



(12) 发明专利

(10) 授权公告号 CN 109411011 B

(45) 授权公告日 2022.05.17

(21) 申请号 201811313752.8

G16B 25/10 (2019.01)

(22) 申请日 2018.11.06

G06K 9/62 (2022.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 109411011 A

(56) 对比文件

CN 107760672 A, 2018.03.06

CN 105925681 A, 2016.09.07

(43) 申请公布日 2019.03.01

CN 103890245 A, 2014.06.25

(73) 专利权人 苏州泓迅生物科技股份有限公司

CN 107964036 A, 2018.04.27

地址 215123 江苏省苏州市苏州工业园区

CN 107435065 A, 2017.12.05

星湖街218号生物纳米园C20栋

CN 103409544 A, 2013.11.27

(72) 发明人 吴婷婷 李彦敏 陈淑美 蔡晓辉  
杨平

审查员 宋鹏飞

(74) 专利代理机构 北京品源专利代理有限公司  
11332

专利代理师 巩克栋

(51) Int. Cl.

G16B 5/00 (2019.01)

权利要求书2页 说明书11页

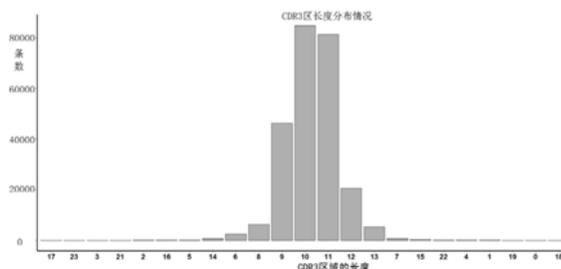
序列表3页 附图2页

(54) 发明名称

一种引物组的设计方法及其应用

(57) 摘要

本发明提供了一种引物组的设计方法及其应用,通过根据扩增长度的需要,巧妙地在抗体FR1起始位点的上游,提取FR1起始位点前的序列,并从5'端的第一个碱基开始,依次移位切取固定长度的引物序列,形成候选引物库,聚类分析后筛选得到免疫组库的引物组,采用该方法得到的引物组能够显著提高多重PCR法免疫组库扩增的覆盖度和实验效率,降低扩增错配率,方法简便,节省成本。



1. 一种引物组的设计方法,其特征在于,包括如下步骤:

(1) 获取物种的抗体重链和/或轻链的种系基因数据和物种的基因组数据;

(2) 引物选择区域定位:将抗体重链和/或轻链的种系基因比对到物种的基因组数据上,找到抗体FR1的起始位点,提取起始位点前的序列;

(3) 候选引物库设计:对步骤(2)提取的起始位点前的序列进行切片,从5'端的第一个碱基开始,依次移位切取固定长度的引物序列,形成候选引物库;

(4) 进行cluster聚类分析,筛选得到初始引物组;

所述聚类分析的方法为:通过Cd-hit软件,设置-c参数,将候选引物库序列按照-c参数设定进行聚类;

所述-c参数设置为0.8-1.0;

所述筛选的方法为:将cluster根据大小排序,选择前10-90的聚类,提取每个cluster的代表序列,筛选得到初始引物组;

(5) 将初始引物组与抗体数据库比对,筛选得到最终引物组。

2. 根据权利要求1所述的方法,其特征在于,步骤(1)所述物种包括人、小鼠、大鼠、兔、恒河猴、绵羊、猪、硬骨鱼类、软骨鱼类、鸭嘴兽、羊驼、食蟹猕猴、牛、狗或鸡中的任意一种或至少两种的组合。

3. 根据权利要求1所述的方法,其特征在于,步骤(2)所述起始位点前的序列的长度范围为1-300bp。

4. 根据权利要求1所述的方法,其特征在于,步骤(3)所述引物序列的固定长度范围为16-26bp。

5. 根据权利要求4所述的方法,其特征在于,步骤(3)所述引物序列的固定长度范围为18-22bp。

6. 根据权利要求1所述的方法,其特征在于,步骤(4)所述-c参数设置为0.9。

7. 根据权利要求1所述的方法,其特征在于,步骤(5)所述比对的方法为:将初始引物组与所对应物种的抗体数据库数据比对,得到每条序列在抗体库的覆盖度,将覆盖度按照大小排序,选取排名前10-40的引物作为最终引物组。

8. 根据权利要求1所述的方法,其特征在于,具体包括如下步骤:

(1) 获取物种的抗体轻链和/或重链的种系基因数据以及物种的基因组数据;

其中,所述物种包括人、小鼠、大鼠、兔、恒河猴、绵羊、猪、硬骨鱼类、软骨鱼类、鸭嘴兽、羊驼、食蟹猕猴、牛、狗或鸡中的任意一种或至少两种的组合;

(2) 引物选择区域定位:将抗体重链和/或轻链的种系基因比对到物种抗体所在染色体的基因组数据上,找到抗体FR1的起始位点,提取起始位点前1-300bp的序列;

(3) 候选引物库设计:对步骤(2)提取的起始位点前的序列进行切片,从5'端的第一个碱基开始,依次移位切取固定长度的引物序列,引物的长度范围为18-22bp,形成候选引物库;

(4) 进行cluster聚类分析,筛选得到初始引物组;

所述聚类分析的方法为:通过Cd-hit软件,设置-c参数,将候选引物库序列按照-c参数设定进行聚类;

所述-c参数的大小设置为0.9;

所述筛选的方法为:将cluster根据大小排序,选择前10-90的聚类,提取每个cluster的代表序列,筛选得到初始引物组;

