



(12) 发明专利申请

(10) 申请公布号 CN 115698321 A

(43) 申请公布日 2023. 02. 03

(21) 申请号 202180039805.5

大卫·伯克哈特

(22) 申请日 2021.06.16

(74) 专利代理机构 上海翼胜专利商标事务所

(30) 优先权数据

(普通合伙) 31218

63/039,886 2020.06.16 US

专利代理师 翟羽

(85) PCT国际申请进入国家阶段日

(51) Int.Cl.

2022.12.02

C12Q 1/6809 (2006.01)

(86) PCT国际申请的申请数据

C12Q 1/6886 (2006.01)

PCT/US2021/037668 2021.06.16

(87) PCT国际申请的公布数据

W02021/257729 EN 2021.12.23

(71) 申请人 格里尔公司

地址 美国加利福尼亚州门洛帕克市

(72) 发明人 露丝·E·蒙兹 马修·拉森

阿卡纳·雪诺伊

阿拉什·詹姆席狄

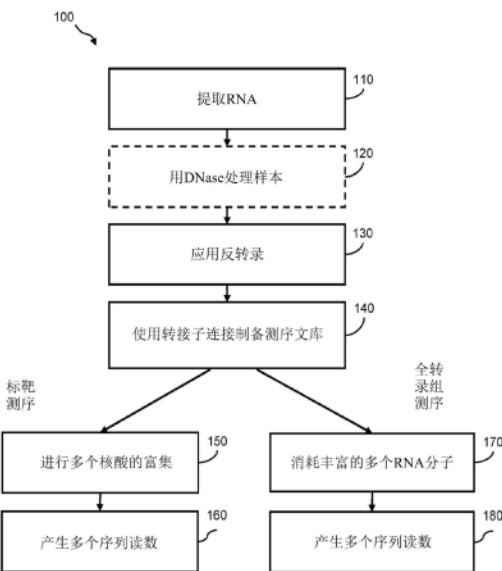
权利要求书3页 说明书74页 附图42页

(54) 发明名称

用于血浆游离RNA分析的多种方法

(57) 摘要

提供了测量多个血浆游离 (cell-free) 核糖核酸 (RNA) 分子的多个亚群的多种方法。在一些实施例中,提供了从获自一受试者的一测试样本中的多个RNA分子生成一测序文库的多种方法,以及用于分析所述测序文库以检测,例如一疾病的所述存在或不存在的多种方法。



1. 一种检测一受试者中癌症的方法,其特征在于:所述方法包括以下步骤:
 - (a) 测量所述受试者的一样本中的多个靶血浆游离RNA分子,其中所述多个靶血浆游离RNA分子选自表11中的一个或多个转录物;及
 - (b) 检测所述癌症,其中检测所述癌症包括检测高于一阈值水平的所述多个靶血浆游离RNA分子中的一个或多个。
2. 如权利要求1所述的方法,其特征在于:所述多个靶血浆游离RNA分子选自表8或12-15中的一者或多者。
3. 如权利要求1所述的方法,其特征在于:所述多个靶血浆游离RNA分子选自来自表8或11-14中的一者或多者的至少5、10、15或20个基因的多个转录物。
4. 如权利要求1所述的方法,其特征在于:所述多个靶血浆游离RNA分子包括来自以下的多个转录物:(i) 表11;(ii) 表2、5和12中的每一者;
 - (iii) 表3、4、6和13中的每一者;(iv) 表14;或(v) 表8。
5. 如权利要求1所述的方法,其特征在于:所述多个靶血浆游离RNA分子包括来自表8或11-15中的一者或多者的至少30个基因的多个转录物。
6. 如权利要求1所述的方法,其特征在于:所述多个靶血浆游离RNA分子是选自表14中的多个转录物。
7. 如权利要求1所述的方法,其特征在于:检测到高于一阈值的所述多个靶血浆游离RNA分子是衍生自选自以下多个基因所组成的群组的多个血浆游离RNA分子:ADIPOQ、AGR3、ANKRD30A、AQP4、BPIFA1、CA12、CEACAM5、CFTR、CXCL17、CYP4F8、FABP7、FOXI1、GGTLC1、GP2、IL20、ITIH6、LDLRAD1、LEMD1、LMX1B、MMP7、NKAIN1、NKX2-1、ROPN1、ROS1、SCGB1D2、SCGB2A2、SFTA2、SFTA3、SLC34A2、SOX9、STK32A、STMND1、TFAP2A、TFAP2B、TFF1、TRPV6、VGLL1及VTCN1。
8. 如权利要求1所述的方法,其特征在于:所述多个靶血浆游离RNA分子是选自表8中的多个转录物。
9. 如权利要求1所述的方法,其特征在于:检测到高于一阈值的所述多个靶血浆游离RNA分子是衍生自选自以下多个基因所组成的群组的多个血浆游离RNA分子:CEACAM5、RHOF、SFTA2、SCGB1D2、IGF2BP1、SFTPA1、CA12、SFTPB、CDH3、MUC6、SLC6A14、HOXC9、AGR3、TMEM125、TFAP2B、IRX2、POTEKP、ARHGEF38、GPR87、LMX1B、ATP10B、NELL1、MUC21、SOX9、LINC00993、STMND1、ERVH48-1、SCTR、MAGEA3、MB、LEMD1、SIX4及NXNL2。
10. 如权利要求1所述的方法,其特征在于:所述多个靶血浆游离RNA分子包括(a) 表8或11-14中的一者或多者的多个转录物、及(b) 表1-6中的一个或多个转录物。
11. 如权利要求1所述的方法,其特征在于:所述多个靶血浆游离RNA分子包括(a) 表8或11-14中的一者或多者的多个转录物、及(b) 表7中的一个或多个转录物。
12. 如权利要求1所述的方法,其特征在于:(i) 所述癌症是肺癌,及(ii) 检测到高于一阈值的所述多个靶血浆游离RNA分子是选自表2、5或12中的一者或多者的多个转录物。
13. 如权利要求1所述的方法,其特征在于:(i) 所述癌症是乳癌,及(ii) 检测到高于一阈值的所述多个靶血浆游离RNA分子是选自表3、4、6或13中的一者或多者的多个转录物。
14. 如权利要求1至13中任一项所述的方法,其特征在于:所述测量包括测序、微阵列分析、反转录PCR、实时PCR、定量实时PCR、数位PCR、数位微滴PCR、数位乳化PCR、多重PCR、杂交

捕获、多种寡核苷酸连接测定或其任意组合。

15. 如权利要求1至13中任一项所述的方法,其特征在于:所述测量的步骤包括对多个血浆游离RNA分子测序,以产生多个血浆游离RNA序列读数。

16. 如权利要求15所述的方法,其特征在于:对所述多个血浆游离RNA分子测序的步骤包括全转录组测序。

17. 如权利要求15所述的方法,其特征在于:对所述多个血浆游离RNA分子测序的步骤包括反转录以产生多个互补DNA分子,以及对所述多个互补DNA分子测序以产生所述多个血浆游离RNA序列读数。

18. 如权利要求15所述的方法,其特征在于:对所述多个血浆游离RNA分子测序的步骤包括富集所述多个靶血浆游离RNA分子或其多个互补DNA分子。

19. 如权利要求1至13中任一项所述的方法,其特征在于:所述样本包括一生物流体。

20. 如权利要求19所述的方法,其特征在于:所述生物流体包括血液、血浆、血清、尿液、唾液、脑脊液、心包液、脑脊液、腹膜液或其任意组合。

21. 如权利要求19所述的方法,其特征在于:所述生物流体包括所述受试者的血液、一血液级分、血浆或血清。

22. 如权利要求1至13中任一项所述的方法,其特征在于:检测高于一阈值水平的所述多个靶血浆游离RNA分子的一个或多个的步骤包括(i)检测、(ii)高于背景检测、或(iii)在一水平检测,所述水平比未患有所述病况的多位受试者中的所述多个靶血浆游离RNA分子的一水平高。

23. 如权利要求1至13中任一项所述的方法,其特征在于:检测高于一阈值水平的所述多个靶血浆游离RNA分子中的一个或多个的步骤包括在一水平检测所述一个或多个靶血浆游离RNA分子,所述水平比未患有所述病况的多位受试者的一水平高至少10倍。

24. 如权利要求15所述的方法,其特征在于:检测高于一阈值水平的所述多个靶血浆游离RNA分子中的一个或多个的步骤包括检测高于每百万个中0.5至5个读数的一阈值。

25. 如权利要求1至13中任一项所述的方法,其特征在于:检测高于一阈值水平的所述多个靶血浆游离RNA分子中的一个或多个包括以下步骤:

(a) 通过将所述多个靶血浆游离RNA分子的每一个的所述表达水平与一RNA组织评分矩阵进行比较,确定每一个靶血浆游离RNA分子的一指标评分;

(b) 汇总每一个靶血浆游离RNA分子的所述多个指标评分;及

(c) 当所述指标评分超过一阈值时检测所述癌症。

26. 如权利要求15所述的方法,其特征在于:检测高于一阈值水平的所述多个靶血浆游离RNA分子中的一个或多个的步骤包括将所述多个序列读数输入到一机器学习或深度学习模型中。

27. 如权利要求26所述的方法,其特征在于:所述机器学习或深度学习模型包括逻辑回归、随机森林、梯度提升机、朴素贝叶斯、神经网络或多项回归。

28. 如权利要求26所述的方法,其特征在于:所述机器学习或深度学习模型通过包括多个学习权重的一函数将所述一个或多个特征的所述多个值转换为所述受试者的所述疾病状态预测。

29. 如权利要求1至13中任一项所述的方法,其特征在于:所述癌症包括:

(i) 一上皮癌、一恶性肉瘤、一骨髓瘤、一白血病、一淋巴瘤、一胚细胞瘤、一生殖细胞瘤或其任意组合；

(ii) 选自以下所组成的群组的一上皮癌：腺癌、鳞状细胞癌、小细胞肺癌、非小细胞肺癌、鼻咽癌、结直肠癌、肛门癌、肝癌、膀胱癌、睾丸癌、子宫颈癌、卵巢癌、胃癌、食道癌、头颈癌、胰脏癌、前列腺癌、肾癌、甲状腺癌、黑色素瘤和乳癌；

(iii) 激素受体阴性乳癌或三阴性乳癌；

(iv) 选自以下所组成的群组的一恶性肉瘤：骨肉瘤、软骨肉瘤、平滑肌肉瘤、横纹肌肉瘤、间皮肉瘤(间皮瘤)、纤维肉瘤、血管肉瘤、脂肪肉瘤、神经胶质瘤和星形细胞瘤；

(v) 选自以下所组成的群组的一白血病：骨髓性白血病、粒细胞性白血病、淋巴性白血病、淋巴球性白血病和淋巴胚细胞性白血病；或者

(vi) 选自以下所组成的群组的一淋巴瘤：霍奇金淋巴瘤和非霍奇金淋巴瘤。

30. 如权利要求1至13中任一项所述的方法，其特征在于：检测所述癌症的步骤包括确定一癌症阶段、确定癌症进展、确定一癌症类型、确定癌症起源组织或其一组合。

31. 如权利要求1至13中任一项所述的方法，其特征在于：所述方法进一步包括根据所述检测到的癌症选择一治疗。

32. 如权利要求31所述的方法，其特征在于：所述治疗包括手术切除、放射疗法或施用一抗癌剂。

33. 如权利要求31所述的方法，其特征在于：所述方法进一步包括用所述选择的治疗治疗所述受试者。

34. 一种测量一样本中多个标靶血浆游离RNA分子的方法，其特征在于：所述方法包括以下步骤：

(a) 富集所述多个标靶血浆游离RNA分子或其多个互补DNA分子，以产生多个多核苷酸的一富集样本；及

(b) 对所述富集样本的所述多个多核苷酸或其多个扩增产物测序；

其中，所述多个标靶血浆游离RNA分子选自表11中的一个或多个转录物。

35. 如权利要求34所述的方法，其特征在于：所述多个标靶血浆游离RNA分子选自表8或12-15中的一者或多者。

36. 如权利要求34所述的方法，其特征在于：所述多个标靶血浆游离RNA分子选自表8。

37. 如权利要求34所述的方法，其特征在于：所述多个标靶血浆游离RNA分子选自表14。

38. 如权利要求34所述的方法，其特征在于：所述多个标靶血浆游离RNA分子选自来自表8或11-14中的一者或多者的至少5、10、15或20个基因的多个转录物。

39. 一种用于执行权利要求1-38中任一项所述方法中的一个或多个步骤的电脑系统。

40. 一种非暂时性电脑可读介质，其上储存有用于执行权利要求1-38中任一项所述方法中的一个或多个步骤的多个电脑可读指令。

用于血浆游离RNA分析的多种方法

[0001] 相关申请案

[0002] 本申请主张2020年6月16日申请的第63/039,886号美国临时专利申请案的优先权,其内容在此通过引用其整体并入本文中以用于所有目的。

背景技术

[0003] 截至2017年,美国每年共有超过160万例新病例,癌症是一个突出的全球公共卫生问题。参见Siegel等人,2017,“癌症统计 (Cancer statistics)”,CA Cancer J Clin.67 (1):7-30。多个筛查计划和早期诊断对提高癌症患者的无病生存率和降低死亡率具有一重要影响。由于早期诊断的多种非侵入性方法可以促进患者的依从性,因此可以将它们包括在多个筛查计划中。

[0004] 多个血浆游离核酸 (cell-free nucleic acid, cfNA) 可在血清、血浆、尿液和其他体液中发现 (Chan等人,“临床生化学家协会临床科学评论委员会血浆、血清和尿液中的游离核酸:一种新的分子诊断工具 (Clinical Sciences Reviews Committee of the Association of Clinical Biochemists Cell-free nucleic acids in plasma, serum and urine: a new tool in molecular diagnosis)”, Ann Clin Biochem.2003;40 (Pt 2):122-130) 代表一“液体活检”,这是一特定疾病的一循环图像。参见De Mattos-Arruda和Caldas,2016,“血浆游离循环肿瘤DNA作为乳腺癌的一液体活检 (Cell-free circulating tumour DNA as a liquid biopsy in breast cancer)”, Mol Oncol.2016;10 (3):464-474。类似地,已提出血浆游离 (cell-free) RNA作为癌症检测的一可能分析物。参见Tzimagiorgis等人,“从多种体液中恢复循环的细胞外或血浆游离RNA (Recovering circulating extracellular or cell-free RNA from bodily fluids)”, Cancer Epidemiology 2011;35 (6):580-589。这些方法代表筛查各种疾病,例如多种癌症的多种潜在非侵入性方法。

[0005] 然而,癌症仍然是全世界一常见的死亡原因。在过去的数十年里治疗方案有所改进,但存活率仍然很低。手术切除和药物治疗的所述成功很大程度上取决于多个早期肿瘤的认识。然而,多种当前的技术,例如成像和多种基于生物标记物 (biomarker) 的方法,通常在所述疾病进入所述更晚期阶段之前无法识别多个肿瘤。

[0006] 因此,当多个治疗干预具有一更大的成功机会时,仍然需要能够在最早阶段识别疾病的多种新的非侵入性检测方式。本发明满足这些和其他需要。

发明内容

[0007] 根据各种实施例,本公开的主题提供了多种测量一样本中的多个靶标血浆游离RNA (cell-free RNA, cfRNA) 分子的方法。在多个实施例中,所述多种方法包括以下步骤:(a) 富集所述多个靶标cfRNA分子或其多个互补DNA (cDNA) 分子,以产生多个多核苷酸的一富集样本;及/或 (b) 对所述富集样本的所述多个多核苷酸或其多个扩增产物测序;其中,所述多个靶标cfRNA分子选自表11中的一个或多个转录物。在多个实施例中,所述多个靶标

cfRNA分子可以选自表8或12-15中的一者或多者,或其任意组合(例如来自表8或11-14中的一者或多者的5、10、15或20个基因的多个转录物)。

[0008] 在一些方面,本主题公开提供了一种检测一受试者中癌症的方法。在多个实施例中,所述多种方法包括以下步骤:(a) 测量所述受试者的一样本中的多个靶血浆游离RNA(cfRNA)分子,其中所述多个靶cfRNA分子选自表11中的多个转录物;及/或(b) 检测所述癌症,其中检测所述癌症包括检测高于一阈值水平的所述多个靶cfRNA分子中的一个或多个。在多个实施例中,检测高于一阈值的所述靶cfRNA分子的一个或多个的步骤包括(i) 检测、(ii) 高于背景检测、或(iii) 在一水平检测,所述水平比未患有所述病况的多位受试者中的多个相应序列读数的一水平高。在多个实施例中,所述多个靶cfRNA分子选自表8或12-15中的一者或多者(例如来自表8或11-14中的一者或多者的5、10、15或20个基因的多个转录物)。

[0009] 在各个方面,本公开提供了检测一受试者的一疾病状态的多种方法和多种组合物。在多个实施例中,所述多种方法包括检测血浆游离核糖核酸(cfRNA)中的一种或多种标记物。在多个实施例中,检测cfRNA的步骤包括对来自一受试者的一生物样本中的cfRNA测序,以产生多个cfRNA读数。在多个实施例中,所述方法进一步包括对来自一受试者的多个细胞中的RNA测序,以产生多个细胞读数,并且过滤所述多个cfRNA读数,以排除对应于一个或多个细胞读数的多个cfRNA读数。在多个实施例中,所述多个细胞是多个血球细胞。在多个实施例中,所述多种方法包括过滤所述多个cfRNA读数,以排除一个或多个核糖体、线粒体及/或血液相关的转录物。在多个实施例中,仅测量与一外显子(exon)-外显子连接重叠的多个cfRNA读数(或多个读数对)。在多个实施例中,测量对应于一种或多种标记物的cfRNA(例如1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、25或多种标记物)。所述一种或多种标记物可以是本文公开的所述多种标记物中的任一者,以任意组合形式。在多个实施例中,所述一种或多种标记物与所述疾病状态相关。在多个实施例中,所述多种方法包括治疗一受试者的所述疾病状态。

[0010] 本公开的多个方面包括多种用于检测一受试者中一疾病状态的方法,所述方法包括:从所述受试者中分离一生物测试样本,其中所述生物测试样本包括多个血浆游离核糖核酸(cfRNA)分子;从所述生物测试样本中提取所述多个cfRNA分子;对所述多个提取的cfRNA分子进行一测序程序,以产生多个序列读数;执行一过滤程序,以产生源自一个或多个健康细胞的多个序列读数的一排除群体以及多个序列读数的一非排除群体;对所述多个非排除的序列读数执行一量化程序;以及当所述量化程序产生超过一阈值的一值时,检测所述受试者中的所述疾病状态。在多个实施例中,检测高于一阈值的一个或多个非排除的序列读数的步骤包括(i) 检测、(ii) 高于背景检测、或(iii) 在一水平检测,所述水平比未患有所述病况的多位受试者中的多个相应序列读数的一水平高。

[0011] 本公开的主题的多个方面进一步包括多种用于识别指示一疾病状态的一个或多个RNA序列的电脑执行方法,所述方法包括:通过一电脑系统获得来自已知患有所述疾病的一受试者的一第一测试样本的多个RNA分子的一第一组的多个序列读数,其中所述第一测试样本包括多个血浆游离RNA(cfRNA)分子;通过一电脑系统获得来自一对照样本的多个RNA分子的一第二组的多个序列读数;通过一电脑系统检测存在于所述第一组的多个序列读数中且不存在于所述第二组的多个序列读数中的一个或多个RNA序列,以识别指示所述

疾病状态的一个或多个RNA序列。

[0012] 在其他多个方面,本主题是有关多种用于检测一受试者中的一个或多个源自肿瘤的RNA分子的电脑执行方法,所述方法包括:通过一电脑系统获得来自已知患有或怀疑患有一肿瘤的一受试者的一第一测试样本的多个RNA分子的一第一组的多个序列读数,其中所述第一测试样本包括多个血浆游离RNA (cfRNA) 分子;通过一电脑系统获得来自所述受试者的多个血球细胞的多个RNA分子的一第二组的多个序列读数;以及通过一电脑系统检测存在于所述第一组的多个序列读数中且不存在于所述第二组的多个序列读数中的一个或多个RNA序列,以检测所述受试者中的所述一个或多个源自肿瘤的RNA分子。

