

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2024 年 12 月 12 日 (12.12.2024)



(10) 国际公布号
WO 2024/250155 A1

(51) 国际专利分类号:

C12Q 1/6806 (2018.01) *C12Q 1/6886* (2018.01)

C12Q 1/6869 (2018.01) *C40B 50/06* (2006.01)

C12Q 1/6883 (2018.01) *C40B 40/06* (2006.01)

(21) 国际申请号: PCT/CN2023/098377

(22) 国际申请日: 2023 年 6 月 5 日 (05.06.2023)

(25) 申请语言: 中文

(26) 公布语言: 中文

(71) 申请人: 清华大学 (TSINGHUA UNIVERSITY) [CN/CN]; 中国北京市海淀区清华园 1 号, Beijing 100084 (CN)。

(72) 发明人: 张强锋 (ZHANG, Qiangfeng); 中国北京市海淀区清华园 1 号, Beijing 100084 (CN)。唐磊 (TANG, Lei); 中国北京市海淀区清华园 1 号, Beijing 100084 (CN)。

(74) 代理人: 北京知联天下知识产权代理事务所 (普通合伙) (BEIJING IPLING INTELLECTUAL PROPERTY LAW FIRM); 中国北京市海淀区知春路 56 号天利大厦 5 层, Beijing 100098 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG,

BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

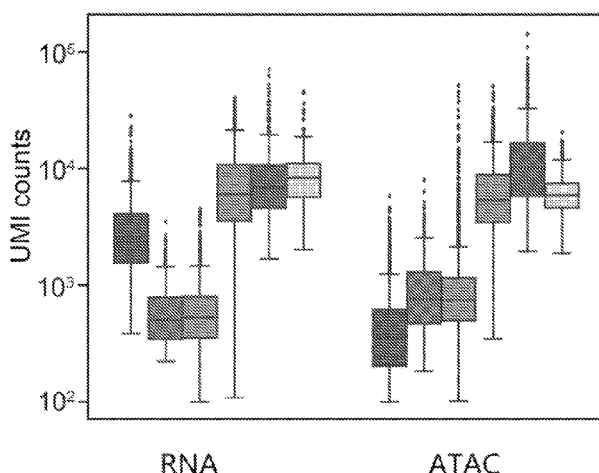
(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

(54) Title: METHOD FOR CONSTRUCTING SINGLE CELL SEQUENCING LIBRARY

(54) 发明名称: 一种构建单细胞测序文库的方法



[图 19]

(57) Abstract: Provided is a method for constructing a single cell sequencing library. Sequencing is performed on a constructed single cell sequencing library to achieve pairwise analysis of RNA and chromatin accessibility in a same single cell.

(57) 摘要: 提供了一种构建单细胞测序文库的方法, 并对构建的单细胞测序文库进行测序, 实现在同一单细胞中的 RNA 和染色质可接近性的配对分析。

WO 2024/250155 A1

说明书

发明名称: 一种构建单细胞测序文库的方法

技术领域

[0001] 本发明涉及基因测序技术领域，具体涉及通过测序对同一单细胞中的RNA和染色质可接近性进行配对分析的方法。

背景技术

[0002] 单细胞测序技术从单细胞RNA-seq发展到超高通量、多模态单细胞测序。G&T-seq检测同一细胞中的单细胞基因组和转录组。ScTrio-seq分析单个哺乳动物细胞的基因组、DNA甲基化和转录组之间的关系，CITE-seq同时测量单细胞中的表位和转录组。

[0003] 在多模式单细胞测序技术中，sci-CAR、SNARE-Seq、Paire-Seq、SHARE-Seq和Chromium单细胞多组学ATAC+基因表达试剂盒可在同一单细胞中定位染色质和RNA。这些方法解剖组织异质性并揭示相关的表观基因组调控元件。然而，sci-CAR条码结合低且碰撞率高，Paired-Seq在每管细胞数量过多时，其标记和逆转录反应效率均不理想。SHARE Seq需要定制测序来读取ATAC Seq库的两个片段，增加了测序成本。SNARE-Seq利用Drop-Seq系统将带有DNA条码微珠的标记细胞封装在纳升液滴中，出现了细胞产量低(每次实验10k)和很高的同一barcoded标记2个以上细胞的比率(11.3%)。各单细胞多组ATAC+基因表达试剂盒获得了最佳的联合分析数据，但成本高，通量与SNARE-Seq相似。

[0004] 在基于液滴的单细胞测序(dsc-seq)方法中，当加载到同一液滴中时，两个细胞将获得相同的条码，称为doublet，这会影响单细胞数据分析。与它产生的液滴相比，Dsc-seq加载的细胞数量要少得多，以避免doublet。例如，10x Genomics Chromium平台可产生约100k个含有条形码微珠和条形码试剂的液滴，但只能以约10%的碰撞率回收10k个单细胞。80%以上的功能液滴从未接收到单个细胞，浪费大部分试剂，并导致其大规模研究的高昂成本。

[0005] 因此，本申请提出了一种超高通量多模式单细胞技术，该技术并行测量同一细胞中的基因表达和染色质可接近性，称为(Parallel-seq)。

[0006] 发明内容

[0007] 本发明提供了一种单细胞超高通量双组学技术(单细胞组合流体标记(scifi)), 可以同时测量同一细胞的基因表达和染色质可接近性。与以往的多模式单细胞分析方法相比, Parallel-Seq仅通过一轮连接反应和两轮扩增反应进行四轮条码索引, Parallel-Split-Seq仅通过两轮连接反应和一轮扩增反应进行四轮条码索引, 实现了在同一单细胞中对开放染色质和基因表达进行联合分析, 可以对调节基因表达的顺式调节元件进行反卷积。用几个人类和小鼠细胞系对Parallel-Seq及Parallel-Split-Seq进行了基准测试, 并将其应用于人类肺癌样本的原代细胞。结果显示, 文库的数据特异性好, 质量高, 捕获的基因数量多。而且, 有很少的doublets, 碰撞率极低。本申请的构建单细胞测序文库的方法具有超标记组合空间能够以更低的成本执行大型细胞图谱项目。

[0008] 本发明的第一方面, 提供了一种构建单细胞测序文库的方法, 所述的方法包括利用转座子切割开放染色质获得携带第一接头的DNA片段; 加入逆转录引物对mRNA进行逆转录获得携带第二接头的cDNA第一条链, 获得在同一细胞中构建染色质DNA文库和转录组文库。

[0009] 优选的, 所述的方法还包括将细胞置于载体上, 利用第一载体特异性接头分别将上述获得的携带第一接头的DNA片段和上述获得的cDNA第一条链连接至载体上。

[0010] 优选的, 所述的方法还包括合成cDNA第二条链。

[0011] 优选的, 所述的方法还包括形成液滴, 裂解细胞并在液滴中进行扩增反应, 优选的, 形成的液滴中过载细胞。

[0012] 优选的, 所述的方法还包括纯化DNA并用引物分别扩增转录组文库的cDNA和染色质DNA。

[0013] 优选的, 所述的方法还包括加入RNA酶。

[0014] 优选的, 所述的方法还包括获得细胞, 将细胞固定并透化。

[0015] 在本发明的一个具体实施方式中, 一种构建单细胞测序文库的方法, 包括:

[0016] a) 利用转座子切割开放染色质获得携带第一接头的DNA片段;

[0017] b) 加入逆转录引物对mRNA进行逆转录获得携带第二接头的cDNA第一条链;

[0018] c)将细胞置于载体上，利用第一载体特异性接头分别将步骤a)获得的携带第一接头的DNA片段和步骤b)获得的cDNA第一条链连接至载体上；

[0019] d)合成cDNA的第二条链；

[0020] e)用引物分别扩增转录组文库的cDNA和染色质DNA。

[0021] 所述的步骤a)与步骤b)可以同时进行，或者先后进行。例如可以先进行步骤a)再进行步骤b)，或者先进行步骤b)再进行步骤a)。

[0022] 优选的，先进行步骤a)再进行步骤b)。

[0023] 优选的，包含大于10个转座子、大于100个转座子、大于1000个转座子、大于10000个转座子等等。

[0024] 所述的转座子包含条形码序列和转座酶。

[0025] 所述的转座酶包括但不限于Tn5转座酶、Mu转座酶、Tn7转座酶或IS5转座酶。在本发明的一个具体实施方式中，所述的转座酶为Tn5转座酶。所述Tn5转座酶携带序列如SEQ ID NO: 1或12所示。

[0026] 所述的条形码序列包含第一接头。进一步优选的，所述的条形码序列包含第一索引。所述的第一接头包含第一索引和转座酶结合位点。

[0027] 所述的第一接头包含至少一个相同或不同的接头。进一步优选的，所述的第一接头包含至少4个相同或不同的接头。在本发明的一个具体实施方式中，包含至少4-96个相同或不同的接头。

[0028] 所述的条形码序列从5' -3' 依次为突出端、第一索引和转座酶结合位点。所述的突出端为与后续引物互补的序列。

[0029] 优选的，所述的第二接头包含至少一个相同或不同的接头。进一步优选的，所述的第二接头包含至少4个相同或不同的接头。

[0030] 所述的逆转录引物包含第二接头，所述的第二接头包含poly(T)和第一索引；优选还包含随机六聚体引物。

[0031] 在本发明的一个具体实施方式中，所述的逆转录引物包含poly(T)和第一索引以及后续引物互补的序列。

[0032] 在本发明的一个具体实施方式中，所述的第一接头与第二接头可以包含相同的后续引物互补的序列。

- [0033] 在本发明的一个具体实施方式中, 所述的第一索引包含AACAAC、ACCGCA、AGTTGG、CCACGT、CGTGTT、GTTCTC、TGACTA、TCAAGG、AACGGT、AAGCCT、ACATGA、ACTCTA、AGAAGT、AGTACC、ATGCGA、CAATAG、CATCCA、CCTGGA、CGAGAC、CGCTCA、GCGTAA、GGATCG、GTGAGG、TCCTTA、TCTGCC、TTAACC或TTAGTG中的至少一个、两个或三个以上的组合。
- [0034] 在本发明的一个具体实施方式中, 所述的条形码序列包含SEQ ID NO: 2分别与SEQ ID NO: 1或12杂交后的至少一个、两个或三个以上的组合。
- [0035] 在本发明的一个具体实施方式中, 所述的逆转录引物包含SEQ ID NO: 3、4中的至少一个、两个或三个以上的组合。
- [0036] 所述的第一载体特异性接头包含第二索引。
- [0037] 所述的第一载体特异性接头包含UMI。
- [0038] 优选的, 所述的第二索引包含AAGACCAA、AAGCTACG、AAGGTCAT、AATAGTGG、AATGCCTT、ACAATAGC、ACAGGATT、ACCGACCT、ACCTAGAT、ACGAGTCC、ACGGACGA、ACGTTCAA、ACTATCTG、ACTCCGAA、AGAACAGA、AGACGCTT、AGATGCGA、AGCCACTC、AGCGAAGC、AGGTAACG、AGTACATC、AGTGATTC、ATAAGAGG、ATATCACG、ATCGCCGT、ATGACGGA、ATGGAATG、ATTCCTAC、CAACGCCA、CAAGTCTG、CACACATC、CACCTTAT、CAGAACCT、CAGCCGAT、CATACTGT、CATCCACC、CATTGAGC、CCAAGCGT、CCACGACT、CCATTGTC、CCGCATGT、CCTACTCC、CCTCCTTG、CCTTAATG、CGAATATC、CGAGAGCA、CGCCTCAA、CGCGTTAC、CGGACTCT、CGGTTGTT、CGTAGCTT、CGTGCCAA、CTACCGGA、CTAGCAGT、CTCAGCCT、CT

CTTCTA、CTGCTGGT、CTGTATTC、CTTCGCTC、GAAGAGTA、GACACCTA、GACGTGAG、GACTTACT、GAGGACAA、GAGTTAAG、GATCCTCG、GCAATCCG、GCAGTGTG、GCCGCTAA、GCGACCAT、GCTAAGAC、GCTGTAGG、GGAAGTGG、GGACAGTT、GGATTGCT、GGTCCTAA、GTACCTGT、GTCAAGGA、GTCTGCTT、GTGCTCCA、GTGTGACC、GTTATTGG、TAATTCGG、TACCAATC、TAGACTCC、TAGTCAAC、TCACGTTG、TCAGAATG、TCCAGCTT、TCCTGCGA、TCGGTTCC、TCTTACCT、TGACATGG、TGCCTATA、TGGTGTGG、TGTACTAG中的至少一个、两个或三个以上的组合。

[0039] 优选的，所述的第一载体特异性接头包含第二索引、UMI以及与逆转录引物或者转座子序列互补的序列。

[0040] 在本发明的一个具体实施方式中，所述的第一载体特异性接头从5' -3' 依次为与逆转录引物或者转座子序列互补的序列、UMI、第二索引和与载体上包含的序列互补的序列。

[0041] 在本发明的一个具体实施方式中，第一载体特异性接头包含SEQ ID NO: 6。

[0042] 在本发明的一个具体实施方式中，所述载体上包含SEQ ID NO: 5。

[0043] 在本发明的另一个具体实施方式中，第一载体特异性接头包含SEQ ID NO: 15。

[0044] 在本发明的另一个具体实施方式中，所述载体上包含SEQ ID NO: 13。

[0045] 所述的方法还包括形成液滴，裂解细胞并在液滴中进行扩增反应的步骤，优选的，形成的液滴中过载细胞。使液滴过载，使所有功能液滴都被使用，大大提高了微流体设备的通量。在液滴中进行线性扩增避免了未扩增产物的纯化，并且可以轻松地结合CRISPR筛选、DNA甲基化分析、蛋白质表达分析，这可能会导致单细胞跨组学测序甚至单个细胞的全组学测序。

[0046] 优选的，所述在液滴中进行扩增反应使用的引物包含第三索引。

[0047] 在本发明的一个具体实施方式中，在液滴中进行扩增反应使用的引物包含SEQ ID NO: 8。

[0048] 优选的，线性扩增后还包括裂解液滴的步骤。在本发明的一个具体实施方式中，所述的裂解液滴为采用破乳剂裂解。

[0049] 优选的，所述的方法包括利用第二载体特异性接头分别将上述获得的携带第一接头的DNA片段和上述获得的cDNA第一条链连接至载体上。优选的，所述的第二载体特异性接头包含第三索引。

[0050] 在本发明的一个具体实施方式中，第二载体特异性接头包含SEQ ID NO: 16。

[0051] 在本发明的一个具体实施方式中，所述载体上包含SEQ ID NO: 14。

[0052] 在本发明的一个具体实施方式中，所述的第三索引包含AACCTCTT、A

ACGTCGC、AAGAATCG、AAGCGGTG、AAGGAGCT
、AATACCGC、AATCTCCA、ACAACTTC、ACACGC
AA、ACCACAGT、ACCGTGTA、ACCTTGCC、ACGC
ATAA、ACGTATGG、ACTAACCA、ACTCAGGT、AC
TTGTTG、AGAAGTAC、AGAGATGA、AGATTAGG、
AGCCTGGT、AGCTCTAA、AGGTGTCT、AGTCCGT
T、AGTTCGCA、ATAAGCTC、ATCCATGA、ATCTA
GCG、ATGCAACC、ATGTGCAG、ATTGGTAG、CAA
GAAGA、CAATGGAC、CACATGCT、CACGGTAG、C
AGAGGTT、CAGTATAG、CATCAAGT、CATGTTCC
、CCAACAAT、CCAATTAC、CCAGTGAA、CCGATC
AG、CCGGTCTT、CGACAACG、CGCCAGTA、CGCG
GAAT、CGGAAGGA、CGGTGAGA、CGTAACAC、CG
TCTATG、CGTTCTCG、CTACTAAG、CTAGTGCG、
CTCTGACA、CTGATGAA、CTGGTACA、CTTACGA
G、GAACTCAA、GAATGTTG、GACGAATT、GACTG
CCA、GAGCTATT、GAGTCGGA、GATAGAAC、GAT
GGTCT、GCAGCACT、GCATTCAT、GCCTCTGT、G

CGCAGAT、GCTCACAA、GCTTGCGT、GTAATGCA
、GTATCGAG、GTCGATCT、GTGAGCGT、GTGGAT
AG、GTTAGCCA、TAAGGTGG、TACACCGG、TACT
CGTC、TAGCTGAG、TCAACAGG、TCACTCAC、TC
ATAGAC、TCCGTACA、TCGGAGTA、TCGTCGGT、
TGAACGCG、TGAGTCTT、TGC GACTG、TGGTTAT
C、TGTGTAAG、TTAGGAAC、TTCAGTGG、TTCTATCC中的至少一个、
两个或三个以上的组合。

- [0053] 优选的，所述的方法还包括纯化DNA的步骤。
- [0054] 所述纯化DNA后进行的扩增反应中的引物包含第四索引。
- [0055] 优选的，所述的第四索引包含P3_{xx}索引中的至少一个、两个或三个以上的组合；
- [0056] 优选的，所述的第四索引包含N7_{xx}中的至少一个、两个或三个以上的组合；
- [0057] 优选的，所述的第四索引包含P5_{xx}中的至少一个、两个或三个以上的组合；
- [0058] 优选的，所述的第四索引包含N5_{xx}中的至少一个、两个或三个以上的组合。

P3xx 索引	N7xx 索引	P5xx 索引	N5xx 索引
TCATCATT	TCGCCTTA	TATAGCCT	TAGATCGC
GCTCCTGT	CTAGTACG	ATAGAGGC	CTCTCTAT
CCTAACGT	TTCTGCCT	CCTATCCT	TATCCTCT
CACGTAGT	GCTCAGGA	GGCTCTGA	AGAGTAGA
GTAAGAGT	AGGAGTCC	AGGCGAAG	GTAAGGAG
TACCTTCT	CATGCCTA	TAATCTTA	ACTGCATA
AAGTCTCT	GTAGAGAG	CAGGACGT	AAGGAGTA
AGAGATCT	CCTCTCTG	GTACTGAC	CTAAGCCT
CCAGCGCT	AGCGTAGC		CGTCTAAT
ATGAACCT	CAGCCTCG		TCTCTCCG
CTATAACT	TGCCTCTT		TCGACTAG
TCTTATAT	TCCTCTAC		TTCTAGCT
GAATCGAT			CCTAGAGT
AATAAGAT			GCGTAAGA
ATICTAAT			CTATTAAG
ATCATATG			AAGGCTAT
GCTTAATG			GAGCCTTA
GGATATGG			TTATGCGA
CTTGATGG			
AAGATCGG			
TTAACCGG			
CTAAGTCG			
TATTCGCG			
GAAGCACG			
TCCAGTAG			
TTGTCTAG			
AGCGCTAG			
CCTGTGAG			
ATAGGAAG			
GATGGTTC			

