



(12) 发明专利申请

(10) 申请公布号 CN 115103909 A

(43) 申请公布日 2022. 09. 23

(21) 申请号 202180011915.0

(22) 申请日 2021.02.01

(30) 优先权数据

62/968,847 2020.01.31 US

62/987,232 2020.03.09 US

62/988,859 2020.03.12 US

(85) PCT国际申请进入国家阶段日

2022.07.29

(86) PCT国际申请的申请数据

PCT/US2021/016089 2021.02.01

(87) PCT国际申请的公布数据

W02021/155374 EN 2021.08.05

(71) 申请人 艾维达生物医学公司

地址 美国加利福尼亚州

(72) 发明人 林盛榕 包赞 王恒 格雷斯·赵

(74) 专利代理机构 北京安信方达知识产权代理

有限公司 11262

专利代理师 刘晓杰 武晶晶

(51) Int.Cl.

C12N 15/11 (2006.01)

C40B 40/06 (2006.01)

C12Q 1/6806 (2006.01)

C12Q 1/6853 (2006.01)

C12Q 1/6874 (2006.01)

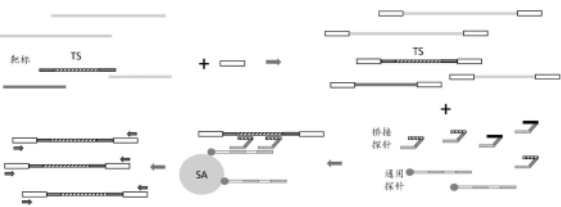
(54) 发明名称

用于靶向核酸捕获的系统和方法

(57) 摘要

本公开提供了用于模板的靶向间接、协同杂交捕获以用于靶序列的扩增和分析的系统和方法。捕获的模板可以用亚硫酸氢盐或其他甲基化试剂进一步处理,以研究模板核酸分子的甲基化模式。

SICON 靶向测序 (TS)



1. 一种方法,所述方法包括:

获得模板核酸分子,在所述模板核酸分子的5'端或3'端包含衔接子;

将第一桥接探针的第一靶特异性区域与所述模板核酸分子的第一靶序列杂交,其中所述第一桥接探针的第一衔接子着陆序列与衔接子锚定探针的第一桥接结合序列结合;并且

将第二桥接探针的第二靶特异性区域与所述模板核酸分子的第二靶序列杂交,其中所述第二桥接探针的第二衔接子着陆序列与所述衔接子锚定探针的第二桥接结合序列结合。

2. 根据权利要求1所述的方法,进一步包括将所述衔接子附接至样本核酸分子的5'端或3'端,从而生成包含所述衔接子的模板核酸分子。

3. 根据权利要求2所述的方法,包括将所述衔接子附接至样本核酸分子的5'端或3'端,并且将衔接子分别附接至包含所述衔接子的所述模板核酸分子的3'端或5'端,从而生成在每一端包含衔接子的模板核酸分子。

4. 根据权利要求2或3所述的方法,进一步包括将衔接子引物与附接至所述模板核酸分子的3'端的衔接子杂交,所述模板核酸分子与所述第一桥接探针和所述第二桥接探针杂交;并且延伸所述衔接子引物的3'端,从而生成延伸产物。

5. 根据权利要求4所述的方法,进一步包括对所述延伸产物进行测序。

6. 根据权利要求1-5中任一项所述的方法,其中在与所述第一靶特异性区域杂交之前,所述第一桥接探针的所述第一衔接子着陆序列与所述衔接子锚定探针的所述第一桥接结合序列结合。

7. 根据权利要求1-5中任一项所述的方法,其中在与所述第一靶特异性区域杂交之后,所述第一桥接探针的所述第一衔接子着陆序列与所述衔接子锚定探针的所述第一桥接结合序列结合。

8. 根据权利要求1-7中任一项所述的方法,其中在与所述第二靶特异性区域杂交之前,所述第二桥接探针的所述第二衔接子着陆序列与所述衔接子锚定探针的所述第二桥接结合序列结合。

9. 根据权利要求1-7中任一项所述的方法,其中在与所述第二靶特异性区域杂交之后,所述第二桥接探针的所述第二衔接子着陆序列与所述衔接子锚定探针的所述第二桥接结合序列结合。

10. 根据权利要求1-9中任一项所述的方法,进一步包括将所述第一桥接探针的所述第一衔接子着陆序列与所述衔接子锚定探针的所述第一桥接结合序列杂交。

11. 根据权利要求1-10中任一项所述的方法,进一步包括将所述第二桥接探针的所述第二衔接子着陆序列与所述衔接子锚定探针的所述第二桥接结合序列杂交。

12. 根据权利要求1-11中任一项所述的方法,其中所述衔接子锚定探针进一步包含位于所述第一桥接结合序列和所述第二桥接结合序列之间的间隔子。

13. 根据权利要求1-12中任一项所述的方法,其中所述衔接子包含分子条形码。

14. 根据权利要求1-13中任一项所述的方法,其中所述衔接子锚定探针包含结合部分。

15. 根据权利要求14所述的方法,其中将所述结合部分附接至支持物。

16. 根据权利要求15所述的方法,其中所述支持物是珠子。

17. 根据权利要求16所述的方法,其中所述珠子是链霉亲和素珠子。

18. 根据权利要求14-17中任一项所述的方法,其中所述结合部分是生物素。

19. 根据权利要求1-18中任一项所述的方法,其中所述第一桥接探针包含结合部分。
20. 根据权利要求19所述的方法,其中将所述结合部分附接至支持物。
21. 根据权利要求20所述的方法,其中所述支持物是珠子。
22. 根据权利要求21所述的方法,其中所述珠子是链霉亲和素珠子。
23. 根据权利要求19-22中任一项所述的方法,其中所述结合部分是生物素。
24. 根据权利要求1-23中任一项所述的方法,其中所述模板核酸分子包含单链DNA。
25. 根据权利要求1-23中任一项所述的方法,其中所述模板核酸分子包含来自生物样本的无细胞核酸。
26. 根据权利要求25所述的方法,其中所述无细胞核酸包含无细胞DNA。
27. 根据权利要求26所述的方法,其中所述无细胞DNA包含循环肿瘤DNA。
28. 根据权利要求1-27中任一项所述的方法,其中所述模板核酸分子包含受损的DNA。
29. 一种方法,所述方法包括:

将第一桥接探针的第一靶特异性区域与模板核酸分子的第一靶序列杂交,其中所述第一桥接探针的第一衔接子着陆序列与衔接子锚定探针的第一桥接结合序列结合;

将第二桥接探针的第二靶特异性区域与模板核酸分子的第二靶序列杂交,其中所述第二桥接探针的第二衔接子着陆序列与衔接子锚定探针的第二桥接结合序列结合,从而生成与所述第一桥接探针和所述第二桥接探针杂交的模板核酸分子;并且

在所述第一靶特异性区域的杂交和所述第二靶特异性区域的杂交之后,用甲基化测定试剂处理所述模板核酸分子。
30. 根据权利要求29所述的方法,其中所述甲基化测定试剂是二硫化物或修饰甲基化胞嘧啶的酶。
31. 根据权利要求29所述的方法,进一步包括将第三桥接探针的第三靶特异性区域与模板核酸分子的第三靶序列杂交,其中所述第三桥接探针的第三衔接子着陆序列与衔接子锚定探针的第三桥接结合序列结合。
32. 根据权利要求31所述的方法,进一步包括将第四桥接探针的第四靶特异性区域与模板核酸分子的第四靶序列杂交,其中所述第四桥接探针的第四衔接子着陆序列与衔接子锚定探针的第四桥接结合序列结合。
33. 根据权利要求29所述的方法,进一步包括在杂交所述第一桥接探针和杂交所述第二桥接探针之前,将衔接子附接至所述模板核酸分子的5'端或3'端。
34. 根据权利要求33所述的方法,进一步包括将衔接子引物与附接至所述模板核酸分子的3'端的衔接子杂交,所述模板核酸分子与所述第一桥接探针和所述第二桥接探针杂交;并且延伸所述衔接子引物的3'端,从而生成延伸产物。
35. 根据权利要求34所述的方法,进一步包括对所述延伸产物进行测序。
36. 根据权利要求34所述的方法,其中所述衔接子引物的杂交在用亚硫酸氢盐处理之前进行。
37. 根据权利要求34所述的方法,其中所述衔接子引物的杂交在用亚硫酸氢盐处理之后进行。
38. 根据权利要求34所述的方法,其中所述衔接子引物基于用亚硫酸氢盐处理之后的所述衔接子设计,其中所述衔接子中的非甲基化胞嘧啶在处理过程中转化为尿嘧啶。

39. 根据权利要求29-38中任一项所述的方法,其中在所述第一靶特异性区域杂交之前,所述第一桥接探针的所述第一衔接子着陆序列与所述衔接子锚定探针的所述第一桥接结合序列结合。

40. 根据权利要求29-38中任一项所述的方法,其中在所述第一靶特异性区域杂交之后,所述第一桥接探针的所述第一衔接子着陆序列与所述衔接子锚定探针的所述第一桥接结合序列结合。

41. 根据权利要求29-40中任一项所述的方法,其中在所述第二靶特异性区域杂交之前,所述第二桥接探针的所述第二衔接子着陆序列与所述衔接子锚定探针的所述第二桥接结合序列结合。

42. 根据权利要求29-40中任一项所述的方法,其中在所述第二靶特异性区域杂交之后,所述第二桥接探针的所述第二衔接子着陆序列与所述衔接子锚定探针的所述第二桥接结合序列结合。

43. 根据权利要求29-42中任一项所述的方法,进一步包括将所述第一桥接探针的所述第一衔接子着陆序列与所述衔接子锚定探针的所述第一桥接结合序列杂交。

44. 根据权利要求29-43中任一项所述的方法,进一步包括将所述第二桥接探针的所述第二衔接子着陆序列与所述衔接子锚定探针的所述第二桥接结合序列杂交。

45. 根据权利要求1-44中任一项所述的方法,其中所述衔接子锚定探针进一步包含位于所述第一桥接结合序列和所述第二桥接结合序列之间的间隔子。

46. 根据权利要求30-45中任一项所述的方法,其中所述衔接子包含分子条形码。

47. 根据权利要求29-46中任一项所述的方法,其中所述衔接子锚定探针包含结合部分。

48. 根据权利要求47所述的方法,其中将所述结合部分附接至支持物。

49. 根据权利要求48所述的方法,其中所述支持物是珠子。

50. 根据权利要求49所述的方法,其中所述珠子是链霉亲和素珠子。

51. 根据权利要求47-50中任一项所述的方法,其中所述结合部分是生物素。

52. 根据权利要求29-50中任一项所述的方法,其中所述第一桥接探针包含结合部分。

53. 根据权利要求52所述的方法,其中将所述结合部分附接至支持物。

54. 根据权利要求53所述的方法,其中所述支持物是珠子。

55. 根据权利要求54所述的方法,其中所述珠子是链霉亲和素珠子。

56. 根据权利要求52-55中任一项所述的方法,其中所述结合部分是生物素。

57. 根据权利要求29-56中任一项所述的方法,其中所述模板核酸分子包含单链DNA。

58. 根据权利要求29-56中任一项所述的方法,其中所述模板核酸分子包含来自生物样本的无细胞核酸。

59. 根据权利要求57所述的方法,其中所述无细胞核酸包含无细胞DNA。

60. 根据权利要求58所述的方法,其中所述无细胞DNA包含循环肿瘤DNA。

61. 根据权利要求29-60中任一项所述的方法,其中所述模板核酸分子包含受损的DNA。

62. 一种试剂盒,所述试剂盒包含:

桥接探针,其包含被配置成与模板核酸分子的靶序列杂交的靶特异性区域;

衔接子锚定探针,其包含被配置成与所述桥接探针的衔接子着陆序列杂交的桥接结合

序列;和

衔接子,其被配置成附接至所述模板核酸分子的5'端或3'端。

63.一种组合物,所述组合物包含:

模板核酸分子,其中所述模板核酸分子的5'端或3'端附接至衔接子;

第一桥接探针,其中第一桥接探针的第一靶特异性区域与所述模板核酸分子的第一靶序列杂交;

第二桥接探针,其中第二桥接探针的第二靶特异性区域与所述模板核酸分子的第二靶序列杂交;和

衔接子锚定探针,其中所述衔接子锚定探针的第一桥接结合序列与所述第一桥接探针的第一衔接子着陆序列结合,并且所述衔接子锚定探针的第二桥接结合序列与所述第二桥接探针的第二衔接子着陆序列结合。

64.一种核酸复合物,所述核酸复合物包含:

模板核酸分子,其中所述模板核酸分子的5'端或3'端附接至衔接子,其中所述模板核酸分子的第一靶序列与第一桥接探针的第一靶特异性区域杂交,并且所述模板核酸分子的第二靶序列与第二桥接探针的第二靶特异性区域杂交,并且其中所述第一桥接探针的第一衔接子着陆序列与衔接子锚定探针的第一桥接结合序列结合,并且所述第二桥接探针的第二衔接子着陆序列与所述衔接子锚定探针的第二桥接结合序列结合。

65.一种包含权利要求64的核酸复合物的组合物。

66.一种顺序富集的方法,所述方法包括:

获得包含多个核酸分子的样本;

进行第一靶标富集以富集包含对应于一个或多个基因组区域的第一基因包的序列的核酸分子,从而生成第一富集样本和剩余样本,所述第一富集样本包含富集有对应于一个或多个基因组区域的所述第一基因包的序列的核酸,所述剩余样本包含耗竭了对应于一个或多个基因组区域的所述第一基因包的序列的核酸;并且

对所述剩余样本进行第二靶标富集以富集包含对应于一个或多个基因组区域的第二基因包的序列的核酸分子,从而生成第二富集样本,所述第二富集样本包含富集有对应于一个或多个基因组区域的所述第二基因包的序列的核酸;其中一个或多个基因组区域的所述第一基因包和一个或多个基因组区域的所述第二基因包不同。

67.根据权利要求66所述的方法,进一步包括对所述第一富集样本进行第一分析和对所述第二富集样本进行第二分析。

68.根据权利要求67所述的方法,其中所述第一分析是序列分析,并且所述第二分析是甲基化分析。

69.根据权利要求67所述的方法,其中所述第一分析是第一序列分析,并且所述第二分析是第二序列分析,其中所述第一序列分析以与所述第二序列分析不同的测序深度进行。

70.根据权利要求66-69中任一项所述的方法,其中所述样本是cfDNA样本。

71.根据权利要求66-69中任一项所述的方法,其中一个或多个基因组区域的所述基因包的基因组区域的靶标富集包括通过杂交的靶标富集。

72.根据权利要求66-69中任一项所述的方法,其中一个或多个基因组区域的所述基因包的基因组区域的靶标富集:

将第一桥接探针的第一靶特异性区域与具有对应于所述基因组区域的序列的分子的第一靶序列杂交,其中所述第一桥接探针的第一衔接子着陆序列与衔接子锚定探针的第一桥接结合序列结合;并且

将第二桥接探针的第二靶特异性区域与具有对应于所述基因组区域的序列的分子的第二靶序列杂交,其中所述第二桥接探针的第二衔接子着陆序列与所述衔接子锚定探针的第二桥接结合序列结合。

73.根据权利要求72所述的方法,其中所述衔接子锚定探针包含结合部分。

74.根据权利要求73所述的方法,进一步包括将所述结合部分附接至支持物,并且将具有附接的结合部分的所述支持物与未结合的核酸分离。

75.根据权利要求66所述的方法,其中基因组区域的所述第一基因包或基因组区域的所述第二基因包包含启动子区域。

76.根据权利要求66或75所述的方法,其中基因组区域的所述第一基因包或基因组区域的所述第二基因包包含内含子区域。

77.根据权利要求66、75或76所述的方法,其中基因组区域的所述第一基因包或基因组区域的所述第二基因包包含外显子区域。

78.根据权利要求66-77中任一项所述的方法,进一步包括将衔接子附接至所述多个核酸分子的核酸分子的5'端或3'端,从而生成包含衔接子的核酸分子文库。

79.根据权利要求66-78中任一项所述的方法,其中所述第二富集样本经亚硫酸氢盐处理并且经受测序反应。

80.根据权利要求79所述的方法,其中所述测序反应的信息读段的数量是可以从所述样本在经受单一靶标富集以富集包含对应于一个或多个基因组区域的第二基因包的序列的核酸分子的情况下所获得的信息读段的数量的至少60%、65%、70%、75%、80%、85%、90%或95%。

81.根据权利要求66所述的方法,进一步包括对第二剩余样本进行第三靶标富集,所述第二剩余样本包含耗竭了对应于一个或多个基因组区域的所述第一基因包和一个或多个基因组区域的所述第二基因包的序列的核酸,以富集包含对应于一个或多个基因组区域的第三基因包的序列的核酸分子,从而生成第三富集样本,所述第三富集样本包含富集有对应于一个或多个基因组区域的所述第三基因包的序列的核酸;其中一个或多个基因组区域的所述第一基因包、一个或多个基因组区域的所述第二基因包和一个或多个基因组区域的所述第三基因包不同。

82.根据权利要求72所述的方法,进一步包括将第三桥接探针的第三靶特异性区域与具有对应于所述基因组区域的序列的所述分子的第三靶序列杂交,其中所述第三桥接探针的第三衔接子着陆序列与所述衔接子锚定探针的第三桥接结合序列结合。

83.根据权利要求72所述的方法,进一步包括将第四桥接探针的第四靶特异性区域与具有对应于所述基因组区域的序列的所述分子的第四靶序列杂交,其中所述第四桥接探针的第四衔接子着陆序列与所述衔接子锚定探针的第四桥接结合序列结合。

用于靶向核酸捕获的系统和方法

交叉引用

[0001] 本申请要求于2020年1月31日提交的美国临时申请号62/968,847、于2020年3月9日提交的美国临时申请号62/987,232和于2020年3月12日提交的美国临时申请号62/988,859的权益,这些申请通过引用并入本文。

[0002] 本申请与以下共同未决专利申请有关:于2019年11月20日提交的国际申请号PCT/US2019/062508,其通过引用并入本文。

背景技术

[0001] 核酸靶标捕获方法可以使特定基因、外显子和其他目的基因组区域富集,例如用于靶向测序。然而,基于靶标捕获的测序方法可能涉及繁琐冗长的方案和昂贵的过程,以及小型捕获基因包 (panel) (例如,少于500个探针) 的低中靶率。此外,由于回收率低,目前的核酸靶标捕获方法可能不适合低输入和受损的DNA。

[0002] 亚硫酸氢盐转化可以是研究核酸分子甲基化模式的有用技术。然而,亚硫酸氢盐转化可以通过例如产生截断的方式损坏核酸。如果用亚硫酸氢盐处理二代测序 (NGS) DNA文库,则大量核酸可能被破坏并且无法在后续的扩增步骤中回收,从而提供低回收率。此外,由于亚硫酸氢盐转化可以产生单链或片段化DNA并降低序列复杂性,因此转化的DNA可能是传统的基于衔接子连接的文库构建的困难输入。考虑到低回收率 (例如,亚硫酸氢盐处理的cfDNA为5%或更低),初始输入通常较小的亚硫酸氢盐处理的无细胞DNA (cfDNA) 或循环肿瘤细胞DNA (ctDNA) 可能带来更大的挑战。还可以进行甲基化敏感酶处理以转化甲基化胞嘧啶。然而,基于酶的方法仍然会在漫长的多步骤过程中遭受甲基化状态的丧失,导致回收率低。

[0003] 无细胞DNA中的甲基化分析对于早期癌症检测具有巨大潜力。在早期癌症患者的血浆中,肿瘤含量估计低于0.1%,通常低至0.01%或更低,因此需要高灵敏度的检测。目前有两种主要的癌症筛查方法:全面方法,包括全基因组亚硫酸氢盐测序 (WGBS)、简并代表性亚硫酸氢盐测序 (RRBS) 或基于亲和力的富集,以及含有10,000个或更多个潜在甲基化标志物的大型靶向基因包。靶向甲基化测序 (TMS) 提供对甲基化标志物最灵敏、特异性的分析。然而,传统TMS的灵敏度和特异性受靶标富集的低效率和低回收率的影响,并进一步受到与大型基因包相关的背景噪声的阻碍。需要使用小型、集中的癌症特异性甲基化生物标志物基因包进行深入分析的方法。

[0004] 因此,需要更有效、易于使用、快速、灵活和实用的靶核酸捕获方法和改进的用于分析亚硫酸氢盐处理的核酸的方法,特别是对于低输入样本 (例如cfDNA)。本文公开的方法可用于极低DNA输入样本的扩增前和亚硫酸氢盐转化前基于杂交的捕获。

发明内容

[0005] 本文公开了一种方法,该方法包括:获得模板核酸分子,在该模板核酸分子的5'端或3'端包含衔接子;将第一桥接探针的第一靶特异性区域与模板核酸分子的第一靶序列杂

