(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 116472356 A
(43)申请公布日 2023.07.21

(21)申请号 202180072261.2

(22)申请日 2021.08.20

(30)优先权数据
2013141.3 2020.08.21 GB

(85)PCT国际申请进入国家阶段日
2023.04.21

(86)PCT国际申请的申请数据
PCT/EP2021/073203 2021.08.20

(87)PCT国际申请的公布数据
WO2022/038291 EN 2022.02.24

(71)申请人 加的夫大学学院咨询有限公司
地址 英国加的夫

(72)发明人 S·利德 F·多布斯
P·范艾伊克

(74)专利代理机构 中国专利代理(香港)有限公司 72001
专利代理师 段菊兰 彭昶

(51)Int.Cl.
C12Q 1/6869(2006.01)

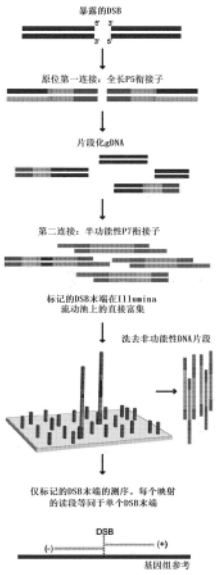
权利要求书9页 说明书40页
序列表24页 附图28页

(54)发明名称

用于分离双链断裂的方法

(57)摘要

本发明涉及用于确定核酸样品中的DNA双链断裂(DSB)的数量和性质的方法,所述核酸样品理想情况下为gDNA;用于实施前述方法的成套试剂盒,其包含至少多个用于连接至所述核酸样品的寡核苷酸,并提供至少一种读段测序引物可与之杂交的至少第一杂交位点(RD1SP或RD2SP);以及用于所述试剂盒和所述方法的寡核苷酸。



CN 116472356 A

1.用于鉴定核酸样品中的DNA双链断裂(DSB)的方法,其包括:

i)在连接条件下将怀疑含有DSB的核酸的样品暴露于第一对寡核苷酸,所述第一对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至所述DSB的第一链的5'结合特征、第一测序引物可与之结合的杂交位点(RD1 SP)和用于从DSB池中分离所述DSB的结合序列;和第二寡核苷酸,其与所述第一对的所述第一寡核苷酸互补并包含使得所述寡核苷酸能够连接至所述DSB的第二链的3'结合特征;其中所述寡核苷酸的一种或两种包含3'和/或5'保护性特征;

ii)将所述样品的所述核酸片段化成片段;

iii)在连接条件下将所述片段暴露于第二对寡核苷酸,所述第二对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至所述片段化核酸的第一链的5'结合特征和第二测序引物可与之结合的杂交位点(RD2 SP);和第二较长的寡核苷酸,其与所述第二对的所述第一寡核苷酸部分互补并包含用于与所述片段化核酸的第二链结合的3'结合特征、与所述杂交位点互补的序列、和另外的序列,其任选为用于实现桥式扩增的结合序列;并且其中所述寡核苷酸的一种或两种包含3'和/或5'保护性特征;

iv)使所述片段变性以提供单链核酸;

v)将部分iv)的链分成两组:A组,已经在第一末端处连接由部分i)的所述寡核苷酸提供的第一杂交位点和结合序列并在另一末端处连接由部分iii)的所述寡核苷酸提供的第二杂交位点和另外的序列的那些片段,以及B组,并未在第一末端处连接由部分i)的所述寡核苷酸提供的杂交位点和结合序列以及并未在另一末端处连接由部分iii)的所述寡核苷酸提供的第二杂交位点和另外的序列的那些片段;和

vi)使用与第一和/或第二杂交位点结合的引物对A组的链进行测序,其中每个序列等同于一个DSB断裂,并且进一步地,其中碱基对缺失的数量和性质可以通过将每个序列与代表从中采集样品的所述物种的基因组进行比较来确定。

2.权利要求1所述的方法,其中包含5'结合特征的所述第二对的所述寡核苷酸不包含用于分离所述片段化核酸的结合序列。

3.权利要求1或权利要求2所述的方法,其中部分i)和部分iii)的所述寡核苷酸互换,由此,在步骤ii)中片段化后,首先将所述核酸暴露于部分iii)的所述寡核苷酸,并随后将所述核酸暴露于部分i)的所述寡核苷酸。

4.权利要求1至3任一项所述的方法,其中所述核酸样品是gDNA。

5.根据前述权利要求任一项所述的方法,其中所述5'和/或3'结合特征包含以下的一种:磷酸基团;三磷酸"T尾",优选脱氧胸苷三磷酸"T尾";三磷酸"A尾",优选脱氧腺苷三磷酸"A尾";至少一个随机N核苷酸和多个N核苷酸。

6.根据前述权利要求任一项所述的方法,其中所述5'和/或3'保护性特征包含提供对以下的任意一个或多个的抗性的特征:磷酸化活性、磷酸酶活性、末端转移酶活性、核酸杂交、核酸内切酶活性、核酸外切酶活性、连接酶活性、聚合酶活性和蛋白结合。

7.根据权利要求6所述的方法,其中所述保护性特征包含硫代磷酸酯连接(phosphorothioate linkage)、双脱氧核苷酸或共价嵌段、亚磷酰胺、C3间隔子亚磷酰胺(3SpC3)。

8.根据前述权利要求任一项所述的方法,其中部分i)的所述第一寡核苷酸的所述5'结

合特征是磷酸基团,并且部分i)的所述第二寡核苷酸的所述3'结合特征是三磷酸尾。

9.根据前述权利要求任一项所述的方法,其中部分i)的所述第一和第二寡核苷酸还包含索引特征,所述索引特征是使得能够确定汇集样品(pooled sample)的来源的特定核苷酸序列。

10.根据前述权利要求任一项所述的方法,其中部分i)的所述第一寡核苷酸,从5'至3'读取,包含5'结合特征和随后任选的保护性特征、测序引物可与之结合的杂交位点(RD1 SP)、索引序列、用于从DSB池中分离所述DSB的结合序列、和3'结合和/或保护性特征。

11.根据权利要求10所述的方法,其中所述结合特征是磷酸基团。

12.根据前述权利要求任一项所述的方法,其中部分i)的所述第二寡核苷酸,从3'至5'读取,包含3'结合特征和随后任选的保护性特征、测序引物可与之结合的杂交序列(RD1 SP)、索引序列、用于从DSB池中分离所述DSB的结合序列、和5'结合和/或保护性特征。

13.根据权利要求12所述的方法,其中所述3'结合特征包含3'脱氧胸苷三磷酸"T尾"以及硫代磷酸酯连接。

14.根据前述权利要求任一项所述的方法,其中部分i)的所述第一寡核苷酸对的所述第一和/或第二寡核苷酸包含两种不同的末端保护性特征。

15.根据前述权利要求任一项所述的方法,其中部分iii)的所述第二寡核苷酸对的所述第一寡核苷酸的所述5'结合特征是磷酸基团,并且部分iii)的所述第二寡核苷酸的所述3'结合特征是三磷酸尾。

16.根据前述权利要求任一项所述的方法,其中部分iii)的所述第一和第二寡核苷酸还包含索引特征,所述索引特征是使得能够确定汇集样品的来源的特定核苷酸序列。

17.根据前述权利要求任一项所述的方法,其中部分iii)的所述第二寡核苷酸,从5'至3'读取,包含5'结合特征和随后任选的保护性特征、任选用于实现桥式扩增的另外的序列、索引序列、测序引物可与之结合的杂交位点(RD2 SP)、和3'结合和/或保护性特征。

18.根据前述权利要求任一项所述的方法,其中部分iii)的该第二寡核苷酸对的所述第一或第二寡核苷酸包含两种不同的末端保护性特征。

19.根据前述权利要求任一项所述的方法,其中部分i)的所述寡核苷酸包含具有SEQ ID NO.1的第一寡核苷酸和具有SEQ ID NO.2的第二寡核苷酸;或与SEQ ID NO.1或2共享至少80%同一性或同源性的寡核苷酸。

20.根据前述权利要求任一项所述的方法,其中部分iii)的所述第二对寡核苷酸包含具有SEQ ID NO.3的第一寡核苷酸和具有SEQ ID NO.4的第二寡核苷酸;或与SEQ ID NO.3或4共享至少80%同一性或同源性的寡核苷酸。

21.根据前述权利要求任一项所述的方法,其中部分iii)的所述第二对寡核苷酸的所述第二寡核苷酸包含以下序列中的任一种:SEQ ID NO.4-28;或与SEQ ID NO.4-28中的一种共享至少80%同一性或同源性的寡核苷酸。

22.根据前述权利要求任一项所述的方法,其中所述样品是哺乳动物或人。

23.根据前述权利要求任一项所述的方法,其中部分i)中的所述连接原位发生或使用细胞或组织样品体外发生。

24.根据前述权利要求任一项所述的方法,其中在进行步骤i)之前,所述样品暴露于透化剂。

25.根据前述权利要求任一项所述的方法,其中在进行步骤i)之前,所述样品暴露于至少一种用于进行精氨酸尾修复的试剂。

26.根据前述权利要求任一项所述的方法,其中部分i)还包括在进行后续步骤之前从所述样品中提取gDNA。

27.根据前述权利要求任一项所述的方法,其中所述方法进一步包括在部分ii)和/或部分iv)之后,除去其尺寸小于大约100bp、或小于大约150bp的片段,和/或保留其尺寸大于大约150bp的片段。

28.根据前述权利要求任一项所述的方法,其中部分v)的所述分离涉及使用由部分i)的所述寡核苷酸提供的所述结合序列来结合配偶体,并因此将部分iv)的A组链与任何其它链分离。

29.权利要求28所述的方法,其中由部分i)的所述寡核苷酸提供的所述结合序列的互补结合链锚定在基底上,并且所述核酸单链流经或流过锚定的互补结合链。

30.根据前述权利要求任一项所述的方法,其中部分vi)涉及桥式扩增,其中在部分v)下分离的单链在已在部分i)的所述第一寡核苷酸的所述结合序列和部分iii)的所述第二寡核苷酸的所述另外的序列的寡核苷酸/结合位点锚定于其上的基底上进行克隆扩增。

31.根据前述权利要求任一项所述的方法,其中在进行要求保护的方法之前,将含有或怀疑含有单链断裂的样品连接或断裂以确保所述单链断裂转化为双链断裂。

32.用于鉴定gDNA样品中的DNA双链断裂(DSB)的成套试剂盒,其包含:

i)第一对寡核苷酸,所述第一对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至所述DSB的第一链的5'结合特征、第一测序引物可与之结合的杂交位点(RD1 SP)和用于从DSB池中分离所述DSB的结合序列;和第二寡核苷酸,其与该第一对的所述第一寡核苷酸互补,并包含用于与所述DSB的第二链结合的3'结合特征;并且其中所述寡核苷酸的一种或两种包含3'和/或5'保护性特征;和

ii)第二对寡核苷酸,所述第二对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至所述DSB的第一链的5'结合特征,和第二测序引物可与之结合的杂交位点(RD2 SP);和第二较长的寡核苷酸,其与该第二对的所述第一寡核苷酸部分互补,并包含用于与所述DSB的第二链结合的3'结合特征、与所述杂交位点互补的序列、和另外的序列,其任选为用于实现桥式扩增的结合序列;并且其中所述寡核苷酸的一种或两种包含3'和/或5'保护性特征。

33.用于鉴定gDNA样品中的DSB的样品制备的试剂盒,其包含:

i)第一对寡核苷酸,所述第一对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至双链核酸的一条链,并包含根据TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的序列的5'结合特征;和第二寡核苷酸,其与所述第一对的所述第一寡核苷酸互补;并且其中所述寡核苷酸的一种或两种分别包含3'和/或5'保护性特征;和

ii)第二对寡核苷酸,所述第二对寡核苷酸的第一寡核苷酸不包含序列ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的超过5、10、15或20个碱基的序列,或不包含序列ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的全部24个碱基;和第二寡核苷酸,其包含使得所述寡核苷酸能够连接至双链核酸的一条链,并包含根据CAAGCAGAAGACGGCATACGAGAT(SEQ ID NO:32)的序列的3'结合特征;并且其中所述寡核苷酸

的一种或两种分别包含3’和/或5’保护性特征。

34.用于鉴定gDNA样品中的DSB的样品制备的试剂盒，其包含：

i)第一对寡核苷酸，所述第一对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至双链核酸的一条链，并包含根据ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的序列的5’结合特征；和第二寡核苷酸，其与所述第一对的所述第一寡核苷酸互补；并且其中所述寡核苷酸的一种或两种分别包含3’和/或5’保护性特征；和

ii)第二对寡核苷酸，所述第二对寡核苷酸的第一寡核苷酸不包含序列TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的超过5、10或15个碱基的序列，或不包含序列TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的全部20个碱基；和第二寡核苷酸，其包含使得所述寡核苷酸能够连接至双链核酸的一条链，并包含根据AATGATACGGCGACCACCGA(SEQ ID NO:34)的序列的3’结合特征；并且其中所述寡核苷酸的一种或两种分别包含3’和/或5’保护性特征。

35.根据权利要求33或34所述的试剂盒，其中所述第一对和/或所述第二对寡核苷酸的所述第一寡核苷酸包含第一测序引物可与之结合的杂交位点。

36.根据权利要求32或35任一项所述的试剂盒，其中部分i)和/或部分ii)的所述第一和第二寡核苷酸还包含索引特征，所述索引特征是使得能够确定汇集样品的来源的特定核苷酸序列。

37.根据权利要求32、35或36任一项所述的试剂盒，其中所述试剂盒进一步包含至少一种引物，其与所述第一和/或第二杂交位点结合以便用于测序目的。

38.根据权利要求32至37所述的试剂盒，其中所述试剂盒进一步包含分别用于将所述核酸片段化和/或变性成片段和/或单链的片段化剂和/或变性剂。

39.用于鉴定核酸样品，诸如gDNA样品中的DNA双链断裂(DSB)的双链衔接子，其包含：

第一寡核苷酸链，其包含使得所述寡核苷酸能够连接至所述DSB的第一链的5’结合特征、测序引物可与之结合的杂交位点(RD1 SP)和用于从DSB池中分离所述DSB的结合序列；和第二寡核苷酸链，其与所述第一寡核苷酸互补并包含用于与所述DSB的第二链结合的3’结合特征；并且其中所述寡核苷酸的一种或两种包含3’和/或5’保护性特征。

40.用于鉴定核酸样品，诸如gDNA样品中的DNA双链断裂(DSB)的双链衔接子，其包含：

第一寡核苷酸链，其包含使得所述寡核苷酸能够连接至所述DSB的第一链的5’结合特征和测序引物可与之结合的杂交位点(RD2 SP)；和第二较长的寡核苷酸链，其与所述第一寡核苷酸部分互补，并包含用于与所述DSB的第二链结合的3’结合特征、与所述杂交位点互补的序列、和另外的序列，其任选为用于实现桥式扩增的结合序列；并且其中所述寡核苷酸的一种或两种包含3’和/或5’保护性特征。

41.权利要求39或40所述的衔接子，其中所述第一和第二寡核苷酸还包含索引特征，所述索引特征是使得能够确定汇集样品的来源的特定核苷酸序列。

42.用于鉴定核酸样品中的DNA双链断裂(DSB)的双链衔接子，其包含：

第一寡核苷酸链，其不包含序列ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的超过5、10、15或20个碱基的序列，或不包含序列ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的全部24个碱基；和第二寡核苷酸，其包含使得所述寡核苷酸能够连接至双链核酸的一条链并包含根据CAAGCAGAAGACGGCATACGAGAT(SEQ ID NO:32)的序列的3’结合特征；并且其中所述

寡核苷酸的一种或两种分别包含3’和/或5’保护性特征。

43.用于鉴定核酸样品中的DNA双链断裂(DSB)的双链衔接子,其包含:

第一寡核苷酸链,其不包含序列TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的超过5、10或15个碱基的序列,或不包含序列TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的全部20个碱基;和第二寡核苷酸,其包含使得所述寡核苷酸能够连接至双链核酸的一条链,并包含根据AATGATACGGCGACCACCGA(SEQ ID NO:34)的序列的3’结合特征;并且其中所述寡核苷酸的一种或两种分别包含3’和/或5’保护性特征。

44.用于鉴定核酸样品中的DNA双链断裂(DSB)的样品制备方法,其中所述制备包括修饰DSB相关核酸以适于与包含固定引物的基底结合,所述方法包括:

a)提供包含多个核酸的样品;

b)在有助于连接的条件下将所述多个核酸暴露于第一衔接子,其中所述第一衔接子包含能够连接至DSB的一条链的3’末端的寡核苷酸且所述寡核苷酸包含能够通过杂交与固定于所述基底的引物结合的序列;

c)片段化所述多个核酸;和

d)在有助于连接的条件下将所述多个核酸暴露于第二衔接子,其中所述第二衔接子包含能够在片段化诱导的断裂处连接至一条链的5’末端但不能连接至所述第一衔接子的寡核苷酸,且所述寡核苷酸不包含能够通过杂交与固定于所述基底的引物结合的序列。

45.权利要求44所述的方法,其中步骤d)的所述寡核苷酸包含与第二引物的区域相同的序列。

46.权利要求44或45所述的方法,其中

所述基底包含第一和第二固定引物;

步骤b)的所述寡核苷酸包含能够通过杂交与所述第一固定引物结合的序列;并且

步骤d)的所述寡核苷酸包含与所述第二固定引物的区域相同的序列。

47.权利要求44至46任一项所述的方法,其中

步骤b)是:

在有助于连接的条件下将所述多个核酸暴露于第一衔接子对,其中所述第一衔接子对能够连接至DSB的一条链的至少3’末端,并且其中所述第一衔接子对包含至少部分互补的第一和第二寡核苷酸,并且所述第一寡核苷酸可连接至3’末端并包含能够通过杂交与固定于所述基底的引物结合的序列;以及

其中步骤d)是:

在有助于连接的条件下将所述多个核酸暴露于第二衔接子对,其中所述第二衔接子对能够在片段化诱导的断裂处连接至一条链的至少5’末端,但不能连接至所述第一衔接子对的所述第一寡核苷酸,其中所述第二衔接子包含第一和第二部分互补的寡核苷酸,并且所述第一寡核苷酸可连接至5’末端并包含与第二引物的区域相同的序列,并且所述第二寡核苷酸不包含与所述与所述第二引物的区域相同的序列互补的序列。

48.权利要求44至47任一项所述的方法,其中所述第二衔接子对包含:

含有根据CAAGCAGAAGACGGCATACGAGAT(SEQ ID NO:32)的序列的第一寡核苷酸,和含有不包含序列ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的超过5、10、15或20个碱基的序列,或不包含序列ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的全部24个碱基的序列的第

二寡核苷酸;或

含有根据AATGATACGGCGACCACCGA(SEQ ID NO:34)的序列的第一寡核苷酸,和含有不包含序列TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的超过5、10或15个碱基的序列,或不包含序列TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的全部20个碱基的序列的第二寡核苷酸。

49.权利要求47或权利要求48所述的方法,其中所述第一衔接子对的所述第一和/或第二寡核苷酸包含3'和/或5'保护性特征;和/或其中所述第二衔接子对的所述第一和/或第二寡核苷酸包含3'和/或5'保护性特征。

50.权利要求44至49任一项所述的方法,其中由于所述第一衔接子的3'修饰的存在,所述第二衔接子不能连接至所述第一衔接子。

51.权利要求44至50任一项所述的方法,其中可连接至5'末端的所述第二衔接子的所述寡核苷酸包含与固定引物的5、10、15、20、21、24或更多个碱基相同的序列。

52.权利要求44至51任一项所述的方法,其进一步包括使所述多个核酸变性以形成多个单链核酸。

53.权利要求44至52任一项所述的方法,其进一步包括在适于所述固定引物与互补核酸杂交的条件下使所述多个核酸与所述包含固定引物的基底接触。

54.权利要求53所述的方法,其进一步包括获得与所述基底杂交的任何核酸的序列信息。

55.权利要求44至54任一项所述的方法,其中所述包含多个核酸的样品是gDNA。

56.权利要求44至55任一项所述的方法,其中:

所述步骤以a)、b)、c)和随后d)的顺序进行;或

所述步骤以a)、c)、d)和随后b)的顺序进行;其中在步骤d)和b)之间将所述样品暴露于能够引起或怀疑能够引起DSB的条件。

57.权利要求56所述的方法,其中将所述样品暴露于能够在所述核酸样品中的目标特征处引起DSB的条件。

58.用于鉴定核酸样品中的DNA双链断裂(DSB)的样品制备方法,其中所述制备包括修饰DSB相关核酸以适于与包含固定的第一引物的基底结合,所述方法包括:

1)提供包含多个核酸的样品;

2)在有助于连接的条件下将所述多个核酸暴露于第一衔接子,其中所述第一衔接子包含能够连接至DSB的一条链的3'末端的寡核苷酸且所述寡核苷酸包含能够与第二引物杂交的序列;

3)片段化所述多个核酸;

4)在有助于连接的条件下将所述多个核酸暴露于第二衔接子,其中所述第二衔接子包含能够在片段化诱导的断裂处连接至一条链的5'末端但不能连接至所述第一衔接子的寡核苷酸,且所述寡核苷酸包含与所述固定的第一引物的区域相同的序列;和

5)在适于所述引物延伸的条件下使所述多个核酸与所述第二引物接触。

59.权利要求58所述的方法,

其中步骤2)是:

在有助于连接的条件下将所述多个核酸暴露于第一衔接子对,其中所述第一衔接子对能够连接至DSB的一条链的至少3'末端,并且其中所述第一衔接子对包含至少部分互补的

第一和第二寡核苷酸,并且所述第一寡核苷酸可连接至3'末端并包含能够与第二引物杂交的序列;以及

其中步骤4)是:

在有助于连接的条件下将所述多个核酸暴露于第二衔接子对,其中所述第二衔接子对能够在片段化诱导的断裂处连接至一条链的至少5'末端但不能连接至所述第一衔接子对的所述第一寡核苷酸,其中所述第二衔接子包含第一和第二部分互补的寡核苷酸,并且所述第一寡核苷酸可连接至5'末端并包含与所述固定的第一引物的区域相同的序列,并且所述第二寡核苷酸不包含与所述与所述固定的第一引物的区域相同的序列互补的序列。

60.权利要求58或权利要求59所述的方法,其中所述第二衔接子对包含:

含有根据AACCCACTACGCCTCCGCTTTCC(SEQ ID NO:40)的序列的第一寡核苷酸;和

不包含序列GGAAAGCGGAGGCGTAGTGGTT(SEQ ID NO:36)的超过5、10、15、20个碱基的序列,或不包含序列GGAAAGCGGAGGCGTAGTGGTT(SEQ ID NO:36)的全部22个碱基的第二寡核苷酸。

61.权利要求59或权利要求60所述的方法,其中所述第一衔接子对的所述第一和/或第二寡核苷酸包含3'和/或5'保护性特征;和/或其中所述第二衔接子对的所述第一和/或第二寡核苷酸包含3'和/或5'保护性特征。

62.权利要求58至61任一项所述的方法,其中由于所述第一衔接子的3'修饰的存在,所述第二衔接子不能连接至所述第一衔接子。

63.权利要求58至62任一项所述的方法,其中可连接至5'末端的所述第二衔接子的所述寡核苷酸包含与固定引物的5、10、15、20、21、24或更多个碱基相同的序列。

64.权利要求58至63任一项所述的方法,其进一步包括使所述多个核酸变性以形成多个单链核酸。

65.权利要求58至64任一项所述的方法,其进一步包括在适于所述固定的第一引物与互补核酸杂交的条件下使所述多个核酸与所述包含所述固定的第一引物的基底接触。

66.权利要求65所述的方法,其进一步包括获得与所述基底杂交的任何核酸的序列信息。

67.权利要求58至66任一项所述的方法,其中所述包含多个核酸的样品是gDNA。

68.权利要求58至67任一项所述的方法,其中:

所述步骤以1)、2)、3)、4)和随后5)的顺序进行;或

所述步骤以1)、3)、4)、2)和随后5)的顺序进行;其中在步骤4)和2)之间将所述样品暴露于能够引起或怀疑能够引起DSB的条件。

69.权利要求68所述的方法,其中将所述样品暴露于能够在所述核酸样品中的目标特征处引起DSB的条件。

70.用于鉴定核酸样品中的目标特征的样品制备方法,其中所述制备包括修饰与目标特征相关的核酸以适于与包含固定引物的基底结合,所述方法包括:

a)提供包含多个核酸的样品,将所述多个核酸暴露于能够在目标特征处切割核酸的至少一条链的条件,并使所述多个核酸变性成单链核酸;

b)在有助于连接的条件下将所述多个核酸暴露于第一衔接子,其中所述第一衔接子包含能够连接至切割位点的一条链的3'末端的寡核苷酸且所述寡核苷酸包含能够通过杂交

与固定于所述基底的引物结合的序列；

c)片段化所述多个核酸；和

d)在有助于连接的条件下将所述多个核酸暴露于第二衔接子，其中所述第二衔接子包含能够在片段化诱导的断裂处连接至一条链的5’末端但不能连接至所述第一衔接子的寡核苷酸，且所述寡核苷酸不包含能够通过杂交与固定于所述基底的引物结合的序列。

71.权利要求70所述的方法，其中所述目标特征是能够被特异性切割的任何特征。

72.权利要求70或权利要求71所述的方法，其中所述目标特征是环丁烷嘧啶二聚体(CPD)、8-氧代鸟嘌呤或脱碱基位点。

73.权利要求70至72任一项所述的方法，其中步骤d)的所述寡核苷酸包含与第二引物的区域相同的序列。

74.权利要求70至73任一项所述的方法，其中

所述基底包含第一和第二固定引物；

步骤b)的所述寡核苷酸包含能够通过杂交与所述第一固定引物结合的序列；以及

步骤d)的所述寡核苷酸包含与所述第二固定引物的区域相同的序列。

75.权利要求70至74任一项所述的方法，其中

步骤b)是：

在有助于连接的条件下将所述多个核酸暴露于第一衔接子对，其中所述第一衔接子对能够连接至切割位点的一条链的至少3’末端，并且其中所述第一衔接子对包含至少部分互补的第一和第二寡核苷酸，并且所述第一寡核苷酸可连接至3’末端并包含能够通过杂交与固定于所述基底的引物结合的序列；以及

其中步骤d)是：

在有助于连接的条件下将所述多个核酸暴露于第二衔接子对，其中所述第二衔接子对能够在片段化诱导的断裂处连接至一条链的至少5’末端但不能连接至所述第一衔接子对的所述第一寡核苷酸，其中所述第二衔接子包含第一和第二部分互补的寡核苷酸，并且所述第一寡核苷酸可连接至5’末端并包含与第二引物的区域相同的序列，并且所述第二寡核苷酸不包含与所述与所述第二引物的区域相同的序列互补的序列。

76.权利要求70至75任一项所述的方法，其中所述第二衔接子对包含：

含有根据CAAGCAGAAGACGGCATACGAGAT(SEQ ID NO:32)的序列的第一寡核苷酸，和含有不包含序列ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的超过5、10、15或20个碱基的序列，或不包含序列ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的全部24个碱基的序列的第二寡核苷酸；或

含有根据AATGATACGGCGACCACCGA(SEQ ID NO:34)的序列的第一寡核苷酸，和含有不包含序列TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的超过5、10或15个碱基的序列，或不包含序列TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的全部20个碱基的序列的第二寡核苷酸。

77.权利要求75或权利要求76所述的方法，其中所述第一衔接子对的所述第一和/或第二寡核苷酸包含3’和/或5’保护性特征；和/或其中所述第二衔接子对的所述第一和/或第二寡核苷酸包含3’和/或5’保护性特征。

78.权利要求70至77任一项所述的方法，其中由于所述第一衔接子的3’修饰的存在，所述第二衔接子不能连接至所述第一衔接子。

79.权利要求70至78任一项所述的方法，其中可连接至5’末端的所述第二衔接子的所述寡核苷酸包含与固定引物的5、10、15、20、21、24或更多个碱基相同的序列。

80.权利要求70至79任一项所述的方法，其进一步包括使所述多个核酸变性以形成多个单链核酸。

81.权利要求70至80任一项所述的方法，其进一步包括在适于所述固定引物与互补核酸杂交的条件下使所述多个核酸与所述包含固定引物的基底接触。

82.权利要求81所述的方法，其进一步包括获得与所述基底杂交的任何核酸的序列信息。

83.权利要求70至82任一项所述的方法，其中所述包含多个核酸的样品是gDNA。

84.权利要求70至83任一项所述的方法，其中：

所述步骤以a)、b)、c)和随后d)的顺序进行；或

所述步骤以c)、d)、a)和随后b)的顺序进行。

85.用于鉴定核酸样品，诸如gDNA样品中的DNA双链断裂(DSB)的双链衔接子，其包含：

第一寡核苷酸链，其不包含序列GGAAAGCGGAGGCGTAGTGGTT(SEQ ID NO:36)的超过5、10、15、20个碱基的序列，或不包含序列GGAAAGCGGAGGCGTAGTGGTT(SEQ ID NO:36)的全部22个碱基；和第二寡核苷酸，其包含使得所述寡核苷酸能够连接至双链核酸的一条链并包含根据AACCCACTACGCCTCCGCTTTCC(SEQ ID NO:40)的序列的3’结合特征；并且其中所述寡核苷酸的一种或两种分别包含3’和/或5’保护性特征。

用于分离双链断裂的方法

技术领域

[0002] 本发明涉及用于确定核酸样品中的DNA双链断裂(DSB)的数量和性质的方法,所述核酸样品理想情况下为基因组DNA (gDNA);用于实施前述方法的成套试剂盒,其包含至少多个用于连接至所述核酸样品的寡核苷酸;以及用于所述试剂盒和所述方法的寡核苷酸。

[0003] 发明背景

核酸中的断裂(其中一条或多条核酸主链被切断)对细胞和生物体特别有害,因为它们可以导致基因组重排。此类断裂包括通常在DNA中的双链断裂(DSB),这是所有DNA损伤中最危险的,因为如果不修复的话它们直接损害基因组的稳定性。除了引起细胞死亡和无法正确地修复外,DSB可以经由形成结构性基因组改变而导致癌变,所述结构性基因组改变包括体细胞中的缺失、插入、DNA易位和有丝分裂重组事件。在健康细胞中,据估计在细胞周期过程中每个细胞形成多达50个内源性DSB。这些低水平的生理性断裂作为正常的细胞过程(诸如DNA复制、转录和染色质环化)的结果而偶尔发生,或是由于为了促进减数分裂期间的过程(诸如V(D)J重组)而由细胞编程的反复断裂形成而以较高水平发生。V(D)J重组是适应性免疫系统的定义性特征。其是在B细胞和T细胞成熟的早期阶段期间在发育的淋巴细胞中发生的遗传重组的独特机制。其涉及体细胞重组,并导致分别在B细胞和T细胞中发现的抗体/免疫球蛋白和T细胞受体(TCR)的高度多样性的谱系(repertoire)。

[0004] 除了通过正常细胞过程获得的DSB外,多种外源物理和化学试剂,诸如电离辐射、化疗药物和更近来的CRISPR基因组编辑技术,也是链断裂(特别是基因组DSB)的有效诱导物。

[0005] 理解在核酸(特别是基因组)中产生断裂的过程和修复它们的机制对于基因组医学是至关重要的。从确定癌症的原因和治愈,到基因组编辑技术的安全发展,对基因组中DNA双链断裂的频率、位置和原因的精确和准确的测量是最重要的。但是,在细胞中同时测量内源和外源/诱导的DSB的全基因组景观(genome-wide landscape)是具有挑战性的,主要是因为相对于频繁诱导的复发事件的极大范围的罕见偶发性断裂事件。

[0006] 下一代测序(NGS)的出现推动了在基因组规模上检测和测量与核酸链断裂(特别是DSB)相关的DNA序列的多种方法的开发。这些方法大致分为三类:

i)使用蛋白作为断裂的代替物的间接断裂标记(例如γ/H2AX ChIP-seq、DISCOVER-seq);

ii)修复的断裂的间接标记(例如GUIDE-seq、HTGTS),和

iii)细胞中未修复的断裂末端的直接标记(例如BLESS、DSBCapture、END-seq、BLISS)。

[0007] 尽管在这些方法的每次迭代中进行了许多技术上的渐进改进,但是它们所有均受困于共同的根本缺陷;在测序前通过PCR标记和富集样品中的DSB之后,依赖于NGS所需的标准DNA文库制备。该方法的PCR扩增阶段在DNA测序文库中引入了显著的偏差,导致样品中存在的DSB的原始模式的间接和失真的表现。这是一种公知的现象,使得不可能确定样品中原始DSB组成的量化。对许多NGS应用,诸如全基因组/外显子组测序,这可能不代表重要的问

题。但是对于特定特征(诸如DSB)的定量的、基因组范围的测量,PCR扩增偏差将高水平的噪声引入到信号(DSB)已经非常低的系统中。

[0008] 为了克服这一缺点,我们设计了新的DNA文库制备方案,该方案避免了在检测DSB之前通过PCR进行断裂序列扩增的需要,并且此外还增强了DSB信号。通过改善DSB检测的信噪比,我们获得了样品中基因组断裂的直接测量,其中一个序列读段(read)等同于样品中存在的一个断裂,理想情况下为DSB。

[0009] 发明概述

在本文中,我们描述了我们的新方法,称为INDUCE-seq,该方法用于直接测量核酸断裂(通常为DSB),此外,我们证明了该方法鉴定断裂模式的能力,所述断裂模式可以表征由多种不同的生理和诱导原因形成的断裂。

[0010] INDUCE-seq由此可以同时检测由生理转录和DNA复制引起的低水平偶发性双链断裂以及由限制酶或基因组编辑核酸酶(诸如CRISPR-Cas9)诱导的更高水平的复发性断裂的存在。INDUCE-seq因此可用于确定DNA断裂形成的起源及其修复机制,以及用于基因组编辑的安全发展。

[0011] 在本发明的一个方面,提供了用于鉴定核酸样品中的DNA双链断裂(DSB)的方法,该方法包括:

i)在连接条件下将怀疑含有DSB的核酸的样品暴露于第一对寡核苷酸,所述第一对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至所述DSB的第一链的5'结合特征、第一测序引物可与之结合的杂交位点(RD1 SP)和用于从DSB池中分离所述DSB的结合序列;和第二寡核苷酸,其与第一对的所述第一寡核苷酸互补并包含使得所述寡核苷酸能够连接至所述DSB的第二链的3'结合特征;其中所述寡核苷酸的一种或两种包含3'和/或5'保护性特征;

ii)将所述样品的核酸片段化成片段;

iii)在连接条件下将所述片段暴露于第二对寡核苷酸,所述第二对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至所述片段化核酸的第一链的5'结合特征和第二测序引物可与之结合的杂交位点(RD2 SP);和第二较长的寡核苷酸,其与第二对的所述第一寡核苷酸部分互补并包含用于与所述片段化核酸的第二链结合的3'结合特征、与所述杂交位点互补的序列、和另外的序列,其任选为用于实现桥式扩增的结合序列;并且其中所述寡核苷酸的一种或两种包含3'和/或5'保护性特征;

iv)使片段变性以提供单链核酸;

v)将部分iv)的链分成两组:A组,已经在第一末端处连接由部分i)的寡核苷酸提供的第一杂交位点和结合序列并在另一末端处连接由部分iii)的寡核苷酸提供的第二杂交位点和另外的序列的那些片段,和B组,并未在第一末端处连接由部分i)的寡核苷酸提供的杂交位点和结合序列以及并未在另一末端处连接由部分iii)的寡核苷酸提供的第二杂交位点和另外的序列的那些片段;和

vi)使用与第一和/或第二杂交位点结合的引物对A组的链进行测序,其中每个序列等同于一个DSB断裂,并且进一步地,其中碱基对缺失的数量和性质可以通过将每个序列与代表从中采集样品的所述物种的基因组进行比较来确定。

[0012] 包含5'结合特征的第二对的寡核苷酸可以不包含用于分离所述片段化核酸的结

合序列。

[0013] 部分i)和部分iii)的寡核苷酸可以互换,由此,在步骤ii)中片段化后,首先将核酸暴露于部分iii)的寡核苷酸,并随后将核酸暴露于部分i)的寡核苷酸。

