(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 118302538 A
(43)申请公布日 2024.07.05

(21)申请号 202480000814.7

(22)申请日 2024.02.20

(66)本国优先权数据
PCT/CN2023/091367 2023.04.27 CN

(85)PCT国际申请进入国家阶段日
2024.04.26

(86)PCT国际申请的申请数据
PCT/CN2024/077716 2024.02.20

(71)申请人 上海茵创昕生物科技有限公司
地址 200000 上海市浦东新区中国(上海)自由贸易试验区临港新片区新杨公路860号10幢

(72)发明人 朱钧 白冰

(74)专利代理机构 北京植众德本知识产权代理有限公司 16083
专利代理师 关巍

(51)Int.Cl.
C12Q 1/6806(2006.01)

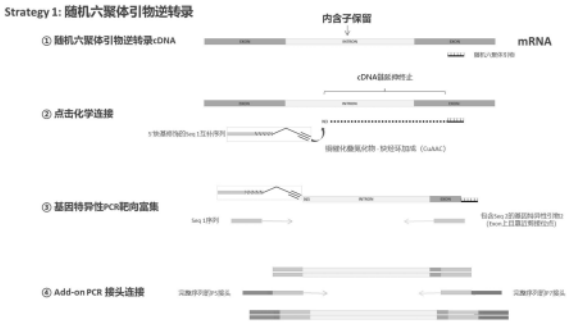
权利要求书2页 说明书25页
序列表（电子公布） 附图3页

(54)发明名称

一种用于检测剪接异构体的靶向高通量测序方法

(57)摘要

一种用于检测剪接异构体的靶向高通量测序方法,包括用于高通量测序的测序文库的建立方法,包括以下步骤:1)采用逆转录引物对样本RNA进行逆转录,加入普通dNTP和3’修饰的dNTP,反应得到cDNA第一链;2)通过点击化学反应将5’端带有炔基修饰的寡核苷酸片段与步骤1)获得的cDNA片段连接;3)以步骤2)的反应产物为模板,PCR扩增;4)获得测序文库。靶向富集引物可在逆转录或PCR扩增步骤中引入。富集步骤反应体系中含有多个基因特异性引物,且每个引物基于该转录本中选择性剪接事件的下游外显子区段设计,保证逆转录产生的随机长度片段覆盖剪接位点。



CN 118302538 A

1.一种用于高通量测序的测序文库的建立方法,其包括以下步骤:

(1)采用逆转录引物对样本RNA进行逆转录,加入普通dNTP和3’修饰的dNTP,反应得到cDNA第一链,所述3’修饰的dNTP选自以下的一种或多种:AzNTP、AmNTP、propargyl-NTP、HalNTP;

(2)通过点击化学反应将5’端带有炔基修饰的寡核苷酸片段与步骤(1)获得的cDNA片段连接,所述寡核苷酸片段包含随机序列和通用测序引物1(seq1)的互补区段序列;

(3)以步骤(2)的反应产物为模板,进行PCR反应扩增;

(4)获得测序文库。

2.根据权利要求1所述的方法,其中,步骤(1)中所述的逆转录引物是随机引物或基于靶向基因的保留内含子下游外显子设计的基因特异性引物群1;

任选地,所述基因特异性引物群1中的引物5’端带有通用测序引物2序列(seq2)。

3.根据权利要求1或2所述的方法,所述步骤(3)采用的PCR扩增引物包括通用测序引物1和基因特异性引物群2,其基因特异性引物群2中的引物的5’端带有通用测序引物2序列。

4.根据权利要求2或3所述的方法,基因特异性引物群1和2均是基于特定靶向基因的保留内含子选择性剪接事件下游的外显子设计,其中,基因特异性引物2的靶向位点比基因特异性引物1的位点向上游上移5-100个碱基,优选为20-50个碱基。

5.根据权利要求2-4任一项所述的方法,基因特异性引物群1和基因特异性引物群2分别靶向的目标基因的个数大于等于1。

6.根据权利要求1-5任一项所述的方法,其中,步骤(1)中加入普通dNTP和3’修饰的dNTP的摩尔浓度比例为1:1-1:100。

7.根据权利要求1-6任一项所述的方法,其中,步骤(1)中加入普通dNTP和3’修饰的dNTP的摩尔浓度比例为1:20。

8.根据权利要求1-7任一项所述的方法,其中,步骤(2)中所述随机序列包含4-16个核苷酸。

9.根据权利要求8任一项所述的方法,其中,步骤(2)中所述随机序列为5个核苷酸的长度。

10.根据权利要求1-9任一项所述的方法,其中步骤(3)的PCR扩增包括采用通用测序引物1和基因特异性引物群2对步骤(2)的反应产物进行PCR扩增后,再对扩增产物添加接头结构的步骤,所述接头结构包括P5/P7接头和核酸条形码,所述核酸条形码连接在引物P5和/或P7接头端。

11.根据权利要求1所述的方法,所述步骤(3)为:以步骤(2)的反应产物为模板,添加P5和P7接头进行PCR反应。

12.根据权利要求1-11任一项所述的方法,其中,所述通用测序引物1序列选自SEQ ID NO:3、SEQ ID NO:5或SEQ ID NO:6中的任一项;优选地,所述通用测序引物1的互补区段序列如SEQ ID NO:4所示。

13.根据权利要求1-11任一项所述的方法,其中,所述通用测序引物2序列(seq2)选自SEQ ID NO:7或SEQ ID NO:8中的任一项。

14.根据权利要求1-13任一项所述的方法构建的高通量测序文库。

15.一种用于检测剪接异构体的靶向高通量测序方法,其包括以下步骤:

(1)提取样本RNA,根据权利要求1-13中任一项所述的方法构建测序文库;

(2)基于上述测序文库,高通量测序获得所述样本中目标基因的测序信息。

16.根据权利要求15所述的方法,其中,所述样本选自如下的细胞样本:MCF10A、MCF7、HeLa、HEK293T和/或MDA-MB-231。

17.根据权利要求15或16所述的方法,其中,所述目标基因包括ATP13A1、CXXC1、ECHDC2、FGFRL1、HMGN3、KLHL17、NAXD、LZTR1、SELENBP1、JMJD8、PSMB1、HIGD2A、HNRNPAB、SMARCC1、ATP5IF1、HIGD2B、RPS21、UQCC5、NFATC3、PCNP和/或OSGEP。

18.权利要求1-13任一项所述的用于高通量测序的测序文库的建立方法、权利要求14所述的高通量测序文库和/或权利要求15-17所述的靶向高通量测序方法在脱靶事件评估中的用途;

优选地,所述脱靶事件评估包括以下:

(1)精准确定反式剪接因子中发生脱靶反式剪接所在的基因组位置;

(2)对目标反式剪接和脱靶反式剪接进行定量分析。

一种用于检测剪接异构体的靶向高通量测序方法

[0001] 相关申请的交叉引用

[0002] 本申请要求于2023年4月27日提交的PCT国际申请PCT/CN2023/091367号的优先权,本申请引用上述PCT国际申请的全文。

技术领域

[0003] 本发明涉及生物技术领域,具体地,本发明涉及一种用于检测剪接异构体的靶向高通量测序方法。

背景技术

[0004] 可变剪接(Alternative Splicing,AS)是指从一个mRNA前体中通过不同的剪接方式(选择不同的剪接位点组合)产生不同的mRNA剪接异构体的过程。可变剪接的发生过程中存在一定程度上的灵活性,最终mRNA的结构可选择性地包含前体RNA的部分外显子和/或内含子片段,从而使蛋白质在序列组成上产生变化。因此,可变剪接是蛋白质组多样性的重要机制。

[0005] 在基因表达过程中,基因转录首先合成pre-mRNA,pre-mRNA包含所有内含子和外显子以及两端非编码区域(5'and 3'Untranslated region)。RNA剪接是pre-mRNA成熟为mRNA的一个必不可少的步骤。在剪接体的作用下切除pre-mRNA中内含子(intron),并连接所有外显子和UTR序列。当发生可变剪接时,同一种pre-mRNA可生成两种或以上的剪接异构体。可变剪接可归纳为五种形式,其中最为常见的两种是内含子保留(intron retention)和外显子跳越(exon skipping)。如下图1所示,内含子保留是在pre-mRNA剪接过程中,内含子没有被切除,在mRNA上仍然保留该内含子;外显子跳越则是在pre-mRNA剪接过程中,外显子被剪接体剪切去除,导致后续生成的mRNA有外显子缺失。

[0006] 目前市面上可提供的商业化RNA建库试剂盒主要以不同片段化方法来构建总mRNA测序文库,无针对剪接异构体的靶向,尤其是未知剪接异构体的靶向富集建库方法。主要原因在于,经典RNA-seq是将总RNA(total RNA)或信使RNA(messenger RNA,mRNA)随机碎片化后,通过随机引物random hexamer和/或oligo(dT)引物逆转录形成双链DNA并完成后续接头连接建库。上述方法只针对全转录组,不能目标性针对某一感兴趣的小部分转录组进行分析,同时成本高昂。由于RNA-seq获得片段位置随机,通常不在剪接交叉处,需要极高的测序深度才能满足准确检测不同剪接异构体(splicing isoforms),尤其对于众多待验证的潜在选择性剪接位点时,因此对于测序的成本是极其巨大的。对于目前市面上可提供的靶向RNA测序商业试剂盒均为已知靶点设计开发的试剂盒,无法满足对未知剪接异构体的靶点测序要求。同时,现有文献报道的靶向RNA测序方法,也难以应用与企业和非科研用途。主要原因在于,需要依托商业公司如Agilent Technologies、IDT等设计、合成用于靶向富集的杂交探针,这一过程通常耗时较长需不断优化,且无法灵活改变(增加、删减)已成型的探针。对于实时发现的新潜在靶点不能及时、高效地完成验证,同时费用成本巨大。即便对于优化成熟的靶向富集体系,建库的运转周期也通常至少需要2天,且动手步骤操作复杂。另

外,做剪接异构体验证时的一个重点是应尽可能包含所有的异构体(isoform),但目前的所有方法对此都可能无法完善。另外一个主要因素是现有方法(Agilent)只能对已知剪接异构体进行分析:引物设计的需求,不同剪接异构体间扩增引物存在相互竞争的可能,即便在模板高度稀释的单微滴情况下。低通量剪接异构体的定量分析的经典方法是qRT-PCR,但不同剪接异构体间不同扩增效率会引起荧光信号的系统性偏差,即便通过繁琐的校准步骤也只能实现相对定量;并且在检测低丰度的剪接异构体时对引物设计、实验操作和设备要求较高,定量结果难以重复。

[0007] 因此开发一款更为高效、精确的RNA多重靶向富集建库方法尤为重要与迫切。

发明内容

[0008] 为了解决现有技术中存在的问题,本公开的目的在于提供一种用于检测剪接异构体的靶向高通量测序方法。

[0009] 为了实现上述目的,本公开采用以下具体技术方案:

[0010] 在一方面,本公开提供了一种用于高通量测序的测序文库的建立方法,其包括以下步骤:

[0011] (1)采用逆转录引物对样本RNA进行逆转录,加入普通dNTP和3’修饰的dNTP,反应得到cDNA第一链,所述3’修饰的dNTP选自以下的一种或多种:AzNTP、AmNTP、propargyl-NTP、HalNTP;

[0012] (2)通过点击化学反应将5’端带有炔基修饰的寡核苷酸片段与步骤(1)获得的cDNA片段连接,所述寡核苷酸片段包含随机序列和通用测序引物1(seq1)的互补区段序列;

[0013] (3)以步骤(2)的反应产物为模板,进行PCR反应扩增;

[0014] (4)获得测序文库。

[0015] 在另一方面,本公开提供了一种根据上述方法构建的高通量测序文库。

[0016] 在另一方面,本公开提供了一种用于检测剪接异构体的靶向高通量测序方法,其包括以下步骤:

[0017] (1)提取目标细胞RNA,根据上述方法构建测序文库;

[0018] (2)基于上述测序文库,高通量测序获得所述细胞样本中目标基因的测序信息。

[0019] 在另一方面,本公开提供了一种利用前述用于高通量测序的测序文库的建立方法、前述的高通量测序文库和/或前述的靶向高通量测序方法在脱靶事件评估中的用途;

[0020] 优选地,所述脱靶事件评估包括以下:

[0021] (1)精准确定反式剪接因子中发生脱靶反式剪接所在的基因组位置;

[0022] (2)对目标反式剪接和脱靶反式剪接进行定量分析。

[0023] 本公开所取得的有益效果至少如下:

[0024] (1)实现多重靶向剪接异构体的分析。传统qRT-PCR方法的通量受限于检测仪器的荧光数目、荧光间的串色效应(bleed through)、以及反应体系中引物间相互干扰。目前常用qRT-PCR的通量为四重,而本公开的HTAS(High-throughput Targeted Alternative Splicing)分析平台可以同时分析5-100(也可更多)pre-mRNA剪接事件;

[0025] (2)实现剪接异构体的精准定量:高通量测序在单分子水平(数字信号)对剪接异构体进行定量分析,而传统qRT-PCR定量依赖于荧光强度(模拟信号)。模拟信号只支持相对

定量,数字信号可实现样品剪接异构体的绝对定量。另外,HTAS的独特优点为逆转录过程中加入一定比例的3’修饰dNTP可随机终止第一链cDNA的延伸,因此同一类型剪接异构体可生成不同长度cDNA产物,在建库过程中不再受到剪接异构体PCR产物长度不同对扩增效率的干扰。最后,在5’接头的设计中引入随机序列(N5;最多12-16个核苷酸)可去除PCR扩增过程中的系统性偏差(PCR skewing);混合样本实验证明HTAS实际检出值和混合样本理论值相关性r>0.99;

[0026] (2)检测灵敏度:实施例中检出最低内含子保留(intron-retained,IR)剪接异构体比例为0.1%。随着测序深度提高,检测IR检出灵敏度预期可达$1/10^5$或更低,其检测灵敏度远高于传统RNA-seq,尤其是针对低丰度转录本;

[0027] (4)可检出未知的剪接异构体:由于组织和细胞的多样性,目前的检测体系中仍有大量的剪接异构体没能在转录组中被注释。传统qRT-PCR方案主要针对已知剪接异构体进行设计,而HTAS平台可同时检出已知和未知的剪接异构体,去除异构体分析中由于注释缺失对结果的干扰,提高定量分析的特异性;

[0028] (5)实现反式剪接(trans-splicing)脱靶事件评估:反式剪接是一个新型技术用于mRNA的定向编辑,但脱靶现象是其在临床应用中主要瓶颈之一。由于脱靶事件预期发生在转录组中任意pre-mRNA,目前尚无方法对脱靶现象进行系统性评估。HTAS只需知道反式剪接分子(Pre-mRNA Trans-splicing Molecule)序列即可对目标(on-target)反式剪接和脱靶(off-target)反式剪接同时进行定性定量分析;

[0029] (6)操作简单,成本低:无需繁琐建库富集步骤,所有建库时间约4-5小时。单个样本成本大幅低于现有产品或方法。

附图说明

[0030] 图1为剪接异构的原理示意图。

[0031] 图2为高通量测序文库构建原理图(方案一)。

[0032] 图3为高通量测序文库构建原理图(方案二)。

[0033] 图4为高通量测序文库的构建原理图(方案三)。

[0034] 图5为不同混合比例的剪接异构体IVT产物实际检出值和混合样本理论值相关性示意图。

具体实施方式

[0035] I.定义

[0036] 在本发明中,除非另有说明,否则本文中使用的科学和技术名词具有本领域技术人员所通常理解的含义。并且,本文中所用的蛋白质和核酸化学、分子生物学、细胞和组织培养、微生物学、免疫学相关术语和实验室操作步骤均为相应领域内广泛使用的术语和常规步骤。同时,为了更好地理解本发明,下面提供相关术语的定义和解释。

[0037] 除非另有定义,本公开所使用的所有的技术和科学术语与属于本公开的技术领域的技术人员通常理解的含义相同。本公开的说明书中所使用的术语只是为了描述具体的实施例的目的,不用于限制本公开。

[0038] 本公开的术语“包括”和“具有”以及它们任何变形,意图在于覆盖不排他的包含。

例如包含了一系列步骤的过程、方法、装置、产品或设备没有限定于已列出的步骤或模块，而是可选地还包括没有列出的步骤，或可选地还包括对于这些过程、方法、产品或设备固有的其它步骤。

[0039] 在本公开中提及的“多个”是指两个或两个以上。“和/或”，描述关联对象的关联关系，表示可以存在三种关系，例如，A和/或B，可以表示：单独存在A，同时存在A和B，单独存在B这三种情况。字符“/”一般表示前后关联对象是一种“或”的关系。同时，为了更好地理解本公开，下面提供相关术语地定义和解释。如本文使用的和除非另作说明，术语“约”或“大约”是指在给定值或范围的加或减10％之内。在需要整数的情况下，该术语是指在给定值或范围的加或减10％之内、向上或向下舍入到最接近的整数。

[0040] 就多肽序列而言，短语“基本相同”可理解为表现出与参照多肽序列至少60％、65％、70％、75％、80％、85％、90％、95％、96％、97％、98％、99％或更多的序列同一性。就核酸序列而言，该术语可理解为表现出与参照核酸序列至少大于60％、65％、70％、75％、80％、85％、90％、95％、96％、97％、98％、99％或更高的序列同一性的核苷酸序列。

[0041] 本公开中使用的术语“Oligo(dT)引物”是指是约12-25个多聚胸腺嘧啶T组成的重复寡核苷酸序列，能够特异性地与真核生物mRNA的ploy(A)尾退火，故不适合缺少ploy(A)尾结构的RNA，如原核生物RNA或miRNA，也不适用于已降解的RNA，如FFPE样品中RNA。Oligo(dT)n VN是由固定且特异序列+(T)20左右+简并碱基VN组成，存在固定且特异的序列是为了便于设计通用的PCR下游引物。Oligo(dT)n VN序列中的VN指在3’端存在锚定碱基。(“V”代表dATP、dGTP或dCTP；“N”代表dATP、dTTP、dGTP、dCTP中的任意一种)。锚定碱基的作用是能够特异性地结合到Poly(A)的5’末端，以免逆转录出过多的T碱基。

[0042] 本公开中使用的术语“点击化学”在本领域中是众所周知的并且通常是指易于纯化和区特异性的快速反应。点击化学是一类允许所选择的底物与特异性分子连接的反应。点击化学不是单一的特异性反应，而是描述了一种按照自然界的实例产生产物的方式，所述方式也通过连接小的模块化单元来产生物质。在许多应用中，点击反应将生物分子和报告分子连接。点击化学不限于生物条件：“点击”反应的概念已用于药理学和各种仿生应用。然而，所述应用在生物分子的检测、定位和鉴定方面特别有用。典型的点击反应是经典的点击反应是形成5元杂原子环的铜催化的叠氮化物与炔烃的反应：Cu(I)催化的叠氮化物-炔烃环加成(CuAAC)。

[0043] 本公开中使用的术语“接触”、“添加”、“反应”、“处理”等意指使一种反应物、试剂、溶剂、催化剂、反应性基团等与另一种反应物、试剂、溶剂、催化剂、反应性基团等接触。可以单独、同时或分开添加并且可以以实现期望结果的任何顺序添加反应物、试剂、溶剂、催化剂、反应性基团等。可以在存在或不存在加热或冷却设备的情况下添加并且可以任选地在惰性气氛下添加所述反应物、所述试剂、所述溶剂、所述催化剂、所述反应性基团等。

[0044] 本公开中使用的术语“互补”是指两个多核苷酸链的区之间或两个核苷酸之间通过碱基配对的序列互补性的广义概念。已知腺嘌呤核苷酸能够与胸腺嘧啶或尿嘧啶核苷酸形成特异性氢键(“碱基配对”)。类似地，已知胞嘧啶核苷酸能够与鸟嘌呤核苷酸碱基配对。

[0045] 本公开中使用的术语“文库”在关于核酸使用时旨在意指具有不同化学组成(例如，不同序列、不同长度等)的核酸的集合。通常，文库中的核酸将是具有某个属或类的共同特征或特性的不同物种，但在其它方面在某种程度上有所不同。例如，文库可以包含核苷酸

序列不同、但在具有糖-磷酸酯主链方面类似的核酸物种。可以使用本领域已知的技术创建文库。本文所例示的核酸可以包含从任何来源,包含例如基因组(例如,人基因组)或基因组混合物的消化获得的核酸。在另一个实例中,核酸可以是从特定环境或生态系统的宏基因组研究中获得的那些核酸。所述术语还包含人工创建的核酸文库,如DNA文库。

[0046] 本公开中使用的术语“随机引物”或“随机六聚体引物”或“Random hexamer”或“Random hexamer primer”在本领域是众所周知的并且通常是指随机序列的短寡脱氧核糖核苷酸(d(N)6),退火至靶标DNA或RNA上的随机互补位点,用作DNA聚合酶或逆转录酶进行DNA合成的引物。

[0047] 本公开中使用的术语“AzNTP(3'-叠氮基-2',3'dNTP)”为叠氮化物脱氧核苷酸,其中,碱基选自腺嘌呤、鸟嘌呤、胞嘧啶和胸腺嘧啶。

[0048] 本公开中使用的术语“Add on PCR”是指参与PCR的引物上除了有与模板互补的序列外,还有一些其他序列,这些其他序列不参与本轮PCR反应,但生成的PCR产物会因为额外附带上这些序列,而可作为模板为下一步PCR反应提供模板。

[0049] II.具体实施方式详述

[0050] 在一些实施方案中,本公开提供了一种用于高通量测序的测序文库的建立方法,其包括以下步骤:

[0051] (1)采用逆转录引物对样本RNA进行逆转录,加入普通dNTP和3'修饰的dNTP,反应得到cDNA第一链,所述3'修饰的dNTP选自以下的一种或多种:AzNTP、AmNTP、propargyl-NTP、HalNTP;