(5) 将初始引物组与所对应物种的抗体数据库比对,得到每条引物在抗体库的覆盖度,将覆盖度按照大小排序,选取排名前10-40的引物作为最终引物组。

9. 一种采用如权利要求1-8中任一项所述方法设计得到的人B细胞免疫组库重链引物组,其特征在于,所述引物组的序列包含如SEQ NO.1-SEQ NO.16所示序列。

10. 根据权利要求9所述的人B细胞免疫组库重链引物组,其特征在于,所述引物组的序列包含如SEQ ID NO.1-SEQ ID NO.16所示序列中的任意十五条引物序列。

11. 根据权利要求10所述的人B细胞免疫组库重链引物组,其特征在于,所述引物组的序列包含如SEQ ID NO.1-SEQ ID NO.16所示序列中的任意十条引物序列。

12. 一种如权利要求1-8中任一项所述方法或权利要求9-11中任一项所述引物组在免疫组库扩增中的应用的装置。

13. 一种如权利要求1-8中任一项所述方法或权利要求9-11中任一项所述引物组用于制备扩增免疫组库的试剂盒的应用的装置。

## 一种引物组的设计方法及其应用

### 技术领域

[0001] 本发明属于生物技术领域,涉及一种引物组的设计方法及其应用,尤其涉及一种免疫组库的引物组的设计方法及其应用。

### 背景技术

[0002] 免疫组库(Immune Repertoire, IR)是指在任何指定时间内,某个个体的循环系统中所有功能多样性B细胞和T细胞的多态性总和。人体负责保卫机体的免疫细胞主要有T细胞、B细胞、巨噬细胞、树突细胞等。T、B细胞是人体主要的淋巴细胞,分别负责细胞免疫和体液免疫,其中,B细胞约占外周淋巴细胞总数的20%。BCR是由两条重链和两条轻链连接而成的,其中重链分为可变区(V区)、恒定区(C区)、跨膜区及胞质区;轻链则只有V区和C区。其中,V区由VH和VL两个结构域组成,各由三个互补决定区(CDR)组成,分别为CDR1、CDR2和CDR3,此三个CDR区共同参与抗体对抗原的识别,共同决定BCR、TCR的抗原特异性。个体中CDR区的氨基酸组成和排列顺序呈现高度多样性,在同一个体内,CDR区多样性可达 $10^9$ - $10^{12}$ ,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。目前新兴的免疫组库研究重点就集中在研究CDR基因的多样性上。因此,需要进行免疫组库的扩增。

[0003] 目前,免疫组库扩增方法主要有:5'-RACE法、多重PCR方法、唯一分子标识符(unique molecular identifiers,简称UID)法等。

[0004] 5'-RACE法即5'末端的快速扩增法,该法用特异性引物进行逆转录后,在cDNA的一链5'末端加入接头,进行二次无偏移的PCR扩增,并通过亲和素磁珠富集,获得含目标区域的序列。该方法扩增引物只需要基因特异性引物如BCR/TCR的C区保守区引物,可以减少多重PCR偏差,但是,该法只能用于扩增RNA并且要分选出所要研究的特定类型的细胞;实验较普通多重PCR复杂,且有基因转录本长度和GC含量的偏好性。该方法使用的引物仅仅设计在C区一端,由于是快速扩增,所以产物长度范围较大。5'RACE法,可实现不同克隆等效扩增,单一克隆数高达 $10^5$ ,可从低丰度的转录本中快速扩增cDNA的5'末端,扩增区域CDR。可以最大程度的避免PCR扩增的偏向性。该方法具有操作繁琐,实验样本打断会丢失部分序列,重复性较差等缺陷。

[0005] 多重PCR法:在同一PCR反应体系里加入两对以上引物,同时扩增出多个核酸片段的PCR反应,其反应原理、反应试剂、操作过程与一般PCR相同。多重PCR与普通PCR相同,不用将样本打断,数据完整,但是扩增具有偏向性。免疫基因多样性的进化通过基因复制、基因突变完成,可通过设计多重PCR扩增引物扩增得到目的BCR/TCR基因。该方法通常在V区和J/C区的保守区域设计引物实现多重PCR扩增,但由于引物不同PCR扩增的效率必然不同造成扩增偏差(amplification bias),有些引物被大量扩增有些却几乎没有扩出来,只有通过几次优化找到最优的引物浓度组合来消除PCR扩增偏差,但这种方法对新的引物不具有通用性,因而增加了寻找最佳引物浓度的复杂性。通常可变区基因的克隆,是参考Kabat数据库中的抗体序列,针对抗体可变区的保守区域设计若干套通用引物,采用RT-PCR法从人的

淋巴B细胞的cDNA文库中扩增可变区基因,该方法简单实用,通常5'端通用引物设计在第一骨架区或前导肽区;3'端通用引物设计在恒定区或J链区,但是该方法得到的抗体序列比较短,FR1前的序列无法测到。

[0006] UID法,即在大规模PCR扩增目标分子前为每个分子加入唯一的UID,这种UID是随机合成的一般12~16个寡聚核苷酸(random barcode),他们的随机组合会产生庞大的数量来为样品中特定的每个分子加上不同的标签,这样即使PCR扩增不均匀也可以通过计算的方法消除这种偏差,同时还能矫正PCR和测序错误。但这种方法需要很长的引物,会导致扩增效率下降和目标基因片段缩短,并且需要极高的通量才能覆盖所有的UID,该方法目前主要应用于IGH/TCR的RNA测序。

