



(12) 发明专利

(10) 授权公告号 CN 108368540 B

(45) 授权公告日 2022.10.11

(21) 申请号 201680058837.9

(74) 专利代理机构 北京市中咨律师事务所

11247

(22) 申请日 2016.08.11

专利代理人 柴云峰 黄革生

(65) 同一申请的已公布的文献号

申请公布号 CN 108368540 A

(51) Int.CI.

C12Q 1/6806 (2018.01)

(43) 申请公布日 2018.08.03

C12Q 1/6869 (2018.01)

(30) 优先权数据

15180705.4 2015.08.12 EP

(56) 对比文件

WO 2014190214 A1, 2014.11.27

15189788.1 2015.10.14 EP

CN 102703426 A, 2012.10.03

(85) PCT国际申请进入国家阶段日

刘开辉等.紫外激光交联和染色质免疫沉淀

2018.04.09

技术及其应用.《安徽农业科学》.2007,第35卷

(86) PCT国际申请的申请数据

(第20期),第6058-6059页.

PCT/EP2016/069121 2016.08.11

Terrence S Furey等.ChIP-seq and

(87) PCT国际申请的公布数据

beyond: new and improved methodologies to

W02017/025594 EN 2017.02.16

detect and characterize protein-DNA

(73) 专利权人 分子医学研究中心责任有限公司

interactions.《Nature Reviews Genetics》

地址 奥地利维也纳

.2012,第13卷(第12期),第840-852页.

(72) 发明人 C·博克 C·施米德尔

审查员 白晓岩

权利要求书2页 说明书45页

序列表7页 附图35页

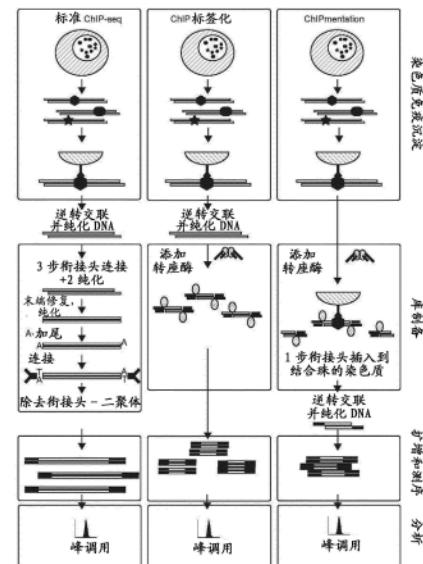
(54) 发明名称

研究核酸的方法

(57) 摘要

CN 108368540 B

本发明提供了用于制备测序文库并研究涉及核酸的分子相互作用的新方法。特别地,本发明涉及用于制备测序文库的方法,所述方法包括将结合染色质的活性剂添加至包含核酸的样品中;分离由所述活性剂结合的染色质;向分离的染色质中添加转座酶;从染色质中分离核酸;并获得测序文库。此外,本发明涉及用于绘制涉及核酸的分子相互作用的方法,所述方法包括将结合染色质的活性剂添加至包含核酸的样品中;分离由所述活性剂结合的染色质;向分离的染色质中添加转座酶;从染色质中分离核酸;扩增核酸;为扩增核酸的测序;并鉴定分子相互作用。



1. 制备测序文库的方法,所述方法包括:

- (a) 将结合染色质的活性剂添加至包含核酸的样品中;
- (b) 分离由所述活性剂结合的染色质;
- (c) 向步骤(b)的分离的染色质中添加转座酶;
- (d) 从染色质中分离核酸;和
- (e) 获得测序文库。

2. 制备测序文库的方法,所述方法包括:

- (a) 将结合染色质的抗体添加至包含核酸的样品中;
- (b) 分离由所述抗体结合的染色质;
- (c) 向步骤(b)的结合和分离的染色质中添加转座酶;
- (d) 从染色质中分离核酸;和
- (e) 获得测序文库。

3. 一种用于非疾病诊断目的的、绘制涉及核酸的分子相互作用的方法,所述方法包括:

- (a) 将结合染色质的活性剂添加至包含核酸的样品中;
- (b) 分离由所述活性剂结合的染色质;
- (c) 向步骤(b)的分离的染色质中添加转座酶;
- (d) 从染色质中分离核酸;
- (e) 扩增核酸;
- (f) 对扩增的核酸测序;和
- (g) 鉴定分子相互作用。

4. 一种用于非疾病诊断目的的、绘制涉及核酸的分子相互作用的方法,所述方法包括:

- (a) 将结合染色质的抗体添加至包含核酸的样品中;
- (b) 分离由所述抗体结合的染色质;
- (c) 向步骤(b)的结合和分离的染色质中添加转座酶;
- (d) 从染色质中分离核酸;
- (e) 扩增核酸;
- (f) 对扩增的核酸测序;和
- (g) 鉴定分子相互作用。

5. 如权利要求1至4所述的方法,其中所述包含核酸的样品已经通过以下步骤制备

(i) 培养并收获细胞,所述细胞不是人胚胎细胞,也不来源于人胚胎阶段;

(ii) 固定细胞;

(iii) 裂解细胞并由此获得包含核酸的第一样品;和

(iv) 对第一样品进行超声处理并由此获得包含核酸的第二样品,其中所述第二样品将用于权利要求1或2的方法中。

6. 如权利要求5所述的方法,其中所述方法还包括逆转在固定细胞期间引入的交联的步骤。

7. 如权利要求1至4中任一项所述的方法,其中所述核酸是DNA。

8. 如权利要求5所述的方法,其中所述核酸是DNA。

9. 如权利要求5所述的方法,其中所述细胞包含核酸-蛋白质复合物。

10. 如权利要求9所述的方法,其中所述细胞是人类细胞、动物细胞、细菌细胞、酵母细胞、植物细胞或病毒。
11. 如权利要求9所述的方法,其中所述细胞是古细菌细胞。
12. 如权利要求10所述的方法,其中所述人类细胞或动物细胞是患病细胞或未患病细胞。
13. 如权利要求10所述的方法,其中所述人类细胞或动物细胞是来自患病或未患病组织的细胞。
14. 如权利要求10所述的方法,其中所述人类细胞或动物细胞是癌细胞、免疫细胞、血细胞或干细胞。
15. 如权利要求14所述的方法,其中所述癌细胞是来自实体癌或血液癌。
16. 如权利要求15所述的方法,其中所述血液癌是白血病。
17. 如权利要求15所述的方法,其中所述实体癌是肿瘤。
18. 如权利要求10的方法,其中所述动物属于稀有物种、濒危物种和/或是模式生物。
19. 如权利要求10的方法,其中所述细胞是胚胎细胞。
20. 如权利要求5所述的方法,其中步骤(i i)包括添加化学物质和/或物理手段。
21. 如权利要求20所述的方法,其中所述化学物质是甲醛或多聚甲醛。
22. 如权利要求20所述的方法,其中所述物理手段包括紫外线或激光。
23. 如权利要求5所述的方法,其中步骤(iv)包括超声处理直到大多数核酸片段长度为20-5000个碱基对。
24. 如权利要求5所述的方法,其中步骤(iv)包括超声处理直到大多数核酸片段长度为200-300个碱基对。
25. 如权利要求1所述的方法,其中结合染色质的活性剂是抗体或化学物质。
26. 如权利要求2或25所述的方法,其中所述抗体与组蛋白、转录因子或与组蛋白和/或转录因子结合的蛋白质特异性结合。
27. 如权利要求26的方法,其中与组蛋白和/或转录因子结合的蛋白质是核酸重塑蛋白质或染色质修饰酶。
28. 如权利要求26所述的方法,其中所述组蛋白是H3.3、H2A.Z、CENP-A、H3.2、H3.3A、H3.3B、H4或H3.1。
29. 如权利要求26的方法,其中所述组蛋白是修饰的组蛋白,其中所述修饰是甲基化、乙酰化、丙酰化、丁酰化、巴豆酰化、2-羟基异丁酰化、丙二酰化、琥珀酰化和/或核糖基化。
30. 如权利要求29的方法,其中所述修饰的组蛋白是H3K4me1/2/3、H2BK5me1、H3K27me1/2/3、H3K9me1/2/3、H4K20me1、H3K79me1、H3K36me3、H2AK5ac、H2AK9ac、H2BK5ac、H2BK12ac、H2BK20ac、H2BK120ac、H3K4ac、H3K9ac、H3K14ac、H3K18ac、H3K23ac、H3K27ac、H3K36ac、H4K5ac、H4K8ac、H4K12ac、H4K16ac、H4K91ac、H2Aub或H2Bub。
31. 如权利要求25所述的方法,其中所述化学物质是药物或工具化合物。
32. 如权利要求25或31所述的方法,其中所述化学物质是生物素化的。
33. 如权利要求1至4所述的方法,其中所述转座酶包括随机DNA序列标签或限定的DNA序列标签。
34. 如权利要求33所述的方法,其中所述转座酶是Tn5转座酶。

研究核酸的方法

[0001] 本发明提供了用于制备测序库并研究涉及核酸的分子相互作用的新方法。特别地,本发明涉及用于制备测序库的方法,所述方法包括将结合染色质的活性剂添加至包含核酸的样品中;分离由所述活性剂结合的染色质;向分离的染色质中添加转座酶;从染色质中分离核酸;以及获得测序库。此外,本发明涉及用于绘制涉及核酸的分子相互作用的方法,所述方法包括将结合染色质的活性剂添加至包含核酸的样品中;分离由所述活性剂结合的染色质;向分离的染色质中添加转座酶;从染色质中分离核酸;核酸扩增;扩增核酸的测序;以及鉴定分子相互作用。

[0002] 核酸与其他化学物质和/或生物分子之间相互作用的知识对研究和医学具有很高的兴趣。研究蛋白质-核酸相互作用的熟知方法是染色质免疫沉淀(ChIP),任选地随后进行大规模平行测序(ChIP-seq)。例如,在染色质中研究与核酸和/或蛋白质的小分子相互作用的方法是Chem-Seq,在下面进一步描述。

[0003] ChIP方法允许研究全基因组DNA-蛋白质相互作用。它对理解染色质组织、组蛋白修饰以及转录因子结合模式(使用X-ChIP)及其对健康和疾病中基因调控的影响作出了重大贡献;参见例如Nature (2012) 489,第 57-74页或Ernst等人(2011) Nature 473,第43-49页。然而,ChIP仍然是相对乏味的方案,特别是当应用于低输入样品时(参见例如Greenleaf, W.J。(2014) Methods)。为了从ChIP制备用于下一代测序(=制备库)的核酸,经典方法包括几个费力的步骤:(i)末端修复纯化的DNA序列以产生具有磷酸化3'端的平端双链DNA片段;(ii)添加A突出端;(iii)将具有互补的T突出端的衔接头连接到具有A突出端的双链末端修复的ChIP-DNA片段。衔接头允许扩增DNA片段,其确保用于质量控制和随后测序的足够量的片段,并且还通过引入用于簇生成的流式细胞末端和条形码序列来准备用于测序程序的片段以进行多重测序实验。经典的方法有几个局限性:(i)通常需要5-10ng的输入材料来产生在低细胞量中不能从ChIP中回收的库。因此,ChIP-seq实验的推荐细胞量在10⁶个细胞的范围内。(ii)库程序依赖于几种酶促反应和DNA纯化,这使得库生成是相对费力的程序。不完善的酶反应以及DNA纯化也降低了回收的库片段的量,这解释了高输入要求。(iii)衔接头可以自连接并且需要从扩增和测序中排除。因此,需要选择大小以针对过量衔接头和衔接头二聚体进行选择。作为衔接头连接的替代方案,开发了其他通过DNA片段的逆转录产生ChIP-seq库的方案;参见Clontech, Mountain View, CA的ChIP Seq Kit。此外,开发了用于低起始材料量的ChIP-seq方案,例如ichIP (Lara-Astiaso等人(2014) Science 345, pp.943-9),linDA (Shankaranarayanan等人(2011) Nature Methods 8,第565-7页)和载体辅助的ChIP (Zwart等人(2013) BMC Genomics 14, 232或Jakobsen等(2015) BMC Genomics 16, 46),但是这些方案需要额外的试剂、动手操作或需要汇集许多样品,这使得它们昂贵、耗时和/或不灵活(参见图1以概述现有技术方法和已知的缺点)。

[0004] 研究小分子与其染色质中蛋白质/核酸靶标相互作用的另一种方法是Chem-seq。该方法采用与大规模平行DNA测序相结合的化学亲和捕获来鉴定小分子与其靶蛋白或DNA相互作用的基因组位点。它首先由Anders等人在Nature Biotechnology (2013), 32 (1), 第92-6页中描述。

[0005] 最近描述了用于库制备核酸的另一种方法。该方法利用了用于DNA 的同时片段化和衔接头标记(“标签化”)的超活性Tn5转座酶的发展(参见Adey等人(2010) *Genome Biol* 11, R119)。它使用转座酶,其预装了测序兼容的衔接头。转座酶将其衔接头负载整合到DNA中,同时将其片段化。仅需要少量转座酶来产生基因组DNA库(Adey等人(2010) *Genome Biol* 11, R119),用于DNA甲基化分析的亚硫酸氢盐转化的DNA (Wang, Q. 等人(2013) *Nature Protocols* 8, 2022-2032),RNA-seq cDNA 或其他核酸导入测序就绪库(Picelli S. 等人(2014) *Genome Res* vol. 24 (12) 第2033-2040页)。在上述Pirelli出版物中,从分离的cDNA样品制备库。然而,纯化的核酸的标签化反应对转座酶与核酸的比率变化极为敏感,因为在标签化反应中尤其转座酶的量决定了核酸片段的最终大小分布。因此,所得方法需要进行工作和成本密集型实验来确定转座酶和核酸的合适比率,以实现用于下一代测序或其他下游应用的期望的核酸大小分布。在某些情况下,核酸片段分布和丰度的测定甚至不可行,因此根据尤其在Furey 等人(2012) *Nature Reviews Genetics* 13 (12),第840-852页综述的现有技术申请中,不可能找到转座酶与核酸的正确比率来制备测序库。此外,向细胞核添加转座酶可恢复开放染色质的区域,并且在基因组的调控区域 中提供核小体定位以及转录因子足迹的信息(Buenrostro JD. 等人(2013) *Nat Meth*, 第10卷 (12) 第1213-1218页)。

[0006] 然而,没有系统地将转座酶描述为继ChIP或Chem-Seq之后适合用于从核酸产生测序库。相反,讨论了这种方法的潜在缺点。如在 W02013/078470或W02014/205296中所述,这些缺点是由ChIP之后对 纯化的DNA进行标签化而产生的。因此,ChIP和标签化组合的主要缺点 是:(1)已经超声处理成小片段(200-700bp)的ChIPed DNA整体上进一步片段化。因此,标签化可导致达到最低约40bp的最小尺寸的非常小 的库片段(Adey等人(2010) *Genome Biol* 11, R119),其难以测序,因为推荐150bp至200bp片段作为Illumina测序的最小长度;(2)通过标 签化的进一步片段化可能产生每个最初沉淀片段的多个测序读段,这可能 妨碍下游分析。作为实例,与300bp片段相比,600bp免疫沉淀DNA片 段可以产生两倍量的库片段,由此人工增加600bp区域中的相对读段量。这在分析ChIP-seq数据时,对于正确的峰值调用可能是有问题的;(3) 使用纯化的ChIP DNA通过标签化来产生测序库的方法是不方便的,因为 需要正确的大小测定和DNA定量来建立标签化反应。由于样品和DNA量之间的超声处理可能有所不同,通常根据用于IP的抗体在数量级上变化,需要在标签化之前针对每个ChIP样品确定这些参数;(4) 使用纯化的 ChIP DNA通过标签化来产生测序库的方法通常也是不可能的,因为ChIP DNA量可能太低而不能进行稳健的定量和大小确定,这两者都是建立稳健 的标签化反应的关键参数;和/或(5) 纯化的ChIP DNA的标签化不保留免疫沉淀靶区域处局部染色质结构的潜在信息。在W02013/078470中描述 的另一种称为TAM-ChIP的方法利用与针对ChIP的抗体缀合的Tn5转座 酶。也就是说,也如W02014/190214 中所述,TAM-ChIP的限制因素是 使用必须在应用之前产生的抗体-寡核苷酸缀合物的限制。这防止了临时使 用主要用于研究蛋白质-DNA相互作用与染色质免疫沉淀的市售抗体。即使在TAM-ChIP中使用第二抗体-寡核苷酸缀合物来克服上述限制,也需要使用两组抗体,这增加了测定的复杂性,同时由于使用通常在ChIP等 应用程序中不正常使用的二抗而增加了成本。因此,如W02013/078470 中所述,TAM-ChIP需要抗体-寡核苷酸-转座酶-复合物的大量优化以输入 染色质比率。由于招募到其识别位点的抗体-寡核苷酸-转座酶复合

物的量 决定了最终的库大小,并且由于识别位点的数量可以依赖于靶抗原而从几百到几十万变化,所以抗体-寡核苷酸-转座酶缀合物的比率必须针对每种 特异性抗体-转座酶缀合物评估以输入染色质。因此,需要一种对转座酶与 输入染色质的定量不敏感的更稳健的方法。TAM-ChIP的另一个缺点是标 签化反应的效率,其范围为0.5%-5%,这是因为只有一个转座体可以被招 募到每个靶抗原的事实。因此,观察到核酸向测序库片段的低转化并可能 增加输入要求。因此,为了在大多数核酸片段的旁边引入测序衔接头,需 要一种可以使用过量转座酶来标记靠近靶位点的几个位点上的核酸的方 法。最后,TAM-ChIP需要大量的 输入染色质。特别是,产生与标准 ChIP-seq可比的测序结果的成功的测序库制备只能用10 μ g输入染色质证 明,这对应于~1.5Mio细胞。因此,调查低丰度和罕见细胞类型,包括原发性患者肿瘤细胞或白血病细胞,来自活检的原发性患者细胞,小群体的 造血细胞类型,从发育模型或胚胎学研究获得的低丰度细胞群是不可行的。因此,稳健地使得使用低输入量,即较少细胞数作为输入来研究分子相互 作用的方法是合乎需要的。

[0007] 鉴于上述情况,需要用于制备测序库和/或绘制涉及核酸的分子相互作 用的改进方法,所述方法与输入量无关并且更快、更容易和/或更便宜。

[0008] 因此,本发明涉及以下条目:

[0009] 1.制备测序库的方法,所述方法包括:

[0010] (a) 将结合染色质的活性剂添加至包含核酸的样品中;

[0011] (b) 分离由所述活性剂结合的染色质;

[0012] (c) 向步骤 (b) 的分离的染色质中添加转座酶;

[0013] (d) 从染色质中分离核酸;和

[0014] (e) 获得测序库。

[0015] 2.制备测序库的方法,所述方法包括:

[0016] (a) 将结合染色质的抗体添加至包含核酸的样品中;

[0017] (b) 分离由所述抗体结合的染色质;

[0018] (c) 向步骤 (b) 的结合和分离的染色质中添加转座酶;

[0019] (d) 从染色质中分离核酸;和

[0020] (e) 获得测序库。

[0021] 3.绘制涉及核酸的分子相互作用的方法,所述方法包括:

[0022] (a) 将结合染色质的活性剂添加至包含核酸的样品中;

[0023] (b) 分离由所述活性剂结合的染色质;

[0024] (c) 向步骤 (b) 的分离的染色质中添加转座酶;

[0025] (d) 从染色质中分离核酸;

[0026] (e) 扩增核酸;

[0027] (f) 对扩增核酸的测序;和

[0028] (g) 鉴定分子相互作用。

[0029] 4.绘制涉及核酸的分子相互作用的方法,所述方法包括:

[0030] (a) 将结合染色质的抗体添加至包含核酸的样品中;

[0031] (b) 分离由所述抗体结合的染色质;

[0032] (c) 向步骤 (b) 的结合和分离的染色质中添加转座酶;

- [0033] (d) 从染色质中分离核酸；
- [0034] (e) 扩增核酸；
- [0035] (f) 对扩增核酸的测序；和
- [0036] (g) 鉴定分子相互作用。
- [0037] 5. 如条目1至4的方法，其中包含核酸的样品已经通过以下方法制备
- [0038] (i) 培养并收获细胞；
- [0039] (ii) 固定细胞；
- [0040] (iii) 裂解细胞并由此获得包含核酸的第一样品；和
- [0041] (iv) 对第一样品进行超声处理并由此获得包含核酸的第二样品，其中所述第二样品将用于条目1或2的方法中。
- [0042] 6. 如条目1至5的方法，其中方法还包括反转在固定细胞期间引入的交联的步骤。
- [0043] 7. 如条目1至6中任一项的方法，其中核酸是DNA。
- [0044] 8. 如条目5至7中任一项的方法，其中细胞包含核酸-蛋白质复合物。
- [0045] 9. 如条目8的方法，其中细胞是人类细胞、动物细胞、细菌细胞、酵母细胞、古细菌细胞、植物细胞或病毒。
- [0046] 10. 如条目9的方法，其中所述人类细胞或动物细胞是患病细胞或未患病细胞或来自患病或未患病组织的细胞。
- [0047] 11. 如条目9或10的方法，其中所述人类细胞或动物细胞是癌细胞、免疫细胞、血细胞或干细胞。
- [0048] 12. 如条目11的方法，其中癌症是实体癌或血液癌。
- [0049] 13. 如条目12的方法，其中血液癌是白血病。
- [0050] 14. 如条目12的方法，其中实体癌是肿瘤。
- [0051] 15. 如条目9、10或11的方法，其中动物属于稀有物种、濒危物种和/或是模式生物。
- [0052] 16. 如条目9、10或11的方法，其中细胞是胚胎细胞。
- [0053] 17. 如条目5的方法，其中步骤(ii)包括添加化学物质和/或物理手段。
- [0054] 18. 如条目17的方法，其中化学物质是甲醛或多聚甲醛。
- [0055] 19. 如条目17的方法，其中物理手段包括紫外线或激光。
- [0056] 20. 如条目5的方法，其中步骤(iv)包括超声处理直到大多数核酸片段长度为20-5000，优选200-300个碱基对。
- [0057] 21. 如条目1或2的方法，其中结合染色质的活性剂是抗体或化学物质。
- [0058] 22. 如条目21的方法，其中抗体特异性结合组蛋白、转录因子或与组蛋白和/或转录因子结合的蛋白质。
- [0059] 23. 如条目22的方法，其中与组蛋白和/或转录因子结合的蛋白质是核酸重塑蛋白或染色质修饰酶。
- [0060] 24. 如条目22或23的方法，其中组蛋白是H3.3、H2A.Z、CENP-A、H3.2、H3.3A、H3.3B、H4或H3.1。
- [0061] 25. 如条目22、23或24的方法，其中组蛋白是修饰的组蛋白，其中所述修饰是甲基化、乙酰化、丙酰化、丁酰化、巴豆酰化、2-羟基异丁酰化、丙二酰化、琥珀酰化和/或核糖基

化。

[0062] 26. 如条目25的方法,其中修饰的组蛋白是H3K4me1/2/3、H2BK5me1、H3K27me1/2/3、H3K9me1/2/3、H4K20me1、H3K79me1、H3K36me3、H2AK5ac、H2AK9ac、H2BK5ac、H2BK12ac、H2BK20ac、H2BK120ac、H3K4ac、H3K9ac、H3K14ac、H3K18ac、H3K23ac、H3K27ac、H3K36ac、H4K5ac、H4K8ac、H4K12ac、H4K16ac、H4K91ac、H2Aub 或H2Bub。

[0063] 27. 如条目21的方法,其中化学物质是药物或工具化合物。

[0064] 28. 如条目21或27的方法,其中化学物质是生物素化的。

[0065] 29. 如条目1至4的方法,其中转座酶包括随机DNA序列标签或限定 的DNA序列标签。

[0066] 30. 如条目29的方法,其中转座酶是Tn5转座酶。

[0067] 因此,本发明提供了用于制备测序库的方法,所述方法包括将结合染色质的活性剂添加至包含核酸的样品中;分离由所述活性剂结合的染色质;向分离的染色质中添加转座酶;从染色质中分离核酸;并获得测序库。

[0068] 本文描述了其他实施方式并在科学部分中进行了举例说明。附图提供了对本发明的说明。然而实施例中以及附图中所示的实验数据不被认为是限制性的。其中包含的技术信息构成本发明的一部分。

[0069] 因此,本发明提供了用于制备测序库的方法和用于绘制涉及核酸,特别是DNA的分子相互作用的方法。从附加实施例中显而易见,本文提供的方法通过组合将结合染色质的活性剂添加至样品中,分离结合的染色质 并以特定的顺序向分离的染色质中添加转座酶的步骤,特别包括测序库的 制备或涉及核酸(特别是DNA)的分子相互作用的绘制。因此,本发明提供了制备测序库的方法,所述方法包括将结合染色质的活性剂添加至包含 核酸的样品中;分离由所述活性剂结合的染色质;向分离的染色质中添加 转座酶;从染色质中 分离核酸;并获得测序库。在分离结合的染色质之后 加入转座酶。优选的是核酸是DNA。本发明的用于绘制涉及核酸的分子相 互作用的方法包括将结合染色质的活性剂添加至包含核酸的样品中;分离 由所述活性剂结合的染色质;向分离的染色质中添加转座酶;从染色质中 分离核酸;核酸扩增;扩增核酸的测序;并鉴定分子相互作用。包含核酸 的样品可以是原代细胞样品或通过培养方法获得的样品。在通过培养方法 获得样品的情况下,优选该方法进一步包括培养并收获细胞;固定细胞; 裂解细胞并由此获得包含核酸的第一样品;以及超声处理第一样品并由此 获得包含核酸的第二样品。所述第二样品优先用于本发明的方法中。此外 优选的是,所述核酸是DNA,特别是双链DNA。当包含核酸的样品是原 代细胞样品时,本发明的方法优先进一步包括固定细胞;裂解细胞并由此 获得包含核酸的第一样品;以及超声处理第一样品并由此获得包含核酸的 第二样品。

[0070] 因此,令人惊讶且出人意料地发现,直接在由染色质特异性活性剂结 合并且在稳健的一步反应中从未结合的染色质中分离出的染色质上进行片 段化导致非常稳健的通用方案,其更快、更便宜、更容易、更稳健、并且 与现有技术中包含的方法的低输入样品更好地兼容。因此,优先在分离由 结合染色质的活性剂,特别是抗体或化学物质,所结合的染色质后添加转 座酶。因此,本发明优先涉及一种方法,其包括将结合染色质的活性剂添 加至包含核酸的样品中;分离由所述活性剂结合的染色质;在分离与结合 染色质的活性剂结合的染色质之后,将转座酶添加至分离的染色质中;从 染色质中分离核酸;并获得测序库。优

选的是核酸是DNA。本发明的用于 绘制涉及核酸的分子相互作用的方法优选地包括将结合染色质的活性剂添加至包含核酸的样品中；分离由所述活性剂结合的染色质；在分离与结合 染色质的活性剂结合的染色质之后，将转座酶添加至分离的染色质中；从 染色质中分离核酸；核酸扩增；扩增核酸的测序；并鉴定分子相互作用。

[0071] 如图2所示，标准ChIP-seq方案包括固定细胞、细胞裂解、染色质超 声处理和用与珠结合的特异性抗体免疫沉淀的步骤。ChIP DNA的纯化之后是逆转交联，然后将其在包括末端修复、纯化、A-加尾、衔接头连接和 大小选择的多步骤程序中进行库制备。在标准 ChIP-seq方案的第一次改进 中，发现了称为ChIP-标签化的方法。在ChIP标签化中，纯化的 ChIP DNA 用于基于标签化的库制备(参见图11)。该方法对不同DNA浓度敏感，因为纯化的DNA的标签化对标签化酶与DNA的比率敏感，并且DNA浓 度在ChIP-seq的许多应用中可以高度变化且太低而不能量化。

[0072] 在这方面，令人惊讶且出乎意料地发现，在使用标签化分离活性剂结 合的染色质，例如使用衔接头负载的Tn5转座酶后，在单个步骤中引入测 序衔接头时，所得方法对不同的DNA浓度不敏感。

[0073] 因此，通过直接对与未结合的染色质分离的活性剂结合的染色质进行 标签化，实现本发明方法的改善的稳健性，其中蛋白质保护核酸免于过度 的标签化。就库的大小分布 (图3)、测序读段的大小分布 (图4b)、作 图性能 (图4c)、跟踪质量 (图4d) 和一致性 (图4e) 而言，本发明方 法的得到的方案(例如如图2中右侧和4a所示) 证明在标签化酶浓度25 倍差异上高度稳健。本发明的方法也得到改进，因为它们不引起测序衔接 头二聚体，并且在标准ChIP方案之后不需要超出标准清除的任何核酸纯 化步骤。

[0074] 本文提供的方法的上述优点使它们相对于现有技术中已知的方法具有 优越性。具体而言，本文提供的方法更灵活、更便宜、更稳健、需要较少 量的样品输入并且允许获得与现有技术中已知的方法相关的附加信息，特 别是如W02014/190214中提供的TAM-ChIP。具体而言，在TAM-ChIP 中，转座酶介导的测序库制备需要特异性抗体-寡核苷酸-缀合物或 抗体-寡 核苷酸-转座酶缀合物。在本发明中，可以特别使用商业上可获得的 ChIP-seq抗体而无需费力和费用高昂的缀合反应。另外，如 W02014/190214中所述的TAM-ChIP方案需要抗体-寡核苷酸-转座酶-缀 合物与输入染色质的比率的广泛优化，而本发明的方法对于 不同的转座酶 与染色质比率是稳健的。

[0075] 如本文提供的方法(通过例如特异性抗体)，在感兴趣的染色质分离 之后添加转 座酶相对于本领域已知的方法具有进一步的益处。具体而言， TAM-ChIP需要确定抗体-寡核苷酸-缀合物的优化比率，以将它们靶向至 它们在染色质中的识别位点。因此，只有紧邻 识别位点的核酸可以被标记， 导致使用转座酶时相对较低的标签频率 (0.5-5%)。相反，本发明的方法 允许标签化，而与用于分离感兴趣的染色质的活性剂无关。因此，就本发 明方法关于转座酶-染色质比率的稳健性而言，在本发明的方法中可以使用 过量的转座酶以使 测序库生成的效率最大化。与本领域已知的方法相比，这一成就显著降低了本发明方法中的 输入要求。

[0076] 此外，在分离感兴趣的染色质之后添加转座酶出乎意料地允许使用少 量转座酶 的有效的测序库制备。这是由于分离感兴趣的染色质时标签化反 应的非特异性模板染色质的存在减少，而其余部分被丢弃。与本领域已知 的方法尤其是W0 2014/205296和W0

2014/190214中描述的方法相比，所需转座酶量的减少是本发明方法的显著成本优势。

[0077] 关于本领域已知的替代方法，例如Picelli等人描述的使用纯化的DNA 作为模板进行测序库制备的方法，使用染色质作为模板的那些，如在本发明的方法中那样，保留了局部染色质背景的高分辨率结构信息；参见例如 图9。

[0078] 因此，本发明的方法首次允许构建和扩增来自染色质的测序库，以研究分子间相互作用，而无需预先纯化或提取核酸，特别是如本文提供的超 快速方法。此外，本发明的方法首次通过提供用于组蛋白标记和转录因子 的全基因组作图的改进的低输入工作流程，实际上允许疾病特别是癌症、组群和临床研究中的大规模染色质可及性作图。鉴于本文提供的染色质谱 分析测定法足够快且直接用于临床测序实验室，染色质失调现在作为例如用于分层癌症治疗的生物标志物的来源是可跟踪的；另参见Rendeiro等 人，(2016) Nature Comm. 7，文章编号11938。

[0079] 本发明的方法验证了五个示例性组蛋白标记 (H3K4me3、H3K27ac、 H3K4me1、 H3K36me3和H3K27me3) 和四种转录因子 (PU.1、CTCF、 GATA1和REST)。在所有情况下，与标准 ChIP-Seq相比，本发明的方 法显示出类似的数据有效性(图5f)。然而，本发明的方法允许细胞输入 的显著减少。特别是，H3K4me3和H3K27me3以及来自10k和100k细胞 的GATA1和 CTCF分别获得了高质量数据，没有任何预扩增(图4f)。就此而言，对于标准ChIP-Seq方案，推荐至少500k/10M细胞 (分别为组 蛋白修饰/转录因子) 作为输入，而使用本发明方法的数据分别使用10k或 100k细胞获得，同时使用1k细胞是可行的。就此而言，使用本领域已知的标准方法从10k细胞纯化DNA后，10k细胞产生约25ng染色质，如通 过**Qubit®**荧光计测定的。因此，当分别与传统ChIP-seq对组蛋白修饰或 转录因子的输入要求的推荐相比时，本发明的新颖和创造性方法允许细胞 输入减少至少50-100倍。因此，与使用纯化的核酸或 cDNA作为模板用于 测序文库制备的标准ChIP-seq和已知转座酶介导的测序文库制备相反，本 文提供的方法可以使用染色质作为模板以产生测序文库。

[0080] 观察到标准ChIP方案和本发明方法之间的良好相关性和峰重叠，并 且还观察到用任一种方法制备的生物学重复样品和低输入样品之间的良好 相关性和峰重叠(图4g、 4h、5、6和7)。总共，使用本发明的方法对52 个文库进行测序，具有标准ChIP-seq的24个文库(左图2) 和具有ChIP- 标签化的9个文库(图2，中心)，并观察到比对率高于95%并且在大多 数实验中独特的读段率约为90% (图8)。鉴于标签化是直接对染色质进 行的，研究了标签化事件的分布是否受局部染色质结构的影响。此外，观 察到提示转录因子足迹(图9a) 和核小体结合(图9b,9c) 的模式。利用 定制的归一化和分析算法，可以推断转录因子足迹，并 且除了常规 ChIP-seq读段之外，预计还可以从根据本发明的方法获得的数据推断核小 体稳定性和/或核小体定位的区域。

[0081] 也在所附实施例中示出的结果将本发明的方法建立为标准ChIP-seq 的通用改进，其更快(10-20分钟，不包括ChIP和最终文库扩增步骤，图 2i)、更成本-有效(图10)、更好地兼容低输入样品(图6) 并且更容易。发现本发明的方法在宽范围的细胞数量和酶浓度，结合染色质的各种活性 剂和不同的ChIP方案上非常稳健，这最小化了对方案适应和优化的需要。因此，本发明的方法非常适合用于测序文库制备和/或涉及核酸(其涉及大 量样品)的分子相互作用的绘制，即作为高通量方法，集中于稀少细胞群 体和/或从快速、经济高效和稳健的实验工作流程中获益。就这一点而言， 也设想本发明的方法以高通量方

式、自动化方式和/或并行方式使用。因此，可以预期用于促进移液/提高再现性的高通量设备和/或机器人用于执行本发明的方法。本领域技术人员将很好地意识到以高通量、自动化和/或并行方式执行本发明的方法的合适手段。例如，可以在本发明的方法中使用多孔板以平行方式进行多个实验。这种多孔板可以例如具有96、384或1536个孔。多孔板也可以与适用于高通量实验的机器人结合使用。已知的机器人系统允许同时执行多个实验，这减少了时间、成本和/或增加实验数据的可靠性。例如，Scic lone NGS工作站 (P/N SG3-31020-0300, Perkin Elmer) 可用于实现自动化高通量测序样品制备。当以高通量方式进行实验时，优选结合染色质的活性剂，特别是抗体或化学物质附着于磁珠，如下面进一步描述的。磁珠在高通量方法中特别有用，因为它们可以容易地用于从未结合的物质中分离结合的颗粒。还设想本发明的方法与微流体装置组合使用。例如，具有简单微流体室特征的聚(二甲基硅氧烷) (PDMS) 装置可以与本发明的方法组合使用。优选的是，微流体室具有一个入口和一个出口，并且出口具有芯片上气动微阀，其可以通过在端口施加压力而部分关闭。用结合染色质的活性剂(特别是抗体或化学物质)包被的磁珠流入微流体室并形成填充床，而气动微阀部分关闭。超声处理的染色质片段然后流过室并吸附到珠表面上。珠之间的空隙小于2μm，并且在小扩散长度下促进靶染色质片段的快速和高效吸附。然后通过在两种不同的洗涤缓冲液中振荡洗涤来洗涤珠以去除非特异性吸附的染色质片段。最后，珠从室中流出并收集用于芯片外 (off-chip) 处理。这种方法，如Cao等人 (2015) *Nature Methods* (可在线获得) 所述，结合本文提供的新颖和创造性方法允许进一步减少实验时间和成本。也设想这里提供的方法与Drop-Seq的组合，如Macosko等人 (2015) *Cell* 161 (5) pp. 1202-14 和Klein等人 (2015) *Cell* 161 (5) pp. 1187-2201 所述。

[0082] 因此，本发明尤其涉及制备测序文库的方法。特别地，用于制备测序文库的方法包括将结合染色质的活性剂添加至包含核酸的样品中；分离由所述活性剂结合的染色质；向分离的染色质中添加转座酶；从染色质中分离核酸；并获得测序文库。

[0083] 关于其中结合染色质的活性剂是除抗体以外的化学物质的方法，已知大量的小分子配体(包括治疗药物)通过结合与基因组相关的特定蛋白质来引发它们的作用。用全基因组的方式绘制这些化学实体与染色质的整体相互作用可以提供洞察小分子影响细胞功能的机制。Chem-seq可用于调查治疗方式的基因组范围的效应，并了解各种生物环境下药物对核结构的影响。从更广泛的意义上讲，这些方法有助于增强对小分子调节基因组相关蛋白功能和活性的治疗机制的理解。通过鉴定药物的细胞靶点，可以在药物开发的早期阶段更好地了解副作用和毒性的原因，这有助于减少开发中的损耗率。

[0084] Chem-seq依赖于创建感兴趣的小分子的生物素化形式以允许下游亲和捕获的能力。Chem-seq可以在体外或体内进行。

[0085] 在体内Chem-seq过程中，培养基中的培养的细胞可以与研究中的生物素化小分子或DMSO(作为对照)以及甲醛同时处理，以交联DNA、蛋白质和小分子。然后从细胞中提取染色质，超声处理并通过与链霉抗生素蛋白磁珠(其对生物素具有非常高的亲和力)温育来富集含有生物素化的感兴趣的分子的区域。然后将富集的染色质分级从珠上洗脱下来，使交联反转，纯化DNA，产生文库并进行下一代测序。相对于对照在Chem-seq文库中富含的基因组区域与正在研究的小分子相关联。因此，本发明还涉及用于绘制小分子和核酸(特别是DNA)之间的相互作用的体内方法。

