



(12) 发明专利申请

(10) 申请公布号 CN 116917496 A

(43) 申请公布日 2023. 10. 20

(21) 申请号 202180094755.0

(22) 申请日 2021.12.29

(30) 优先权数据

63/131722 2020.12.29 US

(85) PCT国际申请进入国家阶段日

2023.08.28

(86) PCT国际申请的申请数据

PCT/EP2021/087813 2021.12.29

(87) PCT国际申请的公布数据

WO2022/144407 EN 2022.07.07

(71) 申请人 比利时意志有限责任公司

地址 比利时伊讷

(72) 发明人 J·V·米卡莱夫

M·E·埃克莱斯顿

D·F·F·帕玛尔特 M·赫尔佐格

(74) 专利代理机构 中国专利代理(香港)有限公司 72001

专利代理师 黄登高 张华

(51) Int.Cl.

C12Q 1/6806 (2006.01)

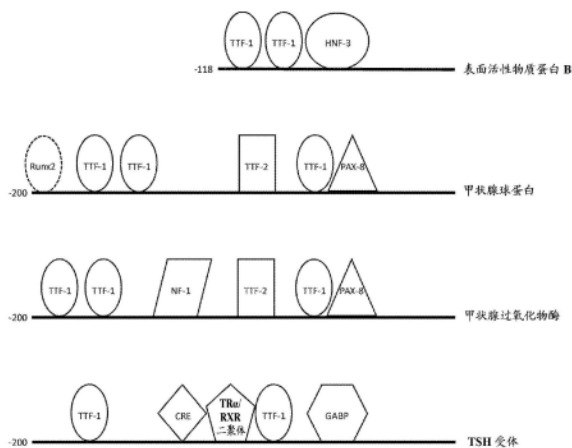
权利要求书2页 说明书47页
序列表2页 附图11页

(54) 发明名称

循环转录因子分析

(57) 摘要

本发明涉及通过用于检测循环染色质片段的微创体液测试来检测受试者中的疾病的方法, 所述染色质片段包括转录因子和缔合DNA序列作为受试者中疾病存在的指示物, 例如癌症。所述方法可以包括对缔合DNA序列测序和/或从体液去除无细胞核小体。



1. 检测在从人或动物受试者获得的体液样品中包含转录因子和DNA片段的无细胞染色质片段的方法,其包括以下步骤:

(i) 使所述体液样品与结合所述转录因子的结合剂接触;

(ii) 检测或测量与所述转录因子缔合的所述DNA片段;和

(iii) 使用所述DNA片段的的存在或量作为所述样品中包含所述转录因子的无细胞染色质片段的量的量度。

2. 根据权利要求1所述的方法,其包括在步骤(ii)中检测所述缔合DNA片段之前,从所述剩余的体液样品中分离在步骤(i)中结合的所述转录因子。

3. 根据权利要求1或权利要求2所述的方法,其中步骤(ii)包括对与所述转录因子缔合的所述DNA片段测序。

4. 根据权利要求1-3中任一项所述的方法,其另外包括提取与所述转录因子缔合的所述DNA片段。

5. 根据权利要求4所述的方法,其另外包括扩增所述提取的DNA片段,例如通过PCR。

6. 根据权利要求1-5中任一项所述的方法,其中通过实时PCR检测和/或测量与所述转录因子缔合的所述DNA片段。

7. 根据权利要求1-6中任一项所述的方法,其另外包括从所述体液样品中去除无细胞核小体。

8. 根据权利要求7所述的方法,其包括在步骤(ii)之前,使所述体液样品与结合核小体或其组分的结合剂接触,并去除与所述结合剂结合的所述样品。

9. 根据权利要求1-8中任一项所述的方法,其中所述无细胞染色质片段由所述转录因子和DNA片段组成。

10. 根据权利要求1-9中任一项所述的方法,其中在步骤(ii)中检测所述缔合DNA片段之前,用含有至少1%浓度的洗涤剂的缓冲溶液洗涤在步骤(i)中被所述结合剂结合的所述转录因子。

11. 检测人或动物受试者中的疾病的方法,其包括以下步骤:

(i) 使从人或动物受试者获得的体液样品与结合转录因子的结合剂接触;

(ii) 检测或测量与所述转录因子缔合的所述DNA;和

(iii) 使用DNA的存在或量作为所述受试者中疾病存在的指示物。

12. 根据权利要求11所述的方法,其包括使用所述转录因子和所述缔合DNA的序列作为指示所述受试者中所述疾病存在的组合生物标志物。

13. 检测人或动物受试者中受疾病影响的组织的方法,其包括以下步骤:

(i) 使从人或动物受试者获得的体液样品与结合转录因子的结合剂接触;

(ii) 对与所述转录因子缔合的所述DNA测序;和

(iii) 使用所述转录因子的存在和所述缔合DNA的序列作为用于确定所述受试者中受所述疾病影响的所述组织的组合生物标志物。

14. 根据权利要求13所述的方法,其中受所述疾病影响的所述组织是起源器官。

15. 根据权利要求11-14中任一项所述的方法,其中所述疾病是癌症或炎症疾病。

16. 根据权利要求1-15中任一项所述的方法,其中结合所述转录因子的所述结合剂是抗体或其片段。

17. 根据权利要求1-16中任一项所述的方法,其中所述体液样品是血液、血清或血浆样品。

18. 根据权利要求1-17中任一项所述的方法,其中所述体液样品是通过以下获得的血浆样品:(1)使全血样品与交联剂接触;(2)使所述交联的样品与钙离子螯合剂接触;和(3)从所述样品中分离血浆。

19. 用于评估动物或人受试者对医学治疗的适合性的方法,其包括以下步骤:

(i) 检测、测量或测序在从所述受试者获得的体液样品中与包含转录因子的无细胞染色质片段缔合的DNA;和

(ii) 使用在步骤(i)中检测的所述缔合DNA水平和/或序列作为参数来选择用于所述受试者的合适的治疗。

20. 用于监测动物或人受试者的治疗的方法,其包括以下步骤:

(i) 检测、测量或测序在从所述受试者获得的体液样品中与包含转录因子的无细胞染色质片段缔合的DNA;

(ii) 在一个或多个时机重复检测、测量或测序在从所述受试者获得的体液样品中与包含所述转录因子的无细胞染色质片段缔合的DNA;和

(iii) 使用与步骤(ii)相比在步骤(i)中检测的所述缔合DNA水平和/或DNA序列的任何变化作为所述受试者病况的任何变化的参数。

21. 根据权利要求20所述的方法,其中所述治疗用于治疗癌症。

22. 根据权利要求1-21中任一项所述的方法,其中作为一组测量之一检测或测量与包含所述转录因子的无细胞染色质片段缔合的所述DNA。

23. 用于检测包含转录因子和DNA片段作为组合生物标志物的无细胞染色质片段的试剂盒,其包含所述转录因子的配体或结合剂,任选地以及用于与所述转录因子缔合的DNA的扩增和/或测序的试剂,和/或核小体的配体或结合剂,和/或根据权利要求1-22中任一项所述的方法使用所述试剂盒的说明书。

24. 在有需要的受试者中治疗癌症的方法,其中所述方法包括以下步骤:

(a) 使从人或动物受试者获得的体液样品与结合转录因子的结合剂接触;

(b) 检测、测量或测序与所述转录因子缔合的DNA片段;和

(c) 使用DNA片段的的存在、量或序列作为所述受试者中癌症存在的指示物;和

(d) 如果在步骤(c)中确定所述受试者患有癌症,则施用治疗。

25. 根据权利要求24所述的方法,其中所述治疗选自:手术、放射疗法、化学疗法、免疫疗法、激素疗法和生物疗法。

26. 检测人或动物胎儿中的疾病的方法,其包括以下步骤:

(i) 从妊娠的人或动物受试者获得体液样品;

(ii) 使所述体液样品与结合转录因子的结合剂接触;

(iii) 检测、测量或测序与所述转录因子缔合的所述DNA;和

(iv) 使用DNA的存在、序列或量作为所述胎儿中疾病存在的指示物。

循环转录因子分析

技术领域

[0001] 本发明涉及通过微创体液测试用于检测受试者中的疾病的方法。本发明还涉及包括转录因子作为受试者中疾病存在的指示物的循环染色质片段的测量或检测。

背景技术

[0002] 癌症是具有高死亡率的常见疾病。疾病的生物学理解为涉及从癌前期状态到I、II、III期和最终IV期癌症的进展。对于大多数癌症疾病，死亡率变化很大，这取决于疾病是在早期局部阶段检测(此时可获得有效的治疗选择)还是在晚期检测(此时疾病可能已在受影响的器官内扩散或超过受影响的器官，此时治疗更困难)。晚期癌症症状是不同的，包括大便中可见的血液、尿液中的血液、咳嗽时排出的血液、从阴道排出的血液、无法解释的体重减轻、持续的无法解释的肿块(例如在乳房中)、消化不良、吞咽困难、变为疣或痣以及取决于癌症类型的许多其它可能的症状。然而，大多数由于这样的症状而诊断的癌症已经是晚期并难以治疗。大多数癌症在早期是无症状的，或者存在不帮助诊断的非特异性症状。因此，理想地，应当使用癌症测试早期检测癌症。

[0003] 为了解决对简单常规癌症血液测试的需要，已经研究了許多血源性蛋白作为潜在的癌症生物标志物，包括CRC的癌胚抗原(CEA)、肝癌的甲胎蛋白(AFP)、卵巢癌的CA125、胰腺癌的CA19-9、乳腺癌的CA15-3和前列腺癌的PSA。然而，它们的临床准确性对于常规诊断用途来说太低，并且它们被认为更好地用于患者监测。

[0004] 最近，本领域工作人员研究了循环肿瘤DNA(ctDNA)作为癌症检测的基于血液的生物标志物。无细胞DNA(cfDNA)在血液中作为染色质片段循环，认为染色质片段源自每天大量细胞的细胞死亡(主要通过凋亡)。在凋亡过程期间，染色质片段化成单核小体和寡核小体，其中一些从细胞释放以作为无细胞核小体循环。每个循环无细胞核小体与长度小于200个碱基对(bp)的小DNA片段缔合。类似地，已经从片段组学分析推断出循环中由DNA结合的转录因子或其它非组蛋白染色质蛋白质组成的无细胞染色质片段。在健康受试者中，认为循环染色质片段是造血起源的，并且水平低。在患有多种病况的受试者中发现了升高水平的循环核小体(并因此cfDNA片段)，所述病况包括许多癌症、自身免疫疾病、炎性病况、中风和心肌梗死(Holdenrieder&Stieber, 2009)。

[0005] 认为癌症患者血液中的至少一些cfDNA源自核小体和其它染色质片段从濒死或死亡的癌细胞释放到循环中(即cfDNA包括一些ctDNA)。对来自癌症患者的匹配的血液和组织样品的研究显示，存在于患者的肿瘤中(但不存在于他/她的健康细胞中)的癌症相关的突变也存在于取自同一患者的血液样品中的cfDNA中(Newman等人, 2014)。类似地，在癌细胞中差异甲基化(通过胞嘧啶残基的甲基化表观遗传学改变)的DNA序列也可以作为循环中cfDNA中的甲基化的序列来检测。此外，由ctDNA组成的循环cfDNA的比例与肿瘤负荷有关，因此疾病进展可以通过存在的ctDNA的比例定量监测，也可以通过其遗传和/或表观遗传组成定性监测。ctDNA的分析可以产生高度有用和临床精确的数据，该数据涉及源自肿瘤内所有或许多不同克隆的DNA，并因此其在空间上整合肿瘤克隆。此外，随着时间的推移重复血

液取样是比例如重复组织活检实用且经济得多的选择。ctDNA的分析具有改革肿瘤的检测和监测的潜力,以及在早期检测复发和获得的耐药性,以通过研究肿瘤DNA而不用侵入性组织活检程序来选择肿瘤的治疗。这样的ctDNA测试可以用于研究所有类型的癌症相关的DNA异常(例如点突变、核苷酸修饰状态、易位、基因拷贝数、微卫星异常和DNA链完整性),并且将适用于常规癌症筛选、常规和更频繁的监测以及最佳治疗方案的常规检查(Zhou等人,2017)。

[0006] 血液血浆通常用作ctDNA测定的底物。从血浆中提取cfDNA片段(包括任何ctDNA)(并因此从与核小体、转录因子或其它蛋白质的结合中去除),并分析核苷酸碱基序列。可以采用任何DNA分析方法,但通常通过使用下一代测序仪仪器的深度测序进行分析。

[0007] 由于DNA异常是所有癌症疾病的特征,并且已经观察到ctDNA用于所有已经研究过的癌症疾病,ctDNA测试在所有癌症疾病中都具有适用性。研究的癌症包括但不限于膀胱癌、乳腺癌、结肠直肠癌、黑素瘤、卵巢癌、前列腺癌、肺癌、肝癌、子宫内膜癌、卵巢癌、淋巴瘤、口腔癌、白血病、头颈癌和骨肉瘤(Crowley等人,2013;Zhou等人,2017;Jung等人,2010)。

[0008] cfDNA分析的一个实例方法涉及受试者的cfDNA片段的起源组织或细胞的鉴定。这种方法的基础是,循环中存在的所有cfDNA片段避免了细胞死亡期间或循环中被核酸酶消化,因为它们受到核小体内蛋白质结合的保护而免受核酸酶作用。该方法涉及测定取自受试者的血液样品中cfDNA的核小体片段化模式,并在参考基因组中定位cfDNA片段的基因组位置。片段化的模式对于不同的细胞类型是不同的,并且可以用于鉴定受试者的cfDNA的起源细胞。

[0009] 该方法涉及从血浆样品中提取cfDNA(包括任何ctDNA)和对DNA的全基因组测序以检测cfDNA片段所展示的核小体结合的DNA模式。通过计算机分析,使用生物信息学,将cfDNA片段的终点序列定位在一个或多个参考基因组内的基因组位置。参考基因组内cfDNA终点的基因组位置提供了基因组的核小体保护的cfDNA覆盖度的图谱。

[0010] 使用生物信息学通过计算机分析,不同细胞类型或组织对受试者中cfDNA的比例贡献也可以通过比较受试者的核小体片段化模式与含有已知相对丰度的来自不同细胞来源的cfDNA的校准样品来确定,如在W02017012592中所述。

[0011] 与含有核小体的染色质片段缔合的cfDNA片段通常长度为120-200bp。然而,cfDNA的蛋白质结合和保护不限于核小体中cfDNA的组蛋白结合。除了核小体以外或在不存在任何核小体下,其它cfDNA片段(包括活性基因启动子序列)被转录因子、辅因子或其它非组蛋白染色质蛋白质结合。在不存在核小体下,这些蛋白质通常结合并保护在35-80bp范围的较短的cfDNA片段。然而,如果所用的DNA片段文库制备方法适于分离、扩增和测序长度小于100个碱基对的短DNA片段,则仅在实验上观察到这些较短的cfDNA片段(Snyder等人,2016)。

[0012] 活细胞中DNA在基因组中的蛋白质结合的模式随细胞类型而变化,因为不同的DNA序列(包括不同的启动子序列和基因序列)在不同的细胞中有活性。任何细胞类型中DNA的蛋白质结合的模式可以通过核酸酶可接近位点作图来确定,其通过用核酸酶消化从细胞中提取的染色质并对所得蛋白质保护的染色质片段中未消化的DNA测序。因此,如果将血液中的cfDNA片段看作是体内核酸酶消化的产物,则发现的cfDNA序列应对应于从中起源cfDNA

的细胞中的蛋白质结合的DNA序列。因此,原则上,血液中cfDNA片段序列的模式应与通过起源细胞的核酸酶可接近位点作图产生的染色质片段序列的模式类似。因此,从血液样品测定的cfDNA序列的片段化模式可以使用生物信息学方法与通过已知组织或癌症类型的细胞的核酸酶可接近位点分析产生的已知DNA片段化模式进行比较,以确定cfDNA的起源组织。取自健康受试者的样品的结果指示cfDNA的起源细胞是造血的。取自癌症患者的样品中该方法的结果指示cfDNA和ctDNA源自包括造血细胞和其它细胞的细胞混合物。在许多情况下,所指示的非造血细胞类型与患者的癌症疾病的组织相关(Snyder等人,2016)。

[0013] 其他工作者使用了类似的cfDNA片段终点分析方法,涉及全基因组cfDNA测序(包括任何ctDNA),但是将生物信息学计算机分析集中在转录因子结合位点(TFBS)序列上。该方法的目的是测定TFBS可接近性,并鉴定取自患有癌症的患者的血浆样品中具有改变的可接近性的TFBS DNA序列(U1z等人,2019)。在该方法中,从受试者采集血浆样品,并使用适于长度小于100bp的小DNA片段的DNA文库制备方法提取和扩增cfDNA。使用下一代测序方法对DNA文库测序。测序数据用于使用生物信息学方法鉴定TFBS附近基因组区中的cfDNA片段化模式。该分析涉及测定cfDNA片段跨越基因启动子序列中TFBS及其侧翼序列的核小体定位特性,以确定TFBS是否与包含cfDNA的染色质片段中的转录因子结合。该方法是复杂的,但可以总结如下:

[0014] 如果在基因组中跨TFBS和侧翼序列的DNA序列中观察到的cfDNA片段化模式显示约200bp的周期性,这涉及DNA的交替的更强的蛋白质结合保护(在核小体结合位置的中心)和更弱的蛋白质结合保护(在DNA未结合和未保护的核小体之间)免于降解。在这种情况下,假定TFBS和侧翼序列已经被包含血浆样品中cfDNA的染色质片段所覆盖。

[0015] 如果存在的cfDNA片段化模式另外显示TFBS及其侧翼序列的蛋白质结合保护,但没有(或减弱的)核小体相关的周期性,这涉及转录调节蛋白在TFBS及其侧翼序列的结合。在这种情况下,假定TFBS已经与包含血浆样品中cfDNA的染色质片段中的一种或多种转录因子和/或其它调节蛋白结合。

[0016] 在健康受试者中,发现的cfDNA片段化模式通常与造血细胞的核酸酶可接近位点实验获得的模式相关。因此,与转录因子结合的TFBS序列或cfDNA中覆盖的核小体与造血细胞中表达或不表达的转录因子相关。在癌症患者中,该模式涉及细胞类型的混合物,其中TFBS可以是在癌细胞类型中结合的转录因子和在造血细胞类型中结合的核小体。由于大多数cfDNA衍生自造血细胞,并且仅少量衍生自癌细胞,与造血信号相比,癌症衍生的片段组学信号小。然而,已经开发了片段组学生物信息学方法,以从存在于造血衍生的cfDNA组分中的大得多的叠加的核小体周期性信号中解开存在于ctDNA中的小转录因子保护的TFBS片段信号。片段组学分析指示,混合模式包括cfDNA TFBS序列,其是与不在造血细胞中表达但由癌组织表达的转录因子结合的转录因子。

[0017] 染色质免疫沉淀然后对染色质缔合DNA测序(ChIP-Seq)是用于对细胞染色质蛋白质的基因组位置作图的分析技术。典型的方法涉及从细胞提取染色质,然后通过物理破坏(例如,超声处理)或通过使用裂解DNA的核酸酶(例如,脱氧核糖核酸酶或微球菌核酸酶)将染色质消化成单核小体或其它染色质片段。然后将片段化的染色质暴露于涂布有抗体的固相支持物,所述抗体被定向结合特定的目的染色质蛋白质,例如特定的修饰的组蛋白。包含特定结构的染色质片段被吸附(免疫沉淀)到固相上。然后从固相提取与吸附的染色质缔合

的DNA并通过聚合酶链反应(PCR)方法扩增。对扩增的DNA片段文库测序以确定基因组内结合目的染色质蛋白质的位置。使用针对转录因子的抗体的ChIP方法也用于鉴定特定转录因子的转录因子结合位点(TFBS)的基因组位置,或者鉴定不同细胞类型中特定TFBS是否被特定转录因子占据。

[0018] 我们先前已经描述了用于检测癌症和其它疾病的循环无细胞核小体的免疫测定测试,所述核小体含有特定表观遗传信号,包括特定翻译后修饰、组蛋白同种型、修饰的核苷酸和非组蛋白染色质蛋白质(如W02005019826、W02013030577、W02013030579和W02013084002中所引用的)。我们还描述了用于检测癌症的染色质片段(包括转录因子结合的DNA)的免疫测定测试(如W02017162755中所引用的)。

[0019] 我们现在报道了具有优异的分析 and 临床特异性和灵敏度的方法,用于分离和直接分析和测量含有一种或多种转录因子以及缔合DNA片段的循环无细胞染色质片段。从多得多的核小体片段中分离转录因子-DNA复合物简化了分析,并且消除从占主导的核小体周期性信号中解开转录因子覆盖的TFBS信号的需要。所述方法可以在血液样品中作为非侵入性或微创血液测试用于疾病,包括癌症、自身免疫疾病和炎性疾病。

发明内容

[0020] 根据本发明的第一方面,提供了检测在从人或动物受试者获得的体液样品中包含转录因子和DNA片段的无细胞染色质片段的方法,其包括以下步骤:

[0021] (i)使所述体液样品与结合所述转录因子的结合剂接触;

[0022] (ii)检测或测量与所述转录因子缔合的所述DNA片段;和

[0023] (iii)使用所述DNA片段的的存在或量作为所述样品中包含所述转录因子的无细胞染色质片段的量的量度。

[0024] 根据本发明的其它方面,提供了检测人或动物受试者中的疾病的方法,其包括以下步骤:

[0025] (i)使从人或动物受试者获得的体液样品与结合转录因子的结合剂接触;

[0026] (ii)检测或测量与所述转录因子缔合的所述DNA;和

[0027] (iii)使用DNA的存在或量作为所述受试者中疾病存在的指示物。

[0028] 根据本发明的其它方面,提供了检测人或动物受试者中受疾病影响的组织的方法,其包括以下步骤:

[0029] (i)使从人或动物受试者获得的体液样品与结合转录因子的结合剂接触;

[0030] (ii)对与所述转录因子缔合的所述DNA测序;和

[0031] (iii)使用所述转录因子的存在和所述缔合DNA的序列作为用于确定所述受试者中受所述疾病影响的所述组织的组合生物标志物。

[0032] 根据本发明的其它方面,提供了用于评估动物或人受试者对医学治疗的适合性的方法,其包括以下步骤:

[0033] (i)检测、测量或测序在从所述受试者获得的体液样品中与包含转录因子的无细胞染色质片段缔合的DNA;和

[0034] (ii)使用在步骤(i)中检测的所述缔合DNA水平和/或序列作为参数来选择用于所述受试者的合适的治疗。

[0035] 根据本发明的其它方面,提供了用于监测动物或人受试者的治疗的方法,其包括以下步骤:

[0036] (i) 检测、测量或测序在从所述受试者获得的体液样品中与包含转录因子的无细胞染色质片段缔合的DNA;

[0037] (ii) 在一个或多个时机重复在从所述受试者获得的体液样品中与包含所述转录因子的无细胞染色质片段缔合的DNA的所述检测、测量或测序;和

[0038] (iii) 使用与步骤(ii)相比在步骤(i)中检测的所述缔合DNA水平和/或DNA序列的任何变化作为所述受试者病况的任何变化的参数。

[0039] 根据本发明的其它方面,提供了用于检测包含转录因子和DNA片段作为组合生物标志物的无细胞染色质片段的试剂盒,其包含所述转录因子的配体或结合剂,任选地还有用于与所述转录因子缔合的DNA的扩增和/或测序的试剂,和/或核小体的配体或结合剂,和/或根据如本文所定义的方法使用所述试剂盒的说明书。

[0040] 根据本发明的其它方面,提供了在有需要的受试者中治疗癌症的方法,其中所述方法包括以下步骤:

[0041] (a) 使从人或动物受试者获得的体液样品与结合转录因子的结合剂接触;

[0042] (b) 检测、测量或测序与所述转录因子缔合的DNA片段;和

[0043] (c) 使用DNA片的存在或量作为所述受试者中癌症的存在的指示物;和

[0044] (d) 如果在步骤(c)中确定所述受试者患有癌症,则施用治疗。

[0045] 检测人或动物胎儿中的疾病的方法,其包括以下步骤:

[0046] (i) 从妊娠的人或动物受试者获得体液样品;

[0047] (ii) 使所述体液样品与结合转录因子的结合剂接触;

[0048] (iii) 检测、测量或测序与所述转录因子缔合的所述DNA;和

[0049] (iv) 使用DNA的存在、序列或量作为所述胎儿中疾病存在的指示物。

附图说明

[0050] 图1:在表面活性物质蛋白B、甲状腺球蛋白、甲状腺过氧化物酶和促甲状腺素受体(TSH受体)基因的启动子位点上各种转录因子的共结合的卡通图说明。CRE:环腺苷单磷酸盐应答元件;GABP:GA-结合蛋白;HNF-3:肝细胞核因子3;NF-1:核因子1;PAX-8:配对盒基因8;Runx2:Runt相关转录因子2;TR α /RXR二聚体:甲状腺激素受体 α /类视黄醇X受体二聚体;TTF-1:甲状腺转录因子1(也称为NK2同源异型盒1,NKX2-1);TTF-2:甲状腺转录因子2。

[0051] 图2:转录复合物的DNA环结构的实例的卡通图,以说明参与转录复合物的各种调节蛋白中的一些的共结合,所述调节蛋白包括但不限于通用转录因子(GTF)、基因特异性转录因子(TF)、辅因子、激活物、阻抑物、介体、DNA弯曲蛋白和RNA聚合酶。调节蛋白与位于基因附近的调节DNA序列以及远离基因的调节序列结合,包括启动子序列、TATA盒序列、增强子序列和阻抑物序列。其它调节蛋白(例如染色质重塑蛋白)以及其它调节序列也是可能的。

[0052] 图3:吸附在涂布有定向结合组蛋白H3的抗体的磁性珠上的重组单核小体的蛋白质印迹分析。结果证明单核小体的剂量依赖性吸附。

[0053] 图4:在核小体的免疫沉淀后,使用未涂布的磁性珠或涂布有定向结合组蛋白H3的

抗体的磁性珠,人血浆样品和重组单核小体溶液的核小体ELISA结果。结果证明溶液中天然存在的人循环核小体和重组核小体两者不受未涂布的磁性珠影响,但使用涂布有定向结合组蛋白H3的抗体的磁性珠,人循环核小体和重组核小体两者通过免疫沉淀定量去除。

[0054] 图5:在诊断患有ER阴性乳腺癌(ER-BC)、卵巢癌或ER阳性乳腺癌(ER+BC)的女性中测量的ER α 的水平,其中ER评分为7或8。

[0055] 图6:用含有0.1% Tween (0.1%)的常规单一洗涤剂洗涤缓冲液或用含有总计1.2%洗涤剂(1.2%)的洗涤剂混合物的强洗涤缓冲液洗涤暴露于从癌症患者获得的血浆样品的磁性聚苯乙烯颗粒的作用。非特异性IgG涂布的颗粒通过使用强洗涤剂洗涤(泳道4和5)显示背景结合的更大降低,而不破坏特异性抗体结合的蛋白(parylated蛋白质的混合物)(泳道6和7)。

