



(12)发明专利申请

(10)申请公布号 CN 109072294 A

(43)申请公布日 2018.12.21

(21)申请号 201680080120.4

(74)专利代理机构 深圳市百瑞专利商标事务所
(普通合伙) 44240

(22)申请日 2016.12.08

代理人 金辉

(30)优先权数据

62/264,822 2015.12.08 US

62/281,917 2016.01.22 US

(51)Int.Cl.

C12Q 1/6869(2018.01)

(85)PCT国际申请进入国家阶段日

2018.07.26

(86)PCT国际申请的申请数据

PCT/US2016/065605 2016.12.08

(87)PCT国际申请的公布数据

W02017/100441 EN 2017.06.15

(71)申请人 特温斯特兰德生物科学有限公司

地址 美国华盛顿州

(72)发明人 L·A·洛布 M·W·施密特

J·J·索尔克

权利要求书17页 说明书37页

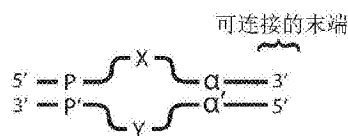
序列表14页 附图21页

(54)发明名称

用于双重测序的改良衔接子、方法和组合物

(57)摘要

本文公开用于测序双链靶核酸的衔接子核酸序列,其包含链限定元件(SDE),例如用于双链衔接子的每一链的条码;双链复合核酸;组合物以及方法,其中通过双重测序应用误差校正。



1. 一种用于测序双链靶核酸分子的衔接子核酸序列对, 包含第一衔接子核酸序列和第二衔接子核酸序列, 其中每一衔接子核酸序列包含:

引物结合结构域,

链限定元件 (SDE),

单个分子标识符 (SMI) 结构域, 以及

连接结构域;

其中所述第一衔接子核酸序列的SDE与所述第二衔接子核酸序列的SDE至少部分地非互补。

2. 根据权利要求1所述的衔接子核酸序列对, 其中所述两个衔接子序列由至少部分地一起退火的两个单独DNA分子组成。

3. 根据权利要求1所述的衔接子核酸序列对, 其中所述第一衔接子核酸序列和所述第二衔接子核酸序列经由连接子结构域连接。

4. 根据权利要求3所述的衔接子核酸序列对, 其中所述连接子结构域由核苷酸构成。

5. 根据权利要求3所述的衔接子核酸序列对, 其中所述连接子结构域含有一种或多种经修饰的核苷酸或非核苷酸分子。

6. 根据权利要求4所述的衔接子核酸序列对, 其中所述一种或多种经修饰的核苷酸或非核苷酸分子选自无碱基位点; 尿嘧啶; 四氢呋喃; 8-氧代-7,8-二氢-2'-脱氧腺苷 (8-氧代-A); 8-氧代-7,8-二氢-2'-脱氧鸟苷 (8-氧代-G); 脱氧肌苷; 5'-硝基咪唑; 5-羟基甲基-2'-脱氧胞苷; 异胞嘧啶; 5'-甲基-异胞嘧啶; 或异鸟苷。

7. 根据权利要求3到5中任一项所述的衔接子核酸序列对, 其中所述连接子结构域形成环。

8. 根据权利要求1到7中任一项所述的衔接子核酸序列对, 其中所述第一衔接子核酸序列的SDE与所述第二衔接子核酸序列的SDE非互补。

9. 根据权利要求1到7中任一项所述的衔接子核酸序列对, 其中所述第一衔接子核酸序列的引物结合结构域与所述第二衔接子核酸序列的引物结合结构域至少部分地互补。

10. 根据权利要求1到9中任一项所述的衔接子核酸序列对, 其中所述第一衔接子核酸序列的引物结合结构域与所述第二衔接子核酸序列的引物结合结构域互补。

11. 根据权利要求1到7中任一项所述的衔接子核酸序列对, 其中所述第一衔接子核酸序列的引物结合结构域与所述第二衔接子核酸序列的引物结合结构域至少部分地非互补。

12. 根据权利要求1到7中任一项所述的衔接子核酸序列对, 其中至少一个SMI结构域为内源性SMI。

13. 根据权利要求12所述的衔接子核酸序列对, 其中所述内源性SMI与剪切点有关。

14. 根据权利要求1到11中任一项所述的衔接子核酸序列对, 其中所述SMI结构域包含至少一种简并或半简并核酸。

15. 根据权利要求1到11中任一项所述的衔接子核酸序列对, 其中所述SMI结构域为非简并的。

16. 根据权利要求1到15中任一项所述的衔接子核酸序列对, 其中所述SMI结构域的序列考虑与对应于连接的DNA的随机或半随机剪切末端的序列结合以获得能够彼此区分单个DNA分子的SMI序列。

17. 根据权利要求1到16中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列的SMI结构域与所述第二衔接子核酸序列的SMI结构域至少部分地互补。

18. 根据权利要求1到17中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列的SMI结构域与所述第二衔接子核酸序列的SMI结构域互补。

19. 根据权利要求1到15中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列的SMI结构域与所述第二衔接子核酸序列的SMI结构域至少部分地非互补。

20. 根据权利要求19所述的衔接子核酸序列对,其中每一SMI结构域包含引物结合位点。

21. 根据权利要求20所述的衔接子核酸序列对,其中每一SMI结构域位于其连接结构域远端。

22. 根据权利要求19所述的衔接子核酸序列对,其中所述第一衔接子核酸序列的SMI结构域与所述第二衔接子核酸序列的SMI结构域非互补。

23. 根据权利要求1到22中任一项所述的衔接子核酸序列对,其中每一SMI结构域包含约1个到约30个之间的简并或半简并核酸。

24. 根据权利要求1到23中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列的连接结构域与所述第二衔接子核酸序列的连接结构域至少部分地互补。

25. 根据权利要求1到24中任一项所述的衔接子核酸序列对,其中每一连接结构域能够连接到双链靶核酸序列的一个链。

26. 根据权利要求1到25中任一项所述的衔接子核酸序列对,其中连接结构域中的一个包含T-突出端、A-突出端、CG-突出端、平末端、或另一可连接的核酸序列。

27. 根据权利要求1到26中任一项所述的衔接子核酸序列对,其中两个连接结构域包含平末端。

28. 根据权利要求1到26中任一项所述的衔接子核酸序列对,其中所述连接结构域中的至少一个包含经修饰的核酸。

29. 根据权利要求28所述的衔接子核酸序列对,其中所述经修饰的核苷酸选自无碱基位点;尿嘧啶;四氢呋喃;8-氧代-7,8-二氢-2'-脱氧腺苷(8-氧代-A);8-氧代-7,8-二氢-2'-脱氧鸟苷(8-氧代-G);脱氧肌苷;5'-硝基吡啶;5-羟基甲基-2'-脱氧胞苷;异胞嘧啶;5'-甲基-异胞嘧啶;或异鸟苷。

30. 根据权利要求1到26中任一项所述的衔接子核酸序列对,其中所述连接结构域中的至少一个包含去磷酸化碱基。

31. 根据权利要求1到26中任一项所述的衔接子核酸序列对,其中所述连接结构域中的至少一个包含去羟基化碱基。

32. 根据权利要求1到26中任一项所述的衔接子核酸序列对,其中所述连接结构域中的至少一个已经化学修饰以便使其不可连接。

33. 根据权利要求1到27中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列的SDE与所述第二衔接子核酸序列的SDE不同和/或在至少一个核苷酸处非互补。

34. 根据权利要求1到27中任一项所述的衔接子核酸序列对,其中至少一个核苷酸通过酶反应从所述第一衔接子核酸序列的SDE或所述第二衔接子核酸的SDE省去。

35. 根据权利要求34所述的衔接子核酸序列对,其中所述酶反应包含聚合酶、核酸内切

酶、糖基化酶或裂解酶。

36. 根据权利要求33所述的衔接子核酸序列对,其中所述至少一种核苷酸为经修饰的核苷酸或包含标记的核苷酸。

37. 根据权利要求37所述的衔接子核酸序列对,其中所述经修饰的核苷酸或包含标记的核苷酸选自无碱基位点;尿嘧啶;四氢呋喃;8-氧代-7,8-二氢-2'-脱氧腺苷(8-氧代-A);8-氧代-7,8-二氢-2'-脱氧鸟苷(8-氧代-G);脱氧肌苷;5'-硝基咪唑;5-羟基甲基-2'-脱氧胞苷;异胞嘧啶;5'-甲基-异胞嘧啶;或异鸟苷。

38. 根据权利要求35所述的衔接子核酸序列对,其中所述第一衔接子核酸序列的SDE包含能够形成发夹环的自互补结构域。

39. 根据权利要求1到38中任一项所述的衔接子核酸序列对,其中第一衔接子核酸序列远离其连接结构域的末端连接到所述第二衔接子核酸序列远离其连接结构域的末端,由此形成环。

40. 根据权利要求39所述的衔接子核酸序列对,其中所述环包含限制酶识别位点。

41. 根据权利要求1到35中任一项所述的衔接子核酸序列对,其中至少所述第一衔接子核酸序列进一步包含第二SDE。

42. 根据权利要求41所述的衔接子核酸序列对,其中所述第二SDE位于所述第一衔接子核酸序列的末端处。

43. 根据权利要求41或权利要求42所述的衔接子核酸序列对,其中所述第二衔接子核酸序列进一步包含第二SDE。

44. 根据权利要求41到43中任一项所述的衔接子核酸序列对,其中所述第二SDE位于所述第二衔接子核酸序列的末端处。

45. 根据权利要求41到44中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列的第二SDE与所述第二衔接子核酸序列的第二SDE至少部分地非互补。

46. 根据权利要求45所述的衔接子核酸序列对,其中所述第一衔接子核酸序列的第二SDE与所述第二衔接子核酸序列的第二SDE不同和/或在至少一个核苷酸处非互补。

47. 根据权利要求46所述的衔接子核酸序列对,其中至少一个核苷酸通过酶反应从所述第一衔接子核酸序列的第二SDE或所述第二衔接子核酸的第二SDE省去。

48. 根据权利要求47所述的衔接子核酸序列对,其中所述酶反应包含聚合酶、核酸内切酶、糖基化酶或裂解酶。

49. 根据权利要求45或权利要求46所述的衔接子核酸序列对,其中所述第一衔接子核酸序列的第二SDE与所述第二衔接子核酸序列的第二SDE非互补。

50. 根据权利要求41到45中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列的SDE直接连接到所述第二衔接子核酸序列的第二SDE。

51. 根据权利要求1到50中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列的引物结合结构域位于第一SDE的5'处。

52. 根据权利要求1到51中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列的第一SDE位于所述SMI结构域的5'处。

53. 根据权利要求1到51中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列的第一SDE位于所述SMI结构域的3'处。

54. 根据权利要求1到53中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列的第一SDE位于所述SMI结构域的5'处且位于所述引物结合结构域的3'处。

55. 根据权利要求1到53中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列的第一SDE位于所述SMI结构域的3'处,所述SMI结构域位于所述引物结合结构域的3'处。

56. 根据权利要求1到54中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列的SMI结构域位于所述连接结构域的5'处。

57. 根据权利要求1到56中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列的3'末端包含所述连接结构域。

58. 根据权利要求1到57中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列从5'到3'包含所述引物结合结构域、所述第一SDE、所述SMI结构域以及所述连接结构域。

59. 根据权利要求1到57中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列从5'到3'包含所述引物结合结构域、所述SMI结构域、所述第一SDE以及所述连接结构域。

60. 根据权利要求1到59中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列或所述第二衔接子核酸序列包含经修饰的核苷酸或非核苷酸分子。

61. 根据权利要求60所述的衔接子核酸序列对,其中所述经修饰的核苷酸或非核苷酸分子为大肠杆菌素E2、Im2、谷胱甘肽、谷胱甘肽-s-转移酶(GST)、镍、聚组氨酸、FLAG-标签、myc-标签或生物素。

62. 根据权利要求61所述的衔接子核酸序列对,其中所述生物素为生物素-16-氨基烯丙基-2'-脱氧尿苷-5'-三磷酸、生物素-16-氨基烯丙基-2'-脱氧胞苷-5'-三磷酸、生物素-16-氨基烯丙基胞苷-5'-三磷酸、N4-生物素-OBFA-2'-脱氧胞苷-5'-三磷酸、生物素-16-氨基烯丙基尿苷-5'-三磷酸、生物素-16-7-脱氮-7-氨基烯丙基-2'-脱氧鸟苷-5'-三磷酸、脱硫生物素-6-氨基烯丙基-2'-脱氧胞苷-5'-三磷酸、5'-生物素-G-单磷酸、5'-生物素-A-单磷酸、5'-生物素-dG-单磷酸或5'-生物素-dA-单磷酸。

63. 根据权利要求61或权利要求62所述的衔接子核酸序列对,其中所述生物素能够与附接到基质上的抗生蛋白链菌素结合。

64. 根据权利要求63所述的衔接子核酸序列对,其中当所述生物素与附接到基质上的抗生蛋白链菌素结合时,所述第一衔接子核酸序列能够与所述第二衔接子核酸序列分离。

65. 根据权利要求1到59中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列或所述第二衔接子核酸序列包含选自小分子、核酸、肽和独特地可结合的部分(其能够与亲和配偶体结合)的亲和标记。

66. 根据权利要求65所述的衔接子核酸序列对,其中当所述亲和配偶体附接到固体基质上且与所述亲和标记结合时,包含所述亲和标记的衔接子核酸序列能够与不包含所述亲和标记的衔接子核酸序列分离。

67. 根据权利要求66所述的衔接子核酸序列对,其中所述固体基质为固体表面、珠粒或另一固定式结构。

68. 根据权利要求65到67中任一项所述的衔接子核酸序列对,其中所述核酸为DNA、RNA

或其组合,且任选地包含肽核酸或锁核酸。

69. 根据权利要求65到68中任一项所述的衔接子核酸序列对,其中所述亲和标记位于衔接子的末端处或与所述第二衔接子核酸序列中的相对结构域不完全互补的所述第一衔接子核酸序列中的结构域内。

70. 根据权利要求1到59中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列或所述第二衔接子核酸序列包含具有磁性特性、电荷特性或不可溶特性的物理基团。

71. 根据权利要求70所述的衔接子核酸序列对,其中当所述物理基团具有磁性特性且施加磁场时,包含所述物理基团的衔接子核酸序列与不包含所述物理基团的衔接子核酸序列分离。

72. 根据权利要求70所述的衔接子核酸序列对,其中当所述物理基团具有电荷特性且施加电场时,包含所述物理基团的衔接子核酸序列与不包含所述物理基团的衔接子核酸序列分离。

73. 根据权利要求70所述的衔接子核酸序列对,其中当所述物理基团具有不可溶特性且所述衔接子核酸序列对含于对于所述物理基团来说不可溶的溶液中时,包含所述物理基团的衔接子核酸序列从不包含所述物理基团的衔接子核酸序列中沉淀出来,保留在溶液中的。

74. 根据权利要求70到73中任一项所述的衔接子核酸序列对,其中所述物理基团位于衔接子的末端处或与所述第二衔接子核酸序列中的相对结构域不完全互补的所述第一衔接子核酸序列中的结构域内。

75. 根据权利要求1到64中任一项所述的衔接子核酸序列对,其中所述第二衔接子核酸序列包含至少一个硫代磷酸酯键。

76. 根据权利要求1到75中任一项所述的衔接子核酸序列对,其中所述双链靶核酸序列为DNA或RNA。

77. 根据权利要求1到76中任一项所述的衔接子核酸序列对,其中每一衔接子核酸序列在其末端中的每一个处包含连接结构域。

78. 根据权利要求1到77中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列或所述第二衔接子核酸序列为至少部分地单链的。

79. 根据权利要求78所述的衔接子核酸序列对,其中所述第一衔接子核酸序列或所述第二衔接子核酸序列为单链的。

80. 根据权利要求79所述的衔接子核酸序列对,其中所述第一衔接子核酸序列和所述第二衔接子核酸序列为单链的。

81. 一种组合物,包含前述权利要求中任一项所述的至少一对衔接子核酸序列和第二对衔接子核酸序列,其中所述第二对衔接子核酸序列的每一链包含至少一个引物结合位点和连接结构域。

82. 根据权利要求81所述的组合物,进一步包含所述第二对衔接子核酸序列的每一链中的SMI结构域。

83. 根据权利要求81或权利要求82所述的组合物,进一步包含所述第二对衔接子核酸序列的每一链中的引物结合位点。

84. 一种组合物,包含前述权利要求中任一项所述的至少两对衔接子核酸序列,其中来

自第一对衔接子核酸序列的第一衔接子核酸序列的SDE与来自至少第二对衔接子核酸序列的第一衔接子核酸序列的SDE不同。

85. 一种组合物, 包含权利要求1到80中任一项所述的至少两对衔接子核酸分子, 其中来自第一对衔接子核酸分子的第一衔接子核酸分子的SMI结构域与来自至少第二对衔接子核酸分子的第一衔接子核酸分子的SMI结构域不同。

86. 根据权利要求85所述的组合物, 其中来自所述第一对单链衔接子核酸分子的第一衔接子核酸分子的SMI结构域与来自所述至少第二对单链衔接子核酸分子的第一单链衔接子核酸分子的SMI结构域长度相同。

87. 根据权利要求85所述的组合物, 其中来自所述第一对单链衔接子核酸分子的第一衔接子核酸分子的SMI结构域具有与来自所述至少第二对单链衔接子核酸分子的第一单链衔接子核酸分子的SMI结构域不同的长度。

88. 根据权利要求85到87中任一项所述的组合物, 其中每一SMI结构域在所述SMI内或侧接所述SMI的位点处包含一个或更多个固定碱基。

89. 一种包含至少一个第一双链复合核酸的组合物, 所述核酸包含: 根据权利要求1到88中任一项所述的第一对衔接子核酸分子, 其连接到双链靶核酸分子的第一末端; 以及根据权利要求1到50中任一项所述的第二对衔接子核酸分子, 其连接到所述双链靶核酸分子的第二末端。

90. 根据权利要求89所述的组合物, 其中所述第一对衔接子核酸分子与所述第二对衔接子核酸分子不同。

91. 根据权利要求90所述的组合物, 其中所述第一对衔接子核酸分子的第一链衔接子标靶核酸分子包含第一SMI结构域, 且所述第二对衔接子核酸分子的第一链衔接子标靶核酸分子包含第二SMI结构域。

92. 根据权利要求89到91中任一项所述的组合物, 包含至少一个第二双链复合核酸。

93. 一种用于测序双链靶核酸分子的衔接子核酸序列对, 包含第一衔接子核酸序列和第二衔接子核酸序列, 其中每一衔接子核酸序列包含:

引物结合结构域, 以及

单个分子标识符 (SMI) 结构域。

94. 根据权利要求93所述的衔接子核酸序列对, 其中所述第一衔接子核酸序列或所述第二衔接子核酸序列中的至少一个进一步包含结构域, 所述结构域包含至少一种经修饰的核苷酸。

95. 根据权利要求93或权利要求94所述的衔接子核酸序列对, 其中所述第一衔接子核酸序列和所述第二衔接子核酸序列进一步包含结构域, 所述结构域包含至少一种经修饰的核苷酸。

96. 根据权利要求93到95中任一项所述的衔接子核酸序列对, 其中所述第一衔接子核酸序列或所述第二衔接子核酸序列中的至少一个进一步包含连接结构域。

97. 根据权利要求93到95中任一项所述的衔接子核酸序列对, 其中所述第一衔接子核酸序列和所述第二衔接子核酸序列包含连接结构域。

98. 根据权利要求94到权利要求97所述的衔接子核酸序列对, 其中所述至少一种经修饰的核苷酸选自无碱基位点; 尿嘧啶; 四氢呋喃; 8-氧代-7,8-二氢-2'-脱氧腺苷 (8-氧代-

A); 8-氧代-7,8-二氢-2'-脱氧鸟苷(8-氧代-G); 脱氧肌苷; 5'-硝基吡啶; 5-羟基甲基-2'-脱氧胞苷; 异胞嘧啶; 5'-甲基-异胞嘧啶; 或异鸟苷。

99. 根据权利要求97到权利要求98中任一项所述的衔接子核酸序列对, 其中所述两个衔接子序列由至少部分地一起退火的两个单独DNA分子组成。

100. 根据权利要求97所述的衔接子核酸序列对, 其中所述第一衔接子核酸序列和所述第二衔接子核酸序列经由连接子结构域连接。

101. 根据权利要求100所述的衔接子核酸序列对, 其中所述连接子结构域由核苷酸构成。

102. 根据权利要求100所述的衔接子核酸序列对, 其中所述连接子结构域含有一种或更多种经修饰的核苷酸或非核苷酸分子。

103. 根据权利要求102所述的衔接子核酸序列对, 其中至少一个种经修饰的核苷酸或非核苷酸分子选自无碱基位点; 尿嘧啶; 四氢呋喃; 8-氧代-7,8-二氢-2'-脱氧腺苷(8-氧代-A); 8-氧代-7,8-二氢-2'-脱氧鸟苷(8-氧代-G); 脱氧肌苷; 5'-硝基吡啶; 5-羟基甲基-2'-脱氧胞苷; 异胞嘧啶; 5'-甲基-异胞嘧啶; 或异鸟苷。

104. 根据权利要求58到61中任一项所述的衔接子核酸序列对, 其中所述连接子结构域形成环。

105. 根据权利要求97到104中任一项所述的衔接子核酸序列对, 其中所述第一衔接子核酸序列的引物结合结构域与所述第二衔接子核酸序列的引物结合结构域至少部分地互补。

106. 根据权利要求97到104中任一项所述的衔接子核酸序列对, 其中所述第一衔接子核酸序列的引物结合结构域与所述第二衔接子核酸序列的引物结合结构域互补。

107. 根据权利要求97到104中任一项所述的衔接子核酸序列对, 其中所述第一衔接子核酸序列的引物结合结构域与所述第二衔接子核酸序列的引物结合结构域非互补。

108. 根据权利要求97到107中任一项所述的衔接子核酸序列对, 其中至少一个SMI结构域为内源性SMI。

109. 根据权利要求108所述的衔接子核酸序列对, 其中所述内源性SMI与剪切点有关。

110. 根据权利要求97到109中任一项所述的衔接子核酸序列对, 其中所述SMI结构域包含至少一种简并或半简并核酸。

111. 根据权利要求97到110中任一项所述的衔接子核酸序列对, 其中所述SMI结构域为非简并的。

112. 根据权利要求97到111中任一项所述的衔接子核酸序列对, 其中所述SMI结构域的序列考虑与对应于连接的DNA的随机或半随机剪切末端的序列结合以获得能够彼此区分单个DNA分子的SMI序列。

113. 根据权利要求97到112中任一项所述的衔接子核酸序列对, 其中所述第一衔接子核酸序列的SMI结构域与所述第二衔接子核酸序列的SMI结构域至少部分地互补。

114. 根据权利要求97到113中任一项所述的衔接子核酸序列对, 其中所述第一衔接子核酸序列的SMI结构域与所述第二衔接子核酸序列的SMI结构域互补。

115. 根据权利要求97到113中任一项所述的衔接子核酸序列对, 其中所述第一衔接子核酸序列的SMI结构域与所述第二衔接子核酸序列的SMI结构域至少部分地非互补。

116. 根据权利要求97到113中任一项所述的衔接子核酸序列对, 其中所述第一衔接子核酸序列的SMI结构域与所述第二衔接子核酸序列的SMI结构域非互补。

117. 根据权利要求97到116中任一项所述的衔接子核酸序列对, 其中每一SMI结构域包含约1个到约30个之间的简并或半简并核酸。

118. 根据权利要求97到117中任一项所述的衔接子核酸序列对, 其中所述第一衔接子核酸序列的连接结构域与所述第二衔接子核酸序列的连接结构域至少部分地互补。

119. 根据权利要求97到118中任一项所述的衔接子核酸序列对, 其中每一连接结构域能够连接到双链靶核酸序列的一个链。

120. 根据权利要求97到119中任一项所述的衔接子核酸序列对, 其中连接结构域中的一个包含T-突出端、A-突出端、CG-突出端、平末端、或另一可连接的核酸序列。

121. 根据权利要求97到120中任一项所述的衔接子核酸序列对, 其中两个连接结构域包含平末端。

122. 根据权利要求97到121中任一项所述的衔接子核酸序列对, 其中每一SMI结构域包含引物结合位点。

123. 根据权利要求97到122中任一项所述的衔接子核酸序列对, 其中至少所述第一衔接子核酸序列进一步包含SDE。

124. 根据权利要求123所述的衔接子核酸序列对, 其中所述第二SDE位于所述第一衔接子核酸序列的末端处。

125. 根据权利要求123或权利要求124所述的衔接子核酸序列对, 其中所述第二衔接子核酸序列进一步包含SDE。

126. 根据权利要求123到124中任一项所述的衔接子核酸序列对, 其中所述SDE位于所述第二衔接子核酸序列的末端处。

127. 根据权利要求123到126中任一项所述的衔接子核酸序列对, 其中所述第一衔接子核酸序列的SDE与所述第二衔接子核酸序列的SDE至少部分地非互补。

128. 根据权利要求127所述的衔接子核酸序列对, 其中所述第一衔接子核酸序列的SDE与所述第二衔接子核酸序列的SDE非互补。

129. 根据权利要求123到128中任一项所述的衔接子核酸序列对, 其中所述第一衔接子核酸序列的SDE直接连接到所述第二衔接子核酸序列的SDE。

130. 根据权利要求123到129中任一项所述的衔接子核酸序列对, 其中所述第一衔接子核酸序列的SDE与所述第二衔接子核酸序列的SDE不同和/或在至少一个核苷酸处非互补。

131. 根据权利要求130所述的衔接子核酸序列对, 其中所述至少一个核苷酸通过酶反应从所述第一衔接子核酸序列的SDE或所述第二衔接子核酸的SDE省去。

132. 根据权利要求131所述的衔接子核酸序列对, 其中所述酶反应包含聚合酶或核酸内切酶。

133. 根据权利要求132所述的衔接子核酸序列对, 其中所述至少一种核苷酸为经修饰的核苷酸或包含标记的核苷酸。

134. 根据权利要求134所述的衔接子核酸序列对, 其中所述经修饰的核苷酸或包含标记的核苷酸选自无碱基位点; 尿嘧啶; 四氢呋喃; 8-氧代-7,8-二氢-2'-脱氧腺苷(8-氧代-A); 8-氧代-7,8-二氢-2'-脱氧鸟苷(8-氧代-G); 脱氧肌苷; 5'-硝基吡啶; 5-羟基甲基-2'-

脱氧胞苷;异胞嘧啶;5'-甲基-异胞嘧啶;或异鸟苷。

135. 根据权利要求123到135中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列的SDE包含能够形成发夹环的自互补结构域。

136. 根据权利要求123到135中任一项所述的衔接子核酸序列对,其中第一衔接子核酸序列远离其连接结构域的末端连接到所述第二衔接子核酸序列远离其连接结构域的末端,由此形成环。

137. 根据权利要求136所述的衔接子核酸序列对,其中所述环包含限制酶识别位点。

138. 根据权利要求93到129中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列的引物结合结构域位于所述SMI结构域的5'处。

139. 根据权利要求93到138中任一项所述的衔接子核酸序列对,其中包含所述第一衔接子核酸序列的至少一种经修饰的核苷酸的结构域位于所述SMI结构域的5'处。

140. 根据权利要求93到138中任一项所述的衔接子核酸序列对,其中包含所述第一衔接子核酸序列的至少一种经修饰的核苷酸的结构域位于所述SMI结构域的3'处。

141. 根据权利要求93到139中任一项所述的衔接子核酸序列对,其中包含所述第一衔接子核酸序列的至少一种经修饰的核苷酸的结构域位于所述SMI结构域的5'处且位于所述引物结合结构域的3'处。

142. 根据权利要求93到138中任一项所述的衔接子核酸序列对,其中包含所述第一衔接子核酸序列的至少一种经修饰的核苷酸的结构域位于所述SMI结构域的3'处,所述SMI结构域位于所述引物结合结构域的3'处。

143. 根据权利要求93到141中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列的SMI结构域位于所述连接结构域的5'处。

144. 根据权利要求93到143中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列的3'末端包含所述连接结构域。

145. 根据权利要求93到143中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列从5'到3'包含所述引物结合结构域、所述包含至少一种经修饰的核苷酸的结构域、所述SMI结构域以及所述连接结构域。