CCACATTC			
TCTTGGTC			
CGAGGATC			
CACTAATC			
TAAGGCGC			
ACTGTTCC			
CTTCCTCC			
TACAGGCC			
TAAATACC			
ATAACACC			
CTACGCAC			
AGGTTAAC			
GTTGCAAC			
CTCAATTA			
ACAACCTA			
CTACCATA			
GACACATA			
CGCGGTGA			
TACTATGA			
TTGTTGGA			
AGTGAGGA			
ATCGCCGA			
CTTATAGA			
CCATGAGA			
TCACCTCA			
ATACTCCA			
GTTCGACA			
CATCATAA			
CACATGAA			
ATGAGGAA			
ATCTATCG			
TTACTGTT			
TAGGAATA			
CCGCACAG			
ATTGCGTT			

TCGTAAGC			
GGACTGCG			
AAGGATGC			
GCAATGGC			
GTATTCTC			
GTCATTAC			
GGTATACT			
TTGCGTGC			
GCCAGCGA			
GGCGCGCA			
CGCCACCT			
CGGTATCA			
ACGCAGCC			
GAGCTGAC			
TGATGGCT			
GCGTGGAG			
ACTTCCGC			
ACATGTAC			
GACGATAT			
CCTGGCCA			
AGCGTGTC			
CCGTTGTT			
TGTAACCA			
AGTAGTTA			
GTCTGCCT			
TGCGCAGC			
TGCCTATA			
TGCTAGTG			
CCGAGCTC			
GA CTGAGG			
TGGAATTC			

[0059] 为增加第四索引(例如P3xx索引), 扩增转录组所用引物为SEQ ID NO:

9、10。

[0060] 为增加第四索引(例如P5xx), 扩增转录组所用引物为SEQ ID NO: 20、18。

[0061] 为增加第四索引(例如N7xx), 扩增开放染色质片段所用引物为SEQ ID NO:

9、11。

- [0062] 为增加第四索引(例如N5xx), 扩增开放染色质片段所用引物为SEQ ID NO: 20、19。
- [0063] 在本发明的一个具体实施方式中, 所述的载体包括孔、管或平板。
- [0064] 优选的, 所述的载体为酶标板例如96孔板。
- [0065] 优选的, 所述的方法还包括加入RNA酶。通过RNase酶切反应, 从第一链cDNA中去除RNA, 然后用随机引物进行第二链合成, 避免了开放染色质片段被0.1N NaOH破坏和RNA-seq库被污染。
- [0066] 优选的, 所述的方法还包括获得细胞, 将细胞固定并透化。
- [0067] 本发明的第二方面, 提供了一种多模式单细胞测序文库的构建方法, 所述的构建方法包括按照上述构建单细胞测序文库的方法。
- [0068] 本发明的第三方面, 提供了一种构建转录组文库的方法, 所述的方法包括加入逆转录引物对mRNA进行逆转录获得携带第二接头的cDNA第一条链; 将细胞置于载体上, 利用第一载体特异性接头将获得的cDNA第一条链连接至载体上; 合成cDNA的第二条链; 纯化并用引物扩增转录组的cDNA。
- [0069] 优选的, 所述的逆转录引物包含第二接头, 所述的第二接头包含poly(T)和第一索引; 优选还包含随机六聚体引物。
- [0070] 优选的, 所述的第一载体特异性接头包含第二索引。
- [0071] 优选的, 所述的方法还包括形成液滴, 裂解细胞并在液滴中进行扩增反应的步骤,
- [0072] 优选的, 形成的液滴中过载细胞;
- [0073] 优选的, 在液滴中进行扩增反应使用的引物包含第三索引。
- [0074] 优选的, 所述的方法包括利用第二载体特异性接头将获得的cDNA第一条链连接至载体上, 优选的, 所述的第二载体特异性接头包含第三索引。
- [0075] 优选的, 所述纯化DNA后进行的扩增反应中的引物包含第四索引。
- [0076] 优选的, 所述的方法还包括加入RNA酶。
- [0077] 本发明的第四方面, 提供了一种构建染色质DNA文库的方法, 所述的方法包括利用转座子切割开放染色质获得携带第一接头的DNA片段; 将细胞置于载

体上，利用第一载体特异性接头将获得的携带第一接头的DNA片段连接至载体上；纯化DNA并用引物分别扩增染色质DNA。

[0078] 优选的，所述的转座子包含条形码序列和转座酶；优选的，所述的条形码序列包含第一接头；进一步优选的，所述的条形码序列还包含第一索引。

[0079] 优选的，所述的第一载体特异性接头包含第二索引。

[0080] 优选的，所述的方法还包括形成液滴，裂解细胞并在液滴中进行扩增反应的步骤，

[0081] 优选的，形成的液滴中过载细胞；

[0082] 优选的，所述在液滴中进行扩增反应使用的引物包含第三索引。

[0083] 优选的，所述的方法包括利用第二载体特异性接头将获得的携带第一接头的DNA片段连接至载体上，优选的，所述的第二载体特异性接头包含第三索引。

[0084] 优选的，扩增染色质DNA所用的引物包含第四索引。

[0085] 本发明的第五方面，提供了一种上述的方法获得的核酸文库。

[0086] 本发明的第六方面，提供了一种核酸文库，所述的核酸文库包含至少一个片段DNA，所述的片段DNA包含至少一个索引，和至少一个独特分子标识。

[0087] 优选的，所述的索引为一个、两个、三个、四个、五个、六个、七个、八个、九个或十个以上。

[0088] 优选的，所述的索引包括第一索引、第二索引、第三索引和/或第四索引。

[0089] 在本发明的一个具体实施方式中，所述的核酸文库包含至少一个从5' 到3' 依次为第四索引、片段DNA、第一索引、第二索引、第三索引。

[0090] 优选的，所述的独特分子标识位于第四索引与片段DNA之间，片段DNA与第一索引之间，第一索引与第二索引之间或者第二索引与第三索引之间。

[0091] 本发明的第七方面，提供了一种测序方法，所述的测序方法包括构建上述的核酸文库。

[0092] 本发明的第八方面，提供了一种上述的核酸文库的应用，所述的应用包括肿瘤靶点筛选、疾病监测或植入前胚胎诊断。

- [0093] 本发明的第九方面，提供了一种同一细胞中分析染色质可接近性和转录的方法，所述的方法包括上述构建单细胞测序文库、上述构建转录组文库、上述构建染色质DNA文库的步骤。
- [0094] 本发明的第十方面，提供了一种单细胞多组学的分析方法，所述的分析方法包括构建单细胞测序文库、上述构建转录组文库、上述构建染色质DNA文库，并进行测序获得染色质可接近性和/或转录组序列信息，然后进行生物信息学分析。
- [0095] 本发明的第十一方面，提供了一种试剂盒，所述的试剂盒包括构建上述核酸文库所用的试剂。
- [0096] 本发明所述的“染色质可接近性”即真核生物染色质DNA在核小体或转录因子等蛋白与其结合后，对其他蛋白能否再结合的开放程度。其中，可以对其他蛋白再结合的区域即为开放染色质。
- [0097] 本发明所述的“载体”可以为任何具有固体支持物表面的物体，其表面可以经过修饰与细胞或核酸分子偶联。其可以为孔玻璃(CPG)、草酰-调孔玻璃、TentaGel支持物-一种氨基聚乙二醇衍生化支持物、聚苯乙烯，Poros(一种聚苯乙烯/二乙烯基苯的共聚物)或可逆交联的丙烯酰胺。很多其它固体支持物市售可得且适用于本发明。在一些实施方式中，可以为聚苯乙烯树脂或聚(甲基丙烯酸甲酯)(PMMA)。也可以是金属。
- [0098] 本发明所述的“液滴”为水包油或油包水结构。不同的液滴可以具有不同的标识。优选为水性混合物与油相合并。优选的，所述的油相为表面活性剂。
- [0099] 本发明所述的“透化”是指在不造成细胞裂解以及不破坏细胞内部有机结构的情况下改变细胞壁和细胞膜的通透性，使得小分子物质和一些较大分子物质能够自由地进出细胞的技术。细胞经过透性化处理后在提高通透性的同时，整体结构保持完整，对胞内酶仍具有相当的保护作用，可保证胞内酶催化作用的充分发挥，并延长酶的使用寿命。
- [0100] 本发明所述的“过载”为超过原有承载量。所述的原有承载量为现有技术中常规的承载量。例如“液滴中过载细胞”代表超过原有液滴中承载细胞的量。现有技术中液滴承载细胞包括空载、承载单个细胞或者过载细胞。其中过载细胞

代表一个液滴中承载的细胞数量超过一个。优选承载两个、三个、四个、五个、六个、七个、八个或九个以上。

[0101] 本发明所述的“接头”与现有技术中的adapter可以互换使用，其可以用于连接片段化的DNA与索引，或者连接索引与索引，或者连接片段化的DNA与片段化的DNA。其优选为一段长度为3-1000个碱基的核苷酸序列。

[0102] 本发明所述的“索引”与现有技术中的index、barcode等可以互换使用。所述的索引可以为一段序列或几段序列的组合。其优选为一段长度为3-1000个碱基的核苷酸序列。

[0103] 本发明所述的“独特分子标识”即Unique Molecular Identifier，简称UMI，其为随机设计的一段核苷酸序列，可以专一性的辨识其偶联的分子，但是并不是所有偶联的分子都具有唯一的UMI，在一个具体实施方式中，其与其他索引组合形成一个唯一的分子标识。

[0104] 本发明所述的“互补”是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。

[0105] 本发明所述的“单细胞”指单个细胞或一个细胞，其可以来自血液样本、细胞培养物，也可以来自特定组织、器官或肿瘤等等。然后再通过现有技术常规的分离方式，将其分离为单个细胞。

[0106] 本发明所述的“doublet”或“doublets”代表两个或三个以上的细胞共用一个标识的情况，所述的标识例如索引、接头、独特分子标识等等或他们的组合。

[0107] 本文所述的“核酸”表示DNA、RNA、单链、双链、或更高度聚集的杂交基序及其任意化学修饰。修饰包括但不限于，提供整合入其它电荷、极化性、氢键、静电相互作用、与核酸配体碱基或核酸配体整体的连接点和作用点的化学基团的那些修饰。这类修饰包括但不限于，肽核酸(PNA)、磷酸二酯基团修饰(例如，硫代磷酸酯、甲基磷酸酯)、2' -位糖修饰、5-位嘧啶修饰、8-位嘌呤修饰、环外胺处的修饰、4-硫尿核苷的取代、5-溴或5-碘-尿嘧啶的取代、骨

架修饰、甲基化、不常见的碱基配对组合如异碱基(iso bases)、异胞苷和异胍(isoguanidine)等。核酸也可包含非天然碱基,如硝基嘌呤。修饰还可包括3' 和5' 修饰,包括但不限于用荧光团(例如,量子点)或其他部分加帽。

[0108] 本发明所述的“和/或”包含该术语所连接的项目的所有组合,应视为各个组合已经单独地在本问列出。例如,“A和/或B”包含了“A”、“A和B”以及“B”。又例如,“A、B和/或C”包含了“A”、“B”、“C”、“A和B”、“A和C”、“B和C”以及“A和B和C”。

[0109] 本发明所述的“包含”或“包括”为开放式写法,当用于描述蛋白质或核酸的序列时,所述蛋白质或核酸可以是由所述序列组成,或者在所述蛋白质或核酸的一端或两端可以具有额外的氨基酸或核苷酸,但仍然具有本发明所述的活性。

[0110] 增加了细胞内第二链合成步骤,以减少交联蛋白抑制的影响,捕获更多的转录本。实现基于液滴标引的线性扩增和提高cDNA捕获效率。同时,提供了cDNA不同于染色质片段的PCR锚定接头,避免了ATAC-seq库污染RNA-seq库。

[0111] Parallel-seq使用多个细胞对液滴进行过载,以充分利用生成的液滴,并进行前后索引以区分一个液滴内的细胞,大大扩展了条形码空间。而且,条形码区域的长度明显降低,使其可以通过条形码和固定核苷酸区域读取150nt测序读取长度内的开放片段。按照设计,Parallel-Seq首先在转座和逆转录过程中用特定于样本的条形码对细胞进行散列,使其可以在一个实验中并行评估多个样本并具有可扩展性。Parallel-Seq在数据质量方面优于现有方法,通量增加(每个实验3600万个细胞),这为构建价格合理的大型单元图谱提供了强大的工具。此外,我们将Parallel-seq应用于肺癌样本,并证明了其在识别特定基因可及区域的顺式调控元件方面的能力。在肿瘤样本中应用了基因表达和染色质可接近性的联合分析,并利用联合分析和新开发的分析方法来识别可能的调控元件,包括致癌基因的增强子和突变。此外,Parallel-Seq易于在实验中处理更多样本,并可扩展到其他组学,如DNA甲基化、蛋白质表达和CRISPR筛选。

附图说明

[0112] 以下,结合附图来详细说明本发明的实施例,其中:

- [0113] 图1: Parallel-Seq的实验设计图, 使用索引、液滴过载来分析同一细胞的scATAC和scRNA, 其中, pool/split代表混合/分散。
- [0114] 图2: 使用NIH/3T3(鼠)、HEK293T(人)和K562(人)细胞的混合物进行Parallel-Seq, 结果映射到人类和小鼠基因组的scRNA-seq(上)和scATAC-seq(下)的UMI计数图, 其中, mm10代表小鼠参考基因组mm10版本。
- [0115] 图3: Parallel-Seq的scATAC-seq部分片段的插入片段长度分布。
- [0116] 图4: TSSs周围scATAC-seq reads的富集。
- [0117] 图5: 散点图显示了K562细胞中Parallel-Seq的scATAC-seq和ENCODE DNase-seq之间 $\log_2(\text{count})$ 的相关性。
- [0118] 图6: 散点图显示了中K562细胞Parallel-Seq的聚集scRNA-seq和ENCODE核RNA-seq之间 $\log_2(\text{TPM}+1)$ 的相关性。
- [0119] 图7: 在K562细胞中分别采用Parallel-Seq、ENCODE DNase-seq捕获染色质可接近性以及分别采用Parallel-Seq、ENCODE RNA-seq捕获RNA的对比结果。
- [0120] 图8: 来自3T3、293T和K562细胞混合的Parallel-Seq配对基因表达数据的均匀流形近似和投影(UMAP)可视化。
- [0121] 图9: 来自3T3、293T和K562细胞混合的Parallel-Seq配对染色质可接近性数据的均匀流形近似和投影(UMAP)可视化。
- [0122] 图10: 箱线图显示了sci-CAR、SNARE-seq、Paired-Seq、SHARE-seq和Parallel-Seq的唯一映射RNA reads的数量和唯一映射的ATAC reads的数量。其中, 横坐标RNA文库框图从左到右依次为sci-CAR、SNARE-Seq、Paired-Seq、SHARE-Seq、Parallel-Seq, ATAC文库框图从左到右依次为sci-CAR、SNARE-Seq、Paired-Seq、SHARE-Seq、Parallel-Seq。
- [0123] 图11: 箱线图显示了sci-CAR、SNARE-seq、Paired-Seq、SHARE-seq和Parallel-Seq中每个细胞捕获的的基因数量。其中, 横坐标RNA文库框图从左到右依次为sci-CAR、SNARE-Seq、Paired-Seq、SHARE-Seq、Parallel-Seq。
- [0124] 图12: Parallel-Split-Seq工作流程示意图。

- [0125] 图13: 映射到人类和小鼠基因组的scRNA-seq(左)和scATAC-seq(右)的UMI计数。该实验使用NIH/3T3(鼠)、HEK293T(人)、HeLa(人)、K562(人)和THP1(人)细胞的混合物进行Parallel-Split-Seq。
- [0126] 图14: Parallel-Split-Seq与Parallel-Seq中scATAC-seq片段的插入片段长度分布。
- [0127] 图15: Parallel-Split-Seq与Parallel-Seq中TSSs周围scATAC-seq reads的富集。
- [0128] 图16: 散点图显示了K562细胞中Parallel-Split-Seq的scRNA-seq和ENCODE核RNA-seq之间 $\log_2(\text{TPM}+1)$ 的相关性(图A)以及scATAC-seq和ENCODE核DNase-seq之间 $\log_2(\text{count})$ 的相关性(图B)。
- [0129] 图17: 来自NIH/3T3、HEK293T、HeLa、K562和THP1细胞的Parallel-Split-Seq配对基因表达(左)和染色质可接近性(右)数据的均匀流形近似和投影(UMAP)可视化。
- [0130] 图18: 在K562细胞中分别采用Parallel-Seq、Parallel-Split-Seq以及ENCODE DNase-seq捕获染色质可接近性的对比结果以及分别采用Parallel-Seq、Parallel-Split-Seq以及ENCODE RNA-seq捕获RNA的对比结果。
- [0131] 图19: 箱线图显示了sci-CAR、SNARE-seq、Paired-Seq、SHARE-seq、Parallel-Seq和Parallel-Split-seq的唯一映射RNA reads的数量和唯一映射的ATAC reads的数量。其中, 横坐标RNA文库框图从左到右依次为sci-CAR、SNARE-Seq、Paired-Seq、SHARE-Seq、Parallel-Seq和Parallel-Split-Seq, ATAC文库框图从左到右依次为sci-CAR、SNARE-Seq、Paired-Seq、SHARE-Seq、Parallel-Seq和Parallel-Split-Seq。
- [0132] 图20: 箱线图显示了sci-CAR、SNARE-seq、Paired-Seq、SHARE-seq、Parallel-Seq和Parallel-Split-seq中每个细胞捕获的的基因数量。其中, 横坐标RNA文库框图从左到右依次为sci-CAR、SNARE-Seq、Paired-Seq、SHARE-Seq、Parallel-Seq和Parallel-Split-Seq。

