendicular to the ground plane. Since in preferred embodiments the microparticles are immobilized in or on a support or substrate, they remain at substantially fixed positions with respect to one another, which facilitates serial acquisition of images and image registration.

[00378] Figs. 24A- J shows schematic diagrams of inventive flow cells or portions thereof, in various orientations. The inventive flow cells can be used for any of a variety of purposes including, but not limited to, analysis methods (e.g., nucleic acid analysis methods such as sequencing, hybridization assays, etc.; protein analysis methods, binding assays, screening assays, etc. The flow cells may also be used to perform synthesis, e.g., to generate combinatorial libraries, etc.

[00379] Fig. 22 shows a schematic diagram of another inventive automated sequencing system. The flow cell is mounted on a temperature-controlled, automated stage (similar to the one described in Example 9) and is attached to a fluid handling system, such as a syringe pump with a multi-port valve. The stage accommodate multiple flow cells in order to allow one flow cell to be imaged while other steps such as extension, ligation, and cleavage are being performed on another flow cell. This approach maximizes utilization of the expensive optical system while increasing the throughput.

[00380] The fluid lines are equipped with optical and/or conductance sensors to detect bubbles and to monitor reagent usage. Temperature control and sensors in the

fluidics system assure that reagents are maintained at an appropriate temperature for long term stability but are raised to the working temperature as they enter the flow cell to avoid temperature fluctuations during the annealing, ligation and cleavage steps. Reagents are preferably prepackaged in kits to prevent errors in loading.

[00381] The optics includes four cameras - each taking one image through one of four filter sets. In order to reduce the effects of photobleaching, the illumination optics may be engineered to illuminate only the area being imaged, to avoid multiple illumination of the edges of the fields. The imaging optics may be built from standard infinity-corrected microscope objectives and standard beam-splitters and filters. Standard 2,000 X 2,000 pixel CCD cameras can be used to acquire the images. The system incorporates appropriate mechanical supports for the optics. Illumination intensity is preferably monitored and recorded for later use by the analysis software. [00382] In order to rapidly acquire a plurality of images (e.g., approximately 1800 or more non-overlapping image fields in a representative embodiment), the system preferably uses a fast auto focus system. Auto focus systems based on analysis of the images themselves are well known in the art. These generally require at least 5 frames per focusing event. This is both slow and costly in terms of the extra illumination required to acquire the focusing images (increases photob leaching). In certain embodiments of the invention an alternate auto focusing system is used, e.g., a system based on independent optics that can focus as quickly as the mechanical systems can respond. Such systems are known in the art and include, for examples the focusing systems used in consumer CD players, which maintain sub-micron focusing in real time as the CD spins.

[00383] In certain embodiments of the invention the system is operated remotely. Scripts for implementing specific protocols may be stored in a central database and downloaded for each sequencing run. Samples can be barcoded to maintain integrity of sample tracking and associating samples with the final data. Central, real-time monitoring will allow quick resolution of process errors. In certain embodiments images gathered by the instruments will immediately be uploaded to a central, multi-terabyte storage system and a bank of one or more processor(s). Using tracking data from the central

database, the processor(s) analyze the images and generate sequence data and, optionally, process metrics, such as background fluorescence levels and bead density, in order, e.g., to track instrument performance. [00384] Control software is used to properly sequence the pumps, stage, cameras, filters, temperature control and to annotate and store the image data. A user interface is provided, e.g., to assist the operator in setting up and maintaining the instrument, and preferably includes functions to position the stage for loading/unloading slides and priming the fluid lines. Display functions may be included, e.g., to show the operator various running parameters, such as temperatures, stage position, current optical filter configuration, the state of a running protocol, etc. Preferably an interface to the database to record tracking data such as reagent lots and sample IDs is included. [00385] K. Image and Data Processing Methods

[00386] The invention provides a variety of image and data processing methods that may be implemented at least in part as computer code (i.e., software) stored on a computer readable medium. Further details are presented in Examples 9 and 10. In addition, in general, both sequencing methods A and B generally employ appropriate computer software to perform the processing steps involved, e.g., keeping track of data gathered in multiple sequencing reactions, assembling such data, generating candidate sequences, performing sequence comparisons, etc.

[00387] L. Computer-readable Media Storing Sequence Information [00388] In addition, the invention provides a computer-readable medium that stores information generated by applying the inventive sequencing methods. Information includes raw data (i.e., data that has not been further processed or analyzed), processed or analyzed data, etc. Data includes images, numbers, etc. The information may be stored in a database, i.e., a collection of information (e.g., data) typically arranged for ease of retrieval, for example, stored in a computer memory. Information includes, e.g., sequences and any information related to the sequences, e.g., portions of the sequence, comparisons of the sequence with a reference sequence, results of sequence analysis, genomic information, such as polymorphism information (e.g., whether a particular template contains a polymorphism) or mutation information, etc., linkage

information (i.e., information pertaining to the physical location of a nucleic acid sequence with respect to another nucleic acid sequence, e.g., in a chromosome), disease association information (i.e., information correlating the presence of or susceptibility to a disease to a physical trait of a subject, e.g., an allele of a subject), etc. The information may be associated with a sample ID, subject ID, etc. Additional information related to the sample, subject, etc., may be included, including, but not limited to, the source of the sample, processing steps performed on the sample, interpretations of the information, characteristics of the sample or subject, etc. The invention also includes a method comprising receiving any of the aforesaid information in a computer-readable format, e.g., stored on a computer-readable medium. The method may further include a step of providing diagnostic, prognostic, or predictive information based on the information, or a step of simply providing the information to a third party, preferably stored on a computer-readable medium. [00389] The following examples are provided for illustrative purposes and are not intended to limit the invention.

Example 1: Efficient Cleavage and Ligation of Phosphorothiolated Oligonucleotides

[00390] This example describes an experiment demonstrating efficient ligation and cleavage of extension oligonucleotides containing a 3'-S phosphorothiolate linkage. [00391] Materials and Methods [00392] Ligation Sequencing Protocol

[00393] Template Preparation. To demonstrate evaluate the potential of sequencing by cycled oligonucleotide ligation and cleavage and to explore the effect of variations in certain aspects of the method, two sets of model bead-based template populations were prepared. In preferred implementations, as described in the Examples, cycled oligonucleotide ligation and cleavage extends strands in the 3 '-^5' direction. Therefore, to evaluate ligation efficiencies, model templates were bound to beads at the 5 ' end and designed with the same binding region at the 3 ' end. One set was comprised of short (70bp) oligonucleotides bound to streptavidin-coated magnetic beads (1 micron) via a dual biotin moiety. Each of these short template populations were designed with an identical primer binding region (40 bp) and a unique sequence region (30 bp) at the 3 'end. The short oligonucleotide template

populations were termed ligation sequencing templates 1-7 (LST 1-7).

[00394] The second set of bead-based template populations were designed from long, PCR-generated DNA fragments (232-bp) derived by inserting 183-bp of spacer sequence (from a human p53 exon) into each template population. Templates were amplified with dual biotin-containing forward primers and reverse primers containing the same 30 base unique 3' end sequence as the short template populations. The templates were made single-stranded by melting off one of the strands with sodium hydroxide-containing buffer. These long template populations were designed to mimic the species generated from short- fragment paired-end libraries described in a copending patent application and were termed long-LSTI-7.

[00395] Primer Hybridization. 2.5 [mu]L of 100 [mu]M FAM-labeled primer was premixed with 100 [mu]L IX Klenow Buffer. This solution was added to a 30 [mu]L aliquot of magnetic beads (10<6>/[mu]L) with attached template after removal of the buffer, and the resulting solution was well mixed. After allowing template/primer hybridization to occur (hybridization reaction was carried out for 2 minutes at 65<0>C, 2 minutes at 4O<0>C and 2 minutes on ice), the primer/buffer was removed, and the beads were washed using 3X Wash IE buffer, and then resuspended in 300 [mu]L (10<6>/mL) in TENT buffer (containing 10 mM Tris, 2 mM EDTA, 30 mM NaOAc, and 0.01% Triton X-100).

[00396] Ligation 1. 2.5 x 10<6> LST7 beads with hybridized LigSeq-FAM were then incubated for 30 minutes at 37<0>C in a mixture containing 1 [mu]L of 100 [mu]M LST7-1 Nonamer, 4 [mu]L 5X T4 Ligase Buffer (Invitrogen), 14 [mu]L of H2O and 1 [mu]L of T4 Ligase (1 u/[mu]L, Invitrogen). [00397] Cleavage 1: The beads were then washed 3 times with 100 [mu]L of LSWashI (containing IX TE, 30 mM sodium acetate, 0.01% Triton XlOO); a 10 [mu]L-aliquot of this solution was removed and saved for analysis. The beads (IX) were then washed in 100 [mu]L of 30 mM sodium acetate. 50 [mu]L of 50 mM AgNO3 was added to this solution and the resulting mixture was incubated at 37<0>C for 20 minutes. AgNO3 was removed, and the beads were washed once in 100 [mu]L of 30 mM sodium acetate. The beads were then washed in 3 times with 100 [mu]L of LSWashI, resuspended in 90

[mu]L Wash (TENT buffer); and a 10 [mu]L-aliquot of this solution was removed and saved for analysis.

[00398] Ligation 2: After removal of the TENT buffer, the beads were resuspended in 14 [mu]L of H2O, and incubated at 37<0>C for 30 minutes with a mixture containing 1 [mu]L of 100 [mu]M LST7-5 Nonamer, 4 [mu]L of 5X T4 Ligase Buffer (Invitrogen) and 1 [mu]L of T4 Ligase (1 u/[mu]L, Invitrogen).

[00399] Cleavage 2: The beads were washed 3 times in 100 [mu]L of LSWashI (IX TE, 30 mM sodium acetate, 0.01% Triton XIOO), and resuspended in 45 [mu]L WashIE. A 15 [mu]L- aliquot of this mixture was removed and saved for analysis. The beads were then washed once with 100 [mu]L of 30 mM sodium acetate and resuspended in 5 [mu]L of 20 mM sodium acetate. 50 [mu]L of 50 mM AgNO3 was added to the beads and the mixture was incubated at 37<0>C for 20 minutes. After removal OfAgNO3, the beads were washed once with 100 [mu]L of 30 mM sodium acetate. The beads were then washed three times in 100 [mu]L of LSWashI, and resuspended in 30 [mu]L WashIE. A 20 [mu] L-aliquot of this mixture was removed and saved for analysis.

[00400] Results

[00401] The experiment will be better understood with reference to Figure 8. The upper section of Fig. 8 shows an overall outline of the experimental procedure. An initializing oligonucleotide (primer) was hybridized to a template (designated LST7), which was attached to a bead via a biotin linkage. The initializing oligonucleotide contained a 5' phosphate and was fluorescently labeled with FAM at its 3 ' end. Two 9mer (nonamer) oligonucleotide probes (1<st> cleavable oligo and 2<nd> cleavable oligo) were synthesized to contain an internal phosphorothiolated thymidine base (sT) (underlined). The first cleavable probe was ligated to the extendable terminus of the primer using T4 DNA ligase and was then cleaved using silver nitrate. Cleavage removed the terminal 5 nucleotides of the extension probe and generated an extendable terminus on the portion of the probe that remained ligated to the primer. The second cleavable probe was then ligated to the extendable terminus and was then similarly cleaved.

[00402] A fluorescent capillary electrophoresis gel shift

assay was used to monitor steps of ligation and cleavage. In this assay, the primer is hybridized to a template strand such that the 5 ' phosphate can serve as a ligation substrate for incoming oligonucleotide probes (the fluorophore serves as a reporter for mobility-based capillary gel electrophoresis). After each step an aliquot of beads was removed for analysis. Following ligation of oligonucleotide probes, the magnetic beads were collected using a magnet and the ligated species consisting of the primer and probe(s) ligated thereto was released from the template beads by heat denaturation and subjected to fluorescent capillary electrophoresis using an automated DNA sequencing instrument (ABI 3730) with labeled size standards (lissamine ladder; size range 15-120 nucleotides; appears as a set of orange peaks in chromatograms, see Fig. 8). In a typical gel shift, the potential peaks include, i) primer peaks (due to no extension or the lack of primer extension), ii) adenylation peaks (due to the attachment of an adenosine residue at the 5' end of a nonproductive ligation junction by the action of DNA ligase - see mechanism in Fig. 8F, see also Lehman, I.R., Science, 186:790-797, 1974), and iii) completion peaks (due to the attachment of an oligo probe). One benefit of using gel shift assays to evaluate ligation efficiency is that the areas under the peaks directly correlate with the concentration of each species.

[00403] Fig. 8A shows a control ligation performed using T4 DNA ligase and an exact match probe containing only phosphodiester linkages (shown to the left of Fig. 8A). Orange peaks represent size markers. The blue peak at the left indicates the position of the primer in the absence of ligation. Ligation of the exact match probe results in a shift to the left (arrow). Fig. 8B shows a ligation performed under the same conditions using a probe containing an internal thiolated T base (shown to the left of Fig. 8B). A shift identical to that observed with the control probe was seen (arrow). Bead-linked template populations containing the ligated phosphorothiolated probes were then incubated with silver nitrate to induce probe cleavage. Gel-shift analysis confirmed efficient cleavage by demonstration of a left-shifted, 4-bp cleavage product (Fig. 8C). The expected cleavage product is shown to the left of Fig. 8C. Cleaved bead-based template populations were then exposed to a second round of ligation and demonstrated productive ligation by the appearance of a right- shifted, 13-bp extension product (Fig. 8D). The

expected cleavage product is shown to the left of Fig. 8D. A second round of cleavage confirmed efficient multiple cleavage steps could be accomplished as demonstrated by the expected left-shifted, 8-bp cleavage product (Fig. 8E). These results demonstrate successful ligation and cleavage of probes containing phosphorothiolate linkages.

[00404] It is evident that ligation did not proceed to 100% completion in these experiments, although a greater dregree of completion was observed in other experiments using T4 DNA ligase (see below). While it is certainly desirable that the ligation proceed to completion it is not a requirement. For example, it is possible to effectively "cap" any unligated 5' ends by treating with a 5 '-phosphatase after the ligation step as described above. In that case, however, there would be a limit to the number of sequential ligations that could be performed, due to attrition of ligatable molecules. With a given number of sequential ligations, the read length will depend on the length of the probe remaining after each ligation/cleavage cycle and on the number of sequencing reactions, each followed by removal of the primer and hybridization of a primer that binds to a different portion of the primer binding site, that can be performed on a given template, also referred to as the number of "resets"). This argues for the use of longer probes with the cleavable linkage located towards the 5' end of the probe. In our experiments, hexamer probes lead to greater amounts of un- ligatable adenylation products than octamers and longer probes. Thus octamers and longer probes will ligate substantially to completion (see below). In addition, adding a fluorescent moiety to the 5 ' end of a hexamer probe seems to reduce the efficiency of ligation, whereas adding a fluorescent moiety to an octamer probe has little or no effect. For these reasons, use of octamers or longer probes is considered preferable.

[00405] Additional experiments (described below) have demonstrated ligation and cleavage of probes containing phosphorothiolate linkages and degeneracy-reducing nucleotides; 3' end specificity and selectivity of ligated extension probes; in-gel ligation and cleavage; sequential cycles of primer hybridization and removal with minimal loss of signal; 100% fidelity for T4 or Taq ligase for 3 '-^5' extensions; and 4-color spectral resolvability of ligated extension probes. An automated system for performing the methods has been

constructed.

Example 2: Efficient Cleavage and Ligation of Phosphorothiolated Oligonucleotides Containing Degeneracy-Reducing Nucleotides [00406] A competing consideration to probe length, however, is the fidelity of the extended oligonucleotide and its effect on subsequent ligation efficiency. The fidelity of T4 DNA ligase has been shown to decrease rapidly following the 5 th base after the junction (Luo et al, Nucleic Acid Res., 24: 3071-3078 and 3079-3085, 1996). If mismatches are introduced at the 5' side of a new ligation junction, the ligation efficiency may be reduced by attrition, however, no dephasing or increase in background signal will be generated (a major obstacle encountered in polymerase-based sequencing by synthesis methods).

[00407] Probe sets should preferably be capable of hybridizing to any DNA sequence in order to permit de novo sequencing of uncharacterized DNA. However, the complexity of a labeled probe set grows exponentially with the length and number of 4-fold degenerate bases. In addition, a complex probe set is more challenging to synthesize while maintaining approximately equal representation of all probe species, and is harder to purify. It also requires a higher concentration of probe mixture to maintain a constant concentration of each species. One way to manage this complexity is to use nucleotides incorporating universal bases, such as deoxyinosine, at certain positions instead of 4-fold degenerate bases.

[00408] Twelve octanucleotide probes were designed with 4-fold degenerate bases (N; equimolar amounts of A, C, G, T) and the universal base inosine (I) at various positions within the octamer (inosine is capable of bi-dentate hydrogen bonding with any of the four canonical bases in B-DNA; the order of stabilities of inosine base pairs is I:C > LA > I:T ~ I:G). One purpose for evaluating these probe designs was to determine how low an octamer complexity could be achieved while still supporting efficient ligation in the presence of inosine bases.

[00409] In initial studies, several oligonucleotide probes were ligated to bead-based templates (long-LSTI) using T4 DNA ligase. Upon ligation, the fluorophore-labeled primer (3'FAM Primer) shifts right in proportion to the

amount of oligonucleotide probe ligated. Probe design NI8-9 showed the highest level of completion, with >99% of the primer population shifting right due to efficient ligation of the probe (see Figure 9). These reactions were conducted at 25[deg.]C; when the reaction temperature was increased to 37[deg.]C, ligation was somewhat less efficient and the completion rates were more variable.

[00410] Closer examination of the data indicated that probes with fewer inosine bases within the first five nucleotides on the 3' side of the junction (underlined) showed higher ligation efficiencies. To investigate further and to evaluate potential sequence context effects on ligation efficiencies, four oligonucleotide probe designs with only a single inosine residue within the first five bases 3' of the ligation junction were screened across all templates. Figure 10 demonstrates ligation completion as evaluated using the gel-shift assay with selected probe compositions on multiple templates using T4 DNA ligase. Data from these initial experiments demonstrated that ligation efficiency, and hence completion, is variable and sequence-dependent when inosine residues are placed within the first five 3 ' positions of the ligation junction (underlined). Efficient ligation of octamers was observed consistently, however, with oligonucleotide probe design NI8-9, as demonstrated here with >99% completion on all templates tested.

[00411] While not wishing to be bound by any theory, this data (including the presence of adenylated intermediates) support the conclusion that unfavorable inosine base pairs within the core DNA binding site for T4 DNA ligase destabilize the DNA protein complex sufficiently to reduce enzyme binding and subsequent ligation. An interesting question, however, was whether such destabilizing inosine base pairs would affect the fidelity of the ligated oligonucleotide probes.

Example 3: Fidelity of Probe Ligation

[00412] Bacterial NAD-dependent ligases, such as Tag DNA ligase, have been reported to have high sequence fidelity across ligation junctions, with mismatches on the 3 ' side having essentially no nick-closure activity, but mismatches on the 5 ' side being tolerated to some degree (Luo et al, Nucleic Acid Res., 24: 3071-3078 and 3079-3085, 1996). T4 DNA ligase, on the other hand,

has been reported to be somewhat less stringent, allowing mismatches on both the 3 '- and 5 '-sides of the junction. It was therefore of interest to evaluate the fidelity of probe ligation with T4 DNA ligase in comparison to Taq DNA ligase in the context of our system.

[00413] We developed two methods to evaluate the sequence fidelity of ligated oligonucleotides using standard ABI sequencing technology. The first method was designed to clone and sequence ligation products. In this method, ligation extension products were attached to adapter sequences, cloned and transformed into bacteria. Individual colonies were picked and sequenced to provide a quantitative assessment of the mismatch frequency at each position across the ligation junction. The second method was designed to sequence of ligation products directly. In that approach, single-stranded ligation products were denatured from bead-based templates and sequenced directly using a complementary primer. Positions with low accuracy display multiple overlapping peaks in the resulting sequence traces, providing a qualitative assessment that is indicative of the sequence fidelity at that position.

[00414] The first method was used to assess the relative fidelity of probe ligation by T4 and Taq DNA ligases. A single bead-based template population (LST1) was hybridized to a universal sequencing primer, which was used as an initializing oligonucleotide. Solution-based ligation reactions were then performed in the presence of a degenerate oligonucleotide probe (N7A, 3 [Alpha]NNNNNNN5', 2000 pmoles) at 37[deg.]C for 30 minutes with either T4 DNA ligase (15U per IxIO<6> beads) or Taq DNA ligase (60U per IxIO<6> beads) (Fig. 11, panel A). The ligation products were cloned and sequenced to evaluate the positional fidelity of each DNA ligase on the 3' side of its ligation junction (Positions 1-8) (Fig. 11, panels B and C). The results indicated that T4 DNA ligase has essentially the same level of fidelity across the first 5 positions as Taq DNA ligase, but lower fidelity in positions 6-8. These results were further substantiated by subsequent cloning experiments that evaluated DNA sequences across ligation junctions of all seven templates (LST 1-7) for three degenerate, inosine- containing probe designs (3'-NNNNNIII-5\ 3'-NNNNNINI-5', and 3'-NNNINNNI-5'). The studies confirmed that T4 DNA ligase has low sequence fidelity across ligation junctions at positions

6-8, however, high fidelity was exhibited across the first 5 positions in all templates tested (data not shown).

[00415] The direct sequencing method was used to assess the fidelity of T4 DNA ligase with degenerate, inosine-containing probes. Oligonucleotide probes were evaluated at 25[deg.]C and 37[deg.]C in ligation reactions that contained T4 DNA ligase and bead-based templates. Oligonucleotide probe ligation efficiencies were evaluated using a gel-shift assay (Fig. 12, panel A). Direct sequencing of the ligation reactions using an ABI3730xl DNA Analyzer was conducted to assess the fidelity of T4 DNA ligase in oligonucleotide probe ligation (Fig. 12, panel B). Ligation of an exact match oligo probe and two representative degenerate inosine-containing oligo probes (NI8-9 and NI8-11) gave >99% completion and a very low frequency of mismatches (absence of multiple peaks in the sequencing traces). The data suggest that probes which are efficiently ligated also give high sequence fidelity.

[00416] In additional experiments, a single bead-based template population (LSTI) was hybridized to a universal sequencing primer that contained 5 'phosphates, which was used as an initializing oligonucleotide. Solution-based ligation reactions were performed at 37C for 30 minutes with T4 DNA ligase (IU per 250,000 beads) in the presence of a degenerate, inosine-containing oligonucleotide probe (3'NNNNNiii5', 3'NNNNNiNi5', or 3'NNNiNNNi5', 600 pmoles). Ligation products were cloned and colonies were picked and sequenced. Sequence fidelity was determined by calculating the number of clones represented for each postion across the ligation junction. Results are tabulated in Fig. 12, panels C-F. These studies demonstrate that 3 '-^5' ligation of degenerate, inosine-containing probes with T4 DNA ligase has high-level fidelity in the first 1-5 positions.

Example 4: In-Gel Ligation and Cleavage

[00417] The initial experiments to explore, develop and optimize methods for cycled oligonucleotide ligation were conducted using bead-based templates in solution, as described above. In a second set of experiments, ligation and cleavage were performed on bead-based templates that were embedded in polyacrylamide gels on slides.

[00418] Slides were prepared by mixing millions of beads, each having a clonal population of single-stranded DNA templates attached thereto, with 5% polyacrylamide and allowing polymerization to occur on a glass slide. A Teflon(R) mask was used to enclose the bead- containing polyacrylamide solution. Fig. 14 (top) shows a fluorescence image of a portion of a slide on which beads with an attached template, to which a Cy3-labeled primer was hybridized, were immobilized within a polyacrylamide gel. (This slide was used in a different experiment, but is representative of the slides used here.) Fig. 14 (bottom) shows a schematic diagram of a slide equipped with a Teflon mask to enclose the polyacrylamide solution.

[00419] Reactants were introduced into slides either by manual dipping of slides into appropriate solutions or by placing the slides in an automated, laminar flow cell. Initial studies confirmed that efficient in-gel ligation could indeed be performed on templates attached to beads immobilized in a polyacryamide matrix on such slides. In the experiment shown in Figure 15, single-stranded DNA template beads were immobized on slides containing acrylamide and DATD. Following polymerization, a universal, 3'fluorophore- labeled, 5 'phosphorylated primer (Seq Primer) was diffused into the gel and allowed to hybridize (panel A). Slides were washed to remove unbound seq primer, overlaid with a ligation cocktail that contained T4 DNA ligase (10U) and an oligonucleotide probe, and incubated at 37[deg.]C for 30 minutes. Slides were then incubated in a buffer containing sodium periodate (0.1M) to digest the acrylamide polymer and to release the bead-based template populations. Ligated products were denatured from the template strand by heat, collected and analyzed using the gel shift assay described above. In-gel ligation reactions performed in the absence of T4 DNA ligase demonstrated a single peak representative of unligated sequencing primer (panel B). Ligation reactions performed with octamer probes in the presence of T4 DNA ligase demonstrated efficient in gel oligonucleotide ligation with >99% of bead-based template populations efficiently ligated (panel C).

Example 5: Four-Color Detection

[00420] To maximize detection efficiency, it is desirable to employ a set of oligonucleotide probes with distinct labels corresponding to each possible base addition

product. This was modeled in our automated sequencing instrument equipped with appropriate excitation and emission filters, as outlined in Fig. 15. Three sets of octamer probes were designed to address issues of probe specificity and selectivity. The first set included four octamers, complementary to four unique template populations, with different 3 ' bases and 5' dye labels. The second set included seven unique octamers with unique 3' bases and 5' dyes. The third set corresponded to a probe design with four degenerate, inosine- containing octamers, each having a unique 3' end base identified by a different 5' dye label.

[00421] To confirm four-color spectral identity, probe set #1 was employed to detect four unique template populations (see Figure 16). Slides were prepared containing four, unique single-stranded template populations attached to beads, which were embedded in polyacrylamide (panel A). Each bead had a clonal population of templates attached thereto. A universal sequencing primer containing 5 ' phosphates was hybridized, in situ, and ligation reactions were performed using an oligonucleotide probe mixture that contained four unique fluorophore probes (Cy5, CAL 610, CAL 560, FAM; 100 pmoles each) and T4 DNA ligase (IOU/slide). Slides were incubated at 37[deg.]C for 30 minutes and washed to remove unbound probes. The slides were imaged in bright light to create a white light base image (panel B) and with fluorescence excitation using the four bandpass filters (FITC, Cy3, TxRed, and Cy5). Fluorescence image capture was conducted pre- and post-ligation. Individual populations were pseudocolored (panel C) and the spectral identity of image values were plotted and confirm minimal signal overlap (panel D).

Example 6: Demonstration of Ligation Specificity and Selectivity in Gels

[00422] To confirm 3 'end specificity, probe set #2, was used to interrogate a single template population (see Fig. 17). Slides were prepared with a beads having a single template population (LST 1.T) attached thereto embedded in a polyacrylamide gel, and were hybridized, in situ, with a universal sequencing primer (panel A). In-gel ligation reactions were conducted with T4 DNA ligase (IOU/slide) using an oligonucleotide probe mixture comprised of four 5 ' end-labeled probes that differed only by a single 3 ' base. Slides were incubated

at 37[deg.]C for 30 minutes and washed to remove unbound probe populations. Slides were imaged in white light to create a base image (panel B) and with fluorescence excitation using four bandpass filters (FITC, Cy3, TxRed, and Cy5). Fluorescence image capture conducted pre- and post-ligation confirmed a single FAM -based probe population (blue spots) present following in-gel ligation with T4 DNA ligase, with no spectral overlap (panels C, D). This data demonstrates that probe specificity with T4 DNA ligase is stringent and is determined by the first 3 'end base of the ligation junction.

[00423] To further substantiate 3 'end specificity and selectivity, probe set #2 was used to identify a mixture of bead-based template populations containing single base differences and present in different amounts. Slides were prepared with mixtures of beads each having one of four template populations, each with a single nucleotide polymorphism (LSTI; A, G, C or T), attached thereto, as indicated in panel A of Figure 18. The beads were embedded in a polyacrylamide gel on the slide. Bead-based template populations were used at various different frequencies, as outlined in panel D. Slides were hybridized, in situ, with universal sequencing primers. In-gel ligation reactions were conducted using T4 DNA ligase (10U/slide) and an oligonucleotide probe mixture containing equimolar amounts (100 pmoles, each) of four 5' end-labeled probes that differed only by a single 3' base. Slides were incubated at 37[deg.]C for 30 minutes and washed to remove unbound probe populations. Slides were imaged in white light to create a base image (panel B) and with fluorescence using four distinct bandpass filters (FITC, Cy3, TxRed, and Cy5). Individual probe images were overlaid and pseudocolored (panel C). Fluorescent images were enumerated using bead-calling software. The results are presented in panel D and confirm that observed ligation frequencies (Obs) correlated with the expected frequencies (Exp). The data demonstrate high probe specificity and probe selectivity after ligation in the presence of multiple templates and demonstrate the capability of detecting single nucleotide polymorphisms (SNPs), i.e., alterations that occur in a single nucleotide base in a stretch of genomic DNA in different individuals of a population, by ligation.

Example 7: Demonstration of Ligation Specificity and

Selectivity in Gels Using Four- color Degenerate Inosine-containing Extension Probes [00424] Another set of experiments were conducted, using probe set #3, to evaluate the specificity and selectivity of probe ligation using four-color degenerate, inosine-containing oligonucleotide probe pools. Results are presented in Figure 19. Bead-based slides were prepared as described above, but with four, unique single-stranded template populations present on beads in different amounts and were then hybridized, in situ, with a universal sequencing primer (panel A). In-gel ligation reactions were performed in the presence of T4 DNA ligase (IOU/slide) using probe pools consisting of octamers designed with five degenerate bases (N; complexity 4<5> = 1024), two universal bases (I, inosine), and single known nucleotide at the 3' end corresponding to a specific 5' fluorophore (G-Cy5, A-CAL 610, T-CAL560, A-FAM; 600 pmoles each). Slides were incubated at 37[deg.]C for 30 minutes and washed to remove unbound probe populations. Slides were imaged in white light to create a base image (panel B) and with fluorescence using four bandpass filters (FITC, Cy3, TxRed, and Cy5). Individual probe images were overlaid and pseudocolored (panel C). Fluorescent images were enumerated and the frequencies of each ligation product tabulated using bead-calling software (panel D); spectral scatter plots of unprocessed raw data and filtered data representing the top 90% of bead signal values are shown in panel E. The data demonstrate that the observed ligation frequencies (Obs) correlated with the expected frequencies (Exp) based on the known concentrations of each template. This confirms that degenerate and universal base-containing probe pools can be used with T4 DNA ligase to afford specific and selective in-gel ligation.

Example 8: Demonstration of Repeated Cycles of Hybridization and Removal of

Initializing Oligonucleotide in Gel

[00425] Experiments conducted on templates immobilized in a gel on a microscope slide mounted in an automated flow cell (see below) confirmed that multiple cycles of annealing and stripping an initializing oligonucleotide could be applied to templates attached to beads embedded in gels on slides with minimal signal loss. A 44 base fluorescently labeled initializing

oligonucleotide was used. As shown in Figure 20, minimal signal loss occurred over 10 cycles. The initializing oligonucleotide is referred to as a primer in Fig. 20. As indicated above, one of the major drawbacks of polymerase based sequencing-by-synthesis procedures is the propensity for both positive and negative dephasing to occur on individual template strands. Positive dephasing occurs when nucleotides are misincorporated in a growing strand, hence causing the base sequence of that particular strand to run ahead of the sequence obtained from the remaining templates and to be out of phase by n+1 base calls. Negative dephasing, which is more common, occurs when strands are not fully extended, resulting in background base calls that run behind the growing strand (n-1). The ability to efficiently strip extension products and to "reset" templates by hybridizing a differentially positioned initializing oligonucleotide allows very long read lengths with little to no signal attrition.

Example 9: Automated Sequencing System

[00426] This example describes a representative inventive automated sequencing system that can be used to gather sequence information from one or more templates. Preferably the templates are located on a substantially planar substrate such as a glass microscope slide. For example, the templates may be attached to beads that are arrayed on the substrate. A photograph of the system is presented in Fig. 21. The system is based on an Olympus epi- fluorescence microscope body (mounted sideways) with an automated, auto-focusing stage and CCD camera. Four filter cubes in a rotating holder permit four-color detection at a variety of excitation and emission wavelengths. A flow cell with peltier temperature control, which can be opened and closed to accept a substrate such as a slide (with a gasket to seal around the edge of an area containing a semi-solid support such as a gel), is mounted on the stage. The vertical orientation of the flow cell is an important aspect of the inventive system and allows air bubbles to escape from the top of the flow cell. The cell can be completely filled with air to eject all reagents prior to each wash step. The flow cell is connected to a fluid handler with two 9-port Cavro syringe pumps, which allow delivery of 4 differentially labeled probe mixtures, cleavage reagent, any other desired reagents, enzyme equilibration buffer, wash buffer and air to the flow cell through a single port. The

operation of the system is completely automated and programmable through control software using a dedicated computer with multiple I/O ports. The Cooke Sensicam camera incorporates a 1.3 megapixel cooled CCD though cameras having lesser or greater sensitivity could also be used (e.g., 4 megapixel, 8 megapixel, etc., can be used). The flow cell utilizes a 0.25 micron stage, with a 1 micron feature size.
Example 10: Image Acquisition and Processing Methods

[00427] This example describes representative methods for acquiring and processing images from arrays of beads having labeled nucleic acids attached thereto. Accurate feature identification and alignment are important for reliable analysis of each acquired image. The features are identified by first discarding all but the most intense pixels for each bead. The pixel values for a given image are plotted in a histogram; pixels corresponding to background are discarded and the remaining pixel values are sorted. In uniform images, where all the beads are roughly the same intensity, the algorithm eliminates the bottom 80-90% of pixel values. Pixels having values in the top 10-20% are then scanned to identify those at a local maximum in a 4 pixel radius. The average intensity in that region as well as the average intensity of the perimeter are then recorded. These values form a normal distribution and pixels whose values fall outside that distribution are then removed. The percentage of pixels initially ignored, the size of the circular region, and the cutoff values that eliminate possible beads in the normal distribution are all parameterized and can be tuned if necessary. Alignment is accomplished by creating feature matrices for each image in the alignment set. The resulting matrices are then searched for the most frequent x,y coordinate offsets to identify the optimal alignment.

[00428] Bead images are collected in the Cy5 channel (corresponding to the sequencing primer) prior to extension probe addition. These images are used to create a feature map marking both positional coordinates and raw signal intensities as fluorescent units (RFU values) for each bead. For each subsequent duplex extension, an image set is acquired both before and after the Cy3 -labeled nucleotides are added. These images are aligned to the original Cy5 images and RFU values are then assigned to each of the beads and recorded. A baseline correction is applied by subtracting

the difference of intensities between the unlabeled (pre-extension) and labeled (fluorescent-addition) images of each base addition. These baseline-subtracted values are then normalized by the intensity found in the Cy5 image for each feature to form the basis by which a bead is considered to have been extended or not (i.e., a bead is considered to be extended if duplexes attached to the bead were extended). Using these methods thousands of features per image with -1,300 images per slide can be analyzed to afford an analysis of 5-100 million template species per experimental run. The algorithms have been designed so that they can be easily ported from MATLAB to C+ at a later date for further efficiency enhancements. Example 11: Bead Alignment and Tracking and Sequence Decoding

[00429] This example describes representative methods for processing images from arrays of beads having labeled nucleic acids attached thereto and for sequence determination from the acquired data.

[00430] Image analysis starts by convolving the image using a zero-integral circular top- hat kernel with a diameter matched to the bead size. This will automatically normalize the background to zero while identifying the centers of individual beads through local maxima. The maxima are located and those which are isolated from other local maxima are used as alignment points. These alignment points are computed for each image in a time-series. For each pair of images, the alignment points are compared and a displacement vector is computed based on the average displacement of all the common alignment points. This provides pair-wise image displacements with sub-pixel resolution.

[00431] For N images, there are N*(N-l)/2 pairwise displacements, but only N-l of these are independent since the rest can be calculated from the independent set. For example, measuring the displacements between images 1 and 2 and between images 1 and 3 implies a displacement between images 2 and 3. If the measured displacement between images 2 and 3 is not the same as the implied displacement, then the measurements are inconsistent. The magnitude of this inconsistency can be used as a metric to gauge how well the alignment algorithm is working. Our initial tests show inconsistencies that are generally less than 0.1 pixel in each dimension (see Figure 23).

[00432] Once a time-series of images is aligned, there are two ways to track the individual beads.

If the bead density is low with most of the beads not touching another bead, the optical center-of-mass of each individual bead can be identified and a region around the bead integrated to compute the bead intensity. If the bead density is so high that most of the beads touch, then it is not possible to identify individual beads by a dark background band around them. However, with all the images aligned to sub-pixel resolution, it is possible to identify pixels belonging to the same bead by computing the correlation, in time, of adjacent pixels. Highly correlated pixel pairs can be confidently assigned to the same bead. A similar technique has been applied to lane tracking in DNA sequencing gels with good results (Blanchard, A. P. Sequence-specific effects on the incorporation of dideoxynucleotides by a modified T7 polymerase, California Institute of Technology, 1993). Once the beads have been tracked through the entire 4-color time-series, the sequence is decoded by knowing which color corresponds to which 3 '-most base of the probe oligonucleotides.

Example 11: Throughput Calculations

[00433] In general, the throughput of the sequencing system is defined primarily by the number of images that the machine can generate per day and the number of nucleotides (bases) of sequence data per image. Since the machine is preferably designed to keep the cameras constantly busy, calculations are based on 100% camera utilization. In implementations in which each bead is imaged in 4 colors to determine the identity of one base, either 4 images by one camera, 2 images by 2 cameras, or one image by 4 cameras can be used. Four-camera imaging permits dramatically higher throughputs than the other options, and preferred systems utilize that approach.

[00434] Our initial tests show that a pixel density of 50 pixels per bead, representing 5.4 square microns, provides a comfortable density for standard image analysis. By using a 4 megapixel CCD camera (now commonplace), a single CCD frame can image ~80,000 beads (based on our current image data). Capturing four images with separate cameras and moving to the next field on the flow cell will take no longer than 1.5

seconds. If 75% of the beads yield useful information, we will be able to collect data from approximately 80,000 beads * 0.75 /1.5 = 40,000 bases/sec of raw sequence data.

[00435] One significant issue in maintaining 100% camera utilization is matching the time it takes to perform one cycle of ligation/cleavage chemistry with the time required to image the entire flow cell. A reasonable estimate for the time taken by a cycle of extension, cleavage, and ligation is 1 Vi hours (5,400 seconds). That 5,400 seconds will accommodate 1,800 image fields, or an area of about 15 mm X 45 mm, which is a comfortable size for a flow cell. A conservative estimate of the throughput of the system utilizing four cameras is 40,000 bases per second with a 15 mm X 45 mm flow cell. This is equivalent to approximately 2,000 ABI3730xl sequencing machines, based on a throughput of 28 runs per day with -650 base read lengths (20 bases /sec), which we have achieved using these machines. A 2.5 fold increase in bead density, to 200,000 per image enables an overall increase in throughput to 100,000 bases per second, approximately equivalent to 5,000 ABI3730xl machines. The total output per day at this throughput level is -8.6 Gb per day, so the time required to complete a 12X human genome sequence would be -4.2 days. [00436] It is noted that the inventive sequencing methods described herein may be practiced using a variety of different sequencing systems, image capture and processing methods, etc. See, e.g., U.S. Pat. Nos. 6,406,848 and 6,654,505 and PCT Pub. No. WO98053300 for discussion.

Example 12: Methods for Preparing Microparticles for Template Synthesis Thereon

[00437] This example describes a protocol preparation of microparticles (in this example, magnetic beads) with amplification primers attached thereto so that a template can be amplified (e.g., by PCR) so as to result in a clonal population of template molecules attached to each microparticle. In general, amplification beads have one primer needed in the clonal PCR reaction attached thereto. This primer can be covalently coupled or, for example, biotin labeled and bound to streptavidin on the bead surface. Beads can be used in a standard PCR reaction (e.g., in wells of a microtiter plate, tubes, etc.), in an emulsion PCR reaction as described in Example

13, etc., to obtain beads having clonal populations of template molecules attached thereto. [00438] Materials
Ix TE: 1OmM Tris (pH 8) ImM EDTA

Ix PCR buffer: (ThermoPol Buffer, NEB)

2OmM Tris-HCl (pH 8.8)

1OmM KCl

1OmM (NH4)2SO4

2mM MgSO4

0.1% Triton X- 100

IM Betaine (add only for Ix PCR-B buffer)

Ix Bind & Wash Buffer 5 mM Tris HCl (pH 7.5) 0.5 mM EDTA 1 M NaCl

DNA Capture Primer (20-mer, 500 [mu]M stock)

Dual Biotin-(HEG)5-PI: 5 '-Dual Biotin-(HEG)5-CTA AGG TAG CGA CTG TCC TA- 3'

(HEG)5 = Hexaethylene glycol linker, an 18 carbon containing spacer, one of a number of different spacer moieties that could be used. Including a spacer is useful, e.g., to raise the PI primer portion of the oligo off the surface of the bead. Any of the primers described herein may incorporate such spacer moieites. 6

Dynal stock magnetic beads (1 [mu]m diameter) = 10mg/ml (7-12 x 10 beads/[mu]l). [00439] Methods

6

1. Remove 50 [mu]l beads (-450 x 10 beads).

2. Add 200 [mu]l Ix TE buffer, mix well. Separate with magnet.

3. Wash Ix with 200 [mu]l Ix TE buffer. Separate with magnet.

4. Resuspend in 100 [mu]l B/W buffer.

5. Add 3[mu]l of PI oligo (500[mu]M stock = 1500 pmol).

6. Rotate at RT for >30 minutes.

7. Wash 3x with 200 [mu]l lx TE buffer.

8. Resuspend in 50[mu]l (initial volume) lx TE buffer.

9. Store DNA capture beads at 4C or place on ice prior to use. Beads should be used within 1 week (beads will tend to clump at storage times >1 week).

Example 13: Methods for Performing PCR on Microparticles in an Emulsion

[00440] This example describes methods that can be used to perform PCR on microparticles in an emulsion to produce microparticles with clonal templates attached thereto. The microparticles (DNA beads in the nomenclature used below) are first functionalized with a first primer (Pl). A second primer (P2) is present in the aqueous phase, where the PCR reaction occurs. If desired, a low concentration of Pl may also be included, e.g., (20-fold less) in the aqueous phase. Doing so allows a rapid build-up of templates in the aqueous phase, which are substrates for additional amplification. As Pl is depleted in solution, the reaction is driven towards utilization of Pl attached to the microparticles. P1 P2 degenlO is an oligonucleotide template (lOObp) that has sequences that hybridize to Pl and P2 to afford amplifcation by PCR and a stretch of approx 10 degenerate bases (incorporated during oligonucleotide synthesis) that give the oligonucleotide population a complexity of 4<10>.

I. Emulsion Protocol (1 [mu]m beads) 1. Prepare oil phase:

Span 80 (7%)

Tween 80 (0.4%)

Prepared in Light Mineral Oil

Use only freshly made oil phase

Total Oil Phase = 450 [mu]l 9

2. Prepare aqueous phase: (Estimated to produce 2 x 10 droplets, 115 fL per droplet)

Reagent (stock) ([mu]l) per reaction Final dmo 156.0 -

MgCh Buffer (10X) 32.0 IX dNTP (IOO mM ea) 11.3 3.5 mM each

MgCl2 (IM) 7.3 23mM

Betaine (5 M) 32.0 0.5M

Pl (Primer I)(IO [mu]M) 1.6 11.25pmole

P2 (Primer 2)(200 [mu]M) 40.0 5625 pmole

Pl P2 degenl0 (100 pM) 6.6 5.9xlO<[Lambda]>7/ul

DNA Beads (8M/[mu]l) 25.0 150M/emulsion

Platinum Taq (5 U/[mu]l) 9.0 0.28U/ul

Total aqueous volume = 320 [mu]l

Final reaction = 255 [mu]l aqueous phase : 450 [mu]l oil phase

3. Transfer aqueous phase tube to ice until addition to emulsion.

4. Add 450 [mu]l oil phase to a 2 ml cryovial.

5. Place cryovial UPRIGHT into foam adapter attached to IKA vortex. Set vortex to 2500 rpm.

6. Aliquot aqueous phase (3 aliquots, 85 [mu]l each = 255 [mu]l) to shaking oil phase. Add monodispersed aqueous phase to the agitating 2 ml cryovial by placing the tip into tube and slowly dispensing the aqueous phase from the tip into the shaking oil phase. Repeat addition 2x with the remaining aqueous phase.

7. Continue shaking emulsion for 24 minutes at 2500 rpm.

8. Transfer -100 [mu]l aliquots of the emulsion into a 96-well plate (total = 4 wells). Also, aliquot remaining aqueous phase (65 [mu]l) into a separate well for a solution-based PCR control reaction. Seal plate and cycle as outlined in next section.

II. Emulsion Amplification (1 [mu]m beads)

1. PCR cycling parameters for l[mu]m bead emulsions (with primer Tms = 62C):

Program: DTB-PCR 94C, 2 min n=l

94C, 15s

57C, 30s n=100

7OC, 60s

55C, 5 min n=l

1OC, for arbitrary time period

2. Cycling time is ~6 hours.

3. Observe emulsions following cycling. Successful emulsions will appear uniformly amber in color with no observable separated aqueous phase. Emulsions that "break" (fall out of solution) will have a distinct aqueous phase at the bottom of the tube. Avoid collecting this phase, as this population of beads will not be clonal.

4. Assess post-cycled emulsions using bright field microscopy. Remove a 2 [mu]l aliquot of the cycled emulsion and drop onto a glass slide. Overlay emulsion sample with a 22 x 60mm glass coverslip.

5. View emulsions using the 2OX objective. Beads should preferably be monodispersed, with the majority of droplets containing single beads.

NOTE: If the emulsion sample contains a high number of multi-bead droplets, pool emulsion reactions into a single 1.5 ml eppendorf tube and spin at 6000 rpm for 15 seconds. Remove the bead suspension that accumulates at bottom of tube. This population will be comprised of both free beads and multi-bead droplets that are heavier than single-bead droplets and thus will settle to the bottom of the tube following a brief spin. This bead population is not clonal and should therefore be avoided prior to subsequent processing. Re-evaluate emulsion by repeating Steps 4 and 5 to confirm integrity of single bead-containing droplets in emulsion sample.

6. Disrupt (break) emulsions using the protocol outlined

in the next section.

III. Emulsion Break and Melt (1 [mu]m beads)

Bead Break Wash (BBW) Buffer

2% Triton X-100 2% Tween 20; 10 mM EDTA

Melt Solution IQOmM NaOH

Ix TE: 1OmM Tris (pH 8) ImM EDTA

Ix Bind & Wash (BAV) Buffer 5 mM Tris-HCl (pH 7.5)
0.5 mM EDTA

1 M NaCl

1. Pool each emulsion set (4 aliquots) into a single 1.5 ml eppendorf tube.

2. Add 800 [mu]l BBW buffer. Break emulsions by vortexing reaction tube for 10 seconds.

3. Spin at 8000 rpm for 2 min.

4. Remove top 800 [mu]l (mainly oil phase). DNA beads will be pelleted at the bottom of tube.

5. Add 800 [mu]l BBW, vortex and spin at 8000 rpm for 2 min. Remove top 600 [mu]l.

6. Wash an additional 2x with 600[mu]l Ix TE using a magnet to exchange each wash.

8. Add 50 [mu]l Melt solution to bead pellet and resuspend sample by vigorous pipetting. Incubate beads in Melt solution for 5 minutes at room temperature, flicking tube intermittently.

9. Place tube in magnet to remove Melt solution. Wash Ix with 100 [mu]l Melt solution to ensure complete removal of second strand.

10. Wash bead pellet 2x with Ix TE and resuspend into 20 [mu]l TE buffer for storage at 4C or 20[mu]l Ix B/W buffer if next step is enrichment. If beads appear to be clumped, exchange into Ix PCR-B buffer.

11. Continue with enrichment protocol (optional).

Example 14: Methods for Enriching for Microparticles Having Clonal Template

Populations Attached Thereto

[00441] This example describes a method for enriching for microparticles on which template amplification has successfully occurred in, e.g, in a PCRemulsion. The method makes use of larger microparticles that have a capture oligonucleotide attached thereto. The capture oligonucleotide comprises a nucleotide region that is complementary to a nucleotide region present in the templates.

I . Emulsion Enrichment (l[mu]m)

A. Preparation of enrichment beads (capture entities)

Enrichment beads:

Spherotech streptavidin-coated polystyrene beads (~6.5um)

Bead stock (0.5% w/v): 33,125 beads/[mu]l

Per Protocol: (33,125 beads/[mu]l) (800[mu]l) = 26.5 x 10 beads

Usage:

119 million beads per emulsion - estimate of emulsion clonality (2%): ~3M template- positive beads per emulsion. Add 2-3 enrichment beads per estimated template-positive emulsion bead = 10 million enrichment beads per emulsion reaction.

Enrichment oligonucleotide (capture agent): P2-enrich (35-mer, Tm=73C)

5 '-Dual biotin-18-carbon spacer-ttaggaccgttatagttaggtgatgcattaccctg 3'

(or)

P2-enrich (e.g., up to 35-mer, Tm=52C)

5 '-Dual Biotin-18-carbon spacer-ggtgatgcattaccctg 3'

Glycerol solution - 60% (v/v) 6 ml glycerol 4 ml

nuclease-free H2O 1. Remove 800 [mu]l of beads and exchange into B/W buffer by centrifugation at 13,000 rpm for 1 minute. Wash lx with 500 [mu]l B/W buffer and resuspend into 100 [mu]l B/W buffer.

2. Add 20 [mu]l enrichment oligo (500 [mu]M stock = 10,000 pmoles per rxn).

3. Rotate bead reaction at room temperature for 1 hour.

4. Wash beads 3x using 500 [mu]l lx TE buffer. Pellet beads between washes by centrifugation at 13,000 rpm for 1 minute.

5. Resuspend beads into 25 [mu]l B/W buffer. Concentration = IM enrichment beads/[mu]l.

NOTE: Pooling four enriched emulsion populations into 20-30 [mu]l lx B/W buffer yields ~ 4OM template-positive beads. Multiple slides can then be run.

B. Enrichment procedure

1. Add 20 [mu]l of the enrichment beads to the tube containing emulsion-derived beads (20 [mu]l). Resuspend bead mixture with gentle pipetting (or use ratios that give rise to 2-3 enrichment beads for every estimated template-positive emulsion bead).

2. If using enrichment beads coated with the biotinylated P2-enrich primer, incubate bead mixture at 65C for 2 minutes. Remove tube to ice for 10 minutes.

NOTE: Initial experiments have suggested that using enrichment beads containing primer sequences used for the 100-cycle PCR (e.g., P2PCR) may be less efficient at enrichment due to the ability to enrich for beads containing primer:dimer species driven to bead in droplets that were devoid of template. If using enrichment beads loaded with the P2-enrich primer described above, incubate bead mixture at 5OC for 2 minutes due to the reduced Tm of this shorter primer.

3. Overlay bead mixture into 1.5 ml eppendorf tube containing 300 [mu]l 60% Glycerol solution.

4. Centrifuge at 13,000 rpm for 1 minute.

5. Following spin, negative beads will pellet to bottom of

tube. Enrichment beads containing attached template beads will float to the top of the glycerol phase. Collect top- phase bead population and transfer to a clean 1.5 ml eppendorf tube.

NOTE: Beads pelleted to the bottom of the tube (beads with no template) can be washed and analyzed using a magnet following the same wash regimen as outlined for template-positive beads.

6. To beads pulled from top phase, add 1 ml nuclease-free H20 to dilute the glycerol concentration. Resuspend bead mixture using gentle pipetting. Spin at 13,000 rpm for 1 minute.

7. Following spin, remove supernatant and wash 2x using 100 [mu]l TE.

8. Add 100 [mu]l Melt solution to the washed bead pellet. Rotate tube for 5 minutes at room temperature.

9. Add an additional 100 [mu]l Melt solution and isolate template beads using a magnet.

10. Remove non-magnetic enrichment beads by washing 2x using 100 [mu]l TE and a magnet to pull DNA beads away from enrichment beads.

11. Resuspend template beads into 10-20 [mu]l lx TE. If beads appear to be clumped, dilute into lx PCR-B buffer.

12. Template-containing beads can be pooled with other enriched populations and loaded onto slides as described in the next Example. Example 15: Methods for Preparing a Microparticle Array Immobilized in or on a

Semi-Solid Support

[00442] This example describes preparation of slides on which microparticles having templates attached thereto are immobilized (e.g., embedded) in a semi-solid support located on the slide. Such slides may be referred to as polony slides. The semi-solid support used in this example is polyacrylamide. One of the protocols employs methods that trap polymerase molecules in the vicinity of templates to enhance amplification. Preparation of Slides

A. Glass Slides: Bind-Silane Treatment

Bind-Silane facilitates the attachment of the acrylamide gel to the glass slide surface. Slides should be pre-treated with Bind-Silane prior to use.

NOTES:

** Store Bind-Silane solution in chemical hood.

** Bind-Silane is an irritant. Work in a chemical when preparing solution.

** Ensure that the stock Bind-Silane solution has not expired.

** Trynot to touch surfaces of slides while transferring to and from racks. Prepare Bind-Silane solution:

1. In a 1 -L plastic container add: 1 L dH2O, 1 Stir bar

Add 220ul concentrated Acetic Acid (to generate pH 3.5) Add 4 ml Bind- Silane reagent Mix solution for > 15minutes using stir plate.

Treat slides:

2. Load slides (facing the same direction) into upside-down plastic 384-well plates.

3. Wash slides by rinsing with dH2O, drain well. 4. Rinse with 100% ethanol, drain well.

5. Rinse again with dH2O, drain well and place in tissue culture hood with vent and UV light running. Allow washed slides to dry (~30 min).

6. Place plate into a plastic container and cover slides with Bind-Silane solution.

7. Allow solution and slides to react for 1 hour. Agitate container intermittently to ensure even coating of Bind-Silane to glass.

8. Following incubation, rinse slides 3x with dH2O.

9. Rinse lx with 100% ethanol, drain well.

10. Allow slides to dry thoroughly prior to use.

11. Store Bind-Silane-treated slides in dessicator.

B. Acrylamide-based Slides (small mask)

- Non-Trapping Protocol

1. Place all reagents on ice. Add the following chilled reagents to a 1.5 ml eppendorf tube:

Reagent amt ([mu]l)

2 slides 1 slide

Ix TE 13 6.5

Beads (1-3M, diluted in Ix TE) 10 5

Rhinohide 1 0.5

40% Acrylamide:Bis (19:1, F/S) 5 2.5

TEMED (5%, in Ix TE) 2 1

APS (0.5%, made fresh) 3 1.5

Total 34 [mu]l 17 [mu]l Pipet mixture vigorously to distribute beads.

Load 17 [mu]l per slide under a glass coverslip.

Polymerize upside down at room temperature for 60 minutes.

Remove coverslip with a clean razorblade.

Soak slide and wash 2x in IE buffer for 15 minutes (to remove unbound beads).

Slides with embedded beads can be stored at 4C in wash IE.

2. Hybridize fluorophore-labeled sequencing primer to embedded bead population.

Equilibrate slide from wash IE to Ix PCR-B buffer by dipping briefly into Coplin jar containing Ix PCR-B buffer.

3. In a 1.5 ml eppendort tube, add 1-6 [mu]l (100 [mu]M stock) primer to 99 [mu]l lx PCR buffer. Over the acrylamide matrix, drop 100 [mu]l primer solution and overlay with a glass coverslip or sealing gasket.

Page 142 of 165

,s Mail Label No.: EL 993747543 US [iota]l l.DOC 4. Hybridize primer to embedded beads by heating slide using <DEVIN> program (65C for 2 minutes, slow anneal to 30C). Wash slide 2x for 2 minutes in wash IE. Slide is ready to be subjected to ligation based sequencing. - Trapping Protocol

1. ssDNA template beads are prepared at lM/[mu]l. [Prepare polony slides with 4-5M beads per slide].

2. Resuspend bead mixture into 30[mu]l lx PCR buffer.

3. Add lul sequencing primer (lOO[mu]M stock), mix well. 4. Heat to 65 C for 2 min.

5. Remove to ice for 5 min.

6. Wash 3x with 80[mu]l lx TE

7. Remove all so ln using a magnet.

8. Add reagents as outlined below:

Reagent amt([mu]l)

2slides

lx buffer 1.5

1Ox buffer 2.0

High cone. (HC) enzyme 16.0

40% Acrylamide:Bis (19:1, F/S) 14.4

Rhinohide 2.0

TEMED (5%, in lx TE) 2.0

APS (0.5%, made fresh) 1.5

Total 39.4 [mu]l

Pipet mixture to distribute beads. Load 17 [mu]l per slide under a glass coverslip.

9. Polymerize, preferably upside down, e.g., using <Pol-I> cycling profile on MJ Research Tetrad PCR machine.

10. Remove coverslip with a clean razorblade. Soak slide and wash 2x in IE buffer for 10 min. (to remove unbound beads). 11. Polony slides are ready to be subjected to ligation-based sequencing.

12. Polony slides with embedded beads can be stored in gaskets at 4C in wash IE.

Example 16: Methods for Preparing a Microparticle Array Attached to a Solid

Support [00443] This example describes preparation of slides on which microparticles having templates attached thereto are attached to a solid support.

[00444] 1. Glass slides prepared with polymer tethers with reactive NHS are stored at - 2OC. (Slide H, Product No. 1070936; Schott Nexterion; Schott North America, Inc., Elmsford, NY)

Page 143 of 165 ?81 [00445] 2. In the presence of dessicant, equilibrate slides to room temperature before use.

[00446] 3. Wash slides in 5Omls lx PBS (30OmM sodium phosphate, pH 8.7) for 5 minutes. Repeat washes 2x.

[00447] 4. Remove slide from solution and cover with an adhesive gasket (to allow sample loading).

[00448] 5. In a separate tube, aliquot 100-400 million protein-coated or DNA-coated beads into lx PBS, pH 8.7. The DNA can be, e.g., DNA templates for sequencing. The

DNA can include, e.g., an amine linker for reaction with NHS.

[00449] 6. Wash bead sample 3x with lx PBS, pH 8.7 by buffer exchange. [00450] 7. Resuspend beads into

125ml lx PBS, pH 8.7.

[00451] 8. Load bead solution into the slide gasket to evenly coat slide surface.

[00452] 9. Enclose slides in a dark chamber and allow reaction to incubate for 1-2 hrs at room temperature.

[00453] 10. Following incubation, remove unbound bead solution and transfer slide to 5Omls lx TE (1OmM Tris, ImM EDTA, pH 8).

[00454] 11. Wash slide 5x using 5Omls lx TE with constant agitation for 15 minutes per wash.

[00455] 12. Slides can be stored in lx TE at 4C for several weeks.

[00456] 13. If desired, bead populations can be assessed by bright field image analysis using white light (WL) or by fluorescence using complementary DNA oligonucleotides attached to fluorophore-based dyes. DNA templates can be sequenced, e.g., using ligation-based sequencing.

[00457] Figure 33 A shows a schematic diagram of the slide with beads attached thereto.

Note that only a small proportion of the DNA template molecules are attached to the slide. One micron beads (Dynabeads MyOne Streptavidin beads; Dynal Biotech, Inc., Product No.

650.01) were used. However, a wide variety of beads could be used.

[00458] Figure 33B shows a population of beads attached to a slide. The lower panels show the same region of the slide under white light (left) and fluorescence microscopy. The upper panel shows a range of bead densities. Example 16: Sequencing by Oligonucleotide Extension and Ligation using a

Gel-free Bead-Based Array

[00459] This example describes preparation of an array of microparticles attached to a substrate (glass slides) via a biotin-streptavidin interaction and demonstrates successful

Page 144 of 165 381 sequencing by cycles of ligation, cleavage, and detection. Microparticles having biotinylated templates are attached thereto were prepared using emulsion PCR and attached to a substrate functionalized with streptavidin via a PEG-containing linkage in the absence of semi-solid medium as described below. The method employs streptavidin-coated beads to which a biotinylated primer was attached prior to amplification. Following amplification and enrichment for particles on which productive template amplification had occurred, the templates were biotinylated. The microparticles having biotinylated templates attached thereto were then incubated with streptavidin-coated slides. Thus a biotin-strepatividin linkage was employed twice in this method. Other approaches would employ alternative means of linking primers to microparticles or linking amplified templates to substrate. [00460] Materials and Methods: [00461] Preparation of BAC Eco v2.1 beads.

[00462] MyOne streptavidin beads (1 -micron) were coated with biotinylated PI primer (see Figures) and used in emulsion PCR to create a population of beads having templates from our BAC-Eco (v 2.1) library attached thereto. The emulsion was broken and beads were purified and treated with exonuclease in a standard way. The beads having fully extended PCR product were enriched by binding to enrichment beads covered by P2 enrichment oligo (see Figures). To improve behavior of enriched beads in solution, they were incubated with biotinylated PI oligo to cover any bead area that had exposed streptavidin coating.

[00463] Deposition of BAC Eco v2.1 beads on slides.

[00464] Enriched BAC-Eco v2.1 beads containing ssDNA were deposited on streptavidin-coated Opti-Chem slides (Accel8 Technology Corporation). To prepare for this process they were incubated with terminal transferase (New England Biolabs) and biotin-11- ddATP (Perkin Elmer) to covalently attach biotin moieties onto the 3 '-ends of DNA template molecules. The beads were mixed with an equal number of MyOne Carboxylic Acid beads (Dynal) and placed in deposition buffer containing 5 mM Tris HCl pH 8.0, 5 mM EDTA, 0.0005% Triton X-100 and 10% PEG 8000 (American Bioanalytical). The suspension was sonicated shortly using Covaris S2 sonicator and deposited onto

streptavidin-coated Opti-Chem slides (Accel8 Technology Corporation). Slides were washed three times with TE buffer and dried with compressed air prior to use. The suspension was covered with a LifterSlip (Erie Scientific Company) to produce even aqueous layer on the slide and reduce evaporation. The slides were incubated for 45 min at room temperature in a

Page 145 of 165

?81 high-humidity chamber to allow the beads to settle and bind to the surface while reducing evaporation on edges. Cover slips were removed by immersing slides in upside down position in a tray filled with TE buffer. Gentle agitation for about one minute removed most of the carboxylic acid beads (as was showed in a separate experiment). The slides were immediately immersed in acetone and dried using compressed air.

[00465] Reagents used in cycled ligation sequencing on gel-less slides were the same as for acrylamide-based gels except for Reset buffer. For non-gel arrays, an alkaline-based Reset buffer was used, containing 10 mM NaOH and 0.1% sodium dodecanesulfonate (Fluka). As demonstrated in Figures 38 and 39, a 300-panel gel-less array (approximately 18x18mm) was seeded with enriched BAC -Eco library beads and placed into an automated small flow cell instrument and exposed to 50 rounds of alkaline reset to validate bead stability in a gel-less environment. Following the 50-cycle flow regimen, the gel-less array contained over 26,000 beads per panel (4Mpixel camera). The gel-less array was then sequenced using cycled ligation and cleavage. Evaluation of cycle 1 data supported the efficient ligation of our 2-base, 4-color probe set as evidenced by high RFU values for each fluorescent channel (Figure 39). Bead populations were subsequently basecalled and plotted on spectral purity plots and demonstrated excellent sequencing performance by Satay analysis and density plot evaluation. [00466]

Equivalents and Scope

[00467] Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. The scope of the present invention is not intended to be

limited to the above Description, but rather is as set forth in the appended claims. In the following claims articles such as "a,", "an" and "the" may mean one or more than one unless indicated to the contrary or otherwise evident from the context. Claims or descriptions that include "or" between one or more members of a group are considered satisfied if one, more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process unless indicated to the contrary or otherwise evident from the context. Use of "optionally" in a claim indicates that the invention includes embodiments in which the optional feature is present and embodiments in which it is absent.

Page 146 of 165

381 [00468] Furthermore, it is to be understood that the invention encompasses all variations, combinations, and permutations in which one or more limitations, elements, clauses, descriptive terms, etc., from one or more of the listed claims is introduced into another claim. In particular, any claim that is dependent on another claim can be modified to include one or more limitations found in any other claim that is dependent on the same base claim.

[00469] In addition, it is to be understood that any one or more embodiments may be explicitly excluded from the claims even if the specific exclusion is not set forth explicitly herein. It should also be understood that where the specification and/or claims disclose a reagent (e.g., a template, microsphere, probe, probe family, etc.) of use in sequencing, such disclosure also encompasses methods for sequencing using the reagent according either to the specific methods disclosed herein, or other methods known in the art unless one of ordinary skill in the art would understand otherwise, or unless otherwise indicated in the specification. In addition, where the specification and/or claims disclose a method of sequencing, any one or more of the reagents disclosed herein may be used in the method, unless one of ordinary skill in the art would understand otherwise, or unless use of the reagent in such method is explicitly excluded in the specification. It should further be understood that where particular components of use in sequencing are disclosed in the specification or claims, the invention encompasses methods for making the reagents also. The term "component" is used broadly to

refer to any item used in sequencing, including templates, microparticles having templates attached thereto, libraries, etc. Furthermore, the figures are an integral part of the specification, and the invention includes structures shown in the figures, e.g., microparticles having templates attached thereto, and methods disclosed in the figures. [00470] Where ranges are given herein, the endpoints are included. Furthermore, it is to be understood that unless otherwise indicated or otherwise evident from the context and understanding of one of ordinary skill in the art, values that are expressed as ranges can assume any specific value or subrange within the stated ranges in different embodiments of the invention, to the tenth of the unit of the lower limit of the range, unless the context clearly dictates otherwise.

Page 147 of 165

381

Espacenet

Claims: JP2009538123 (A) — 2009-11-05

REAGENTS, METHODS, AND LIBRARIES FOR GEL-FREE BEAD-BASED SEQUENCING

Claims not available for JP2009538123 (A)
Claims of corresponding document: WO2007121489 (A2)

A high quality text as facsimile in your desired language may be available amongst the following family members:

CA2649725 (A1) CN101495654 (A) US2009062129 (A1) WO2007121489 (A2)

- Original claims
- Claims tree

The EPO does not accept any responsibility for the accuracy of data and information originating from other authorities than the EPO; in particular, the EPO does not guarantee that they are complete, up-to-date or fit for specific purposes.

We claim:

1. A method for identifying a sequence of nucleotides in a template polynucleotide, the method comprising the steps of:
(a) extending an initializing oligonucleotide along the template polynucleotide by ligating an oligonucleotide probe thereto to form an extended duplex, wherein the oligonucleotide probe is attached to a microparticle, wherein the microparticle is attached to a substrate, and wherein the microparticle is not immobilized in a semi-solid support;
(b) identifying one or more nucleotides of the polynucleotide; and (c) repeating steps (a) and (b) until the sequence of nucleotides is determined.

2. The method of claim 1 , wherein the oligonucleotide probe comprises a phosphorothiolate linkage.

3. The method of claim 1, wherein the step of identifying includes detecting a label attached to the most recently ligated oligonucleotide probe.

4. The method of claim 3, further comprising the step of generating an extendable probe terminus by cleaving

the phosphorothiolate linkage with a cleavage agent comprising an atom selected from the group consisting of: Ag, Hg, Cu, Mn, Zn and Cd.

5. The method of claim 4, wherein the cleavage agent is AgNO3.

6. The method of claim 1, comprising contacting the template polynucleotide with a blocking oligonucleotide prior to extending it.

7. The method of claim 6, wherein the blocking oligonucleotide is not enzymatically extendable.

8. The method of claim 1, wherein the microparticle is attached to the substrate by a linkage comprising biotin and a biotin-binding protein.

9. The method of claim 8, wherein a single-stranded template tethers the microparticle to the substrate by a linkage comprising biotin and a biotin-binding protein.
Page 148 of 165
?81

10. The method of claim 1, wherein the microparticle is attached to the substrate by a linkage comprising biotin and a biotin-binding protein, wherein the biotin-binding protein is attached to the substrate.

11. The method of claim 1 , wherein the microparticle is attached to the substrate by a linkage comprising biotin and a biotin-binding protein, wherein the biotin-binding protein is attached to the substrate and the template comprises biotin.

12. The method of claim 1 , wherein the substrate is substantially planar and rigid.

13. A method for determining a sequence of nucleotides in a template polynucleotide, the method comprising the steps of: (a) providing a probe-template duplex comprising a probe hybridized to a template polynucleotide, the probe having an extendable terminus;
(b) ligating an extension oligonucleotide probe to said extendable terminus, to form an extended duplex containing an extended oligonucleotide probe, wherein the oligonucleotide probe is attached to a microparticle, wherein the microparticle is attached to a substrate, and

wherein the microparticle is not immobilized in a semisolid support;
(c) identifying, in the extended duplex, at least one nucleotide in the template polynucleotide that is either (1) complementary to the just- ligated extension probe or (2) a nucleotide residue in the template polynucleotide which is immediately downstream of the extended oligonucleotide probe;
(d) generating an extendable terminus on the extended oligonucleotide probe, if an extendable terminus is not already present, such that the terminus generated is different from the terminus to which the last extension probe was ligated; and
(e) repeating steps (b), (c) and (d) until a sequence of nucleotides in the template polynucleotide is determined.

14. The method of claim 13, wherein the extension probe comprises a phosphorothiolate linkage;

15. The method of claim 13, wherein each extension probe has a non-extendable moiety at one terminus.
Page 149 of 165

381

16. The method of claim 13, wherein the step of identifying includes detecting a label attached to the most recently ligated extension probe.

17. The method of claim 13, wherein the step of identifying includes removing said non- extendable moiety and extending said extended oligonucleotide probe with a nucleic acid polymerase in the presence of one or more labeled chain-terminating nucleoside triphosphates.

18. The method of claim 13, further including a step of capping an extended oligonucleotide probe whenever no extension probe has ligated to the extendable terminus in the ligation step.

19. The method of claim 13, wherein the generating step comprises cleaving the phosphorothiolate linkage with a cleavage agent comprising an atom selected from the group consisting of: Ag, Hg, Cu, Mn, Zn and Cd.

20. The method of claim 19, wherein the cleavage agent is AgNO3.

21. The method of claim 13, further comprising the steps of: (f) removing the ligated probes and the initializing oligonucleotide from the template; (g) repeating step (a) using a second oligonucleotide that is bound to a different sequence of the template polynucleotide; and (h) repeating steps (b) to (e).

22. The method of claim 21, wherein the method is repeated a plurality of times using initializing oligonucleotides bound to different sequences of the template polynucleotide.

23. The method of claim 22, wherein the extension probes have a non-extendable moiety at one terminus.

24. The method of claim 22, wherein for each repetition, the step of identifying includes detecting a label attached to the most recently ligated extension probe.

25. The method of claim 22, further including a step of capping an extended oligonucleotide probe whenever no extension probe has ligated to the extendable terminus in the ligation step.
Page 150 of 165

381

26. The method of claim 22, wherein the generating step comprises cleaving the phosphorothiolate linkage with a cleavage agent comprising an atom selected from the group consisting of: Ag, Hg, Cu, Mn, Zn and Cd.

27. The method of claim 26, wherein the cleavage agent is AgNO3.

28. The method of claim 13, comprising contacting the template polynucleotide with a blocking oligonucleotide prior to providing the probe-template duplex.

29. The method of claim 28, wherein the blocking oligonucleotide is not enzymatically extendable.

30. The method of claim 13, comprising, prior to providing the probe-template duplex: (a) contacting the template polynucleotide with a blocking oligonucleotide; and
(b) forming a probe-template duplex.

31. A method for identifying a sequence of nucleotides in a template polynucleotide, the method comprising the steps of: (a) providing a template polynucleotide attached to a microparticle that is immobilized in or on a semi-solid support or is attached to a substantially planar, rigid substrate;
(b) contacting the template polynucleotide with a blocking oligonucleotide;
(c) extending an initializing oligonucleotide along the template polynucleotide by ligating an oligonucleotide probe thereto to form an extended duplex, wherein the oligonucleotide probe optionally comprises a scissile linkage;
(d) identifying one or more nucleotides of the polynucleotide; and
(e) repeating steps (c) and (d) until the sequence of nucleotides is determined.

32. The method of claim 31 , wherein the extending step is performed in the semi-solid support.

33. The method of claim 31 , wherein the template is attached to a microparticle that is attached to a substantially planar, rigid substrate.

34. The method of claim 33, wherein the microparticle is attached to the substrate by a linkage comprising biotin and a biotin-binding protein.
Page 151 of 165 381

35. The method of claim 34, wherein a single-stranded template tethers the microparticle to the substrate by a linkage comprising biotin and a biotin-binding protein.

36. The method of claim 33, wherein the microparticle is attached to the substrate by a linkage comprising biotin and a biotin-binding protein, wherein the biotin-binding protein is attached to the substrate.

37. The method of claim 33, wherein a single-stranded template attached to the bead tethers the microparticle to the substrate.

38. A method for determining a sequence of nucleotides in a template polynucleotide, the method comprising the steps of: (a) providing a probe-template duplex comprising a probe hybridized to a template polynucleotide, the probe having an extendable terminus and the template having a blocking

oligonucleotide hybridized thereto, the probe-template duplex being attached to a microparticle that is embedded in or on a semi-solid support or attached to a substrate; (b) ligating an extension oligonucleotide probe to said extendable terminus, to form an extended duplex containing an extended oligonucleotide probe, wherein the extension probe comprises a phosphorothiolate linkage;
(c) identifying, in the extended duplex, at least one nucleotide in the template polynucleotide that is either (1) complementary to the just- ligated extension probe or (2) a nucleotide residue in the template polynucleotide which is immediately downstream of the extended oligonucleotide probe;
(d) generating an extendable terminus on the extended oligonucleotide probe, if an extendable terminus is not already present, such that the terminus generated is different from the terminus to which the last extension probe was ligated; and (e) repeating steps (b), (c) and (d) until a sequence of nucleotides in the template polynucleotide is determined.

39. The method of claim 38, comprising contacting a template with a blocking oligonucleotide prior to step (a).

40. The method of claim 38, wherein the ligating and generating steps are performed in the semi-solid support.
Page 152 of 165 381

41. The method of claim 38, wherein the template is attached to a microparticle that is attached to a substantially planar, rigid substrate.

42. The method of claim 41 , wherein the microparticle is attached to the substrate by a linkage comprising biotin and a biotin-binding protein.

43. The method of claim 41 , wherein the microparticle is attached to the substrate by a linkage comprising biotin and a biotin-binding protein, wherein the biotin-binding protein is attached to the substrate.

44. The method of claim 38, wherein a single-stranded template attached to the microparticle tethers the microparticle to the substrate.

45. A method for determining a sequence of nucleotides in a template polynucleotide, the method comprising the

steps of:
(a) amplifying a template polynucleotide molecule in a compartment of an emulsion in the presence of a microparticle so that a microparticle having a clonal population of template polynucleotides attached thereto is produced; (b) recovering the microparticle from the emulsion;
(c) embedding the microparticle in or on a semi-solid support or attaching the microparticle to a substrate;
(d) extending an initializing oligonucleotide along the template polynucleotide by ligating an oligonucleotide probe thereto to form an extended duplex, wherein the oligonucleotide probe comprises a scissile linkage;
(e) identifying one or more nucleotides of the polynucleotide; and
(f) repeating steps (d) and (e) until the sequence of nucleotides is determined.

46. The method of claim 45, wherein (i) a plurality of template polynucleotide molecules comprising different sequences are amplified in individual compartments of the emulsion; (ii) a plurality of microparticles, each having attached thereto a clonal populations of template polynucleotides, wherein the clonal populations have different sequences, is recovered from the emulsion and embedded in or on the support, and (iii) steps (d), (e), and (f) are performed in parallel on the clonal populations attached to the embedded or attached microparticles so as to determine a plurality of sequences in parallel.
Page 153 of 165 381

47. The method of claim 45, comprising contacting the template polynucleotide with a blocking oligonucleotide prior to step (c).

48. The method of claim 45, wherein the microparticles are attached to the substrate by a linkage comprising biotin and a biotin-binding protein.

49. The method of claim 45, wherein the microparticles are attached to the substrate by a linkage comprising biotin and a biotin-binding protein, wherein the biotin-binding protein is attached to the substrate.

50. The method of claim 45, wherein a single-stranded template attached to the microparticle tethers the microparticle to the substrate.

51. A method for determining information about a sequence of nucleotides in a template polynucleotide using a first collection of at least 2 distinguishably labeled oligonucleotide probe families, the method comprising the steps of:
(a) extending an initializing oligonucleotide along the template polynucleotide by ligating an oligonucleotide probe thereto to form an extended duplex, wherein the oligonucleotide probe is a member of the collection of distinguishably labeled oligonucleotide probe families and has a blocking oligonucleotide hybridized thereto;
(c) detecting a label associated with the oligonucleotide probe; and
(d) repeating steps (b) and (c) until an ordered list of probe family names is obtained; and
(d) using the ordered list of probe family names to eliminate one or more possibilities for the sequence of nucleotides.

52. The method of claim 51 , wherein step (d) comprises decoding the ordered list of probe family names to determine the sequence.

53. The method of claim 51 , wherein the method comprises providing a probe-template duplex comprising an initializing oligonucleotide probe hybridized to a template polynucleotide, the probe having an extendable terminus, and wherein the extending step comprises ligating an oligonucleotide probe to said extendable terminus, to form an extended duplex containing an extended oligonucleotide probe, further
Page 154 of 165 381 comprising the step of capping any remaining extendable termini whenever no oligonucleotide probe has ligated to the extendable terminus in the extending step.

54. The method of claim 51 , wherein the oligonucleotide probes in each probe family comprise a non-extendable moiety at one terminus.

55. The method of claim 51 , further comprising, after each detecting step, (f) producing an extendable terminus on the most recently ligated oligonucleotide probe, if an extendable terminus is not already present, such that the terminus generated is different from the terminus to which the most recently ligated oligonucleotide probe was ligated.

