



(12)发明专利申请

(10)申请公布号 CN 107828876 A

(43)申请公布日 2018.03.23

(21)申请号 201710979254.6

(22)申请日 2017.10.19

(71)申请人 中国科学院生物物理研究所
地址 100101 北京市朝阳区大屯路15号

(72)发明人 俞洋 顾嘉琦

(74)专利代理机构 北京纪凯知识产权代理有限公司 11245

代理人 关畅 张立娜

(51)Int.Cl.

C12Q 1/6869(2018.01)

C12Q 1/68(2018.01)

G01N 33/68(2006.01)

G01N 33/53(2006.01)

权利要求书2页 说明书12页
序列表6页 附图4页

(54)发明名称

可共价结合底物的标签蛋白在CLIP中的应用

(57)摘要

本发明公开了一种可共价结合底物的标签蛋白在CLIP中的应用。本发明提供了使用可共价结合底物的标签融合蛋白通过变性剂纯化方式在CLIP技术中的应用。本发明使用可共价结合底物的标签的融合蛋白通过蛋白变性试剂洗涤的方式得到可以用于测序的RNA文库,结果证明可以很好地重复以前的文献发表的实验结果。在操作上省略了传统CLIP步骤中需要通过转膜并且切胶回收片段的步骤。该方法通过可共价结合底物的标可和固定于介质上的特异性结合物形成共价键的特性,将传统上不能在磁珠上直接进行的剧烈洗涤方法加入到实验操作步骤中。这样在操作上比较简便,节省了操作时间,优化了CLIP步骤,减少了操作过程中可能会造成的损失。

1. 可共价结合底物的标签蛋白在紫外交联免疫共沉淀技术中的应用。
2. 根据权利要求1所述的应用,其特征在於:所述应用中,所述可共价结合底物的标签蛋白作为紫外交联免疫共沉淀技术中靶蛋白的融合标签使用。
3. 根据权利要求1或2所述的应用,其特征在於:所述应用为用于研究RNA与RNA结合蛋白的相互作用,紫外交联免疫共沉淀技术中用非跑胶非转膜的方式纯化RNA-蛋白复合物。
4. 根据权利要求3所述的应用,其特征在於:将由所述可共价结合底物的标签蛋白和所述靶蛋白形成的融合蛋白通过共价键与带有介质的所述可共价结合底物的标签蛋白的特异性结合物相连,然后通过蛋白质变性洗涤的方式纯化所述RNA-蛋白复合物。
5. 一种使用可共价结合底物的标签蛋白并通过非胶纯化方式进行紫外交联免疫共沉淀获取RNA结合蛋白的靶标RNA的方法,包括如下步骤:
 - (1) 使受体细胞表达融合蛋白,所述融合蛋白是由连接肽将可共价结合底物的标签蛋白和靶蛋白连接而形成的融合蛋白;
 - (2) 对步骤(1)中表达了所述融合蛋白的受体细胞进行紫外照射,使RNA-蛋白复合物中的RNA与蛋白之间形成共价键而交联,收集细胞样本;
 - (3) 对步骤(2)所收集的细胞样本进行裂解,然后通过连接有所述可共价结合底物的标签蛋白的特异性结合物的介质进行免疫沉淀,获得含有“靶标RNA-融合蛋白-介质”的沉淀样本;
 - (4) 对步骤(3)所得的沉淀样本进行非变性洗涤,然后再进行去磷酸化处理;
 - (5) 对步骤(4)处理后的样本进行变性洗涤;
 - (6) 对步骤(5)处理后的样本用能够特异性识别所述连接肽的蛋白酶进行酶切处理,从而将由所述靶蛋白和所述靶标RNA形成的RNA-蛋白复合物从所述介质上分离下来;
 - (7) 将步骤(6)所得的由所述靶蛋白和所述靶标RNA形成的RNA-蛋白复合物中的所述靶蛋白去除,从而获得所述靶标RNA。
6. 根据权利要求5所述的方法,其特征在於:步骤(4)中,所述非变性洗涤为:用含有0.1% Triton X-100和500mM NaCl的PBS缓冲液洗涤2次;用含有0.1% Triton X-100的PBS缓冲液洗涤3次;其中%表示体积百分含量;和/或
步骤(5)中,所述变性洗涤为两次变性洗涤;
具体的,所述两次变性洗涤中的第一次变性洗涤为:用Trizol试剂去除碱性磷酸酶;8M盐酸胍洗涤2次;8M尿素洗涤2次;所述两次变性洗涤中的第二次变性洗涤为:用SDS洗涤液洗涤3次;用8M尿素洗涤2次。
7. 根据权利要求5或6所述的方法,其特征在於:所述连接肽和所述“能够特异性识别所述连接肽的蛋白酶”为如下任一:
 - (a) 所述连接肽为Tev酶的特异性识别序列;所述“能够特异性识别所述连接肽的蛋白酶”为Tev酶;
 - (b) 所述连接肽为肠激酶的特异性识别序列;所述“能够特异性识别所述连接肽的蛋白酶”为肠激酶;
 - (c) 所述连接肽为Xa因子的特异性识别序列;所述“能够特异性识别所述连接肽的蛋白酶”为Xa因子;
 - (d) 所述连接肽为人鼻病毒14型的3C蛋白酶的特异性识别序列;所述“能够特异性识别

所述连接肽的蛋白酶”为人鼻病毒14型的3C蛋白酶；

(e) 所述连接肽为凝血酶的特异性识别序列；所述“能够特异性识别所述连接肽的蛋白酶”为凝血酶。

8. 权利要求5-7中任一所述方法在研究蛋白与RNA互作中的应用。

9. 一种鉴定RNA结合蛋白的靶标RNA的方法，包括如下步骤：利用权利要求5-8中任一所述方法获得靶标RNA，然后构建cDNA文库，再对所述cDNA文库进行高通量测序，最后根据测序结果进行基因定位。

10. 根据权利要求1-9中任一所述的应用或方法，其特征在于：所述可共价结合底物的标签蛋白为如下任一：Halo标签蛋白、CLIP标签蛋白、SNAP标签蛋白和Spy标签蛋白；

和/或

所述可共价结合底物的标签蛋白的特异性结合物为如下任一：

(A) 当所述可共价结合底物的标签蛋白为Halo标签蛋白时，所述可共价结合底物的标签蛋白的特异性结合物为含有卤素的配体；

(B) 当所述可共价结合底物的标签蛋白为CLIP标签蛋白时，所述可共价结合底物的标签蛋白的特异性结合物为苄基胞嘧啶衍生物；

(C) 当所述可共价结合底物的标签蛋白为SNAP标签蛋白时，所述可共价结合底物的标签蛋白的特异性结合物为苄基鸟嘌呤；

(D) 当所述可共价结合底物的标签蛋白为Spy标签蛋白时，所述可共价结合底物的标签蛋白的特异性结合物为SpyCatcher蛋白；

和/或

所述介质为磁珠。

可共价结合底物的标签蛋白在CLIP中的应用

技术领域

[0001] 本发明属于RNA生物学以及分子生物学领域,涉及一种可共价结合底物的标签蛋白在进行CLIP(cross-linking immunoprecipitation,紫外交联免疫共沉淀)实验中的应用,具体涉及CLIP技术中,使用带有可共价结合底物的标签的融合蛋白,通过变性试剂洗涤的方式纯化蛋白-RNA共价交联复合物,从而省略了SDS-PAGE(SDS-聚丙烯酰胺)电泳、醋酸纤维素转膜或同位素标记等步骤,最终得到可以用于测序的cDNA文库。

背景技术

[0002] 研究RNA结合蛋白和其结合的RNA的分子生物学方法,简而言之,就是通过纯化RNA和RNA结合蛋白的复合物。纯化复合物的方法既可以是沉淀RNA,也可以是沉淀RNA结合蛋白。由于沉淀RNA的技术有其局限性,因此科研人员更多的通过沉淀RNA结合蛋白的方法,获得RNA和RNA结合蛋白质的复合物。最初,RNA结合蛋白免疫共沉淀(RIP)技术被科研人员广泛运用到研究RNA和蛋白质的互作领域。

[0003] RIP技术首先通过针对靶蛋白的抗体把组织或细胞内的相应的RNA和RNA结合蛋白的复合物沉淀下来。然后经过纯化和分离得到RNA,并通过下游的分析技术得到RNA的信息。RIP技术与ChIP技术的总体思路类似,但是由于研究的对象不同。其相关的步骤有很多变化。而且RNA在体外容易降解,所以对于RNA的操作和实验使用的试剂有更高的要求。RIP技术与microarray技术相结合,就成了RIP-Chip技术;与高通量测序技术相结合就成了RIP-seq技术。RIP技术由于对蛋白和核酸复合物的洗涤比较温和,所以有一定的局限性。在RIP的数据中假阳性的结果比较多,而且也无法得知RNA和蛋白的具体结合位点信息【参见Jayaseelan,S.,F.Doyle,and S.A.Tenenbaum,Profiling post-transcriptionally networked mRNA subsets using RIP-Chip and RIP-Seq.Methods,2014.67(1):p.13-9.】。

[0004] 2003年,Darnell实验室最先发表了通过紫外交联的方法来研究RNA和蛋白质互作的相关文章。最初的CLIP实验步骤结合了RIP和ChIP的实验步骤。在RNA和RNA蛋白复合物进行免疫沉淀之前,运用了研究DNA和蛋白互作的方法,即紫外交联。由于当时技术的局限性,他们无法运用高通量测序技术,因此获得的数据有限。即使如此,他们运用CLIP技术在小鼠的脑组织中鉴定出了神经特异的RNA结合蛋白和剪切因子NOVA1和NOVA2的RNA结合位点【参见Ule,J.,et al.,CLIP identifies Nova-regulated RNA networks in the brain.Science,2003.302(5648):p.1212-5.】。CLIP所得到的结论通过在小鼠脑组织中敲除相关基因而得到验证。2005年,他们采用高通量测序技术,又对CLIP获得的文库进行了深度测序,并根据测序的结果初步绘制了Nova蛋白的全基因组蛋白和RNA的互作网络。他们将CLIP和高通量测序相结合的技术称为HITS-CLIP技术【参见Ule,J.,et al.,Nova regulates brain-specific splicing to shape the synapse.Nat Genet,2005.37(8):p.844-52.】。同一时期,又有课题组运用CLIP和高通量测序技术相结合研究了RbFox2蛋白的全基因组RNA与蛋白互作网络,并且将其命名为CLIP-seq【参见Yeo,G.W.,et al.,An RNA

code for the FOX2splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat Struct Mol Biol*, 2009.16 (2) :p.130-7.】。从此以后高通量测序技术越来越多地被运用到RNA和蛋白全基因组互作网络的研究。2009年, Darnell实验室对Ago2蛋白结合的miRNA-mRNA的互作通过CLIP-seq技术进行了解析, 获得了miRNA-mRNA的之间互作的全基因组信息【参见Chi, S.W., et al., Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*, 2009.460 (7254) :p.479-86.】。同年, Xue, Y等人通过CLIP技术对PTB蛋白在核内对mRNA前体进行可变剪切的机制进行了研究【参见Xue, Y., et al., Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. *Mol Cell*, 2009.36 (6) :p.996-1006.】。

