



(12)发明专利申请

(10)申请公布号 CN 110431237 A

(43)申请公布日 2019.11.08

(21)申请号 201780087116.5

(22)申请日 2017.12.28

(30)优先权数据

62/440,089 2016.12.29 US

(85)PCT国际申请进入国家阶段日

2019.08.21

(86)PCT国际申请的申请数据

PCT/US2017/068672 2017.12.28

(87)PCT国际申请的公布数据

W02018/125982 EN 2018.07.05

(71)申请人 伊鲁米纳公司

地址 美国加利福尼亚州

(72)发明人 R·兰吉 F·J·斯特梅斯

L·克里斯琴森 D·K·波霍洛克

F·张

(74)专利代理机构 北京市金杜律师事务所

11256

代理人 陈文平 袁元

(51)Int.Cl.

C12Q 1/6804(2006.01)

C12N 15/10(2006.01)

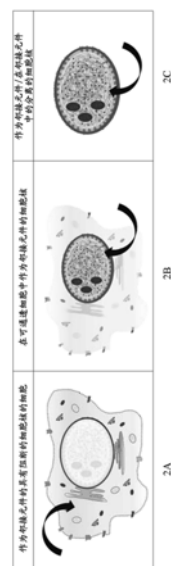
权利要求书4页 说明书23页 附图14页

(54)发明名称

用于在细胞区室中正交接近和标签化生物分子的分析系统

(57)摘要

本发明涉及通过用核通透性增强剂处理细胞核来增强分析性生物分子(例如转座体复合物)对细胞核信息分子(例如DNA、RNA和蛋白质)的接近的系统和方法,以及使用细胞核膜、细胞膜和外部区室化方法作为邻接保持元件的方法。



1. 一种使细胞核信息分子与分析性生物分子反应的方法,其包括使所述细胞核与核通透性增强剂接触并使所述信息分子与所述分析性生物分子反应。

2. 一种分析细胞核信息分子的方法,其包括:

(a) 使包含细胞核信息分子的细胞核与核通透性增强剂和分析性生物分子接触;

(b) 使所述分析性生物分子与所述细胞核信息分子反应以提供分析复合物;和

(c) 分析所述分析复合物,从而检测所述细胞核信息分子。

3. 根据权利要求2所述的方法,其中邻接保持元件包含所述细胞核。

4. 根据权利要求1-3或25中任一项所述的方法,其中所述细胞核信息分子是DNA、RNA、蛋白质或其混合物。

5. 根据权利要求4所述的方法,其中所述细胞核信息分子是DNA。

6. 根据权利要求1-5中任一项所述的方法,其中所述分析性生物分子是转座酶或转座体复合物、或抗体、或寡核苷酸、或核苷酸、或逆转录引物、或酶。

7. 根据权利要求6所述的方法,其中所述寡核苷酸或核苷酸包含至少一种标记的核苷酸。

8. 根据权利要求6所述的方法,其中所述酶是扩增酶、聚合酶、DNA聚合酶、RNA聚合酶、PCR酶、连接酶、Taq DNA聚合酶、Pfu DNA聚合酶、介导体外转录的酶、整合酶或切口酶。

9. 根据权利要求6所述的方法,其中所述分析性生物分子是转座体复合物。

10. 根据权利要求9所述的方法,其中各转座体复合物包含转座酶和两个包含转座子末端序列的转座子末端组合物。

11. 根据权利要求10所述的方法,其中所述细胞核信息分子是靶核酸,其中所述反应包括将所述靶核酸片段化成双链核酸片段,和用来自所述转座子末端组合物的转移链标签化以形成标签化分析复合物。

12. 一种从细胞核靶核酸制备标签化核酸片段的文库的方法:

(a) 提供包含单个核的邻接保持元件,其中所述单个核包含靶核酸;

(b) 使所述邻接保持元件和单个核与核通透性增强剂和分析性生物分子接触;和

(c) 使所述分析性生物分子与所述细胞核信息分子反应以提供分析复合物;和

(d) 分析所述分析复合物,从而检测所述细胞核信息分子。

13. 根据前述权利要求中任一项所述的方法,其还包括在接触之前,在存在核孔阻断剂和不存在核通透性增强剂的情况下使细胞质信息分子与第二分析性生物分子反应,其中方法反应步骤将第一标签引入所述细胞质信息分子,并且将第二标签引入所述细胞核信息分子。

14. 根据前述权利要求中任一项所述的方法,其中至少一种信息分子是蛋白质。

15. 根据前述权利要求中任一项所述的方法,其还包括通过使所述细胞与第二分析性生物分子接触来分析细胞质信息分子,使所述第二分析性生物分子与所述细胞质信息分子反应以提供第二分析复合物,和任选地,分析所述第二分析复合物,从而检测所述细胞质信息分子。

16. 根据前述权利要求中任一项所述的方法,其中所述分析性生物分子是转座体复合物,并且所述方法还包括用聚合酶处理所述分析复合物,其中所述聚合酶任选地是链置换聚合酶。

17. 根据前述权利要求中任一项所述的方法, 其中分析包括对片段化和标签化核酸或其扩增子或其拷贝进行测序。

18. 根据权利要求中任一项所述的方法, 其中所述核通透性增强剂选自: 破坏NPC疏水相互作用的化合物、结合和/或抑制核丝蛋白的化合物、结合和/或抑制网格蛋白的化合物, 和核定位信号肽(NLS)。

19. 根据权利要求18所述的方法, 其中所述核通透性增强剂是网格蛋白抑制剂, 或者是Pitstop-2(也称为N-[5-(4-溴苄基)-4-氧代-4,5-二氢-1,3-噻唑-2-基]萘-1-磺酰胺)、甲基-b-环糊精、吩噻嗪、单丹酰基尸胺、氯喹、莫能菌素、高渗蔗糖、或Dynasore, 或其合成类似物, 或者是Pitstop-2。

20. 根据权利要求18所述的方法, 其中所述核通透性增强剂是疏水性破坏剂, 或是脂肪族醇, 或是C₄₋₁₀-烷基二醇, 或是环状二醇, 或是环烷二醇, 或是邻二醇, 或是反式-1,2-环己二醇、正己烷-1,2-二醇、或1,6-己二醇。

21. 根据权利要求18所述的方法, 其中所述核通透性增强剂是核定位信号肽(NLS), 或是SV40大T抗原(PKKKRKV)、核质蛋白的NLS(KR[PAATKKAGQA]KKKK或AVKRPAATKKAGQAKKKLD)、K-K/R-X-K/R、EGL-13(MSRRRKANPTKLS ENAKKLAKEVEN)、c-Myc(PAAKRVKLD)、TUS蛋白(KLKIKRPVK)、hnRNP A1的酸性M9结构域、来自酵母转录阻遏物Mat α 2的KIPIK、PY-NLS序列、或输入蛋白 β 2的抑制剂, 或是SV40大T抗原。

22. 根据权利要求21所述的方法, 其中所述NLS与所述分析性生物分子共价连接。

23. 根据权利要求21所述的方法, 其中所述NLS不与所述分析性生物分子共价连接。

24. 一种组合物, 其包含一个或多个细胞和/或一个或多个细胞核、核通透性增强剂和分析性生物分子。

25. 根据权利要求24所述的组合物, 其中所述分析性生物分子是转座酶或转座体复合物、或抗体、或任选地包含至少一种标记的核苷酸的寡核苷酸、或任选地标记的核苷酸、或逆转录引物、或酶。

26. 根据权利要求24所述的组合物, 其中所述分析性生物分子是转座体复合物。

27. 根据权利要求24所述的方法, 其中所述分析性生物分子是酶, 其中所述酶是扩增酶、聚合酶、DNA聚合酶、RNA聚合酶、PCR酶、Taq DNA聚合酶、Pfu DNA聚合酶、介导体外转录的酶、整合酶或切口酶。

28. 根据权利要求24-27中任一项所述的组合物, 其中所述核通透性增强剂是Pitstop-2、脂肪族醇、C₄₋₁₀-烷基二醇、环状二醇、环烷二醇、邻二醇、反式-1,2-环己二醇、正己烷-1,2-二醇、1,6-己二醇或洋地黄皂苷。

29. 根据权利要求28所述的组合物, 其中所述核通透性增强剂是Pitstop-2、环己二醇或洋地黄皂苷。

30. 根据权利要求1所述的方法, 其包括:

(a) 使包含细胞核信息分子的细胞核与核通透性增强剂和分析性生物分子接触;

(b) 使所述分析性生物分子与所述细胞核信息分子反应以提供修饰的细胞核信息分子; 和

(c) 检测所述修饰的细胞核信息分子。

31. 根据权利要求30所述的方法, 其中所述细胞核信息分子是靶核酸, 所述分析性生物

分子是转座体复合物,并且所述反应包括在所述转座体复合物和所述细胞核信息分子之间形成复合物,其中所述转座体复合物不使所述靶核酸片段化。

32. 一种使细胞中的信息分子有区别地标签化的方法,其包括:

将包含第一标签的第一分析性生物分子选择性地递送至选自细胞核、细胞质和线粒体的第一细胞区室,其中所述第一细胞区室包含第一信息生物分子;

使所述第一分析性生物分子与所述第一信息分子反应以提供第一标签化信息分子;

将包含第二标签的第二分析性生物分子选择性地递送至选自细胞核、细胞质和线粒体的第二细胞区室,其中所述第二细胞区室包含第二信息分子并且不同于所述第一细胞区室;和

使所述第二分析性生物分子与所述第二信息分子反应以提供第二标签化信息分子,其中所述第一标签和第二标签不同。

33. 根据权利要求32所述的方法,其中将所述第一分析性生物分子选择性地递送至所述第一细胞区室包括用用于所述第一细胞区室的通透性增强剂处理所述细胞。

34. 根据权利要求32或权利要求33所述的方法,其中将所述第一分析性生物分子选择性地递送至所述第一细胞区室包括用用于所述第二细胞区室的通透性阻断剂处理所述细胞。

35. 根据权利要求32-34中任一项所述的方法,其中将所述第二分析性生物分子选择性地递送至所述第二细胞区室包括用用于所述第二细胞区室的通透性增强剂处理所述细胞。

36. 根据权利要求32-35中任一项所述的方法,其中选择性地递送所述第一分析性生物分子是在基本上没有将所述第一分析性生物分子递送至所述第二细胞区室的情况下发生。

37. 根据权利要求32-36中任一项所述的方法,其中所述第一细胞区室是所述细胞质,并且所述第一信息分子是细胞质信息分子,和(a)所述第二细胞区室是所述细胞核,并且所述第二信息分子是细胞核信息分子或(b)所述第二细胞区室是所述线粒体,并且所述第二信息分子是线粒体信息分子。

38. 根据权利要求37所述的方法,其中将所述第一分析性生物分子选择性地递送至所述细胞质是在核孔阻断剂和/或线粒体孔阻断剂的存在下完成。

39. 根据权利要求37所述的方法,其中递送所述第二分析性生物分子是在核通透性增强剂或线粒体通透性增强剂的存在下完成。

40. 根据权利要求37-39中任一项所述的方法,其中所述细胞质信息分子是RNA。

41. 根据权利要求37-39中任一项所述的方法,其中所述细胞核信息分子是DNA或基因组DNA(gDNA)。

42. 根据权利要求32-36中任一项所述的方法,其中:

选择性地递送所述第一分析性生物分子包括使所述第一分析性生物分子与线粒体信息分子反应以产生标签化线粒体信息分子;和

选择性地递送所述第二分析性生物分子包括使所述第二分析性生物分子与细胞核信息分子反应以产生标签化细胞核信息分子。

43. 根据权利要求37-42中任一项所述的方法,其中所述线粒体信息分子是DNA。

44. 根据权利要求37-43中任一项所述的方法,其中所述细胞核信息分子是DNA,或是基因组DNA(gDNA)。

45. 根据权利要求32-44中任一项所述的方法, 其还包括裂解所述细胞以释放所述标签化信息分子。

46. 根据前述方法权利要求中任一项所述的方法, 其中所述分析性生物分子是转座酶或转座体复合物、或抗体、或任选地包含至少一种标记的核苷酸的寡核苷酸、或任选地标记的核苷酸、或逆转录引物、或酶。

47. 根据权利要求45所述的方法, 其中所述酶是扩增酶、聚合酶、DNA聚合酶、RNA聚合酶、PCR酶、Taq DNA聚合酶、Pfu DNA聚合酶、介导体外转录的酶、整合酶或切口酶。

48. 根据权利要求32-46中任一项所述的方法, 其中所述分析性生物分子是转座体复合物。

49. 根据权利要求32-48中任一项所述的方法, 其还包括检测所述标签化信息分子或其扩增子, 其中所述检测任选地包括检测所述标签化信息分子或扩增子的序列。

50. 根据前述权利要求中任一项所述的方法或组合物, 其中所述核通透性增强剂是洋地黄皂苷。

用于在细胞区室中正交接近和标签化生物分子的分析系统

背景技术

[0001] 在生物样本中检测核酸和蛋白质可用于鉴定和分类微生物、诊断传染病、检测和表征遗传异常、鉴定与疾病发作或进展相关的遗传变化、研究对疾病的遗传易感性、以及测量对疾病治疗的响应。

[0002] 对单细胞或单细胞核的分析,或允许细胞质和细胞核区室的离散分析,提供对细胞类型、细胞分化、细胞状态、蛋白质合成和调节、进化、疾病进展和诊断以及对疾病治疗的响应的评估的广泛见解。特别是,在核水平上对DNA和RNA以及组蛋白和其他核蛋白的分析可用于进行完整的基因组测序,以揭示关于细胞的静止状态和开放染色质状态的信息,或提供关于蛋白质调节的实时信息。这样的信息可用于基因编辑、细胞类型转换、蛋白质调节分析和疾病疗法等应用。

[0003] 用于接近细胞核酸和蛋白质内容物以进行后续分析的当前技术通常采用破坏细胞区室的细胞裂解方法。在一些方法中,使用例如离子去垢剂的裂解介质,产生细胞质和核内容物的混合物并阻止区室之间的分子信息的解析(例如,由于细胞质线粒体DNA和核DNA之间的交叉污染)。替代的裂解方法使用温和的非离子去垢剂来破坏细胞膜,同时保持细胞核完整。在另一种方法中,可以用消化酶例如蛋白酶破坏分离的细胞核。

[0004] 下一代测序(NGS)技术通常采用样品或文库制备步骤,其中基因组DNA或RNA被转化为片段化的、可测序的模板的文库。基因组DNA的片段化是用于高通量测序的DNA样品制备中的关键步骤。在一种方法中,转座体复合物用于对靶核酸进行片段化和标签化(tag)。转座酶介导双链DNA的片段化,并在两端连接合成寡核苷酸。附加的寡核苷酸使得能够进行后续的扩增和测序步骤。前文讨论的细胞和核裂解方法与这样的方法不相容,因为它们降解酶(例如,离子去垢剂,消化酶),或者因为酶不能充分穿透完整的细胞核。例如,用于当前的转座片段化(tagmentation)方案中的转座体复合物,例如Nextera Tn5(二聚体~106kDa),由于核被膜的复杂性而不能有效地接近细胞核物质(参见图1)。核被膜由外核膜和内核膜组成,它们一起形成限制生物分子从细胞质扩散到细胞核中的脂质双层。核孔复合物(NPC)跨越细胞核膜并紧密调节生物分子进出细胞核的转运,通常只允许大小小于40kDa的分子穿过。与内核膜相邻的是核纤层,其包括蛋白质细丝,例如骨架/基质附着成分、网格蛋白和其他蛋白质,其创造支持框架以维持核刚性并控制分子到细胞核中的尺寸选择进入。由于酶降解和穿细胞核膜的受限扩散的这些问题,细胞裂解方法将需要另外的纯化和分离步骤以在进一步样品制备可以实现之前分离靶核酸或蛋白质。

[0005] 对单细胞内容物的分析可以通过在离散区室中分离单细胞来完成。在一种技术中,细胞分布到油介质中的水滴中。然而,油妨碍材料和样品制备酶进出水性液滴,因此所有这样的试剂必须存在于初始水性介质中。此外,可以使用区室化方法,例如PCT公开号W02016/013704描述的邻接保持元件(contiguity preserving element,CE),来保持邻接信息。然而,这样的技术没有解决上面讨论的关于离散地接近细胞核和细胞质组分的限制。

[0006] 存在对用于分析细胞组分的样品制备方法的需要,例如,通过下一代测序,其提供对核遗传物质的有效接近,允许独立检测细胞核和细胞质内容物,并产生用于测序的合适

