(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 111406114 A
(43)申请公布日 2020.07.10

(21)申请号 201880049398.4

(22)申请日 2018.05.25

(30)优先权数据
62/512,144 2017.05.29 US

(85)PCT国际申请进入国家阶段日
2020.01.22

(86)PCT国际申请的申请数据
PCT/US2018/034689 2018.05.25

(87)PCT国际申请的公布数据
WO2018/222548 EN 2018.12.06

(71)申请人 哈佛学院董事及会员团体
地址 美国马萨诸塞州

(72)发明人 谢晓亮 A·R·查普曼 D·F·李

(74)专利代理机构 上海专利商标事务所有限公司 31100
代理人 余颖 陈扬扬

(51)Int.Cl.
C12Q 1/6844(2018.01)
C12N 15/10(2006.01)

权利要求书2页 说明书20页 附图7页

(54)发明名称

扩增单个细胞转录组的方法

(57)摘要

本公开提供了用于使用逆转录和基于多次退火和成环的扩增循环的组合扩增RNA的方法。使用引物,从而使所得扩增子包含第一细胞特异性条码序列,第二细胞特异性条码序列和独特分子标识符条码序列。

CN 111406114 A

1.一种扩增RNA模板链的方法,其包括:

使用逆转录酶和逆转录引物序列将所述RNA模板链逆转录成cDNA模板链,所述逆转录引物序列具有与所述RNA模板链的5'多聚(A)序列互补的3'多聚(T)序列,其中所述逆转录引物序列还包含5'自退火序列,条码引物退火位点,具有4-12个核苷酸的第一细胞特异性条码序列和具有10-30个核苷酸的第一独特分子标识符条码序列,其中所述cDNA模板链在其5'包含逆转录引物序列且所述cDNA模板链与所述RNA链杂交,

用酶消化过量的逆转录引物序列,

降解所述RNA链以产生作为单链的cDNA模板链,

使所述逆转录酶失活,

使所述酶失活,

(a)使用DNA聚合酶和延伸引物生成包含逆转录引物序列的所述cDNA模板链的互补链,所述延伸引物在引物的5'端包含所述自退火序列,其中所述互补链在5'端包含所述自退火序列并在3'端包含其互补序列,

(b)使所述cDNA模板链由所述互补链变性,并通过将3'端的所述自退火序列和5'端的其互补序列退火来使所述互补链成环,从而抑制所述互补链的扩增,

重复步骤(a)和(b)数次以由所述cDNA模板链生成数个成环互补链,

使所述数个成环互补链变性,并使用包含所述自退火序列的扩增引物扩增经变性的互补链,以产生包含所述逆转录引物序列的双链扩增子,

使所述双链扩增子变性,并使用下述内容重复扩增经变性的扩增子数次:(1)外部条码引物,其具有与所述条码引物退火位点互补的3'序列,其中所述外部条码序列还包含5'自退火序列,测序引发序列和具有4-12个核苷酸的第二细胞特异性条码序列,和(2)包含3'自退火序列的引物,以产生具有第一细胞特异性序列、第二细胞特异性条码序列和第一独特分子标识符条码序列的所得双链扩增子。

2.如权利要求1所述的方法,其中,所述RNA是信使RNA、转移RNA、核糖体RNA、长链非编码RNA或小干扰RNA。

3.如权利要求1所述的方法,其中,所述RNA来自单个细胞。

4.如权利要求1所述的方法,其中,所述RNA来自异质细胞群内的单个细胞。

5.如权利要求1所述的方法,其中,所述RNA来自单个产前细胞。

6.如权利要求1所述的方法,其中,所述RNA来自单个癌细胞。

7.如权利要求1所述的方法,其中,所述RNA来自单个循环肿瘤细胞。

8.如权利要求1所述的方法,其中,所述逆转录酶是SuperScript II、III或IV,M-MLV逆转录酶,Maxima逆转录酶,Protoscript逆转录酶或Thermoscript逆转录酶。

9.如权利要求1所述的方法,其中,所述3'多聚(T)序列包含10-30个T核苷酸。

10.如权利要求1所述的方法,其中,所述自退火序列是GAT5或GAT1。

11.如权利要求1所述的方法,其中,所述条码引物退火位点是RT3、Read1SP或Read2SP。

12.如权利要求1所述的方法,其中,所述酶是具有链置换活性或者5'-3'外切核酸酶活性的聚合酶。

13.如权利要求1所述的方法,其中,所述酶是Φ29聚合酶,Bst聚合酶,Pyrophage 3173,Vent聚合酶,Deep Vent聚合酶,TOPO Taq DNA聚合酶,Taq聚合酶,T7聚合酶,Vent(外

切-)聚合酶,Deep Vent(外切-)聚合酶,9°Nm聚合酶,DNA聚合酶I的Klenow片段,MMLV逆转录酶,AMV逆转录酶,HIV逆转录酶,缺少3'-5'外切核酸酶活性的T7噬菌体DNA聚合酶的突变形式,Taq聚合酶,Bst DNA聚合酶(全长),大肠杆菌DNA聚合酶,LongAmp Taq聚合酶,OneTaq DNA聚合酶,Q5,Phusion或Kapa HiFi。

14.如权利要求1所述的方法,其中,所述RNA链在75℃-85℃的温度下降解。

15.如权利要求1所述的方法,其中,所述逆转录酶和所述酶在75℃-85℃的温度下失活。

16.如权利要求1所述的方法,其中,所述延伸引物在0℃-10℃的温度下退火至cDNA模板链。

17.如权利要求1所述的方法,其中,所述互补链在10℃-65℃的温度下生成。

18.如权利要求1所述的方法,其中,所述互补链的成环发生在55℃-60℃的温度下。

19.如权利要求1所述的方法,其中,步骤(a)和(b)重复7-12次。

20.如权利要求1所述的方法,其中,扩增经变性的互补链用聚合酶链式反应进行。

21.如权利要求1所述的方法,其中,扩增经变性的互补链用15-20个聚合酶链式反应循环进行。

22.如权利要求1所述的方法,其中,扩增经变性的扩增子用聚合酶链式反应进行。

23.如权利要求1所述的方法,其中,用3-7个PCR循环重复扩增经变性的扩增子。

24.如权利要求1所述的方法,其中,对所得双链扩增子进行加工用于测序。

25.如权利要求1所述的方法,其中,所述第一独特分子标识符条码序列包含半随机序列模式。

26.如权利要求1所述的方法,其中,使用酶消化过量转录引物的步骤包括添加逆转录引物,其具有第二独特分子标识符条码序列,其具有10-30个核苷酸,包含半随机序列模式,并且与所述第一独特分子标识符条码序列不同。

扩增单个细胞转录组的方法

[0001] 相关申请信息

[0002] 本申请要求于2017年5月29日提交的第62/512,144号美国临时申请的优先权,并通过引用将其整体纳入本文用于所有目的。

[0003] 政府权益声明

[0004] 本发明是在国立卫生研究院(National Institutes of Health)的CA174560和CA186693下于政府资助下完成。政府对本发明享有某些权利。

[0005] 背景

发明领域

[0006] 本发明的实施方式一般涉及用于单个细胞信使RNA扩增的方法和组合物,诸如来自单个细胞的信使RNA。

背景技术

[0007] 已知单个细胞RNA测序技术。参见Wen等,Genome Biology(2016)17:17,DOI 10.1186/s13059-016-0941-0;Mortazavi等,Nature Methods DOI:10.1038/nmeth.1226;Chapman等,PLoS ONE 10(3):e0120889,doi:10.1371/journal.pone.0120889(2015);和Sheng等,Nature Methods DOI:10.1038/NMETH.4145(2017)。Tang等(2009)单个细胞的mRNA-Seq全转录组分析(mRNA-Seq whole-transcriptome analysis of a single cell).Nat Methods,6,377-382对于scRNA-seq的首次报道使用多聚-T引物用于cDNA合成,然后进行多聚-A加尾,第二链合成和PCR。后续技术进展包括:添加模板转换以改善RNA回收效率(参见Islam,S.,Kjallquist,U.,Moliner,A.,Zajac,P.,Fan,J.B.,Lonnerberg,P.和Linnarsson,S.(2011)通过高度多重化RNA-seq表征单细胞转录(Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq).Genome Res,21,1160-1167;Picelli,S.,Bjorklund,A.K.,Faridani,O.R.,Sagasser,S.,Winberg,G.和Sandberg,R.(2013)用于在单个细胞中进行灵敏全长转录组概况的Smart-seq2(Smart-seq2 for sensitive full-length transcriptome profiling in single cells).Nat Methods,10,1096-109),细胞特异性条码以允许样品进行多重化(参见Jaitin,D.A.,Kenigsberg,E.,Keren-Shaul,H.,Elefant,N.,Paul,F.,Zaretsky,I.,Mildner,A.,Cohen,N.,Jung,S.,Tanay,A.等.(2014)大规模平行单细胞RNA-seq用于将无标志物分解组织为细胞类型(Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types).Science,343,776-779;Fan,H.C.,Fu,G.K.和Fodor,S.P.(2015)表达概况单个细胞的组合标记用于基因表达细胞计数(Expression profiling.Combinatorial labeling of single cells for gene expression cytometry).Science,347,1258367),经优化的酶促条件(参见Sasagawa,Y.,Nikaido,I.,Hayashi,T.,Danno,H.,Uno,K.D.,Imai,T.和Ueda,H.R.(2013)Quartz-Seq:高度可再生和灵敏的单细胞RNA测序方法,揭示了非遗传性基因表达异质性(Quartz-Seq:a

highly reproducible and sensitive single-cell RNA sequencing method,reveals non-genetic gene-expression heterogeneity).Genome Biol,14,R31),独特分子标识符以标记独特cDNA(参见Islam,S.,Zeisel,A.,Joost,S.,La Manno,G.,Zajac,P.,Kasper,M.,Lonnerberg,P.和Linnarsson,S.(2014)使用独特分子标识符的定量单细胞RNA-seq(Quantitative single-cell RNA-seq with unique molecular identifiers).Nat Methods,11,163-166;Shiroguchi,K.,Jia,T.Z.,Sims,P.A.和Xie,X.S.(2012)数字RNA测序通过优化的单分子条码最大程度地降低了序列依赖性偏倚和扩增噪声(Digital RNA sequencing minimizes sequence-dependent bias and amplification noise with optimized single-molecule barcodes).Proc Natl Acad Sci U S A,109,1347-1352),体外转录cDNA以减少扩增偏倚(参见Hashimshony,T.,Senderovich,N.,Avital,G.,Klochendler,A.,de Leeuw,Y.,Anavy,L.,Gennert,D.,Li,S.,Livak,K.J.,Rozenblatt-Rosen,O.等.(2016)CEL-Seq2:灵敏的高度多重化单细胞RNA-Seq(CEL-Seq2:sensitive highly-multiplexed single-cell RNA-Seq)CEL-Seq2:sensitive highly-multiplexed single-cell RNA-Seq.Genome Biol,17,77)和使用微流体装置的自动化(Zheng,G.X.,Terry,J.M.,Belgrader,P.,Ryvkin,P.,Bent,Z.W.,Wilson,R.,Ziraldo,S.B.,Wheeler,T.D.,McDermott,G.P.,Zhu,J.等.(2017)单个细胞的大规模平行数字转概况(Massively parallel digital transcriptional profiling of single cells).Nat Commun,8,14049;Macosko,E.Z.,Basu,A.,Satija,R.,Nemesh,J.,Shekhar,K.,Goldman,M.,Tirosh,I.,Bialas,A.R.,Kamitaki,N.,Martersteck,E.M.等.(2015)使用纳升液滴对单个细胞进行高度平行全基因组表达概况(Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets).Cell,161,1202-1214;Klein,A.M.,Mazutis,L.,Akartuna,I.,Tallapragada,N.,Veres,A.,Li,V.,Peshkin,L.,Weitz,D.A.和Kirschner,M.W.(2015)液滴条码化用于胚胎干细胞的单细胞转录组学(Droplet barcoding for single-cell transcriptomics applied to embryonic stemcells).Cell,161,1187-1201)。

[0008] 虽然存在这些进展,但是这些方法的一个共有限制是低RNA检测效率,这通常为20%或更低(参见Ziegenhain,C.,Vieth,B.,Parekh,S.,Reinius,B.,Guillaumet-Adkins,A.,Smets,M.,Leonhardt,H.,Heyn,H.,Hellmann,I.和Enard,W.(2017)比较分析单细胞RNA测序方法(Comparative Analysis of Single-Cell RNA Sequencing Methods.)Mol Cell,65,631-643e634;Liu,S.和Trapnell,C.(2016)单细胞转录组测序:最近的进展和剩余的挑战(Single-cell transcriptome sequencing:recent advances and remaining challenges).F1000Res,5)。由于采样噪声,这增加了RNA定量的不确定性,并导致低表达的转录本丢失。另一限制是,尽管添加了UMI,但是RNA定量由于UMI错误计数仍然是不准确的。出现这一现象的原因是包含UMI的逆转录引物在cDNA扩增之前可能不被完全去除,且现有方法无法测量去除效率。最终,对于使用PCR扩增cDNA的方法,指数扩增过程可以导致扩增偏倚。总之,这些问题限制了现有scRNA-seq方法的完整性、准确度和成本效益。因此,存在对于没有一个或多个缺点的情况下扩增(如由单个细胞或一小群细胞扩增)少量RNA的其它方法的需求。

发明内容

[0009] 本公开的实施方式是关于扩增RNA的方法，如少量RNA或有限量RNA，如获自单个细胞或相同细胞类型的多个细胞或来自获自个体或底物的组织、液体或血液样品的RNA。本文所述方法包括使用所述引物逆转录RNA以生成cDNA，然后根据本文所述基于多次退火和成环的扩增循环(multiple annealing and looping based amplification cycle)扩增cDNA(参见述于Zong,C.,Lu,S.,Chapman,A.R.和Xie,X.S.(2012),基因组检测单个人细胞的单核苷酸和拷贝数变异(Genome-wide detection of single-nucleotide and copy-number variations of a single human cell),Science 338,1622-1626中的由单个细胞扩增基因组DNA的方法，其描述了基于多次退火和成环的扩增循环(MALBAC)，通过引用其全部内容纳入本文)以生成具有本文所述第一细胞特异性条码、第二细胞特异性条码和独特分子标识符条码序列的双链扩增子。根据本公开的某些方面，本文所述的方法可以在具有可编程的热循环的单管中进行。

