



## (12) 发明专利申请

(10) 申请公布号 CN 105385756 A

(43) 申请公布日 2016. 03. 09

(21) 申请号 201510755990. 4

(22) 申请日 2011. 04. 05

(30) 优先权数据

61/321, 124 2010. 04. 05 US

13/080, 616 2011. 04. 05 US

(62) 分案原申请数据

201180017696. 3 2011. 04. 05

(71) 申请人 普罗格诺西斯生物科学公司

地址 美国加利福尼亚州

(72) 发明人 马克·S·奇

(74) 专利代理机构 北京安信方达知识产权代理

有限公司 11262

代理人 刘小立 郑霞

(51) Int. Cl.

C12Q 1/68(2006. 01)

G01N 33/68(2006. 01)

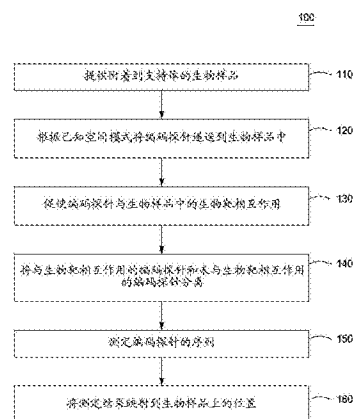
权利要求书2页 说明书18页 附图5页

(54) 发明名称

空间编码的生物学测定

(57) 摘要

本发明提供了用于空间编码的生物学测定的测定及测定系统。本发明提供了包含如下的测定系统：能够高水平地多重检测的测定，其中以特定的空间模式给生物样品提供试剂；能够根据该空间模式控制递送试剂的设备；和提供本质上是数字的读出的解码方案。



1. 一种测定样品中生物靶的丰度或活性或两者的方法,所述方法包括:  
将用于所述生物靶的探针递送至所述样品中的多个位点;  
促使所述探针与所述生物靶在所述多个位点的每个位点相互作用;  
将编码剂递送至所述样品中的所述多个位点的每个位点,其中递送至每个位点的所述编码剂确定所述样品中的所述位点的位置;  
将所述编码剂与递送至所述多个位点的每个位点的所述探针偶联以在所述多个位点的每个位点形成编码探针;  
测定所述编码探针的序列的全部或一部分,从而确定所述探针被递送的位点的位置;  
以及  
将所述多个位点的每个位点的所述生物靶的丰度或活性或两者与所述样品中的所述多个位点的位置相关联,据此测定所述样品中所述生物靶的丰度或活性或两者。
2. 根据权利要求 1 所述的方法,其中所述生物靶是核酸且所述编码剂是寡核苷酸。
3. 根据权利要求 1 或 2 所述的方法,其中至少两种编码剂被递送至所述样品中的所述多个位点的每个位点。
4. 根据权利要求 1 所述的方法,其中所述生物靶是蛋白质,用于所述生物靶的探针包含蛋白质和 / 或所述编码剂包含寡核苷酸。
5. 根据权利要求 4 所述的方法,其中所述生物靶包含酶,和 / 或用于所述生物靶的探针包含酶的底物或推定底物或两者。
6. 根据权利要求 1-5 中任一项所述的方法,其中用于所述生物靶的所述探针包含亲和捕获剂、抗体、适体或小分子。
7. 根据权利要求 1-6 中任一项所述的方法,还包括扩增所述编码探针的序列的全部或一部分。
8. 根据权利要求 1-7 中任一项所述的方法,其中所述测定步骤通过核酸测序或高通量测序进行。
9. 根据权利要求 1-8 中任一项所述的方法,其中所述样品中多种生物靶的丰度或活性或两者的空间模式被确定。
10. 根据权利要求 9 所述的方法,其中被测定的所述多种生物靶与所述样品中的所述多个位点的乘积大于 20、50、75、100、1,000、10,000、100,000 或 1,000,000。
11. 根据权利要求 1-10 中任一项所述的方法,其中至少十万个、至少五十万个或至少一百万个编码探针的序列被并行测定。
12. 根据权利要求 1-11 中任一项所述的方法,其中根据已知的空间模式将用于所述生物靶的所述探针和 / 或所述编码剂递送至所述样品中的所述多个位点。
13. 根据权利要求 12 所述的方法,其中所述已知的空间模式通过所述样品的组织学特征确定。
14. 根据权利要求 1-13 中任一项所述的方法,其中软件程控的硬件进行所述递送步骤、所述偶联步骤、所述测定步骤和所述关联步骤中的至少两个步骤。
15. 根据权利要求 1-14 中任一项所述的方法,还包括将与所述生物靶相互作用的探针和未与所述生物靶相互作用的探针分离。
16. 根据权利要求 15 所述方法,其中所述分离步骤通过用亲和捕获剂捕获与所述生物

靶相互作用的探针实现。

17. 根据权利要求 15 所述的方法,其中所述分离步骤通过洗涤所述样品实现。

18. 根据权利要求 1-17 中任一项所述的方法,其中所述偶联步骤通过连接进行或通过延伸然后连接进行。

19. 根据权利要求 1-18 中任一项所述的方法,其中所述样品附着到支持体。

20. 一种测定样品中核酸靶的丰度或活性或两者的方法,所述方法包括:

将用于所述核酸靶的寡核苷酸探针递送至所述样品中的多个位点;

促使所述寡核苷酸探针与所述核酸靶在所述多个位点的每个位点杂交;

从所述样品中洗掉未杂交的寡核苷酸探针;

将至少两种编码剂递送至所述样品中的所述多个位点的每个位点,其中递送至每个位点的所述至少两种编码剂的组合确定所述样品中的所述位点的位置;

将所述至少两种编码剂与杂交于所述核酸靶的所述寡核苷酸探针偶联以在所述多个位点的每个位点形成编码探针;

测定所述编码探针的序列的全部或一部分,从而确定所述寡核苷酸探针被递送的位点的位置;以及

将所述多个位点的每个位点的所述核酸靶的丰度或活性或两者与所述样品中的所述多个位点的位置相关联,据此测定所述样品中所述核酸靶的丰度或活性或两者。

21. 根据权利要求 20 所述的方法,其中所述样品中多种核酸靶的丰度或活性或两者的空间模式被确定。

22. 根据权利要求 20 或 21 所述的方法,其中所述测定步骤中测定的所述编码探针的序列包含所述寡核苷酸探针和所述编码剂的序列。

23. 根据权利要求 20-22 中任一项所述的方法,还包括确定杂交于所述样品中的所述多个位点的每个位点的所述核酸靶的所述编码探针的量。

24. 根据权利要求 23 所述的方法,其中杂交于所述核酸靶的所述编码探针的量指示所述样品中的所述多个位点的每个位点的所述核酸靶的丰度或活性或两者。

25. 根据权利要求 1-24 中任一项所述的方法,其中所述编码探针还包含用于扩增的一个或多个引发位点和 / 或一个或多个测序接头。

## 空间编码的生物学测定

[0001] 本申请是申请日为 2011 年 4 月 5 日,申请号为 201180017696.3,发明名称为“空间编码的生物学测定”的申请的分案申请。

[0002] 相关申请的交叉引用

[0003] 本申请要求 2011 年 4 月 5 日递交的美国专利申请第 13/080,616 号和 2010 年 4 月 5 日递交的美国临时专利申请第 61/321,124 号的权益并被本文通过引用并入。

### 技术领域

[0004] 本发明涉及生物分子的测定,且更具体地涉及用于同时确定大量生物分子在固体样品中的空间分布的测定。

### 背景技术

[0005] 在以下讨论中,将为了背景和介绍的目的描述某些制品和方法。本部分包含的任何内容都不应被解释为“承认”是现有技术。申请人明确地保留在适当的时候,根据适用的法律条文证实本文述及的制品和方法不构成现有技术的权利。

[0006] 全面的基因表达分析和蛋白质分析已成为理解生物学机制的有用工具。这些工具的使用已促进参与发育和各种疾病例如癌症和自身免疫病的基因和蛋白质的鉴定。传统方法例如不同转录物的原位杂交和其他多重检测已揭示出基因表达的空间模式并已帮助阐明发育和疾病的分子基础。使得能够定量分析每个样品的许多 RNA 序列的其他技术包括微阵列(参见,Shi,等人,Nature Biotechnology,24(9):1151-61(2006);及 Slonim 和 Yanai,Plos Computational Biology,5(10):e1000543(2009));基因表达的系列分析(SAGE)(参见 Velculescu,等人,Science,270(5235):484-87(1995))、qPCR 的高通量实现(参见 Spurgeon,等人,Plos ONE,3(2):e1662(2008))和原位 PCR(参见 Nuovo,Genome Res.,4:151-67(1995))。尽管这些方法是有用的,但是它们不能在样品中的许多空间位置同时测量许多基因的表达或多种蛋白质的存在和/或活性。激光捕获显微切割已允许在少数位置分析许多基因,但是其非常昂贵、费力且不能很好地调整。虽然某些 2D 格式的 PCR 测定保留了空间信息(参见 Armani,等人,Lab on a Chip,9(24):3526-34(2009)),但是这些方法具有低的空间分辨率,因为它们依赖于向孔中物理转移组织,这还阻止了随机接近组织样品和高水平的多重检测(multiplexing)。

[0007] 目前,不存在以高分辨率同时分析大量的基因、蛋白质或其他生物活性分子的空间表达模式的运行方法。因此存在对组织中的生物分子的可重现的、高分辨率的空间图谱的需要。本发明解决了该需要。

### 发明内容

[0008] 提供该概述以便以简化的方式介绍以下在详述中进一步描述的概念的节选。该概述不旨在确定要求保护的主题的关键或必不可少的特征,也不旨在用来限制所要求保护的主题的范围。从随后的包括在附图中示出并在所附权利要求中定义的那些方面的书面详述

中,所要求保护的的主题的其他特征、细节、实用性和优点将是明显的。

[0009] 本发明包括提供组织中生物活性的高分辨率空间图谱的测定系统。该测定系统包含能够高水平多重检测的测定,其中以指定的空间模式为生物样品提供编码探针;能够根据该空间模式控制递送试剂的设备;和提供本质上为数字的读出的解码方案。简言之,本发明提供了观看许多位置中的许多生物靶的能力,提供了以高度并行的测序数据分析对原位杂交的分辨。

[0010] 因此,在某些实施方式中,本发明提供了确定多种生物靶在样品中的多个位点的丰度或活性或两者的空间模式的测定系统,其中该测定系统进行下列步骤:提供附着到支持体的样品;以已知的空间模式将用于多种生物靶的编码探针递送到样品中的多个位点,其中每个编码探针包含可与生物靶相互作用的探针区域和确定编码探针被递送的位点的位置的代码标签(coding tag);促使编码探针与生物靶相互作用;将与生物靶相互作用的编码探针和未与生物靶相互作用的编码探针分离;测定编码探针的序列的全部或一部分,以及将多种生物靶的丰度或活性或两者与样品中的位点的位置相关联。

[0011] 在本发明特定的方面,生物靶包含核酸且编码探针是寡核苷酸,且在某些方面,对于多种核酸靶中的每一种存在两种编码的探针。在某些方面,多种生物靶包含蛋白质,编码探针的探针区域是蛋白质且代码标签包含寡核苷酸。在某些方面,多种生物靶包含酶。在某些方面,编码探针的探针区域包含抗体、适体或小分子。

[0012] 在本发明特定的方面,所述生物靶可以是核酸且所述编码探针可以是寡核苷酸。优选地,其中对于多种核酸靶中的每一种可存在两种编码探针。

[0013] 在本发明特定的方面,所述多种生物靶可以是蛋白质,所述编码探针的所述探针区域可以是蛋白质且所述代码标签可包含寡核苷酸。优选地,其中所述多种生物靶可包含酶。