[0052] (2)通过点击化学反应将5'端带有炔基修饰的寡核苷酸片段与步骤(1)获得的cDNA片段连接,所述寡核苷酸片段包含随机序列和通用测序引物1(seq1)的互补区段序列;

[0053] (3)以步骤(2)的反应产物为模板,进行PCR反应扩增;

[0054] (4)获得测序文库。

[0055] 在一些实施方案中,所述点击化学反应是指CuAAC点击反应,即一价铜离子催化叠氮化物-炔烃环加成反应。

[0056] 在一些实施方案中,所述步骤(1)逆转录的PCR反应包括五个反应阶段:具体反应条件为25℃反应10min,37℃反应10min,50℃反应45min,85℃反应2min,12℃保持。

[0057] 在一些实施方案中,所述步骤(1)中逆转录PCR反应中所使用的酶包括HiScript III Reverse Transcriptase(R302-01,南京诺唯赞生物科技股份有限公司)、SuperScript™ III逆转录酶(18080093,ThermoFisher SCIENTIFIC)、HiFi II M-MLV(H-) Reverse Transcriptase(CW0743,康为世纪生物科技股份有限公司)、EasyScript®Reverse Transcriptase[M-MLV,RNaseH-](AE101-02,北京全式金生物技术股份有限公司)、MutiScript II Reverse Transcriptase(MD311,菲鹏生物股份有限公司)。

[0058] 在一些实施方案中,所述步骤(1)中所述的逆转录引物是随机引物或基于靶向基因的保留内含子下游外显子设计的基因特异性引物群1,所述基因特异性引物群1中的引物5'端带有通用测序引物2序列(seq2)。

[0059] 在一些实施方案中,所述步骤(2)点击化学的反应体系还包括维生素C、铜(II)-TBTA组合物、DMSO。

[0060] 在一些实施方案中,所述步骤(3)中PCR扩增反应包括四个反应阶段:第一PCR扩增

反应为1个循环,具体反应条件为94℃反应1min,60℃反应30s,68℃反应10min;第二PCR扩增反应包括12个循环,具体反应条件为94℃反应30s,60℃反应30s,68℃反应2min;第三PCR扩增反应包括1个循环,具体反应条件为68℃反应5min;第四PCR扩增反应包括1个循环,具体反应条件为12℃。

[0061] 在一些实施方案中,所述步骤(3)中PCR反应体系还包括$MgCl_2$、DMSO、Tris-HCl、EDTA、NaCl、KCl等PCR Buffer溶液。

[0062] 在一些实施方案中,所述步骤(3)中PCR反应所用的酶为Taq DNA聚合酶。

[0063] 在一些实施方案中,所述步骤(3)采用的PCR扩增引物包括通用测序引物1和基因特异性引物群2,其基因特异性引物群2中的引物的5'端带有通用测序引物2序列。

[0064] 在一些实施方案中,所述基因特异性引物群1和2均是基于特定靶向基因的保留内含子选择性剪接事件下游的外显子设计,其中,基因特异性引物2的靶向位点比基因特异性引物1的位点向上游上移5-100个碱基。

[0065] 在一些实施方案中,所述基因特异性引物2的靶向位点比基因特异性引物1的位点向上游上移5、10、20、30、40、50、60、70、80、90或100个碱基。

[0066] 在一些实施方案中,所述基因特异性引物群2的靶向位点比基因特异性引物群1的位置向上游上移20-50个碱基。

[0067] 在一些实施方案中,所述基因特异性引物群1和基因特异性引物群2分别靶向的目标基因的个数大于等于1。

[0068] 在一些实施方案中,所述基因特异性引物群1和2均是基于特定靶向基因保留内含子下游的外显子设计。

[0069] 在一些实施方案中,所述基因特异性引物群1和2的靶向位置可以部分重叠,但不能完全一致。

[0070] 在一些实施方案中,步骤(1)中加入普通dNTP和3'修饰的dNTP的摩尔浓度比例为1:1-1:100,其摩尔浓度比例可以为1:100、1:90、1:80、1:75、1:70、1:65、1:60、1:55、1:50、1:45、1:40、1:35、1:30、1:25、1:20、1:19、1:18、1:17、1:16、1:15、1:14、1:13、1:12、1:11或1:10;在一个优选的实施方案中,步骤(1)中加入普通dNTP和3'修饰的dNTP的摩尔浓度比例为1:50。

[0071] 在一个优选的实施方案中,所述步骤(1)中加入普通dNTP和3'修饰的dNTP的摩尔浓度比例为1:20。

[0072] 在一些实施方案中,步骤(2)中所述随机序列包含4-16个核苷酸。

[0073] 在一些实施方案中,步骤(2)中所述随机序列包含4、5、6、7、8、9、10、11、12、13、14、15或16个核苷酸。

[0074] 在一些实施方案中,步骤(2)中所述随机序列包含5个核苷酸。

[0075] 在一些实施方案中,步骤(3)的PCR扩增包括采用通用测序引物1和基因特异性引物群2对步骤(2)的反应产物进行PCR扩增后,再对扩增产物添加接头结构的步骤,所述接头结构包括P5/P7接头和核酸条形码,所述核酸条形码连接在引物P5和/或P7接头端。

[0076] 在一些实施方案中,所述步骤(3)为:以步骤(2)的反应产物为模板,添加P5和P7接头进行PCR反应。

[0077] 在一些实施方案中,所述P5接头引物如SEQ ID NO:54所示。

[0078] 在一些实施方案中,所述P7接头引物如SEQ ID NO:55-57所示。

[0079] 在一些实施方案中,所述核酸条形码连接在引物P7接头单端。

[0080] 在一些实施方案中,所述核酸条形码连接在引物P5和P7接头双端。

[0081] 在一些实施方式中,所述核酸条形码分为核酸条形码5和核酸条形码7。

[0082] 在一些实施方式中,所述核酸条形码5的核苷酸序列选自SEQ ID NO.10-35任一项所示的序列。

[0083] 在一些实施方式中,所述核酸条形码7的核苷酸序列选自SEQ ID NO.36-53任一项所示的序列。

[0084] 在一些实施方案中,所述步骤(3)中PCR扩增反应包括四个反应阶段:第一PCR扩增反应为1个循环,具体反应条件为94℃反应30s;第二PCR扩增反应包括18个循环,具体反应条件为94℃反应30s,68℃反应30s,72℃反应30s;第三PCR扩增反应包括1个循环,具体反应条件为72℃反应5min;第四PCR扩增反应包括1个循环,具体反应条件为12℃。

[0085] 在一些实施方案中,所述步骤(3)中PCR反应体系还包括$MgCl_2$、DMSO、Tris-HCl、EDTA、NaCl、KCl等PCR Buffer溶液。

[0086] 在一些实施方案中,所述步骤(3)中PCR反应所用的酶为Taq DNA聚合酶。

[0087] 在一些实施方案中,所述通用测序引物1序列选自SEQ ID NO:3、SEQ ID NO:5或SEQ ID NO:6中的任一项。

[0088] 在一些实施方案中,所述通用测序引物1的互补区段序列如SEQ ID NO:4所示。

[0089] 在一些实施方案中,所述通用测序引物2序列(seq2)选自SEQ ID NO:7或SEQ ID NO:8中的任一项。

[0090] 在一些实施方案中,本公开提供了一种根据上述方法构建的高通量测序文库。

[0091] 在一些实施方案中,本公开提供了一种用于检测剪接异构体的靶向高通量测序方法,其包括以下步骤:

[0092] (1)提取样本RNA,根据上述方法构建测序文库;

[0093] (2)基于上述测序文库,高通量测序获得所述细胞样本中目标基因的测序信息。

[0094] 在一些实施方案中,所述样本为组织、细胞和/或体液样本。

[0095] 在一些实施方案中,所述体液样本包括血液、唾液、尿液、母乳、脑脊液、羊水、腹水、胆汁以及胸腔积液中的一种或多种。

[0096] 在一些实施方案中,所述样本为细胞样本。

[0097] 在一些实施方案中,所述细胞样本包括但不限于MCF10A、MCF7、HeLa、HEK293T和/或MDA-MB-231。

[0098] 在一些实施方案中,所述目标基因包括但不限于ATP13A1、CXXC1、ECHDC2、FGFRL1、HMGN3、KLHL17、NAXD、LZTR1、SELENBP1、JMJD8、PSMB1、HIGD2A、HNRNPAB、SMARCC1、ATP5IF1、HIGD2B、RPS21、UQCC5、NFATC3、PCNP和/或OSGEP。

[0099] 在一些实施方案中,本公开提供了一种利用前述的用于高通量测序的测序文库的建立方法、前述的高通量测序文库和/或前述的靶向高通量测序方法在脱靶事件评估中的用途。

[0100] 在一些实施方案中,所述脱靶事件评估包括以下:

[0101] (1)精准确定反式剪接因子中发生脱靶反式剪接所在的基因组位置;

[0102] (2)对目标反式剪接和脱靶反式剪接进行定量分析。

[0103] 在一些实施方案中,所述的高通量测序方法可用于检测转录组中任一种形式的可变剪接和任何转录本的组合。

[0104] 在一些实施方案中,所述用于高通量测序的测序文库的构建过程如下所述,其原理图见附图2:

[0105] 1)逆转录:以RNA为模板通过逆转录合成cDNA第一链,使用随机六聚体引物Random hexamer作为逆转录引物,并且加入普通dNTP和一定浓度的3'修饰dNTP进行逆转录反应,其中,所述3'修饰的dNTP为AzNTP(3'-叠氮基-2',3'dNTP),其能在链延伸时能够阻止添加下一个dNTP结合,同时所述3'修饰基团应可用于下游接头进行化学连接。Random Hexamer是包括6个碱基的随机序列单链DNA,随机序列的特性可帮助其随机结合至RNA的不同片段。逆转录过程中,在逆转录酶的作用下,以RNA为模板在引物3'末端添加dNTPs合成单链cDNA。当3'修饰dNTP替代dNTPs添加至cDNA单链上时,链延伸终止,完成cDNA第一链合成。

[0106] 2)点击化学连接:将上述获得的cDNA与5'端含有炔基修饰的寡核苷酸(oligonucleotides)进行点击化学反应(Click Reaction,即CuAAC点击反应——一价铜离子催化叠氮化物-炔烃环加成反应)。其中,寡核苷酸接头序列包含5'N5+NGS通用测序引物(seq1)互补区段序列,作为后续测序建库中PCR引物结合位点。N5为5个随机碱基组成序列,用于保证测序文库分子多样性以及在结果分析中去除PCR扩增过程中的系统性偏差和测序错误(sequencing error)。随机碱基序列的长度可调整为N4、N5、N6、N7、N8、N9、N10、N11、N12、N13、N14、N15、N16。其中N选自ATCG中的任一碱基。