[0086] 体外Chem-seq开始于培养基中的培养的细胞用甲醛的交联。然后从 细胞中收获细胞核并提取其染色质。将该提取物在与链霉抗生物素蛋白磁 珠一起温育之前进行超声处理,所述磁珠与我们感兴趣的化合物的生物素 化形式结合。这为感兴趣的小分子提供了与其靶染色质区域相互作用的机 会。然后用磁铁分离这些染色质区域并纯化DNA。从DNA中制备文库并 进行下一代测序,然后进行分析以确定富含我们感兴趣的小分子的区域。

[0087] 因此,本发明还涉及如上所述的用于制备测序文库或绘制包含核酸的 分子相互作用的方法,其组合Chem-Seq方法与标签化。

[0088] 在本发明的含义内,术语“测序文库”是指核酸表示,其中每个核酸可 通过例如使用单个序列标签来鉴定。因此,“获得测序文库”需要能够确保 将特定的衔接头序列添加到待分析的核酸片段的末端的方法。这种核酸的 制剂通常被称为“测序文库”。大多数下一代测序应用需要制备测序文库, 在5'和3'末端具有特定衔接头的核酸。例如,Illumina测序工作流程利用 部分互补的衔接头寡核苷酸,其用于引发PCR扩增,并通过桥接PCR引 入用于簇产生所需的特定核苷酸序列,并促进合成测序反应。因此,用于 制备测序文库的本发明方法的所得测序文库适用于标准测序应用,例如, 如下面进一步描述的下一代测序。

[0089] 本发明含义内的“核酸”是由核苷酸组成的任何长度的聚合物,优选具 有大于约 50个核苷酸的长度。本发明的方法允许制备包含核酸的测序文库 和/或绘制涉及核酸的分子相互作用。包含在本发明方法起始样品中的核酸 优选具有约50至约5000个核苷酸,优选 100至约1000,更优选约200至 约700,甚至更优选200至700,最优选200至300个核苷酸。起始样品不 应与包含在用于本发明方法的细胞中的核酸混淆,该方法包括培养并收获 细胞;固定细胞;裂解细胞;并超声处理。就此而言,超声处理用于核酸 包含的并从细胞获得的片段,从而获得起始样品,也称为第二样品,其中 所述方法包括培养并收获细胞;固定细胞;裂解细胞;并超声处理的附加 步骤。就此而言,“核苷酸”旨在包括不仅包含已知的嘌呤 和嘧啶碱基的那 些部分,而且包含已经被修饰的其他杂环碱基的那些部分。这种修饰包括 甲基化嘌呤或嘧啶、酰化嘌呤或嘧啶、烷基化核糖或其他杂环。另外,术 语“核苷酸”包括含有半抗原或荧光标记,并且不仅可以含有常规的核糖和 脱氧核糖,还可以含有其他糖的那些部分。修饰的核苷或核苷酸还包括对 糖部分的修饰,例如其中一个或多个羟基被卤素原子 或脂族基团取代,被 官能化为醚、胺等。然而,在大多数情况下,用于本发明方法的核酸 将包 含天然存在的嘧啶和嘌呤碱基作为可参与沃森-克里克碱基配对相互作用 的脱氧核 糖核苷酸和核糖核苷酸。天然存在的核苷酸包括鸟嘌呤、胞嘧啶、腺嘌呤、胸腺嘧啶和尿嘧啶 (分别为G、C、A、T和U)。核酸可以是 DNA、RNA或任何其他类型的已知核酸。优选的是核酸 是DNA,特别是 双链DNA。

[0090] 术语“结合染色质的活性剂”包括作为包含作为一个结合伴侣的染色质 的结合复合物的成员的任何活性剂。例如,结合染色质的活性剂可以是多 肽,例如蛋白质或其片段,特别是抗体;核酸,例如寡核苷酸,多核苷酸 等;或小分子,例如化学物质。因此,在一个实 施方案中,结合染色质的 活性剂是具有染色质特异性结合结构域的多肽和/或与染色质结合的其他 分子,特别是其他多肽。例如,结合染色质的活性剂可能具有识别染色质 的甲基-CpG结合结构域 (MBD)。结合染色质的活性剂优选多肽,特别 是与染色质结合的抗体,其中抗体特异性结合染色质,蛋白质,例如与染 色质和/或DNA结合的转录因子或组蛋白。

[0091] 就此而言,本文使用的染色质是细胞中发现的大分子的复合物,包含 DNA、蛋白质

和/或RNA。染色质的主要功能是1) 将DNA包装成更小的 体积以适应细胞,2) 增强DNA大分子以允许有丝分裂,3) 防止DNA损伤,以及4) 控制基因表达和DNA复制。染色质的主要蛋白质组分是使 DNA致密的组蛋白。染色质的结构取决于几个因素。整体结构取决于细胞 周期的阶段。在分裂间期期间,染色质在结构上是松散的,以允许接近转 录和复制DNA的RNA和DNA聚合酶。染色质在分裂间期期间的局部结 构取决于存在于DNA上的基因:活跃转录(“打开”)的DNA编码基因被 更松散地包装并且发现与RNA聚合酶(被称为常染色质) 和转录因子结 合,同时发现编码无活性基因的DNA(“关闭”)与结构蛋白结合并且被更 紧密包装(异染色质)。染色质中结构蛋白的表观遗传化学修饰也改变了 局部染色质结构,特别是通过甲基化和乙酰化对组蛋白的化学修饰;参阅 下文。

[0092] 染色质的基本重复元素是核小体,其通过衔接头DNA的部分相互连 接,比溶液中的纯DNA短得多。除了核心组蛋白之外,还有衔接头组蛋白H1,其与核小体上的DNA链的出口/入口接触。核小体核心颗粒与组蛋白H1一起被称为染色小体。具有约20至60个碱基对的衔接头DNA的核 小体可以在非生理条件下形成大约10nm的“串珠”纤维。核小体按照其在普通DNA包装中的功能所要求的非特异性结合DNA。然而,存在控制核 小体定位的大量DNA序列偏好。这主要是由于不同DNA序列的物理性质 不同所致:例如,腺嘌呤和胸腺嘧啶更有利地被压缩到内部小沟中。这意 味着核小体可以在大约每10个碱基对 (DNA的螺旋重复序列) 的一个位 置优先结合,其中DNA旋转以使位于内部小沟中的A和T碱基的数量最 大化。如本发明所提及的,结合染色质的活性剂可以与染色质,常染色质 或异染色质的任何部分结合。例如,结合染色质的活性剂可与染色质中包 含的DNA、RNA或蛋白质相互作用。特别地,结合染色质的活性剂可与 染色质中包含的组蛋白或转录因子和/或与组蛋白、转录因子或染色质相关 的其他蛋白质相互作用。

[0093] 在这方面,组蛋白是在真核细胞核中发现的高碱性蛋白质,其将DNA 包装并有序化为被称为核小体的结构单元(参见上文)。它们是染色质的 主要蛋白质组分,充当DNA缠绕的线轴,并在基因调控中发挥作用。存 在五种主要的组蛋白家族:H1/H5, H2A, H2B, H3和 H4。组蛋白H2A, H2B, H3和H4被称为核心组蛋白,而组蛋白H1和H5被称为衔接头组 蛋白。两个核心组蛋白每个装配形成一个直径约63埃 (螺线管 (DNA) 样颗粒) 的八聚体核小体核心。147个碱基对的DNA在这个核心颗粒周围 以左手超螺旋转角缠绕约1.65次,产生大约100埃的颗粒。接头组蛋白 H1在DNA的入口和出口位点结合核小体,从而将DNA锁定到位并允许形成更高级的结构。最基本的这种形成是10nm纤维或串珠构象。这涉及 用分离每对核小体 (也称为接头DNA) 的大约50个碱基对的DNA包裹核 小体周围的DNA。高级结构包括30纳米纤维 (形成不规则的锯齿形) 和 100纳米纤维,这些都是正常细胞中的结构。在有丝分裂和减数分裂过程 中,浓缩的染色体通过核小体和其他调节蛋白之间的相互作用进行组装。结合染色质,特别是抗体或化学物质的活性剂可以与组蛋白相互作用,即 它们可以特异性结合组蛋白和/或结合与组蛋白相关的其他多肽和/或化学 物质。结合染色质的活性剂,特别是抗体或化学物质优选与组蛋白直接相 互作用。

[0094] 就此而言,已知的人类组蛋白包括五类H1/H5, H2A, H2B, H3和 H4。H1包括H1F0、 H1FNT、H1F00、H1FX、HIST1H1A、HIST1H1B、 HIST1H1C、HIST1H1D、HIST1H1E和HIST1H1T。H2A类包括H2AFB1、 H2AFB2、H2AFB3、H2AFJ、H2AFV、H2AFX、H2AFY、H2AFY2、 H2AFZ、HIST1H2AA、 HIST1H2AB、HIST1H2AC、HIST1H2AD、 HIST1H2AE、HIST1H2AG、HIST1H2AI、HIST1H2AJ、

HIST1H2AK、HIST1H2AL、HIST1H2AM、HIST2H2AA3和HIST2H2AC。H2B类包括H2BFM、H2BFS、H2BFWT、HIST1H2BA、HIST1H2BB、HIST1H2BC、HIST1H2BD、HIST1H2BE、HIST1H2BF、HIST1H2BG、HIST1H2BH、HIST1H2BI、HIST1H2BJ、HIST1H2BK、HIST1H2BL、HIST1H2BM、HIST1H2BN、HIST1H2BO和HIST2H2BE。H3类包括HISTH3A、HISTH3B、HISTH3C、HISTH3D、HISTH3E、HISTH3F、HISTH3G、HISTH3H、HISTH3I、HISTH3J、HIST2H3C和HIST3H3。H4类包括HIST1H4A、HIST1H4B、HIST1H4C、HIST1H4D、HIST1H4E、HIST1H4F、HIST1H4G、HIST1H4H、HIST1H4I、HIST1H4J、HIST1H4K、HIST1H4L 和HIST4H4。结合染色质的活性剂,特别是抗体或化学物质优选结合H3 类组蛋白,特别是H3.3、H3.2、H3.3A、H3.3B或H3.1。另外,结合染色质的活性剂特别是抗体或化学物质优选结合H4、H2A.Z或CENP-A (后 两者含有组蛋白H3相关的组蛋白折叠)。

[0095] 进一步设想结合染色质的活性剂特别是抗体或化学物质对已知组蛋白 的修饰形式是特异性的。已经描述了巨大的组蛋白修饰目录。组蛋白修饰 对于其他结合蛋白和/或其他化学物质的的基因组翻译和DNA可接近性具有特定的含义和结果。因此,设想本发明的方法可用于鉴定可能经历基因 表达改变的经修饰的组蛋白结合的区域,例如,在患病组织/细胞如癌细胞 中。

[0096] 已知的组蛋白修饰包括甲基化、乙酰化、丙酰化、丁酰化、巴豆酰化、2-羟基异丁酰化、丙二酰化、琥珀酰化和核糖基化。特别是赖氨酸甲基化、精氨酸甲基化、赖氨酸乙酰化、丝氨酸/苏氨酸/酪氨酸磷酸化。在这方面,向赖氨酸添加一个、两个或三个甲基基团对组蛋白的化学反应几乎没有影 响;甲基化使赖氨酸的电荷完整并且添加最少数量的原子,因此空间相互 作用大多不受影响。然而,包含Tudor、chromo或PHD结构域的蛋白质 等能够以极高的灵敏度识别赖氨酸甲基化并区分单、二和三甲基赖氨酸到 达对于一些赖氨酸(例如H4K20)单、二和三-甲基化有不同的含义的程 度。正因为如此,赖氨酸甲基化是非常有用的标记,并且支配已知的组蛋白修饰功能。因此,可以设想结合染色质的活性剂对赖氨酸甲基化的组蛋白和/或识别这种修饰的组蛋白的蛋白,例如,含有Tudor、chromo或PHD 结构域的蛋白质是特异的。关于精氨酸甲基化的组蛋白,适用与上述类似 的推理,即一些蛋白结构域-例如Tudor结构域-可以对甲基精氨酸而不是 甲基赖氨酸具有特异性。已知精氨酸是单甲基化或二甲基化的,甲基化可 以是对称或不对称的,可能具有不同的含义。关于赖氨酸乙酰化,添加乙 酰基对赖氨酸具有主要化学作用,因为它中和正电荷。这减少了组蛋白和 带负电的DNA骨架之间的静电引力,使染色质结构松动;高度乙酰化的 组蛋白形成更易接近的染色质并且倾向于与活性转录相关联。赖氨酸乙酰 化的意义似乎不如甲基化精确,因为组蛋白乙酰转移酶倾向于作用于一个 以上的赖氨酸;据推测这反映了需要改变多种赖氨酸以对染色质结构有显 著影响。因此,还设想结合染色质的活性剂对乙酰化赖氨酸和/或与乙酰化 赖氨酸相互作用的蛋白质是特异性的。除上述之外,组蛋白中包含的丝氨 酸/苏氨酸和/或酪氨酸可通过磷酸化修饰。添加带负电荷的磷酸基团可导致 蛋白质结构发生重大变化,从而导致磷酸化在控制蛋白质功能方面充分表 征的作用。组蛋白磷酸化作为翻译后修饰具有明确的功能,并且已经表征 了诸如BRCT (BRCA1C末端结构域) 的结合结构域。因此,也设想这种 修饰的组蛋白,即通过磷酸化修饰的组蛋白,被结合染色质的活 性剂识别。

[0097] 上述组蛋白的修饰和本领域中描述的进一步修饰对转录的控制有影 响。在这方

面,两种已知的组蛋白修饰与活跃转录特别相关:H3赖氨酸4 (H3K4Me3) 的三甲基化和H3赖氨酸36 (H3K36Me3) 的三甲基化。H3K4Me3发生在活性基因的启动子上并由COMPASS复合物进行。修饰 是活性启动子的极好标记,并且基因启动子上这种组蛋白修饰的水平与该 基因的转录活性广泛相关。这种标记的形成以相当复杂的方式与转录关联 在一起:在基因转录早期, RNA聚合酶II经历从“启动”切换到“延长”, 由RNA聚合酶II C末端结构域(CTD) 磷酸化状态的改变为标志。使CTD 磷酸化的同样的酶也使Rad6复合物磷酸化,这又将泛素标记添加到H2B K123 (哺乳动物中的K120)。H2BK123Ub发生在整个转录区域,但该标记是COMPASS在启动子处使H3K4三甲基化所需的。因此,在本发明的 优选方面,结合染色质的活性剂,特别是抗体或化学物质对H3K4Me3是 特异性的。另一方面,结合染色质的活性剂特别是抗体或化学物质对 H3K36Me3是特异性的。这种三甲基化发生在活性基因体内,并由甲基转移酶Set2沉积。该蛋白与延伸RNA聚合酶II相关,并且H3K36Me3指 示活跃转录的基因。H3K36Me3被Rpd3组蛋白脱乙酰酶复合物识别,该 复合物去除周围组蛋白的乙酰修饰,增加染色质的压缩并抑制假转录。增加的染色质压缩防止转录因子接近DNA,并降低在基因体内启动新的转录 事件的可能性。因此这个过程有助于确保转录不会中断。此外,组蛋白H3 (H3K27ac) 的赖氨酸27的乙酰化存在于作为启动子和增强子的活性调节 元件中。在遗传学上,增强子是可以与蛋白质(激活剂)结合以激活基因 转录的短的(50-1500bp) DNA区域。这些蛋白质通常被称为转录因子。增强子通常是顺式作用的,位于远离基因多达1Mbp (1, 000, 000bp) 处, 并且可以位于起始位点的上游或下游,并且在正向或反向方向上。人类基因组中有数十万种增强子。具体而言,H3K27ac被描述为区分活跃的和稳定的调控元件。H3K27ac在这些元件中的富集是表达相关遗传元件的良好 指标。因此,本发明方法中使用的结合染色质的活性剂特别是抗体或化学 物质可以对H3K27ac是特异的。

[0098] 组蛋白修饰也可能与基因表达的抑制有关。例如,已知H3K27Me3、H3K9Me2/3和H4K20Me3与抑制基因有关。H3K27Me3由多梳复合物 PRC2沉积。它是基因抑制的明确标记,并且可能被其他蛋白质结合而发挥抑制功能。另一种多梳复合物PRC1可以结合H3K27Me3 并添加帮助染 色质压缩的组蛋白修饰H2AK119Ub。H3赖氨酸9 (H3K9Me2/3) 的二甲 基化和三甲基化是异染色质的充分表征的标记,因此与基因抑制密切相关。这同样适用于与异染色质紧密相关的H4K20Me3。该标记通过Suv4-20h 甲基转移酶放置,其至少部分由异染色质蛋白1募集。因此,也考虑在本 发明方法中使用的结合染色质的活性剂特异性结合和与之相关的抑制基因 和/或蛋白质相关的这样的修饰的组蛋白。

[0099] 组蛋白的修饰也在DNA修复和染色体浓缩中发挥作用。例如,标记 DNA损伤位点是组蛋白修饰的重要功能。它还可以保护DNA免受太阳紫 外线辐射的破坏。例如,磷酸化的H2AX (也称为 γ H2AX) 是DNA双链 断裂的标记,并且形成对DNA损伤响应的一部分。H2AX在检测到DNA 双链断裂后早期被磷酸化,并在损伤的任一侧形成延伸多个千碱基的结构 域。 γ H2AX可作为蛋白质MDC1的结合位点,而MDC1又可募集关键的 DNA修复蛋白,因此 γ H2AX是确保基因组稳定性的机制的重要组成部分。此外,H3K56Acx是基因组稳定性所必需的。H3K56被p300/Rtt109复合 物乙酰化,但在DNA损伤部位周围迅速脱乙酰化。H3K56乙酰化也需要 稳定停滞的复制叉,防止危险的复制叉崩溃。H3在丝氨酸10 (磷酸-H3S10) 上的磷酸化与浓缩有关,但H3S10磷酸化也存在于有丝分裂外的某些染色 体位点,例如G2期间的细胞的臂间异染色质中。H3S10磷酸化还与在高 度转录位点由R环形成引起的DNA损伤相关。

H2B在丝氨酸10(酵母)或丝氨酸14(哺乳动物)上的磷酸化也与染色质浓缩有关,只是为了在凋亡过程中介导染色体浓缩的非常不同的目的。该标记不仅仅是凋亡中的晚期旁观者,因为携带该残基突变的酵母对过氧化氢诱导的凋亡细胞死亡具有抗性。

[0100] 因此,本发明方法中使用的结合染色质的活性剂,特别是抗体或化学物质可特异性结合组蛋白、修饰的组蛋白和/或其他因子,特别是多肽如酶,与这些组蛋白和/或修饰的组蛋白相互作用。当结合染色质的活性剂,特别是抗体或化学物质与修饰的组蛋白结合时,结合染色质的活性剂特别是抗体或化学物质优选与H3K4me1/2/3、H2BK5me1、H3K27me1/2/3、H3K9me1/2/3、H4K20me1、H3K79me1、H3K36me3、H2AK5ac、H2AK9ac、H2BK5ac、H2BK12ac、H2BK20ac、H2BK120ac、H3K4ac、H3K9ac、H3K14ac、H3K18ac、H3K23ac、H3K27ac、H3K36ac、H4K5ac、H4K8ac、H4K12ac、H4K16ac、H4K91ac、H2Aub或H2Bub结合。

[0101] 在另一个实施方案中,本发明方法中使用的结合染色质的活性剂特别是抗体或化学物质特异性结合转录因子。转录因子(有时称为序列特异性DNA结合因子)是与特定DNA序列结合的蛋白质,从而控制遗传信息从DNA到信使RNA的转录速率。示例性的转录因子包括但不限于AAF、abl、ADA2、ADA-NF1、AF-1、AFP1、AhR、AIIN3、ALL-1、alpha-CBF、alpha-CP1、alpha-CP2a、alpha-CP2b、alphaHo、alphaH2-alphaH3、Alx-4、aMEF-2、AML1、AML1a、AML1b、AML1c、AML1DeltaN、AML2、AML3、AML3a、AML3b、AMY-1L、A-Myb、ANF、AP-1、AP-2alphaA、AP-2alphaB、AP-2beta、AP-2gamma、AP-3(1)、AP-3(2)、AP-4、AP-5、APC、AR、AREB6、Arnt、Arnt(774M形式)、ARP-1、ATBF1-A、ATBF1-B、ATF、ATF-1、ATF-2、ATF-3、ATF-3deltaZIP、ATF-a、ATF-adelta、ATPF1、Barh11、Barh12、Barx1、Barx2、Bcl-3、BCL-6、BD73、beta-连环蛋白、Bin1、B-Myb、BP1、BP2、brahma、BRCA1、Brn-3a、Brn-3b、Brn-4、BTEB、BTEB2、B-TFIID、C/EBPalpha、C/EBPbeta、C/EBPdelta、CACC结合因子、Cart-1、CBF(4)、CBF(5)、CBP、CCAAT-结合因子、CCF、CCG1、CCK-1a、CCK-1b、CD28RC、cdk2、cdk9、Cdx-1、CDX2、Cdx-4、CFF、Chx10、CLIM1、CLIM2、CNBP、CoS、COUP、CP1、CPIA、CPIC、CP2、CPBP、CPE结合蛋白、CREB、CREB-2、CRE-BP1、CRE-BPalpha、CREMalpha、CRF、Crx、CSBP-1、CTCF、CTF、CTF-1、CTF-2、CTF-3、CTF-5、CTF-7、CUP、CUTL1、Cx、细胞周期蛋白A、细胞周期蛋白T1、细胞周期蛋白T2、细胞周期蛋白T2a、细胞周期蛋白T2b、DAP、DAX1、DB1、DBF4、DBP、DbpA、DbpAv、DbpB、DDB、DDB-1、DDB-2、DEF、deltaCREB、deltaMax、DF-1、DF-2、DF-3、Dlx-1、Dlx-2、Dlx-3、Dlx4(长同工型)、Dlx-4(短同工型)、Dlx-5、Dlx-6、DP-1、DP-2、DSIF、DSIF-p14、DSIF-p160、DTF、DUX1、DUX2、DUX3、DUX4、E、E1-2、E2F、E2F+E4、E2F+p107、E2F-1、E2F-2、E2F-3、E2F-4、E2F-5、E2F-6、E47、E4BP4、E4F、E4F1、E4TF2、EAR2、EBP-80、EC2、EF1、EF-C、EGR1、EGR2、EGR3、EIIaE-A、EIIaE-B、EIIaE-Calp、EIIaE-Cbeta、EivF、EIf-1、EIk-1、Emx-1、Emx-2、En-1、En-2、ENH-bind.prot、ENKTF-1、EPAS1、epsilonF1、ER、Erg-1、Erg-2、ERR1、ERR2、ETF、Ets-1、Ets-1deltaVil、Ets-2、Evx-1、F2F、因子2、因子名称、FBP、f-EBP、FKBP59、FKHL18、FKHRL1P2、Fli-1、Fos、FOXB1、FOXC1、FOXC2、FOXD1、FOXD2、FOXD3、FOXD4、FOXE1、FOXE3、FOXF1、FOXF2、FOXG1a、FOXG1b、FOXG1c、FOXH1、FOXI1、FOXJ1a、FOXJ1b、FOXJ2(长同工型)、FOXJ2(短同工型)、FOXJ3、FOXK1a、FOXK1b、FOXK1c、FOXK1、FOXK1a、FOXK1b、FOXK1c、FOXN1、FOXN2、FOXN3、FOXO1a、FOXO1b、FOXO2、FOXO3a、FOXO3b、FOXO4、FOXP1、FOXP3、Fra-1、Fra-2、FTF、FTS、G因子、G6因子、GABP、GABP-alpha、GABP-beta1、GABP-beta2、GADD153、GAF、gammaCMT、gammaCAC1、gammaCAC2、GATA-1、GATA-2、

GATA-3、GATA-4、GATA-5、GATA-6、Gbx-1、Gbx-2、GCF、GCMa、GCNS、GF1、GLI、GLI3、GR alpha、GR beta、GRF-1、Gsc、Gscl、GT-IC、GT-IIA、GT-IIBalpha、GT-IIBbeta、H1TF1、H1TF2、H2RIIBP、H4TF-1、H4TF-2、HAND1、HAND2、HB9、HDAC1、HDAC2、HDAC3、hDaxx、热诱导因子、HEB、HEB1-p67、HEB1-p94、HEF-1B、HEF-1T、HEF-4C、HEN1、HEN2、Hesx1、Hex、HIF-1、HIF-1alpha、HIF-1beta、HiNF-A、HiNF-B、HiNF-C、HiNF-D、HiNF-D3、HiNF-E、HiNF-P、HIP1、HIV-EP2、H1f、HLTF、HLTF (Met123)、HLX、HMBP、HMG I、HMG I (Y)、HMG Y、HMGI-C、HNF-IA、HNF-IB、HNF-IC、HNF-3、HNF-3alpha、HNF-3beta、HNF-3gamma、HNF4、HNF-4alpha、HNF4alpha1、HNF-4alpha2、HNF-4alpha3、HNF-4alpha4、HNF4gamma、HNF-6alpha、hnRNP K、HOX11、HOXA1、HOXA10、HOXA10 PL2、HOXA1 1、HOXA13、HOXA2、HOXA3、HOXA4、HOXA5、HOXA6、HOXA7、HOXA9A、HOXA9B、HOXB-1、HOXB13、HOXB2、HOXB3、HOXB4、HOXB5、HOXB6、HOXA5、HOXB7、HOXB8、HOXB9、HOXC10、HOXC11、HOXC12、HOXC13、HOXC4、HOXC5、HOXC6、HOXC8、HOXC9、HOXD10、HOXD11、HOXD12、HOXD13、HOXD3、HOXD4、HOXD8、HOXD9、Hp55、Hp65、HPX42B、HrpF、HSF、HSF1 (长)、HSF1 (短)、HSF2、hsp56、Hsp90、IBP-1、ICER-II、ICER-1igamma、ICSBP、Id1、Id1 H'、Id2、Id3、Id3/Heir-1、IF1、IgPE-1、IgPE-2、IgPE-3、IkappaB、IkappaB-alpha、IkappaB-beta、IkappaBR、II-1 RF、IL-6RE-BP、11-6RF、INSAF、IPF1、IRF-1、IRF-2、B、IRX2a、Irx-3、Irx-4、ISGF-1、ISGF-3、ISGF3alpha、ISGF-3gamma、1st-1、ITF、ITF-1、ITF-2、JRF、Jun、JunB、JunD、kappa因子、KBP-1、KER1、KER-1、Kox1、KRF-1、Ku自身抗原、KUP、LBP-1、LBP-1a、LBX1、LCR-F1、LEF-1、LEF-IB、LF-A1、LHX1、LHX2、LHX3a、LHX3b、LHX5、LHX6.1a、LHX6.1b、LIT-1、Lmo1、Lmo2、LMX1A、LMX1B、L-My1 (长形式)、L-My1 (短形式)、L-My2、LSF、LXRalpha、LyF-1、Ly1-1、M因子、Mad1、MASH-1、Max1、Max2、MAZ、MAZ1、MB67、MBF1、MBF2、MBF3、MBP-1 (1)、MBP-1 (2)、MBP-2、MDBP、MEF-2、MEF-2B、MEF-2C (433AA形式)、MEF-2C (465AA形式)、MEF-2C (473M形式)、MEF-2C/delta32 (441AA 形式)、MEF-2D00、MEF-2D0B、MEF-2DAO、MEF-2DAO、MEF-2DAB、MEF-2DA'B、Meis-1、Meis-2a、Meis-2b、Meis-2c、Meis-2d、Meis-2e、Meis3、Meox1、Meox1a、Meox2、MHox (K-2)、Mi、MIF-1、Miz-1、MM-1、MOP3、MR、Msx-1、Msx-2、MTB-Zf、MTF-1、mtTF1、Mxi1、Myb、Myc、Myc 1、Myf-3、Myf-4、Myf-5、Myf-6、MyoD、MZF-1、NC1、NC2、NCX、NELF、NER1、Net、NF I11-a、NF NF NF-1、NF-1A、NF-1B、NF-1X、NF-4FA、NF-4FB、NF-4FC、NF-A、NF-AB、NFAT-1、NF-AT3、NF-Atc、NF-Atp、NF-Atx、NfetaA、NF-CLE0a、NF-CLE0b、NFdeltaE3A、NFdeltaE3B、NFdeltaE3C、NFdeltaE4A、NFdeltaE4B、NFdeltaE4C、Nfe、NF-E、NF-E2、NF-E2p45、NF-E3、NFE-6、NF-Gma、NF-GMb、NF-IL-2A、NF-IL-2B、NF-jun、NF-kappaB、NF-kappaB (-like)、NF-kappaB1、NF-kappaB 1前体、NF-kappaB2、NF-kappaB2 (p49)、NF-kappaB2前体、NF-kappaE1、NF-kappaE2、NF-kappaE3、NF-MHCIa、NF-MHCIb、NF-muE1、NF-muE2、NF-muE3、NF-S、NF-X、NF-X1、NF-X2、NF-X3、NF-Xc、NF-YA、NF-Zc、NF-Zz、NHP-1、NHP-2、NHP3、NHP4、NKX2-5、NKX2B、NKX2C、NKX2G、NKX3A、NKX3A v1、NKX3A v2、NKX3A v3、NKX3A v4、NKX3B、NKX6A、Nmi、N-Myc、N-Oct-2alpha、N-Oct-2beta、N-Oct-3、N-Oct-4、N-Oct-5a、N-Oct-5b、NP-TCII、NR2E3、NR4A2、Nrf1、Nrf-1、Nrf2、NRF-2beta1、NRF-2gamma1、NRL、NRSF形式1、NRSF形式2、NTF、O2、OCA-B、Oct-1、Oct-2、Oct-2.1、Oct-2B、Oct-2C、Oct-4A、Oct4B、Oct-5、Oct-6、Octa-因子、八聚体-结合因子、oct-B2、oct-B3、0tx1、0tx2、0ZF、p107、p130、p28调节物、p300、p38erg、p45、p49erg、-p53、p55、p55erg、p65delta、p67、Pax-1、Pax-2、Pax-3、Pax-3A、Pax-4、Pax-5、Pax-6、

Pax-6/Pd-5a、Pax-7、Pax-8、Pax-8a、Pax-8b、Pax-8c、Pax-8d、Pax-8e、Pax-8f、Pax-9、Pbx-1a、Pbx-1b、Pbx-2、Pbx-3a、Pbx-3b、PC2、PC4、PC5、PEA3、PEBP2alpha、PEBP2beta、Pit-1、PITX1、PITX2、PITX3、PKNOX1、PLZF、PO-B、Pontin52、PPARalpha、PPARbeta、PPARgamma1、PPARgamma2、PPUR、PR、PR A、pRb、PRD1-BF1、PRDI-BFc、Prop-1、PSE1、P-TEFb、PTF、PTFalpha、PTFbeta、PTFdelta、PTFgamma、Pu盒结合因子、Pu盒结合因子(B JA-B)、PU.1、PuF、Pur因子、R1、R2、RAR-alpha1、RAR-beta、RAR-beta2、RAR-gamma、RAR-gamma1、RBP60、RBP-Jkappa、Rel、RelA、RelB、RFX、RFX1、RFX2、RFX3、RFXS、RF-Y、RORalpha1、RORalpha2、RORalpha3、RORbeta、RORgamma、Rox、RPF1、RPGalpha、RREB-1、RSRFC4、RSRFC9、RVF、RXR-alpha、RXR-beta、SAP-1a、SAP1b、SF-1、SHOX2a、SHOX2b、SHOXa、SHOXb、SHP、SIII-p1 10、SIII-p15、SIII-p18、SIM'、Six-1、Six-2、Six-3、Six-4、Six-5、Six-6、SMAD-1、SMAD-2、SMAD-3、SMAD-4、SMAD-5、SOX-11、SOX-12、Sox-4、Sox-5、SOX-9、Sp1、Sp2、Sp3、Sp4、Sph因子、Spi-B、SPIN、SRCAP、SREBP-1a、SREBP-1b、SREBP-1c、SREBP-2、SRE-ZBP、SRF、SRY、SRP1、Staf-50、STAT1alpha、STAT1beta、STAT2、STAT3、STAT4、STAT6、T3R、T3R-alpha1、T3R-alpha2、T3R-beta、TAF(I) 110、TAF(I) 48、TAF(I) 63、TAF(II) 100、TAF(II) 125、TAF(II) 135、TAF(II) 170、TAF(II) 18、TAF(II) 20、TAF(II) 250、TAF(II) 250Delta、TAF(II) 28、TAF(II) 30、TAF(II) 31、TAF(II) 55、TAF(II) 70-alpha、TAF(II) 70-beta、TAF(II) 70-gamma、TAF-I、TAF-II、TAF-L、Tal-1、Tal-1beta、Tal-2、TAR因子、TBP、TBX1A、TBX1B、TBX2、TBX4、TBXS(长同工型)、TBXS(短同工型)、TCF、TCF-1、TCF-1A、TCF-1B、TCF-1C、TCF-1D、TCF-1E、TCF-1F、TCF-1G、TCF-2alpha、TCF-3、TCF-4、TCF-4(K)、TCF-4B、TCF-4E、TCFbeta1、TEF-1、TEF-2、tel、TFE3、TFEB、TFIIIA、TFIIIA-alpha/beta前体、TFIIIA-alpha/beta前体、TFIIIA-gamma、TFIIB、TFIID、TFIIE、TFIIE-alpha、TFIIE-beta、TFIIF、TFIIF-alpha、TFIIF-beta、TFIIH、TFIIH*、TFIIH-CAK、TFIIH-细胞周期蛋白H、TFIIH-ERCC2/CAK、TFIIH-MAT1、TFIIH-M015、TFIIH-p34、TFIIH-p44、TFIIH-p62、TFIIH-p80、TFIIH-p90、TFII-I、Tf-LF1、Tf-LF2、TGIF、TGIF2、TGT3、THRA1、TIF2、TLE1、TLX3、TMF、TR2、TR2-11、TR2-9、TR3、TR4、TRAP、TREB-1、TREB-2、TREB-3、TREF1、TREF2、TRF(2)、TTF-1、TXRE BP、TxREF、UBF、UBP-1、UEF-1、UEF-2、UEF-3、UEF-4、USF1、USF2、USF2b、Vav、Vax-2、VDR、vHNF-1A、vHNF-1B、vHNF-1C、VITF、WSTF、WT1、WT1I、WT1I-KTS、WT1I-de12、WT1-KTS、WT1-de12、X2BP、XBP-1、XW-V、XX、YAF2、YB-1、YEBP、YY1、ZEB、ZF1、ZF2、ZFX、ZHX1、ZIC2、ZID、ZNF 174等。

[0102] 转录因子通过促进(作为激活剂)或阻断(作为阻遏物)向特定基因募集RNA聚合酶(执行将遗传信息从DNA转录为RNA的酶)来单独或与复合体中的其他蛋白质一起执行此功能。因此,结合染色质的活性剂可以与转录因子相互作用,该复合物包含一种或多种转录因子和/或与转录因子结合的蛋白质。转录因子的一个明确特征是它们含有一个或多个DNA结合结构域(DBD),它们附着于与其调节的基因相邻的DNA的特定序列。其他蛋白质如共激活因子、染色质重塑因子、组蛋白乙酰化酶、脱乙酰酶、激酶和甲基化酶,虽然在基因调控中起关键作用,但缺乏DNA结合结构域,因此不被归类为转录因子。然而,本发明方法中使用的结合染色质的活性剂也可以与这些蛋白质相互作用。转录因子与其调控的基因相邻的DNA的增强子或启动子区域结合。取决于转录因子,相邻基因的转录被上调或下调。转录因子使用多种机制来调节基因表达。这些机制包括:稳定或阻断RNA聚合酶与DNA的结合;催化组蛋白的乙酰化或脱乙酰化。转录因子可以直接做到这一点,或招募其他

具有这种催化活性的蛋白。许多转录因子使用两种相反机制中的一种或另一种来调节转录：组蛋白乙酰转移酶 (HAT) 活性-乙酰化组蛋白，这减弱了DNA与组蛋白的结合，这使得DNA更易于转录，从而上调转录；和/或组蛋白脱乙酰基酶 (HDAC) 活性-脱乙酰化组蛋白，这增强了DNA与组蛋白的结合，这使得DNA不易转录，从而下调转录。调控基因表达的机制还包括向转录因子DNA复合物招募辅激活蛋白或共抑制蛋白。

[0103] 由于至少两个原因，许多转录因子具有临床意义：(1) 突变可以与特定疾病相关，并且 (2) 它们可以是药物的靶标。因此，本发明方法中使用的结合染色质的活性剂可能与转录因子相关疾病的诊断和/或治疗有关。例如，由于它们在发育、细胞间信号传导和细胞周期中的重要作用，一些人类疾病与转录因子的突变有关。另外，许多转录因子是肿瘤抑制因子或癌基因，因此，它们的突变或异常调节与癌症相关。已知至少三组转录因子在人类癌症中是重要的：(1) NF- κ B 和 AP-1 家族，(2) STAT 家族和 (3) 类固醇受体。下表显示了与人类疾病有关的其他转录因子：

[0104] 表1

病况	描述	基因座
Rett 综合征	MECP2 转录因子中的突变与 Rett 综合征 (神经发育障碍) 相关	Xq28
糖尿病	称为 MODY (年轻人的成熟型糖尿病) 的罕见糖尿病形式可以由肝细胞核因子 (HNF) 或胰岛素启动子因子-1 (IPF1/Pdx1) 突变引起	多个
发育性词汇运用障碍	FOXP2 转录因子中的突变与发育性词汇运用障碍有关，发育性词汇运用障碍是个体无法产生言语所需的精细协调运动的疾病	7q31
自身免疫性疾病	FOXP3 转录因子中的突变导致称为 IPEX 的罕见形式的自身免疫性疾病	Xp11.23-q13.3
Li-Fraumeni 综合征	由肿瘤抑制因子 p53 突变引起	17p13.1
乳腺癌	STAT 家族与乳腺癌有关	多个
多种癌症	HOX 家族参与各种癌症	多个

[0105] 因此，本发明方法中使用的结合染色质的活性剂特别是抗体或化学物质可以与已知与疾病(例如癌症)相关的转录因子相互作用。就此而言，本发明的方法可用于研究患病细胞和/或来自患病组织的细胞中的DNA和转录因子之间的相互作用。而且，本发明的方法可以用于研究药物与DNA/转录因子之间的相互作用。在这方面，目前约10%的处方药直

接靶向核受 体类转录因子。实例包括分别用于治疗乳腺癌和前列腺癌的他莫昔芬和比 卡鲁胺,以及各种类型的抗炎和促蛋白合成类固醇。另外,转录因子通常 通过信号级联由药物间接调节。

