



(12)发明专利申请

(10)申请公布号 CN 111662970 A
(43)申请公布日 2020.09.15

(21)申请号 202010603495.2

(22)申请日 2020.06.29

(71)申请人 武汉菲沙基因信息有限公司
地址 430000 湖北省武汉市东湖高新技术
开发区高新大道666号武汉国家生物
产业基地项目B、C、D区研发楼B1栋

(72)发明人 方涛

(74)专利代理机构 上海精晟知识产权代理有限
公司 31253

代理人 汤蔚莉

(51)Int.Cl.
C12Q 1/6869(2018.01)
C40B 50/06(2006.01)

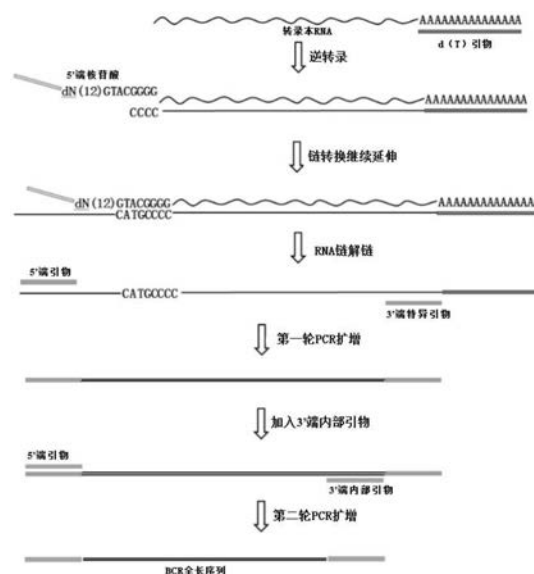
权利要求书2页 说明书11页
序列表7页 附图3页

(54)发明名称

一种BCR免疫组库全长扩增的三代建库测序方法

(57)摘要

本发明公开了一种BCR免疫组库全长扩增的三代建库测序方法,步骤包括:1)根据BCR一致性序列的恒定区设计与合成引物;2)通过Oligo dT引物和BCR Template Switching Oligo引物的作用下合成第一链cDNA;3)BCR cDNA的全长第一轮和第二轮的扩增;4)BCR全长扩增子片段混样;5)文库构建;6)进行三代测序。本发明通过在BCR恒定区的保守区域设计多重引物,覆盖整个BCR的全长序列。在进行BCR序列扩增时,为进一步提高3'端扩增的有效性,设计了多组半巢式引物,通过半巢式引物的匹配筛选即可获得覆盖整个BCR全长的扩增子片段。通过PacBio建库流程,在扩增子的两端添加带有barcode的测序接头,即可进行PacBio测序,通过拆分与数据校正后获得准确率在99%以上的HiFi Reads。



CN 111662970 A

1. 一种BCR免疫组库全长扩增的三代建库测序方法,其特征在于,包括如下步骤:

S1.RNA样本通过BCR Template Switching Oligo引物逆转录合成cDNA第一链,所述BCR Template Switching Oligo引物包含12个碱基UMI标签序列;

S2.在5'端通用引物条件下,利用3'端引物对BCR cDNA全长进行第一轮和第二轮PCR扩增;所述第一轮扩增包含20条引物;所述第二轮扩增包含10组引物,每组2条;

S3.BCR全长扩增子片段混样;

S4.通过末端修复后,连接测序接头,获得文库;

S5.进行三代测序。

2.根据权利要求1所述的一种BCR免疫组库全长扩增的三代建库测序方法,其特征在于:

所述步骤S1中引物及所述S2中5'端通用引物基于BCR免疫组库一致性序列设计并合成;

所述步骤S2中的3'端引物基于BCR免疫组库一致性序列恒定区的保守区域设计并合成,所述引物覆盖整个BCR序列。

3.根据权利要求1所述的一种BCR免疫组库全长扩增的三代建库测序方法,其特征在于:

所述步骤S1中:所述BCR Template Switching Oligo引物的核苷酸序列如SEQ ID NO:2所示;

所述步骤S2中:所述5'端通用引物的核苷酸序列如SEQ ID NO:3所示;所述第一轮扩增的引物的核苷酸序列如SEQ ID NO:4-SEQ ID NO:23所示;所述第二轮扩增的10组引物的核苷酸序列分别为SEQ ID NO:24-25、SEQ ID NO:26-27、SEQ ID NO:28-29、SEQ ID NO:30-31、SEQ ID NO:32-33、SEQ ID NO:34-35、SEQ ID NO:36-37、SEQ ID NO:38-39、SEQ ID NO:40-41、SEQ ID NO:42-43。

4.根据权利要求1所述的一种BCR免疫组库全长扩增的三代建库测序方法,其特征在于,所述步骤S1逆转录前将Oligo dT引物与RNA样本的polyA结合,所述Oligo dT引物核苷酸序列如SEQ ID NO:1所示。

5.根据权利要求1所述的一种BCR免疫组库全长扩增的三代建库测序方法,其特征在于,所述步骤S2中:

所述第一轮扩增利用所述20条引物的混合物进行扩增;

所述第二轮扩增利用所述10组引物分别单独进行扩增。

6.根据权利要求1所述的一种BCR免疫组库全长扩增的三代建库测序方法,其特征在于,所述步骤S2中:

所述第一轮扩增孵育条件为:98℃保持2min;然后98℃保持20s、65℃保持15s、72℃保持45s,共18个循环;72℃保持5min;

所述第二轮扩增孵育条件为:98℃保持2min;然后98℃保持20s、65℃保持15s、72℃保持30s,共20个循环;72℃保持5min。

7.根据权利要求1所述的一种BCR免疫组库全长扩增的三代建库测序方法,其特征在于,所述步骤S3根据Qubit定量结果,将同一样本的不同扩增产物进行等量混样,所述混样后的总量在1ug以上。