[0014] 核酸样品可以是gDNA。

[0015] 5'和/或3'结合特征可包含以下的一种:磷酸基团;三磷酸"T尾",优选脱氧胸苷三磷酸"T尾";三磷酸"A尾",优选脱氧腺苷三磷酸"A尾";至少一个随机N核苷酸和多个N核苷酸。5'和/或3'保护性特征可包含提供对以下的任意一个或多个的抗性的特征:磷酸化活性、磷酸酶活性、末端转移酶活性、核酸杂交、核酸内切酶活性、核酸外切酶活性、连接酶活性、聚合酶活性和蛋白结合。保护性特征可包含硫代磷酸酯连接(phosphorothioate linkage)、双脱氧核苷酸或共价嵌段、亚磷酰胺、C3间隔子亚磷酰胺(3SpC3)。

[0016] 部分i)的所述第一寡核苷酸的5'结合特征可以是磷酸基团,并且部分i)的所述第二寡核苷酸的所述3'结合特征可以是三磷酸尾。部分i)的第一和第二寡核苷酸还可包含索引特征,所述索引特征是使得能够确定汇集样品(pooled sample)的来源的特定核苷酸序列。

[0017] 部分i)的第一寡核苷酸,从5'至3'读取,可包含5'结合特征和随后任选的保护性特征、测序引物可与之结合的杂交位点(RD1 SP)、索引序列、用于从DSB池中分离所述DSB的结合序列、和3'结合和/或保护性特征。结合特征可以是磷酸基团。部分i)的第二寡核苷酸,从3'至5'读取,可包含3'结合特征和随后任选的保护性特征、测序引物可与之结合的杂交序列(RD1 SP)、索引序列、用于从DSB池中分离所述DSB的结合序列、和5'结合和/或保护性特征。3'结合特征可包含3'脱氧胸苷三磷酸"T尾"以及硫代磷酸酯连接。

[0018] 部分i)的第一寡核苷酸对的第一和/或第二寡核苷酸可包含两种不同的末端保护性特征。部分iii)的第二寡核苷酸对的所述第一寡核苷酸的5'结合特征可以是磷酸基团,并且部分iii)的所述第二寡核苷酸的所述3'结合特征可以是三磷酸尾。

[0019] 部分iii)的第一和第二寡核苷酸还可包含索引特征,所述索引特征是使得能够确定汇集样品的来源的特定核苷酸序列。

[0020] 部分iii)的第二寡核苷酸,从5'至3'读取,可包含5'结合特征和随后任选的保护性特征、任选用于实现桥式扩增的另外的序列、索引序列、测序引物可与之结合的杂交位点(RD2 SP)、和3'结合和/或保护性特征。

[0021] 部分iii)的第二寡核苷酸对的第一或第二寡核苷酸可包含两种不同的末端保护性特征。

[0022] 部分i)的寡核苷酸可包含具有SEQ ID NO.1的第一寡核苷酸和具有SEQ ID NO.2的第二寡核苷酸;或与SEQ ID NO.1或2共享至少80%同一性或同源性的寡核苷酸。部分iii)的第二对寡核苷酸可包含具有SEQ ID NO.3的第一寡核苷酸和具有SEQ ID NO.4的第二寡核苷酸;或与SEQ ID NO.3或4共享至少80%同一性或同源性的寡核苷酸。部分iii)的所述第二对寡核苷酸的第二寡核苷酸可包含以下序列中的任一种:SEQ ID NO.4-28;或与SEQ ID NO.4-28中的一种共享至少80%同一性或同源性的寡核苷酸。

[0023] 样品可以是哺乳动物或人。部分i)中的连接可以原位发生或使用细胞或组织样品体外发生。在进行步骤i)之前,样品可以暴露于透化剂。在进行步骤i)之前,样品可以暴露于至少一种用于进行精氨酸尾修复的试剂。

[0024] 部分i)还可包括在进行后续步骤之前从所述样品中提取gDNA。

[0025] 方法可进一步包括在部分ii)和/或部分iv)之后,除去其尺寸小于大约100bp、或小于大约150bp的片段,和/或保留其尺寸大于大约150bp的片段。

[0026] 部分v)的分离可以涉及使用由部分i)的寡核苷酸提供的所述结合序列来结合配偶体,并因此将部分iv)的A组链与任何其它链分离。

[0027] 由部分i)的寡核苷酸提供的所述结合序列的互补结合链可以锚定在基底上,并且所述核酸单链流经或流过锚定的互补结合链。

[0028] 部分vi)可以涉及桥式扩增,其中在部分v)下分离的单链在已在部分i)的第一寡核苷酸的结合序列和部分iii)的第二寡核苷酸的另外的序列的寡核苷酸/结合位点锚定于其上的基底上进行克隆扩增。

[0029] 在进行要求保护的方法之前,可以将含有或怀疑含有单链断裂的样品连接或断裂以确保该单链断裂转化为双链断裂。

[0030] 在本发明的另一方面,提供了用于鉴定gDNA样品中的DNA双链断裂(DSB)的成套试剂盒,其包含:

i)第一对寡核苷酸,所述第一对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至所述DSB的第一链的5'结合特征、第一测序引物可与之结合的杂交位点(RD1 SP)和用于从DSB池中分离所述DSB的结合序列;和第二寡核苷酸,其与第一对的所述第一寡核苷酸互补,并包含用于与所述DSB的第二链结合的3'结合特征;并且其中所述寡核苷酸的一种或两种包含3'和/或5'保护性特征;和

ii)第二对寡核苷酸,所述第二对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至所述DSB的第一链的5'结合特征,和第二测序引物可与之结合的杂交位点(RD2 SP);和第二较长的寡核苷酸,其与第二对的所述第一寡核苷酸部分互补,并包含用于与所述DSB的第二链结合的3'结合特征、与所述杂交位点互补的序列、和另外的序列,其任选为用于实现桥式扩增的结合序列;并且其中所述寡核苷酸的一种或两种包含3'和/或5'保护性特征。

[0031] 在本发明的另一方面,提供了用于鉴定gDNA样品中的DSB的样品制备试剂盒,其包含:

i)第一对寡核苷酸,所述第一对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至双链核酸的一条链,并包含根据TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的序列的5'结合特征;和第二寡核苷酸,其与第一对的所述第一寡核苷酸互补;并且其中所述寡核苷酸的一种或两种分别包含3'和/或5'保护性特征;和

ii)第二对寡核苷酸,所述第二对寡核苷酸的第一寡核苷酸不包含序列ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的超过5、10、15或20个碱基的序列,或不包含序列ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的全部24个碱基;和第二寡核苷酸,其包含使得所述寡核苷酸能够连接至双链核酸的一条链,并包含根据CAAGCAGAAGACGGCATACGAGAT(SEQ ID NO:32)的序列的3'结合特征;并且其中所述寡核苷酸的一种或两种分别包含3'和/或5'保护性特征。

[0032] 在本发明的另一方面,提供了用于鉴定gDNA样品中的DSB的样品制备试剂盒,其包含:

i)第一对寡核苷酸,所述第一对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至双链核酸的一条链,并包含根据ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的序列的5'结合特征;和第二寡核苷酸,其与第一对的所述第一寡核苷酸互补;并且其中所述寡核苷酸的一种或两种分别包含3'和/或5'保护性特征;和

ii)第二对寡核苷酸,所述第二对寡核苷酸的第一寡核苷酸不包含序列TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的超过5、10或15个碱基的序列,或不包含序列TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的全部20个碱基;和第二寡核苷酸,其包含使得所述寡核苷酸能够连接至双链核酸的一条链,并包含根据AATGATACGGCGACCACCGA(SEQ ID NO:34)的序列的3'结合特征;并且其中所述寡核苷酸的一种或两种分别包含3'和/或5'保护性特征。

[0033] 第一对和/或第二对寡核苷酸的第一寡核苷酸可包含第一测序引物可与之结合的杂交位点。部分i)和/或部分ii)的第一和第二寡核苷酸可包含索引特征,所述索引特征是使得能够确定汇集样品的来源的特定核苷酸序列。

[0034] 试剂盒可进一步包含至少一种引物,其与第一和/或第二杂交位点结合以便用于测序目的。试剂盒可进一步包含分别用于将核酸片段化和/或变性成片段和/或单链的片段化剂和/或变性剂。

[0035] 在本发明的另一方面,提供了用于鉴定核酸样品,诸如gDNA样品中的DNA双链断裂(DSB)的双链衔接子,其包含:第一寡核苷酸链,其包含使得所述寡核苷酸能够连接至所述DSB的第一链的5'结合特征、测序引物可与之结合的杂交位点(RD1 SP)和用于从DSB池中分离所述DSB的结合序列;和第二寡核苷酸链,其与所述第一寡核苷酸互补并包含用于与所述DSB的第二链结合的3'结合特征;并且其中所述寡核苷酸的一种或两种包含3'和/或5'保护性特征。

[0036] 在本发明的另一方面,提供了用于鉴定核酸样品,诸如gDNA样品中的DNA双链断裂(DSB)的双链衔接子,其包含:第一寡核苷酸链,其包含使得所述寡核苷酸能够连接至所述DSB的第一链的5'结合特征和测序引物可与之结合的杂交位点(RD2 SP);和第二较长的寡核苷酸链,其与所述第一寡核苷酸部分互补,并包含用于与所述DSB的第二链结合的3'结合特征、与所述杂交位点互补的序列、和另外的序列,其任选为用于实现桥式扩增的结合序列;并且其中所述寡核苷酸的一种或两种包含3'和/或5'保护性特征。

[0037] 衔接子可包含含有索引特征的第一和第二寡核苷酸,所述索引特征是使得能够确定汇集样品的来源的特定核苷酸序列。

[0038] 在本发明的另一方面,提供了用于鉴定核酸样品中的DNA双链断裂(DSB)的双链衔接子,其包含:第一寡核苷酸链,其不包含序列ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的超过5、10、15或20个碱基的序列,或不包含序列ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的全部24个碱基;和第二寡核苷酸,其包含使得所述寡核苷酸能够连接至双链核酸的一条链并包含根据CAAGCAGAAGACGGCATACGAGAT(SEQ ID NO:32)的序列的3'结合特征;并且其中所述寡核苷酸的一种或两种分别包含3'和/或5'保护性特征。

[0039] 在本发明的另一方面,提供了用于鉴定核酸样品中的DNA双链断裂(DSB)的双链衔接子,其包含:第一寡核苷酸链,其不包含序列TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的超过5、10或15个碱基的序列,或不包含序列TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的全部20

个碱基;和第二寡核苷酸,其包含使得所述寡核苷酸能够连接至双链核酸的一条链,并包含根据AATGATACGGCGACCACCGA(SEQ ID NO:34)的序列的3'结合特征;并且其中所述寡核苷酸的一种或两种分别包含3'和/或5'保护性特征。

[0040] 在本发明的另一方面,提供了用于鉴定核酸样品中的DNA双链断裂(DSB)的样品制备方法,其中制备包括修饰DSB相关核酸以适于与包含固定引物的基底结合,该方法包括:

a)提供包含多个核酸的样品;

b)在有助于连接的条件下将多个核酸暴露于第一衔接子,其中第一衔接子包含能够连接至DSB的一条链的3'末端的寡核苷酸且所述寡核苷酸包含能够通过杂交与固定于基底的引物结合的序列;

c)片段化多个核酸;和

d)在有助于连接的条件下将多个核酸暴露于第二衔接子,其中第二衔接子包含能够在片段化诱导的断裂处连接至一条链的5'末端但不能连接至第一衔接子的寡核苷酸,且所述寡核苷酸不包含能够通过杂交与固定于基底的引物结合的序列。

[0041] 步骤d)的寡核苷酸可包含与第二引物的区域相同的序列。基底可包含第一和第二固定引物;步骤b)的寡核苷酸可包含能够通过杂交与第一固定引物结合的序列;并且步骤d)的寡核苷酸可包含与第二固定引物的区域相同的序列。

[0042] 在一个实施方案中,步骤b)是:在有助于连接的条件下将多个核酸暴露于第一衔接子对,其中第一衔接子对能够连接至DSB的一条链的至少3'末端,并且其中第一衔接子对包含至少部分互补的第一和第二寡核苷酸,并且第一寡核苷酸可连接至3'末端并包含能够通过杂交与固定于基底的引物结合的序列;以及步骤d)是:

在有助于连接的条件下将多个核酸暴露于第二衔接子对,其中第二衔接子对能够在片段化诱导的断裂处连接至一条链的至少5'末端,但不能连接至第一衔接子对的第一寡核苷酸,其中第二衔接子包含第一和第二部分互补的寡核苷酸,并且第一寡核苷酸可连接至5'末端并包含与第二引物的区域相同的序列,并且第二寡核苷酸不包含与所述与第二引物的区域相同的序列互补的序列。

[0043] 在方法的一个实施方案中,第二衔接子对包含:

含有根据CAAGCAGAAGACGGCATACGAGAT(SEQ ID NO:32)的序列的第一寡核苷酸,和含有不包含序列ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的超过5、10、15或20个碱基的序列,或不包含序列ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的全部24个碱基的序列的第二寡核苷酸;或

含有根据AATGATACGGCGACCACCGA(SEQ ID NO:34)的序列的第一寡核苷酸,和含有不包含序列TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的超过5、10或15个碱基的序列,或不包含序列TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的全部20个碱基的序列的第二寡核苷酸。

[0044] 第一衔接子对的第一和/或第二寡核苷酸可包含3'和/或5'保护性特征;和/或其中第二衔接子对的第一和/或第二寡核苷酸包含3'和/或5'保护性特征。在一个实施方案中,由于第一衔接子的3'修饰的存在,第二衔接子不能连接至第一衔接子。

[0045] 在一个实施方案中,可连接至5'末端的第二衔接子的寡核苷酸包含与固定引物的5、10、15、20、21、24或更多个碱基相同的序列。

[0046] 方法可进一步包含使多个核酸变性以形成多个单链核酸。方法可进一步包括在适

于固定引物与互补核酸杂交的条件下使多个核酸与包含固定引物的基底接触。方法可进一步包括获得与基底杂交的任何核酸的序列信息。

[0047] 在方法的一个实施方案中,包含多个核酸的样品是gDNA。

[0048] 在一些实施方案中,步骤以a)、b)、c)和随后d)的顺序进行。在另一些实施方案中,步骤以a)、c)、d)和随后b)的顺序进行;其中在步骤d)和b)之间将样品暴露于能够引起或怀疑能够引起DSB的条件。

[0049] 在一些实施方案中,将样品暴露于能够在核酸样品中的目标特征处引起DSB的条件。

[0050] 在本发明的另一方面,提供了用于鉴定核酸样品中的DNA双链断裂(DSB)的样品制备方法,其中制备包括修饰DSB相关核酸以适于与包含固定的第一引物的基底结合,该方法包括:

1)提供包含多个核酸的样品;

2)在有助于连接的条件下将多个核酸暴露于第一衔接子,其中第一衔接子包含能够连接至DSB的一条链的3'末端的寡核苷酸且所述寡核苷酸包含能够与第二引物杂交的序列;

3)片段化多个核酸;

4)在有助于连接的条件下将多个核酸暴露于第二衔接子,其中第二衔接子包含能够在片段化诱导的断裂处连接至一条链的5'末端但不能连接至第一衔接子的寡核苷酸,且所述寡核苷酸包含与固定的第一引物的区域相同的序列;和

5)在适于引物延伸的条件下使多个核酸与第二引物接触。

[0051] 在一个实施方案中,步骤2)是:在有助于连接的条件下将多个核酸暴露于第一衔接子对,其中第一衔接子对能够连接至DSB的一条链的至少3'末端,并且其中第一衔接子对包含至少部分互补的第一和第二寡核苷酸,并且第一寡核苷酸可连接至3'末端并包含能够与第二引物杂交的序列;以及步骤4)是:

在有助于连接的条件下将多个核酸暴露于第二衔接子对,其中第二衔接子对能够在片段化诱导的断裂处连接至一条链的至少5'末端但不能连接至第一衔接子对的第一寡核苷酸,其中第二衔接子包含第一和第二部分互补的寡核苷酸,并且第一寡核苷酸可连接至5'末端并包含与固定的第一引物的区域相同的序列,并且第二寡核苷酸不包含与所述与固定的第一引物的区域相同的序列互补的序列。

[0052] 第二衔接子对可包含含有根据AACCCACTACGCCTCCGCTTTCC(SEQ ID NO:40)的序列的第一寡核苷酸;和不包含序列GGAAAGCGGAGGCGTAGTGGTT(SEQ ID NO:36)的超过5、10、15、20个碱基的序列,或不包含序列GGAAAGCGGAGGCGTAGTGGTT(SEQ ID NO:36)的全部22个碱基的第二寡核苷酸。

[0053] 第一衔接子对的第一和/或第二寡核苷酸可包含3'和/或5'保护性特征;和/或其中第二衔接子对的第一和/或第二寡核苷酸包含3'和/或5'保护性特征。在一个实施方案中,由于第一衔接子的3'修饰的存在,第二衔接子不能连接至第一衔接子。

[0054] 在一个实施方案中,可连接至5'末端的第二衔接子的寡核苷酸包含与固定引物的5、10、15、20、21、24或更多个碱基相同的序列。

[0055] 方法可进一步包括使多个核酸变性以形成多个单链核酸。方法可进一步包括在适

于固定的第一引物与互补核酸杂交的条件下使多个核酸与包含固定的第一引物的基底接触。方法可进一步包括获得与基底杂交的任何核酸的序列信息。包含的样品可以是gDNA。

[0056] 在一个实施方案中，步骤以1)、2)、3)、4)和随后5)的顺序进行。在另一个实施方案中，步骤以1)、3)、4)、2)和随后5)的顺序进行；其中在步骤4)和2)之间将样品暴露于能够引起或怀疑能够引起DSB的条件。

[0057] 在一个实施方案中，将样品暴露于能够在核酸样品中的目标特征处引起DSB的条件。

[0058] 在本发明的另一方面，提供了用于鉴定核酸样品中的目标特征的样品制备方法，其中制备包括修饰与目标特征相关的核酸以适于与包含固定引物的基底结合，该方法包括：

a)提供包含多个核酸的样品，将多个核酸暴露于能够在目标特征处切割核酸的至少一条链的条件，并使多个核酸变性成单链核酸；

b)在有助于连接的条件下将多个核酸暴露于第一衔接子，其中第一衔接子包含能够连接至切割位点的一条链的3'末端的寡核苷酸且所述寡核苷酸包含能够通过杂交与固定于基底的引物结合的序列；

c)片段化多个核酸；和

d)在有助于连接的条件下将多个核酸暴露于第二衔接子，其中第二衔接子包含能够在片段化诱导的断裂处连接至一条链的5'末端但不能连接至第一衔接子的寡核苷酸，且所述寡核苷酸不包含能够通过杂交与固定于基底的引物结合的序列。

[0059] 目标特征可以是能够被特异性切割的任何特征。目标特征可以是环丁烷嘧啶二聚体(CPD)、8-氧代鸟嘌呤或脱碱基位点。

[0060] 步骤d)的寡核苷酸可包含与第二引物的区域相同的序列。基底可包含第一和第二固定引物；步骤b)的寡核苷酸可包含能够通过杂交与第一固定引物结合的序列；以及步骤d)的寡核苷酸可包含与第二固定引物的区域相同的序列。

[0061] 在一个实施方案中，步骤b)是：在有助于连接的条件下将多个核酸暴露于第一衔接子对，其中第一衔接子对能够连接至切割位点的一条链的至少3'末端，并且其中第一衔接子对包含至少部分互补的第一和第二寡核苷酸，并且第一寡核苷酸可连接至3'末端并包含能够通过杂交与固定于基底的引物结合的序列；以及其中步骤d)是：

在有助于连接的条件下将多个核酸暴露于第二衔接子对，其中第二衔接子对能够在片段化诱导的断裂处连接至一条链的至少5'末端但不能连接至第一衔接子对的第一寡核苷酸，其中第二衔接子包含第一和第二部分互补的寡核苷酸，并且第一寡核苷酸可连接至5'末端并包含与第二引物的区域相同的序列，并且第二寡核苷酸不包含与所述与第二引物的区域相同的序列互补的序列。

[0062] 在一个实施方案中，第二衔接子对包含：

含有根据CAAGCAGAAGACGGCATACGAGAT(SEQ ID NO:32)的序列的第一寡核苷酸，和含有不包含序列ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的超过5、10、15或20个碱基的序列，或不包含序列ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的全部24个碱基的序列的第二寡核苷酸；或含有根据AATGATACGGCGACCACCGA(SEQ ID NO:34)的序列的第一寡核苷酸，和含有不包含序列TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的超过5、10或15个碱基的序

列,或不包含序列TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的全部20个碱基的序列的第二寡核苷酸。

[0063] 第一衔接子对的第一和/或第二寡核苷酸可包含3'和/或5'保护性特征;和/或其中第二衔接子对的第一和/或第二寡核苷酸可包含3'和/或5'保护性特征。在一个实施方案中,由于第一衔接子的3'修饰的存在,第二衔接子不能连接至第一衔接子。

[0064] 可连接至5'末端的第二衔接子的寡核苷酸可包含与固定引物的5、10、15、20、21、24或更多个碱基相同的序列。

[0065] 在一个实施方案中,方法进一步包括使多个核酸变性以形成多个单链核酸。在一个实施方案中,方法进一步包括在适于固定引物与互补核酸杂交的条件下使多个核酸与包含固定引物的基底接触。在一个实施方案中,方法进一步包括获得与基底杂交的任何核酸的序列信息。

[0066] 在一个实施方案中,包含多个核酸的样品是gDNA。在一个实施方案中,步骤以a)、b)、c)和随后d)的顺序进行。在另一实施方案中,步骤以c)、d)、a)和随后b)的顺序进行。

[0067] 在本发明的另一方面,提供了用于鉴定核酸样品,诸如gDNA样品中的DNA双链断裂(DSB)的双链衔接子,其包含:第一寡核苷酸链,其不包含序列GGAAAGCGGAGGCGTAGTGGTT(SEQ ID NO:36)的超过5、10、15、20个碱基的序列,或不包含序列GGAAAGCGGAGGCGTAGTGGTT(SEQ ID NO:36)的全部22个碱基;和第二寡核苷酸,其包含使得所述寡核苷酸能够连接至双链核酸的一条链并包含根据AACCCACTACGCCTCCGCTTTCC(SEQ ID NO:40)的序列的3'结合特征;并且其中所述寡核苷酸的一种或两种分别包含3'和/或5'保护性特征。

附图说明

图1.INDUCE-seq的概述。通过将全长化学修饰的P5测序衔接子连接至末端制备好的DSB来进行固定和透化的细胞中的原位断裂标记。随后提取基因组DNA,片段化,末端制备,并使用化学修饰的半功能性P7衔接子连接。所得DNA文库含有功能性DSB标记的片段(P5:P7)和非功能性基因组DNA片段(P7:P7)的混合物。随后对INDUCE-seq文库的测序富集了DNA标记的片段并消除了所有其它非功能性DNA。由于INDUCE-seq文库制备是无PCR的,因次获得的每个测序读段等同于单个标记的DSB末端;

图2.流动池富集衔接子设计的详细示意图。(a)用于测序的完整的衔接子连接的dsDNA片段的结构。3'P5和P7衔接子与流动池杂交。测序引物在第一和第二测序读取过程中与Read 1测序引物(RD1 SP)和Read 2测序引物(RD2 SP)序列结合。索引允许区分文库池中的不同样品。(b)存在于INDUCE-seq文库中的DNA片段的结构。只有DSB连接的片段由测序所需的所有衔接子组件构成。(c)将INUCE-seq文库加载到测序流动池上将经由P5衔接子序列的3'端的杂交来富集DSB连接的片段。没有其它片段可以与流动池相互作用并将被除去;

图3.INDUCE-seq与替代的DSB测序技术相比显示出空前的灵敏度和动态范围。(a)INDUCE-seq以高分辨率同时检测高度复发性诱导DSB和单个内源性DSB。在用限制性核酸内切酶HindIII原位切割后,INDUCE-seq读段的基因组浏览器(Genome browser)视图映射为来自HEK293T细胞的基因组的10mb部分。(上图)在高水平上观察时(10mb,0-1000个读段),高度复发性酶诱导的断裂代表了绝大多数读段。(下图)拉近放大(粉红色高亮,500kb,0-20个读段)显示在未处理样品中和复发性HindIII诱导断裂中存在低水平的单一内源性断裂

(绿色高亮)。(b)HindIII靶位点处的INDUCE-seq读段的映射显示了单核苷酸断裂映射的精确性。该图包括TACTCAAGCTTACCCCTA(SEQ ID NO:35)和GGGGGGTAAGCTTGAGTA(SEQ ID NO:43)。(c)HindIII处理的样品和对照样品的每个细胞测得的断裂的定量。INDUCE-seq定量地检测样品之间跨越3个数量级的每个细胞的断裂。(d和e)在检测由酶HindIII和EcoRV体外切割限制性位点方面,INDUCE-seq和DSBCapture之间的比较。(d)使用INDUCE-seq将更大比例的测序并与基因组比对的读段映射至限制性位点。(e)使用1/800的细胞,INDUCE-seq鉴定了与通过DSBCapture鉴定的EcoRV比例(93.7%)相似的HindIII限制性位点比例(92.7%)。(f)使用INDUCE-seq的诱导DSB检测的动态范围。除了在HindIII在靶序列(AAGCTT)处鉴定的断裂外,还鉴定了多个1bp和2bp错配的脱靶位点。INDUCE-seq测量的诱导断裂事件跨越8个数量级,从在HindIII在靶位点处鉴定的～1.5亿个断裂到在最稀有的脱靶处鉴定的5个断裂。(g)在活DiVA细胞中检测AsiSI诱导的断裂方面,INDUCE-seq、DSBCapture和BLISS之间的比较。将测序的读段的数量(上图)与每个实验鉴定的AsiSI位点的数量(下图)进行比较。INDUCE-seq使用为DSBCapture的1/40的读段和为BLISS的1/23的读段灵敏地检测到了最大数量的AsiSI位点;

图4.INDUCE-seq灵敏地发现和定量CRISPR/Cas9诱导的在靶和脱靶DSB。(a)使用INDUCE-seq鉴定的EMX1 sgRNA的脱靶序列和断裂数量。该图包括GAGTCCGAGCAGAAGAAGAANGG(SEQ ID NO:44)。(b)INDUCE-seq显示了细胞群中形成EMX1诱导的DSB的动力学。对每个样品的每百万个读段检测到的断裂数的定量显示在细胞核转染后立即发生的在靶和脱靶的高Cas9活性。(c)通过INDUCE-seq鉴定的脱靶与成熟的体外方法CIRCLE-seq和DiGenome-seq,以及基于细胞的方法GUIDE-seq、BLISS和HTGTS之间的比较。INDUCE-seq检测了许多以前仅可通过体外方法鉴定的脱靶。与任何目前基于细胞的方法相比鉴定了明显更多的脱靶位点。INDUCE-seq还鉴定了不能通过任何其它方法检测的多个脱靶。(d)进行扩增子测序以测量INDUCE-seq鉴定的脱靶处的插入缺失频率。扩增子测序仅能鉴定60个由INDUCE-seq鉴定的脱靶中的4个,并且受限于0.1%的背景插入缺失错误发现率。这些发现与先前的研究一致,所述研究在EMX1 RNP核转染后48小时测量EMX1脱靶插入缺失频率;

图5.INDUCE-seq与当前的DSB映射工作流程的比较。(a)INDUCE-seq工作流程的概述。INDUCE-seq文库的测序产生定量的输出,其中一个读段等同于一个断裂。(b)DSBCapture、BLISS和END-seq工作流程的概述。在标准文库构建后的测序产生输出,一个读段不等同于单个DSB;

图6.测序的读段数量和对INDUCE-seq和BLISS NGS文库限定的DSB数量之间的比较。(a)测序的INDUCE-seq读段的数量和由单个INDUCE-seq实验限定的断裂数量的散点图。(b)显示来自单个BLISS实验的用正确的read 1(R1)条形码前缀测序的BLISS读段数量和使用UMI校正去除重复后限定的断裂数量的散点图;

图7.HEK293细胞中DSB热点的基因组浏览器视图。(a)ch17 DSB热点的11kb视图。紫色箭头表示右侧(+链)标记的DSB末端,而蓝色箭头表示左侧(-链)标记的DSB末端。复发性DSB均匀地遍布热点区域分布。(b)chr11 DSB热点的5kb视图。在正链和负链的不同位置处可以检测到复发性DSB;

图8.用于INDUCE-seq的脱靶发现流程的示意图。图8中的序列是SEQ ID NO:44;

图9.使用INDUCE-seq的CRISPR脱靶发现是高度可再现的。(a)在r1和r2时程实验中检测到的EMX1脱靶数量之间的比较。(b)显示在两个独立实验中鉴定的CRISPR脱靶位点处发现的断裂数量的散点图；

图10.显示由INDUCE-seq、CIRCLE-seq、GUIDE-seq和BLISS鉴定的脱靶的交叉的韦恩图。(a和b)从独立实验r1(a)和r2(b)计算的单独的0小时至30小时样品的重叠。(c和d)来自集合r1(c)和r2(d)的所有时间点的组合重叠。(e)当组合所有INDUCE-seq样品时，在方法之间计算的重叠；

图11.CRISPR诱导的在靶和脱靶处的DSB模式与编辑结果相关。EMX1在靶(a)和排名靠前的脱靶(显示的序列从上到下的顺序为SEQ ID NO:45-65)、OT-1(显示的序列从上到下的顺序为SEQ ID NO:66-87)(b)和OT-2(显示的序列从上到下的顺序为SEQ ID NO:88-105)(c)跨越180bp的覆盖轨迹。围绕每个靶位点的40bp区域的特写视图显示了独特的1bp突出切割模式，而不是通常的Cas9诱导的平端DSB。每个位点处相应的插入缺失谱显示与观察到的断裂位点相关的突变位置；

图12.示出了示例性化学修饰的示例性实施方案。

[0069] 图13.半功能性衔接子在全功能性衔接子之前连接的示例性实施方案。该图示出了用于检测人工诱导的DSB的实施方案。例如，用于检测由核酸酶引起的脱靶DSB。

[0070] 图14.将本发明的方法应用于微珠(bead)测序技术的示例性工作流程。

[0071] 图15.用于与微珠测序技术一起使用的示例性衔接子。

[0072] 图16.检测影响一条链的DNA损伤的示例性实施方案。该实施方案涉及半功能性衔接子在全功能性衔接子之前的连接。

[0073] 图17.显示CRISPR碱基编辑-胞嘧啶碱基编辑器(CBE)的体外检测的示例性实施方案。

[0074] 图18.显示CRISPR碱基编辑-腺嘌呤碱基编辑器(ABE)的体外检测的示例性实施方案。

[0075] 表1.根据本发明的优选实施方案的实例衔接子序列。

[0075] 发明详述

在本发明的一个方面，提供了用于鉴定核酸样品中的DNA双链断裂(DSB)的方法，该方法包括：

i)在连接条件下将怀疑含有DSB的核酸的样品暴露于第一对寡核苷酸，所述第一对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至所述DSB的第一链的5'结合特征、第一测序引物可与之结合的杂交位点(RD1 SP)和用于从DSB池中分离所述DSB的结合序列；和第二寡核苷酸，其与第一对的所述第一寡核苷酸互补并包含使得所述寡核苷酸能够连接至所述DSB的第二链的3'结合特征；其中所述寡核苷酸的一种或两种包含3'和/或5'保护性特征；

ii)将核酸样品(理想情况下为gDNA)片段化成片段；

iii)在连接条件下将所述片段暴露于第二对寡核苷酸，所述第二对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至所述片段的第一链(理想情况下在远离所述第一对的所述寡核苷酸的结合的位点处)的5'结合特征，和第二测序引物可与之结合的杂交位点(RD2SP)；和第二较长的寡核苷酸，其与第二对的所述第一寡核苷酸部分互补并包含

用于与所述片段或所述DSB的第二链结合的3’结合特征、与所述杂交位点互补的序列、和另外的序列，其任选为用于实现桥式扩增的结合序列；并且其中所述寡核苷酸的一种或两种包含3’和/或5’保护性特征；

iv)使片段变性以提供单一核酸；

v)将部分iv)的链分成两组：A组，已经在第一末端处连接由部分i)的寡核苷酸提供的第一杂交位点和结合序列并在另一末端处连接由部分iii)的寡核苷酸提供的第二杂交位点和另外的序列的那些片段，和B组，并未在第一末端处连接由部分i)的寡核苷酸提供的杂交位点和结合序列以及并未在另一末端处连接由部分iii)的寡核苷酸提供的第二杂交位点和另外的序列的那些片段；和

vi)使用与第一和/或第二杂交位点结合的引物对A组的链进行测序，其中每个序列等同于一个断裂，通常为一个DSB。任选进一步地，其中碱基对缺失的数量和性质可以通过将每个序列与代表从中采集样品的所述物种的基因组进行比较来确定。

[0076] 本发明的方法的步骤不需要在一个时间或在一个位置进行。例如，在一个方面，本发明的方法允许制备核酸样品，所述核酸样品已经以允许选择和分离出DSB的方式被标记。由此，在本发明的一个方面，提供了用于鉴别核酸样品中的DNA双链断裂(DSB)的样品制备方法，其包括：

i)在连接条件下将怀疑含有DSB的核酸的样品暴露于第一对寡核苷酸，所述第一对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至所述DSB的第一链的5’结合特征、第一测序引物可与之结合的杂交位点(RD1 SP)和用于从DSB池中分离所述DSB的结合序列；和第二寡核苷酸，其与第一对的所述第一寡核苷酸互补并包含使得所述寡核苷酸能够连接至所述DSB的第二链的3’结合特征；其中所述寡核苷酸的一种或两种包含3’和/或5’保护性特征；

ii)将核酸样品(理想情况下为gDNA)片段化成片段；和

iii)在连接条件下将所述片段暴露于第二对寡核苷酸，所述第二对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至所述片段的第一链(任选在远离所述第一对的所述寡核苷酸的结合的位点处)的5’结合特征，和第二测序引物可与之结合的杂交位点(RD2 SP)；和第二较长的寡核苷酸，其与第二对的所述第一寡核苷酸部分互补并包含用于与所述片段的第二链结合的3’结合特征、与所述杂交位点互补的序列、和另外的序列，其任选为用于实现桥式扩增的结合序列；并且其中所述寡核苷酸的一种或两种包含3’和/或5’保护性特征。

[0077] 备选地，可以进行本发明的方法，直到链被分成A组和B组，并可以分别进行测序。由此，该方法允许制备其中任何DSB已经被选择并分离出来的核酸样品。由此，在一个实施方案中，提供了用于鉴定核酸样品中的DNA双链断裂(DSB)的样品制备方法，其包括：

i)在连接条件下将怀疑含有DSB的核酸的样品暴露于第一对寡核苷酸，所述第一对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至所述DSB的第一链的5’结合特征、第一测序引物可与之结合的杂交位点(RD1 SP)和用于从DSB池中分离所述DSB的结合序列；和第二寡核苷酸，其与第一对的所述第一寡核苷酸互补并包含使得所述寡核苷酸能够连接至所述DSB的第二链的3’结合特征；其中所述寡核苷酸的一种或两种包含3’和/或5’保护性特征；

ii)将核酸样品(理想情况下为gDNA)片段化成片段;

iii)在连接条件下将所述片段暴露于第二对寡核苷酸,所述第二对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至所述片段的第一链(任选在远离所述第一对的所述寡核苷酸的结合的位点处)的5'结合特征,和第二测序引物可与之结合的杂交位点(RD2 SP);和第二较长的寡核苷酸,其与第二对的所述第一寡核苷酸部分互补并包含用于与所述片段的第二链结合的3'结合特征、与所述杂交位点互补的序列、和另外的序列,其任选为用于实现桥式扩增的结合序列;并且其中所述寡核苷酸的一种或两种包含3'和/或5'保护性特征;

iv)使片段变性以提供单一核酸;和

v)将部分iv)的链分成两组:A组,已经在第一末端处连接由部分i)的寡核苷酸提供的第一杂交位点和结合序列并在另一末端处连接由部分iii)的寡核苷酸提供的第二杂交位点和另外的序列的那些片段,和B组,并未在第一末端处连接由部分i)的寡核苷酸提供的杂交位点和结合序列以及并未在另一末端处连接由部分iii)的寡核苷酸提供的第二杂交位点和另外的序列的那些片段。