146. 根据权利要求93到138中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列从5'到3'包含所述引物结合结构域、所述SMI结构域、所述包含至少一种经修饰的核苷酸的结构域以及所述连接结构域。

147. 根据权利要求93到146中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列或所述第二衔接子核酸序列包含经修饰的核苷酸或非核苷酸分子。

148. 根据权利要求147所述的衔接子核酸序列对,其中所述经修饰的核苷酸或非核苷酸分子为大肠杆菌素E2、Im2、谷胱甘肽、谷胱甘肽-s-转移酶(GST)、镍、聚组氨酸、FLAG-标签、myc-标签或生物素。

149. 根据权利要求148所述的衔接子核酸序列对,其中生物素为生物素-16-氨基烯丙基-2'-脱氧尿苷-5'-三磷酸、生物素-16-氨基烯丙基-2'-脱氧胞苷-5'-三磷酸、生物素-16-氨基烯丙基胞苷-5'-三磷酸、N4-生物素-OBFA-2'-脱氧胞苷-5'-三磷酸、生物素-16-氨基烯丙基尿苷-5'-三磷酸、生物素-16-7-脱氮-7-氨基烯丙基-2'-脱氧鸟苷-5'-三磷酸、脱硫生物素-6-氨基烯丙基-2'-脱氧胞苷-5'-三磷酸、5'-生物素-G-单磷酸、5'-生物素-A-单

磷酸、5'-生物素-dG-单磷酸或5'-生物素-dA-单磷酸。

150. 根据权利要求148或权利要求149所述的衔接子核酸序列对,其中所述生物素能够与附接到基质上的抗生蛋白链菌素结合。

151. 根据权利要求150所述的衔接子核酸序列对,其中当所述生物素与附接到基质上的抗生蛋白链菌素结合时,所述第一衔接子核酸序列能够与所述第二衔接子核酸序列分离。

152. 根据权利要求93到148中任一项所述的衔接子核酸序列对,其中所述第二衔接子核酸序列包含至少一个硫代磷酸酯键。

153. 根据权利要求93到152中任一项所述的衔接子核酸序列对,其中所述双链靶核酸序列为DNA或RNA。

154. 根据权利要求93到153中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列或所述第二衔接子核酸序列包含选自小分子、核酸、肽和独特地可结合的部分(其能够与亲和配偶体结合)的亲和标记。

155. 根据权利要求154所述的衔接子核酸序列对,其中当所述亲和配偶体附接到固体基质上且与所述亲和标记结合时,包含所述亲和标记的衔接子核酸序列能够与不包含所述亲和标记的衔接子核酸序列分离。

156. 根据权利要求155所述的衔接子核酸序列对,其中所述固体基质为固体表面、珠粒或另一固定式结构。

157. 根据权利要求154到156中任一项所述的衔接子核酸序列对,其中所述核酸为DNA、RNA或其组合,且任选地包含肽核酸或锁核酸。

158. 根据权利要求154到157中任一项所述的衔接子核酸序列对,其中所述亲和标记位于衔接子的末端处或与所述第二衔接子核酸序列中的相对结构域不完全互补的所述第一衔接子核酸序列中的结构域内。

159. 根据权利要求93到158中任一项所述的衔接子核酸序列对,其中所述第一衔接子核酸序列或所述第二衔接子核酸序列包含具有磁性特性、电荷特性或不可溶特性的物理基团。

160. 根据权利要求159所述的衔接子核酸序列对,其中当所述物理基团具有磁性特性且施加磁场时,包含所述物理基团的衔接子核酸序列与不包含所述物理基团的衔接子核酸序列分离。

161. 根据权利要求159所述的衔接子核酸序列对,其中当所述物理基团具有电荷特性且施加电场时,包含所述物理基团的衔接子核酸序列与不包含所述物理基团的衔接子核酸序列分离。

162. 根据权利要求157所述的衔接子核酸序列对,其中当所述物理基团具有不可溶特性且所述衔接子核酸序列对含于对于所述物理基团来说不可溶的溶液中时,包含所述物理基团的衔接子核酸序列沉淀远离保留在溶液中的不包含所述物理基团的衔接子核酸序列。

163. 根据权利要求157到162中任一项所述的衔接子核酸序列对,其中所述物理基团位于衔接子的末端处或与所述第二衔接子核酸序列中的相对结构域不完全互补的所述第一衔接子核酸序列中的结构域内。

164. 根据权利要求93到163中任一项所述的衔接子核酸序列对,其中所述第一衔接子

核酸序列或所述第二衔接子核酸序列为至少部分地单链的。

165. 根据权利要求163所述的衔接子核酸序列对, 其中所述第一衔接子核酸序列或所述第二衔接子核酸序列为单链的。

166. 根据权利要求165所述的衔接子核酸序列对, 其中所述第一衔接子核酸序列和所述第二衔接子核酸序列为单链的。

167. 根据权利要求96到166中任一项所述的衔接子核酸序列对, 其中所述连接结构域中的至少一个包含去羟基化碱基。

168. 根据权利要求96到166中任一项所述的衔接子核酸序列对, 其中所述连接结构域中的至少一个已经化学修饰以便使其不可连接。

169. 一种组合物, 包含权利要求93到168中任一项所述的至少两对衔接子核酸分子, 其中来自第一对衔接子核酸分子的第一衔接子核酸分子的SMI结构域与来自至少第二对衔接子核酸分子的第一衔接子核酸分子的SMI结构域不同。

170. 根据权利要求169所述的组合物, 其中来自所述第一对单链衔接子核酸分子的第一衔接子核酸分子的SMI结构域与来自所述至少第二对单链衔接子核酸分子的第一单链衔接子核酸分子的SMI结构域长度相同。

171. 根据权利要求169所述的组合物, 其中来自所述第一对单链衔接子核酸分子的第一衔接子核酸分子的SMI结构域具有与来自所述至少第二对单链衔接子核酸分子的第一单链衔接子核酸分子的SMI结构域不同的长度。

172. 根据权利要求169到171中任一项所述的组合物, 其中每一SMI结构域在所述SMI内或侧接所述SMI的位点处包含一个或更多个固定碱基。

173. 一种包含至少一个第一双链复合核酸的组合物, 所述核酸包含: 根据权利要求93到168中任一项所述的第一对衔接子核酸分子, 其连接到双链靶核酸分子的第一末端; 以及根据权利要求93到168中任一项所述的第二对衔接子核酸分子, 其连接到所述双链靶核酸分子的第二末端。

174. 根据权利要求173所述的组合物, 其中所述第一对衔接子核酸分子与所述第二对衔接子核酸分子不同。

175. 根据权利要求174所述的组合物, 其中所述第一对衔接子核酸分子的第一链衔接子标靶核酸分子包含第一SMI结构域, 且所述第二对衔接子核酸分子的第一链衔接子标靶核酸分子包含第二SMI结构域。

176. 根据权利要求175所述的组合物, 其中所述第一对衔接子核酸分子的第一链衔接子标靶核酸分子包含第一SMI结构域, 且所述第二对衔接子核酸分子的第一链衔接子标靶核酸分子包含第二SMI结构域。

177. 根据权利要求173到176中任一项所述的组合物, 包含至少一个第二双链复合核酸。

178. 一种组合物, 包含根据权利要求1到80中任一项所述的至少一对衔接子核酸分子和根据权利要求93到168中任一项所述的至少一对衔接子核酸分子。

179. 一种包含至少一个第一双链复合核酸的组合物, 所述核酸包含: 根据权利要求1到80中任一项所述的第一对衔接子核酸分子, 其连接到双链靶核酸分子的第一末端; 以及根据权利要求93到168中任一项所述的第二对衔接子核酸分子, 其连接到所述双链靶核酸分

子的第二末端。

180. 一种测序双链靶核酸的方法, 包含以下步骤:

(1) 使根据权利要求1到80中任一项所述的衔接子核酸序列对连接到双链靶核酸分子的至少一个末端, 由此形成包含第一链衔接子靶核酸序列和第二链衔接子靶核酸序列的双链核酸分子;

(2) 使所述第一链衔接子靶核酸序列扩增, 由此产生包含多种第一链衔接子靶核酸序列和其多种互补分子的第一组扩增产物;

(3) 使所述第二链衔接子靶核酸序列扩增, 由此产生包含多种第二链衔接子靶核酸序列和其多种互补分子的第二组扩增产物, 其中所述第二组扩增产物可与所述第一组扩增产物区分开来;

(4) 测序所述第一组扩增产物; 以及

(5) 测序所述第二组扩增产物。

181. 根据权利要求180所述的方法, 其中所述至少一个末端为两个末端。

182. 根据权利要求180或权利要求181所述的方法, 其中扩增通过PCR进行。

183. 根据权利要求180所述的方法, 其中扩增通过多重置换扩增进行。

184. 根据权利要求180所述的方法, 其中扩增通过等温扩增进行。

185. 根据权利要求180到184所述的方法, 其中连接到所述双链靶核酸序列的第一末端的衔接子核酸序列对具有与连接到所述双链靶核酸序列的第二末端的衔接子核酸序列对相同的结构。

186. 根据权利要求180或权利要求185所述的方法, 其中所述第一链衔接子靶核酸序列以5'到3'顺序包含:

(a) 第一衔接子核酸序列;

(b) 所述双链靶核酸的第一链; 以及

(c) 第二衔接子核酸序列。

187. 根据权利要求180到186中任一权利要求所述的方法, 其中所述第二链衔接子靶核酸序列以3'到5'顺序包含:

(a) 第一衔接子核酸序列;

(b) 所述双链靶核酸的第二链; 以及

(c) 第二衔接子核酸序列。

188. 根据权利要求180所述的方法, 其中连接到所述双链靶核酸序列的第一末端的衔接子核酸序列对与连接到所述双链靶核酸序列的第二末端的衔接子核酸序列对不同。

189. 根据权利要求188所述的方法, 其中连接到所述双链靶核酸序列的第一末端的衔接子核酸序列对具有第一SMI结构域, 且连接到所述双链靶核酸序列的第二末端的衔接子核酸序列对具有第二SMI结构域;

其中所述第一SMI结构域可与所述第二SMI结构域不同。

190. 根据权利要求188或权利要求189所述的方法, 其中所述第一链衔接子靶核酸序列以5'到3'顺序包含:

(a) 包含所述第一SDE的第一衔接子核酸序列;

(b) 第一SMI结构域;

(c) 所述双链靶核酸的第一链;以及

(d) 第二衔接子核酸序列。

191. 根据权利要求188到190中任一权利要求所述的方法,其中所述第二链衔接子靶核酸序列以5'到3'顺序包含:

(a) 包含所述第一SDE的第一衔接子核酸序列;

(b) 第二SMI结构域;

(c) 所述双链靶核酸的第二链;以及

(d) 第二衔接子核酸序列。

192. 根据权利要求180到191中任一权利要求所述的方法,其中将所述第一组扩增产物的共有序列与所述第二组扩增产物的共有序列进行比较且所述两个共有序列之间的差异视为伪差。

193. 一种测序双链靶核酸的方法,包含以下步骤:

(1) 使根据权利要求93到168中任一项所述的衔接子核酸序列对连接到双链靶核酸分子的至少一个末端,由此形成包含第一链衔接子靶核酸序列和第二链衔接子靶核酸序列的双链核酸分子;

(2) 使所述第一链衔接子靶核酸分子扩增,由此产生包含多种第一链衔接子靶核酸分子和其多种互补分子的第一组扩增产物;

(3) 使所述第二链衔接子靶核酸分子扩增,由此产生包含多种第二链衔接子靶核酸分子和其多种互补分子的第二组扩增产物;

(4) 测序所述第一组扩增产物,由此获得所述第一组扩增产物的共有序列;以及

(5) 测序所述第二组扩增产物,由此获得所述第二组扩增产物的共有序列。

194. 根据权利要求193所述的方法,其中所述第二组扩增产物可与所述第一组扩增产物区分开来。

195. 根据权利要求193或权利要求194所述的方法,其中扩增通过PCR进行。

196. 根据权利要求193或权利要求194所述的方法,其中扩增通过多重置换扩增进行。

197. 根据权利要求193或权利要求194所述的方法,其中扩增通过等温扩增进行。

198. 根据权利要求193或权利要求194所述的方法,进一步包含在步骤(1)之后使所述双链核酸分子与将所述至少一种经修饰的核苷酸改变成另一化学结构的至少一种酶接触的步骤。

199. 根据权利要求198所述的方法,其中所述酶为糖基化酶。

200. 根据权利要求193到199所述的方法,其中连接到所述双链靶核酸分子的第一末端的衔接子核酸序列对与连接到所述双链靶核酸分子的第二末端的衔接子核酸序列对一致。

201. 根据权利要求193到199所述的方法,其中连接到所述双链靶核酸分子的第一末端的衔接子核酸序列对与连接到所述双链靶核酸分子的第二末端的衔接子核酸序列对不同。

202. 根据权利要求193到199所述的方法,其中衔接子核酸序列对连接到双链靶核酸分子的第一末端,且利用对应于靶DNA分子的DNA序列的一部分的引物以使所述DNA分子扩增。

203. 根据权利要求193到200中任一权利要求所述的方法,其中所述第一链衔接子靶核酸序列以5'到3'顺序包含:

(a) 第一衔接子核酸序列,其包含所述至少一种经修饰的核苷酸或所述至少一个无碱

基位点；

- (b) 所述双链靶核酸的第一链；以及
- (c) 第二衔接子核酸序列。

204. 根据权利要求193到203中任一权利要求所述的方法，其中所述第二链衔接子靶核酸序列以3'到5'顺序包含：

- (a) 第一衔接子核酸序列；
- (b) 所述双链靶核酸的第二链；以及
- (c) 第二衔接子核酸序列。

205. 根据权利要求193所述的方法，其中连接到所述双链靶核酸分子的第一末端的衔接子核酸序列对与连接到所述双链靶核酸分子的第二末端的衔接子核酸序列对不同。

206. 根据权利要求205所述的方法，其中连接到所述双链靶核酸分子的第一末端的衔接子核酸序列对具有第一SMI结构域，且连接到所述双链靶核酸序列的第二末端的衔接子核酸序列对具有第二SMI结构域；

其中所述第一SMI结构域与所述第二SMI结构域不同。

207. 根据权利要求205或权利要求206所述的方法，其中所述第一链衔接子靶核酸序列以5'到3'顺序包含：

(a) 第一衔接子核酸序列，其包含所述至少一种经修饰的核苷酸或所述至少一个无碱基位点和所述第一SMI结构域；

- (b) 所述双链靶核酸的第一链；以及
- (c) 包含所述第二SMI结构域的第二衔接子核酸序列。

208. 根据权利要求207所述的方法，其中当所述至少一种经修饰的核苷酸为8-氧代-G时，且所述第二衔接子核酸序列在对应于所述8-氧代-G的位置处包括胞嘧啶。

209. 根据权利要求205到208中任一权利要求所述的方法，其中所述第二链衔接子靶核酸序列以3'到5'顺序包含：

- (a) 包含所述第一SMI结构域的第一衔接子核酸序列；
- (b) 所述双链靶核酸的第二链；以及
- (c) 包含所述第二SMI结构域的第二衔接子核酸序列。

210. 根据权利要求209所述的方法，其中当所述至少一种经修饰的核苷酸为8-氧代-G时，所述第二衔接子核酸序列在对应于所述8-氧代-G的位置处包含胞苷。

211. 一种方法，其中可区分的扩增产物获自个别DNA分子的两个链中的每一个，且将所述第一组扩增产物的共有序列与所述第二组扩增产物的共有序列进行比较，其中所述两个共有序列之间的差异可视为伪差。

212. 根据权利要求211所述的方法，其中借助于共用相同SMI序列，测定出扩增产物从相同初始DNA分子处产生。

213. 根据权利要求211所述的方法，其中借助于携带已知彼此对应的相异SMI序列，基于在SMI衔接子库合成时产生且与其结合的数据库，测定出扩增产物从相同初始DNA分子处产生。

214. 根据权利要求211到213中任一权利要求所述的方法，其中经由通过SDE引入的至少一个核苷酸序列差异，测定出扩增产物从相同初始双链DNA序列的相异链处产生。

215. 一种方法, 其中可区分的扩增产物获自个别DNA分子的两个链中的每一个, 且将获自对应于单个DNA分子的两个初始DNA链中的一个的扩增产物的序列与对应于所述两个初始DNA链中的第二个的扩增产物, 且所述两个序列之间的差异视为伪差。

216. 一种方法, 其中当将获自对应于单个DNA分子的两个初始DNA链中的一个的扩增产物的序列与对应于两个初始DNA链中的第二个的扩增产物进行比较且未鉴别出所述两个序列之间的差异时, 不可区分的扩增产物获自个别DNA分子的两个链。

217. 根据权利要求215或权利要求216所述的方法, 其中借助于共用相同SMI序列, 基于在SMI衔接子库合成时产生且与其结合的数据库, 测定出扩增产物从相同初始双链DNA分子处产生。

218. 根据权利要求214到217所述的方法, 其中经由通过SDE引入的至少一个核苷酸序列差异, 测定出扩增产物从相同初始双链DNA序列的相异链处产生。

219. 根据权利要求214到218所述的方法, 进一步包含在DNA双螺旋体热或化学熔融成其组分单链之后单分子稀释的步骤。

220. 根据权利要求219所述的方法, 其中所述单链稀释到多个以物理方式隔开的反应腔室中以使得两个共用相同容器的最初配对链的机率较小。

221. 根据权利要求220所述的方法, 其中所述以物理方式隔开的反应腔室选自容器、套管、孔和至少一对非连通液滴。

222. 根据权利要求220所述的方法, 其中对于每一以物理方式隔开的反应腔室, 进行PCR扩增, 优选针对携带不同标签序列的每一腔室使用引物。

223. 根据权利要求222所述的方法, 其中每一标签序列充当SDE。

224. 根据权利要求216到223所述的方法, 其中对应于相同初始DNA的两个链的一系列配对序列彼此进行比较, 且选择一系列产物的至少一个序列作为最可能代表初始DNA分子的正确序列。

225. 根据权利要求224所述的方法, 其中至少部分地归因于具有获自所述两个DNA链的产物之间的最小错配数, 选择最可能代表所述初始DNA分子的正确序列的所选产物。

226. 根据权利要求224或权利要求225所述的方法, 其中至少部分地归因于相对于参考序列具有最小错配数, 选择最可能代表所述初始DNA分子的正确序列的所选产物。

227. 根据权利要求203到210中任一权利要求所述的方法, 其中在步骤(2)或步骤(3)的扩增期间, 所述至少一个无碱基位点在扩增后转化成相应的扩增产物中的胸苷, 从而引入SDE。

228. 根据权利要求193所述的方法, 其中在步骤(2)或步骤(3)的扩增期间, 所述至少一个经修饰的核苷酸位点编码所述相应的扩增产物中的腺苷。

229. 一种包含至少两对衔接子核酸序列的组合物, 其中第一对衔接子核酸序列包含:

引物结合结构域,

链限定元件(SDE), 以及

连接结构域;

其中第二对衔接子核酸序列包含:

引物结合结构域,

单个分子标识符(SMI)结构域, 以及

连接结构域。

230. 一种双链复合核酸, 包含:

(1) 第一对衔接子核酸序列, 其包含:

引物结合结构域, 以及

SDE, 以及

(2) 双链靶核酸; 以及

(3) 第二对衔接子核酸序列, 其包含:

引物结合结构域, 以及

单个分子标识符 (SMI) 结构域

其中所述第一对衔接子核酸分子连接到所述双链靶核酸分子的第一末端, 且所述第二对衔接子核酸分子连接到所述双链靶核酸分子的第二末端。

231. 根据权利要求230所述的双链复合核酸, 其中所述第一对衔接子核酸序列和/或所述第二对衔接子核酸序列进一步包含连接结构域。

232. 一种用于测序双链靶核酸分子的衔接子核酸序列对, 包含第一衔接子核酸序列和第二衔接子核酸序列, 其中每一衔接子核酸序列包含:

引物结合结构域,

SDE,

连接结构域;

其中所述第一衔接子核酸序列的SDE与所述第二衔接子核酸序列的SDE至少部分地非互补。

233. 一种双链环形核酸, 包含根据权利要求1到80中任一项所述的连接到双链靶核酸分子的第一末端且连接到所述双链靶核酸分子的第二末端的衔接子核酸分子对。

234. 一种双链环形核酸, 包含连接到双链靶核酸分子的第一末端且连接到所述双链靶核酸分子的第二末端的根据权利要求93到168中任一项所述的衔接子核酸分子对。

235. 一种双链环形核酸, 包含连接到双链靶核酸分子的第一末端的根据权利要求1到80中任一项所述的衔接子核酸分子对和连接到所述双链靶核酸分子的第二末端的退火的引物结合结构域对;

其中所述退火的引物结合结构域对连接到所述衔接子核酸分子对。

236. 一种双链环形核酸, 包含连接到双链靶核酸分子的第一末端的根据权利要求93到168中任一项所述的衔接子核酸分子对和连接到所述双链靶核酸分子的第二末端的退火的引物结合结构域对;

其中所述退火的引物结合结构域对连接到所述衔接子核酸分子对。

237. 一种双链复合核酸, 包含:

(1) 衔接子核酸序列对, 其包含:

引物结合结构域,

链限定元件 (SDE), 以及

单个分子标识符 (SMI) 结构域;

(2) 双链靶核酸; 以及

(3) 退火的引物结合结构域对;

其中所述衔接子核酸分子对连接到所述双链靶核酸分子的第一末端且所述退火的引物结合结构域对连接到所述双链靶核酸分子的第二末端。

238. 根据权利要求237所述的双链复合核酸, 其中所述衔接子核酸序列对和/或所述退火的引物结合结构域对进一步包含连接结构域。

用于双重测序的改良衔接子、方法和组合物

[0001] 相关申请的交叉引用

[0002] 本申请要求2015年12月8日提交的美国临时申请第62/264,822号和2016年1月22日提交的美国临时申请第62/281,917号的优先权和权益。上文提及的申请中的每一个以全文引用的方式并入本文中。

[0003] 序列表

[0004] 本申请含有一个序列表,其已经由EFS-Web以ASCII格式提交且以全文引用的方式并入本文中。创建于2016年12月8日的所述ASCII拷贝名称为TWIN-001_ST25.txt且大小为11,778个字节。

背景技术

[0005] 双重测序能够通过单独地扩增和测序双螺旋体DNA的两个链来实现高通量DNA测序的准确度的极大改良;因此,扩增和测序错误可在其将典型地存在于两个链中的仅一个上时消除。双重测序首先用引入到与DNA片段末端连接的Y形或“环形”衔接子中的不对称(即,非互补)PCR引物结合位点描述。衔接子自身内存在的不对称引物结合位点从两个DNA链中产生单独产物,其能够使得两个DNA链中的每一个的错误校正。在一些情况下,使用不对称引物结合位点可能不是最佳的;例如Y衔接子的自由端可能倾向于通过核酸外切酶降解,且这些自由端也可退火到其它分子,产生分子的“菊花链”。此外,利用Y形衔接子或“环形”衔接子的双重测序最容易用双端测序途径应用;适用于单端测序的替代途径将简化多种测序平台上双重测序的较宽应用。

[0006] 因此,存在并不涉及使用不对称引物结合位点的双重测序途径的未被满足的需要。

发明内容

[0007] 本文中描述并不需要使用不对称引物结合位点的双重测序的替代和优良途径。实际上,两个链之间的不对称性可通过产生待测序DNA分子中的衔接子或其它地方内的两个链之间的DNA序列中的至少一个核苷酸不同或通过以其它方法区别标记两个链,如使分子附接到所述链中的至少一个(其使得两个链能够物理分离)引入。

[0008] 在第一方面中,本发明涉及用于测序双链靶核酸分子的衔接子核酸序列对,其包括第一衔接子核酸序列和第二衔接子核酸序列,其中每一衔接子核酸序列包括引物结合结构域、链限定元件(SDE)、单个分子标识符(SMI)结构域和连接结构域。第一衔接子核酸序列的SDE可与第二衔接子核酸序列的SDE至少部分地非互补。

[0009] 在第一方面的实施例中,两个衔接子序列可包括两个至少部分地一起退火的单独DNA分子。第一衔接子核酸序列和第二衔接子核酸序列可经由连接子结构域连接。连接子结构域可由核苷酸构成。连接子结构域可包括一个或多个经修饰的核苷酸或非核苷酸分子。一个或多个经修饰的核苷酸或非核苷酸分子可为无碱基位点、尿嘧啶、四氢呋喃、8-氧代-7,8-二氢-2'-脱氧腺苷(8-氧代-A)、8-氧代-7,8-二氢-2'-脱氧鸟苷(8-氧代-G)、脱氧肌

昔、5'-硝基吡啶、5-羟基甲基-2'-脱氧胞昔、异胞嘧啶、5'-甲基-异胞嘧啶或异鸟昔。连接子结构域可形成环。第一衔接子核酸序列的SDE可与第二衔接子核酸序列的SDE非互补。第一衔接子核酸序列的引物结合结构域可与第二衔接子核酸序列的引物结合结构域至少部分地互补。在实施例中,第一衔接子核酸序列的引物结合结构域可与第二衔接子核酸序列的引物结合结构域互补。第一衔接子核酸序列的引物结合结构域可与第二衔接子核酸序列的引物结合结构域至少部分地非互补。在实施例中,至少一个SMI结构域可为内源性SMI,例如涉及剪切点(例如使用剪切点自身,使用剪切点的实际映射位置(例如染色体3,位置1、234、567),使用紧邻剪切点的DNA中规定数目的核苷酸(例如距离剪切点十个核苷酸,远离剪切点开始七个核苷酸的八个核苷酸,且在“C”首次出现之后,在剪切点之后开始的六个核苷酸))。在实施例中,SMI结构域包括至少一种简并或半简并核酸。在实施例中,SMI结构域可为非简并的。在实施例中,SMI结构域的序列可考虑与对应于连接的DNA的随机或半随机剪切末端的序列结合以获得能够彼此区分单个DNA分子的SMI序列。第一衔接子核酸序列的SMI结构域可与第二衔接子核酸序列的SMI结构域至少部分地互补。第一衔接子核酸序列的SMI结构域可与第二衔接子核酸序列的SMI结构域互补。第一衔接子核酸序列的SMI结构域可与第二衔接子核酸序列的SMI结构域至少部分地非互补。在实施例中,每一SMI结构域包括引物结合位点。在实施例中,每一SMI结构域可位于其连接结构域远端。第一衔接子核酸序列的SMI结构域可与第二衔接子核酸序列的SMI结构域非互补。在实施例中,每一SMI结构域包括约1个至约30个之间的简并或半简并核酸。第一衔接子核酸序列的连接结构域可与第二衔接子核酸序列的连接结构域至少部分地互补。在实施例中,每一连接结构域可能连接到双链靶核酸序列的一个链。在实施例中,连接结构域中的一个包括T-突出端、A-突出端、CG-突出端、平末端或另一可连接的核酸序列。在实施例中,两个连接结构域包含平末端。在实施例中,连接结构域中的至少一个包括经修饰的核酸。经修饰的核苷酸可为无碱基位点、尿嘧啶、四氢呋喃、8-氧代-7,8-二氢-2'-脱氧腺昔(8-氧代-A)、8-氧代-7,8-二氢-2'-脱氧鸟昔(8-氧代-G)、脱氧肌昔、5'-硝基吡啶、5-羟基甲基-2'-脱氧胞昔、异胞嘧啶、5'-甲基-异胞嘧啶或异鸟昔。在实施例中,连接结构域中的至少一个包括去磷酸化碱基。在实施例中,连接结构域中的至少一个包括去羟基化碱基。在实施例中,连接结构域中的至少一个已经化学修饰以便使其不可连接。第一衔接子核酸序列的SDE与来自第二衔接子核酸序列的SDE的至少一个核苷酸不同和/或可与其非互补。在实施例中,通过酶反应,至少一种核苷酸可从第一衔接子核酸序列的SDE或第二衔接子核酸的SDE省去。酶反应包括聚合酶、核酸内切酶、糖基化酶或裂解酶。至少一种核苷酸可为经修饰的核苷酸或包括标记的核苷酸。经修饰的核苷酸或包括标记的核苷酸可为无碱基位点、尿嘧啶、四氢呋喃、8-氧代-7,8-二氢-2'-脱氧腺昔(8-氧代-A)、8-氧代-7,8-二氢-2'-脱氧鸟昔(8-氧代-G)、脱氧肌昔、5'-硝基吡啶、5-羟基甲基-2'-脱氧胞昔、异胞嘧啶、5'-甲基-异胞嘧啶或异鸟昔。第一衔接子核酸序列的SDE包括可能形成发夹环的自互补结构域。第一衔接子核酸序列远离其连接结构域的末端可连接到第二衔接子核酸序列可能远离其连接结构域的末端,由此形成环。环包括限制酶识别位点。在实施例中,至少第一衔接子核酸序列进一步包括第二SDE。第二SDE可位于第一衔接子核酸序列末端处。第二衔接子核酸序列进一步包括第二SDE。第二SDE可位于第二衔接子核酸序列末端处。第一衔接子核酸序列的第二SDE可与第二衔接子核酸序列的第二SDE至少部分地非互补。第一衔接子核酸序列的第二SDE与来自第二衔接子核酸