[0007] 针对免疫组库的引物设计,目前技术很难设计一个上游引物用于识别BCR基因的5'端序列,根据FR1的区域进行引物设计会导致抗体的上游信息无法覆盖到,使得测序基因信息不完整;对于文献2(Wang X,Stollar BD.Human immunoglobulin variable region gene analysis by single cell RT-PCR.J Immunol Methods 2000;244:217-25)中的引物,虽然引物的数量较少,但是覆盖度较低。多重引物PCR技术的免疫组库测序方法,仅仅对BCR测序的话,就需要几十对引物,这样,庞大的引物数量会让整个PCR的扩增效率和特异性变的很差。

[0008] 鉴于现有的抗体免疫组库扩增存在上述各种缺陷,因此,研发一种简洁高效、覆盖率高、错配率低的引物组设计方法,具有广阔的应用前景和巨大的市场价值。

## 发明内容

[0009] 针对现有技术的不足及实际的需求,本发明提供一种引物组的设计方法及其应用,巧妙地提取抗体FR1的起始位点前的序列并设计候选引物库,通过生物信息学设计分析,筛选得到引物组,该方法适于所有具有免疫系统的物种免疫组库的引物组的设计,具有方法简单、覆盖度高、错配率低、效率高等优势,并可大大节省免疫组库扩增的成本。

[0010] 为达此目的,本发明采用以下技术方案:

[0011] 第一方面,本发明提供一种引物组的设计方法,包括如下步骤:

[0012] (1) 获取物种的抗体重链和/或轻链的种系基因数据以及物种的基因组数据;

[0013] (2) 引物选择区域定位:将抗体重链和/或轻链的种系基因比对到物种的基因组数据上,找到抗体FR1的起始位点,提取起始位点前的序列;

[0014] (3) 候选引物库设计:对步骤(2)提取的起始位点前的序列进行切片,从5'端的第一个碱基开始,依次移位切取固定长度的引物序列,形成候选引物库;

[0015] (4) 进行cluster聚类分析,筛选得到初始引物组;

[0016] (5) 将初始引物组与抗体数据库比对,筛选得到最终引物组。

[0017] 本发明中,发明人在长期的科研实践中,针对现有技术中免疫组库引物组设计存在的缺陷和弊端,为解决引物数量过多导致的效率降低、引物数量较少产生的覆盖度较低以及无法覆盖上游信息等问题,根据扩增长度的需要,巧妙地在抗体FR1起始位点的上游,提取FR1起始位点前的序列,并从5'端的第一个碱基开始,依次移位切取固定长度的引物序列,形成候选引物库,聚类分析后筛选得到引物组,经过实验验证,采用该方法得到的引物组能够显著提高覆盖度和实验效率,降低错配率,方法简单,节省成本。

[0018] 优选地,步骤(1)所述物种包括人、小鼠、大鼠、兔、恒河猴、绵羊、猪、硬骨鱼类、软骨鱼类、大西洋鳕鱼、海峡鲈鱼、彩虹鳟鱼、斑马鱼、鸭嘴兽、羊驼、食蟹猕猴、牛、狗、鸡或三文鱼中的任意一种或至少两种的组合。

[0019] 优选地,步骤(2)所述起始位点前的序列的长度范围为1-300bp,例如可以是1bp、5bp、10bp、20bp、30bp、40bp、50bp、60bp、70bp、80bp、90bp、100bp、110bp、120bp、130bp、140bp、150bp、160bp、170bp、180bp、190bp、200bp、210bp、220bp、230bp、240bp、250bp、260bp、270bp、280bp、290bp或300bp。

[0020] 优选地,步骤(3)所述引物序列的固定长度范围为16-26bp,例如可以是16bp、17bp、18bp、19bp、20bp、21bp或22bp,优选为18-22bp。

[0021] 优选地,步骤(4)所述聚类分析的方法为:通过Cd-hit软件,设置-c参数,将候选引物库序列按照-c参数设定进行聚类。

[0022] 优选地,所述-c参数设置为0.8-1.0,例如可以是0.8、0.9或1.0,优选为0.9。

[0023] 优选地,步骤(4)所述筛选的方法为:将cluster根据大小排序,选择前10-90的聚类,提取每个cluster的代表序列,根据引物设计原则,筛选得到初始引物组。

[0024] 所述前10-90的聚类例如可以是前10、前20、前30、前40、前50、前60、前70、前80或前90。

[0025] 优选地,步骤(5)所述比对的方法为:将初始引物组与所对应物种的抗体数据库比对,得到每条序列在抗体库的覆盖度,将该值按照大小排序,选取排名前10-40的引物作为最终引物组。

[0026] 所述前10-40的引物例如可以是前10、前16、前20、前25、前30、前35或前40。

[0027] 作为优选技术方案,一种引物组的设计方法,具体包括如下步骤:

[0028] (1) 获取物种的抗体轻链和/或重链的种系基因数据以及物种的基因组数据;

[0029] 其中,所述物种包括人、小鼠、大鼠、兔、恒河猴、绵羊、猪、硬骨鱼类、软骨鱼类、大西洋鳕鱼、海峡鲈鱼、彩虹鳟鱼、斑马鱼、鸭嘴兽、羊驼、食蟹猕猴、牛、狗、鸡或三文鱼中的任意一种或至少两种的组合;

[0030] (2) 引物选择区域定位:将抗体重链和/或轻链的种系基因比对到物种的抗体所在染色体的基因组数据上,找到抗体FR1的起始位点,提取起始位点前1-300bp的序列;

[0031] (3) 候选引物库设计:对步骤(2)提取的起始位点前的序列进行切片,从5'端的第一个碱基开始,依次移位切取固定长度的引物序列,引物的长度范围为18-22bp,形成候选引物库;