[0056] 图7:使用固定在磁性聚苯乙烯珠上的小鼠抗CTCF抗体,通过ChIP从取自诊断患有CRC的患者的4个汇集的交联的EDTA血浆样品免疫沉淀的染色质片段的蛋白质印迹分析,所述磁性聚苯乙烯珠使用强的1.2%洗涤剂混合洗涤缓冲液洗涤。所有4个血浆样品在约140kD处显示对应于CTCF蛋白的带(抗CTCF;泳道3、5、7和9)。使用非特异性小鼠IgG的阴性对照实验显示没有对应于CTCF的带(NS-IgG;泳道2、4、6和8)。实验证明CTCF蛋白从血浆样品中分离,并且使用强洗涤缓冲液导致相对纯的来自血浆的CTCF提取物。

[0057] 图8:显示来自取自患者的交联的EDTA血浆样品中CTCF染色质片段的ChIP的扩增的接头连接的cfDNA片段文库的分析的电泳图谱。在约140bp处的尖锐峰代表接头二聚体,因此175-220bp的接头连接的片段代表35-80bp的cfDNA片段(在电泳图谱上指示)。(a)特异性CTCF ChIP文库含有具有在35-80bp范围的约1000FU的荧光峰的小的cfDNA片段。(b)非特异性对照IgG文库还含有具有约80FU的荧光峰的小的cfDNA片段。

[0058] 图9:9780个公开的CTCF TFBS基因座被转录因子结合的(35-80bp)或核小体结合的(135-155bp或156-180bp)cfDNA片段的标准化覆盖度。(a)通过对于CRC患者获得的cfDNA序列文库的特异性CTCF覆盖度。(b)通过从与小鼠IgG涂布的颗粒非特异性结合的染色质片段获得的cfDNA序列文库的非特异性覆盖度。结果显示,源自血浆循环CTCF-DNA复合物的特异性cfDNA覆盖度的峰与公开的CTCF TFBS基因座相关。在所研究的5kb跨度上,由于核小体结合的预期的振荡覆盖度模式是最小的。在对照样品中,在CTCF结合基因座没有观察到cfDNA覆盖度峰。

[0059] 图10:1041个公开的CTCF TFBS基因座的标准化覆盖度,所述基因座在癌细胞中被CTCF占据,但在正常细胞中未被占据。显示转录因子结合的(35-80bp)或核小体结合的(135-155bp或156-180bp)cfDNA片段的覆盖度。(a)通过对于CRC患者获得的cfDNA序列文库对癌症相关基因座的CTCF占据。结果显示在35-80bp尺寸范围的覆盖度,证实这些1041个位点中的一些或全部的CTCF占据,并因此指示在采集样品的受试者中的癌症。(b)在非特异性对照实验中没有观察到CTCF占据峰。

[0060] 图11:使用固定在磁性聚苯乙烯珠上的小鼠抗AR抗体,通过ChIP从8个交联的EDTA血浆样品免疫沉淀的染色质片段的蛋白质印迹分析,所述磁性聚苯乙烯珠使用强的1.2%洗涤剂混合洗涤缓冲液洗涤。所有8个血浆样品(S1-S8;泳道2-9)在约140kD处显示对应于AR蛋白质的带。对于样品S1和S2观察到最高密度带。泳道10代表使用来自LnCAP前列腺癌细胞的片段化的染色质的阳性对照。

[0061] 图12:显示取自8名前列腺癌患者的交联的EDTA血浆样品(S1-S8)中AR染色质片段的ChIP的扩增的接头连接的cfDNA片段文库的分析的电泳图谱。在约140bp处的尖锐峰代表接头二聚体,因此175-220bp的接头连接的片段代表35-80bp的cfDNA片段。还显示阴性对照(ctrl)的电泳图谱。

具体实施方式

[0062] 转录因子参与癌症,并且占有已知癌基因的约20%(Lambert等人,2018)。我们先前已经描述了含有组织特异性转录因子的染色质片段作为血清中用于检测或诊断受试者中癌症的生物标志物的用途。转录因子的组织特异性可以用于指示癌症的起源组织。例如,据报道转录因子TTF-1在甲状腺和肺组织中表达,而在其它组织中不表达。因此,含有TTF-1的循环染色质片段的的存在指示起源组织是肺或甲状腺。我们还描述了用于测量含有转录因子的循环无细胞染色质片段的免疫测定方法。这种免疫测定涉及双抗体(或其它结合剂)方法,其中一种抗体定向结合转录因子,而另一种抗体定向结合与转录因子缔合的DNA或结合包括在染色质片段中的核小体组分。在所述的一个实施方案中,将靶向结合转录因子的结合剂固定在固相上以分离含有转录因子的染色质片段(即免疫沉淀染色质片段)。然后使用定向结合DNA的第二结合剂检测分离的染色质片段。该免疫测定方法简单、成本低并且是非侵入性的。

[0063] ChIP-Seq是通常应用于细胞染色质提取物的方法,其通过用核酸酶消化或通过超声处理片段化。有一些关于ChIP-Seq方法在EDTA血浆中应用的报道。由于血浆中的染色质已经片段化,不需要对样品核酸酶消化或超声处理。血浆中ChIP-Seq的报道涉及使用抗组蛋白抗体从EDTA血浆中分离组蛋白,然后提取、扩增和测序组蛋白缔合DNA片段(Deligezer等人,2008,Mansson等人,2021,Sadeh等人,2021,Vad-Nielsen等人,2020)。

[0064] 就作者所知,在文献中没有描述用于直接分离、分析或作图完整的循环转录因子-DNA染色质片段和缔合TFBSDNA序列的ChIP-Seq方法。相反,本领域的工作人员已经开发了基于DNA片段分析的间接方法。

[0065] 片段组学是一种这样的间接方法,其中通过生物信息学方法分析从EDTA血浆中提取的cfDNA的深度测序,以鉴定指示起源样品中转录因子-DNA结合的DNA片段化模式(Snyder等人,2016,Ulz等人,2019)。这是间接方法,因为片段化组学的第一步是提取所研究样品中的所有DNA,并且这必然涉及破坏存在的所有转录因子-DNA复合物。这破坏将任何DNA片段或序列与样品中的任何转录因子或其它染色质蛋白质直接连接的所有信息。TFBS的占据从提取的DNA文库中适当序列的短cfDNA片段(35-80bp)的存在推断。然而,附着于DNA片段的染色质蛋白质的同一性(在DNA提取之前)可能是未知的,特别是如图1和2所示,许多蛋白质可能在非常接近目的位点的位置结合。因此片段组学方法的一个缺点是可以推断但不能建立任何特定转录因子在任何特定TFBS处的结合。

[0066] 另一种最近的间接方法涉及EDTA血浆中核小体ChIP-Seq以直接作图无细胞核小体定位,并使用核小体定位数据间接推断转录因子定位(Sadeh等人,2021)。

[0067] 还没有报道用于转录因子-DNA复合物的直接ChIP方法的原因是存在迄今为止还没有解决的显著的技术困难或障碍。这些技术困难包括(i)认识到一些转录因子-DNA复合物在血浆中稳定缔合,而在体内动态缔合的其它转录因子-DNA复合物在血液或其它体液中

解离, (ii) 认识到最常见类型的转录因子-DNA复合物在EDTA血浆中解离, 但可以防止这一点, (iii) 来自细胞或组织材料的核提取物是相对纯的染色质制剂, 其可以以 μg 或 mg 的量获得。相反, 血液、血清或血浆含有非常低水平的非常不纯的染色质, 所述染色质被高水平的其它循环蛋白“污染”, (iv) 存在至少数百种转录因子, 并且任何特定转录因子-DNA复合物将仅是血浆中存在的数千种不同的转录因子-DNA复合物中的一种。进而, cfDNA的总转录因子-DNA级分是总cfDNA(其中大部分包含核小体片段)的一小部分, 并且源自癌细胞的cfDNA的比例是总cfDNA的一小部分。因此, 包括任何特定转录因子的转录因子-DNA复合物是被高水平的其它蛋白质和其它物质污染的小部分的小部分的小部分。这样的一个小结果是在血浆转录因子-DNA CHIP-Seq方法中产生的特异性信号小(小于背景信号), 使得有效的数据分析成为问题。

[0068] 我们现在报道了用于检测含有转录因子-DNA复合物的循环无细胞染色质片段的方法, 所述方法具有优异的分析灵敏度和优异的组织特异性。该方法还扩大了适用的转录因子的用途, 以包括大多数或所有转录因子。

[0069] 我们还报道了由含有转录因子的染色质片段和与所述转录因子缔合的DNA片段的序列组合组成的组合生物标志物用于检测疾病的用途。该组合生物标志物另外具有非常高的组织特异性, 并且可以用作癌症的生物标志物。

[0070] 分析灵敏度对于含有转录因子的循环无细胞染色质片段是重要的, 所述转录因子以低水平、接近或低于免疫测定的检测极限而出现。免疫测定的检测分析极限随测定的设计和所用结合剂(通常是抗体)的亲合力而变化, 但可以在皮摩尔浓度范围。然而, DNA的聚合酶链反应(PCR)检测的分析灵敏度低几个数量级。数字PCR可以检测低至每个样品几个单独分子的浓度。因此, 使用PCR扩增方法检测与转录因子缔合的DNA, 而不是使用定向结合DNA(或核小体表位)的抗体, 能够检测含有极低水平的转录因子的循环染色质片段。

[0071] 除了通过使用PCR检测提高灵敏度之外, 通过处理不包含缔合的核小体转录因子的转录因子的大汇集, 对含有转录因子的染色质片段基于其缔合DNA含量的分析也导致高的分析灵敏度。

[0072] 因此, 根据本发明的第一方面, 提供了检测在从人或动物受试者获得的体液样品中包含转录因子和DNA片段的无细胞染色质片段的方法, 其包括以下步骤:

[0073] (i) 使体液样品与结合转录因子的结合剂接触;

[0074] (ii) 检测或测量与转录因子缔合的DNA片段; 和

[0075] (iii) 使用DNA片段的的存在或量作为样品中包含转录因子的无细胞染色质片段的量的量度。

[0076] 在一个实施方案中, 将在步骤(i)中使用的转录因子的抗体或其它结合剂固定在固相上以从样品中分离转录因子。

[0077] 在一个实施方案中, 该方法包括在检测缔合DNA片段之前, 从剩余的体液样品中分离在步骤(i)中结合的转录因子。例如, 可以将洗涤缓冲液应用于在步骤(i)中与(固相)结合剂结合的样品中的转录因子, 以去除未与结合剂结合的剩余样品。

[0078] 在一个实施方案中, 从转录因子提取转录因子缔合的DNA片段, 用于在步骤(ii)中检测、测量或测序DNA片段。

[0079] 在一个实施方案中, 使用常规DNA结合剂(例如抗DNA抗体)或DNA螯合剂或嵌入剂

(例如溴化乙锭)和花青染料(例如SYBR绿和SYBR金)来检测或测量DNA。

[0080] 在一个实施方案中,步骤(ii)包括对与转录因子缔合的DNA片段测序。测序方法是本领域公知的。

[0081] 根据一些实施方案,在步骤(ii)中检测或测量DNA片段通过DNA片段的扩增进行,例如使用定量PCR方法来确定DNA片段的的存在和/或量。因此,根据本发明的其它方面,提供了检测人或动物受试者中包含转录因子和DNA片段的无细胞染色质片段的方法,其包括以下步骤:

[0082] (i)使从人或动物受试者获得的体液样品与结合转录因子的结合剂接触;

[0083] (ii)分离与转录因子缔合的DNA;

[0084] (iii)扩增DNA;和

[0085] (iv)使用DNA片段的的存在或量作为样品中包含转录因子的无细胞染色质片段的的存在或量的量度。

[0086] 在一个实施方案中,使用DNA杂交方法检测或测量扩增的DNA。

[0087] 在进一步的实施方案中,在将接头寡核苷酸与DNA片段连接后,进行转录因子结合的DNA片段的扩增。接头寡核苷酸可以包括引物序列以促进通过PCR扩增DNA片段,或者可以随后加入引物序列。涉及接头寡核苷酸的方法是本领域公知的,并且常规用于制备用于下一代测序的文库。因此,在本发明的一个实施方案中,提供了检测人或动物受试者中包含转录因子和DNA片段的无细胞染色质片段的方法,其包括以下步骤:

[0088] (i)使从人或动物受试者获得的体液样品与结合转录因子的结合剂接触;

[0089] (ii)分离与转录因子缔合的DNA;

[0090] (iii)将接头寡核苷酸与分离的DNA连接;

[0091] (iv)扩增DNA;和

[0092] (v)使用DNA片段的的存在或量作为样品中包含转录因子的无细胞染色质片段的量的量度。

[0093] 在一个实施方案中,使用设计用于扩增包括一个或多个特定序列的DNA片段的一个或多个特异性序列的PCR引物寡核苷酸进行转录因子结合的DNA片段的扩增。该实施方案促进扩增包括一个或多个TFBS序列和/或一个或多个侧翼序列的选定的DNA片段。该实施方案也是快速、低成本、易于自动化以实现高通量,可以在任何PCR实验室中进行,并且通过将转录因子表达的联合组织特异性与通过分析染色质片段中缔合DNA的TFBS序列和/或侧翼序列鉴定其在基因组中的结合位置的特异性组合,另外进一步增加健康或患病的cfDNA起源组织特异性。因此,在本发明的一个实施方案中,提供了检测人或动物受试者中包含转录因子和DNA片段的无细胞染色质片段的方法,其包括以下步骤:

[0094] (i)使从人或动物受试者获得的体液样品与结合转录因子的结合剂接触;

[0095] (ii)分离与转录因子缔合的DNA;

[0096] (iii)使用序列特异性PCR引物寡核苷酸扩增DNA;和

[0097] (iv)使用DNA片段的的存在或量作为样品中包含转录因子的无细胞染色质片段的量的量度。

[0098] 在一个实施方案中,该方法包括提取与转录因子缔合的DNA片段。在进一步的实施方案中,该方法包括扩增提取的DNA片段。因此,根据本发明的其它方面,提供了检测在从人

或动物受试者获得的体液样品中包含转录因子和DNA片段的无细胞染色质片段的方法,其包括以下步骤:

- [0099] (i)使样品与结合转录因子的结合剂接触;
- [0100] (ii)分离结合的转录因子;
- [0101] (iii)提取与转录因子缔合的DNA;
- [0102] (iv)扩增提取的DNA;
- [0103] (v)检测扩增的提取的DNA;和
- [0104] (vi)使用DNA的存在或量作为样品中包含转录因子的无细胞染色质片段的量的量度。

[0105] 在优选的实施方案中,通过PCR进行转录因子缔合的DNA的扩增。本领域已知有许多PCR方法,包括但不限于定量PCR、实时PCR、逆转录酶PCR、巢式PCR、数字PCR、多重PCR、任意引物PCR、冷PCR(在较低变性温度下的共扩增-PCR)。在一些实施方案中,扩增方法包括DNA定量。

[0106] 根据本发明的其它方面,提供了检测人或动物受试者中的疾病的方法,其包括以下步骤:

- [0107] (i)使从人或动物受试者获得的体液样品与结合转录因子的结合剂接触;
- [0108] (ii)检测或测量与转录因子缔合的DNA;和
- [0109] (iii)使用DNA的存在或量作为受试者中疾病存在的指示物。

[0110] 在本发明的另一方面,提供了检测人或动物受试者中的疾病的方法,其包括以下步骤:

- [0111] (i)使从人或动物受试者获得的体液样品与结合转录因子的结合剂接触;
- [0112] (ii)分离与转录因子缔合的DNA;
- [0113] (ii)使DNA与DNA结合剂接触;
- [0114] (iv)检测DNA结合剂;和
- [0115] (v)使用DNA结合剂的存在或量作为受试者中疾病存在和/或性质的指示物。

[0116] 任何DNA结合剂都可以适用于本发明,包括抗体。DNA结合剂可以直接或间接(例如,通过接头系统,例如生物素/抗生物素蛋白或谷胱甘肽)用可检测部分(例如荧光、酶或放射性部分)标记。

[0117] 在本发明的另一方面,提供了通过检测包含转录因子和缔合DNA片段的无细胞染色质片段来测定特定转录因子占据的基因组TFBS位置(并因此也测定哪些基因被调节)的方法,其中对与转录因子缔合的DNA片段测序以测定转录因子结合的基因组位置。因此,在本发明的另一方面,提供了确定转录因子结合的基因组位置的方法,其包括以下步骤:

- [0118] (i)使样品与结合转录因子的结合剂接触;
- [0119] (ii)分离结合的转录因子;
- [0120] (iii)提取与转录因子缔合的DNA;
- [0121] (iv)扩增提取的DNA;
- [0122] (v)对扩增的提取的DNA测序;和
- [0123] (vi)使用提取的DNA的序列确定TFBS的基因组位置。

[0124] 本发明特别用于分析被转录因子结合的小DNA片段,通常在35-80bp的大小范围。

因此,在一个实施方案中,测序的提取的DNA涉及小DNA片段,例如包含小于约100bp,例如小于约80bp,特别是小于约60bp的DNA片段。注意到这些DNA片段大小与没有接头连接/在接头连接之前的DNA片段有关。在一个实施方案中,测序的提取的DNA包含低于100bp,例如35-80bp(没有接头连接/在接头连接之前)的大小范围的DNA片段。在一个实施方案中,测序的提取的DNA含有多个DNA大小范围,然后将其比较,例如如在图10和11中所示。

[0125] 在优选的实施方案中,样品是体液样品。在进一步的实施方案中,体液样品是血液、血清或血浆样品。

[0126] 在优选的实施方案中,所用的结合剂是定向结合特定转录因子的抗体。因此,在一个实施方案中,结合转录因子的结合剂是抗体或其片段(即结合片段)。

[0127] 在优选的实施方案中,将抗体固定在固相上以促进抗体结合的转录因子-DNA复合物或染色质片段的分离。

[0128] 在循环染色质片段中,转录因子以及已知与转录因子体内一致的序列的缔合DNA片段两者的存在进一步证实转录因子和DNA片段两者的同一性。这种转录因子以及缔合DNA片段的序列的组合是用于诊断或评估多种疾病病况的有效生物标志物组合。此外,存在于健康受试者中的许多转录因子与不同组织中的不同TFBS相结合,因此通过存在的缔合DNA鉴定被转录因子结合的TFBS位置,鉴定染色质片段的起源组织。此外,这同样适用于疾病病况。因此,可以从与通常表达的转录因子结合的TFBS组中鉴定疾病病况的存在(即使转录因子本身在许多或所有组织中表达)。例如,通常表达的转录因子CTCF结合永生化癌细胞中的超过一千个特异性基因组位置,但在其它非癌细胞中不结合(Wang等人,2012,Liu等人,2017)。因此,鉴定循环CTCF-DNA复合物的存在(其中对缔合DNA片段测序并观察到具有与CTCF的癌症特异性TFBS位置之一一致的序列)指示在从中获得样品的受试者中存在癌症疾病。因此,在本发明的高度优选的实施方案中,提供了通过检测包含转录因子和DNA片段的无细胞染色质片段用于检测受试者中疾病状态的方法,它们一起形成组合生物标志物,该生物标志物鉴定从人或动物受试者获得的体液样品中与疾病病况或特定组织一致的基因组中转录因子占据的TFBS位置,其包括以下步骤:

[0129] (i)使样品与结合转录因子的结合剂接触;

[0130] (ii)分离结合的转录因子;

[0131] (iii)提取与转录因子缔合的DNA;

[0132] (iv)扩增提取的DNA;

[0133] (v)对扩增的提取的DNA测序;和

[0134] (vi)使用缔合DNA片段的序列作为染色质片段的起源组织或受试者中的疾病状态的指示物。

[0135] 受试者的疾病状态的确定可以包括例如疾病的检测、诊断、治疗选择、监测或预后或用于疾病。

[0136] 在一个实施方案中,该方法包括使用转录因子和缔合DNA的序列作为指示受试者中疾病存在的组合生物标志物。术语“生物标志物”是指过程、事件或病况的独特的生物学或生物学衍生的指示物。生物标志物物可以用于诊断方法,例如临床筛选和预后评估,以及用于监测疗法结果,鉴定最可能对特定治疗性治疗应答的受试者,药物筛选和开发。这样的生物标志物包括例如与转录因子缔合的DNA的存在(例如,序列)、水平、浓度或量。本文提及的

“组合生物标志物”是指涉及多于一种生物学或生物学衍生的指示物(例如转录因子和缔合DNA)的生物标志物,特别是与DNA的一个或多个特定序列缔合的转录因子的水平、浓度或量。

[0137] 组织特异性是重要的,因为大多数转录因子不具有完美的(单细胞类型)表达特异性。用于含有转录因子的循环染色质片段的免疫测定的组织特异性受到所用抗体的分析特异性和所用转录因子或所用转录因子组的组织特异性的限制。因此,通过将特定的转录因子部分和与其结合的cfDNA片段的序列组合,可以改进组织特异性。

[0138] 其原因是转录因子与不同细胞中基因组中的不同DNA序列结合。基因表达通过转录因子与短TFBSDNA序列(也称为应答元件或结合基序)的特异性结合来调节。TFBS通常但不是必须位于调节的基因的转录起始位点附近的基因启动子区。转录因子通过DNA结合结构域(DBD)以序列特异性方式结合TFBS。通常,TFBS序列在其靶基因的启动子内是5-15bp长,并且转录因子蛋白通常可以以不同程度的结合亲和力结合一组类似的DNA序列。与含有转录因子的循环染色质片段缔合的DNA片段的长度将根据该片段是否还包括被其它转录因子、辅因子、核小体或其它染色质蛋白质结合的其它DNA保护的序列而变化。据报道,许多这样的染色质片段在35-80bp范围出现(Snyder等人,2016)。此外,我们注意到这与通过核酸酶消化从癌症患者的细胞提取的染色质而生产的染色质片段的大小范围一致,并且这个小的约35-80bp片段范围包含比核小体结合的片段更大比例的总染色质片段(Corcess等人,2018)。我们推断这些缔合的DNA片段比典型的DNA应答元件更长,并因此包括侧翼DNA序列。然而,与核小体缔合的DNA片段大小通常超过100bp DNA。因此,我们推断35-80bp DNA片段范围不包括完整的核小体DNA片段。

[0139] 转录因子的应答元件或TFBS序列可以在基因组内的许多位置重复出现,并且对于一些转录因子出现在数千个位置。因此,相同的转录因子有可能结合在细胞的染色质内的许多位置。这意味着单一细胞的死亡原则上可以产生大量含有相同转录因子的循环染色质片段。

[0140] 此外,转录因子倾向于不单独起作用,而是与其它转录因子或辅因子或调节特定基因所需的其它部分协同。因此,转录因子可以结合大量不同基因的启动子中的应答元件,每个基因与不同的转录因子协同。因此,围绕相同TFBS序列或相同转录因子的应答元件的DNA侧翼序列在不同基因的启动子中是不同的,因为它包括转录因子的不同组合的结合基序。这适用于所有或大多数转录因子。

[0141] 此外,应答元件本身的结合序列可以是简并的,使得转录因子可以结合多种不同的基序序列。例如,转录因子TTF-1在健康肺和健康甲状腺组织中以组织特异性方式表达。在肺中,两个蛋白TTF-1因子结合肺特异性表面活性物质蛋白B(SPB)基因的启动子区。SPB的启动子中TTF-1的DNA结合序列或结合基序是GCNCTNAG(SEQ ID NO:1)(其中A、C、G和T分别表示DNA碱基腺嘌呤、胞嘧啶、鸟嘌呤和胸腺嘧啶,并且N表示这些碱基中的任何一个)。TTF-1结合周围的较宽共有启动子DNA序列是(-118)GATCAAGCACCTGGAGGGCTCTCAGAGCAAA GACAAACACTGAGGTCGCTGCCA(-64)(SEQ ID NO:2),其中(-64)表示从SPB转录起始位点的距离,以bp计。在肺组织中的SPB启动子中,TTF-1与转录因子肝细胞核因子3(HNF3)协同结合,如在图1中所示(Matys等人,2006和Bohinski等人,1994)。

[0142] 在甲状腺中,TTF-1调节许多基因,包括甲状腺球蛋白、促甲状腺激素受体和甲状

腺过氧化物酶。甲状腺球蛋白基因的启动子区中TTF-1的共有结合序列不同于肺中的共有结合序列,并报道为TGGCCACACGAGTGGCCCTCA (SEQ ID NO:3)。在甲状腺球蛋白基因的启动子中,TTF-1与TTF-2、PAX8和Runx2转录因子协同结合,并且在5'和3'端处包括50bp侧翼序列的较宽序列是CCCACCCGTTCTGTTCCCCACAGTTTAGACAAGATCCTCATGCTCCACTGGCCACACGAGTGGCCCTCAGGAGTAGACACAGGTGGAGGGAGCTCCTTTTGACCAGCAGAGAAAAC (SEQ ID NO:4)。类似地,TTF-1也与促甲状腺激素受体和甲状腺过氧化物酶基因的启动子区结合,在每种情况下与不同的协同转录因子协同作用。因此,如在图1中所示,不仅甲状腺或肺组织中调节的基因的启动子序列中TTF-1结合位点周围的DNA序列不同,而且与TTF-1缔合的辅因子,以及因此周围的DNA序列,对于与相同组织中不同基因的结合也不同(Matys等人,2006和Maenhaut等人,2015)。这证明,含有TTF-1的循环染色质片段以及与该染色质片段缔合的DNA序列的知识的组合足以鉴定该染色质片段的起源为肺或甲状腺。

[0143] 据认为有约1000-3000种人转录因子(每种因子与基因组中的特定位置结合)导致驱动大量细胞过程的动态转录变化。我们已经关于作为一个实例的TTF-1说明了本发明的原理。然而,原则上任何转录因子都可以用于本发明的方法中。甚至,在许多细胞类型中普遍表达并结合离散DNA序列的转录因子(例如Hox蛋白转录因子)与辅因子协同结合以独特地结合不同序列来调节不同组织中的不同基因(Merabet和Mann,2016,Mann等人,2009)。这意味着所有或大多数转录因子以及它们的TFBS序列(任选地包括侧翼序列)可以用作本发明的方法的组合生物标志物。例如,雌性激素受体- α (ER α) 转录因子结合人基因组中超过一千个结合位点或雌性激素应答元件(ERE),与不同基因组位置的至少60种其它转录因子组合协同(Lin等人,2007)。类似地,雄性激素受体(AR)结合与数千个基因相关的雄性激素应答元件(ARE),与数千个不同的序列基因座处的其它协同转录因子协同。因此,本发明的方法可以通过缔合DNA的序列鉴定含有ER α 或AR的染色质片段的起源组织,即使这些转录因子在多种组织中表达。