的片段化DNA或RNA文库,同时保持细胞质和核区室和/或核邻接。本文讨论的是为在完整细胞核内进行样品制备方法做准备的方法,实质上使用核被膜作为CE。

发明内容

[0007] 本公开涉及通过将细胞和/或细胞核与核通透性增强剂接触来递送分析性生物分子到细胞核中,和使分析性生物分子与细胞核信息分子(例如核酸或蛋白)反应的方法。在一些方面,方法包括使第二分析性生物分子与细胞质信息分子(例如核酸或蛋白质)反应。在一些方面,方法包括促进个体细胞和/或个体细胞核的分析的步骤。在一些方面,单个细胞和/或细胞核内容物用邻接保持元件定位。在一些方面,采用多个邻接保持元件,允许独立分析来自不同细胞区室的分析物。在一些方面,核膜用作邻接保持元件,或用作多个邻接保持元件中的一个。这样的处理还允许从细胞核中除去大的生物分子,例如分析复合物,以促进分离和/或分析。

[0008] 本文描述的是使细胞核信息分子与分析性生物分子反应的方法,其包括使细胞核与核通透性增强剂接触并使所述信息分子与所述分析性生物分子反应。

[0009] 本文描述的是分析细胞核信息分子的方法,其包括:

[0010] (a) 使包含细胞核信息分子的细胞核与核通透性增强剂和分析性生物分子接触;

[0011] (b) 使所述分析性生物分子与所述细胞核信息分子反应以提供分析复合物;和

[0012] (c) 分析所述分析复合物,从而检测所述细胞核信息分子。

[0013] 这样的方法可以包括分析细胞质信息分子。因此,本公开涉及分析核信息分子和细胞质信息分子的方法,其包括:

[0014] (a) 使包含细胞核信息分子和细胞质信息分子的细胞与核通透性增强剂、第一分析性生物分子和第二分析性生物分子接触;

[0015] (b) 使第一分析性生物分子与细胞核信息分子反应以提供第一分析复合物,并使第二分析性生物分子与细胞质信息分子反应以提供第二分析复合物;

[0016] (c) 分析第一分析复合物,从而检测细胞核信息分子;和,任选地,

[0017] (d) 分析第二分析复合物,从而检测细胞质信息分子。

[0018] 本文所述的方法可以利用邻接保持元件。因此,本公开涉及分析细胞核信息分子的方法,其包括:

[0019] (a) 提供包含细胞核的邻接保持元件,其中细胞核包含细胞核信息分子;

[0020] (b) 使邻接保持元件与核通透性增强剂和分析性生物分子接触;和

[0021] (c) 使分析性生物分子与细胞核信息分子反应以提供分析复合物;和

[0022] (d) 分析分析复合物,从而检测细胞核信息分子。

[0023] 本公开还涉及分析细胞核信息分子和细胞质信息分子的方法,其包括:

[0024] (a) 提供包含细胞的邻接保持元件,其中细胞包含细胞核信息分子和细胞质信息分子;

[0025] (b) 使邻接保持元件与核通透性增强剂、第一分析性生物分子和第二分析性生物分子接触;

[0026] (c) 使第一分析性生物分子与细胞核信息分子反应以提供第一分析复合物,并使第二分析性生物分子与细胞质信息分子反应以提供第二分析复合物;

[0027] (d) 分析第一分析复合物,从而检测细胞核信息分子;和,任选地,

[0028] (e) 分析第二分析复合物,从而检测细胞质信息分子。

[0029] 在上述方法中,分析性生物分子可以是转座体复合物。因此,本文描述了使用转座体复合物和核通透性增强剂在细胞核中对靶核酸进行片段化和标签化的方法。用核通透性增强剂处理细胞或细胞核允许将转座体复合物有效递送至细胞核空间。本公开涉及通过用细胞核通透性增强剂处理细胞或细胞核而将转座酶或转座体复合物递送到细胞核中的方法。以这种方式,核遗传物质的转座可以在细胞核内生效。

[0030] 本文描述了从细胞核靶核酸制备标签化(tagged)核酸片段的文库的方法,其包括:

[0031] (a) 使包含靶核酸的细胞核与核通透性增强剂和多个转座体复合物接触,其中各转座体复合物包含转座酶和两个包含转座子末端序列的转座子末端组合物;

[0032] (b) 使靶核酸与多个转座体复合物反应,从而将靶核酸片段化成双链核酸片段,并用来自转座子末端组合物的转移链标签化以形成标签化分析复合物;和

[0033] (c) 分析标签化分析复合物,从而检测标签化核酸片段。

[0034] 本文还描述了从细胞核靶核酸制备标签化核酸片段的文库的方法:

[0035] (a) 提供包含单个核的邻接保持元件,其中所述单个核包含靶核酸;

[0036] (b) 使所述邻接保持元件和单个核与核通透性增强剂和分析性生物分子接触;和

[0037] (c) 使所述分析性生物分子与所述细胞核信息分子反应以提供分析复合物;和

[0038] (d) 分析所述分析复合物,从而检测所述细胞核信息分子。

[0039] 在一些方法中,邻接保持元件包含具有靶核酸的细胞核(分离的或在细胞内的)。所述方法还可以包括如前文所讨论的制备和分析细胞质信息分子。

[0040] 在一些实施方式中,对细胞、细胞质和核细胞组分有区别地接近允许对来自一个或多个细胞或一种或多种细胞质和核细胞组分的RNA、DNA、蛋白质或其任何组合的分析。

[0041] 在一些方法中,化合物或生物分子(例如核孔阻断剂)被用来阻止接近某些细胞质和细胞核区室。

[0042] 本文描述的方法在来自相同的单个细胞的多分析物测定(DNA、RNA、蛋白质等)中是有用的。

[0043] 本文还描述了包含细胞和/或细胞核、核通透性增强剂、和转座酶或转座体复合物的组合物。

附图说明

[0044] 图1A和1B显示细胞(图1A)和细胞核(图1B)的结构。细胞质是蛋白质、RNA、线粒体DNA和其它生物分子的汇合。细胞的细胞核是DNA、RNA、核蛋白和其他生物分子的汇合。

[0045] 图2A、2B、和2C显示接近细胞和核组分不同模式。图2A显示通过阻断核孔而单独接近细胞的细胞质组分。图2B显示通过使用细胞通透性生物分子增强细胞膜的通透性,和/或增强细胞核膜的通透性,而接近细胞核。图2C显示通过使用核通透性增强剂,个体地维持作为邻接元件的分离的细胞核。

[0046] 图3显示如实施例3所述在不同浓度的Pitstop-2的存在下用FAM标记的转座体复合物处理的分离的细胞核的可视化。

[0047] 图4A、4B、和4C显示如实施例5所述用或不用Pitstop-2处理产生的ATAC-seq文库的测序结果。图4A显示Chr 12的一部分上的ATAC-seq谱。图4B报告全基因组测序覆盖的多样性和独特性结果。图4C显示人类基因组的启动子和编码区上的标准化覆盖。

[0048] 图5显示在存在和不存在Pitstop-2的情况下制备的转座DNA文库的琼脂糖凝胶图像。

[0049] 图6显示如实施例7所述在用链置换聚合酶处理后用未纯化的、转座片段化的核酸产生的ATAC-seq文库的琼脂糖凝胶图像。

[0050] 图7A和7B显示实施例8所述的结果。图7A报告基于BioAnalyzer数据的ATAC-seq文库的浓度。图7B显示Chr 6的一部分上的两个文库的ATAC-seq谱。

[0051] 图8A和8B描述如实施例11所讨论的索引引物的产生。图8A是具有与索引PCR引物互补的区域的珠子结合的接种寡核苷酸的示意性表示。图8B报告了使用接种寡核苷酸方法获得的扩增效率。

[0052] 图9显示如实施例13所述采用Bst和Phi29链置换聚合酶的实验的琼脂糖凝胶数据。

[0053] 图10是使用细胞核膜作为CE用于转座以及将转座的细胞核暴露于去离子水或高盐条件的进一步方法的示意图。

[0054] 图11显示在细胞裂解和细胞核易位过程中加入0 μ M、30 μ M、和60 μ M Pitstop-2制备的小鼠 (3T3) ATAC-seq文库的单细胞测序数据。图11A显示Pitstop-2浓度对按不同数量(>1k、>10k和>100k)的独特读取结果分组的三组条形码的独特读取结果的总数的影响。图11B显示含有0 μ M、30 μ M、或60 μ M Pitstop-2的ATAC seq样品在转录起始位点的1kb内的片段的百分比。图11C显示用0 μ M、30 μ M、和60 μ M Pitstop-2处理的样品的通过MACS-2峰发现程序定义的ATAC-seq峰重叠。

具体实施方式

[0055] 在一些实施方式中,细胞核信息分子是DNA、RNA、或蛋白质。在一些实施方式中,它是DNA或mRNA。在一些方面,它是DNA。在一些实施方式中,它是RNA。在一些实施方式中,它是DNA和RNA。在一些实施方式中,它是cDNA或cDNA和DNA。在一些实施方式中,DNA可代表DNA的开放染色质状态。在一些实施方式中,DNA可以呈现全基因组DNA或全基因组DNA的一部分或线粒体DNA。

[0056] 在一些实施方式中,细胞质信息分子是mRNA、DNA、或蛋白质。在一些实施方式中,它是mRNA或DNA。

[0057] 在一些实施方式中,分析性生物分子是转座酶或转座体复合物、或抗体、或寡核苷酸、或核苷酸、或逆转录引物、或酶。在一些实例中,寡核苷酸或核苷酸包含至少一种标记的核苷酸。在一些实施例中,酶是扩增酶、聚合酶、DNA聚合酶、连接酶、RNA聚合酶、PCR酶、Taq DNA聚合酶、Pfu DNA聚合酶、介导体外转录的酶、整合酶、或切口酶。

[0058] 在一些实施方式中,分析性生物分子被索引或条形码化。在一些实施方式中,分析性生物分子是转座酶。在其他实施方式中,它是转座体复合物。在一些实施方式中,转座体复合物包含转座酶和两个包含转座子末端序列的转座子末端组合物。合适的转座方法是本领域已知的,并且描述于例如美国公开号2010/0120098和2014/0194324,以及PCT公开号

W02016/130704。

[0059] 在一些实施方式中,转座子末端组合物包含末端序列和寡核苷酸接头。在一些实施方式中,特定转座子末端序列包含结合转座酶的识别位点的双链区。

[0060] 在一些实施方式中,分析性生物分子是抗体,特别是特异性结合靶信息分子的抗体,所述靶信息分子是蛋白质、蛋白质片段、或肽。任选地,抗体还包含允许后续分离和/或分析的侧链寡核苷酸或其他标签。抗体与靶蛋白的结合产生分析复合物。

[0061] “分析复合物”是分析性生物分子与靶信息分子的反应的产物。这样的复合物可以通过共价或非共价结合、或共价结合和非共价结合两者而形成。例如,转座酶复合物可以用转移链来片段化和标签化靶核酸,产生标签化核酸、转座酶和非转移链的复合物。得到的分析复合物通过共价键和杂交形成。在另一个实例中,分析复合物可以是抗体-靶蛋白或抗体-靶肽复合物。

[0062] 方法和组合物被描述以改善转座体对细胞核中的DNA的可及性(accessibility)。目前的应用不限于转座体,因为各种其他分子可以选择性地插入细胞核或移除。例如,可将聚T捕获探针带入细胞核用于RNA捕获和下游测定(例如,测序)和分析。或者,可以靶向其他孔例如NPC以将分子输入和输出细胞核。

[0063] 细胞质和细胞核区室的实例是,但不限于,线粒体、细胞核、叶绿体、过氧化物酶体、内质网、微管、高尔基体、羧酶体(carboxysome),和代谢体(metabolosome)。

[0064] 核通透性增强剂

[0065] 在一些方面,核通透性增强剂是增加细胞核被膜的通透性而不破坏细胞核膜的化合物(例如,小分子或肽)。在这样的实施方式中,细胞核膜未被除去,但变得更多孔。这样的处理改善了分析性生物分子(例如转座体复合物或抗体)对核遗传物质的接近。因此,在一些实施方式中,通过使细胞核与核通透性增强剂接触而增加分析性生物分子(例如转座体复合物)对核遗传物质的接近。合适的增强剂包括破坏NPC疏水相互作用的化合物、结合和/或抑制核丝蛋白(例如网格蛋白)的化合物,和核定位信号肽。

[0066] 在一些实施方式中,增强剂是网格蛋白抑制剂。已显示网格蛋白外壳组装的抑制剂抑制网格蛋白介导的内吞作用的经典底物的摄取。(参见Liashkovich,I.等,“Clathrin inhibitor Pitstop-2 disrupts the nuclear pore complex permeability barrier,”*Sci.Rep.*2015,5,9994。)这样的化合物可以发挥功能以创建通腔(cavities through)和/或增加核孔复合物的尺寸。合适的网格蛋白抑制剂包括Pitstop-2(也称为N-[5-(4-溴亚苄基)-4-氧代-4,5-二氢-1,3-硫杂唑-2-基]萘-1-磺酰胺)、甲基-β-环糊精、吩噻嗪、单丹磺酰尸胺、氯喹,莫能菌素、高渗蔗糖,和Dynasore,及其合成类似物。(参见Chen,C.-L.等,“Inhibitors of clathrin-dependent endocytosis enhance TGFβ signaling and responses,”*J.Cell.Sci.*2009,122,1863-1871,和其中引用的文献。)如在本文中呈现的数据所示,网格蛋白抑制剂已经被发现实现转座体复合物转运到细胞核中,从而改善转座体复合物对细胞核遗传物质的接近和改善所述物质的易位效率。

[0067] 示例性增强剂还包括疏水性破坏剂例如脂肪族醇,包括C₄₋₁₀-烷基二醇,环状二醇,环烷二醇,或邻二醇(例如反式-1,2-环己二醇、正己烷-1,2-二醇、或1,6-己二醇;参见Ribbeck,K.等,“The permeability barrier of nuclear pore complexes appears to operate via hydrophobic exclusion,”*The EMBO J.*2002,21(11),2664-2671)。在一些实

施方式中,核通透性增强剂是环己二醇、或是1,2-环己二醇,或是反式-1,2-环己二醇。另外的疏水性破坏剂包括表面活性剂如洋地黄皂苷(参见,例如Hagstrom等,J.Cell Sci.1997,110,2323-2331;Tissera等,“Nuclear envelopes show cell-type specific sensitivity for permeabilization with digitonin,”Nature (Protocol Exchange), 2010(可在<https://www.nature.com/protocolexchange/protocols/1994>获得)。

[0068] 示例性增强剂还包括核定位信号(NLS),其是通常用于标签化蛋白质用于通过细胞核转运输入细胞核的氨基酸序列。在一些实施方式中,NLS与分析性生物分子共价或非共价连接(例如,作为复合物施用,或原位形成复合物)。在一些方法中,NLS不与分析性生物分子(例如,转座酶或转座体复合物)共价结合,而是用作分析性生物分子与细胞或细胞核的混合物的添加剂(例如,转座反应)。在一些实施方式中,NLS是SV40大T抗原(PKKRKRV; Creative Peptides, Shirley, NY, Cat#GR1405),核质蛋白的NLS(KR[PAATKKAGQA]KKKK或AVKRPAATKKAGQAKKKLD)(参见Rotello等,Bioconj.Chem.26(6),1004-7),K-K/R-X-K/R(参见Chelsky等,Mol.Cell Biol.1989,9(6),2487-2492;Dingwall等,J.Cell.Biol.1988,107(3),841),EGL-13(MSRRRKANPTKLSNAKKLAKEVEN),c-Myc(PAAKRVKLD),TUS蛋白(KLKI KRPVK),hnRNP A1的酸性M9结构域、来自酵母转录阻遏物Mat α 2的KIPIK, PY-NLS序列(Lee等,Cell 126(3),543-58)和输入蛋白 β 2的抑制剂。在一些实施方式中,NLS是SV40大T抗原。

[0069] 模板核酸的制备

[0070] 一些实施方式包括制备模板核酸的方法。如本文所用,“模板核酸”可以指用于获得序列信息的底物。制备模板核酸的一些方法包括将转座子序列插入靶核酸中,从而制备模板核酸。一些插入方法包括在足以使一个或多个转座子序列整合到靶核酸中的条件下,在酶(例如转座酶或整合酶)存在下使本文提供的转座子序列与靶核酸接触。在一些实施方式中,模板核酸可包括包含至少一个转座子序列、其片段或其任何拷贝的靶核酸、其片段或其任何拷贝。在一些实施方式中,模板核酸可包括含有接头的核酸,所述接头包含适于测序的标签,例如引物位点。

[0071] 在一些实施方式中,细胞可以被固定。在一些实施方式中,本文所述的方法使用完整细胞,并通过用表面活性剂处理来固定细胞。在这样的方法中,分析性生物分子可以进入细胞而不破坏细胞膜。在示例性方法中,用表面活性剂(例如,NP-40、SDS等)、核通透性增强剂和分析性生物分子处理细胞。

[0072] 在一些实施方式中,分析性生物分子被设计用于特定的靶DNA、mRNA、cDNA、DNA,或其任何组合(例如,DNA和cDNA)。例如,某些分析性生物分子可包括设计用于区分DNA和mRNA的索引探针。在一些实施方式中,使用第一分析性分子和第二分析性生物分子,例如第一转座体复合物和第二转座体复合物,其中第一转座体复合物包含用于标签化DNA的转座子末端序列(例如,A14/B15引物序列),并且第二转座体复合物包含用于标签化mRNA的转座子末端序列(例如,包含含有对mRNA聚A尾特异的聚T区的接头)。