序列的第二SDE的至少一个核苷酸不同和/或可与其非互补。在实施例中,通过酶反应,至少一种核苷酸可从第一衔接子核酸序列的第二SDE或第二衔接子核酸的第二SDE省去。酶反应包括聚合酶、核酸内切酶、糖基化酶或裂解酶。第一衔接子核酸序列的第二SDE可与第二衔接子核酸序列的第二SDE非互补。第一衔接子核酸序列的SDE可直接连接到第二衔接子核酸序列的第二SDE。第一衔接子核酸序列的引物结合结构域可位于第一SDE的5'处。第一衔接子核酸序列的第一SDE可位于SMI结构域的5'处。第一衔接子核酸序列的第一SDE可位于SMI结构域的3'处。第一衔接子核酸序列的第一SDE可位于SMI结构域的5'处且可位于引物结合结构域的3'处。第一衔接子核酸序列的第一SDE可位于SMI结构域的3'处(其可位于引物结合结构域的3'处)。第一衔接子核酸序列的SMI结构域可位于连接结构域的5'处。第一衔接子核酸序列的3'末端包括连接结构域。第一衔接子核酸序列从5'到3'包括引物结合结构域、第一SDE、SMI结构域和连接结构域。第一衔接子核酸序列从5'到3'包括引物结合结构域、SMI结构域、第一SDE和连接结构域。在实施例中,第一衔接子核酸序列或第二衔接子核酸序列包括经修饰的核苷酸或非核苷酸分子。经修饰的核苷酸或非核苷酸分子可为大肠杆菌素(Colicin) E2、Im2、谷胱甘肽、谷胱甘肽-s-转移酶(GST)、镍、聚组氨酸、FLAG-标签、myc-标签或生物素。生物素可为生物素-16-氨基烯丙基-2'-脱氧尿苷-5'-三磷酸、生物素-16-氨基烯丙基-2'-脱氧胞苷-5'-三磷酸、生物素-16-氨基烯丙基胞苷-5'-三磷酸、N4-生物素-OBEA-2'-脱氧胞苷-5'-三磷酸、生物素-16-氨基烯丙基尿苷-5'-三磷酸、生物素-16-7-脱氮-7-氨基烯丙基-2'-脱氧鸟苷-5'-三磷酸、脱硫生物素-6-氨基烯丙基-2'-脱氧胞苷-5'-三磷酸、5'-生物素-G-单磷酸、5'-生物素-A-单磷酸、5'-生物素-dG-单磷酸或5'-生物素-dA-单磷酸。生物素可能能够与附接到基质上的抗生蛋白链菌素结合。在实施例中,当生物素与附接到基质上的抗生蛋白链菌素结合时,第一衔接子核酸序列能够与第二衔接子核酸序列分离。在实施例中,第一衔接子核酸序列或第二衔接子核酸序列包括选自小分子、核酸、肽和独特地可结合的部分(其可能能够与亲和配偶体结合)的亲和标记。在实施例中,当亲和配偶体附接到固体基质上且与亲和标记结合时,包括亲和标记的衔接子核酸序列能够与不包括亲和标记的衔接子核酸序列分离。固体基质可为固体表面、珠粒或另一固定式结构。核酸可为DNA、RNA或其组合,且任选地包括肽核酸或锁核酸。亲和标记可位于衔接子的末端处或可能与第二衔接子核酸序列中的相对结构域不完全互补的第一衔接子核酸序列中的结构域内。在实施例中,第一衔接子核酸序列或第二衔接子核酸序列包括具有磁性特性、电荷特性或不可溶特性的物理基团。在实施例中,当物理基团具有磁性特性且施加磁场时,包括物理基团的衔接子核酸序列与不包括物理基团的衔接子核酸序列分离。在实施例中,当物理基团具有电荷特性且施加电场时,包括物理基团的衔接子核酸序列与不包括物理基团的衔接子核酸序列分离。在实施例中,当物理基团具有不可溶特性且衔接子核酸序列对含于对于物理基团来说不可溶的溶液中时,包括物理基团的衔接子核酸序列从不包括物理基团的衔接子核酸序列中沉淀出来,残留于溶液中。物理基团可位于衔接子的末端处或可能与第二衔接子核酸序列中的相对结构域不完全互补的第一衔接子核酸序列中的结构域内。第二衔接子核酸序列包括至少一个硫代磷酸酯键。双链靶核酸序列可为DNA或RNA。在实施例中,每一衔接子核酸序列在其末端中的每一个处包括连接结构域。第一衔接子核酸序列或第二衔接子核酸序列可为至少部分单链的。第一衔接子核酸序列或第二衔接子核酸序列可为单链的。第一衔接子核酸序列和第二衔接子核酸序列可为单链的。

[0010] 在第二方面中,本发明涉及一种包括第一方面的至少一对衔接子核酸序列和第二对衔接子核酸序列的组合物,其中第二对衔接子核酸序列的每一链包括至少一个引物结合位点和连接结构域。

[0011] 第二方面进一步涉及一种包括第一方面的至少两对衔接子核酸序列的组合物,其中来自第一对衔接子核酸序列的第一衔接子核酸序列的SDE与来自至少第二对衔接子核酸序列的第一衔接子核酸序列的SDE不同。

[0012] 第二方面还涉及一种包括第一方面的至少两对衔接子核酸分子的组合物,其中来自第一对衔接子核酸分子的第一衔接子核酸分子的SMI结构域与来自至少第二对衔接子核酸分子的第一衔接子核酸分子的SMI结构域不同。

[0013] 在第二方面的实施例中,所述组合物进一步包括第二对衔接子核酸序列的每一链中的SMI结构域。所述组合物可进一步包括第二对衔接子核酸序列的每一链中的引物结合位点。来自第一对单链衔接子核酸分子的第一衔接子核酸分子的SMI结构域可与来自至少第二对单链衔接子核酸分子的第一单链衔接子核酸分子的SMI结构域长度相同。来自第一对单链衔接子核酸分子的第一衔接子核酸分子的SMI结构域可具有与来自至少第二对单链衔接子核酸分子的第一单链衔接子核酸分子的SMI结构域不同的长度。在实施例中,每一SMI结构域在SMI内或侧接SMI的位点处包括一个或多个固定碱基。在实施例中,至少一个包括第一方面的第一对衔接子核酸分子的第一双链复合核酸连接到双链靶核酸分子的第一末端且第一方面的第二对衔接子核酸分子连接到双链靶核酸分子的第二末端。第一对衔接子核酸分子可与第二对衔接子核酸分子不同。第一对衔接子核酸分子的第一链衔接子靶核酸分子包括第一SMI结构域且第二对衔接子核酸分子的第一链衔接子靶核酸分子包括第二SMI结构域。在实施例中,所述组合物包括至少第二双链复合核酸。

[0014] 在第三方面中,本发明涉及一对用于测序包括第一衔接子核酸序列和第二衔接子核酸序列的双链靶核酸分子的衔接子核酸序列。在第三方面中,每一衔接子核酸序列包括引物结合结构域和单个分子标识符(SMI)结构域。

[0015] 在第三方面的实施例中,第一衔接子核酸序列或第二衔接子核酸序列中的至少一个进一步包括一个包括至少一种经修饰的核苷酸的结构域。第一衔接子核酸序列和第二衔接子核酸序列进一步包含一个包括至少一种经修饰的核苷酸的结构域。在实施例中,第一衔接子核酸序列或第二衔接子核酸序列中的至少一个进一步包括连接结构域。第一衔接子核酸序列和第二衔接子核酸序列可包括连接结构域。至少一种经修饰的核苷酸可为无碱基位点、尿嘧啶、四氢呋喃、8-氧代-7,8-二氢-2'-脱氧腺苷(8-氧代-A)、8-氧代-7,8-二氢-2'-脱氧鸟苷(8-氧代-G)、脱氧肌苷、5'-硝基吡咯、5-羟基甲基-2'-脱氧胞苷、异胞嘧啶、5'-甲基-异胞嘧啶或异鸟苷。两个衔接子序列可包括两个至少部分地一起退火的单独DNA分子。第一衔接子核酸序列和第二衔接子核酸序列可经由连接子结构域连接。连接子结构域可由核苷酸构成。连接子结构域可包括一个或多个经修饰的核苷酸或非核苷酸分子。在实施例中,至少一种经修饰的核苷酸或非核苷酸分子可为无碱基位点、尿嘧啶、四氢呋喃、8-氧代-7,8-二氢-2'-脱氧腺苷(8-氧代-A)、8-氧代-7,8-二氢-2'-脱氧鸟苷(8-氧代-G)、脱氧肌苷、5'-硝基吡咯、5-羟基甲基-2'-脱氧胞苷、异胞嘧啶、5'-甲基-异胞嘧啶或异鸟苷。连接子结构域可形成环。第一衔接子核酸序列的引物结合结构域可与第二衔接子核酸序列的引物结合结构域至少部分地互补。第一衔接子核酸序列的引物结合结构域可与第二

衔接子核酸序列的引物结合结构域互补。第一衔接子核酸序列的引物结合结构域可与第二衔接子核酸序列的引物结合结构域非互补。在实施例中,至少一个SMI结构域为内源性SMI,例如涉及剪切点(例如使用剪切点自身,使用剪切点的实际映射位置(例如染色体3,位置1、234、567),使用紧邻剪切点的DNA中规定数目的核苷酸(例如距离剪切点十个核苷酸,远离剪切点开始七个核苷酸的八个核苷酸,且在“C”首次出现之后,在剪切点之后开始的六个核苷酸))。SMI结构域包括至少一种简并或半简并核酸。SMI结构域可为非简并的。SMI结构域的序列可考虑与对应于连接的DNA的随机或半随机剪切末端的序列结合以获得能够彼此区分单个DNA分子的SMI序列。第一衔接子核酸序列的SMI结构域可与第二衔接子核酸序列的SMI结构域至少部分地互补。第一衔接子核酸序列的SMI结构域可与第二衔接子核酸序列的SMI结构域互补。第一衔接子核酸序列的SMI结构域可与第二衔接子核酸序列的SMI结构域至少部分地非互补。第一衔接子核酸序列的SMI结构域可与第二衔接子核酸序列的SMI结构域非互补。在实施例中,每一SMI结构域包括约1个至约30个之间的简并或半简并核酸。第一衔接子核酸序列的连接结构域可与第二衔接子核酸序列的连接结构域至少部分地互补。在实施例中,每一连接结构域可能够连接到双链靶核酸序列的一个链。在实施例中,连接结构域中的一个包括T-突出端、A-突出端、CG-突出端、平末端或另一可连接的核酸序列。在实施例中,两个连接结构域包含平末端。在实施例中,每一SMI结构域包括引物结合位点。在实施例中,至少第一衔接子核酸序列进一步包括SDE。SDE可位于第一衔接子核酸序列末端处。第二衔接子核酸序列进一步包括SDE。SDE可位于第二衔接子核酸序列末端处。第一衔接子核酸序列的SDE可与第二衔接子核酸序列的SDE至少部分地非互补。第一衔接子核酸序列的SDE可与第二衔接子核酸序列的SDE非互补。第一衔接子核酸序列的SDE可直接连接到第二衔接子核酸序列的SDE。第一衔接子核酸序列的SDE与来自第二衔接子核酸序列的SDE的至少一个核苷酸不同和/或可与其非互补。至少一种核苷酸可通过酶反应从第一衔接子核酸序列的SDE或第二衔接子核酸的SDE省略。酶反应可包括聚合酶或核酸内切酶。至少一种核苷酸可为经修饰的核苷酸或包括标记的核苷酸。经修饰的核苷酸或包括标记的核苷酸可为无碱基位点、尿嘧啶、四氢呋喃、8-氧代-7,8-二氢-2'-脱氧腺苷(8-氧代-A)、8-氧代-7,8-二氢-2'-脱氧鸟苷(8-氧代-G)、脱氧肌苷、5'-硝基吡啶、5-羟基甲基-2'-脱氧胞苷、异胞嘧啶、5'-甲基-异胞嘧啶或异鸟苷。第一衔接子核酸序列的SDE可包含能够形成发夹环的自互补结构域。第一衔接子核酸序列远离其连接结构域的末端可连接到第二衔接子核酸序列远离其连接结构域的末端,由此形成环。环可包括限制酶识别位点。第一衔接子核酸序列的引物结合结构域可位于SMI结构域的5'处。包括第一衔接子核酸序列的至少一种经修饰的核苷酸的结构域可位于SMI结构域的5'处。包括第一衔接子核酸序列的至少一种经修饰的核苷酸的结构域可位于SMI结构域的3'处。包括第一衔接子核酸序列的至少一种经修饰的核苷酸的结构域可位于SMI结构域的5'处且可位于引物结合结构域的3'处。包括第一衔接子核酸序列的至少一种经修饰的核苷酸的结构域可位于SMI结构域的3'处(其可位于引物结合结构域的3'处)。第一衔接子核酸序列的SMI结构域可位于连接结构域的5'处。第一衔接子核酸序列的3'末端可包括连接结构域。在实施例中,第一衔接子核酸序列从5'到3'包括引物结合结构域、包括至少一种经修饰的核苷酸的结构域、SMI结构域和连接结构域。在实施例中,第一衔接子核酸序列从5'到3'包括引物结合结构域、SMI结构域、包括至少一种经修饰的核苷酸的结构域和连接结构域。在实施例中,第一衔接子核酸序列或第二衔接子核

酸序列包括经修饰的核苷酸或非核苷酸分子。经修饰的核苷酸或非核苷酸分子可为大肠杆菌素E2、Im2、谷胱甘肽、谷胱甘肽-s-转移酶 (GST)、镍、聚组氨酸、FLAG-标签、myc-标签或生物素。生物素可为生物素-16-氨基烯丙基-2'-脱氧尿苷-5'-三磷酸、生物素-16-氨基烯丙基-2'-脱氧胞苷-5'-三磷酸、生物素-16-氨基烯丙基胞苷-5'-三磷酸、N4-生物素-OBEA-2'-脱氧胞苷-5'-三磷酸、生物素-16-氨基烯丙基尿苷-5'-三磷酸、生物素-16-7-脱氮-7-氨基烯丙基-2'-脱氧鸟苷-5'-三磷酸、脱硫生物素-6-氨基烯丙基-2'-脱氧胞苷-5'-三磷酸、5'-生物素-G-单磷酸、5'-生物素-A-单磷酸、5'-生物素-dG-单磷酸或5'-生物素-dA-单磷酸。生物素可能能够与附接到基质上的抗生蛋白链菌素结合。在实施例1中，当生物素与附接到基质上的抗生蛋白链菌素结合时，第一衔接子核酸序列能够与第二衔接子核酸序列分离。第二衔接子核酸序列可包括至少一个硫代磷酸酯键。双链靶核酸序列可为DNA或RNA。在实施例1中，第一衔接子核酸序列或第二衔接子核酸序列包括选自小分子、核酸、肽和独特地可结合的部分（其能够与亲和配偶体结合）的亲和标记。在实施例1中，当亲和配偶体附接到固体基质上且与亲和标记结合时，包括亲和标记的衔接子核酸序列能够与不包括亲和标记的衔接子核酸序列分离。固体基质可为固体表面、珠粒或另一固定式结构。核酸可为DNA、RNA或其组合，且任选地包括肽核酸或锁核酸。亲和标记可位于衔接子的末端处或可能与第二衔接子核酸序列中的相对结构域不完全互补的第一衔接子核酸序列中的结构域内。在实施例1中，第一衔接子核酸序列或第二衔接子核酸序列包括具有磁性特性、电荷特性或不可溶特性的物理基团。在实施例1中，当物理基团具有磁性特性且施加磁场时，包括物理基团的衔接子核酸序列与不包括物理基团的衔接子核酸序列分离。在实施例1中，当物理基团具有电荷特性且施加电场时，包括物理基团的衔接子核酸序列与不包括物理基团的衔接子核酸序列分离。在实施例1中，当物理基团具有不可溶特性且衔接子核酸序列对含于对于物理基团来说不可溶的溶液中时，包括物理基团的衔接子核酸序列从不包括物理基团的衔接子核酸序列中沉淀出来，残留于溶液中。物理基团可位于衔接子的末端处或可能与第二衔接子核酸序列中的相对结构域不完全互补的第一衔接子核酸序列中的结构域内。第一衔接子核酸序列或第二衔接子核酸序列可为至少部分单链的。第一衔接子核酸序列或第二衔接子核酸序列可为单链的。第一衔接子核酸序列和第二衔接子核酸序列可为单链的。在实施例1中，连接结构域中的至少一个包括去羟基化碱基。在实施例1中，连接结构域中的至少一个已经化学修饰以便使其不可连接。

[0016] 在第四方面中，本发明涉及一种包括第三方面的至少两对衔接子核酸分子的组合物，其中来自第一对衔接子核酸分子的第一衔接子核酸分子的SMI结构域与来自至少第二对衔接子核酸分子的第一衔接子核酸分子的SMI结构域不同。

[0017] 在第四方面的实施例中，来自第一对单链衔接子核酸分子的第一衔接子核酸分子的SMI结构域可与来自至少第二对单链衔接子核酸分子的第一单链衔接子核酸分子的SMI结构域长度相同。来自第一对单链衔接子核酸分子的第一衔接子核酸分子的SMI结构域可具有与来自至少第二对单链衔接子核酸分子的第一单链衔接子核酸分子的SMI结构域不同的长度。在实施例中，每一SMI结构域在SMI内或侧接SMI的位点处包括一个或多个固定碱基。

[0018] 在第五方面中，本发明涉及一种包括至少一种第一双链复合核酸的组合物，其包括连接到双链靶核酸分子的第一末端的第三方面的第一对衔接子核酸分子以及连接到双

链靶核酸分子的第二末端的第三方面的第二对衔接子核酸分子。

[0019] 在第五方面的实施例中,第一对衔接子核酸分子可与第二对衔接子核酸分子不同。第一对衔接子核酸分子的第一链衔接子靶核酸分子可包括第一SMI结构域且第二对衔接子核酸分子的第一链衔接子靶核酸分子可包括第二SMI结构域。第一对衔接子核酸分子的第一链衔接子靶核酸分子可包括第一SMI结构域且第二对衔接子核酸分子的第一链衔接子靶核酸分子包括第二SMI结构域。在实施例中,所述组合物包括至少第二双链复合核酸。

[0020] 在第六方面中,本发明涉及一种包括第一方面的至少一对衔接子核酸分子和第三方面的至少一对衔接子核酸分子的组合物。

[0021] 在第七方面中,本发明涉及一种包括至少一种第一双链复合核酸的组合物,其包括连接到双链靶核酸分子的第一末端的第一方面的第一对衔接子核酸分子以及连接到双链靶核酸分子的第二末端的第三方面的第二对衔接子核酸分子。

[0022] 在第八方面中,本发明涉及一种测序双链靶核酸的方法,其包括以下步骤:(1)使第一方面的一对衔接子核酸序列连接到双链靶核酸分子的至少一个末端,由此形成包括第一链衔接子靶核酸序列和第二链衔接子靶核酸序列的双链核酸分子,(2)使第一链衔接子靶核酸序列扩增,由此产生包括多种第一链衔接子靶核酸序列和其多种互补分子的第一组扩增产物,(3)使第二链衔接子靶核酸序列扩增,由此产生包括多种第二链衔接子靶核酸序列和其多种互补分子的第二组扩增产物,其中第二组扩增产物可与第一组扩增产物区分开来,(4)测序第一组扩增产物,以及(5)测序第二组扩增产物。

[0023] 在第八方面的实施例中,至少一个末端可为两个末端。扩增可通过PCR,通过多重置换扩增或通过等温扩增进行。连接到双链靶核酸序列的第一末端的衔接子核酸序列对具有与连接到双链靶核酸序列的第二末端的衔接子核酸序列对相同的结构。在第八方面的实施例中,第一链衔接子靶核酸序列以5'到3'排序包括:(a)第一衔接子核酸序列、(b)双链靶核酸的第一链和(c)第二衔接子核酸序列。在第八方面的实施例中,第二链衔接子靶核酸序列可以3'到5'排序包括:(a)第一衔接子核酸序列、(b)双链靶核酸的第二链和(c)第二衔接子核酸序列。连接到双链靶核酸序列的第一末端的衔接子核酸序列对可与连接到双链靶核酸序列的第二末端的衔接子核酸序列对不同。连接到双链靶核酸序列的第一末端的衔接子核酸序列对具有第一SMI结构域,且连接到双链靶核酸序列的第二末端的衔接子核酸序列对具有第二SMI结构域,其中第一SMI结构域可与第二SMI结构域不同。在第八方面的实施例中,第一链衔接子靶核酸序列可以5'到3'排序包括:(a)包括第一SDE的第一衔接子核酸序列、(b)第一SMI结构域、(c)双链靶核酸的第一链和(d)第二衔接子核酸序列。在第八方面的实施例中,第二链衔接子靶核酸序列可以5'到3'排序包括:(a)包括第一SDE的第一衔接子核酸序列、(b)第二SMI结构域、(c)双链靶核酸的第二链和(d)第二衔接子核酸序列。在实施例中,用于第一组扩增产物的共有序列可与用于第二组扩增产物的共有序列进行比较且两个共有序列之间的差异可视为伪差。

[0024] 在第九方面中,本发明涉及一种测序双链靶核酸的方法,其包括以下步骤:(1)使第三方面的一对衔接子核酸序列连接到双链靶核酸分子的至少一个末端,由此形成包括第一链衔接子靶核酸序列和第二链衔接子靶核酸序列的双链核酸分子,(2)使第一链衔接子靶核酸分子扩增,由此产生包括多种第一链衔接子靶核酸分子和其多种互补分子的第一组扩增产物,(3)使第二链衔接子靶核酸分子扩增,由此产生包括多种第二链衔接子靶

靶核酸分子和其多种互补分子的第二组扩增产物, (4) 测序第一组扩增产物, 由此获得用于第一组扩增产物的共有序列, 以及 (5) 测序第二组扩增产物, 由此获得用于第二组扩增产物的共有序列。

[0025] 在第九方面的实施例中, 第二组扩增产物可与第一组扩增产物区分开来。扩增可通过PCR, 通过多重置换扩增或通过等温扩增进行。在第九方面的实施例中, 所述方法在步骤(1)之后进一步包括使双链核酸分子与至少一种将至少一种经修饰的核苷酸改变为另一化学结构的酶(例如糖基化酶)接触的步骤。连接到双链靶核酸分子的第一末端的衔接子核酸序列对可与连接到双链靶核酸分子的第二末端的衔接子核酸序列对一致。连接到双链靶核酸分子的第一末端的衔接子核酸序列对可与连接到双链靶核酸分子的第二末端的衔接子核酸序列对不同。在实施例中, 一对衔接子核酸序列可连接到双链靶核酸分子的第一末端且可利用对应于靶DNA分子的DNA序列的一部分的引物以扩增DNA分子。在第九方面的实施例中, 第一链衔接子靶核酸序列以5'到3'排序包括: (a) 包括至少一种经修饰的核苷酸或至少一个无碱基位点的第一衔接子核酸序列, (b) 双链靶核酸的第一链, 以及(c) 第二衔接子核酸序列。在第九方面的实施例中, 第二链衔接子靶核酸序列以3'到5'排序包括: (a) 第一衔接子核酸序列、(b) 双链靶核酸的第二链和(c) 第二衔接子核酸序列。连接到双链靶核酸分子的第一末端的衔接子核酸序列对可与连接到双链靶核酸分子的第二末端的衔接子核酸序列对不同。连接到双链靶核酸分子的第一末端的衔接子核酸序列对具有第一SMI结构域, 且连接到双链靶核酸序列的第二末端的衔接子核酸序列对具有第二SMI结构域, 其中第一SMI结构域可与第二SMI结构域不同。在第九方面的实施例中, 第一链衔接子靶核酸序列以5'到3'排序包括: (a) 包括至少一种经修饰的核苷酸或至少一个无碱基位点和第一SMI结构域的第一衔接子核酸序列、(b) 双链靶核酸的第一链和(c) 包括第二SMI结构域的第二衔接子核酸序列。在实施例中, 当至少一种经修饰的核苷酸可为8-氧代-G时, 且第二衔接子核酸序列在对应于8-氧代-G的位置处包括胞嘧啶。在第九方面的实施例中, 第二链衔接子靶核酸序列以3'到5'排序包括: (a) 包括第一SMI结构域的第一衔接子核酸序列、(b) 双链靶核酸的第二链和(c) 包括第二SMI结构域的第二衔接子核酸序列。在实施例中, 至少一种经修饰的核苷酸可为8-氧代-G, 第二衔接子核酸序列在对应于8-氧代-G的位置处包括胞苷。在实施例中, 在步骤(2)或步骤(3)的扩增期间, 至少一个无碱基位点可在扩增后转化成相应的扩增产物中的胸苷, 从而引入SDE。在第九方面的实施例中, 在步骤(2)或步骤(3)的扩增期间, 至少一种经修饰的核苷酸位点编码相应的扩增产物中的腺苷。

[0026] 在第十方面中, 本发明涉及一种方法, 其中可区分的扩增产物可获自个别DNA分子的两个链中的每一个, 且用于第一组扩增产物的共有序列可与用于第二组扩增产物的共有序列进行比较, 其中两个共有序列之间的差异可视为伪差。

[0027] 在第十方面的实施例中, 借助于共用相同SMI序列, 可测定出扩增产物从相同初始DNA分子处产生。在实施例中, 借助于携带可能已知彼此对应的相异SMI序列, 基于在SMI衔接子库合成时产生且与其结合的数据库, 可测定出扩增产物从相同初始DNA分子处产生。在实施例中, 经由通过SDE引入的至少一个核苷酸序列差异, 可测定出扩增产物从相同初始双链DNA序列的相异链处产生。

[0028] 在第十一方面中, 本发明涉及一种方法, 其中可区分的扩增产物可获自个别DNA分子的两个链中的每一个, 且对应于单个DNA分子的两个初始DNA链中的一个的获自扩增产物

的序列与对应于两个初始DNA链中的第二个的扩增产物进行比较,且两个序列之间的差异可视为伪差。

[0029] 在第十二方面中,本发明涉及一种方法,其中当将对应于单个DNA分子的两个初始DNA链中的一个的获自扩增产物的序列与对应于两个初始DNA链中的第二个的扩增产物进行比较时,不可区分的扩增产物可获自个别DNA分子的两个链且两个序列之间未鉴别出差异。

[0030] 在第十二方面的实施例中,借助于共用相同SMI序列,基于在SMI衔接子库合成时产生且与其结合的数据库,可测定出扩增产物从相同初始双链DNA分子处产生。在实施例中,经由通过SDE引入的至少一个核苷酸序列差异,可测定出扩增产物从相同初始双链DNA序列的相异链处产生。在实施例中,所述方法进一步包括在DNA双螺旋体热或化学熔融成其组分单链之后单分子稀释的步骤。单链可稀释到多个以物理方式隔开的反应腔室中以使得两个共用相同容器的最初配对链的机率可较小。以物理方式隔开的反应腔室可选自容器、套管、孔和至少一对非连通液滴。在实施例中,对于每一以物理方式隔开的反应腔室,可进行PCR扩增,优选针对携带不同标签序列的每一腔室使用引物。在实施例中,每一标签序列充当SDE。在实施例中,对应于相同初始DNA的两个链的一系列配对序列可彼此进行比较,且可选择一系列产物的至少一个序列作为最可能代表初始DNA分子的正确序列。可至少部分地归因于具有获自两个DNA链的产物之间的最小错配数,选择最可能代表初始DNA分子的正确序列的所选产物。可至少部分地归因于相对于参考序列具有最小错配数,选择最可能代表初始DNA分子的正确序列的所选产物。