具体实施方式

- [0133] 下面将结合本发明实施例中的附图, 对本发明实施例中的技术方案进行清楚、完整地描述, 显然, 所描述的实施例仅是本发明的部分实施例, 而不是全部。

基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0134] 实施例中细胞培养方法：

[0135] HEK293T，HeLa-S3和NIH/3T3细胞在添加10%胎牛血清(P30-3302，PAN BIOTECH)的DMEM(C11995500BT，ThermoFisher)培养基中，37℃和5%CO₂的环境下培养。用PBS(C10010500BT，ThermoFisher)冲洗细胞，并在37℃下用1mL 0.25%胰蛋白酶EDTA(25200114，ThermoFisher)培养3-5分钟以分离细胞。K562细胞在添加10%胎牛血清的RPMI 1640(C11875500BT，ThermoFisher)培养基中，37℃和5%CO₂的环境下培养。通过离心收集分离的HEK293T、HeLa-S3和NIH/3T3细胞以及K562细胞悬浮液，用PBS洗涤并使用Countstar计数。

[0136] 实施例中肺癌样本制备：

[0137] 在解放军总医院采集新鲜非小细胞肺癌实体肿瘤组织，置于预冷MACS组织存储液(130-100-008，Miltenyi Biotec)(2-8℃)中。样品必须被MACS全部覆盖，并从医院运送到实验室。

[0138] 使用了4677 μL DMEM/F-12(11320033，ThermoFisher)、250 μL 2.5mg/ml Liberase TL(05401020001，Sigma Aldrich)至最终浓度250 μg/mL、23 μL 2mg/mL elastase(NC9301601，Worthington)至最终浓度9.2 μg/mL、50 μL 10mg/mL DNA酶(11284932001，Sigma Aldrich)至最终浓度为100 μg/mL的分离混合物。

[0139] 用剪刀将组织在1.5mL Eppendorf微量离心管中切碎成0.4mm以下的小块。解离混合物在37℃下培养，并在90转/分下水平旋转60分钟。通过70 μm细胞过滤器(15-1070，BIOLOGIX)过滤单细胞悬浮液，并在4℃下以500g(离心力)离心5分钟。用1mL PBS和3mL红细胞裂解液(4992957，TIANGEN)重悬细胞。室温孵育5min，4℃下500g离心5min。细胞重悬于500 μL胎牛血清中。取5 μL细胞，与5 μL台班蓝液(15250061，ThermoFisher)混合，用C-Chip一次性血球计(DHC-N01N，As One)计数。用添加二甲基亚砜(D2650，Sigma Aldrich)的胎牛血清稀释单细胞悬液至最终浓度10%。我们冷冻保存细胞，每个试管

含有 1×10^6 的单个细胞。实验前，细胞在 37°C 温和解冻5min，在 4°C 500g离心5min。用80 μL 细胞染色缓冲液(420201, BioLegend)和5 μL Human TruStainFcX(422302, BioLegend)重悬细胞， 4°C 孵育5分钟。分别加入5 μL anti-CD45 PE(304039, BioLegend)、5 μL anti-CD3 BV421(317344, BioLegend)和5 μL anti-EpCam PE/Cy7(324222, BioLegend)抗体，在 4°C 避光孵育15min。用1mL PBS洗涤染色细胞， 4°C 500g离心5min。丢弃上清，用含0.02 μM Calcein-AM(425201, BioLegend)的1mL PBS重悬单个细胞，避光室温孵育15分钟。分别用90 μL Annexin V Binding Buffer、5 μL APC Annexin V(640941, BioLegend)和5 μL 7-AAD活力染色液(420404, BioLegend)重悬细胞。室温孵育10分钟。加入400 μL PBS，用35 μm BD细胞滤器(352235, BD Falcon)过滤细胞。Caicein-AM阳性、7-AAD阴性、Annexin V阴性单细胞用MoFloAstrios EQ Cell Sorter(Beckman Coulter)分选。由于肿瘤细胞和T细胞可能是测序数据的主要部分，因此我们用低于5%的EpCam阳性细胞、40%的T细胞和55%的其他单细胞来平衡样本。

[0140] 实施例制备转座子的方法如下：

[0141] 准备Tn5Merev(/5Phos/CTGTCTCTTATACACATCT(SEQ ID NO: 21)、Tn5ME-A、Tn5ME-B和带条码的R1BxME(x代表1-96)。将10 μM Tn5Merev、10 μM Tn5ME-A(用于Parallel-Split-Seq)或10 μM Tn5ME-B(用于Parallel-Seq)分别与10 μM R1BxME在 95°C ~ 2分钟， $0.1^\circ\text{C}/\text{s}$ 下逐渐退火降至 20°C 和 4°C 。结合2 μL 退火Tn5ME-B(Parallel-Seq)、2 μL 退火的Tn5ME-R1Bx、2 μL 10x TPS、4 μL 转座酶(M0221, Robustnique)和10 μL UltraPure DNase/RNase-Free Water(10977023, ThermoFisher)在室温下孵育30分钟。组装的转座子每管分装4 μL ，在 -20°C 下储存不超过1个月。

[0142] 实施例固定细胞的方法如下：

[0143] 为细胞系数数50k单细胞，或为每管肺癌样本分类50k原代细胞。在 4°C 下以500g离心单细胞5min，并用250 μL PBS重悬单细胞。添加750 μL 含有1.33%无甲醇甲醛的PBS(28906, ThermoFisher)并在冰中孵育10分钟。添加50 μL 20% BSA(V0332-100G, VWR)并在 4°C 下用摆桶离心法在1000g下离心3分钟，然后在预冷

(4°C)的固定角离心机中离心将细胞收集到1.5mL微型离心管(MCT-150-C, Axygen)的一侧,并像omni ATAC一样,通过两个移液步骤去除上清液。结果显示加入BSA后,先用摆桶离心法分离单细胞,可以回收更多的原代细胞。

[0144] 实施例中转座的方法如下:

[0145] 将1mL 1M Tris HCl pH 7.4(T2663-1L, Sigma-Aldrich)、200 μ L 5M NaCl(AM9759, ThermoFisher)、300 μ L 1M MgCl₂(AM9530G, ThermoFisher)和48.5mL超纯DNase/RNase游离蒸馏水混合,制备2x RSB作为Omni ATAC。固定时,制备透化缓冲液,每个样品结合50 μ L 2x RSB、1 μ L核糖锁(EO0384, ThermoFisher)、1 μ L SUPERase • In RNase抑制剂(A M2696, ThermoFisher)、1 μ L 10% Nonidet P40替代物(1133247301, Sigma-Aldrich)、1 μ L 10% tween20(11332465001, Sigma-Aldrich)、1 μ L 1% Digitonin(D141-100MG, Sigma-Aldrich), 5 μ L 20% BSA, 40 μ L超纯DNase/RNase游离蒸馏水。通过将500 μ L 2x RSB、10 μ L 10% TWEEN20、1 μ L RiboLock、50 μ L 20% BSA。去除固定试剂后,立即添加100 μ L透化缓冲液,移液管8次,放入冰中孵育5分钟。透化后,向每个样品中添加1mL洗涤缓冲液。离心细胞并去除上清液。

[0146] 实施例中转座的方法如下:

[0147] 将10 μ L 5xLM缓冲液(M0221, Robustnique)、16.5 μ L PBS、0.5 μ L RiboLock、0.5 μ L SUPERase • In RNase Inhibitor、0.5 μ L 10% Tween 20、0.5 μ L Digitonin.5 μ L、17.5 μ L超纯DNase/RNase游离蒸馏水混合,制备ATAC seq反应溶液。用46 μ L ATAC-seq反应溶液重悬透化单细胞,并向每个试管中添加4 μ L条码特异性转座子。ATAC-seq反应在37°C、550r.p.m.条件下用热盖进行。ATAC-seq反应后,向每根试管中添加949 μ L PBS、10 μ L 10% Triton X-100(93443, Sigma-Aldrich)、1 μ L RiboLock和50 μ L 20% BSA,并离心以去除上清液。

[0148] 实施例中转座的方法如下:

[0149] 混合8 μ L PBS、0.5 μ L RiboLock、0.5 μ L SUPERase • In RNase Inhibitor、7 μ L无核酸酶水,为每个样品制备16 μ L重悬溶液。通过添加8 μ L 5x RT缓冲

液、2 μ L 10mM dNTP(N0447L、NEB)、0.5 μ L RiboLock、0.25 μ L SUPERase • In RNase Inhibitor、4 μ L Maxima H Minus逆转录酶(EP0753和ThermoFisher)和5.25 μ L UltraPure DNase/RNase-Free蒸馏水以制备逆转录混合物。将20 μ L逆转录混合物拆分到每个试管中，并添加条形码匹配的2 μ L随机和2 μ L polyT逆转录引物)。用16 μ L重悬溶液重悬转置细胞，并添加到条形码匹配的PCR管中。充分混匀，50℃反转录10分钟，然后经过3个热循环(8℃ 12秒、15℃ 45秒、20℃ 45秒、30℃ 30秒、42℃ 120秒和50℃ 180秒)，在50℃下孵育5分钟，并在4℃下永久保存。在冰上将逆转录反应合并到一个1.5mL试管中。离心细胞并去除上清液。用补充有10 μ L 10% Triton X-100和50 μ L 20% BSA的1mL PBS再次清洗细胞。离心细胞并去除上清液。

[0150] 实施例连接反应方法如下：

[0151] Parallel Seq使用连接反应添加第二索引。连接接头包含7nt互补链分别连接到转座子和逆转录引物，以及10nt索引链，8nt孔特异性接头，10nt的UMI，以及用于液滴线性扩增的通用PCR锚。在细胞内条形码连接之前，通过在100 μ L反应体积中结合11 μ M连接子链和12 μ M条形码链对连接接头进行退火。将平板在95℃下孵育2分钟，并以每秒-0.1℃的速率冷却至20℃，然后将培养板分成10个连接板，每个孔包含10 μ L连接接头。

[0152] 对于Parallel-Split-Seq，第二和第三索引通过连接反应添加。连接接头包含与接头链互补的10nt序列、8nt孔特异性接头和7nt序列，随后被连接。加入第三索引的连接反应的连接接头包含10nt索引链，8nt孔特异性接头、10nt的UMI、通用PCR引物的P3短序列。按照Parallel-Split-seq方案退火第二和第三轮转接器，并分别分成10个连接板。

[0153] 其中，细胞内连接步骤如下：

[0154] 连接反应按照Split-seq方案进行，不含RNase抑制剂。制备2mL 1xNEBuffer 3.1(B7203S，NEB)和2mL连接溶液(500 μ L 10x T4 DNA连接缓冲液，100 μ L T4 DNA连接酶(M0082，Robustnique)，50 μ L 10% Triton x-100和1350 μ L超纯DNA酶/RNA酶游离蒸馏水)。用1x buffer 3.1重悬组合的单细胞，并与连接溶液充分混合。将连接混合物中的40 μ L细胞添加到连接板的每个孔中。连接反

应在室温下15r.p.m旋转1小时。连接后，向每个孔中添加2 μ L 500 μ M EDTA(A M9260G, ThermoFisher)并合并。合并的细胞加入50 μ l 10% Triton X-100、50 μ L 20% BSA并离心去除上清液。用940 μ L PBS、10 μ L 10% Triton X-100和50 μ L 20% BSA再次清洗细胞。对于Parallel-Split-Seq，在第二索引标记后，通过连接反应添加第三索引。

[0155] 实施例1中RNase消化的方法如下：

[0156] 使用RNase消化反应(40 μ L 5xRT缓冲液、8 μ L RNase Cocktail Enzyme Mix(AM2286、ThermoFisher)、8 μ L RNase H(Y9220L, Enzymatics)和144 μ L UltraPure DNase/无RNase蒸馏水)重悬细胞并在37℃下孵育30分钟，300rpm 15秒后放置在混匀仪上45秒。加入790 μ L PBS和10 μ L 10% Triton X-100洗涤RNase消化反应，离心并去除上清液。此步骤中不要添加BSA。残留的BSA会在下一步中产生带有PEG8000的碎片。

[0157] 实施例1中第二链合成的方法如下：

[0158] 用第二条链合成反应混合物(40 μ L 5xRT缓冲液、48 μ l 50% PEG 8000(B1004SVIAL, NEB)、20 μ L 10mM dNTP、2 μ L 1mM dN-P3短引物(用于Parallel-Seq)或dN-P5短引物(用于Parallel-Split-Seq)、5 μ L Klenow Exo-(M0212L, NEB)和85 μ L UltraPure DNase/RNase-Free蒸馏水)在37℃下孵育1小时，以300r.p.m 15s后放置在混匀仪上45秒。第二条链合成后，用含有0.1% Triton X-100和1% BSA的PBS清洗细胞两次。然后用40 μ L 0.5xPBS重悬细胞，并用C-Chip一次性血细胞计数器在trypan blue中计数。

[0159] 使用10x chromium ATAC-seq kit过载

[0160] 对于Parallel-Seq，将细胞离心并用7 μ L ATAC-seq缓冲液B重悬，并用1xNucleies缓冲液补充至15 μ L。将56.5 μ L Barcoding Reagent B、1.5 μ L Reducing Agent B和2 μ L Barcoding Enzyme(PN-1000176, 10x Genomics)与细胞结合，并加载到Chromium Next GEM Chip H(PN-1000162, 10x Genomics)的一个通道中。GEM生成后，液滴被分成16个管，每个管包含6.25 μ L液滴。线性扩增如下进行：72℃ 5分钟，98℃ 30秒，然后98℃ 10秒，59℃ 30秒，72℃ 1分钟，12个循环。然后15℃保存待用。

[0161] Parallel-Seq文库构建

[0162] 随着Chromium Next GEM Single Cell ATAC Reagent Kits V1.1的缩小, 进行GEM孵育后的清理。将7.8 μ L Recovery Agent添加到每个管中, 轻轻颠倒管10次以混合。短暂离心并加入12.5 μ L Dynabeads Cleanup Mix。移液管混合5次并在室温下孵育10分钟。用81 μ L Elution Solution I洗脱产物并分成两部分, 40 μ L用于ATAC-seq, 40 μ L用于RNA-seq。使用1.2x SPRI beads(B23318, Beckman Coulter)清理ATAC-seq部分, 使用0.8x SPRI beads清理RNA-seq部分。使用SI-PCR引物B(PN-2000128, 10x基因组学)和N7xx引物放大ATAC-seq库。用SI-PCR引物B(PN-2000128)和P3xx引物放大RNA-seq库。扩增后, 用1.2x SPRI beads清理ATAC-seq部分, 用0.8x SPRI beads清理RNA-seq部分。

[0163] Parallel-Split-Seq文库构建

[0164] 将第二条链合成细胞稀释成800个细胞/ μ l, 并将每管2,000个细胞分开。加入2.5 μ l 2x裂解液(0.25 μ l 1M pH 8.0 Tris-HCl, 0.25 μ l 10% IGEPAL CA-630(I8896, Sigma Aldrich), 0.25 μ l 10% Tween 20, 0.5 μ l 291mg/ml QIAGEN Protease(19155, QIAGEN)和1.25 μ l UltraPure DNase/RNase-Free蒸馏水)并在55°C下孵育8小时, 在70°C下孵育15分钟以灭活QIAGEN Protease并在4°C下永久保持。添加45 μ l PCR扩增混合物(25 μ l NEBNext High-Fidelity 2X PCR Master Mix(M0541L, NEB)、2.5 μ l N5xx引物、1.25 μ l P5xx引物、1.25 μ l P3xx引物和15 μ l UltraPure DNase/RNase-Free蒸馏水)以扩增ATAC-seq和RNA-seq片段。循环条件为72°C 5分钟, 98°C 30秒, 然后98°C 10秒, 65°C 30秒, 72°C 1分钟循环5次, 保持在4°C。将PCR混合物分成ATAC-seq部分和RNA-seq部分。分别使用1.0x AMPure XP beads(A63881, Beckman Coulter)和0.8x AMPure XP beads清理RNA-seq部分。用22 μ l UltraPure DNase/RNase-Free蒸馏水洗脱PCR产物。通过添加28 μ l PCR反应混合物进行第二轮PCR扩增(25 μ l NEBNext High-Fidelity 2X PCR Master Mix、1.25 μ l N5xx引物、1.25 μ l P3_end引物、0.5 μ l 25x SYBR Green I(S7563, ThermoFisher)用于ATAC-seq,); 25 μ l NEBNext High-Fidelity 2X PCR Master Mix, 1.25 μ l N5xx引物, 1.25 μ l P3_end引物, 0.5 μ l 25x SYBR Green I, 用于RNA-seq)。我们在

QuantStudio 3实时PCR系统(ThermoFisher)上扩增了ATAC-seq和RNA-seq库的每个子库,跟踪扩增并在荧光单位值达到~100,000时停止每个子库。根据经验,细胞系或肿瘤细胞的循环数约为6-8,其他原代细胞的循环数约为7-10,小于11是可以接受的。分别使用1.0x AMPure XP beads清理ATAC-seq库和使用0.8x AMPure XP beads清理RNA-seq库。用20 μ l洗脱缓冲液(19086, QIAGEN)洗脱文库,并用Qubit dsDNAHS检测试剂盒(Q32851, ThermoFisher)确定浓度。每个子库预计将回收超过20ng产品。使用安捷伦高灵敏度D1000 Screen Tape Assay(5067-5584和5067-5585, 安捷伦)进行质量控制。

[0165] 实施例测序方法如下:

[0166] 使用具有16nt i5索引、8nt i7索引和PE150测序的Illumina NovaSeq 6000测序系统对Parallel-Seq文库进行测序。

