(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 110637084 A
(43)申请公布日 2019.12.31

(21)申请号 201880032804.6

(22)申请日 2018.04.26

(30)优先权数据
62/490,492 2017.04.26 US

(85)PCT国际申请进入国家阶段日
2019.11.18

(86)PCT国际申请的申请数据
PCT/US2018/029641 2018.04.26

(87)PCT国际申请的公布数据
WO2018/200867 EN 2018.11.01

(71)申请人 10X基因组学有限公司
地址 美国加利福尼亚州

(72)发明人 约瑟芬·李 塞缪尔·马尔斯
杰弗里·麦克德莫特
弗朗西斯卡·梅斯基
卢斯·蒙特斯克拉劳斯
凯瑟琳·法伊弗
约瑟夫·弗朗西斯·舒加
杰西卡·米歇尔·特里
索伦戈·B·西拉尔多

(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 武晶晶

(51)Int.Cl.
*C12N 9/12*(2006.01)
*C12N 9/22*(2006.01)
*C12N 15/33*(2006.01)

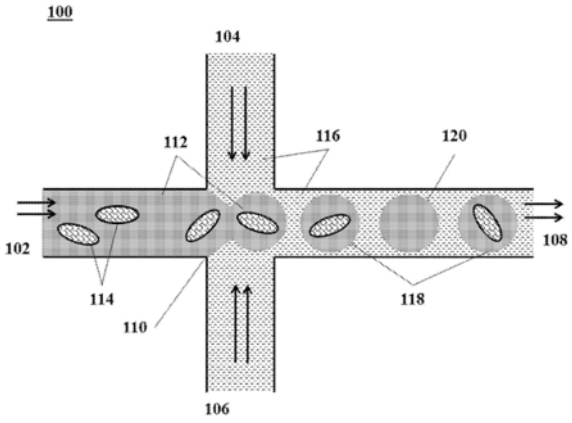
权利要求书3页 说明书71页 附图21页

(54)发明名称

MMLV逆转录酶变体

(57)摘要

本文公开了包含表现出若干所需性质诸如热稳定性、持续逆转录、非模板化碱基添加和模板转换能力的经工程改造逆转录酶的组合物、方法和试剂盒。相较于其他经工程改造逆转录酶，本文所述的经工程改造逆转录酶显示出乎意料的更高的对细胞溶解产物抑制的抗性、更大的捕集全长mRNA转录物的能力，并且在小反应体积的情况下显示出改进的结果。




CN 110637084 A

1.一种经工程改造逆转录酶,其包含与SEQ ID NO:3至少80%同一的氨基酸序列,其中所述氨基酸序列包含:

(i)相对于SEQ ID NO:3,从N末端进行的至少15个氨基酸的截短;和

(ii)一个或多个选自由以下组成的组的突变:相对于SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。

2.如权利要求1所述的经工程改造逆转录酶,其中所述一个或多个突变是相对于SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

3.如权利要求1所述的经工程改造逆转录酶,其中所述氨基酸序列包含多个选自由以下组成的组的突变:相对于SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。

4.如权利要求1所述的经工程改造逆转录酶,其中所述经工程改造逆转录酶包含:(i)三个或更多个选自由以下组成的组的突变:相对于SEQ ID NO:3的L139突变、D200突变、T330突变、P448突变、D449突变、D524突变和L603突变;和(ii)三个或更多个选自由以下组成的组的突变:相对于SEQ ID NO:3的E69突变、E302突变、T306突变、W313突变、L435突变和N454突变。

5.如权利要求4所述的经工程改造逆转录酶,其中所述经工程改造逆转录酶包含:(i)三个或更多个选自由以下组成的组的突变:相对于SEQ ID NO:3的L139P突变、D200N突变、T330P突变、P448A突变、D449G突变、D524N或D524A突变和L603W突变;和(ii)三个或更多个选自由以下组成的组的突变:相对于SEQ ID NO:3的E69K突变、E302R突变、T306K突变、W313F突变、L435G或L435K突变和N454K突变。

6.如权利要求1所述的经工程改造逆转录酶,其中所述经工程改造逆转录酶包含:相对于SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、D524突变、L603突变和E607突变。

7.如权利要求6所述的经工程改造逆转录酶,其中所述经工程改造逆转录酶包含:相对于SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

8.如权利要求1所述的经工程改造逆转录酶,其中所述截短包括相对于SEQ ID NO:3,从所述N末端进行的至少20个氨基酸的截短。

9.如权利要求1所述的经工程改造逆转录酶,其中所述截短包括相对于SEQ ID NO:3,从所述N末端进行的23个氨基酸的截短。

10.如权利要求1所述的经工程改造逆转录酶,其中所述经工程改造逆转录酶在所述N末端或在所述氨基酸序列的C末端进一步包含亲和标签。

11.如权利要求10所述的经工程改造逆转录酶,其中所述亲和标签是至少5个组氨酸氨基酸。

12.如权利要求10所述的经工程改造逆转录酶,其中所述经工程改造逆转录酶进一步包含蛋白酶裂解序列,其中由蛋白酶对所述蛋白酶裂解序列的裂解导致所述亲和标签从所述经工程改造逆转录酶的裂解。

13.如权利要求12所述的经工程改造逆转录酶,其中所述蛋白酶裂解序列是凝血酶裂解序列。

14.如权利要求13所述的经工程改造逆转录酶,其中所述氨基酸序列在所述N末端包含MRSSHHHHHHSSGLVPRGS(SEQ ID NO:7)氨基酸序列。

15.如权利要求1所述的经工程改造逆转录酶,其中所述经工程改造逆转录酶包含根据SEQ ID NO:6的氨基酸序列。

16.如权利要求12所述的经工程改造逆转录酶,其中所述经工程改造逆转录酶包含根据SEQ ID NO:5的氨基酸序列。

17.如权利要求1所述的经工程改造逆转录酶,其中相较于由SEQ ID NO:3组成的逆转录酶,所述经工程改造逆转录酶具有改进的捕集全长转录物的能力。

18.如权利要求1所述的经工程改造逆转录酶,其中相较于由SEQ ID NO:3组成的逆转录酶,所述经工程改造逆转录酶对细胞溶解产物具有更高的抗性。

19.如权利要求1所述的经工程改造逆转录酶,其中相较于由SEQ ID NO:3组成的逆转录酶,所述经工程改造逆转录酶在小于1纳升的反应体积的情况下具有更高的活性。

20.如权利要求1所述的经工程改造逆转录酶,其中相较于由SEQ ID NO:3组成的逆转录酶,所述经工程改造逆转录酶具有增加的热稳定性和逆转录持续性。

21.如权利要求17所述的经工程改造逆转录酶,其中所述经工程改造逆转录酶包含末端转移酶活性和模板转换能力。

22.一种用于核酸样品处理的方法,其包括:

(a)以某一反应体积提供模板核糖核酸(RNA)分子,和

(b)使用经工程改造逆转录酶来使所述RNA分子逆转录成互补性DNA分子,其中所述经工程改造逆转录酶包含与SEQ ID NO:3至少80%同一的氨基酸序列,其中所述氨基酸序列包含:(i)相对于SEQ ID NO:3,从N末端进行的至少15个氨基酸的截短;和(ii)一个或多个选自由以下组成的组的突变:相对于SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。

23.如权利要求22所述的方法,其中所述反应体积小于1纳升。

24.如权利要求23所述的方法,其中所述反应体积小于500皮升。

25.如权利要求22所述的方法,其中所述反应体积是乳液中的小滴。

26.如权利要求22所述的方法,其中所述反应体积是孔。

27.如权利要求22所述的方法,其中所述反应体积进一步包含多个包含条形码序列的核酸条形码分子。

28.如权利要求27所述的方法,其中所述RNA分子是信使RNA(mRNA)分子,其中所述多个核酸条形码分子进一步包含寡(dT)序列,并且其中所述经工程改造逆转录酶使用所述寡(dT)序列来使所述mRNA分子逆转录成所述互补性DNA分子,其中所述互补性DNA分子包含所述条形码序列。

29.如权利要求27所述的方法,其中所述RNA分子是信使RNA(mRNA)分子,其中所述反应体积进一步包含含有寡(dT)序列的核酸分子,其中所述多个核酸条形码分子进一步包含模板转换序列,其中所述经工程改造逆转录酶使用包含所述寡(dT)序列的所述核酸分子来使所述mRNA分子逆转录,并且其中所述经工程改造逆转录酶执行模板转换反应,由此产生所述互补性DNA分子,其中所述互补性DNA分子包含所述条形码序列。

30.如权利要求27所述的方法,其中所述多个核酸条形码分子连接于载体。

31.如权利要求30所述的方法,其中所述核酸条形码分子以可释放方式连接于所述载体。

32.如权利要求30所述的方法,其中所述载体是珠粒。

33.如权利要求32所述的方法,其中所述珠粒是凝胶珠粒。

34.如权利要求32所述的方法,其中所述核酸条形码分子共价连接于所述珠粒。

35.如权利要求32所述的方法,其中所述核酸条形码分子以可释放方式连接于所述珠粒。

36.如权利要求35所述的方法,其中所述核酸条形码分子在施加刺激后被释放。

37.如权利要求36所述的方法,其中所述刺激是化学刺激。

38.如权利要求37所述的方法,其中所述化学刺激是还原剂。

39.如权利要求32所述的方法,其中所述凝胶珠粒是可降解凝胶珠粒。

40.如权利要求39所述的方法,其中所述可降解凝胶珠粒包含可化学裂解的交联。

41.如权利要求40所述的方法,其中所述可化学裂解的交联包括二硫交联。

42.如权利要求22所述的方法,其中所述反应体积包含含有所述RNA分子的细胞。

43.如权利要求42所述的方法,其进一步包括使所述RNA分子从所述细胞释放。

44.一种用于执行逆转录反应的试剂盒,其包括:

(a)经工程改造逆转录酶,其包含(i)相对于SEQ ID NO:3,从N末端进行的至少15个氨基酸的截短;和(ii)一个或多个选自由以下组成的组的突变:相对于SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变;和

(b)关于使用所述经工程改造逆转录酶执行逆转录反应的说明书。

45.如权利要求44所述的试剂盒,其中所述试剂盒进一步包括用于执行所述逆转录反应的反应缓冲液。

46.如权利要求45所述的试剂盒,其中所述试剂盒进一步包括dNTP。

47.如权利要求46所述的试剂盒,其中所述经工程改造逆转录酶、所述缓冲液和所述dNTP被一起提供在主混合物溶液中。

48.如权利要求47所述的试剂盒,其中所述主混合物以所述说明书中指示的用于所述逆转录反应中的工作浓度至少两倍的浓度存在。

49.如权利要求44所述的试剂盒,其中所述试剂盒进一步包括用于引发所述逆转录反应的引物。

50.如权利要求49所述的试剂盒,其中所述引物是多dT引物、随机N聚体引物或靶标特异性引物。

# MMLV逆转录酶变体

[0001] 交叉引用

[0002] 本申请要求2017年4月26日提交的美国临时申请号62/490,492的权益,所述美国临时申请以引用的方式整体并入本文。

## 背景技术

[0003] 在分析和表征生物学和生物化学物质和系统方面的重大进步已导致在了解生命、健康、疾病和治疗的机理方面的空前进步。在这些进步之中,靶向和表征组成生物系统的基因组的技术已产生一些最具开创性的结果,包括在使用和开发遗传扩增技术以及核酸测序技术方面的进步。

[0004] 核酸测序可用于在包括诊断学、预后学、生物工艺学和法医生物学的广泛多种生物医学情形下获得信息。测序可涉及基本方法,包括马克萨姆-吉尔伯特(Maxam-Gilbert)测序和链终止方法;或重新测序方法,包括鸟枪测序和桥式PCR;或下一代方法,包括聚合酶克隆测序、454焦磷酸测序、Illumina测序、SOLiD测序、离子激流半导体测序、HeliScope单分子测序、SMRT®测序以及其他方法。

[0005] 尽管在生物表征方面取得这些进步,但许多挑战仍然未由当前提供的解决方案解决,或由当前提供的解决方案相对不良解决。本公开提供用以解决现有技术的许多短处的新型解决方案和方法。

## 发明内容

[0006] 在一些实施方案中,本文公开经工程改造逆转录酶,其包含与SEQ ID NO:3至少80%同一的氨基酸序列,其中所述氨基酸序列包含:(i)相对于SEQ ID NO:3,从N末端进行的至少15个氨基酸的截短;和(ii)一个或多个选自由以下组成的组的突变:相对于SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。在一些情况下,所述一个或多个突变是相对于SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。在一些情况下,所述氨基酸序列包含多个选自由以下组成的组的突变:相对于SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。在一些情况下,所述经工程改造逆转录酶包含:(i)三个或更多个选自由以下组成的组的突变:相对于SEQ ID NO:3的L139突变、D200突变、T330突变、P448突变、D449突变、D524突变和L603突变;和(ii)三个或更多个选自由以下组成的组的突变:相对于SEQ ID NO:3的E69突变、E302突变、T306突变、W313突变、L435突变和N454突变。在一些情况下,所述经工程改造逆转录酶包含:(i)三个或更多个选自由以下组成的组的突变:相对于SEQ ID NO:3的L139P突变、D200N突变、T330P突变、P448A突变、D449G突变、D524N或D524A突变和L603W突变;和(ii)三个或更多个选自由以下组成的组的突变:相对于SEQ ID NO:3的E69K

突变、E302R突变、T306K突变、W313F突变、L435G或L435K突变和N454K突变。在一些情况下，所述经工程改造逆转录酶包含：相对于SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、D524突变、L603突变和E607突变。在一些情况下，所述经工程改造逆转录酶包含：相对于SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。在一些情况下，所述截短包括相对于SEQ ID NO:3，从所述N末端进行的至少20个氨基酸的截短。在一些情况下，所述截短包括相对于SEQ ID NO:3，从所述N末端进行的23个氨基酸的截短。在一些情况下，所述经工程改造逆转录酶在所述N末端或在所述氨基酸序列的C末端进一步包含亲和标签。在一些情况下，所述亲和标签是至少5个组氨酸氨基酸。在一些情况下，所述经工程改造逆转录酶进一步包含蛋白酶裂解序列，其中由蛋白酶对所述蛋白酶裂解序列的裂解导致所述亲和标签从所述经工程改造逆转录酶的裂解。在一些情况下，所述蛋白酶裂解序列是凝血酶(thrombin)裂解序列。在一些情况下，所述氨基酸序列在所述N末端包含MRSSHHHHHHSSGLVPRGS(SEQ ID NO:7)氨基酸序列。在一些情况下，所述经工程改造逆转录酶包含根据SEQ ID NO:6的氨基酸序列。在一些情况下，所述经工程改造逆转录酶包含根据SEQ ID NO:5的氨基酸序列。在一些情况下，相较于由SEQ ID NO:3组成的逆转录酶，所述经工程改造逆转录酶具有改进的捕集全长转录物的能力。在一些情况下，相较于由SEQ ID NO:3组成的逆转录酶，所述经工程改造逆转录酶对细胞溶解产物具有更高的抗性。在一些情况下，相较于由SEQ ID NO:3组成的逆转录酶，所述经工程改造逆转录酶在小于1纳升的反应体积的情况下具有更高活性。在一些情况下，相较于由SEQ ID NO:3组成的逆转录酶，所述经工程改造逆转录酶具有增加的热稳定性和逆转录持续性。在一些情况下，所述经工程改造逆转录酶包含末端转移酶活性和模板转换能力。

[0007] 在一些实施方案中，本文公开用于核酸样品处理的方法，其包括：(a)以某一反应体积提供模板核糖核酸(RNA)分子，和(b)使用经工程改造逆转录酶来使所述RNA分子逆转录成互补性DNA分子，其中所述经工程改造逆转录酶包含与SEQ ID NO:3至少80％同一的氨基酸序列，其中所述氨基酸序列包含：(i)相对于SEQ ID NO:3，从N末端进行的至少15个氨基酸的截短；和(ii)一个或多个选自由以下组成的组的突变：相对于SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。在一些情况下，所述反应体积小于1纳升。在一些情况下，所述反应体积小于500皮升。在一些情况下，所述反应体积是乳液中的小滴。在一些情况下，所述反应体积是孔。在一些情况下，所述反应体积进一步包含多个包含条形码序列的核酸条形码分子。在一些情况下，所述RNA分子是信使RNA(mRNA)分子，其中所述多个核酸条形码分子进一步包含寡(dT)序列，并且其中所述经工程改造逆转录酶使用所述寡(dT)序列来使所述mRNA分子逆转录成所述互补性DNA分子，其中所述互补性DNA分子包含所述条形码序列。在一些情况下，所述RNA分子是信使RNA(mRNA)分子，其中所述反应体积进一步包含含有寡(dT)序列的核酸分子，其中所述多个核酸条形码分子进一步包含模板转换序列，其中所述经工程改造逆转录酶使用包含所述寡(dT)序列的所述核酸分子来使所述mRNA分子逆转录，并且其中所述经工程改造逆转录酶执行模板转换反应，由此产生所述互补性DNA分子，其中所述互补性DNA分子包含所述条形码序列。在一些情况下，所述多个核

酸条形码分子连接于载体。在一些情况下,所述核酸条形码分子以可释放方式连接于所述载体。在一些情况下,所述载体是珠粒。在一些情况下,所述珠粒是凝胶珠粒。在一些情况下,所述核酸条形码分子共价连接于所述珠粒。在一些情况下,所述核酸条形码分子以可释放方式连接于所述珠粒。在一些情况下,所述核酸条形码分子在施加刺激后被释放。在一些情况下,所述刺激是化学刺激。在一些情况下,所述化学刺激是还原剂。在一些情况下,所述凝胶珠粒是可降解凝胶珠粒。在一些情况下,所述可降解凝胶珠粒包含可化学裂解的交联。在一些情况下,所述可化学裂解的交联包括二硫交联。在一些情况下,所述反应体积包含含有所述RNA分子的细胞。在一些情况下,方法进一步包括使所述RNA分子从所述细胞释放。

[0008] 在一些实施方案中,本文公开了用于执行逆转录反应的试剂盒,其包括:(a)经工程改造逆转录酶,其包含(i)相对于SEQ ID NO:3,从N末端进行的至少15个氨基酸的截短;和(ii)一个或多个选自由以下组成的组的突变:相对于SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变;和(b)关于使用所述经工程改造逆转录酶来进行逆转录反应的说明书。在一些情况下,所述试剂盒进一步包括用于进行所述逆转录反应的反应缓冲液。在一些情况下,所述试剂盒进一步包括dNTP。在一些情况下,所述经工程改造逆转录酶、所述缓冲液和所述dNTP被一起提供在主混合物溶液中。在一些情况下,所述主混合物以是在所述说明书中指示的用于所述逆转录反应中的工作浓度至少两倍的浓度存在。在一些情况下,所述试剂盒进一步包括用于引发所述逆转录反应的引物。在一些情况下,所述引物是多dT引物、随机N聚体引物或靶标特异性引物。

[0009] 以引用的方式并入

[0010] 本说明书中提及的所有出版物、专利和专利申请都以引用的方式并入本文,所述引用的程度就好像已特定地和单独地指出将每个单独的出版物、专利或专利申请以引用的方式并入一样。在以引用的方式并入本文的出版物和专利或专利申请与说明书中含有的公开内容相抵触的情况下,说明书意图替代和/或优先于任何所述抵触材料。

## 附图说明

[0011] 本发明的新型特征特定地阐述于随附权利要求中。对本发明的特征和优势的更充分了解将通过参照阐述其中利用本发明的原理的说明性实施方案的以下具体实施方式以及附图(在本文中也表示为“图(Figure/FIG.)”)来获得,在所述附图中:

[0012] 图1显示用于分配单个细胞或小型细胞群组的微流体通道结构的一实例。

[0013] 图2显示用于共分配细胞和包含额外试剂的珠粒或微囊的微流体通道结构的一实例。

[0014] 图3示意性说明用于细胞的核酸的扩增和条形码化的实例方法。

[0015] 图4提供使用细胞的核酸的条形码化将序列数据归属于单个细胞或细胞群组以用于它们的表征中的示意性说明。

[0016] 图5提供示意性说明细胞与经标记细胞结合配体缔合。

[0017] 图6提供用于使用本文所述的方法进行RNA分析的一实例工作流程的示意性说明。

[0018] 图7提供用于使用本文所述的方法分析核糖核酸(RNA)的一实例条形码化寡核苷酸结构的示意性说明。

[0019] 图8提供单个细胞连同单个携带条形码的珠粒一起共分配的图像。

[0020] 图9A-图9E提供用于分析RNA的实例条形码化寡核苷酸结构和用于进行RNA分析的实例操作的示意性说明。

[0021] 图10提供用于RNA的实例分析中的实例条形码化寡核苷酸结构以及使用序列进行体外转录的示意性说明。

[0022] 图11提供用于分析RNA的一实例条形码化寡核苷酸结构和用于进行RNA分析的实例操作的示意性说明。

[0023] 图12A-图12B提供用于分析RNA的实例条形码化寡核苷酸结构的示意性说明。

[0024] 图13A-图13C提供由分区中的模板转换逆转录和PCR获得的实例产量的说明。

[0025] 图14A-图14B提供在各种细胞数目的情况下由分区中的逆转录和cDNA扩增获得的实例产量的说明。

[0026] 图15提供在各种输入细胞浓度下由cDNA合成和实时定量PCR获得的实例产量的说明以及在固定细胞输入浓度下不同引物浓度对产量的影响。

[0027] 图16提供由体外转录获得的实例产量的说明。

[0028] 图17显示被编程或另外被配置来执行本文提供的方法的一实例计算机控制系统。

[0029] 图18提供相较于可商购获得的对应物，在含有经工程改造RT酶的皮升大小小滴的情况下的实例测序饱和度结果的说明。

[0030] 图19提供相较于可商购获得的对应物，从在含有经工程改造RT酶的皮升大小小滴中制备的cDNA文库获得的每个细胞的实例中值基因(人基因组hg19)的说明。

[0031] 图20提供相较于可商购获得的对应物，从在含有经工程改造RT酶的皮升大小小滴中制备的cDNA文库获得的每个细胞的实例中值基因(小鼠基因组mm10)的说明。

[0032] 图21显示相较于可商购获得的对应物，根据由含有经工程改造RT酶的小滴作出的TCR转录剖析进行的一示例性有效配对比较。

[0033] 图22显示相较于可商购获得的对应物，根据由含有经工程改造RT酶的小滴作出的TCR转录剖析的各种示例性结果。

## 具体实施方式

[0034] 尽管本文已显示和描述本发明的各种实施方案，但将对本领域技术人员明显的是所述实施方案仅通过举例方式来提供。在不脱离本发明下，众多改变、变化和替代可为本领域技术人员所想到。应了解可采用本文所述的本发明的实施方案的各种替代方案。

[0035] 当将数值描述为范围时，应了解所述公开包括对在所述范围内的所有可能子范围以及落在所述范围内的特定数值的公开，不管特定数值或特定子范围是否被明确陈述。

[0036] 如本文所用的术语“条形码”通常是指传达或能够传达关于分析物的信息的标记或标识符。条形码可为分析物的一部分。条形码可与分析物无关。除分析物的内源性特征(例如分析物的大小或末端序列)之外，条形码也可为连接于分析物(例如核酸分子)的标签或标签的组合。条形码可为独特的。条形码可具有多种不同形式。举例来说，条形码可包括：多核苷酸条形码；随机核酸和/或氨基酸序列；和合成核酸和/或氨基酸序列。条形码可以可逆或不可逆方式连接于分析物。条形码可在对样品测序之前、期间和/或之后添加至例如脱氧核糖核酸(DNA)或核糖核酸(RNA)样品的片段中。条形码可允许对单个测序读段的鉴定

和/或定量。

[0037] 术语“衔接头”、“衔接物”和“标签”可以同义方式使用。衔接头或标签可通过包括连接、杂交或其他方法的任何方法来偶联于待“标签化”多核苷酸序列。

[0038] 如本文所用的术语“测序”通常是指用于测定一个或多个多核苷酸中的核苷酸碱基序列的方法和技术。多核苷酸可为例如核酸分子诸如脱氧核糖核酸(DNA)或核糖核酸(RNA),包括其变体或衍生物(例如单链DNA)。测序可通过当前可用的各种系统来进行,诸如不限于Illumina®、Pacific Biosciences(PacBio®)、Oxford Nanopore®或Life Technologies(IonTorrent®)的测序系统。或者或此外,测序可使用核酸扩增、聚合酶链式反应(PCR)(例如数字PCR、定量PCR或实时PCR)或等温扩增来进行。所述系统可提供对应于受试者(例如人)的遗传信息的多个原始遗传数据,如由系统从由所述受试者提供的样品所产生。在一些实例中,所述系统提供测序读段(在本文中也表示为“读段”)。读段可包括对应于核酸分子的已加以测序的序列的一串核酸碱基。在一些情况下,本文提供的系统和方法可与蛋白质组学信息一起使用。

[0039] 如本文所用的术语“珠粒”通常是指粒子。珠粒可为固体或半固体粒子。珠粒可为凝胶珠粒。凝胶珠粒可包括聚合物基质(例如通过聚合或交联形成的基质)。聚合物基质可包括一种或多种聚合物(例如具有不同官能团或重复单元的聚合物)。聚合物基质中的聚合物可为无规排列的,诸如在无规共聚物的情况下,和/或具有有序结构,诸如在嵌段共聚物的情况下。交联可通过共价性、离子性、或诱导性相互作用或物理缠结来达成。珠粒可为大分子。珠粒可由结合在一起的核酸分子形成。珠粒可通过分子(例如大分子)诸如单体或聚合物的共价或非共价组装来形成。所述聚合物或单体可为天然的或合成的。所述聚合物或单体可为或包括例如核酸分子(例如DNA或RNA)。珠粒可由聚合物质形成。珠粒可为磁性的或非磁性的。珠粒可为刚性的。珠粒可为柔性的和/或可压缩的。珠粒可为可破坏的或可溶解的。珠粒可为用包含一种或多种聚合物的涂层覆盖的固体粒子(例如基于金属的粒子,包括但不限于氧化铁、金或银)。所述涂层可为可破坏的或可溶解的。

[0040] 如本文所用的术语“样品”通常是指受试者的生物样品。生物样品可包含许多大分子例如细胞大分子。样品可为细胞样品。样品可为细胞系或细胞培养物样品。样品可包括一个或多个细胞。样品可包括一个或多个微生物。生物样品可为核酸样品或蛋白质样品。生物样品也可为碳水化合物样品或脂质样品。生物样品可源于另一样品。样品可为组织样品,诸如活检体、核心活检体、针抽吸物或细针抽吸物。样品可为流体样品,诸如血液样品、尿样品或唾液样品。样品可为皮肤样品。样品可为面颊拭擦物。样品可为血浆或血清样品。样品可为无细胞(cell-free/cell free)样品。无细胞样品可包括细胞外多核苷酸。细胞外多核苷酸可从可选自由以下组成的组的身体样品分离:血液、血浆、血清、尿、唾液、粘膜排泄物、痰、粪便和泪。

[0041] 如本文所用的术语“受试者”通常是指动物诸如哺乳动物(例如人)或鸟类(例如鸟),或其他生物体诸如植物。举例来说,受试者可为脊椎动物、哺乳动物、啮齿动物(例如小鼠)、灵长类动物、猿猴或人。动物可包括但不限于农场动物、体育动物和宠物。受试者可为健康或无症状个体、患有或被怀疑患有疾病(例如癌症)或具有所述疾病的倾向的个体、和/或需要疗法或被怀疑需要疗法的个体。受试者可为患者。受试者可为微生物(microorganism/microbe)(例如细菌、真菌、古细菌、病毒)。

[0042] 如本文所用的术语“分子标签”通常是指能够结合大分子成分的分子。分子标签可以高亲和力结合大分子成分。分子标签可以高特异性结合大分子成分。分子标签可包含核苷酸序列。分子标签可包含核酸序列。核酸序列可为分子标签的至少一部分或全部。分子标签可为核酸分子或可为核酸分子的一部分。分子标签可为寡核苷酸或多肽。分子标签可包含DNA适体。分子标签可为或包含引物。分子标签可为或包含蛋白质。分子标签可包含多肽。分子标签可为条形码。

[0043] 如本文所用的术语“分区”通常是指可适于含有一种或多种物质或进行一个或多个反应的空间或体积。分区可为实际区室,诸如小滴或孔。分区可使空间或体积与另一空间或体积隔离。小滴可为不与第一相混溶的第二相(例如油)中的所述第一相(例如水相)。小滴可为不与第一相进行相分离的第二相中的所述第一相,诸如像水相中的囊体或脂质体。分区可包含一个或多个其他(内部)分区。在一些情况下,分区可为可由索引(例如索引化文库)跨越多个和/或远程实际区室加以界定和标识的虚拟区室。举例来说,实际区室可包含多个虚拟区室。

[0044] I.单细胞分析

[0045] 先进核酸测序技术已在对生物物质测序方面产生不朽结果,包括提供关于单个生物体和相对纯净生物样品的大量序列信息。然而,这些系统尚未在能够鉴定和表征生物样品中可代表样品的总体组成的较小比例少数部分,但其个性化序列信息可被证明甚至更有价值的细胞子群体方面被证明有效。

[0046] 大多数核酸测序技术从源于组织或其他样品的细胞的集合获得它们进行测序的核酸。细胞可被整体地处理以提取代表平均细胞群体的遗传物质,所述遗传物质可接着被处理成被配置来用于给定测序技术的准备用于测序的DNA文库。如将了解,尽管经常以DNA或核酸加以讨论,但源于细胞的核酸可包括DNA,或可被处理来产生用于例如使用多种RNA-seq方法中的任一者进行测序的cDNA的RNA,包括例如mRNA、总RNA等。由这个处理得出,在不存在细胞特异性标志下,将遗传物质归为由样品中的细胞子组还是由所有细胞贡献在这种全体方法的情况下是实际上不可能的。

[0047] 除不能将特征归于细胞群体的特定子组之外,所述全体样品制备方法也从开始就倾向于主要鉴定和表征细胞样品中的大多数成分,并且未被设计来能够挑选出少数成分,例如由样品中的一个细胞、少许细胞或总体细胞中的小百分比细胞贡献的遗传物质。同样,当分析例如mRNA的表达水平时,全体方法将倾向于呈现来自就表达水平而言是非同质的细胞群体的潜在粗略不准确数据。在一些情况下,当表达在所分析群体中的小比例少数细胞的情况下是较高的,而在群体的大多数细胞的情况下是不存在的时,全体方法将指示整个群体的低水平表达。

[0048] 这个原始大多数偏倚通过用于从这些样品建立测序文库的处理操作而得以进一步放大,乃至占绝对优势。特别地,大多数下一代测序技术依赖于对核酸片段的几何扩增,诸如聚合酶链式反应,以产生足以用于测序文库的DNA。然而,所述几何扩增偏向对样品中的大多数成分的扩增,并且可不保持所述少数组分与大多数组分的起始比率。举例来说,如果样品包括来自样品中特定细胞类型例如宿主组织细胞的95%DNA和来自另一细胞类型例如癌细胞的5%DNA,那么基于PCR的扩增可优先扩增大多数DNA而非少数DNA,两者均随相比而言的指数扩增而变化(较高浓度的重复倍增快速超过较小部分的重复倍增)以及随对扩

增试剂和资源的隔离而变化(当较大部分被扩增时,它优先利用引物和其他扩增试剂)。

[0049] 尽管这些困难中的一些可通过利用不同测序系统诸如不需要扩增的单分子系统来解决,但单分子系统以及其他下一代测序系统的全体测序方法也可关于需要足够大量输入物具有要求。

[0050] II.对细胞的划分和表征

[0051] 然而,本文公开用于表征来自小型细胞群体的核酸,以及在一些情况下,用于表征尤其是在较大细胞群体的情形下来自单个细胞的核酸的方法和系统。方法和系统提供能够以其他下一代系统的高通量提供非扩增单分子方法的归属优势的优势,伴有能够对可源于单个细胞或小型细胞集合的极低量的输入核酸进行处理和测序的额外优势。

[0052] 特别地,本文所述的方法将对单个细胞或小型细胞群体,包括例如来自单个细胞或小型细胞群组的核酸,的分析进行划分,接着使那个分析向回归属于核酸所源于的单个细胞或小型细胞群组。无论细胞群体代表50/50细胞类型混合物、90/10细胞类型混合物、或实际上任何比率的细胞类型,以及不同细胞类型的完全异质混合物,或介于这些之间的任何混合物,这都可实现。不同细胞类型可包括来自个体的不同组织类型,来自不同个体,来自不同属、物种、株系、变体或任何或全部前述各物的任何组合的细胞或生物有机体。举例来说,不同细胞类型可包括来自个体的正常和肿瘤组织,来自环境、法医、微生物群系或其他样品的多个不同细菌物种、株系和/或变体,或多种其他细胞类型混合物中的任一者。

[0053] 在一个方面,本文所述的方法和系统提供将来自含有细胞的样品物质的单个细胞的核酸内含物划分、放置或分配至离散区室或分区(在本文中可互换称为分区)中,其中各分区维持它的自身内含物与其他分区的内含物的分隔。分区可为乳液中的小滴。分区可包含一个或多个其他分区。

[0054] 分区可包括一个或多个细胞。分区可包括一种或多种类型的细胞。分区可包含一个或多个凝胶珠粒。分区可包含一个或多个细胞珠粒。分区可包括单一凝胶珠粒、单一细胞珠粒、或单一细胞珠粒与单一凝胶珠粒两者。分区可包括一种或多种试剂。或者,分区可未被占据。举例来说,分区可不包含珠粒。细胞珠粒可为诸如通过含有细胞的小滴和能够加以聚合或凝胶化的前体的聚合来包裹在凝胶或聚合物基质的内部的细胞。独特标识符诸如条形码可诸如通过微囊(例如珠粒)来在小滴产生之前、之后或与小滴产生并行注射至小滴中,如在本文中其他地方所述。微流体通道网络(例如在芯片上)可用于产生如本文所述的分区。在分配单个细胞的核酸内含物时,也可采用替代性机理,包括水性细胞混合物穿过其来被挤压至非水性流体中的多孔膜。

[0055] 如本文所用,在一些方面,分区是指容器或器皿(诸如孔、微孔、管、纳米阵列(例如BioTrove纳米阵列)基底中的贯穿口、或其他容器)。然而,在许多一些方面,区室或分区包括可在液流内流动的分区。分区可包含例如具有包围内部流体中心或核心的外部屏障的微囊泡。在一些情况下,分区可包含能够在它的基质内夹带和/或保留物质的多孔基质。分区可为在第二相内的第一相的小滴,其中所述第一相和所述第二相是不混溶的。举例来说,分区可为在非水性连续相(例如油相)内的水性流体的小滴。在另一实例中,分区可为在水相内的非水性流体的小滴。在一些实例中,分区可以油包水乳液或水包油乳液形式提供。多种不同器皿描述于例如美国专利申请公布号2014/0155295中,所述美国专利申请公布出于所有目的以引用的方式整体并入本文。用于在非水性或油连续相中创建稳定小滴的乳液系统

描述于例如美国专利申请公布号2010/0105112中，所述美国专利申请公布出于所有目的以引用的方式整体并入本文。

[0056] 在乳液中的小滴的情况下，在一个非限制性实例中，将单个细胞配置至离散分区中可通过以下方式来实现：将水性流体中的流动粒子流引入流动非水性流体流中，以使在两个物流的接合处产生小滴。流体性质(例如流体流速、流体粘度等)、粒子性质(例如体积分数、粒度、粒子浓度等)、微流体构造(例如通道几何结构等)和其他参数可加以调整以控制对所得分区的占据(例如每个分区的细胞数目、每个分区的珠粒数目等)。举例来说，分区占据可通过在某一粒子浓度和/或流速下提供水性物流来控制。为产生单细胞分区，可选择不混溶流体的相对流速以使平均来说，分区可含有每个分区小于一个细胞以确保被占据的那些分区主要以单一方式被占据。在一些情况下，在多个分区之中的分区可含有至多一个生物粒子(例如珠粒、DNA、细胞或细胞物质)。在一些实施方案中，可选择或调整各种参数(例如流体性质、粒子性质、微流体构造等)以使大多数分区被占据，例如从而允许达成仅有小百分比的未占据分区。可控制流动和通道构造以便确保给定数目的经单一占据分区、小于某一水平的未占据分区和/或小于某一水平的经多重占据分区。

[0057] 在某些情况下，微流体通道网络特别适于产生如本文所述的分区。所述微流体装置的实例包括2014年4月4日提交的临时美国专利申请号61/977,804中详述的那些，所述临时美国专利申请的全部公开内容出于所有目的以引用的方式整体并入本文。在分配单个细胞时，也可采用替代性机理，包括水性细胞混合物穿过其来被挤压至非水性流体中的多孔膜。所述系统通常可从例如Nanomi,Inc.获得。

[0058] 图1显示用于分配单个细胞的微流体通道结构100的一实例。通道结构100可包括在通道接合部110处连通的通道区段102、104、106和108。在操作中，包括经混悬单个细胞114的第一水性流体112可沿通道区段102运送至接合部110中，而不与水性流体112混溶的第二流体116从通道区段104和106中的各者递送至接合部110中以创建第一水性流体112的离散小滴118、120，流入通道区段108中，并且从接合部110流走。通道区段108可流体联接于出口储集器，在所述储集器中，离散小滴可加以储存和/或收集。产生的离散小滴可包括单个细胞114(诸如小滴118)。产生的离散小滴可包括超过一个单个细胞114(未显示于图1中)。离散小滴可不含有细胞114(诸如小滴120)。各离散分区可维持它的自身内含物(例如单个细胞114)与其他分区的内含物的分隔。

[0059] 第二流体116可包含包括用于使所得小滴稳定，例如抑制所得小滴118、120的后续聚结的氟表面活性剂的油，诸如氟化油。特别适用的分配流体和氟表面活性剂的实例例如描述于美国专利申请公布号2010/0105112中，所述美国专利申请公布出于所有目的以引用的方式整体并入本文。

[0060] 如将了解，本文所述的通道区段可联接于多种不同流体来源或接收部件中的任一者，包括储集器、管道、歧管、或其他系统的流体部件。如将了解，微流体通道结构100可具有其他几何结构。举例来说，微流体通道结构可具有超过一个通道接合部。举例来说，微流体通道结构可具有2、3、4或5个在通道接合部处相遇的通道区段，各自携带粒子(例如细胞、细胞珠粒和/或凝胶珠粒)。流体可通过一个或多个流体流动单元来被引导沿一个或多个通道或储集器流动。流体流动单元可包括用以控制流体的流动的压缩机(例如提供正压)、泵(例如提供负压)、执行器等。流体也可或另外通过所施加的压差、离心力、电动泵送、真空、毛细

管或重力流动等来控制。

[0061] 产生的小滴可包括两个小滴子组：(1)含有一个或多个细胞114的经占据小滴118，和(2)不含有任何细胞114的未占据小滴120。经占据小滴118可包括经单一占据小滴(具有一个细胞)和经多重占据小滴(具有超过一个细胞)。如在本文中其他地方所述，在一些情况下，大多数经占据分区可包括每个经占据分区至多一个细胞，并且一些产生分区可未被(任何细胞)占据。然而，在一些情况下，一些经占据分区可包括超过一个细胞。在许多情况下，系统和方法用于确保实质性大多数经占据分区(含有一个或多个微囊的分区)包括每个经占据分区至多1个细胞。在一些情况下，可控制分配过程以使少于约25％的经占据分区含有超过一个细胞，并且在许多情况下，少于约20％的经占据分区具有超过一个细胞，而在一些情况下，少于约10％或甚至少于约5％的经占据分区包括每个分区超过一个细胞。

[0062] 在一些情况下，可合乎需要的是使对过量数目的空分区的创建最小化，诸如以降低成本和/或增加效率。尽管这个最小化可通过在分配接合部110处提供足够数目的细胞(例如114)来实现，诸如以确保至少一个细胞被囊封在分区中，但泊松(Poissonian)分布可预期地增加包括多个细胞的分区的数目。因此，当待获得经单一占据分区时，至多约95％、90％、85％、80％、75％、70％、65％、60％、55％、50％、45％、40％、35％、30％、25％、20％、15％、10％、5％或更少的产生分区可未被占据。因此，根据本文所述的方面，控制细胞或被引导至分配区中的其他流体中的一者或多者的流动以使在许多情况下，至多50％的产生分区未被占据，即包括小于1个细胞，至多25％的产生分区、至多10％的产生分区可未被占据。此外，在一些方面，控制这些流动以便呈现经单一占据分区的非泊松分布，同时提供较低水平的未占据分区。再次陈述的是，在一些方面，可实现以上指示范围的未占据分区，同时仍然提供任何上述单一占据率。举例来说，在许多情况下，使用本文所述的系统和方法会创建具有小于25％、小于20％、小于15％、小于10％，并且在许多情况下，小于5％的多重占据率，同时具有小于50％、小于40％、小于30％、小于20％、小于10％，并且在一些情况下，小于5％的未占据分区的所得分区。

[0063] 在一些情况下，可控制生物粒子(例如在通道区段102中)或被引导至分配接合部中的其他流体(例如在通道区段104、106中)中的一者或多者的流动以使在许多情况下，至多约50％的产生分区、至多约25％的产生分区、或至多约10％的产生分区未被占据。可控制这些流动以便呈现经单一占据分区的非泊松分布，同时提供较低水平的未占据分区。可实现以上指示范围的未占据分区，同时仍然提供任何上述单一占据率。举例来说，在许多情况下，使用本文所述的系统和方法可创建具有小于约25％、小于约20％、小于约15％、小于约10％，并且在许多情况下，小于约5％的多重占据率，同时具有小于约50％、小于约40％、小于约30％、小于约20％、小于约10％、小于约5％或更少的未占据分区的所得分区。

[0064] 如将了解，上述占据率也可适用于包括细胞与额外试剂两者的分区，所述额外试剂包括但不限于携带条形码化核酸分子(例如寡核苷酸)的微囊或珠粒(例如凝胶珠粒)(关于图2加以描述)。经占据分区(例如至少约10％、20％、30％、40％、50％、60％、70％、80％、90％、95％或99％的经占据分区)可包括包含条形码化核酸分子的微囊(例如珠粒)与细胞两者。特别地，可合乎需要的是提供至少50％的分区由至少一个细胞和至少一个珠粒占据，或至少75％的分区可被如此占据，或甚至至少80％或至少90％的分区可被如此占据。此外，在其中需要在分区内提供单一细胞和单一珠粒的那些情况下，至少50％的分区可被如此占