[0031] 在第十三方面中,本发明涉及一种包括至少两对衔接子核酸序列的组合物,其中第一对衔接子核酸序列包括:引物结合结构域、链限定元件(SDE)和连接结构域,其中第二对衔接子核酸序列包括:引物结合结构域、单个分子标识符(SMI)结构域和连接结构域。

[0032] 在第十四方面中,本发明涉及一种双链复合核酸,其包括:(1)第一对衔接子核酸序列,其包括:引物结合结构域和SDE;以及(2)双链靶核酸;以及(3)第二对衔接子核酸序列,其包括:引物结合结构域和单个分子标识符(SMI)结构域,其中第一对衔接子核酸分子可连接到双链靶核酸分子的第一末端,且第二对衔接子核酸分子可连接到双链靶核酸分子的第二末端。在第十四方面的实施例中,第一对衔接子核酸序列和/或第二对衔接子核酸序列可进一步包括连接结构域。

[0033] 在第十五方面中,本发明涉及用于测序包括第一衔接子核酸序列和第二衔接子核酸序列的双链靶核酸分子的衔接子核酸序列对,其中每一衔接子核酸序列包括:引物结合结构域、SDE、连接结构域,其中第一衔接子核酸序列的SDE可与第二衔接子核酸序列的SDE至少部分地非互补。

[0034] 在第十六方面中,本发明涉及一种双链环形核酸,其包括连接到双链靶核酸分子的第一末端且连接到双链靶核酸分子的第二第二末端的第一方面的一对衔接子核酸分子。

[0035] 在第十七方面中,本发明涉及一种双链环形核酸,其包括连接到双链靶核酸分子的第一末端且连接到双链靶核酸分子的第二第二末端的第三方面的一对衔接子核酸分子。

[0036] 在第十八方面中,本发明涉及一种双链环形核酸,其包括连接到双链靶核酸分子的第一末端的第一方面的一对衔接子核酸分子以及连接到双链靶核酸分子的第二末端的退火的引物结合结构域对,其中退火的引物结合结构域对可连接到衔接子核酸分子对。

[0037] 在第十九方面中,本发明涉及一种双链环形核酸,其包括连接到双链靶核酸分子的第一末端的第三方面的一对衔接子核酸分子以及连接到双链靶核酸分子的第二末端的退火的引物结合结构域对,其中退火的引物结合结构域对可连接到衔接子核酸分子对。

[0038] 在第二十方面中,本发明涉及一种双链复合核酸,其包括:(1)一对衔接子核酸序列,其包括:引物结合结构域、链限定元件(SDE)和单个分子标识符(SMI)结构域;(2)双链靶核酸;以及(3)退火的引物结合结构域对,其中衔接子核酸分子对可连接到双链靶核酸分子的第一末端且退火的引物结合结构域对可连接到双链靶核酸分子的第二末端。在第二十方面的实施例中,衔接子核酸序列对和/或退火的引物结合结构域对进一步包括连接结构域。

[0039] 双重测序另外描述于W02013142389A1和Schmitt等人,PNAS 2012中,其中的每一个以全文引用的方式并入本文中。

[0040] 上述方面和实施例中的任一个都可以与如本文发明内容、图式和/或实施方式中所公开的任何其它方面或实施例(包括本发明的以下特定非限制性实例/实施例)组合。

[0041] 本发明的其它特征、优点和修改将从图式、实施方式和权利要求书显而易见。前述描述意图说明且不限制本发明的范围。

附图说明

[0042] 从结合附图进行的以下详细描述中将更清楚地理解以上以及其它特征。

[0043] 图1A到图1I说明使用Y形衔接子的最初描述的双重测序。显示示例性Y形衔接子(图1A)、连接到此类衔接子的双链DNA分子(图1B)、由其衍生的PCR产物(图1C和图1D)以及由此产生的测序读段(图1E到图1I)。

[0044] 图2A到图2K说明使用非互补“气泡”衔接子的本发明的双重测序。显示示例性“气泡”衔接子(图2A和图2H到图2K)、连接到图2A的衔接子的双链DNA分子(图2B)、由其衍生的PCR产物(图2C和图2D)以及由此产生的测序读段(图2E到图2G)。

[0045] 图3A到图3G说明使用具有非互补“气泡”形单分子标识符(SMI)(其共同用作分子标识符)以及不对称性引入链限定元件(SDE)的衔接子的本发明的双重测序。显示示例性“气泡”衔接子(图3A)、连接到图3A的衔接子的双链DNA分子(图3B)、由其衍生的PCR产物(图3C和图3D)以及由此产生的测序读段(图3E和图3F)。图3G显示通过特定SMI序列和其相应的非互补配偶体分组的图3E和图3F的测序读段。

[0046] 图4A到图4H说明使用具有核苷酸或核苷酸类似物(其首先形成配对链DNA,但随后在后续生物化学反应之后赋予DNA错配)的衔接子的本发明的双重测序。显示包含8-氧代-7,8-二氢-2'-脱氧鸟苷(8-氧代-G)的示例性衔接子(图4A)、连接到图4A的衔接子的双链DNA分子(图4B),图4C显示在用产生置换8-氧代-G碱基的无碱基位点以及由此衔接子中的错配的糖基化酶处理之后图4B的双链DNA分子;由其衍生的PCR产物(图4D和图4E);以及由此产生的测序读段(图4F到图4H)。

[0047] 图5A到图5H说明使用双重测序衔接子设计的组合以在DNA分子的相对端引入不同引物位点的本发明的双重测序。显示示例性双重测序衔接子(图5A)和“标准”衔接子(图5B);当图5A和图5B的衔接子连接到DNA分子时产生的三种类型的双链DNA分子(图5C到图5E);由其衍生的PCR产物(图5F和图5G);以及由此产生的测序读段(图5H)。

[0048] 图6A到图6I说明使用允许在非双端平台上的两个读段的双重测序衔接子设计的

组合的本发明的双重测序。显示“标准”衔接子(图6A)和示例性双重测序衔接子(图6B);当图6A和图6B的衔接子连接到DNA分子时产生的优选的双链DNA分子(图6C);由其衍生的PCR产物(图6D和图6E);用于测序衍生自“顶部”链(图6F)和“底部”链(图6G)的模板链的布置;以及由此产生的测序读段(图6H和图6I)。

[0049] 图7A到图7I说明使用允许在非双端平台上的两个读段的双重测序衔接子设计的组合的本发明的双重测序。显示另外包括简并或半简并SMI序列的衔接子(图7A)和示例性双重测序衔接子(图7B);当图7A和图7B的衔接子连接到DNA分子时产生的优选的双链DNA分子(图7C);由其衍生的PCR产物(图7D和图7E);用于测序衍生自“顶部”链(图7F)和“底部”链(图7G)的模板链的布置;以及由此产生的测序读段(图7H和图7I)。

[0050] 图8A到图8J说明使用具有不对称SMI的Y形双重测序衔接子的本发明的双重测序。显示示例性双重测序衔接子(图8A)、当图8A的衔接子连接到DNA分子时产生的双链DNA分子(图8B)、由其衍生的PCR产物(图8C和图8D)以及由此产生的测序读段(图8E和图8F)。图8G显示通过特定SMI序列和其相应的非互补配偶体分组的图8E和图8F的测序读段。图8H到图8J显示适用于此实施例中的替代衔接子设计。

[0051] 图9A到图9G说明使用具有位于无单链尾部区域中的不对称SMI的Y形或环形双重测序衔接子的本发明的双重测序。显示示例性双重测序衔接子(图9A)、当图9A的衔接子连接到DNA分子时产生的优选的双链DNA分子(图9B)、由其衍生的PCR产物(图9C和图9D),测序引物位点和索引引物位点的定向显示于图9E和图9F中。图9G显示图9E和图9F中所示方法中获得的分组测序读段。

[0052] 图10A到图10E说明本发明的双重测序,其中双重测序必需的所有元件包括于单个分子而非两个配对衔接子中。图10A显示在双链DNA分子连接之前的此类配置且图10B显示在双链DNA分子连接之后图10A的配置。图10C到图10E显示此实施例的一些替代物。

[0053] 图11A到图11D说明经由不对称化学标记和链分离的双重测序。显示具有化学标签(此处为生物素)的示例性双重测序衔接子(图11A)和第二衔接子(图11B);当图11A的衔接子和图11B的衔接子连接到DNA分子时产生的优选的双链DNA分子(图11C),以及所述方法中的其它步骤,其中包含化学标签的链与其它链分离且各自独立地经扩增和测序(图11D)。

[0054] 图12A到图12M说明本发明的双重测序,其中SDE通过切口平移引入。图12A到图12D显示衔接子设计,其中SDE在切口平移之后损失。显示适用于此实施例中的Ion Torrent™-可相容衔接子(图12E和图12F);当图12E和图12F的衔接子连接到DNA分子时产生的优选的双链DNA分子(图12G);末端核苷酸的错掺入(图12H);由其衍生的延伸产物且其显示错配(图12I);衍生自图12I的分子的PCR产物(图12J和图12K);以及由此产生的测序读段(图12L和图12M)。

[0055] 图13A到图13G说明本发明的双重测序,其中SDE在切口平移之后引入。显示包含去磷酸化5'末端的双重测序衔接子(图13A);当图13A的衔接子连接到DNA分子时产生的双链DNA分子(图13B);已发生链置换合成之后的结构(图13C);图13C的结构的延伸产物(图13D),其显示无错配;在用尿嘧啶DNA糖基化酶和适当的AP核酸内切酶处理之后包括缺口的结构(图13E);在缺口已用错配核苷酸填充空隙且连接关闭之后图13E的结构(图13F);以及由此产生的测序读段(图13G)。

[0056] 图14A到图14I说明本发明的双重测序,其中通过聚合酶延伸到待测序DNA分子中

来引入错配。显示待测序双链DNA分子(图14A)、已经离开5'突出端的核酸内切酶处理的图14A的双链DNA分子(图14B);图14B的部分双链DNA分子经处理以引入两个错配(图14C),图14C的结构延伸产物(图14D),其现在在每一错配处包括“气泡”;显示于图14E中的一对衔接子;当图14E的衔接子连接到图14D的DNA分子时产生的图14F的结构;衍生自图14F的分子的PCR产物(图14G和图14H);以及由此产生的测序读段(图14I)。

具体实施方式

[0057] 首先通过使用用于两个DNA链的单独扩增的不对称引物结合位点描述双重测序。本文中描述并不需要使用不对称引物结合位点的双重测序的替代和优良途径。实际上,可通过在待测序DNA分子中的衔接子内或其它地方产生两个链之间的DNA序列中的至少一种核苷酸的不同(例如错配、额外核苷酸和省略核苷酸),用经修饰的核苷酸(例如不具有碱基或具有非典型性碱基的核苷酸)置换至少一种核苷酸,和/或包括可以物理方式分离两个链的至少一种经标记的核苷酸(例如生物素化核苷酸)来引入两个链之间的不对称性。表1说明如本发明中所公开的用于组装双重测序的衔接子的示例性选项。

[0058] 表1:

[0059]

链限定元件 (SDE)	单分子标识符 (SMI)
至少一种核苷酸的错配存在于衔接子 (即, “气泡衔接子”) 内部	邻近气泡的 SMI 气泡自身内的 SMI SMI 处于第二衔接子中 “内源性 SMI” (剪切点)
“匹配” 序列通过后续步骤转化成错配 --通过酶处理产生错配 (实例: 8-氧代-G) --通过聚合酶引入错配 (实例: 切口平移)	邻近气泡的 SMI 气泡自身内的 SMI SMI 处于第二衔接子中 “内源性 SMI” (剪切点)
不同序列存在于衔接子尾部 (即, “Y 衔接子”) 内	邻近衔接子尾部的 SMI 尾部自身内的 SMI SMI 处于第二衔接子中 “内源性 SMI” (剪切点)
两个链以物理方式分离 (例如利用一个链上的生物素, 而非其它)	衔接子自身内的 SMI SMI 处于第二衔接子中 “内源性 SMI” (剪切点)
两个链为不同长度 --一个衔接子链内的内部 “环” --将额外核苷酸添加到一个链而非其它中	衔接子自身内的 SMI SMI 处于第二衔接子中 “内源性 SMI” (剪切点)

[0060] 注释:

[0061] (i) 这些衔接子设计中的所有可具有所添加的额外任选的元件 (例如两个衔接子链连接在一起且利用各种配置中的PCR引物位点)

[0062] (ii) 每当使用SMI时, 其可为随机/简并、半随机/半简并或预定义的。此外, 如果SMI包含两个链, 那么两个链可为互补、非互补或部分互补的。

[0063] (iii) 含有至少一种SDE和至少一种SMI的完全调适的分子复合物可存在于衔接子中和/或在附接之前连接的DNA可在连接之后产生或可为其组合。

[0064] 本文所描述的用于双重测序的衔接子设计和途径不取决于使用具有互补SMI序列的Y衔接子。

[0065] 一些设计直接适用于单端测序。本文所公开的途径共享两个通用特征: (1) 以使得最终衍生自两个链中的每一个的序列可识别为与相同DNA双螺旋体相关的方式标记个别半个双螺旋体DNA分子的每一单链; 以及 (2) 以使得最终衍生自两个链中的每一个的序列可识

别为不同于衍生自相反链的那些的方式标记个别双螺旋体DNA分子的每一单链。提供这些相应功能的分子特征在本文中命名为单分子标识符(SMI)和链限定元件(SDE)。

[0066] 这是首次公开经由不同形式的内部非互补“气泡”序列引入链限定不对称性。一个此类实施例涉及引入不位于扩增引物位点内的非互补“气泡”序列;与“气泡”的两个链不同的序列将随后产生两个链的单独标记。

[0067] 本文公开经由使用经修饰的DNA碱基作为SDE,可如何类似地将链限定不对称性引入到经调适的DNA分子中。在实例中,不对称性通过包括一种或多种核苷酸类似物引入,所述核苷酸类似物首先产生互补序列,但其可随后转化成非互补序列。

[0068] 还公开了非Y形不对称衔接子设计可应用于测序平台的方法,所述平台需要每一DNA分子的相对端上的不同引物序列。

[0069] 本文中公开了替代方法,其中不同类型的SMI标签和SDE可分布于含有衔接子的两种不同引物位点中以利于阅读长度和SMI标记多样性最大化。

[0070] 本文还公开用于双重测序衔接子的额外设计,所述衔接子包含Y或环形尾部容易适于成对端测序,但其中SMI标签为非互补序列,且因此允许明显的灵活设计。

[0071] 本文证明了此类不对称性的此类引入如何能够出于错误校正的目的,通过双重测序区分产物与两个DNA链。此外,本文证明了一些实施例如何促进在单端读段平台上进行双重测序的描述。

[0072] 进一步公开引入用于单个衔接子双重测序的引物位点和SMI位点和SDE位点以形成环形衔接子-DNA分子复合物的方法。

[0073] 另外,公开一种引入SDE的完全不同途径,其依赖于不对称化学标记,允许配对链物理/机械分离到不同反应隔室中以便非依赖性分析,而非两个链的基于差异序列的分子标记。

[0074] 本文公开了尤其用于Ion Torrent™(Life Technologies®)测序平台的衔接子设计的实例。

[0075] 本文公开了可连接到双螺旋体分子每一端处的两个单链的衔接子变异体以及允许单链连接继而“切口平移”的设计,其在最终制得的分子中保留了必需SMI和SDE元件。

[0076] 本文公开了SDE可如何以独立于衔接子连接的方式掺入到DNA分子自身中。

[0077] 最终,本文公开了用于双重测序的流线型替代算法途径,其可与消除前述单链共有序列(SSCS)产生的需要的任何双螺旋体衔接子设计一起使用。

[0078] 在一些实施例中,核苷酸序列的一部分可为“简并的”。在简并序列中,每一位置可为任何核苷酸,即,由“X”、“N”或“M”表示的每一位置可为腺嘌呤(A)、胞嘧啶(C)、鸟嘌呤(G)、胸腺嘧啶(T)或尿嘧啶(U)或具有碱基配对特性的任何其它天然或非天然DNA或RNA核苷酸或核苷酸类物质或类似物(例如黄苷、肌苷、次黄嘌呤、黄嘌呤、7-甲基鸟嘌呤、7-甲基鸟苷、5,6-二氢尿嘧啶、5-甲基胞嘧啶、二氢尿苷、异胞嘧啶、异鸟嘌呤、脱氧核苷、核苷、肽核酸、锁核酸、二醇核酸和苏糖核酸)。或者,核苷酸序列的一部分可为不完全简并的以使得序列包括至少一种预定义的核苷酸或至少一种预定义的多核苷酸和位置,其可为仅包括可能的核苷酸的子集组合的任何核苷酸或一个或多个位置。可能的核苷酸的子集组合可包括:以下各者中的任何三个:A、C、G和T;以下各者中的任何两个:A、C、G、T和U;或U加以下各者中的任何三个:A、C、G和T。此类子集组合可另外包括具有碱基配对特性的任何其它天然

或非天然DNA或RNA核苷酸或核苷酸类物质或类似物或经其取代。分子群体中的这些核苷酸中的任一种之间的化学计量比可为约1:1或任何其它比率;本文中此类序列称为“半简并的”。在某些实施例中,“半简并”序列是指一组两种或更多种序列,其中两个或更多个序列在至少一个核苷酸位置处存在不同。在实施例中,半简并序列为并非每个核苷酸相对于其相邻核苷酸(紧邻或在两个或更多个核苷酸内)都是随机的序列。在实施例中,如本文所使用,术语简并和半简并可具有与本申请所属领域的普通技术人员通常所理解和本申请所属领域通常所使用相同的意义;此类领域以全文引用的方式并入本文中。

[0079] 在实施例中,序列无需在每一位置处含有所有可能的碱基。简并或半简并n-mer序列可通过聚合酶介导的方法产生,或可通过制备和退火已知序列的个别寡核苷酸库产生。替代地,任何简并或半简并n-mer序列可为来自与靶DNA来源不同的任何替代来源的随机地或非随机地碎片化双链DNA分子。在一些实施例中,替代来源为衍生自细菌、除靶DNA外生物体或此类替代生物体或来源的组合的基因组或质粒。随机或非随机碎片化DNA可引入到SMI衔接子中以充当可变标签。此可经由酶连接或所属领域中已知的任何其它方法实现。

[0080] 除非上下文另外明确规定,否则如本说明书和所附权利要求书中所使用,单数形式“一个(种)(a/an)”和“所述”包括多个指示物。

[0081] 除非尤其陈述或上下文明显可见,如本文所使用,术语“或”理解为包括性的且涵盖“或”和“和”。

[0082] 术语“一个或更多个”、“至少一个”、“多于一个”和类似表述理解为包括但不限于至少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、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、114、115、116、117、118、119、120、121、122、123、124、125、126、127、128、129、130、131、132、133、134、135、136、137、138、139、140、141、142、143、144、145、146、147、148、149或150、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000或更多以及中间的任何数值。

[0083] 相反地,术语“不超过”包括小于所陈述的值的每一值。举例来说,“不超过100个核苷酸”包括100、99、98、97、96、95、94、93、92、91、90、89、88、87、86、85、84、83、82、81、80、79、78、77、76、75、74、73、72、71、70、69、68、67、66、65、64、63、62、61、60、59、58、57、56、55、54、53、52、51、50、49、48、47、46、45、44、43、42、41、40、39、38、37、36、35、34、33、32、31、30、29、28、27、26、25、24、23、22、21、20、19、18、17、16、15、14、13、12、11、10、9、8、7、6、5、4、3、2、1和0个核苷酸。

[0084] 术语“多个”、“至少两个”、“两个或更多个”、“至少第二”和类似表述理解为包括但不限于至少2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、114、115、116、117、118、119、120、121、122、123、124、125、126、127、128、129、130、131、132、133、134、135、136、137、138、139、

140、141、142、143、144、145、146、147、148、149或150、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000或更多以及中间的任何数值。

[0085] 在整个本说明书中,词语“包含 (comprising)”或变化形式,如“包含 (comprises或 comprising)”将理解为暗示包括所陈述的元件、整数或步骤,或元件、整数或步骤的群组,但不排除任何其它元件、整数或步骤,或元件、整数或步骤的群组。

[0086] 除非专门陈述或从上下文显而易见,否则如本文中所使用的术语“约”应理解为在所属领域中的一般公差范围内,例如在平均值的2倍标准差内。约可理解为在陈述值的10%、9%、8%、7%、6%、5%、4%、3%、2%、1%、0.5%、0.1%、0.05%、0.01%或0.001%内。除非另外从上下文清楚可见,否则本文所提供的所有数值都通过术语“约”来修饰。

[0087] 尽管在本发明的实践或测试中可使用与本文所描述的方法和材料类似或等效的方法和材料,但下文描述了适合的方法和材料。本文中提及的所有公开、专利申请、专利和其它参考文献都以全文引用的方式并入。不承认在此所引用的参考文献是所要求的发明的现有技术。倘若有冲突,那么将以本说明书(包括定义)为主。另外,材料、方法以及实例仅仅是说明性的且并不打算是限制性的。

[0088] 除非另外定义,本文所使用的所有技术和科学术语具有与本申请所属领域的普通技术人员通常所理解和本申请所属领域通常所使用相同的意义;此类领域以全文引用的方式并入本文中。

[0089] 上述方面和实施例中的任一个可与如发明内容、图式和/或实施方式章节中所公开的任何其它方面或实施例组合,包括以下实例/实施例。

[0090] 本发明的特定非限制性实例/实施例

[0091] 使用用于双重测序的Y形衔接子的缺点

[0092] 具有Y形衔接子的双重测序最容易地用成对端测序读段进行,如原先所描述(WO2013142389A1和Schmitt等人,PNAS 2012,其中的每一个以全文引用的方式并入本文中)。然而,并非所有测序平台都与成对端测序读段相容。当使用先前所描述的Y或环形衔接子时,其中不对称引物位点位于与衔接子的可连接的末端相反的单链区域中,具有单端测序读取的双重测序需要测序读取以完全延伸穿过DNA分子。这需要捕集分子两端处的SMI标签序列,其需要能够区分测序读取与两个衍生物链。此需求说明如下。

[0093] 先前所描述的Y形双重测序衔接子显示于图1A中。在图1A中,特征A和B代表不同引物结合位点; α 和 α' 代表简并或半简并序列和其反向互补序列; β 表示不同简并或半简并序列;以及 α 和 β 简并或半简并序列池中的两个任意序列。共同地,这些充当单分子标识符(SMI)。

[0094] 如原先所描述(例如WO2013142389A1),SMI用于区分较大池内的个别分子。有必要衔接子库中经编码的足够的足够较大群体以使得在统计学上将用相同SMI序列标记的任何两个DNA分子不大可能。此外,如先前所描述,在库产生期间引入的断裂位点可用作内源性SMI,在某些情形下,独立地或与衔接子序列中经编码的外源性SMI组合。在本发明中,仅外源性SMI结构域显示于不同衔接子设计的实例中;然而,应理解(且包括于本发明中),外源性SMI结构域可经充当内源性SMI的DNA剪切点取代或通过其扩增。

[0095] 在衔接子连接到来自库的双链DNA片段的每一端之后,结构将呈现显示于图1B中。为了追踪后续图式中的衍生物清楚起见,指出特定DNA插入序列的“左侧”和“右侧”末端以

及“顶部”和“底部”链。

[0096] 在PCR之后,衍生自“顶部”链的双链产物显示于图1C中。(L)和(R)指示起始DNA分子的相应“左侧”和“右侧”末端:

[0097] 衍生自“底部”链的双链PCR产物显示于图1D中。

[0098] 应指出 α 和 β 相对于“顶部”链和“底部”链产物中的A和B的不同配置。在成对端测序读段(即,来自每一PCR产物的两个引物位点A和B的读段)的情况下,有可能区分衍生自每一链的产物,因为 α 标签呈现在一个链的A读段中且 β 在B读段中且在其它链中发生互反情况。参见图1E。

[0099] 使用如上文所描述的成对端读段使双螺旋体序列校正为可能的。然而,如果测序读段足够长以捕集两端处的SMI序列,那么在仅使用单端测序读段(即,仅来自引物位点A或引物位点B但不来自特定分子的两个的读段)的情况下,仅有可能获得双螺旋体序列。如果使用测序引物A,那么衍生自不同链的全长测序读段(即,足够长以包括两个SMI序列)将产生图1F中所示的两个序列。类似地,使用具有全长测序读段的测序引物B将产生图1G中所示的以下两个序列。在上述两种情况中,“顶部”和底部链衍生的产物可借助于在相反定向(一个中的 α - β 和另一个中的 β - α)上具有SMI彼此区分。然而,在不具有足够长以捕集两个SMI序列的测序读段的情况下,不易于用单端测序进行双重测序。这是因为两个测序读段并不各自含有 α 和 β 标签两者。看待此问题的另一方式为对于DNA分子末端部分,互补序列可不进行测序,以使得关于第二链的信息不存在,从而进行比较。

[0100] 为了说明这一点,当使用来自引物A的非全长单端测序读段时产生的序列的两种类型显示于图1H中。类似地,当使用来自引物B的非全长单端测序读段时产生的相应的序列显示于图1I中。应注意,对于图1H和图1I中所示的测序读段的两个,每一DNA片段的“左侧”和“右侧”末端仅在给定引物后测序,因此双重测序无法实现。这是因为不存在与其进行比较的相反链序列。因此,即使用两种不同引物中的每一个测序分子的扩增群体,将无关于第二链的信息,这揭示一组来源于相同衍生物分子的特定读段A和B序列。

[0101] 当使用单端测序时,全部DNA分子的“通读”的需要可产生一些测序平台上的技术挑战,其中读段长度受到限制。

[0102] 对于与子全长测序读段相容的双重测序与单端测序,替代衔接子设计为必需的。在Y形衔接子上的成对端测序读段和不对称引物位点不存在下,不对称性的一些其它形式必须引入到能够区分链的经调适的DNA分子中。下文公开此类设计的实例。

[0103] 用非互补“气泡”引入链限定不对称性

[0104] 图2A中公开了非Y形衔接子(本发明)的示例性设计,其允许利用非成对端测序(即,“气泡衔接子”)的双重测序。不同于先前所描述的具有两个引物位点的Y形衔接子,仅存在具有反向互补序列(P')的单个引物位点(P)。 α 和其互补序列 α' 代表简并或半简并单分子标识符(SMI)序列; X 和 Y 代表链限定元件(SDE)的两个半部,其为形成衔接子内相邻互补序列当中的未配对“气泡”的非互补序列的链段。最终,衔接子具有可连接的序列。通过此衔接子设计中的SDE引入的不对称性区分衍生自如图2B到图2G中所示的每一链的测序读段。

[0105] 在类似于图2A中所示的那些的衔接子连接到DNA片段的每一端之后,产生图2B中所示的结构。第二衔接子显示具有SMI序列 β 和 β' 以说明第二连接的衔接子的SMI序列一般与第一衔接子的SMI序列不同。或者,相同衔接子可连接到DNA分子的两端。

[0106] 在PCR扩增之后,衍生自“顶部”链的双链产物显示于图2C中且衍生自“底部”链的双链产物显示于图2D中。

[0107] 因为引物位点序列在本实例中的分子的两端处相同,取决于测序单链半部,两种不同类型的序列测序读段将获自每一链的PCR产物的单端测序读段。衍生自“顶部”链PCR产物的读段显示于图2E中且衍生自“底部”链PCR产物的读段显示于图2F中。

[0108] 对于分析,如图2G中所示,通过含有特定SMI,在此情况下 α 或 β 的那些分组测序读段。从给定单分子DNA出现的序列可借助于具有相同SMI序列分组在一起。显而易见,在每一SMI群组内可见两种类型的序列:一个通过SDE X标记且一个通过SDE Y标记。这些定义衍生自相反链的测序读段(即,“顶部”和“底部”)。举例来说,当具有SMI标签 α 的序列分组在一起时,所获得的序列为X- α -DNA(图2E)和Y'- α -DNA(图2F)。可通过将X- α -DNA序列分组在一起制得由初始DNA分子的“顶部”链产生的序列组成的共有序列。同样地,“底部”链的共有序列可通过将Y'- α -DNA序列分组在一起来制得。最终,两个链的共有序列可通过一起比较由两个链产生的序列制得(即,那些标记的序列X将与用序列Y'标记的那些进行比较)。共同地,这些允许比较作为双重测序分析部分。