[0010] 本文所述用于单细胞RNA扩增的方法可以称之为针对数字转录组的基于多次退火和成环的扩增循环(Multiple Annealing and Looping Based Amplification Cycles for Digital Transcriptomics,MALBAC-DT)，它克服了其他方法的缺点。由于在cDNA扩增期间使用随机引物来退火cDNA，改善了捕获效率，本文所述的MALBAC-DT方法具有较高的RNA检测效率。此外，准线性cDNA扩增减少了扩增偏倚以及因此的转录本丢失(transcript dropout)。此外，由于UMI设计，本文所述MALBAC-DT方法具有较高的准确性。另一方面进一步包括测量cDNA扩增前逆转录引物降解的效率。

[0011] 根据一个方面，使用逆转录引物，其包含与RNA模板链的5'多聚(A)序列互补的3'多聚(T)序列。逆转录酶引物进一步包含5'自退火序列，条码引物退火位点，第一细胞特异性条码序列和第一独特分子标识符条码序列，以产生对应RNA模板的cDNA，其中cDNA还包含逆转录引物。

[0012] 然后，cDNA在第一低温下遇到在引物的5'端具有自退火序列的引物，其中互补链在5'端包含自退火序列并在3'端包含其互补序列，其中引物退火至cDNA。较高温度下的引物延伸在至少一种聚合酶存在的情况下继续，诸如具有5'-3'外切核酸酶活性的一种或多种链置换聚合酶。将延伸产物和cDNA模板分离，然后将混合物置于较低温度下，此时延伸产物的末端退火至自身以形成环，从而产生不能进一步延伸或扩增的延伸产物。然后，cDNA模板以上述方式再一次延伸，然后使延伸产物成环。重复该过程数次以提供成环延伸产物的群体。然后将成环延伸产物去杂交或解链，然后使用包含第二细胞特异性条码序列的引物扩增单链。扩增产生双链扩增子，其包含第一细胞特异性条码序列，第二细胞特异性条码序列和独特分子标识符序列(UMI)，其中UMI具有半随机序列。根据一个方面，进行数个热循环以扩增cDNA并形成成环延伸产物，其抑制延伸产物被进一步延伸或扩增。扩增可以称之为线性扩增或准线性扩增。然后可以使用标准或非标准PCR循环扩增成环延伸产物。某些聚合酶提供了示例性结果。

[0013] 根据某些方面，提供了用于加工至少一个细胞，一个或多个细胞或数个细胞的方法，诸如两个或多个细胞，例如，用于根据本文所述的方法进行RNA扩增。根据示例性的实施方式，分离单个细胞，然后在一流体体积中裂解以获得细胞的RNA。根据示例性的实施方式，可以各自分离多个单个细胞，然后在一流体体积中裂解以获得细胞的RNA，然后可以多重逆

转录并扩增细胞的RNA。

[0014] 本公开的某些实施方式的其他特征和优势将在权利要求中以及以下附图和实施方式的说明下更为显而易见。

附图说明

[0015] 本专利或申请文件包含至少一幅有色附图。本专利或专利申请公开的带彩色附图的副本将根据要求，在支付所需的费用之后由政府机关提供。结合附图，通过以下示例性实施方式的详述能够更全面地理解本发明实施方式的上述和其他特征和其他优点，其中：

[0016] 图1示例性描述了由mRNA转录本制备cDNA的方法。具有UMI模式‘A’(UMI_A)和细胞条码C_n的包含多聚(T)的引物($RT-A_n$)退火至靶mRNA的多聚(A)区域。与SuperScript IV，一种逆转录酶孵育催化cDNA合成。然后添加外切核酸酶I以消化任何剩余的RT引物并防止它们在cDNA扩增期间引发。引物$RT-B_n$的添加能够测量外切核酸酶降解的效率，因为不完整的消化将会产生UMI_A和UMI_B cDNA扩增产物的混合物，所述引物$RT-B_n$具有UMI_B模式而非UMI_A模式。最终，将混合物在80℃孵育以降解RNA并使外切核酸酶I和Superscript IV热失活。

[0017] 图2示例性描述了使用基于多次退火和成环的扩增循环(MALBAC)扩增cDNA的方法。将包含GAT5序列和7核苷酸随机序列的引物(GAT5-7N)随机退火至cDNA。引物还包含B1间隔子序列。与3’->5’外切核酸酶缺陷型Deep Vent，一种DNA聚合酶孵育催化第二链合成。使这些链变性，然后冷却，导致第二链形成稳定的发夹环结构，防止进一步扩增。重复9次以产生多个环并以准线性方式扩增cDNA。这些准线性步骤后，将环变性并使用GAT5-B1引物通过17个PCR循环扩增。最终，MALBAC后，添加外部条码引物，并使用外部条码和GAT5-B1引物进行另外5个PCR循环。

[0018] 图3示例性描述了使用称为标签化(tagmentation)的基于转座子的方法的文库制备方案。使用高活性Tn5转座酶(诸如来自Nextera DNA文库制备盒)的标签化产生多个产物，所需产物具有侧接cDNA的Read1SP和条码序列。于72℃使用DNA聚合酶进行缺口修复后，Illumina测序相容文库通过使用读数1索引衔接子引物(亿明达公司(Illumina)称之为S5XX)和读数2索引衔接子引物的5个PCR循环产生。索引1/索引2是Illumina测序索引，而P5/P7是流动池(flowcell)退火衔接子。

[0019] 图4A描述了HEK293T培养物大约700个经测序的细胞内12,000个持续检测到的基因的mRNA的相关性矩阵(上图)。针对HEK293T数据集，使用t-随机邻居嵌入算法(t-stochastic neighbor embedding algorithm，t-SNE)，图4B描述了基因的聚类(左图)而图4C描述了细胞的聚类(右图)。在图4B的基因聚类图中，各基因簇对应相关性矩阵中的方块。在基因聚类图中，各点是12,000个基因中的一个，而各簇对应相关性矩阵中的方块。在图4C的细胞聚类图中，各点是约700个HEK细胞中的一个，并且不存在可分解的簇。

[0020] 图5描述了HEK293T培养物内12,000个持续检测到的基因中3000个的mRNA相关性矩阵的数据(上图)。图5描述了U-2OS培养物内12,000个持续检测到的基因中3000个的mRNA相关性矩阵的数据(下图)。颜色强度与两种基因之间的皮尔逊相关系数相关。对角线上的每个方块表示观测到强相关性的基因簇。基因簇是可能具有共同转录调节和生物学功能的基因的组。将两个细胞系之间所共有的细胞簇中的两个标记为细胞循环和蛋白质合成簇。

[0021] 图6突出显示了图5中标记的蛋白质合成簇。针对tRNA合成、氨基酸合成、氨基酸转

运以及翻译启动的控制中所涉及的那些富集该簇中的基因,所有这些在蛋白质合成过程中都十分重要。因此,关联的基因簇具有相关的生物学功能和转录调节。

[0022] 图7比较了U-2OS和HEK293T细胞系之间的关联模块。涉及通用细胞功能诸如细胞周期进程和蛋白质合成的一些模块在两种细胞系中是共同的,但是其他模块诸如p53和骨胞外基质模块对于一种细胞类型具有特异性。该细胞类型特异性并不一定反映在差异性表达中。尽管两种细胞系之间存在差异性表达,仍然保留一些模块,同时尽管并未经差异性表达,其他模块并不存在。

具体实施方式

[0023] 除非另有说明,某些实施方式的实践或某些实施方式的特征可以采用分子生物学、微生物学、重组DNA中的常规技术,这些常规技术为本领域普通技术人员所知。这些技术在文献中已有充分描述。参见,例如,Sambrook,Fritsch,和Maniatis,《分子克隆:实验室手册(MOLECULAR CLONING:A LABORATORY MANUAL)》,第二版(1989),《寡核苷酸合成(OLIGONUCLEOTIDE SYNTHESIS)》(M.J.Gait编著,1984),《动物细胞培养(ANIMAL CELL CULTURE)》(R.I.Freshney编著,1987),《酶学方法(METHODS IN ENZYMOLOGY)》丛书(学术出版社有限公司(Academic Press,Inc.));《哺乳动物细胞的基因转移载体(GENE TRANSFER VECTORS FOR MAMMALIAN CELLS)》(J.M.Miller和M.P.Calos编著.1987),《免疫学实验手册(HANDBOOK OF EXPERIMENTAL IMMUNOLOGY)》,(D.M.Weir和C.C.Blackwell编著),《新编分子生物学实验指南(CURRENT PROTOCOLS IN MOLECULAR BIOLOGY)》(F.M.Ausubel,R.Brent,R.E.Kingston,D.D.Moore,J.G.Siedman,J.A.Smith,和K.Struhl编著,1987),《新编免疫学实验指南(CURRENT PROTOCOLS IN IMMUNOLOGY)》(J.E.coligan,A.M.Kruisbeek,D.H.Margulies,E.M.Shevach和W.Strober编著,1991);《免疫学年鉴(ANNUAL REVIEW OF IMMUNOLOGY)》;以及如《免疫学进展(ADVANCES IN IMMUNOLOGY)》等期刊中的专著。本文上下文中提及的所有专利、专利申请和出版物均以参考的方式用全文纳入本文。

[0024] 本文所用的核酸化学、生物化学、遗传学和分子生物学的术语和符号遵循本领域的标准论述和文本中的术语和符号,例如,Kornberg和Baker,DNA Replication(《DNA复制》),第二版(W.H.弗里曼出版社(W.H.Freeman),纽约,1992);Lehninger,Biochemistry(《生物化学》),第二版(沃斯出版社(Worth Publishers),纽约,1975);Strachan和Read,Human Molecular Genetics(《人类分子遗传学》),第二版(WL出版社(Wiley-Liss),纽约,1999);Eckstein编,Oligonucleotides andanalogs:A Practical Approach(《寡核苷酸和类似物:实践方法》)(牛津大学出版社(Oxford University Press),纽约,1991);Gait编,Oligonucleotide Synthesis:A Practical Approach(《寡核苷酸合成:实践方法》)(IRL出版社,牛津,1984);等。

[0025] 本发明部分基于发现由细胞或细胞的集合扩增一个或多个或多种靶RNA序列的方法,其中所得扩增子包含第一细胞特异性条码序列,第二细胞特异性条码序列和独特细胞标识符条码序列。扩增子可以经加工成文库,诸如用于测序。以此方式,可以在单细胞RNA测序的方法中确定一个或多个或多种靶RNA序列,所述单细胞RNA测序的方法被用于表征异质群体内个体细胞的转录组。

[0026] 本公开的一些方面利用长度为10-30个核苷酸的独特分子标识符条码序列(UMI),

示例性长度为20个核苷酸。这样的独特分子标识符条码序列长度减少两个转录本具有相同UMI的机会。因此，本公开的一些方面涉及针对各RNA转录本或其相关cDNA关联不同独特分子标识符条码序列。以此方式，各RNA转录本具有其独特相关的独特分子标识符条码序列。以此方式，数个RNA转录本内的各RNA转录本具有不同于数个RNA转录本其他成员的独特分子标识符条码序列。此外，这样的独特分子标识符条码序列长度能够使得因UMI扩增或测序中错误而产生的假UMI序列(其与真UMI通常仅有一个或两个核苷酸差异)可以被区分，因为UMI序列相距甚远，即UMI之间的汉明(Hamming)距离足以减少将测序误读数(misread)误认为是不同UMI的机会。

[0027] 本公开利用具有本文所述半随机模式的UMI(UMI$_A$和UMI$_B$)。对UMI使用半随机模式允许测序或扩增错误通过对落在模式之外的碱基进行计数来测量，从而提供测序错误率的经验性测量值。具体地，由于半随机模式，UMI插入或缺失错误是显而易见的。知晓错误率对于理解UMI的可靠性十分重要。

[0028] 根据一个方面，UMI$_A$和UMI$_B$均为半随机模式的10-30个碱基对序列，诸如20个碱基对序列。UMI$_A$的模式是[(HBDV)$_5$]，其中H＝不是G，B＝不是A，D＝不是C，和V＝不是T。UMI$_B$的模式是[(VDBH)$_5$]。应当理解的是，可以设计其他半随机模式。该半随机模式提供两个优势。第一，当碱基落在预期的模式之外时，可以检测UMI中的扩增或测序错误，允许经验性测量错误率。第二，因为UMI$_B$可以与UMI$_A$相区分，这允许由纳入UMI$_A$的读数相较于纳入UMI$_B$的读数的比例来确定外切核酸酶降解效率。

[0029] 本公开的一些方面涉及在本文所述逆转录方法期间测量逆转录引物(具有UMI$_A$模式的RT-A)的降解率的方法。外切核酸酶消化通过预防过量逆转录引物结合DNA改善定量准确性。这些引物将以其他方式将多个UMI连接至相同mRNA转录本的拷贝并导致计数过度(overcounting)。根据该方法，逆转录之后和引物降解步骤期间，将具有区别于RT期间所用RT-A引物模式的不同UMI模式的逆转录引物(具有UMI$_B$模式的RT-B)添加到混合物。这允许测量RT引物降解效率，由通过包含UMI$_A$或UMI$_B$模式的产物的读数的最终比例确定。

[0030] 本公开的一些方面涉及使用两种细胞特异性条码来标记源自各个体细胞或样品的RNA。两种条码的使用增加将RNA与细胞或样品关联的可能的条码组合的总数(超过使用单一条码)。两种条码多重化允许由待汇集在一起用于文库制备的多个细胞扩增cDNA。例如，引物纳入分别具有48和48个可能的序列的两种不同的条码序列C$_n$和G$_m$(2304种组合)。这使得需要完成的个体文库制备的数量最小化并降低试剂成本。可能的条码组合与引物的数量成二次方比例。这区别于仅使用一种引物并且其中每个条码需要单独的引物的条码化方案。

[0031] 本公开的一些方面涉及制备与样品中RNA相关联的扩增子的方法，其中，所述扩增子被设计成与标准文库制备试剂盒具有相容性。最终扩增产物的设计对于使用本文所述标准试剂盒的文库制备具有相容性，这不同于单个细胞多重化扩增方法，其需要定制的文库制备方案和定制的测序引物。

[0032] 本公开提供了由诸如来自小样品、单个细胞或小细胞群的RNA进行cDNA合成的方法。然后可以使用基于多次退火和成环的扩增循环来扩增cDNA以产生扩增子，所述扩增子包含第一细胞特异性条码序列，第二细胞特异性条码序列和独特分子标识符条码序列。然后可以对扩增子进行测序，诸如通过加工成测序文库。