8. 根据权利要求1所述的一种BCR免疫组库全长扩增的三代建库测序方法,其特征在于,所述步骤S4为:通过末端修复与加A后,再加入测序接头;所述测序接头为与A末端匹配的带barcode的测序接头。

9. 根据权利要求1所述的一种BCR免疫组库全长扩增的三代建库测序方法,其特征在于,所述步骤S5中所述三代测序前还包括文库质检:取所述步骤S5获得的文库进行Qubit定量,获得文库浓度,取文库进行片段的大小对比,合格后进行测序。

10. 根据权利要求1所述的一种BCR免疫组库全长扩增的三代建库测序方法,其特征在于,所述步骤S5中所述三代测序使用PacBio Sequel II测序平台。

一种BCR免疫组库全长扩增的三代建库测序方法

技术领域

[0001] 本发明涉及基因测序技术领域,尤其涉及一种BCR免疫组库全长扩增的三代建库测序方法。

背景技术

[0002] B细胞受体(B cell receptor,BCR)是B细胞抗原识别决定性表面分子,其本质是一种膜表面免疫球蛋白(membrane immunoglobulin,mIg)。BCR具有抗原结合特异性,每个个体的BCR多样性高达 5×10^{13} ,构成容量巨大的BCR库,赋予个体识别各种抗原、产生特异性抗体的巨大潜能。

[0003] BCR的结构包括2条重链和2条轻链(κ 和 λ)。BCR的重链(heavy chain,H)由65~100种可变区(VH)、2种多变区(DH)、6种结合区(JH)和恒定区(CH)四部分基因片段组成;轻链(light chain,L)由可变区、结合区和恒定区三部分基因片段组成。发育过程中的B细胞在重组酶(RAG1、RAG2)作用下,形成了多样性高达 $1 \sim 2 \times 10^{11}$ 的BCR。同时,由其形成互补决定区(complementarities determining region,CDR):CDR1、CDR2和CDR3区氨基酸序列的多样性,特别是编码CDR3的基因,由于其位于轻链V、J或重链V、D、J片段的连接处,可以通过V(D)J的重排和(或)两个基因片段的连接间丢失或插入数个核苷酸,进一步增加BCR的多样性,而形成具有功能的BCR编码基因(B细胞克隆)。

[0004] BCR测序是通过高通量测序技术检测靶向扩增后的BCR重链和轻链,全面解析BCR基因重排碱基序列,以及各序列的丰度的测序技术。BCR测序常用于评价各种免疫相关疾病和遗传性突变引起的某个物种所有B细胞或特定B细胞激活介导的细胞免疫反应中BCR基因重排碱基序列,以及各序列的丰度,用于研究不同B细胞克隆的转录情况和相互间关系,从而揭示更深层次的B细胞功能特异性,继而解释体液免疫应答耐受以及高频突变在B细胞应答识别抗原异常等相关生命现象。传统的BCR测序都是采用使用Illumina的测序仪进行 $2 \times 300\text{bp}$ 或 $2 \times 150\text{bp}$ 的双端测序方法来对BCR进行测序,该方法的测序准确率高,但对于部分长度超过600bp的BCR序列,该方法仅能获得两端的序列,会存在中间部分关键可变区序列缺失的问题。

[0005] PacBio公司正式推出Sequel II新一代测序系统后,依赖于PacBio SMRT测序技术独有的CCS测序模式,能够通过滚环测序提升reads准确度,同时结合聚合酶试剂的优化,酶读长大幅提升,在获得高精度HiFi reads的同时能够保证10kb以上的插入片段读长,将解决原有的Illumina等二代测序平台下无法完整覆盖整个片段区域的问题。通过进行HiFi测序,可以获得准确性达到99%以上的长读长序列,用于BCR精确分组、BCR重链轻链匹配等应用中。

[0006] 因此有必要针对三代测序平台,研究一种简单快速的BCR免疫组库全长扩增的建库测序方法,利用三代测序的优势,结合特殊的扩增手段,在达到以上目的的同时,以克服传统测序方法带来的缺陷。

发明内容

[0007] 本发明在结合PacBio测序平台的优势基础上,开发出根据BCR免疫组库全长扩增的建库测序方法,经证实该方法快速、特异和有效。

[0008] 为了达到上述目的,本发明提出一种BCR免疫组库全长扩增的三代建库测序方法,包括以下步骤:

[0009] 1.合成cDNA第一链

[0010] 将Oligo dT逆转录引物与样本RNA的poly(A)结合;再通过BCR Template Switching Oligo逆转录合成cDNA。

[0011] 进一步地,所述Oligo dT逆转录引物核苷酸序列如SEQ ID NO:1所示;

[0012] 进一步地,所述逆转录合成过程为poly(A)结合产物在Single Cell RT Buffer、Single Cell RT Enzyme Mix及无核酸酶水条件下42℃孵育75min,再加入BCR Template Switching Oligo引物,42℃孵育15min;

[0013] 进一步地,所述BCR Template Switching Oligo引物基于BCR免疫组库一致性序列设计并合成;

[0014] 进一步地,所述BCR Template Switching Oligo引物的核苷酸序列如SEQ ID NO:2所示。

[0015] 2.BCR cDNA的全长扩增

[0016] cDNA第一链在5'端通用引物和3'端引物组一的条件下进行第一轮扩增反应;然后在5'端通用引物和3'端引物组二的条件下进行第二轮扩增反应。

[0017] 进一步地,所述5'端通用引物和3'端引物均基于BCR免疫组库一致性序列设计;所述3'端引物覆盖整个BCR的全长序列,位置基于BCR恒定区的保守区域;