[0109] 类似结果可通过切换SMI和SDE序列的顺序来实现。此类衔接子的一个实例显示于图2H中。

[0110] 如上文和W02013142389A中所提出,在一些实施例中,含于衔接子序列内的SMI可省略代替包含DNA分子自身的剪切点序列的内源性SMI序列。一个此类衔接子设计的结构将需要所示图2A,但排除 α 和 α' 。

[0111] 在一些应用中,图2H中所示的定向为优选的。举例来说,在一些测序平台中,如Illumina®目前制造的那些,测序操作开始时一定数量的碱基可用于集群鉴别和“不变碱基”,即,读取为与所测序的所有或实质上多个分子相同的碱基可能影响此方法的效率。在此情况下,测序操作开始时紧接着的简并或半简并SMI序列可因此更合乎需要。

[0112] 在其它应用中,图2A中所示的定向为优选的。如双重测序的初始描述(即, W02013142389A1)中所描述,互补双链SMI序列可最适宜地通过用聚合酶引物延伸通过单链简并或半简并序列或通过单独地合成和退火含有不同SMI序列的寡核苷酸,且随后将这些合并在一起以产生不同衔接子库来产生。如果选择聚合酶延伸方法,那么在衔接子的连接结构域末端上具有SMI序列可能有利于延伸反应。在某些测序平台上,如通过Ion Torrent™制造的那些、在衔接子的非可连接的末端处具有经修饰的碱基的3'突出端可能不易于与通过聚合酶合成相容;由此通过聚合酶延伸途径合成衔接子最容易用位于如图2A中所示的衔接子的可连接的末端的SMI序列进行。

[0113] 作为此途径将如何实践的特定实例,考虑Ion Torrent™测序平台,其可所述以下衔接子对:

[0114] 衔接子P1

[0115] 5' CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT 3' (SEQ ID NO:1)

[0116] 3' T*T*GGTGATGCGGAGGCGAAAGGAGAGATACCCGTCAGCCACTA 5' (SEQ ID NO:2)

[0117] 衔接子A

[0118] 5' CCATCTCATCCCTGCGTGTCTCCGACTCAG 3' (SEQ ID NO:3)

[0119] 3' T*T*GGTAGAGTAGGGACGCACAGAGGCTGAGTC 5' (SEQ ID NO:4)

[0120] 星号“*”代表硫代磷酸酯键。

[0121] 测序引物退火到衔接子A,且由此序列信息由始于衔接子A的3'末端的DNA片段读出。衔接子A可在使用以下序列的情况下转化成可应用于图2中的所描绘的途径的形成:

[0122] 5'CCATCTCATCCCTGCGTGTCTCCGACTCAG GCGC NNNNG 3'(SEQ ID NO:5)

[0123] 3'T*T*GGTAGAGTAGGGACGCACAGAGGCTGAGTC ATAT MMMC 5'(SEQ ID NO:6)

[0124] NNNN是指简并或半简并四种核苷酸序列;MMMM是指其互补序列;且在简并序列下游包括G-C碱基对以便于连接,但也可使用连接结构域的其他形式。

[0125] 在此图示中,衔接子P1和衔接子A都连接到待测序靶DNA分子。为简单起见,可忽略连接到DNA分子的两端的相同衔接子。然而,Ion Torrent™衔接子利用分子的每一端上的不同衔接子。在初始连接后,个别DNA分子可与各种配置中的衔接子,例如A-DNA-P1、A-DNA-A或P1-DNA-P1连接。借助于在针对位点A和P1利用引物的乳液PCR中扩增,A-DNA-P1的正确配置可用于测序反应。替代地,可使用所属领域中已知的用于仅连接到两种不同衔接子的分子的其他选择方法。

[0126] 在扩增和测序后,将获得以下产物:

[0127] GCGC NNNN[DNA序列]

[0128] TATA NNNN[DNA序列]

[0129] 应注意,这些对应于产物X- α -DNA和Y'- α -DNA,如图2G中所示。

[0130] 经由如原先所描述的双重测序来自两个链的产物可随后共同地匹配用于数据处理(参见例如W02013142389A1)。具体地说,共有序列可由以序列GCGC NNNN开始的读段制成以获得“顶部”链的共有序列。单独的共有序列可由从序列TATA NNNN开始的读段制成以获得“底部”链的共有序列。可随后比较两个单链共有序列以获得起始DNA分子的双螺旋体共有序列。下文公开替代数据处理途径;参见“用于双重测序的替代数据处理流程”。

[0131] 上述途径能够使得在利用不能够成对端读段的短读段的平台上进行双重测序,如在此实施例中,DNA序列信息仅需要来自DNA片段的两个末端中的一个。

[0132] 此途径的替代实施例将为经由使用双链非互补或部分非互补SMI将不对称性引入到SMI序列自身中。尽管SMI序列自身将不互补,借助于预定形成对,可测定出由非互补SMI序列产生的产物从相同起始双链DNA分子出现。

[0133] 作为此实施例的特定实例,考虑一系列具有以下序列的Ion Torrent™“衔接子A”分子:

[0134] 衔接子1:

[0135] 5'CCATCTCATCCCTGCGTGTCTCCGACTCAG AAAT GCAGC 3'(SEQ ID NO:7)

[0136] 3'T*T*GGTAGAGTAGGGACGCACAGAGGCTGAGTC GGGCCGTCG 5'(SEQ ID NO:8)

[0137] 衔接子2:

[0138] 5'CCATCTCATCCCTGCGTGTCTCCGACTCAG ATATGCAGC 3'(SEQ ID NO:9)

[0139] 3'T*T*GGTAGAGTAGGGACGCACAGAGGCTGAGTC GCGCCGTCG 5'(SEQ ID NO:10)

[0140] 衔接子3:

[0141] 5'CCATCTCATCCCTGCGTGTCTCCGACTCAG TATTGCAGC 3'(SEQ ID NO:11)

[0142] 3'T*T*GGTAGAGTAGGGACGCACAGAGGCTGAGTC GGCCCGTCG 5'(SEQ ID NO:12)

[0143] 衔接子4:

[0144] 5'CCATCTCATCCCTGCGTGTCTCCGACTCAG ATTTGCAGC 3' (SEQ ID NO:13)

[0145] 3'T*T*GGTAGAGTAGGGACGCACAGAGGCTGAGTC CGGGCGTCG 5' (SEQ ID NO:14)

[0146] 为简单起见,上文仅列出四种衔接子,但实际上可能需要具有较大此类衔接子池。应注意,在本实例中,在非互补序列的下游包括互补序列以形成将促进连接到DNA分子的双链区域。

[0147] 个别DNA片段连接到个别衔接子,其导致两个DNA链的不对称标记。确切地说,在测序后,起始DNA分子的“顶部链”的序列将用衔接子的“顶部链”中的序列标记。起始DNA分子的“底部链”的序列将用衔接子的“底部链”中的序列的反向互补序列标记。

[0148] 作为特别的例子,连接到衔接子1的两个DNA链将为标记的AAAT (顶部链) 和CCCG (底部链)。再次,应注意,底部链在测序后产生首先存在于衔接子的底部链中的序列的反向互补序列。同样地,对于连接到其它衔接子的序列,分子标识符可借助于其配对标签共同地配对。计算机程序可随后使用来自衔接子的已知标签序列的表以将其汇编到由单个DNA分子的互补链产生的读段中。表2显示所得序列读段将如何基于上述实例中所显示的特定非互补标识符序列标记。

[0149] 表2.

[0150] 测序读段的前四个核苷酸

	顶部链	底部链
衔接子 1	AAAT	CCCG
[0151] 衔接子 2	ATAT	CGCG
衔接子 3	TATT	CCGG
衔接子 4	ATTT	GCCC

[0152] 这些仅为特定实施例的特定实例。所属领域的技术人员将显而易见,SMI标签可为任何任意长度,SMI可为完全随机的,或由完全预定义序列组成。当SMI序列在双链分子的两个链中时,两个SMI序列可为完全互补的(如上文实例提及的第一例子中所描述)、部分非互补的或完全非互补的。在一些实施例中,根本不需要外源性分子标识符标签。在一些情况下,可使用随机剪切的DNA分子的末端作为独特标识符,只要存在不对称性(包含SDE)的一些分类即可,其允许区分如双链DNA的给定单分子的两个非依赖性链产生的产物。

[0153] 在任何本文所公开的方面或本发明的实施例(且不限于目前描述的实施例)中,在两个单链和双链SMI中,SMI标签集合可设计有不同标签之间的编辑距离以使得合成、扩增或测序SMI序列的误差将不会使得一个SMI序列转化为另一个(参见例如Shiroguchi等人,《美国国家科学院院刊(Proc Nat Acad Sci USA)》,109(4):1347-1352)。在SMI序列之间掺入编辑距离允许例如通过使用识别汉明距离(Hamming distance)、汉明编码(Hamming code)或所属领域中已知的误差校正的另一方法和去除SMI误差。来自集合的所有SMI可为相同长度;或者可在一组SMI内采用两个或更多个不同长度的SMI的混合物。使用SMI长度的混合物可有利于使用SMI序列且另外具有SMI内或侧接SMI的位点处的一个或更多个固定碱基的衔接子设计,因为在集合内利用多于一个长度的SMI将使得在测序期间在相同读段位置处并不都存在不变碱基(参见例如Hummelen R等人,《公共科学图书馆·综合(PLoS

One)》,5(8):e12078(2010))。此途径可避开测序仪平台上可能产生的问题,在特定读段位置处存在不变碱基的情况下,所述平台可能遇到次优的性能(例如集群鉴别困难)。

[0154] 所属领域的技术人员还将显而易见,可在测序衔接子内任何地方在SMI序列之前或之后或在具有此类序列的衔接子设计中的单链“尾部”序列内引入可引入不对称性的序列,包括例如作为如上文所示的内部“气泡”序列。这些序列以及任何相关SMI序列可直接读取为测序读段部分,或者可由非依赖性测序反应测定(例如在索引读段中)。这些序列可另外与Y形衔接子、“环”衔接子或所属领域中已知的任何其它衔接子设计结合使用。

[0155] 实际上,预想具有不同相对定向的SMI序列、SDE序列和引物结合位点的衔接子且包括于本发明中。

[0156] 图2A和图2H中所示的衔接子设计显示非连接末端为平末端的。然而,此末端可为突出的、凹陷的或具有经修饰碱基或化学基团以防止降解或非所需连接。

[0157] 另外,衔接子的两个链可连接形成闭合的“环”,其在一些应用中可以是合乎需要的以防止降解或非所需连接。参见例如图2I。图2I的闭合的“环”连接(标记在位置“S”处)可通过常规磷酸二酯键或通过任何其它天然或非天然化学连接基团来实现。此键可以化学或酶促方式裂解以在进行连接之前、期间或之后获得“打开的”末端;在PCR扩增之前可能需要使所述环裂解以防止滚环型扩增子。此处可使用如尿嘧啶的非标准碱基,且在衔接子连接之前、期间或之后可使用步骤的酶促集合以使磷酸二酯主链裂解。举例来说,在尿嘧啶的情况下,将能够使用尿嘧啶DNA糖基化酶(用以形成无碱基位点)和核酸内切酶VIII(用以使主链裂解)的组合。替代地,此连接位点处的庞大化学基团或其它非可横移修饰碱基可用以防止聚合酶横移超出环末端且用于相同目的。

[0158] 在用于使用双链SMI序列的衔接子设计的本发明的任何本文中所公开的方面或实施例(且不限于目前所描述的实施例)中(无论其是否为互补、部分非互补或完全非互补的),合成衔接子作为退火到“环”形式中的线性分子的特定优点为“顶部”和“底部”链SMI序列将以1:1比率存在于分子自身内。此途径相对于退火个别“顶部”和“底部”寡核苷酸对可为有利的以形成双链SMI,如在此类途径中,如果用于“顶部”和“底部”链的寡核苷酸浓度不为完美的1:1比率,那么可能存在一个衔接子链或另一个的过多分子,且在下游步骤可能成为问题的(例如额外单链寡核苷酸可能在PCR扩增期间造成不当引发,或可能与可能存在的其它单链寡核苷酸一起退火,其可能产生衔接子分子,其中两个SMI链不当地配对)。

[0159] 在一些情况下,可能需要防止全部环序列自身复制,其中经修饰的序列位置可任选地包括为复制阻断。此可为可以酶促方式去除的碱基(例如尿嘧啶,其可通过尿嘧啶DNA糖基化酶去除)或例如部分或完全抑制DNA复制的区域(例如无碱基位点)。

[0160] 或者或另外,可引入限制性核酸内切酶位点(在图2I中的位置“T”处经标记),其可用以获得“开口”结构,最终释放较小发夹片段。

[0161] 应显而易见,两个衔接子链之间的碱基不对称性的不同配置同等地充当链限定元件。当相对于图2J的衔接子中所示的另外互补链嵌入一个核苷酸或大于一个核苷酸时,可在衔接子链中形成气泡。图2K显示衔接子,其中多于一个核苷酸嵌入包括自互补部分;此后一衔接子提供与涉及一个或更多个核苷酸位置的两个链之间的简单差异类似的官能基。

[0162] 使用非互补SMI序列引入链限定不对称性

[0163] 图2A到2K中所示的衔接子设计含有两个实现基于标签的双重测序的关键特征。一

个为独特分子标识符(即,SMI)且另一个为在两个DNA链(即,SDE)中引入不对称性的构件。在双重测序的初始描述中,利用Y形衔接子和成对端测序读段。在两个DNA链中引入不对称性借助于不对称尾部自身实现。如图3A中所示的不同且优良的双重测序衔接子设计包括非互补“气泡”形SMI,其共同用作分子标识符以及不对称性引入SDE。

[0164] 在这个设计中,P和P'分别代表引物位点和其互补序列,且 α_i 和 α_{ii} 代表两个对于其长度的全部或一部分为非互补的简并或半简并序列。衔接子的此形式的合成最易于通过在将这些中的两个或更多个合并在一起以形成多样池之前单独地合成和杂交具有不同简并或半简并序列的寡核苷酸对实现。因为寡核苷酸单独地合成和退火,所以给定 α_i 和 α_{ii} 序列之间的关系将为已知的且记录在数据库中,可检索所述数据库中后测序分析期间相应的配偶体SMI序列。

[0165] 在衔接子连接到双链DNA片段之后,产生图3B中所示的结构。在此结构中, β_i 和 β_{ii} 一对非互补SMI序列一般不同于 α_i 和 α_{ii} ,但相同衔接子结构可连接到两端。

[0166] 在PCR扩增之后,衍生自“顶部”链的双链产物显示于图3C中且衍生自“底部”链的双链产物显示于图3D中。

[0167] 因为引物位点序列在分子(在本实例中)的两端处相同,那么两种不同类型的序列读段将获自每一链的PCR产物的单端测序读段且取决于进行测序的单链。来自“顶部”链PCR产物的单端测序读段显示于图3E中且来自“底部”链的单端测序读段显示于图3F中。

[0168] 在分析期间,读段可随后通过特定SMI序列和其相应的非互补配偶体,基于从SMI衔接子库合成时产生且与其结合的数据库已知的关系分组。如图3G中所示,初始分子的配对的“顶部”和“底部”链序列标记有 α_i 和 α_{ii} (对于在分子一端上发端的读段)以及 β_i 和 β_{ii} (对于在相对端上的那些)。

[0169] 使用经修饰的或非标准核苷酸引入链限定不对称性

[0170] 可将链不对称性引入到双重测序衔接子中的另一方式为通过首先形成配对链DNA,但随后在另一生物化学步骤之后导致错配的核苷酸或核苷酸类似物。其中的一个实例为DNA聚合酶错掺入。错掺入可固有地在扩增期间或在经由化学或酶促步骤转化到错配区域之后发生。

[0171] 对于一些应用,此形式的SDE可优选为上文所公开的“气泡型”序列,因为者避免了可能由无单链区域产生的问题,例如错退火到其它DNA寡核苷酸和核酸外切酶/核酸内切酶降解。

[0172] 所属领域中已知的多种非标准核苷酸可用于此目的。此类经修饰的核苷酸的非限制性实例包括四氢呋喃;8-氧代-7,8-二氢-2'-脱氧腺苷(8-氧代-A);8-氧代-7,8-二氢-2'-脱氧鸟苷(8-氧代-G);脱氧肌苷、5'-硝基咪唑;5-羟基甲基-2'-脱氧胞苷;异胞嘧啶;5'-甲基-异胞嘧啶;以及异鸟苷和所属领域中已知的其它核苷酸。

[0173] 含有8-氧代-G的双重测序衔接子显示于图4A中。8-氧代碱基与互补胞嘧啶碱基相对配对且无气泡形成。正如上文和下文实例的情况,SMI序列(在此情况下为 α)和SDE位点(在此情况下为8-氧代-G位点)的相对顺序可按需要切换。P和P'代表引物位点和其互补序列。

[0174] 在衔接子连接到双链DNA片段之后,产生图4B中所示的结构。

[0175] 可随后用糖基化酶,如氧代鸟嘌呤糖基化酶(OGG1)(潜在地与DNA连接酶结合以修

复具有裂解酶活性的糖基化酶可能发生的所得缺口)进行图4B的双链DNA的处理。在引入无碱基位点的情况下,此处理将产生完整的磷酸二酯DNA主链,如图4C中所示。两个链中的每一个可随后用例如聚合酶拷贝。在适当的反应条件下,某些热稳定聚合酶优先插入A相反无碱基位点(Belousova EA等人,《生物化学与生物物理学报(Biochim Biophys Acta)》2006),产生G→T突变。相比之下,互反链保留连接时衔接子中存在的C核苷酸。此处理导致链不对称性,从而允许区分两个链产物。

[0176] 在PCR或其它形式的DNA扩增期间,在利用特定聚合酶的某些条件下,当拷贝链时,腺嘌呤将优先与无碱基相对插入。随着后续几轮拷贝,此腺嘌呤将与胸腺嘧啶配对,最终用T置换初始8-氧代-G位点。此外,糖基化酶处理不为必选的。在适当的反应条件下,聚合酶可与不具有所示无碱基中间物的8-氧代-G相对插入A(Sikorsky JA等人《生物化学与生物物理学研究通讯(Biochem Biophys Res Commun)》2007)。在任一情况下,在PCR扩增之后,衍生自“顶部”链的双链产物将如图4D中所示,且衍生自“底部”链的双链产物将如图4E中所示。

[0177] 因为引物位点序列在本(非限制性)实例中的分子的两端处相同,取决于进行测序的单链,两种不同类型的序列读段将获自每一链的PCR产物的单端测序读段。衍生自“顶部”链PCR产物的那些PCR产物将如图4F中所示,且衍生自“底部”链的那些PCR产物将如图4G中所示。

[0178] 在分析期间,测序读段可通过含有特定SMI,在此情况下α或β的那些分组。参见图4H。每一SMI分组内的T和G标记的产物限定起点链且允许双螺旋体序列比较。

[0179] 所属领域的技术人员还将显而易见,如上文所描述的经修饰的核苷酸或另一类似物可放入测序衔接子内的任何地方,只要在DNA测序时获自经修饰的核苷酸或其它类似物的序列可回收即可。

[0180] 所属领域的技术人员将显而易见,可利用多种其它核苷酸类似物实现相同目的。其它实例包括四氢呋喃和8-氧代-7,8-二氢-2'-脱氧腺苷(8-氧代-A)。可固有地通过DNA聚合酶产生不同核苷酸的错掺入或可通过酶促或化学步骤或自发地随时间推移转化成错编码病灶或错配碱基的任何核苷酸修饰都可用于此实施例的衔接子中。

[0181] 此外,可掺入非核苷酸分子以不对称标记两个链。举例来说,生物素可掺入到两个衔接子链中的一个中,其将通过利用抗生物素蛋白链菌素以物理方式分离含有生物素的链与不具有的链来促进两个链的单独分析。此实施例详细公开于下文中。

[0182] 使用双重测序衔接子设计的组合以在DNA分子的相对端上引入不同引物位点

[0183] 非Y形衔接子的前述实例显示相同类型的衔接子对称连接到DNA分子的两端。目前,大部分测序平台需要经调适的DNA分子在任一末端上具有不同引物位点以例如允许集群在表面或珠粒上扩增。对于并不常规地使用Y形衔接子产生这些不同引物位点的测序平台(例如Ion Torrent™(Thermo®公司)、SOLiD(Applied Biosystems®公司)和454(Roche®公司)),连接两种不同衔接子的混合物,且随后选择含有每个引物位点中的一个的分子;最常经由基于珠粒的乳液PCR方法。

[0184] 下文说明了一种使用非Y形双重测序衔接子产生不对称引物位点的简单途径。

[0185] 对此,产生一个双螺旋体衔接子和一个标准衔接子的混合物,其中每一衔接子含有不同PCR引物位点。双螺旋体衔接子可为本文中以上或以下所描述或如所属领域中已知

的任何设计。

[0186] 示例性双螺旋体衔接子显示于图5A中,其具有引物位点P与互补序列P',继而由错配序列X和Y构成的SDE,各自包含一个或更多个核苷酸,继而简并或半简并SMI序列 α 。图5B中所示的其它衔接子为含有不同引物位点O与互补序列O'的“标准”衔接子。

[0187] 在此衔接子混合物连接到DNA库之后,产生三种不同类型的产物,如图5C到图5E中所示。平均起来,成功调适的分子中的一半将在每一端上携带不同衔接子序列(图5C),四分之一将具有两个双螺旋体衔接子(图5D),且四分之一将具有两个标准衔接子(图5E)。在适当的选择条件下,仅具有一个引物位点P和一个引物位点O的分子将集群扩增。因此,后两种(非适用的)类型的产物以后可忽略,且后续描述中未示出。

[0188] 在PCR扩增之后,衍生自“顶部”链的双链产物将如图5F中所示,且衍生自“底部”链的双链产物将如图5G中所示。

[0189] 引物位点P测序将产生衍生自“顶部”和“底部”链的以下序列。这些可借助于携带SDE X或Y标记区分。参见图5H。

[0190] 显而易见,本文所描述或如所属领域中已知的任何其它形式的非Y形双螺旋体衔接子可用于与在此实施例中所使用相同的目的。举例来说,替代一个双螺旋体衔接子和一个标准衔接子,有可能使用两个携带不同引物位点的双螺旋体衔接子。在连接和PCR之后,扩增产物可分离且一个部分用引物P测序且另一个用引物O测序。这将使得能够双重测序每一调适分子的两端。因为来自不同引物位点的读段实际上不为成对端的,所以其无法容易地一起联系在一起用于任何特定分子。然而,对于待测序DNA数量极有限的应用,获自分子两端的双重测序的额外序列信息又可为有利的。

[0191] 在非成对端平台上使用两个读段可在双重测序期间使读段长度最大化

[0192] 如在Illumina®仪器上进行的成对端测序一般需要测序平台能够对来自衔接子DNA分子的一端上的引物位点的一个链进行测序,且随后产生反向互补序列链,之后对来自不同引物位点的分子的另一端进行测序。其技术挑战包括互补链生产过程,其为并非所有平台都能容易地与此成对端测序相容的原因。

[0193] 然而,可在有限程度上实现对经调适的DNA分子的两个不同部分进行测序的能力,而不需要产生互补链。此可通过使用相对于第一衔接子附接在DNA分子的相对端处的第二衔接子内所含有的第二引物位点以使得测序读段远离DNA分子和第一衔接子前进,由此产生第二衔接子自身的测序读段来实现。在一些情况下,此类能力可能是合乎需要的。举例来说,因为双重测序所需的SMI和SDE序列消耗可实现的固有地有限读段长度的一部分,当需要最大读段长度时,能够将这些元件移动到第二较短读段期间读取的相对衔接子可能为有帮助的。可通过通常用于样品多路复用的索引条码序列重定位实现类似益处。

[0194] 为了实现此过程,可使用两种不同衔接子。如图6A中所示的第一者含有与其互补序列P'相对的简单引物位点P。

[0195] 如图6B中所示的另一衔接子序列含有双重测序所必需的特征,而无Y形尾部:SMI和SDE。此双螺旋体衔接子可为本文所描述的设计中的任一种且其中SMI和SDE为单独序列元件,合并到相同序列元件中作为未配对SMI,或其中SDE由经修饰碱基构成。

[0196] 在图6B中所示的实例中,SDE需要邻近简并或半简并SMI序列 α 的错配序列X和Y。PCR引物位点O与互补序列O'在衔接子的未连接末端上。此衔接子设计的特点为第二引物位

点P2与邻近可连接的末端但经定向以使得退火引物将延伸到衔接子分子自身中而非朝向DNA片段的互补序列P2'。

[0197] 在此衔接子混合物连接到DNA库之后,产生三种不同产物。在相对端上具有两种相同衔接子类型的那些可忽略,因为仅具有每一衔接子中的一个的产物(含有两个引物位点P和O,如图6C中所示)将成功地集群扩增和测序。

[0198] 在PCR扩增之后,衍生自“顶部”链的双链产物将如图6D中所示,且衍生自“底部”链的双链产物将如图6E中所示。

[0199] 下文显示退火测序引物P1和P2和可逐一测序的区域的定向。这些读段将最适宜地用一个,之后另一个测序。这将通过引入一个测序引物和经受第一测序读取;随后在第一测序读取完成之后引入第二个来实现。如果首先进行“2号读取”(如图6F和图6G中所示),那么可进行测序直到达到分子末端为止且将自终止测序。如果首先进行“1号读段”,那么将需要在添加引物P2以开始“2号读段”之前中止此测序反应。此可通过引入在掺入之后不进一步延伸的经修饰的dNTP或通过使远离模板链的初始测序反应期间合成的链以热或化学方式熔融,且在添加下一测序引物之前将其洗掉来实现。

[0200] 衍生自“顶部”链的测序模板链的布置如图6F中所示,且衍生自“底部”链的测序模板链的布置如图6G中所示。

[0201] 来自“顶部”链衍生的模板的测序读段将如图6H中所示,且来自“底部”链衍生的模板的测序读段将如图6I中所示。

[0202] 显而易见,借助于携带SDE X标记或SDE Y标记,来自不同初始链分子的测序读段对为可区分的。

[0203] 在非成对端平台上使用两个读段使双重测序的标签多样性最大化

[0204] 源于使用上文公开的形式的双重读取的潜在优点不仅仅是简单地保留读取长度。在Y形衔接子的基于标签的双重测序的初始描述中,一个SMI序列附接到经调适的DNA分子的每一端。此设计在某些情形下具有实际优点,以便于高效地产生足够大群体的含有多样SMI的衔接子以确保可独特地标记每个DNA分子。

[0205] 作为示例,如果将完全简并四种核苷酸SMI序列引入到初始Y形衔接子设计中且连接到DNA片段库(如图1B中所示)且用成对端读取进行测序,那么可标记的分子的可能的方法的总数目为 $4^4 \times 4^4 = 65,536$ 。如果将完全简并8个碱基对SMI序列掺入双螺旋体衔接子中且连接到DNA库以便单端读取(如图5C中所示),那么可实现同样65,536个标记组合。当用聚合酶延伸方法产生互补SMI标签时,获得65,536个标记物的这两个手段将为同样可行的,然而当用单独合成的寡核苷酸产生衔接子库时则情况有所不同。在第一情境中,将需要产生总共 $4^4 \times 2 = 512$ 个寡核苷酸。在后一情境中,将需要产生和单独退火 $4^8 \times 2 = 131,072$ 个;这将极大地提高所需财务成本和工作量。

[0206] 对于双重测序的一些实施例,SMI衔接子产生的寡核苷酸合成方法为优选的且含有足够多样SMI的衔接子群体可能几乎不可能用如上文所公开的分子一端上仅单个SMI实现。

[0207] 在非成对端可相容平台上的上文所描述的双重读取方法可通过在两个相同反应步骤中在用于测序的两个衔接子中包括SMI序列用以克服此限制。下文说明了这一点。

[0208] 对此,需要两种类型的衔接子,各自带有不同扩增引物位点。至少一个必须含有

SDE,且在实例中,两个将含有简并或半简并SMI序列。如图7A中所示,第一衔接子类似于图6A的衔接子,不同之处为其另外包括SMI序列(此处鉴别为“B”)。如图7B中所示,第二衔接子类似于图6B中所示的衔接子且含有SMI序列(此处鉴别为“a”)。

[0209] 所属领域中有经验的技术人员将显而易见,两个衔接子的SMI和SDE特征的相对配置可互换以获得相同结果。取而代之,上文后一衔接子中显示的SDE可放置于前者中。此前描述的任何形式的SDE或SMI可经本实例中使用的那些的等效作用取代。