[0108] 根据上文,本发明涉及用于绘制涉及核酸,特别是DNA的分子相互 作用的方法,其中该方法提供关于多肽与核酸,特别是DNA相互作用的 有价值的信息。核酸可以来源于任何来源,例如细胞。具体而言,包含核 酸-蛋白质复合物的细胞。优选细胞是人类细胞、动物细胞、细菌细胞、酵 母细胞、古细菌细胞、植物细胞或病毒。更优选的细胞是人类细胞。然而, 细胞也可能来自非天然来源,例如工程细胞或人工修饰的细胞,特别是基 因修饰细胞。另外,人或动物细胞可以是患病细胞或未患病细胞或来自患 病或未患病组织的细胞。在这方面,人或动物细胞可以是癌细胞、免疫细 胞、血细胞或干细胞。优选的细胞是癌细胞。癌可以是实体瘤或血液瘤, 特别是白血病或肿瘤。已知的与改变的转录,即改变的DNA 可及性、修 饰的组蛋白、修饰的转录因子等有关的癌症由Yeh等人 (2013) *Curr. Opin. Oncol.* 25 (6) 总结。细胞也可以是胚胎细胞。

[0109] 因为本发明的方法对于分析低细胞数目特别有用,所以特别设想具有有限数量的可用作待分析核酸来源的细胞的来源。这些来源包括动物的早期胚胎阶段。在疾病特别是 人类疾病的情况下,细胞数量可能受到疾病性质的限制,例如癌症转移,小原发性肿瘤或小 病变器官,罕见组织和罕见细胞类型。人类临床样品的细胞数量可以通过获得样品的方法 进一步限制,例如穿刺活检或抽血。因此,源自这些来源的样品也被考虑用于本发明的方法 中。另外,细胞数量可能由于诸如受保护动物、珍稀动物、濒危动物等其他限制而受到限制。 此外,本发明的方法在单动物研究中,特别是小动物如线虫或斑马鱼中特别有用。

[0110] 在本发明的方法中,在制备测序文库或绘制涉及核酸的分子相互作用 之前,包含核酸的样品优选通过培养并收获细胞;固定细胞;裂解细胞并 由此获得包含核酸的第一样品;以及超声处理样品并由此获得包含核酸的 第二样品来制备。所述第二样品优选用于本发明的方法中以制备测序文库 或绘制涉及核酸的分子相互作用。当包含核酸的样品是原代细胞样品时, 例如,来自供体的样品,可以省略培养并收获的步骤。因此,在包含核酸 的样品是原代细胞样品的情况下,本发明的方法优选进一步包含固定细胞; 裂解细胞并由此 获得包含核酸的第一样品;以及超声处理第一样品并由此 获得包含核酸的第二样品。

[0111] 因此,包含核酸的样品优选通过包括培养并收获细胞的方法来制备。这可以使用 本领域公知的方法来完成。具体而言,培养方法必须适用于分 析中使用的细胞类型。这些 方法描述于例如Helgason等人 (2005) *Basic Cell Culture Protocols, Methods in Molecular Biology* or Freshney (2010) *Culture of Animal Cells*, Wiley-Blackwell。细胞的收获也通过本领域中描 述的公知方法进行。例如,可以通过离心收获细胞,由此在 所得细胞沉淀 物中发现细胞,同时上清液含有使用的培养基。

[0112] 在收获培养的细胞之后,可以固定细胞。固定用于保存来自衰变的样 品。因此,在 这个过程中,结构被保存在接近天然状态的状态 (化学和结 构上),例如,尽可能在活体组织中。这需要化学固定剂,它可以通过使 组织中的蛋白质和/或核酸不溶解而稳定它们。除了保留这种状态之外,固 定剂还用于交联样品中包含的大分子,特别是蛋白质和/或核酸。

[0113] 因此,交联固定剂通过在大分子,特别是蛋白质和/或核酸之间产生共 价化学键 而起作用。在这方面,众所周知的固定剂是甲醛。它优先用作10 % 中性缓冲福尔马林

(NBF), 即在磷酸盐缓冲盐水中约3.7%-4.0%甲醛。因为甲醛在室温下是气体, 所以在制造前述固定剂时使用溶于水(约37% w/v)的福尔马林甲醛气体。多聚甲醛是甲醛的聚合形式, 通常以细白粉末形式获得, 其在加热时解聚成福尔马林。甲醛通过交联蛋白质来固定组织, 主要是碱性氨基酸赖氨酸的残基。其效果是通过过量水而可逆的, 并避免福尔马林色素沉着。其他益处包括: 长期储存和良好的组织渗透。另一种用于固定的流行醛是戊二醛。它通过引起蛋白质中 α -螺旋结构的变形而以与甲醛类似的方式操作。然而, 戊二醛是更大的分子, 因此它穿过膜的扩散速度比甲醛慢。因此, 较厚的样品上的戊二醛固定可能会受到阻碍, 但这个问题可以通过减小样品的尺寸来克服。戊二醛固定的优点之一是可以提供更刚性或紧密连接的固定产物-其更长的长度和两个醛基使其能够“桥接”并连接更远距离的蛋白质分子对。它引起快速且不可逆转的变化, 迅速固定, 非常适合电子显微镜, 在4°C时固定良好, 并提供最佳的总体细胞质和细胞核细节。然而, 它不是免疫组织化学染色的理想选择。

[0114] 一些固定方案需要甲醛和戊二醛的组合, 以使它们各自的优势相互补充。

[0115] 这些交联固定剂-特别是甲醛-倾向于保留蛋白质的二级结构, 并且也可以保护大量的三级结构。

[0116] 然而, 固定也可以使用替代手段来完成, 例如, 使用物理手段的非化学固定, 特别是如例如Zhang等人(2004) Biochem Biophys Res Commun 322(3), 705-11所述的紫外线。备选地或另外地, 可以使用激光, 特别是UV激光来完成固定, 如例如Benedetti等人(2014) Methods Mol Biol 1204:24-34所述。因此, 优选使用化学物质和/或物理手段进行固定。在这方面, 优选物理手段包括紫外线或UV激光。更优选使用化学物质, 优选甲醛或多聚甲醛进行固定。

[0117] 引入的交联可以在库制备之后, 即在添加转座酶之后和从染色质中分离核酸之前除去。逆转交联可以使用本领域公知的方法完成。

[0118] 例如, 甲醛交联可以通过加热样品来去除。优选将样品加热至约65°C, 优选数小时。具体而言, 可以将样品加热至约65°C持续4小时或更长时间, 例如过夜。或者, 可将样品加热至约95°C约10-15分钟。然而, 为了保持包含核酸的样品的完整性, 加热至较低温度, 特别是至约65°C是优选的。除了加热之外, 可以加入去污剂和/或盐(例如0.5-1% SDS和/或约300mM NaCl)以除去交联。此外, 可以在除去-交联后加入RNase和/或Proteinase K以分别从包含核酸, 特别是DNA的样品中除去蛋白质和/或RNA。作为一个实例, 样品可以在37°C下用0.5 μ l 10mg/ml RNA酶A无DNA酶的RNase处理30min, 然后在55°C下用1 μ l 20mg/ml蛋白酶K处理1-2小时。

[0119] 在本发明方法的超快建立中, 可以将样品加热到高温以逆转交联。具体而言, 可以将样品加热至约95°C以逆转交联。这种高温显著减少了逆转交联所需的时间。具体地说, 逆转交联所需的时间可以从几个小时, 例如约65°C约4小时, 减少到约95°C约10-15分钟。因为本发明方法中使用的转座酶优选包括包含衔接头序列的寡核苷酸, 所以这些衔接头序列可以在逆转交联之前整合, 因为在本发明的方法中加入转座酶之后进行逆转交联。因此, 使用标准的ChIP方案不能使用这种高温。这是因为加热到高温会使ChIP DNA变性, 并且由于ChIP DNA的复杂性, 一些片段(特别是富含AT的序列)不能适当地重新退火。当通过连接双链衔接头来制备文库时, 可能会从最终库中排除不适当重新退火的ChIP DNA

片段，并引入 测序偏差。然而，在本发明的方法中，可以采用高温，如约95°C，以逆转 交联。这显著降低了测定的总体持续时间；参见实施例14。另外，使用高 温(例如约95°C)来逆转交联避免了从珠中洗脱的步骤。避免珠洗脱进一步降低了所用方法的复杂性，并进一步降低了实施本发明方法所需的总时 间。这是因为从珠中洗脱包括使用与随后的PCR步骤不相容的缓冲液，例 如使用SDS和/或高浓度的盐。如果没有事先DNA清除，这种缓冲液使得 文库扩增困难或不可能。因此，本文所述的超快建立不需要DNA纯化。在本发明的方法涉及使用高温，特别是约95°C的温度逆转交联的情况下，所述方法优选还包括在施用高温之前末端修复在转座酶反应过程中引入的 寡核苷酸的步骤，即填充衔接头序列的步骤，特别是在与包含在转座酶反 应过程中引入的寡核苷酸的链相反的反向链上填充衔接头序列。因此，其 中使用高温(如约95°C)来逆转交联的本发明方法优选包括在施加高温之 前添加用于末端修复的PCR成分的步骤。优选地，在末端修复条件下，在 加热之前，使用DNA聚合酶，如 Taq聚合酶，使用PCR MM在珠上进行 末端修复，例如，72°C持续5分钟。但是，最终修复也可以在较低的温度 下使用末端修复混合物来完成。

[0120] 因此，本发明提供了用于制备测序文库的超快速方法。用于制备测序 文库的超快 速方法包括将结合染色质的活性剂添加至包含核酸的样品中，其中样品已通过交联固定；分离由所述活性剂结合的染色质；向由所述活性剂结合的分离的染色质中添加转座酶；填充在转座酶反应期间产生的寡 核苷酸末端；在高温(优选在约95°C)下逆转交联；并获得测序文库。另 外，本发明涉及用于绘制涉及核酸的分子相互作用的超快速方法。用于绘 制分子相互作用的超快速方法包括将结合染色质的活性剂添加至包含核酸 的样品中；分离由所述活性剂结合的染色质；向分离的由所述活性剂结合 的染色质中添加转座酶；填充在转 座酶反应期间产生的寡核苷酸末端；在 高温(优选在约95°C)下逆转交联；核酸扩增；扩增核酸的测序；并鉴定 分子相互作用。在这方面超快速意味着本发明的超快速方法显著减少了对 于已知方法所预期的总体实验时间。特别地，本发明的超快速方法允许分 别在不到一个工作日即小于约10小时内制备测序文库或绘制分子相互作 用。从获得样品制备测序文库所需的总时间，例如从供体获得血液样品以 获得测序文库的总体时间在约15小时的范围内。

[0121] 本发明的方法可以进一步包括裂解细胞的步骤。裂解是指细胞膜的分 解。这可以 通过本领域众所周知的方法来实现。具体而言，裂解可以通过 机械手段或化学手段来实 现。例如，可以使用细胞膜的机械破裂，如通过 反复冻融，超声处理，压力或过滤。然而，优 选通过使用特别是酶或去污 剂或其他离液剂的化学手段来实现裂解。细胞裂解的优 选方 法在Thermo Scientific Pierce Cell Lysis Technical Handbook或Lottspeich, Engels (2012) Bioanalytik, Springer Spektrum中描述。就此而言，进行本发明 方法中使用的裂解以从细胞样品分离核酸，由此获得包含核酸的第一样品。所述第一样品用于使用本发明 的方法进一步分析，即所述第一样品或者 是 用于进一步制备样品(特别是使用超声处理) ，或者 是使用本发明的方法 直接分析以制备测序文库或者绘制涉及核酸的分子相互作用。

[0122] 因此，在一个实施方案中，在细胞裂解之后，本发明的方法可以包括 超声处理步 骤。超声处理具有无数的化学和物理效应。在生物应用中，超 声处理通常用于破坏或失活 生物材料。例如，超声处理通常用于破坏细胞 膜并释放细胞内容物。这个过程被称为声孔 作用 (sonoporation)。超声 处理还用于将核酸分子，特别是DNA片段化，其中经历短暂超声

处理的 核酸特别是DNA被剪切成更小的片段。超声处理还用于将含有核酸和蛋白的分子的复合物,特别是含有核酸,特别是DNA的染色质片段化, 其中复合物经历短暂的超声处理,其中复合物中的核酸内含物,尤其是 DNA,被剪成更小的片段。在这方面,如何调节超声处理强度以产生具有 特定长度和/或其中包含在包含核酸,特别是DNA的样品中的大部分片段 具有特定的长度的核酸片段(特别是DNA)是众所周知的。就此而言,优 选包含核酸,特别是DNA的样品包含具有200至700个碱基对长度的片 段。因此,优选进行超声处理直至大部分核酸片段长200-700个碱基对。如何调整超声强度和持续时间来产生这种片段是众所周知的。此外,如何 确定这些片段的长度以验证超声处理设置是众所周知的。

[0123] 就此而言,超声处理设置可以取决于固定条件和细胞系/组织/细胞类型 /生物以获得核酸样品。另外,超声处理设置可能取决于使用的超声处理设 备。

[0124] 还设想使用替代技术来片段化核酸样品,特别是包含DNA的样品。例如,酶消化可以用于染色质中包含的核酸的片段化。示例性酶是片段酶 (NEB) 或MNase(金黄色葡萄球菌的胞外核酸酶)。化学活性剂或除超 声波之外的其他物理方法也可用于片段化染色质中包含的核酸。

[0125] 超声处理结果可以通过本领域公知的方法来验证。例如,为了验证大 部分核酸,特别是DNA片段是否长200-700个碱基对,可以使用琼脂糖凝 胶电泳测试片段长度。

[0126] 根据上文,本发明的方法包括在上述制备步骤之后作为第一步骤的步 骤:将结合染色质的活性剂特别是抗体或化学物质添加至包含核酸特别是 DNA的样品中。如上所述,优选包含核酸,特别是DNA的样品来源于细 胞。在加入结合染色质的活性剂之后,分离由所述活性剂结合的染色质。特别地,由所述活性剂结合的染色质与未结合的染色质分离。通过这样做,染色质的总量显著减少,这减少了标记事件。染色质的分离可以通过本领 域中描述的各种技术来实现。例如,结合染色质的活性剂,特别是抗体或 化学物质,可以通过亲和相互作用固定在表面上。这些表面优选是颗粒 (珠)。但是,其他表面也是可以设想的,例如,柱。在结合染色质的活性剂是抗体的情况下,抗体的Fc-部分可以通过蛋白A、蛋白G、蛋白L 等结合到珠的表面。就此而言,蛋白A是最初在细菌金黄色葡萄球菌的细 胞壁中发现的42kDa表面蛋白。它由spa基因编码,其调控由DNA拓扑 结构,细胞渗透压和称为Ar1S-Ar1R的双组分系统控制。由于其结合免疫 球蛋白的能力,它常用于生化研究。或者,抗体可经由蛋白G结合到表面, 所述蛋白G是C和G组链球菌中表达的免疫球蛋白结合蛋白,非常像蛋白A但具有不同的结合特异性。它是65-kDa (G148蛋白G) 和58kDa (C40 蛋白G) 细胞表面蛋白,通常用于通过其结合Fab和Fc区来纯化抗体。因此,结合染色质的活性剂(其中所述活性剂是抗体)可以通过蛋白A、蛋白G、蛋白L等与珠结合,以将由所述活性剂特别是抗体结合的染色质 与未结合的染色质分离。

[0127] 在本发明的方法中,染色质也可以通过其他方式,例如与结合染色质 的活性剂相连的亲和标签来分离。例如,亲和标签可以包括分别可以结合 链霉抗生物素蛋白或镍的生物素或His。多组分亲和标签复合物的其他实 例包括配体及其受体,例如抗生物素蛋白-生物素,链霉抗生物素蛋白-生 物素以及生物素,链霉抗生物素蛋白或抗生物素蛋白的衍生物,包括但不 限于2-亚氨基生物素,脱硫生物素,NeutrAvidin,CaptAvidin等;结合 蛋白/ 肽,包括麦芽糖-麦芽糖结合蛋白 (MBP),钙-钙结合蛋白/肽 (CBP) 的;抗原-抗体,包括表位 标签及其相应的抗表位抗体;半抗原,例如二硝 基苯基和地高辛配基及其相应的抗体;适

配体及其相应的靶标；聚-His标签(例如五-His和六-His)及其结合配偶体包括相应的固定化金属离子亲和层析(IMAC)材料和抗聚-His抗体；荧光团和抗荧光团抗体；等等。

[0128] 因此，结合染色质的活性剂优选用生物素标记，其中结合染色质的活性剂是化学物质。

[0129] 珠可以是磁性的、乳胶或琼脂糖基材料等。然后可以通过分离珠来分离固定的靶染色质。这可以通过使用过滤柱的旋转离心来实现，所述过滤柱保留具有与过滤器上的染色质结合的活性剂的珠，而未结合的染色质部分通过过滤器并且可以被丢弃。在磁珠的情况下，将磁力施加到珠以保留在反应容器中，而未结合的染色质部分可以通过例如移液来丢弃。所述活性剂还可以在加入染色质之前预偶联至表面/珠。当预偶联时，活性剂还可以化学交联到表面，并且不完全依赖亲和相互作用来分离染色质。举例来说，可以使用庚二酸亚氨酸二甲酯(DMP)将蛋白质偶联到珠上。染色质的分离通常由洗涤步骤支持以去除染色质与所述活性剂的非特异性相互作用或染色质与反应容器或分离试剂表面的非特异性相互作用。通过所述活性剂分离的染色质或通过上述步骤分离的化学物质的洗涤是通过加入含有包括盐和去污剂的化学品的缓冲水溶液并随后除去它来实现的。因此，本发明的方法可以进一步包括在分离由结合染色质的活性剂结合的染色质之后的洗涤步骤。

[0130] 在分离由结合染色质的活性剂(特别是抗体或化学物质)结合的染色质后，将转座酶加入到分离的染色质中。转座酶是一种与转座子末端结合并通过剪切和粘贴机制或复制转座机制催化转座子向基因组另一部分运动的酶。转座酶以EC号分类为EC 2.7.7。编码转座酶的基因在大多数生物的基因组中广泛存在，并且是已知最丰富的基因。在本发明的上下文中优选的转座酶是转座酶(Tnp)Tn5。Tn5是包含逆转录病毒整合酶的RNA酶超家族成员。Tn5可以在希瓦氏菌和埃希氏菌中找到。转座子编码对卡那霉素和其他氨基糖苷类抗生素的抗生素抗性。Tn5和其他转座酶显著失活。由于DNA转座事件本身具有诱变性，转座酶的低活性对于降低引起宿主致命突变的风险并因此消除可转座元件是必要的。Tn5是非常不反应性的原因之一是因为N端和C端位于彼此相对接近的位置并倾向于相互抑制。这通过表征导致转座酶活性过高的几种突变来阐明。一种这样的突变L372P是Tn5转座酶中氨基酸372的突变。该氨基酸通常是 α -螺旋中间的亮氨酸残基。当这种亮氨酸被脯氨酸残基取代时， α -螺旋被破坏，引入对C-末端结构域的构象变化，将其与N-末端结构域分开以促进蛋白质的更高活性。因此，优选使用这种经修饰的转座酶，其具有比天然存在的Tn5转座酶更高的活性。此外，特别优选用于本发明方法中的转座酶装载有插入靶核酸，特别是靶DNA中的寡核苷酸。

[0131] 例如，由SEQ ID NO:1或2的核酸序列或与SEQ ID NO:1或2中任一具有80、85、90、95、96、97、98或99%序列同一性的核酸序列编码的转座酶可以用于本发明的方法中。就此而言，转座酶可以使用具有如SEQ ID NO:3所示的核酸序列的表达载体或使用包含编码转座酶的序列的表达载体来产生，所述后一转座酶对应于由与SEQ ID NO:1或2中任一具有80、85、90、95、96、97、98或99%序列同一性的核苷酸序列编码的转座酶。

[0132] 因此，优选使用过度激活的Tn5转座酶和Tn5型转座酶识别位点(Goryshin和Reznikoff, J. Biol. Chem., 273:7367 (1998))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, 35:785, 1983; Savilahti, H等人, EMBO J., 14:4893, 1995)。可用于本发明方法的转座系统的更多实例包括金黄色葡萄球菌Tn552(Colegio等

人, J. Bacteriol, 183:2384-8, 2001; Kirby C等人, Mol. Microbiol, 43:173-86, 2002), Ty1 (Devine和Boeke, Nucleic Acids Res., 22:3765-72, 1994和国际公开WO 95/23875), 转座子Tn7 (Craig, NL, Science, 271:1512, 1996; Craig, NL, 综述于: Curr Top Microbiol Immunol, 204:27-48, 1996), Tn/0和IS10 (Kleckner N等人, Curr Top Microbiol Immunol, 204:49-82, 1996), Mariner转座酶 (Lampe DJ 等人, EMBO J., 15:5470-9, 1996), Tel (Plasterk RH, Curr. Topics Microbiol. Immunol, 204:125-43, 1996), P Element (Gloor, GB, Methods Mol. Biol, 260:97-14, 2004), Tn3 (Ichikawa&Ohtsubo, J Biol. Chem. 265: 18829-32, 1990), 细菌插入序列 (Ohtsubo&Sekine, Curr. Top. Microbiol. Immunol. 204:1-26, 1996), 逆转录病毒 (Brown 等人, Proc Natl Acad Sci USA, 86:2525-9, 1989) 和酵母的逆转录转座子 (Boeke&Corces, Annu Rev Microbiol. 43:403-34, 1989)。更多的实例包括转座酶家族酶的IS5, Tn10, Tn903, IS911和工程化形式 (Zhang 等人, (2009) PLoS Genet. 5: e1000689. 电子公开2009年10月16日; Wilson C等人 (2007) J. Microbiol. Methods 71: 332-5) 和美国专利号 5,925,545; 5965443; 6437109; 6159736; 6406896; 7083980; 7316903; 7608434; 6294385; 7,067,644,7,527,966 描述的那些; 以及国际专利公开号 WO2012103545, 其全部通过引用整体并入本文。

[0133] 尽管适用于所用转座酶的任何缓冲液均可用于本发明的方法中, 但优选使用特别适用于所用转座酶的有效酶促反应的缓冲液。就此而言, 包含二甲基甲酰胺的缓冲液特别优先用于本发明的方法中, 特别是在转座酶反应过程中。另外, 可以使用包含替代缓冲体系 (包括TAPS、Tris-乙酸盐 或类似体系) 的缓冲液。此外, 聚集剂如聚乙二醇 (PEG) 对于提高极低量DNA的标签化效率特别有用。Picelli等人 (2014) Genome Res. 24: 2033-2040 描述了用于标签化反应的特别有用的条件。

[0134] 转座酶催化核酸, 特别是靶核酸中的DNA, 特别是靶DNA的插入。用于插入的靶核酸, 特别是靶DNA包含在由结合染色质的活性剂 (特别是抗体或化学物质) 所结合的分离的染色质中。用于本发明方法的转座酶 装载有插入靶核酸, 特别是靶DNA的寡核苷酸。转座酶和寡核苷酸的复合物也被称为转座体。优选地, 转座体是包含用于整合的两种不同寡核苷酸的异二聚体。就此而言, 加载到转座酶上的寡核苷酸包含多个序列。特别地, 寡核苷酸至少包含第一序列和第二序列。第一序列对于将寡核苷酸 加载到转座酶上是必需的。在US2010/0120098中给出了将寡核苷酸加载到转座酶上的示例性序列。第二序列包含扩增期间, 特别是在PCR扩增期间 引物结合所需的衔接头序列。因此, 包含第一和第二序列的寡核苷酸通过转座酶插入靶核酸, 特别是靶DNA中。寡核苷酸可以进一步包含含有条形码序列的序列。条形码序列可以是随机序列或限定的序列。就此而言, 根据本发明的术语“随机序列”应理解为核苷酸序列, 其中每个位置具有独立且相等的概率为任何核苷酸。随机核苷酸可以是以任何顺序的任何核苷酸, 例如G、A、C、T、U或其化学类似物, 其中: G被理解为代表鸟苷酸核苷酸, A腺苷酸核苷酸, T胸苷酸核苷酸, C胞苷酸核苷酸和U尿苷酸核苷酸。本领域技术人员将会理解, 已知的寡核苷酸合成方法可能固有地导致核苷酸G、A、C、T或U的不均等表示。例如, 合成可能导致核苷酸的过度表达, 例如随机化DNA序列中的G。如预期的那样, 这可能导致独特随机序列的数量减少, 如基于核苷酸的相同表示所预期的。用于插入靶核酸, 特别是DNA的寡核苷酸可以进一步包含测序衔接头, 例如适用于纳米孔测序或Roche454测序的衔接头。此外, 寡核苷酸可以包含生物素标签序列。装载到转座酶上的

寡核苷酸优选包含所述第一和第二序列以及用于索引的条形码序列。在转座酶反应期间条形码序列的整合允许在测序分析和/或绘制分子相互作用的过程中独特地鉴定每个核酸片段,特别是DNA片段。

[0135] 本领域技术人员清楚地知晓,使用的转座酶有效整合核酸,特别是在靶核酸中的DNA,特别是靶DNA所需的时间可以根据各种参数而变化,如缓冲液组分、温度等。因此,本领域技术人员很清楚,在发现最佳温育时间之前可以测试/应用各种温育时间。在方面的最佳是指考虑到执行本发明的方法的整合效率和/或所需时间的最佳时间。虽然改变温育时间不一定与所述核酸,特别是在所述靶核酸中的所述DNA,特别是靶DNA的有效整合相关,但优选使用小于10分钟,小于5分钟,优选少于2分钟的温育时间。对于标签化反应,最优选采用1分钟的反应时间。此外,可以优化温度和体积等参数。在这方面,推荐的Tn5转座酶的温育温度约为37°C。因此,优选本发明的方法包括添加转座酶并随后在约37°C下为了标签化而温育的步骤,优选约1min。然而,也可以使用备选反应温度,而优选使用高于约16°C且低于约55°C的温度以维持样品完整性和转座酶效率。

[0136] 在向分离的染色质中添加转座酶后,从包含染色质的样品中分离核酸,特别是DNA,即核酸,特别是DNA与染色质的剩余组分分离。这可以通过本领域已知的各种技术来实现。例如,可以使用柱纯化,苯酚-氯仿提取,接着乙醇沉淀,固相可逆固定化和**Chelex®100**以及本领域已知的其他技术来纯化核酸,特别是DNA。取决于使用的缓冲液的pH和盐含量,柱纯化依赖于核酸,特别是DNA(吸附)与固相(二氧化硅或其他)的结合。离心样品后,变性蛋白质保留在有机相中,而含有核酸,特别是DNA的水相与氯仿混合,从溶液中除去苯酚残留。为了从水相中分离DNA,苯酚-氯仿接着是乙醇或异丙醇沉淀。由于DNA不溶于这些醇,它会聚集,离心后产生沉淀。通过增加离子强度,通常通过添加乙酸钠来改善DNA的沉淀。**Chelex®100**是由Bio-Rad分发的螯合材料,用于通过离子交换纯化其他化合物。它也可以用来纯化DNA。SPRI(固相可逆固定)珠是顺磁性的(磁性仅在磁场中)。每个珠由聚苯乙烯制成,周围覆盖着一层磁石,其上包被羧基分子。在“聚集剂”聚乙二醇(PEG)和盐(通常20%PEG, 2.5M NaCl)存在下,这些是可逆结合DNA的。PEG引起带负电荷的DNA与珠表面上的羧基结合。由于固定依赖于反应中PEG和盐的浓度,因此珠与DNA的体积比是关键的。DNA纯化通常通过向溶液中加入RNA酶和蛋白酶去除RNA和蛋白质来支持。

[0137] 在分离核酸之后,可以如上所述获得测序文库。特别地,与测序相容的核酸文库,特别是DNA文库包含核酸片段(特别是DNA片段),其包含测序所必需的衔接头序列。因此,用于制备测序文库的本发明的方法可以进一步包括用于整合所述衔接头序列的扩增步骤。扩增如下所述完成。衔接头序列根据制备测序文库之后使用的测序方法而变化。例如,在使用 Illumina测序的情况下,可以将i5和i7末端连接至核酸片段。这也可以通过转座酶反应来实现,其中装载到转座酶上的寡核苷酸包含测序相容的衔接头序列。

[0138] 在本发明的方法用于绘制涉及核酸的分子相互作用时,核酸在分离后被扩增。扩增可以通过本领域已知的各种技术来实现。用于核酸,特别是DNA扩增的最公知的技术是聚合酶链式反应(PCR),其中使样品在一定条件下与一对寡核苷酸引物接触,所述条件允许引物与样品中的核酸模板杂交。引物在合适的条件下延伸,从模板中解离,再退火,延伸并解离以扩增核酸的拷贝数。这个循环可以重复。扩增产物可以通过诸如电泳、限制性内

切核酸酶切割模式、寡核苷酸杂交或连接和/或核酸测序的技术来表 征。

[0139] 适用于本发明方法的引物包含序列,所述序列可与包含在用于本发明 方法中的转座体中的寡核苷酸中包含的第二序列杂交。另外,引物可以包 含测序所需的序列。优选的是,在本发明的方法中使用与随后使用的测序 方法相容的特异性引物。就此而言,作为本发明方法中使用的一种优选的 测序方法,Illumina测序与引入流动池末端的引物相容,其可以与聚类扩 增中所需的流动池杂交。在这方面,引物可以引入I5和I7末端用于 Illumina 测序。此外,引物可能会引入用于多路复用的条形码。具体而言,包含在 引物序 列中的条形码可以用作唯一的分子标识符来区分PCR重复和/或用 作限定的条形码以在一次测序运行中组合多个实验。

[0140] 在用于绘制涉及核酸,特别是DNA的分子相互作用的本发明方法中,对扩增的DNA 进行测序。本领域已知有各种测序方法。通常,测序可以 使用固相支持体上的焦磷酸测序 (454测序, Roche),具有可逆终止的合 成测序(**ILLUMINA®**Genome Analyzer)或纳米 孔技术(例如Oxford Nanopore Technologies MinION™)进行测序。在一些实施方案中,例如 通过确定核苷酸序列来分析分离的加标签片段。在一些实例中,使用测序 或杂交技术 在有或没有扩增的情况下测定核苷酸序列。

[0141] 从对核酸进行测序获得的序列信息开始,可以使用本领域已知的工具 来鉴定分子 相互作用。例如,可以使用序列比较软件分析数据,所述序列 比较软件将测序的核酸与 基因组序列比对。基因组序列通常是已知的并可 从可自由访问的数据源获得。在待分析 样品中发现的测序核酸与基因组序 列的匹配可用作所述测序核酸与大分子(例如组蛋白或 转录因子)结合的 指示物,所述大分子在本发明的方法中被结合染色质的活性剂识别。

[0142] 当结合染色质的活性剂是化学物质(例如,药物)时,与待分析的核 酸片段中包含 的基因组序列的匹配表明化学物质(例如,药物)结合到染 色质中包含的特定核酸区域。

[0143] 基于将测序的核酸与基因组序列进行匹配,可以使用统计学计算方法 来确定显 著结合的区域以区分它们与非特异性“背景信号”。通过将它们与 包括基因表达,基因组注 释,基因本体论或其他系统生物学数据集的其他 数据集相关联,可以使用所识别的区域来 进一步推断它们的生物学作用。

[0144] 通过积累来源于与结合染色质的活性剂(特别是化学物质,例如,药 物)结合的所 述核酸的区域,也可以使用计算方法来确定所述区域的显著 序列特征。这种方法可用于发 现已知与特定转录因子结合的特定DNA结 合基序的富集。

[0145] 本发明的方法还可用于鉴定靶核酸中包含的区域,特别是转座酶不能 接近的 DNA。特别地,在使用本发明的方法制备测序文库的情况下,可以 通过在靶染色质中由转座 酶引入测序相容的寡核苷酸来产生测序文库片 段。在包含核酸和蛋白质的染色质中,包含 在染色质中的蛋白质可以在一 定程度上干扰在DNA-蛋白质相互作用的位点处衔接头在靶 核酸,特别是 DNA中的整合,而不干扰测序文库的制备。受保护免于转座酶插入的所述 核 酸区域可以通过计算方法鉴定。这些区域是染色质中包含的蛋白质的“足 迹”,因此揭示了 高分辨率的相互作用。

[0146] 虽然本发明的方法可以在一个管中进行,但优选在反应过程中包括管 转移。特别 地,可以在添加转座酶之前改变反应管。另外或备选地,可以 在转座酶反应完成之后添加 转座酶随后改变管。后者降低了粘附在管壁上 的非特异性染色质片段的标记。因此,本发

明的方法优选包含至少一个，优选两个管转移，其中在从染色质分离核酸或者采用超快方案的逆转交联 之前在转座酶反应之后进行第一个管转移，并且其中如果使用第二次管转移，则第二次管转移在加入转座酶之前进行。

[0147] 本发明还涉及试剂盒，特别是研究试剂盒。本发明的试剂盒包含一种 或多种结合染色质的活性剂，如一种或多种化学物质或一种或多种抗体/ 抗体和转座酶。本发明的试剂盒可以包含过度激活的，优选也是寡核苷酸 负载的转座酶。在一个具体的实施方案中，本发明的试剂盒包含由SEQ ID NO:1或2的核酸序列编码的转座酶或具有SEQ ID NO:3的核酸序列的 表达载体。本发明的试剂盒还可以包含即用形式的转座酶。本发明的试剂 盒可用于诊断疾病等医学病况。所述医疗病况，如疾病，可以是涉及DNA 与其他组分(例如但不限于转录因子/组蛋白等)相互作用的任何病况/疾病。例如，已知和DNA与转录因子/组蛋白相互作用相关的疾病包括但不限于 增殖性疾病，例如癌症。因此，本发明的试剂盒可用于诊断疾病，包括但不限于T细胞急性淋巴细胞白血病等，急性骨髓性白血病，尤因肉瘤，急 性早幼粒细胞白血病，急性淋巴细胞白血病，弥漫性大B细胞淋巴瘤，移 行细胞癌，结肠直肠癌，胰腺癌，乳腺癌，骨髓增生异常综合征，中线癌，乳头状甲状腺癌，肾癌，成神经管细胞瘤，多发性骨髓瘤，骨髓增生异常 综合征，食管癌，卵巢癌，前列腺癌，肺癌，横纹肌样癌，肝细胞癌，家 族性神经鞘瘤病，软骨肉瘤，类肾样骨肉瘤(epethioid sarcoma)，脑膜瘤，脊索瘤，未分化肉瘤，帕金森氏病，亨廷顿病，先天性肌强直性营养 不良，类风湿性关节炎，系统性红斑狼疮，1型糖尿病，免疫缺陷，着丝 粒不稳定和面部异常综合征和ATRX综合征等。

[0148] 此外，试剂盒可用于评估/确定化合物与DNA的相互作用。就此而言，本发明的试剂盒可用于评估/确定个体如患者对用与DNA相互作用的化学 化合物治疗的反应的可能性。治疗可涉及使用已知有效的化学化合物治疗 各种疾病/病况，或者其中待测试有效性时，其中待治疗的病况可能是涉及 DNA相互作用的任何疾病/病况。因此，所述医学病况如疾病可以是涉及 DNA与其他组分(例如但不限于转录因子/组蛋白等)相互作用的任何病况/疾病。例如，已知和DNA与转录因子/组蛋白相互作用相关的疾病包括 但不限于增殖性疾病，例如癌症。因此，本发明的试剂盒可用于诊断疾病，包括但不限于T细胞急性淋巴细胞白血病等，急性骨髓性白血病，尤因肉 瘤，急性早幼粒细胞白血病，急性淋巴细胞白血病，弥漫性大B细胞淋巴 瘤，移行细胞癌，结肠直肠癌，胰腺癌，乳腺癌，骨髓增生异常综合征，中线癌，乳头状甲状腺癌，肾癌，成神经管细胞瘤，多发性骨髓瘤，骨髓 增生异常综合征，食管癌，卵巢癌，前列腺癌，肺癌，横纹肌样癌，肝细 胞癌，家族性神经鞘瘤，软骨肉瘤，类肾样骨肉瘤，脑膜瘤，脊索瘤，未 分化的肉瘤，帕金森氏病，亨廷顿病，先天性肌强直性营 养不良，类风湿 性关节炎，系统性红斑狼疮，1型糖尿病，免疫缺陷，着丝粒不稳定和面 部异常综合征和ATRX综合征等。

[0149] 在本发明的特别优选的实施方案中，本发明的试剂盒(在上下文中制 备)或本发明的方法和用途可以进一步包含或提供(一个或多个)说明手 册。例如，所述说明手册可以指导技术人员(如何)将本发明的试剂盒应 用于本文提供的并根据本发明的诊断用途中。特别地，所述说明手册可以 包括使用或应用在此提供的方法或用途的指导。

[0150] 本发明的试剂盒(在上下文中制备)可以进一步包含实施本发明的方 法和用途所适合/需要的物质/化学品和/或设备。例如，这些物质/化学品和 /或设备是用于稳定和/或

储存(a)本文提供的用途所需的化合物(例如稳定和/或储存化学剂)的溶剂、稀释剂和/或缓冲剂)和/或包含在本发明试剂盒中的转座酶。

[0151] 除非另外定义,否则本文使用的所有技术和科学术语具有与本发明所属领域的普通技术人员通常理解的相同的含义。尽管在本发明的实践或测试中可以使用与本文描述的方法和材料类似或等同的方法和材料,但是下面描述合适的方法和材料。如有冲突,以本说明书(包括定义)为准。另外,这些材料、方法和实施例仅仅是说明性的而不是限制性的。

[0152] 除非另有说明,否则本发明的方法和技术通常根据本领域公知的常规方法进行,并如在本说明书中引用和讨论的各种一般和更具体的参考文献中所述进行。参见例如Sambrook等人,Molecular Cloning:A Laboratory Manual,第2版,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,NY(1989)和Ausubel等人,Current Protocols in Molecular Biology,Greene Publishing Associates 1992),和Harlow和Lane Antibodies:A Laboratory Manual,Cold Spring Harbor Laboratory Press, Cold Spring Harbor,NY(1990)。

[0153] 尽管在附图和前面的描述中详细图示和描述了本发明,但是这样的图示和描述被认为是说明性的或示例性的而不是限制性的。应该理解,在以下权利要求的范围和精神内,普通技术人员可以做出改变和修改。特别地,本发明利用上面和下面描述的不同实施方案的特征的任意组合来覆盖进一步的实施方案。

[0154] 本发明还单独地覆盖附图中所示的所有其他特征,尽管它们可能在之前或以下说明书中未被描述。而且,附图和说明书中描述的实施方案的单个备选方案以及其特征的单个备选方案可以从本发明的另一个方面的主题中被放弃。