【0005】 CLIP技术相比RIP技术有了一些改进。首先, 无论是细胞样品, 还是组织样品, 都需要通过紫外线照射处理。通过紫外照射, RNA和蛋白质之间会形成共价键。打开共价键需要的能量很高, 因此蛋白和RNA的复合体可以经受住强烈的纯化过程。因此, 紫外线的使用可以提高测序得到数据的信噪比。同时, 与ChIP中使用的福尔马林(甲醛)交联相比, UV交联不会引起大分子和多分子之间化学桥键的形成, 干扰信噪比。UV交联只在几个埃的距离内进行交联, 即只会把蛋白质和互作核酸之间进行交联, 形成共价键。其次, 在CLIP的操作步骤中引入了RNA酶的消化。因为没有被蛋白保护的RNA会被或者更容易被RNA酶进行消化, 温和的RNA酶消化反应可以将没有结合蛋白的RNA部分给降解掉, 只保留蛋白结合的RNA部分。截短后的RNA更容易被进行转录和测序以及进行结构域分析, 另外也避免了沉淀不必要的蛋白质RNA蛋白复合体。正因为如此, 后来的RIP操作步骤中也加入了这一步核酸酶的处理。已有文献报道的RNA消化酶有RNase A, RNase T1, RNase I等。他们的应用各有好处和缺点。虽然复合物中RNA被截短了, 但是因为在后续的纯化过程中没有去除RNA酶的步骤, 所以对结果会有一定的干扰。最后, CLIP的样品可以经受住剧烈条件下的洗涤。例如PBST的多次洗涤, SDS-PAGE胶纯化, 转膜到醋酸纤维素薄膜上等。这些步骤都可以去除非特异性地结合在抗体上的蛋白质, 同时可以减少没有结合和没有交联的RNA, 减少数据的背景。

【0006】 CLIP技术是基于ChIP和RIP技术发展而来。传统CLIP的技术如下: 在细胞培养之后, 通过紫外光照射使蛋白和RNA进行交联。裂解细胞后, 对复合物进行RNA酶处理, 使蛋白质没有保护的RNA部分进行降解, 并将RNA的长度修剪到趋于一致。接着通过特异性的抗体在体外免疫沉淀蛋白质, 将RNA和蛋白质的复合物沉淀下来。温和洗涤后, 在RNA的3'末端加上RNA接头。在RNA的5'末端标记上同位素³²P。然后将蛋白质和RNA复合物变性后跑SDS-PAGE胶, 并且转膜到硝酸纤维素薄膜上。因为硝酸纤维素薄膜的特点是只能结合蛋白质, 而不能结合RNA。转膜后通过显影, 将带有同位素标记的蛋白和RNA复合物的条带切割下来, 然后提取RNA和蛋白质复合物。接着用蛋白酶K酶解蛋白质, 纯化RNA。然后在RNA的5'端加上接头序列。通过接头序列将RNA反转录成cDNA。再通过PCR技术将cDNA进行扩增。所获得的cDNA文库, 就可以进行高通量测序, 获取数据并分析【参见Ule, J., et al., CLIP identifies Nova-regulated RNA networks in the brain. *Science*, 2003.302 (5648) :p.1212-5.】。

【0007】 在CLIP技术和高通量测序技术结合运用之后, 很多新的CLIP技术得到发展。PAR-CLIP (Photoactivatable Ribonucleoside-enhanced CLIP) 是广泛得到运用的技术之一【参见Hafner, M., et al., Transcriptome-wide identification of RNA-binding

protein and microRNA target sites by PAR-CLIP. *Cell*, 2010. 141 (1) : p. 129-41.】。PAR-CLIP技术改变了细胞的培养技术。在细胞培养的时候加入了4SU,使得新生的转录本或者体内稳态的RNA的U位点被4SU替代。交联使用的紫外线的波长也从254nm变为365nm。在理论上提高了交联的效率。同时使用了不同的核酸酶对RNA的长度进行修剪。最后得到的cDNA文库中4SU的位置会发生C→T的变换,在生信处理过程中可以被识别。PAR-CLIP技术通过为细胞培养的过程中外源性地加入4-巯基尿嘧啶核苷(4-thiouridine, 4SU)或者6-巯基鸟嘌呤核苷(6-thioguanosine, 6SG),使得U或者G得到标记,同时选择性地运用长波长(一般是UV365nm)进行紫外照射。这样可以提高蛋白质和核酸的交联效率。4SU是更为常用的一个试剂,因为他可以提高UV交联的效率,而且体内核酸中尿嘧啶和蛋白的结合频率极高。RNA中的U被4SU替换后会导致在反转录的时候T变成C的转换。这个转换经常起到交联位点的指示作用,也常被用作减去非交联背景的信号。但是PAR-CLIP中运用的4SU也有很多限制。首先,4SU只能在细胞培养中使用,在组织、动物中或者临床标本中还没有被报道过。其次,4SU对细胞是有毒性的。在培养基中加入4SU之后,细胞还需要培养24到48小时。在这期间,4SU对细胞的代谢可能会有影响,导致细胞的凋亡。最后,虽然有报道说在加了4SU之后得到的mRNA文库并没有显著改变,但是其未知的影响还没有被发现,因此其使用还是需要慎重。

[0008] 在PAR-CLIP技术之后,出现的是iCLIP(individual-nucleotide resolution CLIP)技术【参见Konig, J., et al., iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol*, 2010. 17 (7) : p. 909-15.】。iCLIP技术指的是单个碱基分辨率的CLIP技术。传统的CLIP技术中的在RNA反转录成cDNA的同时,反转录酶在运行至交联位点的时候通常并不能顺利进行转录。因为RNA结合蛋白被降解后,其氨基酸残基和RNA分子相互交联所形成的共价键不容易被打开,反转录酶受到空间位阻后就有可能脱离模板,所以该位点在反转录时变成了反转录的障碍。虽然有这样的障碍存在,但是有时候反转录酶还是会绕过这个屏障进行反转录。这样得到的结果可能会是cDNA在这些位点发生突变或者缺失,但是更多的情况则是反转录的终止。虽然反转录并不完全,但是这些缺失和终止的片段可以提供交联位点的信息。然而,在传统的CLIP步骤中,PCR的扩增需要在cDNA的5'和3'端都加上接头序列。而cDNA的5'和3'接头序列来自于RNA模板的反转录。因此这样反转录得到的cDNA很容易缺失了5'的接头序列,因此会导致建库的失败。为了改进这个步骤,在iCLIP的操作步骤中接头直接加在了反转录的引物上。因此得到的cDNA经过环化和酶切就可以自然而然地在cDNA的两端都加上接头序列。这样可以避免因转录失败而导致的cDNA信息丢失。iCLIP技术改变了传统的建库方式。它通过反转录的cDNA大多数都会在交联位置停顿的特点,采用环化的方式。这样在切开后,测序的5'端的开始即是交联停顿的位置。iCLIP可用于单个核苷酸分辨率的数据分析。

[0009] 虽然iCLIP技术提高了cDNA建库的效率问题,减少了CLIP技术的操作时间和降低了失败可能性。但是由于其建库需要进行单链cDNA的环化,因此这一步中环化酶的效率是极为关键的。如果环化酶的效率低就会影响整个cDNA文库的组成,甚至会影响测序的成功率。为了改善这一点,2016年, Gene, W, Y实验室优化了建库的步骤,发明出了eCLIP(enhanced CLIP)技术【参见Van Nostrand, E.L., et al., Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP(eCLIP). *Nat Methods*, 2016. 13 (6) : p. 508-14.】。eCLIP技术没有改变整体CLIP的操作步骤,但是在建库

方面进行了改变。他们参考了iCLIP的建库步骤,但是没有在反转录后进行环化,而是直接在cDNA的3'末端添加上接头序列。然后再通过PCR的步骤扩增出cDNA文库来。他们指出eCLIP的建库方法增加了cDNA的构建效率,并降低了cDNA文库失败的可能性。

[0010] 另外,irCLIP (infrared-CLIP)技术的出现取代了传统在CLIP技术中同位素的使用。irCLIP技术使用了带有Biotin标记的3' RNA接头。这样就可以在荧光扫描仪上定位RNA的位置【参见Zarnegar, B. J., et al., irCLIP platform for efficient characterization of protein-RNA interactions. Nat Methods, 2016. 13 (6) : p. 489-92.】。

[0011] 虽然CLIP技术经过多次改进,但是操作步骤的过程中需要使用SDS-PAGE电泳和醋酸纤维素转膜。这两个步骤耗时长,操作要求较高,样本容易损失。因此改进CLIP技术的最大出发点在于减少总体操作的时间,减少操作步骤。在操作的过程中,如果整体的效率提高了,那么一开始使用的细胞数量也可以减少。其实,细胞或者组织用量的减少,可以使科研人员对一些很难获得的细胞或者组织进行CLIP实验。使在胚胎早期或者细胞分化周期中某一时期的细胞中靶蛋白和RNA相互作用的研究变成可能。

发明内容

[0012] 本发明的目的是提供一种可共价结合底物的标签蛋白的新用途。这里可共价结合底物的标签蛋白指的是标签蛋白可以与底物反应后形成共价键,结合后无法解离的标签蛋白。

[0013] 本发明所提供的可共价结合底物的标签蛋白的新用途,具体为可共价结合底物的标签蛋白在紫外交联免疫共沉淀技术中的应用。

[0014] 在所述应用中,所述可共价结合底物的标签蛋白作为紫外交联免疫共沉淀技术中靶蛋白的融合标签使用。

[0015] 进一步,在本发明中,所述应用为用于研究RNA与RNA结合蛋白(即所述靶蛋白)的相互作用,紫外交联免疫共沉淀技术中用非跑胶非转膜的方式纯化RNA-蛋白复合物。即与传统CLIP技术相比,本发明中RNA-蛋白复合物的纯化不通过SDS-聚丙烯酰胺凝胶电泳,不通过醋酸纤维素薄膜的转膜,以及不使用同位素。

[0016] 更进一步,本发明中是将由所述可共价结合底物的标签蛋白和所述靶蛋白(即所述RNA结合蛋白)形成的融合蛋白通过共价键与带有介质(如磁珠)的所述可共价结合底物的标签蛋白的特异性结合物相连,然后通过蛋白质变性洗涤的方式纯化所述RNA-蛋白复合物的。

[0017] 其中,所述蛋白质变性洗涤是指通过物理方式(如加热,振荡等)和/或化学方式(如盐酸胍,尿素,十二烷基硫酸钠,Trizol试剂等)破坏蛋白质分子中的氢键,使蛋白质变性后打开原有的高级结构。由于所述可共价结合底物的标签蛋白和固定于介质(如磁珠)上的所述可共价结合底物的标签蛋白的特异性结合物之间形成的是共价键,RNA和蛋白质交联也是形成共价键,因此洗涤过程中,由共价键连接在一起的复合物依旧偶联在介质(如磁珠)上并保持着序列的完整性,非共价结合的物质,如杂质,非特异结合核酸等,就会游离出复合物之外而被洗涤干净。