[0032] (4) 进行cluster聚类分析,筛选得到初始引物组;

[0033] 所述聚类分析的方法为:通过Cd-hit软件,设置-c参数,将候选引物库序列按照-c参数设定进行聚类;

[0034] 所述-c参数的大小设置为0.9;

[0035] 所述筛选的方法为:将cluster根据大小排序,选择前10-90的聚类,提取每个cluster的代表序列,根据引物设计原则,筛选得到初始引物组;

[0036] (5) 将初始引物组与对应物种的抗体数据库比对,得到每条引物在抗体库的覆盖度,将该值按照大小排序,选取排名前10-40的引物作为最终引物组。

[0037] 第二方面,本发明提供一种采用第一方面所述方法设计得到的人B细胞免疫组库

重链引物组,所述引物组的序列包含如SEQ NO.1-SEQ NO.16所示序列;

[0038] 详细序列如下:

[0039] SEQ NO.1:ATGGACATACTTTGTTCCA

[0040] SEQ NO.2:CCATGGAGTTTGGGCTGAGC

[0041] SEQ NO.3:GGCTGAGCTGGGTTTTCTT

[0042] SEQ NO.4:CTGAGCTGGGTTTTCTTGT

[0043] SEQ NO.5:CTCCTGGTGGCAGCTCCCAG

[0044] SEQ NO.6:CCTCCTCCTGGTGGCAGCTC

[0045] SEQ NO.7:CAGCTCCCAGATGTGAGTGT

[0046] SEQ NO.8:GCTGGGTTTTCTTGTGCT

[0047] SEQ NO.9:ATGAAACACCTGTGGTTCTT

[0048] SEQ NO.10:CCTGGAGGATCCTCTTCTTG

[0049] SEQ NO.11:GGTTTTCTTGTGCTATTT

[0050] SEQ NO.12:TGGTGGCAGCTCCCAGATGT

[0051] SEQ NO.13:GGACGTGAGTGAGAGAAACA

[0052] SEQ NO.14:TCCTCACCATGGACTGGACC

[0053] SEQ NO.15:CTTGTTGGTATTTTAAAAGG

[0054] SEQ NO.16:GAGGATCCTCTTCTTGGTGG.

[0055] 优选地,所述引物组的序列包含如SEQ ID NO.1-SEQ ID NO.16所示序列中的任意十五条引物序列。

[0056] 所述任意十五条引物序列为随机从SEQ ID NO.1-SEQ ID NO.16所示序列中选取得到。

[0057] 进一步优选地,所述引物组的序列包含如SEQ ID NO.1-SEQ ID NO.16所示序列中的任意十条引物序列。

[0058] 所述任意十条引物序列为随机从SEQ ID NO.1-SEQ ID NO.16所示序列中选取得到。

[0059] 第三方面,本发明提供一种如第一方面所述方法或第二方面所述引物组在免疫组库扩增中的应用。

[0060] 第四方面,本发明提供一种如第一方面所述方法或第二方面所述引物组用于制备扩增免疫组库的试剂盒的应用。

[0061] 与现有技术相比,本发明具有如下有益效果:

[0062] 本发明提供的方法简洁高效,能显著提高多重PCR的反应效率和覆盖度,覆盖度高达95%-100%,降低错配率,抗体的完整性更高,节省成本。

## 附图说明

[0063] 图1为本发明的人B细胞免疫组库PCR扩增电泳图;

[0064] 图2为本发明的引物组扩增出的CDR3的长度分布图;

[0065] 图3为本发明的IGVH的CDR3的种系基因(germline)分布图。

### 具体实施方式

[0066] 为更进一步阐述本发明所采取的技术手段及其效果,以下结合附图并通过具体实施方式来进一步说明本发明的技术方案,但本发明并非局限在实施例范围内。

[0067] 实施例1人B细胞免疫组库引物组的设计

[0068] (1) 提取人的germline基因

[0069] 从IMGT数据库(<http://imgt.org/vquest/refseqh.html>)下载人的重链germline基因:提取出361条基因,其中1条FASTA序列信息如下:

[0070] >M99641|IGHV1-18\*01|Homo sapiens|F|V-REGION|188..483|296nt|1|||296+24=320|||

[0071] (2) 引物选择区域定位

[0072] 将361条序列比对到14号染色体“hs\_ref\_GRCh38.p12\_chr14.fa”上,提取序列起始位点前150bp的碱基序列组;

[0073] (3) 候选引物库设计

[0074] 按照每20bp形成69430条引物起始数据集,形成候选引物库。

[0075] (4) cluster聚类分析

[0076] 将候选引物库用Cd-hit软件分类,设置-c参数为0.9,进行聚类分析;将cluster根据大小排序,选择前90的聚类,提取每个cluster的代表序列,根据引物设计原则,筛选得到初始引物组;

[0077] 引物原始数据组cluster聚类分析,”>”后面的数字为候选引物库里面的引物序列编号,%表示候选引物序列与代表序列的相似度,聚类的部分结果如下表1:

[0078] 表1

[0079] 

>Cluster 1
------------

[0080]