[0155] 此外,在权利要求中,词语“包括”不排除其他元素或步骤,并且不定冠词“一”或“一个”不排除多个。单个单元可以实现权利要求中记载的若干特征的功能。与属性或值有关的术语“基本上”、“约”、“近似”等也特别地分别准确地定义属性或准确地定义该值。权利要求中的任何参考符号不应被解释为限制范围。

[0156] 以下附图也说明了本发明。

[0157] 图1、现有技术方法的概述及其主要缺点

[0158] 图2、标准ChIP-seq,ChIP-标签化和ChIPmentation的示意性概述

[0159] 通过标签化纯化的ChIP DNA(ChIP-标签化),与标准ChIP-seq和文库制备相比,ChIP的示意性概述。所有三种方案开始于用甲醛固定细胞,细胞裂解,染色质的超声处理和用与珠结合的特异性抗体的免疫沉淀(“染色质免疫沉淀”)。对于标准ChIP-seq(左图),逆转交联之后是ChIP DNA的纯化,然后通过包括末端修复,纯化,A-加尾,衔接头连接和大小选择的多步骤程序进行文库制备。ChIP标签化(中心)使用纯化的ChIP DNA进行基于标签化的库制备,其缺点是该方案对改变DNA浓度敏感。在ChIPmentation(右)中,在免疫沉淀过程中使用具有负载衔接头的Tn5转座酶的标签化而在单个步骤中引入测序衔接头。

[0160] 图3、补充性图3:Tn5酶浓度对ChIPmentation文库大小分布的影响

[0161] 用不同量的Tn5转座酶(来自Illumina Nextera DNA文库制备试剂盒的0.2μl至5μl酶)制备的H3K4me3的ChIPmentation文库的DNA片段大小分布。逆转交联后但文库富集前的片段大小显示为红色,富集PCR后的片段大小显示为绿色,并且经大小选择的最终文

库显示为蓝色。

- [0162] 图4、ChIPmentation:组蛋白标记和转录因子的快速、廉价、低输入 ChIP-seq
- [0163] (a) ChIP的示意性概述(参见图2,标准ChIP-seq,从纯化的ChIP DNA开始的ChIP标签化和ChIPmentation的比较)。
- [0164] (b) 来自用不同Tn5酶浓度获得的H3K4me3ChIPmentation文库的 配对末端测序的绘制的插入物长度的大小分布。
- [0165] (c) 用不同的Tn5酶浓度获得的H3K4me3ChIPmentation文库的绘制的(顶部)和独特的(底部)读段的百分比。
- [0166] (d) 比较用不同Tn5酶浓度获得的H3K4me3ChIPmentation文库的 基因组浏览器屏幕截图。
- [0167] (e) 用不同的Tn5酶浓度获得的H3K4me3ChIPmentation的全基因 组相关热图(1,000bp窗口)。
- [0168] (f) 显示来自不同输入量的五种组蛋白修饰和四种转录因子的 ChIP-seq (“ChIP”) 和ChIPmentation (“CM”) 数据的基因组浏览器屏 幕截图。来自两个生物学重复的数据被组合。
- [0169] (g) 针对不同组蛋白标记和不同细胞输入量的标准ChIP-seq和 ChIPmentation数据的全基因组相关热图(1,000bp窗口)。
- [0170] (h) 不同转录因子和不同细胞输入量(高:10M细胞;低:100k或 500k细胞) 的标准 ChIP-seq和ChIPmentation数据的全基因组相关性 (1,000bp窗口)。
- [0171] (i) 标准ChIP-seq(顶部),市售低输入库制备试剂盒(中) 和 ChIPmentation(底部)的手动操作时间比较。从方案开始直至(但不包 括)最终的文库扩增PCR反应计算手动操作时间。
- [0172] 图5、标准ChIP-seq和ChIPmentation数据的整体比较
- [0173] (a) 沿着所有基因的组蛋白标记分布的复合图,分别显示 ChIPmentation(左) 和标准ChIP-seq(右)。
- [0174] (b) 对于所有测序的库,峰中的读段的比率(FRIP) 和从 ChIPmentation(上图) 和 ChIP-seq(下图)的数据调用的峰的数目。要 注意的是,重复之间测序的深度不同(图8)。
- [0175] 图6、用于低输入ChIPmentation数据的基因组浏览器跟踪
- [0176] 显示单个生物学重复和不同细胞输入量(即10M、500K、100K和10K 细胞) 的 ChIPmentation (“CM”) 数据的基因组浏览器屏幕截图。从1000 万个细胞数据中获得的标准 ChIP-seq (“ChIP”) 被包括在内作为参考。
- [0177] 图7、标准ChIP-seq和ChIPmentation实验之间的峰重叠
- [0178] 峰重叠计算为一个重复/方法/输入量中顶部-X%峰与另一个重复/方法 /输入量中显著峰重叠的百分比。
- [0179] 图8、24个标准ChIP-seq库,52个ChIPmentation库和9个ChIP- 标签化库的测序总结
- [0180] 图9、ChIPmentation数据中的高分辨率模式
- [0181] (a) 在被调用的峰下的各个转录因子的基序周围CTCF, GATA1, PU.1和REST ChIPmentation数据的信号强度(Tn5插入频率)。上图显 示ChIPmentation、ATAC-seq和

DNase-seq的原始信号,而下图显示通过基因组DNA的标签化以及归一化的ChIPmentation和ATAC-seq信号强度定义的背景。通过具有每个峰的背景信号的e到Z分数上的信号来执行归一化。为了可视化目的,归一化的信号对所有峰进行平均,用20bp Hanning窗平滑并将Z分数转换用于比较。

[0182] (b) H3K4me3的ChIPmentation数据中插入事件之间的成对距离的频率。10bp的周期性振荡频率可以与核小体周围的DNA的旋转性质相关联¹。

[0183] (c) 使用NucleoATAC软件定位的核小体周围的 H3K4me1ChIPmentation数据和GM12878细胞的ATAC-seq数据的信号强度(插入频率)。注意在核小体边界处具有更高和周期性插入的结构化模式。

[0184] 图10、标准ChIP-seq和ChIPmentation的试剂成本比较

[0185] 标准ChIP-seq(上),市售低输入库制备试剂盒(中)和ChIPmentation(下)的试剂成本比较。计算文库制备的成本估算值,文库制备包括扩增和索引化,但不包括用于大小选择的试剂,反应纯化和测序前的最终质量控制步骤。

[0186] 图11、从纯化的ChIP DNA开始通过ChIP标签化制备文库

[0187] (a) 使用不同量的纯化的ChIP DNA作为起始材料,外周血单核细胞(PBMC)中H3K4me3的ChIP-标签化概况的代表性UCSC基因组浏览器屏幕截图。

[0188] (b) 使用不同量的纯化DNA作为起始材料,比较标准ChIP-seq(来自1000万个细胞)和外周血单核细胞(PBMC)中H3K4me3的ChIP标签化的成对散点图。

[0189] 图12、使用本发明的方法在PU.1结合位点的不同实验的PU.1读段计数。点划线和虚线分别显示来自使用500k或10mio细胞的实验的数据。M-点划线显示来自使用500k细胞的实验和实施例13中描述的示例性优化设置的PU.1信号。与使用标准方案用10mio细胞的实验相比,优化的方案给出了更高的信噪比。

[0190] 图13、使用本发明的方法在注释转录开始时以超快方式对来自不同实验的H3K27ac读段计数。虚线和点划线对应于实施例14中描述的超快方案,而直线显示使用标准方案的实验。超快方案(其也使用实施例13的优化方案)给出了与使用标准方案用10mio细胞的实验相比相同或更好的信噪比。

[0191] 图14、在一天内生成测序就绪的库。使用500k K562细胞的超快方案,使得能够在一天内为组蛋白标记和转录因子生成测序就绪的库。右边的顶部轨迹显示H3K27ac的密集信号,对应于K562细胞中珠蛋白基因座的超级增强子。

[0192] 图15、使用Tn5转座酶和内部制备的转座酶获得的测序文库制剂的比较。使用H3K4me3抗体和市售的Tn5转座酶或具有由如SEQ ID NO:1和2所示的核酸序列编码的序列的内部转座子对白血病细胞系的细胞进行本发明的方法。可以看出,结果不依赖于使用的转座酶。

[0193] 本发明通过以下说明性非限制性实施例进一步描述,其提供对本发明及其许多优点的更好理解。包括以下实施例以说明本发明的优选实施方式。本领域技术人员应该理解,在下面的实施例中公开的技术代表本发明中使用的技术在本发明的实践中良好地起作用,因此可以认为构成其实践的优选模式。然而,本领域技术人员根据本公开应当认识到,可以在公开的具体实施方式中做出许多改变,并且仍然获得相似或类似的结果,而不偏离本发明的精神和范围。

[0194] 除非另有说明,否则使用重组基因技术的已建立方法如例如在 Sambrook, Russell “Molecular Cloning, A Laboratory Manual”, Cold Spring Harbor Laboratory, N.Y. (2001) 中所述,该文献通过引用整体 并入本文。

[0195] 本文引用了许多文件,包括专利申请,制造商手册和科学出版物。这些文献的公开内容虽然对于本发明的可专利性不被认为是相关的,但其全部内容通过引用并入本文。更具体地说,所有参考文献通过引用并入本文,其程度就好像每个单独文件被具体地和单独地指示通过引用并入。

[0196] 实施例1-ChIPmentation方案

[0197] 收获细胞并固定

[0198] 收获细胞,用PBS洗涤一次并在室温下用1%多聚甲醛在多达1.5ml PBS中固定10分钟。在室温下将甘氨酸加入到0.125M的最终量5分钟以 终止反应。在4°C下以500×g收获细胞10分钟,并用补充有1μM PMSF 的多达1ml冰冷的PBS洗涤两次。

[0199] 裂解和超声处理

[0200] 将沉淀用RIPA缓冲液(10mM Tris-HCl, pH8.0, 1mM EDTA, pH8.0, 140mM NaCl, 1% Triton x-100, 0.1% SDS, 0.1% DOC, 1x蛋白酶抑制剂 (Sigma))裂解并在1ml milliTUBE中在Covaris S220中超声处理30分钟,直到大多数片段长200-700个碱基对(设置:工作期5%, 峰值事件功率140瓦, 对于K562细胞每次爆发周期200)。将裂解物在4°C下全速离心5分钟。然后将含有剪切的染色质的上清液转移到0.5PCR管并保持在 冰上。

[0201] 准备用于IP的珠

[0202] 与超声处理平行,将50μl磁性蛋白质A/G珠(10μl用于低输入的 ChIPmentation)封闭,并通过洗涤并在PBS, 0.5% BSA, 0.5% Tween-20 中重悬2次而与抗体缀合。通过在室温下旋转>1h(或在4°C下>2h)添加 抗体并结合至珠。

[0203] 免疫沉淀和洗涤

[0204] 每ChIP加入50μl封闭抗体缀合的磁性蛋白质A珠,并在4°C温育3 小时。随后用冷的150μl RIPA(两次), RIPA-500(10mM Tris-HCl, pH8.0, 1mM EDTA, pH8.0, 500mM NaCl, 1% Triton x-100, 0.1% SDS, 0.1DOC) (2次) 和RIPA-LiCl(10mM Tris-HCl, pH8.0, 1mM EDTA, pH8.0, 250mM LiCl, 1% Triton X-100, 0.5%DOC, 0.5%NP40(两次)洗涤免疫 沉淀珠。

[0205] 标签化-文库制备

[0206] 用冷Tris-C1 pH8.0将珠洗涤两次以除去去污剂、盐和EDTA。接下来, 将珠重悬于30μl含有来自Nextera DNA样品制备试剂盒(Illumina)的1μl Tagment DNA酶的标签化反应混合物(10mM Tris pH 8.0, 5mM MgCl) 中, 并在热循环仪中于37°C温育10分钟。然后将珠置于磁体上以去除标签化反应物, 然后用RIPA洗涤2次。

[0207] 彻底清洗并洗脱DNA,然后进行逆转交联

[0208] 最后用TE pH 8.0将珠洗涤两次。为了洗脱复合物, 将珠与含有2μl 蛋白酶K (NEB) 的70μl洗脱缓冲液(0.5% SDS, 300mM NaCl, 5mM EDTA, 10mM Tris-HCl pH 8.0) 在55°C温育1小时, 在65°C下8小时以 逆转甲醛交联, 并将上清液转移到新管中。

[0209] 纯化DNA

[0210] 最后,用AMPure XP珠(样品:珠比率1:2) 或Qiagen MinElute 柱纯化DNA。

[0211] 扩增文库

[0212] 在含有0.15 μ M引物(参见Buenrostro等人,Nature Methods-the original ATAC-seq publication-for primer sequences),1x SYBR green和5 μ lKAPA HIFI 2x预混物的10 μ l qPCR反应物中,扩增1 μ l的每个 ChIPmentation反应物,用以下程序估算最佳富集循环次数:72°C 5min, 98°C 30s,24个循环的98°C 10s,63°C 30s,72°C 30s,以及72°C 最终延伸1min。将KAPA HIFI 2x预混物在98°C温育45s,然后制备PCR反应物 以激活用于第一PCR步骤中成功缺口平移的热启动酶。使用0.75 μ M引物 和25 μ l KAPA HIF 2x预混物在50 μ l反应物中进行文库的最终富集。将文 库扩增N个循环,其中N等于qPCR反应中确定的四舍五入Cq值。

[0213] 文库的纯化和大小选择(可选的)

[0214] 富集的文库用大小选择程序使用比率为0.7:1的SPRI AMPure XP珠 (珠:样品) 纯化以除去长片段(>600bp),以2:1(珠:样品)的比率 回收反应物中的剩余DNA。测序由位于CeMM的生物医学测序设施使用 Illumina HiSeq 2000/2500平台进行。

[0215] 实施例2-比较ChIP-Seq、ChIP标签化和ChIPmentation

[0216] 细胞培养和样品采集

[0217] 在补充有10%FCS和抗生素的RPMI培养基中培养K562细胞。用 CASY细胞计数器分析它们以确定细胞数量。经维也纳医科大学伦理委员 会批准,从健康志愿者获得外周血。用EDTA或肝素预防凝血,外周血在 PBS中以1:1 - 1:3稀释,并按照制造商的说明用Lymphoprep密度梯度 (Axis-Shield) 分离外周血单核细胞(PBMC)。将纯化的细胞悬浮于补充有10%FBS和青霉素-链霉素的RPMI中。

[0218] 染色质免疫沉淀

[0219] 与实施例3至5中详细描述的用于进行染色质免疫沉淀的三种不同方 案组合测试了ChIPmentation。

[0220] 标准ChIP-seq文库制备

[0221] 根据制造商的说明,使用NEBNext End Repair Module (NEB) 对纯 化的ChIP DNA进行末端修复。根据制造商的说明使用Ampure XP珠 (Agencourt) 进行清理。使用Klenow (3' → 5' 外切) 聚合酶(Enzymatics) 对片段进行A加尾,并且使用T4DNA连接酶(Enzymatics) 连接TruSeq 相容的衔接头。使用Ampure XP珠来选择最终的文库以去除衔接头二聚 体。

[0222] ChIPmentation文库制备

[0223] ChIPmentation与ChIP的各种不同方案兼容,这使得可以容易地将 ChIPmentation应用于最适合不同ChIP方案的抗体。一般而言,进行选 择的ChIP方案,直到携带免疫沉淀的染色质的珠用含LiCl的洗涤缓冲液 洗涤(如实施例3中用于ChIP的WBIII,如实施例4中用于ChIP的 RIPA-LiCl和如实施例5中用于ChIP的TF-WBIII)。然后用Tris-Cl pH8.0 将珠洗涤两次以除去去污剂,盐和EDTA。随后,将珠重悬于20-30 μ l含 有来自Nextera DNA样品制备试剂盒(Illumina)的1 μ l Tagment DNA酶 的标签化反应缓冲液(10mM Tris pH 8.0,5mM MgCl)中并在热循环仪 中于37°C温育10-20分钟。标签化后,随后用150 μ l的WBI (ChIP实施 例3),RIPA (ChIP实施例4) 或WBI (ChIP实施例5) 洗涤珠两次。之后,相应的ChIP方案继续进行最后一次珠洗涤,从珠洗脱,逆转交联 和DNA纯化。

[0224] 标签化反应的条件依赖于用于染色质分离的活性剂而变化。标签化条件在温度(例如4°C、16°C、55°C等),标签化时间(例如1、2、3、5、15、20、30或60分钟等),Tegment DNA酶浓度(0.001、0.01、0.1、0.2、0.5、1.5、2、3、4、5或10μl等)和反应体积(0.001、0.01、0.1、0.5、1、5、10、20、50、100或200μl等)方面变化。此外,标签化反应缓冲液变化,并且还可以包含添加剂,包括去污剂、盐、溶剂等(例如标签化反应缓冲液可以包含二甲基甲酰胺、聚乙二醇、乙酸锰(II)等)。

[0225] ChIP标签化文库制备

[0226] 使用Qubit荧光计测量来自外周血单核细胞(PBMC)中标准H3K4me3ChIP的纯化的ChIP DNA,然后在补充有0.1%Tween-20至100pg、10pg或2pg总DNA的10mM Tris-HCl pH8.5中稀释。在含有稀释的DNA,5μl 2x标签化缓冲液(Illumina)和1μl(100pg DNA)或0.5μl(10pg和2pg)1:10稀释的Nextera Tegment DNA酶(在预冷的TE/50%甘油中稀释)的10μl反应物中于55°C进行标签化反应5分钟。使用Nextera DNA样品制备试剂盒(Illumina)根据制造商的说明使用以下程序扩增标签化的DNA:72°C 5min, 98°C 30s, 14个循环的98°C 10s, 63°C 30s, 72°C 30s, 最后72°C延伸1min。使用SPRI AMPure XP珠以珠:样品比率为1.5:1来纯化文库。制备纯化的ChIP DNA或来自K562ChIP的去蛋白质化的输入DNA,与PBMCs一样略有修饰:取5ng ChIP DNA进行标签化反应,使用0.5μl的1:10稀释的Tn5酶在5μl反应物中于55°C 5分钟。用MinElute试剂盒(Qiagen)纯化DNA并用KAPA HIFI 2x预混物扩增。

[0227] 标准ChIP-seq、ChIP-标签化和ChIPmentation文库的扩增和测序

[0228] 在含有0.15μM引物,1x SYBR绿和5μl KAPA HIFI 2x预混物的10μl qPCR反应物中扩增1μl每种ChIPmentation反应物以估计富集循环的最佳次数,使用以下程序:72°C 5min, 98°C 30s, 24个循环的98°C 10s, 63°C 30s, 72°C 30s, 最终72°C延伸1min。将KAPA HIFI 2x预混物在98°C下温育45s,然后制备PCR反应物以激活用于第一PCR步骤中成功缺口平移的热启动酶。使用0.75μM引物和25μl KAPA HIFI 2x预混物在50μl反应中进行文库的最终富集。将文库扩增N个循环,其中N等于qPCR反应中确定的四舍五入Cq值。富集的文库用大小选择程序使用比率为0.7:1(珠:样品)的SPRI AMPure XP珠纯化以除去长片段(>600bp),以2:1(珠:样品)的比率回收反应中的剩余DNA。测序由位于CeMM的生物医学测序设施使用Illumina HiSeq 2000/2500平台进行(细节参见图8)。

[0229] ATAC-Seq

[0230] 开放染色质作图用转座酶可及染色质(ATAC-seq)的测定法进行,如前对K562细胞所述,进行较小的改动。在每个实验中,将 1×10^5 个细胞在50μl PBS中洗涤一次,重悬于50μl ATAC-seq裂解缓冲液(10mM Tris-HCl, pH 7.4, 10mM NaCl, 3mM MgCl₂和0.01%IGEPAL CA-630)并在4°C下离心10min。离心后,在转座酶反应混合物(12.5μL 2×TD缓冲液,2μL转座酶(Illumina)和10.5μL无核酸酶的水)中温育之前,在37°C下将沉淀物在50μl MgCl₂缓冲液(10mM Tris pH 8.0, 5mM MgCl₂)中简单洗涤30min。用MinElute试剂盒(Qiagen)纯化DNA后,将1μl洗脱的DNA用于qPCR反应以估计最佳扩增循环次数。文库扩增后进行SPRI大小选择以排除大于1200bp的片段。用Qubit荧光计(Life Technologies)测量DNA浓度。

[0231] 测序数据处理和生物信息学分析

[0232] 读段是用skewer修剪的。修剪的读段与人类基因组的hg19/GRCh37 装配使用 Bowtie2以“-非常灵敏”参数进行比对。对于ChIPmentation和 ATAC-seq数据,调整了读段开始位置以表示转位事件的中心。如前所述², 与正链对齐的读段偏移+4bp,并且与负链对齐的读段偏移-5bp。使用 MACS2调用ChIPmentation,ChIP-seq和ATAC-seq样品上的峰。对于 ChIP和ChIPmentation数据,使用200bp的带宽运行MACS2,并将匹配 的IgG对照作为独立用于生物学重复的背景。对于广泛的组蛋白标记 (H3K27me3、H3K36me3),提供了“--broad”,“--nomodel”,“-extsize 73” 和“--pvalue 1e-3”标记和论据。在确保重复之间的一致性之后,以与所述相同的方式对从合并生物重复中调用的峰进行下游分析。对于 ChIPmentation和ChIP-seq样品的相关性分析,计算全基因组的1,000bp 窗口中的读段计数并相对于非重复读段的总数进行归一化。计算Pearson 相关系数,并为所有窗口绘制信号的底数为2的对数。在后两种基于合并 生物学重复的情况下,进行了生物学重复之间,不同技术(ChIP-seq与 ChIPmentation)之间以及不同数目的细胞之间的比较。调用的峰之间的 比较是通过计算与另一个重复的峰重叠的顶部5%或25%峰的分数来完成 的。在ChIP-seq和ChIPmentation数据之间以及使用具有两个组合重复 的样品的不同数目的细胞产生的ChIPmentation样品之间进行相同的比 较。

[0233] 实施例3-与ChIPmentation兼容的示例性ChIP方案

[0234] 用PBS洗涤细胞一次,并在室温下用1%多聚甲醛在多达1ml PBS中 固定5分钟。加入甘氨酸以停止反应。在4℃下以500×g收获细胞10分钟 (随后的工作在冰上进行并且除非另有说明,否则使用冷的缓冲液和溶液) 并用补充有1μM PMSF的多达1ml冰冷的PBS洗涤两次。将沉淀在细胞 裂解缓冲液 (50mM HEPES/KOH pH 7.4,140mM NaCl,1mM EDTA, 0.5mM EGTA,10%甘油,0.5%NP-40,0.25%Triton X-100,1x蛋白酶抑 制剂(Sigma))在冰上裂解 10分钟。通过在4℃下以1,000xg离心裂解 的细胞10分钟来分离细胞核,弃去上清液,并将沉淀物重悬于超声缓冲液 (10mM Tris-HCl pH 7.6,1mM EDTA,0.1% SDS) 中并且在 Covaris S220 上130μl microTUBE(最多3×10⁶个细胞)中超声处理12分钟,直到大部分片段长200-700个碱基对 (设置:工作期2%,峰值事件功率105瓦特, 每次爆发周期200)。将裂解物在4℃下全速离心5分钟,并将上清液转移 到新管中。用组成为20mM HEPES,0.1% SDS,1%Triton X-100,150mM NaCl,1mM EDTA,0.5mM EGTA的缓冲液将裂解物调整至每IP 为200μl, 并与针对H3K4me3的抗体 (1μg/IP,Diagenode pAb-003-050) 或 H3K27me3 (1μg/IP,Diagenode pAb-195-050) 在旋转器上于4℃过夜温 育。将20μl蛋白质A(或蛋白质G,取决于使用的抗体)磁珠用0.1% BSA 的PBS封闭过夜,第二天在旋转器上于4℃加入到IP中2 小时以捕获免疫 沉淀的片段。随后用WBI (20mM HEPES,150mM NaCl,0.1% SDS,0.1 %DOC, 1%Triton X-100,1mM EDTA,0.5mM EGTA) (两次),WBII (20mM HEPES,500mM NaCl,0.1% SDS,0.1%DOC,1%Triton X-100, 1mM EDTA,0.5mM EGTA) (一次),WBIII (20mM HEPES, 250mM LiCl,0.5%DOC,0.5%NP-40,1mM EDTA,0.5mM EGTA) (一次) 和WBIV (20mM HEPES, 1mM EDTA,0.5mM EGTA) (两次) 洗涤 免疫沉淀的染色质。然后将珠与含有2μl蛋白酶K (NEB) 的70μl洗脱缓 冲液 (0.5% SDS,300mM NaCl,5mM EDTA,10mM Tris-HCl pH 8.0) 在55℃温 育1小时,65℃温育8小时以逆转甲醛交联,并将上清液转移 到新管中。将另外30μl洗脱缓 冲液加入珠中1分钟,合并洗脱液,并与另 外1μl蛋白酶K一起在55℃温育1小时。最后,用 SPRI AMPure XP珠 (样品与珠之比1:2) 或Qiagen MinElute柱纯化DNA。

[0235] 实施例4-与ChIPmentation相容的示例性ChIP方案

[0236] 用PBS洗涤细胞一次,并在室温下用1%多聚甲醛在多达1.5ml PBS 中固定10分钟。加入甘氨酸以停止反应。在4℃下以500×g收获细胞10 分钟(随后的工作在冰上进行并且除非另有说明,否则使用冷的缓冲液和 溶液)并用补充有1μMPMSF的多达1ml冰冷的PBS 洗涤两次。将沉淀 用RIPA缓冲液(10mM Tris-HCl, pH8.0, 1mM EDTA, pH8.0, 140mM NaCl, 1% Triton x-100, 0.1% SDS, 0.1% DOC, 1x蛋白酶抑制剂(Sigma)) 裂解并在1ml MilliTUBE中在Covaris S220中超声处理30分钟,直到大部分片段长200-700个碱基对(设置:工作期5%,峰值事件功率140瓦, 每次爆发周期200)。将裂解物在4℃下全速离心5分钟,并将含有超声处理过的染色质的上清液转移至新管中。平行地,将50μl(10μl低输入 ChIPmentation)磁性蛋白A或蛋白G珠(取决于所用抗体)封闭并通过 洗涤并在PBS, 0.5% BSA, 0.5% Tween-20中重悬两次与抗体缀合。通过 在室温下旋转>1小时添加抗体并结合到珠上。使用的抗体是H3K4me1 (1μg/IP, Diagenode pAb-194-050), H3K36me3 (1μg/IP, Diagenode pAb-192-050) 和REST (10μg/IP, Millipore 07-579)。然后将封闭的抗体缀合的珠放置在磁体上,除去上清液,并将超声裂解的裂解物加入到珠 中,随后在旋转器上在4℃温育3小时。随后用150μl RIPA(两次), RIPA-500 (10mM Tris-HCl, pH8.0, 1mM EDTA, pH8.0, 500mM NaCl, 1% Triton x-100, 0.1% SDS, 0.1% DOC,) (两次), RIPA-LiCl (10mM Tris-HCl, pH 8.0, 1mM EDTA, pH8.0, 250mM LiCl, 1% Triton X-100, 0.5% DOC, 0.5% NP40) 和TE pH 8.0(两次)洗涤珠。然后将珠与含有2μl 蛋白酶K (NEB) 的70μl洗脱缓冲液 (0.5% SDS, 300mM NaCl, 5mM EDTA, 10mM Tris-HCl pH 8.0) 在55℃温育1小时, 65℃温育8小时以 逆转甲醛交联,并将上清液转移到新管中。最后,用SPRI AMPure XP珠 (样品与珠之比1:2) 或Qiagen MinElute柱纯化DNA。

[0237] 实施例5-与ChIPmentation相容的示例性ChIP方案

[0238] 将细胞用PBS洗涤一次,并用1%多聚甲醛在1.5ml PBS中在室温下 固定5-10分钟。加入甘氨酸以停止反应。在4℃下以500×g收获细胞10 分钟(随后的工作在冰上进行并且除非另有说明,否则使用冷的缓冲液和 溶液)并用补充有1μMPMSF的多达1ml冰冷的PBS 洗涤两次。将沉淀 用缓冲液L3B (10mM Tris-HCl, pH8.0, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% 脱氧胆酸钠, 0.5% N- 月桂酰肌氨酸, 1x蛋白酶抑制 剂(Sigma)) 裂解并在 Covaris S220中在1ml milliTUBE中超声处理20 分钟,直到大部分片段长200-700个碱基对(设置:工作期5%,峰值事件 功率140瓦,每次爆发周期200)。用1%Triton-X-100补充裂解物并在4℃ 下全速离心5分钟,并将含有超声处理的染色质的上清液转移到新管中。平行地,封闭珠并通过在含有0.5%BSA的PBS中洗涤它们两次并且在 200μl含有0.5%BSA的 PBS中每IP重悬50μl的磁性蛋白A或蛋白G珠 (10μl珠用于低输入ChIPmentation) (取决于使用的抗体),与抗体缀 合。通过在室温下旋转>1小时或在旋转器中于4℃旋转2小时添加抗体并 结合至珠。使用的抗体是H3K27ac (2μg, Diagenode pAb-196-050), PU.1 (5μg/IP, Santa Cruz sc-352), CTCF (10μl/IP, Millipore 07-729) 和 GATA1/IP (4μl/IP和2μg低输入, Abcam ab11852)。将封闭的抗体缀合 磁珠添加到含有染色质的管中,并在4℃温育3小时。随后用150μl TF-WBI (20mM Tris-HCl/pH 7.4, 150mM NaCl, 0.1% SDS, 1% Triton X-100, 2mM EDTA) (两次), TF-WBIII (250mM LiCl, 1% Triton X-100, 0.7 %DOC, 10mM Tris-HCl, 1mM EDTA) (2次) 和TET (0.2%Tween-20, 10mM Tris-HCl/pH8.0, 1mM EDTA) (2次) 洗涤

珠。然后将珠与含有2μl 蛋白酶K (NEB) 的70μl洗脱缓冲液 (0.5% SDS, 300mM NaCl, 5mM EDTA, 10mM Tris HCl pH 8.0) 在55°C温育1小时, 并在65°C温育8小时以逆转甲醛交联, 并将上清液转移至新管中。将另外30μl洗脱缓冲液加入珠中1分钟, 合并洗脱液并与另外1μl蛋白酶K在55°C下温育1小时。最后, 用SPRI AMPure XP珠 (样品与珠之比1:2) 或Qiagen MinElute 柱纯化DNA。

[0239] 实施例6-示例性超声处理设置

[0240] 对于K562白血病细胞系, 在室温下用1%甲醛进行10分钟固定。将超声处理缓冲液 (10mM Tris-HCl pH 7.6, 1mM EDTA, 0.1% SDS) 中的染色质在Covaris S220 (或类似版本) 上的130μl Covaris microTUBE (多达 3×10^6 个细胞) 中超声处理10-15分钟, 设置: 工作期2%, 峰值事件功率105瓦, 每次爆发周期200, 推荐水温最高8°C, 降压泵打开。作为第二个实例, 染色质可以在RIPA缓冲液 (10mM Tris-HCl, pH8.0, 1mM EDTA, pH8.0, 140mM NaCl, 1% Triton X-100, 0.1% SDS, 0.1% DOC, 1x蛋白酶抑制剂 (Sigma)) 中, 并且在Covaris S220 (或机器的类似版本) 中在1ml Covaris milliTUBE中超声处理25-30分钟, 设置: 工作期5%, 峰值事件功率140瓦, 每次爆发周期200, 推荐水温最高8°C, 减压泵开启。

[0241] 作为实例的生物加速器 (Diagenode), 可以使用其他超声处理装置: 作为适用于几种细胞系的示例性超声处理设置, 染色质可以在裂解缓冲液 (10mM Tris-HCl, pH8.0, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% 脱氧胆酸钠, 0.5% N-月桂酰肌氨酸, 1x蛋白酶抑制剂 (Sigma P8340) , 以设置“高”在Eppies (50μL-200μL) 中2次15分钟, 超声循环: 30秒ON/30秒OFF或在带有共振器的15ml锥形管 (500μl-1.5ml) 中15分钟, 超声循环: 30秒ON/30秒OFF。

[0242] 用于超声处理的另一个装置是探针超声波仪。作为实例, 裂解缓冲液 (10mM Tris-HCl, pH8.0, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% 脱氧胆酸钠, 0.5% N-月桂酰肌氨酸, 1x蛋白酶抑制剂 (Sigma P8340)) 中的染色质可以在~12W输出的情况下在冰上进行6次10秒的超声波处理, 在超声循环之间推荐的30秒停顿以防止使用具有微尖端探针的Branson Sonifier 450过热染色质。

[0243] 实施例7-原发性肿瘤分析

[0244] 使用切片机将冷冻的肿瘤片切成50μm切片, 并转移至冰上的反应管 (取决于肿瘤的大小, 20-50切片足以用于多组蛋白ChIPmentation反应)。将切片用PBS洗涤一次, 并使用1%多聚甲醛在1.5ml PBS中在室温下固定10分钟。加入甘氨酸以停止反应。在4°C以500×g收获细胞10分钟 (除非另有说明, 否则随后的工作在冰上进行并且缓冲液和溶液被冷却), 并用补充有1μM PMSF的多达1ml冰冷的PBS洗涤两次。将沉淀物在缓冲液L3B (10mM Tris-HCl, pH 8.0, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% 脱氧胆酸钠, 0.5% N-月桂酰肌氨酸, 1x蛋白酶抑制剂 (Sigma)) 中裂解并在Covaris S220中的1ml milliTUBE中超声处理35分钟, 直到大多数片段长200-700个碱基对 (设置: 工作期5%, 峰值事件功率140瓦, 每次爆发周期200)。将裂解物补充1% Triton-X-100并在4°C下全速离心5分钟, 将含有超声处理的染色质的上清液转移至新管中。平行地, 将珠封闭并通过在含有0.5% BSA的PBS中洗涤它们两次并且在200μl含有0.5% BSA的PBS中每IP重悬50μl (10μl珠用于低输入ChIPmentation) 磁性蛋白质A或蛋白质G珠 (取决于所使用的抗体), 与抗体缀合。通过在室

温下旋转>1小时添加抗体并结合至珠。抗体的实例是H3K27ac (2 μ g, Diagenode pAb-196-050), PU.1 (5 μ g/IP, Santa Cruz sc-352), CTCF (10 μ l/IP, Millipore 07-729) 和GATA1 (4 μ g/IP, 2 μ g 用于低输入Abcam ab11852)。将封闭的抗体缀合的磁珠添加到含有染色质的管中,并在4℃下温育3小时。随后用150 μ l TF-WBI (20mM Tris-HCl/pH 7.4, 150mM NaCl, 0.1% SDS, 1% Triton X-100, 2mM EDTA) (两次) 和TF-WBIII (250mM LiCl, 1% Triton X-100, 0.7% DOC, 10mM Tris-HCl, 1mM EDTA) (两次) 洗涤珠。用冷的Tris-C1 pH8.0将珠洗涤两次以除去去污剂、盐和EDTA。将珠小心地重悬于30 μ l含有来自Nextera DNA样品制剂盒(Illumina)的1 μ l Tagment DNA酶的标签化反应混合物 (10mM Tris pH 8.0, 5mM MgCl₂) 中,并在37℃下在热循环仪中温育10分钟。通过将反应物置于磁体上并除去上清液来除去标签化反应物,并用TF-WBI将珠洗涤两次。用TET (0.2% Tween-20, 10mM Tris-HCl/pH8.0, 1mM EDTA) 洗涤珠 (两次)。然后将珠与含有2 μ l蛋白酶K (NEB) 的70 μ l洗脱缓冲液 (0.5% SDS, 300mM NaCl, 5mM EDTA, 10mM Tris HCl pH 8.0) 在55℃下温育1小时,并在65℃下温育8小时以逆转甲醛交联,并将上清液转移到新管中。将另外30 μ l洗脱缓冲液加入珠中1分钟,并将洗脱液合并,并与另外1 μ l蛋白酶K一起在55℃温育1小时。最后,用SPRI AMPure XP珠 (样品与珠之比1:2) 或Qiagen MinElute 柱纯化DNA,在11 μ l H₂O中洗脱。在含有0.15 μ M引物,1x SYBR绿和5 μ l KAPA HiFi 2x预混物的10 μ l qPCR反应物中扩增1 μ l每种ChIPmentation 反应物以估计富集循环的最佳数量,所述扩增用以下程序:72℃5min, 98℃30s, 24个循环的98℃10s, 63℃30s, 72℃30s, 最终72℃延伸1min。在制备PCR反应物前,将KAPA HiFi 2x预混物在98℃预温育45s,以激活用于第一个PCR步骤中成功的缺口平移的热启动酶。使用0.75 μ M引物和25 μ l KAPA HiFi 2x预混物在50 μ l反应物中进行文库的最终富集(使用来自ChIP的剩余10 μ l)。文库扩增N个循环,其中N等于qPCR反应中确定的四舍五入Cq值。富集的文库用大小选择程序使用比率为0.7:1 (珠:样品) 的SPRI AMPure XP珠纯化以去除长片段 (>600bp),以2:1 (珠:样品) 的比率回收反应物中的剩余DNA。使用 Illumina HiSeq 2000/2500 平台进行测序。

[0245] 实施例8-在来自临床标本或其他来源的福尔马林固定的,石蜡包埋的样品(FFPE样品)上的ChIPmentation

[0246] 用本领域已知的方法将样品进行福尔马林固定并石蜡包埋。为了在FFPE样品上使用本发明,在室温下通过在1mL hystolemon溶液中连续温育(各10分钟)(六至八次)进行组织样品切片的去石蜡化。然后通过从100% (绝对乙醇) 开始将乙醇浓度降低至95%、70%、50%和20%,用水作为最后步骤(每次再水合步骤在室温下5分钟)使样品再水合。将再水合的FFPE切片在旋转平台中于室温下在0.5mL透化缓冲液 [1x Tris-缓冲盐水 (TBS), 0.5% Tween20, 1mM PMSF和10 μ g/mL RNA酶A] 中温育30min。在+4℃以18,000 \times g离心5分钟后,将样品重悬于200 μ L消化缓冲液 [50mM Tris-HCl (pH 7.4), 0.32M蔗糖, 4mM MgCl₂, 1mM CaCl₂ 和0.1mM PMSF] 中。使用Labsonic L超声仪 (B.Braun, Biotech International) 通过温和超声将FFPE衍生的样品部分片段化,然后在37℃下用终浓度为1U/10 μ g染色质的微球菌核酸酶 (N.70196Y;USB) 消化1min。在+4℃下以18,000 \times g离心5min后,将样品重悬于200 μ L超声处理缓冲液 [1 \times TBS, 0.1% SDS和1mM Na2EDTA (pH 8.0)] 中并进一步片段化。在室温下以8,000 \times g离心5min后,收集第一上清液(体积 \approx 170 μ L)。将沉淀物用50 μ L超声缓冲液洗涤一次,涡旋5s,并且再次离心以获得第二上清液(达到 \approx 220 μ L)