56. The method of claim 55, wherein the oligonucleotide probe comprises a phosphorothiolate linkage and the extendable probe terminus is produced by cleaving the phosphorothiolate linkage with a cleavage agent comprising an atom selected from the group consisting of: Ag, Hg, Cu, Mn, Zn and Cd.

57. The method of claim 56, wherein the cleavage agent is AgNO3.

58. The method of claim 51 , wherein the extending step is performed in or on a semisolid support.

59. The method of claim 51 , wherein the template is attached to a microparticle that is attached to a substantially planar, rigid substrate.

60. The method of claim 51 , wherein the collection comprises 2 distinguishably labeled probe families.

61. The method of claim 51 , wherein the collection comprises 3 distinguishably labeled probe families.

62. The method of claim 51 , wherein the collection comprises 4 distinguishably labeled probe families.

63. The method of claim 51 , wherein the collection comprises more than 4 distinguishably labeled probe families.
Page 155 of 165

381

64. The method of claim 51 , wherein the oligonucleotide probes comprise a constrained portion in which nucleosides are not independently selected, and wherein oligonucleotide probes having constrained portions that differ in sequence are assigned to probe families according to an encoding.

65. The method of claim 51 , wherein oligonucleotide probes are assigned to first, second, third, and fourth probe families according to one of the 24 encodings set forth in Table 1.

66. The method of claim 52, wherein at least one nucleotide in the template has a known identity, and wherein the decoding step comprises: (i) assigning an identity to a nucleotide in the template adjacent to the

nucleotide of known identity by determining which identity is consistent with the identity of the known nucleotide and the possible sequences of the constrained portion of the probe whose proximal nucleotide ligated opposite the nucleotide adjacent to the nucleotide of known identity; (ii) assigning an identity to a succeeding nucleotide by determining which identity is consistent with possible sequences of the constrained portion of the probe whose proximal nucleotide ligated opposite the succeeding nucleotide; and (iii) repeating step (ii) until the sequence is determined.

67. The method of claim 52, further comprising the step of: (a) determining the identity of a nucleotide in the template, so that the nucleotide has a known identity, and wherein the decoding step comprises:
(i) assigning an identity to a nucleotide in the template adjacent to the nucleotide of known identity by determining which identity is consistent with the identity of the known nucleotide and the possible sequences of the constrained portion of the probe whose proximal nucleotide ligated opposite the nucleotide adjacent to the nucleotide of known identity;
(ii) assigning an identity to a succeeding nucleotide by determining which identity is consistent with possible sequences of the constrained portion of the probe whose proximal nucleotide ligated opposite the succeeding nucleotide; and (iii) repeating step (ii) until the sequence is determined.
Page 156 of 165

381

68. The method of claim 67, wherein the determining step comprises contacting a template-probe duplex with a labeled nucleotide in the presence of a polymerase under conditions that allow incorporation of the labeled nucleotide if it is complementary to the template at the position adjacent to the duplex.

69. The method of claim 52, wherein the decoding step comprises: generating at least one candidate sequence from the ordered list of probe family names; and selecting a candidate sequence as the sequence of nucleotides in the template.

70. The method of claim 69, wherein the generating step comprises generating at least 4 candidate sequences.

71. The method of claim 69, wherein the generating step comprises:
(i) assuming an identity for a first nucleotide in the sequence of nucleotides;
(ii) assigning an identity for a nucleotide adjacent to the first nucleotide by determining a possible identity for the adjacent nucleotide based on the name of the probe family that corresponds to the first nucleotide;
(iii) assigning an identity to a succeeding nucleotide by determining a possible identity for the succeeding nucleotide based on the name of the probe family that corresponds to the nucleotide whose identity was most recently assigned;
(iv) repeating step (iii) until a candidate sequence is generated; and (v) repeating steps (i) - (iv), wherein, in each repetition, a different identity is assumed for the first nucleotide, until a desired number of candidate sequences is generated.

72. The method of claim 69, wherein the selecting step comprises comparing at least one candidate sequence with one or more known sequences and selecting a candidate sequence that exhibits a predetermined degree of identity is most nearly identical to to one or more of the known sequences.

73. The method of claim 72, wherein the template is derived from an organism of interest, and wherein the comparing step comprises comparing at least one candidate sequence with sequences in a database that contains sequences obtained from the organism.
Page 157 of 165 381

74. The method of claim 72, wherein the comparing step comprises comparing at least one candidate sequence with sequences in a database that contains a plurality of comparison sequences, each of which comprises an alternative possible sequence for the sequence of polynucleotides to be determined.

75. The method of claim 69, wherein the selecting step comprises:
(i) obtaining a second ordered list of probe family names from the template using a second collection of distinguishably labeled encoded probe families, wherein the probe families in the second collection of probe families are encoded differently to the probe families in the first collection of probe families; (ii) generating at

least one comparison sequence from the second ordered list of probe family names;
(iii) comparing a portion of at least one of the candidate sequences with a portion of at least one of the comparison sequences; and
(iv) selecting a candidate sequence that exhibits a predetermined level of identity or is most nearly identical to a comparison sequence over the portion compared in step (c) as the sequence of nucleotides in the template.

76. The method of claim 75, wherein the portion compared is a single dinucleotide.

77. The method of claim 75, wherein the second ordered list of probe family names contains only a single element.

78. The method of claim 50, wherein the oligonucleotide probes in each probe family have the structure 5'-(XY)(N)kNB*-3' or 3'-(XY)(N)kNB*-5\ wherein N represents any nucleoside, NB represents a moiety that is not extendable by ligase, * represents a detectable moiety, XY is a constrained portion of the probe in which X and Y represent nucleosides that are identical or different but are not independently selected, X and Y are at least 2-fold degenerate, at least one internucleoside linkage is a scissile linkage, and k is between 1 and 100, inclusive, with the proviso that a detectable moiety may be present on Y or on any nucleoside of (N)k instead of, or in addition to, NB.

79. The method of claim 78, wherein the scissile linkage is a phosphorothiolate linkage.
Page 158 of 165

381

80. The method of claim 78, wherein the detectable moiety is attached by a cleavable linker, is photobleachable, or both.

81. The method of claim 80, wherein the cleavable linker comprises a disulfide bond.

82. The method of claim 78, wherein four distinguishably labeled oligonucleotide probe families are used, and wherein oligonucleotide probes having different sequences for the constrained portion of the probe are assigned to first, second, third, and fourth

probe families according to one of the 24 encodings set forth in Table 1.

83. The method of claim 51 , wherein the detecting step comprises acquiring on average 2 bits of information simultaneously from each of at least 2 nucleotides in the template without acquiring two bits of information from any individual nucleotide.

84. The method of claim 51 , wherein the detecting step comprises acquiring less than 2 bits of information simultaneously from each of at least 2 nucleotides in the template.

85. A method for determining information about a sequence of nucleotides in a template polynucleotide using a first collection of at least 2 distinguishably labeled oligonucleotide probe families, the method comprising the steps of:
(a) contacting a probe-template complex comprising a duplex portion with an extendable terminus and a single-stranded portion of interest to be sequenced with at least two distinguishably labeled oligonucleotide probe families, so that hybridization occurs between an oligonucleotide probe that comprises a portion complementary to the portion of the template immediately adjacent to the duplex portion, wherein the template has a blocking oligonucleotide hybridized thereto;
(b) ligating the hybridized oligonucleotide probe to the extendable terminus, thereby producing a probe-template complex comprising an extended duplx; (c) detecting a label associated with the ligated probe;
(d) generating an extendable probe terminus on the extended duplex if one is not already present; and
(e) repeating steps (a) to (d) until an ordered list of probe family names is obtained.
Page 159 of 165

381

86. The method of claim 85, wherein the detecting step comprises acquiring on average 2 bits of information simultaneously from each of at least 2 nucleotides in the template without acquiring two bits of information from any individual nucleotide.

87. The method of claim 85, wherein the detecting step comprises acquiring less than 2 bits of information

simultaneously from each of at least 2 nucleotides in the template.

88. A method for determining information about a sequence of nucleotides in a template polynucleotide using a first collection of oligonucleotide probe families, the method comprising the steps of: (a) performing sequential cycles of extension, ligation, detection, and cleavage, wherein the detection step comprises acquiring on average 2 bits of information simultaneously from each of at least 2 nucleotides in the template without acquiring two bits of information from any individual nucleotide, wherein the template has a blocking oligonucleotide hybridized thereto; and (b) combining the information obtained in step (a) with at least one bit of additional information to determine the sequence.

89. The method of claim 88, wherein the at least one bit of additional information comprises an item selected from the group consisting of: the identity of a nucleotide in the template, information obtained by comparing a candidate sequence with at least one known sequence; and information obtained by repeating the method using a second collection of oligonucleotide probe families.

90. A method of preparing a plurality of template polynucleotides comprising steps of:
(a) contacting a plurality of microparticles with a semi-solid support, wherein at least some of the microparticles have a template attached thereto and wherein the semi-solid support comprises primers attached thereto or embedded therein, so that the templates hybridize to the primers; and
(b) extending the primers to form templates complementary to the templates attached to the microparticles.

91. The method of claim 90, further comprising amplifying the templates generated by extending the primers.
Page 160 of 165

381

92. The method of claim 91 , wherein amplifying comprises performing RCA.

93. The method of claim 90, further comprising

releasing the microparticles from the semi-solid support.

94. The method of claim 90, further comprising sequencing the templates generated by extending the primers, optionally after amplifying the templates.

95. A collection of components for preparing a population of microparticles, the collection comprising:
(a) a population of microparticles wherein individual microparticles have at least first and second populations of primers attached thereto, wherein the primers of the first population have a different sequence from that of the primers of the second population;
(b) a library of nucleic acid fragments, wherein each nucleic acid fragment contains first and second nucleic acid segments of interest, and wherein the first and second primers correspond to universal sequences located external to the first and second nucleic acid segments of interest; and
(c) a blocking oligonucleotide that binds to a common region of the nucleic acid fragments.

96. The collection of components of claim 95, wherein the first and second nucleic acid segments of interest are 5' and 3' tags of a paired tag.

97. The collection of components of claim 95, wherein the nucleic acid fragments contain an internal adaptor comprising one or more primer binding sites for amplification primers so that each of the nucleic acid segments can be amplified using PCR.

98. The collection of components of claim 97, further comprising primers complementary to the primer binding sites in the internal adaptor.

99. A template comprising a population of substantially identical template molecules, wherein the template molecules comprise at least one common region and at least one segment of interest, and wherein at least some of the template molecules have a blocking oligonucleotide hybridized to the common region.
Page 161 of 165

381

100. The template of claim 99, wherein the template molecules are members of a paired- tag library.

101. The template of claim 99, wherein the template molecules are amplified using RCA.

102. The template of claim 99, wherein the template molecules comprise at least two common regions and at least one segment of interest, and wherein at least some of the template molecules have a blocking oligonucleotide hybridized to each of at least two common regions.

103. A support or substrate having the population of claim 99 attached thereto.

104. The support or substrate of claim 101, which is a microparticle.

105. The support or substrate of claim 101, which is a semi-solid support.

106. The support or substrate of claim 101, which is a substantially planar, rigid support.

107. A collection of templates of claim 99, wherein the templates comprise different segments of interest.

108. An array comprising a population of microparticles of claim 104, wherein the microparticles have templates comprising different segments of interest attached thereto.

109. The array of claim 108, wherein the microparticles are immobilized in or on a semisolid support or attached to a substrate.

110. A microparticle attached to the substrate, wherein the microparticle has a template attached thereto.

111. The microparticle of claim 110, wherein the microparticle is attached to the substrate by a linkage comprising biotin and a biotin-binding protein.

112. The microparticle of claim 110, wherein the microparticle is attached to the substrate by a linkage comprising biotin and a biotin-binding protein, wherein the biotin- binding protein is attached to the substrate.
Page 162 of 165

381

113. The microparticle of claim 110, wherein the microparticle is attached to a single- stranded template which is attached to the substrate, whereby the template tethers the microparticle to the substrate.

114. The microparticle of claim 110, wherein the microparticle is attached to a single- stranded template which is attached to the substrate, whereby the template tethers the microparticle to the substrate, and wherein the template is attached to the substrate and the microparticle by linkages that comprise biotin and a biotin-binding protein.

115. A population of microparticles of claim 110, wherein templates comprising different segments of interest and a common sequence are attached to different microparticles.

116. A method for preparing an array comprising steps of: providing a population of microparticles having templates attached thereto, wherein the templates comprise biotin; and contacting the microparticles with a substrate comprising a biotin-binding protein under conditions in which biotin binds to the biotin-binding protein, thereby producing an array of microparticles.

117. The method of claim 116, further comprising hybridizing blocking oligonucleotides to the templates.

118. The method of claim 116, further comprising sequencing the templates.
Page 163 of 165

381

Espacenet

Bibliographic data: JP2015090458 (A) — 2015-05-11

ANALYZER

Inventor(s): KATO KOICHI; SHOJI TOMOHIRO; YAMAMOTO SHUHEI; HIRATSUKA TOSHIAKI ± (加藤 宏一, ; 庄司 智広, ; 山本 周平, ; 平塚 稔章)

Applicant(s): HITACHI HIGH TECH CORP ± (株式会社日立ハイテクノロジーズ)

Classification: - international: ***G01N21/64; G02B21/00; G02B21/28; G02B21/36***

- cooperative:

Application number: JP20130230774 20131107

Global Dossier

Priority number(s): JP20130230774 20131107

Also published as: JP6286183 (B2)

Abstract of JP2015090458 (A)

PROBLEM TO BE SOLVED: To solve a problem that analyzer downsizing is difficult to achieve since a method of optically detecting a liquid on a tip of an immersion objective lens increases cost and needs a plenty of space.SOLUTION: An analyzer includes: an objective lens for observing a sample; a mechanism driving the objective lens in an optical axis direction; a light source illuminating the sample; a photosensor detecting light emission from the sample; a nozzle supplying a liquid between the sample and the objective lens; means for regulating a temperature of the sample; and means for supplying a plurality of reagents to the sample, and the analyzer electrically detects a state of the liquid.

(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2015-90458
(P2015-90458A)
(43)公開日　平成27年5月11日(2015.5.11)

(51)Int.Cl.		F I			テーマコード（参考）
G02B 21/00	***(2006.01)***	G02B	21/00		2G043
G02B 21/28	***(2006.01)***	G02B	21/28		2H052
G02B 21/36	***(2006.01)***	G02B	21/36		
G01N 21/64	***(2006.01)***	G01N	21/64	F	

審査請求 未請求　請求項の数 12　ＯＬ　（全 24 頁）

(21)出願番号　特願2013-230774 (P2013-230774)
(22)出願日　平成25年11月7日 (2013.11.7)

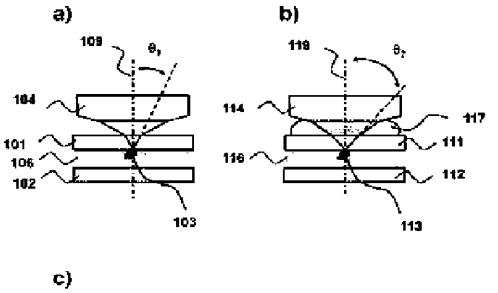
(71)出願人　501387839
株式会社日立ハイテクノロジーズ
東京都港区西新橋一丁目２４番１４号
(74)代理人　100100310
弁理士　井上　学
(74)代理人　100098660
弁理士　戸田　裕二
(74)代理人　100091720
弁理士　岩崎　重美
(72)発明者　加藤　宏一
東京都港区西新橋一丁目２４番１４号　株式会社　日立ハイテクノロジーズ内

最終頁に続く

(54)【発明の名称】分析装置

(57)【要約】
【課題】液浸対物レンズ先端における液体の光学的検出方法はコスト高であり、かつスペースを大きくとるため、装置の小型化が困難である。
【解決手段】サンプルを観察するための対物レンズと、前記対物レンズを光軸方向に駆動する機構と、前記サンプルを照明する光源と、前記対物レンズを介して前記サンプルからの発光を検出する光センサと、前記サンプルと前記対物レンズとの間に液体を供給するノズルと、前記サンプルの温度を調整する手段と、前記サンプルに複数の試薬を供給する手段、を有し、前記液体の状態を電気的に検出する。
【選択図】　図１

図１



対物名称	開口数	媒質の屈折率	最大角度 [°]	信号比	分解能 [μm]
乾燥系対物レンズ	0.75	1.00	48.6	1.00	0.45
水浸対物レンズ	1.2	1.33	64.5	3.27	0.28

【特許請求の範囲】
【請求項１】
サンプルを観察するための対物レンズと、前記対物レンズを光軸方向に駆動する機構と、前記サンプルを照明する光源と、前記対物レンズを介して前記サンプルからの発光を検出する光センサと、前記サンプルと前記対物レンズとの間に液体を供給するノズルと、前記サンプルの温度を調整する手段と、前記サンプルに複数の試薬を供給する手段、を有し、
前記液体の状態を電気的に検出することを特徴とする分析装置。
【請求項２】
請求項１に記載の分析装置において、
前記液体の状態の電気的な検出は、静電容量を測定することにより行うことを特徴とする、分析装置。
【請求項３】
請求項２に記載の分析装置において、
前記静電容量を測定するのは、前記ノズルであることを特徴とする、分析装置。
【請求項４】
請求項３に記載の分析装置において、
さらに、内部に流路を有するフローセルを固定する機構と、該フローセルを水平方向へ駆動する機構とを有し、
前記ノズルは、該ノズルと前記フローセルとの間の静電容量を測定することを特徴とする、分析装置。
【請求項５】
請求項３に記載の分析装置において、
前記ノズルを覆う導電性の電気シールドが設けられていることを特徴とする、分析装置。
【請求項６】
請求項２に記載の分析装置において、
前記対物レンズの周辺に複数の通電センサを備え、
当該通電センサを用いて液体の状態を電気的に検出することを特徴とする、分析装置。
【請求項７】
請求項６に記載の分析装置において、
前記通電センサは円周上に並んで配置されていることを特徴とする、分析装置。
【請求項８】
請求項６に記載の分析装置において、
前記通電センサは等間隔に並んで配置されていることを特徴とする、分析装置。
【請求項９】
請求項１に記載の分析装置において、
前記ノズルの先端が環状であり、複数の液体吐出口を持つことを特徴とする、分析装置。
【請求項１０】
請求項１に記載の分析装置において、
前記サンプルは、光透過性を持つフローチップに保持され、当該フローチップは試薬の注入口、排出口、および流路を備え、フローチップの表面が疎水性であり、純水を弾く性質を持つことを特徴とする、分析装置。
【請求項１１】
請求項１０記載の分析装置において、
前記フローチップの流路の上下面に塩基配列の異なる複数のDNA分子が離散的に固定されていることを特徴とする、分析装置。
【請求項１２】
請求項１０記載の分析装置において、

前記フローチップの表面が透明電極膜で被覆されていることを特徴とする、分析装置。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は分析装置に関わる。より具体的には、液浸対物レンズを用いて光学測定を行う場合にサンプルと対物レンズ間に液体を自動で供給する分析装置に関する。

【背景技術】

【０００２】

対物レンズとサンプルの間に自動で純水あるいは油などのカップリング剤を供給する方法について特許文献１に述べられている。この特許ではカップリング剤である液体を、対物レンズとサンプル間に注入する機構を保持している。特に液体が純水である場合には、長期間に渡る計測においては対物レンズ先端の純水が乾燥するという問題がある。

【先行技術文献】

【特許文献】

【０００３】

【特許文献１】特開２０１０－２６２１８号公報

【発明の概要】

【発明が解決しようとする課題】

【０００４】

これまでに液浸対物レンズ先端における液体の検出は光学的測定を用いて行われてきた。しかし、光学的検出方法はコスト高であり、かつスペースを大きくとるために装置の小型化が困難であった。

【０００５】

より具体的に光学検出における問題点を挙げる。特許文献１に示される赤外線発光光学系は対物レンズの光軸に対して60°程度の角度に入射光路と反射光路を設けなければならない。これはそもそもの対物レンズを用いた光学系のスペースを左右方向に大きく拡張することが必要となる。そのため、コストが高くなるという課題があった。

【課題を解決するための手段】

【０００６】

本発明の分析装置は、サンプルを観察するための対物レンズと、前記対物レンズを光軸方向に駆動する機構と、前記サンプルを照明する光源と、前記対物レンズを介して前記サンプルからの発光を検出する光センサと、前記サンプルと前記対物レンズとの間に液体を供給するノズルと、前記サンプルの温度を調整する手段と、前記サンプルに複数の試薬を供給する手段、を有し、前記液体の状態を電気的に検出することを特徴とする。

【発明の効果】

【０００７】

本発明の分析装置は、コストを抑え、小型化を図ることができる。

【図面の簡単な説明】

【０００８】

【図１】乾燥系対物レンズと液浸対物レンズにおける開口数および分解能について説明図。

【図２】液浸対物レンズにおけるフローチップ流路内の上面と下面の計測方法についての説明図。

【図３】液浸対物レンズを用いたDNAシーケンサの装置全体構成についての説明図。

【図４】静電容量を検知するノズルを用いて対物レンズ下に純水を供給する方法についての説明図。

【図５】静電容量を検知するノズルおよび対物レンズ先端の電通センサを用いて対物レンズ下に純水を供給する方法についての説明図。

【図６】静電容量を検知する、先端が環状であるノズルおよび対物レンズ先端の電通センサを用いて対物レンズ下に純水を供給する方法についての説明図。

【図７】静電容量を検知する対物レンズを用いて対物レンズ下に純水を供給する方法についての説明図。
【図８】対物レンズ先端の電通センサおよびフローチップ表面に塗布した透明電極膜を用いて対物レンズ下に純水を供給する方法についての説明図。
【図９】静電容量を検知するノズルを用いて対物レンズ下に純水を供給する方法についての説明図。
【図１０】静電容量を検知する環状ノズルおよび通電センサを用いて対物レンズ下に純水を供給する方法についての説明図。
【図１１】TDI方式におけるスキャン方式についての説明図。
【図１２】液浸対物レンズを用いて基板上に光学分解能の間隔で格子状に配置した高密度DNA nanoballを検出する方法についての説明図。
【発明を実施するための形態】
【０００９】
（乾燥系対物レンズ、および液浸対物レンズについて）
まず、乾燥系対物レンズ、および液浸対物レンズについて説明する。乾燥系対物レンズと液浸（例えば、水浸）対物レンズを用いた顕微鏡計測方法について図1を用いて説明する。
【００１０】
光軸109に垂直に乾燥系対物レンズ104が配置されている。乾燥系対物レンズ104が集光できる最大角度は下記の式で計算することができる。
$NA = n \sin\theta$
【００１１】
ここでNAは対物レンズの開口数であり、対物レンズの種類により異なる。図１（ｃ）に示すように乾燥系対物レンズ104の開口数を0.75とすると、乾燥系対物レンズ104と上板ガラス101の間の媒質は空気であり、その屈折率は１である。これより乾燥系対物レンズ104が集光可能な最大角度θ１は48.6度と算出することができる。この角度は図１（a）のθ１で表現される。図１（ａ）は本発明の比較例である。図１（ｂ）に示すように、水浸対物レンズ114は上板ガラス111の間の空間は純水117で満たされている。図1（c）に示されるように水浸対物レンズ114の開口数は乾燥系対物レンズ101より大きく、1.2である。水浸対物レンズ101の集光可能な最大角度θ2は64.5°と算出することが可能となる。これらの角度はそれぞれの対物レンズが回収できる光線の最大角度を表現したものである。従って液浸対物レンズの集光率は乾燥系よりも優れており、より鮮明が画像を取得できる。また、大きいNAのため、分解能についても液浸対物レンズのほうが優れていることが分かっている。
【００１２】
一方、計測対象となるフローチップは上板ガラス111と下板ガラス112で溶液116を挟み込むことで流路を形成する。本実施例では上板ガラス111の溶液116に接するガラス面に観察対象となる蛍光物質を含むDNA分子113が固定されている。
【００１３】
DNA分子113に適当な励起光が照射されると、DNA分子113に保持された蛍光色素が発光する。蛍光色素より発せられる光は等方的に発光する。乾燥系対物レンズ114方向に発光した光は上板ガラス111に入射し、更に乾燥系対物レンズ114と上板ガラス101との間に存在する空気に入射する。ここで、乾燥系対物レンズ114が集光できる光は少なくとも乾燥系対物レンズ114方向へ進む光であり、フローチップを横方向に進行する光は集光することができない。これは$n \cdot \sin\theta = 1.33 \cdot \sin 90° = 1.33$の光は集光できないということである。従って集光に寄与できる光の開口数は1.33未満であり、これは蛍光物質の環境として存在する純水の屈折率に限定されるということである。
【００１４】
（水浸対物レンズと油浸対物レンズについて）
次に、水浸対物レンズを用いてフローセルの上面と下面を計測する方法について説明す

る。フローセルの上板ガラス201と下板ガラス204の間には溶液209が挟まれて流路を形成している。溶液209の屈折率は純水と同じ1.33である。従って図２（a）で示される水浸対物レンズ202先端から上板ガラス201の下面までの光路長と、図２（b）で示される下板ガラス209上面までの光路長は等しい。これは純水で満たされた空間で同じ厚さのガラス板を上下に移動させても、対物レンズから計測面までの距離は不変であるということである。

【００１５】

従って図２（a）と図２（b）の光学性能は全く同一となる。これは液浸媒質と流路内の溶液の屈折率が同一であることにより達成される。換言すれば、油浸対物レンズを使用した場合には図２（a）と図２（b）の場合におけるそれぞれの観察面までの光路長は異なる。油浸対物レンズの場合図２（a）で示される状態で対物レンズが設計されているため、観察対象が上板ガラス206下面よりも10μｍ離れると、画像がぼけてしまう。従って水溶液系の液体を保持した流路内の上面と下面の生体物質の挙動を計測する場合、水浸レンズの採用がより好ましい。

【実施例１】

【００１６】

本発明の第１の実施例として、液浸対物レンズを用いてDNAの塩基配列解析を行うDNAシーケンサの装置（分析装置）構成および計測方法（分析方法）について説明する。

【００１７】

フローチップ301はヒートブロック307にエアーチャックで固定される。ヒートブロック307の下面にはペルチェ素子305が配置され、フローセル301の温調を行う。温度制御範囲は4～95℃である。温度制御はフローチップ301で酵素反応による塩基伸長、伸長の足場となるプライマの解離などに必要となる。ヒートブロック307内部には温度センサとして測温抵抗体331が配置され、温度制御のフィードバックに使用される。

【００１８】

ヒートシンク306はペルチェ素子305に密着し、ペルチェ素子305の駆動に伴って発生した熱を放熱する。ヒートシンク306からの放熱は冷却された空気を庫内で循環することにより達成される。酵素、4種類の蛍光試薬、バッファ、ヌクレオチド、洗浄液などは冷却庫308に設置され、4℃に冷却される。冷却庫308は装置内のヒートシンク306に冷気を循環し、排熱が可能なように風路が形成されている。風路内には冷気の循環を促進するファン309が配置される。

【００１９】

風路内を循環し、ヒートシンク306で発生した熱を再び冷却庫308まで運ぶことにより効率のいい温調を行うことができる。また、冷却庫308とヒートブロック307の温調を1つの冷却機構で達成できるため、装置の小型化が可能になる。ヒートブロック307およびフローチップ301はXYステージ324に保持されており、フローチップ301を水浸対物レンズ330に対して水平に移動することができる。水浸対物レンズ330はZステージ319に固定され、測定対象物にフォーカスを合わせるために上下に移動することができる。水浸対物レンズ330とフローチップ301の間には純水332が、ノズル314を介して満たされる。ノズル314は静電容量変化を検出することにより純水332の供給量を制御することができる。純水容器310に補充された純水は、チューブ333を介してシリンジ313により送液することができる。流路には逆止弁311および開閉バルブ312が設置され、純水の供給を制御することができる。

【００２０】

DNA鎖の伸長反応はそれぞれ異なる蛍光色素でラベルされた4種類のヌクレオチドおよびポリメラーゼをフローチップ301上で反応させることで行う。各ヌクレオチドはそれぞれFAM-dCTP, Cy3-dATP, Texas Red -dGTP, Cy5-dTsTPである。各ヌクレオチドの濃度は200nMである。また、反応液は伸張反応が効率よく行なわれるように塩濃度、マグネシウム濃度およびpHが最適化されている。反応溶液中にはポリメラーゼが含まれており、DNA断片に相補的な蛍光ヌクレオチドが1塩基だけ取り込まれる。2塩基目の伸長が発生しないのは、1塩基目の蛍光色素がバルキーであるため、2塩基目の取り込みについて立体障害を起こ

すからである。1塩基が取り込まれた後、浮遊する蛍光ヌクレオチドを洗浄により除去した後、蛍光計測を行う。なお、以降の最小ユニットの反応を行うために、蛍光計測後、解離溶液により塩基から蛍光色素を解離させる工程が必須である。この工程により、最小ユニットの反応における蛍光色素の立体障害をリセットすることができる。これにより反応の継続が可能となる。再び蛍光ヌクレオチドをフローセル内に送液し、反応を繰り返すことにより、の逐次的なシーケンスが可能となる。本実施例で採用している反応方式はSequence By Synthesisと呼ばれる。

【００２１】

2つのLED316,317は蛍光色素を励起するための光源である。LED316および317の中央波長はそれぞれ490、595nmである。LED316はFAM-dCTP, Cy3-dATPを、LED317はTexas Red -dGTP, Cy5-dTsTPの励起光照射に用いる。ダイクロイックミラー315はLED316,317からの光を同一光軸上に揃える。更にダイクロイックミラー318により励起光は対物レンズ330の瞳面に入射させられる。対物レンズ330を介して励起光はフローチップ301内の蛍光色素に照射され、蛍光を発する。等方的に発光した蛍光の一部分はフローチップ301、純水332を経て対物レンズ330に回収される。対物レンズ330を経た光は平行光となり、ダイクロイックミラー319まで直進し、分割される。ダイクロイックミラー319は4色の蛍光波長領域について緩やかな反射特性を持つ。このため、CMOSカメラ322および324の受光面にはフローチップ上の反応場から発光した輝点の蛍光強度比をそれぞれ算出することができる。2つのCMOSカメラ結像面上での比をとることによって、この発光点が4色のいずれに帰属するかを判定することが可能となる。なおダイクロイックミラー319で分割された平行光はそれぞれエミッションフィルタ320、325を経た後、チューブレンズ321、323でCMOSカメラ322、324の受光面に結像する。

【００２２】

次に、図３における液浸対物レンズに自動的に液体を供給する装置構成および方法について図４を用いて説明する。

【００２３】

フローチップ401は真空チャックによりヒートブロック419に保持される。フローチップ401の水浸対物レンズ402側のカバーガラス面は疎水性であり、水滴を弾く特性を持つ。これは水浸対物レンズ402に純水403を滴下したときにガラス表面に純水403が広がらず、ある接触角を保持しながら液浸対物レンズ402と接触するためである。また、XYステージを用いてフローチップ401内の異なる視野を観察するときにも純水403が移動時の振動などでガラス表面に拡がってしまうなどの問題を回避するのに有用な性質となる。

【００２４】

ヒートブロック419は内部に流路を持ち、流路内には蛍光色素を含んだDNA断片を保持する。水浸対物レンズ402はZステージ410に保持され、フローチップ401に対するフォーカスを調整することができる。ノズル404はZステージ410により保持され、水浸対物レンズ402について間接的に固定され、水浸対物レンズ402およびノズル404の相対距離は一定に保たれる。

【００２５】

ノズル404は外周を電気シールド405で覆われ、同軸二重管構造となっている。ノズル404は導電性の電気シールド405で覆われているため、検出電極であるノズル404先端部の露出面積が最小にされ、これにより静電容量の検出感度を向上させることができる。また、シールド405は磁気シールドにもなっているため、Zステージ410の駆動によって生じる外界からのノイズに起因する液面検出の誤動作を防止することができる。

【００２６】

純粋容器406には純水が保持され、チューブ433を介してノズル404に接続されている。開閉バルブ409を開けシリンジ408を駆動させることで純粋容器406内の純水を汲みだし、ノズル404に対して純水を送液できる。なお、逆止弁407はシリンジ408で加圧された純水の逆流を防止している。これら一連の動作により、ノズル404は水浸対物レンズ402とフローチップ401に純水を供給することで、空間を純水で満たすことができる。水浸対物レン

ズ402とフローチップ401の間に純水403が満たされると、ノズル404先端とフローチップ401間の静電容量が急激に大きくなる。これにより水浸対物レンズ402先端に純水403が満たされたかどうかについて直接検知できる。なお、より具体的な静電容量変化の検出は、静電容量検出回路411、液面判定回路412、モーター駆動回路414、制御部413より自動化される。なお、フローチップ401は電気的に接地されており、静電容量を計測する場合のグラウンドとなる。また、本実施例では水浸対物レンズ402を用いて説明を行ったが、これは水浸に限定されるものではなく、油浸対物レンズにも適用することが可能である。静電容量は空間を満たす物質の誘電率に依存するため、本実施例の場合、空気と異なる誘電率を持ち物質であれば、原理的に検出可能である。従ってシリンジ408およびノズル404内を満たす液体は油でもよく、原理的には空気と異なる誘電率を持つ物質であれば、検出が可能である。

【実施例２】

【００２７】

本発明の分析装置第２の実施例として、液浸対物レンズに自動的に液体を供給する装置構成および方法について説明する。本実施例では対物レンズのレンズ周辺部に複数の通電センサを配置することで、対物レンズとフローチップ間における液体の供給精度を高めるものである。液浸対物レンズ502には先端のレンズ周辺に4つの通電センサ521が配置されている。なお、純水517に接触する前の通電センサ間の通電を防止するため、水浸対物レンズ502のレンズ以外の先端部526は絶縁状態になっている。

【００２８】

フローチップ501と水浸対物レンズ502との間の静電容量を計測することによりノズル504は純水517を適量供給することができる。しかし、水浸対物レンズ502先端のレンズが純水517に完全に浸されたかどうかについては正確に判定することができない。もしもレンズの一部のみが純水517と接触し、他の一部が純水517ではなく空気と接触している場合には、屈折率が異常であるため良好な画像を取得することができない。従ってレンズ全体が純水517と接触していることが必要となる。本実施例ではこれを達成する手段として4個の通電センサ522，523，524，525を対物レンズ502のレンズ周辺に配置する。これにより、液浸対物レンズ502のレンズ周辺の4点が確実に純水517に接触している場合には、4個のいずれの通電センサ522，523，524，525も通電する。4個全てが通電していない場合は、対物レンズ502と純水517が完全に接触していないことを検知できる。しかし、この手法では液浸対物レンズ502とフローチップ501との接触を保証するものではなく、あくまでも検知しているのは、通電センサ522，523，524，525が純水517と通電状態である。一方、静電容量をノズル504により検知することにより、水浸対物レンズ502とフローチップ501の接触の有無を検知できるが、水浸レンズ502と純水517との接触は直接検知することができない。従って静電容量による検知方法と、通電方法による検知方法を相補的に用いることでフローチップ501と水浸対物レンズ502との間に正確に純水517を供給することができる。

【実施例３】

【００２９】

本発明の分析装置の第３の実施例として、水浸対物レンズに自動的に液体を供給する装置構成および方法について説明する。本実施例では実施例5に加え、液体供給を行うノズルの先端形状を改変した方法について図6を用いて説明する。

【００３０】

ノズル604は先端が環状になっている。装置にノズル604を配置したときにフローチップ601面に向けた下方に純水617を吐出することができる3個の穴を保持する。本実施例ではノズルは環状となっている。ノズルの先端が線状の場合において、フローチップを移動させると水浸対物レンズ602先端の純水617がフローチップに引きずられ、最悪の場合、対物レンズとフローチップの間に空気が混入してしまう。これに対し、ノズル先端を環状部604の形状にすると、フローチップ駆動時においても環状部が純水を保持する土手として働くため、水浸対物レンズ602先端から純水が解離し、空気層が混入するという問題を回避することができる。

【実施例４】
【００３１】
本発明の分析装置の第４の実施例として、水浸対物レンズに自動的に液体を供給する装置構成および方法について説明する。実施例4ではノズルを用いて静電容量を検知しているのに対して、本実施例では水浸対物レンズ702自体を用いて静電容量を検知する。静電容量を計測するセンサとして対物レンズ702を用いる利点を以下に説明する。静電容量の計測の対象は、「水浸対物レンズ702とフローチップ701の間にある純水の有無」である。しかし、図４を用いて示した計測では、「ノズル706とフローチップ701の間にある純水の有無」を検知している。ノズルを液浸対物レンズ702に変換し、静電容量を直接に計測することにより、より正確な計測を行うことが可能となる。この場合、水浸対物レンズ702は静電容量を計測するため、装置本体のグラウンドとは絶縁されている必要がある。そうでなければ、フローチップ701が接地している状態と水浸対物レンズ702が同じ電位になってしまうからである。従って液浸対物レンズ702は、たとえば対物レンズの胴付面のネジ部が樹脂などの絶縁物質で作製されており、装置本体のグラウンドから絶縁されていることが肝要である。また、水浸対物レンズ702はノイズを防止するためにシールドされていることが望ましい。
【実施例５】
【００３２】
本発明の分析装置の第５の実施例として、液浸対物レンズに自動的に液体を供給する装置構成および方法について説明する。本実施例では対物レンズ先端は通電センサを保持し、更にフローチップ801の上板ガラス面には導電性がある透明電極膜（ITO膜）803が塗布されている。この構成において、ノズル804が純水833をフローチップ801の上面である透明電極膜803に滴下する。液浸対物レンズ802とフローチップ801間に純水が満たされたかどうかについては、通電センサ821、823、824、825と透明電極803の間の通電状態を検知することで確認できる。また、純水833が液浸対物レンズ802の先端部を浸したかどうかについても上述した方法で同様に確認することができる。
【００３３】
（液体の供給について）
液浸対物レンズに自動的に液体を供給する装置構成および方法についてより図9を用いて詳細に説明する。
【００３４】
ここでノズル905は静電容量を検知することができる。
a)ではフローチップ902はヒートブロック901上に真空チャックで固定されている。
b)において、フローチップ902に対して液浸対物レンズ904がZステージにより駆動され、所定の位置まで下方に移動する。
c)ではノズル905とフローチップ902間に純水906の供給が開始される。ここではフローチップ902とノズル905先端は接触していないため、静電容量は予め定められたしきい値を超えず、純水906は供給されていないと判断する。
d)において純水907の供給は継続され、純水907はフローチップ902表面に接し、フローチップ902と対物レンズ904の間の静電容量が急激に増大する。静電容量はしきい値を超え、ノズル905は純水907の供給を停止する
e)表面張力で対物レンズ904に吸着していた液滴907は、フローチップ902と接触すると、水特有の強い表面張力のため、対物レンズ904とフローチップ902の間を満たしたまま、フローチップ902表面に落下し、拡がる。このとき、ノズル905先端が純水909と接触できなくなった場合において、ノズル905は純水909の供給を再開する。純水909がノズル905と接触するまで純水909の供給を継続する。純水909がノズル905と接触すると静電容量がしきい値を超え、ノズル905は純水909の供給を停止する。なお、純水909の形状を安定させるためにXYステージを用いてフローセルをステップ状に駆動させることも可能である。
f)光学計測を長時間継続して行うと、次第に純水910の蒸発が発生する。蒸発が進行すると、純水910とノズル905の接触が離れ、ノズル905は静電容量の急激な低下を検知する。

これを受け、静電容量がしきい値を超えるまでノズル905は再度純水の供給を行う。これにより、たとえば一週間にわたる長期の計測においても自動で液浸対物レンズ904直下に純水910を供給することで、液浸対物レンズ904を用いた自動計測を行うことが可能となる。また、本実施例で説明した方式は液浸対物レンズ904に限定されるものではなく、油浸対物レンズでも同様に行うことができるものである。

液浸対物レンズに自動的に液体を供給する装置構成および方法についてより図10を用いて詳細に説明する。

【００３５】

本実施例では線状のノズルの代わりに実施例6で用いた環状ノズル1005を用いて純水を対物レンズ1004下に供給する。また、液浸対物レンズ1004先端には通電センサ1006が4個配置されている。また、液浸対物レンズ1004先端は絶縁処理されている。第9の実施例と同様にa)からb)で対物レンズがフローチップ1001に対してフォーカスする位置まで降下する。c)で環状ノズル1005が純水1010をフローチップ1001方向に滴下し始める。なお、実施例6で説明したように、環状ノズル1005には3個の吐出口を保持するため、滴下される液滴の数は3つとなる。d)において純水1007はフローチップ1001と接触し、次第に形状を大きくする。e)において純水1009は液浸対物レンズ1004上の通電センサ1006および環状ノズル1005と接触する。これによりフローチップ1001と液浸対物レンズ1004の間に純水1009が存在することと、液浸対物レンズ1004の先端が純水1009内に浸っていることが確認される。f)において純水の蒸発が進行すると、環状ノズル1005と純水1010の接触が断たれ、純水1010の量が足りないことが検知される。環状ノズル1005は純水の供給を再開し、環状ノズル1005が再び純水1009と接触するまで供給を継続する。これを持って純水1010の乾燥にも関わらず安定した長時間にわたる計測の継続が可能となる。

【００３６】

液浸対物レンズに自動的に液体を供給する機構を用いて、TDI（Time Delay Integration）方式によりフローチップ表面をスキャンする装置構成および方法についてより図11を用いて詳細に説明する。

【００３７】

TDI（Time Delay Integration）は、CCDの特殊な読み出し方式で、一定速度で移動する対象物に対して、その移動方向・速度とCCDの電荷転送方向・速度を合わせた撮像を行う手法である。高速性に加え、移動する対象物を繰り返し（CCDの垂直段数分）露光することで高い感度を得ることができる。このため、TDI方式はエリアセンサではなく、ラインセンサを採用する。

【００３８】

また、従来のエリアセンサではフローチップ流路面に固定された反応場の蛍光計測について、ステップアンドリピートを繰り返す必要があった。この方式ではステージ駆動後にステージの振動が収まるまでの90～200ミリ秒の間、像ブレが発生するため、蛍光計測ができないという問題があった。この問題は逐次的なXYステージ停止を伴わないTDI方式では、XYステージの停止振動問題は発生しない。従って、ステージ停止後の振動が減衰する時間を省略し、連続して計測を継続することが可能となる。結果としてスループットが増大するという効果をもたらす。

【００３９】

本実施例では開口数0.95、20倍の液浸対物レンズ1105を用いる。開口数0.95の場合の分解能は0.35μm杜なる。また、視野数は26.5mmである。この場合サンプル面での観察領域はφ1.325mmとなる。

また、フローチップ610内には流路611が4レーンある。流路の形状は80×3.5mmである。また、使用するTDIカメラは4096×24画素を持つ。また画素サイズは6×6μmである。従ってサンプル面に投影されるラインセンサの大きさは1.2×0.077mmとなり、対物レンズの観察領域であるφ1.325mmをほぼカバーできる形状となる。流路611は3回のスキャン動作で全領域をカバーできることになる。また、サンプル面上での画素サイズは0.3×0.3μmとなる。また、TDIカメラは実施例3で詳述したように2個のTDIカメラを用いて、流路内より発

する4種類の蛍光色素の判別を蛍光強度比より行う。またTDIカメラは背面照射タイプの撮像素子構造を持ち、ピクセルクロックは30MHzである。仮にTDIカメラ常時動作した状態で計測が行え、かつ反応場の平均密度が1個/25ピクセルと仮定すると、理論上想定されるスループットは30Mピクセル/秒÷25ピクセル/base×60秒×60分×24時間/日=100Gbase/日となる。

【００４０】

より詳細な計測方法を以下に述べる。フローチップ1110はヒートブロック1101上にバキュームチャックされる。液浸対物レンズ1105とフローチップ1110の間には静電容量による液面検知機能を有するノズル1107を介して純水1106が供給される。フローチップ1101は流路1111の端から端までXYステージにより連続的に停止することなく駆動される。XYステージ駆動あるいは時間経過による純水1106の蒸発により、フローセル1110上の純水1106の量が減少した場合は、ノズル1108はしきい値加減より静電容量が低下したことを検知し、速やかに純水1106を供給する。また、移動に伴って純水1106がフローチップ1110表面に残存し、TDI動作の動作時にそれらの純水が再結合し、純水1106の量が増大し、静電容量がしきい値上限を超えた場合には、純水1106の回収を行う。液面検知機能を保持した純水供給ノズル1107により、液浸対物レンズ1105を用いて円滑にTDIラインスキャンによる蛍光画像取得を行うことができる。

【００４１】

環状ノズル1107は純水1106を保持する効果も持つ。純水1106がフローチップ1110の移動方向とは逆方向に引きずる表面張力が発生する場合でも、環状ノズル1107が外周方向より純水1106を保持しようとする力が働くため、TDIのような連続的な動作環境下でも本実施例で説明した液浸対物レンズによるTDI計測法は問題なく機能する。

【００４２】

基板に格子状に高密度にDNA分子を固定し、液浸対物レンズを用いてシーケンシングする方法について図12を用いて以下に説明する。

【００４３】

基板1201にはシリコンウェハより切り出した、25×75mmのスライドガラスの形状を持つ。基板1201表面には直径300nmのアミノシラン修飾されたDNA結合面1202が390nmピッチで格子状に規則正しく形成されている。アミノシランはDNA分子と効率よく吸着することが知られており、DNAチップなどの表面修飾に積極的に用いられている。また、DNA非結合面1203はDNAと吸着しない性質を持つHMDS(=hexamehyldisilzane)で修飾されている。従ってDNA分子を基板上に注入すると、DNA分子はDNA結合面1204にのみ積極的に結合し、DNA非結合面1203には結合しない。現在の半導体技術では45nmの線幅を露光できるため、上述したDNA結合面1202を390nmピッチに格子状に保持する基板1201の製造は容易に達成することができる。

【００４４】

また、Science, vo327, pp78, 2010に詳しいように試験管内で直径300nmのDNAナノボール1202を2×1010個/mLの濃度で合成することが可能である。これらのDNAナノボール1202はそれぞれ固有の1個のDNA断片が増幅されているため、DNAナノボール1202にプライマを結合させ、塩基伸長を逐次的に進行させることで塩基配列解析が可能であることも知られている。また、DNAナノボール1202は直径300nmの直径を保持している一方、基板1201上で結合できるDNA結合面1204の直径も同様に300nmである。従って１つのDNA結合面1204に2つのDNAナノボール1202は体積排除効果により、結合することができない。つまり、１つのDNA結合面1204には1つのDNAナノボール1202のみ結合する。従ってDNAナノボール1202を基板上に350nmピッチに最密充填した形で光学計測を行うことができる。

【００４５】

本実施例では開口数1.2である60×の液浸対物レンズを採用する。開口数1.2の対物レンズの分解能はδ=0.61×550nm/開口数で計算することができ、この場合分解能は280nmであることが示される。従って390nmピッチで基板に配置されたDNAナノボール1202を光学的に識別することが可能である。乾燥系対物レンズの開口数は最大でも0.9程度であるため、

この場合の分解能は370nmであるため、390nmのピッチで配置されたDNAナノボール1202を光学的に識別できない。従って直径300nmのDNAナノボールを390nmピッチで格子状に配置した場合に光学検出を行うためには、液浸対物レンズを使用することが必須となる。また、開口数0.75の乾燥系対物レンズと比較して開口数1.2の液浸対物レンズを用いると、立体角による集光効率を0.09から0.28まで上昇させるため、信号強度を3.2倍まで高めることができる。これは塩基配列の解析精度向上に寄与する。

【００４６】

光信号を検出するセンサとして、画素サイズ6.5μm、2048×2048画素数を持つCMOSカメラを採用する。センサによる検出を行う。一般に異なる反応場を精度よく判別にするためには反応場の大きさに辺方向について2画素以上を割り当てることが望ましい。本実施例においては、サンプル面に投影される画素サイズは0.11μm/ピクセルとなり、反応場の間隔である350nmについて3ピクセルを割り当てることができるため、良好な光学検出を行うことが可能となる。液浸対物レンズの視野数は26.5mmであるため、サンプル面での光学的に計測可能な領域は26.5/60=φ440μmとなる。サンプル面に投影されるCCD面の大きさは222μm角であるため、全てのCCD面を検出に用いることが可能である。従って0.35μmピッチではCCD一視野あたり403×103の反応場を計測することが可能となる。一反応場あたり80baseの塩基伸長が可能であると仮定すると、一視野あたり32Mbaseの塩基配列解析が可能となる。FOVあたりに励起光を照射する領域を440μm /√2として312μm、かつスライドガラス面の80%の面積にナノボールを固定できると仮定する。この場合、32Mbase/FOV × 12000FOV/slide ≒ 400Gbase/slide程度の読取塩基数を得ることができる。さらに流路の上面および下面にナノボールを固定できるとすると読取塩基数は800Gbase/slideとなる。

【符号の説明】

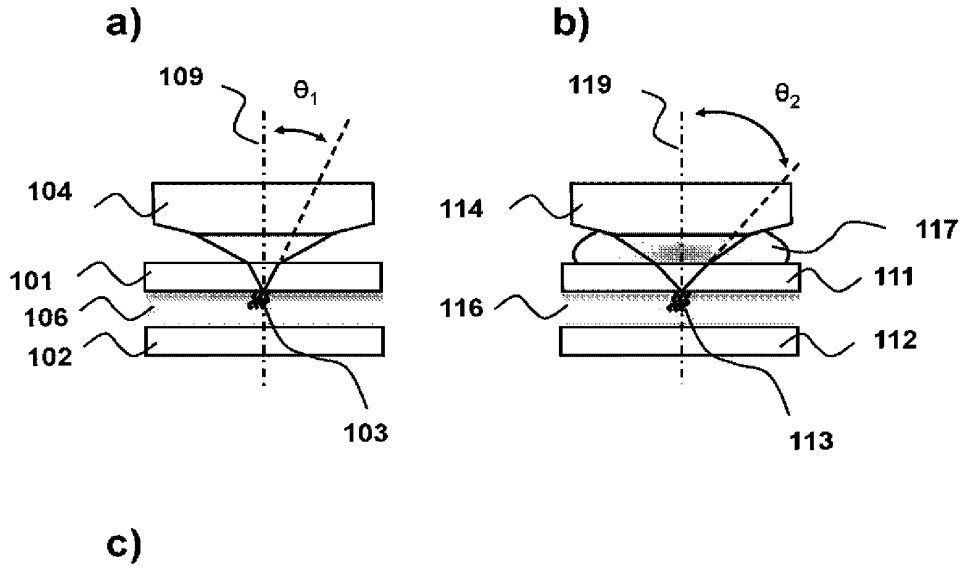
【００４７】

101, 111, 201, 207…上板ガラス
102, 112, 204, 209…下板ガラス
103, 113…DNA分子
104…乾燥系対物レンズ
109…光軸
106, 116, 204, 208…溶液
117, 203, 206,332, 403, 517, 617, 906, 909, 910, 1007, 1009, 1010, 1106…純水
115, 202, 209, 330, 402, 502, 702, 802, 904, 1004, 1105…水浸対物レンズ
301, 401, 501, 601, 701, 801, 902, 1001, 1110…フローチップ
307, 419, 901, 1101…ヒートブロック
305…ペルチェ素子
306…ヒートシンク
308…冷却庫
324…XYステージ
319, 410…Zステージ
314, 404, 504, 706, 804, 905,…ノズル
310, 406…純水容器
333…チューブ
313, 408…シリンジ
311, 407…逆止弁
312, 409…開閉バルブ
316, 317…LED
318, 319…ダイクロイックミラー
322, 324…CMOSカメラ
320, 325…エミッションフィルタ
321, 323…チューブレンズ
405…電気シールド

411…静電容量検出回路
412…液面判定回路
413…制御部
414…モーター駆動回路
521,522, 523, 524, 525, 821, 823, 824, 825, 1006…通電センサ
803…透明電極膜
907…液滴
604,1005, 1107…環状ノズル
1111…流路
1202…DNAナノボール
1204…DNA結合面
1203…DNA非結合面
1201…基板

【図１】

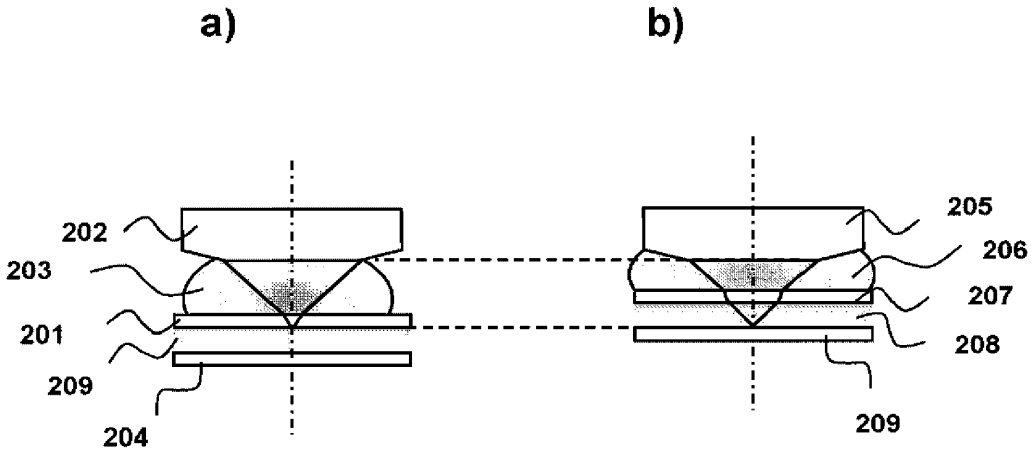
図１



対物名称	開口数	媒質の屈折率	集光角度[°]	信号比	分解能[μm]
乾燥系対物レンズ	0.75	1.00	48.6	1.00	0.45
水浸対物レンズ	1.2	1.33	64.5	3.27	0.28

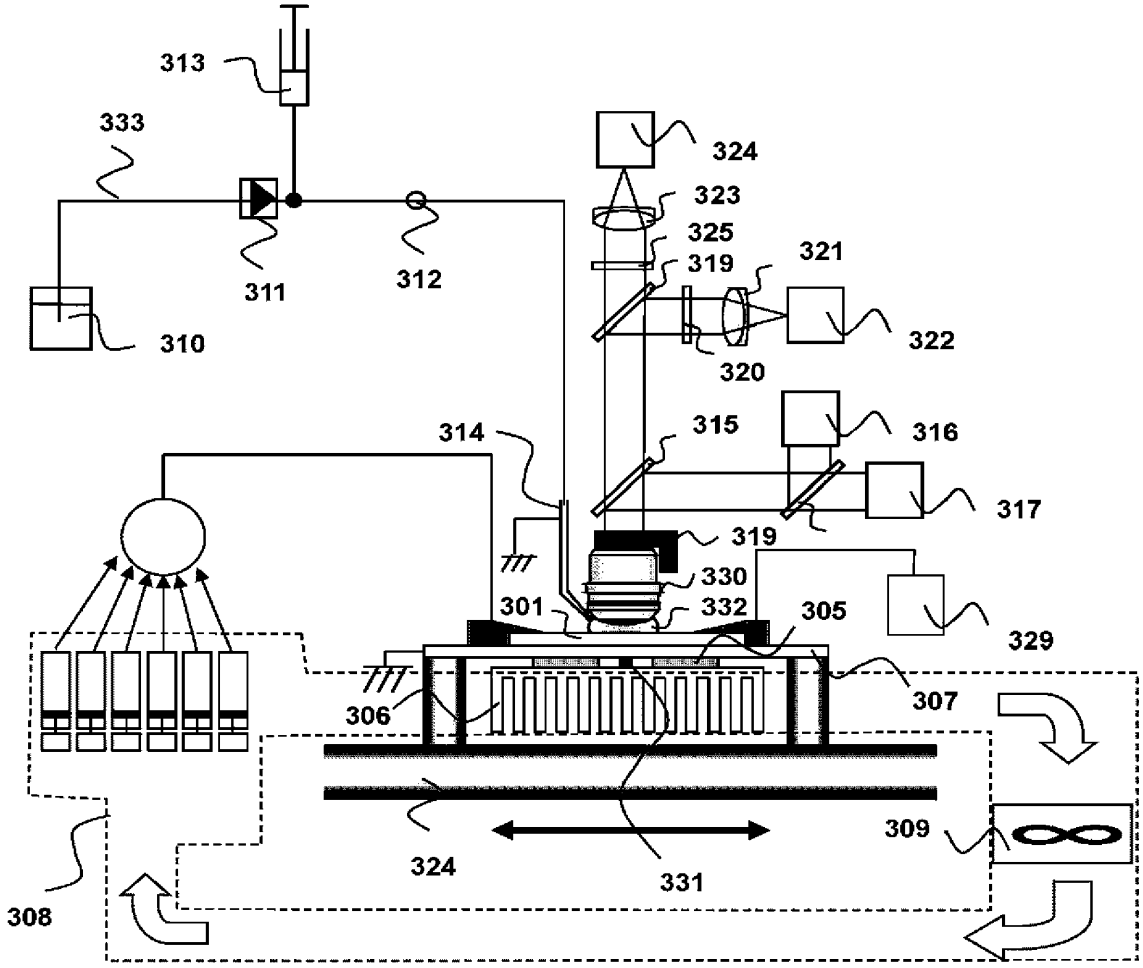
【図２】

図２



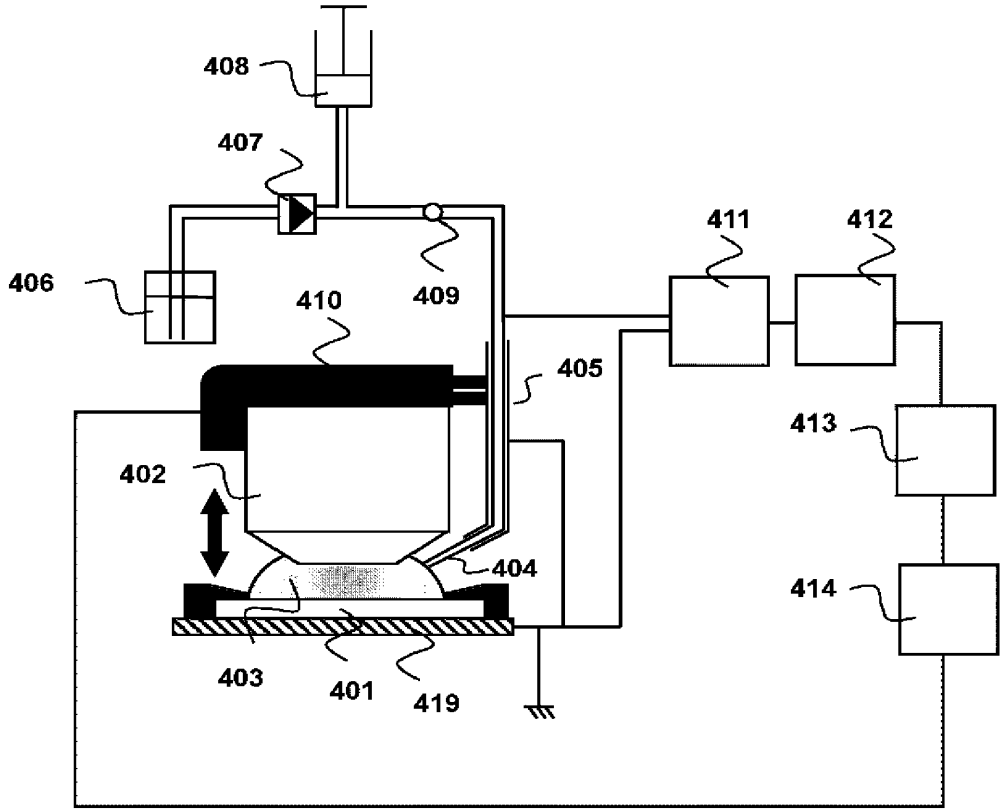
【図３】

図３


313
333
311
312
310
324
323
325
319
321
322
320
315
316
317
314
319
330
332
301
305
329
306
307
309
324
331
308

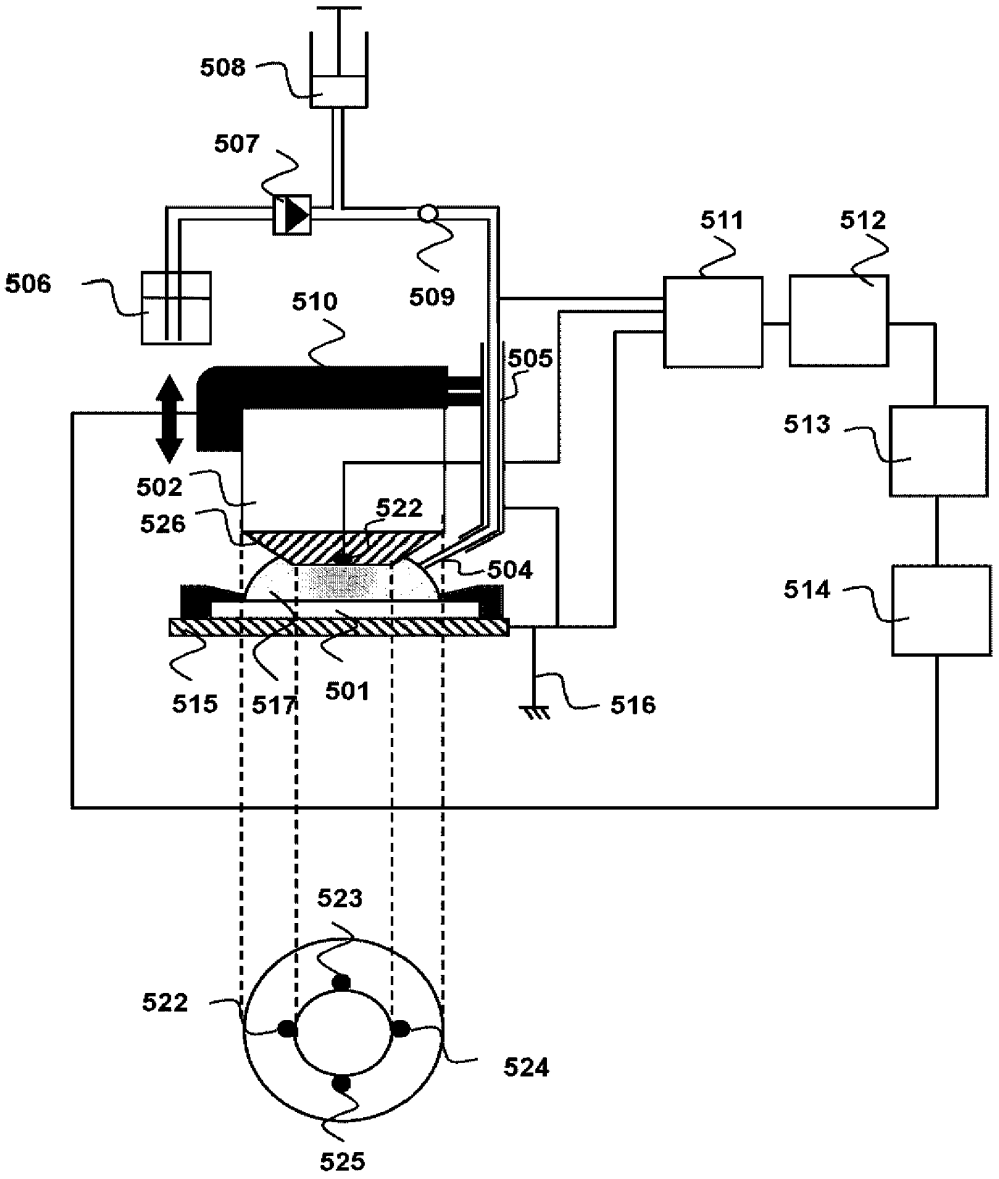
【図４】

図４



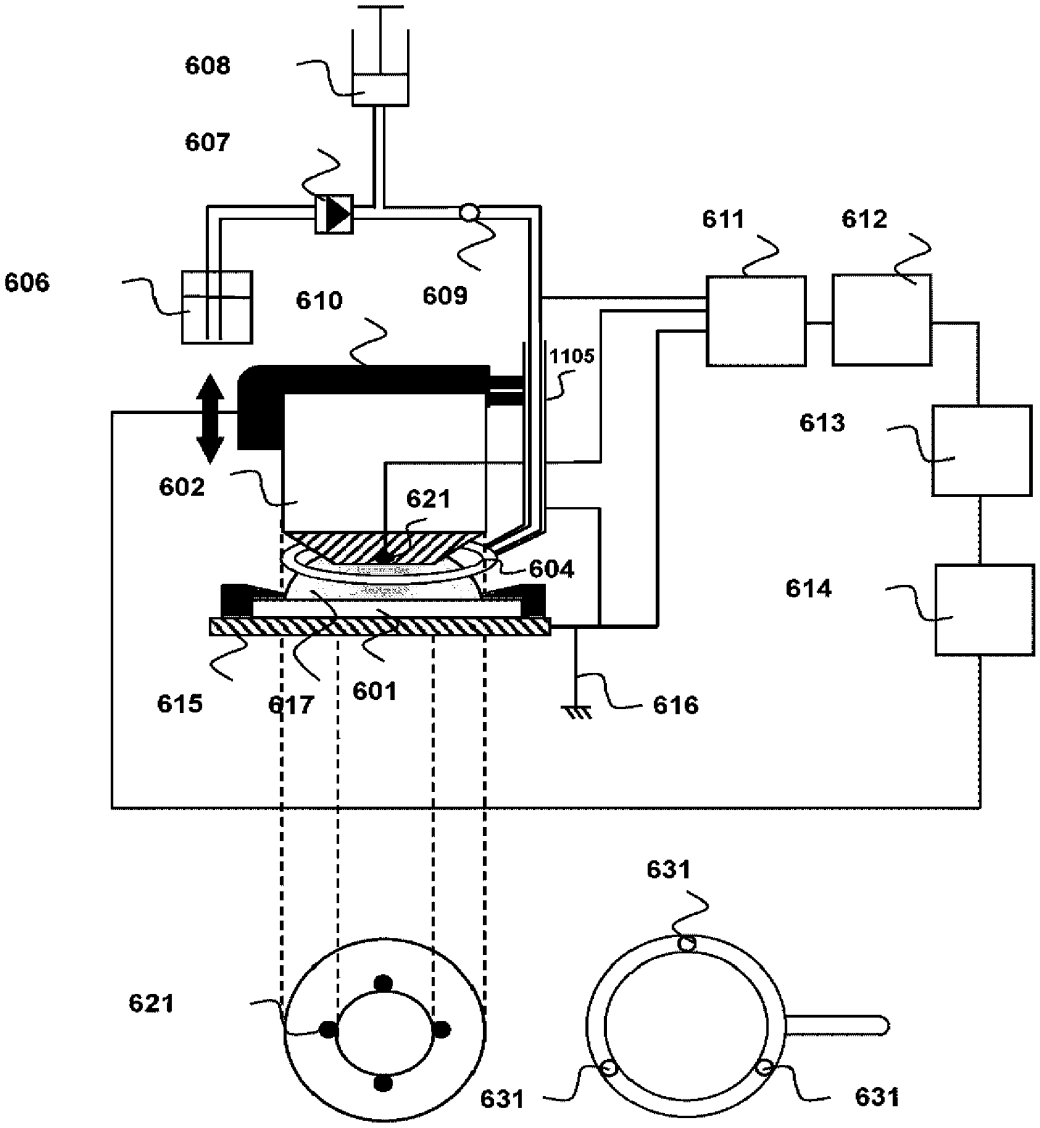
【図５】

図５
（＋電通）


508
507
506
509
510
511
512
505
513
502
526
522
504
514
515
517
501
516
523
522
524
525

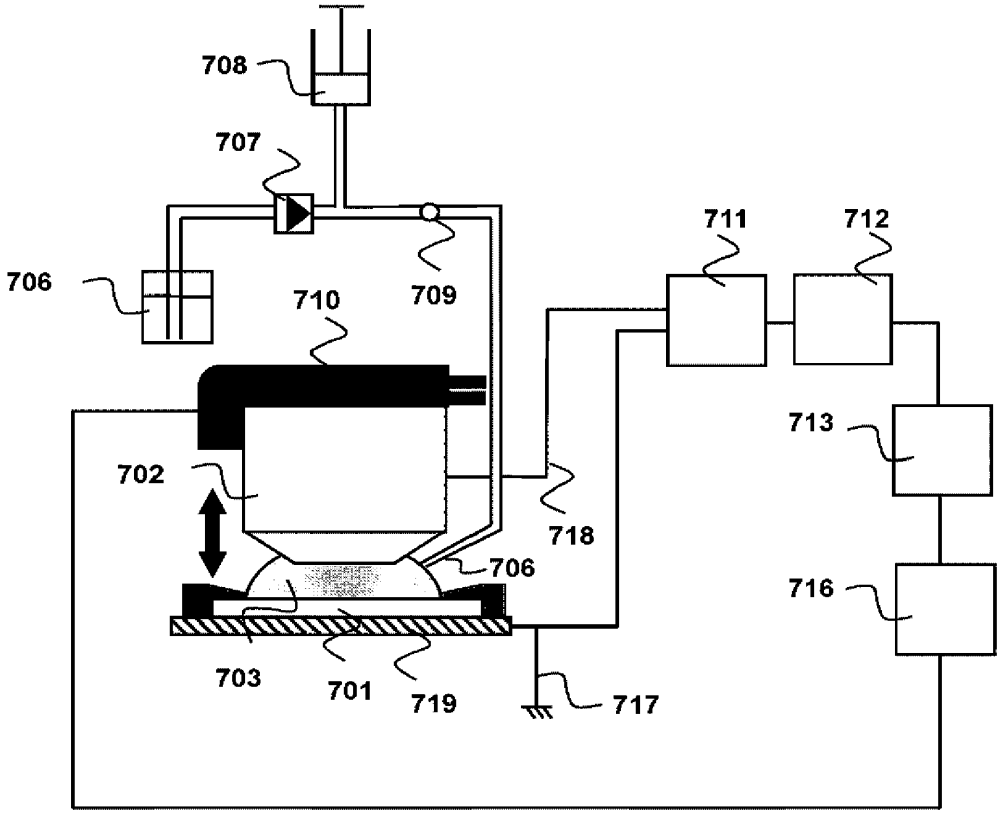
【図６】

図６
（＋電通+環状ノズル）



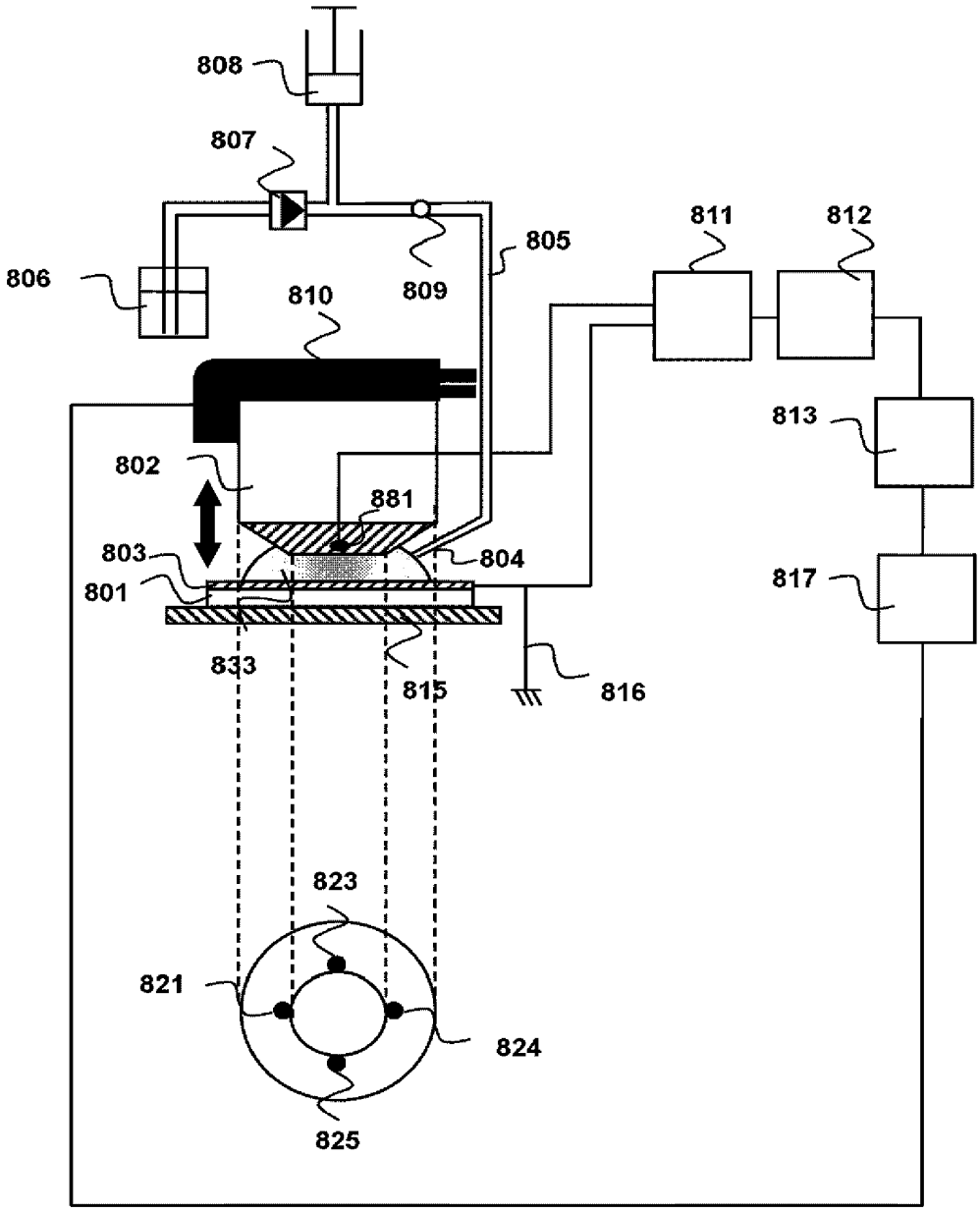
【図７】

図７
（対物静電容量液面検知）



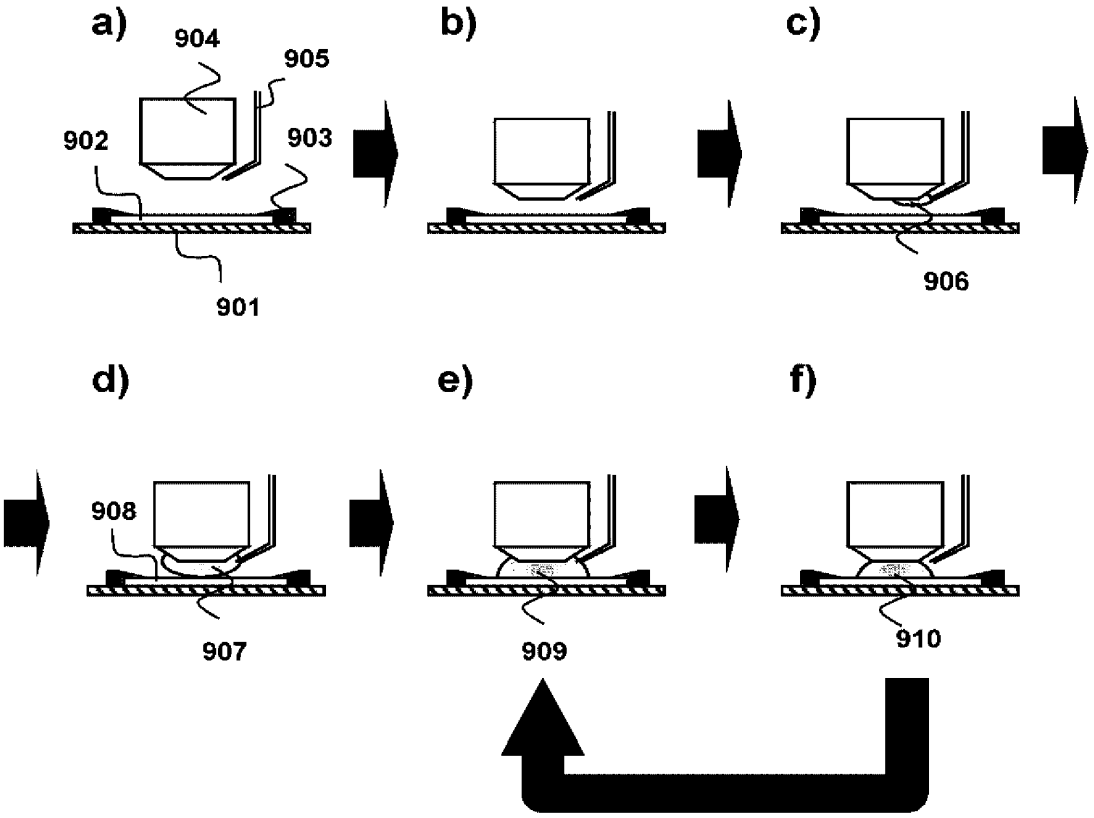
【図８】

図８
（導通法含む）


808
807
806
805
809
810
811
812
813
802
881
803
804
801
817
833
815
816
823
821
824
825

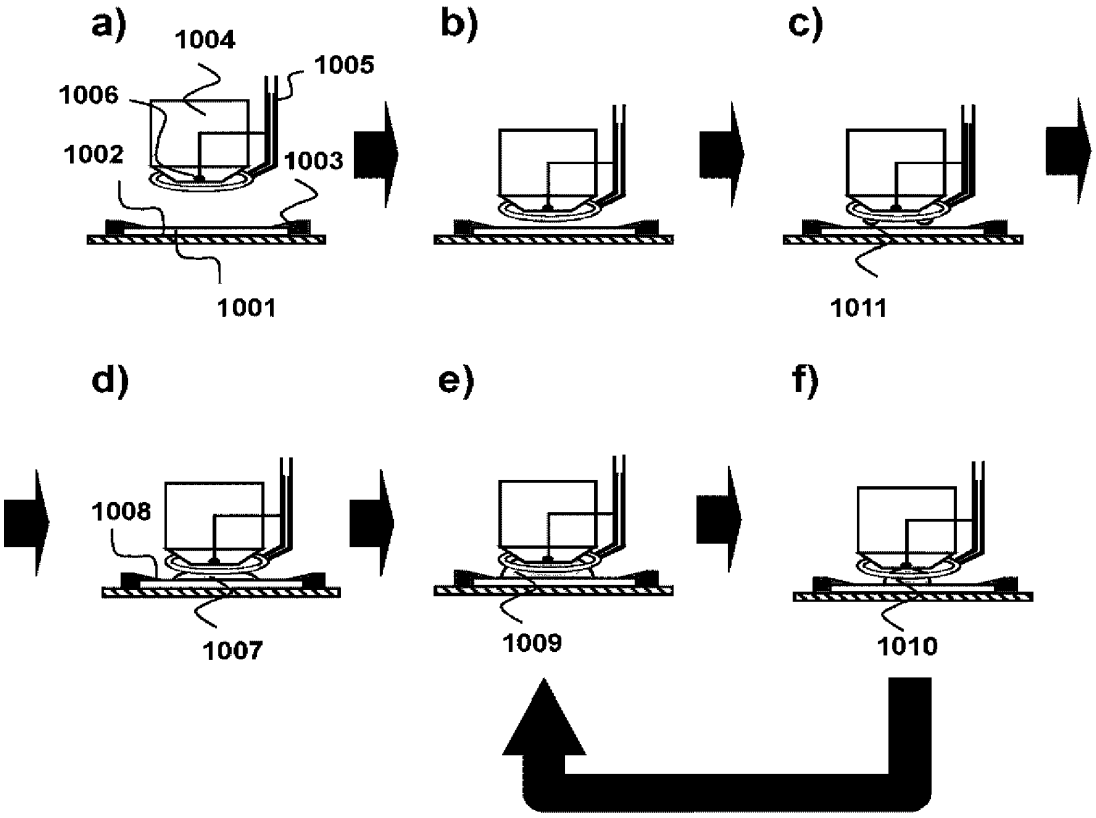
【図９】

図９



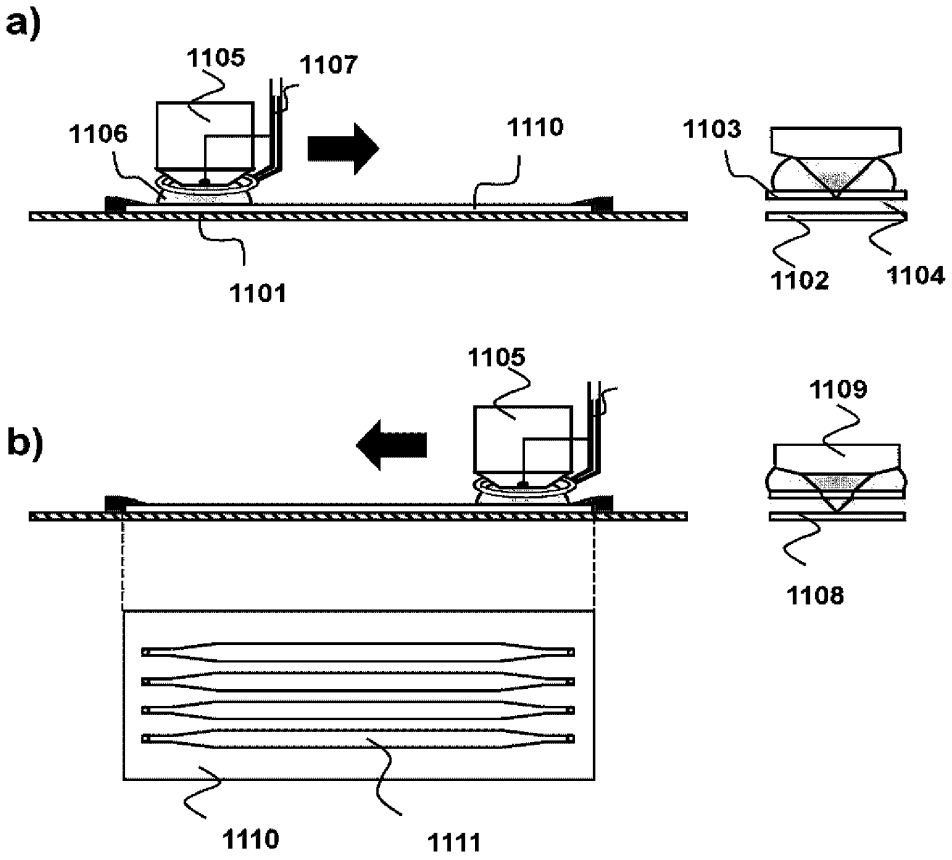
【図１０】

図１０


a)
1004
1005
1006
1002
1003
1001
b)
c)
1011
d)
1008
1007
e)
1009
f)
1010

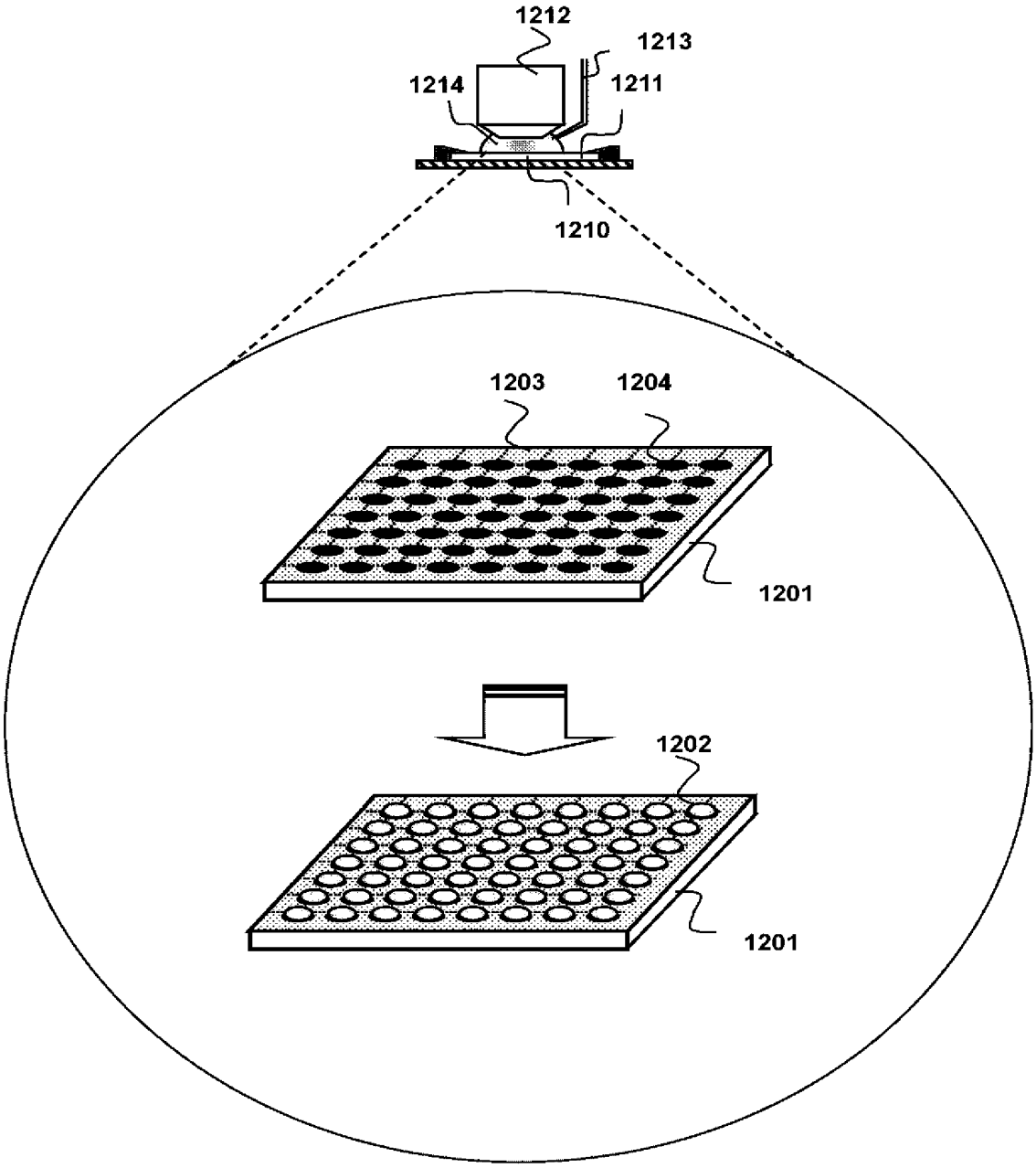
【図１１】

図１１


a)
1105
1107
1106
1110
1103
1101
1102
1104
b)
1105
1109
1108
1110
1111

【図１２】

図１２



フロントページの続き

(72)発明者 庄司 智広
東京都港区西新橋一丁目２４番１４号 株式会社 日立ハイテクノロジーズ内
(72)発明者 山本 周平
東京都港区西新橋一丁目２４番１４号 株式会社 日立ハイテクノロジーズ内
(72)発明者 平塚 稔章
東京都港区西新橋一丁目２４番１４号 株式会社 日立ハイテクノロジーズ内

Ｆターム(参考) 2G043 AA01 BA16 CA03 EA01 GA04 GB01 GB05 GB12 GB16 HA01
HA09 JA02 LA03 MA03
2H052 AA09 AB02 AB17 AC04 AC05 AC14 AC20 AD20 AD23 AD27

Espacenet

Bibliographic data: JP2015090458 (A) — 2015-05-11

ANALYZER

Inventor(s): KATO KOICHI; SHOJI TOMOHIRO; YAMAMOTO SHUHEI; HIRATSUKA TOSHIAKI ± (加藤 宏一, ; 庄司 智広, ; 山本 周平, ; 平塚 稔章)

Applicant(s): HITACHI HIGH TECH CORP ± (株式会社日立ハイテクノロジーズ)

Classification: - international: ***G01N21/64; G02B21/00; G02B21/28; G02B21/36***

- cooperative:

Application number: JP20130230774 20131107

Global Dossier

Priority number(s): JP20130230774 20131107

Also published as: JP6286183 (B2)

Abstract of JP2015090458 (A)

PROBLEM TO BE SOLVED: To solve a problem that analyzer downsizing is difficult to achieve since a method of optically detecting a liquid on a tip of an immersion objective lens increases cost and needs a plenty of space.SOLUTION: An analyzer includes: an objective lens for observing a sample; a mechanism driving the objective lens in an optical axis direction; a light source illuminating the sample; a photosensor detecting light emission from the sample; a nozzle supplying a liquid between the sample and the objective lens; means for regulating a temperature of the sample; and means for supplying a plurality of reagents to the sample, and the analyzer electrically detects a state of the liquid.