交,其中第一桥接探针的第一衔接子着陆序列与衔接子锚定探针的第一桥接结合序列结合;并且将第二桥接探针的第二靶特异性区域与模板核酸分子的第二靶序列杂交,其中第二桥接探针的第二衔接子着陆序列与衔接子锚定探针的第二桥接结合序列结合。方法可以进一步包括将衔接子附接至样本核酸分子的5'端或3'端,从而生成包含衔接子的模板核酸分子。方法可以进一步包括将衔接子附接至样本核酸分子的5'端或3'端,并且将衔接子分别附接至包含衔接子的模板核酸分子的3'端或5'端,从而生成在每一端包含衔接子的模板核酸分子。方法可以进一步包括将衔接子引物与附接至模板核酸分子的3'端的衔接子杂交,该模板核酸分子与第一桥接探针和第二桥接探针杂交;并且延伸衔接子引物的3'端,从而生成延伸产物。方法可以进一步包括对延伸产物进行测序。

[0006] 在第一靶特异性区域杂交之前,第一桥接探针的第一衔接子着陆序列可以与衔接子锚定探针的第一桥接结合序列结合。在第一靶特异性区域杂交之后,第一桥接探针的第一衔接子着陆序列可以与衔接子锚定探针的第一桥接结合序列结合。在第二靶特异性区域杂交之前,第二桥接探针的第二衔接子着陆序列可以与衔接子锚定探针的第二桥接结合序列结合。在第二靶特异性区域杂交之后,第二桥接探针的第二衔接子着陆序列可以与衔接子锚定探针的第二桥接结合序列结合。

[0007] 方法可以进一步包括将第一桥接探针的第一着陆序列与衔接子锚定探针的第一桥接结合序列杂交。方法可以进一步包括将第二桥接探针的第二着陆序列与衔接子锚定探针的第二桥接结合序列杂交。衔接子锚定探针可以进一步包含位于第一桥接结合序列和第二桥接结合序列之间的间隔子。衔接子可以包含分子条形码。

[0008] 衔接子锚定探针可以包含结合部分。可以将结合部分附接至支持物。支持物可以是珠子。珠子可以是链霉亲和素珠子。结合部分可以是生物素。

[0009] 第一桥接探针可以包含结合部分。可以将结合部分附接至支持物。支持物可以是珠子。珠子可以是链霉亲和素珠子。结合部分可以是生物素。

[0010] 模板核酸分子可以包含单链DNA。模板核酸分子可以包含来自生物样本的无细胞核酸。无细胞核酸可以包含无细胞DNA。无细胞DNA可以包含循环肿瘤DNA。模板核酸分子可以包含受损的DNA。

[0011] 本文公开了一种方法,该方法包括:将第一桥接探针的第一靶特异性区域与模板核酸分子的第一靶序列杂交,其中第一桥接探针的第一衔接子着陆序列与衔接子锚定探针的第一桥接结合序列结合;将第二桥接探针的第二靶特异性区域与模板核酸分子的第二靶序列杂交,其中第二桥接探针的第二衔接子着陆序列与衔接子锚定探针的第二桥接结合序列结合,从而生成与第一桥接探针和第二桥接探针杂交的模板核酸分子;并且在第一靶特异性区域杂交和第二靶特异性区域杂交之后,用甲基化测定试剂处理模板核酸分子。甲基化测定试剂可以是二硫化物,或修饰甲基化胞嘧啶的酶。方法可以进一步包括将第三桥接探针的第三靶特异性区域与模板核酸分子的第三靶序列杂交,其中第三桥接探针的第三衔接子着陆序列与衔接子锚定探针的第三桥接结合序列结合。方法可以进一步包括将第四桥接探针的第四靶特异性区域与模板核酸分子的第四靶序列杂交,其中第四桥接探针的第四衔接子着陆序列与衔接子锚定探针的第四桥接结合序列结合。

[0012] 方法可以进一步包括在杂交第一桥接探针和杂交第二桥接探针之前,将衔接子附接至模板核酸分子的5'端或3'端。方法可以进一步包括将衔接子引物与附接至模板核酸分

子的3'端的衔接子杂交,该模板核酸分子与第一桥接探针和第二桥接探针杂交;并延伸衔接子引物的3'端,从而生成延伸产物。方法可以进一步包括对延伸产物进行测序。

[0013] 衔接子引物的杂交可以在用亚硫酸氢盐处理之前进行。衔接子引物的杂交可以在用亚硫酸氢盐处理之后进行。衔接子引物可以基于亚硫酸氢盐处理之后的衔接子设计,其中衔接子中的非甲基化胞嘧啶在处理过程中转化为尿嘧啶。在第一靶特异性区域杂交之前,第一桥接探针的第一衔接子着陆序列可以与衔接子锚定探针的第一桥接结合序列结合。在第一靶特异性区域杂交之后,第一桥接探针的第一衔接子着陆序列可以与衔接子锚定探针的第一桥接结合序列结合。在第二靶特异性区域杂交之前,第二桥接探针的第二衔接子着陆序列可以与衔接子锚定探针的第二桥接结合序列结合。在第二靶特异性区域杂交之后,第二桥接探针的第二衔接子着陆序列可以与衔接子锚定探针的第二桥接结合序列结合。

[0014] 方法可以进一步包括将第一桥接探针的第一着陆序列与衔接子锚定探针的第一桥接结合序列杂交。方法可以进一步包括将第二桥接探针的第二着陆序列与衔接子锚定探针的第二桥接结合序列杂交。衔接子锚定探针可以进一步包含位于第一桥接结合序列和第二桥接结合序列之间的间隔子。衔接子可包含分子条形码。

[0015] 衔接子锚定探针可以包含结合部分。可以将结合部分附接至支持物。支持物可以是珠子。珠子可以是链霉亲和素珠子。结合部分可以是生物素。第一桥接探针可以包含结合部分。可以将结合部分附接至支持物。支持物可以是珠子。珠子可以是链霉亲和素珠子。结合部分可以是生物素。模板核酸分子可以包含单链DNA。模板核酸分子可以包含来自生物样本的无细胞核酸。无细胞核酸可以包含无细胞DNA。无细胞DNA可以包含循环肿瘤DNA。模板核酸分子可以包含受损的DNA。

[0016] 本文公开了一种试剂盒,该试剂盒包含:桥接探针,其包含被配置成与模板核酸分子的靶序列杂交的靶特异性区域;衔接子锚定探针,其包含被配置成与桥接探针的衔接子着陆序列杂交的桥接结合序列;和衔接子,其被配置成附接至该模板核酸分子的5'端或3'端。

[0017] 本文公开了一种组合物,该组合物包含:模板核酸分子,其中模板核酸分子的5'端或3'端附接至衔接子;第一桥接探针,其中第一桥接探针的第一靶特异性区域与模板核酸分子的第一靶序列杂交;第二桥接探针,其中第二桥接探针的第二靶特异性区域与模板核酸分子的第二靶序列杂交;和衔接子锚定探针,其中衔接子锚定探针的第一桥接结合序列与第一桥接探针的第一衔接子着陆序列结合,并且衔接子锚定探针的第二桥接结合序列与第二桥接探针的第二衔接子着陆序列结合。

[0018] 本文公开了一种核酸复合物,该核酸复合物包含:模板核酸分子,其中模板核酸分子的5'端或3'端附接至衔接子,其中模板核酸分子的第一靶序列与第一桥接探针的第一靶特异性区域杂交,并且模板核酸分子的第二靶序列与第二桥接探针的第二靶特异性区域杂交,并且其中第一桥接探针的第一衔接子着陆序列与衔接子锚定探针的第一桥接结合序列结合,并且第二桥接探针的第二衔接子着陆序列与衔接子锚定探针的第二桥接结合序列结合。本文公开了一种包含核酸复合物的组合物。

[0019] 本文公开了一种顺序富集的方法,该方法包括获得包含多个核酸分子的样本;进行第一靶标富集以富集包含对应于一个或多个基因组区域的第一基因包(panel)的序列的

核酸分子,从而生成第一富集样本和剩余样本,第一富集样本包含富集有对应于一个或多个基因组区域的第一基因包的序列的核酸,剩余样本包含耗竭了对应于一个或多个基因组区域的第一基因包的序列的核酸;并且对剩余样本进行第二靶标富集以富集包含对应于一个或多个基因组区域的第二基因包的序列的核酸分子,从而生成第二富集样本,第二富集样本包含富集有对应于一个或多个基因组区域的第二基因包的序列的核酸;其中一个或多个基因组区域的第一基因包和一个或多个基因组区域的第二基因包不同。

[0020] 方法可以进一步包括对第一富集样本进行第一分析和对第二富集样本进行第二分析。

[0021] 第一分析可以是序列分析,并且第二分析可以是甲基化分析。

[0022] 在一些情况下,第一分析是第一序列分析,并且第二分析是第二序列分析,其中第一序列分析与第二序列分析不同的测序深度进行。

[0023] 在一些情况下,样本是cfDNA样本。

[0024] 在一些情况下,一个或多个基因组区域的基因包的基因组区域的靶标富集包括通过杂交的靶标富集。

[0025] 在一些情况下,一个或多个基因组区域的基因包的基因组区域的靶标富集:将第一桥接探针的第一靶特异性区域与具有对应于基因组区域的序列的分子的第一靶序列杂交,其中第一桥接探针的第一衔接子着陆序列与衔接子锚定探针的第一桥接结合序列结合;并且将第二桥接探针的第二靶特异性区域与具有对应于基因组区域的序列的分子的第二靶序列杂交,其中第二桥接探针的第二衔接子着陆序列与衔接子锚定探针的第二桥接结合序列结合。

[0026] 在一些情况下,衔接子锚定探针包含结合部分。

[0027] 权利要求73的方法,进一步包括将结合部分附接至支持物,并且将具有附接的结合部分的支持物与未结合的核酸分离。

[0028] 在一些情况下,基因组区域的第二基因包包含启动子区域。

[0029] 在一些情况下,基因组区域的第一基因包或基因组区域的第二基因包包含内含子区域。

[0030] 权利要求66、75或76的方法,其中基因组区域的第一基因包或基因组区域的第二基因包包含外显子区域。

[0031] 在一些情况下,方法进一步包括将衔接子附接至多个核酸分子的核酸分子的5'端或3'端,从而生成包含衔接子的核酸分子文库。

[0032] 在一些情况下,第二富集样本经亚硫酸氢盐处理并且经受测序反应。

[0033] 在一些情况下,该测序反应的信息读段的数量是可以从所述样本在经受单一靶标富集以富集包含对应于一个或多个基因组区域的第二基因包的序列的核酸分子的情况下所获得的信息读段的数量的至少60%、65%、70%、75%、80%、85%、90%或95%。

[0034] 在一些情况下,方法进一步包括对以及第二剩余样本进行第三靶标富集,该第二剩余样本包含耗竭了对应于一个或多个基因组区域的第一基因包和一个或多个基因组区域的第二基因包的序列的核酸,以富集包含对应于一个或多个基因组区域的第三基因包的序列的核酸分子,从而生成第三富集样本,该第三富集样本包含富集有对应于一个或多个基因组区域的第三基因包的序列的核酸;其中一个或多个基因组区域的第一基因包、一个

或多个基因组区域的第二基因包和一个或多个基因组区域的第三基因包不同。

[0035] 在一些情况下,方法进一步包括将第三桥接探针的第三靶特异性区域与具有对应于基因组区域的序列的分子的第三靶序列杂交,其中第三桥接探针的第三衔接子着陆序列与衔接子锚定探针的第三桥接结合序列结合。

[0036] 在一些情况下,方法进一步包括将第四桥接探针的第四靶特异性区域与具有对应于基因组区域的序列的分子的第四靶序列杂交,其中第四桥接探针的第四衔接子着陆序列与衔接子锚定探针的第四桥接结合序列结合。

援引并入

[0037] 本说明书中提及的所有出版物、专利和专利申请均通过引用并入本文,其程度与每个单独的出版物、专利或专利申请被具体且单独地指示通过引用并入的程度相同。

附图说明

[0038] 在所附权利要求中特别阐述了本发明的新颖性特征。通过参考以下详细描述和附图,将获得对本发明的特征和优点的更好理解,在详细描述中阐述了利用了本发明的原理的说明性实施方案,在附图中:

[0039] 图1图示了模板核酸分子的协同间接杂交捕获的一个实施方案。在该实施方案中,在间接杂交之前构建模板核酸分子文库。

[0040] 图2A-图2B图示了用于甲基化测序的模板核酸分子的协同间接杂交捕获的一个实施方案。图2A显示了模板核酸分子的协同间接杂交捕获,并且图2B显示了捕获的模板核酸分子随后的亚硫酸氢盐转化。

[0041] 图3显示了模板核酸分子的协同间接杂交捕获和靶向甲基化测序(SICON-TMS)的工作流程。

[0042] 图4显示了协同间接杂交的示意图。

[0043] 图5A-图5D显示了不同杂交系统的示意图。图5A图示了非协同直接杂交。图5B图示了协同直接杂交。图5C图示了协同间接杂交。图5D图示了非协同间接杂交。

[0044] 图6A-图6B图示了使用在衔接子锚定探针的桥接结合序列之间具有或不具有间隔子的衔接子锚定探针的协同间接杂交的示意图。图6A显示了与包含间隔子的衔接子锚定探针的协同间接杂交的示意图。图6B显示了与缺少间隔子的衔接子锚定探针的协同间接杂交。

[0045] 图7显示了使用协同间接捕获方法的15个靶标基因包的测序覆盖度。

[0046] 图8A-图8B显示了使用两种不同杂交方法的一组76个人基因靶标(人ID)的测序覆盖度。图8A显示了通过协同间接杂交进行的扩增前捕获的覆盖度。图8B显示了通过直接杂交进行的扩增后捕获的覆盖度。

[0047] 图9显示了在协同间接捕获从非癌个体中提取的cfDNA后靶向甲基化测序测定的结果。

[0048] 图10图示了靶向甲基化测序测定的结果,显示了加标(spike-in)甲基化DNA的预期量与测量值之间的线性关系。

[0049] 图11A和11B分别显示了正常结肠组织和结肠癌组织基因组DNA中DMR1的分子甲基化散射模式。

[0050] 图12A和12B分别显示了正常结肠组织和结肠癌组织基因组DNA中DMR2的分子甲基化散射模式。

[0051] 图13A和13B分别显示了健康个体血浆cfDNA和结肠癌患者血浆cfDNA中DMR1和DMR2的分子甲基化散射模式。

[0052] 图14图示了用于从样本进行顺序靶标富集的示意图。

[0053] 图15图示了实施例11中在CRC cfDNA样本中识别的突变。

[0054] 图16图示了来自独立和双重分析TMS的甲基化评分。

[0055] 图17图示了来自独立和双重分析TMS的信息分子计数。

[0056] 图18图示了在个性化基因包分析中变异等位基因检测的灵敏度。

[0057] 图19图示了Point-n-SeqTM技术的实施。

具体实施方式

[0058] 使用甲基化和突变分析的基于cfDNA的液体活检可用于癌症早期检测和管理。本文提供了用于对有限量的核酸样本进行组合分析的系统和方法。例如,本文提供了用于对有限的DNA样本进行组合的靶向甲基化测序(TMS)和突变分析的系统和方法。这些系统和方法可以特别适用于数量可能较少的cfDNA样本。

[0059] 癌症基因组中广泛但组织特异性的甲基化变化可用于灵敏检测早期或复发癌症患者血浆中的循环肿瘤(ctDNA)。然而,甲基化分析的灵敏度可能因在该过程中回收甲基化标志物的效率低而受到影响,并且特异性有时进一步受到包括噪声非特异性标志物以补偿检测灵敏度低的方法的阻碍。此外,虽然甲基化分析对于早期癌症检测具有优势,但可操作的突变可以直接提供信息来指导治疗选择并进一步提高检测特异性。来自有限临床血液样本的cfDNA产量可能是低量,这可能是对一个样本进行多重分析的重大挑战,因此可以检测甲基化和突变的检测可以为临床研究和诊断检测提供改进。

[0060] 本公开提供了一种改进的技术,其经设计用于cfDNA中靶向甲基化和突变组合分析:Point-n-Seq,其特征在于在胞嘧啶转化和扩增之前,直接从cfDNA富集靶分子。该技术可以实现小型聚焦基因包,以查询至少10个、100个、1000个或超过1000个标志物的甲基化或突变状态。本文提供了结肠直肠癌(CRC)基因包,其经设计涵盖来自22个基因的100个甲基化标志物和>350个热点突变。Point-n-Seq TMS可用于使用cfDNA进行小型聚焦甲基化和突变组合基因包测序。Point-n-Seq TMS可用于开发实用且具有成本效益的甲基化测定,以供研究和临床使用。

[0061] 利用超高效的转化前/扩增前捕获Point-n-Seq可用于聚焦疾病的甲基化和突变基因包富集。Point-n-Seq TMS能够使用cfDNA分析小型聚焦甲基化和突变基因包。Point-n-Seq TMS可用于实用且具有成本效益的甲基化测定,以供研究和临床使用。

[0062] 本文还提供了用于协同间接捕获核酸以进行测序的系统和方法(SICON-SEQ,也称为Point-n-SEQ)。本文公开的系统和方法使得有效捕获和富集核酸材料。SICON-SEQ/Point-n-SEQ可以在文库构建后将衔接子附接至模板核酸材料上进行捕获富集。在一些实施方案中,SICON-SEQ可以在文库构建之前进行。SICON-SEQ可以在无需文库构建通过衔接子附接进行。本文公开的SICON-SEQ方法可以实现较短的周转时间和简单的工作流程。SICON-SEQ可用于处理低输入样本,例如无细胞DNA(cfDNA),因此可适用于甲基化测序分

析。

[0063] 本文公开了包括通过一种或多种桥接探针与模板核酸杂交而使模板核酸分子与衔接子锚定探针间接杂交的方法。可以设计一种或多种桥接探针以与模板核酸分子中的特定靶序列杂交,从而可以与靶模板杂交。衔接子锚定探针又可以设计成与一种或多种桥接探针杂交,从而生成三种或更多种杂交核酸分子的组装体。多结构杂交组装体可以协同作用,为组装体提供更高的稳定性。随后可以用亚硫酸氢盐处理杂交的模板核酸分子以进行甲基化测序。

[0064] 本文公开了一种试剂盒,该试剂盒包含:桥接探针,其包含与模板核酸分子的靶序列杂交的靶特异性区域;衔接子锚定探针,其包含与桥接探针的衔接子着陆序列杂交的桥接结合序列;和衔接子,其被配置成附接至该模板核酸分子的5'端或3'端。

I. 杂交间接捕获

[0065] 可以通过模板核酸和形成杂交组装体的两种或更多种探针的协同相互作用来促进靶探针杂交。多复合物组装体可以稳定模板与靶探针(例如桥接探针)之间的杂交相互作用。桥接探针可以包含与模板的靶区域杂交的靶特异性区域和与衔接子锚定探针的桥接结合序列(BBS)杂交的衔接子着陆序列(ALS)。模板与桥接探针之间的杂交以及桥接探针与衔接子锚定探针之间的杂交可以形成多复合体组装体。

[0066] 在本文公开的方法中,每个靶区域可以使用多于两个桥接探针。例如,至少2、3、4、5、6、7、8、9、10、25、50、75、100个或更多个桥接探针可用于桥接模板和衔接子锚定探针。用于测序的协同间接捕获核酸(SICON-SEQ)方法可以进一步包括将第二桥接探针的第二靶特异性区域与模板核酸分子的第二靶序列杂交,其中第二桥接探针的第二衔接子着陆序列可以与衔接子锚定探针的第二桥接结合序列结合(图1)。在一些情况下,SICON-SEQ可以在衔接子附接至模板核酸分子之后进行,以生成文库(图1)。该文库可以是二代测序(NGS)文库。

[0067] 桥接探针可以进一步包含连接靶特异性区域和衔接子着陆序列的接头。衔接子锚可以在桥接结合序列之间包含一个或多个间隔子。一个或多个间隔子的存在可以提高杂交捕获的效率并增加捕获的特异性。

[0068] 模板核酸可以从低输入样本(例如无细胞DNA(cfDNA)和循环肿瘤DNA(ctDNA))捕获和富集。捕获和富集可以通过与桥接探针杂交而与衔接子锚定探针间接缔合来完成。桥接探针和/或衔接子锚定探针可以包含一个或多个结合部分。结合部分可以是生物素。可以将结合部分附接至支持物。支持物可以是珠子。珠子可以是链霉亲和素珠子。

[0069] 本文公开了一种试剂盒,该试剂盒包含:桥接探针,其包含与模板核酸分子的靶序列杂交的靶特异性区域;衔接子锚定探针,其包含与桥接探针的衔接子着陆序列杂交的桥接结合序列;和衔接子,其被配置成附接至该模板核酸分子的5'端或3'端。