据，至少60％、至少70％、至少80％或甚至至少90％的分区可被如此占据。

[0065] 在另一方面，除基于小滴的分配之外或作为基于小滴的分配的替代方案，可将细胞囊封在包含其中夹带一个或多个单个细胞或小型细胞群组的外部壳体、层或多孔基质的微囊内。微囊可包括其他试剂。对细胞的囊封可通过多种方法来进行。所述方法可将含有细胞的水性流体与可能够在向聚合物前体施加特定刺激后形成凝胶或其他固体或半固体基质的聚合前体物质组合。所述刺激可包括例如热刺激(例如加热或冷却)、光刺激(例如通过光固化)、化学刺激(例如通过交联、前体的聚合引发(例如通过所添加引发剂))、机械刺激或其组合。

[0066] 对包含细胞的微囊的制备可通过多种方法来进行。举例来说，气刀小滴或气雾发生器可用于将前体流体的小滴分配至胶凝溶液中以形成包括单个细胞或小型细胞群组的微囊。同样，基于膜的囊封系统可用于产生如本文所述的包含囊封细胞的微囊。本公开的诸如显示于图1中的微流体系统可容易地用于囊封如本文所述的细胞。特别地，并且参照图1，使包含(i)单个细胞114和(ii)聚合物前体物质(未显示)的水性流体112流入通道接合部110中，在所述通道接合部中，它通过非水性流体116的流动来被分配成小滴118、120。在囊封方法的情况下，非水性流体116也可包括引发剂(未显示)以导致聚合物前体的聚合和/或交联以形成包括经夹带细胞的微囊。聚合物前体/引发剂对的实例包括美国专利申请公布号2014/0378345中所述的那些，所述美国专利申请公布出于所有目的以引用的方式整体并入本文。

[0067] 举例来说，在其中聚合物前体物质包含线性聚合物物质诸如线性聚丙烯酰胺、PEG或其他线性聚合物质的情况下，活化剂可包括交联剂或使所形成小滴内的交联剂活化的化学物质。同样，对于包含可聚合单体的聚合物前体，活化剂可包括聚合引发剂。举例来说，在某些情况下，当聚合物前体包含丙烯酰胺单体与N,N'-双-(丙烯酰基)胱胺(BAC)共聚单体的混合物时，可引发丙烯酰胺和BAC共聚成交联聚合物网络或水凝胶的试剂诸如四乙基甲二胺(TEMED)可提供在通道区段104和106中的第二流体流116内。

[0068] 在第二流体流116与第一流体流112在接合部110处接触后，在小滴的形成期间，TEMED可从第二流体116扩散至包含线性聚丙烯酰胺的水性流体112中，所述TEMED将使小滴118、120内的聚丙烯酰胺的交联活化，从而导致形成呈夹带细胞114的固体或半固体珠粒或粒子形式的胶凝(例如水凝胶)微囊。尽管以聚丙烯酰胺囊封加以描述，但其他'可活化'囊封组合物也可在本文所述的方法和组合物的情形下采用。举例来说，形成海藻酸盐小滴继之以暴露于二价金属离子(例如$Ca^{2+}$离子)可在使用所述方法的情况下用作囊封方法。同样，琼脂糖小滴也可通过基于温度的胶凝(例如在冷却后，等)来转变成囊体。

[0069] 在一些情况下，囊封细胞可选择性地从微囊释放，诸如通过时间推移或在施加特定刺激后，所述时间推移或特定刺激使微囊充分降解以允许细胞或它的其他内含物从微囊释放，诸如进入分区(例如小滴)中。举例来说，在上述聚丙烯酰胺聚合物的情况下，微囊的降解可通过引入适当还原剂诸如DTT等以裂解使聚合物基质交联的二硫键来实现。参见例如美国专利申请公布号2014/0378345，其出于所有目的以引用的方式整体并入本文。

[0070] 可使细胞经受足以使前体聚合或胶凝的其他条件。足以使前体聚合或胶凝的条件可包括暴露于加热、冷却、电磁辐射和/或光照。足以使前体聚合或胶凝的条件可包括足以使前体聚合或胶凝的任何条件。在聚合或胶凝之后，聚合物或凝胶可围绕细胞来形成。聚合

物或凝胶可为化学或生物化学试剂所扩散渗透。聚合物或凝胶可不可为细胞的大分子成分所扩散渗透。以这个方式，聚合物或凝胶可起允许细胞经受化学或生物化学操作的作用，同时在空间上将核酸局限于小滴的由聚合物或凝胶界定的区域。聚合物或凝胶可包括以下中的一者或多者：二硫交联聚丙烯酰胺、琼脂糖、海藻酸盐、聚乙烯醇、聚乙二醇(PEG)-二丙烯酸酯、PEG-丙烯酸酯、PEG-硫醇、PEG-叠氮化物、PEG-炔、其他丙烯酸酯、壳聚糖、透明质酸、胶原、纤维蛋白(fibrin)、明胶或弹性蛋白(elastin)。聚合物或凝胶可包括任何其他聚合物或凝胶。

[0071] 聚合物或凝胶可被官能化来结合所靶向分析物，诸如核酸、蛋白质、碳水化合物、脂质或其他分析物。聚合物或凝胶可通过被动机理来聚合或胶凝。聚合物或凝胶可在碱性条件下或在高温下稳定。聚合物或凝胶可具有与珠粒的机械性质类似的机械性质。举例来说，聚合物或凝胶可与珠粒具有类似尺寸。聚合物或凝胶可具有与珠粒的机械强度类似的机械强度(例如拉伸强度)。相比于油，聚合物或凝胶可具有较低密度。聚合物或凝胶可具有大致类似于缓冲液的密度的密度。聚合物或凝胶可具有可调谐孔尺寸。可选择孔尺寸以例如保留变性核酸。可选择孔尺寸以维持对外源性化学物质诸如氢氧化钠(NaOH)和/或内源性化学物质诸如抑制剂的扩散可渗透性。聚合物或凝胶可为生物可相容的。聚合物或凝胶可维持或增强细胞活力。聚合物或凝胶可为生物化学可相容的。聚合物或凝胶可以热方式、化学方式、酶促方式和/或光学方式加以聚合和/或解聚。

[0072] 聚合物可包括以二硫连键交联的聚(丙烯酰胺-共-丙烯酸)。对聚合物的制备可包括两步反应。在第一活化步骤中，可使聚(丙烯酰胺-共-丙烯酸)暴露于酰化剂以使羧酸转化成酯。举例来说，可使聚(丙烯酰胺-共-丙烯酸)暴露于氯化4-(4,6-二甲氧基-1,3,5-三嗪-2-基)-4-甲基吗啉鎓(DMTMM)。可使聚丙烯酰胺-共-丙烯酸暴露于4-(4,6-二甲氧基-1,3,5-三嗪-2-基)-4-甲基吗啉鎓的其他盐。在第二交联步骤中，可使在第一步骤中形成的酯暴露于二硫交联剂。举例来说，可使酯暴露于胱胺(2,2'-二硫基双(乙胺))。在两步之后，细胞可由通过二硫桥来连接在一起的聚丙烯酰胺束围绕。以这个方式，细胞可被包裹在凝胶或基质(例如聚合物基质)的内部或包含凝胶或基质(例如聚合物基质)以形成“细胞珠粒”。细胞珠粒可含有细胞或单个细胞的核酸(例如RNA、DNA)。细胞珠粒可包括单一细胞或多个细胞，或单一细胞或多个细胞的衍生物。举例来说，在溶解和洗涤细胞之后，可将来自细胞溶解产物的抑制性组分洗除，并且核酸可作为细胞珠粒加以结合。本文公开的系统和方法可适用于含有单个细胞的细胞珠粒(和/或小滴或其他分区)与含有单个细胞的核酸的细胞珠粒(和/或小滴或其他分区)两者。

[0073] 囊封细胞或细胞群体可提供比基于小滴的经分配细胞更可储存以及更便携的某些潜在优势。此外，在一些情况下，可合乎需要的是使细胞在分析之前持续所选时期进行孵育，诸如以表征所述细胞在存在或不存在不同刺激下随时间的变化。在所述情况下，相比于在乳液小滴的情况下分配，囊封可允许达成较长久孵育，但在一些情况下，也可持续不同时期孵育小滴经分配细胞，例如至少10秒、至少30秒、至少1分钟、至少5分钟、至少10分钟、至少30分钟、至少1小时、至少2小时、至少5小时或至少10小时或更久。对细胞的囊封可构成对向其中共分配其他试剂的细胞的分配。或者或此外，囊封细胞可易于放置至如上所述的其他分区(例如小滴)中。

[0074] 珠粒

[0075] 分区可包含一个或多个独特标识符诸如条形码。条形码可先前、随后或并行递送至容纳经划分或经分配细胞的分区中。举例来说，条形码可在小滴产生之前、之后或与小滴产生并行注射至小滴中。向特定分区中递送条形码允许稍后将单个细胞的特征归属于特定分区。条形码可例如通过任何适合机理来于核酸分子(例如寡核苷酸)上递送至分区中。条形码化核酸分子可通过微囊来递送至分区中。在一些情况下，微囊可包括珠粒。在以下进一步详细描述珠粒。

[0076] 在一些情况下，条形码化核酸分子可初始与微囊缔合，接着从微囊释放。条形码化核酸分子的释放可为被动的(例如通过扩散出微囊)。此外或替代地，从微囊的释放可在施加允许条形码化核酸分子从微囊解离或释放的刺激后。所述刺激可破坏微囊、使条形码化核酸分子偶联于微囊或在微囊内的相互作用、或两者。所述刺激可包括例如热刺激、光刺激、化学刺激(例如pH变化或使用还原剂)、机械刺激、辐射刺激；生物刺激(例如酶)或其任何组合。

[0077] 图2显示用于将携带条形码的珠粒递送至小滴中的微流体通道结构200的一实例。通道结构200可包括在通道接合部210处连通的通道区段201、202、204、206和208。在操作中，通道区段201可将包括多个珠粒214(例如具有核酸分子、寡核苷酸、分子标签)的水性流体212沿通道区段201运送至接合部210中。多个珠粒214可来源于珠粒混悬液。举例来说，通道区段201可连接于包含珠粒214的水性混悬液的储集器。通道区段202可将包括多个细胞216的水性流体212沿通道区段202运送至接合部210中。多个细胞216可来源于细胞混悬液。举例来说，通道区段202可连接于包含细胞216的水性混悬液的储集器。在一些情况下，第一通道区段201或第二通道区段202中或两个区段中的水性流体212可包括一种或多种如以下进一步所述的试剂。不与水性流体212混溶的第二流体218(例如油)可从通道区段204和206中的各者递送至接合部210中。在来自通道区段201和202中的各者的水性流体212和来自通道区段204和206中的各者的第二流体218在通道接合部210处相遇后，水性流体212可被分配成第二流体218中的离散小滴220，并且沿通道区段208从接合部210流走。通道区段208可将离散小滴递送至流体联接于通道区段208的出口储集器中，在所述储集器中，它们可加以收集。

[0078] 作为一替代方案，通道区段201和202可在接合部210的上游的另一接合部处相遇。在所述接合部处，珠粒和细胞可形成沿另一通道被引导至接合部210中以产生小滴220的混合物。混合物可以交替方式提供珠粒和细胞，以使例如小滴包含单一珠粒和单一细胞。

[0079] 珠粒、细胞和小滴可以实质上规则流动概况(例如在规则流速下)沿通道流动。所述规则流动概况可容许小滴包括单一珠粒和单一细胞。所述规则流动概况可容许小滴具有大于5％、10％、20％、30％、40％、50％、60％、70％、80％、90％或95％的占据率(例如小滴具有珠粒和细胞)。所述规则流动概况和可用于提供所述规则流动概况的装置例如提供于美国专利公布号2015/0292988中，所述美国专利公布以引用的方式整体并入本文。

[0080] 第二流体218可包含包括用于使所得小滴稳定，例如抑制所得小滴220的后续聚结的氟表面活性剂的油，诸如氟化油。

[0081] 产生的离散小滴可包括单个细胞216。产生的离散小滴可包括携带条形码或其他试剂的珠粒214。产生的离散小滴可包括单个细胞与携带条形码的珠粒两者，诸如小滴220。在一些情况下，离散小滴可包括超过一个单个细胞或不包括细胞。在一些情况下，离散小滴

可包括超过一个珠粒或不包括珠粒。离散小滴可未被占据(例如无珠粒、无细胞)。

[0082] 有益地,分配细胞和携带条形码的珠粒的离散小滴可有效允许将条形码归属于分区内细胞的核酸。某一分区的内含物可保持与其他分区的内含物离散。

[0083] 如将了解,本文所述的通道区段可联接于多种不同流体来源或接收部件中的任一者,包括储集器、管道、歧管、或其他系统的流体部件。如将了解,微流体通道结构200可具有其他几何结构。举例来说,微流体通道结构可具有超过一个通道接合部。举例来说,微流体通道结构可具有2、3、4或5个在通道接合部处相遇的通道区段,各自携带珠粒。流体可通过一个或多个流体流动单元来被引导沿一个或多个通道或储集器流动。流体流动单元可包括用以控制流体的流动的压缩机(例如提供正压)、泵(例如提供负压)、执行器等。流体也可或另外通过所施加的压差、离心力、电动泵送、真空、毛细管或重力流动等来控制。

[0084] 珠粒可为多孔的、非多孔的、固体的、半固体的、半流体的、流体的和/或其组合。在一些情况下,珠粒可为可溶解的、可破坏的和/或可降解的。在一些情况下,珠粒可不是可降解的。在一些情况下,珠粒可为凝胶珠粒。凝胶珠粒可为水凝胶珠粒。凝胶珠粒可由分子前体诸如聚合或单体物质形成。半固体珠粒可为脂质体珠粒。固体珠粒可包含金属,包括氧化铁、金和银。在一些情况下,珠粒可为二氧化硅珠粒。在一些情况下,珠粒可为刚性的。在其他情况下,珠粒可为柔性的和/或可压缩的。

[0085] 珠粒可具有任何适合形状。珠粒形状的实例包括但不限于球形、非球形、卵形、椭圆形、无定形、圆形、圆柱形及其变化形式。

[0086] 珠粒可具有均一尺寸或不均一尺寸。在一些情况下,珠粒的直径可为至少约10纳米(nm)、100nm、500nm、1微米(μm)、5μm、10μm、20μm、30μm、40μm、50μm、60μm、70μm、80μm、90μm、100μm、250μm、500μm、1mm或更大。在一些情况下,珠粒可具有小于约10nm、100nm、500nm、1μm、5μm、10μm、20μm、30μm、40μm、50μm、60μm、70μm、80μm、90μm、100μm、250μm、500μm、1mm或更小的直径。在一些情况下,珠粒可具有在约40-75μm、30-75μm、20-75μm、40-85μm、40-95μm、20-100μm、10-100μm、1-100μm、20-250μm、或20-500μm的范围内的直径。

[0087] 在某些方面,珠粒可以具有相对单分散尺寸分布的珠粒群体或多个珠粒形式提供。当可合乎需要的是在分区内提供相对一致数量的试剂时,维持相对一致珠粒特征诸如尺寸可促进总体一致性。特别地,本文所述的珠粒可具有在它们的横截面尺寸的情况下具有小于50%、小于40%、小于30%、小于20%,并且在一些情况下,小于15%、小于10%、小于5%或更小的变异系数的尺寸分布。

[0088] 珠粒可包含天然和/或合成物质。举例来说,珠粒可包含天然聚合物、合成聚合物、或天然聚合物与合成聚合物两者。天然聚合物的实例包括蛋白质和糖诸如脱氧核糖核酸、橡胶、纤维素、淀粉(例如直链淀粉、支链淀粉)、蛋白质、酶、多糖、丝、聚羟基链烷酸酯、壳聚糖、右旋糖酐、胶原、角叉菜胶、卵叶车前子(ispaghula)、阿拉伯胶、琼脂、明胶、虫胶、苹婆胶(sterculia gum)、黄原胶、玉米糖胶、瓜尔胶、刺梧桐胶、琼脂糖、海藻酸、海藻酸盐或其天然聚合物。合成聚合物的实例包括丙烯酸、尼龙、硅酮、斯潘德克斯(spandex)、粘胶人造丝、聚羧酸、聚乙酸乙烯酯、聚丙烯酰胺、聚丙烯酸酯、聚乙二醇、聚氨酯、聚乳酸酸、二氧化硅、聚苯乙烯、聚丙烯腈、聚丁二烯、聚碳酸酯、聚乙烯、聚对苯二甲酸乙二酯、聚(三氟氯乙烯)、聚(氧化乙烯)、聚(对苯二甲酸乙二酯)、聚乙烯、聚异丁烯、聚(甲基丙烯酸甲酯)、聚(甲醛)、聚甲醛、聚丙烯、聚苯乙烯、聚(四氟乙烯)、聚(乙酸乙烯酯)、聚(乙烯醇)、聚(氯乙

烯)、聚(偏二氯乙烯)、聚(偏二氟乙烯)、聚(氟乙烯)和/或其组合(例如共聚物)。珠粒也可由除聚合物以外的物质形成,包括脂质、胶束、陶瓷、玻璃-陶瓷、物质复合物、金属、其他无机物质以及其他物质。

[0089] 在一些情况下,珠粒可含有分子前体(例如单体或聚合物),其可通过分子前体的聚合来形成聚合物网络。在一些情况下,前体可为能够经受通过例如化学交联连键发生进一步聚合的已聚合物质。在一些情况下,前体可包括丙烯酰胺或甲基丙烯酰胺单体、寡聚物或聚合物中的一者或多者。在一些情况下,珠粒可包含预聚合物,其是能够进一步聚合的寡聚物。举例来说,聚氨酯珠粒可使用预聚合物来制备。在一些情况下,珠粒可含有可被进一步聚合在一起的单种聚合物。在一些情况下,珠粒可通过不同前体的聚合来产生,以使它们包含混合聚合物、共聚物和/或嵌段共聚物。在一些情况下,珠粒可在聚合前体(例如单体、寡聚物、线性聚合物)、核酸分子(例如寡核苷酸)、引物和其他实体之间包含共价键或离子键。在一些情况下,共价键可为碳-碳键、硫醚键或碳-杂原子键。

[0090] 视所用特定交联剂而定,交联可为永久的或可逆的。可逆交联可允许聚合物在适当条件下达成线性化或解离。在一些情况下,可逆交联也可允许可逆连接结合于珠粒的表面的物质。在一些情况下,交联剂可形成二硫连键。在一些情况下,形成二硫连键的化学交联剂可为胱胺或经改性胱胺。

[0091] 在一些情况下,二硫连键可在分子前体单元(例如单体、寡聚物或线性聚合物)或并入珠粒中的前体以及核酸分子(例如寡核苷酸)之间形成。举例来说,胱胺(包括经改性胱胺)是包含二硫键的有机剂,其可用作珠粒的单种单体或聚合前体之间的交联剂。聚丙烯酰胺可在胱胺或包含胱胺的物质(例如经改性胱胺)存在下聚合以产生包含二硫连键的聚丙烯酰胺凝胶珠粒(例如包含可化学还原交联剂的可化学降解珠粒)。在使珠粒暴露于还原剂后,二硫连键可容许珠粒被降解(或溶解)。

[0092] 在一些情况下,壳聚糖,作为一种线性多糖聚合物,可通过亲水链来与戊二醛交联以形成珠粒。壳聚糖聚合物的交联可通过由热、压力、pH变化和/或辐射引发的化学反应来实现。

[0093] 在一些情况下,珠粒可包含acrydite部分,其在某些方面可用于使一个或多个核酸分子(例如条形码序列、条形码化核酸分子、条形码化寡核苷酸、引物或其他寡核苷酸)连接于珠粒。在一些情况下,acrydite部分可指由acrydite与一种或多种物质的反应产生的acrydite类似物,所述反应诸如在聚合反应期间acrydite与其他单体和交联剂的反应。Acrydite部分可被改性来与待连接的物质诸如核酸分子(例如条形码序列、条形码化核酸分子、条形码化寡核苷酸、引物或其他寡核苷酸)形成化学键。Acrydite部分可用能够形成二硫键的硫醇基团改性,或可用已包含二硫键的基团改性。硫醇或二硫化物(通过二硫交换)可用作待连接物质的锚定点,或acrydite部分的另一部分可用于连接。在一些情况下,连接可为可逆的,以使当二硫键断裂(例如在还原剂存在下)时,所连接物质从珠粒释放。在其他情况下,acrydite部分可包含可用于连接的反应性羟基。

[0094] 对珠粒进行的用于连接核酸分子(例如寡核苷酸)的官能化可通过广泛范围的不同方法来实现,包括使聚合物内的化学基团活化,在聚合物结构中并入活性或可活化官能团,或在珠粒产生中的预聚合物或单体阶段时进行连接。

[0095] 举例来说,聚合以形成珠粒的前体(例如单体、交联剂)可包含acrydite部分,以使

当产生珠粒时，珠粒也包含acrydite部分。可使acrydite部分连接于核酸分子(例如寡核苷酸)，所述核酸分子可包括引发序列(例如用于扩增靶标核酸的引物、随机引物、用于信使RNA的引物序列)和/或一个或多个条形码序列。一个或多个条形码序列可包括对于偶联于给定珠粒的所有核酸分子都相同的序列和/或在偶联于给定珠粒的所有核酸分子之间不同的序列。核酸分子可并入珠粒中。

[0096] 在一些情况下，核酸分子可包含功能性序列，例如用于连接于测序流动槽，诸如像用于Illumina®测序的P5序列。在一些情况下，核酸分子或其衍生物(例如从核酸分子产生的寡核苷酸或多核苷酸)可包含另一功能性序列，诸如像用于连接于Illumina测序的测序流动槽的P7序列。在一些情况下，核酸分子可包含条形码序列。在一些情况下，引物可进一步包含独特分子标识符(UMI)。在一些情况下，引物可包含用于Illumina测序的R1引物序列。在一些情况下，引物可包含用于Illumina测序的R2引物序列。如可与本公开的组合物、装置、方法和系统一起使用的所述核酸分子(例如寡核苷酸、多核苷酸等)及其使用的实例提供于美国专利公布号2014/0378345和2015/0376609中，所述美国专利公布各自以引用的方式整体并入本文。

[0097] 在操作中，细胞可连同携带条形码的珠粒一起共分配。条形码化核酸分子可在分区中从珠粒释放。举例来说，在分析样品RNA的情形下，一个经释放核酸分子的多dT(多脱氧胸腺嘧啶，也被称为寡(dT))区段可杂交于mRNA分子的多A尾部。逆转录可产生mRNA的cDNA转录物，但所述转录物包括核酸分子的各序列区段。因为核酸分子包含锚定序列，所以它将更可能杂交于mRNA的多A尾部的序列末端，并且在所述序列末端处引发逆转录。在任何给定分区内，单个mRNA分子的所有cDNA转录物都可包括共同条形码序列区段。然而，由给定分区内的不同mRNA分子制得的转录物可在独特分子标识序列区段(例如UMI区段)处变化。有益地，甚至在对给定分区的内含物的任何后续扩增之后，不同UMI的数目也可指示来源于给定分区，因此来源于细胞的mRNA的数量。如上所指示，转录物可被扩增，净化以及测序以鉴定mRNA的cDNA转录物的序列，以及对条形码区段和UMI区段测序。尽管描述的是多dT引物序列，但其他靶向或随机引发序列也可用于引发逆转录反应。同样，尽管描述为使条形码化寡核苷酸释放至分区中，但在一些情况下，结合于珠粒(例如凝胶珠粒)的核酸分子可用于在珠粒的固相上杂交和捕集mRNA，例如以有助于使RNA与其他细胞内含物分离。

[0098] 在一些情况下，包含具有反应性或能够被活化以使它变得具有反应性的官能团的前体可与其他前体聚合以产生包含经活化或可活化官能团的凝胶珠粒。官能团可接着用于使额外物质(例如二硫接头、引物、其他寡核苷酸等)连接于凝胶珠粒。举例来说，包含羧酸(COOH)基团的一些前体可与其他前体共聚合以形成也包含COOH官能团的凝胶珠粒。在一些情况下，丙烯酸(一种包含游离COOH基团的物质)、丙烯酰胺和双(丙烯酰基)胱胺可一起共聚合以产生包含游离COOH基团的凝胶珠粒。凝胶珠粒的COOH基团可被活化(例如通过1-乙基-3-(3-二甲基氨基丙基)碳二亚胺(EDC)和N-羟基丁二酰亚胺(NHS)或氯化4-(4,6-二甲氧基-1,3,5-三嗪-2-基)-4-甲基吗啉鎓(DMTMM))以使它们具有反应性(例如与胺官能团具有反应性，其中EDC/NHS或DMTMM用于活化)。经活化COOH基团可接着与包含待连接于珠粒的部分的适当物质(例如包含胺官能团的物质，其中羧酸基团被活化以与胺官能团具有反应性)反应。

[0099] 在它们的聚合网络中包含二硫连键的珠粒可通过使一些二硫连键还原成游离硫

醇来用额外物质官能化。二硫连键可通过例如还原剂(例如DTT、TCEP等)的作用来还原以产生游离硫醇基团,而不使珠粒溶解。珠粒的游离硫醇可接着与某一物质的游离硫醇或包含另一二硫键的物质反应(例如通过硫醇-二硫交换)以使所述物质可被连接于珠粒(例如通过产生的二硫键)。在一些情况下,珠粒的游离硫醇可与任何其他适合基团反应。举例来说,珠粒的游离硫醇可与包含acrydite部分的物质反应。珠粒的游离硫醇基团可通过迈克尔(Michael)加成化学来与acrydite反应,以使包含acrydite的物质连接于珠粒。在一些情况下,不受控制的反应可通过包括硫醇封盖剂诸如N-乙基顺丁烯二酰胺或碘乙酸酯来阻止。

[0100] 可控制珠粒内二硫连键的活化以使仅少量二硫连键被活化。控制可例如通过控制用于产生游离硫醇基团的还原剂的浓度和/或用于在珠粒聚合中形成二硫键的试剂的浓度来施加。在一些情况下,低浓度(例如还原剂的分子:凝胶珠粒比率小于或等于约1:100,000,000,000、小于或等于约1:10,000,000,000、小于或等于约1:1,000,000,000、小于或等于约1:100,000,000、小于或等于约1:10,000,000、小于或等于约1:1,000,000、小于或等于约1:100,000、小于或等于约1:10,000)的还原剂可用于还原。控制还原成游离硫醇的二硫连键的数目可适用于确保在官能化期间的珠粒结构完整性。在一些情况下,光学活性剂诸如荧光染料可通过珠粒的游离硫醇基团来偶联于珠粒,并且用于定量存在于珠粒中的游离硫醇的数目和/或追踪珠粒。

[0101] 在一些情况下,在凝胶珠粒形成之后将部分添加至凝胶珠粒中可为有利的。举例来说,在凝胶珠粒形成之后添加寡核苷酸(例如条形码化寡核苷酸)可避免在可发生在聚合期间的链转移终止期间的物质损失。此外、较小前体(例如不包含侧链基团和所连接部分的单体或交联剂)可用于聚合,并且可最小程度地由于粘滞效应而被阻碍使链末端增长。在一些情况下,在凝胶珠粒合成之后的官能化可使物质(例如寡核苷酸)由于暴露而装载有潜在损害剂(例如自由基)最小化和/或使物质(例如寡核苷酸)对化学环境的暴露最小化。在一些情况下,产生的凝胶可具有可容许对珠粒达成温度驱动的溶胀和崩塌的上临界溶解温度(UCST)。在用寡核苷酸对珠粒的后续官能化期间,所述功能性可有助于寡核苷酸(例如引物)浸润至珠粒中。产生后官能化也可适用于控制物质在珠粒中的装载比率,以使例如装载比率可变性得以最小化。物质装载也可以批式方法进行以使多个珠粒可在单一批次中用物质官能化。

[0102] 注射至分区中或以其他方式引入分区中的珠粒可包含以可释放方式、可裂解方式或可逆方式连接的条形码。注射至分区中或以其他方式引入分区中的珠粒可包含可活化条形码。注射至分区中或以其他方式引入分区中的珠粒可为可降解、可破坏或可溶解珠粒。

[0103] 条形码可以可释放方式、可裂解方式或可逆方式连接于珠粒,以使条形码可通过条形码分子与珠粒之间的连键的裂解而释放或可释放,或通过下层珠粒自身的降解来释放,从而允许条形码为其他试剂所及或可及,或两种情况均存在。在非限制性实例中,裂解可通过使二硫键还原、使用限制酶、光活化裂解,或通过其他类型的刺激(例如化学、热、pH、酶等)和/或诸如在本文中其他地方所述的反应达成的裂解来实现。可释放条形码可有时被称为可活化,因为它们一旦释放即可用于反应。因此,举例来说,可活化条形码可通过使条形码从珠粒(或本文所述的其他适合类型的分区)释放来活化。在所述方法和系统的情形下,也设想其他可活化配置。

[0104] 除珠粒与缔合分子诸如含有条形码的核酸分子(例如条形码化寡核苷酸)之间的

可裂解连键之外，或作为所述可裂解连键的替代方案，珠粒可自发地或在暴露于一种或多种刺激(例如温度变化、pH变化、暴露于特定化学物质或化学相、暴露于光照、还原剂等)后可降解、可破坏或可溶解。在一些情况下，珠粒可为可溶解的，以使珠粒的物质组分在暴露于特定化学物质或环境变化诸如温度变化或pH变化时溶解。在一些情况下，凝胶珠粒可在高温下和/或在碱性条件下降解或溶解。在一些情况下，珠粒可为可热降解的，以使当所述珠粒暴露于适当温度变化(例如热)时，所述珠粒降解。结合于物质(例如核酸分子，例如条形码化寡核苷酸)的珠粒的降解或溶解可导致所述物质从所述珠粒释放。

[0105] 如将根据以上公开内容所了解，珠粒的降解可指所结合或经夹带物质从珠粒解离，伴有以及不伴有使实体珠粒自身发生结构降解。举例来说，珠粒的降解可涉及通过一种或多种在本文中其他地方所述的物质和/或方法来使可裂解连键裂解。在另一实例中，经夹带物质可通过归因于例如化学环境改变的渗透压差来从珠粒释放。举例来说，归因于渗透压差的珠粒孔尺寸的改变可通常在无珠粒自身的结构降解下发生。在一些情况下，归因于珠粒的渗透溶胀的孔尺寸增加可容许所述珠粒内的经夹带物质的释放。在其他情况下，珠粒的渗透收缩可由于孔尺寸紧缩而导致珠粒更好保留经夹带物质。

[0106] 可将可降解珠粒引入分区诸如乳液的小滴或孔中，以使当施加适当刺激时，所述珠粒在所述分区内降解，并且任何缔合物质(例如寡核苷酸)在所述小滴内释放。游离物质(例如寡核苷酸、核酸分子)可与分区中含有的其他试剂相互作用。举例来说，可使包含胱胺，并且通过二硫键来连接于条形码序列的聚丙烯酰胺珠粒与还原剂在油包水乳液的小滴内组合。在小滴内，还原剂可使各个二硫键断裂，从而导致珠粒降解以及条形码序列释放至小滴的水性内部环境中。在另一实例中，在碱性溶液中加热包含珠粒结合的条形码序列的小滴也可导致珠粒降解以及所连接条形码序列释放至所述小滴的水性内部环境中。

[0107] 任何适合数目的分子标签分子(例如引物、条形码化寡核苷酸)都可与珠粒缔合，以使在从所述珠粒释放后，分子标签分子(例如引物，例如条形码化寡核苷酸)在预定浓度下存在于分区中。可选择所述预定浓度以有助于分区内用于产生测序文库的某些反应例如扩增。在一些情况下，引物的预定浓度可受产生携带核酸分子(例如寡核苷酸)的珠粒的方法限制。

[0108] 在一些情况下，珠粒可用一种或多种试剂以非共价方式装载。珠粒可通过例如以下方式来以非共价方式装载：使珠粒经受足以使珠粒溶胀的条件，允许有足够时间来供试剂扩散至珠粒的内部中，以及使珠粒经受足以使珠粒去溶胀的条件。珠粒的溶胀可例如通过以下方式来实现：将珠粒放置在热力学上有利的溶剂中，使珠粒经受较高或较低温度，使珠粒经受较高或较低离子浓度，和/或使珠粒经受电场。珠粒的溶胀可通过各种溶胀方法来实现。珠粒的去溶胀可例如通过以下方式来实现：将珠粒转移于热力学上不利的溶剂中，使珠粒经受较低温度或高温，使珠粒经受较低或较高离子浓度，和/或移除电场。珠粒的去溶胀可通过各种去溶胀方法来实现。转移珠粒可导致珠粒中的孔收缩。收缩可接着阻碍珠粒内的试剂扩散出珠粒的内部。阻碍可归因于试剂与珠粒的内部之间的空间相互作用。转移可以微流体方式实现。举例来说，转移可通过使珠粒从一个同向流动溶剂流移动至不同同向流动溶剂流来实现。珠粒的溶胀性和/或孔尺寸可通过改变珠粒的聚合物组成来调整。

[0109] 在一些情况下，连接于前体的acrydite部分、连接于前体的另一物质、或前体自身可包含不稳定键诸如化学敏感、热敏感或光敏感键，例如二硫键、UV敏感键等。一旦包含不

稳定键的acrydite部分或其他部分并入珠粒中，所述珠粒即也可包含所述不稳定键。不稳定键可例如适用于以可逆方式使物质(例如条形码、引物等)连接(例如共价连接)于珠粒。在一些情况下，热不稳定键可包括基于核酸杂交的连接，例如其中寡核苷酸杂交于连接于珠粒的互补序列，以使杂交物的热熔融使所述寡核苷酸例如含有条形码的序列从珠粒或微囊释放。

[0110] 将多个类型的不稳定键添加至凝胶珠粒中可导致产生能够对不同刺激起响应的的珠粒。各类型的不稳定键可对相关刺激(例如化学刺激、光照、温度、酶等)敏感，以使通过各不稳定键来连接于珠粒的物质的释放可通过施加适当刺激来控制。所述功能性可适用于物质从凝胶珠粒的控制释放。在一些情况下，包含不稳定键的另一物质可在凝胶珠粒形成之后通过例如凝胶珠粒的如上所述的经活化官能团来连接于所述凝胶珠粒。如将了解，以可释放方式、可裂解方式或可逆方式连接于本文所述的珠粒的条形码包括以下条形码：所述条形码通过条形码分子与珠粒之间的连键的裂解而释放或可释放，或通过下层珠粒自身的降解来释放，从而允许条形码为其他试剂所及或可及，或两种情况均存在。

[0111] 除可热裂解键、二硫键和UV敏感键之外，可偶联于前体或珠粒的不稳定键的其他非限制性实例包括酯连键(例如可用酸、碱或羟胺裂解)、邻二醇连键(例如可通过过碘酸钠来裂解)、狄尔斯-阿德耳(Diels-Alder)连键(例如可通过热来裂解)、砜连键(例如可通过碱来裂解)、甲硅烷基醚连键(例如可通过酸来裂解)、糖苷连键(例如可通过淀粉酶来裂解)、肽连键(例如可通过蛋白酶来裂解)、或磷酸二酯连键(例如可通过核酸酶(例如DNA酶)来裂解)。键可通过如以下进一步所述的其他核酸分子靶向酶诸如限制酶(例如限制核酸内切酶)来裂解。

[0112] 物质可在珠粒产生期间(例如在前体的聚合期间)囊封在珠粒中。所述物质可或可不参与聚合。所述物质可进入聚合反应混合物中，以使在珠粒形成后，产生的珠粒包含所述物质。在一些情况下，可在形成之后将所述物质添加至凝胶珠粒中。所述物质可包括例如核酸分子(例如寡核苷酸)、用于核酸扩增反应的试剂(例如引物、聚合酶、dNTP、辅因子(例如离子性辅因子)、缓冲剂)，包括本文所述的那些，用于酶促反应的试剂(例如酶、辅因子、底物、缓冲剂)、用于核酸修饰反应诸如聚合、连接或消化的试剂、和/或用于针对一个或多个测序平台进行模板制备(例如标签化)(例如针对Illumina®的Nextera®)的试剂。所述物质可包括一种或多种本文所述的酶，包括不限于聚合酶、逆转录酶、限制酶(例如核酸内切酶)、转座酶、连接酶、蛋白酶K、DNA酶等。所述物质可包括一种或多种在本文中其他地方所述的试剂(例如溶解剂、抑制剂、失活剂、螯合剂、刺激物)。对所述物质的圈闭可通过在前体的聚合期间产生的聚合物网络密度，通过控制凝胶珠粒内的离子电荷(例如通过连接于聚合物质的离子性物质)，或通过其他物质的释放来控制。囊封物质可在珠粒降解后和/或通过施加能够使物质从珠粒释放的刺激来从珠粒释放。或者或此外，物质可在分区形成期间或之后被分配在分区(例如小滴)中。所述物质可包括不限于以上提及的也可囊封在珠粒中的物质。

[0113] 可降解珠粒可包含一种或多种具有不稳定键的物质，以使当使珠粒/物质暴露于适当刺激时，所述键断裂，并且所述珠粒降解。不稳定键可为化学键(例如共价键、离子键)，或可为另一类型的物理相互作用(例如范德华(van der Waals)相互作用、偶极-偶极相互作用等)。在一些情况下，用于产生珠粒的交联剂可包含不稳定键。在暴露于适当条件后，不

稳定键可断裂，并且珠粒降解。举例来说，在包含胱胺交联剂的聚丙烯酰胺凝胶珠粒暴露于还原剂后，胱胺的二硫键可断裂，并且珠粒降解。

[0114] 相较于不降解的珠粒，可降解珠粒可适用于在将适当刺激施加于珠粒时更快速使所连接物质(例如核酸分子、条形码序列、引物等)从珠粒释放。举例来说，对于结合于多孔珠粒的内表面的物质或在囊封物质的情况下，在所述珠粒降解后，所述物质在溶液中可具有较大移动性以及其他物质可及性。在一些情况下，物质也可通过可降解接头(例如二硫接头)来连接于可降解珠粒。可降解接头与可降解珠粒可对同一刺激起响应，或两种可降解物质可对不同刺激起响应。举例来说，条形码序列可通过二硫键来连接于包含胱胺的聚丙烯酰胺珠粒。在条形码化珠粒暴露于还原剂后，珠粒降解，并且条形码序列在条形码序列与珠粒之间的二硫连键以及珠粒中胱胺的二硫连键都断裂后释放。

[0115] 当提供可降解珠粒时，可有益的是在给定时间之前避免所述珠粒暴露于一种或多种导致所述降解的刺激，以例如避免过早珠粒降解以及由所述降解产生的问题，包括例如不良流动特征和聚集。举例来说，当珠粒包含可还原交联基团诸如二硫基团时，将合乎需要的是避免使所述珠粒与还原剂例如DTT或其他二硫裂解试剂接触。在所述情况下，对本文所述的珠粒的处理将在一些情况下在不含还原剂诸如DTT下提供。因为还原剂经常提供于商业酶制剂中，所以可合乎需要的是在处理本文所述的珠粒时提供不含还原剂(或不含DTT)的酶制剂。所述酶的实例包括例如聚合酶酶制剂、逆转录酶酶制剂、连接酶酶制剂以及可用于处理本文所述的珠粒的许多其他酶制剂。术语“不含还原剂”或“不含DTT”的制剂可指制剂具有的所述物质的量是所述物质用于使珠粒降解的下部范围的小于约1/10、小于约1/50、或甚至小于约1/100。举例来说，对于DTT，不含还原剂的制剂可具有小于约0.01毫体积摩尔浓度(mM)、0.005mM、0.001mM DTT、0.0005mM DTT、或甚至小于约0.0001mM DTT。在许多情况下，DTT的量可为不可检测的。

[0116] 众多化学触发可用于触发珠粒的降解。这些化学变化的实例可包括但不限于珠粒内组分的完整性的由pH介导的变化、珠粒的组分的通过交联键的裂解而达成的降解、以及珠粒的组分的解聚。

[0117] 在一些实施方案中，珠粒可由包含可降解化学交联剂诸如BAC或胱胺的物质形成。所述可降解交联剂的降解可通过许多机理来实现。在一些实例中，可使珠粒与可诱导氧化、还原或其他化学变化的化学降解剂接触。举例来说，化学降解剂可为还原剂诸如二硫苏糖醇(DTT)。还原剂的额外实例可包括β-巯基乙醇、(2S)-2-氨基-1,4-二巯基丁烷(二硫丁胺或DTBA)、三(2-羧乙基)膦(TCEP)或其组合。还原剂可使在形成珠粒的凝胶前体之间形成的二硫键降解，因此，使珠粒降解。在其他情况下，溶液的pH变化诸如pH增加可触发珠粒的降解。在其他情况下，暴露于水溶液诸如水可触发水解降解，因此触发珠粒的降解。在一些情况下，刺激的任何组合都可触发珠粒的降解。举例来说，pH变化可使得化学剂(例如DTT)能够变为有效还原剂。

[0118] 在施加热刺激后，珠粒也可被诱导来释放它们的内含物。温度变化可导致珠粒的多种变化。举例来说，热可导致固体珠粒液化。热变化可导致珠粒的熔融，以使所述珠粒的一部分降解。在其他情况下，热可使珠粒组分的内部压力增加，以使珠粒破裂或爆裂。热也可作用于用作构建珠粒的物质的热敏感聚合物。

[0119] 任何适合试剂都可使珠粒降解。在一些实施方案中，温度或pH变化可用于使珠粒

内的热敏感或pH敏感键降解。在一些实施方案中，化学降解剂可用于通过氧化、还原或其他化学变化来使珠粒内的化学键降解。举例来说，化学降解剂可为还原剂诸如DTT，其中DTT可使在交联剂与凝胶前体之间形成的二硫键降解，由此使珠粒降解。在一些实施方案中，可添加还原剂以使珠粒降解，此可或可不导致珠粒释放它的内含物。还原剂的实例可包括二硫苏糖醇(DTT)、β-巯基乙醇、(2S)-2-氨基-1,4-二巯基丁烷(二硫丁胺或DTBA)、三(2-羧乙基)膦(TCEP)或其组合。还原剂可在约0.1mM、0.5mM、1mM、5mM、10mM的浓度下存在。还原剂可在至少约0.1mM、0.5mM、1mM、5mM、10mM、或大于10mM的浓度下存在。还原剂可在至多约10mM、5mM、1mM、0.5mM、0.1mM或更小的浓度下存在。

[0120] 尽管图1和图2已在以上以提供大致上经单一占据分区加以描述，但在某些情况下，可合乎需要的是提供经多重占据分区，例如在单一分区内含有两个、三个、四个或更多个细胞和/或包含条形码化核酸分子(例如寡核苷酸)的微囊(例如珠粒)。因此，如上所指示，可控制含有细胞和/或珠粒的流体和分配流体的流动特征以提供所述经多重占据分区。特别地，可控制流动参数以提供分区的处于大于约50％、大于约75％、并且在一些情况下，大于约80％、90％、95％或更高的给定占据率。