Europäisches Patentamt
European Patent Office
Office européen des brevets

Patent Translate
Powered by EPO and Google

Notice

This translation is machine-generated. It cannot be guaranteed that it is intelligible, accurate, complete, reliable or fit for specific purposes. Critical decisions, such as commercially relevant or financial decisions, should not be based on machine-translation output.

DESCRIPTION JP2015090458A

Analysis equipment

[0001]

The present invention relates to an analyzer. More specifically, the present invention relates to an analyzer that automatically supplies a liquid between a sample and an objective lens when performing optical measurement using an immersion objective lens.

[0002]

Patent Document 1 describes a method of automatically supplying a coupling agent such as pure water or oil between an objective lens and a sample. This patent retains a mechanism for injecting a liquid, which is a coupling agent, between the objective lens and the sample. In particular, when the liquid is pure water, there is a problem that the pure water at the tip of the objective lens dries in the measurement over a long period of time.

[0003]

<patcit num="1"> <text> 特開２０１０－２６２１８号公報 </text> </patcit>

[0004]

So far, the detection of liquid at the tip of the immersion objective lens has been performed using optical measurement. However, the optical detection method is expensive, and it is difficult to miniaturize the device because it takes a large space.

[0005]

More specifically, the problems in optical detection will be mentioned.

The infrared emission optical system shown in Patent Document 1 must be provided with an incident optical path and a reflected optical path at an angle of about 60 ° with respect to the optical axis of the objective lens. This requires a large expansion of the space of the optical system using the objective lens in the left-right direction. Therefore, there is a problem that the cost becomes high.

[0006]

The analyzer of the present invention emits light from the sample via an objective lens for observing the sample, a mechanism for driving the objective lens in the optical axis direction, a light source for illuminating the sample, and the objective lens. The liquid has a light sensor for detection, a nozzle for supplying a liquid between the sample and the objective lens, a means for adjusting the temperature of the sample, and a means for supplying a plurality of reagents to the sample. It is characterized by electrically detecting the state of.

[0007]

The analyzer of the present invention can reduce the cost and reduce the size.

[0008]

Explanatory drawing about numerical aperture and resolution in dry objective lens and immersion objective lens. It is explanatory drawing about the measurement method of the upper surface and the lower surface in the flow chip flow path in the immersion objective lens.

An explanatory diagram of the entire device configuration of a DNA sequencer using an immersion objective lens.

An explanatory diagram of a method of supplying pure water under an objective lens using a nozzle that detects capacitance.

An explanatory diagram of a method of supplying pure water under an objective lens using a nozzle for detecting capacitance and a current sensor at the tip of the objective lens.

An explanatory diagram of a method of supplying pure water under an objective lens by using a nozzle having an annular tip and a Dentsu sensor at the tip of the objective lens to detect capacitance.

An explanatory diagram of a method of supplying pure water under an objective lens using an objective lens that detects capacitance.

Explanatory drawing of the method of supplying pure water under an objective lens by using the electric conduction sensor at the tip of an objective lens and the transparent electrode film coated on the surface of a flow chip.

An explanatory diagram of a method of supplying pure water under an objective lens using a nozzle that detects capacitance.

An explanatory diagram of a method of supplying pure water under an objective lens using an annular nozzle for detecting capacitance and an energization sensor.

Explanatory drawing about the scan method in the TDI method.

An explanatory diagram of a method of detecting a high-density DNA nanoball arranged in a grid pattern on a substrate at intervals of optical resolution using an immersion objective lens.

[0009]

(Dry system objective lens and immersion objective lens) First, a dry system objective lens and an immersion objective lens will be described.

A microscope measurement method using a dry objective lens and a liquid immersion (for example, water immersion) objective lens will be described with reference to FIG.

[0010]

The dry objective lens 104 is arranged perpendicular to the optical axis 109.

The maximum angle that the dry objective lens 104 can collect can be calculated by the following formula.

NA = nsinθ

[0011]

Here, NA is the numerical aperture of the objective lens and varies depending on the type of the objective lens.

As shown in FIG. 1 (c), assuming that the numerical aperture of the dry objective lens 104 is 0.75, the medium between the dry objective lens 104 and the upper plate glass 101 is air, and its refractive index is 1.

From this, the maximum angle θ1 that the dry objective lens 104 can collect can be calculated as 48.6 degrees.

This angle is represented by θ1 in FIG. 1 (a).

FIG. 1A is a comparative example of the present invention.

As shown in FIG. 1 (b), the space between the upper glass 111 of the water-immersed objective lens 114 is filled with pure water 117.

As shown in FIG. 1 (c), the numerical aperture of the water-immersed objective lens 114 is larger than that of the dry objective lens 101, which is 1.2.

The maximum condensing angle θ2 of the water immersion objective lens 101 can be calculated as 64.5 °.

These angles represent the maximum angles of light rays that can be collected by each objective lens.

Therefore, the focusing rate of the immersion objective lens is superior to that of the dry system, and a clearer image can be obtained.

It is also known that the immersion objective lens is superior in resolution because of its large NA

[0012]

On the other hand, the flow chip to be measured forms a flow path by sandwiching the solution 116 between the upper plate glass 111 and the lower plate glass 112.

In this embodiment, the DNA molecule 113 containing the fluorescent substance to be observed is fixed on the glass surface of the upper glass 111 in contact with the solution 116.

[0013]

When the DNA molecule 113 is irradiated with an appropriate excitation light, the fluorescent dye retained in the DNA molecule 113 emits light.

The light emitted from the fluorescent dye emits isotropically.

The light emitted in the direction of the dry objective lens 114 is incident on the upper plate glass 111, and further incident on the air existing between the dry objective lens 114 and the upper plate glass 101.

Here, the light that can be focused by the dry objective lens 114 is at least the light that travels in the direction of the dry objective lens 114, and the light that travels laterally through the flow chip cannot be focused.

This means that light with n · sin θ = 1.33 · sin 90 ° = 1.33 cannot be focused.

Therefore, the numerical aperture of light that can contribute to light collection is less than 1.33, which means that it is limited to the refractive index of pure water that exists as an environment for fluorescent substances.

[0014]

(Regarding the water-immersed objective lens and the oil-immersed objective lens) Next, a method of measuring the upper surface and the lower surface of the flow cell using the water-immersed objective lens will be described.

A solution 209 is sandwiched between the upper glass 201 and the lower glass 204 of the flow cell to form a flow path.

The refractive index of solution 209 is 1.33, which is the same as that of pure water

Therefore, the optical path length from the tip of the water-immersed objective lens 202 shown in FIG. 2 (a) to the lower surface of the upper plate glass 201 is equal to the optical path length to the upper surface of the lower plate glass 209 shown in FIG. 2 (b).

This means that the distance from the objective lens to the measurement surface does not change even if a glass plate of the same thickness is moved up and down in a space filled with pure water.

[0015]

Therefore, the optical performances of FIGS. 2 (a) and 2 (b) are exactly the same.

This is achieved by having the same refractive index of the immersion medium and the solution in the flow path.

In other words, when an oil-immersed objective lens is used, the optical path lengths to the observation surfaces in FIGS. 2 (a) and 2 (b) are different.

In the case of an oil-immersed objective lens Since the objective lens is designed in the state shown in FIG. 2A, the image is blurred when the observation target is 10 μm away from the lower surface of the upper glass 206.

Therefore, when measuring the behavior of biological substances on the upper and lower surfaces in a flow path holding an aqueous solution-based liquid, it is more preferable to use a water immersion lens.

[0016]

As a first embodiment of the present invention, a device (analytical device) configuration and a measurement method (analysis method) of a DNA sequencer that analyzes a DNA base sequence using an immersion objective lens will be described.

[0017]

The flow tip 301 is fixed to the heat block 307 with an air chuck.

A Peltier element 305 is arranged on the lower surface of the heat block 307 to control the temperature of the flow cell 301.

The temperature control range is 4 to 95 ° C.

Temperature control is required for base elongation by enzymatic reaction in the flow chip 301, dissociation of the prima that serves as a scaffold for elongation, and the like.

A resistance temperature detector 331 is arranged inside the heat block 307 as a temperature sensor and is used for feedback of temperature control.

[0018]

The heat sink 306 is in close contact with the Peltier element 305 and dissipates heat generated by driving the Peltier element 305.

The heat dissipation from the heat sink 306 is achieved by circulating the cooled air in the refrigerator.

Enzymes, 4 types of fluorescent reagents, buffers, nucleotides, cleaning solutions, etc. are installed in the cooler 308 and cooled to 4 ° C.

The cool box 308 circulates cold air to the heat sink 306 in the apparatus, and an air passage is formed so that heat can be exhausted.

A fan 309 that promotes the circulation of cold air is placed in the air passage.

[0019]

Efficient temperature control can be performed by circulating in the air passage and carrying the heat generated by the heat sink 306 to the cooling chamber 308 again.

Further, since the temperature control of the cooling chamber 308 and the heat block 307 can be achieved by one cooling mechanism, the device can be miniaturized.

The heat block 307 and the flow tip 301 are held on the XY stage 324, and the flow tip 301 can be moved horizontally with respect to the water immersion objective lens 330.

The water immersion objective lens 330 is fixed to the Z stage 319 and can be moved up and down to focus on the object to be measured.

Pure water 332 is filled between the water immersion objective lens 330 and the flow tip 301 via the nozzle 314.

The nozzle 314 can control the supply amount of pure water 332 by detecting the change in capacitance.

The pure water replenished in the pure water container 310 can be sent by the syringe 313 via the tube 333.

A check valve 311 and an on-off valve 312 are installed in the flow path to control the supply of pure water.

[0020]

The DNA strand elongation reaction is carried out by reacting four kinds of nucleotides and polymerases labeled with different fluorescent dyes on the flow chip 301.

Each nucleotide is FAM-dCTP, Cy3-dATP, Texas Red -dGTP, Cy5-dTsTP, respectively.

The concentration of each nucleotide is 200 nM.

In addition, the salt concentration, magnesium concentration and pH of the reaction solution are optimized so that the stretching reaction can be efficiently carried out.

The reaction solution contains a polymerase, and only one fluorescent nucleotide complementary to the DNA fragment is incorporated.

2

The extension of the base does not occur because the fluorescent dye of the first base is bulky and causes steric hindrance to the uptake of the second base.

1After the base is taken in, the floating fluorescent nucleotides are removed by washing, and then the fluorescence is measured.

In addition, in order to carry out the reaction of the minimum unit thereafter, a step of dissociating the fluorescent dye from the base with a dissociation solution after measuring the fluorescence is indispensable.

This step can reset the steric hindrance of the fluorescent dye in the reaction of the smallest unit.

This allows the reaction to continue.

By sending the fluorescent nucleotides into the flow cell again and repeating the reaction, a sequential sequence of is possible.

The reaction method adopted in this example is called Sequence By Synthesis.

[0021]

2

The two LEDs 316 and 317 are light sources for exciting fluorescent dyes.

The median wavelengths of LEDs 316 and 317 are 490 and 595 nm, respectively.

LED316 uses FAM-dCTP and Cy3-dATP, and LED317 uses Texas Red -dGTP and Cy5-dTsTP for excitation light irradiation.

The dichroic mirror 315 aligns the light from LEDs 316 and 317 on the same optical axis.

Further, the excitation light is incident on the pupil surface of the objective lens 330 by the dichroic mirror 318.

The excitation light is applied to the fluorescent dye in the flow chip 301 through the objective lens 330 to emit fluorescence.

A part of the isotropically emitted fluorescence is recovered by the objective lens 330 via the flow chip 301 and the pure water 332.

The light that has passed through the objective lens 330 becomes parallel light, travels straight to the dichroic mirror 319, and is divided.

The dichroic mirror 319 has a gentle reflection characteristic in the fluorescence wavelength region of four colors.

Therefore, it is possible to calculate the fluorescence intensity ratio of the bright spots emitted from the reaction field on the flow chip on the light receiving surfaces of the CMOS cameras 322 and 324, respectively.

2By taking the ratio on the image plane of the two CMOS cameras, it is possible to determine which of the four colors the emission point belongs to.

The parallel light divided by the dichroic mirror 319 passes through the emission filters 320 and 325, respectively, and then is imaged on the light receiving surfaces of the CMOS cameras 322 and 324 by the tube lenses 321, 323.

[0022]

Next, a device configuration and method for automatically supplying the liquid to the immersion objective lens in FIG. 3 will be described with reference to FIG.

[0023]

The flow tip 401 is held in the heat block 419 by a vacuum chuck.

The cover glass surface of the flow chip 401 on the water-immersed objective lens 402 side is hydrophobic and has the property of repelling water droplets.

This is because when pure water 403 is dropped on the immersion objective lens 402, the pure water 403 does not spread on the glass surface and comes into contact with the immersion objective lens 402 while maintaining a certain contact angle.

In addition, when observing different fields of view in the flow chip 401 using the XY stage, it is a useful property to avoid problems such as the pure water 403 spreading on the glass surface due to vibration during movement.

[0024]

The heat block 419 has a flow path inside, and holds a DNA fragment containing a fluorescent dye in the flow path.

The water immersion objective lens 402 is held by the Z stage 410, and the focus on the flow chip 401 can be adjusted.

The nozzle 404 is held by the Z stage 410 and indirectly fixed to the water immersion objective lens 402, and the relative distance between the water immersion objective lens 402 and the nozzle 404 is kept constant.

[0025]

The outer circumference of the nozzle 404 is covered with an electric shield 405, and it has a coaxial double tube structure.

Since the nozzle 404 is covered with a conductive electric shield 405, the exposed area of the tip of the nozzle 404, which is a detection electrode, is minimized, which can improve the detection sensitivity of the capacitance.

Further, since the shield 405 is also a magnetic shield, it is possible to prevent a malfunction of the liquid level detection due to noise from the outside world caused by driving the Z stage 410.

[0026]

Pure water is held in the pure container 406 and is connected to the nozzle 404 via a tube 433.

By opening the open / close valve 409 and driving the syringe 408, the pure water in the pure container 406 can be pumped out and the pure water can be sent to the nozzle 404.

The check valve 407 prevents the backflow of pure water pressurized by the syringe 408.

Through these series of operations, the nozzle 404 can fill the space with pure water by supplying pure water to the water immersion objective lens 402 and the flow tip 401.

When pure water 403 is filled between the water-immersed objective lens 402 and the flow tip 401, the capacitance between the tip of the nozzle 404 and the flow tip 401 increases sharply.

As a result, it is possible to directly detect whether or not the tip of the water-immersed objective lens 402 is filled with pure water 403.

More specific detection of the change in capacitance is automated by the capacitance detection circuit 411, the liquid level determination circuit 412, the motor drive circuit 414, and the control unit 413.

The flow chip 401 is electrically grounded and serves as a ground when measuring the capacitance.

Further, although the description has been given using the water-immersed objective lens 402 in this embodiment, this is not limited to water immersion, and can be applied to an oil-immersed objective lens.

Since the capacitance depends on the dielectric constant of the substance that fills the space, in the case of this embodiment, any substance having a dielectric constant different from that of air can be detected in principle.

Therefore, the liquid filling the syringe 408 and the nozzle 404 may be oil, and in principle, any substance having a dielectric constant different from that of air can be detected.

[0027]

As a second embodiment of the analyzer of the present invention, a device configuration and method for automatically supplying a liquid to the immersion objective lens will be described.

In this embodiment, by arranging a plurality of energization sensors around the lens of the objective lens, the accuracy of liquid supply between the objective lens and the flow chip is improved.

In the immersion objective lens 502, four energization sensors 521 are arranged around the lens at the tip.

In order to prevent energization between the energization sensors before contacting the pure water 517, the tip portion 526 other than the lens of the water immersion objective lens 502 is in an insulated state.

[0028]

By measuring the capacitance between the flow tip 501 and the water immersion objective lens 502, the nozzle 504 can supply an appropriate amount of pure water 517.

However, it cannot be accurately determined whether or not the lens at the tip of the water-immersed objective lens 502 is completely immersed in pure water 517.

If only a part of the lens is in contact with pure water 517 and the other part is in contact with air instead of pure water 517, a good image cannot be obtained because the refractive index is abnormal. ..

Therefore, it is necessary that the entire lens is in contact with pure water 517.

In this embodiment, four energization sensors 522, 523, 524, 525 are arranged around the lens of the objective lens 502 as a means for achieving this.

As a result, when the four points around the lens of the immersion objective lens 502 are surely in contact with the pure water 517, any of the four energization sensors 522, 523, 524, 525 is energized.

4

When all the lenses are not energized, it can be detected that the objective lens 502 and the pure water 517 are not in complete contact.

However, this method does not guarantee the contact between the immersion objective lens 502 and the flow chip 501, and it is only detecting that the energization sensors 522, 523, 524, 525 are energized with pure water 517. ..

On the other hand, by detecting the capacitance with the nozzle 504, the presence or absence of contact between the water immersion objective lens 502 and the flow tip 501 can be detected, but the contact between the water immersion lens 502 and pure water 517 cannot be directly detected. ..

Therefore, by complementarily using the detection method based on the capacitance and the detection method based on the energization method, pure water 517 can be accurately supplied between the flow chip 501 and the water immersion objective lens 502.

[0029]

As a third embodiment of the analyzer of the present invention, a device configuration and method for automatically supplying a liquid to a water immersion objective lens will be described.

In this embodiment, in addition to the fifth embodiment, a method of modifying the tip shape of the nozzle for supplying the liquid will be described with reference to FIG.

[0030]

The tip of the nozzle 604 is annular.

When the nozzle 604 is placed in the device, it holds three holes that can discharge pure water 617 downward toward the flow tip 601 surface.

In this embodiment, the nozzle is annular.

When the tip of the nozzle is linear, moving the flow tip drags the pure water 617 at the tip of the water-immersed objective lens 602 to the flow tip, and in the worst case, air is mixed between the objective lens and the flow tip. It ends up.

On the other hand, when the tip of the nozzle is shaped like the annular portion 604, the annular portion acts as a bank to hold pure water even when the flow tip is driven, so that the pure water dissociates from the tip of the water immersion objective lens 602 and the air layer is formed. The problem of mixing can be avoided.

[0031]

As a fourth embodiment of the analyzer of the present invention, a device configuration and method for automatically supplying a liquid to a water immersion objective lens will be described.

In the fourth embodiment, the capacitance is detected by using the nozzle, whereas in this embodiment, the capacitance is detected by using the water immersion objective lens 702 itself.

The advantages of using the objective lens 702 as a sensor for measuring capacitance will be described below.

The target of the capacitance measurement is "presence or absence of pure water between the water immersion objective lens 702 and the flow chip 701".

However, in the measurement shown with reference to FIG. 4, "presence or absence of pure water between the nozzle 706 and the flow tip 701" is detected.

By converting the nozzle to an immersion objective lens 702 and directly measuring the capacitance, more accurate measurement can be performed.

In this case, since the water-immersed objective lens 702 measures the capacitance, it needs to be insulated from the ground of the main body of the device.

Otherwise, the potential of the water-immersed objective lens 702 will be the same as that of the state where the flow tip 701 is in contact with the ground.

Therefore, in the immersion objective lens 702, for example, it is important that the threaded portion of the body surface of the objective lens is made of an insulating material such as resin and is insulated from the ground of the main body of the apparatus.

Further, it is desirable that the water immersion objective lens 702 is shielded in order to prevent noise.

[0032]

As a fifth embodiment of the analyzer of the present invention, an apparatus configuration and method for automatically supplying a liquid to an immersion objective lens will be described.

In this embodiment, the tip of the objective lens holds the energization sensor, and the upper glass surface of the flow chip 801 is coated with a conductive transparent electrode film (ITO film) 803.

In this configuration, the nozzle 804 drops pure water 833 onto the transparent electrode film 803, which is the upper surface of the flow chip 801.

Whether or not the immersion objective lens 802 and the flow chip 801 are filled with pure water can be confirmed by detecting the energization state between the energization sensors 821, 823, 824, 825 and the transparent electrode 803.

Further, it can be similarly confirmed by the above-mentioned method whether or not the pure water 833 has immersed the tip portion of the immersion objective lens 802.

[0033]

(Regarding the supply of liquid) The configuration and method of automatically supplying the liquid to the immersion objective lens will be described in detail with reference to FIG.

[0034]

Here, the nozzle 905 can detect the capacitance.

In a), the flow tip 902 is fixed on the heat block 901 with a vacuum chuck.

In b), the immersion objective lens 904 is driven by the Z stage with respect to the flow chip 902 and moves downward to a predetermined position.

In c), the supply of pure water 906 is started between the nozzle 905 and the flow tip 902.

Here, since the flow tip 902 and the tip of the nozzle 905 are not in contact with each other, the capacitance does not exceed a predetermined threshold value, and it is determined that pure water 906 is not supplied.

In d), the supply of pure water 907 is continued, the pure water 907 comes into contact with the surface of the flow chip 902, and the capacitance between the flow chip 902 and the objective lens 904 increases sharply.

The capacitance exceeds the threshold value, and the nozzle 905 stops supplying pure water 907. e) The droplet 907 adsorbed on the objective lens 904 by surface tension becomes strong peculiar to water when it comes into contact with the flow chip 902. Due to surface tension, it falls to the surface of the flow chip 902 and spreads while filling the space between the objective lens 904 and the flow chip 902.

At this time, when the tip of the nozzle 905 cannot come into contact with the pure water 909, the nozzle 905 resumes the supply of the pure water 909.

The supply of pure water 909 is continued until the pure water 909 comes into contact with the nozzle 905.

When the pure water 909 comes into contact with the nozzle 905, the capacitance exceeds the threshold value, and the nozzle 905 stops supplying the pure water 909.

It is also possible to drive the flow cell in steps using an XY stage in order to stabilize the shape of the pure water 909.

f) If the optical measurement is continued for a long time, the pure water 910 gradually evaporates.

As the evaporation progresses, the pure water 910 and the nozzle 905 are separated from each other, and the nozzle 905 detects a sudden decrease in capacitance.

In response to this, the nozzle 905 supplies pure water again until the capacitance exceeds the threshold value.

As a result, for example, even in a long-term measurement over a week, by automatically supplying pure water 910 directly under the immersion objective lens 904, it becomes possible to perform automatic measurement using the immersion objective lens 904.

Further, the method described in this embodiment is not limited to the immersion objective lens 904, and can be similarly applied to the oil immersion objective lens.

The configuration and method of automatically supplying the liquid to the immersion objective lens will be described in more detail with reference to FIG.

[0035]

In this embodiment, pure water is supplied under the objective lens 1004 by using the annular nozzle 1005 used in Example 6 instead of the linear nozzle.

In addition, four energization sensors 1006 are arranged at the tip of the immersion objective lens 1004.

In addition, the tip of the immersion objective lens 1004 is insulated.

As in the ninth embodiment, the objective lens descends from a) to b) to the position where the objective lens is focused on the flow chip 1001.

At c), the annular nozzle 1005 begins to drip pure water 1010 in the direction of the flow tip 1001.

As described in Example 6, since the annular nozzle 1005 holds three ejection ports, the number of droplets to be dropped is three.

In d), the pure water 1007 comes into contact with the flow chip 1001 and gradually increases in shape.

In e), the pure water 1009 comes into contact with the energization sensor 1006 and the annular nozzle 1005 on the immersion objective lens 1004.

This confirms that pure water 1009 exists between the flow chip 1001 and the immersion objective lens 1004, and that the tip of the immersion objective lens 1004 is immersed in the pure water 1009.

As the evaporation of pure water progresses in f), the contact between the annular nozzle 1005 and the pure water 1010 is cut off, and it is detected that the amount of pure water 1010 is insufficient.

The annular nozzle 1005 resumes the supply of pure water and continues to supply until the annular nozzle 1005 comes into contact with the pure water 1009 again.

With this, it is possible to continue stable measurement for a long time despite the drying of pure water 1010.

[0036]

A device configuration and method for scanning the surface of the flow chip by the TDI (Time Delay Integration) method using a mechanism that automatically supplies the liquid to the immersion objective lens will be described in detail with reference to FIG.

[0037]

TDI (Time Delay Integration) is a special readout method for CCD, and is a method for imaging an object moving at a constant speed by combining the moving direction / speed with the charge transfer direction / speed of the CCD.

In addition to high speed, high sensitivity can be obtained by repeatedly exposing a moving object (for the number of vertical stages of the CCD).

Therefore, the TDI method adopts a line sensor instead of an area sensor.

[0038]

Further, in the conventional area sensor, it is necessary to repeat the step and repeat for the fluorescence measurement of the reaction field fixed to the flow chip flow path surface.

This method has a problem that fluorescence measurement cannot be performed because image blurring occurs for 90 to 200 milliseconds after the stage is driven until the vibration of the stage subsides.

This problem does not occur in the XY stage stop vibration problem in the TDI method without sequential XY stage stop.

Therefore, it is possible to omit the time for the vibration to be attenuated after the stage is stopped and to continue the measurement continuously.

The result is an increase in throughput.

[0039]

In this embodiment, an immersion objective lens 1105 having a numerical aperture of 0.95 and a magnification of 20 is used.

When the numerical aperture is 0.95, the resolution is 0.35 μm.

The number of fields of view is 26.5 mm.

In this case, the observation area on the sample surface is φ1.325 mm.

Further, there are four lanes of the flow path 611 in the flow chip 610.

The shape of the flow path is 80 × 3.5 mm.

In addition, the TDI camera used has 4096 x 24 pixels.

The pixel size is 6 × 6 μm.

Therefore, the size of the line sensor projected on the sample surface is 1.2 × 0.077 mm, which is a shape that can almost cover the observation area of the objective lens, φ1.325 mm.

The flow path 611 can cover the entire area with three scanning operations.

The pixel size on the sample surface is 0.3 × 0.3 μm.

Further, as the TDI camera, as described in detail in Example 3, two TDI cameras are used to discriminate between the four types of fluorescent dyes emitted from the flow path based on the fluorescence intensity ratio.

The TDI camera has a back-illuminated image sensor structure, and the pixel clock is 30 MHz.

Assuming that the TDI camera can be measured while the camera is always in operation and the average density of the reaction field is 1 piece / 25 pixels, the theoretically assumed throughput is 30 Mpixels / sec ÷ 25 pixels / base x 60 seconds x 60 minutes. × 24 hours / day = 100 Gbase / day.

[0040]

A more detailed measurement method will be described below.

The flow tip 1110 is vacuum chucked onto the heat block 1101.

Pure water 1106 is supplied between the immersion objective lens 1105 and the flow chip 1110 via a nozzle 1107 having a liquid level detection function based on capacitance.

The flow tip 1101 is continuously driven by the XY stage from end to end of the flow path 1111 without stopping.

When the amount of pure water 1106 on the flow cell 1110 decreases due to XY stage drive or evaporation of pure water 1106 due to the passage of time, the nozzle 1108 detects that the electrostatic capacity has decreased from the threshold value adjustment and promptly. Supply pure water 1106.

In addition, pure water 1106 remains on the surface of the flow chip 1110 as it moves, and those pure water recombine during the operation of TDI operation, the amount of pure water 1106 increases, and the capacitance reaches the upper threshold limit. If it exceeds, the pure water 1106 is recovered.

The pure water supply nozzle 1107, which retains the liquid level detection function, enables smooth fluorescence image acquisition by TDI line scanning using the immersion objective lens 1105.

[0041]

The annular nozzle 1107 also has the effect of retaining pure water 1106.

Even when surface tension is generated in which the pure water 1106 drags in the direction opposite to the moving direction of the flow chip 1110, the annular nozzle 1107 exerts a force to hold the pure water 1106 from the outer peripheral direction, so that it is continuous like TDI. The TDI measurement method using the immersion objective lens described in this embodiment works without any problem even in a favorable operating environment.

[0042]

A method of fixing DNA molecules at high density in a lattice pattern on a substrate and sequencing using an immersion objective lens will be described below with reference to FIG. 12.

[0043]

The substrate 1201 has the shape of a slide glass of 25 x 75 mm cut out from a silicon wafer.

On the surface of the substrate 1201, aminosilane-modified DNA binding surfaces 1202 having a diameter of 300 nm are regularly formed in a grid pattern at a pitch of 390 nm.

Aminosilanes are known to efficiently adsorb to DNA molecules and are actively used for surface modification of DNA chips and the like.

In addition, the DNA non-binding surface 1203 is modified with HMDS (= hexamethyldisilzane), which has the property of not adsorbing to DNA.

Therefore, when the DNA molecule is injected onto the substrate, the DNA molecule actively binds only to the DNA binding surface 1204 and not to the DNA non-binding surface 1203.

Since the current semiconductor technology can expose a line width of 45 nm, the production of the substrate 1201 that holds the above-mentioned DNA binding surface 1202 in a grid pattern at a pitch of 390 nm can be easily achieved.

[0044]

In addition, as detailed in Science, vo327, pp78, 2010, it is possible to synthesize DNA nanoballs 1202 with a diameter of 300 nm in vitro at a concentration of 2 × 1010 cells / mL.

Since each of these DNA nanoballs 1202 has one unique DNA fragment amplified, it is also known that base sequence analysis is possible by binding a prima to the DNA nanoballs 1202 and sequentially advancing base elongation. Has been done.

Moreover, while the DNA nanoball 1202 holds a diameter of 300 nm, the diameter of the DNA binding surface 1204 that can be bound on the substrate 1201 is also 300 nm.

Therefore, two DNA nanoballs 1202 cannot be bound to one DNA binding surface 1204 due to the volume exclusion effect.

That is, only one DNA nanoball 1202 is bound to one DNA binding surface 1204.

Therefore, it is possible to perform optical measurement in the form of DNA nanoballs 1202 packed tightly on a substrate at a pitch of 350 nm.

[0045]

In this embodiment, a 60 × immersion objective lens having a numerical aperture of 1.2 is used.

The resolution of an objective lens with a numerical aperture of 1.2 can be calculated at δ = 0.61 × 550 nm / numerical aperture, in which case the resolution is shown to be 280 nm.

Therefore, it is possible to optically identify the DNA nanoballs 1202 placed on the substrate at a pitch of 390 nm.

Since the numerical aperture of the dry objective lens is about 0.9 at the maximum, the resolution in this case is 370 nm, so that the DNA nanoballs 1202 arranged at a pitch of 390 nm cannot be optically identified.

Therefore, it is essential to use an immersion objective lens in order to perform optical detection when DNA nanoballs with a diameter of 300 nm are arranged in a grid pattern at a pitch of 390 nm.

In addition, when an immersion objective lens with a numerical aperture of 1.2 is used compared to a dry objective lens with a numerical aperture of 0.75, the light collection efficiency due to the solid angle is increased from 0.09 to 0.28, so that the signal strength can be increased up to 3.2 times. can.

This contributes to improving the accuracy of base sequence analysis.

[0046]

A CMOS camera with a pixel size of 6.5 μm and a pixel count of 2048 x 2048 is used as a sensor to detect optical signals.

Detect with a sensor.

In general, it is desirable to allocate two or more pixels in the side direction to the size of the reaction field in order to accurately discriminate between different reaction fields.

In this embodiment, the pixel size projected on the sample surface is 0.11 μm / pixel, and 3 pixels can be allocated for the reaction field spacing of 350 nm, so that good optical detection can be performed.

Since the field of view of the immersion objective lens is 26.5 mm, the optically measurable area on the sample surface is 26.5 / 60 = φ440 μm.

Since the size of the CCD surface projected on the sample surface is 222 μm square, it is possible to use all the CCD surfaces for detection.

Therefore, at a pitch of 0.35 μm, it is possible to measure a reaction field of 403 × 103 per CCD field of view.

Assuming that 80 bases of base can be extended per reaction field, 32 Mbases of base sequence can be analyzed per field of view.

It is assumed that the region to be irradiated with the excitation light per FOV is 440 μm / √2 and the nanoballs can be fixed to 312 μm and 80% of the area of the slide glass surface.

In this case, the number of read bases of about 32Mbase / FOV × 12000FOV / slide ≒ 400Gbase / slide can be obtained.

Furthermore, if nanoballs can be fixed on the upper and lower surfaces of the flow path, the number of read bases will be 800 Gbase / slide.

[0047]

101, 111, 201, 207... Upper plate glass 102, 112, 204, 209... Lower plate glass 103, 113... DNA molecule 104... Dry objective lens 109. . Optical axis 106, 116, 204, 208.. Solution 117, 203, 206, 332, 403, 517, 617, 906 , 909, 910, 1007, 1009, 1010, 1106... Pure water 115, 202, 209, 330, 402, 502, 702, 802, 904, 1004, 1105... Immersion objective lens 301, 401, 501, 601, 701 , 801, 902, 1001, 1110 .. Flow chip 307, 419, 901, 1101... Heat block 305... Peiche element 306... Heat sink 308... Cooler 324... XY stage 319, 410... Z stage 314, 404, 504, 706, 804, 905,... Nozzle 310, 406... Pure water container 333... Tube 313, 408... Syringe 311, 407... Check valve 312, 409... Open / close valve 316, 317... LED 318, 319... Dicroic mirror 322, 324... CMOS Camera 320, 325... Emission filter 321, 323... Tube lens 405... Electric shield 411... Capacitive detection circuit 412.. Liquid level determination circuit 413... Control unit 414... Motor drive circuit 521, 522, 523, 524, 525, 821, 823 , 824, 825, 1006... Energization sensor 803... Transparent electrode film 907.. Droplet 604,1005, 1107.. Circular nozzle 1111.. Channel 1202... DNA nanoball 1204... DNA binding surface 1203... DNA non-binding surface 1201... Substrate

Europäisches Patentamt
European Patent Office
Office européen des brevets

Patent Translate
Powered by EPO and Google

Notice

This translation is machine-generated. It cannot be guaranteed that it is intelligible, accurate, complete, reliable or fit for specific purposes. Critical decisions, such as commercially relevant or financial decisions, should not be based on machine-translation output.

CLAIMS JP2015090458A

1.

An objective lens for observing a sample, a mechanism for driving the objective lens in the optical axis direction, a light source for illuminating the sample, an optical sensor for detecting light emission from the sample via the objective lens, and the above. It has a nozzle for supplying a liquid between the sample and the objective lens, a means for adjusting the temperature of the sample, and a means for supplying a plurality of reagents to the sample, and electrically detects the state of the liquid. An analyzer characterized by doing.

2.

The analyzer according to claim 1, wherein the electric detection of the state of the liquid is performed by measuring the capacitance.

3.

The analyzer according to claim 2, wherein it is the nozzle that measures the capacitance.

4.

The analyzer according to claim 3 further has a mechanism for fixing a flow cell having a flow path inside and a mechanism for driving the flow cell in the horizontal direction, and the nozzle is a nozzle and the flow cell. An analyzer characterized by measuring the capacitance between.

5.

The analyzer according to claim 3, wherein a conductive electric shield covering the nozzle is provided.

6.

The analyzer according to claim 2, wherein a plurality of energization sensors are provided around the objective lens, and the state of the liquid is electrically detected by using the energization sensors.

7.

The analyzer according to claim 6, wherein the energization sensors are arranged side by side on the circumference.

8.

The analyzer according to claim 6, wherein the energization sensors are arranged side by side at equal intervals.

9.

The analyzer according to claim 1, wherein the tip of the nozzle is annular and has a plurality of liquid discharge ports.

10.

In the analyzer according to claim 1, the sample is held by a light-transmitting flow chip, the flow chip includes a reagent inlet, an outlet, and a flow path, and the surface of the flow chip is hydrophobic. An analyzer characterized by having the property of repelling pure water.

11.

The analyzer according to claim 10, wherein a plurality of DNA molecules having different base sequences are discretely fixed on the upper and lower surfaces of the flow path of the flow chip.

12.

The analyzer according to claim 10, wherein the surface of the flow chip is covered with a transparent electrode film.

(19)日本国特許庁(JP)　(12)公表特許公報(A)　(11)特許出願公表番号
特表2007-526772
(P2007-526772A)
(43)公表日　平成19年9月20日(2007.9.20)

(51)Int.Cl.		F I		テーマコード（参考）
C12N 15/09	(2006.01)	C12N 15/00	F	4B024
G01N 33/53	(2006.01)	G01N 33/53	ZNAM	4B029
C12M 1/00	(2006.01)	C12M 1/00	A	
G01N 37/00	(2006.01)	G01N 37/00	102	

審査請求 未請求　予備審査請求 未請求　（全 41 頁）

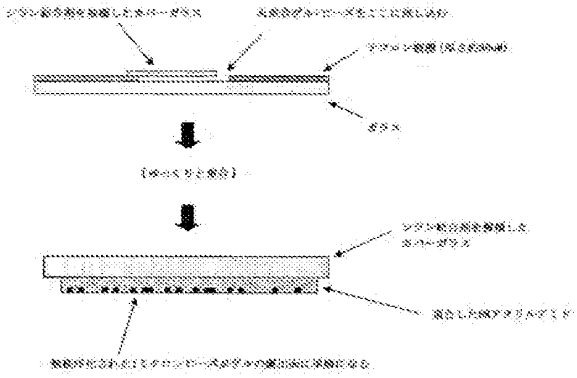
(21)出願番号　特願2007-500806 (P2007-500806)
(86)(22)出願日　平成17年2月28日 (2005.2.28)
(85)翻訳文提出日　平成18年9月29日 (2006.9.29)
(86)国際出願番号　PCT/US2005/006425
(87)国際公開番号　WO2005/082098
(87)国際公開日　平成17年9月9日 (2005.9.9)
(31)優先権主張番号　60/548,631
(32)優先日　平成16年2月27日 (2004.2.27)
(33)優先権主張国　米国 (US)

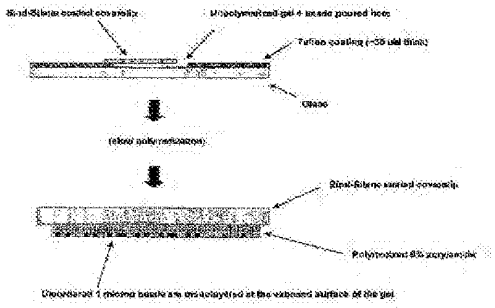
(71)出願人　592257310
プレジデント・アンド・フェロウズ・オブ・ハーバード・カレッジ
アメリカ合衆国０２１３８マサチューセッツ州ケンブリッジ、クウィンシー・ストリート１７
(74)代理人　100102978
弁理士　清水　初志
(74)代理人　100128048
弁理士　新見　浩一
(72)発明者　チャーチ　ジョージ　エム、
アメリカ合衆国　マサチューセッツ州　ブルックリン　ケント　ストリート　２１８

最終頁に続く

(54)【発明の名称】インサイチュー配列決定用ポロニー蛍光ビーズ

(57)【要約】
小型化された、高密度のビーズ系アレイを提供する。クローンビーズを製造および使用する方法、ならびに小型化された、高密度のビーズ系アレイを製造および使用する方法も提供する。





【特許請求の範囲】
【請求項１】
多数のビーズを含むアレイであって、ここで個々のビーズはビーズに付着した実質的に同一の核酸配列の集団を有し、集団は他の個々のビーズに付着した実質的に同一の核酸配列の集団と配列が異なり、かつ多数のビーズは半流動培地中に固定化されてアレイを形成する、アレイ。
【請求項２】
半流動培地が固体支持体に付着している、請求項1記載のアレイ。
【請求項３】
支持体がスライドグラスまたはフローセルである、請求項2記載のアレイ。
【請求項４】
個々のビーズがビーズに付着した実質的に同一の核酸配列の2つ、3つまたは4つの異なる集団を含む、請求項1記載のアレイ。
【請求項５】
個々のビーズの少なくとも40％がビーズに付着した実質的に同一の核酸配列の集団を含む、請求項1記載のアレイ。
【請求項６】
ビーズが単層として固定化されている、請求項1記載のアレイ。
【請求項７】
半流動培地がx、y、およびz軸を有し、多数のビーズがxおよびy軸に対してランダムに配置されている、請求項1記載のアレイ。
【請求項８】
半流動培地が上面および底面を有し、多数のビーズが上面近くに固定化されている、請求項1記載のアレイ。
【請求項９】
半流動培地が上面および底面を有し、固体支持体が底面に付着している、請求項2記載のアレイ。
【請求項１０】
半流動培地が、ポリアクリルアミド、セルロース、ポリアミド、架橋アガロース、架橋デキストラン、および架橋ポリエチレングリコールからなる群より選択される、請求項1記載のアレイ。
【請求項１１】
多数のビーズが多数のクローンビーズ（clonal beads）を含む、請求項1記載のアレイ。
【請求項１２】
多数のビーズがライブラリーを含む、請求項1記載のアレイ。
【請求項１３】
以下の段階を含む、アレイを製造する方法：
a）多数のビーズを提供する段階であって、ここで個々のビーズがビーズに付着した実質的に同一の核酸配列の集団を有し、集団が他の個々のビーズに付着した実質的に同一の核酸配列の集団と配列が異なる段階；および
b）アレイを形成するためにビーズを半流動培地中に固定化する段階。
【請求項１４】
段階b）において固体支持体に半流動培地を付着させる段階をさらに含む、請求項13記載の方法。
【請求項１５】
支持体が、スライドグラスまたはフローセルである、請求項14記載の方法。
【請求項１６】
多数のビーズがビーズに付着した実質的に同一の核酸配列の2つ、3つまたは4つの異なる集団を含む、請求項13記載の方法。

【請求項１７】
個々のビーズの少なくとも40％がビーズに付着した実質的に同一の核酸配列の集団を含む、請求項13記載の方法。
【請求項１８】
ビーズが単層として固定化されている、請求項13記載の方法。
【請求項１９】
半流動培地がx、y、およびz軸を有し、多数のビーズがxおよびy軸に対してランダムに配置されている、請求項13記載の方法。
【請求項２０】
半流動培地が上面および底面を有し、多数のビーズが上面近くに固定化されている、請求項13記載の方法。
【請求項２１】
半流動培地が上面および底面を有し、固体支持体が底面に付着している、請求項14記載の方法。
【請求項２２】
半流動培地が、ポリアクリルアミド、セルロース、ポリアミド、架橋アガロース、架橋デキストラン、および架橋ポリエチレングリコールからなる群より選択される、請求項13記載の方法。
【請求項２３】
多数のビーズが多数のクローンビーズを含む、請求項13記載の方法。
【請求項２４】
多数のビーズがライブラリーを含む、請求項13記載の方法。
【請求項２５】
以下の段階を含む、アレイを製造する方法：
a）多数のビーズを提供する段階であって、ここで個々のビーズがビーズに付着した実質的に同一の核酸配列の集団を有し、集団が他の個々のビーズに付着した実質的に同一の核酸配列の集団と配列が異なる段階；
b）アレイを形成するためにビーズを半流動培地中に固定化する段階；および
c）ビーズに付着した実質的に同一の核酸配列の増幅された集団を有する多数のビーズを形成するために、実質的に同一の核酸配列の集団を増幅する段階。
【請求項２６】
半流動培地が増幅プライマーを含む、請求項25記載の方法。
【請求項２７】
半流動培地が半流動培地中に間隙を形成する添加剤を含む、請求項25記載の方法。
【請求項２８】
添加剤が、陽イオン性脂質、ポリアミン、およびポリカチオンからなる群より選択される、請求項27記載の方法。
【請求項２９】
以下の段階を含む、アレイを製造する方法：
a）多数のビーズを提供する段階であって、ここで個々のビーズがビーズに付着した実質的に同一の核酸配列の集団を有し、集団が他の個々のビーズに付着した実質的に同一の核酸配列の集団と配列が異なる段階；
b）ビーズに付着した実質的に同一の核酸配列の増幅された集団を有する多数の固定化ビーズを形成するために、実質的に同一の核酸配列の集団を増幅する段階；および
c）アレイを形成するためにビーズを半流動培地中に固定化する段階。
【請求項３０】
増幅がエマルジョンPCRにより行われる、請求項29記載の方法。
【請求項３１】
以下の段階を含む、アレイを製造する方法：
a）多数のビーズを提供する段階であって、ここで個々のビーズがビーズに付着した実質

的に同一の核酸配列の集団を有し、集団が他の個々のビーズに付着した実質的に同一の核酸配列の集団と配列が異なる段階；
b）ビーズに付着した実質的に同一の核酸配列の増幅された集団を有する多数のビーズを形成するために、実質的に同一の核酸配列の集団を増幅する段階；
c）ビーズの濃縮された集団を形成するために、ビーズに付着した実質的に同一の核酸配列の増幅された集団を有する多数のビーズを濃縮する段階；および
d）アレイを形成するためにビーズを半流動培地中に固定化する段階。

【請求項３２】
以下の段階を含む、ビーズに付着した第一の核酸配列を有するビーズの集団を濃縮するための方法：
a）ビーズの集団を提供する段階であって、個々のビーズがビーズに付着した実質的に同一の核酸配列の集団を有し、集団が他の個々のビーズに付着した実質的に同一の核酸配列の集団と配列が異なる段階；
b）ビーズの集団を第一の核酸配列に相補的な第二の核酸配列と接触させる段階；
c）ハイブリダイゼーションが起こりハイブリダイズしたビーズの集団およびハイブリダイズしないビーズの集団を形成するようにビーズの集団および第二の核酸配列を一緒にインキュベートする段階；ならびに
d）ハイブリダイズしたビーズの集団をハイブリダイズしないビーズの集団から分離する段階。

【請求項３３】
第二の核酸配列が捕捉ビーズ上に固定化されている、請求項32記載の方法。

【請求項３４】
ハイブリダイズしたビーズの集団がハイブリダイズしないビーズの集団から密度または親和性により分離される、請求項32記載の方法。

【請求項３５】
半流動培地中に固定化された多数のビーズを含むアレイを含むキットであって、多数のビーズがビーズに付着した実質的に同一のアミノ酸の集団を有する個々のビーズを含み、集団が他の個々のビーズに付着した実質的に同一の核酸配列の集団と配列が異なる、キット。

【請求項３６】
実質的に同一の核酸配列がプライマーである、請求項1記載の方法。

【請求項３７】
実質的に同一の核酸配列が増幅された核酸配列である、請求項1記載の方法。

【請求項３８】
実質的に同一の核酸配列がプライマーである、請求項13記載の方法。

【請求項３９】
実質的に同一の核酸配列が増幅された核酸配列である、請求項13記載の方法。

【請求項４０】
実質的に同一の核酸配列がプライマーである、請求項25記載の方法。

【請求項４１】
実質的に同一の核酸配列が増幅された核酸配列である、請求項25記載の方法。

【請求項４２】
実質的に同一の核酸配列がプライマーである、請求項29記載の方法。

【請求項４３】
実質的に同一の核酸配列が増幅された核酸配列である、請求項29記載の方法。

【請求項４４】
実質的に同一の核酸配列がプライマーである、請求項31記載の方法。

【請求項４５】
実質的に同一の核酸配列が増幅された核酸配列である、請求項31記載の方法。

【請求項４６】

実質的に同一の核酸配列がプライマーである、請求項32記載の方法。
【請求項４７】
実質的に同一の核酸配列が増幅された核酸配列である、請求項32記載の方法。
【発明の詳細な説明】
【技術分野】
【０００１】
発明の分野
本発明は、小型化された高密度のビーズ系アレイ、小型化された高密度のビーズ系アレイを製造および使用する方法、ならびに該アレイの構成要素を製造および使用する方法に関する。なお、本出願は、2004年2月27日出願の米国特許仮出願第60/548,631号の優先権の恩典を主張し、事実上その全体が参照として本明細書に組み入れられる。
【０００２】
政府の権益の記載
本発明は、DARPAが与える助成金番号F30602-01-2-0586およびエネルギー省が与える助成金番号DE-FG02-02ER63445の政府の援助を受けて成されたものである。政府は本発明に一定の権利を有する。
【背景技術】
【０００３】
発明の背景
ポリメラーゼコロニー（ポロニー（polony））法は、単分子増幅法であり、その方法により、それぞれ個々の分子の配列を高度に並列的に解明することが可能となる。しかし、ポロニー法のスループットは、数十から数千ミクロンの範囲である個々のコロニーの大きさに反比例する。
【０００４】
クローンミクロスフェア（すなわちクローン増幅されたDNAを載せるビーズ）の集団を発生させるための方法は、当技術分野で公知である（例えば、Dressman (2003) Proc. Natal. Acad. Sci. USA 100:8817（非特許文献1）；Brenner et al. (2000) Nat. Biotech. 18:630（非特許文献2））。しかし、これらの方法はいくつかの欠点に悩んでいる。Dressmanらの著作では、稼働に費用がかかり、数億個のビーズをを処理するには低すぎるスループット（すなわち1秒あたり70,000イベント未満）に悩む蛍光標示式細胞分取（FACS）によりビーズが分析される。Brennerらの著作では、物理的充填がビーズの散乱を制限するようにビーズが操作され、充填された平面アレイが形成される。
【０００５】
【非特許文献１】Dressman (2003) Proc. Natal. Acad. Sci. USA 100:8817
【非特許文献２】Brenner et al. (2000) Nat. Biotech. 18:630
【発明の開示】
【０００６】
概要
本発明は、効率的で、費用効率の高いビーズ系アレイの製造のための新規な方法を見出したことに一部基づく。そのようなアレイは、遺伝学の研究および診断への適用に特に有用である。本明細書に記載した方法および組成物は、数千万から数十億の別個の核酸配列を妥当な時間で経済的なやり方で問い合わせできるようにする。
【０００７】
本発明の態様は、ビーズ系アレイおよびビーズ系アレイを作製する方法を目的とする。ある態様によると、多数のビーズを有するアレイが提供され、ここで、個々のビーズは、それらに付着した実質的に同一の核酸配列の集団を有し、実質的に同一の核酸配列の集団は、他のビーズに付着した実質的に同一の核酸配列の集団と配列が異なる。多数のビーズを半流動培地中に固定化してアレイを形成する。ポリアクリルアミド、セルロース、ポリアミド、架橋アガロース、架橋デキストラン、または架橋ポリエチレングリコールから半流動培地を作製できる。ある局面では、半流動培地はx、y、およびz軸を有し、多数のビ

ーズはそのxおよびy軸に対してランダムに配置される。それらのビーズを、例えば半流動培地の上面近くに単層として固定化できる。

【０００８】

ある局面では、半流動培地を、顕微鏡用スライドまたはフローセルのような固体支持体に付着できる。固体支持体を半流動培地の底面に付着できる。

【０００９】

他の局面では、実質的に同一の核酸配列の2つ、3つ、または4つの異なる集団を、それらのビーズに付着できる。他の局面では、それらのビーズはクローンビーズである。なお他の局面では、それらのビーズはライブラリーを含む。

【００１０】

他の態様では、

実質的に同一の核酸配列の集団が付着した多数のビーズを提供する段階；

それらのビーズを半流動培地中に固定化してアレイを形成する段階；

その実質的に同一の核酸配列の集団を増幅して、実質的に同一の核酸配列の増幅された集団が付着した多数のビーズを形成する段階；

を含むビーズ系アレイを作製する方法を提供する。ある局面では、半流動培地は増幅プライマーを含む。他の局面では、半流動培地は、陽イオン性脂質、ポリアミン、またはポリカチオンのようなその半流動培地中で間隙を形成する添加剤を含む。

【００１１】

他の態様では、

実質的に同一の核酸配列の集団が付着した多数のビーズを提供する段階；

その実質的に同一の核酸配列の集団を増幅して、実質的に同一の核酸配列の増幅された集団が付着した多数の固定化ビーズを形成する段階；

を含むビーズ系アレイを作製する方法を提供する。次に、それらのビーズを半流動培地中に固定化してアレイを形成する。ある局面では、エマルジョンPCRにより増幅を行う。

【００１２】

なお他の態様では、

実質的に同一の核酸配列の集団が付着した多数のビーズを提供する段階；

その実質的に同一の核酸配列の集団を増幅して、実質的に同一の核酸配列の増幅された集団が付着した多数のビーズを形成する段階；

を含むビーズ系アレイを作製する方法を提供する。実質的に同一の核酸配列の増幅された集団が付着した多数のビーズを濃縮して、ビーズの濃縮された集団を形成し、それらのビーズを半流動培地中に固定化してアレイを形成する。

【００１３】

本発明の態様は、第一の核酸配列が付着したビーズの集団を濃縮するための方法を目的とする。これらの方法は、ビーズの集団を提供する段階を含み、ここで、その集団の少なくとも一部は第一の核酸配列が付着したビーズを含む。そのビーズの集団を第一の核酸配列に相補的な第二の核酸配列と接触させ、そのビーズの集団および第二の核酸配列を一緒にインキュベートし、ハイブリダイゼーションが起こり、ハイブリダイズしたビーズの集団とハイブリダイズしないビーズの集団とを形成するようにする。次に、ハイブリダイズしたビーズの集団をハイブリダイズしないビーズの集団と分離する。ある局面では、第二の核酸を捕捉ビーズ上に固定化する。他の局面では、ハイブリダイズしたビーズの集団を、ハイブリダイズしないビーズの集団と密度または親和性により分離する。

【００１４】

別の態様によると、多数のビーズが半流動培地中に固定化したアレイを含むキットが提供され、そのキットでは、実質的に同一の核酸配列の集団がそれらのビーズに付着している。

【００１５】

詳細な説明

本発明は、増幅されたDNAを載せる多くのミクロスフェアを調製する方法およびそのア

レイを形成する方法を提供する。次に、単一で少量の試薬を使用して、アレイのビーズの増幅されたDNAに様々なハイブリダイゼーションおよび酵素法を同時に適用できる。ある局面では、核酸の並列分析、例えばDNA配列決定およびRNA発現プロファイリングを本発明のビーズと組み合わせることができる。本明細書に記載するビーズならびにビーズ系高密度アレイを作製および使用する方法は、様々な遺伝学系研究および診断への適用に有用であり、それらについて以下でさらに論じる。

【００１６】

本発明は、当技術分野で公知であるビーズ系方法に勝る利点を提供する。例えば、本明細書に記載する方法は、1ミクロンのスケールまで読み取りを減少させることにより、現行の配列決定法の分解能を大きく増加させる。本発明のある方法の別の利点は、ポリマーまたはゲルにビーズを埋め込むことにより、現行の「ポリメラーゼトラッピング」法が増強されることである。本明細書に記載するある方法のなお別の利点は、ビーズが0.4ミクロンよりも良い精度で固定化されたままであることから、埋め込みが画像レジストレーションを援助できることである。本発明は、増幅されていないビーズと比較して増幅されたビーズを濃縮するための簡便な方法も好都合に提供する。本発明の他の局面では、単層を含む固定化ビーズはz軸に対して秩序化されいるが、x軸およびy軸に対して完全に無秩序化されている。これは、ビーズのパターンに秩序を発生させる支持体（例えばビーズが納まるためのエッチングされたウェルを有する表面）を使用する必要なしにビーズのアレイを発生できるという利点を提供する。ビーズを埋め込み、またはそうでなく付着させるための、本明細書に記載される方法のさらなる利点は、その方法により取得系への組み込みが可能になることであり、その取得系では、ビーズが検出手段に対して移動して、その検出手段を用いて概して使用されうるであろう要素よりも大きい要素からデータを収集することが可能になる。

【００１７】

本発明は、ビーズおよびビーズ系アレイを提供する。本発明に使用する用語「ビーズ」は、球形（例えばミクロスフェア）であるか、または不規則な形状を有しうる別個の粒子を指す。ビーズは直径約0.1μｍと小さいか、または直径約数ミリメータと大きい場合がある。ビーズは、サイズが概して直径約0.1μｍから200μｍ、直径約0.25μｍから100μｍ、直径約0.5μｍから50μｍ、直径約0.6μｍ~40μｍ、直径約0.7μｍから30μｍ、直径約0.8μｍから20μｍ、直径約0.9μｍから10μｍ、または直径約1μｍから9μｍの範囲でありうる。ある局面では、本発明のビーズは、直径約1μｍ、2μｍ、3μｍ、4μｍ、5μｍ、6μｍ、7μｍ、8μｍまたは9μｍである。ビーズは、常磁性材料、セラミック、プラスチック、ガラス、ポリスチレン、メチルスチレン、アクリルポリマー、チタン、ラテックス、セファロース、セルロース、ナイロンなどを含む様々な材料を含みうるが、それに限定されるわけではない。

【００１８】

ある例によると、ビーズは表面に官能基を有する場合があり、その官能基を使用してビーズに核酸配列を結合できる。ハイブリダイゼーション（例えばポリマーに対する結合）、共有付着、磁気付着、親和性付着などにより核酸配列をビーズに付着できる。例えばビーズをストレプトアビジンで被覆でき、核酸配列はビオチン部分を含みうる。ビオチンはビーズ上のストレプトアビジンと結合でき、よってビーズに核酸配列を付着できる。ストレプトアビジン、オリゴdT、およびヒスチジンタグ結合基質で被覆したビーズが商業的に入手可能である（Dynal Biotech, Brown Deer, WI）。例えば、カルボキシル、アミノ、およびヒドロキシル基のような核酸アレイを発生させるための化学法、または官能化シリコン化合物のような当技術分野で公知の固相化学法を使用してビーズを官能化することもできる（例えば、米国特許第5,919,523号参照、全体がすべての趣旨で参照として本明細書に組み入れられる）。

【００１９】

支持体にオリゴヌクレオチドを固定化する方法が記載され、当技術分野で公知であり（ビーズ: Dressman et al. (2003) Proc. Natl. Acad. Sci. USA 100:8817, Brenner et al. (2000) Nat. Biotech. 18:630, Albretsen et al. (1990) Anal. Biochem. 189:40お

よびLang et al. Nucleic Acids Res. (1988) 16:10861; ニトロセルロース: Ranki et al. (1983) Gene 21:77; セルロース： (Goldkorn (1986) Nucleic Acids Res. 14:9171; ポリスチレン： Ruth et al. (1987) Conference of Therapeutic and Diagnostic Applications of Synthetic Nucleic Acids, Cambridge U.K.; テフロン-アクリルアミド: Duncan et al. (1988) Anal. Biochem. 169:104; ポリプロピレン: Polsky-Cynkin et al. (1985) Clin. Chem. 31:1438; ナイロン: Van Ness et al. (1991) Nucleic Acids Res. 19:3345; アガロース:Polsky-Cynlcin et al., Clin. Chem. (1985) 31:1438; ならびにセファクリル: Langdale et al. (1985) Gene 36:201; ラテックス: Wolf et al. (1987) Nucleic Acids Res. 15:2911; 全体がすべての趣旨で参照として本明細書に組み入れられる）、本明細書にさらに記載される。

【００２０】

本明細書に使用する用語「付着する」は、共有相互作用および非共有相互作用の両方を指す。共有相互作用は、1対の電子（すなわち単結合）、2対の電子（すなわち二重結合）または3対の電子（すなわち三重結合）の共用により形成される2個の原子またはラジカルの間の化学結合である。共有相互作用は、電子対相互作用または電子対結合としても当技術分野で公知である。非共有相互作用には、ファンデルワールス相互作用、水素結合、弱化学結合（すなわち短距離非共有結合力を介する）、疎水性相互作用、イオン結合などがあるが、それに限定されるわけではない。Alberts et al., in Molecular Biology of the Cell, 3d edition, Garland Publishing, 1994に、非共有相互作用の総説を見出すことができ、全体がすべての趣旨で参照として本明細書に組み入れられる。

【００２１】

本発明の態様は、核酸配列（例えばオリゴヌクレオチド配列またはポリヌクレオチド配列）の1個から数百万個のコピーが付着したビーズを提供する。一局面では、ビーズは、それに付着した単一核酸配列の多数のコピーを有しうる（すなわちクローンビーズ）。別の局面では、そのビーズは、それに付着した2、3、4、5、10個またはそれよりも大きい化学種の核酸配列を有しうる。例えば、一局面では、両方向の遺伝子配列（すなわち正のストランドおよび負のストランド）がビーズに付着している場合がある。

【００２２】

ある態様では、実質的に同一の核酸配列の集団が付着したビーズを提供する。本明細書に使用する用語「実質的に同一の核酸配列」は、相互に対して少なくとも、50%、55%、60%、65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%または99.9%の配列同一性を有する核酸配列を含むことを意図するが、それに限定されるわけではない。ある局面では、実質的に同一は、100%の配列同一性を含む。用語、実質的に同一は、ビーズに付着したすべての核酸配列、ビーズに付着したプライマー、および/またはビーズに付着した増幅産物にあてはまりうる。

【００２３】

本明細書に使用する用語「オリゴヌクレオチド」は、概して合成手段により調製された一本鎖DNAまたはRNA分子を含むことを意図するが、それに限定されるわけではない。本発明のヌクレオチドは、アデノシン、グアノシン、ウリジン、シチジンおよびチミジン由来のヌクレオチドのような概して天然ヌクレオチドであろう。オリゴヌクレオチドを「二本鎖」と称する場合、当業者は、水素結合性ヘリカルアレイ状に、例えばDNAと概して会合して一対のオリゴヌクレオチドが存在することを了解している。100%相補形態の二本鎖オリゴヌクレオチドに追加して、本明細書に使用するような用語「二本鎖」は、湾曲およびループのような構造特性を含む形態も含むことを意味する（Stryer, Biochemistry, Third Ed. (1988) 参照、全体がすべての趣旨で参照として本明細書に組み入れられる）。本明細書に使用する用語「ポリヌクレオチド」は、様々な異なるサイズでありうる核酸鎖を指す。ポリヌクレオチドはオリゴヌクレオチドと同じサイズでありうるし、またオリゴヌクレオチドのサイズの2倍、3倍、4倍、5倍、10倍、もしくはそれよりも大きい場合がある。オリゴヌクレオチドおよびポリヌクレオチドには、本明細書に記載された任意の方法

を使用してビーズに付着されたものと、増幅により作製されたもの（すなわち「増幅産物」）とが含まれる。

【００２４】

オリゴヌクレオチドおよび/またはポリヌクレオチドを天然供給源から単離でき、あるいは商業的な供給源から購入できる。オリゴヌクレオチドおよび/またはポリヌクレオチド配列を、任意の適切な方法、例えばBeaucageおよびCarruthers（(1981) Tetrahedron Lett. 22: 1859）に記載されたホスホルアミダイト法もしくはMatteucci et al. (1981) J. Am. Chem. Soc. 103:3185）によるトリエステル法、共に全体がすべての趣旨で参照として本明細書に組み入れられるか、または市販の自動オリゴヌクレオチド合成装置もしくは本明細書に記載されるハイスループット、高密度アレイ法および当技術分野で公知の同方法のいずれかを使用する他の化学法により調製できる（米国特許第5,602,244号、第5,574,146号、第5,554,744号、第5,428,148号、第5,264,566号、第5,141,813号、第5,959,463号、第4,861,571号および第4,659,774号参照、全体がすべての趣旨で参照として本明細書に組み入れられる）。予備合成されたオリゴヌクレオチドを様々な売り手から商業的に獲得することもできる。

【００２５】

本発明のある態様では、当技術分野で公知の様々なマイクロアレイ技法を使用してオリゴヌクレオチドおよび/またはポリヌクレオチドを調製できる。予備合成されたオリゴヌクレオチドおよび/またはポリヌクレオチド配列を支持体に付着できるし、あるいは光指向方法、フローチャネルおよびスポッティング法、インジェクト法、ピン系方法、およびビーズ系方法を使用してその場で合成でき、以下の参照に記述されている：McGall et al. (1996) Proc. Natl. Acad. Sci. U.S.A. 93:13555; Synthetic DNA Arrays In Genetic Engineering, Vol. 20:111, Plenum Press (1998); Duggan et al. (1999) Nat. Genet. S21:10; Microarrays: Making Them and Using Them In Microarray Bioinformatics, Cambridge University Press, 2003; 米国特許出願公開第2003/0068633号および第2002/0081582号; 米国特許第6,833,450号、第6,830,890号、第6,824,866号、第6,800,439号、第6,375,903号及び第5,700,637号; ならびにPCT出願第WO 04/031399号、第WO 04/031351号、第WO 04/029586号、第WO 03/100012号、第WO 03/066212号、第WO 03/065038号、第WO 03/064699号、第WO 03/064027号、第WO 03/064026号、第WO 03/046223号、第WO 03/040410号および第WO 02/24597号；全体がすべての趣旨で参照として本明細書に組み入れられる。

【００２６】

ある態様では、本発明のビーズは、ライブラリー、例えばゲノムライブラリー、cDNAライブラリーなどを分析するために有用である。分子ライブラリーの合成のための方法の例は、当技術分野、例えば下記に見出すことができる：DeWitt et al. (1993) Proc. Natl. Acad. Sci. USA 90:6909; Erb et al. (1994) Proc. Natl. Acad. Sci. USA 91:11422; Zuckermann et al. (1994) J. Med. Chem. 37: 2678; Cho et al. (1993) Science 261:1303; Carrell et al. (1994) Angew. Chem. Int. Ed. Engl. 33:2059; Carell et al. (1994) Angew. Chem. Int. Ed. Engl. 33:2061; およびin Gallop et al. (1994) J. Med. Chem. 37: 1233、全体がすべての趣旨で参照として本明細書に組み入れられる。本明細書でさらにライブラリーについて説明する。

【００２７】

ある態様では、本発明のビーズを半流動培地中に固定化する。半流動培地は、有機および無機物質の両方を含み、半流動培地には、ポリアクリルアミド、セルロースおよびポリアミド（ナイロン）、ならびに架橋アガロース、デキストランまたはポリエチレングリコールがあるが、それに限定されるわけではない。例えば本明細書に記載するビーズをポリマーゲル中に物理的に固定化できる。そのゲルはz方向の寸法（例えば約30ミクロン）よりもxおよびy方向の寸法（例えば数センチメートル）の方が大きい場合があり、ここで、z方向の寸法はそのゲルの中に固定化されたビーズよりも実質的に大きい（例えばゲル30ミクロン対ビーズ1ミクロン）。

【００２８】

ある局面では、半流動培地中に固定化されたビーズの少なくとも、約5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、またはそれよりも大きい割合は、それらのビーズに付着した核酸配列（例えば増幅された核酸配列）を有する。すなわち、半流動培地中に固定化されたビーズ中には空席の（すなわちそれに付着した核酸配列を有さない）ものがある場合もあれば、増幅プライマーのみを有する（すなわちそれに付着した増幅された核酸配列を有さない）ものがある場合もあり、かつ/または核酸の不均一集団（すなわち実質的に同一ではない配列）が付着しているものがある場合もある。

【００２９】

他の局面では、付着した核酸配列を含むそれぞれの固定化ビーズは、他の固定化ビーズとは異なる核酸配列を有するであろう。すなわち、50%、45%、40%、35%、30%、25%、20%、15%、14%、13%、12%、11%、10%、9%、8%、7%、6%、5%、4%、3%、2%、1%、またはそれよりも小さい割合の固定化ビーズが、一つまたは複数の他の固定化ビーズに付着した核酸配列と実質的に同一の核酸配列に付着した核酸配列（例えば増幅された核酸配列）を有するであろう。ビーズに付着した核酸配列は、ビーズに付着したすべての核酸配列、ビーズに付着したプライマーおよび/またはビーズに付着した増幅産物にあてはまりうる。

【００３０】

なお他の局面では、本発明の半流動培地を固体支持体と共に使用する。例えば、上記の節に記載したゲルの一表面が固体支持体（例えばガラス表面）に付着される一方で、他方の表面が露出されるように、そのゲルを重合できる。ある局面では、ビーズが単層を形成し、単層がそのゲルの露出した表面近くに存在するように、ゲルを流し込むことができる。

【００３１】

本発明の固体支持体を様々な形状に形作ることができる。ある態様では、その固体支持体は実質的に平面である。固体支持体の例には、スライド、マイクロタイター平板、フローセル、カバーガラス、マイクロチップなどのような平板、および微量遠沈管、試験管などのような容器、および管類、シート、パッド、フィルムなどがある。追加的に、固体支持体は、例えば生物学的、非生物学的、有機、無機、またはそれらの組み合わせでありうる。ある態様では、ビーズがその固体支持体に結合されうるようにそのビーズおよび/または固体支持体を官能化できる。官能基について本明細書にさらに論じる。