[0018] 相应的,本发明提供了一种使用可共价结合底物的标签蛋白并通过非胶纯化方式

进行紫外交联免疫共沉淀获取RNA结合蛋白的靶标RNA的方法。

[0019] 本发明所提供的可使用可共价结合底物的标签蛋白并通过非胶纯化方式进行紫外交联免疫共沉淀获取RNA结合蛋白的靶标RNA的方法,明确说来,其中RNA-蛋白复合物的纯化不通过SDS-聚丙烯酰胺凝胶电泳,不通过醋酸纤维素薄膜的转膜,以及不使用同位素。

[0020] 该方法具体可包括如下步骤:

[0021] (1) 使受体细胞表达融合蛋白,所述融合蛋白是由连接肽将可共价结合底物的标签蛋白和靶蛋白连接而形成的融合蛋白;

[0022] (2) 对步骤(1)中表达了所述融合蛋白的受体细胞进行紫外照射,使RNA-蛋白复合物中的RNA与蛋白之间形成共价键而交联,收集细胞样本;

[0023] (3) 对步骤(2)所收集的细胞样本进行裂解,然后通过连接有所述可共价结合底物的标签蛋白的特异性结合物的介质(如磁珠)进行免疫沉淀,获得含有“靶标RNA-融合蛋白-介质”的沉淀样本;

[0024] (4) 对步骤(3)所得的沉淀样本进行非变性洗涤,然后再进行去磷酸化处理(如用碱性磷酸酶进行处理);

[0025] (5) 对步骤(4)处理后的样本进行变性洗涤;

[0026] (6) 对步骤(5)处理后的样本用能够特异性识别所述连接肽的蛋白酶进行酶切处理,从而将由所述靶蛋白和所述靶标RNA形成的RNA-蛋白复合物从所述介质上分离下来;

[0027] (7) 将步骤(6)所得的由所述靶蛋白和所述靶标RNA形成的RNA-蛋白复合物中的所述靶蛋白去除(如通过蛋白酶消化处理去除所述靶蛋白),从而获得所述靶标RNA。

[0028] 进一步,步骤(1)中所述融合蛋白,其中所述可共价结合底物的标签蛋白可位于N端也可位于C端。

[0029] 进一步,所述步骤(1)为:将可共价结合底物的标签蛋白的编码基因与靶蛋白(即RNA结合蛋白)的编码基因通过连接肽的编码基因进行连接,得到融合基因;所述融合基因能够表达由所述连接肽将所述可共价结合底物的标签蛋白和所述靶蛋白连接而形成的融合蛋白;将所述融合基因导入受体细胞,得到重组细胞;培养所述重组细胞,使导入的所述融合基因表达出所述融合蛋白。

[0030] 在所述方法中,在步骤(3)中,对步骤(2)所收集的细胞样本进行裂解后还包括进行RNA酶(或者能够水解RNA的其他核酸酶)处理的步骤。

[0031] 在本发明中,步骤(4)中,所述非变性洗涤具体为:用含有0.1% Triton X-100和500mM NaCl的PBS缓冲液洗涤2次,用含有0.1% Triton X-100的PBS缓冲液洗涤3次;其中%表示体积百分含量。

[0032] 在本发明中,步骤(5)中,所述变性洗涤可为一次或多次变性洗涤。在本发明中,具体为两次变性洗涤。

[0033] 更加具体的,所述两次变性洗涤中的所述第一次变性洗涤具体为:用Trizol试剂(Invitrogen)去除碱性磷酸酶;8M盐酸胍洗涤2次(每次5分钟);8M尿素洗涤2次(每次5分钟)。所述第二次变性洗涤具体为:用SDS洗涤液(配方:10% SDS;50mM Tris-HCl pH 7.0;1mM EDTA;1mM DTT;余量为水;%表示质量体积百分含量,即10%表示10g/100ml)洗涤3次(每次在65°C条件下振荡5分钟);用8M尿素洗涤2次(每次室温洗涤5分钟)。

[0034] 在所述方法中,所述两次变性洗涤之间还可包括对第一次变性洗涤后的样本进行

RNA的3'末端加接头反应的步骤。当然这个步骤也可以在从RNA-蛋白复合物中成功分离出靶标RNA以后再进行。

[0035] 在所述方法中,所述“能够特异性识别所述连接肽的蛋白酶”可为Tev酶、肠激酶、Xa因子、人鼻病毒14型的3C蛋白酶(HRV 3C蛋白酶或PreScission蛋白酶)、凝血酶等,相应的,所述连接肽具体为一个或多个(如2-3个)这些酶的特异性识别序列。其中,Tev酶识别的连接肽,其具体序列为Glu-Asn-Leu-Tyr-Phe-Gln-Gly/Ser(“/”表示“或”)。肠激酶识别的连接肽,其具体序列为Asp-Asp-Asp-Asp-Lys。Xa因子识别的连接肽,其具体序列为Ile-Glu/Asp-Gly-Arg。人鼻病毒14型的3C蛋白酶(HRV 3C蛋白酶或PreScission蛋白酶)识别连接肽,其具体序列为Leu-Glu-Val-Leu-Phe-Gln-Gly-Pro。凝血酶识别的连接肽,其具体序列为A-B-Pro-Arg-X-Y(其中,A和B为疏水氨基酸,X和Y为非酸性氨基酸),常见的识别序列为Leu-Val-Pro-Arg-Gly-Ser。

[0036] 在本发明的一个实施例中,所述连接肽具体为3个串联的Tev酶的特异性识别序列。所述Tev酶的特异性识别序列具体为Glu-Asn-Leu-Tyr-Phe-Gln-Gly/Ser(“/”表示“或”)。更加具体的,所述连接肽为序列表中序列1的第2322-2396位所示DNA片段编码所得的氨基酸序列。相应的,所述“能够特异性识别所述连接肽的蛋白酶”具体为Tev酶。本发明研究表明Tev酶可以耐受2M尿素条件。

[0037] 前文所述方法在研究蛋白与RNA互动中的应用也属于本发明的保护范围。

[0038] 本发明还进一步保护一种鉴定RNA结合蛋白的靶标RNA的方法。

[0039] 本发明所保护的鉴定RNA结合蛋白的靶标RNA的方法,命名为GoldCLIP,具体可包括如下步骤:利用前文所述的方法获得靶标RNA,然后构建cDNA文库,再对所述cDNA文库进行高通量测序,最后根据测序结果进行基因定位。

[0040] 在本发明中,获得靶标RNA后是接着通过iCLIP文献【参见Konig,J.,et al.(2011).“iCLIP--transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution.”J Vis Exp(50).】中介绍的方法构建所述cDNA文库的。大致如下:RNA反转录并加上每个样本独立的标签序列之后,跑TEB-Urea变性胶纯化。反转录用到引物序列中不仅包含有独立的标签序列,还有5'端的接头序列和3'端的接头序列,以及接头序列之间的BamHI酶切位点。得到单链cDNA通过ssDNA环化酶环化,然后通过PCR仪将一条与引物互补配对的序列退火到引物序列上。接着通过BamHI酶切,得到产物具有5'端的接头序列,标签序列和3'端的接头序列。再通过PCR扩增后,将得到的cDNA进行高通量测序。

[0041] 前文所述可共价结合底物的标签蛋白可为如下任一: Halo标签蛋白, CLIP标签蛋白, SNAP标签蛋白, Spy标签蛋白等。Halo标签是Promega公司经过基因工程改造的微生物紫红球菌的脱卤素酶。该酶的底物是含有卤素的配体。SNAP标签来自人烷基鸟嘌呤-DNA烷基转移酶,是NEB公司经过基因工程编辑的一种DNA修复蛋白。该酶的底物是苄基鸟嘌呤。CLIP标签是由SNAP标签衍生而来。该酶的底物是苄基胞嘧啶衍生物。Spy标签来自于酿脓链球菌纤连蛋白结合蛋白(fibronectin-binding protein, FbaB) CnaB2结构域的11个氨基酸的短肽。该标签的底物是SpyCatcher蛋白。

[0042] 相应的,所述可共价结合底物的标签蛋白的特异性结合物为如下任一:

[0043] (A) 当所述可共价结合底物的标签蛋白为Halo标签蛋白时,所述可共价结合底物

的标签蛋白的特异性结合物为含有卤素的配体；

[0044] (B) 当所述可共价结合底物的标签蛋白为CLIP标签蛋白时,所述可共价结合底物的标签蛋白的特异性结合物为苄基胞嘧啶衍生物；

[0045] (C) 当所述可共价结合底物的标签蛋白为SNAP标签蛋白时,所述可共价结合底物的标签蛋白的特异性结合物为苄基鸟嘌呤；

[0046] (D) 当所述可共价结合底物的标签蛋白为Spy标签蛋白时,所述可共价结合底物的标签蛋白的特异性结合物为SpyCatcher蛋白。

[0047] 在本发明的一个实例中,所述可共价结合底物的标签蛋白具体为Halo标签蛋白。更加具体的,所述Halo标签蛋白的氨基酸序列具体如序列表中序列1所示。

[0048] 在本发明的一个实例中,所述靶蛋白具体为PTB蛋白。所述Halo标签蛋白位于PTB蛋白的N端。所述受体细胞为HEK293细胞。

[0049] 一般生理条件下,Halo标签蛋白的活性中心只能与含有卤素的配体以共价键的形式相互结合,而且该共价键不能打开。因此,当将靶蛋白和Halo标签结合后,可以通过在体外加入含荧光的Halo配体使之进行共价结合。如果将Halo配体连接在磁珠上,就可以将Halo标签蛋白质共价偶联到磁珠上。Halo标签的分子量为33kDa,可以结合在靶蛋白的N端或者C端。该融合蛋白可以在原核或真核表达系统中表达。本发明利用Halo标签可以与配体共价结合的特点来设计新的CLIP方案。

[0050] 本发明涉及使用Halo标签融合蛋白通过变性剂纯化的方式在CLIP(紫外交联免疫共沉淀)技术中的应用。传统CLIP步骤需要通过SDS-PAGE电泳、醋酸纤维素转膜和同位素对蛋白质RNA复合物进行纯化。其耗时长,样本易损失。本发明使用Halo标签的融合蛋白通过蛋白变性试剂洗涤的方式得到可以用于测序的RNA文库。本发明技术步骤如下:对带有Halo标签融合蛋白的细胞系进行紫外交联收集样本。细胞裂解后,用核酸酶修剪RNA的长度。然后通过Magne® HaloTag® Beads进行免疫沉淀。接着对免疫沉淀后的RNA进行去磷酸化反应,并在RNA的3'末端加上接头序列。然后使用强蛋白变性剂洗涤去除残留的RNA污染物。最后将蛋白消化后,得到纯化后的RNA。将RNA构建成cDNA文库后即可进行高通量测序。本发明又称为GoldCLIP技术(Gel-Omitted and Ligation-Dependent Cross-Linking and ImmunoPrecipitation)。

[0051] 本发明以PTB蛋白作为靶蛋白,验证所提供的GoldCLIP技术,结果证明可以很好地重复以前的文献发表的实验结果。在操作上省略了传统CLIP步骤中需要通过转膜并且切胶回收片段的步骤。该方法通过Halo标签可以和磁珠配体形成共价键的特性,将传统上不能在磁珠上直接进行的剧烈洗涤的方法加入到了实验操作步骤中。这样在操作上也比较简便,节省了操作的时间,优化了CLIP的步骤,减少了操作过程中可能会造成的损失。