[0073] 在一些实施方式中,人工孔可以插入到细胞或细胞核区室中以促进分析性生物分子转移进或出区室。

[0074] 在一些实施方式中,方法包括分离细胞核。细胞核的分离可以使用本领域已知的标准方法完成,条件是该方法保留细胞核的完整性。在一些情况下,可以使用细胞裂解方

法。在暴露于核通透性增强剂之前,可以从细胞质组分中分离或纯化细胞核。在一些实施方式中,维持细胞质级分并单独分析。

[0075] 对于核酸分析,样品制备通常包括将基因组核酸片段化成可测序的长度,并将接头连接到片段以得到用于后续纯化,和扩增(如果需要的话)的模板。使用引物将模板转化为“接种”模板,将其扩增并进行测序方案。

[0076] 在溶液中将DNA和/或RNA转化成接头修饰的模板准备用于扩增和测序所需的步骤的数目可以通过使用转座酶介导的片段化和标签化来最小化。该过程在本文中称为“转座片段化”,通常涉及通过转座体复合物修饰DNA或mRNA,所述转座体复合物包含与包含转座子末端序列的接头复合的转座酶。转座片段化导致DNA的同时片段化和接头连接至双链体片段的两条链的5'端。纯化步骤可用于除去转座酶。可以填充双链产物中的任何缺口,并且可以通过PCR将另外的序列添加到接上接头的片段的末端。

[0077] 在一些实施方式中,本文提供的多个转座子序列插入到靶核酸中。一些实施方式包括选择足以实现多个转座子序列整合到靶核酸中的条件,使得各个整合的转座子序列之间的平均距离包含靶核酸中的一些数量的连续核苷酸。

[0078] 制备模板核酸的一些实施方式可以包括复制包含靶核酸的序列。例如,一些实施方式包括将引物与整合到靶核酸中的转座子序列的引物位点杂交。在一些这样的实施方式中,引物可以与引物位点杂交并延伸。复制的序列可包含至少一个条形码序列和靶核酸的至少一部分。在一些实施方式中,复制的序列可包含第一条形码序列、第二条形码序列和置于其间的靶核酸的至少一部分。在一些实施方式中,至少一个复制的核酸可以包含第一复制核酸的至少第一条形码序列,其可以被鉴定或指定为与第二复制核酸的第二条形码序列配对。在一些实施方式中,引物可包含测序引物。在一些实施方式中,使用测序引物获得测序数据。在更多的实施方式中,可以将包含引物位点的接头连接到核酸的各个末端,并从这样的引物位点扩增核酸。

[0079] 制备模板核酸的一些实施方式可以包括扩增包含一个或多个转座子序列的至少一部分和靶核酸的至少一部分的序列。在一些实施方式中,可以使用与整合到靶核酸中的整合转座子序列的引物位点杂交的引物扩增靶核酸的至少一部分。在一些这样的实施方式中,扩增的核酸可包含第一条形码序列和第二条形码序列,具有置于其间的靶核酸的至少一部分。在一些实施方式中,至少一种扩增的核酸可以包含第一扩增核酸的至少第一条形码序列,其可以被鉴定为与第二扩增序列的第二条形码序列配对。

[0080] 在一些实施方式中,各模板核酸包含至少一个通用引物位点是有利的。例如,模板核酸可包含含有第一通用引物位点的第一末端序列和含有第二通用引物位点的第二末端序列。通用引物位点可具有各种应用,例如用于扩增、测序和/或鉴定一种或多种模板核酸。第一和第二通用引物位点可以相同、基本上相似、相似或不同。可以通过本领域公知的各种方法将通用引物位点引入核酸中,例如,将引物位点连接至核酸、使用加尾引物扩增核酸、以及插入包含通用引物位点的转座子序列。

[0081] 转座体

[0082] 如本文所用,术语“转座体复合物”通常是指转座酶(transposition enzyme)(例如,整合酶或转座酶(transposase))和包含整合识别位点,例如转座酶识别位点的双链核酸。在本文提供的实施方式中,转座酶可以与转座酶识别位点形成功能复合物,其能够催化

转座反应。转座酶可以与转座酶识别位点结合并通过“转座片段化”将转座酶识别位点插入到靶核酸中。在一些这样的插入事件中,转座酶识别位点的一条链可以转移到靶核酸中。在一个实例中,转座体包含含有两个亚基和两个非邻接转座子序列的二聚体转座酶。在另一个实例中,转座酶包含含有两个亚基,和邻接转座子序列的二聚体转座酶。在一些实施方式中,通过在支持非共价复合物形成的条件下将转座酶与双链转座子DNA一起孵育来形成复合物。

[0083] 双链转座子DNA可包括,但不限于,Tn5DNA、Tn5DNA的一部分、转座子末端组合物、转座子末端组合物的混合物、或能够与转座酶例如超活性(hyperactive) Tn5转座酶相互作用的其他双链DNA。

[0084] “转座酶”是指能够与含转座子末端的组合物(例如,转座子、转座子末端、转座子末端组合物)形成功能性复合体,并催化含转座子末端的组合物插入或转座到与其一起孵育的双链靶DNA中,例如,在体外转座反应中。本文呈现的转座酶还可包括来自逆转录转座子和逆转录病毒的整合酶。

[0085] 转座酶、转座体和转座体复合物对本领域技术人员通常是已知的,如美国公开号2010/0120098和PCT公开号2016/130704的公开内容所示例的,其通过引用并入本文。转座酶包括但不限于Tn5转座酶、Mu转座酶和哈氏弧菌转座酶,及其变体,例如超活性Tn5转座酶。转座酶可以是多聚体,例如二聚体、三聚体或四聚体,例如Tn5二聚体。能够出于其预期目的以足够的效率插入转座子末端以对靶DNA进行标签化和片段化的任何转座系统可用于本文所述的方法中。在特定的实施方式中,优选的转座系统能够以随机或几乎随机的方式插入转座子末端以使靶DNA 5'-标签化和片段化。在特定的实施方式中,是超活性Tn5转座酶和Tn5型转座酶识别位点(Goryshin和Reznikoff, J. Biol. Chem. 1983, 273:7367),或MuA转座酶和包含R1和R2末端序列的Mu转座酶识别位点(Mizuuchi, K., Cell 1983, 35:785; Savilahti, H. et al., EMBO J. 1995, 14:4893)。镶嵌末端(Mosaic end, ME)序列也可以由技术人员优化使用。上述参考文献通过引用并入本文。W02016/130704中描述了另外的示例性转座系统,例如金黄色葡萄球菌Tn552、Ty1、转座子Tn7、Tn/0和IS10, Mariner转座酶、Tc1、P元件、Tn3、细菌插入序列、逆转录病毒、逆转录病毒整合酶(例如来自HIV-1、HIV-2、SIV、PFV-1和RSV的整合酶)、酵母的逆转录转座子、IS5、Tn10、Tn903或IS911,及其工程化变体。

[0086] 在一些实施方式中,转座酶是Tn5转座酶, Mu转座酶或弧菌(例如哈氏弧菌)转座酶。在一些实施方式中,转座酶是Tn5转座酶。在一些实施方式中,转座酶是超活性Tn5转座酶。在一些实施方式中,转座酶是二聚体。在一些实施方式中,转座酶是Tn5二聚体。在一些实施方式中, Tn5二聚体是超活性的。

[0087] 转座子序列

[0088] 术语“转座子末端”是指包含至少一个转座识别位点的双链核酸DNA(“转座子末端序列”),其与转座酶或整合酶形成复合物以提供用于体外转座反应的转座体复合物。可用于本文提供的方法和组合物的转座子序列在美国公开号2012/0208705和2012/0208724,和PCT公开号W02012/061832中提供,其全部内容各自通过引用并入。在一些实施方式中,转座子末端能够在转座反应中与转座酶形成功能性复合物。作为非限制性实例,转座子末端可包含野生型或突变体Tn5转座酶识别的19-bp外端(“OE”)转座子末端、内端(“IE”)转座子末端或“镶嵌末端”(“ME”)转座子末端,或如美国公开号2010/0120098中描述的R1和R2转座子

末端。转座子末端可包含适于在体外转座反应中与转座酶或整合酶形成功能性复合物的任何核酸或核酸类似物。例如,转座子末端可包含DNA、RNA、修饰碱基、非天然碱基和/或修饰骨架,并且可以在一条或两条链中包含切口。尽管在整个本公开中术语“DNA”用于与转座子末端的组合物有关,但应理解,任何合适的核酸或核酸类似物可用于转座子末端。

[0089] “转移链(transferred strand)”是指转座子末端两者的被转移的部分。类似地,术语“非转移链(non-transferred strand)”是指“转座子末端”两者的非转移部分。转移链的3'末端在体外转座反应中连接或转移至靶DNA。表现出与转移的转座子末端序列互补的转座子末端序列的非转移链在体外转座反应中不连接或转移至靶DNA。

[0090] 如本文所用的术语“接头(adaptor)”是指包含至少一个标签的多核苷酸区域。一些实施方式包括转座体复合物,其包含具有包含转座子末端序列的3'部分的多核苷酸,和包含至少一个标签的接头区域。接头序列可包括连接子(linker)区域。

[0091] 如本文所用的术语“标签”是指具有适合用于期望预期目的或应用的序列的寡核苷酸区域。标签可包含在插入靶核酸时有用的一种或多种序列,例如片段化位点(可在确定的时间化学地、生物化学地或光化学地切割的序列)、引物位点、条形码(用于鉴定一种或多种特定分析物)、亲和标签、识别位点和/或报告部分(可以发射信号,例如荧光、化学发光、生物发光、磷光、放射性、量热,电子或其他信号的部分)。应当理解,可以将任何其他合适的特征引入标签中。在一些实施方式中,标签包含长度为5至200bp,或10至100bp,或20至50bp,或长度为约5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、150、或200bp的序列。

[0092] 在一些实施方式中,标签包含适合于与引物(其可以被结合至固体表面例如珠子或流动细胞)杂交用于扩增反应例如簇扩增和/或序列反应的一个或多个引物位点。引物结合位点的示例性序列包括但不限于:

[0093] AATGATACGGCGACCACCGAGATCTACAC (P5序列)

[0094] 和

[0095] CAAGCAGAAGACGGCATACGAGAT (P7序列)

[0096] 及其互补体。在一些情况下,引物序列包含修饰碱基以允许后续酶促或化学切割,例如邻二醇或8-氧代鸟嘌呤。

[0097] 在一些实施方式中,标签包含在模板核酸的制备中使用的条形码。如将被理解的,大量可用条形码允许各模板核酸分子包含独特的标识。模板核酸的混合物中各分子的独特标识可用于若干应用中。例如,独特标识的分子可用于在具有多个染色体的样品中、在基因组中、在细胞中、在细胞类型中、在细胞疾病状态中、和在物种中,例如在单倍型测序中、在亲本等位基因区分中、在宏基因组测序中,和在基因组的样品测序中鉴定个体核酸分子。示例性条形码序列包括但不限于TATAGCCT、ATAGAGGC、CCTATCCT、GGCTCTGA、AGGCGAAG、TAATCTTA、CAGGACGT和GTACTGAC。

[0098] 制备转座子末端序列、转座子、和转座体复合物的方法在本领域中是已知的。

[0099] 如本文所用,术语“核酸”和“寡核苷酸”是指共价连接在一起的至少两个核苷酸单体。核酸通常含有磷酸二酯键和天然碱基,但在一些实施方式中可包含非天然骨架和/或碱基,如例如PCT公开号W02016/130704中所描述的。核酸可以是DNA,例如基因组或cDNA、RNA或杂合体,其来自单个细胞、单个细胞核、多个细胞、多个细胞核或来自多个物种,如同宏基因组样品,例如来自环境样品,进一步来自混合样品,例如混合组织样品,或相同物种的不

同个体的混合样品,疾病样品例如癌症相关核酸,等等。核酸可含有脱氧核糖核苷酸和核糖核苷酸的任何组合,以及碱基的任何组合,包括天然碱基、或非天然的、类似的或合成的碱基。

[0100] 如本文中所使用的,术语“检测”和/或其语法等同形式可以是指鉴定分析物的存在 (presence) 或存在 (existence), 鉴定分析物的个体组分,例如序列信息、和/或定量这样的分析物的量。在一些实施方式中,检测包括检测通过本文描述的方法产生的标签化信息分子或扩增子的序列。

[0101] 在一些实施方式中,靶信息分子、细胞、和/或细胞核可以获自包含DNA和/或mRNA的任何生物样品,包括但不限于生物样品或患者样品。细胞和/或细胞核的来源可以是例如细菌、植物、寄生虫、昆虫、动物或哺乳动物(例如,大鼠、小鼠、猴子、非人灵长类动物、人)或其混合物。如本文所用的术语“生物样品”或“患者样品”包括样品,例如一种或多种细胞、组织或体液。样品可包括天然或加工的样品,例如浸软的组织或裂解物。“体液”可包括但不限于血液、血清、血浆、唾液、脑脊髓液、支气管抽吸物、胸膜液、法氏囊液 (bursa fluid)、滑液、泪液、乳管液、淋巴液、粘液、痰液、尿液、粪便、羊水或精液、或其混合物。组织可包括活组织检查样品、肿瘤样品或皮肤。样品可包括包含少于约1% (w/w) 的全细胞物质的体液,例如血浆或血清。样品可包括天然或合成来源的样品(即制成无细胞的细胞样品)。

[0102] 在一些实施方式中,方法还包括用聚合酶例如链置换聚合酶(例如,Bst聚合酶或phi29聚合酶)处理转座片段化核酸以从转座片段化材料去除转座酶。在这样的实施方式中,细胞核内的转座片段化核酸可以用聚合酶(例如,链置换聚合酶)和十二烷基硫酸钠处理以允许PCR扩增。在其他实施方式中,标签化信息分子可以用聚合酶(例如链置换聚合酶)处理以允许PCR扩增。

[0103] 在一些实施方式中,方法还包括用去离子水接触转座细胞核。因此,方法还可以包括通过用去离子水处理细胞或细胞核而从经处理的细胞或细胞核去除分析复合物,例如标签化核酸或抗体-蛋白质/肽复合物。这样的处理造成水的渗透性流入,从而使核体膨胀。在一些实施方式中,用核通透性增强剂处理细胞或细胞核包括用去离子水处理。这样的膨胀用于进一步增加核孔径以允许转座体复合物进一步流入核空间。(见图10)

[0104] 在一些实施方式中,用核通透性增强剂进行处理是在高盐缓冲液的存在下发生。缓冲液在核膜上产生渗透差异,使得核物质被排出核体。这样的处理将允许对核膜的破坏的时间控制(见图10)。

[0105] 在不使用通透性增强剂处理细胞核的情况下,少量的转座体复合物确实可以接近细胞核物质。因此,在一些实施方式中,方法包括在存在核孔阻断剂和不存在核通透性增强剂的情况下使细胞质信息分子与第一分析性生物分子反应。在随后的步骤中,任选地去除核孔阻断剂和/或细胞质物质,并且在存在核通透性增强剂的情况下用第二分析性生物分子处理剩余的细胞物质。以这种方式,细胞质和细胞核组分被正交标签化。合适的核孔阻断剂包括试剂,例如小麦胚芽凝集素、细霉素 (leptomycin) B、特异于NPC的抗体(参见,例如,Adam, S.A.等, J. Cell Biol. 1990, 111 (3), 807-816; Moore等, Cell 1992, 69 (6), 939-950), 或与NPC或门控离子通道结合的其他配体、酶或生物分子。

[0106] 邻接保持元件

[0107] 在一些实施方式中,邻接保持元件 (CE) 被用于保持单个细胞和/或单个细胞核信

息。在一些实施方式中,通过个体细胞核或个体细胞的区室化来保持单个细胞核的邻接信息。在一些实施方式中,通过将单个细胞核或细胞定位在物理区室(例如微孔、微液滴或微囊)中,或通过使用水凝胶固定遗传物质或在微珠上固定遗传物质或在微珠内固定遗传物质来实现这样的区室化。这样的过程在转座期间有效地维持细胞和/或细胞核邻接,同时给出更直接地接近细胞核物质。这样的方法还减少了细胞核材料与单独的细胞核的交叉污染。在一些实施方式中,CE可以包含靶核酸。

[0108] 在一些实施方式中,方法包括对保持、嵌入、或包含在邻接保持元件内的核酸进行测序。特别地,本文提供的方法和组合物的实施方式涉及制备核酸模板并从中获得序列数据。本文提供的方法和组合物涉及美国公开号2012/0208705和2012/0208724以及PCT公开号W02012/061832中提供的方法和组合物,其全部内容各自通过引用并入。本文提供的一些实施方式涉及制备在邻接保持元件内的DNA以从靶核酸获得定相(phasing)和序列装配信息,并从这样的模板获得定相和序列装配序列信息。本文提供的具体实施方式涉及整合酶(例如转座酶)用于维持片段化核酸的相关末端的物理临近的用途;以及组合索引用于从各邻接保持元件创建个体库的用途。获得单倍型信息包括区分靶核酸中的不同等位基因(例如,SNP、遗传异常等)。这样的方法可用于表征靶核酸中的不同等位基因,并降低序列信息中的错误率。