[0210] 在此衔接子混合物连接到DNA库之后,产物成功地与将如图7C中所示的每一衔接子类型中的一种结合。

[0211] 在PCR扩增之后,衍生自“顶部”链的双链产物将如图7D中所示,且衍生自“底部”链的双链产物将如图7E中所示。

[0212] 如前述实施例中所描述,测序引物位点P1和P2的定向和通过各自测序的区域如图7F中所示(对于“顶部”链)和如图7G中所示(对于底部链)。

[0213] “顶部”链衍生的模板的读段将如图7H中所示且衍生自“底部”链衍生的模板的读段将如图7I中所示。

[0214] 再次,两个链的产物可容易地借助于其不同X和Y SDE标记物区分。对于双重测序分析,SMI_a和SMI_B的序列可组合为单个鉴别标签序列。

[0215] Y形双重测序衔接子中的不对称SMI

[0216] 若干目前可用的测序平台需要DNA分子的相对端上的不同引物位点以允许集群扩增和测序。此可用具有不对称引物结合位点的Y或气泡形衔接子或经由紧接着前述三个实施例中说明的两种衔接子连接方法实现。Y形衔接子已最常用于成对端测序可相容平台,如通过Illumina®制造的那些;然而,其可在其它平台上使用。

[0217] 库制备的Y或“气泡形”衔接子的一个通用优点理论上为每个双重调适的DNA分子将能够经受测序。然而,在使用两种不同衔接子的方法的情况下,仅一半产生的分子将能够经受测序,其具有每一衔接子类型中的一种,而产生的另一半分子将具有相同衔接子的两个拷贝。在某些情形下,例如在输入DNA受限制的情况下,Y形衔接子的更高转化率可为合乎需要的。

[0218] 然而,如上文所描述的第一实施例(原先描述的双重测序方法)中所示,在无成对端读段或完全通读的能力的情况下,原先描述的Y形双螺旋体衔接子并不易于允许具有单端读段的双重测序。

[0219] 然而,使用Y形衔接子的互补“茎”序列中的测序引物位点允许用于双重测序的单端读段,但仅在通过衔接子序列中的至少一个SDE其它地方引入不对称性时才可。简要图示如下。

[0220] 在图8A中,显示Y形衔接子,其含有包含序列a_i和a_{ii}的未配对SMI。这个设计中的此序列将还充当SDE。存在三个引物位点:A和B,其为游离尾部上的PCR引物;以及C(和C'),其包括测序引物位点(和其互补序列)。

[0221] 在衔接子连接到DNA片段之后,产生图8B中所示的结构,其中具有两个不同非互补SMI的两个衔接子附连到任一末端。

[0222] 在使用与位点A和B的引物互补的PCR扩增之后,衍生自“顶部”链的双链产物将如图8C中所示,且衍生自“底部”链的双链PCR产物将如图8D中所示。

[0223] 在由引物位点C测序之后,两种不同类型的测序读段将获自每一链的PCR产物的单端读段,取决于其测序进行的一半单链。来自“顶部”链PCR产物的测序读段如图8E中所示,且衍生自“底部”链PCR产物的测序读段如图8F中所示。

[0224] 在分析期间,测序读段可通过特定SMI序列和其相应的非互补配偶体,基于从SMI衔接子库合成时产生且与其结合的数据库已知的关系分组。其中,如图8G中所示,初始分子的配对的“顶部”和“底部”链序列标记有 α_i 和 α_{ii} (对于在分子一端上发端的读段)以及 β_i 和 β_{ii} (对于在相对端上的那些)。

[0225] 可因此进行双螺旋体序列分析。类似于上文实施例中标题为“使用非互补SMI序列引入链限定不对称性”所描述的分析。

[0226] 此类型Y形衔接子的如图8H中例示的替代设计包括闭环,其有利于防止核酸外切酶消化或潜在地非特异性连接到Y的游离臂以及游离臂的“菊链”。闭合的“环”键(以箭头标记)可通过常规的磷酸二酯键或通过任何其它天然或非天然化学连接基团来实现。此连接可以化学或酶促方式裂解以在连接已进行之后获得“开口”端,如在PCR扩增之前将通常所需以防止滚环型扩增子。替代地,此连接位点处的庞大化学基团或经修饰的核苷酸可用以防止聚合酶横移超出环末端且用于相同目的。替代地,如图8I中所例示,在环(以箭头标记)内的发夹互补区域处引入限制性核酸内切酶识别位点;此可用以获得“开口”结构,最终释放较小发夹片段。

[0227] 在一些情况下,优选不需要在衔接子连接之后,在PCR之前进行额外酶促步骤。在额外步骤的需要不存在下,如图8J中例示的衔接子设计(其中衔接子尾部为互补,但不共价连接的)仍可克服由游离未配对DNA尾部引起的问题。

[0228] Y形双重测序衔接子中的不对称SMI

[0229] 关于Y形或环形衔接子中的未配对SMI的概念的另一变化包括位于PCR引物位点与互补茎之间的游离单链尾部区域中的这些未配对SMI。此设计的一个优点为其允许SMI完全测序作为“双重索引”读段的部分,如在选择Illumina®测序系统上可用的(Kircher等人(2012)《Nucleic Acid Res. (核酸研究)》第40卷,第1期,e3)。对于尤其需要长读段的应用,SMI不包括于主测序读段中将使DNA插入序列的读段长度最大化。一个实例如下。

[0230] 图9A显示含有未配对PCR引物位点A和B的Y形双螺旋体测序衔接子。 α_i 和 α_{ii} 代表一对至少部分地非互补简并或半简并SMI。 P 和 P' 为测序引物位点和其互补序列。

[0231] 在衔接子连接到DNA片段之后,产生图9B中所示的结构,由此具有两个至少部分地非互补SMI的两个衔接子附连到任一末端。

[0232] 在使用与位点A和B的引物互补的PCR扩增之后,衍生自“顶部”链的双链产物将如图9C中所示,且衍生自“底部”链的双链产物将如图9D中所示。

[0233] 在Illumina®平台上,作为实例,当使用具有双重索引的成对端测序时,在完成一个测序读取和一个索引读取之后,可产生互补链且可进行另一链的相应测序和索引读取。

[0234] 然而,应注意,成对端测序或双重索引技术本身允许双重测序。尽管给定PCR产物的两个单链有效地一起测序,每一PCR产物衍生自初始DNA双螺旋体的两个链中的仅一个,且因此对PCR产物的两个链进行测序不等同于对初始DNA双螺旋体的两个链进行测序。

[0235] 测序引物和索引引物以及其测序区域的可能的相对定向显示于图9E中(对于在两个方向上衍生自“顶部”链的PCR产物的读段)且显示于图9F中(对于在两个方向上衍生自

“底部”链的PCR产物的读段)。

[0236] 其还将满足对单个测序读段而非两个单独读段中的SMI和序列自身两者进行测序。显而易见,可利用多种不同配置和数量的引物来对SMI和读段序列进行测序。在一些实施例中,如纳米孔测序,SMI和/或DNA序列的测序可能根本不需要特异性引物位点。此外,此实例描述使用PCR,但此和其它实施例可通过所属领域中已知的任何其它方法,包括滚环扩增和其它途径扩增。参见Kircher等人(2012)。

[0237] 当在衍生自“顶部”和“底部”链(如图9G中所示)的那些方面比较所有四个读段中的不同模式的序列时,显而易见,其可彼此区别,因为一个携带SMI标签 α_i' 和 β_i 且另一个携带标签 α_{ii} 和 β_{ii}' 。尽管在此非限制性实例中两个链并不共享任何公共标签,但其可又彼此相关,因为 α_i 与 α_{ii} 之间和 β_i 与 β_{ii} 之间的关系在制备衔接子时从作为分析组件的数据库获知且可由此由数据库查找。

[0238] 使用单个环形载体引入用于双重测序的引物位点、SMI和SDE

[0239] 图10中说明了在单个分子而非两个配对衔接子中引入双重测序必需的所有元件的替代结构。

[0240] 在此实施例中,环形结构通过线性双链分子(包含双重测序必需的元件)的两个末端与具有可相容连接位点的DNA片段的两个末端附接形成。

[0241] 在图10A中,A/A'和B/B'代表两种不同引物位点及其反向互补序列; α 和 α' 需要简并或半简并SMI序列;且X和Y为SDE的相应非互补一半。

[0242] 在双链DNA片段连接到图10A的双链分子中之后,产生闭环,如图10B中所示。

[0243] 在产生图10B的连接产物之后,使用PCR从引物位点进行扩增。替代地,可首先进行滚环扩增。未连接的库和衔接子的选择性破坏可为有利的且用5'-3'或3'-5'核酸外切酶实现。环形设计独特地提供这些机会,其在多种其它设计的情况下不太可能容易实现。

[0244] 显而易见,上文和下文所描述的SMI和SDE的形式中的任一种可取代所示那些或其重排顺序。

[0245] 作为另一实施例的实例,如图10C中所示,可使用接近用作SMI和SDE的一个连接位点的单个元件,如标题为“使用非互补SMI序列引入链限定不对称性”的实施例中所论述。

[0246] 替代地,如图10D和图10E中所示,SDE和SMI可设计到接近衔接子连接位点中的每一个的序列中以便于成对端测序。

[0247] 在这个设计中,应注意,相对链上的SMI序列不必需为互补的(如图10E中所示),只要已知相应的序列(即 α_i 和 α_{ii})之间的关系即可且在分析期间可在数据库中查找。

[0248] 经由不对称化学标记和链分离的双重测序

[0249] 如上文所论述,双重测序根本上依赖于以其可区分的方式对DNA双螺旋体的两个链进行测序。在原先描述的双重测序的实施例中(在W02013142389A1中),两个链可与发夹序列连接在一起以共同地对配对链进行测序。W02013142389A1以及在上文所公开的多个实施例中描述了方法,其中独特DNA双螺旋体的两个链可使用DNA标记区分。此后一途径涉及用独特DNA序列标记每一DNA分子(包含DNA片段的一个或两个末端的坐标的内源性SMI或包含简并或半简并序列的外源性SMI)且经由至少一种形式的SDE引入链限定不对称性(例如具有成对端读取的不对称引物位点、“气泡”序列、非互补SMI序列和非标准核苷酸,其以天然或化学方式转化成错配)。

[0250] 下文公开了用于进行双重测序的另一途径,其包括双螺旋体中的两个链的不对称化学标记以使得对于非依赖性反应中的测序,其可以物理方式分离。其一个实例如下。

[0251] 如图11A中所示,使用两种不同衔接子。第一衔接子含有引物位点P与互补序列P'和SMI序列 α 与互补序列 α' 。第一衔接子的一个链另外携带能够结合或与已知物质(例如固体表面、珠粒、固定式结构和结合搭配物)结合的化学标签,以另一DNA链不会如此的方式。如图11A中所示,化学标签为生物素,其具有结合搭配物和对抗生蛋白链菌素的亲和性。

[0252] 可使用所属领域中已知的其它结合搭配物对,优选呈小分子、肽或任何其它独特地可结合的部分形式。此标记还可呈核酸序列(例如DNA、RNA或其组合;以及经修饰的核酸,如肽核酸或锁核酸)形式,优选呈单链形式,其中附连到固体基质(例如固体表面、珠粒或类似其它固定式结构)的基本上互补的“饵”序列可用以结合且选择性地捕集和隔离衔接子连接的分子的一个链与另一链。

[0253] 第二衔接子不携带此非限制性实例中的化学标签。如图11B中所示,第二衔接子带有不同引物位点O与互补序列O'。

[0254] 在图11A和图11B的衔接子连接到DNA片段之后,产生图11C中所示的(优选)结构。

[0255] 另外,将产生两种其它类型的结构:一个具有含有两个引物位点P的衔接子且另一个连接到含有两个引物位点O的衔接子。如上文标题为“使用双重测序衔接子设计的组合以在DNA分子的相对端上引入不同引物位点”的实施例中所论述,相对于其它两种类型结构,优选的结构增浓可常规地用特定扩增条件在测序之前实现以使得可忽略其它两种类型的结构。

[0256] 如图11D中所示,在连接之后,DNA链可以热或化学方式熔融分开,且随后带有化学标签(其具有对特定结合搭配物(在此情况下为抗生蛋白链菌素,例如与顺磁珠粒结合)的选择性亲和性)的链可与另一链分离。现分离的两个链可独立地测序,任选地利用前述步骤,其中两个分离链独立地扩增(测序可在物理上不同的反应中或在每一链应用不同索引之后相同反应中发生,例如用经标记的PCR引物和重组)。

[0257] 或者,两个链可用对两种不同类型的饵具有亲和力的不同化学标签标记。一个测序反应或索引基团中可见的标签可随后与另一群体中相应的标签进行比较且进行双重测序分析。在本实例中,又使用SDE,但其需要可用于以物理方式分离所述链的不对称附着的化学标签。其物理上不同的区室化允许两个链单独地测序或在合并之前进行后续差异标记步骤(例如利用在其尾部上携带不同索引序列的引物的PCR)且合并的测序可稍后利用波射束进行去卷积。

[0258] 此概念的另一实施例将为使用具有其它特性的标记物(即,物理基团),其允许通过除化学亲和性外的手段实现链分离。举例来说,包含具有较强正电荷的分子(例如具有电荷特性的物理基团)的核酸链可优先经由施加电场(例如通过电泳)与其配对的未标记配对链分离,或包含具有较强导磁系数的分子(例如具有磁性特性的物理基团)的核酸链可优先经由施加磁场与其配对的未标记配对链分离。当在溶液中,在某些应用条件下时,可优先分离的包含对沉淀敏感的化学基团(例如具有不可溶特性的物理基团)的核酸链为配对的未标记配对链以使得DNA自身为可溶性的,但包含物理基团的DNA为不可溶的。

[0259] 在施加SMI(作为连接的衔接子测序仪内的外源性标签或包含DNA片段的独特剪切点的内源性SMI)之后配对链的物理分离的概念的又一变化为在DNA双螺旋体热或化学熔融

成其组分单链之后使用稀释。将单链稀释到多个(即,两个或更多个)以物理方式分离的反应腔室中,以使得共用相同容器的两个原先配对的链的机率较小,而非向一个链施加可纯化的化学标记,从而将其与另一个分离。举例来说,如果在一百个容器中通过随机分离混合物,那么仅约1%的配偶体链将放置于同一容器中。容器可需要一组物理容器,如容器、试管或微孔板中的孔或以物理方式分离的非连通液滴,例如水相/疏水相乳液。可使用任何其它方法,其中防止两个或更多个空间不同体积的含有核酸分子的流体或固体的内含物与核酸分子基本上互混。在每一容器中,可优选使用各自携带不同标签序列的引物,进行PCR扩增。通过每个容器中的不同引物添加的此独特标签序列将最适宜地位于在测序索引读段期间可记录的位置处(例如参见图9E)。这些标记物将充当SDE。在本实例中,约99%的携带相同SMI标记的配偶体链将指定为与其配偶体链不同的SDE标记。仅约1%将指定为相同标记。双重测序分析和共有序列制得可使用SMI和这些SDE照常进行。在少数情况下(其中配偶体链偶然获得相同SDE),在双螺旋体分析期间将固有地忽略这些分子且将不提供假突变。

[0260] 在切口平移期间引入SDE

[0261] 在一些设置中,如在用于Ion Torrent™平台的衔接子连接的可商购试剂盒中,双链衔接子连接到待测序的双链靶DNA分子。然而,此处,靶DNA分子的两个链中的仅一个连接到衔接子。当连接结构域的5'链为非磷酸化的时,这是通用实施例。在通常已知为“切口平移”的过程中,具有链置换活性的聚合酶随后用于从连接的链将序列拷贝到未连接的链上。如果以此方式使用本文所公开的衔接子设计且无修饰,那么在许多情况下SDE将在切口平移步骤期间损失;由此,预防双重测序。这例示于下文中。

[0262] 图12A中显示一种类型的双重测序衔接子。N'代表简并或半简并SMI序列;与GG相对的TT为非互补SDE区域;且星号系统表示非连接的去磷酸化5'碱基:

[0263] 在图12A的衔接子连接到双链DNA分子之后,保留一个未连接的切口,如图12B中所示。

[0264] 利用标准“切口平移”途径,链移动聚合酶用于延伸库DNA分子的3'末端且置换衔接子的未连接的链。这显示于图12C中。在延伸之后,非互补SDE如图12D中所示损失。当SDE损失时,双重测序无法进行,因为链为不可区分的。

[0265] 允许使用衔接子连接的切口平移方法且保留SDE的一个途径如下。

[0266] 图12E中显示已经修饰的以包括简并或半简并SMI序列的Ion Torrent™衔接子“A”的一个实例。应注意,不存在SDE。“A”为引物位点。星号系统表示非磷酸化5'碱基。图12F中显示Ion Torrent™P1引物的一个实例。P1表示引物位点。星号系统指示去磷酸化5'碱基。

[0267] 在图12E和图12F的每一衔接子连接到双链DNA之后,形成图12G的结构。未示出具有两个P1或两个A引物位点的产物,因为其将不集群扩增。为了清楚起见,未连接的衔接子链也未示出。

[0268] 随后,链移动聚合酶根据典型的切口平移方案添加(例如Bst聚合酶,如一些工业试剂盒中所使用,归因于其较强链置换活性)。然而,如图12H中所示,在本实例dGTP中,首先添加四种dNTP中的仅一个,且由此将发生T-dGTP错掺入(值得注意的是,此错掺入事件可用多种DNA聚合酶在适当的反应条件下发生;参见例如McCulloch和Kunkel,《细胞研究(Cell Research)》18:148-161(2008)和其中引用的参考文献)。

[0269] 尽管错配掺入可能在某些条件下极其有效,第二错配的错配延伸和产生极其低效

(McCulloch和Kunkel, 2008)。因此,在适当条件下,核苷酸掺入将在错配发生之后中止。此时,可添加剩余三种dNTP以使得聚合酶可利用所有四种dNTP。将拷贝衔接子序列的其余部分以形成图12I中所示的结构,其具有非互补位置以使得“顶部”链的扩增产物将可与“底部”链的扩增产物区分。

[0270] 在PCR之后,由初始“顶部”链产生的产物将如图12J中所示,且由初始“底部”链产生的PCR产物将如图12K中所示。

[0271] “顶部”链产物的测序将产生图12L中所示的结构,且“底部”链产物的测序将产生图12M中所示的结构。

[0272] 应注意,测序产物可基于引入的错配彼此区分。

[0273] 下文显示削弱此概念以用Ion Torrent™衔接子实践的特定实例。

[0274] Ion Torrent™衔接子可使用以下序列:

[0275] 衔接子P1

[0276] 5'CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT 3' (SEQ ID NO:15)

[0277] 3'T*T*GGTGATGCGGAGGCGAAAGGAGAGATACCCGTCAGCCACTA 5' (SEQ ID NO:16)

[0278] 衔接子A

[0279] 5'CCATCTCATCCCTGCGTGTCTCCGACTCAG 3' (SEQ ID NO:17)

[0280] 3'T*T*GGTAGAGTAGGGACGCACAGAGGCTGAGTC 5' (SEQ ID NO:18)

[0281] 星号系统“*”表示硫代磷酸酯键。

[0282] 衔接子A的序列可如下经修饰。NNNN指示简并或半简并SMI序列(显示四种核苷酸,但此序列长度为任意的),且MMMM指示NNNN的互补序列。如先前所描述,可不利用SMI序列进行双重测序,但此处显示SMI作为施加双链分子标记的概念的特定实例。

[0283] 经修饰的衔接子A

[0284] 5'CCATCTCATCCCTGCGTGTCTCCGACTCAG NNNN AAC 3' (SEQ ID NO:19)

[0285] 3'T*T*GGTAGAGTAGGGACGCACAGAGGCTGAGTC MMMM TTG 5' (SEQ ID NO:20)

[0286] 衔接子A和P1附接到待测序DNA分子的相对端上。为简单起见,仅显示分子的衔接子A末端,且同样为简单起见,两个链分别显示为X'和Y'。可使用任何长度的任何DNA序列,只要测序片段的长度与所使用的测序方法相容。

[0287] “顶部”链为连接的,但“底部”链不为连接的,留下切口(显示为|)

[0288] 5'CCATCTCATCCCTGCGTGTCTCCGACTCAGNNNNAACXXXXXXXXX 3' (SEQ ID NO:21)

[0289] 3'T*T*GGTAGAGTAGGGACGCACAGAGGCTGAGTCMMMMTTG|YYYYYYYYY 5' (SEQ ID NO:22)

[0290] 链置换聚合酶与dGTP一起添加。在5'-3'方向上遇到的第一位置处(与C相对的G的正确掺入)以及遇到的第二位置处(与A相对的G的不正确掺入)掺入G。因为不正确碱基延伸在错配之后为低效的,在聚合酶浓度、反应时间和缓冲液条件的适当条件下,且聚合酶停止运作且进一步掺入不发生。应注意,“底部”衔接子链的前两个核苷酸在此反应期间被置换,且下文显示以下示意图中的衔接子-DNA构建体。新掺入的碱基以粗体指示。

[0291] 5'CCATCTCATCCCTGCGTGTCTCCGACTCAGNNNNAACXXXXXXXXX 3'

[0292] 3' T*T*GGTAGAGTAGGGACGCACAGAGGCTGAGTCMMMMT**GG**YYYYYYYYY 5'

[0293] \

[0294] TG

[0295] (顶部:SEQ ID NO:23和底部:SEQ ID NO:24)

[0296] 现在,将dCTP、dATP和dTTP添加到反应中,以使得所有四种核苷酸可供聚合酶使用。出于说明的目的,可用下文显示的中间产物进行链置换合成:

[0297] 5'CCATCTCATCCCTGCGTGTCTCCGACTCAGNNNNAACXXXXXXXXX 3'

[0298] 3'T*T*GGTAGAGTAGGGACGCACAGAGGCTGAGTCMMMMTGGYYYYYYYYY 5'

[0299] \

[0300] GCTGAGTCMMMMTTG

[0301] (顶部:SEQ ID NO:25,中间:SEQ ID NO:26,以及底部:SEQ ID NO:27)

[0302] 在达到模板的末端之后,衔接子的初始“底部”链完全置换(未示出)且完全合成的“底部”链用非互补单个碱基对存在(A:G碱基对,带下划线)

[0303] 5'CCATCTCATCCCTGCGTGTCTCCGACTCAGNNNNAACXXXXXXXXX 3' (SEQ ID NO:28)

[0304] 3'GGTAGAGTAGGGACGCACAGAGGCTGAGTCMMMMTGGYYYYYYYYY 5' (SEQ ID NO:29)

[0305] 此构建体可随后用于PCR扩增且根据典型的Ion Torrent™方案进行测序。值得注意的是,PCR扩增由“顶部”和“底部”链产生产物,且这些产物可借助于在切口平移期间引入的非互补碱基对彼此区分。

[0306] 由“顶部”链产生的产物将具有以下形式(碱基错配的位置带下划线):

[0307] TCAGNNNNAACXXXXXXXXX (SEQ ID NO:30)

[0308] 相比之下,由“底部”链产生的产物将具有以下形式(碱基错配的位置带下划线):

[0309] TCAGNNNNACCXXXXXXXXX (SEQ ID NO:31)

[0310] 应注意,“底部”链产物为首先存在于衔接子连接的DNA的“底部”链中的序列的反向互补序列(且因此,在测序为C核苷酸期间读出G核苷酸,其为在切口平移期间引入的碱基错插入)。

[0311] 现在,针对误差校正,由两个链中的每一个产生的扩增复制可彼此进行比较。由双链DNA的给定分子产生的“顶部链”产物将具有标签序列NNNNAAC。相比之下,“底部”链产物将具有标签序列NNNNACC。因此,出于误差校正的目的,两个链的复制品可分解,如先前所描述(Schmitt等人,PNAS 2012)。

[0312] 在切口平移之后引入错配

[0313] 上述替代途径将为利用所有存在的四种核苷酸的完全切口平移,且随后将模板链中的碱基改变为不同碱基。

[0314] 含有引物序列和其互补序列(P/P')、U-A碱基对(U=尿嘧啶)以及单链SMI序列和其互补序列(α/α')的衔接子显示于图13A中;星号系统表示去磷酸化5'末端。

[0315] 在图13A的衔接子连接到待测序双链DNA分子之后,单链切口保留在去磷酸化位点处,如图13B中所示。此处,“顶部”链借助于靶DNA分子中的5'磷酸酯连接,但“底部”链归因于不具有衔接子中的5'磷酸酯,不连接到靶DNA,留下切口。

[0316] 链置换合成可用聚合酶(例如Bst聚合酶)和所有四种dNTP进行,产生图13C中所示的结构。

[0317] 所得延伸产物现在正如在初始衔接子中一般再次出现。如图13D中所示,但不存在不对称性位点。

[0318] 可进行纯化步骤以去除聚合酶和dNTP。通过添加尿嘧啶DNA糖基化酶和适当的AP核酸内切酶,尿嘧啶可随后从“顶部”链(图13D中所示的结构)去除,产生如图13E中所示的单核苷酸缺口。

[0319] 随后,添加非链置换聚合酶(例如硫化叶菌属DNA聚合酶IV,其高度易错的且有助于碱基错掺入)以及单核苷酸,例如dGTP,但无其它核苷酸。在本实例中,这将产生G相对于A的错掺入。所得切口可用DNA连接酶密封,产生在衔接子中具有错配的产物,如图13F中所示。

[0320] 如图13G中所示,在扩增和测序之后,基于携带相同SMI序列的测序读段,借助于具有G或T,由“顶部”链产生的产物可与由“底部”链产生的那些区分。

[0321] 此实例用G-A错配产生示出,但将显而易见,分子中任何位置处的一个或多个碱基的任何其它错配将具有相同作用。

[0322] 下文显示在Ion Torrent™平台上应用此概念的特定实例。

[0323] 考虑以下“经修饰的衔接子A”,其中以粗体添加标准序列(U=尿嘧啶):

[0324]

5' CCATCTCATCCCTGCGTGTCTCCGACTCAG **UNNNN C** 3' (SEQ ID NO: 32)

3' T*T*GGTAGAGTAGGGACGCACAGAGGCTGAGTC **AMMM G** 5' (SEQ ID NO: 33)

[0325] 衔接子连接到如上所述的靶DNA分子,其中切口位置显示为“|”:

[0326] 5' CCATCTCATCCCTGCGTGTCTCCGACTCAGUNNNNCXXXXXXXXXX 3' (SEQ ID NO: 34)

[0327] 3' T*T*GGTAGAGTAGGGACGCACAGAGGCTGAGTCAMMMG|YYYYYYYYY 5' (SEQ ID NO: 35)

[0328] 现在,在所有四种dNTP存在下使用链置换聚合酶以允许衔接子的“底部链”的全部链置换(新掺入的碱基为粗体,初始底部衔接子链被置换且不显示):

[0329] 5' CCATCTCATCCCTGCGTGTCTCCGACTCAGUNNNNCXXXXXXXXXX 3' (SEQ ID NO: 36)

[0330]

3' **GGTAGAGTAGGGACGCACAGAGGCTGAGTCAMMMG**YYYYYYYYY 5' (SEQ ID NO: 37)

[0331] 纯化产物以去除dNTP,随后添加尿嘧啶DNA糖基化酶和AP核酸内切酶以从“顶部”链去除尿嘧啶,留下单核苷酸缺口:

[0332] 5' CCATCTCATCCCTGCGTGTCTCCGACTCAG NNNCXXXXXXXXXX 3' (SEQ ID NO: 38)

[0333] 3' GGTAGAGTAGGGACGCACAGAGGCTGAGTCAMMMGYYYYYYYYY 5' (SEQ ID NO: 39)

[0334] 随后,非链置换易错聚合酶(例如硫化叶菌属DNA聚合酶IV)与dGTP一起添加,其在单核苷酸缺口处导致G相对于A的掺入;可随后添加连接酶以在“顶部”链上产生完整的衔接子-DNA产物。此导致非互补碱基对(位置带下划线)。

[0335] 5' CCATCTCATCCCTGCGTGTCTCCGACTCAGGNNNCXXXXXXXXXX 3' (SEQ ID NO: 40)

[0336] 3' GGTAGAGTAGGGACGCACAGAGGCTGAGTCAMMMGYYYYYYYYY 5' (SEQ ID NO: 41)