附图说明

[0052] 图1为使用Halo标签融合蛋白并通过非胶纯化方式进行紫外交联免疫共沉淀的技术流程图。带有Halo标签的目标蛋白通过病毒感染的方法转入细胞系后,进行紫外交联收集样本。细胞裂解并进行核酸酶处理,修剪结合蛋白的RNA。然后通过连有Halo配体的磁珠进行免疫沉淀。接着通过温柔的方式洗涤磁珠,将非特异的条带洗去,然后再通过碱性磷酸酶去除3'端的磷酸基团后,通过Trizol试剂去除碱性磷酸酶,同时进行第一次蛋白变性洗

涤。接着,在RNA的3'末端加上接头,并通过蛋白变性洗涤等强烈的方式进行第二次变性洗涤,去除残留的RNA污染物。进行Tev酶切释放蛋白核酸复合物,然后再将蛋白消化掉,纯化得到的RNA。接着通过iCLIP文献中介绍的方法构建cDNA文库,将得到的cDNA进行高通量测序。

[0053] 图2为质粒MSCV-NHalo-3xTev-PTB-T2-puro的图谱。

[0054] 图3为Halo-PTB免疫沉淀结果图。Halo-PTB UV254和UV365样品通过Magne[®] HaloTag[®] Beads进行免疫沉淀后,用HaloTag[®] Alexa Fluor[®]660Ligand进行检测,PAGE电泳后,通过Odyssey双色红外荧光成像系统进行检测。in:免疫沉淀前的带有Halo标签的蛋白;ub:免疫沉淀后未结合Magne[®] HaloTag[®] Beads的Halo标签的蛋白;星号:Halo-PTB (~98kD);三角:Halo-YFP (~65kD)。

[0055] 图4为PTB高通量测序得到的数据量。PTB高通量数据在经过Pyicoclip软件处理后的有效序列数据。依次显示了PTB UV365,PTB UV254和YFP UV254得到的两组重复实验数据量的结果。

[0056] 图5为PTB序列的全基因组序列比对结果。PTB UV254,PTB UV365数据在全基因组比对后,数据在基因组功能区域的分布百分比图。

[0057] 图6为PTB各重复数据间的相关性分析。A.PTB UV254两组重复数据间的相关性比较;B.PTB UV365两组重复数据间的相关性比较。

[0058] 图7为PTB数据的Motif分析。A.PTB UV365rep1数据中的前5个Motif序列结果;B.PTB UV365rep2数据中的前5个Motif序列结果;C.PTB UV254rep1数据中的前5个Motif序列结果;D.PTB UV254rep2数据中的前5个Motif序列结果;E.YFP UV254rep1数据中的前5个Motif序列结果;F.YFP UV254rep2数据中的前5个Motif序列结果。

具体实施方式

[0059] 下述实施例中所使用的实验方法如无特殊说明,均为常规方法。

[0060] 下述实施例中所用的材料、试剂等,如无特殊说明,均可从商业途径得到。

[0061] 本发明所提供的GoldCLIP技术的操作步骤大致如下:带有Halo标签的目标蛋白通过病毒感染的方法转入细胞系后,进行紫外交联收集样本。细胞裂解并进行核酸酶处理,修剪结合蛋白的RNA。然后通过连有Halo配体的磁珠进行免疫沉淀。接着通过温柔的方式洗涤磁珠,将非特异的条带洗去,然后再通过碱性磷酸酶去除3'端的磷酸基团后,通过Trizol试剂(第一次变性洗涤)去除碱性磷酸酶。接着,在RNA的3'末端加上接头,并通过蛋白变性洗涤(第二次变性洗涤)等强烈的方式,去除残留的RNA污染物。然后通过Tev蛋白酶,将蛋白消化后,纯化得到的RNA。接着通过iCLIP文献中介绍的方法构建cDNA文库。RNA反转录并加上每个样本独立的标签序列之后,跑TEB-Urea变性胶纯化。反转录用到引物序列中不仅包含有独立的标签序列,还有5'端的接头序列和3'端的接头序列,以及接头序列之间的BamHI酶切位点。得到单链cDNA通过ssDNA环化酶环化,然后通过PCR仪将一条与引物互补配对的序列退火到引物序列上。接着通过BamHI酶切,得到产物具有5'端的接头序列,标签序列和3'端的接头序列。再通过PCR扩增后,将得到的cDNA进行高通量测序。具体技术流程图见图1。

[0062] 实施例1、带有Halo标签的PTB蛋白在HEK293细胞系中进行CLIP实验

[0063] 本实施例将通过对带有Halo标签的PTB蛋白在HEK293细胞系中进行CLIP实验,进

而研究本发明所提供的GoldCLIP技术在RNA结合蛋白中的应用。

[0064] (1) 在HEK293细胞系中表达Halo-PTB融合蛋白

[0065] 将带有Halo-PTB融合蛋白编码基因的质粒MSCV-NHalo-3xTev-PTB-T2-puro (Halo标签蛋白和PTB蛋白通过3个串联的Tev酶的特异性识别序列Glu-Asn-Leu-Tyr-Phe-Gln-Gly/Ser相连,该质粒的图谱如图2所示,全序列如序列表中序列2所示,其中序列2的第1419-2309位编码序列表中序列1所示的Halo标签蛋白,第2322-2342位、第2349-2369位以及第2376-2396位编码3个串联的Tev酶的特异性识别序列,第2598-4193位编码Human PTB蛋白)转入HEK293细胞系中,通过Puromycin药筛稳定细胞系。通过HaloTag® Alexa Fluor®660Ligand (Promega,G8471)检测细胞系中Halo-PTB的表达。

[0066] 实验同时设置在HEK293细胞系中表达Halo-YFP融合蛋白的对照。其中,导入HEK293细胞系中的质粒为将序列表中序列1的第2598-4193位替换为YFP编码基因(序列3),其余操作同上。

[0067] (2) 收集UV交联的细胞

[0068] 将带有Halo-PTB的细胞系扩大培养到2盘100mm细胞培养皿上。其中一盘细胞铺满100mm培养皿后,进行UV254nm紫外交联。另一盘细胞铺满80%100mm培养皿后,加入4SU(终浓度200 μ M),继续培养24小时后,进行UV365nm紫外交联。将带有Halo-YFP的细胞系扩大培养到1盘100mm细胞培养皿上。细胞铺满100mm培养皿后,进行UV254nm紫外交联。细胞进行紫外交联后,用细胞刮刀刮下细胞,用预冷的PBS洗涤两次,收集到15ml离心管中。收集后的细胞或置于冰上直接进行后续实验,或放置于-80 $^{\circ}$ C保存。

[0069] (3) 利用Magne® HaloTag® Beads对Halo-PTB进行免疫共沉淀

[0070] 细胞通过裂解液(配方:50mM Tris-HCl pH 7.5;100mM NaCl;1mM DTT;2mM CaCl₂;10%甘油;1 \times 蛋白酶抑制剂[Promega];0.2%TritonX-100;0.5U/ μ l Micrococcal Nuclease [NEB];余量为水;%表示体积百分含量)重悬细胞后,通过Dounce工具在冰上对细胞进行破核后,冰上静置10分钟。然后37 $^{\circ}$ C反应3分钟后,立刻放置于冰上。加入终浓度2mM EDTA和2mM EGTA后,高速离心10分钟。将上清液转移到新离心管中,收集RNA input和用于western blot的input,并加入300 μ l Magne® HaloTag® Beads,在4 $^{\circ}$ C免疫沉淀过夜。第二天通过HaloTag® Alexa Fluor®660Ligand (Promega,G8471)检测免疫沉淀的效率。由于Halo标签和磁珠形成共价键,因此通过比较沉淀后的样品中Halo标签蛋白的表达量(unbound)相对于未沉淀中Halo标签蛋白的表达量(input)的减少比例,可以估算中免疫沉淀的效率和效果。结果参见图3。

[0071] (4) On-beads去磷酸化处理

[0072] 用PBST洗涤(即非变性洗涤,具体为“用含有0.1%Triton X-100和500mM NaCl的PBS缓冲液洗涤2次,用含有0.1%Triton X-100的PBS缓冲液洗涤3次;其中%表示体积百分含量”)结合了Halo-PTB蛋白或Halo-YFP的Magne® HaloTag® Beads(磁珠)后,用1 \times NEB Cutsmart缓冲液洗涤3次,然后加入80 μ l碱性磷酸酶(CIP)和DNase I反应体系(8 μ l 10 \times NEB CutSmart;8 μ l 10mM CaCl₂;5 μ l CIP [NEB,M0290S];5 μ l DNase I [Promega];2 μ l RNase inhibitor [Promega];52 μ l H₂O),使用Thermomixer仪器,在37 $^{\circ}$ C反应30分钟(其中每3分钟,1000rpm混匀15秒)。

[0073] (5) 第一次蛋白变性洗涤

[0074] 去磷酸化反应结束后加入750 μ l Trizol LS (Invitrogen), 200 μ l CHCl_3 后, 涡旋15秒。接着磁珠用8M盐酸胍洗涤2次(每次5分钟), 8M尿素洗涤2次(每次5分钟), 最后用PNK缓冲液(配方: 20mM Tris-HCl pH 7.0; 10mM MgCl_2 ; 0.1% TritonX-100; 余量为水; %表示体积百分含量)洗涤3次。

[0075] (6) On-beads 3' RNA接头连接

[0076] 向磁珠中加入40 μ l RNA接头预混液(配方: 2 μ l RNase抑制剂; 2 μ l 3' RNA Linker [100 pmol/ μ l]; 36 μ l H_2O 。其中, 3' RNA Linker的具体序列为: /5' P/AGGTCCGAAGAGCGGTTTCAG/3' ddC/, 由IDT公司合成), 再加入55 μ l RNA连接反应预混液(配方: 10 μ l 10 \times T4RNA ligase buffer[NEB]; 10 μ l BSA; 10 μ l 10mM ATP; 10 μ l DMSO; 10 μ l T4RNA ligase[Ambion, AM2141]; 5 μ l 0.1M DTT)。使用Thermomixer仪器, 在16 $^\circ\text{C}$ 反应过夜(其中每3分钟, 1000rpm混匀15秒)。第二天, 再加入12 μ l连接反应预混液(配方: 1 μ l 3' RNA Linker [100pmol/ μ l]; 5 μ l T4RNA ligase[Ambion, AM2141]; 5 μ l 10mM ATP; 1 μ l 0.1M DTT)在25 $^\circ\text{C}$ 继续反应3小时(其中每3分钟, 1000rpm混匀15秒)。

[0077] (7) 第二次蛋白变性洗涤

[0078] 去除磁珠中的连接反应体系, 用PBST洗涤1次, 用SDS洗涤液(配方: 10% SDS; 50mM Tris-HCl pH 7.0; 1mM EDTA; 1mM DTT; 余量为水; %表示质量体积百分含量, 即10%表示10g/100ml)洗涤3次, 每次使用Thermomixer仪器, 在65 $^\circ\text{C}$ 以1000rpm的转速振荡5分钟。接着用8M尿素洗涤2次, 每次室温振荡5分钟。