【００３２】

ある態様では、標準的な落射蛍光顕微鏡でビーズのアレイを撮像できる。固定化ビーズの使用は、多サイクルのハイブリダイゼーション/酵素系操作の後で、ビーズ上に固定化されたDNAとハイブリダイズした分子上の視覚的に検出可能な標識、またはそのビーズに固定化されたDNA自体に組み込まれた視覚的に検出可能な標識の撮像を行うことにそのアレイを供することを可能にする。本明細書でさらに記載する配列決定アッセイに、様々な検出可能な標識を使用できる。本発明に使用するための標識の例には、フルオレセイン（例えばFITC）、ローダミン（例えばTRITC、RITC）、DAPI、BODIPY、Cy3、Cy5、Alexa、テキサスレッド、カスケードブルー、緑色蛍光タンパク質（GFP）、黄色蛍光タンパク質（YFP）、シアン蛍光タンパク質（CFP）、ホースラディッシュペルオキシダーゼ、アルカリホスファターゼ、アビジン、ビオチン、ルシフェラーゼ（例えばウミシイタケ（renilla）ルシフェラーゼ、ホタルルシフェラーゼ）などのような視覚的に検出可能な標識がある。多くの適切な標識は当技術分野で公知であり、例えばMolecular Probes（Eugene, OR）およびSigma-Aldrich（St. Louis, MO）のカタログから注文することができ、全体がすべての趣旨で参照として本明細書に組み入れられる。

【００３３】

本発明の態様は、ビーズ上の核酸配列の増幅をさらに目的とする。ある局面では、オリゴヌクレオチドを増幅する方法には、以下でさらに記載するエマルジョンPCRがある。核

酸配列を増幅する他の方法には、反応生成物を捕捉するためのプライマーを有するビーズを使用したエマルジョン中でのローリングサークル型増幅（超分岐または直鎖）；水溶液中でのローリングサークル型増幅（超分岐または直鎖）後のビーズ上へのクローンの「捕捉」；エマルジョン中でのヘリカーゼ置換増幅（HDA）；および例えばSiO_2表面オリゴまたはゲル固定化オリゴ薄層を使用したその場のローリングサークル型増幅があるが、それに限定されるわけではない。

【００３４】

ある局面では、オリゴヌクレオチドの増幅法は、アンカーPCRまたはRACE PCRのようなPCRの使用を伴うか、または連結鎖反応（ligation chain reaction）（LCR）（例えばLandegran et al. (1988) Science 241:1077-1080；およびNakazawa et al. (1994) Proc. Natl. Acad. Sci. U.S.A. 91:360-364参照；全体がすべての趣旨で参照として本明細書に組み入れられる）に関与する。代替的な増幅法には、自立配列複製（self sustained sequence replication）（Guatelli et al. (1990) Proc. Natl. Acad. Sci. USA 87:1874、全体がすべての趣旨で参照として本明細書に組み入れられる）、転写増幅系（Kwoh et al. (1989) Proc. Natl. Acad. Sci. US. 86:1173、全体がすべての趣旨で参照として本明細書に組み入れられる）、Q-ベータレプリカーゼ（Lizardi et al. (1988) BioTechnology 6:1197、全体がすべての趣旨で参照として本明細書に組み入れられる）、反復PCR（Jaffe et al. (2000) J. Biol. Chem. 275:2619；およびWilliams et al. (2002) J. Biol. Chem. 277:7790；全体がすべての趣旨で参照として本明細書に組み入れられる）または当業者に周知の技法を使用した他の任意の核酸増幅法がある。様々な増幅法が、米国特許第6,391,544号、第6,365,375号、第6,294,323号、第6,261,797号、第6,124,090号及び第5,612,199号に記載されており、全体がすべての趣旨で参照として本明細書に組み入れられる。

【００３５】

本発明の態様は、本明細書に記載する増幅法を使用してオリゴヌクレオチドを増幅する方法を目的とする。ある態様では、従来法を使用してオリゴヌクレオチドの3'末端での増幅部位に増幅プライマーを選択的にハイブリダイズすることによってオリゴヌクレオチドを増幅する。増幅プライマーは、6から100までであり、1,000までさえあるヌクレオチド長であるが、概して10から40ヌクレオチド長である。しかし、種々の長さのオリゴヌクレオチドが使用される。増幅プライマーは、例えばエマルジョンPCRなどでは溶液中に存在しうるし、かつ/または本明細書に記載する半流動培地中に存在しうる

【００３６】

概して、二つの核酸配列が少なくともヌクレオチド14から25個の伸長にわたり実質的に相補的、すなわち少なくとも、約65%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%、または100%相補的である場合に、選択的ハイブリダイゼーションが起こる。Kanehisa, M., 1984, Nucleic Acids Res. 12: 203参照、全体がすべての趣旨で参照として本明細書に組み入れられる。

【００３７】

全般的に、五つの要因が第二の核酸分子に対するプライマーのハイブリダイゼーションの効率性および選択性に影響する。これらの要因は、（ｉ）プライマー長、（ii）ヌクレオチドの配列および/または組成、（iii）ハイブリダイゼーション温度、（iv）緩衝液の化学的性質、ならびに（ｖ）プライマーがハイブリダイズする必要のある領域での立体障害のポテンシャルであるが、ランダムではないプライミング配列を設計する場合には重要な検討材料である。

【００３８】

プライマー長と、プライマーが標的配列にアニールするであろう効率性および正確度の両方との間に正の相関が存在し、長鎖配列は、短鎖配列よりも高いT_mを有し、所与の標的配列内に繰り返しがある見込みが低いことによって、雑多なハイブリダイゼーションが削減される。高GC含量を有するか、またはパリンドローム配列を含むプライマー配列は、それらの目的となる標的部位と同様に、自己とハイブリダイズする傾向がある。それは、二

分子よりもむしろ単分子のハイブリダイゼーション動態が一般に溶液中で有利であるからである。同時に、その標的配列に緊密に結合するG-Cヌクレオチドの対形成を十分な数だけ含むプライマーを設計することが重要である。それは、そのような対のそれぞれが3個の水素結合で結合するのに比べ、AおよびT塩基が対形成する場合は2個の水素結合がみられるからである。ハイブリダイゼーション温度はプライマーのアニーリング効率と逆に変動し、ハイブリダイゼーション混合物に含まれうる有機溶媒、例えばホルムアミドの濃度はこれと同様であるが、塩濃度の増加は結合を促進する。ストリンジェントなハイブリダイゼーション条件下で、長鎖プローブは短鎖プローブよりも効率的にハイブリダイズするが、より許容性の条件では短鎖プローブが十分である。ストリンジェントなハイブリダイゼーション条件は、概して約1M未満、より通常には約500mM未満、好ましくは約200mM未満の塩濃度を含む。ハイブリダイゼーション温度は、低くは0℃から、22℃よりも大きい、約30℃よりも大きい、約37℃を上回る（最も頻繁には）までの範囲である。長鎖フラグメントは、特異的ハイブリダイゼーションにさらに高いハイブリダイゼーション温度を必要としうる。数個の要因がハイブリダイゼーションのストリンジェンシーに影響することから、パラメータの組み合わせが任意の一つのパラメータ単独の絶対的な尺度よりも重要である。ハイブリダイゼーション条件は当業者に公知であり、Current Protocols in Molecular Biology, John Wiley & Sons, N. Y. (1989), 6.3.1-6.3.6に見出すことができ、全体がすべての趣旨で参照として本明細書に組み入れられる。

【００３９】

上記の最初の四つの検討材料を念頭において、プライマーを設計する。非常に多くの配列の相対的真価の推定を心の中で行う一方で、これらのいくつかのパラメータの評価およびプライマー配列の最適化を援助するためにコンピュータプログラムが設計されている（例えば、Hoover et al. (2002) Nucleic Acids Res. 30:e43およびRouillard et al. (2004) Nucleic Acids Res. 32:W176参照、全体がすべての趣旨で参照として本明細書に組み入れられる）。

【００４０】

ある例によると、対象となる核酸配列が付着したビーズを濃縮するための方法が提供される。対象となる核酸配列と相補的核酸配列とのハイブリダイゼーションを可能にする条件下で、ビーズの集団を、対象となる核酸配列に相補的な核酸配列（すなわち相補的核酸配列）と接触させることにより、対象となる核酸配列を有するビーズを濃縮できる（ここで、少なくとも一つのビーズはそれに付着した対象となる核酸配列を有する）。次に、当技術分野で公知の方法を使用して、ビーズに付着した対象となる核酸配列を含むハイブリダイズしたビーズから、ハイブリダイズしないビーズを分離する。ある局面では、初発のビーズ集団での対象となる核酸配列が付着したビーズの割合に対して、対象となる核酸配列が付着したビーズを少なくとも、約5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、またはそれよりも大きく濃縮する。ある態様では、相補的核酸配列を支持体上に固定化する。適切な支持体には、合成ポリマー支持体、例えばポリスチレン、ポリプロピレン、置換ポリスチレン（例えばアミノ化またはカルボキシル化ポリスチレン）、ポリアクリルアミド、ポリアミド、ポリビニルクロリドなど、ポリマービーズ、磁気ビーズ、ガラスビーズ、セファロース、アガロース、セルロース、またはアフィニティクロマトグラフィーに有用な任意の材料があるが、それに限定されるわけではない。対象となる核酸配列を有するビーズを濃縮する方法を、本明細書にさらに記載する。

【００４１】

ある例によると、ビーズ上の核酸配列を配列決定する方法が提供される。可逆ターミネータ（reversible terminator）を用いた伸長による配列決定、蛍光インサイチュー配列決定（FISSEQ）、ピロシーケンシング、大規模並列シグニチャシーケンシング（massively parallel signature sequencing）（MPSS）など（Shendure et al. (2004) Nat. Rev. 5:335に記載、全体として参照により本明細書に組み入れられる）のような当技術分野で

公知の一般的な配列決定法は、本明細書に記載されるビーズおよびビーズ系アレイを用いた使用に適する。可逆終止法（reversible termination method）は合成による段階的配列決定という生化学反応を使用し、その生化学反応は可逆終止および除去可能な蛍光と連結している（上記Shendure et al.および米国特許第5,750,341号および第6,306,597号、参照により本明細書に組み入れられる）。FISSEQは、

反応物にただ一つの種類の蛍光標識ヌクレオチド三リン酸を添加する段階；

組み込まれていないヌクレオチドを洗浄して除く段階；

蛍光を測定することによりヌクレオチドへの取り込みを検出する段階；および

そのサイクルを繰り返す段階；

を含む、DNAが伸長する方法である。各サイクルで、以前のサイクルからの蛍光を退色させるか、もしくはデジタル減算するか、またはフルオロホアをヌクレオチドから開裂させて洗浄して除く。FISSEQはさらにMitraら (2003) Anal. Biochem. 320:55に記載されており、全体がすべての趣旨で参照として本明細書に組み入れられる。ピロシーケンシングは、ヌクレオチド取り込みの各イベントの間に（すなわちヌクレオチドが成長中のポリヌクレオチド配列に加えられた場合に）ピロリン酸塩（PPi）が放出される方法である。DNAポリメラーゼ触媒反応で放出されたPPiを、ATPスルフリラーゼおよびルシフェラーゼにより共役反応で検出する。これを視覚的に検出できる。添加したヌクレオチドは、ヌクレオチド分解酵素により連続的に分解する。最初に添加したヌクレオチドが分解した後で、次のヌクレオチドを添加できる。この手順を繰り返すにしたがい、より長い伸長のテンプレート配列が推定される。ピロシーケンシングはさらにRonaghi et al. (1998) Science 281: 363に記載されており、全体がすべての趣旨で参照として本明細書に組み入れられる。MPSSは、連結に基づくDNA配列決定をマイクロビーズ上で同時に利用する。すべての可能性のあるオーバーハングを含む標識アダプターの混合物を四つのヌクレオチドの標的配列にアニールする。アダプターの連結に成功した場合に標識を検出する。次に、制限酵素を使用してDNAテンプレートを開裂して、次の4塩基を露出させる。MPSSはさらにBrenner et al. (2000) Nat. Biotech. 18:630に記載されており、全体がすべての趣旨で参照として本明細書に組み入れられる。

【００４２】

本発明を、以下の実施例によりさらに例示するが、これらの実施例を限定するものとして解釈してはならない。本出願にわたり引用したすべての参照、特許および公表された特許出願の内容は、全体としてすべての趣旨で参照として本明細書に組み入れられる。

【００４３】

実施例I

クローンビーズ

段階1：ビーズへのオリゴヌクレオチドのカップリング

Integrated DNA Technologies（Coralville, IA）から得た、商業的に入手可能なオリゴヌクレオチドを使用した。それらのオリゴヌクレオチドの5'末端の二重ビオチン部分を介して、ストレプトアビジンを被覆した常磁性ビーズ（1μ M MYONE（商標）ビーズ（Dynal Biotech, Brown Deer, WI））にそれらのオリゴヌクレオチドを付着した（図1）。オリゴヌクレオチドの配列は、下記のライブラリー発生プロトコルに記載するライブラリー分子のPR1-F区域と同一であった（実施例VI）。この段階の後で、本明細書に記載するテンプレートライブラリーを増幅する間に、ビーズに固定化されたオリゴヌクレオチドがPCRプライマーとして作用することがこれにより可能になる。本明細書に記載するプロトコルでは、この配列は「フォワード」PCRプライマーと称される。

【００４４】

フォワードプライマーを載せたビーズを発生させるための段階は、以下の通りであった。

1）微量遠沈管の側面にビーズを引きつける磁界を使用して、1× 10^{7}個のMYONE（商標）ストレプトアビジン常磁性ビーズ（原液100μ L）をTE 200ulで一回洗浄した。

2）ビーズをBind & Wash緩衝液180μ l（5mM Tris-HCl（pH7.5）、0.5mM EDTA、1.0M NaCl

）に再懸濁した。
3）100μｌ（2nmole）の5'二重ビオチン化フォワードPCRプライマー
PR1-F-2BIO: 2Bio-CCACTACGCCTCCGCTTT
CCTCTCTATGGGCAGTCGGTGAT (SEQ ID NO:1)

20μｌと共にビーズを25℃で30分間インキュベートした。
4）微量遠沈管の側面にビーズを引きつける磁界を使用して、ビーズをTE 200μｌで2回洗浄した。
5）ビーズををTE 200μｌ（ビーズ5× 10^9個/μｌ）に再懸濁して、次の段階で使用するまで4℃で保存した。
【００４５】
段階2：マイクロエマルジョンの調製
油中水型エマルジョン中でPCR反応を実施した。エマルジョンの使用は、エマルジョンのごく少量の水性構成要素を相互に隔離した。その結果、PCR試薬類およびPCR反応産物類は温度サイクリング反応にわたって相互に隔離された。二つのPCRプライマーの一つとして、ビーズ（例えば段階1からのビーズ）に固定化したオリゴヌクレオチドを使用することにより、PCR反応の終了時にそれらのビーズに物理的に固定化されたPCR産物が生じた。そのPCR反応に使用したテンプレートは、ポリヌクレオチドの複合混合物からなり、それらのポリヌクレオチドは、分子により異なる介入配列を有する二つの隣接「共通」領域を含んだ。本実施例では、大腸菌（E. coli）ゲノムライブラリーを使用した。このライブラリーの特有の領域に隣接する共通配列を、本明細書では「PR1-F」および「PR1-R」と称する。ビーズを含むエマルジョンの大部分の水性構成要素がPCR反応の開始時に0個または1個のテンプレートのいずれかを含むように、低濃度のテンプレートをPCRに使用した。PCR反応の終了時に、多くのビーズはテンプレートを全く含まない区画に存在したことから、「空席」であった。他のビーズは、同じPCR産物の数千のコピーを含む点で「クローン」であった。各クローンビーズのPCR産物は、他のクローンビーズのPCR産物と異なった。あまり一般的でないのは、二つ以上のテンプレートを有する区画に存在するビーズであったが、これらのビーズもまた存在した。第三の種類の「非クローン」ビーズは、一つよりも多いテンプレートから得られたPCR産物の混合物を含んだ。このように、ビーズの分布は、空席、クローン、および非クローンの三つの区分に入り、これをポアソン分布を用いてモデル化できた。
【００４６】
A.エマルジョンの油相の調製
驚くことに、エマルジョンの安定性は油相の各成分の比率に非常に敏感であることが見出された。したがって、ある実験から次の実験への可変性が最小であることを確実にするために、各成分を測定する場合には特に注意を払った（例えばピペット洗い込み、容積形シリンジによる測定など）。10%（v/v）Span 80を量るためにシリンジを使用すると、エマルジョンPCRの一貫した性能が可能になった。
1）10mlまたは1ml容のシリンジと16ゲージの針を用いて、鉱物油に溶かした10%（v/v）Span 80を調製した。
a）軽油9ml
b）SPAN（登録商標）80（Sigma-Aldrich, St. Louis, MO）1ml
2）ピペット洗い込みにより以下を微量遠沈管に加えた。
a）軽油（Sigma-Aldrich）545μｌ
b）鉱物油に溶かした10%SPAN（登録商標）80 450μｌ
c）TWEEN（登録商標）80（Sigma-Aldrich）4μｌ
d）TRITON（登録商標）X-100（Sigma-Aldrich）0.5μｌ
3）得られた溶液をボルテックスミキサーに30秒間かけて完全に混合した。
【００４７】

B.エマルジョンの水相の調製

ヌクレオチドの濃度を下記に示した高い量に増加させることで、本明細書に記載するビーズ-ゲル撮像系に、より大きいシグナルが生じることが見出された。$MgCl_2$濃度の同時増加も、ヌクレオチド：$MgCl_2$の比を10：1に十分近づけて保つために必要であった。

【００４８】

以下のものを微量遠沈管に加えて、静かにピペッティングすることで混合した。10× $MgCl_2$-PLATINUM（登録商標）Taq PCR緩衝液8.0μｌ（Invitrogen, Carlsbad, CA）、50mM $MgCl_2$30μｌ（18.75mM）（Invitrogen）、（各）25mM dNTP混合物11.3μｌ（3.5mM）（Invitrogen）、2mM未修飾逆PCRプライマーPR1-R 1.0μｌ（Integrated DNA Technologies (IDT), Coralville, IA）：

CTGCCCCGGGTTCCTCATTCTCT (SEQ ID NO:2)

、10μM未修飾短鎖フォワードPCRプライマーPR1-3LF：IDT 0.4μｌ

CCTCTCTATGGGCA GTCGGTGAT (SEQ ID NO:3)

、PR1-Fフォワードプライマーを載せる1ミクロンビーズ5μｌ、滅菌dH_2O 20.5μｌ、PLATINUM（登録商標）Taq 4.5μｌ（Invitrogen、5U/μｌ）、および1nMテンプレートDNA 0.25μｌ。

【００４９】

C.油中水型エマルジョンの調製

以下のエマルジョンの調製は、ビーズ5μｌ（約2.5× 10^9個）用である。使用したテンプレートの量から、一般に10%の増幅されたビーズが生じた。概して、スライド1枚あたりビーズ5～50μｌを濃縮の非存在下で使用するか、またはスライド1枚あたり濃縮したビーズ50～200μｌを使用した。水相に対する油相の比がエマルジョンの安定性に影響することが見出され、1：6の水相：油相比が一貫して安定なエマルジョンをもたらすことが決定された。

1）油相400μｌを、磁気微量撹拌バーを入れて1400RPMに調整した閉ループ磁気撹拌プレートに置いた2ml容の丸底凍結バイアルに加えた。

2）撹拌中の油相に水相75μｌを滴下した。

3）得られた混合物をVWR Scientific565型磁気撹拌装置の上で磁気微量撹拌バー（no.58948-353, VWR Scientific, West Chester, PA）で1400RPMで30分間撹拌した。

4）その管の内容物を200μｌ容試験管× 8（各50μｌ）に分けた。

【００５０】

D.温度サイクリング

PCRサイクル数の増加および伸長時間の増加がシグナルを増加することが見出された。よって、75秒の伸長期を有する120回のPCRサイクルを使用した。

【００５１】

以下のプログラムによりエマルジョンの温度サイクリングを行った（分：秒）。

a）94℃、2：00

b）94℃、0：15

c）57℃、0：30

d）70℃、1：15

e）段階b）～d）をさらに119回繰り返す

f）72℃、2：00

g）使用まで4℃

【００５２】

E.エマルジョンからのビーズの回収

以下のプロトコルを使用してエマルジョンからビーズを回収した。

1）200μｌ容PCR管× 8の内容物を1本の1.5μｌ容微量遠沈管にプールした。

2）NX緩衝液（100mM NaCl；1% TRITON（登録商標）X-100；10mM Tris-HCl（pH7.5）；1m

M EDTA）800μlを加えた。
3）その管を30秒間ボルテックスミキサーで撹拌した。
4）その管を11,000RPMで1.5分間遠心分離した。
5）ペレットを乱すことなく上清約1150μlを取り出した。
6）段階2~5をさらに2回繰り返した。
7）微量遠沈管の側面にビーズを引きつける磁気分離を使用して残りの液体を取り除いた。
8）ビーズをTE 50μlで2回洗浄した。
9）ビーズをTE 5μlに再懸濁した。
【００５３】
F.伸長していないフォワードプライマーをビーズから取り除くためのエキソヌクレアーゼ処理
エマルジョンPCRの後で、「増幅された」ビーズに、PCR産物（1本の鎖がビーズに固定化された2本鎖）と、残りの伸長していないフォワードプライマー（1本鎖）との両方を載せた。残りの伸長していないフォワードプライマーは核酸分析のその後の段階でバックグラウンドシグナルの供給源となりうることが見出された。よって、それを除去することが理想的であった。エキソヌクレアーゼIを使用して残りの伸長していないプライマーを選択的に消化した。
1）以下を混合した。
ビーズ+dH_2O 86μl
10× エキソヌクレアーゼI反応緩衝液（New England Biolabs, Beverly, MA） 10μl
エキソヌクレアーゼI（20単位/μl、New England Biolabs） 4μl
2）得られた混合物を37℃で1時間インキュベートし、30分の時点で1回混合してビーズの沈降を打ち消した。
3）その混合物を80℃で20分間インキュベートしてエキソヌクレアーゼを不活性化した。
4）磁気分離を使用して、ビーズをNX緩衝液200μl× 5で洗浄した。
5）使用までビーズを4℃で保存した。
【００５４】
G.水酸化ナトリウムによるビーズに結合した産物の1本鎖化
本段階は、ビーズに固定化したPCR産物の効率的な1本鎖化を可能にするために重要であることが見出された。
1）磁気分離によりビーズからすべての液体を除去
2）0.1M NaOH 50μlを加え、ビーズと混合
3）25℃で10分間インキュベート
4）0.1M NaOH 50μlで1回洗浄
5）TEで2回洗浄
6）TE 50μlにビーズを再懸濁
【００５５】
このプロトコルは、当技術分野で公知のビーズ系方法を使用して発生したライブラリーよりも数桁複雑さの大きいライブラリーの増幅を可能にする。
【００５６】
2ml容丸底凍結バイアル（no.430661, Corning）に入れた油相400μlに水相200μlを滴下することにより、油中水型マイクロエマルジョンを調製できる。VWRモデル565磁気撹拌装置の上で磁気微量撹拌バー（no.58948-353, VWR Scientific）でその混合物を1,400rpmで撹拌しながら、滴下を1分間かけて実施できる。水相の添加の後で、得られた混合物を合計30分間撹拌できる。磁気撹拌装置の中央に置いたラックに2本の管を配置することにより、一度に2個のエマルジョンを作製できる。
【００５７】
実施例II
ビーズの濃縮

以下のプロトコルを使用して、クローンビーズから空席のビーズを分離できる。エマルジョンPCR反応での低テンプレート濃度および濃縮プロトコルの併用は、その他で可能であろうよりも高い割合の「クローン増幅された」ビーズをもたらした。濃縮の基本は、第二セットの大型（直径3ミクロン）、非磁性ビーズ（すなわち「捕捉ビーズ」）を使用することであり、その大型ビーズは「リバース」PCRプライマー配列（PR1-R）と同一の配列を有するプライマーを含む。リバースPCRプライマーに相補的配列は、PCR反応産物であるDNA鎖にのみ存在するであろうことから、増幅されたビーズはこれらの大型捕捉ビーズと選択的にハイブリダイズしたが、空席のビーズは捕捉ビーズと効率的にハイブリダイズしなかった。密度の差に基づいて（例えば密度勾配溶液を通過してビーズを遠心分離することにより）、1ミクロンの増幅ビーズとハイブリダイズした3ミクロンの捕捉ビーズを、1ミクロンの空席のビーズから分離した。

【００５８】

捕捉ビーズを以下のように調製した。SPHERO（商標）ポリスチレンストレプトアビジン被覆ビーズ50μl（非常磁性、直径3ミクロンのビーズ、Spherotech, Libertyville, IL）を1.5μl容微量遠沈管にピペットで入れ、13.2krpmで30秒間遠心分離してペレットにし、Bind & Wash緩衝液50μl中に再懸濁した。ビーズをもう一度遠心分離し、すべての液体をビーズのペレットから取り出し、そのビーズをBind & Wash緩衝液49.5μlに再懸濁した。1mMビオチン修飾「捕捉プライマー」

PR1-R-BioXL: ビオチン-5'-

cgtaccccgcttggtctttctccccgtaccccgcttggtctttctccCTGCCCCGGGTTCCTCATTCT CT

(SEQ ID NO:4)

0.5μlを加えた。時々混合しながら、ビーズを室温で20分間インキュベートした。ビーズを30秒間遠心分離してペレットにし、液体を除き、Bind & Wash緩衝液50μlを加えた。この洗浄段階を繰り返し、ビーズをBind & Wash緩衝液10μl中に再懸濁した。得られた捕捉ビーズを使用まで4℃で保存した。

【００５９】

TEに入れた1ミクロンの加工されたビーズ40μl（上節で発生した増幅されたビーズと増幅されていないビーズとの混合物）をBind & Wash緩衝液20μlに再懸濁した。3ミクロンの捕捉ビーズ10μlを1ミクロンの加工されたビーズ20μlに加え、それらのビーズをピペッティングにより混合した。その混合物を56℃で10分間インキュベートすることにより捕捉ビーズを加工されたビーズとハイブリダイズした。次に、1.5μl容微量遠沈管に入れた60%（v/v）グリセロール150μlの上面にその混合物を注意深くピペットでのせた。微量遠沈管を13.2krpmで1分間遠心分離した。3ミクロンの非磁性ビーズと1ミクロンの磁性ビーズとの密度の差が原因で、3ミクロンのビーズは（それらのビーズとハイブリダイズした増幅された1ミクロンのビーズと共に）上清に残ったが、ハイブリダイズしない、未増幅の1ミクロン磁性ビーズは、管の底にペレットを形成した。得られた液体を微量遠沈管から取り出し（管の底にあるビーズのペレットを乱すことなく）、新しい微量遠沈管にピペットで入れた。捕捉ビーズとハイブリダイズして増幅されたビーズは、この上清中に濃縮された。

【００６０】

次に、増幅された1ミクロンのビーズの濃縮画分を精製して捕捉ビーズを除いた。その上清を含む新しい管に水1mLを加え、ピペッティングにより混合し、13.2krpmで2分間遠心分離した。20から30μlの液体以外のすべてを取り除き、真水50μlを加え、混合し、13.2krpmで2分間もう一度遠心分離した。すべての液体をビーズのペレットから除き、そのビーズを0.1M水酸化ナトリウム50μlに再懸濁した。1ミクロンのビーズを3ミクロンの捕捉ビーズと解離させるために、時々混合しながらビーズを10分間インキュベートした。磁界を微量遠沈管に適用して、管の側面に1ミクロンの磁気ビーズを引きつけ、すべての上清を取り除いた。3ミクロンの捕捉ビーズを含む上清は白濁していた。1ミクロンのビーズを

0.1M NaOHで1回、1× PCR緩衝液で3回洗浄した。1ミクロンのビーズをTE 5μ Lに再懸濁して4℃で保存した。

【００６１】

図2A～2Cは、上記のようなプロトコルを経たビーズのセットを示す。ビーズを様々な段階で取り出して、空席のビーズに対する増幅されたビーズの割合を調べた。これらの疑似色表示の図で、緑色のビーズは空席であり、赤色のビーズは増幅されていた。エマルジョンPCRにより発生した加工されたが未濃縮の1ミクロンビーズである出発材料のうち、約8％のビーズが増幅された（図2A）。上記手順により除去されたビーズのセットでは、1％未満が増幅されていた（図2C）。増幅されたビーズに関して濃縮された材料の画分では、増幅されたビーズの割合は8％から43％に上昇し（図2B）、これは5.5倍の濃縮であった。

【００６２】

実施例III

無秩序化された固定化ビーズの単層

高分解能顕微鏡の視野深度に関連して、単層化の決定的な性質を理解することができる。例えば、分解能の点からみて商業的に入手可能な最高の対物レンズの一つである20× Plan Apo（NA＝0.75）対物レンズの視野深度は1.9ミクロンである。クローンミクロスフェアが直径わずか1ミクロンの場合、単層からの重大な変位により、アレイを形成したビーズの撮像が不可能な結果となるであろうが、所与の視野内ですべてのビーズに対する焦点は維持される。

【００６３】

以下の段階を増幅の前後実施できよう（例えばクローンミクロスフェアの集団を作製するための方法によって）、また場合によりビーズの濃縮プロトコルが行われた後でもよい。ビーズを単層化した後で、並列的な配列決定または他の形態の環状核酸分析を行うであろう。

【００６４】

単層のビーズを形成するために、以下のプロトコルを使用した。以下の試薬を混合した。（所望の密度の）ビーズ3.00μ l、dH_2O 5.10μ l、アクリルアミド：ビス（38％アクリルアミド、2％ビス-アクリルアミド、Roche）1.50μ l、RHINOHIDE（商標）（ポリアクリルアミドゲル補強剤）0.60μ l（Molecular Probes, Eugene, OR）、5％ N,N,N',N'-テトラメチルエチレンジアミン（TEMED）1.20μ l、および過硫酸アンモニウム溶液（APS）（0.5％）1.80μ l。

【００６５】

テフロン被覆したスライドグラスとカバーガラスとの間に混合した試薬を流し込み、重合させた結果、厚さ30ミクロンのゲルをもたらした。カバーガラスまたはスライドグラスのいずれかにシラン結合剤（Bind Silane）（3-メタクリルオキシプロピルトリメトキシシラン）を被覆して、ゲルが重合後にスライドグラスまたはカバーガラスのいずれかに接着するようにした。シラン結合剤中でガラス表面を被覆するためのプロトコルは、例えばMitraら（同上）に記載されている。簡潔には、酢酸220μ lおよびシラン結合剤試薬4mLをdH_2O 1リットルに入れて混合した。静かに振盪しながら1時間浸漬することにより、ガラス表面をこの溶液に曝露した。dH_2Oに3回、100％エタノールに1回浸漬することによりガラス表面を洗浄した。スライドを風乾して、乾燥しながら保存した。

【００６６】

重合の間に、ビーズがゲルの所望の側に沈降するような方向にそのスライドグラスおよび重合中のゲルを配置した。このように、ビーズはゲルの露出面に対して「上層」または「底層」のいずれかを形成するであろう（図3）。ビーズに固定化したDNAに対してハイブリダイゼーションおよび酵素反応を行うために、「上層」のゲルの中のビーズの方がそれらのビーズに適用された酵素およびオリゴヌクレオチドにはるかに接近可能であり、はるかに効率的な反応動態を許容することが見出された。これは意外なことではない。それは、ゲルに問い合わせするために使用する多くの酵素反応は酵素によって行われるが、ビーズがゲル内部深くに実際存在するならば、それらの酵素はビーズに接近するのに困難を有

するであろうからである。ゲルは厚さ30ミクロンであるが、ビーズはわずか厚さ1ミクロンであった。ゲルの孔は大きかったが、ビーズが例えばゲルの底に位置するのであれば、ゲルを通過する酵素および/またはオリゴヌクレオチドの拡散は遅いであろう。

【００６７】

最終産物は、アクリルアミドゲルがシラン結合剤試薬を介して付着されたガラス表面からなった（図3）。ゲルは厚さ約30ミクロンであった。ゲルの上面に単層になって（すなわちシラン結合剤を介してゲルが付着するカバーガラスから最も離れて）ビーズは存在した。

【００６８】

本実施例は、触媒（APSおよびTEMED）の濃度が、当技術分野でゲルの重合に概して使用される濃度よりもはるかに低いので、ゲルを流し込んだ後でゲルは低速（約1時間）で重合し、ビーズは重力で沈降して、ゲルの片面に単層になるであろうことを提供している。重合工程が速すぎるならば、ビーズは単層になって沈降する時間を有さず、その結果として単一の焦点面に視覚化できないビーズの層となることが確定された。ゲルの露出面に単層になってビーズが沈降するように重合中のゲルを上下逆さにする行為は、酵素反応を可能にするために有利であり、ビーズ上のDNAを操作するためにその酵素反応をその後使用する。

【００６９】

ビーズまたは他の支持体の「秩序化された」アレイ（すなわち当技術分野のビーズ系アレイ） を使用するための動機づけは、直交座標を有する秩序化されたアレイの形態を概してとり、同じ特性を繰り返し独立して観察する欲求または「アドレス指定能力」の欲求であった。本実施例に記載したアレイは、単層である点でz軸に対して秩序化しているが、x軸およびy軸に対しては無秩序化している。個々のビーズが高密度で視覚化および識別され、ゲルの所与の小区分が繰り返し独立して観察され、その観察結果をゲルの同じ小区分の以前の観察結果に対して整列することによって、各小区分内で個々のビーズに多数の独立した観察を行うことが可能になった。ビーズの位置は、たとえ多数の酵素操作、95℃の熱への曝露などを受けても不変のままであった。このように、無秩序化された単層の使用は、秩序化されたアレイを発生することに伴う多くの困難を回避するが、その一方で、同じ目標を実現する。

【００７０】

実施例IV

一塩基伸長によるプライマーハイブリダイゼーションおよび配列決定

プロトコルのこの段階で、ビーズはアクリルアミドポリマーゲルの露出した表面に単層になって存在し、そのゲル自体はカバーガラスに付着している。よって、ビーズ表面のDNAを、ゲル中での絶対位置を乱すことなく様々な試薬および条件に容易に曝露できる。配列特異的オリゴヌクレオチドハイブリダイゼーション、ポリメラーゼ指令プライマー伸長、制限エンドヌクレアーゼ指令配列特異的開裂、リガーゼ指令オリゴヌクレオチド連結、エキソヌクレアーゼ指令DNA分解などを含むが、それらに限定されるわけではない、多様な酵素反応および化学反応を、この形式に固定化したDNAに関して行えることが見出された。

【００７１】

これらの酵素反応の例として、オリゴヌクレオチドプライマーをハイブリダイズしてから、蛍光ヌクレオチドでポリメラーゼ指令一塩基伸長を行うためのプロトコルを記載する。

【００７２】

ビーズ上のPCR産物が一本鎖であることから、PCR産物は、今では5'から3'方向にかけて、ビーズに鎖を固定化している二重ビオチン部分、PR1-Fフォワードプライマー配列、未知配列（ライブラリーが構築された材料に依存してビーズ毎に可変である）、およびPR1-Rプライマーに相補的な配列からなる。ビーズに固定化したDNAにPR1-Rプライマーをハイブリダイズするために、ハイブリダイゼーション緩衝液（0.01% TRITON（登録商標）X-1

00を有する6× SSPE）に溶かした1μ M PR1-R 100μ lをゲルに適用し、フレームシール（FrameSeal）チャンバー125μ lを用いて密封して、液体をゲルと接触させた。スライドを56℃で10分間加熱し、フレームシールチャンバーを取り除き、スライドを1× 洗浄緩衝液（10mM Tris（pH7.5）；50mM KCl；2mM EDTA；0.01% TRITON（登録商標）X-100）に浸漬した。洗浄ではスライドを振盪しながら2分間インキュベートし、1× 洗浄緩衝液中でさらに2回洗浄した。このように、PR1-RプライマーをDNAとハイブリダイズした。PR1-Rプライマーの3'末端を、ビーズに固定化した分子の未知配列の直近に配置した。

【００７３】

ポリメラーゼ指令単塩基伸長による第一未知塩基の同一性を問い合わせるために、ポリメラーゼとフルオロホア標識ddNTPとの両方を含む以下の混合物を調製した。1× ThermoSequenase（商標）緩衝液122μ l（Amersham Biosciences, Piscataway, NJ）、ThermoSequenase（商標）1μ l（4u/uL, Amersham Biosciences）、R110-ddGTP 0.5μ l（100μ M、PerkinElmer, Wellesley, MA）、Cy5-ddCTP 0.5μ l（100μ M、PerkinElmer）、Cy3-ddUTP 0.5uL（100μ M、PerkinElmer）、およびテキサスレッドddATP 0.5μ l（100μ M、PerkinElmer）。

【００７４】

フレームシールチャンバー125μ lを使用して上記のように混合物をスライドに適用し、42℃で5分間インキュベートして伸長させた。浸漬物を室温でゆっくり撹拌しながら1× 洗浄緩衝液中で2分間洗浄し、その洗浄を1から2回繰り返してから撮像した。キセノンランプ光源と本明細書に使用したフルオロホアに適したフィルターセットとを備える落射蛍光顕微鏡で撮像を行った。四つのフルオロホアの一つだけに対応するフィルターにより各ビーズは蛍光を発し、これはDNA配列に取り込まれた塩基の同一性、よって未知の位置の同一性を明らかにした。

【００７５】

実施例V

落射蛍光顕微鏡での撮像

カバーガラスに固定したゲルにビーズを固定化した。ゲルおよびガラスを両方見通して撮像できた。使用した%でアクリルアミドゲルは重大な自己蛍光も生じず、光の透過も妨害しなかったことが確定された。

【００７６】

ビーズに固定化したDNAに組み込まれたか、またはそのDNAに会合した分子上の蛍光部分を撮像するために、自動X-Yステージ（Prior）および焦点コントロールを備える落射蛍光顕微鏡（Nikon TE2000）を使用した。落射蛍光照明は、水銀アークランプ、キセノンアークランプ、またはハロゲン化水銀アークランプのいずれかによった。顕微鏡に取り付けたCCD系検出器により画像を取得した。典型的な倍率が10× ~ 40× である長作動距離対物レンズまたは高開口数対物レンズのいずれかを使用した。概して、スライドグラスに付着されたゲルにビーズを固定化した。または、時にはフローセルに取り付けたカバーガラスまたはフローセルに取り付けたスライドグラスに固定化ビーズ-ゲルを付着した。フローセルまたはスライドグラスのいずれかは、細部までステージに適合する。スライドグラスの場合、画像取得のサイクルの間に顕微鏡からそのスライドを取り外して実験プロトコル（例えば配列決定）を行うことができた。画像取得サイクルの間にスライドをステージから取り外すことに伴う一つの問題は、ミクロン単位の精度で顕微鏡のステージにそのスライドを正確に再配置できないことである。典型的には、100ミクロン以上の再配置誤差が予想されるであろう。40倍の倍率で7ミクロンのCCD画素サイズでは、この誤差は、あるサイクルから次のサイクルへの画像のずれが原因で、75%以上のデータ喪失をもたらすであろう。この問題は、スライドまたはフローセルを顕微鏡から取り外さずにスライドまたはフローセルの温度のゆらぎを伴う実験プロトコルを行ったとしても、より少ない程度で同じように起こりうることに留意すべきである。そのような温度変化は、スライドまたはフローセルの幾何形状に機械的変化を誘導するであろうし、画像のずれをもたらしうる。撮像およびスライドの取り外しの連続サイクルの後で、XおよびYで数ミクロン、Zで1ミクロン

未満以内になるようにX、YおよびZ軸への顕微鏡のステージの正確な再配置を可能にするために、以下の段階を使用できる。

【００７７】

段階1

実験の開始時に、撮像する表面の「焦点地図」を発生させた。この焦点地図は、顕微鏡が各取得サイクルの間に視察するためのビーズアレイx、y、z座標のリストである。この焦点地図は、対応するz位置を発生するためにそのリストでの各x、y位置での「自動焦点」ルーチンを遂行することにより通常は発生する。使用した自動焦点アルゴリズムは、METAMORPH（登録商標）取得ソフトウェア（Universal Imaging Corporation, Downingtown, PA）により実行され、透過明視野像、反射明視野像、または対象となるビーズを蛍光分子または蛍光分子類で標識した場合は落射蛍光像のいずれかを使用する。

【００７８】

段階2

各撮像サイクルの開始時に、ソフトウェアは焦点地図の最初のx、y、zハードウェア位置に戻った。自動焦点ルーチンをその特定の視野に関して遂行して、z軸での位置決定誤差を補正した。次に、画像を取得し、画像レジストレーションソフトウェアプログラムに渡した。そのプログラムは、新しい画像をそのフレームの本来の画像とレジストレーションする目的で、その新しい画像を変換するためにxおよびy軸での適切なオフセットを見出した（1画素以内まで；その画素のサイズでミクロン単位の距離、通常は約7μmを与える）。ビーズを固定化したため、ビーズは、アレイに追加の特性を導入する必要を避ける基準マーカー（fiduciary marker）として働くことができた。このように、無秩序化されたビーズを有することは、精密な画像アライメントを行おうとする場合に大きな利点であった。次に、顕微鏡のステージをこれらのオフセットだけxおよびy方向に移動した。次に、ビーズアレイは段階1の本来の位置に近いx、y、z位置であった。この方法を使用してサブミクロンの分解能で画像をアライメントできることが見出された。第二の自動焦点を実施して、新しいx、y位置に焦点が合っていることを確認した。この新しい位置に原点をリセットした。

【００７９】

段階3

本来の焦点地図で各位置を視察することにより画像を取得した。これらの画像は以前のサイクルすべてと一致しており、各フレームのビーズの大部分またはすべてに関してデータの抽出が可能であった。典型的な「焦点地図」のプロット座標の例を図4に示す。この図は、二つの大きな穴を有する円形ゲル（直径約1cm）を示す。赤色の点は、個々のフレームの相対XYZ座標を示す。図5は、透過明視野で高密度で撮像した単層ビーズのサンプル画像を示す。

【００８０】

実施例VI

対タグのインビトロライブラリーの構築

直鎖dsDNA分子のプールをもたらすインビトロライブラリー構築プロトコルを開発した。そのライブラリーでは、各分子は長さ約134～136bpであり、対象ゲノムから得られた17～18bpのタグの特有の対を含む。これらの特有のタグは、ユニバーサルプライマー（PR1-RおよびPR1-F）に相補的な配列のセットに隣接し、追加のユニバーサルスペーサー配列（「T30」）がこれらのタグを分離している（図11）。特有のタグは、対象ゲノムに同一に配向して、対象ゲノムでのタグの分離が制約された範囲（例えば1000+/-100塩基）に入る点で、それらのタグは「対」になっている。このプロトコルは、完全にインビトロである点で特有であり、大腸菌（E. coli）への形質転換は必要ない。

【００８１】

インビトロ法により構築されたこの形式のライブラリーは、従来のゲノムショットガン配列決定に比べて以下の利点を提供する：

（a）エマルジョンPCRプロトコルの方が短鎖配列を増幅するための効率がかなり高いこ

とから、ライブラリー中の各増幅可能な分子の合計長を最小化する動機づけが存在する；
（ｂ）サンガー（Sanger）に基づくゲノムプロジェクトの経験は、対形成した読み取りは、特にゲノムが繰り返しエレメントを含む場合に、ゲノムの再配列決定に大いに有用であることを指示している；
（ｃ）特有の塩基を配列決定するために本明細書に記載する方法は、ユニバーサル「アンカー」配列を介して位置決定することに依存している。
このライブラリー形式では、二つのタグに関して近位および遠位アンカーの両方が存在し、そのことは、各タグについて独立して配列決定法を適用することによって、読み取り長を二倍にすることが効率的に可能になる。

【００８２】

このライブラリー構築プロトコルは以下の段階を有した。
1. ゲノムDNAの精製
2. フラグメントを発生するためのゲノムDNAの切断
3. DNAフラグメントの末端修復およびA尾部形成
4. PAGEによる切断したフラグメントのサイズ選択
5. T尾部スペーサーオリゴヌクレオチド（「T30」）を用いた環化
6. ランダムヘキサマーを用いたローリングサークル型増幅（RCA）
7. MmeI（IIs型）を用いて消化して対タグを放出
8. タグ-T30-タグライブラリーのPAGE精製
9. タグ-T30-タグライブラリーの末端修復
10. FDV2（PR1-F）およびRDV2（PR1-R）プライマーオリゴヌクレオチドの連結
11. PAGEによる対タグライブラリーのサイズ選択
12. ニックトランスレーションを行ってdsDNAライブラリーのニックを除去
13. 対タグライブラリーのPCR増幅
14. 対タグライブラリーのPAGEサイズ選択
15. クローニングおよびサンガー配列決定によるライブラリーの妥当性確認

【００８３】

「Ｍ」および「Ｒ」大腸菌株に関する対タグライブラリーを構築するために上記段階を実施したときの、上記段階のそれぞれに関する詳細なプロトコルを下記に示す。

【００８４】

ゲノムDNAの精製

大腸菌株「Ｍ」および「Ｒ」それぞれに関して、LB 3mL中で培養物を一晩成長させ、製造業者のプロトコルの通りQiagen DNEASY（登録商標）組織キットを用いて単離した。各ゲノムDNAの精製物の収量は約30μｇであった（Qiagen Inc., Valencia, CA）。

【００８５】

広いサイズ分布のフラグメントを発生させるためのゲノムDNAの切断

両方の株からのゲノムDNAの切断をAgencourt Bioscience Corporation（Beverly, MA）により行った。下記のゲルにみられるように、結果として得られたDNAフラグメントのサイズ分布は極めて広範囲であった（図12）。

【００８６】

DNAフラグメントの末端修復およびA尾部形成

特記のない限り、Nanodrop ND-1000分光光度計でDNAの定量をすべて行った。

【００８７】

切断されたゲノム「Ｍ」DNAは57ng/μｌと定量され、切断されたゲノム「Ｒ」DNAは55ng/μｌと定量された。切断されたDNAフラグメントをEpiCentre END-IT（商標）DNA末端修復キット（Madison, WI）で末端修復した。各ライブラリーについて、以下の混合物を調製した：切断された大腸菌DNA（約9～10μｇ）170μｌ；10× END-IT（商標）緩衝液25μｌ；10× END-IT（商標）ATP 25μｌ；10× END-IT（商標）dNTP 25μｌ；およびEND-IT（商標）酵素5μｌ、全体積250μｌ。

【００８８】

反応物を室温で1時間インキュベートした。PCR産物の精製に関する製造業者の助言の通りQiagen QIAQUICK（登録商標）カラムでDNAを精製した。溶出に約90μlの緩衝液EB（10mM Tris・Cl、pH8.5）を使用した。「M」DNAを96.5ng/μlと定量し、「R」DNAを75.1ng/μlと定量した。各体積を約22μlの試験管4本に分割した。残余の酵素活性を除去するために、試験管を70℃に15分間加熱した。A尾部形成用マスター混合物を以下のように調製した。10× PCR緩衝液100μl（$MgCl_2$を有さない）（Invitrogen, Carlsbad, CA）、50mM $MgCl_2$ 60μl（Invitrogen）（終濃度3mM）、100mM dATP 5μl（Invitrogen）（終濃度0.5mM）、Taq（5U/uL）5μl（New England Biolabs, Beverly, MA）、およびdH_2O 610μl。

【００８９】

熱不活性化した後で、切断され末端修復されたDNAフラグメント22μlを含む各試験管にマスター混合物78μlを加えた。温度サイクリング機で試験管を70℃で30分間インキュベートした。サイクリングプログラムは、試験管を4℃に冷却することで終了した。次に、これらの試験管を温度サイクラーから直接氷中に移した。

【００９０】

DNAを以下のようにフェノール-クロロホルム抽出およびエタノール沈殿（P：C：P）により精製した。

1. 等量のフェノール：クロロホルム：イソアミルアルコール（25：24：1）を加えた。
2. 0.1容の3M NaOAc（pH5.2）を加えた。
3. グリコーゲン（20μg/μl）1.0μlを加えた。
4. 冷100％エタノール2.5容（-20℃で保存した瓶から）を加えた。
5. 上下逆にして混合した。
6. 試験管を-70℃に約30～60分間入れた。
7. 4℃の室内で微量遠心分離器で10分間最高速度で回転した。
8. 上清を除去した。
9. 80％エタノール1mlを加えた。
10. 室温で微量遠心分離器で5分間最高速度で回転した。
11. 上清を除去した。
12. Speed-Vacに試験管を約5分間入れた。
13. 緩衝液EBまたはTE 40μlにペレットを再懸濁した。

【００９１】

<u>PAGEによる切断されたフラグメントのサイズ選択</u>

各ライブラリーからの材料の半分をプレキャスト6％TBE-PAGEゲル（Invitrogen, Carlsbad, CA）にロードした。DNA 20μlを5× 高密度試料緩衝液5uL（Novex, San Diego, CA）と混合した。これと同じローディング緩衝液を、このプロトコルでのその後のPAGEゲルのすべてに使用した。試料/ローディング緩衝液混合物を1レーンあたり12.5μlロードした（ライブラリー一つあたりレーン2本）。ゲルを標準的な装置で泳動し、約1000塩基対フラグメントに対応する領域をUVへの最小の曝露で切り出した。ゲルフラグメントをカミソリでサイコロ状に切り、各ライブラリーフラグメントをPAGE溶出緩衝液600μl（10mM Tris-HCl (pH7.5)、50mM NaCl、1mM EDTA (pH8.0)）に移した。試験管を37℃で一晩インキュベートした。

【００９２】

翌日、溶出液を微量遠心分離器で最高速度で1分間回転し、上清を新しい試験管に移した。回収率を改善するために、追加のPAGE緩衝液200uLでゲルフラグメントを洗浄した。DNAをP：C：Pプロトコルを使用して精製し、各ペレットを緩衝液EB 22μlに再懸濁した。

【００９３】

A尾部形成段階を繰り返して、A尾部の潜在的部分分解の影響を最小にした。その分解は、ゲル精製中に起こった恐れがあった。末端修復およびA尾部形成の両方をPAGEに基づくサイズ選択の前よりもむしろ後に行ってもよい。A尾部形成用マスター混合物を以下のように調製した。10× PCR緩衝液（$MgCl_2$なし）25.00μl、50mM $MgCl_2$ 15.00μl（終濃度3mM）、100mM dATP 1.25μl（終濃度0.5mM）、Taq（5U/μl）1.25μl、およびdH_2O 152.50μ

l。

【００９４】

A末端尾部形成用マスター混合物78μｌを22μｌに再懸濁したDNAを含む各試験管に加え、総容積100μｌとした。各ライブラリーについて、これを2本の温度サイクラー適合性試験管にそれぞれ50μｌに分割した。試験管を温度サイクリング機に入れて70℃で30分間インキュベートした。試験管を4℃に冷却することにより、サイクリングプログラムは終了した。次に、これらの試験管を温度サイクラーから直接氷中に移した。DNAをＰ：Ｃ：Ｐプロトコルにより精製した。各ライブラリーを緩衝液EB 10μｌに再懸濁して、氷中に入れた。

【００９５】

回収率を定量して回収したフラグメントのサイズ範囲を推定するために、精製した材料の20％を使用してプレキャスト6％ TBE PAGEゲル（Invitrogen）を泳動した（図13）。ゲルに基づく定量では、「Ｍ」＝43ng（2μｌ中、または全量の20％）；範囲＝約850～1150、平均＝約1000。ゲルに基づく定量では、「Ｒ」＝18ng（2μｌ中、または全量の20％）；範囲＝約900～1250；平均＝約1075。各ライブラリーについて確保した残りの約8μｌの体積には「Ｍ」フラグメント約171ngおよび「Ｒ」フラグメント約73ngが存在した。

【００９６】

T尾部スペーサーオリゴヌクレオチド（「T30」）を用いた環化

次に、A尾部を有するライブラリーフラグメントを、T尾部スペーサーオリゴヌクレオチド「T30」を使用して環化した。二つの32bpオリゴヌクレオチドをアニールすることによりT30セグメントを調製して、1塩基の「Ｔ」突出を有する30bpのdsDNAフラグメントを発生させた。

5′ GTCGGAGGCCAAGGCGGCCGTACGTCCAACT 3′(SEQ ID NO:5)

3′ TCAGCCTCCGGTTCCGCCGGCATGCAGGTTG 5′(SEQ ID NO:6)

T30セグメントは外向きMmeI部位に隣接している。

【００９７】

混合して各オリゴについて終濃度50μＭにし、温度サイクラーで95℃に10分間加熱し、温度サイクラーを止め、その混合物を1時間かけて室温にゆっくりと冷却させることにより、オリゴのアニーリングを行った。

【００９８】

T30フラグメントとA尾部を有するライブラリーフラグメントとの連結を、QUICK LIGATION（商標）キット（New England Biolabs, Beverly, MA）を使用して実施した。反応物を以下のように調製した。「Ｍ」ライブラリー環化反応（合成体積80μｌ）：「Ｒ」試験管からのA末端を有するフラグメント8.0μｌ（1000bpで約171ng＝0.2599pmol）；dH_2O 27.2μｌ；T30 0.8μｌ（初発濃度1μＭ（0.8pmol、3倍モル過剰））；2× QUICK LIGATION（商標）緩衝液40.0μｌ；およびQUICK（商標）T4 DNAリガーゼ4.0μｌ。「Ｒ」ライブラリー環化反応（総容積30μｌ）：「Ｒ」試験管からのA末端を有するフラグメント8.0μｌ（1075bpで約73ng＝0.1032pmol）；dH_2O 5.2μｌ；T30 0.3μｌ（初発濃度1μＭ（0.3pmol、3倍モル過剰））；2× QUICK LIGATION（登録商標）緩衝液15.0μｌ；QUICK（商標）T4 DNAリガーゼ1.5μｌ。各試験管に酵素を加える前後に各反応物を十分に混合した。反応物を室温で10分間インキュベートしてから、氷中に移した。

【００９９】

リガーゼを温度サイクラーで65℃で10分間熱不活性化した。すべての非環化材料を分解するために、エキソヌクレアーゼ混合物を加えた。このエキソヌクレアーゼ混合物を以下のように調製する：エキソヌクレアーゼI（20U/μｌ）4.0μｌ（New England Biolabs, Beverly, MA）；エキソヌクレアーゼIII（100U/μｌ）0.4μｌ（New England Biolabs, Beverly, MA）；およびTE 35.6μｌ。エキソヌクレアーゼ混合物10μｌを「Ｍ」反応物80μｌに加え、エキソヌクレアーゼ混合物3.75μｌを「Ｒ」反応物30μｌに加えた。試験管を37℃の温度サイクラーで45分間、その後に80℃で20分間インキュベートして、エキソヌクレアーゼを

熱不活性化した。この材料を全く精製せずに、次の段階のRCA反応に直接使用した。

【０１００】

ランダムヘキサマーを用いたローリングサークル型増幅（RCA）

REPLIPHI（商標）phi29キット（EpiCentre, Madison, WI）を使用してライブラリー材料の量を増幅するために、超分岐RCAを行った。以下のようにマスター混合物を調製した。dNTP混合物32.0μl（各25μM）、10× REPLIPHI（商標）phi29反応緩衝液80.0μl、ランダムDNAヘキサマー40.0μl（1mM、5'-NNNN*N*N-3'として合成、ここで「*」はホスホロチオエート結合を示す）、dH_2O 552.0μl、および5× SybrGreen 16.0μl。そのマスター混合物を試験管2本に270μlずつ分割し、その試験管に「M」または「R」材料のいずれか30μlを混合して総容積300μlにした。次に、各試験管を50μlずつ6本に分割した。環化テンプレートを変性させるために、試験管を95℃に5分間加熱してから、4℃に速やかに冷却した。氷冷した各試験管にphi29酵素2.5μlを加えて、試験管1本あたり総容積52.5μlとした。試験管を十分に混合して、氷冷を続けた。温度サイクラーで試験管を30℃で一晩インキュベートした。

【０１０１】

RCA反応をリアルタイムPCR機で実施し、反応物に存在するSybrGreen色素を介して増幅を観察した。dsDNA含量は上昇し、2時間の時点で頭打ちになった。これは、一晩反応を実施する必要がないおそれがあることを示している。

【０１０２】

MICROCON-30（登録商標）カラム（Millipore, Billerica, MA）でTE合計1mLで洗浄してDNAを精製した。ペレットは十分量であった。MICROCON-30（登録商標）膜の数回の洗浄を使用して回収率を最大にした。50℃での加熱と追加の再懸濁緩衝液（緩衝液EB）の添加との組み合わせを使用してDNAを再懸濁した。各ライブラリーの約750μLを回収した。

【０１０３】

試料をNANODROP（登録商標）装置（NanoDrop Technologies, Wilmington, DE）で定量した：「M」＝230ng/uLおよび「R」＝204ng/uL。このように、RCA反応によって各ライブラリー約150μgがもたらされた。

【０１０４】

対タグを放出するためのMmeI（IIs型）を用いた消化

各ライブラリー約40μgをMmeIで消化した。MmeI部位はT30セグメントにあるその認識部位から離れたところを切断し、T30セグメントの末端のいずれかに外向きMmeI部位が存在することから、この消化は2bpの突出を有し約18bpのタグが隣接するT30セグメント（長さ約70bp）を放出すると予測した。MmeI消化前は、ゲノムフラグメントはT30を有して環化していたことから、これらのタグはそれらの起点に関して対形成していると予想した。

【０１０５】

反応物を以下のように調製した。MmeI、S-アデノシルメチオニン（SAM）、およびNEBuffer 4（10× ）をNew England Biolabs（Beverly, MA）から得た。32mM SAMを1× NEBuffer 4で1：20に希釈した（->1.6mM）。

【０１０６】

【表１】

	反応物「M」	反応物「R」
DNA	173.9	196.0
dH_2O	664.5	642.4
NE緩衝液4 (10x)	100.0	100.0
1.6 mM SAM	1.6	1.6
MmeI (2U/μl)	60.0	60.0
総容積 (μl)	1000.0	1000.0

【０１０７】

反応物を氷冷して調製し、酵素を加える前に試薬を十分に混合した。各反応物を125μｌずつ試験管8本に分割し、37℃で30分間温度サイクラーでインキュベートした。エタノール2.5容の代わりにわずか2容しか使用せず、最初に熱不活性化を行わない以外は上記のように、Ｐ：Ｃ：Ｐ精製を行った。各ライブラリーについて消化されたフラグメントをTE 80μｌに再懸濁した。約70bpのバンドは予想通り「Ｍ」および「Ｒ」レーンに観察された（図14）。

【０１０８】

PAGEによるタグ-T30-タグライブラリーの精製

ライブラリー一つあたり4本のレーンを使用して、各ライブラリー全量を上記のような10レーンのプレキャスト6% PAGEゲルを泳動した（ライブラリー20μｌおよび5× 色素5μｌ）。約70塩基対の明瞭なバンドを切り出した。各種類のライブラリーのすべてのレーンからのフラグメントを合わせ、溶出が約3時間であった以外は上記のようにゲルからの抽出を行った。Ｐ：Ｃ：Ｐ回収後に、各ライブラリーについてのDNAをTE約20μｌに再懸濁した。回収した材料を定量するために診断ゲルを泳動した（図15）。「Ｍ」および「Ｒ」ライブラリーの両方は約12.5ng/μｌと推定され、それぞれ18μｌがこの時点で残っていた。

【０１０９】

タグ-T30-タグライブラリーの末端修復

タグ-T30-タグ分子は、MmeI消化の結果、2bpの3'突出を含んだ。上記のようにEpiCentre END-IT（商標）DNA末端修復キット（Madison, WI）を使用して末端を修復した。反応物を以下のように調製した。「Ｍ」または「Ｒ」フラグメント8.50μｌ（12.5ng/μｌ、約100ng）、10× END-IT（商標）緩衝液1.25μｌ、10× END-IT（商標）ATP 1.25μｌ、10× END-IT（商標）dNTP 1.25μｌ、およびEND-IT（商標）酵素0.25μｌ、総容積12.5μｌ。

【０１１０】

反応物を室温で45分間インキュベートしてから、4℃に直接移した。TE 40μｌを加えることにより体積を50μｌに増やし、反応物を上記のようにＰ：Ｃ：Ｐ抽出した。沈殿段階を-70℃で一晩行った。回収したDNAをTE 8μｌに再懸濁した。

【０１１１】

FDV2（PR1-F）およびRDV2（PR1-R）プライマーオリゴヌクレオチドの連結

1：1に混合（終濃度50μ Ｍ）して、95℃で10分間加熱し、かつその反応物を1時間かけてゆっくりと冷却させることによって、完全に相補的なオリゴヌクレオチド（100μ Ｍ、HPLCで精製）をアニールすることにより、プライマー-アダプター（dsDNA、FDV2およびRDV2）を調製した。

【０１１２】

「アニールした」形式では、FDV2およびRDV2は以下の通りであった。

FDV2:

5′-AACCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT (SEQ ID NO:7)

3′-TTGGTGATGCGGAGGCGAAAGGAGAGATACCCGTCAGCCACTA (SEQ ID NO:8)

RDV2:

5′-AACTGCCCCGGGTTCCTCATTCTCT (SEQ ID NO:9)

3′-TTGACGGGGCCCAAGGAGTAAGAGA (SEQ ID NO:10)

【０１１３】

FDV2およびRDV2分子はリン酸化しておらず、よって、自己-自己連結も相互連結もできないと予想された。末端修復され連結した分子はリン酸化しており、よって、過剰のFDV2およびRDV2を使用してライブラリー分子についてのコンカタマー化イベントを最小にした。ライブラリー分子へのプライマー-アダプターの連結は平滑末端-平滑末端であり、よって、連結効率を向上するためにポリエチレングリコール（PEG）の存在下で連結を実行した。各反応を以下のように設定した。dH_2O 12.3μｌ、精製した平滑末端ライブラリーフラグメント（「Ｍ」または「Ｒ」約100ng、すなわち約2pmol）8.0μｌ、RDV2 1.0μｌ（50μＭ、50pmol）、FDV2 1.0μｌ（50μＭ、50pmol）、10× T4リガーゼ緩衝液2.5μｌ（New England Biolabs, Beverly, MA）、40% PEG 21.2μｌ（40%ポリエチレングリコール8000）、およびT4リガーゼ2.0μｌ（2000U/μＬ）（New England Biolabs）。PEGおよびリガーゼ以外のすべての試薬を混合することにより、反応物を室温で調製した。次にPEGを加えて混ぜ合わせ、リガーゼを加えて混ぜ合わせた。反応物を16℃で一晩インキュベートした。精製するために、反応体積をTEで100μｌに増やし、Ｐ：Ｃ：Ｐ精製した。ペレットを緩衝液EB 10μｌに再懸濁した。

【０１１４】

PAGEによる対タグライブラリーのサイズ選択

反応全体を10ウェルの6% PAGEゲルと適切なラダーとで実施した。未連結のRDV2およびFDV2（ライブラリーに対して大きくモル過剰で存在）がゲルの外まで泳動したと予想されるように、十分な距離だけゲルを泳動させた。予想通り適切なサイズの三重バンドを観察した（RDV2/ライブラリーフラグメント/RDV2連結、RDV2/ライブラリーフラグメント/FDV2連結、またはFDV2/ライブラリーフラグメント/FDV2連結に起因する）。完全な三重バンドを含む領域を切り出して、溶出が3時間であり、エタノール沈殿が-70℃で一晩であったことを除いて、上記のようにゲルから精製した。試料をそれぞれ緩衝液EB 20μｌに再懸濁した。

【０１１５】

dsDNAライブラリーでのニックを除去するためのニックトランスレーション

ライブラリー分子だけが連結反応で5'-リン酸化されたので、連結産物は修復しなければならないニックを含むと予想された。残りの材料の半分と取り組んで、以下のようにニックトランスレーションを実施した。ライブラリー10.0μｌ（回収率100%と仮定、これはタグ-T30-タグ分子50ngと連結したプライマー-アダプターの量との合計のはずである）、dNTP混合物0.5μｌ（25mM、よって各ヌクレオチドの終濃度は500μＭ）、10× NEBuffer-2 2.5μｌ（New England Biolabs）、大腸菌DNAポリメラーゼI（10U/μｌ）1.0μｌ（New England Biolabs）、およびdH_2O 11.0μｌ。反応物を調製し、氷冷しながら混合してから、16℃で30分間インキュベートした。精製するために、反応体積をTEで100μｌに増やし、反応物をＰ：Ｃ：Ｐ精製して、TE 10μｌに再懸濁した。

【０１１６】

対タグライブラリーのPCR増幅

PCRをこの場面で行って、1）我々が取り組まねばならないライブラリー材料の量を増やし、かつ2）ワンステップで外来連結産物を除去した。理論に拘束するつもりはないが、この段階で記載されたPCRから生じるはずである連結産物だけが、いずれかの側に正しく配向したRDV2およびFDV2が隣接したタグ-T30-タグを有する（T30セグメント自体は対称ではなく、よって、RDV2およびFDV2セグメントに対してどちらの配向でもどちらの方向でもありうることに注意）。

【０１１７】

複合混合物をPCR増幅しようとしていたことから、プライマー分子を使い果たす前にPCR反応を停止させることが重大であった。意図されたプライマーがいったん使い果たされたならば、ライブラリー分子が相互にプライマーとして働き始める事実、およびライブラリー分子が十分な類似性（約134個中約100個の同一塩基）を含むため、変性後に所定の一本鎖ライブラリー分子が厳密に相補的なパートナーと再アニールしそうにないという事実にこれは起因する。変性してプライマーが使い果たされた後で再アニールした、結果として生じたライブラリーは、多くの「ハイブリッド」ライブラリー分子を含みうる。

【０１１８】

リアルタイムPCR機でPCR増幅を行った（OPTICON（登録商標）2, MJ Research, Bio Rad, Waltham, MA）。

【０１１９】

【表２】

	50μlあたり	×20（総容積1000μl）
10x PCR緩衝液	5	100
25 mM (各) dNTPs	0.4	8
50 mM $MgCl_2$	1.5	30
プラチナ Taq	0.2 μl	4
水	42.7	853
RDV2-T (100 μM)	0.1	2
FDV2-T (100 μM)	0.1	2
SybrGreen (200x)	0.0025	0.5

【０１２０】

マスター混合物（表2に示す）を499.5μlずつ2本の試験管に分割し、ライブラリー材料（「M」または「R」）0.5μlをそれぞれに加えた。各ライブラリーのPCRをそれぞれ50μlずつ試験管8本に分割して（総容積400μl）、温度サイクラーにかけた。温度サイクリングを以下のように行った。

1. 94℃、2分間
2. 94℃、30秒間
3. 55℃、30秒間
4. 72℃、90秒間

段階2に戻る

【０１２１】

DNAの量がプラトーになり始めたことから、15サイクルが経過した後に反応を停止した。各ライブラリーからの反応物を個々の試験管に合わせ、PCR産物の精製のために製造業

者の助言の通りQIAQUICK（登録商標）（Qiagen, Valencia, CA）カラムで精製した。緩衝液EB 100μl中で再懸濁した。これは次の段階には体積が大きすぎると判断し、試料をエタノールで沈殿させ（フェノール抽出を行わなかった以外は、P：C：Pプロトコルの通り）、洗浄してTE 10μlに再懸濁した。

【０１２２】

PAGEによる対タグライブラリーのサイズ選択

6% PAGEゲルで反応を実施した。「最終」ライブラリーのバンド（約135塩基対での明瞭なバンド）を切り出し、溶出させ、前の通り精製した。これが最終精製段階であり、よってできるだけゲル精製を努めて厳密に行って、存在するおそれのある、いかなる非ライブラリー分子の混入も最小にすることが重大であった。PAGEゲルをラダーなしで泳動した。それは、これらラダーの分子は混入物にもなりうることが多いからであり、メスよりもカミソリの刃をギロチン型動作で使用した。

【０１２３】

一晩沈殿させてP：C：P精製した後で、回収したライブラリー材料をTE 10μlに再懸濁した。ライブラリーを定量するために、適切なマーカーを有する6%PAGEゲルを泳動した（図16）。ライブラリーおよびラダーのバンドの相対強度に基づき、濃度を2ng/μlと推定し、その結果、残余の各ライブラリーは約9× 2 = 18ngであった。そのライブラリーが135bpであったならば、濃度は約23nMとなる。エマルジョンPCRに使用するためにライブラリーをTEに様々な濃度に希釈して、本来のライブラリーとそれらの希釈物との両方を-20℃で保存した。

【０１２４】

クローニングおよびサンガー配列決定によるライブラリーの妥当性確認

ライブラリーのタグが大腸菌由来で対形成しているであろうという予想の妥当性を確認するために、「R」ライブラリーフラグメントをInvitrogen TOPO（商標）-4キット（Invitrogen, Carlsbad, CA）を用いてクローニングしてから、M13F/M13Rを使用するPCRおよびサンガー配列決定（クローン一つあたり1つの読み取り）を行った。96個のPCR産物を配列決定に供したが、これらのうち20個は不正確な読み取りであるか、またはベクターもしくは混入物に関係するかのいずれかとして戻ってきた。残りの76個の挿入物は、予想通りRDV2およびFDV2セグメントが適切に隣接していると思われた。これら76個のうち：1個は6bpの挿入（TTATCA）；1個は大腸菌ゲノムフラグメント（長さ65bp；BLASTで大腸菌MG1655ゲノムと63/63の100%の一致）；1個は大腸菌ゲノムフラグメント（長さ70bp；BLASTで大腸菌MG1655ゲノムと69/69の100%の一致）であり；1個はNCBIデータベースに有意な一致を有さない約27bpが隣接するRDV2プライマーを含み；72個は予想通りT30セグメントによって分離された二つのタグを含んだ。

【０１２５】

これら72個のうち、タグの長さは以下の分布を有した：1個のタグは9bp；1個のタグは11bp；2個のタグは13bp；1個のタグは14bp；1個のタグは15bp；2個のタグは16bp；73個のタグは17bp；62個のタグは18bp；かつ1個のタグは22bpであった。対形成に関して、タグは大腸菌のゲノムと以下のように一致した：4個は、大腸菌ゲノムと完全に一致するタグが1個または全く存在しない状態であった（おそらく配列決定の誤りまたは非標準（non-canonical）配列が原因）；1個は、両タグが特有の位置と一致したが、同じゲノム領域に由来しないと思われた点で「不対」であり；67個は対タグとして大腸菌ゲノムと一致した（予想された制約内のタグ間距離で同じ配向）。これらの67個の対タグについて、対タグの距離の分布は951+/-90bpであった。最小距離は729bpであり、最大距離は1162bpであった。

【０１２６】

このように、68個の読み取りについての対形成率は67/68、すなわち約98.5%であった。対形成した大腸菌タグを表し、T30セグメントがそれらのタグを分離しているライブラリーでのエマルジョンPCR増幅可能な分子の割合の最小推定値は67/76、すなわち88%であった。一方または両方のタグが一致不可能であった4個の読み取りが、サンガーの配列決

定の誤りまたは「R」株と標準ゲノム配列参照との間の差異が原因で一致不可能であった対形成した読み取りを実際に表すならば、実際の割合はやや高くなりうるであろう。

【０１２７】

標本サイズは小さいが（n＝72）、タグ配列で25/25/25/25頻度からの偏差が観察され、これは有意な傾向でありうる（表3に示す）。第一列の数は、あるプライマーもしくは他のプライマーとの、またはT30セグメントとのいずれかの接合部に関するタグ塩基の位置を表す。塩基の頻度を表にした基の鎖は、プライマーまたはT30セグメントからの伸長により（5'から3'に）配列決定するならば、これらの頻度の出現が予想されるような鎖である。カッコ内の数は、実際のカウント数（頻度とは対照的に）である。

【０１２８】

【表３】

プライマー/タグ接合部

	A	G	C	T
+1	0.315 (45)	0.315 (45)	0.154 (22)	0.217 (31)
+2	0.340 (49)	0.104 (15)	0.208 (30)	0.347 (50)
+3	0.292 (42)	0.208 (30)	0.146 (21)	0.354 (51)
+4	0.229 (33)	0.250 (36)	0.264 (38)	0.257 (37)
+5	0.333 (48)	0.194 (28)	0.243 (35)	0.229 (33)

T30/タグ接合部

	A	G	C	T
+1	0.299 (43)	0.194 (28)	0.326 (47)	0.181 (26)
+2	0.299 (43)	0.319 (46)	0.146 (21)	0.236 (34)
+3	0.312 (45)	0.222 (32)	0.222 (32)	0.243 (35)
+4	0.188 (27)	0.236 (34)	0.264 (38)	0.312 (45)
+5	0.252 (36)	0.196 (28)	0.273 (39)	0.280 (40)

【０１２９】

実施例VII

ビーズに連結したオリゴヌクレオチドに関する並列的な配列決定

ビオチンを介してストレプトアビジン被覆超常磁性ビーズ（直径8.8ミクロン）に連結したオリゴヌクレオチドに5から8塩基対の読み取りが得られた実験の結果を図6に示す。ビーズをアクリルアミドに固定化した。図6に示したビーズは、以前に公知の方法を使用した場合よりも、当業者が単一スライド上で単位面積あたり10,000倍多い特性を配列決定するのを可能にする。

【０１３０】

さらなる小型化に向けて、1ミクロンのビーズ上に短い配列決定用読み取り（テンプレート1個あたり4塩基対）を発生させるための実験を行った（図7）。それぞれのビーズがおよそ1から4個のアクティブ画素によってのみ表されることから、これらの結果は「1画素あたり1配列読み取り」の分解能に近づいている。この実証されたビーズのサイズおよび密度で、配列決定用読み取りを3インチのスライドグラスで1インチあたり3千万個を超

える独立したビーズから得られうる。
【０１３１】
実施例VIII
ミクロスフェアの単層
実施例IIIに示したプロトコルの変形は、自己集合した単層での親和性を利用して、増幅されたビーズの濃縮を提供する。これは、増幅のための現行のプロトコルに関連する問題を回避する。そのようなプロトコルでは、希釈標的DNAを使用してビーズ1個あたりのアンプリコンの重複を避け、よって標的DNAがゼロのビーズの数は、ポアソン分布に近似される。
【０１３２】
本発明は、スライド上の核酸のパターン（例えば組織切片、マイクロアレイ、伸長した染色体中のRNA）の上に「その場で」密封チャンバーを形成するために（増幅の拡散を制限するための）類似のエマルジョンを使用することも含む。固定化エマルジョンの使用は、エマルジョン液滴のサイズを制約し、安定化し、かつそれをさらに一様な分布にすることも助けるであろう。
【０１３３】
3インチのスライドグラスの1インチあたり数億から数十億個の独立した配列決定用読み取りを得るために、同じ1ミクロンのビーズを高密度で充填できる。超常磁性ビーズの自己組織化単層（SOM）を発生した（図8）。横磁場を使用して多層も構築できる。1ミクロンのビーズの単層で完全に被覆したスライド1枚から30塩基対の読み取りを得ることは、約560億の塩基の配列をもたらすであろう。
【０１３４】
実施例IX
エマルジョン非存在下での粒子上の核酸増幅
理論に拘束するつもりはないが、ミクロンまたはサブミクロンサイズにポロニーの形状を減少させると、スループットが少なくとも1000倍改善するであろう。これは、ポリマービーズ（1μ m常磁性ビースなど、Dynal, Oslo, Norwayから入手可能）、ポリマーネットワーク、またはポリマーナノ構造に一つの増幅プライマーを制約することにより実現しうる。このように、ポロニーの形状のサイズは、ビーズまたはポリマーマトリックスのサイズによって決定され、よって、ポリマーマトリックス（アクリルアミドゲル）の孔サイズおよび/または架橋含量ならびに標的アンプリコンの長さとは無関係である。従来のポロニーPCRに類似して、ポロニービーズについての実際の増幅段階は、ポリマーマトリックス（例えばアクリルアミドゲル）中で実施され、それは個々のビーズの位置を保持することを助け、アンプリコンの拡散を部分的に制約する。本発明の重要な一構成要素は、添加剤（例えば陽イオン性脂質、ポリアミン、ポリカチオンなど）の包含であり、それはポリマービーズを囲むゲル内間隙、例えばミセルまたは凝集体を形成し、固相での効率的な増幅を可能にする（図9A～9D）。この改善の主な利点は、ビーズ付近への標的アンプリコンおよび他の増幅プライマーの分画が異なることであり、それは増幅効率および/またはビーズとの近接を増加させる。それは、エマルジョンを形成する必要性を回避し、よって、より単純なマルチプレックス増幅と、増幅すべき初発セットの分子の2Dまたは3D配置の保存さえも可能にする。
【０１３５】
代替として、第二増幅プライマーをポリマーマトリックスに固定化できる（例えばアクリダイト修飾または他の架橋法により）。このように、遊離の第二プライマーは増幅にほとんどまたは全く必要ない。これは、増幅の間にビーズ同士の交差を排除し、よって、クローン性を強化するであろう。ポリマーマトリックス（例えばアクリルアミドゲル）に追加して、ミセルまたは凝集体の形成は、たとえあったとしても（すなわち、残余の遊離第二増幅プライマーがあるか、または遊離の第二プライマーが増幅中に使用されるにしても）遊離の増幅産物の拡散をさらに制約しうる。ポリカチオン層（または網目構造）内のポリアニオン（例えばDNA）の移動性は、そのポリカチオンからの自由拡散よりも迅速であ

り得、よって、その反応はカチオン層が満たされるまでは指数的に近く、次にほとんど停止して、その後の画像の鮮鋭度がさらに増強しうる。

【０１３６】

図9A～9Dは、ビオチン-ストレプトアビジン相互作用により1μｍDynaビーズに予備連結した一増幅プライマーのコピーを示す。他の増幅プライマーをアクリダイト修飾して重合の際にポリアクリルアミドゲルマトリックスに取り込めるようにする（左）。対象となる核酸増幅テンプレートと共に二つのプライマーをアクリルアミドマトリックスに集合させる。PCR増幅を行ってから、Cy3標識プローブの変性およびハイブリダイゼーションを行った。増幅産物をほとんど検出できなかった（右）。（左）とは対照的に、集合した混合物に陽イオン性脂質を加えて15分間インキュベートすると、より強いシグナルが実証したように、増幅効率は大きく増強した。