[0144] 此外,转录因子与DNA基因座的全基因组结合在癌症中被重编程,并且在癌细胞中表达的转录因子和它们结合的TFBS与在相同组织的健康细胞中结合的那些不同,因此鉴定在循环中含有转录因子的染色质片段与缔合DNA片段的序列数据的组合使得能够鉴定患有癌症的受试者以及鉴定癌症类型(例如前列腺癌或肺癌等)两者(Pomerantz等人,2015)。这是能够实现的,因为染色质在肿瘤发生期间被重塑,并且这种重塑涉及通过癌细胞中的重塑的转录因子结合模式上调肿瘤相关蛋白。因为如此,许多转录因子的表达在癌细胞中被上调。这是广泛的现象,但可以由几个非限制性的实例来举例说明。例如,在大多数癌症中,公知的癌症相关转录因子c-Myc和p53被上调。在前列腺癌中被AR结合的结合位点序列大大变化(Pomerantz等人2015)。类似地,癌细胞中与转移和对疗法的抗性相关的上皮间质转化(EMT)涉及转录因子Jun/Fos家族(包括Fos11、Fosb、Fos和Junb)的上调。还已经发现转录因子的ETS(E26转化特异性)家族以及Runx1、Tead和Nfkb转录因子在肿瘤细胞的开放染色质中高度富集。此外,据报道,p63、Klf、Grhl和Cepba在肿瘤细胞中上调,并且它们的结合位点富集在开放染色质区。Klf5和p63转录因子与癌相关,并且在肺癌和头颈癌中作为驱动因子起作用。与EMT相关的其它转录因子包括bHLH、Runx、Nfat、Tbx1、Tcf711和Smad2(Latil等人,2017)。

[0145] 真核基因转录的调节涉及与多个调节DNA序列结合的多个调节蛋白,所述调节DNA

序列在转录复合物中位于基因的转录起始位点(TSS)附近并远离基因组中的TSS,例如如在图2中说明的。DNA中的远端调节序列可以位于距离TSS几百至超过百万个碱基或可能更远。转录复合物通常涉及DNA环,其可以涉及DNA弯曲蛋白,其中更远端的调节序列以及与其结合的调节蛋白与结合到更靠近TSS的调节序列的蛋白接触,例如也如在图2中说明的。TATA盒如此命名是因为其含有重复的胸腺嘧啶/腺嘌呤核苷酸的序列,其结合转录所需的一般转录因子。其它基因特异性转录因子也是特定基因表达所需的(例如,表达表面活性物质蛋白B、甲状腺球蛋白、甲状腺过氧化物酶和TSH受体基因所需的转录因子,如在图1中所示)。此外,多种其它蛋白质是必需的,例如包括但不限于辅因子、介体、激活物、辅激活物、阻抑物、辅阻抑物、染色质重塑蛋白、DNA弯曲蛋白、绝缘子、RNA聚合酶部分、延伸因子、染色质重塑因子、STAT部分或细胞因子或与STAT部分结合的细胞因子相关因子、上游结合因子(UBF)或任何其它与这样的基因调节或转录复合物相关的部分。这样的复合物也可以包括一定长度的核小体保护的DNA。转录复合物可以是稳定的以促进高容量转录。因此,健康和/或疾病起源的循环染色质片段可以包括大的蛋白质/DNA复合物,其包含可能对核酸酶活性有抗性的多种蛋白质。如在图2中说明的,涉及近端和远端调节序列的一些大转录复合物称为超增强子。超增强子是具有高水平的转录因子结合的大簇,并且是驱动参与控制细胞同一性的基因表达的核心。超增强子也是刺激癌症中癌基因转录的核心。癌细胞获得超增强子,并且癌性表型依赖于超增强子驱动的异常转录。因此,通过本文所述的方法检测包括对应于超增强子的近端和远端调节序列的cfDNA片段序列的全部或部分的超增强子复合物和/或组合的染色质片段的存在提供了鉴定包括起源癌细胞的染色质片段的细胞起源的方法。我们还通过其性质推论,超增强子复合物可能包含稳定结合而非瞬时结合的转录因子。

[0146] 这样的衍生自转录复合物的染色质片段中的DNA环原则上可以是完整的,或者可以在一个或多个位置被消化,导致(i)对应于近端和远端调节序列的两个循环染色质片段;或(ii)含有两个DNA片段的大染色质片段。因此,cfDNA可以包括对应于基因的近端和远端调节序列两者的小DNA片段。

[0147] 根据本发明的其它方面,提供了检测人或动物受试者中的疾病的方法,其包括以下步骤:

[0148] (i)使从人或动物受试者获得的体液样品与结合转录因子的结合剂接触;

[0149] (ii)确定与转录因子缔合的一个或多个DNA片段的序列;和

[0150] (iii)使用转录因子的存在和缔合DNA的序列作为用于确定受试者中疾病存在和/或性质的组合生物标志物。

[0151] 应当理解,在健康和患病受试者中结合DNA并且其cfDNA结合模式不同的任何非组蛋白染色质蛋白质都适用于本发明的方法,包括转录因子以及其它非组蛋白染色质蛋白质,包括染色质修饰蛋白质、遗传和表观遗传的阅读、书写和删除蛋白质、参与RNA转录的蛋白质(例如RNA聚合酶分子)和构造或结构染色质蛋白质(例如DNA弯曲蛋白质)。

[0152] 因此,根据本发明的其它方面,提供了检测人或动物受试者中的疾病的方法,其包括以下步骤:

[0153] (i)使从人或动物受试者获得的体液样品与结合非组蛋白染色质蛋白质的结合剂接触;

[0154] (ii)确定与非组蛋白染色质蛋白质缔合的一个或多个DNA片段的序列;和

[0155] (iii) 使用非组蛋白染色质蛋白质的存在和缔合DNA的序列作为用于确定受试者中疾病存在和/或性质的组合生物标志物。

[0156] 在优选的实施方案中,非组蛋白染色质蛋白质是RNA聚合酶,特别是RNA聚合酶II。RNA聚合酶II是负责转录基因的DNA序列以产生RNA拷贝的DNA结合酶。RNA拷贝可以通过核糖体导致相应蛋白质产生的信使RNA(mRNA)分子,或者可以是不翻译成蛋白质的非编码RNA(ncRNA)分子。因此,其循环染色质片段中RNA聚合酶II的存在指示该片段衍生自在其起源细胞中有活性的基因。因此,衍生自与RNA聚合酶II缔合的染色质片段的DNA片段序列文库提供了存在于样品中的活性动态基因的文库。在健康的人中,该文库主要对应于存在于造血组织中的活性基因。在患病的人中,文库另外包括在受疾病影响的一个或多个组织中有活性的基因。这可以是受疾病影响的任何组织。例如,在肝或肾细胞中有活性的基因可以表现在由取自患有肝或肾病患者的样品产生的RNA聚合酶II文库中,其中这样的基因不存在于健康人的文库中。类似地,在癌症中上调的基因可以表现在由取自患有癌症疾病的患者的样品产生的RNA聚合酶II文库中,其中这样的基因不存在于健康人的文库中。在本发明的该方面,RNA聚合酶II的使用允许鉴定样品中代表的活性动态基因。这允许检测癌症疾病以及确定受癌症影响的一个或多个组织。

[0157] 因此,根据本发明的其它方面,提供了检测人或动物受试者中的疾病的方法,其包括以下步骤:

[0158] (i) 使从人或动物受试者获得的体液样品与结合RNA聚合酶的结合剂接触;

[0159] (ii) 确定与RNA聚合酶缔合的一个或多个DNA片段的序列;和

[0160] (iii) 使用RNA聚合酶缔合DNA片段的序列作为用于确定受试者中疾病存在和/或性质的生物标志物。

[0161] 在一个实施方案中,疾病选自癌症、自身免疫疾病或炎性疾病。在进一步的实施方案中,疾病是癌症。在进一步的实施方案中,自身免疫疾病选自:系统性红斑狼疮(SLE)和类风湿性关节炎。在进一步的实施方案中,炎性疾病选自:克罗恩氏病、结肠炎、子宫内膜异位和慢性阻塞性肺病(COPD)。

[0162] 在优选的实施方案中,疾病是癌症。在进一步的实施方案中,癌症选自:乳腺癌、膀胱癌、结肠直肠癌、皮肤癌(例如黑素瘤)、卵巢癌、前列腺癌、肺癌、胰腺癌、肠癌、肝癌、子宫内膜癌、淋巴瘤、口腔癌、头颈癌、白血病和骨肉瘤。

[0163] 在进一步的实施方案中,疾病是胎儿疾病或疾况。本领域公知,胎儿起源的染色质片段(例如含有源自(XY)雄性胎儿的Y染色体DNA序列)在妊娠的动物和人(XX)母亲的血液中循环。据报道在妊娠受试者中循环的cfDNA包含预期长度的核小体保护的DNA片段(约160bp)的cfDNA片段以及50bp以上范围的较短cfDNA片段两者。此外,已报道长度小于140bp的母体cfDNA片段富集胎儿起源的cfDNA(Hu等人;2019)。因此,本发明的方法不仅适用于从中采集样品的受试者的疾病状态,而且适用于母体血液样品中胎儿病况的产前研究或测试。

[0164] 因此,根据本发明的其它方面,提供了检测人或动物胎儿中的疾病的方法,其包括以下步骤:

[0165] (i) 从妊娠的人或动物受试者获得体液样品;

[0166] (ii) 使体液样品与结合转录因子的结合剂接触;

[0167] (iii)检测、测量或测序与转录因子缔合的DNA;和

[0168] (iv)使用DNA的存在、序列或量作为胎儿中疾病存在的指示物。

[0169] 根据本发明的其它方面,提供了检测人或动物受试者中受疾病影响的组织的方法,其包括以下步骤:

[0170] (i)使从人或动物受试者获得的体液样品与结合转录因子的结合剂接触;

[0171] (ii)确定与转录因子或染色质片段缔合的DNA的DNA碱基序列;和

[0172] (iii)使用组合的转录因子/DNA序列生物标志物作为受试者中受疾病影响的组织的指示物。

[0173] 在优选的实施方案中,疾病是癌症。在另一个实施方案中,受疾病影响的组织是起源器官,例如癌症的起源器官。

[0174] 在本发明的另一方面,提供了检测人或动物受试者中的疾病的方法,其包括以下步骤:

[0175] (i)使从人或动物受试者获得的体液样品与结合转录因子的结合剂接触;

[0176] (ii)分离与转录因子缔合的DNA;

[0177] (iii)通过PCR方法扩增分离的DNA;

[0178] (iv)确定扩增的DNA的序列;和

[0179] (v)使用转录因子的存在和缔合DNA的序列作为用于确定受试者中疾病存在和/或性质的组合生物标志物。

[0180] 本领域技术人员还将清楚,可以获得对应于被特定转录因子结合的各种基因基因座的多个序列,并且可以整合关于各种序列的数据以确定疾病和/或受疾病影响的组织的性质。

[0181] 在本发明的另一方面,提供了检测人或动物受试者中的疾病的方法,其包括以下步骤:

[0182] (i)使从人或动物受试者获得的体液样品与结合转录因子的结合剂接触;

[0183] (ii)分离与转录因子缔合的DNA;

[0184] (iii)通过PCR方法扩增分离的DNA,例如使用序列特异性引物;

[0185] (iv)检测扩增的DNA;和

[0186] (v)使用扩增的DNA的存在、量和/或序列作为受试者中疾病存在和/或性质的指示物。

[0187] 在一个实施方案中,在接头寡核苷酸与DNA片段连接后进行分离的转录因子结合的DNA片段的扩增。因此,在本发明的一个实施方案中,提供了检测人或动物受试者中的疾病的方法,其包括以下步骤:

[0188] (i)使从人或动物受试者获得的体液样品与结合转录因子的结合剂接触;

[0189] (ii)分离与转录因子缔合的DNA;

[0190] (iii)将接头寡核苷酸与分离的DNA连接;

[0191] (iv)扩增DNA;和

[0192] (v)使用DNA片段的存在、量和/或序列作为受试者中疾病存在和/或性质的指示物。

[0193] 在本发明的另一方面,提供了检测人或动物受试者中的疾病的方法,其包括以下

步骤:

[0194] (i)使从人或动物受试者获得的体液样品与结合转录因子的结合剂接触;

[0195] (ii)分离与转录因子缔合的DNA;

[0196] (iii)使用序列特异性引物寡核苷酸扩增分离的DNA;

[0197] (iv)检测扩增的DNA;和

[0198] (v)使用扩增的DNA的存在、量和/或序列作为受试者中疾病存在和/或性质的指示物。

[0199] 该方面利用本发明的组合的转录因子/DNA序列生物标志物的组织特异性,同时避免DNA片段接头文库的制备和用于生物标志物目的的通过PCR扩增包括一个或多个TFBS序列和/或一个或多个目的侧翼序列的选定DNA片段的下一代DNA测序。该方法快速、低成本、易于自动化以实现高通量,并且可以在任何PCR实验室中进行。

[0200] 在步骤(i)或(ii)中分离的DNA序列可以通过本领域已知的任何方法扩增。在一些实施方案中,使用PCR方法扩增分离的DNA,所述PCR方法采用与DNA片段连接的接头。在其它实施方案中,PCR引物用于DNA扩增。引物可以设计成扩增在步骤(i)或(ii)中分离的所有DNA序列,或者可以设计成扩增与转录因子的应答元件的序列缔合的特定DNA序列,任选地还包括侧翼区。

[0201] 在本发明的另一方面,提供了检测人或动物受试者中的疾病的方法,其包括以下步骤:

[0202] (i)使从人或动物受试者获得的体液样品与结合转录因子的结合剂接触;

[0203] (ii)分离(并任选地扩增)与转录因子缔合的DNA;

[0204] (iii)通过杂交方法检测DNA;和

[0205] (iv)使用杂交的DNA的存在、量和/或序列作为受试者中疾病存在和/或性质的指示物。

[0206] 该方面利用本发明的组合的转录因子/DNA序列生物标志物的组织特异性,同时避免通过对包括一个或多个TFBS序列和/或一个或多个侧翼序列的DNA片段的的选择性DNA杂交的昂贵的下一代DNA测序。该方法成本低,并且可以在任何PCR实验室中进行。

[0207] 在优选的实施方案中,在杂交前扩增分离的DNA。在优选的实施方案中,杂交方法是DNA微阵列方法(也称为DNA芯片方法)。

[0208] 本发明的方法也可以用于测量转录因子和序列缔合DNA的组合生物标志物。

[0209] 转录因子的选择

[0210] 真核生物体中的基因转录的调节是高度复杂的,并且可以涉及DNA的弯曲和成环,使在调节转录复合物中通过多个调节蛋白结合的多个调节DNA序列在一起,如在图2中说明的。因此,如本文所用的术语“转录因子”是指直接或间接结合基因组中的基因调节序列以调节基因转录的调节蛋白,包括但不限于与一个或多个特定基因的调节相关的一般转录因子和特异性转录因子,以及与STAT部分、UBF或与这样的基因调节或转录复合物相关的任何其它部分结合的增强子、辅增强子、阻抑物、辅阻抑物、介体、激活物、辅激活物、阻抑物、辅阻抑物、染色质重塑蛋白、DNA弯曲蛋白、绝缘子、RNA聚合酶部分、延伸因子、STAT部分、细胞因子或细胞因子相关因子。类似地,如本文所用的术语“转录因子结合位点”(TFBS)是指与基因的转录调节相关的调节蛋白的DNA结合位点,包括但不限于远端或近端增强子和阻抑

物序列,如在图2中所示。

[0211] 众所周知,转录因子表达在疾病中改变。因此,本发明的方法可以涉及转录因子,其表达在疾病中上调和/或在疾病组织(例如癌组织)中不适当地表达,其通常在所述(健康)组织中不高度表达。因此,体液样品中存在的转录因子的水平可以用作疾病的生物标志物。

[0212] 还众所周知,转录因子的TFBS占据的特性在不同的细胞类型和疾病中改变(Wang等人,2012)。因此,体液样品中存在的转录因子的TFBS占据的特性可以用作疾病的生物标志物。

[0213] 健康受试者的循环中存在的染色质片段主要是造血起源的。因此,本发明的方法还涉及检测包含转录因子以及缔合DNA的染色质片段的不适当存在,其通常不在造血组织中表达(但可以在非造血组织中表达)。

[0214] 例如,许多癌症疾病衍生自上皮组织。上皮GRHL2转录因子在许多上皮组织以及许多上皮组织衍生的癌症疾病中表达,但不在造血组织中表达。GRHL2在循环中的存在指示上皮衍生的癌症的存在,例如结肠直肠癌、前列腺癌、肺癌或乳腺癌。因此,本发明的方法可以用于检测癌症本身的存在,以及使用谱系特异性转录因子和/或转录因子与缔合DNA序列的谱系特异性组合来鉴定癌症的起源器官。因此,任何转录因子都可以用于本发明的方法中。在优选的实施方案中,包括所选转录因子的染色质片段的水平在患病受试者的体液中升高(超过在其它受试者中发现的水平),部分或完全是组织和/或疾病特异性的,和/或在基因组中具有多个应答元件。

[0215] 因此,在一个实施方案中,转录因子是疾病特异性的(即,在疾病中上调包括转录因子的循环染色质片的水平)。在一个实施方案中,转录因子是组织特异性的。在一个实施方案中,转录因子在基因组中的多于一个位置结合,例如在基因组中的超过5个、超过10个、超过100个、超过1000个或超过10,000个位置结合。一些转录因子结合位置在一些组织类型中被占据,但在其它组织类型中未被占据。一些转录因子结合位置在患病细胞中被占据,但在相同组织的健康细胞中未被占据。

[0216] 转录因子可以通过结合结构域分类(例如参见Vaquerizas等人,2009,其通过引用并入本文)。在一个实施方案中,转录因子包含选自以下的DNA结合结构域:同源结构域、HLH、bZip、NHR、叉头、P53、HMG、ETS、aIPT/TIG、POU、MAD、SAND、IRF、TDP、DM、热休克、STAT、CP2、RFX、AP2或锌指(例如锌指C₂H₂或锌指GATA)结合结构域。在一个实施方案中,转录因子包含非锌指DNA结合结构域。

[0217] 合适的转录因子可以通过实验确定,例如使用经典的核酸酶可接近位点作图方法来鉴定一个或多个目的组织中的目的转录因子。在典型的实验中,从目的细胞(例如癌细胞、相同组织的健康细胞和造血细胞)提取染色质,并使用合适的核酸酶消化。将通过消化产生的染色质片段暴露于与转录因子结合的抗体,并分离抗体结合的DNA片段并测序以鉴定被转录因子结合的一个或多个TFBS序列(任选地包括侧翼序列)。结果可以用于选择本发明中使用的转录因子。例如,在患病细胞中升高但在造血细胞中低或不存在的转录因子和转录因子/TFBS(任选地包括侧翼序列)组合可用于本发明的方法。经典的核酸酶可接近性方法最近已经得到改进,并且现在在本领域包括方法,例如CUT&RUN和其它方法,其更容易进行并且提供改进的结果(Skene和Henikoff,2017)。任何这样的方法都将适用于鉴定用于本

发明的合适的转录因子。

[0218] 已经进行了许多这样的实验和类似的实验,并因此在本领域中可以获得合适的转录因子。文献中有许多关于转录因子和癌症的出版物列出可用于本发明的方法的转录因子。例如Lambert等人,2018列出了294种已知的致癌转录因子和调节剂。Gurel等人,2010描述了转录因子NKX3.1作为前列腺癌的标志物。Darnell,2002列出了许多致癌转录因子,包括TAT3、5、STAT-STAT、GR、IRF、TCF/LEF、 β -联蛋白、NF-KB、NOTCH(NICD)、GLI、c-JUN、bZip蛋白(包括c-JUN、JUNB、JUND、c-FOS、FRA、ATF和CREB-CREM家族)、cEBP家族、ETS蛋白和MAD-box家族。Vaquerizas等人,2009描述了许多可用于本发明的方法的组织特异性转录因子。Ulz等人,2019描述了转录因子,例如上皮转录因子GRHL2,其存在于许多癌症类型中但不存在于血液学组织中,以及AR(雄性激素受体)、NKX3-1和HOXB13。Corces等人,2018描述了许多癌症特异性和组织特异性转录因子,包括NR5A1、TP63、GRHL1、FOXA1、GATA3、NFIC、CDX2、RFX2、ASCL1、PAX2、HNF1A、NKX2.A、PHOX2B、DRGX、HOXB13、AR、MITF、HNF4和POU5F1。使用ChIP-Seq,Wang等人,2012在19种不同的细胞类型中鉴定了77,811个不同的转录因子CTCF的结合位点,包括7种永生癌细胞系和12种正常细胞类型。在这77,811个CTCF TFBS中,发现1236个位点在癌细胞中被差异占据。发现195个位点的占据发生在正常细胞类型中,但不发生在癌细胞中。发现1041个位点的占据发生在癌细胞中,但不发生在正常细胞类型中(Liu等人,2017)。通过ChIP-Seq,对应于体液中癌症特异性TFBS的CTCF相关的cfDNA片段的发现指示所研究受试者中癌症疾病存在,并且可以以这种方式用作生物标志物。所述参考文献通过引用并入本文。

[0219] 用于本发明的方法的合适的转录因子还可以使用各种转录因子、癌症和基因组数据库来选择,例如ENSEMBL数据库,其提供许多物种(包括人)的注释基因组序列、DNA元件百科全书或(ENCODE)数据库(<https://www.encodeproject.org>)、转录因子(TRANSFAC)数据库(Matys等人,2006)、基因转录调节数据库(GTRD)版本18.01(<http://gtrd.biouml.org>)、人转录因子数据库版本1.01(<http://humantfs.cabr.utoronto.ca>)、NIH基因组数据通讯数据库(<https://gdc.cancer.gov>)、癌症基因组图谱(TCGA)(<https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>)、UCSC Xena Browser(<https://atacseq.xenahubs.net>)和人蛋白图谱数据库(<https://www.proteinatlas.org>),其提供关于在其中表达转录因子的健康组织以及其在癌症疾病中的表达的数据以及其它数据库。

[0220] 这些数据库用于表征本发明的方法中所用转录因子和相关TFBS序列和侧翼序列的用途可以参考这些数据库中的一些作为实例进行说明。TRANSFAC数据库提供了关于数千人和其它真核转录因子的数据。对每个转录因子提供的细节包括其在基因组中结合的TFBS的数目、其调节转录的基因列表、与每个调节的基因相关的TFBS的序列和基因组位置、以协同方式与其操作以调节转录的其它转录因子的细节、共有TFBSDNA序列、DBD细节和癌症相关性。为了说明的目的,以下举例说明在本发明的上下文中该数据用于转录因子CDX2和c-JUN的用途。TRANSFAC数据库列出了48种调节26种特定基因的人CDX2 TFBS。提供了CDX2 TFBS序列以及它们的基因组位置和各自调节的基因。每个CDX2 TFBS的侧翼序列可以通过参考在每个基因组位置的序列的ENSEMBL人基因组数据库来确定。还提供了共有CDX2 TFBS序列。类似地,TRANSFAC数据库列出了265种调节166种特定基因的人c-JUN TFBS。提供了c-

JUN TFBS序列以及它们的基因组位置和各自调节的基因。每个c-JUN TFBS的侧翼序列可以通过参考在每个基因组位置的序列的ENSEMBL人基因组数据库来确定。还提供了共有c-JUN TFBS序列。

[0221] 因此,转录因子和/或TFBS可以通过实验选择或从文献和/或从数据库(例如人蛋白质图谱数据库)中选择,如在本发明的方法中有用的。转录因子可以关于(i)在其中表达的健康和患病组织,(ii)在这些细胞或组织中调节的基因,(iii)在这些组织中与其结合的TFBS序列(任选地包括侧翼序列)和(iv)通过与在TFBS上共结合而协作进行转录调节的其它因子来表征。通过本文所述的方法,该表征可以用于鉴定体液样品中染色质片段和/或转录因子缔合的cfDNA片段的起源的健康或患病的组织或细胞。

[0222] 类似地,与体液样品中染色质片段和/或cfDNA序列缔合的实验数据可以使用这些数据库来解释,以鉴定cfDNA片段中包含的TFBS序列的全部或部分,任选地包括侧翼序列。然后,该数据可以用于鉴定cfDNA片段的起源组织或细胞。

[0223] 目前有三组主要的转录因子被认为在癌症中特别重要。第一组是核激素受体组,其包括雌性激素受体、雄性激素受体、孕酮受体、糖皮质激素受体、甲状腺受体和视黄酸受体。转录因子的核激素受体组是细胞表面受体,其可以被认为是可以通过配体结合而激活的无活性或潜在转录因子。例如,雌性激素受体通过与雌性激素结合而被激活。配体结合导致核激素受体迁移到细胞核,在细胞核中它结合靶DNA序列(例如,雌性激素受体结合雌性激素应答元件),并上调或下调与DNA靶序列缔合的基因(例如,雌性激素调节的基因)。

[0224] 已知在癌症的起始和发展中重要的第二组转录因子是信号转导物和转录激活物(STAT)。这些是潜在的细胞质转录因子,其可以被细胞质中和/或细胞表面处的多种分子触发剂所激活。STAT激活通常涉及细胞质中的生物化学事件级联,例如激酶反应、蛋白水解反应和蛋白质-蛋白质相互作用,其导致进入调节靶基因转录的蛋白质或蛋白质复合物的细胞核。通常导致转录激活的生物化学级联反应由配体在细胞表面处的受体结合触发,包括例如细胞因子部分被细胞因子受体结合,或生长因子(例如表皮生长因子或血小板衍生的生长因子)被生长因子受体结合,或通过肽或蛋白与G蛋白偶联的受体的结合。

[0225] 在癌症中重要的第三组转录因子是固有的核蛋白,其转录作用通常由涉及丝氨酸激酶反应的生物化学事件级联激活。有数百个丝氨酸激酶部分和数百个核蛋白,所述核蛋白是丝氨酸激酶的靶。

[0226] 本领域技术人员将清楚,包含(即包括或含有)参与癌症的起始、发展或维持的任何转录因子(例如上述三组转录因子)的无细胞染色质片段将可用于本发明的方法。在癌症中具有已知作用或已知在癌症疾病中升高的一些转录因子或转录因子家族例如包括但不限于STAT(特别是STAT3、STAT5和STAT-STAT二聚体部分)、**NF- κ B**、 β -联蛋白、 γ -联蛋白、Notch和notch胞内结构域(NICD)、GLI、c-JUN、JUNB、JUND、c-FOS、FRA、ATF、CREB-CREM、cEBP、ETS、MYC、N-MYC、MAX、E2F、干扰素调节因子(IRF)、T-细胞因子(TCF)、淋巴细胞增强子因子(LEF)、EN2、GATA3、CDX2、PAX8、WT1、NKX3.1、P63(TP63)或P40和螺旋-环-螺旋蛋白(Darnell, 2002)。所有这样的转录因子都可用于本发明的方法。

[0227] 已经发现许多转录因子是谱系特异性的,并且与特定组织相关,并因此可以认为是组织特异性转录因子,即在某些组织或癌症中总是或通常表达而在其它组织或癌症中很少或从不表达的转录因子。本发明的方法可以与组织特异性转录因子一起使用,其中组合

检测缔合DNA提供了增强的特异性和/或灵敏度。

[0228] 甲状腺转录因子1 (TTF-1) 在胚胎发生期间在甲状腺、间脑和呼吸上皮中选择性表达。TTF-1在取自神经内分泌和非神经内分泌肺癌的组织样品中表达,但其表达频率在不同的组织学亚型中显著不同。因此,本发明的方法也可以用于通过测量含有转录因子及其缔合DNA序列的染色质片段来鉴定癌症类型和亚型。