[0033] 根据一个方面，实施方式提供了可以在单管或在微滴定板中(例如)以高通量方式进行的三步过程。第一步涉及使用引物、逆转录酶、核酸酶和本文所述或以其他方式为本领域技术人员所知的其他合适的试剂和介质将RNA逆转录成cDNA以产生连接引物序列的cDNA。在第二步中，使用线性或准线性扩增方法扩增cDNA以产生在各末端具有引物序列的成环延伸产物。在第三步骤中，扩增成环延伸产物，例如，使用PCR引物，本文所述或为本领域技术人员已知的试剂和条件，以产生双链扩增子，所述双链扩增子具有第一细胞特异性条码序列、第二细胞特异性条码序列和独特分子标识符条码序列。反应混合物中的cDNA样品通过至少一种DNA聚合酶进行延伸或扩增，其中引物退火至DNA以允许DNA聚合酶由引物的3'端合成互补DNA链以产生DNA产物。如果需要，通过DNA聚合酶进行DNA扩增的步骤使DNA产物变性；将引物退火至DNA以形成DNA-引物杂交体；并在核碱基存在的情况下孵育DNA-引物杂交体以允许DNA聚合酶延伸引物并合成DNA产物。

[0034] 根据一个方面，用于逆转录、延伸或扩增的反应混合物形成单链核酸分子/引物混合物，其是包含至少一种单链核酸分子的混合物，其中至少一种引物与所述单链核酸分子中的区域杂交，如本文所述。在具体实施方式中，多个引物与单链核酸分子的多个位置杂交。在其他具体实施方式中，混合物包含多个具有与之杂交的多个简并引物的单链核酸分子。在另一些具体实施方式中，单链核酸分子是cDNA或RNA。

[0035] 对于扩增，反应混合物经历多个热循环。在具体热循环中，反应混合物经历也称为退火温度的第一温度第一时间段以允许将引物充分退火至cDNA序列。根据该方面，引物在第一步骤中以低于约30℃的温度退火至cDNA序列，诸如约0℃-约10℃。然后，反应混合物经历也称为扩增温度的第二温度第二时间段以允许扩增cDNA序列。根据该方面，在第二步骤中以高于约10℃的温度扩增cDNA序列，诸如约10℃-约65℃。本领域技术人员将理解的是，扩增发生的温度将取决于所用特定聚合酶。例如，Φ29聚合酶在约30℃下具有完全活性，而Bst聚合酶和pyrophage 3173聚合酶(外切-)在约62℃下具有完全活性。然后，双链DNA在也称为解链温度的第三温度下解链，持续第三时间段，以提供可以用作扩增模板的单链DNA扩增子。根据该方面，双链DNA在第三步骤中以高于约90℃的温度去杂交成单链DNA，诸如约90℃-约100℃。

[0036] 根据一个方面，可以在也称为成环温度、约55℃-约60℃的第四温度下进行在各末端具有自退火序列的延伸产物的成环，只要延伸产物的自退火末端一起退火以形成环。示例性的温度为约58℃。

[0037] 当反应混合物处于解链温度以产生用于进一步加工、扩增或测序的扩增子时，终止最终扩增循环。根据该方面，如果处于充足的量，可以进一步加工扩增子，用于本文所述测序。根据另一方面，可以进一步扩增扩增子，例如，使用采用本领域技术人员已知缓冲液、引物和聚合酶的标准PCR程序。根据另一方面，如果处于充足的量，可以对扩增子使用本领域技术人员已知的高通量测序方法进行测序。

[0038] 根据某些方面，待扩增的RNA首先经变性，这通过加热反应混合物至约65℃-约85℃，并且示例性地为约72℃，持续约10秒-约5分钟，并且示例性地为3分钟。该步骤中，引物可以存在于反应混合物中。或者，可以在热变性之前或在变性步骤期间的任何时间或热变性步骤后添加引物到包含待扩增的RNA样品的反应混合物中。

[0039] 然后冷却反应混合物并将引物退火。反应混合物的温度低于允许引物退火至单链

RNA的温度。引物的退火温度应当为约0℃-约30℃，示例性的为约0℃-约10℃，或约4℃，持续约10秒-5分钟的时间段。然后，将反应温度增加到特定逆转录被激活并开始合成cDNA的温度。不同的逆转录酶可能在不同的温度下起作用，因此循环可以升高或增加温度，因此可以连续激活逆转录酶以开始合成cDNA。整个孵育期可以为约2分钟-约15分钟，更优选约10分钟。应当理解的是，逆转录步骤的温度、孵育期和升高次数在不显著改变cDNA产生效率的情况下可能与本文所提供的值不同。本领域技术人员基于本公开将理解，参数可以是不同的。反应条件和参数的较小变化包括在本公开的范围之内。

[0040] 将第一组反应中待扩增的cDNA加热至约70℃-约90℃，并且示例性地加热至约80℃，持续约10秒-约5分钟，并且示例性地为2分钟，以降解RNA。该步骤中，引物可以存在于反应混合物中。或者，RNA降解后，可以添加引物至包含cDNA样品的应混合物。

[0041] 对于成环延伸产物的扩增，增加反应混合物的温度以使成环延伸产物变性成单链形式。该温度低于允许引物退火至cDNA的温度。引物的退火温度为约0℃-约30℃，示例性的为约0℃-约10℃，持续约10秒-约5分钟的时间段。然后，将反应温度增加到特定DNA聚合酶变得具有活性并开始合成DNA的温度。不同的DNA聚合酶可能在不同的温度下起作用，因此循环可以升高或增加温度，因此可以连续激活不同的DNA聚合酶以开始合成DNA。整个孵育期可以为约2分钟-约7分钟，更优选约5分钟。

[0042] 应当理解的是，DNA扩增步骤的温度、孵育期和升高次数在不显著改变DNA扩增效率的情况下可能与本文所提供的值不同。本领域技术人员基于本公开将理解，参数可以是不同的。反应条件和参数的较小变化包括在本公开的范围之内。

[0043] 然后，所得扩增子可以经加工用于测序，如本文所述或如本领域技术人员所已知。

[0044] RNA、细胞类型和样品

[0045] 本文所用术语“RNA”可以为本领域技术人员所理解，指在编码，解码，调节和表达基因的各种生物学作用中至关重要的聚合分子。类似于DNA，RNA是核酸。RNA组装成核苷酸链，并且通常以单链存在，自身折叠成二级结构。RNA通常包含核苷酸G、U、A和C以表示氮基鸟嘌呤、尿嘧啶、腺嘌呤和胞嘧啶。RNA的类型包括信使RNA，转移RNA，核糖体RNA，长非编码RNA，小干扰RNA和本领域技术人员已知的其他RNA类型。

[0046] 根据一个方面，RNA是信使RNA或来自待测试的自然或人工来源的其他RNA。在另一优选实施方式中，RNA样品是哺乳动物RNA、植物RNA、酵母RNA、病毒RNA或原核生物RNA。在又一优选实施方案中，RNA样品获自人、牛、猪、羊、马、啮齿动物、禽、鱼、虾、植物、酵母、病毒或细菌。优选RNA样品是来自单个细胞的信使RNA。

[0047] 根据一个方面，RNA来自单个细胞。根据一个方面，RNA来自异质细胞群内的单个细胞。根据一个方面，RNA来自单个产前细胞。根据一个方面，RNA来自单个癌细胞。根据一个方面，RNA来自单个循环肿瘤细胞。

[0048] 术语“分离的RNA”(例如，“分离的mRNA”)指这样的RNA分子，通过重组技术产生时，其基本不含其它细胞物质或培养基，或通过化学方法合成时，基本不含化学物质前体或其它化学物质。

[0049] 根据一个方面，样品可以是体外的。术语“体外”具有其本技术领域公认的含义，例如，涉及纯化的试剂或提取物，例如，细胞提取物。

[0050] 本文所用术语“生物样品”旨在包括但不限于从对象中分离的组织、细胞、生物液

体和分离物，以及对象中存在的组织、细胞和液体。

[0051] 通过本文所述方法加工的RNA可以获自任何有用的来源，例如，人样品。样品可以是来自人的任何样品，如血液、血清、血浆、脑脊液、脸颊刮擦物、乳头抽吸物、活组织检查、精液(可以称为射精液)、尿液、粪便、毛囊、唾液、汗液、免疫沉淀或物理分离的染色质等。在具体的实施方式中，样品包括单个细胞。在具体的实施方式中，样品仅包括单个细胞。

[0052] 在特定实施方式中，扩增自样品的核酸分子提供诊断或预后信息。例如，由样品制备的核酸分子可提供基因组拷贝数和/或序列信息、等位基因变异信息、癌症诊断、产前诊断、亲子信息、疾病诊断、检测、监测和/或治疗信息、序列信息等。

[0053] 本文所用“单个细胞”指一个细胞。可用于本文所述方法中的单个细胞可获自感兴趣组织，或活组织检查，血液样本，或细胞培养物。此外，可以获得来自特定器官、组织、肿瘤、赘生物等的细胞并将其用于本文所述的方法中。此外，通常，来自任何群体的细胞都可以用于所述方法中，如原核或真核单细胞生物体的群体，包括细菌或酵母。使用本领域已知的标准方法，可以获得单个细胞悬浮液，包括例如使用胰蛋白酶或木瓜蛋白酶酶促消化蛋白质，所述蛋白质在组织样品中连接细胞，或在培养中释放贴壁细胞，或在样品中机械地分离细胞。可以将单细胞置于任何合适的反应容器中，在其中可以单独处理单个细胞。例如96孔板，从而将各单个细胞置于单个孔中。

[0054] 本公开范围内的细胞包括任何类型的细胞，对于其中RNA内容物的理解被本领域技术人员认为是有用的。根据本公开的细胞包括任何类型的癌细胞、肝细胞、卵母细胞、胚胎、干细胞、iPS细胞、ES细胞、神经元、红细胞、黑素细胞、星形胶质细胞、生殖细胞、少突胶质细胞、肾细胞等。根据一个方面，本发明的方法使用来自单个细胞的细胞RNA进行。多个细胞包括约2至约1,000,000个细胞，约2至约10个细胞，约2至约100个细胞，约2至约1,000个细胞，约2至约10,000个细胞，约2至约100,000个细胞，约2个至约10个细胞或约2至约5个细胞。

[0055] 用于操作单个细胞的方法是本领域已知的，并且包括荧光激活细胞分选术(FACS)、流式细胞术(Herzenberg.,PNAS USA 76:1453-55 1979)、显微操作以及使用半自动细胞选择器(picker)(例如，来自Stoelting有限公司的Quixell™细胞转移系统)。例如，可以基于通过显微镜观察可检测的特征(如位置、形态或报告基因表达)单独选择个体细胞。此外，还可以使用梯度离心和流式细胞术的组合来增加分离或分选效率。

[0056] 一旦鉴定到所需细胞，使用本领域技术人员已知的方法将细胞裂解以释放包含RNA的细胞内容物。细胞内容物被包含在容器或收集体积内。在本发明的一些方面，细胞内容物(如RNA)可通过裂解细胞从细胞释放。例如，裂解可以通过这样实现，加热细胞，或通过使用洗涤剂或其它化学方法，或通过这些方法的组合。然而，可以使用本领域已知的任何合适的裂解方法。例如，在存在吐温20的情况下，于72℃加热细胞2分钟足以将细胞裂解。或者，可以将细胞于65℃水中加热10分钟(Esumi等.,Neurosci Res 60(4):439-51(2008))；或于70℃在补充有0.5％NP-40的PCR缓冲液II(应用生物系统公司(Applied Biosystems))中90秒(Kurimoto等.,Nucleic Acids Res 34(5):e42(2006))；或裂解可以使用蛋白酶实现，如蛋白酶K，或通过使用离液盐，如异硫氰酸胍(美国公布号2007/0281313)。根据本文所述方法扩增RNA可以直接在细胞裂解物上进行，从而使得可以将反应混合物添加到细胞裂解物。或者，可以使用本领域技术人员已知的方法将细胞裂解物分成两个或更多个体积，如

分到两个或更多个容器、管或区域，其中各体积容器、管或区域包含细胞裂解物的一部分。然后，通过本文所述方法或本领域技术人员已知的方法，可以扩增包含在各容器、管或区域中的RNA。

[0057] 由RNA合成cDNA

[0058] 本文所述方法利用“逆转录酶PCR”(“RT-PCR”)，其是其中起始材料是mRNA的一类PCR。使用逆转录酶将起始mRNA酶促转化为互补DNA或“cDNA”。然后，将cDNA用作PCR反应的模板。

[0059] 根据一个方面，cDNA由RNA产生，其中所得cDNA包含第一细胞特异性条码序列和第一独特分子标识符条码序列。根据一个方面，cDNA由RNA模板合成，诸如获自(即，裂解)单个细胞的mRNA模板。在反应容器中，RNA模板由其二级结构经变性至单链形式。添加逆转录引物序列，其具有与RNA模板链5’多聚(A)序列互补的3’多聚(T)序列。逆转录引物序列进一步包含5’自退火序列，条码引物退火位点，具有4-12个核苷酸的第一细胞特异性条码序列和具有10-30个核苷酸的第一独特分子标识符条码序列。对于给定的mRNA，可以包含10-30个T核苷酸的逆转录引物序列的3’多聚(T)序列与RNA模板链的5’多聚(A)序列杂交。

[0060] 在逆转录酶存在并处于合适的条件和试剂的情况下，RNA模板链经逆转录以产生cDNA模板链，其在cDNA模板链的5’包含逆转录引物序列。cDNA模板链与RNA链杂交。消化过量的逆转录引物序列，诸如使用消化酶。降解RNA链以产生作为单链的cDNA模板链。逆转录酶是失活的。消化酶是失活的。然后，扩增所得cDNA。

[0061] 逆转录酶(RT)是用于由RNA模板生成互补DNA(cDNA)的酶，称为逆转录的过程。根据一个方面，示例性且有用的逆转录酶可以商购和/或为本领域技术人员所知。逆转录酶以称之为逆转录聚合酶链式反应(RT-PCR)的技术将聚合酶链式反应技术应用于RNA。逆转录酶在本公开中用于由mRNA产生cDNA文库。示例性的逆转录酶可以商购，例如SuperScript II、III或IV，M-MLV逆转录酶，Maxima逆转录酶，Protoscript逆转录酶，Thermoscript逆转录酶，或许多其他相容、已知或可商购的逆转录酶。