[0109] 在一些实施方式中,CE包含细胞或单个细胞。在一些实施方式中,CE包含来自细胞或来自单个细胞的核酸,例如DNA、mRNA或cDNA;细胞或单个细胞的大分子,包括蛋白质、多糖、脂质和核酸,以及来自细胞或来自单个细胞的小分子,例如初级代谢产物、次级代谢产物和天然产物。在一些实施方式中,核酸在形成包含核酸的CE之前经历扩增,例如PCR或全基因组扩增。在一些实施方式中,DNA和mRNA的分析可以并行进行。在一些实施方式中,细胞膜是CE。

[0110] 在一些实施方式中,CE包含细胞核或单个细胞核。在一些实施方式中,CE包含来自细胞核或来自单个细胞核的核酸。在一些实施方式中,核膜是CE。

[0111] 如图2所示,本公开考虑接近维持邻接信息的细胞和细胞核区室的各种模式。在一些实施方式中,细胞是CE元件,其中细胞核(例如核孔复合物)被阻断允许细胞核信息分子与分析性生物分子的相互作用所需的分析性生物分子或试剂进入(图2A)。因此,细胞核膜是允许与细胞质内容物相对于细胞核内容物的有差别的反应的CE。在这样的情况下,细胞膜可以用作第二CE,或者细胞可以在另外的CE中区室化。在一些实施方式中,通过增强细胞膜的通透性、使用细胞可通透生物分子、和/或增强细胞核膜的通透性来接近细胞核内容物(图2B)。在其他实施方式中,当用核通透性增强剂处理时,分离的细胞核的细胞核膜可以用作CE(图2C)。所有的这些方法可以与另外的CE结合使用,例如微滴、微珠(水凝胶)、微孔,和其他种类的区室,以维持信息邻接。

[0112] 在一些实施方式中是从RNA、DNA、或其混合物制备文库和获得RNA、DNA、或RNA和DNA两者的单个细胞数据的方法。在一些实施方式中,是从RNA、DNA、蛋白质或其任何组合制备文库的方法。

[0113] 在一些实施方式中,细胞在CE内裂解使得单个细胞内的多个靶信息分子在CE内释放。在一些情况下,细胞被裂解但细胞核未被裂解,使得细胞质组分在CE内释放并且细胞核保持完整。

[0114] 在一些实施方式中,多个CE与组合标签化方法结合使用,使得核酸或其他信息分子被物理地划分和/或正交地标签化。

[0115] 在一个实施方式中,靶标可以被稀释到CE例如液滴中。可以使用任选的全基因组扩增,并且可以从相当于靶核酸的大约单倍体等同物的模板核酸的量获得序列信息。

[0116] 在一些实施方式中,除核酸之外,多组合标记化方案还可以被用于单个细胞内的组分,例如,蛋白质、细胞器、脂质、或细胞膜,使得单个细胞或细胞核内的组分可以从来自不同单个细胞或细胞核的组分鉴定。在一些实施方式中,CE可包含单个细胞内或单个细胞核内的组分。在一些实施方式中,CE内的单个细胞和/或单个细胞核的组分将具有不同于不同CE内的单个细胞和/或细胞核的组分的可识别的独特标记。

[0117] 在一些实施方式中,多组合条形码化方案可以被采用。在一些实施方式中,这样的组合条形码化和组合标记化可以在包含单个细胞的CE内进行。在一些实施方式中,可以对包含单个细胞或细胞核的多个CE平行进行这样的组合条形码化和组合标记化。

[0118] 邻接保持元件(CE)是物理实体,其通过一个或多个测定步骤保持至少两个、或更多个,或所有分析物处于紧密临近(proximity)(或邻接(contiguity)),并提供对测定试剂的接近(access),并且可以被合并和分开多次而不失去对分析物的临近。

[0119] 在一些实施方式中,CE可以是固体支持物。在一个实施方式中,CE可以是乳液或液滴。在一些实施方式中,CE是凝胶、水凝胶或凝胶珠。在一些实施方式中,CE可以包含固体支持物,例如珠子。在一些实施方式中,珠子可以进一步包含抗体、寡核苷酸和/或条形码。在另一个实施方式中,CE可以构成由WGA、RCA或任何核酸试剂的缩合产生的DNA纳米球。

[0120] 在一些实施方式中,CE可以通过将来自细胞或来自单个细胞的核酸,或其扩增产物(来自WGA等),嵌入到聚合物基质中例如琼脂糖、聚丙烯酰胺、藻酸盐等来制备。在一些实施方式中,CE内细胞或单个细胞的内容物的邻接性通过借助包封(例如在聚合物基质中)、固定在珠子上或截留而保持组分彼此的物理临近,通过多轮合并和再分配有效保持CE内的邻接信息。CE的收集可以独立地合并和分开,与测定试剂反应,再次合并和分开,然而保持构成个体CE的分析物的邻接性的特征,使得能够通过不同的分开和合并步骤进行组合索引。

[0121] 在一些实施方式中,在邻接保持元件中的分析物可接近测定试剂,包括水溶液、酶(例如,片段化酶、聚合酶、连接酶、转座酶、激酶、限制性内切酶、蛋白酶、磷酸酶、或脂肪酶)、核酸接头、核酸条形码和/或标签。

[0122] 在一些实施方式中,CE的一种或多种分析物标记有一种或多种标记。示例性标记包括但不限于DNA条形码或索引、荧光标记、化学发光标记、RNA条形码或索引、放射性标记、包含标记的抗体,包含标记的珠子。

[0123] 在一些实施方式中,方法可以包括以下步骤:(a)将包含靶核酸的CE区室化为多个第一容器;(b)向各第一容器的靶核酸提供第一索引,从而获得第一索引核酸;(c)组合第一索引核酸;(d)将第一索引模板核酸区室化为多个第二容器;(e)向各第二容器的第一索引模板核酸提供第二索引,从而获得第二索引核酸。步骤a-e可以用a-e系列的一个或多个步骤的另外循环继续,以得到另外的虚拟区室。这种组合索引的方法可用于从有限数量的物理区室有效地创建大量虚拟区室。

[0124] 在一些实施方式中,方法可包括以下步骤:(a)提供包含具有附着的核酸报告子的

非核酸分析物(例如蛋白质)的CE; (b) 将CE区室化为多个第一容器; (c) 向各第一容器的靶核酸报告子提供第一索引, 从而获得第一索引靶核酸报告子; (d) 组合第一索引核酸报告子; (e) 将第一索引CE区室化为多个第二容器; (f) 向各第二容器的第一索引核酸报告子提供第二索引, 从而获得第二索引核酸报告子。步骤a-f可以用a-f系列的一个或多个步骤的另外循环继续, 以得到另外的虚拟区室。区室化步骤还可以包括核酸扩增或捕获步骤, 例如PLA、PEA、或捕获或扩增核酸的其他技术。

[0125] 在一些实施方式中, 一个或多个核酸可以嵌入在将核酸限制在限定空间但允许试剂接近以进行以下步骤的基质中, 包括但不限于, 扩增(PCR、全基因组扩增、随机引物延伸等)、连接、转座、杂交、限制性消化和DNA诱变。诱变的实例包括但不限于易错延伸、烷基化、亚硫酸氢盐转化和活化诱导的(胞苷)脱氨酶等。

[0126] 在一些实施方式中, CE中的感兴趣的分析物是蛋白质。蛋白质可以用条形码或替代标记进行标记。可以使用传统阵列或基于序列的方法读出条形码或标记。临近连接方法和抗体-索引序列可用于检测蛋白质(Fredriksson等, *Nature Biotechnology* 20, 473-477 (2002), 通过引用并入本文)以及条形码序列的检测以确定各个体细胞中蛋白质的同一性和丰度。蛋白质可以通过技术人员已知的各种方法(www.piercenet.com/cat/protein-antibody-labeling)进行标记, 包括体内和体外位点特异性化学标记策略。

[0127] 在一些实施方式中, 邻接保持元件可以包含单个细胞, 并且来自该细胞的核酸可以被扩增。随后, 可以通过组合索引方案独特地索引每个邻接保持元件。在一些实施方式中, 可以使用组合索引方法。例如, 通过使用片段化(酶促)和接头连接的标准文库制备技术, 或通过使用转座酶复合物的转座片段化, 将初始索引附着于基因组DNA或cDNA。随后的索引通过连接或PCR附着到文库。连接是优选的, 因为易于以顺序方式添加索引的接头。最后一步可以仅涉及索引PCR或连接和PCR。

[0128] 细胞生物分子的有区别地索引: 含有生物分子例如DNA、RNA、线粒体DNA、蛋白质等的细胞和亚细胞区室, 可以通过保持细胞邻接性而被特异地和正交地索引用于基因组测定。这可以通过亚细胞区室(细胞膜、细胞核膜和线粒体膜)的膜蛋白的有区别地靶向来实现, 以允许选择性地将离子或其他蛋白质转运到某些区室中但不转移到其他区室中。例如, 核膜蛋白或线粒体蛋白可以用阻断剂分子阻断, 例如小麦胚芽凝集素、细霉菌B和对NPC特异的抗体, 而环孢菌素A、寡霉素和其他化合物阻断线粒体膜蛋白。该方法允许在细胞的亚细胞区室内索引遗传物质(不限于转座、剪接-连接等), 从而提供细胞内特异性靶向的更好、精确的测序读取结果。

[0129] 本文中在一些方面描述的是有区别地索引来自细胞的不同区室中的信息分子的方法。在一些方面, 这样的方法包括:

[0130] 将包含第一标签的第一分析性生物分子选择性地递送至选自细胞核、细胞质和线粒体的第一细胞区室, 其中所述第一细胞区室包含第一信息生物分子;

[0131] 使所述第一分析性生物分子与所述第一信息分子反应以提供第一标签化信息分子;

[0132] 将包含第二标签的第二分析性生物分子选择性地递送至选自细胞核、细胞质和线粒体的第二细胞区室, 其中所述第二细胞区室包含第二信息分子并且不同于所述第一细胞区室; 和

[0133] 使所述第二分析性生物分子与所述第二信息分子反应以提供第二标签化信息分子,其中所述第一标签和第二标签不同。

[0134] 在一些实施方式中是有区别地索引细胞质和细胞核信息分子的方法,包括:(a) 将包含第一标签的第一分析性生物分子递送至细胞的细胞质而基本上不将第一分析性生物分子递送至细胞的细胞核;(b) 使第一分析性生物分子与细胞质信息分子反应,从而将第一标签或其互补物添加到细胞质信息分子中以提供标签化细胞质信息分子;(c) 用核通透性增强剂和包含第二标签的第二分析性生物分子处理细胞,从而将第二分析性生物分子递送至细胞的细胞核;和(d) 使第二分析性生物分子与细胞核信息分子反应,从而将第二标签或其互补物添加到细胞核信息分子中,以提供标签化细胞核信息分子。在一些实施方式中,递送至细胞质还包括用核孔阻断剂和/或线粒体孔阻断剂处理细胞。在一些实施方式中,方法还包括裂解细胞以释放标签化分子。在一些实施方式中,第一和第二标签是正交的,使得它们在检测时是可区分的,或者可以在后续操纵步骤(例如扩增)中单独靶向。在一些实施方式中,细胞质信息分子是RNA。在一些实施方式中,细胞核信息分子是DNA或基因组DNA(gDNA)。在一些实施方式中,分析性生物分子是转座体。在一些实施方式中,方法还包括检测标签化分子或其扩增子。在一些实施方式中,方法还包括选择性扩增标签化分子(例如,第一标签化相对于第二标签化或相反)。

[0135] 在其它方面,方法包括:(a) 使包含第一标签的第一分析性生物分子与线粒体信息分子反应,以产生标签化线粒体信息分子;(b) 使包含第二标签的第二分析性生物分子与细胞核信息分子反应,以产生标签化细胞核信息分子。在一些实施方式中,方法还包括:通过用线粒体膜通透性增强剂和第一分析性生物分子处理细胞而将第一分析性生物分子递送至线粒体。在一些实施方式中,递送至线粒体还包括用核孔阻断剂处理细胞。在一些实施方式中,递送至线粒体不包括将第一分析性生物分子基本上递送至细胞的细胞核。在一些方面,线粒体信息分子是DNA。在一些方面,细胞核信息分子是DNA,或是基因组DNA(gDNA)。在一些实施方式中,第一和第二标签是正交的,使得它们在检测时是可区分的,或者可以在后续操纵步骤(例如扩增)中单独靶向。在一些实施方式中,分析性生物分子是转座体。在一些实施方式中,方法还包括检测标签化分子或其扩增子。

[0136] 在一些方面,将第一分析性生物分子选择性地递送至第一细胞区室包括用用于所述第一细胞区室的通透性增强剂处理细胞。在一些实施方式中,将第一分析性生物分子选择性地递送至第一细胞区室包括用用于所述第二细胞区室的通透性阻断剂处理细胞。在一些实施方式中,将第二分析性生物分子选择性地递送至第二细胞区室包括用用于所述第二细胞区室的通透性增强剂处理细胞。在一些实施方式中,选择性地递送第一分析性生物分子是在基本上没有将所述第一分析性生物分子递送至第二细胞区室的情况下发生。

[0137] 在一些实施方式中,第一细胞区室是细胞质,并且第一信息分子是细胞质信息分子,和(a) 第二细胞区室是细胞核,并且第二信息分子是细胞核信息分子或(b) 第二细胞区室是线粒体,并且第二信息分子是线粒体信息分子。在一些实施方式中,将第一分析性生物分子选择性地递送至细胞质是在核孔阻断剂和/或线粒体孔阻断剂的存在下完成。在一些实施方式中,递送第二分析性生物分子是在核通透性增强剂或线粒体通透性增强剂的存在下完成。

[0138] 如本文所使用的,“线粒体通透性增强剂”增强了线粒体对试剂例如酶的通透性。

这样的增强剂不裂解线粒体。示例性线粒体通透性增强剂包括增加通过线粒体孔的流动的试剂。实例包括无机多磷酸盐(参见,例如,Seidlmayer等,J.Gen.Physiol.2012,139(5), 321-331)。

[0139] 如本文所使用的,术语“选择性递送”或其语法变体是指试剂向一个细胞区室的递送相对于向另一个细胞区室的递送增强。选择性递送可以通过相对于未处理时的区室的通透性增强一个区室的通透性或阻断一个区室的通透性来实现。选择性递送不需要100%选择性,但可以包括增加来自所得文库中不同区室的标签化信息分子的相对分数。

[0140] 分析和测序

[0141] 本文提供的一些方法包括分析核酸的方法。这样的方法包括制备靶核酸的模板核酸的文库,从模板核酸的文库获得序列数据,和从这样的序列数据组装靶核酸的序列呈现。

[0142] 靶核酸和模板核酸可以使用本领域公知的各种方法针对感兴趣的特定序列富集。这样的方法的实例在PCT公开号W02012/108864中提供,其全部内容通过引用并入本文。在一些实施方式中,可以在制备模板文库的方法期间进一步富集核酸。例如,核酸可以在转座片段化之前、转座片段化之后和/或核酸扩增之后针对某些序列富集。

[0143] 本文中所描述的技术的一些实施方式包括分析模板核酸的方法。在这样的实施方式中,可以从模板核酸获得测序信息,并且该信息可以用于产生一种或多种靶核酸的序列呈现。

[0144] 在本文所述的测序方法的一些实施方式中,可以使用连接的读取策略。连接的读取策略可以包括鉴定连接至少两个测序读取结果的测序数据。例如,第一测序读取结果可以包含第一标志物,并且第二测序读取结果可以包含第二标志物。第一和第二标志物可以鉴定在靶核酸的序列呈现中相邻的来自各测序读取结果的测序数据。在本文所述的组合物和方法的一些实施方式中,标志物可以包含第一条形码序列和第二条形码序列,其中第一条形码序列可与第二条形码序列配对。在其他实施方式中,标志物可以包含第一宿主标签和第二宿主标签。在更多的实施方式中,标志物可以包含具有第一宿主标签的第一条形码序列和具有第二宿主标签的第二条形码序列。

[0145] 用于测序模板核酸的方法的示例性实施方式可以包括以下步骤:(a)使用杂交至第一引物位点的测序引物对第一条形码序列进行测序;和(b)使用杂交至第二引物的测序引物对第二条形码序列进行测序。结果是有助于将模板核酸与其基因组邻居连接的两个序列读取结果。考虑到足够长的读取结果和足够短的文库片段,这两个读取结果可以在信息上合并以制造覆盖整个片段的一个长读取结果。使用条形码序列读取结果和从插入存在的9核苷酸重复序列,现在可以将读取结果与其基因组邻居连接以在计算机中形成长得多的“连接读取结果”。