[0107] 3)靶向富集:以点击化学反应产物为模板,通过PCR反应靶向富集目标基因片段。PCR体系中的引物包括一个通用测序引物1(seq1,其3'端序列与上述炔基引物中3'端互补)和基因特异性引物群。引物群基于每个靶向基因的保留内含子下游的外显子而设计,并在引物5'端添加通用测序引物2序列(seq2)。CLICK产物中含有叠氮化物-炔烃连接键,因此,DNA聚合酶需使用Taq或其他可扩增该种模板的聚合酶。

[0108] 4)测序接头连接及富集:以上述富集片段为模板,进行PCR反应为富集片段添加完整的P5和P7接头,用于后续NGS测序。并加入核酸条形码(sample barcode)。核酸条形码可实现高通量测序过程中的混样,提高检测通量、降低样本测序过程中的仪器误差。

[0109] 在一些实施方案中,所述高通量测序文库构建过程中逆转录引物为基于靶向基因的保留内含子下游外显子设计的基因特异性引物群1,所述特异性引物5'端添加通用测序引物序列(seq2);在一些实施方案中,在实际构建过程中可直接在点击化学连接产物纯化后进行测序接头连接及富集;在一些实施方案中,所述用于高通量测序的测序文库的构建过程如下所述,其原理图见附图3:

[0110] 1)逆转录:以RNA为模板通过逆转录合成cDNA第一链,使用基于靶向基因的保留内含子下游外显子设计的基因特异性引物群1,作为逆转录引物,并在5'端添加通用测序引物序列(seq2),加入普通dNTP和一定浓度的3'修饰dNTP进行逆转录反应,其中,所述3'修饰的dNTP为AzNTP(3'-叠氮基-2',3'dNTP),其能在链延伸时能够阻止添加下一个dNTP结合,同时所述3'修饰基团应可用于下游接头进行化学连接。逆转录过程中,在逆转录酶的作用下,以RNA为模板在引物3'末端添加dNTPs合成单链cDNA。当3'修饰dNTP替代dNTPs添加至cDNA单链上时,链延伸终止,完成cDNA第一链合成。

[0111] 2)点击化学连接:将上述获得的cDNA与5’端含有炔基修饰的寡核苷酸(oligonucleotides)进行点击化学反应(Click Reaction,即CuAAC点击反应——一价铜离子催化叠氮化物-炔烃环加成反应)。其中,寡核苷酸接头序列包含5’N5+NGS通用测序引物(seq1)互补区段序列,作为后续测序建库中PCR引物结合位点。N5为5个随机碱基组成序列,用于保证测序文库分子多样性以及在结果分析中去除PCR扩增过程中的系统性偏差和测序错误(sequencing error)。随机碱基序列的长度可调整为N4、N5、N6、N7、N8、N9、N10、N11、N12、N13、N14、N15、N16。

[0112] 3)测序接头连接及富集:以点击化学反应产物为模板,添加完整的P5和P7接头,用于后续PCR扩增。P5接头可与点击化学反应产物中的Seq 1互补序列进行互补,P7接头可与PCR反应过程中产生的含有Seq 2互补序列进行互补。并加入核酸条形码(sample barcode)。核酸条形码可实现高通量测序过程中的混样,提高检测通量、降低样本测序过程中的仪器误差。经过数轮PCR反应后实现P5和P7接头的连接以及待测序片段的富集,以便用于后续NGS测序。CLICK产物中含有叠氮化物-炔烃连接键,因此,DNA聚合酶需使用Taq或其他可扩增该种模板的聚合酶。

[0113] 在一些实施方案中,所述高通量测序文库构建过程中逆转录引物为基于靶向基因的保留内含子下游外显子设计的基因特异性引物群1;在一些实施方案中,在点击化学连接反应后进行特异性基因PCR,以去除模板中的核糖体RNA对测序数据的影响;在一些实施方案中,所述用于高通量测序的测序文库的构建过程如下所述,其原理图见附图4:

[0114] 1)逆转录:以RNA为模板通过逆转录合成cDNA第一链,使用基于靶向基因的保留内含子下游外显子设计的基因特异性引物群1,作为逆转录引物,并在5’端添加通用测序引物序列(seq2),加入普通dNTP和一定浓度的3’修饰dNTP进行逆转录反应,其中,所述3’修饰的dNTP为AzNTP(3’-叠氮基-2’,3’dNTP),其能在链延伸时能够阻止添加下一个dNTP结合,同时所述3’修饰基团应可用于下游接头进行化学连接。逆转录过程中,在逆转录酶的作用下,以RNA为模板在引物3’末端添加dNTPs合成单链cDNA。当3’修饰dNTP替代dNTPs添加至cDNA单链上时,链延伸终止,完成cDNA第一链合成。

[0115] 2)点击化学连接:将上述获得的cDNA与5’端含有炔基修饰的寡核苷酸(oligonucleotides)进行点击化学反应(Click Reaction,即CuAAC点击反应——一价铜离子催化叠氮化物-炔烃环加成反应)。其中,寡核苷酸接头序列包含5’N5+NGS通用测序引物(seq1)互补区段序列,作为后续测序建库中PCR引物结合位点。N5为5个随机碱基组成序列,用于保证测序文库分子多样性以及在结果分析中去除PCR扩增过程中的系统性偏差和测序错误(sequencing error)。随机碱基序列的长度可调整为N4、N5、N6、N7、N8、N9、N10、N11、N12、N13、N14、N15、N16。

[0116] 3)靶向富集:以点击化学反应产物为模板,通过PCR反应靶向富集目标基因片段。PCR体系中的引物包括一个通用测序引物1(seq1,其3’端序列与上述炔基引物中3’端互补)和基因特异性引物群2。引物群2基于每个靶向基因的保留内含子下游的外显子而设计,但应比靶向逆转录引物群1位置向上游上移5-100个碱基,甚至可以部分重叠,但不能完全一致,并在引物5’端添加通用测序引物2序列(seq2)。CLICK产物中含有叠氮化物-炔烃连接键,因此,DNA聚合酶需使用Taq或其他可扩增该种模板的聚合酶。

[0117] 4)测序接头连接及富集:以上述富集片段为模板,进行PCR反应为富集片段添加完

整的P5和P7接头,用于后续NGS测序。并加入核酸条形码(sample barcode)。核酸条形码可实现高通量测序过程中的混样,提高检测通量、降低样本测序过程中的仪器误差。

[0118] 在一些实施方案中,所述构建后的高通量测序文库中的片段均包括以下元件:P5序列、样本标签5、通用测序引物1(seq1)、插入DNA片段、通用测序引物2(seq2)、样本标签7、P7接头。

[0119] 在一些实施方案中,所述P5序列和P7的核苷酸序列分别如SEQ ID NO:1(AATGATACGGCGACCACCGAGATCTACAC)和SEQ ID NO:2(CAAGCAGAAGACGGCATACGAGA T)所示。

[0120] 在一些实施方案中,所述通用测序引物包括两种类型,PE接头和Nextera接头。

[0121] 在一些实施方案中,所述通用测序引物1和2选自以下表1中序列的任意组合。

[0122] 表1通用测序引物的序列信息

[0123]

名称	核酸序列	所包含核酸序列编号
Seq1 (PE 接头)	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	SEQ ID NO:3
Seq1 互补序列	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	SEQ ID NO:4
Seq1 (Nextera 接头)	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	SEQ ID NO:5
PE1_p26 (Seq1_p26)	ACACTCTTTCCCTACACGACGCTCTT	SEQ ID NO:6
Seq2 (PE 接头)	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	SEQ ID NO:7
Seq2 (Nextera 接头)	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	SEQ ID NO:8

[0124] 在一些实施方案中,所述寡核苷酸Hex_N5_Seq1rc的核苷酸序列如SEQ ID NO:9所示(NNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT)。

[0125] 在一些实施方案中,所述样本标签5和样本标签7选自下表2中序列的任意组合。

[0126] 表2序列信息

[0127]

标签 7 名称	序列	所包含核酸序列编号	标签 5 名称	序列	所包含核酸序列编号
701	TCGCCTTA	SEQ ID NO:10	501	TAGATCGC	SEQ ID NO:36
702	CTAGTACG	SEQ ID NO:11	502	CTCTCTAT	SEQ ID NO:37
703	TTCTGCCT	SEQ ID NO:12	503	TATCCTCT	SEQ ID NO:38
704	GCTCAGGA	SEQ ID NO:13	504	AGAGTAGA	SEQ ID NO:39
705	AGGAGTCC	SEQ ID NO:14	505	GTAAGGAG	SEQ ID NO:40
706	CATGCCTA	SEQ ID NO:15	506	ACTGCATA	SEQ ID NO:41
707	GTAGAGAG	SEQ ID NO:16	507	AAGGAGTA	SEQ ID NO:42
708	CCTCTCTG	SEQ ID NO:17	508	CTAAGCCT	SEQ ID NO:43

[0128]

标签7名称	序列	所包含核酸序列编号	标签5名称	序列	所包含核酸序列编号
709	AGCGTAGC	SEQ ID NO:18	510	CGTCTAAT	SEQ ID NO:44
710	CAGCCTCG	SEQ ID NO:19	511	TCTCTCCG	SEQ ID NO:45
711	TGCCTCTT	SEQ ID NO:20	513	TCGACTAG	SEQ ID NO:46
712	TCCTCTAC	SEQ ID NO:21	515	TTCTAGCT	SEQ ID NO:47
714	TCATGAGC	SEQ ID NO:22	516	CCTAGAGT	SEQ ID NO:48
715	CCTGAGAT	SEQ ID NO:23	517	GCGTAAGA	SEQ ID NO:49
716	TAGCGAGT	SEQ ID NO:24	518	CTATTAAG	SEQ ID NO:50
718	GTAGCTCC	SEQ ID NO:25	520	AAGGCTAT	SEQ ID NO:51
719	TACTACGC	SEQ ID NO:26	521	GAGCCTTA	SEQ ID NO:52
720	AGGCTCCG	SEQ ID NO:27	522	TTATGCGA	SEQ ID NO:53
721	GCAGCGTA	SEQ ID NO:28			
722	CTGCGCAT	SEQ ID NO:29			
723	GAGCGCTA	SEQ ID NO:30			
724	CGCTCAGT	SEQ ID NO:31			
726	GTCTTAGG	SEQ ID NO:32			
727	ACTGATCG	SEQ ID NO:33			
728	TAGCTGCA	SEQ ID NO:34			
729	GACGTCGA	SEQ ID NO:35			

[0129] 在一些实施方案中,所述完整序列的P5和P7接头的序列信息如表3所示。

[0130] 表3序列信息

[0131]

名称	核酸序列	所包含核酸序列编号
P5_PE1_N501（P5_Seq1_N501）	AATGATACGGCGACCACCGAGATCTACACTAGATCGCACACTCTTTCCCTACACGACGCTCTTCCGATCT	SEQ ID NO:54
P7_PE2_N701（P7_Seq2_N701）	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	SEQ ID NO:55
P7_PE2_N702（P7_Seq2_N702）	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	SEQ ID NO:56
P7_PE2_N703（P7_Seq2_N703）	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	SEQ ID NO:57