[0062] 用于消化引物的酶为本领域技术人员所熟知并且是可商购的。示例性的消化酶包括外切核酸酶I，外切核酸酶I与虾碱性磷酸酶，外切核酸酶T和其他合适的核酸酶等。

[0063] 根据上述cDNA合成方法，反应容器中的反应介质经历数个温度以实现该方法的各方面。例如，RNA链在75℃-85℃的温度下降解。逆转录酶和酶在75℃-85℃的温度下失活。

[0064] 使用基于多次退火和成环的扩增循环的cDNA扩增

[0065] 然后，使用基于多次退火和成环的扩增循环来扩增所得单链cDNA分子。根据一个方面，使用DNA聚合酶在合适条件和试剂，包括在引物的5’端包含自退火序列的延伸引物下，生成包含逆转录引物序列的cDNA模板链的互补链。所得互补链在5’端包含自退火序列并在3’端包含其互补序列。cDNA模板链由互补链变性而来，并且互补链通过使3’端的自退火序列与其5’端的互补序列退火来成环。一旦成环，成环互补链将被抑制进行扩增。将生成cDNA模板的互补链并由互补链使cDNA链变性然后使互补链成环的步骤重复数次，如7-12次，以由各cDNA模板链生成多个成环互补链。

[0066] 数个成环互补链经变性，然后使用包含自退火序列的扩增引物进行扩增，以产生包含逆转录引物序列的双链扩增子。双链扩增子经变性并使用下述内容重复扩增数次：(1)具有与条码引物退火位点互补的3’序列的外部条码引物，其中外部条码引物还包含5’自退

火序列,测序引发序列和具有4-12个核苷酸的第二细胞特异性条码序列,和(2)包含5'自退火序列的引物。所得双链扩增子包含第一细胞特异性条码序列,第二细胞特异性条码序列和第一独特分子标识符条码序列。对所得双链扩增子经加工用于测序。

[0067] 根据一个方面,第一独特分子标识符条码序列可以具有半随机序列模式。

[0068] 本领域技术人员已知示例性的自退火序列并且包括GAT5和GAT1等。

[0069] 本领域技术人员已知示例性的条码引物退火位点序列并且包括RT3、Read2SP、Read1SP等。

[0070] 根据一个方面,提供了逆转录自一个或多个或多种RNA序列的一个或多个或多种cDNA序列、引物和至少一种聚合酶的反应混合物。根据一个方面,提供了具有链置换活性或者5'-3'外切核酸酶活性的聚合酶。链置换聚合酶是随着其延伸将会使下游片段离开原来位置的聚合酶。链置换聚合酶包括:Φ29聚合酶,Bst聚合酶,Pyrophage 3173,Vent聚合酶,Deep Vent聚合酶,TOPO Taq DNA聚合酶,Taq聚合酶,T7聚合酶,Vent(外切-)聚合酶,Deep Vent(外切-)聚合酶,9°Nm聚合酶,DNA聚合酶I的Klenow片段,MMLV逆转录酶,AMV逆转录酶,HIV逆转录酶,缺少3'-5'外切核酸酶活性的T7噬菌体DNA聚合酶的突变形式,或其混合物。具有5'侧翼内切核酸酶或5'-3'外切核酸酶活性的一种或多种聚合酶,诸如Taq聚合酶,Bst DNA聚合酶(全长),大肠杆菌DNA聚合酶,LongAmp Taq聚合酶,OneTaq DNA聚合酶或其混合物,可以用于去除由于不均匀引发而引起的残基偏倚(residual bias)。不具有链置换活性的其他聚合酶是有用的,如Q5、Phusion和Kapa HiFi。

[0071] 本领域技术人员已知并且可以商购能够用于制备测序文库的测序引发序列、衔接子序列、测序标识符、流动池退火衔接子,并且包括Read1SP、Read2SP、Index1、Index2、P5和P7。

[0072] 下表1中提供了示例性的序列。所有序列以5'-3'列出。H=不是G,B=不是A,D=不是C,V=不是T。Read1SP、Read2SP、Index1、Index2、P5和P7的序列为本领域技术人员所知并且来自亿明达公司(Illumina)和亿明达公司公开的信息。

[0073]

序列名称	核苷酸序列
GAT5	GTAGGTGTGAGTGATGGTTGAGGTAGT
B1	GAGGAG
GAT1	GTGAGTGATGGTTGAGGTAGTGTGGAG
RT3	AGTCGCTTGGGTGTAGTGC
UMI_A	HBDVHBDVHBDVHBDVHBDV
UMI_B	VDBHVDBHVDBHVDBHVDBH
C_n	GTTGTT, GTTAAA, GTTTGG, AGGGTT, AGGAAA, AGGTGG, TAATGG, GGAGAG, GGAAGT, GGATTA, AATGAG, AATAGT, AATTTA, TTGGAG, TTGAGT,TTGTTA,ATAATG,ATATAT,ATAGGA,TGTATG,TGTTAT,TGTGGA, GAGATG, GAGTAT, GAGGGA, GTTGAG, GTTAGT, GTTTTA, AGGGAG, AGGAGT, AGGTTA, TAAGAG, TAAAGT, TAATTA, GTTATG, GTTTAT, GTTGGA, AGGATG, AGGTAT, AGGGGA, TAAATG, TAATAT, TAAGGA, GGAGTT, AATGTT, TTGGTT, GGAAAA, AATAAA
G_m	GATATG, ATACG, CCGTCTG, TGCG, GAACTCG, ATGTAG, CCCG, TGTAG, GAGTAAG, ATCG, CCTAG, TGACCG, GACG, ATTAG, CCACTG, TGGTCTG, GTTTACG, ACAG, CGGAG, TACCTG, GTAG, ACGACG, CGCCG, TATTAAG, GTGATCG, ACCCG, CGTTCG, TAAG, GTCCG, ACTTATG, CGAG, TAGATG, GCTCAG, AGATG, CAGG, TTCACAG, GCAATCG, AGGCCG, CACTG, TTTG, GCGG, AGCAG, CATCTG, TTATATG, GCCTG, AGTG, CAAACG, TTGCAAG

[0074] 根据上述基于多次退火和成环的扩增循环方法,反应容器中的反应介质经历数个

温度以实现该方法的各方面。例如，延伸引物在0℃-10℃的温度下退火至cDNA模板链。互补链在10℃-65℃的温度下生成。互补链的成环发生在55℃-65℃的温度下。

[0075] 根据一个方面，扩增变性的互补链的步骤使用聚合酶链式反应进行，如使用15-20个聚合酶链式反应循环。

[0076] 根据一个方面，扩增变性的扩增子的步骤使用聚合酶链式反应进行，如使用3-7个聚合酶链式反应循环。

[0077] 根据一个方面，测序引发序列是Read2SP或Read1SP。

[0078] 测量逆转录引物降解效率

[0079] 根据一个方面，提供了用于测量或以其他方式确定逆转录引物降解效率的效率的方法。该方法包括在消化酶存在的情况下添加逆转录引物，其具有10-30个核苷酸的第二独特分子标识符条码序列。第二独特分子标识符条码序列包含不同于第一独特分子标识符条码序列的半随机序列模式。以此方式，依照包含第一独特分子标识符条码序列和第二独特分子标识符条码序列的产物的最终比例可以测量RT引物降解效率。

[0080] 扩增

[0081] 在某些方面，扩增使用PCR实现。PCR是这样一种反应，其中使用由上游和下游引物组成的一组引物或一对引物和聚合催化剂(如DNA聚合酶，通常为热稳定的聚合酶)，由靶多核苷酸制备复制拷贝。PCR的方法在本领域是公知的，并且在例如MacPherson等.(1991)PCR 1:使用方法(PCR 1:A Practical Approach)牛津大学出版社(Oxford University Press)IRL出版社(IRL Press)中教导。Mullis(美国专利号4,683,195、4,683,202和4,965,188)的术语“聚合酶链式反应”(“PCR”)指无需克隆或纯化即可提高靶序列区段浓度的方法。用于扩增靶序列的该方法包括提供具有所需靶序列的寡核苷酸引物和扩增试剂，然后在聚合酶(例如，DNA聚合酶)存在的情况下进行准确的一连串热循环。引物与双链靶序列各自对应的链(“引物结合序列”)互补。总之，为了进行扩增，将双链靶序列变性，然后将引物退火至靶分子中的其互补序列。退火后，用聚合酶延伸引物，从而形成一对新的互补链。变性、引物退火和聚合酶延伸步骤可以重复多次(即，变性、退火或延伸组成一个“循环”；可以存在许多“循环”)以获得高浓度的所需靶序列的扩增区段。所需靶序列的扩增区段的长度由引物相对于彼此的相对位置确定，并且因此，长度是可控参数。由于该过程的重复，该方法被称为“聚合酶链式反应”(下文称之为“PCR”)并且靶序列被称为是“PCR扩增的”。

[0082] 术语“PCR产物”、“PCR片段”和“扩增产物”指在变性、退火和延伸PCR步骤的两个或更多个循环完成后得到的化合物混合物。这些术语包括已经扩增了一个或多个靶序列的一个或多个区段的情况。

[0083] 任何寡核苷酸或多核苷酸都可以用适当的引物分子组进行扩增。用于进行PCR的方法和试剂盒是本技术领域已熟知的。产生多核苷酸复制拷贝的所有方法(如PCR或基因克隆)在本文中统称为复制。

[0084] “扩增”或“进行扩增”这样的表达指通过其将形成特定多核苷酸的额外或多个拷贝的过程。扩增包括诸如PCR、连接扩增(或连接酶链反应，LCR)和其他扩增方法的方法。这些方法在本领域中是已知且广泛应用的。参见，例如，美国专利号4,683,195和4,683,202，以及Innis等.，“PCR方法:方法和应用的指南”(PCR protocols:a guide to method and applications)学术出版社股份有限公司(Academic Press,Incorporated)(1990)(针对

PCR)；和Wu等.(1989)Genomics4:560-569(针对LCR)。通常，PCR过程描述了一种基因扩增方法，其包括(i)引物与DNA样品(或文库)中特定基因的序列特异性杂交，(ii)随后的扩增，涉及使用DNA聚合酶的多轮退火、延伸和变性，和(iii)筛选正确大小条带的PCR产物。使用的引物是具有足够长度和适当序列的寡核苷酸以引发聚合，即特异性地设计各引物，使其与待扩增的基因组基因座的各条链互补。

[0085] 进行扩增反应的试剂和硬件是市售可得的。用于从特定基因区域扩增序列的引物优选与目标区域或其侧接区中的序列互补并与其特异性杂交，并且可以使用本领域技术人员已知道的方法制备。通过扩增生成的核酸序列可以直接进行测序。

[0086] 当杂交以两个单链多核苷酸之间的反平行构型发生时，该反应被称为“退火”，并且这些多核苷酸被描述为“互补的”。如果杂交可以发生在第一多核苷酸的一条链与第二多核苷酸的链之间，那么双链多核苷酸可以与另一多核苷酸互补或同源。根据普遍接受的碱基配对规则，互补性或同源性(一个多核苷酸与另一个多核苷酸互补的程度)可依据相对链中预计将彼此之间形成氢键的碱基的比例来定量。

[0087] 术语“扩增试剂”可以指除了引物、核酸模板和扩增酶以外扩增所需的那些试剂(脱氧核糖核苷三磷酸，缓冲液等)。通常，将扩增试剂与其他反应组分一起放置并容纳在反应容器中(试管，微孔等)。扩增方法包括本领域技术人员已知的PCR方法，并且还包括滚环扩增(Blanco等.,J.Biol.Chem.,264,8935-8940,1989)、超支化滚环扩增(Lizard等.,Nat.Genetics,19,225-232,1998)和环介导的等温扩增(Notomi等.,Nuc.Acids Res.,28,e63,2000)，其各自通过引用将其全部内容纳入本文。

[0088] 其他扩增方法，如英国专利申请号GB 2,202,328以及PCT专利申请号PCT/US89/01025中所述的方法，各自通过引用纳入本文，可以依据本公开使用。可以根据本公开使用乳液PCR。其它合适的扩增方法包括“race”和“单侧PCR”(Frohman，述于《PCR方案：方法和引用的指南(PCR Protocols:A Guide To Methods And Applications)》，学术出版社，纽约，1990，其各自通过引用纳入本文)。基于在具有所得“二寡核苷酸”序列的核酸存在的情况下连接两个(或多个)寡核苷酸并因此扩增该二寡核苷酸的方法也可以用于根据本公开扩增DNA(Wu等.,Genomics 4:560-569,1989，通过引用纳入本文)。

[0089] 待扩增的RNA可以获得自单个细胞或小细胞群。本文所述方法允许由反应混合物中的任何物种或生物体扩增RNA，如在单个反应容器中进行的单一反应混合物。在一个方面中，本文所述方法包括由任何来源进行RNA的序列非依赖性扩增，所述来源包括但不限于人、动物、植物、酵母、病毒、真核和原核RNA。

[0090] 引物

[0091] 本文所用术语“引物”通常包括这样的天然或合成的寡核苷酸，其与多核苷酸模板形成双链体时能够用作核酸合成的起点(如测序引物)并从其3’末端沿模板延伸以形成延伸的双链体。引物包括延伸引物，扩增引物或逆转录引物。

[0092] 在延伸过程中添加的核苷酸序列由模板多核苷酸的序列决定。通常，引物通过DNA聚合酶或逆转录酶延伸。引物通常具有这样范围内的长度：3-36个核苷酸、5-24个核苷酸或14-36个核苷酸。本发明范围内的引物还包括正交引物、扩增引物、构建引物等。成对的引物可以侧接于感兴趣的序列或一组感兴趣的序列。引物和探针可以按顺序简并或准简并(quasi-degenerate)。本发明范围内的引物结合于靶序列邻近处。“引物”可以被认为是短

多核苷酸，通常具有游离的3'-OH基团，其通过与靶标杂交结合潜在地存在于感兴趣样品中的模板或靶标，并在此后促进与该靶标互补的多核苷酸的聚合。本发明的引物由核苷酸组成，其范围在17-30个核苷酸。在一个方面，引物是至少17个核苷酸、又或者至少18个核苷酸、又或者至少19个核苷酸、又或者至少20个核苷酸、又或者至少21个核苷酸、又或者至少22个核苷酸、又或者至少23个核苷酸、又或者至少24个核苷酸、又或者至少25个核苷酸、又或者至少26个核苷酸、又或者至少27个核苷酸、又或者至少28个核苷酸、又或者至少29个核苷酸、又或者至少30个核苷酸、又或者至少50个核苷酸、又或者至少75个核苷酸又或者至少100个核苷酸。