【０１３７】

実施例X

ポロニー配列決定を使用したDNA修飾の分析

CpGジヌクレオチドのシトシン残基の5'炭素にメチル（CH_3）基を付着すること（遺伝子外修飾と称する）は、細胞性遺伝子発現を制御するための重要なメカニズムを構成する。ゲノミクスメチル化パターンの動力学に関する定量的知識は、進化制御および病理状態を含む生物学に大きな影響を有するであろう。全ゲノムからの異常DNAメチル化の検出は、メチル化特異的酵素を用いた制限消化の後のディファレンシャルディスプレイ（MCA-RDA; Ueki et al. (2001) Cancer Res. 61:8540、全体がすべての趣旨で参照として本明細書に組み入れられる）またはマイクロアレイ分析（DMH (Yan et al. (2000) Clin. Cancer Res. 6:1432）、ECIST（Shi et al. (2002) Cancer Res. 62:3214、全体がすべての趣旨で参照として本明細書に組み入れられる）のいずれかに依存している。メチル化に差のある候補遺伝子座を同定できるが、それらのどれも任意の所定のCpGアイランドに典型的である組み合わせメチル化イベントを監視できない。多数のメチル化の関係をシスで（in cis）問い合わせることができる唯一の技法は、メチル化特異的配列決定である。しかし、亜硫酸水素塩で変換した個々のゲノムフラグメントのクローニングおよび従来の配列決定手順の性能は、その取り組みを労働集約的および低スループットとしている。

【０１３８】

一態様では、本発明は、メチル化特異的配列決定をポロニープラットフォームに適用して、従来のメチル化配列決定法に比べていくつかの利点を与えることを目的とする。（１）亜硫酸水素塩で変換したDNAフラグメントを固相または半固相（この場合、アクリルアミド）中で直接増幅でき、よって、骨の折れるクローニング段階の必要性およびクローニング効率の差によって導入される潜在的バイアスを撤廃できる。（２）数千から数百万の単分子に同時に問い合わせするようにメチル-dCの検出を高度に並列的に実現できる。二つの取り組み、すなわち繰り返しプロービングおよびFISSEQ（Zhu et al. (2003) Science 301:836; Mitra et al. (2003) Analyt. Biochen1. 320:55、全体がすべての趣旨で参照として本明細書に組み入れられる）を適用してメチル化イベントの差を監視できる。繰り返しプロービングは、一握りの公知の遺伝子座を研究するのに適しているが、FISSEQはゲノム全体の研究に確実に利点を有する（下記（４）参照）。（３）ポロニープラットフォームは単分子系技法であることから、投入材料はほとんど必要ない。よって、単一の細胞および/または染色体を利用できる。個々の細胞および/または染色体（例えば染色体の伸展）をポリマーマトリックスまたはゲル（例えばポリアクリルアミド）に包埋してから、亜硫酸水素塩による固相変換および増幅を行うことができる。これは、DNA精製段階を排除して、試料調製する間に、試料の損失および汚染の可能性を回避するであろう。（４）（３）の延長として、マルチプレックスPCRおよび/またはゲノム全体の増幅（例えばPhi29を用いた）を、亜硫酸水素塩で変換した細胞および/または染色体を用いて行える。これによって、ゲノムDNA全体のメチル化状態を単一細胞および/または単一染色体レベルで決定することが可能になる。

【０１３９】

実施例XI
蛍光共鳴エネルギー移動（FRET）
シングルFISSEQステップのためのFRETのドナー-アクセプター対、例えばCy3-dATPドナーおよびCy5-dATPアクセプターを形成する二つの異なるフルオロホア標識dNTPを使用することによって、一続きのホモポリマー中の0、1および2個の同一の塩基を識別できる。一つはドナーdNTPの励起波長で最初に励起し、ドナーの発光波長およびアクセプターの波長の両方で発光を捜す。次に一つは、アクセプターの励起波長で励起し、アクセプターの発光波長で発光を捜す。一つは、順序を逆にして最初にアクセプターを励起してFRETの結果としてアクセプターフルオロホアの光退色を最小にしうる。
【０１４０】
例えば、所与のサイクルで導入される0個のdA残基が存在するならば、無（または低）シグナルが観察されるであろう。1個のdAは、FRETが生じる結果として励起された各フルオロホアに対応する発光波長の検出をもたらすであろう。2個のdAはFRETが生じる結果として、ドナーの励起波長で励起したときにアクセプターの波長での発光をもたらすであろう。3個以上のdAは、2個のdAの場合の量的な変形として決定されうる。
【０１４１】
実施例XII
ライブラリー
<u>ローリングサークル/増幅用のライブラリー</u>
ゲノムDNAから直接、環の大きなライブラリーを発生する一方法は、
NNNNNN...（共通プライマー1）..（切断部位）..（共通プライマー2）...NNNNNN
のように末端にランダムなN残基重合体を使用して、マイクロアレイハイブリダイゼーション用に内部タグを使用しないことであろう。これらのN残基重合体を使用してDNAにランダムにハイブリダイズできる。当技術分野での系とは異なり、このN残基重合体は1塩基だけ離れているよりも、むしろ相互にある距離を置いて位置できるであろう。この距離は、利用するオリゴヌクレオチドの長さに制約されるであろう。
【０１４２】
ポリメラーゼ、リガーゼおよびdNTPを使用してギャップを埋めて連結して、「南京錠をかけた（padlocked）」プローブを発生させることができる。再環化により南京錠を開けるよりもむしろ制限酵素（オリゴヌクレオチドに存在しない部位で）およびエキソヌクレアーゼでゲノムDNAを消化して、ゲノム塩基の伸長を含む環化プローブのみを残すことができる。それらの環をローリングサークル法に使用できる。または、エマルジョン法を使用するならば、環を開裂させることによりその南京錠を開けて、共通プライマーに隣接したゲノム配列を残すことができる。
【０１４３】
<u>ジャンピングライブラリー</u>
本実施例は、ポロニーの準備が整ったライブラリーをphi29ローリングサークル型アンプリマーにより発生できる一方法を示す。MmeI（またはEcoP15I）を使用してジャンピングライブラリーを発生させるために、以下の段階を実施できよう：
（1）エンドヌクレアーゼ（例えば、NlaIII（CATG^）またはCvijI**（NG^CN）またはDNアーゼさえ用いたゲノムの完全または部分開裂；4bpの突出は開始に最もすっきりした場所でありうる）；
（2）任意のサイズ選択（サイズ選択が精密なほど複合繰り返し領域での配列の集合が精密である）；
（3）環が形成するように二重のタグと二重のMmeIアダプターとの連結；
（4）任意のサイズ選択；
（5）MmeIで切断、平滑末端にして再度環化；
（6）ランダムなニック形成、配列特異的ニック形成酵素、DNA、PNA、RNAプライマーを用いた鎖への侵入、またはRNAポリメラーゼもしくはプリマーゼ開始により、環にローリングサークルを開始できる。

【０１４４】
この方法は、各末端から27塩基対までの配列情報と、さらに最初のエンドヌクレアーゼ部位に関する情報（NlaIIIについて合計62塩基対）とを与え、さらに重要にはより長いデノボ配列にわたる集合を可能にする。
【０１４５】
非ローリングサークルライブラリーに使用するための特定の制限酵素および本発明の非ローリングサークルライブラリーの非限定的な例は、以下の通りである：SAGE: BsmFI、New England Biolabs (NEB)より入手可能、Beverly, MA (Velculescu et al. (1995) Science 5235:484、全体がすべての趣旨で参照として本明細書に組み入れられる）；LongSage: MmeI、NEBより入手可能（Saha et al. (2002) Nat Biotechnol. 20:508、全体がすべての趣旨で参照として本明細書に組み入れられる）； CAGE: MmeI、NEBより入手可能（Shiraki et al. (2003) Proc. Natl. Acad.Sci. U.S.A. 100:15776、全体がすべての趣旨で参照として本明細書に組み入れられる）； SuperSage: EcoP15I (Matsumura et al. (2003) Proc. Natl. Acad. Sci. U.S.A. 100:15718、全体がすべての趣旨で参照として本明細書に組み入れられる）； SAGS: BsaXI、NEBより入手可能（Torstein and Meyerson, Serial Analysis of Genome Subsets）。BsmFIおよびEcoP15Iは5'突出を有するが、MmeIは3'突出を有する。MmeIについて、3'の2塩基対の情報は相補的な対を連結することにより保存される）；および2塩基認識エンドヌクレアーゼCviJIを利用した迅速ショットガンクローニング（Fitzgerald et al. (1992) Nucleic Acids Res. 20:3753、全体がすべての趣旨で参照として本明細書に組み入れられる）。
【０１４６】
実施例XIII
排除体積の拡大
本実施例は、PCRの必要なしに、増幅されたビーズ：反応体積の高い比でクローン増幅されたビーズを発生できる手段を提供する。別のポリマーと相互作用しているあるポリマー（または粒子）は、近くの追加の粒子とのさらなる相互作用を防止できる。ローリングサークル型増幅によりテンプレートを予備増幅して、長鎖のかさ高いコンカテマーを得る。このように、過剰のそのようなテンプレート（図10）を用いて、ビーズ1個あたり多数のコンカテマーが結合するイベントなしに試験管内のすべてのビーズを飽和させることができる。
【０１４７】
例えば、30塩基対のユニバーサルタグとゲノムライブラリーからの30塩基対の挿入配列とをそれぞれ有する60塩基長のニックの入った環のライブラリーを用いる。本来の環から下がる1'タンデム一本鎖リピートに近くなるまでphi29ローリングサークル（ニック以外に全くプライマーを用いずに）を約1時間使用する。低塩でこれは約6'bpの長さ（＝200ミクロンの伸長）および数立方ミクロンの排除体積を有するであろう。テンプレートの希釈溶液が、相補的ユニバーサルタグを保有する1ミクロンのビーズと相互作用するならば、ビーズと接触する第一テンプレートは多数の点で結合し、結合から他のテンプレートを排除するであろう（図10に示す）。ビーズを低塩で洗浄して（他のテンプレートとの凝集体を減らし）、次にローリングサークル型増幅を（ビーズと結合したプライマーだけを用いて）再び開始する。理論に拘束するつもりはないが、これは配列決定用の十分量の一本鎖テンプレートを産生するはずである（各プライマーは多くのタンデム一本鎖繰り返しと結合しているであろうことから、PCRで得られるであろう量よりもいっそう多くがこれによって産生されうる）。この方法は、温度サイクリングまたはビーズソーティングの必要性を回避するはずである。
【０１４８】
実施例XIV
参照
各参照は、全体がすべての趣旨で参照として本明細書に組み入れられる。

Aach and Church (2004) *J. Theor. Biol.* 228:31

Merritt et al. (2003) *Nucleic Acids Res.* 31:e84

Mikkilineni et al. (2003) *Biotechnol. Bioeng.* 86:117

Mitra et al. (2003) *Anal. Biochem.* 320(1):55

Zhu et al. (2003) *Science* 301(5634):836

Butz et al. (2003) *BMC Biotechnol.* 31:11

Denkov et al. (1992) *Langmuir* 8:3183

Dressman et al. (2003) *Proc. Natl. Acad. Sci. USA* 100:8817

Brenner et al. (2000) *Nat. Biotechnol.* 18:630

Mitra et al. (2003) *Proc. Natl. Acad. Sci. USA* 100:5926

Mitra et al. (2003) *Anal. Biochem.* 320:55

Zhu et al. (2003) *Science* 301:836

【図面の簡単な説明】

【０１４９】

本発明の上記および他の特性ならびに利点は、添付の図面と共に解釈される例示的な態様の以下の詳細な説明からさらに十分に理解されるであろう。

【０１５０】

【図１】ビオチン化プライマー（SEQ ID NO：１）およびストレプトアビジン被覆ビーズの略図を示す。

【図２】本明細書に記載した濃縮プロトコルを経たビーズを示す。（Ａ）は、未濃縮のビーズを示す。増幅配列を含む（「増幅された」）ビーズは赤色で、増幅された配列を有さない（「空席の」）ビーズは緑色である。（Ｂ）は濃縮された増幅されたビーズを示す。（Ｃ）は空席のビーズを大きな割合で示すペレット画分からのビーズを示す。

【図３】本明細書に記載した単層化プロトコルの略図を示す。

【図４】焦点地図を示す。

【図５】高密度で透過明視野で単層ビーズの画像を示す。

【図６】8.8ミクロンのビーズに連結したオリゴヌクレオチドに関する並列的な配列決定を示す。5つの80塩基長オリゴヌクレオチドの一つをそれぞれ載せたビーズの集団をアクリルアミド中に固定化して、5~8塩基対の読み取りが得られるまで多ラウンドの蛍光インサイチュー配列決定（FISSEQ）に供した。この加工後の画像は、スライドの領域を示し、疑似色は個々の配列を表す（暗青色は「ノイズシグニチャ」を表す）。600× 600画素領域を示し、ここで分解能は各寸法1画素あたり約0.5ミクロンである。倒立落射蛍光顕微鏡で画像を取得した。

【図７】1ミクロンビーズに連結したオリゴヌクレオチドに関する並列的な配列決定を示す。それぞれが三つの80塩基長オリゴヌクレオチドのうち一つを載せた1ミクロンビーズ

の集団をアクリルアミド中に固定化して、4塩基対の読み取りが得られるまで多ラウンドのFISSEQに供した。より大きいビーズ（2.8ミクロン）を混ぜ込んで、画像レジストレーション用の基準マーカーとして作用させた。正しい配列シグニチャは疑似色の赤色、白色、黄色であり、ノイズシグニチャは疑似色の暗青色であり、基準マーカーは疑似色の緑色である。200× 200画素の領域を示し、ここで、分解能は各寸法で1画素あたりおよそ0.5ミクロンである。倒立落射蛍光顕微鏡で画像を捕捉した。

【図８】対流自己集合により調製した常磁性ポリスチレンミクロスフェア（直径1ミクロン）の単層を示す図である。加える界面活性剤の量を変動することにより、その単層を粗充填（Ａ）またはさらに密に充填（Ｂ）できる。

【図９】図9A～図9Dは、エマルジョン非存在下の粒子上での核酸の増幅を示す。

【図１０】排除体積の取り組みの略図を示す。

【図１１】ライブラリー構築の最終産物の略図を示す。

【図１２】切断後の切断フラグメントのサイズ選択のためのポリアクリルアミドゲル電気泳動（PAGE）ゲルを示す。

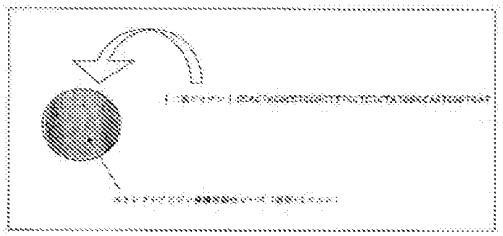
【図１３】サイズ選択したフラグメントを定量するためのPAGEゲルを示す。

【図１４】MmeIで消化したローリングサークル型増幅（RCA）材料の診断6% PAGEゲルを示す。

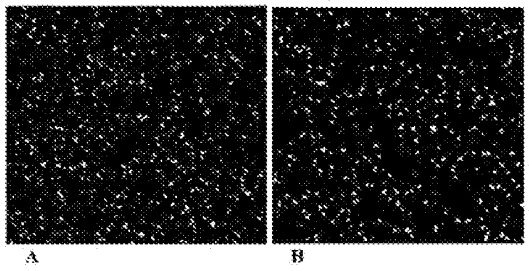
【図１５】ゲル精製したプライマーを有さないライブラリーの診断6% PAGEゲルを示す。

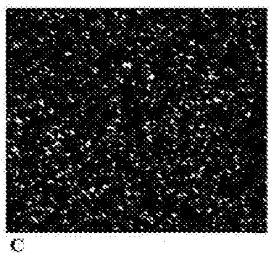
【図１６】最終ライブラリーの診断6% PAGEゲルを示す。

【図１】

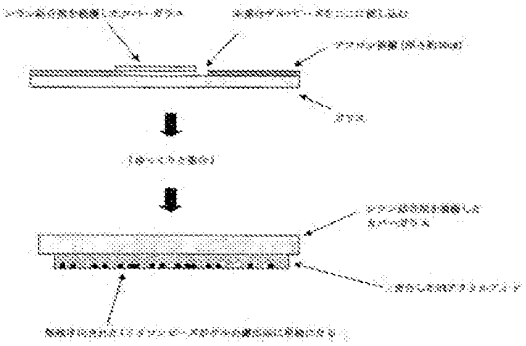


【図２】

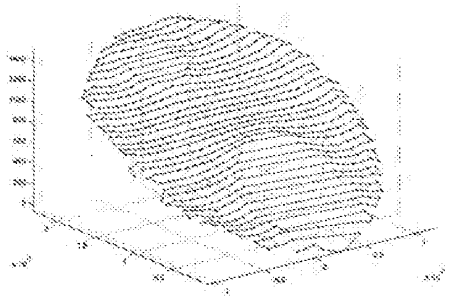




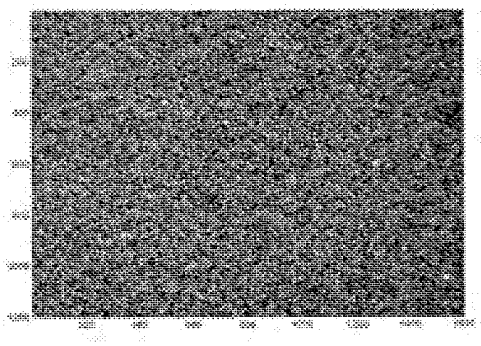
【図３】



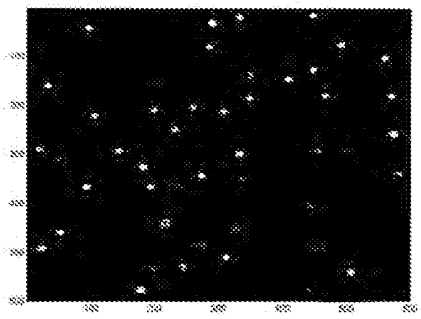
【図４】



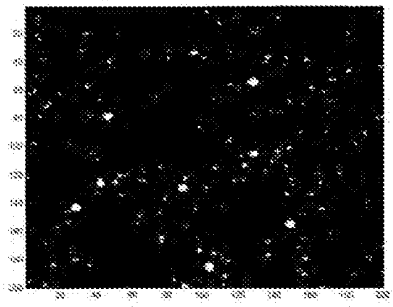
【図５】



【図６】

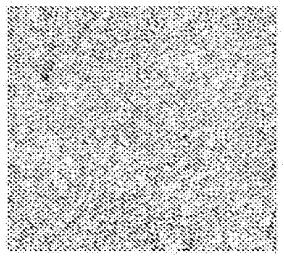


【図７】

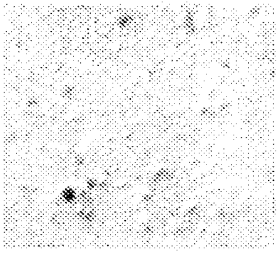


【図８】

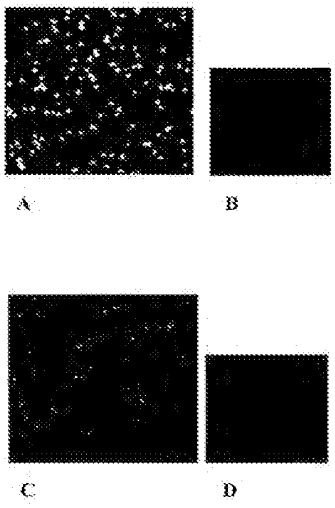
A



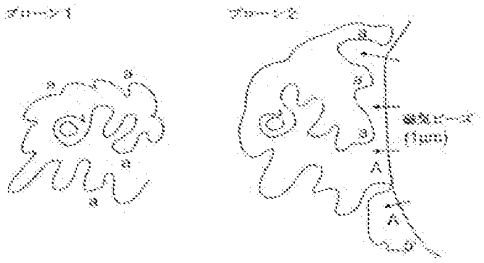
B



【図９】



【図１０】



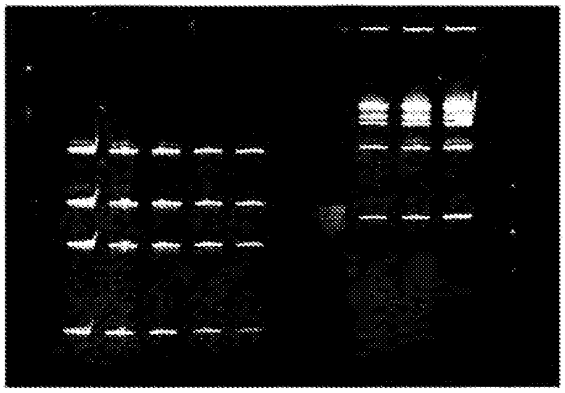
【図１１】



【図１２】



【図１３】



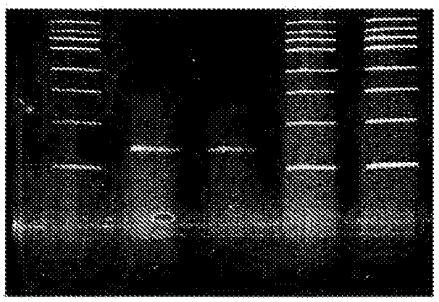
【図１４】



【図１５】



【図１６】



【配列表】
2007526772000001.app

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US05/06425

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) : C12Q 1/68; C12M 1/34; C07H 21/04; C12N 15/11
US Cl. : 435/6, 287.2; 536/23.1, 24.3
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/6, 287.2; 536/23.1, 24.3

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EAST, STN, bead#, array?, microarray?, solid, semisolid, dna#, rna#, nucleic acid#, oligonucleotide#

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DRESSMAN, D. et al, Transforming sing DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations, Proceedings of the Nations Academy of Sciences, USA, 22 July 2003, Vol. 100, No. 15, pages 8817-8822, see especially page 8819, Figure 1 and legend.	32-34
A	US 6,156,962 A (SHI et al) 24 October 2000 (24.10.2000), column 10, lines 31-44.	1-47
A	US 2003/0108867 A1 (CHEE et al) 12 June 2003 (12.06.2003), paragraph 0131.	1-47

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
27 September 2005 (27.09.2005)	27 OCT 2005
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (571) 273-3201	Authorized officer James Martinell, Valerie Bell-Harris Telephone No. (571)272-1600

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

フロントページの続き

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

（特許庁注：以下のものは登録商標）
１．テフロン

(72)発明者 シェンドュア ジェイ
アメリカ合衆国 オハイオ州 シャグリン フォールズ ディープビュー ドライブ １７６４０
(72)発明者 ポーレカ グレゴリー ジェー．
アメリカ合衆国 マサチューセッツ州 ケンブリッジ アンドリュー ストリート ３０ アパートメント １
(72)発明者 ジュー ジュン
アメリカ合衆国 ノースカロライナ州 ダーハム ダグラス ストリート ６１０ アパートメント ３０４エー

Ｆターム(参考) 4B024 AA11 AA20 CA04 CA09 HA14
4B029 AA23 BB20 CC03 CC13

Espacenet

Bibliographic data: JP2007526772 (A) — 2007-09-20

POLONY FLUORESCENT IN SITU SEQUENCING BEADS

Inventor(s):

Applicant(s):

Classification: **- international:** ***C07H21/04; C12M1/00; C12M1/34; C12N15/09; C12N15/10; C12N15/11; C12Q1/68; G01N33/53; G01N37/00***

- cooperative: **C12N15/1093 (EP, US); C12Q1/6834 (EP, US); C12Q1/6874 (EP, US)** more

Application number: JP20070500806 20050228

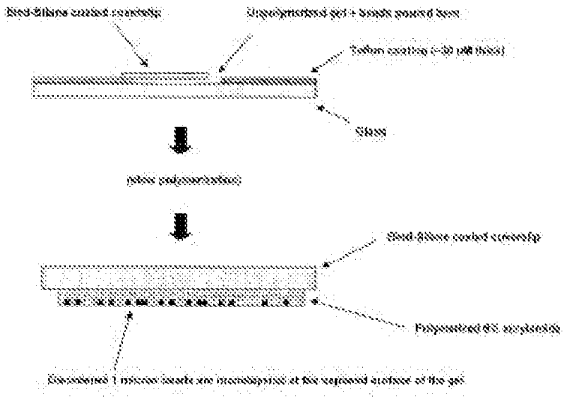
Global Dossier

Priority number(s): US20040548631P 20040227 ; WO2005US06425 20050228

Also published as: AU2005216549 (A1) CA2557841 (A1) CN1950519 (A) EP1718769 (A2) EP1718769 (A4) more

Abstract not available for JP2007526772 (A)
Abstract of corresponding document: WO2005082098 (A2)

Miniaturized, high-density, bead-based arrays are provided. Methods of producing and using clonal beads and producing and using miniaturized, high density, bead-based arrays are also provided.



Espacenet

Description: JP2007526772 (A) — 2007-09-20

POLONY FLUORESCENT IN SITU SEQUENCING BEADS

Description not available for JP2007526772 (A)
Description of corresponding document: WO2005082098 (A2)

A high quality text as facsimile in your desired language may be available amongst the following family members:

CA2557841 (A1) CN1950519 (A) US2007087362 (A1) WO2005082098 (A2)

The EPO does not accept any responsibility for the accuracy of data and information originating from other authorities than the EPO; in particular, the EPO does not guarantee that they are complete, up-to-date or fit for specific purposes.

POLONY FLUORESCENT IN SITU SEQUENCING BEADS

RELATED APPLICATIONS

[001] This application claims priority to U.S. Provisional Patent Application No. 60/548,631 filed on February 27, 2004, hereby incorporated by reference in its entirety for all purposes.

STATEMENT OF GOVERNMENT INTERESTS

[002] This invention was made with Government support under grant number F30602-01-2- 0586 awarded by DARPA and grant number DE-FG02-02ER63445 awarded by the Department of Energy. The Government has certain rights in the invention.

FIELD OF THE INVENTION

[003] The present invention relates to miniaturized, high density, bead-based arrays, methods of producing and using miniaturized, high density, bead-based arrays and methods of producing and using the components thereof.

BACKGROUND OF THE INVENTION

[004] Polymerase colony (polony) technology is a single-molecule amplification technology which allows the sequence of each individual molecule to be elucidated in a highly parallel manner. However, the throughput of polony technology is inversely proportional to the size of individual colonies, which ranges from tens to thousands of microns.

[005] Methods for generating populations of clonal microspheres (i.e., beads bearing clonally amplified DNA) are known in the art (e.g., Dressman (2003) Proc. Natl. Acad. Sci. USA 100:8817; Brenner et al. (2000) Nat. Biotech. 18:630). However, these methods suffer from several drawbacks. In Dressman et al., beads are analyzed via fluorescence activated cell sorting (FACS), which is expensive to operate and suffers too low of a throughput (i.e., less than 70,000 events per second) to process hundreds of millions of beads. In Brenner et al., beads are manipulated to form a packed, planar array, such that the physical packing limits scattering of the beads. SUMMARY

[006] The present invention is based in part on the discovery of a novel method for the efficient, cost-effective production of bead-based arrays. Such arrays are particularly useful for genetics research and diagnostic applications. The methods and compositions described herein allow for tens of millions to billions of discrete nucleic acid sequences to be queried in a reasonable time in an economical manner.

[007] Embodiments of the present invention are directed to bead-based arrays and methods of making bead-based arrays. In accordance with certain embodiments, arrays having a plurality of beads wherein an individual bead has a population of substantially identical nucleic acid sequences attached to them and wherein the population of substantially identical nucleic acid sequences differs in sequence from the population of substantially identical nucleic acid sequences attached to other beads are provided. The plurality of beads is immobilized in a semi-solid medium to form an array. The semi-solid medium can be made from poryacrylamide, cellulose, polyamide, cross- linked agarose, cross-linked dextran or cross-linked polyethylene glycol. In certain aspects, the semi-solid

medium has x, y and z axes, and the plurality of beads is randomly arranged relative to the x and y axes. The beads can be immobilized as a monolayer, for example, near the top surface of the semi-solid medium.

[008] In certain aspects, the semi-solid medium can be attached to a solid support such as a microscope slide or a flow cell. The solid support can be attached to the bottom surface of the semi-solid medium.

[009] In other aspects, two, three or four different populations of substantially identical nucleic acid sequences can be attached to the beads. In other aspects, the beads are clonal beads. In still other aspects, the beads include a library.

[010] In other embodiments, methods of making bead-based arrays including providing a plurality of beads having a population of substantially identical nucleic acid sequences attached, immobilizing the beads in a semi-solid medium to form an array, and amplifying the population of substantially identical nucleic acid sequences to form a plurality of beads having an amplified population of substantially identical nucleic acid sequences attached are provided. In certain aspects, the semi-solid medium includes an amplification primer. In other aspects, the semi-solid medium includes an additive that forms voids in the semi-solid medium, such as a cationic lipid, polyamine or polycation.

[Oil] In other embodiments, methods of making bead-based arrays including providing a plurality of beads having a population of substantially identical nucleic acid sequences attached, and amplifying the population of substantially identical nucleic acid sequences to form a plurality of immobilized beads having an amplified population of substantially identical nucleic acid sequences attached are provided. The beads are then immobilized in a semi-solid medium to form an array. In certain aspects, amplifying is performed by emulsion PCR.

[012] In still other embodiments, methods of making bead-based arrays including providing a plurality of beads having a population of substantially identical nucleic acid sequences attached, and amplifying the population of substantially identical nucleic acid sequences to form a plurality of beads having an

amplified population of substantially identical nucleic acid sequences attached are provided. The plurality of beads having an amplified population of substantially identical nucleic acid sequences attached thereto is enriched to form an enriched population of beads, and the beads are immobilized in a semi-solid medium to form an array.

[013] Embodiments of the present invention are directed to methods for enriching a population of beads having a first nucleic acid sequence attached. These methods include providing a population of beads wherein at least a portion of the population includes a bead having a first nucleic acid sequence attached. The population of beads is contacted with a second nucleic acid sequence that is complementary to the first nucleic acid sequence, and the population of beads and the second nucleic acid sequence are incubated together such that hybridization occurs to form a population of hybridized beads and a population of unhybridized beads. The population of hybridized beads is then separated from the population unhybridized beads. In certain aspects, the second nucleic acid is immobilized on a capture bead. In other aspects, the population of hybridized beads are separated from the population of unhybridized beads by density or affinity. [014] In accordance with another embodiment, kits are provided containing an array having a plurality of beads immobilized in a semi-solid medium, in which a population of substantially identical nucleic acid sequences are attached to the beads.

BRIEF DESCRIPTION OF THE DRAWINGS

[015] The foregoing and other features and advantages of the present invention will be more fully understood from the following detailed description of illustrative embodiments taken in conjunction with the accompanying drawings.

[016] Figure 1 depicts a schematic of a biotinylated primer (SEQ ID NO:I) and a streptavidin coated bead.

[017] Figures 2A-2C depict beads that have undergone the emichment protocol described herein. (A) depicts unenriched beads. Beads comprising amplified sequence ('amplified') are red and beads not having amplified sequence ('empty') are green. (B) depicts enriched amplified beads. (C) depicts beads from the pellet fraction showing a large percentage of empty

beads.

[018] Figure 3 depicts a schematic of the monolayering protocol described herein.

[019] Figure 4 depicts a focal map.

[020] Figure 5 depicts an image of monolayered beads with transmitted bright-field at high density.

[021] Figure 6 depicts parallel sequencing on oligonucleotides coupled to 8.8-micron beads. A population of beads, each bearing one of five 80-mer oligonucleotides was immobilized in acrylamide and subjected to multiple rounds of fluorescent in situ sequencing (FISSEQ) until five to eight base-pair reads were obtained. This postprocessing image shows a region of the slide, with pseudo-colors representing individual sequences (dark blue representing "noise signatures"). A 600x600-pixel region is shown, where the resolution is approximately 0.5 micron per pixel in each dimension. Images were acquired on an inverted epifluorescence microscope.

[022] Figure 7 depicts parallel sequencing on oligonucleotides coupled to 1 micron beads. A population of 1 micron beads, each bearing one of three 80-mer oligonucleotides was immobilized in acrylamide and subjected to multiple rounds of FISSEQ until four base-pair reads were obtained. Larger beads (2.8 micron) were mixed in to serve as fiduciary markers for image registration. Correct sequence signatures are pseudocolored red, white, yellow; noise signatures are pseudocolored dark blue; and fiduciary markers are pseudocolored green. A 200x200-pixel region is shown, where the resolution is approximately 0.5 micron per pixel in each dimension. Images were captured on an inverted epifluorescence microscope.

[023] Figure 8 depicts monolayers of paramagnetic polystyrene microspheres (1 micron diameter) prepared by convective self-assembly. The monolayers can be loosely packed (A) or more tightly packed (B) by varying the amount of detergent added.

[024] Figures 9A-9D depicts amplification of nucleic acids on particles without emulsions.

[025] Figure 10 depicts a schematic of an excluded

volume approach.

[026] Figure 11 depicts a schematic of end-products of library construction.

[027] Figure 12 depicts a polyacrylamide gel electrophoresis (PAGE) gel for size selection of sheared fragments, post-cutting.

[028] Figure 13 depicts a PAGE gel for quantitating size-selected fragments.

[029] Figure 14 depicts a diagnostic 6% PAGE gel of MmeI digested rolling circle amplification (RCA) material.

[030] Figure 15 depicts a diagnostic 6% PAGE gel of a gel purified, primerless library.

[031] Figure 16 depicts a diagnostic 6% PAGE gel of the final libraries.

DETAILED DESCRIPTION

[032] The present invention provides methods of preparing and arraying large numbers of microspheres bearing amplified DNA. A variety of hybridization and enzymatic- based methods can then be applied to the amplified DNA the beads of the array simultaneously using a single small reagent volume. In certain aspects, parallel nucleic acid analysis, e.g., DNA sequencing and RNA expression profiling, may be combined with the beads of the invention. The beads and methods of making and using bead-based, high-density arrays described herein are useful for a variety of genetics-based research and diagnostic applications, which are discussed further below.

[033] The present invention provides advantages over bead-based methods known in the art. For example, methods described herein greatly increase resolution of current methods of sequencing by decreasing readout to the scale of a single micron. Another advantage of certain methods of the invention is that imbedding beads in a polymer or gel enhances current methods of "polymerase-trapping." Yet another advantage of certain of the methods described herein is that imbedding can aid image registration since the beads stay immobilized with better than 0.4 micron precision. The present

invention also advantageously provides a facile method for enriching amplified beads relative to unamplified beads. In other aspects of the invention, immobilized beads comprising a monolayer are ordered with respect to the Z axis, but are entirely disordered with respect to the X and Y axes. This provides the benefit of allowing one to generate bead arrays without the need for using a substrate that generates order in the bead pattern (e.g., a surface with etched wells for beads to fall into). A further advantage of methods described herein for embedding or otherwise attaching a bead is that it allows for integration into an acquisition system in which the beads are moved relative to the detecting means to allow for collection of data from a larger element than could typically be used with the detecting means.

[034] The present invention provides beads and bead-based arrays. As used herein, the term "bead" refers to a discrete particle that may be spherical (e.g., microspheres) or have an irregular shape. Beads may be as small as approximately 0.1 μm in diameter or as large approximately several millimeters in diameter. Beads typically range in size from approximately 0.1 μm to 200 μm in diameter, from approximately 0.25 μm to 100 μm in diameter, from approximately 0.5 μm to 50 μm in diameter, from approximately 0.6 μm to 40 μm in diameter, from approximately 0.7 μm to 30 μm in diameter, from approximately 0.8 μm to 20 μm in diameter, from approximately 0.9 μm to 10 μm in diameter or from approximately 1 μm to 9 μm in diameter. In certain aspects, beads of the invention are approximately 1 μm, 2 μm, 3 μm, 4 μm, 5 μm, 6 μm, 7 μm, 8 μm or 9 μm in diameter. Beads may comprise a variety of materials including, but not limited to, paramagnetic materials, ceramic, plastic, glass, polystyrene, methylstyrene, acrylic polymers, titanium, latex, sepharose, cellulose, nylon and the like.

[035] In accordance with certain examples, beads may have functional groups on their surface which can be used to bind nucleic acid sequences to the bead. Nucleic acid sequences can be attached to a bead by hybridization (e.g., binding to a polymer), covalent attachment, magnetic attachment, affinity attachment and the like. For example, the bead can be coated with streptavidin and the nucleic acid sequence can include a biotin moiety. The biotin is capable of binding streptavidin on the bead, thus attaching the nucleic acid sequence to the bead. Beads coated with streptavidin,

oligo-dT, and histidine tag binding substrate are commercially available (Dynal Biotech, Brown Deer, WI). Beads may also be functionalized using, for example, solid-phase chemistries known in the art, such as those for generating nucleic acid arrays, such as carboxyl, amino, and hydroxyl groups, or functionalized silicon compounds (see, for example, U.S. Pat. No. 5,919,523, incorporated herein by reference in its entirety for all purposes).

[036] Methods of immobilizing oligonucleotides to a support are described are known in the art (beads: Dressman et al. (2003) Proc. Natl. Acad. Sci. USA 100:8817, Brenner et al. (2000) Nat. Biotech. 18:630, Albretsen et al. (1990) Anal Biochem. 189:40, and Lang et al. Nucleic Acids Res. (1988) 16:10861; nitrocellulose: Ranki et al. (1983) Gene 21 :77; cellulose: (Goldkorn (1986) Nucleic Acids Res. 14:9171; polystyrene: Ruth et al. (1987) Conference of Therapeutic and Diagnostic Applications of Synthetic Nucleic Acids, Cambridge U.K.; teflon-acrylamide: Duncan et al. (1988) Anal. Biochem. 169:104; polypropylene: Polsky-Cynkin et al. (1985) Clin. Chem. 31 :1438; nylon: Van Ness et al. (1991) Nucleic Acids Res. 19:3345; agarose: Polsky-Cynkin et al., Clin. Chem. (1985) 31:1438; and sephacryl: Langdale et al. (1985) Gene 36:201; latex: Wolf et al. (1987) Nucleic Acids Res. 15:2911; incorporated by reference herein in their entirety for all purposes) and are described further herein.

[037] As used herein, the term "attach" refers to both covalent interactions and noncovalent interactions. A covalent interaction is a chemical linkage between two atoms or radicals fonned by the sharing of a pair of electrons (i.e., a single bond), two pairs of electrons (i.e., a double bond) or three pairs of electrons (i.e., a triple bond). Covalent interactions are also known in the art as electron pair interactions or electron pair bonds. Noncovalent interactions include, but are not limited to, van der Waals interactions, hydrogen bonds, weak chemical bonds (i.e., via short-range noncovalent forces), hydrophobic interactions, ionic bonds and the like. A review of noncovalent interactions can be found in Alberts et al., in Molecular Biology of the Cell, 3d edition, Garland Publishing, 1994, incorporated herein by reference in its entirety for all purposes.

[038] Embodiments of the present invention provide a

bead upon which one to millions of copies of a nucleic acid sequence (e.g., an oligonucleotide sequence or a polynucleotide sequence) is attached. In one aspect, the bead can have multiple copies of a single nucleic acid sequence attached thereto (i.e., clonal beads). In another aspect, the bead may have two, three, four, five, ten or more species of nucleic acid sequences attached thereto. For example, in one aspect, both orientations of a genetic sequence (i.e., the positive and the negative strands) may be attached to a bead.

[039] In certain embodiments, beads are provided that have a population of substantially identical nucleic acid sequences attached thereto. As used herein, the term "substantially identical nucleic acid sequence" is intended to include, but is not limited to, nucleic acid sequences having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity to one another. In certain aspects, substantially identical includes 100% sequence identity. The term substantially identical to may apply to all nucleic acid sequences attached to a bead, to primers attached to a bead and/or to amplification products attached to a bead.

[040] As used herein, the term "oligonucleotide" is intended to include, but is not limited to, a single-stranded DNA or RNA molecule, typically prepared by synthetic means. Nucleotides of the present invention Avill typically be the naturally-occurring nucleotides such as nucleotides derived from adenosine, guanosine, uridine, cytidine and thymidine. When oligonucleotides are referred to as "double-stranded," it is understood by those of skill in the art that a pair of oligonucleotides exists in a hydrogen-bonded, helical array typically associated with, for example, DNA. In addition to the 100% complementary form of double-stranded oligonucleotides, the term "double-stranded" as used herein is also meant to include those form which include such structural features as bulges and loops (see Stryer, Biochemistry, Third Ed. (1988), incorporated herein by reference in its entirety for all purposes). As used herein, the term "polynucleotide" refers to a strand of nucleic acids that can be a variety of different sizes. Polynucleotides may be the same size as an oligonucleotide, or may be two-times, three-times, four-times, five-times, ten-times, or greater than the size of

an oligonucleotide. Oligonucleotides and polynucleotides include those attached to beads and made by amplification (i.e., "amplification products") using any of the methods described herein.

[041] Oligonucleotides and/or polynucleotides may be isolated from natural sources or purchased from commercial sources. Oligonucleotide and/or polynucleotide sequences may be prepared by any suitable method, e.g., the phosphoramidite method described by Beaucage and Carruthers ((1981) Tetrahedron Lett. 22: 1859) or the triester method according to Matteucci et al. (1981) J. Am. Chem. Soc. 103:3185), both incorporated herein by reference in their entirety for all purposes, or by other chemical methods using either a commercial automated oligonucleotide synthesizer or high-throughput, high-density array methods described herein and known in the art (see U.S. Patent Nos. 5,602,244, 5,574,146, 5,554,744, 5,428,148, 5,264,566, 5,141,813, 5,959,463, 4,861,571 and 4,659,774, incorporated herein by reference in its entirety for all purposes). Pre-synthesized oligonucleotides may also be obtained commercially from a variety of vendors.

[042] In certain embodiments of the invention oligonucleotides and/or polynucleotides may be prepared using a variety of microarray technologies known in the art. Pre- synthesized oligonucleotide and/or polynucleotide sequences may be attached to a support or synthesized in situ using light-directed methods, flow channel and spotting methods, inkjet methods, pin-based methods and bead-based methods set forth in the following references: McGall et al. (1996) Proc. Natl. Acad. Sci. U.S.A. 93:13555; Synthetic DNA Arrays In Genetic Engineering, Vol. 20:111, Plenum Press (1998); Duggan et al. (1999) Nat. Genet. S21 : 10; Microarray s: Making Them and Using Them In Microarray Bioinformatics, Cambridge University Press, 2003; U.S. Patent Application Publication Nos. 2003/0068633 and 2002/0081582; U.S. Patent Nos. 6,833,450, 6,830,890, 6,824,866, 6,800,439, 6,375,903 and 5,700,637; and PCT Application Nos. WO 04/031399, WO 04/031351, WO 04/029586, WO 03/100012, WO 03/066212, WO 03/065038, WO 03/064699, WO 03/064027, WO 03/064026, WO 03/046223, WO 03/040410 and WO 02/24597; incorporated herein by reference in their entirety for all purposes.

[043] In certain embodiments, the beads of the invention are useful for analyzing libraries, e.g., genomic libraries, cDNA libraries and the like. Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt et al. (1993) Proc. Natl. Acad. Sci. USA 90:6909; Erb et al. (1994) Proc. Natl. Acad. Sci. USA 91:11422; Zuckermann et al. (1994) J. Med. Chem. 37:2678; Cho et al. (1993) Science 261 :1303; Carrell et al. (1994) Angew. Chem. Int. Ed. Engl 33:2059; Carell et al. (1994) Angew. Chem. Int. Ed. Engl. 33.2061; and in Gallop et al. (1994) J. Med. Chem. 37:1233, incorporated herein by reference in their entirety for all purposes. Libraries are described further herein.

[044] In certain embodiments, the beads of the invention are immobilized in a semi-solid medium. Semi-solid media comprise both organic and inorganic substances, and include, but are not limited to, polyacrylamide, cellulose and polyamide (nylon), as well as cross-linked agarose, dextran or polyethylene glycol. For example, beads described herein can be physically immobilized in a polymer gel. The gel can be larger in its X and Y dimensions (e.g., several centimeters) than its Z-dimension (e.g., approximately 30 microns), wherein the Z-dimension is substantially thicker than the beads that are immobilized within it (e.g., 30 micron gel versus one micron beads).

[045] In certain aspects, at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more of the beads immobilized in a semi-solid medium have nucleic acid sequences (e.g., amplified nucleic acid sequences) attached thereto. That is, some of the beads immobilized in the semi-solid medium can be empty (i.e., do not have nucleic acid sequences attached thereto), some can have only amplification primer (i.e., do not have amplified nucleic acid sequences attached thereto) and/or some can have a heterogeneous population of nucleic acids (i.e., not substantially identical sequences) attached thereto.

[046] In other aspects, each immobilized bead that contains a nucleic acid sequences attached thereto will have different nucleic acid sequences than the other immobilized beads. That is, 50%, 45%, 40%, 35%, 30%,

25%, 20%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1% or less of the immobilized beads will have nucleic acid sequences (e.g., amplified nucleic acid sequences) attached thereto that are substantially identical to the nucleic acid sequences attached to one or more of the other immobilized beads. Nucleic acid sequences attached to a bead may apply to all nucleic acid sequences attached to a bead, to primers attached to a bead and/or to amplification products attached to a bead.

[047] In still other aspects, a semi-solid medium of the invention is used in conjunction with a solid support. For example the gel described in the paragraph above can be polymerized in such a way that one surface of the gel is attached to a solid support (e.g., a glass surface), while the other surface of the gel is exposed. In certain aspects, the gel can be poured in such a way that the beads form a monolayer that resides near the exposed surface of the gel.

[048] Solid supports of the invention may be fashioned into a variety of shapes. In certain embodiments, the solid support is substantially planar. Examples of solid supports include plates such as slides, microtitre plates, flow cells, coverslips, microchips, and the like, containers such as microfuge tubes, test tubes and the like, tubing, sheets, pads, films and the like. Additionally, the solid supports may be, for example, biological, nonbiological, organic, inorganic, or a combination thereof. In certain embodiments, beads and/or the solid supports may be functionalized such that the beads may be bound to the solid support. Functional groups are discussed further herein.

[049] In certain embodiments, an array of beads can be imaged on a standard epifluorescence microscope. The use of immobilized beads allows the array to be subjected to multiple cycles of hybridization/enzymatic-based manipulations followed by imaging of visually detectable labels on molecules hybridized to the DNA immobilized on the beads or visually detectable labels incorporated into the bead- immobilized DNA itself. A variety of deteciable labels can be used with the sequencing assays described further herein. Examples of labels for use in the present invention include visually detectable labels such as fluorescein (e.g., FITC), rhodamine (e.g., TRITC, RITC), DAPI, BODIPY, Cy3, Cy5, Alexa, Texas red, Cascade blue, green fluorescent

protein (GFP), yellow fluorescent protein (YFP), cyan fluorescent protein (CFP), horseradish peroxidase, alkaline phosphatase, avidin, biotin, luciferase (e.g., renilla luciferase, firefly luciferase), and the like. Many suitable labels are known in the art and can be ordered, for example, from the catalogs from Molecular Probes (Eugene, OR) and Sigma-Aldrich (St. Louis, MO), incorporated herein by reference in their entirety for all purposes.

[050] Embodiments of the present invention are further directed to the amplification of nucleic acid sequences on beads. In certain aspects, methods of amplifying oligonucleotides include emulsion PCR, which is described further below. Other methods of amplifying nucleic acid sequences include, but are not limited to, rolling circle amplification (hyper-branched or linear) in emulsions using beads with primers to capture reaction products; rolling circle amplification (hyper-branched or linear) in aqueous solution, followed by clonal 'capture' on beads; helicase displacement amplification (HDA) in emulsions; and rolling circle amplification in situ using, for example, SiO -surface oligos, or a thin gel-immobilized oligo layer.

[051] In certain aspects, methods of amplifying oligonucleotides involves the use of PCR, such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran et al. (1988) Science 241:1077-1080; and Nakazawa et al. (1994) Proc. Natl. Acad. Sci. U.S.A. 91:360-364; incorporated herein by reference in their entirety for all purposes). Alternative amplification methods include: self sustained sequence replication (Guatelli et al. (1990) Proc. Natl. Acad. Sci. USA 87:1874, incorporated herein by reference in its entirety for all purposes), transcriptional amplification system (Kwoh et al. (1989) Proc. Natl. Acad. Sci. L/S. 86:1173, incorporated herein by reference in its entirety for all purposes), Q-Beta Replicase (Lizardi et al. (1988) BioTechnology 6:1197, incorporated herein by reference in its entirety for all purposes), recursive PCR (Jaffe et al. (2000) J. Biol. Chem. 275:2619; and Williams et al. (2002) J. Biol Chem. 277:7790; incorporated herein by reference in their entirety for all purposes) or any other nucleic acid amplification method using techniques well known to those of skill in the art. A variety of amplification methods are described in U.S. Patent Nos. 6,391,544, 6,365,375, 6,294,323, 6,261,797, 6,124,090 and

5,612,199, incorporated herein by reference in their entirety for all purposes.

[052] Embodiments of the present invention are directed to methods of amplifying oligonucleotides using the amplification methods described herein. In certain aspects, oligonucleotides are amplified by selectively hybridizing an amplification primer to an amplification site at the 3' end of an oligonucleotide using conventional methods. Amplification primers are 6 to 100, and even up to 1,000, nucleotides in length, but typically from 10 to 40 nucleotides, although oligonucleotides of different length are of use. Amplification primers may be present in solution, such as with emulsion PCR, and/or present in the semi-solid media described herein.

[053] Typically, selective hybridization occurs when two nucleic acid sequences are substantially complementary, i.e., at least about 65% 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9% or 100% complementary over a stretch of at least 14 to 25 nucleotides. See Kanehisa, M., 1984, Nucleic Acids Res. 12: 203, incorporated herein by reference in its entirety for all purposes.

[054] Overall, five factors influence the efficiency and selectivity of hybridization of the primer to a second nucleic acid molecule. These factors, which are (i) primer length, (ii) the nucleotide sequence and/or composition, (iii) hybridization temperature, (iv) buffer chemistry and (v) the potential for steric hindrance in the region to which the primer is required to hybridize, are important considerations when non-random priming sequences are designed.

[055] There is a positive correlation between primer length and both the efficiency and accuracy with which a primer will anneal to a target sequence; longer sequences have a higher Tmthan do shorter ones, and are less likely to be repeated within a given target sequence, thereby cutting down on promiscuous hybridization. Primer sequences with a high G-C content or that comprise palindromic sequences tend to self-hybridize, as do their intended target sites, since unimolecular, rather than bimolecular, hybridization kinetics are generally favored in solution; at the same time, it is important to design a primer containing

sufficient numbers of G-C nucleotide pairings to bind the target sequence tightly, since each such pair is bound by three hydrogen bonds, rather than the two that are found when A and T bases pair. Hybridization temperature varies inversely with primer annealing efficiency, as does the concentration of organic solvents, e.g., formamide, that might be included in a hybridization mixture, while increases in salt concentration facilitate binding. Under stringent hybridization conditions, longer probes hybridize more efficiently than do shorter ones, which are sufficient under more permissive conditions. Stringent hybridization conditions typically include salt concentrations of less than about 1M, more usually less than about 500 mM and preferably less than about 200 mM. Hybridization temperatures range from as low as 0 °C to greater than 22 °C, greater than about 30 °C, and (most often) in excess of about 37 °C. Longer fragments may require higher hybridization temperatures for specific hybridization. As several factors affect the stringency of hybridization, the combination of parameters is more important than the absolute measure of any one alone. Hybridization conditions are known to those skilled in the art and can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6, incorporated herein by reference in its entirety for all purposes.

[056] Primers are designed with the above first four considerations in mind. While estimates of the relative merits of numerous sequences are made mentally, computer programs have been designed to assist in the evaluation of these several parameters and the optimization of primer sequences (see, e.g., Hoover et al. (2002) Nucleic Acids Res. 30:e43, and Rouillard et al. (2004) Nucleic Acids Res. 32:W176, incorporated by reference herein in their entirety for all purposes).

[057] In accordance with certain examples, methods for emiching beads having a nucleic acid sequence of interest attached thereto are provided. Beads having a nucleic acid sequence of interest may be enriched by contacting a population of beads (wherein at least one of the beads has a nucleic acid sequence of interest attached thereto) with a nucleic acid sequence complementary to the nucleic acid sequence of interest (i.e., a complementary nucleic acid sequence) under conditions that allow hybridization of the nucleic acid sequence of interest and the complementary nucleic

acid sequence.

The un-hybridized beads are then separated from the hybridized beads which contain the nucleic acid sequence of interest attached thereto using methods known in the art.

In certain aspects, beads having a nucleic acid sequence of interest attached thereto are enriched at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%,

55%, 60%, 65%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%,

98%, 99% or more over the percentage of beads having a nucleic acid sequence of interest attached thereto in the starting population of beads. In certain embodiments, the complementary nucleic sequence is immobilized on a support. Suitable supports include, but are not limited to, synthetic polymer supports, e.g., polystyrene, polypropylene, substituted polystyrene (e.g., aminated or carboxylated polystyrene), polyacrylamides, polyamides, polyvinylchloride, and the like, polymeric beads, magnetic beads, glass beads, sepharose, agarose, cellulose, or any material useful in affinity chromatography. Methods of enriching beads having a nucleic acid sequence of interest are described further herein.

In accordance with certain examples, methods of sequencing nucleic acid sequences on beads are provided. General sequencing methods known in the art, such as sequencing by extension with reversible terminators, fluorescent in situ sequencing (FISSEQ), pyrosequencing, massively parallel signature sequencing (MPSS) and the like (described in Shendure et al. (2004) Nat. Rev. 5:335, incorporated herein by reference in its entirety), are suitable for use with the beads and bead-based arrays described herein. Reversible termination methods us step-wise sequencing-by- synthesis biochemistry that coupled with reversible termination and removable fluorescence (Shendure et al. supra ands U.S. Patent Nos. 5,750,341 and 6,306,597, incorporated herein by reference. FISSEQ is a method whereby DNA is extended by adding a single type of fluorescently-labelled nucleotide triphosphate to the reaction, washing away unincorporated nucleotide, detecting incorporation of the

nucleotide by measuring fluorescence, and repeating the cycle. At each cycle, the fluorescence from previous cycles is bleached or digitally subtracted or the fluorophore is cleaved from the nucleotide and washed away. FISSEQ is described further in Mitra et al. (2003) Anal. Biochem. 320:55, incorporated herein by reference in its entirety for all purposes. Pyrosequencing is a method in which the pyrophosphate (PPi) released during each nucleotide incorporation event (i.e., when a nucleotide is added to a growing polynucleotide sequence). The PPi released in the DNA polymerase-catalyzed reaction is detected by ATP sulfurylase and luciferase in a coupled reaction which can be visibly detected. The added nucleotides are continuously degraded by a nucleotide-degrading enzyme. After the first added nucleotide has been degraded, the next nucleotide can be added. As this procedure is repeated, longer stretches of the template sequence are deduced. Pyrosequencing is described further in Ronaghi et al. (1998) Science 281:363, incorporated herein by reference in its entirety for all purposes. MPSS utilizes ligation-based DNA sequencing simultaneously on microbeads. A mixture of labelled adaptors comprising all possible overhangs is annealed to a target sequence of four nucleotides. The label is detected upon successful ligation of an adaptor. A restriction enzyme is then used to cleave the DNA template to expose the next four bases. MPSS is described further in Brenner et al. (2000) Nat. Biotech. 18:630, incorporated herein by reference in its entirety for all purposes.

[059] This invention is further illustrated by the following examples, which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application are hereby incorporated by reference in their entirety for all purposes.

EXAMPLE I Clonal Beads Step 1 : Coupling Oligonucleotides to Beads [060] Commercially available oligonucleotides obtained from Integrated DNA Technologies were used (Coralville, IA). The oligonucleotides were attached to streptavidin-coated paramagnetic beads (1 μM MYONE™ beads (Dynal Biotech, Brown Deer, WI)) via the double-biotin moieties at their 5' ends (Figure 1). The sequence of the oligonucleotide was identical to the PR1-F section of the library molecules described in the library generation protocol below (Example VI). Subsequent to this step,

this allows the bead-immobilized oligonucleotide to act as a PCR primer during amplification of a template library as described herein. In the protocols described herein, this sequence is referred to as the 'forward' PCR primer. [061] Steps for generating forward-primer-loaded beads were as follows:

1) Using a magnetic field to the pull beads to the side of a microcentrifuge tube, lxl 0<9>MYONE™ Streptavidin paramagnetic beads (100 µL of stock solution) were washed once in 200µl TE.

2) The beads were resuspended in 180µl Bind & Wash Buffer (5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1.0M NaCl)

3) Beads were incubated for 30 minutes at 25 °C with 20 µl of 100 µM (2 nmole) 5' dual- biotinylated forward PCR primer PR1-F-2BIO: 2Bio-CCACTACGCCTCCGCTTT CCTCTCTATGGGCAGTCGGTGAT (SEQ ID NO:I)

4) Using a magnetic field to pull the beads to side of a microcentrifuge tube, the beads were washed twice with 200 µl TE.

5) The beads were resuspended in 200 µl TE (5x10<s>beads/µl) and stored at 4 °C until used in subsequent steps.

Step 2: Preparing Microemulsions [062] PCR reactions in water-in-oil emulsions were performed. The use of an emulsion isolated very small volumes of the aqueous components of the emulsion from one another, consequently, the PCR reagents and products of the PCR reaction were isolated from one another over the course of the thermocyling reaction. The use of oligonucleotides immobilized to beads (e.g., beads from step 1) as one of the two PCR primers resulted in PCR products that were physically immobilized to the beads at the end of the PCR reaction. The template used for the PCR reaction consisted of a complex mixture of polynucleotides that contained two flanking 'common' regions, with an intervening sequence that varied from molecule to molecule. In this example, an E. coli genomic library was used. The common sequences flanking the unique region of this library are referred to herein as "PR1-F" and "PR1-R". A low concentration of template was used for the PCR, such that most

aqueous compartments of the emulsion that contained a bead contained either zero or one template at the beginning of the PCR reaction. At the end of the PCR reaction, many beads were 'empty', as they were in a compartment that did not contain any template. Other beads were 'clonal', in that they contained thousands of copies of the same PCR product. The PCR product of each clonal bead differed from the PCR product of other clonal beads. Less common, but also present, were beads that were present in compartments with two or more templates. This third kind of 'non-clonal' bead contained a mixture of PCR products deriving from more than one template. Thus, the distribution of beads fell into three categories: empty, clonal and non-clonal, which could be modeled using Poisson distribution.

A. Preparation of emulsion oil phase

[063] Surprisingly, it has been discovered that the stability of the emulsion was very sensitive to the proportions of each component of the oil phase. Accordingly, extra care was taken when measuring each component to ensure minimal variability from one experiment to the next (e.g. using reverse pipetting, measuring by positive- displacement syringe and the like). The use of syringes to measure out 10% (v/v) Span 80 enabled consistent performance of the emulsion PCR.

1) 10% (v/v) Span 80 in mineral oil solution was prepared, measuring with a 10 ml or 1 ml syringe and 16 gauge needle: a) 9 ml light mineral oil b) 1 ml SPAN<®>80 (Sigma- Aldrich, St. Louis, MO) 2) The following was added to a microcentrifuge tube by reverse pipetting: a) 545 µl light mineral oil (Sigma-Aldrich) b) 450 µl 10% SPAN<®>80 in mineral oil c) 4 µl TWEEN<®>80 (Sigma-Aldrich) d) 0.5 µl TRITON<®>X-100 (Sigma-Aldrich) 3) The solution was vortexed for 30 seconds to mix thoroughly.

B. Preparation of emulsion aqueous phase

[064] It was discovered that increasing the concentration of nucleotides to the high amount set forth below yielded more signal in the bead-gel-imaging system described herein. A concurrent increase in MgCl2concentration was also necessary to keep the ratio of nucleotides : MgCl sufficiently close to 10:1. [065] The following were added to a microcentrifuge tube, and were mixed by

pipetting gently: 8.0 µl IOx MgCl2- PLATINUM<®>Taq PCR buffer (Invitrogen, Carlsbad, CA); 30 µl 50 mM MgCl2(18.75 mM) (Invitrogen); 11.3 µl 25 mM (each) dNTP mix (3.5 mM) (Invitrogen); 1.0 µl 2 mM unmodified reverse PCR primer PR1-R (Integrated DNA Technologies (IDT), Coralville, IA): CTGCCCCGGGTTCCTCATTCTCT (SEQ ID NO:2); 0.4 µl 10 µM unmodified short forward PCR primer PR1-3LF: IDT, CCTCTCTATGGGCA GTCGGTGAT (SEQ ID NO:3); 5 µl PR1-F forward-primer bearing 1 -micron beads; 20.5 µl sterile dH2O; 4.5 µl PLATINUM<®>Taq (Invitrogen; 5U/µl); and 0.25 µl 1 iiM template DNA

C. Preparation of water-in-oil emulsion

[066] The following emulsion preparation is for 5 µl (approximately 2.5x10<9>) beads. The amount of template used generally yielded 10% amplified beads. Typically, 5-50 µl beads per slide were used in the absence of emichment, or 50-200 µl enriched beads per slide were used. It has been discovered that the ratio of oil phase to aqueous phase affects the stability of the emulsion, and it has been determined that a ratio of 1:6 aqueous: oil resulted in consistently stable emulsions.

1) 400 µl oil phase was added to a 2 ml round-bottom cryogenic vial on a closed-loop magnetic stir plate set to 1400 RPM with a magnetic microstir bar.

2) 75 µl aqueous phase was added dropwise to the stirring oil phase.

3) The mixture was stined for 30 minutes at 1400 RPM with a magnetic microstir bar (no. 58948-353, VWR Scientific, West Chester, PA) on a VWR Scientific model 565 magnetic stirrer.

4) The contents of the tube were split into 8 x 200 µl tubes (50 µl each) D. Thermal cycling

[067] It was discovered that increasing the number of PCR cycles and increasing the extension time increased signal. Accordingly 120 PCR cycles were used with a 75 second extension phase.

[068] Emulsions were thermo-cycled according to the following program (minutes : seconds): a) 94 °C 2:00 b) 94 °C 0:15 c) 57 °C 0:30 d) 70 °C I :15 e) repeat steps b) - d), 119 more times f) 72 °C 2:00 g) 4 °C until use

E. Beads were recovered from the emulsion

[069] The following protocol was used to recover beads from the emulsion:

1) Contents of 8 x 200 µl PCR tubes were pooled into a single 1.5 µl microcentrifuge tube. 2) 800 µl NX Buffer (lOOmM NaCl; 1% TRITON<®>X-100; 10 mM Tris-HCl pH 7.5; 1 mM EDTA) was added. 3) The tube was vortexed for 30 seconds. 4) The tube was centrifuged at 11,000 RPM for 1.5 minutes. 5) Approximately 1150 µl supernatant was removed without disturbing pellet. 6) Step 2-5 were repeated two more times. 7) The remaining liquid was removed liquid using magnetic separation to pull beadso side of microcentrifuge tube.

8) The beads were washed twice with 50 µl TE. 9) The beads were resuspended in 5 µl TE. F. Exonuclease treatment to remove unextended forward primer from beads

[070] After emulsion PCR, 'amplified' beads bore both PCR product (double-stranded, with one strand immobilized to be beads), and residual, unextended forward primer (single-stranded). It has been discovered that the residual, unextended forward primer could be the source of background signal in subsequent steps of nucleic acid analysis. Therefore, it was desirable to eliminate it. Exonuclease I was used to selectively digest the residual unextended primer.

1) The following was mixed: 86 µl Beads + dH2O 10 µl lOx Exonuclease I reaction buffer (New England Biolabs, Beverly, MA) 4 µl of Exonuclease I (20 units/µl, New England Biolabs) 2) The mixture was incubated at 37 °C for 1 hour, mixing once at the 30 minute time-point to disrupt bead settling. 3) The mixture was incubated at 80 °C for 20 minutes to inactivate the exonuclease. 4) Using magnetic separation, the beads were washed 5x with 200 µl of NX Buffer. 5) The beads were stored at 4 °C until use.

G. Single-stranding of bead-bound product with sodium hydroxide

[071] It has been discovered that this step is important to allow efficient single-stranding of the bead-immobilized PCR products.

1) Remove all liquid from beads by magnetic separation 2) Add 50 µl 0.1M NaOH and mix with beads 3) Incubate 10 minutes at 25 °C 4) Wash once with 50 µl 0.1M NaOH 5) Wash twice with TE 6) Resuspend beads in 50 µl TE

[072] This protocol enables amplification of libraries that are several orders of magnitude more complex than libraries generated using bead-based methods known in the art. [073] Water-in-oil microemulsions may be prepared by dropwise addition of 200 µl of the aqueous phase to 400 µl of the oil phase in a 2-ml round-bottom cryogenic vial (no. 430661, Corning). The dropwise addition may be performed over 1 minute while the mixture is being stirred at 1,400 rpm with a magnetic microstir bar (no. 58948-353, VWR Scientific) on a VWR model 565 magnetic stirrer. After the addition of the aqueous phase, the mixture may be stirred for a total time of 30 minutes. Two emulsions can be made at once by placing two tubes in a rack placed at the center of the magnetic stirrer.

EXAMPLE II Enrichment of Beads

[074] The following protocol can be used to separate empty beads from clonal beads. The combined use of a low template concentration in the emulsion PCR reaction and the emichment protocol resulted in a higher fraction of 'clonally amplified' beads than would otherwise be possible. The basis for emichment was the use of a second set of large (3 micron diameter), non-magnetic beads (i.e., 'capture beads') that included a primer having a sequence identical to the 'reverse' PCR primer sequence (PR1-R). As sequence complementary to the reverse PCR primer will only be present on strands of DNA that are the product of the PCR reaction, amplified beads selectively hybridized to these large capture beads, whereas empty beads did not hybridize to capture beads efficiently. The 3 micron capture beads, hybridized to 1 micron amplified beads, were separated from the 1 micron empty beads on the basis of their differential density (e.g., by centrifuging beads through a density gradient solution).

[075] Capture beads were prepared as follows: 50 µl of SPHERO™ polystyrene streptavidin-coated beads (non-paramagnetic, 3 -micron diameter beads, Spherotech, Libertyville, IL) were pipetted into a 1.5 µl

microcentrifuge tube, centrifuged at 13.2 krpm for 30 seconds to pellet, and resuspended in 50 µl of Bind & Wash buffer. The beads were centrifuged again, all liquid was drawn from the bead pellet, and the bead was resuspended in 49.5 µl of Bind & Wash buffer. 0.5 µl of 1 mM biotin-modified 'capture primer' PRI-R-BioXL: Biotin-5'-cgtaccccgcttggtctttctcccgtaccccgcttggtctttctccCTGCCCCGGGTTCCTCATTCT CT (SEQ ID NO:4) was added. The beads were incubated for 20 minutes at room temperature with occasional mixing. The beads were centrifuged for 30 seconds to pellet, the liquid was removed, and 50 µl of Bind & Wash buffer was added. This wash step was repeated, and the beads were resuspended in 10 µl of Bind & Wash buffer. The capture beads were stored at 4 °C until use.

[076] 40 µl of 1 -micron, processed beads in TE (mixture of amplified and unamplified beads generated in the above sections) was resuspended in 20 µl of Bind & Wash buffer. 10 µl of 3-micron capture beads was added to the 20 µl of 1 -micron processed beads, and the beads were mixed by pipetting. Capture beads were hybridized to processed beads by incubating the mixture at 56 °C for 10 minutes. The mix was then carefully pipetted onto the top of 150 µl 60% glycerol (v/v) in a 1.5 µl microcentrifuge tube. The microcentrifuge tube was centrifuged for 1 minute at 13.2 krpm. Because of the differential density of the non-magnetic 3 micron beads and the magnetic 1 micron beads, the 3 micron beads remained in the supernatant (along with amplified 1 micron beads hybridized to them), while unhybridized, unamplified magnetic one micron beads formed a pellet at the bottom of the tube. The liquid was drawn out of the microcentrifuge (without disturbing pellet of beads at bottom of tube) and pipetted into a new microcentrifuge tube. The amplified beads, hybridized to the capture beads, were emiched in this supernatant.

[077] Next, the emiched fraction of amplified one-micron beads was purified away from the capture beads. To the new tube containing the supernatant, 1 mL of water was added, mixed by pipetting, and centrifuged for 2 minutes at 13.2 krpm. All but 20 to 30 µl of liquid was drawn off, 50 µl of fresh water was added, mixed, and centrifuged again at 13.2 krpm for 2 minutes. All liquid was removed from the pellet of beads, and the beads were resuspended in 50 µl of 0.1M sodium hydroxide. The beads were incubated for 10 minutes with occasional

mixing in order to dissociate the 1 micron beads from the 3 micron capture beads. A magnetic field was applied to the microcentrifuge tube to draw magnetic 1 micron beads to the side of the tube, and all supernatant was removed. The supernatant, which contained the 3 micron capture beads, was cloudy white. The 1 micron beads were washed once in 0.1 M NaOH and three times in lx PCR buffer. The 1 -micron beads were resuspended in 5 uL of TE and stored at 4 °C. [078] Figures 2A-2C depict a set of beads that went through the protocol as described above. The beads were taken from various steps to examine the fraction of beads that were amplified relative to those that were empty. In these pseudocolored figures, green beads were empty and red beads were amplified. In the starting material of processed but unenriched one micron beads generated by emulsion PCR, approximately 8% of beads were amplified (Figure 2A). In the set of beads that were eliminated by the above procedure, less than 1% were amplified (Figure 2C). In the fraction of material emiched for amplified beads, the fraction of amplified beads had risen from 8% to 43% (Figure 2B), a 5.5-fold emichment.

EXAMPLE III Monolayers of Disordered, Immobilized Beads

[079] The critical nature of monolayering can be understood in the context of the depth of field of high resolution microscopes. For example, the depth of field with a 20x Plan Apo (NA = 0.75) objective, one of the best commercially available objectives in terms of resolution is 1.9 microns. When clonal microspheres are only one micron in diameter, significant deviations from a monolayer would thus result in an inability to image arrayed beads while maintaining focus on all beads within a given field of view.

[080] The following steps could take place before or after amplification (e.g., by a method for making populations of clonal microspheres), and optionally, after the bead enrichment protocol had been performed. Parallel sequencing or other forms of cyclic nucleic acid analysis would be carried out after the beads had been monolayered.

[081] To form a monolayers of beads, the following protocol was used. The following reagents were mixed: 3.00 µl of beads (at desired density); 5.10 µl of dH2O;

1.50 µl of acrylamide:bis (38% acrylamide, 2% bis-acrylamide; Roche); 0.60 µl of RHINOHIDE™ (polyacrylamide gel strengthener) (Molecular Probes, Eugene, OR); 1.20 µl of 5% N,N,N',N'-tetramethylethylenediamine (TEMED); and 1.80 µl of ammonium persulfate solution (APS) (0.5%).

[082] The mixed reagents were poured between a teflon-coated microscope slide and a coverslip, and allowed by polymerize, resulting in a 30 micron thick gel. Either the coverslip or the microscope slide was coated with Bind Silane (3-Methacryloxypropyltrimethoxysilane), such that the gel will stick to either the microscope slide or to the coverslip after polymerizing. The protocol for coating a glass surface in Bind Silane is described in Mitra et al. (Id.), for example. In brief, 220 µl of acetic acid and 4 mL of Bind Silane reagent were mixed into 1 liter of dH2O. The glass surface was exposed to this solution via immersion for 1 hour with gentle shaking. Glass surfaces were washed by immersion three times in dH2O and once in 100% ethanol. Slides were allowed to air-dry and were stored dessicated.

[083] During polymerization, the microscope slide and polymerizing gel were placed in an orientation such that the beads would settle towards the desired side of the gel. Thus, the beads would form either a 'top layer' or a 'bottom layer' with respect to the exposed surface of the gel (Figure 3). For performing hybridization and enzymatic reactions on DNA immobilized to the beads, it has been discovered that the beads in a 'top layered' gel are much more accessible to enzymes and oligonucleotides applied to them, permitting much more efficient reaction kinetics. This is not surprising, as many of the enzymatic reactions used to query the gel are performed by enzymes that would have difficulty accessing the beads if they were actually deep within the gel. The gel was 30 microns thick while the beads were only one micron thick. Although the gel pores were large, diffusion of enzymes and or oligonucleotides through them would be slow if the beads sat at the bottom of the gel, for example.

[084] The final product consisted of a glass surface to which the acrylamide gel was attached via the Bind Silane reagent (Figure 3). The gel was approximately 30 microns thick. The beads were present in a monolayer at the top of the gel (i.e. most distal from the glass

coverslip to which the gel is attached via the Bind-Silane).

[085] This example provides that the concentrations of the catalyzing agents (APS and TEMED) are much lower than what is typically used in the art to polymerize gels, such that after the gel is poured, it will polymerize at a slow rate (about one hour) allowing the beads to settle via gravity into a monolayer at one surface of the gel. It has been determined that if the polymerization process is too rapid, the beads do not have time to settle into a monolayer, and the result is a layer of beads that cam ot be visualized in a single focal plane. The act of inverting the polymerizing gel such that the beads settle in a monolayer at the exposed surface of the gel is advantageous for enabling enzymatic reactions that are subsequently used to manipulate the DNA on the beads.

[086] A motivation for using 'ordered' arrays of beads or other substrates (i.e., the bead- based arrays of the art) has been the desire to make repeated independent observations of the same feature, or desire for 'addressability', typically taking the form of an ordered array with Cartesian coordinates. The arrays described in this example are ordered with respect to the Z-axis, in that they are monolayered, but disordered with respect to the X and Y axes. Individual beads have been visualized and distinguished at high densities and repeated, independent observations have been made of a given subsection of the gel, which have been aligned to previous observations of the same subsection of the gel, thus allowing multiple independent observations of individual beads within each subsection. Bead positions remained invariant even with numerous enzymatic manipulations, exposure to 95 °C heat, and the like. The use of disordered monolayers thus avoids many of the difficulties associated with generating ordered arrays while accomplishing the same goal.

EXAMPLE IV Primer Hybridization and Sequencing by Single Base Extension

[087] At this step in the protocol, the beads reside in a monolayer at the exposed surface of an acrylamide polymer gel that is itself attached to a glass coverslip. The DNA on the surface of the beads can thus be easily exposed to a variety of reagents and conditions without disturbing their absolute positions in the gel. It has been

discovered that a variety of enzymatic and chemical reactions can be performed on DNA immobilized in this format including, but not limited to, sequence-specific oligonucleotide hybridization, polymerase-driven primer-extension, restriction endonuclease-driven sequence-specific cleavage, ligase-driven oligonucleotide ligation, exonuclease- driven DNA degradation and the like.

[088] As an example of these enzymatic reactions, the protocol for hybridizing a oligonucleotide primer and then performing a polymerase-driven single-base-extension with a fluorescent nucleotide is described.
[089] As the PCR products on the bead have been single-stranded, they now consist, from 5' to 3', of the double-biotin moiety immobilizing the strand to the beads, the PR1-F forward primer sequence, unknown sequence (dependent on material from which library was constructed, and variable from bead to bead), and sequence complementary to the PR1-R primer. To hybridize the PR1-R primer to the bead- immobilized DNA, 100 µl of 1 µM PR1-R in Hybridization Buffer (6x SSPE with 0.01% TRITON<®>X-100) was applied to the gel, and the liquid was sealed into contact with the gel using a 125 µl FrameSeal chamber. The slide was heated at 56 °C for 10 minutes, the FrameSeal chamber was removed, and the slide was immersed in lx Wash Buffer (10 mM Tris, pH 7.5; 50 mM KC1; 2 mM EDTA; 0.01% TRITON<®>X- 100). The slide was incubated with shaking for 2 minutes in wash, and washed twice more in lx Wash Buffer. The PR1-R primer was thus hybridized to DNA. The 3' end of the PR1-R primer was positioned immediately adjacent to the unknown sequence of the bead-immobilized molecules.

[090] To query the identity of the first unknown base by polymerase-driven single-base- extension, the following mix was prepared, containing both polymerase and fluorophore-labeled ddNTPs: 122 µl of lx ThermoSequenase™ Buffer (Amersham Biosciences, Piscataway, NJ); 1 µl of ThermoSequenase™ (4 u/uL, Amersham Biosciences); 0.5 µl of RI IO-ddGTP (100 µM, PerkinElmer, Wellesley, MA); 0.5 µl of Cy5-ddCTP (100 µM, PerkinElmer); 0.5 uL of Cy3-ddUTP (100 µM, PerkinElmer); and 0.5 µl of Texas-Red ddATP (100 µM, PerkinElmer).

[091] The mixture was applied to the slide as above using a 125 µl FrameSeal chamber, and incubated at 42 °C for 5 minutes to extend. The immersion was washed

in Ix Wash Buffer for 2 minutes with slow shaking at room temperature, and the wash was repeated 1 to 2 times before imaging. Imaging was performed on an epifluorescence microscope equipped with the a xenon lamp light-source and the appropriate set of filters for the fluorophores used herein. Each bead emitted fluorescence with filters corresponding to only one of the four fluorophores, revealing the identity of the incorporated base and thus of that unknown position in the DNA sequence. EXAMPLE V Imaging on Epifluorescence Microscope

[092] The beads were immobilized on a gel which was immobilized to a glass coverslip. The gel and the glass could both be imaged through. It has been determined that the acrylamide gel, at the % that was used did not cause significant autofluorescence or interference with light transmission.

[093] To image fluorescent moieties incorporated into or on molecules associated with DNA immobilized to the beads, an epifluorescence microscope (Nikon TE2000) with an automated X-Y stage (Prior) and focus control was used. Epifluorescent illumination was by either a mercury arc lamp, a xenon arc lamp, or a mercury halide arc lamp. Images were acquired by a CCD-based detector mounted on the microscope. Either long working distance objectives or high-numerical aperture objectives were used, with typical magnifications from IOx to 40x. Typically, the beads were immobilized in a gel which was attached to a microscope slide. Alternatively, the immobilized bead-gel was occasionally attached to a glass coverslip which mounted in a flowcell or a microscope slide which mounts in a flowcell. Either the flowcell or microscope slide fit into a detail in the stage. In the case of the microscope slide, the slide could be removed from the microscope between cycles of image acquisition to perform experimental protocol (e.g., sequencing). One problem associated with removing the slide from the stage between image acquisition cycles is the inability to accurately reposition the slide on the microscope stage with micron accuracy. Typically, one could expect a repositioning error of 100 microns or more. This error, at 40x magnification with 7 micron CCD pixel size, would result in a data loss of 75% or more due to image misalignment from one cycle to the next. It should be noted that this problem can also occur to a lesser extent if the slide or flowcell is not removed from the

microscope, but an experimental protocol is performed which involves fluctuations in temperature of the slide or flowcell. Such temperature changes will induce mechanical changes in the geometry of the slide or flowcell and can result in image misalignment. To allow accurate re-positioning of the microscope stage in the X, Y, and Z axes after successive cycles of imaging and slide removal to within several microns in X and Y, and less than a micron in Z the following steps may be used: Step I

[094] At the start of the experiment, a 'focal map' of the surface to be imaged was generated. This focal map is a list of x, y, z coordinates on the bead array for the microscope to visit during each acquisition cycle. This focal map is normally generated by executing an 'auto focus' routine at each x, y position in the list to generate the corresponding z position. The auto focus algorithm used is implemented by the METAMORPH<®>acquisition software (Universal Imaging Corporation, Downingtown, PA) and uses either transmitted brightfield images, reflected brigbtfield images, or epifluorescence images where beads of interest have been labeled with a fluorescent molecule or molecules.