[0121] 在一些情况下，额外微囊可用于将额外试剂递送至分区中。在所述情况下，可有利的是将不同珠粒引入共同通道或小滴产生接合部中，通过不同通道进口从不同珠粒来源(例如含有不同缔合试剂)进入所述共同通道或小滴产生接合部(例如接合部210)中。在所述情况下，可控制不同珠粒进入通道或接合部中的流动和频率以提供某一比率的来自各来源的微囊，同时确保进入分区中的所述珠粒与给定数目的细胞的给定配对或组合(例如每个分区一个细胞和一个珠粒)。

[0122] 本文所述的分区可包含小体积，例如小于约10微升(μL)、5μL、1μL、900皮升(pL)、800pL、700pL、600pL、500pL、400pL、300pL、200pL、100pL、50pL、20pL、10pL、1pL、500纳升(nL)、100nL、50nL或更小。

[0123] 举例来说，在基于小滴的分区的情况下，小滴可具有小于约1000pL、900pL、800pL、700pL、600pL、500pL、400pL、300pL、200pL、100pL、50pL、20pL、10pL、1pL或更小的总体积。当与微囊共分配时，应了解，分区内例如包括共分配生物粒子和/或珠粒的样品流体体积可为上述体积的小于约90％，上述体积的小于约80％、小于约70％、小于约60％、小于约50％、小于约40％、小于约30％、小于约20％或小于约10％。

[0124] 如在本文中其他地方所述，分配物质可产生分区群体或多个分区。在所述情况下，可产生或另外提供任何适合数目的分区。举例来说，可产生或另外提供至少约1,000个分区、至少约5,000个分区、至少约10,000个分区、至少约50,000个分区、至少约100,000个分区、至少约500,000个分区、至少约1,000,000个分区、至少约5,000,000个分区至少约10,000,000个分区、至少约50,000,000个分区、至少约100,000,000个分区、至少约500,000,000个分区、至少约1,000,000,000个分区或更多个分区。此外，多个分区可包括未占据分区(例如空分区)与经占据分区两者。

[0125] 试剂

[0126] 根据某些方面，细胞可连同溶解试剂一起分配以在分区内释放细胞的内含物。在所述情况下，可与将细胞诸如通过一个或多个在通道接合部的上游的额外通道来引入分配接合部/小滴产生区(例如接合部210)中并行或紧靠在所述引入之前，使溶解剂与细胞混悬

液接触。有益地，当共分配溶解试剂和生物粒子时，溶解试剂可有助于生物粒子的内含物在分区内的释放。在分区中释放的内含物可保持与其他分区的内含物离散。

[0127] 溶解剂的实例包括生物活性试剂，诸如用于使不同细胞类型例如革兰氏阳性或阴性细菌、植物、酵母、哺乳动物等溶解的溶解酶，诸如溶菌酶、无色肽酶(achromopeptidase)、溶葡萄球菌素(lysostaphin)、labiase、kitalase、溶壁酶(lyticase)和可从例如Sigma-Aldrich,Inc.(St Louis,MO)获得的多种其他溶解酶以及其他可商购获得的溶解酶。其他溶解剂可另外或替代地与细胞共分配以导致细胞的内含物释放至分区中。举例来说，在一些情况下，基于表面活性剂的溶解溶液可用于使细胞溶解，但这些溶液可较不合乎基于乳液的系统需要，其中表面活性剂可干扰稳定乳液。在一些情况下，溶解溶液可包括非离子表面活性剂诸如像曲通X-100(TritonX-100)和吐温20(Tween 20)。在一些情况下，溶解溶液可包括离子表面活性剂诸如像N-月桂酰基肌氨酸钠和十二烷基硫酸钠(SDS)。电穿孔、热、声学或机械细胞破坏也可在某些情况下使用，例如非基于乳液的分配，诸如可除小滴分配之外也采用或替代小滴分配加以采用的细胞囊封，其中囊封物的任何孔尺寸都是足够小的以致在细胞破坏之后保留具有给定尺寸的核酸片段。

[0128] 或者或除与上述细胞共分配的溶解剂之外，其他试剂也可与细胞共分配，包括例如DNA酶和RNA酶失活剂或抑制剂诸如蛋白酶K、螯合剂诸如EDTA、以及用于移除或另外降低不同细胞溶解产物组分对后续核酸处理的负性活性或影响的其他试剂。此外，在囊封细胞的情况下，可使细胞暴露于适当刺激以使细胞或它们的内含物从共分配微囊释放。举例来说，在一些情况下，化学刺激可连同囊封细胞一起共分配以允许达成微囊的降解以及细胞或它的内含物释放至较大分区中。在一些情况下，这个刺激可与在本文中其他地方所述的刺激相同以使核酸分子(例如寡核苷酸)从它们的相应微囊(例如珠粒)释放。在替代性方面，这可为不同和非重叠刺激，以允许囊封细胞在与核酸分子释放至分区中不同的时间释放至同一分区中。

[0129] 如将了解，许多其他试剂可连同细胞、珠粒、溶解剂和化学刺激一起共分配，包括例如保护性试剂如蛋白酶K、螯合剂、核酸延伸、复制、转录或扩增试剂诸如聚合酶、逆转录酶、可用于基于转座子的方法(例如Nextera)的转座酶、三磷酸核苷或NTP类似物、引物序列和用于所述反应中的额外辅因子诸如二价金属离子、连接反应试剂诸如连接酶和连接序列、染料、标记或其他加标签试剂。

[0130] 额外试剂也可与细胞共分配，诸如用以使细胞的DNA碎裂的核酸内切酶、用于扩增细胞的核酸片段以及使条形码分子标签连接于经扩增片段的DNA聚合酶和dNTP。可共分配其他酶，包括不限于聚合酶、转座酶、连接酶、蛋白酶K、DNA酶等。额外试剂也可包括逆转录酶，包括具有末端转移酶活性的酶；引物和寡核苷酸以及可用于模板转换的转换寡核苷酸(在本文中也称为“转换寡聚物”或“模板转换寡核苷酸”)。在一些情况下，模板转换可用于使cDNA的长度增加。在一些情况下，模板转换可用于使预定核酸序列附接于cDNA。在模板转换的一实例中，cDNA可由模板例如细胞mRNA的逆转录产生，其中具有末端转移酶活性的逆转录酶可以模板非依赖性方式将额外核苷酸例如多C添加至cDNA中。转换寡聚物可包括互补于额外核苷酸的序列例如多G。cDNA上的额外核苷酸(例如多C)可杂交于转换寡聚物上的额外核苷酸(例如多G)，借此转换寡聚物可由逆转录酶用作模板以进一步延伸cDNA。模板转换寡核苷酸可包含杂交区域和模板区域。杂交区域可包含能够杂交于靶标的任何序列。在

一些情况下，如先前所述，杂交区域包含一系列G碱基以互补于在cDNA分子的3’末端的突出C碱基。所述一系列的G碱基可包含1个G碱基、2个G碱基、3个G碱基、4个G碱基、5个G碱基或超过5个G碱基。模板序列可包括待并入cDNA中的任何序列。在一些情况下，模板区域包含至少1个(例如至少2、3、4、5个或更多个)标签序列和/或功能性序列。转换寡聚物可包含脱氧核糖核酸；核糖核酸；经修饰核酸，包括2-氨基嘌呤、2,6-二氨基嘌呤(2-氨基-dA)、反转dT、5-甲基dC、2’-脱氧肌苷、Super T(5-羟基丁炔基-2’-脱氧尿苷)、Super G(8-氮杂-7-脱氮杂鸟苷)、锁定核酸(LNA)、解锁核酸(UNA，例如UNA-A、UNA-U、UNA-C、UNA-G)、异dG、异dC、2’氟代碱基(例如氟代C、氟代U、氟代A和氟代G)或任何组合。

[0131] 在一些情况下，转换寡聚物的长度可为2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、114、115、116、117、118、119、120、121、122、123、124、125、126、127、128、129、130、131、132、133、134、135、136、137、138、139、140、141、142、143、144、145、146、147、148、149、150、151、152、153、154、155、156、157、158、159、160、161、162、163、164、165、166、167、168、169、170、171、172、173、174、175、176、177、178、179、180、181、182、183、184、185、186、187、188、189、190、191、192、193、194、195、196、197、198、199、200、201、202、203、204、205、206、207、208、209、210、211、212、213、214、215、216、217、218、219、220、221、222、223、224、225、226、227、228、229、230、231、232、233、234、235、236、237、238、239、240、241、242、243、244、245、246、247、248、249、250个核苷酸或更长。

[0132] 在一些情况下，转换寡聚物的长度可为至少约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、114、115、116、117、118、119、120、121、122、123、124、125、126、127、128、129、130、131、132、133、134、135、136、137、138、139、140、141、142、143、144、145、146、147、148、149、150、151、152、153、154、155、156、157、158、159、160、161、162、163、164、165、166、167、168、169、170、171、172、173、174、175、176、177、178、179、180、181、182、183、184、185、186、187、188、189、190、191、192、193、194、195、196、197、198、199、200、201、202、203、204、205、206、207、208、209、210、211、212、213、214、215、216、217、218、219、220、221、222、223、224、225、226、227、228、229、230、231、232、233、234、235、236、237、238、239、240、241、242、243、244、245、246、247、248、249或250个核苷酸或更长。

[0133] 在一些情况下，转换寡聚物的长度可为至多约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、

111、112、113、114、115、116、117、118、119、120、121、122、123、124、125、126、127、128、129、130、131、132、133、134、135、136、137、138、139、140、141、142、143、144、145、146、147、148、149、150、151、152、153、154、155、156、157、158、159、160、161、162、163、164、165、166、167、168、169、170、171、172、173、174、175、176、177、178、179、180、181、182、183、184、185、186、187、188、189、190、191、192、193、194、195、196、197、198、199、200、201、202、203、204、205、206、207、208、209、210、211、212、213、214、215、216、217、218、219、220、221、222、223、224、225、226、227、228、229、230、231、232、233、234、235、236、237、238、239、240、241、242、243、244、245、246、247、248、249或250个核苷酸。

[0134] 一旦细胞的内含物被释放至它们的相应分区中，其中含有的核酸即可在分区内被进一步处理。根据本文所述的方法和系统，单个细胞的核酸内含物可被提供以独特标识符，以使在表征那些核酸后，它们可被归属为已源于上述一个或多个细胞。将特征归属于单个细胞或细胞群组的能力通过对单个细胞或细胞群组特定指定独特标识符来提供，此是本文所述的方法和系统的另一有利方面。特别地，指定例如呈核酸条形码形式的独特标识符，或使其与单个细胞或细胞群体相关联，以用独特标识符对细胞的组分(以及因此，对它的特征)加标签或进行标记。这些独特标识符接着用于将细胞的组分和特征归属于单个细胞或细胞群组。

[0135] 在一些方面，这通过将单个细胞或细胞群组与诸如以上所述的独特标识符共分配来进行(参照图2)。在一些方面，独特标识符以包含核酸条形码序列的核酸分子(例如寡核苷酸)形式提供，所述核酸条形码序列可连接于单个细胞的核酸内含物或另外与单个细胞的核酸内含物缔合，或连接于细胞的其他组分，并且特别是连接于那些核酸的片段。对核酸分子进行分配以使在给定分区中的核酸分子之间，其中含有的核酸条形码序列是相同的，但在不同分区之间，核酸分子可并且确实具有不同条形码序列，或至少代表在给定分析中跨越所有分区的许多不同条形码序列。在一些方面，仅一个核酸条形码序列可与给定分区相关联，但在一些情况下，可存在两个或更多个不同条形码序列。

[0136] 在核酸分子(例如寡核苷酸)的序列内，核酸条形码序列可包括约6至约20个或更多个核苷酸。核酸条形码序列可包括约6至约20、30、40、50、60、70、80、90、100个或更多个核苷酸。在一些情况下，条形码序列的长度可为约6、7、8、9、10、11、12、13、14、15、16、17、18、19、20个核苷酸或更长。在一些情况下，条形码序列的长度可为至少约6、7、8、9、10、11、12、13、14、15、16、17、18、19、20个核苷酸或更长。在一些情况下，条形码序列的长度可为至多约6、7、8、9、10、11、12、13、14、15、16、17、18、19、20个核苷酸或更短。这些核苷酸可为完全连续的，即在单一邻近核苷酸链段中，或它们可被分成两个或更多个由1个或更多个核苷酸分隔的单独子序列。在一些情况下，经分隔条形码子序列的长度可为约4至约16个核苷酸。在一些情况下，条形码子序列可为约4、5、6、7、8、9、10、11、12、13、14、15、16个核苷酸或更长。在一些情况下，条形码子序列可为至少约4、5、6、7、8、9、10、11、12、13、14、15、16个核苷酸或更长。在一些情况下，条形码子序列可为至多约4、5、6、7、8、9、10、11、12、13、14、15、16个核苷酸或更短。

[0137] 共分配寡核苷酸也可包含适用于处理来自共分配细胞的核酸的其他功能性序列。这些序列包括例如用于扩增来自分区内单个细胞的基因组DNA，同时连接相关条形码序列的靶向或随机/通用扩增引物序列，测序引物或引物识别位点、例如用于鉴定序列的存在性

或用于下拉条形码化核酸的杂交或探测序列，或许多其他潜在功能性序列中的任一者。也可采用对寡核苷酸进行共分配的其他机理，包括例如两个或更多个小滴的聚结，其中一个小滴含有寡核苷酸，或将寡核苷酸微分配至分区例如微流体系统内的小滴中。

[0138] 在一实例中，提供微囊诸如珠粒，其各自包括以可释放方式连接于珠粒的大量上述条形码化核酸分子(例如条形码化寡核苷酸)，其中连接于特定珠粒的所有核酸分子都将包括相同核酸条形码序列，但其中跨越所用珠粒群体代表许多不同条形码序列。在一些实施方案中，例如包含聚丙烯酰胺聚合物基质的水凝胶珠粒用作核酸分子进入分区中的固体载体和递送载体，因为它们能够携带大量核酸分子，并且可被配置来在暴露于特定刺激后使那些核酸分子释放，如在本文中其他地方所述。在一些情况下，珠粒群体提供多样性条形码序列文库，其包括至少约1,000个不同条形码序列、至少约5,000个不同条形码序列、至少约10,000个不同条形码序列、至少约50,000个不同条形码序列、至少约100,000个不同条形码序列、至少约1,000,000个不同条形码序列、至少约5,000,000个不同条形码序列、或至少约10,000,000个不同条形码序列或更多。另外，各珠粒可被提供以大量所连接核酸(例如寡核苷酸)分子。特别地，在单个珠粒上，包括条形码序列的核酸分子的分子数可为至少约1,000个核酸分子、至少约5,000个核酸分子、至少约10,000个核酸分子、至少约50,000个核酸分子、至少约100,000个核酸分子、至少约500,000个核酸、至少约1,000,000个核酸分子、至少约5,000,000个核酸分子、至少约10,000,000个核酸分子、至少约50,000,000个核酸分子、至少约100,000,000个核酸分子、至少约250,000,000个核酸分子，并且在一些情况下，至少约10亿个核酸分子或更多。给定珠粒的核酸分子可包括相同(或共同)条形码序列、不同条形码序列或两者组合。给定珠粒的核酸分子可包括多组核酸分子。给定组的核酸分子可包括相同条形码序列。相同条形码序列可不同于另一组的核酸分子的条形码序列。

[0139] 此外，当对珠粒群体进行分配时，所得分区群体也可包括多样性条形码文库，其包括至少约1,000个不同条形码序列、至少约5,000个不同条形码序列、至少约10,000个不同条形码序列、至少约50,000个不同条形码序列、至少约100,000个不同条形码序列、至少约1,000,000个不同条形码序列、至少约5,000,000个不同条形码序列、或至少约10,000,000个不同条形码序列。另外，群体的各分区可包括至少约1,000个核酸分子、至少约5,000个核酸分子、至少约10,000个核酸分子、至少约50,000个核酸分子、至少约100,000个核酸分子、至少约500,000个核酸、至少约1,000,000个核酸分子、至少约5,000,000个核酸分子、至少约10,000,000个核酸分子、至少约50,000,000个核酸分子、至少约100,000,000个核酸分子、至少约250,000,000个核酸分子，并且在一些情况下，至少约10亿个核酸分子。

[0140] 在一些情况下，可合乎需要的是在给定分区内并入多个不同条形码，连接于所述分区内的单一珠粒或多个珠粒。举例来说，在一些情况下，一组混合但已知的条形码序列可在后续处理中提供更大鉴定保证，例如通过对给定分区提供更强条形码地址或归属，作为对来自给定分区的输出信息的重复或独立确认。

[0141] 核酸分子(例如寡核苷酸)可在向珠粒施加特定刺激后从珠粒释放。在一些情况下，刺激可为光刺激，例如通过对光不稳定连键的裂解，所述裂解使核酸分子释放。在其他情况下，可使用热刺激，其中珠粒环境的温度的升高将导致连键的裂解或另外从珠粒释放核酸分子。在其他情况下，可使用使核酸分子与珠粒的连键裂解，或另外导致核酸分子从珠粒释放的化学刺激。在一个情况下，所述组合物包括用于囊封生物粒子的上述聚丙烯酰胺

基质，并且可通过暴露于还原剂诸如DTT而降解以释放所连接核酸分子。

[0142] 在一些方面，提供用于达成控制分配的系统和方法。小滴尺寸可通过调整通道构造(例如微流体通道构造)中的某些几何特征来控制。举例来说，可调整通道的扩展角、宽度和/或长度以控制小滴尺寸。

[0143] 图8显示在油包水乳液中，单个Jurkat细胞连同含有条形码寡核苷酸的珠粒一起在水性小滴中共分配的图像。如所说明，单个细胞可易于与单个珠粒共分配。如将了解，对单个细胞装载量的优化可通过许多方法来进行，包括通过向微流体系统中提供细胞群体稀释液以实现如在本文中其他地方所述的所需每个分区细胞装载量。

[0144] 在操作中，一旦溶解，单个细胞的核酸内含物即接着可用于在分区内进行进一步处理，包括例如碎裂、扩增和条形码化以及连接其他功能性序列。如上所指示，碎裂可通过共分配剪切酶诸如核酸内切酶以使核酸碎裂成较小片段来实现。这些核酸内切酶可包括限制核酸内切酶，包括II型和IIs型限制核酸内切酶，以及其他核酸裂解酶诸如切口核酸内切酶等。在一些情况下，可不需要碎裂，并且可使全长核酸保留在分区内，或在囊封细胞或细胞内含物的情况下，碎裂可在分配之前例如通过酶促方法例如本文所述的那些，或通过机械方法例如机械、声学或其他剪切来进行。

[0145] 一旦共分配，并且细胞被溶解以释放它们的核酸，安置在珠粒上的寡核苷酸即可用于使那些核酸的片段条形码化和扩增。关于使用这些条形码寡核苷酸使样品核酸的片段扩增和条形码化的一特别精致方法详述于2014年6月26日提交，并且以引用的方式并入本文的美国专利公布号US 2014/0378345中。简要来说，在一个方面，存在于与细胞共分配的珠粒上的寡核苷酸从它们的珠粒释放至具有细胞的核酸的分区中。寡核苷酸可连同条形码序列一起在它的5’末端包括引物序列。这个引物序列可为意图随机引发细胞的核酸上众多不同区域的随机寡核苷酸序列，或它可为在细胞的基因组的特定所靶向区域的上游进行靶向引发的特异性引物序列。

[0146] 一旦释放，寡核苷酸的引物部分即可退火于细胞的核酸的互补性区域。也与细胞和珠粒共分配的延伸反应试剂例如DNA聚合酶、三磷酸核苷、辅因子(例如$Mg^{2+}$或$Mn^{2+}$)，接着使用细胞的核酸作为模板来使引物序列延伸，以产生细胞的由引物所退火的核酸链的互补性片段，所述互补性片段包括寡核苷酸和它的相关条形码序列。多个引物退火于细胞的核酸的不同部分以及延伸将产生核酸的重叠互补性片段的大型汇合物，所述片段各自具有它的指示它被创建所处的分区的自身条形码序列。在一些情况下，这些互补性片段可自身用作由存在于分区中的寡核苷酸引发的模板以产生所述互补物的再次包括条形码序列的互补物。在一些情况下，配置这个复制过程以使当第一互补物被复制时，它在它的末端处或附近产生两个互补序列，以允许形成使分子成为产生进一步迭代拷贝的基础的能力降低的发夹结构或部分发夹结构。如本文所述，细胞的核酸可包括细胞内的任何所需核酸，包括例如细胞的DNA例如基因组DNA、RNA例如信使RNA等。举例来说，在一些情况下，本文所述的方法和系统用于表征所表达mRNA，包括所述mRNA的存在和定量，并且可包括RNA测序过程作为表征过程。或者或另外，连同细胞一起分配的试剂可包括用于使mRNA转变成cDNA的试剂，例如逆转录酶和试剂，以有助于其中采用DNA测序的测序过程。在一些情况下，当待表征的核酸包含RNA例如mRNA时，这个表征的一个实例的示意性说明显示于图3中。

[0147] 如所示，包括条形码序列的寡核苷酸连同样品核酸304一起在例如处于乳液中的

小滴302中共分配。如在本文中其他地方所指示，寡核苷酸308可被提供在与样品核酸304共分配的珠粒306上，所述寡核苷酸可从珠粒306释放，如图版A中所示。除一个或多个功能性序列例如序列310、314和316之外，寡核苷酸308也包括条形码序列312。举例来说，寡核苷酸308显示为包含条形码序列312以及可充当给定测序系统的连接或固定序列的序列310，例如用于在Illumina Hiseq®或Miseq®系统的流动槽中连接的P5序列。如所示，寡核苷酸也包括引物序列316，其可包括用于引发样品核酸304的各部分的复制的随机或靶向N聚体。也包括在寡核苷酸308内的是可提供测序引发区域诸如“读段1”或R1引发区域的序列314，所述测序引发区域用于引发在测序系统中通过合成反应来达成的聚合酶介导的模板引导测序。如将了解，可选择功能性序列以可与多种不同测序系统例如454测序、Ion Torrent Proton或PGM、Illumina X10等及其要求相容。在许多情况下，条形码序列312、固定序列310和R1序列314可为连接于给定珠粒的所有寡核苷酸所共有。引物序列316对于随机N聚体引物可变化，或可为给定珠粒上的寡核苷酸所共有以达成某些靶向应用。

[0148] 如将了解，在一些情况下，功能性序列可包括适用于RNA-seq应用的引物序列。举例来说，在一些情况下，对于RNA-seq，寡核苷酸可包括用于引发RNA的逆转录的多dT引物。在其他情况下，除共同条形码序列之外，给定分区中的例如包括在单个珠粒上的寡核苷酸也可包括多种类型的引物序列，诸如DNA测序引物与RNA测序引物两者，例如包括在偶联于珠粒的寡核苷酸内的多dT引物序列。在所述情况下，可使单一经分配细胞经受DNA测序过程与RNA测序过程两者。

[0149] 基于引物序列316的存在，寡核苷酸可如图版B中所示来引发样品核酸，此允许使用也与珠粒306和样品核酸304共分配的聚合酶和其他延伸试剂来使寡核苷酸308和308a延伸。如图版C中所示，在对于随机N聚体引物来说将退火于样品核酸304的多个不同区域的寡核苷酸的延伸之后；创建核酸的多个重叠互补物或片段，例如片段318和320。尽管包括互补于样品核酸的各部分的序列部分例如序列322和324，但这些构建体通常在本文中称为包含样品核酸304的片段，具有所连接条形码序列。

[0150] 可接着使条形码化核酸片段经受表征，例如通过序列分析，或它们可在过程中加以进一步扩增，如图版D中所示。举例来说，也从珠粒306释放的额外寡核苷酸例如寡核苷酸308b可引发片段318和320。这对于片段318加以显示。特别地，再次，基于寡核苷酸308b中随机N聚体引物316b(其在许多情况下可不同于给定分区中的其他随机N聚体例如引物序列316)的存在，寡核苷酸与片段318退火，并且延伸以创建片段318的至少一部分的互补物326，其包括包含样品核酸序列的一部分的重复的序列328。寡核苷酸308b的延伸继续直至它已复制通过片段318的寡核苷酸部分308。如在本文中其他地方所指示，以及如图版D中所说明，寡核苷酸可被配置来促使在由聚合酶达成的复制中在所需地点处终止，例如在复制通过包括在片段318内的寡核苷酸308的序列316和314之后。如本文所述，这可通过不同方法来实现，包括例如并入不能够由所用聚合酶处理的不同核苷酸和/或核苷酸类似物。举例来说，这可包括在序列区域312内包括含有尿嘧啶的核苷酸以阻止非尿嘧啶耐受性聚合酶来停止那个区域的复制。因此，创建片段326，其在一个末端包括全长寡核苷酸308b，包括条形码序列312、连接序列310、R1引物区域314和随机N聚体序列316b。在序列的另一末端，可包括第一寡核苷酸308的随机N聚体的互补物316’，以及R1序列的全部或一部分的互补物，显示为序列314’。R1序列314和它的互补物314’接着能够杂交在一起以形成部分发夹结构

328。如将了解，因为随机N聚体在不同寡核苷酸之间不同，所以这些序列和它们的互补物预期将不参与发夹形成，例如是随机N聚体316的互补物的序列316’预期将不互补于随机N聚体序列316b。这对于其他应用将不是事实，例如靶向引物，其中N聚体在给定分区内的寡核苷酸之间将是共同的。

[0151] 通过形成这些部分发夹结构，允许从进一步复制移除样品序列的一级重复，例如防止对各拷贝的迭代拷贝。部分发夹结构也提供适用于对所创建片段例如片段326的后续处理的结构。

[0152] 一般来说，进行细胞的核酸的扩增直至分区内条形码化重叠片段构成细胞的基因组的特定部分或全部的至少1X覆盖、目标基因组或它的相关部分的至少2X、至少3X、至少4X、至少5X、至少10X、至少20X、至少40X或更大覆盖。一旦产生条形码化片段，它们即可直接在适当测序系统例如Illumina Hiseq®、Miseq®或X10系统上测序，或可使它们经受额外处理，诸如进一步扩增、连接其他功能性序列例如用于反向读段的第二测序引物、样品索引序列等。

[0153] 可接着汇合来自多个不同分区的所有片段以在如本文所述的高通量测序仪上测序，其中汇合片段包含源于不同细胞或小型细胞群体的核酸的许多片段，但其中来自给定细胞的核酸的片段将共有相同条形码序列。特别地，因为各片段关于它的来源分区，并且因此关于它的单一细胞或小型细胞群体加以编码，所以那个片段的序列可基于条形码的存在来向回归属于那个细胞或那些细胞，此也将有助于将来自多个分区的各种序列片段应用于对不同细胞的单个基因组的组装。这示意性地说明于图4中。如一个实例中所示，来自第一细胞400的第一核酸404和来自第二细胞402的第二核酸406各自连同它们自身的各组条形码寡核苷酸一起如上所述加以分配。核酸可包含染色体、整个基因组或来自细胞的其他大型核酸。

[0154] 在各分区内，各细胞的核酸404和406接着被处理以单独提供第一片段的重叠第二片段组，例如第二片段组408和410。这个处理也提供具有对于源于特定第一片段的第二片段中的各者是相同的条形码序列的第二片段。如所示，第二片段组408的条形码序列由“1”表示，而片段组410的条形码序列由“2”表示。多样性条形码文库可用于以差异性方式使大量不同片段组条形码化。然而，对于来自不同第一片段的每个第二片段组，不必要用不同条形码序列加以条形码化。实际上，在许多情况下，可并行处理多个不同第一片段以包括相同条形码序列。多样性条形码文库在本文中其他地方详细描述。

[0155] 可接着汇合例如来自片段组408和410的条形码化片段以使用例如通过可从Illumina或Thermo-Fisher,Inc.的Ion Torrent部门获得的合成技术来获得的序列进行测序。一旦测序，序列读段412可至少部分地基于包括的条形码，并且在一些情况下，部分地基于片段自身的序列来归属于它们的相应片段组，例如如以聚集读段414和416所示。接着组装各片段组的归属序列读段以提供各细胞的核酸的组装序列，例如序列418和420，其转而可归属于单个细胞例如细胞400和402。

[0156] 尽管以分析存在于细胞内的遗传物质加以描述，但通过允许向单个细胞分配试剂，以及提供对响应于那些试剂的那些细胞的可归属分析或表征，本文所述的方法和系统可具有广泛得多的适用性，包括能够表征单个细胞或细胞群体的其他方面。这些方法和系统在能够出于例如研究、诊断、病原体鉴定和许多其他目的来表征细胞方面特别具有价值。

举例来说，广泛范围的不同细胞表面特征例如细胞表面蛋白如分化簇或CD蛋白在表征疾病如癌症方面具有重大诊断相关性。

[0157] 在一个特别适用应用中，本文所述的方法和系统可用于表征细胞特征，诸如细胞表面特征，例如蛋白质、受体等。特别地，本文所述的方法可用于使报道体分子连接于这些细胞特征，所述报道体分子在如上所述加以分配时可被条形码化，并且例如使用DNA测序技术来分析，以探明存在性，并且在一些情况下，探明单个细胞或细胞群体内所述细胞特征的相对丰度或数量。

[0158] 在一特定实例中，可提供与第一组核酸报道体分子缔合的潜在细胞结合配体例如抗体、抗体片段、细胞表面受体结合分子等的文库，例如其中不同报道体寡核苷酸序列与特定配体缔合，因此能够结合特定细胞表面特征。在一些方面，文库的不同成员可通过不同寡核苷酸序列标记的存在性来表征，例如针对第一类型的细胞表面蛋白或受体的抗体将具有与它缔合的第一已知报道体寡核苷酸序列，而针对第二受体蛋白质的抗体将具有与它缔合的不同已知报道体寡核苷酸序列。在共分配之前，将使细胞与可代表抗体的配体文库一起孵育，所述抗体针对一组广泛不同细胞表面特征例如受体、蛋白质等，并且包括它们的缔合报道体寡核苷酸。将未结合配体从细胞洗涤，并且细胞接着连同上述条形码寡核苷酸一起共分配。因此，分区将包括一个或多个细胞，以及所结合配体和它们的已知缔合报道体寡核苷酸。

[0159] 在不需要使分区内的细胞溶解下，可接着使报道体寡核苷酸经受以上对于细胞核酸所述的条形码化操作，以产生条形码化报道体寡核苷酸，其中存在报道体寡核苷酸可指示存在特定细胞表面特征，并且条形码序列将允许基于与那个细胞或细胞群体共分配的条形码序列来将不同细胞表面特征的范围归属于给定单个细胞或细胞群体。因此，可产生较广泛细胞群体内细胞表面特征的逐一细胞概况。本文所述的方法和系统的这个方面在以下更详细描述。

[0160] 这个实例示意性地说明于图5中。如所示，使由细胞502和504表示的细胞群体与细胞表面相关试剂例如抗体、细胞表面结合蛋白、配体等的文库一起孵育，其中各不同类型的结合基团包括与它缔合的缔合核酸报道体分子，显示为配体和缔合报道体分子506、508、510和512(其中报道体分子由差异阴影化圆圈指示)。当细胞表达由文库结合的表面特征时，配体和它们的缔合报道体分子可变得与细胞表面缔合或偶联。接着将单个细胞连同它们的缔合配体/报道体分子以及如在本文中其他地方所述的单个条形码寡核苷酸珠粒例如珠粒522和524一起分别分配至单独分区例如小滴514和516中。如同本文所述的其他实例一样，条形码化寡核苷酸从珠粒释放，并且用于使条形码序列连接于存在于各分区内的报道体分子，其中条形码为给定分区所共有，但其在不同分区之间广泛变化。举例来说，如图5中所示，在分区514中，与细胞502缔合的报道体分子用条形码序列518加以条形码化，而在分区516中，与细胞504缔合的报道体分子用条形码520加以条形码化。因此，提供反映细胞的表面配体的寡核苷酸文库，所述表面配体如由报道体分子所反映，但借助于共同条形码序列而大致上可归属于单个细胞，从而允许对细胞的表面特征进行单一细胞水平剖析。如将了解，这个过程不限于细胞表面受体，而是可用于鉴定广泛多种特定细胞结构、化学物质或其他特征的存在。

[0161] III.单细胞分析的应用

[0162] 存在本文所述的单细胞处理和分析方法和系统的广泛多种不同应用，包括分析特定单个细胞、分析不同细胞类型的群体内的不同细胞类型、分析和表征用于环境、人类健康、流行病学、法医应用或广泛多种不同应用中的任一者的大型细胞群体。

[0163] 本文所述的单细胞分析方法的一特别有价值的应用是在对癌细胞测序和表征方面。特别地，常规分析技术，包括以上提到的全体测序方法，不高度擅长于抓取癌细胞的基因组组成的小变化，特别是当那些小变化存在于大量正常组织细胞中时。此外，即使在肿瘤细胞之间，广泛变化也可存在，并且可由于全体测序方法而掩蔽(参见例如Patel等，Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma,Science DOI:10.1126/science.1254257(2014年6月12日在线发表))。癌细胞可源于实体肿瘤、血液恶性肿瘤、细胞系或以循环肿瘤细胞形式获得，并且经受上述分配方法。在分析后，可将单个细胞序列鉴定为源于单一细胞或小型细胞群组，并且将那些序列与正常组织细胞序列区分。此外，如2015年6月26日提交的，其全部公开内容据此出于所有目的以引用的方式整体并入本文的共同待决的美国专利申请号14/752,589(US20150376700)中所述，也可获得来自各细胞的分段序列信息，从而允许更明确表征癌细胞内的单倍型变体。单细胞分析方法特别适用于涉及低量输入核酸的系统和方法，如2015年6月26日提交的，其全部公开内容据此出于所有目的以引用的方式整体并入本文的共同待决的美国专利申请号14/752,602(US20150376605)中所述。

[0164] 如同癌细胞分析一样，通过使用常规技术分析胎儿细胞来分析和诊断胎儿健康或异常是一项艰难任务。特别地，在不存在相对侵袭性程序诸如羊膜穿刺下，获得胎儿细胞样品可采用从母性循环收集那些细胞。如将了解，所述循环胎儿细胞组成那个循环的总体细胞群体的极小部分。因此，进行复合分析以表征与母性细胞相对比，有什么所得数据可能源于胎儿细胞。然而，通过采用本文所述的单细胞表征方法和系统，可将遗传组成归属于单个细胞，并且基于它们的相应遗传组成来将那些细胞分类为母性或胎儿。此外，胎儿细胞的遗传序列可用于鉴定许多遗传病症中的任一者，包括例如非整倍性诸如唐综合征(Down syndrome)、爱德华综合征(Edwards syndrome)和巴陶综合征(Patau syndrome)。

[0165] 能够从较大多样性细胞群体表征单个细胞也在环境测试方面以及在法医分析方面均具有重大价值，其中样品就它们的性质来说，相对于对样品进行测试所针对的细胞，可由多样性细胞群体和使样品“污染”的其他物质组成，所述细胞例如用于例如环境和食物安全性测试的环境指示生物体、毒性生物体等、针对性侵犯和其他暴力犯罪的法医分析中的受害者和/或犯罪者细胞等。

[0166] 上述单细胞测序和表征方法的额外适用应用在神经科学研究和诊断领域。特别地，神经细胞可包括长散布核元件(LINE)或可在基因组周围移动的‘跳跃’基因，其导致各神经元不同于它的近邻细胞。研究已显示人脑中的LINE的数目超过其他组织例如心脏和肝组织的LINE的数目，具有80个与300个之间的独特插入(参见例如Coufal,N.G.等Nature 460,1127-1131(2009))。这些差异已被假定为与个人对神经病症的易感性相关联(参见例如Muotri,A.R.等Nature 468,443-446(2010))，或对脑提供对激发起响应所具有的多样性。因此，本文所述的方法可用于对单个神经细胞的测序和表征。

[0167] 本文所述的单细胞分析方法也适用于分析基因表达，如上所指示，在鉴定RNA转录物与它们的定量两方面。特别地，使用本文所述的单细胞水平分析方法，可分离和分析存在

于单个细胞、细胞群体或细胞群体的子组中的RNA转录物。特别地，在一些情况下，条形码寡核苷酸可被配置来引发、复制以及因此产生来自单个细胞的RNA的条形码化片段。举例来说，在一些情况下，条形码寡核苷酸可包括mRNA特异性引发序列，例如允许在逆转录反应中引发和复制mRNA的多dT引物区段或其他靶向引发序列(例如基因特异性引物)。或者或另外，可使用条形码寡核苷酸的随机N聚体引物区段来进行随机RNA引发。

[0168] 图6提供用于使用本文所述的方法来进行单个细胞中的RNA表达分析的一个实例方法的示意图。如所示，在操作602时，分选含有细胞的样品中的活细胞，将所述活细胞定量和稀释以用于后续分配。在操作604时，将单个细胞单独与如本文所述的携带条形码化寡核苷酸的凝胶珠粒共分配。在操作606时，细胞被溶解，并且条形码化寡核苷酸释放至分区中，在操作608时，它们在分区中与mRNA相互作用并杂交于mRNA，例如借助于互补于mRNA的多A尾部的多dT引物序列。在操作610时，使用多dT条形码寡核苷酸作为引发序列，利用本文所述的经工程改造逆转录酶来进行逆转录反应以合成mRNA的包括条形码序列的cDNA转录物。接着在操作612时，例如使用PCR方法来使条形码化cDNA转录物经受额外扩增，在操作614时纯化，随后将它们放置在核酸测序系统上以测定cDNA序列和它的缔合条形码序列。在一些情况下，如所示，操作602至608可在试剂保持在它们的原始小滴或分区中时发生，而操作612至616可散开发生(例如在分区的外部)。在其中分区是乳液中的小滴的情况下，可破坏所述乳液，并且汇合所述小滴的内含物以完成操作612至616。在一些情况下，条形码寡核苷酸可在破坏乳液之后用核酸外切酶消化。在引物消化之后，核酸外切酶活性可用乙二胺四乙酸(EDTA)抑制。在一些情况下，操作610可基于逆转录混合物例如逆转录酶和相关试剂的共分配来在分区内进行，或它可散开进行。

[0169] 如在本文中其他地方所指示，除寡核苷酸条形码序列之外，条形码寡核苷酸的结构也可包括许多序列元件。用于如上所述的RNA分析中的条形码寡核苷酸的一个实例显示于图7中。如所示，总体寡核苷酸702由可释放连键706诸如二硫接头偶联于珠粒704。寡核苷酸可包括用于后续处理中的功能性序列，诸如功能性序列708，其可包括以下中的一者或多者：测序仪特异性流动槽连接序列例如用于Illumina测序系统的P5序列、以及测序引物序列例如用于Illumina测序系统的R1引物。条形码序列710包括在结构内以用于使样品RNA条形码化。mRNA特异性引发序列诸如多dT序列712也包括在寡核苷酸结构中。可包括锚定序列区段714以确保多dT序列在mRNA的序列末端杂交。这个锚定序列可包括随机短核苷酸序列例如1聚体、2聚体、3聚体或更长序列，其将确保多dT区段更可能在mRNA的多A尾部的序列末端杂交。另一序列区段716可被提供在寡核苷酸序列内。在一些情况下，这个另一序列提供独特分子序列区段，例如作为在偶联于单一珠粒的单个寡核苷酸之间变化的随机序列(例如像随机N聚体序列)，而条形码序列710可在栓系于单个珠粒的寡核苷酸之间恒定。这个独特序列用于提供所捕集的起始mRNA分子的独特标识符，以允许定量原始表达RNA的数目。如将了解，尽管显示为栓系于珠粒的表面的单一寡核苷酸，但单个珠粒可包括数万至数十万或甚至数百万个单个寡核苷酸分子，其中如所指示，对于给定珠粒，条形码区段可为恒定的或相对恒定的，但其中可变或独特序列区段将跨越单个珠粒而变化。在寡核苷酸的序列内，这个独特分子序列区段可包括5至约8个或更多个核苷酸。在一些情况下，独特分子序列区段的长度可为2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸或更长。在一些情况下，独特分子序列区段的长度可为至少2、3、4、5、6、7、8、9、10、11、12、13、14、15、

16、17、18、19或20个核苷酸或更长。在一些情况下，独特分子序列区段的长度可为至多2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸或更短。

[0170] 在操作中，以及参照图6和图7，将细胞连同携带条形码的珠粒一起共分配并且加以溶解，同时条形码化寡核苷酸从珠粒释放。经释放条形码寡核苷酸的多dT部分接着杂交于mRNA的多A尾部。多dT区段接着引发mRNA的逆转录以产生mRNA的cDNA转录物，但所述cDNA转录物包括条形码寡核苷酸的序列区段708-716中的各者。再次，因为寡核苷酸702包括锚定序列714，所以它将更可能杂交于mRNA的多A尾部的序列末端，并且在所述序列末端处引发逆转录。在任何给定分区内，单个mRNA分子的所有cDNA转录物都将包括共同条形码序列区段710。然而，通过包括独特随机N聚体序列，由给定分区内的不同mRNA分子制得的转录物将在这个独特序列处变化。这提供可甚至在对给定分区的内含物的任何后续扩增之后可鉴定的定量特征，例如与共同条形码缔合的独特区段的数目可指示来源于单一分区，因此来源于单一细胞的mRNA的数量。如上所指示，转录物接着被扩增，净化以及测序以鉴定mRNA的cDNA转录物的序列，以及对条形码区段和独特序列区段测序。

[0171] 如在本文中其他地方所指示，尽管描述的是多dT引物序列，但其他靶向或随机引发序列也可用于引发逆转录反应。同样，尽管描述为条形码化寡核苷酸连同经溶解细胞的内含物一起释放至分区中，但应了解，在一些情况下，凝胶珠粒结合的寡核苷酸可用于在凝胶珠粒的固相上杂交和捕集mRNA，以有助于使RNA与其他细胞内含物分离。