[0018] 进一步地,所述3'端引物组一包含20条引物;所述引物组二包含10组引物,每组2条;

[0019] 进一步地,所述5'端通用引物的核苷酸序列如SEQ ID NO:3所示;

[0020] 进一步地,所述3'端引物组一的20条引物的核苷酸序列如SEQ ID NO:4-SEQ ID NO:23所示;所述3'端引物组二的核苷酸序列如SEQ ID NO:24-43所示,所述10组引物分别为SEQ ID NO:24-25、SEQ ID NO:26-27、SEQ ID NO:28-29、SEQ ID NO:30-31、SEQ ID NO:32-33、SEQ ID NO:34-35、SEQ ID NO:36-37、SEQ ID NO:38-39、SEQ ID NO:40-41、SEQ ID NO:42-43;

[0021] 进一步地,所述第一轮扩增反应为在20条混合引物条件下进行扩增:在HiFi Fidelity Buffer、dNTP Mix、HiFi Enzyme及无核酸酶水条件下进行孵育:98℃保持2min;然后98℃保持20s、65℃保持15s、72℃保持45s,共18个循环;72℃保持5min;

[0022] 进一步地,所述第二轮扩增过程将所述引物组二的10组引物分别单独并进行扩增;所述第二轮扩增反应为在HiFi Fidelity Buffer、dNTP Mix、HiFi Enzyme及无核酸酶水条件下进行孵育:98℃保持2min;然后98℃保持20s、65℃保持15s、72℃保持30s,共20个循环;72℃保持5min。

[0023] 3.BCR全长扩增子片段混样

[0024] 根据Qubit定量结果,将同一样本的不同扩增产物进行等量混样,所述混样后的总量在1ug以上,用于后续建库。

[0025] 4. 文库构建

[0026] 末端修复并加A后,连接与A末端匹配的带barcode的测序接头,消化不完整的文库,纯化后获得目的文库。

[0027] 进一步地,所述末端修复加A步骤为取全基因组扩增子样本,在DNA Damage Repair Buffer、End Prep Enzyme体系下,20℃孵育30min;然后进行磁珠纯化、洗脱获得;

[0028] 进一步地,所述消耗不完整文库方法为使用外切酶III和外切酶VII进行消耗。

[0029] 5. 进行三代测序

[0030] 对获得的文库进行Qubit定量,获得文库浓度,取文库进行片段大小对比;分析合格后的文库在PacBio Sequel II测序平台上进行混合测序。

[0031] 与现有技术相比,本发明的有益效果为:

[0032] 1) 本发明的方法直接使用总RNA进行逆转录反应,无需先富集具有poly(A)结构的mRNA,在样本总要求上可以达到非常低的水平,样本量达到10ng以上即可进行逆转录与后续的扩增测序;

[0033] 2) 本方法使用链转换原理,在使用d(T)引物进行逆转录至5'末端后,可转换延伸链至5'端核苷酸,在cDNA链上引入5'端引物结合位点,通过一步反应简化了实验流程。同时,在5'端核苷酸上加入了12碱基的UMI序列,可以校正扩增反应导致的偏好性,实现准确定量;

[0034] 3) 本方法针对不同类型的BCR序列,基于BCR的恒定区设计了可覆盖整个BCR基因的半巢式扩增引物,在第一轮扩增时使用的3'端特异性引物可以特异性的扩增富集特定类型的BCR片段,在第二轮扩增时,使用3'端内部引物,可以进一步的富集BCR序列,使扩增富集效果更优异,经过两轮半巢式PCR扩增富集后,测序数据中BCR序列的比例可以达到70%以上,极大的提供了测序数据中BCR序列的比例;

[0035] 4) 本方法通过将第一轮扩增的3'端特异性引物混合为引物混池的方法进行扩增,将需要10个单独的扩增反应简化为仅需要1个扩增反应,极大的简化了实验操作,提高了实验效率。同时,在第二轮扩增时采用单独的引物进行可以防止不同引物扩增效率产生扩增偏差,使测序数据可以均匀的覆盖各目标区域;

[0036] 5) 在测序平台上采用PacBio平台,可以充分发挥PacBio三代测序的长读长优势,同时准确率也可以达到Q20以上,无需拼接即可实现BCR序列的直接读取,可精确进行BCR序列的分类与BCR序列轻重链的匹配。

附图说明

[0037] 图1为BCR免疫组库全长扩增原理示意图。

[0038] 图2为BCR全长扩增序列PacBio建库流程示意图。

[0039] 图3为不同BCR片段扩增电泳图。

[0040] 图4为文库检测峰图。

具体实施方式

[0041] 下面结合实施例对本发明作进一步详细的描述,但本发明的实施方式不限于此。

[0042] 本发明的BCR免疫组库全长扩增流程及原理示意图如图1所示,具体原理如下:

[0043] 1.采用逆转录引物d(T)与RNA转录本的3'端的poly(A)结合,进行逆转录反应,逆转录至5'末端时会额外添加3~5个C碱基,额外添加的C碱基可以和加入的5'端核苷酸序列结合;

[0044] 2.在5'端核苷酸结合后,逆转录酶会继续以5'端为模板进行延伸,至5'端核苷酸的末端;

[0045] 3.以cDNA第一链为模板,加入3'端特异引物与5'端引物后进行第一轮扩增富集;

[0046] 4.加入3'端内部引物与5'端引物后进行第二轮扩增富集,获得BCR全长扩增序列。

[0047] 基于以上原理的BCR全长扩增序列PacBio建库流程示意图如图2所示。

[0048] 实施例1:引物的合成

[0049] 本发明设计的引物序列可以在各引物合成公司进行合成,如上海生工,提供的引物序列如表1所示。

[0050] 表1:设计引物的序列

ID	引物名称	序列(5'to3')
SEQ ID NO: 1	Oligo dT 引物	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT

[0052]

ID	引物名称	序列(5'to3')	
SEQ ID NO: 2	BCR Template Switching Oligo 引物	GTGACTGGAGTTCAGACGTGTGCTCTTCC GATCT[dN(12)]GTACGGGGG	
SEQ ID NO: 3	5'通用引物	TCAGACGTGTGCTCTTCCGATCT	
SEQ ID NO: 4	3'特异性引物组一	TGAGTTCCACGACACCGTCA	
SEQ ID NO: 5		AGTCCTGAGGACTGTAGGACA	
SEQ ID NO: 6		AAGCCCCGGGTGCTGCTGATGT	
SEQ ID NO: 7		TTCTTTGTTGCCGTTGGGGTGCT	
SEQ ID NO: 8		GGCCTCTCTGGGATAGAAGTT	
SEQ ID NO: 9		GTCCTTGCTGTCCTGCTC	
SEQ ID NO: 10		CTGCTATCTGCCTTCCAGGC	
SEQ ID NO: 11		GTKTGGTGGTCTCCACTC	
SEQ ID NO: 12		GTGATGGAGTCGGGAAGGAAGT	
SEQ ID NO: 13		TCTTTGTTGCCGTTGGGGTGCTG	
SEQ ID NO: 14		ACGCTGCTGAGGGAGTAGAGT	
SEQ ID NO: 15		TTGTCCACCTTGGTGTTGCTGG	
SEQ ID NO: 16		GAAGTTTCTGGCGGTCACG	
SEQ ID NO: 17		TCAGCTGGCTGCTCGTGGTGTACA	
SEQ ID NO: 18		GTTGAAGCTCTTTGTGACGG	
SEQ ID NO: 19		CTGATGGGTGACTTCGCAG	
SEQ ID NO: 20		3'特异性引物组一	TCAGGCTCAGRTAGCTGCTG
SEQ ID NO: 21			TTGGAGGGTKTGGTGGTCTCCA
SEQ ID NO: 22			GCAAGCTGATGGTGGCATAG
SEQ ID NO: 23			GGCTGGTAAGGTCATAGTTGTC
SEQ ID NO: 24	3'特异性引物组二-1	AAGACSGATGGGCCCTTGGT	

ID	引物名称	序列(5'to3')
SEQ ID NO: 25		AAGTAGTCCTTGACCAGGCAG
SEQ ID NO: 26	3'特异性引物组二-2	ACGGGAATTCTCACAGGAGA
SEQ ID NO: 27		TGATGGAGTCGGGAAGGAAGT
SEQ ID NO: 28	3'特异性引物组二-3	AAGACAGATGGTGCAGCCACA
SEQ ID NO: 29		ACAACAGAGGCAGTTCAGAT
SEQ ID NO: 30	3'特异性引物组二-4	AGGAGGGYGGGAACAGAGTGA
SEQ ID NO: 31		AGACACACCAGTGTGGCCTTGT
SEQ ID NO: 32	3'特异性引物组二-5	GTTGGGGCGGATGCACTCC
SEQ ID NO: 33		AATTCTCACAGGAGACGAGG
SEQ ID NO: 34	3'特异性引物组二-6	AGTAGTCCTTGACCAGGCAGCC
SEQ ID NO: 35		GGAAGGTGTGCACGCCGCTGGTC
SEQ ID NO: 36	3'特异性引物组二-7	GCGGGAAGACCTTGGGG
SEQ ID NO: 37		CGCTCCAGGTCACACTGAGTGG
SEQ ID NO: 38	3'特异性引物组二-8	CTGCTTTGCTCAGCGTCAGG
SEQ ID NO: 39		CTGTAGGTGCTGTCCTTGCT
SEQ ID NO: 40	3'特异性引物组二-9	ATCTGCCTTCCAGGCCACTGTCA
SEQ ID NO: 41		TGTTGGCTTGRAGCTCCTCAG
SEQ ID NO: 42	3'特异性引物组二-10	GGTCACCATCACCGGCTCCG
SEQ ID NO: 43		GGCAGCCCAGAGTCACGG

[0053] 其中,BCR Template Switching Oligo引物、5'端通用引物均为基于不同类型的BCR免疫组库恒定区的一致性序列设计;3'端特异性引物基于保守区域设计得来的,覆盖整个BCR免疫组库的全长序列。

[0054] 另外,Oligo dT引物中V为简并碱基,代表G、A、C三种之一;BCR Template Switching Oligo引物中[dN(12)]表示为12个碱基UMI标签序列,由12个随机碱基组成,为用于校正扩增的偏好性,即Template Switching Oligo引物核苷酸序列为:

[0055] GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNGTACGGGGG,N代表A、T、C、G四种之一。

[0056] 以上引物中其他简并碱基代表的碱基分别如下:

[0057]	简并碱基	正常碱基
--------	------	------

R	A/G
Y	C/T
K	G/T
S	G/C

[0059] 实施例2:总RNA中cDNA第一链的合成

[0060] 实验步骤中使用到的引物合成于上海生工,将引物统一稀释至10uM。第一步进行cDNA第一链合成的项目实施例操作如下:

[0061] 1) Oligo dT逆转录引物与poly(A)结合

[0062] 表2:

	试剂	生产商	体积
[0063]	总 RNA 样本 (0.1-2ug)	-	7 μ L
	Oligo dT Primer	-	2 μ L
	总体积	-	9 μ L

[0064] 轻弹混匀,瞬时离心,70 $^{\circ}$ C孵育5min后立即置于冰上。

[0065] 2) 逆转录合成cDNA第一链

[0066] 配制如下反应:

[0067] 表3:

	试剂	生产商	体积
	上步产物	-	9 μ L
	NEBNext Single Cell RT Buffer	NEB	5 μ L
[0068]	NEBNext Single Cell RT Enzyme Mix	NEB	2 μ L
	无核酸酶水	NEB	3 μ L
	总体积	-	19 μ L

[0069] 轻弹混匀,瞬时离心,42 $^{\circ}$ C孵育75min。反应完成后立即置于冰上,加入1uL的BCR Template Switching Oligo,轻弹混匀,瞬时离心,42 $^{\circ}$ C孵育15min。

[0070] 实施例3:BCR cDNA的全长扩增

[0071] BCR cDNA的全长扩增包括两轮半巢式扩增反应,第一轮扩增进行BCR序列的初步富集,扩增序列的条带单一性较差,第二轮扩增时采用内部的巢式引物,进一步提高扩增的

特异性,使扩增条带单一。

[0072] 1) BCR cDNA全长的第一轮PCR扩增

[0073] 取一新的0.2mL PCR管,加入如下试剂:

[0074] 表4:

	Reagent	Volume	说明
[0075]	cDNA 第一链产物	20 μ L	
	5' 通用引物 (10 μ M)	2 μ L	
	3'特异性引物一	2 μ L	该引物为多条引物组成的混合引物
	KAPA HiFi Fidelity Buffer (5X)	10 μ L	
[0076]	KAPA dNTP Mix (10 mM)	1.5 μ L	
	KAPA HiFi Enzyme (1U/ μ L)	1 μ L	
	Nuclease-free water	13.5 μ L	
	Total Volume	50 μ L	

[0077] 充分混匀,瞬时离心,置于PCR仪上进行PCR反应:98 $^{\circ}$ C 孵育2min;98 $^{\circ}$ C 孵育20s、65 $^{\circ}$ C 孵育15s、72 $^{\circ}$ C 孵育45s,共18个循环;然后72 $^{\circ}$ C 孵育5min。

[0078] 反应结束后可以在-20 $^{\circ}$ C冰箱长期保存。

[0079] 在第一次扩增PCR反应开始时,温度达到98 $^{\circ}$ C,此时温度已经达到了RNA链解链所需的温度,因此RNA链在扩增反应开始时发生了解链。

[0080] 2) BCR cDNA全长的第二轮PCR扩增

[0081] 取一新的0.2mL PCR管,加入如下试剂:

[0082] 表5:

	Reagent	Volume	说明
	第一轮 PCR 扩增产物	5 μ L	
	5' 通用引物 (10 μ M)	2 μ L	
[0083]	3'特异性引物二	2 μ L	该引物包含 10 组混合引物, 每组混合引物均单独进行扩增
	KAPA HiFi Fidelity Buffer (5X)	10 μ L	
	KAPA dNTP Mix (10 mM)	1.5 μ L	
	KAPA HiFi Enzyme (1U/ μ L)	1 μ L	
[0084]	Nuclease-free water	28.5 μ L	
	Total Volume	50 μ L	

[0085] 10组混合引物均需要单独进行扩增,可以提高扩增反应的稳定性。

[0086] 充分混匀,瞬时离心,置于PCR仪上进行PCR反应:98 $^{\circ}$ C 孵育2min;98 $^{\circ}$ C 孵育20s、65 $^{\circ}$ C 孵育15s、72 $^{\circ}$ C 孵育30s、共20个循环;然后72 $^{\circ}$ C 孵育5min。

[0087] 反应结束后进行电泳检测图如图3所示,可以看到明显的单一扩增条带。

[0088] 按照AMPure磁珠的使用说明对扩增产物进行磁珠纯化,最后使用10 μ L洗脱缓冲液洗脱,取1 μ L纯化产物,使用无核酸酶水稀释5倍后进行Qubit定量。

[0089] 实施例4:BCR全长扩增子片段混样

[0090] 根据Qubit定量结果,将同一样本的不同扩增产物进行等量混样,要求混样后的总量在1 μ g以上,用于后续建库。对于仅关注某些特殊BCR全长序列的应用,可以仅挑选部分扩增产物进行混样后测序。

[0091] 实施例5:文库构建

[0092] 1) 末端修复

[0093] 取1 μ g的全基因组扩增子样本,进行末端修复反应体系的配制,配制如下反应:

[0094] 表6:

	试剂	生产商	体积
	全基因组扩增子样本(1ug)	-	20.5 μ L
[0095]	DNA Damage Repair Buffer	Vazyme	2.5uL
	End Prep Enzyme	Vazyme	2uL
	总体积	-	25 μ L

[0096] 充分混匀,瞬时离心,20 $^{\circ}$ C孵育30min。

[0097] 反应完成后,按照AMPure磁珠的使用说明进行1X磁珠纯化,去除反应时加入的酶与Buffer,最后使用14 μ L洗脱缓冲液洗脱,获得片段末端加A的粘性末端。

[0098] 2) 连接带barcode的测序接头

[0099] 通过末端修复与加A后,在加入具有与A末端匹配的带barcode的测序接头,在连接酶的作用下即可实现接头的连接。同时,由于酶切与连接反应存在一定的效率问题,需要使用外切酶III和外切酶VII将结构不完整的文库消化掉。具体实施例的操作流程如下:

[0100] 进行连接反应,反应体系如下:

[0101] 表7:

	试剂	生产商	体积
	上步纯化产物	-	14 μ L
[0102]	Rapid Ligation Buffer 3	Vazyme	6uL
	Rapid DNA Ligase	Vazyme	2.5uL
	Barcode Adapter (20uM)	-	2uL
	Total Volume	-	24.5 μ L