[0229] PAX8是参与甲状腺、肾和缪勒氏管系统的胚胎发生的转录因子。PAX8在取自非粘蛋白性卵巢癌、浆液性、子宫内膜样、透明细胞和移行细胞癌的组织样品中显示高水平的表达。PAX8也在子宫内膜样腺癌、子宫浆液性癌、子宫内膜透明细胞癌以及导管和小叶乳腺癌组织中表达。

[0230] CDX2是谱系特异性转录因子,在控制肠上皮细胞的增殖和分化中具有关键作用,并且在几乎所有结肠直肠癌组织样品中表达。

[0231] NKX3.1是正常前列腺发育所必需的,并且是在几乎所有前列腺癌中表达的已知标志物。

[0232] GATA3早在人妊娠第四周就具有转录活性。GATA3在取自乳腺癌的组织样品(特别是雌性激素受体阳性乳腺癌组织样品)和尿道上皮癌和移行细胞癌中高度表达。

[0233] WT1在胚胎发育中起重要作用。WT1是卵巢癌组织的良好标志物,并且由非常有限范围的健康成人组织表达。

[0234] EN2在胚胎发育中起作用,并且在一系列癌症中表达,但在极少数的成人健康组织中表达。尿中EN2的存在已用作检测前列腺癌的尿测试的基础。

[0235] 其它转录因子可以用于本发明的方法中。例如,UBF是结合核糖体RNA基因启动子并激活由RNA聚合酶I介导的转录的转录因子。已知UBF表达在一些癌症的组织中升高。许多其它这样的实例无疑存在并且是用于本发明的方法的合适的转录因子。此外,RNA聚合酶I和RNA聚合酶III在癌症中也升高。这些部分负责tRNA和核糖体RNA基因的转录,以提供癌细胞和组织的升高和快速的蛋白质产生、生长和细胞复制特征所需的细胞机器。在本发明的其它实施方案中,提供了检测或测量包含UBF、RNA聚合酶I或RNA聚合酶III的无细胞染色质片段的方法。

[0236] 在备选的实施方案中,转录因子不是组织特异性转录因子。本发明的方法也能够检测通常表达的转录因子,即在超过5种、超过10种、超过15种、超过20种或超过30种组织类型中表达的转录因子。通过将检测与缔合DNA序列组合(即组合生物标志物),本发明的方法可以检测通常表达的转录因子以提供临床上有用的结果。核激素受体转录因子是实例。如上所讨论的,CTCF也是本文进一步研究的实例。

[0237] 转录因子以与许多其它因子高度协同的方式结合其DNA靶序列,所述其它因子包括其它转录因子、辅因子、辅激活物、辅阻抑物、RNA聚合酶部分、延伸因子、染色质重塑因子、介体、STAT部分、UBF等。这意味着通过本发明检测的循环转录因子可以包括作为较大基因调节复合物的一部分的其它部分,包括任何或所有具有缔合DNA的核小体、核激素受体、类固醇或其它与核激素受体结合的激素、其它转录因子、辅因子、辅激活物、辅阻抑物、RNA聚合酶部分、延伸因子、染色质重塑因子、介体、STAT部分或与STAT部分结合的细胞因子或细胞因子相关因子、上游结合因子(UBF)或与在无细胞染色质片段中出现的这样的基因调节或转录复合物相关的任何其它部分。

[0238] 含有转录因子部分的无细胞染色质片段可以包括或不包括复合物中完整核小体或任何组蛋白的存在。所有这样的无细胞染色质复合物都将在本发明中有益，并且包括在本发明中。

[0239] 在优选的实施方案中，转录因子选自：STAT、NF- κ B、 β -联蛋白、 γ -联蛋白、Notch、notch胞内结构域(NICD)、GLI、c-JUN、JUNB、JUND、c-FOS、FRA、ATF、CREB-CREM、cEBP、ETS、MYC、MAX、E2F、干扰素调节因子(IRF)、T-细胞因子(TCF)、淋巴细胞增强子因子(LEF)和螺旋-环-螺旋蛋白、HOX蛋白、EN2、GATA3、CDX2、TTF-1、PAX8、WT1、NKX3.1、P63(或TP63)、P40或CTCF。在进一步的实施方案中，转录因子选自：EN2、CDX2或TTF-1。在另一个实施方案中，转录因子是CTCF。

[0240] 这些转录因子中的大多数不是100%组织特异性的，而是可以在一些癌症以及一些成人组织类型中表达。通过使用分析灵敏的检测一个或多个缔合DNA片段的方法，增强血液中含有转录因子的染色质片段的检测。通过将转录因子的同一性和与其缔合的DNA的一个或多个特定序列组合，增强方法的疾病和/或组织特异性。

[0241] 在一个实施方案中，使取自受试者的体液样品与一种或多种转录因子结合剂接触，选择所述转录因子结合剂以在多重测定中测试一种或多种疾病病况。例如，任选地除了许多癌症中表达的转录因子之外，测试各自对一种或多种癌症疾病特异的多种转录因子，使得能够在单次血液测试中除了鉴定癌症组织之外还检测许多不同的癌症疾病。多重测试的方法是本领域公知的，例如但不限于Luminex Corporation的多重珠系统可以用于在单个样品中进行大量的多重测定(Dunbar, 2006)。

[0242] 根据本发明的其它方面，提供了检测人或动物受试者中的疾病的方法，其包括以下步骤：

[0243] (i) 使从人或动物受试者获得的体液样品与结合多个转录因子的多个结合剂接触；

[0244] (ii) 分析与不同的转录因子缔合的DNA；和

[0245] (iii) 使用与多个转录因子结合的DNA的存在和/或量和/或模式来确定受试者中疾病存在和/或性质。

[0246] 根据本发明的其它方面，提供了检测人或动物受试者中的疾病的方法，其包括以下步骤：

[0247] (i) 使从人或动物受试者获得的体液样品与结合两种或更多种(例如多种)转录因子的两种或更多种(例如多种)结合剂接触；

[0248] (ii) 确定与在步骤(i)中结合的转录因子缔合的DNA的序列；和

[0249] (iii) 使用与转录因子结合的DNA的存在和/或量和/或模式和/或一个或多个序列来确定受试者中疾病存在和/或性质。

[0250] 在一个实施方案中，多个转录因子中的每一个都附着于单独的固相支持物，使得每个转录因子都可以被分离用于分析或测序其缔合的DNA片段。例如，Luminex多重珠系统由多种珠类型组成，每种珠类型可以用不同的转录因子结合剂涂布，所述转录因子结合剂可以暴露于单一样品并随后彼此分离用于独立地对与每种转录因子缔合的DNA进行(单独)测序。

[0251] 转录因子-DNA染色质片段

[0252] 存在于循环中的染色质片段源自多种来源。一种来源是通过在细胞死亡后将染色质释放到循环中,所述细胞可以包括患病细胞,例如癌细胞。在一些情况下,可能存在染色质进入循环的主动释放。

[0253] 循环中染色质片段的主要来源通过称作NETosis的过程产生嗜中性粒细胞胞外陷阱(NET)而衍生自嗜中性粒细胞。在该过程中,嗜中性粒细胞将染色质材料(NET)喷射到胞外基质中以及在感染部位局部捕获和中和病原体。NET及其代谢物主要由寡核小体和单核小体组成,其具有大小 ≥ 150 bp的组分DNA片段。

[0254] 从血液中提取的cfDNA的大小分布揭示cfDNA的主要组分是大小分布峰为约160-170bp的单核小体,其范围为约130-200bp,对应于具有不同长度的缔合接头DNA的单核小体。可以存在对应于各种大小的寡核苷酸的另外的峰,包括例如二核小体(约340bp)、三核小体(510bp)等。在受NETosis影响的样品中,也可能存在与长度至多几千bp的大的染色质片段相关的宽峰。

[0255] 转录因子结合短DNA序列,并且转录因子-DNA复合物含有35-80bp范围的短得多的DNA片段(Snyder等人,2016)。在双链血浆cfDNA文库的典型大小特性图中,几乎没有或没有对应于长度 < 100 bp的cfDNA片段长度可见的材料。然而,单链文库制备物含有更多的35-80bp范围的cfDNA片段(Snyder等人,2016)。这种蛋白质结合的35-80bp cfDNA组分是总循环染色质片段的次要组分。

[0256] 在本发明的上下文中,转录因子-DNA结合的另一个重要方面涉及转录因子-DNA结合的动力学稳定性。一些转录因子在TFBS处与DNA体内稳定结合。其它转录因子在TFBS处在体内瞬时结合,它们以动态方式缔合、解离和缔合。在使用基于细胞和组织的底物的ChIP-Seq方法中,这不是问题,因为两者都可以使用交联技术检测。动态结合的转录因子在结合形式和游离形式之间自然交替,但是当交联时,它们变成以结合形式被“捕获”。因此,使用短的时间导致稳定结合的转录因子的高检测,但动态结合的转录因子的检测较少。相反,使用较长的交联时间导致动态结合的转录因子的检测增加,因为更多的转录因子随着时间的推移通过交联以相关形式被“捕获”(Poorey等人,2013)。

[0257] 然而,基于动力学考虑,我们推论出动态结合的转录因子不可能存在于血液或其它体液的循环中。体内存在相对高核浓度的染色质和转录因子两者,使得动态结合的转录因子-DNA复合物缔合、解离和再缔合。然而,体液中转录因子-DNA复合物的水平被高度稀释,并且以如此低的浓度存在,以致一旦解离,任何瞬时或动态结合的转录因子和DNA组分不可能再缔合。我们推论,血浆中的交联因此仅与稳定结合的转录因子缔合,并因此总是快速的(因为较慢的交联瞬时结合的转录因子将被解离,并且可以忽略)。因此,在本发明的一个实施方案中,提供了检测人或动物受试者中的疾病的方法,其包括以下步骤:

[0258] (i)使从人或动物受试者获得的体液样品与结合动力学稳定的转录因子-DNA复合物的结合剂接触;

[0259] (ii)确定在所述动力学稳定的转录因子-DNA复合物中与转录因子缔合的一个或多个DNA片段的序列;和

[0260] (iii)使用转录因子的存在和缔合DNA的序列作为用于确定受试者中疾病存在和/或性质的组合生物标志物。

[0261] 转录因子DNA结合结构域

[0262] 转录因子可以根据其DNA结合结构域(DBD)分类。Vaquerizas等人,2009研究了1391种已知的转录因子,并基于DBD鉴定了超过24种不同类型的转录因子。鉴定的最常见的转录因子是具有锌指DBD的那些,并且这些占有所有转录因子的几乎一半(48.5%)。

[0263] 用于分析cfDNA、ctDNA或核小体的优选的样品类型是EDTA血浆。血浆血液收集管中EDTA或柠檬酸盐的功能是螯合(chelate)和螯合(sequester)血液中的钙离子以防止凝结(血液中的凝结级联需要钙离子的存在)。管的离心将血液的细胞组分与血浆上清液分离,可以将血浆上清液去除并用作许多临床诊断目的样品基质。

[0264] 锌指转录因子与其DNA TFBS的结合依赖于锌离子的存在。然而,用于血浆血液收集管中的钙螯合剂也螯合锌离子。从锌指转录因子中螯合并去除锌离子可以导致丧失与DNA结合的转录因子(Ralston,2008)。锌螯合剂和锌指转录因子的相互作用意味着该家族的转录因子在EDTA血浆中的行为与通过其它DBD类型结合DNA的转录因子不同。

[0265] 还没有直接证明血液中锌指转录因子-DNA复合物的存在。我们推论,虽然存在这样的复合物,它们尚未被分离,因为它们是血液的小循环染色质片段组分的一小部分(绝大多数循环染色质片段是核小体),而且在本领域工作人员使用的血浆样品中解离。如本文所述,我们已经解决了这两个问题,并证明了是锌指转录因子的CTCF的血浆ChIP-Seq。

[0266] 转录因子结合剂

[0267] 优选的转录因子结合剂包括定向结合转录因子的抗体,或寡核苷酸,例如TFBS的DNA序列(任选地包括侧翼序列)。优选的结合剂对转录因子具有高亲和力,使得结合将在低转录因子浓度下发生,以及对转录因子的结合具有高特异性,使得其它蛋白质的非特异性结合最小。

[0268] 结合剂可以涂布在固体支持物(例如琼脂糖凝胶、葡聚糖凝胶、塑料或磁性珠)上。在一个实施方案中,所述固体支持物包含多孔材料。在另一个实施方案中,结合剂被衍生化以包括标记或接头,所述标记或接头可以用于将结合剂附着于合适的支持物,所述支持物已被衍生化以结合标记。许多这样的标记和支持物是本领域已知的(例如Sortag、点击化学、生物素/链霉亲和素、his-标记/镍或钴、GST-标记/GSH、抗体/表位标记和许多更多)。然后,可以在结合剂与转录因子反应之前、同时或之后进行结合剂的分离。为了易于使用,涂布的支持物可以包括在装置(例如微流体装置)内。可以以多重测定形式使用多种固相结合剂,以在单一测试中在单一体液样品中同时测试含有不同转录因子的多种染色质片段的的存在。

[0269] 在其它实施方案中,将结合剂加入到溶液中,并通过交联和用沉淀剂(例如聚乙二醇(PEG))沉淀结合的核小体来分离。然后,例如通过离心或过滤,可以将沉淀的沉淀作为单独的相分离出来。许多免疫沉淀方法是本领域已知的,并且任何这样的方法都可以用于本发明的方法。

[0270] 在一些实施方案中,与转录因子结合的DNA通过DNA结合剂结合。DNA结合剂可以附着于固相(例如塑料颗粒、磁性颗粒、琼脂糖等)。DNA结合剂可以直接或间接(例如通过接头系统,如生物素/抗生物素蛋白或谷胱甘肽)附着于固相。

[0271] 我们使用了商业上可获得的定向结合转录因子的抗体。对于ChIP-Seq,我们将抗体固定在商业上可获得的磁性聚苯乙烯颗粒上。因此,在本发明的优选的实施方案中,转录因子结合剂是固定在磁性聚苯乙烯颗粒上的固相抗转录因子抗体(或其一部分)。

[0272] DNA文库制备

[0273] 本发明的一些实施方案包括制备与染色质片段中的转录因子缔合的cfDNA片段的文库。可以扩增文库使得易于使用PCR方法检测和测序。原则上,任何文库制备方法都适用于本发明的方法。

[0274] DNA片段文库制备方法是本领域公知的,并且通常涉及将接头寡核苷酸连接到DNA片段。接头连接的DNA片段文库的扩增通常通过PCR进行。PCR引物也可以用于DNA扩增,并且可以是简并的以扩增文库中存在的所有序列,或者可以使用本领域已知的软件设计以扩增与转录因子的应答元件的序列缔合的特定DNA序列,任选地还包括侧翼区。

[0275] 文库制备方法可以涉及cfDNA片段的单链或双链接头连接。优选的文库制备方法涉及单链cfDNA接头连接。优选的文库制备方法对于长度小于100bp的小DNA片段的扩增和分离具有高效率。许多这样的文库制备方法是本领域已知的,包括例如(i) TruSeq DNA样品制备试剂盒(Illumina),根据制造商的方案使用20-25个PCR循环,5-10ng输入DNA(Ulz等人,2019), (II) 使用MagMAX cfDNA分离试剂盒(Applied Biosystems),然后使用NEBNextUltra IIDNA Library Prep Kit(New England Biolabs)进行文库制备(Ulz等人,2019),或(iii) 使用血液和体液方案,用于Qiagen QIAamp DSP DNABloodMini Kit,使用Life technologies Ion Plus FragmentLibraryKit进行PCR扩增(Hu等人,2019)。其它方法包括Sanchez等人,2018、Skene和Henikoff,2017、Snyder等人,2016和Liu等人,2019描述的那些。在本文提供的实施例中,我们使用了商业上可获得的单链DNA文库制备试剂盒(ClaretBio SRSly NGS Library Prep Kit)。

[0276] 本领域技术人员清楚,对于本发明的实施方案,其中(仅)进行转录因子缔合的DNA的PCR扩增以增加转录因子检测或定量的灵敏度,然后单独的应答元件序列而没有侧翼序列的扩增就足够了。

[0277] 转录因子-DNA复合物的免疫沉淀

[0278] 免疫沉淀原则上是简单的方法。在典型的方法中,将定向特异性结合目的蛋白质的抗体涂布在固体支持物上,并暴露于含有该蛋白质的生物样品。目的蛋白质被抗体结合,并因此被吸附到固相的表面,而其它蛋白质和其它物质保留在溶液中。从样品中分离出固相并洗涤,留下附着于固体支持物的目的蛋白质的纯样品。

[0279] 基于细胞和组织的ChIP-Seq方法在本领域中有充分描述。通常使用20-30 μ g从组织或培养的细胞提取的消化的或超声处理的染色质作为底物材料。由于染色质由约40%的DNA组成,这代表约8-24 μ g的底物DNA。然而,循环cfDNA的浓度低,并且已在健康人受试者中测量为30 \pm 14ng/ml,并且在胃癌患者中测量为71 \pm 55ng/ml(Park等人,2012)。因此,1ml血浆样品将产生是在ChIP-Seq中通常使用的染色质材料的约200-500分之一的染色质材料。

[0280] 由于循环无细胞染色质的大部分由核小体组成,可用的循环无细胞转录因子-DNA染色质片段物质是极小的。此外,可用的循环无细胞转录因子-DNA染色质片段材料将包含数千个转录因子。因此,由单个转录因子代表的用于通过本发明的方法分析的可用的底物材料将是存在于循环中的少量循环无细胞转录因子-DNA材料的一小部分。

[0281] 此外,来自细胞的染色质提取物是相对纯的染色质材料。相反,体液(例如血液、血清或血浆)含有少量染色质,但含有较高浓度的大量蛋白质和其它化合物,其中任何一种都可能通过非特异性地粘附于所用的固相转录因子抗体或其它结合剂而干扰本发明的方法。

来自血液、血清或血浆的循环转录因子DNA复合物的免疫沉淀的额外复杂性在于,背景非特异性结合因此相对于与固相支持物上的特异性结合剂结合的少量靶转录因子而言是高的,并且可能使其检测模糊。

[0282] 由于所有这些困难,在文献中很少报道血浆或其它血液样品基质中的ChIP-Seq。当已经描述了血浆中的ChIP-Seq时,已经发现核小体和核小体组蛋白的水平是高的(相对于单个转录因子的水平)。

[0283] 我们已经通过使用高亲和力抗体和通过使用适当的固相支持物将固相支持物上其它蛋白的非特异性结合降低至极低的水平,并结合使用含有高浓度强洗涤剂的溶液严格洗涤固相,解决了这些困难。

[0284] 因此,在提取转录因子缔合的DNA之前,可以用强(例如至少1%,例如1.2%的浓度)洗涤剂或洗涤剂混合物洗涤抗体结合的转录因子-DNA复合物。在一个实施方案中,在检测缔合DNA片段之前,用含有至少1%浓度的洗涤剂的缓冲溶液洗涤在步骤(i)中被结合剂结合的转录因子。有非常大量的洗涤剂可以用于该目的。一些常见的实例包括但不限于Triton洗涤剂(例如TritonX-100)、Tween洗涤剂(例如Tween 20和Tween 80)、脱氧胆酸钠、十二烷基硫酸钠、辛基苯氧基聚乙氧基乙醇(IGEPAL CA-630)、二十二碳烷基三甘醇十二烷基醚(Brij)、正十二烷基- β -麦芽糖糖苷、辛基- β -葡糖苷、辛基硫代葡糖苷、3-((3-胆酰胺丙基)二甲基铵)-1-丙磺酸盐(CHAPS)和更多。

[0285] 我们使用磁性聚苯乙烯微珠,并用含有1%辛基苯氧基聚乙氧基乙醇、0.1%脱氧胆酸钠和0.1%十二烷基硫酸钠的混合物的洗涤溶液重复洗涤(5次洗涤)。

[0286] 在本发明的优选的实施方案中,固相支持物是聚苯乙烯颗粒,例如磁性聚苯乙烯颗粒。所用的抗体(或转录因子的其它结合剂)可以直接或间接地附着于支持物。

[0287] 在本发明的优选的实施方案中,用含有至少0.25%、或至少0.5%或至少1%洗涤剂或表面活性剂的溶液洗涤在固相支持物上分离的固相结合转录因子-DNA复合物。所用的洗涤剂可以由单一洗涤剂或如本文所述的洗涤剂的混合物组成。

[0288] 在本发明的一个实施方案中,所用的固相转录因子结合剂支持物包含多重系统,例如多重珠系统(例如Luminex Corporation提供的系统)。在该固体支持物系统中,可以基于荧光区分的多个珠可以各自用不同转录因子的不同的特异性结合剂涂布,并且同时用于研究单个样品中的多个转录因子-DNA复合物(Dunbar, 2006)。

[0289] DNA测序

[0290] 本领域已知有许多方法来分析、定量或鉴定DNA序列,并且任何DNA分析方法都可以用于本发明的方法,包括但不限于下一代测序方法、等温DNA扩增、冷PCR(在较低变性温度下的共扩增-PCR)、MAP(MIDI活化的焦磷酸解作用)、PARE(重排末端的个性化分析)、DNA杂交方法(包括基因芯片方法和原位杂交方法)。此外,还可以通过表观遗传DNA测序分析来分析基因序列的表观遗传改变的DNA序列(例如,对于含有5-甲基胞嘧啶的序列,使用未修饰的胞嘧啶至尿嘧啶的亚硫酸氢盐转化)。因此,在一个实施方案中,使用DNA测序分析缔合DNA,所述DNA测序例如选自以下的测序方法:下一代测序(靶向或全基因组)和甲基化的DNA测序分析、BEAMing、PCR(包括数字PCR和冷PCR(在较低变性温度下的共扩增-PCR))、等温扩增、杂交、MIDI活化的焦磷酸解作用(MAP)或重排末端的个性化分析(PARE)。

[0291] 本文所述的实例使用IlluminaNovaSeq测序。因此,在本发明的优选的实施方案

中,通过下一代测序分析从分离的转录因子中提取的DNA。

[0292] 样品制备

[0293] 样品可以是其中可以检测染色质片段的任何体液。已知染色质片段存在于血液、粪便、尿和脑脊液中。我们还检测了痰中的染色质片段。在优选的实施方案中,体液样品是血液、血清或血浆样品。这些样品可以用于测量和分析含有转录因子和DNA片段的循环无细胞染色质片段。

[0294] 当血液样品用于本发明的方法时,其可以是全血、血清样品或血浆样品。全血或血清样品可以用作分析任何(稳定结合的)转录因子-DNA染色质片段的底物,所述片段涉及具有任何DBD类型的转录因子。

[0295] 血浆样品(例如EDTA血浆样品)也可以用于本发明的方法中。在典型的血浆样品收集方法中,将全血收集到柠檬酸盐或EDTA血液收集管中并在2小时内离心。所得上清液血浆可以新鲜使用或可以冷冻直至分析。然而,用作血液收集管添加剂以产生血浆的钙离子螯合剂引起循环锌指转录因子-DNA复合物的解离。如上所述,最常见类型的转录因子是锌指转录因子。

[0296] 有许多方式来克服这个困难,包括但不限于:(i)避免使用锌指转录因子和使用具有其它DBD类型的转录因子,(ii)使用血清样品,(iii)使用肝素血浆或不涉及钙螯合的其它血浆样品类型,或(iv)防止转录因子-DNA复合物解离,例如通过交联血液样品中染色质片段中的蛋白质和/或DNA。

[0297] 在一个实施方案中,体液样品是血清样品。认为血清含有衍生自白细胞的污染染色质材料(例如NET)。这种污染干扰cfDNA的分析,并因此血浆是ctDNA方法最常用的样品基质。然而,在DNA分析前从存在的其它染色质材料中分离含有转录因子的染色质片段去除这样的干扰。此外,血清被染色质材料污染是由通过凝血(NETosis的已知诱导物)触发的血液样品中的嗜中性粒细胞形成嗜中性粒细胞胞外陷阱(NET)的结果。条件是含有全血的血清样品收集管以及时的方式处理,例如在静脉穿刺后15-60分钟,污染性NET材料将是大的染色质而不是小的染色质片段,并且将不干扰小转录因子-DNA复合物的分析。因此,拓宽可以使用的样品类型是本发明的方法的进一步优点。

[0298] 通过向血清血液收集管中加入NETosis抑制剂,可以进一步最小化或消除血清中污染NET的存在。这防止NETosis,并因此使血清样品中存在的背景染色质水平最小化。NETosis的许多抑制剂是本领域已知的。优选的抑制剂包括葱环类药物,特别是阿霉素。因此,在本发明的一个实施方案中,提供了检测从人或动物受试者获得的血清样品中包含转录因子和DNA片段的无细胞染色质片段的方法,其包括以下步骤:

[0299] (i)在血清血液收集管中从受试者获得全血样品;

[0300] (ii)使全血样品与NETosis抑制剂接触;

[0301] (iii)从全血样品中分离血清样品;

[0302] (iv)使血清样品与结合转录因子的结合剂接触;

[0303] (v)检测或测量与转录因子缔合的DNA片段;和

[0304] (vi)使用DNA片段的存在或量或序列作为血清样品中包含转录因子的无细胞染色质片段的量的量度。

[0305] 应当理解,本发明的该实施方案也可以用于提供信息作为受试者的疾病状态的指

示物,如本文先前所述。

[0306] 在一个实施方案中,体液样品是任何血浆样品,包括使用钙螯合剂生产的气浆样品,例如EDTA血浆或柠檬酸盐血浆,其中血浆样品通过使全血样品与交联剂接触而获得。交联剂可以在方法的第一步骤中与全血接触,该方法涉及:(1)使全血样品与交联剂接触;(2)使交联的样品与钙离子螯合剂接触;和(3)从样品中分离血浆。

[0307] 交联是本领域的公知技术。最常用的交联试剂是甲醛,其将蛋白质分子彼此结合并结合至DNA。然而,过度交联可能导致转录因子中抗体结合表位的结构改变(并因此导致丧失抗体结合),并且甚至导致转录因子交联以分离蛋白质分子或复合物。为了防止这种情况,通常在加入甲醛后数秒或数分钟猝灭交联,例如通过加入过量的甘氨酸或三(羟甲基)氨基甲烷(TRIS),以停止进一步的交联。因此,在本发明的一个方面,提供了检测、分析或测量在取自人或动物受试者的血液样品中含有转录因子和缔合DNA片段的染色质片段的方法,其包括以下步骤:

[0308] (i)使从受试者获得的血液样品与交联剂接触;

[0309] (ii)任选地加入猝灭剂以停止进一步的交联;

[0310] (iii)使样品与钙离子螯合剂接触;

[0311] (iv)从样品中分离血浆;

[0312] (v)使血浆样品与结合转录因子的结合剂接触;

[0313] (vi)分离含有转录因子的结合的染色质片段;和

[0314] (vii)分析分离的染色质片段(例如通过本文所述的方法)。

[0315] 在本发明的另一方面,提供了检测人或动物受试者中的疾病的方法,其包括以下步骤:

[0316] (i)使从受试者获得的血液样品与交联试剂接触;

[0317] (ii)任选地加入猝灭剂以停止进一步的交联;

[0318] (iii)使样品与钙离子螯合剂接触;