[0172] 用于包括信使RNA(mRNA，包括从细胞获得的mRNA)分析的RNA分析中的条形码寡核苷酸的另一实例显示于图9A中。如所示，总体寡核苷酸902可由可释放连键906诸如二硫接头偶联于珠粒904。寡核苷酸可包括用于后续处理中的功能性序列，诸如功能性序列908，其可包括测序仪特异性流动槽连接序列例如用于Illumina测序系统的P5序列，以及功能性序列910，其可包括测序引物序列例如用于Illumina测序系统的R1引物结合位点。条形码序列912包括在结构内以用于使样品RNA条形码化。RNA特异性(例如mRNA特异性)引发序列诸如多dT序列914也包括在寡核苷酸结构中。可包括锚定序列区段(未显示)以确保多dT序列在mRNA的序列末端杂交。另一序列区段916可被提供在寡核苷酸序列内。这个另一序列可提供独特分子序列区段，例如作为在偶联于单一珠粒的单个寡核苷酸之间变化的随机N聚体序列，而条形码序列912可在栓系于单个珠粒的寡核苷酸之间恒定。如在本文中其他地方所述，这个独特序列可用于提供所捕集的起始mRNA分子的独特标识符，以允许定量原始表达RNA的数目，例如mRNA计数。如将了解，尽管显示为栓系于珠粒的表面的单一寡核苷酸，但单个珠粒可包括数万至数十万或甚至数百万个单个寡核苷酸分子，其中如所指示，对于给定珠粒，条形码区段可为恒定的或相对恒定的，但其中可变或独特序列区段将跨越单个珠粒而变化。

[0173] 在细胞RNA(例如mRNA)分析的一实例方法中，以及参照图9A，将细胞连同携带条形码的珠粒、转换寡聚物924以及其他试剂诸如逆转录酶、还原剂和dNTP一起共分配至分区(例如乳液中的小滴)中。在操作950中，细胞被溶解，同时条形码化寡核苷酸902从珠粒释放(例如通过还原剂的作用)，并且经释放条形码寡核苷酸的多dT区段914接着杂交于从细胞释放的mRNA 920的多A尾部。接着，在操作952中，使用mRNA作为模板，多dT区段914在逆转录反应中被延伸以产生互补于mRNA的cDNA转录物922，并且也包括条形码寡核苷酸的序列区段908、912、910、916和914中的各者。逆转录酶的末端转移酶活性可将额外碱基(例如多C)

添加至cDNA转录物中。转换寡聚物924可接着与添加至cDNA转录物中的额外碱基杂交，并且有助于模板转换。互补于转换寡聚物序列的序列可接着通过使用转换寡聚物924作为模板达成的cDNA转录物922的延伸来被并入cDNA转录物922中。在任何给定分区内，单个mRNA分子的所有cDNA转录物都将包括共同条形码序列区段912。然而，通过包括独特随机N聚体序列916，由给定分区内的不同mRNA分子制得的转录物将在这个独特序列处变化。如在本文中其他地方所述，这提供可甚至在对给定分区的内含物的任何后续扩增之后可鉴定的定量特征，例如与共同条形码缔合的独特区段的数目可指示来源于单一分区，因此来源于单一细胞的mRNA的数量。在操作952之后，在操作954中，cDNA转录物922接着用引物926(例如PCR引物)扩增。接着，在操作956中，接着纯化(例如通过固相可逆固定(SPRI))经扩增产物。在操作958时，接着剪切经扩增产物，连接于额外功能性序列，并且进一步扩增(例如通过PCR)。功能性序列可包括测序仪特异性流动槽连接序列930例如用于Illumina测序系统的P7序列，以及功能性序列928，其可包括例如用于Illumina测序系统的R2引物的测序引物结合位点，以及功能性序列932，其可包括样品索引，例如用于Illumina测序系统的i7样品索引序列。在一些情况下，操作950和952可在分区中发生，而操作954、956和958可在散开溶液中(例如在分区的外部的汇合混合物中)发生。在其中分区是乳液中的小滴的情况下，可破坏所述乳液，并且汇合所述小滴的内含物以完成操作954、956和958。在一些情况下，操作954可在分区中完成。在一些情况下，条形码寡核苷酸可在破坏乳液之后用核酸外切酶消化。在引物消化之后，核酸外切酶活性可用乙二胺四乙酸(EDTA)抑制。尽管以用于某些测序系统例如Illumina系统的特定序列提及物加以描述，但应了解对这些序列的提及仅出于说明目的，并且本文所述的方法可被配置来与并有用于那些系统中的特定引发、连接、索引和其他操作序列的其他测序系统一起使用，所述其他测序系统例如可从Ion Torrent、Oxford Nanopore、Genia、Pacific Biosciences、Complete Genomics等获得的系统。

[0174] 在用于如图9A中所示的RNA(例如细胞RNA)分析中的条形码寡核苷酸的一替代性实例中，功能性序列908可为P7序列，并且功能性序列910可为R2引物结合位点。此外，功能性序列930可为P5序列，功能性序列928可为R1引物结合位点，并且功能性序列932可为用于Illumina测序系统的i5样品索引序列。由这种条形码寡核苷酸产生的构建体的构型可帮助使在测序期间对多dT序列的测序最小化(或避免)。

[0175] 图9B中显示用于包括细胞mRNA分析的RNA分析的另一实例方法。在这个方法中，连同试剂诸如逆转录酶、还原剂和dNTP一起将转换寡聚物924与单个细胞和条形码化珠粒共分配至分区(例如乳液中的小滴)中。转换寡聚物924可用另一标签934例如生物素标记。在操作951中，细胞被溶解，同时条形码化寡核苷酸902(例如如图9A中所示)从珠粒释放(例如通过还原剂的作用)。在一些情况下，序列908是P7序列，并且序列910是R2引物结合位点。在其他情况下，序列908是P5序列，并且序列910是R1引物结合位点。接着，经释放条形码寡核苷酸的多dT区段914杂交于从细胞释放的mRNA920的多A尾部。在操作953中，使用mRNA作为模板，多dT区段914接着在逆转录反应中被延伸以产生互补于mRNA的cDNA转录物922，并且也包括条形码寡核苷酸的序列区段908、912、910、916和914中的各者。逆转录酶的末端转移酶活性可将额外碱基(例如多C)添加至cDNA转录物中。转换寡聚物924可接着与cDNA转录物杂交，并且有助于模板转换。互补于转换寡聚物序列的序列可接着通过使用转换寡聚物924作为模板达成的cDNA转录物922的延伸来被并入cDNA转录物922中。接着，分离操作960可用

于使cDNA转录物922与分区中的试剂和寡核苷酸分离。可使另一标签934例如生物素与可连接于磁性珠粒938的相互作用标签936例如链霉亲和素接触。在操作960时，cDNA可用下拉操作(例如通过磁性分离、离心)来分离，随后在操作955中扩增(例如通过PCR)，继之以在操作957中纯化(例如通过固相可逆固定(SPRI))，并且在操作959中进一步处理(剪切、连接序列928、932和930、以及后续扩增(例如通过PCR))。在其中序列908是P7序列，并且序列910是R2引物结合位点的一些情况下，序列930是P5序列，并且序列928是R1引物结合位点，并且序列932是i5样品索引序列。在其中序列908是P5序列，并且序列910是R1引物结合位点的一些情况下，序列930是P7序列，并且序列928是R2引物结合位点，并且序列932是i7样品索引序列。在一些情况下，如所示，操作951和953可在分区中发生，而操作960、955、957和959可在散开溶液中(例如在分区的外部的汇合混合物中)发生。在其中分区是乳液中的小滴的情况下，可破坏所述乳液，并且汇合所述小滴的内含物以完成操作960。操作955、957和959可接着在操作960之后在使转录物汇合以进行处理之后进行。

[0176] 图9C中显示用于包括细胞mRNA分析的RNA分析的另一实例方法。在这个方法中，连同试剂诸如逆转录酶、还原剂和dNTP一起将转换寡聚物924与单个细胞和条形码化珠粒共分配于分区(例如乳液中的小滴)中。在操作961中，细胞被溶解，同时条形码化寡核苷酸902(例如如图9A中所示)从珠粒释放(例如通过还原剂的作用)。在一些情况下，序列908是P7序列，并且序列910是R2引物结合位点。在其他情况下，序列908是P5序列，并且序列910是R1引物结合位点。接着，经释放条形码寡核苷酸的多dT区段914接着杂交于从细胞释放的mRNA 920的多A尾部。接着，在操作963中，使用mRNA作为模板，多dT区段914接着在逆转录反应中被延伸以产生互补于mRNA的cDNA转录物922，并且也包括条形码寡核苷酸的序列区段908、912、910、916和914中的各者。逆转录酶的末端转移酶活性可将额外碱基(例如多C)添加至cDNA转录物中。转换寡聚物924可接着与cDNA转录物杂交，并且有助于模板转换。互补于转换寡聚物序列的序列可接着通过使用转换寡聚物924作为模板达成的cDNA转录物922的延伸来被并入cDNA转录物922中。在操作961和操作963之后，在操作962中，使mRNA 920和cDNA转录物922变性。在操作964时，第二链从具有另一标签942例如生物素的引物940延伸，并且杂交于cDNA转录物922。也在操作964中，可使经生物素标记第二链与可连接于磁性珠粒938的相互作用标签936例如链霉亲和素接触。cDNA可用下拉操作(例如通过磁性分离、离心)来分离，随后在操作965中扩增(例如通过聚合酶链式反应(PCR))，继之以在操作967中纯化(例如通过固相可逆固定(SPRI))，并且在操作969中进一步处理(剪切、连接序列928、932和930、以及后续扩增(例如通过PCR))。在其中序列908是P7序列，并且序列910是R2引物结合位点的一些情况下，序列930是P5序列，并且序列928是R1引物结合位点，并且序列932是i5样品索引序列。在其中序列908是P5序列，并且序列910是R1引物结合位点的一些情况下，序列930是P7序列，并且序列928是R2引物结合位点，并且序列932是i7样品索引序列。在一些情况下，操作961和963可在分区中发生，而操作962、964、965、967和969可散开发生(例如在分区的外部)。在其中分区是乳液中的小滴的情况下，可破坏所述乳液，并且汇合所述小滴的内含物以完成操作962、964、965、967和969。

[0177] 图9D中显示用于包括细胞mRNA分析的RNA分析的另一实例方法。在这个方法中，连同试剂诸如逆转录酶、还原剂和dNTP一起将转换寡聚物924与单个细胞和条形码化珠粒共分配。在操作971中，细胞被溶解，同时条形码化寡核苷酸902(例如如图9A中所示)从珠粒释

放(例如通过还原剂的作用)。在一些情况下,序列908是P7序列,并且序列910是R2引物结合位点。在其他情况下,序列908是P5序列,并且序列910是R1引物结合位点。接着,经释放条形码寡核苷酸的多dT区段914接着杂交于从细胞释放的mRNA 920的多A尾部。接着,在操作973中,使用mRNA作为模板,多dT区段914接着在逆转录反应中被延伸以产生互补于mRNA的cDNA转录物922,并且也包括条形码寡核苷酸的序列区段908、912、910、916和914中的各者。逆转录酶的末端转移酶活性可将额外碱基(例如多C)添加至cDNA转录物中。转换寡聚物924可接着与cDNA转录物杂交,并且有助于模板转换。互补于转换寡聚物序列的序列可接着通过使用转换寡聚物924作为模板达成的cDNA转录物922的延伸来被并入cDNA转录物922中。在操作966中,可使mRNA 920、cDNA转录物922和转换寡聚物924变性,并且可使cDNA转录物922与用另一标签946例如生物素标记的捕集寡核苷酸944杂交。在这个操作时,可使杂交于cDNA转录物的经生物素标记捕集寡核苷酸944与可连接于磁性珠粒938的相互作用标签936例如链霉亲和素接触。在使用下拉操作(例如通过磁性分离、离心)与其他物质(例如过量条形码化寡核苷酸)分离之后,在操作975时,cDNA转录物可用引物926扩增(例如通过PCR),继之以在操作977中纯化(例如通过固相可逆固定(SPRI)),并且在操作979中进一步处理(剪切、连接序列928、932和930、以及后续扩增(例如通过PCR))。在其中序列908是P7序列,并且序列910是R2引物结合位点的一些情况下,序列930是P5序列,并且序列928是R1引物结合位点,并且序列932是i5样品索引序列。在其中序列908是P5序列,并且序列910是R1引物结合位点的其他情况下,序列930是P7序列,并且序列928是R2引物结合位点,并且序列932是i7样品索引序列。在一些情况下,操作971和973可在分区中发生,而操作966、975、977(纯化)和979可散开发生(例如在分区的外部)。在其中分区是乳液中的小滴的情况下,可破坏所述乳液,并且汇合所述小滴的内含物以完成操作966、975、977和979。

[0178] 图9E中显示用于包括细胞RNA分析的RNA分析的另一实例方法。在这个方法中,将单个细胞连同携带条形码的珠粒、转换寡聚物990和其他试剂诸如逆转录酶、还原剂和dNTP一起共分配至分区(例如乳液中的小滴)中。在操作981中,细胞被溶解,同时条形码化寡核苷酸(例如如图9A中所示的902)从珠粒释放(例如通过还原剂的作用)。在一些情况下,序列908是P7序列,并且序列910是R2引物结合位点。在其他情况下,序列908是P5序列,并且序列910是R1引物结合位点。接着,经释放条形码寡核苷酸的多dT区段接着杂交于从细胞释放的mRNA 920的多A尾部。接着,在操作983时,多dT区段接着在逆转录反应中被延伸以产生互补于mRNA的cDNA转录物922,并且也包括条形码寡核苷酸的序列区段908、912、910、916和914中的各者。逆转录酶的末端转移酶活性可将额外碱基(例如多C)添加至cDNA转录物中。转换寡聚物990可接着与cDNA转录物杂交,并且有助于模板转换。互补于转换寡聚物序列,并且包括T7启动子序列的序列可被并入cDNA转录物922中。在操作968时,合成第二链,并且在操作970时,T7启动子序列可由T7聚合酶用于在体外转录中产生RNA转录物。在操作985时,RNA转录物可被纯化(例如通过固相可逆固定(SPRI)),逆转录以形成DNA转录物,并且可合成各DNA转录物的第二链。在一些情况下,在纯化之前,可使RNA转录物与DNA酶(例如DNA酶I)接触以使残余DNA分解。在操作987时,接着使DNA转录物碎裂,并且连接于额外功能性序列诸如序列928、932和930,并且在一些情况下,加以进一步扩增(例如通过PCR)。在其中序列908是P7序列,并且序列910是R2引物结合位点的一些情况下,序列930是P5序列,并且序列928是R1引物结合位点,并且序列932是i5样品索引序列。在其中序列908是P5序列,并且序列

910是R1引物结合位点的一些情况下，序列930是P7序列，并且序列928是R2引物结合位点，并且序列932是i7样品索引序列。在一些情况下，在移除DNA转录物的一部分之前，可使DNA转录物与RNA酶接触以使残余RNA分解。在一些情况下，操作981和983可在分区中发生，而操作968、970、985和987可散开发生(例如在分区的外部)。在其中分区是乳液中的小滴的情况下，可破坏所述乳液，并且汇合所述小滴的内含物以完成操作968、970、985和987。

[0179] 用于包括信使RNA(mRNA，包括从细胞获得的mRNA)分析的RNA分析中的条形码寡核苷酸的另一实例显示于图10中。如所示，总体寡核苷酸1002由可释放连键1006诸如二硫接头偶联于珠粒1004。寡核苷酸可包括用于后续处理中的功能性序列，诸如功能性序列1008，其可包括测序仪特异性流动槽连接序列例如P7序列，以及功能性序列1010，其可包括测序引物序列例如R2引物结合位点。条形码序列1012包括在结构内以用于使样品RNA条形码化。RNA特异性(例如mRNA特异性)引发序列诸如多dT序列1014可包括在寡核苷酸结构中。可包括锚定序列区段(未显示)以确保多dT序列在mRNA的序列末端杂交。另一序列区段1016可被提供在寡核苷酸序列内。这个另一序列可提供如在本文中其他地方所述的独特分子序列区段。可包括另一功能性序列1020以用于体外转录，例如T7 RNA聚合酶启动子序列。如将了解，尽管显示为栓系于珠粒的表面的单一寡核苷酸，但单个珠粒可包括数万至数十万或甚至数百万个单个寡核苷酸分子，其中如所指示，对于给定珠粒，条形码区段可为恒定的或相对恒定的，但其中可变或独特序列区段将跨越单个珠粒而变化。

[0180] 在细胞RNA分析的一实例方法中，以及参照图10，将细胞连同携带条形码的珠粒和其他试剂诸如逆转录酶、还原剂和dNTP一起共分配至分区(例如乳液中的小滴)中。在操作1050中，细胞被溶解，同时条形码化寡核苷酸1002从珠粒释放(例如通过还原剂的作用)，并且经释放条形码寡核苷酸的多dT区段1014接着杂交于mRNA 1020的多A尾部。接着，在操作1052时，使用mRNA作为模板，多dT区段接着在逆转录反应中被延伸以产生mRNA的cDNA转录物1022，并且也包括条形码寡核苷酸的序列区段1020、1008、1012、1010、1016和1014中的各者。在任何给定分区内，单个mRNA分子的所有cDNA转录物都将包括共同条形码序列区段1012。然而，通过包括独特随机N聚体序列，由给定分区内的不同mRNA分子制得的转录物将在这个独特序列处变化。如在本文中其他地方所述，这提供可甚至在对给定分区的内含物的任何后续扩增之后可鉴定的定量特征，例如与共同条形码缔合的独特区段的数目可指示来源于单一分区，因此来源于单一细胞的mRNA的数量。在操作1054时，合成第二链，并且在操作1056时，T7启动子序列可由T7聚合酶用于在体外转录中产生RNA转录物。在操作1058时，使转录物碎裂(例如剪切)，连接于额外功能性序列，并且加以逆转录。功能性序列可包括测序仪特异性流动槽连接序列1030例如P5序列，以及功能性序列1028，其可包括测序引物例如R1引物结合序列，以及功能性序列1032，其可包括样品索引例如i5样品索引序列。在操作1060时，RNA转录物可被逆转录成DNA，扩增DNA(例如通过PCR)，并且测序以鉴定mRNA的cDNA转录物的序列，以及对条形码区段和独特序列区段测序。在一些情况下，操作1050和1052可在分区中发生，而操作1054、1056、1058和1060可散开发生(例如在分区的外部)。在其中分区是乳液中的小滴的情况下，可破坏所述乳液，并且汇合所述小滴的内含物以完成操作1054、1056、1058和1060。

[0181] 在用于如图10中所示的RNA(例如细胞RNA)分析中的条形码寡核苷酸的一替代性实例中，功能性序列1008可为P5序列，并且功能性序列1010可为R1引物结合位点。此外，功

能性序列1030可为P7序列，功能性序列1028可为R2引物结合位点，并且功能性序列1032可为i7样品索引序列。

[0182] 用于包括信使RNA(mRNA，包括从细胞获得的mRNA)分析的RNA分析中的条形码寡核苷酸的另一实例显示于图11中。如所示，总体寡核苷酸1102由可释放连键1106诸如二硫接头偶联于珠粒1104。寡核苷酸可包括用于后续处理中的功能性序列，诸如功能性序列1108，其可包括测序仪特异性流动槽连接序列例如P5序列，以及功能性序列1110，其可包括测序引物序列例如R1引物结合位点。在一些情况下，序列1108是P7序列，并且序列1110是R2引物结合位点。条形码序列1112包括在结构内以用于使样品RNA条形码化。另一序列区段1116可被提供在寡核苷酸序列内。在一些情况下，这个另一序列可提供如在本文中其他地方所述的独特分子序列区段。可包括另一序列1114以有助于模板转换，例如多G。如将了解，尽管显示为栓系于珠粒的表面的单一寡核苷酸，但单个珠粒可包括数万至数十万或甚至数百万个单个寡核苷酸分子，其中如所指示，对于给定珠粒，条形码区段可为恒定的或相对恒定的，但其中可变或独特序列区段将跨越单个珠粒而变化。

[0183] 在细胞mRNA分析的一实例方法中，以及参照图11，将细胞连同携带条形码的珠粒、多dT序列以及其他试剂诸如逆转录酶、还原剂和dNTP一起共分配至分区(例如乳液中的小滴)中。在操作1150中，细胞被溶解，同时条形码化寡核苷酸从珠粒释放(例如通过还原剂的作用)，并且多dT序列杂交于从细胞释放的mRNA 1120的多A尾部。接着，在操作1152中，使用mRNA作为模板，多dT序列接着在逆转录反应中被延伸以产生互补于mRNA的cDNA转录物1122。逆转录酶的末端转移酶活性可将额外碱基(例如多C)添加至cDNA转录物中。添加至cDNA转录物中的额外碱基例如多C可接着与条形码化寡核苷酸的1114杂交。这可有助于模板转换，并且互补于条形码寡核苷酸的序列可被并入cDNA转录物中。转录物可被进一步处理(例如扩增、移除各部分、添加额外序列等)以及如在本文中其他地方所述加以表征，例如通过测序。由这种方法产生的构建体的构型可帮助使在测序期间对多dT序列的测序最小化(或避免)。

[0184] 用于包括细胞RNA分析的RNA分析中的条形码寡核苷酸的另一实例显示于图12A中。如所示，总体寡核苷酸1202由可释放连键1206诸如二硫接头偶联于珠粒1204。寡核苷酸可包括用于后续处理中的功能性序列，诸如功能性序列1208，其可包括测序仪特异性流动槽连接序列例如P5序列，以及功能性序列1210，其可包括测序引物序列例如R1引物结合位点。在一些情况下，序列1208是P7序列，并且序列1210是R2引物结合位点。条形码序列1212包括在结构内以用于使样品RNA条形码化。另一序列区段1216可被提供在寡核苷酸序列内。在一些情况下，这个另一序列可提供如在本文中其他地方所述的独特分子序列区段。如将了解，尽管显示为栓系于珠粒的表面的单一寡核苷酸，但单个珠粒可包括数万至数十万或甚至数百万个单个寡核苷酸分子，其中如所指示，对于给定珠粒，条形码区段可为恒定的或相对恒定的，但其中可变或独特序列区段将跨越单个珠粒而变化。在使用这个条形码进行细胞RNA分析的一实例方法中，将细胞连同携带条形码的珠粒和其他试剂诸如RNA连接酶和还原剂一起共分配至分区(例如乳液中的小滴)中。细胞被溶解，同时条形码化寡核苷酸从珠粒释放(例如通过还原剂的作用)。条形码化寡核苷酸可接着在分区中通过RNA连接酶来连接于mRNA转录物的5’末端。后续操作可包括纯化(例如通过固相可逆固定(SPRI))以及进一步处理(剪切、连接功能性序列和后续扩增(例如通过PCR))，并且这些操作可散开发生

(例如在分区的外部)。在其中分区是乳液中的小滴的情况下,可破坏所述乳液,并且汇合所述小滴的内含物以用于额外操作。

[0185] 用于包括细胞RNA分析的RNA分析中的条形码寡核苷酸的另一实例显示于图12B中。如所示,总体寡核苷酸1222由可释放连键1226诸如二硫接头偶联于珠粒1224。寡核苷酸可包括用于后续处理中的功能性序列,诸如功能性序列1228,其可包括测序仪特异性流动槽连接序列例如P5序列,以及功能性序列1230,其可包括测序引物序列例如R1引物结合位点。在一些情况下,序列1228是P7序列,并且序列1230是R2引物结合位点。条形码序列1232包括在结构内以用于使样品RNA条形码化。引发序列1234(例如随机引发序列)也可包括在寡核苷酸结构中,例如随机六聚体。另一序列区段1236可被提供在寡核苷酸序列内。在一些情况下,这个另一序列提供如在本文中其他地方所述的独特分子序列区段。如将了解,尽管显示为栓系于珠粒的表面的单一寡核苷酸,但单个珠粒可包括数万至数十万或甚至数百万个单个寡核苷酸分子,其中如所指示,对于给定珠粒,条形码区段可为恒定的或相对恒定的,但其中可变或独特序列区段将跨越单个珠粒而变化。在使用图12B的条形码寡核苷酸进行细胞mRNA分析的一实例方法中,将细胞连同携带条形码的珠粒和额外试剂诸如逆转录酶、还原剂和dNTP一起共分配至分区(例如乳液中的小滴)中。细胞被溶解,同时条形码化寡核苷酸从珠粒释放(例如通过还原剂的作用)。在一些情况下,序列1228是P7序列,并且序列1230是R2引物结合位点。在其他情况下,序列1228是P5序列,并且序列1230是R1引物结合位点。随机六聚体引发序列1234可随机杂交细胞mRNA。使用来自细胞的mRNA作为模板,随机六聚体序列可接着在逆转录反应中被延伸以产生互补于mRNA的cDNA转录物,并且也包括条形码寡核苷酸的序列区段1228、1232、1230、1236和1234中的各者。后续操作可包括产生扩增产物、纯化(例如通过固相可逆固定(SPRI))、进一步处理(例如剪切、连接功能性序列和后续扩增(例如通过PCR))。这些操作可散开发生(例如在分区的外部)。在其中分区是乳液中的小滴的情况下,可破坏所述乳液,并且汇合所述小滴的内含物以用于额外操作。可连同携带条形码的珠粒一起共分配的额外试剂可包括待用以阻断核糖体RNA(rRNA)的寡核苷酸和用以消化来自细胞的基因组DNA的核酸酶。或者,rRNA移除剂可在额外处理操作期间施加。由这种方法产生的构建体的构型可帮助使在测序期间对多dT序列的测序最小化(或避免)和/或对多核苷酸序列的5'末端测序。扩增产物例如第一扩增产物和/或第二扩增产物可经受测序以进行序列分析。在一些情况下,扩增可使用用于测序的部分发夹扩增(PHASE)方法来进行。本文所述的单细胞分析方法也可适用于全转录物组的分析中。回头参照图12B的条形码,引发序列1234可为随机N聚体。在一些情况下,序列1228是P7序列,并且序列1230是R2引物结合位点。在其他情况下,序列1228是P5序列,并且序列1230是R1引物结合位点。在使用这个条形码进行全转录物组分析的一实例方法中,将单个细胞连同携带条形码的珠粒、多dT序列以及其他试剂诸如逆转录酶、聚合酶、还原剂和dNTP一起共分配至分区(例如乳液中的小滴)中。在这个方法的一操作中,细胞被溶解,同时条形码化寡核苷酸从珠粒释放(例如通过还原剂的作用),并且多dT序列杂交于细胞mRNA的多A尾部。在使用mRNA作为模板进行的逆转录反应中,可产生细胞mRNA的cDNA转录物。RNA可接着用RNA酶降解。条形码化寡核苷酸中的引发序列1234可接着随机杂交于cDNA转录物。类似于如图3中所示,寡核苷酸可使用与珠粒和细胞共分配的聚合酶和其他延伸试剂来延伸以产生扩增产物(例如条形码化片段),类似于图3中所示的实例扩增产物(图版F)。在一些情况下,可使条形码化核酸片段经

受进一步处理(例如扩增、添加额外序列、净化过程等,如在本文中其他地方所述),例如通过序列分析来表征。在这个操作中,测序信号可来自全长RNA。

[0186] 尽管已单独地讨论用各种条形码设计进行的操作,但单个珠粒可包括供同时使用的具有各种设计的条形码寡核苷酸。

[0187] 除表征单个细胞或来自较大群体的细胞子群体之外,本文所述的方法和系统也可用于表征单个细胞作为一种用以提供细胞或其他生物体群体的总体概况的方式。多种应用需要评估细胞群体内不同细胞或生物体类型的存在以及对其定量,包括例如微生物群系分析和表征、环境测试、食物安全性测试、流行病学分析,例如在追踪污染方面等。特别地,上述分析方法可用于对群体内的大量单个细胞进行单独表征、测序和/或鉴定。这个表征可接着用于组装来源群体的总体概况,此可提供重要预后和诊断信息。

[0188] 举例来说,包括例如消化道、口颊、表皮微生物群系等的人微生物群系的转变已被鉴定为可对不同疾患或一般性健康状态进行诊断和预后。使用本文所述的单细胞分析方法和系统,可再次对总群体中的单个细胞进行表征、测序和鉴定,并且鉴定那个群体内可指示诊断相关因素的转变。举例来说,对细菌16S核糖体RNA基因的测序已作为一种高度准确方法用于对细菌的分类学分类。使用上述靶向扩增和测序方法可提供对细胞群体内单个细胞的鉴定。可进一步定量群体内不同细胞的数目以鉴定当前状态或状态随时间的转变。参见例如Morgan等,PLoS Comput.Biol.,第12章,2012年12月,8(12):e1002808以及Ram等,Syst.Biol.Reprod.Med.,2011年6月,57(3):162-170,其各自出于所有目的以引用的方式整体并入本文。同样,对感染或潜在感染的鉴定和诊断也可受益于本文所述的单细胞分析,例如以鉴定存在于其他细胞或其他生物物质、细胞和/或核酸的大型混合物中,包括存在于上述环境以及任何其他诊断相关环境例如脑脊髓液、血液、粪便或肠样品等中的微生物物种。

[0189] 先前分析也可特别适用于通过对不同抗性标志物/突变跨越给定样品中细胞群体的分布分析和剖析来表征不同细胞例如癌细胞、细菌性病原体等的潜在药物抗性。另外,对这些标志物/突变跨越细胞群体随时间的转变的表征可提供对特征在于所述药物抗性问题的多种疾病的进展、改变、预防和治疗的有价值见地。

[0190] 尽管以细胞加以描述,但应了解多种单个生物有机体或生物体的组分中的任一者都涵盖在本描述内,包括例如细胞、病毒、细胞器、细胞包涵物、囊泡等。另外,当提及细胞时,应了解所述提及包括任何类型的细胞,包括不限于原核细胞、真核细胞、细菌、真菌、植物、哺乳动物或其他动物细胞类型、支原体、正常组织细胞、肿瘤细胞或任何其他细胞类型,无论是源于单细胞还是多细胞生物体。

[0191] 类似地,分析不同环境样品以剖析存在于所述样品内的微生物有机体、病毒或其他生物污染物可提供关于疾病流行病学的重要信息,并且潜在有助于预测疾病爆发、流行病和大流行病。

[0192] 如上所述,本文所述的方法、系统和组合物也可用于分析和表征单个细胞或细胞群体的其他方面。在一个实例方法中,提供含有待关于它们的细胞表面蛋白加以分析和表征的细胞的样品。也提供抗体、抗体片段或对细胞被表征所针对的细胞表面蛋白或抗原(或其他细胞特征)具有结合亲和力的其他分子的文库(在本文中也称为细胞表面特征结合基团)。为便于讨论,这些亲和基团在本文中被称为结合基团。结合基团可包括指示结合基团

与其结合的细胞表面特征的报道体分子。特别地，对一种类型的细胞表面特征具有特异性的结合基团类型将包含第一报道体分子，而对不同细胞表面特征具有特异性的结合基团类型将具有与它缔合的不同报道体分子。在一些方面，这些报道体分子将包含寡核苷酸序列。基于寡核苷酸的报道体分子提供能够在序列方面产生显著多样性，同时也可易于连接于大多数生物分子例如抗体等，以及例如使用测序或阵列技术而易于检测的优势。在实例方法中，结合基团包括连接于它们的寡核苷酸。因此，第一结合基团类型例如针对第一类型的细胞表面特征的抗体将具有与它缔合的具有第一核苷酸序列的报道体寡核苷酸。不同结合基团类型例如对其他不同细胞表面特征具有结合亲和力的抗体将具有与其缔合的包含不同核苷酸序列，例如具有部分或完全不同核苷酸序列的报道体寡核苷酸。在一些情况下，对于各类型的细胞表面特征结合基团例如抗体或抗体片段，报道体寡核苷酸序列可为已知的，并且可易于鉴定为与已知细胞表面特征结合基团缔合。这些寡核苷酸可直接偶联于结合基团，或它们可连接于珠粒，分子晶格例如线性、球状、交联或其他聚合物，或与结合基团连接或另外与结合基团缔合的其他框架，此允许多个报道体寡核苷酸连接于单一结合基团。

[0193] 在多个报道体分子偶联于单一结合基团的情况下，所述报道体分子可包含相同序列，或特定结合基团将包括一组已知报道体寡核苷酸序列。在例如对不同细胞表面特征具有特异性的不同结合基团之间，报道体分子可不同以及可归属于特定结合基团。

[0194] 报道基团与结合基团的连接可通过多种直接或间接共价或非共价缔合或连接中的任一者来实现。举例来说，在寡核苷酸报道基团与基于抗体的结合基团缔合的情况下，可使用化学缀合技术(例如可从Innova Biosciences获得的Lightning-Link®抗体标记试剂盒)来使所述寡核苷酸共价连接于抗体或抗体片段的一部分，以及使用其他非共价连接机理，例如采用亲和素或链霉亲和素接头，使用生物素化抗体和寡核苷酸(或偶联于寡核苷酸的包括一个或多个生物素化接头的珠粒)。抗体和寡核苷酸生物素化技术是可用的(参见例如Fang等，Fluoride-Cleavable Biotinylation Phosphoramidite for 5′-end-Labeling and Affinity Purification of Synthetic Oligonucleotides，Nucleic Acids Res.2003年1月15日;31(2):708-715;可从Thermo Scientific获得的DNA 3’末端生物素化试剂盒;以及可从Thermo Fisher Scientific获得的SiteClick™抗体标记系统，其全部公开内容出于所有目的以引用的方式整体并入本文)。同样，蛋白质和肽生物素化技术已被开发，并且可易于获得(参见例如美国专利号6,265,552，其全部公开内容出于所有目的以引用的方式整体并入本文)。

[0195] 视给定分析所需的报道体分子的多样性、采用的序列检测流程等而定，可提供具有任何不同长度范围的报道体寡核苷酸。在一些情况下，这些报道体序列的长度可大于约5个核苷酸、大于约10个核苷酸、大于约20、30、40、50、60、70、80、90、100、120、150个或甚至200个核苷酸。在一些情况下，这些报道体核苷酸的长度可小于约250个核苷酸、小于约200、180、150、120 100、90、80、70、60、50、40个或甚至30个核苷酸。在许多情况下，可选择报道体寡核苷酸以提供已确定大小，并且另外被配置来在测序系统上加以分析的条形码化产物。举例来说，这些序列可以理想地创建具有为特定测序系统所需的长度的可测序产物的长度提供。同样，除报道体序列之外，这些报道体寡核苷酸也可包括额外序列元件，诸如测序仪连接序列、测序引物序列、扩增引物序列、或这些序列元件中的任一者的互补物。

[0196] 在操作中，对于需要分析的任何细胞表面特征，使含有细胞的样品与结合分子和

它们的缔合报道体寡核苷酸一起孵育。在孵育之后，洗涤细胞以移除未结合结合基团。在洗涤之后，将细胞连同上述携带条形码的珠粒一起分配至单独分区例如小滴中，其中各分区包括有限数目的细胞，例如在一些情况下，包括单一细胞。在使条形码从珠粒释放后，它们将引发报道体寡核苷酸的扩增和条形码化。如上所指示，报道体分子的条形码化复制物可另外包括功能性序列诸如引物序列、连接序列等。

[0197] 接着使条形码化报道体寡核苷酸经受序列分析以鉴定哪些报道体寡核苷酸结合于分区内的细胞。此外，通过也对缔合条形码序列测序，可鉴定给定细胞表面特征与报道体序列包括相同条形码序列的其他不同细胞表面特征可能来自同一细胞，即它们源于同一分区。以基于条形码序列的存在而发源于单个分区的报道体分子为基础，可接着创建来自细胞群体的单个细胞的细胞表面概况。可将单个细胞或细胞群体的概况与来自其他细胞例如‘正常’细胞的概况进行比较，以鉴定细胞表面特征的变化，此可提供诊断相关信息。特别地，这些概况可特别适用于诊断特征在于细胞表面受体的变化的多种病症，诸如癌症和其他病症。

[0198] IV.经工程改造逆转录酶变体

[0199] cDNA合成反应中的一个主要挑战是cDNA合成中来自RNA二级结构的干扰。尽管较高反应温度可从模板RNA移除二级结构，但在不使用高效热稳定RT酶的情况下，高温通常导致较低逆转录酶(RT)酶活性。另外，抑制剂诸如细胞溶解产物和相关试剂可降低RT酶活性。

[0200] 野生型(WT)莫洛尼鼠白血病病毒(MMLV)逆转录酶是一种通常在较高温度下失活的RT酶。然而，若干可商购获得的突变MMLV RT酶表现出改进的热稳定性、保真度、底物亲和力和/或降低的末端脱氧核苷酸基转移酶活性。

[0201] 在一些实施方案中，本文公开包含与SEQ ID NO:3至少80％同一的氨基酸序列的经工程改造逆转录酶，其中所述氨基酸序列包含：(a)相对于SEQ ID NO:3，从N末端进行的至少15个氨基酸的截短；和(b)一个或多个选自由以下组成的组的突变：相对于SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。在一些情况下，(b)中的一个或多个突变选自由以下组成的组：相对于SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0202] 在一些实施方案中，经工程改造逆转录酶在所述N末端或在所述氨基酸序列的C末端进一步包含亲和标签。在一些情况下，所述亲和标签包括但不限于白蛋白结合蛋白(ABP)、AU1表位、AU5表位、T7标签、V5标签、B标签、氯霉素乙酰基转移酶(CAT)、二氢叶酸还原酶(DHFR)、AviTag、钙调蛋白标签、聚谷氨酸标签、E标签、FLAG标签、HA标签、Myc标签、NE标签、S标签、SBP标签、Doftag 1、Softag 3、Spot标签、四半胱氨酸(TC)标签、Ty标签、VSV标签、Xpress标签、生物素羧基载体蛋白(BCCP)、绿色荧光蛋白标签、HaloTag、Nus标签、硫氧还蛋白标签、Fc标签、纤维素结合结构域、几丁质结合蛋白(CBP)、胆碱结合结构域、半乳糖结合结构域、麦芽糖结合蛋白(MBP)、辣根过氧化物酶(HRP)、Strep标签、HSV表位、酮类固醇异构酶(KSI)、KT3表位、LacZ、荧光素酶、PDZ结构域、PDZ配体、聚精氨酸(Arg标签)、聚天冬氨酸(Asp标签)、聚半胱氨酸(Cys标签)、聚苯丙氨酸(Phe标签)、Profinity eXact、蛋白C、S1标签、S1标签、葡萄球菌蛋白A(蛋白A)、葡萄球菌蛋白G(蛋白G)、小泛素样修饰物(SUMO)、

串联亲和纯化(TAP)标签、TrpE、泛素、通用标签、谷胱甘肽-S-转移酶(GST)和聚(His)标签。在一些情况下,所述亲和标签是至少5个组氨酸氨基酸。

[0203] 在一些实施方案中,经工程改造逆转录酶进一步包含蛋白酶裂解序列,其中由蛋白酶对所述蛋白酶裂解序列的裂解导致所述亲和标签从所述经工程改造逆转录酶的裂解。在一些情况下,蛋白酶裂解序列是由包括但不限于以下的蛋白酶识别的蛋白酶裂解序列:丙氨酸羧基肽酶、蜜环菌虾红素(Armillaria mellea astacin)、细菌亮氨酰基氨基肽酶、癌促凝物质、组织蛋白酶B(cathepsin B)、梭菌蛋白酶(clostripain)、胞质液丙氨酰基氨基肽酶、弹性蛋白酶(elastase)、胞内蛋白酶Arg-C(endoproteinase Arg-C)、肠激酶(enterokinase)、胃亚蛋白酶(gastricsin)、明胶酶(gelatinase)、Gly-X羧基肽酶、甘氨酰基内肽酶、人鼻病毒3C蛋白酶、皮蝇素C(hypodermin C)、Iga特异性丝氨酸内肽酶、亮氨酰基氨基肽酶、亮氨酰基内肽酶、lysC、溶酶体pro-X羧基肽酶、赖氨酰基氨基肽酶、甲硫氨酰基氨基肽酶、粘球菌(myxobacter)、苯乙肼裂解酶(nardilysin)、胰腺内肽酶E、细小核糖核酸病毒内肽酶2A(picornain 2A)、细小核糖核酸病毒内肽酶3C、内肽酶原(proendopeptidase)、脯氨酰基氨基肽酶、原蛋白转化酶I、原蛋白转化酶II、拉塞尔溶素(russellysin)、糖胃蛋白酶(saccharopepsin)、semenogelase、T-纤维蛋白溶酶原活化因子、凝血酶、组织激肽释放酶、烟草蚀刻病毒(TEV)、批膜病毒素(togavirin)、色氨酰基氨基肽酶、U-纤维蛋白溶酶原活化因子、V8、蛇毒凝血酶样酶A(venombin A)、蛇毒凝血酶样酶AB和Xaa-pro氨基肽酶。在一些情况下,所述蛋白酶裂解序列是凝血酶裂解序列。

[0204] 在一些实施方案中,本文公开经工程改造逆转录酶变体,其包含与氨基酸(多肽)序列SEQ ID NO:3至少80%同一的氨基酸序列,其中所述氨基酸序列的特征在于具有以下中的两者或更多者:(i)从所述氨基酸序列的N末端进行的至少15个氨基酸的截短;(ii)位于所述氨基酸序列的所述N末端的具有至少5个组氨酸氨基酸的序列;(iii)凝血酶裂解识别位点;和(iv)一个或多个选自由以下组成的组的突变:相对于SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。在一些实施方案中,本文公开经工程改造逆转录酶变体,其包含与氨基酸(多肽)序列SEQ ID NO:3至少80%同一的氨基酸序列,其中所述氨基酸序列的特征在于具有以下中的两者或更多者:(i)从所述氨基酸序列的N末端进行的至少15个氨基酸的截短;(ii)位于所述氨基酸序列的所述N末端的具有至少5个组氨酸氨基酸的序列;(iii)凝血酶裂解识别位点;和(iv)一个或多个选自由以下组成的组的突变:相对于SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0205] 野生型MMLV表达含1738个氨基酸的多肽链(参见例如UniProt P03355),其由病毒蛋白酶p14加工成许多成熟蛋白质,包括野生型MMLV p80逆转录酶(参见例如SEQ ID NO:3)。