[0146] 如将理解的,包含模板核酸的文库可以包含重复的核酸片段。对重复核酸片段进行测序在包括产生重复片段的共有序列的方法中是有利的。这样的方法可以提高为模板核酸和/或模板核酸的文库提供共有序列的准确性。

[0147] 在一些实施方式中,分析可以涉及来自单个细胞,或来自一个或多个细胞和/或细胞核区室的DNA或RNA或蛋白质或其任何组合。在一些实施方式中,分析是在10、100、1000、10,000、100,000、1,000,000、10,000,000、100,000,000、1,000,000,000或更多细胞,或细胞和/或细胞核区室中进行。

[0148] 在本文所述的测序技术的一些实施方式中,实时地进行序列分析。例如,可以通过同时获取和分析测序数据来进行实时测序。在一些实施方式中,可以在各个点终止用于获得测序数据的测序过程,包括在获得至少一部分的靶核酸序列数据之后或在测序整个核酸读取结果之前。示例性方法、系统和另外的实施方式在国际专利公开号W02010/062913中提供,其全部公开内容通过引用并入本文。

[0149] 在用于使用连接读取策略组装短测序读取结果的方法的示例性实施方式中,包含条形码的转座子序列被插入基因组DNA中,制备文库并且获得用于模板核酸的文库的测序数据。可以通过识别成对的条形码来组装模板的块(block),然后组装更大的重叠群(contig)。在一个实施方式中,通过使用重叠读取结果的代码配对,可以将组装的读取结果进一步组装成更大的重叠群。

[0150] 本文描述的测序技术的一些实施方式包括错误检测和校正特性。错误的实例可以包括在测序过程期间在碱基调用中的错误,以及将片段组装成更大的重叠群中的错误。可以理解,错误检测可以包括检测数据集中错误的存在或可能性,因此,可以不需要检测错误的位置或错误的数量。对于错误校正,关于数据集中错误的位置和/或错误的数量的信息是有用的。用于错误校正的方法在本领域中是公知的。实例包括汉明距离的使用和校验和算法的使用(参见,例如,美国专利申请公开号2010/0323348;美国专利号7,574,305;和美国专利号6,654,696,其全部公开内容通过引用并入本文)。

[0151] 在一些实施方式中,细胞和/或细胞核蛋白可以被分析、检测和/或测序。在一些实施方式中,这样的蛋白质被独特地标记。在一些实施方式中,蛋白质被保存、嵌入、固定或包含在一个或多个CE中。可以使用本领域已知的方法进行鉴定和测序。在一些实施方式中,蛋白质的鉴定和/或测序可以与收集核酸的序列信息一起进行。

[0152] 在一些实施方式中,索引可以通过在PCT期间从包含共同文库接头的反向互补序列的珠上寡核苷酸产生索引引物来进行(参见实施例3,下文)。在PCT期间,P7引物例如杂交至其在珠上寡核苷酸的3'末端的互补序列,并且其延伸导致产生完全功能性的索引寡核苷酸。索引寡核苷酸的量可以通过在适当的退火温度下的PCR循环的数量来特定地控制。然后合成的PCR引物与通过转座引入的文库片段的B15区域杂交,从而允许发生扩增。P7引物可以是生物素化的,以便在PCR后富集ATAC-seq片段用于测序。

[0153] 固体支持物

[0154] 本文所述的固体支持物可以用于测序和分析。固体支持物也可用作邻接保持元件。固体支持物可以是二维或三维的,并且可以包含平坦表面(例如载玻片)或可以成形。固体支持物可以包括玻璃(例如,受控孔玻璃(CPG))、石英、塑料(例如聚苯乙烯(低交联和高交联聚苯乙烯)、聚碳酸酯、聚丙烯和聚(甲基丙烯酸甲酯))、丙烯酸共聚物、聚酰胺、硅、金属(例如,烷硫醇盐衍生的金)、纤维素、尼龙、乳胶、葡聚糖、凝胶基质(例如,硅胶)、聚丙烯醛或复合物。

[0155] 合适的三维固体支持物包括,例如,球、微粒、珠子、纳米颗粒、聚合物基质例如琼脂糖、聚丙烯酰胺、藻酸盐、膜、载玻片、平板、微机械加工的芯片、管(例如,毛细管)、微孔、微流体装置、通道、过滤器、流动细胞、适于固定核酸、蛋白质或细胞的结构。固体支持物可包括能够具有包含模板核酸或引物的群的区域的平面阵列或基质。实例包括核苷衍生的CPG和聚苯乙烯载玻片;衍生的磁性载玻片;用聚乙二醇接枝的聚苯乙烯,等等。

[0156] 在一些实施方式中,固体支持物包含微球体或珠子。本文中的“微球体”或“珠子”或“颗粒”或语法等同物是指小的离散颗粒,其可以是球形、非球形或不规则形状。合适的珠子组合物包括但不限于塑料、陶瓷、玻璃、聚苯乙烯、甲基苯乙烯、丙烯酸聚合物、顺磁性材料、氧化钽溶胶、碳石墨、二氧化钛、乳胶或交联葡聚糖例如Sephadex、纤维素、尼龙、交联胶束和Teflon,以及本文概述的用于固体支持物的任何其他材料,都可以使用。在某些实施方式中,微球体或珠子是磁性和/或颜色编码的,以允许分离和操纵。珠子或微球体可以是实心的或中空的,并且可以是多孔的。通过适当选择材料和形成方法,可以根据需要调整珠子或微球体的孔隙率。珠子尺寸范围从纳米,即100nm,到毫米,即1mm,优选珠子为约0.2微米至约200微米,特别优选约0.5至约5微米,尽管在一些实施方式中可以使用更小或更大的珠子。

[0157] 在一些实施方式中,珠子可以包含抗体或其它亲和性探针(参见,Immobilized Biomolecules in Analysis.A Practical Approach.Cass T,Ligler F S,eds.Oxford University Press,New York,1998.pp 1-14,通过引用并入本文,用于典型的附着方案)。在一些实施方式中,抗体可以是单克隆的,并且在其他实施方式中,抗体可以是多克隆的。在一些实施方式中,抗体可以对细胞表面表位具有特异性。在一些实施方式中,抗体可以对细胞内的蛋白质具有特异性。

[0158] 在一些实施方式中,本文提供的核酸模板可以附着到固体支持物。在本领域中公知的各种方法可用于将核酸连接、锚定、或固定到固体支持物的表面上。

[0159] 在一些实施方式中,固体支持物可以包封细胞、细胞和细胞的细胞和细胞核区室,细胞的细胞和细胞核区室,或细胞的细胞核区室。

[0160] 实施例

[0161] 下面的实施例用于描述而不是限制本文所提供的公开内容。

[0162] 实施例1:用细胞通透性网格蛋白抑制剂对细胞核DNA进行转座片段化

[0163] 使用标准的细胞培养方案,从组织培养瓶中收集 1×10^6 K562细胞(人淋巴瘤细胞系)用于转座片段化反应。所有细胞沉淀/离心步骤均在4℃下以 $300 \times g$ 进行3分钟。将新鲜分离的细胞离心并在冰冷的PBS(磷酸盐缓冲盐水)中重悬。将细胞再次沉淀并在冰上孵育,同时悬浮于200μL冰冷细胞裂解缓冲液(Tris/NaCl/辛基苯氧基聚乙氧基乙醇(IGEPAL CA-630去垢剂))中5分钟。在此之后,细胞核在 $300 \times g$ 下旋下3分钟,用200μL裂解缓冲液洗涤一次,并且沉淀的细胞核在55℃下在 $1 \times$ Tagment DNA缓冲液(PN:15027866)中用Nextera Tn5转座酶(TDE1)转座30分钟,所述 $1 \times$ Tagment DNA缓冲液补充有(a)在DMSO中的Pitstop-2以获得30μM的最终Pitstop-2浓度(测试细胞核)或(b)2%DMSO(对照细胞核)。沉淀来自每个样品的细胞核,并且小心移除大部分上清液。

[0164] 实施例2.用链置换聚合酶从转座片段化DNA中去除转座酶

[0165] 将来自前面实施例(测试和对照池)的转座片段化细胞核重悬浮于含有50mM Tris-HCl pH 7.5、10mM MgCl₂、10mM (NH₄)₂SO₄、4mM DTT、0.005% SDS、200μM各种dNTP和10单位的Phi29DNA聚合酶(New England Biolabs)的Phi29反应混合物中,并且将所得的混合物在30℃下在热循环仪中孵育30分钟。然后,细胞核在台式条管离心机中通过温和离心沉淀,并且重悬浮于30μL终体积的在Tris缓冲液中的30%Optiprep(Sigma-Aldrich)。

[0166] 实施例3.用FAM标记的复合物使转座子对细胞核的接近可视化

[0167] 使用实施例1的方法将FAM标记的转座体复合物输入到细胞核区室中。使用与在寡核苷酸合成期间在5'末端用FAM(6-FAM, Integrated DNA Technologies)标记的转座子序列复合的Nextera Tn5转座酶,并且使用2%DMSO对照,和不同浓度的Pitstop-2(10 μ M、30 μ M和60 μ M)如实施例1所述分离并处理细胞核。如图3所示,用Pitstop-2处理增加了FAM标记的转座体进入核空间的流入,特别是在细胞核周边(箭头所示)。

[0168] 实施例4.Pitstop-2处理对DNA文库产生的影响

[0169] 如在前述实施例中所述,在不同浓度的Pitstop-2(0至80 μ M)的存在下,用Nextera Tn5试剂盒转座分离的细胞核(50,000)。通过短暂离心除去过量的转座体,并将得到的细胞核沉淀用Qiagen蛋白酶(ID19155)在50 $^{\circ}$ C下处理1小时,然后在80 $^{\circ}$ C下处理10分钟。在Bio-Rad Thermocycler上用KAPA SYBR Fast主混合物和退火至转座子的镶嵌端(ME)部分的引物用25周期的qPCR分析得到的裂解物。对应于不同Pitstop-2浓度的Cq值和Pitstop-2样品与对照样品(无Pitstop-2)之间的Delta Cq显示在表1中。一个Cq值大致对应于DNA输入的两倍差异。如表中所示,与未处理对照相比,用Pitstop-2处理完整细胞核增加了文库产量。一个Cq值大致对应于DNA输入的两倍差异。

[0170] 表1A.Pitstop-2增加了来自转座细胞核的文库产量

[0171]	Pitstop-2 浓度(μM)	Cq-Pitstop- 2	% DMSO	Cq-DMSO	ΔCq-Pitstop-2 相对于对照
	0	20.94	0	20.18	0.755
	5	19.91	0.05	20.11	-0.2
	10	20.00	0.10	20.64	-0.65
	20	19.36	0.20	20.84	-1.48
	40	19.94	0.40	21.20	-1.26
	80	19.24	0.80	21.30	-2.06

[0172] 使用不同浓度的Pitstop-2和Tn5和TsTn5转座酶进行另外的实验。这些实验的数据显示在表1B中。

[0173] 表1B.

[0174]	转座体	Pitstop (μM)	Cq
	Tn5	300	23.67
	Tn5	200	23.105
	Tn5	100	23.17
	Tn5	50	22.21

[0175]	Tn5	0	24.21
	TsTn5	300	19.475
	TsTn5	200	18.55
	TsTn5	100	18.52
	TsTn5	50	18.32
	TsTn5	0	20.97

[0176] 实施例5.转座文库的测序

[0177] 使用如在前述实施例中所述的方法的从K562 (人淋巴瘤细胞系) 细胞核内容物产生DNA文库。对文库进行测序, 并且结果证实Pitstop-2处理增强了对转座酶可接近基因间和开放染色质区域的接近。如图4A中针对Chr 12的一部分上的ATAC-seq谱所示, 文库在所有Pitstop-2浓度下已经主要靶向基因间区域。如图4B所示, 对测序数据的更详细分析显示全基因组测序覆盖的基本度量, 这表明Pitstop-2处理的样品相对于对照更好的多样性和独特性, 与采用Pitstop-2的改善的文库产量一致。图4C显示了人类基因组的启动子和编码区上的标准化覆盖, 并且数据显示用30和60 μ M Pitstop-2处理改善了对照上的ATAC-seq谱, 显示了基因间区域上增加的覆盖。这些结果证实, Pitstop-2处理增强了转座体复合物进入细胞核并改善了细胞核转座的效率。

[0178] 实施例6. 在Pitstop-2存在下对转座细胞核DNA进行索引

[0179] 如前所述在Pitstop-2的存在下 (在0、10、30、和60 μ M的浓度下) 用Nextera Tn5Kit转座的基因组DNA从细胞核中释放, 并且被捕获到含有寡核苷酸捕获序列的索引珠子上。将珠子捕获的DNA进一步间隙填充 (gap-fill) 连接以将基因组DNA的两条链与共同序列 (包含独特的索引) 共价连接并通过PCR扩增。当通过琼脂糖凝胶电泳进行分析时, 扩增的文库表现出通过使用Pitstop-2, 转座产量明显增加 (图5)。条带强度增加与Pitstop-2浓度的增加相关。由得到的文库中的不同数量的核小体覆盖的相应DNA片段包含“核小体条带”, 表明在没有核小体的区域成功的ATAC-seq转座片段化。

[0180] 在一个单独的实验中, 使用如前所述的方法, 通过Pitstop-2 (在0、10、30、和60 μ M的浓度下) 处理, 获得FAM-标记的转座细胞核物质, 并在具有捕获寡核苷酸的珠子上富集, 并且通过流式细胞术分析珠子以评估捕获效率。与先前的结果一致, 用Pitstop-2处理制备的FAM标记的转座文库表现出比对照池增强的珠子捕获率 (表2)。

[0181] 表2.

[0182]

Pitstop-2浓度	空白	0 μ M	10 μ M	30 μ M	60 μ M
平均 (MFU)	1044	1342	1297	1439	1488

[0183] 实施例7. 通过链置换聚合酶去除转座酶

[0184] 在某些转座文库制备方法中, 从DNA除去转座酶造成转座片段化DNA的进一步片段化。相比之下, 使用链置换扩增去除转座酶保持了转座细胞核和细胞核的所有单个细胞相关个体文库的邻接性。

[0185] 作为转座片段化步骤之后和扩增之前的gDNA纯化的替代, 链置换聚合酶可以被用

于实现在单一步骤中除去转座酶和扩增。在30μM Pitstop-2存在下如在前述实施例中所述制备的完整细胞核 (24ng) 中未纯化的细胞核转座片段化产物用Illumina NPM PCR主混合物、链置换聚合酶 (Bst) 和低浓度SDS (0.0025%-0.04%) 进行处理。如图6所示,Bst的添加改善了转座片段化DNA的扩增效率。

[0186] 在进一步的实验中,使用链置换聚合酶以从转座片段化DNA中除去转座酶,从转座的沉淀的细胞核产生测序文库。在链置换扩增后在上清液中没有观察到测序文库,证实没有DNA片段从细胞核“泄漏”到上清液中。细胞核可用作“区室”以进行反应和测定步骤,而不丢失关于单个细胞的组分的边界或直接接触的状态的邻接信息。

[0187] 实施例8. 来自未纯化细胞核的ATAC-seq文库的产量

[0188] 如在前述实施例中所述用Pitstop-2 (10和100μM) 处理细胞核增加了从未纯化的细胞核样品产生的ATAC-seq文库的产量,如在BioAnalyzer仪器上的高灵敏度芯片上的1μL各ATAC-seq文库的分析所示。各个文库的浓度在BioAnalyzer上确定为对各Pitstop-2浓度绘制的150-1000bp文库范围之间 (图7A)。在不同的Pitstop-2浓度 (5、10、30、50、100和200μM) 下的另外的实验运行产生了相对于不存在Pitstop-2时产生的文库增强的文库产量,如分光光度分析所示 (表3)。

[0189] 表3.

[0190]

Pitstop-2浓度 (μM)	%DMSO	产量 (nM) 150-1000bp
0	3	2.9
5	3	7.3
10	3	9.3
20	3	20.6
50	3	23.7
100	3	15.7
200	3	8.7

[0191] 图7B显示对于使用0和100μM Pitstop-2的研究在Chr 6的一部分上的ATAC-seq谱。

[0192] 实施例9. 来自细胞核的RNA输出

[0193] Pitstop-2处理通过改善来自细胞核mRNA的cDNA合成来增加RNA输出。用0.1% NP40裂解K562细胞。将细胞核和细胞质进行分离,并且细胞核被洗涤两次,然后用 (a) 无添加剂、(b) 0.1% SDS、或 (c) 30μM Pitstop-2进行逆转录/定量PCR方案 (ThermoFisher Maxima HT RT试剂盒,Kapa qPCR试剂盒,RPLPOPCR引物)。无逆转录和逆转录之间的 ΔCt用作报告相对cDNA合成效率的度量。Pitstop-2处理使来自细胞核的cDNA合成相对于未处理的细胞核提高4-5倍 (表4)。

[0194] 表4.