的最终体积)。通过Qubit (Invitrogen) 荧光定量染色质。染色质的免疫选择在ChIP缓冲液 [30mM Tris-HCl (pH7.4) , 50mM NaCl, 5mM Na2EDTA和0.1mM PMSF] 中进行,对于每个测定使用260-600ng染色质(取决于每个实验中从FFPE样品提取的染色质的量 或进行ChIP测定的次数)并且在+4℃下在具有所需抗体的旋转平台中温育16h。将二十微升50% vol/vol的浆液rec-Protein G-Sepharose 4B Conjugate (在+4℃下与在ChIP缓冲液中的1mg/mL的BSA预温育16小时; Zymed) 加入到每个ChIP测定中并在+4℃温育3h。在4℃以2000xg 离心5min后,随后用2mL洗涤缓冲液A[50mM Tris-HCl (pH7.4) , 1 % TritonX-100, 50mM NaCl, 5mM Na2EDTA, 和0.1mM PMSF] 和2mL 洗涤缓冲液B[50mM Tris-HCl (pH 7.4) , 1% TritonX-100, 100mM NaCl, 5mM Na2EDTA和0.1mM PMSF] 洗涤沉淀物。用冷的Tris-C1 pH8.0将珠洗涤两次以除去去污剂、盐和EDTA。将珠小心地重悬于30μl含有来自 Nextera DNA样品制备试剂盒(Illumina)的1μl Tagment DNA酶的标签化反应混合物(10mM Tris pH 8.0, 5mM MgCl)中,并在37℃下在热循环仪上温育10分钟。通过将反应物置于磁体上并除去上清液来去除标签化反应物,并用洗涤缓冲液A和10mL洗涤缓冲液C[50mM Tris-HCl (pH 7.4) , 1% TritonX-100, 150mM NaCl, 5mM Na2EDTA和0.1mM PMSF] 洗涤珠两次。通过加入200μL洗脱缓冲液[1×Tris-EDTA (TE) /1% SDS] 并在室温下在旋转平台中温育30分钟来进行洗脱。在室温下以1,200×g 离心5min后,保存上清液并重复洗脱以获得400μL的最终体积(结合部分)。DNA分离。通过在65℃在洗脱缓冲液/0.2M NaCl中过夜温育,然后用 80μg/mL蛋白酶K(在+45℃下3h) 消化来进行去交联。用三分之一体积 的苯酚:氯仿(1:1),一体积苯酚:氯仿(1:1)和一体积氯仿依次萃取 来分离DNA。DNA在-20℃过夜沉淀。离心后,将DNA沉淀物重悬于11μL 的LTE缓冲液(储存于-20℃)中。在含有0.15μM引物,1x SYBR绿和5μl KAPA HIFI 2x预混物的10μl qPCR反应物中扩增1μl每种ChIPmentation 反应物以估计富集循环的最佳数量,使用以下程序:72℃5min, 98℃30s, 24个循环的98℃10s, 63℃30s, 72℃30s, 最终72℃延伸1min。在制备 PCR反应物前,将KAPA HIFI 2x预混物在98℃预温育45s,以激活用于第一个PCR步骤中成功缺口平移的热启动酶。使用0.75μM引物和25μl KAPA HIFI 2x预混物在50μl反应物中进行文库的最终富集(使用来自 ChIP的剩余10μl)。文库扩增N个循环,其中N等于qPCR反应中确定 的四舍五入Cq值。富集的文库用大小选择程序使用SPRI AMPure XP珠 以比率为0.7:1(珠:样品)纯化以去除长片段 (>600bp),以2:1(珠: 样品)的比率回收反应中的剩余DNA。使用Illumina HiSeq 2000/2500平台进行测序。

[0247] 实施例9-人类白血病的ChIPmentation

[0248] B细胞慢性淋巴细胞白血病(B-CLL) 也称为慢性淋巴性白血病 (CLL),是成人中最常见的白血病类型(白血细胞癌的一种类型)。CLL 影响源自骨髓的B细胞淋巴细胞,在淋巴结中发展,并且通常通过产生抗体来抵抗感染。

[0249] 通过Robosep分离原发患者CLL样品和CD2+细胞的阴性选择

[0250] 将丁酸钠(NaB) 加入到新鲜的外周血中至最终浓度为5mM。将其在 室温下铺在等体积的Ficoll(GE Healthcare,Amersham,UK) 上,并在 4℃以13.8g离心20min,无制动。提取PBMC层并在20mL含有5mM NaB 的PBS中洗涤两次。获得的细胞在完全培养基中洗涤并重悬于250μl Robosep缓冲液中,并根据制造商的说明(EasySep Human CD2Positive Selection Kit,目录号18657;StemCell Technologies,Grenoble,France) 设置用于阴性选择的Robosep仪器。收集不结合柱的细胞(即CD2-群), 重悬于培养基中并直接用于本发

明中描述的免疫沉淀程序。

[0251] 实际上,可以使用本领域已知的方法冷冻来自患者的衍生的白血病样 品。

[0252] 冷冻细胞

[0253] 细胞应以 1×10^8 个细胞/ml浓度冷冻。在5min的时间段中向白血病细 胞(最大浓度 1×10^8 细胞/ml,冷冻培养基RPMI/50%FCS+10%DMSO的 最终浓度)中滴加并重复混合等 体积的RPMI/50%FCS+20%DMSO。将 1ml冷冻培养基中的细胞转移到无菌标记的冷冻管中。将冷冻管转移至室 温下的冷冻保存箱中,并将冷冻保存箱置于-80°C冰箱中。在24小时内 将 冷冻管转移到氮气罐中。

[0254] 解冻细胞

[0255] 在37°C水浴中迅速解冻细胞。打开前用70%乙醇擦拭小瓶。立即将 1ml细胞转移 到含有37°C预热的RPMI/10%FCS的15ml试管中。在室温 下以 $250 \times g$ 离心细胞5分钟。在2ml 37°C预热的RPMI/10%FCS中小心 地重悬细胞沉淀物。继续免疫沉淀。

[0256] 免疫沉淀

[0257] 用PBS洗涤细胞一次,并在室温下用1%多聚甲醛在多达1ml PBS中 固定5分钟。加 入甘氨酸以停止反应。在4°C下以 $500 \times g$ 收获细胞10分 钟(除非另有说明,否则随后的工作 在冰上进行并且使用冷缓冲液和溶液) 并用补充有 $1\mu M$ PMSF的多达1ml冰冷的PBS洗涤两 次。将沉淀物在细 胞裂解缓冲液(50mM HEPES/KOH pH 7.4,140mM NaCl,1mM EDTA, 0.5mM EGTA,10%甘油,0.5%NP-40,0.25%Triton X-100,1x蛋白酶抑 制剂(Sigma))中在冰上裂 解10分钟。通过将裂解的细胞在4°C以 $1,000 \times g$ 离心10分钟来分离核,弃去上清液,并将沉 淀物重悬于超声缓冲液(10mM Tris-HCl pH 7.6,1mM EDTA,0.1%SDS)并在Covaris S220 上的130μl microTUBE(多达 3×10^6 个细胞)中超声处理15分钟,直到大部分片段长 200- 700个碱基对(设置:工作期2%,峰值事件功率105瓦特,每次爆发 周期200)。将裂解物在4 °C下全速离心5分钟,并将上清液转移到新管中。用20mM HEPES,0.1%SDS,1%Triton X- 100,150mM NaCl,1mM EDTA,0.5mM EGTA的缓冲液组合物将裂解物调节至每IP 200μl,并与 选择的抗体一起在旋转器上于4°C温育过夜。用PBS中的0.1%BSA将20μl 蛋白质A(或蛋白 质G,取决于所用抗体)磁珠封闭过夜,第二天在旋转 器上于4°C加入到IP中2小时以捕获免 疫沉淀的片段。随后用WBI (20mM HEPES,150mM NaCl,0.1%SDS,0.1%DOC,1%Triton X- 100,1mM EDTA,0.5mM EGTA) (两次),WBII (20mM HEPES,500mM NaCl, 0.1%SDS,0.1%DOC, 1%Triton X-100,1mM EDTA,0.5mM EGTA) (一次) 和WBIII (20mM HEPES,250mM LiCl, 0.5%DOC,0.5%NP-40, 1mM EDTA,0.5mM EGTA) (一次) 洗涤免疫沉淀的染色质。用冷的 Tris-C1 pH8.0将珠洗涤两次以除去去污剂、盐和EDTA。将珠小心地重悬 于30μl含有来自 Nextera DNA样品制备试剂盒(Illumina)的1μl Tagment DNA酶的标签化反应混合物(10mM Tris pH 8.0,5mM MgCl)中,并在 37°C下在热循环仪上温育10分钟。通过将反应物置于磁 体上并除去上清液 来除去标签化反应物,并用WBI洗涤珠两次。用WBIV (20mM HEPES, 1mM EDTA,0.5mM EGTA)洗涤珠(两次)。然后将珠与含有2μl蛋白 酶K (NEB) 的70μl洗脱缓冲液 (0.5%SDS,300mM NaCl,5mM EDTA, 10mM Tris-HCl pH 8.0) 在55°C温育1小时和65°C温育 8小时以逆转甲 醛交联,并将上清液转移至新管中。再向珠中加入30μl洗脱缓冲液1分钟, 将洗脱液合并,并与另外1μl蛋白酶K一起在55°C温育1小时。最后,用 SPRI AMPure XP珠 (样品-珠比率1:2) 或Qiagen MinElute柱纯化DNA。DNA在11μl的EB缓冲液(10mM Tris-HCl

pH 8.5) 中洗脱。在含有0.15 μ M 引物,1x SYBR绿和5 μ l KAPA HIFI 2x预混物的10 μ l qPCR 反应物中扩 增1 μ l每种ChIPmentation反应物以估计富集循环的最佳数量,使用以下 程序:72°C 5min,98°C 30s,24个循环的98°C 10s,63°C 30s,72°C 30s, 最终72°C 延伸1min。在制备PCR反应物前,将KAPA HIFI 2x预混合物 在98°C 预温育45s,以激活用于第一个PCR步骤中成功缺口平移的热启动 酶。使用0.75 μ M引物和25 μ l KAPA HIF 2x预混物在50 μ l反应物中进行 文库的最终富集(使用来自ChIP的剩余10 μ l)。文库扩增N个循环,其 中N等于qPCR 反应中确定的四舍五入Cq值。富集的文库用大小选择程 序使用SPRI AMPure XP珠以比率 0.7:1(珠:样品)纯化以去除长片段 (>600bp),以2:1(珠:样品)的比率回收反应物中的剩余DNA。使用 Illumina HiSeq 2000/2500平台进行测序。

[0258] 实施例10-在细胞系上的ChIPmentation

[0259] K562细胞是第一个要建立的人永生髓系白血病细胞系。K562细胞属 于红白血病类型,并且该细胞系来源于处于原始细胞危象的53岁女性 CML患者。细胞是非粘附且圆形的,对于bcr:abl融合基因是阳性的, 并且与未分化的粒细胞和红细胞都具有一些蛋白质组相似性。

[0260] 用PBS洗涤细胞一次,并在室温下用1%多聚甲醛在最多1ml PBS中 固定5分钟。加入甘氨酸以停止反应。在4°C下以500xg收获细胞10分钟 (除非另有说明,随后的工作在冰上进行并且使用冷却缓冲液和溶液) 并 用补充有1 μ M PMSF的多达1ml冰冷的PBS洗涤两次。将沉淀在细胞裂 解缓冲液(50mM HEPES/KOH pH 7.4,140mM NaCl,1mM EDTA,0.5mM EGTA, 10%甘油,0.5%NP-40,0.25%Triton X-100,1x蛋白酶抑制剂 (Sigma))在冰上裂解10分钟。通过将裂解的细胞在4°C以1,000xg离 心10分钟来分离核,弃去上清液,并将沉淀重悬于超声缓冲液(10mM Tris-HCl pH 7.6,1mM EDTA,0.1%SDS) 和在Covaris S220上的130 μ l microTUBE (多达 3×10^6 个细胞)超声处理15分钟,直到大部分片段长 200-700个碱基对(设置:工作期2%,峰值事件功率105瓦特,每次爆发 周期200)。将裂解物在4°C下全速离心5分钟,并将上清液转移到新管中。用20mM HEPES,0.1%SDS,1%Triton X-100,150mM NaCl, 1mM EDTA,0.5mM EGTA的缓冲液组合物将裂解物调节至每IP 200 μ l,并与 选择的抗体一起在旋转器上于4°C温育过夜。用PBS中的0.1%BSA将20 μ l 蛋白质A(或蛋白质G,取决于所用抗体)磁珠封闭过夜,第二天在旋转 器上于4°C加入到IP中2小时以捕获免疫沉淀的片段。随后用WBI (20mM HEPES,150mM NaCl,0.1%SDS,0.1%DOC,1%Triton X-100,1mM EDTA, 0.5mM EGTA) (两次),WBII (20mM HEPES,500mM NaCl, 0.1%SDS,0.1%DOC,1%Triton X-100,1mM EDTA,0.5mM EGTA) (一次) 和WBIII (20mM HEPES,250mM LiCl,0.5%DOC,0.5% NP-40, 1mM EDTA,0.5mM EGTA) (一次)洗涤免疫沉淀的染色质。用冷的 Tris-C1 pH8.0将 珠洗涤两次以除去去污剂、盐和EDTA。将珠小心地重悬 于30 μ l含有来自Nextera DNA样品 制备试剂盒(Illumina)的1 μ l Tagment DNA酶的标签化反应混合物(10mM Tris pH 8.0, 5mM MgCl) 中,并于 37°C下在热循环仪中温育10分钟。通过将反应物置于磁体上并除去上 清液 来除去标签化反应物,并用WBI洗涤珠两次。用WBIV (20mM HEPES, 1mM EDTA,0.5mM EGTA)洗涤珠 (两次)。然后将珠与含有2 μ l蛋白 酶K(NEB) 的70 μ l洗脱缓冲液(0.5%SDS, 300mM NaCl,5mM EDTA, 10mM Tris-HCl pH 8.0)在55°C温育1小时和65°C温育8小时以逆 转甲 醛交联,并将上清液转移至新管中。再向珠中加入30 μ l洗脱缓冲液1分钟, 将洗脱液 合并,并与另外1 μ l蛋白酶K一起在55°C温育1小时。最后,用 SPRI AMPure XP珠(样品-珠比

率1:2)或Qiagen MinElute柱纯化DNA。DNA在11 μ l EB缓冲液(10mM Tris-HCl pH 8.5)中洗脱。在含有0.15 μ M 引物,1x SYBR绿和5 μ l KAPA HIFI 2x预混物的10 μ l qPCR反应物中扩增1 μ l每种ChIPmentation反应物以估计富集循环的最佳数量,使用以下 程序:72°C 5min, 98°C 30s, 24个循环的98°C 10s, 63°C 30s, 72°C 30s, 最终72°C 延伸1min。在制备PCR反应物前,将KAPA HIFI 2x预混合物 在98°C 预温育45s,以激活用于第一个PCR步骤中成功缺口平移的热启动 酶。使用0.75 μ M引物和25 μ l KAPA HIF 2x预混物在50 μ l反应物中进行 文库的最终富集(使用来自ChIP的剩余10 μ l)。文库扩增N个循环,其 中N等于qPCR反应中确定的四舍五入Cq值。富集的文库用大小选择程 序使用SPRI AMPure XP珠以比率0.7:1(珠:样品) 纯化以去除长片段 (>600bp),以2:1(珠:样品)的比率回收反应中的剩余DNA。使用 Illumina HiSeq 2000/2500平台进行测序。

[0261] 实施例11 - 对可能具有低细胞数量的模式生物或模式生物的部分进行 ChIPmentation

[0262] 本发明的方法仅需要低输入量来分析基因组范围的组蛋白-DNA相互 作用。预计本发明允许分析由低细胞数组成的个体动物的早期发育阶段。一个实例是斑马鱼 (*Danio rerio*),它是一种属于鲤鱼目鲤科(Cyprinidae) 的热带淡水鱼。原产于喜马拉雅地区,它是一种流行的水族馆鱼,经常以 商品名斑马鱼 (*zebra danio*) 出售。斑马鱼也是科学研 究中重要的脊椎动 物模式生物。受精后,卵子分裂,4小时后胚胎已经由数千个细胞组成,这可能足以使用本发明分析单个胚胎中的组蛋白修饰。早期发育阶段的胚 胎汇集也可用于增加细胞数目,因此本发明可用于该细胞。当使用本领域 众所周知的方法分离期望的细胞类型/组织/发育阶段时,建议的方案可适用 于所有类型的其他生物体,例如小鼠。在某些情况下,很难一次获得该技 术所需的大量胚胎,即在过度表达或敲除实验中。在这些情况下,胚胎加 工和固定可以分批进行,可以在液氮中冷冻并在-80°C 储存直到收集所需的 胚胎总数。为确保在相同发育阶段收集所有胚胎,仅将斑马鱼雌性和雄性 交配15分钟,在具有胚胎培养基(E3培养基:5mM NaCl, 0.17mM KC1, 0.4mM CaCl₂, 0.16mM MgSO₄)的培养皿 中收集胚胎并将它们在28°C下 培育直至达到所需的发育阶段。将胚胎收集在0.50ml E3 中,并以30mg/ml 加入5 μ l链霉蛋白酶(Roche,参考号10165921001)。轻轻摇动并在28°C 温育胚胎。大约15分钟用于绒毛膜软化。用立体显微镜检查它们,直到检 测到没有绒毛膜的第一个胚胎。立即用E3培养基彻底冲洗胚胎(三次) 以完全去除链霉蛋白酶。要从绒毛膜释 放胚胎,用移液器将它们在E3培 养基中小心吸入。用移液器将胚胎转移至0.5ml管中并除去所有E3。向胚 胎中加入0.46ml E3和0.4ml 4%PFA (4%PFA (SIGMA P6148), 磷酸 盐缓冲液200mM, pH 7.4, NaOH 0.02N) 并在室温下轻轻摇动15min。加入甘氨酸(Merck, 1.00590.1000)至终浓度为0.125M以淬灭甲醛并在 室温下轻轻摇动5min。取出上清液并在 冰冷的1X PBS中冲洗胚胎三次。取出PBS并进行细胞裂解或在液氮中冷冻并将沉淀储存在-80°C。从现在 起在4°C的冷室里工作。彻底混合Dyna蛋白G磁珠。取10 μ l(每种抗体), 用1ml新鲜封闭液(1X PBS中的0.5% BSA;可在4°C保存一周)在1.5ml 安全锁管中洗涤。以3000rpm旋转3min收集珠。再次洗涤1.5ml封闭溶 液中的珠两次。每次洗涤后,用力地重悬珠。用磁力架(DYNAMag-Spin, Invitrogen 123.20D)收集珠并丢弃上清液。将珠重悬在10 μ L封闭溶液中, 并添加抗体。一些抗体的可能的量:1 μ l抗H3K4me1Ab (Diagenode, 目录号CS-037-100,浓度未测定), 1 μ l抗H3K4me3Ab (Diagenode, 目录 号pAb-003-050, 1.1 μ g/ μ l),

1μl抗H3K27ac抗体 (Abcam, Cat. No. ab4729, 0.80mg/ml) 和1μl抗H3K27me3 (Millipore 07-449, 1mg/ml)。在4℃温育抗体至少4h或在旋转平台上过夜温育。用磁力架收集珠并除去上清液。在冷室中用0.2ml封闭溶液洗涤珠。重复此步骤两次。在10μl封闭溶液中重悬珠。临用前将蛋白酶抑制剂 (**Complete** 片, Roche 11 697 498 001) 加入所有裂解缓冲液中。(

Complete 的50X存储液 (1片/ml 1X PBS) 可以保持在-20℃两个月)。在0.13ml细胞裂解缓冲液 (50mM Tris-HCl pH 7.5, 10mM EDTA, 1% SDS) 中重悬交联的胚胎。上下吸移并挤压胚胎以使其破裂。将试管置于冰上温育10min。在冰上放置样品15min, 并更新冰水浴。在 Covaris S220上的130μl microTUBE (Covaris, 多达 3×10^6 个 细胞) 超声处理10-60分钟 (取决于使用多少胚胎, 需要根据经验确定), 直到大部分片段为200-700碱基对长 (设置: 工作期2%, 峰值事件功率 105瓦, 每次爆发周期200)。

[0263] 加入2倍体积的IP稀释缓冲液 (16.7mM Tris-HCl pH 7.5, 167mM NaCl, 1.2mM EDTA, 0.01% SDS)。加入1% Triton X-100超声处理染色质。在4℃下以14,000rpm离心10min并将染色质转移到新管中。向每个等分的超声处理染色质中添加10μl的抗体/磁珠混合物。在4℃的旋转平台上过夜温育管。用磁力架从管中收集珠并除去上清液。向每个管中加入 0.2ml RIPA洗涤缓冲液 (50mM HEPES pH 7.6, 1mM EDTA, 0.7% DOC, 1% Igepal, 0.5M LiCl)。轻轻摇动试管以重悬珠。用磁力架收集珠并除去上清液。再重复上一步三次。用冷的Tris-C1 pH8.0将珠洗涤两次以除去去污剂、盐和EDTA。将珠小心地重悬于30μl标记反应混合物 (这里: 来自 Nextera DNA样品制备试剂盒 (Illumina) 的2μl Tagment DNA酶, 来自 Nextera DNA样品制备试剂盒的15μl 2×Tagment DNA缓冲液和13μl无核酸酶水) 并在37℃下在热循环仪中温育3分钟。通过将反应置于磁体上并除去上清液来去除标签化反应物, 并用RIPA将珠洗涤两次。用1ml 1X TBS (50mM Tris pH 7.5, 150mM NaCl) 洗涤一次, 用磁力架收集珠并除去上清液。在200μl 1X TBS中重悬珠。以3000rpm旋转3min并吸出任何残留的TBS。加入60μl洗脱缓冲液 (50mM NaHCO3, 1% SDS)。在65℃从珠中洗脱DNA-蛋白质复合物10-15min, 每2分钟进行短暂涡旋。以14,000rpm离心珠1min。将650μl上清液转移至1.5ml安全锁管中。加入 300mM NaCl。在65℃下逆转甲醛交联6小时或过夜。加入RNase A至终浓度为0.33μg/μl, 并在37℃温育2h。加入1体积苯酚/氯仿/异戊醇 (25:24:1, AMRESCO 0883), 混合并旋转5min。将上层相转移至新的 1.5ml安全锁管。加入1μg糖原。加入1/10 3M NaAc和两倍体积的100% EtOH。以14,000rpm的速度旋转沉降DNA 10min。用500μl 75%冷-EtOH 洗涤沉淀物并在4℃以14,000rpm旋转5min。在室温下空气干燥沉淀物并重悬于70μl 10mM Tris-HCl, pH 8中。使用QIAquick PCR纯化试剂盒 (Qiagen 28104) 纯化DNA (来自输入和ChIP反应) (按照试剂盒提供的说明)。在11μl的EB缓冲液 (10mM Tris-HCl pH 8.5) 中洗脱。在含有0.15μM引物, 1x SYBR绿和5μl KAPA HIFI 2x预混物的10μl qPCR 反应物中扩增1μl每种ChIPmentation反应物以估计富集循环的最佳数量, 使用以下程序: 72℃5min, 98℃30s, 24个循环的98℃10s, 63℃30s, 72℃30s, 最终72℃延伸1min。在制备PCR反应物前, 将KAPA HIFI 2x 预混合物在98℃预温育45s, 以激活用于第一个PCR步骤中成功缺口平移的热启动酶。使用0.75μM引物和25μl KAPA HIFI 2x预混物在50μl反应物中进行文库的最终富集 (使用来自ChIP的剩余10μl)。文库扩增N个循环, 其中N等于qPCR反应中确定的四舍五入Cq值。富集的文库用大小选择程序使用SPRI AMPure XP珠以比率0.7:1(珠:样品) 纯化以去除长片段 (>600bp), 以2:1(珠:样品) 的比率回收反

应中的剩余DNA。使用Illumina HiSeq 2000/2500平台进行测序。

[0264] 实施例12-使用化学物质作为结合染色质的活性剂的本发明方法

[0265] 生物素化JQ1的体内全基因组占据分析(体内Chem-seq)。将指数生长的MM1.S细胞(每个样品 2×10^8 个细胞)与细胞培养基中的5 μ M生物素化JQ1(Bio-JQ1)或DMSO(载体)和1%甲醛同时处理20min。化学交联通过加入TRIS缓冲液(pH7.5)至终浓度为300mM TRIS终止。使用硅刮刀收获细胞,离心,并将得到的沉淀物用PBS洗涤三次。如下制备细胞核:细胞在50mM HEPES, pH7.5, 140mM NaCl, 1mM EDTA, 10%甘油, 0.5%NP-40, 0.25%Triton X-100加蛋白酶抑制剂混合物“完全”(Roche)中裂解,并且用10mM Tris-HCL, pH8.0, 200mM NaCl, 1mM EDTA, 0.5mM EGTA和蛋白酶抑制剂洗涤细胞核一次。在冰上在50mM HEPES-KOH, pH7.5, 140mM NaCl, 1mM EDTA, 1mM EGTA, 1%Triton X-100, 0.1%脱氧胆酸钠, 0.1%SDS(超声处理缓冲液)和蛋白酶抑制剂混合物中重悬并以18W超声处理细胞核进行10次循环(每次30s),循环之间间隔30s。通过离心澄清经超声处理的裂解物,并在4℃下用磁性链霉抗生物素蛋白Dynabeads (MyOne Streptavidin T1, Invitrogen)温育16-20h(在该温育步骤之前,珠在含有0.5%BSA的PBS中封闭)。在核超声裂解物中温育后,将珠在超声缓冲液中洗涤两次,在含有500mM NaCl的超声处理缓冲液中洗涤一次,在LiCl缓冲液(20mM Tris-HCl, pH8.0, 1mM EDTA, 250mM LiCl, 0.5%NP-40, 0.5%脱氧胆酸钠)中洗涤一次。用冷的Tris-C1 pH8.0将珠洗涤两次以除去去污剂、盐和EDTA。将珠小心地重悬于30 μ l标记反应混合物(这里:来自Nextera DNA样品制备试剂盒(Illumina)的2 μ l Tagment DNA酶,来自Nextera DNA样品制备试剂盒的15 μ l 2×Tagment DNA缓冲液和13 μ l无核酸酶水)并在37℃下在热循环仪中温育3分钟。通过将反应物置于磁体上并除去上清液来除去标签化反应,并用超声缓冲液将珠洗涤两次。然后用10mM TRIS, pH7.5, 0.1mM EDTA洗涤珠一次。随后将结合的蛋白-DNA复合物在50mM Tris-HCL, pH8.0, 10mM EDTA, 1%SDS中在65℃下洗脱15min,并通过在65℃过夜温育洗脱液来逆转交联。污染的RNA和蛋白质分别通过添加RNA酶和蛋白酶K进行消化,并如先前所述纯化DNA。最后,纯化的DNA片段大量平行测序。

[0266] 生物素化JQ1的体外全基因组占据分析(体外Chem-seq)。指数生长的未处理的MM1.S细胞用1%甲醛在细胞培养基中固定20min。终止化学交联,制备细胞核和超声处理如上所述的核裂解物。然而,与体内方案不同,链霉抗生物素蛋白Dynabeads在含有0.5%BSA以及或是200 μ M生物素化药物或是载体(DMSO)的PBS中预温育6h。随后在PBS/0.5%BSA中洗涤药物结合的珠四次以去除未结合的药物,并在4℃下在核超声裂解物中温育16-20h。以下所有步骤与上述相同(体内Chem-seq方法)。

[0267] 使用生物素化的AT7519的体外全基因组占据分析(体外Chem-seq)。指数生长的未处理的MM1.S细胞用0.5%甲醛在细胞培养基中固定5min。化学交联通过加入TRIS缓冲液(pH7.5)至终浓度为300mM TRIS终止。将细胞在PBS中洗涤3次,并如下制备细胞核:细胞核在50mM HEPES, pH7.5, 140mM NaCl, 1mM EDTA, 10%甘油, 0.5%NP-40, 0.25%Triton X-100加蛋白酶抑制剂混合物“完全”(Roche)中裂解,细胞核用10mM Tris-HCL, pH8.0, 200mM NaCl, 1mM EDTA, 0.5mM EGTA和蛋白酶抑制剂洗涤一次。在50mM HEPES-KOH, pH 7.5, 140mM NaCl, 1mM EDTA, 1mM EGTA, 0.5%NP-40, 0.5%Triton-X(超声缓冲液)中重悬细胞核并超声处理。在Misonix超声波仪上在冰上将沉淀物在9-12W超声处理4个循环(每次30s),循环之间具有1min的休息间隔。将药物结合的珠加入到澄清的超声处理物中,并使沉淀进行

12-18h。随后在超声缓冲液中将药物结合的珠洗涤三次。用冷的Tris-C1 pH8.0将珠洗涤两次以除去去污剂、盐和EDTA。将珠小心地重悬于30μl标记反应混合物(这里:来自Nextera DNA样品制备试剂盒(Illumina)的2μl Tagment DNA酶,来自Nextera DNA样品制备试剂盒的15μl 2×Tagment DNA缓冲液和13μl 无核酸酶水)并在37℃下在热循环仪中温育3分钟。通过将反应置于磁体上并除去上清液来除去标签化反应,并用超声缓冲液将珠洗涤两次。蛋白质在1% SDS中洗脱,并通过在65℃在1% SDS中温育洗脱液过夜来逆转交联。通过与RNA酶A和蛋白酶K连续温育消化污染的RNA和蛋白质,并如前所述纯化DNA。对纯化的DNA片段进行大规模平行测序。

[0268] 通过Chem-seq的生物素化补骨脂素的全基因组占据分析

[0269] 使用Nuclei EZ制备试剂盒(Sigma)由指数生长的MM.S细胞制备细胞核。然后将细胞核重悬于冰冷的PBS中,并直接与5μM生物素化的补骨脂素或载体(DMSO)一起在4℃温育30min。核在PBS中洗涤一次并立即在冰上在360nm下照射30min(Stratalinker)。在冰上在50mM HEPES-KOH, pH7.5, 140mM NaCl, 1mM EDTA, 1mM EGTA, 1% Triton X-100, 0.1% 脱氧胆酸钠, 0.1% SDS(超声缓冲液)和蛋白酶抑制剂混合物中重悬细胞核并在18W超声处理10次循环(每次30s),并在循环之间具有30s的间隔。通过离心澄清超声破碎的裂解物,并在4℃下用磁性链霉抗生物素蛋白Dynabeads (MyOne Streptavidin T1, Invitrogen)温育16-20h(在该温育步骤之前将珠在含有0.5% BSA的PBS中封闭)。在核超声裂解物中温育后,将珠在超声缓冲液中洗涤两次,在含有500mM NaCl的超声处理缓冲液中一次,在LiCl缓冲液(20mM Tris-HCl, pH8.0, 1mM EDTA, 250mM LiCl, 0.5% NP-40, 0.5% 脱氧胆酸钠)中一次。用冷的Tris-C1 pH8.0将珠洗涤两次以除去去污剂、盐和EDTA。将珠小心地重悬于30μl标记反应混合物(这里:来自Nextera DNA样品制备试剂盒(Illumina)的2μl Tagment DNA酶,来自Nextera DNA样品制备试剂盒的15μl 2×Tagment DNA缓冲液和13μl 无核酸酶水)并在37℃下在热循环仪中温育3分钟。通过将反应物置于磁体上并除去上清液来除去标记反应,并用超声缓冲液将珠洗涤两次,然后在10mM TRIS, pH7.5, 0.1mM EDTA中洗涤一次。然后将结合的蛋白-DNA复合物在50mM Tris-HCl, pH8.0, 10mM EDTA, 1% SDS和10mM生物素中洗脱,并将洗脱液在65℃过夜温育。通过分别加入RNA酶和蛋白酶K消化污染的RNA和蛋白质,并如前所述纯化DNA。最后,将纯化的DNA样品在254nm照射5min(Stratalinker)以逆转补骨脂素-DNA交联,然后进行库制备,大规模平行的DNA测序。

[0270] 实施例13-用于提高的信噪比的参数优化方案

[0271] 使用本发明的方法,可以优化多个参数以提高实验中的信噪比。本文提供的示例性方案通过使用管转移和特定标签化缓冲液将标签化时间减少至约1分钟。该方案与ChIP的各种方案兼容,分别与实施例3、4和5中描述的方案类似。这使得将方案应用于最适用于某个ChIP方案的抗体更容易。通过遵循实施例3、4或5的方案观察到最佳的信噪比,直至携带免疫沉淀的染色质的珠用含LiCl的洗涤缓冲液(实施例3的WBIII,实施例4的RIPA-LiCl和实施例5的TF-WBIII)洗涤。然后,用冷Tris-C1 pH8.0将珠洗涤一次以除去去污剂、盐和EDTA。随后,再次用冷Tris-C1 pH8.0洗涤珠,但不将反应物放置在磁体上而立即丢弃上清液。相反,包括珠的整个反应被转移到新的管中,然后放在磁铁上以除去上清液。这减少了粘附在管壁上的非特异性染色质片段的标签化。然后,将珠小心地重悬于含有1μl来自Nextera DNA样品制备试剂盒(Illumina)的Tagment DNA酶的25μl标签化反应混合物

(10mM Tris pH 8.0, 5mM MgCl₂, 10% w/v 二甲基甲酰胺) 并在热循环仪中37°C温育1分钟。二甲基甲酰胺作为极性非质子溶剂, 其可以增强亲核反应, 因此可能有利于转座反应, 因为转座酶 使用水分子用于亲核性连接作为将寡核苷酸整合到DNA中的机制。或者, 可以使用来自Nexera试剂盒的标记DNA缓冲液。去除标签化反应并用 WBI (实施例3), RIPA (实施例4) 或TF-WBI (实施例5) 洗涤珠两次。然后按照ChIP方案, 但如先前步骤中已经描述的那样, 当用WBIV (实施例3), TE (实施例4) 或TET (实施例5) 第二次洗涤时, 反应再次转移到新管中。这减少了粘附在管壁上的标记的非特异性片段的携带。使用500k细胞进行实验, 并与分别使用500k细胞或10Mio细胞、没有管转移也没有使用二甲基甲酰胺的标准方案比较。结果如图12所示。优化的方案与使用本文所述方法的标准方案使用10mio细胞的实验相比产生相等或更高的信噪比。

[0272] 实施例14-优化的测定持续时间

[0273] 为了进一步优化测定持续时间, 可以增加逆转交联的加热温度。因此, 本发明的方法通过“末端修复”的附加步骤在反向链上填充衔接头序列来进行, 例如, 在Nextera方案中, 建议在最初变性前进行5分钟72°C PCR步骤。为此, 加入PCR主混合物并将样品加热至72°C, 然后在95°C逆转交联。由于转座酶可以“粘”到DNA上, 方案建议在末端修复前用EDTA“剥离”(striping)转座酶, 用MgCl₂猝灭EDTA以允许末端修复(EDTA会干扰任何PCR)。进行了几种超快速方案, 包括包含ChIPmentation的方案直到最终洗涤步骤, 而不是立即用洗脱缓冲液从珠洗脱染色质, 通过直接添加用于末端修复的PCR主混合物来重悬珠(下文a); EDTA以确保从染色质完全剥离转座酶, 然后用MgCl₂猝灭, 然后加入用于末端修复的PCR主混合物(下文b); 或者如上所述, 只需要更快的程序(下文c)。随后, 反应物在72°C温育5分钟以修复末端(=填充第二链的衔接头末端), 交联在95°C下反转10-15分钟并且反应用新鲜PCR主混合物注满(或不在方案2c的情况下), 加入引物并扩增文库。在(a)的情况下, 遵循ChIPmentation程序, 直到最后一次洗涤携带染色质的珠。丢弃上清液并将珠重悬于冷的14μl 1x KAPA HiFi Hot Start ReadyMix(预热至95°C 30s)中。反应物在72°C温育5分钟, 然后在95°C温育10分钟。之后, 将反应物冷却至4°C。对于随后的PCR, 加入1.5μl正向和反向引物(各25μM), 15μl H₂O和18μl 2x KAPA HiFi Hot Start ReadyMix(预热至95°C 30s), 并扩增DNA。在(b)的情况下, 继续进行ChIP处理直到最后一次洗涤携带染色质的珠并丢弃上清液。将珠重悬于8μl 50mM EDTA中并在50°C温育30分钟。然后, 加入2μl 200mM MgCl₂并在50°C温育30分钟, 然后加入10μl 2x KAPA HiFi Hot Start ReadyMix(预热至95°C 30s)并在72°C下温育5分钟。随后, 将反应物在95°C温育10分钟, 然后冷却至4°C。最后, 加入1.5μl 25μM每种正向和反向引物, 12μl H₂O和15μl 2x KAPA HiFi Hot Start ReadyMix(预热至95°C 30s), 根据ChIPmentation文库扩增参数使用12-18个循环扩增反应。在(c)的情况下, 遵照ChIPmentation处理, 直到最后一次洗涤携带染色质的珠并丢弃上清液。然后, 加入11μl 20mM EDTA并在50°C温育10分钟。随后, 加入11μl 20mM MgCl₂+25μl 2x KAPA HiFi Hot Start ReadyMix(预热至95°C 30s)并在72°C温育5分钟。将反应物在95°C温育10分钟, 然后冷却至4°C。然后加入1.5μl 25μM每种正向和反向引物, 根据ChIPmentation文库扩增参数在12-18个循环中扩增反应。结果在图15中显示。因此, 使用如本文所述的超快速程序导致可比的数据有效性, 同时显著减少实验时间。特别是在(c)的情况下, 从细胞培养到扩增文库的整个程序可以在一个工作日内完成。更具体地

说,实验步骤包括收获细胞并制备珠(按照实施例5) (20min),用甲醛固定细胞并洗涤沉淀(45min),裂解和超声处理细胞(40min),用抗体分离染色质(=免疫沉淀步骤,当按照实施例5时为3小时),洗涤染色质(30min),加入转座酶(10min),随后洗涤(10min),末端修复和逆转交联(30min),扩增文库(45min)并纯化文库(45min),导致总时间<8小时。这里,从患者中分离特定细胞(例如CD4+T细胞)仍然可以纳入长时间工作日的时间线(抽血和CD4+分离为+90分钟)。如果第二个操作员正在准备合适的测序仪(例如Illumina MiSeq),则可以对样品进行过夜测序,从而产生从患者抽血到测序的约24小时的完整工作流程,允许在临床时间范围内个性化表观基因组。更具体地说,当使用MiSeq试剂盒v3在Illumina MiSeq上对ChIPmentation样品进行测序时,可以在5-6小时内完成来自约2500万个簇的50个测序循环,这导致从抽血至测序的ChIPmentation实验的总共~15小时的时间框。