II. 甲基化分析工作流程

[0070] 本文提供了用于甲基化分析核酸的方法。甲基化分析可以通过亚硫酸氢盐处理进行。亚硫酸氢盐处理的核酸可用于研究核酸的甲基化。亚硫酸氢盐处理可以将未甲基化的胞嘧啶转化为尿嘧啶。胞嘧啶(例如5'-甲基胞嘧啶)的甲基化可以防止亚硫酸氢盐将甲基化胞嘧啶转化为尿嘧啶。

[0071] 模板核酸分子可以在使用捕获探针或桥接探针/衔接子锚定探针杂交捕获之前或之后用亚硫酸氢盐处理。在一些情况下,杂交的模板核酸分子可以用亚硫酸氢盐处理。双链

序列(例如,在模板的TS和捕获探针的TSR之间)的形成可以保护在亚硫酸氢盐处理过程中杂交区域中的胞嘧啶不被转化为尿嘧啶。通过捕获探针与模板或桥接探针与模板和衔接子锚定探针的杂交形成的双链序列可以提供保护,使杂交区域中的胞嘧啶亚硫酸氢盐不被转化为尿嘧啶。此外,由于亚硫酸氢盐处理可以将非甲基化胞嘧啶转化为尿嘧啶,因此在TS区域保护胞嘧啶不被转化为尿嘧啶可以实现使用经设计用于退火至非亚硫酸氢盐转化的DNA的扩增引物。对于亚硫酸氢盐转化前的捕获,还可以针对未转化的序列设计探针。退火至未转化的胞嘧啶的探针和引物可以更直接地设计并提供更好的杂交。

在一些情况下,可以进行酶处理以进行甲基化分析。酶可以是甲基化敏感酶或甲基化依赖性酶。酶可以是限制酶。这些酶可以是甲基化敏感的限制性内切酶。在其他情况下,甲基化分析可以通过使用与甲基化位点特异性结合的特异性抗体或蛋白质来富集甲基化核酸。

a. 模板核酸杂交捕获之后的甲基化处理或富集

[0072] 模板核酸(例如,DNA)可用于如本文所述的协同间接杂交和后续测序(SICON-SEQ)(参见例如图3)。模板核酸(例如DNA)可以是例如基因组DNA或cfDNA。模板核酸(例如,DNA)可以与捕获探针直接杂交或通过桥接探针杂交而间接结合至衔接子锚定探针(或通用锚定探针),例如,如本文所述,例如,如图1和2A所示。杂交捕获的模板核酸(例如,DNA)可以用亚硫酸氢盐处理、延伸和后续扩增(图2B),例如,用于靶向甲基化测序(SICON-TMS)。在一些情况下,可以用甲基化敏感酶处理捕获的模板核酸。在另一种情况下,可以通过与靶向模板核酸分子中的甲基化CpG位点的抗体或蛋白质特异性结合来富集捕获的模板核酸分子的甲基化核酸。SICON-TMS可与大范围内的核物质质量的临床样本相容。在一些情况下,SICON-TMS可用于核酸分子小于5ng、小于4ng、小于3ng、小于2ng或小于1ng的序列样本。

[0073] 捕获探针或桥接探针的靶特异性序列或靶特异性区域(TSR)可以基于模板核酸分子的靶序列进行设计,并且模板核酸分子的靶序列在亚硫酸氢盐处理之后可保留非甲基化胞嘧啶。

[0074] 在一些情况下,亚硫酸氢盐处理可以在桥接探针的靶特异性序列脱离之前进行。在捕获探针或桥接探针的TS和TSR与模板杂交之后发生的亚硫酸氢盐处理过程中,可以保护TS和TSR位点中的未甲基化的胞嘧啶不被转化为尿嘧啶。随后,杂交的模板可以用亚硫酸氢盐处理,在此期间杂交的TSR-TS区域中的非甲基化胞嘧啶不被转化为尿嘧啶,而单链区域中的非甲基化胞嘧啶转化为尿嘧啶。防止TS区域的胞嘧啶转化为尿嘧啶可以实现使用经设计用于退火至非亚硫酸氢盐转化的DNA的探针。

[0075] 在一些情况下,亚硫酸氢盐处理可以在捕获探针或桥接探针从模板核酸序列脱离之后进行。引物结合位点(例如,衔接子和/或模板中)的一个或多个胞嘧啶残基可以不受亚硫酸氢盐转化的保护。在亚硫酸氢盐转化之后,衔接子中的引物结合位点可以包含一个或多个尿嘧啶。可以设计引物以与包含一个或多个尿嘧啶的衔接子序列互补。引物可以与包含一个或多个尿嘧啶的衔接子序列100%互补,或与包含一个或多个尿嘧啶的衔接子序列的互补性小于100%。

[0076] 在亚硫酸氢盐处理之后,模板可以包含一个或多个尿嘧啶。退火至衔接子的引物可以使用包含一个或多个尿嘧啶的模板用于链延伸。延伸链可以包含与一个或多个尿嘧啶碱基配对的一个或多个腺嘌呤。延伸产物可以从模板变性。引物可以退火至包含一个或多

个腺嘌呤的区域中的延伸产物并延伸。引物可用于模板的扩增,例如,使用衔接子引物。

[0077] 甲基化处理或富集可以在附接衔接子之前应用于模板核酸分子。甲基化处理或富集可以在附接衔接子之后应用于模板核酸分子。甲基化处理或富集可以在第一衔接子附接至模板之后应用于模板核酸分子。甲基化处理或富集可以在第二衔接子附接至模板之后应用于模板核酸分子。

b. 模板核酸杂交捕获之前的甲基化处理或富集

[0078] 模板核酸分子可以在与捕获探针或桥接探针杂交之前进行亚硫酸氢盐处理。DNA可以用亚硫酸氢盐处理以将未甲基化的胞嘧啶转化为尿嘧啶。亚硫酸氢盐处理的DNA可用作协同间接杂交和后续测序(SICON-SEQ)的输入。探针的TSR可以设计成退火至模板,其中现有的非甲基化胞嘧啶已转化为尿嘧啶。在杂交捕获之后,可以进行延伸,随后进行靶标扩增。在一些情况下,可以用甲基化敏感酶处理捕获的模板核酸。在另一种情况下,可以通过与靶向模板核酸分子中的甲基化CpG位点的抗体或蛋白质特异性结合来富集捕获的模板核酸分子的甲基化核酸。

[0079] 可以在附接衔接子之前对模板核酸分子进行甲基化处理或富集。甲基化处理或富集可以在附接衔接子之后应用于模板核酸分子。甲基化处理或富集可以在第一衔接子附接至模板之后应用于模板核酸分子。甲基化处理或富集可以在第二衔接子附接至模板之后应用于模板核酸分子。

III. 固相萃取

[0080] 本文提供了例如在衔接子锚定探针与模板连接之前,选择与桥接探针杂交的模板(或通过桥接探针与衔接子锚定探针缔合的模板)的方法。该方法可以采用固相萃取。本文提供了将桥接探针或衔接子锚定探针与固体支持物结合的方法。衔接子锚定探针可以在非依赖于桥接探针的情况下以附接(例如,连接)至模板的可能性引入次优特异性。为了减少此类非特异性连接产物以及未结合的探针,可以使用标签(例如,生物素)和捕获部分(例如,链霉亲和素珠子)。

[0081] 桥接探针或衔接子锚定探针可以包含标签。公开的方法可以进一步包括通过标签捕获桥接探针、衔接子锚定探针或包含模板核酸分子、桥接探针和衔接子锚定探针的杂交复合物。标签可以是生物素。标签可以是核酸序列,例如多聚A或多聚T,或特定序列。核酸序列的长度可以为约5至30个碱基。核酸序列可以包含DNA和/或RNA。标签可以位于桥接探针或衔接子锚定探针的3'端。标签可以是肽,或可以被抗体(例如5-溴尿苷和生物素)识别的修饰核酸。标签可以通过反应(例如“点击”化学)与桥接探针或衔接子锚定探针缀合。“点击”化学可以实现报告分子(如荧光染料)与生物分子(如DNA)缀合。点击化学可以是叠氮化物和炔烃之间的反应,并且可以产生共价产物(例如,1,5-二取代的1,2,3-三唑)。铜可以用作催化剂。

[0082] 可以在固体支持物上捕获标签。固体支持物可以是磁性的。固体支持物可以包括珠子、流动池、玻璃、板、包含一个或多个微流体通道的装置或柱。固体支持物可以是磁珠。

[0083] 固体支持物(例如,珠子)可以包含(例如,通过包被具有)可以结合标签的一个或多个捕获部分。捕获部分可以是链霉亲和素,并且链霉亲和素可以结合生物素。捕获部分可以是抗体。抗体可以结合标签。捕获部分可以是核酸,例如,包含DNA和/或RNA的核酸。核酸捕获部分可以结合例如衔接子锚定探针或桥接探针上的序列。在一些情况下,与固体表面

结合的抗RNA/DNA杂合抗体可用作捕获部分。

[0084] 标签和捕获部分可以通过一个或多个共价或非共价键结合。在固体支持物上捕获桥接探针、衔接子锚定探针或杂交复合物之后,可以洗涤固体支持物以从样本去除例如未结合的模板。在一些情况下,不进行洗涤步骤。洗涤可以是严格的或温和的。例如,当标签为生物素且捕获部分为链霉亲和素时,通过向样本添加游离生物素,可以洗脱捕获的与模板核酸分子杂交的桥接探针或衔接子锚定探针。

[0085] 延伸步骤(例如,退火至衔接子的衔接子引物的延伸)可以在桥接探针或衔接子锚定探针被捕获在固体支持物上时或在桥接探针(和杂交模板)或衔接子锚定探针(和间接杂交的模板)从固体支持物上洗脱之后进行。

[0086] 在模板、桥接探针和衔接子锚定探针杂交之后,可以使用链霉亲和素珠子进行清理,其中衔接子锚定探针的3'端被生物素化。杂交复合物和游离的衔接子锚定衔接子均可以与珠子结合。未结合的模板和桥接探针可以被冲离。第一和或第二桥接探针的5'端或3'端可以被生物素化。链霉亲和素珠子可用于去除未杂交的衔接子锚定衔接子和模板,这可以防止衔接子锚定探针和模板的随机连接。

IV. 模板核酸分子

[0087] 模板核酸可以是DNA或RNA。DNA可以是基因组DNA(gDNA)、线粒体DNA、病毒DNA、cDNA、cfDNA或合成DNA。DNA可以是双链DNA、单链DNA、片段化DNA或受损的DNA。RNA可以是mRNA、tRNA、rRNA、微小RNA、snRNA、piRNA、小非编码RNA、多聚核糖体RNA、内含子RNA、前mRNA、病毒RNA或无细胞RNA。

[0088] 模板核酸可以是天然存在的或合成的。模板核酸可以具有修饰的杂环碱基。修饰可以是甲基化的嘌呤或嘧啶、酰基化的嘌呤或嘧啶、烷基化的核糖或其他杂环。模板核酸可以具有修饰的糖部分。修饰的糖部分可以包括肽核酸。模板核酸可以包括肽核酸。模板核酸可以包括苏糖核酸。模板核酸可以包括锁核酸。模板核酸可以包括己糖醇核酸。模板核酸可以是柔性核酸。模板核酸可以包括甘油核酸。

[0089] 模板核酸分子可以从低输入(例如1ng的核酸材料)样本(例如无细胞DNA(cfDNA)和循环肿瘤DNA(ctDNA))捕获和富集。低输入样本可以具有1ng、2ng、3ng、4ng、5ng、6ng、7ng、8ng、9ng、10ng或更多的核酸材料。低输入样本可以具有少于10ng、9ng、8ng、7ng、6ng、5ng、4ng、3ng、2ng、1ng或更少的核酸材料。低输入样本可以具有200pg至10ng的核酸材料。低输入样本可以具有少于10ng的核酸材料。低输入样本可以具有少于10ng、5ng、1ng、100pg、50pg、25pg或更少的核酸材料。在一些情况下,输入样本可以具有1ng、10ng、20ng、30ng、40ng、50ng或更多的核酸分子。输入样本可以具有少于50ng、40ng、30ng、20ng、10ng、1ng或更少的核酸材料。捕获和富集可以通过靶探针杂交来完成。靶探针可以是捕获探针、桥接探针和/或衔接子锚定探针。靶探针可以包含一个或多个结合部分。结合部分可以是生物素。可以将结合部分附接至支持物。支持物可以是珠子。珠子可以是链霉亲和素珠子。

[0090] 模板核酸可以是受损的。受损的核酸可以包含改变的或缺失的碱基,和/或修饰的主链。模板核酸可以受到氧化、辐射或随机突变的损害。模板核酸可以受到亚硫酸氢盐处理的损害。

[0091] 对于受损的DNA,本公开可以消除双链DNA修复步骤,提供更高的转化率和改进的灵敏度,因为该过程中较少的步骤导致较少的DNA损失。

[0092] 受损的dsDNA(带有缺口)或ssDNA可用作文库构建的模板。对于受损的dsDNA,可以将dsDNA变性,因此可以使用至少一条未受损的链作为模板。然后可以将模板与捕获探针杂交和附接,并使用各种引物进行扩增。

[0093] 模板可以源自无细胞DNA(cfDNA)或循环肿瘤DNA(ctDNA)。cfDNA的来源可以是胎儿或肿瘤。模板可以源自对象的液体活检、固体活检或固定组织。模板可以是cDNA并且可以通过反转录生成。模板核酸可以源自流体样本,包括但不限于血浆、血清、痰液、唾液、尿液或汗液。可以对流体样本进行亚硫酸氢盐处理,以研究模板核酸的甲基化模式和/或确定模板核酸的组织来源。模板核酸可以源自肝脏、食道、肾脏、心脏、肺、脾、膀胱、结肠或脑。模板核酸可以用亚硫酸氢盐处理,以分析模板核酸来源器官的甲基化模式。对象可能患有甲基化相关疾病,例如自身免疫性疾病、心血管疾病、动脉粥样硬化、神经失常和癌症。

[0094] 模板核酸可以源自男性或女性对象。对象可以是婴儿。对象可以是青少年。对象可以是年轻的成年人。对象可以是老年人。

[0095] 模板核酸可以来源于人、大鼠、小鼠、其他动物或特定植物、细菌、藻类、病毒等。模板核酸可以来源于灵长类动物。灵长类动物可以是黑猩猩或大猩猩。其他动物可以是恒河猴。模板还可以来自不同物种(包括宿主-病原体、细菌种群等)的基因组混合物。模板可以由两个或更多个物种的基因组表达的RNA制成的cDNA。

[0096] 模板核酸可以包含靶序列。靶序列是外显子。靶序列可以是内含子。靶序列可以包含启动子。靶序列可以是先前已知的。靶序列可以是先前部分已知的。靶序列可以是先前未知的。靶序列可以包含染色体、染色体臂或基因。基因可以是与病况(例如癌症)相关的基因。模板核酸分子可以在杂交之前去磷酸化以例如降低自连接率。

V. 桥接探针

[0097] 桥接探针可用于将模板核酸分子与靶序列和衔接子锚定探针杂交。桥接探针可以进一步使得衔接子锚定探针和模板间接缔合,从而促进它们的附接。由于相互作用的随机性,游离的衔接子锚定探针和模板的连接率可能非常低。但与游离的衔接子锚定探针相比,杂交的桥接探针可以增加衔接子锚定探针和模板之间的连接概率。桥接探针可以包含DNA。桥接探针可以包含RNA。桥接探针可以包含尿嘧啶和甲基化胞嘧啶。桥接探针可以不包含尿嘧啶。

[0098] 桥接探针可以包含与靶序列杂交的靶特异性区域(TSR)。桥接探针可以包含与衔接子锚定探针的桥接结合序列杂交的衔接子着陆序列(ALS)。桥接探针可以包含连接TSR和ALS的接头。TSR可以位于桥接探针的3'-部分。TSR可以位于桥接探针的5'-部分。

[0099] 桥接探针可以包含一个或多个分子条形码。桥接探针可以包含一个或多个结合部分。结合部分可以是生物素。可以将结合部分附接至支持物。支持物可以是珠子。珠子可以是链霉亲和素珠子。

[0100] 桥接探针可以包含约400个核苷酸、约300个核苷酸、约200个核苷酸、约120个核苷酸、约100个核苷酸、约90个核苷酸、约80个、约70个核苷酸、约50个核苷酸、约40个核苷酸、约30个核苷酸、约20个核苷酸、或约10个核苷酸。

[0101] 多个桥接探针可用于退火至样本中的多个靶序列。桥接探针可以设计成具有相似的解链温度。一组桥接探针的解链温度可以在约15℃内、约10℃内、约5℃内或约2℃内。一种或多种桥接探针的解链温度可以是约75℃、约70℃、约65℃、约60℃、约55℃、约50℃、约

45℃或约40℃。桥接探针的解链温度可以是约40℃至约75℃、约45℃至约70℃、45℃至约60℃、或约52℃至约58℃。

[0102] 衔接子锚定探针与围绕特定桥接探针的一个或多个桥接探针一起使用可以通过协同效应有助于稳定特定桥接探针与其靶序列的杂交。形成多桥接探针组装体的杂交温度可以高于单个桥接探针的解链温度。较高的温度可以通过减少可能在较低温度下发生的非特异性杂交来产生更好的捕获特异性。杂交温度可以比单个桥接探针的解链温度高约5℃、约10℃、约15℃或约20℃。杂交温度可以比桥接探针的解链温度高约5℃至约20℃,或比多个桥接探针的平均解链温度高约5℃至约20℃。

[0103] 多个桥接探针的杂交温度可以是约75℃、约70℃、约65℃、约60℃、约55℃或约50℃。多个桥接探针的杂交温度可以是约50℃至约75℃、55℃至约75℃、60℃至约75℃、或65℃至约75℃。

[0104] 桥接探针可以进一步包含标签。标签可以是荧光的。荧光标签可以是有机荧光染料、金属螯合物、碳纳米管、量子点、金颗粒或荧光矿物。标签可以是放射性的。标签可以是生物素。桥接探针可以与标记的核酸结合分子结合。核酸结合分子可以是抗体、抗生素、组蛋白、抗体或核酸酶。

[0105] 桥接探针可以包含接头。接头可以包含约30个核苷酸、约25个核苷酸、约20个核苷酸、约15个核苷酸、约10个核苷酸或约5个核苷酸。接头可以包含约5个至约20个核苷酸。

[0106] 接头可以包含非核酸聚合物(例如,碳串)。接头非核苷酸聚合物可以包含约30个单元、约25个单元、约20个单元、约15个单元、约10个单元或约5个单元。

[0107] 桥接探针可以在3'端和/或5'端被阻断。桥接探针可以缺少5'磷酸盐。桥接探针可以缺少3'OH。桥接探针可以包含3'ddC、3'倒置dT、3'C3间隔子、3'氨基或3'磷酸化。

VI. 衔接子锚定探针

[0108] 衔接子锚定探针或通用锚定探针可以包含一个或多个桥接结合序列,其与一个或多个桥接探针的衔接子着陆序列杂交。

[0109] 衔接子锚定探针可以包含在BBS之间的间隔子。一个或多个间隔子的存在可以提高杂交捕获的效率并增加捕获的特异性。

[0110] 衔接子锚定探针可以包含分子条形码(MB)。衔接子锚定探针可以包含一个或多个桥接探针可以与之杂交的桥接结合序列(BBS)。衔接子锚定探针可以包含1至100个BBS。衔接子锚定探针可以包含用于区分样本的索引。分子条形码或索引可以是衔接子序列的5'和BBS的5'。

[0111] 衔接子锚定探针可以包含约400个核苷酸、约200个核苷酸、约120个核苷酸、约100个核苷酸、约90个核苷酸、约80个核苷酸、约70个核苷酸、约50个核苷酸、约40个核苷酸、约30个核苷酸、约20个核苷酸、或约10个核苷酸。衔接子锚定探针可以是约20个至约70个核苷酸。

[0112] 衔接子锚定探针对桥接探针的解链温度可以是约65℃、约60℃、约55℃、约50℃、约45℃,或约45℃至约70℃。

[0113] 衔接子锚定探针可以包含标签。标签可以是荧光的。荧光标签可以是有机荧光染料、金属螯合物、碳纳米管、量子点、金颗粒或荧光矿物。标签可以是放射性的。标签可以是生物素。衔接子锚定探针可以与标记的核酸结合分子结合。核酸结合分子可以是抗体、抗生

素、组蛋白、抗体或核酸酶。

VII. 衔接子/衔接子引物

[0114] 一个或多个衔接子可以附接至多个模板核酸用于构建文库。该文库可以是二代测序(NGS)文库。可以将一个衔接子附接至模板核酸分子的5'端或3'端。可以将两个衔接子附接至模板核酸分子的5'端和3'端。一个或多个衔接子可以通过连接附接至模板核酸。一个或多个衔接子的附接可以在模板核酸和靶探针杂交之前进行。在一些情况下,可以在杂交之后将衔接子添加到捕获的模板核酸中。一个或多个衔接子可以包含分子条形码(MB)。