[0195]	运行	DeltaCt
--------	----	---------

[0196]	细胞质	8.84
	细胞核 (未处理)	4.63
	细胞核 (+ 0.1% SDS)	2.47
	细胞核 (+ 30 μ M Pitstop-2)	7.19

[0197] 实施例10:索引珠子的制备

[0198] 使得能够进行寡核苷酸偶联(例如肼-醛,环氧化物-胺等)的30 μ m直径和 1×10^7 珠子密度的表面官能化珠子在去离子水中洗涤。将洗涤过的珠子以浓度为 1×10^5 个珠子/50 μ L加入到96孔板的每个孔中。然后,将500pmole独特的5' 醛索引oligo_1 (/5FormInd/CCGAGCCCACGAGAC INDEX1GACTTGTC) 分别加入到每个孔中至每孔最终体积为50 μ L反应混合物。将平板密封以避免蒸发并在37 $^{\circ}$ C下在旋转培养箱中孵育过夜。合并珠子并用0.1 $\times$ TE/0.1%Tween-20洗涤三次。洗涤的珠子以与先前所描述的相同的方式分配到96孔板的个体孔中,独特的5' 磷酸化的索引oligo_2 (/5Phos/TAGAGCAT INDEX2ATCTCGTATGCCGTCTTCTGCTTG),夹板寡聚物(splint oligo),与悬浮于T4DNA连接酶缓冲液的T4DNA连接酶一起加入其中。使连接反应在室温下过夜进行。然后最后将珠子从所有孔合并在一起并用0.1 $\times$ TE/0.1%Tween-20洗涤三次。

[0199] 实施例11.“在飞行中(on the fly)”索引引物的产生

[0200] 在PCR过程中用细胞核的混合物或在液滴中进行索引需要一组在组合合成珠子上呈现的个体索引PCR引物。这些引物可以在用于PCR之前从珠子上切割下来,或者可以是具有实际PCR引物的反向互补序列的寡核苷酸(“接种寡聚物”)。在后一种情况下,PCR混合物中的共同引物(例如P5或P7)将与珠子上的索引“接种寡聚物”杂交,并产生PCR中使用的索引引物。可以调节“接种寡聚物”的密度以优化PCR,并且每轮PCR将产生另外量的索引PCR引物。如图8A所示,将具有索引PCR引物的反向互补序列的接种寡聚物偶联至珠子。PCR混合物补充有共同引物,其退火至“接种寡聚物”,并且在每个PCR循环中延伸时产生用于文库扩增的引物。

[0201] 在一个实验中,将含有在其3' 端上具有P7的反向互补序列的索引寡聚物(B15', 核索引,P7')的缀合珠子以250个珠子/ μ L加入到PCR主混合物。将由Nextera PCR Master Mix (NPM)、0.01%SDS、0.5 μ M样品索引引物(P5,样品索引和A14')、0.5 μ M P7引物和30%Optiprep组成的主混合物加入到滴式发生芯片上的孔1中。将细胞核(6×10^4)与悬浮于30%Optiprep的250个珠子/ μ L一起加入孔2中。索引引物在PCR期间在P7与其3' 端杂交、延伸并从珠子上变性脱离时从珠子上的寡聚物产生。P7引物也可以被生物素化以便在PCR后富集ATAC-seq片段用于测序。然后合成的PCR引物与通过转座产生的文库片段的B15区域杂交,从而允许发生扩增。遵循以下PCR参数:在73 $^{\circ}$ C初始延伸3分钟,98 $^{\circ}$ C30秒,98 $^{\circ}$ C25个循环10秒,56 $^{\circ}$ C30秒和73 $^{\circ}$ C30秒。

[0202] 在其它实验中,接种寡核苷酸以不同浓度附着在珠子上并且这些珠子用于在PCR中作为PCR引物之一的来源,与含有PCR引物两者的阳性对照、和只有一种PCR引物的阴性对照并排。如图8B中所呈现,用具有偶联的“接种寡聚物”的珠子产生的文库提供了足量的第二PCR引物以支持有效的扩增。图6B显示用两种PCR寡核苷酸(+对照)或一种PCR寡核苷酸和

具有接种寡核苷酸的珠子 (5、50和500pmol珠子) 和实验中使用的试剂扩增文库。

[0203] 实施例12.液滴产生

[0204] 使用QX200Bio-Rad液滴发生器产生包含细胞、珠子、和油入口的液滴。简言之,将10 μ L引发溶液(例如水性缓冲液或细胞培养基)加入细胞入口和珠子入口,并使其充满通道1分钟,然后取出。将细胞核和珠子溶液(各30 μ L)加入到各自的孔中。在将盒插入到液滴发生器之前将大约70 μ L的油加入到各油入口。在液滴产生后,从出口两者将液滴合并在一起,并进行PCR。PCR之后,油包水乳液通过用乳化剂处理而破坏,并且水相用于测序。

[0205] 实施例13.细胞核作为酶反应的物理区室

[0206] 活细胞具有多种细胞器,其具有不同机制来进行单独的酶反应,例如在细胞核中的基因组维护、线粒体中的能量产生、溶酶体中的代谢过程等。为了利用由活细胞提供的天然区室化,进行说明细胞核可以提供物理屏障用于基因组DNA的分离的体外反应的试验。

[0207] 细胞核(150,000)用含有0.1% IGEPAL CA-630的标准细胞核分离缓冲液提取,并且细胞核沉淀用含有或不含有30 μ M Pitstop-2的Tn5转座体在55 $^{\circ}$ C下转座30分钟。完成该步骤时细胞核的完整性通过光学显微镜监测(数据未显示)。用温和离心沉淀细胞核,除去过量未使用的转座体。细胞核被重悬浮在含有强链置换DNA聚合酶(Bst或Phi 29)的反应混合物中,并且在适当的温度(65 $^{\circ}$ C,BST;30 $^{\circ}$ C,phi29)下孵育30分钟以从转座片段化染色质中除去Tn5,并且使得DNA片段可用于PCR扩增。反应后完成,细胞核通过温和离心从上清分离并且重悬浮在PCR主混合物中。

[0208] 用Nextera索引试剂盒引物扩增细胞核和上清液两者(25个循环),并且反应产物通过琼脂糖凝胶电泳进行分析,如图9所示。没有文库在任一组条件下从上清产生,这表明DNA片段在转座片段化和Tn5置换反应期间不从细胞核扩散出来。

[0209] 数据显示,细胞核被成功地用作转座片段化和其它酶促过程的物理区室(参见图1、图2和10)。因此,本方法实现细胞核内转座产物的区室特异性扩增。酶,例如Tn5转座体复合物和聚合酶,以及其他反应混合物组分,能够进入并在Pitstop-2处理的细胞核内发挥作用。DNA和文库元件在反应期间和反应后保留在细胞核中。因此,可以在细胞核内完成多个酶促反应而不破坏核膜的完整性,并且细胞核可以经历多种试剂交换,多个过程和/或各种操纵(例如酶促反应、染色、分选等)而没有DNA从细胞核空间泄漏。

[0210] 实施例14.Pitstop-2对液滴中单个细胞ATAC-seq度量的影响

[0211] 小鼠(3T3)细胞核在细胞裂解和细胞核转座期间用0 μ M、30 μ M和60 μ M Pitstop-2处理。Pitstop-2用于打开细胞核孔并且允许开放染色质的转座增加。将确保单个细胞核占用的限定数量的转座细胞核组加载到具有设计以产生单个细胞ATAC-seq文库的索引珠子的QX200Bio-Rad液滴发生器中。单个细胞文库在液滴中进行PCR扩增,然后进行纯化,并在NextSeq[®]550测序系统中测序。图11显示各组处理细胞(0 μ M、30 μ M、和60 μ M Pitstop-2)的单个细胞测序数据。如图11A所示,增加Pitstop-2的浓度导致按独特读取结果的不同数量(>1k,>10k和>100k)分组的所有三组条形码的独特读取结果的总数增加。

[0212] 用Pitstop-2处理细胞核允许更有效的转座和增加的开放染色质的独特覆盖。在存在Pitstop-2的情况下,转座酶靶向转录起始位点(TSS)之外的另外的开放染色质区域(见图11B)。图11B显示对于含有0 μ M、30 μ M、或60 μ M Pitstop-2的ATAC seq样品,在转录起始位点的1kb内的片段的百分比。在两个基因组位置(GAPDH和RPL10基因)的读取结果比对的

详细检查表明在具有潜在开放的染色质结构的区域(例如增强子和开放阅读框)处的文库产生,表明测定的更高灵敏度(数据未显示)。图11C显示对于用0 μ M、30 μ M、和60 μ M Pitstop-2处理的样品,由更保守的MACS-2峰发现软件程序定义的ATAC-seq峰重叠。对于含或不含Pitstop-2的样品,检测到非常强的峰重叠,表明Pitstop-2没有改变测定的整体完整性(参见图11C)。也观察到了由0 μ M、30 μ M、和60 μ M Pitstop-2在染色体6处的GAPDH基因周围产生的合并的ATAC-seq读取结果的比对和由0 μ M、30 μ M、和60 μ M Pitstop-2在染色体X处的RPL10基因周围产生的合并的ATAC-seq读取结果的比对(数据未显示)。