[0093] 测序

[0094] 例如，使用本领域技术人员已知的高通量测序方法对扩增子进行测序。使用本领域已知的多种测序方法可以确定感兴趣的核酸序列的序列，所述方法包括但不限于通过杂交测序(SBH)，通过连接测序(SBL)(Shendure等.(2005)Science 309:1728)，定量增量荧光核苷酸加法测序(QIFNAS)，逐步连接和切割，荧光共振能量转移(FRET)，分子信标，TaqMan报告探针消化，焦磷酸测序，荧光原位测序(FISSEQ)，FISSEQ珠(美国专利号7,425,431)，摇摆测序(PCT/US05/27695)，多重测序(美国系列号12/027,039，提交于2008年2月6日；Porreca等(2007)Nat.Methods 4:931)，聚合集落(POLONY)测序(美国专利号6,432,360、6,485,944和6,511,803，以及PCT/US05/06425)；纳米网格滚环测序(ROLONY)(美国系列号12/120,541，2008年5月4日提交)、等位基因特异性寡聚体连接试验(例如，寡聚体连接试验(OLA)，使用连接的线性探针和滚环扩增(RCA)读出的单模板分子OLA，连接的锁式探针，和/或使用连接的环状锁式探针和滚环扩增(RCA)读出的单模板分子OLA)等。也可以利用高通量测序方法，例如，使用诸如Roche 454、Illumina Solexa、AB-SOLiD、Helicos、Polonator平台等的平台。本领域已知各种基于光的测序技术(Landegren等.(1998)Genome Res.8:769-76；Kwok(2000)Pharmacogenomics 1:95-100；以及Shi(2001)Clin.Chem.47:164-172)。

[0095] 扩增的DNA可以通过任何合适的方法进行测序。具体而言，可以使用高通量筛选方法扩增的DNA进行测序，如应用生物系统公司(Applied Biosystems)的SOLiD测序技术或亿明达公司的基因组分析仪.在本发明的一个方面，可以对扩增的DNA进行鸟枪法测序。读数的数量可以是至少10,000、至少100万、至少1000万、至少1亿或至少10亿。在另一个方面，读数的数量可以是10,000-100,000，或者100,000-100万，或者100万-1000万，或者1000万-1亿、或者1亿到10亿。“读数(read)”是通过测序反应获得的连续核酸序列的长度。

[0096] “鸟枪法测序”是指用于非常大量DNA(如整个基因组)测序的方法。在该方法中，首先将待测序的DNA切碎成较小的片段，可以对其进行单独测序。然后根据这些片段的重叠序列将这些片段的序列重组为它们的原始顺序，从而产生完整的序列。可以使用多种不同的技术来完成DNA的“切碎”，包括限制酶消化或机械剪切。重叠序列通常由适当编程的计算机对齐。鸟枪法测序cDNA文库的方法和程序在本领域中是公知的。

[0097] 扩增和测序方法在预测医学领域是有用的，其中诊断试验、预后试验、药物基因组学和监测临床试验用于预后(预测)目的，从而预防性地治疗个体。相应地，本发明的一个方面涉及诊断试验，其用于确定RNA以便确定个体是否处于患病症和/或疾病的风险中。这样的试验可用于预后或预测目的，从而因此在病症和/或疾病发作之前预防性治疗个体。相应

地，在某些示例性实施方式中，提供了使用本文所述一种或多种表达谱方法来诊断和/或预测一种或多种疾病和/或病症的方法。

[0098] 互补性和杂交

[0099] 本文所用术语“互补”和“互补性”用于指通过碱基配对规则相关联的核苷酸序列。例如，序列5'-AGT-3'与序列5'-ACT-3'互补。互补性可以是部分的或完全的。部分互补性发生在当一个或多个核酸碱基根据碱基配对规则不匹配时。核酸间完全或完整互补性发生在每个核酸碱基各自在碱基配对规则下与另一个碱基匹配时。核酸链间的互补性程度对于核酸链间杂交的效率和强度有显著影响。

[0100] 术语“杂交”是指互补核酸的配对。杂交和杂交的强度(即核酸之间关联的强度)受诸如如下因素的影响：核酸之间的互补性程度，涉及条件的严谨性，形成的杂交体的T_m和核酸内G:C比例。认为在其结构中包含互补核酸配对的单个分子是“自交的”。

[0101] 术语“T_m”指核酸的解链温度。解链温度是双链核酸分子群体一半解离成单链的温度。计算核酸T_m的等式是本领域熟知的。如标准参考文献所示，当核酸处于1M NaCl水性溶液中时，通过T_m＝81.5+0.41(%G+C)等式可以简单估计T_m值(参见，例如，Anderson和Young，定量滤膜杂交(Quantitative Filter Hybridization)，Nucleic Acid Hybridization (1985))。其他参考文献包括更复杂的计算，它们将结构以及序列特性考虑到T_m的计算中。

[0102] 术语“严谨性”指进行核酸杂交的温度，离子强度和存在其他化合物(如有机溶剂)的条件。

[0103] 当述及核酸杂交时，“低严谨性条件”包括等同于使用约500个核苷酸长度的探针时，在42℃的溶液中结合或杂交的如下条件，所述溶液由5x SSPE(43.8g/l NaCl、6.9g/l $NaH_2PO_4(H_2O)$和1.85g/l EDTA，用NaOH将pH调至7.4)、0.1％SDS、5x Denhardt试剂(50x Denhardt试剂，其每500ml含：5g Ficoll(400型，法玛西亚公司(Pharmacia))、5g BSA(组分V；西格玛公司(Sigma)))和100mg//ml变性的鲑鱼精DNA组成，然后在42℃包括5x SSPE、0.1％SDS的溶液中洗涤。

[0104] 当述及核酸杂交时，所用的“中严谨性条件”包括等同于使用约500个核苷酸长度的探针时，在42℃的溶液中结合和杂交的如下条件，所述溶液由5x SSPE(43.8g/l NaCl、6.9g/l $NaH_2PO_4(H_2O)$和1.85g/l EDTA、用NaOH将pH调节至7.4)、0.5％SDS、5x Denhardt试剂和100mg/ml变性的鲑鱼精DNA组成，然后在42℃包括1.0x SSPE、1.0％SDS的溶液中洗涤。

[0105] 当述及核酸杂交时，所用的“高严谨性条件”包括等同于使用约500个核苷酸长度的探针时，在42℃的溶液中结合和杂交的如下条件，所述溶液由5x SSPE(43.8g/l NaCl、6.9g/l $NaH_2PO_4(H_2O)$和1.85g/l EDTA、用NaOH将pH调节至7.4)、0.5％SDS、5x Denhardt试剂和100mg/ml变性的鲑鱼精DNA组成，然后在42℃包括0.1x SSPE、1.0％SDS的溶液中洗涤。

[0106] 元件和电子设备和介质

[0107] 在某些示例性实施方式中，提供了包含本文所述的一种或多种RNA或cDNA序列的电子设备可读介质。本文所用“电子设备可读介质”指用于存储、携带或保持可由电子设备直接读取和访问的数据或信息的任何合适的介质。这样的介质可以包括但不限于磁存储介质、如软盘，硬盘存储介质和磁带；光存储介质，如光盘；电子存储介质，如RAM，ROM，EPROM，EEPROM等；普通硬盘和这些类别的混合物，如磁/光存储介质。介质适用于或被配制成用于以使其上记录有本文描述的一个或多个表达谱。

[0108] 本文所用术语“电子设备”旨在包括被配置成或适用于存储数据或信息的任何合适的计算或处理设备或其他设备。适合用于本发明的电子设备的示例包括独立计算设备;网络,包括局域网(LAN)、广域网(WAN)互联网、内联网和外联网;电子设备,如个人数字助理(PDA)、蜂窝电话、寻呼机等;和本地和分布式处理系统。

[0109] 本文所用“记录的”指用于在电子设备可读介质上存储或编码信息的过程。本领域技术人员可以容易地采用任何目前已知用于在已知介质上记录信息的方法来生成包含本文描述的一个或多个表达概况的制品。

[0110] 可使用各种软件程序和格式来将本发明的RNA或cDNA信息存储在电子设备可读介质上。例如,核酸序列可以用文字处理文本文件来表示,以如WordPerfect和微软Word等市售可得软件对其进行格式化,或以ASCII文件的形式表示,存储在数据库应用程序,诸如DB2、Sybase、Oracle等,以及以其他形式。可使用任何数量的数据处理器结构格式(例如,文本文件或数据库),从而获得或创建其上记录有本文所述一个或多个表达谱的介质。

[0111] 应理解的是,已描述的本发明的实施方式仅用于说明本发明的一些应用和原理。基于本文的教导,本领域技术人员可进行多种修改而不偏离本发明的真正精神与范围。贯穿本发明中所引用的所有参考文献、专利和公开专利申请的内容通过引用全文纳入本文并用于所有目的。

[0112] 以下实施例是本发明的代表。这些实施例并不构成对本发明范围的限制,因为这些和其他等价实施方式将对于本发明、附图和所附权利要求而言是显而易见的。

[0113] 实施例I

[0114] 由mRNA模板合成cDNA

[0115] 图1显示了用于由mRNA模板合成cDNA的一个示例性方法。将悬浮于4μl的细胞裂解缓冲液(1X SuperScript IV缓冲液(赛默飞世尔科技公司(Thermo Fisher Scientific)),0.5%IGEPAL CA-630(西格玛-奥德里奇公司(Sigma-Aldrich)),500mM dNTP,6mM $MgSO_4$,1M甜菜碱,1U SUPERase In RNA酶抑制剂(赛默飞世尔科技公司),2.5μM‘RT-A’逆转录引物(IDT))的裂解的RNA加热至72℃持续3分钟以使RNA二级结构变性。加热后,冷却混合物至4℃,以使逆转录酶引物(RT-A)退火至mRNA转录本的多聚(A)段(tract)。该RT-A引物包含(从5’端开始):GAT5序列,其被用于在cDNA扩增期间产生自退火环,B1间隔子序列,RT3序列,其被用作最终PCR步骤期间外部条码引物的退火位点,C_n序列,其是通过≥3汉明距离分离的“n”个不同的6核苷酸细胞特异性条码之一,UMI_A序列,其是复杂性降低的(即半随机的)20聚体,具有大约35亿(3^{20})可能的组合以独特地标记各转录本,和12核苷酸的多聚(T)区段(参见表1)。添加2μl的逆转录酶混合物(1X SuperScript IV缓冲液,0.1M DTT,1U SUPERase In RNA酶抑制剂,60U SuperScript IV(赛默飞世尔科技公司)),并将混合物于55℃孵育10分钟以催化cDNA合成。为了防止过量RT-A引物在后续cDNA扩增期间退火,添加2μl引物消化混合物(1X外切核酸酶I缓冲液(NEB)、12U外切核酸酶I(NEB),2.5uM“RT-B”逆转录引物(IDT)),并于37℃孵育30分钟以消化逆转录引物。根据一个方面,添加第二逆转录引物(“RT-B”),并且其与RT-A相同,除了其包含UMI_B模式而非UMI_A模式以外(参见表1),这允许测量外切核酸酶消化效率,因为不完整的消化将产生具有UMI_A和UMI_B条码混合物的cDNA扩增产物。消化后,加热混合物至80℃持续20分钟以降解RNA并使外切核酸酶I和SuperScript IV热失活。

[0116] 实施例II

[0117] cDNA扩增

[0118] 图2显示了使用基于多次退火和成环的扩增循环(MALBAC)扩增实施例1的cDNA以形成成环延伸产物,然后PCR扩增成环延伸产物。MALBAC过程述于Zong,C.,Lu,S.,Chapman,A.R.和Xie,X.S.(2012)单个人细胞的单核苷酸和拷贝数变异的基因组范围检测(Genome-wide detection of single-nucleotide and copy-number variations of a single human cell).Science,338,1622-1626;和Chapman,A.R.,He,Z.,Lu,S.,Yong,J.,Tan,L.,Tang,F.和Xie,X.S.(2015)使用MALBAC的单个细胞转录组扩增(Single cell transcriptome amplification with MALBAC).PLoS One,10,e0120889,其各自通过引用其全部内容纳入本文。

[0119] 对于MALBAC,将22μl的cDNA扩增混合物(1X ThermoPol缓冲液(NEB),200μM dNTP,1.25mM $MgSO_4$,50μM“GAT5-B1-7N”引物(IDT),50μM“GAT5-B1”引物(IDT),2U Deep Vent(外切-)DNA聚合酶(NEB))添加到cDNA合成混合物。加热混合物至95℃持续5分钟,然后通过重复下述孵育程序10次进行准线性cDNA扩增:4℃持续50s,10℃持续50s,20℃持续50s,30℃持续50s,40℃持续45s,50℃持续45s,65℃持续4分钟,95℃持续20s,58℃持续20s。该孵育程序首先冷却混合物,以允许GAT5-B1-7N引物沿着cDNA随机退火。增加至65℃允许Deep Vent(外切-)催化第二链合成。95℃时的变性分离第二链,而冷却至58℃允许第二链(延伸产物)的互补5’和3’序列形成稳定的环并防止进一步的扩增。准线性扩增后,使用GAT5引物进行17个循环的PCR扩增。MALBAC后,添加0.4μl的50μM外部条码引物,并用OB_m和GAT5-B1进行另外5个循环的PCR以产生最终产物。外部条码引物包含(从5’端开始):Read2SP序列,其是Illumina读数2测序引发序列,G_m序列,是通过≥2汉明距离分离的“m”个不同的4-7核苷酸细胞特异性条码之一,和RT3序列,其退火至MALBAC cDNA产物。添加外部条码产生总计m x n种可能的条码。使用0.8x Amazi珠(Aline生物科学公司(Aline Biosciences))纯化产物以去除<150碱基对的引物二聚体。

