



(12)发明专利申请

(10)申请公布号 CN 108026524 A

(43)申请公布日 2018.05.11

(21)申请号 201680052330.2

(74)专利代理机构 北京安信方达知识产权代理有限公司 11262

(22)申请日 2016.09.08

代理人 李平 郑霞

(30)优先权数据

62/217,220 2015.09.11 US

(51)Int.Cl.

C12N 15/10(2006.01)

(85)PCT国际申请进入国家阶段日

2018.03.09

C12Q 1/6869(2018.01)

(86)PCT国际申请的申请数据

PCT/US2016/050694 2016.09.08

(87)PCT国际申请的公布数据

W02017/044574 EN 2017.03.16

(71)申请人 赛卢拉研究公司

地址 美国加利福尼亚州

(72)发明人 克雷格·贝茨 格伦·弗

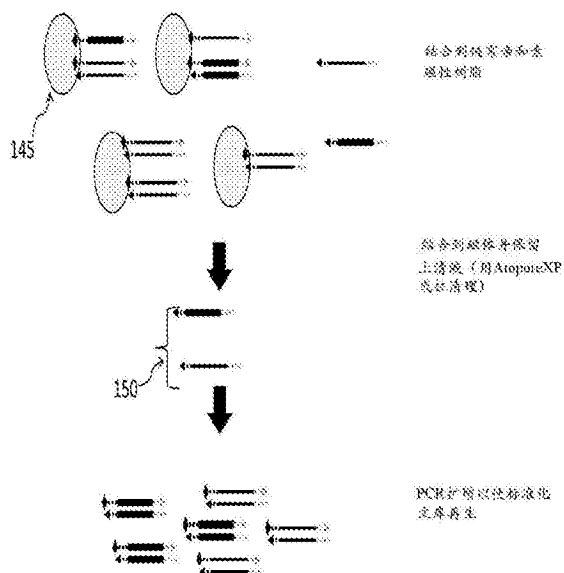
权利要求书7页 说明书35页 附图8页

(54)发明名称

用于核酸文库标准化的方法和组合物

(57)摘要

本披露提供了用于标准化核酸文库例如测序文库的方法、组合物和试剂盒。



1. 一种从多个核酸分子中除去高丰度种类的方法,所述方法包括:
使包括结合部分的多个第一寡核苷酸与第一多个核酸分子杂交,其中所述第一多个核酸分子包括至少一个高丰度种类;
延伸所述多个第一寡核苷酸以产生包括所述结合部分的所述第一多个核酸分子的多条互补链;
使包括所述第一多个核酸分子的所述多条互补链的多个双链核酸分子变性;
使所述第一多个核酸分子的所述多条互补链部分再退火;并且
通过固定在一个或多个固体支持物上的捕获分子除去所述第一多个核酸分子的所述再退火互补链以产生第二多个核酸分子,其中所述捕获分子与所述结合部分特异性结合,
凭此与所述第一多个核酸分子中的所述至少一个高丰度种类的含量相比,所述第二多个核酸分子中的所述至少一个高丰度种类的含量减少。
2. 如权利要求1所述的方法,其中所述结合部分是选自下组的官能团,该组由以下各项组成:生物素、链霉亲和素、肝素、适体、点击化学部分、地高辛、一个或多个伯胺、一个或多个羧基、一个或多个羟基、一个或多个醛、一个或多个酮和其任何组合。
3. 如权利要求2所述的方法,其中所述结合部分是生物素。
4. 如权利要求3所述的方法,其中所述捕获分子是链霉亲和素。
5. 如权利要求1-4中任一项所述的方法,所述方法还包括合成所述第一多个核酸分子的所述多条互补链中的至少一个的第二链以产生所述多个双链核酸分子中的一个或多个,所述多个双链核酸分子包括所述第一多个核酸分子的所述多条互补链。
6. 如权利要求5所述的方法,其中所述合成包括使多个第二寡核苷酸与所述第一多个核酸分子的所述多条互补链杂交并延伸所述多个第二寡核苷酸。
7. 如权利要求6所述的方法,其中所述多个第一寡核苷酸或所述多个第二寡核苷酸包括通用引物结合位点。
8. 如权利要求1-7中任一项所述的方法,所述方法还包括扩增所述多个双链核酸分子。
9. 如权利要求1-8中任一项所述的方法,其中所述第一多个核酸分子包括多个高丰度种类。
10. 如权利要求1-9中任一项所述的方法,其中所述至少一个高丰度种类代表所述第一多个核酸分子的至少50%。
11. 如权利要求1-9中任一项所述的方法,其中所述至少一个高丰度种类代表所述第一多个核酸分子的至少60%。
12. 如权利要求1-9中任一项所述的方法,其中所述至少一个高丰度种类代表所述第一多个核酸分子的至少70%。
13. 如权利要求1-12中任一项所述的方法,其中所述至少一个高丰度种类的含量减少是至少80%。
14. 如权利要求1-12中任一项所述的方法,其中所述至少一个高丰度种类的含量减少是至少90%。
15. 如权利要求1-12中任一项所述的方法,其中所述至少一个高丰度种类的含量减少是至少95%。
16. 如权利要求1-12中任一项所述的方法,其中所述至少一个高丰度种类的含量减少

少是至少99%。

17. 如权利要求9-16中任一项所述的方法,其中所述第二多个核酸分子包括所述多个高丰度种类。

18. 如权利要求17所述的方法,其中所述第二多个核酸分子中的所述多个高丰度种类代表所述第二多个核酸分子的少于50%。

19. 如权利要求17所述的方法,其中所述第二多个核酸分子中的所述多个高丰度种类代表所述第二多个核酸分子的少于40%。

20. 如权利要求17所述的方法,其中所述第二多个核酸分子中的所述多个高丰度种类代表所述第二多个核酸分子的少于30%。

21. 如权利要求1-20中任一项所述的方法,其中所述第一多个核酸分子包括多个低丰度种类。

22. 如权利要求21所述的方法,其中所述多个低丰度种类代表所述第一多个核酸分子的少于10%。

23. 如权利要求21所述的方法,其中所述多个低丰度种类代表所述第一多个核酸分子的少于5%。

24. 如权利要求21所述的方法,其中所述多个低丰度种类代表所述第一多个核酸分子的少于1%。

25. 如权利要求21-24中任一项所述的方法,其中所述第二多个核酸分子包括所述多个低丰度种类。

26. 如权利要求25所述的方法,其中所述第二多个核酸分子中的所述多个低丰度种类代表所述第二多个核酸分子的至少5%。

27. 如权利要求25所述的方法,其中所述第二多个核酸分子中的所述多个低丰度种类代表所述第二多个核酸分子的至少10%。

28. 如权利要求25所述的方法,其中所述第二多个核酸分子中的所述多个低丰度种类代表所述第二多个核酸分子的至少20%。

29. 如权利要求1-28中任一项所述的方法,其中所述第一多个核酸分子中的每个或所述第二多个核酸分子中的每个包括随机条形码。

30. 如权利要求1-29中任一项所述的方法,所述方法还包括对所述第二多个核酸分子进行测序以生成多个测序读数。

31. 如权利要求30所述的方法,其中所述多个高丰度种类的测序读数小于全部测序读数的50%。

32. 如权利要求30所述的方法,其中所述多个高丰度种类的测序读数小于全部测序读数的40%。

33. 如权利要求30所述的方法,其中所述多个高丰度种类的测序读数小于全部测序读数的30%。

34. 如权利要求30-33中任一项所述的方法,其中所述多个低丰度种类的测序读数为总测序读数的至少5%。

35. 如权利要求30-33中任一项所述的方法,其中所述多个低丰度种类的测序读数为总测序读数的至少10%。

36. 如权利要求30-33中任一项所述的方法,其中所述多个低丰度种类的测序读数为总测序读数的至少20%。

37. 如权利要求1-36中任一项所述的方法,所述方法还包括在所述部分再退火步骤中添加多种阻滞剂。

38. 如权利要求37所述的方法,其中所述多种阻滞剂与所述第一寡核苷酸的通用引物结合位点或所述第二寡核苷酸的通用引物结合位点杂交。

39. 如权利要求38所述的方法,其中所述多种阻滞剂阻止所述第一寡核苷酸的通用引物结合位点或所述第二寡核苷酸的通用引物结合位点与其互补序列之间的杂交。

40. 一种产生标准化核酸文库的方法,所述方法包括:

使包括结合部分的多个第一寡核苷酸与多个核酸靶标杂交;

延伸所述多个第一寡核苷酸以产生包括所述结合部分的所述多个核酸靶标的多条互补链;

使包括所述多个核酸靶标的所述多条互补链的多个双链核酸分子变性;

使所述多个核酸靶标的所述多条互补链部分再退火;并且

通过固定在一个或多个固体支持物上的捕获分子除去所述多个核酸靶标的所述再退火互补链,其中所述捕获分子与所述结合部分特异性结合,

凭此产生所述多个核酸靶标的标准化核酸文库。

41. 如权利要求40所述的方法,其中所述结合部分是选自下组的官能团,该组由以下各项组成:生物素、链霉亲和素、肝素、适体、点击化学部分、地高辛、一个或多个伯胺、一个或多个羧基、一个或多个羟基、一个或多个醛、一个或多个酮和其任何组合。

42. 如权利要求41所述的方法,其中所述结合部分是生物素。

43. 如权利要求42所述的方法,其中所述捕获分子是链霉亲和素。

44. 如权利要求40-43中任一项所述的方法,所述方法还包括合成所述多个核酸靶标的所述多条互补链中的一个或多个的第二链,以产生所述多个双链核酸分子中的一个或多个,所述多个双链核酸分子包括所述多个核酸靶标的所述多条互补链。

45. 如权利要求44所述的方法,其中所述合成包括使多个第二寡核苷酸与所述多个核酸靶标的所述多条互补链杂交并延伸所述多个第二寡核苷酸。

46. 如权利要求45所述的方法,其中所述多个第一寡核苷酸或所述多个第二寡核苷酸包括通用引物结合位点。

47. 如权利要求44-46中任一项所述的方法,所述方法还包括扩增所述多个双链核酸分子。

48. 如权利要求40-47中任一项所述的方法,其中所述多个核酸靶标包括多个低丰度核酸靶标。

49. 如权利要求48所述的方法,其中所述多个低丰度核酸靶标代表所述多个核酸靶标的少于10%。

50. 如权利要求48所述的方法,其中所述多个低丰度核酸靶标代表所述多个核酸靶标的少于5%。

51. 如权利要求48所述的方法,其中所述多个低丰度核酸靶标代表所述多个核酸靶标的少于1%。

52. 如权利要求48-51中任一项所述的方法,其中所述多个核酸靶标的所述标准化核酸文库包括所述多个低丰度核酸靶标。

53. 如权利要求52所述的方法,其中所述标准化核酸文库中的所述多个低丰度核酸靶标代表所述标准化核酸文库中的所述多个核酸靶标的至少5%。

54. 如权利要求52所述的方法,其中所述标准化核酸文库中的所述多个低丰度核酸靶标代表所述标准化核酸文库中的所述多个核酸靶标的至少10%。

55. 如权利要求52所述的方法,其中所述标准化核酸文库中的所述多个低丰度核酸靶标代表所述标准化核酸文库中的所述多个核酸靶标的至少20%。

56. 如权利要求40-55中任一项所述的方法,其中所述多个核酸靶标包括多个高丰度核酸靶标。

57. 如权利要求56所述的方法,其中所述多个高丰度核酸靶标代表所述多个核酸靶标的至少50%。

58. 如权利要求56所述的方法,其中所述多个高丰度核酸靶标代表所述多个核酸靶标的至少60%。

59. 如权利要求56所述的方法,其中所述多个高丰度核酸靶标代表所述多个核酸靶标的至少70%。

60. 如权利要求56-59中任一项所述的方法,其中所述标准化核酸文库中的所述多个高丰度种类的含量减少至少80%。

61. 如权利要求56-59中任一项所述的方法,其中所述标准化核酸文库中的所述多个高丰度种类的含量减少至少90%。

62. 如权利要求56-59中任一项所述的方法,其中所述标准化核酸文库中的所述多个高丰度种类的含量减少至少95%。

63. 如权利要求56-59中任一项所述的方法,其中所述标准化核酸文库中的所述多个高丰度种类的含量减少至少99%。

64. 如权利要求56-63中任一项所述的方法,其中所述多个核酸靶标的所述标准化核酸文库包括所述多个高丰度核酸靶标。

65. 如权利要求64所述的方法,其中所述标准化核酸文库中的所述多个高丰度核酸靶标代表所述标准化核酸文库中的所述多个核酸靶标的少于50%。

66. 如权利要求64所述的方法,其中所述标准化核酸文库中的所述多个高丰度核酸靶标代表所述标准化核酸文库中的所述多个核酸靶标的少于40%。

67. 如权利要求64所述的方法,其中所述标准化核酸文库中的所述多个高丰度核酸靶标代表所述标准化核酸文库中的所述多个核酸靶标的少于30%。

68. 如权利要求40-67中任一项所述的方法,其中所述多个第一寡核苷酸中的每个或所述多个第二寡核苷酸中的每个包括随机条形码。

69. 如权利要求40-68中任一项所述的方法,所述方法还包括对所述标准化核酸文库进行测序以生成多个测序读数。

70. 如权利要求69所述的方法,其中所述多个高丰度核酸靶标的测序读数小于全部测序读数的50%。

71. 如权利要求69所述的方法,其中所述多个高丰度核酸靶标的测序读数小于全部测

序读数的40%。

72. 如权利要求69所述的方法,其中所述多个高丰度核酸靶标的测序读数小于全部测序读数的30%。

73. 如权利要求69-72中任一项所述的方法,其中所述多个低丰度核酸靶标的测序读数为总测序读数的至少5%。

74. 如权利要求69-72中任一项所述的方法,其中所述多个低丰度核酸靶标的测序读数为总测序读数的至少10%。

75. 如权利要求69-72中任一项所述的方法,其中所述多个低丰度核酸靶标的测序读数为总测序读数的至少20%。

76. 如权利要求40-75中任一项所述的方法,所述方法还包括在所述部分再退火步骤中添加多种阻滞剂。

77. 如权利要求76所述的方法,其中所述多种阻滞剂与所述第一寡核苷酸的通用引物结合位点或所述第二寡核苷酸的通用引物结合位点杂交。

78. 如权利要求77所述的方法,其中所述多种阻滞剂阻止所述第一寡核苷酸的通用引物结合位点或所述第二寡核苷酸的通用引物结合位点与其互补序列之间的杂交。

79. 如权利要求40-78中任一项所述的方法,其中所述多个核酸靶标包括mRNA。

80. 如权利要求79所述的方法,其中所述多个核酸靶标包括线粒体mRNA。

81. 如权利要求79所述的方法,其中所述多个核酸靶标包括核糖体蛋白mRNA。

82. 如权利要求79-81中任一项所述的方法,其中所述低丰度核酸靶标包括7,000个具有最低数目的转录物的基因。

83. 如权利要求79-81中任一项所述的方法,其中所述低丰度核酸靶标包括4,000个具有最低数目的转录物的基因。

84. 如权利要求79-81中任一项所述的方法,其中所述低丰度核酸靶标包括2,000个具有最低数目的转录物的基因。

85. 如权利要求40-84中任一项所述的方法,其中所述多个第一寡核苷酸包括靶标特异性引物。

86. 如权利要求40-84中任一项所述的方法,其中所述多个第一寡核苷酸包括非靶标特异性引物。

87. 如权利要求40-78中任一项所述的方法,其中所述多个核酸靶标包括cDNA。

88. 如权利要求40-78中任一项所述的方法,其中所述多个核酸靶标包括基因组DNA。

89. 如权利要求88所述的方法,其中所述高丰度核酸靶标包括短串联重复序列。

90. 如权利要求88所述的方法,其中所述高丰度核酸靶标包括端粒序列。

91. 如权利要求88所述的方法,其中所述高丰度核酸靶标包括着丝粒序列。

92. 如权利要求40-91中任一项所述的方法,其中所述多个核酸靶标来自单细胞。

93. 一种产生标准化核酸文库的方法,所述方法包括:

使包括结合部分的多个第一寡核苷酸与未标准化核酸文库中的多个核酸靶标杂交;

延伸所述多个第一寡核苷酸以产生包括所述结合部分的所述多个核酸靶标的多条互补链;

使包括所述多个核酸靶标的所述多条互补链的多个双链核酸分子变性;

使所述多个核酸靶标的所述多条互补链部分再退火;并且
除去所述多个核酸靶标的所述再退火互补链,
凭此产生所述多个核酸靶标的标准化核酸文库。

94. 如权利要求93所述的方法,其中所述未标准化核酸文库包括一个或多个高丰度核酸靶标和一个或多个低丰度核酸靶标。

95. 如权利要求94所述的方法,其中所述一个或多个高丰度核酸靶标代表所述未标准化核酸文库的至少50%。

96. 如权利要求94所述的方法,其中所述一个或多个高丰度核酸靶标代表所述未标准化核酸文库的至少60%。

97. 如权利要求94所述的方法,其中所述一个或多个高丰度核酸靶标代表所述未标准化核酸文库的至少70%。

98. 如权利要求94-97中任一项所述的方法,其中所述标准化核酸文库中的所述一个或多个高丰度核酸靶标的含量减少至少80%。

99. 如权利要求94-97中任一项所述的方法,其中所述标准化核酸文库中的所述一个或多个高丰度核酸靶标的含量减少至少90%。

100. 如权利要求94-97中任一项所述的方法,其中所述标准化核酸文库中的所述一个或多个高丰度核酸靶标的含量减少至少95%。

101. 如权利要求94-97中任一项所述的方法,其中所述标准化核酸文库中的所述一个或多个高丰度核酸靶标的含量减少至少99%。

102. 如权利要求94-101中任一项所述的方法,其中所述一个或多个低丰度核酸靶标代表所述未标准化核酸文库的小于10%。

103. 如权利要求94-101中任一项所述的方法,其中所述一个或多个低丰度核酸靶标代表所述未标准化核酸文库的小于5%。

104. 如权利要求94-101中任一项所述的方法,其中所述一个或多个低丰度核酸靶标代表所述未标准化核酸文库的小于1%。

105. 如权利要求94-104中任一项所述的方法,其中所述一个或多个低丰度核酸靶标代表所述标准化核酸文库的至少5%。

106. 如权利要求94-104中任一项所述的方法,其中所述一个或多个低丰度核酸靶标代表所述标准化核酸文库的至少10%。

107. 如权利要求94-104中任一项所述的方法,其中所述一个或多个低丰度核酸靶标代表所述标准化核酸文库的至少20%。

108. 如权利要求94-107中任一项所述的方法,其中所述一个或多个高丰度核酸靶标代表所述标准化核酸文库的少于50%。

109. 如权利要求94-107中任一项所述的方法,其中所述一个或多个高丰度核酸靶标代表所述标准化核酸文库的少于40%。

110. 如权利要求94-107中任一项所述的方法,其中所述一个或多个高丰度核酸靶标代表所述标准化核酸文库的少于30%。

111. 如权利要求93-110中任一项所述的方法,其中所述未标准化核酸文库是cDNA文库。

112. 如权利要求93-110中任一项所述的方法,其中所述未标准化核酸文库是基因组文库。

113. 如权利要求93-112中任一项所述的方法,其中所述未标准化核酸文库是单细胞核酸文库。

用于核酸文库标准化的方法和组合物

相关申请的交叉引用

[0001] 本申请是基于2015年9月11日提交的美国临时专利申请序列号62/217,220并且根据35U.S.C. §119(e) 要求其权益。出于所有目的,将所述相关申请的内容以其全文通过引用并入本文。

背景技术

[0002] 在许多应用(特别是测序和文库筛选)中,研究人员想要评估样品或文库中的整个核酸目录。然而,样品(例如cDNA文库)中的核酸常处于范围几个数量级的宽范围浓度内。这大大增加了全面评估样品所需的测序或筛选量。为了克服这一点,研究人员常想用标准化文库来工作,其中不同核酸的浓度都处于相似的浓度,这样以来需要更少的多余努力来探询新的分子。

发明内容

[0003] 本文披露的一些实施例提供了从多个核酸分子中除去高丰度种类的方法,所述方法包括:使包括结合部分的多个第一寡核苷酸与第一多个核酸分子杂交,其中所述第一多个核酸分子包括至少一个高丰度种类;延伸所述多个第一寡核苷酸以产生包括所述结合部分的所述第一多个核酸分子的多条互补链;使包括所述第一多个核酸分子的所述多条互补链的多个双链核酸分子变性;使所述第一多个核酸分子的所述多条互补链部分再退火;并且通过固定在一个或多个固体支持物上的捕获分子除去所述第一多个核酸分子的所述再退火互补链以产生第二多个核酸分子,其中所述捕获分子与所述结合部分特异性结合,凭此与所述第一多个核酸分子中的所述至少一个高丰度种类的含量相比,所述第二多个核酸分子中的所述至少一个高丰度种类的含量减少。

[0004] 在一些实施例中,所述结合部分是选自下组的官能团,该组由以下各项组成:生物素、链霉亲和素、肝素、适体、点击化学部分、地高辛、一个或多个伯胺、一个或多个羧基、一个或多个羟基、一个或多个醛、一个或多个酮和其任何组合。在一些实施例中,所述结合部分是生物素。在一些实施例中,所述捕获分子是链霉亲和素。在一些实施例中,所述方法还包括合成所述第一多个核酸分子的所述多条互补链中的至少一个的第二链以产生所述多个双链核酸分子中的一个或多个,所述多个双链核酸分子包括所述第一多个核酸分子的所述多条互补链。在一些实施例中,所述合成包括使多个第二寡核苷酸与所述第一多个核酸分子的所述多条互补链杂交并延伸所述多个第二寡核苷酸。在一些实施例中,所述多个第一寡核苷酸或所述多个第二寡核苷酸包括通用引物结合位点。在一些实施例中,所述方法还包括扩增所述多个双链核酸分子。在一些实施例中,所述第一多个核酸分子包括多个高丰度种类。在一些实施例中,所述至少一个高丰度种类代表所述第一多个核酸分子的至少50%。在一些实施例中,所述至少一个高丰度种类代表所述第一多个核酸分子的至少60%。在一些实施例中,所述至少一个高丰度种类代表所述第一多个核酸分子的至少70%。在一些实施例中,所述至少一个高丰度种类的含量减少是至少80%。在一些实施例中,所述至