[0079] (8) Tev酶切反应将PTB蛋白从磁珠上分离出来

[0080] 用1 \times Tev酶切反应液(配方: 50mM Tris-HCl pH 8.0; 1mM EDTA; 1mM DTT; 1% TritonX-100; 2M Urea; 余量为水; %表示体积百分含量)洗涤3次。用200 μ l 1 \times Tev酶切反应液重悬磁珠, 加入5 μ l RNase抑制剂, 5 μ l Tev酶(5 $\mu\text{g}/\mu\text{l}$), 使用Thermomixer仪器, 25 $^\circ\text{C}$ 反应2小时(其中每3分钟, 1000rpm混匀15秒)。将反应液转移到新的1.5mL离心管中。用100 μ l 1 \times Tev酶切反应液重悬磁珠, 加入2 μ l RNase抑制剂, 2 μ l Tev酶(5 $\mu\text{g}/\mu\text{l}$), 使用Thermomixer仪器, 30 $^\circ\text{C}$ 反应2小时(其中每3分钟, 1000rpm混匀15秒)。将反应液和上次的反应液合并。

[0081] (9) RNA的提取

[0082] 在合并后的Tev反应液中加入100 μ l PK反应液(配方: 4 μ l 5M NaCl; 20 μ l 1M Tris-HCl pH 7.0; 8 μ l 10% SDS; 8 μ l H_2O ; 60 μ l 蛋白酶K[NEB, P8102S]), 37 $^\circ\text{C}$ 水浴30分钟。用400 μ l 酚氯仿溶液(Sigma)抽提RNA, 涡旋离心后, 上层液体转移到新的1.5mL离心管中, 加入50 μ l 3M NaAc pH 5.5, 1 μ l glycoblue, 1mL 1:1(体积比)乙醇:异丙醇, 在-20 $^\circ\text{C}$ 静置过夜。第二天离心后, 将沉淀物用75%乙醇洗涤2次后, 用16 μ l H_2O 重悬后即得到RNA文库(即所收集的与PTB蛋白互作的靶标RNA的集合)。

[0083] (10) cDNA文库的构建

[0084] cDNA的文库构建参考iCLIP的方法【参见Konig, J., et al. (2011). "iCLIP—transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution." J Vis Exp (50).】。简而言之, 首先使用Superscript III (Invitrogen) 将RNA逆转录成cDNA。使用的逆转录引物为/5Phos/NN[Index]

NNNAGATCGGAAGAGCGTCGTGgataCTGAACCGC,其中Index为AACC,ACAA,ATTG,AGGT,CGCC,CCGG,CTAA,CATT,GCCA,GACC,GGTT,GTGG,TCCG,TGCC,TATT,TAA。然后将cDNA跑6%TEB-urea gel (Invitrogen),将85nt-200nt的目的片段切胶回收,并酒精沉淀过夜。回收后的cDNA用Circligase II (Epicentre)进行环化。接着加入30 μ l oligo anneal mix,通过PCR退火的方式将oligo与目的序列结合。其中oligo序列为5'-GTTCAGGATCCACGACGCTCTTCAAAA-3'。然后用BamHI-HF (NEB)酶切反应1小时。最后通过2 \times Phusion Mix扩增cDNA文库。所用的扩增正向引物为5'-AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTTCCGATCT-3',反向引物为5'-CAAGCAGAAGACGGCATAACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT-3'。

[0085] (11) 高通量测序

[0086] 将得到的cDNA文库送测序平台测序,主要测序平台有illumina HiSeq 2500,illumina HiSeq X10。选择SE50或PE250等方式测序。

[0087] (12) 数据处理

[0088] 通过Cutadapt软件(v1.10<http://journal.embnet.org/index.php/embnetjournal/article/view/200;cutadapt-a AGATCGGAAGAGCGGTTCAG-e 0.2-q 20-m 24>)过滤接头序列以及修剪3'末端低质量碱基的序列,并将每条序列的碱基数少于24bp的去掉。然后将重复序列合并后取出5'端的条码序列和随机序列。经过修剪后的数据再通过Pyicoclip程序(Pyicoclip:-P-value 0.001-region*.bed)寻找peak。经过Pyicoclip程序后得到的结果如图4所示。由图可见,经过UV365照射的含有Halo-PTB的细胞(实验组)的两个生物学重复经本发明方法后得到了大约10万个peak。经过UV254照射的含有Halo-PTB的细胞(实验组)的两个生物学重复经本发明方法后得到了大约20万个peak。而经过紫外照射的含有Halo-YFP细胞(对照组)的两个生物学重复经过本发明方法几乎没有peak。

[0089] 接着使用Bowtie(v1.1.2(REF);bowtie-f-p 8-v 2-k 1--best--sam--un)将每条序列与人的基因组进行比对(hg19)。PTB序列的全基因组序列比对结果如图5所示。由图可见,无论是经过UV365照射还是UV254照射的Halo-PTB细胞,内含子(intron)部分peak的比重占所有peak的60%以上。其次是intergenic区域,约占10%的比重。

[0090] (13) 数据相关性分析

[0091] 通过Pearson相关系数来比较Halo-PTB UV254,UV365和NoUV数据的两组实验之间的相关性。结果如图6所示,显示PTB UV254的两组数据的相关性 $R^2=0.97$,PTB UV365两组数据的相关性 $R^2=0.99$ 。说明实验具有重复性,数据可靠。

[0092] (14) Motifs分析

[0093] 通过HOMER软件(-p 8-rna-S 10-len 5,6,7)寻找Peak序列中的motif。方法是通过序列信息和背景进行比较。背景信息通过在同一基因区域的三组随机序列而产生的。图7所示结果列举了找寻到的Motif序列中的前5个序列。结果显示无论是PTB UV254还是PTB UV365,通过该方法进行的实验都能够找寻到UC rich的Motif序列,并且与发表的结果一致【参见Xue,Y.,et al.,Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping.Mol Cell,2009.36(6):p.996-1006.】。

[0094] 本实施例的结果显示:使用Halo标签融合蛋白并通过非胶纯化方式进行紫外交联免疫共沉淀的技术的CLIP方法可以很好地重复以前的文献发表的实验结果。在操作上省略

了传统CLIP步骤中需要通过转膜并且切胶回收片段的步骤。该方法通过Halo标签可以和磁珠配体形成共价键的特性,将传统上不能在磁珠上直接进行的剧烈洗涤的方法加入到了实验操作步骤中。这样在操作上也比较简便,节省了操作的时间,优化了CLIP的步骤,减少了操作过程中可能会造成的损失。

Leu Pro Ile Ala Gly Glu Pro Ala Asn Ile Val Ala Leu Val Glu Glu
 210 215 220
 Tyr Met Asp Trp Leu His Gln Ser Pro Val Pro Lys Leu Leu Phe Trp
 225 230 235 240
 Gly Thr Pro Gly Val Leu Ile Pro Pro Ala Glu Ala Ala Arg Leu Ala
 245 250 255
 Lys Ser Leu Pro Asn Cys Lys Ala Val Asp Ile Gly Pro Gly Leu Asn
 260 265 270
 Leu Leu Gln Glu Asp Asn Pro Asp Leu Ile Gly Ser Glu Ile Ala Arg
 275 280 285
 Trp Leu Ser Thr Leu Glu Ile Ser Gly
 290 295

<210> 2

<211> 9099

<212> DNA

<213> 人工序列

<220>

<223>

<400> 2

tgaagacc cactgtagg tttggcaagc tagcttaagt aacgccattt tgcaaggcat 60
 ggaaaataca taactgagaa tagagaagtt cagatcaagg ttaggaacag agagacagca 120
 gaatatgggc caaacaggat atctgtggta agcagttcct gccccggctc agggccaaga 180
 acagatggtc cccagatgcg gtcccgcct cagcagtttc tagagaacca tcagatgttt 240
 ccagggtgcc ccaaggacct gaaatgacc tgtgccttat ttgaactaac caatcagttc 300
 gcttctcgt tctgttcgcg cgcttctgct ccccgagctc aataaaagag cccacaacc 360
 ctactcggc gcgccagtc tccgatagac tgcgtcgccc ggtaccctg attccaata 420
 aagcctcttg ctgtttgcat ccgaatcgtg gactcgtga tccttgggag ggtctcctca 480
 gattgattga ctgcccacct cgggggtctt tcatttggag gttccaccga gatttggaga 540
 cccctgcca gggaccaccg accccccgc cgggaggtaa gctggccagc ggtcgtttcg 600
 tgtctgtctc tgtctttgtg cgtgtttgtg ccggcatcta atgtttgcgc ctgcgtctgt 660
 actagttagc taactagctc tgtatctggc ggaccctgg tggaactgac gagttctgaa 720
 caccggccg caacctggg agacgtcca gggactttgg gggccgtttt tgtggcccga 780
 cctgaggaag ggagtcgat tggaatccga cccgctcagg atatgtggtt ctggtaggag 840
 acgagaacct aaaacagtt ccgcctcct ctgaatttt gctttcgggt tggaaccgaa 900
 gccgcgcgtc ttgtctgctg cagcgtgca gcatcgttct gtgttctctc tgtctgactg 960
 tgtttctgta tttgtctgaa aattagggc agactgttac cactccetta agtttgacct 1020
 taggtcactg gaaagatgct gagcggatcg ctcaacca gtcggtagat gtcaagaaga 1080
 gacgttgggt taccttctgc tctgcagaat ggccaacct taacgtcgga tggccgcgag 1140
 acggcacct taaccgagac ctcatcacc aggttaagat caaggtcttt tcacctggcc 1200