[0115] 一个或多个衔接子引物可以与附接至模板核酸分子的一个或多个衔接子杂交。在一些情况下,衔接子被掺入衔接子锚定探针或捕获探针。在某些情况下,附接、添加或掺入的衔接子可以为引物杂交提供位点以进行扩增。第一衔接子(AD1)可以通过捕获探针或衔接子锚定探针附接至模板。针对AD1的引物可用于合成与模板互补的链。第二衔接子(AD2)可以附接至模板的5'端和/或互补链的3'端,以进一步扩增模板。可以使用AD1引物和AD2引物构建文库。可以使用AD1引物和针对TSR或其侧翼区域的引物进行选择性的扩增。

[0116] 衔接子可以是单链核酸。衔接子可以是双链核酸。衔接子可以是部分双链体,长链比短链长,或两条链长度相等。

VIII. 酶

[0117] 可用于本文所述的方法和试剂盒的DNA聚合酶的实例包括Klenow聚合酶、Bst DNA聚合酶、Bca聚合酶、phi 29DNA聚合酶、Vent聚合酶、Deep Vent聚合酶、Taq聚合酶、T4聚合酶、T7聚合酶或大肠杆菌DNA聚合酶I。

[0118] 可用于本文所述的方法和试剂盒的连接酶的实例包括CircLigase、CircLigase II、大肠杆菌DNA连接酶、T3 DNA连接酶、T4 DNA连接酶、T7 DNA连接酶、DNA连接酶I、DNA连接酶II、DNA连接酶III、DNA连接酶IV、Taq DNA连接酶或Tth DNA连接酶。

[0119] 可用于本文所述的方法和试剂盒的甲基化敏感酶或甲基化依赖性限制酶的实例包括Aat II、Acc II、Aor13H I、Aor51H I、BspT104 I、BssH II、Cfr10 I、Cla I、Cpo I、Eco52 I、Hae II、Hap II、Hha I、Mlu I、Nae I、Not I、Nru I、Nsb I、PmaC I、Psp1406 I、Pvu I、Sac II、Sal I、Sma I和SnaB I。

IX. 扩增产物的下游分析

[0120] 使用本文所述的方法产生的扩增产物可以使用各种方法进一步分析,所述方法包括Southern印迹、聚合酶链反应(PCR)(例如,实时PCR(RT-PCR)、数字PCR(dPCR)、微滴数字PCR(ddPCR)、定量PCR(Q-PCR)、nCounter分析(纳米串技术)、凝胶电泳、DNA微阵列、质谱(例如,串联质谱、基质辅助激光解吸电离飞行时间质谱(MALDI)-TOF MS)、链终止测序(Sanger测序)或二代测序。

[0121] 二代测序可以包含454测序(ROCHE)(使用焦磷酸测序)、使用可逆终止染料的测序(ILLUMINA测序)、半导体测序(THERMOFISHER ION TORRENT)、单分子实时(SMRT)测序(PACIFIC BIOSCIENCES)、纳米孔测序(例如,使用来自OXFORD NANOPORE或GENIA的技术)、使用焦磷酸解作用(pyrophosphorolysis)(BASE4)的微滴单分子测序、单分子电子检测测序(例如,在核酸(DNA/RNA)通过纳米间隙时测量通过纳米电极的隧道电流并计算电流差(来自QUANTUM BIOSYSTEMS的QUANTUM测序))、GenapSys基因电子纳米集成超灵敏(GENIUS)技术(GENAPYS)、来自QIAGEN的GENEREADER、使用由特定荧光团识别的中心确定碱基(或碱基

对)的部分随机寡核苷酸的顺序杂交和连接进行测序(SOLiD测序)。测序可以是双端测序。

[0122] 可以使用本文所述的方法测序的来自样本的靶序列的数量可以是约5、10、15、25、50、100、1000、10,000、100,000或1,000,000个,或约5至约100个、约100至约1000个、约1000至约10,000个、约10,000至约100,000个、或约100,000至约1,000,000个。

[0123] 使用本文所述的方法生成的核酸文库可以从多于一个样本产生。每个文库可以具有与样本缔合的不同索引。例如,捕获探针或衔接子锚定探针可以包含可用于将核酸识别为来自同一样本的索引(例如,包含相同第一索引的第一组捕获探针或衔接子锚定探针可用于从来自第一对象的第一样本生成第一文库,以及包含相同第二索引的第二组捕获探针或衔接子锚定探针可用于从来自第二对象的第二样本生成第二文库,第一文库和第二文库可以合并、测序,并且索引可用于辨别经测序的核酸源自哪个样本)。使用本文所述的方法生成的扩增产物可用于从至少2、5、10、25、50、100、1000或10,000个样本生成文库,每个文库具有不同的索引,并且文库可以合并和测序,例如,使用二代测序技术。

[0124] 测序可以生成至少100、1000、5000、10,000、100,000、1,000,000或10,000,000个序列读段。测序可以生成约100个序列读段至约1000个序列读段、约1000个序列读段至约10,000个序列读段、约10,000个序列读段至约100,000个序列读段、约100,000个序列读段至约1,000,000个序列读段、或约1,000,000个序列读段至约10,000,000个序列读段。

[0125] 测序深度可以是约1x、5x、10x、50x、100x、1000x或10,000x。测序深度可以为约1x至约10x、约10x至约100x、约100x至约1000x、或约1000x至约10000x。

X. 生物信息学分析

本文提供了用于测序数据的生物信息学分析的方法。例如,排除亚硫酸氢盐转化不完全的分子的方法,以及分析疾病分子含量非常低的样本中的甲基化模式的方法。

a. 排除亚硫酸氢盐转化不完全的分子

[0126] 排除具有不完全C>T转化的分子的过滤技术用于增强分子计数和甲基化分数数据的稳健性。

[0127] 映射到每个差异甲基化区域(DMR)的测序读段可以使用基因组中的读段开始和结束核苷酸位置以及唯一分子标识符信息进行去重。也可以单独使用开始和结束位置信息以较低的准确度完成去重。

[0128] 根据CH背景中未转换的C的数量过滤去重的读段,其中C代表胞嘧啶,并且H代表三种核苷酸中的任何一种:C(胞嘧啶)、A(腺嘌呤)或T(胸腺嘧啶)。在CH背景下存在未转化为T的C表明分子的亚硫酸氢盐或酶处理不完全的可能性很高。当在CH背景下未转换的C的数量大于预设阈值时,弃去读段。在一些情况下,在CH背景下未转换的C的阈值数量为1、2、3、4、5、6、7、8、9或10。在一些情况下,如果在CH背景下未转换的C的百分比(作为在CH背景下C总数的百分比)大于1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、12%、14%、16%、18%、20%、25%、30%、35%、40%或50%,可弃去读段。

b. SICON TMS分析

[0129] 当前分析甲基化测序数据的方法可以涉及计算用于下游分析的两个指标中的一个或两个:(1)单个CpG位点的甲基化分数;(2)目的基因组区域的甲基化密度。对于(1),CpG位点上甲基化C的数量可以除以覆盖该CpG位点的分子总数。对于(2),可以计算定义的基因组区域中CpG位点的所有甲基化分数的平均值。作为对上述概念的轻微修改,可以引入甲基

化单倍型负载 (MHL) 以努力考虑区域分子中甲基化模式的差异。从本质上讲, MHL 代表对分子混合物的平均测量, 加上权重以说明单元长度。这些方法对所有测序分子中的 DNA 分子进行平均测量, 包括疾病来源材料和健康正常来源材料。

[0130] 在组织测序数据中, 对所有分子取平均值通常是充分且必要的方法。例如, 在肿瘤活检组织的情况下, 肿瘤含量可能适度偏高 (例如 20% 或更多)。肿瘤和正常组织之间的甲基化水平的显著差异可以反映在肿瘤-正常混合组织的平均值和纯正常组织的平均值上。由于大多数亚硫酸氢盐测序数据在每个基因组区域的复杂性较低, 因此通常不需要进行平均。例如, 30x 可能被认为是在全基因组亚硫酸氢盐测序中的深度覆盖, 而许多研究的覆盖度要低得多。该区域中许多 CpG 位点的平均值消除了由于覆盖度低导致的变异性, 并可能增强测量的稳健性。在疾病分子含量非常低的样本中, 例如使用来自肿瘤患者的血浆 cfDNA 进行的液体活检, 在肿瘤含量通常低于 0.1% 的情况下, 健康正常分子和疾病来源分子的混合物的平均值可能以正常分子为主。换言之, 肿瘤来源的甲基化信息在取平均值的作用下被正常来源的分子所掩盖。

[0131] 一种分析甲基化测序数据的方法在此描述为“SICON TMS 分析”。简言之, 对每个测序分子上 CpG 位点的数量进行计数, 并计算这些位点的甲基化分数。由 CpG 计数和甲基化分数组成的数据对代表下游分类模型中的一个数据点。与基于平均值的方法相比, 未对来自疾病来源分子和正常来源分子的甲基化信息取平均值。因此, 疾病来源分子和正常细胞来源分子的甲基化谱可以保持分离。每个生成的读段可以包含来自通过测定捕获的独特 DNA 分子的 CpG 甲基化信息。从每个读段收集两个指标:

1) N: 读段中的 CpG 总数;

2) M: 读段中的甲基化 CpG 的数目。

根据 1) 和 2), 第三个指标计算如下:

3) $f = M/N$, 在当前读段中甲基化的 CpG 分数。

[0132] 针对测定中所有 DMR 上的每个分子收集数据对 (N, f)。可以针对 DMR 生成显示 f (y 轴) 与 N (x 轴) 的散点图, DMR 中的每个读段在图中显示为点。例如, 图 11 显示了 DMR1 在正常结肠组织 (图 11A) 和结肠癌组织基因组 DNA (图 11B) 中的分子甲基化散射模式。它展示了一种 DMR, 其中正常结肠组织中没有超甲基化 DNA 分子, 而结肠癌组织中存在大量超甲基化分子。图 12A 和 12B 分别显示了 DMR2 在正常结肠组织和结肠癌组织基因组 DNA 中的分子甲基化散射模式。它展示了一种 DMR, 其中正常结肠组织中存在一些超甲基化 DNA 分子 (图 12A), 而结肠癌组织中存在大量超甲基化分子 (图 12B)。图 13 显示了来自健康个体 (图 13A) 和结肠癌患者 (图 13B) 的血浆 cfDNA 中 DMR1 和 DMR2 的分子甲基化散射模式。图 13B 上部所示的来自每个 DMR 的超甲基化分子的计数是从液体活检中检测疾病的基础。

[0133] 可以进行一些进一步的分析。例如, 可以应用过滤器来对超甲基化分子进行计数。过滤超甲基化分子: 可以选择阈值 f_0 来计数 $f > f_0$ 的所有分子 (即在散点图的上部)。这些读段是超甲基化读段, 是疾病组织 (例如结肠癌) 的特征。超甲基化过滤阈值 (f_0) 可以设置为 0.2、0.3、0.4、0.5、0.6、0.7、0.8 或 0.9。在一些情况下, 超甲基化过滤阈值 (f_0) 可以基于对正常组织或来自健康对象的样本中的甲基化的分析来设置。例如, 超甲基化过滤阈值 (f_0) 可以设置为与正常组织样本或来自健康对象的样本中的平均甲基化分数的 0.5、1、1.5、2、2.5 或 3 个标准差。

[0134] 还可以过滤分子以用于稳健的信号。过滤具有稳健信号的分子：可以选择另外的阈值N0以仅保留N>N0的读段，以增强分子计数的稳健性。阈值N0可以设置为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或30。

[0135] 过滤超甲基化分子和稳健信号可以确保对于每个DMR仅对稳健的超甲基化分子进行计数。这可以提高分析质量和/或灵敏度。

[0136] 在一些情况下，阈值f0和N0在所有DMR中均是相同的。在一些情况下，可以针对每个单独的DMR自定义阈值f0和N0。在一些情况下，阈值f0可以在所有DMR中均相同，并且可以针对每个单独的DMR自定义阈值N0。在一些情况下，阈值N0可以在所有DMR中均相同，并且可以针对每个单独的DMR自定义阈值f0。在一些情况下，可以针对每个单独的DMR自定义阈值f0和N0。

[0137] 可以将测定中所有DMR的稳健超甲基化分子计数输入模型，以使用机器学习分类器方法确定样本的疾病状态。

XI. 顺序靶标富集

本公开提供了一种基于顺序杂交的富集方法，该方法可用于从相同的DNA输入富集两个或更多个基因包的序列而不进行分割。图14图示了进行顺序富集的方法。在一些情况下，顺序富集的方法可以包括获得包含多个核酸分子的样本并且进行第一靶标富集以富集包含对应于一个或多个基因组区域的第一基因包的序列的核酸分子，从而生成第一富集样本，其包含富集有对应于一个或多个基因组区域的第一基因包的序列的核酸。第一靶标富集还可以生成剩余样本（或第一剩余样本），其包含耗竭了对应于一个或多个基因组区域的第一基因包的序列的核酸。该剩余样本可用于对剩余样本进行第二靶标富集以富集包含对应于一个或多个基因组区域的第二基因包的序列的核酸分子，从而生成第二富集样本，其包含富集有对应于一个或多个基因组区域的第二基因包的序列的核酸。一个或多个基因组区域的第一基因包和一个或多个基因组区域的第二基因包通常不同。在一些情况下，可以用基因组区域的第三基因包、第四基因包或更多基因包进行第三轮、第四轮或更多轮的靶标富集。

[0138] 例如，一个或多个基因组区域的基因包可以包含与突变热点、癌基因、肿瘤抑制基因、癌基因外显子、肿瘤抑制外显子或调控区域相关的1-50,000个、5-10000个或5-5000个基因组区域的基因包。在另一个实例中，一个或多个基因组区域的基因包可以包含与差异甲基化区域、表观遗传修饰、内含子、启动子或其他调控序列相关的5-5000个基因组区域的基因包。在一些情况下，基因包包含50-500个与癌症中的超甲基化相关的基因组区域。

[0139] 因为Point-n-Seq是扩增前和转化前富集技术，富集的样本可以通过测序进行分析，或者可以在测序之前进行二硫化物处理（或酶处理）以评估甲基化。在一些情况下，可以通过测序来分析第一富集样本以评估突变，而在测序之前对第二富集样本进行二硫化物（或酶促）处理以评估甲基化。在一些情况下，第一富集样本和第二富集样本均通过直接测序进行评估，以获取基因组改变，但是可以以不同深度对样本进行测序。在一些情况下，在进行第二靶标富集步骤之前可以对第一富集样本进行分析。第一富集样本的分析结果可用于为第二富集步骤选择第二基因包。

[0140] 靶标富集可以包括本文公开的或本领域已知的任何方法。在一些情况下，靶标富集包括将第一桥接探针的第一靶特异性区域与具有对应于基因组区域的序列的分子的第

一靶序列杂交,其中第一桥接探针的第一衔接子着陆序列与衔接子锚定探针的第一桥接结合序列结合;并且将第二桥接探针的第二靶特异性区域与具有对应于基因组区域的序列的分子的第二靶序列杂交,其中第二桥接探针的第二衔接子着陆序列与衔接子锚定探针的第二桥接结合序列结合。如本文所述,锚定探针可以包含结合部分。该方法通常包括将衔接子附接至多个核酸分子的核酸分子的5'端或3'端,从而生成包含衔接子的核酸分子文库。

[0141] 本文所述的顺序靶标富集可以是高效的。例如,当第二富集样本经过亚硫酸氢盐处理并经受测序反应时,该测序反应的信息读段的数量可以是可以从所述样本在经受单一靶标富集以富集包含对应于一个或多个基因组区域的第二基因包的序列的核酸分子的情况下所获得的信息读段的数量的至少60%、65%、70%、75%、80%或85%。

[0142] 本文所述的顺序靶标富集方法可以推广到任何核酸样本。这些方法对于分析有限的核酸样本可能特别有用。

XII. 应用

a. 核酸特征检测

[0143] 可以分析使用本文所述的方法和试剂盒生成的扩增的核酸产物的一种或多种核酸特征。一种或多种核酸特征可以是一个或多个甲基化事件。甲基化可以是CpG二核苷酸中胞嘧啶的甲基化。甲基化的碱基可以是5-甲基胞嘧啶。非CpG背景中的胞嘧啶可以是甲基化的。甲基化或未甲基化的胞嘧啶可以在CpG岛中。CpG岛可以是具有高频率CpG位点的基因组区域。CpG岛可以是至少200bp,或约300至约3000bp。CpG岛可以是至少60%的CpG二核苷酸含量。CpG岛可以在基因的启动子区域中。甲基化可以是5-hmC (5-羟甲基胞嘧啶)、5-fC (5-甲酰基胞嘧啶)或5-caC (5-羧基胞嘧啶)。本文所述的方法和试剂盒可用于检测甲基化模式,例如来自实体组织或来自生物流体(例如,包括例如无细胞DNA的血浆、血清、尿液或唾液)的DNA的甲基化模式。

[0144] 一种或多种核酸特征可以是新生突变、无义突变、错义突变、沉默突变、移码突变、插入、置换、点突变、单核苷酸多态性(SNP)、单核苷酸变异(SNV)、新生单核苷酸变异、缺失、重排、扩增、染色体易位、中间缺失、染色体倒位、杂合性丢失、功能丧失、功能获得、显性阴性或致死突变。可以分析扩增的核酸产物以检测种系突变或体细胞突变。一种或多种核酸特征可以与病况(例如癌症、自身免疫疾病、神经系统疾病、感染(例如病毒感染)或代谢疾病)相关。

b. 诊断/检测/监测

[0145] 公开的方法和试剂盒还可用于诊断或检测疾病或病况。该疾病或病况可能与甲基化异常有关。该病况可以是心理障碍。病况可能老龄化。该病况可能是疾病。病况(例如疾病)可以是癌症、神经系统疾病(例如阿尔茨海默病、自闭症谱系障碍、雷特综合征、精神分裂症)、免疫缺陷、皮肤病、自身免疫性疾病(例如,眼白塞病、系统性红斑狼疮(SLE)、类风湿性关节炎(RA)、多发性硬化症、感染(例如病毒感染)或代谢疾病(例如高血糖、高脂血症、2型糖尿病)。癌症可以是例如结肠癌、乳腺癌、肝癌、膀胱癌、Wilms癌、卵巢癌、食管癌、前列腺癌、骨癌或肝细胞癌、胶质母细胞瘤、乳腺癌、鳞状细胞肺癌、甲状腺癌或白血病(参见例如,Jin和Liu(2018)DNA methylation in human disease.Genes&Diseases,5:1-8)。该病况可以是Beckwith-Wiedemann综合征、Prader-Willi综合征或Angelman综合征。

[0146] 使用本文提供的方法和试剂盒生成的无细胞DNA的甲基化模式可用作癌症的标志

物(参见例如,Hao等人,DNA methylation markers for diagnosis and prognosis of common cancers.Proc.Natl.Acad.Sci.2017;国际PCT申请公开号W02015116837)。无细胞DNA的甲基化模式可用于确定DNA来源的组织(参见例如国际PCT申请公开号W02005019477)。本文所述的方法和试剂盒可用于确定甲基化单倍型信息,并可用于确定无细胞DNA的组织或细胞来源(参见例如Seioighe等人,(2018) DNA methylation haplotypes as cancer markers.Nature Genetics 50,1062-1063;国际PCT申请公开号W02015116837;美国专利申请公开号20170121767)。本文所述的方法和试剂盒可用于检测患有癌症的对象和未患有癌症的对象中的甲基化水平,例如,无细胞DNA的甲基化水平(参见例如,Vidal等人,A DNA methylation map of human cancer at single base-pair resolution.Oncogenomics 36,5648-5657;国际PCT申请公开号W02014043763)。本文所述的方法和试剂盒可用于确定甲基化水平或确定不同组织对无细胞DNA混合物的贡献比例(参见例如国际PCT申请公开号W02016008451)。本文所述的方法和试剂盒可用于无细胞DNA来源的组织,例如血浆中,例如基于甲基化单倍型的模式和丰度的比较(参见例如Tang等人,(2018) Tumor origin detection with organization-specific miRNA and DNAmethylation marker.Bioinformatics 34,398-406;国际PCT申请公开号W02018119216)。本文所述的方法和试剂盒可用于区分癌细胞和正常细胞,并根据其起源组织对不同的癌症类型进行分类(参见例如美国专利申请公开号20170175205A1)。本文提供的方法和试剂盒可用于使用母体样本检测胎儿DNA或胎儿异常(参见例如,Poon等人(2002) Differential DNA Methylation between Fetus and Mother as a Strategy for Detecting Fetal DNA in Maternal Plasma.Clinical Chemistry,48:35-41)。