少一个高丰度种类的含量的减少是至少90%。在一些实施例中,所述至少一个高丰度种类的含量的减少是至少95%。在一些实施例中,所述至少一个高丰度种类的含量的减少是至少99%。在一些实施例中,所述第二多个核酸分子包括所述多个高丰度种类。在一些实施例中,所述第二多个核酸分子中的所述多个高丰度种类代表所述第二多个核酸分子的少于50%。在一些实施例中,所述第二多个核酸分子中的所述多个高丰度种类代表所述第二多个核酸分子的少于40%。在一些实施例中,所述第二多个核酸分子中的所述多个高丰度种类代表所述第二多个核酸分子的少于30%。在一些实施例中,所述第一多个核酸分子包括多个低丰度种类。在一些实施例中,所述多个低丰度种类代表所述第一多个核酸分子的少于10%。在一些实施例中,所述多个低丰度种类代表所述第一多个核酸分子的少于5%。在一些实施例中,所述多个低丰度种类代表所述第一多个核酸分子的少于1%。在一些实施例中,所述第二多个核酸分子包括所述多个低丰度种类。在一些实施例中,所述第二多个核酸分子中的所述多个低丰度种类代表所述第二多个核酸分子的至少5%。在一些实施例中,所述第二多个核酸分子中的所述多个低丰度种类代表所述第二多个核酸分子的至少10%。在一些实施例中,所述第二多个核酸分子中的所述多个低丰度种类代表所述第二多个核酸分子的至少20%。在一些实施例中,所述第一多个核酸分子中的每个或所述第二多个核酸分子中的每个包括随机条形码。在一些实施例中,所述方法还包括对所述第二多个核酸分子进行测序以生成多个测序读数。在一些实施例中,所述多个高丰度种类的测序读数小于全部测序读数的50%。在一些实施例中,所述多个高丰度种类的测序读数小于全部测序读数的40%。在一些实施例中,所述多个高丰度种类的测序读数小于全部测序读数的30%。在一些实施例中,所述多个低丰度种类的测序读数为总测序读数的至少5%。在一些实施例中,所述多个低丰度种类的测序读数为总测序读数的至少10%。在一些实施例中,所述多个低丰度种类的测序读数为总测序读数的至少20%。在一些实施例中,所述方法还包括在部分再退火步骤中添加多种阻滞剂。在一些实施例中,所述多种阻滞剂与所述第一寡核苷酸的通用引物结合位点或所述第二寡核苷酸的通用引物结合位点杂交。在一些实施例中,所述多种阻滞剂阻止所述第一寡核苷酸的通用引物结合位点或所述第二寡核苷酸的通用引物结合位点与其互补序列之间的杂交。

[0005] 本文披露的一些实施例提供了产生标准化核酸文库的方法,所述方法包括:使包括结合部分的多个第一寡核苷酸与多个核酸靶标杂交;延伸所述多个第一寡核苷酸以产生包括所述结合部分的所述多个核酸靶标的多条互补链;使包括所述多个核酸靶标的所述多条互补链的多个双链核酸分子变性;使所述多个核酸靶标的所述多条互补链部分再退火;并且通过固定在一个或多个固体支持物上的捕获分子除去所述多个核酸靶标的所述再退火互补链,其中所述捕获分子与所述结合部分特异性结合,凭此产生所述多个核酸靶标的标准化核酸文库。

[0006] 在一些实施例中,所述结合部分是选自下组的官能团,该组由以下各项组成:生物素、链霉亲和素、肝素、适体、点击化学部分、地高辛、一个或多个伯胺、一个或多个羧基、一个或多个羟基、一个或多个醛、一个或多个酮和其任何组合。在一些实施例中,所述结合部分是生物素。在一些实施例中,所述捕获分子是链霉亲和素。在一些实施例中,所述方法还包括合成所述多个核酸靶标的所述多条互补链中的一个或多个的第二链,以产生所述多个双链核酸分子中的一个或多个,所述多个双链核酸分子包括所述多个核酸靶标的所述多条

互补链。在一些实施例中,所述合成包括使所述多个第二寡核苷酸与所述多个核酸靶标的所述多条互补链杂交并延伸所述多个第二寡核苷酸。在一些实施例中,所述多个第一寡核苷酸或所述多个第二寡核苷酸包括通用引物结合位点。在一些实施例中,所述方法还包括扩增所述多个双链核酸分子。在一些实施例中,所述多个核酸靶标包括多个低丰度核酸靶标。在一些实施例中,所述多个低丰度核酸靶标代表所述多个核酸靶标的少于10%。在一些实施例中,所述多个低丰度核酸靶标代表所述多个核酸靶标的少于5%。在一些实施例中,所述多个低丰度核酸靶标代表所述多个核酸靶标的少于1%。在一些实施例中,所述多个核酸靶标的所述标准化核酸文库包括所述多个低丰度核酸靶标。在一些实施例中,所述标准化核酸文库中的所述多个低丰度核酸靶标代表所述标准化核酸文库中的所述多个核酸靶标的至少5%。在一些实施例中,所述标准化核酸文库中的所述多个低丰度核酸靶标代表所述标准化核酸文库中的所述多个核酸靶标的至少10%。在一些实施例中,所述标准化核酸文库中的所述多个低丰度核酸靶标代表所述标准化核酸文库中的所述多个核酸靶标的至少20%。在一些实施例中,所述多个核酸靶标包括多个高丰度核酸靶标。在一些实施例中,所述多个高丰度核酸靶标代表所述多个核酸靶标的至少50%。在一些实施例中,所述多个高丰度核酸靶标代表所述多个核酸靶标的至少60%。在一些实施例中,所述多个高丰度核酸靶标代表所述多个核酸靶标的至少70%。在一些实施例中,所述标准化核酸文库中的所述多个高丰度种类的含量减少至少80%。在一些实施例中,所述标准化核酸文库中的所述多个高丰度种类的含量减少至少90%。在一些实施例中,所述标准化核酸文库中的所述多个高丰度种类的含量减少至少95%。在一些实施例中,所述标准化核酸文库中的所述多个高丰度种类的含量减少至少99%。在一些实施例中,所述多个核酸靶标的所述标准化核酸文库包括所述多个高丰度核酸靶标。在一些实施例中,所述标准化核酸文库中的所述多个高丰度核酸靶标代表所述标准化核酸文库中的所述多个核酸靶标的少于50%。在一些实施例中,所述标准化核酸文库中的所述多个高丰度核酸靶标代表所述标准化核酸文库中的所述多个核酸靶标的少于40%。在一些实施例中,所述标准化核酸文库中的所述多个高丰度核酸靶标代表所述标准化核酸文库中的所述多个核酸靶标的少于30%。在一些实施例中,所述多个第一寡核苷酸中的每个或所述多个第二寡核苷酸中的每个包括随机条形码。在一些实施例中,所述方法还包括对所述标准化核酸文库进行测序以生成多个测序读数。在一些实施例中,所述多个高丰度核酸靶标的测序读数小于全部测序读数的50%。在一些实施例中,所述多个高丰度核酸靶标的测序读数小于全部测序读数的40%。在一些实施例中,所述多个高丰度核酸靶标的测序读数小于全部测序读数的30%。在一些实施例中,所述多个低丰度核酸靶标的测序读数为总测序读数的至少5%。在一些实施例中,所述多个低丰度核酸靶标的测序读数为总测序读数的至少10%。在一些实施例中,所述多个低丰度核酸靶标的测序读数为总测序读数的至少20%。在一些实施例中,所述方法还包括在部分再退火步骤中添加多种阻滞剂。在一些实施例中,所述多种阻滞剂与所述第一寡核苷酸的通用引物结合位点或所述第二寡核苷酸的通用引物结合位点杂交。在一些实施例中,所述多种阻滞剂阻止所述第一寡核苷酸的通用引物结合位点或所述第二寡核苷酸的通用引物结合位点与其互补序列之间的杂交。在一些实施例中,所述多个核酸靶标包括mRNA。在一些实施例中,所述多个核酸靶标包括线粒体mRNA。在一些实施例中,所述多个核酸靶标包括核糖体蛋白mRNA。在一些实施例中,低丰度核酸靶标包括7,000个具有最低数目的转录物的基因。所

述在一些实施例中,低丰度核酸靶标包括4,000个具有最低数目的转录物的基因。所述在一些实施例中,低丰度核酸靶标包括2,000个具有最低数目的转录物的基因。所述在一些实施例中,所述多个第一寡核苷酸包括靶标特异性引物。在一些实施例中,所述多个第一寡核苷酸包括非靶标特异性引物。在一些实施例中,所述多个核酸靶标包括cDNA。在一些实施例中,所述多个核酸靶标包括基因组DNA。在一些实施例中,所述高丰度核酸靶标包括短串联重复序列。在一些实施例中,所述高丰度核酸靶标包括端粒序列。在一些实施例中,所述高丰度核酸靶标包括着丝粒序列。在一些实施例中,所述多个核酸靶标来自单细胞。

[0007] 本文披露的一些实施例提供了产生标准化核酸文库的方法,所述方法包括:使包括结合部分的多个第一寡核苷酸与未标准化核酸文库中的多个核酸靶标杂交;延伸所述多个第一寡核苷酸以产生包括所述结合部分的所述多个核酸靶标的多条互补链;使包括所述多个核酸靶标的所述多条互补链的多个双链核酸分子变性;使所述多个核酸靶标的所述多条互补链部分再退火;并且除去所述多个核酸靶标的所述再退火互补链,凭此产生所述多个核酸靶标的标准化核酸文库。

[0008] 在一些实施例中,所述未标准化核酸文库包括一个或多个高丰度核酸靶标和一个或多个低丰度核酸靶标。在一些实施例中,所述一个或多个高丰度核酸靶标代表所述未标准化核酸文库的至少50%。在一些实施例中,所述一个或多个高丰度核酸靶标代表所述未标准化核酸文库的至少60%。在一些实施例中,所述一个或多个高丰度核酸靶标代表所述未标准化核酸文库的至少70%。在一些实施例中,所述标准化核酸文库中的所述一个或多个高丰度核酸靶标的含量减少至少80%。在一些实施例中,所述标准化核酸文库中的所述一个或多个高丰度核酸靶标的含量减少至少90%。在一些实施例中,所述标准化核酸文库中的所述一个或多个高丰度核酸靶标的含量减少至少95%。在一些实施例中,所述标准化核酸文库中的所述一个或多个高丰度核酸靶标的含量减少至少99%。在一些实施例中,所述一个或多个低丰度核酸靶标代表所述未标准化核酸文库的小于10%。在一些实施例中,所述一个或多个低丰度核酸靶标代表所述未标准化核酸文库的小于5%。在一些实施例中,所述一个或多个低丰度核酸靶标代表所述未标准化核酸文库的小于1%。在一些实施例中,所述一个或多个低丰度核酸靶标代表所述标准化核酸文库的至少5%。在一些实施例中,所述一个或多个低丰度核酸靶标代表所述标准化核酸文库的至少10%。在一些实施例中,所述一个或多个低丰度核酸靶标代表所述标准化核酸文库的至少20%。在一些实施例中,所述一个或多个高丰度核酸靶标代表所述标准化核酸文库的少于50%。在一些实施例中,所述一个或多个高丰度核酸靶标代表所述标准化核酸文库的少于40%。在一些实施例中,所述一个或多个高丰度核酸靶标代表所述标准化核酸文库的少于30%。在一些实施例中,所述未标准化核酸文库是cDNA文库。在一些实施例中,所述未标准化核酸文库是基因组文库。在一些实施例中,所述未标准化核酸文库是单细胞核酸文库。

[0009] 在一个方面,本披露提供了核酸文库标准化的方法,所述方法包括:产生不对称标记的双链cDNA文库,其中所述cDNA文库的双链cDNA的一条链包括结合部分;使所述文库中所述双链cDNA的链变性和部分再退火,从而产生包括所述结合部分的再退火cDNA分子、包括所述结合部分的单链分子和缺乏所述结合部分的单链分子的混合物;并且除去包括所述结合部分的分子,同时留下缺乏所述结合部分的单链分子,从而产生标准化文库。在一些实施例中,所述产生包括使用包括所述结合部分的引物将mRNA逆转录成第一cDNA链。在一些

实施例中,所述引物包括随机条形码。在一些实施例中,所述方法还包括产生与所述第一cDNA链互补的第二cDNA链,从而产生所述双链cDNA。在一些实施例中,所述产生包括使用包括所述结合部分的引物对DNA分子进行引物延伸,从而产生所述双链cDNA。在一些实施例中,所述结合部分选自下组,该组由以下各项组成:生物素和链霉亲和素。在一些实施例中,所述结合部分附接至所述双链cDNA的正义链。在一些实施例中,所述结合部分附接至所述双链cDNA的反义链。在一些实施例中,所述变性包括加热所述双链cDNA。在一些实施例中,所述变性包括使所述双链cDNA的至少50%变性。在一些实施例中,所述变性包括使所述双链cDNA的至少90%变性。在一些实施例中,所述再退火的cDNA来自高丰度核酸种类。在一些实施例中,将所述再退火的cDNA的链以至少两倍于来自丰度较小核酸种类的双链cDNA文库的链的速度进行再退火。在一些实施例中,将所述再退火的cDNA的链以至少五倍于来自丰度较小核酸种类的双链cDNA文库的链的速度进行再退火。在一些实施例中,将所述再退火的cDNA的链以比来自丰度较小核酸种类的双链cDNA文库的链更高的丰度进行再退火。在一些实施例中,所述除去包括使所述分子与固体支持物接触。在一些实施例中,所述固体支持物包括与所述结合部分结合的捕获部分。在一些实施例中,所述捕获部分选自下组,该组由以下各项组成:生物素和链霉亲和素。在一些实施例中,所述固体支持物是珠。在一些实施例中,所述固体支持物是磁性的。在一些实施例中,所述再退火的cDNA是双链的。在一些实施例中,缺乏所述结合部分的所述单链分子是在除去期间除去的单链的互补体。在一些实施例中,所述方法还包括扩增所述标准化文库。在一些实施例中,所述扩增包括在所述标准化文库中从单链核酸产生双链核酸。在一些实施例中,所述方法还包括向所述双链核酸添加衔接子。在一些实施例中,添加衔接子包括将所述衔接子连接至所述双链核酸。在一些实施例中,添加衔接子包括通过扩增将所述衔接子引至所述双链核酸。在一些实施例中,所述衔接子包括测序流动池序列。在一些实施例中,所述方法还包括对所述双链核酸进行测序。在一些实施例中,所述标准化文库中的丰度最高的种类比所述标准化文库中的丰度最低的种类的丰度多至多20倍。在一些实施例中,所述标准化文库中的丰度最高的种类比所述标准化文库中的丰度最低的种类的丰度多至多10倍。在一些实施例中,所述标准化文库中的丰度最高的种类比所述标准化文库中的丰度最低的种类的丰度多至多5倍。

附图说明

[0010] 本发明的新颖特征在所附权利要求书中具体阐述。通过参考对说明性实施例进行阐述的以下详细说明,将获得对本发明的特征和优点的更好理解,在所述实施例中利用了本发明的原理,并且在所述附图中:

[0011] 图1展示了本披露的文库标准化方法的示例性实施例。

[0012] 图2展示了本披露的随机条形码化方法的示例性实施例。

[0013] 图3展示了本披露的用于以结合部分不对称地标记扩增子的扩增方法的示例性实施例。

[0014] 图4展示了阻滞剂在本披露的方法中的使用的示例性实施例。

[0015] 图5A和5B描绘了使用通用衔接子引物将随机条形码添加到靶标的方法的示例性实施例。

具体实施方式

[0016] 出于所有目的,本说明书中提到的所有出版物、专利和专利申请都通过引用并入本文,如同每一单独的出版物、专利或专利申请具体且单独地指明通过引用并入。

[0017] 大多数文库标准化策略是基于两个原则之一。第一个是将文库与另一组核酸杂交,其中序列被均匀地代表,如来自源生物体的基因组DNA,并保留杂交的部分。另一个方法依赖于溶液杂交的浓度依赖性。当一组dsDNA分子变性时,它们将以与其原始浓度的平方成比例的速率重新杂交。在一些实施例中,本文披露的方法、组合物和试剂盒利用此特性进行文库标准化,其通过使混合物变性并且仅允许其部分再退火;成比例地,多得多的低浓度种类将与dsDNA重新杂交,而丰度较小的种类仍将主要为单链的。

[0018] 在一些实施例中,本文披露的方法、组合物和试剂盒可避免在文库标准化期间使用ssDNA和dsDNA部分的物理和酶促分离。在最初的文库制备过程中,文库在一端用结合部分进行不对称标记(例如通过用一个5'生物素化引物和另一个未标记引物进行PCR)。变性和部分再退火后,将所有标记的链捕获在支持基质如顺磁性链霉亲和素珠上,并分离结合和未结合的部分。高丰度序列将主要重新杂交,并且在结合的部分中两条链都将被除去。然而,低丰度序列将不太可能重新杂交,因此标记的链的互补体将存在于未结合的部分中。未结合的部分将代表标准化文库,并且可以直接使用或进一步扩增以用于下游应用。

定义

[0019] 除非另有定义,否则本文使用的所有技术术语具有与本披露所属领域的本领域普通技术人员通常理解的相同的含义。如本说明书和所附权利要求书中使用的,除非上下文另有明确指示,否则单数形式“一个/一种(a/an)”和“所述(the)”包括复数的提及物。除非另有说明,否则本文中“或”的任何提及旨在包含“和/或”。

[0020] 如本文使用的,术语“关联”或“与……相关联”可意指两个或更多个种类可以被识别为在某个时间点处同位。关联可意指两个或更多个种类在或曾经在相似的容器内。关联可以是信息学关联,其中例如关于两个或更多个种类的数字信息被存储并且可以用于确定所述种类中的一个或多个在某个时间点处同位。关联也可以是物理关联。在一些情况下,两个或更多个相关联的种类彼此之间或与共同的固体或半固体表面是“连接的”、“附接的”或“固定的”。关联可以指用于将标记附接到固体或半固体支持物(如珠)上的共价或非共价方式。关联可以包括靶标和标记之间的杂交。

[0021] 如本文使用的,术语“互补性”可以指两个核苷酸之间精确配对的能力。例如,如果核酸的在给定位置的核苷酸能够与另一个核酸的核苷酸以氢键结合,则两个核酸被认为在所述位置处是彼此互补的。两单链核酸分子之间的互补性可以是“部分的”,其中所述核苷酸中仅一些结合,或者当所述单链分子之间存在完全互补性时,这种互补性可以是完全的。如果第一核苷酸序列与第二核苷酸序列互补,则可以认为第一核苷酸序列是第二序列的“互补体”。如果第一核苷酸序列互补于和第二序列相反的序列(即,核苷酸顺序相反),则可以认为第一核苷酸序列是第二序列的“反向互补体”。如本文使用的,术语“互补体”、“互补”和“反向互补体”可以互换使用。从本披露可以理解,如果一个分子可以与另一个分子杂交,则其可以是杂交的分子的互补体。

[0022] 如本文使用的,术语“数字计数”可以指用于估计样品中靶分子数目的方法。数字

计数可以包括确定已经与样品中的靶标相关联的独特标记的数目的步骤。这种随机方法将计数分子的问题从相同分子的定位和鉴定之一转化为有关检测到一组预定义标记的一系列是/否数字问题。

[0023] 如本文使用的,术语(多个)“标记”可以指与样品中的靶标相关联的核酸代码。标记可以是例如核酸标记。标记可以是完全或部分可扩增的标记。标记可以是完全或部分可测序的标记。标记可以是可鉴定为有区别的天然核酸的一部分。标记可以是已知的序列。标记可以包括核酸序列的接点,例如天然和非天然序列的接点。如本文使用的,术语“标记”可以与术语“索引”、“标签”或“标记-标签”互换使用。标记可以传达信息。例如,在各种实施例中,可以使用标记来确定样品的身份、样品的来源、细胞的身份和/或靶标。

[0024] 如本文使用的,术语“非耗尽性储库(non-depleting reservoir)”可以指由许多不同标记组成的随机条形码池。非耗尽性储库可以包括大量不同的随机条形码,使得当非耗尽性储库与靶标池相关联时,每个靶标可能与独特的随机条形码相关联。每个标记的靶分子的独特性可以通过随机选择的统计来确定,并且取决于与多样的标记相比在集合中相同的靶分子的拷贝数。所得的标记的靶分子集合的大小可以通过条形码化处理的随机性质来确定,然后对检测到的随机条形码的数目的分析允许计算原始集合或样品中存在的靶分子的数目。当存在的靶分子的拷贝数与独特的随机条形码的数目的比率低时,标记的靶分子是高度独特的(即,用给定的标记来标记多于一个靶分子的概率非常低)。

[0025] 如本文使用的,“核酸”通常可指多核苷酸序列或其片段。核酸可以包括核苷酸。核酸对于细胞可以是外源的或内源的。核酸可以存在于无细胞环境中。核酸可以是基因或其片段。核酸可以是DNA。核酸可以是RNA。核酸可以包括一种或多种类似物(例如改变的骨架、糖或核碱基)。类似物的一些非限制性实例包括:5-溴尿嘧啶,肽核酸、异种核酸、吗啉代、锁核酸、二醇核酸、苏糖核酸、双脱氧核苷酸、虫草素、7-脱氮杂-GTP、荧光团(例如与糖连接的罗丹明或荧光素)、含硫醇的核苷酸、生物素连接的核苷酸、荧光碱类似物、CpG岛、甲基-7-鸟苷、甲基化核苷酸、肌苷、硫尿苷、假尿苷、二氢尿苷、癸苷和怀俄苷。“核酸”、“多核苷酸”、“靶多核苷酸”和“靶核酸”可以互换使用。

[0026] 核酸可以包括一种或多种修饰(例如,碱基修饰、骨架修饰),以为核酸提供新的或增强的特征(例如,改进的稳定性)。核酸可以包括核酸亲和标签。核苷可以是碱基-糖组合。核苷的碱基部分可以是杂环碱基。此类杂环碱基的两个最常见的类别是嘌呤和嘧啶。核苷酸可以是还包含与核苷的糖部分共价连接的磷酸基团的核苷。对于包含呋喃戊糖的那些核苷,磷酸基团可以连接到糖的2'、3'或5'羟基部分。在形成核酸中,磷酸基团可以将相邻的核苷彼此共价连接以形成线性高分子化合物。转而此线性高分子化合物的各自末端可以进一步接合而形成环状化合物;然而,线性化合物通常是合适的。此外,线性化合物可以具有内部核苷酸碱基互补性,并且因此可以按产生完全或部分双链化合物的方式折叠。在核酸中,所述磷酸基团通常可以被称为形成核酸的核苷间骨架。核酸的键或骨架可以是3'至5'磷酸二酯键。