[0167] 使用Illumina HiSeq X 10系统或NovaSeq 6000测序系统对Parallel-Split-Seq文库进行测序,标准PE150测序具有8nt i5索引和8nt i7索引。

[0168] Parallel-Seq和Parallel-Split-Seq数据的预处理

[0169] 我们使用read1对Parallel-Seq的细胞条形码和连接接头进行测序。为了平衡测序核苷酸组合,我们在barcode2和连接接头之间添加相位核苷酸,第1-24位不添加任何核苷酸,第25-48位添加T,第49-72位添加CA,第73-96位添加ACA。对于barcode2的1-24th, Parallel-Seq的barcode1、barcode2、barcode3和barcode4应该位于36-41st、11-18th、i5索引、i7索引内。对于barcode2的25-48th、49-72nd和73-96th位,只需要改变barcode1的位置一个核苷酸步长。唯一的分子标识符位于read1的1-10th内。位于read1的Tn5ME后面的序列是一个Tn5切割位点,而read2为ATAC-seq提供另一Tn5切割位点。RNA-seq文库的Read2从第二链合成退火位点开始,与靶基因的RNA序列相同。

[0170] 对于Parallel-Split-Seq,使用read2对细胞条码和连接接头进行测序。barcode1、barcode2、barcode3和barcode4应位于第61-66、36-43、11-18、i7索引内。位于read2的Tn5ME后面的序列是一个Tn5切割位点,而read1为ATAC-seq提供了另一个片段。RNA-seq文库的Read1提供了目标基因的RNA序列。

- [0171] 原始读数用cutadapt修剪。条形码由FREE Difference软件解析，每轮条形码只允许一次编辑。从RNA文库中筛选出具有嵌入的末端序列的数据，并从ATAC文库中筛选出不具有嵌入的末端序列的数据。使用STAR将数据与hg38、mm10或组合基因组比对。
- [0172] 对于单细胞RNA-seq，使用来自Split-seq管道的修改后的python脚本来折叠UMI并生成数字基因表达矩阵。对于单细胞ATAC-seq，删除了线粒体读数。如前所述计算TSS可访问性的丰富性，以评估数据质量。丢弃TSS富集<6的细胞。然后计算整个基因组中2-kb bins(区间)上的Tn5插入。
- [0173] 丢弃表达<200个基因、<200bin的细胞。我们使用Scrublet来预测doublet概率并通过默认阈值去除双峰。
- [0174] 对于混合细胞系数据，将<90%UMI映射到一个物种的细胞视为混合细胞。
- [0175] 实施例中对照方法如下：
- [0176] sci-CAR的步骤参见文献：Cao, J.et al.Joint profiling of chromatin accessibility and gene expression in thousands of single cells.Science361, 1380–1385(2018)。
- [0177] paired-Seq的步骤参见文献：Zhu, C.et al.An ultra high-throughput method for single-cell joint analysis of open chromatin and transcriptome.Nat Struct Mol Biol26, 1063–1070(2019)。
- [0178] SNARE-Seq的步骤参见文献：Chen, S., Lake, B.B.&Zhang, K.High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell.Nat Biotechnol37, 1452–1457(2019)。
- [0179] SHARE-Seq的步骤参见文献：Ma, S.et al.Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin.Cell183, 1103–1116 e20(2020)。
- [0180] 本申请代表序列中碱基的“X...X”、“N...N”均可以代表任意天然或修饰或现有技术中已知的碱基类型，其中“X”与“N”可替换使用，包括但不限于A、T、C、G或U。V代表A、C或G。B代表C、G、T或U。当然，当“X”代表氨基酸时则代表现有技术中已知的天然或经修饰的氨基酸类型。
- [0181] 实施例1 Parallel-Seq分析多个样本的同一单细胞中的RNA和开放染色质

[0182] Parallel-Seq的实验设计如图1所示。具体步骤如下：

[0183] (1)Parallel-Seq从27个不同的样品开始，每个样品50,000个单细胞；

[0184] (2)对每个样品的细胞进行固定、透化并使用条码Tn5转座子，用转座子特有的条形码标记开放染色质；其中，转座子特有的条码序列Tn5ME-B如SEQ ID NO: 1所示，带有第一索引的Tn5ME-x(x代表1-27)序列如SEQ ID NO: 2所示，序列中XXXXXX代表第一索引，见表1。

[0185] Tn5ME-B: GTCTCGTGGGCTCGGAGATGTGTATAAGAGAC
AG(SEQ ID NO: 1)

[0186] Tn5ME-x: /5Phos/TGCAGTAXXXXXXAGATGTGTATAAGAGAC
AG(SEQ ID NO: 2)

[0187] (3)用条码匹配的poly(T)引物R1BxT15VN和随机六聚体引物R1BxN6对每个样本的mRNA进行逆转录，逆转录引物R1BxT15VN和R1BxN6(x代表1-27)如SEQ ID NO: 3和SEQ ID NO: 4所示，序列中XXXXXX代表第一索引，见表1；

[0188] R1BxT15VN: /5Phos/TGCAGTAXXXXXXTTTTTTTTTTTTTTTVN
(SEQ ID NO: 3)。

[0189] R1BxN6: /5Phos/TGCAGTAXXXXXXNNNNNNN(SEQ ID NO: 4)

[0190] (4)将不同样本的细胞组合并随机分配到96孔板中，每孔中包含dscB' 序列(SEQ ID NO: 5)，将孔特异性接头序列dscBx连接到转座染色质或cDNA第一链，其中，孔特异性接头序列dscBx(x代表1-96)如SEQ ID NO: 6所示，序列中“NNNNNNNNNN”为UMI，第二索引和UMI之间添加相位核苷酸，第dscB1-dscB24不添加任何核苷酸，dscB25-dscB48位添加T，dscB49-dscB72位添加CA，dscB73-dscB96位添加ACA，序列中XXXXXXXXX代表第二索引，见表2；

[0191] dscB' 序列: TACTGCACTCAGTGACT(SEQ ID NO: 5)

[0192] dscBx序列: TCGTCGGCAGCGTCAGATGTGTATAA
GAGACAGNNNNNNNNNNXXXXXXXXXAGTCACTGAG(SEQ ID NO: 6)

[0193] (5)用RNA酶消化RNA；

- [0194] (6)随机引物的第二链合成将第二个PCR锚点附加到cDNA上,其中,第二链合成所用引物为p3短引物。
- [0195] p3短引物: CAGACGTGTGCTCTTCCGATCTNNNGGNNNB(SEQ ID NO: 7)
- [0196] (7)所有细胞汇集在一起并过载进入Chromium scATAC-seq芯片的一个通道;
- [0197] (8)裂解细胞,在液滴内线性扩增添加液滴特异性标记p5 adapter即第三索引,见表2,其中,线性扩增引物如(SEQ ID NO: 8)所示,其中XXXXXXXXXXXXXXXXXX为第三索引信息,为每个液滴中beads的特异性索引;
- [0198] 5' -AATGATACGGCGACCACCGAGATCTACAC-XXXXXXXXXXXXXXXXXX-TCGTCGGCAGCGTC-3' (SEQ ID NO: 8)
- [0199] (9)进一步将液滴分到16个PCR管中进行PCR纯化;
- [0200] (10)将每个PCR管中纯化产物分为两部分,分别用相应的引物扩增转录组和开放染色质片段,其中,扩增转录组采用引物SI-PCR引物B(SEQ ID NO: 9)和P3xx引物(SEQ ID NO: 10),序列中XXXXXXXXXX代表扩增转录组所需引物序列的第四索引,见表2中P3xx索引;扩增开放染色质片段采用引物SI-PCR引物B(SEQ ID NO: 9)和N7xx引物(SEQ ID NO: 11),序列中XXXXXXXXXX代表扩增开放染色质片段所需引物序列的第四索引,见表1中N7xx索引;
- [0201] SI-PCR引物B: AATGATACGGCGACCACCGAGA(SEQ ID NO: 9)
- [0202] P3xx引物: CAAGCAGAAGACGGCATACGAGATXXXXXXXXXXGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT(SEQ ID NO: 10)
- [0203] N7xx引物: CAAGCAGAAGACGGCATACGAGATXXXXXXXXXXGTCTCGTGGGCTCGG(SEQ ID NO: 11)
- [0204] (11)经过测序和条码解析分析后,4轮索引相同组合的基因表达和染色质可接近性图谱代表了一个单细胞的成对图谱。原则上,经过4轮索引后,将条码空间很大程度上扩展到 $(96 \times 96 \times 100000 \times 16 \approx 1.47 \times 10^{10})$,这使得Parallel-Seq能够在一次实验中评估超过100万个细胞,碰撞率极低。
- [0205] 表1

第一索引	N7xx 索引
AACAAC	TCGCCTTA
ACCGCA	CTAGTACG
AGTTGG	TTCTGCCT
CCACGT	GCTCAGGA
CGTGTT	AGGAGTCC
GTTCTC	CATGCCTA
TGACTA	GTAGAGAG
TCAAGG	CCTCTCTG
AACGGT	AGCGTAGC
AAGCCT	CAGCCTCG
ACATGA	TGCCTCTT
ACTCTA	TCCTCTAC
AGAAGT	
AGTACC	
ATGCGA	
CAATAG	
CATCCA	
CCTGGA	
CGAGAC	
CGCTCA	
GCGTAA	
GGATCG	
GTGAGG	
TCCTTA	
TCTGCC	
TTAACC	
TTAGTG	

[0206] 表2

第二索引	P3xx 索引	Parallel-Split-Seq 第三索引
AAGACCAA	TCATCATT	AACCTCTT
AAGCTACG	GCTCCTGT	AACGTCGC
AAGGTCAT	CCTAACGT	AAGAATCG

AATAGTGG	CACGTAGT	AAGCGGTG
AATGCCTT	GTAAGAGT	AAGGAGCT
ACAATAGC	TACCTTCT	AATACCGC
ACAGGATT	AAGTCTCT	AATCTCCA
ACCGACCT	AGAGATCT	ACAACTTC
ACCTAGAT	CCAGCGCT	ACACGCAA
ACGAGTCC	ATGAACCT	ACCACAGT
ACGGACGA	CTATAACT	ACCGTGTA
ACGTTCAA	TCTTATAT	ACCTTGCC
ACTATCTG	GAATCGAT	ACGCATAA
ACTCCGAA	AATAAGAT	ACGTATGG
AGAACAGA	ATTCTAAT	ACTAACCA
AGACGCTT	ATCATATG	ACTCAGGT
AGATGCGA	GCTTAATG	ACTTGTTG
AGCCACTC	GGATATGG	AGAAGTAC
AGCGAAGC	CTTGATGG	AGAGATGA
AGGTAACG	AAGATCGG	AGATTAGG
AGTACATC	TTAACCGG	AGCCTGGT
AGTGATTC	CTAAGTCG	AGCTCTAA
ATAAGAGG	TATTCGCG	AGGTGTCT
ATATCACG	GAAGCACG	AGICCGTT
ATCGCCGT	TCCAGTAG	AGTTCGCA
ATGACGGA	TTGTCTAG	ATAAGCTC
ATGGAATG	AGCGCTAG	ATCCATGA
ATTCCTAC	CCTGTGAG	ATCTAGCG
CAACGCCA	ATAGGAAG	ATGCAACC
CAAGTCTG	GATGGTTC	ATGTGCAG
CACACATC	CCACATTC	ATTGGTAG
CACCTTAT	TCTTGGTC	CAAGAAGA
CAGAACCT	CGAGGATC	CAATGGAC
CAGCCGAT	CACTAATC	CACATGCT
CATACTGT	TAAGGCGC	CACGGTAG
CATCCACC	ACTGTTCC	CAGAGGTT
CATTGAGC	CTTCCTCC	CAGTATAG
CCAAGCGT	TACAGGCC	CATCAAGT

CCACGACT	TAATTACC	CATGTTCC
CCATTGTC	ATAACACC	CCAACAAT
CCGCATGT	CTACGCAC	CCAATTAC
CCTACTCC	AGGTTAAC	CCAGTGAA
CCTCCTTG	GTTGCAAC	CCGATCAG
CCTTAATG	CTCAATTA	CCGGTCTT
CGAATATC	ACAACCTA	CGACAACG
CGAGAGCA	CTACCATA	CGCCAGTA
CGCCTCAA	GACACATA	CGCGGAAT
CGCGTTAC	CGCGGTGA	CGGAAGGA
CGGACTCT	TACTATGA	CGGTGAGA
CGGTTGTT	TTGTTGGA	CGTAACAC
CGTAGCTT	AGTGAGGA	CGTCTATG
CGTGCCAA	ATCGCCGA	CGTTCTCG
CTACCGGA	CTTATAGA	CTACTAAG
CTAGCAGT	CCATGAGA	CTAGTGCG
CTCAGCCT	TCACCTCA	CTCTGACA
CTCTTCTA	ATACTCCA	CTGATGAA
CTGCTGGT	GTTCGACA	CTGGTACA
CTGTATTC	CATCATAA	CTTACGAG
CTTCGCTC	CACATGAA	GAACTCAA
GAAGAGTA	ATGAGGAA	GAATGTTG
GACACCTA	ATCTATCG	GACGAATT
GACGTGAG	TTACTGTT	GACTGCCA
GACTTACT	TAGGAATA	GAGCTATT
GAGGACAA	CCGCACAG	GAGTCGGA
GAGTTAAG	ATTGCGTT	GATAGAAC
GATCCTCG	TCGTAAGC	GATGGTCT
GCAATCCG	GGACTGCG	GCAGCACT
GCAGTGTG	AAGGATGC	GCATTCAT
GCCGCTAA	GCAATGGC	GCCTCTGT
GCGACCAT	GTATTCTC	GCGCAGAT
GCTAAGAC	GTCATTAC	GCTCACAA
GCTGTAGG	GGTATACT	GCTTGCGT
GGAACTGG	TTGCGTGC	GTAATGCA

GGACAGTT	GCCAGCGA	GTATCGAG
GGATTGCT	GGCGCGCA	GTCGATCT
GGTCCTAA	CGCCACCT	GTGAGCGT
GTACCTGT	CGGTATCA	GTGGATAG
GTCAAGGA	ACGCAGCC	GTTAGCCA
GTCTGCTT	GAGCTGAC	TAAGGTGG
GTGCTCCA	TGATGGCT	TACACCGG
GTGTGACC	GCGTGAG	TACTCGTC
GTTATTGG	ACTTCCGC	TAGCTGAG
TAATTCGG	ACATGTAC	TCAACAGG
TACCAATC	GACGATAT	TCACTCAC
TAGACTCC	CCTGGCCA	TCATAGAC
TAGTCAAC	AGCGTGTC	TCCGTACA
TCACGTTG	CCGTTGTT	TCGGAGTA
TCAGAATG	TGTAACCA	TCGTCGGT
TCCAGCTT	AGTAGTTA	TGAACCGC
TCCTGCGA	GTCTGCCT	TGAGTCTT
TCGGTTCC	TGCGCAGC	TGCGACTG
TCTTACCT	TGCCTATA	TGGTTATC
TGACATGG	TGCTAGTG	TGTGTAAG
TGCCTATA	CCGAGCTC	TTAGGAAC
TGGTGTGG	GACTGAGG	TTCAGTGG
TGTACTAG	TGGAATTC	TTCTATCC

[0207] 实施例2 Parallel-Seq性能验证

[0208] 用NIH/3T3(小鼠)、HEK293T(人)和K562(人)细胞的混合物进行了Parallel-Seq(步骤同实施例1)，经质量筛选获得了2200个细胞的转录组和染色质可接近性，其中scRNA序列部分的平均UMI为7014，scATAC序列部分的平均UMI为10103。来自人类和小鼠细胞的Reads在转录组和染色质谱中都被很好地分离，其中转录组分配给802个小鼠细胞和1398个人类细胞，染色质谱分配给805个小鼠细胞和1398个人类细胞，均有很少的doublet，两个图谱的碰撞率分别为0.2%、0.1%(图2)。

- [0209] 聚合后的scATAC-seq数据的插入片段大小分布显示出清晰的核小体结合模式(图3), 测序reads的TSS富集得分高达14(图4), 说明scATAC-seq数据是合格的。
- [0210] Parallel-Seq生成的聚集单细胞染色质可接近性和转录组谱分别与ENCODE中K562细胞的大量DNA酶序列(ENCFF156LGK, $R=0.79$)(图5)和核RNA序列(ENCFF631TDY, $R=0.81$)(图6)密切相关(图7)。此外, 每个细胞的表达和染色质可接近性谱在细胞类型内聚在一起, 彼此分离(图8-9)。总之, 这些数据证明了Parallel-Seq的高特异性和高质量。
- [0211] 进一步比较了Parallel-Seq与sci-CAR、paired-Seq、SNARE-Seq和SHARE-Seq的数据质量。Parallel Seq在两个文库上显示出优于最先进的方法SHARE Seq的数据质量(图10-11), 其ATAC片段和RNA的UMI的数量更多, 捕获的基因数量也较其他方法多, 具有更大的带宽。
- [0212] 实施例3 Parallel-Split-Seq及性能验证
- [0213] 为了使其更易于使用, 进一步降低成本, 开发了Parallel-Split-Seq, 将Parallel-Seq中第三索引的加入位置改变, 即从液滴中线性扩增增加第三索引改为在板上增加一轮连接反应增加第三索引。依然包含液滴中线性扩增的步骤, 只是在这步中不增加第三索引, 条码空间为 $24 \times 96 \times 96 \times 96 \sim 2.12 \times 10^7$ (图12)。
- [0214] 本实施例用NIH/3T3(小鼠)、HEK293T(人)、Hela(人)、K562(人)及THP1(人)细胞的混合物进行了Parallel-Split-Seq(步骤同实施例1), 具体步骤如下:
- [0215] (1)对每个样品的细胞进行固定、透化并使用条码Tn5转座子, 用转座子特有的条形码标记开放染色质; 其中, 转座子特有的条码序列Tn5ME-A如SEQ ID NO: 12所示, 带有第一索引的Tn5ME-x(x代表1-27)序列如SEQ ID NO: 2所示, 序列中XXXXXX代表第一索引, 见表1;
- [0216] Tn5ME-A: TCGTCGGCAGCGTCAGATGTGTATAAGAGACA
G(SEQ ID NO: 12)
- [0217] (2)用条码匹配的poly(T)引物R1BxT15VN和随机六聚体引物R1BxN6对每个样本的mRNA进行逆转录, 逆转录引物R1BxT15VN和R1BxN6(x代表1-27)如SEQ ID NO: 3和SEQ ID NO: 4所示, 序列中XXXXXX代表第一索引, 见表1;