[0147] 公开的方法可用于监测病况。病况可以是疾病。疾病可以是癌症、神经系统疾病(例如,阿尔茨海默病)、免疫缺陷、皮肤病、自身免疫性疾病(例如,眼白塞病)、感染(例如病毒感染)或代谢疾病。癌症可以缓解。由于公开的方法可以使用cfDNA和ctDNA来检测低水平的异常,因此本公开可以提供相对无创的疾病监测方法。公开的方法可用于监测治疗或疗法。治疗或疗法可用于病况,例如疾病,例如癌症,或用于本文公开的任何病况。

本文所述的方法可以实现在亚硫酸氢盐转化和扩增之前直接从cfDNA富集靶分子。这些方法还可以开发小型聚焦基因包,用于查询给定疾病的1至~1000个标志物的甲基化状态。在一些情况下,可以针对查询给定疾病的1至约10000个差异甲基化区域的甲基化状态的基因包生产试剂盒。

实施例

实施例1

协同间接杂交捕获

[0148] 使用具有不同序列的两个桥接探针和衔接子锚定探针/通用锚定探针(UP,SEQ ID NO:1)进行用于测序的协同间接捕获核酸(SICON-SEQ)实验。设计两个桥接探针(EGFR-BP2,SEQ ID NO:2和EGFR-BP3,SEQ ID NO:3)以靶向EGFR基因组序列。每个桥接探针包含约25bp的靶向序列(TS1或TS2)区域、包含至少15个胸腺嘧啶的接头和设计成在衔接子锚定探针上与桥接结合序列互补的20bp着陆序列(LS1或LS2,斜体)。衔接子锚定探针包含两个桥接结合序列(BBS1或BBS2),它们被设计成与桥接探针的任一着陆序列杂交。衔接子锚定探针在核酸序列的5'进一步生物素化。图4提供了协同间接杂交的示意图。

表1.序列表

SEQ ID NO.	ID	类型	序列
1	UP	衔接子锚定探针	5'-TTTTTTTTTTTGGCACCAGACTTAATCTAA GCAGAGAACATGATAAGAGA-3'
2	EGFR-B P2	桥接探针	3'-TTAGATTAAAGTCTGGTGCCATTTTTTTTTTTTTTTTC AAGGAATTAAGAGAAGCAACATC-5'
3	EGFR-B P3	桥接探针	3'-TCTCTTATCATGTTCTCTGCTTTTTTTTTTTTTTTGA AAGCCAACAAGGAAATCCTCGAT-5'
4	EGFR Fw	EGFR 正向 引物	5'-CCCGTCGCTATCAAGGAATTAAGA-3'
5	EGFR Rev	EGFR 反向 引物	5'-CCACACAGCAsAAGCAGAAACTCAC-3'

[0149] 对于杂交捕获,将20ng片段化(峰大小160bp) DNA与针对EGFR的两个桥接探针(各1fmole)以及一个通用锚定探针(200fmole)混合,最终溶液体积为20 μ l。DNA输入和杂交探针在杂交缓冲液中于95℃下变性30min,然后使其逐渐冷却至65℃。杂交复合物在热循环仪上于65℃下温育1小时。最终的杂交缓冲液包含100ng/ μ l封闭DNA、1 μ g/ μ l牛血清白蛋白(BSA)、1 μ g/ μ l Ficoll、1 μ g/ μ l聚乙烯吡咯烷酮(PVP)、0.075M柠檬酸钠、0.75M NaCl、5x SSC和1X Denhardt's溶液。

[0150] 为了捕获/清理,杂交组装体与链霉亲和素珠子(Thermo Fisher Dynabeads M270 链霉亲和素)在室温下温育10min。用三次洗涤进行清理(洗涤1:5X SSPE,1% SDS;洗涤2:2X SSPE,0.1% SDS;洗涤3:0.1X SSPE,0.01% triton)。

[0151] 使用针对EGFR靶向序列的引物(SEQ ID NO.4和5)通过qPCR评估富集的DNA。将捕获的EGFR DNA的qPCR结果与没有捕获富集的gDNA的相同部分进行比较。65%至超过90%的EGFR被回收。

实施例2

通过不同杂交方案捕获

[0152] 为了确定各种杂交系统的捕获性能,测试了四种杂交方案:非协同直接杂交(图5A)、协同直接杂交(图5B)、协同间接杂交(图5C)和非协同间接杂交(图5D)。

[0153] 非协同直接方法涉及包含靶特异性序列(阴影线,图5A)的生物素化捕获探针(120bp,SEQ ID NO.6)的杂交。协同直接方法涉及四个短生物素化捕获探针(SEQ ID NOS.7-10)的杂交,每个探针包含25bp的靶特异性序列(阴影线,图5B)。协同间接方法使用四个不含生物素的短桥接探针(SEQ ID NOS.12-15)(图5C),并且每个包含与协同直接方法中使用的捕获探针之一相同的靶特异性序列。每个桥接探针(BP)包含两个不同的着陆序列(虚线和垂直阴影线)中的一个,其被设计成与通用锚定探针(SEQ ID NO.11)中的桥接结合序列中的一个互补。通过使用与用于协同直接杂交的相同通用锚定探针配对的短桥接探针(SEQ ID NO.16)来测试非协同间接方法(图5D)。实验中使用的捕获探针或通用锚定探针(UP)在5'端被生物素化。

表2.序列表

	SEQ ID NO.	ID	序列
非协同，直接	6	EGFR-bio	生物素- AGAAGGTGAGAAAGTTAAAATTCCCGTCGCTAT CAAGGAATTAAGAGAAGCAACATCTCCGAAAGC CAACAAGGAAATCCTCGATGTGAGTTTCTGCTTT
			GCTGTGTGGGGGTCCATGGC
协同，直接	7	EGFR-bioP1	生物素 -TTTTTTTTTTGGTGAGAAAGTTAAAATTCCCGTC G
	8	EGFR-bioP2	生物素 -TTTTTTTTTTTCAAGGAATTAAGAGAAGCAACAT C
	9	EGFR-bioP3	生物素 -TTTTTTTTTTGAAAGCCAACAAGGAAATCCTCGA T
	10	EGFR-bioP4	生物素 -TTTTTTTTTTAGTTTCTGCTTTGCTGTGTGGGGGT
协同，间接	11	UP	生物素 -TTTTTTGGCACCAGACTTAATCTAATTTGCAGAG AACATGATAAGAGATTTTGGCACCAGACTTAATC TAATTTGCAGAGAACATGATAAGAGA
	12	EGFR-BP1	TCTCTTATCATGTTCTCTGCTTTTTTTTTTTTTTG GTGAGAAAGTTAAAATTCCCGTCG
	13	EGFR-BP2	TTAGATTAAGTCTGGTGCCATTTTTTTTTTTTTTT CAAGGAATTAAGAGAAGCAACATC
	14	EGFR-BP3	TCTCTTATCATGTTCTCTGCTTTTTTTTTTTTTTG AAAGCCAACAAGGAAATCCTCGAT
	15	EGFR-BP4	TTAGATTAAGTCTGGTGCCATTTTTTTTTTTTTTT AGTTTCTGCTTTGCTGTGTGGGGGT
非协同，间接	11	UP	生物素 -TTTTTTGGCACCAGACTTAATCTAATTTGCAGAG AACATGATAAGAGATTTTGGCACCAGACTTAATC TAATTTGCAGAGAACATGATAAGAGA
	16	EGFR-BP2	TTAGATTAAGTCTGGTGCCATTTTTTTTTTTTTTT CAAGGAATTAAGAGAAGCAACATC
	17	EGFR Fw	CCCGTCGCTATCAAGGAATTAAGA
	18	P7 引物	CAAGCAGAAGAC GGCATACGAGAT

	19	P5 引物	AATGATACGGCGACCACCGA
--	----	-------	----------------------

[0154] 在杂交反应之前,按照随附方案中的步骤,使用NEBNext^{ultra} II DNA文库制备试剂盒使用10ng的cfDNA构建NGS文库。文库构建之后,使用连接混合物无需珠子纯化直接进行基于杂交的捕获,以富集文库。然后对富集的文库进行qPCR分析。

[0155] 通过比较捕获前后EGFR存在的百分比来评价捕获效率。将捕获之后的ct与2.5ng的人gDNA文库(捕获输入的适当部分)进行比较。通过使用针对EGFR (SEQ ID NO.17) 和NGS衔接子P7序列 (SEQ ID NO.18) 设计的引物进行捕获效率PCR。通过qPCR使用可以扩增所有DNA文库的引物 (SEQ ID NOS.18,19) 评价背景(总DNA存在)。所有背景delta ct均归一化为从“C”探针设计中获得的平均CT。

[0156] 间接协同杂交捕获显示出优于任何非协同方法和直接方法的杂交灵敏度和特异性(表3)。协同间接探针设计展示了最高捕获效率(平均为~91%)和最低背景噪声。在高得多的(300x)桥接探针浓度下,非协同直接杂交显示未达到14.87%的回收率,但背景增加超过200倍。降低杂交温度对捕获效率没有帮助,反而显著增加背景噪声。对于协同但非间接的设计,桥接探针浓度的增加和杂交的降低均无助于捕获效率。对于间接非协同方法,未检测到捕获富集。

表3. 各种杂交方案的捕获性能。

探针浓度	50μl 中 10 fmole 探针		50μl 中 3 pmole 探针		50μl 中 3 pmol 探针	
	60°C 杂交		60°C 杂交		55°C 杂交	
	捕获效率	背景	捕获效率	背景	捕获效率	背景
非协同直	N/D	1.4X	14.87%	256.0X	N/D	128.0X

接	1.0%	1.4X	9.81%	294.1X	1.27%	137.2X
协同直接	N/D	1.4X	N/D	1.3X	N/D	1.2X
	0.6%	1.3X	0.70%	1.1X	1.03%	1.1X
协同间接	94.0%	0.9X				
	76.3%	1.1X				
	90.1%	0.9X				
	84.1%	1.1X				
	107.2%	1.0X				
	100.0%	1.0X				
非协同间接	0.0%	1.1X				
	0.1%	1.1X				

实施例3

通过具有或不具有间隔子的通用锚定探针间接捕获

[0157] 进行研究以了解在通用锚定探针 (UP) 上的两个或更多个桥接结合序列之间存在间隔子是否影响间接协同杂交捕获的捕获性能。在这两种情况下均使用相同的桥接探针。

[0158] 表4列出了使用的桥接探针和UP的序列。图6A显示了使用具有间隔子的UP的协同间接杂交的示意图。图6B显示了使用不具有间隔子的UP的协同间接杂交。

表4. 序列表

SEQ ID NO.	ID	着陆序列之间的间隔子	序列
20	UP-间隔子	是	生物素 -TTTTTTGGCACCAGACTTAATCTAATTTGCAGAGAACATGATAAGAGATTTTGGCACCAGACTTAATCTAATTTGCAGAGAACATGATAAGAGA
21	UP-无间隔子	否	生物素 -TTTTTTGGCACCAGACTTAATCTAAGCAGAGAACATGATAAGAGATGGCACCAGACTTAATCTAAGCAGAGAACATGATAAGAGA
22	EGFR-BP1	TCTCTTATCATGTTCTCTGCTTTTTTTTTTTTTTTGGTGAGAAAGTTAAA ATTCCCGTCG	
23	EGFR-BP2	TTAGATTAAAGTCTGGTGCCATTTTTTTTTTTTTTTCAAGGAATTAAGA GAAGCAACATC	
24	EGFR-BP3	TCTCTTATCATGTTCTCTGCTTTTTTTTTTTTTTTGAAAGCCAACAAGG AAATCCTCGAT	
25	EGFR-BP4	TTAGATTAAAGTCTGGTGCCATTTTTTTTTTTTTTTAGTTTCTGCTTTGCT GTGTGGGGGT	

[0159] 确定任一杂交捕获的捕获效率和背景噪声。通过将qPCR结果归一化为平均背景信号来计算背景噪声。捕获效率不受间隔子存在的很大影响,但没有间隔子的捕获杂交的背景噪声比有间隔子的捕获高约100倍(表5)。因此,这表明通用锚定探针中的间隔子在实现高度特异性(低背景)捕获方面发挥了重要作用。

表5. 使用具有或不具有间隔子的通用锚定探针的杂交捕获性能

	捕获效率	背景
UP-间隔子	75.8%	1.1X
	70.7%	0.9X
UP-无间隔子	66.0%	93.7X
	66.0%	107.6X

实施例4

使用协同间接捕获方法确定NGS指标

[0160] 确定使用3、15和76个靶标基因包的二代测序(NGS)指标。映射率计算为与人类基因组比对的测序读段的百分比。3、15和76个靶标基因包的映射率分别为97%、94%、95%(表6)。使用捕获探针和100bp侧翼覆盖的区域上的去重(dedupe)映射读段计算中靶率。对

于例如3、15和76个靶标的小型基因包,传统的基于杂交的DNA富集是不可行的。然而,该研究显示,与超过50kb的标准靶标基因包相比,15个和76个靶标基因包的中靶率相当高,分别为83.6%和85.3%。

[0161] 此外,基因包的均匀性很高(>99%的位置读段高于平均覆盖度的0.2x,并且超过95%为0.5x覆盖度)。0.2X或0.5X覆盖度不适合具有3个靶标的微型基因包。15个靶标基因包的高均匀性也反映在GC含量高的区域的均匀覆盖(图7)。GC含量为80%的区域覆盖度高于平均覆盖度的0.5x。

表6. 使用协同间接捕获方法的NGS指标

	3-靶标 (n=3)	15-靶标 (n=5)	76-靶标 (n=6)
映射率	97.0%	93.8%	95.7%
中靶率	14.3%	83.6%	85.3%
0.2X 覆盖度	NA	98.6%	99.2%
0.5X 覆盖度	NA	88.6%	98.2%

实施例5

使用协同间接捕获法确定人SNP的NGS指标

[0162] 进行协同间接杂交分析以覆盖76个人ID单核苷酸多态性(SNP)。对20ng的人无细胞DNA(cfDNA)进行扩增前杂交。将结果与使用市售IDT xGen杂交和洗涤试剂盒的扩增后杂交结果进行比较。覆盖相同76个ID SNP的xGen人ID研究基因包V1.0用于捕获。使用xGEN人ID基因包按照商业方案对使用20ng cfDNA作为原始输入构建的NGS文库进行基于杂交的捕获。

[0163] 确定使用76个靶标基因包的二代测序(NGS)指标(表7)。扩增后捕获的靶标率很低,为中靶率的30.7%。相反,覆盖相同基因组区域的SICON-MAS基因包的中靶率为88%。

表7. 使用协同间接捕获方法的NGS指标

捕获	SICON-MAS 扩增前	IDT xGEN 扩增后
映射率	99.5%	97.7%
中靶率	88%	30.7%
0.2X 覆盖度	100%	100.0%
0.5X 覆盖度	96%	94%

实施例6

SICON-SEQ与扩增后方法的比较

[0164] 对一组76个人基因靶标进行用于测序的协同间接捕获核酸(SICON-SEQ),由10ng cfDNA输入提供了>80%中靶率,仅1小时的扩增前捕获。使用公司“I”试剂盒的扩增后捕获用于相同基因包,在16小时的扩增后捕获中,从双倍量输入(20ng cfDNA)的1M读段仅产生6-30%中靶率。使用公司I试剂盒进行扩增前捕获,但未能产生任何结果。

[0165] 图8A-图8B显示了SICON-SEQ和IDT xGen杂交和洗涤试剂盒对不同百分比的GC含量的区域的覆盖。对于SICON-SEQ测定,从低GC含量(<30%)到高GC含量(>50%)的区域的覆盖度非常均匀(图8A)。对于使用未产生文库富集的IDT xGEN试剂盒(图8B)的捕获方案,具有不同CG含量的区域的覆盖度存在系统偏差。

实施例7

SICON-TMS甲基化测定

[0166] 如图2A和2B所示进行SICON靶向甲基化测序(SICON-TMS)测定。从来自不同非癌症个体的3-5ml血浆提取样本cfDNA,并针对120个不同的差异甲基化区域(DMR)进行查询。读出显示与输入接近线性($R^2=0.9474$)关系,甚至低至1ng的cfDNA输入(图9)。

实施例8

通过SICON-TMS检测cfDNA中的甲基化DNA

[0167] 进行SICON-TMS测定以查询60个不同的差异甲基化区域(DMR)。

[0168] 按照NEBNextultra II试剂盒手册,首先使用cfDNA构建二代测序(NGS)文库。输入文库DNA(以0.01%、0.1%、1%、10%或100%的比例加标甲基化DNA的cfDNA)用于杂交捕获。将20ng未扩增的DNA与探针混合,将文库/探针混合物在杂交缓冲液中于95℃下变性30min。使混合物逐渐冷却至60℃。杂交混合物在热循环仪上于60℃下温育1小时。最终的杂交缓冲液含有100ng/μl鲑精DNA、1μg/μl牛血清白蛋白(BSA)、1μg/μl Ficoll、1μg/μl聚乙烯吡咯烷酮(PVP)、0.075M柠檬酸钠、0.75M NaCl、5x SSC和1X Denhardt's的溶液。

[0169] 对于清理,将捕获的组装体与链霉亲和素珠子(Thermo Fisher Dynabeads M270链霉亲和素)在室温下温育10min,然后进行3次洗涤(洗涤1:5X SSPE、1%SDS;洗涤2:2X SSPE,0.1%;洗涤3:0.1X SSPE,0.01%triton)。清理后的组装体用亚硫酸氢盐处理以进行甲基化分析。

[0170] 图10显示了预期加标和测量值之间的关系。SICON-TMS测定显示低至0.01%甲基化的分析灵敏度和线性。甲基化百分比与预期值高度相关, R^2 为0.99,表明该测定的准确度很高。

实施例9

通过SICON-TMS检测cfDNA中的癌症甲基化模式

[0171] 对来自正常结肠组织和结肠癌组织的样本,以及来自健康个体和结肠癌患者的血浆cfDNA样本进行亚硫酸氢盐处理和测序。将测序读段映射到每个差异甲基化区域(DMR),并进行去重。每个得到的读段均包含来自通过测定捕获的独特DNA分子的CpG甲基化信息。然后针对每个读段计算两个指标:

1) N:读段中的CpG总数;

2) M:读段中的甲基化CpG的数目。

根据1)和2),第三个指标计算如下:

3) $f = M/N$, 在当前读段中甲基化的CpG分数。

[0172] 结果显示为散点图, 显示每个DMR的 f (y轴) 与 N (x轴), DMR中的每个读段在图中显示为点。图11显示了DMR1在正常结肠组织 (图11A) 和结肠癌组织基因组DNA (图11B) 中的分子甲基化散射模式。它展示了一种DMR, 其中正常结肠组织中没有超甲基化DNA分子, 而结肠癌组织中存在大量超甲基化分子。

[0173] 图12A和12B分别显示了DMR2在正常结肠组织和结肠癌组织基因组DNA中的分子甲基化散射模式。这些图显示了一种DMR, 其中正常结肠组织中存在一些超甲基化DNA分子, 而结肠癌组织中存在大量超甲基化分子。

[0174] 图13A和13B分别显示了健康个体血浆cfDNA和结肠癌患者血浆cfDNA中DMR1和DMR2的分子甲基化散射模式。图13B上部所示的来自每个DMR的超甲基化分子的计数可以用作从液体活检中检测疾病的基础。

实施例10

通过SICON-TMS检测cfDNA中的癌症甲基化模式

[0175] 覆盖100个甲基化标志物的Point-n Seq结肠直肠癌 (CRC) 基因包分三步设计。首先, 从公共数据库识别约1000个CRC特异性标志物。其次, 淘汰健康人群基线cfDNA中背景信号高的标志物。最后, 最终确定列表以包含患者和健康cfDNA之间最具区分性的标志物。SICON CRC基因包的捕获效率很高, 从而实现了高均匀性 ($94\% > 0.5X$, $100\% > 0.2X$) 和中靶率 ($> 80\%$)。对于20ng cfDNA输入, 尽管GC含量高 ($> 80\%$), 但平均每个标志物获得超过1000个去重读段。信息读段的输出与1ng至40ng的cfDNA输入呈线性关系。在滴定研究中, 在cfDNA背景下可靠地检测到20ng cfDNA (0.003%) 中的0.6pg (0.2X基因组当量) 甲基化DNA。在一项使用来自结肠直肠癌患者的血浆样本的初步临床研究中——早期 (I, $n=7$; II, $n=7$)、晚期 (III, $n=11$; IV, $n=3$) 和对照个体 ($n=105$), 对照、I、II、III、IV期相应的甲基化信号的平均分数分别为0.0034%、0.013%、0.09%、0.17%、0.29%。I期样本的甲基化分数与对照组具有显著差异 ($P < 0.001$)。通过使用甲基化分数的简单截断, Point-n Seq CRC基因包在I期达到86%的灵敏度, 在 (II-IV) 期达到100%, 特异性为91%, $AUC = 0.96$ 。