[0027] 核酸可以包括修饰的骨架和/或修饰的核苷间键。修饰的骨架可以包括在骨架中保留磷原子和在骨架中不具有磷原子的那些。其中含有磷原子的合适修饰的核酸骨架可以包含例如硫代磷酸酯;手性硫代磷酸酯;二硫代磷酸酯;磷酸三酯;氨基烷基磷酸三酯;甲基磷酸酯和其他烷基磷酸酯,如3'-亚烷基磷酸酯、5'-亚烷基磷酸酯;手性磷酸酯;亚磷酸酯;

包括3'-氨基磷酰胺酯和氨基烷基磷酰胺酯的磷酰胺酯；磷二酰胺酯；硫代羰基磷酰胺酯；硫代羰基烷基磷酸酯；硫代羰基烷基磷酸三酯；硒代磷酸酯；以及具有正常3'-5'键的硼烷磷酸酯，2'-5'连接的类似物和具有反向极性的那些，其中一个或多个核苷酸间键是3'至3'、5'至5'或2'至2'键。

[0028] 核酸可以包括由短链烷基或环烷基核苷间键、混合杂原子、和烷基或环烷基核苷间键或者一个或多个短链杂原子的或杂环的核苷间键形成的多核苷酸骨架。这些可包括具有以下结构的那些：吗啉代键（从核苷的糖部分部分地形成）；硅氧烷骨架；硫化物、亚砷和砷骨架；甲酰乙酰基和硫代甲酰乙酰基骨架；亚甲基甲酰乙酰基和硫代甲酰乙酰基骨架；核糖乙酰基骨架；含烯的骨架；氨基磺酸盐骨架；亚甲亚氨基和亚甲胍基骨架；磺酸酯和磺酰胺骨架；酰胺骨架；和具有混合的N、O、S和CH₂组分部分的其他骨架。

[0029] 核酸可以包括核酸模拟物。术语“模拟物”可以旨在包括其中只有呋喃糖环或呋喃糖环和核苷酸间键两者被非呋喃糖基团替代的多核苷酸，仅替代呋喃糖环也可以称为糖替代物。可以保持杂环碱基部分或修饰的杂环碱基部分以便与适当的靶核酸杂交。一种这样的核酸可以是肽核酸（PNA）。在PNA中，多核苷酸的糖骨架可以被含酰胺的骨架（特别是氨基乙基甘氨酸骨架）替代。核苷酸可以被保持并且直接或间接地结合至骨架的酰胺部分的氮杂氮原子上。PNA化合物中的骨架可以包括两个或更多个连接的氨基乙基甘氨酸单元，其给予PNA含酰胺的骨架。杂环碱基部分可以直接或间接地结合到骨架的酰胺部分的氮杂氮原子上。

[0030] 核酸可以包括吗啉代骨架结构。例如，核酸可以包括代替核糖环的6元吗啉代环。在这些实施例的一些中，磷二酰胺酯或其他非磷酸二酯核苷间键可替代磷酸二酯键。

[0031] 核酸可以包括具有附接到吗啉代环上的杂环碱基的连接的吗啉代单元（即吗啉代核酸）。连接基团可以连接吗啉代核酸中的吗啉代单体单元。非离子型基于吗啉代的寡聚化合物可以与细胞蛋白具有较少的不希望的作用。基于吗啉代的多核苷酸可以是核酸的非离子模拟物。吗啉代类别中的多种化合物可以使用不同的连接基团连接。另一类多核苷酸模拟物可称为环己烯基核酸（CeNA）。通常存在于核酸分子中的呋喃糖环可以被环己烯基环替代。可以制备CeNA DMT保护的亚磷酰胺单体，并用于使用亚磷酰胺化学的寡聚化合物合成。将CeNA单体并入核酸链可以增加DNA/RNA杂交体的稳定性。CeNA寡聚腺苷酸可以与具有和天然复合物相似的稳定性的核酸互补体形成复合物。另外的修饰可以包括锁核酸（LNA），其中2'-羟基基团连接到糖环的4'碳原子，从而形成2'-C,4'-C-氧亚甲基键，由此形成双环糖部分。所述键可以是桥连2'氧原子和4'碳原子的基团亚甲基（—CH₂—），其中n是1或2。LNA和LNA类似物可以显示与互补核酸非常高的双重热稳定性（T_m=+3°C到+10°C）、对于3'-外切核酸水解降解的稳定性和良好的溶解特性。

[0032] 核酸还可以包括核碱基（通常简称为“碱基”）修饰或取代。如本文使用的，“未修饰的”或“天然的”核碱基可以包括嘌呤碱基（例如腺嘌呤（A）和鸟嘌呤（G））、以及嘧啶碱基（例如胸腺嘧啶（T）、胞嘧啶（C）和尿嘧啶（U））。修饰的核碱基可以包括其他合成以及天然的核碱基，如5-甲基胞嘧啶（5-me-C），5-羟甲基胞嘧啶，黄嘌呤，次黄嘌呤，2-氨基腺嘌呤，腺嘌呤和鸟嘌呤的6-甲基以及其他烷基衍生物，腺嘌呤和鸟嘌呤的2-丙基以及其他烷基衍生物，2-硫尿嘧啶，2-硫胸腺嘧啶以及2-硫胞嘧啶，5-卤代尿嘧啶以及胞嘧啶，5-丙炔基（—C≡C—CH₃）尿嘧啶及胞嘧啶以及嘧啶碱基的其他炔基衍生物，6-偶氮基尿嘧啶，胞嘧啶以及

胸腺嘧啶,5-尿嘧啶(假尿嘧啶),4-硫尿嘧啶,8-卤基、8-氨基、8-巯基、8-硫烷基、8-羟基以及其他8-取代的腺嘌呤和鸟嘌呤,5-卤基特别是5-溴、5-三氟甲基以及其他5-取代的尿嘧啶和胞嘧啶、7-甲基鸟嘌呤和7-甲基腺嘌呤、2-F-腺嘌呤、2-氨基腺嘌呤、8-氮杂鸟嘌呤和8-氮杂腺嘌呤、7-脱氮杂鸟嘌呤和7-脱氮杂腺嘌呤、以及3-脱氮杂鸟嘌呤和3-脱氮杂腺嘌呤。修饰的核碱基可以包括三环嘧啶如吩噻嗪胞苷(1H-嘧啶并(5,4-b)(1,4)苯并噻嗪-2(3H)-酮)、吩噻嗪胞苷(1H-嘧啶并(5,4-b)(1,4)苯并噻嗪-2(3H)-酮),G-夹(clamp)如取代的吩噻嗪胞苷(例如9-(2-氨基乙氧基)-H-嘧啶并(5,4-(b)(1,4)苯并噻嗪-2(3H)-酮),咪唑胞苷(2H-嘧啶并(4,5-b)咪唑-2-酮),吡啶并咪唑胞苷(H吡啶并(3',':4,5)吡咯并[2,3-d]嘧啶-2-酮)。

[0033] 如本文使用的,术语“准对称的随机条形码化的核酸”可以指包括本披露的随机条形码和末端的分子,所述末端足够对称以杂交在一起而形成锅柄结构(例如,用于抑制PCR)但可以不是相同的。准对称的随机条形码化的核酸可以表现为类似于对称的核酸,但具有不对称的序列。

[0034] 如本文使用的,术语“样品”可以指包括靶标的组合物。用于通过披露的方法、装置和系统进行分析的合适样品包括细胞、单细胞、组织、器官或生物体。

[0035] 如本文使用的,术语“采样装置”或“装置”可以指可以取一部分样品和/或将所述部分放置在基底上的装置。采样装置可以指例如荧光激活细胞分选(FACS)机、细胞分选机、活检针、活检装置、组织切片装置、微流体装置、叶栅和/或超薄切片机。

[0036] 如本文使用的,术语“固体支持物”可以指可以附接多个随机条形码的离散固体或半固体表面。固体支持物可以包括任何类型的实心的、多孔的或空心的球体、球、承座、圆柱体或其他类似配置,其由塑料、陶瓷、金属或高分子材料(例如,水凝胶)构成,其上可以固定核酸(例如,共价地或非共价地)。固体支持物可以包括可以是球形的(例如,微球)或具有非球形或不规则形状的离散颗粒,所述形状是如立方形、长方形、锥形、圆柱形、圆锥形、椭圆形或圆盘形等。以阵列间隔开的多个固体支持物可以不包括基底。固体支持物可以与术语“珠”互换使用。如本文使用的,“固体支持物”和“基底”可以互换使用。

[0037] 如本文使用的,术语“随机条形码”可以指包括本披露标记的多核苷酸序列。随机条形码可以是可用于随机条形码化的多核苷酸序列。随机条形码可用于对样品中的靶标定量。随机条形码可用于控制标记与靶标相关联后可能发生的错误。例如,随机条形码可用于评估扩增或测序错误。与靶标相关联的随机条形码可以称为随机条形码-靶标或随机条形码-标签-靶标。

[0038] 如本文使用的,术语“随机条形码化”可以指核酸的随机标记(例如,条形码化)。随机条形码化可以利用递归泊松策略来关联并对与靶标相关的标记进行定量。如本文使用的,术语“随机条形码化”可以与“随机标记”互换使用。

[0039] 如本文使用的,术语“靶标”可以指可与随机条形码相关联的组合物。用于通过披露的方法、装置和系统进行分析的示例性合适靶标包括寡核苷酸、DNA、RNA、mRNA、微小RNA、tRNA等。靶标可以是单链的或双链的。在一些实施例中,靶标可以是蛋白质。在一些实施例中,靶标是脂质。如本文使用的,“靶标”可以与“种类”互换使用。

[0040] 术语“逆转录酶”可以指具有逆转录酶活性(即,催化从RNA模板合成DNA)的一组酶。通常,这样的酶包括但不限于逆转录病毒逆转录酶、逆转录转座子逆转录酶、逆转录质

粒逆转录酶、逆转录子逆转录酶、细菌逆转录酶、II型内含子衍生的逆转录酶,及其突变体、变体或衍生物。非逆转录病毒逆转录酶包括非LTR逆转录转座子逆转录酶、逆转录质粒逆转录酶、逆转录子逆转录酶和II型内含子逆转录酶。II型内含子逆转录酶的实例包括乳酸乳球菌L1.LtrB内含子逆转录酶、细长嗜热聚球藻 (*Thermosynechococcus elongatus*) TeI4c 内含子逆转录酶或嗜热脂肪土芽孢杆菌GsI-IIIC内含子逆转录酶。其他类别的逆转录酶可以包括许多类型的非逆转录病毒逆转录酶(即,逆转录子、II型内含子、以及多样性产生型逆转录元件等等)。

除去高丰度种类的方法

[0041] 本文披露的一些实施例提供了从多个核酸分子中除去高丰度种类的方法。在一些实施例中,本文披露的方法可以减少来自多个核酸分子的高丰度种类的含量,而没有从所述多个核酸分子中显著除去低丰度种类或中等丰度种类。如本文使用的,“显著除去”是指从多个核酸分子中除去低丰度种类或中等丰度种类的至少10%、至少20%、至少30%、至少40%、至少50%或更多。在一些实施例中,本文披露的方法可以从多个核酸分子中除去高丰度种类和中等丰度种类,而没有从所述多个核酸分子中显著除去低丰度种类。

[0042] 如本文使用的,“种类”是指所述多个核酸分子中的多核苷酸(例如,单链多核苷酸),它们彼此相同或互补,或者能够彼此杂交,或者是来自相同遗传基因座的转录物,或者编码相同蛋白质或其片段等。在一些实施例中,一个种类的成员彼此或与其互补体至少80%、至少90%、至少95%、至少98%、至少99%或100%同源。在一些实施例中,一个种类的成员可以在高严格杂交条件下彼此杂交。在一些实施例中,一个种类的成员可以在中等严格杂交条件下彼此杂交。在一些实施例中,一个种类的成员可以在低严格杂交条件下彼此杂交。在一些实施例中,一个种类的成员是来自相同遗传基因座的转录物,并且所述转录物可以具有相同或不同的长度。在一些实施例中,所述种类是cDNA或mRNA。

[0043] 如本文使用的,“高丰度种类”是指在所述多个核酸中以高量存在的种类,例如所述种类可以代表所述多个核酸分子的至少1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、20%、30%、40%、50%或更多。在一些实施例中,所述多个核酸分子可以包括至少1个、至少2个、至少3个、至少4个、至少5个、至少10个、至少20个、至少50个、至少100个、至少200个、至少500个、至少1,000个或更多个高丰度种类。在一些实施例中,所有高丰度种类的总体代表所述多个核酸分子的至少10%、至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少80%或更多。在一些实施例中,高丰度种类可以包括编码一种或多种核糖体蛋白的多核苷酸。在一些实施例中,高丰度种类可以包括编码一种或多种线粒体蛋白的多核苷酸。在一些实施例中,高丰度种类可以包括编码一种或多种持家蛋白的多核苷酸。

[0044] 如本文使用的,“中等丰度种类”是指以低于所述多个核酸中的至少一个种类且高于所述多个核酸中的至少一个其他种类的量存在于所述多个核酸中的种类。在一些实施例中,中等丰度种类可以代表所述多个核酸分子的约10%、5%、4%、3%、2%、1%、0.1%、0.01%或任意两个上述值之间的范围。在一些实施例中,所述多个核酸分子可以包括至少1个、至少2个、至少3个、至少4个、至少5个、至少10个、至少20个、至少50个、至少100个、至少200个、至少500个、至少1,000个或更多个中等丰度种类。在一些实施例中,所有中等丰度种类的总体代表所述多个核酸分子的约1%、约2%、约3%、约4%、约5%、约10%、约20%、约30%或任意两个上述值之间的范围。

[0045] 如本文使用的,“低丰度种类”是指在所述多个核酸中以低量存在的种类,例如所述种类可以代表所述多个核酸分子的少于1%、0.1%、0.01%、0.001%、0.0001%或更少。在一些实施例中,所述多个核酸分子可以包括至少1个、至少2个、至少3个、至少4个、至少5个、至少10个、至少20个、至少50个、至少100个、至少200个、至少500个、至少1,000个或更多个低丰度种类。在一些实施例中,所有低丰度种类的总体代表所述多个核酸分子的少于20%、少于10%、少于5%、少于4%、少于3%、少于2%、少于1%、少于0.1%或更少。在一些实施例中,低丰度种类可以包括编码一种或多种转录因子的多核苷酸。在一些实施例中,高丰度种类可以包括编码一种或多种T细胞受体的多核苷酸。在一些实施例中,高丰度种类可以包括编码一种或多种抗体的多核苷酸。

[0046] 在一些实施例中,本文披露的方法和组合物可以减少来自所述多个核酸分子的一个或多个高丰度种类的含量。例如,本文披露的方法和组合物可以减少至少1个、至少2个、至少3个、至少4个、至少5个、至少10个、至少20个、至少50个、至少100个、至少200个、至少500个、至少1,000个或更多个高丰度种类的含量。在一些实施例中,本文披露的方法和组合物可以将来自所述多个核酸分子的一个或多个高丰度种类中每个的含量减少至少10%、至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%、至少95%或100%。在一些实施例中,本文披露的方法和组合物可以将来自所述多个核酸分子的一个或多个高丰度种类中至少一个的含量减少至少10%、至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%、至少95%或100%。在一些实施例中,本文披露的方法和组合物可以将来自所述多个核酸分子的全部高丰度种类的含量减少至少10%、至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%、至少95%或100%。

[0047] 在一些实施例中,本文披露的方法和组合物可以减少来自所述多个核酸分子的一个或多个高丰度种类的含量,而没有从所述多个核酸分子中显著除去低丰度种类或中等丰度种类。在一些实施例中,本文披露的方法和组合物可以将来自所述多个核酸分子的一个或多个高丰度种类中每个的含量减少至少10%、至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%、至少95%或100%,而没有从所述多个核酸分子中显著除去低丰度种类或中等丰度种类。在一些实施例中,本文披露的方法和组合物可以将来自所述多个核酸分子的全部高丰度种类的含量减少至少10%、至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%、至少95%或100%,而没有从所述多个核酸分子中显著除去低丰度种类或中等丰度种类。在一些实施例中,本文披露的方法和组合物可以减少来自所述多个核酸分子的一个或多个高丰度种类的含量,同时保持一个或多个低丰度种类中每个的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%或100%。在一些实施例中,本文披露的方法和组合物可以减少来自所述多个核酸分子的一个或多个高丰度种类的含量,同时保持一个或多个低丰度种类中至少一个的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%或100%。在一些实施例中,本文披露的方法和组合物可以减少来自所述多个核酸分子的一个或多个高丰度种类的含量,同时保持全部低丰度种类的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%或100%。在一些实施例中,本文披露的方法和组合物可以减少来自所述多个核酸分子的一个或多个高丰度种类的含量,同时保持一个或多个中等丰度种类中至少一个的至少50%、至

少60%、至少70%、至少80%、至少90%、至少95%或100%。在一些实施例中,本文披露的方法和组合物可以减少来自所述多个核酸分子的一个或多个高丰度种类的含量,同时保持全部中等丰度种类的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%或100%。在一些实施例中,本文披露的方法和组合物可以减少来自所述多个核酸分子的一个或多个高丰度种类的含量,同时保持来自所述多个核酸分子的中等丰度种类中每个的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%或100%。

多个核酸分子

[0048] 本文披露的多个核酸分子可以包括多种核酸分子。在一些实施例中,所述多个核酸分子可以包括DNA分子、RNA分子、基因组DNA分子、cDNA分子、mRNA分子、rRNA分子、siRNA分子或其组合,并且可以是双链的或单链的。在一些实施例中,所述多个核酸分子包括至少100个、至少1,000个、至少10,000个、至少20,000个、至少30,000个、至少40,000个、至少50,000个、至少100,000个、至少1,000,000个或更多种类。在一些实施例中,所述多个核酸分子可以来自样品,如单细胞或多个细胞。在一些实施例中,所述多个核酸分子可以从多个样品(如多个单细胞)中合并。

[0049] 在一些实施例中,所述多个核酸分子包括未标准化核酸文库、部分标准化核酸文库或通过其他方法标准化的核酸文库,如cDNA文库、基因组DNA文库等。在一些实施例中,所述多个核酸分子可以包括合并的非标准化核酸文库,例如由各自代表单细胞的多个非标准化核酸文库构建的合并的非标准化核酸文库。在一些实施例中,所述未标准化核酸文库是cDNA文库。在一些实施例中,所述未标准化核酸文库是基因组文库。在一些实施例中,所述未标准化核酸文库是单细胞核酸文库。如本文使用的,“单细胞核酸文库”意指源自单细胞的核酸分子(如基因组DNA或mRNA分子)的集合。在一些实施例中,单细胞核酸文库可以指源自多个单细胞的核酸分子的集合,其中所述核酸分子包括细胞标记以鉴定所述核酸分子源头的单细胞。

[0050] 在一些实施例中,可以在除去高丰度种类之前对所述多个核酸分子进行扩增。例如,所述多个核酸分子可以包括扩增的核酸文库。在一些实施例中,所述多个核酸分子可以包括每个核酸分子的至少2个、至少4个、至少8个、至少16个、至少100个、至少1,000个或更多拷贝。

结合部分

[0051] 在一些实施例中,本文披露的方法包括使包括结合部分的多个第一寡核苷酸与所述多个核酸分子杂交。可将多种结合部分用于本文披露的方法和组合物。例如,结合部分可以是结合对的一部分。在一些实施例中,所述结合部分可以是添加至寡核苷酸的官能团。在一些实施例中,所述结合部分可以是生物素、链霉亲和素、肝素、适体、点击化学部分、地高辛、一个或多个伯胺、一个或多个羧基、一个或多个羟基、一个或多个醛、一个或多个酮或其任何组合。

[0052] 本文披露的结合部分能够结合到捕获部分如捕获分子。在一些实施例中,所述结合部分和捕获分子可以是结合对的成员,例如生物素/链霉亲和素。所述捕获分子可以被固定在固体支持物(如珠、微粒或纳米颗粒)上。

[0053] 在一些实施例中,所述第一寡核苷酸可被延伸以产生包括所述结合部分的所述多个核酸靶标的多条互补链。在一些实施例中,第二链可以使用与互补链上的结合位点结合

的引物来合成以产生双链核酸分子。

通过变性/部分再退火减少高丰度种类的含量

[0054] 在一些实施例中,减少高丰度种类的含量可以包括变性,随后对双链核酸分子进行部分再退火,随后通过固定在一个或多个固体支持物上的捕获分子除去所述多个核酸靶标的再退火互补链,其中所述捕获分子与所述结合部分特异性结合。

[0055] 变性可以通过多种方法进行,包括加热双链核酸分子、用有机溶剂(例如DMS或甲酰胺)处理双链核酸分子、改变双链核酸分子的盐浓度和/或改变双链核酸分子的pH。

[0056] 变性后,单链核酸分子可以部分再退火。可以通过任何方法进行部分再退火,例如在冰上快速冷却、改变盐浓度(例如,将盐浓度从用于变性的量逆转)和/或改变pH(例如,将pH从用于变性的水平逆转)等。

[0057] 可以理解,再退火的程度可以根据各种因素进行调整,包括但不限于待除去种类的类型(例如高丰度种类和/或中等丰度种类)、待除去的高丰度种类的希望百分比和/或待保留的中等或低丰度种类的百分比。不受任何特定理论的限制,相信在相同的退火条件下,丰度更大的种类(例如,高丰度种类)比丰度更低的种类(例如,中等丰度种类和低丰度种类)更快退火。例如,通过改变再退火步骤的温度、盐浓度、pH和/或持续时间,可以调整待除去的高丰度种类的百分比和/或待保留的中等丰度种类或低丰度种类的百分比。在一些实施例中,可以调整再退火步骤的温度、盐浓度、pH和/或持续时间,使得高丰度种类的至少10%、至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%、至少95%或至少99%再退火。在一些实施例中,可以调整再退火步骤的温度、盐浓度、pH和/或持续时间,使得中等丰度种类或低丰度种类的至少10%、至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%、至少95%或100%保持单链。

[0058] 图1描绘了本披露的方法的示例性实施例。样品可以包括多个核酸。一些核酸可以是丰度较高的106,并且一些核酸可以是丰度较低的105。所述核酸可以被转化为用结合部分120不对称标记的双链cDNA文库125。例如,核酸105/106可以是逆转录的mRNA。第二链合成可以使用包括结合部分120的引物(例如,基因特异性引物或随机多聚体引物)进行,从而产生不对称标记的双链核酸文库125。在另一个实例中,核酸105/106可以是DNA。可使用包括结合部分120的引物延伸DNA,从而产生不对称标记的双链cDNA文库125。所述双链cDNA文库可包括高丰度双链cDNA种类116和低丰度双链cDNA种类115。双链cDNA文库125可以连接到衔接子130/135,从而产生非标准化文库125。