[0120] 实施例III

[0121] 文库制备

[0122] 图3显示了由实施例II的扩增子制备用于测序的文库的方法。实施例II的扩增子产物可以使用多种化学物质制备为Illumina测序相容文库。对于文库制备,将高活性Tn5转座酶,诸如来自Nextera DNA文库制备试剂盒(亿明达公司)的高活性Tn5转座酶,用于将读数1测序衔接子的部分与扩增子连接,然后用全长测序衔接子进行PCR以产生Illumina相容测序文库(图3)。使用Nextera试剂盒的标签化产生多个产物,并且所需产物包含侧接cDNA的读数1测序引发序列(Read1SP)和条码序列。将标签化的产物添加到50μl的PCR扩增混合物(1X Kapa HiFi热启动主混合物,0.5μM S5XX引物(亿明达公司),0.5μM读数2索引衔接子引物(IDT)),并使用下述孵育程序扩增:72℃持续3分钟,98℃持续30s,然后以98℃持续10s、63℃持续30s和72℃持续3分钟进行5个循环。最终测序文库再次使用0.8x Amazi珠进行纯化,然后在测序前使用生物分析仪(安捷伦公司(Agilent))进行大小调整以调节浓度。

[0123] 实施例IV

[0124] 确定均质人细胞培养物内组织特异性转录调节模型

[0125] 如下所示,在两个人细胞系上进行针对数字转录组的基于多次退火和成环的扩增

循环MALBAC-DT。U2-OS骨骨肉瘤和HEK293T胚胎肾细胞系获自美国典型培养物保藏中心(the American Type Culture Collection,ATCC,罗克维尔市)。将U2-OS和HEK293T细胞维持于补充有10％胎牛血清和100U/ml盘尼西林-链霉素的达氏改良伊氏培养基中(ATCC)。为了收集,使用0.05％胰蛋白酶-EDTA(赛默飞世尔科技公司)使细胞悬浮,然后用1X PBS洗涤并重悬于补充有10％胎牛血清、2μg/ml碘化丙啶(赛默飞世尔科技公司)和1μM钙黄绿素AM(BD生物科学公司(BD Bioscience))的达氏改良伊氏培养基中。使用MoFlo Astrios(贝克曼库尔特公司(Beckman Coulter))将具有阳性钙绿黄素信号和阴性碘化丙啶信号的存活单个细胞分选到96孔板中,其中各孔包含3μl的裂解缓冲液(1X SuperScript IV缓冲液(赛默飞世尔科技公司),0.5％IGEPAL CA-630(西格玛-奥德里奇公司),500mM dNTP,6mM $MgSO_4$,1M甜菜碱,1U SUPERase In RNA酶抑制剂(赛默飞世尔科技公司),2.5μM“RT-A”逆转录引物(IDT),$2.4x10^7$ ERCC的稀释物)。该RT-A引物包含(从5’端开始):GAT5序列,其被用于在cDNA扩增期间产生自退火环,B1间隔子序列,RT3序列,其被用作最终PCR步骤期间外部条码引物的退火位点,C_n序列,其是通过≥3汉明距离分离的“n”个不同的6核苷酸细胞特异性条码之一,UMI_A序列,其是复杂性降低的20聚体,具有大约35亿(3^{20})可能的组合以独特地条码化各转录本,和12核苷酸的多聚(T)区段(表1)。

[0126] 对于cDNA合成,将平板离心,于72℃孵育3分钟以使RNA二级结构变性,然后冷却至4℃以允许引物退火。添加1μl的逆转录混合物(1X SuperScript IV缓冲液,0.1M DTT,1U SUPERase In RNA酶抑制剂,60U SuperScript IV(赛默飞世尔科技公司)),并将混合物于55℃孵育10分钟以催化cDNA合成。为了防止过量RT-A引物在后续cDNA扩增期间退火,添加2μl引物消化混合物(1X外切核酸酶I缓冲液(NEB)、12U外切核酸酶I(NEB),2.5uM“RT-B”逆转录引物(IDT)),并于37℃孵育30分钟以消化逆转录引物。RT-B引物与RT-A相同,除了其包含UMI_B模式而非UMI_A模式以外(表1),这允许测量外切核酸酶消化效率,因为不完整的消化将产生具有UMI_A和UMI_B条码混合物的cDNA扩增产物。消化后,加热混合物至80℃持续20分钟以降解RNA并使外切核酸酶I和SuperScript IV热失活。

[0127] 使用基于多次退火和成环的扩增循环(MALBAC)扩增所得cDNA(图2)。对于MALBAC,将24μl的cDNA扩增混合物(1X ThermoPol缓冲液(NEB),200μM dNTP,1.25mM $MgSO_4$,50μM“GAT5-B1-7N”引物(IDT),50μM“GAT5-B1”引物(IDT),2U Deep Vent(外切-)DNA聚合酶(NEB))添加到cDNA合成混合物。通过加热混合物至95℃持续5分钟然后重复10次下述循环进行准线性cDNA扩增:4℃持续50s、10℃持续50s、20℃持续50s、30℃持续50s、40℃持续45s、50℃持续45s、65℃持续4分钟、95℃持续20s、58℃持续20s。准线性扩增后,PCR扩增通过这样进行:加热至98℃持续1分钟,然后重复下述孵育程序17次:95℃持续20s、58℃持续30s、72℃持续3分钟。MALBAC后,添加0.4μl的50μM外部条码序列(参见表1所示序列),并通过这样进行另一轮PCR:加热至95℃持续1分钟,将95℃持续20s、58℃持续30s和72℃持续3分钟这样的循环重复5次,然后于72℃孵育5分钟。外部条码引物包含(从5’端开始):Read2SP序列,其是Illumina读数2测序引发序列,G_m序列,是通过≥2汉明距离分离的“m”个不同的4-7核苷酸细胞特异性条码之一,和RT3序列,其退火至MALBAC cDNA产物。添加外部条码产生总计m x n种可能的条码。使用0.8x Amazi珠(Aline生物科学公司(Aline Biosciences))纯化产物以去除<150碱基对的引物二聚体。

[0128] 使用Nextera DNA文库制备试剂盒(亿明达公司)将产物制备为Illumina测序相容

文库。使用Nextera试剂盒的标签化产生多个产物，并且所需产物包含位于cDNA一侧的条码序列和读数1测序引发序列(Read1SP)以及位于另一侧的N5XX序列。将标签化的产物添加到PCR扩增混合物以产生50μl的总PCR混合物(1X Kapa HiFi热启动主混合物，0.5μM S5XX引物(亿明达公司)，0.5μM读数2索引衔接子引物(IDT))，并通过下述方式扩增：加热至72℃持续3分钟，98℃持续30s，然后以98℃持续10s、63℃持续30s和72℃持续3分钟进行5个循环。使用0.8X Amazi珠纯化产物，洗脱至20ul，然后使用E-Gel SizeSelect2％琼脂糖凝胶(飞世尔公司(Fisher))针对300-500bp的条带进行大小选择，然后在加样至HiSeq 4000(亿明达公司)用于测序之前使用生物分析仪(安捷伦公司)进行定量用于调节浓度。

[0129] 使用10^6读数/细胞的平均测序深度对约700个均质培养的HEK293T细胞和约700个均质培养的U-2OS细胞进行测序。80％的读段映射到外显子组，这表明文库准确反映了转录组。在该深度下，持续检测到12,000个基因。图4A显示了针对HEK293T的基因表达相关性矩阵。对角线上的每个方块表示观测到强相关性的基因簇。这些观测值来自处于非平衡稳态的培养物中的波动。12,000个基因中存在总计约100-200个簇。针对HEK293T数据集，使用t-随机邻居嵌入算法(t-stochastic neighbor embedding algorithm，t-SNE)，图4B描述了基因的聚类(左图)而图4C描述了细胞的聚类(右图)。在图4B的基因聚类图中，各基因簇对应相关性矩阵中的方块。在基因聚类图中，各点是12,000个基因中的一个，而各簇对应相关性矩阵中的方块。在图4C的细胞聚类图中，各点是约700个HEK细胞中的一个，并且不存在可分解的簇。这意味着基因簇不是因为表型不同的细胞的簇。图5显示了基因簇的比较，12,000个基因中的3000个针对HEK293T(上图)。图5显示了基因簇的比较，12,000个基因中的3000个针对U-2OS(下图)。两种细胞系之间存在一些共有簇，诸如涉及细胞循环和蛋白质合成的那些。然而，也存在不同的基因簇，其可能是细胞类型特异性转录调节过程。图6突出显示了图5中标记的蛋白质合成簇。针对tRNA合成、氨基酸合成、氨基酸转运以及翻译启动的控制中所涉及的那些富集该簇中的基因，所有这些在蛋白质合成过程中都十分重要。因此，关联的基因簇具有相关的生物学功能和转录调节。

[0130] 实施例V

[0131] 试剂盒

[0132] 公开的逆转录和扩增方法所需的材料和试剂可在试剂盒中组装在一起。本公开的试剂盒通常将至少包括本文所述进行所要求保护的方法所必需的逆转录酶，和逆转录引物，降解酶，核苷酸，DNA聚合酶以及延伸和扩增引物。在优选实施方式中，试剂盒还将包括用于逆转录RNA至cDNA并扩增cDNA的说明。在各种情况中，试剂盒将优选具有对各种单独试剂、酶或反应物不同的容器。通常，将各物质在其各自的容器中适当分装。试剂盒的容器装置通常包括至少一个小瓶或试管。也可以是能够将试剂放置并分装于其中的细颈瓶、瓶子和其他容器装置。试剂盒的单个容器将优选保持密闭状态以用于商业销售。合适的较大容器可包括注塑或吹塑的塑料容器，其中保留所需小管。优选说明书与试剂盒一起提供。

[0133] 实施方式

[0134] 本公开提供了扩增RNA模板链的方法，其包括使用逆转录酶和逆转录引物序列将所述RNA模板链逆转录成cDNA模板链，所述逆转录引物序列具有与所述RNA模板链的5’多聚(A)序列互补的3’多聚(T)序列，其中所述逆转录引物序列还包含5’自退火序列，条码引物退火位点，具有4-12个核苷酸的第一细胞特异性条码序列和具有10-30个核苷酸的第一独

特分子标识符条码序列,其中所述cDNA模板链在所述cDNA模板的5'包含逆转录引物序列且所述cDNA模板链与所述RNA链杂交,用酶消化过量的逆转录引物序列,降解所述RNA链以产生作为单链的cDNA模板链,使所述逆转录酶失活,使所述酶失活,(a)使用DNA聚合酶和延伸引物生成包含逆转录引物序列的所述cDNA模板链的互补链,所述延伸引物在引物的5'端包含所述自退火序列,其中所述互补链在5'端包含所述自退火序列并在3'端包含其互补序列,(b)使所述cDNA模板链由所述互补链变性,并通过将3'端的所述自退火序列和5'端的其互补序列退火来使所述互补链成环,从而抑制所述互补链的扩增,重复步骤(a)和(b)数次以由所述cDNA模板链生成数个成环互补链,使所述数个成环互补链变性,并使用包含所述自退火序列的扩增引物扩增经变性的互补链,以产生包含所述逆转录引物序列的双链扩增子,使所述双链扩增子变性,并使用下述内容重复扩增经变性的扩增子数次:(1)外部条码引物,其具有与所述条码引物退火位点互补的3'序列,并且所述外部条码引物还包含5'自退火序列,测序引发序列和具有4-12个核苷酸的第二细胞特异性条码序列,和(2)包含3'自退火序列的引物,以产生具有第一细胞特异性序列、第二细胞特异性条码序列和第一独特分子标识符条码序列的双链扩增子。根据一个方面,RNA是信使RNA、转移RNA、核糖体RNA、长链非编码RNA或小干扰RNA。根据一个方面,RNA来自单个细胞。根据一个方面,RNA来自异质细胞群内的单个细胞。根据一个方面,RNA来自单个产前细胞。根据一个方面,RNA来自单个癌细胞。根据一个方面,RNA来自单个循环肿瘤细胞。根据一个方面,逆转录酶是SuperScript II、III或IV,M-MLV逆转录酶,Maxima逆转录酶,Protoscript逆转录酶或Thermoscript逆转录酶。根据一个方面,3'多聚(T)序列包含10-30个T核苷酸。根据一个方面,自退火序列是GAT5或GAT1。根据一个方面,条码引物退火位点是RT3、Read1SP或Read2SP。根据一个方面,酶是具有链置换活性或者5'-3'外切核酸酶活性的聚合酶。根据一个方面,酶是Φ29聚合酶,Bst聚合酶,Pyrophage 3173,Vent聚合酶,Deep Vent聚合酶,TOPO Taq DNA聚合酶,Taq聚合酶,T7聚合酶,Vent(外切-)聚合酶,Deep Vent(外切-)聚合酶,9°Nm聚合酶,DNA聚合酶I的Klenow片段,MMLV逆转录酶,AMV逆转录酶,HIV逆转录酶,缺少3'-5'外切核酸酶活性的T7噬菌体DNA聚合酶的突变形式,Taq聚合酶,Bst DNA聚合酶(全长),大肠杆菌DNA聚合酶,LongAmpTaq聚合酶,OneTaq DNA聚合酶,Q5,Phusion或Kapa HiFi。根据一个方面,RNA链在75℃-85℃的温度下降解。根据一个方面,逆转录酶和酶在75℃-85℃的温度下失活。根据一个方面,延伸引物在0℃-10℃的温度下退火至cDNA模板链。根据一个方面,互补链在10℃-65℃的温度下生成。根据一个方面,互补链的成环发生在55℃-60℃的温度下。根据一个方面,步骤(a)和(b)重复7-12次。根据一个方面,扩增经变性的互补链用聚合酶链式反应进行。根据一个方面,用15-20个聚合酶链式反应循环扩增经变性的互补链。根据一个方面,用聚合酶链式反应扩增经变性的扩增子。根据一个方面,用3-7个PCR循环重复扩增经变性的扩增子。根据一个方面,对所得双链扩增子进行加工用于测序。根据一个方面,第一独特分子标识符条码序列包含半随机序列模式。根据一个方面,使用酶消化过量转录引物的步骤包括添加逆转录引物,其具有第二独特分子标识符条码序列,其具有10-30个核苷酸,包含半随机序列模式,并且与所述第一独特分子标识符条码序列不同。