[0014] 测定系统的某些方面还在分离步骤和测定步骤之间包含扩增步骤。在某些方面,测定步骤通过核酸测序进行,且在优选的方面,测序是高通量数字核酸测序(high throughput digital sequencing)。

[0015] 在本发明的某些方面,被测定的多种生物靶和样品中的多个位点的乘积大于 20,在某些方面,被测定的多种生物靶和样品中的多个位点的乘积大于 50,在某些方面,被测定的多种生物靶和样品中的多个位点的乘积大于 75、100、150、500、750、1,000、5,000、10,000、25,000、50,000、100,000、500,000 或 1,000,000 或更大。在其他方面,并行测定至少五万个编码探针的序列,在其他方面,并行测定至少十万个编码探针的序列,在某些方面,并行测定至少五十万个编码探针的序列,且在某些方面,并行测定至少一百万、一千万、一亿、十亿、一百亿、一千亿或更多个编码探针的序列。

[0016] 在某些方面,已知的空间模式通过样品的组织学特征确定。还在某些方面,软件程控的硬件进行递送步骤、分离步骤、测定步骤和关联步骤中的至少两个步骤。

[0017] 在某些方面,编码探针的探针区域是蛋白质且分离步骤通过用亲和捕获剂捕获与生物靶相互作用的编码探针实现。在某些方面,编码探针的探针区域是核酸且分离步骤通过洗涤样品实现。

[0018] 在其他实施方式中,提供了测定多种核酸靶在样品中的多个位点的丰度或活性或两者的空间模式的测定系统,其中该测定系统进行以下步骤:提供附着到支持体的样品;

以已知空间模式将用于多种核酸靶的寡核苷酸探针递送到样品中的多个位点；促使寡核苷酸探针与核酸靶杂交；从样品中洗掉未杂交的编码寡核苷酸探针；根据已知空间模式将一种或多种编码剂 (encoding agent) 递送到样品中的多个位点的位置，其中递送到每个位点的编码剂的组合是不同的；将编码剂和寡核苷酸探针偶联以形成编码探针；使用高通量测序测定编码探针的序列的全部或一部分，以及将多种生物靶的丰度或活性或两者与样品中多个位点的位置相关联。

[0019] 在某些方面，被测定的所述多种生物靶与所述样品中的所述多个位点的乘积可大于 20。

[0020] 在某些方面，被测定的所述多种生物靶与所述样品中的所述多个位点的乘积可大于 50。

[0021] 在某些方面，被测定的所述多种生物靶与所述样品中的所述多个位点的乘积可大于 75。

[0022] 在某些方面，被测定的所述多种生物靶与所述样品中的所述多个位点的乘积可大于 10,000。

[0023] 在某些方面，被测定的所述多种生物靶与所述样品中的所述多个位点的乘积可大于 100,000。

[0024] 在某些方面，被测定的所述多种生物靶与所述样品中的所述多个位点的乘积可大于 1,000,000。

[0025] 在某些方面，至少十万个编码探针的序列可被并行测定。

[0026] 在某些方面，至少一百万个编码探针的序列可被并行测定。

[0027] 在某些方面，对于所述多种核酸靶中的每一种可递送两种寡核苷酸探针。

[0028] 在某些方面，所述偶联步骤可通过连接进行。

[0029] 在某些方面，所述偶联步骤可通过延伸然后连接进行。

[0030] 在某些方面，还可在所述偶联步骤和所述测定步骤之间包含扩增步骤。

[0031] 本发明的其他实施方式提供了测定多种蛋白质靶在样品中的多个位点的丰度或活性或两者的空间模式的测定系统，其中该测定系统进行以下步骤：提供附着到支持体的样品；以已知空间模式将用于多种蛋白质靶的编码探针递送到样品中的多个位点，其中每个编码探针包含可与蛋白质靶相互作用的蛋白质探针区域和确定编码探针被递送的位点的位置的代码标签，且代码标签是编码探针的蛋白质探针区域的一部分；促使编码探针与蛋白质靶相互作用；将与蛋白质靶相互作用的编码探针和未与蛋白质靶相互作用的编码探针分离；通过高通量测序测定编码探针的序列的全部或一部分，以及将多种蛋白质靶的丰度或活性或两者与样品中的多个位点的位置相关联。

[0032] 在某些方面，被测定的所述多种生物靶与所述样品中的所述多个位点的乘积可大于 20。

[0033] 在某些方面，被测定的所述多种生物靶与所述样品中的所述多个位点的乘积可大于 50。

[0034] 在某些方面，被测定的所述多种生物靶与所述样品中的所述多个位点的乘积可大于 75。

[0035] 在某些方面，被测定的所述多种生物靶与所述样品中的所述多个位点的乘积可大

于 100。

[0036] 在某些方面,被测定的所述多种生物靶与所述样品中的所述多个位点的乘积可大于 500。

[0037] 在某些方面,被测定的所述多种生物靶与所述样品中的所述多个位点的乘积可大于 1000。

[0038] 在某些方面,至少一万个编码探针的序列可被并行测定。

[0039] 在某些方面,至少十万个编码探针的序列可被并行测定。

[0040] 在某些方面,至少一百万个编码探针的序列可被并行测定。

[0041] 在某些方面,还可在所述偶联步骤和所述测定步骤之间包含扩增步骤。

[0042] 在某些方面,所述蛋白质靶可以是酶,且所述编码探针的所述探针区域可以是酶的底物、推定底物或两者。

[0043] 在某些方面,所述编码探针的所述探针区域可以是亲和捕获剂。优选地,所述编码探针的所述探针区域可以是抗体。优选地,所述编码探针的所述探针区域可以是适体。

[0044] 在某些方面,将与所述蛋白质靶相互作用的编码探针和未与所述蛋白质靶相互作用的编码探针分离可通过亲和捕获剂实现,所述亲和捕获剂区分与所述蛋白质靶相互作用的编码探针和未与所述蛋白质靶相互作用的编码探针。

[0045] 在某些方面,所述代码标签可以是寡核苷酸。

[0046] 其他实施方式提供了测定多种生物靶在样品中的多个位点的丰度或活性或两者的空间模式的测定系统,其中该测定系统进行以下步骤:提供附着到支持体的样品;以已知空间模式将用于多种生物靶的编码探针递送到样品中的多个位点,其中每个编码探针包含可与生物靶相互作用的探针区域和确定编码探针被递送的位点的位置并鉴定生物靶的代码标签;促使编码探针与生物靶相互作用;测定编码探针的序列的全部或一部分,以及将多种生物靶的丰度或活性或两者与样品中位点的位置相关联。

[0047] 基于待检测的分子和该检测系统所需要的试剂,本发明的测定系统可采用多种检测机制。以下更详细地描述了可用于本发明的测定系统的示例性的方法。

## 附图说明

[0048] 图 1 提供了本发明的测定系统的简化概述。

[0049] 图 2 提供了用于检测核酸的本发明的测定系统的一个实施方式的简化概述。

[0050] 图 3 是图 2 中概述的测定的一个实施方式的代表性描绘。

[0051] 图 4 示出了用于本发明的测定系统的组合编码方案的一个实施方式的一般机制。

[0052] 图 5 提供了图 4 中所示的组合编码方案的实施方式的简化的、具体的实例。

[0053] 定义

[0054] 本文使用的术语旨在具有本领域普通技术人员所理解的普通且常用的含义。除非特别指明,以下定义旨在帮助读者理解本发明,而不旨在改变或以其他方式限制这些术语的含义。

[0055] 如本文所用的术语“抗体”旨在表示完整的免疫球蛋白或抗体或能够特异性结合抗原的免疫球蛋白分子的任何功能片段(抗体和抗原是本文所定义的“结合配偶体(binding partner)”)。如本文所用的“抗体”意在包括完整的抗体以及能够结合目的抗

原或抗原片段的任何抗体片段。这些肽的实例包括完整的抗体分子、抗体片段例如 Fab、F(ab')<sub>2</sub>、CDR、VL、VH 和抗体的能够特异性结合抗原的任何其他部分。用于本发明的测定的抗体对在本发明的测定中所检测的蛋白质（即，生物靶）或用于检测的蛋白质（即，探针）具有免疫反应性或免疫特异性，且因此与它们特异性且选择性地结合。

[0056] 如本文所用的术语“结合剂”是指特异性结合目的生物分子的任何剂。

[0057] “互补的”或“基本上互补的 (substantially complementary)”是指核苷酸或核酸之间，诸如例如双链 DNA 分子的两条链之间或寡核苷酸引物和单链核酸上的引物结合位点之间的杂交、或碱基配对或双链体的形成。通常，互补的核苷酸是 A 和 T（或 A 和 U），或 C 和 G。当核苷酸的一条链（经最佳比对和比较且具有适当的核苷酸插入或缺失）与另一条链的至少约 80%，通常至少约 90% 至约 95%，和甚至约 98% 至约 100% 配对时，两条单链 RNA 或 DNA 分子被表述为是基本上互补的。

[0058] “杂交”是指其中两条单链多核苷酸非共价地结合形成稳定的双链多核苷酸的过程。所得的（通常）双链的多核苷酸是“杂交体 (hybrid)”或“双链体”。“杂交条件”将通常包括约小于 1M，通常小于约 500mM 且可以小于约 200mM 的盐浓度。“杂交缓冲液”是缓冲盐溶液，例如 5% SSPE 或其他的本领域已知的这类缓冲液。杂交温度可以低至 5℃，但是通常大于 22℃，且更通常地大于约 30℃，且通常超过 37℃。杂交通常在严格条件，即在这种条件下引物将与其靶子序列 (subsequence) 杂交而不会与其他的非互补序列杂交的条件下进行。严格条件是序列依赖性的且在不同的环境中是不同的。例如，对于特异性杂交，较长的片段可能需要比短片段高的杂交温度。因为其他因素可能影响杂交的严格性，包括互补链的碱基组成和长度、有机溶剂的存在和碱基错配的程度，所以参数的组合比单独的任何一个参数的绝对测量更重要。通常严格条件被选择为比在特定的离子强度和 pH 下特定序列的  $T_m$  低约 5℃。在约 7.0 至约 8.3 的 pH 下和至少 25℃ 的温度下，示例性的严格条件包括至少 0.01M 至不大于 1M 的钠离子浓度（或其他盐）的盐浓度。例如，5xSSPE (pH 7.4 下的 750mM NaCl、50mM 磷酸钠、5mM EDTA) 和约 30℃ 的温度的条件适合等位基因特异性的杂交，但是合适的温度取决于杂交区域的长度和 / 或 GC 含量。

[0059] “连接”是指在温度驱动的反应中在两种或更多种核酸，例如寡核苷酸和 / 或多核苷酸的末端之间形成共价键或连接 (linkage)。键或连接的性质可极大地不同且连接可酶促地或化学地进行。如本文所用的，连接通常酶促地进行以在一个寡核苷酸的 5' 碳末端核苷酸与另一个核苷酸的 3' 碳之间形成磷酸二酯连接。

[0060] 本文所用的“核酸”、“寡核苷酸”、“oligo”或语法等价形式通常是指共价连接在一起的至少两个核苷酸。核酸通常会包含磷酸二酯键，但是在某些情况下可包括具有可选骨架，例如亚磷酰胺、二硫代磷酸酯或甲基亚磷酰胺 (methylphosphoroamidite) 连接；或肽核酸骨架和连接的核酸类似物。其他类似物核酸包括具有双环结构（包括锁核酸）、刚性骨架 (positive backbone)、非离子骨架和非核糖骨架的类似物核酸。可对核糖 - 磷酸骨架进行修饰以增加分子的稳定性；例如，在某些环境中 PNA:DNA 杂交体可显示出更高的稳定性。