[0206] 表1.序列

[0207]

| | |
|---|---|
| SEQ ID NO: 1 | ATGGGTAGCTCACATCACCATCATCATCATTCTTCTG<br>GTCTGGTCCCACGCGGCAGCACTTGGCTGTCTGATTT<br>CCCTCAGGCGTGGGCCGAAACGGGTGGCATGGGTCT<br>GGCAGTGCGTCAGGCACCGCTGATTATTCCGCTGAA<br>AGCGACGTCGACCCCGGTGAGCATCAAGCAATATCC<br>GATGTCCCAAGAGGCGCGCTTAGGTATTAAGCCGCA<br>CATTCAGCGTCTGCTGGATCAAGGTATTCTGGTTCCG<br>TGTCAGAGCCCGTGGAATACCCCGCTTCTCCCGGTGA<br>AGAAACCGGGCACGAACGATTACCGTCCAGTCCAAG<br>ACTTGCGCGAAGTTAACAAGCGCGTTGAAGATATTC<br>ACCCGACCGTCCCGAACCCGTACAATCTGCTGAGCG<br>GTCTGCCGCCAAGCCACCAATGGTACACCGTGCTGG<br>ATCTGAAAGATGCTTTCTTCTGTCTGCGTCTGCACCC<br>AACCAGCCAGCCTCTGTTTGCATTTGAGTGGCGTGAC<br>CCTGAGATGGGTATTAGCGGCCAGCTGACGTGGACC<br>CGCCTGCCGCAAGGTTTTAAGAATTCCCCTACGCTGT<br>TTGACGAAGCGCTGCACCGTGACCTGGCGGATTTCCG<br>TATCCAGCACCCGGACCTGATCTTGCTGCAGTACGTT<br>GATGACCTGTTGCTGGCGGCGACGAGCGAGCTGGAT<br>TGCCAACAGGGCACCCGTGCGCTGTTGCAGACCTTG<br>GGTAACCTGGGTTATCGCGCTAGCGCGAAGAAAGCG<br>CAGATTTGCCAAAAACAAGTTAAGTATCTGGGCTAC<br>CTGTTAAAGGAAGGCCAACGTTGGCTGACCGAAGCC<br>CGCAAAGAAACTGTCATGGGTCAGCCGACCCCGAAA<br>ACGCCACGCCAACTGCGTGAGTTCTTGGGCACCGCG<br>GGTTTCTGCCGCCTGTGGATCCCGGGCTTTGCCGAAA<br>TGGCAGCCCCGCTGTATCCGTTGACCAAGACCGGCA<br>CCCTGTTCAACTGGGGTCCGGACCAGCAGAAAGCGT<br>ACCAAGAAATTAAACAAGCACTGCTGACGGCACCGG<br>CGCTGGGTCTGCCGGACCTGACCAAGCCGTTTGAGCT<br>GTTCGTGGATGAGAAGCAAGGTTACGCGAAGGGCGT<br>GTTGACCCAGAAATTGGGTCCGTGGCGTCGTCCGGTT<br>GCATACCTGTCCAAGAAACTGGACCCGGTTGCTGCTG<br>GTTGGCCGCCTTGCCTGCGCATGGTTGCCGCTATCGC<br>GGTGCTGACTAAAGACGCGGGTAAGCTGACGATGGG<br>TCAACCGCTGGTGATCAAGGCACCGCATGCAGTCGA<br>GGCCCTTGTTAAGCAACCGCCAGATAGATGGCTGAG<br>CAACGCGCGTATGACGCATTACCAGGCACTGCTGTTG<br>GACACCGATCGTGTGCAGTTTGGCCCGGTCGTTGCGC<br>TCAACCCGGCGACCCTGCTGCCGCTCCCGGAAGAAG<br>GCTTGCAGCACAACTGTTTGGACATCCTGGCAGAGG<br>CGCACGGCACTCGCCCGGATCTGACGGACCAGCCGC<br>TGCCGGACGCCGATCATACCTGGTATACGAATGGTA |

[0208]

| | |
|---|---|
| | GCAGCCTGTTGCAAGAGGGTCAGCGTAAGGCCGGTG<br>CCGCGGTCACCACCGAGACTGAAGTGATTTGGGCTA<br>AAGCATTGCCTGCGGGTACCAGCGCGCAGCGTGCCG<br>AGCTGATCGCACTGACCCAAGCGCTGAAAATGGCTG<br>AGGGTAAGAAACTGAATGTGTACACGGATAGCCGTT<br>ATGCCTTTGCGACCGCCCACATTCACGGCGAGATCTA<br>TCGCCGTCGCGGCCTGCTGACGTCCGAAGGCAAAGA<br>GATCAAGAATAAAGACGAAATTCTGGCGCTGCTGAA<br>AGCGCTGTTCCTGCCGAAACGTCTGTCGATCATCCAT<br>TGCCCGGGTCACCAGAAAGGCCACAGCGCAGAGGCG<br>CGTGGTAATCGCATGGCTGACCAGGCTGCGCGTAAA<br>GCCGCAATTACCGAAACCCCGGACACCAGCACGCTG<br>CTGATCGAGAATAGCAGCCCGAACAGCCGTCTGATC<br>AATTGATAA |
| SEQ ID NO: 2 | MGSSHHHHHHSSGLVPRGSTWLSDFPQAWAETGGMG<br>LAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQR<br>LLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREV<br>NKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFF<br>CLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN<br>SPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSE<br>LDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLG<br>YLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTA<br>GFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAY<br>QEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLT<br>QKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVL<br>TKDAGKLTMGQPLVIKAPHAVEALVKQPPDRWLSNAR<br>MTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQH<br>NCLDILAEAHGTRPDLTDQPLPDADHTWYTNGSSLLQE<br>GQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQA<br>LKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSE<br>GKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEA<br>RGNRMADQAARKAAITETPDTSTLLIENSSPNSRLIN |
| SEQ ID NO: 3 | TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGG<br>MGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHI<br>QRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLR<br>EVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDA<br>FFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF<br>KNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAAT<br>SELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKY<br>LGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLG<br>TAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQK<br>AYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGV<br>LTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIA<br>VLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSN<br>ARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQ<br>HNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQ<br>EGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQ<br>ALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTS |

[0209]

| | |
|---|---|
| | EGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAE<br>ARGNRMADQAARKAAITETPDTSTLL |
| SEQ ID NO: 4 | ATGGGTAGCTCACATCACCATCATCATCATTCTTCTG<br>GTCTGGTCCCACGCGGCAGCACTTGGCTGTCTGATTT<br>CCCTCAGGCGTGGGCCGAAACGGGTGGCATGGGTCT<br>GGCAGTGCGTCAGGCACCGCTGATTATTCCGCTGAA<br>AGCGACGTCGACCCCGGTGAGCATCAAGCAATATCC<br>GATGTCCCAAAAGGCGCGCTTAGGTATTAAGCCGCA<br>CATTCAGCGTCTGCTGGATCAAGGTATTCTGGTTCCG<br>TGTCAGAGCCCGTGGAATACCCCGCTTCTCCCGGTGA<br>AGAAACCGGGCACGAACGATTACCGTCCAGTCCAAG<br>ACTTGCGCGAAGTTAACAAGCGCGTTGAAGATATTC<br>ACCCGACCGTCCCGAACCCGTACAATCTGCTGAGCG<br>GTCCGCCGCCAAGCCACCAATGGTACACCGTGCTGG<br>ATCTGAAAGATGCTTTCTTCTGTCTGCGTCTGCACCC<br>AACCAGCCAGCCTCTGTTTGCATTTGAGTGGCGTGAC<br>CCTGAGATGGGTATTAGCGGCCAGCTGACGTGGACC<br>CGCCTGCCGCAAGGTTTTAAGAATTCCCCTACGCTGT<br>TTAACGAAGCGCTGCACCGTGACCTGGCGGATTTCCG<br>TATCCAGCACCCGGACCTGATCTTGCTGCAGTACGTT<br>GATGACCTGTTGCTGGCGGCGACGAGCGAGCTGGAT<br>TGCCAACAGGGCACCCGTGCGCTGTTGCAGACCTTG<br>GGTAACCTGGGTTATCGCGCTAGCGCGAAGAAAGCG<br>CAGATTTGCCAAAAACAAGTTAAGTATCTGGGCTAC<br>CTGTTAAAGGAAGGCCAACGTTGGCTGACCGAAGCC<br>CGCAAAGAAACTGTCATGGGTCAGCCGACCCCGAAA<br>ACGCCACGCCAACTGCGTAGGTTCTTGGGCAAAGCG<br>GGTTTCTGCCGCCTGTTCATCCCGGGCTTTGCCGAAA<br>TGGCAGCCCCGCTGTATCCGTTGACCAAGCCGGGCA<br>CCCTGTTCAACTGGGGTCCGGACCAGCAGAAAGCGT<br>ACCAAGAAATTAAACAAGCACTGCTGACGGCACCGG<br>CGCTGGGTCTGCCGGACCTGACCAAGCCGTTTGAGCT<br>GTTCGTGGATGAGAAGCAAGGTTACGCGAAGGGCGT<br>GTTGACCCAGAAATTGGGTCCGTGGCGTCGTCCGGTT<br>GCATACCTGTCCAAGAAACTGGACCCGGTTGCTGCTG<br>GTTGGCCGCCTTGCCTGCGCATGGTTGCCGCTATCGC<br>GGTGCTGACTAAAGACGCGGGTAAGCTGACGATGGG<br>TCAACCGCTGGTGATCAAGGCACCGCATGCAGTCGA<br>GGCCCTTGTTAAGCAACCGGCAGGCAGATGGCTGAG<br>CAAGGCGCGTATGACGCATTACCAGGCACTGCTGTT<br>GGACACCGATCGTGTGCAGTTTGGCCCGGTCGTTGCG<br>CTCAACCCGGCGACCCTGCTGCCGCTCCCGGAAGAA<br>GGCTTGCAGCACAACTGTTTGGACATCCTGGCAGAG<br>GCGCACGGCACTCGCCCGGATCTGACGGACCAGCCG<br>CTGCCGGACGCCGATCATACCTGGTATACGAATGGT<br>AGCAGCCTGTTGCAAGAGGGTCAGCGTAAGGCCGGT<br>GCCGCGGTCACCACCGAGACTGAAGTGATTTGGGCT<br>AAAGCATTGCCTGCGGGTACCAGCGCGCAGCGTGCC<br>GAGCTGATCGCACTGACCCAAGCGCTGAAAATGGCT<br>GAGGGTAAGAAACTGAATGTGTACACGGATAGCCGT |

[0210]

| | |
|---|---|
| | TATGCCTTTGCGACCGCCCACATTCACGGCGAGATCT<br>ATCGCCGTCGCGGCTGGCTGACGTCCAAAGGCAAAG<br>AGATCAAGAATAAAGACGAAATTCTGGCGCTGCTGA<br>AAGCGCTGTTCCTGCCGAAACGTCTGTCGATCATCCA<br>TTGCCCGGGTCACCAGAAAGGCCACAGCGCAGAGGC<br>GCGTGGTAATCGCATGGCTGACCAGGCTGCGCGTAA<br>AGCCGCAATTACCGAAACCCCGGACACCAGCACGCT<br>GCTGATCGAGAATAGCAGCCCGAACAGCCGTCTGAT<br>CAATTGATAA |
| SEQ ID NO: 5 | MGSSHHHHHHSSGLVPRGSTWLSDFPQAWAETGGMG<br>LAVRQAPLIIPLKATSTPVSIKQYPMSQKARLGIKPHIQR<br>LLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREV<br>NKRVEDIHPTVPNPYNLLSGPPPSHQWYTVLDLKDAFF<br>CLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN<br>SPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSE<br>LDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLG<br>YLLKEGQRWLTEARKETVMGQPTPKTPRQLRRFLGKA<br>GFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQ<br>EIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQ<br>KLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLT<br>KDAGKLTMGQPLVIKAPHAVEALVKQPAGRWLSKAR<br>MTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQH<br>NCLDILAEAHGTRPDLTDQPLPDADHTWYTNGSSLLQE<br>GQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQA<br>LKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTS<br>KGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAE<br>ARGNRMADQAARKAAITETPDTSTLLIENSSPNSRLIN |
| SEQ ID NO: 6 | TWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIK<br>QYPMSQKARLGIKPHIQRLLDQGILVPCQSPWNTPLLPV<br>KKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGP<br>PPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEM<br>GISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPD<br>LILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRA<br>SAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMG<br>QPTPKTPRQLRRFLGKAGFCRLFIPGFAEMAAPLYPLTK<br>PGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFE<br>LFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAA<br>GWPPCLRMVAAIAVLTKDAGKLTMGQPLVIKAPHAVE<br>ALVKQPAGRWLSKARMTHYQALLLDTDRVQFGPVVA<br>LNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPD<br>ADHTWYTNGSSLLQEGQRKAGAAVTTETEVIWAKALP<br>AGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFAT<br>AHIHGEIYRRRGWLTSKGKEIKNKDEILALLKALFLPKR<br>LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDT<br>STLL |

[0211] 如本文所用，当最优对准以进行比较时，术语“变体”可与某一多肽序列具有至少约45％、至少约50％、至少约55％、至少约60％、至少约65％、至少约70％、至少约75％、至少约80％、至少约85％、至少约88％、至少约90％、至少约91％、至少约92％、至少约93％、至少

约94％、至少约95％、至少约96％、至少约97％、至少约98％、至少约99％或至少约99.5％序列同一性。

[0212] 如本文所用，多肽与另一序列具有某一百分比(例如至少80％、至少85％、至少90％、至少95％或至少99％)的序列同一性意指当比对时，那个百分比的碱基或氨基酸残基在比较两个序列时是相同的。这个比对和同源性或同一性百分比可使用本领域中已知的任何适合软件程序来测定，例如CURRENT PROTOCOLS IN MOLECULAR BIOLOGY,Ausubel等人编,1987,补编30,章节7.7.18中所述的那些。代表性程序包括Vector NTI Advance™ 9.0(Invitrogen Corp.Carlsbad,CA)、GCG Pileup、FASTA(Pearson等(1988)Proc.Natl Acad.ScL USA 85:2444-2448)和BLAST(BLAST手册、Altschul等,Nat'l Cent.Biotechnol.Inf.,Nat'l Lib.Med.(NCIB NLM NIH),Bethesda,Md.以及Altschul等,(1997)Nucleic Acids Res.25:3389-3402)程序。另一典型比对程序是ALIGN Plus(Scientific and Educational Software,PA)，通常使用缺省参数。适用的另一序列软件程序是可在6.0版序列软件包(Sequence Software Package Version 6.0,Genetics Computer Group,University of Wisconsin,Madison,WI)中获得的TFASTA数据搜索程序。

[0213] 在一些实施方案中，经工程改造逆转录酶包含与WT MMLV RT酶(SEQ ID NO:3)至少80％同一的氨基酸序列。在一些实施方案中，经工程改造逆转录酶包含与根据SEQ ID NO:1的核苷酸序列至少80％同一的核苷酸序列。在一些实施方案中，经工程改造逆转录酶包含与SEQ ID NO:2至少80％同一的氨基酸序列。在一些实施方案中，经工程改造逆转录酶包含与根据SEQ ID NO.1的核苷酸序列至少80％、至少81％、至少82％、至少83％、至少84％、至少85％、至少86％、至少87％、至少88％、至少89％、至少90％、至少91％、至少92％、至少93％、至少94％、至少95％、至少96％、至少97％、至少98％或至少99％同一的核苷酸序列。在一些实施方案中，经工程改造逆转录酶包含与SEQ ID NO.2至少80％、至少81％、至少82％、至少83％、至少84％、至少85％、至少86％、至少87％、至少88％、至少89％、至少90％、至少91％、至少92％、至少93％、至少94％、至少95％、至少96％、至少97％、至少98％或至少99％同一的氨基酸序列。在一些实施方案中，经工程改造逆转录酶包含与SEQ ID NO.3至少80％、至少81％、至少82％、至少83％、至少84％、至少85％、至少86％、至少87％、至少88％、至少89％、至少90％、至少91％、至少92％、至少93％、至少94％、至少95％、至少96％、至少97％、至少98％或至少99％同一的氨基酸序列。

[0214] 在一些实施方案中，经工程改造逆转录酶包含在至少150个氨基酸残基的跨度内，与WT MMLV RT酶(SEQ ID NO:3)至少80％同一的氨基酸序列。在一些实施方案中，经工程改造逆转录酶包含在至少450个氨基酸残基的跨度内，与根据SEQ ID NO:1的核苷酸序列至少80％同一的核苷酸序列。在一些实施方案中，经工程改造逆转录酶包含在至少150个氨基酸残基的跨度内，与SEQ ID NO:2至少80％同一的氨基酸序列。在一些实施方案中，经工程改造逆转录酶包含在至少450个核苷酸的跨度内，与根据SEQ ID NO.1的核苷酸序列至少80％、至少81％、至少82％、至少83％、至少84％、至少85％、至少86％、至少87％、至少88％、至少89％、至少90％、至少91％、至少92％、至少93％、至少94％、至少95％、至少96％、至少97％、至少98％或至少99％同一的核苷酸序列。在一些实施方案中，经工程改造逆转录酶包含在至少150个氨基酸残基的跨度内，与SEQ ID NO.2至少80％、至少81％、至少82％、至少83％、至少84％、至少85％、至少86％、至少87％、至少88％、至少89％、至少90％、至少91％、

至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%同一的氨基酸序列。在一些实施方案中，经工程改造逆转录酶包含在至少150个氨基酸残基的跨度内，与SEQ ID NO.3至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%同一的氨基酸序列。

[0215] 在一些实施方案中，历经至少300个氨基酸残基的跨度，经工程改造逆转录酶包含与WT MMLV RT酶(SEQ ID NO:3)至少80%同一的氨基酸序列。在一些实施方案中，经工程改造逆转录酶包含在至少900个核苷酸的跨度内，与根据SEQ ID NO:1的核苷酸序列至少80%同一的核苷酸序列。在一些实施方案中，历经至少300个氨基酸残基的跨度，经工程改造逆转录酶包含与SEQ ID NO:2至少80%同一的氨基酸序列。在一些实施方案中，经工程改造逆转录酶包含在至少900个核苷酸的跨度内，与根据SEQ ID NO.1的核苷酸序列至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%同一的核苷酸序列。在一些实施方案中，经工程改造逆转录酶包含在至少300个氨基酸残基的跨度内，与SEQ ID NO.2至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%同一的氨基酸序列。在一些实施方案中，经工程改造逆转录酶包含在至少300个氨基酸残基的跨度内，与SEQ ID NO.3至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%同一的氨基酸序列。

[0216] 在一些实施方案中，经工程改造逆转录酶包含在至少450个氨基酸残基的跨度内，与WT MMLV RT酶(SEQ ID NO:3)至少80%同一的氨基酸序列。在一些实施方案中，经工程改造逆转录酶包含在至少1,350个核苷酸的跨度内，与根据SEQ ID NO:1的核苷酸序列至少80%同一的核苷酸序列。在一些实施方案中，经工程改造逆转录酶包含在至少450个氨基酸残基的跨度内，与SEQ ID NO:2至少80%同一的氨基酸序列。在一些实施方案中，经工程改造逆转录酶包含在至少1,350个核苷酸的跨度内，与根据SEQ ID NO.1的核苷酸序列至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%同一的核苷酸序列。在一些实施方案中，经工程改造逆转录酶包含在至少450个氨基酸残基的跨度内，与SEQ ID NO.2至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%同一的氨基酸序列。在一些实施方案中，经工程改造逆转录酶包含在至少450个氨基酸残基的跨度内，与SEQ ID NO.3至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%同一的氨基酸序列。

[0217] 在一些实施方案中，经工程改造逆转录酶包含在至少600个氨基酸残基的跨度内，与WT MMLV RT酶(SEQ ID NO:3)至少80%同一的氨基酸序列。在一些实施方案中，经工程改

造逆转录酶包含在至少1,800个核苷酸的跨度内,与根据SEQ ID NO:1的核苷酸序列至少80%同一的核苷酸序列。在一些实施方案中,经工程改造逆转录酶包含在至少600个氨基酸残基的跨度内,与SEQ ID NO:2至少80%同一的氨基酸序列。在一些实施方案中,经工程改造逆转录酶包含在至少1,800个核苷酸的跨度内,与根据SEQ ID NO.1的核苷酸序列至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%同一的核苷酸序列。在一些实施方案中,经工程改造逆转录酶包含在至少600个氨基酸残基的跨度内,与SEQ ID NO.2至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%同一的氨基酸序列。在一些实施方案中,经工程改造逆转录酶包含在至少600个氨基酸残基的跨度内,与SEQ ID NO.3至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%同一的氨基酸序列。

[0218] 在一些实施方案中,经工程改造逆转录酶包含相较于WT MMLV RT酶(例如SEQ ID NO:3),从经工程改造逆转录酶的N末端进行的至少15个氨基酸的截短。在一些实施方案中,经工程改造逆转录酶包含相较于WT MMLV RT酶,从经工程改造逆转录酶的N末端进行的至少20个氨基酸的截短。在一些实施方案中,经工程改造逆转录酶包含相较于WT MMLV RT酶,从经工程改造逆转录酶的N末端进行的至少21个氨基酸的截短。在一些实施方案中,经工程改造逆转录酶包含相较于WT MMLV RT酶,从经工程改造逆转录酶的N末端进行的至少25个氨基酸的截短。在一些实施方案中,经工程改造逆转录酶包含相较于WT MMLV RT酶,从经工程改造逆转录酶的N末端进行的至少30个氨基酸的截短。在一些实施方案中,经工程改造逆转录酶包含相较于WT MMLV RT酶,从经工程改造逆转录酶的N末端进行的至少35个氨基酸的截短。在一些实施方案中,经工程改造逆转录酶包含相较于WT MMLV RT酶,从经工程改造逆转录酶的N末端进行的至少40个氨基酸的截短。在一些实施方案中,经工程改造逆转录酶包含相较于WT MMLV RT酶(SEQ ID NO:3),从经工程改造逆转录酶的N末端进行的至少15、至少16、至少17、至少18、至少19、至少20、至少21、至少22、至少23、至少24、至少25、至少26、至少27、至少28、至少29、至少30、至少31、至少32、至少33、至少34、至少35、至少36、至少37、至少38、至少39或至少40个氨基酸的截短。在一些实施方案中,相较于WT MMLV RT,经工程改造逆转录酶中的N末端截短使蛋白质溶解性增加。

[0219] 在一些实施方案中,经工程改造逆转录酶在该酶的N末端包含具有至少5个组氨酸氨基酸的序列。在一些实施方案中,经工程改造逆转录酶在经工程改造逆转录酶的N末端包含6个组氨酸氨基酸。在一些实施方案中,经工程改造逆转录酶包含凝血酶裂解识别位点。在一些实施方案中,经工程改造逆转录酶包含位于该酶的N末端的具有至少5个组氨酸氨基酸的序列以及凝血酶裂解识别位点。在一些实施方案中,经工程改造逆转录酶在经工程改造逆转录酶的N末端包含6个组氨酸氨基酸和凝血酶裂解识别位点。在一些实施方案中,位于经工程改造逆转录酶的N末端的6个组氨酸氨基酸和凝血酶裂解识别位点具有氨基酸序列MRSSHHHHHHSSGLVPR(SEQ ID NO:7)。

[0220] 在一些实施方案中,除上述N末端截短之外,经工程改造逆转录酶也包含位于经工

程改造逆转录酶的N末端的至少5个组氨酸氨基酸和/或位于经工程改造逆转录酶的N末端的凝血酶裂解序列。举例来说，在一些实施方案中，经工程改造逆转录酶包含(a)相较于WT MMLV RT酶，从经工程改造RT酶的N末端进行的至少15个氨基酸的截短，(b)位于经工程改造逆转录酶的N末端的至少5个组氨酸氨基酸，和(c)位于该酶的N末端的凝血酶裂解识别位点。在一些实施方案中，经工程改造逆转录酶包含(a)相较于WT MMLV RT酶，从经工程改造RT酶的N末端进行的至少21个氨基酸的截短，(b)位于经工程改造逆转录酶的N末端的至少5个组氨酸氨基酸，和(c)位于所述酶的N末端的凝血酶裂解识别位点。在一些实施方案中，经工程改造逆转录酶包含(a)相较于WT MMLV RT酶，从经工程改造RT酶的N末端进行的至少25个氨基酸的截短，(b)位于经工程改造逆转录酶的N末端的至少5个组氨酸氨基酸，和(c)位于所述酶的N末端的凝血酶裂解识别位点。

[0221] 在一些实施方案中，经工程改造逆转录酶包含相较于WT MMLV RT酶，从经工程改造RT酶的N末端进行的至少15个氨基酸的截短，以及位于经工程改造逆转录酶的N末端的MRSSHHHHHHSSGLVPR氨基酸序列。在一些实施方案中，经工程改造逆转录酶包含相较于WT MMLV RT酶，从经工程改造RT酶的N末端进行的至少21个氨基酸的截短，并且进一步包含位于经工程改造逆转录酶的N末端的MRSSHHHHHHSSGLVPR氨基酸序列。在一些实施方案中，经工程改造逆转录酶包含相较于WT MMLV RT酶，从经工程改造RT酶的N末端进行的至少25个氨基酸的截短，并且进一步包含位于经工程改造逆转录酶的N末端的MRSSHHHHHHSSGLVPR氨基酸序列。

[0222] 在一些实施方案中，经工程改造逆转录酶包含一个或多个选自由以下组成的组的突变：相对于氨基酸序列SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。

[0223] 在一些实施方案中，经工程改造逆转录酶包含两个或更多个选自由以下组成的组的突变：相对于氨基酸序列SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。

[0224] 在一些实施方案中，经工程改造逆转录酶包含三个或更多个选自由以下组成的组的突变：相对于氨基酸序列SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。

[0225] 在一些实施方案中，经工程改造逆转录酶包含四个或更多个选自由以下组成的组的突变：相对于氨基酸序列SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。

[0226] 在一些实施方案中，经工程改造逆转录酶包含五个或更多个选自由以下组成的组的突变：相对于氨基酸序列SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。

[0227] 在一些实施方案中，经工程改造逆转录酶包含六个或更多个选自由以下组成的组

的突变:相对于氨基酸序列SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。

[0228] 在一些实施方案中,经工程改造逆转录酶包含七个或更多个选自由以下组成的组的突变:相对于氨基酸序列SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。

[0229] 在一些实施方案中,经工程改造逆转录酶包含八个或更多个选自由以下组成的组的突变:相对于氨基酸序列SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。

[0230] 在一些实施方案中,经工程改造逆转录酶包含九个或更多个选自由以下组成的组的突变:相对于氨基酸序列SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。

[0231] 在一些实施方案中,经工程改造逆转录酶包含十个或更多个选自由以下组成的组的突变:相对于氨基酸序列SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。

[0232] 在一些实施方案中,经工程改造逆转录酶包含十一个或更多个选自由以下组成的组的突变:相对于氨基酸序列SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。

[0233] 在一些实施方案中,经工程改造逆转录酶包含十二个或更多个选自由以下组成的组的突变:相对于氨基酸序列SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。

[0234] 在一些实施方案中,经工程改造逆转录酶包含十三个或更多个选自由以下组成的组的突变:相对于氨基酸序列SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。

[0235] 在一些实施方案中,经工程改造逆转录酶包含相对于氨基酸序列SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。

[0236] 在一些实施方案中,经工程改造逆转录酶包含一个或多个选自由以下组成的组的突变:相对于氨基酸序列SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0237] 在一些实施方案中,经工程改造逆转录酶包含两个或更多个选自由以下组成的组

的突变:相对于氨基酸序列SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0238] 在一些实施方案中,经工程改造逆转录酶包含三个或更多个选自由以下组成的组的突变:相对于氨基酸序列SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0239] 在一些实施方案中,经工程改造逆转录酶包含四个或更多个选自由以下组成的组的突变:相对于氨基酸序列SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0240] 在一些实施方案中,经工程改造逆转录酶包含五个或更多个选自由以下组成的组的突变:相对于氨基酸序列SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0241] 在一些实施方案中,经工程改造逆转录酶包含六个或更多个选自由以下组成的组的突变:相对于氨基酸序列SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0242] 在一些实施方案中,经工程改造逆转录酶包含七个或更多个选自由以下组成的组的突变:相对于氨基酸序列SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0243] 在一些实施方案中,经工程改造逆转录酶包含八个或更多个选自由以下组成的组的突变:相对于氨基酸序列SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0244] 在一些实施方案中,经工程改造逆转录酶包含九个或更多个选自由以下组成的组的突变:相对于氨基酸序列SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0245] 在一些实施方案中,经工程改造逆转录酶包含十个或更多个选自由以下组成的组的突变:相对于氨基酸序列SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0246] 在一些实施方案中,经工程改造逆转录酶包含十一个或更多个选自由以下组成的组的突变:相对于氨基酸序列SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0247] 在一些实施方案中，经工程改造逆转录酶包含十二个或更多个选自由以下组成的组的突变：相对于氨基酸序列SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0248] 在一些实施方案中，经工程改造逆转录酶包含十三个或更多个选自由以下组成的组的突变：相对于氨基酸序列SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0249] 在一些实施方案中，经工程改造逆转录酶包含相对于氨基酸序列SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0250] 在一些实施方案中，经工程改造逆转录酶包含(i)从N末端进行的至少15个氨基酸的截短；(ii)位于N末端的具有至少5个组氨酸氨基酸的序列；(iii)凝血酶裂解识别位点；和(iv)一个或多个选自由以下组成的组的突变：相对于氨基酸序列SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。在一些情况下，(iv)中的一个或多个突变选自由以下组成的组：相对于SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0251] 在一些实施方案中，经工程改造逆转录酶包含(i)从N末端进行的至少21个氨基酸的截短；(ii)位于N末端的具有至少5个组氨酸氨基酸的序列；(iii)凝血酶裂解识别位点；和(iv)一个或多个选自由以下组成的组的突变：相对于氨基酸序列SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。在一些情况下，(iv)中的一个或多个突变选自由以下组成的组：相对于SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0252] 在一些实施方案中，经工程改造逆转录酶包含(i)从N末端进行的至少21个氨基酸的截短；(ii)位于经工程改造逆转录酶的N末端的MRSSHHHHHHSSGLVPR氨基酸序列；和(iii)一个或多个选自由以下组成的组的突变：相对于氨基酸序列SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。在一些情况下，(iii)中的一个或多个突变选自由以下组成的组：相对于SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0253] 在一些实施方案中，经工程改造逆转录酶包含(i)从N末端进行的至少21个氨基酸的截短；(ii)位于经工程改造逆转录酶的N末端的MRSSHHHHHHSSGLVPR氨基酸序列；和(iii)相对于氨基酸序列SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313

突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。在一些情况下，(iii)中的一个或多个突变选自由以下组成的组：相对于SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0254] 在一些实施方案中，经工程改造逆转录酶包含：(i)一个或多个选自由以下组成的组的突变：相对于SEQ ID NO:3的L139突变、D200突变、T330突变、P448突变、D449突变、D524突变和L603突变；和(ii)一个或多个选自由以下组成的组的突变：相对于SEQ ID NO:3的E69突变、E302突变、T306突变、W313突变、L435突变和N454突变。在一些情况下，(i)和(ii)中的突变选自由以下组成的组：相对于SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0255] 在一些实施方案中，经工程改造逆转录酶包含：(i)两个或更多个选自由以下组成的组的突变：相对于SEQ ID NO:3的L139突变、D200突变、T330突变、P448突变、D449突变、D524突变和L603突变；和(ii)两个或更多个选自由以下组成的组的突变：相对于SEQ ID NO:3的E69突变、E302突变、T306突变、W313突变、L435突变和N454突变。在一些情况下，(i)和(ii)中的突变选自由以下组成的组：相对于SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0256] 在一些实施方案中，经工程改造逆转录酶包含：(i)三个或更多个选自由以下组成的组的突变：相对于SEQ ID NO:3的L139突变、D200突变、T330突变、P448突变、D449突变、D524突变和L603突变；和(ii)三个或更多个选自由以下组成的组的突变：相对于SEQ ID NO:3的E69突变、E302突变、T306突变、W313突变、L435突变和N454突变。在一些情况下，(i)和(ii)中的突变选自由以下组成的组：相对于SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0257] 在一些实施方案中，经工程改造逆转录酶包含：(i)四个或更多个选自由以下组成的组的突变：相对于SEQ ID NO:3的L139突变、D200突变、T330突变、P448突变、D449突变、D524突变和L603突变；和(ii)四个或更多个选自由以下组成的组的突变：相对于SEQ ID NO:3的E69突变、E302突变、T306突变、W313突变、L435突变和N454突变。在一些情况下，(i)和(ii)中的突变选自由以下组成的组：相对于SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0258] 在一些实施方案中，经工程改造逆转录酶包含：(i)五个或更多个选自由以下组成的组的突变：相对于SEQ ID NO:3的L139突变、D200突变、T330突变、P448突变、D449突变、D524突变和L603突变；和(ii)五个或更多个选自由以下组成的组的突变：相对于SEQ ID NO:3的E69突变、E302突变、T306突变、W313突变、L435突变和N454突变。在一些情况下，(i)和(ii)中的突变选自由以下组成的组：相对于SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G

突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0259] 在一些实施方案中，经工程改造逆转录酶包含：(i)六个或更多个选自由以下组成的组的突变：相对于SEQ ID NO:3的L139突变、D200突变、T330突变、P448突变、D449突变、D524突变和L603突变；和(ii)五个或更多个选自由以下组成的组的突变：相对于SEQ ID NO:3的E69突变、E302突变、T306突变、W313突变、L435突变和N454突变。在一些情况下，(i)和(ii)中的突变选自由以下组成的组：相对于SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0260] 在一些实施方案中，经工程改造逆转录酶具有根据SEQ ID NO:4的核苷酸序列。在一些实施方案中，经工程改造逆转录酶具有根据SEQ ID NO:5的氨基酸序列。

[0261] 在一些实施方案中，经工程改造逆转录酶被工程改造成具有降低和/或消除的RNA酶活性。在一些实施方案中，经工程改造逆转录酶被工程改造成具有降低和/或消除的RNA酶H活性。在一些实施方案中，被工程改造成具有降低和/或消除的RNA酶H活性的经工程改造逆转录酶包含D524突变。在一些实施方案中，被工程改造成具有降低和/或消除的RNA酶H活性的经工程改造逆转录酶包含D524A或D524N突变。

[0262] 本公开的经工程改造逆转录酶变体与其他所述和/或可商购获得的酶相比，出乎意料地具有各种改进益处，诸如改进的热稳定性、持续逆转录、非模板化碱基添加和模板转换能力。此外，相比于可商购获得的RT酶，本文所述的经工程改造逆转录酶变体也表现出对细胞溶解产物的出乎意料更高的抗性(即受细胞溶解产物抑制的程度较小)。最后，相较于其他所述和可商购获得的MMLV逆转录酶，本公开的经工程改造逆转录酶变体具有出乎意料更大的捕集全长转录物的能力(例如在T细胞受体成对转录剖析中)。

[0263] V.使用经工程改造逆转录酶进行核酸样品处理

[0264] 在一些实施方案中，本文公开用于核酸样品处理的方法，其包括：以某一反应体积提供模板核糖核酸(RNA)分子；以及使用经工程改造逆转录酶来使所述RNA分子逆转录成互补性DNA分子，其中所述经工程改造逆转录酶包含与SEQ ID NO:3至少80％同一的氨基酸序列，其中所述氨基酸序列包含：(i)相对于SEQ ID NO:3，从N末端进行的至少15个氨基酸的截短；和(ii)一个或多个选自由以下组成的组的突变：相对于SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。在一些情况下，(ii)中的突变选自由以下组成的组：相对于SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0265] 在一些实施方案中，模板核酸分子是核糖核酸(RNA)分子。在一些实施方案中，RNA分子是信使RNA(mRNA)。在一些实施方案中，本文所述的经工程改造逆转录酶在小于1nL的反应体积的情况下使用。在一些实施方案中，本文所述的经工程改造逆转录酶在小于500pL的反应体积的情况下使用。在一些实施方案中，反应体积包含在分区内。在一些实施方案中，反应体积包含在乳液中的小滴内。在一些实施方案中，反应体积包含在具有小于1nL的反应体积的小滴乳液内。在一些实施方案中，反应体积包含在具有小于500pL的反应体积的小滴乳液内。在一些实施方案中，反应体积包含在孔内。在一些实施方案中，反应体积包含

在具有小于1nL的反应体积的孔内。在一些实施方案中，反应体积包含在孔内。在一些实施方案中，反应体积包含在具有小于500pL的反应体积的孔内。在一些实施方案中，反应体积包含在一系列孔中的具有经提取核酸分子的孔内，并且其中所述模板核酸分子是所述经提取核酸分子。在一些实施方案中，反应体积包含在一系列孔中的具有包含模板核酸分子的细胞的孔内，并且其中所述模板核酸分子从所述细胞释放。

[0266] 在一些实施方案中，反应体积进一步包含粒子，所述粒子包含分子标签(例如条形码)。在一些实施方案中，粒子是凝胶珠粒。在一些实施方案中，所述分子标签以可释放方式连接于所述凝胶珠粒。在一些实施方案中，凝胶珠粒包含聚丙烯酰胺聚合物。

[0267] 在一些实施方案中，凝胶珠粒的截面小于约100μm。在一些实施方案中，凝胶珠粒的截面小于约60μm。在一些实施方案中，凝胶珠粒的截面小于约50μm。在一些实施方案中，凝胶珠粒的截面小于约40μm。在一些实施方案中，凝胶珠粒的截面小于约100μm、小于约99μm、小于约98μm、小于约97μm、小于约96μm、小于约95μm、小于约94μm、小于约93μm、小于约92μm、小于约91μm、小于约90μm、小于约89μm、小于约88μm、小于约87μm、小于约86μm、小于约85μm、小于约84μm、小于约83μm、小于约82μm、小于约81μm、小于约80μm、小于约79μm、小于约78μm、小于约77μm、小于约76μm、小于约75μm、小于约74μm、小于约73μm、小于约72μm、小于约71μm、小于约70μm、小于约69μm、小于约68μm、小于约67μm、小于约66μm、小于约65μm、小于约64μm、小于约63μm、小于约62μm、小于约61μm或小于约60μm。

[0268] 在一些实施方案中，分子标签(例如条形码寡核苷酸)包括独特分子标识符(UMI)。在一些实施方案中，UMI包含寡核苷酸。在一些实施方案中，分子标签偶联于引发序列。在一些实施方案中，所述引发序列各自包含随机N聚体序列。在一些实施方案中，随机N聚体序列互补于所述RNA分子的3’序列。在一些实施方案中，引发序列包含具有至少5个碱基的长度的多dT序列。在一些实施方案中，引发序列包含具有至少10个碱基的长度的多dT序列。在一些实施方案中，引发序列包含具有至少5个碱基、至少6个碱基、至少7个碱基、至少8个碱基、至少9个碱基、至少10个碱基的长度的多dT序列。

[0269] 在一些实施方案中，反应体积进一步包含细胞，并且模板核酸分子来自所述细胞。在一些实施方案中，反应体积进一步包含多个细胞，并且模板核酸分子来自所述多个细胞。

[0270] 在一些实施方案中，逆转录由所述引发序列与所述RNA分子的杂交引发，并且由经工程改造逆转录酶以模板引导方式延伸。在一些实施方案中，在使来自所述核酸分子的核糖核酸分子逆转录后，逆转录酶添加多个非模板寡核苷酸。在一些实施方案中，逆转录反应产生单链互补性脱氧核糖核酸(cDNA)分子，各自在其5’末端上具有来自所述分子标签的分子标签，继之以扩增cDNA以产生双链cDNA，所述双链cDNA在所述双链cDNA的5’末端和3’末端上具有分子标签。

[0271] 在一些实施方案中，本文公开的用于核酸样品处理的方法利用本文所述的经工程改造逆转录酶。在一些实施方案中，本文公开的用于核酸样品处理的方法利用包含与氨基酸序列SEQ ID NO:3至少80％同一的氨基酸序列的经工程改造逆转录酶，其中所述氨基酸序列的特征在于具有以下中的两者或更多者：(a)从所述氨基酸序列的N末端进行的至少15个氨基酸的截短；(b)位于所述氨基酸序列的所述N末端的具有至少5个组氨酸氨基酸的序列；(c)凝血酶裂解识别位点；和(d)一个或多个选自由以下组成的组的突变：相对于氨基酸序列SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突

变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。在一些情况下，(d)中的一个或多个突变选自由以下组成的组：相对于SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0272] 在一些实施方案中，本文公开的用于核酸样品处理的方法利用包含与氨基酸序列SEQ ID NO:3至少80％同一的氨基酸序列的经工程改造逆转录酶，其中所述氨基酸序列的特征在于具有以下中的三者或更多者：(a)从所述氨基酸序列的N末端进行的至少15个氨基酸的截短；(b)位于所述氨基酸序列的所述N末端的具有至少5个组氨酸氨基酸的序列；(c)凝血酶裂解识别位点；和(d)一个或多个选自由以下组成的组的突变：相对于氨基酸序列SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。在一些情况下，(d)中的一个或多个突变选自由以下组成的组：相对于SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0273] 在一些实施方案中，本文公开的用于核酸样品处理的方法利用包含与氨基酸序列SEQ ID NO:3至少80％同一的氨基酸序列的经工程改造逆转录酶，其中所述氨基酸序列的特征在于具有：(a)从所述氨基酸序列的N末端进行的至少15个氨基酸的截短；(b)位于所述氨基酸序列的所述N末端的具有至少5个组氨酸氨基酸的序列；(c)凝血酶裂解识别位点；和(d)一个或多个选自由以下组成的组的突变：相对于氨基酸序列SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。在一些情况下，(d)中的一个或多个突变选自由以下组成的组：相对于SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0274] 在一些实施方案中，本文公开的核酸样品处理方法利用包含与氨基酸序列SEQ ID NO:3至少80％同一的氨基酸序列的经工程改造逆转录酶，其中所述氨基酸序列包含(a)从N末端进行的至少21个氨基酸的截短；(b)位于经工程改造逆转录酶的N末端的MRSSHHHHHHSSGLVPR氨基酸序列；和(c)一个或多个选自由以下组成的组的突变：相对于氨基酸序列SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。在一些情况下，(c)中的一个或多个突变选自由以下组成的组：相对于SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0275] 在一些实施方案中，本文公开的核酸样品处理方法利用具有根据SEQ ID NO:5的氨基酸序列的经工程改造逆转录酶。

[0276] VI.使用经工程改造逆转录酶在亚纳升大小的小滴中进行核酸样品处理

[0277] 当反应体积小于约1nL时，来自单一细胞的mRNA的逆转录可被抑制。迄今为止，尚无公开研究描述如何克服这个效应。

[0278] 在一些实施方案中，本文公开用于核酸样品处理的方法，其包括：(a)在乳液中产

生多个小滴，其中所述多个小滴中的单个小滴包含(i)包含分子标签的粒子、和(ii)具有核酸分子的细胞，其中所述粒子的体积与所述单个小滴的体积的比率小于0.9，并且其中所述单个小滴的所述体积小于1纳升；(b)在条形码化反应中使用所述分子标签来使所述核酸分子条形码化，所述条形码化反应具有偏离对照反应速率至多约20％的速率，所述对照反应速率如对于具有1纳升的对照小滴体积，并且包含单一细胞的对照小滴中的对照条形码化反应所测定，由此提供条形码化核酸分子；和(c)使所述条形码化核酸分子经受核酸测序以产生关于至少所述核酸分子的子组的序列信息。