[0078] 在本发明的一个方面,提供了用于鉴定核酸样品中的DNA双链断裂(DSB)的方法,其包括:

i)在连接条件下将怀疑含有DSB的核酸的样品暴露于第一对寡核苷酸,所述第一对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至所述DSB的第一链的5'结合特征,和用于从DSB池中分离所述DSB的结合序列;和第二寡核苷酸,其与第一对的所述第一寡核苷酸互补;其中所述寡核苷酸的一种或两种包含3'和/或5'保护性特征;

ii)将核酸样品(例如gDNA)片段化成片段;

iii)在连接条件下将所述片段暴露于第二对寡核苷酸,所述第二对寡核苷酸的第一寡核苷酸与第二对的第二寡核苷酸部分互补;和第二寡核苷酸,其包含用于与所述片段的第二链结合的3'结合特征、和任选的用于实现桥式扩增的结合序列;并且其中所述寡核苷酸的一种或两种包含3'和/或5'保护性特征;

iv)使片段变性以提供单一核酸;

v)将部分iv)的链分成两组:A组,已经在第一末端处连接由部分i)的寡核苷酸提供的结合序列并任选在另一末端处连接由部分iii)的寡核苷酸提供的用于实现桥式扩增的结合序列的那些片段,和B组,并未在第一末端处连接由部分i)的寡核苷酸提供的结合序列以及任选并未在另一末端处连接由部分iii)的寡核苷酸提供的用于实现桥式扩增的结合序列的那些片段;和

vi)对A组的链进行测序,其中每个序列等同于一个断裂,通常为一个DSB,并且任选进一步地,其中碱基对缺失的数量和性质可以通过将每个序列与代表从中采集样品的所述物种的基因组进行比较来确定。如上所述,方法可包含步骤i)、ii)和iii);或i)、ii)、iii)、iv)和v);或所有步骤。

[0079] 在特定实施方案中,包含5'结合特征的第二对的寡核苷酸不包含用于分离所述片段化核酸的结合序列。

[0080] 在本发明的实施方案中,部分i)和部分iii)的寡核苷酸可以互换,由此首先将核酸暴露于部分iii)的寡核苷酸,并随后暴露于部分i)的寡核苷酸。

[0081] 在本发明的一个实施方案中，部分i)和部分iii)的寡核苷酸可以互换，由此，首先将核酸暴露于部分iii)的寡核苷酸，并在片段化后，随后将核酸片段暴露于部分i)的寡核苷酸。

[0082] 涉及部分i)和部分iii)的互换的实施方案尤其与用于检测已经诱导的DSB的方法有关。在此类实施方案中，样品可以片段化，并且随后可以进行步骤iii)。随后，可以对样品进行处理以潜在地诱导DSB，并且在诱导后，可以进行步骤i)。在本文中进一步讨论了其中引入DSB的实施方案，并且包括用于检测核酸酶的脱靶效应的实施方案等。图13、17和18示出了特定实施方案。

[0083] 在本发明的方法中，所述寡核苷酸对的任一种可以是或可以被称为衔接子。

[0084] 在本文中提及衔接子(或衔接头)是指遗传工程中的接头，并且它是可以连接至其它DNA分子末端的寡核苷酸。衔接子可以是双链的。双链衔接子可以合成为在两个末端都具有平末端，或者一个末端为粘性末端而另一个为平末端。衔接子可以是短的，并且可以是化学合成的。

[0085] 第一对寡核苷酸可以配置为允许经由与固定于基底的寡核苷酸杂交而结合到所述基底上，其中固定的寡核苷酸被定向，使得5’端在固定点的近端，而3’端在固定点的远端。例如，连接至DSB的一条链的3’末端的第一对的寡核苷酸(即包含5’结合特征的寡核苷酸)可以与固定的寡核苷酸至少部分互补。互补性程度可允许经由杂交与固定的寡核苷酸结合。第二对寡核苷酸可以配置为不允许经由与固定于基底的寡核苷酸杂交而结合到所述基底上，其中固定的寡核苷酸被定向，使得5’端在固定点的近端，而3’端在固定点的远端。例如，连接至片段化位点的一条链的3’末端的第二对的寡核苷酸(即包含5’结合特征的寡核苷酸)可与任何固定的寡核苷酸都互补性不足以致不能经由杂交结合。在一些实施方案中，例如为了允许后续的桥式扩增，与片段化位点的一条链的5’末端结合的第二对的寡核苷酸(即包含3’结合特征的寡核苷酸)可以是与固定的寡核苷酸至少部分相同的序列。在另一个实施方案中，例如为了允许微珠乳液扩增，可以采用其它布置(例如参见图14)。

[0086] 在一个实施方案中，提供了用于鉴定核酸样品中的DNA双链断裂(DSB)的样品制备方法，其中制备包括修饰DSB相关核酸以适于与包含固定的寡核苷酸的基底结合，该方法包括：

i)在连接条件下将怀疑含有DSB的核酸的样品暴露于第一对寡核苷酸，所述第一对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至所述DSB的第一链的5’结合特征、第一测序引物可与之结合的杂交位点(RD1 SP)和用于从DSB池中分离所述DSB的结合序列，其中结合序列与固定的寡核苷酸至少部分互补；和第二寡核苷酸，其与第一对的所述第一寡核苷酸互补并包含使得所述寡核苷酸能够连接至所述DSB的第二链的3’结合特征；其中所述寡核苷酸的一种或两种包含3’和/或5’保护性特征；

ii)将核酸样品(理想情况下为gDNA)片段化成片段；和

iii)在连接条件下将所述片段暴露于第二对寡核苷酸，所述第二对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至所述片段的第一链(理想情况下在远离所述第一对的所述寡核苷酸的结合的位点处)的5’结合特征、第二测序引物可与之结合的杂交位点(RD2 SP)，并且其中第一寡核苷酸不包含能够与固定的寡核苷酸杂交的序列；和第二较长的寡核苷酸，其与第二对的所述第一寡核苷酸部分互补并包含用于与所述片段的第二

链结合的3’结合特征、与所述杂交位点互补的序列、和另外的序列，其任选为用于实现桥式扩增的结合序列；并且其中所述寡核苷酸的一种或两种包含3’和/或5’保护性特征。

[0087] DSB相关核酸是位于DSB的一侧的核酸。因此，DSB相关核酸的测序使得能够鉴定DSB例如在基因组中的位置。

[0088] 在一个实施方案中，本发明的方法设计为与Illumina P5和P7衔接子一起使用。由此，在一个特定实施方案中，第二寡核苷酸对不包含序列ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的超过5、10、15或20个碱基的序列，或不包含序列ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的全部24个碱基。在另一个实施方案中，第二寡核苷酸对不包含序列TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的超过5、10或15个碱基的序列，或不包含序列TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的全部20个碱基。

[0089] 由此，在一个特定实施方案中，步骤iii)是：

在连接条件下将所述片段暴露于第二对寡核苷酸，所述第二对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至所述片段的第一链(理想情况下在远离所述第一对的所述寡核苷酸的结合的位点处)的5’结合特征，和第二测序引物可与之结合的杂交位点(RD2 SP)；和第二较长的寡核苷酸，其与第二对的所述第一寡核苷酸部分互补并包含用于与所述DSB的第二链结合的3’结合特征、与所述杂交位点互补的序列、和另外的序列，其任选为用于实现桥式扩增的结合序列；并且其中所述寡核苷酸的一种或两种包含3’和/或5’保护性特征；并且其中第二对寡核苷酸不包含SEQ ID NO:30的超过5、10、15、20、或24个碱基的序列和/或不包含SEQ ID NO:31的超过5、10、15、或20个碱基的序列。

[0090] 在本发明的优选方法中，所述寡核苷酸对的所述5’或3’结合特征包含以下之一：磷酸基团；三磷酸“T尾”，优选脱氧胸苷三磷酸“T尾”；三磷酸“A尾”，优选脱氧腺苷三磷酸“A尾”；至少一个随机N核苷酸，优选多个N核苷酸，或允许所述衔接子连接至所述DSB的任何其它已知结合基团。

[0091] 核酸样品可以是能够包含DSB的任何DNA样品。在本发明的一个优选实施方案中，所述核酸样品是gDNA。

[0092] 更理想情况下，部分i)的所述第一寡核苷酸的所述5’结合特征是磷酸基团，而部分i)的所述第二寡核苷酸的所述3’结合特征是三磷酸尾。

[0093] 在本发明的一个优选实施方案中，所述第一对寡核苷酸的所述5’和/或3’保护性特征包含提供对以下的任意一个或多个的抗性的特征：磷酸化活性、磷酸酶活性、末端转移酶活性、核酸杂交、核酸内切酶活性、核酸外切酶活性、连接酶活性、聚合酶活性和蛋白结合。这可以通过本领域技术人员已知的任何手段来实现，诸如但不限于硫代磷酸酯连接、亚磷酰胺间隔子、磷酸基团、2’-O-甲基基团、反向脱氧或双脱氧T修饰、锁核酸碱基、双脱氧核苷酸等。这些特征提供的活性的实例显示在表2中。

[0094] 优选地，部分i)的所述第一寡核苷酸包含提供对核酸外切酶活性的抗性的3’保护性特征，诸如硫代磷酸酯连接。此外或备选地，所述3’保护性特征还提供对连接酶活性和/或聚合酶活性(理想情况下为5’>3’聚合酶活性)的抗性，并且例如是双脱氧核苷酸或物理嵌段(理想情况下为亚磷酰胺形式，特别是C3间隔子亚磷酰胺(3SpC3))，或本领域技术人员已知的任何其它保护性特征，诸如表2中提供对核酸外切酶活性和/或连接酶活性和/或聚合酶活性的抗性的那些。

[0095] 更优选地，部分i)的所述第一和第二寡核苷酸还包含索引特征，其是使得能够确定汇集的测序文库的来源的特定核苷酸序列(例如GATCT)，换句话说，其使得能够对汇集的测序文库进行拆分(demultiplexing)。理想情况下，索引特征位于所述杂交位点和所述结合序列之间。

[0096] 最优选地，部分i)的所述第一寡核苷酸，从5’至3’读取，包含5’结合特征和随后任选的保护性特征(理想情况下，结合特征是磷酸基团)、第一测序引物可与之结合的杂交位点(RD1 SP)、索引序列、用于从DSB池中分离所述DSB的结合序列、和3’结合和/或保护性特征。优选地，所述保护性特征提供对以下的任意一个或多个的抗性：核酸外切酶活性、连接酶活性和/或聚合酶活性。

[0097] 最优选地，部分i)的所述第二寡核苷酸，从3’至5’读取，包含3’结合特征和随后任选的保护性特征、第一测序引物可与之结合的杂交序列(RD1 SP)、索引序列、用于从DSB池中分离所述DSB的结合序列、和5’结合和/或保护性特征。优选地，所述保护性特征提供对以下的任意一个或多个的抗性：核酸外切酶活性、连接酶活性和/或聚合酶活性。理想情况下，结合特征是3’的。

[0098] 在本发明的一个优选实施方案中，部分i)的所述第一寡核苷酸是第一寡核苷酸对中的一个，并且该第一寡核苷酸对的第二寡核苷酸与所述第一寡核苷酸互补并包含5’和3’保护性特征，优选地，所述5’和3’保护性特征提供对以下的任意一个或多个的抗性：核酸外切酶活性、连接酶活性和/或聚合酶活性。由此，在某些实施方案中，用于实施本发明的寡核苷酸缺少5’或3’磷酸基团。

[0099] 在本发明的一个优选实施方案中，部分i)的所述第一寡核苷酸含有用于对连接的DSB进行测序的杂交位点(/序列)和用于从DSB池中分离所述DSB的结合序列。但是，本领域技术人员将理解，部分i)的第二寡核苷酸可以含有用于对连接的DSB进行测序的杂交位点(/序列)和用于从DSB池中分离所述DSB的结合序列。这是因为测序和分离都将由用于测序的引物的性质和用于分离的寡核苷酸的性质决定。最典型地，所有寡核苷酸的取向，即至少第一和/或第二测序引物可与之结合的杂交位点(RD1 SP和RD2 SP)、和用于从DSB池中分离所述DSB的结合序列、以及任选用于实现桥式扩增的另外的序列的取向使得当进行部分v)的分离时，仅A组链可以使用用于从DSB池中分离所述DSB的结合序列提取。此外，这种取向也使得当进行部分vi)的测序时，仅A组链可被桥接(当存在桥式扩增序列时)。

[0100] 更理想情况下，部分i)的第一寡核苷酸对的第一或第二寡核苷酸包含两种不同的末端保护性特征。

[0101] 再更优选地，部分i)的该第一寡核苷酸对的第二寡核苷酸包含3’脱氧胸苷三磷酸“T尾”以提供用于连接至“A加尾”DNA片段的基底，并还理想情况下包含硫代磷酸酯连接，由此赋予对核酸外切酶活性的抗性。

[0102] 在本发明的再进一步优选的方法中，部分iii)的第二寡核苷酸对的所述第一寡核苷酸的所述5’结合特征是磷酸基团，并且部分iii)的第二寡核苷酸对的所述第二寡核苷酸的所述3’结合特征是三磷酸尾。

[0103] 在本发明的一个优选实施方案中，所述第二对寡核苷酸的所述5’和/或3’保护性特征包含提供对以下的任意一个或多个的抗性的特征：磷酸化活性、磷酸酶活性、末端转移酶活性、核酸杂交、核酸内切酶活性、核酸外切酶活性、连接酶活性、聚合酶活性和蛋白结

合。这可以通过本领域技术人员已知的任何手段来实现，诸如但不限于硫代磷酸酯连接、亚磷酰胺间隔子、磷酸基团、2'-O-甲基基团、反向脱氧或双脱氧T修饰、锁核酸碱基、双脱氧核苷酸等。这些特征提供的活性的实例显示在表2中。

[0104] 优选地，部分iii)的所述第一寡核苷酸还包含提供对核酸外切酶活性的抗性的3'保护性特征，诸如亚磷酰胺间隔子。此外或备选地，所述3'保护性特征还提供对连接酶活性和/或聚合酶活性(理想情况下为5'>3'聚合酶活性)的抗性，并且例如是双脱氧核苷酸或物理嵌段(理想情况下为亚磷酰胺形式，特别是C3间隔子亚磷酰胺(3SpC3))，或本领域技术人员已知的任何其它保护性特征，诸如表2中提供对核酸外切酶活性和/或连接酶活性和/或聚合酶活性的抗性的那些。

[0105] 更优选地，部分iii)的所述第一和第二寡核苷酸还包含索引特征，其是使得能够确定汇集的测序文库的来源的特定核苷酸序列(例如GATCT)，换句话说，其使得能够对汇集的测序文库进行拆分。理想情况下，索引特征位于所述杂交位点和所述另外的序列(当存在时)之间。

[0106] 更理想情况下，部分iii)的该第二寡核苷酸对的第一或第二寡核苷酸包含两种不同的末端保护性特征。

[0107] 最优选地，部分iii)的所述第二寡核苷酸，从5'至3'读取，包含5'结合和/或保护性特征、任选用于实现桥式扩增的另外的序列、索引序列、测序引物可与之结合的杂交位点(RD2 SP)、和3'结合和/或保护性特征。优选地，一种或两种保护性特征提供对以下的任意一个或多个的抗性：核酸外切酶活性、连接酶活性和/或聚合酶活性。

[0108] 在本发明的一个优选实施方案中，部分iii)的所述第二寡核苷酸与该寡核苷酸对的所述第一寡核苷酸互补并包含5'和3'保护性特征，优选地，所述5'和3'保护性特征提供对以下的任意一个或多个的抗性：核酸外切酶活性、连接酶活性和/或聚合酶活性。由此，在某些实施方案中，用于实施本发明的寡核苷酸缺少5'或3'磷酸基团。

[0109] 再更优选地，部分iii)的该第二寡核苷酸包含3'脱氧胸苷三磷酸“T尾”以提供用于连接至“A加尾”DNA片段的基底，理想情况下具有硫代磷酸酯连接，由此赋予对核酸外切酶活性的抗性。

[0110] 在本发明的一个优选实施方案中，部分i)中的所述连接原位发生或使用细胞或组织悬浮液在体外发生，并因此发生在完整细胞中。再更优选地，通过化学、电子、机械或生理透化细胞或组织由此待连接的所述寡核苷酸可以接近DSB位点，并经由其末端结合特征(例如磷酸)，其连接至DSB位点来促进该步骤。优选地，通过在裂解缓冲液中温育来透化细胞。再更优选地，所述DSB位点在与所述寡核苷酸连接之前进行精氨酸尾修复。

[0111] 如将理解的那样，在进一步加工之前将所述第一对寡核苷酸/衔接子连接至DSB确保了鉴定的DSB是真实事件而不是后续加工步骤的结果(并因此代表加工的假象)。

[0112] 在本发明的再进一步优选的实施方案中，部分i)还包括在进行后续步骤之前使用任何常规手段(诸如提取缓冲液等)从所述细胞中提取gDNA。

[0113] 在本发明的一个优选实施方案中，步骤ii)包含通过任何本领域已知的手段(诸如超声处理或标签化(tagmentation))将该gDNA片段化为更小的片段。

[0114] 在本发明的再进一步优选的实施方案中，所述方法进一步包含在部分ii)和/或部分iv)之后除去其尺寸小于大约100bp，更优选小于大约150bp的片段并保留其尺寸大于大

约150bp的片段的任选步骤。本领域技术人员将理解的是，该步骤有利地除去可能已经形成的否则随后会促成序列假象的任何寡核苷酸链/二聚体。理想情况下，这可以使用常规手段(诸如使用Bioruptor超声波仪)进行，以及使用SPRI微珠(GC Biotech,CNGS-0005)进行尺寸选择以除去<150bp的片段。或者选择优选范围内的片段，诸如但不限于150-1000bp，理想情况下为200-800bp，更理想情况下为250-750bp以及最优选为300-500bp。

[0115] 在本发明的再进一步优选的实施方案中，部分v)的所述分离涉及使用由部分i)的寡核苷酸提供的所述结合序列结合配偶体，并因此将部分iv)的A组链与任何其它链分离。通常，由部分i)的寡核苷酸提供的所述结合序列的互补结合链锚定在基底上，并且所述核酸单链流经或流过锚定的互补结合链。

[0116] 在本发明的再进一步优选的方法中，部分vi)涉及桥式扩增，其中在部分v)下分离的单链在已在部分i)的第一寡核苷酸的结合序列和部分iii)的第二寡核苷酸的另外的序列的寡核苷酸/结合位点锚定于其上的基底上进行克隆扩增。以这种方式，A组片段的单链可以两端都结合在所述基底上以促进桥式扩增。再更优选地，所述测序可以通过采用使用标记的核苷酸的合成测序进行，所述标记的核苷酸每个发射特征信号，当序列延伸时读取该特征信号以提供序列信息的读出。通常，在平行过程中对许多链进行测序。如果愿意，索引序列可以与DSB分开测序，由此在对DSB进行测序之前提供核酸来源的指示。备选地，可以对两个索引进行测序并从而一起读取。

[0117] 在一个特定实施方案中，部分i)的第一对寡核苷酸包含含有根据TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的序列的第一寡核苷酸，和含有根据AATGATACGGCGACCACCGA(SEQ ID NO:34)的序列的第二寡核苷酸。

[0118] 在可与前述段落的主题内容结合的另一个实施方案中，第二对寡核苷酸包含不包含序列ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的超过5、10、15或20个碱基的序列，或不包含序列ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的全部24个碱基的序列的第一寡核苷酸，和包含根据CAAGCAGAAGACGGCATACGAGAT(SEQ ID NO:32)的序列的第二寡核苷酸。

[0119] 在一个特定实施方案中，部分i)的第一对寡核苷酸包含含有根据ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的序列的第一寡核苷酸，和含有根据CAAGCAGAAGACGGCATACGAGAT(SEQ ID NO:32)的序列的第二寡核苷酸。

[0120] 在可与前述段落的主题内容结合的另一个实施方案中，第二对寡核苷酸可包含不包含序列TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的超过5、10或15个碱基的序列，或不包含序列TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的全部20个碱基的第一寡核苷酸，和包含根据AATGATACGGCGACCACCGA(SEQ ID NO:34)的序列的第二寡核苷酸。

[0121] SEQ ID NO:30、31、32和34可包含1至12、1至10、1至8、1至5、1至3、2或1个修饰，诸如取代、缺失或插入。在一个实施方案中，修饰是取代。

[0122] 在本发明的进一步优选的实施方案中，部分i)的所述第一对寡核苷酸包含具有SEQ ID NO.1的第一寡核苷酸和具有SEQ ID NO.2的第二寡核苷酸或与其共享至少80％同一性或同源性，且更优选地，以提高的优先顺序，与其共享81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99或100％同一性或同源性的寡核苷酸。根据SEQ ID NO:1和SEQ ID NO:2的寡核苷酸可包含1至12、1至10、1至8、1至5、1至3、2或1个修饰，诸如取代、缺失或插入。在一个实施方案中，修饰是取代。本文中所用的同源性可以称为相似性。

[0123] 在本发明的进一步优选的实施方案中，部分i)的所述第一对寡核苷酸包含序列GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT[索引]TCGGTGGTCGCCGTATCATTC的第一寡核苷酸，或者所述第一寡核苷酸包含1至7、1至6、1至5、1至4、1至3、2或1个修饰，诸如取代、缺失或插入。部分i)的第一对寡核苷酸可进一步包含序列AATGATACGGCGACCACCGA[索引]ACACTCTTTCCCTACACGACGCTCTTCCGATCT的第二寡核苷酸，或者所述第二寡核苷酸包含1至7、1至6、1至5、1至4、1至3、2或1个修饰，诸如取代、缺失或插入。

[0124] 在本发明的进一步优选的实施方案中，部分i)的所述第一对寡核苷酸包含序列GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT(SEQ ID NO:37)、索引、和TCGGTGGTCGCCGTATCATTC(SEQ ID NO:38)的第一寡核苷酸。部分i)的第一对寡核苷酸可进一步包含序列AATGATACGGCGACCACCGA(SEQ ID NO:34)、索引、ACACTCTTTCCCTACACGACGCTCTTCCGATCT(SEQ ID NO:39)的第二寡核苷酸。索引可以是任何碱基(n)，并且长度可以为例如5至15或6至10个碱基对。SEQ ID NO:34、37、38和39可包含1至7、1至6、1至5、1至4、1至3、2或1个修饰，诸如取代、缺失或插入。在一个实施方案中，修饰是取代。

[0125] 在本发明的进一步的实施方案中，部分i)的所述第一对寡核苷酸包含含有SEQ ID NO.3、索引、和序列ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的第一寡核苷酸。部分i)的第一对寡核苷酸可进一步包含第二寡核苷酸，所述第二寡核苷酸以5’至3’的顺序包含CAAGCAGAAGACGGCATACGAGAT(SEQ ID NO:32)的序列、索引、和序列GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT(SEQ ID NO:33)。索引可以是任何碱基(n)，并且长度可以为例如5至15或6至10个碱基对。SEQ ID NO:3、30、32和33可包含1至7、1至6、1至5、1至4、1至3、2或1个修饰，诸如取代、缺失或插入。在一个实施方案中，修饰是取代。

[0126] 在本发明的进一步优选的实施方案中，部分iii)的所述第二对寡核苷酸包含具有SEQ ID NO.3的第一寡核苷酸和序列CAAGCAGAAGACGGCATACGAGAT[索引]GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT的第二寡核苷酸或与其共享至少80％同一性或同源性，且更优选地，以提高的优先顺序，与其共享81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99或100％同一性或同源性的寡核苷酸。根据SEQ ID NO:3的寡核苷酸可包含1至7、1至6、1至5、1至4、1至3、2或1个修饰，诸如取代、缺失或插入。根据CAAGCAGAAGACGGCATACGAGAT[索引]GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT的寡核苷酸可包含1至12、1至10、1至8、1至5、1至3、2或1个修饰，诸如取代、缺失或插入。在一个实施方案中，修饰是取代。

[0127] 在本发明的进一步优选的实施方案中，部分iii)的所述第二对寡核苷酸包含具有SEQ ID NO.3的第一寡核苷酸，和以5’至3’的顺序具有CAAGCAGAAGACGGCATACGAGAT(SEQ ID NO:32)的序列、索引、和序列GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT(SEQ ID NO:33)的第二寡核苷酸。索引可以是任何碱基(n)，并且长度可以为例如5至15或6至10个碱基对。寡核苷酸可包含与SEQ ID NO:3、32或33的任一种具有80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99或100％同一性或同源性的序列。SEQ ID NO:3、32或33可包含1至7、1至6、1至5、1至4、1至3、2或1个修饰，诸如取代、缺失或插入。在一个实施方案中，修饰是取代。

[0128] 在本发明的再进一步优选的实施方案中，部分iii)的所述第二对寡核苷酸的所述第二寡核苷酸包含以下序列SEQ ID NO:4-28、30-34、37-39或41-42的任一种或与其共享至少80％同一性或同源性，且更优选地，以提高的优先顺序，与其共享81、82、83、84、85、86、

87、88、89、90、91、92、93、94、95、96、97、98、99或100％同一性或同源性的寡核苷酸。SEQ ID NO:4-28、30-34、37-39或41-42可包含1至12、1至10、1至8、1至5、1至3、2或1个修饰，诸如取代、缺失或插入。在一个实施方案中，修饰是取代。

[0129] 在一个优选实施方案中，所述样品是哺乳动物样品，理想情况下为人样品。

[0130] 根据本发明的另一方面，提供了理想情况下适于鉴定gDNA样品中的DNA双链断裂(DSB)的成套试剂盒，其包含：

i)第一对寡核苷酸，所述第一对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至所述DSB的第一链的5’结合特征、第一测序引物可与之结合的杂交位点(RD1 SP)和用于从DSB池中分离所述DSB的结合序列；和第二寡核苷酸，其与该第一对的所述第一寡核苷酸互补，并包含用于与所述DSB的第二链结合的3’结合特征；并且其中所述寡核苷酸的一种或两种分别包含3’和/或5’保护性特征；和

ii)第二对寡核苷酸，所述第二对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至所述DSB的第一链的5’结合特征，和第二测序引物可与之结合的杂交位点(RD2 SP)；和第二较长的寡核苷酸，其与该第二对的所述第一寡核苷酸部分互补，并包含用于与所述DSB的第二链结合的3’结合特征、与所述杂交位点互补的序列、和另外的序列，其任选为用于实现桥式扩增的结合序列；并且其中所述寡核苷酸的一种或两种分别包含3’和/或5’保护性特征。

[0131] 在本发明的优选试剂盒中，所述试剂盒进一步包含至少一种引物，其与第一和/或第二杂交位点结合以便用于测序目的。

[0132] 更优选地，所述试剂盒进一步包含分别用于将核酸片段化和/或变性成片段和/或单链的片段化剂和/或变性剂。

[0133] 在本发明的另一方面，提供了适于鉴定gDNA样品中的DNA双链断裂(DSB)的成套试剂盒，其包含：

i)第一对寡核苷酸，所述第一对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至所述DSB的第一链的5’结合特征和用于从DSB池中分离所述DSB的结合序列；和第二寡核苷酸，其与该第一对的所述第一寡核苷酸互补；并且其中所述寡核苷酸的一种或两种分别包含3’和/或5’保护性特征；和

ii)第二对寡核苷酸，所述第二对寡核苷酸的第一寡核苷酸与第二对的第二寡核苷酸部分互补；和第二寡核苷酸，其包含用于与所述DSB的第二链结合的3’结合特征、和用于实现桥式扩增的结合序列；并且其中所述寡核苷酸的一种或两种分别包含3’和/或5’保护性特征。

[0134] 在本发明的一个方面，提供了适于用于鉴定gDNA样品中的DSB的样品制备的试剂盒，其包含：

i)第一对寡核苷酸，所述第一对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至双链核酸的一条链，并包含根据TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的序列的5’结合特征；和第二寡核苷酸，其与第一对的所述第一寡核苷酸互补；并且其中所述寡核苷酸的一种或两种分别包含3’和/或5’保护性特征；和

ii)第二对寡核苷酸，所述第二对寡核苷酸的第一寡核苷酸不包含序列ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的超过5、10、15或20个碱基的序列，或不包含

序列ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的全部24个碱基;和第二寡核苷酸,其包含使得所述寡核苷酸能够连接至双链核酸的一条链,并包含根据CAAGCAGAAGACGGCATACGAGAT(SEQ ID NO:32)的序列的3'结合特征;并且其中所述寡核苷酸的一种或两种分别包含3'和/或5'保护性特征。

[0135] 在本发明的另一方面,提供了适于用于鉴定gDNA样品中的DSB的样品制备的试剂盒,其包含:

i)第一对寡核苷酸,所述第一对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至双链核酸的一条链,并包含根据ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的序列的5'结合特征;和第二寡核苷酸,其与第一对的所述第一寡核苷酸互补并任选包含3'结合特征;并且其中所述寡核苷酸的一种或两种分别包含3'和/或5'保护性特征;和

ii)第二对寡核苷酸,所述第二对寡核苷酸的第一寡核苷酸不包含序列TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的超过5、10或15个碱基的序列,或不包含序列TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的全部20个碱基,并任选包含5'结合特征;和第二寡核苷酸,其包含使得所述寡核苷酸能够连接至双链核酸的一条链,并包含根据AATGATACGGCGACCACCGA(SEQ ID NO:34)的序列的3'结合特征;并且其中所述寡核苷酸的一种或两种分别包含3'和/或5'保护性特征。

[0136] SEQ ID NO:30、31、32和34可包含1至12、1至10、1至8、1至5、1至3、2或1个修饰,诸如取代、缺失或插入。在一个实施方案中,修饰是取代。

[0137] 部分i)的第二寡核苷酸可包含3'结合特征,以及部分ii)的第一寡核苷酸可包含5'结合特征。第一和第二对寡核苷酸的第一寡核苷酸可包括测序引物可与之结合的杂交位点(RD1 SP)。

[0138] 根据本发明的另一方面,提供了理想情况下用于鉴定核酸样品,诸如gDNA样品中的DNA双链断裂(DSB)的双链衔接子,其包含:

第一寡核苷酸链,其包含使得所述寡核苷酸能够连接至所述DSB的第一链的5'结合特征、测序引物可与之结合的杂交位点(RD1 SP)和用于从DSB池中分离所述DSB的结合序列;和第二寡核苷酸链,其与所述第一寡核苷酸互补并包含用于与所述DSB的第二链结合的3'结合特征;并且其中所述寡核苷酸的一种或两种分别包含3'和/或5'保护性特征。

[0139] 根据本发明的另一方面,提供了理想情况下用于鉴定核酸样品,诸如gDNA样品中的DNA双链断裂(DSB)的双链衔接子,其包含:

第一寡核苷酸链,其包含使得所述寡核苷酸能够连接至所述DSB的第一链的5'结合特征和测序引物可与之结合的杂交位点(RD2 SP);和第二较长的寡核苷酸链,其与所述第一寡核苷酸部分互补,并包含用于与所述DSB的第二链结合的3'结合特征、与所述杂交位点互补的序列、和另外的序列,其任选为用于实现桥式扩增的结合序列;并且其中所述寡核苷酸的一种或两种分别包含3'和/或5'保护性特征。

[0140] 根据本发明的另一方面,提供了适于鉴定核酸样品,诸如gDNA样品中的DNA双链断裂(DSB)的双链衔接子,其包含:

第一寡核苷酸链,其不包含序列ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的超过5、10、15或20个碱基的序列,或不包含序列ATCTCGTATGCCGTCTTCTGCTTG(SEQ ID NO:30)的全部24个碱基;和第二寡核苷酸,其包含使得所述寡核苷酸能够连接至双链核酸的一条

链并包含根据CAAGCAGAAGACGGCATACGAGAT(SEQ ID NO:32)的序列的3'结合特征;并且其中所述寡核苷酸的一种或两种分别包含3'和/或5'保护性特征。第一寡核苷酸可包含5'结合特征。

[0141] 根据本发明的另一方面,提供了适于鉴定核酸样品,诸如gDNA样品中的DNA双链断裂(DSB)的双链衔接子,其包含:

第一寡核苷酸链,其不包含序列TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的超过5、10或15个碱基的序列,或不包含序列TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的全部20个碱基;和第二寡核苷酸,其包含使得所述寡核苷酸能够连接至双链核酸的一条链,并包含根据AATGATACGGCGACCACCGA(SEQ ID NO:34)的序列的3'结合特征;并且其中所述寡核苷酸的一种或两种分别包含3'和/或5'保护性特征。第一寡核苷酸可包含5'结合特征。

[0142] 在本发明的一个方面,提供了用于鉴定核酸样品中的DSB的样品制备方法,其中制备包括修饰DSB相关核酸以适于与包含固定引物的基底结合,该方法包括:

a)提供包含多个核酸的样品;

b)在有助于连接的条件下将多个核酸暴露于第一衔接子,其中第一衔接子包含能够连接至DSB的一条链的3'末端的寡核苷酸且所述寡核苷酸包含能够通过杂交与固定于基底的引物结合的序列;

c)片段化多个核酸;和

d)在有助于连接的条件下将多个核酸暴露于第二衔接子,其中第二衔接子包含能够在片段化诱导的断裂处连接至一条链的5'末端但不能连接至第一衔接子的寡核苷酸,且所述寡核苷酸不包含能够通过杂交与固定于基底的引物结合的序列。能够在片段化诱导的断裂处连接至一条链的5'末端的第二衔接子的寡核苷酸可包含与第二引物的区域相同的序列。

[0143] 在一些实施方案中,包含多个核酸的样品中的核酸在步骤b)至d)的过程中是双链的。在另一些实施方案中,核酸在步骤b)至d)的过程中可以是单链的。在再进一步的实施方案中,核酸对于一些步骤可以是双链的,并对于另一些是单链的(例如参见图16)。在涉及双链衔接子连接至单链核酸的实施方案中,衔接子可包含"夹板寡核苷酸(splint oligo)"。夹板寡核苷酸可包含随机核苷酸,诸如6-8个随机核苷酸,并位于不连接至核酸样品的衔接子的寡核苷酸的3'端处。夹板寡核苷酸可有助于连接过程。

[0144] 在一个实施方案中,提供了用于鉴定核酸样品中的DSB的样品制备方法,其中制备包括修饰DSB相关核酸以适于包含第一固定引物的基底结合并适于包括使用所述第一固定引物和第二引物的扩增,该方法包括:

a)提供包含多个核酸的样品;

b)在有助于连接的条件下将多个核酸暴露于第一衔接子,其中第一衔接子包含能够连接至DSB的一条链的3'末端的寡核苷酸,并且所述寡核苷酸包含能够通过杂交与固定于基底的引物结合的序列;

c)片段化多个核酸;和

d)在有助于连接的条件下将多个核酸暴露于第二衔接子,其中第二衔接子包含能够在片段化诱导的断裂处连接至一条链的5'末端但不能连接至第一衔接子的寡核苷酸,并且所述寡核苷酸包含与第二引物的区域相同的序列。

[0145] 适于包括使用第一固定引物和第二引物的扩增的核酸样品是能够通过与第一固定引物序列相同的引物和与第二引物序列相同的引物来扩增的样品。可以在溶液中确定扩增的适合性。

[0146] 在一些实施方案中，第二引物并未固定于基底。例如，基底可以是微珠，并且扩增可以经由微珠乳液扩增来进行。

[0147] 在另一些实施方案中，第二引物固定于基底。例如，基底可以是流动池，并且扩增可以经由桥式扩增来进行。

[0148] 在一些实施方案中，衔接子是单链寡核苷酸。

[0149] 在另一些实施方案中，衔接子包含至少部分互补的第一和第二寡核苷酸。在此类实施方案中，第一衔接子对能够连接至DSB的一条链的至少3’末端，且第一衔接子对包含至少部分互补的第一和第二寡核苷酸，其中第一寡核苷酸可连接至3’末端并包含能够通过杂交与固定于基底的引物结合的序列。此外，在此类实施方案中，第二衔接子对能够在片段化诱导的断裂处连接至一条链的至少5’末端，但不能连接至第一衔接子对的第一寡核苷酸，其中第二衔接子包含第一和第二部分互补的寡核苷酸，其中第一寡核苷酸可连接至5’末端并包含与第二引物的区域相同的序列，并且第二寡核苷酸不包含与所述与第二引物的区域相同的序列互补的序列。