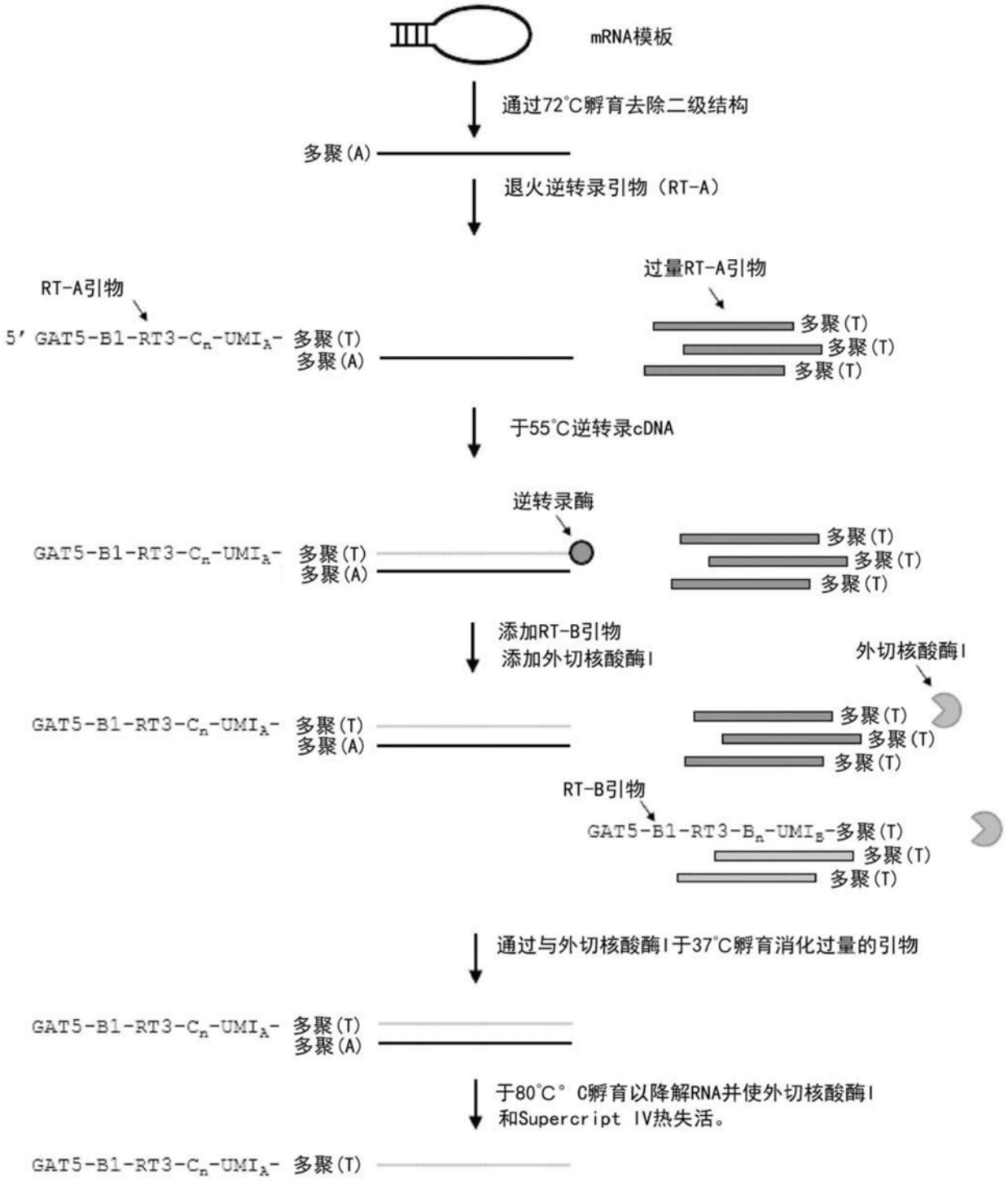

mRNA模板
通过72°C孵育去除二级结构
多聚(A)
退火逆转录引物（RT-A）
RT-A引物
5′ GAT5-B1-RT3-Cn-UMIA- 多聚(T)
多聚(A)
过量RT-A引物
多聚(T)
多聚(T)
多聚(T)
于55°C逆转录cDNA
逆转录酶
GAT5-B1-RT3-Cn-UMIA- 多聚(T)
多聚(A)
多聚(T)
多聚(T)
多聚(T)
添加RT-B引物
添加外切核酸酶I
外切核酸酶I
GAT5-B1-RT3-Cn-UMIA- 多聚(T)
多聚(A)
多聚(T)
多聚(T)
多聚(T)
RT-B引物
GAT5-B1-RT3-Bn-UMIB-多聚(T)
多聚(T)
多聚(T)
通过与外切核酸酶I于37°C孵育消化过量的引物
GAT5-B1-RT3-Cn-UMIA- 多聚(T)
多聚(A)
于80°C° C孵育以降解RNA并使外切核酸酶I
和Supercript IV热失活。
GAT5-B1-RT3-Cn-UMIA- 多聚(T)

图1

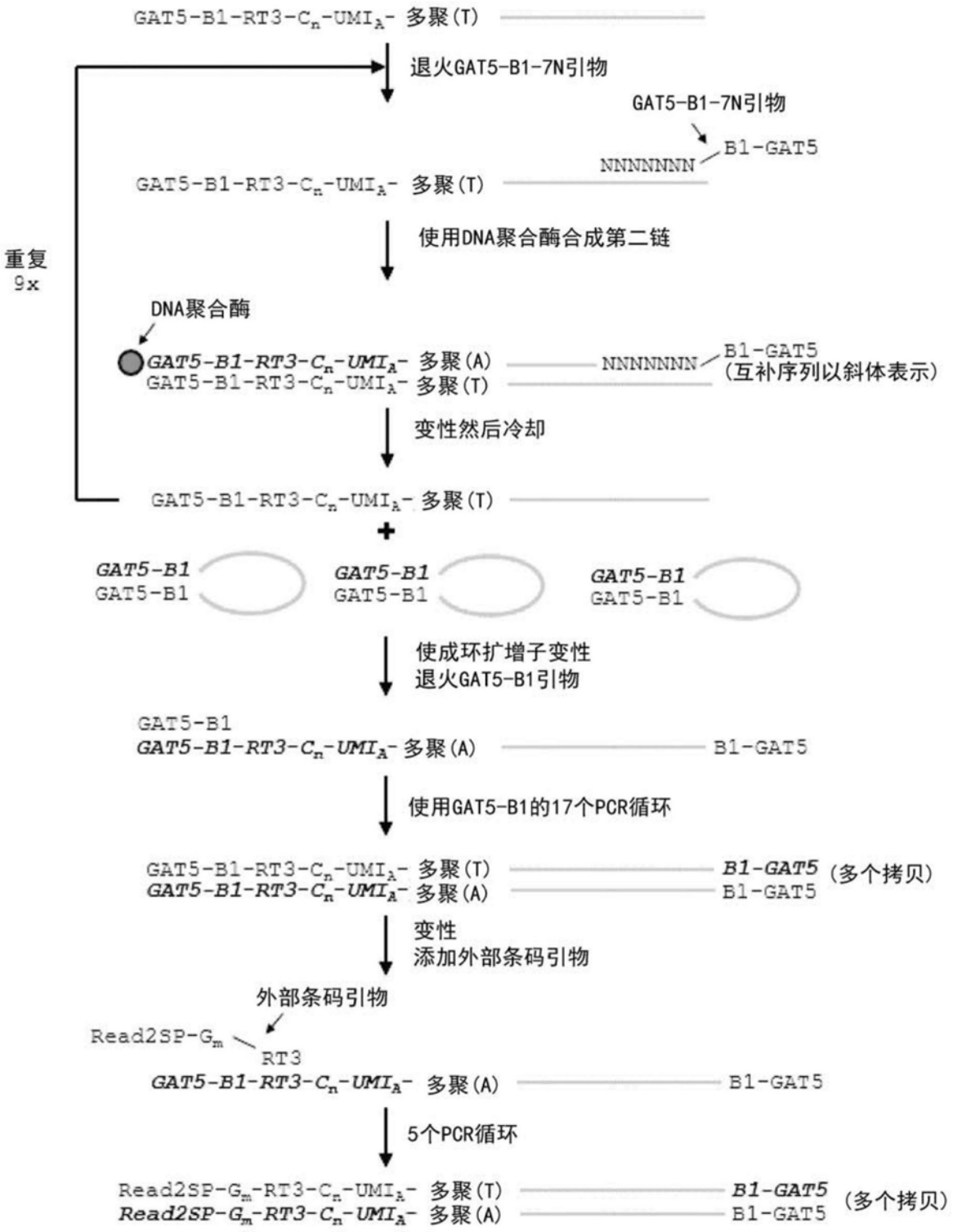

GAT5-B1-RT3-Cn-UMIA- 多聚(T)
退火GAT5-B1-7N引物
GAT5-B1-7N引物
NNNNNNN
B1-GAT5
GAT5-B1-RT3-Cn-UMIA- 多聚(T)
使用DNA聚合酶合成第二链
重复
9x
DNA聚合酶
GAT5-B1-RT3-Cn-UMIA- 多聚(A)
GAT5-B1-RT3-Cn-UMIA- 多聚(T)
NNNNNNN
B1-GAT5
(互补序列以斜体表示)
变性然后冷却
GAT5-B1-RT3-Cn-UMIA- 多聚(T)
GAT5-B1
GAT5-B1
GAT5-B1
GAT5-B1
GAT5-B1
GAT5-B1
使成环扩增子变性
退火GAT5-B1引物
GAT5-B1
GAT5-B1-RT3-Cn-UMIA- 多聚(A)
B1-GAT5
使用GAT5-B1的17个PCR循环
GAT5-B1-RT3-Cn-UMIA- 多聚(T)
GAT5-B1-RT3-Cn-UMIA- 多聚(A)
B1-GAT5
B1-GAT5
(多个拷贝)
变性
添加外部条码引物
外部条码引物
Read2SP-Gm
RT3
GAT5-B1-RT3-Cn-UMIA- 多聚(A)
B1-GAT5
5个PCR循环
Read2SP-Gm-RT3-Cn-UMIA- 多聚(T)
Read2SP-Gm-RT3-Cn-UMIA- 多聚(A)
B1-GAT5
B1-GAT5
(多个拷贝)

图2

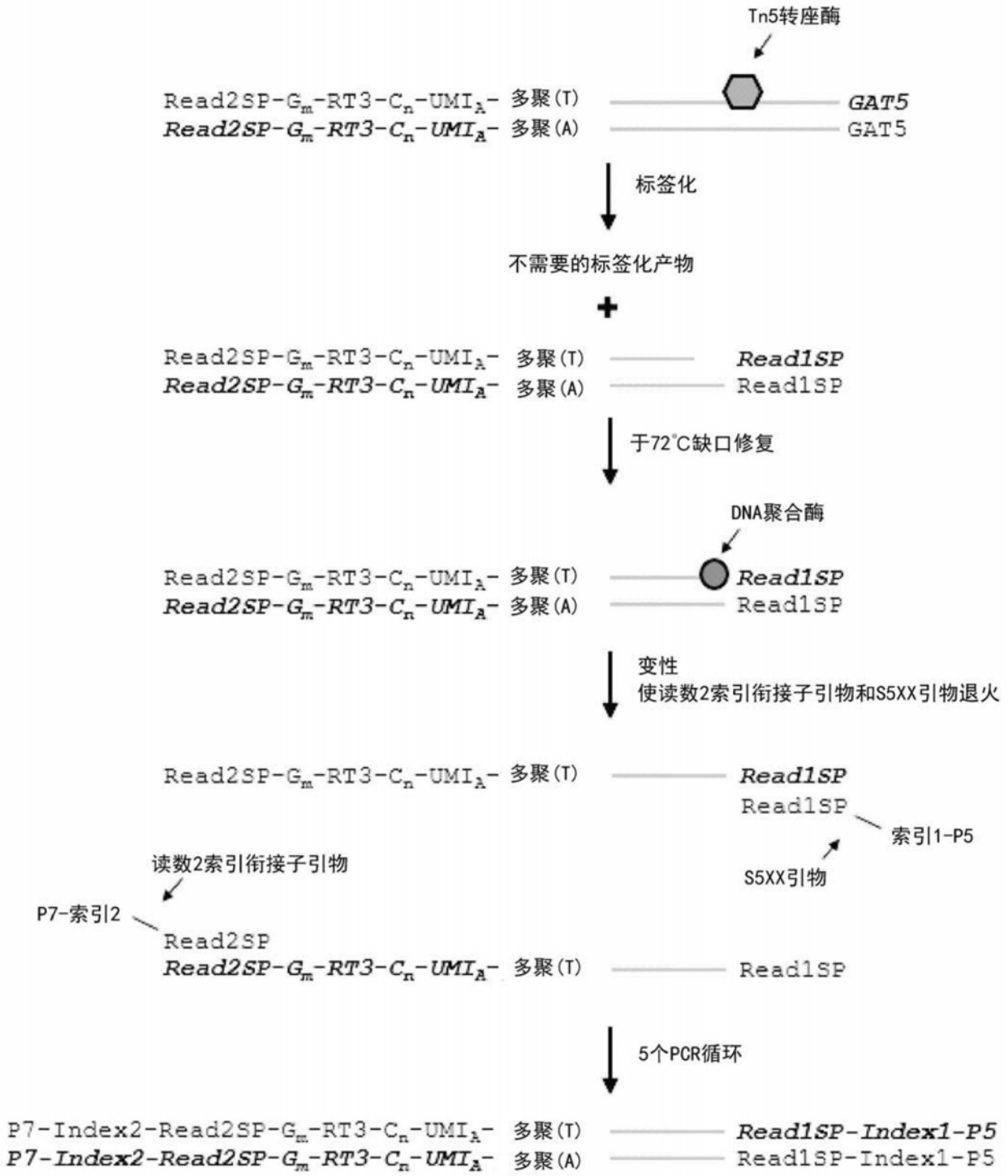

Tn5转座酶
Read2SP-Gm-RT3-Cn-UMIA- 多聚(T)
Read2SP-Gm-RT3-Cn-UMIA- 多聚(A)
GAT5
GAT5
标签化
不需要的标签化产物
Read2SP-Gm-RT3-Cn-UMIA- 多聚(T)
Read2SP-Gm-RT3-Cn-UMIA- 多聚(A)
Read1SP
Read1SP
于72°C缺口修复
DNA聚合酶
Read2SP-Gm-RT3-Cn-UMIA- 多聚(T)
Read2SP-Gm-RT3-Cn-UMIA- 多聚(A)
Read1SP
Read1SP
变性
使读数2索引衔接子引物和S5XX引物退火
Read2SP-Gm-RT3-Cn-UMIA- 多聚(T)
Read1SP
Read1SP
索引1-P5
S5XX引物
读数2索引衔接子引物
P7-索引2
Read2SP
Read2SP-Gm-RT3-Cn-UMIA- 多聚(T)
Read1SP
5个PCR循环
P7-Index2-Read2SP-Gm-RT3-Cn-UMIA- 多聚(T)
P7-Index2-Read2SP-Gm-RT3-Cn-UMIA- 多聚(A)
Read1SP-Index1-P5
Read1SP-Index1-P5