[0279] 在一些实施方案中，所述多个小滴的至少1％包含细胞。在一些实施方案中，所述多个小滴的至少10％包含细胞。在一些实施方案中，所述多个小滴的至少20％包含细胞。在一些实施方案中，所述多个小滴的至少1％、至少2％、至少3％、至少4％、至少5％、至少6％、至少7％、至少8％、至少9％、至少10％、至少11％、至少12％、至少13％、至少14％、至少15％、至少16％、至少17％、至少18％、至少19％或至少20％包含细胞。在一些实施方案中，所述多个小滴中的单个小滴包含多个细胞。

[0280] 在一些实施方案中，所述多个小滴中的单个小滴包含凝胶珠粒，所述凝胶珠粒包含分子标签。在一些实施方案中，所述分子标签以可释放方式连接于所述凝胶珠粒。在一些实施方案中，所述多个小滴中的单个小滴包含凝胶珠粒，所述凝胶珠粒包含聚丙烯酰胺聚合物。在一些实施方案中，所述多个小滴中的单个小滴包含可降解凝胶珠粒。在一些实施方案中，所述分子标签包括独特分子标识符(UMI)。在一些实施方案中，所述UMI包含寡核苷酸。

[0281] 在一些实施方案中，凝胶珠粒的截面小于约100μm。在一些实施方案中，凝胶珠粒的截面小于约60μm。在一些实施方案中，凝胶珠粒的截面小于约50μm。在一些实施方案中，凝胶珠粒的截面小于约40μm。在一些实施方案中，凝胶珠粒的截面小于约100μm、小于约99μm、小于约98μm、小于约97μm、小于约96μm、小于约95μm、小于约94μm、小于约93μm、小于约92μm、小于约91μm、小于约90μm、小于约89μm、小于约88μm、小于约87μm、小于约86μm、小于约85μm、小于约84μm、小于约83μm、小于约82μm、小于约81μm、小于约80μm、小于约79μm、小于约78μm、小于约77μm、小于约76μm、小于约75μm、小于约74μm、小于约73μm、小于约72μm、小于约71μm、小于约70μm、小于约69μm、小于约68μm、小于约67μm、小于约66μm、小于约65μm、小于约64μm、小于约63μm、小于约62μm、小于约61μm或小于约60μm。

[0282] 在一些实施方案中，所述粒子的体积与所述单个小滴的体积的比率小于0.9。在一些实施方案中，所述粒子的体积与所述单个小滴的体积的比率小于0.8。在一些实施方案中，所述粒子的体积与所述单个小滴的体积的比率小于0.7。在一些实施方案中，所述粒子的体积与所述单个小滴的体积的比率小于0.2。在一些实施方案中，所述粒子的体积与所述单个小滴的体积的比率小于约0.9、小于约0.85、小于约0.8、小于约0.75、小于约0.7、小于约0.65、小于约0.6、小于约0.55、小于约0.5、小于约0.45、小于约0.4、小于约0.35、小于约0.3、小于约0.25或小于约0.2。

[0283] 在一些实施方案中，对照条形码化反应是对来自所述单一细胞的核酸分子进行的逆转录反应。在一些实施方案中，使用具有如以SEQ ID NO:3阐述的氨基酸序列的逆转录酶在所述对照小滴中进行逆转录反应。在一些实施方案中，使用具有如以SEQ ID NO.5阐述的序列的逆转录酶在所述对照小滴中进行逆转录反应。在一些实施方案中，对照小滴包含另

一粒子，所述粒子包含分子标签。在一些实施方案中，条形码化反应是酶促反应。在一些实施方案中，条形码化反应是逆转录扩增反应，其在使所述细胞的核糖核酸(RNA)分子逆转录后产生互补性脱氧核糖核酸(cDNA)分子。在一些实施方案中，RNA分子从所述细胞释放。在一些实施方案中，通过使所述细胞溶解来从所述细胞释放RNA分子。在一些实施方案中，RNA分子是信使RNA(mRNA)。

[0284] 在一些实施方案中，分子标签偶联于引发序列，并且条形码化反应由所述引发序列与所述RNA分子的杂交引发。在一些实施方案中，所述引发序列各自包含随机N聚体序列。在一些实施方案中，所述随机N聚体序列互补于所述细胞的核糖核酸分子的3'序列。在一些实施方案中，所述随机N聚体序列包含具有至少5个碱基的长度的多dT序列。在一些实施方案中，所述随机N聚体序列包含具有至少10个碱基的长度的多dT序列。在一些实施方案中，条形码化反应通过使用用于逆转录的试剂，以模板引导方式使所述引发序列各自延伸来进行。在一些实施方案中，用于逆转录的试剂包括逆转录酶、缓冲剂和核苷酸混合物。在一些实施方案中，在使来自所述核酸分子的核糖核酸分子逆转录后，逆转录酶添加多个非模板寡核苷酸。在一些实施方案中，逆转录酶是如本文公开的经工程改造逆转录酶。

[0285] 在一些实施方案中，条形码化反应产生单链互补性脱氧核糖核酸(cDNA)分子，各自在其5'末端上具有来自所述分子标签的分子标签，继之以扩增cDNA以产生双链cDNA，所述双链cDNA在所述双链cDNA的5'末端和3'末端上具有分子标签。

[0286] 在一些实施方案中，在对条形码化核酸分子测序之前，使来自单个小滴的包含双链cDNA的条形码化核酸分子释放。在一些实施方案中，在对条形码化核酸分子测序之前，将来自所述多个小滴的条形码化核酸分子汇合。

[0287] 在一些实施方案中，其中单个小滴小于1纳升的本文公开的用于核酸样品处理的方法利用本文所述的经工程改造逆转录酶。在一些实施方案中，其中单个小滴小于1纳升的本文公开的用于核酸样品处理的方法利用包含与氨基酸序列SEQ ID NO:3至少80％同一的氨基酸序列的经工程改造逆转录酶，其中所述氨基酸序列的特征在于具有以下中的两者或更多者：(a)从所述氨基酸序列的N末端进行的至少15个氨基酸的截短；(b)位于所述氨基酸序列的所述N末端的具有至少5个组氨酸氨基酸的序列；(c)凝血酶裂解识别位点；和(d)一个或多个选自由以下组成的组的突变：相对于氨基酸序列SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。在一些情况下，(d)中的一个或多个突变选自由以下组成的组：相对于SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0288] 在一些实施方案中，其中单个小滴小于1纳升的本文公开的用于核酸样品处理的方法利用包含与氨基酸序列SEQ ID NO:3至少80％同一的氨基酸序列的经工程改造逆转录酶，其中所述氨基酸序列的特征在于具有以下中的三者或更多者：(a)从所述氨基酸序列的N末端进行的至少15个氨基酸的截短；(b)位于所述氨基酸序列的所述N末端的具有至少5个组氨酸氨基酸的序列；(c)凝血酶裂解识别位点；和(d)一个或多个选自由以下组成的组的突变：相对于氨基酸序列SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和

E607突变。在一些情况下，(d)中的一个或多个突变选自由以下组成的组：相对于SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0289] 在一些实施方案中，其中单个小滴小于1纳升的本文公开的用于核酸样品处理的方法利用包含与氨基酸序列SEQ ID NO:3至少80％同一的氨基酸序列的经工程改造逆转录酶，其中所述氨基酸序列的特征在于具有：(a)从所述氨基酸序列的N末端进行的至少15个氨基酸的截短；(b)位于所述氨基酸序列的所述N末端的具有至少5个组氨酸氨基酸的序列；(c)凝血酶裂解识别位点；和(d)一个或多个选自由以下组成的组的突变：相对于氨基酸序列SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。在一些情况下，(d)中的一个或多个突变选自由以下组成的组：相对于SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0290] 在一些实施方案中，其中单个小滴小于1纳升的本文公开的核酸样品处理方法利用包含与氨基酸序列SEQ ID NO:3至少80％同一的氨基酸序列的经工程改造逆转录酶，其中所述氨基酸序列包含(a)从N末端进行的至少21个氨基酸的截短；(b)位于经工程改造逆转录酶的N末端的MRSSHHHHHHSSGLVPR氨基酸序列；和(c)一个或多个选自由以下组成的组的突变：相对于氨基酸序列SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。在一些情况下，(c)中的一个或多个突变选自由以下组成的组：相对于SEQ ID NO:3的E69K突变、L139P突变、D200N突变、E302R突变、T306K突变、W313F突变、T330P突变、L435G或L435K突变、P448A突变、D449G突变、N454K突变、D524N或D524A突变、L603W突变和E607K突变。

[0291] 在一些实施方案中，其中单个小滴小于1纳升的本文公开的核酸样品处理方法利用具有根据SEQ ID NO:5的氨基酸序列的经工程改造逆转录酶。

[0292] VII.装置和系统

[0293] 本文也提供如上所述的用于分配细胞的微流体装置。所述微流体装置可包括用于进行分配过程的通道网络，如图1和图2中阐述的那些。特别适用的微流体装置的实例描述于2014年4月4日提交，并且出于所有目的以引用的方式整体并入本文的美国临时专利申请号61/977,804中。简要来说，这些微流体装置可包括用于将细胞分配至单独分区中，以及将所述细胞与例如安置在珠粒上的寡核苷酸条形码文库成员共分配的通道网络，诸如本文所述的那些。这些通道网络可被安置在通道被限定于其中的固体主体例如玻璃、半导体或聚合物主体结构内，在所述固体主体中，那些通道在它们的末端与用于接收各种输入流体以及用于最终存放来自通道网络的输出物的经分配细胞等的储集器连通。

[0294] 也提供例如通过所施加的压差、离心力、电动泵送、毛细管或重力流动等来控制这些流体通过通道网络进行流动的系统。

[0295] VIII.试剂盒

[0296] 本文也提供了用于执行逆转录反应的试剂盒，所述试剂盒包括(a)本公开的经工

程改造逆转录酶和(b)关于使用所述经工程改造逆转录酶来进行逆转录反应的说明书。经工程改造逆转录酶可包含(i)相对于SEQ ID NO:3,从N末端进行的至少15个氨基酸的截短;和(ii)一个或多个选自由以下组成的组的突变:相对于SEQ ID NO:3的E69突变、L139突变、D200突变、E302突变、T306突变、W313突变、T330突变、L435突变、P448突变、D449突变、N454突变、D524突变、L603突变和E607突变。试剂盒也可包括被公开用于进行本公开的方法的适合反应缓冲液、dNTP、一种或多种引物、一种或多种对照试剂、或任何其他试剂。经工程改造逆转录酶、反应缓冲液和dNTP可被单独提供,或可被一起提供在主混合物溶液中。在经工程改造逆转录酶、反应缓冲液和dNTP被提供在主混合物中的情况下,所述主混合物以是在说明书中指示的用于逆转录反应中的工作浓度至少两倍的浓度存在。在其他情况下,主混合物可以是指示的工作浓度至少三倍、至少四倍、至少五倍、至少六倍、至少七倍、至少八倍、至少九倍、或至少十倍的浓度存在。试剂盒中的引物可为多dT引物、随机N聚体引物、或靶标特异性引物。

[0297] 试剂盒可进一步包括以下中的一者、两者、三者、四者、五者或更多者、直至全部:分配流体,包括水性缓冲液与非水性分配流体或油两者;如本文所述的以可释放方式与珠粒缔合的核酸条形码文库;微流体装置;用于破坏细胞、扩增核酸以及在细胞核酸或其复制物的片段上提供额外功能性序列的试剂;以及关于在本文所述的方法中使用前述各物中的任一者的说明书。

[0298] 关于使用任何方法的说明书通常记录在适合记录介质上(例如印刷在基材诸如纸张或塑料上)。因此,说明书可以包装插页形式存在于试剂盒中,存在于试剂盒或其组成部分(即与包装或子包装相伴的组成部分)的容器的标签中。在一些情况下,说明书可以存在于适合计算机可读储存介质上的电子储存数据文件形式存在,所述介质例如CD-ROM、磁盘等。在其他情况下,实际说明书可不存在于试剂盒中,而是可提供用于例如通过因特网来从远程源头获得说明书的手段。举例来说,试剂盒包括可察看说明书所处的和/或可从其下载说明书的网址。如同说明书一样,用于获得说明书的这个手段被记录在适合基材上。

[0299] IX.计算机控制系统

[0300] 本公开提供被编程来执行本公开方法的计算机控制系统。图17显示计算机系统1701,其被编程或另外被配置来执行本公开方法,包括核酸测序方法、对核酸测序数据的解释和对细胞核酸诸如RNA(例如mRNA)的分析、以及根据测序数据对细胞的表征。计算机系统1701可为用户的电子装置或相对于所述电子装置加以远程定位的计算机系统。电子装置可为移动电子装置。

[0301] 计算机系统1701包括中央处理单元(CPU,在本文中也称为“处理器”和“计算机处理器”)1705,其可为单核或多核处理器或用于并行处理的多个处理器。计算机系统1701也包括存储器或存储位置1710(例如随机存取存储器、只读存储器、闪速存储器)、电子存储单元1715(例如硬盘)、用于与一个或多个其他系统通信的通信接口1720(例如网络适配器)、以及外围装置1725,诸如高速缓冲存储器、其他存储器、数据存储和/或电子显示适配器。存储器1710、存储单元1715、接口1720和外围装置1725通过诸如主板的通信总线(实线)与CPU1705通信。存储单元1715可为用于存储数据的数据存储单元(或数据储存库)。借助于通信接口1720,可使计算机系统1701可操作地耦合于计算机网络(“网络”)1730。网络1730可为因特网、内联网和/或外联网、或与因特网通信的内联网和/或外联网。在一些情况下,网

络1730是电信和/或数据网络。网络1730可包括一个或多个计算机服务器，其可使得能够达成分布计算诸如云计算。在一些情况下，借助于计算机系统1701，网络1730可执行对等网络，其可使得耦合于计算机系统1701的装置能够充当客户端或服务器。

[0302] CPU 1705可执行可包括在程序或软件中的机器可读指令的序列。指令可存储在存储位置诸如存储器1710中。指令可被导向CPU1705，其可随后对CPU 1705编程或另外配置以执行本公开的方法。由CPU 1705进行的操作的实例可包括提取、解码、执行和写回。

[0303] CPU 1705可为电路诸如集成电路的一部分。系统1701的一个或多个其他部件可被包括在电路中。在一些情况下，电路是专用集成电路(ASIC)。

[0304] 存储单元1715可存储文件，诸如驱动程序、文库和保存程序。存储单元1715可存储用户数据，例如用户偏好和用户程序。在一些情况下，计算机系统1701可包括一个或多个在计算机系统1701的外部，诸如位于通过内联网或因特网来与计算机系统1701通信的远程服务器上的额外数据存储单元。

[0305] 计算机系统1701可通过网络1730来与一个或多个远程计算机系统通信。举例来说，计算机系统1701可与用户的远程计算机系统通信。远程计算机系统的实例包括个人计算机(例如便携式PC)、平板(slate/tablet)PC(例如Apple® iPad、Samsung® Galaxy Tab)、电话、智能电话(例如Apple® iPhone、支持Android的装置、Blackberry®)或个人数字助理。用户可通过网络1730来访问计算机系统1701。

[0306] 如本文所述的方法可通过存储在计算机系统1701的电子存储位置上，诸如像存储在存储器1710或电子存储单元1715上的机器(例如计算机处理器)可执行代码来执行。机器可执行或机器可读代码可以软件的形式提供。在使用期间，代码可由处理器1705执行。在一些情况下，代码可从存储单元1715撷取，并且存储在存储器1710上以达成准备由处理器1705访问。在一些情况下，可排除电子存储单元1715，并且将机器可执行指令存储在存储器1710上。

[0307] 代码可被预编译以及配置供与具有适合于执行代码的处理器的机器一起使用，或可在运行时间期间加以编译。代码可以可被选择以使得代码能够以预编译或期间编译方式执行的编程语言提供。

[0308] 本文提供的系统和方法诸如计算机系统1701的各个方面可在编程中体现。技术的各个方面可被认为是通常呈在某一类型的机器可读介质上携带或在某一类型的机器可读介质中包括的机器(或处理器)可执行代码和/或相关数据形式的“产品”或“制品”。机器可执行代码可存储在电子存储单元诸如存储器(例如只读存储器、随机存取存储器、闪速存储器)或硬盘上。“存储”型介质可包括计算机、处理器等的任何或所有有形存储器或其相关模块，诸如各种半导体存储器、磁带驱动器、磁盘驱动器等，其可在任何时间为软件编程提供非暂时存储。软件的全部或各部分可有时通过因特网或各种其他电信网络来通信。所述通信例如可使得能够将来自一个计算机或处理器的软件装载至另一计算机或处理器中，例如从管理服务器或主计算机装载至应用服务器的计算机平台中。因此，可携带软件要素的另一类型的介质包括光波、电波和电磁波，诸如在本地装置之间的物理接口之间，通过有线和光学地线网络，以及经各种空中链路来使用。携带所述波的物理元件诸如有线或无线链路、光链路等，也可被视为携带软件的介质。如本文所用，除非局限于非暂时有形“存储”介质，

否则诸如计算机或机器“可读介质”的术语是指参与对处理器提供执行指令的任何介质。

[0309] 因此，机器可读介质诸如计算机可执行代码可采用许多形式，包括但不限于有形存储介质、载波介质或物理传输介质。非易失性存储介质包括例如光盘或磁盘，诸如任何计算机等中的任何存储装置，诸如附图中所示的可用于执行数据库等的存储装置。易失性存储介质包括动态存储器，诸如这种计算机平台的主存储器。有形传输介质包括同轴电缆；铜丝和光纤，包括构成计算机系统内的总线的丝线。载波传输介质可采用电信号或电磁信号或声波或光波的形式，诸如在射频(RF)和红外(IR)数据通信期间产生的那些。因此，计算机可读介质的常见形式包括例如：软盘、软磁盘、硬盘、磁带、任何其他磁性介质、CD-ROM、DVD或DVD-ROM、任何其他光学介质、穿孔卡纸带、具有孔样式的任何其他物理存储介质、RAM、ROM、PROM和EPROM、FLASH-EPROM、任何其他存储器芯片或盒体、传送数据或指令的载波、运送这种载波的电缆或链路、或计算机可从其读取编程代码和/或数据的任何其他介质。这些形式的计算机可读介质中的许多可涉及将一个或多个指令的一个或多个序列携带至处理器以用于执行。

[0310] 计算机系统1701可包括电子显示器1735或与电子显示器1735通信，所述电子显示器包括用于提供例如核酸测序结果、核酸测序数据分析、核酸测序样品表征、细胞表征等的用户接口(UI)1740。UI的实例包括不限于图形用户接口(GUI)和基于网络的用户接口。

[0311] 本公开的方法和系统可通过一种或多种算法来执行。算法可在由中央处理单元1705执行软件后，通过软件来执行。算法可例如启动核酸测序，处理核酸测序数据，解释核酸测序结果，表征核酸样品，表征细胞等。

[0312] 实施例

[0313] 实施例1.使用乳液进行细胞RNA分析

[0314] 在一实例中，以如图9A中所示的操作，在乳液小滴的情况下进行伴有模板转换和cDNA扩增(通过PCR)的逆转录。被分配来进行逆转录和cDNA扩增(通过PCR)的反应混合物包括1,000个细胞或10,000个细胞或10ng RNA、携带条形码化寡核苷酸的珠粒/0.2％Tx-100/5x Kapa缓冲液、2x Kapa HS HiFi Ready Mix、4μM转换寡聚物和Smartscribe。当存在细胞时，将混合物分配以使大多数或所有小滴包含单一细胞和单一珠粒。细胞被溶解，同时条形码化寡核苷酸从珠粒释放，并且条形码化寡核苷酸的多dT区段杂交于从细胞释放的mRNA的多A尾部，如在操作950中。如在操作952中，多dT区段在逆转录反应中被延伸，并且如在操作954中，cDNA转录物被扩增。热循环条件是42℃持续130分钟；98℃持续2分钟；以及35个以下循环：98℃持续15秒，60℃持续20秒，并且72℃持续6分钟。在热循环之后，破坏乳液，并且用Dynabead和0.6x SPRI纯化转录物，如在操作956中。

[0315] 由在乳液的情况下进行模板转换逆转录和PCR获得的产量显示在图13A中(对于1,000个细胞)和图13C中(对于10,000个细胞)以及图13B中(对于10ng RNA)(Smartscribe线)。将由在乳液的情况下针对10ng RNA进行的RT和PCR获得的cDNA转录物剪切，并且连接于功能性序列，用0.8x SPRI净化，并且通过PCR来进一步扩增，如在操作958中。用0.8x SPRI净化扩增产物。由这个处理获得的产量显示于图13B中(SSII线)。

[0316] 实施例2.使用乳液进行细胞RNA分析

[0317] 在另一实例中，以如图9A中所示的操作，在乳液小滴的情况下进行伴有模板转换和cDNA扩增(通过PCR)的逆转录。被分配来进行逆转录和cDNA扩增(通过PCR)的反应混合物

包括Jurkat细胞、携带条形码化寡核苷酸的珠粒/0.2％曲通X-100/5x Kapa缓冲液、2x Kapa HS HiFi Ready Mix、4μM转换寡聚物和Smartscribe。将混合物分配以使大多数或所有小滴包含单一细胞和单一珠粒。细胞被溶解，同时条形码化寡核苷酸从珠粒释放，并且条形码化寡核苷酸的多dT区段杂交于从细胞释放的mRNA的多A尾部，如在操作950中。如在操作952中，多dT区段在逆转录反应中被延伸，并且如在操作954中，cDNA转录物被扩增。热循环条件是42℃持续130分钟；98℃持续2分钟；以及35个以下循环：98℃持续15秒，60℃持续20秒，并且72℃持续6分钟。在热循环之后，破坏乳液，并且用Dynabead和0.6x SPRI净化转录物，如在操作956中。由用各种细胞数目(625个细胞、1,250个细胞、2,500个细胞、5,000个细胞和10,000个细胞)进行的反应获得的产量显示于图14A中。这些产量以图14B中所示的GADPH qPCR测定结果加以确认。

[0318] 实施例3.使用乳液进行RNA分析

[0319] 在另一实例中，以与如图9C中所示的方式类似的方式，在乳液小滴的情况下进行逆转录，并且散开进行cDNA扩增。被分配来进行逆转录的反应混合物包括携带条形码化寡核苷酸的珠粒、10ng Jurkat RNA(例如Jurkat mRNA)、5x第一链缓冲液和Smartscribe。条形码化寡核苷酸从珠粒释放，并且条形码化寡核苷酸的多dT区段杂交于RNA的多A尾部，如在操作961中。多dT区段在逆转录反应中被延伸，如在操作963中。用于逆转录的热循环条件是在42℃下持续2小时的一个循环和在70℃下持续10分钟的一个循环。在热循环之后，破坏乳液，并且使RNA和cDNA转录物变性，如在操作962中。接着采用具有生物素标签的引物，通过引物延伸来合成第二链，如在操作964中。用于这个引物延伸的反应条件包括作为第一链的cDNA转录物和浓度在0.5-3.0μM的范围内的生物素化延伸引物。热循环条件是在98℃下持续3分钟的一个循环和98℃持续15秒、60℃持续20秒并且72℃持续30分钟的一个循环。在引物延伸之后，用Dynabead MyOne链霉亲和素C1和T1下拉第二链，并且用Agilent SureSelect XT缓冲液净化。如在操作965中，采用以下循环条件，通过PCR来预扩增第二链-在98℃下持续3分钟的一个循环和98℃持续15秒、60℃持续20秒并且72℃持续30分钟的一个循环。对于各种浓度的生物素化引物(0.5μM、1.0μM、2.0μM和3.0μM)，产量显示于图15中。

[0320] 实施例4.使用乳液进行RNA分析

[0321] 在另一实例中，通过T7聚合酶进行的体外转录用于产生RNA转录物，如图10中所示。被分配来进行逆转录的混合物包括携带条形码化寡核苷酸的珠粒，所述条形码化寡核苷酸也包括T7 RNA聚合酶启动子序列；10ng人RNA(例如人mRNA)、5x第一链缓冲液和Smartscribe。将混合物分配以使大多数或所有小滴包含单一珠粒。条形码化寡核苷酸从珠粒释放，并且条形码化寡核苷酸的多dT区段杂交于RNA的多A尾部，如在操作1050中。多dT区段在逆转录反应中被延伸，如在操作1052中。热循环条件是在42℃下持续2小时的一个循环和在70℃下持续10分钟的一个循环。在热循环之后，破坏乳液，并且散开进行剩余操作。接着通过引物延伸来合成第二链，如在操作1054中。用于这个引物延伸的反应条件包括作为模板的cDNA转录物和延伸引物。热循环条件是在98℃下持续3分钟的一个循环和98℃持续15秒、60℃持续20秒并且72℃持续30分钟的一个循环。在这个引物延伸之后，用0.6x SPRI纯化第二链。如在操作1056中，接着进行体外转录以产生RNA转录物。过夜进行体外转录，并且用0.6x SPRI纯化转录物。由体外转录获得的RNA产量显示于图16中。

[0322] 实施例5.小于1nL的小滴的RNA分析

[0323] 一堆明确证据显示当反应体积小于约1nL时，来自单一细胞的mRNA的逆转录由存在于细胞溶解产物中的未知组分抑制。为克服这个抑制以及有助于利用较小反应体积来达成样品通量增加，产生了如本文公开的经工程改造MMLV RT酶，并且在含有皮升大小反应体积的小滴的情况下测试。相较于可商购获得的突变MMLV RT酶(CA-MMLV)，一种所述经工程改造MMLV RT酶，即酶42B(SEQ ID NO:5)，显示在350pL反应体积的情况下，逆转录抑制降低。

[0324] 表2.样品条件

[0325]

| 条件 | 样品 | 小滴体积 | 酶 |
|---|---|---|---|
| 1 | GEM-U | 1.1 nL | CA-MMLV (对照) |
| 2 | GEM-L | 350 pL | CA-MMLV |
| 3 | GEM-L | 350 pL | 42B |

[0326] 如图18-图20和表3中所示，在350pL小滴体积的情况下，当使用可商购获得的突变MMLV RT酶时，RT反应被抑制。相反，在350pL小滴体积的情况下，当使用酶42B时，未观察到抑制，从而指示相较于可商购获得的突变MMLV RT酶，酶42B在小滴在体积方面<1nL的情况下更耐受于细胞溶解产物抑制，并且提供更大文库复杂性。

[0327] 表3.在350pL小滴的情况下的比较结果

[0328]

| | 规格 | 1.1 nL (GEM-U) CA-MMLV | 350 pL (GEM-L) CA-MMLV | 350 pL (GEM-L) 42B |
|---|---|---|---|---|
| **缩减取样量度** | | | | |
| **每个细胞的 hg19 中值基因(50K RRPC)** | | 4,598 | 3,671 | 5,386 |
| **每个细胞的 hg19 中值 UMI 计数(50K RRPC)** | | 24,473 | 13,481 | 21,478 |

[0329]

| | | | | |
|---|---|---|---|---|
| 每个细胞的 mm10 中值基因(50K RRPC) | | 3,100 | 2,150 | 3,484 |
| 每个细胞的 mm10 中值 UMI 计数(50K RRPC) | | 12,268 | 5,959 | 11,958 |
| 平均 UMI 计数纯度 (50K RRPC) | ≥ 98% | 98.50% | 99.30% | 99.40% |
| cDNA PCR 重复率(50K RRPC) | | 32.20% | 56.80% | 15.60% |
| 多态率(每 1000 个细胞) | ≤ 2.0% | | | |
| 定位率量度 | | | | |
| 可用读段的分数 | | 56.40% | 46.00% | 43.50% |
| 具有引物或同聚物序列的读段的分数 | | 3.50% | 4.50% | 7.60% |
| rRNA 读段的分数 | | 0.10% | 0.10% | 0.50% |
| mtRNA 读段的分数 | | 4.30% | 3.70% | 6.60% |
| 确信定位于转录物组的读段的分数 | ≥ 50.0% | 68.00% | 61.50% | 52.70% |
| 确信定位于以下的人(hg19)读段 | | | | |
| 转录物组 | | 39.00% | 36.30% | 31.20% |
| 外显子区域 | | 41.60% | 39.10% | 33.80% |
| 内含子区域 | | 8.90% | 12.80% | 18.00% |
| 基因间区域 | | 2.40% | 3.00% | 4.80% |
| 确信定位于以下的小鼠(mm10)读段 | | | | |
| 转录物组 | | 29.00% | 25.20% | 21.50% |
| 外显子区域 | | 30.30% | 26.50% | 22.50% |
| 内含子区域 | | 5.40% | 7.50% | 7.90% |
| 基因间区域 | | 1.50% | 1.80% | 1.90% |

[0330] 实施例6.在单细胞转录剖析的情况下分析经工程改造MMLV RT酶变体

[0331] 收集细胞,并且洗涤以移除污染物。如本文所公开来产生包含单一细胞、单一凝胶珠粒和RT主混合物的小滴。含有引物的条形码化凝胶珠粒从凝胶珠粒释放,并且与细胞溶解产物和主混合物混合以产生准备用于下一代测序的条形码化全长cDNA。条形码化cDNA使用如本文公开的经工程改造MMLV RT酶(例如酶42B)或可商购获得的突变MMLV RT酶(CA-MMLV)来产生。

[0332] 表4.CA-MMLV结果

[0333]

| CA-MMLV 酶 | | | | |
|---|---|---|---|---|
| 样品 | 10 U/μL | 15 U/μL | 20 U/μL | 25 U/μL |
| 样品 ID | 27325 | 27326 | 27327 | 27328 |
| 每个细胞的 hg19 中值 UMI 计数(20K RRPC) | 10,478 | 10,738 | 10,749 | 10,577 |
| 每个细胞的 mm10 中值 UMI 计数 (20K RRPC) | 6,683 | 6,982 | 6,893 | 6,859 |
| <500 nt 的基因的 UMI 计数的分数 | 0.8% | 1.0% | 1.1% | 1.1% |
| 500-1000 nt 基因的 UMI 计数的分数 | 29.4% | 30.9% | 32.4% | 32.7% |
| 1000-1500 nt 基因的 UMI 计数的分数 | 29.0% | 28.7% | 28.3% | 28.5% |
| >1500 nt 的基因的 UMI 计数的分数 | 40.3% | 39.4% | 38.2% | 37.7% |
| 核糖体蛋白质 UMI 计数的分数 | 26.4% | 28.5% | 29.8% | 30.0% |
| 线粒体 UMI 计数的分数 | 2.1% | 2.1% | 2.1% | 1.9% |

[0334] 表5.CA-MMLV结果

[0335]

| CA-MMLV 酶 | | | | |
|---|---|---|---|---|
| 样品 | 10 U/μL | 15 U/μL | 20 U/μL | 25 U/μL |
| 样品 ID | 27325 | 27326 | 27327 | 27328 |
| 每个细胞的平均读段 | 51,623 | 87,837 | 93,780 | 57,701 |
| 细胞中读段的 hg19 分数 | 80.2% | 82.6% | 81.1% | 84.4% |
| 细胞中读段的 mm10 分数 | 84.6% | 85.4% | 85.2% | 86.6% |
| 可用读段的分数 | 54.1% | 54.2% | 53.5% | 58.8% |
| mtRNA 读段的分数 | 1.9% | 1.9% | 2.0% | 1.9% |
| 具有引物或同聚物序列的读段的分数 | 2.2% | 2.1% | 2.0% | 2.1% |
| cDNA PCR 重复(20K RRPC) | 20.2% | 17.2% | 15.4% | 15.1% |
| 每个细胞的 hg19 中值基因(20K RRPC) | 3,050 | 3,035 | 3,010 | 2,939 |
| 每个细胞的 mm10 中值基因(20K RRPC) | 2,257 | 2,255 | 2,190 | 2,186 |

[0336] 表6.经工程改造RT突变体结果

[0337]

| 突变体 42B | | | | |
|---|---|---|---|---|
| 样品 | 6 U/μL | 9 U/μL | 12 U/μL | 15 U/μL |
| 样品 ID | 27329 | 27330 | 27331 | 27332 |

[0338]

| | | | | |
|---|---|---|---|---|
| 每个细胞的平均读段 | 55,525 | 53,843 | 28,539 | 47,768 |
| 细胞中读段的 hg19 分数 | 89.4% | 87.3% | 90.1% | 91.5% |
| 细胞中读段的 mm10 分数 | 90.5% | 90.7% | 90.8% | 91.6% |
| 可用读段的分数 | 60.2% | 58.6% | 59.2% | 58.8% |
| mtRNA 读段的分数 | 8.5% | 7.9% | 7.7% | 9.1% |
| 具有引物或同聚物序列的读段的分数 | 1.6% | 1.6% | 1.6% | 1.5% |
| cDNA PCR 重复(20K RRPC) | 36.5% | 28.6% | 26.0% | 19.3% |
| 每个细胞的 hg19 中值基因(20K RRPC) | 2,985 | 3,060 | 3,148 | 3,176 |
| 每个细胞的 mm10 中值基因(20K RRPC) | 2,167 | 2,341 | 2,367 | 2,242 |

[0339] 表7.经工程改造RT突变体结果

[0340]

| 突变体 42B | | | | |
|---|---|---|---|---|
| 样品 | 约 6 U/μL | 约 9 U/μL | 约 12 U/μL | 约 15 U/μL |
| 样品 ID | 27329 | 27330 | 27331 | 27332 |
| 每个细胞的 hg19 中值 UMI 计数(20K RRPC) | 8,742 | 9,484 | 10,070 | 11,253 |
| 每个细胞的 mm10 中值 UMI 计数(20K RRPC) | 5,322 | 6,185 | 6,312 | 6,630 |
| <500 nt 的基因的 UMI 计数的分数 | 0.5% | 0.6% | 0.5% | 0.9% |
| 500-1000 nt 基因的 UMI 计数的分数 | 23.6% | 24.4% | 24.2% | 30.1% |
| 1000-1500 nt 基因的 UMI 计数的分数 | 26.1% | 26.7% | 27.2% | 27.7% |
| >1500 nt 的基因的 UMI 计数的分数 | 49.8% | 28.4% | 48.1% | 41.4% |
| 核糖体蛋白质 UMI 计数的分数 | 17.3% | 18.5% | 18.5% | 22.3% |
| 线粒体 UMI 计数的分数 | 9.4% | 8.7% | 8.1% | 9.7% |

[0341] 如表4-表7中所见,尽管经工程改造RT酶42B在给定测序深度下产生大致相等的文库复杂性和大致相等的重复率,但包括酶42B的反应产生偏向于较长基因的文库。

[0342] 实施例7.在单细胞全长成对V(D)J转录剖析的情况下分析经工程改造MMLV RT酶变体

[0343] 收集淋巴细胞,并且分配至小滴乳液中,以使得产生包含单一细胞、包含条形码寡核苷酸的单一凝胶珠粒、和如在本文中其他地方所述的用于逆转录的试剂的小滴。使用经工程改造MMLV RT酶42B(SEQ ID NO:5)或类似可商购获得的MMLV RT酶(CA-MMLV)来产生条形码化cDNA(参见例如图11和随附正文)以分析酶42B在表征淋巴细胞T细胞受体(TCR)α和β链方面的性能。

[0344] 表8.根据靶向细胞回收的有效配对

[0345]

| 靶向细胞回收 | RT酶 | 具有有效V-J跨越(TRA、TRB)配对的细胞 |
|---|---|---|
| 1K T细胞 | CA-MMLV | 24.5% |
| | 42B | 39.4% |
| 6k T细胞 | CA-MMLV | 24.0% |
| | 42B | 44.7% |

[0346] 如表8中所见，相较于CA-MMLV，酶42B显示对低细胞计数的灵敏性改进。另外，如图21和图22中所见，含有酶42B的样品展现较高分数的具有全长有效TCR-α和TCR-β配对的细胞，同时展现较低部分pR1多联体副产物。总之，由于检测到的TCR mRNA分子的数目较大以及引物多联体的背景降低，所以使用酶42B导致TCR组装极大增强。

[0347] 实施例8.在转录剖析的情况下分析经工程改造MMLV RT酶变体

[0348] 收集人PBMC，并且分配至小滴乳液中，以使得产生包含单一细胞、包含条形码寡核苷酸的单一凝胶珠粒、和如本文先前所述的用于逆转录的试剂的小滴。使用经工程改造MMLV RT酶42B(SEQ ID NO:5)或类似可商购获得的MMLV RT酶(CA-MMLV)来产生条形码化cDNA以在3’逆转录测定(参见例如图9A和随附正文)中以及在5’逆转录测定(参见例如图11和随附正文)中分析酶42B的性能。包含可释放条形码化寡核苷酸(其包含多dT序列(3’测定)或模板转换序列(5’测定))的凝胶珠粒用于产生全长cDNA，接着如在本文中其他地方所述将所述cDNA汇合和处理以用于下一代测序和分析。

[0349] 表9.在3’测定中在CA-MMLV与突变酶42B之间的比较

[0350]

| | CA-MMLV 酶 | | 突变 42B 酶 | |
|---|---|---|---|---|
| 描述 | SC3'v2-Maxima-重复实验 1 | SC3'v2-Maxima-重复实验 2 | SC3'v2-42B-重复实验 1 | SC3'v2-42B-重复实验 2 |
| 细胞装载量 | 1000 | 1000 | 1000 | 1000 |
| 每个细胞的平均读段 | 257,168 | 288,724 | 260,854 | 256,403 |
| 有效条形码 | 98.30% | 98.30% | 98.30% | 98.20% |
| 确信定位于转录物组的读段 | 59.20% | 59.80% | 52.50% | 52.70% |
| 确信定位于基因间区域的读段 | 4.20% | 4.00% | 4.70% | 4.70% |
| 确信定位于内含子区域的读段 | 21.30% | 20.80% | 25.60% | 25.50% |
| 确信定位于外显子区域的读段 | 62.30% | 63.00% | 55.00% | 55.20% |
| 反义定位于基因的读段 | 1.10% | 1.00% | 0.90% | 0.90% |
| rRNA 读段分数 | 0.10% | 0.10% | 0.10% | 0.10% |
| mtRNA 读段分数 | 2.40% | 2.30% | 6.30% | 6.60% |
| 未定位读段分数 | 9.20% | 9.30% | 12.00% | 11.90% |
| 每个细胞的中值基因(每个细胞 50k 原始读段) | 987 | 1,016 | 1,311 | 1,343 |
| 每个细胞的中值基因(每个细胞 50k 经定位细胞读段) | 1,034 | 1,064 | 1,392 | 1,428 |
| 每个细胞的中值 UMI 计数(每个细胞 50k 原始读段) | 2,914 | 3,042 | 3,530 | 3,591 |
| 每个细胞的中值 UMI 计数(每个细胞 50k 经定位细胞读段) | 3,057 | 3,182 | 3,783 | 3,830 |
| 检测到的>1500 nt 的总基因(每个细胞 50k 原始读段) | 11,411 | 11,359 | 11,803 | 11,880 |
| 检测到的>1500 nt 的总基因(每个细胞 50k 经定位细胞读段) | 11,587 | 11,542 | 12,045 | 12,110 |
| <500 nt 的基因的 UMI 计数的分数 | 4.00% | 3.60% | 3.20% | 3.40% |
| 500-1000 nt 基因的 UMI 计数的分数 | 31.70% | 31.70% | 25.00% | 25.00% |
| 1000-1500 nt 基因的 UMI 计数的分数 | 24.90% | 24.80% | 20.00% | 20.20% |
| >1500 nt 的基因的 UMI 计数的分数 | 39.40% | 40.00% | 51.80% | 51.50% |
| 核糖体蛋白质 UMI 计数的分数 | 37.80% | 37.80% | 23.20% | 22.80% |

[0351] 表10.在5’测定中在CA-MMLV与突变酶42B之间的比较

[0352]

| | CA-MMLV 酶 | | 突变 42B 酶 | |
|---|---|---|---|---|
| 样品 ID | 41621 | 41622 | 41623 | 41624 |

[0353]

| 描述 | SC5'-Maxima-重复实验 1 | SC5'-Maxima-重复实验 2 | SC5'-42B-重复实验 1 | SC5'-42B-重复实验 2 |
|---|---|---|---|---|
| 每个细胞的平均读段 | 280,128 | 269,343 | 280,848 | 257,745 |
| 有效条形码 | 91.30% | 91.60% | 84.80% | 86.60% |
| 确信定位于转录物组的读段 | 58.80% | 60.20% | 50.00% | 52.70% |
| 确信定位于基因间区域的读段 | 3.70% | 3.70% | 7.60% | 6.70% |
| 确信定位于内含子区域的读段 | 11.60% | 11.30% | 9.60% | 9.60% |
| 确信定位于外显子区域的读段 | 67.50% | 68.70% | 58.30% | 60.50% |
| 反义定位于基因的读段 | 5.60% | 5.40% | 5.40% | 4.90% |
| rRNA 读段分数 | 2.10% | 2.00% | 6.50% | 5.60% |
| mtRNA 读段分数 | 0.80% | 0.80% | 2.80% | 2.30% |
| 未定位读段分数 | 10.80% | 9.90% | 10.10% | 10.40% |
| 每个细胞的中值基因(每个细胞 50k 原始读段) | 602 | 651 | 1,365 | 1,367 |
| 每个细胞的中值基因(每个细胞 50k 经定位细胞读段) | 631 | 675 | 1,470 | 1,460 |
| 每个细胞的中值 UMI 计数(每个细胞 50k 原始读段) | 1,341 | 1,452 | 3,744 | 3,825 |
| 每个细胞的中值 UMI 计数(每个细胞 50k 经定位细胞读段) | 1,399 | 1,517 | 4,111 | 4,146 |
| 检测到的>1500 nt 的总基因(每个细胞 50k 原始读段) | 9,451 | 9,522 | 10,567 | 10,646 |
| 检测到的>1500 nt 的总基因(每个细胞 50k 经定位细胞读段) | 9,562 | 9,631 | 10,731 | 10,801 |
| <500 nt 的基因的 UMI 计数的分数 | 3.90% | 4.60% | 4.30% | 4.00% |
| 500-1000 nt 基因的 UMI 计数的分数 | 37.70% | 37.60% | 34.60% | 34.60% |
| 1000-1500 nt 基因的 UMI 计数的分数 | 25.50% | 25.20% | 25.00% | 24.90% |
| >1500 nt 的基因的 UMI 计数的分数 | 32.90% | 32.70% | 36.10% | 36.50% |
| 核糖体蛋白质 UMI 计数的分数 | 36.60% | 36.30% | 30.60% | 30.20% |