0	20aa, >3... *
1:	20aa, >4... at 95.00%
2 :	20aa, >132... at 95.00%
3 :	20aa, >133... at 95.00%
4:	20aa, >263... at 100.00%
5:	20aa, >264... at 95.00%
6:	20aa, >394... at 95.00%
7 :	20aa, >525... at 95.00%
8 :	20aa, >657... at 95.00%
9 :	20aa, >658... at 95.00%
10 :	20aa, >919... at 95.00%
11:	20aa, >920... at 95.00%
12 :	20aa, >1049... at 95.00%
13 :	20aa, >1182... at 95.00%
14:	20aa, >1705... at 95.00%
15 :	20aa, >1706... at 100.00%
16:	20aa, >1707... at 95.00%
17 :	20aa, >2228... at 100.00%
18:	20aa, >2229... at 95.00%
19 :	20aa, >2359... at 100.00%
20 :	20aa, >2360... at 95.00%
21:	20aa, >3147... at 100.00%
22 :	20aa, >3148... at 95.00%
23:	20aa, >4198... at 95.00%
24 :	20aa, >4199... at 100.00%
25:	20aa, >4200... at 95.00%
26 :	20aa, >4324... at 100.00%
27 :	20aa, >4325... at 95.00%

[0081]	28 : 20aa, >4455... at 95.00%
	29: 20aa, >4586... at 100.00%
	30: 20aa, >4587... at 95.00%
	31: 20aa, >4717... at 100.00%
	32: 20aa, >4718... at 95.00%
	33 : 20aa, >6682... at 95.00%
	34 : 20aa, >7730... at 95.00%
	35: 20aa, >7864... at 95.00%

[0082] 由表1可知, cluster1中的序列和代表序列的相似度, 即通过cd-hit筛选后, cluster中的95%以上的集合中前35个序列。

[0083] (5) 将初始引物组与人的抗体数据库比对, 得到每条序列在抗体库的覆盖度, 将该值按照大小排序, 选取排名前20, 分析GC含量、退火温度等, 最终选择16条作为最终重链引物组进行下一步实验验证。

[0084] 选择的16条的引物序列 (SEQ NO.1-SEQ NO.16) 即筛选出的人B细胞免疫组库引物组, 详细序列如下 (5' -3') :

[0085] SEQ NO.1: ATGGACATACTTTGTTCCA

[0086] SEQ NO.2: CCATGGAGTTTGGGCTGAGC

[0087] SEQ NO.3: GGCTGAGCTGGGTTTTCCTT

[0088] SEQ NO.4: CTGAGCTGGGTTTTCCTTGT

[0089] SEQ NO.5: CTCCTGGTGGCAGCTCCCAG

[0090] SEQ NO.6: CCTCCTCCTGGTGGCAGCTC

[0091] SEQ NO.7: CAGCTCCCAGATGTGAGTGT

[0092] SEQ NO.8: GCTGGGTTTTCCTTGTTGCT

[0093] SEQ NO.9: ATGAAACACCTGTGGTTCTT

[0094] SEQ NO.10: CCTGGAGGATCCTCTTCTTG

[0095] SEQ NO.11: GGTTTTCTTGTTGCTATTT

[0096] SEQ NO.12: TGGTGGCAGCTCCCAGATGT

[0097] SEQ NO.13: GGACGTGAGTGAGAGAAACA

[0098] SEQ NO.14: TCCTCACCATGGACTGGACC

[0099] SEQ NO.15: CTTGTTGGTATTTTAAAAGG

[0100] SEQ NO.16: GAGGATCCTCTTCTTGTTGG

[0101] 后续引物组性能评估时以SEQ NO.1-SEQ NO.16作为上游引物扩增免疫组库; 以SEQ NO.17作为下游引物扩增免疫组库, 下游引物序列SEQ NO.17为 (5' -3') :

[0102] SEQ NO.17: GGGGAAGACCGATGGGCCCTTGTTGG

[0103] 实施例2引物组性能评估

[0104] (1) 样本制备

[0105] 使用人淋巴细胞分离液分离外周血B淋巴细胞,提取RNA,逆转录得cDNA,步骤如下:

[0106] 1) 收集的新鲜外周血样(5个)各10毫升(mL),按LymphoPrep试剂盒说明书操作,获得相对较纯的外周血单核细胞(PBMC);

[0107] 2) 提取RNA,并用Nanodrop2000测定RNA的浓度及纯度,使用TIANScriptM-MLV逆转录酶试剂盒(货号ER104),逆转录后得cDNA(即样本:ST-1、ST-2、ST-3、ST-4、ST-5)为扩增模板,备用;

[0108] 逆转录反应步骤如下:

[0109] a) 体系配置:

[0110] 按照表2配置逆转录反应体系。

[0111] 表2

组分	用量
oligo (dT) <sub>12-18</sub>	2 $\mu$ L
总RNA	1 $\mu$ g
dNTP	2 $\mu$ L
ddH <sub>2</sub> O	用ddH <sub>2</sub> O将体系补足至15 $\mu$ L

[0113] b) 65 $^{\circ}$ C加热5min后迅速在冰上冷却2min,简短离心收集反应液后加入4 $\mu$ L5 $\times$ First-Strand Buffer及RNasin(40U/ $\mu$ L);

[0114] c) 加入1 $\mu$ L(200U)TIANScript M-MLV并轻轻用移液器混匀;

[0115] d) 42 $^{\circ}$ C温浴1h10min;

[0116] e) 85 $^{\circ}$ C加热5min终止反应置于冰上用于后续实验。

[0117] (2) 多重PCR扩增

[0118] 1) 体系配置

[0119] 加入上游引物以及下游引物,配置多重PCR体系,其中,该体系中的上游引物组(SEQ ID NO:1-SEQ ID NO:16)的等摩尔混合物,引物总浓度是20 $\mu$ M;下游引物(SEQ NO.17)总浓度是20 $\mu$ M;cDNA(即样本:ST-1、ST-2、ST-3、ST-4、ST-5)为扩增模板,多重PCR反应体系按照表3配置;