Step 2

[095] At the start of each imaging cycle, the software returned to the first x, y, z hardware position in the focal map. The auto focus routine was executed for that particular field of view to compensate for positioning error in the z axis. An image was then acquired and passed to an image registration software program which found the appropriate offsets in the x and y axes to translate the new image in order to bring it into registration with the original image of that frame (to within a pixel; distance in microns is given by the pixel size, normally approximately 7 μm). Since the beads were immobilized, they could serve as fiduciary markers, obviating the need for introduction of additional features in the anay. Thus, having disordered beads was a strong advantage when one is performing precise image alignment. The microscope stage was then moved in x and y by these offsets. The bead array was then at an x, y, z position close to the original position from Step 1. It has been discovered that images can be brought into alignment with sub-micron resolution using this method. A second auto focus was performed to

ensure that the new x, y position was in focus. The origin was reset at this new position.

Step 3

[096] Images were acquired by visiting each position in the original focal map. These images were be in register with all previous cycles, allowing extraction of data for most or all beads in every frame. An example of plotted coordinates of a typical 'focal map' is set forth in Figure 4, which shows a circular gel (approximately 1 cm in diameter) with two large holes. Red points indicate the relative XYZ coordinates of individual frames. Figure 5 depicts a sample image of monolayered beads at imaged with transmitted bright-field at high density.

EXAMPLE VI Paired-Tag In Vitro Library Construction

[097] An in vitro library construction protocol has been developed that results in a pool of linear dsDNA molecules, where each molecule is approximately 134-136 bp in length, and includes a unique pair of 17-18 bp tags derived from the genome-of- interest. These unique tags are flanked by a set of sequences complementary to universal primers (PR1-R and PR1-F) and separated by an additional universal spacer sequence ("T30") (Figure 11). The unique tags are "paired" in that they are identically oriented on the genome-of-interest and their separation on the genome-of- interest falls within a constrained range (e.g. 1000 +/- 100 bases). The protocol is unique in that it is entirely in vitro; no transformation into E. coli is required.

[098] A library in this format constructed by in vitro methods, provides the following benefits over conventional genomic shotgun sequencing: (a) the emulsion PCR protocol is considerably more efficient for amplifying short sequences; thus there is a motivation to minimize the total length of each amplifiable molecule in the library; (b) the experience of Sanger-based genome projects indicate that paired reads are immensely useful for genomic re-sequencing, especially when a genome contains repetitive elements; (c) the methods described herein to sequence the unique bases rely on the positioning via an universal 'anchor' sequence. In this library format there are both proximal and distal anchors for two tags, effectively enabling one to double read-lengths by applying the sequencing methods independently for each tag.

[099] The library construction protocol had the following steps:

1. purification of genomic DNA 2. shearing of genomic DNA to generate fragments 3. end-repair and A-tailing of DNA fragments 4. PAGE size-selection of sheared fragments 5. circularization with T-tailed spacer oligonucleotide ("T30") 6. rolling circle amplification (RCA) with random hexamers 7. digestion with MmeI (type IIs) to release paired tags 8. PAGE purification of tag-T30-tag library 9. end-repair of tag-T30-tag library 10. ligation of FDV2 (PR1-F) and RDV2 (PR1-R) primer oligonucleotides 11. PAGE size-selection of paired-tag library 12. nick translation to eliminate nicks in dsDNA library 13. PCR amplification of paired-tag library 14. PAGE size-selection of paired-tag library 15. Library validation via cloning and S anger sequencing

[0100] Detailed protocols for each of the above steps are given below, as they were performed for constructing paired-tag libraries for the "M" and "R" E. coli strains.

Purification of Genomic DNA

[0101] For each of E. coli strains "M" and "R", cultures were grown overnight in 3 mL of LB and isolated with the Qiagen DNEASY<®>Tissue kit as per the manufacturer's protocol. Yield for each genomic DNA purification was approximately 30 µg (Qiagen Inc., Valencia, CA).

Shearing of Genomic DNA to Generate Fragments with a Broad Size Distribution

[0102] Shearing of genomic DNA from both strains was performed by Agencourt Bioscience Corporation (Beverly, MA). The size-distribution of the resultant DNA fragments was quite broad, as can be seen on the gel below (Figure 12).

End-Repair and A-Tailing of DNA Fragments

[0103] Unless stated otherwise, all DNA quantitation was performed on a Nfanodrop ND- 1000 Spectrophotometer. [0104] Sheared genomic "M" DNA was quantitated at 57 ng/µl, and sheared genomic "R" DNA was quantitated at 55 ng/ µl. Sheared DNA

fragments were end-repaired with the EpiCentre END-IT™ DNA End Repair Kit (Madison, WI). For each library, the following mix was prepared: 170 µl of sheared E. coli DNA (-9-10 µg); 25 µl of lOx END-IT™ Buffer; 25 µl of lOx END-IT™ ATP; 25 µl of lOx END-IT™ dNTPs; and 5 µl of END-IT™ Enzyme for a total volume of 250 µl.

[0105] Reactions were incubated at room temperature for 1 hour. DNA was purified on a Qiagen QIAQUICK® column as per manufacturer's recommendations for PCR product purification. Approximately 90 µl of Buffer EB (10 mM Tris-Cl, pH 8.5) was used for elution. "M" DNA was quantitated at 96.5 ng/µl, and "R" DNA was quantitated at 75.1 ng/µl. Each volume was split to 4 tubes of approximately 22 µl. To eliminate residual enzyme activity, tubes were heated to 70 °C for 15 minutes. An A-tailing master-mix was prepared as follows: 100 µl of lOx PCR buffer (no MgCl2) (Invitrogen, Carlsbad, CA); 60 µl of 50 mM MgCl2(Invitrogen) (final concentration of 3 mM); 5 µl of 100 mM dATP (Invitrogen) (final concentration of 0.5 mM); 5 µl of Taq (5 U/uL) (New England Biolabs, Beverly, MA); and 610 µl of dH2O.

[0106] After heat-inactivation, 78 µl of the master-mix was added to each tube containing 22 µl of sheared, end-repaired DNA fragments. Tubes were incubated at 70 °C for 30 minutes in a thermal cycling machine. The cycling program ended by cooling the tubes to 4 °C. The tubes were and then transferred from the thermal cycler directly to ice.

[0107] DNA was purified by phenol-chloroform extraction and ethanol precipitation (P : C : P) as follows:

1. Added an equal volume of phenol : chloroform : isoamyl alcohol (25 : 24 : 1) 2. Added 0.1 volume of 3M NaOAc (pH 5.2) 3. Added 1.0 µl of glycogen (20 µg/µl) 4. Added 2.5 volumes of cold 100% ethanol (from bottle stored at -20 °C) 5. Mixed by inverting 6. Put tube at -70 °C for -30-60 minutes 7. Spun at maximum speed on microcentrifuge in 4 °C room for 10 minutes 8. Removed supernatant 9. Added 1 ml of 80% ethanol 10. Spun at maximum speed on microcentrifuge at room temperature for 5 minutes 11. Removed supernatant 12. Placed tube on Speed- Vac for -5 minutes 13. Resuspended pellet in 40 µl Buffer EB or TE

PAGE Size-Selection of Sheared Fragments

[0108] Half of the material from each library was loaded in a pre-cast 6% TBE-PAGE gel (Invitrogen, Carlsbad, CA). 20 µl of DNA was mixed with 5 uL of 5x High-Density Sample Buffer (Novex, San Diego, CA). The same loading buffer was used for all subsequent PAGE gels in this protocol. 12.5 µl of the sample/loading buffer mixture was loaded per lane (two lanes per library). The gel was run on standard apparatus and a region corresponding to approximately 1000 base pair fragments was cut out with minimal exposure to UV. Gel fragments were diced with razor and each library's fragments transferred to 600 µl of PAGE elution buffer (10 mM Tris-HCl (pH 7.5), 50 mM NaCl, 1 mM EDTA (pH 8.0)). Tubes were incubated at 37 °C overnight.

[0109] The following day, the elutions were spun for one minute at maximum speed in a microcentrifuge and the supernatants were transferred to new tubes. To improve recovery, washed gel fragments with an additional 200 uL of PAGE buffer. DNA was purified using the P:C:P protocol, and each pellet was resuspended in 22 µl of Buffer EB.

[0110] The A-tailing step was repeated to minimize the impact of potential partial degradation of the A-tails that may have occurred during the gel purification. Both end-repair and A-tailing may be performed after the PAGE-based size-selection, rather than before. An A-tailing master-mix was prepared as follows: 25.00 µl of lOx PCR buffer (no MgCl2); 15.00 µl of 50 mM MgCl2(final concentration of 3 mM); 1.25 µl of 100 mM dATP (final concentration of 0.5 mM); 1.25 µl of Taq (5 U/ µl); and 152.50 µl of dH2O.

[0111] 78 µl of the A-tailing master-mix was added to each tube containing DNA resuspended in 22 µl for total volume of 100 µl. For each library, this was split into two thermal-cycler compatible tubes of 50 µl each. Tubes were incubated at 70 °C for 30 minutes in a thermal cycling machine. The cycling program ended by cooling the tubes to 4 °C. The tubes were then transferred from the thermal cycler directly to ice. DNA was purified by P : C : P protocol. Each library was resuspended in 10 µl of Buffer EB and put on ice.

[0112] To quantitate recovery and estimate the size-

range of recovered fragments, a pre-cast 6% TBE PAGE gel (Invitrogen) was run using 20% of the purified material (Figure 13). Gel-based quantitation of "M" = 43 ng (in 2 µl, or 20% of the total); range = approximately 850-1150, mean = approximately 1000. Gel-based quantitation of "R" = 18 ng (in 2 µl, or 20% of the total); range = approximately 900-1250; mean = approximately 1075. In the remaining approximately 8 µl volume retained for each library, there was approximately 171 ng of "M" fragments and approximately 73 ng of "R" fragments.

Circularization with T-Tailed Spacer Oligonucleotide ("T30")

[0113] Next, the A-tailed library fragments were circularized using the T-tailed spacer oligonucleotide "T30". The T30 segment was prepared by annealing two 32-bp oligonucleotides to generate a 30 bp dsDNA fragment with single base "T" overhangs:

5 ' GTCGGAGGCCAAGGCGGCCGTACGTCCAACT 3 ' (SEQ ID NO:5) 3 ' TCAGCCTCCGGTTCCGCCGGCATGCAGGTTG 5 ' (SEQ ID NO:6) The T30 segment is flanked by outward-facing MmeI sites.

[0114] Annealing of the oligos was performed by mixing to a final concentration of 50 µM for each oligo, heating to 95 °C for 10 minutes in a thermal cycler, shutting the thermal cycler off and allowing the mixture to cool slowly back to room temperature over the course of an hour.

[0115] Ligation of the T30 fragment with the A-tailed library fragments was performed using a QUICK LIGATION™ kit (New England Biolabs, Beverly, MA). Reactions were prepared as follows. "M" library circularization reaction (total volume of 80 µl): 8.0 µl of A-tailed fragments from "R" tube (~171 ng at 1000 bp = 0.2599 pmol); 27.2 µl of dH2O; 0.8 µl of T30 (1 µM starting concentration (0.8 pmol, 3 -fold molar excess)); 40.0 µl of 2x QUICK LIGATION™ Buffer; and 4.0 µl of QUICK™ T4 DNA ligase. "R" library circularization reaction (total volume of 30 µl): 8.0 µl of A- tailed fragments from "R" tube (~73 ng at 1075 bp = 0.1032 pmol); 5.2 µl of dH2O; 0.3 µl of T30 (1 µM starting concentration (0.3 pmol, 3 -fold molar excess)); 15.0 µl of 2x QUICK LIGATION™ Buffer; 1.5 µl of QUICK™ T4 DNA ligase. Each reaction was mixed well before and

after adding enzyme to each tube. The reactions were incubated for 10 minutes at room temperature, then moved to ice.

[0116] The ligase was heat-inactivated on a thermal cycler at 65 °C for 10 minutes. To destroy all non-circularized material, an exonuclease mix was added. The exonuclease mix is prepared as follows: 4.0 µl of Exonuclease I (20 U/µl) (New England Biolabs, Beverly, MA); 0.4 µl of Exonuclease III (100 U/µl) (New England Biolabs, Beverly, MA); and 35.6 µl of TE. 10 µl of exonuclease mix was added to the 80 µl "M" reaction, and 3.75 µl of exonuclease mix was added to the 30 µl "R" reaction. The tubes were incubated for 45 minutes on a thermal cycler at 37°C, followed by 80 °C for 20 minutes to heat-inactivate the exonucleases. This material was used directly in the RCA reaction of the next step without any purification.

Rolling Circle Amplification (RCA) with Random Hexamers

[0117] Hyperbranched RCA was performed to amplify the amount of library material using the REPLIPHI™ phi29 kit (EpiCentre, Madison, WI). A master-mix was prepared as follows: 32.0 µl of dNTP mix (25 µM each); 80.0 µl of lOx REPLIPHI™ phi29 reaction buffer; 40.0 µl of random DNA hexamers (1 mM, synthesized as 5'-NNNN*N*N-3', where "*" indicates phosphorothioate linkage); 552.0 µl of dH2O; and 16.0 µl of 5x SybrGreen. The master-mix was split into two tubes of 270 µl, to which either 30 µl of "M" or "R" material was mixed for a total volume of 300 µl. Each tube was then split into 6 tubes of 50 µl. To denature circularized template, the tubes were heated to 95 °C for 5 minutes, followed by rapid cooling to 4 °C. 2.5 µl of phi29 enzyme was added to each tube on ice for total volume of 52.5 µl per tube. The tubes were mixed well and kept on ice. The tubes were incubated overnight at 30 °C in the thermal cycler.

[0118] The RCA reactions were run on a real-time PCR machine, and amplification was observed via the SybrGreen dye present in the reaction. The dsDNA content had risen and leveled off by the 2-hour time-point, indicating that running the reaction overnight may not be necessary.

[0119] DNA was purified with a MICROCON-

30<®>column (Millipore, Billerica, MA), washing with a total of 1 mL of TE. The pellet was substantial. Several washings of the MICROCON-30<®>membrane were used to maximize recovery. A combination of heating at 50 °C and adding additional resuspension buffer (Buffer EB) was used to resuspend the DNA. Approximately 750 uL of each library was recovered.

[0120] Samples were quantitated on a NANODROP<®>instrument (NanoDrop Technologies, Wilmington, DE): "M"= 230 ng/uL and "R"= 204 ng/uL. The RCA reaction thus resulted in ~150 µg of each library.

Digestion with MmeI (type IIs to Release Paired Tags

[0121] Approximately 40 µg of each library was digested with MmeI. As the MmeI site cuts at a distance from its recognition site in the T30 segment, and there are outward- facing MmeI sites at either end of the T30 segment, this digestion was expected to release the T30 segment flanked by ~18 bp tags with 2 bp overhangs (~70 bp in length). Because genomic fragments were circularized with T30 prior to MmeI digestion, these tags were expected to be paired with respect to the positions of their origin.

[0122] Reactions were prepared as follows. MmeI, S-adenosylmethionine (SAM), and NEBuffer 4 (lOx) were obtained from New England Biolabs (Beverly, MA). 32 mM SAM was diluted 1 :20 (-> 1.6 mM) in lx NEBuffer 4. Reaction ' 'M" Reaction "R" DNA 173.9 196.0 dH2O 664.5 642.4 NEBuffer 4 (1 Ox) 100.0 100.0 1.6 mM SAM 1.6 1.6 MmeI (2U/µl) 60.0 60.0 total volume (µl) 1000.0 1000.0 Table 1.

[0123] Reactions were prepared on ice, and reagents were well-mixed prior to adding enzyme. Each reaction was split to 8 tubes of 125 µl, and incubated on a thermal- cycler for 30 minutes at 37 °C. P:C:P purification as described in above, except using only 2 volumes of ethanol instead of 2.5 volumes, without heat-inactivating first. Digested fragments for each library were resuspended in 80 µl of TE. A band at approximately 70 bp was observed in the "M" and "R" lanes, as expected (Figure 14).

PAGE Purification of Tag-T30-Tag Library

[0124] The full amount of each library was run on a 10-lane pre-cast 6% PAGE gel as above, using 4 lanes per library (20 µl of library and 5 µl of 5x dye). A sharp band at approximately 70 base pairs was cut. Fragments from all lanes of each library type were combined, and the gel-extraction was carried out as described above, except elution was for approximately 3 hours. After P:C:P recovery, the DNA for each library was resuspended in approximately 20 µl of TE. A diagnostic gel was run to quantitate the recovered material (Figure 15). Both the "M" and "R" libraries were estimated at approximately 12.5 ng/µl, and 18 µl of each remained at this point.

End-Repair of Tag-T30-Tag Library

[0125] The tag-T30-tag molecules contained 2 bp 3'-overhangs, consequent to MmeI digestion. The ends were repaired using EpiCentre END-IT™ DNA End Repair Kit as described above (Madison, WI). Reactions were prepared as follows: 8.50 µl of "M" or "R" fragments (12.5 ng/µl: approximately 100 ng); 1.25 µl of lOx END-IT™ Buffer; 1.25 µl of lOx END-IT™ ATP; 1.25 µl of lOx END-IT™ dNTPs; and 0.25 µl of END-IT™ Enzyme for a total volume of 12.5 µl.

[0126] Reactions were incubated at room-temperature for 45 minutes, then moved directly to 4 °C. Volumes were increased to 50 µl by adding 40 µl of TE, and the reactions were P : C : P extracted as described above. The precipitation step was allowed to go overnight at -70 °C. Recovered DNA was resuspended in 8 µl of TE.

Ligation of FDV2 fPRI-F and RDV2 (PR1-R) Primer Oligonucleotides

[0127] The primer-adaptors (dsDNA, FDV2 and RDV2) were prepared by annealing fully complementary oligonucleotides (100 µM, HPLC-purified) by mixing 1:1 (final concentration of 50 µM), heating to 95 °C for 10 minutes, and allowing the reaction to cool slowly over the course of an hour.

[0128] In "annealed" format, FDV2 and RDV2 were as follows:

FDV2:

5 '

-AACCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT (SEQ ID NO:7) 3' -TTGGTGATGCGGAGGCGAAAGGAGAGATACCCGTCAGCCACTA (SEQ ID NO:8) RDV2:

5 ' -AACTGCCCCGGGTTCCTCATTCTCT (SEQ ID NO:9) 3' -TTGACGGGGCCCAAGGAGTAAGAGA (SEQ IDNO:10)

[0129] The FDV2 and RDV2 molecules were unphosphorylated, and therefore were not expected to be able to self-self ligate nor to ligate to one another. The end-repaired ligated molecules were phosphorylated, and therefore, an excess of FDV2 and RDV2 was used to minimize concatamerization events for library molecules. Ligations of primer-adaptors to the library molecules were blunt-blunt and therefore conducted in the presence of polyethylene glycol (PEG) to improve ligation efficiency. Each reaction was set up as follows: 12.3 µl of dH O; 8.0 µl of purified, blunted library fragments ("M" or "R"; approximately 100 ng, i.e., approximately 2 pmol); 1.0 µl of RDV2 (50 µM, 50 pmol); 1.0 µl of FDV2 (50 µM, 50 pmol); 2.5 µl of lOx T4 Ligase Buffer (New England Biolabs, Beverly, MA); 21.2 µl of 40% PEG (40% polyethylene glycol 8000); and 2.0 µl of T4 Ligase (2000 U/µL) (New England Biolabs). Reactions were prepared at room-temperature by mixing all reagents except the PEG and ligase. The PEG was then added and mixed in, and the ligase was added and mixed in. Reactions were incubated at 16 °C overnight. To purify, reaction volumes were increased to 100 µl with TE and P : C : P purified. Pellets were resuspended in 10 µl of Buffer EB.

PAGE Size-Selection of Paired-Tag Library

[0130] Entire reactions were run on a 10 well 6% PAGE gel along with the appropriate ladders. The gel was run far enough such that unligated RDV2 and FDV2 (present in great molar excess relative to the library) were expected to have run off the gel. A triplet of bands of the appropriate size was observed as expected (resulting from RDV2/library fragment/RDV2 ligation, RDV2/library fragment/FDV2 ligation, or FDV2/library-fragment/FDV2 ligation). The regions containing the full triplets were cut and gel-purified as described above, except elution was for three hours, and ethanol precipitation was overnight at -70 °C. Samples were each resuspended in 20 µl of Buffer EB.

Nick Translation to Eliminate Nicks in dsDNA Library

[0131] As only the library molecules were 5 '-phosphorylated in the ligation, the ligation products were expected to contain nicks that must be repaired. Moving forward with half of the remaining material, nick translation was performed as follows: 10.0 µl of library (assuming 100% recovery, this should be 50 ng of tag-T30-tag molecules plus the mass of ligated primer-adaptors); 0.5 µl of dNTP mix (25 mM, so final concentration of 500 µM for each nucleotide); 2.5 µl of lOx NEBuffer-2 (New England Biolabs); 1.0 µl of E. coli DNA polymerase I (10 U/µl) (New England Biolabs); and 11.0 µl of dH2O. Reactions were prepared and mixed on ice, and then incubated at 16 °C for 30 minutes. To purify, the volume of the reactions was increased to 100 µl with TE, the reactions were P:C:P purified, and resuspended in 10 µl ofTE. PCR Amplification of Paired-Tag Library

[0132] PCR was performed at this stage to 1) increase the amount of library material that we have to work with, and 2) to eliminate extraneous ligation products in a single step. Without intending to be bound by theory, the only ligation products that should result from the PCR described in this step have tag-T30-tag flanked by properly oriented RDV2 and FDV2 on either side (note that the T30 segment itself is not symmetric, and therefore may be in either oriented in either direction relative to RDV2 and FDV2 segments).

[0133] As a complex mixture was being PCR-amplified, it was critical to stop the PCR reaction before primer molecules were exhausted. This is due to the fact that library molecules will begin to serve as primers for one another once the intended primers have run out, and that the library molecules contain enough similarity (-100 out of -134 identical bases) such that after denaturing, it is unlikely that a given single- stranded library molecule will reanneal to its exactly complementary partner. The resultant library that has denatured and reannealed after primer has been exhausted can contain many "hybrid" library molecules.

[0134] PCR amplification was performed on a real-time PCR machine (OPTICON™ 2, MJ Research, Bio Rad, Waltham, MA).

Per 50 µl x20 (total volume of 1000 uD

1 Ox PCR Buffer 5 100 25 mM (each) dNTPs 0.4 8 50 mM MgCl21.5 30 Platinum Taq 0.2 µl 4 Water 42.7 853 RDV2-T (100 µM) 0.1 2 FDV2-T (100 µM) 0.1 2 SybrGreen (200x) 0.0025 0.5 Table 2. [0135] The master mix (set forth in Table 2) was split into two tubes of 499.5 µl, and 0.5 µl of library material ("M" or "R") was added to each. Each library's PCR was split into 8 tubes of 50 µl each (total volume of 400 µl) to run on the thermal cycler. Thermal cycling was performed as follows:

1. 94 °C for 2 minutes 2. 94 °C for 30 seconds 3. 55 °C for 30 seconds 4. 72 °C for 90 seconds Go to step 2

[0136] Reactions were stopped after 15 cycles because the quantity of DNA was beginning to plateau. Reactions from each library were combined to individual tubes and purified with QIAQUICK® (Qiagen, Valencia, CA) columns as per manufacturer's recommendations for PCR product purification. Resuspension was in 100 µl of Buffer EB. It was decided this was too high a volume for the next step, so the samples were ethanol precipitated (as in the P:C:P protocol except no phenol-extraction was done), washed and resuspended in 10 µl of TE.

PAGE Size-Selection of Paired-Tag Library

[0137] Reactions were run on a 6% PAGE gel. The 'final' library bands (sharp bands at approximately 135 base pairs) were cut, eluted and purified as previously. This was the last purification step—it was therefore critical to try and get as tight of a gel- purification as possible to minimize contamination from any non-library molecules that might be present. The PAGE gel was run with no ladders, as these molecules can also be frequent contaminants, and using a razor blade in a guillotine-type motion, rather than a scalpel.

[0138] After P:C:P purification with overnight precipitation, recovered library material was resuspended in 10 µl of TE. To quantify the library, a 6% PAGE gel with appropriate markers was run (Figure 16). Based on relative intensities of library and ladder bands, concentrations were estimated at 2 ng/ µl, resulting in approximately 9*2 = 18 ng of each library remaining. If the library was 135 bp, then the

concentration was approximately 23 nM. The libraries were diluted in TE to various levels for use in emulsion PCR, and both the original libraries and their dilutions were stored at -20 °C.

Library Validation Via Cloning and Sanger Sequencing

[0139] To validate the expectation that library tags would be E. coli derived and paired, "R" library fragments were cloned with the Invitrogen TOPO™-4 kit (Invitrogen, Carlsbad, CA) followed by PCR using M13F/M13R, and Sanger sequencing (single read per clone). Although 96 PCR products were submitted for sequencing, 20 of these came back as either garbage reads, or vector- or contaminant related. The remaining 76 inserts appeared be appropriately flanked by the RDV2 and FDV2 segments, as expected. Of these 76: one was a 6 bp insert (TTATCA); one was an E. coli genomic fragment (65 bp in length; 63/63 100% match to E. coli MG1655 genome on BLAST); one was an E. coli genomic fragment (70 bp in length; 69/69 100% match to E. coli MG1655 genome on BLAST); one contained the RDV2 primer flanked by ~27 bp with no significant matches in the NCBI data-base; and Seventy- two contained two tags separated by the T30 segment, as expected.

[0140] Of these 72, the tag lengths had the following distribution: 1 tag was 9 bp; 1 tag was 11 bp; 2 tags were 13 bp; 1 tag was 14 bp; 1 tag was 15 bp; 2 tag was 16 bp; 73 tags were 17 bp; 62 tags were 18 bp; and 1 tag was 22 bp. In terms of pairing, tags matched the E. coli genome as follows: 4 were situations where one or neither tag had any perfect matches to the E. coli genome (likely due to sequencing errors or non-canonical sequence); 1 was "unpaired" in that tags both matched unique locations but did not appear to originate from the same genomic regions; and 67 were matched the E. coli genome as paired tags (identically oriented with inter-tag distance falling within expected constraints). For these 67 paired tags, the distance distribution of the paired tags was 951 +/- 90 bp. The minimum distance was 729 bp and the maximum distance was 1162 bp.

[0141] Thus, the pairing rate for the 68 reads was 67/68, i.e., approximately 98.5%. A minimal estimate of the fraction of emulsion-PCR-amplifiable molecules in the library that represented paired E. coli tags with a T30

segment separating them was 67/76, i.e., 88%. The actual fraction could be slightly higher if the four reads where one or both tags were unmatchable actually did represent paired reads that were not matchable due to Sanger sequencing enors or differences between the "R" strain and the canonical genome sequence reference.

[0142] Although the sample size was small (n = 72), deviations from 25/25/25/25 frequencies were observed in the tag sequences that may be significant trends (set forth in Table 3). The numbers in the first column represent the tag base position relative to either its junction with one or the other primer or with the T30 segment. The strand from which the base frequencies were tabulated is such that these frequencies are what one would expect to see if sequencing by extension from the primer or T30 segment (5' to 3'). Numbers in parentheses are the actual count-numbers (as opposed to frequencies).

PRIMER/TAG JUNCTION A G C T +1 0.315(45) 0.315 (45) 0.154(22) 0.217(31) +2 0.340(49) 0.104(15) 0.208 (30) 0.347(50) +3 0.292(42) 0.208(30) 0.146(21) 0.354 (51) +4 0.229(33) 0.250(36) 0.264(38) 0.257(37) +5 0.333(48) 0.194(28) 0.243(35) 0.229(33) T30/TAG JUNCTION A G C T +1 0.299(43) 0.194(28) 0.326(47) 0.181(26) +2 0.299(43) 0.319(46) 0.146(21) 0.236(34) +3 0.312(45) 0.222(32) 0.222(32) 0.243(35) +4 0.188 (27) 0.236(34) 0.264(38) 0.312(45) +5 0.252(36) 0.196 (28) 0.273(39) 0.280(40) Table 3. EXAMPLE VII Parallel Sequencing on Oligonucleotides Coupled to Beads

[0143] The results of an experiment in which 5 to 8 base-pair reads were obtained on oligonucleotides coupled via biotin to streptavidin-coated superparamagnetic beads (8.8 micron in diameter) are set forth in Figure 6. The beads were immobilized in acrylamide. The beads set forth in Figure 6 enable one of skill in the art to sequence 10,000 times more features per unit area on a single slide than using previously known methods.

[0144] Towards further miniaturization, experiments were performed to generate short sequencing reads (4 base pairs per template) on 1 -micron beads (Figure 7). These results approach the resolution of "one sequence-read per pixel," as each bead was only represented by approximately 1 to 4 active pixels. At

this demonstrated bead size and density, sequencing reads can be obtained from over 30 million independent beads per 1 inch by 3 inch microscope slide.

EXAMPLE VIII Monolayers of Microspheres

[0145] A variation of the protocol set forth in Example III utilizes affinity in the self- assembled monolayers to provide for enrichment of the amplified beads. This avoids problems associated with current protocols for amplification. In such protocols, dilute target DNA is used to avoid double amplicons per bead and thus the number of beads with zero target DNA is approximated by a Poisson distribution.

[0146] The present invention also includes the use of similar emulsions to form sealed chambers (to limit spread of amplification) "in situ" on patterns of nucleic acids on slides (e.g., RNA in tissue sections, microanays, stretched chromosomes). The use of an immobilized emulsion will also help constrain, stabilize and make more uniform distribution of emulsion droplet sizes.

[0147] To obtain hundreds of millions to billions of independent sequencing reads per 1 inch by 3 inch microscope slide, the same 1 -micron beads can be packed at high densities. Self-organizing monolayers (SOM) of superparamagnetic beads were generated (Figure 8). Multilayers may also be constructed using transverse magnetic fields. Obtaining 30 base-pair reads from a slide perfectly coated with a monolayer of 1 micron beads would yield -56 billion bases of sequence.

EXAMPLE IX Amplification of Nucleic Acids on Particles without Emulsions

[0148] Without intending to be bound by theory, reducing feature size of polonies to micron or submicron sizes will improve the throughput at least 1000 fold. This can be achieved by restricting one amplification primer to polymer beads (such as lμm paramagnetic beads, available from Dynal, Oslo, Norway), polymer networks, or polymer nano structures. Thus, the polony feature size is determined by the size of beads or polymeric matrix, and therefore is independent of the pore size and/or cross- linking content of polymeric matrix (acrylamide gel), and the length of target amplicons.

Similar to conventional polony PCR, the actual amplification step for polony beads is carried out in a polymeric matrix (e.g. acrylamide gel), which helps to retain the location of individual beads and partially restrict the diffusion of amplicon. One important component of the present invention is the inclusion of additive (such as cationic lipid, polyamine, polycation, etc), which forms in-gel voids, e.g., micelles or aggregates, surrounding polymer beads and allows efficient amplification in a solid phase (Figures 9A-9D). The main advantage of this improvement is differential partitioning of target amplicon and the other amplification primer to the vicinity of the beads, which increases amplification efficiency and/or proximity to the bead. It avoids the necessity of forming emulsions and hence allows for simpler multiplex amplification and even the preservation of the 2D or 3D arrangement of the initial set of molecules to be amplified.

[0149] As an alternative, the second amplification primer can be immobilized to polymeric matrix (e.g., by acrydite modification or other crosslinidng methods). Thus, little or no free second primer is required for amplification. This will eliminate crossover between beads during amplification and therefore enforce clonality. In addition to polymeric matrix (e.g., acrylamide gel), the formation of micelle or aggregate may further restrict the diffusion of free amplification products, if there are any (i.e., if there is residual free second amplification primer, or a free second primer is used during amplification). The mobility of a polyanion (e.g. DNA) within a polycationic layer (or network) can be fast relative to free diffusion away from the polycation and hence the reaction can be close to exponential until the cation layer is filled and then nearly come to a stop, further enhancing the sharpness of the subsequent images.

[0150] Figures 9A-9D depict copies of one amplification primer pre-coupled to 1-μm Dyna beads through biotin-streptavidin interaction. The other amplification primer is acrydite modified to permit it to be incorporated into a polyacrylamide gel matrix upon polymerization (left). Two primers, along with a nucleic acid amplification template of interest, were assembled in acrylamide matrix. PCR amplification was performed followed by denaturation and hybridization of Cy3-labelled probe. Few amplification products could be detected (right). In contrast to (left), when cationic lipid was added to the assembled mix and incubated for 15 minutes, the

amplification efficiency was greatly enhanced as demonstrated by brighter signals.

EXAMPLE X Analysis of DNA Modifications Using Polony Sequencing

[0151] The attachment of a methyl (CH3) group to the 5' carbon of the cytosine residue in a CpG dinucleotide (which is referred to as an epigenetic modification) constitutes an important mechanism for controlling cellular gene expression. Quantitative knowledge on the dynamics of genomics methylation patterns will have great impact on biology, including developmental control and pathological states. Genome-wide detection of abnormal DNA methylation relies on restriction digestion with methylation-specific enzyme, followed by either differential display (MCA-RDA; Ueki et al. (2001) Cancer Res. 61:8540, incorporated herein by reference in its entirety for all purposes) or microanay analysis (DMH (Yan et al. (2000) Clin. Cancer Res. 6:1432), ECIST (Shi et al. (2002) Cancer Res. 62:3214, incorporated herein by reference in its entirety for all purposes). Though candidate differential methylation loci can be identified, none of them can monitor the combinatorial methylation events that are typical for any given CpG island. The only technology that can query the relationship of multiple methylations in cis is methylation-specific sequencing. However, the cloning of individual bisulfite-converted genomic fragments and performance of conventional sequencing procedures make that approach labor-intensive and low-throughput.
[0152] In one embodiment, the present invention is directed to adapting methylation-specific sequencing to the polony platform, imparting several advantages over the traditional methylation sequencing method. (1) Bisulfite-converted DNA fragments can be directly amplified in a solid or semi-solid phase (in this case, acrylamide), thus eliminating the requirement of laborious cloning step and potential bias introduced by differential cloning efficiency. (2) The detection of methyl-dC can be achieved in a highly-parallel manner such that thousands to millions of single molecules are queried simultaneously. Two approaches, repetitive probing and FISSEQ (Zhu et al. (2003) Science 301 :836; Mitra et al. (2003) Analyt. Biochem. 320:55, incorporated herein by reference in their entirety for all purposes) can be applied to monitor differential methylation events. Whereas repetitive probing is suitable to study a handful of known loci, FISSEQ

certainly has advantages for whole genome studies (see (4), below). (3) Because the polony platform is a single-molecule-based technology, little input material is required. Thus single cells and/or chromosomes may be utilized. Individual cells and/or chromosomes (e.g., a chromosome spread) can be embedded in polymer matrix or gel (e.g., polyacrylamide), followed by a solid-phase bisulfite conversion and amplification. This will eliminate DNA purification steps, and avoid the possible sample loss and contamination during sample preparation. (4) As an extension of (3), multiplex PCR and/or whole genome amplification (with, e.g., Phi29) can be performed with bisulfite-converted cells and/or chromosomes. This allows the determination of methylation status of the entire genomic DNA at a single cell and/or single chromosome level.

EXAMPLE XI Fluorescence Resonance Energy Transfer (FRET)

[0153] By using two different fluorophore-labelled dNTPs which form a FRET donor- acceptor pair for a single FISSEQ step, e.g. Cy3-dATP donor plus Cy5-dATP acceptor, one can distinguish among zero, one and two identical bases in a homopolymer run. One excites first with the excitation wavelength of the donor dNTP and looks for emission at both the donor emission wavelength and the acceptor wavelength. One then excites with the acceptor excitation wavelength and looks for emission at the acceptor emission wavelength. One could reverse the order and first excite the acceptor to minimize photo-bleaching of the acceptor fluorophore as a result of FRET.

[0154] For example, if there are zero dA residues to be incorporated in a given cycle, then one will observe no (or low) signal. One dA will result in detection of the corresponding emission wavelength for each fluorophore excited as a result of FRET occurring. Two dAs will result in emission at the acceptor wavelength when excited at the donor excitation wavelength, as a result of FRET occurring. Three or more dAs can be determined as quantitative variants of the 2 dAs case.

EXAMPLE XII Libraries Libraries For Rolling Circles/ Amplification [0155] One method of generating a large library of circles directly from genomic DNA would be to use random N-mers at the ends and no internal tag for microanay hybridization such as:

NNNNNN... (common primer I)..(cut-site)..(common primer 2)..NNNNNN

These N-mers may be used to hybridize randomly to DNA. Unlike the systems in the art, the N-mers could sit down at some distance from one another rather than with a single-base separation. This distance would be constrained by the length of the oligonucleotides utilized.

[0156] Polymerases, ligases and dNTPs may be used to gap-fill and ligate to generate a "padlocked" probe. Rather than releasing the pad-lock by recircularization, the genomic DNA may be digested with a restriction enzyme (with site not present in oligonucleotide) and exonuclease, leaving only circularized probes that contain a stretch of genomic bases. The circles could be used for the rolling circle method. Alternatively if the the emulsion method is being used, the padlock could be released via cleavage of the circle, leaving with genomic sequence flanked by common primers. Jumping Libraries [0157] This example sets forth one method by which a polony-ready library may be generated via phi29 rolling-circle amplimers. To generated a jumping library using MmeI (or EcoP15I) the following steps could be performed:

(1) Complete or partial cleavage of the genome with an endonuclease (e.g. NlaIII (CATG<^>) or CviJI** (NG<^>CN) or even DNase; the 4 bp overhang may be the cleanest place to start);

(2) Optional size selection (the more precise the size selection, then the more precision in the assembly of sequences in complex repeat regions);

(3) Ligate a double-tag double-MmeI adapter, such that circles are formed;

(4) Optional size selection;

(5) Cut with MmeI, blunt and circularize again; and

(6) The circles can be initiated into rolling circles by nicking randomly, by a sequence specific nicking enzyme, by strand invasion with DNA, PNA, RNA primers, or by RNA polymerase or Primase initiation.

[0158] This method will yield up to 27 base pairs of sequence information from each end, in addition to the information on the initial endonuclease sites (62 base pairs total for NlaIII) and more importantly allow assembly across longer de novo sequences.

[0159] Non-limiting examples of special restriction enzymes for used with non-rolling-circle libraries and non-rolling-circle libraries of the invention are as follows: SAGE: BsmFI, available from New England Biolabs (NEB), Beverly, MA (Velculescu et al. (1995) Science 5235:484, incorporated herein by reference in its entirety for all purposes); LongSage: MmeI, available from NEB (Sal a et al. (2002) Nat Biotechnol. 20:508, incorporated herein by reference in its entirety for all purposes); CAGE: MmeI, available from NEB (Shiraki et al. (2003) Proc. Natl. Acad. Sci. U.S.A. 100:15776, incorporated herein by reference in its entirety for all purposes); SuperSage: EcoP15I (Matsumura et al. (2003) Proc. Natl. Acad. Sci. U.S.A. 100:15718, incorporated herein by reference in its entirety for all purposes); SAGS: BsaXI, available from NEB (Torstein and Meyerson, Serial Analysis of Genome Subsets). BsmFI and EcoP15I have 5' overhangs, while MmeI has a 3' overhang. For MmeI the 3' 2 base pair information is preserved by ligating complementary pairs); and rapid shotgun cloning utilizing the two base recognition endonuclease CviJI (Fitzgerald et al. (1992) Nucleic Acids Res. 20:3753, incorporated herein by reference in its entirety for all purposes).

EXAMPLE XIII Extensions of Excluded Volume

[0160] This example provides a means where clonal amplified beads can be generated with a high amplified bead : reaction volume ratio, without requiring PCR. One polymer (or particle) interacting with another can prevent further interactions with additional nearby particles. Template is pre-amplified by rolling circle amplification to yield long, voluminous concatemers. Thus, with excess of such template (Figure 10) one can saturate all beads in the tube without multiple concatamer binding events per bead.

[0161] For example, with a library of 60-mer nicked circles, each having 30 base pairs of universal tags and 30 base pairs of insert sequence from a genomic library. A phi29 rolling circle (without any primers other than the nick) is used for about an hour until there are close to

l<4>tandem single stranded repeats hanging off of the original circles. At low salt this will have a length of about 6<5>bp (=200 microns extended) and an excluded volume of a few cubic microns. If a dilute solution of template interacts with 1 -micron beads that cany complementary universal tags, then the first template to contact a bead will be bound at numerous points and exclude other templates from binding (set forth in Figure 10). One can wash the beads at low salt (to reduce aggregation with other templates), then restart the rolling circle amplification (with only the bead bound primers). Without intending to be bound by theory, this should produce sufficient of single stranded template for sequencing (this may even produce more would be obtained by PCR since each primer will be bound to many tandem single-stranded repeats). This method should obviate the need for thermal cycling or bead sorting. EXAMPLE XIV References Each reference is incorporated herein by reference in its entirety for all purposes.

Aach and Church (2004) J Theor. Biol. 228:31

Menitt et al. (2003) Nucleic Acids Res. 31 :e84

Mikkilineni et al. (2003) Biotechnol. Bioeng. 86:117

Mitra et al. (2003) Anal Biochem. 320(1):55

Zhu et al. (2003) Science 301(5634):836

Butz et al. (2003) BMC Biotechnol. 31:11

Denkov et al. (1992) Langmuir 8:3183

Dressman et al. (2003) Proc. Natl. Acad. Sci. USA 100:8817

Brenner et al. (2000) N t. Biotechnol 18:630

Mitra et al. (2003) Proc. Natl. Acad. Sci. USA 100:5926

Mitra et al. (2003) Anal. Biochem. 320:55

Zhu et al. (2003) Science 301:836

Espacenet

Claims: JP2007526772 (A) — 2007-09-20

POLONY FLUORESCENT IN SITU SEQUENCING BEADS

Claims not available for JP2007526772 (A)
Claims of corresponding document: WO2005082098 (A2)

A high quality text as facsimile in your desired language may be available amongst the following family members:

CA2557841 (A1) US2007087362 (A1) WO2005082098 (A2)

- Original claims
- Claims tree

The EPO does not accept any responsibility for the accuracy of data and information originating from other authorities than the EPO; in particular, the EPO does not guarantee that they are complete, up-to-date or fit for specific purposes.

What is claimed is:
1. An array comprising a plurality of beads, wherein an individual bead has a population of substantially identical nucleic acid sequences attached thereto that differs in sequence from the population of substantially identical nucleic acid sequences attached to other individual beads, and wherein the plurality of beads is immobilized in a semi-solid medium to form an array.
2. The array of claim 1, wherein the semi-solid medium is attached to a solid support.
3. The anay of claim 2, wherein the support is a microscope slide or a flow cell.
4. The array of claim 1, wherein the individual beads comprise two, three or four different populations of substantially identical nucleic acid sequences attached thereto .
5. The array of claim 1, wherein at least 40% of the individual beads include the population of substantially identical nucleic acid sequences attached thereto.
6. The anay of claim 1, wherein the beads are immobilized as a monolayer.
7. The array of claim 1, wherein the semi-solid medium has x, y and z axes, and the plurality of beads is randomly arranged relative to the x and y axes.
8. The anay of claim 1, wherein the semi-sold medium has top and bottom surfaces and the plurality of beads

is immobilized near the top surface.
9. The anay of claim 2, wherein the semi-sold medium has top and bottom surfaces and the solid support is attached to the bottom surface.
10. The anay of claim 1, wherein the semi-solid medium is selected from trie group consisting of: polyacrylamide, cellulose, polyamide, cross-linked agarose, cross-linked dextran and cross-linked polyethylene glycol.
11. The anay of claim 1, wherein the plurality of beads comprises a plurality of clonal beads.
12. The anay of claim 1, wherein the plurality of beads comprises a library.
13. A method of producing an anay comprising the steps of: a) providing a plurality of beads wherein an individual bead has a population of substantially identical nucleic acid sequences attached thereto that differs in sequence from the population of substantially identical nucleic acid sequences attached to other individual beads; and b) immobilizing the beads in a semi-solid medium to form an array.
14. The method of claim 13, further comprising attaching the semi-solid medium to a solid support in step b).
15. The method of claim 14, wherein the support is a microscope slide or a flow cell.
16. The method of claim 13, wherein the plurality of beads comprises two, three or four different populations of substantially identical nucleic acid sequences attached thereto.
17. The method of claim 13, wherein at least 40% of the individual beads include the population of substantially identical nucleic acid sequences attached thereto.
18. The method of claim 13, wherein the beads are immobilized as a monolayer.
19. The method of claim 13, wherein the semi-solid medium has x, y and z axes, and the plurality of beads is randomly ananged relative to the x and y axes.
20. The method of claim 13, wherein the semi-sold medium has top and bottom surfaces and the plurality of beads is immobilized near the top surface.
21. The method of claim 14, wherein the semi-sold medium has top and bottom surfaces and the solid support is attached to the bottom surface.
22. The method of claim 13, wherein the semi-solid medium is selected from the group consisting of: polyacrylamide, cellulose, polyamide, cross-linked agarose, cross-linked dextran and cross-linked polyethylene glycol.
23. The method of claim 13, wherein the plurality of

beads comprises a plurality of clonal beads.
24. The method of claim 13, wherein the plurality of beads comprises a library.
25. A method of producing an anay comprising the steps of: a) providing a plurality of beads wherein an individual bead has a population of substantially identical nucleic acid sequences attached thereto that differs in sequence from the population of substantially identical nucleic acid sequences attached to other individual beads; b) immobilizing the beads in a semi-solid medium to form an anay; and c) amplifying the population of substantially identical nucleic acid sequences to form a plurality of beads having an amplified population of substantially identical nucleic acid sequences attached thereto.
26. The method of claim 25, wherein the semi-solid medium includes an amplification primer.
27. The method of claim 25, wherein the semi-solid medium includes an additive that forms voids in the semi-solid medium.
28. The method of claim 27, wherein the additive is selected from the group consisting of: cationic lipid, polyamine and polycation.
29. A method of producing an anay comprising the steps of: a) providing a plurality of beads wherein an individual bead has a population of substantially identical nucleic acid sequences attached thereto that differs in sequence from the population of substantially identical nucleic acid sequences attached to other individual beads; b) amplifying the population of substantially identical nucleic acid sequences to form a plurality of immobilized beads having an amplified population of substantially identical nucleic acid sequences attached thereto; and c) immobilizing the beads in a semi-solid medium to form an array.
30. The method of claim 29, wherein the amplifying is performed by emulsion PCR.
31. A method of producing an anay comprising the steps of: a) providing a plurality of beads wherein an individual bead has a population of substantially identical nucleic acid sequences attached thereto that differs in sequence from the population of substantially identical nucleic acid sequences attached to other individual beads; b) amplifying the population of substantially identical nucleic acid sequences to form a plurality of beads having an amplified population of substantially identical nucleic acid sequences attached thereto; c) enriching the plurality of beads having an amplified population of substantially identical nucleic

acid sequences attached thereto to form an emiched population of beads; and d) immobilizing the beads in a semi-solid medium to form an anay.
32. A method for enriching a population of beads having a first nucleic acid sequence attached thereto comprising the steps of: a) providing a population of beads wherein an individual bead has a population of substantially identical nucleic acid sequences attached thereto that differs in sequence from the population of substantially identical nucleic acid sequences attached to other individual beads; b) contacting the population of beads with a second nucleic acid sequence that is complementary to the first nucleic acid sequence; c) incubating the population of beads and the second nucleic acid sequence together such that hybridization occurs to form a population of hybridized beads and a population of unhybridized beads; and d) separating the population of hybridized beads from the population unhybridized beads.
33. The method of claim 32, wherein the second nucleic acid is immobilized on a capture bead.
34. The method of claim 32, wherein the population of hybridized beads are separated from the population of unhybridized beads by density or affinity.
35. A kit containing an anay comprising a plurality of beads immobilized in a semi- solid medium, wherein the plurality of beads includes an individual bead having a population of substantially identical amino acids attached thereto that differs in sequence from the population of substantially identical nucleic acid sequences attached to other individual beads.
36. The method of claim 1, wherein the substantially identical nucleic acid sequences are primers.
37. The method of claim 1, wherein the substantially identical nucleic acid sequences are amplified nucleic acid sequences.
38. The method of claim 13, wherein the substantially identical nucleic acid sequences are primers.
39. The method of claim 13, wherein the substantially identical nucleic acid sequences are amplified nucleic acid sequences.
40. The method of claim 25, wherein the substantially identical nucleic acid sequences are primers.
41. The method of claim 25, wherein the substantially identical nucleic acid sequences are amplified nucleic acid sequences.
42. The method of claim 29, wherein the substantially identical nucleic acid sequences are primers.
43. The method of claim 29, wherein the substantially

identical nucleic acid sequences are amplified nucleic acid sequences.
44. The method of claim 31, wherein the substantially identical nucleic acid sequences are primers.
45. The method of claim 31, wherein the substantially identical nucleic acid sequences are amplified nucleic acid sequences.
46. The method of claim 32, wherein the substantially identical nucleic acid sequences are primers.
47. The method of claim 32, wherein the substantially identical nucleic acid sequences are amplified nucleic acid sequences.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

NOTICE OF ALLOWANCE AND FEE(S) DUE

22907 7590 06/21/2022

BANNER & WITCOFF, LTD.
1100 13th STREET, N.W.
SUITE 1200
WASHINGTON, DC 20005-4051

EXAMINER
THOMAS, DAVID C

ART UNIT	PAPER NUMBER
1637	

DATE MAILED: 06/21/2022

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
17/395,534	08/06/2021	Evan R. Daugharthy	010498.01437\US	5147

TITLE OF INVENTION: Method and Apparatus for Volumetric Imaging

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	UNDISCOUNTED	$1200	$0.00	$0.00	$1200	09/21/2022

THE APPLICATION IDENTIFIED ABOVE HAS BEEN EXAMINED AND IS ALLOWED FOR ISSUANCE AS A PATENT. PROSECUTION ON THE MERITS IS CLOSED. THIS NOTICE OF ALLOWANCE IS NOT A GRANT OF PATENT RIGHTS. THIS APPLICATION IS SUBJECT TO WITHDRAWAL FROM ISSUE AT THE INITIATIVE OF THE OFFICE OR UPON PETITION BY THE APPLICANT. SEE 37 CFR 1.313 AND MPEP 1308.

THE ISSUE FEE AND PUBLICATION FEE (IF REQUIRED) MUST BE PAID WITHIN THREE MONTHS FROM THE MAILING DATE OF THIS NOTICE OR THIS APPLICATION SHALL BE REGARDED AS ABANDONED. THIS STATUTORY PERIOD CANNOT BE EXTENDED. SEE 35 U.S.C. 151. THE ISSUE FEE DUE INDICATED ABOVE DOES NOT REFLECT A CREDIT FOR ANY PREVIOUSLY PAID ISSUE FEE IN THIS APPLICATION. IF AN ISSUE FEE HAS PREVIOUSLY BEEN PAID IN THIS APPLICATION (AS SHOWN ABOVE), THE RETURN OF PART B OF THIS FORM WILL BE CONSIDERED A REQUEST TO REAPPLY THE PREVIOUSLY PAID ISSUE FEE TOWARD THE ISSUE FEE NOW DUE.

HOW TO REPLY TO THIS NOTICE:

I. Review the ENTITY STATUS shown above. If the ENTITY STATUS is shown as SMALL or MICRO, verify whether entitlement to that entity status still applies.

If the ENTITY STATUS is the same as shown above, pay the TOTAL FEE(S) DUE shown above.

If the ENTITY STATUS is changed from that shown above, on PART B - FEE(S) TRANSMITTAL, complete section number 5 titled "Change in Entity Status (from status indicated above)".

For purposes of this notice, small entity fees are 1/2 the amount of undiscounted fees, and micro entity fees are 1/2 the amount of small entity fees.

II. PART B - FEE(S) TRANSMITTAL, or its equivalent, must be completed and returned to the United States Patent and Trademark Office (USPTO) with your ISSUE FEE and PUBLICATION FEE (if required). If you are charging the fee(s) to your deposit account, section "4b" of Part B - Fee(s) Transmittal should be completed. If an equivalent of Part B is filed, a request to reapply a previously paid issue fee must be clearly made, and delays in processing may occur due to the difficulty in recognizing the paper as an equivalent of Part B.

III. All communications regarding this application must give the application number. Please direct all communications prior to issuance to Mail Stop ISSUE FEE unless advised to the contrary.

IMPORTANT REMINDER: Maintenance fees are due in utility patents issuing on applications filed on or after Dec. 12, 1980. It is patentee's responsibility to ensure timely payment of maintenance fees when due. More information is available at www.uspto.gov/PatentMaintenanceFees.

PART B - FEE(S) TRANSMITTAL

Complete and send this form, together with applicable fee(s), by mail or fax, or via EFS-Web.

By mail, send to: Mail Stop ISSUE FEE
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

By fax, send to: (571)-273-2885

INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 5 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Use Block 1 for any change of address)

22907 7590 06/21/2022
BANNER & WITCOFF, LTD.
1100 13th STREET, N.W.
SUITE 1200
WASHINGTON, DC 20005-4051

Note: A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

Certificate of Mailing or Transmission

I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Mail Stop ISSUE FEE address above, or being transmitted to the USPTO via EFS-Web or by facsimile to (571) 273-2885, on the date below.

(Typed or printed name)
(Signature)
(Date)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
17/395,534	08/06/2021	Evan R. Daugharthy	010498.01437\US	5147

TITLE OF INVENTION: Method and Apparatus for Volumetric Imaging

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	UNDISCOUNTED	$1200	$0.00	$0.00	$1200	09/21/2022

EXAMINER	ART UNIT	CLASS-SUBCLASS
THOMAS, DAVID C	1637	435-006120

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363).

❑ Change of correspondence address (or Change of Correspondence Address form PTO/AIA/122 or PTO/SB/122) attached.

❑ "Fee Address" indication (or "Fee Address" Indication form PTO/AIA/47 or PTO/SB/47; Rev 03-02 or more recent) attached. **Use of a Customer Number is required.**

2. For printing on the patent front page, list
(1) The names of up to 3 registered patent attorneys or agents OR, alternatively,
(2) The name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed.

1 _____
2 _____
3 _____

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. If an assignee is identified below, the document must have been previously recorded, or filed for recordation, as set forth in 37 CFR 3.11 and 37 CFR 3.81(a). Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE (B) RESIDENCE: (CITY and STATE OR COUNTRY)

Please check the appropriate assignee category or categories (will not be printed on the patent) : ❑ Individual ❑ Corporation or other private group entity ❑ Government

4a. Fees submitted: ❑ Issue Fee ❑ Publication Fee (if required) ❑ Advance Order - # of Copies _____

4b. Method of Payment: *(Please first reapply any previously paid fee shown above)*

❑ Electronic Payment via EFS-Web ❑ Enclosed check ❑ Non-electronic payment by credit card (Attach form PTO-2038)

❑ The Director is hereby authorized to charge the required fee(s), any deficiency, or credit any overpayment to Deposit Account No. _____

5. **Change in Entity Status** (from status indicated above)

❑ Applicant certifying micro entity status. See 37 CFR 1.29

❑ Applicant asserting small entity status. See 37 CFR 1.27

❑ Applicant changing to regular undiscounted fee status.

NOTE: Absent a valid certification of Micro Entity Status (see forms PTO/SB/15A and 15B), issue fee payment in the micro entity amount will not be accepted at the risk of application abandonment.
NOTE: If the application was previously under micro entity status, checking this box will be taken to be a notification of loss of entitlement to micro entity status.
NOTE: Checking this box will be taken to be a notification of loss of entitlement to small or micro entity status, as applicable.

NOTE: This form must be signed in accordance with 37 CFR 1.31 and 1.33. See 37 CFR 1.4 for signature requirements and certifications.

Authorized Signature _____ Date _____

Typed or printed name _____ Registration No. _____

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
17/395,534	08/06/2021	Evan R. Daugharthy	010498.01437\US	5147

22907 7590 06/21/2022

BANNER & WITCOFF, LTD.
1100 13th STREET, N.W.
SUITE 1200
WASHINGTON, DC 20005-4051

EXAMINER	
THOMAS, DAVID C	

ART UNIT	PAPER NUMBER
1637	

DATE MAILED: 06/21/2022

Determination of Patent Term Adjustment under 35 U.S.C. 154 (b)
(Applications filed on or after May 29, 2000)

The Office has discontinued providing a Patent Term Adjustment (PTA) calculation with the Notice of Allowance.

Section 1(h)(2) of the AIA Technical Corrections Act amended 35 U.S.C. 154(b)(3)(B)(i) to eliminate the requirement that the Office provide a patent term adjustment determination with the notice of allowance. See Revisions to Patent Term Adjustment, 78 Fed. Reg. 19416, 19417 (Apr. 1, 2013). Therefore, the Office is no longer providing an initial patent term adjustment determination with the notice of allowance. The Office will continue to provide a patent term adjustment determination with the Issue Notification Letter that is mailed to applicant approximately three weeks prior to the issue date of the patent, and will include the patent term adjustment on the patent. Any request for reconsideration of the patent term adjustment determination (or reinstatement of patent term adjustment) should follow the process outlined in 37 CFR 1.705.

Any questions regarding the Patent Term Extension or Adjustment determination should be directed to the Office of Patent Legal Administration at (571)-272-7702. Questions relating to issue and publication fee payments should be directed to the Customer Service Center of the Office of Patent Publication at 1-(888)-786-0101 or (571)-272-4200.

OMB Clearance and PRA Burden Statement for PTOL-85 Part B

The Paperwork Reduction Act (PRA) of 1995 requires Federal agencies to obtain Office of Management and Budget approval before requesting most types of information from the public. When OMB approves an agency request to collect information from the public, OMB (i) provides a valid OMB Control Number and expiration date for the agency to display on the instrument that will be used to collect the information and (ii) requires the agency to inform the public about the OMB Control Number's legal significance in accordance with 5 CFR 1320.5(b).

The information collected by PTOL-85 Part B is required by 37 CFR 1.311. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 30 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, Virginia 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, Virginia 22313-1450. Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether disclosure of these records is required by the Freedom of Information Act.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspection or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

Notice of Allowability	Application No. 17/395,534	Applicant(s) Daugharthy et al.	
	Examiner DAVID C THOMAS	Art Unit 1637	AIA (FITF) Status Yes

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. ☑ This communication is responsive to the submission of 03 June 2022.
   ☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2. ☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

3. ☑ The allowed claim(s) is/are 9-17 and 19-45 . As a result of the allowed claim(s), you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information , please see **http://www.uspto.gov/patents/init_events/pph/index.jsp** or send an inquiry to **PPHfeedback@uspto.gov.**

4. ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

   Certified copies:

   a) ☐ All b) ☐ Some* c) ☐ None of the:

   1. ☐ Certified copies of the priority documents have been received.
   2. ☐ Certified copies of the priority documents have been received in Application No. _____ .
   3. ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a)).

   * Certified copies not received: _____ .

Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file a reply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application.
THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.

5. ☐ CORRECTED DRAWINGS (as "replacement sheets") must be submitted.
   ☐ including changes required by the attached Examiner's Amendment / Comment or in the Office action of Paper No./Mail Date _____ .

   Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).

6. ☐ DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

Attachment(s)

1. ☐ Notice of References Cited (PTO-892)
2. ☑ Information Disclosure Statements (PTO/SB/08), Paper No./Mail Date 06/07/2022, 03/01/2022.
3. ☐ Examiner's Comment Regarding Requirement for Deposit of Biological Material _____.
4. ☐ Interview Summary (PTO-413), Paper No./Mail Date. _____.
5. ☐ Examiner's Amendment/Comment
6. ☑ Examiner's Statement of Reasons for Allowance
7. ☐ Other _____.

/DAVID C THOMAS/
Primary Examiner, Art Unit 1637

REASONS FOR ALLOWANCE

1. The following is an examiner's statement of reasons for allowance: In view of the claim amendments and attorney's arguments, the claimed invention is novel and unobvious over the closest prior art of Keller et al. (U.S. Patent Pub. No. 2015/0098126) and the secondary references of Callahan et al. (U.S. Patent Pub. No. 2011/0216953) and Frisen et al. (U.S. Patent Pub. No. 2019/0203275), and no additional prior art was identified that teaches or suggests the systems comprising a stage, a fluidics system, an optical system, a Z-axis motion control system and a computing system as currently claimed. While Keller teaches a microscope system for imaging a live biological specimen comprising a specimen chamber, a specimen holder, a perfusion system, a microscope system configured for performing volumetric imaging, and an optical scanner device (see paragraphs 26, 153-155, 106 and 159 and Figures 3 and 5A and B), the reference is silent with regards to a computing system comprising at least one computer readable storage medium having instructions that are executable by at least one processor to perform a method of determining a plurality of optically-encoded signals from a plurality of light signals obtained from a plurality of read-out cycles, the plurality of optically-encoded signals being associated with one or more nucleic acid molecules within a volume of a sample, and determining three-dimensional positional information of the one or more nucleic acid molecules using the plurality of optically-encoded signals. The secondary references of Callahan and Frisen are also silent with regard to the computing system as currently claimed.

Other references of particular interest to the currently claimed invention include Luo et al. (U.S. Patent Pub. No. 2012/0040397, cited on IDS of 08/06/2021) and Green

et al. (U.S. Patent Pub. No. 2014/0220587, cited on IDS of 08/06/2021) which teach methods or preparing and analyzing nucleic acids that are coupled to a three-dimensional matrix. However, neither reference teaches the systems as currently claimed, including the computing system configured for determining three-dimensional positional information of the one or more nucleic acid molecules using a plurality of optically-encoded signals.

Any comments considered necessary by applicant must be submitted no later than the payment of the issue fee and, to avoid processing delays, should preferably accompany an issue fee. Such admissions should be clearly labeled "Comments on Statement of Reasons for Allowance."

Correspondence

2. Any inquiry concerning this communication or earlier communications from the examiner should be directed to David C. Thomas whose telephone number is 571-272-3320 and whose fax number is 571-273-3320. The examiner can normally be reached on 5 days, 9-5:30.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Gary Benzion can be reached on 571-272-0782. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

/David C Thomas/
Primary Examiner, Art Unit 1637

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	17/395,534	Daugharthy et al.
	Examiner	**Art Unit**
	DAVID C THOMAS	1637

CPC						
Symbol					**Type**	**Version**
C12Q	/	1	/	6869	F	2013-01-01
C12Q	/	1	/	68	I	2013-01-01
C12Q	/	1	/	6841	I	2013-01-01

CPC Combination Sets								
Symbol					**Type**	**Set**	**Ranking**	**Version**
C12Q	/	1	/	6841	I	1	1	2013-01-01
C12Q	/	2543	/	101	I	1	2	2013-01-01
C12Q	/	2563	/	159	I	1	3	2013-01-01
C12Q	/	2565	/	601	I	1	4	2013-01-01
C12Q	/	1	/	6869	A	2	1	2013-01-01
C12Q	/	2543	/	101	A	2	2	2013-01-01
C12Q	/	2563	/	159	A	2	3	2013-01-01
C12Q	/	2565	/	601	A	2	4	2013-01-01

NONE (Assistant Examiner)	(Date)	**Total Claims Allowed:** 36	
/DAVID C THOMAS/ Primary Examiner, Art Unit 1637 (Primary Examiner)	10 June 2022 (Date)	O.G. Print Claim(s) 1	O.G. Print Figure None

U.S. Patent and Trademark Office — Part of Paper No.: 20220610

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	17/395,534	Daugharthy et al.
	Examiner	**Art Unit**
	DAVID C THOMAS	1637

INTERNATIONAL CLASSIFICATION				
CLAIMED				
C12Q	/	1	/	68
C12Q	/	1	/	6869
C12Q	/	1	/	6841

NON-CLAIMED				
	/		/	

US ORIGINAL CLASSIFICATION	
CLASS	**SUBCLASS**

CROSS REFERENCES(S)						
CLASS	**SUBCLASS (ONE SUBCLASS PER BLOCK)**					

NONE		**Total Claims Allowed:**	
(Assistant Examiner)	(Date)	36	
/DAVID C THOMAS/ Primary Examiner, Art Unit 1637	10 June 2022	O.G. Print Claim(s)	O.G. Print Figure
(Primary Examiner)	(Date)	1	None

U.S. Patent and Trademark Office — Part of Paper No.: 20220610

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	17/395,534	Daugharthy et al.
	Examiner	**Art Unit**
	DAVID C THOMAS	1637

☑ Claims renumbered in the same order as presented by applicant ☐ CPA ☑ T.D. ☐ R.1.47

CLAIMS

Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original

NONE (Assistant Examiner)	(Date)	**Total Claims Allowed:** 36	
/DAVID C THOMAS/ Primary Examiner, Art Unit 1637 (Primary Examiner)	10 June 2022 (Date)	O.G. Print Claim(s) 1	O.G. Print Figure None

U.S. Patent and Trademark Office Part of Paper No.: 20220610

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	17/395,534	Daugharthy et al.
	Examiner	**Art Unit**
	DAVID C THOMAS	1637

✓	Rejected	-	Cancelled	N	Non-Elected	A	Appeal
=	Allowed	÷	Restricted	I	Interference	O	Objected

CLAIMS

☑ Claims renumbered in the same order as presented by applicant ☐ CPA ☑ T.D. ☐ R.1.47

CLAIM		DATE								
Final	Original	06/10/2022								
	1	-								
	2	-								
	3	-								
	4	-								
	5	-								
	6	-								
	7	-								
	8	-								
1	9	=								
2	10	=								
3	11	=								
4	12	=								
5	13	=								
6	14	=								
7	15	=								
8	16	=								
9	17	=								
	18	-								
10	19	=								
11	20	=								
12	21	=								
13	22	=								
14	23	=								
15	24	=								
16	25	=								
17	26	=								
18	27	=								
19	28	=								
20	29	=								
21	30	=								
22	31	=								
23	32	=								
24	33	=								
25	34	=								
26	35	=								
27	36	=								
28	37	=								
29	38	=								
30	39	=								
31	40	=								
32	41	=								
33	42	=								

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	17/395,534	Daugharthy et al.
	Examiner	Art Unit
	DAVID C THOMAS	1637

CLAIM		DATE								
Final	Original	06/10/2022								
34	43	=								
35	44	=								
36	45	=								

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	17/395,534	Daugharthy et al.
	Examiner	Art Unit
	DAVID C THOMAS	1637

CPC - Searched*		
Symbol	Date	Examiner
C12Q 1/68	02/25/2022	DCT
C12Q 1/6841	02/25/2022	DCT
C12Q 1/6869	02/25/2022	DCT
C12Q 2565/518	02/25/2022	DCT
C12Q 2565/601	02/25/2022	DCT

CPC Combination Sets - Searched*		
Symbol	Date	Examiner

US Classification - Searched*			
Class	Subclass	Date	Examiner

* See search history printout included with this form or the SEARCH NOTES box below to determine the scope of the search.

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	17/395,534	Daugharthy et al.
	Examiner	**Art Unit**
	DAVID C THOMAS	1637

Search Notes		
Search Notes	**Date**	**Examiner**
EAST	12/09/2021	DCT
EAST	02/25/2022	DCT
EAST	03/01/2022	DCT
Google-search terms: optical device, fluorescence microscope, objective, pump, flow cell, xyz motion, volumetric imaging, fluidics dispenser, multi-axis motion control, detectors, filters, lasers, emission, excitation	02/25/2022	DCT
PALM, inventor name search	02/25/2022	DCT
EAST-updated	06/09/2022	DCT
PALM, inventor name search-updated	06/09/2022	DCT
EAST-updated	06/13/2022	DCT

Interference Search			
US Class/CPC Symbol	**US Subclass/CPC Group**	**Date**	**Examiner**
	Claim search performed	06/09/2022	DCT

Doc code: IDS
Doc description: Information Disclosure Statement (IDS) Filed

PTO/SB/08a (01-22)
Approved for use through 05/31/2024. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	17395534
Filing Date	2021-08-06
First Named Inventor	Evan R. Daugharthy
Art Unit	1637
Examiner Name	D. C. Thomas
Attorney Docket Number	010498.01437\US

U.S.PATENTS Remove

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Patent citation information please click the Add button. Add

U.S.PATENT APPLICATION PUBLICATIONS Remove

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
/D.C.T/	1	20080180790	A1	2008-07-31	Tafas et al.	

If you wish to add additional U.S. Published Application citation information please click the Add button. Add

FOREIGN PATENT DOCUMENTS Remove

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2]	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1							

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17395534
	Filing Date	2021-08-06
	First Named Inventor	Evan R. Daugharthy
	Art Unit	1637
	Examiner Name	D. C. Thomas
	Attorney Docket Number	010498.01437\US

	1		

If you wish to add additional non-patent literature document citation information please click the Add button Add

EXAMINER SIGNATURE

Examiner Signature	/DAVID C THOMAS/	Date Considered	06/10/2022

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17395534
	Filing Date	2021-08-06
	First Named Inventor	Evan R. Daugharthy
	Art Unit	1637
	Examiner Name	D. C. Thomas
	Attorney Docket Number	010498.01437\US

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

See attached certification statement.

The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/John P. Iwanicki/	Date (YYYY-MM-DD)	2022-03-01
Name/Print	John P. Iwanicki	Registration Number	34,628

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

Doc code: IDS
Doc description: Information Disclosure Statement (IDS) Filed

PTO/SB/08a (01-22)
Approved for use through 05/31/2024. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	17395534
Filing Date	2021-08-06
First Named Inventor	Evan R. Daugharthy
Art Unit	1637
Examiner Name	D. C. Thomas
Attorney Docket Number	010498.01437\US

U.S.PATENTS Remove

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	7989166		2011-08-02	Koch et al.	
	2	9551032		2017-01-24	Landegren et al.	

If you wish to add additional U.S. Patent citation information please click the Add button. Add

U.S.PATENT APPLICATION PUBLICATIONS Remove

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	20090280559	A1	2009-11-12	McCarthy	
	2	20170067096	A1	2017-03-09	Wassie et al.	
	3	20130288249	A1	2013-10-31	Gullberg et al.	

If you wish to add additional U.S. Published Application citation information please click the Add button. Add

FOREIGN PATENT DOCUMENTS Remove

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17395534
	Filing Date	2021-08-06
	First Named Inventor	Evan R. Daugharthy
	Art Unit	1637
	Examiner Name	D. C. Thomas
	Attorney Docket Number	010498.01437\US

	1	2009-538123	JP	A	2009-11-05	Applera Corp.		☒
	2	2015-090458	JP	A	2015-05-11	Hitachi High Tech Corp		☒
	3	2007-526772	JP	A	2007-09-20			☒

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T5
	1	BIBIKOVA et al. "Quantitative Gene Expression Profiling in Formalin-Fixed, Paraffin-Embedded Tissues Using Universal Bead Arrays" American Journal of Pathology, Vol. 165, No. 5, November 2004	
	2	CAPODIECI et al. "Gene expression profiling in single cells within tissue" Nature Methods, September 14, 2005, 2(9) pp. 663-665	
	3	CONZE et al. "Single molecule analysis of combinatorial splicing" Nucleic Acids Research, June 29, 2010, Vol. 38, No. 16; e163	
	4	FEMINO et al. "Visualization of Single RNA Transcripts in Situ" Science, April 24, 1998, Vol. 280, pp. 585-90	
	5	GAVRILOVIC et al. "Automated Classification of Multicolored Rolling Circle Products in Dual-Channel Wide-Field Fluroescence Microscopy" Cytometry Part A, July 2011, 79(7), pp. 518-27	
	6	GEISS et al. "Direct multiplexed measurement of gene expression with color-coded probe pairs" Nature Biotechnology, Vol. 26, No. 3, March 2008, pp. 317-25	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17395534
	Filing Date	2021-08-06
	First Named Inventor	Evan R. Daugharthy
	Art Unit	1637
	Examiner Name	D. C. Thomas
	Attorney Docket Number	010498.01437\US

	7	GUNDERSON et al. "Decoding Randomly Ordered DNA Arrays" Genome Research, May 2004, 14(5), pp. 870-7	
	8	HAN et al. "Quantum-dot-tagged microbeads for multiplexed optical coding of biomolecules" Nature Biotechnology, July 2001, Vol. 19, No. 7, pp. 631-5	
	9	ITZKOVITZ et al. "Validating Transcripts with Probes and Imaging Technology" Nature Methods, April 2011, 8(4 Suppl): S12-S19	
	10	ITZKOVITZ et al. "Single molecule transcript counting of stem cell markers in the mouse intestine" Nat Cell Biol., November 2011, 14(1), pp. 106-114	
	11	LAGUNAVICIUS et al. "Novel application of Phi29 DNA polymerase: RNA detection and analysis in vitro and in situ by target RNA-primed RCA" RNA, May 2009, 15(5), pp. 765-771	
	12	LARSSON et al. "In situ detection and genotyping of individual mRNA molecules" Nature Methods, Vol. 7, No. 5, May 2010, pp. 395-7	
	13	LEVSKY et al. "Fluorescence in situ hybridization: past, present and future" Journal of Cell Science, July 15, 2003, 116 (Pt 14), pp. 2833-8	
	14	LEVSKY et al. "Single-Cell Gene Expression Profiling" Science, August 2, 2002, 297(5582), pp. 836-40	
	15	MAIERHOFER et al. "Multicolor Deconvolution Microscopy of Thick Biological Specimens" American Journal of Pathology, Vol. 162, No .2, February 2003, pp. 373-9	
	16	MEADE et al. "Multiplexed DNA detection using spectrally encoded porous" Analytical Chemistry, April 1, 2009, 81(7), pp. 2618-2625	
	17	RAJ et al. "Imaging individual mRNA molecules using multiple singly labeled probes" Nature Methods, October 2009, 5 (10), pp. 877-879	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17395534
	Filing Date	2021-08-06
	First Named Inventor	Evan R. Daugharthy
	Art Unit	1637
	Examiner Name	D. C. Thomas
	Attorney Docket Number	010498.01437\US

	18	SUN et al. "Composite Organic-Inorganic Nanoparticles as Raman Labels for Tissue Analysis" Nano Letters, February 2007, Vol. 7, No. 2, pp. 351-356	
	19	WAHLBY et al. "Sequential Immunofluorescence Staining and Image Analysis for Detection of Large Numbers of Antigens in Individual Cell Nuclei" Cytometry, January 1, 2002, 47(1), pp. 32-41	
	20	WEIBRECHT et al. "Simultaneous Visualization of Both Signaling Cascade Activity and End-Point Gene Expression in Single Cells" PLoS One, May 2011, Vol. 6, Issue 5, e20148	
	21	WILSON et al. "Encoded Microcarriers For High-Throughput Multiplexed Detection" Angewandte Chemie International Edition, September 18, 2006, 45(37), pp. 6104-17	
	22	ZHAO et al. "Advances of multiplex and high throughput biomolecular detection technologies based on encoding microparticles" Science China Chemistry, August 2011, Vol. 54, No. 8, pp. 1185-1201	

If you wish to add additional non-patent literature document citation information please click the Add button Add

EXAMINER SIGNATURE

Examiner Signature	/DAVID C THOMAS/	Date Considered	06/09/2022

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17395534
	Filing Date	2021-08-06
	First Named Inventor	Evan R. Daugharthy
	Art Unit	1637
	Examiner Name	D. C. Thomas
	Attorney Docket Number	010498.01437\US

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

See attached certification statement.