[0013] 在一些方面,本公开的主题是有关多种用于在已知患有或怀疑患有一癌症的一受试者中检测一癌症的一存在、确定一癌症阶段、监测一癌症进展、及/或确定一癌症类型或癌症亚型的方法,所述方法包括:(a) 从所述受试者中获得一生物测试样本,其中所述生物测试样本包括多个血浆游离核糖核酸(cfRNA) 分子;(b) 定量地检测所述生物测试样本中衍生自一个或多个靶RNA分子的一个或多个核酸序列的所述存在,以确定一肿瘤RNA评分,其中所述一个或多个靶RNA分子选自表8或11-14中任一者所列的所述多个靶RNA分子;及/或(c) 当所述肿瘤RNA评分超过一阈值时,在所述受试者中检测所述癌症的所述存在、确定所述癌症阶段、监测所述癌症进展、及/或确定所述癌症类型或亚型。

[0014] 根据各个方面,本公开是有关多种用于检测一受试者中一癌症的所述存在的电脑执行方法,所述方法包括:在包括一处理器和一电脑可读介质(computer-readable medium)的一电脑中接收一数据集,其中所述数据集包括从来自所述受试者的一生物测试样本中的多个核糖核酸(RNA) 分子获得的多个序列读数,并且其中所述电脑可读介质包括多个指令,当由所述处理器执行这些指令时,使所述电脑:确定所述生物测试样本中多个靶RNA分子的一表达水平;将所述多个靶RNA分子中的每一个的所述表达水平与一RNA组织评分矩阵进行比较,以确定所述多个靶RNA分子中的每一个的一癌症指标评分;汇总所述多个靶RNA分子中的每一个的所述癌症指标评分,以产生所述生物测试样本的一癌症指标评分;及/或当所述生物测试样本的所述癌症指标评分超过一阈值时,检测所述受试者中所述癌症的所述存在。

[0015] 在各个方面,本公开的所述主题是有关多种用于构建一RNA组织评分矩阵的方法,所述方法包括:编译从多位受试者中获得的多个RNA序列读数,以产生一RNA表达矩阵;以及使用一组织特异性RNA表达矩阵对所述RNA表达矩阵进行归一化,以构建所述RNA组织评分矩阵。在一些实施例中,从多位患有一已知癌症类型的受试者中获得所述多个RNA序列读数,以构建一癌症RNA组织评分矩阵。

[0016] 在一些方面,本公开的主题提供了多种测量一受试者的多个血浆游离RNA(cfRNA) 分子的方法。在多个实施例中,所述方法包括(a) 对所述多个cfRNA分子测序,以产生多个cfRNA序列读数;(b) 对从所述受试者的多个细胞中提取的细胞RNA测序,以产生多个细胞序列读数;(c) 执行一过滤程序,以产生多个cfRNA序列读数的一非排除群体,其中所述过滤的步骤包括排除匹配所述多个细胞序列读数的一个或多个的多个cfRNA序列读数;及/或(d) 量化所述多个非排除序列读数的一个或多个。

[0017] 在一些方面,本公开提供了多种识别从一名或多名受试者中收集的多个样本中的多种癌症生物标记物(在本文中也称为“标记物”)的方法。在多个实施例中,所述方法包括:

(a) 对从未患有癌症的多位受试者中收集的生物流体的cfRNA测序,以产生多个非癌症测序读数; (b) 对于从患有一癌症的一名或多名受试者中收集的多个匹配样本: (i) 对从一匹配样本的一癌症组织收集的DNA及RNA测序,以产生所述癌症组织的多个测序读数; (ii) 对从所述匹配样本的一匹配生物流体中收集的cfDNA和cfRNA测序,以产生所述匹配生物流体的多个测序读数; (iii) 通过将所述匹配生物流体的多个cfDNA测序读数的多个计数与所述癌症组织的多个DNA测序读数的多个相应计数相关联来测量一肿瘤分数; 及/或 (iv) 通过将一种或多种候选生物标记物的所述多个RNA测序读数的一计数乘以所述肿瘤分数来测量所述一种或多种候选生物标记物的肿瘤含量,其中所述一种或多种候选生物标记物在所述匹配生物流体中的表达水平高于从未患有癌症的所述多位受试者中收集的所述生物流体中的表达水平; (c) 使用所述肿瘤含量作为一共变量(covariate)模拟cfRNA中的所述一种或多种候选生物标记物的表达; 及/或 (d) 基于所述模拟从所述一种或多种候选生物标记物中识别一种或多种cfRNA癌症生物标记物。

[0018] 在一些方面,本公开提供了多种电脑系统,用于在本文公开的所述各个方面的任一方面的多种方法中执行一个或多个步骤。

[0019] 在一些方面,本公开的主题提供了多种非暂时性电脑可读介质,其上储存有用于在本文公开的所述各个方面的任一方面的多种方法中执行一个或多个步骤的多个电脑可读指令。

附图说明

[0020] 图1是根据一个实施例的一种用于制备用于测序的一核酸样本的方法的一流程图。

[0021] 图2是说明根据本发明的一个实施例的一种用于识别指示一疾病状态的一个或多个RNA序列的方法的一流程图。

[0022] 图3是说明根据本发明的一个实施例的一种用于识别一个或多个源自肿瘤的RNA序列的方法的一流程图。

[0023] 图4是说明根据本发明的一个实施例的一种在一受试者中检测癌症的所述存在、确定癌症的一状态、监测癌症进展及/或确定癌症类型的方法的一流程图。

[0024] 图5是说明根据本发明的一个实施例的一种从源自一个或多个靶RNA分子的一个或多个序列读数检测一疾病状态的方法的一流程图。

[0025] 图6是说明根据本发明的一个实施例的一种用于基于一癌症指标评分来检测一受试者中癌症存在的方法的一流程图。

[0026] 图7显示了根据一实施例的多个样本分类方案的灵敏度和特异性的多个示例结果。

[0027] 图8A至8C显示了根据一实施例的多个样本分类方案的灵敏度和特异性的多个示例结果。

[0028] 图9描绘了肺癌中20个暗通道基因(dark channel gene)的所述表达水平,其中癌性和非癌性样品之间的表达水平比率最高。每百万个中个读数(reads per million,RPM)绘制为多个暗通道基因的一函数。在每个图中,从左到右的多个点的所述多个列分别对应于所述顶部图例中从左到右指示的多个组(类别、肛门直肠癌、乳癌、结直肠癌、肺癌和非癌

症)。

[0029] 图10是使用从多个暗通道基因汇总的一组织评分的所述决策树分类器(decision tree classifier)的一ROC曲线。

[0030] 图11是说明根据一些实施例的一种方法的一流程图。

[0031] 图12A是第III期TCGA(癌症基因组图谱计划)FFPE(福马林固定后包埋在石蜡块中)组织RNA-seq数据的一示例PCA(主成分分析)的一散点图。多个基因表达水平以每百万个中个读数绘制。

[0032] 图12B是散点图,显示CCGA(循环血浆游离基因组图谱)肿瘤组织RNA-seq数据的多个示例结果,投射在TCGA PCA轴上。多个基因表达水平以每百万个中个读数绘制。

[0033] 图12C是散点图,显示投射在TCGA PCA轴上的CCGA癌症血浆游离RNA(cfRNA) RNA-seq数据的多个示例结果。多个基因表达水平以每百万个中个读数绘制。

[0034] 图13是多个示例暗通道生物标记物基因的一热图。每列描绘一个cfRNA样本,每一行描绘一个基因。所述多个行的所述颜色编码组织特异性(从上到下,所述多个组织分别是:乳房、肺和非特异性)。所述多个列的所述颜色编码所述多个样本组(从左到右,所述多个癌症类型分别是:肛门直肠癌、乳癌、结直肠癌、肺癌和非癌症)。

[0035] 图14A显示描绘多个不同样本:HER2+、HR+/HER2-、三阴性乳癌(TNBC)或非癌症样本中两个示例乳房暗通道生物标记物(dark channel biomarker,DCB)基因(FABP7和SCGB2A2)的cfRNA表达水平和组织表达水平的多个箱形图。

[0036] 图14B显示描绘多个不同样本:腺癌、小细胞肺癌、鳞状细胞癌或非癌症样本中四个示例肺DCB基因(SLC34A2、R0S1、SFTPA2和CXCL17)的cfRNA表达水平和组织表达水平的多个箱形图。

[0037] 图15A显示森林图,描绘了两个乳房DCB基因(FABP7和SCGB2A2)对于具有匹配肿瘤组织的多个乳癌样本的所述可检测性。所述多个样本ID是根据它们在血浆游离DNA(cfDNA)中的相对肿瘤分数绘制的(95% CI)。在样本4653、4088、2037、3116和1202中检测到FABP7。在样本1656、2419、3911、2367、2037、1039、2139和3162中检测到SCGB2A2。从来自所述cfDNA富集测定的多个SNV等位基因分数中测量cfDNA中的肿瘤分数。

[0038] 图15B显示森林图,描绘了两个乳房DCB基因(FABP7和SCGB2A2)对于具有匹配肿瘤组织的多个乳癌样本的所述可检测性。多个样本ID绘制为肿瘤含量的一函数(肿瘤分数*肿瘤组织表达)。在样本4088、1202、3116和2037中检测到FABP7。在样本1656、2419、2367、3911、1039、2139、3162和2037中检测到SCGB2A2。从来自所述cfDNA富集测定的多个SNV等位基因分数中测量cfDNA中的肿瘤分数。从匹配的肿瘤组织的RNA-seq数据测量组织表达。

[0039] 图16A至16D显示了对于患有乳癌、肺癌或无癌症(正常)的多位受试者的所述多个指示基因在cfRNA和匹配组织中的DCB基因表达的多个示例测序结果。多个读数计数的所述数量表示在所述y轴上。

[0040] 图17A至17B显示了示例分类器工作流程。

[0041] 图18A至18C显示了显示多个示例分类方案的灵敏度和特异性的ROC图。

[0042] 图19显示了根据本发明的一个实施例的一种样本处理和参数确定方法。

[0043] 图20A至20B显示了根据一实施例的选择多个乳房和肺特异性生物标记物的所述多个分布,显示了衍生自乳癌和肺癌的(分别)cfRNA与衍生自非癌症的cfRNA相比增加的信

号。从多位乳癌、肺癌和非癌症CCGA参与者的所述cfRNA制备多个全转录组(transcriptome)样本。

[0044] 图21显示了来自多个全转录组CCGA乳癌样本的匹配血浆和组织基因表达。结果显示,组织中的高表达不一定会在血浆中产生高脱落率。

[0045] 图22显示说明CCGA血浆中暗通道表达与多个乳癌的CCGA肿瘤组织表达相关的一散点图。出于多个可视化目的,平均血浆或组织表达为零的多个基因在此被转化为 $1e-4$ 。

[0046] 图23是一散点图,说明CCGA血浆中暗通道表达与多个肺癌的CCGA肿瘤组织表达相关。出于多个可视化目的,平均血浆或组织表达为零的多个基因在此被转化为 $1e-4$ 。

[0047] 图24是显示多个CCGA血浆样本中的多个肿瘤特异性标记物的一图。基于从所有癌症血浆到所有非癌症血浆的多个观察计算每个基因的所述血浆对数可能性比率(log odds ratio)。所述多个显示的基因表示多个示例暗通道生物标记物。

[0048] 图25是一文氏图(Venn diagram),显示表15的多个cfRNA生物标记物按来源和识别方法分组的所述分布。表14中提供了所述图表中所有分组中存在的所述38个生物标记物。过滤多个基因以优化二元检测和优化起源组织(tissue-of-origin,T00)。在一CCGA血浆中观察到为优化二元检测而过滤的所述多个基因,其一对数可能性比率 >0.1 ,并且在乳癌和肺癌中具有高TCGA表达($>5RPM$)的所述多个基因。T00优化过滤的所述多个基因是通过多类随机森林方法(forest method)从TCGA组织中选择所述多个基因,并且在人类蛋白质图谱(Human Protein Atlas)中注释为乳房/肺肿瘤或组织特异性的所述多个基因。

[0049] 图26A至26D显示了根据一实施例与多位非癌症受试者相比,在乳癌及/或肺癌中检测到的多个选择的生物标记物的多个水平。

具体实施方式

[0050] 在更详细地描述本发明之前,应当理解本发明不限于所描述的多个特定实施例,因为这些实施例当然可以变化。还应理解,本文使用的所述术语仅用于描述多个特定实施例的所述目的,而不是限制性的,因为本发明的所述范围将仅由所附权利要求限制。

[0051] 在提供多个数值的一范围的情况下,应当理解,除非上下文另有明确规定,否则每个中间值至下限的所述单位的十分之一,在所述范围的上限和下限与所述规定范围内的任何其他规定的或中间值之间,以及所述范围的所述多个提供的端点的每一个,都包含在本发明内。这些较小范围的所述上限和下限可以独立地包括在本发明所涵盖的所述多个较小范围内,但受制于所述规定范围内的任何具体排除的限制。

[0052] 除非另有定义,否则本文使用的所有技术及/或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含义。Singleton等人,微生物学和分子生物学词典第2版(Dictionary of Microbiology and Molecular Biology 2nd ed.),J.Wiley&Sons(New York,NY 1994)为本领域技术人员提供了本申请中使用的所述许多术语的一般指南,如下所述,其每一个都通过引用其整体并入本文中:Kornberg和Baker,DNA复制(DNA Replication),第二版(W.H.Freeman,New Tork,1992);Lehninger,生物化学(Biochemistry),第二版(Worth Publishers,New Tork,1975);Strachan和Read,人类分子遗传学(Human Molecular Genetics),第二版(Wiley-Liss,New Tork,1999);Abbas等人,细胞和分子免疫学(Cellular and Molecular Immunology),第6版(Saunders,2007)。

[0053] 本文提及的所有出版物均通过引用明确地并入本文,以公开和描述与引用所述多个出版物相关的所述多种方法及/或多种材料。

[0054] 所述多个术语“多核苷酸”、“核酸”和“寡核苷酸”可互换使用。它们是指任何长度的多个核苷酸的一聚合形式,多个脱氧核糖核苷酸或多个核糖核苷酸或其多个类似物。多个多核苷酸可以具有任何三维结构,并且可以执行任何已知或未知的功能。以下是多个多核苷酸的多个非限制性示例:一基因或基因片段的编码区或非编码区、连锁分析定义的多个基因座(loci)(基因座(locus))、多个外显子、多个内含子(intron)、信使RNA(messenger RNA,mRNA)、传递RNA(transfer RNA,tRNA)、核糖体RNA(ribosomal RNA,rRNA)、短干扰RNA(short interferingRNA,siRNA)、短发夹RNA(short-hairpin RNA,shRNA)、小分子RNA(micro-RNA,miRNA)、多个核酶(ribozyme)、cDNA、多个重组多核苷酸、多个分支多核苷酸、多个质体(plasmid)、多个载体(vector)、任何序列的分离DNA、任何序列的分离RNA、多个核酸探针和多个引子(primer)。一多核苷酸可以包括一个或多个修饰的核苷酸,例如多个甲基化核苷酸和多个核苷酸类似物。如果存在,可以在所述聚合物的组装之前或之后赋予对所述核苷酸结构的多个修饰。多个核苷酸的所述序列可以被多个非核苷酸组分打断。一多核苷酸可在聚合后进一步修饰,例如通过与一标记组分结合。

[0055] 一般而言,所述术语“标靶多核苷酸”是指具有一标靶序列的多个核酸分子的一起始群体中的一核酸分子或多核苷酸,所述标靶序列的存在、数量及/或核苷酸序列,或这些中的一个或多个的多个变化是需要被确定的。一般而言,所述术语“标靶序列”是指一单股核酸上的一核酸序列。所述标靶序列可以是一基因的一部分、一调控序列、基因组DNA、cDNA、RNA,包括mRNA、miRNA、rRNA或其他。所述标靶序列可以是来自一样本的一标靶序列或一次级标靶,例如一扩增反应的一产物。

[0056] 所述术语“标记物”和“生物标记物”在本文中可互换使用,是指一多核苷酸(例如一基因或其一可识别的序列片段),其所述水平或浓度与一特定生物状态(例如一疾病状态,如一般癌症的存在,或一特定的癌症类型及/或阶段)。在多个实施例中,一标记物是一特定基因的一cfRNA,其所述水平的多个变化可以通过测序来检测。多个cfRNA生物标记物在本文中可以参考所述cfRNA所源自的所述基因,但不需要检测所述整个基因转录物。在多个实施例中,仅检测到一特定基因转录物的多个片段。在多个实施例中,检测一特定基因的所述存在及/或水平包括检测一个或多个cfRNA片段,所述一个或多个cfRNA片段包括源自相同基因的多个转录物的多个不同序列片段(重叠或非重叠),其可以作为相同“生物标记物”的部分进行综合评分。与多个引用的基因名称有关的附加信息,包括序列信息(例如DNA、RNA和多个氨基酸序列)、通常通过基因符号识别的多个基因的多个全名等,可在本领域技术人员已知的多个可公开进入的数据库中获得,例如可从所述国家生物技术信息中心(www.ncbi.nlm.nih.gov/)获得的多个数据库,包括GenBank(www.ncbi.nlm.nih.gov/genbank/)和所述NCBI蛋白质数据库(www.ncbi.nlm.nih.gov/protein/)和UniProt(www.uniprot.org/)。

[0057] 如本文所用,所述术语“扩增子(amplicon)”是指一多核苷酸扩增反应的所述产物;即,多个多核苷酸的一克隆群体,其可以是单股或双股的,其从一个或多个起始序列复制。所述一个或多个起始序列可以是相同序列的一个或多个拷贝,或者它们可以是多个不同序列的一混合物。优选地,多个扩增子由一单个起始序列的所述扩增形成。多个扩增子可

以通过各种扩增反应产生,其多个产物包括所述一种或多种起始或靶核酸的复制品。在一个方面,产生多个扩增子的多种扩增反应是“模板驱动的”,因为多个反应物(多个核苷酸或多个寡核苷酸)的碱基配对在一模板多核苷酸中具有产生多个反应产物所需的多个互补物。在一个方面,多个模板驱动的反应是使用一核酸聚合酶的多个引子延伸,或使用一核酸连接酶的多个寡核苷酸连接。此类反应包括,但不限于多个聚合酶连锁反应(polymerase chain reaction,PCR)、多个线性聚合酶反应(linear polymerase reaction)、多个基于核酸序列的扩增(nucleic acid sequence-based amplification,NASBA)、多个滚环式扩增(rolling circle amplification)等,在以下参考文献中公开,其中的每一篇文献在此都通过引用其整体并入本文中:Mullis等人,美国第4,683,195号、第4,965,188号、第4,683,202号、第4,800,159号专利(PCR);Gelfand等人,美国第5,210,015号专利(带有多个“taqman”探针的实时PCR(real-time PCR with “taqman” probes));Wittwer等人,美国第6,174,670号专利;Kacian等人,美国第5,399,491号专利(“NASBA”);Lizardi,美国第5,854,033号专利;Aono等人,日本第JP 4-262799号专利公开案(滚环式扩增)等。在一个方面,本发明的多个扩增子通过多个PCR产生。如果可以使用允许在所述扩增反应进行时测量一反应产物的一检测化学物质,则一扩增反应可以是一“实时”扩增,例如“实时PCR(real-time PCR)”或“实时NASBA”,如Leone等人,核酸研究(Nucleic Acids Research),26:2150-2155 (1998)和类似参考文献中所述的。

[0058] 所述术语“扩增”是指进行一扩增反应。一“反应混合物”是指包含进行一反应所需的所有所述反应物的一溶液,其可以包括,但不限于在一反应期间将pH保持在一选择水平的多个缓冲剂、多个盐、多个辅助因子、多个清除剂等。