实施例11

对CRC血浆样本的Point-n-Seq SNV+甲基双重捕获分析

[0176] 在晚期CRC患者的血浆样本 (1ml) 中, 通过统一的Point-n-Seq测定检测遗传和表观遗传变化。设计Point-n-Seq结肠直肠癌 (CRC) 基因包, 涵盖来自22个基因的甲基化标志物和>350个热点突变。

[0177] 使用甲基化标志物基因包和突变热点基因包通过本文所述的协同间接杂交捕获进行两轮连续的靶标富集。简言之, 将20 μ L每个cfDNA样本添加到PCR管中。对于小于20 μ L的DNA体积, 添加IDTE或Buffer EB至最终体积为20 μ L。对于每个样本, 添加2.8 μ L末端制备缓冲液和1.2 μ L末端制备酶。通过温和涡旋将管充分混合, 然后短暂离心。试管在带加热盖的热循环仪中在20 $^{\circ}$ C温度下运行30min, 然后在65 $^{\circ}$ C下运行30min。随后向其中加入2.5 μ L的衔接子溶液, 以及13 μ L的连接混合物, 并将混合物在20 $^{\circ}$ C下温育30min。

[0178] 将样本结合珠子平衡至室温持续至少15分钟, 然后涡旋以重悬。将48 μ L ($\sim 1.2x$ 体积) 的文库结合珠子添加到39.5 μ L连接反应中。通过移液至少10次将它们彻底混合并短暂离心。将混合物在室温下温育10min, 然后放置在磁铁上至少2min或直至溶液澄清。去除并

弃去上清液。在磁铁上,在不干扰珠子的情况下,将150 μ L样本洗涤缓冲液添加到珠子中,温育2min,弃去上清液。

[0179] 对于靶标捕获,添加包含突变捕获基因包和探针结合混合物的杂交混合物,并通过温和涡旋或轻弹充分混合。将混合物加热至98 $^{\circ}$ C保持2min,然后以2.5 $^{\circ}$ C/s的速率降温至60 $^{\circ}$ C,并在60 $^{\circ}$ C下温育60min。杂交60min后,将样本置于磁铁上30秒,小心地将上清液转移到标记的试管中,并保存用于第二杂交步骤。将珠子洗涤3次并重悬,DNA在珠子上扩增。

[0180] 将上述保存的上清液与包含TMS捕获基因包的杂交混合物混合,并如突变捕获基因包进行捕获杂交。捕获的TMS DNA经过二硫化物处理、修复并从珠子上洗脱,然后进行指数PCR。两个扩增的DNA样本均准备用于测序,并在Illumina平台上进行测序。

[0181] 图14图示了顺序靶标富集。表8列出了每个患者样本的DNA输入量、甲基化信号的分数和突变信号的分数。检测到的突变的详情显示在图15中。如表8所示,Point-n-Seq CRC突变和甲基化基因包的捕获效率很高,从而可以检测到来自各种起始DNA量的超甲基化和突变。此外,使用来自CRC患者的血浆cfDNA进行的甲基化和突变组合分析显示,从甲基化状态和驱动突变等位基因频率估计的肿瘤含量一致。

[0182] 表8

	血浆体积 (ml)	DNA 输入 (ng)	甲基化信号%	突变%
CRC 1	1	2.74	1.09%	4.40%
CRC 2	1	6	0.34%	0.00%
CRC 3	1	6	5.12%	11.00%
CRC 4	1	4.95	0.19%	1.80%
CRC 5	1	49.8	4.38%	0.00%

实施例12

双重分析的甲基化信号与独立甲基化 (TMS) 分析相当

[0183] 为了评估源自顺序靶标富集方法的甲基化信号,使用加标到对照cfDNA中的来自细胞系HCT116的gDNA进行滴定实验。HCT116gDNA的加标浓度范围为0.001%至10%。相同的DNA输入单独进行TMS分析或通过顺序SICON进行突变-TMS双重分析,其中首先进行突变分析的富集步骤,然后进行TMS分析的富集步骤,如图14所示。如图16所示,来自独立分析和双重分析的甲基化评分相当,表明甲基化测定的灵敏度并没有因为连续捕获双重分析中的第二捕获受到损害。图17显示第二捕获TMS回收率(来自每个差异甲基化区域 (DMR) 的测序的信息分子计数)是第一捕获TMS的约85%。

实施例13

肿瘤知情的个性化基因包分析

[0184] 对CRC肿瘤gDNA进行全外显子测序,并选择114个单核苷酸变体来制作个性化基因包。在滴定实验中,CRC肿瘤gDNA以0.001%、0.003%、0.01%、0.03%和0.1%的浓度加标到对照cfDNA中。如图18所示,以0.003%加标的样本可以与0%分离,这表明对于特定的基于个性化杂交的测定,检测限为0.003%。预期更大型基因包将产生更低检测限。

[0185] 虽然本文已经示出和描述了本发明的优选实施方案,但是对于本领域技术人员而

言显而易见的是,这些实施方案仅作为示例提供。在不脱离本发明的情况下,本领域技术人员现在将想到许多变化、改变和替换。应当理解,在实施本发明时可以采用对本文所述的本发明的实施方案的各种替代方案。以下权利要求旨在限定了本发明的范围,并由此覆盖这些权利要求范围内的方法和结构及其等同物。

SICON 靶向测序 (TS)

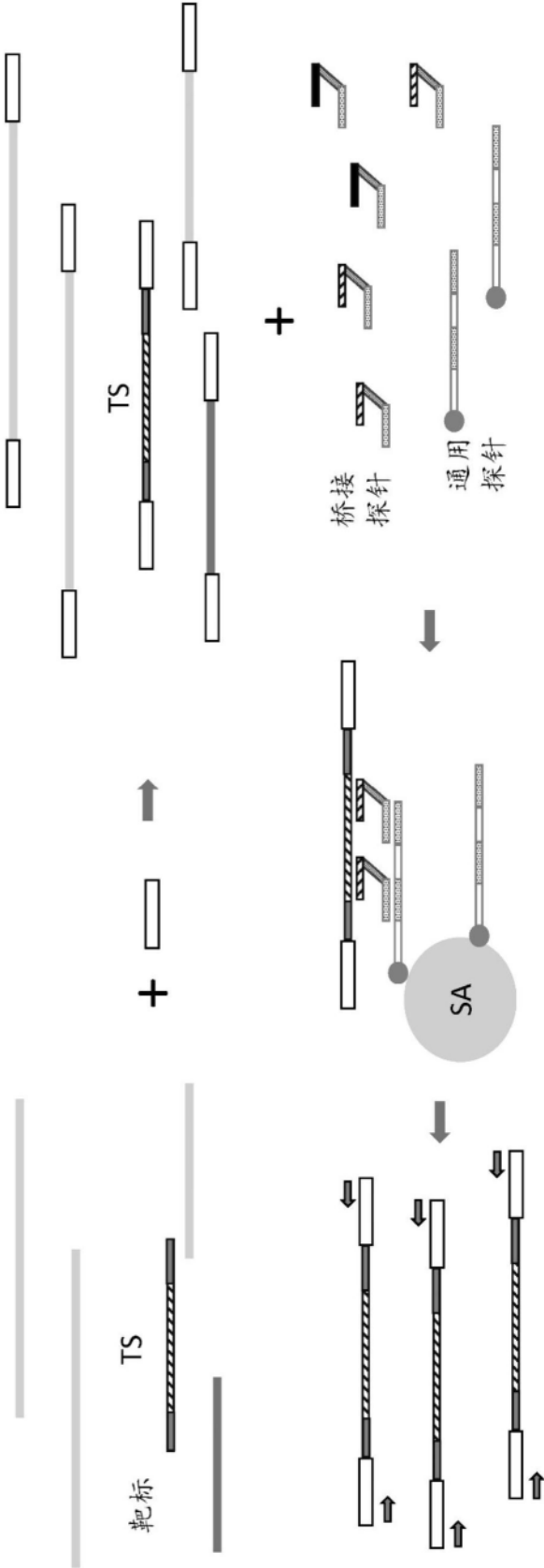


图1

SICON 靶向甲基化测序 (TMS)