cgcatggaca cccagaccag gtcccctaca tcgtgacctg ggaagccttg gcttttgacc 1260
 cccctccctg ggtaagccc ttgtacacc ctaagcctcc gcctcctctt cctccatccg 1320
 ccccgctctct ccccttgaa cctcctcgtt cgaccccgcc tcgatcctcc ctttatccag 1380
 ccctcactcc ttctctaggc gccggaatta gatccacat ggacagaaatc ggtactggct 1440
 ttccattcga cccccattat gtggaagtcc tgggcgagcg catgcactac gtcgatgttg 1500
 gtccgcgcga tggcaccctt gtgctgttcc tgcacggtaa cccgacctcc tcctacgtgt 1560
 ggcgcaacat catcccgcac gttgcaccga cccatcgctg cattgctcca gacctgatcg 1620
 gtatgggcaa atccgacaaa ccagacctgg gttatttctt cgacgaccac gtccgcttca 1680
 tggatgcctt catcgaagcc ctgggtctgg aagaggtcgt cctggctcatt cacgactggg 1740
 gctccgctct gggtttccac tgggccaagc gcaatccaga gcgctcaaaa ggtattgcat 1800
 ttatggagtt catccgcctt atcccgacct gggacgaatg gccagaattt gcccgcgaga 1860
 ccttccagge ctccgcacc accgacgtcg gccgcaagct gatcatgat cagaacgttt 1920
 ttatcgaggg tacgtgccg atgggtgtcg tccgccgct gactgaagtc gagatggacc 1980
 attacgcga gccgttctg aatcctgttg accgcgagcc actgtggcgc tcccacaaacg 2040
 agctgccaat cgccggtgag ccagcgaaca tcgtcgcgct ggtcgaagaa tacatggact 2100
 ggctgcacca gtcccctgtc ccgaagctgc tgttctgggg caccacagge gttctgatcc 2160
 caccggccga agccgctcgc ctggccaaaa gcctgcctaa ctgcaaggct gtggacatcg 2220
 gcccgggtct gaatctgctg caagaagaca acccgacct gatcggcagc gagatcgcgc 2280
 gctggctgtc gacgtcagc atttccggcg agccaaccac tgaggatctg tactttcaga 2340
 gcatcgatga aaatctgtac ttccagggga tcgatgagaa cctgtacttt caggggagcg 2400
 cctggtcca ccccagttc gaaaagggcg gcggaagcgg aggaggctcc ggaggttccg 2460
 ctggtcca cccgagttc gagaagaccg gcgccattta tcaacaagt ttgtacaaaa 2520
 aagcaggctc cacatggga accaattcag tcgactggat cctcatggct agcatgactg 2580
 gtggacagca aatgggtatg gacggcattg tccagatat agccgttggt acaaagcggg 2640
 gatctgacga gcttttctt acttgtgtca ctaacggacc gtttatcatg agcagcaact 2700
 cggcttctgc agcaaacgga aatgacagca agaagttcaa aggtgacagc cgaagtgcag 2760
 gcgtcccctc tagagtgat cacatccgga agtccccat cgacgtcacg gagggggaag 2820
 tcactccct ggggtgcc tttgggaagg tcaccaacct cctgatgctg aaggggaaaa 2880
 accaggcctt catcgagatg aacacggagg aggtgcca caccatggtg aactactaca 2940
 cctcggtagc cctgtgctg cgcggccagc ceatctacat ccagttctct aaccacaagg 3000
 agctgaagac cgacagctct cccaaccagg cgcgggcccc ggccggcctg caggcgggtga 3060
 actcgggtcca gtcggggaac ctggccttgg ctgcctcggc ggccggcctg gacgcaggga 3120
 tggcgatggc cgggcagagc cctgtgtctc ggatcatcgt ggagaacctc ttctaccctg 3180
 tgaccctgga tgtgtgca cagatttctt ccaagttcgg cacagtgttg aagatcatca 3240
 ccttaccaca gaacaaccag ttccaggccc tgctgcagta tgcggacccc gtgagcgc 3300
 agcacgcaa gctgtcgtg gacgggcaga acatctacaa cgctgtctgc acgtgcgca 3360
 tcgactttc caagtcacc agcctcaacg tcaagtacaa caatgacaag agccgtgact 3420
 acacagccc agacctgct tccggggaca gccagccctc gctggaccag accatggccc 3480
 cggccttcgg ctttccgtt ccgaacgtcc acggcgcctt ggccccctg gccatccct 3540

cggcggcggc ggcagctgcg gcggcaggtc ggatecgcct cccgggcctg gcgggggcag 3600
 gaaattctgt attgctggtc agcaacctca acccagagag agtcacaccc caaagcctct 3660
 ttattctttt cggcgtctac ggtgacgtgc agcgcgtgaa gatcctgttc aataagaagg 3720
 agaacgccct agtgcagatg gcggacggca accaggceca gctggccatg agccacctga 3780
 acgggcacaa gctgcacggg aagcccatcc gcatcacgct ctcgaagcac cagaacgtgc 3840
 agctgccccg cgagggccag gaggaccagg gcctgaccaa ggactacggc aactcacccc 3900
 tgcaccgctt caagaagccg ggtccaaga acttcagaa catattcccg ccctcggcca 3960
 ctctgcacct ctccaacatc ccgcctcag tctccagga ggatctcaag gtctgtttt 4020
 ccagcaatgg gggcgtcgtc aaaggattca agttcttcca gaaggaccgc aagatggcac 4080
 tgatccagat gggctccgtg gaggaggcgg tccaggeect cattgacctg cacaaccacg 4140
 acctcgggga gaaccaccac ctgcgggtct cttctccaa gtccaccatc tagctcgaga 4200
 tatctagacc cagctttctt gtacaaagtg gttcgataaa ttgacgtaag ctagtctaga 4260
 cggaattcta ccgggtaggg gaggcgcttt tccaagga gtctggagca tgcgctttag 4320
 cagccccgct gggcaacttg cgtacacaa gtggcctctg gcctcgaca cattccacat 4380
 ccaccgtag gcgccaaccg gctccgttct ttggtggccc cttegcgcca cttctactc 4440
 ctcccctagt caggaagttc cccccgcc cgcagctcgc gtcgtgcagg acgtgacaaa 4500
 tggaaagtag acgtctcact agtctcgtgc agatggacag caccgctgag caatggaagc 4560
 gggtaggcct ttggggcagc ggccaatagc agctttgctc cttegtttc tgggctcaga 4620
 ggctgggaag ggggtgggtcc gggggcgggc tcagggcgcg gctcaggggc ggggcgggcg 4680
 cccgaaggtc ctccggagge ccggcattct gcacgcttca aaagcgcacg tctgcccgcg 4740
 tgttctctc ttctcatct ccgggccttt cgacctgcag cccaagctta ccatgaccga 4800
 gtacaagccc acggtgcgcc tcgccaccg cgacgacgtc cccagggccg tacgcacct 4860
 cgccgccgcg ttcgccgact accccgccac gcgccacacc gtcgatccgg accgccacat 4920
 cgagcgggtc accgagctgc aagaactctt cctcacgcgc gtcgggctcg acatcggcaa 4980
 ggtgtgggtc gcggacgacg gcgcccggtt ggcggtctgg accacgccgg agagcgtcga 5040
 agcgggggcg gtgttcgccg agatcggccc gcgcatggcc gagttgagcg gttcccggct 5100
 ggccgcgcag caacagatgg aaggcctcct ggccgcgcac cggcccaagg agcccgcgtg 5160
 gttcttgcc accgtcggcg tctgcccga ccaccaggc aagggtctgg gcagcgcctg 5220
 cgtgctcccc ggagtggagg cggccgagcg cgcgggggtg cccgccttcc tggagacctc 5280
 cgcgccccgc aacctcccct tctacgagcg gctcggttcc accgtcaccg ccgacgtcga 5340
 ggtgcccga ggaccgcgca cctggtgcat gaccgcaag cccggtgect gacgcccgcc 5400
 ccacgaccg cagcgcgca ccgaaaggag cgcacgacc catgcatcga taaaataaaa 5460
 gattttattt agtctccaga aaaagggggg aatgaaagac cccacctgta ggtttggcaa 5520
 gctagcttaa gtaacgcat tttgcaaggc atggaaaata cataactgag aatagagaag 5580
 ttcagatcaa ggtaggaac agagagacag cagaatatgg gccaaacagg atatctgtgg 5640
 taagcagttc ctgccccg ctagggccaa gaacagatgg tccccagatg cgggtcccgc 5700
 ctcagcagtt tctagagaac catcagatgt ttccagggtg cccaaggac ctgaaatgac 5760
 cctgtgcctt atttgaacta accaatcagt tegettctcg cttctgttctg cgcgcttctg 5820
 ctccccgagc tcaataaaaag agcccacaac ccctcactcg gcgcgccagt cctccgatag 5880

actgcgtcgc ccgggtaccc gtgtatccaa taaacctctt tgcagttgca tccgacttgt 5940
 ggtctcgcctg ttccttggga gggctcctc tgagtgattg actaccctgc agcgggggctc 6000
 tttcatgggt aacagtttct tgaagttgga gaacaacatt ctgagggtag gagtcgaata 6060
 ttaagtaatc ctgactcaat tagccactgt tttgaatcca catactccaa tactcctgaa 6120
 atagttcatt atggacagcg cagaaagagc tggggagaat tgtgaaattg ttatccgctc 6180
 acaattccac acaacatacg agccggaagc ataaagtgt aagcctgggg tgcctaataga 6240
 gtgagctaac tcacattaat tgcgttgcgc tcaactgccc ctttccagtc gggaaacctg 6300
 tcgtgccagc tgcattaatg aatcggccaa cgcgcgggga gaggcggttt gcgtattggg 6360
 cgctcttccg ctctctcgt cactgactcg ctgcgctcgg tcgttcggct gcggcgagcg 6420
 gtatcagctc actcaaagge ggtaatacgg ttatccacag aatcagggga taacgcagga 6480
 aagaacatgt gagcaaaaagg ccagcaaaaag gccaggaacc gtaaaaaggc cgcgttgctg 6540
 gcgtttttcc ataggetccg cccccctgac gagcatcaca aaaatcgacg ctcaagtcag 6600
 aggtggcgaa acccgacagg actataaaga taccaggcgt tccccctgg aagctccctc 6660
 gtgcgctctc ctgttccgac cctgccgctt accggatacc tgtccgctt tctccctteg 6720
 ggaagcgtgg cgttttctca tagctcagc tgtaggtatc tcagttcggg ttaggtcgtt 6780
 cgctccaagc tgggctgtgt gcacgaacc cccgttcagc ccgaccctg cgccttatcc 6840
 ggtaactatc gtcttgagtc caaccggta agacacgact tatcgccact ggcagcagcc 6900
 actggtaaca ggattagcag agcgaggtat gtaggcgggt ctacagagtt cttgaagtgg 6960
 tggcctaact acggctacac tagaagaaca gtatttggt tctgcgctct gctgaagcca 7020
 gttaccttcg gaaaaagagt tggtagctct tgatccggca acaaaccac cgctggtagc 7080
 ggtgggtttt ttgtttgcaa gcagcagatt acgcgcagaa aaaaaggatc tcaagaagat 7140
 cctttgatct tttctacggg gtctgacgct cagtggaacg aaaactcacg ttaagggatt 7200
 ttggatcatga gattatcaaa aaggatctt acctagatcc ttttaaatta aaaatgaagt 7260
 tttaaatcaa tctaaagtat atatgagtaa acttggctctg acagttacca atgcttaatc 7320
 agtgaggcac ctatctcagc gatctgtcta tttcgttcat ccatagttgc ctgactcccc 7380
 gtcgtgtaga taactacgat acgggagggc ttaccatctg gccccagtgc tgcaatgata 7440
 ccgcgagacc cacgctcacc ggtccagat ttatcagcaa taaaccagcc agccggaagg 7500
 gccgagcgca gaagtgttc tgcacttta tccgctcca tccagtctat taattgttgc 7560
 cgggaagcta gagtaagtag ttcgccagtt aatagtttgc gcaacgttgt tgccattgct 7620
 acaggcatcg tgggtgcacg ctgctcgtt ggtatggctt cattcagctc cggttcccaa 7680
 cgatcaagge gagttacatg atccccatg ttgtgcaaaa aagcggttag ctcttcgggt 7740
 cctccgatcg ttgtcagaag taagttggcc gcagtgttat cactcatggt tatggcagca 7800
 ctgcataatt ctcttactgt catgccatc gtaagatgct tttctgtgac tggtgagtac 7860
 tcaaccaagt cattctgaga atagtgtatg cggcgaccga gttgctcttg cccggcgtea 7920
 atacgggata ataccgccc acatagcaga actttaaag tgctcatcat tggaaaacgt 7980
 tcttcggggc gaaaactctc aaggatctta ccgctgttga gatccagttc gatgtaacce 8040
 actcgtgcac ccaactgatc ttcagcatct tttactttca ccagcgttcc tgggtgagca 8100
 aaaacaggaa ggcaaaaatgc cgcaaaaaag ggaataaggc cgacacggaa atgttgaata 8160
 ctcaactct tccttttca atattattga agcatttate aggttattg tctcatgagc 8220