[0059] 双链cDNA文库125可以被热变性,从而分离所述双链cDNA的链。所述热变性文库可以被再退火(例如,部分再退火)140。丰度较大的核酸106可以比丰度较小的核酸105更快地退火。变性和部分再退火可得到种类混合物141,其包括:包括所述结合部分的再退火双链cDNA 142、包括所述结合部分的单链分子143和缺乏所述结合部分的单链分子144。

[0060] 所述变性和再退火的文库可与固体支持物145接触。所述固体支持物可包括可结合到不对称标记的双链cDNA 115/116的结合部分120的捕获部分。包括结合部分120的核酸可以被所述固体支持物结合。这些核酸可以包括:包括所述结合部分的再退火的双链cDNA 142、包括所述结合部分的单链分子143。可以使用磁体来除去与核酸结合的固体支持物145。

[0061] 留下的核酸150代表不退火的种类的互补体(例如,这是包括结合部分并被固体支

持物除去的双链的互补体,但其浓度足够低,使得它不与那个单链分子退火)。留下的核酸150可能不包含结合部分120。留下的核酸150可以被扩增和/或与衔接子连接以用于测序。这些核酸代表标准化文库。在一些情况下,文库标准化可能不需要使用酶学来标准化所述文库(例如,核酸酶)。

文库标准化方法

[0062] 本披露提供了用于文库标准化的方法。可以在核酸样品上进行本披露的方法。所述核酸样品可以包括核酸。所述样品可以来自本披露的样品。所述样品可以是单细胞(例如,所述核酸样品可以是来自单细胞的核酸)。所述核酸样品可以包括RNA、DNA或RNA和DNA两者。所述样品的核酸可以是单链的、双链的或单链和双链两者的混合物。在一些情况下,所述样品的全部或大部分核酸都是单链的。

[0063] 所述核酸样品可以包括具有不同丰度的核酸靶标(即核酸种类,在本文中与核酸靶标可互换使用)。例如,所述核酸样品可以包括高丰度靶标(例如,肌动蛋白、GapDH、球蛋白、持家基因)。所述核酸样品可以包括低丰度靶标(例如,来自干细胞或循环肿瘤细胞(CTC)的稀少靶标,或低表达基因)。所述核酸样品可以包括高丰度靶标和低丰度靶标的混合物。

[0064] 所述样品的核酸可以与包括结合部分的引物接触以产生双链cDNA。在一些实施例中,所述核酸可以是RNA并且所述引物可以是逆转录引物。所述逆转录引物可以逆转录RNA,从而产生RNA-cDNA杂交体(例如,第一链合成)。可以使用标准的第二链合成技术产生第二链。所述第一cDNA链可以包括所述结合部分。所述第二链可以是所述第一链的互补体。所述第二链可能不包括所述结合部分。此cDNA可以被称为不对称标记的cDNA(例如,cDNA的一条链被所述结合部分标记)。

[0065] 当所述样品的核酸是DNA时,可以使包括结合部分的引物与所述核酸接触以产生第一链(例如,DNA模板的互补链)。所述第一链可以包括所述结合部分。在一些情况下,所述引物可以产生与所述第一链互补的第二链。所述第二链可能不包括所述结合部分。

[0066] 用RNA或DNA起始模板,结果可以是不对称标记的双链cDNA分子,其中一条链包括结合部分,并且另一条链是互补体。一组不对称标记的双链cDNA分子可以被称为非标准化文库。

[0067] 在一些实施例中,所述结合部分通过在产生文库(例如,第一链、第二链、衔接子连接)后发生的扩增反应来添加。如图3中所示,核酸(例如,mRNA、DNA) 305可以用包括分子标记(即,分子索引) 310、样品标记(即,细胞标记、样品索引) 315和通用标记(即,通用引物结合序列) 320的引物进行逆转录或延伸。所述产物可以经历第一巢式PCR扩增反应325,其中使用巢式基因特异性引物330和结合通用标记320的通用引物335来扩增所述产物,从而产生扩增子。在一些情况下,所述扩增子可以经历使用巢式PCR引物345和包括结合通用标记320的结合部分355的通用引物350的第二轮扩增。所述扩增反应可以产生用结合部分不对称标记的扩增子(例如,巢式PCR扩增子)。图3描绘了添加结合部分的代表性扩增方案。所述结合部分可以在任何步骤添加,如逆转录时、第二链合成时、衔接子连接之前、衔接子连接之后和/或任何PCR扩增步骤时。包括所述结合部分的引物可用于文库制备方案的多个步骤中。

[0068] 所述结合部分可以是具有结合伴侣的任何小分子。示例性的结合部分包括生物

素、链霉亲和素、肝素、适体、点击化学部分、蛋白质结合区段或结构以及核酸结合区段或结构等。

[0069] 本文所述的方法和组合物解决了文库标准化期间ssDNA和dsDNA部分的物理和酶促分离的挑战。如图5A和5B中所示,未标准化文库500包含高丰度种类505和低丰度种类510。在初始文库制备期间,所述文库或所述文库的一部分在一端用结合部分515(如生物素等)不对称标记。将标记的双链核酸分子520进行变性530以产生包括高丰度种类535和低丰度种类540的单链核酸分子。将所述单链核酸分子部分再退火550以形成高丰度种类555的双链分子,而低丰度种类保持单链560。变性和部分再退火后,将所有标记的链捕获在支持基质565如顺磁性链霉亲和素珠上,并分离结合和未结合的部分570。高丰度序列将主要重新杂交,并且在结合的部分中两条链都将被除去。然而,低丰度序列将不会再退火,因此标记的链的互补体将存在于未结合的部分中。包括单链高丰度种类575和单链低丰度种类580的未结合的部分将代表标准化库585,并且可以直接使用或进一步扩增590用于下游应用。

变性和部分再退火

[0070] 所述未标准化文库(例如,包括核酸靶cDNA)可被变性。可以通过多种方法进行变性,包括加热样品、用有机溶剂(例如,DMS或甲酰胺)处理样品、改变样品的盐浓度和/或改变样品的pH。

[0071] 在一些情况下,通过加热所述样品进行变性。变性可以在至少50°C、60°C、70°C、80°C、90°C或95°C或更高的温度下进行。变性可以在至多50°C、60°C、70°C、80°C、90°C或95°C或更高的温度下进行。变性可以进行至少1、2、3、4、5、6、7、8、9或10或更多分钟。变性可以进行至多1、2、3、4、5、6、7、8、9或10或更多分钟。变性可以导致至少50%、60%、70%、80%、90%或100%的cDNA变性。变性可以导致至多50%、60%、70%、80%、90%或100%的cDNA变性。变性可以导致至少50%、60%、70%、80%、90%或100%的核酸呈单链形式。

[0072] 所述变性样品可以再退火。所述变性样品可以部分再退火。可以通过任何方法进行部分再退火,例如在冰上快速冷却、改变盐浓度(例如,将盐浓度从用于变性的量逆转)和/或改变pH(例如,将pH从用于变性的水平逆转)等。在一些情况下,部分再退火包括冷却所述变性样品(例如,在冰上)。部分再退火可以包括所述变性样品的至少10%、20%、30%、40%、50%、60%、70%、80%、90%或100%的链的再退火。部分再退火可以包括所述变性样品的至多10%、20%、30%、40%、50%、60%、70%、80%、90%或100%的链的再退火。至少10%、20%、30%、40%、50%、60%、70%、80%、90%或100%来自高丰度核酸的链可在部分再退火步骤期间再退火。至多10%、20%、30%、40%、50%、60%、70%、80%、90%或100%来自高丰度核酸的链可在部分再退火步骤期间再退火。至少10%、20%、30%、40%、50%、60%、70%、80%、90%或100%来自较低丰度核酸的链可在部分再退火步骤期间再退火。至多10%、20%、30%、40%、50%、60%、70%、80%、90%或100%来自较低丰度核酸的链可在部分再退火步骤期间再退火。

[0073] 来自较高丰度种类的链可以比来自较低丰度种类的链再退火快至少50%、100%、150%、200%、250%、300%、350%、400%、450%或500%或更多。来自较高丰度种类的链可以比来自较低丰度种类的链再退火快至多50%、100%、150%、200%、250%、300%、350%、400%、450%或500%或更多。来自较高丰度种类的链可以比来自较低丰度种类的链再退火至少1、2、3、4、5、6、7、8、9或10倍或更多。来自较高丰度种类的链可以比来自较低丰度种类

的链再退火至多1、2、3、4、5、6、7、8、9或10倍或更多。

[0074] 再退火之后,所述样品可以包括多个双链分子、包括结合部分的多个单链分子以及可能不包括所述结合部分的多个单链分子。所述双链分子可以包括可包括所述结合部分的一条链和不包括所述结合部分的一条链。

[0075] 可以使所述再退火的样品与固体支持物接触。所述固体支持物可以包括捕获部分。捕获部分可结合到所述结合部分。例如,如果所述结合部分是生物素,则所述捕获部分可以是链霉亲和素。所述固体支持物可以是例如树脂、浆液、珠、树脂、水凝胶、半固体支持物、不溶性支持物和/或半固体支持物。在一些情况下,所述固体支持物是树脂。

[0076] 所述固体支持物可以具有物理性质。例如,在某些pH条件、盐条件和/或温度条件下,所述固体支持物可以是可溶的。所述固体支持物可以是磁性的、铁磁性的和/或顺磁性的。

[0077] 所述固体支持物可结合到包括所述结合部分的再退火样品的分子。所述固体支持物可结合到双链cDNA(例如,包括所述结合部分)。所述固体支持物可结合到单链未退火的链(例如,包括所述结合部分)。

[0078] 所述固体支持物可用于纯化所述样品。所述固体支持物可以与所述样品(即,上清液)分离(例如通过离心、磁力)。剩余样品(即,上清液)可被称为标准化文库。所述标准化文库可以包括可能不包括结合部分的单链核酸分子。所述单链核酸分子可以是在所述方法的部分再退火步骤期间不退火的分子。

[0079] 与未标准化文库相比,所述标准化文库可以包括相对较多的相当量的较低丰度种类和较高丰度种类。例如,与未标准化文库相比,在标准化文库中较高丰度种类可以大于较低丰度种类至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20倍或更多倍。与未标准化文库相比,在标准化文库中较高丰度种类可以大于较低丰度种类至多1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20倍或更多倍。与未标准化文库相比,在标准化文库中较低丰度种类可以大于较高丰度种类至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20倍或更多倍。与未标准化文库相比,在标准化文库中较低丰度种类可以大于较高丰度种类至多1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20倍或更多倍。

阻滞剂的使用

[0080] 在一些实施例中,本披露的方法提供了在变性和部分再退火期间阻滞剂的使用。如本文使用的,阻滞剂是指可以与文库的通用序列(例如,通用引物序列、通用测序流动池序列)杂交的寡核苷酸序列。阻滞剂可用于阻止来自不同基因的靶标/扩增子在进行部分再退火期间通过它们的通用区域一起退火,而不考虑基因序列。

[0081] 图4中示出了阻滞剂的示例性使用。第一靶标405和第二靶标410可各自包括通用序列415/420和结合部分421。在一些情况下,在变性和部分再退火期间,具有差异序列的靶标405/410可通过各自包括415/420的通用序列一起退火。阻滞剂可用于阻止这种情况发生。可以使所述样品与阻滞剂430接触425。阻滞剂430可以与所述靶标的一个或多个通用序列杂交。可以使用一种或多种不同类型的阻滞剂430。所述阻滞剂可以通过强制链通过其基因序列缔合(例如,杂交)来帮助部分再退火。以这种方式,可以使用阻滞剂来帮助本披露的文库标准化方法。

[0082] 阻滞剂的长度可以是至少1、2、3、4、5、6、7、8、9或10个或更多个核苷酸。阻滞剂的长度可以是至多1、2、3、4、5、6、7、8、9或10个或更多个核苷酸。可以将阻滞剂与其靶标以至少10%、20%、30%、40%、50%、60%、70%、80%、90%或100%互补性进行杂交。可以将阻滞剂与其靶标以至多10%、20%、30%、40%、50%、60%、70%、80%、90%或100%互补性进行杂交。

[0083] 在一些实施例中,本披露的文库标准化方法可以在固体支持物上进行。例如,可以产生文库,其中所述文库的扩增子用参与点击化学的分子之一(例如,叠氮化物、炔,用于叠氮化物-炔环加成)不对称地标记。所述固体支持物可以包括点击化学中的另一分子。例如,所述扩增子可以包括炔并且所述固体支持物可以包括叠氮化物。所述扩增子可以附接到所述固体支持物上(例如,通过点击化学)。可以加热所述固体支持物,从而诱导附接的扩增子的变性。在部分再退火期间,丰度较大的扩增子可以再退火至附接于固体支持物的分子。丰度较小的扩增子可以留在溶液中(例如,通过离心、磁力、层析)。可以将所述固体支持物从溶液中除去,从而留下标准化文库。

扩增

[0084] 可以进行一个或多个核酸扩增反应以产生标准化的靶核酸分子的多个拷贝。扩增能以多重方式进行,其中多个靶核酸序列同时进行扩增。所述扩增反应可用于将测序衔接子添加至所述核酸分子。所述扩增反应可以包括扩增至少一部分的样品标记(如果存在的话)。所述扩增反应可以包括扩增至少一部分的细胞和/分子标记。所述扩增反应可以包括扩增至少一部分的样品标签、细胞标记、空间标记、分子标记、靶核酸或其组合。所述扩增反应可以包括扩增所述多个核酸的至少1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、97%或100%。所述方法还可以包括进行一个或多个cDNA合成反应,从而产生含有样品标记、细胞标记、空间标记和/或分子标记的靶标-条形码分子的一个或多个cDNA拷贝。

[0085] 在一些实施例中,可以使用聚合酶链式反应(PCR)进行扩增。如本文使用的,PCR可以指用于通过DNA的互补链的同时引物延伸使特异性DNA序列体外扩增的反应。如本文使用的,PCR可包括所述反应的派生形式,包括但不限于RT-PCR、实时PCR、巢式PCR、定量PCR、多重PCR、数字PCR和组装PCR。

[0086] 标记的核酸的扩增还可以包括基于非PCR的方法。基于非PCR的方法的实例包括但不限于多重置换扩增(MDA)、转录介导的扩增(TMA)、全转录组扩增(WTA)、全基因组扩增(WGA)、基于核酸序列的扩增(NASBA)、链置换扩增(SDA)、实时SDA、滚环扩增或环到环扩增(circle-to-circle amplification)。其他基于非PCR的扩增方法包括DNA依赖性RNA聚合酶驱动的RNA转录扩增或RNA指导的DNA合成和转录的多个循环以扩增DNA或RNA靶标、连接酶链式反应(LCR)和QB复制酶(QB)方法、回文探针的使用、链置换扩增、使用限制性内切核酸酶的寡核苷酸驱动的扩增、使引物杂交至核酸序列并且将所得双链体在延伸反应和扩增之前切割的扩增方法、使用缺乏5'外切核酸酶活性的核酸聚合酶的链置换扩增、滚环扩增和分支延伸扩增(RAM)。在一些情况下,所述扩增可能不产生环化转录物。

[0087] 抑制PCR可用于本披露的扩增方法。抑制PCR可以指选择性地排除小于某一尺寸的侧翼为末端反向重复序列的分子,这是因为当用于扩增的一个或多个引物对应于整个重复

序列或重复序列的一部分时,所述分子无效扩增。对此原因可在于片段互补端的生产性PCR引物退火和非生产性自退火之间的平衡。在侧翼末端反向重复序列的尺寸固定的情况下,插入物越短,抑制效果越强,反之亦然。同样,在插入物尺寸固定的情况下,末端反向重复序列越长,抑制效果越强。

[0088] 抑制PCR可以使用在PCR扩增之前连接到DNA片段末端的衔接子。在解链和退火时,在链的5'端和3'端具有自互补衔接子的单链DNA片段可形成抑制PCR过程中片段扩增的抑制性“网球拍”形结构。

[0089] 在一些情况下,本文披露的方法还包括对标记的核酸(例如,标记的RNA、标记的DNA、标记的cDNA)进行聚合酶链式反应,以产生随机标记的扩增子。所述随机标记的扩增子可以是双链分子。所述双链分子可以包括或是双链RNA分子、双链DNA分子或者与DNA分子杂交的RNA分子。所述双链分子的一条或两条链可以包括样品标记、空间标记、细胞标记和/或分子标记。所述随机标记的扩增子可以是单链分子。所述单链分子可以包括DNA、RNA或其组合。本披露的核酸可以包括合成的或改变的核酸。

[0090] 扩增可以包括使用一种或多种非天然核苷酸。非天然核苷酸可以包括光不稳定或可触发的核苷酸。非天然核苷酸的实例可以包括但不限于肽核酸(PNA)、吗啉代和锁核酸(LNA)以及二醇核酸(GNA)与苏糖核酸(TNA)。可以将非天然核苷酸添加至扩增反应的一个或多个循环中。添加所述非天然核苷酸也可以用于鉴定扩增反应中特定循环或时间点的产物。

[0091] 进行一个或多个扩增反应可以包括使用一个或多个引物。所述一个或多个引物可以包括至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、或15个或更多个核苷酸。所述一个或多个引物可以包括至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、或15个或更多个核苷酸。所述一个或多个引物可以包括少于12-15个核苷酸。所述一个或多个引物可以退火至所述多个随机标记的靶标的至少一部分。所述一个或多个引物可以退火至所述多个随机标记的靶标的3'末端或5'末端。所述一个或多个引物可以退火至所述多个随机标记的靶标的内部区域。所述内部区域可以是来自所述多个随机标记的靶标的3'末端的至少约50、100、150、200、220、230、240、250、260、270、280、290、300、310、320、330、340、350、360、370、380、390、400、410、420、430、440、450、460、470、480、490、500、510、520、530、540、550、560、570、580、590、600、650、700、750、800、850、900或1000个核苷酸。所述一个或多个引物可以包括一组固定的引物。所述一个或多个引物可以包括至少一个或多个定制引物。所述一个或多个引物可以包括至少一个或多个对照引物。所述一个或多个引物可以包括至少一个或多个基因特异性引物。

[0092] 所述一个或多个引物可以包括本披露的任何通用引物。所述通用引物可以退火至通用引物结合位点。所述一个或多个定制引物可以退火至第一样品标记、第二样品标记、空间标记、细胞标记、分子标记、靶标或其任何组合。所述一个或多个引物可以包括通用引物和定制引物。可以将所述定制引物设计成扩增一个或多个靶标。所述靶标可以包括在一个或多个样品中的总核酸的亚组。所述靶标可以包括在一个或多个样品中的总随机标记的靶标的亚组。所述一个或多个引物可以包括至少96个或更多个定制引物。所述一个或多个引物可以包括至少960个或更多个定制引物。所述一个或多个引物可以包括至少9600个或更多个定制引物。所述一个或多个定制引物可以退火至两个或更多个不同的标记的核酸。所

述两个或更多个不同的标记的核酸可以对应于一个或多个基因。

[0093] 可以在本披露的方法中使用任何扩增方案。例如,在一个方案中,第一轮PCR可以使用基因特异性引物和针对通用亿明达(Illumina)测序引物1序列的引物来扩增分子(例如,附接到珠上)。第二轮PCR可以使用侧翼于亿明达测序引物2序列的巢式基因特异性引物和针对通用亿明达测序引物1序列的引物扩增第一PCR产物。第三轮PCR添加P5和P7以及样品索引,以便使PCR产物进入亿明达测序文库。使用150bp x 2测序的测序可以揭示读段1上的细胞标记和分子索引、读段2上的基因以及索引1读段上的样品索引。

[0094] 扩增可以进行一轮或多轮。在一些情况下,存在多轮扩增。扩增可以包括两轮或更多轮扩增。第一次扩增可以是X'外的延伸以产生基因特异性区域。当样品核酸与新产生的链杂交时,可发生第二次扩增。

[0095] 在一些实施例中,杂交不需要发生在核酸分子的末端。在一些实施例中,将较长核酸的完整链内的靶核酸杂交并扩增。例如在基因组DNA或mRNA的较长部分内的靶标。靶标可以是离多核苷酸的一端多于50nt、多于100nt、或多于1000nt。

包括衔接子连接的文库制备

[0096] 可以制备未标准化文库(或标准化文库)的单链分子用于测序,其可以例如包括产生双链分子和并入流动池测序衔接子(例如,通过连接和/或杂交和PCR)。

[0097] 在一些实施例中,衔接子可以连接至所述双链核酸。衔接子可以包括本披露的第一通用引物序列、本披露的第二通用引物序列和限制性内切核酸酶结合位点或其任何组合。在一些情况下,所述衔接子包括本披露的第二通用引物序列和限制性内切核酸酶结合位点。

[0098] 本文使用的术语“衔接子”是指可附接到核酸末端的具有至少10、15、20或25个碱基的单链的或双链的寡核苷酸。衔接子序列可以使用例如引发位点、引发位点的互补体、和内切核酸酶的识别位点、共同序列和启动子来合成。所述衔接子可以完全是或基本上是双链的。双链衔接子可以包括至少部分互补的两个寡核苷酸。所述衔接子可以在一条或两条链上磷酸化或未磷酸化。所述衔接子可以具有与突出端(例如,由限制性酶或聚合酶产生的)完全或部分互补的双链部分和单链突出部分。所述衔接子中的突出端可以是例如4到8个碱基。例如,当用限制性酶EcoRI消化DNA时,所得双链片段的任一末端的侧翼为单链突出端5'-AATT-3',携带单链突出端5'-AATT-3'的衔接子可以通过突出区域之间的互补性与所述片段杂交。所述衔接子与所述片段的这种“粘性末端”杂交促进所述衔接子与所述片段的连接,然而,平端连接也是可能的。可使用例如Klenow片段的外切核酸酶活性将平端转化为粘性末端。例如当用PvuII消化DNA时,通过在dTTP和dCTP的存在下将所述片段与Klenow一起孵育,可以将平端转化成两个碱基对突出端。突出端也可以通过填充突出端或除去突出端而转化为平端。