[0132] 实施例

[0133] 下面通过具体实施的方式来进一步说明本公开的技术方案。本领域技术人员应该明了,所述实施例仅仅是帮助理解本公开,不应视为对本公开的具体限制。

[0134] 实施例1:检测剪接异构体的靶向高通量测序平台的构建

[0135] 1、RNA提取

[0136] 用商业化试剂盒对目标细胞的RNA进行提取,例如济凡生物科技(北京)有限公司的FineProtect通用型RNA提取试剂盒。提取过程中需保证RNA的提取质量,以进行后续试

验。

[0137] 2、RNA定量

[0138] 采用商业化试剂盒对RNA进行定量,并利用荧光染料技术特异性的检测RNA浓度。例如Invitrogen公司的Qubit™ RNA高灵敏度(HS)定量试剂盒。

[0139] 3、逆转录

[0140] 根据上述测定的定量浓度,选择相同质量的RNA作为合成cDNA第一链的模板,以随机六聚体引物random hexamer作为逆转录引物,在逆转录体系中加入普通dNTP和1/20摩尔浓度比例的3’修饰的dNTP进行逆转录反应。

[0141] 其中逆转录反应体系组分见表4,逆转录反应条件见表5。

[0142] 表4:转录反应体系(20μL/反应)

[0143]

反应体系组分	主要成分	加入量(μL)
酶系	逆转录酶 HiScript III Reverse Transcriptase R302-01(南京诺唯赞生物科技股份有限公司)	1
3’修饰 dNTP 混合液	普通 dNTP(10mM)、AzNTP(3'-叠氮基-2',3'dNTP) (0.5mM)	1
随机引物	6 个碱基的随机序列单链 DNA (100μM)	1
第一链合成缓冲液	$MgCl_2$、Tris-Hcl、KCl	4
模板	提取得到的 RNA 溶液	5
纯化水		补至 20μL

[0144] 表5:逆转录反应条件

[0145]

温度	时间
25℃	10min
37℃	10min
50℃	45min
85℃	2min
12℃	Hold

[0146] 4、cDNA纯化

[0147] 完成First-strand cDNA合成后的产物采用商业化试剂盒进行DNA纯化。例如Zymo公司的DNA Clean&Concentrator-5试剂盒。纯化产物使用纯水洗脱至10μL。

[0148] 5、Click ligation(点击化学连接)

[0149] 以上述纯化后的cDNA与5’端含有炔基修饰的寡核苷酸(oligonucleotides)在催化剂Copper(II)-TBTA complex存在的条件下,室温进行Click Reaction(点击化学反应,即CuAAC点击反应——一价铜离子催化叠氮化物-炔烃环加成反应)1小时。

[0150] 其中点击化学反应体系组分见表6,寡核苷酸(oligonucleotides)的核酸序列见

SEQ ID NO:9所示。

[0151] 表6:点击化学反应体系(37.8μL/反应)

[0152]

反应体系组分	主要成分/浓度	加入量（μL）
5’端含有炔基修饰的寡核苷酸	包含5’N5+NGS通用测序引物（seq1）互补区段序列(5μM)	3
vitamin C（维生素C）(50 mM)		0.8
Copper(II)-TBTA complex（铜（II）-TBTA组合物）		4
DMSO 100%		20
模板	以上步纯化后的cDNA	10

[0153] 6、Click ligation产物纯化

[0154] 点击化学反应产物采用商业化试剂盒进行纯化。例如Zymo公司的DNA Clean& Concentrator-5试剂盒。纯化产物使用纯水洗脱至10μL。

[0155] 7、靶向富集

[0156] 以上述纯化后的产物为模板,以靶向基因各保留内含子下游外显子设计的引物作为靶向基因特异性引物组合与一个通用测序引物(PE1_p26)进行PCR反应。

[0157] 其中PCR反应体系组分见表7,引物组合混合液的配制见表8,PE1_p26引物的核酸序列如SEQ ID NO:6所示,qPCR反应条件见表9。

[0158] 表7:PCR反应体系(25μL/反应)

[0159]

反应体系组分	主要成分	加入量（μL）
酶系	Taq DNA聚合酶	1
PCR缓冲液	MgCl2、DMSO、Tris-Hcl、EDTA、NaCl、KCl	5
dNTPs	dATP/dTTP/dCTP/dGTP (10mM)	0.5
PE1_p26 (10 μM)	包含seq1序列，其3’端序列与上述炔基引物中3’端互补	0.5
引物组合混合液 (10 μM)	5’端添加通用测序引物（seq2）序列	0.5
模板	转化得到的DNA溶液	5
纯化水		补至25μL

[0160] 表8:引物组合混合液组分表

[0161]

组分	加入量(μL)
各靶向基因引物(各100μM)	各10
纯化水	补至100

[0162] 表9:PCR反应条件

[0163]

循环数	温度	时间
1	94°C	1 min
	60°C	30 s
	68°C	10 min
12	94°C	30 s
	60°C	30 s
	68°C	2 min
1	68°C	5 min
1	12°C	保持

[0164] 8、靶向富集产物纯化

[0165] 靶向富集PCR产物采用商业化试剂盒进行纯化。例如Zymo公司的DNA Clean& Concentrator-5试剂盒。纯化产物使用纯水洗脱至10μL。

[0166] 9、测序接头连接及富集

[0167] 以上述获得的靶向富集片段为模板,以完整的P5和P7接头为引物,进行PCR反应,以便用于后续的NGS测序。其中,P5接头可与click反应产物中含有Seq 1互补区段序列的序列进行互补,P7接头可与PCR反应过程中产生的含有Seq 2互补序列进行互补。在单一Add-on PCR引物或两个Add-on PCR引物中加入核酸条形码(sample barcode),实现高通量测序过程中的混样。

[0168] 其中PCR反应体系组分见表10,完整序列的P5、P7接头引物的核酸序列见表3,qPCR反应条件见表11。

[0169] 表10:PCR反应体系(25μL/反应)

[0170]

反应体系组分	主要成分	加入量(μL)
酶系	Taq DNA 聚合酶	1
PCR 缓冲液	$MgCl_2$、DMSO、Tris-HCl、EDTA、NaCl、KCl	5
dNTPs	dATP/dTTP/dCTP/dGTP (10mM)	0.5
P5 接头 (10 μM)		0.5
P7 接头 (10 μM)		0.5
模板	转化得到的 DNA 溶液	5
纯化水		补至 25μL

[0171] 表11:PCR反应条件

[0172]

循环数	温度	时间
1	94°C	30 s
18	94°C	30 s
	68°C	30 s
	72°C	30 s
1	72°C	5 min
1	12°C	Hold

[0173] 10.高通量测序及结果分析

[0174] 基于上述步骤扩增得到的测序文库,高通量测序获得所述细胞样本中目标基因的测序信息,包括但不限于多重靶向剪接异构体的分析、剪接异构体的精准定量、未知的剪接异构体和反式剪接脱靶事件评估等。

[0175] 实施例2:MCF10A、MCF7、MDA-MB-231细胞系中剪接异构体检测

[0176] 利用实施例1所述的高通量测序平台检测MCF10A、MCF7、MDA-MB-231细胞系中的剪接异构体。

[0177] 其中,步骤1-6同实施例1中的操作。

[0178] 步骤7的靶向富集过程中,以ATP13A1基因、CXXC1基因、ECHDC2基因、FGFRL1基因、HMGN3基因、KLHL17基因、OSGEP基因各保留内含子下游外显子设计的引物作为靶向富集引物组合和通用测序引物PE1_p26进行PCR反应。引物组合混合液的组分表见表12,各靶向基因的引物序列见表13。

[0179] 表12:引物组合混合液组分表

[0180]

组分	加入量(μL)
ATPE7R_PE2(100μM)	10
KLHE11R_PE2(100μM)	10
CXXE12R_PE2(100μM)	10
ECHE6R_PE2(100μM)	10
FGFE8R_PE2(100μM)	10
HMGE6R_PE2(100μM)	10
OSGE5R_PE2(100μM)	10
纯化水	补至100

[0181] 表13:引物的核酸序列

[0182]

名称	核酸序列	所包含核酸序列编号
ATPE7R_PE2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGAAGCACGTCACATGGCACC	SEQ ID NO: 58
KLHE11R_PE2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGGCTCAGCATGGACGCCA	SEQ ID NO: 59
CXXE12R_PE2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCA	SEQ ID

[0183]

	AAGGACGTCTGGCTCTCA	NO: 60
ECHE6R_PE2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCCTCGCGTGGTCTCAATCA	SEQ ID NO: 61
FGFE8R_PE2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGACGAGGAGGCCACAGGTG	SEQ ID NO: 62
HMGE6R_PE2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTCCCTCGTTATCTACAGATTCAGT	SEQ ID NO: 63
OSGE5R_PE2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGCTTTGCCATCTGTTCAATGT	SEQ ID NO: 64

[0184] 步骤8-10同实施例1中的操作。

[0185] 高通量测序结果显示MCF10A、MCF7、MDA-MB-231细胞系目标基因均可见正常剪接体与剪接异构体(内含子保留),并成功定量检测靶向基因中剪接异构体的比例,具体试验结果见表14。以上数据说明,本公开构建的靶向高通量测序平台可有效实现对剪接异构体

的检测。

[0186] 表14:MCF10A、MCF7、MDA-MB-231细胞系目标基因的检测结果

[0187]

样本	基因	正常剪接读数	内含子保留读数	与正常剪接的比例(%)
MCF10A	ATP13A1	67,204	4,402	6.55%
	CXXC1	69,983	0	0.00%
	ECHDC2	7,052	225	3.19%
	FGFRL1	60,805	7,556	12.43%
	KLHL17	7,349	980	13.34%
	OSGEP	2,945	0	0.00%
MCF7	ATP13A1	92,105	4,071	4.42%
	CXXC1	101,210	788	0.78%
	ECHDC2	6,080	1,609	26.46%
	FGFRL1	105,533	8,254	7.82%
	KLHL17	13,994	854	6.10%
	OSGEP	6,208	0	0.00%
MDA-MB-231	ATP13A1	107,975	3,435	3.18%
	CXXC1	102,050	3	0.00%
	ECHDC2	9,613	1,223	12.72%
	FGFRL1	99,454	10,070	10.13%
	KLHL17	6,115	1,500	24.53%
	OSGEP	5,168	0	0.00%

[0188] 实施例3:体外转录(In-vitro Transcription,IVT)RNA模拟不同比例的剪接异构体检测

[0189] 制备体外RNA合成样本(IVT产物),包括两种剪接异构体序列,即正常剪接体与剪接异构体(内含子保留)。将IVT产物模拟不同比例的剪接异构体混合,使用本公开构建的靶向高通量测序平台检测,并计算混合样本实际检出值和混合样本理论值相关性。具体实验细节如下:

[0190] 步骤1、体外转录获得目的RNA合成样本

[0191] 已合成好的正常剪接体与剪接异构体(内含子保留)DNA使用商业化试剂盒进行体外转录反应,如诺唯赞T7 High Yield RNA Transcription Kit(TR101)。

[0192] 体外转录反应体系见表15。

[0193] 表15:体外转录体系(20μL体系)

[0194]