☒ The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/John P. Iwanicki/	Date (YYYY-MM-DD)	2022-06-06
Name/Print	John P. Iwanicki	Registration Number	34,628

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/

EAST Search History

EAST Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	1	("20080180790").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2022/06/13 10:41
S1	1	("11118220").PN.	USPAT; USOCR	OR	OFF	2021/12/09 18:40
S2	1	("10266888").PN.	USPAT; USOCR	OR	OFF	2021/12/09 18:58
S3	1	("20210363579").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2021/12/09 19:00
S4	1	("6232067").PN.	USPAT; USOCR	OR	OFF	2022/02/25 17:08
S5	1	("20070190543").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2022/02/25 17:09
S6	1	("20020029979").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2022/02/25 17:09
S7	1	("20050064435").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2022/02/25 17:10
S8	1	("20170262984").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2022/02/25 17:11
S9	1	("20200034347").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2022/02/25 17:14
S10	1	("20100049448").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2022/02/25 17:15
S11	1	("10370698").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2022/02/25 17:20
S12	1	("20030165852").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2022/02/25 17:21
S13	1	("20160002704").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2022/02/25 17:21
S14	1	("10030261").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2022/02/25 17:24
S15	1	("8431691").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2022/02/25 17:25
S16	1	("6194148").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2022/02/25 17:26
S17	1	("20110040397").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2022/02/25 17:26
S18	1	("20120040397").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2022/02/25 17:27

S19	1	("20140220587").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2022/02/25 17:28
S20	1,058	c12q 1/68	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/02/25 17:29
S21	76	c12q 1/6841	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/02/25 17:29
S22	877	c12q 1/6869	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/02/25 17:30
S23	32	c12q 2565/518	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/02/25 17:30
S24	18	c12q 2565/601	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/02/25 17:32
S25	3,940	volumetric imaging	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/02/25 17:40
S26	1	S25 and c12q 1/68	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/02/25 17:40
S27	2	S25 and c12q 1/6869	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/02/25 17:40
S28	27,090	3d imaging	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/02/25 17:41
S29	9	S28 and c12q 1/68	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/02/25 17:41
S30	17	S28 and c12q 1/6869	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/02/25 17:43

S31	0	S25 and biolgical sample	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/02/25 17:46
S32	230	S25 and biological sample	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/02/25 17:46
S33	19,040,946	S32 and device or apparatus	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/02/25 17:47
S34	221	S32 and device	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/02/25 17:48
S35	176	S34 and light source	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/02/25 17:48
S36	39	S35 and xyz	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/02/25 17:48
S37	24	S36 and pump	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/02/25 17:49
S38	13	S37 and reagent$1	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/02/25 17:49
S39	13	S38 and detector	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/02/25 17:49
S40	44	S35 and pump	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/02/25 18:23
S41	9	S40 and pump fluid	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	WITH	ON	2022/02/25 18:24
S42	5,101	S28 and biological	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/02/25 18:26

S43	95	S42 and pump fluid	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/02/25 18:27
S44	66	S43 and light source	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/02/25 18:27
S45	424	imaging biological specimen	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/02/25 18:55
S46	0	S27 and reagent pump	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/02/25 18:55
S47	2	S27 and reagent pump	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	WITH	ON	2022/02/25 18:55
S48	0	S42 and pump reagent	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/02/25 18:56
S49	47	S42 and pump reagent	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	WITH	ON	2022/02/25 18:56
S50	39	S49 and light source	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	WITH	ON	2022/02/25 18:57
S51	1	("20150098126").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2022/02/25 19:13
S52	1	("20150098126").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2022/02/25 19:13
S53	1	("20110216953").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2022/02/25 19:14
S54	1	("5635352").PN.	USPAT; USOCR	OR	OFF	2022/03/01 10:36
S55	424	imaging biological specimen	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/03/01 10:39
S56	11	S55 and flow cell	US-PGPUB; USPAT; USOCR; FPRS; EPO;	NEAR	ON	2022/03/01 10:39

			JPO; DERWENT; IBM_TDB			
S57	2,522,352	imaging	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/03/01 10:47
S58	9,453	S57 and biological specimen	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/03/01 10:47
S59	1,542	S58 and flow cell	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/03/01 10:47
S60	582	S59 and objective	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/03/01 10:47
S61	6	S60 and open flow cell	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/03/01 10:48
S62	13	S58 and open flow cell	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/03/01 10:48
S63	1	("20090280559").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2022/06/09 15:32
S64	1	("20170067096").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2022/06/09 15:33
S65	1	("20130288249").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2022/06/09 15:36
S66	1	("7989166").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2022/06/09 15:37
S67	1	("9551032").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2022/06/09 15:38
S68	4,053	volumetric imaging	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/06/09 15:39
S69	6	S68 and optically-encoded	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/06/09 15:39
S70	5	S68 and read-out cycle$1	US-PGPUB; USPAT; USOCR; FPRS; EPO;	NEAR	ON	2022/06/09 15:40

			JPO; DERWENT; IBM_TDB			
S71	174	S68 and light signal$1	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/06/09 15:41
S72	87	S71 and optical system	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/06/09 15:42
S73	13	S72 and sample holder	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2022/06/09 15:42

6/13/2022 10:46:32 AM
C:\Users\dthomas1\Documents\EAST\Workspaces\17395534.wsp

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

BIB DATA SHEET

CONFIRMATION NO. 5147

SERIAL NUMBER	FILING or 371(c) DATE RULE	CLASS	GROUP ART UNIT	ATTORNEY DOCKET NO.
17/395,534	08/06/2021	435	1637	010498.01437\US

APPLICANTS

President and Fellows of Harvard College, Cambridge, MA;

INVENTORS

Evan R. Daugharthy, Cambridge, MA;
Richard C. Terry, Carlisle, MA;
Je-Hyuk Lee, Allston, MA;
George M. Church, Brookline, MA;
Benjamin W. Pruitt, Cambridge, MA;

**** CONTINUING DATA ***************************

This application is a CON of 16/386,337 04/17/2019 PAT 11118220
which is a CON of 15/969,118 05/02/2018 PAT 10266888
which is a CON of PCT/US16/60279 11/03/2016
which claims benefit of 62/250,182 11/03/2015

**** FOREIGN APPLICATIONS ***************************

**** IF REQUIRED, FOREIGN FILING LICENSE GRANTED ****

08/18/2021

Foreign Priority claimed ❑ Yes ☑ No 35 USC 119(a-d) conditions met ❑ Yes ❑ No Verified and Acknowledged /DAVID C THOMAS/ Examiner's Signature	❑ Met after Allowance Initials	STATE OR COUNTRY	SHEETS DRAWINGS	TOTAL CLAIMS	INDEPENDENT CLAIMS
		MA	13	8	2

ADDRESS

BANNER & WITCOFF, LTD.
1100 13th STREET, N.W.
SUITE 1200
WASHINGTON, DC 20005-4051
UNITED STATES

TITLE

Method and Apparatus for Volumetric Imaging

FILING FEE RECEIVED		
3420	FEES: Authority has been given in Paper No._____ to charge/credit DEPOSIT ACCOUNT No._____ for following:	❑ All Fees ❑ 1.16 Fees (Filing) ❑ 1.17 Fees (Processing Ext. of time) ❑ 1.18 Fees (Issue) ❑ Other _____ ❑ Credit

BIB (Rev. 05/07).

EAST Search History

EAST Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	6,206	sample holder.clm.	US-PGPUB; USPAT	NEAR	ON	2022/06/09 15:46
L2	47,048	temperature control.clm.	US-PGPUB; USPAT	NEAR	ON	2022/06/09 15:46
L3	10,871	motion control.clm.	US-PGPUB; USPAT	NEAR	ON	2022/06/09 15:46
L4	23,016	fluid$3 system.clm.	US-PGPUB; USPAT	NEAR	ON	2022/06/09 15:47
L5	97,419	optical system.clm.	US-PGPUB; USPAT	NEAR	ON	2022/06/09 15:47
L6	44	optically encoded signal$1.clm.	US-PGPUB; USPAT	NEAR	ON	2022/06/09 15:49
L7	81	read-out cycle$1.clm.	US-PGPUB; USPAT	NEAR	ON	2022/06/09 15:49
L8	169,229	three-dimensional.clm.	US-PGPUB; USPAT	NEAR	ON	2022/06/09 15:49
L9	86	l1 same l2	US-PGPUB; USPAT	NEAR	ON	2022/06/09 15:50
L10	5	l3 same l4	US-PGPUB; USPAT	NEAR	ON	2022/06/09 15:50
L11	0	l5 same l6	US-PGPUB; USPAT	NEAR	ON	2022/06/09 15:50
L12	0	l7 same l8	US-PGPUB; USPAT	NEAR	ON	2022/06/09 15:50
L13	0	l9 same l10	US-PGPUB; USPAT	NEAR	ON	2022/06/09 15:51
L14	0	l9 same l4	US-PGPUB; USPAT	NEAR	ON	2022/06/09 15:51
L15	0	l9 same l3	US-PGPUB; USPAT	NEAR	ON	2022/06/09 15:51
L16	1	l9 same l5	US-PGPUB; USPAT	NEAR	ON	2022/06/09 15:51
L17	0	l9 same l6	US-PGPUB; USPAT	NEAR	ON	2022/06/09 15:52
L18	0	l9 same l7	US-PGPUB; USPAT	NEAR	ON	2022/06/09 15:52
L19	0	l9 same l8	US-PGPUB; USPAT	NEAR	ON	2022/06/09 15:52

6/9/2022 3:53:19 PM
C:\Users\dthomas1\Documents\EAST\Workspaces\17395534-2.wsp

UNITED STATES PATENT AND TRADEMARK OFFICE
COMMISSIONER FOR PATENTS
P.O.BOX 1450
ALEXANDRIA VA 22313-1451

PRESORTED
FIRST-CLASS MAIL
U.S. POSTAGE PAID
POSTEDIGITAL
NNNNN

BANNER & WITCOFF, LTD.
1100 13th STREET, N.W.
SUITE 1200
WASHINGTON, DC 20005-4051

Courtesy Reminder for
Application Serial No: 17/395,534

Attorney Docket No: 010498.01437\US
Customer Number: 22907
Date of Electronic Notification: 06/21/2022

This is a courtesy reminder that new correspondence is available for this application. If you have not done so already, please review the correspondence. The official date of notification of the outgoing correspondence will be indicated on the form PTOL-90 accompanying the correspondence.

An email notification regarding the correspondence was sent to the following email address(es) associated with your customer number:

eofficeaction@bannerwitcoff.com
GPD@bannerwitcoff.com

To view your correspondence online or update your email addresses, please visit us anytime at **https://ppair-my.uspto.gov/pair/PrivatePair.**
If you have any questions, please email the Electronic Business Center (EBC) at EBC@uspto.gov or call 1-866-217-9197.

PTO/SB/08a (01-22)
Approved for use through 05/31/2024. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT
(Not for submission under 37 CFR 1.99)

Application Number	17395534
Filing Date	2021-08-06
First Named Inventor	Evan R. Daugharthy
Art Unit	1637
Examiner Name	D. C. Thomas
Attorney Docket Number	010498.01437\US

U.S.PATENTS Remove

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	10669569		2020-06-02	Gullberg et al.	
	2	10815519		2020-10-27	Husain et al.	
	3	10457980		2019-10-29	Cai et al.	
	4	10510435		2019-12-17	Cai et al.	
	5	10872679		2020-12-22	Cai et al.	
	6	10501777		2019-12-10	Beechem et al.	
	7	10246700		2019-04-02	Dunaway et al.	
	8	10640816		2020-05-05	Beechem et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17395534
	Filing Date	2021-08-06
	First Named Inventor	Evan R. Daugharthy
	Art Unit	1637
	Examiner Name	D. C. Thomas
	Attorney Docket Number	010498.01437\US

	9	10415080		2019-09-17	Dunaway et al.	
	10	10844426		2020-11-24	Daugharthy et al.	
	11	10746981		2020-08-18	Tomer et al.	
	12	10495554		2019-12-03	Deisseroth et al.	
	13	10309879		2019-06-04	Chen et al.	
	14	10526649		2020-01-07	Chen et al.	
	15	10317321		2019-06-11	Tillberg et al.	
	16	10364457		2019-07-30	Wassie et al.	
	17	9778155		2017-10-03	Gradinaru et al.	
	18	10794802		2020-10-06	Gradinaru et al.	
	19	10802262		2020-10-13	Tomer et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17395534
	Filing Date	2021-08-06
	First Named Inventor	Evan R. Daugharthy
	Art Unit	1637
	Examiner Name	D. C. Thomas
	Attorney Docket Number	010498.01437\US

	20	8462981		2013-06-11	Determan et al.	
	21	9909167		2018-03-06	Samusik et al.	
	22	9783841		2017-10-10	Nolan et al.	
	23	9376717		2016-06-28	Gao et al.	
	24	10580128		2020-03-03	Miller	
	25	10126242		2018-11-13	Miller et al.	
	26	10032064		2018-07-24	Hoyt	
	27	9541504		2017-01-10	Hoyt	
	28	8330087		2012-12-11	Domenicali	
	29	7555155		2009-06-30	Levenson et al.	
	30	10964001		2021-03-30	Miller	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17395534
	Filing Date	2021-08-06
	First Named Inventor	Evan R. Daugharthy
	Art Unit	1637
	Examiner Name	D. C. Thomas
	Attorney Docket Number	010498.01437\US

	31	8199999		2012-06-12	Hoyt et al.	
	32	7655898		2010-02-02	Miller	
	33	7534991		2009-05-19	Miller et al.	
	34	9727810		2017-08-08	Fodor et al.	
	35	9273349		2016-03-01	Nguyen et al.	
	36	10774372		2020-09-15	Chee et al.	
	37	10774374		2020-09-15	Frisen et al.	

If you wish to add additional U.S. Patent citation information please click the Add button. Add

U.S.PATENT APPLICATION PUBLICATIONS Remove

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	20120270214	A1	2012-10-25	BERNITZ et al.	
	2	20170009278	A1	2017-01-12	SODERBERG et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17395534
	Filing Date	2021-08-06
	First Named Inventor	Evan R. Daugharthy
	Art Unit	1637
	Examiner Name	D. C. Thomas
	Attorney Docket Number	010498.01437\US

	3	20180237864	A1	2018-08-23	Imler et al.	
	4	20180208967	A1	2018-07-26	Larman et al.	
	5	20200140920	A1	2020-05-07	Pierce et al.	
	6	20210017587	A1	2021-01-21	CAI et al.	
	7	20210115504	A1	2021-04-22	CAI et al.	
	8	20170220733	A1	2017-08-03	Zhuang et al.	
	9	20190264270	A1	2019-08-29	Zhuang et al.	
	10	20190276881	A1	2019-09-12	Zhuang et al.	
	11	20190271028	A1	2019-09-05	KHAFIZOV et al.	
	12	20200010891	A1	2020-01-09	BEECHEM et al.	
	13	20190119735	A1	2019-04-25	DEISSEROTH et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17395534
	Filing Date	2021-08-06
	First Named Inventor	Evan R. Daugharthy
	Art Unit	1637
	Examiner Name	D. C. Thomas
	Attorney Docket Number	010498.01437\US

	14	20160305856	A1	2016-10-20	Boyden et al.	
	15	20170081489	A1	2017-03-23	Rodriques et al.	
	16	20190032128	A1	2019-01-31	Chen et al.	
	17	20170253918	A1	2017-09-07	KOHMAN	
	18	20180052081	A1	2018-02-22	KOHMAN	
	19	20200239946	A1	2020-07-30	DEWAL	
	20	20200354782	A1	2020-11-12	DEWAL	
	21	20160108458	A1	2016-04-21	Frei et al.	
	22	20180080876	A1	2018-03-22	Rockel et al.	
	23	20190017106	A1	2019-01-17	FRISEN et al.	
	24	20160376642	A1	2016-12-29	LANDEGREN et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17395534
	Filing Date	2021-08-06
	First Named Inventor	Evan R. Daugharthy
	Art Unit	1637
	Examiner Name	D. C. Thomas
	Attorney Docket Number	010498.01437\US

	25	20170010672	A1	2017-04-13	Luo et al.	

If you wish to add additional U.S. Published Application citation information please click the Add button. Add

FOREIGN PATENT DOCUMENTS Remove

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1	2017143155	WO	A2	2017-08-24	Harvard College		

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]
	1	LEUCHOWIUS et al. "Parallel visualization of multiple protein complexes in individual cells in tumor tissue" Molecular & Cellular Proteomics; Vol. 12; No. 6; pp. 1563-1571; June 2013; published online February 22, 2013	
	2	MARBLESTONE et al. "Rosetta Brains: A strategy for molecularly-annotated connectomics" arXiv:1404.5103v1 -q-bio.NC], https://doi.org/10.48550/arXiv.1404.5103, pp. 1-18; April 21, 2014	

If you wish to add additional non-patent literature document citation information please click the Add button Add

EXAMINER SIGNATURE

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17395534
	Filing Date	2021-08-06
	First Named Inventor	Evan R. Daugharthy
	Art Unit	1637
	Examiner Name	D. C. Thomas
	Attorney Docket Number	010498.01437\US

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

See attached certification statement.

☒ The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/John P. Iwanicki/	Date (YYYY-MM-DD)	2022-08-01
Name/Print	John P. Iwanicki	Registration Number	34,628

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

Electronic Patent Application Fee Transmittal

Application Number:	17395534
Filing Date:	06-Aug-2021
Title of Invention:	Method and Apparatus for Volumetric Imaging
First Named Inventor/Applicant Name:	Evan R. Daugharthy
Filer:	John P. Iwanicki/Jennifer Hazzard
Attorney Docket Number:	010498.01437\US

Filed as Large Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Miscellaneous:				
SUBMISSION- INFORMATION DISCLOSURE STMT	1806	1	260	260
		Total in USD ($)		**260**

Electronic Acknowledgement Receipt	
EFS ID:	46334144
Application Number:	17395534
International Application Number:	
Confirmation Number:	5147
Title of Invention:	Method and Apparatus for Volumetric Imaging
First Named Inventor/Applicant Name:	Evan R. Daugharthy
Customer Number:	22907
Filer:	John P. Iwanicki/Jennifer Hazzard
Filer Authorized By:	John P. Iwanicki
Attorney Docket Number:	010498.01437\US
Receipt Date:	03-AUG-2022
Filing Date:	06-AUG-2021
Time Stamp:	15:13:30
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	DA
Payment was successfully received in RAM	$260
RAM confirmation Number	E202283F16000624
Deposit Account	
Authorized User	
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Foreign Reference	WO2017143155A2_22034633.pdf	2433857 866912be932e9a74d073c5d132a6950d60280bc7	no	57
Warnings:					
Information:					
2	Non Patent Literature	LEUCHOWIUS_et_al_Molecular__Cellular_Proteomics_2013_22025823.PDF	6200169 e9c69bdc9d3817677057589625bb9d03b1799346	no	9
Warnings:					
Information:					
3	Non Patent Literature	MARBLESTONE_et_al_2014_22025824.PDF	374357 adb293378d545d706e1748b97920af7da47c0259	no	18
Warnings:					
Information:					
4	Information Disclosure Statement (IDS) Form (SB08)	Information_Disclosure_Statement_Fillable_PDF_22041232.pdf	1265251 52fac5a98581d0b6de3af56be182794aef345c5c	no	9
Warnings:					
Information:					
5	Fee Worksheet (SB06)	fee-info.pdf	38524 f537cbefab9133e6636faf97826912bf3261fd26	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	10312158		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 August 2017 (24.08.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/143155 A2

(51) **International Patent Classification**:
C40B 30/04 (2006.01)

(21) **International Application Number**:
PCT/US2017/018322

(22) **International Filing Date**:
17 February 2017 (17.02.2017)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:
62/296,746 18 February 2016 (18.02.2016) US

(71) **Applicant**: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, Massachusetts 02138 (US).

(72) **Inventors**: **CHURCH, George M.**; 218 Kent Street, Brookline, Massachusetts 02446 (US). **IYER, Eswar**; 107 Perkins Street, Apt. 3R, Somerville, Massachusetts 02145 (US). **CAMPLISSON, Conor**; 1071 Borden Road, San Marcos, California 92069 (US).

(74) **Agent**: **IWANICKI, John**; Banner and Witcoff, LTD, 28 State Street, Suite 1800, Boston, Massachusetts 02109-1701 (US).

(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

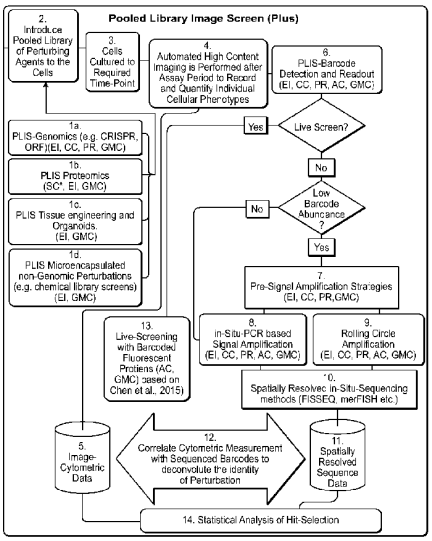
(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) **Title**: MULTIPLEX ALTERATION OF CELLS USING A POOLED NUCLEIC ACID LIBRARY AND ANALYSIS THEREOF

(57) **Abstract**: A method of altering cells with a pooled library of bar-coded nucleic acids is provided including combining a group of cells with the pooled library of bar coded nucleic acids under conditions which promote uptake of one or more barcoded nucleic acids from the pooled library into the cells, analyzing an individual cell for phenotype, sequencing the individual cell to identify the one or more barcoded nucleic acids from the pooled library, and correlating the phenotype for the individual cell with the one or more barcoded nucleic acids from the pooled library.

FIG. 1



WO 2017/143155 A2

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

MULTIPLEX ALTERATION OF CELLS USING A POOLED NUCLEIC ACID LIBRARY AND ANALYSIS THEREOF

STATEMENT OF GOVERNMENT INTERESTS

This invention was made with government support under 5P50HG005550-05 awarded by NIH/NHRGRI. The government has certain rights in the invention.

FIELD

The present invention relates in general to methods of altering cells using a pooled library of nucleic acids, determining phenotype of the altered cells and sequencing of the altered cells to determine genotype.

BACKGROUND

Methods of altering cells and measuring biological activity are known.

SUMMARY

The disclosure provides a method of altering cells with a pooled library of bar coded nucleic acids including combining a group of cells with the pooled library of bar coded nucleic acids under conditions which promote uptake of one or more barcoded nucleic acids from the pooled library into the cells, wherein the barcodes are functional or nonfunctional, analyzing an individual cell for phenotype, sequencing the individual cell to identify the one or more barcoded nucleic acids from the pooled library, and correlating the phenotype for the individual cell with the one or more barcoded nucleic acids from the pooled library. The disclosure provides that the steps of analyzing, sequencing and correlating are conducted for multiple cells within the group of cells. The disclosure provides that one or more barcoded nucleic acids from the pooled library are delivered to the cells using viral or non-viral

methods. The disclosure provides that the individual cell is analyzed for phenotype using automated imaging and one or more barcoded nucleic acids within the individual cell is sequenced using FISSEQ. The disclosure provides that the cells are analyzed for phenotype using high-content imaged based analysis and one or more barcoded nucleic acids within the individual cell is sequenced using FISSEQ. The disclosure provides that the phenotype is correlated with the one or more barcoded nucleic acids from the pooled library using a bioinformatic pipeline to merge cytometric data with sequence information at the single cell level to allow statistical deconvolution of the effect of a nucleic acid from the library with resulting phenotype. The disclosure provides that the one or more barcoded nucleic acids are amplified in situ before using FISSEQ. The disclosure provides that the one or more barcoded nucleic acids are amplified in situ using isothermal or non-isothermal amplification before using FISSEQ. The disclosure provides that the nucleic acid from the library element is delivered using an expression vector where the nucleic acid is flanked by a first common sequence and a second common sequence and promoter which facilitates expression of RNA copies of the nucleic acid, wherein the RNA is reversed transcribed, circularized and amplified by rolling circle amplification before using FISSEQ.

The disclosure provides a method of altering cells with a pooled library of drugs including combining a group of cells with the pooled library of drugs under conditions which promote uptake of one or more drugs from the pooled library into the cells, analyzing an individual cell for phenotype, analyzing the individual cell to identify the one or more drugs from the pooled library, and correlating the phenotype for the individual cell with the one or more drugs from the pooled library.

The disclosure provides a method of altering cells with a pooled library of nucleic acids having fluorescent barcodes including combining a group of cells with the pooled

library of nucleic acids having fluorescent barcodes under conditions which promote uptake of one or more nucleic acids having fluorescent barcodes from the pooled library into the cells, analyzing an individual cell for phenotype, analyzing the individual cell to identify the fluorescent barcode, and correlating the phenotype for the individual cell with the fluorescent barcode.

Further features and advantages of certain embodiments of the present invention will become more fully apparent in the following description of embodiments and drawings thereof, and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee. The foregoing and other features and advantages of the present embodiments will be more fully understood from the following detailed description of illustrative embodiments taken in conjunction with the accompanying drawings in which:

Fig. 1 depicts in schematic a flow chart describing the pooled library image screen methods.

Fig. 2A and Fig. 2B depict in schematic contructs used in the pooled library image screen methods.

Fig. 3 depicts images of results of sgRNA targeted reverse transcription.

DETAILED DESCRIPTION

The present disclosure provides methods of altering a plurality of cells with a pooled library of nucleic acids. One or more nucleic acids of the pooled library are introduced into

the cells. The cells are analyzed for phenotype. Nucleic acids within the cells are sequenced. The cell phenotype is correlated with the one or more nucleic acids from the pooled library. The one or more nucleic acids identified from sequencing results in the observed phenotype.

The disclosure provides a library of nucleic acids. Such a library of nucleic acids may be generated by in silico pooled oligonucleotide synthesis technologies to generate large genomic libraries, such as for screens such as for pooled selection screens. See Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. Science 343, 80–84 (2014); Gilbert, L. A. et al. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell 159, 647–661 (2014); Shalem, O., Sanjana, N. E. & Zhang, F. High-throughput functional genomics using CRISPR-Cas9. Nat Rev Genet 16, 299–311 (2015); Shalem, O. et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science 343, 84–87 (2014) each of which is hereby incorporated by reference in its entirety.

The disclosure provides synthesis of the library of nucleic acids using methods known to those of skill in the art. The library or subcomponents thereof, is delivered to the cells, such as by methods known to those in the art such as viral and non-viral methods. Such methods include lentiviral transduction. The disclosure provides that the cells may be cultured, such as under selective pressure and lysed and sequencing, such as deep-sequencing, may be performed to identify one or more nucleic acids from the library that are introduced into the cells using methods known to those of skill in the art. See Shalem, O., Sanjana, N. E. & Zhang, F. High-throughput functional genomics using CRISPR-Cas9. *Nat Rev Genet* **16,** 299–311 (2015); Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343,** 84–87 (2014); Agrotis, A. & Ketteler, R. A new age in functional genomics using CRISPR/Cas9 in arrayed library screening. *Front Genet* **6,** 300

(2015) each of which is hereby incorporated by reference in its entirety. Statistically-overrepresented nucleic acids may be determined using methods associated with pooled and arrayed screens. See Agrotis, A. & Ketteler, R. A new age in functional genomics using CRISPR/Cas9 in arrayed library screening. *Front Genet* **6,** 300 (2015) hereby incorporated by reference in its entirety.

The disclosure provides analysis of the cells to determine phenotype using methods known to those of skill in the art. Automated microscopy based High Content Screening (HCS) measures biological activity in single cells or whole organisms that have been subjected to libraries of compound or nucleic acids, such as a chemical compound library, an siRNA library, a guide RNA library, as ORFeome library and the like. See Buchser, W. *et al.* Assay Guidance Manual. (2004); Bray, M. A. & Carpenter, A. Assay Guidance Manual. (2004) each of which is hereby incorporated by reference in its entirety.

The disclosure provides a method of pooled library image screening including screening cells with a pooled library of compounds, such as nucleic acids. The cells are analyzed for phenotype. The cells are analyzed to determine presence on or within the cells of one or more members of the library of compounds. The phenotype of a cell is correlated with the identification of the one or more members of the library of compounds on or within a cell. The disclosure provides a method of a pooled screen, cytological profiling and in situ sequencing. See Lee, J. H. *et al.* Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat Protoc* **10,** 442–458 (2015); Lee, J. H. *et al.* Highly multiplexed subcellular RNA sequencing in situ. *Science* **343,** 1360–1363 (2014); Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348,** aaa6090 (2015); Ke, R. *et al.* In situ sequencing for RNA analysis in preserved tissue and cells. *Nat Methods* **10,**

857–860 (2013) each of which are hereby incorporated by reference in its entirety.

The disclosure provides that a pooled library of nucleic acids is delivered to cells using physical, chemical, viral or non-viral methods using methods known to those of skill in the art. See Nayerossadat, N., Maedeh, T. & Ali, P. A. Viral and nonviral delivery systems for gene delivery. *Adv Biomed Res* **1,** 27 (2012) hereby incorporated by reference in its entirety. One or more of the cells, such as each cell, randomly uptakes single or multiple library elements, i.e. one or more of the members of the library. The biological activity, i.e. phenotype, resulting from the introduction of the one or more of the members of the library on or into the cell is measured, such as with high precision using automated imaging and high-content image-based analysis. See Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol* **7,** R100 (2006) hereby incorporated by reference in its entirety. The cells are then subjected to analysis of the presence on or within the cell of one or more library elements using methods known to those of skill in the art. An exemplary method is sequencing of nucleic acids within the cell, such as spatially-resolved in-situ sequencing which is used to identify unique library elements uptaken by individual cells. See Crosetto, N., Bienko, M. & van Oudenaarden, A. Spatially resolved transcriptomics and beyond. *Nat Rev Genet* **16,** 57–66 (2015) hereby incorporated by reference in its entirety. Other methods such as spatially resolved sequencing methods may be used. See Lee, J. H. *et al.* Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat Protoc* **10,** 442–458 (2015); Lee, J. H. *et al.* Highly multiplexed subcellular RNA sequencing in situ. *Science* **343,** 1360–1363 (2014); Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348,** aaa6090 (2015); Ke, R. *et al.* In situ sequencing for RNA analysis in preserved tissue and cells. *Nat Methods* **10,**

857–860 (2013); Larsson, C., Grundberg, I., Söderberg, O. & Nilsson, M. In situ detection and genotyping of individual mRNA molecules. *Nat Methods* **7,** 395–397 (2010); Weibrecht, I. *et al.* In situ detection of individual mRNA molecules and protein complexes or post-translational modifications using padlock probes combined with the in situ proximity ligation assay. *Nat Protoc* **8,** 355–372 (2013); Chen, R. *et al.* A Barcoding Strategy Enabling Higher-Throughput Library Screening by Microscopy. *ACS Synth Biol* **4,** 1205–1216 (2015) each of which are hereby incorporated by reference in its entirety. Useful aspects of identification methods include high sensitivity of detection of nucleic acid barcodes, high signal-to-noise to allow rapid imaging at lower magnification objective to allow fast data collection, scalable aspects and cost-effectiveness. Exemplary methods include spatially-resolved in-situ sequencing methods using padlock probes (see Ke, R. *et al.* In situ sequencing for RNA analysis in preserved tissue and cells. *Nat Methods* **10,** 857–860 (2013) hereby incorporated by reference in its entirety); multiplexed error-robust fluorescence in situ hybridization (MERFISH; see Chen et al., Science; 2015; 348 (6233) aaa6090, doi: 10.1126/science.aaa6090 hereby incorporated by reference in its entirety); and fluorescence in-situ sequencing (FISSEQ; Lee, J. H. *et al.* Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat Protoc* **10,** 442–458 (2015); Lee, J. H. *et al.* Highly multiplexed subcellular RNA sequencing in situ. *Science* **343,** 1360–1363 (2014) each of which is hereby incorporated by reference in its entirety.)

Data or information related to biological activity is correlated with data or information regarding presence on or within the cell of one or more library elements using methods known to those of skill in the art. The disclosure provides bioinformatic methods that are used to merge cytometric data with sequence information at the single cell level. The disclosure provides statistical deconvolution of the effect of each library element or pertubant

with resulting phenotype. The disclosure provides methods with the sensitivity of arrayed formats (where each perturbing agent is physically or fluidically separated) and with the mutiplexability and cost-reduction of pooled screens.

The disclosure provides a method of pooled library imaging screening using a single vessel including the library of components (which may include on the order of from 1000 to 1,000,000 library components) and the cells (which may include 1000 cells to 10,000,000 cells).

The disclosure provides amplification or pre-amplification methods that may be used with sequencing methods described herein. The disclosure provides PCR-based amplification or pre-amplification of target nucleic acid sequences. Such amplification or pre-amplification may be isothermal or non-isothermal. In-situ PCR (see Bagasra, O. Protocols for the in situ PCR-amplification and detection of mRNA and DNA sequences. *Nat Protoc* **2,** 2782–2795 (2007); Mitra, R. D. & Church, G. M. In situ localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Res* **27,** e34 (1999) each of which is hereby incorporated by reference in its entirety) or similar methods may be used to amplify a barcoded region of a nucleic acid of the library several orders of magnitude, such that the sequences become available for more ready detection.

The disclosure provides biological amplification and targeted reverse transcription for use with in situ sequencing. Genomic libraries are introduced into cells in an expression vector where a high number of RNA copies are produced within each cell, such as by using a promoter, such as U6, to create a large number of RNA copies within each cell. These RNA molecules, which themselves may serve a biological function (CRISPR, ORF etc.) also serve as barcodes, i.e. "expressed barcodes" which may be targeted for reverse transcription using a common set of primers that bind to a sequence found commonly in all library elements.

Cells according to the present disclosure include any cell into which foreign nucleic acids can be introduced and expressed as described herein. It is to be understood that the basic concepts of the present disclosure described herein are not limited by cell type. Cells according to the present disclosure include eukaryotic cells, prokaryotic cells, animal cells, plant cells, fungal cells, bacteria cells, archael cells, eubacterial cells and the like. Cells include eukaryotic cells such as yeast cells, plant cells, and animal cells. Particular cells include mammalian cells and human cells. Particular cells include stem cells, such as pluripotent stem cells, such as human induced pluripotent stem cells.

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described in Sambrook, J., Fritsch, E.F. and Maniatis, T., *Molecular Cloning: A Laboratory Manual*, 2nd ed.; Cold Spring Harbor Laboratory: Cold Spring Harbor, N.Y., (1989) and by Silhavy, T.J., Bennan, M.L. and Enquist, L.W., *Experiments with Gene Fusions*; Cold Spring Harbor Laboratory: Cold Spring Harbor, N.Y., (1984); and by Ausubel, F.M. et. al., *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience (1987) each of which are hereby incorporated by reference in its entirety.

Additional useful methods are described in manuals including Advanced Bacterial Genetics (Davis, Roth and Botstein, Cold Spring Harbor Laboratory, 1980), Experiments with Gene Fusions (Silhavy, Berman and Enquist, Cold Spring Harbor Laboratory, 1984), Experiments in Molecular Genetics (Miller, Cold Spring Harbor Laboratory, 1972) Experimental Techniques in Bacterial Genetics (Maloy, in Jones and Bartlett, 1990), and A Short Course in Bacterial Genetics (Miller, Cold Spring Harbor Laboratory 1992) each of which are hereby incorporated by reference in its entirety.

In certain aspects of the invention, vectors and plasmids useful for transformation of a variety of host cells are provided. Vectors and plasmids are common and commercially

available from companies such as Invitrogen Corp. (Carlsbad, CA), Stratagene (La Jolla, CA), New England Biolabs, Inc. (Beverly, MA) and Addgene (Cambridge, MA).

Certain aspects of the invention pertain to vectors, such as, for example, expression vectors. As used herein, the term "vector" refers to a nucleic acid sequence capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid," which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. By way of example, but not of limitation, a vector of the invention can be a single-copy or multi-copy vector, including, but not limited to, a BAC (bacterial artificial chromosome), a fosmid, a cosmid, a plasmid, a suicide plasmid, a shuttle vector, a P1 vector, an episome, YAC (yeast artificial chromosome), a bacteriophage or viral genome, or any other suitable vector. The host cells can be any cells, including prokaryotic or eukaryotic cells, in which the vector is able to replicate.

Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors." In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

10

Typically, the vector or plasmid contains sequences directing transcription and translation of a relevant gene or genes, a selectable marker, and sequences allowing autonomous replication or chromosomal integration. Suitable vectors comprise a region 5' of the gene which harbors transcriptional initiation controls and a region 3' of the DNA fragment which controls transcription termination. Both control regions may be derived from genes homologous to the transformed host cell, although it is to be understood that such control regions may also be derived from genes that are not native to the species chosen as a production host.

Initiation control regions or promoters, which are useful to drive expression of the relevant pathway coding regions in the desired host cell are numerous and familiar to those skilled in the art. Virtually any promoter capable of driving these genetic elements is suitable for the present invention including, but not limited to, lac, ara, tet, trp, IP_L, IP_R, T7, tac, and trc (useful for expression in *Escherichia coli* and *Pseudomonas*); the amy, apr, npr promoters and various phage promoters useful for expression in *Bacillus subtilis*, and *Bacillus licheniformis*; nisA (useful for expression in Gram-positive bacteria, Eichenbaum et al. *Appl. Environ. Microbiol.* 64(8):2763-2769 (1998)); and the synthetic P11 promoter (useful for expression in *Lactobacillus plantarum*, Rud et al., *Microbiology* 152:1011-1019 (2006)). Termination control regions may also be derived from various genes native to the preferred hosts.

In certain exemplary embodiments, the recombinant expression vectors comprise a nucleic acid sequence in a form suitable for expression of the nucleic acid sequence in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant

expression vector, "operably linked" is intended to mean that the foreign nucleic acid sequence encoding a plurality of ribonucleic acid sequences described herein is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleic acid sequence. The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, Calif. (1990). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

Cells according to the present disclosure include any cell into which foreign nucleic acids can be introduced and expressed as described herein. It is to be understood that the basic concepts of the present disclosure described herein are not limited by cell type. Cells according to the present disclosure include eukaryotic cells, prokaryotic cells, animal cells, plant cells, insect cells, fungal cells, archaeal cells, eubacterial cells, a virion, a virosome, a virus-like particle, a parasitic microbe, an infectious protein and the like. Cells include eukaryotic cells such as yeast cells, plant cells, and animal cells. Particular cells include bacterial cells. Other suitable cells are known to those skilled in the art.

Foreign nucleic acids (i.e., those which are not part of a cell's natural nucleic acid composition) may be introduced into a cell using any method known to those skilled in the art for such introduction. Such methods include transfection, transduction, infection (e.g., viral transduction), injection, microinjection, gene gun, nucleofection, nanoparticle bombardment, transformation, conjugation, by application of the nucleic acid in a gel, oil, or cream, by electroporation, using lipid-based transfection reagents, or by any other suitable transfection method. One of skill in the art will readily understand and adapt such methods using readily identifiable literature sources.

As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection (e.g., using commercially available reagents such as, for example, LIPOFECTIN® (Invitrogen Corp., San Diego, CA), LIPOFECTAMINE® (Invitrogen), FUGENE® (Roche Applied Science, Basel, Switzerland), JETPEI™ (Polyplus-transfection Inc., New York, NY), EFFECTENE® (Qiagen, Valencia, CA), DREAMFECT™ (OZ Biosciences, France) and the like), or electroporation (e.g., *in vivo* electroporation). Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (*Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

Typically, the vector or plasmid contains sequences directing transcription and translation of a relevant gene or genes, a selectable marker, and sequences allowing autonomous replication or chromosomal integration. Suitable vectors comprise a region 5' of the gene which harbors transcriptional initiation controls and a region 3' of the DNA

fragment which controls transcription termination. Both control regions may be derived from genes homologous to the transformed host cell, although it is to be understood that such control regions may also be derived from genes that are not native to the species chosen as a production host.

Initiation control regions or promoters, which are useful to drive expression of the relevant pathway coding regions in the desired host cell are numerous and familiar to those skilled in the art. Virtually any promoter capable of driving these genetic elements is suitable for the present invention including, but not limited to, lac, ara, tet, trp, IP_L, IP_R, T7, tac, and trc (useful for expression in *Escherichia coli* and *Pseudomonas*); the amy, apr, npr promoters and various phage promoters useful for expression in *Bacillus subtilis*, and *Bacillus licheniformis*; nisA (useful for expression in gram positive bacteria, Eichenbaum et al. *Appl. Environ. Microbiol.* 64(8):2763-2769 (1998)); and the synthetic P11 promoter (useful for expression in *Lactobacillus plantarum*, Rud et al., *Microbiology* 152:1011-1019 (2006)). Termination control regions may also be derived from various genes native to the preferred hosts.

According to certain aspect of the invention, phages and their genetic material are provided. As used herein, the terms "phage" and "bacteriophage" are used interchangeably. Phage can be distinguished from each another based on their genetic composition and/or their virion morphology. Some phage have double stranded DNA genomes, including phage of the corticoviridae, lipothrixviridae, plasmaviridae, myrovridae, siphoviridae, sulfolobus shibate, podoviridae, tectiviridae and fuselloviridae families. Other phage have single stranded DNA genomes, including phage of the microviridae and inoviridae families. Other phage have RNA genomes, including phage of the leviviridae and cystoviridae families. Exemplary *bacteriophage* include, but are not limited to, Wphi, Mu, T1, T2, T3, T4, T5, T6,

T7, P1, P2, P4, P22, fd, phi6, phi29, phiC31, phi80, phiX174, SP01, M13, MS2, PM2, SSV-1, L5, PRD1, Qbeta, lambda, UC-1, HK97, HK022 and the like.

Exemplary libraries useful in the disclosed methods include CRISPR libraries such as those at world wide website addgene.org/crispr/libraries; human ORFeome such as those at world wide website dharmacon.gelifesciences.com/gene-expression-cdnas-orfs/mammalian-orfs/horfeome-v81-library; siRNA libraries, microRNA libraries, phagemid antibody libraries, ribosome display libraries and the like. Other libraries useful in the methods described herein will become apparent to those of skill in the art based on the present disclosure.

Guide RNA as described herein refer to the guide RNA component of a CRISPR system known to those of skill in the art. The guide RNA includes a spacer sequence of about 20 nucleotides which is complementary to a target nucleic acid sequence, referred to as a protospacer.

The following examples are set forth as being representative of the present disclosure. These examples are not to be construed as limiting the scope of the present disclosure as these and other equivalent embodiments will be apparent in view of the present disclosure, figures and accompanying claims.

EXAMPLE I

Pooled Library Image Screen (PLIS)

The disclosure provides a method of pooled library image screening as generally depicted in Fig. 1. Pooled and barcoded libraries can be introduced inside biological systems (e.g. cells, iPSC derived human organoids or whole organism) using a suitable physical, chemical or biological method such that each cell discretely and probabilistically uptakes one to several library elements based upon the delivery conditions (See Fig. 1, steps 1a-1c and

step 2.). Cells are cultured until the required time-point (See Fig. 1, step 3). The resulting phenotypic consequence of the introduced library element is assayed via high-content microscopic analysis (See Fig. 1, step 4 where automated high content imaging is performed after the assay period to record and quantify individual cellular phenotype). Images are segmented and analyzed to identify and measure cytometric changes in each cell to obtain cytometric data (See Fig. 1, step 5). In order to identify the library element introduced into each cell, targeted and/or un-targeted in-situ-sequencing is used to identify barcodes and in addition can be used to simultaneously measure transcriptomic changes (See Fig. 1, step 6). In cases that require pre-amplification of a target nucleic acid sequence such as when a barcode is in low abundance, either in-situ PCR based signal amplification or rolling circle amplification or other amplification methods known to those of skill in the art can be used to amplify the target nucleic acid sequence (See Fig. 1, steps 7-9). Spatially resolved In-situ sequencing methods known to those of skill in the art (e.g. FISSEQ, MERFISH) is performed on the same cells to read the introduced barcodes associated with each perturbation event (See Fig. 1, step 10) and spatially resolved sequence data is collected (See Fig. 1, step 11). Sequenced reads are computationally mapped to cellular and sub-cellular compartments identified from step 5. Cytometric measurements or data from step 5 are correlated with sequence barcodes from step 11 to deconvolute the identity of perturbation, i.e. the phenotype is matched with the library element on or within the cell. (See Fig. 1, step 12.) Bioinformatic methods or pipelines are used to perform statistical analyses and hit-selection (See Fig. 1, step 14).

EXAMPLE II

Common Sequences in Expressed Libraries are Used in Targeted In Situ Sequencing

Library elements can be introduced into the cells as expression plasmids which

contain a high-expressivity promoter and express several RNA copies of a variable library region, such as a spacer sequence for guide RNA (CRISPR N20) or an ORFeome. As depicted in Fig. 2A, the disclosure provides an expression vector (step 1) which includes a variable library element having a 5' first common sequence (common sequence 1) and a 3' second common sequence (common sequence 2). The first common sequence and the second common sequence flank the variable library element. As depicted in Fig. 2A, the common sequences in the expressed RNA can be used to perform targeted-cDNA conversion (step 2) using universal RT Primers containing FISSEQ adapters (red) to generate linear cDNA molecules containing both the variable region and FISSEQ adapter. In Fig, 2A, step 3, the cDNA molecules are enzymatically circularized using CircLigase and amplified using rolling circle amplification to generate hundreds of tandem copies of library sequence, which can then be sequenced with FISSEQ.

The disclosure provides an example of a vector using a 20 nucleotide spacer sequence of a guide RNA as a variable region. As depicted in Fig. 2B, sgRNA has 20-nucleotide variable region (Orange) at the 5' end that determines the DNA target site, and also contains a common backbone region (green). Targeted-RT is performed using universal primers targeting the common-region (green), and cDNA molecules containing the variable N20 region are circularized and sequenced using FISSEQ. The variable also serves as an expressed barcode and is used to map the identity of perturbation within a given cell.

EXAMPLE III

Transfection of a Pooled Library of CRISPR sgRNA in Human PGP1 Fibroblasts

A pooled-library of CRISPR sgRNA was transfected in human PGP1 fibroblasts and a high-density of targeted rolling circle amplicons were generated (See Fig. 3), compatible with FISSEQ for sequencing and detection. The same approach was used to detect the

ORFeome library. One advantage of this approach is that it does not require additional modifications to the plasmid or vectors, and is compatible with existing libraries, as long as there is a common sequence to perform targeted-RT. Targeted primers were designed for the experiments using methods known to those of skill in the art.

The disclosure provides a Pooled Library Image Screen using targeted amplification and sequencing of a Synergistic Activation Mediator (SAM) library sgRNA in primary human fibroblasts from the Personal Genome Project (PGP1F). Cells were transiently transfected using a lentivirus with the pooled SAM sgRNA library including over 70,290 guide spacer sequences targeting all human RefSeq coding isoforms. See Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517,** 583–588 (2015) hereby incorporated by reference in its entirety. The reverse transcription primer targets expressed sgRNAs and adds an adapter sequence (T2S) which in turn serves as a common priming site for amplification, fluorescent *in situ* hybridization (FISH), and sequencing using FISSEQ. FISH was used to verify successful amplification and to quantify amplicon density. Using the T2S_N sequencing primer, the first two bases of the targeted sequence are interrogated. Amplicons appear in all four channels due to the diversity of the SAM sgRNA library. Using the T2S_N-2 sequencing primer, the last two bases of the T2S adapter sequence are interrogated, which are the same for every amplicon, which serves as a sequencing control.

The disclosure provides that a pooled library of sgRNA or ORFeome can be lentivirally delivered to cultured cells. Using targeted spatially-resolved sequencing (as described above), the identity of each library element can be deconvoluted (as shown in Fig. 2). The disclosure provides targeted generation of rolling-circle products from PGP1 fibroblast cells transiently transfected with Synergistic activation mediator (SAM) sgRNA

library consisting of over 70,000 unique guide RNA sequences.

EXAMPLE IV

Micropatterned Cells

The disclosure provides methods of cell micropatterning in combination with High Content Screening. See Harkness, T. *et al.* High-Content Imaging with Micropatterned Multiwell Plates Reveals Influence of Cell Geometry and Cytoskeleton on Chromatin Dynamics. *Biotechnol J* (2015). doi:10.1002/biot.201400756 hereby incorporated by reference in its entirety. Pooled library image screening is combined with cell-micropatterning to perform highly-sensitive pooled library image screens.

The disclosure provides a method to perform pooled library imaging screening on engineered human organs. An EB-Grid is combined with pooled library imaging screening to create gene deletions (or expression controls) of key regulatory genes which result in the development of fully vascularized organoids or systems of organs, without forming a human embryo. The gridded organ-systems are used for Protein, RNA or drug-library screens with high-content imaging and/or FISSEQ. The specified layout of the organs in a 3D pattern enable reproducibility, imaging and easy data interpretation.

EXAMPLE V

Pooled Image Screens on Live Cells

The disclosure provides methods of performing pooled library image screening using fluorescent proteins for barcoding in live cells. See Fig. 1, step 13 and Chen, R. *et al.* A Barcoding Strategy Enabling Higher-Throughput Library Screening by Microscopy. *ACS Synth Biol* **4,** 1205–1216 (2015) hereby incorporated by reference in its entirety. The use of fluorescent reports allows pooled library image screening to be performed on live cells,

without in-situ sequencing.

EXAMPLE VI

Pooled Library Image Drug Screening

The disclosure provides methods of pooled library image screening using drugs. Chemical compounds are encapsulated or coated on micro or nanoparticles (see Bao, G., Mitragotri, S. & Tong, S. Multifunctional nanoparticles for drug delivery and molecular imaging. *Annu Rev Biomed Eng* **15,** 253–282 (2013) hereby incorporated by reference in its entirety) along with a nucleic acid barcode such as a DNA barcode, and delivered to cells or tissues, in culture or whole organisms. Pooled library image screening is used to deconvolute the identity of compounds uptaken by cells, and the resulting phenotype.

EXAMPLE VII

High-Throughput Cell and Tissue Engineering

The disclosure provides methods of pooled library image screening to create in parallel large number of genomic variants in single cells or engineered organoids, which can be used to perform pooled forward/ reverse genetic screens. The disclosure provides that mutant human iPSCs can be parallel generated with PLIS, to screen for factors that allow growth of 3-D vascularized human organoids. For example, human iPSC cells exposed to genomic library are dissociated and cultured in matrigel in low density to permit them to grow into 3-dimentional embryioid bodies and organoids. See Meinhardt, A. *et al.* 3D reconstitution of the patterned neural tube from embryonic stem cells. *Stem Cell Reports* **3,** 987–999 (2014). 3-D phenotypes are identified via screening and the identity of the pertubing library element is deconvoluted with in-situ sequencing. This process is iterated to identify factor promoting growth of desired vascularized tissue type. This process screens for factors

that can enable the growth of vascularized organoids.

EXAMPLE VII

Materials and Methods

Cell Culture and Lentiviral Transduction: Cells are cultured under standard culture conditions. Brifefly, PGP1 Fibroblasts, or HeLa-Cas9 cells, are cultured in DMEM with 4.5 g/L D-glucose, GlutaMAX™ Supplement, and 110 mg/L sodium pyruvate, supplemented with 10% FBS (Gibco) and 1% penicillin-streptomycin. Cells are seeded on glass-bottom dishes (MatTek) and subsequently transduced with a lentiviral library containing genetic perturbations (sgRNAs, TFs, etc.). In order to perform statistically analysis, the cell seeding and viral titer delivery should be optimized to yield at least 10-100X coverage (each library element is represented in at least 10-100 cells). Optionally, drug selection can be performed to eliminate cells that were not transduced.

in situ Sequencing Library Construction: Once lentiviral transduction and selection are complete, cDNA libraries are constructed for *in situ* sequencing. Cells are fixed with 4% paraformaldehyde for 15 minutes, permeabilized with 70% ethanol for 5 minutes, and treated with 0.1N HCl for 1 minute. A targeted reverse transcription primer with common adapter sequence is then used to capture expressed library elements (e.g. sgRNA molecules). Reverse transcription (RT), circularization, and rolling circle amplification (RCA) are performed as previously described in Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Yang JL, Ferrante TC, et al. Highly multiplexed subcellular RNA sequencing in situ. Science. 2014 Mar 21;343(6177):1360–3. and Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Ferrante TC, Terry R, et al. Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. Nat Protoc. 2015 Mar;10(3):442–58 each of which is hereby incorporated by reference in its entirety (See Appendix 1: FISSEQ Library Construction

Protocol). Alternatively, *in situ* amplification and sequencing can be accomplished using Padlock Probes (see Weibrecht I, Lundin E, Kiflemariam S, Mignardi M, Grundberg I, Larsson C, et al. In situ detection of individual mRNA molecules and protein complexes or post-translational modifications using padlock probes combined with the in situ proximity ligation assay. Nat Protoc. 2013 Feb;8(2):355–72; Larsson C, Grundberg I, Söderberg O, Nilsson M. In situ detection and genotyping of individual mRNA molecules. Nat Methods. 2010 May;7(5):395–7 each of which are hereby incorporated by reference in its entirety (see Appendix 2: Padlock Probe Amplification Protocol) or Polony *in situ* PCR which combines Polony PCR (see Mitra RD, Church GM. In situ localized amplification and contact replication of many individual DNA molecules. Nucleic Acids Res. 1999 Dec 15;27(24):e34 hereby incorporated by reference in ite entirety with in-situ PCR (see Bagasra O. Protocols for the in situ PCR-amplification and detection of mRNA and DNA sequences. Nat Protoc. 2007;2(11):2782–95 hereby incorporated by reference in its entirety. In brief, cells are fixed, permeabilized and embedded in a layer of gel (for example Acrylamide). The primers are optionally cross-linked to the gel and in-situ PCR amplification if performed (see Appendix 3: Polony Amplification Protocol).

Primers used:

T2S-IDT-gRNA-RT-Primer-1	/5Phos/ACTTCAGCTGCCCCGGGTGAAGATGATAACGGACTAGC [HPLC] [order 250 nmole]
T2S-IDT-gRNA-RT-Primer-2	/5Phos/ACTTCAGCTGCCCCGGGTGAAGACTAGCCTTATTTTAA [HPLC] [order 250 nmole]

gRNA_Cy3_Probe (AKA gRNA-FISH-probe)	/5Cy3/AACTTGCTATTTCTAGCTCTAAAAC [HPLC] [order 250 nmole]

T2S_IDT_gRNA_RCA_P1	GTTTTAGAGCTAGAAATAGCA*A*G
gRNA-Sequencing Anchor	/5Phos/AACTTGCTATTTCTAGCTCTAAAAC

Fluorescent *in situ* Sequencing: Once the library preparation is complete, the quality of sequencing libraries is verified by fluorescence *in situ* hybridization (FISH). For example, a fluorescently labeled oligo-probe complementary to an adapter sequence added during reverse transcription. This provides the density and subcellular localization of amplicons. The probe is then stripped and replaced with a sequencing primer. After successful library construction, either SOLiD (see Appendix 4: SOLiD Sequencing Protocol) or fluorescent 9-mer sequencing by ligation (SBL) chemistries can be used to determine identity of the sgRNA present in each cell (see 7. Ke R, Mignardi M, Pacureanu A, Svedlund J, Botling J, Wählby C, et al. In situ sequencing for RNA analysis in preserved tissue and cells. Nat Methods. 2013 Sep;10(9):857–60 hereby incorporated by reference in its entirety (see Appendix 5: Sequencing by ligation using 9-mer probes). Sequencing can be performed with either manual or automated fluidic exchanges.

Cytological Profiling: The goal of this step is to identify the corresponding phenotypic changes associated with modification by a library element. Once the library construction is successfully completed, the sample is mounted on a microscope with appropriate settings. In order to measure the phenotypic state of each cell, the cells are subjected to phenotypic screening. This can be performed prior to sequencing, or immediately post sequencing. However it is exemplary that the different image sets are perfectly aligned within the X-Y-Z coordinates to enable cell data merging. In order to achieve this, the samples cannot be moved once they are mounted on the microscopes until high-content screening and sequencing both have taken place. The cells are subjected to standard high-content image based screening assays based on the biological problem of

interest. For example, tagged proteins, small molecules and antibodies can be used to label and image the cellular features of interest (see Buchser W, Collins M, Garyantes T, Guha R, Haney S, Lemmon V, et al. Assay Development Guidelines for Image-Based High Content Screening, High Content Analysis and High Content Imaging. In: Sittampalam GS, Coussens NP, Nelson H, Arkin M, Auld D, Austin C, et al., editors. Assay Guidance Manual [Internet]. Bethesda (MD): Eli Lilly & Company and the National Center for Advancing Translational Sciences; 2004; Gustafsdottir SM, Ljosa V, Sokolnicki KL, Anthony Wilson J, Walpita D, Kemp MM, et al. Multiplex cytological profiling assay to measure diverse cellular states. PLoS ONE. 2013 Dec 2;8(12):e80999 each of which is hereby incorporated by reference in its entirety (see Appendix 6: Sample protocol for cell staining).

Image Processing and Bioinformatics Pipeline: Images are passed through a custom bioinformatics pipeline which aligns images, performs base calling. For example the images can be analyzed using CellProfiler software (see Bray M-A, Vokes MS, Carpenter AE. Using CellProfiler for Automatic Identification and Measurement of Biological Objects in Images. Curr Protoc Mol Biol. 2015 Jan 5;109:14.17.1–14.17.13; Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol. 2006 Oct 31;7(10):R100 each of which are hereby incorporated by reference in its entirety. Briefly illumination correction, background subtraction and feature segmentation is performed. The nuclei, cytoplasm and other stained features are identified. The image analysis pipeline can then measure various parameters like cell size, shape, texture, intensity, local density of object of interest within certain cellular compartments (eg. nucleus vs. cytoplasm). Images are also processed to identify and read the barcoded sequences using another pipeline (an example pipeline is described in Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Ferrante TC, Terry R, et al. Fluorescent in situ

sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. Nat Protoc. 2015 Mar;10(3):442–58 hereby incorporated by reference in its entirety. The identified barcodes are eventually mapped at single cell level to individual cellular and subcellular compartments identified from the first image-processing pipeline, since both datasets are aligned in common X-Y-Z coordinates.

Appendix 1: FISSEQ Library Construction Protocol

Protocols from *Lee et al.* 2014 may be used for the remaining steps of the library construction protocol after reverse transcription. After this protocol, the sample is ready for sequencing. For Example, Lee et al 2014 provides that the sample is washed using 1x PBS and cross-linked using BS(PEG)9 (Thermo Scientific), diluted to 50 mM in PBS, for 1 hour at 25°C. 1 M Tris (G Biosciences) is added to quench the reaction for 30 minutes at 25°C. A mixture of DNase-free RNases (Roche Diagnostics) and RNase H (Enzymatics) is added to degrade residual RNA for 1 hour at 37°C. A 100 uL circularization reaction mixture (1x reaction buffer, 2.5 mM MnCl2, 1 M Betaine and 5 uL CircLigase II from Illumina/Epicentre) is then added to the sample well and incubated at 60°C for 2 hours. After circularization, the sample is washed using H20 and incubated with a 200 uL mixture containing 0.1 uM RCA primer (TCTTCAGCGTTCCCGA*G*A from IDT) in 2x SSC and 30% formamide for 15 minutes at 60°C. The sample is washed using 2x SSC, and a 200 uL amplification mixture containing 500 U 3 Phi29 DNA polymerase (Enzymatics), 250 uM dNTP, 40 uM, and aminoallyl dUTP is added. The sample is incubated in a dry 30°C chamber overnight and cross-linked using BS(PEG)9 diluted to 50 mM in PBS for 1 hour at 25°C. After a rinse with PBS, 1 M Tris is added to quench the reaction for 30 minutes.

Appendix 2: Padlock Probe Amplification Protocol

Protocols from *Larsson et al.* 2010 provide an alternative strategy for generating RCA amplicons, via the hybridization, ligation, and amplification of padlock probes. After this protocol, the sample is ready for sequencing. For example, Larsson et al. provides that to make the target cDNA strands available for padlock probe hybridization, the RNA portion of the created RNA-DNA hybrids was degraded with ribonuclease H. This was performed in the same step as the padlock probe hybridization and ligation. For most reactions, Ampligase (Epicentre) was used for ligation. Samples were first preincubated in Ampligase buffer (20 mM Tris-HCl, pH 8.3, 25 mM KCl, 10 mM $MgCl_2$, 0.5 mM NAD and 0.01% Triton X-100). Ligation was then carried out with 100 nM of each padlock probe in a mix of 0.5 U μl^{-1} Ampligase, 0.4 U μl^{-1} RNase H (Fermentas), 1 U μl^{-1} RiboLock RNase Inhibitor, Ampligase buffer, 50 mM KCl and 20% formamide. Incubation was performed first at 37 °C for 30 min, followed by 45 min at 45 °C. For detection of actin transcript isoforms in mouse embryonic tissue sections, ligation was instead carried out using T4 DNA ligase (Fermentas). Samples were then first preincubated in T4 DNA ligase buffer (Fermentas). Then 100 nM of each padlock probe was added with 0.1 U μl^{-1} T4 DNA ligase, 0.4 U μl^{-1} RNase H, 1 U μl^{-1} RiboLock RNase Inhibitor and 0.2 $\mu g\ \mu l^{-1}$ BSA in T4 DNA ligase buffer supplemented with 0.5 mM ATP and 250 mM NaCl. Slides were then incubated at 37 °C for 30 min. After ligation with Ampligase or T4 DNA ligase, slides were washed in DEPC-treated 2× SSC with 0.05% Tween-20 at 37 °C for 5 min and rinsed in PBS-T. Slides were preincubated briefly in Φ29 DNA polymerase buffer (Fermentas). RCA was then performed with 1 U μl^{-1} Φ29 DNA polymerase (Fermentas) in the supplied reaction buffer, 1 U μl^{-1} RiboLock RNase Inhibitor, 250 μM dNTPs, 0.2 $\mu g\ \mu l^{-1}$ BSA and 5% glycerol. Incubation was carried out for 60 min at 37 °C. The incubation was followed by a wash inPBS-T."

Appendix 3: Polony Amplification Protocol

An alternate exemplary protocol for generating amplicons wherein modified PCR primers are covalently crosslinked to a polyacrylamide gel matrix, fixing them in space is provided below. See also see (Mitra and Church, 1999). After this protocol, the sample is ready for sequencing.

CAST ACRYLAMIDE GELS

1. Remove 8 Bind-Silane treated slides from storage box in dessicator and place face up in AirClean hood.

2. Turn on UV lamp in AirClean hood for 15 minutes.

3. Label slides with numbers and date using a SHARPIE pen (other inks will wash off in hexane).

4. Almost completely cover slides with coverslips (Fisherbrand, 18mm x 30mm, #1, untreated). Keep a small area of oval exposed for subsequent gel loading.

5. Prepare fresh "ABD mix" in a 15 ml polystyrene conical:

- 9 mL IEF 40% Acrylamide
- 1 mL Acrylamide / Bis (19:1; 38%:2%)
- 200 mg DATD

6. Using a 3-cc syringe and a 0.22 micron filter, filter ~1 mL of the ABD mix into a 1.5 mL microcentrifuge tube.

7. Remove Acrydite-Modified Amplification Primer from pre-PCR area of -20'C freezer to thaw.

8. Prepare fresh 5% APS. Notably, the APS bottle should be stored a room-temperature dessicator such as a large plastic screw-top container.

- 0.5 g APS →10 mL dH20

9. Prepare fresh 5% TEMED

- 2 uL TEMED → 38 uL dH20.

10. Put a few drops of 30% BSA into a 1.5 mL tube (BSA should be stored at 4'C)

11. Prepare the gel-casting mix (200 uL total volume). Do not add APS until immediately prior to casting the gels. This recipe is for a 10% gel with sufficient mix for at least 8 (identical) slides.

- 50 uL A-B-D mix **(FILTERED)**
- 1.33 uL 30% BSA
- 2 uL **JCF-AC** (100 uM)
- 136.66 uL dH20
- 4 uL 5% TEMED
- 4 uL 5% APS
- 0.5 uL Template (at appropriate concentration)

12. Gel Loading:

Suck up 18 microliters of the mix and pipet it into the small exposed area of the oval, such that surface tension pulls the liquid into the space between the coverslip and glass to cover the full surface area of the oval. The liquid should distribute under the coverslip such that only a small amount (1 to 2 uL) cannot fit. Slide the coverslip over such that the oval is completely covered.

As many as 8 gels from a single mix are loaded (after adding APS). For more slides, the master mix may be split prior to adding APS (so that polymerization doesn't proceed too far in the tube before all of the gels are cast). The master-mix can also be split prior to adding APS if needed to have a different template on each slide, for instance.

13. Place the slides on a flat tray and load the tray into the argon chamber (and fill with argon). Allow gels to polymerize for ~30 minutes.

14. Remove slides from argon chamber. Use a razor-blade to cleanly remove cover-slips from polymerized gels by sticking the edge of the blade under the coverslip and gently "popping" it up.

15. To wash off the excess acrylamide monomer, place slides in a dH20 filled plastic Coplin jar and incubate for 30 minutes at RT with slow shaking.

16. Remove slides from Coplin jar and place face up in PCR hood.

DIFFUSE IN PCR REAGENTS

17. Prepare diffuse-in mix while allowing slides to dry, while not allowing the slides to over-dry. A thin, shrinking film of liquid on the surface of each gel may be observed. Generally, the diffuse-in mix is added within 5 minutes or so of this shrinking film's disappearance. This generally means about 30 minutes of drying.

18. Prepare diffuse-in mix (200 uL total volume):

- 152.33 uL dH20
- 10 uL 5 mM unlabeled dNTP mix
- 20 uL 10x Taq Buffer (with MgCl2)
- 1.33 uL 30% BSA
- 2 uL 10% Tween-20
- 1 uL unmodified amplification primer (100 uM)
- 13.33 uL Jumpstart TAQ (2.5 units / uL)

19. Pipet 25 uL of diffuse-in mix to the center of each gel. Assuming the gel is newly dried, the liquid will ball up rather than spreading over the surface.

20. Apply 18 x 30 mm cover-slip by resting one edge of the coverslip on the Teflon coat and allowing the coverslip to "fall" onto the gel. The liquid should spread to evenly cover the surface of the gel.

21. Apply an orange SecureSeal chamber. Seal down the edges with the blunt end of a set of tweezers.

22. Fill chamber with mineral oil. Use a P1000 pipetman filled with ~550 uL of mineral oil. Use a Chemwipe to dry off the tip of the pipet tip.

23. Wipe off any excess mineral oil from the surface of the SecureSeal, if neccessary. Seal holes.

PCR AMPLIFICATION

24. Place slides in thermocycler (with labels facing out).

25. To PCR, cycle slides as follows:

- 1. 94 C x 3 minutes
- 2. 94 C x 45 seconds
- 3. 58 C x 30 seconds
- 4. 72 C x 3 minutes
- 5. Goto Step 2, 43 more times
- 6. 72 C x 6 minutes
- 7. 4 C forever

Both the annealing temperature and the extension time can be adjusted to optimize for the set of amplification primers being used and the length of the PCR products being formed, respectively. The above protocol may be used for an 800 bp template.

SLIDE CLEAN UP

26. Fill a GLASS Coplin jar with Hexane

27. Fill a PLASTIC Coplin jar with Wash 1E

28. Without removing cover-slips, place the slides in the hexane-filled glass Coplin jar and leave for 5 minutes.

29. One-by-one, remove the slides and cleanly pop off coverslips with a razor-blade as above. Quick wave in the air is sufficient to evaporate off residual hexane. Remove all residual adhesive from the cover-chamber using a razor-blade, and place the slide in the plastic Coplin containing Wash 1E.

30. Wash 2 x 4' in Wash 1E (meaning, 4 minutes shaking in Wash 1E, then replace the solution in the same Coplin jar with fresh Wash 1E and allow it to go another 4 minutes on the shaker).

31. Gels can remain indefinitely (at least a week or two) before proceeding.

Appendix 4: SOLiD Sequencing Protocol

Protocols from *Lee et al.* 2014 may be used to perform SOLiD sequencing. For example, five sequencing primers are designed specific to a universal adaptor (N, N-1, N-2, N-3, N-4, where N-x is recessed at the 5' end by x-nt). These sequencing primers are annealed to the sample sequentially, and k ligation reactions are performed for each primer (k+1 ligation reactions for primers N-2, N-3, and N-4). Each sequencing primer is annealed to the sample at 2.5 uM in 200 uL 80°C 5X SASC (0.75 M sodium acetate and 75 mM tri-sodium citrate, pH 7.5), incubating for 10 minutes at 25°C. The sample is washed twice for one minute each with 1 mL 1x Instrument Buffer (SOLiD Instrument Buffer Kit, Applied Biosystems Cat# 4389784). 200 uL sequencing mix is freshly prepared on ice using 165 uL nuclease-free H20, 20 uL T4 DNA ligase buffer, 10 uL T4 DNA Ligase (Enzymatics), and 5 uL SOLiD sequencing oligos (the dark purple tubes from the SOLiD ToP Sequencing Kit Fragment Library F3 Tag MM50 Cat# 4449388). After aspirating the Instrument Buffer, the

sequencing mix is added to the sample and incubated at 25ºC for 45 minutes. The sequencing mix is aspirated, and the sample is washed with 1x Instrument Buffer (four 1 mL washes for 5 minutes each). Imaging is done in 1 mL 1x Instrument Buffer. After aspirating the Instrument Buffer, the fluorophore is cleaved to allow for subsequent ligation. The sample is incubated twice for 5 minutes each in 200 uL 1x Cleave Solution 1 (SOLiD ToP Instrument Buffer Kit Component 4406489), followed by two incubations for five minutes each in 200 uL 1X Cleave Mix 2.1 (SOLiD ToP Instrument Buffer Kit Component 4445677, prepared fresh with 106.7 uL Cleave 4 Solution 2.1 Part 1 and 293.3 uL Cleave Solution Part 2). After the second incubation with Cleave Mix 2.1, the sample is washed three times for 5 minutes each with 1x Instrument Buffer. After repeating the cyclic ligation process k (or k+1) times, the ligated strands are stripped by four 5 minute washes in 80ºC strip buffer (80% formamide, 0.01% Triton-X100). Another sequencing primer is annealed, and the cyclic ligation process is repeated."

Appendix 5: Sequencing by Ligation using 9-mer probes

Protocols from *Ke R. et al.,* 2013 may be used for sequencing by ligation. For example, before the sequencing is performed, the detection probes are stripped off. The slides are first incubated through an ethanol series to remove the mounting medium and dried at room temperature. For detection probes without uracils, the samples are first washed with DEPC-PBS-T and then incubated three times with 65% formamide for 30 s, which is followed by washing twice with DEPC-PBS-T. Detection probes that contained uracils are first treated with UNG treating buffer (1× phi29 polymerase buffer (Fermentas), 0.2 μg/μl BSA, 0.02 U/μl UNG (Fermentas)) for 10 min and washed twice with DEPC-PBS-T before the formamide incubation. A mix containing 500 nM of corresponding anchor primers in 2×

SSC and 20% formamide is then added to the sample and incubated at RT for 30 min, and the incubation is followed by two brief washes with DEPC-PBS-T. A ligation mix containing each interrogation probe, 1× T4 ligase buffer (Fermentas), 1 mM ATP (Fermentas) and 0.1 U/μl of T4 ligase (Fermentas) is applied to the samples and incubated for 30 min at RT. The unligated probes are washed away by 3× 1 min incubation with DEPC-PBS-T. The slides are mounted in Vectashield mounting medium containing 100 ng/ml of DAPI for counterstaining the nuclei. The concentration of each interrogation probe is 100 nM for cell cultures and 500 nM for tissue sections. After imaging, the slides are prepared for the next sequencing cycle by UNG treatment buffer as described above followed by repeating the hybridization, ligation and imaging processes. To sequence each base, the same procedures are applied. For evaluation of loss of sequencing substrates (RCA products), cell cytoplasm is stained after the last sequencing cycle with Alexa Fluor 488 phalloidin (Invitrogen) in PBS at a final concentration of 0.15 μM, and the cells are incubated for 10 min at RT.

Appendix 6: Example of sample staining for High-content image analysis

Protocols from *Gustafsdottir SM et al., 2013* may be used for sample staining for high-content image analysis. For example, samples may be stained as follows.

Step 1: MitoTracker and Wheat Germ Agglutinin staining.

MitoTracker Deep Red (#M22426, Invitrogen) is dissolved in DMSO to 1 mM. Wheat Germ Agglutinin (WGA) Alexa594 conjugate (#W11262, Invitrogen) is dissolved in dH2O to 1 mg/mL. A 500 nM MitoTracker, 60 μg/mL WGA solution is prepared in prewarmed media (DMEM, 10% FBS, 1% penicillin/streptomycin). Media is removed from plates; residual volume is 10 μL in each well. 30 μL of staining solution is added to wells and incubated for 30 min at 37 °C.

Step 2: Fixation.

10 µL of 16% methanol-free paraformaldehyde (#15710-S, Electron Microscopy Services) is added to wells for a final concentration of 3.2%. The plates are then incubated at room temperature for 20 min. Wells are washed once with 70 µL 1xHBSS (#14065-056, Invitrogen).

Step 3: Permeabilization.

A 0.1% solution of Triton X-100 (T8787-100mL, Sigma) is prepared in 1x HBSS. 30 µL of the solution is added to the wells and incubated for 10–20 min. Wells are washed twice with 70 µL 1x HBSS.

Step 4: Phalloidin, ConcanavalinA, Hoechst, and SYTO 14 staining.

Concanavalin A Alexa488 conjugate (#C11252, Invitrogen) is dissolved to 1 mg/mL in 0.1 M sodium bicarbonate (SH30033.01, HyClone), and Phalloidin Alexa594 conjugate (#A12381, Invitrogen) is dissolved in 1.5 mL methanol (67-56-1, BDH) per vial. A 0.025 µL phalloidin/µL solution, 100 µg/mL ConcanavalinA, 5 µg/mL Hoechst33342 (#H3570, Invitrogen), and 3 µM SYTO14 green fluorescent nucleic acid stain (#S7576, Invitrogen) solution is prepared in 1x HBSS, 1% BSA. 30 µL of staining solution is added to wells and incubated for 30 min. Wells ware washed three times with 70 µL 1xHBSS, no final aspiration. Plates are sealed with blue Remp thermal seal, at 171 °C for 4 s.

EMBODIMENTS

Aspects of the present disclosure are directed to a method of altering cells with a pooled library of barcoded nucleic acids including combining a group of cells with the pooled library of bar coded nucleic acids under conditions which promote uptake of one or more barcoded nucleic acids from the pooled library into the cells, analyzing an individual cell for phenotype, sequencing the individual cell to identify the one or more barcoded nucleic acids from the pooled library, wherein the barcodes are functional or nonfunctional, and correlating the phenotype for the individual cell with the one or more barcoded nucleic acids from the pooled library. According to one aspect, the steps of analyzing, sequencing and correlating are conducted for multiple cells within the group of cells. According to one aspect, one or more barcoded nucleic acids from the pooled library are delivered to the cells using viral or non-viral methods. According to one aspect, the individual cell is analyzed for phenotype using automated imaging and one or more barcoded nucleic acids within the individual cell is sequenced using in situ sequencing. According to one aspect, the individual cell is analyzed for phenotype using automated imaging and one or more barcoded nucleic acids within the individual cell is sequenced using FISSEQ. According to one aspect, the cells are analyzed for phenotype using high-content imaged based analysis and one or more barcoded nucleic acids within the individual cell is sequenced using FISSEQ. According to one aspect, the phenotype is correlated with the one or more barcoded nucleic acids from the pooled library using a bioinformatic pipeline to merge cytometric data with sequence information at the single cell level to allow statistical deconvolution of the effect of a nucleic acid from the library with resulting phenotype. According to one aspect, the one or more barcoded nucleic acids are amplified in situ before using in situ sequencing. According to one aspect, the one or more barcoded nucleic acids are amplified in situ using isothermal or non-isothermal amplification before using in situ sequencing. According to one aspect, the one or more

barcoded nucleic acids are delivered using an expression vector where the nucleic acid is flanked by a first common sequence and a second common sequence and promoter which facilitates expression of RNA copies of the nucleic acid, wherein the RNA is reversed transcribed, circularized and amplified by rolling circle amplification before using in situ sequencing. According to one aspect, the cells are eukaryotic cells, prokaryotic cells, animal cells, plant cells, yeast cells, fungal cells, bacteria cells, archaeal cells, or eubacterial cells. According to one aspect, the cells are mammalian cells. According to one aspect, the cells are human cells. According to one aspect, the cells are stem cells, pluripotent stem cells, or human induced pluripotent stem cells. According to one aspect, the cells are of a human organoid, an engineered human organ, an engineered organoid, an embryoid body or a whole organism.

The disclosure provides a method of altering cells with a pooled library of drugs including combining a group of cells with the pooled library of drugs under conditions which promote uptake of one or more drugs from the pooled library into the cells, analyzing an individual cell for phenotype, analyzing the individual cell to identify the one or more drugs from the pooled library, and correlating the phenotype for the individual cell with the one or more drugs from the pooled library. According to one aspect, the cells are eukaryotic cells, prokaryotic cells, animal cells, plant cells, yeast cells, fungal cells, bacteria cells, archaeal cells, or eubacterial cells. According to one aspect, the cells are mammalian cells. According to one aspect, the cells are human cells. According to one aspect, the cells are stem cells, pluripotent stem cells, or human induced pluripotent stem cells. According to one aspect, the cells are of a human organoid, an engineered human organ, an engineered organoid, an embryoid body or a whole organism.

The disclosure provides a method of altering cells with a pooled library of nucleic acids having fluorescent barcodes including combining a group of cells with the pooled library of nucleic acids having fluorescent barcodes under conditions which promote uptake of one or more nucleic acids having fluorescent barcodes from the pooled library into the cells, analyzing an individual cell for phenotype, analyzing the individual cell to identify the fluorescent barcode, and correlating the phenotype for the individual cell with the fluorescent barcode.

The disclosure provides a method of altering stem cells with a pooled library of bar coded nucleic acids Including combining a group of stem cells with the pooled library of bar coded nucleic acids under conditions which promote uptake of one or more barcoded nucleic acids from the pooled library into the stem cells, culturing the stem cells into embryoid bodies or organoids, analyzing individual cells from the embryoid bodies or organoids for phenotype, sequencing the individual cells to identify the one or more barcoded nucleic acids from the pooled library, wherein the barcodes are functional or nonfunctional, and correlating the phenotype for the individual cells with the one or more barcoded nucleic acids from the pooled library. According to one aspect, one or more barcoded nucleic acids from the pooled library are delivered to the cells using viral or non-viral methods. According to one aspect, the individual cells are analyzed for phenotype using automated imaging and one or more barcoded nucleic acids within the individual cells is sequenced using in situ sequencing. According to one aspect, the individual cells are analyzed for phenotype using automated imaging and one or more barcoded nucleic acids within the individual cells is sequenced using FISSEQ. According to one aspect, the individual cells are analyzed for phenotype using high-content imaged based analysis and one or more barcoded nucleic acids within the individual cells is sequenced using in situ sequencing. According to one aspect, the

phenotype is correlated with the one or more barcoded nucleic acids from the pooled library using a bioinformatic pipeline to merge cytometric data with sequence information at the single cell level to allow statistical deconvolution of the effect of a nucleic acid from the library with resulting phenotype. According to one aspect, the one or more barcoded nucleic acids are amplified in situ before using in situ sequencing. According to one aspect, the one or more barcoded nucleic acids are amplified in situ using isothermal or non-isothermal amplification before using in situ sequencing. According to one aspect, the one or more barcoded nucleic acids are delivered using an expression vector where the nucleic acid is flanked by a first common sequence and a second common sequence and promoter which facilitates expression of RNA copies of the nucleic acid, wherein the RNA is reversed transcribed, circularized and amplified by rolling circle amplification before using in situ sequencing.