[0319] (iv)从样品中分离血浆;

[0320] (v)使血浆样品与结合转录因子的结合剂接触;

[0321] (vi)分离与转录因子缔合的DNA;

[0322] (vii)任选地通过PCR方法扩增分离的DNA;

[0323] (viii)确定DNA的量和/或序列;和

[0324] (ix)使用转录因子的存在和/或缔合DNA的序列作为用于检测受试者的疾病状态的生物标志物。

[0325] 在优选的实施方案中,甲醛或甲醛释放剂用作交联剂。在一个实施方案中,EDTA用作钙离子的螯合剂以防止凝结。在优选的实施方案中,在收集全血样品之后立即将甲醛加入到全血中,例如通过将全血样品加入到已经含有甲醛的管中。将该管放置足够的时间以进行交联反应,然后通过加入猝灭剂来停止反应以防止血浆组分的过度交联。猝灭剂通常是胺化合物,例如与甲醛反应的甘氨酸或TRIS。猝灭剂可以与EDTA一起加入,例如通过加入甘氨酸和EDTA在TRIS缓冲液中的溶液。然后离心全血样品,并分离含有交联的转录因子结合的DNA复合物的血浆,用于通过本发明的方法进行分析。

[0326] 在本发明的另一方面,提供了检测人或动物受试者中的疾病的方法,其包括以下

步骤:

- [0327] (i)使从人或动物受试者获得的血液样品与交联试剂接触;
- [0328] (ii)使全血样品与猝灭试剂和钙离子螯合剂接触;
- [0329] (iii)分离在步骤(ii)中由样品生产的血浆;
- [0330] (iv)使血浆样品与结合转录因子的结合剂接触;
- [0331] (v)分离与转录因子缔合的DNA;
- [0332] (vi)任选地扩增分离的DNA;
- [0333] (vii)确定DNA的量和/或序列;和
- [0334] (viii)使用转录因子的存在和/或缔合DNA的序列作为用于检测受试者中疾病存在和/或性质的生物标志物。

[0335] 如上所述,存在于循环中的转录因子最可能是稳定结合DNA的那些转录因子,而不是与DNA瞬时缔合并以动态方式解离的那些转录因子。对于包括转录因子的最稳定结合的DNA循环染色质片段,在完整培养的细胞或组织样品中与甲醛交联是快速的,并且花费少于1或2分钟。我们推论,虽然甲醛扩散和进入细胞可能需要1或2分钟,然后进入细胞核,然后染色质交联,但在全血环境中这一时间可以减少,其中染色质片段在溶液中是游离的并且可立即交联。所用的交联试剂可以是甲醛或甲醛释放剂(releasing agent)(也称为甲醛释放剂(releaser)、甲醛供体或甲醛释放防腐剂)。甲醛释放剂是缓慢释放甲醛的部分。许多甲醛释放剂是本领域已知的,并且通常在化妆品工业中用作抗微生物防腐剂,例如在皮肤护理和毛发护理产品中,其中由于毒性避免了高水平的甲醛,但是通过释放保持低的保护水平。因此,在一个实施方案中,交联剂是甲醛释放剂。

[0336] 我们推论,全血(与细胞或组织相反)中无细胞循环转录因子-DNA复合物的交联是快速的,并且可以比锌指蛋白的锌消耗更快地发生。因此,在本发明的一个实施方案中,交联试剂可以与钙离子螯合剂同时加入。含有EDTA和甲醛释放剂两者的血液收集管(BCT)是商业上可获得的,例如,可从Streck Inc获得的无细胞DNA BCT。将加入到这样的管中的全血同时暴露于EDTA和交联剂。

[0337] 我们使用不同的EDTA样品制备方法进行了许多实验。例如,雌性激素受体(ER)是锌指转录因子。我们使用ELISA方法测量(常规)EDTA血浆样品中存在的ER的水平。ER是可检测的,如在图5中所示。我们从EDTA血浆样品中免疫沉淀ER,提取与固相结合DNA,并扩增提取物中存在的DNA。然而,在扩增的样品中没有观察到DNA。我们推论这是因为ER-DNA复合物在EDTA血浆中解离。

[0338] CTCF(也称为CCCTC-结合因子)是进化上保守的锌指转录因子,其通过11个锌指的组合结合基因组中的大量位点,并且在基因组功能中具有关键作用。对人基因组中CTCF结合位点的研究鉴定了19种不同细胞类型中77,811个不同的结合位点(Wang等人,2012)。发现77,811个结合位点中的27,662个在所有19种研究的细胞类型中被占据。剩余50,149个结合位点的CTCF结合表现出组织特异性。研究的19种细胞类型包括12种正常细胞类型和7种癌症或EBV永生化的细胞系,代表结肠直肠癌(Caco-2)、宫颈癌(HeLa-S3)、肝细胞癌(HepG2)、成神经细胞瘤(SK-N-SH_RA)、成视网膜细胞瘤(WERI-RB-1)和EBV转化的淋巴质体(GM06990)。发现在1,236个结合位点的CTCF结合对癌细胞系是特异性的,并且这些结合位点的占据将永生和癌细胞系与正常细胞(包括上皮、成纤维细胞和内皮)区分开来(Liu等

人,2017)。

[0339] 我们使用小鼠抗CTCF抗体免疫沉淀来自4个汇集的交联的EDTA血浆样品(在Streck cfDNA BCT中收集)的CTCF-DNA,所述血浆样品从18个诊断患有癌症的受试者收集。我们在固相支持物上通过ChIP进行分离的蛋白质的蛋白质印迹分析。图7中的结果显示,在所有4个汇集的样品中都存在对应于分子量为约140kD的CTCF的蛋白质带(但在采用非特异性小鼠IgG代替抗CTCF抗体的对照实验中不存在)。约50kD的带对应于用于蛋白质印迹的标记的抗小鼠IgG抗体与用于ChIP的小鼠抗CTCF抗体的重链的结合。

[0340] 然后我们使用从诊断患有乳腺癌的受试者收集的交联的EDTA血浆样品(在Streck cfDNA BCT中收集)重复ChIP方法以免免疫沉淀CTCF-DNA复合物。我们从固相支持物提取cfDNA片段,将提取的DNA片段连接到接头寡核苷酸,并扩增存在的cfDNA。通过电泳分析扩增的cfDNA文库,并且所得电泳图谱(图8)显示文库含有35-80bp范围的小片段(其对应于在x轴上175-220bp之间的峰,以说明接头连接的片段)。在约50bp长度处观察到接头连接的cfDNA片段的主峰(其对应于在x轴上190bp处的峰,以说明接头连接的片段长度)。尽管扩增的cfDNA文库含有35-80bp范围的小片段,不是所有这些片段都与样品中的CTCF结合,因为从涂布有非特异性小鼠IgG的固体支持物中扩增的提取物也获得小DNA片段。然而,用特异性抗CTCF抗体ChIP(1000荧光单位(FU))获得的特异性峰高于非特异性IgG峰(80FU)。

[0341] 通过下一代测序方法对使用抗CTCF免疫沉淀分离的扩增的cfDNA文库测序。从诊断患有CRC的患者收集的交联的EDTA血浆样品(在Streck cfDNA BCT中收集)制备的扩增的文库的结果示于图9中。我们观察到9780种公开的CTCF TFBS序列的小cfDNA片段结合的富集(Kelly等人,2012)。相反,结合非特异性小鼠IgG获得的cfDNA文库未显示富集。参考输入的非特异性对照,cfDNA片段序列的峰型调用导致CTCF作为转录因子,具有最多的TFBS序列片段。我们得出结论,本发明的方法成功用于血浆中转录因子的ChIP-Seq。

[0342] 雄性激素受体(AR)是前列腺癌中目的锌指转录因子。为了显示本发明的方法可以应用于比CTCF丰度更低的转录因子,我们将相同的方法应用于AR。我们使用小鼠抗AR抗体从来自8个诊断患有前列腺癌的受试者的交联的EDTA血浆样品(在Streck cfDNA BCT中收集)免疫沉淀AR。我们使用来自LnCAP前列腺癌细胞系细胞的AR作为阳性对照,在固相支持物上进行通过ChIP分离的蛋白质的蛋白质印迹分析。图11的结果显示,在所有8个样品中都存在对应于分子量为约10kD的AR的蛋白质带,并且在2个样品中特别强(图11的泳道2和3)。约50kD的带对应于标记的抗小鼠IgG抗体与用于ChIP的小鼠抗AR抗体的重链的结合。然后,我们从固相支持物提取DNA,将提取的DNA片段与接头寡核苷酸连接,并扩增存在的DNA。图12中的结果显示扩增的cfDNA文库含有35-80bp范围的小片段(如上所述,对于接头连接的片段,在175-220bp处显示的峰)。尽管扩增的cfDNA文库含有35-80bp范围的小片段,但不是所有这些片段都与样品中的AR结合,因为从涂布有非特异性小鼠IgG的固体支持物的扩增的提取物也获得了小DNA片段。然后通过下一代测序对通过蛋白质印迹观察到的AR水平最高的2个样品获得的扩增的cfDNA文库测序。

[0343] 解离的转录因子-DNA复合物

[0344] 本发明的前述方面是检测、测量或表征包括直接或间接结合DNA的转录因子的染色质片段的方法。在本发明的一个实施方案中,存在用于检测取自受试者的体液样品中未与DNA结合的转录因子(即游离或未结合的转录因子)的方法。游离转录因子的检测可以通

通过使用包括转录因子的TFBS DNA序列(任选地包括侧翼序列)的寡核苷酸作为游离转录因子的结合剂来进行。然后可以检测寡核苷酸结合的游离转录因子,例如使用标记的抗转录因子抗体(例如参见Active Motif,2006)。转录因子最初可以以无活性形式产生,其随后可以例如通过磷酸化而被翻译后激活。活性转录因子形式与包括其TFBS序列的寡核苷酸结合。非活性转录因子形式不结合包括其TFBS序列的寡核苷酸(Lee等人,2007)。因此,可以使用涉及将游离转录因子结合到寡核苷酸的测定来检测体液样品中的活性游离转录因子,所述寡核苷酸包括转录因子结合的DNA序列,例如转录因子的TFBS序列,然后加入第二转录因子结合剂,例如定向特异性结合转录因子的抗转录因子抗体,并使用抗体结合的存在或程度作为样品中存在的活性游离转录因子的存在或量的量度。因此,在本发明的一个实施方案中,提供了检测人或动物受试者中的游离转录因子的方法,其包括以下步骤:

[0345] (i)使从人或动物受试者获得的体液样品与结合转录因子的寡核苷酸接触;

[0346] (ii)分离寡核苷酸结合的转录因子;

[0347] (iii)使分离的转录因子与结合转录因子的第二结合剂接触;和

[0348] (iv)使用第二结合剂与转录因子结合的存在或程度作为样品中无细胞转录因子的量的度量。

[0349] 在优选的实施方案中,用于结合游离转录因子的寡核苷酸包括TFBS序列。在优选的实施方案中,用于结合游离转录因子的寡核苷酸附着于固相支持物。在优选的实施方案中,第二结合剂是抗体。在优选的实施方案中,第二结合剂被标记,使得其与固相寡核苷酸结合的转录因子的结合可以容易地检测和/或定量。

[0350] 在一个实施方案中,将锌离子加入到样品中以促进寡核苷酸与锌指转录因子结合。锌离子可以在步骤(i)中加入寡核苷酸的同时加入,或在步骤(i)之前加入。

[0351] 在本发明的另一方面,提供了检测人或动物受试者中的疾病的方法,其包括以下步骤:

[0352] (i)使从人或动物受试者获得的体液样品与结合转录因子的寡核苷酸接触;

[0353] (ii)分离寡核苷酸结合的转录因子;

[0354] (iii)使分离的转录因子与结合转录因子的第二结合剂接触;和

[0355] (iv)使用第二结合剂与转录因子结合的存在或程度作为受试者中疾病存在和/或性质的指示物。

[0356] 在一个实施方案中,使取自受试者的体液样品与一种或多种寡核苷酸(例如,对结合一种或多种转录因子特异性的TFBS序列)接触以鉴定疾病存在和/或性质。在进一步的实施方案中,使用测试一种或多种疾病的多重测定(即包含多于一种寡核苷酸,优选其中每种寡核苷酸对不同的转录因子具有特异性)进行所述方法。例如,任选地除了在许多癌症中表达的转录因子之外,测试各自对一种或多种癌症疾病特异性的多种转录因子,使得能够在单次血液测试中除了鉴定癌症组织之外还检测许多不同的癌症疾病。多重测试的方法是本领域公知的,例如但不限于DNA微阵列方法或Luminex Corporation的多重珠系统,其可以用于在单个样品中进行大量的多重测定(Dunbar,2006)。

[0357] 在优选的实施方案中,疾病是癌症。在另一个实施方案中,疾病的性质是受癌症影响的组织。

[0358] 雌性激素受体(ER)是配体激活的核激素受体锌指转录因子。我们推论,包括锌指

转录因子和DNA片段的血液中循环染色质片段可能在EDTA血浆样品中被破坏。我们对取自涉及雌性激素受体过表达的妇科癌症患者以及取自患有ER阴性乳腺癌患者的血浆样品中的游离(即,未与DNA结合)雌性激素受体 α (ER α)进行了酶联免疫吸附测定(ELISA)测量。ER参与大量基因转录的调节,并在雌性生殖组织和生殖癌组织中高度表达。ER在造血细胞中以低水平表达,但在ER阳性乳腺癌和卵巢癌细胞中高度表达。ER阳性癌细胞具有雌性激素受体,对雌性激素敏感,并且它们的生长受雌性激素刺激。ER阴性癌细胞不具有雌性激素受体,并且对雌性激素不敏感。约80%的卵巢癌和乳腺癌是ER阳性的。和ER阴性癌症相比,ER阳性癌症与更好的预后相关。由于ER阳性癌症应答雌性激素而生长,它们适合于激素治疗,包括他莫昔芬和芳香酶抑制剂,其通过结合雌性激素抑制雌性激素受体的活化并因此预防癌症生长。

[0359] 通过手术切除的癌组织的免疫组织化学测试来确定癌症的ER阳性或阴性状态。通常,将结合ER的标记的抗体与癌细胞/组织一起孵育,并且观察到的抗体染色水平决定状态。将ER阳性癌症指定ER评分。测量激素受体测试为阳性的癌细胞的比例以及染色的强度。将两个参数组合以在0-8的等级上对样品评分。具有更多在较高强度下可见的受体的样品评分较高。

[0360] 由于核激素受体是细胞蛋白,预期ER不存在于循环中。我们假设,血浆中存在的任何游离ER α 必须源自包括ER α 的循环染色质片段,但在加入EDTA以产生血浆时解离以从DNA结合释放游离ER α 。我们预期这样的染色质片段的水平是非常低的,并因此预期发现血浆中游离ER α 的水平对于ELISA方法是不可检测的,并且低于所用ELISA的最小灵敏度(0.8pg/ml)。令人惊奇的是,我们发现游离ER α 以高达20pg/ml的水平存在于血浆中(图5)。为了将这种情况放在本上下文中,白细胞介素-6和肿瘤坏死因子通常是测量的血液生物标志物,其正常范围分别为约5-15pg/ml和高达8pg/ml。此外,测量的ER α 水平在卵巢癌和ER阳性乳腺癌中高于ER阴性乳腺癌,指示ER α 的肿瘤起源。

[0361] 因此,在本发明的另一方面,提供了检测生物样品中锌指转录因子的存在或水平测量的方法,其包括以下步骤:

[0362] (i)使样品与锌离子螯合试剂接触;和

[0363] (ii)分析样品中置换的锌指转录因子的存在或水平。

[0364] 在一个实施方案中,生物样品是体液样品,例如血液、血清或血浆。在进一步的实施方案中,锌离子螯合剂是EDTA。可以将EDTA加入到体液样品中以破坏锌指-DNA结合。

[0365] 在优选的实施方案中,生物样品是全血样品,并且锌离子螯合试剂是EDTA,将EDTA加入到全血样品中以破坏锌指-DNA结合,以及防止血液凝固,并因此产生含有游离锌指转录因子的血浆样品。任何方法都可以用于分析样品中的转录因子。在优选的实施方案中,采用的分析方法是免疫测定,并且特别是2-位点“夹心”免疫测定。因此,在本发明的优选的实施方案中,提供了用于检测在取自受试者的全血样品中含有锌指转录因子的循环染色质片段的的存在或水平测量的方法,其包括以下步骤:

[0366] (i)使全血样品与EDTA接触以生产血浆样品;和

[0367] (ii)使用免疫测定方法分析血浆样品中锌指转录因子的存在或水平。

[0368] 锌指转录因子家族是最丰富的转录因子家族。因此,本发明的该方面可以用于检测大多数目的转录因子。术语“锌指转录因子”是指含有锌指-结合结构域的任何转录因子。

[0369] 循环锌指转录因子可以用作用于检测疾病的生物标志物,例如妇科癌症的检测、诊断、治疗选择、监测或预后。因此,在本发明的一个实施方案中,提供了用于确定受试者的疾病状态的方法,例如用于受试者中疾病的检测、诊断、治疗选择、监测或预后或用于疾病,其包括以下步骤:

[0370] (i)使从受试者获得的血液样品与锌螯合剂接触以生产血浆样品;

[0371] (ii)分析血浆样品中锌指转录因子的存在或水平;和

[0372] (iii)使用样品中锌指转录因子的存在或水平作为受试者的疾病状态的指示物。

[0373] 本发明的该方面也适用于细胞培养方法。用于转录因子的染色质免疫沉淀(ChIP)方法是复杂的、困难的、耗时的和不稳健的。典型的ChIP方法涉及从细胞中提取染色质材料,通过DNA消化或使用物理方法(例如超声处理)使染色质片段化,使用抗体分离染色质片段,提取与抗体缔合的DNA并测定提取的DNA的DNA序列。使用本发明的方法,可以通过将染色质材料从细胞提取到含有EDTA(或其它锌螯合剂)的流体中并测量游离锌指转录因子(例如通过ELISA)来建立锌指转录因子的存在或量。

[0374] 任何方法都可以用于分析样品中锌指转录因子的存在或量,包括但不限于质谱法和任何免疫化学方法。在优选的实施方案中,用于分析样品中锌指转录因子的存在或量的方法是免疫测定。

[0375] 由于我们已经发现,向含有包括锌指转录因子的染色质片段的样品中加入锌离子螯合剂导致这些染色质片段的破坏,以生产游离锌指转录因子,并且EDTA是锌(以及钙)离子的强螯合剂,将清楚的是,在EDTA血浆样品中不能使用涉及用抗体(或其它转录因子结合剂)分离转录因子并分析与转录因子缔合的DNA的方法来研究锌指转录因子结合的DNA,因为DNA将不再与转录因子缔合。

[0376] 应当理解,破坏锌指转录因子与DNA的结合将导致游离锌指转录因子以及游离DNA片段两者,所述游离DNA片段包括TFBS序列和基因组中的侧翼DNA序列。因此,在本发明的进一步的方面,提供了用于鉴定受试者中含有锌指转录因子或与该锌指转录因子结合的DNA片段的一个或多个序列的循环染色质片段的存在的方法,其包括以下步骤:

[0377] (i)使从受试者获得的血液样品与锌螯合剂接触以生产血浆样品;和

[0378] (ii)分析血浆样品中含有包括转录因子结合位点序列或与锌指转录因子结合位点的侧翼序列的DNA序列的游离DNA片段的存在或水平。

[0379] 含有转录因子和缔合TFBS的染色质片段的存在可以用于临床目的,包括用于本文所述疾病的检测、监测、预后或治疗选择或用于本文所述疾病。因此,在本发明的一个方面,提供了用于确定受试者的疾病状态的方法,例如用于疾病的检测、监测、预后或治疗选择或用于疾病,其包括以下步骤:

[0380] (i)使从受试者获得的血液样品与锌螯合剂接触以生产血浆样品;

[0381] (ii)分析血浆样品中含有包括转录因子结合位点序列或与锌指转录因子结合位点的侧翼序列的DNA序列的游离DNA片段的存在或水平;和

[0382] (iii)使用样品中DNA片段的存在和/或水平和/或序列作为受试者的疾病状态的指示物。

[0383] 血浆或其它样品中核小体或其它蛋白质结合的DNA片段中游离DNA片段的存在和/或序列可以通过许多方式来确定,包括使用互补DNA序列来结合样品中的DNA片段。这可以

例如通过使用DNA芯片来实现,所述芯片促进同时探测样品的多个序列。本发明的另一个实施方案涉及使用外源锌指转录因子作为特异性DNA结合剂。在该方法中,去除锌螯合剂以促进锌指转录因子与DNA的结合。这可以通过缓冲液交换进行,例如通过透析或通过使用尺寸排阻层析法,例如使用sephadex尺寸排阻层析柱。可以分离含有锌指转录因子的TFBS的DNA片段,例如通过使用固相结合转录因子作为含有TFBS的游离DNA的结合剂。可以分析分离的DNA的序列和/或DNA片段长度。重组转录因子蛋白可以用于本发明的目的。重组锌指转录因子蛋白可以与固相支持物连接或可以含有接头部分,并且转录因子可以以液体形式使用并通过连接系统分离。许多这样的连接样品是本领域已知的,例如锌指转录因子可以使用固相链霉亲和素生物素化和分离。因此,在本发明的一个实施方案中,提供了用于鉴定受试者中含有锌指转录因子和/或与该锌指转录因子结合的DNA片段的一个或多个序列的循环染色质片段的存在的方法,其包括以下步骤:

[0384] (i) 使从受试者获得的血液样品与锌螯合剂接触以生产血浆样品;

[0385] (ii) 从样品中去除锌螯合剂;

[0386] (iii) 使样品与外源锌指转录因子接触;和

[0387] (iv) 分析被外源转录因子结合的DNA片段。

[0388] 作为另外一种选择或除此之外,锌螯合剂可以在样品中简单地失活。在本发明的一个实施方案中,在与外源转录因子接触之前,通过加入过量的离子,优选锌离子,使锌螯合剂失活。因此,在本发明的一个实施方案中,提供了用于鉴定受试者中含有锌指转录因子和/或与该锌指转录因子结合的DNA片段的一个或多个序列的循环染色质片段的存在的方法,其包括以下步骤:

[0389] (i) 使从受试者获得的血液样品与锌螯合剂接触以生产血浆样品;

[0390] (ii) 通过加入过量的锌或其它离子使样品中的锌螯合剂失活;

[0391] (iii) 使样品与外源锌指转录因子接触;和

[0392] (iv) 分析被外源转录因子结合的DNA片段。

[0393] 含有转录因子和缔合TFBS的染色质片段的的存在可以用于临床目的,包括用于本文所述疾病的检测、监测、预后或治疗选择或用于本文所述疾病。因此,在本发明的一个方面,提供了用于确定受试者的疾病状态的方法,例如用于疾病的检测、监测、预后或治疗选择或用于疾病,其包括以下步骤:

[0394] (i) 使从受试者获得的血液样品与锌螯合剂接触以生产血浆样品;

[0395] (ii) 去除或使样品中的锌螯合剂失活;

[0396] (iii) 使样品与外源锌指转录因子接触;

[0397] (iv) 分析被外源转录因子结合的DNA片段,和

[0398] (v) 使用样品中DNA片段的的存在和/或水平和/或序列作为受试者的疾病状态的指示物。

[0399] 无细胞核小体的去除

[0400] 样品制备也可以任选地涉及预纯化步骤以在分析前从样品中去除大部分核小体和核小体结合的DNA。这降低了背景信号,改进分离和扩增目的转录因子结合的DNA片段的效率,并且可以改进本发明的方法的分析和临床灵敏度。因此,在一个实施方案中,所述方法另外包括从体液样品中去除无细胞核小体。在采用本文所述的本发明的方法之前,可以

从样品中去除包含核小体染色质片段(任选地单独分析)。该制备步骤的目的是从样品中去除大部分DNA片段,以降低它们在分析中可能产生的任何背景信号。这可以例如但不限于通过使样品与结合核小体的结合剂接触来完成,所述结合剂例如固相抗核小体结合剂,包括例如抗体或核小体结合蛋白,例如在W02021038010中所述的蛋白质。抗体可以选择性地结合组蛋白,例如核心组蛋白(例如H2A、H2B、H3或H4)或接头组蛋白(例如H1)。提及组蛋白时,包括组蛋白翻译后修饰和组蛋白变体或同种型。核小体结合蛋白可以选自:结合接头DNA的染色质结合蛋白或结合核小体缔合接头DNA的蛋白。例如,结合接头DNA的染色质结合蛋白可以选自:色域解旋酶DNA结合(CHD)蛋白;DNA(胞嘧啶-5)-甲基转移酶(DNMT)蛋白;高迁移率族盒蛋白(HMGB)蛋白;聚[ADP-核糖]聚合酶(PARP)蛋白;或甲基-CpG-结合结构域(MBD)蛋白,例如MBD1、MBD2、MBD3、MBD4或甲基CpG结合蛋白2(MECP2)。结合核小体缔合接头DNA的蛋白质可以选自组蛋白H1、macroH2A(mH2A)、或其片段或工程类似物。

[0401] 可以吸附(例如,吸附到固相上)样品中存在的全部或大部分核小体材料,并因此从样品中去除。因此,在一个实施方案中,该方法包括使体液样品与结合核小体或其组分的结合剂接触,并在使样品与转录因子结合剂接触之前去除与结合剂结合的样品。

[0402] 已经报道,血浆中长度小于100bp的短cfDNA片段的大的部分或大部分不是衍生自包括调节蛋白的染色质片段,而是衍生自在一条或两条DNA链中有切口或断裂的核小体缔合的DNA。在这种情况下,短cfDNA片段可以代表例如与核小体缔合的150bp DNA片段,所述核小体在一个或多个位置被切口以产生两个或更多个更小的cfDNA片段(例如两个75bp的片段),而不是单个150bp cfDNA片段(Sanchez等人,2018)。因此,在将样品暴露于转录因子结合剂之前从样品中去除核小体具有去除源自核小体缔合的切口DNA的小于100bp的短cfDNA片段的另外的优点。例如,与通过凝胶分离方法对提取的cfDNA片段的大小分离相比,这进一步降低了样品中核小体缔合的cfDNA的背景。

[0403] 我们已经证明使用抗H3抗体从人血浆样品定量去除含有核小体的染色质片段。

[0404] 在优选的实施方案中,磁性珠用作固相支持物,但可以使用任何合适的材料。类似地,可以使用W02016067029、W02017068371和W02021038010中描述的用于核小体结合的任何方法作为去除核小体的方法。因此,在一个实施方案中,本发明的方法中使用的样品不包含核小体。在进一步的实施方案中,通过本发明的方法检测的无细胞染色质片段由转录因子和DNA片段组成。

[0405] 在本发明的一个实施方案中,提供了检测人或动物受试者中的疾病的方法,其包括以下步骤:

[0406] (i) 从由人或动物受试者获得的体液样品中去除无细胞核小体;

[0407] (ii) 使样品与结合转录因子的结合剂接触;

[0408] (iii) 分离与转录因子缔合的DNA;

[0409] (iv) 通过PCR方法扩增分离的DNA;