- [0218] (3)将不同样本的细胞组合并随机分配到96孔板中,进行2次连接反应,增加第二索引及第三索引,其中增加第二索引时每孔中包含R2' 序列(SEQ ID NO: 13),增加第三索引时每孔中包含R3' 序列(SEQ ID NO: 14),将孔特异性接头序列连接到转座染色质或cDNA第一链,其中,带有第二索引的孔特异性接头序列如R2Bx(SEQ ID NO: 15)所示,序列中XXXXXXXXXX代表第二索引,见表2;带有的第三索引的孔特异性接头序列如R3Bx(SEQ ID NO: 16)所示,序列中XXXXXXXXXX代表第三索引,见表2;
- [0219] R2' 序列: TACTGCAGCTGAACCTC(SEQ ID NO: 13)
- [0220] R3' 序列: TCTCCAAAGCTGTGGAC(SEQ ID NO: 14)
- [0221] R2Bx序列: /5Phos/TTGGAGAXXXXXXXXXXXGAGG TTCAGC(SEQ ID NO: 15)
- [0222] R3Bx序列: CAGACGTGTGCTCTTCCGATCTNNNNNNNNNNN
XXXXXXXXXXGTCCACAGCT(SEQ ID NO: 16)。
- [0223] (4)用RNA酶消化RNA;
- [0224] (5)随机引物的第二链合成将第二个PCR锚点附加到cDNA上,其中,第二链合成所用引物为p5短引物。
- [0225] P5短引物: ACACGACGCTCTTCCGATCTNNNGGNNNB(SEQ ID NO: 17)
- [0226] (6)所有细胞汇集在一起,计数并稀释到800细胞/ul,分装到PCR管中,每管2.5ul细胞;
- [0227] (7)在PCR管中裂解细胞,并直接加入PCR扩增体系(包含P5xx(SEQ ID NO: 18)、N5xx(SEQ ID NO: 19)和P3xx(SEQ ID NO: 10))进行扩增,并增加第四索引,见表2;
- [0228] (8)对每个PCR进行纯化,将纯化产物分为两部分,分别用相应的引物扩增转录组和可接近性染色质片段,其中,扩增转录组采用p3 end引物(SEQ ID NO: 20)和P5xx(SEQ ID NO: 18),P5xx序列中XXXXXXXXXX为扩增转录组的第四索引,见表3;扩增开放染色质片段采用引物p3 end(SEQ ID NO: 20)和N5xx(SEQ ID NO: 19),N5xx序列中XXXXXXXXXX为扩增开放染色质片段的第四索引,见表3。

[0229] p3 end序列：CAAGCAGAAGACGGCATACGAGAT(SEQ ID NO：20)

[0230] P5xx序列：

[0231] AATGATACGGCGACCACCGAGATCTACACXXXXXXXXX
ACACTCTTTCCCTACACGACGCTCTTCCGATCT(SEQ ID NO：18)

[0232] N5xx序列：AATGATACGGCGACCACCGAGATCTACACXXXXXXXXX
TCGTCGGCAGCGTC(SEQ ID NO：19)

[0233] 表3

P5xx 索引	N5xx 索引
TATAGCCT	TAGATCGC
ATAGAGGC	CTCTCTAT
CCTATCCT	TATCCTCT
GGCTCTGA	AGAGTAGA
AGGCGAAG	GTAAGGAG
TAATCTTA	ACTGCATA
CAGGACGT	AAGGAGTA
GTACTGAC	CTAAGCCT
	CGTCTAAT
	TCTCTCCG
	TCGACTAG
	TTCTAGCT
	CCTAGAGT
	GCGTAAGA
	CTATTAAG
	AAGGCTAT
	GAGCCTTA
	TTATGCGA

[0234] 结果显示，Parallel-Split-Seq的特异性较好，碰撞率低，与大量数据相关性高(见图13-18)。而且，Parallel-Split-Seq与Parallel-Seq的性能相当，优于现有的方法(见图19-20)。

[0235] 以上详细描述了本发明的优选实施方式，但是，本发明并不限于上述实施方式中的具体细节，在本发明的技术构思范围内，可以对本发明的技术方案进行多

种简单变型，这些简单变型均属于本发明的保护范围。另外需要说明的是，在上述具体实施方式中所描述的各个具体技术特征，在不矛盾的情况下，可以通过任何合适的方式进行组合，为了避免不必要的重复，本发明对各种可能的组合方式不再另行说明。

权利要求书

- [权利要求 1] 一种构建单细胞测序文库的方法，其特征在于，包括：
- a) 利用转座子切割开放染色质获得携带第一接头的DNA片段；
 - b) 加入逆转录引物对mRNA进行逆转录获得携带第二接头的cDNA第一条链；
 - c) 将细胞置于载体上，利用第一载体特异性接头分别将步骤a)获得的携带第一接头的DNA片段和步骤b)获得的cDNA第一条链连接至载体上；
 - d) 合成cDNA的第二条链；
 - e) 并用引物分别扩增转录组文库的cDNA和染色质DNA。
- [权利要求 2] 根据权利要求1所述的方法，其特征在于，所述的转座子包含条形码序列和转座酶；
- 优选的，所述的条形码序列包含第一接头；
- 优选的，所述的条形码序列包含第一索引。
- [权利要求 3] 根据权利要求1或2所述的方法，其特征在于，所述的逆转录引物包含第二接头，所述的第二接头包含poly(T)和第一索引；优选还包含随机六聚体引物。
- [权利要求 4] 根据权利要求1-3任一所述的方法，其特征在于，所述的第一索引包含AACAAC、ACCGCA、AGTTGG、CCACGT、CGTGTT、GTTCTC、TGACTA、TCAAGG、AACGGT、AAGCCT、ACATGA、ACTCTA、AGAAGT、AGTACC、ATGCGA、CAATAG、CATCCA、CCTGGA、CGAGAC、CGCTCA、GCGTAA、GGATCG、GTGAGG、TCCCTTA、TCTGCC、TTAACC或TTAGTG中的至少一个、两个或三个以上的组合。
- [权利要求 5] 根据权利要求1-4任一所述的方法，其特征在于，所述的第一载体特异性接头包含第二索引，优选的，所述的第二索引包含AAGACCAA、AAGCTACG、AAGGTCAT、AATAGTGG、AA

TGCCTT、ACAATAGC、ACAGGATT、ACCGACCT、ACCTAGAT、ACGAGTCC、ACGGACGA、ACGTTCA A、ACTATCTG、ACTCCGAA、AGAACAGA、AGACG CTT、AGATGCGA、AGCCACTC、AGCGAAGC、AGG TAACG、AGTACATC、AGTGATTTC、ATAAGAGG、A TATCACG、ATCGCCGT、ATGACGGA、ATGGAATG 、ATTCCTAC、CAACGCCA、CAAGTCTG、CACACA TC、CACCTTAT、CAGAACCT、CAGCCGAT、CATA CTGT、CATCCACC、CATTGAGC、CCAAGCGT、CC ACGACT、CCATTGTC、CCGCATGT、CCTACTCC、 CCTCCTTG、CCTTAATG、CGAATATC、CGAGAGC A、CGCCTCAA、CGCGTTAC、CGGACTCT、CGGTT GTT、CGTAGCTT、CGTGCCAA、CTACCGGA、CTA GCAGT、CTCAGCCT、CTCTTCTA、CTGCTGGT、C TGTATTC、CTTCGCTC、GAAGAGTA、GACACCTA 、GACGTGAG、GACTTACT、GAGGACAA、GAGTTA AG、GATCCTCG、GCAATCCG、GCAGTGTG、GCCG CTAA、GCGACCAT、GCTAAGAC、GCTGTAGG、GG AACTGG、GGACAGTT、GGATTGCT、GGTCCTAA、 GTACCTGT、GTCAAGGA、GTCTGCTT、GTGCTCC A、GTGTGACC、GTTATTGG、TAATTCGG、TACCA ATC、TAGACTCC、TAGTCAAC、TCACGTTG、TCA GAATG、TCCAGCTT、TCCTGCGA、TCGGTTCC、T CTTACCT、TGACATGG、TGCCTATA、TGGTGTGG 、TGTACTAG中的至少一个、两个或三个以上的组合。

[权利要求 6]

根据权利要求1-5任一所述的方法，其特征在于，所述的方法还包括形成液滴，裂解细胞并在液滴中进行扩增反应的步骤，优选的，形成的液滴中过载细胞。

- [权利要求 7] 根据权利要求6所述的方法，其特征在于，所述在液滴中进行扩增反应使用的引物包含第三索引。
- [权利要求 8] 根据权利要求1-5任一所述的方法，其特征在于，所述的方法包括利用第二载体特异性接头分别将权利要求1获得的携带第一接头的DNA片段和权利要求1获得的cDNA第一条链连接至载体上，优选的，所述的第二载体特异性接头包含第三索引。
- [权利要求 9] 根据权利要求7或8所述的方法，其特征在于，所述的第三索引包含AACCTCTT、AACGTCGC、AAGAATCG、AAGCGGTG、AAGGAGCT、AATACCGC、AATCTCCA、ACAACTTC、ACACGCAA、ACCACAGT、ACCGTGTA、ACCTTGCC、ACGCATAA、ACGTATGG、ACTAACCA、ACTCAGGT、ACTTGTTG、AGAAGTAC、AGAGATGA、AGATTAGG、AGCCTGGT、AGCTCTAA、AGGTGTCT、AGTCCGTT、AGTTCGCA、ATAAGCTC、ATCCATGA、ATCTAGCG、ATGCAACC、ATGTGCAG、ATTGGTAG、CAAGAAGA、CAATGGAC、CACATGCT、CACGGTAG、CAGAGGTT、CAGTATAG、CATCAAGT、CATGTTCC、CCAACAAT、CCAATTAC、CCAGTGAA、CCGATCAG、CCGGTCTT、CGACAACG、CGCCAGTA、CGCGGAAT、CGGAAGGA、CGGTGAGA、CGTAACAC、CGTCTATG、CGTTCTCG、CTACTAAG、CTAGTGCG、CTCTGACA、CTGATGAA、CTGGTACA、CTTACGAG、GAACTCAA、GAATGTTG、GACGAATT、GACTGCCA、GAGCTATT、GAGTCGGA、GATAGAAC、GATGGTCT、GCAGCACT、GCATTCAT、GCCTCTGT、GCGCAGAT、GCTCACAA、GCTTGCGT、GTAATGCA、GTATCGAG、GTCGATCT、GTGAGCGT、GTGGATAG、GTTAGCCA、TAAGGTGG、

TACACCGG、TACTCGTC、TAGCTGAG、TCAACAG
G、TCACTCAC、TCATAGAC、TCCGTACA、TCGGA
GTA、TCGTCGGT、TGAACGCG、TGAGTCTT、TGC
GACTG、TGGTTATC、TGTGTAAG、TTAGGAAC、T
TCAGTGG、TTCTATCC中的至少一个、两个或三个以上的组
合。

[权利要求 10]

根据权利要求1-9任一所述的方法，其特征在于，所述步骤e)进行的扩增反应中的引物包含第四索引；

优选的，所述的第四索引包含P3_{xx}索引中的至少一个、两个或三个以上的组合；

优选的，所述的第四索引包含N7_{xx}索引中的至少一个、两个或三个以上的组合；

优选的，所述的第四索引包含P5_{xx}索引中的至少一个、两个或三个以上的组合；

优选的，所述的第四索引包含N5_{xx}索引中的至少一个、两个或三个以上的组合；

其中，所述的P3_{xx}索引、N7_{xx}索引、P5_{xx}索引或N5_{xx}索引如下所示：

P3xx 索引	N7xx 索引	P5xx 索引	N5xx 索引
TCATCATT	TCGCCTTA	TATAGCCT	TAGATCGC
GCTCCTGT	CTAGTACG	ATAGAGGC	CTCTCTAT
CCTAACGT	TTCTGCCT	CCTATCCT	TATCCTCT
CACGTAGT	GCTCAGGA	GGCTCTGA	AGAGTAGA
GTAAGAGT	AGGAGTCC	AGGCGAAG	GTAAGGAG
TACCTTCT	CATGCCTA	TAATCTTA	ACTGCATA
AAGTCTCT	GTAGAGAG	CAGGACGT	AAGGAGTA
AGAGATCT	CCTCTCTG	GTACTGAC	CTAAGCCT
CCAGCGCT	AGCGTAGC		CGTCTAAT
ATGAACCT	CAGCCTCG		TCTCTCCG
CTATAACT	TGCCTCTT		TCGACTAG
TCTTATAT	TCCTCTAC		TTCTAGCT
GAATCGAT			CCTAGAGT
AATAAGAT			GCGTAAGA
ATTCTAAT			CTATTAAG
ATCATATG			AAGGCTAT
GCTTAAATG			GAGCCTTA
GGATATGG			TTATGCGA
CTTGATGG			
AAGATCGG			
TTAACCGG			
CTAAGTCG			
TATTCGCG			
GAAGCACG			
TCCAGTAG			

TTGTCTAG			
AGCGCTAG			
CCTGTGAG			
ATAGGAAG			
GATGGTTC			
CCACATTC			
TCTTGGTC			
CGAGGATC			
CACTAATC			
TAAGGCGC			
ACTGTTCC			
CTTCCTCC			
TACAGGCC			
TAATTACC			
ATAACACC			
CTACGCAC			
AGGTTAAC			
GTGCAAC			
CTCAATTA			
ACAACCTA			
CTACCATA			
GACACATA			
CGCGGTGA			
TACTATGA			
TTGTTGGA			
AGTGAGGA			
ATCGCCGA			
CTIATAGA			
CCATGAGA			
TCACCTCA			
ATACTCCA			
GTTCGACA			
CATCATAA			
CACATGAA			
ATGAGGAA			

ATCTATCG			
TTACTGTT			
TAGGAATA			
CCGCACAG			
ATTGCGTT			
TCGTAAGC			
GGACTGCG			
AAGGATGC			
GCAATGGC			
GTATTCTC			
GTCATTAC			
GGTATACT			
TTGCGTGC			
GCCAGCGA			
GGCGCGCA			
CGCCACCT			
CGGTATCA			
ACGCAGCC			
GAGCTGAC			
TGATGGCT			
GCGTGGAG			
ACTTCCGC			
ACATGTAC			
GACGATAT			
CCTGGCCA			
AGCGTGTC			
CCGTTGTT			
TGTAACCA			
AGTAGTTA			
GTCTGCCT			
TGCGCAGC			
TGCCTATA			
TGCTAGTG			
CCGAGCTC			
GACTGAGG			

TGGAATTC			
----------	--	--	--

- [权利要求 11]

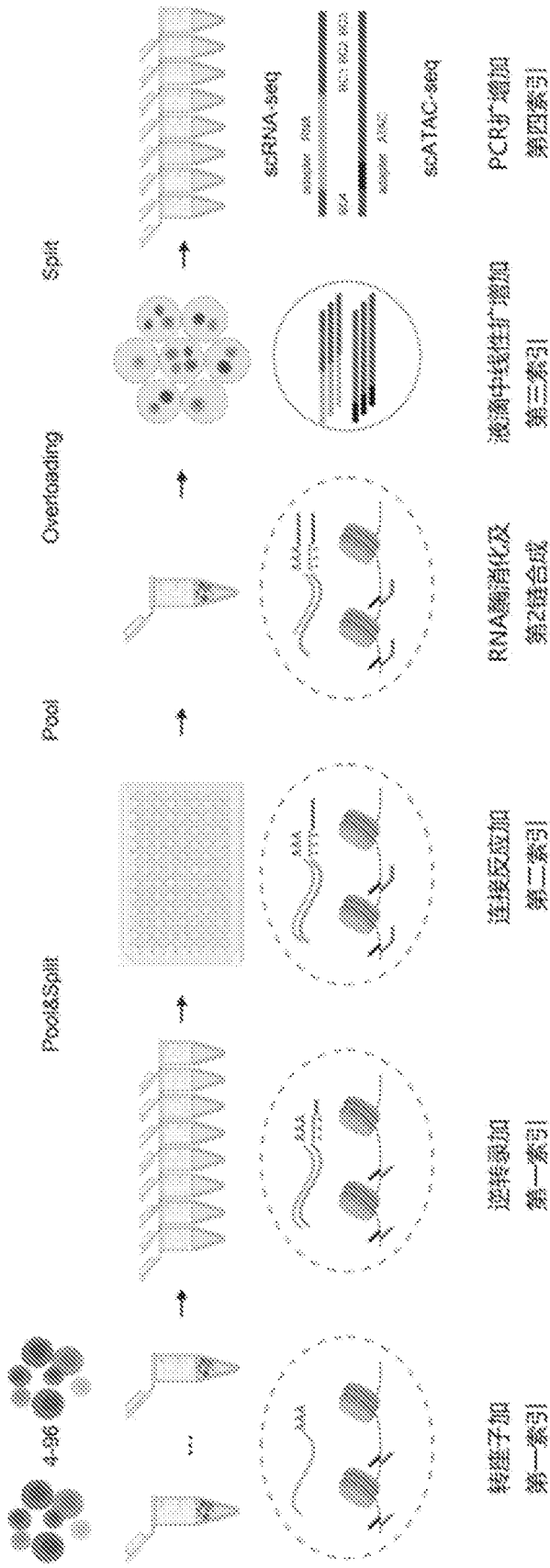
根据权利要求2-10任一所述的方法，其特征在于，所述的载体包括孔、管或平板。
- [权利要求 12]