[0274] 实施例15- 使用备选转座酶的超快测定

[0275] 根据Picelli等人(2014) *Genome Research* 24:2033-40制备备选转座酶。简而言之,根据Picelli等人(如上所述)并且使用具有SEQ ID NO: 3序列的表达载体,生产并储存Tn5酶,然后将Tn5的等分试样在Tn5稀释缓冲液(取决于Tn5的活性)中稀释,制备寡核苷酸以装载到Tn5上,然后用寡核苷酸装载稀释的Tn5(如前所述的所有步骤),所述Tn5酶具有由SEQ ID NO:1或2的序列编码的氨基酸序列,其中SEQ ID NO:1涉及编码含有C端内含肽标签和几丁质结合域的自制转座酶的核酸并且SEQ ID NO:2涉及核心转座酶。“自制”Tn5可以用作直接替代品。使用以下缓冲液:Tn5稀释缓冲液(50mM Tris-HCl, pH 7.5; 100mM NaCl; 0.1mM EDTA; 50%甘油; 0.1%Triton X-100和1mM DTT(在稀释前总是新鲜加入-每1ml缓冲液添加1μl 1M DTT)。对于嵌合末端寡核苷酸的预退火,按照下面的程序:(1)制备Tn5ME-A+Tn5MErev和Tn5ME-B+Tn5MErev寡核苷酸的100μM等摩尔混合物;使用Tn5MErev:5'-[磷酸]CTGTCTCTTATACACATCT-3'(SEQ ID NO:4);Tn5ME-A: 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3'(SEQ ID NO:5)和Tn5ME-B: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3'(SEQ ID NO:6),其中Tn5MErev在5'末端被磷酸化;(2)两种混合物在95°C温育3分钟并在关闭的热块中温育过夜以冷却,随后两者混合1:1(3)转座体的产生是通过1:1地向Tn5稀释液中加入0.143体积的预退火的寡核苷酸并在R/T温育1小时并在冰上温育直至需要。

[0276] 对于每个实验,将300,000个K562白血病细胞系细胞用PBS洗涤一次,并在室温下用1%多聚甲醛在1.5ml PBS中固定10分钟。加入甘氨酸以停止反应。在4°C下以500×g收获细胞10分钟(除非另有说明,随后的工作在冰上进行,使用冷的缓冲液和溶液)并用补充有1μM PMSF的多达1ml冰冷的PBS洗涤两次。将沉淀物用超声处理缓冲液(10mM Tris-HCl, pH 8.0, 1mM EDTA, pH 8.0, 0.25% SDS, 1x蛋白酶抑制剂(Sigma))裂解并在Covaris S220中在1ml milliTUBE中超声处理20分钟直至大多数片段长200-700个碱基对(设置:工作期5%,峰值事件功率140瓦特,每次爆发周期200)。将裂解物调整至RIPA条件(10mM Tris-HCl, pH 8.0, 1mM EDTA, pH 8.0, 140mM NaCl, 1% Triton x-100, 0.1% SDS, 0.1% DOC, 1x蛋白酶抑制剂(Sigma))。将裂解物在4°C下全速离心5分钟,并将含有超声处理过的染色质的上清液转移至新管中。平行地,将10μl磁性蛋白质A或蛋白质G珠(取决于使用的抗体)封闭并通过洗涤并且在PBS, 0.5% BSA, 0.5% Tween-20中重悬两次来与抗体缀合。通过在室温下旋转>1小时添加抗体并结合到珠上。使用的抗体是H3K4me3(1μg/IP, Diagenode)。然后

将封闭的抗体缀合的珠放置在磁体上,除去上清液,并 将超声裂解的裂解物加入到珠中,随后在旋转器上在4℃温育3小时。随后用150μl RIPA (两次), RIPA-500 (10mM Tris-HCl, pH8.0, 1mM EDTA, pH8.0, 500mM NaCl, 1% Triton x-100, 0.1% SDS, 0.1% DOC,) (两 次), RIPA-LiCl (10mM Tris-HCl, pH8.0, 1mM EDTA, pH8.0, 250mM LiCl, 1% Triton X-100, 0.5% DOC, 0.5% NP40) 和Tris pH8.0 (两次) 洗涤珠。将11μl 20mM EDTA加入珠中并在50℃下温育10分钟。随后, 加入11μl 20mM MgCl₂+25μl 2x KAPA HiFi Hot Start ReadyMix (预热至 95℃30s) 并在72℃温育5分钟。将反应物在95℃温育10分钟,然后冷却至4℃。然后加入1.5μl 25μM每种正向和反向引物,根据ChIPmentation 文库扩增参数在10-14个循环中扩增反应物。结果在图15中显示,显示了 使用市售Illumina Tn5转座体或来自2个不同来源的“自制”Tn5转座体对 H3K4me3的ChIP化测序文库。该实验使用市售的Illumina转座酶 (tn5) 或自制转座酶展示相同的结果。

序列表

<110> 分子医学研究中心责任有限公司

<120> 研究核酸的方法

<130> Y2067 PCT S3

<150> EP15 18 9788.1

<151> 2015-10-14

<150> EP15 18 0705.4

<151> 2015-08-12

<160> 6

<170> BiSSAP 1.3

<210> 1

<211> 2211

<212> DNA

<213> 人工序列

<220>

<223> 编码含有C-末端内蛋白标签和几丁质结合结构域（CBD）的自制转座酶的核酸序列

<400> 1

atgattacca	gtgactgca	tcgtcgccg	gattggcg	aaagcg	ttctagtg	60
gcgcgtgg	atccgcgt	taccgcgt	ctggtaat	ttgcggcg	actggccaa	120
tatagcgg	aaagcattac	cattagcag	gaaggcag	aagccatg	ggaaggcg	180
tatcg	tttta	tgcgttat	gaacgtg	gcggaa	ttcgtaa	240
cagaccgt	gaaactgg	ccca	gaaatttcc	gaactgct	ggatgaaga	300
ctgagctat	gtcatcagg	gtcggaa	ctggcaaa	tggtagcat	tcaggataaa	360
agccgtgg	ttgtgg	gtca	tagcgt	ctgctgg	cgaccac	420
ggcgtgc	atcaagaat	gtggatcg	ccggatg	cgccggat	ggatgaaaa	480
gaaagcgg	aatggctgg	cgctgtc	acttcgc	tgagaatgg	cagcatgat	540
agcaacgt	ttgcgtgt	cgatcg	cgatcgat	atcgatct	gcaagataaa	600
ctggccata	acgaacgtt	tgtgg	tcgt	agcaaacatc	cgctaaaga	660
ggcgtgtat	catacgtt	cctgaaaa	cagccgaa	tggcg	tcagattagc	720
atccgcaga	aaggcgtgg	ggataaacgt	ggcaaacgt	aaaaccgt	ggcgtaaa	780
gcgacgt	gcctgcgt	cgccgtatt	accctgaa	aggcaacat	taccctgaa	840
gcggcgt	ccgaagaaat	taatcccg	aaaggcgg	ccccgt	atggctgt	900
ctgaccagcg	agccgtgg	aagtctgg	caagcg	gtgtgatt	tatttatacc	960
catcggtgc	gcattgaa	atttcacaaa	gcgtgg	cgggtgcgg	tgcgaaacgt	1020
cagcgatgg	aagaaccg	taacctgg	cgtatgg	gcattctg	ctttgtggc	1080
gtgcgtctgc	tgcaactgc	tgaatctt	actccgc	aagcactgc	tgcgaggc	1140
ctgctgaaag	aagcggaa	cgttggaa	cagagcgg	aaaccgt	gacccggat	1200
gaatgccaac	tgctgg	tctggataaa	ggcaaacg	aacgcaa	aaaagcggc	1260
acgctgaat	ggcgtat	ggcgtat	cggtgg	gtttatgg	tagcaaacgt	1320
accggcatt	cgagctgg	tgcgtgt	gaagg	aagcgctg	aagcaaactg	1380
gatggctt	tggccg	agacctg	gcccgg	ttaaaatct	catcacgg	1440
gatgcactag	ttgcctt	cgaggc	cgag	tcggta	tcgcccacat	1500

[0001]

gcgcggccca acagtacaa cgcacatcgac ctgaaagtcc ttgaccggca tggcaatccc	1560
gtgtcgccg accggctgtt ccactccggc gagcatccgg tgtacacgggt gcgtacggc	1620
gaaggctctgc gtgtgacggg caccgcgaac caccgggtgt tgggtttgggt cgacgtcgcc	1680
ggggtgcgcga ccctgctgtg gaagctgatc gacgaaatca agccggcgaa ttacgggtg	1740
attcaacgca ggcattcag cgtcactgt gcagggttgc cccggggaa accggattt	1800
gcgcaccaa cctacacagt cggcgccct ggactgggtc gtttcttggaa agcacaccac	1860
cgagacccgg acgccaagc tatcgccgac gagctgaccc acggcggtt ctactacgc	1920
aaagtgcgcga gtgtcaccga cggcggtgt cagccgggtt atagccttcg tgcgacacgc	1980
gcagaccacg cgtttatcac gaacgggttc gtcagccacg ctactggcct caccggctg	2040
aactcaggcc tcacgacaaa tcctgggtta tccgcttggc aggtcaacac agcttatact	2100
gcgggacaat tggcacata taacggcaag acgtataaat gttgcagcc ccacacctcc	2160
ttggcaggat gggaccatc caacgttccgccttgc agcttcaatg a	2211

<210> 2

<211> 1428

<212> DNA

<213> 人工序列

<220>

<223> 编码自制转座酶的核酸序列

<400> 2

atgattacca gtgcactgca tcgtcgccg gattggcgaa aagcggttt ttcttagtgc	60
gcgcgtgggt atccgcgtcg taccgcgcgt ctggtaatgg ttgcggcga actggccaaa	120
tatagcgccaa aaagcattac cattagcagc gaaggcagca aagccatgca ggaaggcg	180
tatcgtttta ttcgtaatcc gaacgtgagc gcggaaagcgaa ttctggtaatgg gggccatg	240
cagaccgtga aactggccca ggaattcccg gaactgtgg caattgaaga taccacctct	300
ctgagctatc gtcatcagggt ggcggaaagaa ctggccaaac tgggttagcat tcaggataaa	360
agccgtgggtt ggtgggtgc tagcgtgctg ctgcgtggaa cggccacattt tcgtaccgt	420
ggccgtgtgc atcaagaatg gtggatgcgt cggatgatc cggccgtatgc ggtggaaaaaa	480
gaaagcggca aatggctggc cgctgtgcacttcgcgtc tgagaatggg cagcatgtg	540
agcaacgtga ttgcgggtgt cgatcgtgaa cggatatttc atgcgtatct gcaagataaa	600
ctggccata acgaacgttt tgggtgcgt agcaaacatc cggtaaaga tggaaagc	660
ggccgtatc tggatgatca cctgaaaaac cagccggaaac tggccgtata tcaggatgg	720
atccgcaga aaggcggtgtt ggataaacgtt ggcaaaacgtt aaaaccgtcc ggcgttaaa	780
gcgagcgtga ccgtcgtag cggccgtatt accctgaaac agggcaacat taccctgaa	840
gcgggtgtgg ccgaagaaat taatccggc aaaggcgaaa ccccgctgaa atggctgt	900
ctgaccagcg agccgggtga aagtctggcc caagcgctgc gtgtgattga tatttatacc	960
catcggttgc gcattgaaga atttcacaaa cgggtggaaa cgggtgcggg tgcggaaacgt	1020
cagcgatgg aagaaccgga taacctggaa cgtatggta cgtatgttgc tttgtggcg	1080
gtgcgtctgc tgcaactgca tggatctttt actccgcgc aagcactgca tgcgcaggc	1140
ctgctgaaag aagcggaaca cgttgaaagc cagagcgccgg aaaccgtgtt gacccggat	1200
gaatgccaac tgctgggtta tctggataaa ggcaaaacgc aacgcaaaaga aaaagcg	1260

[0002]

agcctgcaat gggcgtatat ggcgattgcg cgtctggcg gctttatgga tagcaaacgt	1320
accggcattt cgagctggg tgcgctgtgg gaagggtggg aagcgctgca aagcaaactg	1380
gatggcttcc tggcccgaa agacctgatg ggcgaggca taaaatc	1428
 <210> 3	
<211> 8079	
<212> DNA	
<213> 人工序列	
 <220>	
<223> 用于制备自制转座酶的载体的核酸序列	
 <400> 3	
aactacgtca ggtggcactt ttccggaaa tgtgcgcga acccctattt gtttattttt	60
ctaaaatacat tcaaataatgt atccgctcat gagacaataa ccctgataaa tgcttcaata	120
atattgaaaa aggaagagta tgagtattca acatttccgt gtgcgcctta ttccctttt	180
tgccgcattt tgccttcctg tttttgcata cccagaaacg ctggtaaaag taaaagatgc	240
tgaagatcgat ttgggtgcac gagtggttta catcgaaactg gatctcaaca gcgtaagat	300
cctttagatgtt ttccggcccg aagaacgttc tccaaatgtatg agcactttt aagttctgt	360
atgtggcgcg gtattatccc gtgttgacgc cggcaagag caactcggtc gcccataca	420
ctattctcg aatgacttgg ttgagtactc accagtcaaca gaaaagcata ttacggatgg	480
catgacagta agagaattat gcagtgcgtc cataaccatg agtgataaca ctgcggccaa	540
cttacttctg acaacgatcg gaggaccgaa ggagcttacc gcttttttgc acaacatgg	600
ggatcatgtt actcgccctt atcggtggg accggagctt aatgaagcca taccaaacga	660
cgagcggtgac accacgatgc ctgttagata ggcaacaacg ttgcgcacaa tattaaactgg	720
cgaactactt actctagctt cccggcaaca attaatagac tggatggagg cggataaaat	780
tgcaggacca ctctcgctt cggcccttcc ggctggctgg tttattgttgc attaaatctgg	840
agccggtagt cgtgggtctc ggggtatcat tgcagactg gggccagatg gtaagccctc	900
ccgtatcgta gttatctaca cgacggggag tcaggcaact atggatgaac gaaatagaca	960
gatcgcttagt ataggtgcct cactgattaa gcattggtaa ctgtcagacc aagttactc	1020
atataatactt tagattgatt taccgggtt gataatcaga aaagcccaa aaacaggaag	1080
attgtataag caaatattta aattgttaaaat gttatattt tgtaaaaattt cgcgttaaat	1140
ttttgttaaa tcaagtcatt ttttaaccaaa tagggcggaaa tcggcaaaat cccttataaa	1200
tcaaaagaat agcccgagat agggttgagt gttgttccag tttggaaacaa gagtccacta	1260
ttaaaagaacg tggactccaa cgtcaaaaggcg cggaaaacccg tctatcaggcg cgtggccca	1320
ctacgtgaac cateacccaa atcaagttt ttgggttgcg ggtgcgtaa agcactaaat	1380
cgaaacccta aaggggcccccc ccgatttgcgat gcttgcggg gaaagccggc gaaacgtggcg	1440
agaaaggaag ggaagaaagc gaaaggagcg ggcgctaggcg cgtggcaag tggatgggtc	1500
acgctgcgcg taaccaccac accccggcgctt cttatgcgc cgctacaggcg cgcgtaaaag	1560
gatcttagtg aagatccctt ttgataatctt catgaccaaa atcccttaac tggatgtttc	1620
gttccactga gcgtcagacc ccgtgaaaaat gatcaaagga tcttcttgcgatcctttt	1680
tctgcgcgtt atctgtgtt tgcaaaacaaa aaaaccacccg ctaccagccg tggttgttt	1740
gccggatcaa gagtacccaa ctctttcc gaaaggtaact ggcttcggca gaggccagat	1800

[0003]

[0004]	accaaatact gtccttctag tgttagccgt a gtttaggcac cacttcaaga actctgttagc accgcctaca tacctcgctc tgctaatcct gttaccatg gctgctgcca gtggcgataa gtcgtgtctt accgggttgg actcaagacg atagttaccg gataaggcgc agcggcggg ctgaacgggg ggttcgtgca cacagcccag cttggagcga acgacactaca cccgaaactgag atacctacag cgtgagctat gagaagcgc cacgcttccc gaagggagaa aggcggacag gtatccggta agcggcaggg tcggaacagg agagcgcacg agggagcttc cagggggaaa cgccctggat ctatagtc ctgtcggtt tcgcccaccc tgacttgcg gtcgattttt gtgatgctcg tcagggggc ggagcctatg gaaaaacgcc agcaacgcgg ccttttacg gttctggcc ttttgcgtgc cttttgcgtca catgttctt cctgcgttat cccctgattc tgtggataac cgtattaccg cctttgatgt agctgatacc gctgcccga gccgaacgac cgagcgcagc gagtcagtga gcgaggaagc tatggtgac tctcagataca atctgctcg atgcgcata gttaaagccag tatacactcc gctatcgta cgtgactggg tcatggctgc gcggcggacac cggccaaacac cggctgacgc gcccgtacgg gcttgcgtc tccggcattc cgcttacaga caagctgtga cggctccgg gagctgcatg tgtcagagggt tttcaccgtc atcaccgaaa cgcgcggagc agctcggtt aagctcatca gctggcgatc gcaagcgattc acagatgtct gcctgttcat cccgtccag ctgcgttgcgt ttctccagaa gcttaatgt ctggcttctg ataaagcggg ccatgttaag ggcggtttt tcctgtttgg tcaactgtatgc ctccgtgtaa gggggatttc tttcatggg ggtaatgata ccgtaaac gagagaggat gctcacgata cgggttactg atgatgaaaca tgcccggtt ctggaaacgtt gtgagggtaa acaactggcg gtatggatgc ggcgggacca gaaaaaaatc actcagggtc aatgccagcc gaacgcgcgc aagacgttagc ccagcgcgtc gggccgcattt cccgcgatata tggcctgctt ctggccgaaa cgttgggtt cgggaccagt gacgaaggct tggcgaggc cgtcaagat tccgaataacc gcaagcgaca gggccatcat cgtcgccgtc cagcgaaagc ggtctcgcc gaaaatgacc cagagcgcgtc cgggcacctgt tcctacgatg tgcgtatgata agaagacagt cataagtgcg ggcgcgtatgc tcatgccccg cggccacccgg aaggagctga ctgggttggaa ggctctcaag ggcgcgtc gagatccgg tgcctaatgatg tgcgtatgaa ttacattat tgcgttgcgc tcactgccccg ctggccatgtt gggaaacccgt tcgtgcgc tgcattatgt aatcggccaa cgcgcgggaa gggcggtt gcttgcgtt cggccagggtt gttttctt tcaccagtga gacggcaac agctgattgc ctttcacccgc ctggccctgt gatgttgc gcaagcggtc cacgtgggtt tggccatgtt ggcggaaatc ctgtttgtatg gtgggttacg gccccatata acatgagctg ttttcgtatcc cgtcgatcc cactaccgg atatccgcac caacgcgcgtc cccggactcg gtaatggcg gcattgcgc cagcgccatc tgcgttgg caaccagcat cgcgtggaa acgtggccct cattcagatcat ttgcgtatgtt tggaaac cgacatggc actccagtcg cttcccggtt cggctatcg ctggatgtt gatgttgc gatattttatg ccagccagcc agacgcagac ggcgcggagac agaacttaat gggccgtt acagcgcat tttgcgttgc cccatgttgc ccagatgttc cacggccatgtt cgcgttacgt cttcgttgcgtt gaaaataata ctgttgcgtt gatgttgcgtt agagacatca agaaataacg ccggaaacatt agtgcggca gcttccacagc caatggcatc ctggatcc acggatagt	1860 1920 1980 2040 2100 2160 2220 2280 2340 2400 2460 2520 2580 2640 2700 2760 2820 2880 2940 3000 3060 3120 3180 3240 3300 3360 3420 3480 3540 3600 3660 3720 3780 3840 3900 3960 4020 4080
--------	--	--

cgctaaaga tggaaagc ggcgttatac tgtatgatca cctgaaaaac cagccggac	6420
tggcgctca tcagattagc attccgcaga aaggcggtgt ggataaacgt ggcaaacgt	6480
aaaaccgtcc ggcgttaaa gcgagcctga gcctgcgtt cggccgtatt accctgaaac	6540
aggcaacat taccctgaac gcgggtgtgg ccgaagaaat taatccggc aaaggcgaaa	6600
cccgctgaa atggcgctg ctgaccagcg agccgtgga aagtctggcc caagcgctgc	6660
gtgtgattga tatttatacc catcggttgc gcattgaaga atttcacaaa gcgtggaaaa	6720
cgggtcgccc tgcggAACGT cagcgatgg aagaaccgga taacctggaa cgtatggta	6780
gcattctgag ctttgcggcg gtgcgtctgc tgcaactgcg tgaatcttt actccggcgc	6840
aagcaactgcg tgcgcaggc ctgcgtgaaag aagcggaaaca cgttggaaagc cagagcgccg	6900
aaaccgtgt gacccggat gaatgcacac tgctggcata tctggataaa ggcaaacgca	6960
aacgcaaaaga aaaagcgccc agcctgcaat gggcgatata ggcgattgcg cgtctggcg	7020
gctttatgga tagcaaacgt accggcattt cgagctgggg tgcgtgtgg gaagggtggg	7080
aagcgctgca aagcaactg gatggcttgc tgccgcgaa agacctgtat ggcgcaggca	7140
ttaaaatctg cateacgggaa gatgeactag ttgccttacc cgaggcggag teggtacgca	7200
tcgcgcacat cgtccgggt ggcgcggccaa acagtgcacaa cgcgcacatgc ctgaaagtcc	7260
ttgaccggca tggcaatccc gtgcgtccgc accggctgtt ccactccggc gacatccgg	7320
tgtacacggt gcgtacggc gaaggctctgc gtgtgacggg caccgcgaa caccgttgt	7380
tgtgtttgtt cgacgtcgcc ggggtgcgaa ccctgtgtt gaagctgtatc gacaaatca	7440
agccggcga ttacgcgggtt attcaacgcg ggcatttgc cgtcgcactgtt gcaagggttttgc	7500
cccgccggaa acccgaaattt ggcgcacaa cctacacagt cgccgtccct ggactgggtgc	7560
gtttcttggaa agcacaaccac cgagaccggc acggccaaacg tgcgcacac gagctgaccg	7620
acggcggtt ctactacgcg aaagtgcacca gtgtcaccga cgcggcggtt cagccgtgt	7680
atagccttcg tgcgtacacgc gcaagaccacg cgtttatcac gacgggttc gtcagccacg	7740
ctactggcctt caccggctcg aactcaggcc tcacgcacaa tccgtgttgc tccgttggc	7800
aggtcaacac agcttataact ggggacaaat tggcacata taacggcaag acgtataaat	7860
gtttcagcc ccacacccctt ttggcaggat gggacccatc caacgttccct gccttgc	7920
agcttcaatg actgcaggaa gggatccgg ctgctaacaa agccggaaag gaagctgtat	7980
tggcgctgc caccgcgttgc caataacttag cataacccct tgggcctctt aaacgggtct	8040
tgagggtttt ttgtgaaagg gggactatc tatccggat	8079

<210> 4

<211> 19

<212> DNA

<213> 人工序列

<220>

<223> Tn5M^Erev

<400> 4

ctgtcttta tacacatct

19

<210> 5

<211> 33

<212> DNA

<213> 人工序列

<220>		
<223>	Tn5ME-A	
<400> 5		
tcgtcggcag	cgtcagatgt	gtataagaga
	cag	33
[0007]	<210> 6	
	<211> 34	
	<212> DNA	
	<213> 人工序列	
<220>		
<223>	Tn5ME-B	
<400> 6		
gtctcgtgg	ctcggagatg	tgtataagag
	acag	34

	Nano ChIP-seq	LinDA	小规模 ChIP	运载体 ChIP-seq	细菌运载体 ChIP-seq	iChIP	低细胞 ChIP	ChIP-exo	ChIP-nexus	ChIPmentation
出版物标题	Genome-wide chromatin maps derived from limited numbers of hematopoietic progenitors	Single-tube linear DNA amplification (LinDA) for robust ChIP-seq	In vivo epigenomic profiling of germ cells reveals germ cell molecular signatures from breast cancer core needle biopsy samples	A carrier-assisted ChIP-seq method for estrogen receptor-chromatin interactions for small-cell-number transcription factor ChIP-seq	Amplification of pico-scale DNA mediated by bacterial carrier DNA for small-cell-number transcription factor ChIP-seq	Chromatin state dynamics during blood formation	Bivalent chromatin marks during developmental regulation	Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution	ChIP-nexus enables improved detection of in vivo transcription factor binding footprints	ChIPmentation: fast, cheap, low-input ChIP-seq for histones and transcription factors
参考文献	Adli et al., Nature Methods 7, 615-618, 2010	Shankaranarayanan et al., Nature Methods 8, 565-567, 2011	Ng et al., Developmental Cell, 24, 324-333, 2013	Zwart et al., BMC Genomics 14, 232, 2013	Jakobsen et al., BMC Genomics 16, 2015	Lara-Astiaso et al., Science Reports 3, 1777-1784, 2013	Sachs et al., Cell 147, 1408-1419, 2011	Rhee et al., Cell Reports 3, 943-949, 2014	He et al., Nature Biotechnology 33, 395-401, 2015	this study
ChIP-seq对于组氨酸的标记所需的细胞数	10k	5k	1k (H3K27me3)	n/a	10k	500 (在多路反应中)	12.5k	n/a	n/a	10k

	Nano ChIP-seq	linDA	小规模 ChIP	运载体 ChIP-seq	细菌 DNA 运载体 ChIP-seq	iChIP	低细胞 ChIP	ChIP-exo	ChIP-nexus	ChIPmentation
ChIP-seq 对于转录因子所需的细胞数	n/a	5k	n/a	10k (测试了 1个因子)	10k (测试了 1个因子；方案需要多路化)	10k (测试了 1个因子；方案需要多路化)	n/a	2.5M	10M	100k
方案	多步文库制备；方案包括预扩增	多步文库制备；方案包括预扩增	方案使用运载体 RNA 和重组蛋白	方案使用运载体 RNA 和重组蛋白	方案使用额外的利用 H3 组蛋白的抗体的沉淀	市售低输入文库制盒 (Rubicon Genomics)	复杂的，多步备方案	复杂的，多步备方案	复杂的，多步备方案	简单的文库制备
文库制备所需的时间	2 天	常规模接头常连接介导的文库制备，额外的 4-5 小时	常规模接头常连接介导的文库制备，额外的 1.5 天	常规模接头常连接介导的文库制备，额外的 1 天	~70 分钟	5-6 小时文库制备 +DNA 纯化	7-8 小时 (ChIP-exo 处理 & nexus 文库制备)	10-20 分钟文库制备		

	Nano ChIP-seq	linDA ChIP	小规模 ChIP	运载体 ChIP-seq	细菌 DNA 运载体 ChIP-seq	iChIP	低细胞 ChIP	ChIP-exo	ChIP-nexus	ChIPmentation
优点	低细胞数 用蛋白修饰 蛋白和转录因子	低细胞数 用蛋白修饰 蛋白和转录因子	极低细胞数 蛋白修饰 蛋白和转录因子	低细胞数 蛋白修饰 蛋白和转录因子	极低细胞数 蛋白修饰 蛋白和转录因子	低细胞数 蛋白修饰 蛋白和转录因子	低细胞数 蛋白修饰 蛋白和转录因子	转录因子 结合背景图 低高分辨率 (相对于 ChIP-exo 改进)	转录因子 结合背景图 低高分辨率 (相对于 ChIP-exo 改进)	快速、简便、 低细胞数 和细胞录在率 转录因子 结合背景图 低高分辨率 (相对于 ChIP-exo 改进)
缺点	耗时方案； 需要试剂； 由于扩增的 PCR偏差	耗时方案； 需要试剂； 由于扩增的 PCR偏差	耗时方案； 需要试剂； 由于扩增的 PCR偏差							

	Nano ChIP-seq	linDA	小规模 ChIP	运载体 ChIP-seq	细菌 DNA 运载体 ChIP-seq	iChIP	低细胞 ChIP	ChIP-exo	ChIP-nexus	ChIPmentation