[0059] 如本文可互换使用的所述术语“片段(fragment)”或“区段(segment)”是指一较大核苷酸分子的一部分。例如,一多核苷酸可以通过多个自然过程,例如可以在一生物样本中自然出现的多个cfDNA片段的情况,或通过体外操纵分解或片段化成多个区段。使核酸片段化的各种方法在本领域中是众所周知的。这些方法本质上可以是,例如化学的或物理的或酶促的。酶促片段化(fragmentation)可能包括用一脱氧核糖核酸分解酶(DNase)进行部分降解;用酸部分去嘌呤(depuration);多个限制酶的所述使用;多个内含子编码的核酸内切酶;多种基于DNA的切割方法,例如三股螺旋(triplex)和杂合体(hybrid)形成方法,其依赖于一核酸区段的所述特异性杂交,以将一切割剂定位到所述核酸分子中的一特定位置;或在已知或未知位置切割一多核苷酸的其他多个酶或多个化合物。多种物理片段化方法可能涉及使一多核苷酸经受一高剪切速率。可以,例如通过使DNA移动通过带有多个凹坑或多个尖刺的一腔室或通道,或迫使一DNA样本通过一尺寸受限的流动通道,例如一横截面尺寸在所述微米或次微米范围内的一孔,产生多个高剪切速率。其他多种物理方法包括超音波震荡(sonication)和雾化(nebulization)。同样可以使用多种物理和化学片段化方法的多种组合,例如通过热和离子介导的水解进行的片段化。参见,例如Sambrook等人,“分子克隆:一实验室手册(Molecular Cloning:A Laboratory Manual)”,第3版.Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y. (2001) (Sambrook等人),出于所有目的通过引用将其并入本文中。可以优化这些方法,以将一核酸消化成一选择大小范围的多个片段。

[0060] 如本文可互换使用的所述术语“聚合酶连锁反应”或“PCR”是指通过DNA的多个互

补股的所述同时引子延伸来体外扩增多个特定DNA序列的一反应。换言之,PCR是一种用于制造侧面为多个引子结合位点的一靶核酸的多个拷贝或多个复制品的反应,所述反应包括以下多个步骤的一个或多个重复:(i)使所述靶核酸变性,(ii)将多个引子黏合至所述多个引子结合位点,及(iii)在多个核苷三磷酸的所述存在下通过一核酸聚合酶延伸所述多个引子。通常,所述反应通过针对一热循环仪仪器中的每个步骤优化的多个不同温度进行循环。每个步骤的多个特定温度、多个持续时间和步骤之间的多个变化率取决于本领域普通技术人员众所周知的许多因素,例如以下参考文献举例说明:McPherson等人,编辑,PCR:一实用方法和PCR2:一实用方法(PCR:A Practical Approach and PCR2:A Practical Approach)(分别为IRL Press,Oxford,1991and 1995)。例如,在使用Taq DNA聚合酶的一常规PCR中,一双股靶核酸可以在>90℃的一温度下变性,多个引子在50-75℃的所述范围的一温度下黏合,并且多个引子在72-78℃的所述范围的一温度下延伸。所述术语“PCR”涵盖所述反应的多个衍生形式,包括但不限于反转录酶PCR(Rreverse Transcriptase PCR, RT-PCR)、实时PCR、巢式聚合酶连锁反应(nested PCR)、定量PCR(quantitative PCR)、多重PCR(multiplexed PCR)等。本领域技术人员可以从一申请案的上下文中辨别所采用的所述特定PCR的格式。多个反应体积的范围可以从几百纳升,例如200nL,到几百μL,例如200μL。“反转录PCR”或“RT-PCR”是指在一反转录反应之前进行的一PCR,所述反转录反应将一靶RNA转换为一互补的单股DNA,然后将其扩增,其一示例在Tecott等人,美国第5,168,038号专利中描述,其公开的全部内容通过引用并入本文中。“实时PCR”是指随着所述反应进行而监测反应产物(即,扩增子)的所述数量的一PCR。存在多种形式的实时PCR,其主要区别在于用于监测所述反应产物的所述多个检测化学物质,例如Gelfand等人,美国第5,210,015号专利(“taqman”);Wittwer等人,美国第6,174,670号和第6,569,627号专利(多个嵌入染料);Tyagi等人,美国第5,925,517号专利(多个分子信标);其公开的全部内容通过引用并入本文中。Mackay等人,核酸研究(Nucleic Acids Research),30:1292-1305(2002)中回顾了实时PCR的多个检测化学物质,其也通过引用并入本文中。“巢式PCR”是指一两阶段PCR,其中一第一次PCR的所述扩增子成为使用一组新的多个引子进行一第二次PCR的所述样本,其中至少一个引子与所述第一扩增子的一内部位置结合。如本文所用,关于一巢式扩增反应的“多个初始引子”是指用于产生一第一扩增子的所述多个引子,以及“多个次级引子”是指用于产生一第二或巢式扩增子的所述一个或多个引子。“不对称PCR(asymmetric PCR)”是指其中使用的所述两个引子之一的浓度大大过量的一PCR,使得所述反应主要是一线性扩增,其中一靶核酸的所述两股之一被优先复制。多个不对称PCR引子的所述过量浓度可以表示为一浓度比。多个典型的比率在10到100的所述范围内。“多重PCR”是指其中多个靶序列(或一单个靶序列和一个或多个参考序列)在相同反应混合物中同时进行的一PCR,例如Bernard等人,Anal.Biochem.,273:221-228(1999)(双色实时PCR)。通常,对每个被扩增的序列使用多个不同组的多个引子。典型地,一多重PCR中多个靶序列的所述数量在2到50、或2到40、或2到30的所述范围内。“定量PCR”是指设计用于测量一样本或样品中一个或多个特定靶序列的所述丰度的一PCR。定量PCR包括此类靶序列的绝对定量和相对定量。使用一个或多个参考序列或多个内部标准进行多个定量测量,这些参考序列或内部标准可以单独地或与一靶序列一起分析。所述参考序列对于一样本或样品可以是内源的或外源的,并且在后者情况下,可以包括一个或多个竞争模板。多个典型的内源参考序列包

括以下多个基因的多个转录物的多个区段： β -肌动蛋白、GAPDH、 β_2 -微球蛋白、核糖体RNA等。用于定量PCR的多种技术对于本领域普通技术人员来说是众所周知的，如以下参考文献中所举例说明的，这些文献通过引用其整体并入本文中：Freeman等人，多种生物技术 (Biotechniques), 26:112-126 (1999)；Becker-Andre等人，核酸研究 (Nucleic Acids Research), 17:9437-9447 (1989)；Zimmerman等人，多种生物技术 (Biotechniques), 21:268-279 (1996)；Diviacco等人，基因 (Gene), 122:3013-3020 (1992)；以及Becker-Andre等人，核酸研究 (Nucleic Acids Research), 17:9437-9446 (1989)。

[0061] 如本文所用，所述术语“引子”是指天然或合成的一寡核苷酸，其在与一多核苷酸模板形成一双股螺旋 (duplex) 后，能够充当核酸合成的一初始点并从其3'端沿所述模板延伸，从而形成一延伸的双股螺旋。一引子的延伸通常用一核酸聚合酶进行，例如一DNA或RNA聚合酶。所述延伸过程中添加的多个核苷酸的所述序列由所述模板多核苷酸的所述序列决定。通常，多个引子由一DNA聚合酶延伸。多个引子的一长度通常在14至40个核苷酸的所述范围内，或在18至36个核苷酸的所述范围内。多个引子用于各种核酸扩增反应，例如，使用一单个引子的多个线性扩增反应，或使用两个或多个引子的多个聚合酶连锁反应。为多个特定应用选择多个引子的所述多个长度和多个序列的指导对于本领域普通技术人员来说是众所周知的，如以下参考文献所证明的，所述参考文献通过引用其整体并入本文中：Dieffenbach, 编辑, PCR引子：一实验室手册 (PCR Primer: A Laboratory Manual), 第2版 (Cold Spring Harbor Press, New York, 2003)。

[0062] 所述术语“受试者”和“患者”在本文中可互换使用，是指已知患有或可能患有一医学病况或病症，例如癌症的一人类或非人类动物。

[0063] 如本文所用，所述术语“序列读数”是指来自获自一受试者的一样本的一核酸分子的部分或所有的一串的多个核苷酸。一序列读数可以从一核酸片段测序的一短串的多个核苷酸 (例如，20-150)，在一核酸片段的一端或两端的一短串的多个核苷酸，或存在于所述生物样本中的所述整个核酸片段的测序。多个序列读数可以通过本领域已知的各种方法获得。例如，可以通过各种方式获得一序列读数，例如使用多个测序技术或使用多个探针，例如在多个杂交阵列或捕获多个探针中，或多个扩增技术，例如聚合酶连锁反应 (PCR) 或使用一单个引子的线性扩增或恒温扩增。

[0064] 如本文所用，所述术语“读数区段 (read segment)”或“读数”是指任何核苷酸序列，包括从一受试者获得的多个序列读数及/或衍生自一样本的一初始序列读数的多个核苷酸序列。例如，一读数区段可以指一比对的序列读数、一折叠的序列读数或一拼接的读数。此外，一读数区段可以指一单独的核苷酸碱基，例如一单个核苷酸变异体。

[0065] 如本文所用，所述术语“富集”是指增加一样本中一个或多个靶核酸的一比例。因此，一“富集的”样本或测序文库是其中一个或多个靶核酸相对于所述样本中的多个非靶核酸的一比例已增加的一样本或测序文库。

[0066] 通常，应用于多个多核苷酸 (例如“血浆游离RNA”和“血浆游离DNA”) 的所述多个术语“血浆游离”、“循环”和“细胞外”可互换使用，以指存在于来自一受试者或其部分的一样本中的多个多核苷酸，其可以被分离或以其他方式操纵，而无需对最初收集的所述样本应用一裂解步骤 (例如在用于从多个细胞或多个病毒中提取的裂解中)。因此，即使在收集所述受试者的一样本之前，多个细胞游离多核苷酸也未包封或“游离”于它们来源的多个细胞

或多个病毒中。多个细胞游离多核苷酸可以作为细胞死亡(例如细胞凋亡或坏死)或细胞脱落的一副产物产生,将多个多核苷酸释放到周围的多个体液或进入循环中。因此,多个细胞游离多核苷酸可以从血液的一非细胞级分(例如血清或血浆)、从其他多个体液(例如尿液)或从其他多个类型的多个样本的多个非细胞级分中分离。所述术语“细胞游离RNA”或“cfRNA”是指在一受试者体内(例如血流)循环的多个核糖核酸片段,并且可能源自一种或多种健康细胞及/或一种或多种癌细胞。同样地,“细胞游离DNA”或“cfDNA”是指在一受试者体内(例如血流)循环的多个脱氧核糖核酸分子,可能源自一种或多种健康细胞及/或一种或多种癌细胞。

[0067] 所述术语“循环肿瘤RNA”或“ctRNA”是指源自多种肿瘤细胞或其他多个类型的癌细胞的多个核糖核酸片段,其可作为多个生物过程的一结果释放到一受试者体内(例如血流),例如多个垂死细胞的细胞凋亡或坏死,或可以由多个活的肿瘤细胞主动释放。

[0068] 如本文所用,所述术语“暗通道RNA”或“暗通道cfRNA分子”或“暗通道基因”是指其在多个健康细胞中的表达非常低或不存在的一RNA分子或基因。因此,多个暗通道RNA(cfRNA)分子的识别、检测及/或定量在一疾病状态,例如癌症的评估中提高了信噪比(signal-to-noise),并提高了灵敏度和特异性。

[0069] 如本文所用的“治疗(treating或treat)”包括用于在一受试者的病况中获得多个有益或期望结果的任何方法,包括多个临床结果。多个有益或期望的临床结果可以包括,但不限于减轻或改善一种或多种症状或病况、减轻一疾病的所述程度、稳定(即,不恶化)疾病的所述状态、预防一疾病的传染或传播、延迟或减缓疾病进展、改善或缓和所述疾病状态、减少疾病的所述复发、和缓解,无论是部分的还是全部的,无论是可检测的还是不可检测的。换言之,如本文所用,“治疗(treatment)”包括一疾病的任何治愈、改善或预防。治疗可以预防所述疾病的发生;抑制所述疾病的传播;缓解所述疾病的多种症状,完全或部分消除所述疾病的根本原因,缩短一疾病的持续时间,或将这些事情结合起来。

[0070] 如本文所用的“治疗(treating或treat)”包括预防性治疗。多种治疗方法包括向一受试者施用一治疗上有效量的一活性剂。所述施用步骤可以由一单次施用组成或可以包括一系列的多次施用。所述治疗期的所述长短取决于各种因素,例如所述病况的所述严重程度、所述患者的所述年龄、活性剂的所述浓度、所述治疗中使用的所述多种组合物的所述活性、或其一组合。还应理解,用于所述治疗或预防的一药剂的所述有效剂量可以在一特定治疗或预防方案的所述过程中增加或减少。剂量的多个变化可以通过本领域已知的多个标准诊断测定来产生并变得明显。在一些情况下,可能需要长期施用。例如,将多种组合物以足以治疗所述患者的一用量和一持续时间施用于所述受试者。在多个实施例中,所述治疗不是预防性治疗。

[0071] 所述术语“预防(prevent)”与一受试者的一疾病或病况有关,是指所述受试者中一种或多种相应症状的所述发生的一减少。如上所述,所述预防可以是完全的(没有可检测到的多种症状)或部分的,使得观察到的症状更少,及/或发生率低于没有治疗时可能发生的情况。

[0072] “抗癌剂(anti-cancer agent或anticancer agent)”按照其简单的普通含义使用,是指具有多个抗肿瘤性质或能够抑制多个细胞的所述生长或增殖的一组合物(例如化合物、药物、拮抗剂、抑制剂、调节剂)。在一些实施例中,一抗癌剂是一化学治疗剂。在一些

实施例中,一抗癌剂是本文识别的在治疗癌症的多种方法中具有效用的一试剂。在一些实施例中,一抗癌剂是经所述FDA或所述美国以外一国家的类似监管机构批准用于治疗癌症的一试剂。多种抗癌剂的多个示例包括,但不限于多个MEK(例如MEK1、MEK2、或MEK1和MEK2)抑制剂(例如XL518、CI-1040、PD035901、司美替尼(selumetinib)/AZD6244、GSK1120212/曲美替尼(trametinib)、GDC-0973、ARRY-162、ARRY-300、AZD8330、PD0325901、U0126、PD98059、TAK-733、PD318088、AS703026、BAY 869766)、多个烷化剂(例如环磷酰胺(cyclophosphamide)、异环磷酰胺(ifosfamide)、苯丁酸氮芥(chlorambucil)、白消安(busulfan)、氮芥苯丙胺酸(melphalan)、甲基二(氯乙基)胺(mechlorethamine)、尿嘧啶氮芥(uramustine)、噻替哌(thiotepa))、多个亚硝基尿(nitrosourea)、多个氮芥子气(nitrogen mustard)(例如甲基二(氯乙基)胺、环磷酰胺、氮芥苯丁酸、melphalan)、次乙亚胺(ethylenimine)和多个三聚氯胺(methylmelamine)(例如六甲三聚氰胺(hexamethylmelamine)、沙奥特帕)、多个烷基磺酸盐(例如硫酸布他卡因)、多个亚硝基尿(例如双氯乙基亚硝脲(carmustine)、洛莫司汀(lomustine)、司莫司汀(semustine)、链尿佐菌素(streptozocin))、多个三氮烯(triazene)(达卡巴仁(dacarbazine))、多个抗代谢物(例如5-硫唑嘌呤(5-azathioprine)、甲醯四氢叶酸(leucovorin)、凯希得平(capecitabine)、氟达拉滨(fludarabine)、吉西他滨(gemcitabine)、培美曲塞(pemetrexed)、雷替曲塞(raltitrexed)、类叶酸(例如胺甲叶酸(methotrexate))、或多个嘧啶类似物(例如氟尿嘧啶(flourouracil)、氟尿苷(floxouridine)、阿糖胞苷(Cytarabine))、多个嘌呤类似物(例如硫醇嘌呤(mercaptapurine)、硫鸟嘌呤(thioguanine)、喷司他丁(pentostatin))等)、多个植物碱(例如长春新碱(vincristine)、长春花碱(vinblastine)、长春瑞滨(vinorelbine)、长春地辛(vindesine)、鬼臼毒素(podophyllotoxin)、太平洋紫杉醇(paclitaxel)、多烯紫杉醇(docetaxel)等)、多个拓扑异构酶(topoisomerase)抑制剂(例如伊立诺替康(irinotecan)、拓普替康(topotecan)、安吡啶(amsacrine)、依托泊苷(etoposide;VP16)、磷酸依托泊苷(etoposide phosphate)、替尼泊苷(teniposide)等)、多个抗癌抗生素(例如艾霉素(doxorubicin)、阿霉素(adriamycin)、道诺霉素(daunorubicin)、表柔比星(epirubicin)、放线菌素(actinomycin)、博来霉素(bleomycin)、丝裂霉素(mitomycin)、米托蒽醌(mitoxantrone)、普卡霉素(plicamycin)等)、多个基于铂的化合物(例如二氯二胺铂(cisplatin)、奥沙利铂(oxaloplatin)、卡铂(carboplatin))、蒽醌(anthracenedione)(例如米托蒽醌)、取代的尿素(例如羟基尿素(hydroxyurea))、甲胍(methyl hydrazine)衍生物(例如丙卡巴胍(procarbazine))、肾上腺皮质抑制剂(例如米托坦(mitotane)、胺赭精(aminogluthetimide))、多个表鬼臼毒素(epipodophyllotoxin)(例如依托泊苷)、多个抗生素(例如道诺霉素、艾霉素、博来霉素)、多个酶(例如L-天门冬酰胺酶(L-asparaginase))、多个丝裂原活化蛋白激酶信号传导的抑制剂(例如U0126、PD98059、PD184352、PD0325901、ARRY-142886、SB239063、SP600125、BAY 43-9006、渥曼青霉素(wortmannin)、或LY294002、多个Syk抑制剂、多个mTOR抑制剂、多个抗体(例如利妥昔(rituxan))、棉酚(gossypol)、genasense、多酚E(polyphenol E)、Chlorofusin、全反式维甲酸(all trans-retinoic acid, ATRA)、苔藓虫素(bryostatin)、肿瘤坏死因子相关细胞凋亡诱导配体(tumor necrosis factor-related apoptosis-inducing ligand, TRAIL)、