[0150] 由此，在一个实施方案中，提供了用于鉴定核酸样品中的DNA DSB的样品制备方法，其中制备包括修饰DSB相关核酸以适于与包含第一固定引物的基底结合并适于包括使用所述第一固定引物和第二引物的扩增，该方法包括，该方法包括：

a)提供包含多个核酸的样品；

b)在有助于连接的条件下将多个核酸暴露于第一衔接子对，其中第一衔接子对能够连接至DSB的一条链的至少3’末端，并且其中第一衔接子对包含至少部分互补的第一和第二寡核苷酸，并且第一寡核苷酸可连接至3’末端并包含能够通过杂交与固定于基底的引物结合的序列；

c)片段化多个核酸；和

d)在有助于连接的条件下将多个核酸暴露于第二衔接子对，其中第二衔接子对能够在片段化诱导的断裂处连接至一条链的至少5’末端，但不能连接至第一衔接子对的第一寡核苷酸，其中第二衔接子包含第一和第二部分互补的寡核苷酸，其中第一寡核苷酸可连接至5’末端并包含与第二引物的区域相同的序列，并且第二寡核苷酸不包含与所述与第二引物的区域相同的序列互补的序列。

[0151] 在一些实施方案中，基底包含第一固定引物和第二固定引物。在一些实施方案中，固定引物可以适于在桥式扩增过程中充当引物。

[0152] 由此，在一个实施方案中，提供了用于鉴定核酸样品中的DNA DSB的样品制备方法，其中制备包括修饰DSB相关核酸以适于与包含第一和第二固定引物的基底结合，该方法包括：

a)提供包含多个核酸的样品；

b)在有助于连接的条件下将多个核酸暴露于第一衔接子，其中第一衔接子包含能够连接至DSB的一条链的3’末端的寡核苷酸且所述寡核苷酸包含能够通过杂交与第一固定引物结合的序列；

c)片段化多个核酸;和

d)在有助于连接的条件下将多个核酸暴露于第二衔接子,其中第二衔接子包含能够在片段化诱导的断裂处连接至一条链的5'末端但不能连接至第一衔接子的寡核苷酸,且所述寡核苷酸包含与第二固定引物的区域相同的序列。

[0153] 在另一些实施方案中,衔接子是包含至少部分互补的第一和第二寡核苷酸的衔接子对。由此,提供了用于鉴定核酸样品中的DNA DSB的样品制备方法,其中制备包括修饰DSB相关核酸以适于与包含第一和第二固定引物的基底结合,该方法包括:

a)提供包含多个核酸的样品;

b)在有助于连接的条件下将多个核酸暴露于第一衔接子对,其中第一衔接子对能够连接至DSB的一条链的至少3'末端,并且其中第一衔接子对包含至少部分互补的第一和第二寡核苷酸,并且其中第一寡核苷酸可连接至3'末端并包含能够通过杂交与第一固定引物结合的序列;

c)片段化多个核酸;和

d)在有助于连接的条件下将多个核酸暴露于第二衔接子对,其中第二衔接子对能够在片段化诱导的断裂处连接至一条链的至少5'末端但不能连接至第一衔接子对的第一寡核苷酸,其中第二衔接子包含第一和第二部分互补的寡核苷酸,并且可连接至5'末端的寡核苷酸包含与第二固定引物的区域相同的序列,并且其中另一个寡核苷酸不包含与所述与第二固定引物的区域相同的序列互补的序列。

[0154] 如本文中所述,在一些实施方案中,样品中的核酸在步骤a)至d)的过程中可以保持为双链分子。由此,在特定实施方案中,方法可进一步包括:

使多个双链核酸变性以形成多个单链核酸。这在一些实施方案中可以是"步骤e)"。

[0155] 在一个特定实施方案中,方法可进一步包括:

在适于固定引物与互补核酸杂交的条件下,使多个单链核酸与包含固定引物的基底接触。这在一些实施方案中可以是"步骤f)"。

[0156] 结合本文中公开的方法的步骤i)公开的特征也适用于步骤b)。结合本文中公开的方法的步骤ii)公开的特征也适用于步骤c)。结合本文中公开的方法的步骤iii)公开的特征也适用于步骤d)。结合本文中公开的方法的步骤iv)公开的特征也适用于步骤e)。结合本文中公开的方法的步骤v)公开的特征也适用于步骤f)。

[0157] 基底可以是固体表面,诸如流动池、微珠、载玻片或膜的表面。特别地,基底可以是流动池。基底可以是图案化或未图案化的流动池。基底可包含玻璃、石英、二氧化硅、金属、陶瓷或塑料。基底表面可包含聚丙烯酰胺基质或涂层。

[0158] 本文中使用的术语"流动池"意在具有本领域,特别是在通过合成测序的领域中的普通含义。示例性的流动池包括但不限于在核酸测序装置中使用的那些,诸如由Illumina,Inc.(San Diego,Calif.)商业化的Genome Analyzer®、MiSeq®、NextSeq®、HiSeq®或NovaSeq®平台的流动池;或由Life Technologies(Carlsbad,Calif.)商业化的SOLiD™或Ion Torrent™测序平台的流动池。示例性的流动池及其制造和使用方法还描述在例如WO2014/142841A1;美国专利申请公开号2010/0111768A1和美国专利号8,951,781中。

[0159] 基底可包含固定引物,例如两种类型的引物,其可以一起充当桥式扩增的正向和

反向引物。固定于基底意味着即使在将使双链核酸变性的条件下引物也结合在基底上。例如，引物可以共价结合在基底上。将引物定向，使得5’端在固定点的近端，而3’端在固定点的远端。此类布置在本领域中是标准的。

[0160] 步骤可以按以下顺序进行：步骤c)、步骤d)和随后步骤b)。与用于检测预先存在的DSB的实施方案相反，该顺序特别与其中DSB在样品中潜在地诱导的实施方案相关。参照以罗马数字限定的本发明的特征步骤的陈述，步骤可以按以下顺序进行：步骤ii)、步骤iii)和随后步骤i)。在此类实施方案中，可在连接第二衔接子之后和在连接第一衔接子之前诱导DSB。例如，参见图13。DSB的诱导可包括将样品暴露于能够引起或怀疑能够引起DSB的条件。

[0161] 在其中在步骤b)之前进行步骤d)的实施方案中，步骤d)的衔接子可以包含3’和/或5’保护性特征以防止步骤b)的衔接子连接至步骤d)的那些。保护性特征可以使第一衔接子不能连接至第二衔接子。在此类实施方案中，第二衔接子保持不能连接至第一衔接子，因为在第一衔接子存在之前发生连接。

[0162] 在一些实施方案中，步骤以步骤b)、步骤c)和步骤d)的顺序进行，其中步骤b)在细胞中进行或原位进行。在另一些实施方案中，步骤以步骤c)、步骤d)和随后步骤b)的顺序进行，其中步骤c)在分离核酸样品后体外进行。

[0163] 由此，在一个实施方案中，提供了用于鉴定核酸样品中的DSB的样品制备方法，其中制备包括修饰DSB相关核酸以适于与包含固定引物的基底结合，该方法依次包括：

a)提供包含多个核酸的样品；

c)片段化多个核酸；

d)在有助于连接的条件下将多个核酸暴露于第二衔接子，其中第二衔接子包含能够在片段化诱导的断裂处连接至一条链的5’末端的寡核苷酸且所述寡核苷酸不包含能够通过杂交与固定于基底的引物结合的序列；任选包含与第二引物的区域相同的序列；和

b)在有助于连接的条件下将多个核酸暴露于第一衔接子，其中第一衔接子包含能够连接至DSB的一条链的3’末端但不能连接至第二衔接子的寡核苷酸，且所述寡核苷酸包含能够通过杂交与固定于基底的引物结合的序列。

[0164] 在步骤a)中，包含多个双链核酸的样品可以是能够包含DSB的任何DNA样品，诸如gDNA。

[0165] 对于任何本文中公开的方法，样品可含有DSB或可以以可能或确实引入DSB的方式处理过。

[0166] 例如，本文中公开的方法可用于检测仅限于单链的DNA损伤或变化。由此，本文中公开的方法可以用于检测和/或定量核酸样品中的目标特征。例如，双链DNA的一条链中的损伤可以酶促转化为DSB，其随后可以通过本文中公开的方法来检测。

[0167] 在一些实例中，损伤可以是单链断裂。在另一些实例中，损伤是核酸样品中的碱基变化。例如，本文中公开的方法可用于检测CRISPR/Cas诱导的碱基编辑，诸如胞嘧啶碱基编辑器或腺苷碱基编辑器。此类编辑可以转化为DSB并如本文中公开的那样检测。图17和图18示出了示例性实施方案。

[0168] 由此，在一个实施方案中，方法包括提供包含多个核酸的样品，所述核酸包含或怀疑包含目标特征或损伤；并将样品暴露于能够将目标特征或损伤转化为DSB的条件。

[0169] 对于此类实施方案，方法可以按以下顺序：将损伤转化为DSB，b)，c)，和d)；或可以按以下顺序：c)，d)，将损伤转化为DSB，和b)。

[0170] 在另一些实施方案中，样品已经用核酸酶处理，所述核酸酶诸如类转录激活因子效应物核酸酶(TALEN)、CRISPR/Cas核酸内切酶、锌指核酸酶、大范围核酸酶或任何限制性核酸内切酶，并且方法可以用于检测脱靶效应。样品可以用药剂(诸如潜在的治疗剂)处理以确定所述药剂是否能够引起DSB或脱靶DSB。

[0171] 本文中公开的方法可用于鉴定蛋白与核酸结合的位点。例如，样品可以与对目标蛋白特异性的蛋白结合剂(诸如抗体)接触，其中目标蛋白可能与DNA结合。DNA可以是已经与目标蛋白接触的样品。蛋白结合剂可以直接或间接与核酸酶缔合，因此在结合的目标蛋白的任何位点处形成DSB。随后可以通过本文中公开的方法来检测任何DSB。方法可以是核酸酶靶向切割和释放(CUT&RUN)方法。由此，方法可包括：步骤a)；使样品与目标蛋白接触；使样品与能够直接或间接与目标蛋白缔合的核酸酶接触以便在蛋白结合的核酸中形成DSB；步骤b)，步骤c)，和随后步骤d)。这种顺序特别可用于其中在细胞中进行方法直到步骤b)的实施方案。备选地，方法可包括：步骤a)；步骤c)；步骤d)；使样品与目标蛋白接触；使样品与能够直接或间接与目标蛋白缔合的核酸酶接触以便在蛋白结合的核酸中形成DSB；和随后步骤b)。这种顺序特别可用于体外实施方案。

[0172] 在另一些实施方案中，可以在已知位点或靶位点中有意诱导DSB。例如，靶向核酸酶，诸如CRISPRR/Cas或TALEN，可用于在目标序列中诱导DSB，使得本发明的方法可用于分离所述目标序列以便进一步分析。目标序列可以是特定基因、全外显子组或基因组中的特定基因座。由此，在一个实施方案中，方法包括使样品与能够在目标序列中靶向诱导DSB的核酸酶接触。

[0173] 本发明的方法可用于检测病毒或细菌插入事件或由病毒或细菌插入事件引起的DNA损伤。这些事件可以借助与将插入基因组中的细菌和病毒相关的独特遗传序列来测量。本发明的方法因此可以揭示这些外来DNA插入(其可经由DSB中间结构发生)的位点。

[0174] 在其它实施方案中，本文中公开的方法可用于评估与治疗剂(诸如基因疗法)相关的风险。例如，本文中公开的方法可应用于取自患者的样品，其中该样品已经用治疗剂处理，以确定所述患者的脱靶效应的风险、性质或频率。由此，本文中公开的方法通过提供由感兴趣的试剂诱导的DSB的患者特异性模式及其频率，可用于个体化医疗。由此，在一个实施方案中，方法可包括从受试者获得样品并将所述样品暴露于药剂，诸如治疗剂。方法可包括在暴露后确定样品中任何损伤或DSB的性质和/或频率。本发明的步骤的顺序可以是本文中描述的任一种。

[0175] 本文中公开的方法可用于检测样品被任何能够引起DSB的试剂的污染。

[0176] 本文中公开的方法可用于测量人工组装或合成的基因组的稳定性。

[0177] 本文中公开的方法可用于遗传毒理学中的下一代风险评估(NGRA)。NGRA定义为一种暴露导向的、假设驱动的风险评估方法，其集成了新技术方法(NAM)以确保安全而无需使用动物试验。DSB是遗传毒性暴露的直接量度，并可以通过本发明的方法来定量。因此，本发明的方法可用于需要定量遗传毒性暴露的风险评估。

[0178] 总之，怀疑含有DSB的核酸的样品可由于天然发生的DNA损伤、由于用潜在的DSB引发剂处理、由于目标位点处DSB的有意诱导或出于任何其它原因而含有所述DSB。

[0179] 在步骤b)中,条件使得第一衔接子能够连接至DSB。第一衔接子包含能够连接至DSB的一条链的3’末端的寡核苷酸,并还可包含能够连接至DSB的一条链的5’末端的另一寡核苷酸。由此,第一衔接子共价连接至DSB。连接可以是直接或间接的。例如,可以连接至在DSB处引入的另外的核苷酸。备选地,可以将寡核苷酸或寡核苷酸对连接至DSB,并可将第一衔接子连接至所述一个或多个寡核苷酸。连接方法可以是本文中公开的任何连接方法。第一衔接子对的两个寡核苷酸可以是完全互补的。

[0180] 第一衔接子对的第一和/或第二寡核苷酸可包含3’和/或5’保护性特征。这些保护性特征可以是本文中公开的任何保护性特征,特别是结合就本文中公开的方法的步骤i)论述的第一对寡核苷酸公开的任何保护性特征。

[0181] 第一衔接子对的第一和/或第二寡核苷酸可包含结合就本文中公开的方法的步骤i)论述的第一对寡核苷酸公开的任何特征。特别地,第一衔接子对可包括或不包括结合就步骤i)论述的第一对寡核苷酸公开的序列。

[0182] 第一衔接子包括可连接至3’末端并包含能够通过杂交与固定引物结合的序列的寡核苷酸。这意味着,当连接的第一衔接子对变性为单链时,一条链与固定于基底的引物互补。衔接子包括足够长度的互补序列,以实现在洗涤或聚合步骤过程中不会释放的结合。互补区的长度可以为5、10、15、20、21、24或更多个碱基。备选地,互补区可包括5、10、15、20、21、24或更多个互补碱基。

[0183] 步骤c)可包含本文中所述的任何片段化方法。

[0184] 在步骤d)中,条件使得核酸能够连接至第二衔接子。第二衔接子包含能够在片段化位点处连接至5’末端的寡核苷酸,并可包含能够在片段化位点处连接至3’末端的寡核苷酸。连接方法可以是本文中公开的任何方法。

[0185] 第二衔接子不能连接至第一衔接子。在一个特定实施方案中,这种防止是由于在第一衔接子上包含3’保护性特征。在相关实施方案中,这种防止是由于在第一衔接子对上包含5’和/或3’保护性特征。特别地,第一衔接子的第一寡核苷酸(即能够与3’末端结合的寡核苷酸)可包含3’保护性特征(诸如间隔子C3 3’链终止子),以防止衔接子连接至该链(参见图12)。

[0186] 第二衔接子对的第一和/或第二寡核苷酸可包含3’和/或5’保护性特征。这些保护性特征可以是本文中公开的任何保护性特征,特别是结合就本文中公开的方法的步骤iii)论述的第二对寡核苷酸公开的任何保护性特征。

[0187] 在其中步骤d)在步骤b)之前进行的实际方案中,第二衔接子对的第一和/或第二寡核苷酸可包含3’和/或5’保护性特征,并且至少一个保护性特征可防止步骤b)的衔接子连接至步骤d)的那些。例如,能够结合至3’末端的第二衔接子的寡核苷酸可包含3’保护性特征,诸如间隔子C3 3’链终止子。

[0188] 第二衔接子对的第一和/或第二寡核苷酸可包含结合就本文中公开的方法的步骤iii)论述的第二对寡核苷酸公开的任何特征。特别地,第二衔接子对可包括或不包括结合就步骤iii)论述的第二对寡核苷酸公开的序列。

[0189] 第二衔接子对可通过标签化引入。由此,第二衔接子的片段化和连接可以是同一步骤。例如,可以组合步骤c)和d)。

[0190] 第二衔接子包括可连接至5’末端的寡核苷酸且所述寡核苷酸包含与固定至基底

的引物的至少一部分相同的序列。因此，当该寡核苷酸在聚合过程中充当模板时，新的链将包括能够与所述引物杂交的序列。由此，衔接子序列的相关区域和引物的相关区域的长度应足以满足该功能。相同区域的长度可以为5、10、15、20、21、24或更多个碱基。第二衔接子中的另一寡核苷酸不包括与该所谓的相同区域互补的序列。由此，第二衔接子不能经由杂交与基底结合。

[0191] 在步骤f)之后，方法可进一步包括在适于固定引物延伸的条件下使任何杂交的核酸与聚合酶接触以合成核酸，所述核酸是与杂交的核酸互补的核苷酸的链。随后可以扩增新形成的核酸。在一些实施方案中，用于扩增的引物也固定到基底上，并例如可以适于桥式扩增。该过程是本领域已知的，并形成核酸的克隆簇。在另一些实施方案中，例如对于其中基底为微珠的一些实施方案，用于扩增(implication)的引物可以在溶液中。扩增的核酸可以随后以通常方式进行测序，例如通过边合成边测序。衔接子可包含用于测序引物的结合的位点以协助该过程。衔接子还可包含本文中公开的索引。

[0192] 由此，在一个实施方案中，方法可进一步包括：

g)获得在步骤f)中与基底杂交的任何核酸的序列信息。

[0193] 包括步骤g)的方法可以被称为用于鉴定核酸样品中的DNA DSB的方法。

[0194] 结合本文中公开的方法的步骤vi)公开的特征也适用于步骤g)。

[0195] 在包含含有AATGATACGGCGACCACCGA(SEQ ID NO:34)的寡核苷酸或其变体的产品或方法的任何实施方案中，寡核苷酸可包含AATGATACGGCGACCACCGAGATCTACAC(SEQ ID NO:41)，或对SEQ ID NO:34所定义的变体。在包含含有SEQ ID NO:31或含有序列TCGGTGGTCGCCGTATCATT(SEQ ID NO:31)的不超过5、10、15或20个碱基的寡核苷酸的产品或方法的任何实施方案中，寡核苷酸可包含SEQ ID NO:38或42，或可包含序列GTGTAGATCTCGGTGGTCGCCGTATCATT(SEQ ID NO:42)的不超过5、10、15、20、25或29个碱基，或序列TCGGTGGTCGCCGTATCATTC(SEQ ID NO:38)的不超过5、10、15或19个碱基。

[0196] 图16公开了其中通过本发明的方法鉴定目标特征(诸如双链DNA的单链中的损伤)的实施方案。目标特征可以是能够被特异性切割的任何特征。例如，将导致双链DNA的单链在目标特征的位点处切割的任何特征。目标特征可以是例如环丁烷嘧啶二聚体(CPD)、8-氧代鸟嘌呤、脱碱基位点及其任何组合。在实施方案中，可以切割包含目标特征的DNA链，并使双链样品变性，以产生衔接子可以连接的3’末端。

[0197] 由此，在本发明的一个方面，提供了用于鉴定核酸样品中的目标特征的样品制备方法，其中制备包括修饰与目标特征相关的核酸以适于与包含固定引物的基底结合，该方法包括：

a)提供包含多个核酸的样品，将多个核酸暴露于能够在目标特征处切割核酸的至少一条链的条件，并使多个核酸变性为单链核酸；

b)在有助于连接的条件下将多个核酸暴露于第一衔接子，其中第一衔接子包含能够连接至切割位点的一条链的3’末端的寡核苷酸且所述寡核苷酸包含能够通过杂交与固定于基底的引物结合的序列；

c)片段化多个核酸；和

d)在有助于连接的条件下将多个核酸暴露于第二衔接子，其中第二衔接子包含能够在片段化诱导的断裂处连接至一条链的5’末端但不能连接至第一衔接子的寡核苷酸，且

所述寡核苷酸不包含能够通过杂交与固定于基底的引物结合的序列。能够在片段化诱导的断裂处连接至一条链的5'末端的第二衔接子的寡核苷酸可包含与第二引物的区域相同的序列。

[0198] 结合鉴定DNA DSB的方法所公开的所有特征也与用于鉴定目标特征的方法相关。特别地,用于鉴定核酸样品中的目标特征的样品制备方法的步骤b)、c)和d)可以与用于鉴定核酸样品中的DNA DSB的样品制备方法中公开的步骤b)、c)和d)相同。衔接子可以与鉴定DSB所公开的相同,并且方法可用于制备适于与本文中公开的任何基底杂交的文库。对于包括双链衔接子连接至单链核酸的实施方案,可以包括本文中公开的夹板寡核苷酸。

[0199] 在一些实施方案中,步骤可以按以下顺序进行:c)、d)、a)和随后b)。这是图16中显示的顺序。由此,在一个实施方案中,提供了用于鉴定核酸样品中的目标特征的样品制备方法,其中制备包括修饰与目标特征相关的核酸以适于与包含固定引物的基底结合,该方法依次包括:

c)将包含多个核酸的样品片段化;和

d)在有助于连接的条件下将多个核酸暴露于第二衔接子,其中第二衔接子包含能够在片段化诱导的断裂处连接至一条链的5'末端的寡核苷酸且所述寡核苷酸不包含能够通过杂交与固定于基底的引物结合的序列;任选包含与第二引物的区域相同的序列;

a)将多个核酸暴露于能够在目标特征处切割核酸的至少一条链的条件,并将多个核酸变性为单链核酸;

b)在有助于连接的条件下将多个核酸暴露于第一衔接子,其中第一衔接子包含能够连接至切割位点的一条链的3'末端但不能连接至第二衔接子的寡核苷酸,且所述寡核苷酸包含能够通过杂交与固定于基底的引物结合的序列。

[0200] 结合鉴定DNA DSB的方法公开的所有下游步骤和特征也与用于鉴定目标特征的方法相关。特别地,方法可包括将样品变性为单链核酸,例如以使任何双链衔接子变性。方法可进一步包括在适于固定引物与互补核酸杂交的条件下使多个核酸与包含固定引物的基底接触。此外,方法可进一步包括获得与基底杂交的任何核酸的序列信息。

[0201] 在另一个实施方案中,提供了用于鉴定核酸样品中的目标特征的样品制备方法,其中制备包括修饰与目标特征相关的核酸以适于与包含固定引物的基底结合,该方法包括:

α)将包含多个核酸的样品暴露于能够在目标特征处切割核酸的至少一条链的条件,并使多个核酸变性为单链核酸;

β)在连接条件下将样品暴露于第一对寡核苷酸,所述第一对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至所述DSB的第一链的5'结合特征、第一测序引物可与之结合的杂交位点(RD1 SP)和用于从DSB池中分离所述DSB的结合序列;和第二寡核苷酸,其与第一对的所述第一寡核苷酸互补;其中所述寡核苷酸的一种或两种包含3'和/或5'保护性特征;

γ)将所述样品的核酸片段化成片段;和

δ)在连接条件下将所述片段暴露于第二对寡核苷酸,所述第二对寡核苷酸的第一寡核苷酸包含第二测序引物可与之结合的杂交位点(RD2 SP);和第二较长的寡核苷酸,其与第二对的所述第一寡核苷酸部分互补并包含用于与所述片段化核酸的第二链结合的3'

结合特征、与所述杂交位点互补的序列、和另外的序列，其任选为用于实现桥式扩增的结合序列；并且其中所述寡核苷酸的一种或两种包含3’和/或5’保护性特征。

[0202] 在一些实施方案中，步骤的顺序可以为步骤γ)、步骤δ)、步骤α)和随后步骤β)。由此，在另一个实施方案中，提供了用于鉴定核酸样品中的目标特征的样品制备方法，其中制备包括修饰与目标特征相关的核酸以适于与包含固定引物的基底结合，该方法依次包括：

γ)将核酸样品片段化成核酸片段；

δ)在连接条件下将所述片段暴露于第二对寡核苷酸，所述第二对寡核苷酸的第一寡核苷酸包含第二测序引物可与之结合的杂交位点(RD2 SP)；和第二较长的寡核苷酸，其与第二对的所述第一寡核苷酸部分互补并包含用于与所述片段化核酸的第二链结合的3’结合特征、与所述杂交位点互补的序列、和另外的序列，其任选为用于实现桥式扩增的结合序列；并且其中所述寡核苷酸的一种或两种包含3’和/或5’保护性特征；

α)将样品暴露于能够在目标特征处切割核酸的至少一条链的条件，并使多个核酸变性为单链核酸；和

β)在连接条件下将样品暴露于第一对寡核苷酸，所述第一对寡核苷酸的第一寡核苷酸包含使得所述寡核苷酸能够连接至所述DSB的第一链的5’结合特征、第一测序引物可与之结合的杂交位点(RD1 SP)和用于从DSB池中分离所述DSB的结合序列；和第二寡核苷酸，其与第一对的所述第一寡核苷酸互补；其中所述寡核苷酸的一种或两种包含3’和/或5’保护性特征。

[0203] 方法可进一步包括：

ε)使片段变性以提供单链核酸。

[0204] 方法可进一步包括：

ζ)将部分ε)的链分成两组：A组，已经在第一末端处连接由部分β)的寡核苷酸提供的第一杂交位点和结合序列并在另一末端处连接由部分δ)的寡核苷酸提供的第二杂交位点和另外的序列的那些片段，和B组，并未在第一末端处连接由部分β)的寡核苷酸提供的杂交位点和结合序列以及并未在另一末端处连接由部分δ)的寡核苷酸提供的第二杂交位点和另外的序列的那些片段。

[0205] 方法可进一步包括：

η)使用与第一和/或第二杂交位点结合的引物对A组的链进行测序。

[0206] 在本发明的另一方面，方法可调整为特别适于用于基于微珠的系统。例如Ion Torrent测序。在图14和15中例示了特定实施方案。

[0207] 由此，在本发明的一个方面，提供了用于鉴定核酸样品中的DNA DSB的样品制备方法，其中制备包括修饰DSB相关核酸以适于与包含固定的第一引物的基底结合，该方法包括：

1)提供包含多个核酸的样品；

2)在有助于连接的条件下将多个核酸暴露于第一衔接子，其中第一衔接子包含能够连接至DSB的一条链的3’末端的寡核苷酸且所述寡核苷酸包含能够与第二引物杂交的序列；

3)将多个核酸片段化；

4)在有助于连接的条件下将多个核酸暴露于第二衔接子，其中第二衔接子包含能

够在片段化诱导的断裂处连接至一条链的5’末端但不能连接至第一衔接子的寡核苷酸，且所述寡核苷酸包含与固定的第一引物的区域相同的序列；和

5)在适于引物延伸的条件下使多个核酸与第二引物接触。

[0208] 包含多个核酸的样品可以是本文中公开的任何样品。特别地，样品可以怀疑含有DSB，可以用能引起怀疑能引起DSB的药剂处理，或者可以包含或怀疑包含可以转化为DSB的目标特征。样品可包括能够被特异性切割的目标特征。

[0209] 样品可以是能够包含DSB的任何DNA样品，例如gDNA。

[0210] 在一个特定实施方案中，步骤2)是：

在有助于连接的条件下将多个核酸暴露于第一衔接子对，其中第一衔接子对能够连接至DSB的一条链的至少3’末端，且其中第一衔接子对包含至少部分互补的第一和第二寡核苷酸，并且第一寡核苷酸可连接至3’末端并包含能够与第二引物杂交的序列；和

其中步骤4)是：

在有助于连接的条件下将多个核酸暴露于第二衔接子对，其中第二衔接子对能够在片段化诱导的断裂处连接至一条链的至少5’末端但不能连接至第一衔接子对的第一寡核苷酸，其中第二衔接子包含第一和第二部分互补的寡核苷酸，且第一寡核苷酸可连接至5’末端并包含与固定的第一引物的区域相同的序列，并且第二寡核苷酸不包含与所述与固定的第一引物的区域相同的序列互补的序列。

[0211] 在特定实施方案中，可连接至5’末端的第二衔接子对的寡核苷酸包含根据AACCCACTACGCCTCCGCTTTCC(SEQ ID NO:40)的序列的5、10、15、20或全部23个碱基。另一寡核苷酸具有不包含序列GGAAAGCGGAGGCGTAGTGGTT(SEQ ID NO:36)的超过5、10、15、20个碱基的序列或不包含序列GGAAAGCGGAGGCGTAGTGGTT(SEQ ID NO:36)的全部22个碱基的序列。

[0212] 根据SEQ ID NO:36和SEQ ID NO:40的寡核苷酸可包含1至12、1至10、1至8、1至5、1至3、2或1个修饰，诸如取代、缺失或插入。在一个实施方案中，修饰是取代。

[0213] 第一衔接子对的第一和/或第二寡核苷酸可包含3’和/或5’保护性特征。这些保护性特征可以是本文中公开的任何保护性特征，特别是结合就本文中公开的方法的步骤i)论述的第一对寡核苷酸公开的任何保护性特征。

[0214] 第二衔接子对的第一和/或第二寡核苷酸可包含3’和/或5’保护性特征。这些保护性特征可以是本文中公开的任何保护性特征，特别是结合就本文中公开的方法的步骤iii)论述的第二对寡核苷酸公开的任何保护性特征。

[0215] 在特定实施方案中，由于第一衔接子的3’修饰的存在，第二衔接子不能连接至第一衔接子。例如C3 3’链终止子。

[0216] 可连接至5’末端的第二衔接子的寡核苷酸包含与第二引物的区域相同的序列，使得当生成互补链时，该互补链包含引物可通过杂交与之结合的区域。与第二引物的区域相同的该序列可以与固定引物的5、10、15、20、21、24或更多个碱基相同。

[0217] 方法可进一步包括使多个核酸变性以形成多个单链核酸。该步骤的特征可以是结合本发明的其它实施方案在本文中公开的任何特征。

[0218] 方法可进一步包括在适于固定的第一引物与互补核酸杂交的条件下使多个核酸与包含固定的第一引物的基底接触。例如，具有连接的衔接子的核酸的样品可以随后与基底(诸如包含固定引物的微珠)结合。引物可以固定到珠粒上，使得5’端是固定点的近端，而

3’端是固定点的远端。

[0219] 可以使用常规技术,使得基底(诸如微珠)展示多个拷贝的具有相同序列的核酸。

[0220] 方法可进一步包括获得与基底杂交的任何核酸的序列信息。衔接子可包含用于测序引物的结合的位点,以协助该过程。衔接子可包含索引序列以协助该过程。

[0221] 根据本发明的另一方面,提供了适用于鉴定核酸样品,诸如gDNA样品中的DNA双链断裂(DSB)的双链衔接子,其包含:

含有根据AACCCACTACGCCTCCGCTTTCC(SEQ ID NO:40)的序列的第一寡核苷酸;和不包含序列GGAAAGCGGAGGCGTAGTGGTT(SEQ ID NO:36)的超过5、10、15、20个碱基的序列,或不包含序列GGAAAGCGGAGGCGTAGTGGTT(SEQ ID NO:36)的全部22个碱基的序列的第二寡核苷酸;其中所述寡核苷酸的一种或两种分别包含3’和/或5’保护性特征。

[0222] 在以下权利要求和本发明的前述描述中,除了上下文因明示语言或必要含义而另有要求,否则词语“包含(comprises)”或诸如“包含(comprises)”或“包含(comprising)”的变体以包括性意义使用,即指定本发明的各种实施方案中所述特征的存在,但不排除另外的特征的存在或增加。

[0223] 本说明书中引用的所有参考文献,包括任何专利或专利申请,均通过引用特此并入。不进行任何参考文献构成现有技术的承认。此外,不进行任何现有技术构成本领域中公知常识的一部分的承认。

[0224] 本发明的每个方面的优选特征可以如结合任何其它方面所描述的那样。

[0225] 结合发明特征步骤i)、ii)、iii)等的陈述公开的任何特征可以与结合发明特征步骤a)、b)、c)等的陈述公开的特征组合。结合发明特征步骤a)、b)、c)等的陈述公开的任何特征可以与结合发明特征步骤i)、ii)、iii)等的陈述公开的特征组合。

[0226] 本发明的其它特征将从以下实施例变得显而易见。一般而言,本发明延及本说明书(包括所附权利要求和附图)中公开的特征的任何新的一个或任何新的组合。由此,结合本发明的特定方面、实施方案或实施例所描述的特征、整数、特性、化合物或化学部分应理解为适用于本文中描述的任何其它方面、实施方案或实施例,除非与其不相容。

[0227] 此外,除非另行说明,本文中公开的任何特征可由用于相同或类似目的的替代特征代替。

[0228] 现在将仅通过示例,特别参考以下附图来描述本发明,其中:

实施例

[0229] 方法

细胞培养和处理

在37℃下在5％CO_2下在补充有10％FBS(Life Technologies)的DMEM(Life Technologies)中培养HEK293、HEK293T和U2OS DIvA细胞。使用具有脉冲编码CM-130的Lonza 4D-Nucleofector X单元以224pmol核糖核蛋白(RNP)/3.5×10^5个细胞对HEK293细胞进行核转染。在核转染后0、7、12、24和30小时收获细胞用于INDUCE-seq加工。为了刺激AsiSI依赖性DSB诱导,用300nM 4OHT(Sigma,H7904)处理DIvA细胞4小时。

[0230] Cas9蛋白和sgRNA

靶向EMX1的向导RNA(GAGTCCGAGCAGAAGAAGAA;SEQ ID NO:29)被合成为全长的未

修饰的sgRNA寡核苷酸(Synthego)。Cas9蛋白是内部生产的(AstraZeneca),且含有N-末端6xHN标签。

[0231] INDUCE-seq方法

将细胞以～1×10^5/孔的密度接种到预涂覆有聚-D-赖氨酸的96孔板(Greiner Bio-One,655940)中并在室温(rt)下在4%多聚甲醛(PFA)(Pierce,28908)中交联10分钟。细胞在1×PBS中洗涤以除去甲醛,并在4℃下储存至多30天。INDUCE-seq方法由透化细胞开始。在温育步骤之间,细胞在1×PBS中在室温下洗涤。通过在室温下在裂解缓冲液1(10mM Tris-HCL pH8,10mM NaCl,1mM EDTA,0.2%Triton X-100,pH 8,在4℃下)中温育一小时,接着在37℃下在裂解缓冲液2(10mM Tris-HCL,150mM NaCl,1mM EDTA,0.3%SDS,pH 8,在25℃下)中温育一小时来透化细胞。将透化的细胞在1×CutSmart®缓冲液(NEB,B7204S)中洗涤三次,并使用NEB快速平端化试剂盒(E1201L)+100μg/mL BSA以50μL的最终体积在室温下平端修复一小时。随后将细胞在1×CutSmart®缓冲液中洗涤三次,并使用NEBNext®dA加尾模块(NEB,E6053L)以50μL的最终体积在37℃下进行A加尾30分钟以便将单个dATP添加至双链DNA的3'端。A加尾细胞在1×CutSmart®缓冲液中洗涤三次,随后在室温下在1×T4 DNA连接酶缓冲液(NEB,B0202S)中温育5分钟。使用T4 DNA连接酶(NEB,M0202M)+0.4μM P5衔接子以50μL的最终体积在16℃下连接16-20小时来标记A加尾末端。在连接后,通过在室温下在洗涤缓冲液(10mM Tris-HCL,2MNaCl,2mM EDTA,0.5%Triton X-100,pH 8,在25℃下)中洗涤细胞10次来除去过量的P5衔接子,每个洗涤步骤温育2分钟。细胞在PBS中洗涤一次,并随后在无核酸酶的H_2O(IDT,11-05-01-04)中洗涤一次。通过将细胞在DNA提取缓冲液(10mM Tris-HCL,100mM NaCl,50mM EDTA,1.0%SDS,pH 8,在25℃下)+1mg/mL蛋白酶K(Invitrogen,AM2584)以100μL的最终体积在室温下温育5分钟来提取基因组DNA。将细胞裂解物转移至1.5mL Eppendorf RNA/DNA LoBind管(Fisher Scientific,13-698-792)中并在65℃下温育1小时,在800rpm下摇动。使用基因组DNA Clean&Concentrator™-10(Zymo Research,D4010)纯化DNA,并使用100μL洗脱缓冲液洗脱。在进行文库制备之前,使用1μL样品和Qubit DNA HS试剂盒(Invitrogen,Q32854)评估DNA产率。使用Bioruptor超声波仪将基因组DNA片段化至300-500bp,并使用SPRI微珠(GC Biotech,CNGS-0005)进行尺寸选择以除去＜150bp的片段。使用NEBNext®Ultra™II DNA模块(NEB,E7546L)对片段化和尺寸选择的DNA进行末端修复。使用NEBNext®Ultra™II连接模块(NEB,E7595L)根据制造商的说明书使用7.5μM修饰的半功能性P7衔接子并省略USER酶添加,将片段化和末端修复的DNA直接加入连接反应。使用SPRI微珠纯化连接的测序文库。使用SPRI微珠将文库再纯化两次,并进行尺寸选择以除去＜200bp的片段,以便除去残留的衔接子DNA。使用Illumina®平台的KAPA文库定量试剂盒(Roche,07960255001),通过qPCR定量最终的清洁文库。汇集样品并浓缩至所需体积以便使用SpeedVac进行测序。使用1×75bp高容量流动池在Illumina NextSeq 550上进行测序。