[0410] (v) 确定扩增的DNA的序列;和

[0411] (vi) 使用转录因子的存在和缔合DNA的序列作为用于确定受试者中疾病存在和/或性质的组合生物标志物。

[0412] 在本发明的一些实施方案中,可以在不分离DNA的情况下确定与无细胞转录因子或染色质片段缔合的DNA片段的存在或序列。这可以通过多种方法完成,包括但不限于不需

要DNA分离的扩增方法。

[0413] 如本文所用的术语“结合剂”是指能够特异性结合生物标志物(即特异性转录因子)的配体或结合剂,例如天然存在的或化学合成的化合物。根据本发明的配体或结合剂可以包含能够特异性结合生物标志物的肽、抗体或其片段,或合成配体(例如塑料抗体)或适体或寡核苷酸或分子印迹表面或装置。抗体可以是单克隆抗体或能够特异性结合靶的其片段。根据本发明的配体或结合剂可以用可检测标记物(例如发光、荧光、酶或放射性标记物)标记;作为另外一种选择或除此之外,根据本发明的配体可以用亲和力标记(例如生物素、抗生物素蛋白、链霉亲和素或His(例如六His)标记)来标记。在一个实施方案中,结合剂选自:抗体、抗体片段或适体。在进一步的实施方案中,所用的结合剂是抗体。术语“抗体”、“结合剂(binding agent)”或“结合剂(binder)”在本文中可互换使用。

[0414] 在一个实施方案中,样品是生物流体(其与本文的术语“体液”可互换使用)。任何体液样品类型都可以用于本发明,包括但不限于血液、血浆、经血、子宫内膜流体、粪便、尿、唾液、粘液、精液和呼气,例如作为浓缩的呼气,或从其提取物或纯化物,或其稀释物。生物样品还包括来自活的受试者或死后取得的试样。例如,可以适当稀释或浓缩来制备样品,并以通常的方式储存。在优选的实施方案中,生物流体样品选自:血液或血清或血浆。本领域技术人员清楚,体液中染色质片段的检测具有是不需要活检的微创方法的优点。

[0415] 在一个实施方案中,受试者是哺乳动物受试者。在进一步的实施方案中,受试者选自人或动物(例如伴侣动物或小鼠)受试者。在又进一步的实施方案中,受试者是人受试者。在一个实施方案中,人受试者是非胚胎受试者(即处于发育的任何阶段的人,而不是胚胎)。在进一步的实施方案中,人受试者是成年受试者,即大于16岁,例如大于18、21或25岁。在备选的实施方案中,受试者是动物受试者。在进一步的实施方案中,动物受试者选自啮齿动物(例如小鼠、大鼠、仓鼠、沙鼠或花栗鼠)、猫科动物(即猫)、犬科动物(即狗)、马科动物(即马)、猪科动物(即猪)或牛科动物(即母牛)受试者。

[0416] 应当理解,本发明的用途和方法可以在体外或离体进行。

[0417] 根据本发明的其它方面,提供了用于检测或诊断动物或人受试者中癌症的方法,其包括以下步骤:

[0418] (i)检测或测量在从受试者获得的体液样品中与包含转录因子的无细胞染色质片段缔合的DNA;和

[0419] (ii)使用在步骤(i)中检测的缔合DNA水平和/或DNA序列来鉴定受试者的疾病状态。

[0420] 根据本发明的其它方面,提供了用于检测或诊断动物或人受试者中的炎症性疾病的方法,其包括以下步骤:

[0421] (i)检测或测量在从受试者获得的体液样品中与包含转录因子的无细胞染色质片段缔合的DNA;和

[0422] (ii)使用在步骤(i)中检测的缔合DNA水平和/或DNA序列来鉴定受试者的炎症性疾病状态。

[0423] 在本发明的一个实施方案中,样品中包含转录因子和DNA片段的无细胞染色质片段的的存在用于确定需要这样的治疗的受试者的最佳治疗方案。

[0424] 根据本发明的其它方面,提供了用于评估动物或人受试者对医学治疗的适合性的

方法,其包括以下步骤:

[0425] (i) 检测、测量或测序在从受试者获得的体液样品中与包含转录因子的无细胞染色质片段缔合的DNA;和

[0426] (ii) 使用在步骤(i)中检测的缔合DNA水平和/或DNA序列作为参数来选择用于受试者的合适的治疗。

[0427] 根据本发明的其它方面,提供了用于监测动物或人受试者的治疗的方法,其包括以下步骤:

[0428] (i) 检测、测量或测序在从受试者获得的体液样品中与包含转录因子的无细胞染色质片段缔合的DNA;

[0429] (ii) 在一个或多个时机重复在受试者的体液中与包含转录因子的无细胞染色质片段缔合的DNA的检测、测量或测序;和

[0430] (iii) 使用与步骤(ii)相比在步骤(i)中检测的缔合DNA水平和/或DNA序列的任何变化作为受试者病况的任何变化的参数。

[0431] 相对于在早先取自同一测试受试者的先前测试样品中检测的水平或序列,在测试样品中检测的与含有转录因子的无细胞染色质片段缔合的测量的DNA水平和/或DNA序列的水平的变化可以指示所述疗法对病症或疑似病症的有益作用,例如稳定或改进。此外,一旦治疗已完成,可以周期性地重复本发明的方法以监测疾病的复发。

[0432] 应当理解,本发明的这些方面可以与本文公开的方法组合使用,例如步骤(i)包括使体液样品与结合转录因子的结合剂接触,然后检测或测量与所述转录因子缔合的DNA。

[0433] 在一个实施方案中,检测或测量包含转录因子和DNA片段的无细胞染色质片段(即与包含转录因子的无细胞染色质片段缔合的DNA)作为一组测量之一。例如,与其它无细胞染色质转录因子标志物或与任何其它生物标志物组合。

[0434] 根据本发明的其它方面,提供了用于检测、测量或测序包含转录因子和DNA片段的无细胞染色质片段的方法,所述方法单独或作为一组测量的一部分,用于确定或评估动物或人受试者对医学治疗的适合性,或用于监测动物或人受试者的治疗,用于患有实际或疑似癌症或良性肿瘤的受试者的目的。

[0435] 应当理解,通过本发明的方法进行的测量或测定可以包括使用参考材料作为校准物或阳性对照以提供标准,可以将测定的输出与该标准进行比较或校准,和/或证实或监测测定的化学的正确功能。合适的参考材料可以包括生物来源的含有转录因子的染色质片段或重组染色质片段,包括但不限于重组转录因子-DNA复合物。

[0436] 如本文所用,术语“检测”和“诊断”涵盖疾病状态的鉴定、证实和/或表征。根据本发明的检测、监测和诊断方法可用于证实疾病存在,以通过评估发病和进展来监测疾病的发展,或评估疾病的改善或消退。检测、监测和诊断的方法也用于评估临床筛选、预后、疗法的选择、治疗益处的评价的方法中,即用于药物筛选和药物开发。

[0437] 应当理解,检测和测量包括测序。如本文所用的术语“测序”包括确定DNA片段的全部或部分的核苷碱基序列(通常为腺嘌呤、鸟嘌呤、胸腺嘧啶和胞嘧啶碱基序列)。

[0438] 有效的诊断和监测方法通过建立正确的诊断、允许快速鉴定最合适的治疗(因此减少对有害药物副作用的不必要暴露)和降低复发率,提供了具有改进的预后潜力的非常有力的“患者解决方案”。

[0439] 应当理解,鉴定和/或定量可以通过适于鉴定在来自患者的生物样品或生物样品的纯化物或提取物或其稀释物中特定蛋白质或DNA片段序列的存在和/或量的任何方法来进行。在本发明的方法中,定量可以通过测序或通过测量一个或多个样品中生物标志物的浓度来进行。可以在本发明的方法中测试的生物样品包括如上文所定义的那些。例如,可以适当稀释或浓缩来制备样品,并以通常的方式储存。

[0440] 生物标志物的鉴定和/或定量可以通过检测生物标志物或其片段,例如具有C-末端截短或具有N-末端截短的片段来进行。片段的长度适宜地大于4个氨基酸,例如5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个氨基酸。

[0441] 生物标志物可以例如通过SELDI或MALDI-TOF直接检测。或者,生物标志物可以直接或间接地经由与一个或多个配体的相互作用来检测,所述配体例如抗体或其生物标志物-结合片段,或能够特异性结合生物标志物的其它肽或配体,例如适体或寡核苷酸。配体或结合剂可以具有可检测的标记,例如发光、荧光或放射性标记,和/或亲和力标记。

[0442] 例如,检测和/或定量可以通过选自以下的一种或多种方法进行:SELDI(-TOF)、MALDI(-TOF)、基于1-D凝胶的分析、基于2-D凝胶的分析、质谱(MS)、反相(RP)LC、尺寸渗透(凝胶过滤)、离子交换、亲和力、HPLC、UPLC和其它基于LC或LC MS的技术。合适的LC MS技术包括ICAT® (Applied Biosystems, CA, USA) 或iTRAQ® (Applied Biosystems, CA, USA)。也可以使用液相色谱法(例如高压液相色谱法(HPLC)或低压液相色谱法(LPLC))、薄层色谱法、NMR(核磁共振)光谱法。

[0443] 应当理解,检测和/或测量DNA可以包括例如如本文所述的杂交或测序。

[0444] 根据本发明的诊断或监测方法可以包括通过SELDI TOF或MALDI TOF分析样品以检测生物标志物的存在或水平。这些方法也适用于临床筛选、预后、监测疗法结果、鉴定最可能对特定治疗性治疗应答的患者、用于药物筛选和开发、以及鉴定用于药物治疗的新靶。

[0445] 鉴定和/或定量分析物生物标志物可以使用免疫学方法进行,涉及能够特异性结合生物标志物的抗体或其片段。

[0446] 根据本发明的其它方面,提供了用于鉴定包含转录因子和DNA片段作为组合生物标志物的无细胞染色质片段用于检测或诊断动物或人受试者中的疾病的方法,其包括以下步骤:

[0447] (i) 检测和/或测量和/或测序在患病受试者的体液样品中包含转录因子和DNA片段组合生物标志物的无细胞染色质片段;

[0448] (ii) 检测和/或测量和/或测序在健康受试者或对照受试者的体液样品中包含转录因子和DNA片段组合生物标志物的无细胞染色质片段;和

[0449] (iii) 使用在患病和健康或对照受试者中检测的水平和/或DNA序列之间的差异来鉴定包含转录因子和DNA片段组合生物标志物的无细胞染色质片段是否可用作疾病状态的生物标志物。

[0450] 应当理解,本发明的该方面可以与本文所述的方法组合,即,可以使用如本文所定义的方法来进行步骤(i)和/或(ii)。

[0451] 根据本发明的其它方面,提供了通过本文所述的方法鉴定的生物标志物或组合生物标志物。

[0452] 本文提供了用于进行本发明的方法的诊断或监测试剂盒。用于检测和/或定量生

物标志物或组合生物标志物的这样的试剂盒将适当地包含转录因子的配体或结合剂,和任选地用于与所述转录因子缔合的DNA的扩增和/或测序的试剂,和任选地核小体的配体或结合剂,任选地以及试剂盒的使用说明书。生物标志物监测方法、生物传感器和试剂盒作为患者监测工具也是至关重要的,以使医师能够确定复发是否是由于病症的恶化。如果药理学治疗被评估为不充分,则疗法可以恢复或增加;如果合适,可以改变疗法。由于生物标志物对病症状态敏感,它们提供了药物疗法影响的指示。

[0453] 根据本发明的其它方面,提供了用于检测包含转录因子和DNA片段作为组合生物标志物的无细胞染色质片段的试剂盒,其包含转录因子的配体或结合剂,任选地用于与所述转录因子缔合的DNA的扩增和/或测序的试剂,和任选地核小体的配体或结合剂,任选地以及根据本文所述的方法使用试剂盒的说明书。

[0454] 本发明的其它方面是用于检测疾病状态的存在的试剂盒,其包含能够检测和/或定量一种或多种如本文所定义的生物标志物的生物传感器。

[0455] 根据其它方面,提供了如本文所定义的试剂盒用于诊断癌症的用途。根据其它方面,提供了如本文所定义的试剂盒用于诊断炎症性疾病的用途。根据其它方面,提供了如本文所定义的试剂盒用于诊断产前疾病的用途。

[0456] 根据其它方面,提供了在有需要的受试者中治疗疾病的方法,其中所述方法包括以下步骤:

[0457] (a)使从人或动物受试者获得的体液样品与结合转录因子的结合剂接触;

[0458] (b)检测、测量或测序与转录因子缔合的DNA片段;和

[0459] (c)使用DNA片段的存在、量或序列作为受试者中疾病存在的指示物;和

[0460] (d)如果在步骤(c)中确定受试者患有疾病,则施用治疗。

[0461] 在一个实施方案中,疾病是癌症。在备选的实施方案中,疾病是炎症性疾病。根据其它方面,提供了如本文所定义的试剂盒用于诊断妊娠受试者的胎儿的产前疾病的用途。

[0462] 在一个实施方案中,施用的治疗选自:手术、放射疗法、化学疗法、免疫疗法、激素疗法和生物疗法。

[0463] 根据本发明的其它方面,提供了在有需要的受试者中治疗癌症的方法,其中所述方法包括以下步骤:

[0464] (a)根据本文所述的方法检测或诊断受试者中的癌症;然后

[0465] (b)对所述个体施用抗癌症疗法、手术或药剂。

[0466] 在一个实施方案中,受试者是人或动物受试者。

[0467] 现在我们用以下实施例来说明本发明。

[0468] 实施例

[0469] 实施例1

[0470] 将定向特异性结合转录因子TTF-1(也称为NKX2-1)的抗体涂布在磁性珠上用于生物磁性分离(例如,商业上可获得的Dynabeads)。TTF-1是同源异型盒螺旋-转角-螺旋转录因子。

[0471] 将抗TTF-1抗体涂布的磁性珠加入到从诊断患有IV期肺癌、IV期甲状腺癌的人受试者和从健康受试者收集的EDTA血浆样品中。孵育(轻轻旋转以保持磁性颗粒的悬浮)后,从血浆样品中去除磁性颗粒,并用测定缓冲液洗涤。使用Qiagen QiaAMP循环核酸试剂盒从

磁性固相分离TTF-1缔合的DNA片段。通过Snyder等人,2016(其通过引用并入本文)中描述的文库方法,将接头寡核苷酸连接到分离的DNA片段,以生产对于每个血浆样品与TTF-1缔合的DNA序列的单链DNA文库。

[0472] 通过实时定量PCR扩增为每个受试者产生的片段文库。使用下一代测序方法对扩增的文库测序,并比较每个文库中的DNA的量和缔合的序列。在健康样品中,35-80bp范围的小cfDNA片段对TTF-1TFBS基因座的覆盖度低,因为在健康样品中TTF-1缔合的DNA的量低或不可检测。相反,在癌症样品中35-80bp范围的小cfDNA片段对TTF-1TFBS基因座的覆盖度将是高的,因为在患有IV期肺癌或IV期甲状腺癌的患者样品中TTF-1缔合的DNA的量较高。在甲状腺癌样品中测定的缔合的TTF-1DNA的序列将与甲状腺细胞中TTF-1调节的基因启动子的已知序列缔合。类似地,在肺癌样品中测定的缔合的TTF-1DNA的序列将与甲状腺细胞中TTF-1调节的基因启动子的已知序列相关联。在此基础上,大多数或所有健康样品、甲状腺癌样品和肺癌样品从通过实验产生的数据中可鉴定。

[0473] 实施例2

[0474] 重复实施例1中所述的实验,但在与涂布有抗TTF1抗体的磁性颗粒孵育之前,将涂布有抗核小体抗体的磁性珠加入到血浆样品中以预清除核小体和核小体结合的DNA片段的样品。在孵育(轻轻旋转以保持磁性颗粒的悬浮)之后,从血浆样品中去除磁性颗粒。然后如在实施例1中所述使用剩余样品完成实验,除了在健康样品中发现的DNA的背景水平甚至低于实施例1所述的之外,具有类似的结果。

[0475] 实施例3

[0476] 将抗TTF-1抗体涂布的磁性珠加入到从患有IV期肺癌、IV期甲状腺癌的人受试者和从健康受试者收集的EDTA血浆样品中。孵育(轻轻旋转以保持磁性颗粒的悬浮)后,从血浆样品中去除磁性颗粒,并用测定缓冲液洗涤。使用Qiagen QiaAMP循环核酸试剂盒从磁性固相提取TTF-1缔合的DNA片段。使用本领域已知的用于引物设计的典型软件设计特异性序列引物,以扩增与人基因组的SPB、促甲状腺激素受体和甲状腺过氧化物酶基因启动子区中的TTF-1结合位点缔合的特异性序列的DNA片段加上侧翼DNA。引物用于通过实时定量PCR扩增DNA片段。测量每个血浆样品中每个序列的存在的DNA的量。取自健康受试者的样品的结果将是低的或不可检测的。大多数取自肺癌患者的样品含有可检测量的SPB基因启动子序列DNA片段。大多数取自甲状腺癌患者的样品含有可检测量的促甲状腺激素受体和/或甲状腺过氧化物酶基因启动子序列DNA片段。在此基础上,大多数或所有健康样品、甲状腺癌样品和肺癌样品从通过实验产生的数据中可鉴定。

[0477] 实施例4

[0478] 重复实施例3中所述的实验,但是在与涂布有抗TTF-1抗体的磁性颗粒孵育之前,将涂布有抗核小体抗体的磁性珠加入到血浆样品中以预清除核小体和核小体结合的DNA片段的样品。在孵育(轻轻旋转以保持磁性颗粒的悬浮)之后,从血浆样品中去除磁性颗粒。然后如实施例3所述使用剩余样品完成实验,除了在健康样品中发现的DNA背景水平甚至低于实施例3所述的之外,具有类似的结果。

[0479] 实施例5

[0480] 通过在从健康男性和诊断患有IV期前列腺癌男性收集的血浆样品中测试,对螺旋-转角-螺旋转录因子NKX3.1重复与上述实施例所述的那些类似的实验。取自健康受试者

的样品的结果将是低的或不可检测的。取自前列腺癌患者的大多数样品含有可检测量的35-80bp大小范围的NKX3.1基因启动子序列DNA片段。在此基础上,大多数或所有健康样品和前列腺癌样品从通过实验产生的数据中可鉴定。

[0481] 实施例6

[0482] 通过在从健康女性和诊断患有IV期卵巢癌的女性收集的血清样品中测试,对锌指转录因子WT1重复与上述实施例中所述的那些类似的实验。取自健康受试者的样品的结果将是健康受试者中35-80bp大小范围的WT1缔合的cfDNA片段对WT1 TFBS基因座覆盖度低或不可检测。取自卵巢癌患者的大多数样品将显示35-80bp大小范围的WT1缔合的cfDNA片段对WT1 TFBS基因座覆盖度更高,因为它们含有可检测量的WT1基因启动子序列35-80bp cfDNA片段。在此基础上,大多数或所有健康样品和卵巢癌样品从通过实验产生的数据中可鉴定。

[0483] 实施例7

[0484] 我们用定向结合位于氨基酸30-33位的组蛋白H3表位的抗体涂布Dynabeads M280 甲苯磺酰基活化的磁性珠。该抗体选自许多测试的抗体,因为观察到它结合含有完整组蛋白尾的核小体和具有剪切的组蛋白尾的核小体两者。

[0485] 我们将抗H3抗体涂布的磁性珠(1mg)加入到含有一定浓度范围的购自Active Motif的重组单核小体(0.5ml)的溶液中。将珠与核小体在室温下孵育1小时,轻轻滚动管以维持珠悬浮。磁性分离珠并洗涤。然后通过洗脱去除吸附到珠的核小体,并通过蛋白质印迹分析。结果证明核小体以图3所示的剂量依赖性方式被磁性珠从溶液中吸附。

[0486] 实施例8

[0487] 如实施例7所述制备和使用抗H3抗体涂布的磁性珠。我们将抗H3抗体涂布的磁性珠以及未涂布的珠加入到8个人EDTA血浆样品以及含有一定浓度范围的重组单核小体的溶液中。选择重组单核小体浓度的范围以包括通常在人临床样品中观察到的水平。

[0488] 我们使用用于具有光密度(OD)读数的核小体的ELISA测试了在与磁性珠孵育后保留在溶液中的核小体的存在。图4中所示的结果证明,在用抗H3抗体涂布的磁性珠吸附后,保留在溶液中的重组单核小体的水平是不可检测的(具有与不含核小体的对照溶液类似的OD),而与未涂布的磁性珠孵育的溶液中的水平不受影响,导致正常的ELISA剂量应答曲线。类似地,在用抗H3抗体涂布的磁性珠吸附后,测试的8个人血浆样品中保留在溶液中的核小体的水平也是低的或不可检测的,但不受与未涂布的磁性珠孵育的影响。这些结果证明从人血浆样品中定量去除核小体。

[0489] 实施例9

[0490] 根据制造商的方案,不同颜色的Luminex珠涂布有定向结合转录因子TTF-1、NKX3.1、GATA-3、CDX-2和GRHL2的抗体。将取自健康受试者和诊断患有多种癌症的受试者的血浆样品与所有珠的混合物接触。通过PCR方法或通过下一代测序来测量覆盖与每个珠-结合的转录因子结合的相应转录因子TFBS的35-80bp范围的cfDNA的量或覆盖度。结果将显示,与涂布有定向结合NKX3.1和GRHL2的抗体的珠结合的35-80bp cfDNA的NKX3.1和GRHL2 TFBS覆盖度在取自前列腺癌患者的样品中升高,而与其它珠(涂布有抗TTF-1、GATA-3或CDX-2抗体)的转录因子结合低。类似地,在取自肺癌患者的样品中,与涂布有定向结合TTF-1和GRHL2的抗体的珠结合的短的35-80bp cfDNA片段的量将升高,而与其它珠(涂布有抗

NKX3.1、GATA-3或CDX-2抗体)的转录因子结合低。类似地,与涂布有定向结合GATA-3和GRHL2的抗体的珠结合的短35-80bp cfDNA片段的量在取自乳腺癌患者的样品中将升高,而与其它珠粒(涂布有抗TTF-1、NKX3.1或CDX-2抗体)的结合低。相反,在取自健康受试者的样品中,短35-80bp cfDNA片段与所有珠的结合将是低的。

[0491] 实施例10

[0492] 根据制造商的方案,磁性珠涂布有定向结合RNA聚合酶II的抗体。使取自健康受试者和诊断患有多种癌症的受试者的血浆样品与珠接触。洗涤珠以去除未结合的染色质片段。

[0493] 提取与珠结合的DNA,与接头寡核苷酸连接,并且对文库测序以发现存在于受试者的样品中的活性基因的组。结果显示,存在于取自健康受试者的样品中的活性基因是在造血细胞中具有活性的基因的代表。在取自患有癌症的患者的样品中也存在相同的序列,但是发现这些样品另外含有RNA聚合酶II结合的DNA序列,其代表在造血细胞中无活性但在疾病组织的细胞中有活性的基因,包括在相关组织的(健康或患病)细胞中通常有活性的和/或在癌细胞中上调的基因。

[0494] 实施例11

[0495] 根据制造商的方案,磁性珠涂布有定向结合RNA聚合酶II的抗体。使取自健康受试者和诊断患有多种癌症的受试者的血浆样品与珠接触。洗涤珠以去除未结合的染色质片段。

[0496] 使用PCR引物扩增序列,分析与珠结合的DNA中特定DNA序列的存在。选择待分析的序列以与结肠直肠癌特异性结合。结果将显示该序列存在于取自患有结肠直肠癌的受试者的样品中,但不存在于取自健康受试者或取自患有其它癌症的受试者的样品中。

[0497] 实施例12

[0498] 从6名诊断患有卵巢癌的女性、2名诊断患有ER阴性乳腺癌的女性和8名诊断患有ER阳性乳腺癌的女性收集EDTA血浆样品,其中4名女性诊断ER评分为7,并且4名女性诊断ER评分为8。使用商业ER α ELISA试剂盒测定EDTA血浆样品的ER α 。所用的ELISA试剂盒的定量检测范围为3-200pg/ml,其中ER α 的检测下限为0.8pg/ml。ER α 的平均测量水平对于ER阴性受试者是低的,而对于诊断患有卵巢癌或ER阳性乳腺癌的受试者是较高的。此外,对于诊断患有ER阳性乳腺癌的受试者,测量的平均水平对于ER评分较高的那些女性较高(图5)。我们得出结论,从取自女性的全血样品制备的EDTA血浆样品中ER α 的存在可用作妇科疾病(包括妇科癌症)的生物标志物。

[0499] 实施例13

[0500] 乳腺癌的孕酮受体状态为PR阳性或PR阴性在诊断和治疗妇科癌症中同样重要。我们进一步得出结论,在从取自女性的全血样品制备的EDTA血浆样品中孕酮受体(PR)水平的测量类似地可用作妇科疾病(包括妇科癌症)的生物标志物。

[0501] 实施例14

[0502] 前列腺癌的雄性激素受体状态在前列腺癌的诊断和治疗中同样重要。我们进一步推断,由取自男性的全血样品制备的EDTA血浆样品中的雄性激素受体(AR)水平的测量可用作包括前列腺癌的前列腺疾病的生物标志物。

[0503] 实施例15

[0504] 通过使用考马斯蓝染色进行显影的蛋白质印迹评估从对小鼠IgG涂布的磁性颗粒非特异(非特异性)的血浆样品吸附的蛋白质的背景水平。在用含有0.1% Tween 20洗涤剂的典型免疫化学洗涤缓冲液或用含有高水平的1.2%的洗涤剂混合物(包含1%辛基苯氧基聚乙氧基乙醇洗涤剂、0.1%脱氧胆酸钠和0.1%十二烷基硫酸钠)的洗涤缓冲液5次洗涤颗粒后,评估背景。结果(图6)显示通过使用强洗涤剂,背景染色大大降低。

[0505] 对特异性吸附到小鼠抗聚ADP抗体(其结合任何大小的parlylated蛋白质)的蛋白质施用相同的实验。在这种情况下,染色受到较少的影响,这显示洗涤去除了非特异性结合的蛋白质,但不影响(或具有较少的影响)附着于抗体的特异性结合的蛋白质。

[0506] 实施例16

[0507] 我们使用标准方法将定向特异性结合转录因子CTCF的单克隆抗体涂布在磁性珠(MyOne Tosyls Activated Dynabeads™)上。简言之,将0.86mg单克隆抗体与29mg磁性珠(30 μg抗体/mg珠)在滚动瓶中于37°C下在含有1M硫酸铵的2.9ml 0.1M硼酸盐缓冲液pH 9.5中孵育18小时以维持珠的悬浮。沉降珠,并倾析上清液。将珠重悬,并在37°C下在含有0.1% Tween 20和1%牛血清白蛋白(BSA)的2.9mL磷酸盐缓冲盐水pH7.4的封闭缓冲液(PBS)中孵育1小时。然后沉降珠,用含有0.1% Tween 20和1% BSA的3mL PBS洗涤两次,并储存在含有0.1% Tween 20、1% BSA和防腐剂的2.9mL PBS中。类似地,将非特异性小鼠IgG涂布到磁性珠作为非特异性对照试剂。