[0099] 衔接子可以连接到本披露的双链cDNA。连接方法可以包括使用T4DNA连接酶,其催化在具有平端和粘性末端的双链体DNA或RNA中并列的5'磷酸和3'羟基末端之间形成磷酸二酯键;使用Taq DNA连接酶,其催化在与互补靶DNA杂交的两个相邻寡核苷酸中并列的5'磷酸和3'羟基末端之间形成磷酸二酯键;使用大肠杆菌DNA连接酶,其催化在含有粘性末端的双链体DNA中并列的5'-磷酸和3'-羟基末端之间形成磷酸二酯键;以及使用T4RNA连接酶,其催化通过形成3'到5'磷酸二酯键使末端为5'磷酸基的核酸供体与末端为3'羟基的核

酸受体连接,底物包括单链RNA和DNA以及二核苷焦磷酸;或者本领域中描述的任何其他方法。不同的酶产生不同的突出端,并且所述衔接子的突出端可以靶向连接到由选择的限制性酶产生的片段。

[0100] 在一些实施例中,使用双链衔接子,并且将所述衔接子的仅一条链连接到所述双链cDNA。可以选择性地阻断衔接子的一条链的连接。为了阻断连接,例如,可以将所述衔接子的一条链设计为在所述衔接子的此链的5'端和所述靶核酸的3'端之间引入一个或多个核苷酸的缺口。从衔接子的5'端缺少磷酸酯可以阻断此5'端与可用的3'OH的连接。

测序

[0101] 确定不同的随机标记的核酸的数目可以包括确定标记的靶标、空间标记、分子标记、样品标记和细胞标记或其任何产物(例如标记的扩增子、标记的cDNA分子)的序列。扩增的靶标可以经受测序。确定所述随机标记的核酸或其任何产物的序列可以包括进行测序反应以确定样品标记、空间标记、细胞标记、分子标记的至少一部分和/或所述随机标记的靶标、其互补体、其反向互补体的至少一部分、或其任何组合的序列。

[0102] 可以使用多种测序方法确定核酸的序列(例如,扩增的核酸、标记的核酸、标记的核酸的cDNA拷贝等),所述方法包括但不限于合成测序(SBS)、杂交测序(SBH)、连接法测序(SBL)、量化增量荧光核苷酸附加测序(quantitative incremental fluorescent nucleotide addition sequencing,QIFNAS)、分段连接与断裂、荧光共振能量转移(FRET)、分子信标、TaqMan报告探针消化、焦磷酸测序、荧光原位测序(FISSEQ)、FISSEQ珠、摆动测序(wobble sequencing)、多重测序、聚合集群(polymerized colony,POLONY)测序;纳米格滚环测序(nanogrid rolling circle sequencing,ROLONY)、等位基因特异性寡核苷酸连接检验(allele-specific oligo ligation assay)(例如,寡核苷酸连接检验(OLA)、使用连接的线性探针和滚环扩增(RCA)读出、连接的持锁探针的单模板分子(single template molecule)OLA、或使用连接的环形持锁探针和滚环扩增(RCA)读出的单模板分子OLA)等。

[0103] 在一些情况下,确定所述标记的核酸或其任何产物的序列包括配对末端测序、纳米孔测序、高通量测序、鸟枪法测序、染料终止剂测序、多重引物DNA测序、引物步移、桑格双脱氧测序法、马克西姆-吉尔伯特(Maxim-Gilbert)测序、焦磷酸测序、真正的单分子测序或其任何组合。可替代地,可以通过电子显微镜分析法或化学敏感场效应晶体管(chemFET)阵列来确定所述标记的核酸或其任何产物的序列。

[0104] 也可以使用高通量测序方法,如使用平台(如Roche 454,Illumina Solexa,ABI-SOLiD,ION Torrent,Complete Genomics,Pacific Bioscience,Helicos、或Polonator平台)的循环阵列测序。测序可以包括MiSeq测序。测序可以包括HiSeq测序。

[0105] 所述随机标记的靶标可以包括代表来自生物体基因组基因的约0.01%至生物体基因组基因的约100%的核酸。例如,可以使用包括多个多聚体的靶标互补区域,通过从所述样品中捕获含有互补序列的基因,对约0.01%的生物体基因组基因至约100%的生物体基因组基因进行测序。在一些实施例中,所述标记的核酸包括代表从约0.01%的生物体转录组转录物至约100%的生物体转录组转录物的核酸。例如,可以使用包括聚T尾的靶标互补区域,通过从所述样品中捕获mRNA,对约0.501%的生物体转录组转录物至约100%的生物体转录组转录物进行测序。

[0106] 测序可以包括对所述标记的核酸和/或随机条形码的至少约10、20、30、40、50、60、

70、80、90、100个或更多个核苷酸或碱基对进行测序。测序可以包括对所述标记的核酸和/或随机条形码的至多约10、20、30、40、50、60、70、80、90、100个或更多个核苷酸或碱基对进行测序。测序可以包括对所述标记的核酸和/或随机条形码的至少约200、300、400、500、600、700、800、900、1,000个或更多个核苷酸或碱基对进行测序。测序可以包括对所述标记的核酸和/或随机条形码的至多约200、300、400、500、600、700、800、900、1,000个或更多个核苷酸或碱基对进行测序。测序可以包括对所述标记的核酸和/或随机条形码的至少约1,500;2,000;3,000;4,000;5,000;6,000;7,000;8,000;9,000;或10,000个或更多个核苷酸或碱基对进行测序。测序可以包括对所述标记的核酸和/或随机条形码的至多约1,500;2,000;3,000;4,000;5,000;6,000;7,000;8,000;9,000;或10,000个或更多个核苷酸或碱基对进行测序。

[0107] 测序可以包括至少约200、300、400、500、600、700、800、900、1,000个或更多个测序读数/运行。测序可以包括至多约200、300、400、500、600、700、800、900、1,000个或更多个测序读数/运行。在一些情况下,测序包括至少约1,500;2,000;3,000;4,000;5,000;6,000;7,000;8,000;9,000;或10,000个或更多个测序读数/运行的测序。在一些情况下,测序包括至多约1,500;2,000;3,000;4,000;5,000;6,000;7,000;8,000;9,000;或10,000个或更多个测序读数/运行的测序。测序可以包括至少1000万、5000万、1亿、1.5亿、2亿、2.5亿、3亿、3.5亿、4亿、4.5亿、5亿、5.5亿、6亿、6.5亿、7亿、7.5亿、8亿、8.5亿、9亿、9.5亿、或10亿或更多个测序读数/运行的测序。测序可以包括至多1000万、5000万、1亿、1.5亿、2亿、2.5亿、3亿、3.5亿、4亿、4.5亿、5亿、5.5亿、6亿、6.5亿、7亿、7.5亿、8亿、8.5亿、9亿、9.5亿、或10亿或更多个测序读数/运行的测序。测序可以包括总计为至少1亿、2亿、3亿、4亿、5亿、6亿、7亿、8亿、9亿、10亿、11亿、12亿、13亿、14亿、15亿、16亿、20亿、30亿、40亿、或50亿或更多个测序读数的测序。测序可以包括总计为至多1亿、2亿、3亿、4亿、5亿、6亿、7亿、8亿、9亿、10亿、11亿、12亿、13亿、14亿、15亿、16亿、20亿、30亿、40亿、或50亿或更多个测序读数的测序。测序可以包括小于或等于约1,600,000,000个测序读数/运行。测序可以包括小于或等于约200,000,000个读数/运行。

[0108] 在通过本披露的方法产生的标准化文库中,与未标准化文库相比,可以更容易地鉴定丰度更小(例如,更罕见)的转录物。相比在未标准化文库中的情况,在标准化文库中丰度更小的转录物的测序读数可以包括总读数的更大部分。与未标准化文库中同一转录物的读数相比,标准化文库中丰度较小的转录物的测序读数可包括至少50%、100%、150%、200%、250%、300%、350%、400%、450%或500%或更多的读数。标准化文库中丰度较小的转录物的测序读数可以是未标准化文库中同一转录物的测序读数的至少1、2、3、4、5或6倍或更多倍。

随机条形码

[0109] 本文使用的随机条形码是指可用于随机标记(例如,条形码化、标签化)靶标的多核苷酸序列。随机条形码可以包括一个或多个标记。示例性标记包括但不限于通用标记、细胞标记、分子标记、样品标记、板标记、空间标记和/或前空间标记。随机条形码可以包括可将随机条形码连接至固体支持物的5' 胺。所述随机条形码可以包括一个或多个通用标记、一个或多个维度标记、一个或多个空间标记、一个或多个细胞标记、和/或一个或多个分子标记。所述随机条形码中各个不同标记的位置可以变化。例如,所述通用标记可以是最5' 的

标记。所述分子标记可以是最3'的标记。所述空间标记、维度标记和细胞标记可以处于任何顺序。在一些实施例中,所述通用标记、所述空间标记、所述维度标记、所述细胞标记和所述分子标记处于任何顺序。所述随机条形码可以包括靶结合区。所述靶结合区可以与样品中的靶标(例如,靶核酸、RNA、mRNA、DNA)相互作用。例如,靶结合区可以包括可以与mRNA的聚A尾相互作用的寡聚dT序列。在一些情况下,所述随机条形码的标记(例如,通用标记、维度标记、空间标记、细胞标记和分子标记)可以由1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、或20个或更多个核苷酸分开。

[0110] 随机条形码可以包括一个或多个通用标记。对于所述随机条形码组中的所有随机条形码(例如,附接到给定的固体支持物上的),所述一个或多个通用标记可以是相同的。在一些实施例中,对于附接到多个珠上的所有随机条形码,所述一个或多个通用标记可以是相同的。在一些实施例中,通用标记可以包括能够与测序引物杂交的核酸序列。测序引物可以用于对包括通用标记的随机条形码进行测序。测序引物(例如,通用测序引物)可以包括与高通量测序平台相关联的测序引物。在一些实施例中,通用标记可以包括能够与PCR引物杂交的核酸序列。在一些实施例中,所述通用标记可以包括能够与测序引物和PCR引物杂交的核酸序列。能够与测序或PCR引物杂交的通用标记的核酸序列可以被称为引物结合位点。通用标记可以包括可用于引发随机条形码转录的序列。通用标记可以包括可用于延伸随机条形码或随机条形码内的区域的序列。通用标记的长度可以是至少约1、2、3、4、5、10、15、20、25、30、35、40、45、50个或更多个核苷酸。通用标记可以包括至少约10个核苷酸。通用标记的长度可以是至多约1、2、3、4、5、10、15、20、25、30、35、40、45、50个或更多个核苷酸。在一些实施例中,可切割接头或修饰的核苷酸可以是所述通用标记序列的一部分,以使所述随机条形码能够从所述支持物上被切割下来。如本文使用的,通用标记可以与“通用PCR引物”互换使用。

[0111] 随机条形码可以包括维度标记。维度标记可以包括提供关于随机标记发生的维度的信息的核酸序列。例如,维度标记可以提供关于对靶标进行随机条形码化的时间的信息。维度标记可以与样品中随机条形码化的时间相关联。维度标记可以在随机标记的时间处被激活。不同的维度标记可以在不同的时间被激活。所述维度标记提供关于靶标、靶标组和/或样品被随机条形码化的顺序的信息。例如,在细胞周期的G0期可以对细胞群进行随机条形码化。在细胞周期的G1期,可以用随机条形码对所述细胞再次进行脉冲处理。在细胞周期的S期,可以用随机条形码对所述细胞再次进行脉冲处理,等等。在每个脉冲期(例如,细胞周期的每个时期),随机条形码可以包括不同的维度标记。以这种方式,所述维度标记提供关于哪些靶标在细胞周期的哪个时期被标记的信息。维度标记可以探询许多不同的生物阶段。示例性的生物学时间可以包括但不限于细胞周期、转录(例如,转录起始)和转录物降解。在另一个实例中,样品(例如,细胞、细胞群)可以在用药物和/或疗法治疗之前和/或之后随机标记。不同靶标的拷贝数的变化可以指示样品对药物和/或疗法的反应。

[0112] 维度标记可以是可激活的。可以在特定时间点激活可激活的维度标记。所述可激活的维度标记可以被组成性地激活(例如,不关闭)。所述可激活的维度标记可以被可逆地激活(例如,所述可激活的维度标记可以打开和关闭)。所述维度标记可以被可逆地激活至少1、2、3、4、5、6、7、8、9或10次或更多次。所述维度标记可以被可逆地激活至少1、2、3、4、5、6、7、8、9或10次或更多次。所述维度标记可以用荧光、光、化学事件(例如,切割、连接另一分

子、添加修饰(例如,聚乙二醇化、sumo化、乙酰化、甲基化、脱乙酰化、去甲基化)、光化学事件(例如,光笼蔽)、以及引入非天然核苷酸来激活。

[0113] 所述维度标记对于附接到给定的固体支持物(例如,珠)上的所有随机条形码可以是相同的,但对于不同的固体支持物(例如,珠)是不同的。在一些实施例中,同一固体支持物上的至少60%、70%、80%、85%、90%、95%、97%、99%或100%的随机条形码可以包括相同的维度标记。在一些实施例中,同一固体支持物上的至少60%的随机条形码可以包括相同的维度标记。在一些实施例中,同一固体支持物上的至少95%的随机条形码可以包括相同的维度标记。

[0114] 多个固体支持物(例如,珠)可以表现多达 10^6 个或更多个独特维度标记序列。维度标记的长度可以是至少约1、2、3、4、5、10、15、20、25、30、35、40、45、50个或更多个核苷酸。维度标记的长度可以是至多约300、200、100、90、80、70、60、50、40、30、20、15、12、10、9、8、7、6、5、4个或更少或更多个核苷酸。维度标记可以包括从约5至约200个核苷酸。维度标记可以包括从约10至约150个核苷酸。维度标记的长度可以包括从约20至约125个核苷酸。

[0115] 随机条形码可以包括空间标记。空间标记可以包括提供与随机条形码相关联的靶分子的空间取向的信息的核酸序列。空间标记可以与样品中的坐标相关联。所述坐标可以是固定的坐标。例如可以参考基底固定坐标。空间标记可以参考二维或三维网格。可以参考界标固定坐标。在空间中界标是可被识别的。界标可以是可被成像的结构。界标可以是生物学结构,例如解剖学界标。界标可以是细胞界标,例如细胞器。界标可以是非天然界标,如具有可识别标识符(如色码、条形码、磁性、荧光、放射性或独特尺寸或形状)的结构。空间标记可以与物理分区(例如,孔、容器或液滴)相关联。在一些情况下,将多个空间标记一起用于编码在空间中的一个或多个位置。

[0116] 所述空间标记对于附接到给定的固体支持物(例如,珠)上的所有随机条形码可以是相同的,但对于不同的固体支持物(例如,珠)是不同的。在一些实施例中,同一固体支持物上的至少60%、70%、80%、85%、90%、95%、97%、99%或100%的随机条形码可以包括相同的空间标记。在一些实施例中,同一固体支持物上的至少60%的随机条形码可以包括相同的空间标记。在一些实施例中,同一固体支持物上的至少95%的随机条形码可以包括相同的空间标记。

[0117] 多个固体支持物(例如,珠)可以表现多达 10^6 个或更多个独特空间标记序列。空间标记的长度可以是至少约1、2、3、4、5、10、15、20、25、30、35、40、45、50个或更多个核苷酸。空间标记的长度可以是至多约300、200、100、90、80、70、60、50、40、30、20、15、12、10、9、8、7、6、5、4个或更少或更多个核苷酸。空间标记可以包括从约5至约200个核苷酸。空间标记可以包括从约10至约150个核苷酸。空间标记的长度可以包括从约20至约125个核苷酸。

[0118] 随机条形码可以包括细胞标记(即,样品标记)。如本文使用的,术语“样品标记”和“细胞标记”可以互换地使用。细胞标记可以包括提供用于确定哪个靶核酸来自哪个细胞的信息的核酸序列。在一些实施例中,所述细胞标记对于附接到给定的固体支持物(例如,珠)上的所有随机条形码是相同的,但对于不同的固体支持物(例如,珠)是不同的。在一些实施例中,同一固体支持物上的至少60%、70%、80%、85%、90%、95%、97%、99%或100%的随机条形码可以包括相同的细胞标记。在一些实施例中,同一固体支持物上的至少60%的随机条形码可以包括相同的细胞标记。在一些实施例中,同一固体支持物上的至少95%的随

机条形码可以包括相同的细胞标记。

[0119] 多个固体支持物(例如,珠)可以表现多达 10^6 个或更多个独特细胞标记序列。细胞标记的长度可以是至少约1、2、3、4、5、10、15、20、25、30、35、40、45、50个或更多个核苷酸。细胞标记的长度可以是至多约300、200、100、90、80、70、60、50、40、30、20、15、12、10、9、8、7、6、5、4个或更少或更多个核苷酸。细胞标记可以包括从约5至约200个核苷酸。细胞标记可以包括从约10至约150个核苷酸。细胞标记的长度可以包括从约20至约125个核苷酸。

[0120] 随机条形码可以包括分子标记。分子标记可以包括为与随机条形码杂交的特定类型的靶核酸种类提供识别信息的核酸序列。分子标记可以包括如下核酸序列,所述核酸序列为与随机条形码(例如,靶结合区)杂交的靶核酸种类的特定出现提供计数器。在一些实施例中,将一组不同的分子标记附接到给定的固体支持物(例如,珠)上。在一些实施例中,可以存在多达 10^6 个或更多个附接至给定的固体支持物(例如,珠)的独特分子标记序列。在一些实施例中,可以存在多达 10^5 个或更多个附接至给定的固体支持物(例如,珠)的独特分子标记序列。在一些实施例中,可以存在多达 10^4 个或更多个附接至给定的固体支持物(例如,珠)的独特分子标记序列。在一些实施例中,可以存在多达 10^3 个或更多个附接至给定的固体支持物(例如,珠)的独特分子标记序列。在一些实施例中,可以存在多达 10^2 个或更多个附接至给定的固体支持物(例如,珠)的独特分子标记序列。分子标记的长度可以是至少约1、2、3、4、5、10、15、20、25、30、35、40、45、50个或更多个核苷酸。分子标记的长度可以是至多约300、200、100、90、80、70、60、50、40、30、20、15、12、10、9、8、7、6、5、4个或更少个核苷酸。

[0121] 随机条形码可以包括靶结合区。在一些实施例中,所述靶结合区可包括与靶标(例如,靶核酸、靶分子,例如待分析的细胞核酸)特异性杂交(例如与特定基因序列杂交)的核酸序列。在一些实施例中,靶结合区可以包括可附接(例如,杂交)至特定靶核酸的特定位置的核酸序列。在一些实施例中,所述靶结合区可以包括能够与限制性位点突出端(例如EcoRI粘端突出端)特异性杂交的核酸序列。然后所述随机条形码可以连接到包括与所述限制性位点突出端互补的序列的任何核酸分子。

[0122] 随机条形码可以包括靶结合区。靶结合区可以与感兴趣的靶标杂交。例如,靶结合区可以包括寡聚dT,所述寡聚dT可以与含有聚腺苷酸化末端的mRNA杂交。靶结合区可以是基因特异性的。例如,可以将靶结合区配置为与靶标的特定区域杂交。靶结合区的长度可以是至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、或30个或更多个核苷酸。靶结合区的长度可以是至多1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、或30个或更多个核苷酸。靶结合区的长度可以为从5-30个核苷酸。当随机条形码包括基因特异性靶结合区时,所述随机条形码可以称为基因特异性随机条形码。

[0123] 靶结合区可以包括非特异性靶核酸序列。非特异性靶核酸序列可以指独立于靶核酸的特定序列可与多个靶核酸结合的序列。例如,靶结合区可以包括与mRNA分子上的聚A尾杂交的随机多聚体序列或寡聚dT序列。随机多聚体序列可以是例如随机二聚体、三聚体、四聚体、五聚体、六聚体、七聚体、八聚体、九聚体、十聚体或任何长度的更高多聚体序列。在一些实施例中,对于附接至给定珠的所有随机条形码,所述靶结合区是相同的。在一些实施例中,对于附接至给定珠的所述多个随机条形码,所述靶结合区可以包括两个或更多个不同的靶结合序列。靶结合区的长度可以是至少约5、10、15、20、25、30、35、40、45、50个或更多个

核苷酸。靶结合区的长度可以是至多约5、10、15、20、25、30、35、40、45、50个或更多个核苷酸。

[0124] 随机条形码可以包括定向特性,所述定向特性可以用于定向(例如,比对)所述随机条形码。随机条形码可以包括用于等电聚焦的部分。不同的随机条形码可以包括不同的等电聚焦点。当这些随机条形码被引入样品中时,所述样品可以经历等电聚焦,以便于将所述随机条形码定位成已知的方式。以这种方式,所述定向特性可以用于开发样品中随机条形码的已知的映射。示例性定向特性可以包括电泳迁移率(例如,基于随机条形码的尺寸)、等电点、自旋、电导率和/或自组装。例如,随机条形码可以包括自组装的定向特性,或当激活时可以自组装成特定定向(例如,核酸纳米结构)。

[0125] 随机条形码可以包括亲和力特性。空间标记可以包括亲和力特性。亲和力特性可以被包括在化学和/或生物部分中,所述特性可以促进所述随机条形码与另一种实体(例如,细胞受体)的结合。例如,亲和力特性可以包括抗体。抗体对于样品上的特定部分(例如,受体)可以是特异性的。抗体可以将所述随机条形码引导至特定的细胞类型或分子。可以对在特定的细胞类型或分子处和/或在其附近的靶标进行随机标记。除了空间标记的核苷酸序列,亲和力特性还可以提供空间信息,因为所述抗体可以将所述随机条形码引导至特定位置。抗体可以是治疗性抗体。抗体可以是单克隆抗体。抗体可以是多克隆抗体。抗体可以是人源化的。抗体可以是嵌合的。抗体可以是裸抗体。抗体可以是融合抗体。

[0126] 抗体可以指全长(即,天然存在的或通过正常免疫球蛋白基因片段重组过程形成的)免疫球蛋白分子(例如,IgG抗体)或免疫球蛋白分子的免疫活性(即,特异性结合)部分(像抗体片段)。

[0127] 抗体可以是抗体片段。抗体片段可以是抗体的一部分,如F(ab')₂、Fab'、Fab、Fv、sFv等。抗体片段可以与由全长抗体识别的相同的抗原结合。抗体片段可以包括由抗体的可变区组成的分离的片段,如由重链和轻链的可变区组成的“Fv”片段和其中轻链和重链可变区通过肽接头连接的重组单链多肽分子(“scFv蛋白”)。示例性抗体可以包括但不限于癌细胞抗体、病毒抗体、结合至细胞表面受体(CD8、CD34、CD45)的抗体和治疗性抗体。