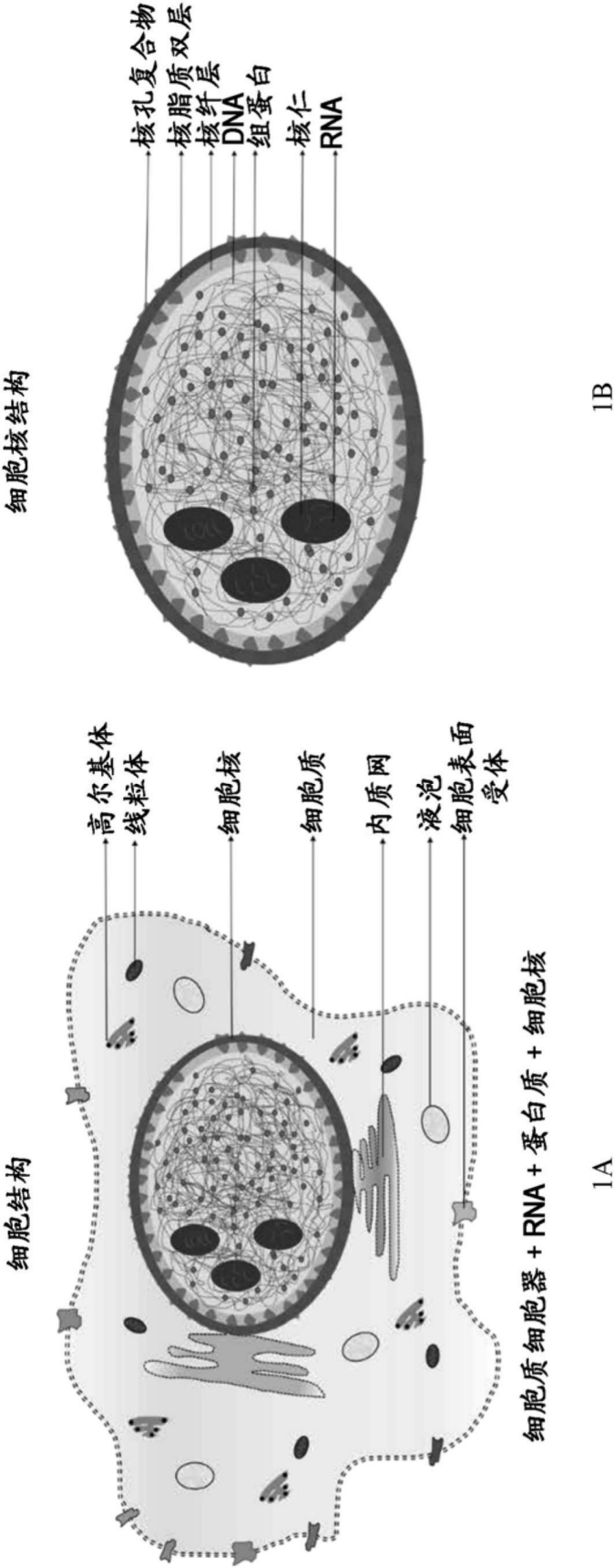


图1

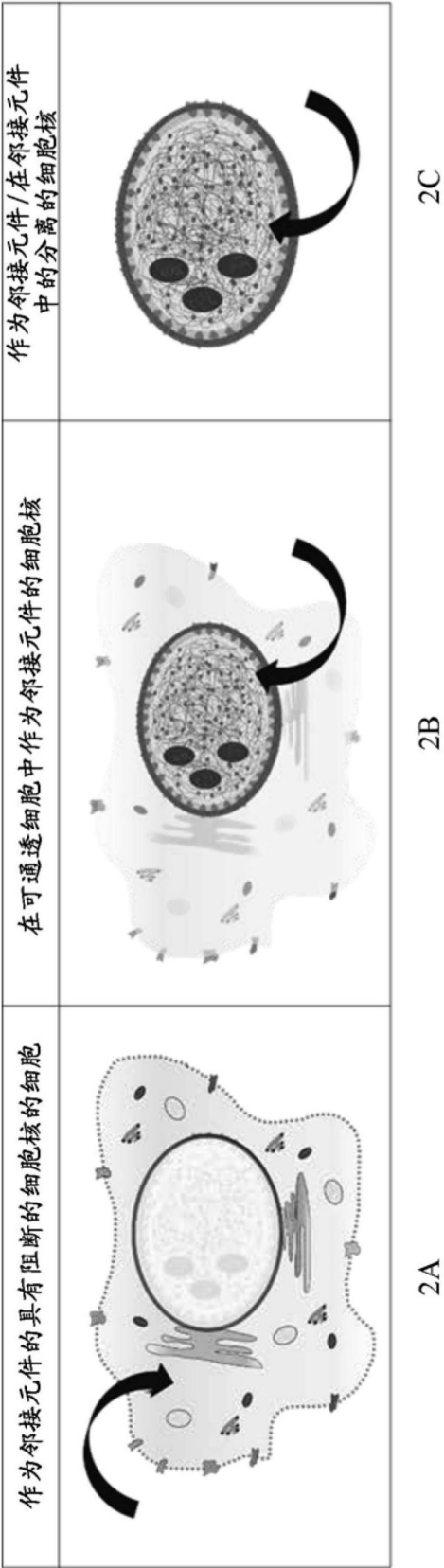


图2

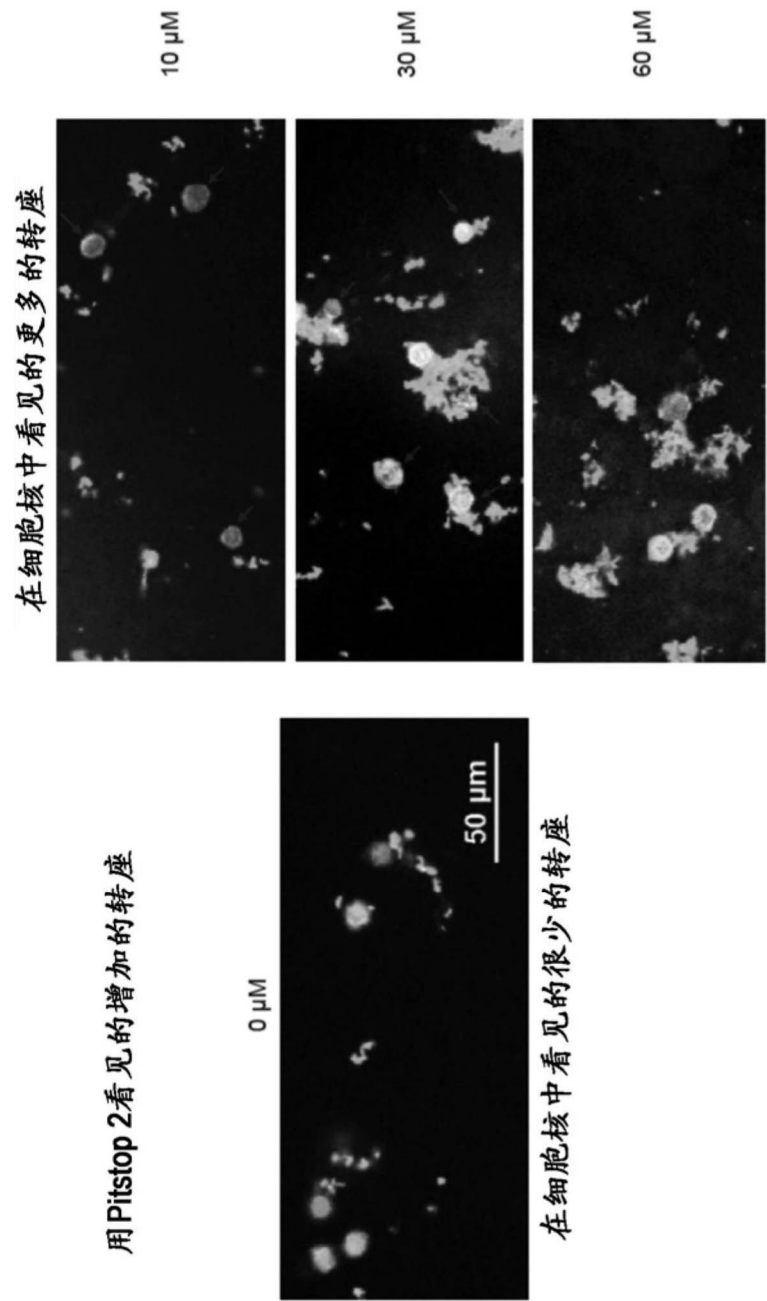


图3

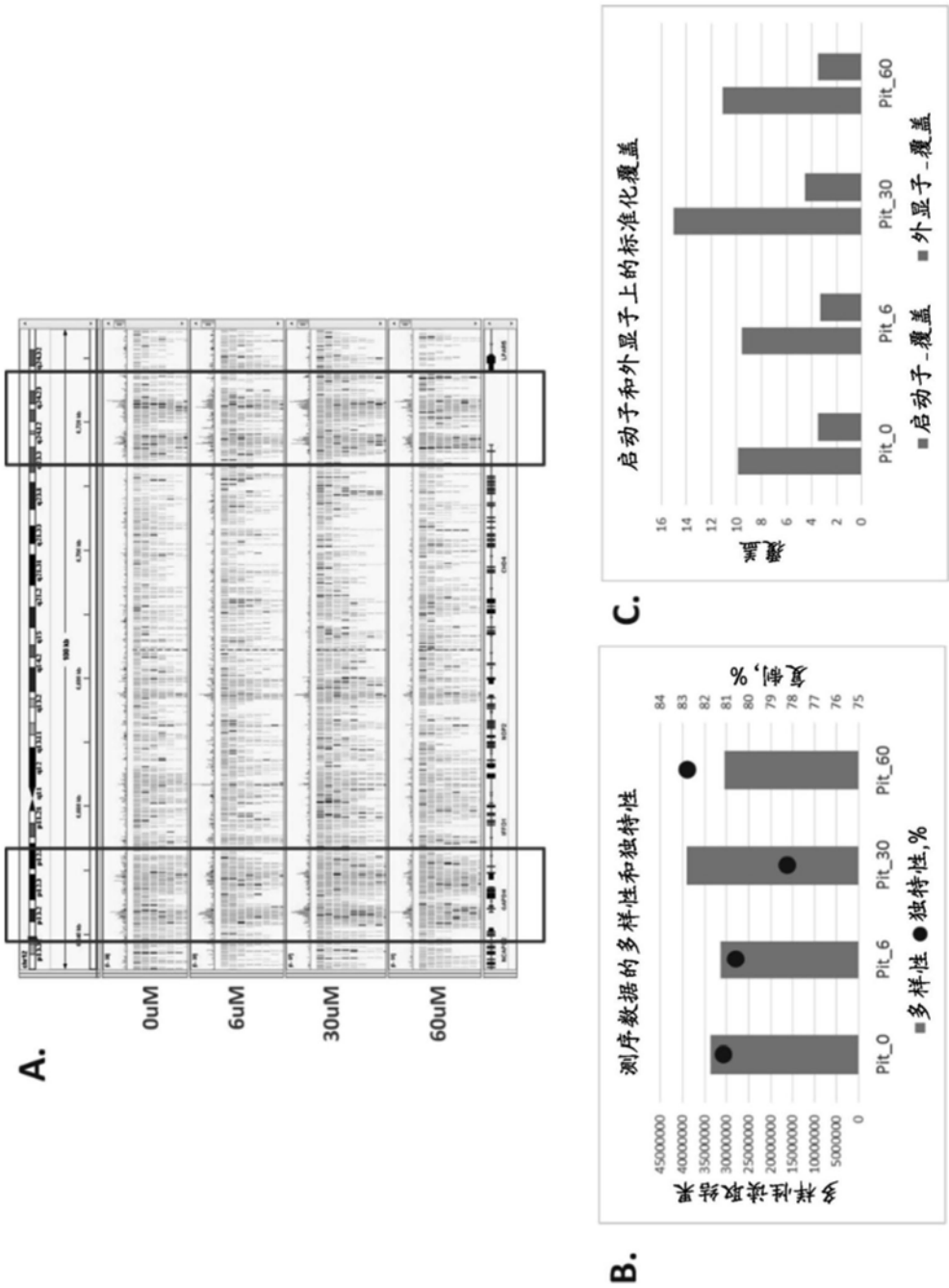


图4

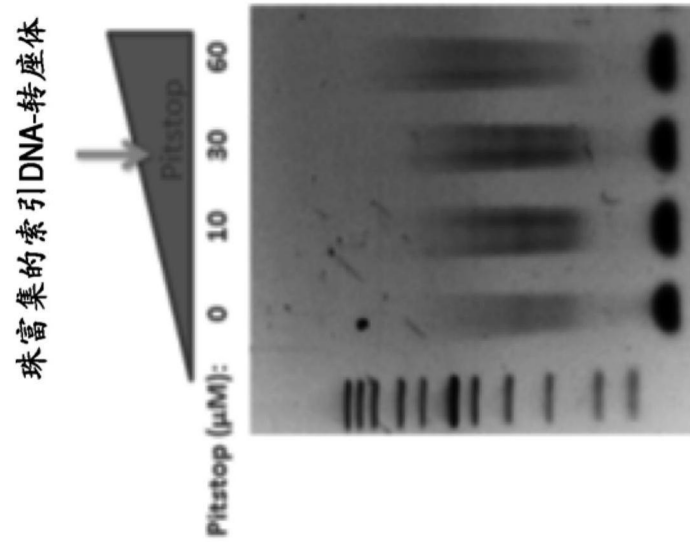


图5

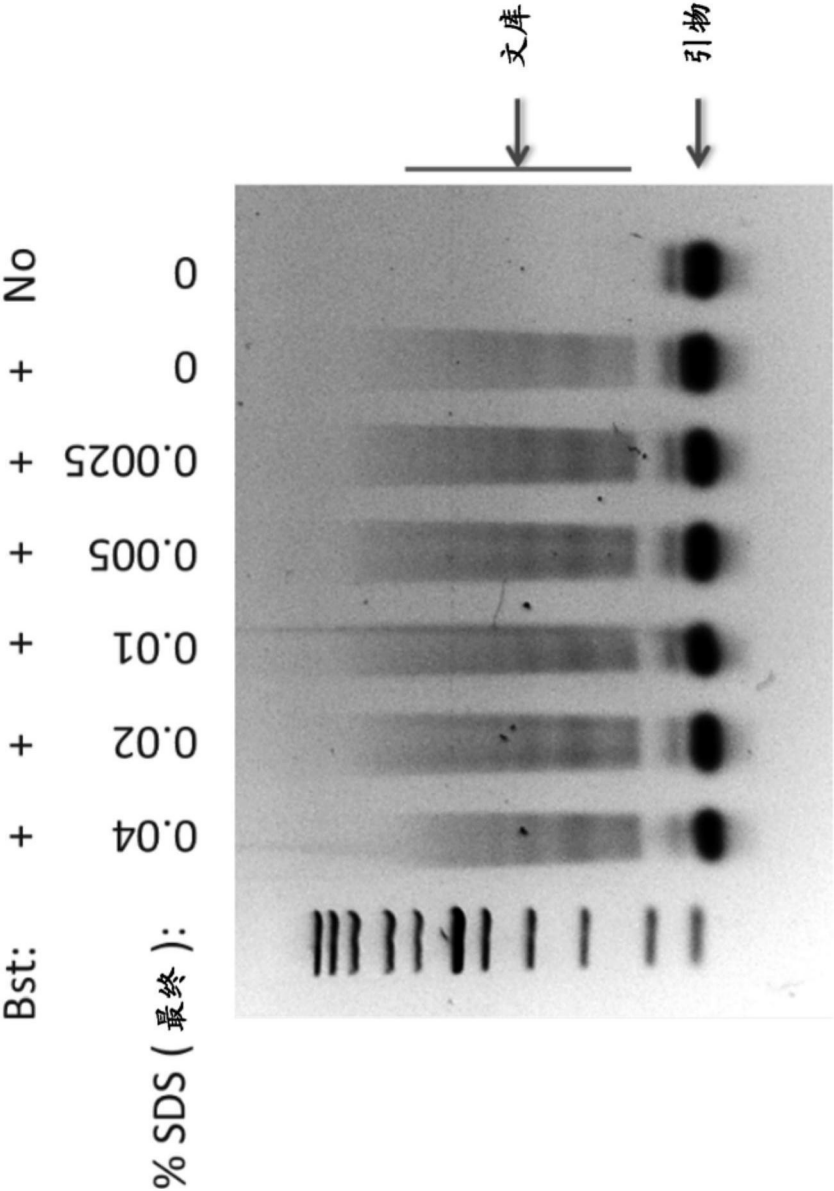


图6

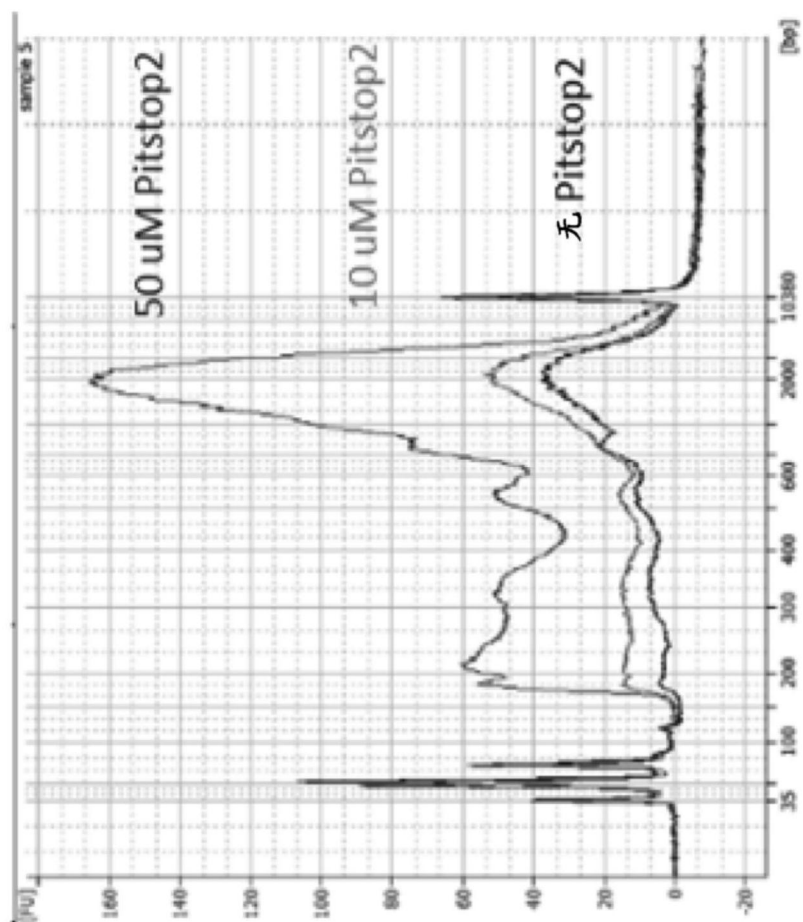


图7A