ggatacatat ttgaatgtat ttagaaaaat aaacaaatag gggttccgcg cacatttccc 8280
 cgaaaagtgc cacctgacgt ctaagaaacc attattatca tgacattaac ctataaaaaat 8340
 aggcgatatca cgaggccctt tcgtctcgcg cgtttcgggtg atgacgggtga aaacctctga 8400
 cacatgcagc tcccggagac ggtcacagct tgtctgtaag cggatgcccg gagcagacaa 8460
 gcccgtcagg gcgcgtcagc ggggtgtggc ggggtgtcggg gctggcttaa ctatgctgca 8520
 tcagagcaga ttgtactgag agtgcacat atgcgggtgtg aaataccgca cagatgctga 8580
 aggagaaaaat accgcatcag gcgccattcg ccattcaggc tgcgcaactg ttgggaaggg 8640
 cgatcgggtgc gggcctcttc gctattacgc cagctggcga aagggggatg tgctgcaagg 8700
 cgattaagtt gggtaacgcc agggttttcc cagtcacgac gttgtaaaac gacggcgcaa 8760
 ggaatgggtgc atgcaaggag atggcgccca acagtcccc ggccacgggg cctgccacca 8820
 taccacgcc gaaacaagcg ctcatgagcc cgaagtggcg agcccgatct tccccatcgg 8880
 tgatgtcggc gatataggcg ccagcaaccg cacctgtggc gccgggtgat cgggccacga 8940
 tgcgtccggc gtagaggcga ttagtccaat ttgttaaaga caggatatca gtggtccagg 9000
 ctctagtttt gactcaacaa taccaccagc tgaagcctat agagtacgag ccatagataa 9060
 aataaaagat tttatttagt ctccagaaaa aggggggaa 9099

<210> 3

<211> 720

<212> DNA

<213> 人工序列

<220>

<223>

<400> 3

atgggtgagca agggcgagga gctgttcacc ggggtgtgtc ccatcctggt cgagctggac 60
 ggcgacgtaa acggccacaa gttcagcgtg tccggcgagg gcgagggcga tgccacctac 120
 ggcaagctga ccctgaagtt catctgcacc accggcaagc tgcccgtgcc ctggcccacc 180
 ctcgtgacca cttcggcta cggcctgcaa tgcttcgcc gctaccccga ccacatgaag 240
 ctgcacgact tcttcaagtc cgccatgccc gaaggctacg tccaggagcg caccatcttc 300
 ttcaaggacg acggcaacta caagaccgc gccgaggtga agttcgaggg cgacaccctg 360
 gtgaaccgca tcgagctgaa gggcatcgac ttcaaggagg acggcaacat cctggggcac 420
 aagctggagt acaactacaa cagecacaac gtctatatac tggccgacaa gcagaagaac 480
 ggcatcaagg tgaactcaa gatecgcacc aacatcgagg acggcagcgt gcagctcgcc 540
 gaccactacc agcagaacac ccccatcggc gacggccccg tgctgctgcc cgacaaccac 600
 tacctgagct accagtccgc cctgagcaaa gaccccaacg agaagcgcga tccatgggtc 660
 ctgctggagt tcgtgaccgc cgccgggatc actctcgcca tggacgagct gtacaagtaa 720