5-氮杂-2'-脱氧胞啶 (5-aza-2'-deoxycytidine)、全反式维甲酸、艾霉素、长春新碱、依托泊苷、吉西他滨、伊马替尼 (imatinib; Gleevec.RTM.)、格尔德霉素 (geldanamycin)、17-N-烯丙氨基-17-去甲氧基格尔德霉素 (17-N-Allylamino-17-Demethoxygeldanamycin, 17-AAG)、黄酮哌啶醇 (flavopiridol)、LY294002、硼替佐米 (bortezomib)、曲妥珠单抗 (trastuzumab)、BAY 11-7082、PKC412、PD184352、20-epi-1、25二羟基维生素D3 (25dihydroxyvitamin D3); 5-乙炔基尿嘧啶 (5-ethynyluracil); 阿比特龙 (abiraterone); 阿柔比星 (acliarubicin); 醯基富烯 (acylfulvene); 阿的培诺 (adecypenol); 阿多来新 (adozelesin); 阿地白介素 (aldesleukin); 多个ALL-TK拮抗剂; 六甲蜜胺 (altretamine); 氨莫司汀 (ambamustine); 醯胺醇 (amidox); 阿米福汀 (amifostine); 氨基乙醯丙酸 (aminolevulinic acid); 氨柔比星 (amrubicin); 安吡啶; 安阁灵 (anagrelide); 阿那曲唑 (anastrozole); 穿心莲内酯 (andrographolide); 多个血管新生抑制剂; 拮抗剂D; 拮抗剂G; 安他利 (antarelix); 抗背侧型蛋白-1 (anti-dorsalizing morphogenetic protein-1); 抗雄性素 (antiandrogen); 摄护腺癌 (prostatic carcinoma); 抗雌激素 (antiestrogen); 抗瘤酮 (antineoplaston); 多个反义寡核苷酸 (antisense oligonucleotide); 阿非迪霉素 (aphidicolin glycinate); 多个细胞凋亡基因调节剂; 多个细胞凋亡调控剂; 类嘌呤酸 (apurinic acid); ara-CDP-DL-PTBA; 精氨酸脱亚氨酶 (arginine deaminase); 奥沙那宁 (asulacrine); 阿他美坦 (atamestane); 阿莫司汀 (atrimustine); 阿新司坦汀1 (axinastatin 1); 阿新司坦汀2 (axinastatin 2); 阿新司坦汀3 (axinastatin 3); 阿扎司琼 (azasetron); 毒伞肽 (azatoxin); 重氮酪胺酸 (azatyrosine); 多个巴卡亭III (baccatin III) 衍生物; 巴诺醇 (balanol); 巴马司他 (batimastat); 多个BCR/ABL拮抗剂; 多个苯并氯 (benzochlorin); 苯甲醯基十字孢碱 (benzoylstauroporine); 多个β-内醯胺类 (beta lactam) 衍生物; β-阿立辛 (beta alethine); β-克拉霉素 (betaclamycin B); 桦木酸 (betulinic acid); bFGF抑制剂; 比卡鲁胺 (bicalutamide); 比生群 (bisantrene); 双氮丙啶精胺 (bisaziridinylspermine); 双萘胺 (bisnafide); 双曲汀A (bistratene A); 比折来新 (bizelesin); 比锐来特 (breflate); 溴匹立明 (bropirimine); 布度盘太 (budotitane); 丁硫氨酸亚砷亚胺 (buthionine sulfoximine); 卡泊三醇 (calcipotriol); 卡弗他丁C (calphostin C); 多个喜树碱 (camptothecin) 衍生物; 金丝雀痘介白素-2 (canarypox IL-2); 凯希得平 (capecitabine); 甲醯胺-氨基-三氮唑 (carboxamide-amino-triazole); 羧胺三唑 (carboxyamidotriazole); CaRest M3; CARN 700; 软骨衍生抑制剂; 卡折来新 (carzelesin); 多个酪蛋白激酶抑制剂 (casein kinase inhibitor, ICOS); 栗精胺 (castanospermine); 天蚕素B (cecropin B); 西曲瑞克 (cetorelix); 克罗林 (chlorins); 氯代喹啉磺醯胺 (chloroquinoxaline sulfonamide); 西卡前列素 (cicaprost); 顺式卟啉 (cis-porphyrin); 克拉屈滨 (cladribine); 多个氯米芬类似物 (clomifene analogue); 克霉唑 (clotrimazole); 柯利霉素A (collismycin A); 柯利霉素B (collismycin B); 康柏斯达汀A4 (combretastatin A4); 康柏斯达汀类似物 (combretastatin analogue); 康纳京尼 (conagenin); 卡那贝西汀816 (crambescidin 816); 克里斯奈托 (crisnatol); 自念珠藻环肽8 (cryptophycin8); 多个自念珠藻环肽A衍生物 (cryptophycin A derivative); 卡拉新A (curacin A); 多个环戊蒽醌 (cyclopentantraquinone); 环普兰姆 (cycloplatam); 西匹霉

素(cypemycin);阿糖胞苷十八烷基磷酸钠(cytarabine ocfosfate);溶细胞因子(cytolytic factor);细胞抑素(cytostatin);达昔单抗(dacliximab);地西他滨(decitabine);去氢膜海鞘素B(dehydrodidemnin B);地洛瑞林(deslorelin);地塞米松(dexamethasone);右异环磷酰胺(dexifosfamide);右雷佐生(dexrazoxane);右维拉帕米(dexverapamil);地吡酮(diaziquone);膜海鞘素B(didemnin B);地多西(didox);二乙基去甲精胺(diethylmorspermine);二氢-5-氮杂胞核苷(dihydro-5-azacytidine);9-二氢紫杉酚(9-dioxamycin);二苯基螺莫司汀(diphenyl spiromustine);多可沙诺(docosanol);多拉司琼(dolasetron);去氧氟尿苷(doxifluridine);屈洛昔芬(droloxifene);屈大麻酚(dronabinol);多卡米新SA(duocarmycin SA);依布硒啉(ebselen);依考莫司汀(ecomustine);依地福新(edelfosine);依决洛单抗(edrecolomab);依氟鸟胺酸(eflornithine);榄香烯(elemene);乙嘧替氟(emitofur);表柔比星;爱普列特(epristeride);雌莫司汀类似物(estrामustine analogue);多个雌激素促效剂(estrogen agonist);多个雌激素拮抗剂(estrogen antagonist);依他硝唑(etanidazole);磷酸依托泊苷;依西美坦(exemestane);法屈唑(fadrozole);法扎拉滨(fazarabine);芬维A胺(fenretinide);非格司亭(filgrastim);非那雄安(finasteride);夫拉平度(flavopiridol);氟卓斯汀(flezelastine);夫斯特隆(flusterone);氟达拉滨;盐酸氟道诺霉素(fluorodaunorubicin hydrochloride);福酚美克(forfenimex);福美坦(formestane);福司曲星(fostriecin);福莫司汀(fotemustine);钆德卟啉(gadolinium texaphyrin);硝酸镓(gallium nitrate);加洛他滨(galocitabine);加尼瑞克(ganirelix);多个明胶酶(gelatinase)抑制剂;吉西他滨;多个谷胱甘肽(glutathione)抑制剂;和普苏姆(hepsulfam);和瑞古林(heregulin);六亚甲基二乙酰胺(hexamethylene bisacetamide);金丝桃素(hypericin);伊班膦酸(ibandronic acid);伊达比星(idarubicin);艾多昔芬(idoxifene);伊决孟酮(idramantone);伊莫福新(ilmofosine);伊洛马司他(ilomastat);多个咪唑吡啶酮(imidazoacridone);咪喹莫特(imiquimod);多个免疫刺激肽(immunostimulant peptide);胰岛素样生长因子-1受体抑制剂;多个干扰素促效剂;多个干扰素;多个介白素;碘苄胍(iobenguane);碘阿霉素(iododoxorubicin);4-甘薯醇(ipomeanol,4-);伊罗普拉(iroplact);伊索拉定(irsogladine);异苯胍唑(isobengazole);太平洋海绵素B(isohomohalicondrin B);伊他司琼(itasetron);杰斯普拉克立德(jasplakinolide);卡哈拉立得F(kahalalide F);三乙酸层状素-N(lamellarin-N triacetate);兰瑞肽(lanreotide);雷那霉素(leinamycin);来格司亭(lenograstim);硫酸香菇多糖(lentinan sulfate);立托斯坦汀(leptolstatin);来曲唑(letrozole);白血病抑制因子;白血球 α 干扰素;亮丙立德+雌激素+孕酮(leuprolide+estrogen+progesterone);亮丙瑞林(leuprorelin);左旋咪唑(levamisole);利阿唑(liarazole);线性聚胺类似物;亲脂双糖肽;多个亲脂铂化合物;立索克林酰胺7(lissoclinamide 7);洛铂(lobaplatin);蚯蚓磷脂(lombricine);洛美曲索(lometrexol);氯尼达明(lonidamine);洛索萘酮(losoxantrone);洛伐他汀(lovastatin);洛索立宾(loxoribine);勒托替康(lurtotecan);镧德卟啉(lutetium texaphyrin);立索茶碱(lysofylline);多个裂解肽(lytic peptide);美坦新(maitansine);麦洛坦汀A(mannostatin A);马立马司他(marimastat);马索罗酚(masoprocol);马司非(maspin);多个基质溶素(matrilysin)抑制

剂;多个基质金属蛋白酶(matrix metalloproteinase)抑制剂;美诺立尔(menogaril);麦尔巴隆(merbarone);美替瑞林(meterelin);蛋胺酶(methioninase);甲氧氯普胺(metoclopramide);MIF抑制剂;米非司酮(mifepristone);米替福新(miltefosine);米立司亭(mirimostim);错配的双股RNA;米托胍脲(mitoguazone);二溴卫矛醇(mitolactol);多个丝裂霉素(mitomycin)类似物;米托萘胺(mitonafide);米托毒素纤维母细胞生长因子-沙泊宁(mitotoxin fibroblast growth factor-saporin);米托蒽醌;莫法罗汀(mofarotene);莫拉司亭(molgramostim);单克隆抗体(monoclonal antibody)、人绒毛膜促性腺激素(human chorionic gonadotrophin);单磷酸基脂A+分枝杆菌细胞壁sk(monophosphoryl lipid A+myobacterium cell wall sk);莫哌达醇(mopidamol);多重耐药基因抑制剂;基于多重肿瘤抑制因子1的疗法;芥末抗癌剂;美卡普罗B(mycaperoxide B);分枝杆菌细胞壁提取物;美瑞普龙(myriaporone);N-乙酰基地那林(N-acetyldinaline);多个N取代之苯甲酰胺(N-substituted benzamide);那法瑞林(nafarelin);纳格瑞替(nagrestip);纳洛酮+喷他佐辛(naloxone+pentazocine);纳普维(napavin);萘特非(naphterpin);那托司亭(nartograstim);奈达铂(nedaplatin);奈莫柔比星(nemorubicin);奈立膦酸(neridronic acid);中性肽链内切酶(neutral endopeptidase);尼鲁米特(nilutamide);尼沙霉素(nisamycin);多个氧化氮调节剂;氧氮化物抗氧化剂;尼挫林(nitrullyn);06-苄基鸟嘌呤(06-benzylguanine);奥曲肽(octreotide);奥克恩(okicenone);多个寡聚核苷酸;奥那司酮(onapristone);昂丹司琼(ondansetron);昂丹司琼;奥拉新(oracin);口服细胞因子诱导剂;奥马铂(ormaplatin);奥沙特隆(osaterone);奥赛力铂(oxaliplatin);厄诺霉素(oxaunomycin);帕诺明(palauamine);棕榈醯基根瘤菌素(palmitoylrhizoxin);帕米膦酸(pamidronic acid);人参三醇(panaxytriol);帕诺米芬(panomifene);帕拉贝新(parabactin);帕折普汀(pazelliptine);培门冬酶(pegaspargase);皮地新(peldesine);戊聚糖聚硫酸钠(pentosan polysulfate sodium);喷司他丁;喷曲唑(pentrozole);全氟溴烷(perflubron);培磷醯胺(perfosfamide);紫苏醇(perillyl alcohol);吩嗪霉素(phenazinomycin);苯乙酸盐(phenylacetate);多个磷酸酶抑制剂;沙培林(picibanil);盐酸毛果芸香碱(pilocarpine hydrochloride);吡柔比星(pirarubicin);吡曲克辛(piritrexim);普来司汀A(placetin A);普来司汀B;血浆素原活化剂抑制剂(plasminogen activator inhibitor);铂错合物;多个铂化合物;铂-三胺错合物(platinum-triamine complex);卟吩姆钠(porfimer sodium);泊非霉素(porfiromycin);泼尼松(prednisone);丙基双吡啶酮(propylbis-acridone);前列腺素J2(prostaglandin J2);多个蛋白酶体抑制剂;基于蛋白质A之免疫调节剂;蛋白激酶C抑制剂;多个微藻蛋白激酶C抑制剂;多个蛋白酪氨酸磷酸酶抑制剂;多个嘌呤核苷磷酸化酶抑制剂;多个紫红素(purpurin);吡唑啉吡啶(pyrazoloacridine);吡哆醛化血红蛋白聚氧化乙烯结合(pyridoxylated hemoglobin polyoxyethylene conjugate);多个raf拮抗剂;雷替曲赛;雷莫司琼ramosetron;多个ras法呢基蛋白转移酶(ras farnesyl protein transferase)抑制剂;多个ras抑制剂;ras-GAP抑制剂;脱甲基瑞替普汀(retelliptine demethylated);依替膦酸铼Re 186(rhenium Re 186etidronate);根瘤菌素(rhizoxin);多个核糖核酸酶(ribozyme);RII维甲醯胺(RII retinamide);罗谷亚胺(rogletimide);罗希吐碱(rohitukine);罗莫肽(romurtide);罗喹

美克(roquinimex);鲁滨吉隆B1(rubiginone B1);鲁泊塞(ruboxyl);沙芬戈(safingol);圣持平(saintopin);SarCNU;沙卡弗托A(sarcophytol A);沙格司亭(sargramostim);多个Sdi 1模拟剂(Sdi 1mimetics);司莫司汀;衰老源性抑制剂1(senescence derived inhibitor 1);多个正义寡核苷酸(sense oligonucleotide);多个信号转导抑制剂;多个信号传导调节剂;单链抗原结合蛋白;西佐喃(sizofuran);索布佐生(sobuzoxane);硼卡钠(sodium borocaptate);苯基乙酸钠(sodium phenylacetate);索佛罗(solverol);促生长因子结合蛋白(somatomedin binding protein);索纳明(sonermin);斯帕福斯酸(sparfosic acid);斯皮卡霉素D(spicamycin D);螺莫司汀;脾脏五肽(splenopentin);海绵抑素1(spongistatin 1);角鲨胺(squalamine);干细胞抑制剂;多个干细胞分裂抑制剂;斯替皮米德(stipiamide);多个基质溶素(stromelysin)抑制剂;索非诺新(sulfinosine);超活性血管活性肠肽拮抗剂(superactive vasoactive intestinal peptide antagonist);磺化偏端霉素(suradista);苏拉明(suramin);苦马豆素(swainsonine);多个合成的糖胺聚多糖(glycosaminoglycan);他莫司汀(tallimustine);他莫昔芬甲碘化物(tamoxifen methiodide);牛磺莫司汀(tauromustine);他扎罗汀(tazarotene);替康兰钠(tecogalan sodium);替加氟(tegafur);碲吡喃鎓(tellurapyrylium);多个端粒酶抑制剂;替莫泊芬(temoporfin);替莫唑胺(temozolomide);替尼泊昔;四氯十氧化物(tetrachlorodecaoxide);替唑明(tetrazomine);噻立拉斯汀(thaliblastine);噻考瑞林(thiocoraline);血小板生成素(thrombopoietin);血小板生成素模拟剂(thrombopoietin mimetic);胸腺法新(thymalfasin);胸腺生长素受体促效剂(thymopoietin receptor agonist);胸腺曲南(thymotrinan);促甲状腺激素;锡乙基初卟啉(tin ethyl etiopurpurin);替拉扎明(tirapazamine);二氯二茂钛(titanocene bichloride);托普升替(topsentin);托瑞米芬(toremifene);全能干细胞因子;多个转译抑制剂;维甲酸(tretinoin);三乙酰尿苷(triacetyluridine);曲西立滨(triciribine);曲美沙特(trimetrexate);曲普瑞林(triptorelin);托烷司琼(tropisetron);妥罗雄脲(turosteride);多个酪氨酸激酶抑制剂;多个替伏汀(tyrphostin);多个UBC抑制剂;乌苯美司(ubenimex);泌尿生殖窦衍生之生长抑制因子(urogenital sinus-derived growth inhibitory factor);多个尿激酶(urokinase)受体拮抗剂;伐普肽(vapreotide);凡瑞林B(variolin B);载体系统,红血球基因疗法;维拉雷琐(velaresol);凡拉明(veramine);多个凡啉(verdin);维替泊芬(verteporfin);长春瑞滨;维萨汀(vinxaltine);维他欣(vitaxin);伏罗唑(vorozole);扎诺特隆(zanoterone);折尼铂(zeniplatin);亚苄维C(zilascorb);净司他丁斯酯(zinostatinstimalamer)、阿霉素、放线菌素(Dactinomycin)、博来霉素、长春花碱、二氯二胺铂、阿西维辛(acivicin);阿柔比星;盐酸阿卡哒唑(acodazole hydrochloride);阿克罗宁(acronine);阿多来新(adozelesin);阿地白介素;六甲蜜胺;安波霉素(ambomycin);乙酸阿美蒽醌(ametanttrone acetate);胺麸精;安吡啶;阿那曲唑;安曲霉素(anthramycin);天门冬酰胺酶;曲林霉素(asperlin);阿扎胞苷(azacitidine);阿扎替帕(azetepa);阿佐霉素(azotomycin);巴马司他;苯唑替哌(benzodepa);比卡鲁胺;盐酸比生群(bisantrene hydrochloride);二甲磺酸双奈法德(bisnafidedimesylate);比折来新;硫酸博来霉素(bleomycin sulfate);布喹那钠(brequinar sodium);溴匹立明;白消安;放线菌素C(cactinomycin);卡普唑酮

(calusterone);卡拉醯胺(caracemide);卡贝替姆(carbetimer);卡铂;双氯乙基亚硝脒;盐酸卡柔比星(carubicin hydrochloride);卡折来新;西地芬戈(cedefingol);苯丁酸氮芥;西罗霉素(cirolemycin);克拉屈滨;甲磺酸雷斯托(crisnatol mesylate);环磷醯胺;阿糖胞苷;达卡巴仁;盐酸道诺霉素(daunorubicin hydrochloride);地西他滨;右奥马铂(dexormaplatin);地扎胍宁(dezaguanine);甲磺酸地扎胍宁(dezaguanine mesylate);地吡酮;艾霉素;盐酸艾霉素(doxorubicin hydrochloride);屈洛昔芬;柠檬酸屈洛昔芬(droloxifene citrate);丙酸曲他雄酮(dromostanolone propionate);偶氮霉素(duazomycin);依达曲沙(edatrexate);盐酸依氟鸟胺酸(eflornithine hydrochloride);依沙芦星(elsamitrucin);恩洛铂(enloplatin);恩普氨酯(enpromate);依匹哌啉(epipropidine);盐酸表柔比星(epirubicin hydrochloride);厄布洛唑(erbulozole);盐酸依索比星(esorubicin hydrochloride);雌莫司汀;雌莫司汀磷酸钠;依他硝唑;依托泊苷;磷酸依托泊苷;埃托宁(etoprine);盐酸法屈唑(fadrozole hydrochloride);法扎拉滨;芬维A胺;氟尿苷(floxuridine);磷酸氟达拉滨(fludarabine phosphate);氟尿嘧啶;氟西他滨(fluorocitabine);磷喹酮(fosquidone);福司曲星钠(fostriecin sodium);吉西他滨;盐酸吉西他滨(gemcitabine hydrochloride);羟基尿素;盐酸伊达比星(idarubicin hydrochloride);异环磷醯胺;伊莫福新;介白素II(包括重组介白素II,或rIL.sub.2),干扰素 α -2a;干扰素 α -2b;干扰素 α -n1;干扰素 α -n3;干扰素 β -1a;干扰素 γ -1b;伊罗夫文(iproplatin);盐酸伊立诺替康(irinotecan hydrochloride);乙酸兰瑞肽(lanreotide acetate);来曲唑;乙酸亮丙立德(leuprolide acetate);盐酸利阿唑(liarozole hydrochloride);洛美曲索钠(lometrexol sodium);洛莫司汀;盐酸洛索蒽醌(losoxantrone hydrochloride);马索罗酚;美登素(maytansine);盐酸甲氮芥(mechlorethamine hydrochloride);乙酸甲地孕酮(megestrol acetate);乙酸甲烯磁醇(melengestrol acetate);氮芥苯丙胺酸;美诺立尔;硫醇嘌呤;胺甲叶酸;胺甲叶酸钠(methotrexate sodium);氯苯氨啉(metoprime);美妥替哌(meturedopa);米丁度胺(mitindomide);米托卡星(mitocarcin);丝裂红素(mitocromin);米托洁林(mitogillin);丝裂马菌素(mitomalcin);丝裂霉素;丝裂帕菌素(mitosper);米托坦;盐酸米托蒽醌(mitoxantrone hydrochloride);麦考酚酸(mycophenolic acid);诺考达唑(nocodazole);诺加霉素(nogalamycin);奥马铂;奥昔舒仑(oxisuran);培门冬酶;佩里霉素(peliomycin);戊氮芥(pentamustine);硫酸培洛霉素(peplomycin sulfate);培磷醯胺;哌泊溴烷(pipobroman);哌泊舒凡(piposulfan);盐酸吡罗蒽醌(piroxantrone hydrochloride);普卡霉素;普洛美坦(plomestane);吡吩姆钠;泊非霉素;泼尼莫司汀(prednimustine);盐酸丙卡巴肼(procarbazine hydrochloride);嘌呤霉素(puromycin);盐酸嘌呤霉素(puromycin hydrochloride);吡唑呋喃菌素(pyrazofurin);利波腺苷(riboptine);罗谷亚胺;沙芬戈;盐酸沙芬戈(safingol hydrochloride);司莫司汀;双曲秦(simtrazene);磷乙醯天冬胺酸钠(sparfosate sodium);稀疏霉素(sparsomycin);盐酸螺锗(spirogermanium hydrochloride);螺莫司汀;螺铂(spiroplatin);链黑霉素(streptonigrin);链尿佐菌素;磺氯苯脲(sulofenur);他利霉素(talisomycin);替康兰钠;替加氟;盐酸替螺酮(teloxantrone hydrochloride);替莫泊芬;替尼泊苷;替罗昔隆(teroxirone);睾内酯(testolactone);硫咪嘌呤(thiamiprine);硫鸟嘌呤;噻替哌;噻唑