[0061] “引物”是指天然或合成的寡核苷酸，其与多核苷酸模板形成双链体后能够充当核酸合成的起始点并从其 3' 端沿模板延伸致使形成延伸的双链体。延伸过程期间添加的核苷酸的序列由模板多核苷酸的序列决定。通常通过 DNA 聚合酶延伸引物。

[0062] 术语“SNP”或“单核苷酸多态性”指个体之间的遗传变异；例如，生物体的 DNA 中可



变的单一含氮碱基的位置。发现 SNP 遍布基因组；个体之间的许多遗传变异归因于 SNP 位点的变异，且通常该遗传变异引起个体之间的表型变异。用于本发明的 SNP 及其相应的等位基因可衍生自任意几种来源，例如公共数据库 (U. C. Santa Cruz Human Genome Browser Gateway (<http://genome.ucsc.edu/cgi-bin/hgGateway>) 或 NCBI dbSNP 网址 (<http://www.ncbi.nlm.nih.gov/SNP/>) 或可如美国专利第 6, 969, 589 号；和名称为“Human Genomic Polymorphisms”的美国公布第 2006/0188875 号中所述经实验确定。虽然在本文提供的某些实施方式中描述了 SNP 的使用，但是将应理解其他双等位基因或多等位基因遗传标志也可使用。双等位基因遗传标志是具有两种多态形式或等位基因的遗传标志。如以上提及的，对于与性状相关联的双等位基因遗传标志，与对照组相比在实例组的遗传组成中更丰富的等位基因被称作“关联等位基因 (associated allele)”，而其他等位基因可被称为“非关联等位基因 (unassociated allele)”。因此，对于与特定性状（例如，疾病或药物反应）相关联的每种双等位基因多态性，存在相应的关联等位基因。可被用于本文提供的方法的其他双等位基因多态性包括但不限于多种核苷酸改变、插入、缺失和转位。还将应理解，本文述及的 DNA 可包括基因组 DNA、线粒体 DNA、游离体 DNA 和 / 或 DNA 衍生物例如扩增子、RNA 转录物、cDNA、DNA 类似物等。在关联研究中筛选到的多态位点可处于二倍体或单倍体状态且理想地，将来自遍布基因组的位点。

[0063] 当提及结合配偶体（例如，蛋白质、核酸、抗体或其他亲和捕获剂等等）时，如本文所用的术语“选择性地结合”、“选择性的结合”及类似术语是指两个或更多个结合配偶体的具有高亲和力 / 或互补性的结合反应以确保在指定的测定条件下的选择性杂交。通常，特异性结合将是背景信号的标准差的至少三倍。因此，在指定条件下，结合配偶体结合其特有的“靶”分子而不大量地结合样品中存在的其他分子。

[0064] “测序”、“序列测定”及类似术语是指与核酸的核苷酸碱基序列相关的信息的测定。该信息可包括核酸的部分以及全部序列信息的鉴定或测定。序列信息可根据不同程度的统计学可信度或置信度确定。在一个方面，该术语包括对核酸中多个连续核苷酸的同源性或顺序的测定。“高通量数字测序”或“下一代测序 (next generation sequencing)”是指使用以本质上并行的方式测定许多（通常数千至数十亿）核酸序列的方法测定序列，即，其中不是一次一个地，而是在大批量程序中准备用于测序的 DNA 模板，且其中优选地并行或可选择地使用本身可能并行化的超高通量系列程序读出许多序列。这些方法包括但不限于焦磷酸测序（例如，如由 454Life Sciences, Inc., Branford, CT 商售）；通过连接测序（例如，如在 SOLiD™ technology, Life Technology, Inc., Carlsbad, CA 商售）；通过使用修饰的核苷酸合成测序（例如，在由 Illumina, Inc., San Diego, CA 的 TruSeq™ and HiSeq™ technology、由 Helicos Biosciences Corporation, Cambridge, MA 的 HeliScope™ 和 Pacific Biosciences of California, Inc., Menlo Park, CA 的 PacBio RS 商售），通过离子检测技术测序 (Ion Torrent, Inc., South San Francisco, CA)；DNA 纳米球测序 (Complete Genomics, Inc., Mountain View, CA)；基于纳米孔的测序技术（例如，如由 Oxford Nanopore Technologies, LTD, Oxford, UK 开发）及类似的高并行化测序方法。

[0065] 术语“ $T_m$ ”被用来指“熔化温度”。熔化温度是一群双链核酸分子被半数解离成单链的温度。用于计算核酸的  $T_m$  的多种等式是本领域熟知的。如由标准参考文献指出的，当核酸在 1M NaCl 的水溶液中时，对  $T_m$  值的简单估算可通过等式  $T_m = 81.5 + 0.41(\% G+C)$  来

计算（参见，例如 Anderson 和 Young, Quantitative Filter Hybridization, in Nucleic Acid Hybridization(1985)）。其他参考文献（例如，Allawi 和 SantaLucia, Jr., Biochemistry, 36:10581-94(1997)）包括可选的计算方法，对于  $T_m$  的计算这些方法考虑了结构和环境以及序列的特征。

### 具体实施方式

[0066] 除非另外指明，本文描述的技术的实践可采用有机化学的、聚合物技术、分子生物学（包括重组技术）、细胞生物学、生物化学和测序技术的传统技术和说明，这些都在从事本领域的人员的技术范围内。这些传统技术包括聚合物阵列合成、多核苷酸的杂交和连接，以及使用标记检测杂交。对合适技术的具体阐述可参考本文的实例获得。然而，当然也可使用其他等效的传统技术。这些传统技术及说明可在标准实验室手册如 Green, 等人，编辑，Genome Analysis:A Laboratory Manual Series(第 I-IV 卷)(1999)；Weiner, Gabriel, Stephens, 编辑，Genetic Variation:A Laboratory Manual(2007)；Dieffenbach, Dveksler, 编辑，PCR Primer:A Laboratory Manual(2003)；Bowtell 和 Sambrook, DNA Microarrays:A Molecular Cloning Manual(2003)；Mount, Bioinformatics:Sequence and Genome Analysis(2004)；Sambrook 和 Russell, Condensed Protocols from Molecular Cloning:A Laboratory Manual(2006)；及 Sambrook 和 Russell, Molecular Cloning:A Laboratory Manual(2002)（全部来自 Cold Spring Harbor Laboratory Press）；Stryer, Biochemistry(第 4 版)(1995) W. H. Freeman, New York N. Y.；Gait, “Oligonucleotide Synthesis:A Practical Approach”(2002) IRL Press, London；Nelson 和 Cox, Lehninger, Principles of Biochemistry(2000) 第 3 版，W. H. Freeman Pub., New York, N. Y.；及 Berg, 等人，Biochemistry(2002) 第 5 版，W. H. Freeman Pub., New York, N. Y. 中找到，为了所有目的本文通过引用将所有这些文献全部并入。

[0067] 注意除非上下文另外清楚地指明，否则如在本文和所附权利要求中使用的，单数形式“一个 (a)”、“一个 (an)”和“该 (the)”包括复数的所指对象。因此，例如，述及“核酸”是指一种或多种核酸，而述及“该测定”包括述及本领域技术人员已知的等效步骤和方法，等等。

[0068] 除非另外定义，本文所用的所有技术和科学术语具有与本发明所属领域的普通技术人员通常理解相同含义。为了描述并公开可与目前描述的发明结合使用的装置、制剂和方法学的目的，本文提及的所有出版物被通过引用并入。

[0069] 当提供了一个范围的值时，应理解在该范围的上限和下限之间的每个中间值和在该指定的范围内的任何其他指定值或中间值包括在本发明内。这些较小范围的上限和下限可被独立地包括在这些较小的范围内，且也包括在本发明内，服从该指定范围内任何特别排除的限值。当指定的范围包括限值中的一个或两个时，排除了所包括的那些限值中的任何一个或两个的范围也包括在本发明内。

[0070] 在以下说明中，阐述了许多具体细节以提供对本发明的更透彻的理解。但是，将对于本领域技术人员明显的是，没有这些具体细节中的一个或多个时，可实施本发明。在其他实例中，未描述本领域技术人员熟知的熟知特征和步骤以避免使本发明难以理解。

[0071] 大体发明

[0072] 本发明的测定系统提供了空间编码的多重测定,其包含 1) 能够用有效的空间编码方案高水平地多重检测的测定;2) 能够根据空间模式递送试剂的设备;和 3) 通过本质上为数字的读出确定的解码。本发明的测定系统检测生物样品中生物靶或指示生物靶的生物活性的存在或不存在和相对量以及该生物靶或活性的位置,生物样品例如布置在支持体例如显微镜载玻片或组织培养皿上的组织切片或其他生物学结构。

[0073] 该测定系统还提供了具有以空间上界定的模式递送试剂的能力的设备。该设备,连同软件、试剂和说明书,为本发明的高度创新的测定系统提供了关键组成部分,允许在目的空间环境中测量许多生物靶或活性,包括基因表达和肽定位。用于这些测定系统中的编码方案允许人们在多重测定的产物被从生物样品中移除并被汇集用于分析之后确定生物靶或活性在生物样品中的位置(或其缺乏)。对编码方案的解码可通过例如下一代测序进行,下一代测序容易地以低的成本提供数百万或数万亿的数据点。然后可将测定结果例如生物靶的量或活性映射回生物样品中的特定位置。该测定系统打开了通往生物样品中细胞功能和调控的复杂空间模式的新分析窗口。

[0074] 图 1 提供了本发明的测定系统 100 的简化概述。在步骤 110 处,提供了附着到支持体的生物样品。该生物样品包含目的生物靶。生物靶可包含任何目的分子,例如核酸(包括,例如 RNA 转录物、基因组 DNA 序列、cDNA、扩增子或其他核酸序列)和蛋白质、酶及类似物。在步骤 120 处,根据已知的空间模式将编码探针递送到生物样品中。编码探针包括可与目的生物靶相互作用的探针和确定被测定的生物靶在样品中的位置并由此可用来将测定结果反向连接到样品中的位置的代码标签。大多数实施方式中的代码标签是寡核苷酸。然而,代码标签还可以是质量标签(mass tag)、荧光标记或其他部分。

[0075] 在某些实施方式中,将编码探针的探针部分和代码标签部分在向生物样品递送之前预偶联。例如,在其中编码探针是寡核苷酸的实例中,可将探针和代码标签序列合成为单一的寡核苷酸。可选择地,可合成或单独获得编码探针的探针部分和代码标签部分并在向生物样品递送之前组合(例如,可合成两个单独的寡核苷酸并通过例如连接来偶联;或可单独地制备抗体和寡核苷酸并在向生物样品递送之前共轭)。另外,如图 2-5 中描述的,单独地合成探针和代码标签(在编码寡核苷酸中)并在测定中的不同步骤将其递送到生物样品(例如,先探针且其后代码标签或反之亦然)。

[0076] 在步骤 130 处,促使编码探针与生物靶反应或相互作用,即,提供促使例如寡核苷酸与核酸靶杂交、酶催化与蛋白质靶的反应、抗体结合表位等的条件。在其中生物靶为核酸的实例中,编码探针通常是寡核苷酸并与靶核酸杂交。在生物靶为蛋白质的实例中,编码探针通常是通过与靶蛋白结合或通过靶蛋白反应来与靶蛋白相互作用的适体、小分子或寡核苷酸共轭的蛋白质(也就是,蛋白质中的一种是另一种的底物)。编码寡核苷酸可经合适的基团通过共轭、化学或光交联及类似的手段与探针(蛋白质)偶联。