[0128] 本披露的细胞标记和/或任何标记还可以包括一组独特的具有限定长度的核酸子序列,例如,每个为7个核苷酸(相当于一些汉明纠错码中使用的位数),其被设计为提供纠错能力。可以设计包括7个核苷酸序列的纠错子序列组,使得所述组中的序列的任何成对组合展现出定义的“遗传距离”(或错配碱基数),例如一组纠错子序列可被设计为展现3个核苷酸的遗传距离。在一些实施例中,用于产生纠错码的核酸子序列的长度可以变化,例如它们的长度可以是至少3个核苷酸、至少7个核苷酸、至少15个核苷酸、或至少31个核苷酸。在一些实施例中,可以使用其他长度的核酸子序列来产生纠错码。

[0129] 本披露的随机条形码在其中可以包括用于纠错的纠错序列(例如,汉明码)。汉明码可以参考基于能够纠正单个位错误的固有冗余鉴别独特的二进制码的运算过程。例如,汉明码可以与核酸条形码匹配,以便筛选在核酸扩增期间发生的单核苷酸错误。通过使用汉明码鉴别单核苷酸错误,从而可以允许核酸条形码的纠正。

[0130] 当随机条形码包括多于一种类型的标记(例如,多于一种细胞标记或多于一种分子标记)时,所述标记可以穿插着接头标记序列。接头标记序列的长度可以是至少约5、10、15、20、25、30、35、40、45、50个或更多个核苷酸。接头标记序列的长度可以是至多约5、10、

15、20、25、30、35、40、45、50个或更多个核苷酸。在一些情况下,接头标记序列的长度是12个核苷酸。可以将接头标记序列用于促进随机条形码的合成。所述接头标记可以包括纠错(例如,汉明)码。

固体支持物

[0131] 本文披露的随机条形码可以附接到固体支持物(例如,珠、基底)上。如本文使用的,术语“拴系”、“附接”和“固定”可互换使用,并且可以指用于将随机条形码附接到固体支持物上的共价或非共价方式。可以将多种不同的固体支持物中的任何一种用作固体支持物,以用于附接预先合成的随机条形码或用于随机条形码的原位固相合成。

[0132] 在一些情况下,固体支持物是珠。珠可以包括任何类型的实心的、多孔的或空心的球体、球、承座、圆柱体或其他类似配置,其由塑料、陶瓷、金属或高分子材料构成,其上可以固定核酸(例如,共价地或非共价地)。珠可以包括可以是球形的(例如,微球)或具有非球形或不规则形状的离散颗粒,所述形状是如立方体、长方形、锥形、圆柱形、圆锥形、椭圆形或圆盘形等。珠的形状可以是非球形的。

[0133] 珠可以包括多种材料,包括但不限于顺磁性材料(例如镁、钼、锂和钽)、超顺磁性材料(例如铁氧体(Fe_3O_4 ;磁铁矿)纳米颗粒)、铁磁材料(例如,铁、镍、钴,其一些合金,以及一些稀土金属化合物)、陶瓷、塑料、玻璃、聚苯乙烯、二氧化硅、甲基苯乙烯、丙烯酸聚合物、钛、胶乳、交联琼脂糖、琼脂糖、水凝胶、聚合物、纤维素、尼龙及其任何组合。

[0134] 所述珠的直径可以变化,例如可以是至少约 $5\mu\text{m}$ 、 $10\mu\text{m}$ 、 $20\mu\text{m}$ 、 $25\mu\text{m}$ 、 $30\mu\text{m}$ 、 $35\mu\text{m}$ 、 $40\mu\text{m}$ 、 $45\mu\text{m}$ 或 $50\mu\text{m}$ 。所述珠的直径可以是至多约 $5\mu\text{m}$ 、 $10\mu\text{m}$ 、 $20\mu\text{m}$ 、 $25\mu\text{m}$ 、 $30\mu\text{m}$ 、 $35\mu\text{m}$ 、 $40\mu\text{m}$ 、 $45\mu\text{m}$ 或 $50\mu\text{m}$ 。所述珠的直径可以与所述基底的孔的直径相关。例如,所述珠的直径可以比所述孔的直径长或短至少10%、20%、30%、40%、50%、60%、70%、80%、90%或100%。所述珠的直径可以比所述孔的直径长或短至多10%、20%、30%、40%、50%、60%、70%、80%、90%或100%。所述珠的直径可以与细胞(例如,由基底的孔截留的单细胞)的直径相关。所述珠的直径可以比所述细胞的直径长或短至少10%、20%、30%、40%、50%、60%、70%、80%、90%、100%、150%、200%、250%、或300%或更多。所述珠的直径可以比所述细胞的直径长或短至多10%、20%、30%、40%、50%、60%、70%、80%、90%、100%、150%、200%、250%、或300%或更多。

[0135] 可以将珠附接至和/或嵌入本披露的基底。可以将珠附接至和/或嵌入凝胶、水凝胶、聚合物和/或基质中。可以使用存在于可用作位置地址的珠上的随机条形码上的空间标记来鉴别基底(例如,凝胶、基质、支架或聚合物)内的珠的空间位置。

[0136] 珠的实例可以包括但不限于链霉亲和素珠、琼脂糖珠、磁珠、Dynabeads®、MACS®微珠、缀合抗体的珠(例如,抗免疫球蛋白微珠)、缀合A蛋白的珠、缀合G蛋白的珠、缀合A/G蛋白的珠、缀合L蛋白的珠、缀合寡聚dT的珠、二氧化硅珠、二氧化硅样珠、抗生物素微珠、抗荧光染料微珠以及BcMag™羧基封端的磁珠。

[0137] 珠可以关联有(例如浸渍有)量子点或荧光染料,以使其在一个荧光光通道或多个光通道中是荧光的。珠可以关联有氧化铁或氧化铬,以使其具有顺磁性或铁磁性。珠可以是可识别的。使用照相机可以将珠成像。珠可以具有与所述珠相关联的可检测的代码。例如,珠可以包括RFID标签。珠可以包括任何可检测的标签(例如,UPC代码、电子条形码、蚀刻标识符)。珠的尺寸可以变化,例如由于在有机或无机溶液中的溶胀。珠可以是疏水的。珠可以

是亲水的。珠可以是生物相容的。

[0138] 可以使固体支持物(例如,珠)可视化。固体支持物可以包括可视化标签(例如,荧光染料)。可以用标识符(例如,数字)将固体支持物(例如,珠)蚀刻。通过对所述固体支持物(例如,珠)成像可以将所述标识符可视化。

[0139] 固体支持物可以指不溶性、半溶性或不溶性材料。当固体支持物包括接头、支架、结构单元或附接至其上的其他反应性部分时,它可以被称为“官能化的”,而当固体支持物缺少附接至其上的这样一个反应性部分时,它可以被称为“非官能化的”。固体支持物可以在溶液中不受约束地利用,如以微量滴定孔形式;以流通形式,如在柱中;或在试纸条(dipstick)中。

[0140] 固体支持物可以包括膜、纸、塑料、涂覆的表面、平表面、玻璃、载玻片、芯片或其任何组合。固体支持物可以采取树脂、凝胶、微球或其他几何构型的形式。固体支持物可以包括二氧化硅芯片;微粒;纳米颗粒;平板;阵列;毛细管;平支持物,如玻璃纤维过滤器,玻璃表面,金属表面(钢、金、银、铝、硅以及铜),玻璃支持物,塑料支持物,硅支持物,芯片,过滤器,膜,微孔板,载玻片;塑料材料包括多孔板或膜(例如,由聚乙烯、聚丙烯、聚酰胺、聚偏二氟乙烯形成);和/或晶片;梳状物(comb);针或针头(例如,适于组合合成或分析的针阵列);或平表面(如晶片(例如,硅晶片)、带有具有或不具有滤底的凹陷的晶片)的凹陷或纳升孔的阵列中的珠。

[0141] 所述固体支持物可以包括聚合物基质(例如,凝胶、水凝胶)。所述聚合物基质可能能够渗透细胞内间隙(例如,细胞器周围)。所述聚合物基质可能能够贯穿循环系统进行泵送。

[0142] 在一些实施例中,固体支持物可以是生物分子。例如,固体支持物可以是核酸、蛋白质、抗体、组蛋白、细胞区室、脂质、碳水化合物等。作为生物分子的固体支持物可被扩增、翻译、转录、降解和/或修饰(例如,聚乙二醇化、sumo化、乙酰化、甲基化)。除了附接至生物分子的空间标记之外,作为生物分子的固体支持物可以提供空间和时间信息。例如,生物分子可以在未修饰时包括第一构象,但是在修饰时可以改变为第二构象。所述不同的构象可以将本披露的随机条形码暴露给靶标。例如,生物分子可以包括由于生物分子折叠而不可接受的随机条形码。在修饰生物分子(例如,乙酰化)时,所述生物分子可以改变构象以暴露所述随机标记。修饰的时间设置可以为本披露的随机条形码化的方法提供另一时间维度。

[0143] 在另一个实例中,包括本披露的随机条形码的生物分子可以位于细胞的细胞质中。激活时,所述生物分子可以移动到细胞核,于此可以进行随机条形码化。以这种方式,所述生物分子的修饰可以编码由所述随机条形码鉴别的靶标的另外的空间-时间信息。

[0144] 维度标记可以提供关于生物事件(例如,细胞分裂)的空间-时间的信息。例如,可以将维度标记添加到第一细胞,所述第一细胞可以分裂产生第二子细胞,所述第二子细胞可以包括所述维度标记的全部、一些或不包括所述维度标记。所述维度标记可以在原始细胞和子细胞中激活。以这种方式,所述维度标记可以提供关于随机条形码在不同空间中的时间的信息。

样品

[0145] 如本文所述,所述多个核酸分子可以从样品例如细胞样品获得或衍生。用于在本披露的方法中使用的样品可以包括一个或多个细胞。样品可以指一个或多个细胞。在一些

实施例中,所述细胞是从癌组织切除的癌细胞,例如乳腺癌、肺癌、结肠癌、前列腺癌、卵巢癌、胰腺癌、脑癌、黑素瘤和非黑素瘤皮肤癌等。在一些情况下,所述细胞源自癌症,但是从体液收集(例如,循环肿瘤细胞)。癌症的非限制性实例可以包括腺瘤、腺癌、鳞状细胞癌、基底细胞癌、小细胞癌、大细胞未分化癌、软骨肉瘤以及纤维肉瘤。

[0146] 在一些实施例中,所述细胞是已经被病毒感染并含有病毒寡核苷酸的细胞。在一些实施例中,所述病毒感染可以由选自下组的病毒引起,该组由以下各项组成:双链DNA病毒(例如腺病毒、疱疹病毒、痘病毒)、单链(+链或“正义”)DNA病毒(例如细小病毒)、双链RNA病毒(例如呼肠孤病毒)、单链(+链或正义)RNA病毒(例如微小核糖核酸病毒、披膜病毒)、单链(-链或反义)RNA病毒(例如正粘病毒、弹状病毒)、单链(+链或正义)RNA病毒,在其生命周期中具有DNA中间体)RNA-RT病毒(例如逆转录病毒)和双链DNA-RT病毒(例如嗜肝DNA病毒)。示例性病毒可以包括但不限于SARS、HIV、冠状病毒、埃博拉病毒、疟疾、登革热、丙型肝炎、乙型肝炎和流感。

[0147] 在一些实施例中,所述细胞是细菌。这些可以包括革兰氏阳性细菌或革兰氏阴性细菌。可以使用所披露的方法、装置和系统分析的细菌的实例包括但不限于马杜拉放线菌属(*Actinomedurae*)、衣氏放线菌、炭疽杆菌、蜡样芽孢杆菌、肉毒梭菌、艰难梭菌、产气荚膜梭菌、破伤风梭菌、棒状杆菌属、粪肠球菌、单核细胞增生李斯特菌、诺卡氏菌属、痤疮丙酸杆菌、金黄色葡萄球菌、表皮葡萄球菌(*Staphylococcus epiderm*)、变异链球菌、肺炎链球菌等。革兰氏阴性细菌包括但不限于猫阿菲波菌(*Afipia felis*)、类杆菌(*Bacteriodes*)、杆状巴尔通体、百日咳杆菌(*Bordetella pertussis*)、伯氏疏螺旋体、回归热疏螺旋体、布鲁氏菌属、肉芽肿鞘杆菌、弯曲杆菌属、大肠杆菌、土拉热弗朗西丝菌、阴道加德纳菌、埃及嗜血杆菌(*Haemophilus aegyptius*)、杜克雷嗜血杆菌(*Haemophilus ducreyi*)、流感嗜血杆菌(*Haemophilus influenzae*)、幽门螺杆菌、嗜肺军团菌、问号钩端螺旋体、脑膜炎奈瑟菌、牙龈卟啉单胞菌、斯氏普罗维登斯菌(*Providencia sturti*)、铜绿假单胞菌、肠炎沙门菌(*Salmonella enteridis*)、伤寒沙门菌、粘质沙雷菌、鲍氏志贺氏菌、念珠状链杆菌、化脓链球菌、梅毒螺旋体、霍乱弧菌、小肠结肠炎耶尔森菌、鼠疫耶尔森菌等。其他细菌可以包括鸟分枝杆菌(*Myobacterium avium*)、麻风分枝杆菌(*Myobacterium leprae*)、结核分枝杆菌(*Myobacterium tuberculosis*)、汉式巴尔通体(*Bartonella henselae*)、鸚鵡热衣原体、沙眼衣原体、伯氏考克斯体、肺炎支原体、螨立克次体、普氏立克次体、立氏立克次体、姜虫病立克次体、斑疹伤寒立克次体、尿素分解脲原体、肺炎双球菌、查菲埃立克体(*Ehrlichia chafensis*)、尿肠球菌、脑膜炎球菌等。

[0148] 在一些实施例中,所述细胞是真菌。可以使用所披露的方法、装置和系统分析的真菌的非限制性实例包括但不限于曲霉属、假丝酵母属(*Candidae*)、白色念珠菌、粗球孢子菌、隐球菌属(*Cryptococci*)及其组合。

[0149] 在一些实施例中,所述细胞是原生动物或其他寄生虫。待使用本披露的方法、装置和系统分析的寄生虫的实例包括但不限于结肠小袋纤毛虫、微小隐孢子虫、卡晏环孢子虫(*Cyclospora cayatanensis*)、脑炎微孢子虫属(*Encephalitozoa*)、溶组织内阿米巴(*Entamoeba histolytica*)、比氏肠孢虫(*Enterocytozoon bieneusi*)、兰伯贾第虫、利什曼原虫(*Leishmaniae*)、疟原虫(*Plasmodii*)、鼠弓形体、锥体虫(*Trypanosomae*)、梯形阿米巴(*trapezoidal amoeba*)、蠕虫(worm)(例如,蠕虫(helminthes)),尤其是寄生蠕虫,包括但

不限于线虫纲(蛔虫(roundworm),例如鞭虫、钩虫、蛲虫、蛔虫(ascarid)、丝状虫(filarid)等)、绦虫纲(例如,绦虫)。

[0150] 如本文使用的,术语“细胞”可以指一个或多个细胞。在一些实施例中,所述细胞是正常细胞,例如不同发育阶段的人类细胞或来自不同器官或组织类型的人类细胞(例如白细胞,红细胞,血小板,上皮细胞,内皮细胞,神经元,神经胶质细胞,成纤维细胞,骨骼肌细胞,平滑肌细胞,配子或来自心脏、肺、脑、肝、肾、脾、胰腺、胸腺、膀胱、胃、结肠、小肠的细胞)。在一些实施例中,所述细胞可以是未分化的人类干细胞或已被诱导分化的人类干细胞。在一些实施例中,所述细胞可以是人类胎儿细胞。人类胎儿细胞可以获得自怀有胎儿的母亲。在一些实施例中,所述细胞是罕见细胞。罕见细胞可以是例如循环肿瘤细胞(CTC)、循环上皮细胞、循环内皮细胞、循环子宫内膜细胞、循环干细胞、干细胞、未分化干细胞、癌干细胞、骨髓细胞、祖细胞、泡沫细胞、间充质细胞、滋养层、免疫系统细胞(宿主或移植体)、细胞片段、细胞器(例如线粒体或细胞核)、病原体感染的细胞等。

[0151] 在一些实施例中,所述细胞是非人类细胞,例如其他类型的哺乳动物细胞(例如小鼠、大鼠、猪、狗、牛或马)。在一些实施例中,所述细胞是其他类型的动物或植物细胞。在其他实施例中,所述细胞可以是任何原核或真核细胞。

[0152] 在一些实施例中,第一细胞样品获得自未患疾病或病症的人,并且第二细胞样品获得自患有所述疾病或病症的人。在一些实施例中,所述人是不同的。在一些实施例中,所述人是相同的,但是细胞样品是在不同时间点取得。在一些实施例中,所述人是患者,并且所述细胞样品是患者样品。所述疾病或病症可以是癌症、细菌感染、病毒感染、炎症性疾病、神经退行性疾病、真菌性疾病、寄生虫病、遗传性障碍或其任何组合。

[0153] 在一些实施例中,适用于当前披露方法的细胞尺寸可以为直径在从约2微米至约100微米的范围中。在一些实施例中,所述细胞可以具有至少2微米、至少5微米、至少10微米、至少15微米、至少20微米、至少30微米、至少40微米、至少50微米、至少60微米、至少70微米、至少80微米、至少90微米、或至少100微米的直径。在一些实施例中,所述细胞可以具有至多100微米、至多90微米、至多80微米、至多70微米、至多60微米、至多50微米、至多40微米、至多30微米的直径、至多20微米、至多15微米、至多10微米、至多5微米、或至多2微米的直径。所述细胞可以具有在一定范围内(例如从约5微米至约85微米)的任何值的直径。在一些实施例中,所述细胞具有约10微米的直径。

[0154] 在一些实施例中,在将细胞与珠关联之前和/或在微孔中对所述细胞进行分选。例如,可以通过荧光激活细胞分选或磁激活细胞分选,或例如通过流式细胞术来分选所述细胞。所述细胞可以按尺寸进行过滤。在一些情况下,滞留物含有与所述珠相关联的细胞。在一些情况下,流过物(flow through)含有与所述珠相关联的细胞。

[0155] 在一些实施例中,所述样品包括免疫细胞。免疫细胞可以包括例如T细胞、B细胞、淋巴干细胞、骨髓祖细胞、淋巴细胞、粒细胞、B细胞祖细胞、T细胞祖细胞、自然杀伤细胞、Tc细胞、Th细胞、浆细胞、记忆细胞、嗜中性粒细胞、嗜酸性粒细胞、嗜碱性粒细胞、肥大细胞、单核细胞、树突细胞和/或巨噬细胞或其任何组合。

[0156] T细胞可以是T细胞克隆,其可以指源自单一T细胞的T细胞或具有相同TCR的那些。T细胞可以是可包括T细胞克隆和具有不同TCR的混合T细胞群体的T细胞系的一部分,所述TCR全部都可以识别相同的靶标(例如,抗原、肿瘤、病毒)。T细胞可以从许多来源获得,包括

外周血单核细胞、骨髓、淋巴结组织、脾组织和肿瘤。可以从收集自受试者的血液单位获得T细胞,如使用Ficoll分离。来自个体的循环血液的细胞可以通过单采术或白细胞分离术获得。单采产物可以包括淋巴细胞,包括T细胞、单核细胞、粒细胞、B细胞、其他有核白细胞、红细胞和血小板。可以将所述细胞洗涤并重悬浮于培养基中以分离感兴趣的细胞。

[0157] 可以通过裂解红细胞并消耗单核细胞而从外周血淋巴细胞中分离T细胞,例如通过PERCOLL™梯度离心。可以通过阳性或阴性选择技术进一步分离T细胞的特定子群,如CD28⁺、CD4⁺、CD45RA⁺和CD45RO⁺T细胞。例如,可以通过用缀合抗CD3/抗CD28(即,3×28)的珠(如DYNABEADS®M-450CD3/CD28T或XCYTE DYNABEADS™)孵育足以阳性选择所希望的T细胞的一段时间来分离T细胞。免疫细胞(例如,T细胞和B细胞)可以是抗原特异性的(例如,对于肿瘤是特异性的)。

[0158] 在一些实施例中,所述细胞可以是抗原呈递细胞(APC),如B细胞、来自淋巴结的活化B细胞、淋巴母细胞样细胞、静息B细胞或瘤性B细胞(例如,来自淋巴瘤)。APC可以指B细胞或滤泡树突细胞,在其表面上表达至少一种BCRC蛋白。

随机条形码化和文库标准化方法

[0159] 本披露提供了用于样品的文库标准化的方法。文库标准化方法可以与任何文库样品制备方法组合。文库标准化方法可以与本披露的随机条形码化方法组合。随机条形码化可以用于以独特的条形码对单独的核酸分子(例如,mRNA分子)进行索引,使得可以对特定靶标的分子进行追踪和/或计数。

[0160] 随机条形码化方法可以包括将所述随机条形码靠近所述样品放置,裂解所述样品,将不同靶标与所述随机条形码相关联,对所述靶标进行扩增和/或对所述靶标进行数字计数。图2展示了本披露的随机条形码化方法的示例性实施例。可以使样品(例如,样品、薄片和/或细胞的部分)与含有随机条形码的固体支持物接触。所述样品中的靶标可以与所述随机条形码相关联。可以收集所述固体支持物。可以在所述固体支持物上进行cDNA合成。cDNA合成可以在所述固体支持物之外进行。cDNA合成可以将来自所述随机条形码中的标记的标记信息并入正在合成的新cDNA靶分子中,从而产生靶标-条形码分子。可以使用PCR扩增所述靶标-条形码分子。在所述靶标-条形码分子上的靶标和随机条形码的标记的序列可以通过测序方法确定。

接触样品和随机条形码

[0161] 可以使包括例如细胞、器官或组织薄片的样品与随机条形码接触。所述固体支持物可以自由浮动。所述固体支持物可嵌入半固体或固体阵列中。所述随机条形码可以不与固体支持物关联。所述随机条形码可以是单独核苷酸。所述随机条形码可以与基底相关联。当随机条形码靠近靶标时,所述靶标可以与所述随机条形码进行杂交。可以使所述随机条形码以不可消耗的比率接触,使得每个不同的靶标可以与本披露的不同随机条形码相关联。为了确保靶标与随机条形码之间的有效关联,可以将所述靶标与所述随机条形码交联。