呋林(tiazofurin);替拉扎明;柠檬酸托瑞米芬(toremifene citrate);乙酸曲托笼(trestolone acetate);磷酸曲西立滨(triciribine phosphate);曲美沙特;葡萄糖醛酸曲美沙特(trimetrexate glucuronate);曲普瑞林;盐酸妥布氯唑(tubulozole hydrochloride);尿嘧啶氮芥(uracil mustard);乌瑞替帕(uredepa);伐普肽;维替泊芬;硫酸长春花碱(vinblastine sulfate);硫酸长春新碱(vincristine sulfate);长春地辛;硫酸长春地辛(vindesine sulfate);硫酸长春匹定(vinepidine sulfate);硫酸长春甘酯(vinglycinatate sulfate);硫酸长春罗新(vinleurosine sulfate);酒石酸长春瑞滨(vinorelbine tartrate);硫酸长春罗定(vinrosidine sulfate);硫酸长春利定(vinzolidine sulfate);伏罗唑;折尼铂;净司他丁;盐酸佐柔比星(zorubicin hydrochloride)、将多个细胞阻滞在所述G2-M期及/或调节多个微小管的所述形成或稳定性的多个试剂、(例如Taxol.TM(即,太平洋紫杉醇))、Taxotere.TM、包含所述紫杉烷(taxane)骨架的多个化合物、厄布洛唑(即,R-55104)、尾海兔素10(Dolastatin 10)(即,DLS-10和NSC-376128)、羟乙基磺酸米伏布林(Mivobulin isethionate)(即,作为CI-980)、长春新碱、NSC-639829、圆皮海绵内酯(Discodermolide)(即,作为NVP-XX-A-296)、ABT-751(Abbott,即E-7010)、多个奥托里汀(Altorhyrtin)(例如奥托里汀A和奥托里汀C)、多个海绵抑制素(Spongistatin)(例如海绵抑制素1、海绵抑制素2、海绵抑制素3、海绵抑制素4、海绵抑制素5、海绵抑制素6、海绵抑制素7、海绵抑制素8、和海绵抑制素9)、盐酸西马多丁(Cemadotin hydrochloride)(即,LU-103793和NSC-D-669356)、多个埃博霉素(Epothilone)(例如埃博霉素A、埃博霉素B、埃博霉素C(即,脱氧糖埃博霉素A(desoxyepothilone A)或dEpoA)、埃博霉素D(即,KOS-862、dEpoB和脱氧糖埃博霉素B)、埃博霉素E、埃博霉素F、埃博霉素B N-氧化物、埃博霉素A N-氧化物、16-氮杂-埃博霉素B(16-aza-epothilone B)、21-胺基埃博霉素B(21-aminoepothilone B)(即,BMS-310705)、21-羟基埃博霉素D(21-hydroxyepothilone D)(即,羟基埃博霉素F和dEpoF)、26-氟埃博霉素(26-fluoroepothilone)、澳瑞他汀PE(Auristatin PE)(即,NSC-654663)、舒碧利婷(Soblidotin)(即,TZT-1027)、LS-4559-P(Pharmacia,即LS-4577)、LS-4578(Pharmacia,即LS-477-P)、LS-4477(Pharmacia)、LS-4559(Pharmacia)、RPR-112378(Aventis)、硫酸长春新碱、DZ-3358(Daiichi)、FR-182877(Fujisawa,即WS-9885B)、GS-164(Takeda)、GS-198(Takeda)、KAR-2(Hungarian Academy of Sciences)、BSF-223651(BASF,即ILX-651和LU-223651)、SAH-49960(Lilly/Novartis)、SDZ-268970(Lilly/Novartis)、AM-97(Armad/Kyowa Hakko)、AM-132(Armad)、AM-138(Armad/Kyowa Hakko)、IDN-5005(Indena)、Cryptophycin 52(即,LY-355703)、AC-7739(Ajinomoto,即AVE-8063A和CS-39.HCl)、AC-7700(Ajinomoto,即AVE-8062、AVE-8062A、CS-39-L-Ser.HCl和RPR-258062A)、Vitilevuamide、微管溶素A(Tubulysin A)、卡那丹醇(Canadensol)、矢车菊黄素(Centaureidin)(即,NSC-106969)、T-138067(Tularik,即T-67、TL-138067和TI-138067)、COBRA-1(Parker Hughes研究所,即DDE-261和WHI-261)、H10(堪萨斯州立大学)、H16(堪萨斯州立大学)、Oncocidin A1(即,BTO-956和DIME)、DDE-313(Parker Hughes研究所)、非加诺德B(Fijianolide B)、Laulimalide、SPA-2(Parker Hughes研究所)、SPA-1(Parker Hughes研究所,即T SPIKET-P)、3-IAABU(Cytoskeleton/西奈山医学院(Mt.Sinai School of Medicine),即MF-569)、诺司卡品(Narcosine,也称作NSC-5366)、那可汀(Nascapine)、

D-24851 (Asta Medica)、A-105972 (Abbott)、Hemiasterlin、3-BAABU (Cytoskeleton/西奈山医学院,即MF-191)、TMPN (亚利桑那州立大学)、基丙酮酸瓦纳多 (Vanadocene acetylacetonate)、T-138026 (Tularik)、Monsatrol、乐诺辛 (lnanocine,即NSC-698666)、3-IAABE (Cytoskeleton/西奈山医学院)、A-204197 (Abbott)、T-607 (Tuiarik,即T-900607)、RPR-115781 (Aventis)、多个五加素 (Eleutherobin,例如去甲基五加素 (Desmethyleleutherobin)、十乙基五加素 (Desaetyeleleutherobin)、异五加素A (lsoeleutherobin A)和Z-五加素 (Z-Eleutherobin))、卡利巴昔 (Caribaeoside)、卡里贝奥林 (Caribaeolin)、软海绵素B (Halichondrin B)、D-64131 (Asta Medica)、D-68144 (Asta Medica)、含氯环肽A (Diazonamide A)、A-293620 (Abbott)、NPI-2350 (Nereus)、根薯酮内酯A (Taccalonolide A)、TUB-245 (Aventis)、A-259754 (Abbott)、Diozostatin、(-)-Phenylahistin (即,NSCL-96F037)、D-68838 (Asta Medica)、D-68836 (Asta Medica)、肌基质蛋白B (Myoseverin B)、D-43411 (Zentaris,即D-81862)、A-289099 (Abbott)、A-318315 (Abbott)、HTI-286 (即,SPA-110,三氟乙酸盐 (trifluoroacetate salt)) (Wyeth)、D-82317 (Zentaris)、D-82318 (Zentaris)、SC-12983 (NCI)、瑞舒伐他汀磷酸钠 (Resverastatin phosphate sodium)、BPR-0Y-007 (国家卫生研究院 (National Health Research Institutes))、及SSR-250411 (Sanofi)、多个类固醇 (例如地塞米松)、非那雄安、多个芳香环转化酶 (aromatase) 抑制剂、多个性腺激素释放素促进剂 (gonadotropin-releasing hormone agonist, GnRH) 例如戈舍促性腺激素释放激素 (goserelin) 或亮丙立德、多个肾上腺皮质类固醇 (adrenocorticosteroid) (例如泼尼松)、多个黄体素 (progestin) (例如己酸羟孕酮 (hydroxyprogesterone caproate)、乙酸甲地孕酮、乙酸甲羟孕酮 (medroxyprogesterone acetate))、多个雌激素 (例如己烯雌酚 (diethylstilbestrol)、乙炔雌二醇 (ethinyl estradiol))、抗雌激素 (例如他莫昔芬)、多个雄性素 (例如丙酸睾酮 (testosterone propionate)、氟羟甲基睾酮 (fluoxymesterone))、抗雄性素 (例如氟塔醯胺 (flutamide))、多个免疫刺激剂 (immunostimulant) (例如卡介苗 (Bacillus Calmette-Guérin, BCG)、左旋咪唑、介白素-2、 α -干扰素等)、多个单克隆抗体 (例如抗CD20、抗HER2、抗CD52、抗HLA-DR和抗VEGF单克隆抗体)、多个免疫毒素 (immunotoxin) (例如抗CD33单克隆抗体-卡奇霉素 (calicheamicin) 偶联物、抗CD22单克隆抗体-假单胞菌属 (pseudomonas) 外毒素 (exotoxin) 偶联物等)、放射免疫疗法 (例如与¹¹¹In、⁹⁰Y或¹³¹I等偶联的抗CD20单克隆抗体)、雷公藤甲素 (triptolide)、高三尖杉酯碱 (homoharringtonine)、放线菌素、艾霉素、表柔比星、拓普替康、伊曲康唑 (itraconazole)、长春地辛、西立伐他汀 (cerivastatin)、长春新碱、脱氧腺苷 (deoxyadenosine)、舍曲林 (sertraline)、匹伐他汀 (pitavastatin)、伊立诺替康、氯法齐明 (clofazimine)、5-壬基氧基色胺草酸盐 (5-nonyloxytryptamine)、维莫非尼 (vemurafenib)、达拉非尼 (dabrafenib)、厄洛替尼 (erlotinib)、吉非替尼 (gefitinib)、多个EGFR抑制剂、表皮生长因子受体 (epidermal growth factor receptor, EGFR) 靶标疗法或治疗性 (例如吉非替尼 (IressaTM)、厄洛替尼 (TarcevaTM)、西妥昔单抗 (cetuximab) (ErbixTM)、拉帕替尼 (lapatinib) (TykerbTM)、帕尼单抗 (panitumumab) (VectibixTM)、凡德他尼 (vandetanib) (CaprelsaTM)、阿法替尼 (afatinib) /BIBW2992、CI-1033/卡奈替尼 (canertinib)、来那替尼 (neratinib) /HKI-272、CP-724714、TAK-285、AST-1306、ARRY334543、ARRY-380、AG-1478、达克替尼 (dacomitinib) /PF299804、OSI-420/去甲

基厄洛替尼 (desmethyl erlotinib)、AZD8931、AEE788、培利替尼 (pelitinib)/EKB-569、CUDC-101、WZ8040、WZ4002、WZ3146、AG-490、XL647、PD153035、BMS-599626)、索拉非尼 (sorafenib)、伊马替尼、舒尼替尼 (sunitinib)、达沙替尼 (dasatinib) 等。

[0073] 如本文所用,一“表观遗传 (epigenetic) 抑制剂”是指一表观遗传过程的一抑制剂,例如DNA甲基化 (一DNA甲基化抑制剂) 或多个组蛋白 (histone) 的修饰 (一组蛋白修饰抑制剂)。一表观遗传抑制剂可以是一组蛋白去乙酰酶 (histone-deacetylase, HDAC) 抑制剂、一DNA甲基转移酶 (DNA methyltransferase, DNMT) 抑制剂、一组蛋白甲基转移酶 (histone methyltransferase, HMT) 抑制剂、一组蛋白去甲基酶 (histone demethylase, HDM) 抑制剂或一组蛋白乙酰转移酶 (histone acetyltransferase, HAT)。多种HDAC抑制剂的多个示例包括伏立诺他 (Vorinostat)、罗米地辛 (romidepsin)、CI-994、贝利司他 (Belinostat)、帕比司他 (Panobinostat)、吉诺司他 (Givinostat)、恩替诺特 (Entinostat)、莫西妥他汀 (Mocetinostat)、SRT501、CUDC-101、JNJ-26481585或PCI24781。多种DNMT抑制剂的多个示例包括阿扎胞苷和地西他滨。多种HMT抑制剂的多个示例包括EPZ-5676。多种HDM抑制剂的多个示例包括巴吉林 (pargyline) 和反苯环丙胺 (tranylcypromine)。多种HAT抑制剂的多个示例包括CCT077791和山竹醇 (garcinol)。

[0074] 一“多重激酶抑制剂”是至少一种蛋白激酶的一小分子抑制剂,包括多个酪氨酸蛋白激酶和多个丝氨酸/苏氨酸激酶。一多重激酶抑制剂可以包括一单个激酶抑制剂。多个多重激酶抑制剂可阻断磷酸化。多个多重激酶抑制剂可以作为多个蛋白激酶的多个共价修饰剂。多个多重激酶抑制剂可以结合激酶活性位点或抑制蛋白激酶活性的一二级或三级位点。一多重激酶抑制剂可以是一抗癌多重激酶抑制剂。多个示例性抗癌多重激酶抑制剂包括达沙替尼、舒尼替尼、厄洛替尼、贝伐珠单抗 (bevacizumab)、瓦他拉尼 (vatalanib)、维莫非尼、凡德他尼、卡博替尼 (cabozantinib)、普纳替尼 (poatinib)、阿昔替尼 (axitinib)、鲁索利替尼 (ruxolitinib)、瑞戈非尼 (regorafenib)、克唑替尼 (crizotinib)、伯舒替尼 (bosutinib)、西妥昔单抗、吉非替尼、伊马替尼、拉帕替尼、乐伐替尼 (lenvatinib)、木利替尼 (mubritinib)、尼洛替尼 (nilotinib)、帕尼单抗、帕唑帕尼 (pazopanib)、曲妥珠单抗或索拉非尼。

[0075] 如本文所用,所述术语“约”是指包括所述指定值的多个值的一范围,本领域普通技术人员会认为其合理地类似于所述指定值。在多个实施例中,约表示使用本领域通常可接受的多个测量值在一标准偏差内。在多个实施例中,约表示延伸至所述指定值的 $\pm 10\%$ 的一范围。在多个实施例中,约包括所述指定值。

[0076] 本公开主题的多个方面包括用于基于从一受试者中对一个或多个RNA分子的分析检测所述受试者中的一疾病状态 (例如癌症的一存在或不存在) 及/或所述疾病的一起源组织的多种方法。在一些实施例中,一种用于检测一受试者中的一疾病状态的方法,包括从所述受试者中分离一生物测试样本,其中所述生物测试样本包括多个血浆游离核糖核酸 (cfRNA) 分子、从所述生物测试样本中提取所述多个cfRNA分子、对所述多个提取的cfRNA分子测序,以产生多个序列读数、执行一过滤程序,以产生源自一个或多个健康细胞的多个序列读数的一排除群体以及多个序列读数的一非排除群体、及/或对所述多个非排除的序列读数执行一量化程序。在多个实施例中,所述多种方法包括当所述量化程序产生超过一阈值的一值时,检测所述受试者中的所述疾病状态。在多个实施例中,检测高于一阈值的一个

或多个非排除的序列读数的步骤包括 (i) 检测、(ii) 高于背景检测、及/或 (iii) 在一水平检测, 所述水平比未患有所述病况的多位受试者中的多个相应序列读数的一水平高。在各种实施例中, 所述阈值是一整数, 范围从大约或正好1至大约或正好10, 例如大约或正好2、3、4、5、6、7、8或大约或正好9。在一些实施例中, 所述阈值是一非整数值, 范围从大约或正好0.1至大约或正好0.9, 例如大约或正好0.2、0.3、0.4、0.5、0.6、0.7或大约或正好0.8。

[0077] 在一些实施例中, 所述多种方法涉及使用测序程序来检测和定量从一生物测试样本中提取的所述多个cfRNA分子。例如, 在各种实施例中, 一测序程序涉及对所述多个cfRNA分子执行一反转录程序, 以产生多个cDNA/RNA杂合分子、将所述多个杂合分子的所述RNA降解, 以产生多个单股cDNA分子模板、合成来自所述多个单股cDNA分子模板的多个双股DNA分子、将多个双股DNA转接子(adapter) 连接到所述多个双股DNA分子上产生一测序文库、以及对所述测序文库的至少一部分进行一测序程序, 以获得多个序列读数。在各种实施例中, 合成所述多个双股DNA分子的步骤涉及执行一股置换(strand-displacement) 反转录酶程序。

[0078] 在一些实施例中, 所述多种方法利用多个全转录组测序程序。在其他多个实施例中, 一测序程序涉及一标靶测序程序, 其中在制备一测序文库之前从所述生物测试样本中富集所述多个cfRNA分子的一个或多个。根据此实施例, 一个或多个指示所述疾病状态的cfRNA分子被标靶用于富集。例如, 在一些实施例中, 所述一个或多个标靶cfRNA分子衍生自选自自由以下基因所组成的群组的一个或多个基因: AGR2、BPIFA1、CASP14、CSN1S1、DISP2、EIF2D、FABP7、GABRG1、GNAT3、GRHL2、HOXC10、IDI2-AS1、KRT16P2、LALBA、LINC00163、NKX2-1、OPN1SW、PADI3、PTPRZ1、ROS1、S100A7、SCGB2A2、SERPINB5、SFTA3、SFTPA2、SLC34A2、TFF1、VTCN1、WFDC2、MUC5B、SMIM22、CXCL17、RNU1-1和KLK5, 并且可以包括其任意组合。在一些实施例中, 一个或多个标靶RNA分子衍生自选自自由以下基因所组成的群组的一个或多个基因: ROS1、NKX2-1、GGTLC1、SLC34A2、SFTPA2、BPIFA1、SFTA3、GABRG1、AGR2、GNAT3、MUC5B、SMIM22、CXCL17和WFDC2, 并且可以包括其任意组合。在一些实施例中, 一个或多个标靶RNA分子衍生自选自自由以下基因所组成的群组的一个或多个基因: SCGB2A2、CSN1S1、VTCN、FABP7、LALBA、RNU1-1、OPN1SW、CASP14、KLK5和WFDC2, 并且可以包括其任意组合。在一些实施例中, 一个或多个标靶RNA分子衍生自选自自由以下基因所组成的群组的一个或多个基因: CASP14、CRABP2、FABP7、SCGB2A2、SERPINB5、TRGV10、VGLL1、TFF1和AC007563.5, 并且可以包括其任意组合。在还有其他多个实施例中, 所述标靶RNA分子衍生自所述AKR1B10、C3及/或PIEX02基因。

[0079] 本公开主题的多个方面涉及一个或多个暗通道RNA分子的分析, 其在多位健康受试者的所述血浆中的表达非常低或不存在。由于它们或多位健康受试者的所述血浆中的低表达水平, 多个暗通道RNA分子提供了可以与本发明多种方法结合使用的一高信噪比。

[0080] 本公开主题的一些方面涉及多个过滤程序, 这些过滤程序用于产生源自一个或多个健康细胞的多个序列读数的一排除群体, 以及用于多个后续分析的多个序列读数的一非排除群体。在各种实施例中, 所述过滤程序涉及将从所述生物测试样本中提取的所述多个cfRNA分子的每个序列读数与多个RNA序列的一对照数据集进行比较、识别与多个RNA序列的所述对照数据集中的多个序列读数相匹配的一个或多个序列读数、以及将与多个RNA序列的所述对照数据集中的所述一个或多个序列读数相匹配的每个序列读数置于多个序列读数的所述排除群体中。

[0081] 在一些实施例中,多个RNA序列的一对照数据集包括从一位或多位健康受试者中获得多个序列读数。在一些实施例中,多个RNA序列的一对照数据集包括从所述受试者的多个血球细胞中获得多个序列读数。例如,在一些实施例中,从一受试者的多个白血球细胞(white blood cell,WBC)中获得多个序列读数。

[0082] 多个生物样本

[0083] 在各种实施例中,本公开涉及从一受试者中获得一测试样本,例如一生物测试样本,例如一组织及/或体液样本,用于分析其中的多个核酸(例如多个cfRNA分子)。可以以任何临床上可接受的方式收集根据本发明多个实施例的多个样本。任何怀疑含有多个核酸的样本都可以与本发明的所述多种方法结合使用。在一些实施例中,一样本可以包括一组织、一体液或其一组合。在一些实施例中,从一健康受试者中收集一生物样本。在一些实施例中,从已知患有一特定疾病或病症(例如一特定癌症或肿瘤)的一受试者中收集一生物样本。在一些实施例中,从怀疑患有一特定疾病或病症的一受试者中收集一生物样本。