[0120] 表3

组分	体积
10 $\times$ PCR缓冲液	5 $\mu$ L
上游引物组	2 $\mu$ L
下游引物	2 $\mu$ L
dNTP	1 $\mu$ L
Taq酶	0.5 $\mu$ L
cDNA	5 $\mu$ L
ddH <sub>2</sub> O	34.5 $\mu$ L
总计	50 $\mu$ L

[0122] 2) PCR反应

[0123] 按多重PCR的条件设置PCR仪器程序,进行多重PCR:PCR结束后,4℃保存PCR产物并电泳检测,挑选片段长度约为500bp的片段,割胶回收,得到纯化后的抗体DNA片段,胶回收步骤采用QIAGEN公司QIAquick胶纯化试剂盒,按常规实验操作进行;Nanodrop 2000测试DNA浓度,并进行高通量测序。

[0124] 多重PCR条件如下表4:

[0125] 表4

	温度	时间	循环数
	95℃	15min	1个循环
[0126]	94℃	30s	35个循环
	60℃	30s	
	72℃	45s	
	72℃	10min	1个循环

[0127] PCR扩增后经凝胶电泳得到人B细胞免疫组库PCR扩增电泳图见图(1),由图(1)可知,在500bp附近有明显的条带,且本发明引物组1(即SEQ ID NO:1-SEQ ID NO:16)与引物组2(文献1(1.High-throughput isolation of immunoglobulin genes from single human B cells and expression as monoclonal antibodies.Liao HX1,Levesque MC, Nagel A,et al.J Virol Methods.2009;158(1-2):171-9.)公开引物)扩增出的条带范围基本一致。可以初步判断,本发明设计的引物组(SEQ ID NO:1-SEQ ID NO:16)可以扩增出人B细胞免疫组库的重链序列。

[0128] (3) 引物利用率分析

[0129] 采用生物信息学bowtie软件,分析引物(SEQ ID NO:1-SEQ ID NO:16)与5个样本(ST-1、ST-2、ST-3、ST-4、ST-5)的序列匹配情况,以5' RACE方法为对照(即该方法扩增出的抗体序列为基准),分析得到的引物的利用率如下表5所示;

[0130] 表5

	样本				
	ST-1	ST-2	ST-3	ST-4	ST-5
[0131] 0 错配 (%)	95	85	95	95	90
1 错配 (%)	100	90	100	100	95
2 错配 (%)	100	90	100	100	100

[0132] 上表中:

[0133] 0错配表示不允许引物序列和抗体序列发生错配;

[0134] 1错配表示允许引物序列的1个碱基和抗体序列发生错配;

[0135] 2错配表示允许引物序列的2个碱基和抗体序列发生错配。

[0136] 上表结果显示:设计的引物组(SEQ ID NO:1-SEQ ID NO:16)与样本具有很好的匹配。

[0137] (4) 覆盖度分析

[0138] 以样本ST-1、ST-2、ST-3、ST-4、ST-5为扩增模板,然后分别用本发明设计的引物

(SEQ NO.1-SEQ NO.16,即表中引物组1)和文献1中的引物(表中引物组2)进行PCR扩增,以5' RACE方法为对照(即该方法扩增出的抗体序列为基准),分析本发明引物扩增出的抗体序列的覆盖度,覆盖度分析结果如下表6:

[0139] 表6

样本名	5'RACE法 抗体总数	覆盖度 (%)	
		引物组 1	引物组 2
ST-1	153089	94.84	88.90
ST-2	70976	95.35	91.14
ST-3	106381	94.95	88.84
ST-4	150000	95.04	89.13
ST-5	150000	95.12	89.29

[0141] 覆盖度等于“PCR法”除以“5' RACE法”再乘以百分之百。

[0142] 其中:“PCR法”指多重PCR法扩增出的与5' RACE法扩增出的抗体序列相一致的条数;“5' RACE法”指采用5' RACE法扩增出的抗体序列的总条数。

[0143] 上表中,对测试的5个样本,引物组1扩增出的的抗体序列的平均覆盖度在95.01%以上,比文献中引物组2的覆盖平均为89.35%,本发明设计引物组1比文献中公开的引物的覆盖度平均高出5.66%,本发明设计的引物组具有很高的覆盖度。

[0144] 实施例3免疫组库生物信息分析

[0145] 用设计的引物组(SEQ ID NO:1-SEQ ID NO:16)为上游引物,SEQ ID NO:17为下游引物,扩增出的人B细胞免疫组库,进行后续生物信息学分析,分析结果如下:

[0146] (1) CDR3区域长度分析

[0147] CDR3氨基酸序列中,1个氨基酸(aa)的变化可能导致受体构象的改变,因此,CDR3氨基酸序列长度的变化可以反映出CDR3基因连接区的多样性,通过Igbblast软件分析免疫组库的序列信息后,得到免疫组库的CDR3长度分布情况(见图(2)),和CDR3区域中种系基因(germline)使用情况(见图(3))。

[0148] 由图2可以发现,CDR3的氨基酸长度主要集中在为9-12个氨基酸之间。

[0149] (2) CDR3区域中种系基因(germline)使用情况

[0150] 图3为IGVH的CDR3的Germline分布和使用分布;横坐标为不同的Germline,纵坐标为某种Germline的数量,结果见图(3);

[0151] 由图(3)可知,不同Germline的丰度非常不一致,有些片段丰度显著高于其他片段,该样本的VH的Germline主要集中在IGVH3-7\*01。

[0152] (3) CDR3区域中氨基酸的种类,以及氨基酸序列出现的频率和长度情况见表7所示;