根据权利要求1-11任一所述的方法，其特征在于，所述的方法还包括加入RNA酶。
- [权利要求 13]

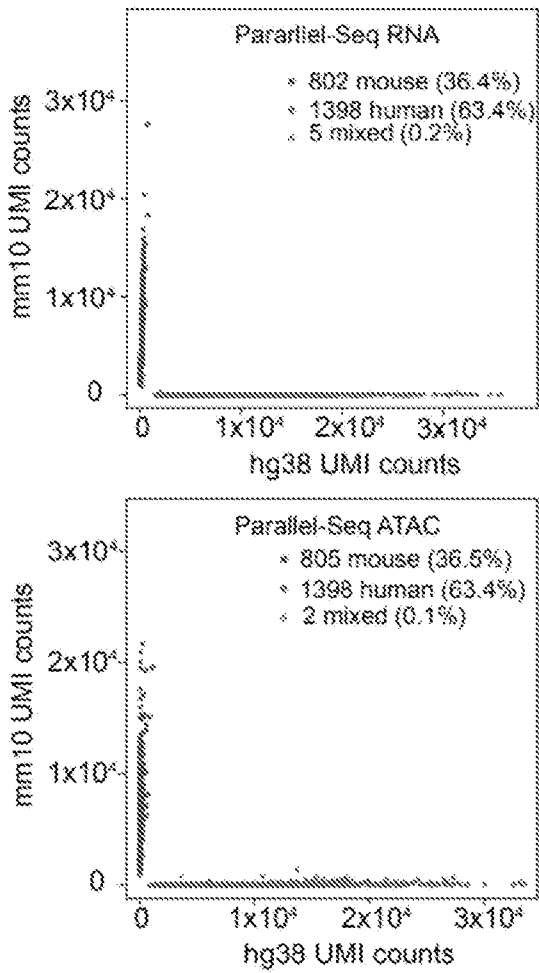
根据权利要求1-12任一所述的方法，其特征在于，所述的方法还包括获得细胞，将细胞固定并透化。

- [权利要求 14] 一种构建转录组文库的方法，其特征在于，所述的方法包括加入逆转录引物对mRNA进行逆转录获得携带第二接头的cDNA第一条链；将细胞置于载体上，利用第一载体特异性接头将获得的cDNA第一条链连接至载体上；合成cDNA的第二条链；纯化并用引物扩增转录组的cDNA。
- [权利要求 15] 一种构建染色质DNA文库的方法，其特征在于，所述的方法包括利用转座子切割开放染色质获得携带第一接头的DNA片段；将细胞置于载体上，利用第一载体特异性接头将获得的携带第一接头的DNA片段连接至载体上；纯化DNA并用引物分别扩增染色质DNA。
- [权利要求 16] 一种权利要求1-15任一所述的方法获得的核酸文库。
- [权利要求 17] 一种测序方法，其特征在于，所述的测序方法包括构建权利要求16所述的核酸文库。
- [权利要求 18] 一种权利要求16所述的核酸文库的应用，其特征在于，所述的应用包括肿瘤靶点筛选、疾病监测或植入前胚胎诊断。

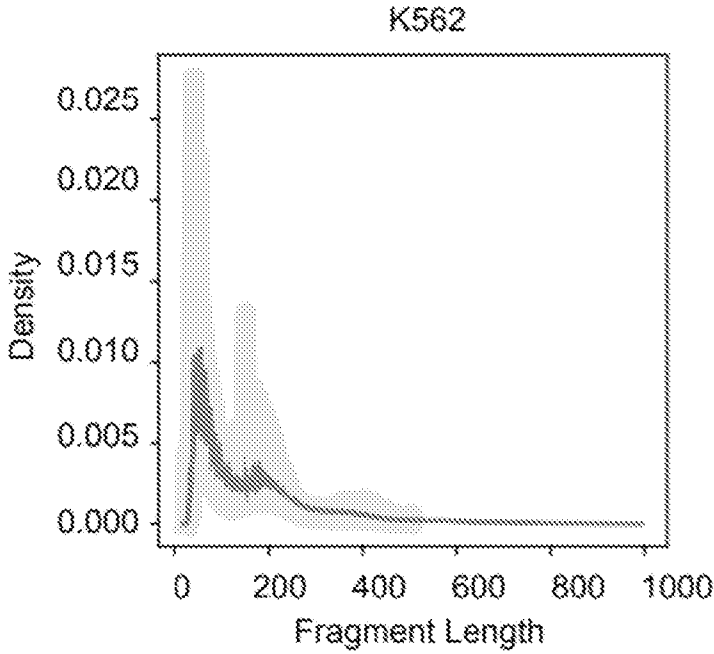
[图 1]



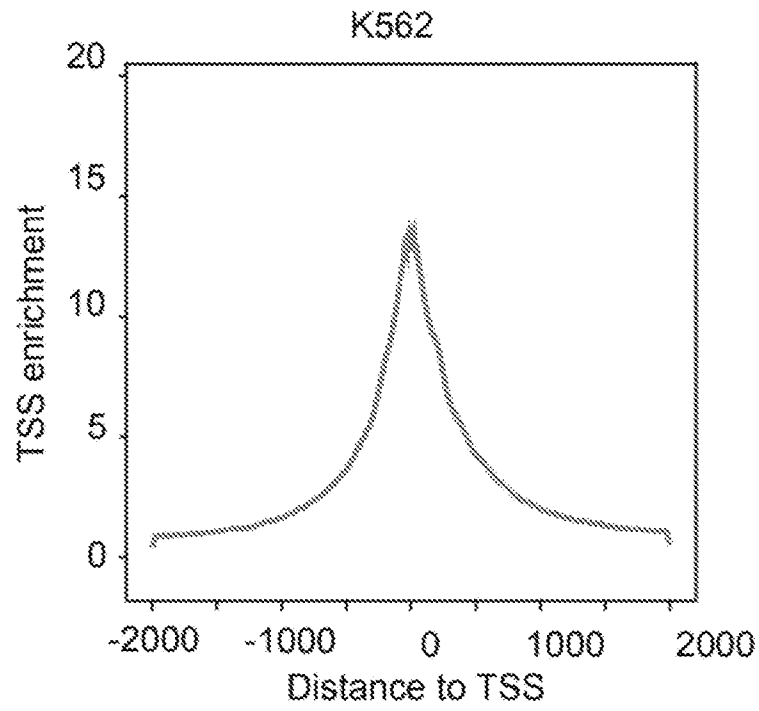
[图 2]



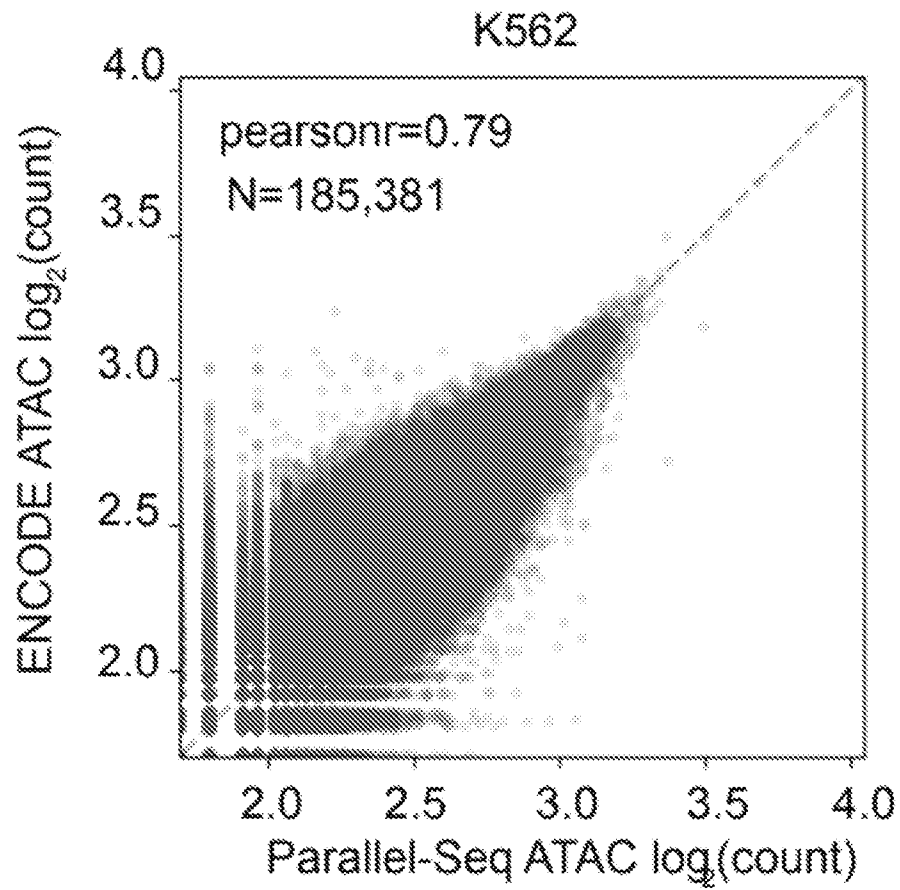
[图 3]



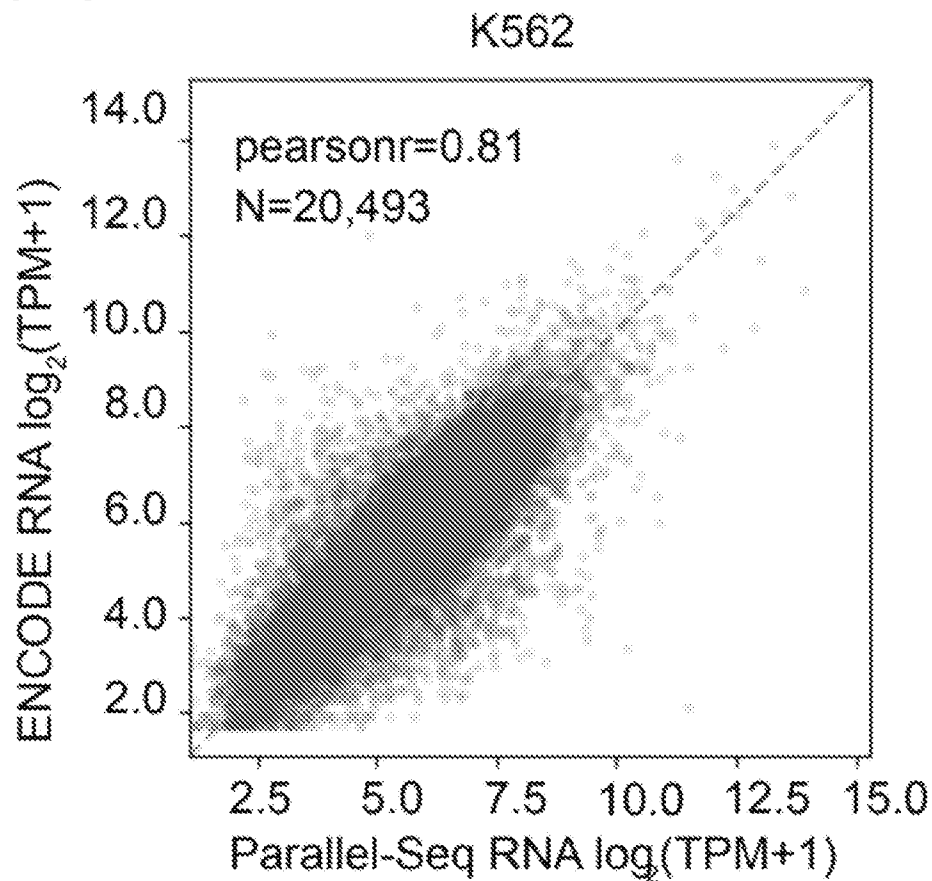
[图 4]



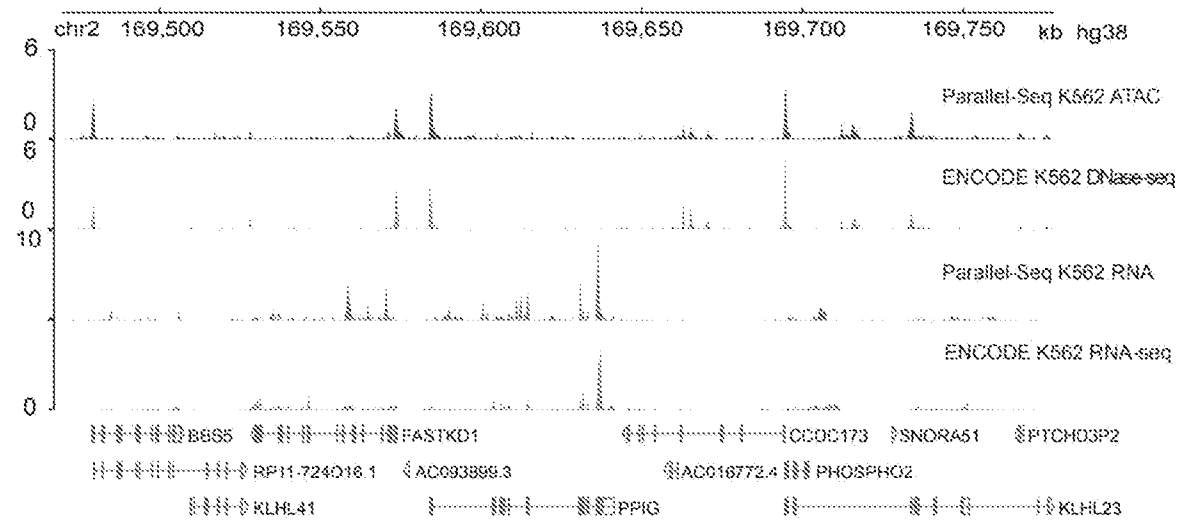
[图 5]



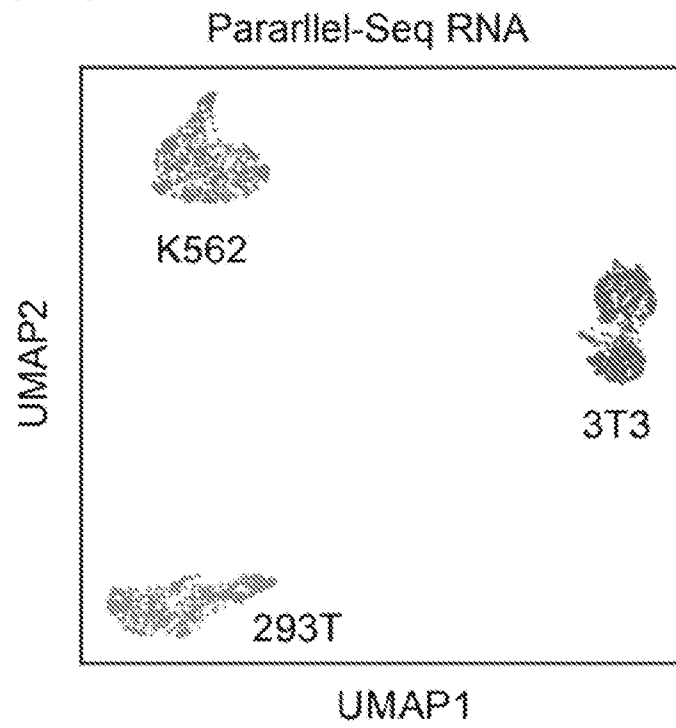
[图 6]



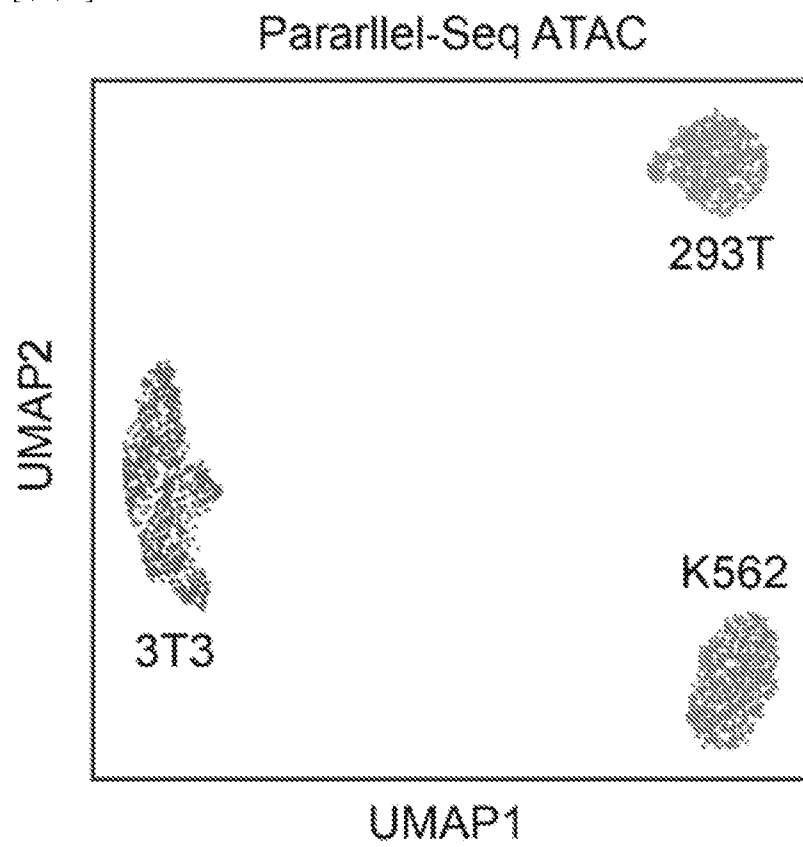
[图 7]