[0354] 如表9和表10中所见，在3’测定形式与5’测定形式两者中，经工程改造RT酶42B均显示灵敏性增加，如由对每个细胞检测的中值UMI和对每个细胞检测的中值基因所度量。另外，酶42B显示在偏向于长基因方面增加，如由长度>1,500个核苷酸的基因的UMI计数的分数所度量。总的来说，数据显示酶42B出乎意料地展现对各种类型的RNA的不同定位率、不同长度偏倚，并且重要的是，相比于类似对应物CA_MMLV，尤其是在用于5’mRNA测定中时，导致产生更复杂的cDNA文库(更多基因/细胞和更多UMI/细胞)。

[0355] 尽管本文已显示和描述本发明的一些实施方案，但将对本领域技术人员明显的是

所述实施方案仅通过举例方式来提供。不意图本发明受限于说明书内提供的特定实例。尽管本发明已参照以上提及的说明书加以描述，但对本文实施方案的描述和说明不意图以限制性意义加以解释。在不脱离本发明下，众多改变、变化和替代现将为本领域技术人员所想到。此外，应了解本发明的所有方面都不限于本文阐述的取决于多种条件和变量的特定描绘、配置或相对比例。应了解本文所述的本发明的实施方案的各种替代方案可用于实施本发明。因此，预期本发明也将涵盖任何所述替代方案、修改、变化或等效物。意图以下权利要求限定本发明的范围，并且在这些权利要求的范围内的方法和结构以及它们的等效物由所述权利要求涵盖。

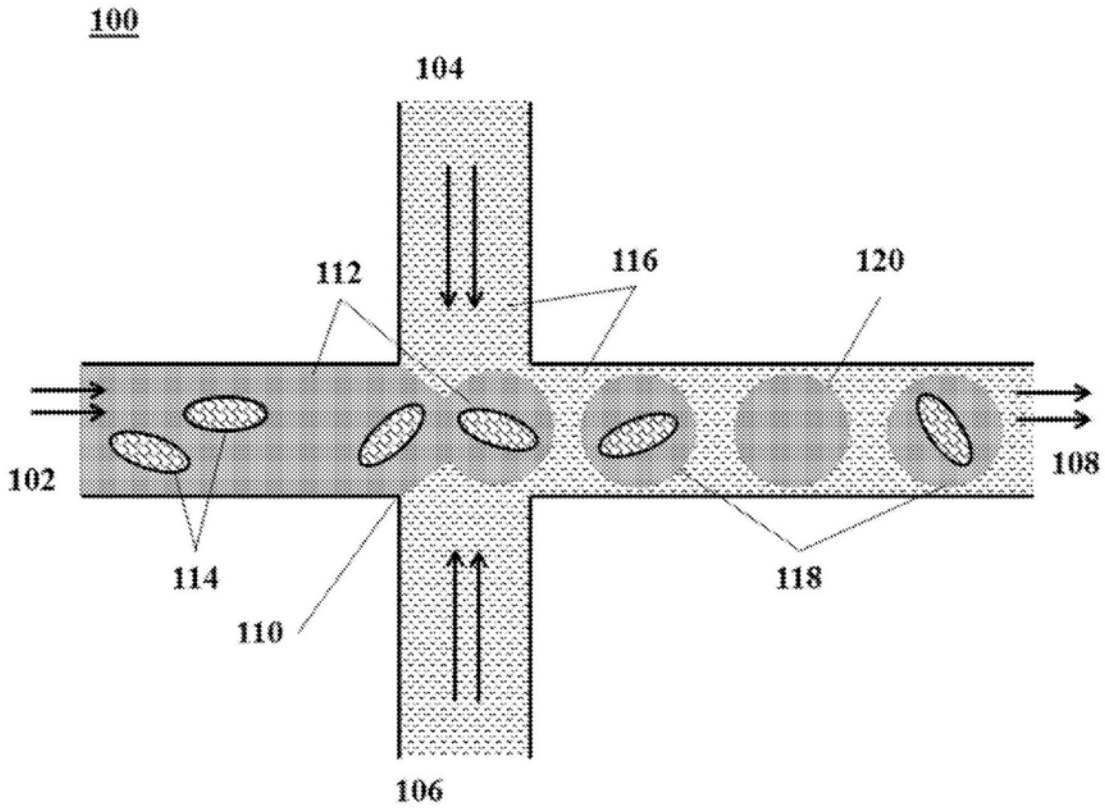

100
104
112
116
120
102
108
114
118
110
106

图1

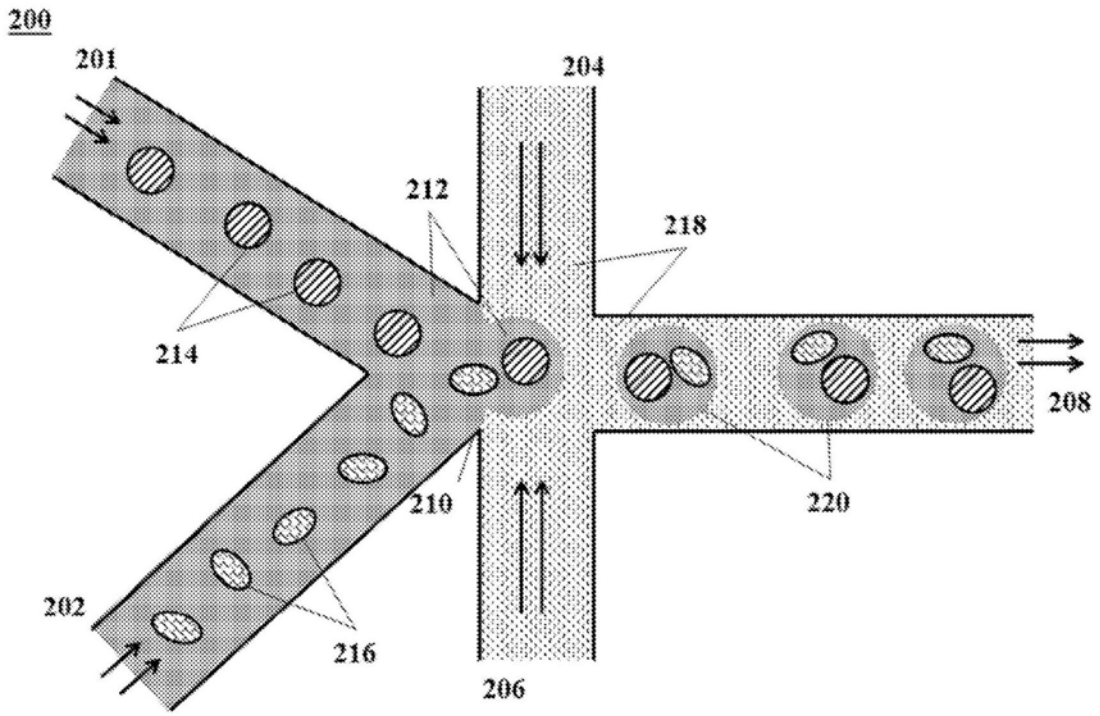

200
201
204
212
218
214
208
210
220
202
216
206

图2

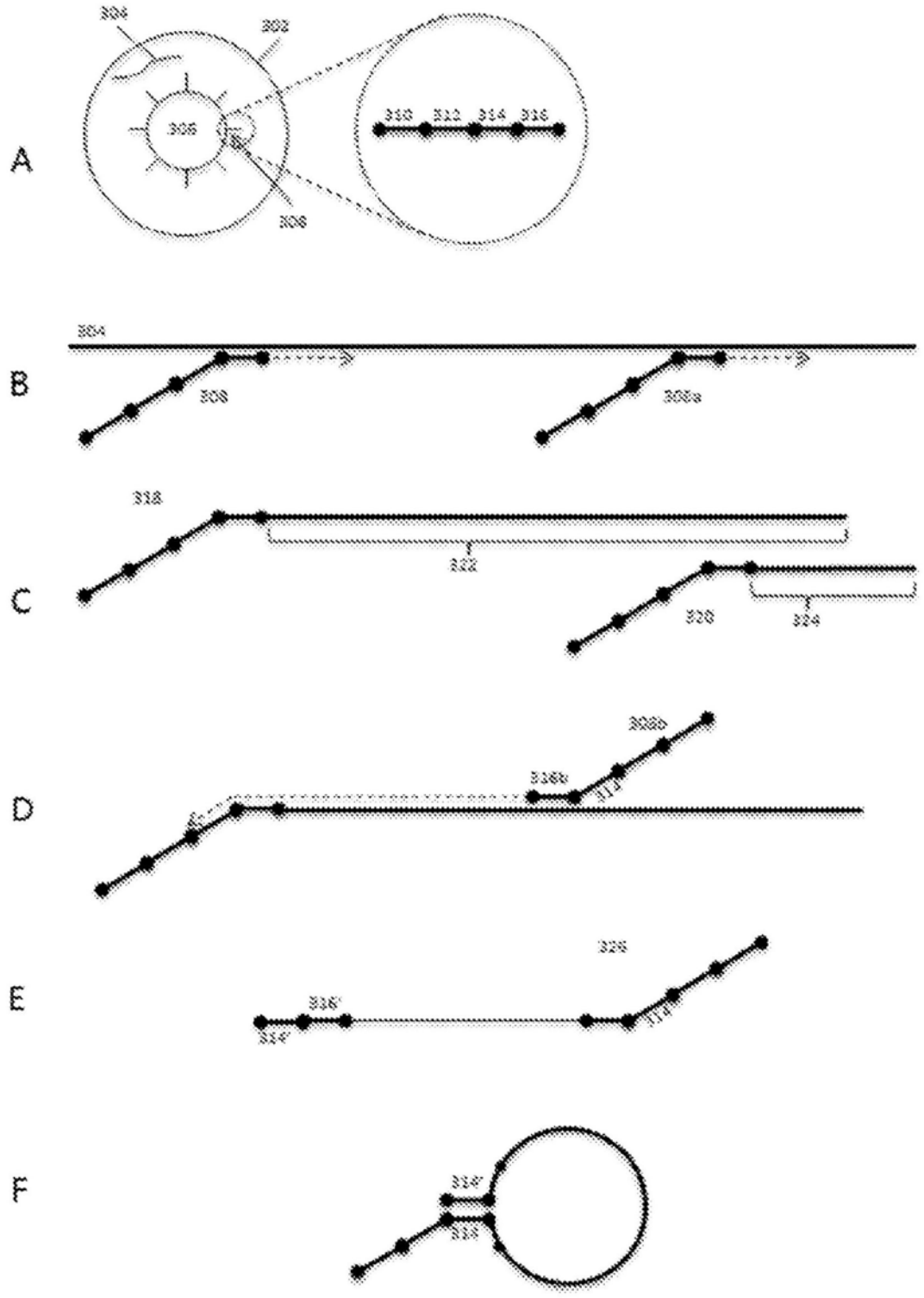

A
304
302
306
310
312
314
316
B
304
308
308a
C
318
322
320
324
D
308b
316b
E
326
316'
314'
F
314'
312'

图3

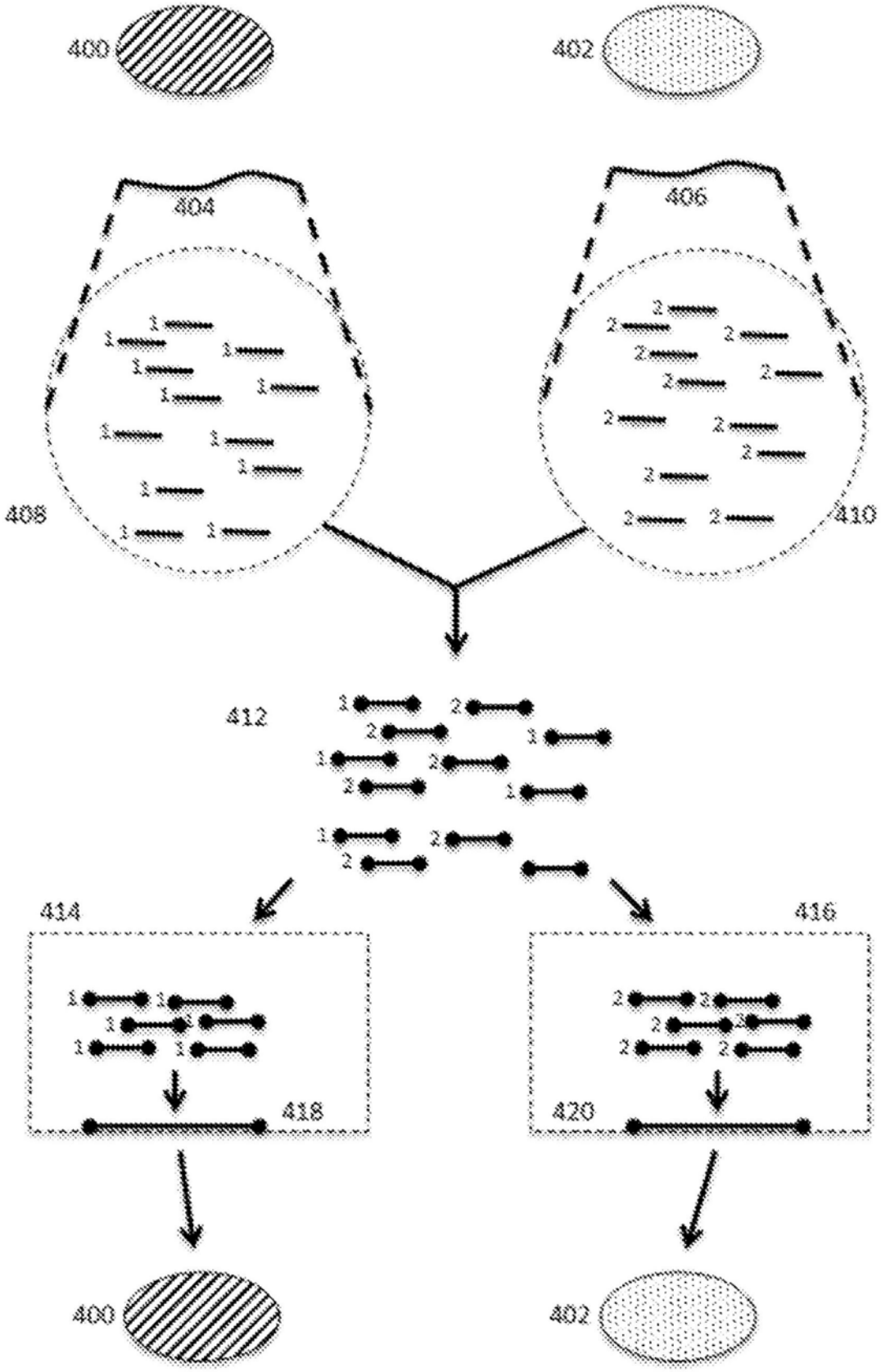



图4

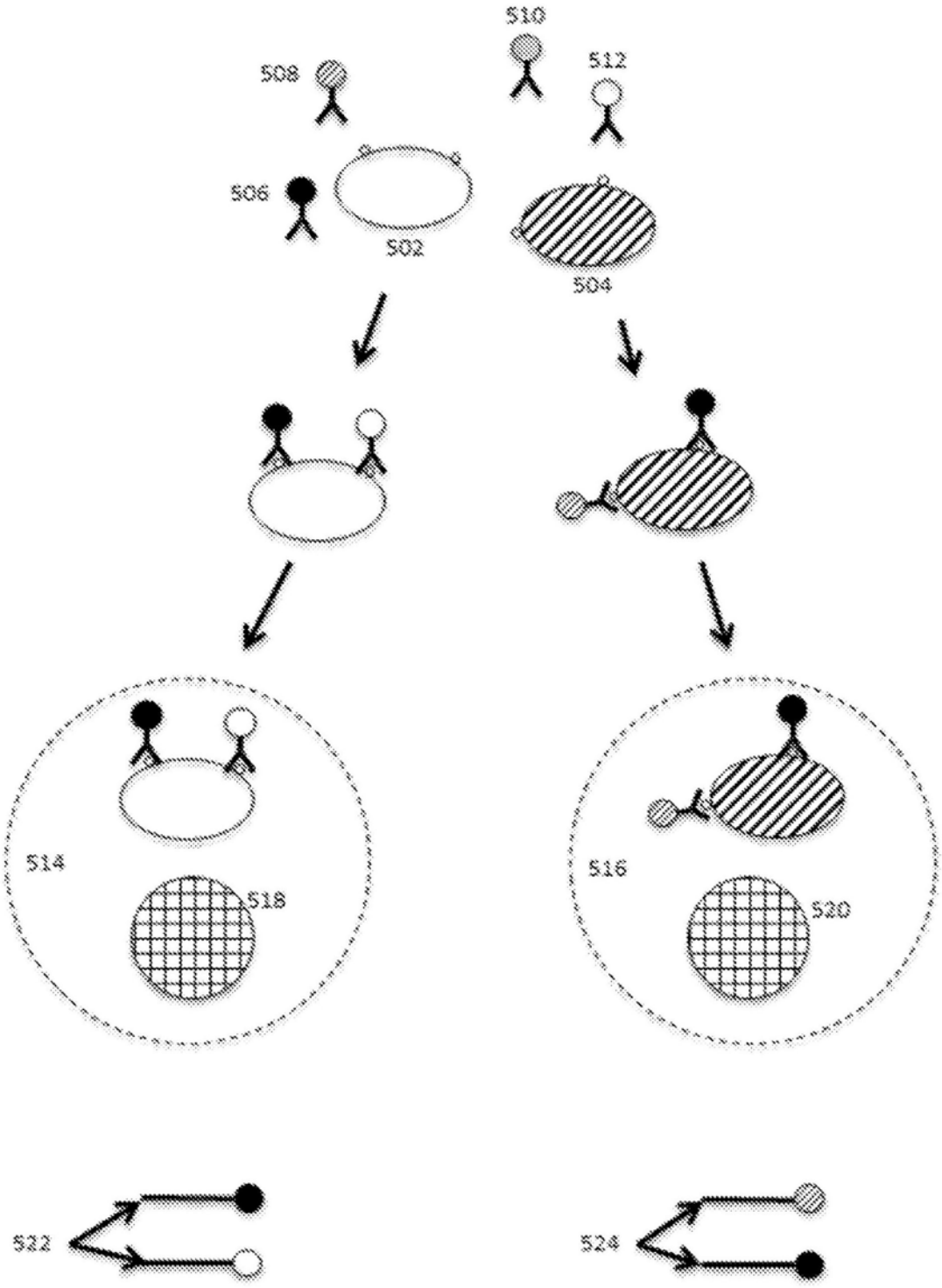



图5

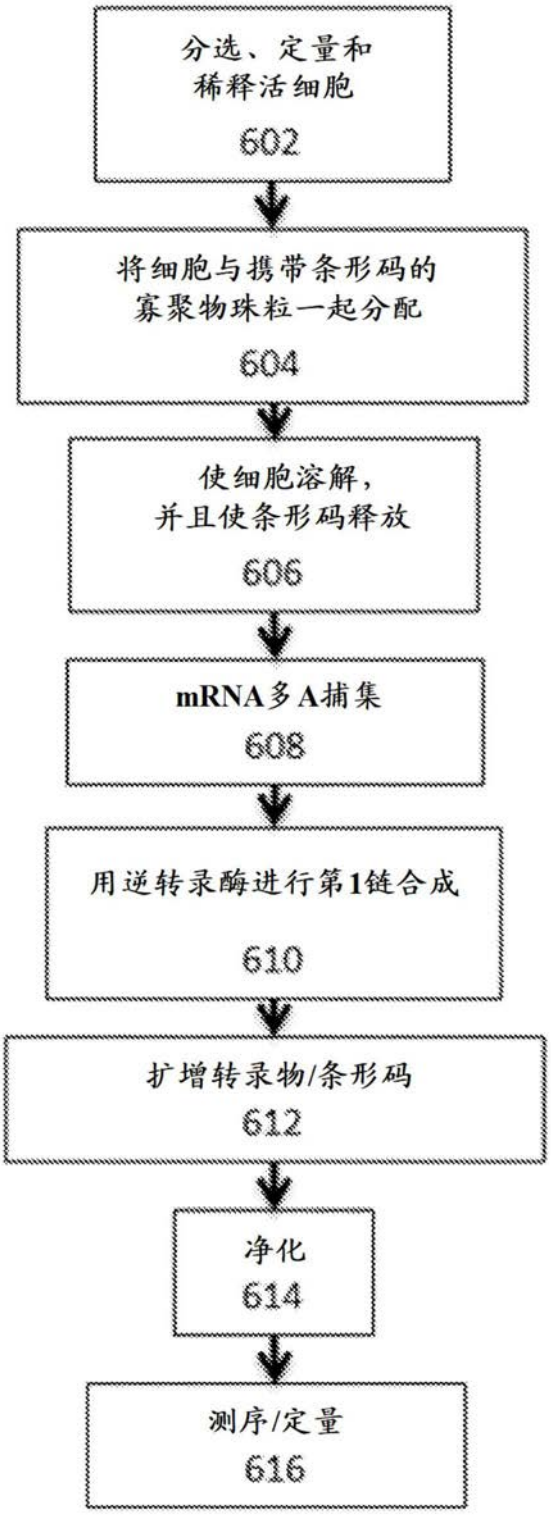

分选、定量和
稀释活细胞
602
将细胞与携带条形码的
寡聚物珠粒一起分配
604
使细胞溶解，
并且使条形码释放
606
mRNA多A捕集
608
用逆转录酶进行第1链合成
610
扩增转录物/条形码
612
净化
614
测序/定量
616

图6

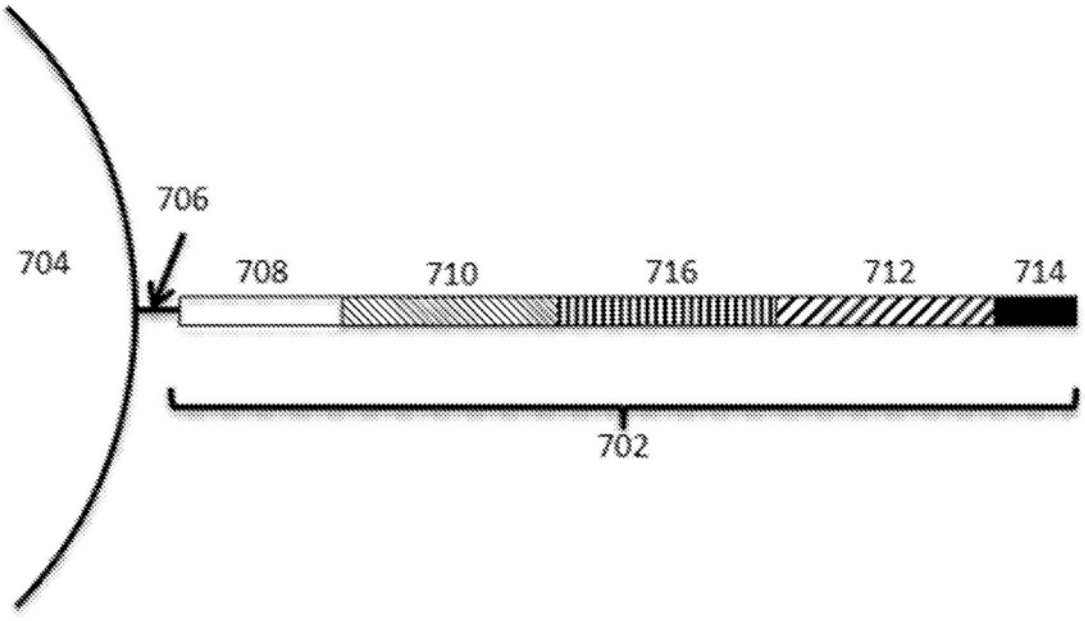



图7

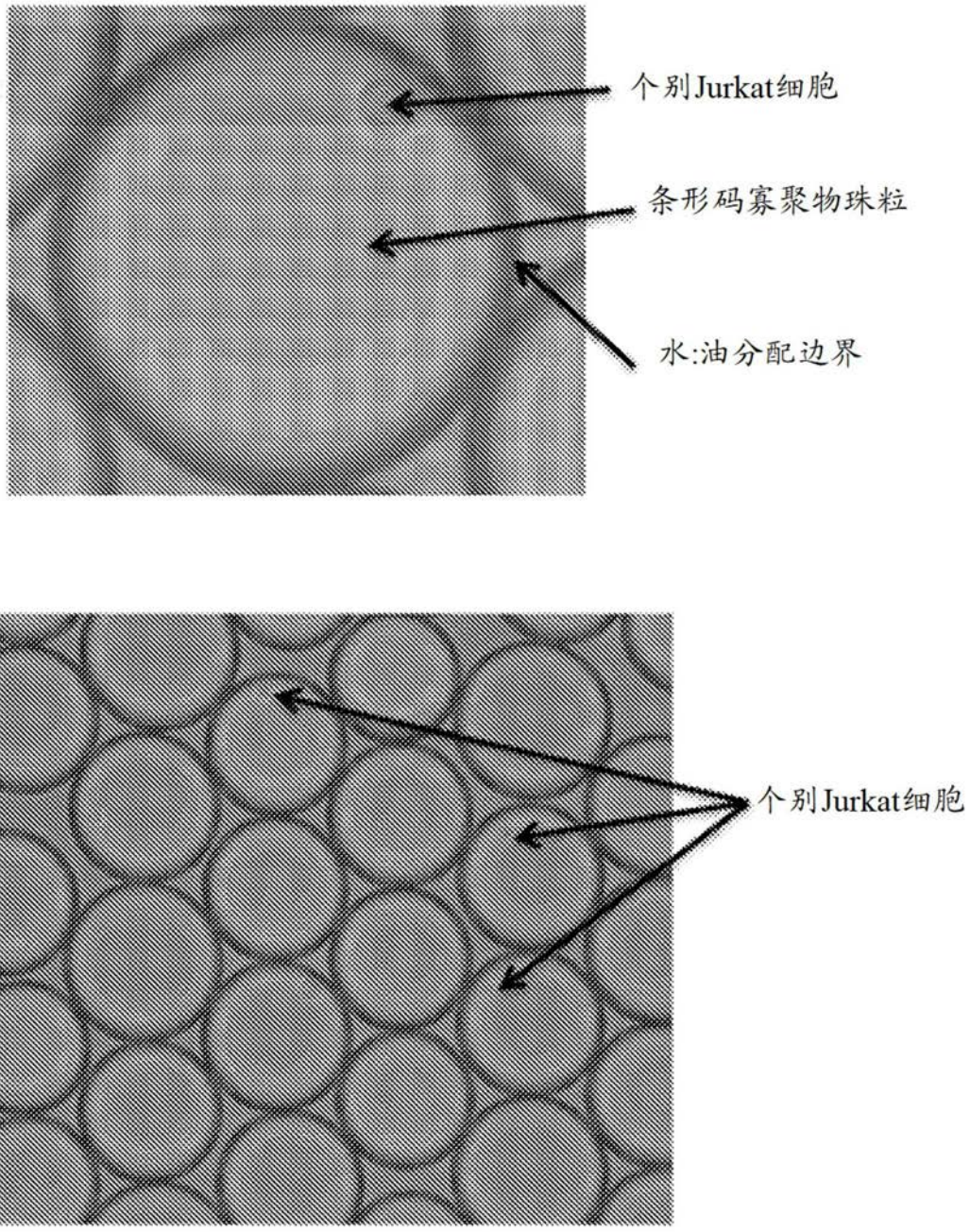



图8

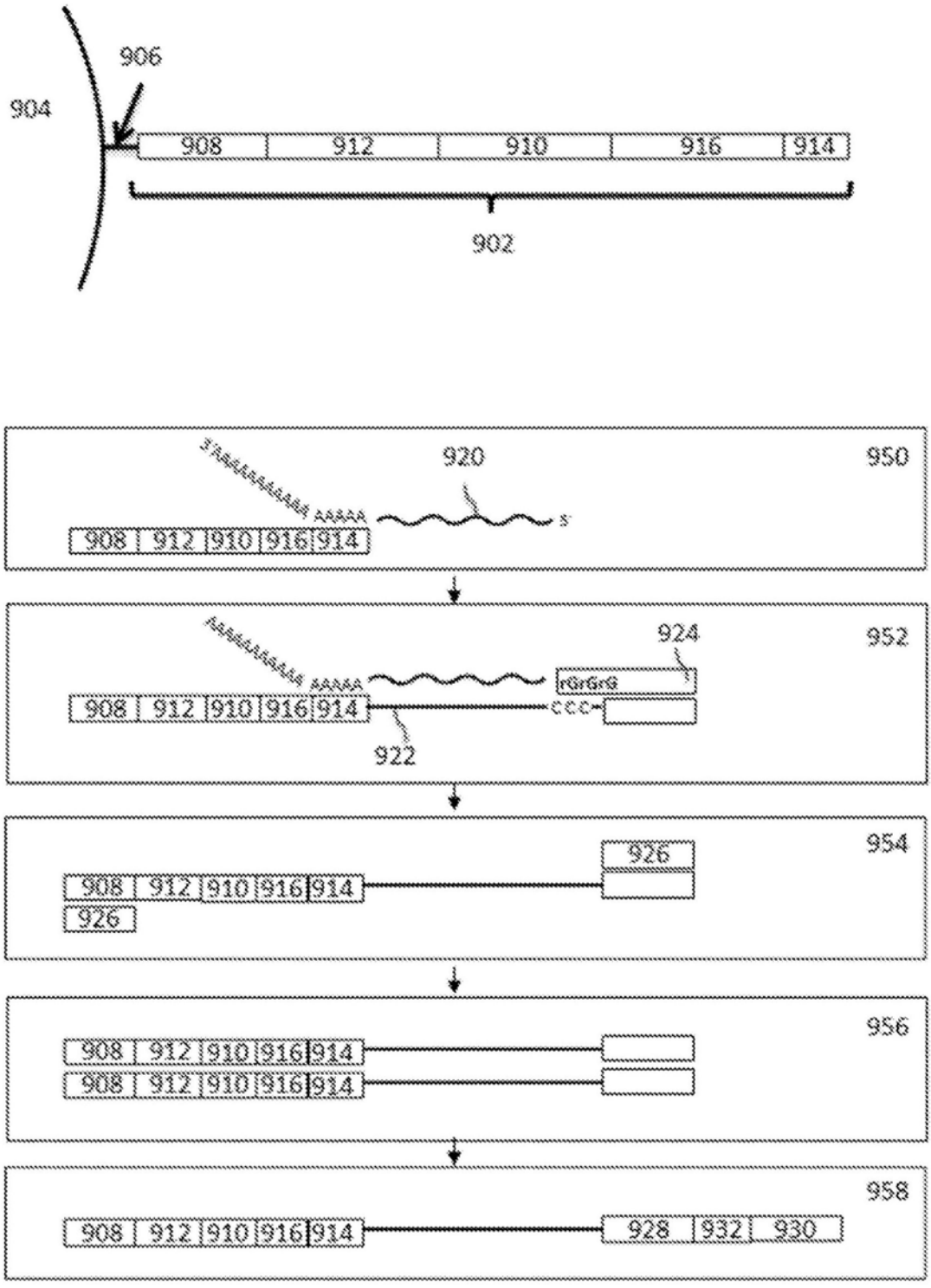

906
904
908
912
910
916
914
902
950
920
AAAAA
908 912 910 916 914
952
924
AAAAA
908 912 910 916 914
922
954
926
908 912 910 916 914
926
956
908 912 910 916 914
908 912 910 916 914
958
908 912 910 916 914
928 932 930

图9A

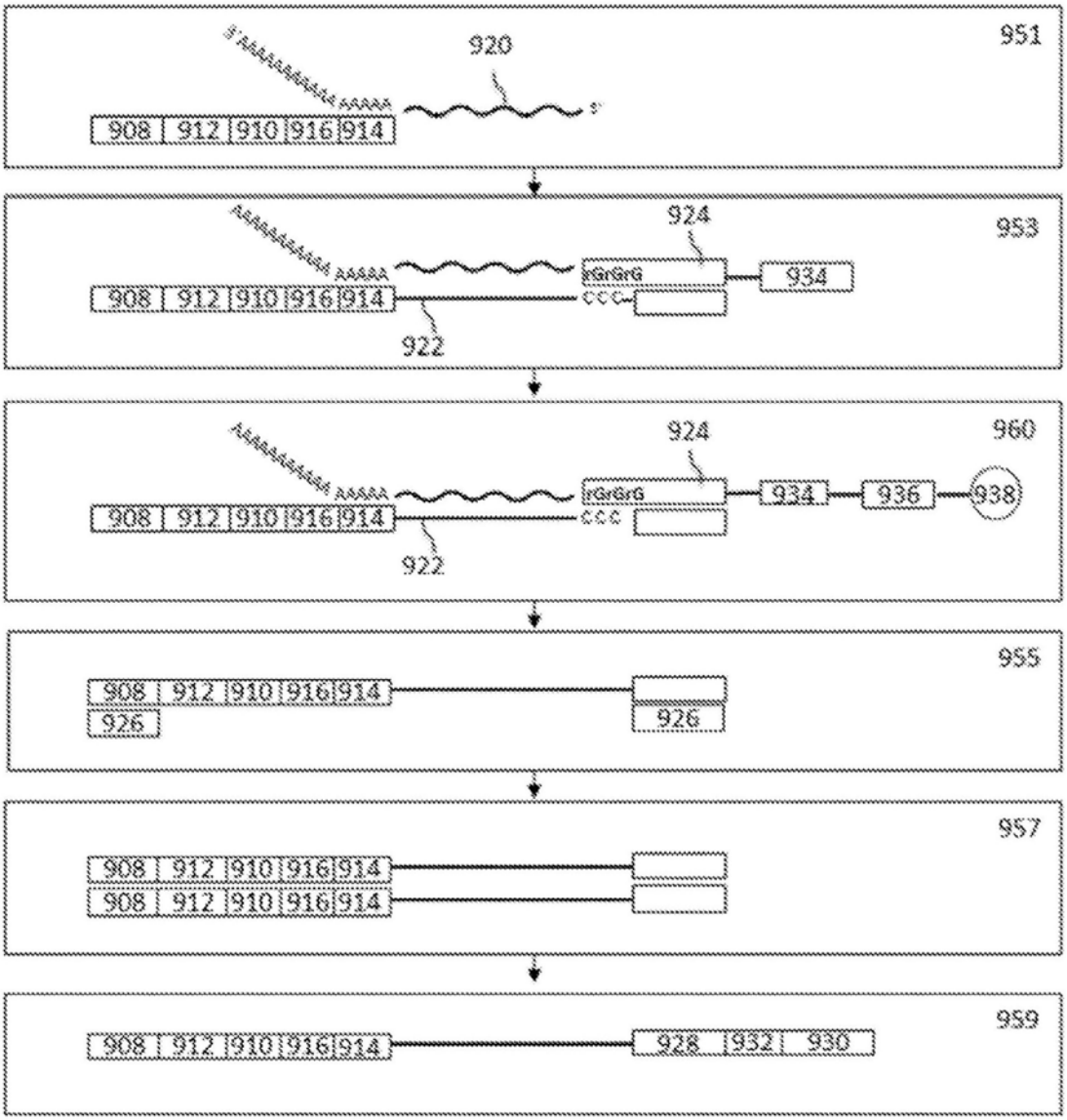

951
920
908
912
910
916
914
953
924
rGrGrG
934
922
960
936
938
955
926
957
959
928
932
930

图9B

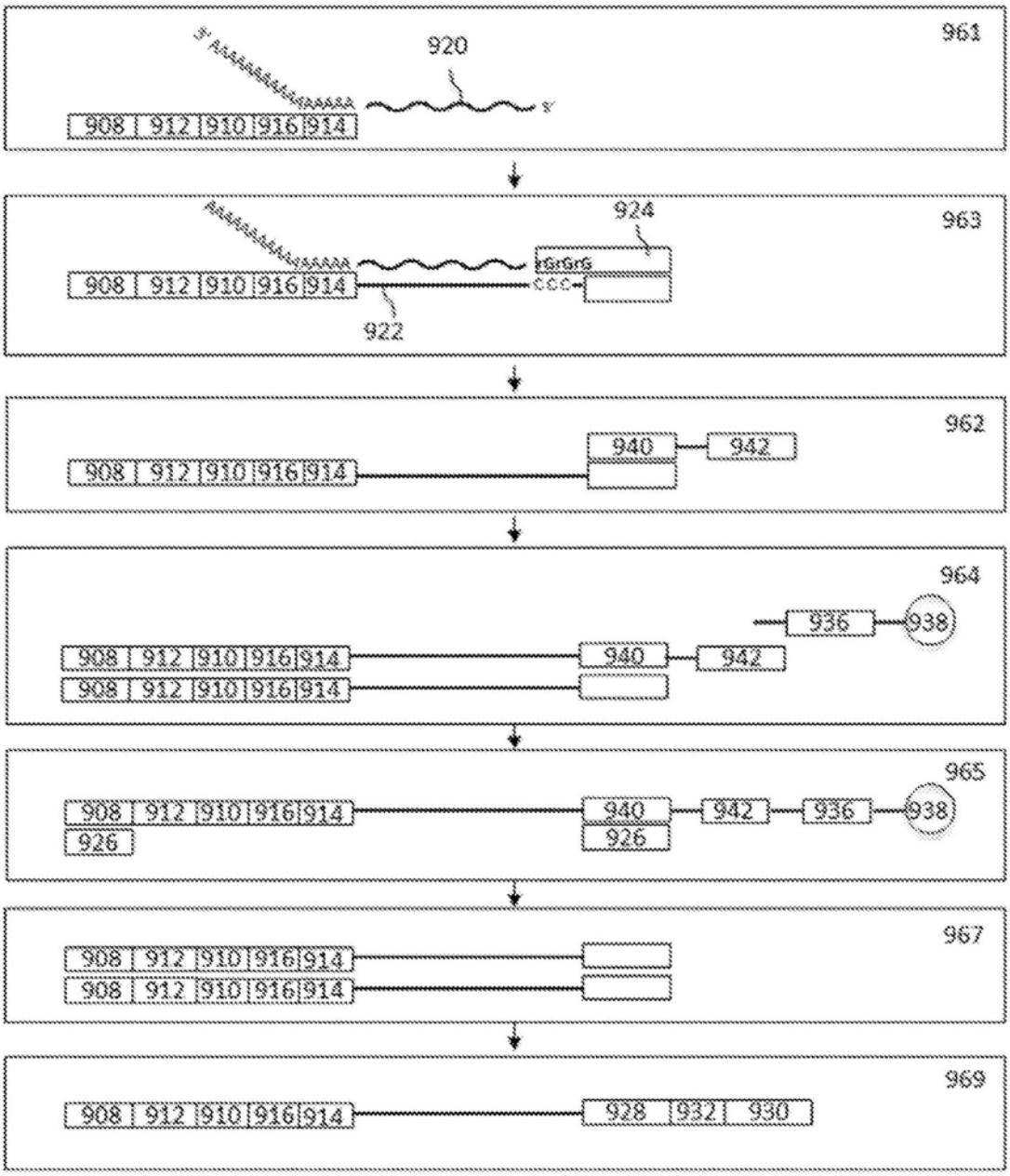

961
920
908 912 910 916 914
963
924
rGrGrG
908 912 910 916 914
922
962
940
942
908 912 910 916 914
964
936
938
908 912 910 916 914
940
942
908 912 910 916 914
965
908 912 910 916 914
926
940
926
942
936
938
967
908 912 910 916 914
908 912 910 916 914
969
908 912 910 916 914
928 932 930

图9C

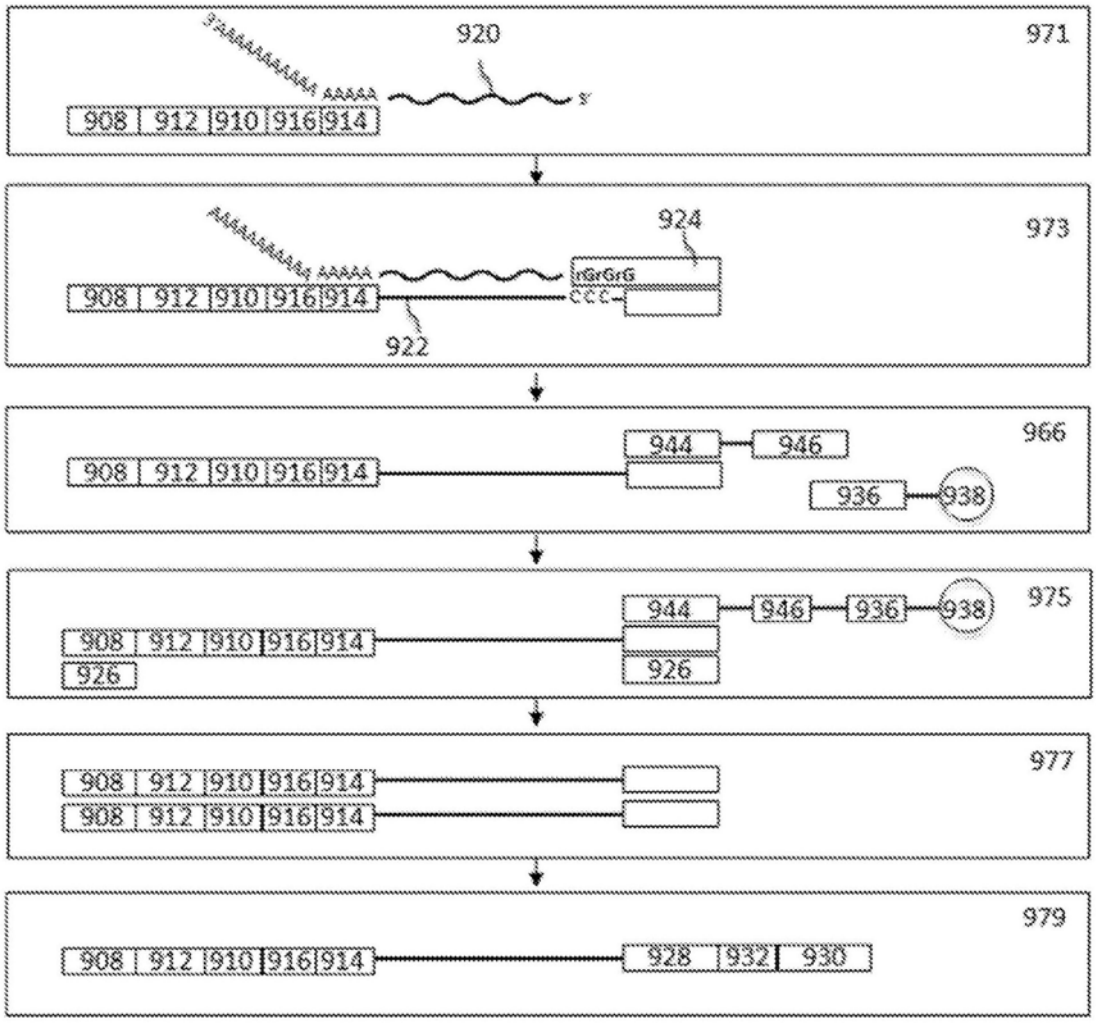

971
920
908 912 910 916 914
973
924
922
rGrGrG
CCC
966
944
946
936
938
975
926
977
979
928
932
930