[0084] 如本文所用,所述术语“组织”是指大量连接的细胞及/或细胞外基质材料。通常与本发明多种方法结合使用的多种组织的多个非限制性示例包括皮肤、毛发、多个指甲、子宫内膜组织、鼻道组织、中枢神经系统(central nervous system,CNS)组织、神经组织、眼组织、肝组织、肾组织、胎盘组织、乳腺组织、胃肠组织、肌肉骨骼组织、泌尿生殖组织、骨髓等,其衍生自例如,一人类或非人类哺乳动物。根据本发明的多个实施例的多个组织样本可以以本领域已知的任何组织样本类型的所述形式制备和提供,例如但不限于福马林固定后包埋在石蜡块中(formalin-fixed paraffin-embedded,FFPE)、新鲜的和新鲜冷冻的(fresh frozen,FF)组织样本。

[0085] 如本文所用,所述术语“流体”和“生物流体”是指衍生自一受试者,例如一人类或非人类哺乳动物的一液体材料。通常与本发明多种方法结合使用的多种体液的多个非限制性示例包括黏液、血液、血浆、血清、多个血清衍生物、滑液、淋巴液、胆汁、痰、唾液、汗液、泪液、痰液、羊水、月经液、阴道液、精液、尿液、脑脊液(cerebrospinal fluid,CSF),例如腰椎或心室CSF、胃液、包括一种或多种衍生自一鼻腔、咽喉或口腔拭子的材料的一液体样本、包括一种或多种衍生自一灌洗程序的材料的一液体样本,例如一腹膜、胃、胸或导管灌洗程序等。

[0086] 在一些实施例中,一样本可以包括一细针吸出物或活检组织。在一些实施例中,一样本可以包括含有多个细胞或生物材料的多个培养基。在一些实施例中,一样本可以包含一血凝块,例如,在去除所述血清后从全血中获得的一血凝块。在一些实施例中,一样本可以包括粪便。在一个优选实施例中,从全血中抽取一样本。在一个方面,仅使用一全血样本的一部分,例如血浆、多个红血球细胞、多个白血球细胞和多个血小板。在一些实施例中,结合本发明多种方法将一样本分离成两个或多个组成部分。例如,在一些实施例中,将一全血样本分离成血浆、红血球细胞、白血球细胞和血小板组分。

[0087] 在一些实施例中,一样本不仅包括来自提取所述样本的所述受试者的多个核酸,还包括来自一个或多个其他生物体,例如在取样时所述受试者体内存在的病毒DNA/RNA的多个核酸。

[0088] 可以根据本领域已知的任何合适的方法从一样本中提取核酸,并且所述提取的核酸可以与本文所述的多种方法结合使用。参见例如Maniatis等人,分子克隆:一实验室手册

(Molecular Cloning: A Laboratory Manual), Cold Spring Harbor, N.Y., pp. 280-281, 1982, 其内容通过引用其整体并入本文中。在一个优选实施例中, 从一样本中提取血浆游离核糖核酸(例如cfRNA)。

[0089] 在多个实施例中, 所述样本是一“匹配的”或“配对的”样本。一般而言, 所述术语“匹配的样本”和“配对的样本”是指从相同受试者中收集的多个不同类型的一对的多个样本, 优选地是在大约相同时间(例如作为一个单个程序或办公室访问的部分, 或在同一天)。在多个实施例中, 所述多个不同类型是一组织样本(例如癌症组织, 如一切除或活检样本中的)和一生物流体样本(例如血液或一血液级分)。所述术语还可用于指衍生自所述匹配的样本的多个多核苷酸(例如从一癌症组织中提取的多个多核苷酸, 与来自一匹配的生物流体样本的多个血浆游离多核苷酸配对)或其多个测序读数。在多个实施例中, 分析多个配对的样本, 例如在识别多个癌症生物标记物时。多个配对的样本可以来自在多个不同时间收集的相同个体(例如来自一癌症早期的一配对的样本中的、以及来自一癌症晚期的一配对的样本中的)、来自相同或多个不同的多个不同个体、或这些的一组合。在多个实施例中, 所述多个匹配的样本来自多位不同的受试者。在多个实施例中, 所述多个匹配的样本来自具有相同癌症类型和任选相同癌症阶段的多位受试者。

[0090] 示例测定方案

[0091] 图1是根据一个实施例的一种用于制备用于测序的一核酸样本的方法100的流程图。所述方法100包括但不限于以下多个步骤。例如, 所述方法100的任何步骤可以包括用于质量控制的一量子步骤或本领域技术人员已知的其他多个实验室测定程序。

[0092] 在步骤110中, 从一受试者中提取一核糖核酸(RNA)样本。所述RNA样本可以包括所述整个人类转录组, 或所述人类转录组的任何子集。所述样本可以从已知患有或怀疑患有一疾病(例如癌症)的一受试者中提取。所述样本可以包括血液、血浆、血清、尿液、粪便、唾液、多个其他类型的体液或其任意组合。在一些实施例中, 用于抽取一血液样本的多种方法(例如注射器或手指点刺)可能比用于获得一组织活检的多个程序侵入性更小, 这可能需要手术。所述提取的样本还可包括cfDNA。如果一受试者患有一疾病(例如癌症), 则一提取的样本中的cfRNA可能以一可检测水平存在以用于诊断。

[0093] 在步骤120中, 包括多个RNA分子的所述核酸样本可选地用一DNase酶处理。所述DNase可以从所述核酸样本中去除多个DNA分子, 以减少所述多个RNA分子的DNA污染。多个RNA分子转换为DNA后, 可能很难区分所述核酸样本中最初发现的所述RNA转换的DNA和基因组DNA。应用所述DNase可以对源自cfRNA的多个分子进行靶扩增。所述DNase过程可以包括添加一DNase缓冲液、使用一离心机将施加的所述样本与DNase混合以及培养的多个步骤。在一些实施例中, 步骤120包括基于所述Qiagen QIAamp循环核酸手册(Qiagen QIAamp Circulating Nucleic Acid Handbook)中描述的所述DNase处理方案的一个或多个过程。

[0094] 在步骤130中, 一反转录酶用于将所述核酸样本中的所述多个RNA分子转变为互补DNA(complementary DNA, cDNA)。所述反转录酶过程可以包括一第一股合成步骤(通过反转录产生一cDNA股)、降解所述RNA股以产生一单股cDNA分子, 以及使用一聚合酶从所述单股cDNA分子合成一双股DNA分子。在第一股合成期间, 一引子与一RNA分子的所述3'端黏合。在第二股合成期间, 一不同的引子与所述cDNA分子的所述3'端黏合。

[0095] 在步骤140中, 制备一测序文库。例如, 如本领域所周知的, 多个转接子可以连接到

一dsDNA分子的一端或两端,以制备用于测序的一文库。在一个实施例中,使用的所述多个转接子可以包括一个或多个测序寡核苷酸,用于随后的丛集(cluster)生成及/或测序(例如用于合成式测序(sequencing by synthesis,SBS)(Illumina,圣地亚哥,加利福尼亚州)的已知P5和P7序列)。在另一实施例中,所述转接子包括一样本特异性指标序列,使得在文库制备之后,所述文库可以与由多个单独样本制备的一个或多个其他文库组合,从而允许多重测序。所述样本特异性指标序列可以包括一长度为约或正好2nt至约或正好20nt、约或正好2nt至约或正好10nt、约或正好2nt至约或正好8nt,或约或正好2nt至约或正好6nt。在另一实施例中,所述样本特异性指标序列可以包括长度大于约或正好2、3、4、5、6、7或8个核苷酸(nt)的一短寡核苷酸序列。

[0096] 可选地,在文库制备期间,可以通过转接子连接将多个唯一分子标识符(unique molecular identifier,UMI)添加到所述样本中的所述多个核酸分子。所述多个UMI是多个短核酸序列(例如4-10个碱基对),在转接子连接期间添加到多个核酸片段的一端或两端。在一些实施例中,多个UMI是多个简并碱基对,其充当一独特标签,可用于识别源自一特定核酸片段的多个序列读数。在转接子连接后的PCR扩增期间,所述多个UMI与所述附接的核酸片段一起复制,这提供了一种在下游分析中识别来自相同原始核酸分子的多个序列读数的方式。

[0097] 对于包括RNA的标靶测序的多个实施例,在步骤150中,多个标靶核酸序列从所述文库中富集。在富集期间,多个杂交探针(在本文中也称为“多个探针”)用于标靶和下拉多个核酸片段,这些核酸片段可提供有关一疾病(例如癌症)的所述存在或不存在、疾病状态(例如癌症状态)、或一疾病分类(例如癌症类型或起源组织)的信息。对于一给定的工作流程,所述多个探针可以设计为与一目标(互补)核酸股(例如从RNA转变的一DNA股)黏合(或杂合)。所述多个探针的长度范围可以从10s、100s或1000s碱基对。在一个实施例中,所述多个探针是基于基因套组(gene panel)设计的,以分析怀疑对应于某些癌症或其他多个类型的多个疾病的所述基因组(例如所述人类或另一种生物体的基因组)的多个特定目标区域。此外,所述多个探针可以覆盖一目标区域的多个重叠部分。在其他多个实施例中,可以在使用反转录酶(未显示)将所述多个RNA分子转变为多个cDNA股之前使用多个杂交探针来富集多个标靶RNA分子。一般而言,本领域中任何已知的方法都可以用于分离和富集多个探针杂交的标靶核酸。例如,如本领域所周知的,可以将一生物素部分添加到所述多个探针的所述5'-端(即,生物素化的(biotinylated)),以促进使用一链霉亲和素(streptavidin)包覆的表面(例如多个链霉亲和素包覆的珠子)分离与多个探针杂交的多个标靶核酸。

[0098] 另外地,对于标靶测序,在步骤160中,从所述富集的核酸样本中产生多个序列读数。测序数据可以通过本领域已知的方法从所述多个富集的DNA序列(即,从多个RNA序列衍生或转变的多个DNA序列)获得。例如,所述方法100可以包括多个次世代测序(next generation sequencing,NGS)技术,包括合成技术(Illumina)、焦磷酸测序(pyrosequencing;454Life Sciences)、离子半导体技术(Ion Torrent测序)、单分子实时测序(single-molecule real-time sequencing;Pacific Biosciences)、连接测序(SOLiD测序)、纳米孔测序(Oxford Nanopore Technologies)或双端测序(paired-end sequencing)。在一些实施例中,使用具有多个可逆染料终止子(terminator)的合成式测序进行大规模平行测序。

[0099] 在其他多个实施例中,例如,在全转录组测序方法中(例如代替标靶测序),在步骤170中,从所述核酸样本中消耗丰富的RNA种类。例如,在一些实施例中,可以消耗核糖体RNA(ribosomal RNA,rRNA)及/或传递RNA(transfer RNA,tRNA)种类。多个可用的商业试剂盒,例如RiboMinusTM(ThermoFisher Scientific)或AnyDeplete(NuGen),可用于消耗丰富的RNA种类。在一实施例中,在消耗衍生自丰富的多个RNA分子的多个核酸(例如转变的DNA)之后,在步骤180中产生多个序列读数。

[0100] 在一实施例中,可以使用本领域已知的多种方法将所述多个序列读数与一参考基因组比对以确定比对位置信息。所述比对位置信息可以指示所述参考基因组中对应于一给测序列读数的一开始核苷酸碱基和结束核苷酸碱基的一区域的一开始位置和一结束位置。比对位置信息还可以包括序列读数长度,其可以从所述开始位置和结束位置确定。所述参考基因组中的一区域可以与一基因或一基因的一区段相关联。所述参考基因组可以包括所述整个转录组,或其任何部分(例如多个标靶转录物)。在另一实施例中,所述参考基因组可以是来自被测试的一生物体的所述整个基因组,并且衍生自(或反转录自)多个提取的RNA分子的多个序列读数与所述参考基因组比对以确定位置、片段长度及/或开始和结束位置。例如,在一个实施例中,多个序列读数与人类参考基因组hg19比对。所述人类参考基因组hg19的所述序列可从参考基因组体协会(Genome Reference Consortium)获得,一参考号为GRCh37/hg19,并且也可从Santa Cruz Genomics Institute提供的基因组浏览器(Genome Browser)获得。所述比对位置信息可以指示所述参考基因组中对应于一给测序列读数的一开始核苷酸碱基和结束核苷酸碱基的一区域的一开始位置和一结束位置。比对位置信息还可以包括序列读数长度,其可以从所述开始位置和结束位置确定。所述参考基因组中的一区域可以与一基因或一基因的一区段相关联。

[0101] 多个暗通道RNA分子的识别

[0102] 本公开的多个方面包括多种用于识别指示一受试者中的一疾病状态的一个或多个RNA序列(或“多个暗通道RNA分子”)的电脑执行方法。在一些实施例中,所述多种方法涉及通过一电脑系统获得来自已知患有所述疾病的一受试者的一第一测试样本的多个RNA分子的一第一组的多个序列读数,其中所述第一测试样本包括多个血浆游离RNA(cfRNA)分子,以及获得来自一对照样本的多个RNA分子的一第二组的多个序列读数,检测存在于所述第一组的多个序列读数中且不存在于所述第二组的多个序列读数中的一个或多个RNA序列,以识别指示所述疾病状态的一个或多个RNA序列。在一些实施例中,从所述患者中获得的所述第一测试样本包括一体液(例如血液、血浆、血清、尿液、唾液、肋膜液、心包液、脑脊液(CSF)、腹膜液或其任意组合)。在一个优选实施例中,从所述患者中获得的一测试样本是一血浆样本。在一些实施例中,所述对照样本包括从所述受试者的多个健康细胞(例如多个白血球细胞)中获得的多个RNA分子。

[0103] 图2是说明根据本公开的一个实施例的一种用于识别指示一疾病状态的一个或多个RNA序列的方法的一流程图。如图2所示,在步骤210中,从包括多个血浆游离RNA(cfRNA)分子的一生物测试样本中获得一第一组的多个序列读数。所述含有血浆游离的生物测试样本可以是任何一体液,例如血液、血浆、血清、尿液、肋膜液、脑脊液、泪液、唾液或腹水。根据此实施例,所述cfRNA生物测试样本获得自己知患有或怀疑患有一疾病的一测试受试者,从所述样本中提取的所述多个cfRNA分子和确定的多个序列读数(如本文别处所述)。例如,在

一个实施例中,使用一反转录步骤合成一互补DNA股,生成一cDNA/RNA杂合分子,降解所述RNA分子,使用一聚合酶从所述cDNA股合成一双股DNA分子,制备一测序文库,以及使用一测序平台确定的多个序列读数。所述测序步骤可以使用本领域中任何已知的测序平台进行,例如任何大规模并行测序平台,包括一合成式测序平台(例如Illumina的HiSeq X)或一连接测序平台(例如所述Life Technologies SOLiD平台)、所述Ion Torrent/Ion Proton、半导体测序、Roche 454、多个单分子测序平台(例如Helicos、Pacific Biosciences和nanopore),如前所述。替代地,可以使用用于检测和定量所述多个序列读数的其他方法,例如,基于阵列的杂交、基于探针的溶液内杂交、多个基于连接的测定、多个引子延伸反应测定,可以用于确定来自多个DNA分子(例如从多个RNA分子转变而来)的多个序列读数,如本领域技术人员所容易理解的。

[0104] 在步骤220,一第二组的多个序列读数是从一健康对照样本中获得的。在一个实施例中,所述健康对照样本来自相同受试者并且包括多个细胞RNA分子。例如,所述对照样本可以是多个血球细胞,例如多个白血球细胞,并且所述多个序列读数衍生自从所述多个血球细胞中提取的多个RNA分子。根据此实施例,所述多个RNA分子从所述健康对照样本(例如多个血球细胞)中提取,转变为DNA,制备一测序文库,并确定所述第二组的多个序列读数(如本文别处所述)。在其他多个实施例中,所述健康对照样本可以从一健康受试者或从多个健康细胞中获得的多个RNA序列确定的序列数据的一数据库。

[0105] 在步骤230,比较来自所述第一组的多个序列读数和所述第二组的多个序列读数的多个序列读数,以识别指示一疾病状态的一个或多个RNA分子。此外,存在于所述第一组的多个序列读数中并且不存在于所述第二组的多个序列读数中的一个或多个序列读数(衍生自多个RNA分子)被识别为衍生自指示一疾病状态的多个RNA分子。例如,所述第一组的多个序列读数可以包括衍生自从已知患有或怀疑患有一疾病(例如癌症)的一受试者中获得的一血浆样本的多个cfRNA分子的多个序列读数。并且所述第二组的多个序列读数可以包括衍生自从多个健康细胞(例如多个白血球细胞)的多个RNA分子的多个序列读数。通过将衍生自多个健康细胞的所述第二组的多个序列读数与衍生自一血浆游离RNA样本的所述第一组的多个序列读数进行比较和去除,可以识别衍生自一疾病状态(例如癌症)的所述多个序列读数。

[0106] 在一些实施例中,多个RNA序列的一对照数据集包括从一位或多位健康受试者中获得的多个序列读数。在各种实施例中,所述第二组的多个序列读数包括从一公共数据库中获得的RNA序列信息。可以根据本发明的多个实施例使用的多个公共数据库包括所述组织RNA-seq数据库GTEx(可在gtexportal.org/home获得)。在一些实施例中,多个RNA序列的一对照数据集包括从所述受试者的多个血球细胞中获得的多个序列读数。例如,在一些实施例中,多个序列读数是从一受试者的多个白血球细胞(WBC)中获得的。

[0107] 多个源自肿瘤的RNA分子的检测

[0108] 本公开的多个方面包括多种用于检测一受试者中的一个或多个源自肿瘤的RNA分子的电脑执行方法。在一些实施例中,所述多种方法包括:通过一电脑系统获得来自已知患有一肿瘤的一受试者的一第一测试样本的多个RNA分子的一第一组的多个序列读数,其中所述第一测试样本包括多个血浆游离RNA(cfRNA)分子;通过一电脑系统获得来自所述受试者的多个血球细胞的多个RNA分子的一第二组的多个序列读数;及/或通过一电脑系统检测

存在于所述第一组的多个序列读数中且不存在于所述第二组的多个序列读数中的一个或多个RNA序列,以检测所述受试者中的所述一个或多个源自肿瘤的RNA分子。

[0109] 在一些实施例中,从所述患者中获得的所述第一测试样本包括血液、血浆、血清、尿液、唾液、肋膜液、心包液、脑脊液(CSF)、腹膜液或其任意组合。在一个优选实施例中,从所述患者中获得的一测试样本是一血浆样本。在一些实施例中,从所述受试者中获得的所述多个血球细胞是多个白血球细胞(WBC)。

[0110] 图3是说明根据本发明的一个实施例的一种用于识别一个或多个源自肿瘤的RNA序列的方法的一流程图。在步骤310,从包括多个血浆游离RNA(cfRNA)分子的一生物测试样本中获得一第一组的多个序列读数。根据此实施例,所述cfRNA生物测试样本获自己知患有或怀疑患有一疾病的一测试受试者,从所述样本中提取的所述多个cfRNA分子和确定的多个序列读数(如本文别处所述)。例如,在一个实施例中,使用一反转录步骤合成一互补DNA股,生成一cDNA/RNA杂合分子,降解所述RNA分子,使用一聚合酶从所述cDNA股合成一双股DNA分子,制备一测序文库,以及使用一测序平台确定的多个序列读数。如前所述,所述测序步骤可以使用本领域任何已知的测序平台进行。替代地,可以使用用于确定所述多个序列读数的其他方法,例如,基于阵列的杂交、基于探针的溶液内杂交、多个基于连接的测定、多个引子延伸反应测定,可以用于检测及/或定量从多个DNA分子(例如从多个RNA分子转变而来)获得的多个序列读数,如本领域技术人员所容易理解的。