[0508] 在从癌症患者获得的4个汇集的交联的EDTA血浆样品(1.6mL,在Streck Cell-Free DNA BCT中收集)中进行CTCF-DNA片段的染色质免疫沉淀(ChIP)。用0.4mL商业上可获得的放射免疫沉淀测定缓冲液稀释每个汇集的样品,并加入1mg抗CTCF涂布的磁性颗粒。将混合物在室温下滚动孵育1小时以保持珠的悬浮。然后沉降珠,并用含有1% Triton X-100洗涤剂、0.1%脱氧胆酸钠和0.1%十二烷基硫酸钠的混合物的强洗涤剂洗涤溶液洗涤5次,并储存在0.1mL缓冲液中。平行地,通过将1.6mL的每种汇集的血浆样品与非特异性小鼠IgG涂布的磁性珠一起孵育来进行对照实验。

[0509] 在将磁性颗粒与汇集的血浆样品一起孵育后,将磁性颗粒结合的蛋白质悬浮在变性的1%十二烷基硫酸钠(SDS)缓冲液中,并使用抗CTCF抗体和标记的抗小鼠抗体通过蛋白质印迹分析变性的蛋白质用于检测。在蛋白质印迹实验中,由130-140kD的带的存在指示CTCF的存在(Klenova等人,1997)。蛋白质印迹分析的结果示于图7中。简言之,当暴露于涂布有抗CTCF抗体(抗CTCF)的磁性颗粒时,在约140kD的对应于CTCF转录因子的存在的蛋白质带对于所有4个样品是可见的。相反,暴露于涂布有非特异性小鼠IgG(NS-IgG)的磁性颗粒的相同的4个样品中的任何一个都看不到带。这指示采用的ChIP方法能够从所有4个测试的汇集的样品中选择性分离循环转录因子CTCF。它还证明由所采用的洗涤方案产生的干净的背景。

[0510] 实施例17

[0511] CTCF是锌指转录因子。在从诊断患有乳腺癌的受试者获得的交联的EDTA血浆样品(2.4mL,在Streck Cell-Free DNA BCT中收集)中进行CTCF-DNA片段的染色质免疫沉淀(ChIP)。如以上在实施例16中所述进行ChIP,不同的是用0.6mL放射免疫沉淀测定缓冲液稀释2.4mL样品,并加入1.5mg抗CTCF涂布的磁性颗粒。在平行对照实验中,将2.4mL交联的EDTA血浆样品与涂布有非特异性小鼠IgG的磁性珠一起孵育。将磁性珠分成2个级分。一个

级分用于通过蛋白质印迹分析,蛋白质印迹使用来自MCF7乳腺癌细胞的片段化的染色质作为阳性对照证实珠上CTCF蛋白的存在。

[0512] 第二级分(测试和对照)珠用于DNA提取和分析。磁性珠缔合的染色质片段与缔合的DNA的交联通过在95°C下加热15分钟而逆转。然后使用商业上可获得的DNA提取试剂盒(Qiagen QIAamp DSP循环NA试剂盒)根据制造商的说明书提取与磁性珠缔合的DNA。

[0513] 使用商业上可获得的试剂盒(Claret Bio SRSLYNGS Library Prep Kit),根据制造商的说明书,使用16个扩增循环扩增提取的cfDNA,以产生用于测序的单链文库。使用Bioanalyzer仪器通过电泳分析扩增的测试和非特异性cfDNA片段文库。结果(图8)显示从特异性抗CTCF涂布的磁性颗粒获得的扩增的cfDNA文库含有35-80bp范围的小片段。注意到电泳图谱中约140bp处的尖峰代表接头二聚体,因此175-220bp的接头连接的片段代表35-80bp的cfDNA片段。在约190bp处观察到接头连接的cfDNA片段的主峰,其对应于长度为约50bp的cfDNA片段。尽管扩增的cfDNA文库含有35-80bp范围的小片段,不是所有这些片段都与样品中的CTCF结合,因为从涂布有非特异性小鼠IgG的固体支持物中扩增的提取物也获得了小DNA片段。然而,用特异性抗CTCF抗体ChIP(1000荧光单位[FU])获得的特异性峰高于非特异性IgG峰(80FU)。将该样品送去测序。

[0514] 实施例18

[0515] 通过如上在实施例17中所述的抗CTCF免疫沉淀,从诊断患有结肠直肠癌(CRC)的患者收集的交联的EDTA血浆样品(在Streck cfDNA BCT中收集)制备扩增的cfDNA文库。通过下一代Illumina NovaSeq测序对使用抗CTCF免疫沉淀分离的扩增的cfDNA文库测序。

[0516] 使用Illumina DRAGEN Bioinformatics流水线(<https://emea.illumina.com/products/by-type/informatics-products/dragen-bio-it-platform.html>)将每个代表cfDNA片段的测序的读数与人参考基因组GRCh38/hg38比对。丢弃任何未比对的读数。使用序列比对/图谱SAMTools(Li等人,2009),将所得比对BAM文件用于产生不同片段大小(35-80bp、135-155bp和156-180bp)的子集。使用1bp的二进制大小(可能的最高分辨率)计算读数覆盖度(发现覆盖特定基因座的片段的数目)。使用deepTools bamCoverage,将读数覆盖度标准化为用RPGC绘制到人基因组的读数(每个基因组覆盖度的读数)的总数。使用deepTools plotProfile(Ram-Rez等人,2016)为每个片段大小产生覆盖度分布图(图9和10)。

[0517] 在图9(a)中显示与较长的cfDNA片段的覆盖度相比,与CTCF缔合的短的35-80bp片段在9780个公开的CTCF结合位点的基因座处的覆盖度的结果(Kelly等人,2012),该较长的cfDNA片段与循环单核小体缔合预期的大小(135-155bp和156-180bp)一致。在5000bp范围显示覆盖度,包括CTCF结合位点位置上游和下游的2500个碱基。我们观察到小的35-80bp cfDNA片段结合的覆盖度的强峰精确地在Kelly等人,2012年报道的CTCF TFBS基因座的基因组位置处结合。因为测序的文库直接从附着于在具有低背景抗CTCF涂布的磁性珠上分离的CTCF蛋白的cfDNA产生,cfDNA文库含有很少的核小体并且核小体定位信号低。该特征产生清晰的35-80bp信号,并消除对混合样品(例如含有源自造血和癌组织的混合cfDNA片段的样品)中竞争信号去卷积的需要。相反,结合非特异性小鼠IgG获得的cfDNA文库在CTCF TFBS基因座处未显示峰(图9(b))。

[0518] 大量蛋白质可以结合至或接近TFBS,包括转录因子,或多种协同结合转录因子、转

录增强子、阻抑物或其它调节蛋白的任意组合。本发明的方法的主要优点是已知CTCF TFBS基因座的小cfDNA片段覆盖度仅涉及与CTCF缔合的cfDNA片段。与本领域的方法相反,例如Snyder等人,2016和Ulz等人,2019的片段组学方法对从EDTA血浆中提取的所有大小的所有cfDNA片段作图,并推断在任何特定的基因组位置上是否发生蛋白质结合。不能知道涉及哪种蛋白质,因为所有这样的方法的第一步是提取cfDNA,其需要样品中的所有核蛋白染色质片段(包括核小体和转录因子-DNA复合物)的解离,并因此破坏将任何特定cfDNA序列与任何特定转录因子或其它蛋白质连接的任何直接信息。

[0519] 参考输入非特异性对照,cfDNA片段序列的峰值识别导致CTCF作为具有最多的TFBS序列片段的转录因子。使用MACS2(Zhang等人,2008)窄峰对BAM文件进行峰值识别。峰文件用于使用来自Homer软件包(Heinz等人,2010)的findMotifGeneome工具检测转录因子结合位点。

[0520] 然后我们重复永生化的癌细胞中占据的1041个CTCF TFBS的富集的分析(Liu等人,2017)。图10(a)所示的结果显示,存在明显的35-80bp CTCF缔合的cfDNA片段峰,其结合1041个癌症特异性CTCF TFBS序列。与片段组学不同,有助于分析的cfDNA片段仅源自CTCF-DNA复合物,并且如果它们不包括CTCF,则不源自其它转录因子-DNA或辅因子-DNA复合物。这证明癌症特异性基因座的CTCF占据,并因此也指示那些cfDNA片段和它们所衍生的CTCF-DNA复合物的肿瘤细胞起源。更长的(核小体大小)cfDNA片段没有峰。结合非特异性小鼠IgG获得的cfDNA文库未显示峰(图10(b))。

[0521] 通过ChIP-Seq证明CTCF缔合的cfDNA片段结合体液中癌症特异性TFBS基因座指示所研究的受试者中癌症疾病存在,并且可以以这种方式用作生物标志物。我们得出结论,本发明的方法成功用于血浆中转录因子的ChIP-Seq并作为疾病的生物标志物。

[0522] 实施例19

[0523] 雄性激素受体(AR)是前列腺癌中目的锌指转录因子。我们将实施例17中对CTCF所述的相同的方法应用于AR。我们使用小鼠抗AR抗体从来自8个诊断患有前列腺癌的受试者的交联的EDTA血浆样品(在Streck cfDNA BCT中收集)免疫沉淀AR。我们使用来自LnCAP前列腺癌细胞系细胞的AR作为阳性对照,在固相支持物上进行通过ChIP分离的蛋白质的蛋白质印迹分析。图11的结果显示,在所有8个样品中都存在对应于分子量为约100kD的AR的蛋白质带,并且在2个样品中是高水平(图11的泳道2和3)。约50kD的带对应于标记的抗小鼠IgG抗体与用于ChIP的小鼠抗AR抗体的重链的结合。然后,我们从固相支持物提取DNA,将提取的DNA片段与接头寡核苷酸连接,并扩增存在的DNA。结果(图12)显示,对于所有8个样品,扩增的cfDNA文库都含有35-80bp范围的小片段(175-220bp的接头连接的片段)。尽管扩增的cfDNA文库含有35-80bp范围的小片段,不是所有这些片段都与样品中的AR结合,因为从涂布有非特异性小鼠IgG的固体支持物的扩增的提取物也获得小DNA片段。然后通过下一代测序对通过蛋白质印迹具有最高观察到的AR水平的2个样品获得的扩增cfDNA文库测序。

[0524] 参考文献

[0525] Active Motif,Nat.Methods 3:658(2006),doi:10.1038/NMETH907

[0526] Bohinski等人Molecular and Cellular Biology,14(9):5671(1994)

[0527] Corces等人Science,362(6413):eaav1898(2018),doi:10.1126/science.aav1898。

- [0528] Crowley等人*Nat.Rev.Clin.Oncol.*10:472-484(2013),doi:10.1038/nrclinonc.2013.110
- [0529] Darnell,*Nat.Rev.Cancer* 2:740-749(2002),doi:10.1038/nrc906
- [0530] Deligezer等人*Clinical Chemistry* 54:71125-1131(2008)
- [0531] Dunbar,*Clinica Chimica Acta* 363(1-2):71-82(2006),doi.org/10.1016/j.cccn.2005.06.023
- [0532] Gurel等人*Am J Surg Pathol*,34(8):1097-105(2010),doi:10.1097/PAS.0b013e3181e6cbf3.
- [0533] Heinz等人*Mol.Cell* 38(4):576-89(2010),doi:10.1016/j.molcel.2010.05.004。
- [0534] Holdenrieder&Stieber,*Crit.Rev.Clin.Lab.Sci.*46(1):1-24(2009),doi:10.1080/10408360802485875
- [0535] Hu等人*J.Trans.Med.*17:124(2019),doi:10.1186/s12967-019-1871-x
- [0536] Jung等人*Clin.Chim.Acta* 411(21-22):1611-24(2010),doi:10.1016/j.cca.2010.07.032
- [0537] Kelly等人*Genome Res.*22:2497-2506(2012),doi:10.1101/gr.143008.112。
- [0538] Klenova等人*Nucleic Acids Res.*25(3):466-473(1997),doi.org/10.1093/nar/25.3.466
- [0539] Lambert等人*Cell* 172(4):650-665(2018),doi:10.1016/j.cell.2018.01.029
- [0540] Latil等人*Cell Stem Cell* 20(2):191-204.e5(2017),doi:10.1016/j.stem.2016.10.018。
- [0541] Lee等人*J.Mol.Med. (Berl)* .85(12):1393-404(2007),doi:10.1007/s00109-007-0237-7
- [0542] Li等人*Bioinformatics* 25(16):2078-2079(2009),doi:10.1093/bioinformatics/btp352
- [0543] Lin等人*PLoS Genet.*3(6):e87(2007),
- [0544] doi:10.1371/journal.pgen.0030087.eor
- [0545] Liu等人*Oncotarget* 8(69):114183-114194(2017),doi:10.18632/oncotarget.23172
- [0546] Liu等人*EBioMedicine* 41:345-356(2019),doi:10.1016/j.ebiom.2019.02.010
- [0547] Maenhaut等人2015In:Feingold,Anawalt,Boyce,等人编辑Endotext.<https://www.ncbi.nlm.nih.gov/books/NBK285554/>
- [0548] Mann等人*Curr.Top Dev.Biol.*88:63-101(2009),doi:10.1016/S0070-2153(09)88003-4。
- [0549] Mansson等人*Mol.Oncol.*15(11):2868-2876(2021),doi:10.1002/1878-0261.13093
- [0550] Matys等人*Nucleic Acids Res.*34:D108-D110(2006),doi:10.1093/nar/gkj143
- [0551] Merabet和Mann,*Trends Genet.*32(6):334-347(2016),doi:10.1016/j.tig.2016.03.004。

- [0552] Newman等人*Nat.Med.*20(5):548-54(2014),doi:10.1038/nm.3519
- [0553] Park等人*Oncol.Lett.*3(4):921-926(2012),doi:10.3892/ol.2012.592
- [0554] Pomerantz等人*Nat.Genet.*47(11):1346-51(2015),doi:10.1038/ng.3419。
- [0555] Poorey等人*Science* 342(6156):369-72(2013),doi:10.1126/science.1242369。
- [0556] Ramírez等人*Nucleic Acids Res.*44(W1):W160-5(2016),doi:10.1093/nar/gkw257
- [0557] Ralston,Do transcription factors actuallybindDNA?DNA footprinting and gel shift assays.*Nature Education* 1(1):121(2008)
- [0558] Sadeh等人*Nat.Biotechnol.*39:586-598(2021),doi.org/10.1038/s41587-020-00775-6
- [0559] Sanchez等人*NPJ Genom.Med.*3:31(2018),doi:10.1038/s41525-018-0069-0
- [0560] Skene和Henikoff,*eLife* 6:e21856(2017),doi:10.7554/eLife.21856.002
- [0561] Snyder等人*Cell* 164(1-2):57-68(2016),
- [0562] doi:10.1016/j.cell.2015.11.050
- [0563] Ulz等人*Nat.Commun.*10(1):4666(2019),doi:10.1038/s41467-019-12714-4
- [0564]
- [0565] Vad-Nielsen等人*Lung Cancer* 147:P244-251(2020),doi.org/10.1016/j.lungcan.2020.07.023
- [0566] Vaquerizas等人*Nat.Rev.Genet.*10(4):252-63(2009),doi:10.1038/nrg2538
- [0567] Wang等人*Genome Res.*22(9):1680-8(2012),doi:10.1101/gr.136101.111
- [0568] Zhang等人*Genome Biol.*9(9):R137(2008),doi:10.1186/gb-2008-9-9-r137
- [0569] Zhou等人*BMC Genomics* 18(1):724(2017),doi:10.1186/s12864-017-4115-6。

序列表

<110>	比利时意志有限责任公司	
<120>	循环转录因子分析	
<130>	VOL-C-P2959PCT	
<150>	63/131,722	
<151>	2020-12-29	
<160>	4	
<170>	PatentIn版本3.5	
<210>	1	
<211>	9	
<212>	DNA	
<213>	智人	
<220>		
<221>	misc_特征	
<222>	(3) .. (3)	
<223>	n为a、c、g或t	
<220>		
<221>	misc_特征	
<222>	(6) .. (7)	
<223>	n为a、c、g或t	
<400>	1	
	gcncctnnag	9
<210>	2	
<211>	55	
<212>	DNA	
<213>	智人	
<400>	2	
	gatcaagcac ctggagggt cttcagagca aagacaaaca ctgaggtcgc tgcca	55
<210>	3	
<211>	20	
<212>	DNA	
<213>	智人	
<400>	3	
	tggccacacg agtgccctca	20
<210>	4	
<211>	120	
<212>	DNA	
<213>	智人	

<400> 4

cccaccccgt tctgttcccc cacagtttag acaagatcct catgctccac tggccacacg 60

agtgccctca ggaggagtag acacaggtgg aggagctcc ttttgaccag cagagaaaac 120

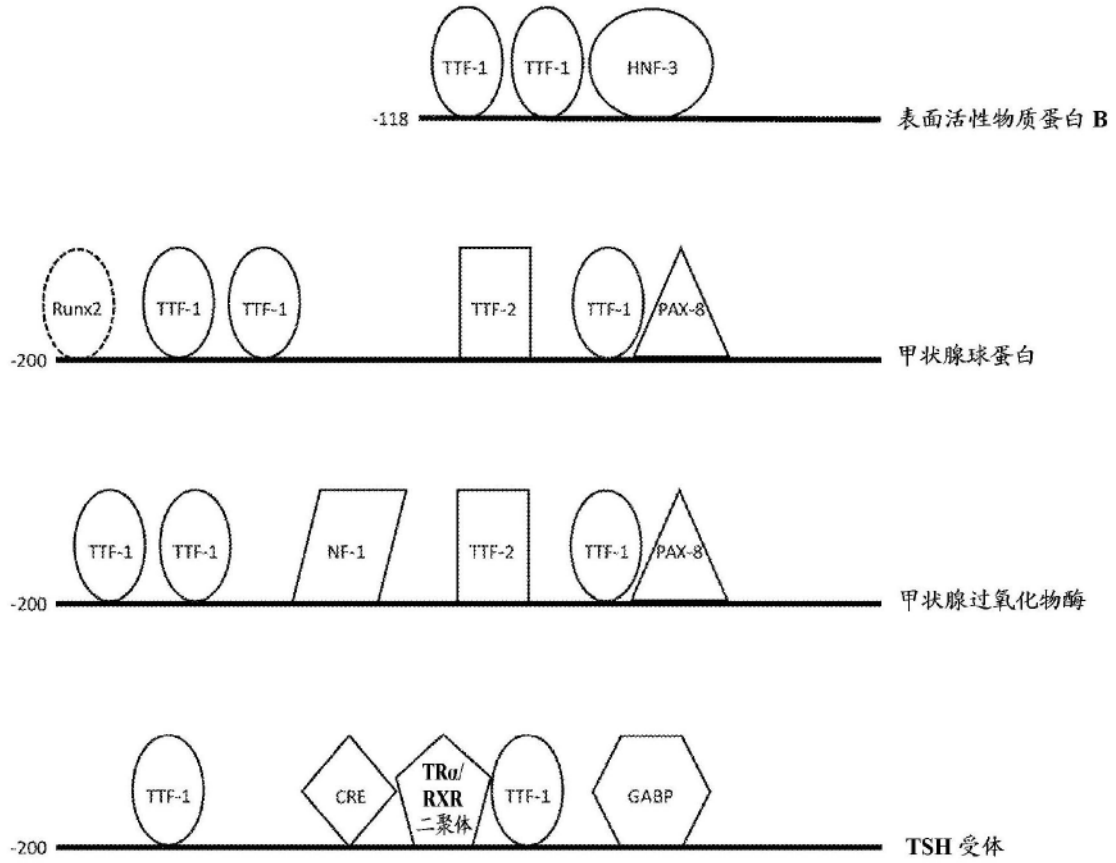


图1

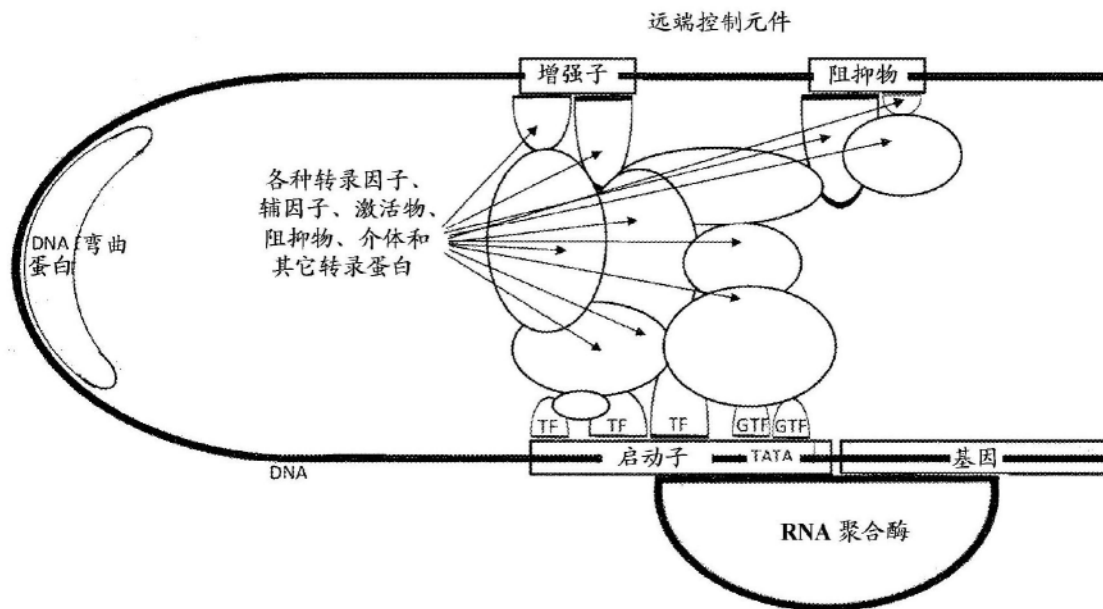


图2

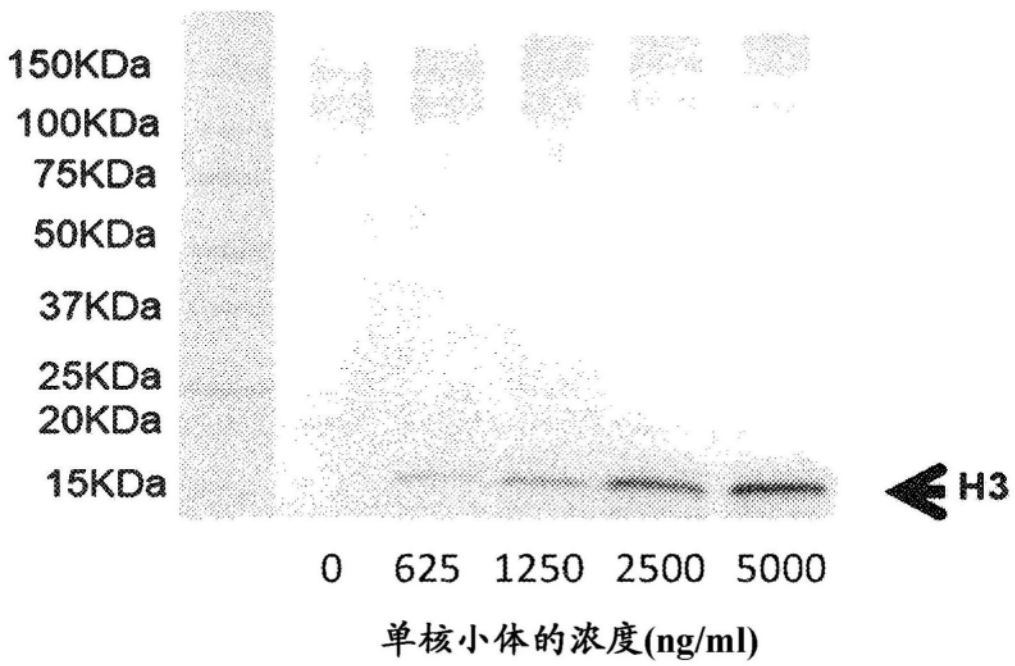
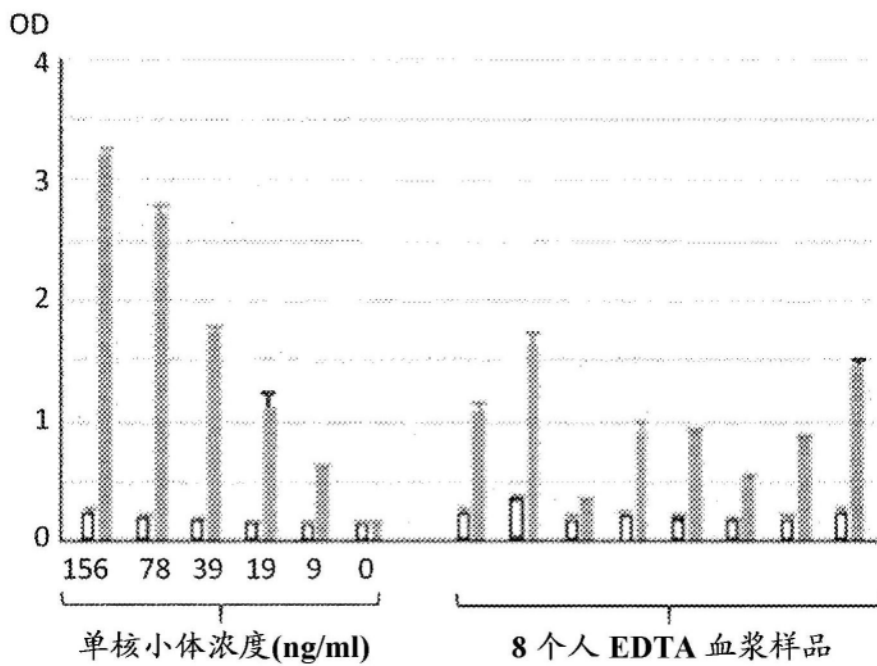