[0162] 样品的两个不同靶标可以接触相同的独特随机条形码的概率可以是至少 10^{-6} 、 10^{-5} 、 10^{-4} 、 10^{-3} 、 10^{-2} 、或 10^{-1} 或更多。样品的两个不同靶标可以接触相同的独特随机条形码的概率可以是至多 10^{-6} 、 10^{-5} 、 10^{-4} 、 10^{-3} 、 10^{-2} 、或 10^{-1} 或更多。来自相同细胞的相同基因的两个靶标可以接触相同的随机条形码的概率可以是至少 10^{-6} 、 10^{-5} 、 10^{-4} 、 10^{-3} 、 10^{-2} 、或 10^{-1} 或更多。来自相同细胞的相同基因的两个靶标可以接触相同的随机条形码的概率可以是至多

10^{-6} 、 10^{-5} 、 10^{-4} 、 10^{-3} 、 10^{-2} 、或 10^{-1} 或更多。

[0163] 在一些情况下,来自细胞群的细胞可以分离(例如,隔离)到本披露的基底的孔中。所述细胞群可以在分离前稀释。可以将所述细胞群进行稀释,使得至少1%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、或100%的基底的孔接收单细胞。可以将所述细胞群进行稀释,使得至多1%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、或100%的基底的孔接收单细胞。可以将所述细胞群进行稀释,使得在经稀释的细胞群中的细胞数目是至少%、1%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、或100%在基底上的孔的数目。可以将所述细胞群进行稀释,使得在经稀释的细胞群中的细胞数目是至少%、1%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、或100%在基底上的孔的数目。在一些情况下,将所述细胞群进行稀释,使得细胞的数目是基底中孔的数目的约10%。

[0164] 将单细胞分布到基底的孔中可以遵循泊松分布。例如,可以至少有0.1%、0.5%、1%、2%、3%、4%、5%、6%、7%、8%、9%或10%或更多的概率,使得所述基底的一个孔具有多于一个细胞。可以至少有0.1%、0.5%、1%、2%、3%、4%、5%、6%、7%、8%、9%或10%或更多的概率,使得所述基底的一个孔具有多于一个细胞。将单细胞分布到基底的孔中可以是随机的。将单细胞分布到基底的孔中可以是非随机的。可以将所述细胞进行分离,使得所述基底的孔仅接收一个细胞。

细胞裂解

[0165] 在细胞和随机条形码的分布之后,可以裂解所述细胞以释放所述靶分子。可以通过各种手段,例如通过化学的或生化手段、通过渗透冲击、或通过热裂解手段、机械裂解或光学裂解来完成细胞裂解。可以通过添加包含洗涤剂(例如SDS、十二烷基硫酸锂、Triton X-100、Tween-20或NP-40)的细胞裂解缓冲液、有机溶剂(例如甲醇或丙酮)或消化酶(例如蛋白酶K、胃蛋白酶或胰蛋白酶)或其任何组合来裂解细胞。为了增加靶标与随机条形码的关联性,可以通过例如降低温度和/或增加裂解物的粘度来改变所述靶分子的扩散速率。

将随机条形码附接至靶核酸分子

[0166] 在细胞裂解和核酸分子从中释放之后,所述核酸分子可以随机地与共定位的固体支持物的随机条形码相关联。关联可以包括随机条形码的靶识别区与所述靶核酸分子的互补部分的杂交(例如,所述随机条形码的寡聚dT可与靶标的聚A尾相互作用)。可以选择用于杂交的测定条件(例如缓冲液pH、离子强度、温度等)以促进形成特定的稳定的杂交体。

[0167] 附接可以还包括随机条形码的靶识别区和所述靶核酸分子的一部分的连接。例如,所述靶结合区可以包括能够与限制性位点突出端(例如EcoRI粘端突出端)特异性杂交的核酸序列。所述测定程序还可以包括用限制性酶(例如EcoRI)处理所述靶核酸以产生限制性位点突出端。然后所述随机条形码可以连接到包括与所述限制性位点突出端互补的序列的任何核酸分子。连接酶(例如,T4DNA连接酶)可以用于连接两个片段。

[0168] 可以随后合并来自多个细胞(或多个样品)(例如,靶标-条形码分子)的标记的靶标,例如通过检索所述随机条形码和/附接有所述靶标-条形码分子的珠。基于固体支持物的附接的靶标-条形码分子的集合的收回可以通过使用磁珠和外部施加的磁场来实现。一

且所述靶标-条形码分子已经合并,所有进一步的处理可以在单个反应容器中进行。进一步的处理可以包括例如逆转录反应、扩增反应、切割反应、解离反应和/或核酸延伸反应。进一步的处理反应可以在微孔内进行,即,没有来自多个细胞的标记的靶核酸分子的第一合并。

逆转录

[0169] 本披露提供了使用逆转录来产生随机的靶标-条形码缀合物的方法。所述随机的靶标-条形码缀合物可以包括所述随机条形码以及所述靶核酸(即,随机条形码化的cDNA分子)的全部或部分的互补性序列。相关联的RNA分子的逆转录可以通过添加逆转录引物连同逆转录酶而进行。所述逆转录引物可以是寡聚dT引物、随机六核苷酸引物、或靶标特异性寡核苷酸引物。寡聚dT引物的长度可以是例如12-18个核苷酸,并与哺乳动物mRNA的3'端的内源性聚A尾结合。随机六核苷酸引物可在多个互补位点结合至mRNA。靶标特异性寡核苷酸引物通常选择性地引发感兴趣的mRNA。

[0170] 可以使所述随机条形码化的cDNA分子经历下游方法,如扩增(例如通过通用和/或基因特异性引物)和本披露的文库标准化方法。

试剂盒

[0171] 本文披露了用于进行本披露的文库标准化方法的试剂盒。试剂盒可以包括含结合部分的第二链合成引物。试剂盒可以包括固体支持物,所述固体支持物包括可结合到第二链合成引物上的结合部分的捕获部分。试剂盒可以包括用来捕获固体支持物的磁体。试剂盒可以包括用于清理扩增反应的试剂(例如,AmpureXP珠和/或纯化旋转柱)。试剂盒可以包括含测序流动池序列的衔接子和/或引物。所述试剂盒还可以包括用于进行核酸延伸反应(例如,逆转录反应和引物延伸反应)的试剂(例如,酶、引物、dNTP、NTP、RNA酶抑制剂、或缓冲液)。所述试剂盒还可以包括用于进行扩增反应以制备测序文库的试剂(例如,酶、通用引物、测序引物、靶标特异性引物、或缓冲液)。

[0172] 本文披露了用于进行随机条形码化测定的试剂盒。所述试剂盒可以包括一种或多种固体支持物悬浮液,其中在悬浮液中的所述单独固体支持物包括本披露的多个附接的随机条形码。所述试剂盒可以包括可能不附接至固体支持物的随机条形码。所述试剂盒还可以包括用于进行随机条形码化测定的试剂,例如裂解缓冲液、漂洗缓冲液或杂交缓冲液。所述试剂盒还可以包括用于进行核酸延伸反应(例如,逆转录反应和引物延伸反应)的试剂(例如,酶、引物、dNTP、NTP、RNA酶抑制剂、或缓冲液)。所述试剂盒还可以包括用于进行扩增反应以制备测序文库的试剂(例如,酶、通用引物、测序引物、靶标特异性引物、或缓冲液)。

[0173] 本披露的试剂盒通常可以包括用于实施本文所述的一种或多种方法的说明书。所述试剂盒中包含的说明书可以贴在包装材料上,或者可以作为包装插入物而被包含。虽然说明书通常是书面材料或印刷材料,但它们不限于此。任何能够存储此类说明书并将其传送给最终使用者的介质都被本披露所涵盖。这样的介质可以包括但不限于电子存储介质(例如,磁盘、磁带、柱体、芯片)、光学介质(例如,CD ROM)、RF标签等。如本文所用的,术语“说明书”可以包括提供指令的因特网站点的地址。

[0174] 虽然已经在本文示出并描述了本发明的优选实施例,但是对本领域的普通技术人员而言应该显而易见的是这样的实施例仅以举例方式提供。在不背离本发明的情况下众多变化、改变和替代现在将被本领域的普通技术人员想到。应该理解的是,本文所述的本发明的实施例的不同替代方案可以用于实践本发明。预期的是以下权利要求书限定了本发明的

范围以及由此覆盖在这些权利要求和它们的等效物的范围内的方法和结构。

实例

实例1:用巢式PCR进行文库标准化

[0175] 本实例提供了用巢式PCR进行文库标准化的方法。使用包括寡聚dT序列、分子标记、样品标记和通用标记的引物,将多个mRNA逆转录成多个cDNA。在第一扩增反应中使用基因特异性反向引物和通用引物(例如,结合至通用标记的通用引物)扩增cDNA,从而产生第一组扩增子。通用引物可以包括生物素部分。在第二扩增反应中,可以使用第二基因特异性巢式PCR引物和包括生物素部分的通用引物来扩增第一组扩增子。此反应产生一端包括生物素部分的不对称标记的扩增子。将文库热变性。将文库冷却以诱导部分再退火。在部分再退火期间,高丰度扩增子将比低丰度扩增子更快地再退火。

[0176] 可以使部分再退火的文库与包括链霉亲和素部分的固体支持物接触。链霉亲和素可以结合至文库的链上的生物素部分。再退火的扩增子将包括生物素,并将被链霉亲和素除去。包括生物素的未经再退火的链也将被链霉亲和素除去。剩余的链代表丰度较低的序列,并且是具有生物素的链的互补体。这些链代表标准化文库。

[0177] 用PCR引物使文库再生。PCR引物可以包括测序流动池引物序列。对标准化文库进行测序。

实例2:用全转录组扩增进行文库标准化

[0178] 可以对由全转录组扩增产生的文库进行文库标准化。可以使用衔接子连接方法进行全转录组扩增。靶标包括聚A尾。靶标是mRNA。将靶标与随机条形码杂交。随机条形码包括数个标记。例如,随机条形码包括靶标特异性区域(例如,用于结合至mRNA的聚A尾的寡聚dT)、分子标记、细胞标记和第一通用标记。随机条形码是使用逆转录酶逆转录的,从而产生标记的cDNA分子。多余的随机条形码用降解酶处理。降解酶是外切核酸酶。

[0179] 标记的cDNA分子经历第二链合成,从而产生双链标记的cDNA分子。第二链合成是通过以下方式进行:使标记的cDNA分子-mRNA杂交体与切口酶(例如,RNA酶)接触,所述切口酶可以将和标记的cDNA分子杂交的mRNA切割从而产生切割的mRNA。切割的mRNA用作引物并使用聚合酶(例如,DNA Pol I)延伸,从而并入第一链的序列。聚合酶包括5'-3'外切核酸酶活性。聚合酶降解充当第二链合成的引物的下游mRNA切口。使用连接酶将延伸的序列连接在一起,从而产生第二链(例如,双链标记的cDNA分子)。

[0180] 双链标记的cDNA分子包括与第一通用标记互补的序列。使双链标记的cDNA分子与衔接子接触。衔接子是双链的。衔接子包括限制性内切核酸酶切割位点。衔接子包括第二通用引物序列(其与第一通用引物序列相同)。衔接子包括3'突出端。衔接子包括可连接到双链标记的cDNA分子的3'羟基的游离5'磷酸(P)。衔接子连接到双链标记的cDNA分子1558的两条链上。

[0181] 使用一种或多种WTA扩增引物扩增产物。WTA扩增引物之一包括生物素部分。WTA扩增引物之一不包括生物素部分。扩增产物,使得一条链线性扩增,而一条链指数扩增。线性扩增的链在一端包括可扩增的通用序列。可以指数方式扩增的链在两端包括通用序列,其中一个末端包括生物素部分。使WTA扩增的产物经历如实例1中所述的文库标准化方案。

实例3:阻滞剂在文库标准化中的使用

[0182] 本实例提供了用阻滞剂进行文库标准化的方法。使用包括寡聚dT序列、分子标记、

样品标记和通用标记的引物,将多个mRNA逆转录成多个cDNA。在第一扩增反应中使用基因特异性反向引物和通用引物(例如,结合至通用标记的通用引物)扩增cDNA,从而产生第一组扩增子。通用引物可以包括生物素部分。在第二扩增反应中,使用第二基因特异性巢式PCR引物和包括生物素部分的通用引物来扩增第一组扩增子。此反应产生一端包括生物素部分的不对称标记的扩增子。将阻滞剂添加到文库中。将文库热变性。阻滞剂与扩增子的通用标记结合。将文库冷却以诱导部分再退火。在部分再退火期间,高丰度扩增子将比低丰度扩增子更快地再退火。与由引物中的任何序列(例如,分子标记、样品标记、通用标记)驱动的相比,部分再退火将由靶序列更多地驱动。

[0183] 使部分再退火的文库与包括链霉亲和素部分的固体支持物接触。链霉亲和素可以结合至文库的链上的生物素部分。再退火的扩增子将包括生物素,并将被链霉亲和素除去。包括生物素的未经再退火的链也将被链霉亲和素除去。剩余的链代表丰度较低的序列,并且是具有生物素的链的互补体。这些链代表标准化文库。

[0184] 用PCR引物使文库再生。PCR引物可以包括测序流动池引物序列。对标准化文库进行测序。

实例4:在固体支持物上进行文库标准化

[0185] 本实例提供了用固体支持物进行文库标准化的方法。使用包括寡聚dT序列、分子标记、样品标记和通用标记的引物,将多个mRNA逆转录成多个cDNA。在第一扩增反应中使用基因特异性反向引物和通用引物(例如,结合至通用标记的通用引物)扩增cDNA,从而产生第一组扩增子。通用引物可以包括叠氮化物或炔部分。在第二扩增反应中,可以使用第二基因特异性巢式PCR引物和包括叠氮化物或炔部分的通用引物来扩增第一组扩增子。此反应产生一端包括叠氮化物或炔部分的不对称标记的扩增子。

[0186] 使用点击化学将文库附接到固体支持物上。固体支持物包括参与点击化学的互补分子。例如,如果扩增子包括炔,则固体支持物包括叠氮化物。将文库热变性。在一些情况下,引入阻滞剂。将文库冷却以诱导部分再退火。在部分再退火期间,高丰度扩增子将比低丰度扩增子更快地再退火。

[0187] 通过点击化学将再退火的扩增子附接到固体支持物上。除去固体支持物(例如,通过离心或磁力)。剩余的链(例如未经再退火的链)将不包括叠氮化物或炔部分。剩余的链代表丰度较低的序列,并且是具有叠氮化物或炔的链的互补体。这些链代表标准化文库。

[0188] 用PCR引物使文库再生。PCR引物可以包括测序流动池引物序列。对标准化文库进行测序。

[0189] 在至少一些先前描述的实施例中,在一个实施例中使用的一个或多个元件可以互换地用于另一个实施例中,除非这种替换在技术上不可行。本领域技术人员将理解,在不脱离所要求保护的题目的范围的情况下,可以对上述方法和结构进行各种其他的省略、添加和修改。所有此类修改和改变都旨在落在由所附权利要求书限定的题目的范围内。

[0190] 关于本文中基本上任何复数和/或单数术语,在对于背景和/或应用适当的情况下,本领域技术人员可以从复数转换为单数和/或从单数转换为复数。为了清楚起见,可以在本文明确阐述各种单数/复数排列。

[0191] 本领域技术人员将理解,一般来说,本文使用的术语,尤其是所附权利要求书(例如,所附权利要求书的主体)中的术语,通常旨在作为“开放性的”术语(例如,术语“包括

(including)”应解释为“包括但不限于(including but not limited to)”,术语“具有(having)”应解释为“具有至少(having at least)”,术语“包括(includes)”应解释为“包括但不限于(includes but is not limited to)”等。本领域技术人员将进一步理解,如果预期到所介绍的权利要求陈述的特定数目,这样的预期将明确地陈述于权利要求中,并且在不存在这种陈述的情况下没有这种意图存在。例如,作为对理解的帮助,以下所附权利要求书可以包含介绍性短语“至少一个”和“一个或多个”的使用,以介绍权利要求陈述。然而,此类短语的使用不应解读为意味着由不定冠词“一个”或“一种”介绍权利要求陈述会将任何包含这种介绍的权利要求陈述的具体权利要求限制到包含仅一个这种陈述的实施例中,甚至当相同的权利要求包括介绍性短语“一个或多个”或“至少一个”以及不定冠词如“一个”或“一种”时也是如此(例如,“一个”和/或“一种”应解释为意指“至少一个”或“一个或多个”);这对于使用定冠词来介绍权利要求陈述同样适用。此外,即使明确地陈述了介绍的权利要求陈述的特定数目,本领域技术人员将认识到,这种陈述应解释为意指至少所陈述的数字(例如,仅陈述“两个陈述”而没有其他修饰词意指至少两个陈述、或两个或更多个陈述)。此外,在使用类似于“A、B和C等中的至少一个”的惯例的那些情况下,通常这种句法结构是在本领域技术人员将理解所述惯例的意义上预期(例如,“具有A、B和C中的至少一个的系统”将包括但不限于仅具有A,仅具有B,仅具有C,A和B一起,A和C一起,B和C一起,和/或A、B、和C一起等的系统)。在使用类似于“A、B或C等中的至少一个”的惯例的那些情况下,通常这种句法结构是在本领域技术人员将理解所述惯例的意义上预期(例如,“具有A、B或C中的至少一个的系统”将包括但不限于仅具有A,仅具有B,仅具有C,A和B一起,A和C一起,B和C一起,和/或A、B、和C一起等的系统)。本领域技术人员将进一步理解,实际上,无论在说明书、权利要求书还是在附图中,呈现两个或更多个替代术语的任何分离性词语和/或短语应被理解为考虑到包括术语之一、任一术语或两个术语的可能性。例如,短语“A或B”应理解为包括“A”或“B”或“A和B”的可能性。

[0192] 此外,当本披露的特征或方面以马库什组(Markush group)描述时,本领域技术人员将意识到本披露还由此以马库什组的任何单独的成员或成员子组描述。

[0193] 如本领域技术人员将理解的,出于任何和所有目的,如在提供书面描述方面,本文披露的所有范围还包括任何和所有可能的它的子范围和子范围组合。任何列出的范围都可以容易地被识别为充分地描述并使同一范围能被分解成至少相等的两等份、三等份、四等份、五等份、十等份等。作为非限制性实例,本文讨论的每个范围都可以容易地分解成下三分之一、中三分之一和上三分之一等。如本领域技术人员还将理解的,所有语言,如“多至”、“至少”、“大于”、“小于”等包括所陈述的数字,并且指代可以随后分解为如上讨论的子范围的范围。最后,如本领域技术人员将理解的,范围包括每个单独的成员。因此,例如,具有1-3个物品的组是指具有1、2或3个物品的组。类似地,具有1-5个物品的组指代具有1、2、3、4或5个物品的组,等等。