[0111] 在步骤315,一第二组的多个序列读数是来自多个血球细胞(例如多个白血球细胞或肤色血球层)中获得的。在一个实施例中,所述多个血球细胞是从相同受试者中获得的,而多个RNA分子是从中提取的。根据此实施例,从所述多个血球细胞中提取所述多个RNA分子,将其转变为DNA,制备一测序文库,并确定所述第二组的多个序列读数(如本文别处所述)。一般而言,本领域任何已知的方法都可用于从所述测试样本中提取和纯化多个血浆游离核酸。例如,可以使用一种或多种已知的市售方案或试剂盒,例如所述QIAamp循环核酸试剂盒(Qiagen)提取和纯化多个血浆游离核酸。

[0112] 在步骤320,当一个或多个RNA序列存在于所述第一组的多个序列读数中且不存在于所述第二组的多个序列读数中时,检测到一个或多个源自肿瘤的RNA分子。此外,存在于所述第一组的多个序列读数中且不存在于所述第二组的多个序列读数中的一个或多个序列读数(衍生自多个RNA分子)被识别为衍生自指示一疾病状态的多个RNA分子。例如,所述第一组的多个序列读数可以包括衍生自从已知患有或怀疑患有一疾病(例如癌症)的一受试者中获得的一血浆样本的多个cfRNA分子的多个序列读数。并且所述第二组的多个序列读数可以包括衍生自从多个血球细胞(例如多个白血球细胞)的多个RNA分子的多个序列读数。通过将衍生自多个血球细胞的所述第二组的多个序列读数与衍生自一血浆游离RNA样本的所述第一组的多个序列读数进行比较和去除,可以识别衍生自一癌症的所述多个序列读数。

[0113] 使用一暗通道的多个RNA分子检测一疾病状态

[0114] 图4是说明根据本发明的一个实施例的一种在一受试者中检测癌症的所述存在、确定癌症的一状态、监测癌症进展及/或确定癌症类型的方法的一流程图。在步骤410,从一受试者中提取一生物测试样本。如前所述,在一个实施例中,所述测试样本可以是一体液(例如血液、血浆、血清、尿液、唾液、肋膜液、心包液、脑脊液(CSF)、腹膜液或其任意组合),

所述体液包括多个血浆游离RNA分子。

[0115] 在步骤415,从所述测试样本中提取多个血浆游离RNA分子,别且制备一测序文库。一般而言,本领域任何已知的方法都可用于从所述测试样本中提取和纯化多个血浆游离核酸。例如,可以使用一种或多种已知的市售方案或试剂盒,例如所述QIAamp循环核酸试剂盒(Qiagen)提取和纯化多个血浆游离核酸(多个cfRNA分子)。提取后,所述多个cfRNA分子用于制备一测序文库。在一个实施例中,使用一反转录步骤以产生多个cDNA/RNA杂合分子、所述RNA降解以产生一单股cDNA分子、从所述多个单股cDNA分子模板合成一第二股以产生多个双股DNA分子、以及与所述多个双股DNA分子连接的多个DNA转接子以生成一测序文库。如前所述,所述多个DNA转接子可以包括一个或多个测序寡核苷酸,用于随后的从集生成及/或测序(例如用于合成式测序(SBS)(Illumina, 圣地亚哥, 加利福尼亚州)的已知P5和P7序列)。在另一实施例中,所述转接子包括一样本特异性指标序列,使得在文库制备之后,所述文库可以与由多个单独样本制备的一个或多个其他文库组合,从而允许多重测序。在另一实施例中,通过转接子连接添加多个唯一分子标识符(UMI)。

[0116] 在步骤420,进行一测序反应以产生多个序列读数。一般而言,可以使用本领域已知的任何方法从所述测序文库中获得序列数据或多个序列读数。例如,在一个实施例中,可以使用次世代测序(NGS)获得来自所述测序文库的测序数据或多个序列读数。多种次世代测序方法包括,例如合成技术(Illumina)、焦磷酸测序(454)、离子半导体技术(Ion Torrent测序)、单分子实时测序(Pacific Biosciences)、连接测序(SOLiD测序)、及纳米孔测序(Oxford Nanopore Technologies)。在一些实施例中,测序是使用具有多个可逆染料终止子的合成式测序进行大规模平行测序。在其他多个实施例中,测序是连接测序。在还有其他多个实施例中,测序是单分子测序。在又另一实施例中,测序是双端测序。可选地,可以在测序之前进行一扩增步骤。

[0117] 在步骤425,过滤从所述cfRNA样本获得的多个序列读数以生成多个非排除序列读数的一列表,并且在步骤430定量所述多个非排除序列读数。例如,如本文别处所述,可以过滤从所述cfRNA样本获得的所述多个序列读数,以排除已知存在于多个健康细胞中的序列。在一个实施例中,对从多个健康细胞(例如多个白血球细胞)中提取的多个RNA分子测序,导出从所述多个cfRNA衍生的序列读数中排除的多个序列读数,以获得多个非排除的序列读数。在另一实施例中,来自一数据库(例如一公共数据库)的RNA测序数据可用于过滤或排除已知存在于多个健康细胞的多个读数中的多个序列,以获得多个非排除的序列读数。

[0118] 在步骤435,当所述多个量化的非排除序列读数超过一阈值时,检测一疾病状态。在各种实施例中,所述阈值是一整数,范围从大约或正好1至大约或正好10,例如大约或正好2、3、4、5、6、7、8或大约或正好9。在一些实施例中,所述阈值是一非整数值,范围从大约或正好0.1至大约或正好0.9,例如大约或正好0.2、0.3、0.4、0.5、0.6、0.7或大约或正好0.8。

[0119] 本公开的多个方面是有关多种用于在已知患有或怀疑患有一癌症的一受试者中检测癌症的一存在、确定一癌症阶段、监测一癌症进展及/或确定一癌症类型的方法。在一些实施例中,所述多种方法涉及:(a)从所述受试者的一生物测试样本中的多个cfRNA分子获得多个序列读数;(b)定量地检测所述生物测试样本中衍生自一个或多个RNA标记物的一个或多个序列的所述存在,以确定一肿瘤RNA评分,其中所述一个或多个RNA标记物选自由一个或多个靶RNA分子组成的所述群组;及(c)当所述肿瘤RNA评分超过一阈值时,在所述

受试者中检测所述癌症的所述存在、确定所述癌症阶段、监测所述癌症进展、及/或确定所述癌症类型。在各种实施例中,所述阈值是一整数,范围从大约或正好1至大约或正好10,例如大约或正好2、3、4、5、6、7、8或大约或正好9。在一些实施例中,所述阈值是一非整数值,范围从大约或正好0.1至大约或正好0.9,例如大约或正好0.2、0.3、0.4、0.5、0.6、0.7或大约或正好0.8。

[0120] 根据本公开的多个实施例的多种定量检测方法可以包括多个核酸测序程序,例如次世代测序。在各种实施例中,测序可涉及全转录组测序。在各种实施例中,测序可涉及在进行所述测序程序之前针对一个或多个感兴趣的标靶RNA序列富集一样本。替代地,可以使用其他检测和定量多个序列读数的方法,例如,基于阵列的杂交、基于探针的溶液内杂交、多个基于连接的测定、多个引子延伸反应测定,可以用于确定来自多个DNA分子(例如从多个RNA分子转变而来)的多个序列读数,如本领域技术人员所容易理解的。

[0121] 图5是说明根据本发明的一个实施例的一种从源自一个或多个标靶RNA分子的一个或多个序列读数检测一疾病状态的方法的一流程图。在步骤510,获得包括多个血浆游离RNA分子的一生物测试样本。在一个实施例中,所述生物测试样本是一体液(例如血液、血浆、血清、尿液、唾液、肋膜液、心包液、脑脊液(CSF)、腹膜液样本或其任意组合)。

[0122] 在步骤515,检测并定量所述生物测试样本中衍生自一个或多个标靶RNA分子的一个或多个核酸序列的所述存在,以确定一肿瘤RNA评分。如本文别处所述,可以使用本领域任何已知的方法检测和定量衍生自多个RNA分子的多个核酸。例如,根据一个实施例,使用一测序程序,例如次世代测序平台(例如HiSeq或NovaSeq,Illumina,圣地亚哥,加利福尼亚州)检测和定量衍生自多个RNA分子的多个核酸。在其他多个实施例中,使用一微阵列(microarray)、反转录PCR(reverse transcription PCR)、实时PCR(real-time PCR)、定量实时PCR(quantitative real-time PCR)、数位PCR(digital PCR)、数位微滴PCR(digital droplet PCR)、数位乳化PCR(digital emulsion PCR)、多重PCR(multiplex PCR)、杂交捕获(hybrid capture)、多种寡核苷酸连接测定或其任意组合。如别处所述,在一个实施例中,可以使用一种或多种已知的市售方案或试剂盒,例如所述QIAamp循环核酸试剂盒(Qiagen)提取和纯化多个血浆游离核酸(多个cfRNA分子)。提取后,使用所述多个cfRNA分子以制备一测序文库。在一个实施例中,使用一反转录步骤以产生多个cDNA/RNA杂合分子、所述RNA降解以产生一单股cDNA分子、从所述多个单股cDNA分子模板合成一第二股以产生多个双股DNA分子。可选地,在一个实施例中,一个或多个标靶RNA分子(或从其衍生的多个DNA分子)在检测和定量之前被富集,如本文别处所述。

[0123] 在一个实施例中,所述肿瘤RNA评分是检测到的多个标靶RNA分子(或从衍生自所述多个标靶RNA分子的多个DNA分子中获得的多个序列读数)的所述数量或计数。在另一实施例中,所述肿瘤RNA评分包括检测到的多个标靶RNA分子(或从衍生自所述多个标靶RNA分子的多个DNA分子中获得的多个序列读数)的所述总数除以多个RNA分子标靶的多个基因的所述总数的一均数、一众数或一平均值。在还有其他多个实施例中,通过将所述多个序列读数输入到一预测模型来确定所述肿瘤RNA评分,并且所述肿瘤RNA评分输出为一可能性或概率,如本文别处所述。

[0124] 在步骤520,当所述肿瘤RNA评分超过一阈值时,检测癌症的所述存在、确定癌症的一状态、监测癌症进展、及/或确定一受试者中的一癌症类型。所述阈值是一整数,范围从大

约或正好1至大约或正好10,例如大约或正好2、3、4、5、6、7、8或大约或正好9。在一些实施例中,所述阈值是一非整数值,范围从大约或正好0.1至大约或正好0.9,例如大约或正好0.2、0.3、0.4、0.5、0.6、0.7或大约或正好0.8。替代地,当从一预测模型输出所述肿瘤RNA评分时,所述输出可以只是一可能性或概率,指示所述受试者患有癌症或一癌症类型的所述可能性或概率。

[0125] 癌症指标评分

[0126] 本公开的多个方面是有关多种用于检测一患者中一癌症的所述存在的电脑执行方法。在一些实施例中,所述多种方法涉及:在包括一处理器和一电脑可读介质的一电脑中接收一数据集,其中所述数据集包括通过对衍生自从所述患者的一生物测试样本中的多个靶核糖核酸(RNA)分子的多个核酸分子(例如DNA分子)测序而获得的多个序列读数,并且其中所述电脑可读介质包括多个指令,当所述处理器执行这些指令时,使所述电脑:确定所述生物测试样本中的所述多个靶RNA分子的一表达水平;将所述多个靶RNA分子中的每一个的所述表达水平与一RNA组织评分矩阵进行比较,以确定每一个靶RNA分子的一癌症指标评分;汇总每一个靶RNA分子的所述癌症指标评分,以产生所述生物测试样本的一癌症指标评分;以及当所述生物测试样本的所述癌症指标评分超过一阈值时,检测所述患者中所述癌症的所述存在。

[0127] 在一些实施例中,所述多个靶RNA分子在患有一已知癌症状态的多位患者中的一表达水平超过其在多位健康患者中的表达水平。在各种实施例中,患有一已知癌症状态的一患者中一靶RNA分子的一表达水平比一健康患者中的所述靶RNA分子的所述表达水平高大约或正好2倍至大约或正好10倍,例如比一健康患者中的所述靶RNA分子的所述表达水平高大约或正好3、4、5、6、7、8或大约或正好9倍。在各种实施例中,在来自一健康患者的一生物测试样本中检测不到一靶RNA分子,即,所述靶RNA分子具有一检测不到的表达水平。

[0128] 在一些实施例中,所述生物测试样本中的多个靶RNA分子的所述数量范围从大约或正好1至大约或正好2000、大约或正好10至大约或正好1000、大约或正好10至大约或正好500、或大约或正好10至大约或正好500。在其他多个实施例中,多个靶RNA分子的所述数量范围从大约或正好1至大约或正好50、大约或正好1至大约或正好40、大约或正好1至大约或正好30、或大约或正好1至大约或正好20,例如大约或正好2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、或大约或正好20。

[0129] 在一些实施例中,所述癌症指标评分包括从所述生物测试样本中检测到的多个靶RNA分子(或从衍生自所述多个靶RNA分子的多个DNA分子中获得的多个序列读数)的所述总数的一汇总。在另一实施例中,所述癌症指标评分包括检测到的多个靶RNA分子(或从衍生自所述多个靶RNA分子的多个DNA分子中获得的多个序列读数)的所述总数除以多个RNA分子靶的多个基因的所述总数的一均数、一众数或一平均值。在还有其他多个实施例中,通过将所述多个序列读数输入到一预测模型来确定所述肿瘤RNA评分,并且所述肿瘤RNA评分输出为一可能性或概率,如本文别处所述。

[0130] 在一些实施例中,所述阈值是一整数,范围从大约或正好1至大约或正好10,例如大约或正好2、3、4、5、6、7、8或大约或正好9。在一些实施例中,所述阈值是一非整数值,范围从大约或正好0.1至大约或正好0.9,例如大约或正好0.2、0.3、0.4、0.5、0.6、0.7或大约或

正好0.8。在其他多个实施例中,所述阈值的范围从每百万个中大约或正好0.5个读数至大约或正好5个读数(RPM),例如大约或正好1、1.5、2、2.5、3、3.5、4或大约或正好4.5RPM。所述癌症定位器评分阈值可以基于在一对照样本中检测到的多个标靶RNA分子(或从其衍生的多个序列读数)的所述数量来确定,例如一健康受试者或患有一已知疾病状态的一受试者。替代地,当所述癌症定位器评分从一预测模型输出时,所述输出可以只是一可能性或概率,指示所述受试者患有癌症或一癌症类型的所述可能性或概率。

[0131] 图6是说明根据本发明的一个实施例的一种用于基于一癌症指标评分来检测一受试者中癌症的所述存在的方法的一流程图。在步骤610,接收一数据集,其包括衍生自一生物测试样本中的多个cfRNA分子的多个序列读数。例如,如本文所述,可以确定从一生物测试样本中提取的多个cfRNA分子的多个序列读数。此外,反转录多个cfRNA分子以产生多个DNA分子,并对所述多个DNA分子测序以产生多个序列读数。

[0132] 在步骤615,确定所述生物测试样本中多个标靶RNA分子的一表达水平。例如,在一个实施例中,多个标靶RNA分子的所述表达水平可以基于检测到的衍生自一个或多个感兴趣的标靶RNA分子的多个序列读数的量化来确定。

[0133] 在步骤620,将所述多个标靶RNA分子的每一个的所述表达水平与一RNA组织评分矩阵进行比较以确定每一个标靶RNA分子的一癌症指标评分。所述RNA组织评分矩阵可以从包括多个序列读数的一训练集中确定,所述多个序列读数衍生自多个具有已知癌症状态的癌症训练样本。

[0134] 在步骤625,汇总每一个标靶RNA分子的所述多个癌症指标评分以产生一癌症指标评分。在一些实施例中,所述癌症指标评分包括从所述生物测试样本中检测到的多个标靶RNA分子(或从衍生自所述多个标靶RNA分子的多个DNA分子中获得的多个序列读数)的所述总数的一汇总。在另一实施例中,所述癌症指标评分包括检测到的多个标靶RNA分子(或从衍生自所述多个标靶RNA分子的多个DNA分子中获得的多个序列读数)的所述总数除以多个RNA分子标靶的多个基因的所述总数的一均数、一众数或一平均值。

[0135] 在步骤630,当所述测试样本的所述癌症指标评分超过一阈值时,检测一受试者中癌症的所述存在。如前所述,在一个实施例中,所述阈值是一整数,范围从大约或正好1至大约或正好10,例如大约或正好2、3、4、5、6、7、8或大约或正好9。在一些实施例中,所述阈值是一非整数值,范围从大约或正好0.1至大约或正好0.9,例如大约或正好0.2、0.3、0.4、0.5、0.6、0.7或大约或正好0.8。在其他多个实施例中,所述阈值的范围从每百万个中大约或正好0.5个读数至大约或正好5个读数(RPM),例如大约或正好1、1.5、2、2.5、3、3.5、4或大约或正好4.5RPM。

[0136] 本公开的多个方面包括多种用于基于所述多个标靶RNA分子的一个或多个的所述表达水平、所述多个标靶RNA分子的一个或多个的所述癌症指标评分、所述生物测试样本的所述癌症指标评分或其任意组合,来确定所述患者中所述癌症的一癌细胞类型或起源组织的方法。在各种实施例中,所述多种方法进一步涉及基于所述多个标靶RNA分子的一个或多个的所述表达水平、所述多个标靶RNA分子的一个或多个的所述癌症指标评分、所述生物测试样本的所述癌症指标评分或其任意组合,将一患者治疗性地分类为多个治疗类别中的一者或多者。

[0137] 在各种实施例中,所述电脑被配置为生成一报告,所述报告包括所述多个标靶RNA

分子的一个或多个的一表达水平、所述多个靶RNA分子的一个或多个的一癌症指标评分、所述生物测试样本的一癌症指标评分、所述患者中所述癌症的所述存在或不存在的一指示、所述患者中所述癌症的起源组织的所述癌细胞类型的一指示、所述患者的一治疗分类或其任意组合。

[0138] RNA组织矩阵评分

[0139] 本公开的多个方面包括多种用于构建一RNA组织评分矩阵的方法。在一些实施例中,所述多种方法涉及编译从多位患者中获得的多个RNA序列读数,以产生一RNA表达矩阵、及/或使用一组织特异性RNA表达矩阵对所述RNA表达矩阵进行归一化,以构建所述RNA组织评分矩阵。在一些实施例中,所述组织特异性RNA表达矩阵包括多个参考人类组织。在各种实施例中,从多位健康患者中获得所述多个RNA序列读数,以构建一健康RNA组织评分矩阵。在各种实施例中,从多位患有一已知癌症类型的患者中获得所述多个RNA序列读数,以构建一癌症RNA组织评分矩阵。

[0140] 多个RNA标记物和分析技术

[0141] 可以对多个cfRNA分子及/或多个ctRNA分子执行根据本公开的一些实施例的多种方法。在一些实施例中,用于所述多种主题方法的多个RNA分子包括来自多个癌细胞和多个非癌细胞的多个RNA分子。

[0142] 在多个实施例中,多种方法包括(a)富集所述多个靶cfRNA分子或其多个cDNA分子,以产生多个多核苷酸的一富集样本;及/或(b)对所述富集样本的所述多个多核苷酸或其多个扩增产物测序;其中所述多个靶cfRNA分子选自表11中的一个或多个转录物。在多个实施例中,所述多个靶cfRNA分子选自表8或12-15中的一者或多者(例如来自表8或11-14中的一者或多者的5、10、15或20个基因的多个转录物)。在多个实施例中,测量所述多个cfRNA分子包括在检测或测量之前富集所述多个cfRNA分子(或其多个cDNA分子),例如通过测序。在多个实施例中,测量的所述多个靶cfRNA分子来自500个或更少个基因(例如400个或更少个、300个或更少个、200个或更少个、100个或更少个、或50个或更少个基因)。

[0143] 在多个实施例中,多种方法包括:(a)测量所述受试者的一样本中的多个靶血浆游离RNA(cfRNA)分子,其中所述多个靶cfRNA分子选自表8或11-14中的多个转录物;以及(b)检测所述癌症,其中检测所述癌症包括检测高于一阈值水平的所述多个靶cfRNA分子中的一个或多个。在多个实施例中,所述多个靶cfRNA分子是选自表8或11-14的一者或多者中列出的至少2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、50或多个基因的多个转录物。多个靶cfRNA分子可以来自选自这些表中的任何一个或其任意组合的多个基因。在多个实施例中,选自表8或11-14中多个表的所述数量为2、3、4或所有表,或所述多个表的任意组合。在多个实施例中,测量所述多个cfRNA分子不包括全转录组测序。在多个实施例中,测量所述多个cfRNA分子包括在检测或测量之前富集所述多个cfRNA分子(或其多个cDNA分子),例如通过测序。在多个实施例中,测量的所述多个靶cfRNA分子来自少于500个基因(例如少于400、300、200、100或50个基因)。