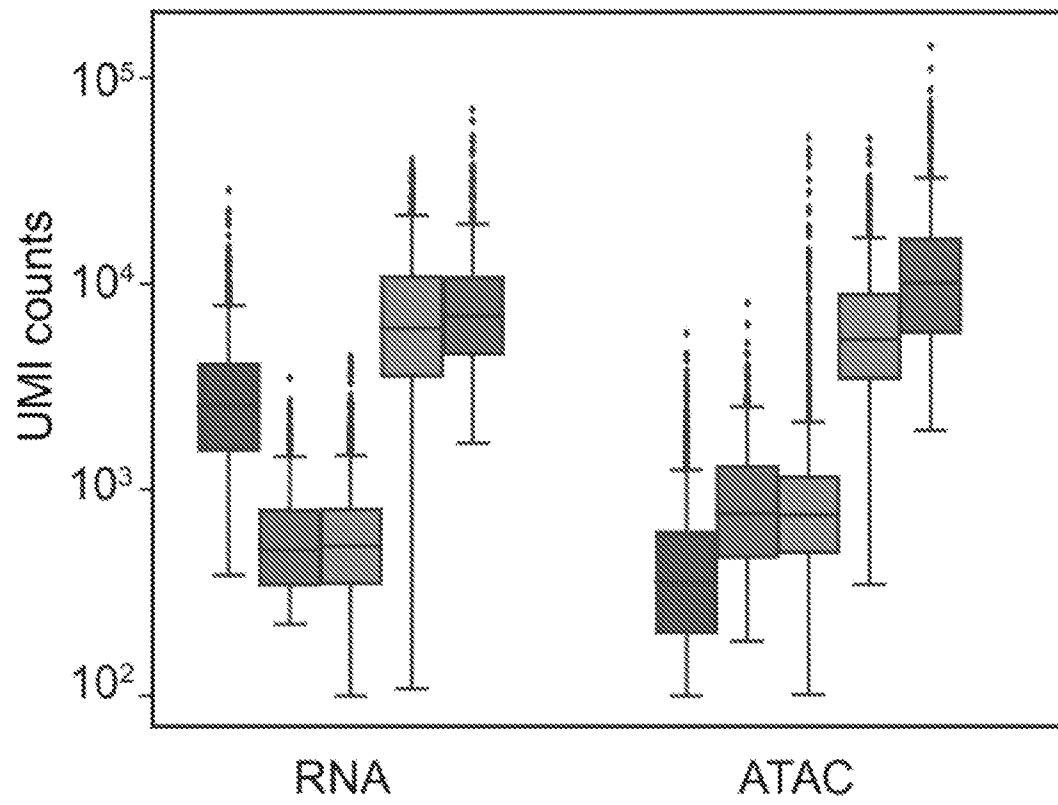
[图 8]



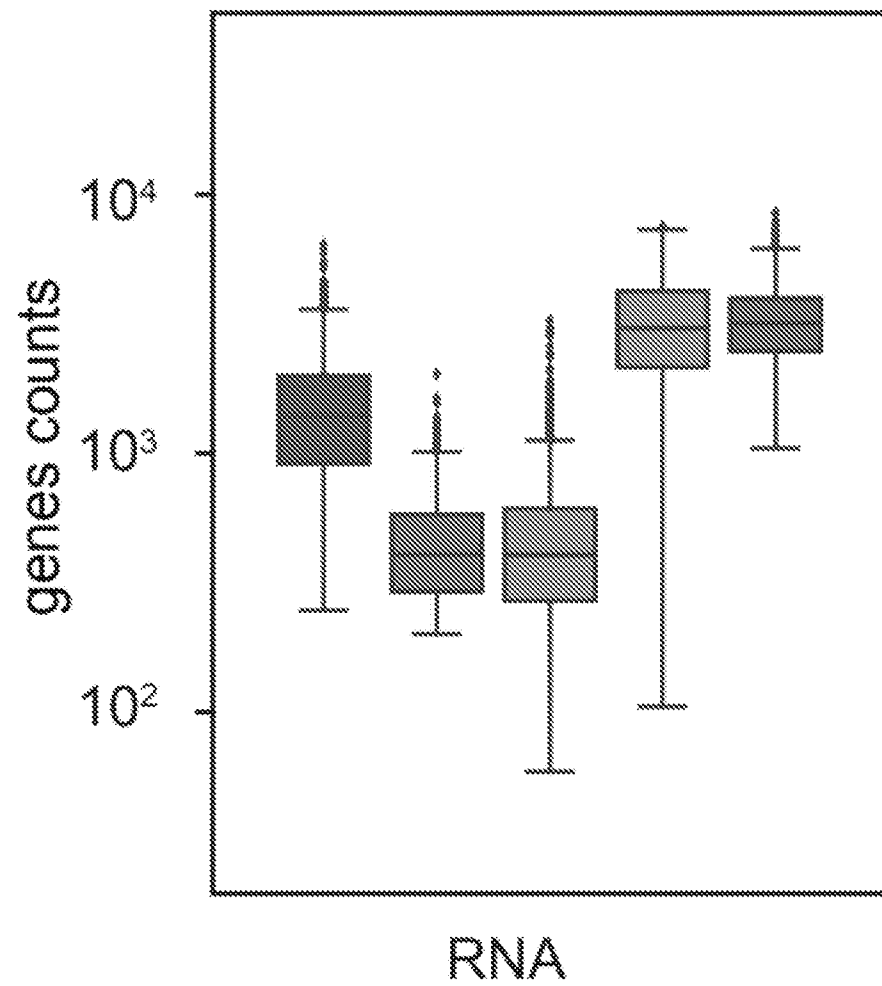
[图 9]



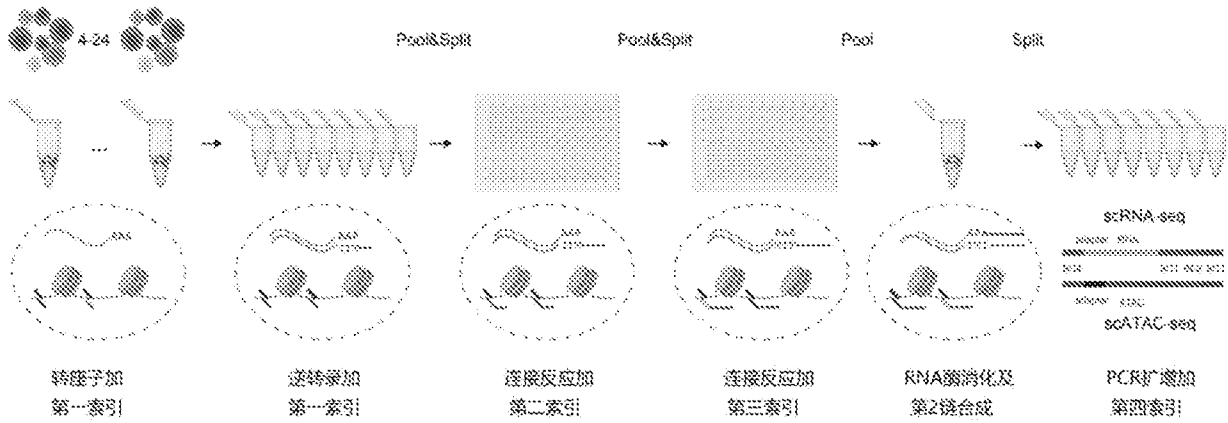
[图 10]



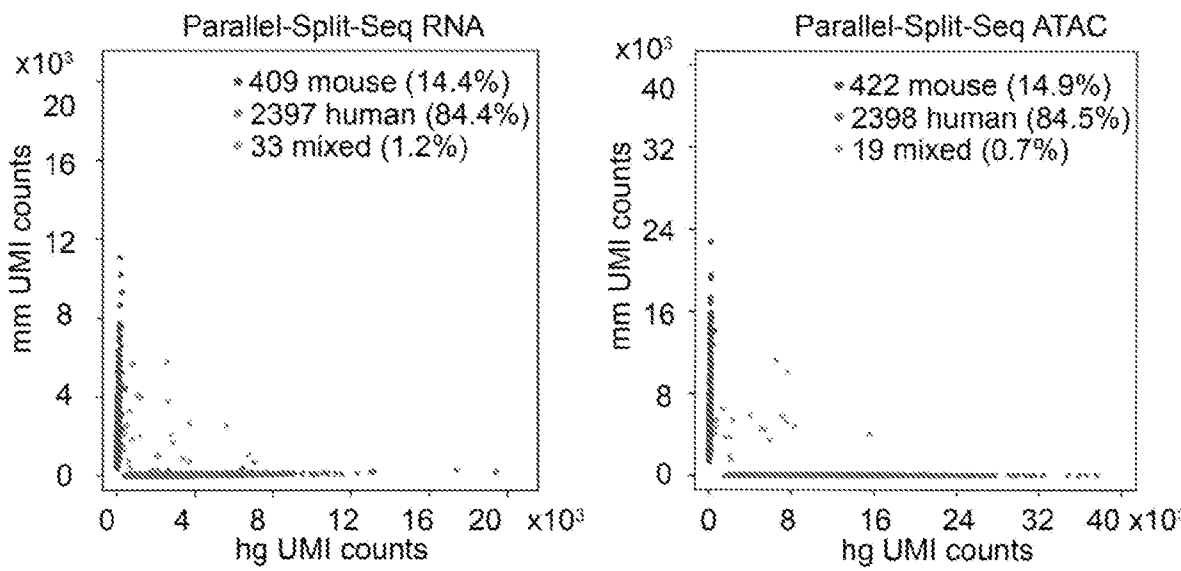
[图 11]



[图 12]

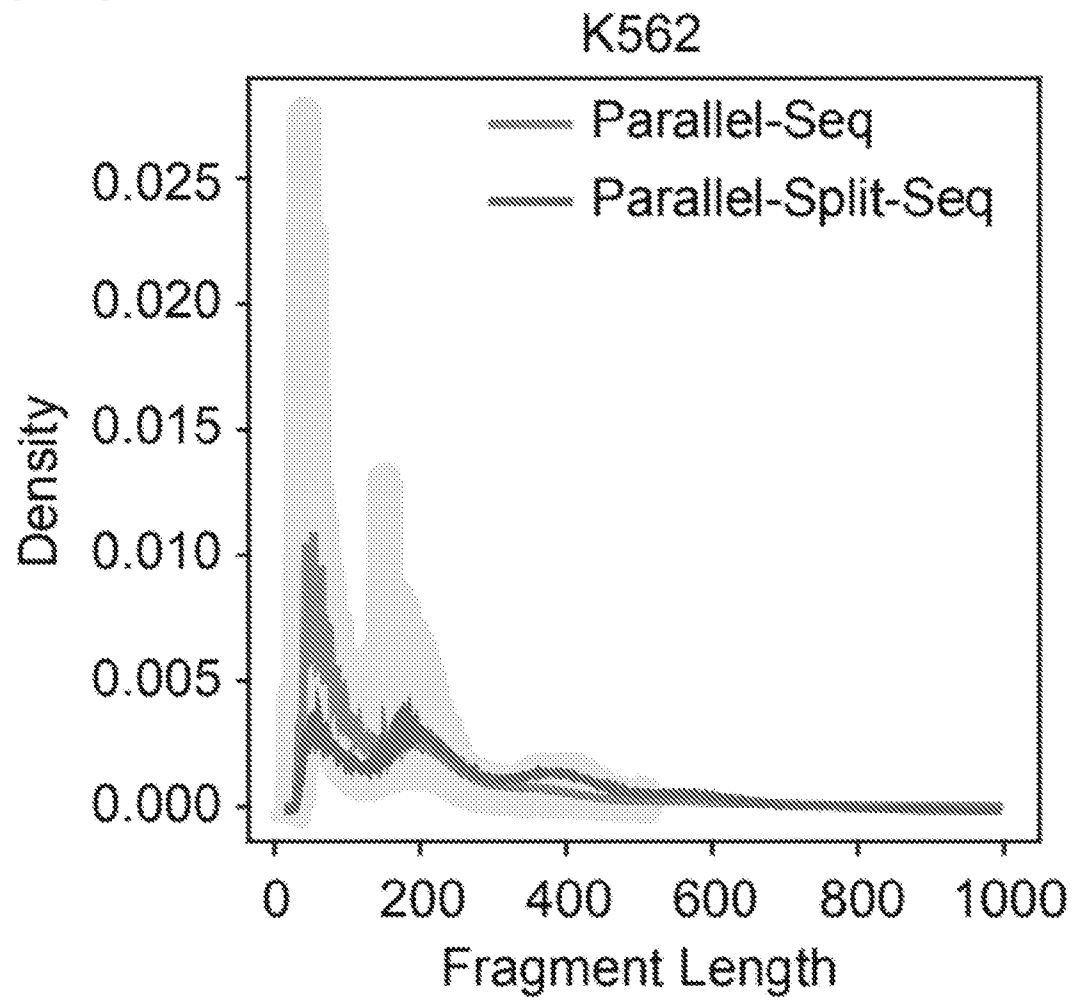


[图 13]

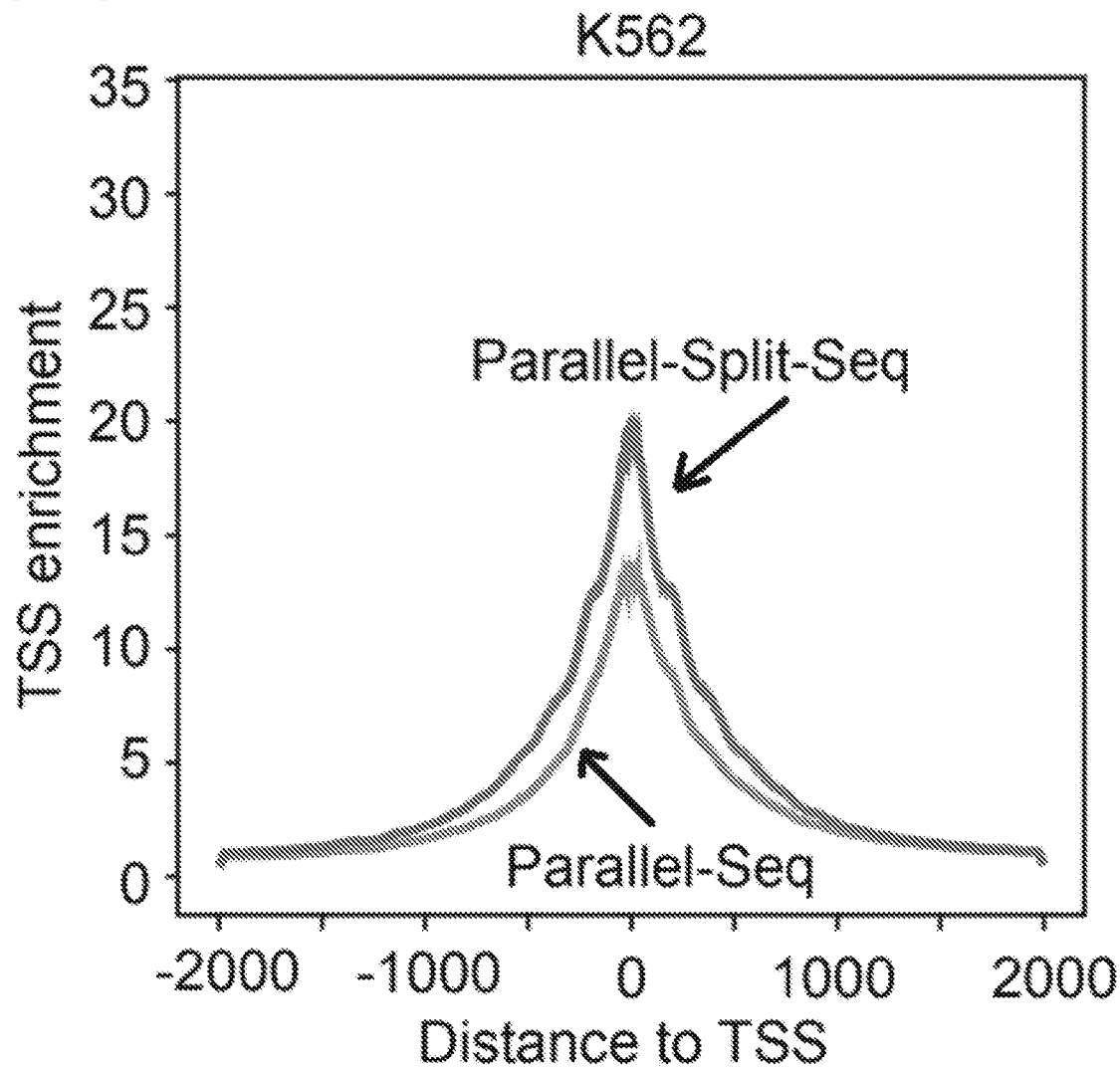


细则 26,
28.06.2023

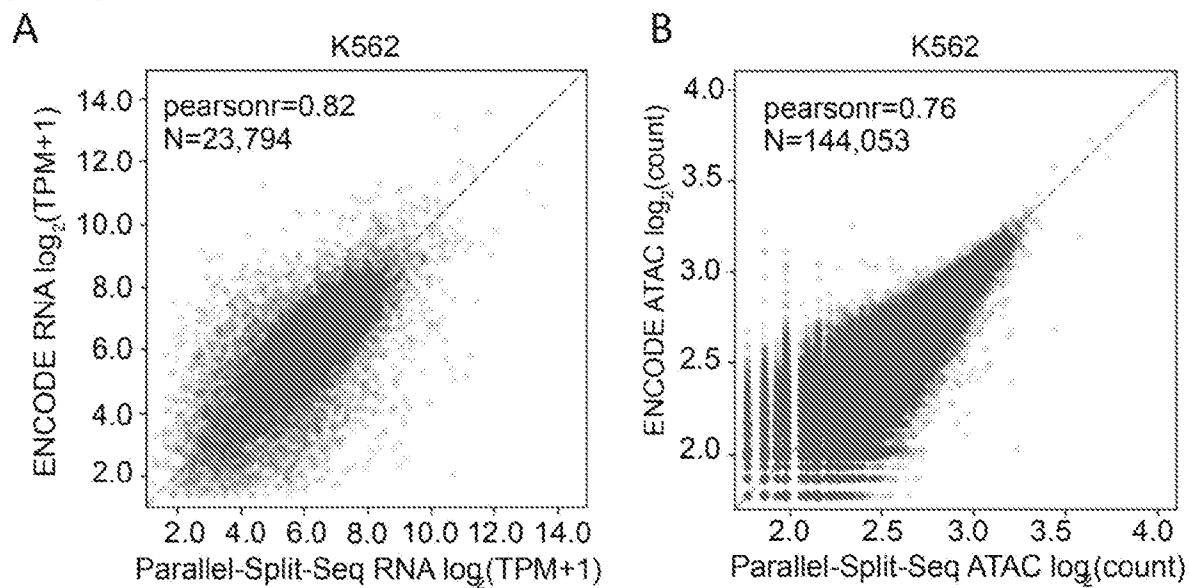
[图 14]