[0103] 混匀,瞬时离心,20 $^{\circ}$ C孵育60min,反应结束后65 $^{\circ}$ C孵育10min,置于冰上。

[0104] 进行外切酶消化,反应体系如下:

[0105] 表8:

	试剂	生产商	体积
[0106]	连接产物	-	24.5 μ L
	外切酶 III	Vazyme	0.5 μ L

[0107]	外切酶 VII	Vazyme	0.5 μ L
	Total Volume	-	25.5 μ L

[0108] 混匀,瞬时离心,37 $^{\circ}$ C孵育60min,置于冰上。按照AMPure磁珠的使用说明进行磁珠纯化,最后使用20 μ L洗脱缓冲液洗脱,获得的适用于PacBio测序平台的哑铃型的环状文库。

[0109] 实施例6:文库质检与上机测序

[0110] 取1 μ L文库进行Qubit定量,获得文库浓度;取1 μ L文库进行安捷伦2100的片段大小分析,文库检测峰图如图4所示,从图中可以看出,PCR扩增出来的DNA目的条带长度与目的片段的大小一致,文库大小也与目的片段大小相符,说明制备的文库合格。

[0111] 由以上实施例获得的BCR全长扩增文库在PacBio Sequel II测序平台上进行混合测序,每个样本获得了约30G的测序数据,说明文库质量合格且可获得符合测序要求的数据。

[0112] 将测序的CCS数据进行一致性校正后,获得了质量值在Q20以上的BCR全长一致性序列,经过与BCR数据库进行比对后,获得了每个样本测序数据中的BCR序列比例。从以下统计列表可以看出,在未进行BCR扩增富集时,测序数据中的BCR序列在1%以下,当经过本发明的方法进行富集后,测序获得的数据有70%以上为BCR的序列,表明本发明可以显著的对BCR序列进行富集,实现BCR全长序列的测序。

[0113] 附本实施例中BCR全长扩增测序数据分析的比例数据,见表9。

[0114] 表9:

[0115]	样本名称	未富集数据中 BCR 的比例	扩增富集后 BCR 数据的比例	扩增富集倍数
	全血一	0.03%	79.1%	2637
	全血二	0.01%	73.2%	2440
[0116]	全血三	0.01%	72.1%	2403
	全血四	0.02%	76.5%	2550
	全血五	0.01%	73.5%	2450

[0117] 本发明并不仅仅限于说明书和实施方式中所描述,因此对于熟悉领域的人员而言可容易地实现另外的优点和改进,故在不背离权利要求及等同范围所限定的一般概念的精神和范围的情况下,本发明并不限于特定的细节、代表性的方案和这里示出与描述的图示示例。

序列表

<110> 武汉菲沙基因信息有限公司

<120> 一种BCR免疫组库全长扩增的三代建库测序方法

<141> 2020-06-24

<160> 43

<170> SIPOSequenceListing 1.0

<210> 1

<211> 31

<212> DNA

<213> 人工序列(Artificial Sequence)

<400> 1

tttttttttt tttttttttt tttttttttt v 31

<210> 2

<211> 55

<212> DNA

<213> 人工序列(Artificial Sequence)

<400> 2

gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnngtac ggggg 55

<210> 3

<211> 23

<212> DNA

<213> 人工序列(Artificial Sequence)

<400> 3

tcagacgtgt gctcttccga tct 23

<210> 4

<211> 20

<212> DNA

<213> 人工序列(Artificial Sequence)

<400> 4

tgagttccac gacaccgtca 20

<210> 5

<211> 21

<212> DNA

<213> 人工序列(Artificial Sequence)

<400> 5

agtcctgagg actgtaggac a 21

<210> 6

<211> 22

<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 6
aagccccggg tgctgctgat gt 22
<210> 7
<211> 23
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 7
ttctttgttg ccgttgggt gct 23
<210> 8
<211> 21
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 8
ggcctctctg ggatagaagt t 21
<210> 9
<211> 18
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 9
gtccttgctg tcctgctc 18
<210> 10
<211> 20
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 10
ctgctatctg ccttccaggc 20
<210> 11
<211> 18
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 11
gktkggtggt ctccactc 18
<210> 12
<211> 22
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 12

gtgatggagt cgggaaggaa gt 22
<210> 13
<211> 23
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 13
tctttgttgc cgttggggtg ctg 23
<210> 14
<211> 21
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 14
acgctgctga gggagtagag t 21
<210> 15
<211> 22
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 15
ttgtccacct tgggttgct gg 22
<210> 16
<211> 19
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 16
gaagtttctg gcggtcacg 19
<210> 17
<211> 24
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 17
tcagctggct gctcgtggtg taca 24
<210> 18
<211> 20
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 18
gttgaagctc tttgtgacgg 20
<210> 19
<211> 19

<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 19
ctgatgggtg acttcgcag 19
<210> 20
<211> 20
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 20
tcaggctcag rtagctgctg 20
<210> 21
<211> 23
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 21
tttggagggt ktggtgtct cca 23
<210> 22
<211> 20
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 22
gcaagctgat ggtggcatag 20
<210> 23
<211> 22
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 23
ggctggtaag gtcatagttg tc 22
<210> 24
<211> 20
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 24
aagacsgatg ggcccttgg 20
<210> 25
<211> 21
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 25

aagtagtcct tgaccaggca g 21
<210> 26
<211> 21
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 26

acggggaatt ctcacaggag a 21
<210> 27
<211> 21
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 27

tgatggagtc gggaaggaag t 21
<210> 28
<211> 21
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 28

aagacagatg gtcagccac a 21
<210> 29
<211> 21
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 29

acaacagagg cagttccaga t 21
<210> 30
<211> 21
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 30

aggaggygg gaacagagtg a 21
<210> 31
<211> 22
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 31