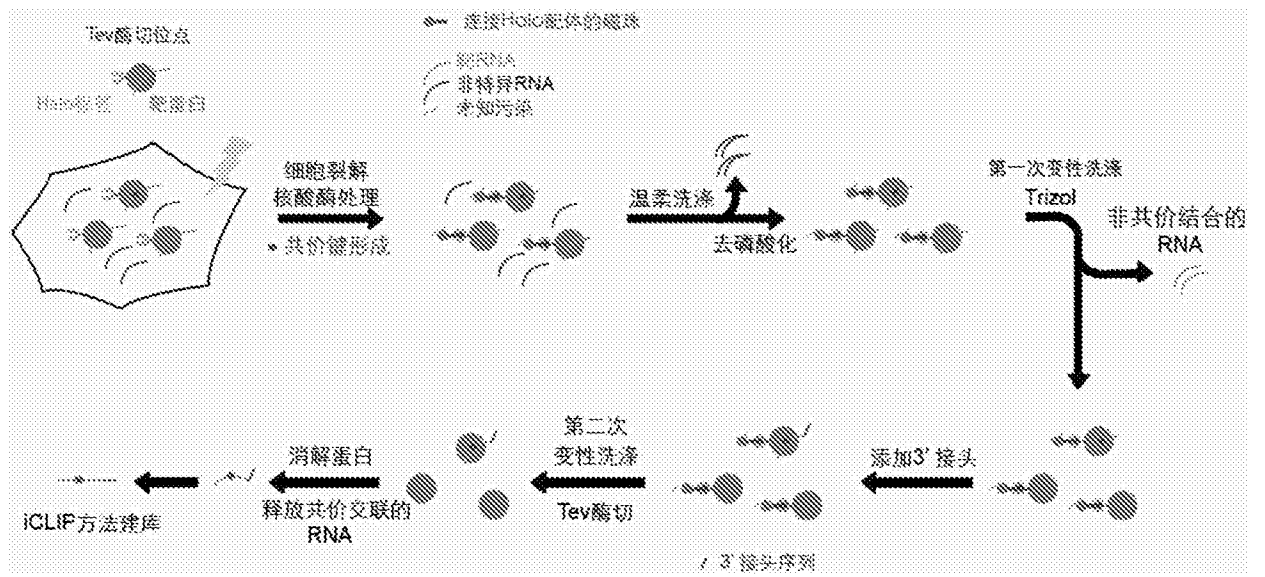


图1

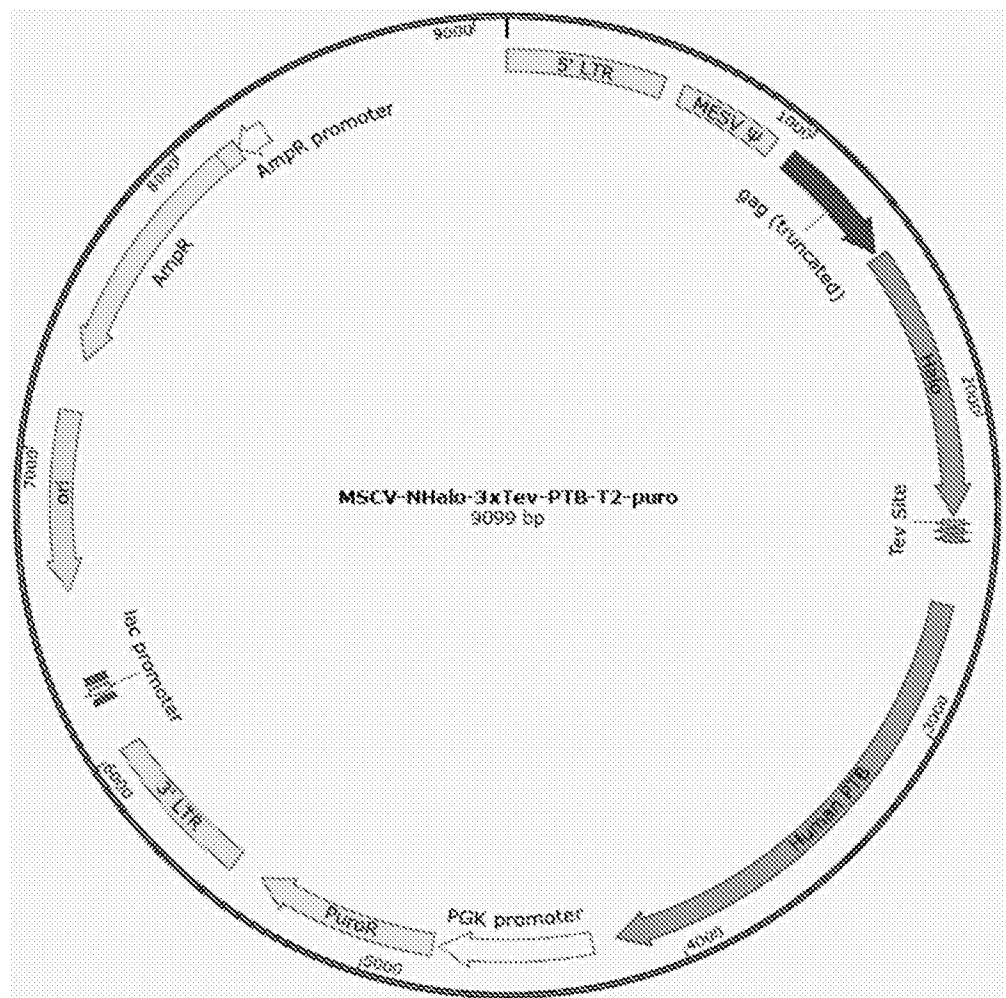


图2

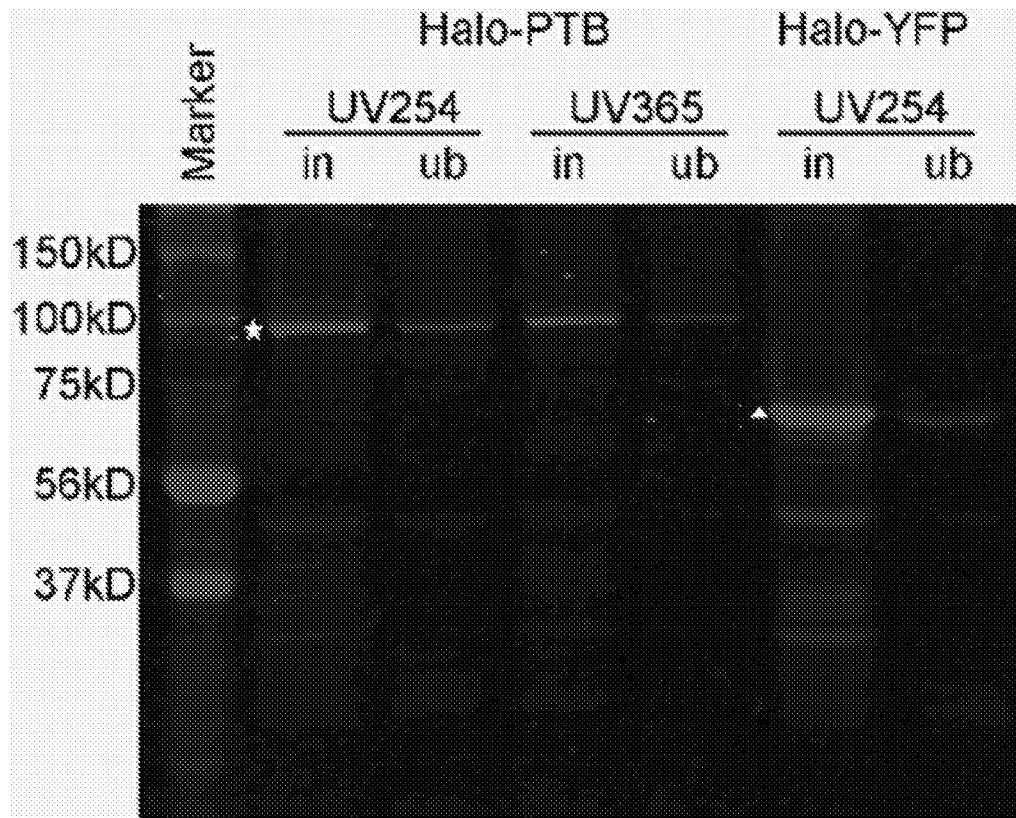


图3

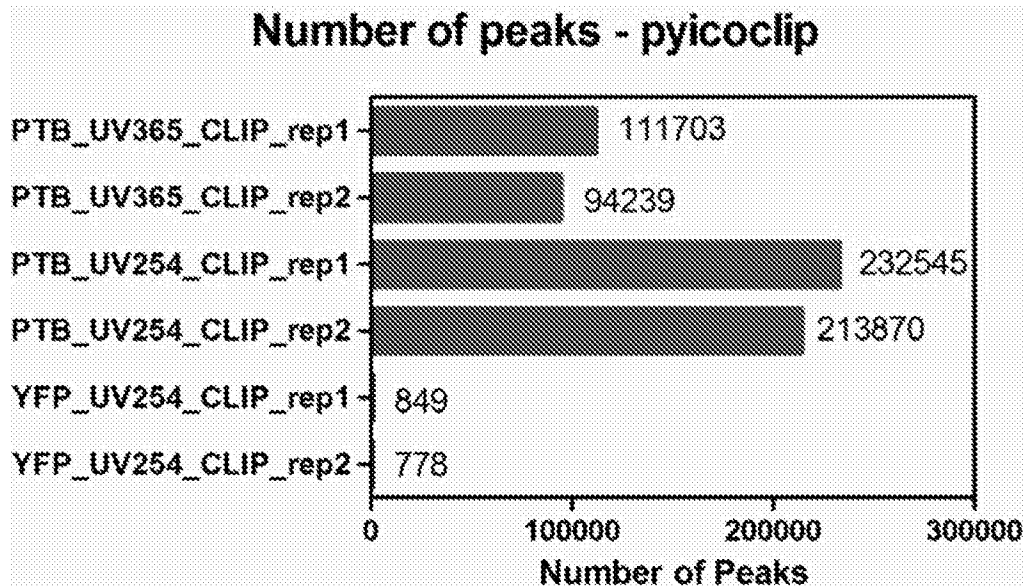


图4

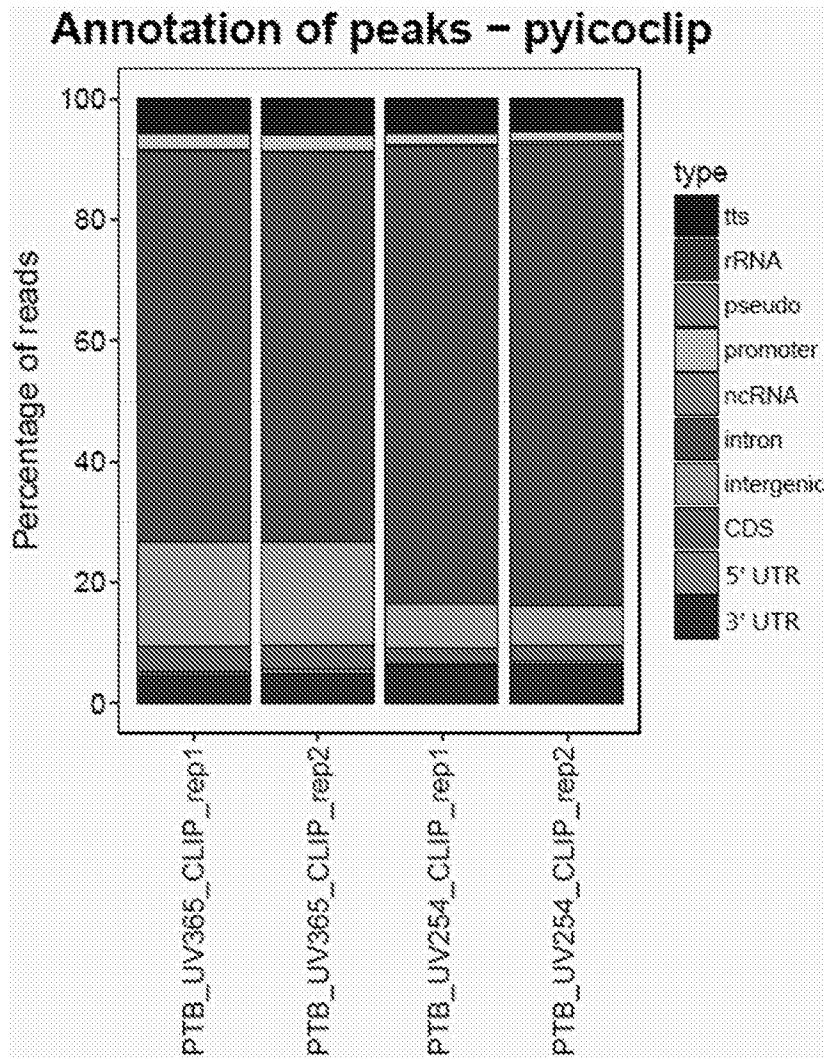


图5

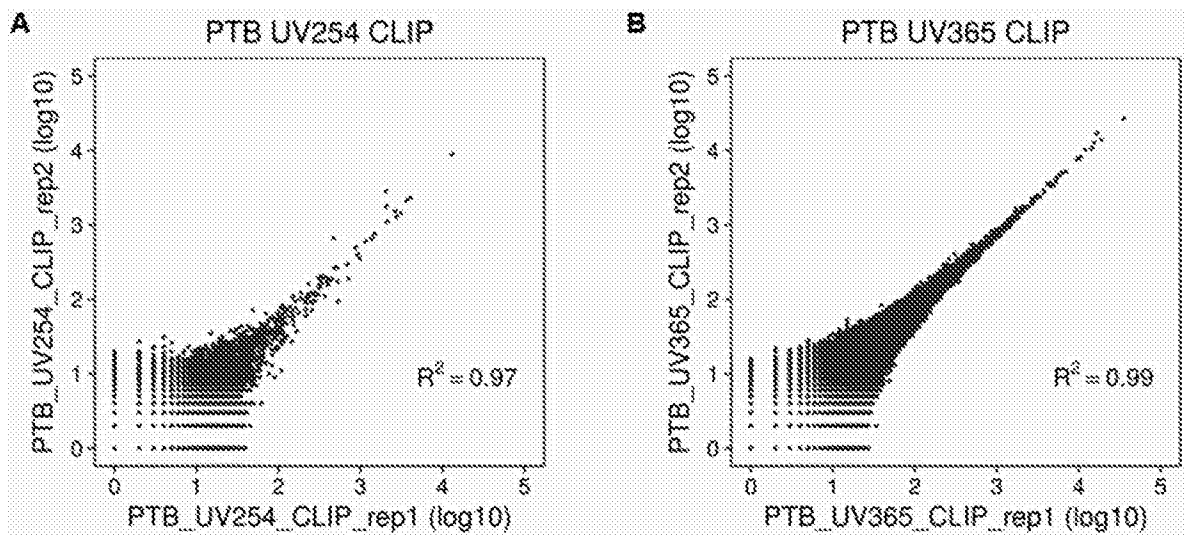


图6

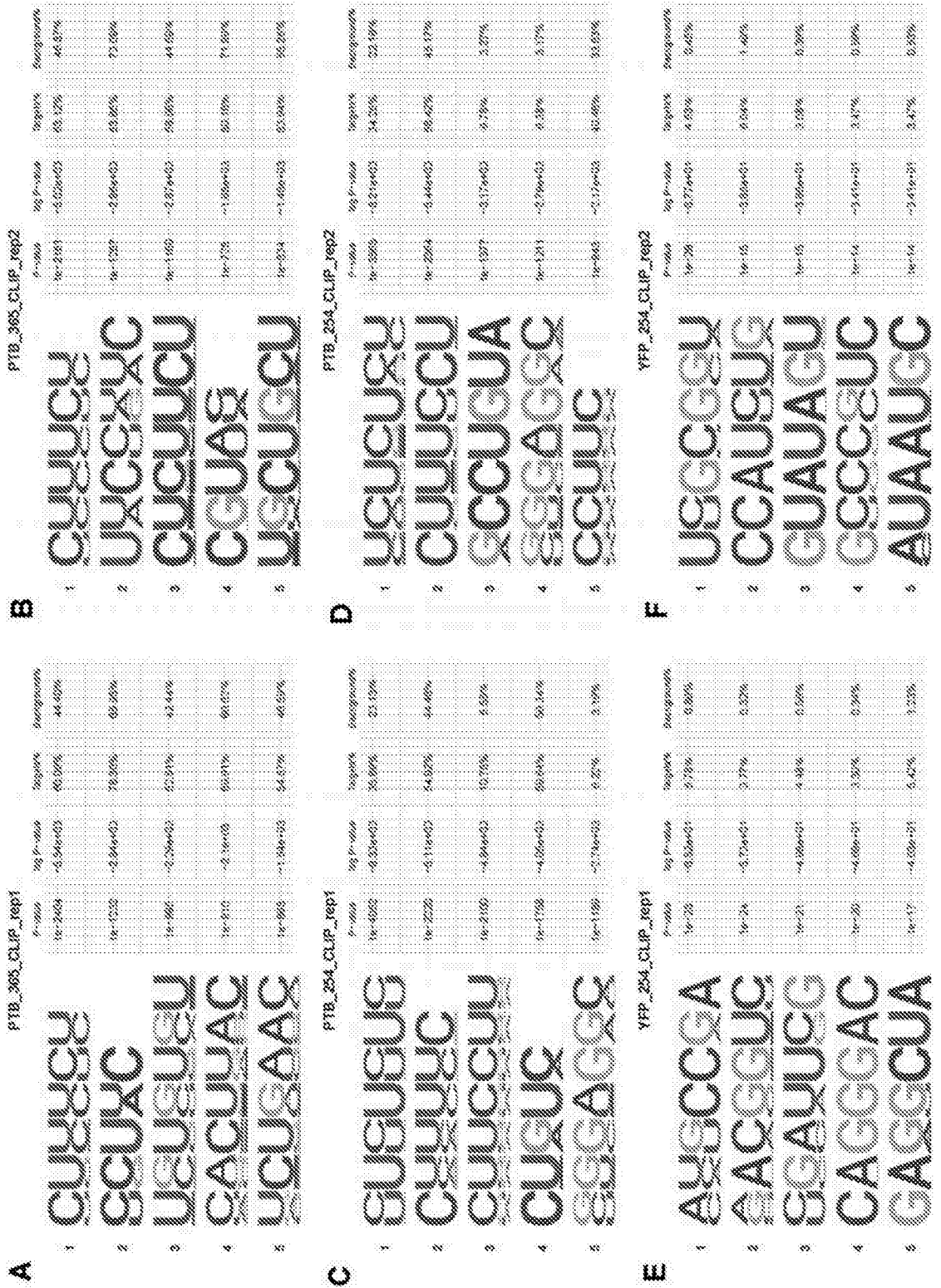


图7