[0232] INDUCE-seq衔接子

所有修饰的INUCE-seq衔接子寡核苷酸购自IDT。通过加热至95℃持续5分钟并使用热循环仪缓慢冷却至25℃,在不含核酸酶的双链体缓冲液(IDT,11-01-03-01)中将单链寡核苷酸以10μM的最终浓度进行退火。在图2中提供了衔接子寡核苷酸的结构的概述。

[0233] 用HindIII的原位DSB诱导

通过使用限制酶HindIII-HF®(NEB,R3104S)在HEK293T细胞中原位诱导DSB来进行INDUCE-seq预实验。该过程与完整的INDUCE-seq方法中所述相同,在末端平端化前加入DSB诱导。细胞透化后,使用50U HindIII-HF®在1×CutSmart®缓冲液中以50μL的最终体积诱导DSB。在37℃下进行消化18小时。

[0234] INDUCE-seq数据分析流程

获得拆分的FASTQ文件并使用默认设置,通过Trim Galore!(Krueger 2015)以除去读段的3'端处的衔接子序列。测序数据的质量用FastQC(Andrews 2010)来评估。在使用BWA-mem(Li and Durbin 2009)与人类参考基因组(GRCh37/hg19)进行读段比对之后,使用SAMtools(Li等人,2009)除去以低比对得分(MAPQ<30)映射的比对,并使用定制AWK脚本过滤软剪切读段以确保准确的DSB分配。使用bedtools bam2bed函数(Quinlan和Hall,2010)将所得BAM文件转换为BED文件,随后使用映射性差的区域、染色体末端和不完整的参考基因组重叠群过滤读段坐标的列表,以便从数据中去除这些特征。DSB位置指定为相对于链取向的读段上游的第一个5'核苷酸,并输出为"breakends"BED文件。小心地去除光学重复,同时保留真实的复发性DSB事件。通过保持每个读段ID,使用定制AWK脚本,将流动池X和Y位置信息用于滤出光学重复。最终的输出是含有定量的单个核苷酸断裂位置列表的BED文件。

[0235] HEK293T细胞中的HindIII诱导的DSB分析

hg19中HindIII靶位点的位置首先使用工具SeqKit locate(Shen等人,2016)在计算机上进行预测,允许与HindIII靶序列AAGCTT有2bp的最大错配。使用bedtools intersect计算与这些预测位点重叠的断裂的数量。为了与DSBCaptrure EcoRV实验(Lensing等人,2016)进行比较,使用≥5个断裂/位点的相同覆盖阈值来限定每个HindIII诱导的断裂位点。

[0236] DIvA细胞中AsiSI诱导的DSB检测和分析

AsiSI靶位点的位置以与HindIII相同的方式计算,但不允许错配,并使用序列GCGATCGC。由于DIvA细胞是雌性的,Y染色体上存在的位点被除去,留下1211个位点用于chr1-X。为了严格计算真实的AsiSI诱导的断裂,在预测的断裂位置处将8bp AsiSI位点减少到1bp基因组间隔。这将每个8bp基因组间隔减少为两个1bp间隔;在正链上的位置6和负链上的位置3处。随后使用bedtools intersect计算1bp断裂末端位置和预测的AsiSI断裂位点之间的直接重叠。对于每个重叠,需要匹配的链取向才被认为是真实的AsiSI诱导的断裂位点。

[0237] CRISPR脱靶分析流程

首先使用Cas-OFFinder的命令行版本(Bae等人,2014)预测hg19中EMX1的两组潜在脱靶位点,对于第一组允许组合的间隔子和标准PAM中多达6个错配,而对于第二组允许多达7个错配。接着,基于种子区中的错配数量过滤两组预测序列,所述种子区定义为PAM近端的12个核苷酸。对于种子中多达2、3、4和5个错配,对每个组进行过滤,生成具有不同的错配过滤参数的8个文件的集合。为了限定CRISPR诱导的DSB,首先将每个23bp预测位点减少至预期的CRISPR断裂位置侧翼的2bp间隔,PAM上游3bp。随后使用bedtools intersecT(Quinlan和Hall,2010)计算这些2bp预期断裂区和INDUCE-seq 1bp断裂末端位置之间的重叠,返回一组在预期的CRISPR断裂位点处鉴定的DSB。最后,基于位点错配数量和在位点处检测到的断裂数量,过滤与CRISPR位点重叠的DSB。具有>n的错配的位点需要具有超过1个

DSB重叠才保留为真实的脱靶位点。使用＞2、＞3、＞4和＞5的错配值过滤每组断裂重叠，得到每个INDUCE-seq样品的总共32个过滤条件和脱靶数据集。

[0238] 计算CRISPR脱靶检测方法之间的重叠

将EMX1脱靶位点与替代方法CIRCLE-seq、Digenome-seq、GUIDE-seq、BLISS和HTGTS进行比较。为每个相应的脱靶检测方法生成基因组间隔文件。使用bedtools intersecT(Quinlan和Hall,2010)计算通过每种方法检测到的EMX1脱靶的重叠。

[0239] 突变结果的扩增子测序验证

扩增子测序DNA文库使用一组定制的rhAmpSeq RNA酶-H依赖性引物(IDT)来制备，所述引物位于EMX1的INDUCE-seq鉴定的脱靶的侧翼。根据制造商的说明书使用rhAmpSeq HotStart Master Mix 1、定制引物混合物和10ng的基因组DNA进行多重PCR。使用SPRI微珠纯化PCR产物，并通过使用rhAmpSeq HotStart Master Mix 2的第二多重PCR掺入Illumina测序P5和P7索引序列。汇集所得测序文库，并使用Illumina NextSeq 550Mid输出流动池用2x 150bp chemistry来进行测序。使用采用以下参数的CRISPresso软件(Pinello等人，2016)v2.0.32确定在靶和脱靶处的编辑结果：CRISPRessoPooled-q30-ignore_substitutions--max_paired_end_reads_overlap 151。使用CRISPressoCompare比较插入缺失频率。

[0240] 结果

通过INDUCE-seq的断裂测量经由两阶段、无PCR文库制备来实现(图1和5)。阶段一由经由全长的、化学修饰的P5衔接子连接至DSB末端来原位标记末端制备的DSB组成。在阶段二中，使用第二化学修饰的半功能性P7衔接子连接提取的、片段化的和末端制备的gDNA。所得包含P5和半功能性P7衔接子的DSB标记的DNA片段可与Illumina流动池相互作用，并随后使用单端测序进行测序。不具有P5衔接子的DNA片段保持非功能性，因为它们缺乏与流动池杂交所需的序列。该方法使得能够富集功能性标记的DSB序列并消除所有其它基因组DNA片段，否则这些片段将促成系统噪音。避免断裂扩增产生了其中单个测序读段等同于单个标记的DSB末端的测序输出(比较图5a和5b，参见图8a)。这种创新使得能够通过测序直接测量并由此定量DSB，代表了基因组中DSB的准确测量的重大进展。

[0241] 在原位断裂标记后，目前可用的DSB检测方法BLISS、DSBCapture和END-seq均采用富集方案以便将DSB标记的DNA片段与剩余的基因组DNA分离。这随后是基于PCR的文库制备和测序(图5b)。这种基于扩增的方法产生其中单个读段不等同于单个断裂的测序输出，其中需要PCR误差校正方案(诸如独特分子标识符(UMI))以试图定量DSB(图5b和图6b)(Yan等人,2017)。重要的是，新的INDUCE-Seq文库制备和衔接子组合与任何公开的原位DSB标记方案都兼容。

[0242] 为了证明INDUCE-seq方法的特征，我们首先检查了在使用高保真HindIII限制性核酸内切酶在固定和透化的细胞中诱导限定的DSB后，INDUCE-seq如何测量基因组范围的DSB。该方法先前已经用于检测方法BLISS、END-seq和DSBCapture。如图3a中所示，INDUCE-seq除了从同一样品中检测到数十万个低水平内源性DSB外还同时检测到数亿个高度复发性的HindIII诱导的DSB，而不需要任何形式的误差校正。这首次使得能够精确定量和表征基因组中的内源性DSB。图7显示了在HEK293细胞中使用INDUCE-seq进行内源DSB检测的实例。总之，这些观察结果显示出用INDUCE-seq可实现的显著较宽的动态范围和灵敏度。图3b

显示了INDUCE-seq检测预期的HindIII切割模式,其中正向和反向测序读段的两个半重叠对称区段映射至正向和反向链上的已知HindIII切割位点。这证实了INDUCE-seq可用于以单个核苷酸分辨率精确测量DSB末端结构。我们测量到用HindIII处理细胞后每个细胞的断裂显著增加,从未经处理的细胞的少于10个内源性断裂到每个处理过的细胞的超过3000个诱导断裂。这证明INDUCE-seq能够定量地测量跨越三个数量级的每个细胞的断裂(图3c)。与使用方法DSBCaptrure检测EcoRV诱导的断裂的等效实验相比,我们发现更大比例的总和和比对的INDUCE-seq读段映射至限制性位点。显著地,96.7%的比对读段映射至限制性位点,这表示断裂检测的保真度比DSBCapture改善了25%(图3d)。重要的是,INDUCE-seq使用了为DSBCapture实验的1/800的细胞,并且鉴定了与通过DSBCapture鉴定的EcoRV比例(93.7%)相似的HindIII限制性位点比例(92.7%)(图3e)。除了包含序列AAGCTT的在靶HindIII位点外,我们在相差一个或两个错配碱基的各种HindIII脱靶序列处鉴定了大量的DSB(图3f)。通过INDUCE-seq测量的HindIII诱导的DSB的总数量范围为在靶位点处的～150,000,000至最低排名的脱靶位点处的仅五个DSB。因此INDUCE-seq定量地检测了跨越八个数量级的断裂,这证明了断裂检测的动态范围比其它方法大大增强。当进行测量活DiVA细胞中AsiSI位点处诱导的DSB的类似实验时,也观察到增强的灵敏度。尽管对为相应的DSBcapture实验的1/40的读段和为BLISS的1/23的读段进行测序,但INDUCE-seq检测到在230AsiSI位点处断裂的存在。这表示超过通过BLISS检测的214个位点和通过DSBCapture检测的121个位点的增加。这表明INDUCE-seq比任何当前的断裂检测方法明显更灵敏、有效,并因此更经济(图3g)。

[0243] 已经通过INDUCE-seq确立了断裂检测的特征,接下来我们将其应用于检测基因组中CRISPR/Cas9诱导的在靶和脱靶DSB。该分析在开发基于CRISPR的疗法的安全性概况分析方面具有重要意义。用广泛表征的EMX1 sgRNA对HEK293细胞进行RNP核转染后,如定制的数据分析流水线所定义的那样,在核转染后0、7、12、24和30小时测定脱靶(图4a、图8和图9)。图3a还显示了在EMX1的在靶和60个脱靶(与靶位点相比错配为2bp至6bp不等)的每一个处检测到的断裂数。

[0244] 该实验揭示了通过EMX1引导的断裂诱导的动力学概况,提供了对CRISPR/Cas9编辑过程的机制的认识。如图4b中所示,在核转染后立即和在时程的早期阶段过程中观察到大部分的在靶和脱靶活性。这些结果表明通过该sgRNA编辑的快速性。当与现有技术比较时,我们发现INDUCE-seq显著优于替代的基于细胞的方法GUIDE-seq、BLISS和HTGTS,以及捕获了先前仅使用体外脱靶发现方法CIRCLE-seq和DiGenome-seq鉴定的若干位点。重要的是,INDUCE-seq还检测到先前通过任何其它方法都未检测到的新的脱靶断裂位点(图4c和图10)。

[0245] 最后,使用来自相同样品的DNA,我们使用扩增子测序测量了在通过INDUCE-seq鉴定的这些在靶和脱靶断裂位点处的编辑结果。扩增子测序仅能检测高于0.1%的背景错误发现率的插入缺失。因此,在在靶位点处检测到编辑的证据,并仅在60个脱靶位点中的四个处检测到编辑的证据,所述四个脱靶位点在整个时程中被鉴定(图4d)。这一观察结果与相同的先前研究一致,所述先前的研究在用EMX1 RNP对HEK293细胞进行核转染后48小时鉴定了插入缺失频率＞0.1%的五个脱靶(图4d,最右栏,48小时)。该数据表明INDUCE-seq可以以比使用扩增子测序检测插入缺失可能的高得多的灵敏度来检测和定量CRISPR诱导的脱

靶DSB。该观察结果凸显了对更灵敏的检测编辑结果的方法的需要，以评估基因组编辑的安全性。引人关注的是，在在靶位点和前两个脱靶位点处仔细检查断裂模式和随后的插入缺失概况揭示了反映在编辑结果中并与编辑结果相关的sgRNA特异性切割模式(图11)。这提高了使用在CRISPR诱导的断裂位点观察到的DSB模式来预测在靶和脱靶位点处的最终编辑结果的可能性。

[0246] 讨论

我们开发了一种新的无PCR方法来制备DNA文库，以用于基因组中DSB的下一代测序。这首次使得能够直接测量细胞中的断裂末端。我们的方法克服了与通常用于标准NGS文库制备的PCR扩增相关的DSB的信噪比不佳的问题。新的INUCE-seq衔接子设计基本上允许测序流动池用于富集标记的DSB序列，无需对它们进行扩增。通过过滤在样品制备过程中产生的与细胞中发现的真实的生理DSB不相关的噪声断裂末端，实现了对信噪比的改善。我们证明了INDUCE-seq在一系列不同应用中测量基因组DSB的特征。我们揭示了它同时灵敏地和定量地检测低水平内源性和高水平限制酶诱导的断裂的能力。先前在不需要具有其自身限制和缺点的复杂且昂贵的误差校正方法的情况下这是不可能的。我们将我们的结果与目前可用的断裂检测方法进行比较，以证明它如何对这些进行改进，不仅在其准确性和灵敏度方面，而且在其简单性、可扩展性、易用性和成本效益方面。这些都是可用于评估CRISPR基因组编辑的合成指导的安全性概况的测定的基本特征。我们证明了对于检测通过EMX1 sgRNA的在靶和脱靶编辑，与几种当前的DSB检测方法相比，INDUCE-seq如何表现。我们揭示了，除检测到通过五种其它方法累积测量的许多脱靶位点之外，INDUCE-seq还鉴定了当前方法不能检测的显著数量的新脱靶位点。我们指出，INDUCE-seq可能是用于安全性概况分析和合成向导RNA设计的一种非常重要的方法，用于未来开发基因组编辑作为治疗方式。此类特征包括全基因组突变、单链断裂和缺口，以及其它类型的可转化为断裂的DNA损伤。INDUCE-seq及其衍生测定的开发在一系列不同的生物医学应用中具有重要的意义。

[0247] 表1.

引物对	序列(5'-3')	标识符
第一引物对(P5衔接子)	P-GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATTC /3SpC3/	SEQ ID NO: 1
	A*ATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T	SEQ ID NO: 2
第二引物对	P-GATCGGAAGAGCACACGTCTGAACTCCAGTCAC/3SpC3/	SEQ ID NO: 3

(P7 衔接子)	C*AAGCAGAAGACGGCATACGAGAT[索引]GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	SEQ ID NO: 4
	C*AAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	SEQ ID NO: 5
	C*AAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	SEQ ID NO: 6
	C*AAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	SEQ ID NO: 7
	C*AAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	SEQ ID NO: 8
	C*AAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	SEQ ID NO: 9
	C*AAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	SEQ ID NO: 10
	C*AAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	SEQ ID NO: 11
	C*AAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	SEQ ID NO: 12
	C*AAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	SEQ ID NO: 13
	C*AAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	SEQ ID NO: 14
	C*AAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	SEQ ID NO: 15
	C*AAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	SEQ ID NO: 16
	C*AAGCAGAAGACGGCATACGAGATTGTTGACTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	SEQ ID NO: 17
	C*AAGCAGAAGACGGCATACGAGATACGGAACTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	SEQ ID NO: 18
	C*AAGCAGAAGACGGCATACGAGATTCTGACATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	SEQ ID NO: 19
	C*AAGCAGAAGACGGCATACGAGATCGGGACGGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	SEQ ID NO: 20
	C*AAGCAGAAGACGGCATACGAGATGTGCGGACGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	SEQ ID NO: 21
	C*AAGCAGAAGACGGCATACGAGATCGTTTCACGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	SEQ ID NO: 22
	C*AAGCAGAAGACGGCATACGAGATAAGGCCACGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	SEQ ID NO: 23
	C*AAGCAGAAGACGGCATACGAGATTCCGAAACGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	SEQ ID NO: 24
	C*AAGCAGAAGACGGCATACGAGATTACGTACGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	SEQ ID NO: 25
	C*AAGCAGAAGACGGCATACGAGATATCCACTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	SEQ ID NO: 26
	C*AAGCAGAAGACGGCATACGAGATATATCAGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	SEQ ID NO: 27
	C*AAGCAGAAGACGGCATACGAGATAAAGGAATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	SEQ ID NO: 28

*硫代磷酸酯连接对核酸外切酶活性的抗性

/3SpC3/C3间隔子亚磷酰胺(共价嵌段)对核酸外切酶、连接酶和5'>3'聚合酶活性的抗性

-P 5'磷酸基团促进5'连接

INDEX Illumina测序索引序列能够拆分汇集的测序文库

*T具有硫代磷酸酯连接的3'脱氧胸苷三磷酸"T尾"提供用于连接至"A加尾"DNA片段的基底

表2

寡核苷酸修饰(P5 或 P7 衔接子)	活性
硫代磷酸酯连接(寡核苷酸的 5'和 3'端)	抵抗 5'>3'和 3'>5'核酸外切酶活性。
硫代磷酸酯连接(在整个寡核苷酸中)	抵抗核酸内切酶活性。
C3 间隔子亚磷酰胺 (3')	抵抗核酸外切酶、连接酶、末端转移酶、5'>3'聚合酶活性。
3' 磷酸基团	抑制通过某些 3'-核酸外切酶的降解，可用于阻断通过 DNA 聚合酶进行的延伸，并抵抗连接。
2'-O-甲基(2'OMe)	包括该修饰的 DNA 寡核苷酸对 DNA 酶的敏感性通常为未修饰的 DNA 的 1/10 至 1/5。
反向 dT (3')	寡核苷酸的 3'端处的反向 dT 导致 3'-3'连接，这将抑制通过 3'核酸外切酶的降解或通过 DNA 聚合酶进行的延伸。
5'反向 ddT	在寡核苷酸的 5'末端处的 5'反向 ddT 抵抗连接，并可以提供对某些形式的核酸外切酶的抗性。
锁核酸碱基	提高寡核苷酸的解链温度，并防止假性杂交，提高结合特异性。
双脱氧胞苷(ddC) (3')	防止通过聚合物进行的 3'延伸的 3'链终止子。

参考文献

Andrews,S.(2010).FastQC:a quality control tool for high throughput sequence data.Available at:http://www.bioinformatics.babraham.ac.uk/projects/fastqc[Accessed:2019].

Bae,S.,Park,J.和Kim,J.S.(2014).Cas-OFFinder:a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases.Bioinformatics 30(10):1473-1475.

Krueger,F.(2015).Trim Galore:A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files.Available at:http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/[Accessed:2019].

Lensing,S.V.,Marsico,G.,Hansel-Hertsch,R.,Lam,E.Y.,Tannahill,D.和Balasubramanian,S.(2016).DSBCapture:in situ capture and sequencing of DNA breaks.Nature Methods 13(10):855-+.

Li,H.和Durbin,R.(2009).Fast and accurate short read alignment with Burrows-Wheeler transform.Bioinformatics 25(14):1754-1760.

Li,H.,Handsaker,B.,Wysoker,A.,Fennell,T.,Ruan,J.,Homer,N.,...Genome Project Data,P.(2009).The Sequence Alignment/Map format and SAMtools.Bioinformatics 25(16):2078-2079.

Pinello,L.,Canver,M.C.,Hoban,M.D.,Orkin,S.H.,Kohn,D.B.,Bauer,D.E.和Yuan,G.C.(2016).Analyzing CRISPR genome-editing experiments with

CRISPResso.Nature Biotechnology34(7):695-697.

Quinlan,A.R.和Hall,I.M.(2010).BEDTools:a flexible suite of utilities for comparing genomic features.Bioinformatics 26(6):841-842.

Shen,W.,Le,S.,Li,Y.和Hu,F.Q.(2016).SeqKit:A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation.Plos One 11(10).

Yan,W.X.,Mirzazadeh,R.,Garnerone,S.,Scott,D.,Schneider,M.W.,Kallas,T.,...Crosetto,N.(2017).BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks.Nature Communications 8.

<110> University College Cardiff Consultants Ltd
Broken String Biosciences Limited
<120> 用于分离双链断裂的方法
<130> BRSB-001/01WO 343608-2004
<150> GB2013141.3
<151> 2020-08-21
<160> 111
<170> PatentIn version 3.5
<210> 1
<211> 58
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 1
gatcggaaga gcgtcgtgta gggaaagagt gtagatctcg gtggtcgccg tatcattc 58
<210> 2
<211> 58
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 2
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatct 58
<210> 3
<211> 33
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 3
gatcggaaga gcacacgtct gaactccagt cac 33
<210> 4
<211> 64
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列

<220>
<221> misc_feature
<222> (25)..(30)
<223> n是a、c、g或t
<400> 4
caagcagaag acggcatacg agatnnnnnn gtgactggag ttcagacgtg tgctcttccg 60
atct 64
<210> 5
<211> 64
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 5
caagcagaag acggcatacg agatcgtgat gtgactggag ttcagacgtg tgctcttccg 60
atct 64
<210> 6
<211> 64
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 6
caagcagaag acggcatacg agatacatcg gtgactggag ttcagacgtg tgctcttccg 60
atct 64
<210> 7
<211> 64
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 7
caagcagaag acggcatacg agatgcctaa gtgactggag ttcagacgtg tgctcttccg 60
atct 64
<210> 8
<211> 64
<212> DNA
<213> 人工序列
<220>

<223> 衔接子序列
<400> 8
caagcagaag acggcatacg agattggtca gtgactggag ttcagacgtg tgctcttccg 60
atct 64
<210> 9
<211> 64
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 9
caagcagaag acggcatacg agatcactgt gtgactggag ttcagacgtg tgctcttccg 60
atct 64
<210> 10
<211> 64
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 10
caagcagaag acggcatacg agatattggc gtgactggag ttcagacgtg tgctcttccg 60
atct 64
<210> 11
<211> 64
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 11
caagcagaag acggcatacg agatgatctg gtgactggag ttcagacgtg tgctcttccg 60
atct 64
<210> 12
<211> 64
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 12
caagcagaag acggcatacg agattcaagt gtgactggag ttcagacgtg tgctcttccg 60

atct 64

<210> 13

<211> 64

<212> DNA

<213> 人工序列

<220>

<223> 衔接子序列

<400> 13

caagcagaag acggcatacg agatctgatc gtgactggag ttcagacgtg tgctcttccg 60

atct 64

<210> 14

<211> 64

<212> DNA

<213> 人工序列

<220>

<223> 衔接子序列

<400> 14

caagcagaag acggcatacg agataagcta gtgactggag ttcagacgtg tgctcttccg 60

atct 64

<210> 15

<211> 64

<212> DNA

<213> 人工序列

<220>

<223> 衔接子序列

<400> 15

caagcagaag acggcatacg agatgtagcc gtgactggag ttcagacgtg tgctcttccg 60

atct 64

<210> 16

<211> 64

<212> DNA

<213> 人工序列

<220>

<223> 衔接子序列

<400> 16

caagcagaag acggcatacg agattacaag gtgactggag ttcagacgtg tgctcttccg 60

atct 64

<210> 17

<211> 66

<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 17
caagcagaag acggcatacg agattgttga ctgtgactgg agttcagacg tgtgctcttc 60
cgatct 66
<210> 18
<211> 66
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 18
caagcagaag acggcatacg agatacggaa ctgtgactgg agttcagacg tgtgctcttc 60
cgatct 66
<210> 19
<211> 66
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 19
caagcagaag acggcatacg agattctgac atgtgactgg agttcagacg tgtgctcttc 60
cgatct 66
<210> 20
<211> 66
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 20
caagcagaag acggcatacg agatcgggac gggtgactgg agttcagacg tgtgctcttc 60
cgatct 66
<210> 21
<211> 66
<212> DNA
<213> 人工序列
<220>

<223> 衔接子序列
<400> 21
caagcagaag acggcatacg agatgtgcgg acgtgactgg agttcagacg tgtgctcttc 60
cgatct 66
<210> 22
<211> 66
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 22
caagcagaag acggcatacg agatcgtttc acgtgactgg agttcagacg tgtgctcttc 60
cgatct 66
<210> 23
<211> 66
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 23
caagcagaag acggcatacg agataaggcc acgtgactgg agttcagacg tgtgctcttc 60
cgatct 66
<210> 24
<211> 66
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 24
caagcagaag acggcatacg agattccgaa acgtgactgg agttcagacg tgtgctcttc 60
cgatct 66
<210> 25
<211> 66
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 25
caagcagaag acggcatacg agattacgta cggtgactgg agttcagacg tgtgctcttc 60

cgatct 66

<210> 26

<211> 66

<212> DNA

<213> 人工序列

<220>

<223> 衔接子序列

<400> 26

caagcagaag acggcatacg agatatccac tcgtgactgg agttcagacg tgtgctcttc 60

cgatct 66

<210> 27

<211> 66

<212> DNA

<213> 人工序列

<220>

<223> 衔接子序列

<400> 27

caagcagaag acggcatacg agatatatca gtgtgactgg agttcagacg tgtgctcttc 60

cgatct 66

<210> 28

<211> 66

<212> DNA

<213> 人工序列

<220>

<223> 衔接子序列

<400> 28

caagcagaag acggcatacg agataaagga atgtgactgg agttcagacg tgtgctcttc 60

cgatct 66

<210> 29

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 靶向EMX1的gRNA

<400> 29

gagtccgagc agaagaagaa 20

<210> 30

<211> 24

<212> DNA

<213> 人工序列
<220>
<223> 衔接子序列——任选不存在
<400> 30
atctcgtatg ccgtcttctg cttg 24
<210> 31
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列——任选不存在
<400> 31
tcggtggtcg ccgtatcatt 20
<210> 32
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 32
caagcagaag acggcatacg agat 24
<210> 33
<211> 34
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 33
gtgactggag ttcagacgtg tgctcttccg atct 34
<210> 34
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 34
aatgatacgg cgaccaccga 20
<210> 35
<211> 18

<212> DNA
<213> 人工序列
<220>
<223> HindIII靶位点
<400> 35
tactcaagct tacccctа 18
<210> 36
<211> 22
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列——任选可以不存在
<400> 36
ggaaagcgga ggcgtagtgg tt 22
<210> 37
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 37
gatcggaaga gcgtcgtgta gggaaagagt gt 32
<210> 38
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 38
tcggtggtcg ccgtatcatt c 21
<210> 39
<211> 33
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 39
acactctttc cctacacgac gctcttccga tct 33
<210> 40

<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 40
aacccactac gcctccgctt tcc 23
<210> 41
<211> 29
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 41
aatgatacgg cgaccaccga gatctacac 29
<210> 42
<211> 29
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 42
gtgtagatct cggtggtcgc cgtatcatt 29
<210> 43
<211> 18
<212> DNA
<213> 人工序列
<220>
<223> HindIII靶位点
<400> 43
gggggtaag cttgagta 18
<210> 44
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> EMX1 sgRNA靶位点
<220>
<221> misc_feature

<222> (21)..(21)
<223> n是a、c、g或t
<400> 44
gagtccgagc agaagaagaa ngg 23
<210> 45
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 ON读段
<400> 45
cctgagtccg agcagaagaa gaagggctcc catcacatca 40
<210> 46
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 ON读段
<400> 46
cctgagtccg agcagaagaa agaagggctc ccatcacatc 40
<210> 47
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 ON读段
<400> 47
cctgagtccg agcagaagaa gaagggctcc catcacatca 40
<210> 48
<211> 34
<212> DNA
<213> 人工序列
<220>
<223> EMX1 ON读段
<400> 48
cctgagtccg agcagaaggg ctcccatcac atca 34
<210> 49
<211> 37
<212> DNA

<213> 人工序列
<220>
<223> EMX1 ON读段
<400> 49
cctgagtccg agcagaagaa gggctcccat cacatca 37
<210> 50
<211> 39
<212> DNA
<213> 人工序列
<220>
<223> EMX1 ON读段
<400> 50
cctgagtccg agcagaagag aagggctccc atcacatca 39
<210> 51
<211> 38
<212> DNA
<213> 人工序列
<220>
<223> EMX1 ON读段
<400> 51
cctgagtccg agcagaagaa agggctccca tcacatca 38
<210> 52
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 ON读段
<400> 52
cctgagtccg agcagaagaa gcagaagggc tcccatcaca 40
<210> 53
<211> 29
<212> DNA
<213> 人工序列
<220>
<223> EMX1 ON读段
<400> 53
cctgagtccg agcagaaggg ctcccatca 29
<210> 54
<211> 37

<212> DNA
<213> 人工序列
<220>
<223> EMX1 ON读段
<400> 54
cctgagtccg agcacaggaa gggctcccat cacatca 37
<210> 55
<211> 18
<212> DNA
<213> 人工序列
<220>
<223> EMX1 ON读段
<400> 55
cctgagtcca tcacatca 18
<210> 56
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> EMX1 ON读段
<400> 56
cctgagctcc catcacatca 20
<210> 57
<211> 27
<212> DNA
<213> 人工序列
<220>
<223> EMX1 ON读段
<400> 57
cctgaaagaa gggctcccat cacatca 27
<210> 58
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 ON读段
<400> 58
cctgagtccg agcagaagaa gagaagggct cccattacat 40
<210> 59

<211> 36
<212> DNA
<213> 人工序列
<220>
<223> EMX1 ON读段
<400> 59
cctgagtccg agcagagaag ggctcccatc acatca 36
<210> 60
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 ON读段
<400> 60
cctgagtccg agcagaagaa aagaagggct cccatcacat 40
<210> 61
<211> 29
<212> DNA
<213> 人工序列
<220>
<223> EMX1 ON读段
<400> 61
cctgagtcag aagggctccc atcacatca 29
<210> 62
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 ON读段
<400> 62
cctgagtccg agcagaagaa gagaagggct cccatcacat 40
<210> 63
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 ON读段
<400> 63
cctgagtccg agcagaagaa gaagaagaag aagaagaaga 40

<210> 64
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 ON读段
<400> 64
cctgagtccg agcggaagaa agaagggctc ccatcacatc 40
<210> 65
<211> 18
<212> DNA
<213> 人工序列
<220>
<223> EMX1 ON读段
<400> 65
cctgctccca tcacatca 18
<210> 66
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-1读段
<400> 66
tcagagttag agcagaagaa gaaaggcatg gagtaaaggc 40
<210> 67
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-1读段
<400> 67
tcagagttag agcagaagaa agaaaggcat ggagtaaagg 40
<210> 68
<211> 37
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-1读段
<400> 68

tcagagttag agcagaagaa aggcatggag taaaggc 37

<210> 69

<211> 26

<212> DNA

<213> 人工序列

<220>

<223> EMX1 OT-1读段

<400> 69

tcagagttag agcatggagt aaaggc 26

<210> 70

<211> 34

<212> DNA

<213> 人工序列

<220>

<223> EMX1 OT-1读段

<400> 70

tcagagttag agcagaaagg catggagtaa aggc 34

<210> 71

<211> 27

<212> DNA

<213> 人工序列

<220>

<223> EMX1 OT-1读段

<400> 71

tcagagttag agcagaagag taaaggc 27

<210> 72

<211> 33

<212> DNA

<213> 人工序列

<220>

<223> EMX1 OT-1读段

<400> 72

tcagagttag agcagaaggc atggagtaaa ggc 33

<210> 73

<211> 31

<212> DNA

<213> 人工序列

<220>

<223> EMX1 OT-1读段

<400> 73
tcagagttag agaaaggcat ggagtaaagg c 31
<210> 74
<211> 26
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-1读段
<400> 74
tcagaagaaa ggcatggagt aaaggc 26
<210> 75
<211> 35
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-1读段
<400> 75
tcagagttag agcagaagaa gcatggagta aaggc 35
<210> 76
<211> 33
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-1读段
<400> 76
tcagagttag agagaaaggc atggagtaaa ggc 33
<210> 77
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-1读段
<400> 77
tcagagttag agcagagaaa ggcatggagt aaaggc 36
<210> 78
<211> 29
<212> DNA
<213> 人工序列
<220>

<223> EMX1 OT-1读段
<400> 78
tcagagttag aaaggcatgg agtaaaggc 29
<210> 79
<211> 29
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-1读段
<400> 79
tcagagttag agcagaagaa agtaaaggc 29
<210> 80
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-1读段
<400> 80
tcagagttag agcagaagaa gaaaggtatg gagtaaaggc 40
<210> 81
<211> 37
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-1读段
<400> 81
tcagagttag agcatgagaa aggcatggag taaaggc 37
<210> 82
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-1读段
<400> 82
tcagagttag agcagaagaa aagaaaggca tggagtaaag 40
<210> 83
<211> 40
<212> DNA
<213> 人工序列

<220>
<223> EMX1 OT-1读段
<400> 83
tcagagttag agcagaagaa gaaaggcatg gggtaaaggc 40
<210> 84
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-1读段
<400> 84
tcagagttag agcagaagaa aaccactacc tgagcaccca 40
<210> 85
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-1读段
<400> 85
tcagagttag agcagaagaa gaagaaaggc atggagtaaa 40
<210> 86
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-1读段
<400> 86
tcagagttag agcagaagaa ctgagcaccc agtccgccct 40
<210> 87
<211> 33
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-1读段
<400> 87
tcagagttag agcagaagaa atggagtaaa ggc 33
<210> 88
<211> 40
<212> DNA

<213> 人工序列
<220>
<223> EMX1 OT-2读段
<400> 88
caagagtcta agcagaagaa gaagagagcc actacccaac 40
<210> 89
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-2读段
<400> 89
caagagtcta agcagaagaa agaagagagc cactacccaa 40
<210> 90
<211> 37
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-2读段
<400> 90
caagagtcta agcagaagaa gagagccact acccaac 37
<210> 91
<211> 39
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-2读段
<400> 91
caagagtcta agcagaagag aagagagcca ctacccaac 39
<210> 92
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-2读段
<400> 92
caagagccta agcagaagaa gaagagagcc actacccaac 40
<210> 93
<211> 40

<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-2读段
<400> 93
caagagtcta agcagaagaa gaggagagcc actacccaac 40
<210> 94
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-2读段
<400> 94
cgagagtcta agcagaagaa gaagagagcc actacccaac 40
<210> 95
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-2读段
<400> 95
caagggtcta agcagaagaa gaagagagcc actacccaac 40
<210> 96
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-2读段
<400> 96
caagagtcta agcagaagaa ggagagagcc actacccaac 40
<210> 97
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-2读段
<400> 97
caagagtcta agcagaggaa gaagagagcc actacccaac 40
<210> 98

<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-2读段
<400> 98
caagagtcta agcaggagaa gaagagagcc actacccaac 40
<210> 99
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-2读段
<400> 99
caagagtcta agcagaagaa gaagagggcc actacccaac 40
<210> 100
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-2读段
<400> 100
caagagtcta agcagaagag gaagagagcc actacccaac 40
<210> 101
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-2读段
<400> 101
caagagtcta agcagaagga gaagagagcc actacccaac 40
<210> 102
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-2读段
<400> 102
caggagtcta agcagaagaa gaagagagcc actacccaac 40