反应体系组分	主要成分	加入量(μL)
T7 RNA Polymerase Mix	T7 RNA Polymerase	2
10×反应缓冲液	Tris-HCl、MgCl2、NaCl、spermidine、DTT	2
ATP solution	100 mM ATP	2
GTP solution	100 mM GTP	2
UTP solution	100 mM UTP	2
CTP solution	100 mM CTP	2
模板	已合成好正常剪接体与剪接异构体的DNA片段，浓度50ng/μL	8

[0195] 上述体系37℃孵育2小时后,加入1μL DNase I,继续37℃孵育15min,以去除DNA模板。

[0196] 步骤2、RNA回收纯化

[0197] 体外转录合成的RNA,使用苯酚-乙醇沉淀法进行回收。将最终沉淀RNA产物使用双蒸水进行溶解,使用Nano-Drop测定RNA浓度,并将RNA稀释至2μM,0.2μM,0.02μM,0.002μM后备用。

[0198] 步骤3、混合不同比例的剪接异构体IVT产物

[0199] 如表16以1:1的体积比例混合IVT产物。

[0200] 表16:IVT产物混样

[0201]

正常剪接体IVT产物:剪接异构体IVT产物	1000:1	100:1	10:1	1:1	1:10	1:100	1:1000
正常剪接体IVT产物浓度	2μM	2μM	2μM	2μM	0.2μM	0.02μM	0.002μM
剪接异构体IVT产物浓度	0.002μM	0.02μM	0.2μM	2μM	2μM	2μM	2μM

[0202] 步骤4、逆转录

[0203] 将上述混合好的IVT产物,分别稀释100倍后作为合成cDNA第一链的模板,以目的内含子下游外显子上反向引物作为逆转录引物,在逆转录体系中加入普通dNTP和1/15摩尔浓度比例的3’修饰的dNTP进行逆转录反应。

[0204] 其中逆转录反应体系组分见表17,逆转录反应条件见表18。

[0205] 表17:转录反应体系(20μL/反应)

[0206]

反应体系组分	主要成分	加入量(μL)
酶系	逆转录酶 HiScript III Reverse	1

	Transcriptase R302-01（南京诺唯赞生物科技股份有限公司）	
3’修饰 dNTP 混合液	普通 dNTP (10mM)、AzNTP（(3'-叠氮基-2',3'dNTP） (0.5mM)	1
IVT 目的基因特异性引物 1	目的内含子下游外显子上反向引物 (2μM)	1
第一链合成缓冲液	$MgCl_2$、Tris-Hcl、KCl	4
模板	提取得到的 RNA 溶液	1
纯化水		补至 20μL

[0207]

[0208] 表18:逆转录反应条件

[0209]

温度	时间
25℃	10min
37℃	10min
50℃	45min
85℃	2min
12℃	Hold

[0210] 其中,步骤5-6同实施例1中的4-6操作。

[0211] 步骤8、靶向富集

[0212] 以上述纯化后的产物为模板,以目的内含子下游外显子设计的引物作为靶向基因特异性引物组合与一个通用测序引物(PE1_p26)进行PCR反应。

[0213] 其中PCR反应体系组分见表19,PE1_p26引物的核酸序列如SEQ ID NO:6所示,qPCR反应条件见表20。引物序列见表21。

[0214] 表19:PCR反应体系(25μL/反应)

反应体系组分	主要成分	加入量（μL）
酶系	Taq DNA 聚合酶	1
PCR 缓冲液	MgCl2、DMSO、Tris-Hcl、EDTA、NaCl、KCl	5
dNTPs	dATP/dTTP/dCTP/dGTP (10mM)	0.5
PE1_p26 (10 μM)	包含 seq1 序列，其 3’端序列与上述炔基引物中 3’端互补	0.5
IVT 目的基因特异性引物 2 (10 μM)	目的内含子下游外显子设计的引物，5’端添加通用测序引物（seq2）序列	0.5
模板	转化得到的 DNA 溶液	5
纯化水		补至 25μL

[0215]

[0216] 表20:PCR反应条件

[0217]

循环数	温度	时间
1	94°C	1 min

[0218]

	60°C	30 s
	68°C	10 min
15	94°C	30 s
	60°C	30 s
	68°C	2 min
1	68°C	5 min
1	12°C	保持

[0219] 表21:引物的核酸序列

[0220]

名称	核酸序列	所包含核酸序列编号
IVT 目的基因特异性引物 1	CACAGCGTCATACAGTCTGC	SEQ ID NO：65
IVT 目的基因特异性引物 2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTTCCGGTAGCCATGGATGAG	SEQ ID NO：66

[0221] 其中,步骤9-10同实施例1中的8-9操作。

[0222] 高通量测序结果显示不同混合比例的剪接异构体IVT产物实际检出值和混合样本理论值相关性r>0.99,具体试验结果见表22和图5。

[0223] 表22:RNA模拟不同比例的剪接异构体的检测结果

[0224]

理论比例		正常剪接（reads 数）	内含子保留（reads 数）	实测比例	
正常剪接	内含子保留			正常剪接	内含子保留
1	1,000	12,243	3,992,598	1	326.11
1	100	43,959	3,810,962	1	86.69
1	10	352,587	3,665,838	1	10.40
1	1	1,981,202	2,688,795	1	1.36
10	1	3,130,548	692,311	4.52	1
100	1	3,763,067	50,387	74.68	1
1000	1	3,874,960	8,916	434.61	1

[0225] 实施例4:HeLa细胞系中剪接异构体检测

[0226] 利用实施例1所述的高通量测序平台检测HeLa细胞系中的剪接异构体。

[0227] 其中,步骤1-2同实施例1中的操作。

[0228] 步骤3的靶向富集逆转录过程中,以基因特异性引物作为逆转录引物,在逆转录体系中加入普通dNTP和1/15摩尔浓度比例的3’修饰的dNTP进行逆转录反应。

[0229] 其中基因特异性引物序列见表23,基因特异性引组合混合液配制见表24,逆转录反应体系组分见表25,逆转录反应条件见表26。

[0230] 表23:逆转录反应基因特异性引物序列

[0231]

名称	引物序列 5'→3'	所包含核酸序列编号
ATP-e7-R	CATGGCCTCGTCTACGAT	SEQ ID NO：67
CXXC1-e13-R	CCTGGAGCCGCTTACAGTAT	SEQ ID NO：68

[0232]

ECH-e6-R	TCGCGTGGTCTCAATCAG	SEQ ID NO：69
FGF-e7-R	GGTGCCCAGGATGAAGAC	SEQ ID NO：70
HMG-e6-R	AAAGTTGTCCCAAGAGATACAT	SEQ ID NO：71
KLH-e11-R	TGCGTCGGCTCAGCAT	SEQ ID NO：72
OSG-e6-R	TATGGCAGCTCAACTAGCTT	SEQ ID NO：73
NAXD-e7-R	CGTCATACAGTCTGCTGAACT	SEQ ID NO：74
LZTR1-e19-R	CAGGTTCTGCTTGCAGTAC	SEQ ID NO：75
SEL1-e4-R	GAGTGATGCAGCTCGTCCTTC	SEQ ID NO：76
SEL1-e3-R	CAGGGCAGGTAGACGATCTCT	SEQ ID NO：77
JMJD8-e3-R	TTGTCCGTGAGTCCCTGCAG	SEQ ID NO：78
JMJD8-e4-R	GTGGTAGGAGTAGGTGTTGGC	SEQ ID NO：79

[0233] 表24:基因特异性引物混合物配制

[0234]

组分	加入量
单个引物(200μM)	1μL
双蒸水	补至100μL

[0235] 表25:转录反应体系(20μL/反应)

[0236]

反应体系组分	主要成分	加入量（μL）
酶系	逆转录酶 HiScript III Reverse Transcriptase R302-01（南京诺唯赞生物科技股份有限公司）	1
3'修饰 dNTP 混合液	普通 dNTP (10mM)、AzNTP（(3'-叠氮基-2',3'dNTP）(0.5mM)	1
基因特异性引物组合混合液	目的内含子下游外显子设计的引物混合物，单个引物浓度为 2μM	1
第一链合成缓冲液	$MgCl_2$、Tris-Hcl、KCl	4
模板	提取得到的 RNA 溶液，200ng/μL	5
纯水		补至 20μL

[0237] 表26:逆转录反应条件

[0238]

温度	时间
25℃	10min
37℃	10min
50℃	45min
85℃	2min
12℃	Hold

[0239] 其中,步骤4-6同实施例1中的操作。

[0240] 步骤7的靶向富集过程中,以ATP13A1基因、CXXC1基因、ECHDC2基因、FGFRL1基因、HMGN3基因、KLHL17基因、OSGEP基因、NAXD基因、LZTR1基因、SELENBP1基因、JMJD8基因各保留内含子下游外显子设计的引物作为靶向富集引物组合和通用测序引物PE1_p26进行PCR反应。相较于逆转录中的基因特异性引物,靶向富集PCR的引物更加靠近内含子约20-50bp。其中各靶向基因的引物序列见表27,引物混合液配制见表28,PCR反应体系组分见表29,PE1_p26引物的核酸序列如SEQ ID NO:6所示,PCR反应条件见表30。

[0241] 表27:引物组合混合液配制

[0242]

名称	引物序列 5'→3'	所包含核酸序列编号
ATP-e7-R-PE2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCAGCAGAAGCACGTCACATG	SEQ ID NO: 80
CXXE-12-R-PE2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCAAAGGACGTCTGGCTCTCA	SEQ ID NO: 81
ECH-e6-R-PE2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTGGTCTCAATCAGTCCCATGACT	SEQ ID NO: 82
FGF-e7-R-PE2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGCCACGGCAGGCTAGT	SEQ ID NO: 83
HMG-e6-R-PE2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTCCCTCGTTATCTACAGATTCAGT	SEQ ID NO: 84
KLH-e11-R-PE2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCACGGGCGACCACACGTT	SEQ ID NO: 85
OSG-e5-R-PE2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGCTTTGCCATCTGTTCAATGT	SEQ ID NO: 86
NAXD-e7-R-PE2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTTCCGGTAGCCATGGATGAG	SEQ ID NO: 87
LZTR1-e19-R-PE2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGCAGCCGGTTGTTGTAGAAG	SEQ ID NO: 88
SEL1-e4-R-PE2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTTCAGGTTGGGCATGGG	SEQ ID NO: 89
SEL1-e3-R-PE2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTAGACGATCTCTTCCCTGGGT	SEQ ID NO: 90
JMJD8-e3-R-PE2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGTGAGTCCCTGCAGGATGAC	SEQ ID NO: 91
JMJD8-e4-R-PE2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCACTCTGTCCCCAAACGAAG	SEQ ID NO: 92

[0243] 表28:引物混合物配制

[0244]

组分	加入量
单个引物(200μM)	1μL
双蒸水	补至20μL

[0245] 表29:PCR反应体系(25μL/反应)

[0246]