[0153] 表7

	氨基酸序列	频率	长度
[0154]	AVDSNYQLI	0.001229	9
	AVNSGNTPLV	0.001158	10
	AVNRDDKII	0.000982	9
	AVEYGNKLV	0.000957	9
	AATDSWGKLQ	0.000952	10
	AVTSGSRLT	0.000944	9
	[0155]	SRTL	0.000924
AVQGGSEKLV		0.000904	10
AVTDSWGKLQ		0.000857	10
AVNTGNQFY		0.000814	9
VVNNARLM		0.000805	8
AVTGGGNKLT		0.000784	10
AANTNAGKST		0.000778	10
AVPTGTASKLT		0.000769	11
LVGDTGRRALT		0.000738	11

[0156] 由表7可知,CDR3的序列变化很大,有些片段丰度如“AVDSNYQLI”等高于其他片段,可以根据CDR3的丰度,判断免疫组库中的特征抗体,为后续研究和治疗等提供思路。

[0157] 综上所述,本发明提供一种引物组的设计方法,根据扩增长度的需要,巧妙地在抗体FR1起始位点的上游,提取FR1起始位点前的序列,并从5'端的第一个碱基开始,依次移位切取固定长度的引物序列,形成候选引物库,聚类分析后筛选得到引物组,采用该方法得到的引物组能够显著提高覆盖度和实验效率,降低错配率,方法简单,节省成本。

[0158] 申请人声明,本发明通过上述实施例来说明本发明的详细方法,但本发明并不局限于上述详细方法,即不意味着本发明必须依赖上述详细方法才能实施。所属技术领域的技术人员应该明了,对本发明的任何改进,对本发明产品各原料的等效替换及辅助成分的添加、具体方式的选择等,均落在本发明的保护范围和公开范围之内。

		序列表	
[0001]			
[0002]	<110>	苏州泓迅生物科技股份有限公司	
[0003]	<120>	一种引物组的设计方法及其应用	
[0004]	<130>	2018年	
[0005]	<141>	2018-11-06	
[0006]	<160>	17	
[0007]	<170>	SIP0SequenceListing 1.0	
[0008]	<210>	1	
[0009]	<211>	19	
[0010]	<212>	DNA	
[0011]	<213>	人工合成()	
[0012]	<400>	1	
[0013]		atggacatac tttgttcca	19
[0014]	<210>	2	
[0015]	<211>	20	
[0016]	<212>	DNA	
[0017]	<213>	人工合成()	
[0018]	<400>	2	
[0019]		ccatggagtt tgggctgagc	20
[0020]	<210>	3	
[0021]	<211>	20	
[0022]	<212>	DNA	
[0023]	<213>	人工合成()	
[0024]	<400>	3	
[0025]		ggctgagctg ggttttcctt	20
[0026]	<210>	4	
[0027]	<211>	20	
[0028]	<212>	DNA	
[0029]	<213>	人工合成()	
[0030]	<400>	4	
[0031]		ctgagctggg ttttccttgt	20
[0032]	<210>	5	
[0033]	<211>	20	
[0034]	<212>	DNA	
[0035]	<213>	人工合成()	
[0036]	<400>	5	
[0037]		ctcctggtgg cagctcccag	20
[0038]	<210>	6	

[0039]	<211>	20	
[0040]	<212>	DNA	
[0041]	<213>	人工合成()	
[0042]	<400>	6	
[0043]	cctcctcctg	gtggcagctc	20
[0044]	<210>	7	
[0045]	<211>	20	
[0046]	<212>	DNA	
[0047]	<213>	人工合成()	
[0048]	<400>	7	
[0049]	cagctcccag	atgtgagtgt	20
[0050]	<210>	8	
[0051]	<211>	20	
[0052]	<212>	DNA	
[0053]	<213>	人工合成()	
[0054]	<400>	8	
[0055]	gctgggtttt	ccttgttgct	20
[0056]	<210>	9	
[0057]	<211>	20	
[0058]	<212>	DNA	
[0059]	<213>	人工合成()	
[0060]	<400>	9	
[0061]	atgaaacacc	tgtggttctt	20
[0062]	<210>	10	
[0063]	<211>	20	
[0064]	<212>	DNA	
[0065]	<213>	人工合成()	
[0066]	<400>	10	
[0067]	cctggaggat	cctcttcttg	20
[0068]	<210>	11	
[0069]	<211>	20	
[0070]	<212>	DNA	
[0071]	<213>	人工合成()	
[0072]	<400>	11	
[0073]	ggttttcctt	gttgctattt	20
[0074]	<210>	12	
[0075]	<211>	20	
[0076]	<212>	DNA	
[0077]	<213>	人工合成()	

[0078]	<400> 12	
[0079]	tggtggcagc tcccagatgt	20
[0080]	<210> 13	
[0081]	<211> 20	
[0082]	<212> DNA	
[0083]	<213> 人工合成()	
[0084]	<400> 13	
[0085]	ggacgtgagt gagagaaaca	20
[0086]	<210> 14	
[0087]	<211> 20	
[0088]	<212> DNA	
[0089]	<213> 人工合成()	
[0090]	<400> 14	
[0091]	tcctcaccat ggactggacc	20
[0092]	<210> 15	
[0093]	<211> 20	
[0094]	<212> DNA	
[0095]	<213> 人工合成()	
[0096]	<400> 15	
[0097]	cttgttgta ttttaaagg	20
[0098]	<210> 16	
[0099]	<211> 20	
[0100]	<212> DNA	
[0101]	<213> 人工合成()	
[0102]	<400> 16	
[0103]	gaggatcctc ttcttggtgg	20
[0104]	<210> 17	
[0105]	<211> 26	
[0106]	<212> DNA	
[0107]	<213> 人工合成()	
[0108]	<400> 17	
[0109]	ggggaagacc gatgggcct tggtgg	26