<210> 103
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-2读段
<400> 103
caagagtcta agcagaagaa gaagggagcc actacccaac 40
<210> 104
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-2读段
<400> 104
caagagtcta ggcagaagaa gaagagagcc actacccaac 40
<210> 105
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> EMX1 OT-2读段
<400> 105
caagagtcta agcggaagaa gaagagagcc actacccaac 40
<210> 106
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 引物
<400> 106
ccatctcatc cctgcgtgtc tcc 23
<210> 107
<211> 44
<212> DNA
<213> 人工序列
<220>
<223> 衔接子序列
<400> 107

ccatctcatc cctgcgtgtc tccgactcag ctaaggtaac gatt 44

<210> 108

<211> 43

<212> DNA

<213> 人工序列

<220>

<223> 衔接子序列

<400> 108

atcgttacct tagctgagtc ggagacacgc agggatgaga tgg 43

<210> 109

<211> 21

<212> DNA

<213> 人工序列

<220>

<223> 衔接子序列

<400> 109

atcaccgact gcccatagag a 21

<210> 110

<211> 45

<212> DNA

<213> 人工序列

<220>

<223> 衔接子序列

<400> 110

aacccactac gcctccgctt tcctctctat gggcagtcgg tgatt 45

<210> 111

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 衔接子序列

<400> 111

aacccactac gcctccgctt tcc 23

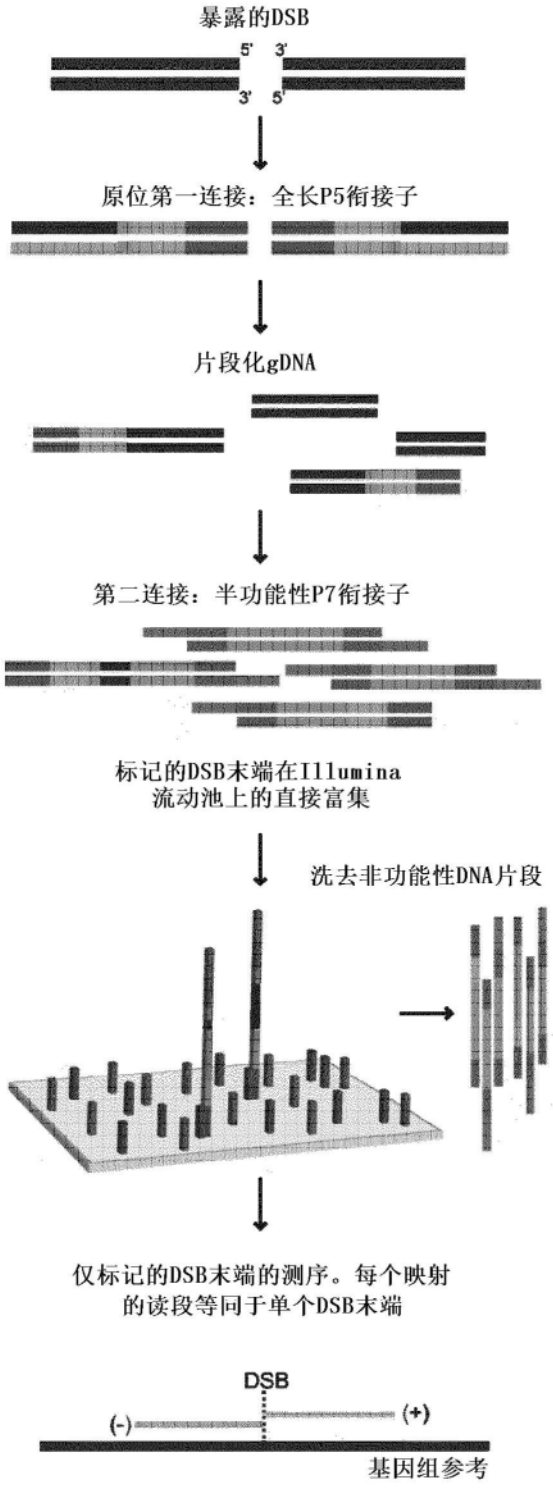

暴露的DSB
5'
3'
3'
5'
原位第一连接：全长P5衔接子
片段化gDNA
第二连接：半功能性P7衔接子
标记的DSB末端在Illumina
流动池上的直接富集
洗去非功能性DNA片段
仅标记的DSB末端的测序。每个映射
的读段等同于单个DSB末端
DSB
(-)
(+)
基因组参考

图1

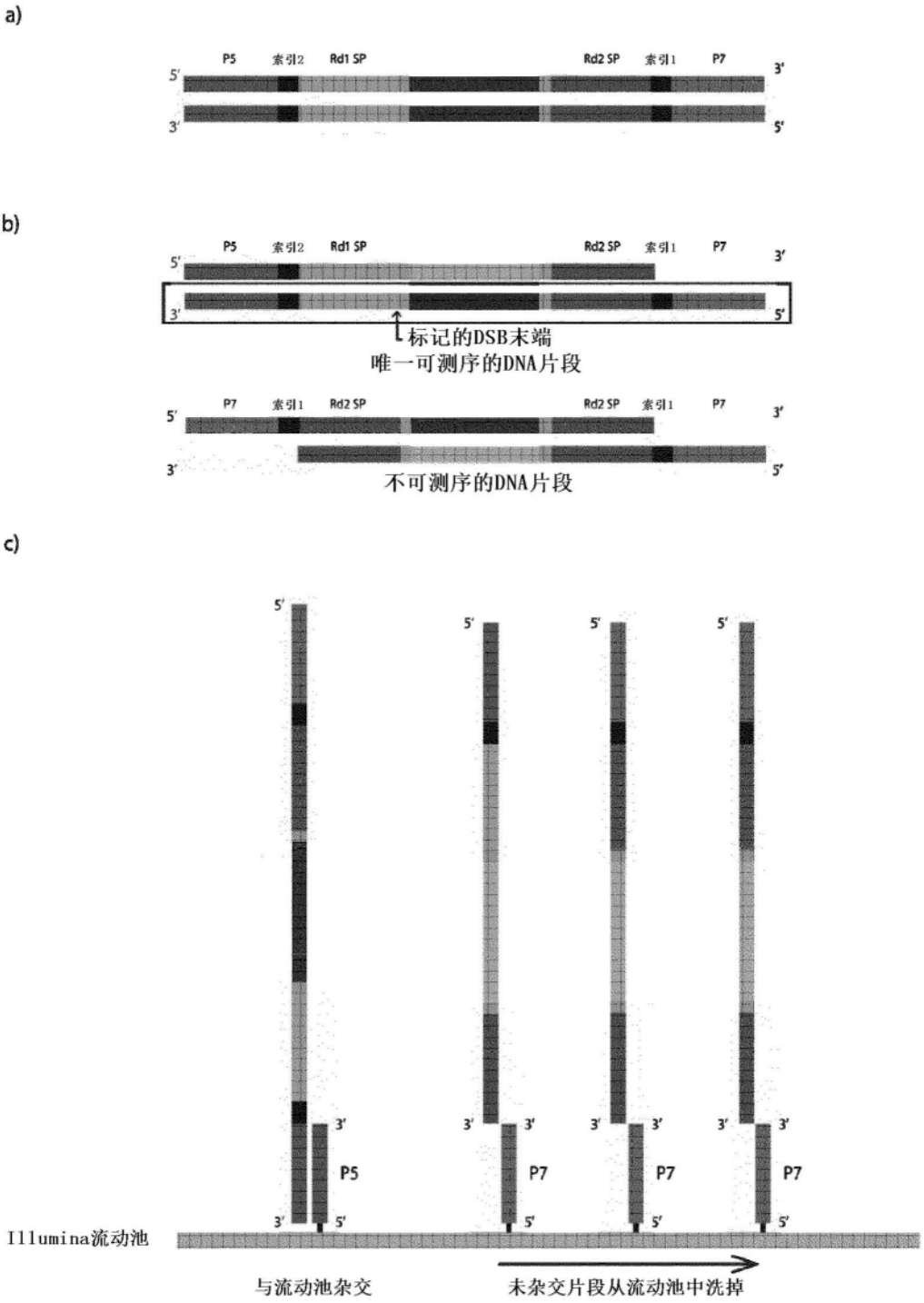

a)
P5
索引2
Rd1 SP
Rd2 SP
索引1
P7
5′
3′
b)
标记的DSB末端
唯一可测序的DNA片段
不可测序的DNA片段
c)
P5
P7
Illumina流动池
与流动池杂交
未杂交片段从流动池中洗掉

图2

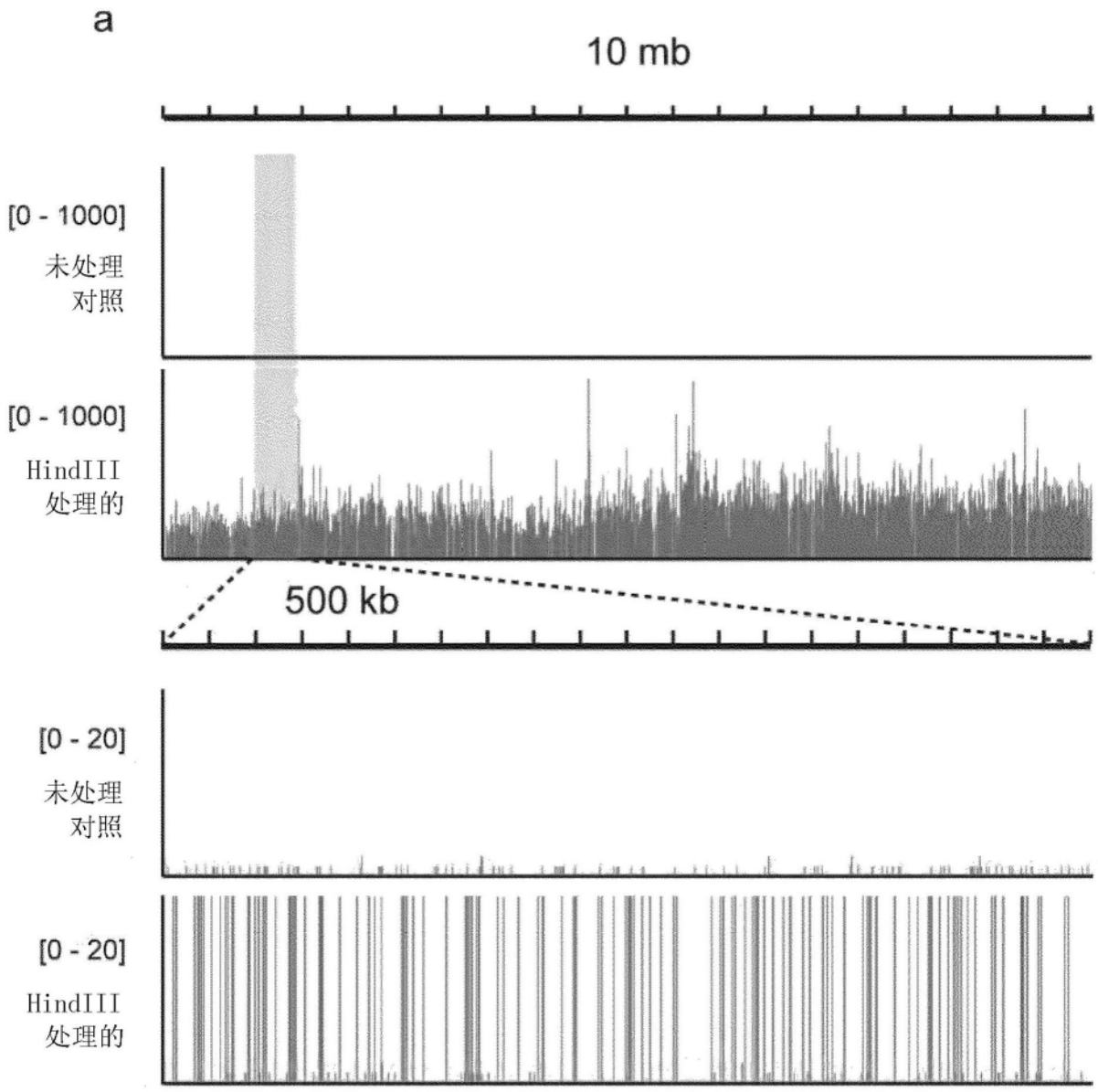

a
10 mb
[0 - 1000]
未处理
对照
[0 - 1000]
HindIII
处理的
500 kb
[0 - 20]
未处理
对照
[0 - 20]
HindIII
处理的

图3

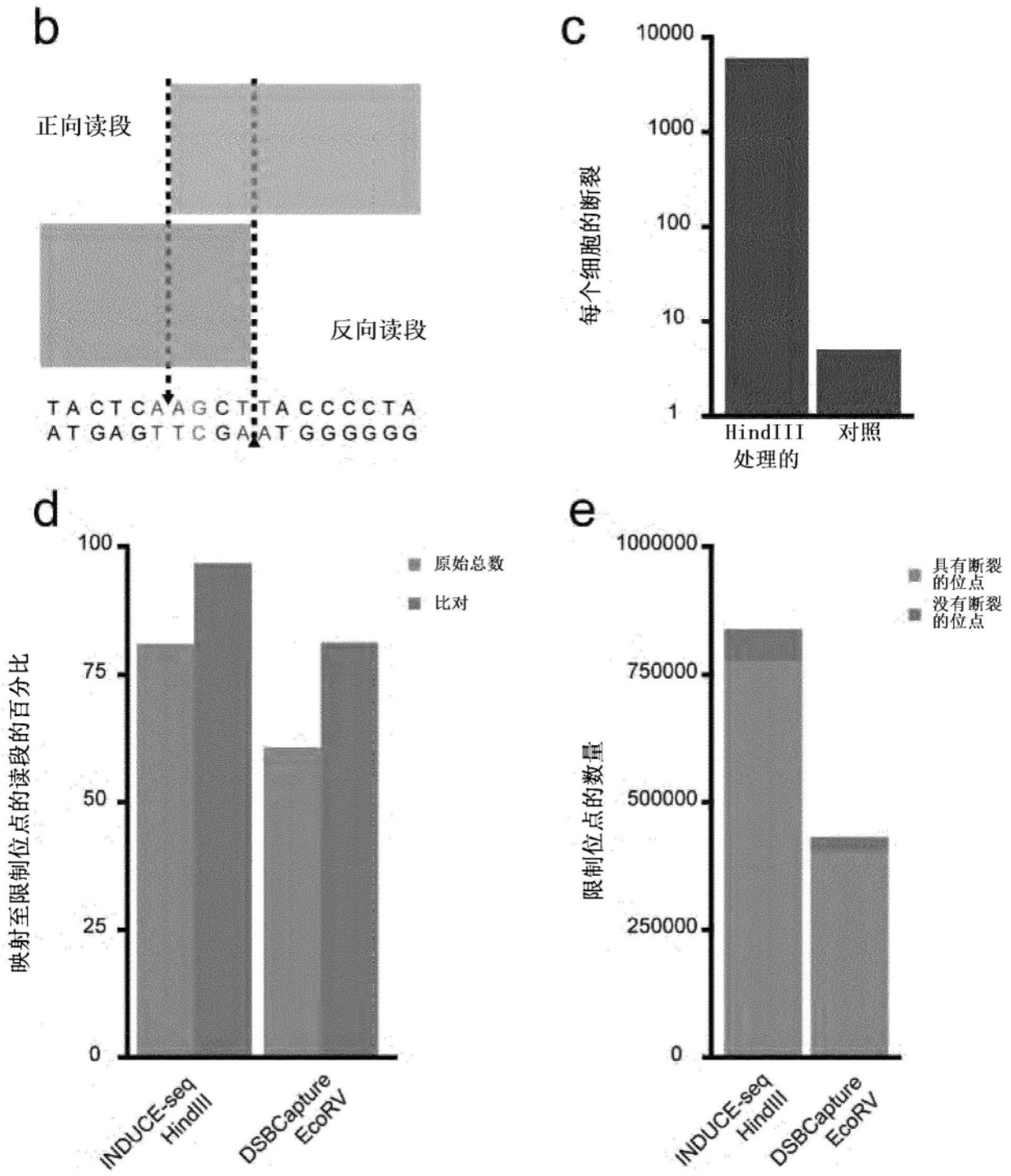

b
正向读段
反向读段
TACTCAAGCTTACCCCTA
ATGAGTTCGAATGGGGGG
c
每个细胞的断裂
10000
1000
100
10
1
HindIII
处理的
对照
d
映射至限制位点的读段的百分比
100
75
50
25
0
原始总数
比对
INDUCE-seq HindIII
DSBCapture EcoRV
e
限制位点的数量
1000000
750000
500000
250000
0
具有断裂的位点
没有断裂的位点
INDUCE-seq HindIII
DSBCapture EcoRV

图3(续)

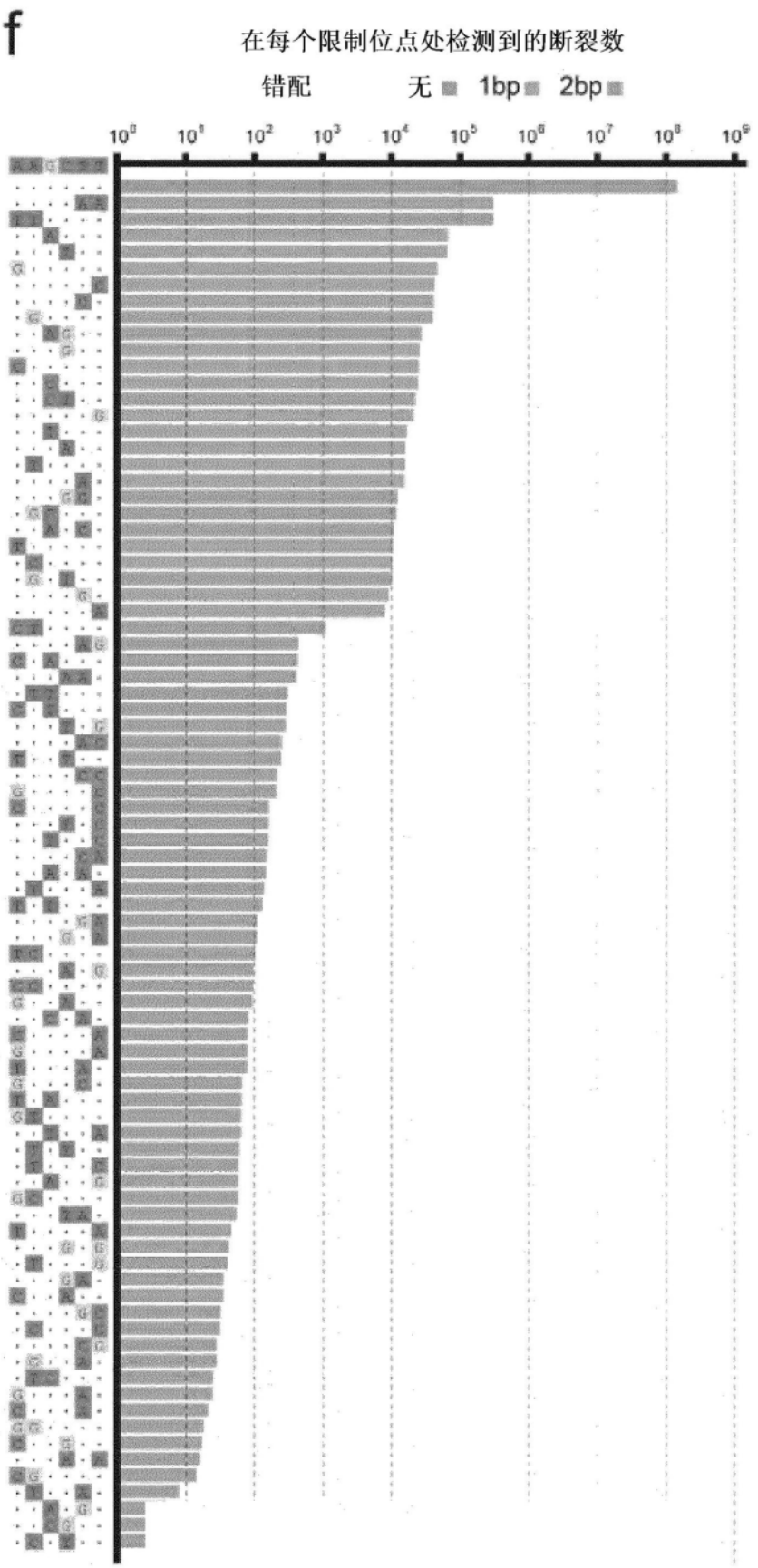

f
在每个限制位点处检测到的断裂数
错配
无
1bp
2bp
10^0
10^1
10^2
10^3
10^4
10^5
10^6
10^7
10^8
10^9

图3(续)

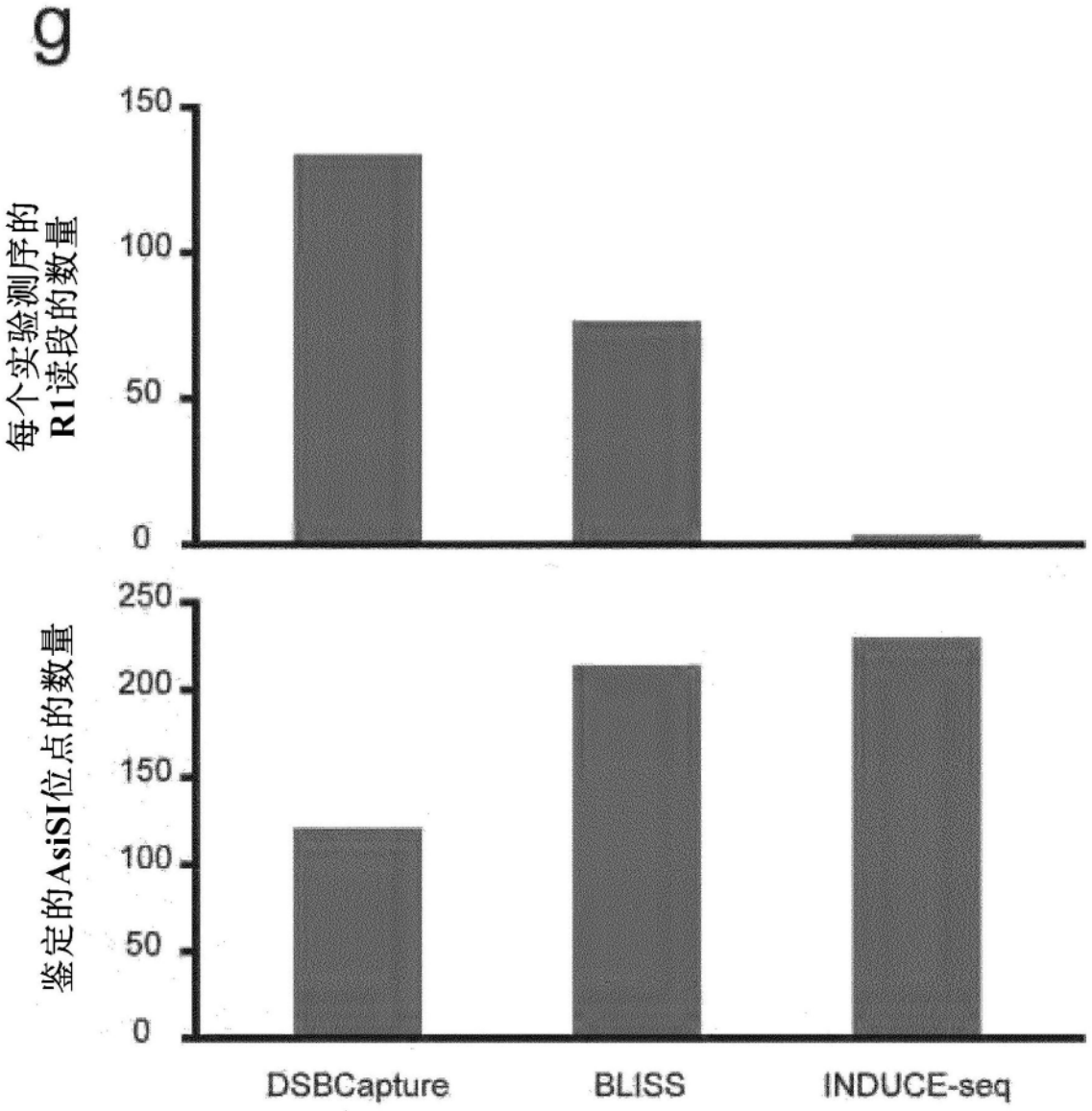
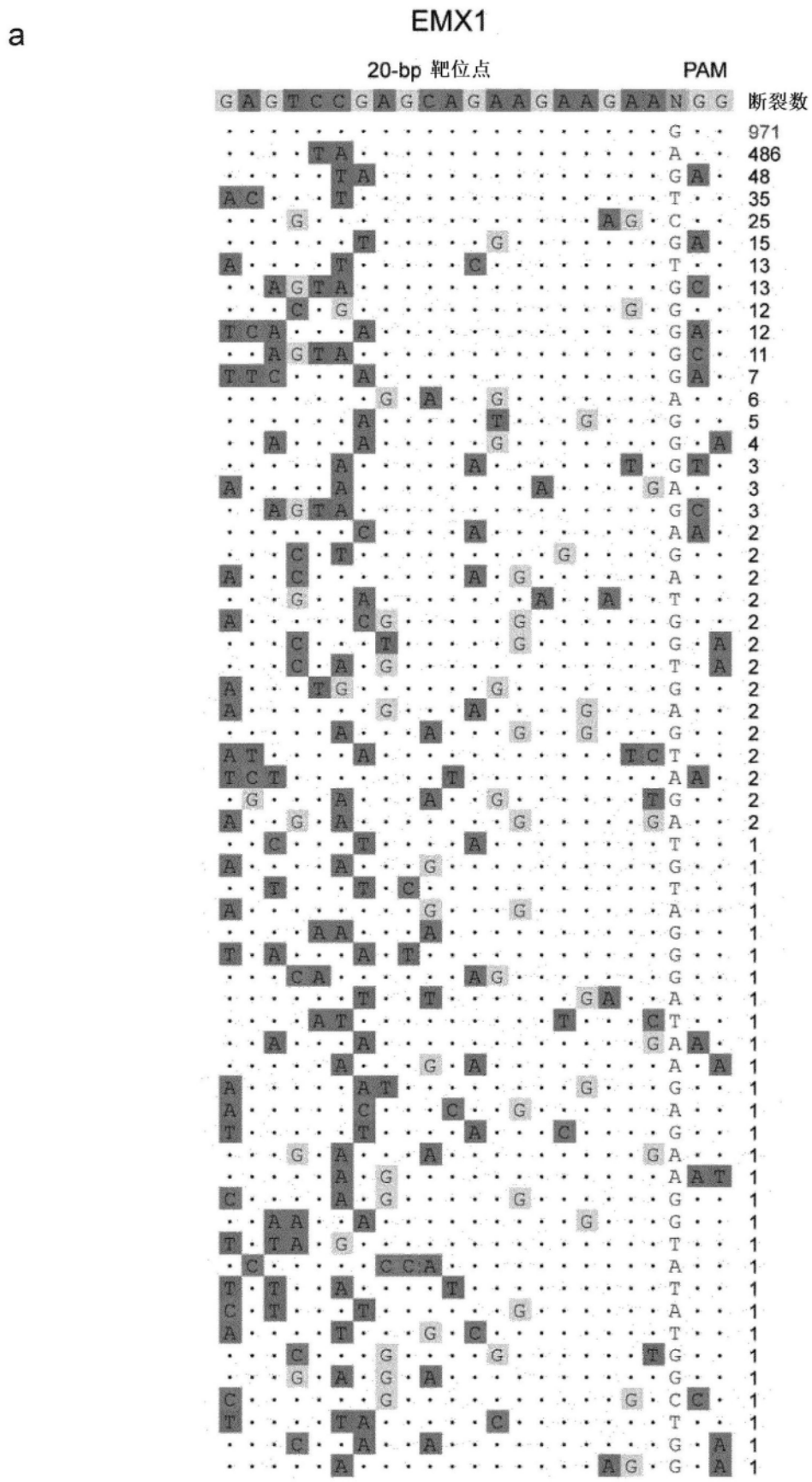

g
每个实验测序的
R1读段的数量
150
100
50
0
鉴定的AsiSI位点的数量
250
200
150
100
50
0
DSBCapture
BLISS
INDUCE-seq

图3(续)


a
EMX1
20-bp 靶位点
PAM
GAGTCCGAGCAGAAGAAGAANGG
断裂数
971
486
48
35
25
15
13
13
12
12
11
7
6
5
4
3
3
3
2
2
2
2
2
2
2
2
2
2
2
2
2
2
1
1
1

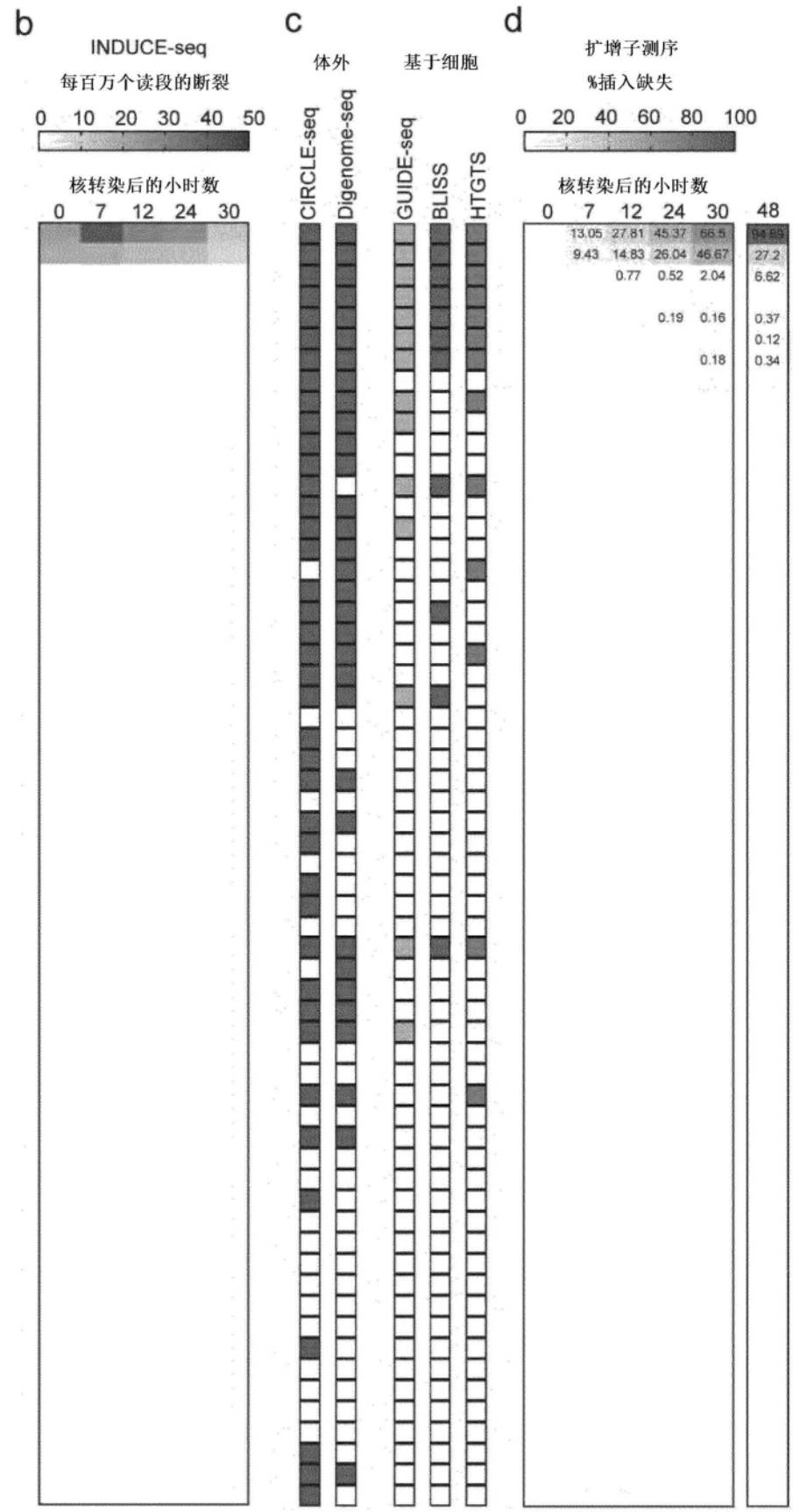

b
INDUCE-seq
每百万个读段的断裂
0 10 20 30 40 50
核转染后的小时数
0 7 12 24 30
c
体外
基于细胞
CIRCLE-seq
Digenome-seq
GUIDE-seq
BLISS
HTGTS
d
扩增子测序
%插入缺失
0 20 40 60 80 100
核转染后的小时数
0 7 12 24 30 48
13.05 27.81 45.37 66.5 94.59
9.43 14.83 26.04 46.67 27.2
0.77 0.52 2.04 6.62
0.19 0.16 0.37
0.12
0.18 0.34

图4(续)

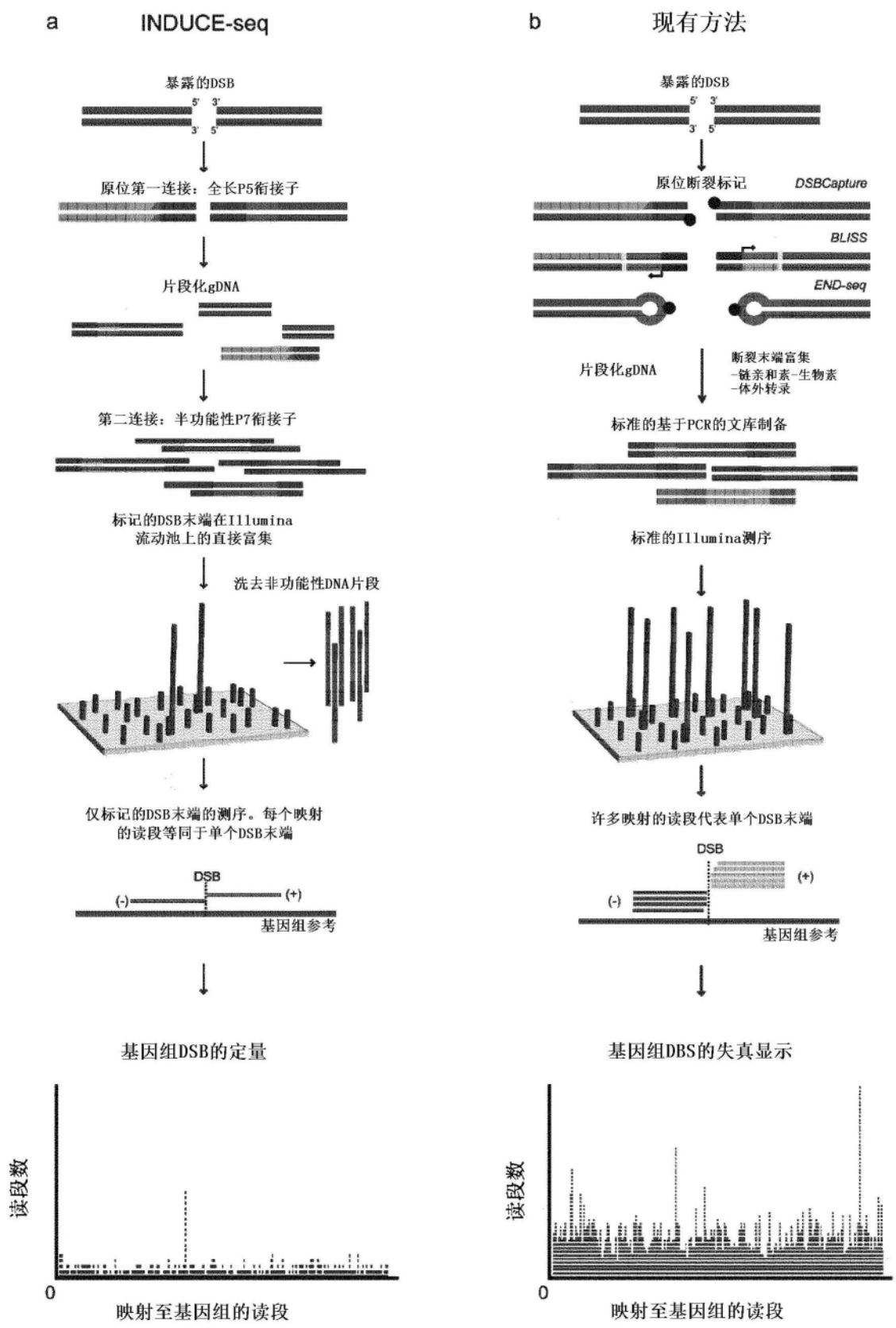

a
INDUCE-seq
暴露的DSB
5'
3'
3'
5'
原位第一连接：全长P5衔接子
片段化gDNA
第二连接：半功能性P7衔接子
标记的DSB末端在Illumina
流动池上的直接富集
洗去非功能性DNA片段
仅标记的DSB末端的测序。每个映射
的读段等同于单个DSB末端
DSB
(-)
(+)
基因组参考
基因组DSB的定量
读段数
0
映射至基因组的读段
b
现有方法
暴露的DSB
原位断裂标记
DSBCapture
BLISS
END-seq
片段化gDNA
断裂末端富集
-链亲和素-生物素
-体外转录
标准的基于PCR的文库制备
标准的Illumina测序
许多映射的读段代表单个DSB末端
DSB
(-)
(+)
基因组参考
基因组DBS的失真显示
读段数
0
映射至基因组的读段