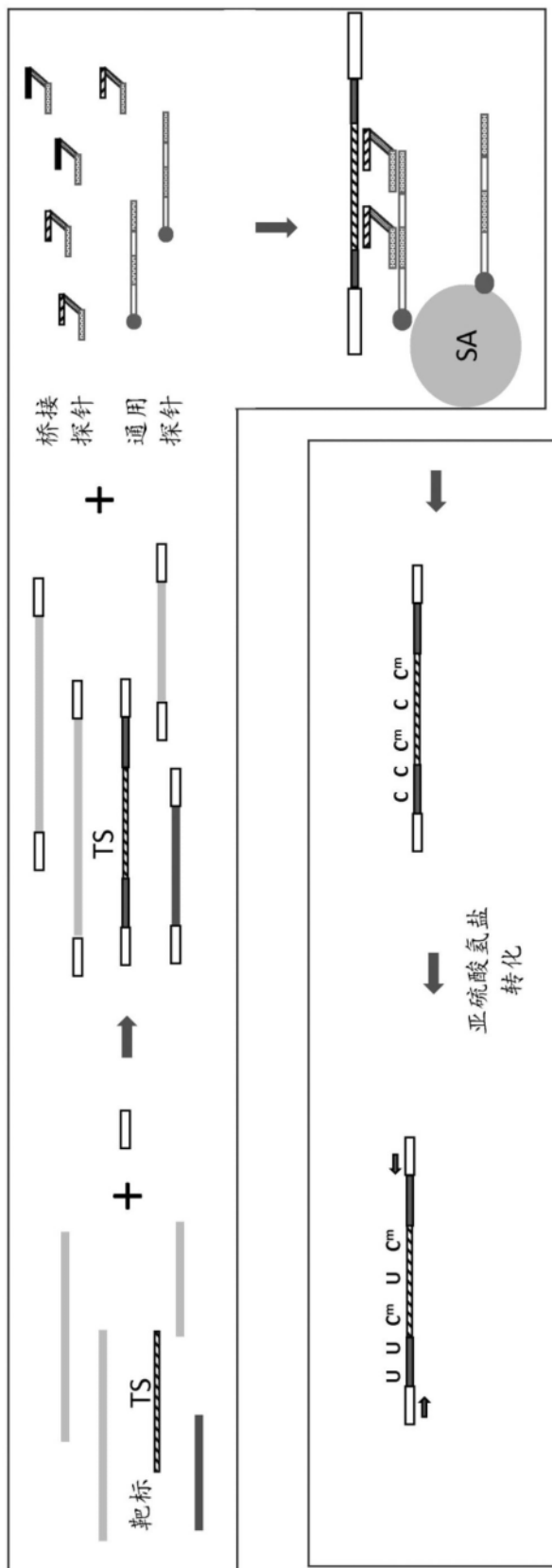


图 2A

图 2B

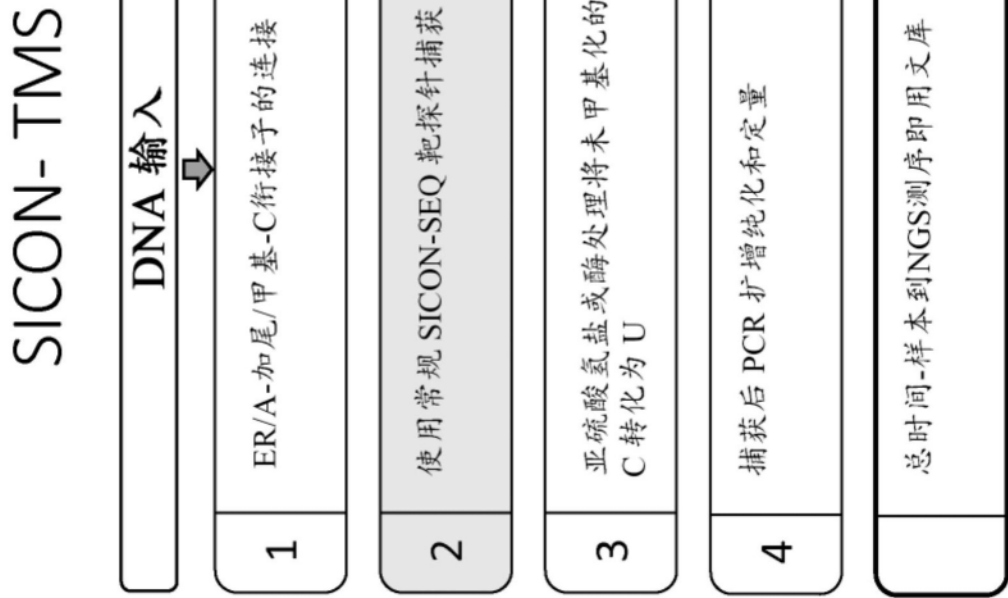


图3

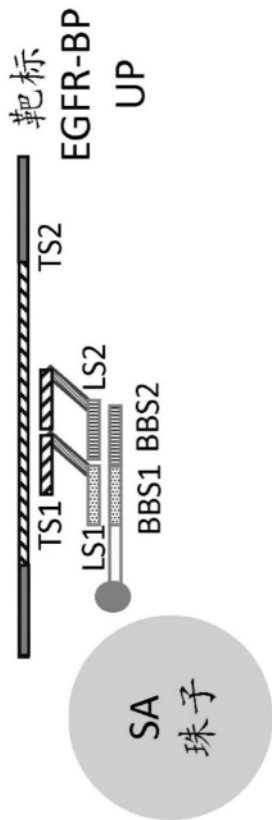


图4

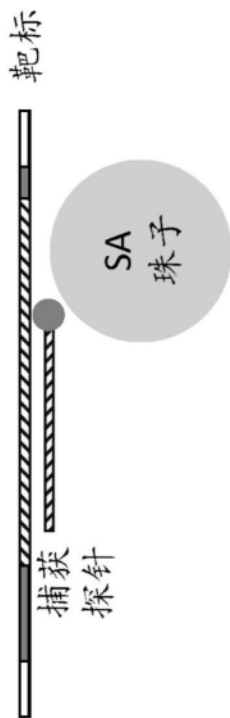


图5A

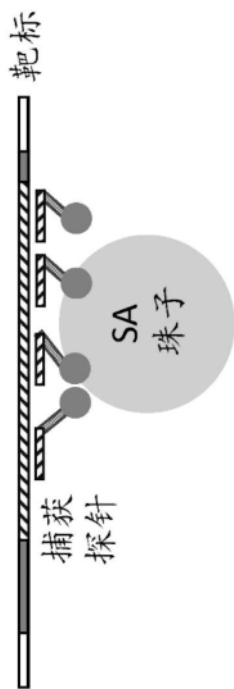


图5B

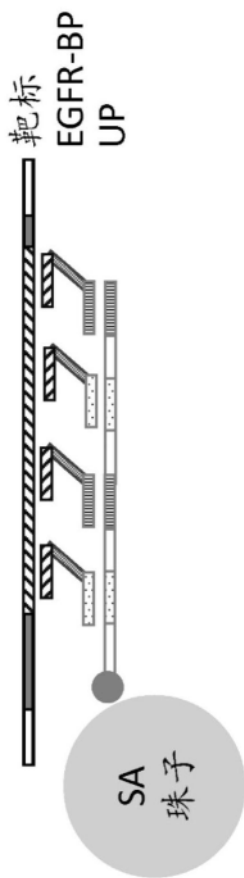


图5C

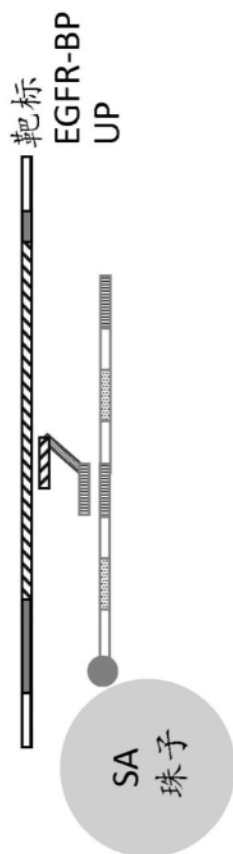


图5D

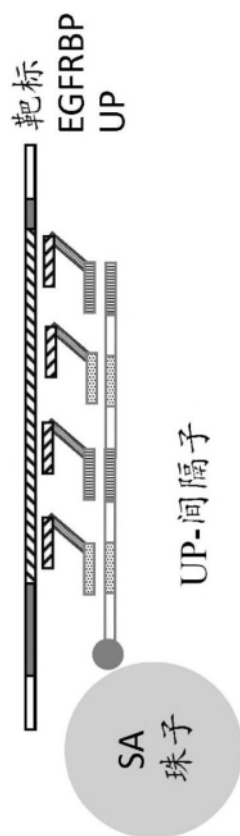


图6A

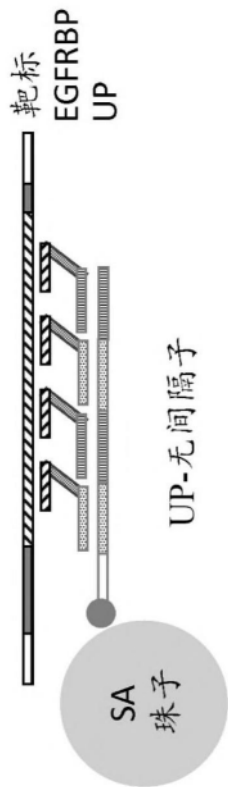


图6B

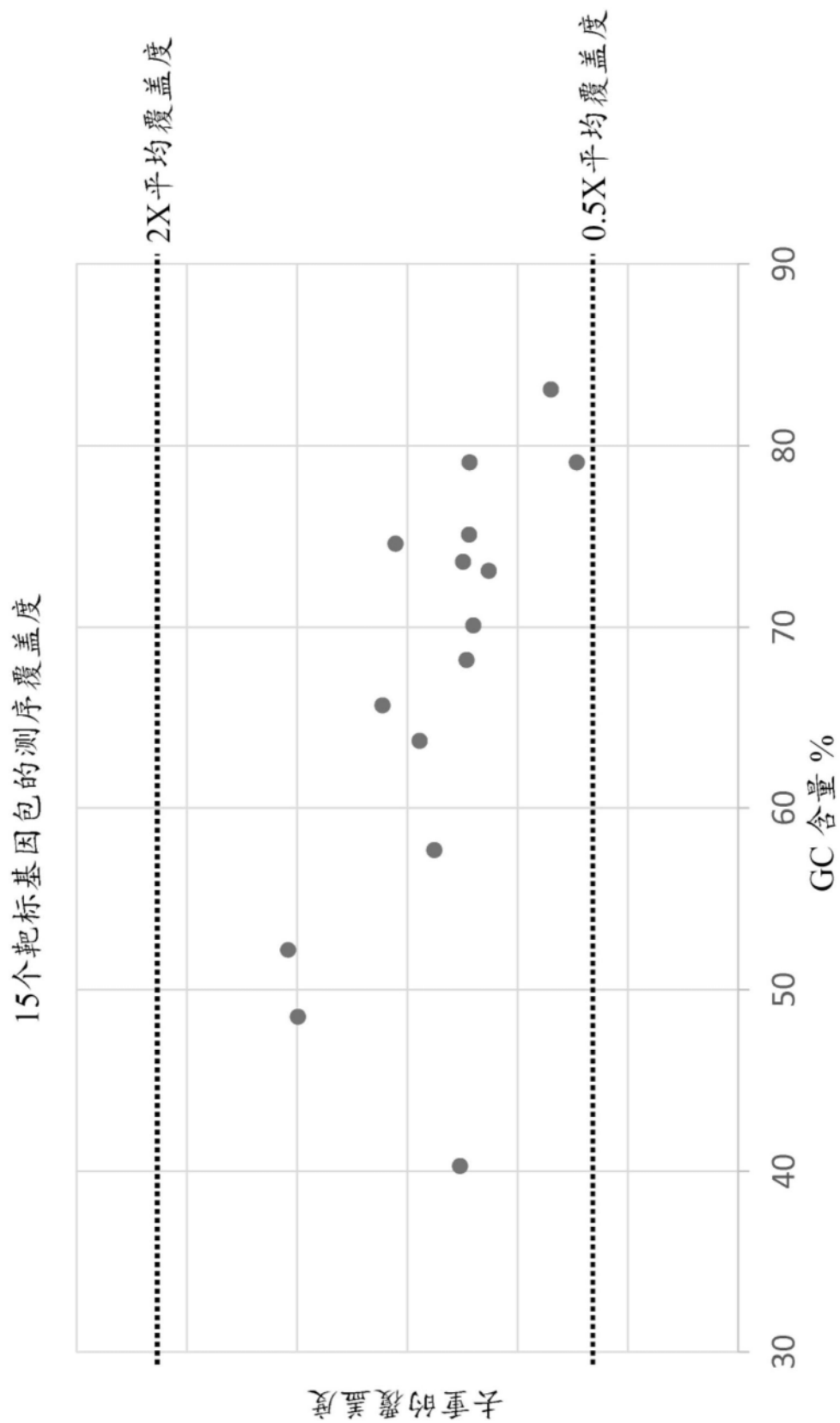


图7

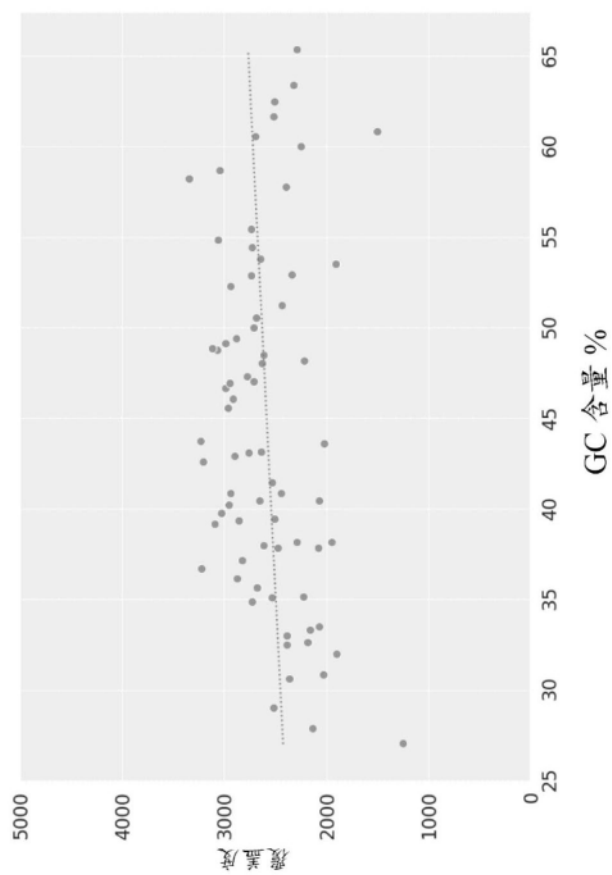


图8A

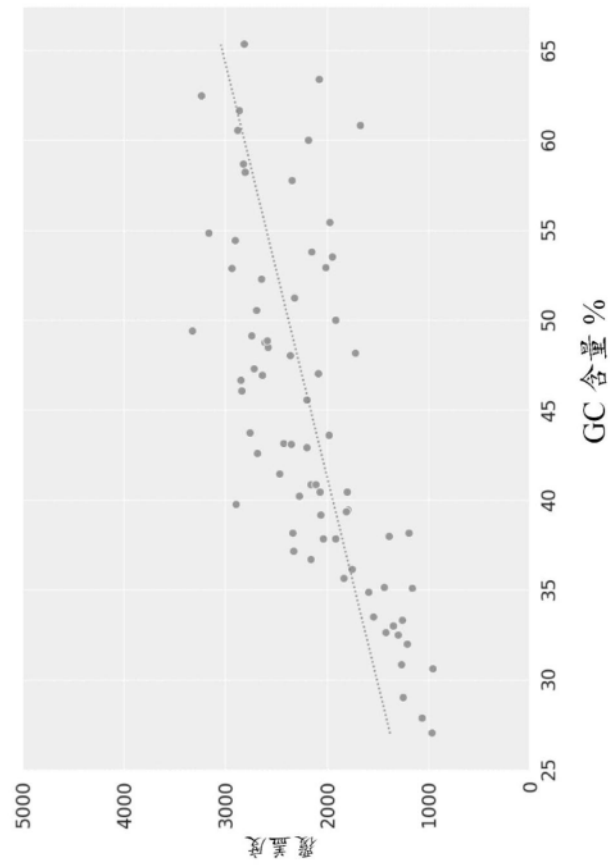


图8B

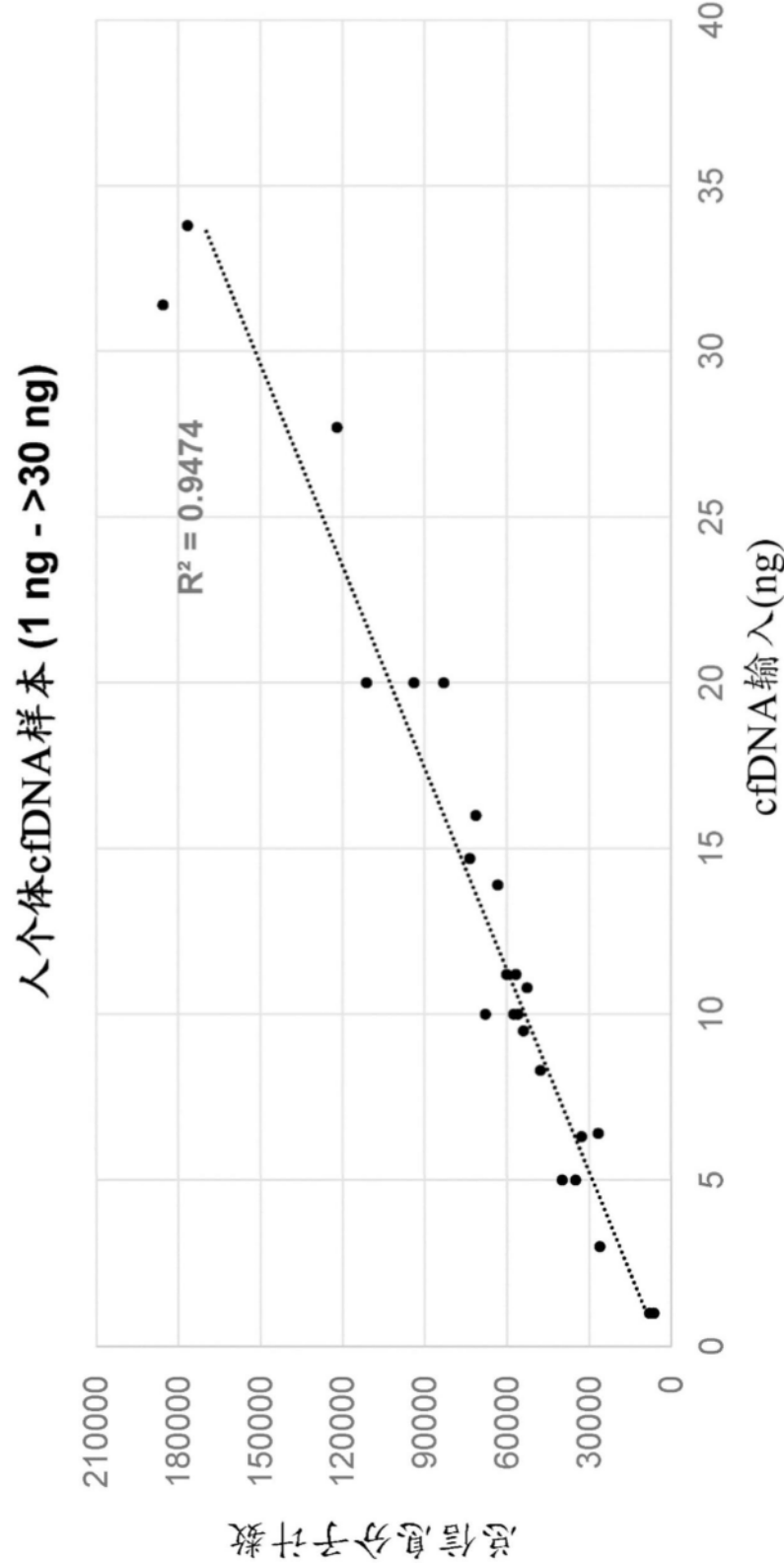


图9

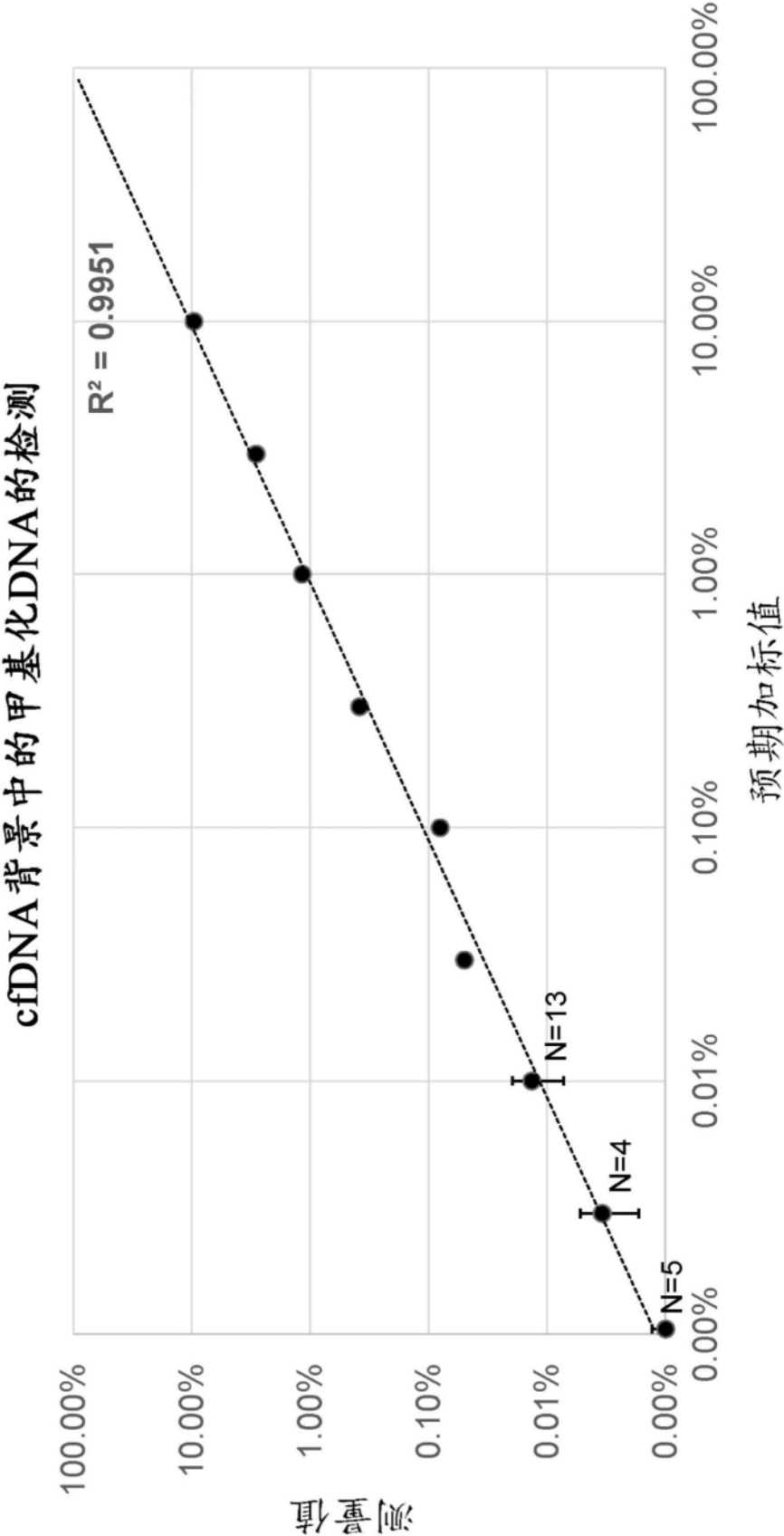


图10

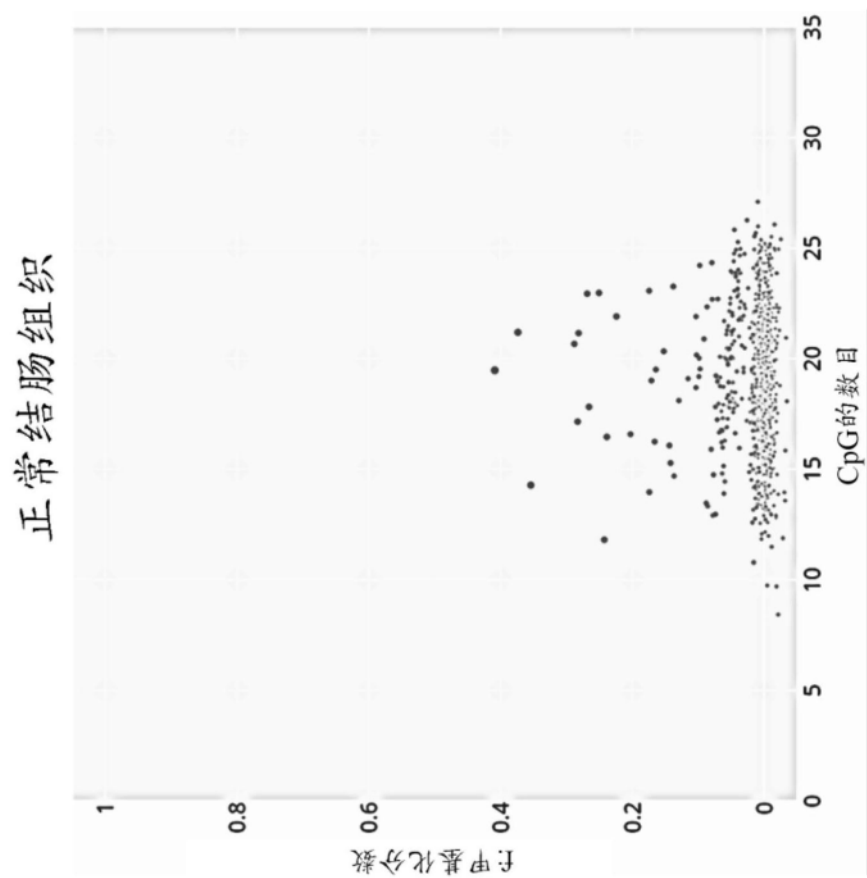


图11A

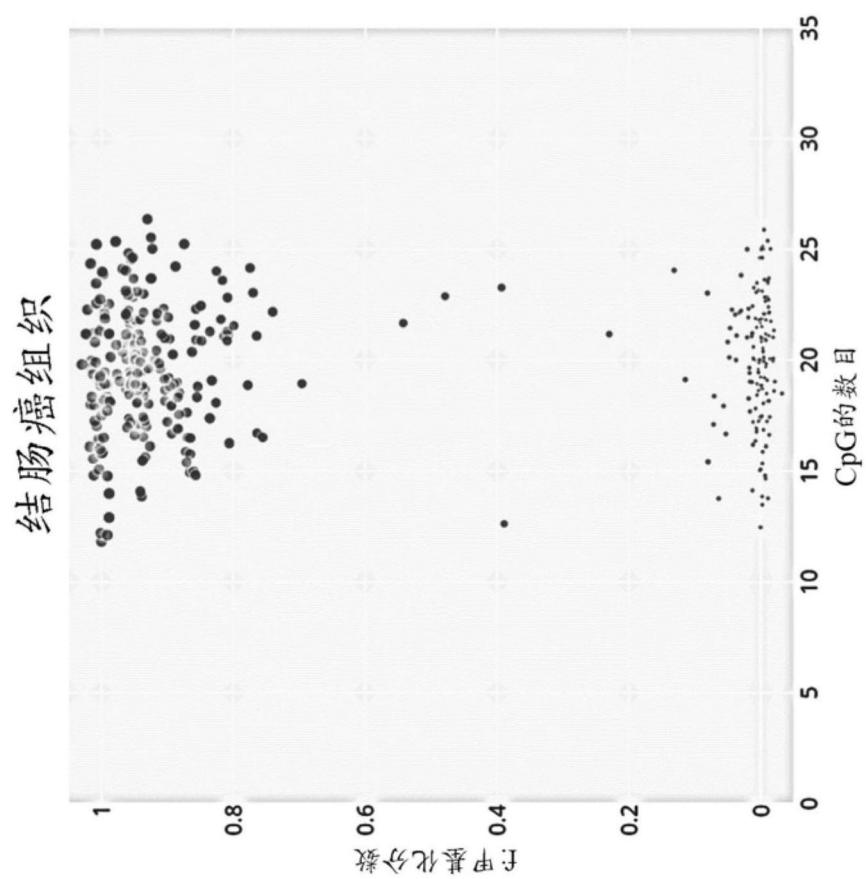


图11B

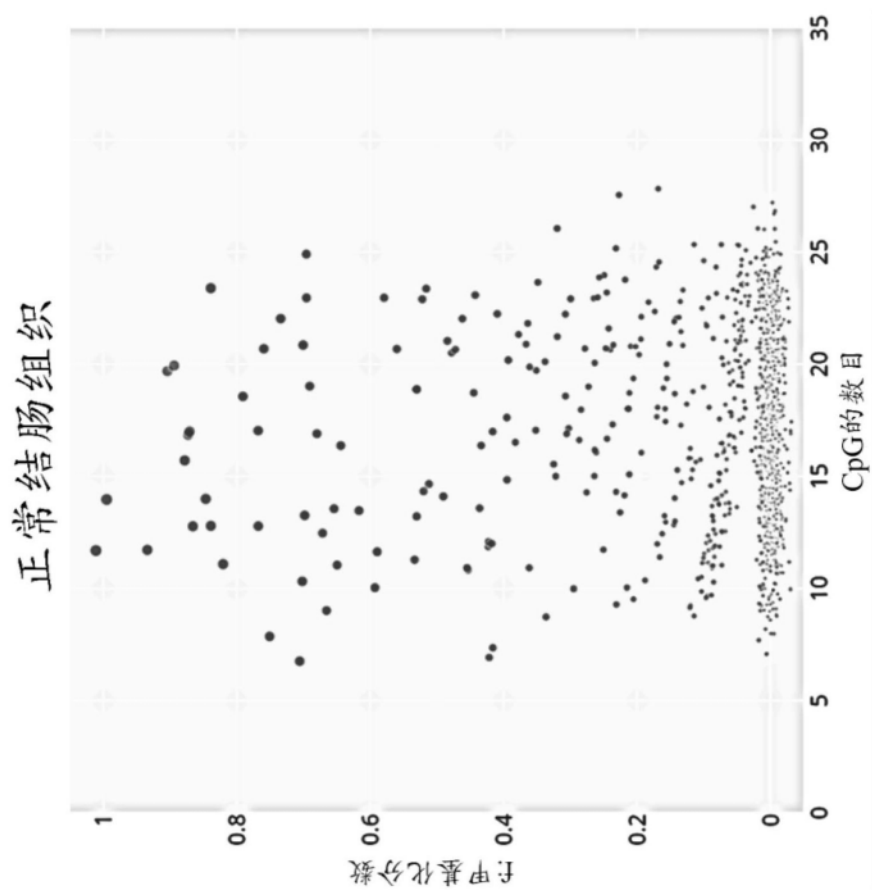


图12A

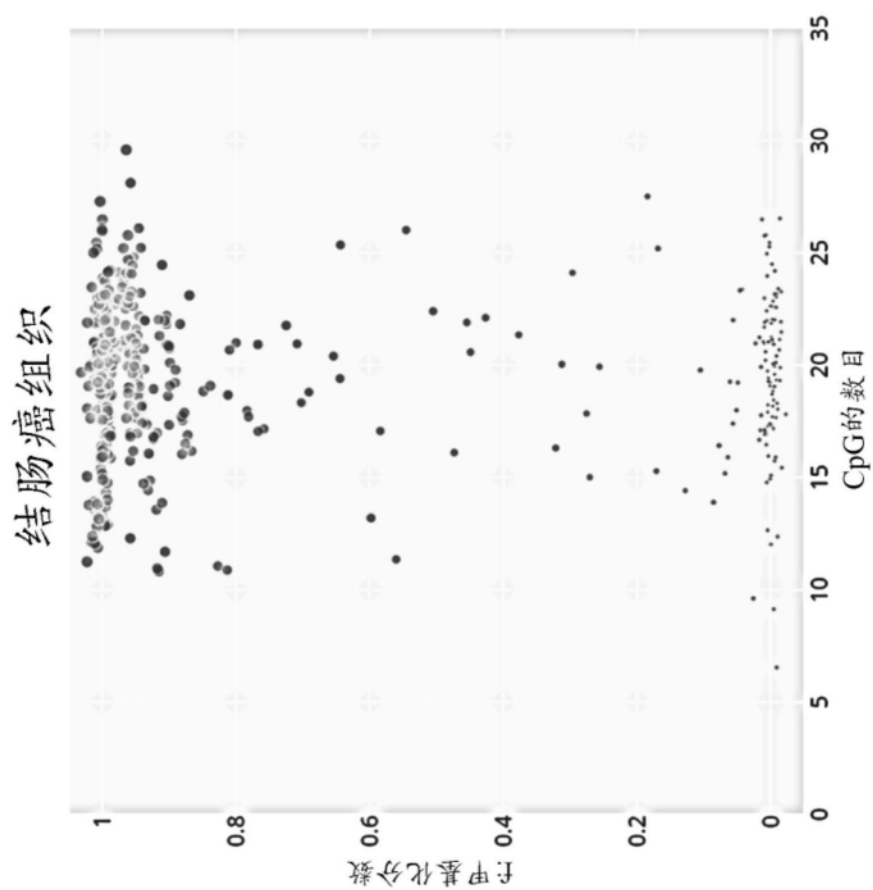


图12B