agacacacca gtgtggcett gt 22
<210> 32
<211> 19

<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 32
gttggggcgg atgcactcc 19
<210> 33
<211> 20
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 33
aattctcaca ggagacgagg 20
<210> 34
<211> 22
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 34
agtagtcctt gaccaggcag cc 22
<210> 35
<211> 23
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 35
ggaaggtgtg cacgccgctg gtc 23
<210> 36
<211> 17
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 36
gcgggaagac cttgggg 17
<210> 37
<211> 22
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 37
cgctccaggt cacactgagt gg 22
<210> 38
<211> 20
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 38

ctgctttgct cagcgtcagg 20
<210> 39
<211> 20
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 39
ctgtaggtgc tgccttgct 20
<210> 40
<211> 23
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 40
atctgccttc caggccactg tca 23
<210> 41
<211> 21
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 41
tgttggcttg ragctcctca g 21
<210> 42
<211> 20
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 42
ggcaccatc accggctccg 20
<210> 43
<211> 18
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 43
ggcagcccag agtcacgg 18

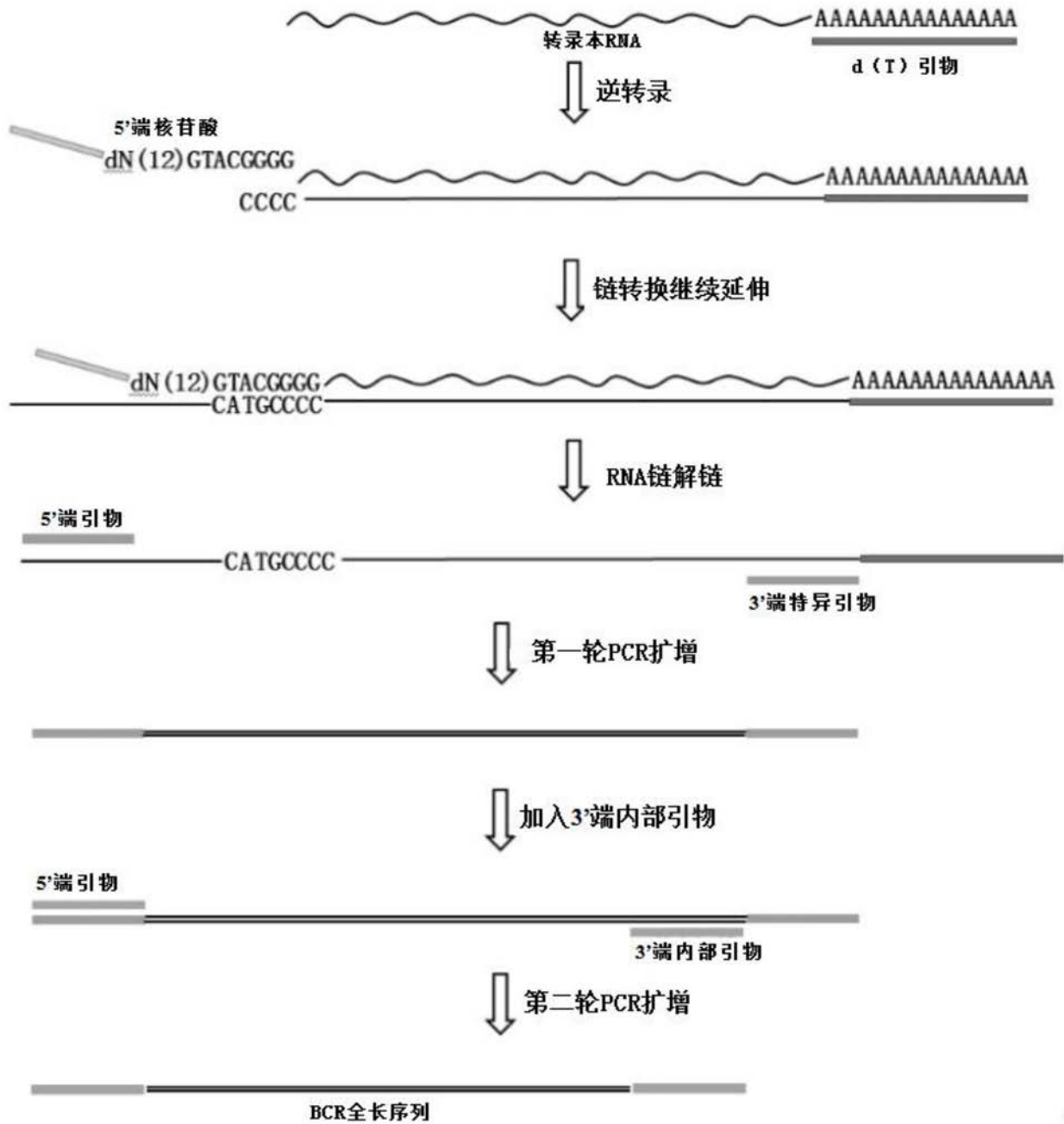


图1

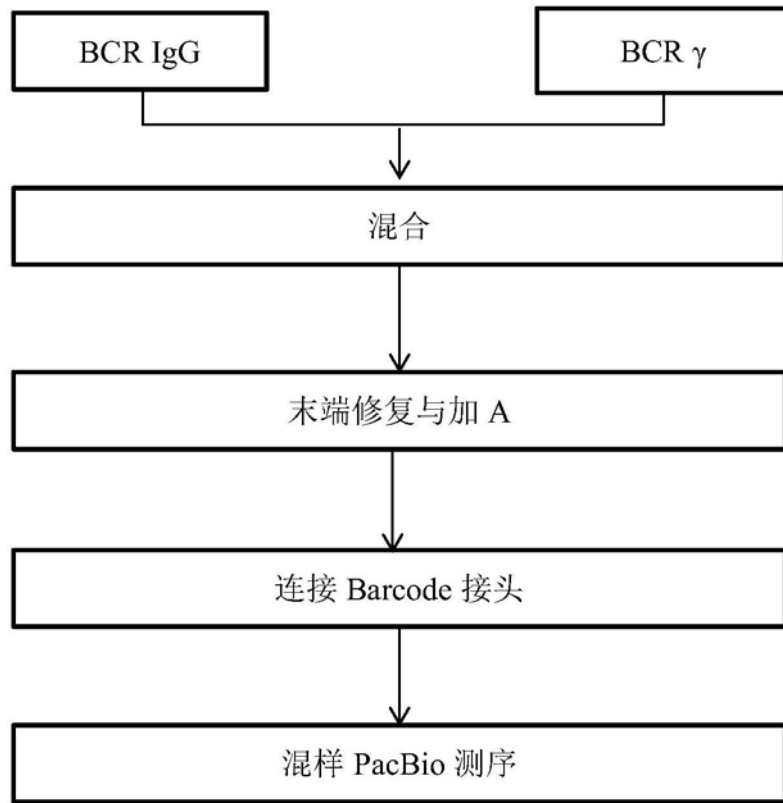


图2

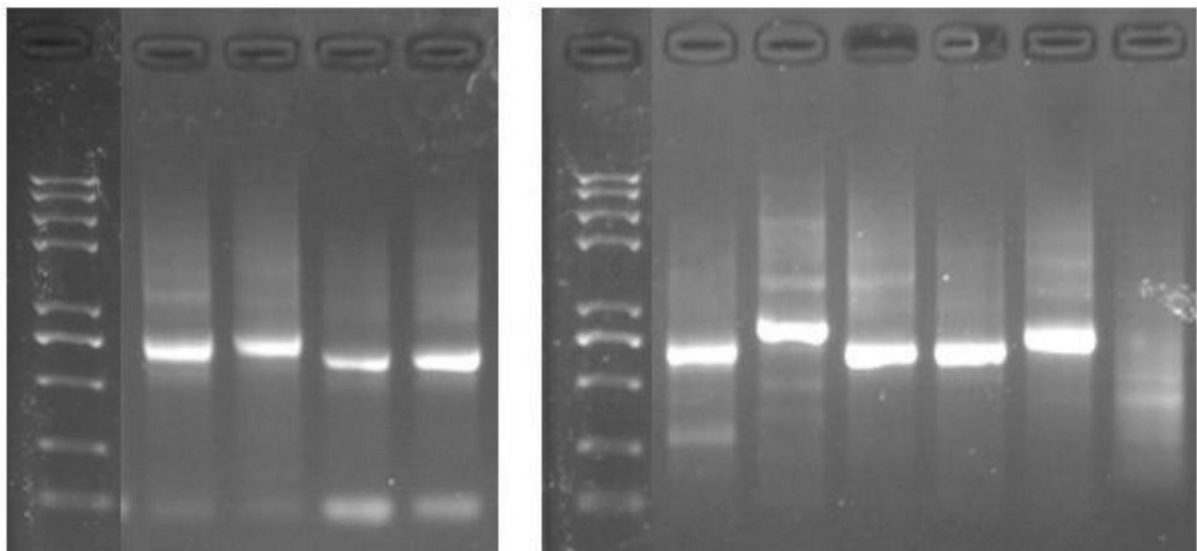


图3

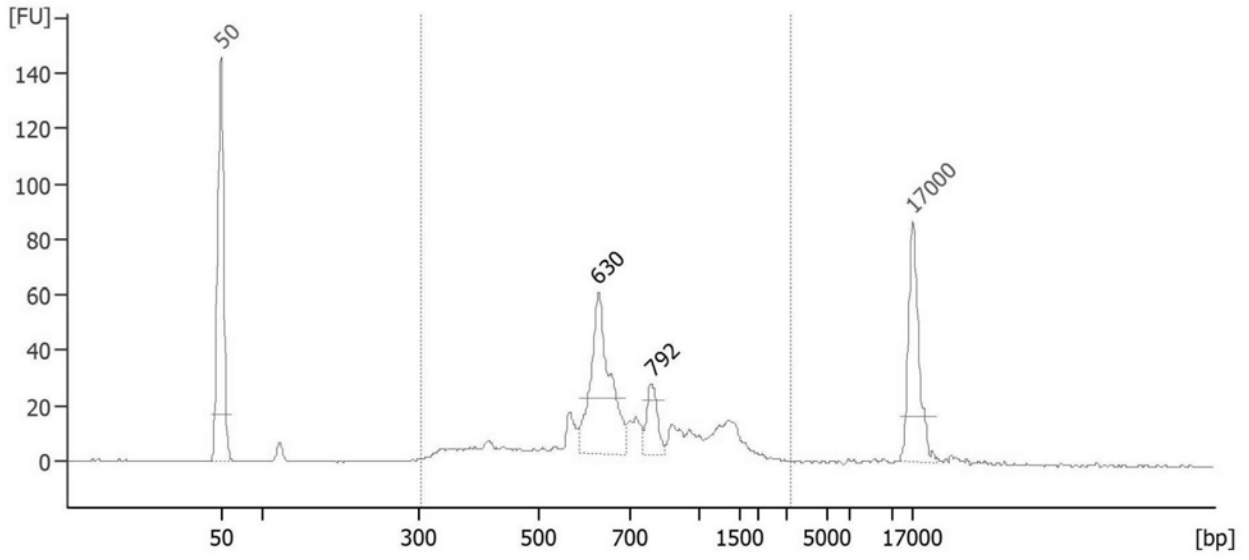


图4