The disclosure provides a method of altering cells of an organoid or embryoid body with a pooled library of bar coded nucleic acids including combining the organoid or embryoid body with the pooled library of bar coded nucleic acids under conditions which promote uptake of one or more barcoded nucleic acids from the pooled library into individual cells of the organoid or embryoid body, analyzing the individual cells of the organoid or embryoid body for phenotype, sequencing the individual cells of the organoid or embryoid body to identify the one or more barcoded nucleic acids from the pooled library, wherein the barcodes are functional or nonfunctional, and correlating the phenotype for the individual cells of the organoid or embryoid body with the one or more barcoded nucleic acids from the pooled library. According to one aspect, one or more barcoded nucleic acids from the pooled library are delivered to the individual cells of the organoid or embryoid body using viral or non-viral methods. According to one aspect, the individual cells of the organoid or embryoid

body are analyzed for phenotype using automated imaging and one or more barcoded nucleic acids within the individual cells of the organoid or embryoid body is sequenced using in situ sequencing. According to one aspect, the individual cells of the organoid or embryoid body are analyzed for phenotype using high-content imaged based analysis and one or more barcoded nucleic acids within the individual cells of the organoid or embryoid body is sequenced using in situ sequencing. According to one aspect, the phenotype is correlated with the one or more barcoded nucleic acids from the pooled library using a bioinformatic pipeline to merge cytometric data with sequence information at the single cell level to allow statistical deconvolution of the effect of a nucleic acid from the library with resulting phenotype. According to one aspect, the one or more barcoded nucleic acids are amplified in situ before using in situ sequencing. According to one aspect, the one or more barcoded nucleic acids are amplified in situ using isothermal or non-isothermal amplification before using in situ sequencing. According to one aspect, the one or more barcoded nucleic acids are delivered using an expression vector where the nucleic acid is flanked by a first common sequence and a second common sequence and promoter which facilitates expression of RNA copies of the nucleic acid, wherein the RNA is reversed transcribed, circularized and amplified by rolling circle amplification before using in situ sequencing.

The disclosure provides a method of altering cells of a plurality of embryoid bodies with a pooled library of bar coded nucleic acids including combining the embryoid bodies which are placed within a grid with the pooled library of bar coded nucleic acids under conditions which promote uptake of one or more barcoded nucleic acids from the pooled library into individual cells of the embryoid bodies, analyzing the individual cells of the embryoid bodies for phenotype for developing organoids or systems of organs, sequencing the individual cells of the embryoid bodies to identify the one or more barcoded nucleic acids

from the pooled library, wherein the barcodes are functional or nonfunctional, and correlating the phenotype for the individual cells of the embryoid bodies with the one or more barcoded nucleic acids from the pooled library.

The disclosure provides a method of altering cells of a plurality of embryoid bodies with a pooled library of bar coded nucleic acids including combining the embryoid bodies which are placed within a grid with the pooled library of bar coded nucleic acids under conditions which promote uptake of one or more barcoded nucleic acids from the pooled library into individual cells of the embryoid bodies, culturing the embryoid bodies to form organoids, analyzing individual cells of the organoids for phenotype, sequencing the individual cells of the organoids to identify the one or more barcoded nucleic acids from the pooled library, wherein the barcodes are functional or nonfunctional, and correlating the phenotype for the individual cells of the organoids with the one or more barcoded nucleic acids from the pooled library.

REFERENCES

The following references are incorporated herein by reference in their entireties.

1. Buchser, W. *et al.* Assay Guidance Manual. (2004).

2. Bray, M. A. & Carpenter, A. Assay Guidance Manual. (2004).

3. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343,** 80–84 (2014).

4. Gilbert, L. A. *et al.* Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159,** 647–661 (2014).

5. Shalem, O., Sanjana, N. E. & Zhang, F. High-throughput functional genomics using CRISPR-Cas9. *Nat Rev Genet* **16,** 299–311 (2015).

6. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343,** 84–87 (2014).

7. Agrotis, A. & Ketteler, R. A new age in functional genomics using CRISPR/Cas9 in arrayed library screening. *Front Genet* **6,** 300 (2015).

8. Lee, J. H. *et al.* Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat Protoc* **10,** 442–458 (2015).

9. Lee, J. H. *et al.* Highly multiplexed subcellular RNA sequencing in situ. *Science* **343,** 1360–1363 (2014).

10. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348,** aaa6090 (2015).

11. Ke, R. *et al.* In situ sequencing for RNA analysis in preserved tissue and cells. *Nat Methods* **10,** 857–860 (2013).

12. Nayerossadat, N., Maedeh, T. & Ali, P. A. Viral and nonviral delivery systems for gene delivery. *Adv Biomed Res* **1,** 27 (2012).

13. Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol* **7,** R100 (2006).

14. Crosetto, N., Bienko, M. & van Oudenaarden, A. Spatially resolved transcriptomics and beyond. *Nat Rev Genet* **16,** 57–66 (2015).

15. Larsson, C., Grundberg, I., Söderberg, O. & Nilsson, M. In situ detection and genotyping of individual mRNA molecules. *Nat Methods* **7,** 395–397 (2010).

16. Weibrecht, I. *et al.* In situ detection of individual mRNA molecules and protein complexes or post-translational modifications using padlock probes combined with the in situ proximity ligation assay. *Nat Protoc* **8,** 355–372 (2013).

17. Chen, R. *et al.* A Barcoding Strategy Enabling Higher-Throughput Library Screening by Microscopy. *ACS Synth Biol* **4,** 1205–1216 (2015).

18. Bagasra, O. Protocols for the in situ PCR-amplification and detection of mRNA and DNA sequences. *Nat Protoc* **2,** 2782–2795 (2007).

19. Mitra, R. D. & Church, G. M. In situ localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Res* **27,** e34 (1999).

20. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517,** 583–588 (2015).

21. Harkness, T. *et al.* High-Content Imaging with Micropatterned Multiwell Plates Reveals Influence of Cell Geometry and Cytoskeleton on Chromatin Dynamics. *Biotechnol J* (2015). doi:10.1002/biot.201400756

22. Bao, G., Mitragotri, S. & Tong, S. Multifunctional nanoparticles for drug delivery and molecular imaging. *Annu Rev Biomed Eng* **15,** 253–282 (2013).

23. Meinhardt, A. *et al.* 3D reconstitution of the patterned neural tube from embryonic stem cells. *Stem Cell Reports* **3,** 987–999 (2014).

24. Mitra, R. D., Shendure, J., Olejnik, J., Edyta-Krzymanska-Olejnik & Church, G. M. Fluorescent in situ sequencing on polymerase colonies. *Anal Biochem* **320,** 55–65 (2003).

Claims:

1. A method of altering cells with a pooled library of barcoded nucleic acids comprising

combining a group of cells with the pooled library of bar coded nucleic acids under conditions which promote uptake of one or more barcoded nucleic acids from the pooled library into the cells,

analyzing an individual cell for phenotype,

sequencing the individual cell to identify the one or more barcoded nucleic acids from the pooled library, wherein the barcodes are functional or nonfunctional, and

correlating the phenotype for the individual cell with the one or more barcoded nucleic acids from the pooled library.

2. The method of claim 1 wherein the steps of analyzing, sequencing and correlating are conducted for multiple cells within the group of cells.

3. The method of claim 1 wherein one or more barcoded nucleic acids from the pooled library are delivered to the cells using viral or non-viral methods.

4. The method of claim 1 wherein the individual cell is analyzed for phenotype using automated imaging and one or more barcoded nucleic acids within the individual cell is sequenced using in situ sequencing.

5. The method of claim 1 wherein the individual cell is analyzed for phenotype using automated imaging and one or more barcoded nucleic acids within the individual cell is sequenced using FISSEQ.

6. The method of claim 1 wherein the cells are analyzed for phenotype using high-content imaged based analysis and one or more barcoded nucleic acids within the individual cell is sequenced using FISSEQ.

7. The method of claim 1 wherein the phenotype is correlated with the one or more barcoded nucleic acids from the pooled library using a bioinformatic pipeline to merge cytometric data with sequence information at the single cell level to allow statistical deconvolution of the effect of a nucleic acid from the library with resulting phenotype.

8. The method of claim 1 wherein the one or more barcoded nucleic acids are amplified in situ before using in situ sequencing.

9. The method of claim 1 wherein the one or more barcoded nucleic acids are amplified in situ using isothermal or non-isothermal amplification before using in situ sequencing.

10. The method of claim 1 wherein the one or more barcoded nucleic acids are delivered using an expression vector where the nucleic acid is flanked by a first common sequence and a second common sequence and promoter which facilitates expression of RNA

copies of the nucleic acid, wherein the RNA is reversed transcribed, circularized and amplified by rolling circle amplification before using in situ sequencing.

11. The method of claim 1 wherein the cells are eukaryotic cells, prokaryotic cells, animal cells, plant cells, yeast cells, fungal cells, bacteria cells, archaeal cells, or eubacterial cells.

12. The method of claim 1 wherein the cells are mammalian cells.

13. The method of claim 1 wherein the cells are human cells.

14. The method of claim 1 wherein the cells are stem cells, pluripotent stem cells, or human induced pluripotent stem cells.

15. The method of claim 1 wherein the cells are of a human organoid, an engineered human organ, an engineered organoid, an embryoid body or a whole organism.

16. A method of altering cells with a pooled library of drugs comprising

combining a group of cells with the pooled library of drugs under conditions which promote uptake of one or more drugs from the pooled library into the cells,

analyzing an individual cell for phenotype,

analyzing the individual cell to identify the one or more drugs from the pooled library, and

correlating the phenotype for the individual cell with the one or more drugs from the pooled library.

17. The method of claim 16 wherein the cells are eukaryotic cells, prokaryotic cells, animal cells, plant cells, yeast cells, fungal cells, bacteria cells, archaeal cells, or eubacterial cells.

18. The method of claim 16 wherein the cells are mammalian cells.

19. The method of claim 16 wherein the cells are human cells.

20. The method of claim 16 wherein the cells are stem cells, pluripotent stem cells, or human induced pluripotent stem cells.

21. The method of claim 16 wherein the cells are of a human organoid, an engineered human organ, an engineered organoid, an embryoid body or a whole organism.

22. A method of altering cells with a pooled library of nucleic acids having fluorescent barcodes comprising

combining a group of cells with the pooled library of nucleic acids having fluorescent barcodes under conditions which promote uptake of one or more nucleic acids having fluorescent barcodes from the pooled library into the cells,

analyzing an individual cell for phenotype,

analyzing the individual cell to identify the fluorescent barcode, and

correlating the phenotype for the individual cell with the fluorescent barcode.

23. A method of altering stem cells with a pooled library of bar coded nucleic acids comprising

combining a group of stem cells with the pooled library of bar coded nucleic acids under conditions which promote uptake of one or more barcoded nucleic acids from the pooled library into the stem cells,

culturing the stem cells into embryoid bodies or organoids,

analyzing individual cells from the embryoid bodies or organoids for phenotype,

sequencing the individual cells to identify the one or more barcoded nucleic acids from the pooled library, wherein the barcodes are functional or nonfunctional, and

correlating the phenotype for the individual cells with the one or more barcoded nucleic acids from the pooled library.

24. The method of claim 23 wherein one or more barcoded nucleic acids from the pooled library are delivered to the cells using viral or non-viral methods.

25. The method of claim 23 wherein the individual cells are analyzed for phenotype using automated imaging and one or more barcoded nucleic acids within the individual cells is sequenced using in situ sequencing.

26. The method of claim 23 wherein the individual cells are analyzed for phenotype using automated imaging and one or more barcoded nucleic acids within the individual cells is sequenced using FISSEQ.

27. The method of claim 23 wherein the individual cells are analyzed for phenotype using high-content imaged based analysis and one or more barcoded nucleic acids within the individual cells is sequenced using in situ sequencing.

28. The method of claim 23 wherein the phenotype is correlated with the one or more barcoded nucleic acids from the pooled library using a bioinformatic pipeline to merge cytometric data with sequence information at the single cell level to allow statistical deconvolution of the effect of a nucleic acid from the library with resulting phenotype.

29. The method of claim 23 wherein the one or more barcoded nucleic acids are amplified in situ before using in situ sequencing.

30. The method of claim 23 wherein the one or more barcoded nucleic acids are amplified in situ using isothermal or non-isothermal amplification before using in situ sequencing.

31. The method of claim 23 wherein the one or more barcoded nucleic acids are delivered using an expression vector where the nucleic acid is flanked by a first common sequence and a second common sequence and promoter which facilitates expression of RNA copies of the nucleic acid, wherein the RNA is reversed transcribed, circularized and amplified by rolling circle amplification before using in situ sequencing.

32. A method of altering cells of an organoid or embryoid body with a pooled library of bar coded nucleic acids comprising

combining the organoid or embryoid body with the pooled library of bar coded nucleic acids under conditions which promote uptake of one or more barcoded nucleic acids from the pooled library into individual cells of the organoid or embryoid body,

analyzing the individual cells of the organoid or embryoid body for phenotype,

sequencing the individual cells of the organoid or embryoid body to identify the one or more barcoded nucleic acids from the pooled library, wherein the barcodes are functional or nonfunctional, and

correlating the phenotype for the individual cells of the organoid or embryoid body with the one or more barcoded nucleic acids from the pooled library.

33. The method of claim 32 wherein one or more barcoded nucleic acids from the pooled library are delivered to the individual cells of the organoid or embryoid body using viral or non-viral methods.

34. The method of claim 32 wherein the individual cells of the organoid or embryoid body are analyzed for phenotype using automated imaging and one or more barcoded nucleic acids within the individual cells of the organoid or embryoid body is sequenced using in situ sequencing.

35. The method of claim 32 wherein the individual cells of the organoid or embryoid body are analyzed for phenotype using high-content imaged based analysis and one

or more barcoded nucleic acids within the individual cells of the organoid or embryoid body is sequenced using in situ sequencing.

36. The method of claim 32 wherein the phenotype is correlated with the one or more barcoded nucleic acids from the pooled library using a bioinformatic pipeline to merge cytometric data with sequence information at the single cell level to allow statistical deconvolution of the effect of a nucleic acid from the library with resulting phenotype.

37. The method of claim 32 wherein the one or more barcoded nucleic acids are amplified in situ before using in situ sequencing.

38. The method of claim 32 wherein the one or more barcoded nucleic acids are amplified in situ using isothermal or non-isothermal amplification before using in situ sequencing.

39. The method of claim 32 wherein the one or more barcoded nucleic acids are delivered using an expression vector where the nucleic acid is flanked by a first common sequence and a second common sequence and promoter which facilitates expression of RNA copies of the nucleic acid, wherein the RNA is reversed transcribed, circularized and amplified by rolling circle amplification before using in situ sequencing.

40. A method of altering cells of a plurality of embryoid bodies with a pooled library of bar coded nucleic acids comprising

combining the embryoid bodies which are placed within a grid with the pooled library of bar coded nucleic acids under conditions which promote uptake of one or more barcoded nucleic acids from the pooled library into individual cells of the embryoid bodies,

analyzing the individual cells of the embryoid bodies for phenotype for developing organoids or systems of organs,

sequencing the individual cells of the embryoid bodies to identify the one or more barcoded nucleic acids from the pooled library, wherein the barcodes are functional or nonfunctional, and

correlating the phenotype for the individual cells of the embryoid bodies with the one or more barcoded nucleic acids from the pooled library.

41. A method of altering cells of a plurality of embryoid bodies with a pooled library of bar coded nucleic acids comprising

combining the embryoid bodies which are placed within a grid with the pooled library of bar coded nucleic acids under conditions which promote uptake of one or more barcoded nucleic acids from the pooled library into individual cells of the embryoid bodies,

culturing the embryoid bodies to form organoids,

analyzing individual cells of the organoids for phenotype,

sequencing the individual cells of the organoids to identify the one or more barcoded nucleic acids from the pooled library, wherein the barcodes are functional or nonfunctional, and

correlating the phenotype for the individual cells of the organoids with the one or more barcoded nucleic acids from the pooled library.

FIG. 1

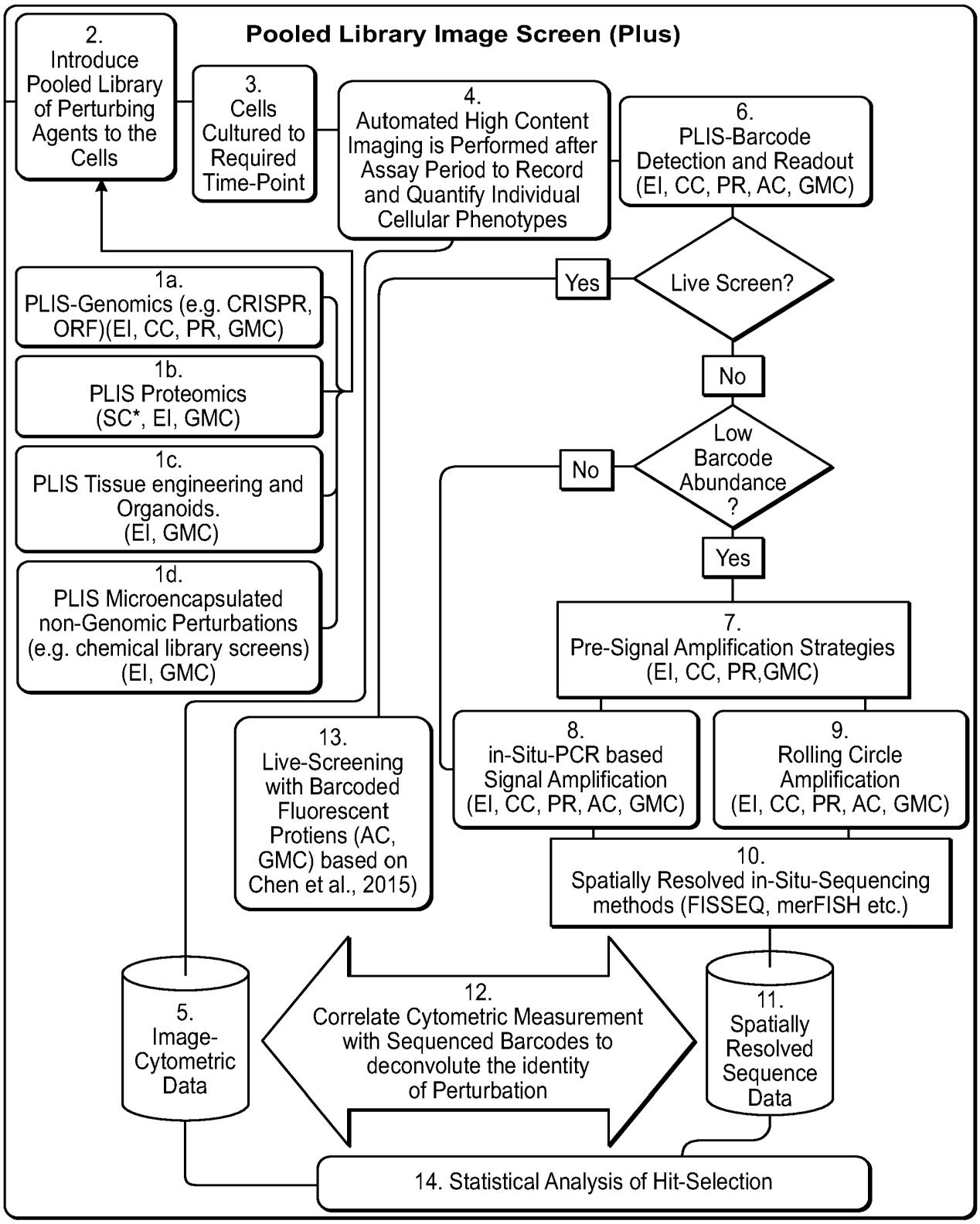

Pooled Library Image Screen (Plus)
2. Introduce Pooled Library of Perturbing Agents to the Cells
3. Cells Cultured to Required Time-Point
4. Automated High Content Imaging is Performed after Assay Period to Record and Quantify Individual Cellular Phenotypes
6. PLIS-Barcode Detection and Readout (EI, CC, PR, AC, GMC)
Live Screen?
Yes
No
1a. PLIS-Genomics (e.g. CRISPR, ORF)(EI, CC, PR, GMC)
1b. PLIS Proteomics (SC*, EI, GMC)
1c. PLIS Tissue engineering and Organoids. (EI, GMC)
1d. PLIS Microencapsulated non-Genomic Perturbations (e.g. chemical library screens) (EI, GMC)
Low Barcode Abundance ?
No
Yes
7. Pre-Signal Amplification Strategies (EI, CC, PR,GMC)
13. Live-Screening with Barcoded Fluorescent Protiens (AC, GMC) based on Chen et al., 2015)
8. in-Situ-PCR based Signal Amplification (EI, CC, PR, AC, GMC)
9. Rolling Circle Amplification (EI, CC, PR, AC, GMC)
10. Spatially Resolved in-Situ-Sequencing methods (FISSEQ, merFISH etc.)
5. Image-Cytometric Data
12. Correlate Cytometric Measurement with Sequenced Barcodes to deconvolute the identity of Perturbation
11. Spatially Resolved Sequence Data
14. Statistical Analysis of Hit-Selection

FIG. 2A

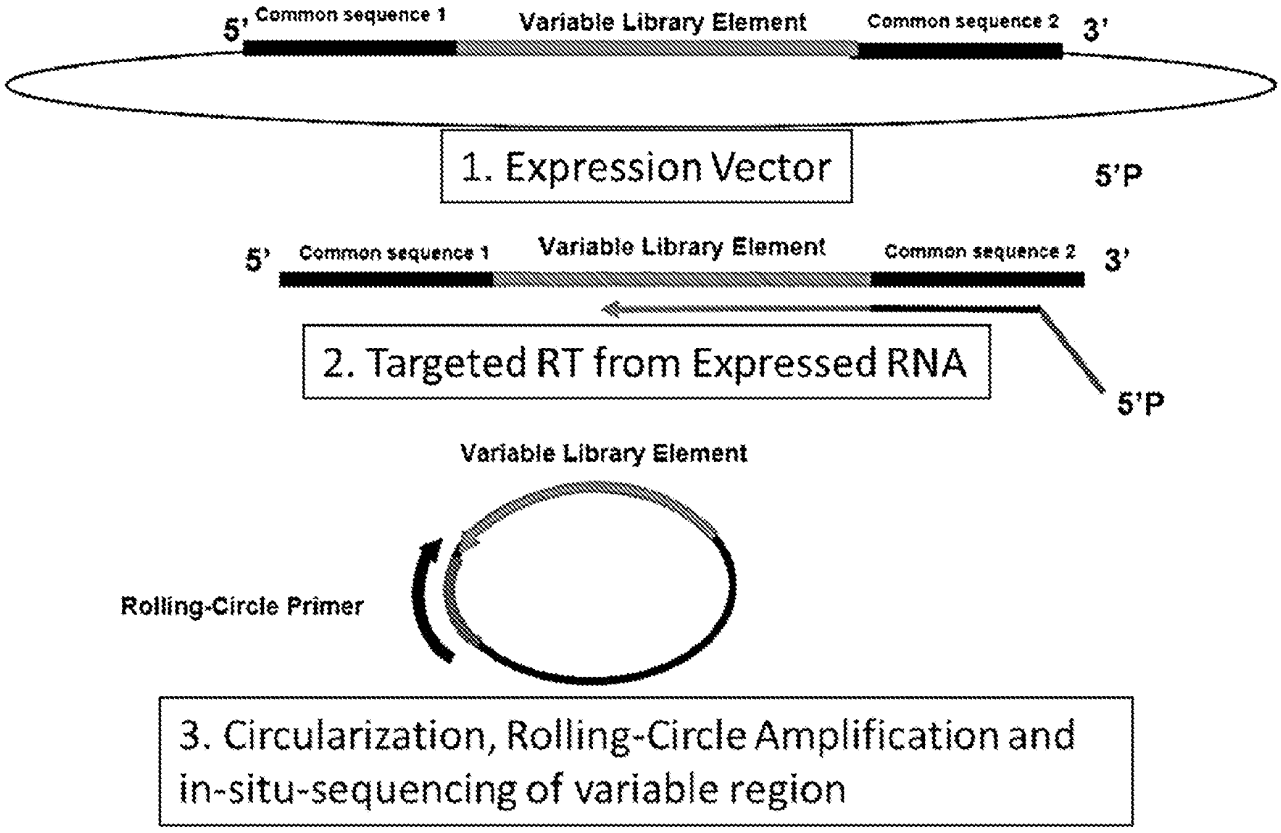

5' Common sequence 1
Variable Library Element
Common sequence 2 3'
1. Expression Vector
5'P
5' Common sequence 1
Variable Library Element
Common sequence 2 3'
2. Targeted RT from Expressed RNA
5'P
Variable Library Element
Rolling-Circle Primer
3. Circularization, Rolling-Circle Amplification and in-situ-sequencing of variable region

FIG. 2B

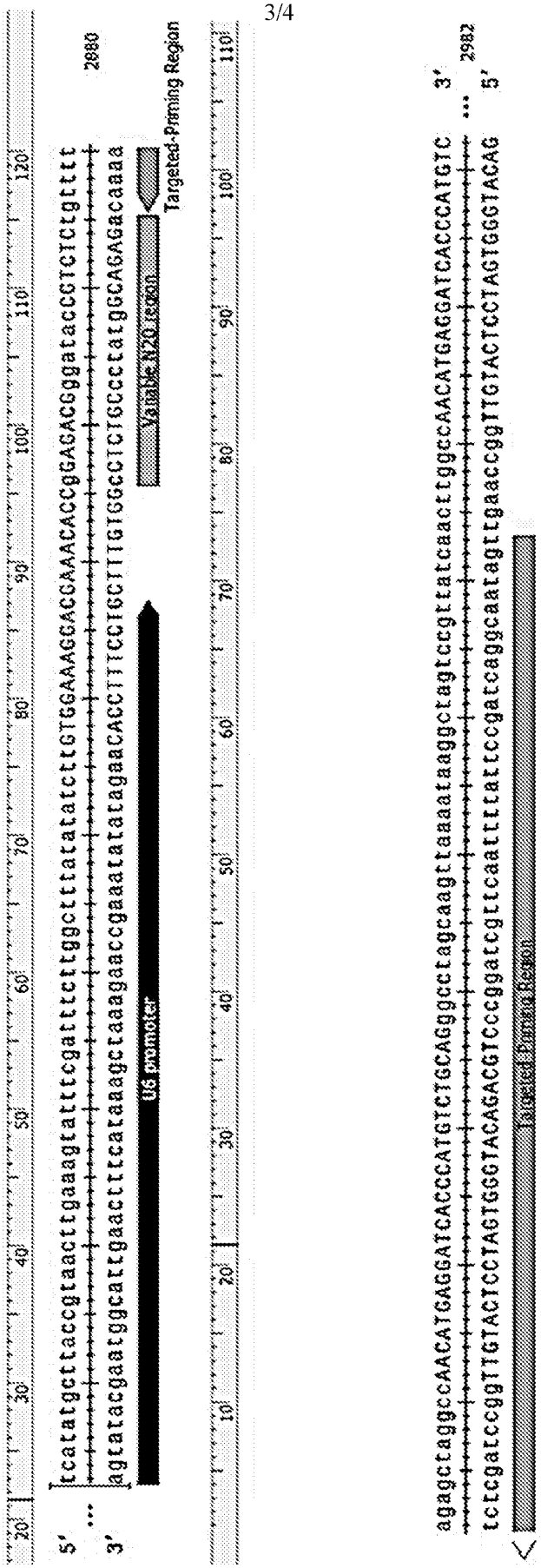
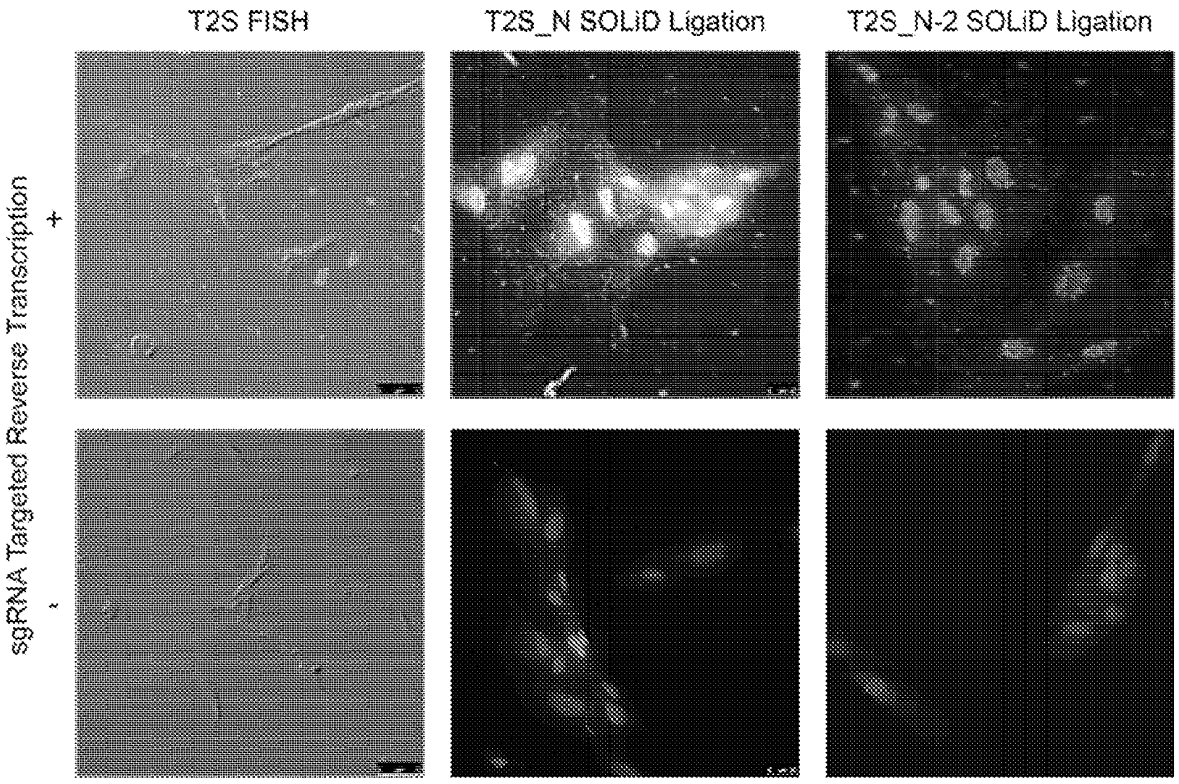

5'
3'
2880
U6 promoter
Variable N20 region
Targeted-Priming Region
2982
3'
5'
Targeted Priming Region

FIG. 3


T2S FISH
T2S_N SOLiD Ligation
T2S_N-2 SOLiD Ligation
sgRNA Targeted Reverse Transcription
+
-

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
17/395,534	08/06/2021	Evan R. Daugharthy	010498.01437\US	5147

7590 08/11/2022

BANNER & WITCOFF, LTD.
1100 13th STREET, N.W.
SUITE 1200
WASHINGTON, DC 20005-4051

EXAMINER	
THOMAS, DAVID C	

ART UNIT	PAPER NUMBER
1637	

NOTIFICATION DATE	DELIVERY MODE
08/11/2022	ELECTRONIC

NOTICE OF NON-COMPLIANT INFORMATION DISCLOSURE STATEMENT

An Information Disclosure Statement (IDS) filed 8.3.2022 in the above-identified application fails to meet the requirements of 37 CFR 1.97(d) for the reason(s) specified below. Accordingly, the IDS will be placed in the file, but the information referred to therein has not been considered.

The IDS is not compliant with 37 CFR 1.97(d) because:

☒ The IDS lacks a statement as specified in 37 CFR 1.97(e).

☐ The IDS lacks the fee set forth in 37 CFR 1.17(p).

☐ The IDS was filed after the issue fee was paid. Applicant may wish to consider filing a petition to withdraw the application from issue under 37 CFR 1.313(c) to have the IDS considered. See MPEP 1308.

N. Horne, for

571-272-4200 or 1-888-786-0101
Application Assistance Unit
Office of Data Management

FORM PTOM327-B (Rev. 02/08)

Electronic Patent Application Fee Transmittal

Application Number:	17395534
Filing Date:	06-Aug-2021
Title of Invention:	Method and Apparatus for Volumetric Imaging
First Named Inventor/Applicant Name:	Evan R. Daugharthy
Filer:	Pei Wu/Jennifer Hazzard
Attorney Docket Number:	010498.01437\US

Filed as Large Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
CLAIMS IN EXCESS OF 20	1202	34	100	3400
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
UTILITY APPL ISSUE FEE	1501	1	1200	1200
Extension-of-Time:				
Miscellaneous:				
		Total in USD ($)		**4600**

Electronic Acknowledgement Receipt	
EFS ID:	46652522
Application Number:	17395534
International Application Number:	
Confirmation Number:	5147
Title of Invention:	Method and Apparatus for Volumetric Imaging
First Named Inventor/Applicant Name:	Evan R. Daugharthy
Customer Number:	22907
Filer:	Pei Wu/Jennifer Hazzard
Filer Authorized By:	Pei Wu
Attorney Docket Number:	010498.01437\US
Receipt Date:	20-SEP-2022
Filing Date:	06-AUG-2021
Time Stamp:	12:52:23
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	DA
Payment was successfully received in RAM	$4600
RAM confirmation Number	E20229JC53230644
Deposit Account	190733
Authorized User	Jennifer Hazzard
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows: 37 CFR 1.16 (National application filing, search, and examination fees) 37 CFR 1.17 (Patent application and reexamination processing fees)	

37 CFR 1.19 (Document supply fees)

37 CFR 1.20 (Post Issuance fees)

37 CFR 1.21 (Miscellaneous fees and charges)

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Issue Fee Payment (PTO-85B)	Issue_Fee_Transmittal_22125352.pdf	38062 7419bad97f2f235663532f37fad475211c63513c	no	1
Warnings:					
Information:					
2	Amendment after Notice of Allowance (Rule 312)	312_Amendment_after_Allowance_22125382.pdf	45038 25b1f330366d5f4c951a5f0d919d0d804e008318	no	14
Warnings:					
Information:					
3	Fee Worksheet (SB06)	fee-info.pdf	40795 9b1f888a6ec82678921adc2b079c09e513c90de0	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	123895		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

<u>New Applications Under 35 U.S.C. 111</u>
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.
<u>National Stage of an International Application under 35 U.S.C. 371</u>
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.
<u>New International Application Filed with the USPTO as a Receiving Office</u>
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

PART B–FEE(S) TRANSMITTAL

Complete and send this form, together with the applicable fee(s), by mail or fax, or via EFS-Web.

By mail, send to: Mail Stop ISSUE FEE
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

By fax, send to: (571) 273-2885

INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 5 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Use Block 1 for any change of address)

BANNER & WITCOFF, LTD.
28 State Street
Suite 1800
Boston, Massachusetts 02109

Note: A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

Certificate of Mailing or Transmission

I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Mail Stop ISSUE FEE address above, or being transmitted to the USPTO via EFS-Web or by facsimile to (571) 273-2885, on the date below.

(Typed or printed name)

(Signature)

(Date)

APPLICATION NO	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO	CONFIRMATION NO
17/395,534	08/06/2021	Evan R. Daugharthy	010498.01437\US	5147

TITLE OF INVENTION: Method and Apparatus for Volumetric Imaging

APPLN: TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	UNDISCOUNTED	$1,200.00			$1,200.00	09/21/2022

EXAMINER	ART UNIT	CLASS-SUBCLASS
D. C. Thomas	1637	

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363)

☐ Change of correspondence address (or Change of Correspondence Address form PTO/SB/122) attached.

☐ "Fee Address" indication (or "Fee Address" Indication form PTO/SB/47; Rev 03-09 or more recent) attached. **Use of a Customer Number is required.**

2. For printing on the patent front page, list

(1) The names of up to 3 registered patent attorneys or agents OR, alternatively,

(2) The name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed.

1. Banner & Witcoff, Ltd.
2.
3.

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. If an assignee is identified below, the document must have been previously recorded, or filed for recordation, as set forth in 37 CFR 3.11 and 37 CFR 3.81(a). Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE: President and Fellows of Harvard College

(B) RESIDENCE: (CITY and STATE or COUNTRY): Cambridge, Massachusetts

Please check the appropriate assignee category or categories (will not be printed on the patent): ☐ Individual ☒ Corporation or other private group entity ☐ Government

4a. Fees Submitted: ☒ Issue Fee ☐ Publication Fee (if required) ☐ Advance Order - # of Copies _____

4b. Method of Payment (*Please first reapply any previously paid fee shown above*):

☒ Electronic Payment via EFS-Web ☐ Enclosed check ☐ Non-electronic payment by credit card (Attach form PTO-2038)

☒ The Director is hereby authorized to charge the required fee(s), any deficiency, or credit any overpayment to Deposit Account No. 19-0733

5. Change of Entity Status (from status indicated above)

☐ Applicant certifying micro entity status. See 37 CFR 1.29. NOTE: Absent a valid Certification of Micro Entity Status (see forms PTO/SB/15A and 15B), issue fee payment in the micro entity amount will not be accepted at the risk of application abandonment.

☐ Applicant asserting small entity status. See 37 CFR 1.27. NOTE: If the application was previously under micro entity status, checking this box will be taken as a notification of loss of entitlement to micro entity status.

☐ Applicant changing to regular undiscounted fee status. NOTE: Checking this box will be taken as a notification of loss of entitlement to small or micro entity status, as applicable.

NOTE: This form must be signed in accordance with 37 CFR 1.31 and 1.33. See 37 CFR 1.4 for signature requirements and certifications.

Authorized Signature /John P. Iwanicki/ Date September 20, 2022

Typed or printed name John P. Iwanicki Registration No. 34,628

PATENT
ATTORNEY DOCKET NO. 010498.01437\US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:	)	
	)	
Evan R. Daugharthy	)	Examiner: David C. Thomas
	)	
Serial No.: 17/395,534	)	Art Unit: 1637
	)	
Filed: August 6, 2021	)	Conf. No.: 5147
	)	
Title: Method and Apparatus for Volumetric Imaging of a Three-Dimensional Nucleic Acid Containing Matrix	)	

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

AMENDMENT AFTER ALLOWANCE UNDER 37 C.F.R. § 1.312

Dear Commissioner:

In response to the Notice of Allowance mailed June 21, 2022, and prior to payment of the issue fee and issuance of the patent, Applicant respectfully requests entry of this amendment under 37 C.F.R. 1.312 for the above-captioned patent application. The Commissioner is hereby authorized to charge such extension fee or credit any overpayment of fees for this submission to Deposit Account No. 19-0733.

A Summary of the Interview begins at page 2 of this paper

Amendments to the Claims begin at page 3 of this paper.

Remarks begin at page 14 of this paper.

SUMMARY OF THE INTERVIEW

Applicant is appreciative of Primary Examiner David C. Thomas for extending the courtesy of a telephonic interview to Applicant's representative, Pei Wu, on September 19, 2022. During the interview, the claim amendments submitted herewith were discussed.

Amendments to the Claims

The following claim listing replaces all prior claim listings in the subject application.

Listing of Claims:

1. – 8. (Cancelled)

9. (Previously Presented) A system, comprising:

(a) a stage comprising:

(i) a sample holder configured to receive a substrate comprising a sample having a volume,

(ii) a temperature control apparatus operatively coupled to said sample holder; and

(iii) an XY multi-axis motion control system;

(b) a fluidics system comprising:

(i) one or more reservoirs configured to contain one or more fluid reagents; and

(ii) a fluid dispenser operationally coupled to a pump, which pump is configured to provide said one or more fluid reagents from said one or more reservoirs to said sample holder; and

(c) an optical system comprising:

(i) an objective;

(ii) one or more light sources, wherein each light source of said one or more light sources is configured to deliver light to said volume; and

(iii) one or more detectors configured to receive light signals from said volume; and

(d) a Z-axis motion control system configured for Z-motion of an objective focal plane through said volume; and

(e) a computing system comprising at least one computer readable storage medium having program instructions stored thereon, which program instructions are executable by at least one processor of said computing system to cause said at least one processor to perform a method comprising:

determining a plurality of optically-encoded signals from a plurality of light signals, wherein said plurality of light signals is obtained from a plurality of read-out cycles, wherein said plurality of optically-encoded signals is associated with one or more nucleic acid molecules

within said volume; and

determining three-dimensional positional information of said one or more nucleic acid molecules using said plurality of optically-encoded signals.

10. (Previously Presented) The system of claim 9, wherein said substrate comprises a flow cell.

11. (Previously Presented) The system of claim 10, wherein said flow cell is a closed flow cell.

12. (Previously Presented) The system of claim 11, wherein said closed flow cell comprises a fluid inlet and a fluid outlet, and wherein said fluid inlet is fluidically coupled to said fluid dispenser.

13. (Previously Presented) The system of claim 10, wherein said flow cell is an open flow cell.

14. (Previously Presented) The system of claim 9, wherein said objective is a water immersion objective.

15. (Previously Presented) The system of claim 9, wherein said objective is a water dipping objective.

16. (Previously Presented) The system of claim 9, wherein said temperature control apparatus comprises a heat sink or a Peltier element.

17. (Currently Amended) The system of claim 9, further comprising a thermal control system.

18. (Cancelled)

19. (Previously Presented) The system of claim 9, wherein said pump is a syringe pump.

20. (Previously Presented) The system of claim 9, wherein said fluidics system further comprises one or more valves downstream of said one or more reservoirs and upstream of said fluid dispenser, wherein, upon actuation, said one or more valves facilitate dispensation of fluid to said sample holder.

21. (Previously Presented) The system of claim 20, wherein said one or more valves comprise a rotary valve.

22. (Previously Presented) The system of claim 20, wherein said one or more valves comprise a solenoid valve or a microfluidic valve.

23. (Previously Presented) The system of claim 13, wherein said fluidics system further comprises an aspiration system comprising: (1) an aspiration tube configured to remove fluid from said open flow cell and a pressure source operatively coupled to said aspiration tube.

24. (Previously Presented) The system of claim 23, wherein said aspiration system further comprises a motor operatively coupled to said aspiration tube, wherein said motor is configured to adjust a motion axis of said aspiration tube to control an interface between said aspiration tube and said open flow cell.

25. (Previously Presented) The system of claim 24, wherein said motor is a stepper motor.

26. (Currently Amended) The system of claim 9, wherein said Z-axis motion control system comprises a Z-axis linear motor configured for Z-motion of said objective focal plane through said volume.

27. (Currently Amended) The system of claim 26, further comprising a linear encoder

configured to provide Z-motion feedback to said Z-axis linear motor.

28. (Previously Presented) The system of claim 9, further comprising an objective wetting apparatus.

29. (Currently Amended) The system of claim 9, wherein said one or more light sources comprise[[s]] one or more light emitting diodes (LEDs).

30. (Currently Amended) The system of claim 9, wherein said one or more light sources comprise[[s]] one or more lasers.

31. (Previously Presented) The system of claim 9, wherein said optical system further comprises one or more excitation filters configured to selectively filter one or more wavelengths of light from said one or more light sources to said volume.

32. (Previously Presented) The system of claim 31, wherein at least one excitation filter of said one or more excitation filters is configured to selectively filter light having a wavelength of about 480 nm, about 530 nm, about 590 nm, or about 640 nm from said one or more light sources to said volume.

33. (Previously Presented) The system of claim 9, wherein said optical system further comprises one or more emission filters configured to selectively filter one or more wavelengths of light emitted from said volume to said one or more detectors.

34. (Previously Presented) The system of claim 9, wherein said optical system is mechanically, electrically, or electromechanically coupled to a controller system configured to synchronize one or more imaging events within said system.

35. (Previously Presented) The system of claim 9, further comprising a vibration control rigid foundation.

36. (Previously Presented) The system of claim 9, further comprising a fluidics controller operatively coupled to said fluid dispenser and configured to automate or control dispensation of said one or more fluid reagents to said sample holder.

37. (Previously Presented) The system of claim 17, wherein said temperature control apparatus further comprises a temperature sensor configured to provide a feedback signal to said thermal control system.

38. (Previously Presented) The system of claim 33, wherein at least one of said one or more emission filters is configured to selectively filter light having a wavelength of about 510 nm, about 570 nm, about 620 nm, or about 680 nm to said one or more detectors from said volume to said one or more detectors.

39. (Currently Amended) The system of claim 9, wherein said one or more light sources comprises two or more light sources, wherein said two or more light sources excite fluorescence in unique color channels such that~~wherein~~ each light source of said two or more light sources is configured to excite fluorescence in a unique color channel of said unique color channels.

40. (Previously Presented) The system of claim 39, wherein each optically-encoded signal of said plurality of optically-encoded signals comprises at least one of said unique color channels.

41. (Previously Presented) The system of claim 9, wherein at least one of the plurality of optically-encoded signals comprises two or more colors detected in series.

42. (Previously Presented) The system of claim 39, wherein said two or more light sources comprise four light sources.

43. (Previously Presented) The system of claim 9, wherein said optical system is configured for wide-field microscopy.

44. (Previously Presented) The system of claim 9, wherein a numerical aperture of said objective is about 1.0.

45. (Previously Presented) A system comprising:

(a) a stage comprising:

(i) a sample holder configured to receive a substrate comprising a sample having a volume;

(ii) a temperature control apparatus operatively coupled to said sample holder; and

(iii) an XY motion control system;

(b) a fluidics system comprising:

(i) one or more reservoirs configured to contain one or more fluid reagents; and

(ii) a fluid dispenser operationally coupled to a pump, which pump is configured to provide said one or more fluid reagents from said one or more reservoirs to said sample holder; and

(c) an optical system comprising:

(i) an objective;

(ii) two or more light sources, wherein each light source of said two or more light sources is configured to deliver light to said volume, wherein said two or more light sources are configured to emit two or more colors of light; and

(iii) one or more detectors configured to receive light signals from said volume;

(d) a Z-axis motion control system configured for Z-motion of an objective focal plane through said volume; and

(e) a computing system comprising at least one computer readable storage medium having program instructions stored thereon, which program instructions are executable by at least one processor of said computing system to cause said at least one processor to perform a method comprising:

receiving a plurality of images of at least a portion of said volume from a plurality

of read-out cycles, wherein said plurality of images comprises a plurality of light signals received from said volume during said plurality of readout cycles, wherein said plurality of light signals is associated with one or more nucleic acid molecules within said volume, wherein each light signal of said plurality of light signals has a three-dimensional spatial position within said volume;

determining a plurality of optically-encoded signals from said plurality of light signals obtained from said plurality of read-out cycles, wherein each optically-encoded signal is associated with a nucleic acid molecule of said one or more nucleic acid molecules, wherein said plurality of optically-encoded signals comprises said two or more colors of light detected by said one or more detectors in series;

identifying said one or more nucleic acid molecules based on said plurality of optically-encoded signals; and

determining three-dimensional positional information of said one or more nucleic acid molecules.

46. (New) The system of claim 45, wherein said substrate comprises a flow cell.

47. (New) The system of claim 46, wherein said flow cell is a closed flow cell.

48. (New) The system of claim 47, wherein said closed flow cell comprises a fluid inlet and a fluid outlet, and wherein said fluid inlet is fluidically coupled to said fluid dispenser.

49. (New) The system of claim 46, wherein said flow cell is an open flow cell.

50. (New) The system of claim 45, wherein said objective is a water immersion objective.

51. (New) The system of claim 45, wherein said objective is a water dipping objective.

52. (New) The system of claim 45, wherein said temperature control apparatus comprises a heat sink or a Peltier element.

53. (New) The system of claim 45, further comprising a thermal control system.

54. (New) The system of claim 45, wherein said pump is a syringe pump.

55. (New) The system of claim 45, wherein said fluidics system further comprises one or more valves downstream of said one or more reservoirs and upstream of said fluid dispenser, wherein, upon actuation, said one or more valves facilitate dispensation of fluid to said sample holder.

56. (New) The system of claim 55, wherein said one or more valves comprise a rotary valve.

57. (New) The system of claim 55, wherein said one or more valves comprise a solenoid valve or a microfluidic valve.

58. (New) The system of claim 49, wherein said fluidics system further comprises an aspiration system comprising: (1) an aspiration tube configured to remove fluid from said open flow cell and a pressure source operatively coupled to said aspiration tube.

59. (New) The system of claim 58, wherein said aspiration system further comprises a motor operatively coupled to said aspiration tube, wherein said motor is configured to adjust a motion axis of said aspiration tube to control an interface between said aspiration tube and said open flow cell.

60. (New) The system of claim 59, wherein said motor is a stepper motor.

61. (New) The system of claim 45, wherein said Z-axis motion control system comprises a Z-axis linear motor configured for Z-motion of said objective focal plane through said volume.

62. (New) The system of claim 61, further comprising a linear encoder configured to provide Z-motion feedback to said Z-axis linear motor.

63. (New) The system of claim 45, further comprising an objective wetting apparatus.

64. (New) The system of claim 45, wherein said two or more light sources comprise one or more light emitting diodes (LEDs).

65. (New) The system of claim 45, wherein said two or more light sources comprise one or more lasers.

66. (New) The system of claim 45, wherein said optical system further comprises one or more excitation filters configured to selectively filter one or more wavelengths of light from said one or more light sources to said volume.

67. (New) The system of claim 66, wherein at least one excitation filter of said one or more excitation filters is configured to selectively filter light having a wavelength of about 480 nm, about 530 nm, about 590 nm, or about 640 nm from said one or more light sources to said volume.

68. (New) The system of claim 45, wherein said optical system further comprises one or more emission filters configured to selectively filter one or more wavelengths of light emitted from said volume to said one or more detectors.

69. (New) The system of claim 45, wherein said optical system is mechanically, electrically, or electromechanically coupled to a controller system configured to synchronize one or more imaging events within said system.

70. (New) The system of claim 45, further comprising a vibration control rigid foundation.

71. (New) The system of claim 45, further comprising a fluidics controller operatively coupled to said fluid dispenser and configured to automate or control dispensation of said one or more fluid reagents to said sample holder.

72. (New) The system of claim 53, wherein said temperature control apparatus further comprises a temperature sensor configured to provide a feedback signal to said thermal control system.

73. (New) The system of claim 68, wherein at least one of said one or more emission filters is configured to selectively filter light having a wavelength of about 510 nm, about 570 nm, about 620 nm, or about 680 nm to said one or more detectors from said volume to said one or more detectors.

74. (New) The system of claim 45, wherein said two or more light sources excite fluorescence in unique color channels such that each light source of said two or more light sources is configured to excite fluorescence in a unique color channel of said unique color channels.

75. (New) The system of claim 74, wherein each optically-encoded signal of said plurality of optically-encoded signals comprises at least one of said unique color channels.

76. (New) The system of claim 45, wherein at least one of the plurality of optically-encoded signals comprises two or more colors detected in series.

77. (New) The system of claim 45, wherein said two or more light sources comprise four light sources.

78. (New) The system of claim 45, wherein said optical system is configured for wide-field microscopy.

79. (New) The system of claim 45, wherein a numerical aperture of said objective is about 1.0.

REMARKS

Claims 9-17 and 19-45 were pending prior to entry of the abovementioned amendments. This amendment amends claims 17, 26-27, 29-30 and 39 for clarity and adds new claims 46-79. The amendments are fully supported by the application as originally filed. No new matter is added.

Respectfully submitted,

Dated: September 20, 2022

/Pei Wu/

Pei Wu, Reg. No. 60,023
BANNER & WITCOFF, LTD.
28 State Street, Suite 1800
Boston, MA 02109
(617) 720-9655 (direct dial)

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
17/395,534	08/06/2021	Evan R. Daugharthy	010498.01437\US	5147

22907 7590 09/23/2022

BANNER & WITCOFF, LTD.
1100 13th STREET, N.W.
SUITE 1200
WASHINGTON, DC 20005-4051

EXAMINER	
THOMAS, DAVID C	

ART UNIT	PAPER NUMBER
1637	

NOTIFICATION DATE	DELIVERY MODE
09/23/2022	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

GPD@bannerwitcoff.com
eofficeaction@bannerwitcoff.com

Applicant-Initiated Interview Summary	**Application No.** 17/395,534	**Applicant(s)** Daugharthy et al.		
	Examiner DAVID C THOMAS	**Art Unit** 1637	**AIA (First Inventor to File) Status** Yes	**Page** 1 of 1

All Participants (applicant, applicants representative, PTO personnel)	**Title**	**Type**
DAVID C THOMAS	Primary Examiner	Telephonic
Pei Wu	Attorney	

Date of Interview: 19 September 2022

Issues Discussed:

Proposed Amendment(s)

A Notice of Allowance was previously mailed for this application. Applicant's agent proposed several new claims that Applicant wishes to submit as a 312 Amendment. Several claims dependent from independent claim 9 would be corrected for clarity or grammar. In addition, new claims 46-79, dependent from independent claim 45 would be very similar to the dependent claims from claim 9. The examiner indicated that all the proposed claims appear to be proper, as they are further limiting to the subject matter of the base claims, and there are no issues with antecedent basis. The examiner looks forward to Applicant's reply.

/DAVID C THOMAS/ Primary Examiner, Art Unit 1637	

Applicant is reminded that a complete written statement as to the substance of the interview must be made of record in the application file. It is the applicants responsibility to provide the written statement, unless the interview was initiated by the Examiner and the Examiner has indicated that a written summary will be provided. See MPEP 713.04
Please further see:
MPEP 713.04
Title 37 Code of Federal Regulations (CFR) § 1.133 Interviews, paragraph (b)
37 CFR § 1.2 Business to be transacted in writing

Applicant recordation instructions: The formal written reply to the last Office action must include the substance of the interview. (See MPEP section 713.04). If a reply to the last Office action has already been filed, applicant is given a non-extendable period of the longer of one month or thirty days from this interview date, or the mailing date of this interview summary form, whichever is later, to file a statement of the substance of the interview.

Examiner recordation instructions: Examiners must summarize the substance of any interview of record. A complete and proper recordation of the substance of an interview should include the items listed in MPEP 713.04 for complete and proper recordation including the identification of the general thrust of each argument or issue discussed, a general indication of any other pertinent matters discussed regarding patentability and the general results or outcome of the interview, to include an indication as to whether or not agreement was reached on the issues raised.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
17/395,534	08/06/2021	Evan R. Daugharthy	010498.01437\US	5147

22907 7590 10/12/2022

BANNER & WITCOFF, LTD.
1100 13th STREET, N.W.
SUITE 1200
WASHINGTON, DC 20005-4051

EXAMINER	
THOMAS, DAVID C	

ART UNIT	PAPER NUMBER
1637	

NOTIFICATION DATE	DELIVERY MODE
10/12/2022	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

GPD@bannerwitcoff.com
eofficeaction@bannerwitcoff.com

Notice of Allowability	Application No. 17/395,534	Applicant(s) Daugharthy et al.	
	Examiner DAVID C THOMAS	Art Unit 1637	AIA (FITF) Status Yes

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. ☑ This communication is responsive to the submission of 20 September 2022.
   ☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2. ☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

3. ☑ The allowed claim(s) is/are 9-17 and 19-79 . As a result of the allowed claim(s), you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information , please see **http://www.uspto.gov/patents/init_events/pph/index.jsp** or send an inquiry to **PPHfeedback@uspto.gov.**

4. ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

   Certified copies:

   a) ☐ All b) ☐ Some* c) ☐ None of the:

   1. ☐ Certified copies of the priority documents have been received.
   2. ☐ Certified copies of the priority documents have been received in Application No. _____ .
   3. ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a)).

   * Certified copies not received: _____ .

Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file a reply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application.
THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.

5. ☐ CORRECTED DRAWINGS (as "replacement sheets") must be submitted.
   ☐ including changes required by the attached Examiner's Amendment / Comment or in the Office action of Paper No./Mail Date _____ .

   Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).

6. ☐ DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

Attachment(s)

1. ☐ Notice of References Cited (PTO-892)
2. ☐ Information Disclosure Statements (PTO/SB/08), Paper No./Mail Date _____.
3. ☐ Examiner's Comment Regarding Requirement for Deposit of Biological Material _____.
4. ☐ Interview Summary (PTO-413), Paper No./Mail Date. _____.
5. ☐ Examiner's Amendment/Comment
6. ☐ Examiner's Statement of Reasons for Allowance
7. ☐ Other _____.

/DAVID C THOMAS/
Primary Examiner, Art Unit 1637

Response to Rule 312 Communication	Application No.	Applicant(s)	
	17/395,534	Daugharthy et al.	
	Examiner	Art Unit	AIA (FITF) Status
	DAVID C THOMAS	1637	Yes

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

1. ☑ The amendment filed on 20 September 2022 under 37 CFR 1.312 has been considered, and has been:

a) ☐ entered.

b) ☑ entered as directed to matters of form not affecting the scope of the invention.

c) ☐ disapproved because the amendment was filed after the payment of the issue fee.
Any amendment filed after the date the issue fee is paid must be accompanied by a petition under 37 CFR 1.313(c)(1) and the required fee to withdraw the application from issue.

d) ☐ disapproved. See explanation below.

e) ☐ entered in part. See explanation below.

f) ☐ not entered because the supplemental or corrected Application Data sheet (ADS)

☐ was not accompanied by a petition to accept an unintentionally delayed claim under 37 CFR 1.55 or 27 CFR 1.78;

☐ did not identify the information being changed in accordance with 37 CFR 1.76(c)(2);

☐ was not properly signed in accordance with 37 CFR 1.76(e) (or 37 CFR 1.33(b) for applications filed prior to September 16, 2012).

/DAVID C THOMAS/
Primary Examiner, Art Unit 1637

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	17/395,534	Daugharthy et al.
	Examiner	**Art Unit**
	DAVID C THOMAS	1637

CPC						
Symbol					**Type**	**Version**
C12Q	/	1	/	6869	F	2013-01-01
C12Q	/	1	/	68	I	2013-01-01
C12Q	/	1	/	6841	I	2013-01-01

CPC Combination Sets								
Symbol					**Type**	**Set**	**Ranking**	**Version**
C12Q	/	1	/	6841	I	1	1	2013-01-01
C12Q	/	2543	/	101	I	1	2	2013-01-01
C12Q	/	2563	/	159	I	1	3	2013-01-01
C12Q	/	2565	/	601	I	1	4	2013-01-01
C12Q	/	1	/	6869	A	2	1	2013-01-01
C12Q	/	2543	/	101	A	2	2	2013-01-01
C12Q	/	2563	/	159	A	2	3	2013-01-01
C12Q	/	2565	/	601	A	2	4	2013-01-01

NONE (Assistant Examiner) (Date)	**Total Claims Allowed:** 70	
/DAVID C THOMAS/ Primary Examiner, Art Unit 1637 (Primary Examiner) (Date) 06 October 2022	O.G. Print Claim(s) 1	O.G. Print Figure None

U.S. Patent and Trademark Office Part of Paper No.: 20221006

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	17/395,534	Daugharthy et al.
	Examiner	**Art Unit**
	DAVID C THOMAS	1637

INTERNATIONAL CLASSIFICATION				
CLAIMED				
C12Q	/	1	/	68
C12Q	/	1	/	6869
C12Q	/	1	/	6841

NON-CLAIMED				
	/		/	

US ORIGINAL CLASSIFICATION	
CLASS	**SUBCLASS**

CROSS REFERENCES(S)						
CLASS	**SUBCLASS (ONE SUBCLASS PER BLOCK)**					

NONE (Assistant Examiner) (Date)	**Total Claims Allowed:** 70	
/DAVID C THOMAS/ Primary Examiner, Art Unit 1637 (Primary Examiner) (Date) 06 October 2022	O.G. Print Claim(s) 1	O.G. Print Figure None

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	17/395,534	Daugharthy et al.
	Examiner	**Art Unit**
	DAVID C THOMAS	1637

☑ Claims renumbered in the same order as presented by applicant ☐ CPA ☑ T.D. ☐ R.1.47

CLAIMS

Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original

NONE (Assistant Examiner)	(Date)	**Total Claims Allowed:** 70	
/DAVID C THOMAS/ Primary Examiner, Art Unit 1637 (Primary Examiner)	06 October 2022 (Date)	O.G. Print Claim(s) 1	O.G. Print Figure None

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	17/395,534	Daugharthy et al.
	Examiner	Art Unit
	DAVID C THOMAS	1637

✓	Rejected	-	Cancelled	N	Non-Elected	A	Appeal
=	Allowed	÷	Restricted	I	Interference	O	Objected

CLAIMS

☑ Claims renumbered in the same order as presented by applicant ☐ CPA ☑ T.D. ☐ R.1.47

CLAIM		DATE								
Final	Original	06/10/2022	10/06/2022							
	1	-	-							
	2	-	-							
	3	-	-							
	4	-	-							
	5	-	-							
	6	-	-							
	7	-	-							
	8	-	-							
1	9	=	=							
2	10	=	=							
3	11	=	=							
4	12	=	=							
5	13	=	=							
6	14	=	=							
7	15	=	=							
8	16	=	=							
9	17	=	=							
	18	-	-							
10	19	=	=							
11	20	=	=							
12	21	=	=							
13	22	=	=							
14	23	=	=							
15	24	=	=							
16	25	=	=							
17	26	=	=							
18	27	=	=							
19	28	=	=							
20	29	=	=							
21	30	=	=							
22	31	=	=							
23	32	=	=							
24	33	=	=							
25	34	=	=							
26	35	=	=							
27	36	=	=							
28	37	=	=							
29	38	=	=							
30	39	=	=							
31	40	=	=							
32	41	=	=							
33	42	=	=							

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	17/395,534	Daugharthy et al.
	Examiner	**Art Unit**
	DAVID C THOMAS	1637

CLAIM		DATE								
Final	Original	06/10/2022	10/06/2022							
34	43	=	=							
35	44	=	=							
36	45	=	=							
37	46		=							
38	47		=							
39	48		=							
40	49		=							
41	50		=							
42	51		=							
43	52		=							
44	53		=							
45	54		=							
46	55		=							
47	56		=							
48	57		=							
49	58		=							
50	59		=							
51	60		=							
52	61		=							
53	62		=							
54	63		=							
55	64		=							
56	65		=							
57	66		=							
58	67		=							
59	68		=							
60	69		=							
61	70		=							
62	71		=							
63	72		=							
64	73		=							
65	74		=							
66	75		=							
67	76		=							
68	77		=							
69	78		=							
70	79		=							

PATENT
ATTORNEY DOCKET NO. 010498.01437\US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:	)	
	)	
Evan R. Daugharthy	)	Examiner: David C. Thomas
	)	
Serial No.: 17/395,534	)	Art Unit: 1637
	)	
Filed: August 6, 2021	)	Conf. No.: 5147
	)	
Title: Method and Apparatus for Volumetric Imaging of a Three-Dimensional Nucleic Acid Containing Matrix	)	

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

AMENDMENT AFTER ALLOWANCE UNDER 37 C.F.R. § 1.312

Dear Commissioner:

In response to the Notice of Allowance mailed June 21, 2022, and prior to payment of the issue fee and issuance of the patent, Applicant respectfully requests entry of this amendment under 37 C.F.R. 1.312 for the above-captioned patent application. The Commissioner is hereby authorized to charge such extension fee or credit any overpayment of fees for this submission to Deposit Account No. 19-0733.

A Summary of the Interview begins at page 2 of this paper

Amendments to the Claims begin at page 3 of this paper.

Remarks begin at page 14 of this paper.

"OK to enter"

/D.C.T/

Doc code: IDS
Doc description: Information Disclosure Statement (IDS) Filed

PTO/SB/08a (02-18)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	
Filing Date	
First Named Inventor	Evan R. Daugharthy
Art Unit	N/A
Examiner Name	Not Yet Assigned
Attorney Docket Number	010498.01437\US

U.S.PATENTS Remove

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Patent citation information please click the Add button. Add

U.S.PATENT APPLICATION PUBLICATIONS Remove

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Published Application citation information please click the Add button. Add

FOREIGN PATENT DOCUMENTS Remove

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1							

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	
	First Named Inventor	Evan R. Daugharthy
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	010498.01437\US

	1	Terry R, Porreca G, McCarthy K, Church GM. 2008. Polonator Instrument https://web.archive.org/web/20080905133800/http://www.polonator.org; Wayback Machine (September 5, 2008) "Polony Sequence by Ligation Protocol"	
	2	Terry R, Porreca G, McCarthy K, Church GM. 2008. Polonator Instrument https://web.archive.org/web/20080705172604/http://www.polonator.org; Wayback Machine (July 5, 2008) "Polony Sequence Protocols"	
	3	Terry R, Porreca G, McCarthy K, Church GM. 2008. Polonator Instrument https://web.archive.org/web/20080905133818/http://www.polonator.org; Wayback Machine (September 5, 2008) "Help Wanted"	
	4	Terry R, Porreca G, McCarthy K, Church GM. 2008. Polonator Instrument https://web.archive.org/web/20080807155529/http://www.polonator.org/software.aspx; Wayback Machine (August 7, 2008) "Software"	
	5	Terry R, Porreca G, McCarthy K, Church GM. 2008. Polonator Instrument https://web.archive.org/web/20080807155404/http://www.polonator.org; Wayback Machine (August 7, 2008) "Reagent Kits"	
	6	Terry R, Porreca G, McCarthy K, Church GM. 2008. Polonator Instrument https://web.archive.org/web/20080905133541/http://www.polonator.org/reagentkits/run.aspx; Wayback Machine (September 5, 2008) "Run Kits"	
	7	Terry R, Porreca G, McCarthy K, Church GM. 2008. Polonator Instrument https://web.archive.org/web/20080905133601/http://www.polonator.org/reagentkits.pairedtag.aspx; Wayback Machine (September 5, 2008) "Paired-Leg Library Kits"	
	8	Terry R, Porreca G, McCarthy K, Church GM. 2008. Polonator Instrument https://web.archive.org/web/20080905133741/http://www.polonator.org/reagentkits.emulsion.aspx; Wayback Machine (September 5, 2008) "Emulsion PCR/Bead Capping Kits"	
	9	Terry R, Porreca G, McCarthy K, Church GM. 2008. Polonator Instrument https://web.archive.org/web/20080905133702/http://www.polonator.org/reagentkits/enrichment.aspx; Wayback Machine (September 5, 2008) "Enrichment Kits"	
	10	Terry R, Porreca G, McCarthy K, Church GM. 2008. Polonator Instrument https://web.archive.org/web/20080807155316/http://www.polonator.org/flowcells.aspx; Wayback Machine (August 7, 2008) "Flow Cells"	
	11	Church GM. 2006. "Genomes for all" Sci Am 294: 46-54	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	
	First Named Inventor	Evan R. Daugharthy
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	010498.01437\US

	12	de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. 2005. "Efficiency and power in genetic association studies" Nat Genet 37: 1217-23.	
	13	Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KC, Taylor J, Burnett E, Gut I, Farrall M, Lathrop GM, Abecasis GR, Cookson WO. 2007. "A genome-wide association study of global gene expression" Nat Genet 39: 1202-7	
	14	Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J, Carlson S, Helgason A, Walters GB, Gunnarsdottir S, Mouy M, Steinthorsdottir V, Eiriksdottir GH, Bjornsdottir G, Reynisdottir I, Gudbjartsson D, Helgadottir A, Jonasdottir A, Jonasdottir A, Styrkarsdottir U, Gretarsdottir S, Magnusson KP, Stefansson H, Fossdal R, Kristjansson K, Gislason HG, Stefansson T, Leifsson BG, Thorsteinsdottir U, Lamb JR, Gulcher JR, Reitman ML, Kong A, Schadt EE, Stefansson K. 2008; "Genetics of gene expression and its effect on disease" Nature 452: 423-8.	
	15	Risch N, Merikangas K. 1996. "The future of genetic studies of complex human diseases" Science 273: 1516-7	
	16	Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH. 2003. "Genetics of gene expression surveyed in maize, mouse and man" Nature 422: 297-302	
	17	Altshuler D, Daly MJ, Lander ES. 2008. "Genetic mapping in human disease" Science 322: 881-8	
	18	Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M. 2009. "Mapping complex disease traits with global gene expression" Nat Rev Genet 10: 184-94	
	19	International HapMap C. 2005. "A haplotype map of the human genome" Nature 437: 1299-320. PMC ID: PMC1880871	
	20	Klein RJ. 2007. "Power analysis for genome-wide association studies" BMC Genet 8: 58. PMC ID: PMC2042984	
	21	Kwan T, Benovoy D, Dias C, Gurd S, Provencher C, Beaulieu P, Hudson TJ, Sladek R, Majewski J. 2008. "Genome-wide analysis of transcript isoform variation in humans" Nat Genet 40: 225-31	
	22	McCarroll SA. 2008. "Extending genome-wide association studies to copy-number variation" Hum Mol Genet 17: R135-42	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	
	First Named Inventor	Evan R. Daugharthy
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	010498.01437\US

	23	Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, Cheung VG. 2004. "Genetic analysis of genome-wide variation in human gene expression" Nature 430: 743-7
single nucleotide polymorphisms. Nature 409:928-933	24	Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D, International SNPMWG. 2001. "A map of human genome sequence variation containing 1.42 million
	25	Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, Kasarskis A, Zhang B, Wang S, Suver C, Zhu J, Millstein J, Sieberts S, Lamb J, GuhaThakurta D, Derry J, Storey JD, Avila-Campillo I, Kruger MJ, Johnson JM, Rohl CA, van Nas A, Mehrabian M, Drake TA, Lusis AJ, Smith RC, Guengerich FP, Strom SC, Schuetz E, Rushmore TH, Ulrich R. 2008. "Mapping the genetic architecture of gene expression in human liver" PLoS Biol 6: e107. PMC ID: PMC2365981
	26	Serre D, Gurd S, Ge B, Sladek R, Sinnett D, Harmsen E, Bibikova M, Chudin E, Barker DL, Dickinson T, Fan JB, Hudson TJ. 2008. "Differential allelic expression in the human genome: a robust approach to identify genetic and epigenetic cis-acting mechanisms regulating gene expression" PLoS Genet 4: e1000006. PMC ID: PMC2265535
	27	Ball MP, Li JB, Gao Y, Lee J, LeProust E, Park I-H, Xie B, Daley GQ, Church GM. 2009. "Targeted and whole-genome methylomics reveals gene-body signatures in human cell lines" Nat Biotechnol 27: 361-8
	28	Brenner S, Williams SR, Vermaas EH, Storck T, Moon K, McCollum C, Mao JI, Luo S, Kirchner JJ, Eletr S, DuBridge RB, Burcham T, Albrecht G. 2000. "In vitro cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs" Proc Natl Acad Sci U S A 97: 1665-70. PMC ID: PMC26493
	29	Chiang DY, Getz G, Jaffe DB, O'Kelly MJ, Zhao X, Carter SL, Russ C, Nusbaum C, Meyerson M, Lander ES. 2009. "High-resolution mapping of copy-number alterations with massively parallel sequencing" Nat Methods 6: 99-103. PMC ID: PMC2630795
	30	Choy E, Yelensky R, Bonakdar S, Plenge RM, Saxena R, De Jager PL, Shaw SY, Wolfish CS, Slavik JM, Cotsapas C, Rivas M, Dermitzakis ET, Cahir-McFarland E, Kieff E, Hafler D, Daly MJ, Altshuler D. 2008. "Genetic analysis of human traits in vitro: drug response and gene expression in lymphoblastoid cell lines" PLoS Genet 4: e1000287. PMC ID: PMC2583954
	31	Christian AT, Pattee MS, Attix CM, Reed BE, Sorensen KJ, Tucker JD. 2001. "Detection of DNA point mutations and mRNA expression levels by rolling circle amplification in individual cells" Proc Natl Acad Sci U S A 98: 14238-43. PMC ID: PMC64666
	32	Church GM, Porreca GJ, Terry RC, Lares M. 2008. "High-Speed Imaging for DNA Sequencing" Biophotonics (<http://www.photonics.com/Content/ReadArticle.aspx?ArticleID=33989>)
	33	Deng J, Shoemaker R, Xie B, Gore A, LeProust EM, Antosiewicz-Bourget J, Egli D, Maherali N, Park IH, Yu J, Daley GQ, Eggan K, Hochedlinger K, Thomson J, Wang W, Gao Y, Zhang K. 2009. "Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming" Nat Biotechnol 27: 353-60

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	
	First Named Inventor	Evan R. Daugharthy
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	010498.01437\US

	34	Eberwine J, Kacharmina JE, Andrews C, Miyashiro K, McIntosh T, Becker K, Barrett T, Hinkle D, Dent G, Marciano P. 2001. "mRna expression analysis of tissue sections and single cells" J Neurosci 21: 8310-4	
	35	Kolb HC, Finn MG, B. SK. 2001. "Click Chemistry: Diverse Chemical Function from a Few Good Reactions" Angew. Chem. Int. 40: 2004-21	
	36	Kwiatkowski M, Fredriksson S, Isaksson A, Nilsson M, Landegren U. 1999. "Inversion of in situ synthesized oligonucleotides: improved reagents for hybridization and primer extension in DNA microarrays" Nucleic Acids Res 27: 4710-4. PMC ID: PMC148770	
	37	Li JB, Gao Y, Aach J, Zhang K, Kryukov GV, Xie B, Ahlford A, Yoon J-K, Rosenbaum AM, Wait-Zaranek A, LeProust E, Sunyaev S, Church GM. 2009. "Multiplex padlock capture and sequencing reveal human hypermutable CpG variations" Genome Res in press	
	38	Mitra RD, Butty VL, Shendure J, Williams BR, Housman DE, Church GM. 2003. "Digital genotyping and haplotyping with polymerase colonies" Proc Natl Acad Sci U S A 100: 5926-31. PMC ID: PMC156303	
	39	Pan X, Urban AE, Palejev D, Schulz V, Grubert F, Hu Y, Snyder M, Weissman SM. 2008. "A procedure for highly specific, sensitive, and unbiased whole-genome amplification" Proc Natl Acad Sci U S A 105: 15499-504. PMC ID: PMC2563063	
	40	Stougaard M, Lohmann JS, Zajac M, Hamilton-Dutoit S, Koch J. 2007. "In situ detection of non-polyadenylated RNA molecules using Turtle Probes and target primed rolling circle PRINS" BMC Biotechnol 7: 69. PMC ID: PMC2203993	
	41	Wang Z, Gerstein M, Snyder M. 2009. "RNA-Seq: a revolutionary tool for transcriptomics" Nat Rev Genet 10: 57-63	
	42	Wu J, Zhang S, Meng Q, Cao H, Li Z, Li X, Shi S, Kim DH, Bi L, Turro NJ, Ju J. 2007. "3'-O-modified nucleotides as reversible terminators for pyrosequencing" Proc Natl Acad Sci U S A 104: 16462-7. PMC ID: PMC2034218	
	43	Zhang K, Li JB, Gao Y, Egli D, Xie B, Deng J, Li Z, Lee J, Aach J, Leproust E, Eggan K, Church GM. 2009. "Digital RNA Allelotyping Reveals Tissue-specific and Allele-specific Gene Expression in Human" (submitted to Nature Methods)	
	44	Bakal C, Aach J, Church G, Perrimon N. 2007. "Quantitative morphological signatures define local signaling networks regulating cell morphology" Science 316: 1753-6	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	
	First Named Inventor	Evan R. Daugharthy
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	010498.01437\US

	45	Bang D, Church GM. 2008. "Gene synthesis by circular assembly amplification" Nat Methods 5: 37-9
	46	Bell J. 2004. "Predicting disease using genomics" Nature 429: 453-6
from single polymerase molecules. Science 323: 133-138	47	Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S. 2009. "Real-time DNA sequencing
	48	Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, Braslavsky I, Causey M, Colonell J, Dimeo J, Efcavitch JW, Giladi E, Gill J, Healy J, Jarosz M, Lapen D, Moulton K, Quake SR, Steinmann K, Thayer E, Tyurina A, Ward R, Weiss H, Xie Z. 2008. "Single-molecule DNA sequencing of a viral genome" Science 320: 106-9
	49	Kim JB, Porreca GJ, Song L, Greenway SC, Gorham JM, Church GM, Seidman CE, Seidman JG. 2007. "Polony multiplex analysis of gene expression (PMAGE) in mouse hypertrophic cardiomyopathy" Science 316: 1481-4
	50	Kurimoto K, Yabuta Y, Ohinata Y, Saitou M. 2007. "Global single-cell cDNA amplification to provide a template for representative high-density oligonucleotide microarray analysis" Nat Protoc 2: 739-52

If you wish to add additional non-patent literature document citation information please click the Add button Add

EXAMINER SIGNATURE

Examiner Signature	/DAVID C THOMAS/	Date Considered	02/25/2022

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	
	First Named Inventor	Evan R. Daugharthy
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	010498.01437\US

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

See attached certification statement.

The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/John P. Iwanicki/	Date (YYYY-MM-DD)	2021-08-05
Name/Print	John P. Iwanicki	Registration Number	34,628

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	ISSUE DATE	PATENT NO.	ATTORNEY DOCKET NO.	CONFIRMATION NO.
17/395,534	01/03/2023	11542554	010498.01437\US	5147

22907 7590 12/14/2022

BANNER & WITCOFF, LTD.
1100 13th STREET, N.W.
SUITE 1200
WASHINGTON, DC 20005-4051

ISSUE NOTIFICATION

The projected patent number and issue date are specified above.

Determination of Patent Term Adjustment under 35 U.S.C. 154 (b)
(application filed on or after May 29, 2000)

The Patent Term Adjustment is 0 day(s). Any patent to issue from the above-identified application will include an indication of the adjustment on the front page.

If a Continued Prosecution Application (CPA) was filed in the above-identified application, the filing date that determines Patent Term Adjustment is the filing date of the most recent CPA.

Applicant will be able to obtain more detailed information by accessing the Patent Application Information Retrieval (PAIR) WEB site (http://pair.uspto.gov).

Any questions regarding the Patent Term Extension or Adjustment determination should be directed to the Office of Patent Legal Administration at (571)-272-7702. Questions relating to issue and publication fee payments should be directed to the Application Assistance Unit (AAU) of the Office of Patents Stakeholder Experience (OPSE), Stakeholder Support Division (SSD) at (571)-272-4200.

INVENTOR(s) (Please see PAIR WEB site http://pair.uspto.gov for additional inventors):

Evan R. Daugharthy, Cambridge, MA;
Richard C. Terry, Carlisle, MA;
Je-Hyuk Lee, Allston, MA;
George M. Church, Brookline, MA;
Benjamin W. Pruitt, Cambridge, MA;

APPLICANT(s) (Please see PAIR WEB site http://pair.uspto.gov for additional applicants):

President and Fellows of Harvard College, Cambridge, MA;

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The USA offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to encourage and facilitate business investment. To learn more about why the USA is the best country in the world to develop technology, manufacture products, and grow your business, visit SelectUSA.gov.