[0077] 一旦编码探针与生物靶相互作用,必须在步骤 140 处将与生物靶相互作用的编码探针和未与生物靶相互作用的编码探针分离。在其中生物靶是核酸且编码探针是寡核苷酸的实例中,分离可通过例如从样品中洗掉未杂交的编码探针来实现。相似地,对于基于亲和结合的其他测定,包括使用适体、小分子和蛋白质探针的测定,洗涤步骤可用来去除低亲和力的结合剂。在其中探针经与靶相互作用而被转化的实例中,例如,在经例如蛋白酶裂解或

经激酶磷酸化的肽的实例中,收集所有的编码探针即与生物靶相互作用并被转化的编码探针和未被转化的编码探针二者是方便的。在收集或汇集之后,抗体或其他亲和捕获剂可用来捕获通过添加某个部分(例如,磷酸基团)而被转化的探针。在其中探针已通过裂解而被转化的实例中,可通过例如借助于在转化期间从转化的探针中去除的标签(例如,通过裂解)来捕获未转化的探针或通过裂解位点添加新的标签来分离转化的探针。

[0078] 一旦反应的(转化的)或相互作用的编码探针与未反应的或未相互作用的编码探针分离,优选通过测序来测定反应的和/或相互作用的编码探针的序列。编码探针的序列允许将测定结果映射回在生物样品中的位置。

[0079] 图2提供了体现出用于编码空间信息的组合代码方案的有效实施的本发明的测定系统的简化概述。为了该概述的目的,探针是寡核苷酸,但是如其他地方解释的,也可使用其他类型的探针。在步骤210中,提供了附着到支持体的生物样品,例如组织样品或其他生物结构。在步骤220中,向该生物样品递送一个或多个寡核苷酸探针,其中寡核苷酸探针能够与生物样品中的生物靶杂交。在步骤230中,促使寡核苷酸探针与核酸靶相互作用(杂交);也就是说,提供其中寡核苷酸探针可与靶核酸杂交的合适条件。

[0080] 在步骤240中,去除未与靶核酸杂交的寡核苷酸探针,并从而与未与靶核酸杂交的寡核苷酸探针分离。在该实施方式中,分离可通过例如,洗涤样品去除未杂交的寡核苷酸探针来实现。接下来,在步骤250中,根据所选择的模式向生物样品递送编码探针(编码剂),其中编码寡核苷酸包含用来编码生物靶在生物样品中的位置的代码标签。注意与图1的测定系统相比,此处是在分开的步骤中递送探针和编码剂(编码寡核苷酸)。在步骤260中,编码寡核苷酸与寡核苷酸探针偶联以制备编码探针。在其中探针是寡核苷酸的实例中,编码寡核苷酸可通过例如连接与寡核苷酸探针偶联。可选择地,可通过使用DNA聚合酶来延长起引物作用的探针寡核苷酸而转移编码寡核苷酸中的信息,并从而复制和并入编码寡核苷酸的序列。

[0081] 在步骤270中,测定了编码探针中代码标签的序列以及探针本身的序列或一部分序列,并在步骤280中,将靶核酸映射回生物样品。在某些实施方式中,序列的丰度揭示出生物靶在该位置的相对量。虽然该实施方式以特定的顺序展示了单独的步骤,但是为了更好地解释本发明,这些步骤的精确顺序可改变。例如,可组合步骤220和步骤250,以根据所选择的模式递送探针与编码寡核苷酸的混合物。然后可在组合的步骤220和步骤250之后立即进行或与这两个步骤同时进行偶联步骤260。在该实例中,则步骤240将在步骤260之后发生。因此可理解这一系列的步骤的两个关键结果,即,探针分子的位置特异性编码和基于探针分子与相应的靶分子的相互作用的能力分离探针分子可在特定步骤的实施中以一定灵活性实现。相似地,在代码方案的设计方面存在极大的灵活性。如下文描述的,本发明的测定特别适用于组合方法。

[0082] 因此,本发明提供了观看许多位置中的许多生物靶的能力,提供了以高度并行的测序数据分析对原位杂交的分辨。在某些实施方式中,被测试的多种生物靶和生物样品中的多个位点的和大于20,在其他实施方式中,被测试的多种生物靶和生物样品中的多个位点的和大于50,在其他实施方式中,被测试的多种生物靶和生物样品中的多个位点的和大于100、大于500、1,000、10,000、25,000、100,000、500,000、1,000,000。应理解,由于本发明的空间编码尺寸,可设想甚至大得多的数字。例如,测定每个位置10,000靶 $\times$ 10,000

位置将产生  $10^8$  种不同的测定,且可容易地设想甚至比这些数更大的数,特别是如果使用具有单细胞分辨率级别的分辨率的空间位置的话。另外,在其中采用高通量数字测序的实施方式中,通常并行测定至少 1,000 个编码探针的序列。更通常地,使用数字读出,对于每个测定期望获得多个序列读数(由探针和空间位置码定义)。取决于实验的设计和测定的需要,,期望获得每个测定至少 3 拷贝的平均值,且更通常地,每个测定至少 10 或至少 30 拷贝的平均值。对于具有合适动态范围的量化读出,每个测定可能期望获得至少 1,000 个读数。因此,如果进行 1,000,000 测定,序列读数的数字将是十亿或更大。对于高通量数字测序,并考虑到冗余,并行测定至少 10,000 编码探针的序列,或并行测定至少 100,000、500,000、1,000,000、10,000,000、100,000,000、1,000,000,000 或更多的编码探针的序列。

#### [0083] 测定

[0084] 本发明的测定系统的测定部分包含下列一般步骤:递送探针和编码剂,其中根据已知的空间模式向样品递送编码剂(在某些实施方式中与探针预偶联),促使探针与样品中的生物靶相互作用或反应,且如果探针和编码剂未被预偶联,将编码剂与探针偶联。

[0085] 本发明的样品几乎包括可附着到支持体或主要以二维方式提供的任何生物样品或多个样品,其中将所测定的生物靶或活性反向系于在生物样品内的位置的能力是重要的。示例性的生物样品包括组织切片(例如,包括完整的动物切片和组织活检)、载玻片或组织培养皿上的细胞群及类似物。本发明的测定系统是特别有利的,因为其和许多生物样品类型,包括新鲜样品,例如原代组织切片,和储藏的样品包括但不限于冷冻样品和福尔马林固定的(paraformalin-fixed)、石蜡包埋的(FFPE)样品相容。本发明的测定系统的重要的方面是将生物样品固定到具有不连续的、可独立测量的区域的基材表面上。

[0086] 待检测的生物靶可以是任何生物分子包括但不限于蛋白质、核酸、脂质、碳水化合物、离子或包含以上任何一种的多组分复合物。亚细胞靶的实例包括细胞器,例如线粒体、高尔基体、内质网、叶绿体,内吞囊泡、胞吐囊泡、液泡、溶酶体等。

[0087] 在某些特定的实施方式中,使用该测定系统例如通过基因型分型、DNA 拷贝数或 RNA 转录物的定量、定位样品内的特定转录物及类似手段来分析核酸。图 3 示出了可用于本发明的测定系统的用于例如,检测单核苷酸多态性(SNP)的示例性测定的整体方案。在图 3 中,提供了两个寡核苷酸探针。每个寡核苷酸探针包含在 305 和 307 看到的靶特异性区域(位于待分析的 SNP 的任何一侧)和在 301 和 303 看到的连接区域。促使寡核苷酸探针与生物样品中的靶核酸(未示出)杂交。在步骤 302 处,寡核苷酸探针之一被延伸以并入 SNP 序列并与另一个探针连接以形成包含靶核酸区域 309 和连接区域 301 和 303 的延伸的探针。

[0088] 将包含代码标签(在 315 和 317 处看到)、连接区域(在 311 和 313 处看到)和引物区域(在 319 和 321 处看到)的两种编码剂在步骤 304 与延伸的探针相组合并连接以形成编码的靶特异性寡核苷酸。再次地,与图 1 相比,在分开的步骤中递送探针和编码剂。这样做允许使用下文描述的组合实施方式。在优选的实施方式中,一对编码寡核苷酸内的编码寡核苷酸与靶序列的一侧或另一侧(即,靶序列的 5' 或 3' 端)特异性连接。另外,通常编码寡核苷酸和探针的连接区域和引物区域是通用的;也就是,在构建探针和编码寡核苷酸中使用的连接区域和引物区域的组合是恒定的,且只有探针的靶特异性区域和编码核苷酸的代码标签不同。但是,又在替代实施方式中,连接区域和引物区域不是通用的,并在探

针和编码剂之间不同。

[0089] 连接以后,洗脱、汇集编码探针,并任选地通过 PCR 向编码探针加入测序接头。在替代实施方式中,测序引物可与编码寡核苷酸连接,或测序引物序列可被包括作为编码寡核苷酸的一部分。如图 3 中所看到的,每个测序接头包含与编码探针上的引物区域 319 和 321 相容的引物区域 319 或 321。现在准备将包含第一接头 327、第一引物区域 319、第一代代码标签 315、连接区域 311 和 301、靶区域 309、连接区域 313 和 303、第二代代码标签 317、第二引物区域 325 和第二接头 329 的最终构建体输入到数字高通量测序程序中。

[0090] 图 3 中例示了延伸反应和连接反应的组合,但是应当理解多种反应可用来偶联编码寡核苷酸和靶特异性寡核苷酸,包括仅连接(例如,对于与靶核酸序列的连续部分杂交的寡核苷酸)。可选择地,可采用使用了另外的寡核苷酸的测定,例如在 GOLDENGATE® 测定中(参见 Fan, 等人, Cold Spring Symp. Quant. Biol., 68:69-78(2003); (Illumina, Inc., San Diego, CA))。

[0091] 在其他实施方式中,本发明的测定系统还可用来分析生物样品中的肽或蛋白质、抗体的存在、酶活性和其他蛋白质的活性、翻译后修饰、肽的活性和非活性形式以及肽的异构体(isoform)。因此,探针可包含酶的活性区域、免疫球蛋白的结合结构域、蛋白质的特定结构域、完整的蛋白质、合成的肽、引入突变的肽、适体和类似物。

[0092] 在某些方面,探针是酶或酶原(proenzyme)例如激酶、磷酸酶、酶原(zymogen)、蛋白酶或其片段的底物。在某些方面,探针是用来检测参与一种或多种信号转导通路的蛋白质例如激酶或磷酸酶的磷酸化底物。在本发明的另一个特定的方面,探针是仅与单独的蛋白酶或蛋白酶种类结合的特定蛋白酶底物。在其他方面,探针是酶的不同加工形式、异构体和/或结构域。基于蛋白质的探针通常与寡核苷酸编码剂共轭或以其他方式与其连接。在该实例中的寡核苷酸编码剂还将包括促进蛋白质探针鉴定的核苷酸序列组分。

[0093] 在某些方面,本发明提供了用于评价样品中的不同位置之间和/或样品之间生物靶的量和/或活性的差异的测定。该方法包括确定来自生物样品的多个编码结果和评价生物靶在生物样品中的每个位置的数量的差异。

[0094] 组合实施方式

[0095] 为了将编码效率最大化,可使用在编码寡核苷酸中利用成对代码标签的组合方法。通过将靶特异性信息和编码标签去偶联,所需要的寡核苷酸的数目急剧减少,伴随成本减少。