反应体系组分	主要成分	加入量（μL）
酶系	Taq DNA 聚合酶	1
PCR 缓冲液	MgCl2、DMSO、Tris-Hcl、EDTA、NaCl、KCl	5
dNTPs	dATP/dTTP/dCTP/dGTP (10mM)	0.5
PE1_p26 (10 μM)	包含 seq1 序列，其 3’端序列与上述炔基引物中 3’端互补	0.5
引物组合混合液 (10 μM)	5’端添加通用测序引物（seq2）序列，单个引物浓度为 10μM	0.5
模板	转化得到的 DNA 溶液	5
纯化水		补至 25μL

[0247] 表30:PCR反应条件

[0248]

循环数	温度	时间
1	94°C	1 min
	60°C	30 s
	68°C	10 min
15	94°C	30 s
	60°C	30 s
	68°C	2 min
1	68°C	5 min
1	12°C	保持

[0249] 步骤8-10同实施例1中的操作。

[0250] 最终产物进行高通量测序,高通量测序结果显示HeLa细胞系目标基因均可见正常剪接体与剪接异构体(内含子保留),并成功定量检测靶向基因中剪接异构体的比例,具体试验结果见表31。以上数据说明,本公开构建的靶向高通量测序平台可有效实现对剪接异构体的检测。

[0251] 表31:HeLa细胞系目标基因的检测结果

[0252]

样本	基因	正常剪接读数	内含子保留读数	与正常剪接的比例(%)
HeLa	ATP13A1	177,706	18,900	10.64%
	CXXC1	983,777	36,144	3.67%
	ECHDC2	23,600	0	0.00%
	FGFRL1	69,696	7,301	10.48%
	HMGN3	63,319	1,925	3.04%
	KLHL17	74,230	4,912	6.62%
	OSGEP	50,740	2,096	4.13%
	NAXD	231,478	45,530	19.67%

[0253]

	LZTR1	33,546	303	0.90%
	SELENBP1	219,337	7,965	3.63%
	SELENBP1	525,666	48,319	9.19%
	JMJD8	130,949	31,480	24.04%
	JMJD8	110,469	42,081	38.09%

[0254] 实施例5:反式剪接(trans-splicing)脱靶事件评估

[0255] 利用实施例1所述的高通量测序平台检测HEK293T细胞转染迷你基因(mini-Gene)与反式剪接因子(Pre-mRNA Trans-splicing Molecule)后发生目标(on-target)反式剪接和脱靶(off-target)反式剪接事件。

[0256] 其中,步骤1-2同实施例1中的操作。

[0257] 步骤3的靶向富集逆转录过程中,以反式剪接分子、迷你基因特异性引物作为逆转录引物,在逆转录体系中加入普通dNTP和1/15摩尔浓度比例的3’修饰的dNTP进行逆转录反应。

[0258] 其中基因特异性引物序列见表32,基因特异性引组合混合液配制见表33,逆转录反应体系组分见表34,逆转录反应条件见表35。

[0259] 表32:逆转录反应基因特异性引物序列

[0260]

名称	引物序列5’→3’	所包含核酸序列编号
HTAS_SP_ts_R1	GTTGCCATCCTCCTTGAAAT	SEQ ID NO:93
HTAS_SP_cis_R1	CCACCTAGTGCTTTAGCAATAT	SEQ ID NO:94

[0261] 表33:基因特异性引物混合物配制

[0262]

组分	加入量
单个引物(200μM)	1μL
双蒸水	补至100μL

[0263] 表34:转录反应体系(20μL/反应)

[0264]

反应体系组分	主要成分	加入量(μL)
酶系	逆转录酶 HiScript III Reverse Transcriptase R302-01（南京诺唯赞生物科技股份有限公司）	1
3’修饰 dNTP 混合液	普通 dNTP (10mM)、AzNTP（(3'-叠氮基-2',3'dNTP）(0.5mM)	1
基因特异性引物组合混合液	目的内含子下游外显子设计的引物混合物，单个引物浓度为 2μM	1
第一链合成缓冲液	$MgCl_2$、Tris-Hcl、KCl	4
模板	提取得到的 RNA 溶液，200ng/μL	5
纯水		补至 20μL

[0265] 表35:逆转录反应条件

[0266]

温度	时间
25℃	10min

37℃	10min
50℃	45min
85℃	2min
12℃	Hold

[0267] 其中,步骤4-6同实施例1中的操作。

[0268] 步骤7的靶向富集过程中,以反式剪接分子、迷你基因特异性引物作为靶向富集引物组合和通用测序引物PE1_p26进行PCR反应。相较于逆转录中的特异性引物,反式剪接分子、迷你基因靶向富集PCR的引物分别更加靠近3’剪接位点(3’splicing site)和5’剪接位点(5’splicing site)约20-50bp。其中各靶向基因的引物序列见表36,引物混合液配制见表37,PCR反应体系组分见表38,PE1_p26引物的核酸序列如SEQ ID NO:6所示,PCR反应条件见表39。

[0269] 表36:引物组合混合液配制

[0270]

名称	引物序列 5’→3’	所包含核酸序列编号
HTAS_SP_tsR_PE2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCGATGCGATTCACCAGGGT	SEQ ID NO：95
HTAS_SP_cisR_PE2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGACAGGGCTGCCAAGAACAT	SEQ ID NO：96

[0271] 表37:引物混合物配制

[0272]

组分	加入量
单个引物(200μM)	1μL
双蒸水	补至20μL

[0273] 表38:PCR反应体系(25μL/反应)

[0274]

反应体系组分	主要成分	加入量(μL)
酶系	Taq DNA 聚合酶	1
PCR 缓冲液	MgCl2、DMSO、Tris-Hcl、EDTA、NaCl、KCl	5
dNTPs	dATP/dTTP/dCTP/dGTP (10mM)	0.5
PE1_p26 (10 μM)	包含 seq1 序列，其 3’端序列与上述炔基引物中 3’端互补	0.5
引物组合混合液 (10 μM)	5’端添加通用测序引物（seq2）序列，单个引物浓度为 10μM	0.5
模板	转化得到的 DNA 溶液	5
纯化水		补至 25μL

[0275] 表39:PCR反应条件

[0276]

循环数	温度	时间
1	94°C	1 min
	60°C	30 s
	68°C	10 min
15	94°C	30 s
	60°C	30 s
	68°C	2 min
1	68°C	5 min
1	12°C	保持

[0277] 步骤8-10同实施例1中的操作。

[0278] 最终产物进行高通量测序,高通量测序结果显示HEK293T细胞转染迷你基因(mini-Gene)与反式剪接因子(Pre-mRNA Trans-splicing Molecule)后可见目标反式剪接产物与脱靶反式剪接见产物。同时,精准确定反式剪接因子发生脱靶反式剪接所在基因组位置(事件较多,未全部列举),具体试验结果见表40。以上数据说明,本公开构建的靶向高通量测序平台可有效实现对反式剪接脱靶事件评估,并对目标反式剪接和脱靶反式剪接同时进行定性定量分析。

[0279] 表40:HEK293T细胞反式剪接和脱靶反式剪接的检测结果

[0280]

样本	HEK293T
目标反式剪接读数	20,044
目标反式剪接占比 (目标反式剪接产物读数与反式剪接分子靶向富集测序总读数比例%)	8.67%
脱靶反式剪接读数	139,844
脱靶反式剪接占比 (脱靶反式剪接产物读数与反式剪接分子靶向富集测序总读数比例%)	60.50%
脱靶反式剪接所在基因组位置	PSMB1 基因, 6 号染色体
	HIGD2A 基因, 5 号染色体
	HNRNPAB 基因, 5 号染色体
	SMARCC1 基因, 3 号染色体
	ATP5IF1 基因, 1 号染色体
	HIGD2B 基因, 15 号染色体
	RPS21 基因, 20 号染色体
	UQCC5 基因, 3 号染色体
	NFATC3 基因, 16 号染色体
	PCNP 基因, 3 号染色体

[0281] 以上所述实施例仅表达了本公开的几种实施方式,其描述较为具体和详细,但并不能因此而理解为对发明专利范围的限制。应当指出的是,对于本领域的普通技术人员来说,在不脱离本公开构思的前提下,还可以做出若干变形和改进,这些都属于本公开的保护范围。因此,本公开专利的保护范围应以所附权利要求为准。

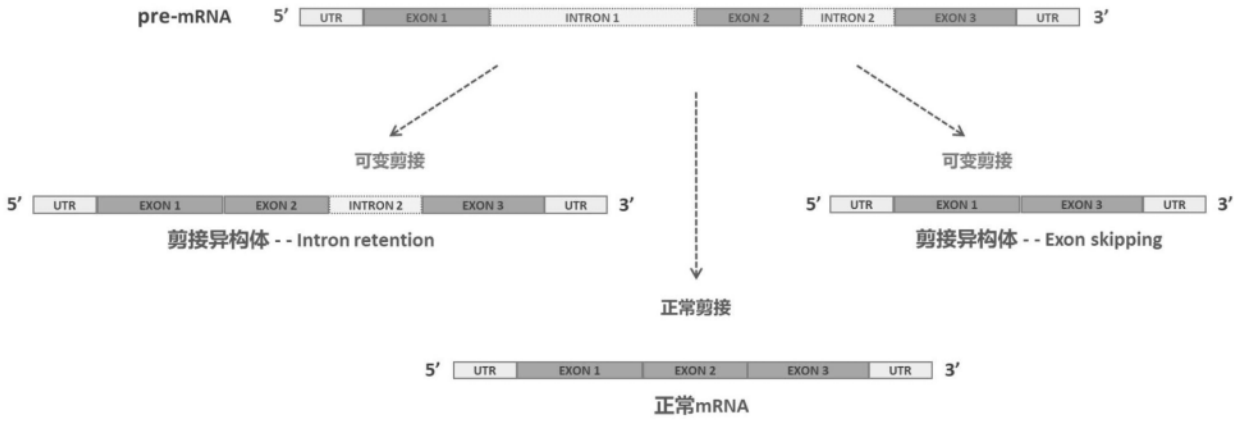

pre-mRNA
5′
UTR
EXON 1
INTRON 1
EXON 2
INTRON 2
EXON 3
UTR
3′
可变剪接
可变剪接
剪接异构体 - - Intron retention
剪接异构体 - - Exon skipping
正常剪接
正常mRNA