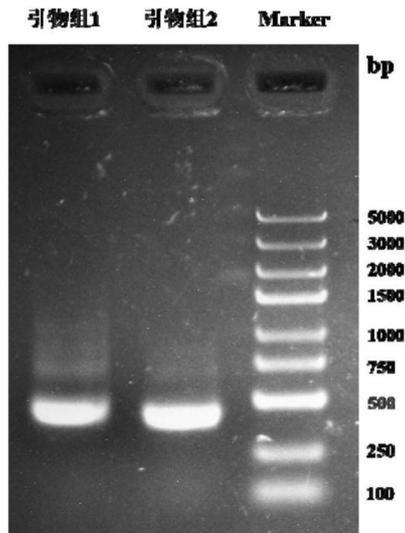


图1

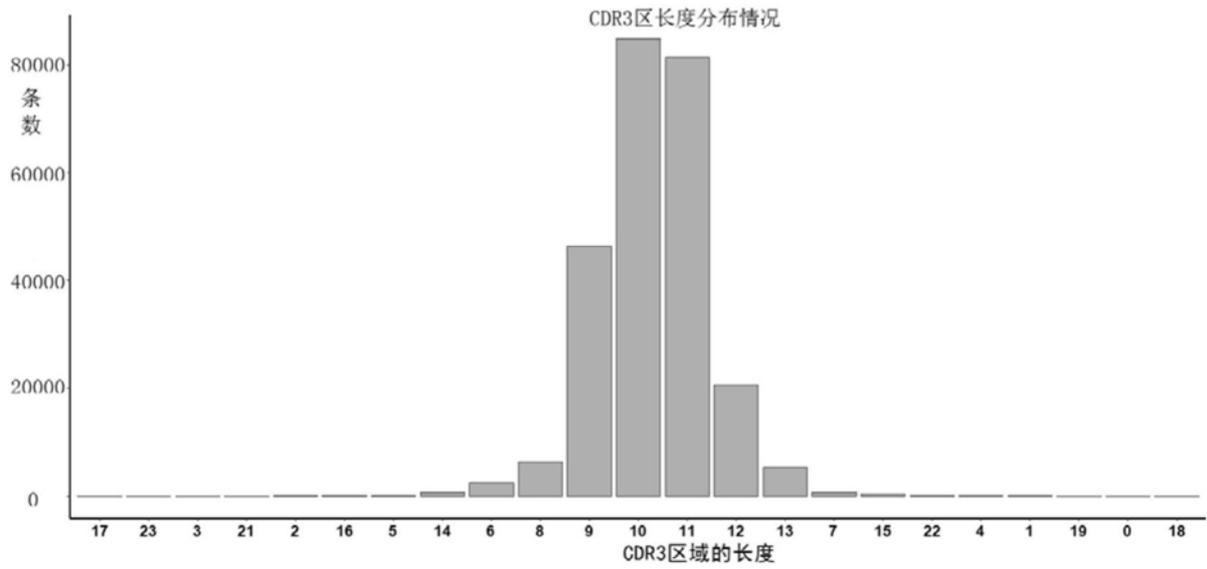


图2

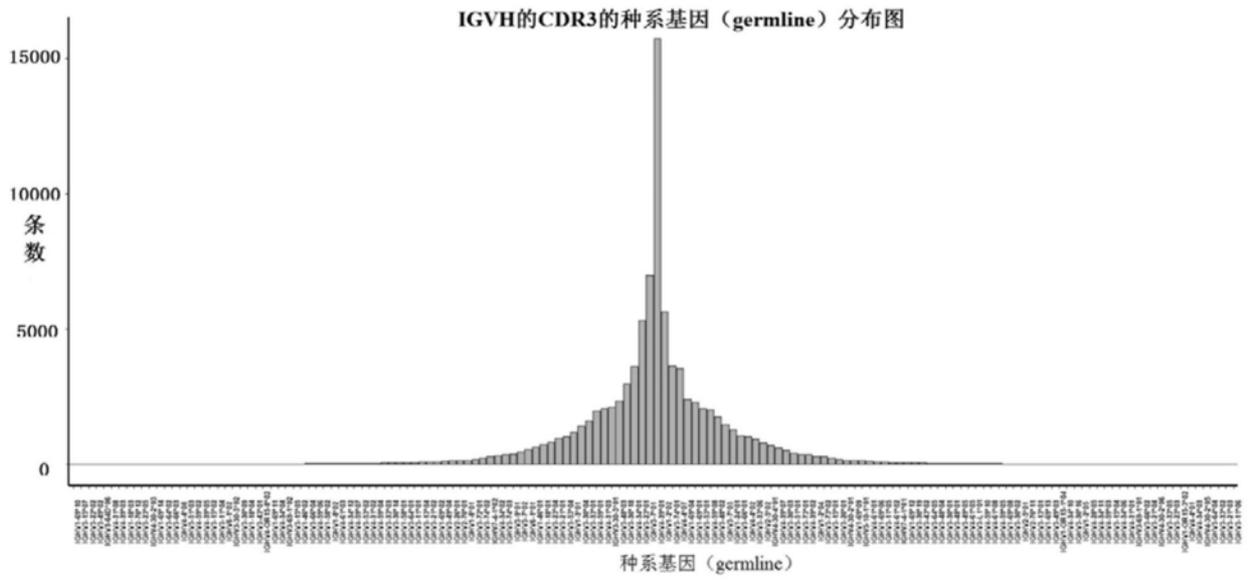


图3