图1

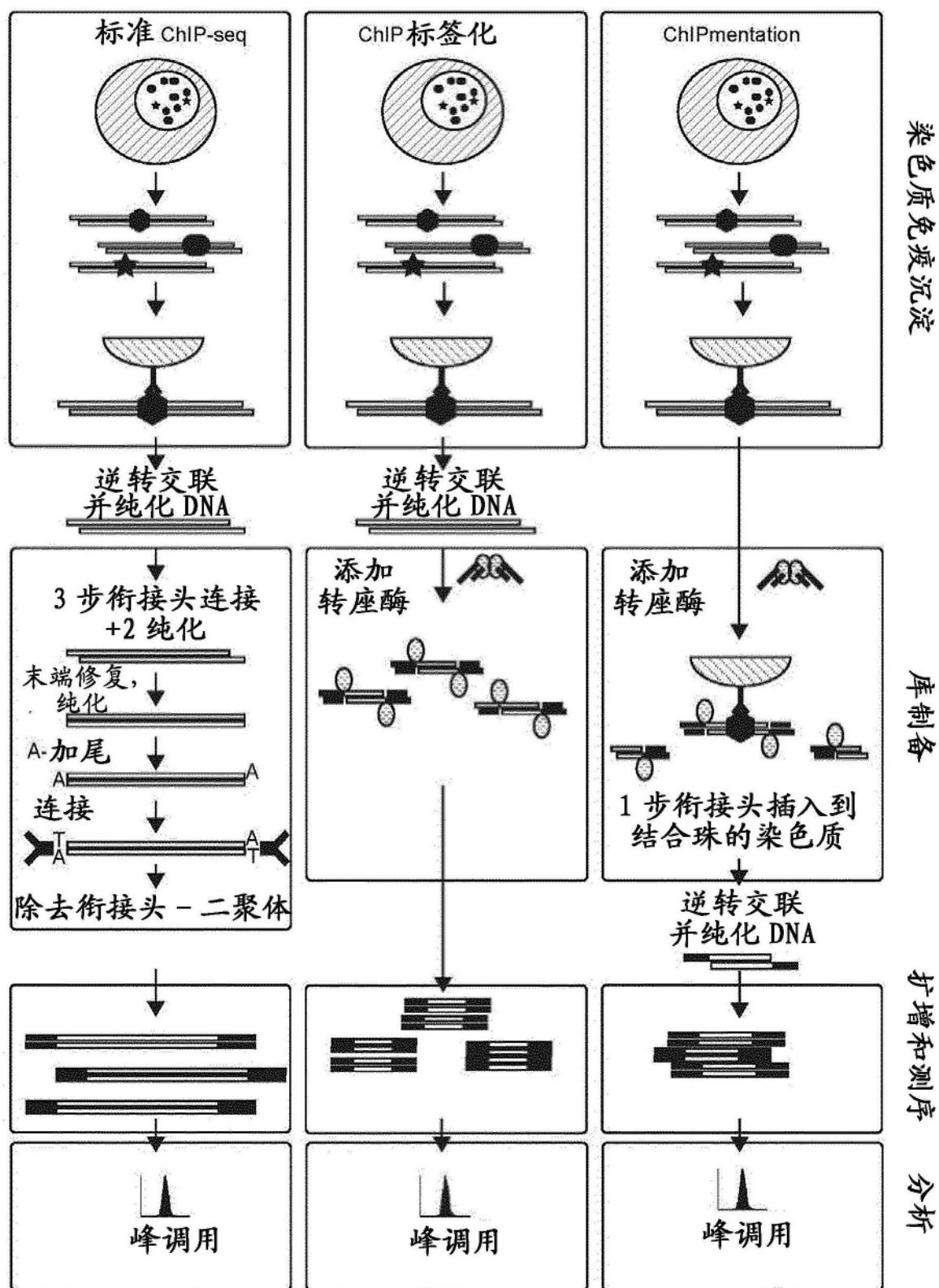


图2

用不同的 $Tn5$ 转座酶制备的 H3K4me3 ChIPmentation 文库的 PCR 之前和之后的片段大小分布

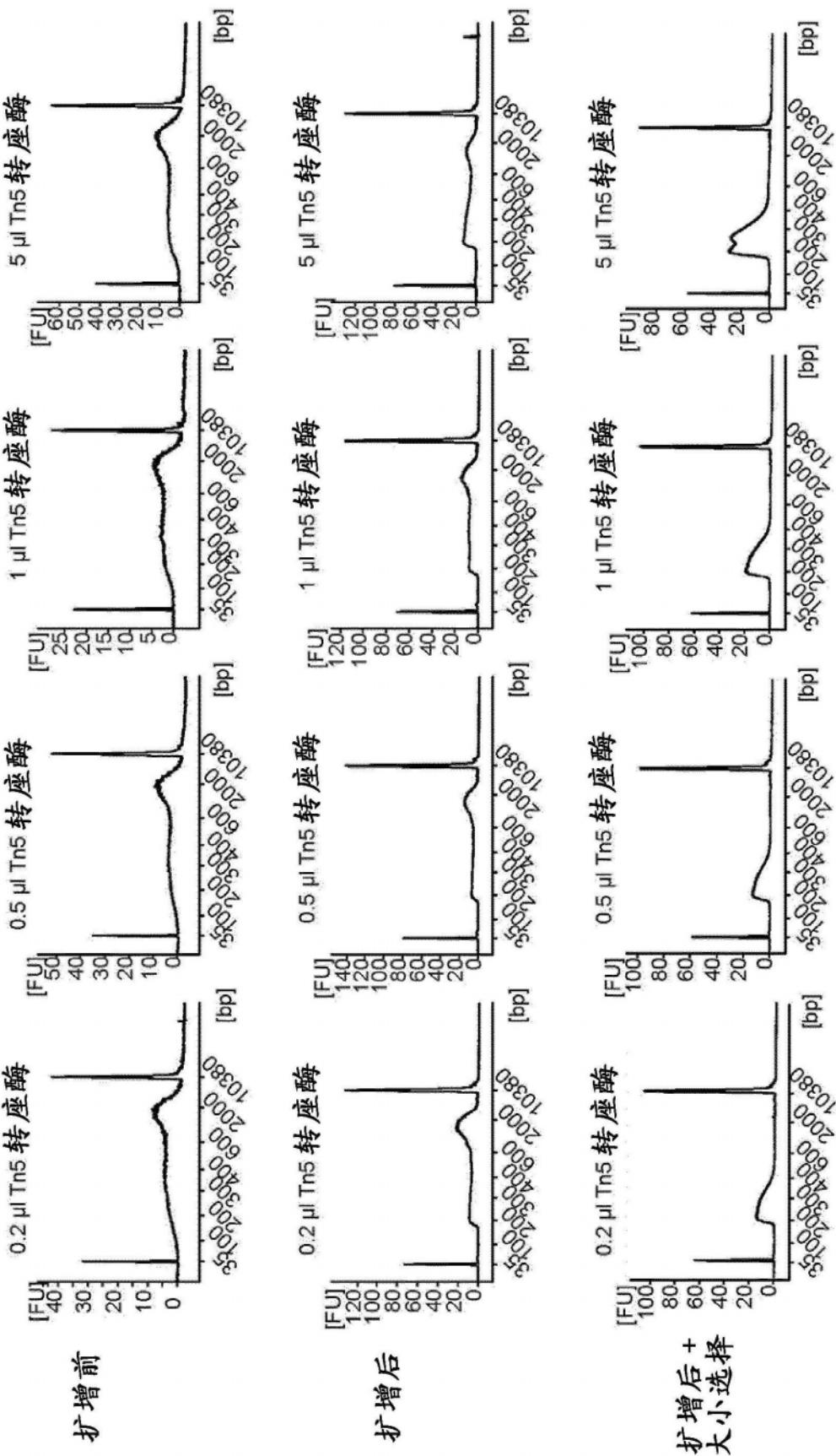


图3

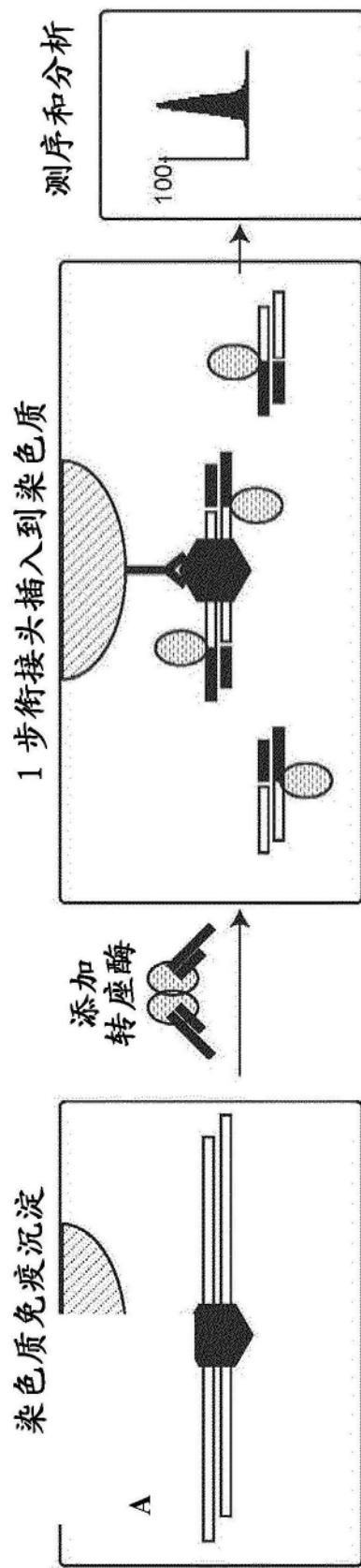


图4

B

H3K4m3 读段的大小分布

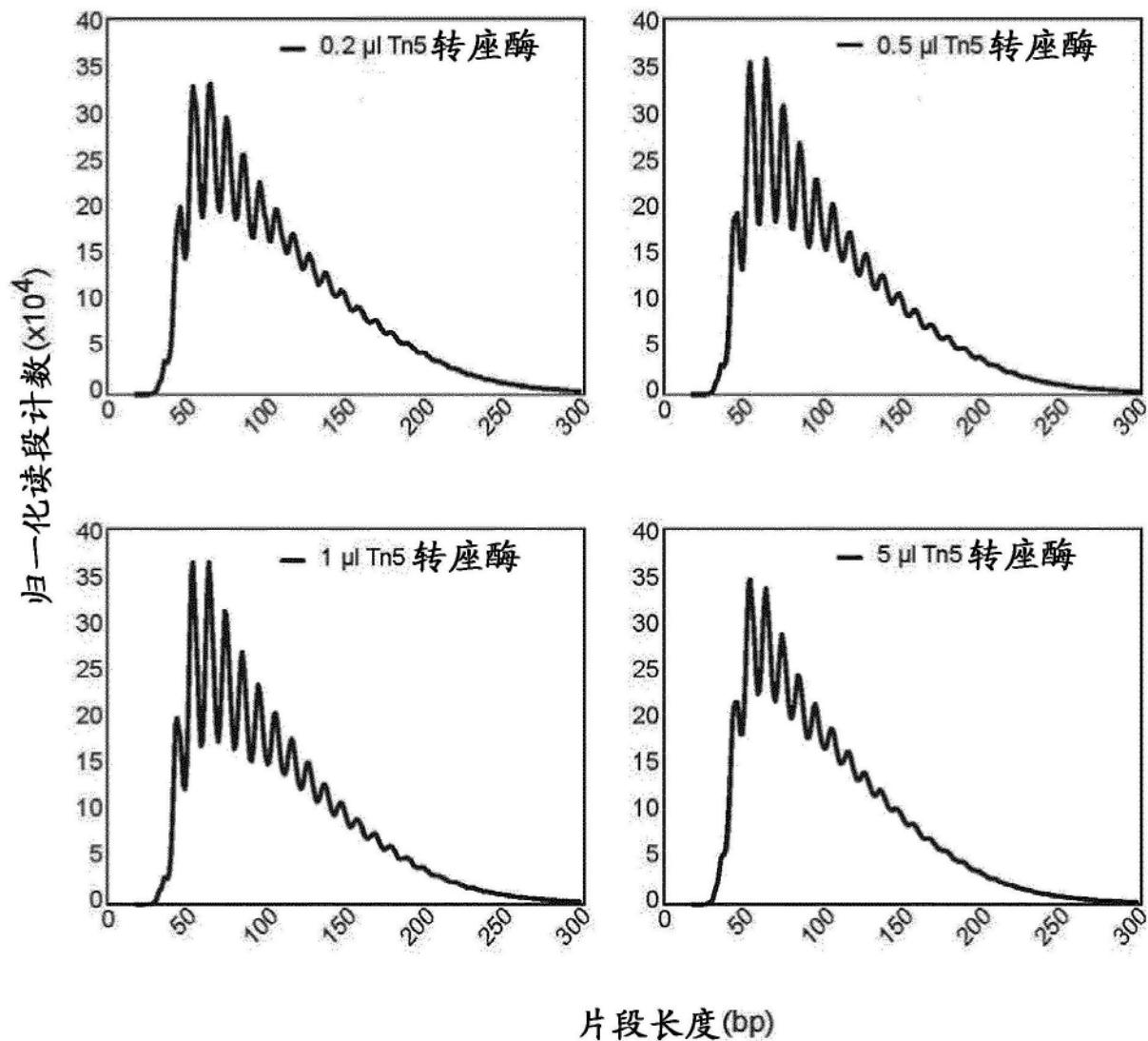


图4续

C

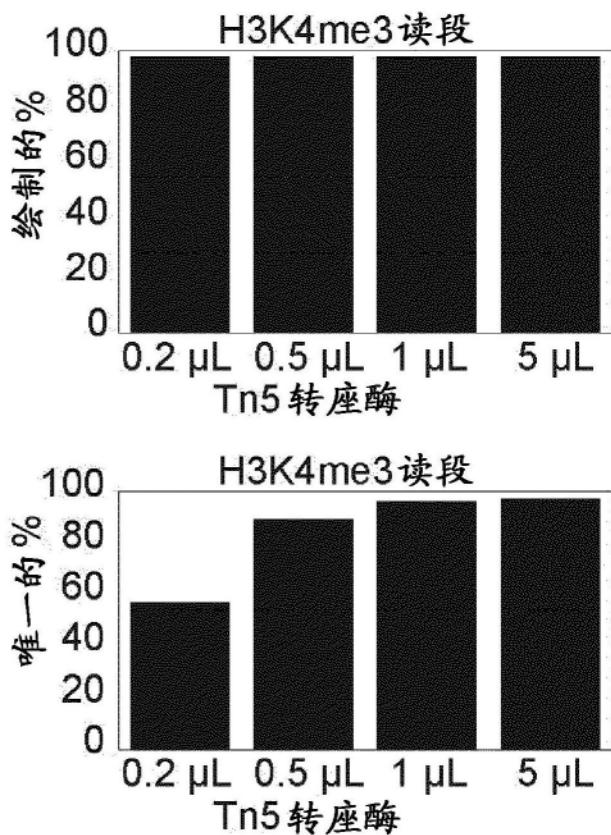


图4续

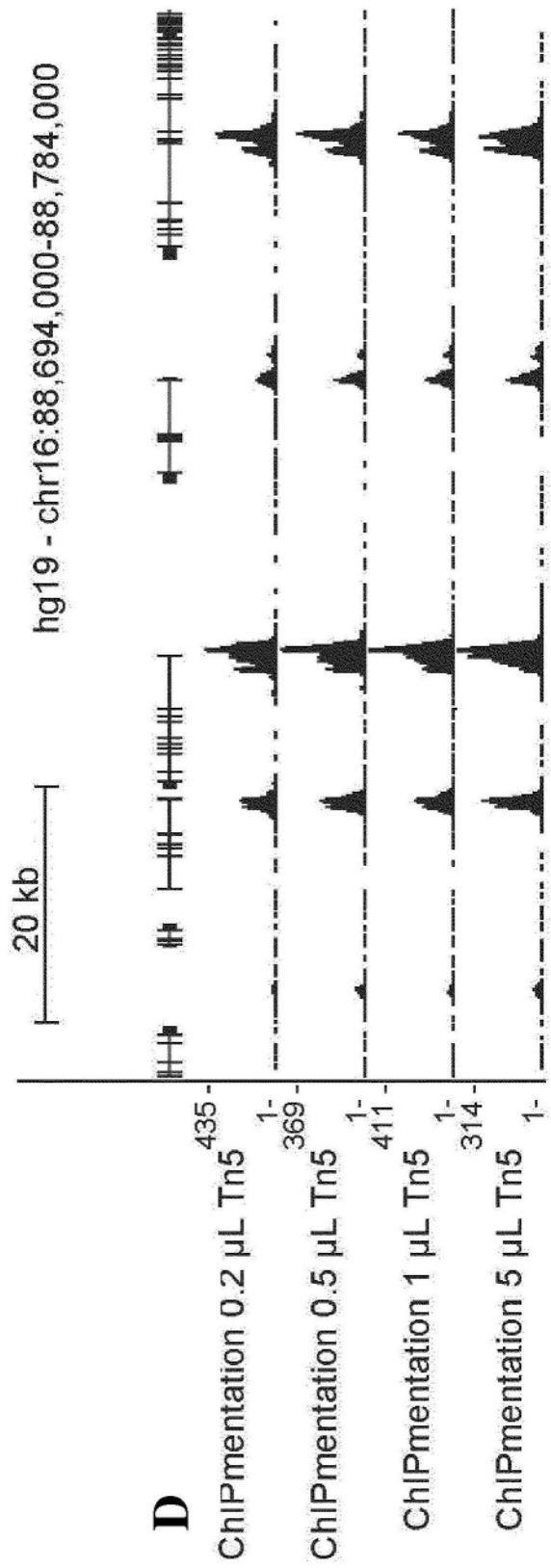


图4续

E

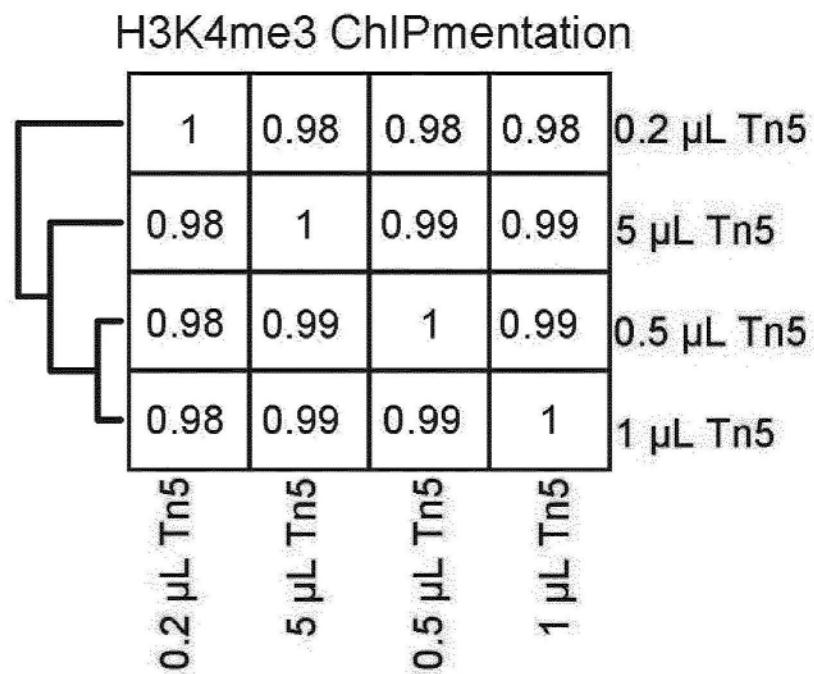


图4续

F

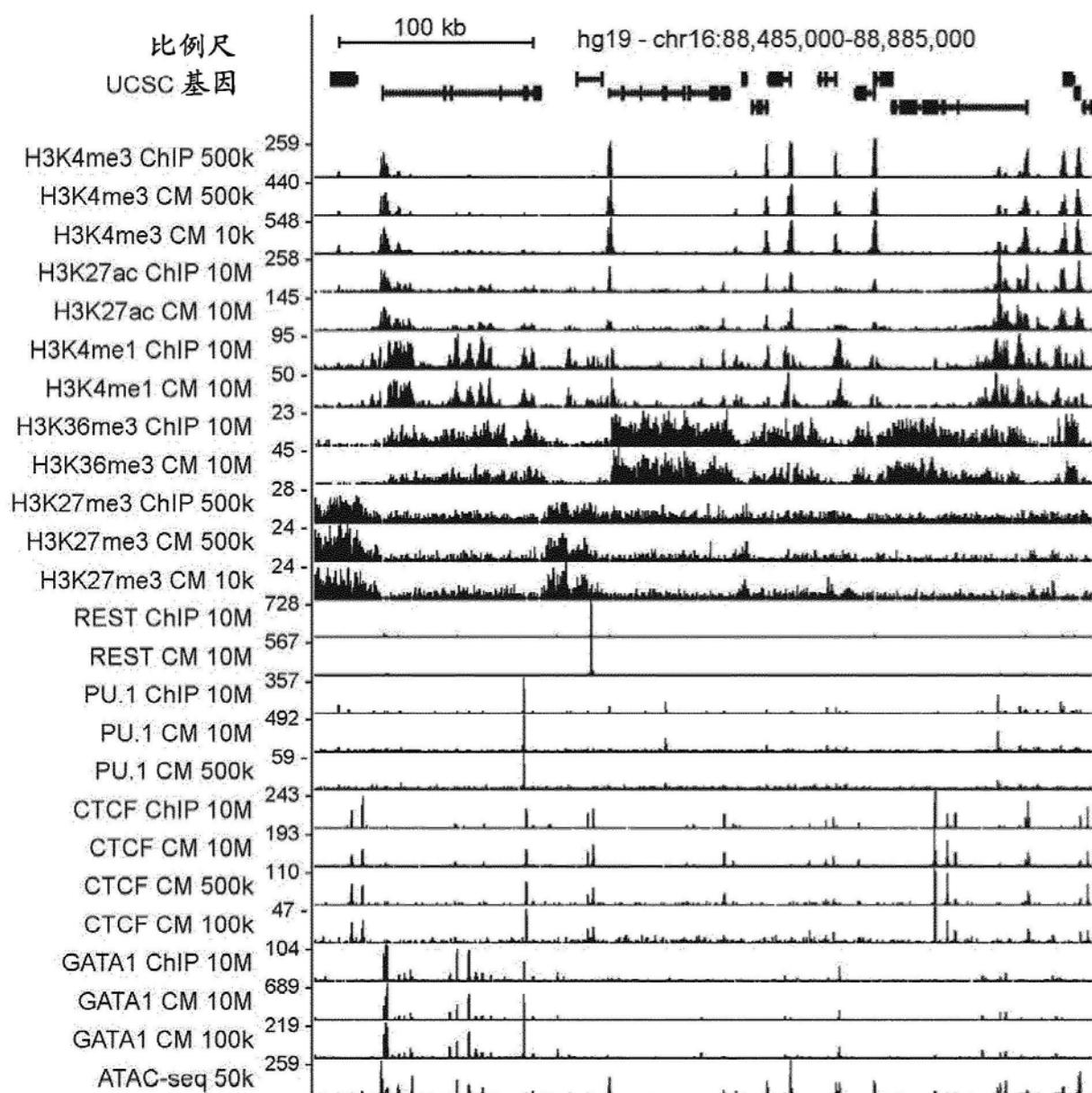


图4续

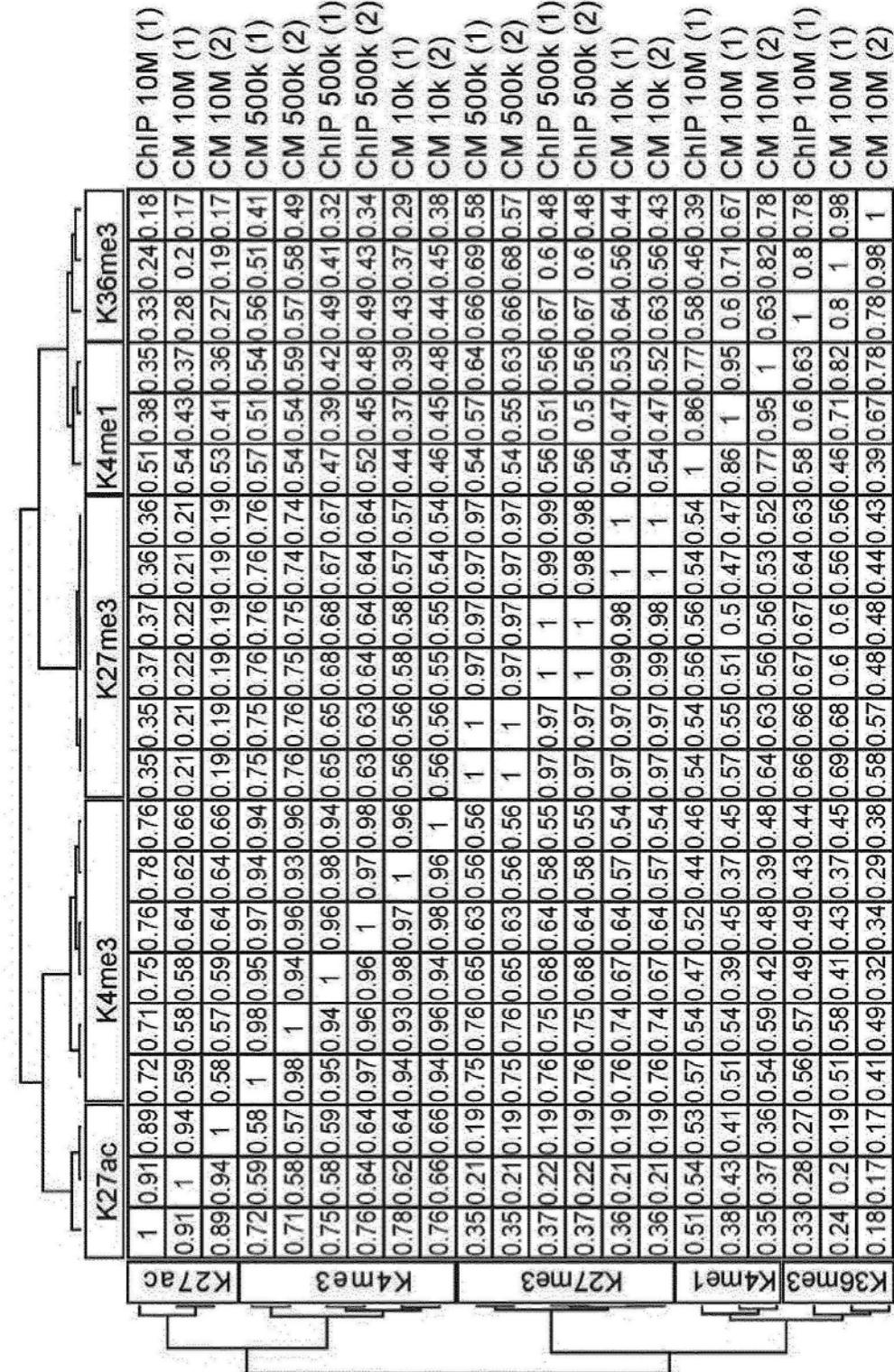


图4续

H

重复 1vs2		CM vs ChIP
	高细胞数	低细胞数
CTCF	0.94	0.98
GATA1	0.99	0.90
PU1	0.96	0.97
REST	0.97	0.98

I

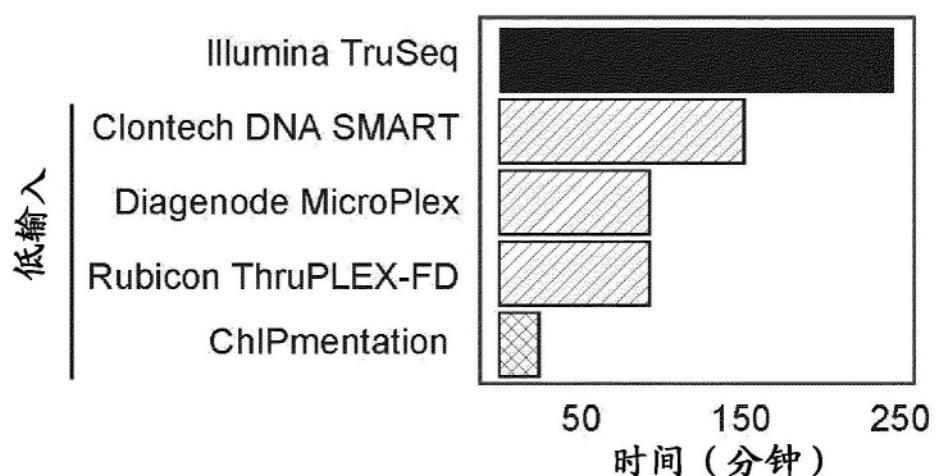


图4续

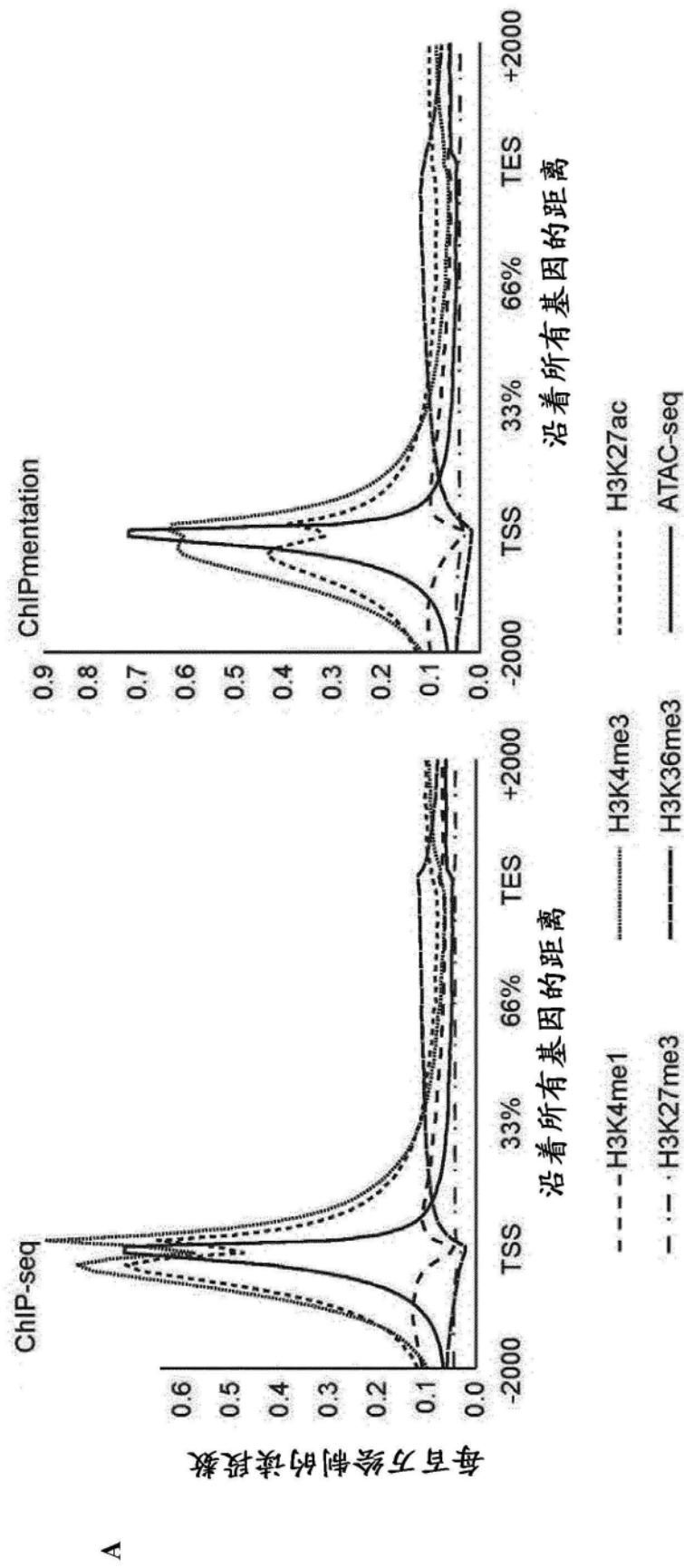


图5

B

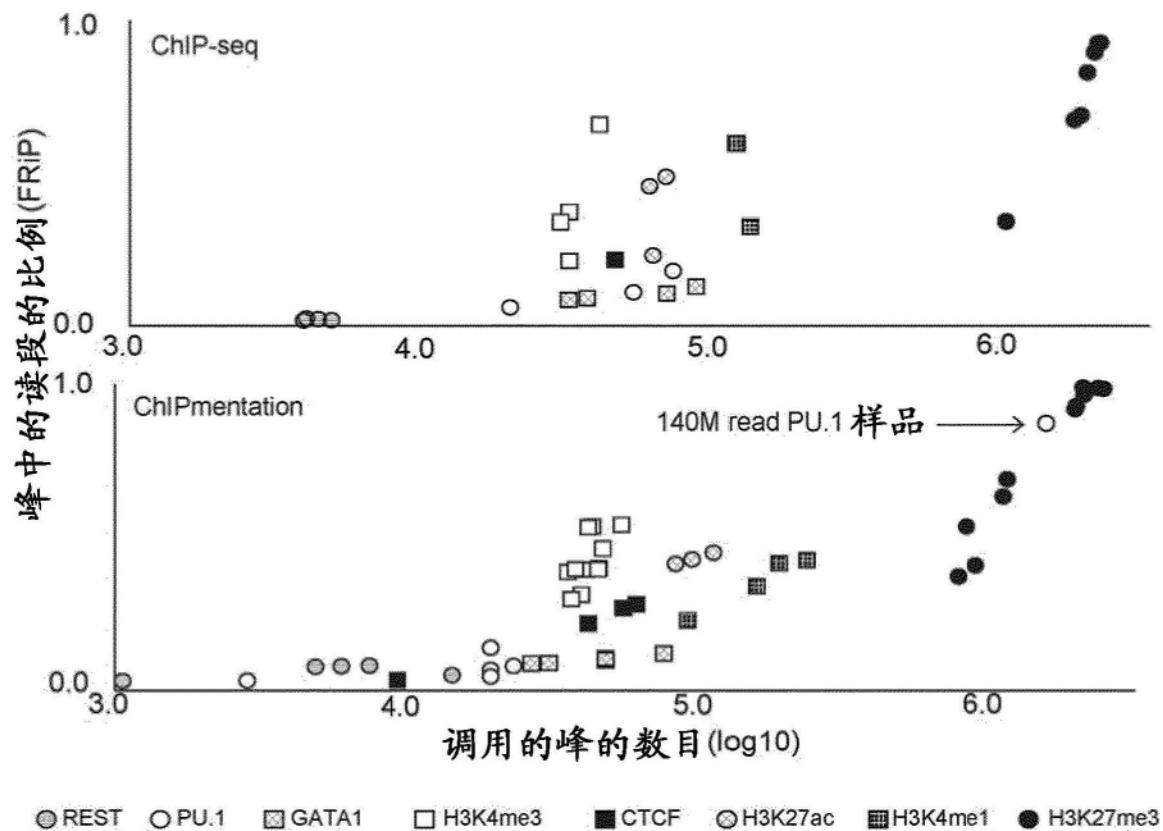


图5续

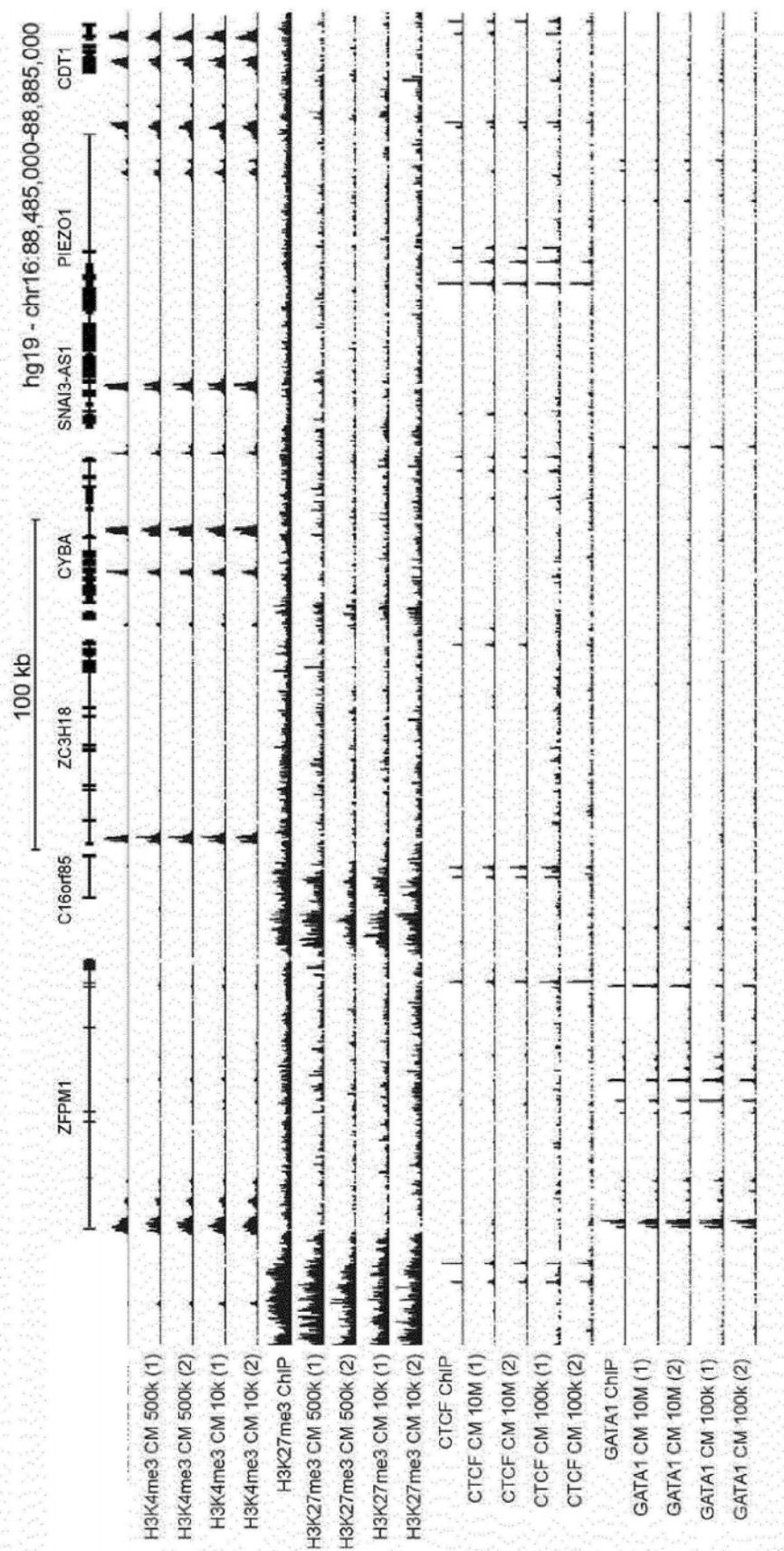


图6

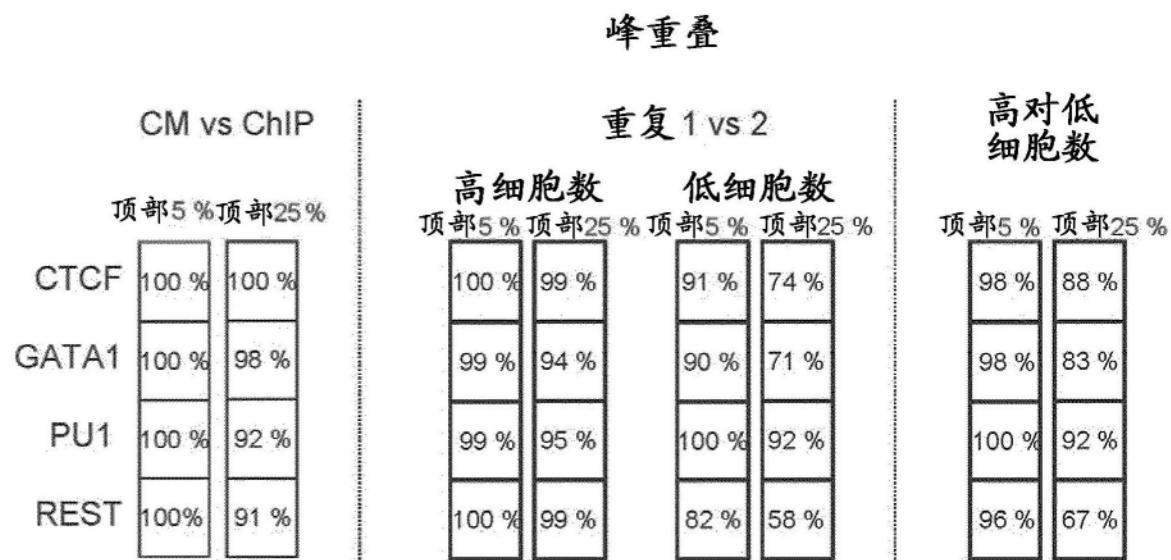


图7

id	细胞类型	输入量	方案	抗体	重复	测序仪	读段类型	读段长度	读段计数	基因组	比对率	唯一率	内部样品 id
1	K562	10M 细胞	ChIP-seq	H3K4me1	1	HiSeq 2000	单读段	50	24,523,943	hg19	99	94	K562_10M_ChIP-seq_H3K4ME1_nan_1 hg19
2	K562	10M 细胞	ChIP-seq	H3K4me1	2	HiSeq 2000	单读段	50	25,448,971	hg19	99	97	K562_10M_ChIP-seq_H3K4ME1_nan_2 hg19
3	K562	500K 细胞	ChIP-seq	H3K4me3	1	HiSeq 2000	单读段	50	18,610,366	hg19	99	94	K562_500K_ChIP-seq_H3K4ME3_nan_1 hg19
4	PBMC	500K 细胞	ChIP-seq	H3K4me3	1	HiSeq 2000	单读段	50	39,543,850	hg19	86	31	PBMC_nan_ChIP-seq_H3K4ME3_nan_1 hg19
5	K562	500K 细胞	ChIP-seq	H3K4me3	2	HiSeq 2000	单读段	50	37,576,419	hg19	99	99	K562_500K_ChIP-seq_H3K4ME3_nan_2 hg19
6	K562	10M 细胞	ChIP-seq	H3K27ac	1	HiSeq 2000	单读段	50	27,503,214	hg19	98	45	K562_10M_ChIP-seq_H3K27AC_nan_1 hg19
7	K562	500K 细胞	ChIP-seq	H3K27ac	1	HiSeq 2000	单读段	50	7,714,041	hg19	98	95	K562_500K_ChIP-seq_H3K27AC_nan_1 hg19
8	K562	500K 细胞	ChIP-seq	H3K27me3	1	HiSeq 2000	单读段	50	33,279,554	hg19	98	96	K562_500K_ChIP-seq_H3K27ME3_nan_1 hg19
9	K562	500K 细胞	ChIP-seq	H3K27me3	2	HiSeq 2000	单读段	50	31,605,383	hg19	98	95	K562_500K_ChIP-seq_H3K27ME3_nan_2 hg19
10	K562	10M 细胞	ChIP-seq	H3K36me3	1	HiSeq 2000	单读段	50	24,453,845	hg19	98	96	K562_10M_ChIP-seq_H3K36ME3_nan_1 hg19
11	K562	10M 细胞	ChIP-seq	CTCF	1	HiSeq 2000	单读段	50	16,731,630	hg19	96	80	K562_10M_ChIP-seq_H3K36ME3_nan_1 hg19

id	细胞类型	输入量	方案	抗体	重复	测序仪	读段类型	读段长度	读段计数	基因组比对率	唯一率	内部样品id
12	K562	10M 细胞	ChIP-seq	CTCF	2	HiSeq 2000	单读段	50	13.292.391	hg19	96	seq_CTCF_nan_1_hg19
13	K562	10M 细胞	ChIP-seq	GATA1	1	HiSeq 2000	单读段	50	20.908.847	hg19	98	K562_10M_ChIP-seq_GATA1_nan_1_hg19
14	K562	10M 细胞	ChIP-seq	GATA1	2	HiSeq 2000	单读段	50	17.734.841	hg19	98	K562_10M_ChIP-seq_GATA1_nan_2_hg19
15	K562	10M 细胞	ChIP-seq	PU.1	1	HiSeq 2000	单读段	50	13.663.300	hg19	97	K562_10M_ChIP-seq_PU1_nan_1_hg19
16	K562	10M 细胞	ChIP-seq	PU.1	2	HiSeq 2000	单读段	50	13.453.300	hg19	97	K562_10M_ChIP-seq_PU1_nan_2_hg19
17	K562	10M 细胞	ChIP-seq	REST	1	HiSeq 2000	单读段	50	27.078.470	hg19	98	K562_10M_ChIP-seq_REST_nan_1_hg19
18	K562	10M 细胞	ChIP-seq	REST	2	HiSeq 2000	单读段	50	18.994.605	hg19	98	K562_10M_ChIP-seq_REST_nan_2_hg19
19	K562	10M 细胞	ChIP-seq	免疫球蛋白	1	HiSeq 2000	单读段	50	31.818.701	hg19	95	K562_10M_ChIP-seq_IgG_nan_1_hg19
20	K562	500k 细胞	ChIP-seq	免疫球蛋白	1	HiSeq 2000	单读段	50	32.545.198	hg19	98	K562_500k_ChIP-seq_IgG_nan_1_hg19
21	K562	10M 细胞	ChIP-seq	免疫球蛋白	2	HiSeq 2000	单读段	50	12.999.527	hg19	93	K562_10M_ChIP-seq_IgG_nan_2_hg19
22	K562	500k 细胞	ChIP-seq	免疫球蛋白	2	HiSeq 2000	单读段	50	36.366.641	hg19	90	K562_500k_ChIP-seq_IgG_nan_2_hg19
23	K562	10M 细胞	ChIP-seq	免疫球蛋白	3	HiSeq 2000	单读段	50	14.298.197	hg19	96	K562_10M_ChIP-seq_IgG_nan_3_hg19
24	PBMC	500k 细胞	ChIP-seq	Input	1	HiSeq 2000	单读段	50	4.674.578	hg19	93	PBMC_nan_ChIP-seq_Input_nan_1_hg19
25	K562	10M 细胞	ChIPmentation	H3K4me1	1	HiSeq 2000	单读段	50	14.683.831	hg19	97	K562_10M_ChIPmentation_H3K4ME1_nan_1_hg19

id	细胞类型	输入量	方案	抗体	重复	测序仪	读段类型	读段长度	读段计数	基因组	比对率	唯一率	内部样品 id
26	K562	10M 细胞	ChIPmentation	H3K4me1	1	HiSeq 2000	成对末端	100	25.372.361	hg19	98	91	K562_10M_ChIPmentation_on_H3K4ME1_nan_1_h_g19
27	K562	10M 细胞	ChIPmentation	H3K4me1	1	HiSeq 2000	成对末端	100	17.989.151	hg19	98	92	K562_10M_ChIP- Fragmentation_H3K4ME1_nan_1_hg19
28	K562	10M 细胞	ChIPmentation	H3K4me1	2	HiSeq 2000	单读段	50	13.216.471	hg19	97	89	K562_10M_ChIPmentation_on_H3K4ME1_nan_2_h_g19
29	K562	500K 细胞	ChIPmentation	H3K4me3	1	HiSeq 2000	单读段	50	28.139.774	hg19	98	79	K562_500K_ChIPmentation_H3K4ME3_nan_1_h_g19
30	K562	500K 细胞	ChIPmentation	H3K4me3	2	HiSeq 2000	单读段	50	14.306.608	hg19	99	87	K562_500K_ChIPmentation_H3K4ME3_nan_2_h_g19
31	K562	10K 细胞	ChIPmentation	H3K4me3	3	HiSeq 2000	单读段	50	37.048.339	hg19	98	34	K562_10K_ChIPmentation_on_H3K4ME3_nan_3_h_g19
32	K562	10K 细胞	ChIPmentation	H3K4me3	4	HiSeq 2000	单读段	50	35.357.419	hg19	97	40	K562_10K_ChIPmentation_on_H3K4ME3_nan_4_h_g19
33	K562	10M 细胞	ChIPmentation	H3K27ac	1	HiSeq 2000	单读段	50	19.377.579	hg19	98	83	K562_10M_ChIPmentation_on_H3K27AC_nan_1_h_g19
34	K562	10M 细胞	ChIPmentation	H3K27ac	2	HiSeq 2500	单读段	50	10.356.848	hg19	99	92	K562_10M_ChIPmentation_on_H3K27AC_nan_2_h_g19
35	K562	10K 细胞	ChIPmentation	H3K27me3	1	HiSeq 2000	单读段	50	23.965.157	hg19	98	86	K562_10K_ChIPmentation_on_H3K27ME3_nan_1_h_g19
36	K562	500K 细胞	ChIPmentation	H3K27me3	1	HiSeq 2000	单读段	50	22.064.368	hg19	97	86	K562_500K_ChIPmentation_on_H3K27ME3_nan_1_h_g19

id	细胞类型	输入量	方案	抗体	重复	测序仪	读段类型	读段长度	读段计数	基因组比对率	唯一率	内部样品id
37	K562	10K细胞	ChIPmentation	H3K27me3	2	HiSeq 2000	单读段	50	26.019.592	hg19	98	hg19
38	K562	500k细胞	ChIPmentation	H3K27me3	2	HiSeq 2000	单读段	50	24.286.257	hg19	98	K562_500K_ChIPmentation_H3K27ME3_nan_2_hg19
39	K562	10M 细胞	ChIPmentation	H3K36me3	1	HiSeq 2000	单读段	50	21.628.106	hg19	97	K562_10M_ChIPmentation_H3K36ME3_nan_1_hg19
40	K562	10M 细胞	ChIPmentation	H3K36me3	2	HiSeq 2500	单读段	50	18.738.407	hg19	98	K562_10M_ChIPmentation_H3K36ME3_nan_2_hg19
41	K562	100k细胞	ChIPmentation	CTCF	1	HiSeq 2000	单读段	50	11.497.781	hg19	82	K562_100K_ChIPmentation_CTCF_nan_1_hg19
42	K562	10M 细胞	ChIPmentation	CTCF	1	HiSeq 2000	单读段	50	11.830.997	hg19	98	K562_10M_ChIPmentation_CTCF_nan_1_hg19
43	K562	500k细胞	ChIPmentation	CTCF	1	HiSeq 2500	单读段	50	10.464.133	hg19	98	K562_500K_ChIPmentation_CTCF_nan_1_hg19
44	K562	100k细胞	ChIPmentation	CTCF	2	HiSeq 2500	单读段	50	10.725.153	hg19	98	K562_100K_ChIPmentation_CTCF_nan_2_hg19
45	K562	10M 细胞	ChIPmentation	CTCF	2	HiSeq 2000	单读段	50	10.404.401	hg19	96	K562_10M_ChIPmentation_CTCF_nan_2_hg19
46	K562	500k细胞	ChIPmentation	CTCF	2	HiSeq 2000	单读段	50	10.772.645	hg19	97	K562_500K_ChIPmentation_CTCF_nan_2_hg19
47	K562	100k细胞	ChIPmentation	GATA1	1	HiSeq 2500	单读段	50	12.073.362	hg19	97	K562_100K_ChIPmentation_GATA1_nan_1_hg19
48	K562	10M 细胞	ChIPmentation	GATA1	1	HiSeq 2000	单读段	50	18.086.796	hg19	97	K562_10M_ChIPmentation_GATA1_nan_1_hg19
49	K562	100k细胞	ChIPmentation	GATA1	2	HiSeq 2500	单读段	50	10.182.490	hg19	98	K562_100K_ChIPmentation_GATA1_nan_2_hg19