图3

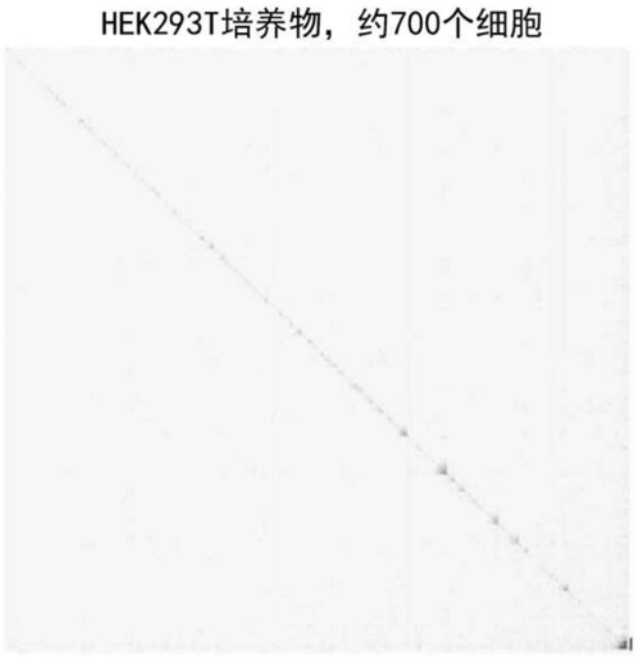


图 4A

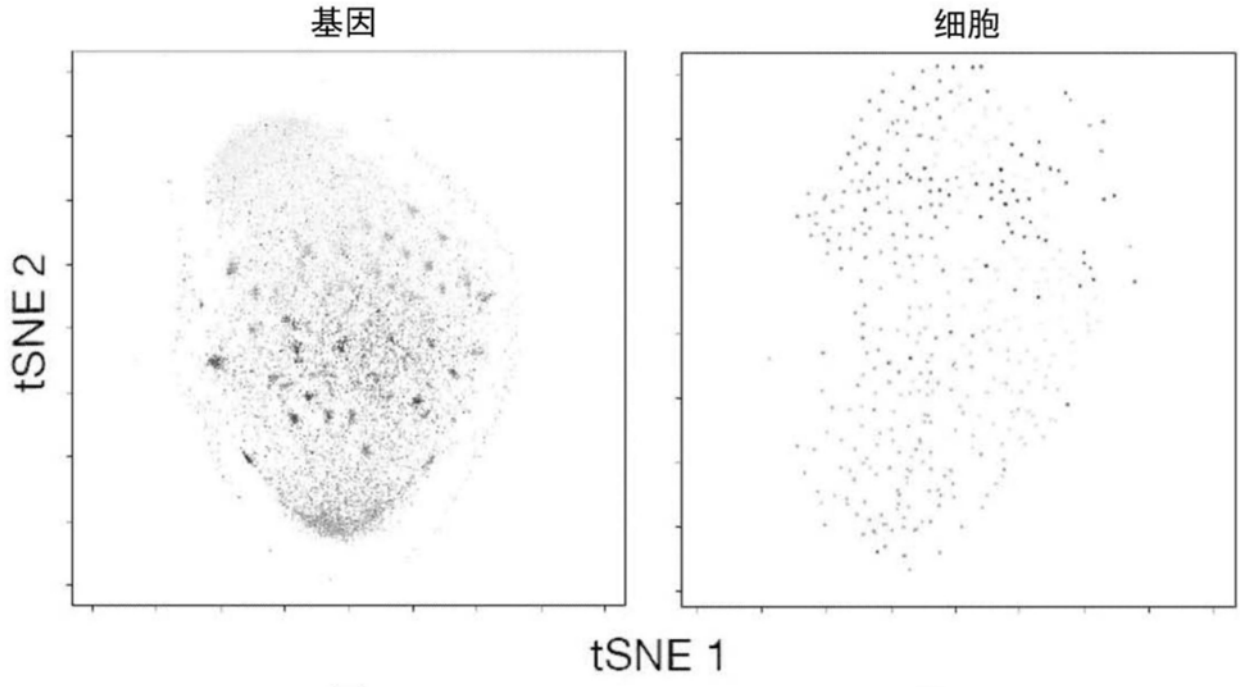


图 4B 图 4C

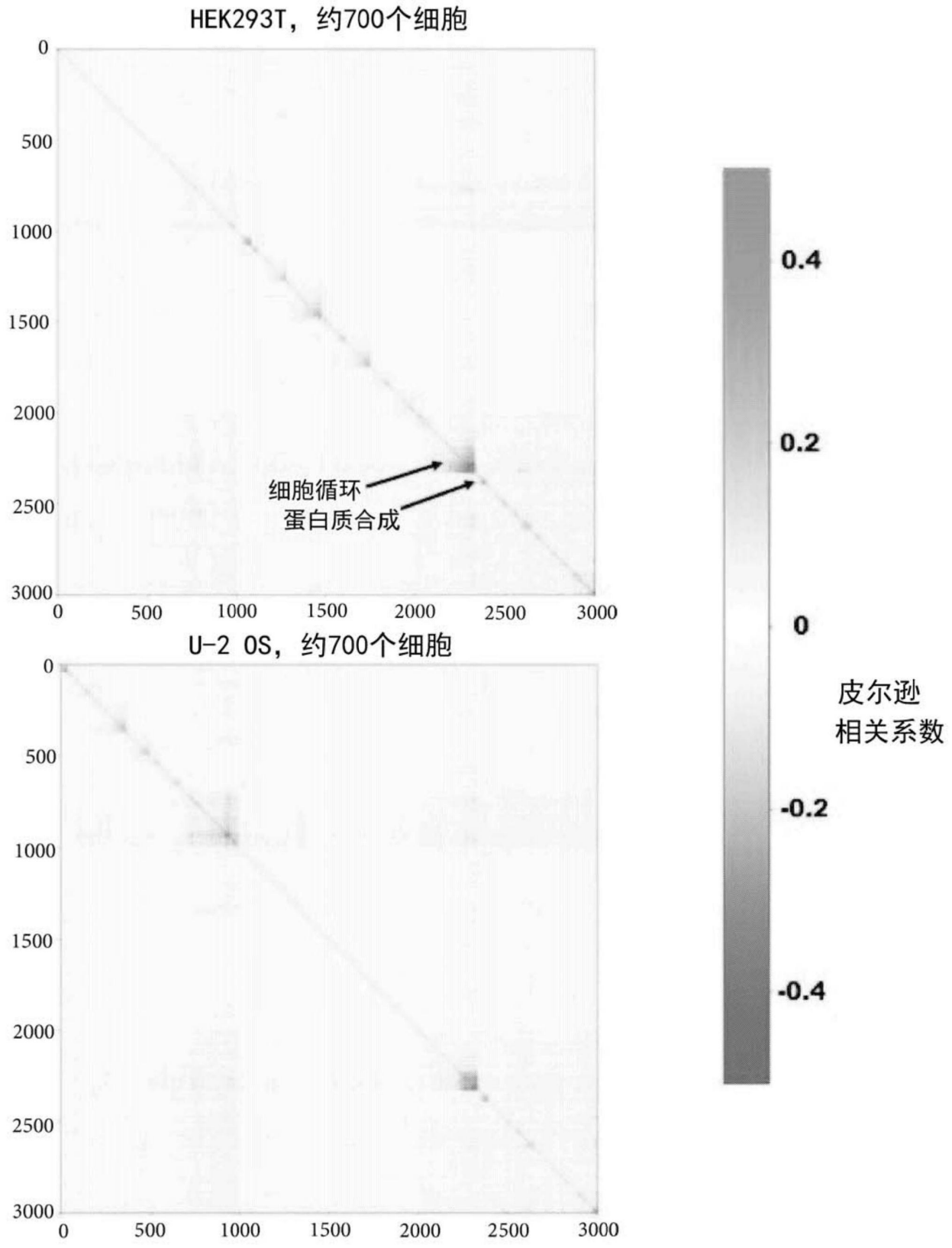
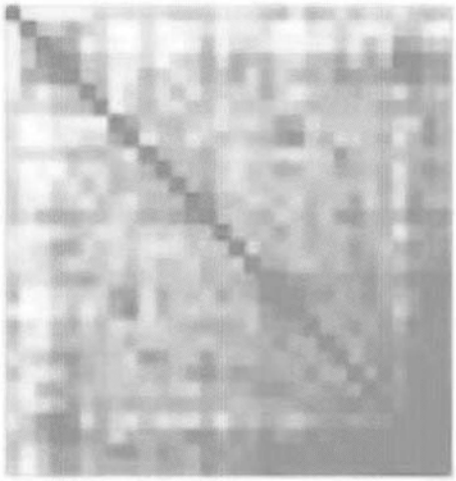

HEK293T，约700个细胞
0
500
1000
1500
2000
2500
3000
细胞循环
蛋白质合成
U-2 OS，约700个细胞
0.4
0.2
0
-0.2
-0.4
皮尔逊
相关系数

图5

蛋白质合成



氨基酸转运

SLC1A5	溶质运载体家族1成员5
SLC38A1	溶质运载体家族38成员1
SLC38A2	溶质运载体家族38成员2
SLC6A9	溶质运载体家族6成员9
SLC7A11	溶质运载体家族7成员11

tRNA合成

EPRS	谷氨酰脯氨酰tRNA合成酶
GARS	甘氨酰-tRNA合成酶
IARS	异亮氨酰tRNA合成酶
LARS	亮氨酰tRNA合成酶
MARS	甲硫氨酰-tRNA合成酶
NARS	天冬酰胺基tRNA合成酶
SARS	丝氨酰tRNA合成酶
TARS	苏氨酰-tRNA合成酶
YARS	酪氨酰tRNA合成酶

氨基酸分析

ASNS	天冬酰胺合成酶（谷氨酰胺水解）
BCAT1	支链氨基酸转氨酶1
PSAT1	磷酸丝氨酸转氨酶1
SHMT2	丝氨酸羟甲基转移酶2
CHAC1	ChaC谷胱甘肽特异性γ-谷氨酰环转移酶1
GPT2	谷氨酸丙酮酸转氨酶2
PHGDH	磷酸甘油酸脱氢酶
MTHFD1L	亚甲基四氢叶酸脱氢酶1样
MTHFD2	亚甲基四氢叶酸脱氢酶2

翻译起始的控制

EIF1	真核翻译起始因子1
EIF4EBP1	真核翻译起始因子4E结合蛋白1
CEBPG	CCAA T/增强子结合蛋白γ

图6

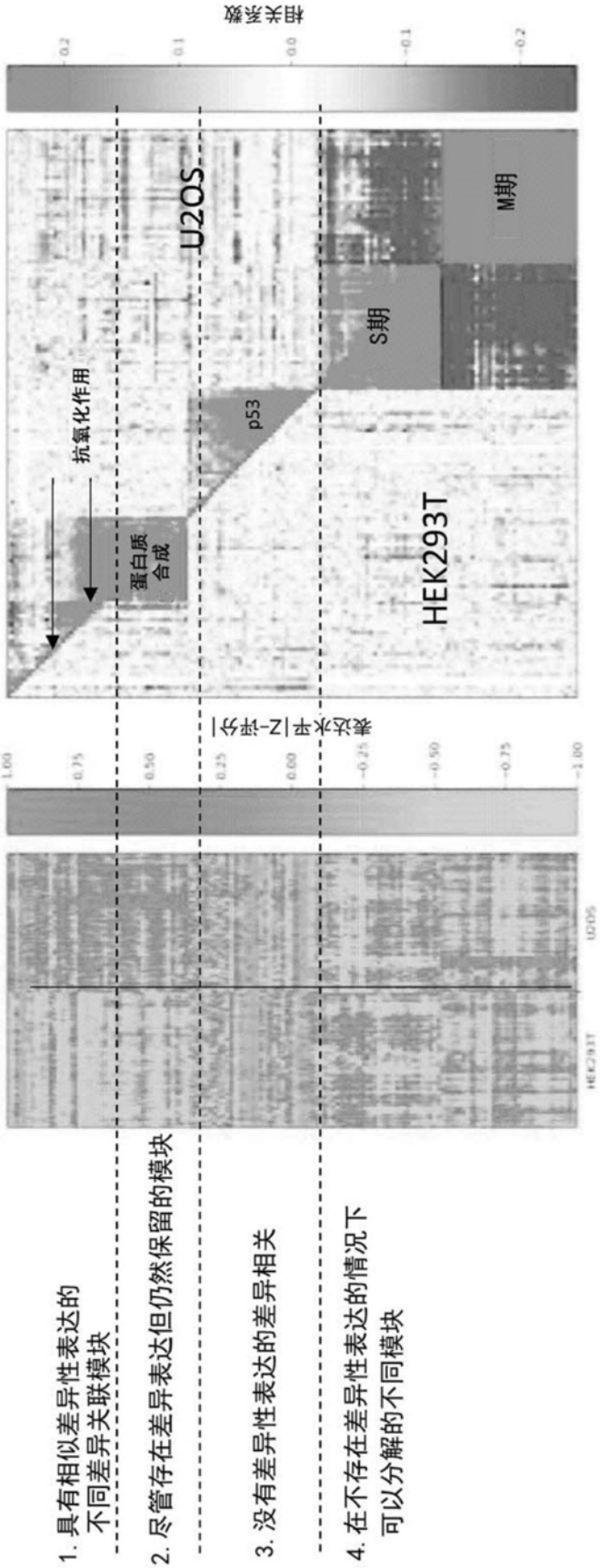

相关系数
抗氧化作用
蛋白质合成
U2OS
p53
S期
M期
HEK293T
细胞周期
1. 具有相似差异性表达的不同差异关联模块
2. 尽管存在差异表达但仍然保留的模块
3. 没有差异性表达的差异相关
4. 在不存在差异性表达的情况下可以分解的不同模块

图7