图5

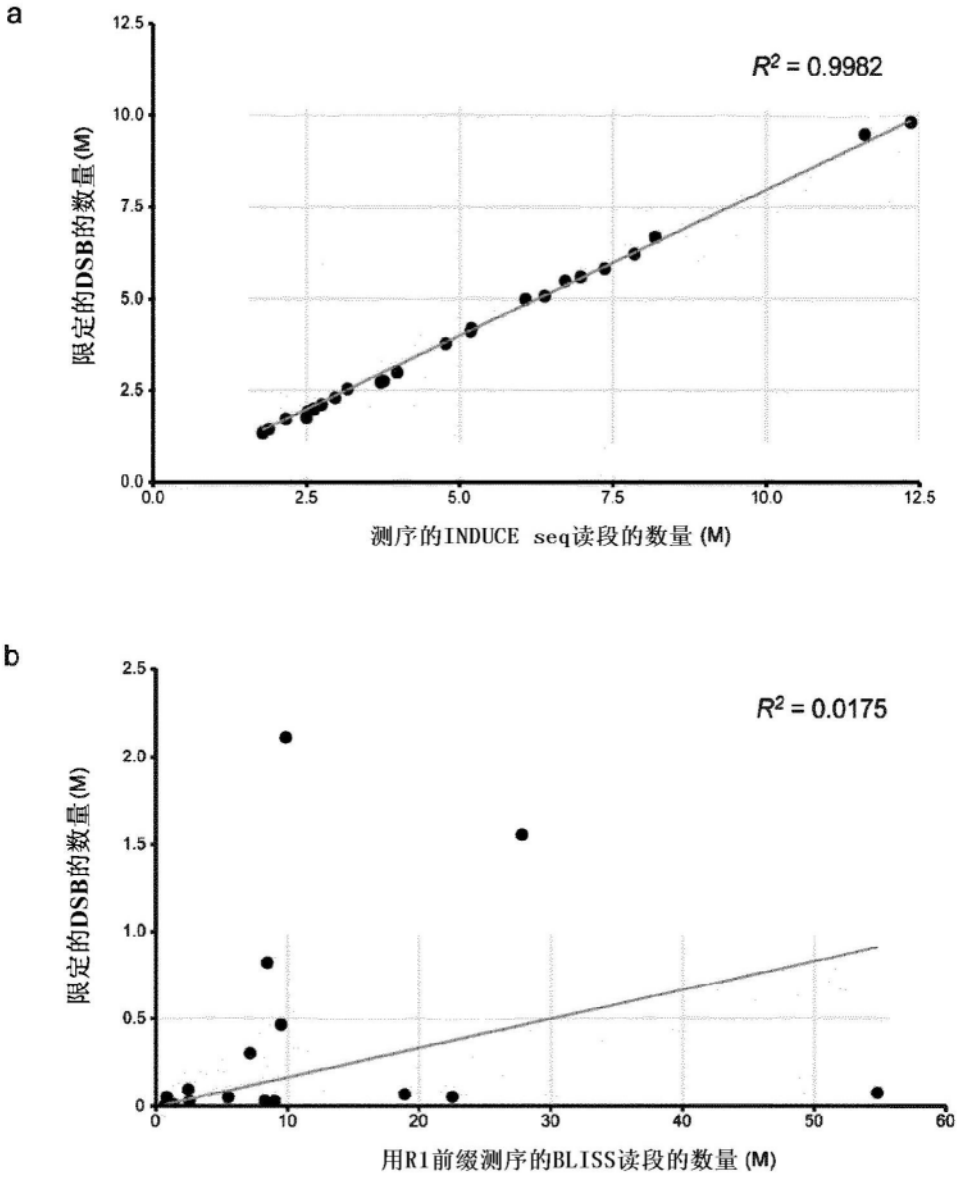

a
R^2 = 0.9982
限定的DSB的数量(M)
测序的INDUCE seq读段的数量(M)
b
R^2 = 0.0175
限定的DSB的数量(M)
用R1前缀测序的BLISS读段的数量(M)

图6

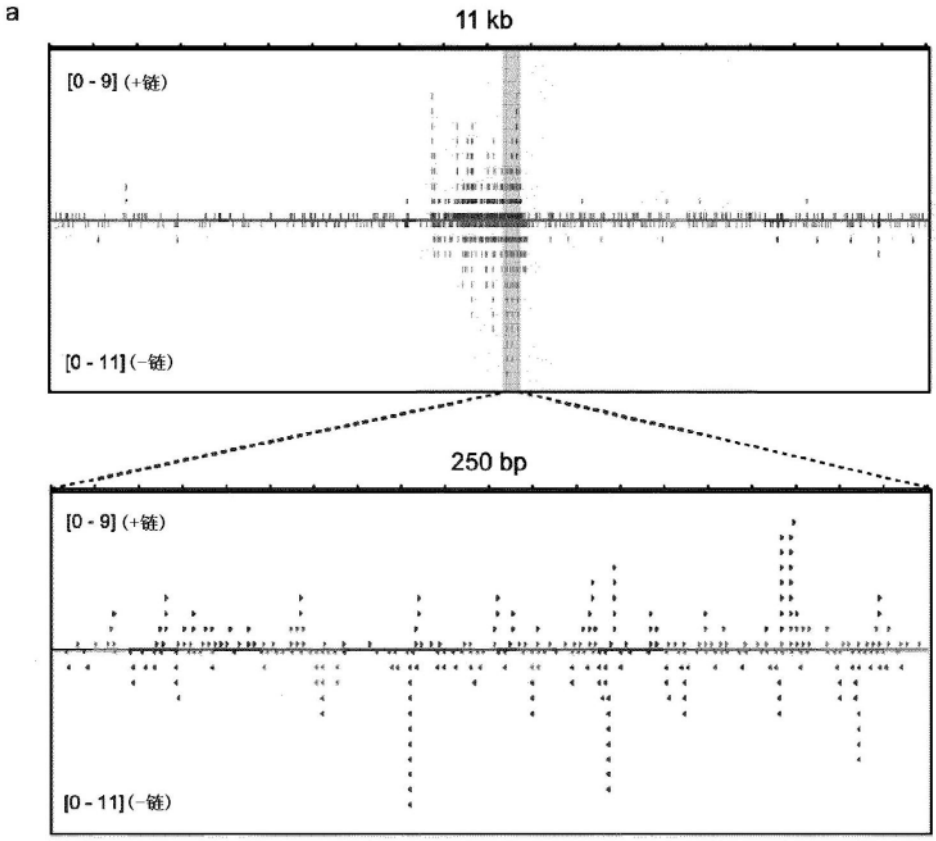


图7

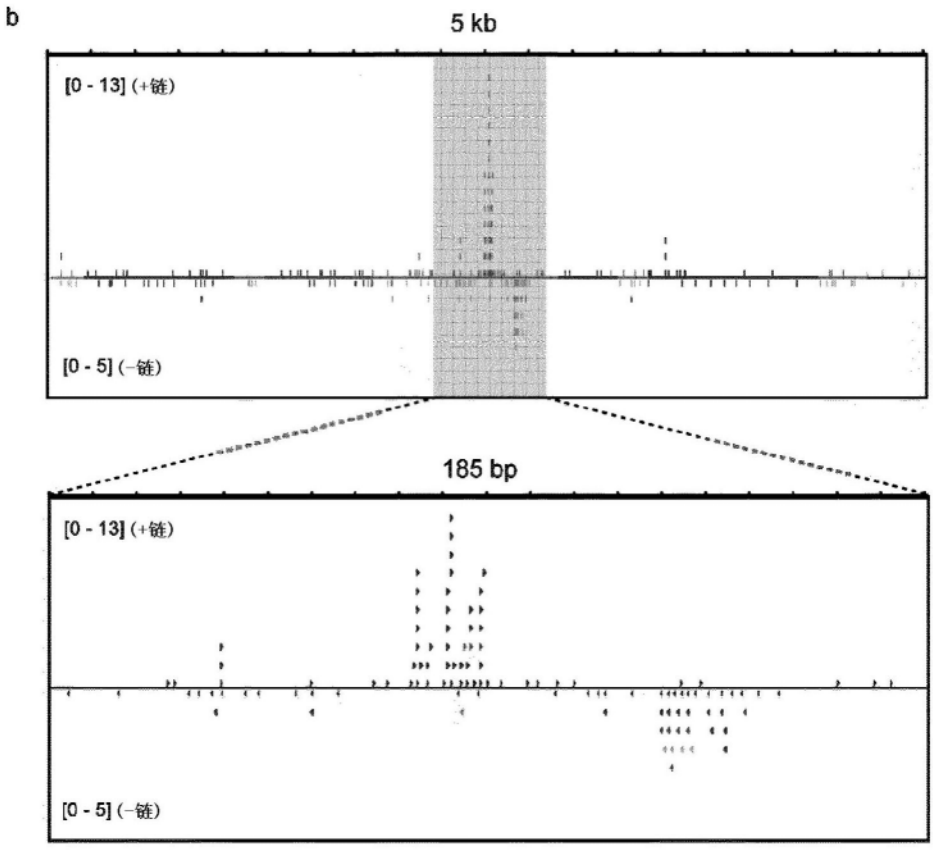


图7(续)

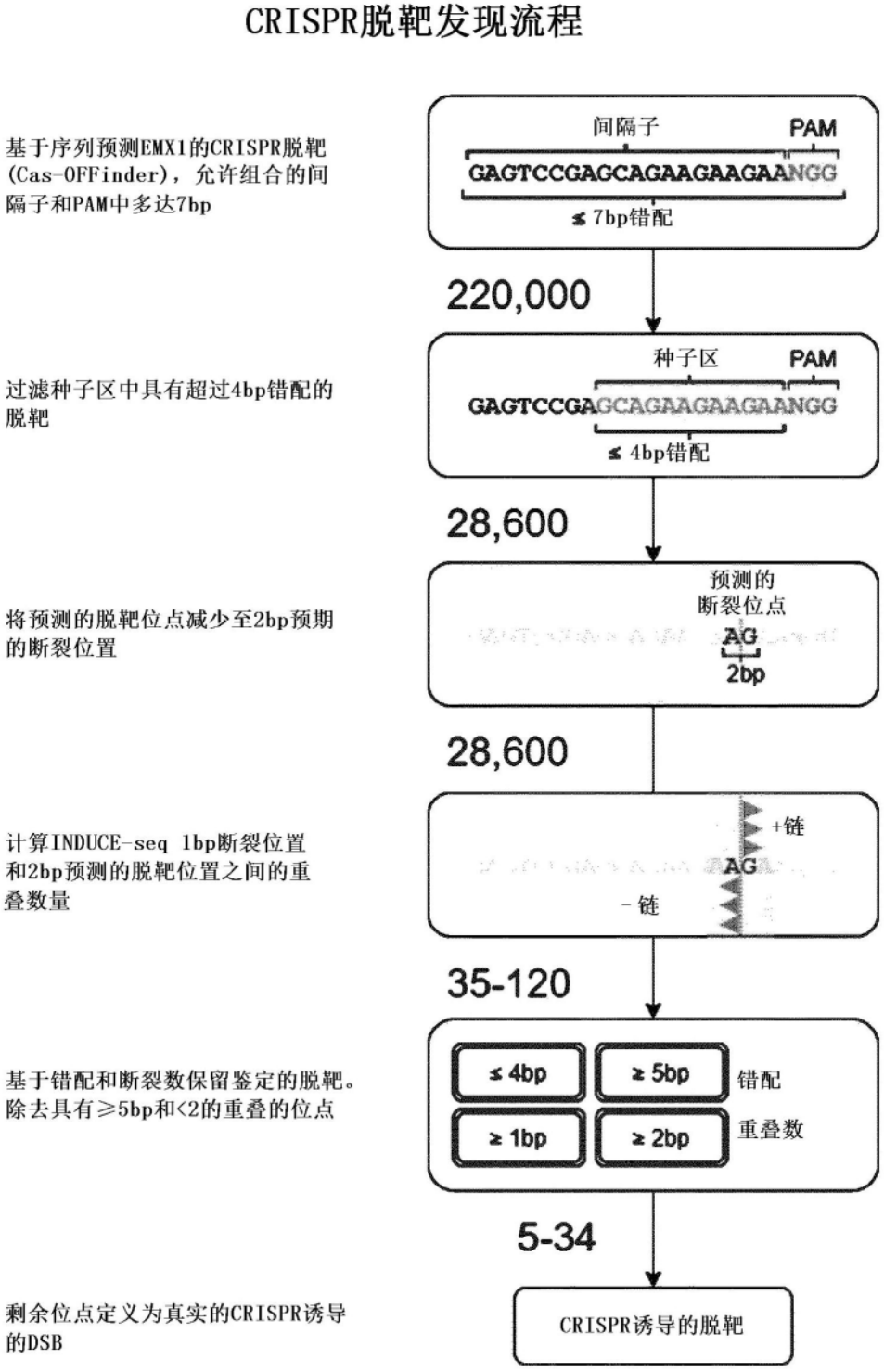

CRISPR脱靶发现流程
基于序列预测EMX1的CRISPR脱靶(Cas-OFFinder)，允许组合的间隔子和PAM中多达7bp
间隔子
PAM
GAGTCCGAGCAGAAGAAGAANGG
≤ 7bp错配
220,000
过滤种子区中具有超过4bp错配的脱靶
种子区
PAM
GAGTCCGAGCAGAAGAAGAANGG
≤ 4bp错配
28,600
将预测的脱靶位点减少至2bp预期的断裂位置
预测的断裂位点
AG
2bp
28,600
计算INDUCE-seq 1bp断裂位置和2bp预测的脱靶位置之间的重叠数量
+链
AG
-链
35-120
基于错配和断裂数保留鉴定的脱靶。除去具有≥5bp和<2的重叠的位点
≤ 4bp
≥ 5bp
错配
≥ 1bp
≥ 2bp
重叠数
5-34
剩余位点定义为真实的CRISPR诱导的DSB
CRISPR诱导的脱靶

图8

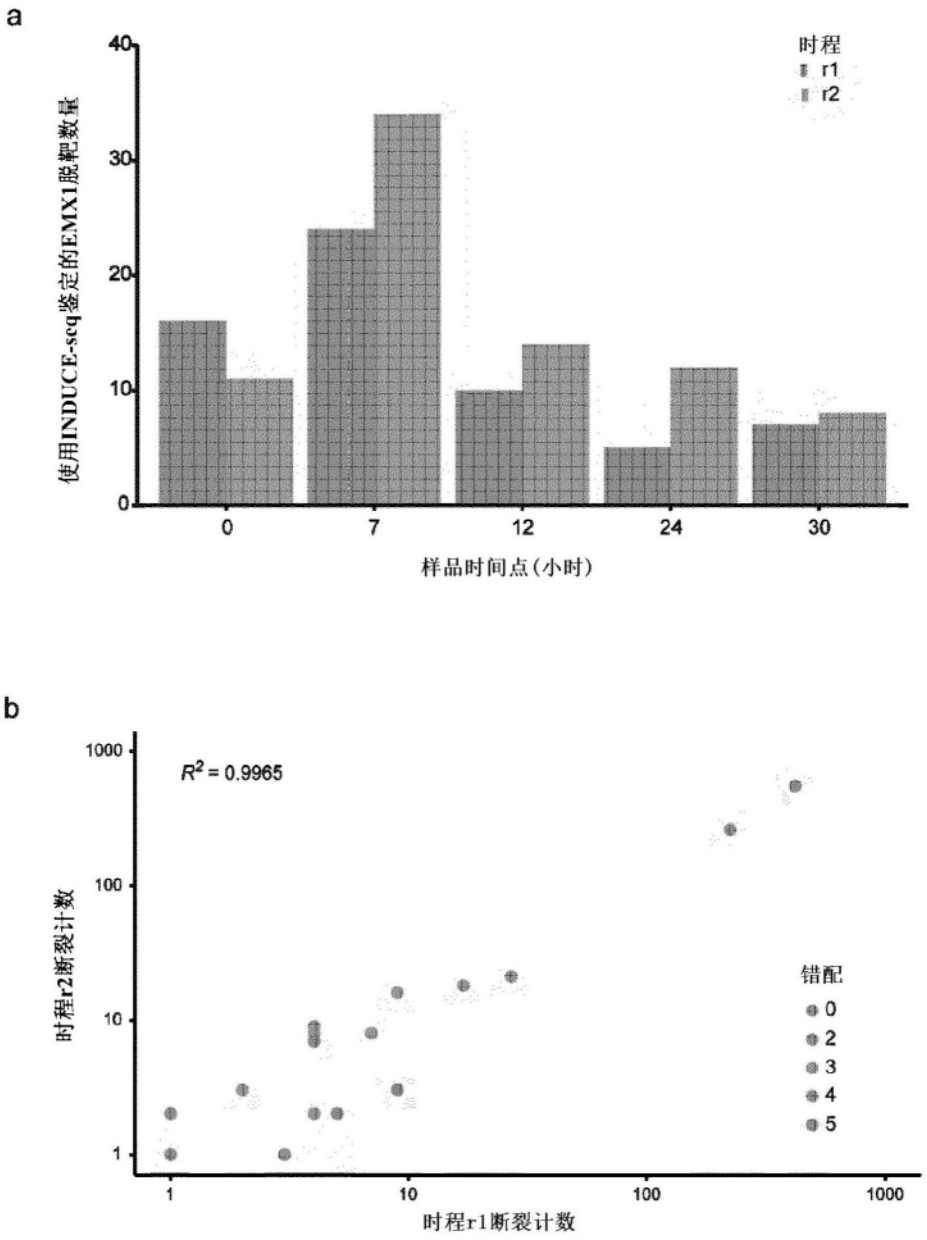

a
使用INDUCE-seq鉴定的EMX1脱靶数量
40
30
20
10
0
时程
r1
r2
0
7
12
24
30
样品时间点(小时)
b
时程r2断裂计数
1000
100
10
1
R^2 = 0.9965
错配
0
2
3
4
5
1
10
100
1000
时程r1断裂计数

图9

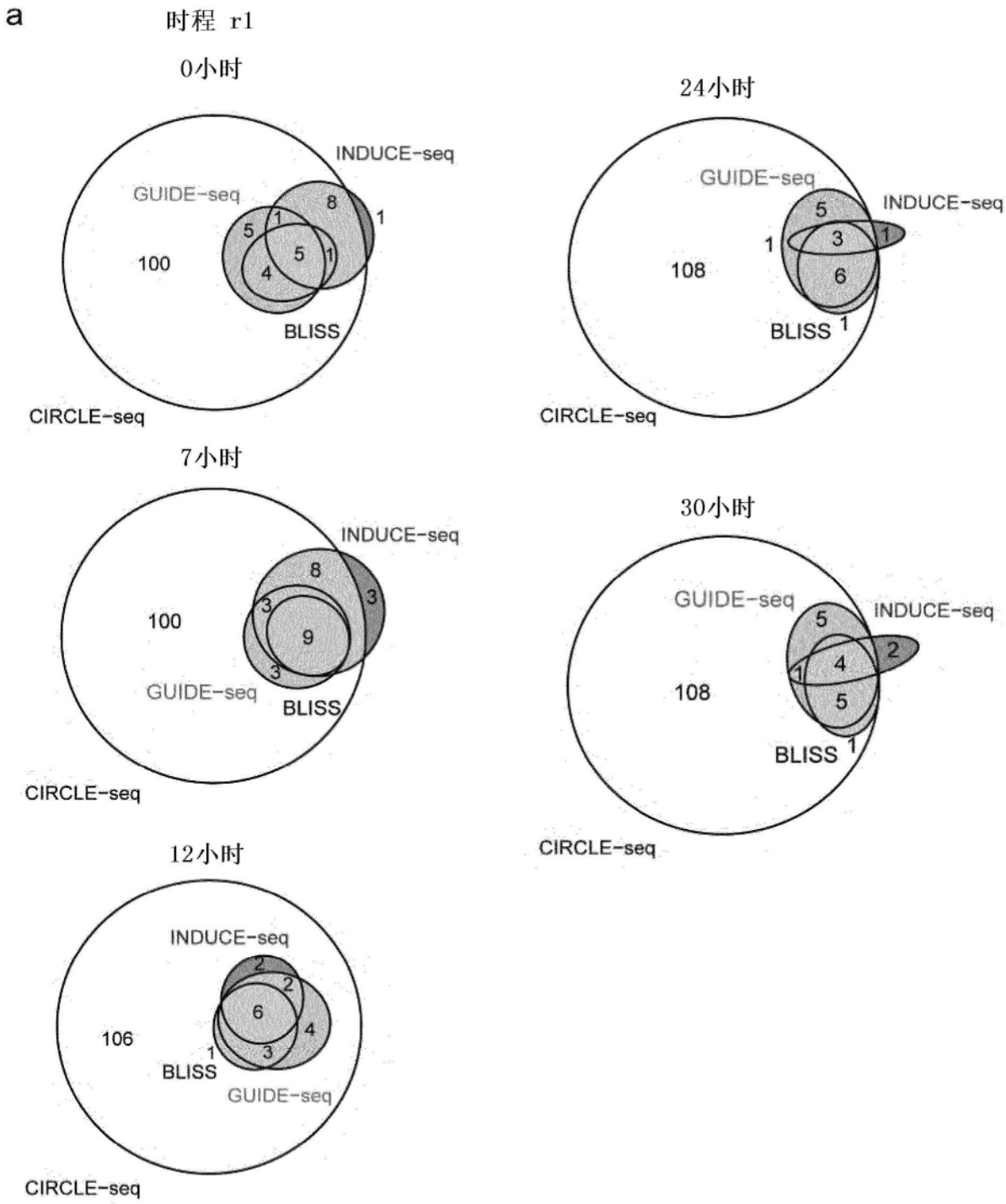

a
时程 r1
0小时
INDUCE-seq
GUIDE-seq
100
BLISS
CIRCLE-seq
24小时
GUIDE-seq
INDUCE-seq
108
BLISS
CIRCLE-seq
7小时
INDUCE-seq
100
GUIDE-seq
BLISS
CIRCLE-seq
30小时
GUIDE-seq
INDUCE-seq
108
BLISS
CIRCLE-seq
12小时
INDUCE-seq
106
BLISS
GUIDE-seq
CIRCLE-seq

图10

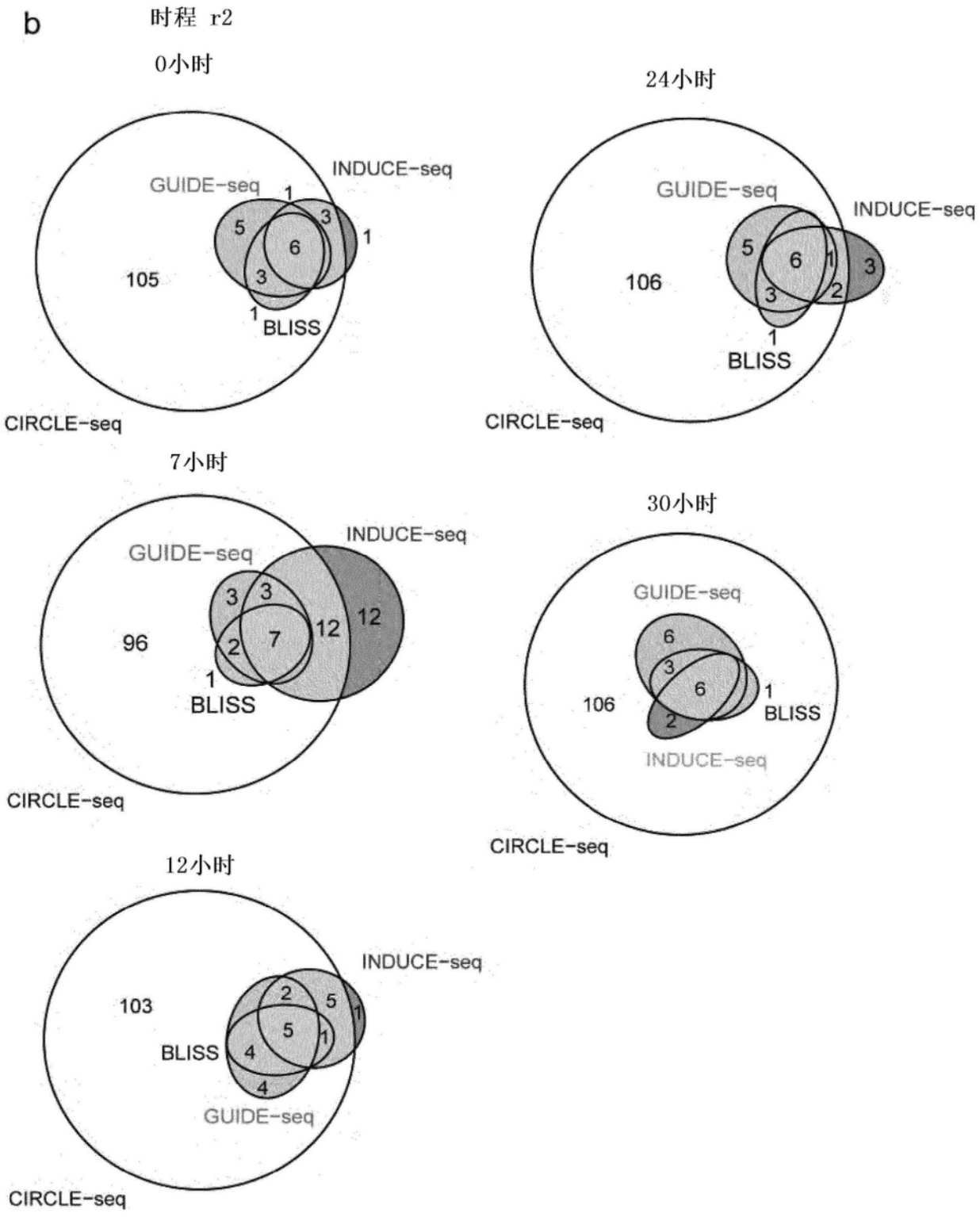
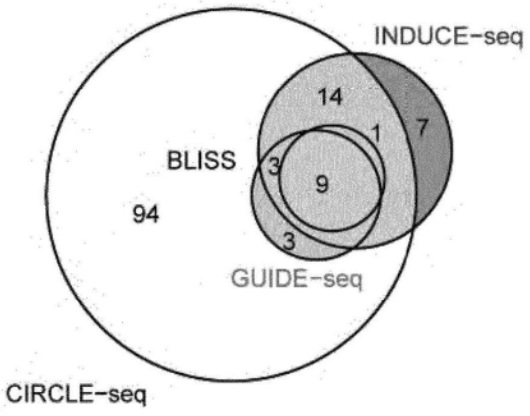
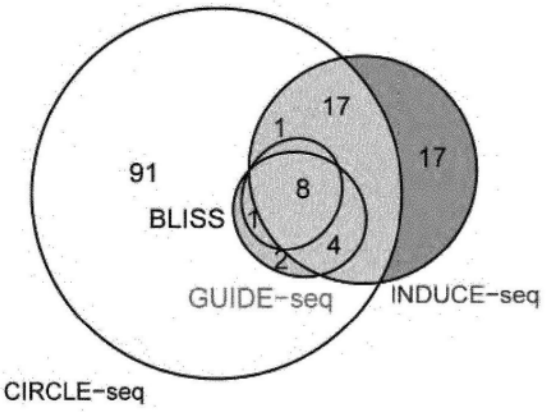

b
时程 r2
0小时
GUIDE-seq
INDUCE-seq
CIRCLE-seq
BLISS
105
24小时
106
7小时
96
30小时
103

图10(续)

c 时程r1 0小时-30小时



d 时程r2 0小时-30小时



e 时程r1和r2 0小时-30小时

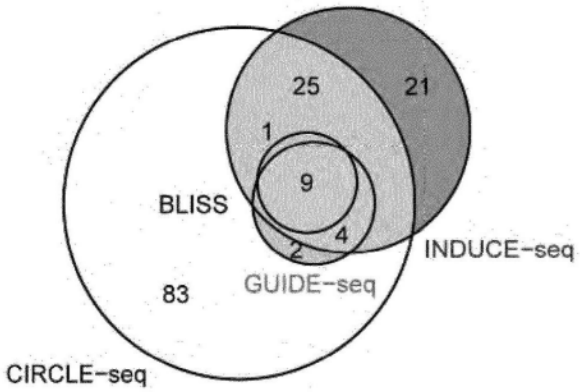


图10(续)

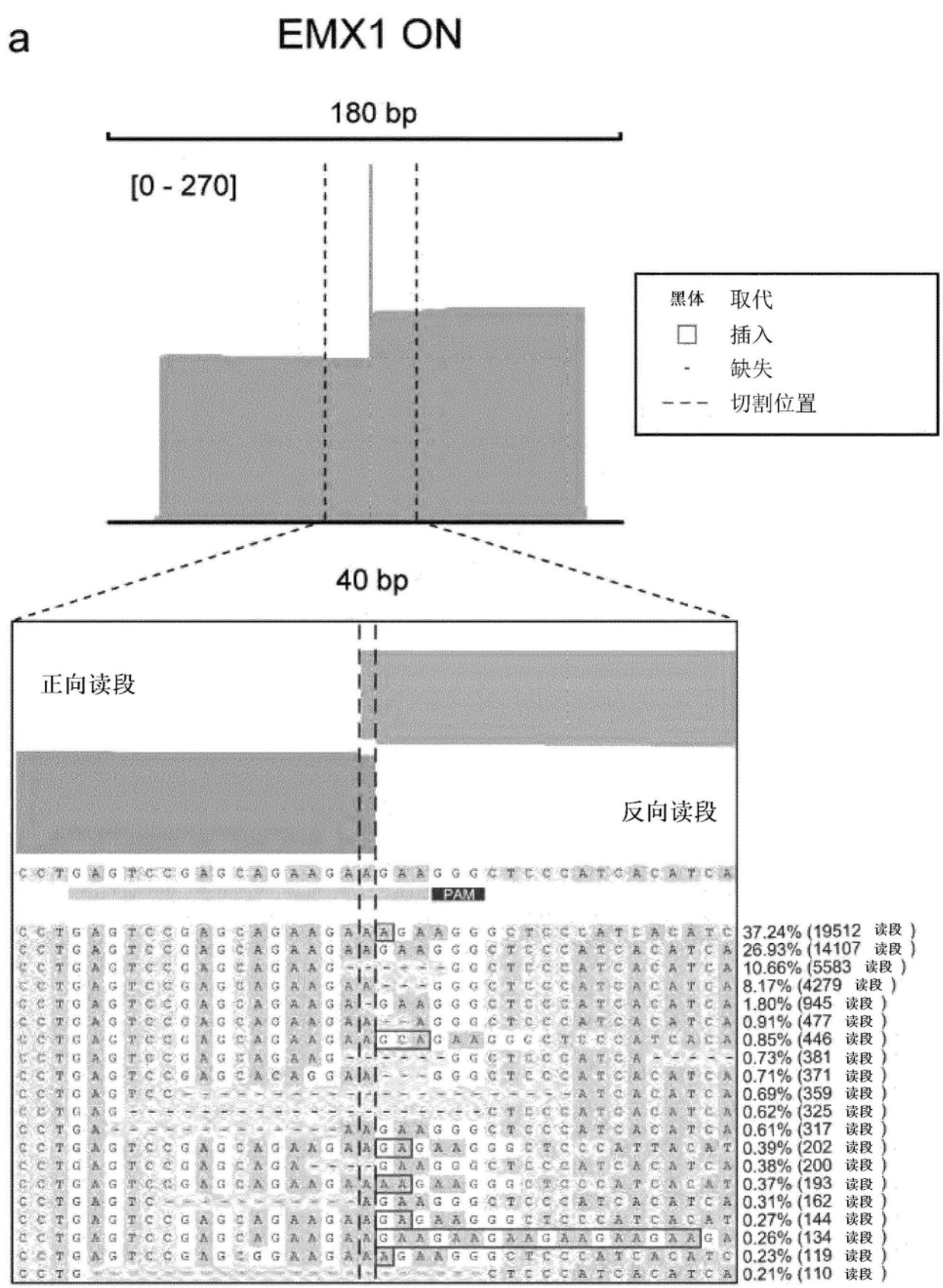

a
EMX1 ON
180 bp
[0 - 270]
黑体 取代
插入
缺失
切割位置
40 bp
正向读段
反向读段
PAM
37.24% (19512 读段)
26.93% (14107 读段)
10.66% (5583 读段)
8.17% (4279 读段)
1.80% (945 读段)
0.91% (477 读段)
0.85% (446 读段)
0.73% (381 读段)
0.71% (371 读段)
0.69% (359 读段)
0.62% (325 读段)
0.61% (317 读段)
0.39% (202 读段)
0.38% (200 读段)
0.37% (193 读段)
0.31% (162 读段)
0.27% (144 读段)
0.26% (134 读段)
0.23% (119 读段)
0.21% (110 读段)

图11

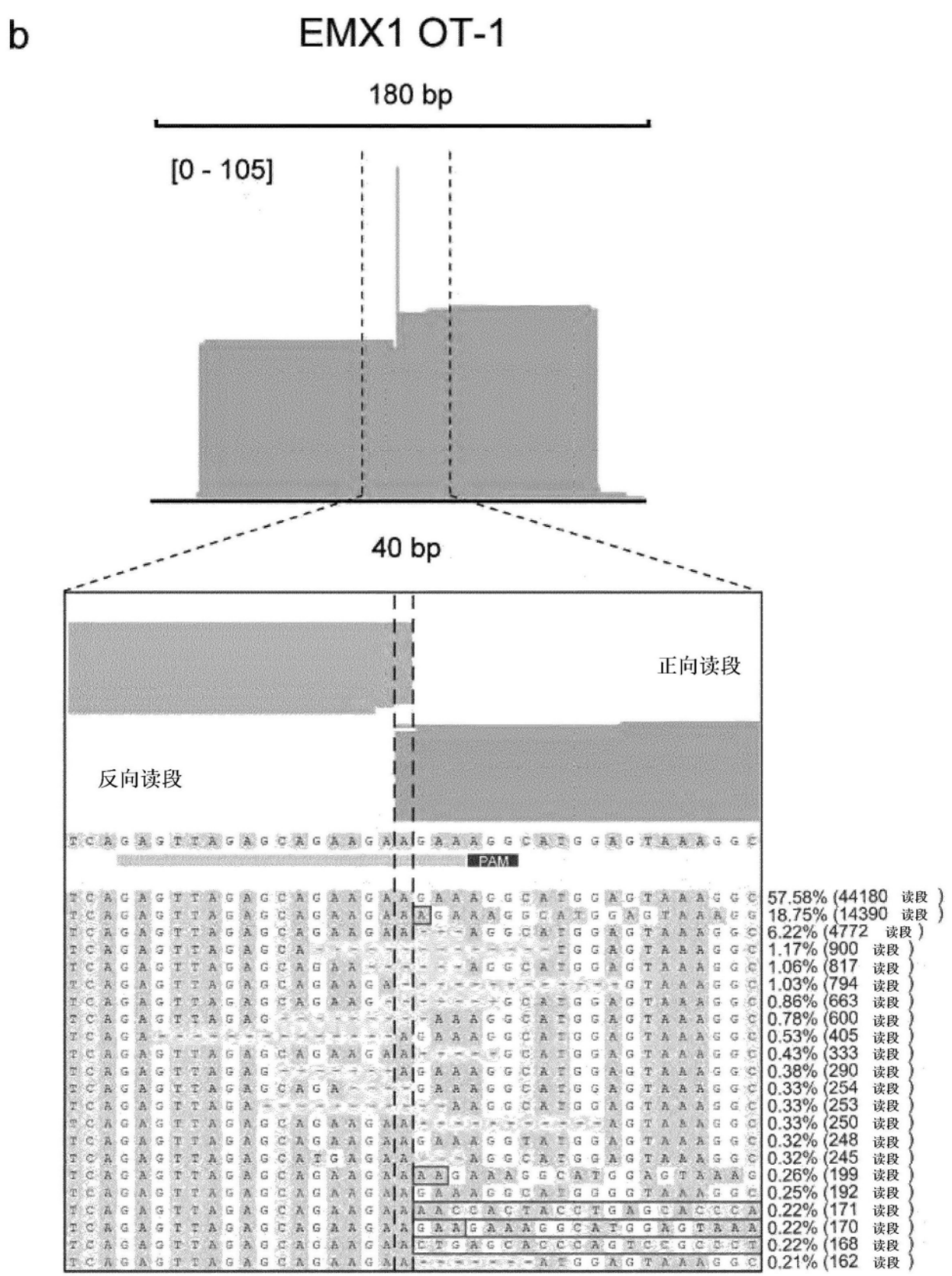

b
EMX1 OT-1
180 bp
[0 - 105]
40 bp
正向读段
反向读段
PAM
57.58% (44180 读段)
18.75% (14390 读段)
6.22% (4772 读段)
1.17% (900 读段)
1.06% (817 读段)
1.03% (794 读段)
0.86% (663 读段)
0.78% (600 读段)
0.53% (405 读段)
0.43% (333 读段)
0.38% (290 读段)
0.33% (254 读段)
0.33% (253 读段)
0.33% (250 读段)
0.32% (248 读段)
0.32% (245 读段)
0.26% (199 读段)
0.25% (192 读段)
0.22% (171 读段)
0.22% (170 读段)
0.22% (168 读段)
0.21% (162 读段)