图3



□ 抗 H3 抗体涂布的磁性珠
■ 未涂布的颗粒

图4

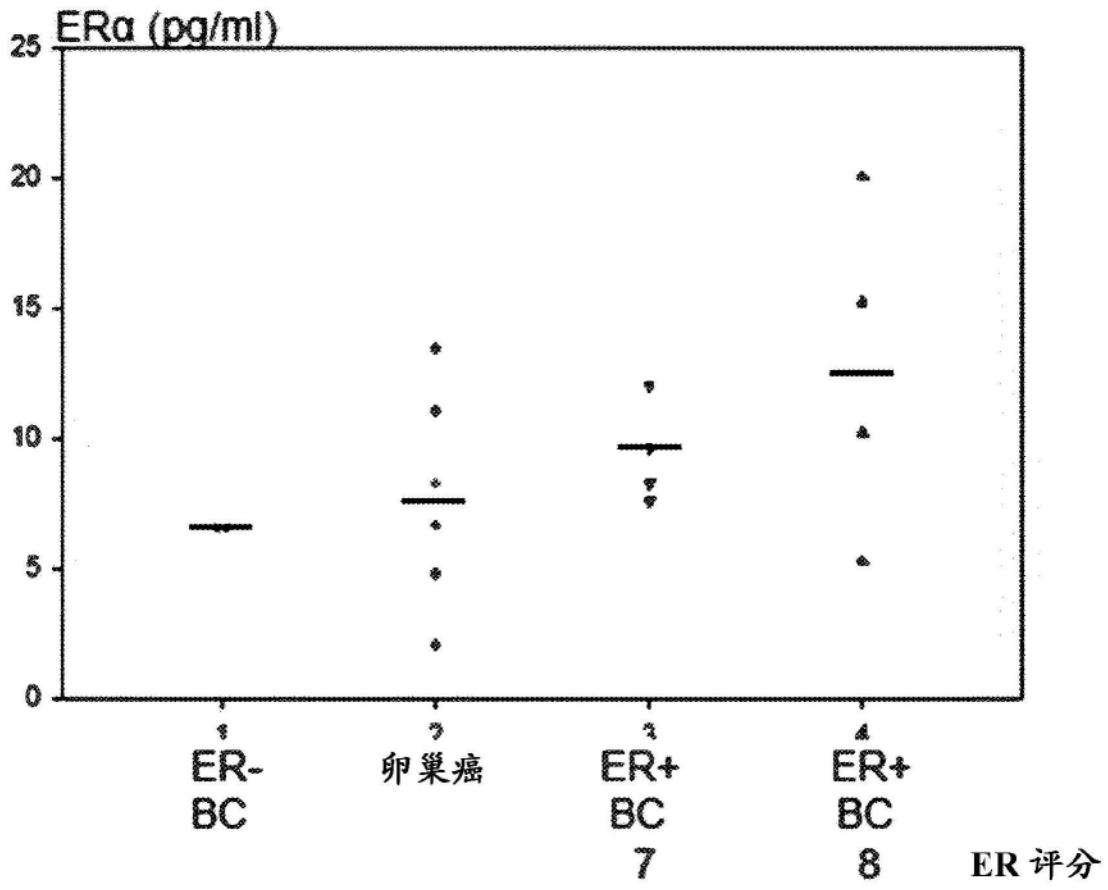


图5

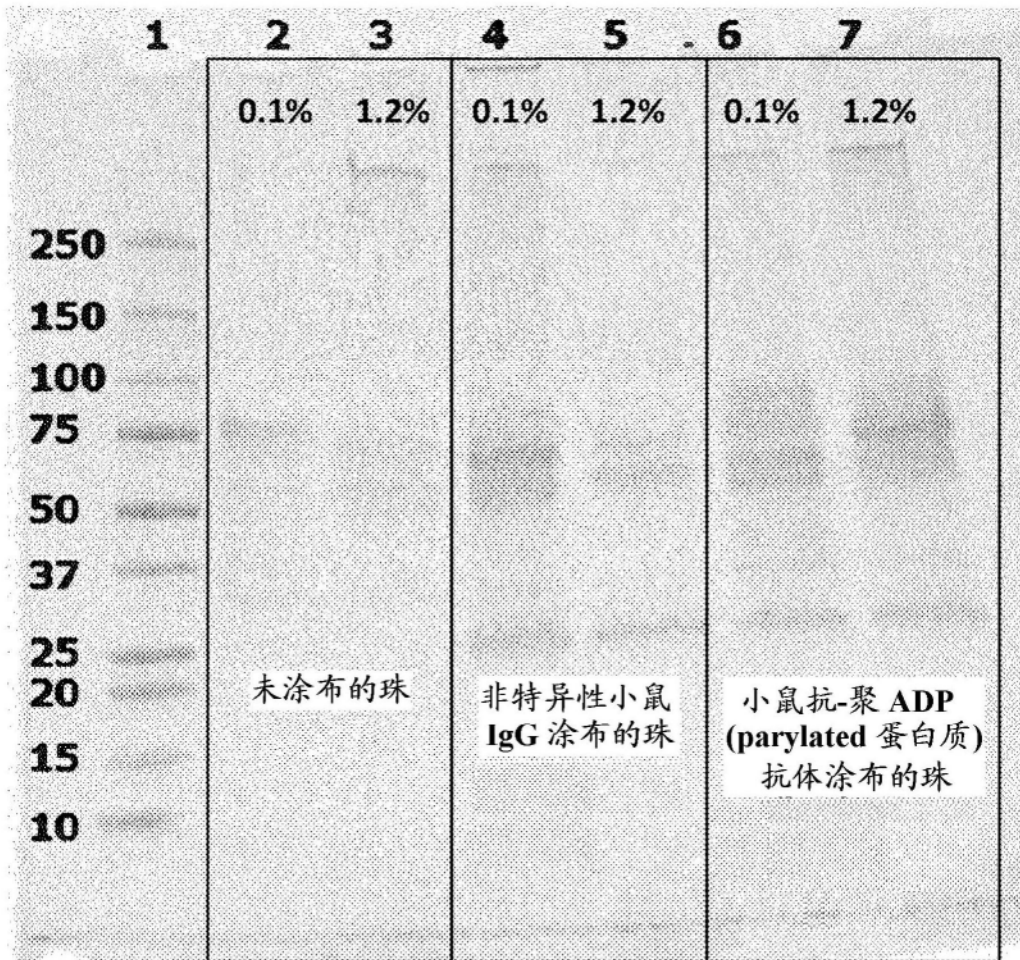


图6

1	样品 1		样品 2		样品 3		样品 4	
阶段	NS-IgG	抗 CTCF	NS-IgG	抗 CTCF	NS-IgG	抗 CTCF	NS-IgG	抗 CTCF

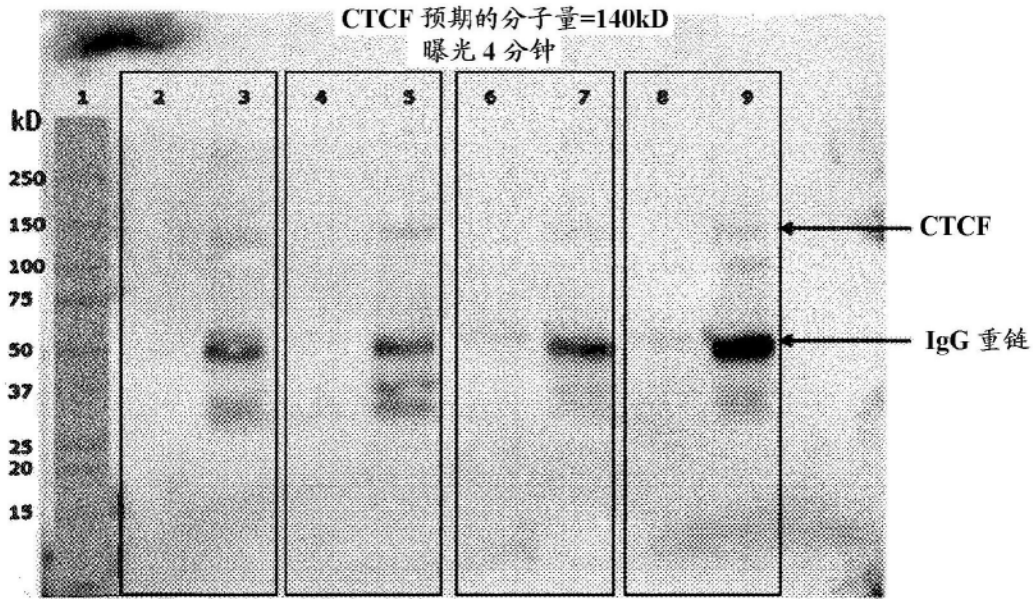


图7

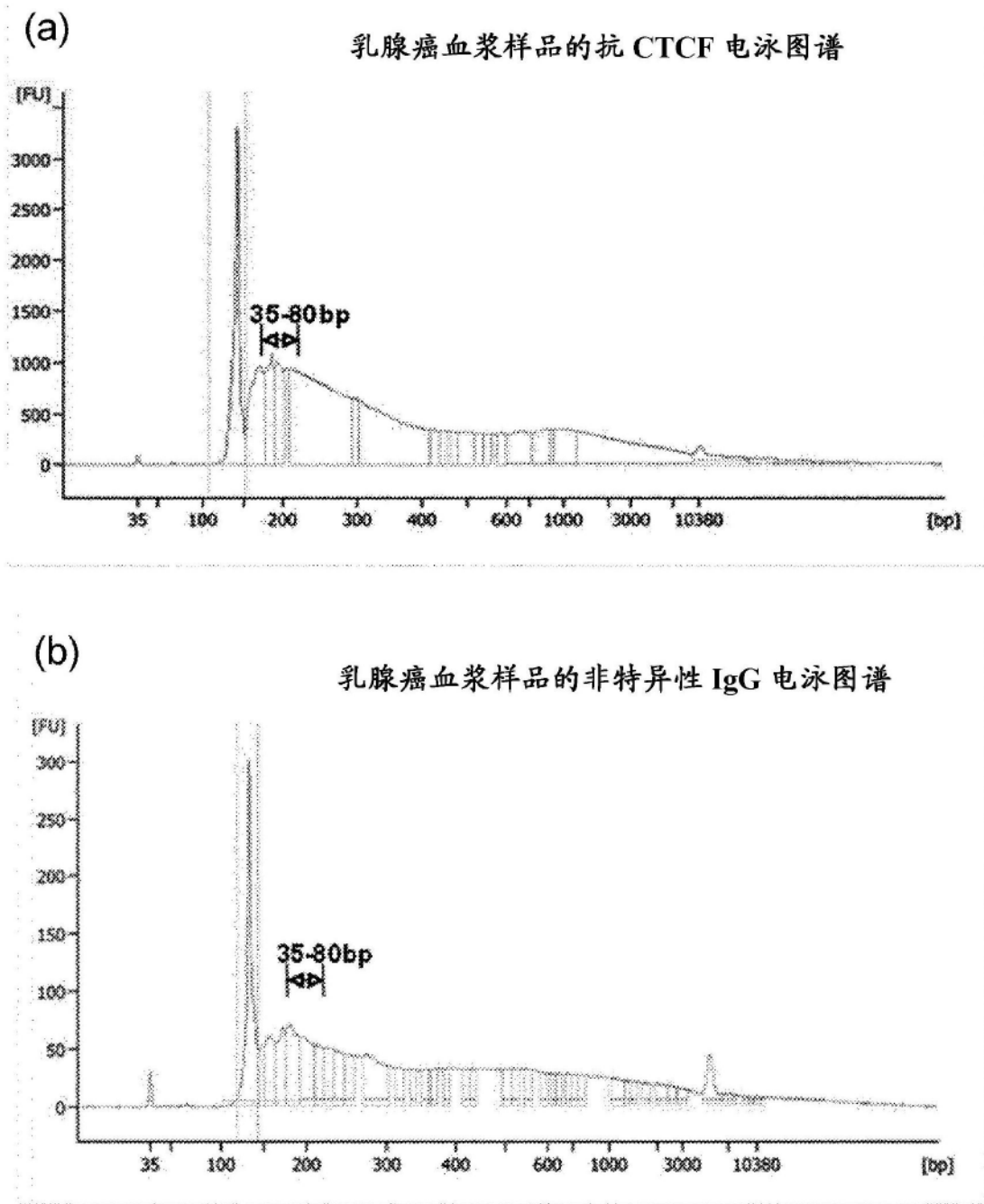


图8

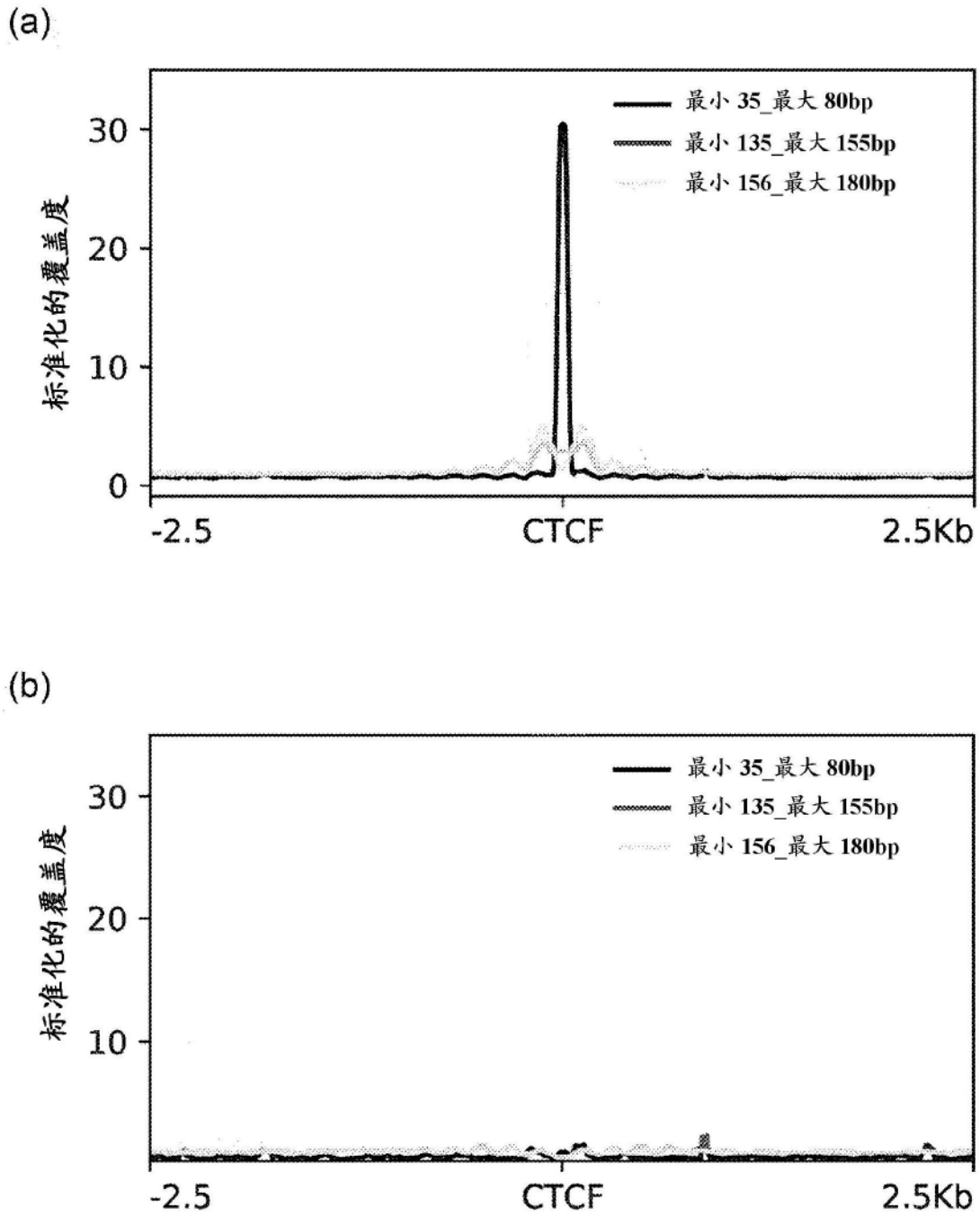


图9

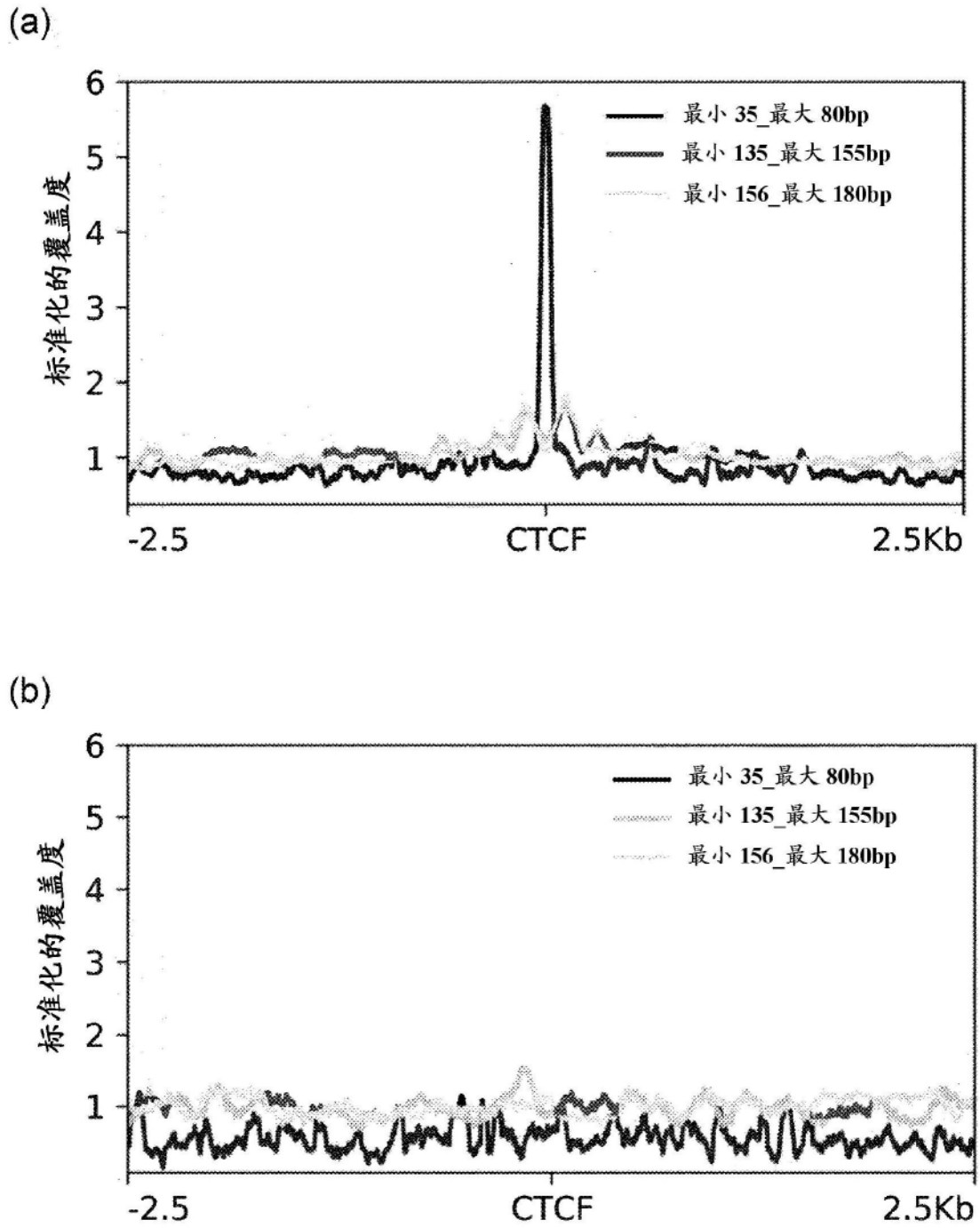


图10

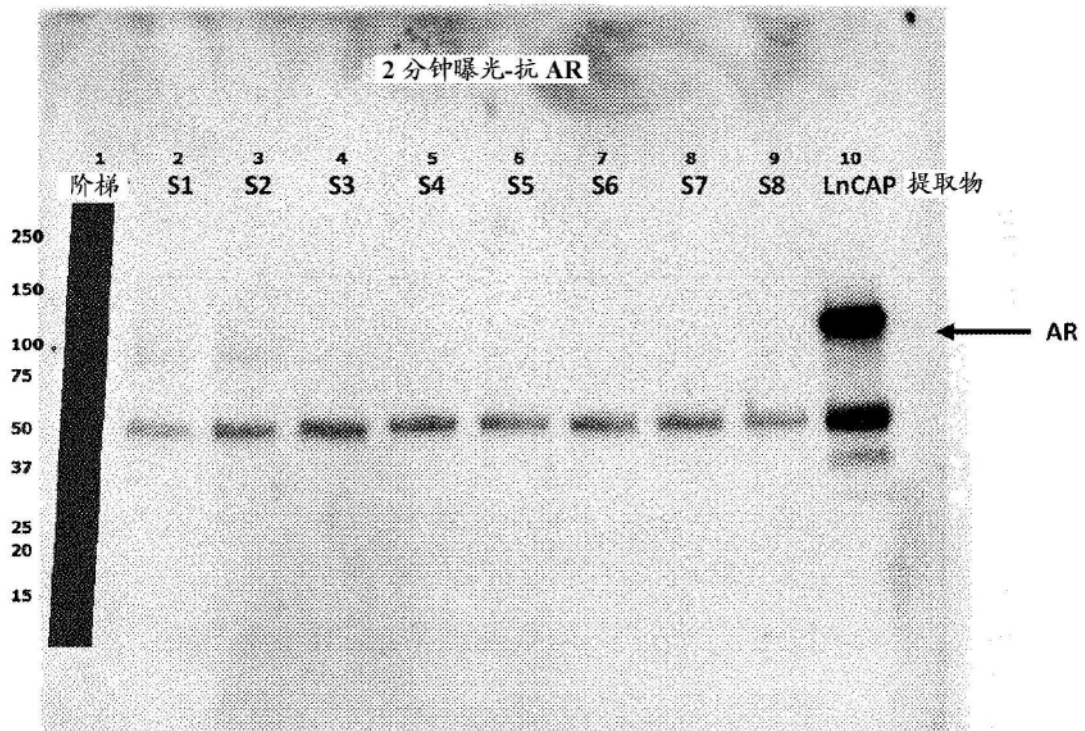


图11

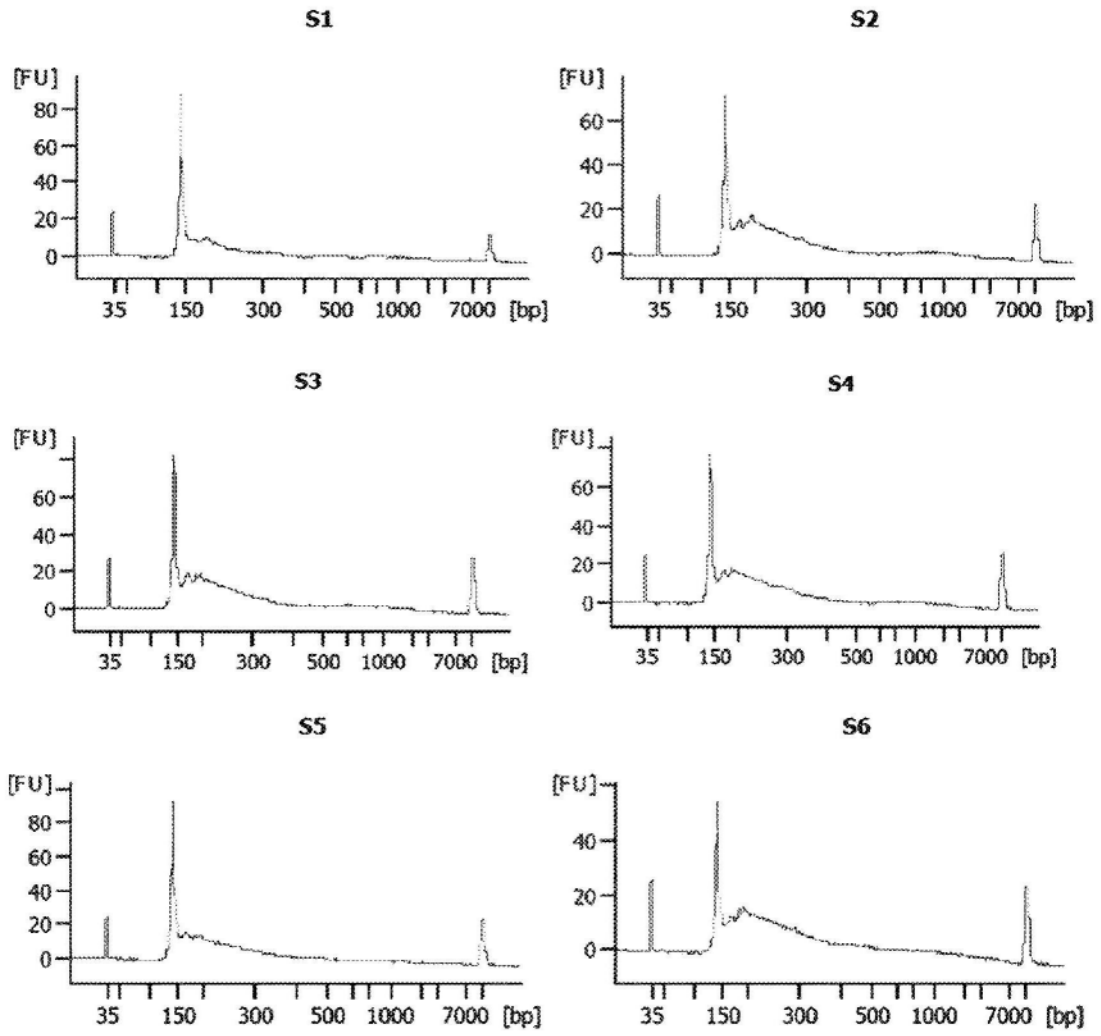


图12

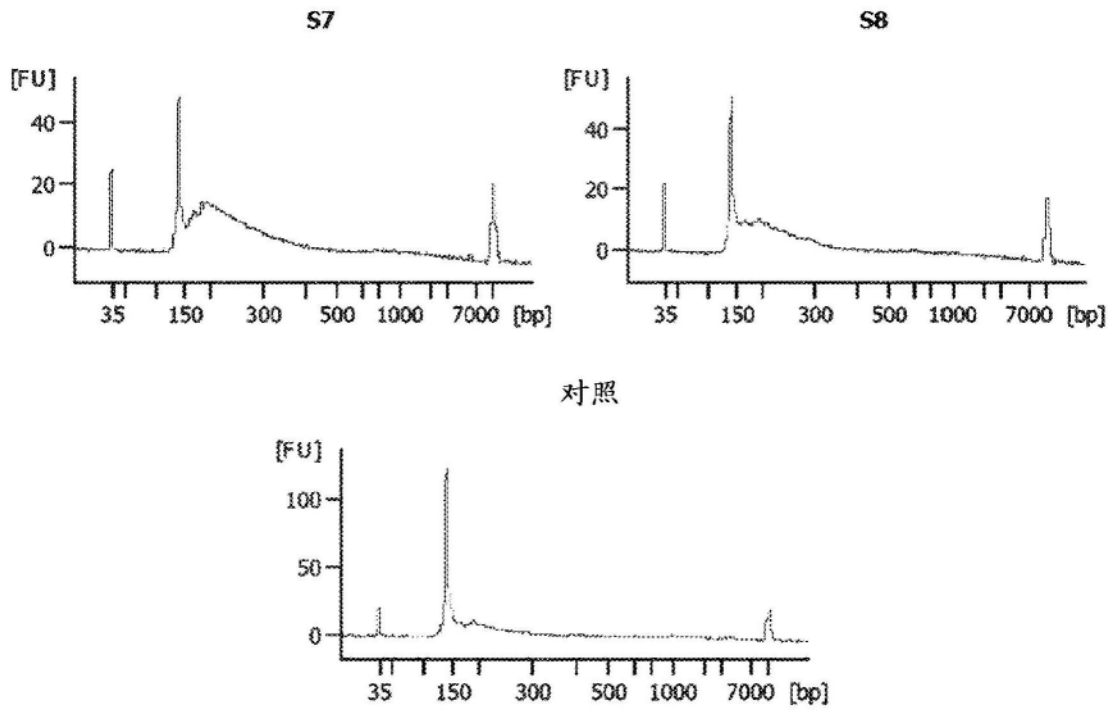


图12(续)