[0337] 此产物可用于误差校正,利用与紧接在前的实施例中描述的类似的方法。

[0338] 在切口平移之后引入错配

[0339] 标题为“在切口平移期间引入SDE”的实施例显示不对称SDE如何可在切口平移期间在衔接子序列内引入。相同原理可应用于DNA分子库自身以使得不对称位点(SDE)掺入到库分子中,可能甚至在衔接子添加之前。此可实现多种方法。以下仅仅为一个实例。

[0340] 具有“顶部”和“底部”链的双链DNA分子显示于图14A中。DNA分子可通过多种方法碎片化以便库制备。如血浆中的无细胞DNA的一些DNA来源已呈小碎片形式且不需要单独的碎片化步骤。声波剪切为通常使用的方法。可使用半随机酶促剪切方法。在规定的识别位点处切割的非随机核酸内切酶为另一方法。在本实例中,留下5'突出端的核酸内切酶用于产生类似地5'突出片段的库,如图14B中所示。

[0341] 通过使用聚合酶,在仅与通过聚合酶拷贝的第一核苷酸不互补的单核苷酸存在下,此不对称状态可转化成序列不对称性。在本实例中,使用dGTP,其将导致T-dGTP错掺入(此类错掺入可在适当的反应条件下用多种DNA聚合酶进行;参见McCulloch和Kunkel,《细胞研究》18:148-161(2008),和其中引用的参考文献)。包括两个错配的部分双链DNA分子显示于图14C中。

[0342] 随后,将所有四种核苷酸添加到反应中且持续拷贝以延伸DNA分子末端直到DNA分子为双链的为止。错配气泡在每一片段末端上产生,形成两个SDE,如图14D中所示。

[0343] 双重测序衔接子可随后连接到DNA分子。图14E中所示的示例性衔接子具有引物位点P与互补序列P'、不同引物位点O与互补序列O'以及简并或半简并SMI α 与互补序列 α' 。

[0344] 在图14D的双链DNA分子与图14E的衔接子之间进行连接以产生图14F的结构。如前述实施例中所论述,可忽略连接到两个相同衔接子序列的产物,因为在适当条件下其将不扩增。

[0345] 在PCR之后,衍生自“顶部”链的产物如图14G中所示,且衍生自“底部”链的产物如图14H中所示。

[0346] 使用引物P的测序将由图14I中所示的相应链产生以下序列。

[0347] 应注意,在SMI序列之后,C与T的存在允许“顶部”链读段与衍生自“底部”链的那些区分开来。

[0348] 类似SDE标记可类似地通过使用诱变核苷酸类似物填补3'凹入端缺口或其它方法实现。

[0349] 可使用其它剪切方法且在填充之前以产生SDE的方式用核酸外切酶产生3'凹入端。

[0350] 从广义上来说,此实例说明SDE可以独立于衔接子自身的方式引入。对于发生的双重测序,每一最终调适的分子中的仅某一形式的SMI和SDE允许衍生自双螺旋体的每一链的序列彼此相关,但还决定性地彼此区分。这些元件以如上文所考虑的多种形式出现,且可在衔接子连接之前、期间或之后引入。

[0351] 组装适于双重测序的分子的变化

[0352] 上文所公开的实施例说明用于双重测序的改良方法,其中组装的最终分子包含至少一个链限定元件(SDE)和至少一个单分子标识符(SMI)序列;SDE和SMI都附接到待测序DNA的双链或部分双链分子上。然而,SMI和SDE不需要包括于单个衔接子中;其简单地需要存在于最终分子中,理想地在任何扩增和/或测序步骤之前或期间。

[0353] 举例来说,SDE可在经由酶反应连接之后在衔接子中产生,如图4D中所示。类似地,如原先所描述(在W02013142389A1中),在一些实施例中,个别DNA库片段的剪切点处的特定序列可充当内源性SMI序列,而不需要添加衔接子内包括的外源性SMI。“剪切点”可视为当片段与参考基因组比对时,DNA片段的任一末端的映射坐标。任一端或两端的坐标可用作

“内源性SMI”以单独或与一个或多个外源性SMI序列的序列组合使不同DNA分子彼此区分。

[0354] 以下清单包括此类衔接子的非限制性变异体：

[0355] --SDE存在于两个链中，但SMI和引物结合位点存在于仅一个衔接子链中。这些元件随后用聚合酶拷贝到另一链。

[0356] --不存在SDE；SMI和引物结合位点在仅一个链中。聚合酶与存在的仅一个不正确dNTP一起使用以产生SDE，且随后添加剩余dNTP以允许聚合酶制得SMI和引物结合位点双链。

[0357] --连接结构域仅存在于一个衔接子链中（以使得第二衔接子链不附接）。新第二衔接子链随后用聚合酶由第一衔接子链拷贝。此产生SMI和引物结合结构域。如上所述，首先仅添加一个不正确dNTP以产生SDE；随后添加剩余dNTP。此途径显示于上文公开的实施例中。

[0358] --连接结构域仅存在于一个衔接子链（以使得第二衔接子链不附接）中；此衔接子链包括尿嘧啶。新第二衔接子链随后用聚合酶，用存在的所有四种核苷酸由第一衔接子链拷贝。随后，初始衔接子链中的尿嘧啶碱基以酶促方式用尿嘧啶DNA糖基化酶和适当的AP核酸内切酶去除。随后，DNA聚合酶与存在的单个不正确核苷酸一起使用以将错配插入到DNA中的缺口中，且随后缺口用DNA连接酶连接。此途径更详细地显示于上文关于图4公开的实施例中。

[0359] --第一附接的衔接子单独在两个链中具有SMI结构域。第二衔接子随后附接到其上，其具有引物结合结构域和SDE，同样在两个链中。

[0360] --第一附接的衔接子在两个链中具有SMI和SDE结构域。附接第二衔接子，其在两个链中具有引物结合结构域。

[0361] --第一附接的衔接子在两个链中具有SMI结构域。随后附接第二“Y衔接子”，其具有两个非互补或部分非互补引物结合结构域。

[0362] --第一附接的衔接子在两个链以及单链区域，同样与连接结构域中具有SMI。寡核苷酸退火且连接到单链区域中；在产生SDE结构域的寡核苷酸内包括错配。

[0363] --在其它实施例中，可改变气泡的位置，可改变n-mer的长度，n-mer可与鉴别的每一链而非DNA分子末端处的剪切点的复制品一起消除。变异体核苷酸或核苷酸类分子可在DNA（例如锁核酸（LNA）和肽核酸（PNA）以及RNA）内使用。

[0364] 本文所公开的变异体中的每一个包括于本发明中。

[0365] 在这些变异体中的每一个中，适用相同通用概念：双重测序的最终分子包含连接到待测序DNA的链段的SDE和SMI的核心元件。还应注意，相同通用概念适用于双重测序的初始描述（在W02013142389A1中），其中用包含两个不对称引物结合位点的衔接子进行双重测序（例如在“Y”配置中）（其在此情况下充当SDE）和附接到双链DNA分子上的SMI序列。这些组分可通过多种方式组装到靶DNA分子上，只要必需的组分存在于最终分子中即可，理想地在任何扩增或测序步骤之前或期间。

[0366] 用于双重测序的替代数据处理流程

[0367] 可通过获得由两个个别DNA链中的每一个产生的扩增复制品的“共有序列”进行双重测序以获得两个单链共有序列，随后比较所得单链共有序列以获得双螺旋体共有序列。

在一些设置下(例如如果在严重受损DNA中的给定位置处可能发生的复发性扩增误差),可能不需要逐位置单个分子的扩增复制品的“平均化”序列的此途径且可因此在不同数据处理流程的一些设置下获得可靠的结果。

[0368] 替代途径包括以下:

[0369] --在具有对应于“顶部”和“底部”链的给定标签序列的分子中,任意挑选一个“顶部”链和一个“底部”链,且比较两个链的序列。保持两个链都一致的位置;标记不一致位置为非限定的。判读所得序列读段为双螺旋体读段。

[0370] --针对共用相同标签序列的任意选择的“顶部”和“底部”链重复此方法以获得一系列“双螺旋体读段”。

[0371] --在具有给定标签序列的所得“双螺旋体读段”中,选择具有例如相对于参考序列最少的序列变化和/或读段内最少的非限定位置的双螺旋体读段。此读段可随后被视为最可能代表起始DNA双螺旋体的真实序列的读段。

[0372] 在一个实施例中,此类途径可尤其用下文所描述的算法实现。应理解,这仅是出于例示的目的的单个实例,且多种其它算法可用以形成双螺旋体共有序列读段。另外,显示双重测序的特定实施例的实例,但可制备适于双重测序的多种其它实施例的类似实例。

[0373] 以下步骤可用于本文所公开的实施例,其使用“气泡”序列产生标记GCGC的每一双螺旋体的“顶部”链和标记TATA的“底部”链,其中两个链共用相同单分子标识符(SMI)序列。

[0374] 1.制作含有来自实验的所有测序读段的文件;

[0375] 2.将所述文件分成两个文件:一个文件称作标记GCGC的含有“GCGC”的读段,且第二个文件称作标记TATA的含有“TATA”的读段;

[0376] 3.在“GCGC”文件中挑选一个任意读段,读取其SMI标签,且在“TATA”文件中搜索匹配的SMI标签;

[0377] 4.如果发现匹配,那么由这两个序列产生一个新序列。在新序列中,维持一致的读段内的所有序列位置,且将两个读段中所有不一致的位置标记为非限定的。将此新序列写入称作“双螺旋体”的文件中,且从“GCGC”和“TATA”文件去除两个序列

[0378] 如果未发现匹配:那么从“GCGC”文件去除序列且将其写入称作“不匹配”的文件中;

[0379] 5.从“GCGC”文件挑选另一个任意读段,且再次进行步骤3到4;以及

[0380] 6.继续直到“GCGC”文件中无读段剩余为止。

[0381] 在所得“双螺旋体”文件内,设想所有读段具有匹配的SMI标签序列。在一些情况下,可能存在多个具有相同SMI标签的“双螺旋体”读段(这些可归因于例如单个起始DNA分子的多个PCR复制品)。这些可通过以下途径中的任一种转化成单个双螺旋体读段:

[0382] --在这些读段中,选择相对于参考基因组序列具有最少错配的读段且丢弃剩余读段。

[0383] --替代地,选择相对于参考基因组序列具有最少非限定位置的读段且丢弃剩余读段。

[0384] --替代地,在具有共用SMI标签序列的读段中产生共有序列以产生双螺旋体共有序列读段。

[0385] 所属领域的技术人员将显而易见,上述选项的组合可用于产生双螺旋体共有序列

读段,或可使用上文未描述的若干其它方法。

[0386] 其它实施例

[0387] 尽管本发明已结合其实施方式描述,但前述描述意图说明且不限制由所附权利要求书的范围定义的本发明的范围。其它方面、优点和修改在所附权利要求书的范围内。

序列表

<110> 双链生物科技公司

迈克尔·施密特

劳伦斯·罗卜

杰西·索尔克

<120> 用于双重测序的改良衔接子、方法和组合物

<130> TWIN-001

<150> US 62/264,822

<151> 2015-12-08

<150> US 62/281,917

<151> 2016-01-22

<160> 41

<170> PatentIn version 3.5

<210> 1

<211> 41

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 1

ccactacgcc tccgctttcc tctctatggg cagtcggtga t 41

<210> 2

<211> 43

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 2

ttggatgatgc ggaggcgaaa ggagagatac ccgtcagcca cta 43

<210> 3

<211> 30

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 3

ccatctcatc cctgcgtgtc tccgactcag 30

<210> 4

<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 4
ttggtagagt agggacgcac agaggctgag tc 32
<210> 5
<211> 39
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<220>
<221> 少见特征
<222> (35) .. (38)
<223> n为a、c、g或t
<400> 5
ccatctcatc cctgcgtgtc tccgactcag gcgcnnnng 39
<210> 6
<211> 41
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<220>
<221> 少见特征
<222> (37) .. (40)
<223> n为a、c、g或t
<400> 6
ttggtagagt agggacgcac agaggctgag tcatatnnnn c 41
<210> 7
<211> 39
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 7
ccatctcatc cctgcgtgtc tccgactcag aaatgcagc 39

<210> 8
<211> 41
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 8
ttggtagagt agggacgcac agaggctgag tcgggccgtc g 41
<210> 9
<211> 39
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 9
ccatctcatc cctgcgtgtc tccgactcag atatgcagc 39
<210> 10
<211> 41
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 10
ttggtagagt agggacgcac agaggctgag tcgcgccgtc g 41
<210> 11
<211> 39
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 11
ccatctcatc cctgcgtgtc tccgactcag tattgcagc 39
<210> 12
<211> 41
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 12

ttggttagagt agggacgcac agaggctgag tcggcccgtc g 41

<210> 13

<211> 39

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 13

ccatctcatc cctgcgtgtc tccgactcag atttgcagc 39

<210> 14

<211> 41

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 14

ttggttagagt agggacgcac agaggctgag tccgggcgtc g 41

<210> 15

<211> 41

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 15

ccactacgcc tccgctttcc tctctatggg cagtcggtga t 41

<210> 16

<211> 43

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 16

ttggtgatgc ggaggcgaaa ggagagatac ccgtcagcca cta 43

<210> 17

<211> 30

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 17
ccatctcatc cctgcgtgtc tccgactcag 30
<210> 18
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 18
ttggtagagt agggacgcac agaggctgag tc 32
<210> 19
<211> 37
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<220>
<221> 少见特征
<222> (31) .. (34)
<223> n为a、c、g或t
<400> 19
ccatctcatc cctgcgtgtc tccgactcag nnnnaac 37
<210> 20
<211> 39
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<220>
<221> 少见特征
<222> (33) .. (36)
<223> n为a、c、g或t
<400> 20
ttggtagagt agggacgcac agaggctgag tennnnttg 39
<210> 21
<211> 46
<212> DNA
<213> 人工序列
<220>

<223> 合成多核苷酸

<220>

<221> 少见特征

<222> (31) .. (34)

<223> n为a、c、g或t

<220>

<221> 少见特征

<222> (38) .. (46)

<223> n为a、c、g或t

<400> 21

ccatctcatc cctgcgtgtc tccgactcag nnnnaacnnn nnnnnn 46

<210> 22

<211> 48

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<220>

<221> 少见特征

<222> (33) .. (36)

<223> n为a、c、g或t

<220>

<221> 少见特征

<222> (40) .. (48)

<223> n为a、c、g或t

<400> 22

ttggtagagt agggacgcac agaggctgag tcnnnnnnttgn nnnnnnnn 48

<210> 23

<211> 46

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<220>

<221> 少见特征

<222> (31) .. (34)

<223> n为a、c、g或t

<220>

<221> 少见特征

<222> (38) .. (46)
<223> n为a、c、g或t
<400> 23
ccatctcatc cctgcgtgtc tccgactcag nnnnaacnnn nnnnnn 46
<210> 24
<211> 48
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<220>
<221> 少见特征
<222> (33) .. (36)
<223> n为a、c、g或t
<220>
<221> 少见特征
<222> (40) .. (48)
<223> n为a、c、g或t
<400> 24
ttggtagagt agggacgcac agaggctgag tcnnnnntggn nnnnnnnn 48
<210> 25
<211> 46
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<220>
<221> 少见特征
<222> (31) .. (34)
<223> n为a、c、g或t
<220>
<221> 少见特征
<222> (38) .. (46)
<223> n为a、c、g或t
<400> 25
ccatctcatc cctgcgtgtc tccgactcag nnnnaacnnn nnnnnn 46
<210> 26
<211> 48
<212> DNA

<213> 人工序列
<220>
<223> 合成多核苷酸
<220>
<221> 少见特征
<222> (33) .. (36)
<223> n为a、c、g或t
<220>
<221> 少见特征
<222> (40) .. (48)
<223> n为a、c、g或t
<400> 26
ttggtagagt agggacgcac agaggctgag tcnnnntggn nnnnnnnn 48
<210> 27
<211> 15
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<220>
<221> 少见特征
<222> (9) .. (12)
<223> n为a、c、g或t
<400> 27
gctgagtcnn nnttg 15
<210> 28
<211> 46
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<220>
<221> 少见特征
<222> (31) .. (34)
<223> n为a、c、g或t
<220>
<221> 少见特征
<222> (38) .. (46)
<223> n为a、c、g或t

<400> 28
ccatctcatc cctgcgtgtc tccgactcag nnnnaacnnn nnnnnn 46
<210> 29
<211> 46
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<220>
<221> 少见特征
<222> (31) .. (34)
<223> n为a、c、g或t
<220>
<221> 少见特征
<222> (38) .. (46)
<223> n为a、c、g或t
<400> 29
ggtagagtag ggacgcacag aggctgagtc nnnntggnnn nnnnnn 46
<210> 30
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<220>
<221> 少见特征
<222> (5) .. (8)
<223> n为a、c、g或t
<220>
<221> 少见特征
<222> (12) .. (20)
<223> n为a、c、g或t
<400> 30
tcagnnnnnaa cnnnnnnnnnn 20
<210> 31
<211> 20
<212> DNA
<213> 人工序列
<220>

<223> 合成多核苷酸

<220>

<221> 少见特征

<222> (5) .. (8)

<223> n为a、c、g或t

<220>

<221> 少见特征

<222> (12) .. (20)

<223> n为a、c、g或t

<400> 31

tcagnnnnnac cnnnnnnnnnn 20

<210> 32

<211> 36

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<220>

<221> 少见特征

<222> (32) .. (35)

<223> n为a、c、g、t或u

<400> 32

ccatctcatc cctgcgtgtc tccgactcag unnnnc 36

<210> 33

<211> 38

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<220>

<221> 少见特征

<222> (34) .. (37)

<223> n为a、c、g或t

<400> 33

ttggtagagt agggacgcac agaggctgag tcannnnng 38

<210> 34

<211> 45

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<220>

<221> 少见特征

<222> (32) .. (35)

<223> n为a、c、g、t或u

<220>

<221> 少见特征

<222> (37) .. (45)

<223> n为a、c、g、t或u

<400> 34

ccatctcatc cctgcgtgtc tccgactcag unnnnnennnn nnnnn 45

<210> 35

<211> 47

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<220>

<221> 少见特征

<222> (34) .. (37)

<223> n为a、c、g或t

<220>

<221> 少见特征

<222> (39) .. (47)

<223> n为a、c、g或t

<400> 35

ttggtagagt agggacgcac agaggctgag tcannnnngnn nnnnnnn 47

<210> 36

<211> 45

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<220>

<221> 少见特征

<222> (32) .. (35)

<223> n为a、c、g、t或u

<220>

<221> 少见特征
<222> (37) .. (45)
<223> n为a、c、g、t或u
<400> 36
ccatctcatc cctgcgtgtc tccgactcag unnnncnnnn nnnnn 45
<210> 37
<211> 45
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<220>
<221> 少见特征
<222> (32) .. (35)
<223> n为a、c、g、t或u
<220>
<221> 少见特征
<222> (37) .. (45)
<223> n为a、c、g、t或u
<400> 37
ggtagagtag ggacgcacag aggctgagtc annnngnnnn nnnnn 45
<210> 38
<211> 44
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<220>
<221> 少见特征
<222> (31) .. (34)
<223> n为a、c、g或t
<220>
<221> 少见特征
<222> (36) .. (44)
<223> n为a、c、g或t
<400> 38
ccatctcatc cctgcgtgtc tccgactcag nnnncnnnnnn nnnn 44
<210> 39
<211> 45

<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<220>
<221> 少见特征
<222> (32) .. (35)
<223> n为a、c、g或t
<220>
<221> 少见特征
<222> (37) .. (45)
<223> n为a、c、g或t
<400> 39
ggtagagtag ggacgcacag aggctgagtc annnngnnnn nnnnn 45
<210> 40
<211> 45
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<220>
<221> 少见特征
<222> (32) .. (35)
<223> n为a、c、g或t
<220>
<221> 少见特征
<222> (37) .. (45)
<223> n为a、c、g或t
<400> 40
ccatctcatc cctgcgtgtc tccgactcag gnnnnnnnnnn nnnnn 45
<210> 41
<211> 45
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<220>
<221> 少见特征
<222> (32) .. (35)

<223> n为a、c、g或t

<220>

<221> 少见特征

<222> (37) .. (45)

<223> n为a、c、g或t

<400> 41

ggtagagtag ggacgcacag aggctgagtc annnngnnnn nnnnn 45

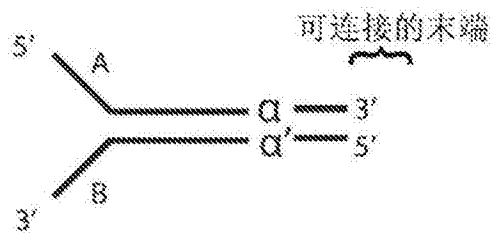


图1A

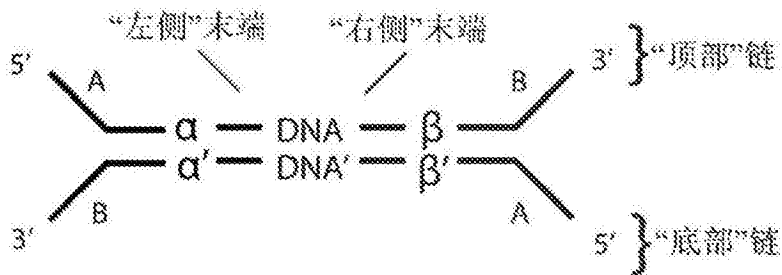


图1B

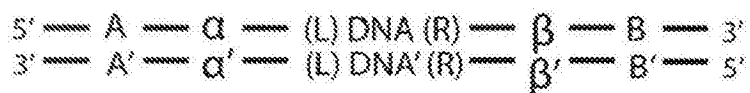


图1C

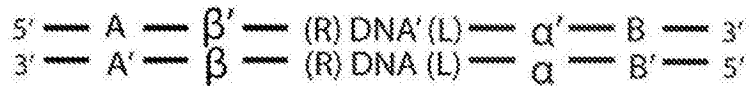


图1D



图1E

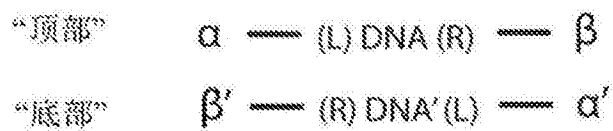


图1F

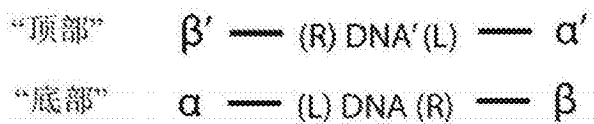


图1G

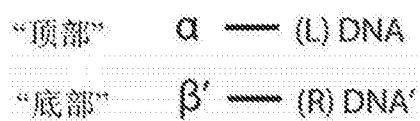


图1H

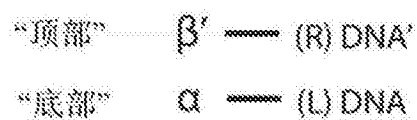


图1I

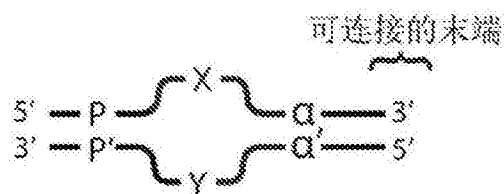


图2A

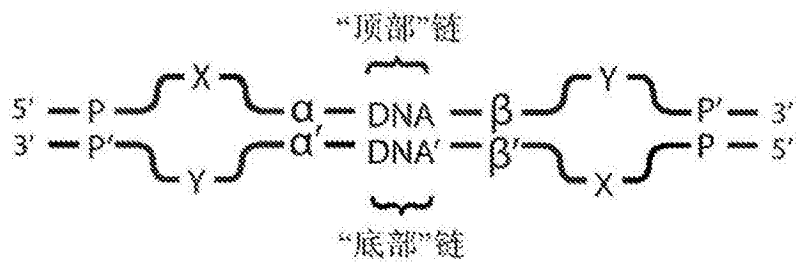


图2B

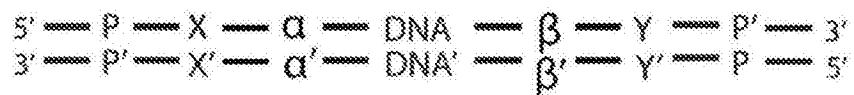


图2C

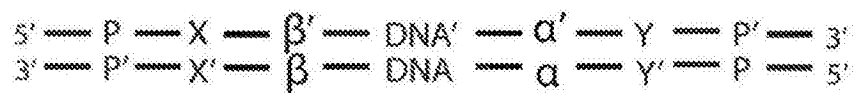


图2D

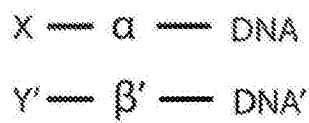


图2E

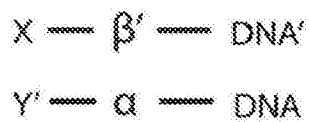


图2F



图2G

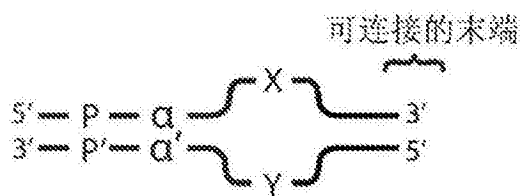


图2H

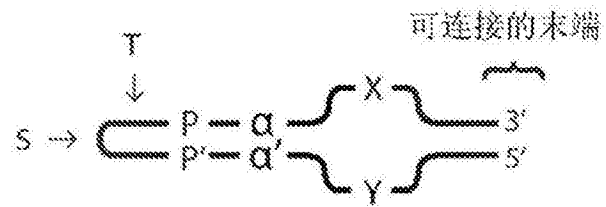


图2I

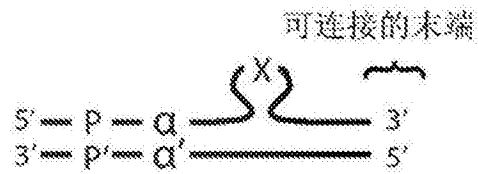


图2J

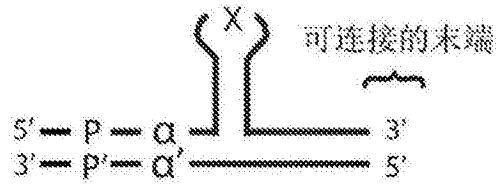


图2K



图3A

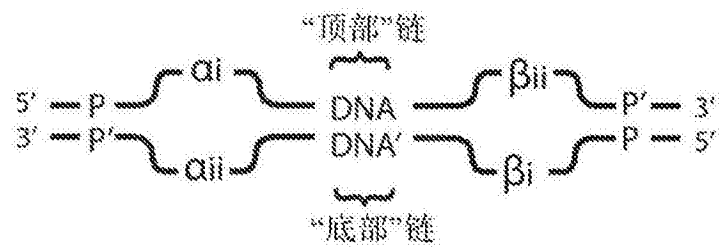


图3B

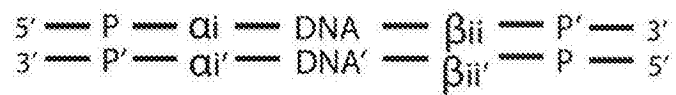


图3C

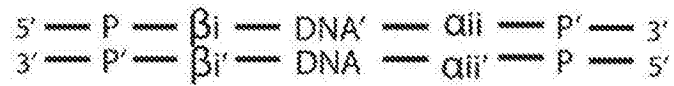


图3D

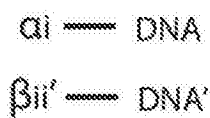


图3E

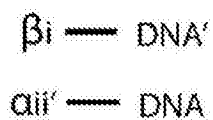


图3F



图3G

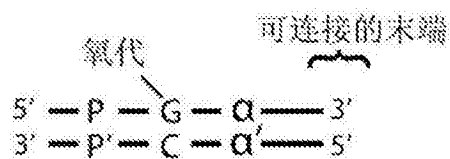


图4A



图4H

图5A

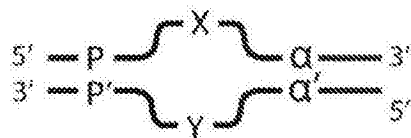


图5B

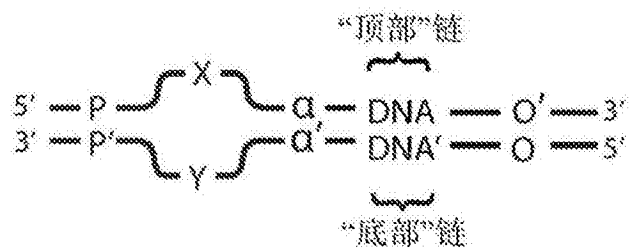
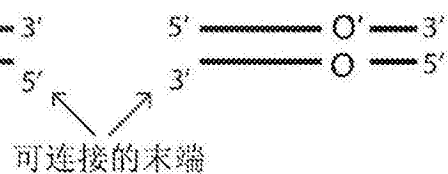