图11(续)

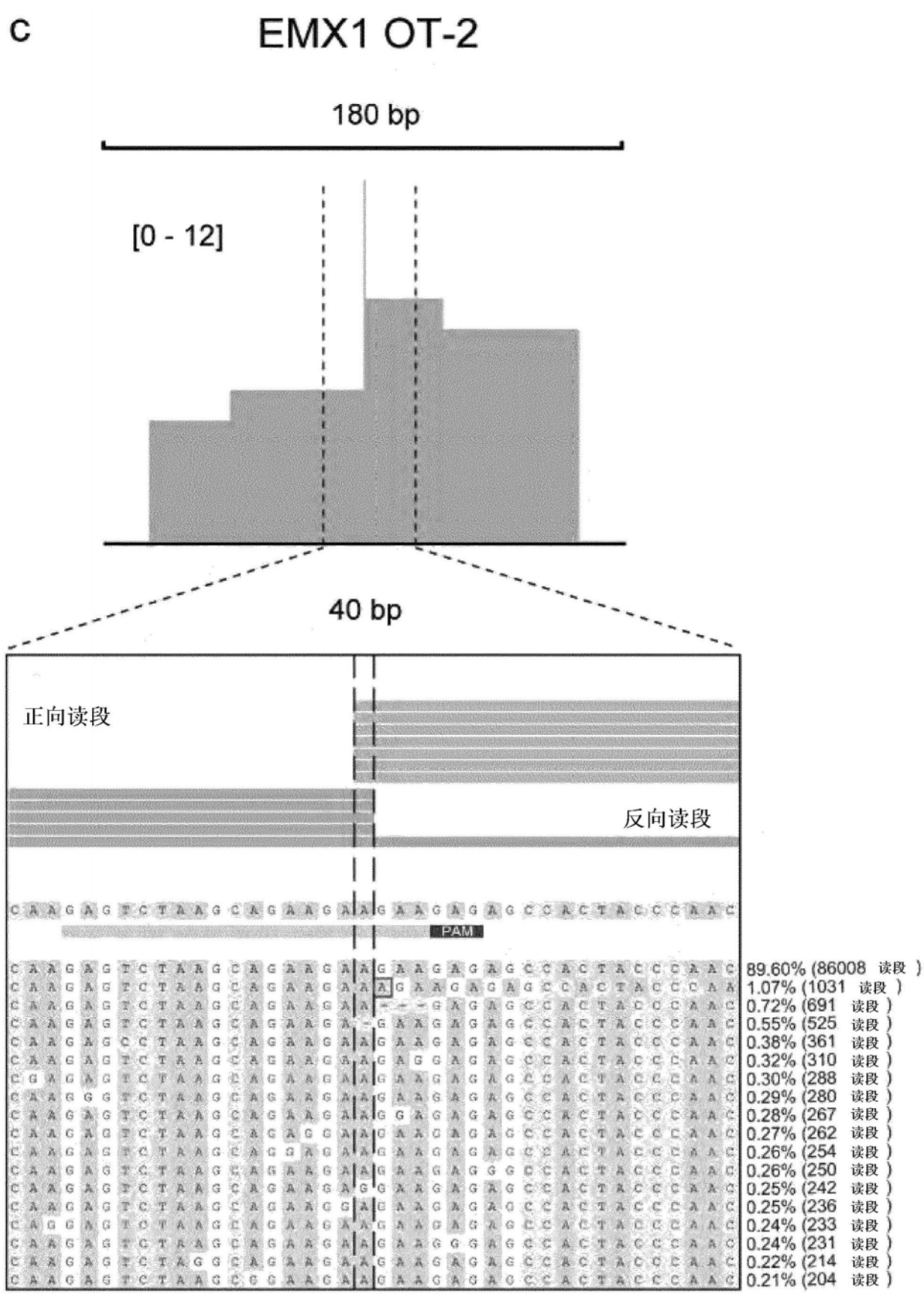

c
EMX1 OT-2
180 bp
[0 - 12]
40 bp
正向读段
反向读段
PAM
89.60% (86008 读段)
1.07% (1031 读段)
0.72% (691 读段)
0.55% (525 读段)
0.38% (361 读段)
0.32% (310 读段)
0.30% (288 读段)
0.29% (280 读段)
0.28% (267 读段)
0.27% (262 读段)
0.26% (254 读段)
0.26% (250 读段)
0.25% (242 读段)
0.25% (236 读段)
0.24% (233 读段)
0.24% (231 读段)
0.22% (214 读段)
0.21% (204 读段)

图11(续)

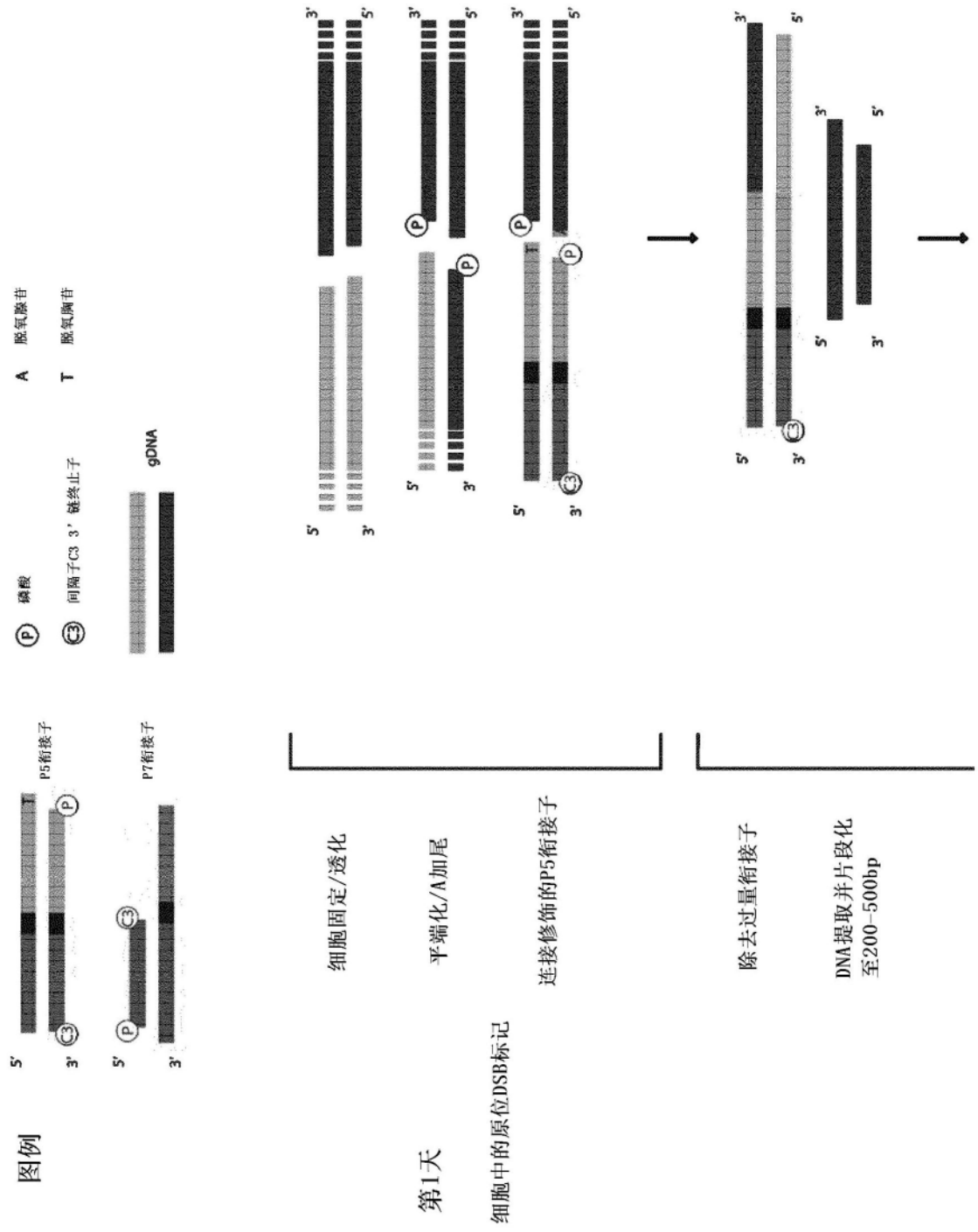

图例
5'
3'
P5衔接子
P7衔接子
磷酸
间隔子C3 3' 链终止子
A 脱氧腺苷
T 脱氧胸苷
gDNA
细胞中的原位DSB标记
第1天
细胞固定/透化
平端化/A加尾
连接修饰的P5衔接子
除去过量衔接子
DNA提取并片段化
至200-500bp

图12

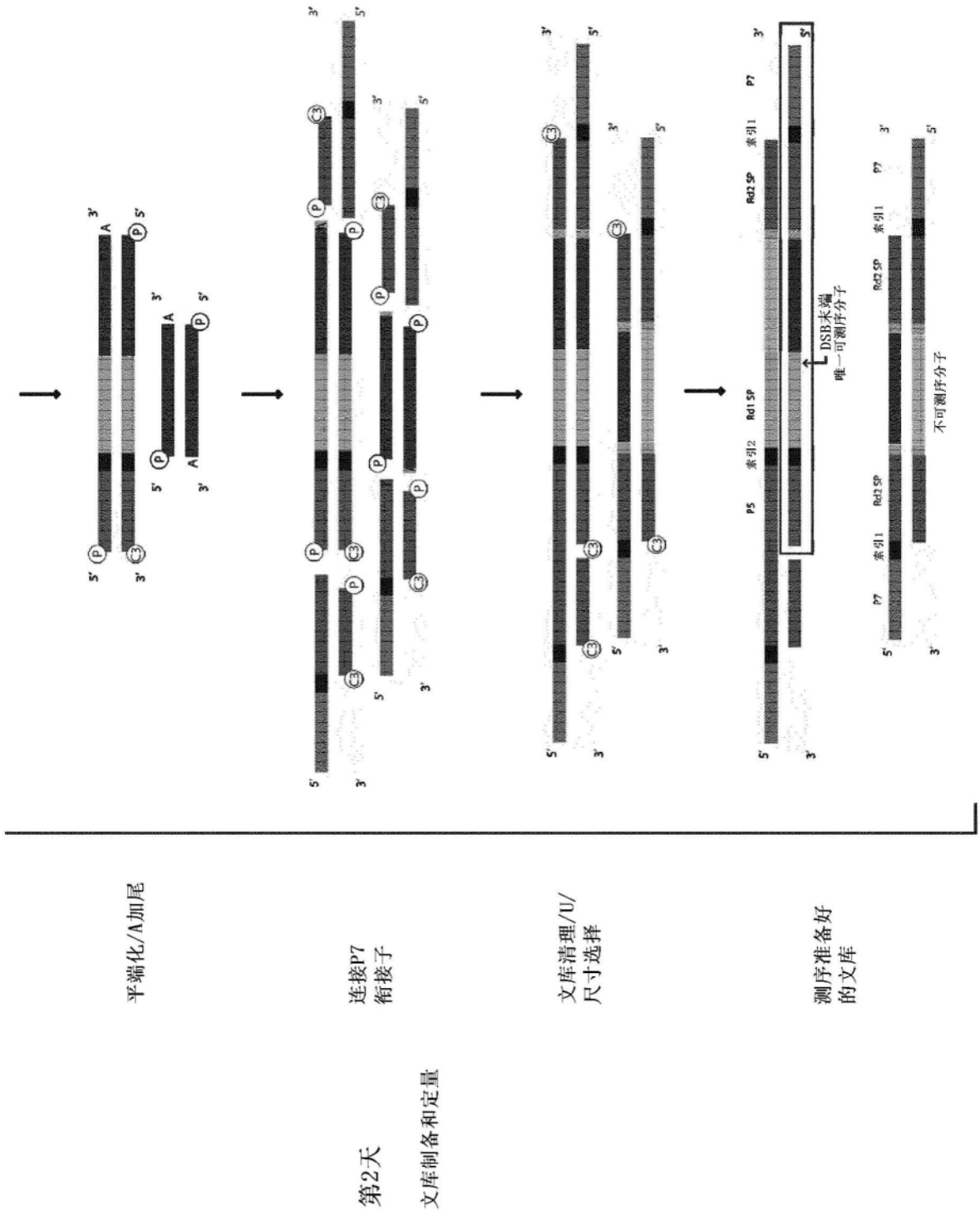

平端化/A加尾
连接P7
衔接子
文库清理/U/
尺寸选择
测序准备好
的文库
第2天
文库制备和定量
DSB末端
唯一可测序分子
不可测序分子

图12(续)

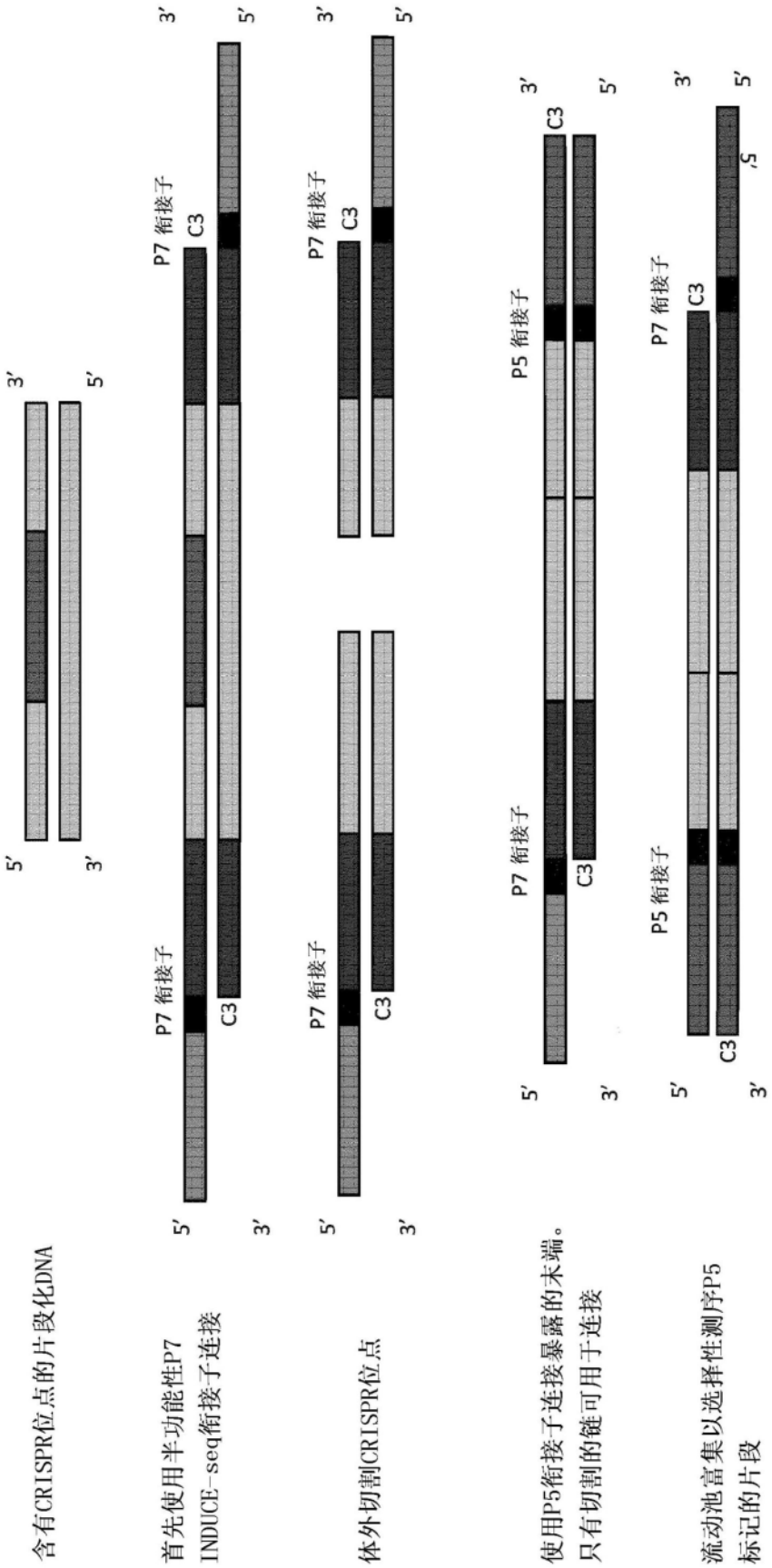

含有CRISPR位点的片段化DNA
首先使用半功能性P7
INDUCE-seq衔接子连接
体外切割CRISPR位点
使用P5衔接子连接暴露的末端。
只有切割的链可用于连接
流动池富集以选择性测序P5
标记的片段
P7 衔接子
C3
P5 衔接子
5′
3′

图13

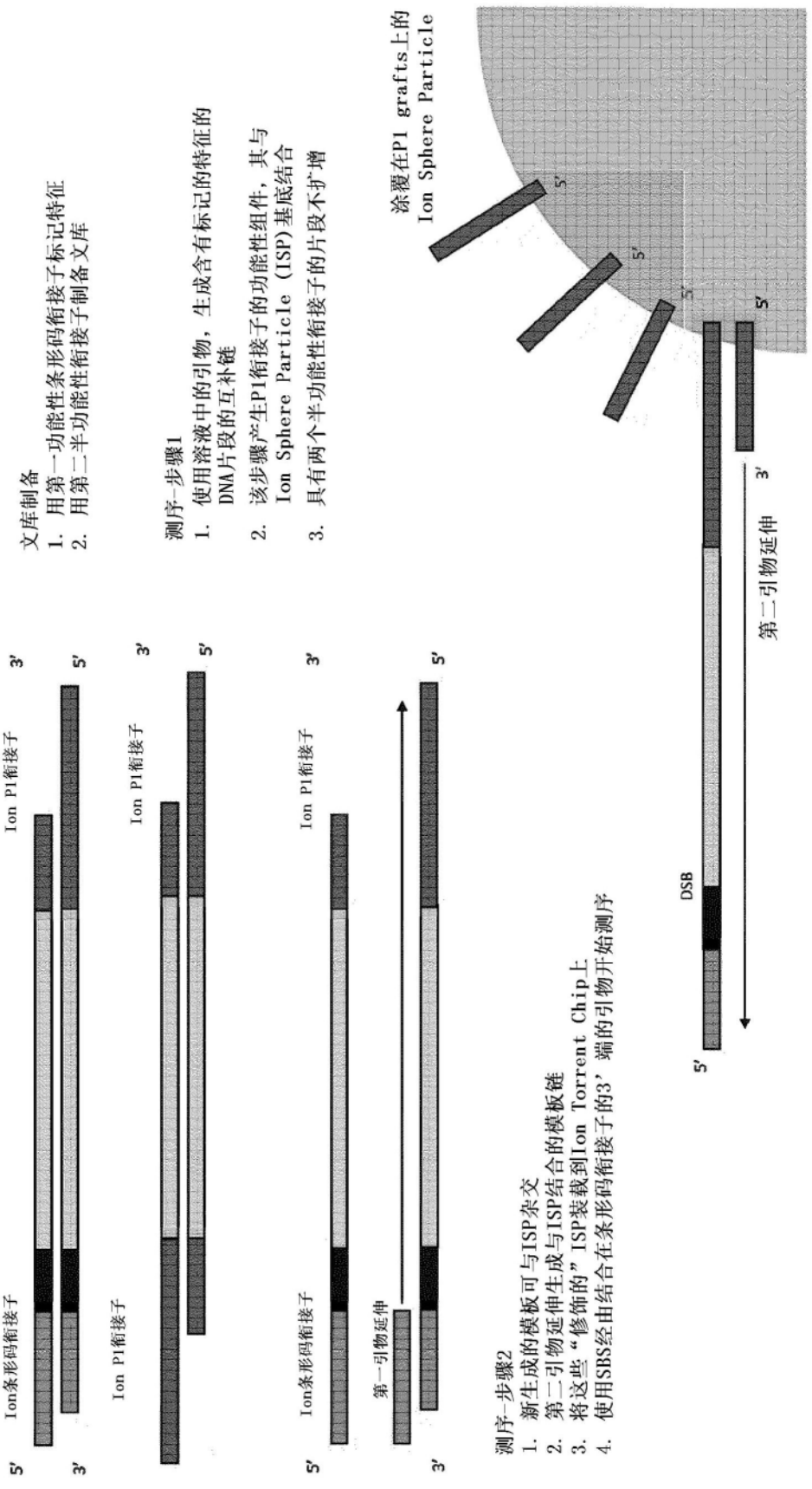

文库制备
1. 用第一功能性条形码衔接子标记特征
2. 用第二半功能性衔接子制备文库
测序-步骤1
1. 使用溶液中的引物，生成含有标记的特征的DNA片段的互补链
2. 该步骤产生P1衔接子的功能性组件，其与Ion Sphere Particle (ISP)基底结合
3. 具有两个半功能性衔接子的片段不扩增
涂覆在P1 grafts上的Ion Sphere Particle
Ion条形码衔接子
Ion P1衔接子
第一引物延伸
第二引物延伸
DSB
5'
3'
测序-步骤2
1. 新生成的模板可与ISP杂交
2. 第二引物延伸生成与ISP结合的模板链
3. 将这些“修饰的”ISP装载到Ion Torrent Chip上
4. 使用SBS经由结合在条形码衔接子的3’端的引物开始测序

图14

INDUCE-seq Torrent衔接子

Ion Xpress™条形码1（黑体下划线的条形码的5’-3’取向）inc T尾

引物1 CCATCTCATCCCT-G-CGTGTCTCC (SEQ ID NO: 106)

5’ CCATCTCATCCCT*G*CGTGTCTCCGACTCAGCTAAGGTAACGAT*T P-NNNNNNNNNNNNNNNNNNNN-A 3’ (SEQ ID NO: 107)

3’ /3SpC3/GGTAGAGTAGGGA*C*GCACAGAGGCTGAGTCGATTCCATTGCTA-P A-NNNNNNNNNNNNNNNNNNNN-P 5’ (SEQ ID NO: 108)

^ DSB位置

引物2

与ISP graft相同

半功能性Ion P1衔接子(inc T尾)

缺少的序列

P-ATCACCGACTGCCCATAGAGA/3SpC3/ 3’ (SEQ ID NO: 109)

T*TAGTGGCTGACGGGTATCTCTCCTTTCGCCTCCGCATCACCCAA 5’ (SEQ ID NO: 110)

CCTTTCGCCTCCGCATCACCCAA (SEQ ID NO: 111)

图15

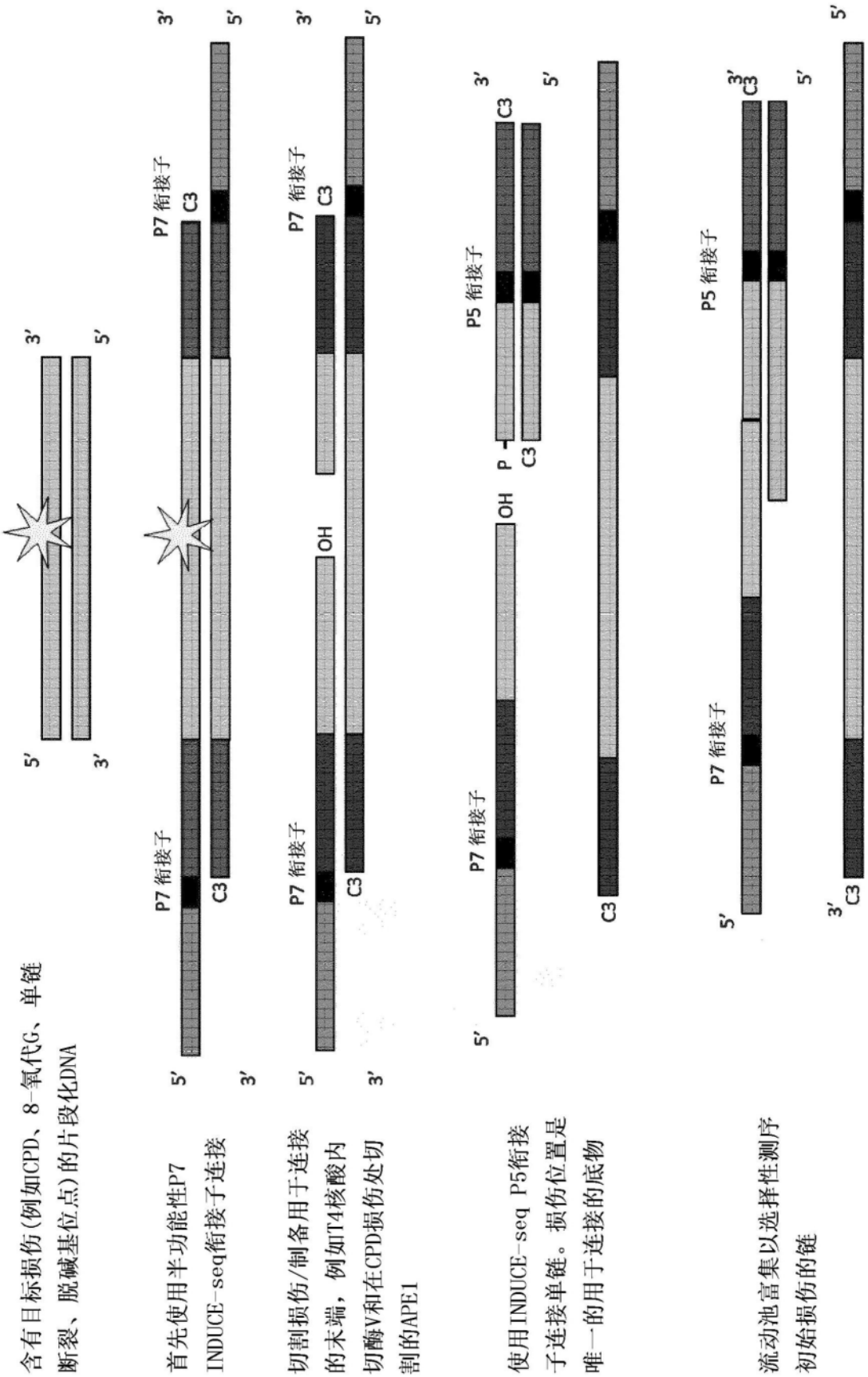

含有目标损伤(例如CPD、8-氧代G、单链断裂、脱碱基位点)的片段化DNA
5′
3′
3′
5′
首先使用半功能性P7 INDUCE-seq衔接子连接
P7 衔接子
C3
5′
3′
3′
5′
切割损伤/制备用于连接的末端，例如T4核酸内切酶V和在CPD损伤处切割的APE1
P7 衔接子
C3
OH
5′
3′
使用INDUCE-seq P5衔接子连接接单链。损伤位置是唯一的用于连接的底物
P5 衔接子
P7 衔接子
C3
OH
P
C3
5′
3′
流动池富集以选择性测序初始损伤的链
P5 衔接子
P7 衔接子
C3
5′
3′
3′
5′
C3

图16

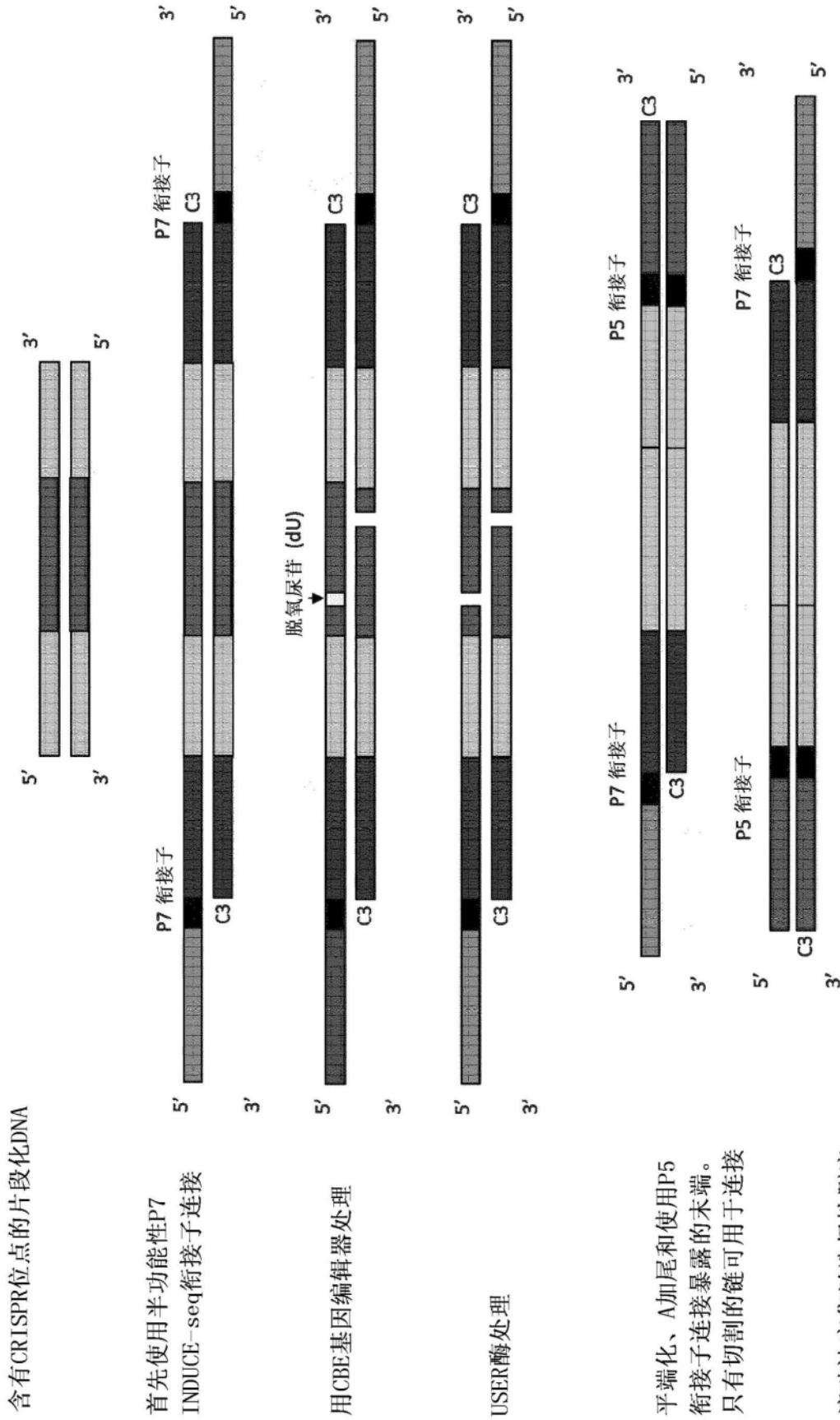

含有CRISPR位点的片段化DNA
首先使用半功能性P7
INDUCE-seq衔接子连接
用CBE基因编辑器处理
USER酶处理
平端化、A加尾和使用P5
衔接子连接暴露的末端。
只有切割的链可用于连接
流动池富集以选择性测序
P5标记的片段
P7 衔接子
P5 衔接子
C3
脱氧尿苷 (dU)
5′
3′

图17

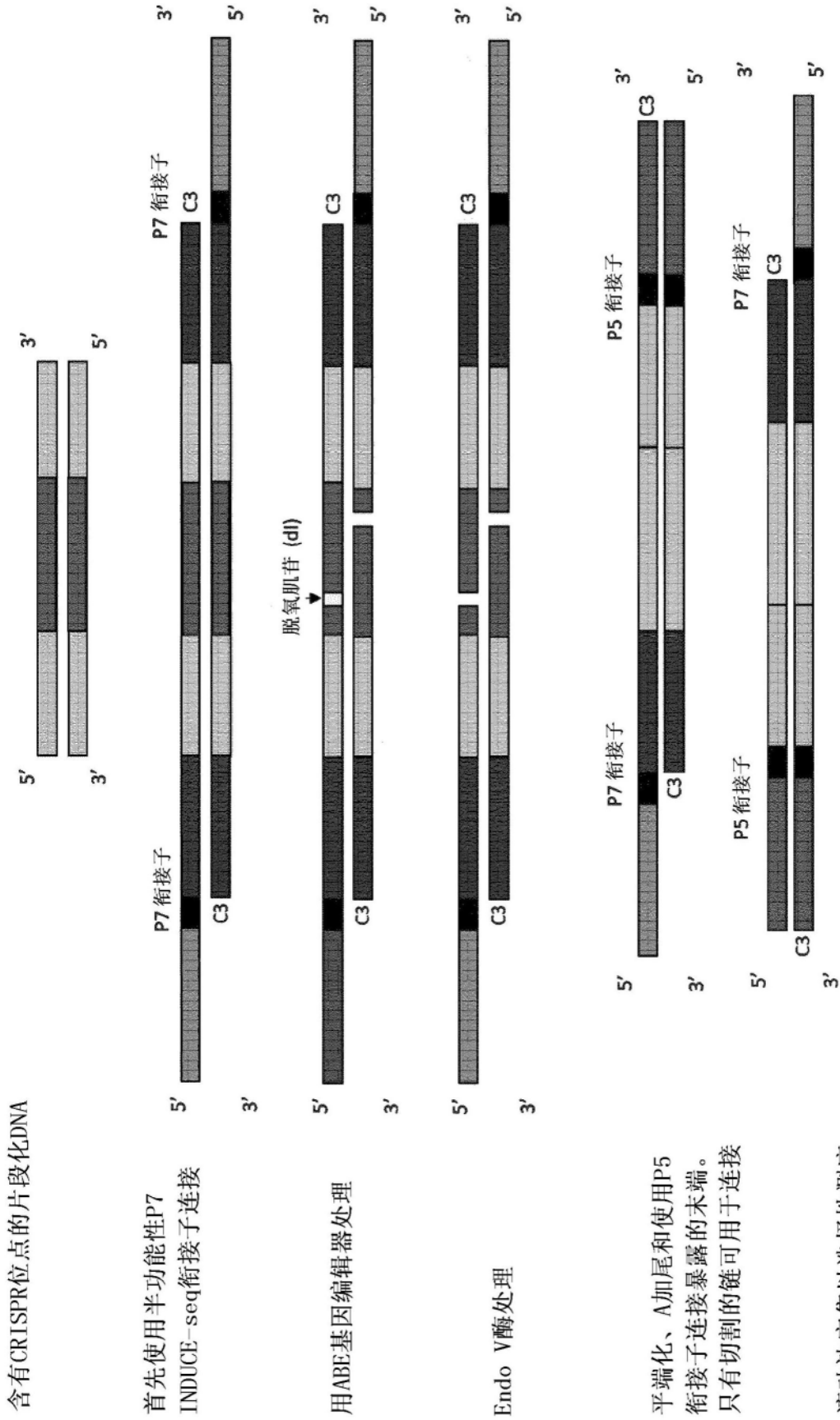

含有CRISPR位点的片段化DNA
5′
3′
3′
5′
首先使用半功能性P7
INDUCE-seq衔接子连接
P7 衔接子
C3
5′
3′
用ABE基因编辑器处理
脱氧肌苷 (dI)
Endo V酶处理
平端化、A加尾和使用P5
衔接子连接暴露的末端。
只有切割的链可用于连接
P5 衔接子
流动池富集以选择性测序
P5标记的片段

图18