图1

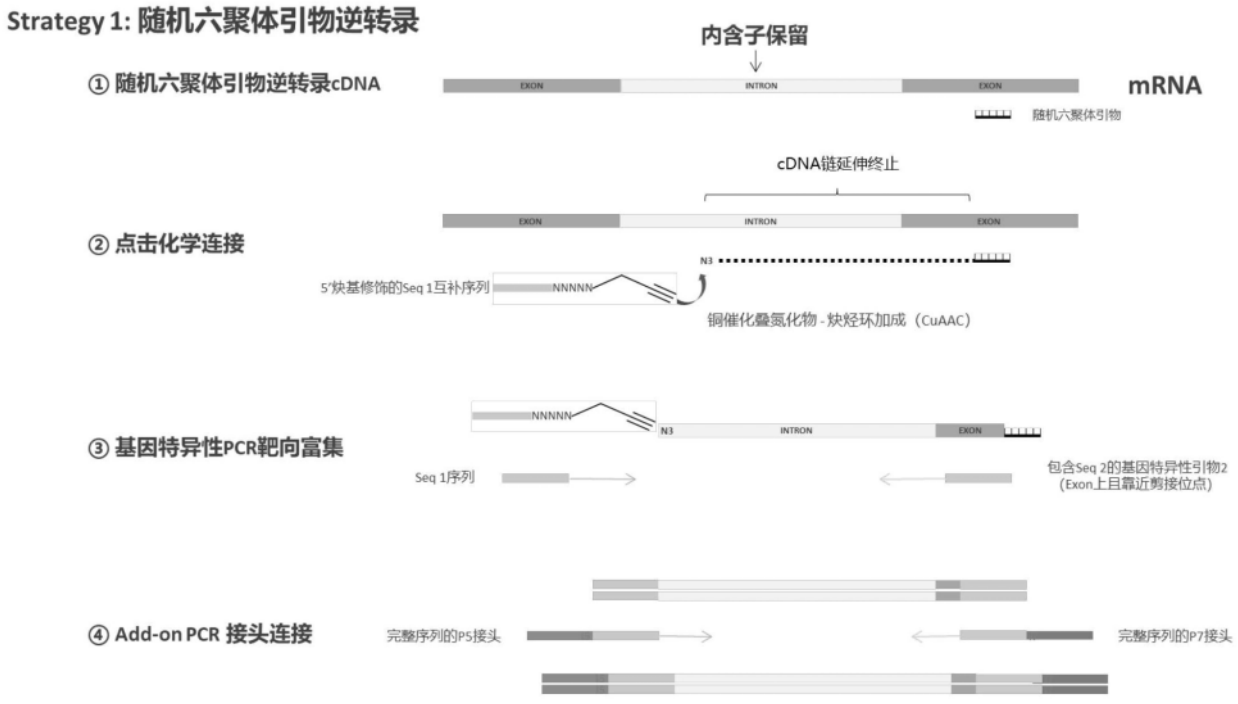

Strategy 1: 随机六聚体引物逆转录
内含子保留
① 随机六聚体引物逆转录cDNA
EXON
INTRON
mRNA
随机六聚体引物
cDNA链延伸终止
② 点击化学连接
N3
5′炔基修饰的Seq 1互补序列
NNNNN
铜催化叠氮化物-炔烃环加成（CuAAC）
③ 基因特异性PCR靶向富集
Seq 1序列
包含Seq 2的基因特异性引物2
(Exon上且靠近剪接位点)
④ Add-on PCR 接头连接
完整序列的P5接头
完整序列的P7接头

图2

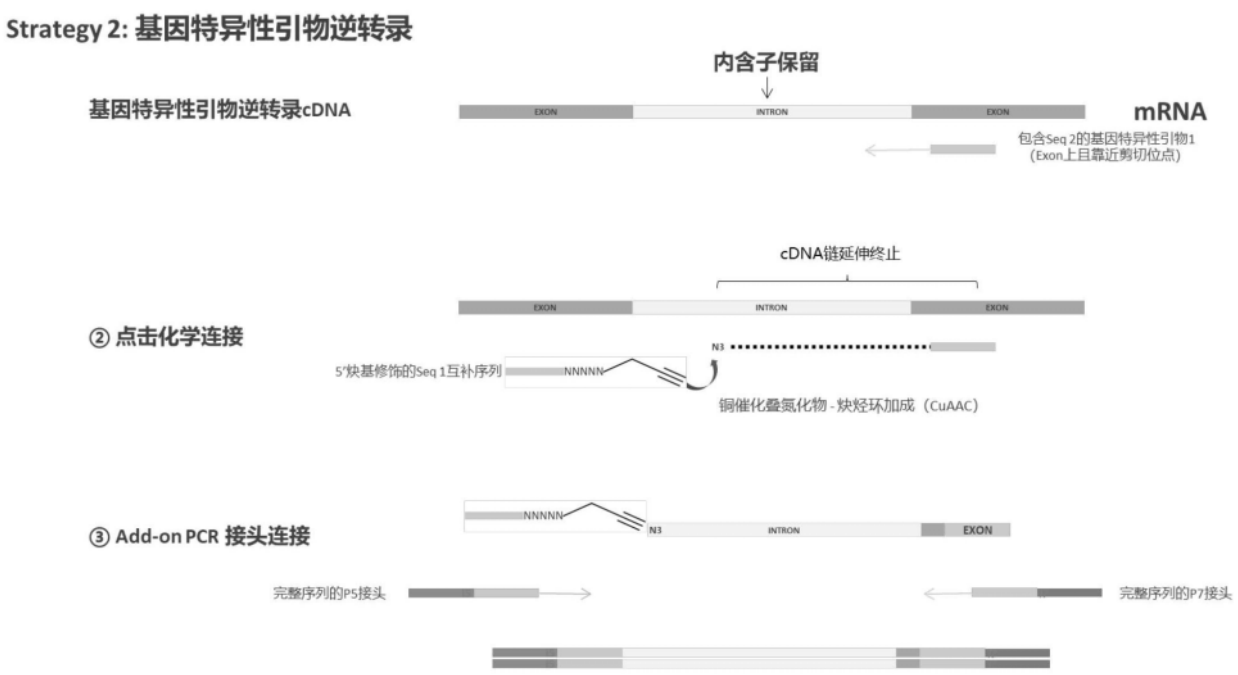

Strategy 2: 基因特异性引物逆转录
内含子保留
基因特异性引物逆转录cDNA
EXON
INTRON
EXON
mRNA
包含Seq 2的基因特异性引物1
(Exon上且靠近剪切位点)
cDNA链延伸终止
② 点击化学连接
5'炔基修饰的Seq 1互补序列
NNNNN
N3
铜催化叠氮化物-炔烃环加成（CuAAC）
③ Add-on PCR 接头连接
完整序列的P5接头
完整序列的P7接头

图3

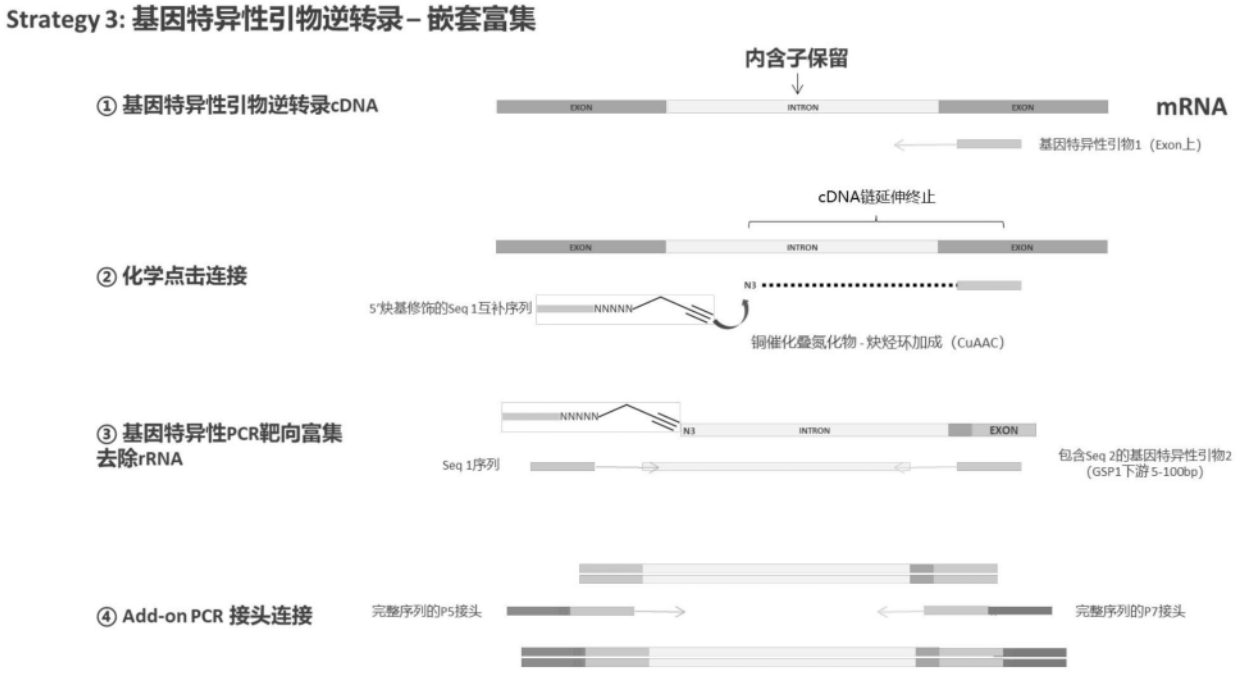

Strategy 3: 基因特异性引物逆转录 – 嵌套富集
内含子保留
① 基因特异性引物逆转录cDNA
EXON
INTRON
EXON
mRNA
基因特异性引物1（Exon上）
cDNA链延伸终止
② 化学点击连接
5'炔基修饰的Seq 1互补序列
NNNNN
N3
铜催化叠氮化物-炔烃环加成（CuAAC）
③ 基因特异性PCR靶向富集
去除rRNA
Seq 1序列
包含Seq 2的基因特异性引物2
(GSP1下游 5-100bp)
④ Add-on PCR 接头连接
完整序列的P5接头
完整序列的P7接头

图4

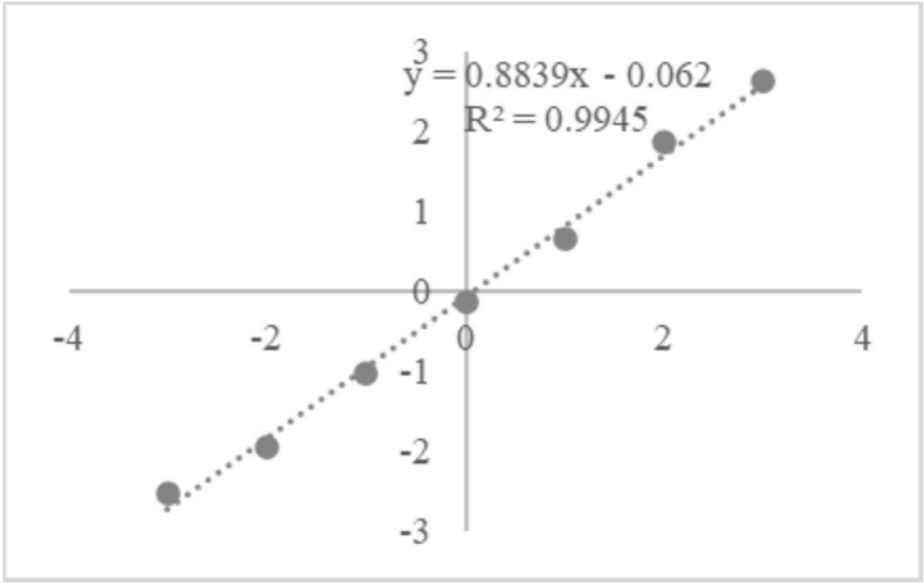


图5