[0096] 图 4 示出了用于本发明的测定系统的组合编码方案的一个实施方式的一般机制,其中测定了代表性组织切片中的核酸(在 416 处示出)。图 4 在 A 处示出了特异性结合目的靶核酸 402 的两个靶特异性/编码寡核苷酸构建体 420 和 422(例如,在图 3 的步骤 302 和 304 之间形成)。第一编码探针 420 包括与例如用于扩增测定产物的通用引发位点或接头相关联的代码标签 408 以使得能够使用测序技术 404 鉴定代码识别剂(coding identifier)。第二编码探针 422 包括与例如用于扩增测定产物的通用引发位点或接头相关联的代码标签 406 以使得能够使用测序技术 410 鉴定编码识别剂。

[0097] 图 4 在 B 处示出了可用于二十种不同的代码标签 a1 到 a10(编码探针 420 上的代码标签 406)和 b1 到 b10(编码探针 422 上的代码标签 408)的空间模式。例如代码标签 a1 安置在生物样品上 10 个离散的区域或斑点(在 412 中示为第一水平线的斑点)中。代码

标签 a2 安置在生物样品上在 412 中第二水平线上的 10 个斑点中。代码标签 a3 安置在生物样品上在 412 中第三水平线上的 10 个斑点中,等等。鉴于“a”标签安置在 10 个水平的排中,如 414 中所示,“b”标签安置在 10 个垂直的排中。例如,代码标签 b1 安置在生物样品上在 414 的第一垂直排的十个离散的斑点中,代码标签 b2 安置在生物样品上在 414 的第二垂直排的十个离散的斑点中,等等。使用该设置允许 20 个代码标签唯一地界定生物样品上 100 个不同的位置。

[0098] 图 4 在 C 处示出了与代码标签网格 418 一致的代表性组织切片 416。箭头表明了“a”代码标签和“b”代码标签如何安置在与组织切片 416 一致的网格 418 上。如果一旦测序,代码标签 a1 和 b4 例如与靶核酸序列相关联,那么该靶核酸序列(即,生物靶)呈现在组织切片中的位置 a1、b4 处。

[0099] 图 5 提供了用于本发明的测定系统的编码方案的简化的特定实例。图 5 示出了包含 a1、a2、a3、a4 和 b1、b2、b3 及 b4 的编码寡核苷酸 510。在 520 处示出了靶特异性寡核苷酸(TSO)(探针)1 和 2。在 530 处示出了安置或分配方案。如图 4 中例示的网格一样,编码寡核苷酸 a1 至 a4 以某种模式(此处,以竖直模式)安置在斑点中,且编码寡核苷酸 b1 至 b4 以某种模式(此处,以水平模式)安置在斑点中。网格虽然被示为具有斑点的正方形,但实际上是在生物样品(未示出)例如图 4 中所示的组织切片 416 上的安置模式。

[0100] 将靶特异性寡核苷酸递送到生物样品中,其中靶特异性寡核苷酸与生物样品中的靶核酸杂交(如果存在靶核酸的话)。然后通过例如洗涤去除未杂交的靶特异性寡核苷酸。然后根据在 530 处所示的空间模式,将编码寡核苷酸递送到生物样品中。编码寡核苷酸与生物样品中的靶核酸杂交的任何靶特异性寡核苷酸连接(或,例如,延伸并连接),然后将连接的构建体从生物样品中洗脱出来,汇集并通过例如 PCR 或连接加入测序接头(如果该序列之前未被包含在编码寡核苷酸内的话)。通过例如高通量或“下一代”测序对连接的构建体测序。在 540 处示出了所得序列的集合(pool)。仅在 a4b1、a4b2、a1b3、a2b3、a3b3、a4b3 和 a4b4(用水平线示出的位置)获得了靶特异性寡核苷酸 1 的序列读出。仅在 a1b1(用竖直线示出的位置)获得了靶特异性寡核苷酸 2 的序列读出。在位置 a2b1、a3b1、a1b2、a2b2 和 a3b2 处(用交叉影线示出的位置)获得了靶特异性寡核苷酸 1 和 2 的序列读出。未在 a1b4、a2b4 或 a3b4 处(无阴影示出的位置)获得任一个靶特异性寡核苷酸的序列读出。因此,在进行测定的生物样品中,在生物样品的左侧的大部分并在底部检测到第一靶核酸,仅在生物样品的左上部分检测到第二靶核酸,且未在生物样品的右上部分检测到任何一种靶核酸。现在可将两种靶核酸的不同表达映射回生物样品并映射生物样品的这些位置中的生物结构或细胞类型。除了位置信息,可获得与编码标签的相对丰度相关的信息。例如,如果发现与 a4T1b2 序列相比多达十倍的 a4T1b1 序列出现在数据组中,这将表明靶核酸序列 1 在 a4T1b1 位置比在 a4T1b2 位置丰富十倍。在如图 3 所示的核苷酸分析的实例中,通过将代码标签与靶特异性寡核苷酸直接连接,对于 n 个靶仅需要 2n 个靶特异性寡核苷酸。例如,使用图 2 中概括的组合方法,在 10,000 个空间位置测定 100 个不同的靶将需要  $2 \times 100$  个靶特异性寡核苷酸和  $2 \times 100$  个编码寡核苷酸。不计算通用引物,测定寡核苷酸的总数将仅为 400(200 个靶特异性的和 200 个编码的寡核苷酸)。相比之下,如果代码寡核苷酸未与靶特异性寡核苷酸去偶联,将需要  $(n \times X \text{ 个位置代码}) + (n \times Y \text{ 个位置代码})$ ,或在上述实例中,不计算通用引物序列,需要 20,000 寡核苷酸。此外,虽然图 2-5 中所

示的实施方式描绘了使用两种编码剂（代码标签）的组合方案，但是可使用三种、四种或更多种编码剂和代码标签，并通过不同的方法在不同的步骤组合中与探针连接或互相连接。

[0101] 由于本发明的测定系统的空间编码方面，用甚至数目适中的测定就可产生大量的信息。例如，在样品中的 5 个或更多个位置测定的 5 种或更多种生物靶产生 25 种或更多种组合。使用数字测序作为读出，每种组合的序列读数的最佳数目取决于所需要的灵敏度和动态范围，并可调整。例如，如果对于每种组合抽出平均 100 个读数，对于 25 种组合的总数是 2500 个读数。如果以平均 1,000 的取样深度在 1,000 个位置测定 1,000 个靶，则需要  $10^9$  个读数。这些数字，虽然大，但是在本质上并行的数字测序方法的能力之内，这些方法可在合理的时间表中并以每个读数非常低的成本产生数十亿或甚至数万亿读数的数据集。因此，通过改变待检测的位置或待测定的生物靶的数目或两者都改变，并使用数字测序，可获得大量的信息。在特定的方面，对于两种或更多种生物分子检测多个位置。

#### [0102] 试剂递送系统

[0103] 本发明的试剂递送系统包括允许向生物样品的离散部分递送试剂的设备，维持编码方案的空间模式的完整性。本发明的测定系统的试剂递送系统包含任选的成像装置、试剂递送硬件和控制软件。试剂递送可以多种不同的方式实现。应当注意试剂递送可能一次针对许多不同的生物样品。本文已例示了单一的组织切片；然而，可同时附着并分析多种生物样品。例如，可并行分析组织样品的连续切片并组合数据以建立 3D 图谱。

[0104] 对于本发明的测定系统不可缺的是允许试剂到达生物样品上的空间模式的设备。用于配制并递送生物分子（例如，寡核苷酸或抗体）和化学试剂（例如，小分子或 dNTP）的技术是本领域已知的且这些设备系统的使用是本领域技术人员已知的并容易适用本发明的测定系统。合适的试剂递送系统的一个实例是 Labcyte™Echo 声波移液设备（Labcyte™Echo acoustic liquid handler），其可用来以高的准确度和重复性递送包含生物分子的纳升（nanoliter）尺度的小滴。使用软件来具体说明试剂应被递送的位置，本领域技术人员可以将该试剂递送装置集成到整个系统中。

[0105] 可用于将剂和 / 或代码识别剂安置到生物样品上的其他设备包括但不限于喷墨点涂（ink jet spotting）；通过大头针、笔或毛细管的机械点涂；微接触印刷；光化学或光刻方法；及类似方法。对于多种应用，将生物样品的某些区域分割或隔离成用于不同的试剂分布和 / 或生物靶测定的一个或多个测定区域可能是优选的。可使用屏障或管道将这些测定区域物理地分隔开。

[0106] 在一个示例性的方面，试剂递送系统可以是基于流的系统（flow based system）。在本发明中用于试剂递送的基于流的系统可包括例如一个或多个泵、阀、流体贮池（fluid reservoir）、管道和 / 或试剂储存池的设备。试剂递送系统被设置成使流体移动以接触生物样品的离散部分。这些试剂的移动可由布置在例如，流体试剂下游的泵来驱动。泵可将每种流体试剂驱送到（并经过）反应区室。可选择地，可通过重力驱使试剂穿过流体。美国公布第 20070166725 和 20050239192 号公开了可用于本发明的测定系统的某些一般目的的应用流体学（fluidics）工具，允许用相对简单的硬件精确地操纵气体、液体和固体以完成非常复杂的分析操作。

[0107] 在更具体的实例中，一个或多个流通池（flow cell）可与来自上文的附着基材的生物样品连接。流通池可包括与其连接的入口管和出口管并且任选地，外部泵用来将试剂



递送到流通池并经过生物样品。流通池被设置为仅将试剂递送到生物样品的某些部分,限制递送到生物样品的任何具体部分的试剂的量和类型。

[0108] 在另一个方面,可将微流体系统集成到其上布置生物样品的基材中或从外部附接到基材的顶部。用于容纳并携带流体的微流体管道可通过与基材邻接的应用流体层形成在平面基材上和/或其上方。可根据选择性地开放和关闭布置在试剂池之间的阀来选择和递送流体试剂。

[0109] 泵通常包括用于移动流体和/或设置在流体中的试剂的任何机构。在某些实例中,泵可被设置成使流体和/或试剂移动经过具有小体积的管道(即,微流体结构)。除了其他方法外,泵可通过对流体和/或携带流体的结构施加正压或负压来机械地操作,或通过电场的适当应用来电动地操作,或两者兼有。示例性的机械泵可包括注射泵、蠕动泵、旋转泵、加压气体、移液器等。机械泵可被微型机械化、模制等。示例性的电动泵可包括电极且可通过电泳、电内渗(electroendoosmosis)、电毛细管、双向电泳(包括其行波形式)和/或类似的方法操作。

[0110] 阀通常包括用于调控流体通过管道的任何机构。阀可包括例如,可被选择性地变形以部分或完全地关闭管道的可变形的构件,可选择性地延伸进入通道以部分或完全地堵塞管道的可移动的突出部,电毛细管结构和/或类似的结构。

[0111] 开口垫圈可与生物样品的顶部附接并可样品和试剂注射到垫圈中。合适的垫圈材料包括但不限于氯丁橡胶、腈和硅橡胶。可选择地,防水反应室可由夹在基材上的生物样品和化学上惰性的防水材料例如但不限于黑色电镀铝、热塑性塑料(例如,聚苯乙烯、聚碳酸酯等)、玻璃等之间的垫圈形成。

[0112] 在可选的实施方式中,测定系统包括确定目的生物样品的特征和组织的成像装置。所获得的图像例如可用来设计试剂的安置模式。成像装置是任选的,因为人们可使用例如显微镜来代替观察生物样品,分析生物样品的组织并具体说明递送测定试剂的空间模式。递送系统(如果包括的话)可包含微电路构造,所述微电路构造包括成像器例如基于 CCD 或 IGFET(例如,基于 CMOS)的成像器和例如通过引用并入本文的美国公布第 20090197326 号中描述的用于试剂递送的超声喷雾器。另外,应注意虽然图 4 和 5 使用 x,y 网格构型阐述,可使用其他构型,诸如,例如遵循组织样品的拓扑学;靶向组织中的某些组的细胞、细胞层和/或细胞类型,及类似的构型。