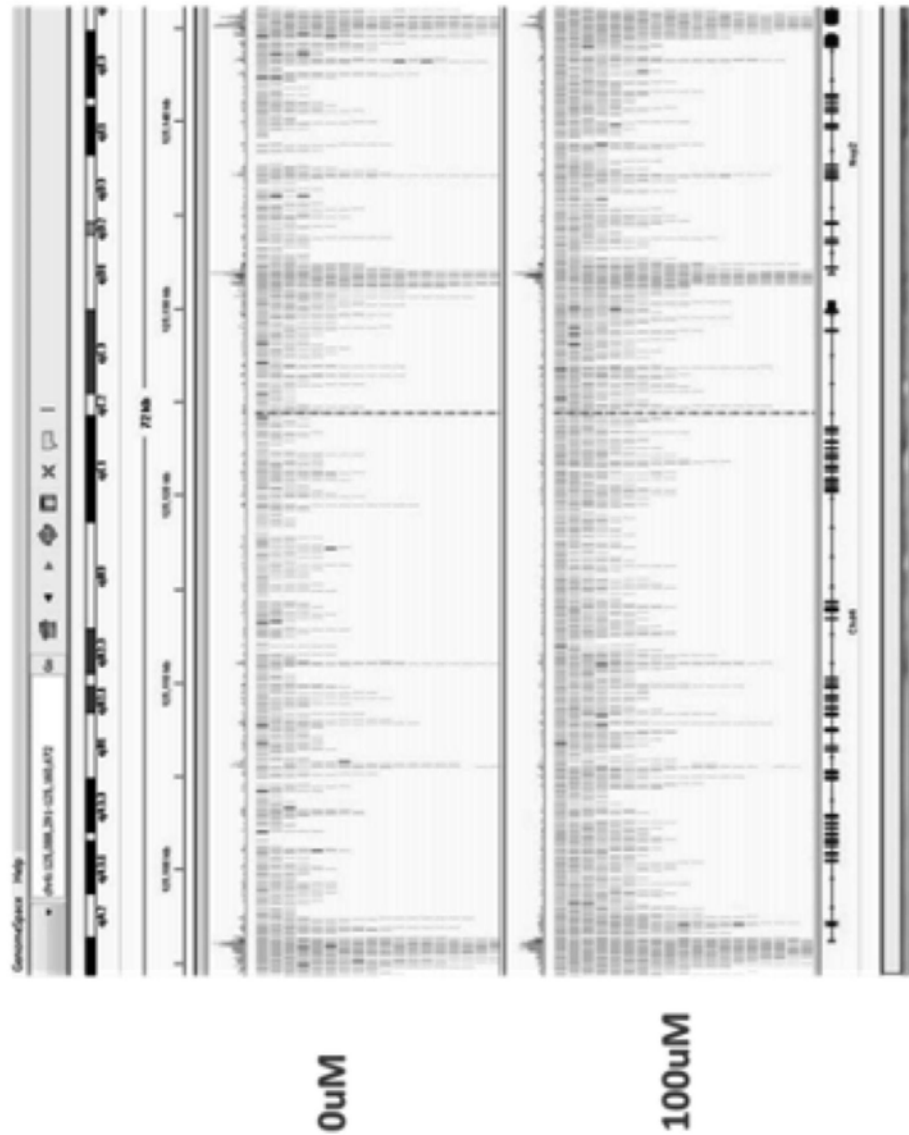


图7B

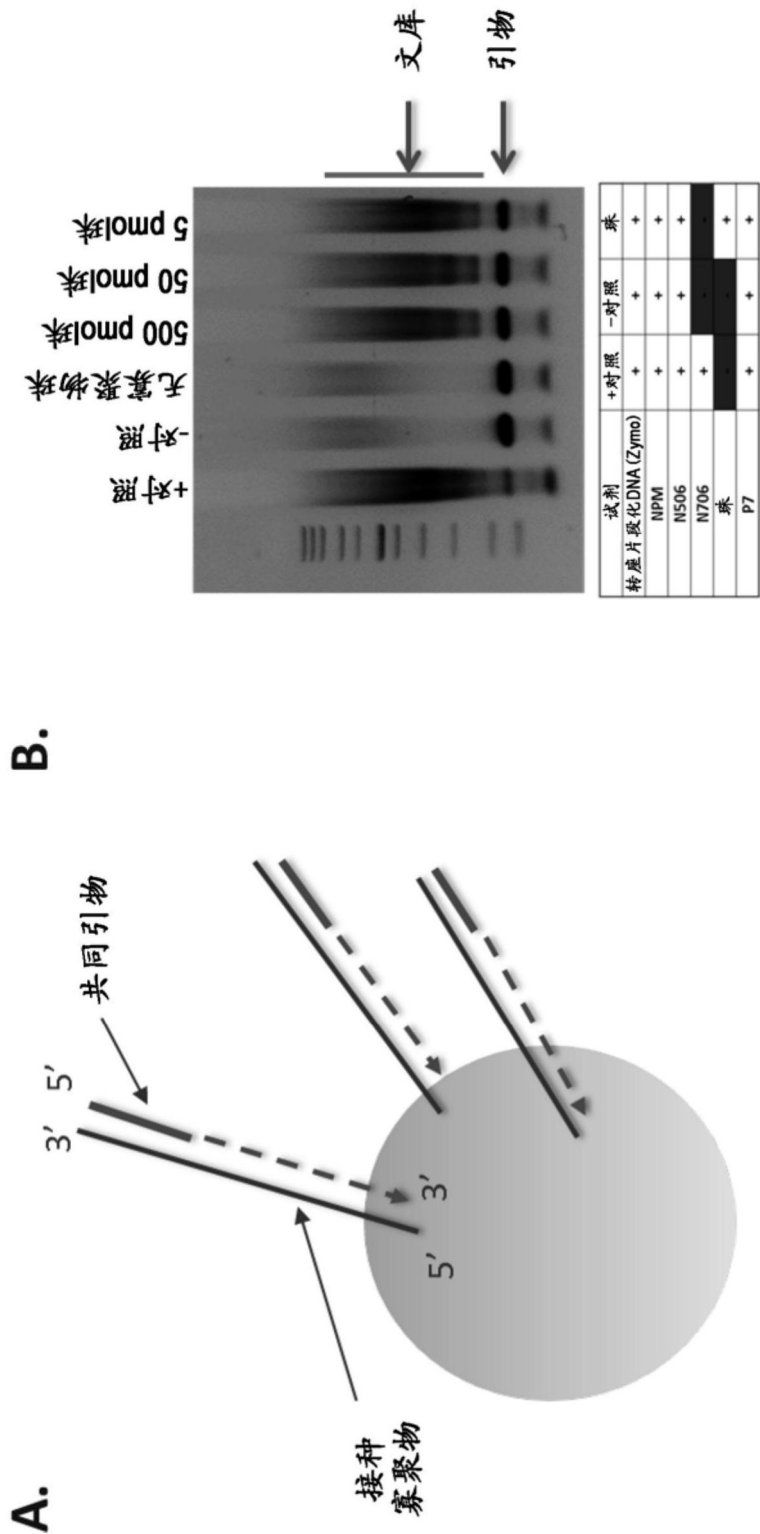


图8

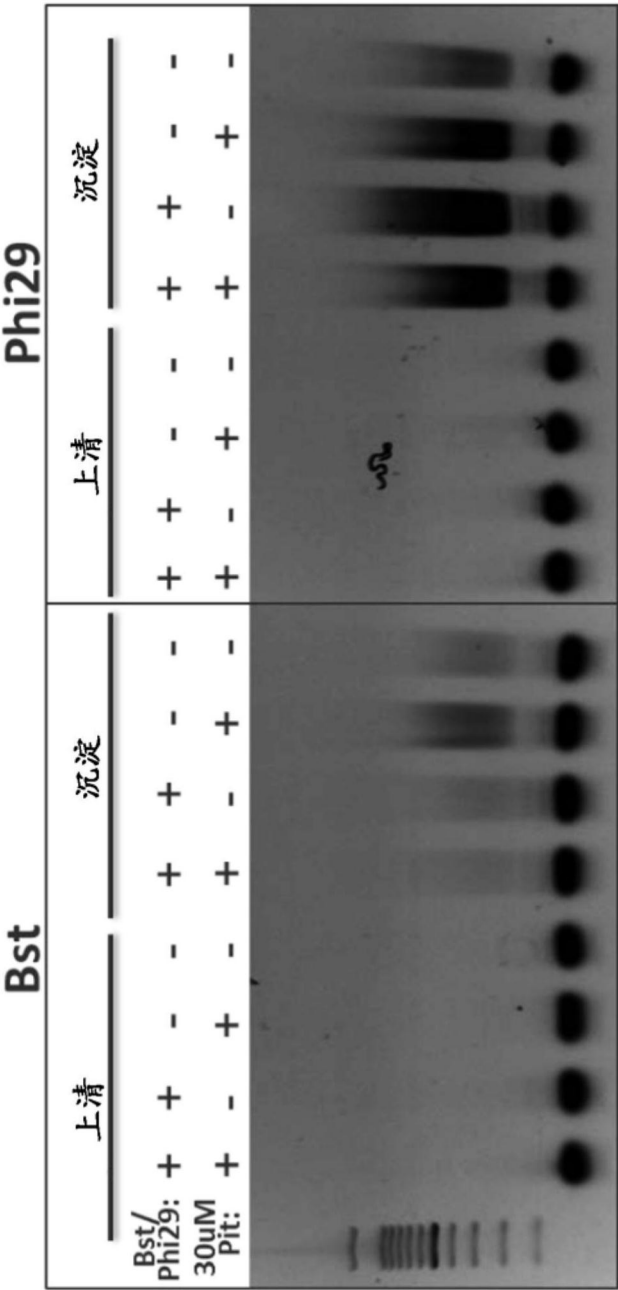


图9

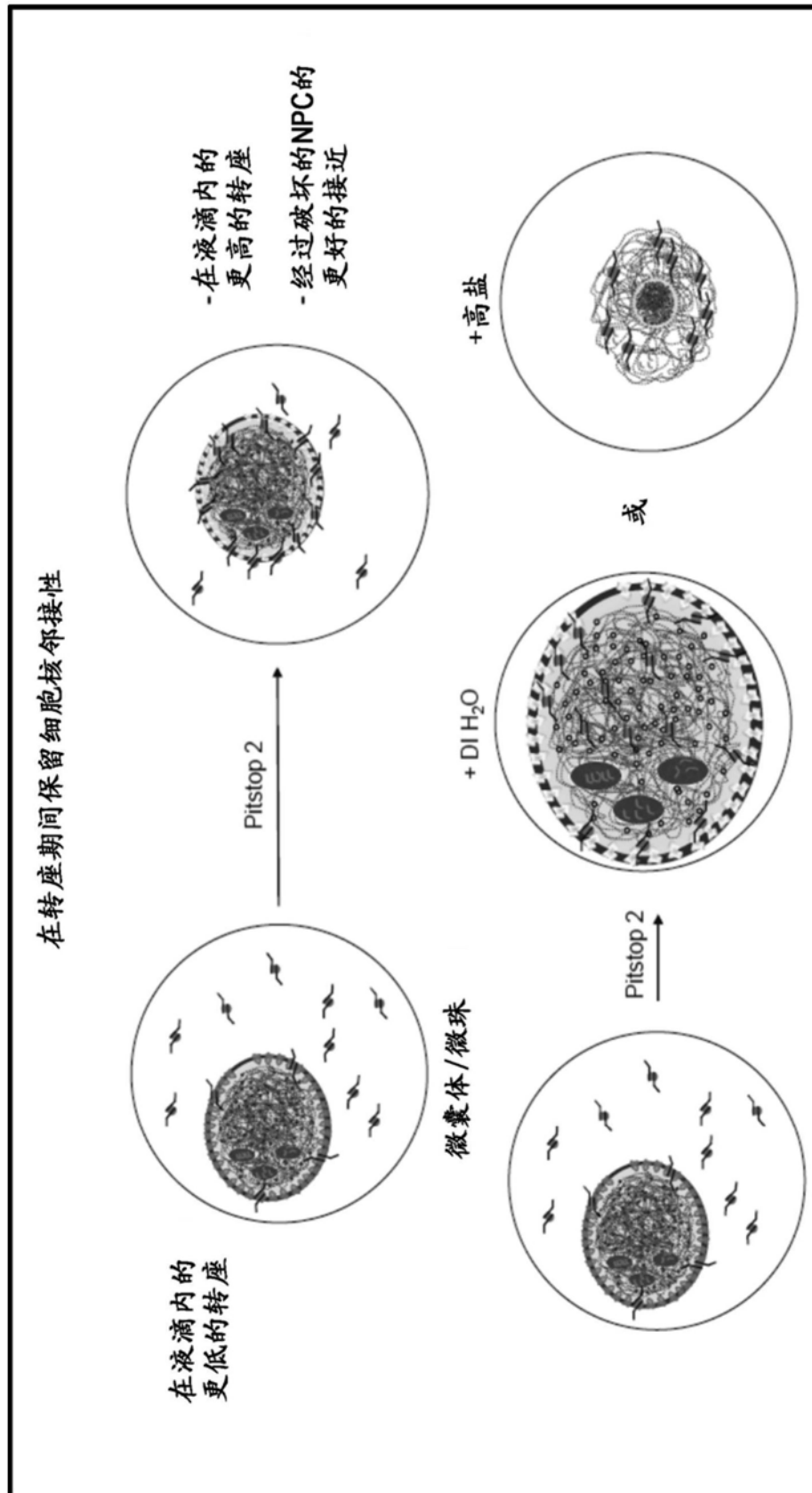


图10

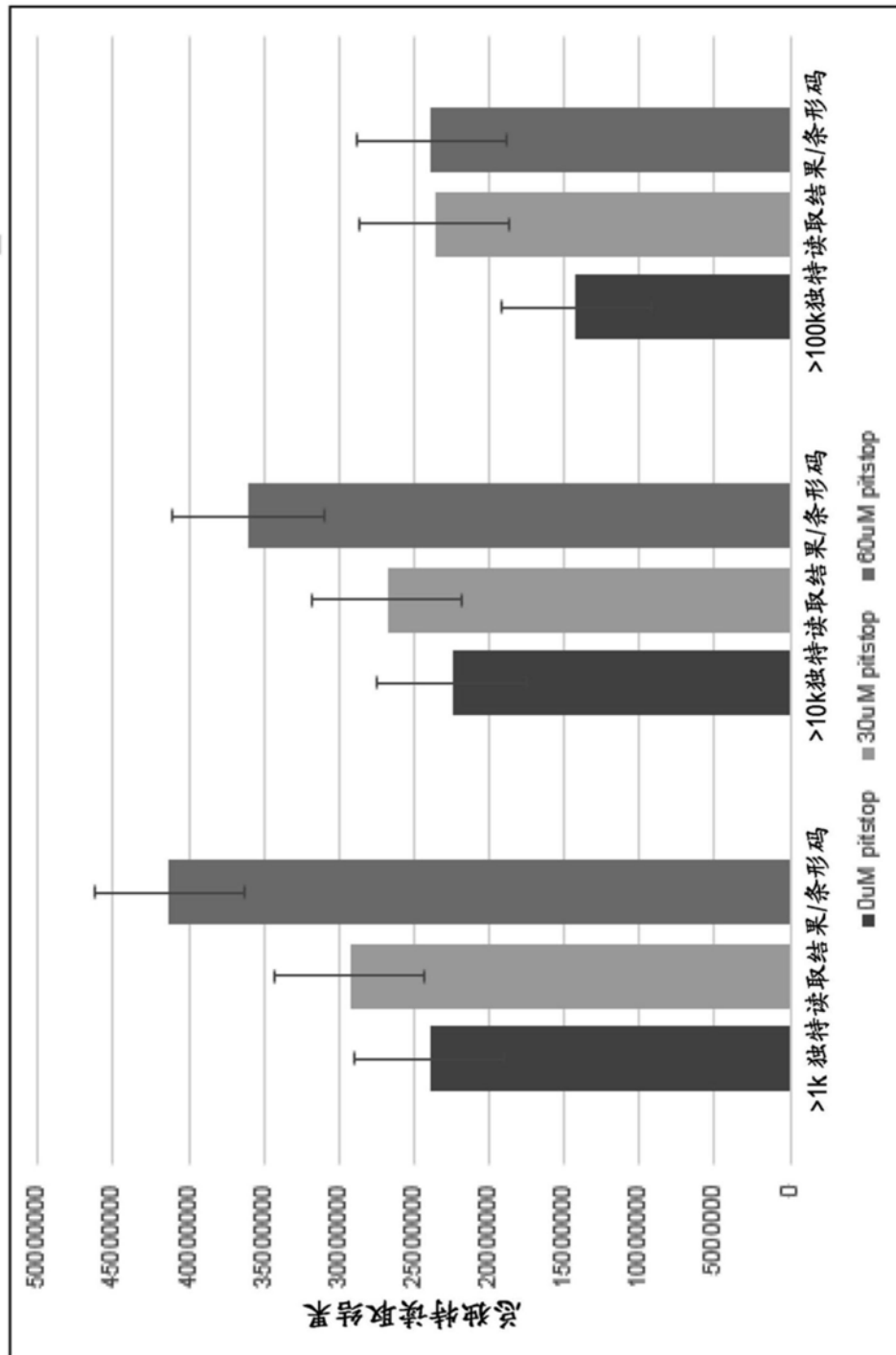


图11A

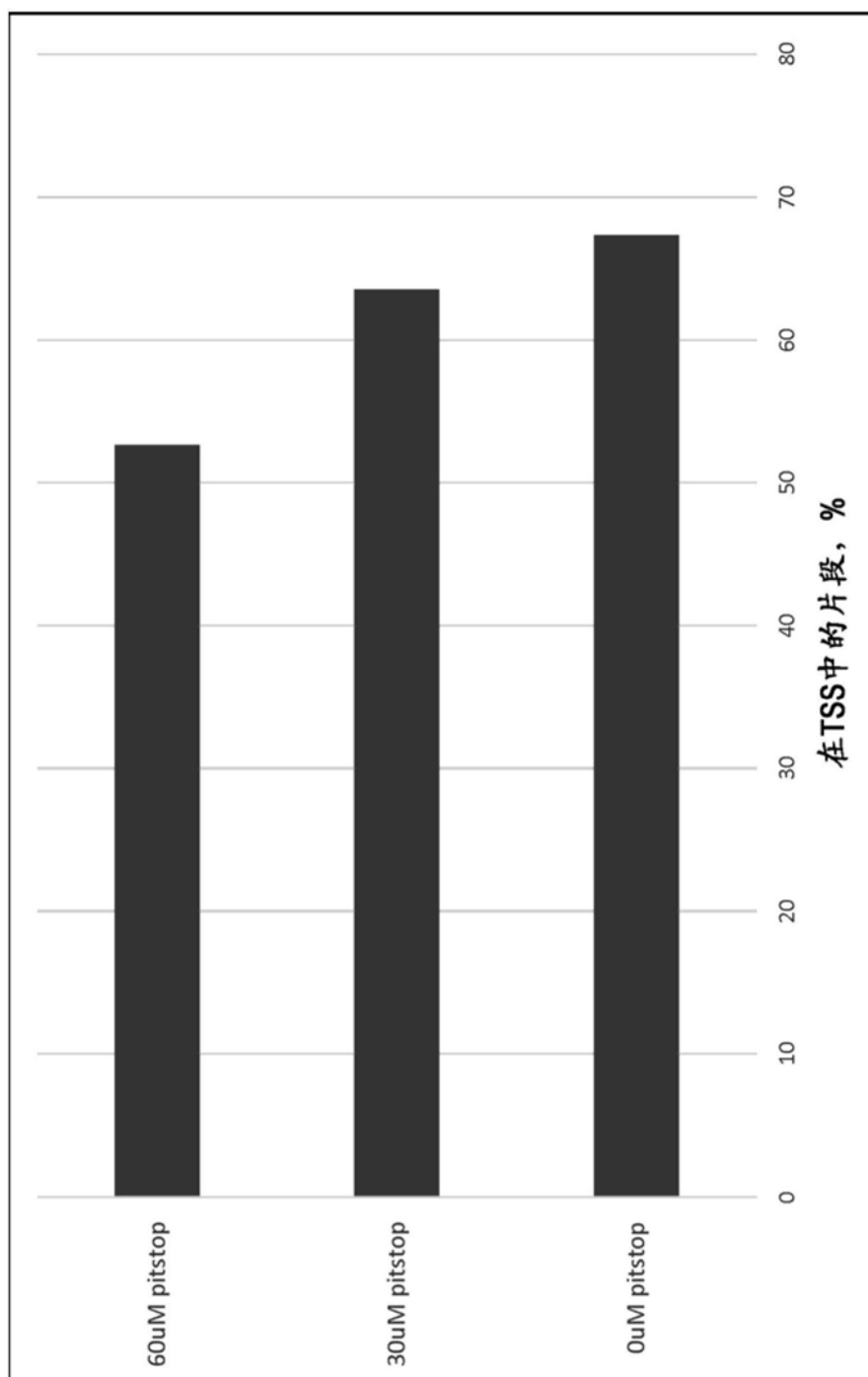


图11B

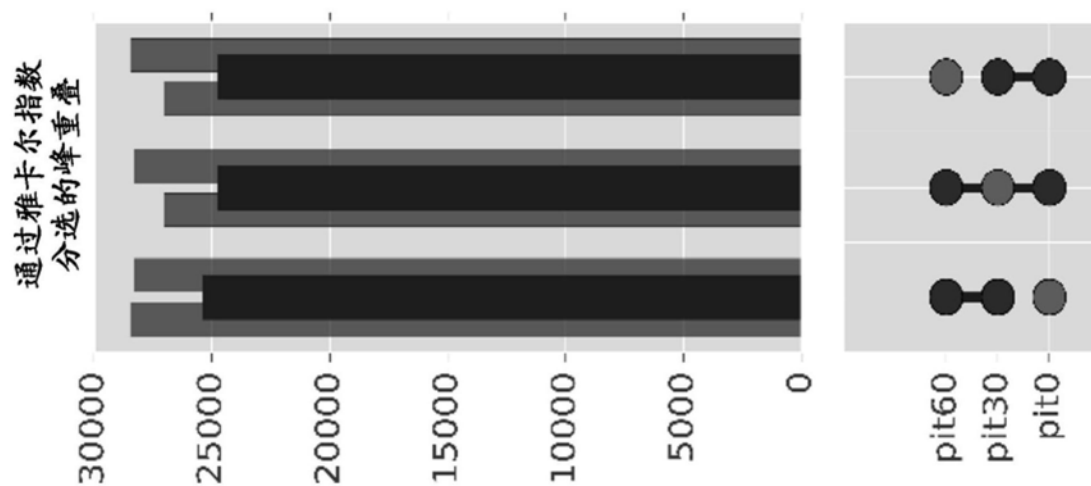


图11C