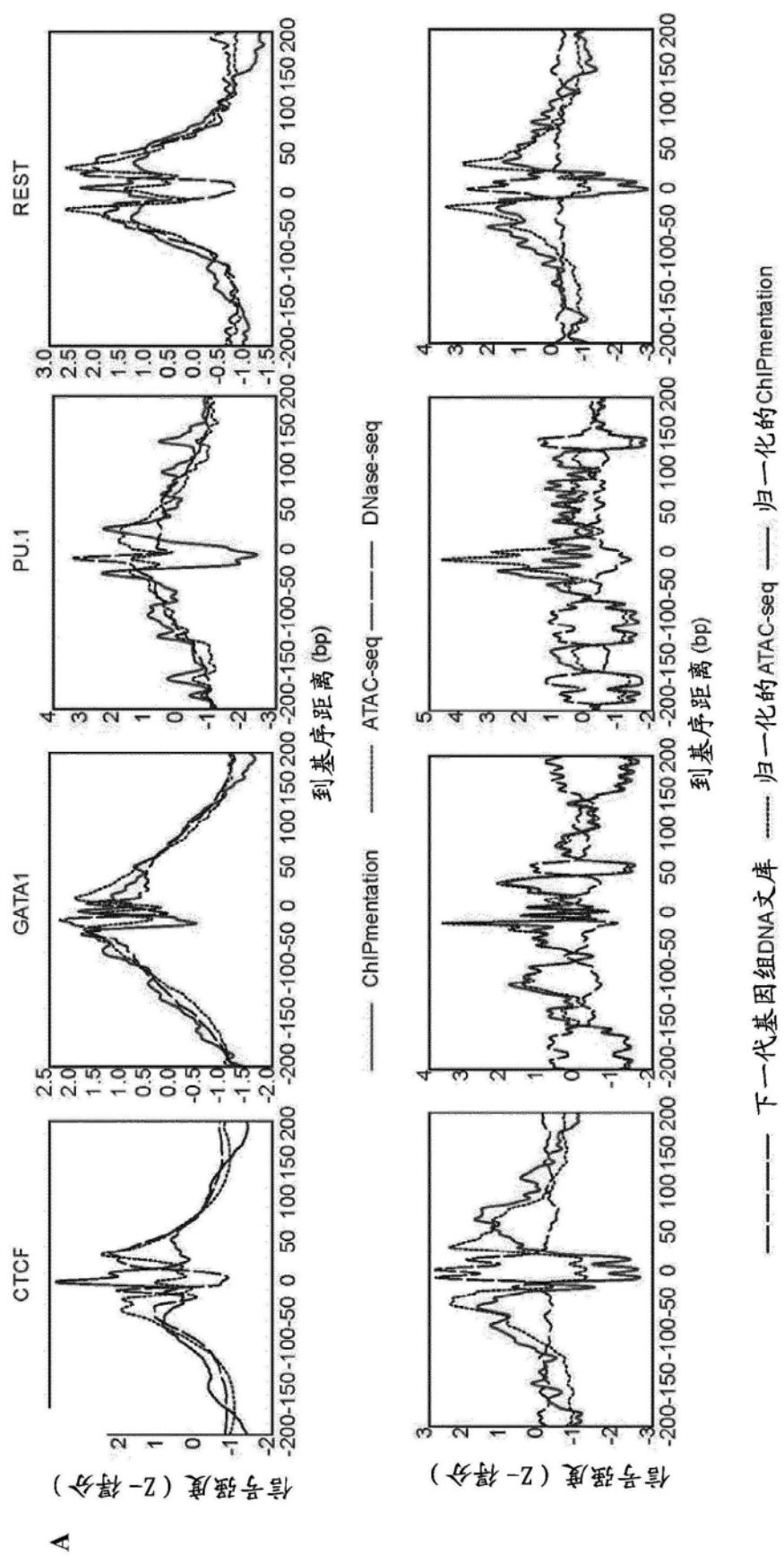
id	细胞类型	输入量	方案	抗体	重复	测序仪	测序类型	读段长度	读段计数	基因组	比对率	唯一率	内部样品 id
50	K562	10M细胞	ChIPmentation	GATA1	2	HiSeq 2000	单读段	50	16.679.429	hg19	98	92	K562_10M_ChIPmentation on GATA1_nan_2_hg19
51	K562	10M细胞	ChIPmentation	PU.1	1	HiSeq 2000	单读段	50	65.539.103	hg19	98	84	K562_10M_ChIPmentation on PU1_nan_1_hg19
52	K562	10M细胞	ChIPmentation	PU.1	1	HiSeq 2000	成对末端	100	27.976.709	hg19	97	87	K562_10M_ChIPmentation on PU1_nan_1_hg19
53	K562	500k细胞	ChIPmentation	PU.1	1	HiSeq 2000	单读段	50	9.937.246	hg19	98	93	K562_500K_ChIPmentation on PU1_nan_1_hg19
54	K562	10M细胞	ChIPmentation	PU.1	1	HiSeq 2000	成对末端	100	13.476.614	hg19	96	95	K562_10M_ChIPmentation on PU1_nan_2_hg19
55	K562	10M细胞	ChIPmentation	PU.1	2	HiSeq 2000	单读段	50	13.231.293	hg19	98	93	K562_10M_ChIPmentation on PU1_nan_1_hg19
56	K562	500k细胞	ChIPmentation	PU.1	2	HiSeq 2000	单读段	50	4.121.056	hg19	96	95	K562_500K_ChIPmentation on PU1_nan_2_hg19
57	K562	500k细胞	ChIPmentation	PU.1	3	HiSeq 2500	单读段	50	10.438.095	hg19	98	87	K562_500K_ChIPmentation on PU1_nan_3_hg19
58	K562	100k细胞	ChIPmentation	REST	1	HiSeq 2500	单读段	50	10.260.884	hg19	94	91	K562_100K_ChIPmentation on REST_nan_1_hg19
59	K562	10M细胞	ChIPmentation	REST	1	HiSeq 2000	单读段	50	16.482.402	hg19	97	81	K562_10M_ChIPmentation on REST_nan_2_hg19
60	K562	100k细胞	ChIPmentation	REST	2	HiSeq 2500	单读段	50	10.233.072	hg19	92	83	K562_100K_ChIPmentation on REST_nan_1_hg19
61	K562	10M细胞	ChIPmentation	REST	2	HiSeq 2000	单读段	50	24.670.236	hg19	97	74	K562_100K_ChIPmentation on REST_nan_2_hg19
62	K562	10M细胞	ChIPmentation	免疫球蛋白G	1	HiSeq 2000	单读段	50	13.561.441	hg19	98	86	K562_10M_ChIPmentation on IgG_nan_1_hg19
63	K562	500k细胞	ChIPmentation	免疫球蛋白G	1	HiSeq 2000	成对末端	100	14.672.785	hg19	96	57	K562_500K_ChIPmentation on IgG_1ULTN5_1_hg19
64	K562	500k细胞	ChIPmentation	免疫球蛋白G	1	HiSeq 2000	单读段	50	34.263.776	hg19	97	50	K562_500K_ChIPmentation

id	细胞类型	输入量	方案	抗体	重复	测序仪	读段类型	读段长度	读段计数	基因组	比对率	唯一率	内部样品id
65	K562	10M 细胞	ChIPmentation	G	免疫球蛋白	2	HiSeq 2000	单读段	50	12.858 450	hg19	96	ion_I GG_nan_1_hg19
66	K562	500k 细胞	ChIPmentation	G	免疫球蛋白	2	HiSeq 2000	单读段	50	18.412 113	hg19	96	K562_10M_ChIPmentati on_I GG_nan_2_hg19
67	K562	10k 细胞	ChIPmentation	G	免疫球蛋白	3	HiSeq 2000	单读段	50	17.612 836	hg19	94	K562_500K_ChIPmentati on_I GG_nan_2_hg19
68	K562	500k 细胞	ChIPmentation	G	免疫球蛋白	3	HiSeq 2000	单读段	50	7.474 797	hg19	83	K562_10K_ChIPmentati on_I GG_nan_3_hg19
69	K562	10k 细胞	ChIPmentation	G	免疫球蛋白	4	HiSeq 2000	单读段	50	45.937 406	hg19	92	K562_10M_ChIPmentati on_I GG_nan_4_hg19
70	K562	10M 细胞	ChIPmentation	G	免疫球蛋白	4	HiSeq 2000	单读段	50	5.670 213	hg19	91	K562_10K_ChIPmentati on_I GG_nan_4_hg19
71	K562	10k 细胞	ChIPmentation	G	免疫球蛋白	5	HiSeq 2000	单读段	50	12.200 893	hg19	87	K562_10M_ChIPmentati on_I GG_nan_5_hg19
72	K562	10M 细胞	ChIPmentation	G	免疫球蛋白	5	HiSeq 2000	单读段	50	10.716 675	hg19	97	K562_10K_ChIPmentati on_I GG_nan_5_hg19
73	K562	500k 细胞	ChIP-fragmentation	0.1μl Tn5	输入	1	HiSeq 2000	成对末端	100	13.668 707	hg19	97	K562_500K_ChIP-fragmentation_input_01
74	K562	500k 细胞	ChIP-fragmentation	0.1μl Tn5	H3K4me3	1	HiSeq 2000	成对末端	100	11.615 337	hg19	98	K562_500K_ChIP-fragmentation_H3K4ME3_01ULTN
75	K562	500k 细胞	ChIPmentation	0.2μl Tn5	H3K4me3	1	HiSeq 2000	成对末端	100	13.425 746	hg19	98	K562_500K_ChIPmentati on_H3K4ME3_02ULTN
76	K562	500k 细胞	ChIPmentation	0.5μl Tn5	H3K4me3	1	HiSeq 2000	成对末端	100	13.052 416	hg19	98	K562_500K_ChIPmentati on_H3K4ME3_05ULTN
77	K562	500k 细胞	ChIPmentation	1μl Tn5	H3K4me3	1	HiSeq 2000	成对末端	100	12.767 587	hg19	98	K562_500K_ChIPmentati on_H3K4ME3_1ULTN5

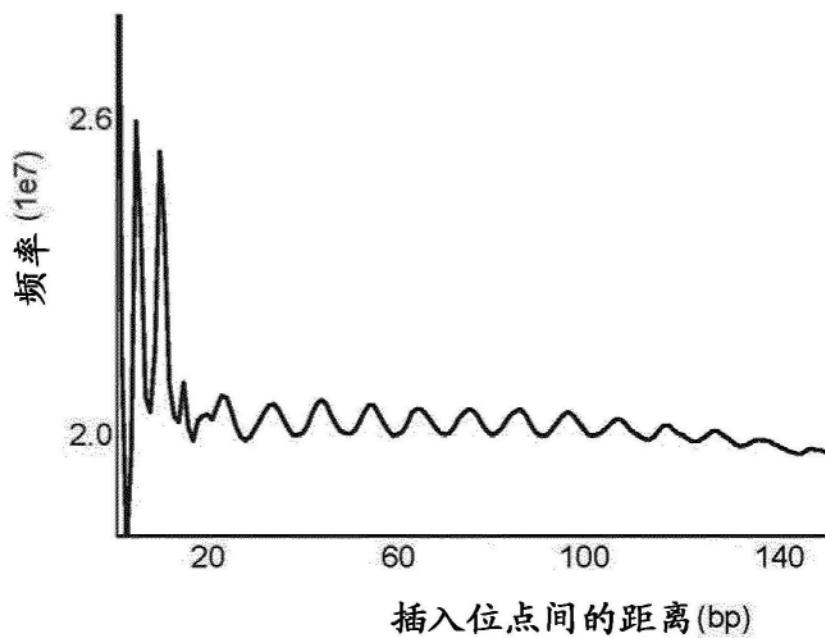
id	细胞类型	输入量	方案	抗体	重複	测序仪	读段类型	读段长度	读段计数	基因组	比对率	唯一率	内部样品 id
78	K562	500k 细胞	ChIPmentation	H3K4me3	1	HiSeq 2000	成对末端	100	11.743.183	hg19	98	97	K562_500K_ChIPmental_H3K4ME3_5ULTN5_1_hg19
79	PBMC	100pg DNA	ChIP-mentation	输入	1	HiSeq 2000	单读段	50	18.751.591	hg19	96	73	PBMC_100pg_DNA_ChIP-DNA_ChIP-mentation_Input_nan_1_hg19
80	PBMC	10pg DNA	ChIP-mentation	输入	1	HiSeq 2000	单读段	50	9.031.248	hg19	90	99	PBMC_10pg_DNA_ChIP-DNA_ChIP-mentation_Input_nan_1_hg19
81	PBMC	2pg DNA	ChIP-mentation	输入	1	HiSeq 2000	单读段	50	12.031.802	hg19	90	99	PBMC_2pg_DNA_ChIP-DNA_ChIP-mentation_Input_nan_1_hg19
82	PBMC	2pg DNA	ChIP-mentation	H3K4me3	1	HiSeq 2000	单读段	50	11.960.898	hg19	90	99	PBMC_2pg_DNA_ChIP-DNA_ChIP-mentation_H3K4ME_3_nan_1_hg19
83	PBMC	100pg DNA	ChIP-mentation	H3K4me3	1	HiSeq 2000	单读段	50	3.626.047	hg19	99	99	PBMC_100pg_DNA_ChIP-DNA_ChIP-mentation_H3K4ME_3_nan_1_hg19
84	PBMC	100pg DNA	ChIP-mentation	H3K4me3	1	HiSeq 2000	单读段	50	8.393.968	hg19	43	90	PBMC_100pg_DNA_ChIP-DNA_ChIP-mentation_H3K4ME_3_nan_1_hg19
85	PBMC	10pg DNA	ChIP-mentation	H3K4me3	1	HiSeq 2000	单读段	50	11.741.222	hg19	92	99	PBMC_10pg_DNA_ChIP-DNA_ChIP-mentation_H3K4ME_3_nan_1_hg19
86	K562	50k 细胞	ATAC-seq	N/A	1	HiSeq 2000	单读段	50	51.962.519	hg19	98	34	K562_50K_ATAC-seq_nan_nan_1_hg19
87	K562	50k 细胞	ATAC-seq	N/A	2	HiSeq 2000	单读段	50	32.581.411	hg19	98	66	K562_50K_ATAC-seq_nan_nan_1_hg19

id	细胞类型	输入量	方案	抗体	重复	测序仪型号	读段类型	读段长度	读段计数	基因组	比对率	唯一率	内部样品 id
													seq_nan_nan_2_hg19

图8



B



C

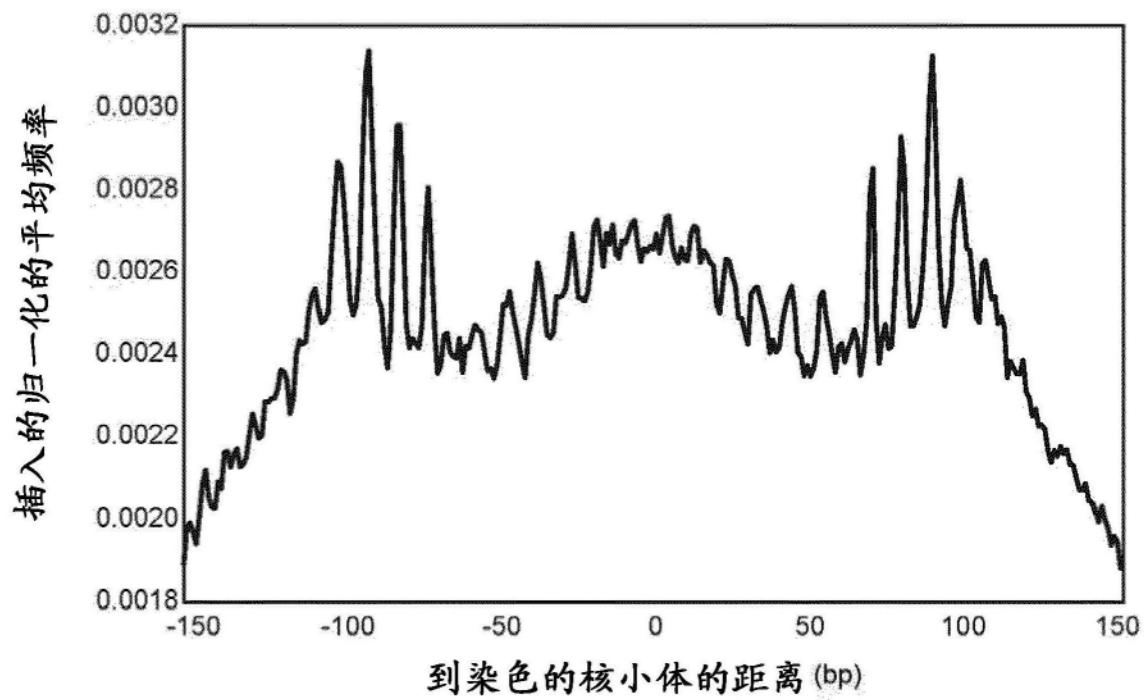


图9续

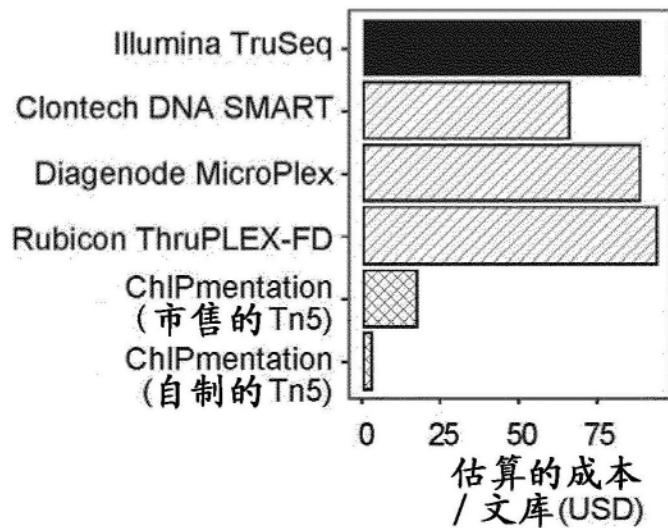
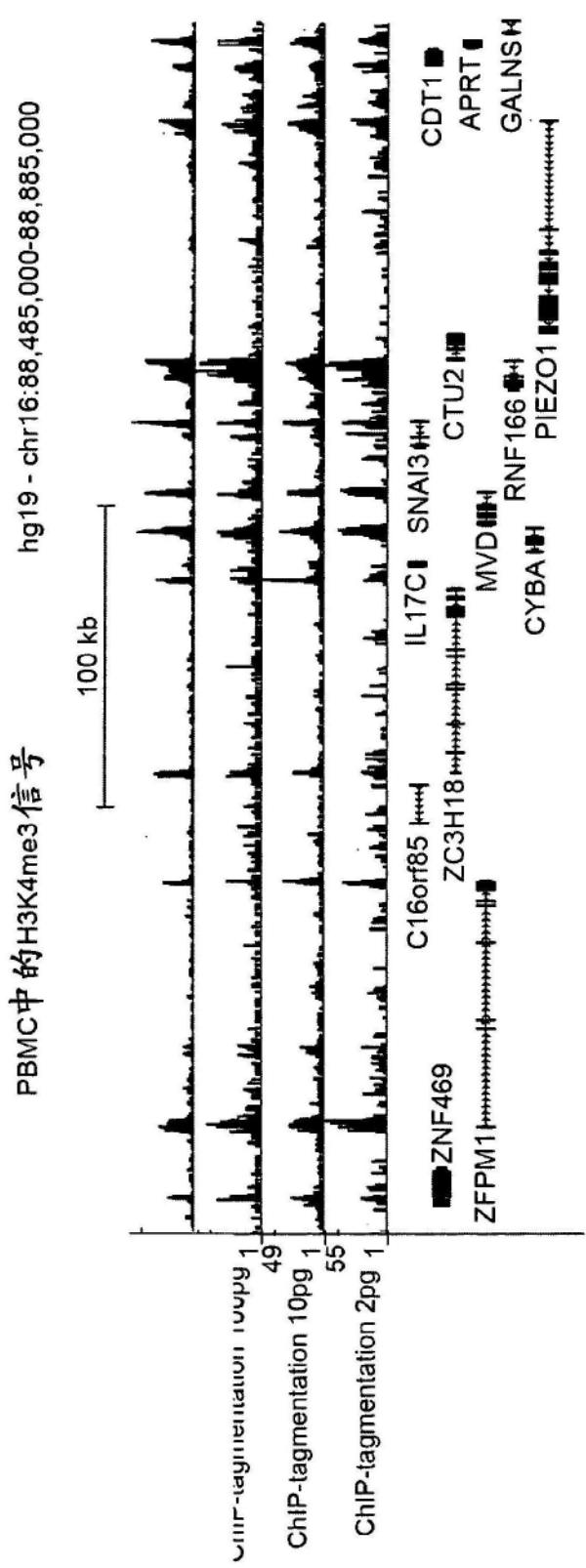


图10



B

来自具有不同量的输入 DNA 的标准 ChIP-seq
和 ChIP 标签化的 H3K4me3 测序的校正

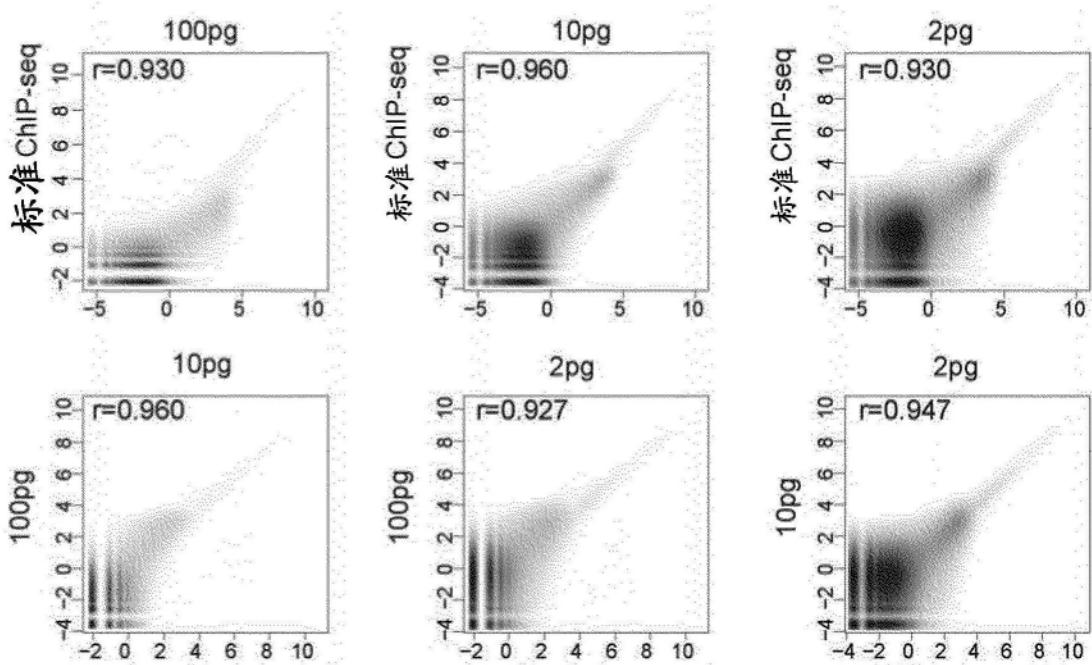


图11续

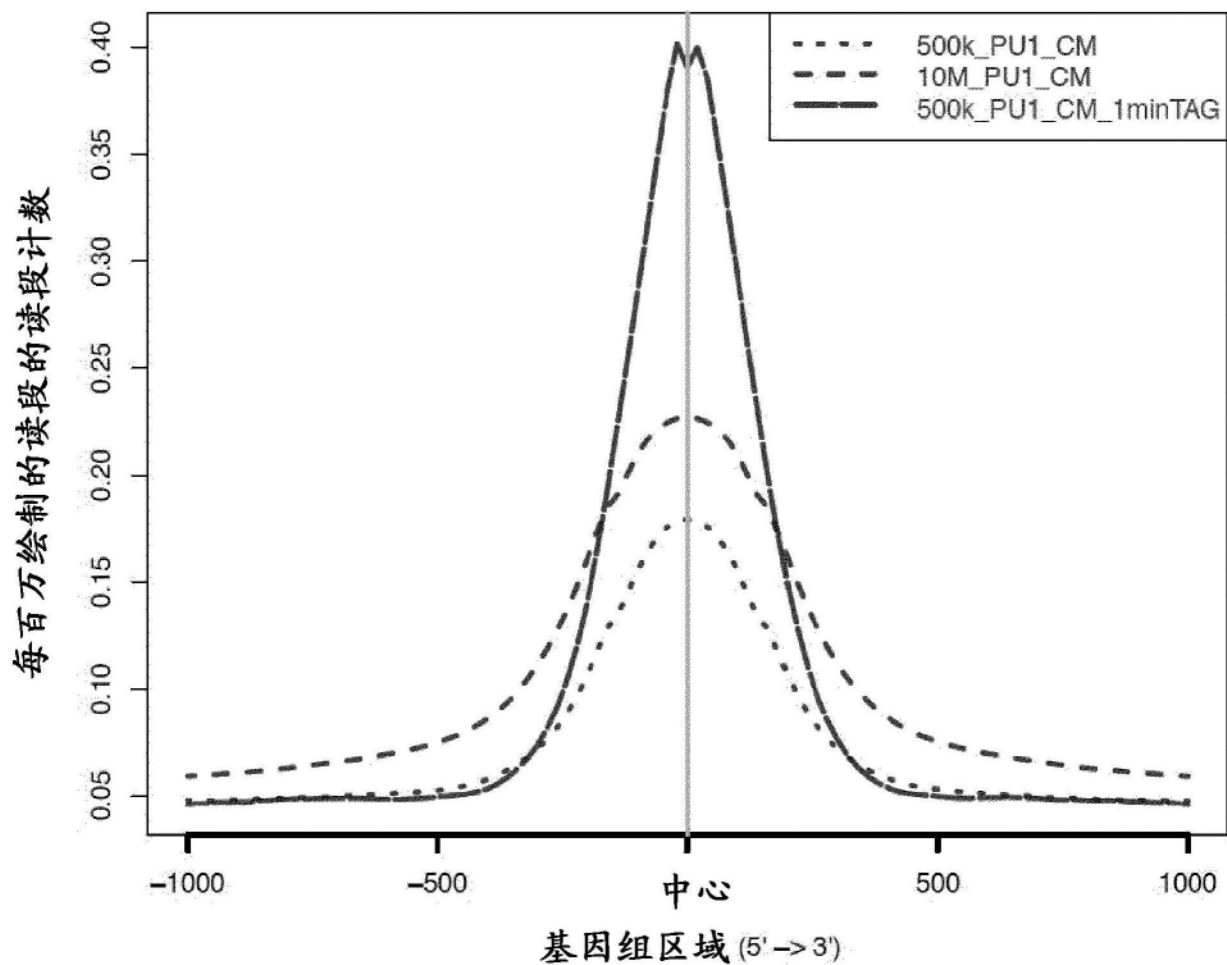


图12

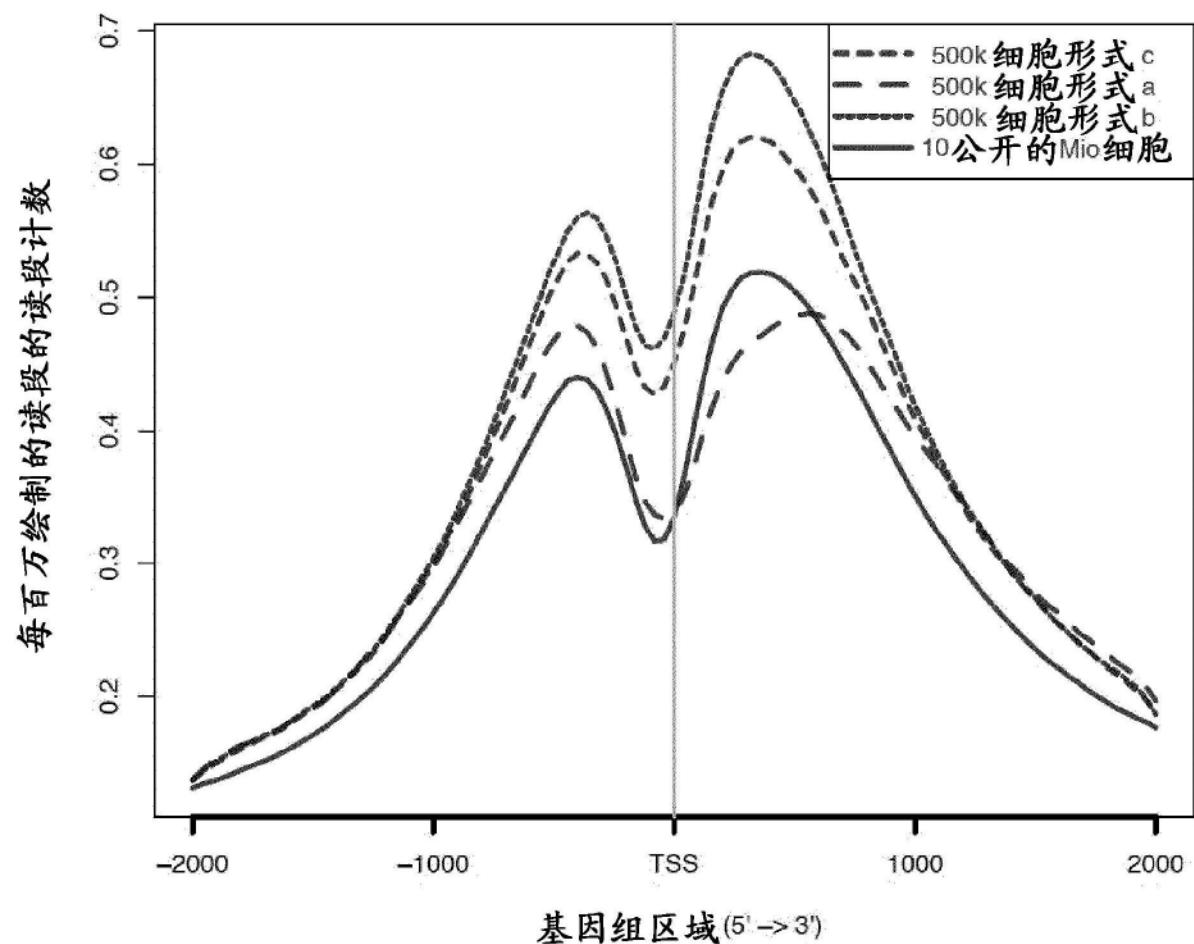


图13

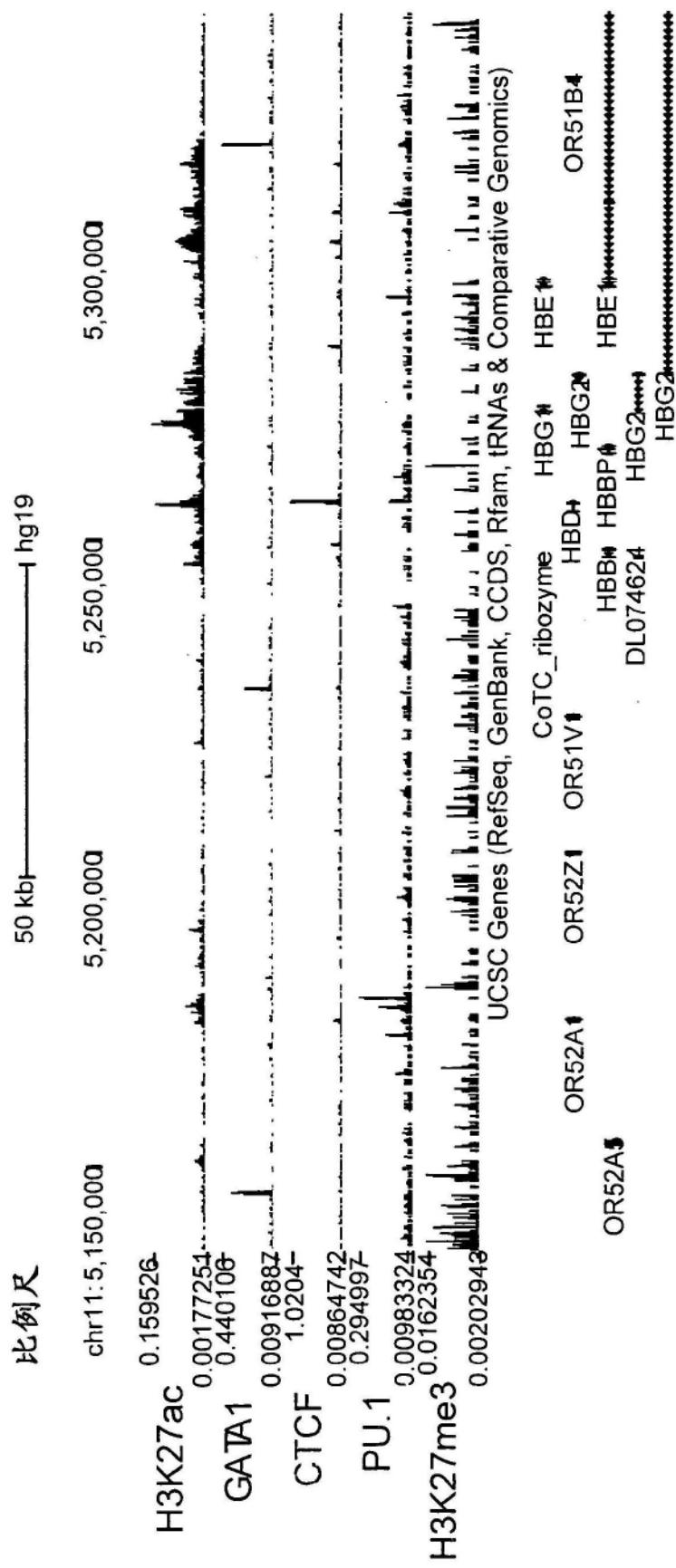


图14

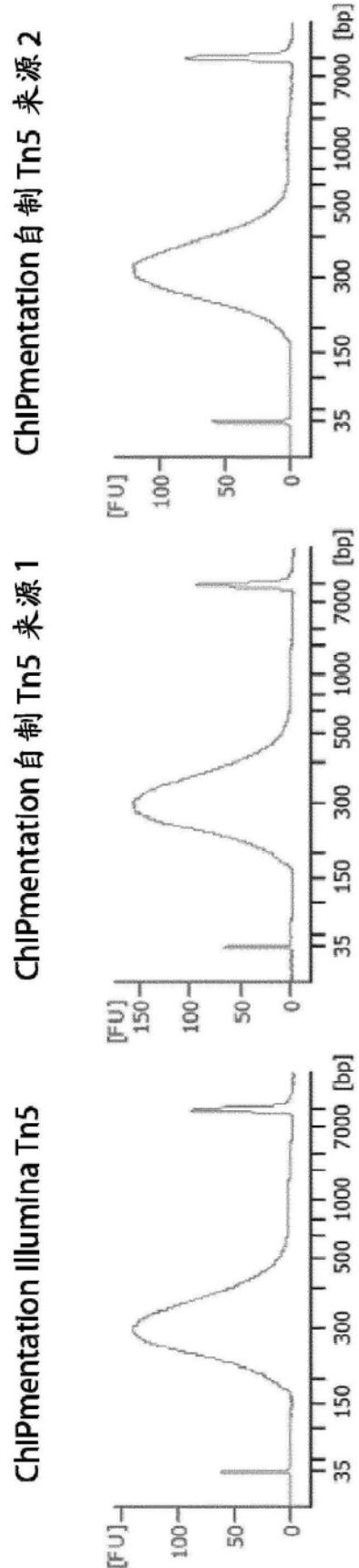


图15