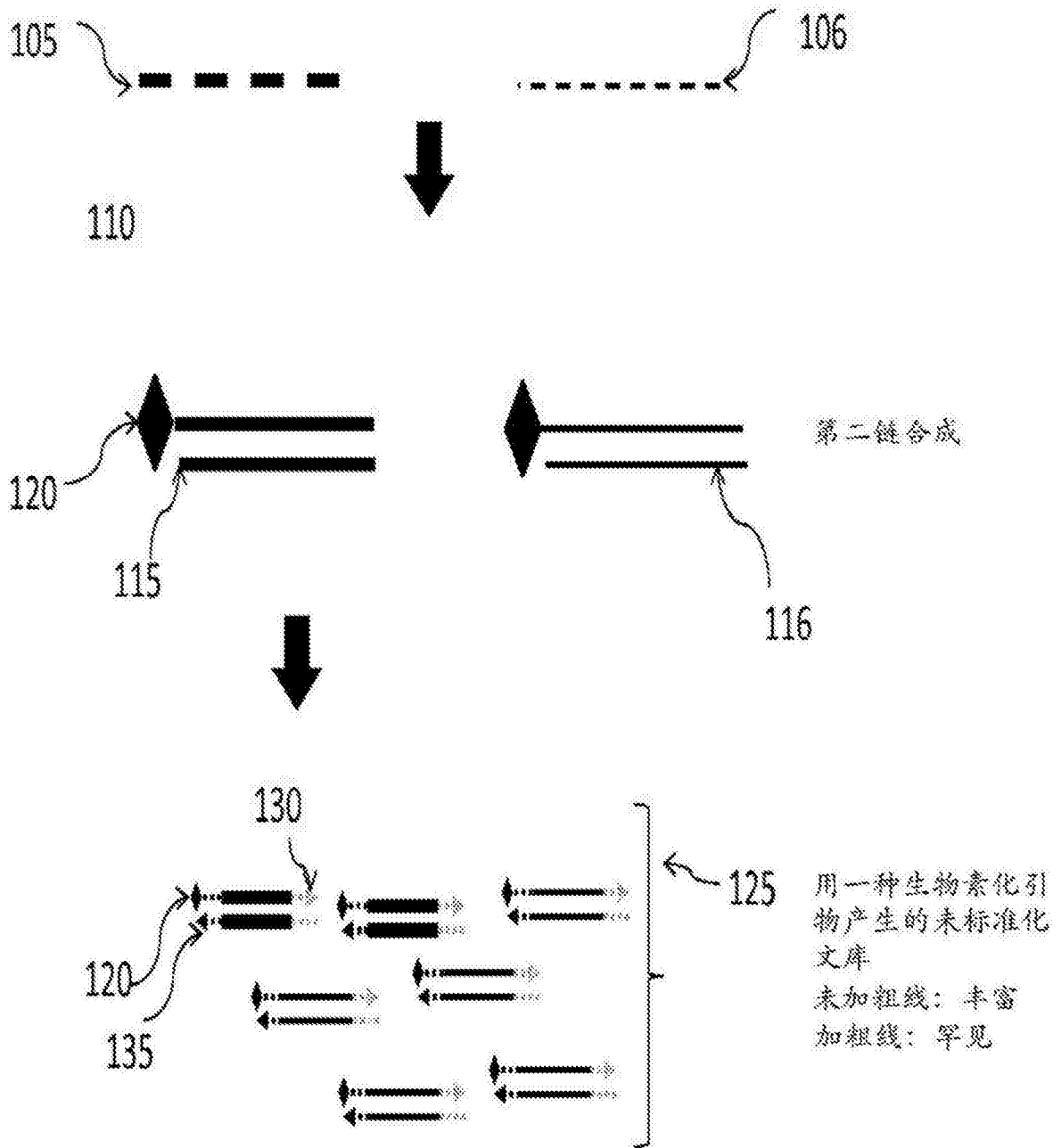


图1

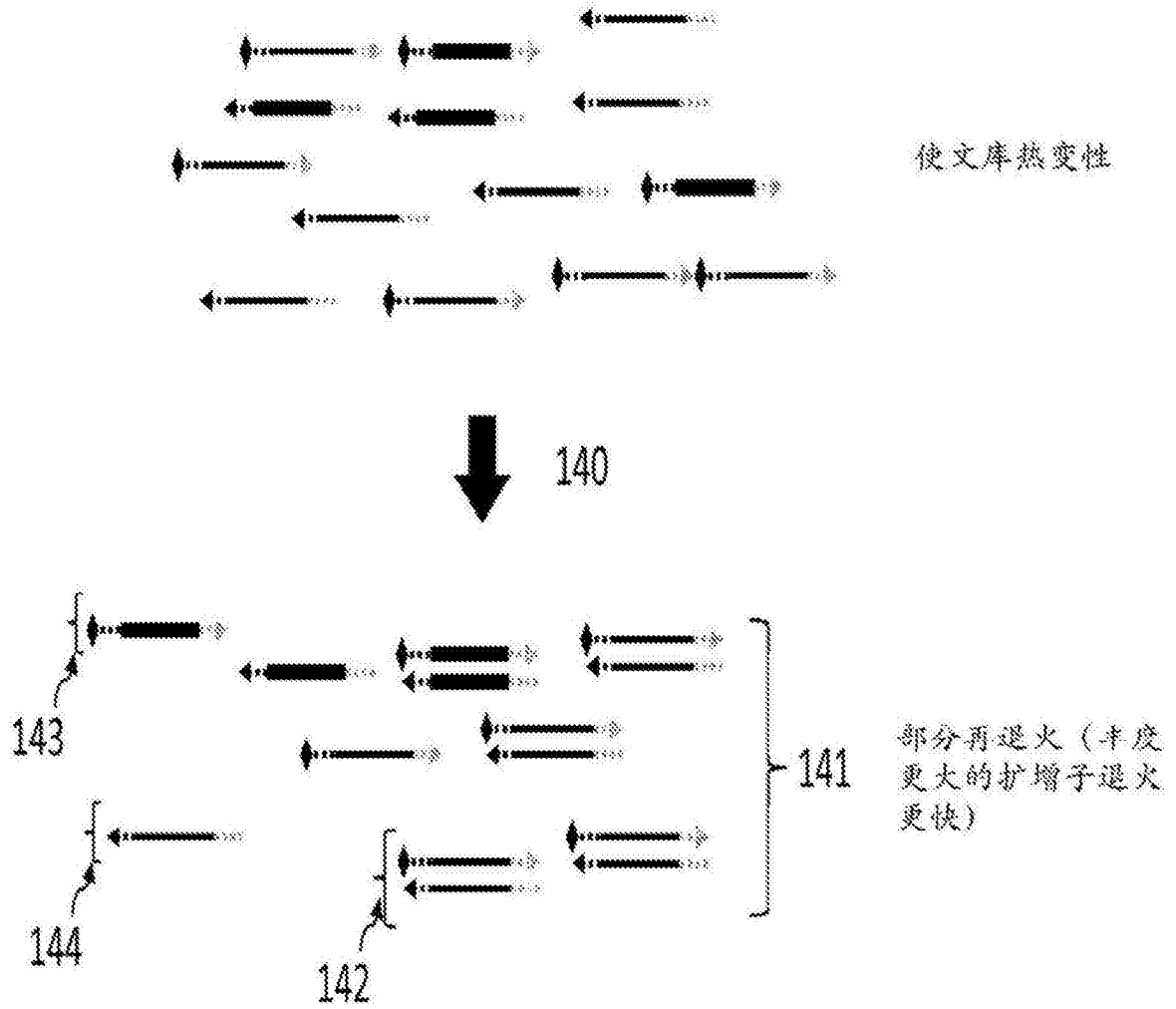


图1 (续)

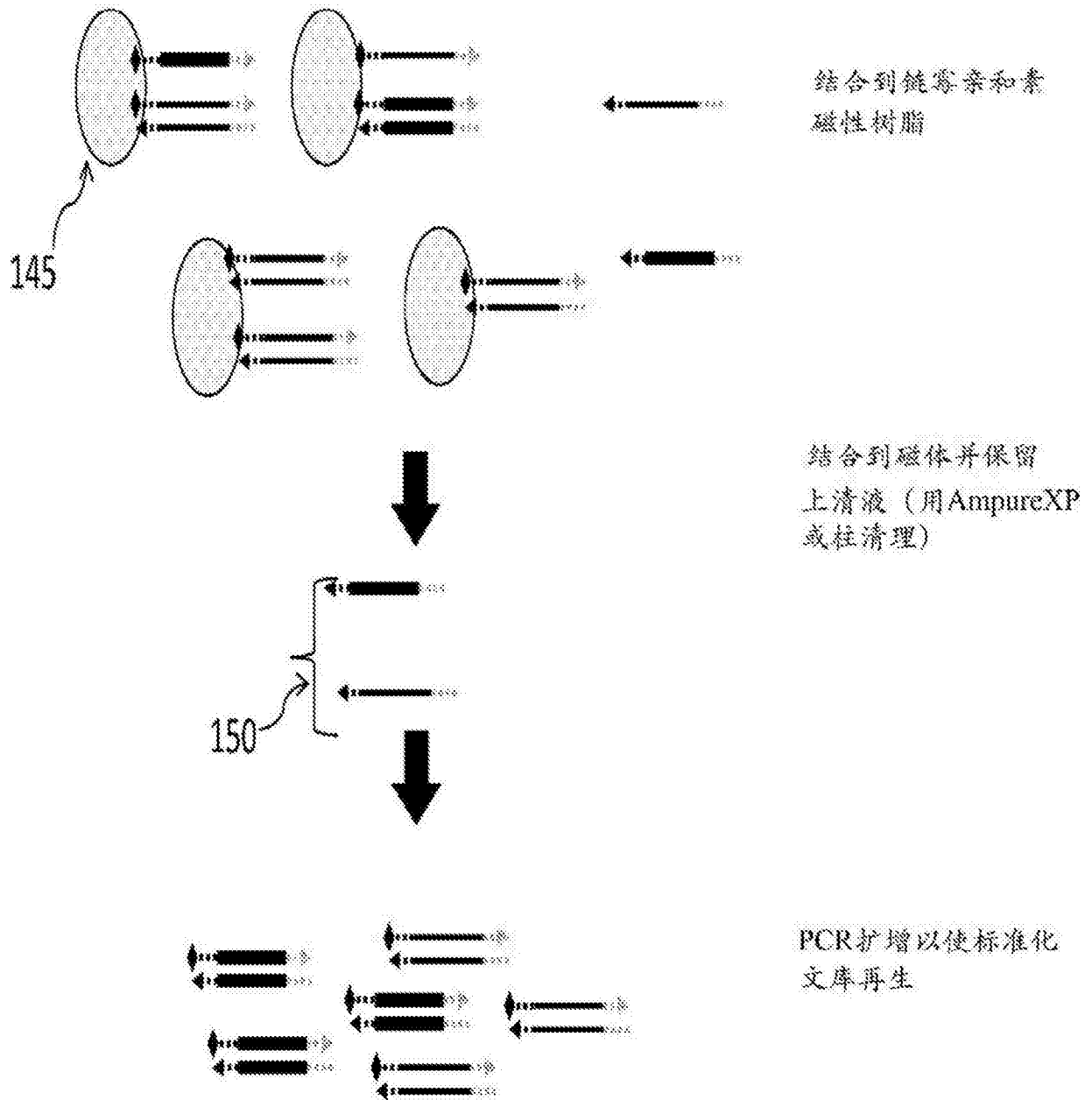


图1(续)

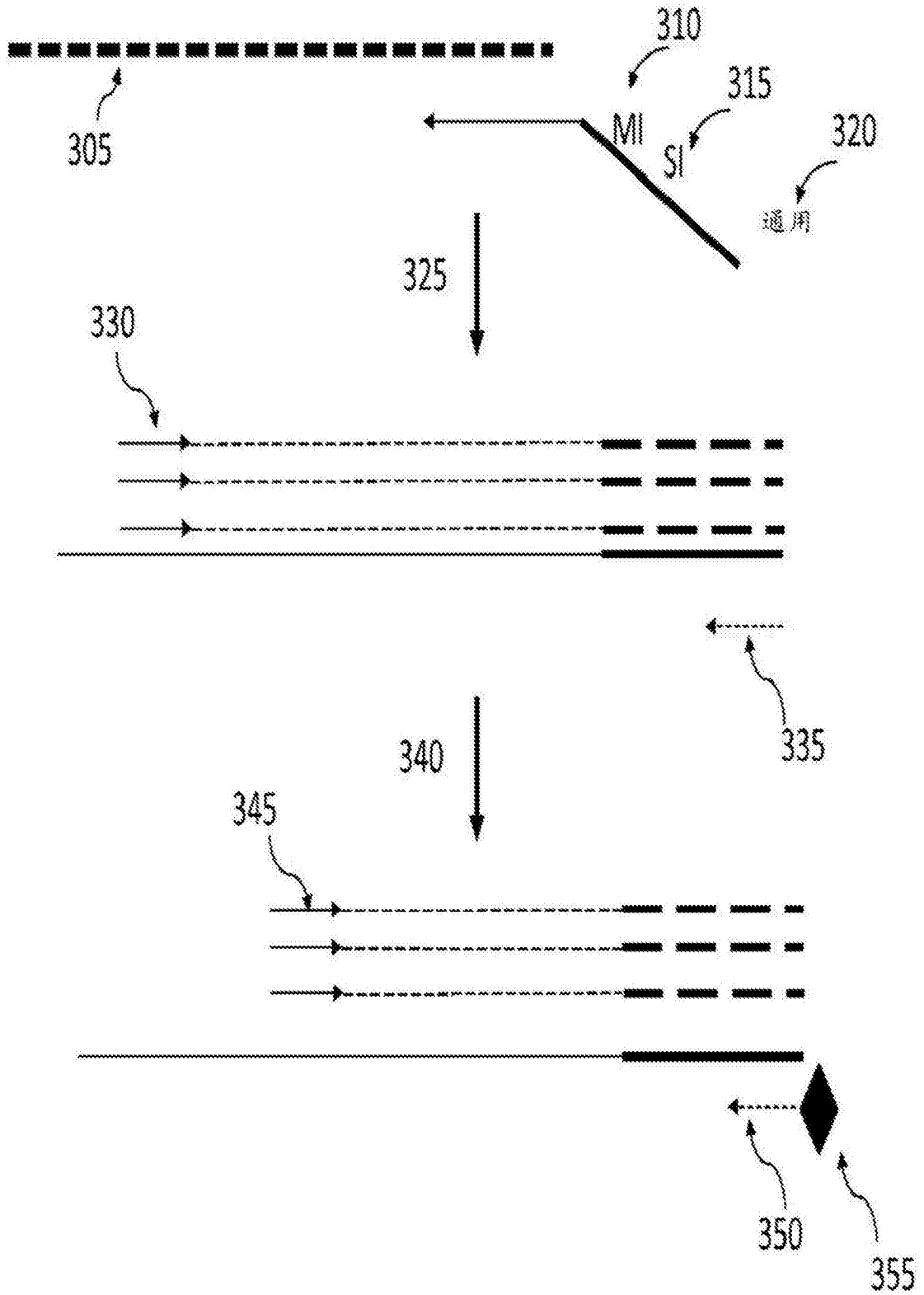


图3

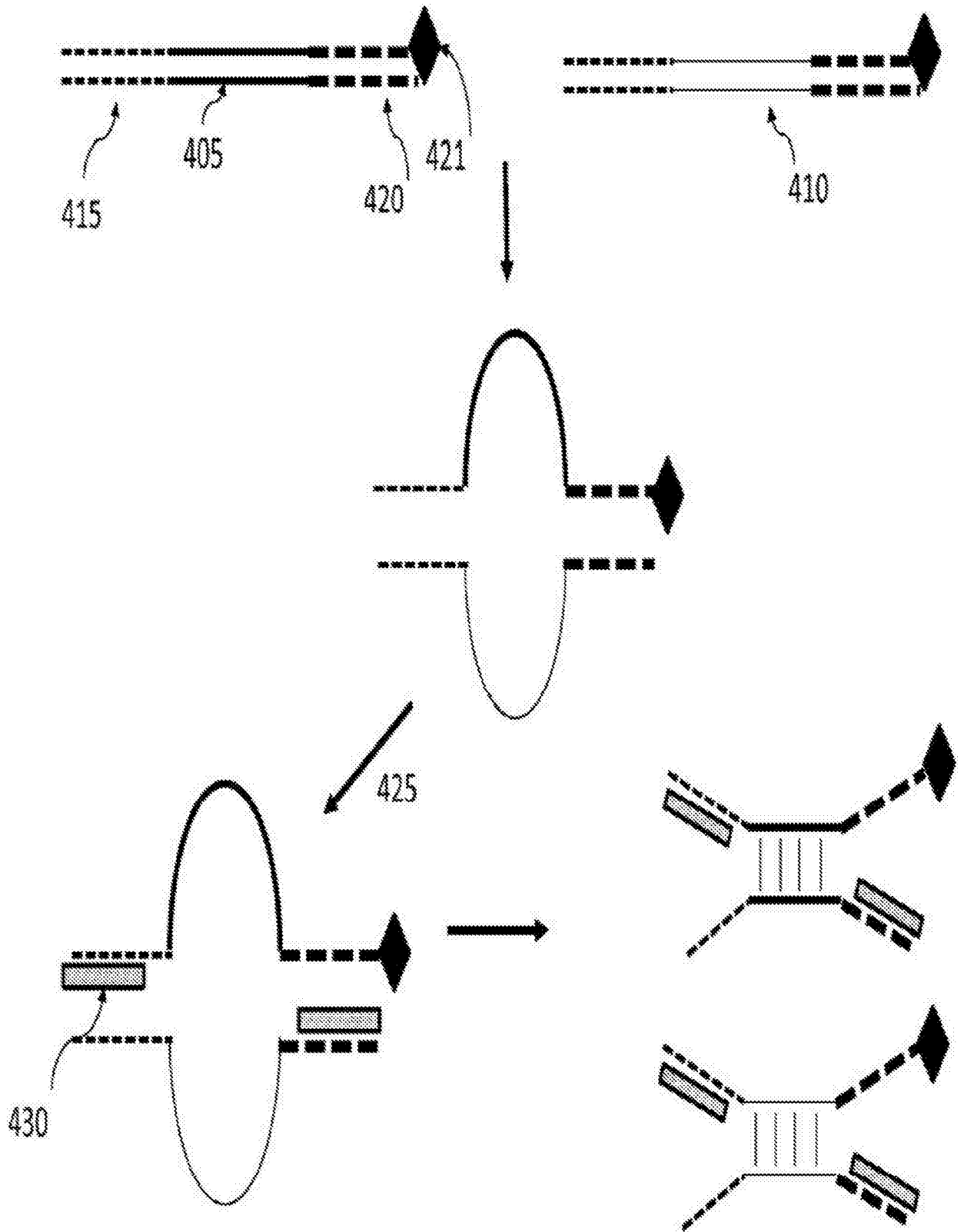


图4

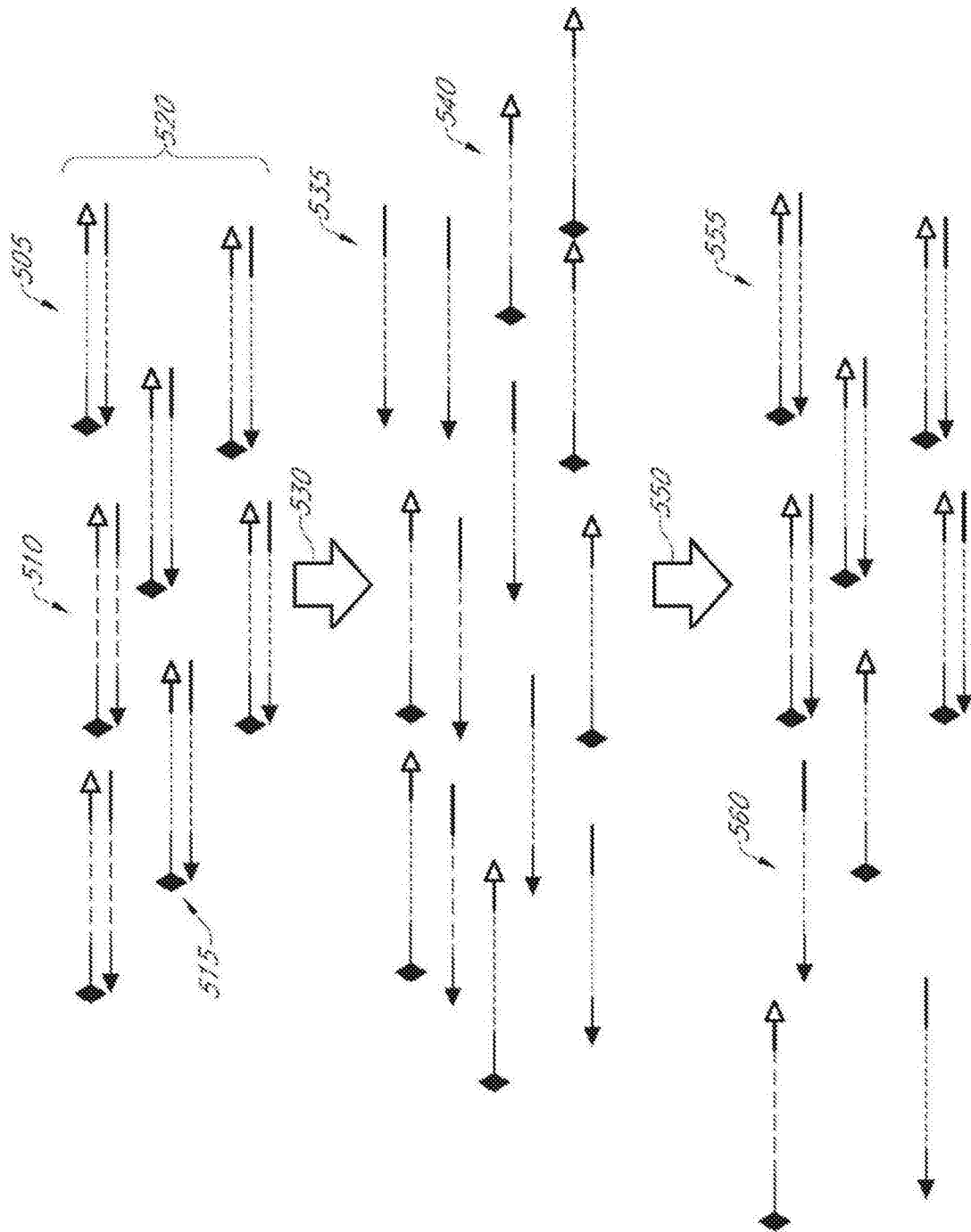


图5A

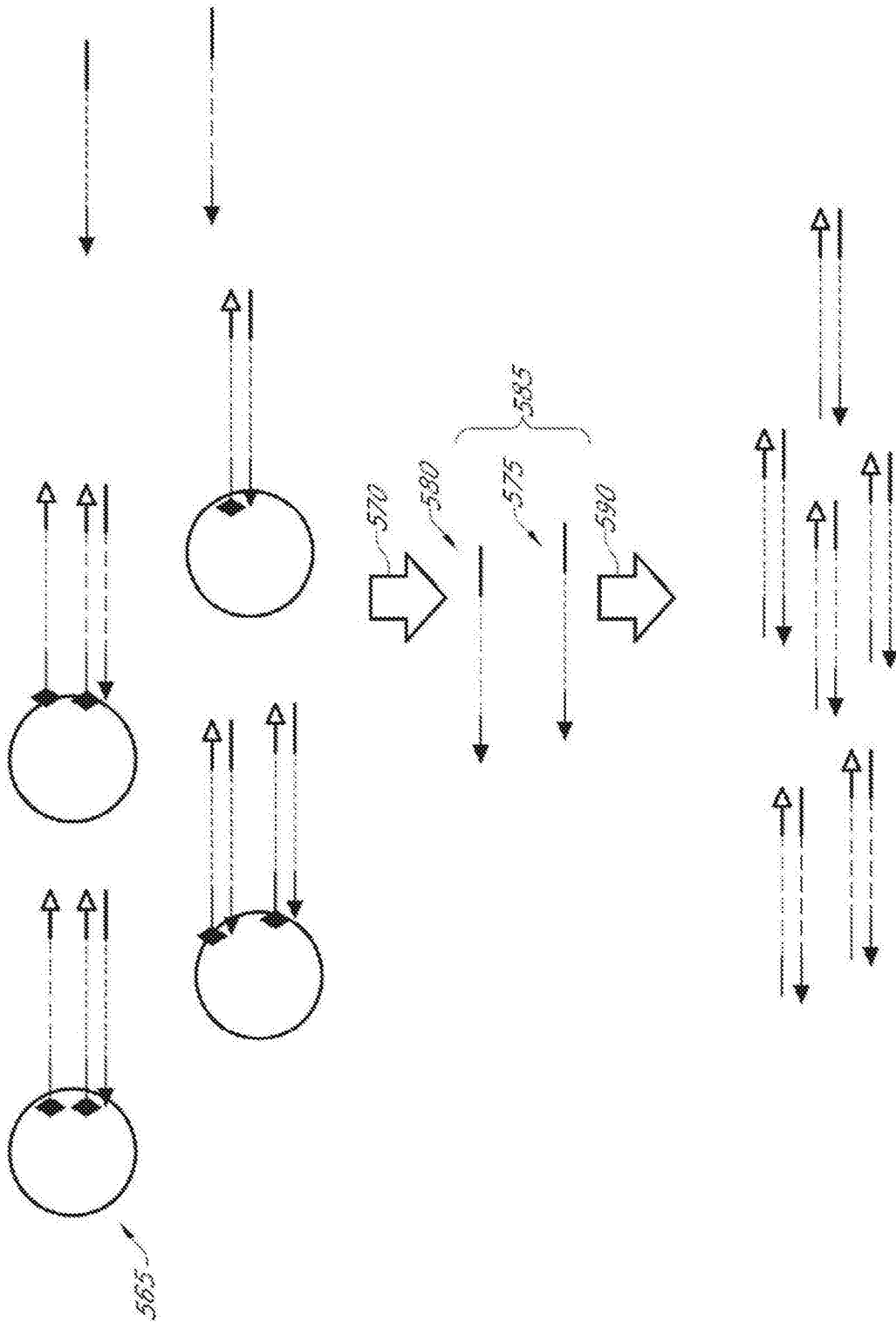


图5B