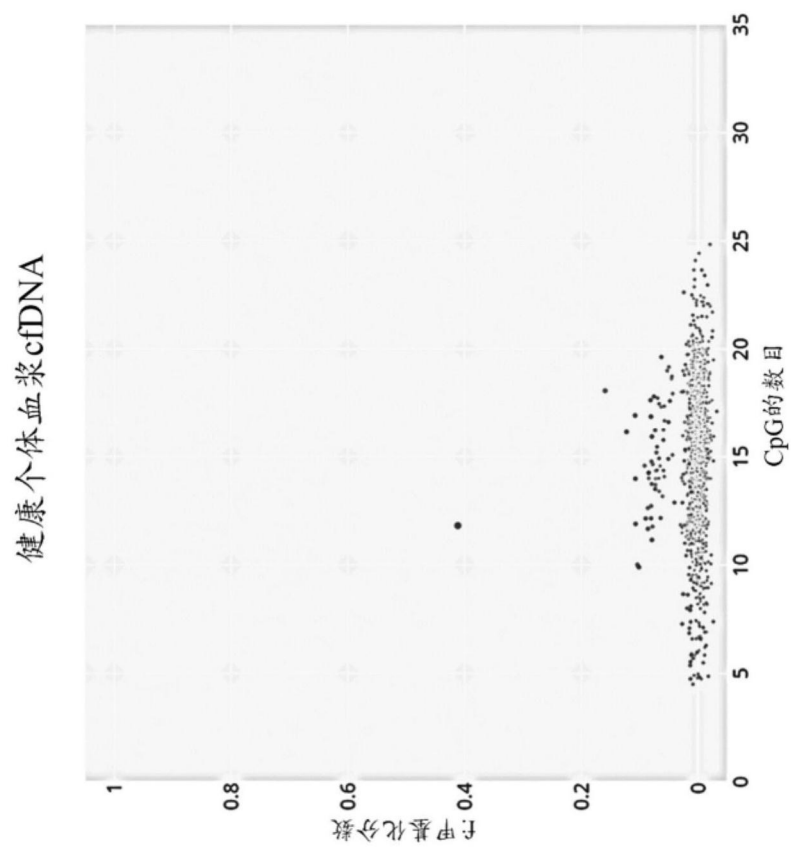


图13A

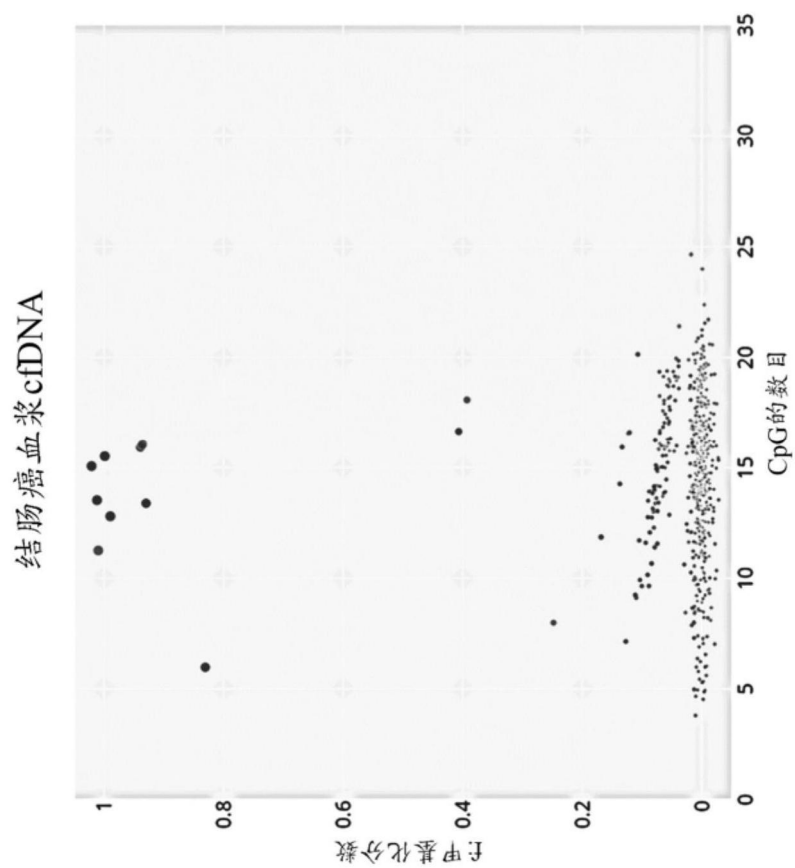


图13B

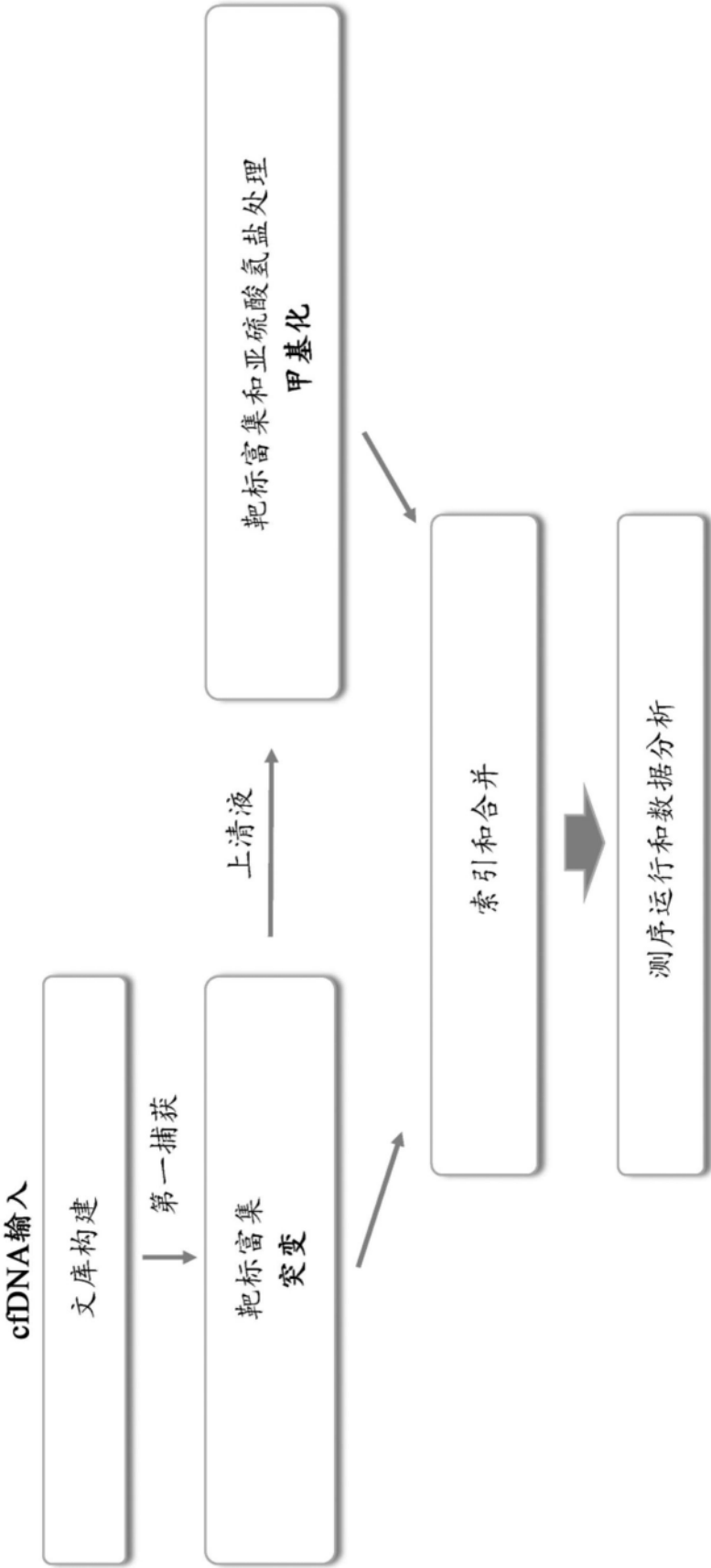


图14

样本	基因:Pshort	Trans_id:cDNA	VAF	效果_Pred
CRC1	APC:p.R1314fs	NM_000038:c.3932_3939dup TTGGAAC	0.0141	移码
	KRAS:p.G12V	NM_033360:c.35G>T	0.0228	错义
	TP53:p.V173M	NM_000546:c.517G>A	0.0317	错义
CRC3	ERBB2:p.V842I	NM_004448:c.2524G>A	0.0956	错义
	PIK3CA:p.H1047R	NM_006218:c.3140A>G	0.1016	错义
	TP53:p.R273C	NM_000546:c.817C>T	0.0172	错义
CRC4	CTNNB1:p.T41A	NM_001098209:c.121A>G	0.0134	错义
	TP53:p.R175H	NM_000546:c.524G>A	0.0092	错义
	PTEN:p.H296Y	NM_001304717:c.886C>T	0.0042	错义
	GNAS:p.V857E	NM_080425:c.2570T>A	0.0036	错义
	KRAS:p.G12D	NM_033360:c.35G>A	0.0036	错义

图15

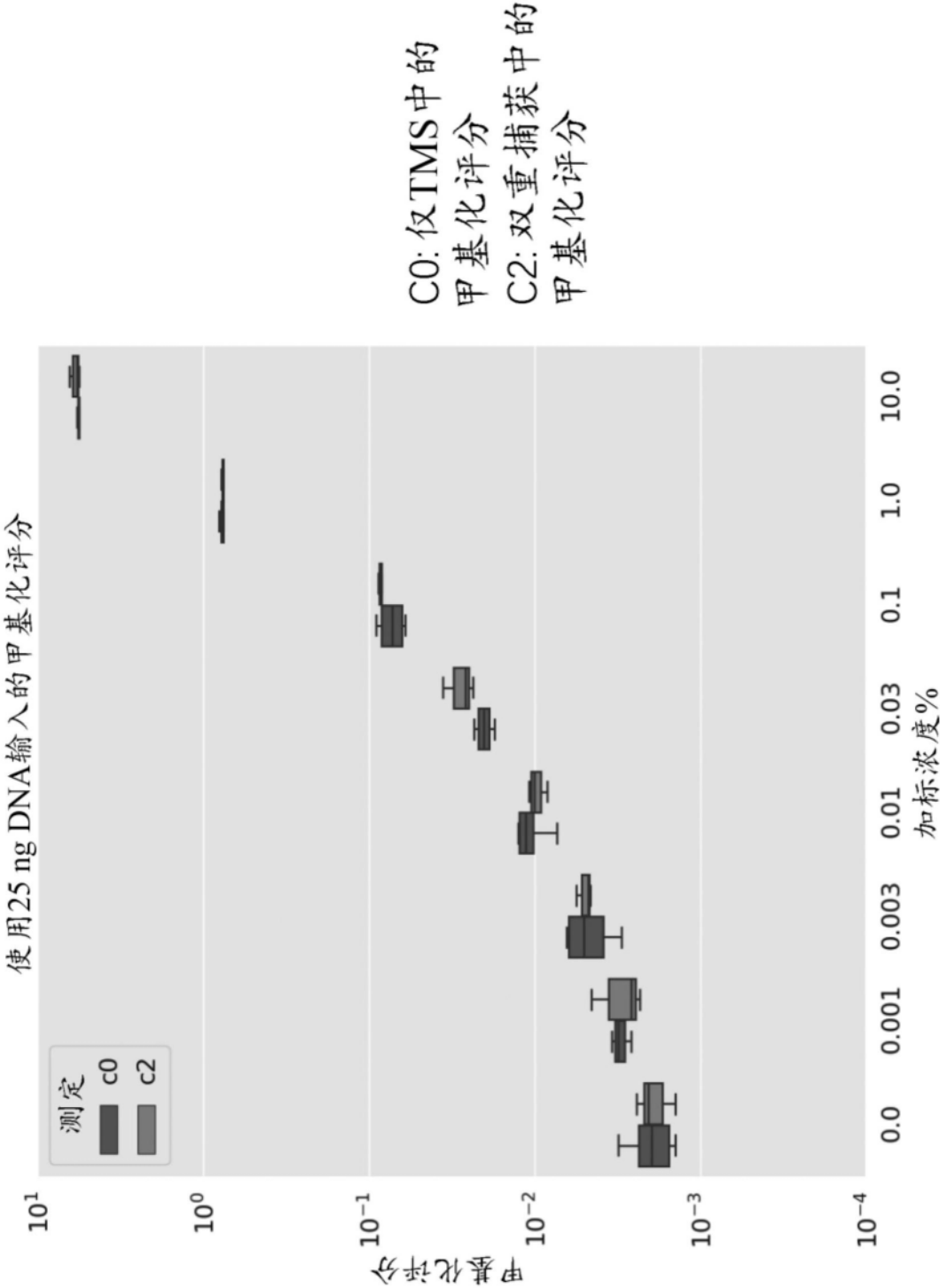


图16

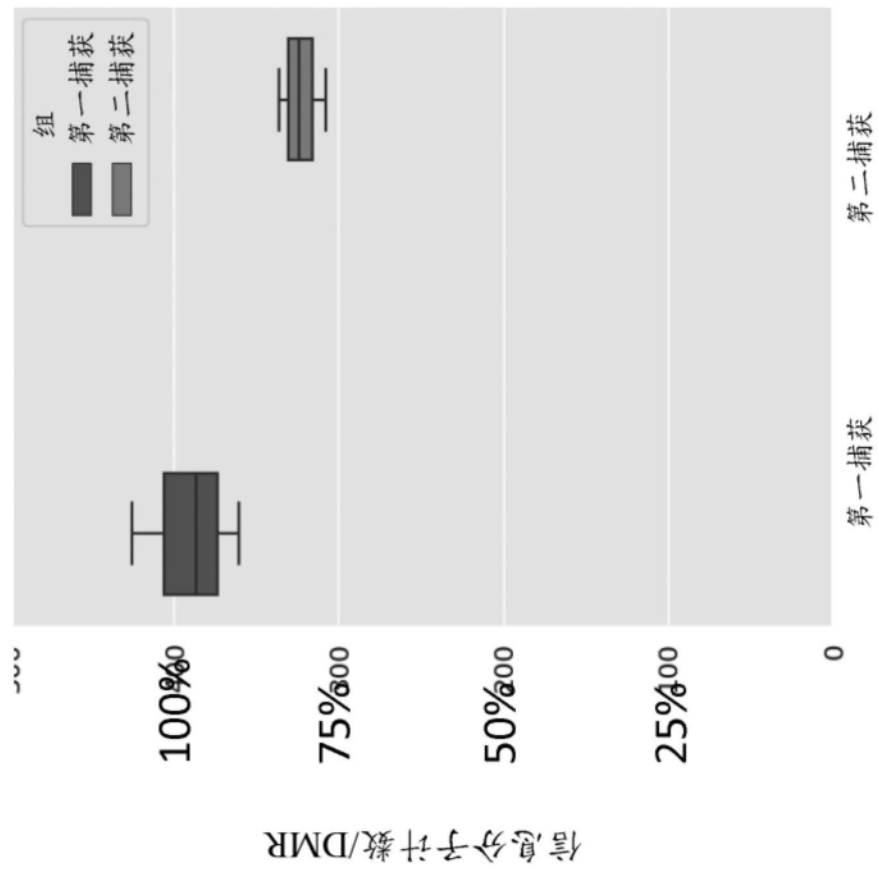


图17

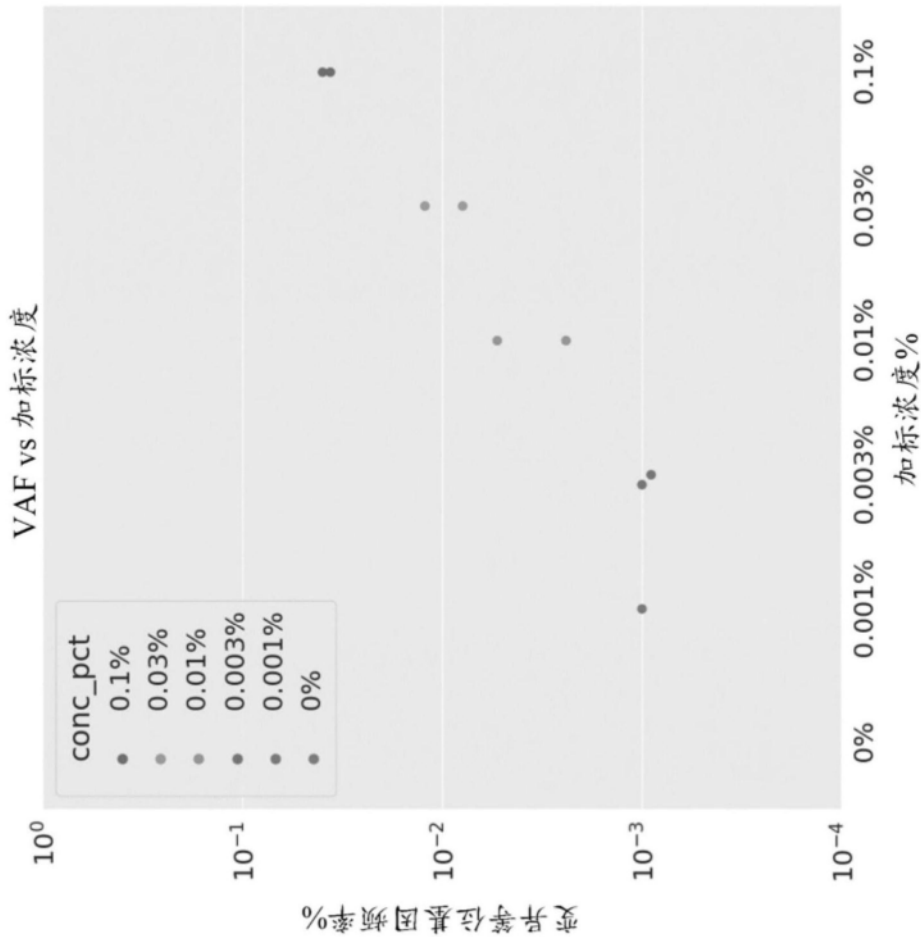


图18

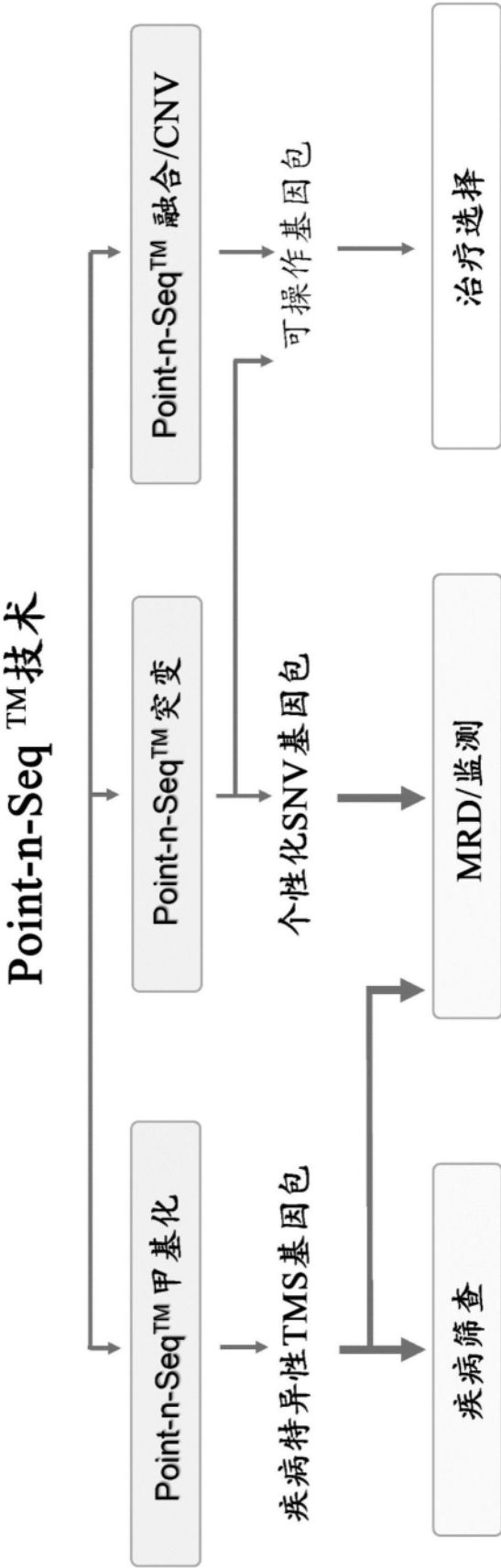


图19