[0113] 在又一个替代方案中,试剂递送系统使用半导体技术例如掩蔽和喷涂控制试剂递送到生物样品表面上的具体模式。通过使用掩蔽物保护特定区域免于暴露,可避免生物样品的特定区域暴露于试剂。可使用传统技术例如喷涂或流体流将试剂引入到生物样品。掩蔽递送的使用促成基材表面上模式化的递送方案。

[0114] 在本发明的优选的方面,试剂递送的设备基于喷墨印刷技术。存在多种不同的喷墨机制(例如,热式、压电式)且已显示与含水墨制剂和有机墨制剂的相容性。数组独立驱动的喷嘴可用来同时递送多种试剂,并达到非常高的分辨率。

[0115] 为了靶向特定的目的位点,待测定的生物样品的信息图像可用来辅助试剂递送方法及相关的编码方案。可使用图像处理(例如,通过免疫组织化学或其他染色化学方法区分的细胞类型的图像)结合测定系统的其他特征来鉴定生物样品的取样区域。在某些方面,使用软件将图像信息自动翻译成试剂递送模式。因此非常准确地记录并比对生物样品

的试剂递送的机构 (mechanism) 是本发明的测定系统的重要组成部分。机构例如在载玻片上的基准标志 (fiducial marker) 和 / 或其他非常精准的物理定位系统的使用可适用于该目的。

[0116] 本发明优选地包括一整套适合该测定系统的软件。任选地,寡核苷酸设计软件用来设计待进行特定测定的编码核苷酸 (且在其中测定核酸的实施方式中,为靶特异性寡核苷酸) 并可被集成成为该系统的一部分。还任选地,可将用于试剂递送和数据分析 (即,序列分析) 的算法和软件集成在一起以确定测定结果。集成数据分析是特别有用的,因为由于规模的结果,所产生的数据组的类型可能是巨大的。为分析通过该测定系统产生的空间上关联的数据而专门设计的算法和软件工具,包括模式分析软件和可视工具增强了由该测定系统所产生的数据的价值。

[0117] 在某些方面,该测定系统包含用于进行和实施试剂质量控制例如,寡核苷酸库的完整性和序列保真性的程序。特别地,根据例如挥发性、在关键温度下的稳定性及化学相容性的因素配制试剂以与试剂递送设备相容并可通过集成在测定系统内的设备来分析这些试剂。

#### [0118] 测序

[0119] 许多方法可用来鉴定本发明的测定系统的编码探针中的代码标签和探针序列。可使用例如质谱 (例如,LC-MS/MS 的 MalDI-T)、核磁共振成像或优选地,核酸测序的技术来检测代码标签。用于对本发明的代码标签解码的技术的实例可在例如通过引用并入本文的美国公布第 20080220434 号中找到。例如,代码标签可以是寡核苷酸质量标签 (OMT 或质量标签)。例如,在通过引用整体并入本文的美国公布第 20090305237 号中描述了这些标签。在又一个替代方案中,可扩增编码探针并与微阵列杂交。这将需要进行单独的扩增反应,其中每次扩增是对特定的空间代码或代码的子集特异的,通过使用代码特异性引物来实现。每次扩增还将并入不同的可分辨标记 (例如,荧光团)。杂交后,可通过可分辨标记的相对丰度测定映射到样品中不同的空间位置的特定靶的相对量。

[0120] 在一个特别优选的方面,所得的根据该测定系统的代码标签是高通量、下一代测序的底物,并使用高并行下一代测序方法来确认代码标签的序列,例如用 SOLiD™ 技术 (Life Technologies, Inc.) 或 Genome Analyzer (Illumina, Inc.)。该下一代测序方法可使用例如一轮测序 (one pass sequencing) 方法或使用双末端测序进行。下一代测序方法包括但不限于诸如在例如 Drmanac 美国专利第 6,864,052 ;6,309,824 ;和 6,401,267 号 ;和 Drmanac 等人,美国专利公布 2005/0191656 中公开的基于杂交的方法 ;通过合成测序 (sequencing-by-synthesis) 的方法,例如美国专利第 6,210,891 ;6,828,100 ;6,969,488 ;6,897,023 ;6,833,246 ;6,911,345 ;6,787,308 ;7,297,518 ;7,462,449 和 7,501,245 号 ;美国公布申请第 20110059436 ;20040106110 ;20030064398 和 20030022207 号 ;Ronaghi, 等人, Science, 281:363-365(1998) ;和 Li, 等人, Proc. Natl. Acad. Sci., 100:414-419(2003) ;基于连接的方法,例如美国专利第 5,912,148 和 6,130,073 号 ;和美国专利申请第 20100105052、20070207482 和 20090018024 号 ;纳米孔测序例如美国专利申请第 20070036511 ;20080032301 ;20080128627 ;20090082212 号 ;和 Soni 和 Meller, Clin Chem 53:1996-2001(2007)), 以及其他方法,例如美国专利申请第 20110033854 ;20090264299 ;20090155781 和 20090005252 号 ;还参见 McKernan, 等

人, *Genome Res.*, 19:1527-41 (2009) 和 Bentley, 等人, *Nature* 456:53-59 (2008), 为了所有目的将所有这些文献通过引用整体并入本文。

#### [0121] 测定系统的应用

[0122] 在阅读本公开内容后将对于本领域技术人员明显的是:存在将受益于可同时测量生物靶在生物样品中的量和空间位置的高通量多重测定系统的许多重要的生物研究、诊断和药物开发的领域。例如,组合估计不同 RNA 转录物的相对丰度的能力和重新构建遍及许多位置的丰度的空间模式的图像(可能和组织中的单个细胞一样小或甚至比它还小)的能力使得许多不同的基础研究领域成为可能。以下是示例性的用途且绝不意味着在范围上是限制性的。

[0123] 在一个实例中,以与 CT 扫描中的重新构建图像的方式相似的方式通过分析一系列的切片确定基因表达的 3 维模式。该方法可用来测量疾病病理中例如在癌性组织和/或经损伤、炎症或感染的组织中基因表达的变化。通过本发明的测定系统,获得了关于复杂组织中的基因表达和蛋白质定位的更详细的信息,导致对正常和疾病状态中的功能和调控的新认识,并提供了可被检验的新假设。例如,本发明的测定系统可使得从许多单独的研究和较大的项目如 ENCODE (Birney, 等人, *Nature*, 447:799-816 (2007)) 和 modENCODE 获得的认识能够在组织水平上整合。该测定系统还辅助计算机尝试来模仿系统生物学领域基因表达的相互作用网络。

[0124] 该测定系统还提供了分析体细胞变异例如癌症中的体细胞突变或响应于受感染的生物体的变异性的新方法。例如,肿瘤通常是高度异质性的,包含癌细胞以及在异常局部环境中的遗传学上正常的细胞。癌细胞进行突变和选择,且在该过程中形成局部克隆并不是不常见的。在肿瘤环境中鉴定相对稀少的体细胞突变可使得研究关键突变在克隆变异体的选择中的作用成为可能。可分析在癌细胞和遗传学上正常的细胞中与血管生成、炎症或其他癌相关过程有关的转录模式以获得对癌症生物学的认识并有助于开发用于治疗癌症的新型治疗剂。在另一个实例中,个体对感染性生物体具有不同的敏感性,且本发明的测定系统可用来研究微生物与组织或组织内的多种细胞类型之间的相互作用。

[0125] 重要的是,因为在任何给定的反应中由于只测定了少数位置,信噪比 (signal to noise) 可急剧增加,所以除了提供空间上相关的信息,本发明允许检测稀少突变的灵敏性大幅增加。在对混合的样品中的稀有突变的典型测定中,批量处理该样品,即,将核酸从许多细胞提取到单一的池中。因此,如果突变存在于 10,000 个细胞中的一个细胞内,必须针对来自~10,000 个细胞的正常 DNA 的背景检测该突变。相比之下,通过本发明的测定系统,可分析许多细胞,但是单独的细胞或少数几组细胞将被空间代码系统鉴定。因此,在本发明的测定系统中,背景成数量级地减少,极大地增加了灵敏性。此外,可观察突变细胞的空间组织,这对于在癌症组织切片中检测关键突变可能是特别重要的。已有的分子组织学分析正在产生对癌症生物学的认识,且可能具有用于诊断的潜力。本发明的技术有希望极大地增加这些方法的力量。

#### [0126] 实施例

[0127] 提供以下实施例以给本领域普通技术人员提供如何实施和使用本发明的完整公开和描述,且这些实施例不意在限制发明人所认为的其发明的范围,这些实施例也不意在代表或意味着以下实验是所进行的全部实验或仅有的实验。本领域技术人员应理解可如具

体实施方式中展现的那样对本发明进行许多改变和 / 或修改而不偏离宽泛地描述的本发明的精神或范围。因此,无论在哪个方面,应认为该实施方式是说明性而不是限制性的。

[0128] 已经进行努力以确保关于所使用的数字(例如,量、温度等)的准确性但是应考虑某些实验误差和偏差。除非另外指明,份是以重量计的份,分子量是重均分子量,温度以摄氏度计,且压力是处于或接近大气压。

[0129] 实施例 1:编码方案概念的初步验证

[0130] 作为概念的初步验证,使用微阵列开发模型系统来阐释工作的单重测定(working single-plex assay)。基础设计证实了该测定的概念,并在解决涉及分析更复杂的生物样品的问题之前建立了工作的测定。使用传统测序作为这种概念验证的读出。

[0131] 使用微阵列作为组织切片的代用品。充分地说明了微阵列的靶序列,以致于靶的组成是已知的并可系统地改变。合成的寡核苷酸模板通过 5' 氨基修饰与玻璃载玻片附接。每个载玻片具有单一的寡核苷酸模板序列,且所进行的测定可采用连接或延伸然后连接,因为这在测定某些多态性中可能是有用的。

[0132] 一旦该测定的原位部分完成,洗脱反应产物并通过 qPCR 分析以确定产物的存在或不存在并估算收率,并通过传统测序来确定所测定产物的结构。所测试的单重测定包括合适的阳性和阴性对照,和检验区分单碱基变体的能力的单核苷酸变体(SNV)。

[0133] 实施例 2:可扩展性

[0134] 增加测定系统的复杂性以确定该测定用于高通量研究的可扩展性。空间编码和测定系统的可扩展性通过使用微阵列模型系统进行 24-重  $\times$  24-位点测定来阐释。

[0135] 生物靶,此处为 DNA 靶序列在每一个测定位置的量在微阵列基材上系统地变化。例如,在具有 50 微米斑点尺寸(中心到中心)的微阵列上,1mm<sup>2</sup>面积包含 $\sim$ 400 个斑点。每个位点周围的区域任选地被缺乏这些斑点的区域占据以允许靶序列各自的可分辨性。可选择地,斑点可簇集起来,两个或更多个直接相邻的斑点被缺乏靶序列的区域围绕或与之相邻。

[0136] 为了表明空间编码是准确的,这些位点包含不同的靶组成以显示测定读出与每个位点的预期组成匹配。用 24 个靶序列制备简单的数字模式,每个位点具有 12 个靶存在和 12 个靶不存在的不同集合以制备二进制代码(0 = 不存在,1 = 存在)。然后确定测定读出以表明空间解码后所检测的区域与预期信号匹配。在这个特别的实例中,代码空间是足够大的(2<sup>24</sup>)以致于即使几个误差也不会导致不同的代码被混淆。此外,该设计允许误差的确定且允许不仅估计空间编码准确性而且估计宣称靶序列的存在或不存在的准确性。