图5C

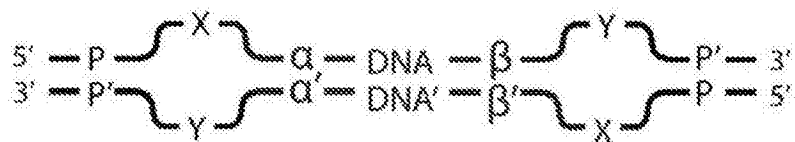


图5D

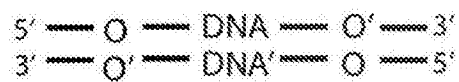


图5E

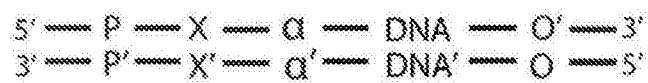


图5F

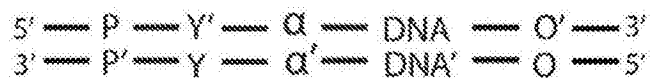


图5G



图5H

图6A

图6B

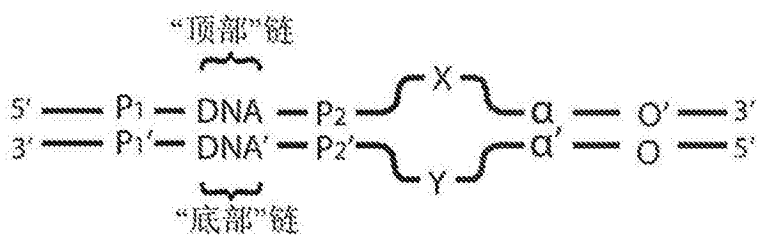
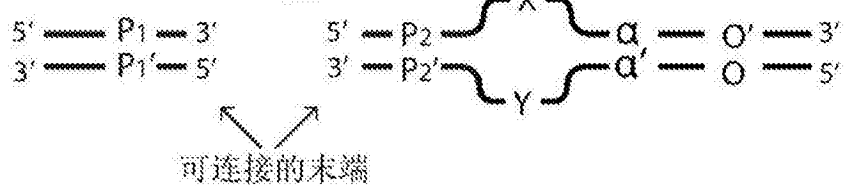


图6C

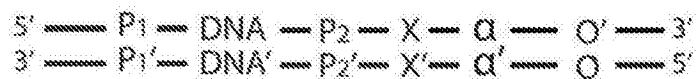


图6D

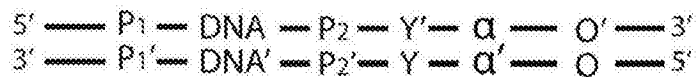


图6E

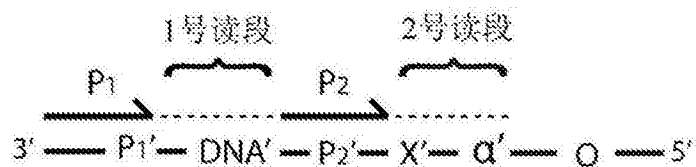


图6F

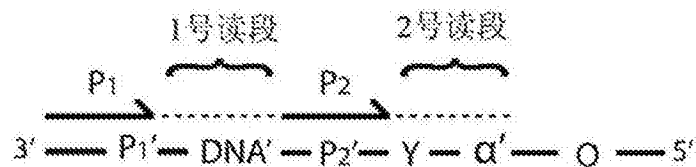


图6G

1号读段 DNA

2号读段 X — α

图6H

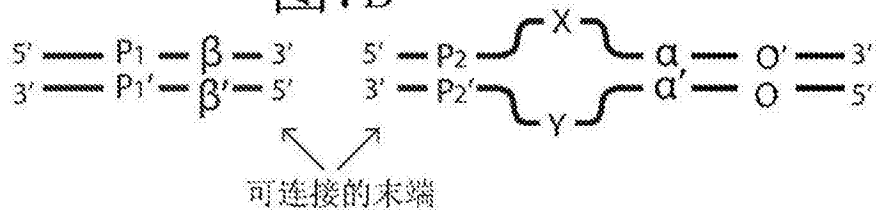
1号读段 DNA

2号读段 Y' — α

图6I

图7A

图7B



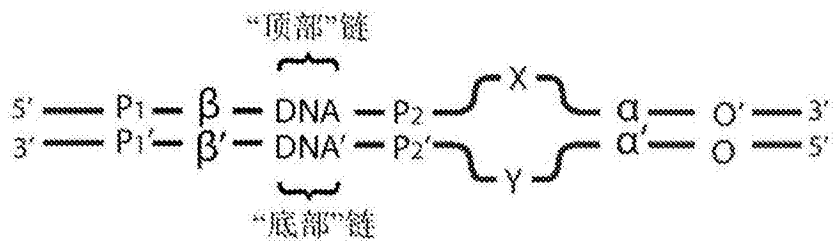


图7C

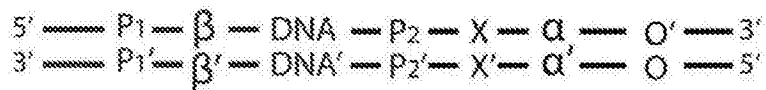


图7D

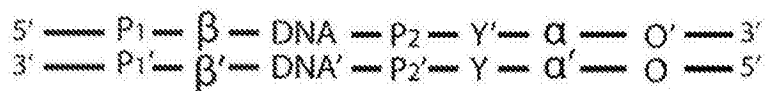


图7E

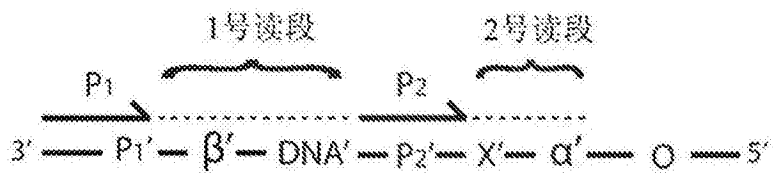


图7F

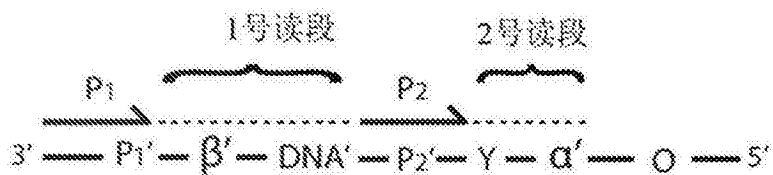


图7G

1号读段 β — DNA

2号读段 X — α

图7H

1号读段 β —DNA

2号读段 γ' — α

图7I

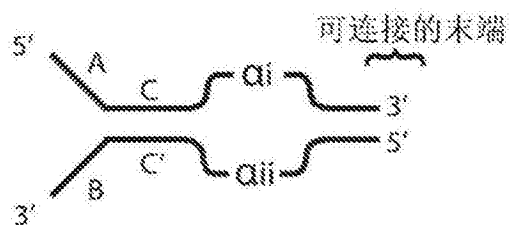


图8A

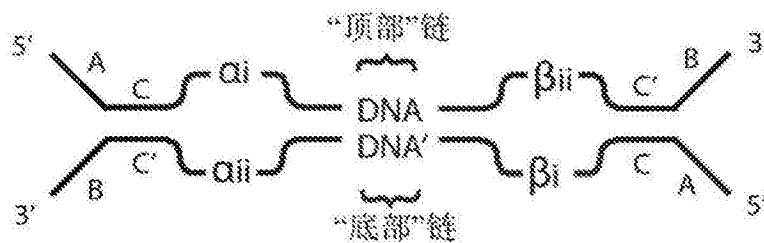


图8B

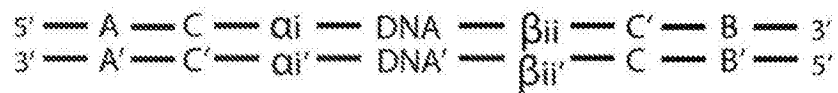


图8C

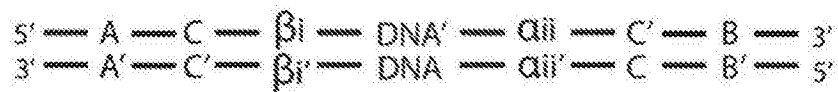


图8D



图8E

β_i — DNA'

 α_{ii}' — DNA

图8F

 α_i — DNA

 β_i — DNA'

 α_{ii}' — DNA

 β_{ii}' — DNA'

图8G



图8H

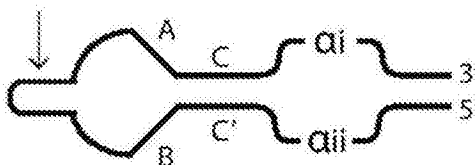


图8I

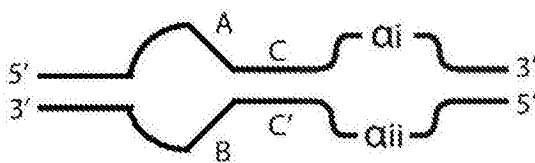


图8J

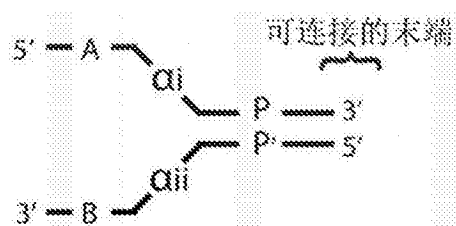


图9A

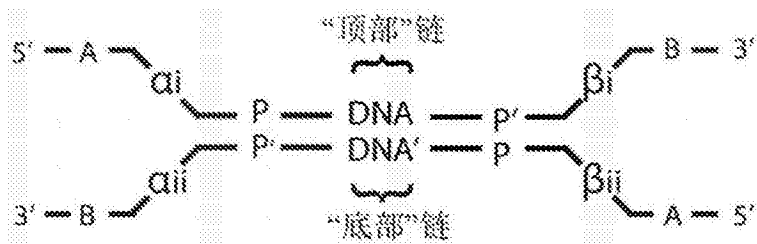


图9B

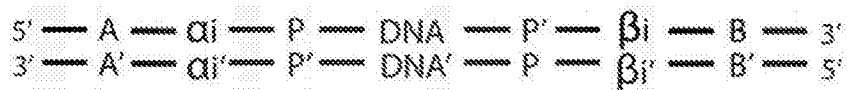


图9C

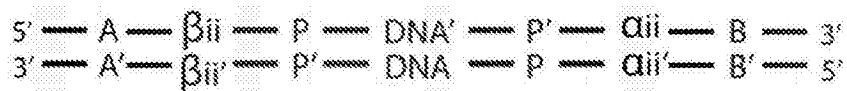


图9D

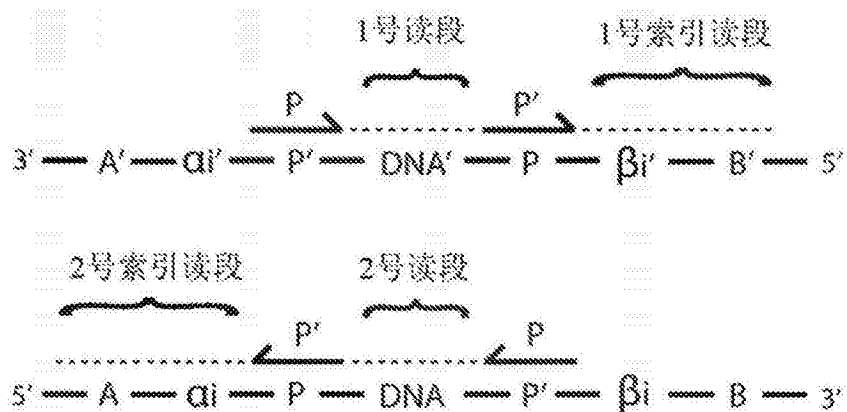


图9E

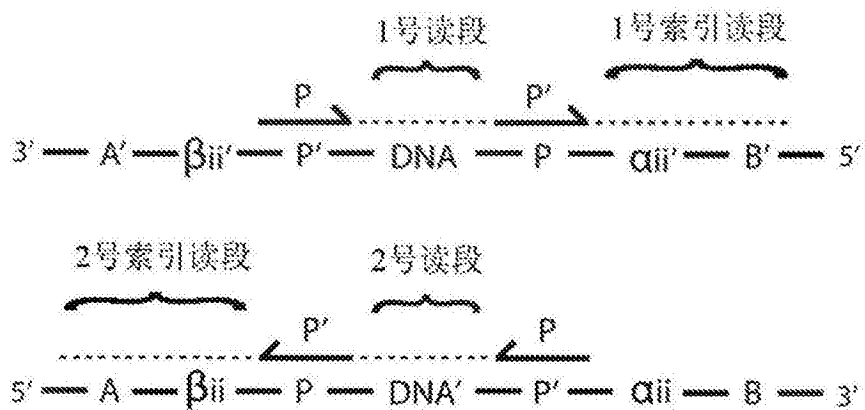


图9F

衍生自“顶部”链		衍生自“底部”链	
1号读段	DNA	1号读段	DNA'
2号读段	DNA'	2号读段	DNA
1号索引	β_i	1号索引	α_{ii}
2号索引	α_i'	2号索引	β_{ii}'

图9G

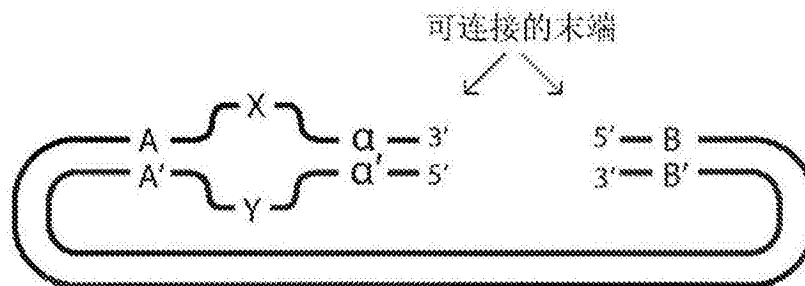


图10A

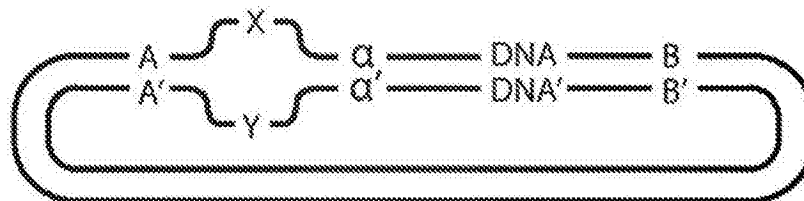


图10B

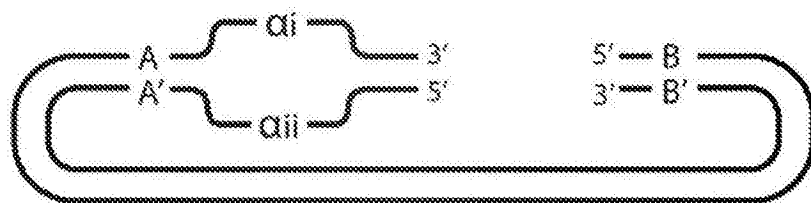


图10C

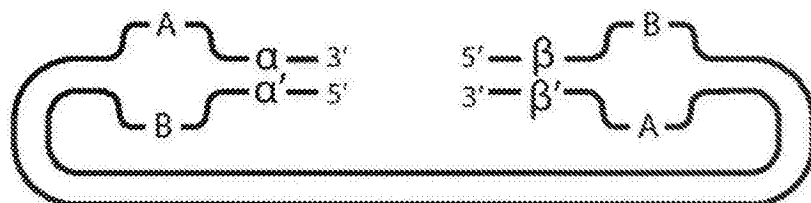


图10D

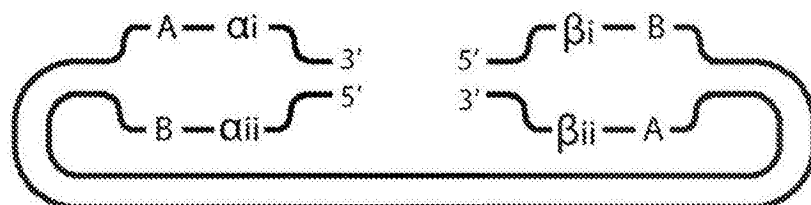


图10E

图11A

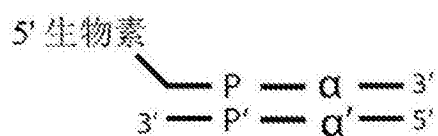
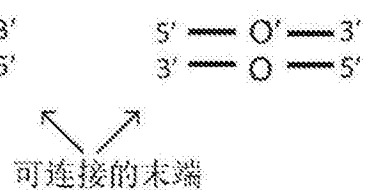


图11B



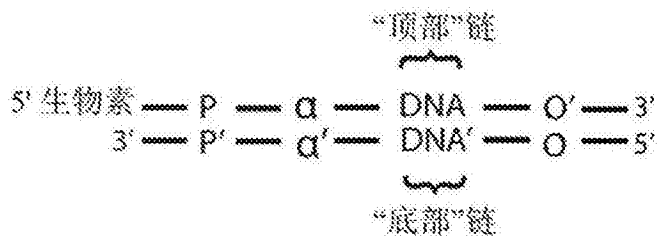


图11C

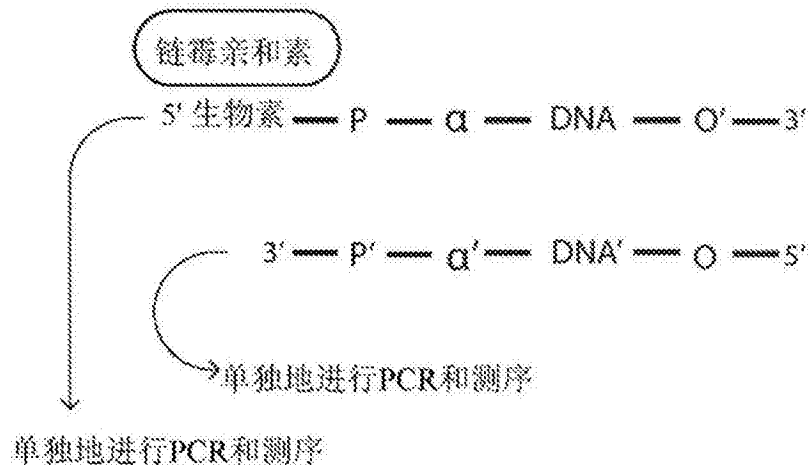


图11D

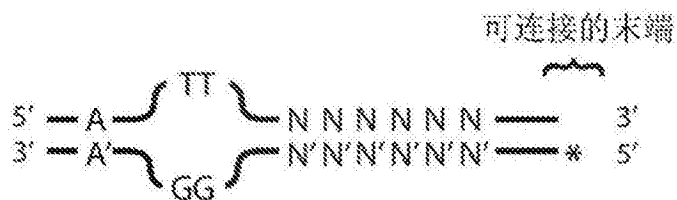


图12A

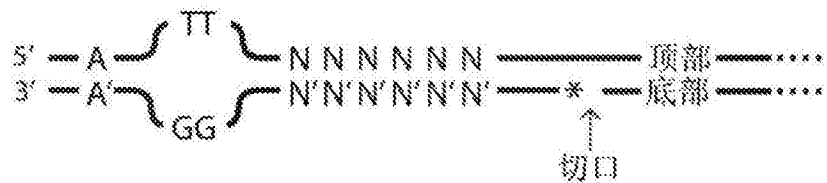


图12B

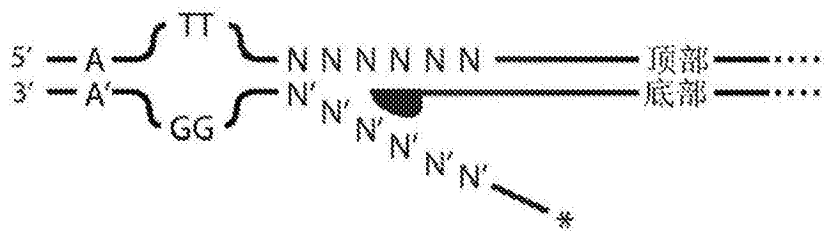


图12C

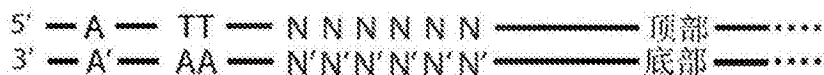


图12D



图12E

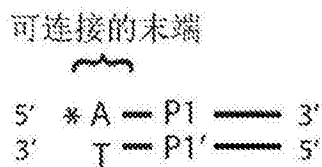


图12F

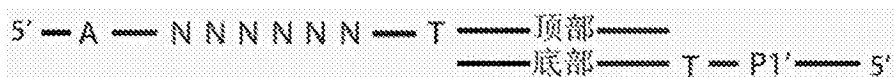


图12G

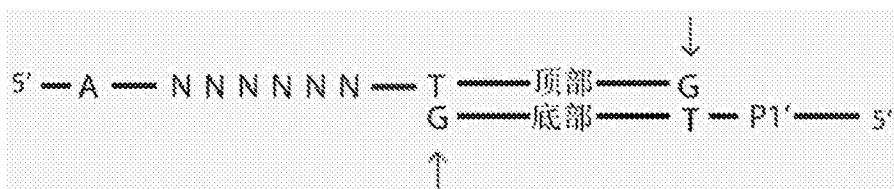


图12H

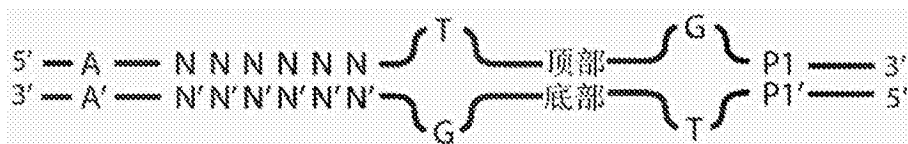


图12I

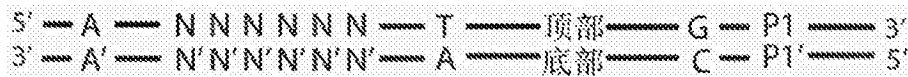


图12J

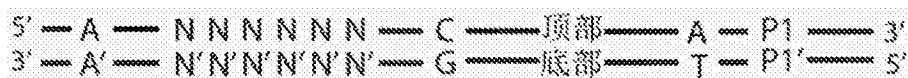


图12K

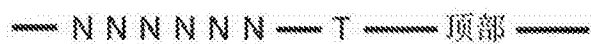


图12L



图12M

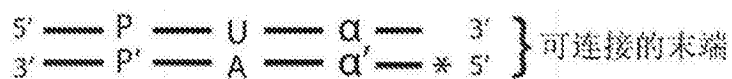


图13A

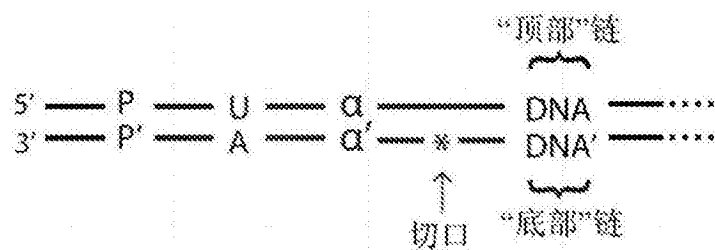


图13B

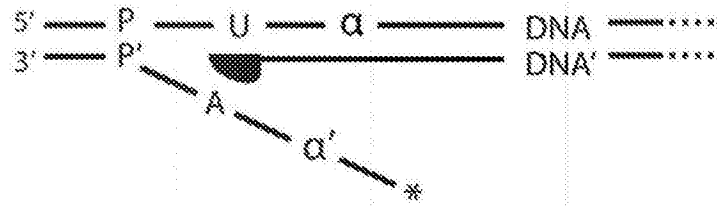


图13C

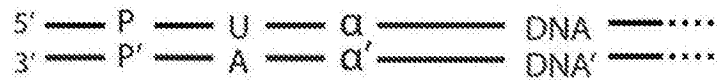


图13D

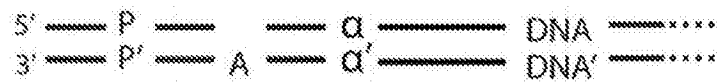


图13E

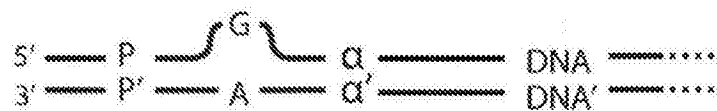


图13F

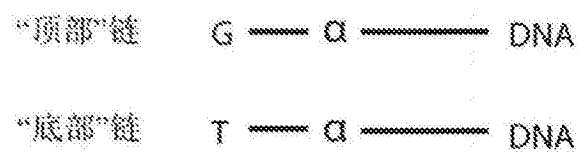


图13G

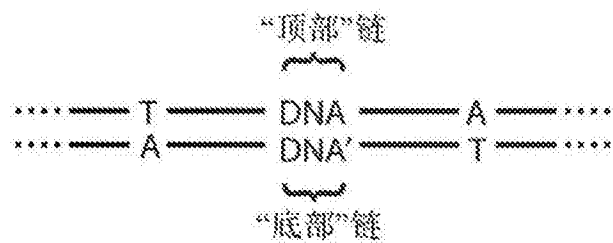


图14A

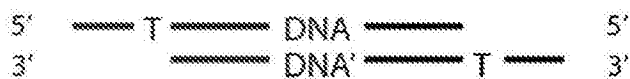


图14B

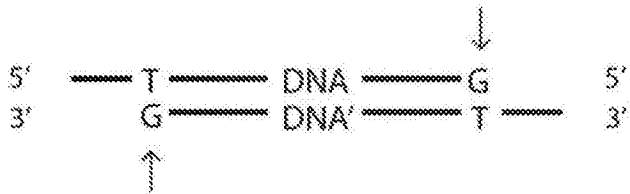


图14C

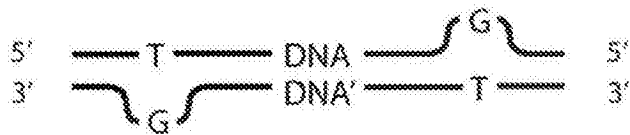


图14D

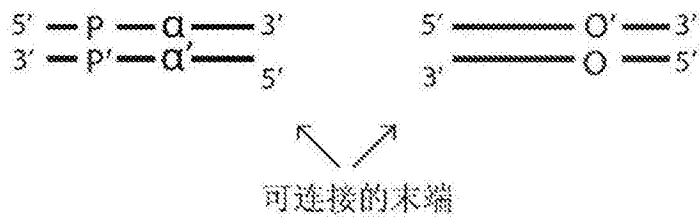


图14E

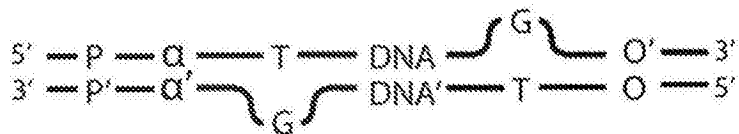


图14F

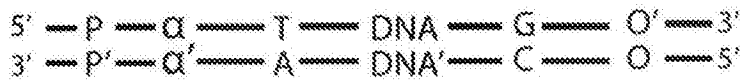


图14G

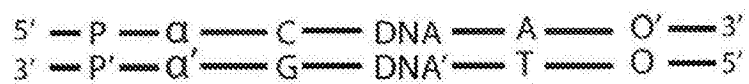


图14H

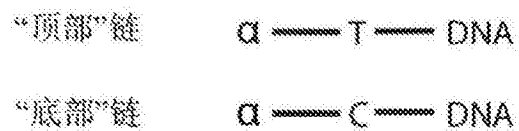


图14I