图9D

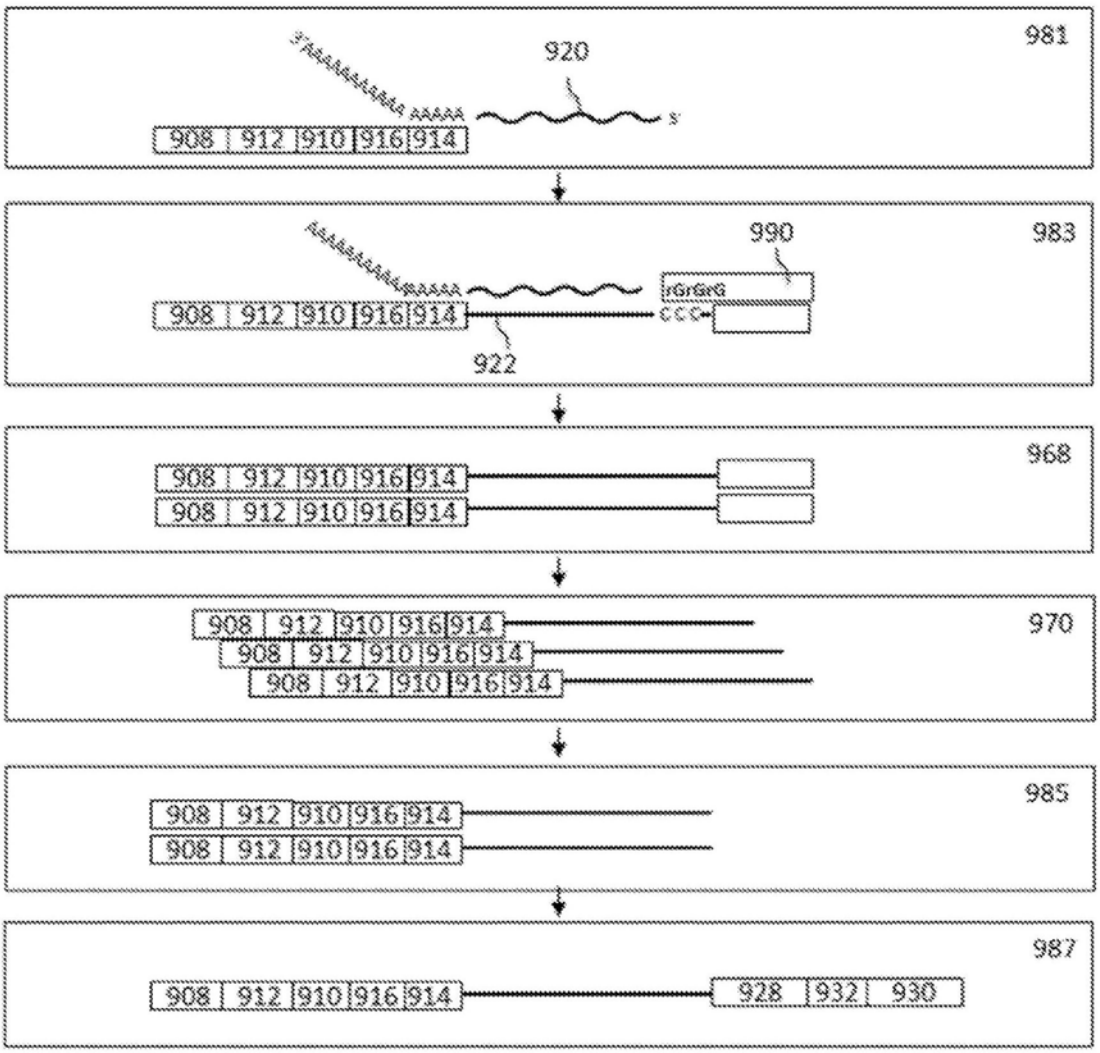

981
920
908 912 910 916 914
983
990
rGrGrG
CCC
922
968
908 912 910 916 914
908 912 910 916 914
970
908 912 910 916 914
908 912 910 916 914
908 912 910 916 914
985
908 912 910 916 914
908 912 910 916 914
987
908 912 910 916 914
928 932 930

图9E

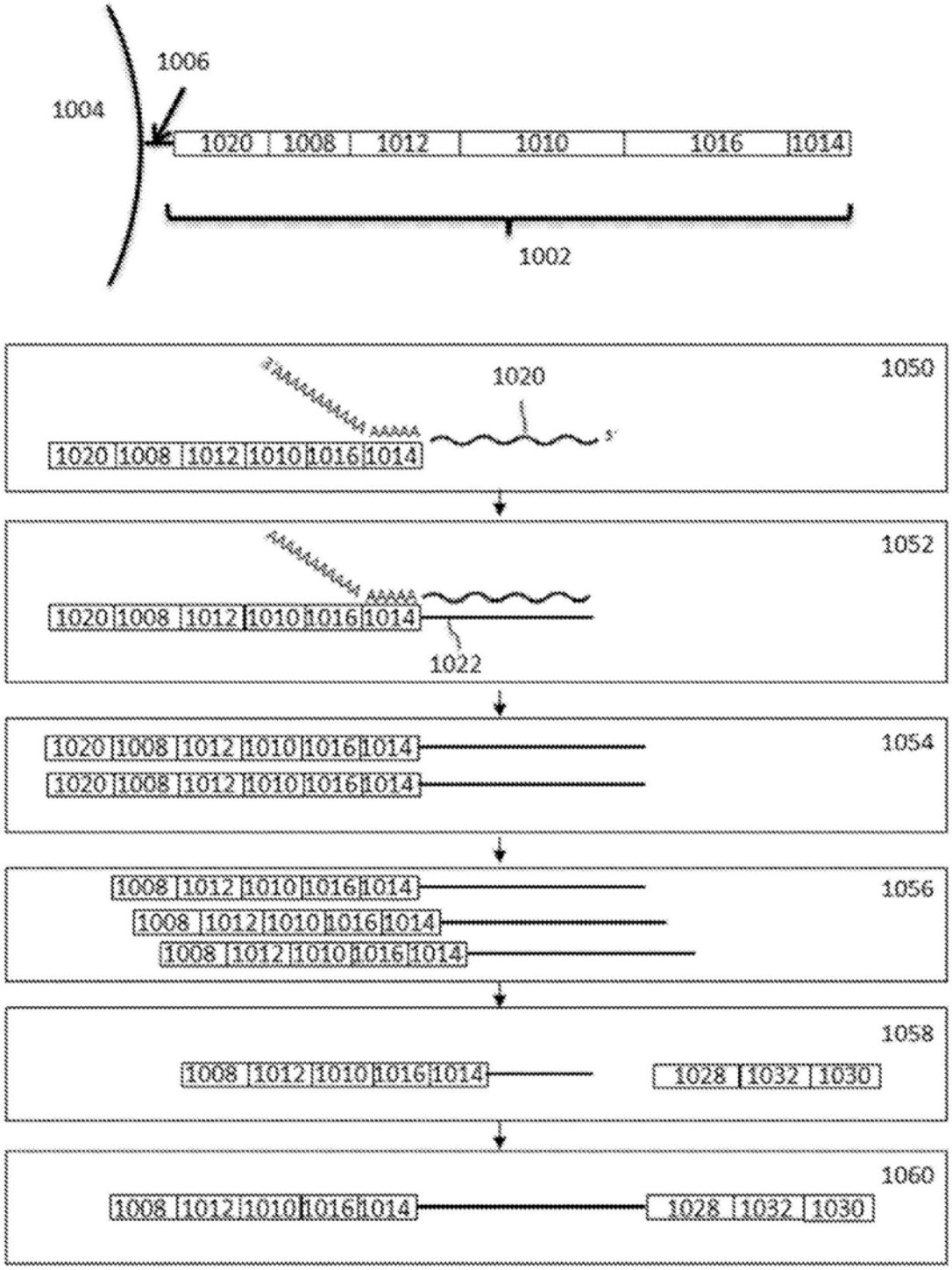

1006
1004
1020
1008
1012
1010
1016
1014
1002
1050
1020
1020 1008 1012 1010 1016 1014
1052
1020 1008 1012 1010 1016 1014
1022
1054
1020 1008 1012 1010 1016 1014
1020 1008 1012 1010 1016 1014
1056
1008 1012 1010 1016 1014
1008 1012 1010 1016 1014
1008 1012 1010 1016 1014
1058
1008 1012 1010 1016 1014
1028 1032 1030
1060
1008 1012 1010 1016 1014
1028 1032 1030

图10

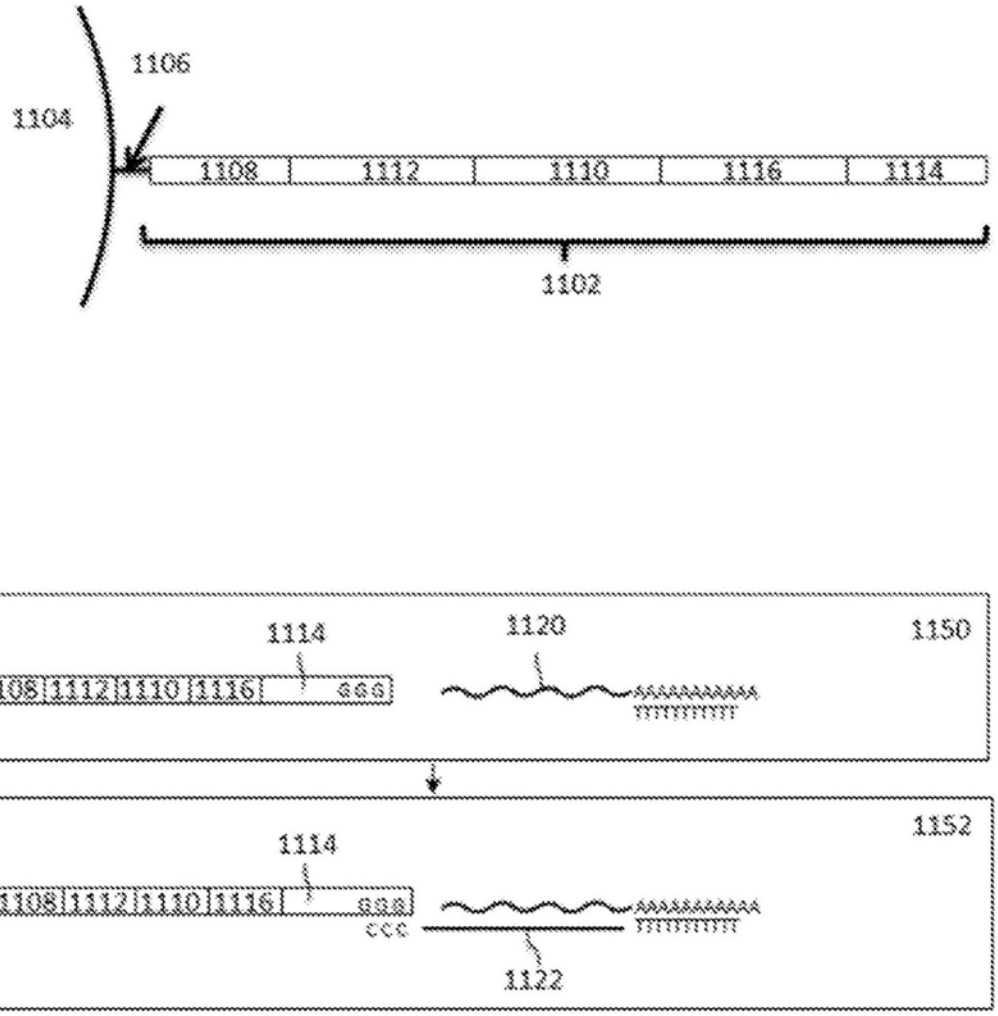

1106
1104
1108
1112
1110
1116
1114
1102
1150
1114
1120
1108
1112
1110
1116
aaa
1152
1114
1108
1112
1110
1116
aaa
ccc
1122

图11

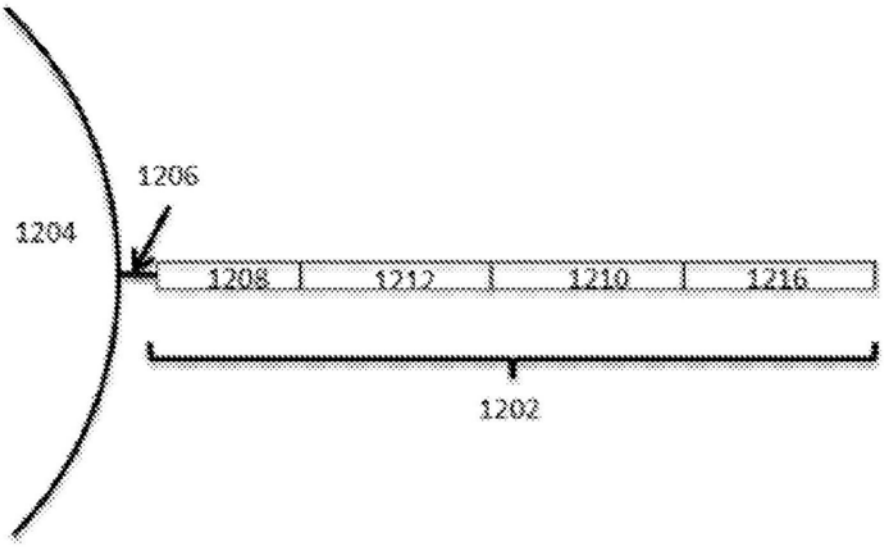

1206
1204
1208
1212
1210
1216
1202

图12A

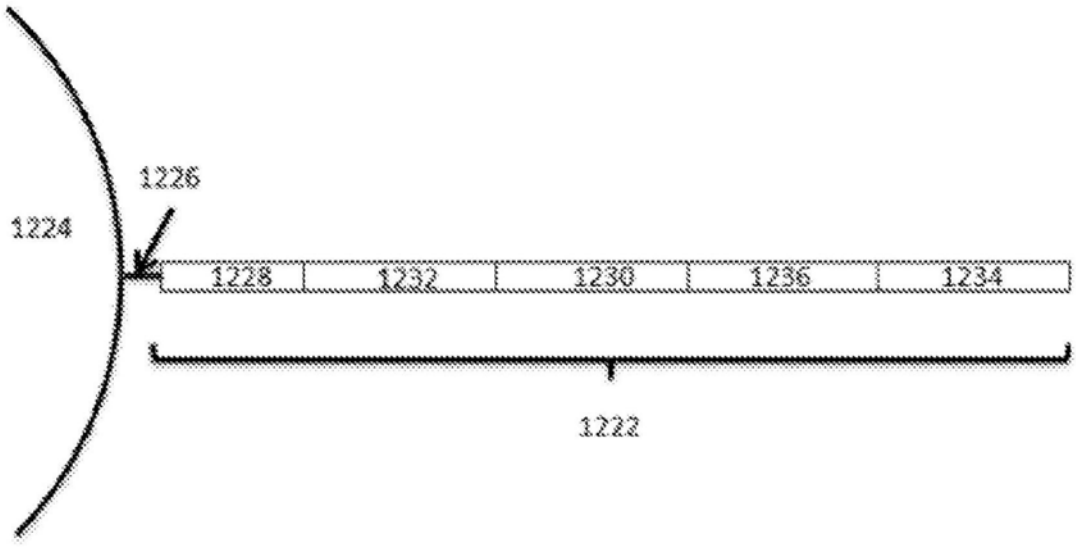



图12B

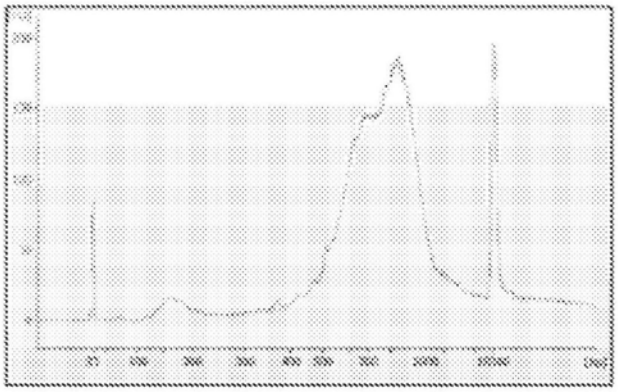


图13A

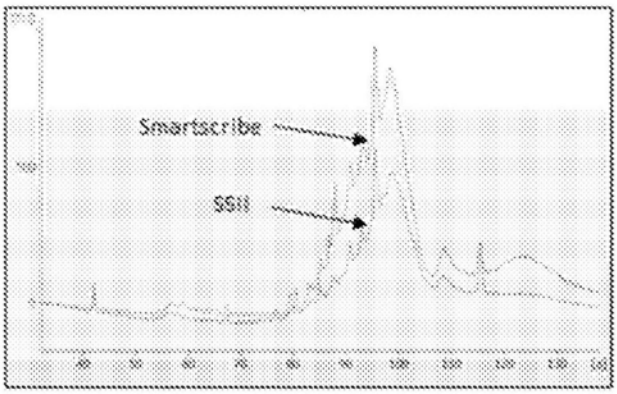



图13B

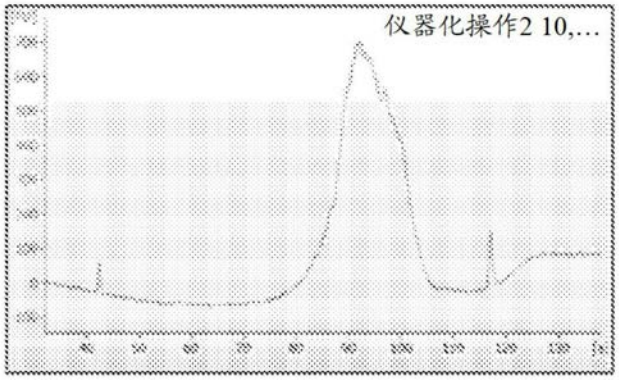

仪器化操作2 10,...

图13C

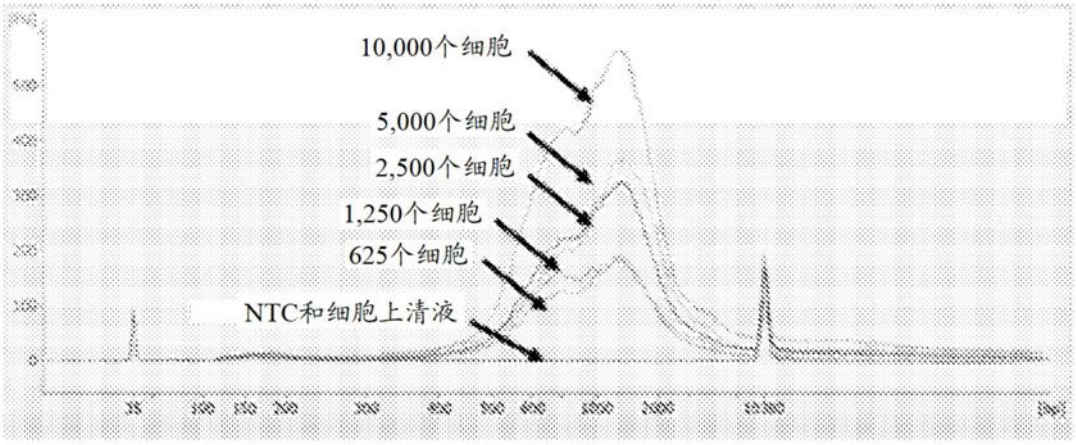

10,000个细胞
5,000个细胞
2,500个细胞
1,250个细胞
625个细胞
NTC和细胞上清液

图14A

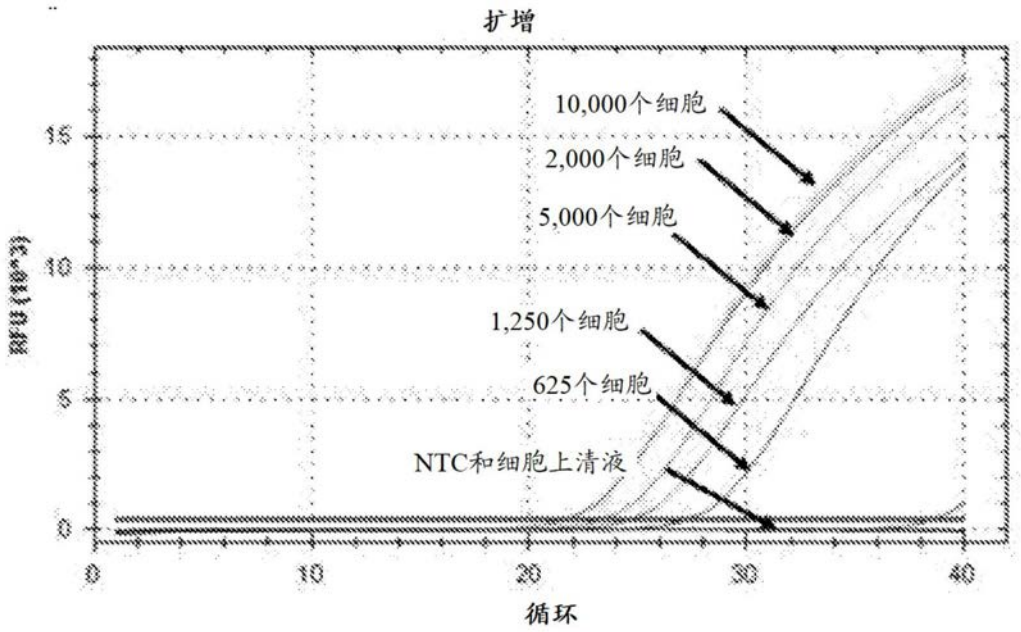

扩增
10,000个细胞
2,000个细胞
5,000个细胞
1,250个细胞
625个细胞
NTC和细胞上清液
循环

图14B

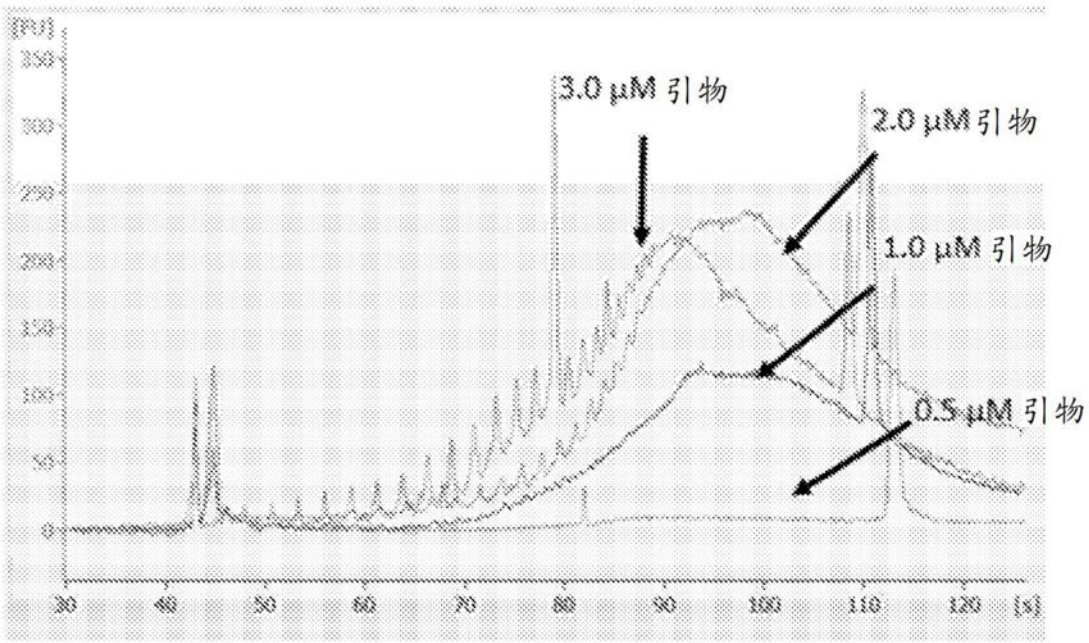

3.0 μM 引物
2.0 μM 引物
1.0 μM 引物
0.5 μM 引物

图15

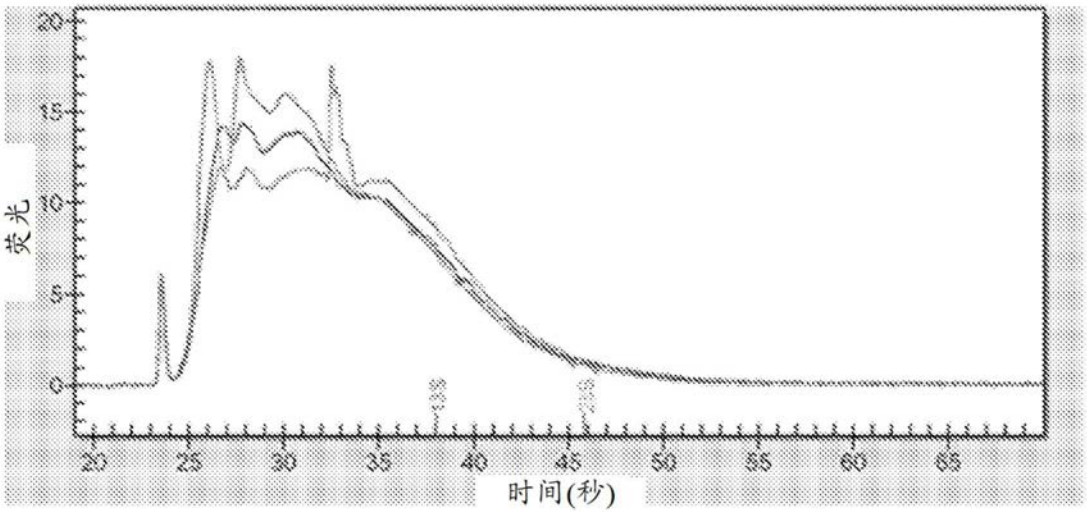

荧光
0
5
10
15
20
20
25
30
35
40
45
50
55
60
65
时间(秒)

图16

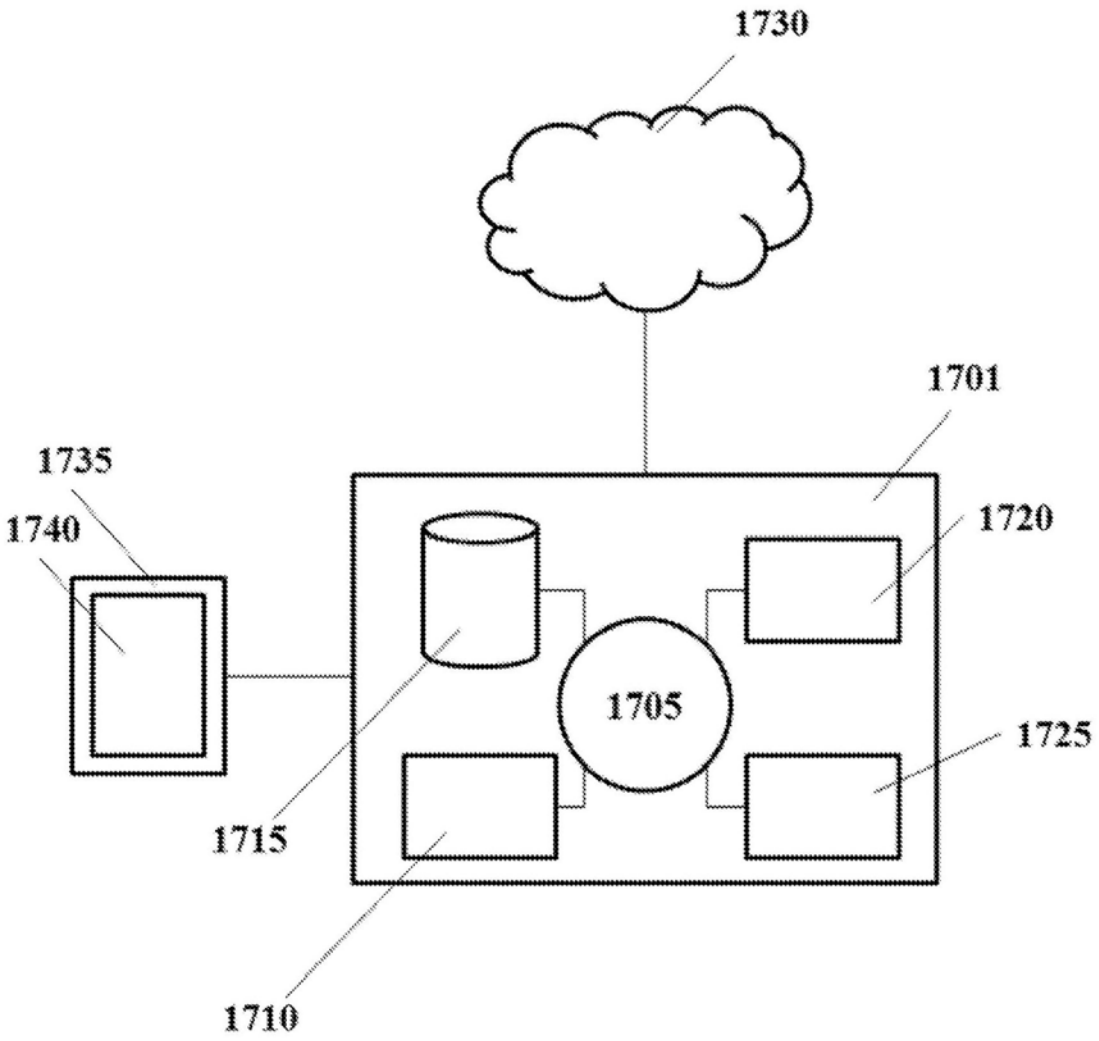

1730
1701
1735
1740
1720
1705
1725
1715
1710

图17

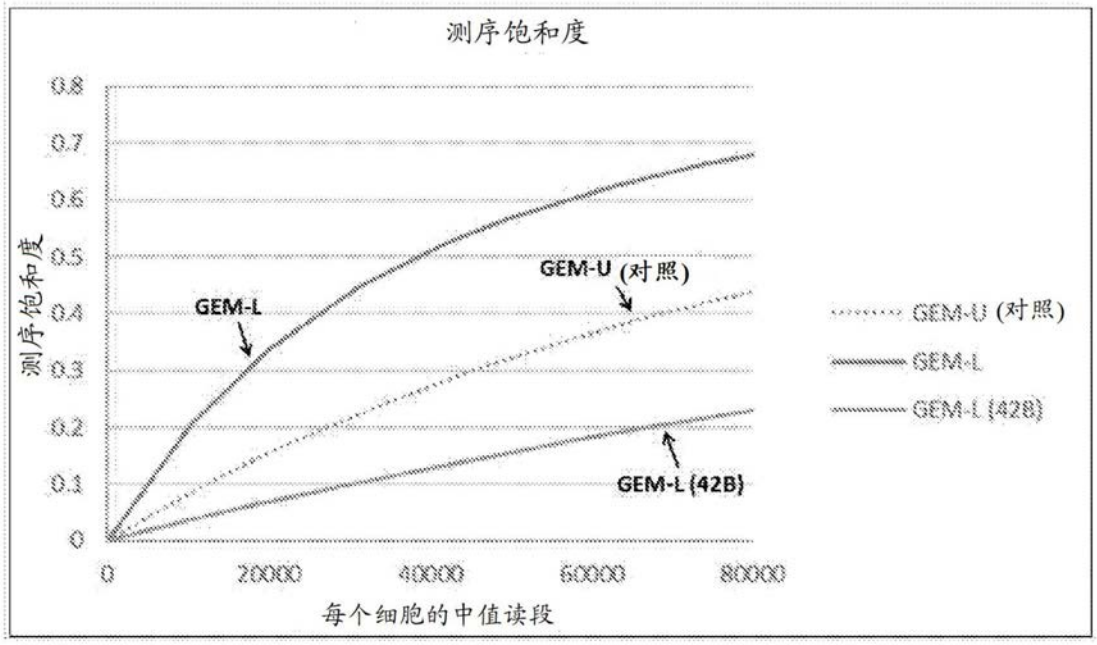

测序饱和度
测序饱和度
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
20000
40000
60000
80000
每个细胞的中值读段
GEM-L
GEM-U (对照)
GEM-L (42B)
GEM-U (对照)
GEM-L
GEM-L (42B)

图18

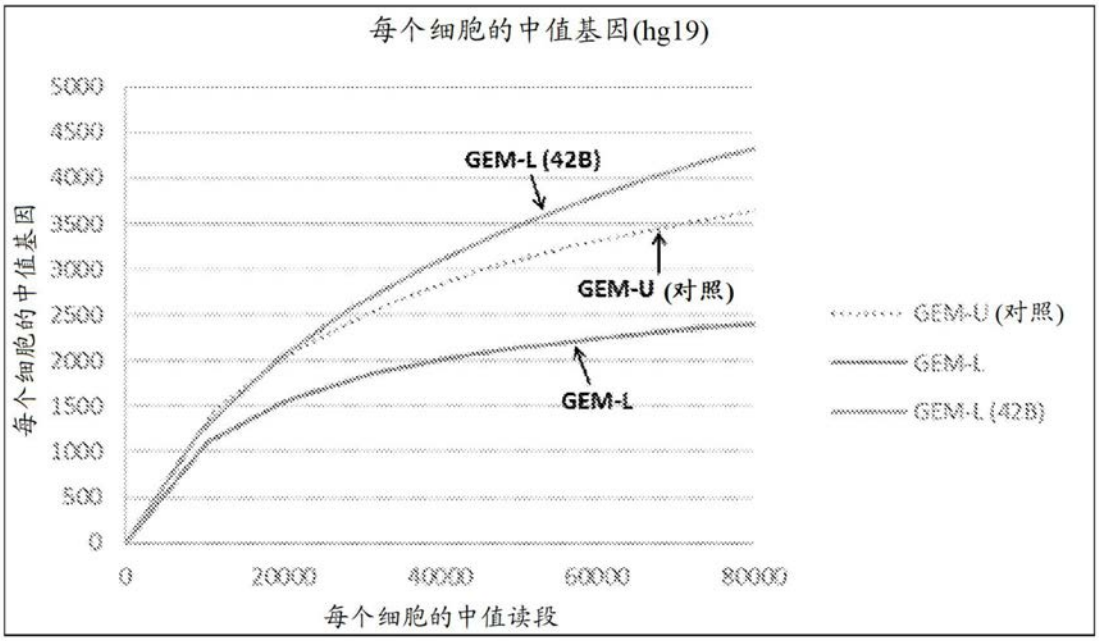

每个细胞的中值基因(hg19)
每个细胞的中值基因
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
0
20000
40000
60000
80000
每个细胞的中值读段
GEM-L (42B)
GEM-U (对照)
GEM-L
GEM-U (对照)
GEM-L
GEM-L (42B)

图19

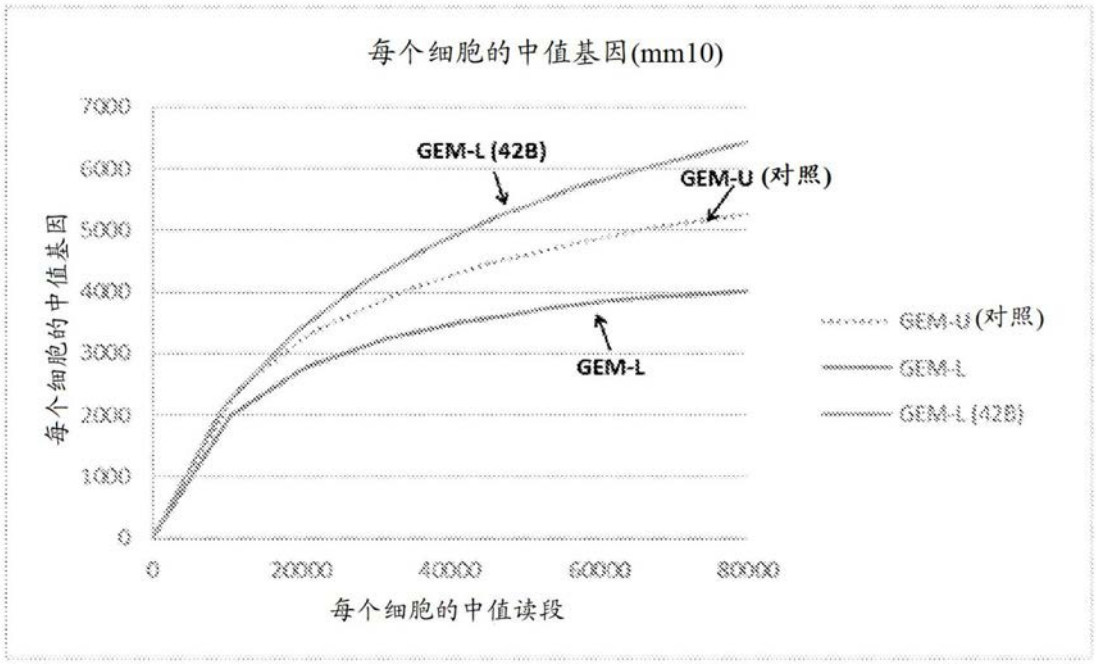

每个细胞的中值基因(mm10)
每个细胞的中值基因
每个细胞的中值读段
GEM-L (42B)
GEM-U (对照)
GEM-L
GEM-U (对照)
GEM-L
GEM-L (42B)
7000
6000
5000
4000
3000
2000
1000
0
0
20000
40000
60000
80000

图20

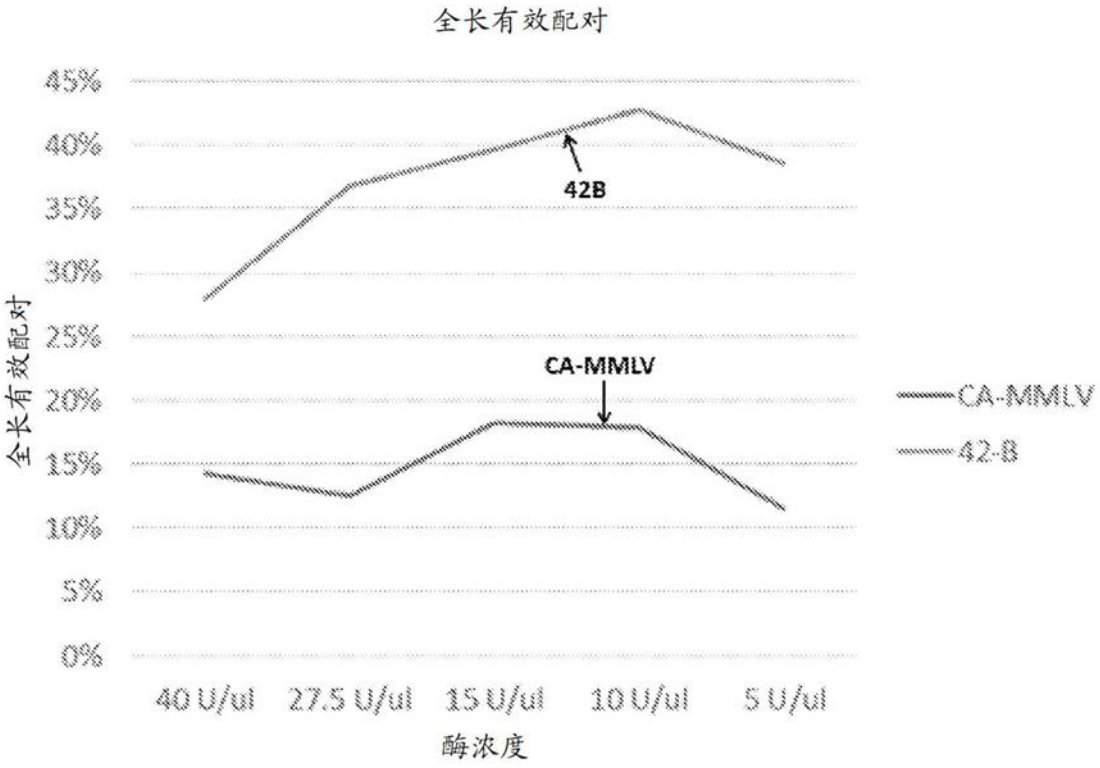

全长有效配对
全长有效配对
酶浓度
42B
CA-MMLV
CA-MMLV
42-B
45%
40%
35%
30%
25%
20%
15%
10%
5%
0%
40 U/ul
27.5 U/ul
15 U/ul
10 U/ul
5 U/ul

图21

| 描述 | 775pL GEMs CA-MMLV | 1.1nL GEMs CA-MMLV | 775pL GEMs 428 | 1.1nL GEMs 428 |
|---|---|---|---|---|
| 估计细胞数目 | 5119 | 4911 | 5325 | 5922 |
| 检测细胞与预期细胞的相对差异 | -0.15 | -0.18 | -0.11 | -0.01 |
| 每个细胞的平均组装读段 | 4804 | 6148 | 4251 | 3559 |
| 具有全长有效配对的细胞的经调整TRA TRB分数 | 0.28 | 0.22 | 0.42 | 0.46 |
| 定位于重组物组的读段的分数 | 0.86 | 0.86 | 0.71 | 0.74 |
| 有效条形码 | 0.91 | 0.89 | 0.81 | 0.84 |
| 低支持UMI读段分数 | 0.05 | 0.05 | 0.06 | 0.05 |
| RNA读段中的Q30碱基 | 0.97 | 0.97 | 0.98 | 0.98 |
| VDJ分数修正的条形码 | 0.06 | 0.06 | 0.05 | 0.05 |
| rRNA读段分数 | 0.010 | 0.006 | 0.096 | 0.082 |
| mtRNA读段分数 | 0.001 | 0.001 | 0.033 | 0.025 |
| 具有任何R1序列的读段的分数 | 0.013 | 0.024 | 0.006 | 0.008 |
| 具有>1个CDR3的CDR3+细胞的经调整TRA分数 | 0.09 | 0.07 | 0.14 | 0.13 |
| 具有>1个CDR3的CDR3+细胞的经调整TRB分数 | 0.12 | 0.09 | 0.17 | 0.16 |
| 中值重叠群长度 | 562 | 565 | 562 | 562 |
| 未注释重叠群的分数 | 0.04 | 0.03 | 0.04 | 0.05 |
| 处于有效、全长和配对克隆型的细胞的经调整TRA TRB分数(推断) | 0.35 | 0.29 | 0.50 | 0.54 |
| TRA TRB克隆型数目(推断) | 4183 | 3898 | 4357 | 4928 |

图22