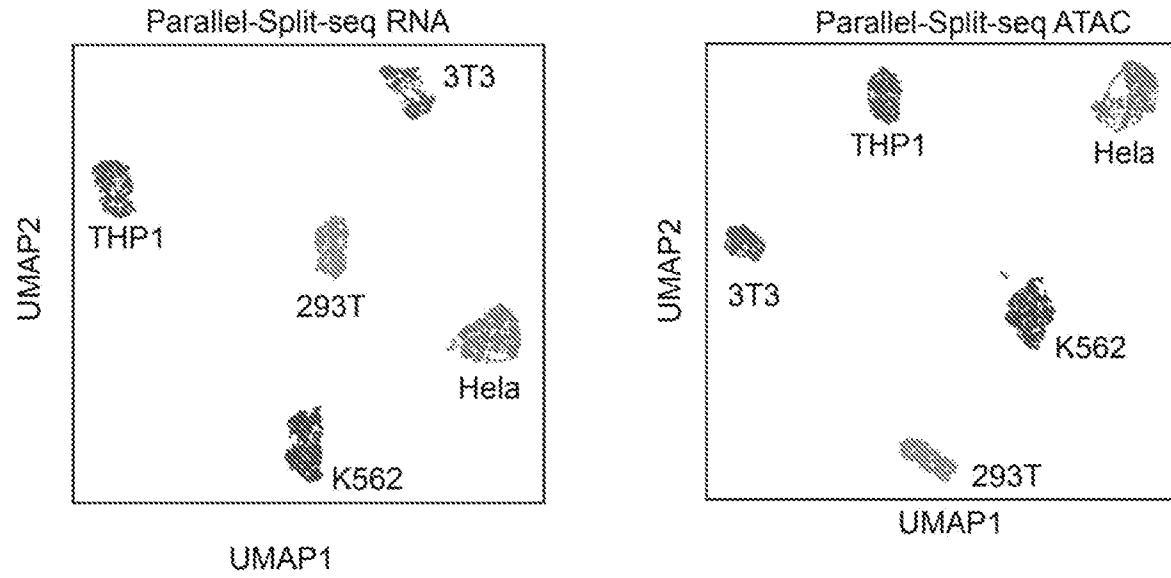
[图 15]



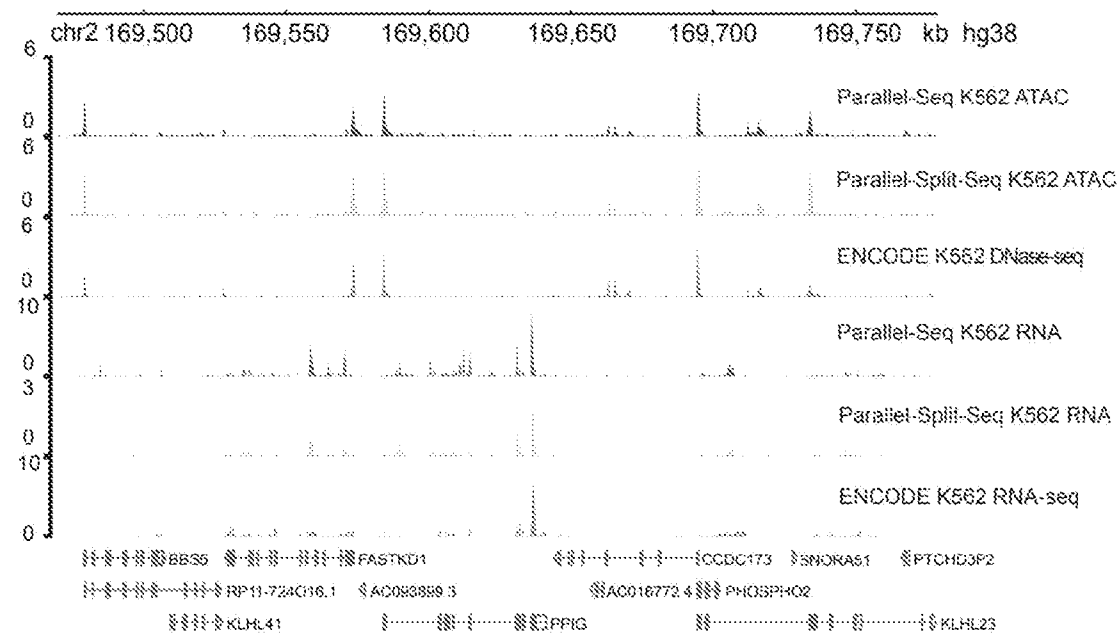
[图 16]



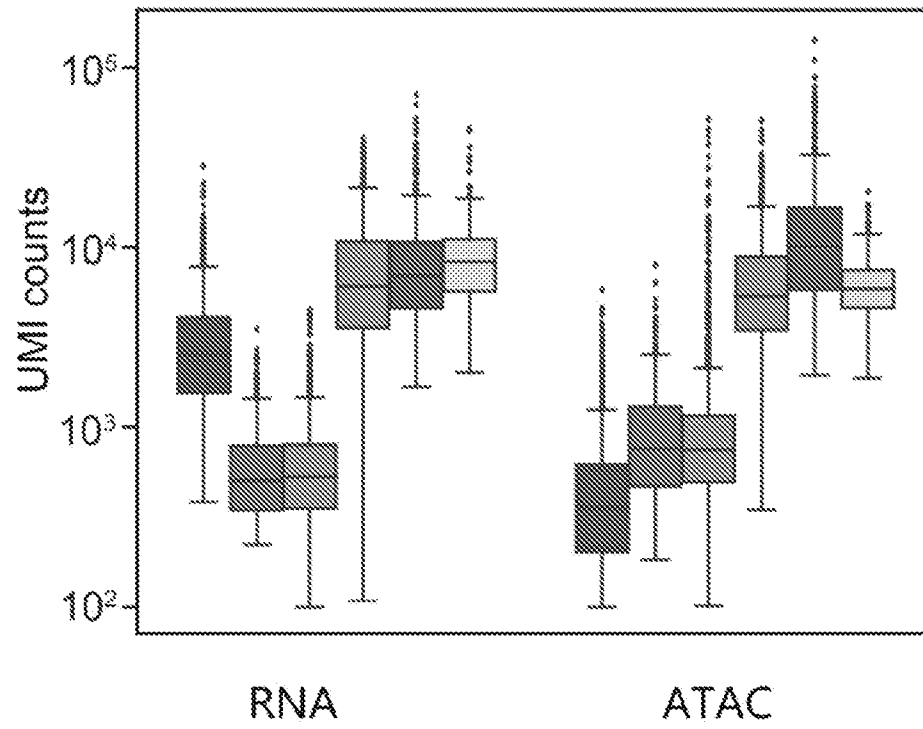
[图 17]



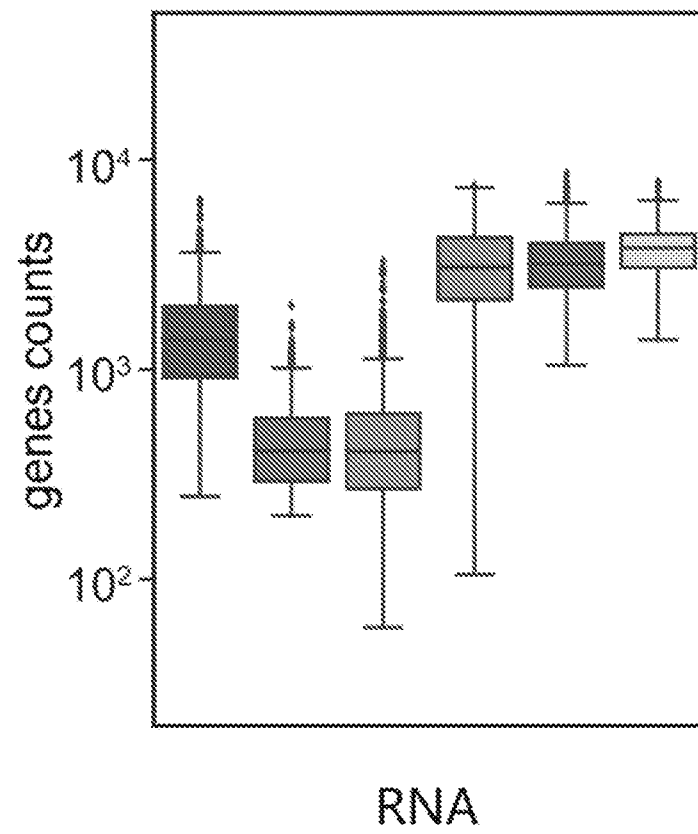
[图 18]



[图 19]



[图 20]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/098377

A. CLASSIFICATION OF SUBJECT MATTER

C12Q1/6806(2018.01)i; C12Q1/6869(2018.01)i; C12Q1/6883(2018.01)i; C12Q1/6886(2018.01)i; C40B50/06(2006.01)i; C40B40/06(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: C12Q, C40B

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNMED, CNABS, DWPI, WPABS, WPABSC, CJFD, AUABS, TWMED, ILABS, HKABS, MOABS, CNTXT, ENTXT, ENTXTC, WOTXT, USTXT, VCN, VEN, GBTXT, EPTXT, CATXT, CNKI, PubMed, EMBL, Web of Science, GoogleScholar: 单细胞, 测序, 文库, 转座子, 染色质, 逆转录, 条形码序列, single-cell, sequencing, library, transposon, chromatin, reverse transcription, barcode, Parallel-seq; 中国专利生物序列检索系统, Chinese Patent Biological Sequence Retrieval System, NCBI, EMBL, STN: 基于SEQ ID NO: 1-6的序列检索, sequence search on the basis of SEQ ID NO: 1-6

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
E	CN 116949132 A (TSINGHUA UNIVERSITY) 27 October 2023 (2023-10-27) claims 1-18	1-18
A	CN 115537408 A (XIAMEN UNIVERSITY) 30 December 2022 (2022-12-30) claims 1-4	1-18
A	CN 109996892 A (SHENZHEN MGI TECHNOLOGY CO., LTD.) 09 July 2019 (2019-07-09) claims 1-3	1-18
A	CN 115478098 A (UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA) 16 December 2022 (2022-12-16) abstract	1-18
A	EP 3553180 A1 (SHENZHEN MGI TECH CO., LTD.) 16 October 2019 (2019-10-16) abstract	1-18



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance
 “D” document cited by the applicant in the international application
 “E” earlier application or patent but published on or after the international filing date
 “L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
 “O” document referring to an oral disclosure, use, exhibition or other means
 “P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

01 February 2024

Date of mailing of the international search report

11 February 2024

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
China No. 6, Xitucheng Road, Jimenqiao, Haidian District,
Beijing 100088

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/098377**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2022259586 A1 (10X GENOMICS, INC.) 18 August 2022 (2022-08-18) abstract	1-18
A	US 2022356461 A1 (ILLUMINA, INC.; UNIVERSITY OF WASHINGTON) 10 November 2022 (2022-11-10) abstract	1-18
A	WO 2021189679 A1 (ARMY MEDICAL UNIVERSITY) 30 September 2021 (2021-09-30) abstract	1-18

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/098377

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
 - a. ☐ forming part of the international application as filed.
 - b. ☒ furnished subsequent to the international filing date for the purposes of international search (Rule 13ter.1(a)),
☒ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.
2. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.
3. Additional comments:

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2023/098377

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	116949132	A	27 October 2023	None			
CN	115537408	A	30 December 2022	None			
CN	109996892	A	09 July 2019	DK	3553180	T3	25 July 2022
				US	2019338279	A1	07 November 2019
				US	11434483	B2	06 September 2022
				WO	2018103025	A1	14 June 2018
				ES	2922281	T3	12 September 2022
				EP	3553180	A1	16 October 2019
				EP	3553180	A4	12 August 2020
				EP	3553180	B1	04 May 2022
CN	115478098	A	16 December 2022	None			
EP	3553180	A1	16 October 2019	None			
US	2022259586	A1	18 August 2022	US	11773389	B2	03 October 2023
				EP	3445876	A1	27 February 2019
				EP	3445876	A4	11 December 2019
				EP	3445876	B1	05 July 2023
				SG	11201901822	QA	28 March 2019
				EP	4230746	A2	23 August 2023
				EP	4230746	A3	01 November 2023
US	2022356461	A1	10 November 2022	BR	112021019640	A2	21 June 2022
				AU	2020407641	A1	23 September 2021
				EP	3927824	A2	29 December 2021
				IL	286643	A	01 December 2021
				CA	3134746	A1	24 June 2021
				WO	2021127436	A2	24 June 2021
				WO	2021127436	A3	29 July 2021
				MX	2021011847	A	17 November 2021
				SG	11202109486	QA	29 September 2021
				JP	2023508792	A	06 March 2023
				KR	20220118295	A	25 August 2022
WO	2021189679	A1	30 September 2021	None			

A. 主题的分类

C12Q1/6806(2018.01)i; C12Q1/6869(2018.01)i; C12Q1/6883(2018.01)i; C12Q1/6886(2018.01)i; C40B50/06(2006.01)i; C40B40/06(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

IPC: C12Q, C40B

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNMED, CNABS, DWPI, WPABS, WPABSC, CJFD, AUABS, TWMED, ILABS, HKABS, MOABS, CNTXT, ENTXT, ENTXTC, WOTXT, USTXT, VCN, VEN, GBTXT, EPTXT, CATXT, CNKI, PubMed, EMBL, Web of Science, GoogleScholar: 单细胞, 测序, 文库, 转座子, 染色质, 逆转录, 条形码序列, single-cell, sequencing, library, transposon, chromatin, reverse transcription, barcode, Parallel-seq 中国专利生物序列检索系统, NCBI, EMBL, STN: 基于SEQ ID NO: 1-6的序列检索

C. 相关文件

类型*	引用文件, 必要时, 指明相关段落	相关的权利要求
E	CN 116949132 A (清华大学) 2023年10月27日 (2023 - 10 - 27) 权利要求1-18	1-18
A	CN 115537408 A (厦门大学) 2022年12月30日 (2022 - 12 - 30) 权利要求1-4	1-18
A	CN 109996892 A (深圳华大智造科技有限公司) 2019年7月9日 (2019 - 07 - 09) 权利要求1-3	1-18
A	CN 115478098 A (中国科学技术大学) 2022年12月16日 (2022 - 12 - 16) 摘要	1-18
A	EP 3553180 A1 (MGI TECH CO LTD) 2019年10月16日 (2019 - 10 - 16) 摘要	1-18
A	US 2022259586 A1 (10X GENOMICS INC) 2022年8月18日 (2022 - 08 - 18) 摘要	1-18

☒ 其余文件在C栏的续页中列出。

☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“D” 申请人在国际申请中引证的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2024年2月1日

国际检索报告邮寄日期

2024年2月11日

ISA/CN的名称和邮寄地址

中国国家知识产权局
中国北京市海淀区蓟门桥西土城路6号 100088

授权官员

张艳霞

电话号码 (+86) 010-62089438

PCT/ISA/210 表(第2页) (2022年7月)

第I栏

核苷酸和/或氨基酸序列(续第1页第1.c项)

1. 关于国际申请中所公开的任何核苷酸和/或氨基酸序列,国际检索是基于下列序列表进行的:

a. ☐ 作为国际申请的一部分提交的:

b. ☒ 为国际检索的目的在国际申请日之后提交(细则13之三.1(a)),

☒ 附有说明序列表不超出所提交国际申请公开范围的声明。

2. ☐ 本报告是在没有收到符合WIPO ST.26标准的序列表的情况下,考虑了国际申请中披露的任何核苷酸和/或氨基酸序列,在可进行有意义检索的范围内做出的。

3. 补充意见:

C. 相关文件		
类 型*	引用文件，必要时，指明相关段落	相关的权利要求
A	US 2022356461 A1 (ILLUMINA INC;UNIV WASHINGTON;) 2022年11月10日 (2022 - 11 - 10) 摘要	1-18
A	WO 2021189679 A1 (UNIV ARMY MEDICAL) 2021年9月30日 (2021 - 09 - 30) 摘要	1-18

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2023/098377

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	116949132	A	2023年10月27日	无			
CN	115537408	A	2022年12月30日	无			
CN	109996892	A	2019年7月9日	DK	3553180	T3	2022年7月25日
				US	2019338279	A1	2019年11月7日
				US	11434483	B2	2022年9月6日
				WO	2018103025	A1	2018年6月14日
				ES	2922281	T3	2022年9月12日
				EP	3553180	A1	2019年10月16日
				EP	3553180	A4	2020年8月12日
				EP	3553180	B1	2022年5月4日
CN	115478098	A	2022年12月16日	无			
EP	3553180	A1	2019年10月16日	无			
US	2022259586	A1	2022年8月18日	US	11773389	B2	2023年10月3日
				EP	3445876	A1	2019年2月27日
				EP	3445876	A4	2019年12月11日
				EP	3445876	B1	2023年7月5日
				SG	11201901822	QA	2019年3月28日
				EP	4230746	A2	2023年8月23日
				EP	4230746	A3	2023年11月1日
US	2022356461	A1	2022年11月10日	BR	112021019640	A2	2022年6月21日
				AU	2020407641	A1	2021年9月23日
				EP	3927824	A2	2021年12月29日
				IL	286643	A	2021年12月1日
				CA	3134746	A1	2021年6月24日
				WO	2021127436	A2	2021年6月24日
				WO	2021127436	A3	2021年7月29日
				MX	2021011847	A	2021年11月17日
				SG	11202109486	QA	2021年9月29日
				JP	2023508792	A	2023年3月6日
				KR	20220118295	A	2022年8月25日
WO	2021189679	A1	2021年9月30日	无			