[0137] 通过对在 24 位点测定中的每个位点进行的 24 个测定中的每一个产生剂量反应曲线来评价检测定量差异的能力。这允许对检测限、动态范围和贯穿该范围检测给定倍数变化的能力的估计。

[0138] 在一个方面,通过改变对于每个靶的特征的数目,使用拉丁方设计代表不同比例的单个靶。换句话说,对于某位点内的多个斑点,分配到 24 个靶序列中的每一个上的斑点的数目可改变,且 24 个位点中的每一个可具有不同的组成。1 $\times$ 3 英寸的微阵列对于允许多个重复是足够大的。这种较大的 24 个序列的组将需要去卷积,且这使用高通量技术例如下一代测序技术(例如,SOLiD™技术(Life Technologies, Inc., Carlsbad, CA) 或 Genome Analyzer(Illumina, Inc., San Diego, CA)) 来实现。使用 24 重测定证实了空间编码与解码

的准确性和该测定系统的定量反应。

[0139] 实施例 3 :该测定对储藏的样品的适应。

[0140] 测定基因组 DNA 作为用于测定 RNA 的概念的验证,因为它提供了建立单拷贝参考信号的一种方式。一旦发展用于 FFPE 样品的运行的测定,其适用于 RNA 测定。为了这个目的,测定了测定寡核苷酸的浓度以确保与高度多重检测的相容性。假定细胞直径为 10 微米,并向单个细胞递送 10 微米直径的试剂小滴,小滴的体积将是  $\sim 500 \mu\text{l}$  且在  $1 \mu\text{M}$  浓度下可包含  $\sim 3 \times 10^{11}$  个分子。为测定 10,000 个细胞中的 1,000 个靶序列,在一个小滴中将需要  $\sim 2,000$  个靶向寡核苷酸。所以,每个小滴可包含  $\sim 160,000,000$  拷贝的每个测定 oligo,大大超过细胞内的几千个靶序列。

[0141] 增强对从非常少或受损的样品产生的绝对数量少的产物分子的处理以应对回收效率低的组织;也就是说,洗脱是有效的并避免了由表面对分子的吸收导致的损失。解决后一个问题的方法是必须包括载体材料,例如糖原或载体核酸。

[0142] 实施例 4 :使该测定适应生物样品。

[0143] 将对照 RNA 模板固定到固体支持体以制备人工系统。使用 T4DNA 连接酶进行测定,T4DNA 连接酶可修复 DNA/RNA 杂交体中的切口。在相匹配的载玻片或相同载玻片的不同切片上进行测定,其中在一个实例中,测定 gDNA 而在另一个实例中,测定 RNA。当测定 gDNA 时,可用 RNA 酶预处理载玻片且当测定 RNA 时,用 DNA 酶预处理载玻片。通过提取 gDNA 或 RNA 并分别通过 PCR 或 RT-qPCR 量化相对量来确认测定结果。

[0144] 为了使组织切片 RNA 测定尽可能多地提供信息,在测定设计中使用了关于特定组织中跨越一定丰度范围的靶转录物的表达水平的已存在的信息。高丰度转录物以及某些中丰度和低丰度转录物被靶向以使得对测定的定量性能特征的初步评价成为可能。

[0145] 以上仅阐释了本发明的原理。应理解,本领域技术人员将能够设计本文虽未清楚地描述或展现,但是体现了本发明的原理并被包含在其精神和范围内的多种方案。此外,本文提到的所有实例和条件性语言原则上意在帮助读者理解本发明的原理和发明人为深化本领域所贡献的概念,且将应解释为不局限于这些具体提及的实例和条件。此外,本文提及本发明的原理、方面和实施方式以及其具体实例的所有声明意在包括其结构和功能等价物。另外,期望这些等价物包括目前已知的等价物和未来开发的等价物,即不管结构如何,表现同样的功能的开发的任何元件。因此,不期望本发明的范围局限于本文展现和描述的示例性实施方式。更确切地,本发明的范围和精神通过所附权利要求来体现。在随附的权利要求中,除非使用了术语“手段”,否则根据 35 U.S.C. § 112, ¶ 6,其中所提及的任何特征或元件都不应解释为手段加功能 (means-plus-function) 限制。

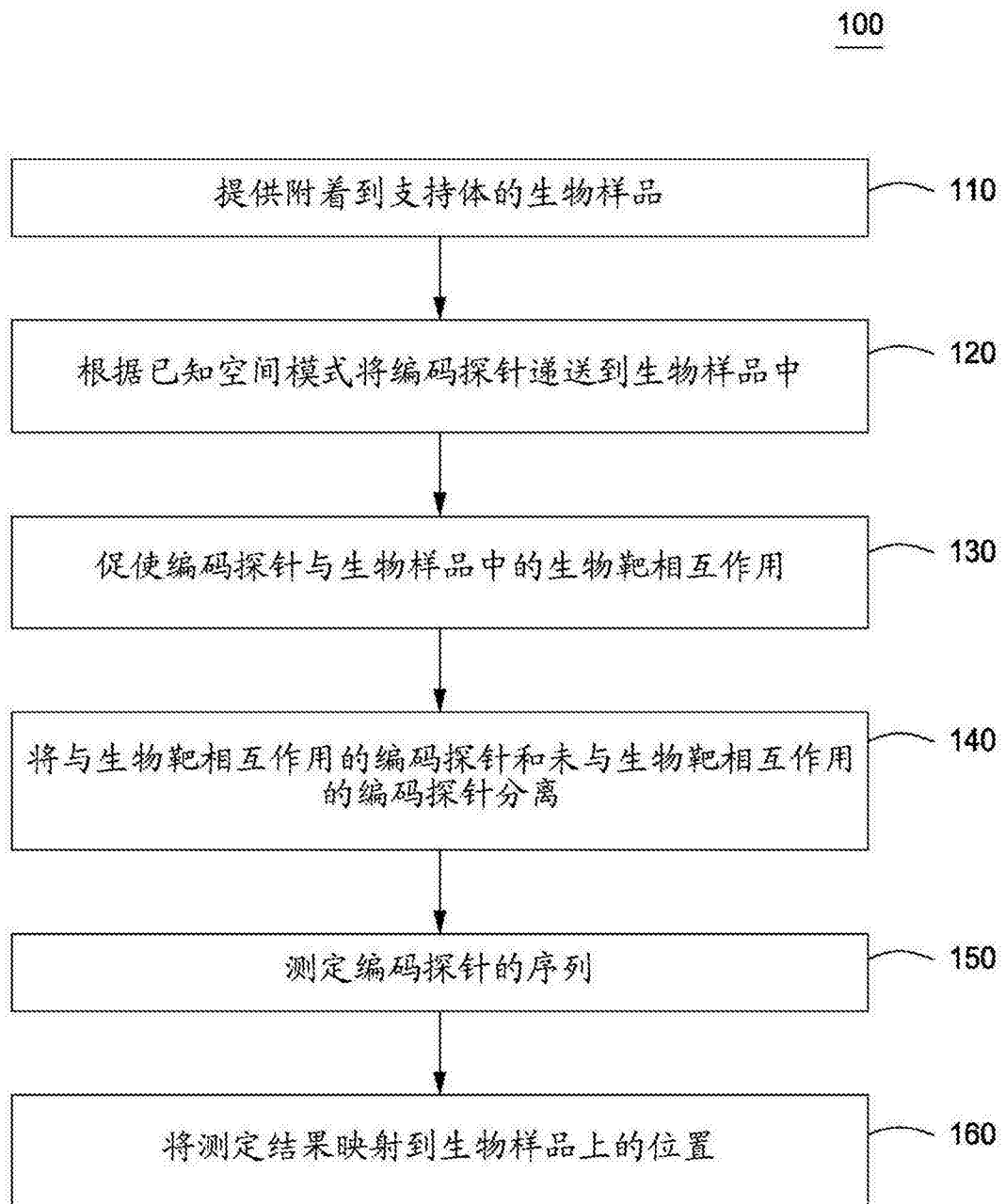


图 1

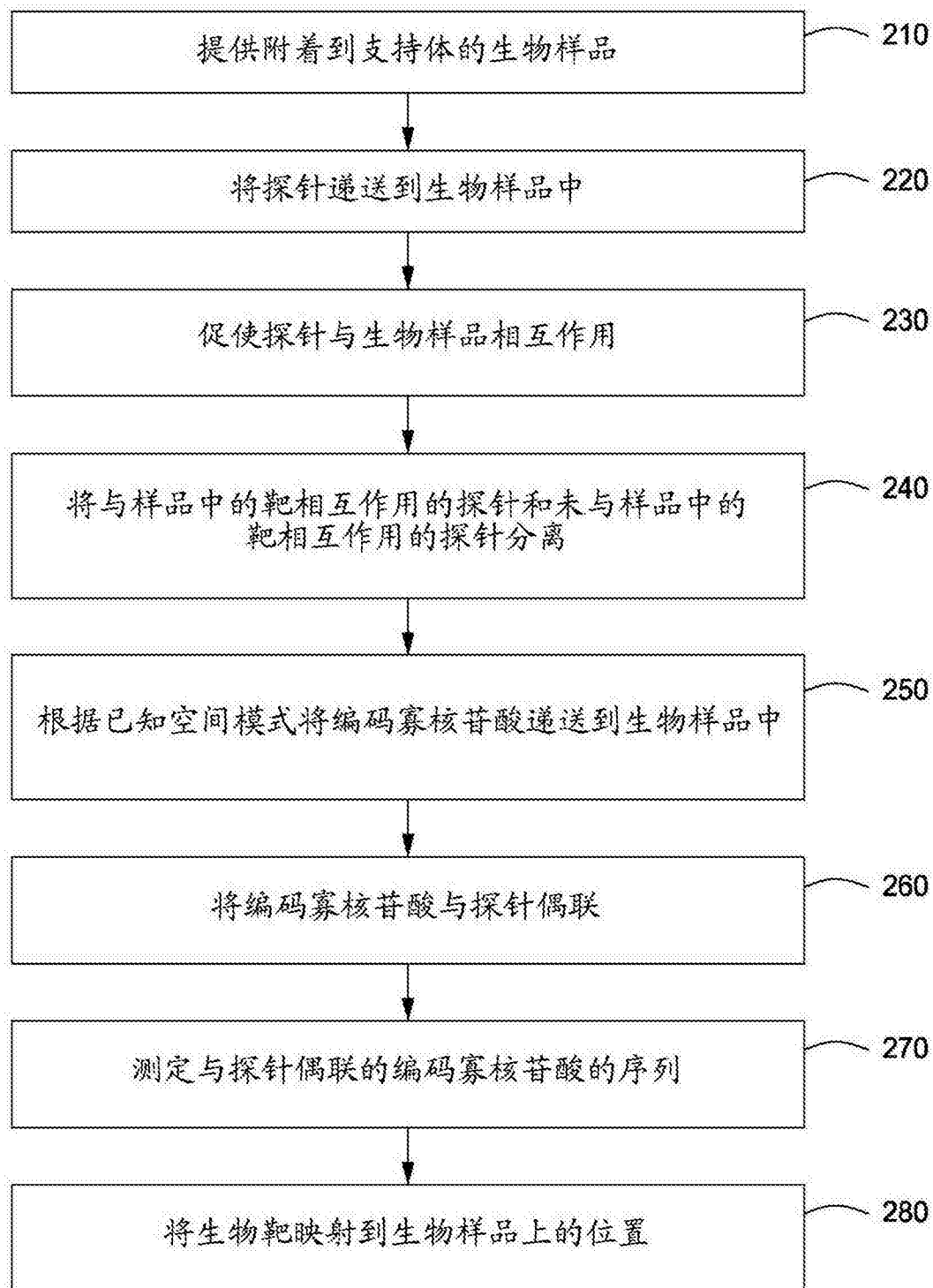
200

图 2

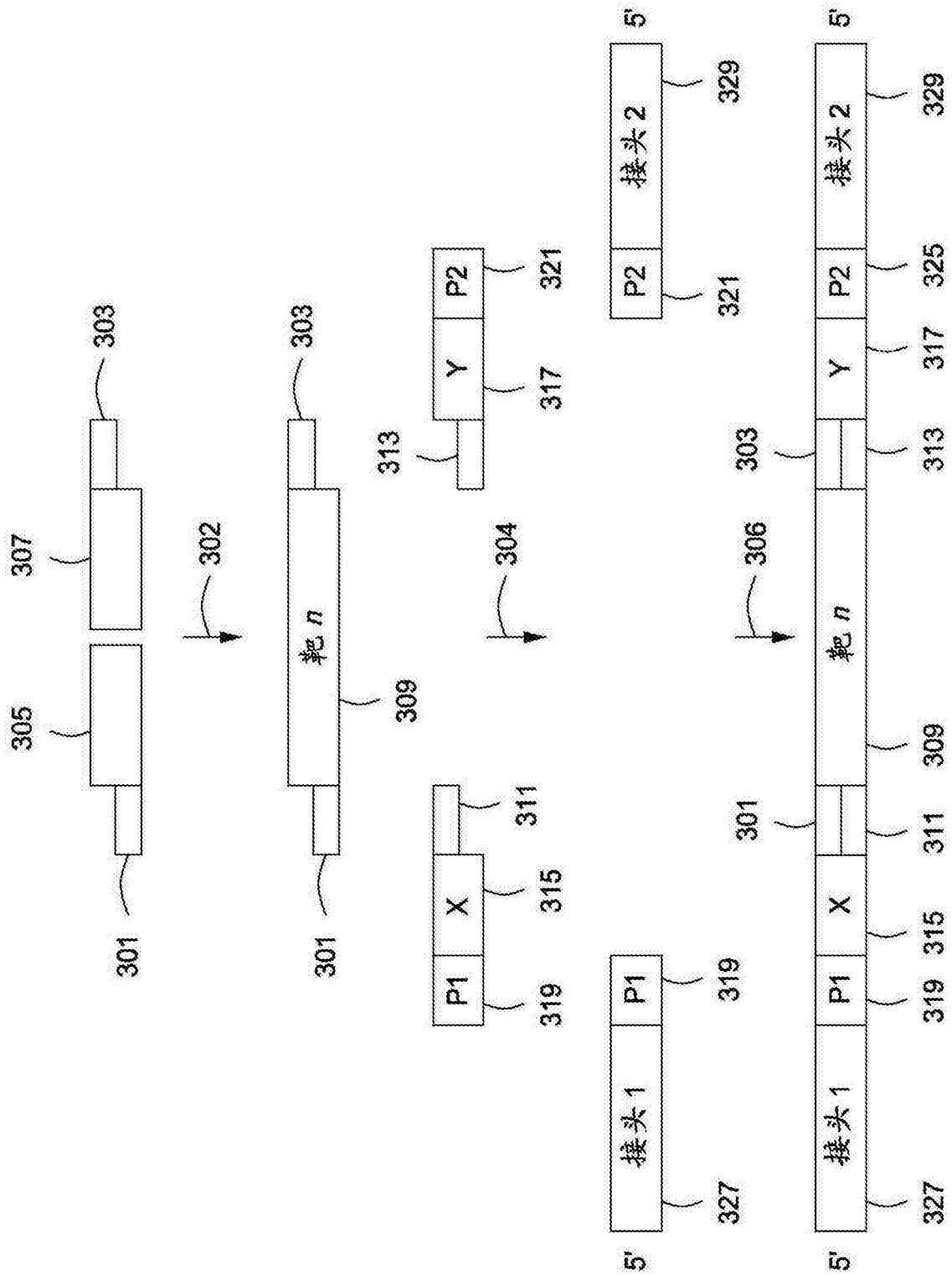


图 3



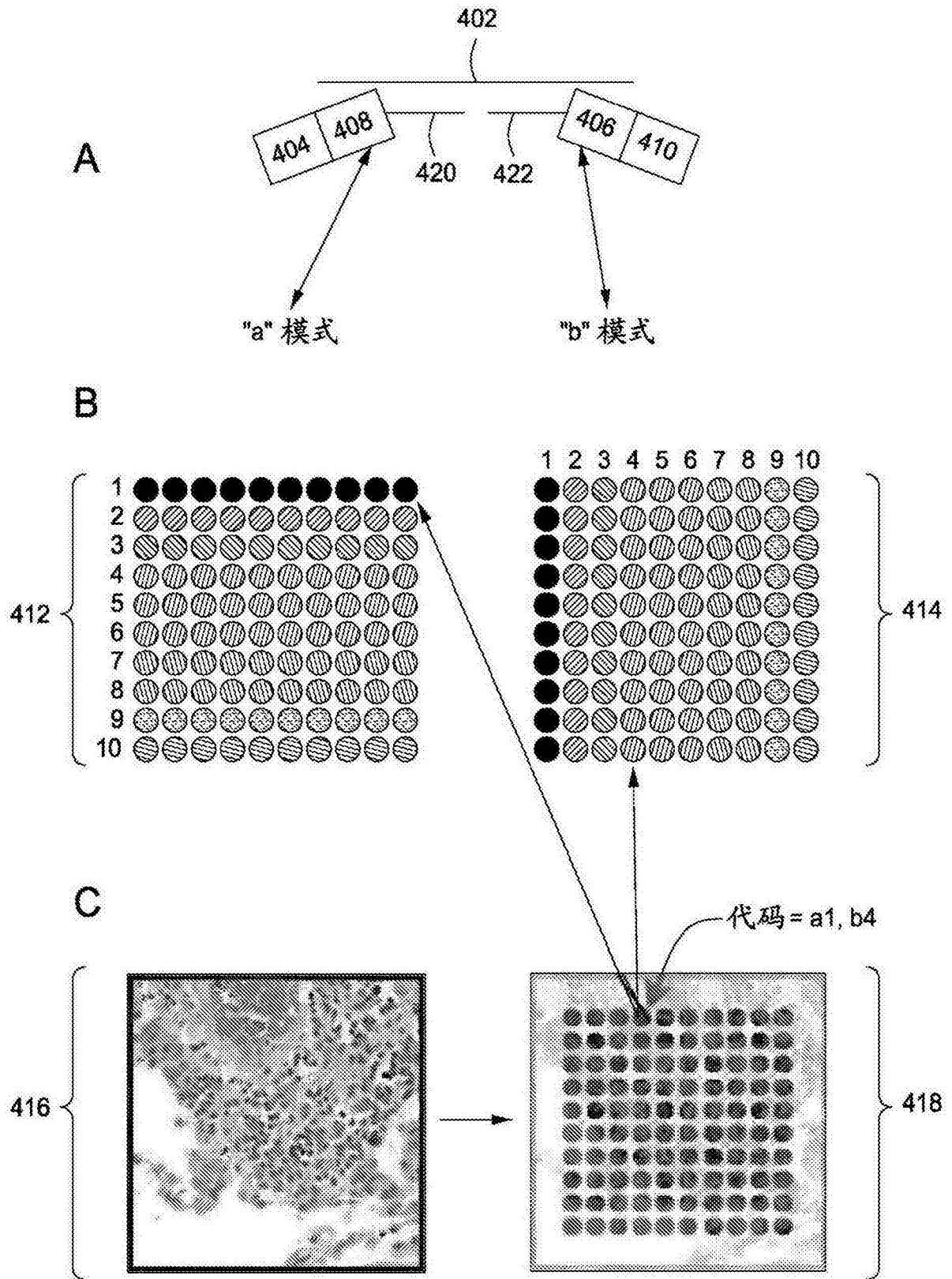


图 4

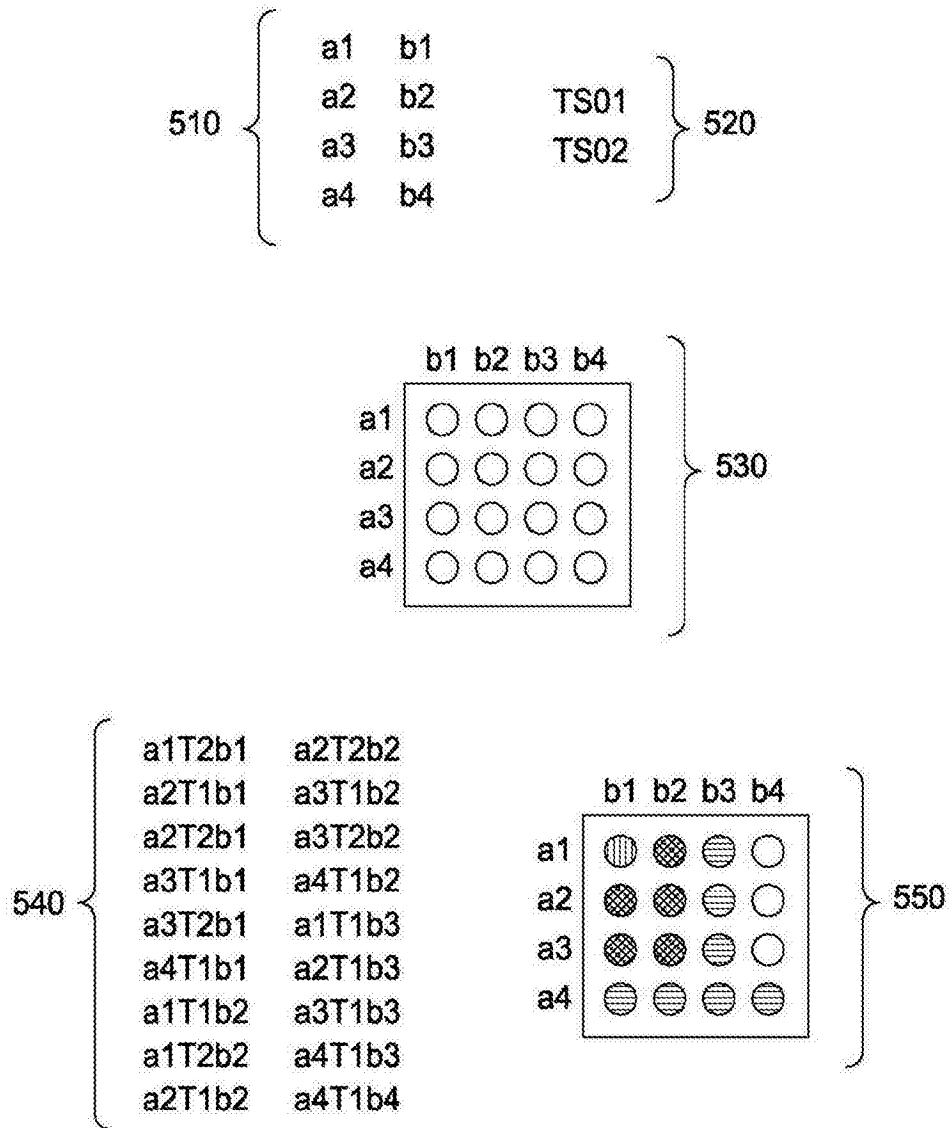


图 5

## **Abstract**

The present invention provides assays and assay systems for use in spatially encoded biological assays. The invention provides an assay system comprising an assay capable of high levels of multiplexing where reagents are provided to a biological sample in defined spatial patterns; instrumentation capable of controlled delivery of reagents according to the spatial patterns; and a decoding scheme providing a readout that is digital in nature.