[0144] 在一些实施例中,一个或多个靶cfRNA分子衍生自选自表1中所列的所述多个基因的一个或多个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表1中的至少2、3、4、5、6、7、8、9、10、15、20、25或30个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表1中的至少5个基因。在多个实施例中,所述一个或多个靶cfRNA分

子包括来自表1中的至少10个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表1中的所述多个基因的所有。在多个实施例中,所述一个或多个靶cfRNA分子包括表1的所述前5个基因中的至少一者 (AGR2、HOXC10、S100A7、BPIFA1及/或IDI2-AS1),以及可选地来自表1中的一个或多个另外的基因。在多个实施例中,所述一个或多个靶cfRNA分子包括所述AGR2基因的多个转录物。在多个实施例中,所述一个或多个靶cfRNA分子包括AGR2、HOXC10、S100A7、BPIFA1和IDI2-AS1的多个转录物。在多个实施例中,测量的所述多个靶cfRNA分子来自少于500个基因(例如少于400、300、200、100或50个基因)。下表1提供了多个癌症暗通道生物标记物的多个示例。

表 1		
AGR2	HOXC10	S100A7
BPIFA1	IDI2-AS1	SCGB2A2
CASP14	KRT16P2	SERPINB5
CSN1S1	LALBA	SFTA3
DISP2	LINC00163	SFTPA2
EIF2D	NKX2-1	SLC34A2
FABP7	OPN1SW	TFF1
GABRG1	PADI3	VTCN1
GNAT3	PTPRZ1	WFDC2
GRHL2	ROS1	MUC5B
SMIM22	CXCL17	RNU1-1
KLK5		

[0146] 在一些实施例中,一个或多个靶cfRNA分子衍生自选自表2中所列的所述多个基因的一个或多个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表2中的至少2、3、4、5、6、7、8、9或10个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表2中的至少5个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表2中的至少10个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表2中的所述多个基因的所有。在多个实施例中,所述一个或多个靶cfRNA分子包括表2的所述前5个基因中的至少一者 (ROS1、NKX2-1、GGTLC1、SLC34A2和SFTPA2),以及可选地来自表2中的一个或多个另外的基因。在多个实施例中,所述一个或多个靶cfRNA分子包括所述ROS1基因的多个转录物。在多个实施例中,所述一个或多个靶cfRNA分子包括ROS1、NKX2-1、GGTLC1、SLC34A2和SFTPA2的多个转录物。在多个实施例中,测量的所述多个靶cfRNA分子来自少于500个基因(例如少于400、300、200、100或50个基因)。下表2提供了多个暗通道肺癌生物标记物的多个示例。

[0147]	表 2
	ROS1
	NKX2-1
	GGTLC1
	SLC34A2
	SFTPA2
	BPIFA1
	SFTA3
	GABRG1
	AGR2
	GNAT3
	MUC5B
	SMIM22
	CXCL17
[0148]	WFDC2

[0149] 在一些实施例中,一个或多个靶cfRNA分子衍生自选自表3中所列的所述多个基因的一个或多个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表3中的至少2、3、4、5、6、7、8或9个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表3中的至少5个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表3中的所述多个基因的所有。在多个实施例中,所述一个或多个靶cfRNA分子包括表3的所述前5个基因中的至少一者(SCGB2A2、CSN1S1、VTCN1、FABP7和LALBA),以及可选地来自表3中的一个或多个另外的基因。在多个实施例中,所述一个或多个靶cfRNA分子包括所述SCGB2A2基因的多个转录物。在多个实施例中,所述一个或多个靶cfRNA分子包括SCGB2A2、CSN1S1、VTCN1、FABP7和LALBA的多个转录物。在多个实施例中,测量的所述多个靶cfRNA分子来自少于500个基因(例如少于400、300、200、100或50个基因)。下表3提供了多个乳腺癌通道生物标记物的多个示例。

[0150]	表 3
	SCGB2A2
	CSN1S1
	VTCN1
	FABP7
	LALBA
	CASP14
	KLK5
	WFDC2
[0151]	OPN1SW

[0152] 在一些实施例中,一个或多个靶cfRNA分子衍生自选自表4中所列的所述多个基因的一个或多个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表4中的至少2、3、4或5个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表4中的至少5个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表4中的所述多个基因的所有。在多个实施例中,所述一个或多个靶cfRNA分子包括表4的所述前5个基因中的至少一者(CASP14、CRABP2、FABP7、SCGB2A2和SERPINB5),以及可选地来自表4中的一个或多个另外的基因。在多个实施例中,所述一个或多个靶cfRNA分子包括所述CASP14基因的多个转录物。在多个实施例中,所述一个或多个靶cfRNA分子包括CASP14、CRABP2、FABP7、SCGB2A2和SERPINB5的多个转录物。在多个实施例中,测量的所述多个靶cfRNA分子来自少于500个基因(例如少于400、300、200、100或50个基因)。下表4提供了使用一heteroDE方法识别的多个乳癌生物标记物的多个示例,如本文所述。

[0153]	表 4
	CASP14
	CRABP2
	FABP7
	SCGB2A2
	SERPINB5
	TRGV10
	VGLL1
	TFF1

[0154]

AC007563.5

[0155] 在一些实施例中,一个或多个靶cfRNA分子衍生自选自表5中所列的所述多个基因的一个或多个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表5中的至少2、3、4、5、6、7、8、9、10、15、20或25个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表5中的至少5个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表5中的至少10个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括表5中的所述多个基因的所有。在多个实施例中,所述一个或多个靶cfRNA分子包括表5的所述前5个基因中的至少一者 (PTPRZ1、AGR2、SHANK1、PON1和MYO16_AS1),以及可选地来自表5中的一个或多个另外的基因。在多个实施例中,所述一个或多个靶cfRNA分子包括所述PTPRZ1基因的多个转录物。在多个实施例中,所述一个或多个靶cfRNA分子包括PTPRZ1、AGR2、SHANK1、PON1和MYO16_AS1的多个转录物。在多个实施例中,测量的所述多个靶cfRNA分子来自少于500个基因(例如少于400、300、200、100或50个基因)。下表5提供了使用一信息增益方法识别的多个肺癌生物标记物的多个示例,如本文所述。

表 5

[0156]

PTPRZ1	AGR2	SHANK1
PON1	MYO16_AS1	NPAS3
LINC00407	LMO3	KRT15
ELFN2	MUC5B	SAA2
SLIT3	NALCN	LUM
GDA	LINC01498	TMEM178A
RCVRN	XKRX	ROS1

[0157]

NBPF7	ACSM5	SLC10A3
SAA1	CYP3A4	LINC00643
GLP1R	TRAV8_5	GNAT3

[0158] 在一些实施例中,一个或多个靶cfRNA分子衍生自选自表6中所列的所述多个基因的一个或多个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表6中的至少2、3、4、5、6、7、8、9、10、15、20或25个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表6中的至少5个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表6中的至少10个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括表6中的所述多个基因的所有。在多个实施例中,所述一个或多个靶cfRNA分子包括表6的所述前5个基因中的至少一者 (ADARB2、HORMAD2、SPDYE18、RPS19和CYP4F35P),以及可选地来自表6中的一个或多个另外的基因。在多个实施例中,所述一个或多个靶cfRNA分子包括所述ADARB2基因的多个转录物。在多个实施例中,所述一个或多个靶cfRNA分子包括ADARB2、HORMAD2、SPDYE18、RPS19和CYP4F35P的多个转录物。在多个实施例中,测量的所述

多个靶cfRNA分子来自少于500个基因(例如少于400、300、200、100或50个基因)。下表6提供了使用一信息增益方法识别的多个乳癌生物标记物的多个示例,如本文所述。

表 6		
[0159]	ADARB2	HORMAD2
	RPS19	CYP4F35P
	SLC34A2	MUC5B
	TLX2	IDI2
[0160]	SPDYE18	MIR503HG
	IGKVID_16	PDPK2P
	ACTBP2	TTPA
	RIMKLA	WNT6
	RANBP6	FHOD3
	CTF1	GSTA9P
	FAM9C	SMIM2_AS1
	FAM171A2	GRIA2
	LINC01140	TRBV6_4
	LINC00856	FOXC1
	CCDC188	GABRR2

[0161] 在一些实施例中,一个或多个靶cfRNA分子衍生自选自表7中所列的所述多个基因的一个或多个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表7中的至少2、3、4、5、6、7、8、9或10个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表7中的至少5个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表7中的至少10个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表7中的所述多个基因的所有。在多个实施例中,所述一个或多个靶cfRNA分子包括表7的所述前5个基因中的至少一者(S100A7、FOXA1、BARX2、MMP7和PLEKHG4B),以及可选地来自表7中的一个或多个另外的基因。在多个实施例中,所述一个或多个靶cfRNA分子包括所述S100A7基因的多个转录物。在多个实施例中,所述一个或多个靶cfRNA分子包括S100A7、FOXA1、BARX2、MMP7和PLEKHG4B的多个转录物。在多个实施例中,测量的所述多个靶cfRNA分子来自少于500个基因(例如少于400、300、200、100或50个基因)。下表7提供了在癌症组织中以多个相对高水平表达的多个暗通道癌症生物标记物的多个示例。

表 7		
[0162]	S100A7	FOXA1
	MMP7	PLEKHG4B
[0163]	BARX2	TFAP2A
	TOX3	VTCN1
	COL22A1	FDCSP
	MATN3	TFF1
	ANKRD30A	LAMA1
	VGLL1	

[0164] 在一些实施例中,一个或多个靶cfRNA分子衍生自选自表11中所列的所述多个

基因的一个或多个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表11中的至少2、3、4、5、10、25、50、100、150、200、300或400个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表11中的至少5个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表11中的至少25个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表11中的至少100个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表11中的至少200个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表11中的至少300个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表11中的所述多个基因的所有。在多个实施例中,测量的所述多个靶cfRNA分子来自少于500个基因(例如少于400、300、200、100或50个基因)。下

[0165] 表11提供了多个癌症生物标记物的多个示例。

表 11					
AARD	CKMT1A	FGFR1	KRT14	PIP	SLC6A17
ABCA12	CKMT1B	FGFR2	KRT23	PITX2	SLC9A6
ABCC11	CLCA2	FGFR3	KRT6B	PLA2G12B	SLITRK4
ABCC8	CLDN10	FGFR4	KRT83	PLA2G1B	SLITRK6
ACTL8	CLDN18	FKBP10	LAMA4	PLA2G4E	SMIM17
ADAMTS15	CLDN6	FKBPL	LDLRAD1	POPDC3	SMR3A
ADAMTS8	CLEC3A	FLRT3	LEMD1	POTEKP	SNTN
ADGRF1	CLIC1	FOLR1	LENG1	POU5F1	SOWAHA
ADH7	CLPSL1	FOXA2	LEP	PPIP5K2	SOX2
ADIPOQ	CLPSL2	FOXI1	LILRB1	PPP1R11	SOX21
ADRB1	CLSTN2	FOXJ1	LINC00052	PPP1R14BP3	SOX9
AFP	CNGA3	FOXQ1	LINC00261	PPP1R14D	SP8

[0167]

AGER	CNOT3	FUT2	LINC00511	PPP1R18	SPDEF
AGR3	COL6A5	FUT3	LINC00641	PRIMA1	SPINK8
AGTR1	COL6A6	FZD7	LINC00707	PRND	SPON1
AGTR2	COPG2	FZD9	LINC00993	PRR15	SPRR2D
AIF1	CRYAB	GAL3ST1	LINC01016	PRR15L	SRSF12
AKR1B15	CSF2	GATA3-AS1	LINC01087	PRSS8	STAC2
ALG9	CSN3	GCNT3	LMX1B	PSMB8	STC2
ALK	CSNK2B	GCNT4	LRRC31	PTCHD1	STK32A
ALPG	CST1	GDF15	LRRN4	PYDC1	STMND1
ALPP	CST9	GFRA1	LY6D	RAB6C	STOML3
ANKRD30B	CT62	GGT6	MAGEA3	RABL2B	SUN3
ANKRD35	CTSE	GIN1	MAGEA6	RASD2	SURF6
ANXA8	CWC15	GJB5	MAPT	RBBP8NL	SUV39H1
AQP4	CYP21A2	GJC3	MB	RET	SYNM
ARHGEF38	CYP27C1	GKN2	MBOAT7	RGMA	SYP
ARL14	CYP4F23P	GNG4	MET	RHOV	SYT9
ART3	CYP4F8	GNL1	MEX3A	RHPN1-AS1	TAF15
ATF6B	CYP4Z2P	GP2	MIA	ROPN1	TAP2
ATP10B	DCX	GPR12	MICA	ROPN1B	TBX4
ATP6V0A4	DHX16	GPR143	MILR1	RRAD	TCF21
ATP7A	DIXDC1	GPR39	MIR205HG	RTL8B	TFAP2B
AZGP1P1	DLX1	GPR87	MKX	RTN4RL1	TMC3
B3GNT3	DLX3	GPSM3	MMP10	RXR8	TMEM125
B3GNT6	DMRTA2	GPX2	MMP12	S100A1	TMEM198
BMP5	DNAJB13	HAPLN1	MOCS3	S100A7A	TMEM59L
BPIFA2	DOC2A	HHIP	MSLN	SCARNA5	TMSB15A
BPIFB2	DSCAM-AS1	HOTAIR	MSX2	SCGB1A1	TNRC18P1
BRINP2	DSTNP2	HOXB13	MTM1	SCGB1D2	TRARG1
BRINP3	DUOXA1	HOXC11	MUC15	SCGB2A1	TRH
C2CD4A	DXO	HOXC13	MUC21	SCN7A	TRIM27
C4B	DYDC2	HOXC6	MUC6	SCNN1G	TRIM31
C5orf30	ECEL1	HOXC8	MUCL1	SCTR	TRIM39
C5orf46	EGFR	HOXC9	MUCL3	SEC14L6	TRIM48
C5orf49	EHMT2	HS6ST3	NCMAP	SEMA3B	TRPV6
C9orf116	ELF5	IGF2BP1	NDNF	SERPINA11	TSEN34
C9orf152	EMX1	IGSF1	NDUFA3	SERPINB4	TSPY26P
CA12	EN1	IL20	NELL1	SERPINB7	TTC30B
CACNG1	EPN3	INHA	NKAIN1	SEZ6	TTC36
CACNG4	EPOP	INSL4	NKAIN4	SEZ6L2	TTC6
CALCA	EPYC	IP6K3	NKX1-2	SFRP1	TTYH1
CALML5	ERBB2	IRX1	NKX2-1-AS1	SFRP5	UCN3
CCDC125	ERBB4	IRX2	NKX2-2	SFTA1P	UCP1
CCDC160	ERN2	IRX4	NKX6-1	SFTA2	UGT2B15

[0168]

CCKBR	ERVH48-1	IRX5	NNAT	SFTPA1	UPK1A
CCNO	ESR1	ITGA10	NTRK1	SFTPB	USP41
CCT8P1	ESRP1	ITGB6	NXNL2	SHH	VAR5
CD99	ESYT3	ITIH6	OBP2B	SHISA3	VN1R53P
CDH3	ETV3L	IVL	ODAM	SHISA9	VSTM2A
CDKL2	EXTL1	KCNC2	OSCAR	SIM2	WFDC10B
CEACAM5	F7	KCNJ11	OVOL2	SIX4	WNT3A
CEP41	FAM198A	KCNJ3	PAEP	SLC16A6	YBX1P10
CFTR	FAM19A3	KCNK15	PAX7	SLC22A31	ZBTB12
CGA	FAM216B	KCNK2	PCP4L1	SLC25A48	ZFP57
CHGA	FAXC	KCNK3	PDE4C	SLC26A3	ZNF737
CHIA	FBN3	KIFC1	PFDN6	SLC26A9	ZNRD1
CHRM1	FBXL19-AS1	KLHL38	PGC	SLC37A4	
CHST9	FEZF1-AS1	KLK8	PGR	SLC44A4	
CIDEA	FGFBP1	KRAS	PIK3CA	SLC6A14	

[0169] 在一些实施例中,一个或多个靶cfRNA分子衍生自选自表12中所列的所述多个基因的一个或多个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表12中的至少2、3、4、5、10、20、30、40、50或60个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表12中的至少5个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表12中的至少10个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表12中的至少25个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表12中的至少50个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表12中的所述多个基因的所有。在多个实施例中,测量的所述多个靶cfRNA分子来自少于500个基因(例如少于400、300、200、100或50个基因)。下表12提供了多个肺癌生物标记物的多个示例。

[0170]

表 12

AGTR2	CHIA	GCNT3	MET	SCN7A	SLC6A14
AQP4	COL6A5	GDF15	MMP12	SCTR	SLC9A6
ATP10B	CRYAB	GFRA1	MUC21	SEC14L6	SOX2

[0171]

B3GNT3	CTSE	GKN2	NDNF	SFRP1	SOX9
B3GNT6	DSTNP2	GPR39	NKX2-1-AS1	SFTA1P	STC2
BMP5	EPN3	HHIP	PCP4L1	SFTA2	STK32A
BPIFA1	ERN2	IRX5	PDE4C	SFTPA1	SYNM
CALCA	ESYT3	KRT14	PLA2G4E	SFTPB	TBX4
CCT8P1	FLRT3	KRT23	RASD2	SHH	TCF21
CDKL2	FOXA2	KRT6B	RRAD	SHISA3	UCN3
CEACAM5	FZD7	LINC00261	RTN4RL1	SLC22A31	
CFTR	GAL3ST1	LRRN4	SCGB1A1	SLC26A9	

[0172] 在各种实施例中,一个或多个靶cfRNA分子衍生自选自表13中所列的所述多个基因的一个或多个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表13中的至少2、3、4、5、10、20、30、40、50、60或70个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表13中的至少5个基因。在多个实施例中,所述一个或多个靶cfRNA分

子包括来自表13中的至少10个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表13中的至少25个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表13中的至少50个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表13中的所述多个基因的所有。在多个实施例中,测量的所述多个靶cfRNA分子来自少于500个基因(例如少于400、300、200、100或50个基因)。下表13提供了多个乳癌生物标记物的多个示例。

表 13					
[0173]	ABCC8	CRABP2	FUT3	KRT6B	PPP1R18
	ADAMTS15	CSN3	GIN1	LINC00511	PRR15
	AGR3	DSCAM-AS1	GP2	LINC01087	RET
	ART3	ELF5	HOXC13	LMX1B	RGMA
	AZGP1P1	ERBB2	HOXC6	LY6D	RHPN1-AS1
	B3GNT3	ERBB4	HOXC9	MAPT	ROPN1B
[0174]	BPIFB2	ESR1	IRX1	MB	RTN4RL1
	C2CD4A	EXTL1	IRX5	MET	S100A1
	CA12	F7	ITIH6	MEX3A	SEMA3B
	CCDC125	FBXL19-AS1	KCNJ11	MMP12	SFRP1
	CDH3	FOLR1	KCNK15	MSX2	SLC16A6
	CEACAM5	FOXA1	KIFC1	OBP2B	SLC44A4
[0175]	CLSTN2	FOXJ1	KRT23	PIK3CA	SOWAHA

[0175] 在一些实施例中,一个或多个靶cfRNA分子衍生自选自表14中所列的所述多个基因的一个或多个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表14中的至少2、3、4、5、10、15、20或30个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表14中的至少5个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表14中的至少10个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表14中的至少25个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表14中的所述多个基因的所有。在多个实施例中,测量的所述多个靶cfRNA分子来自少于500个基因(例如少于400、300、200、100或50个基因)。在多个实施例中,检测到高于一阈值的所述多个靶cfRNA分子是衍生自选自以下多个基因所组成的群组的多个cfRNA分子:ADIPOQ、AGR3、ANKRD30A、AQP4、BPIFA1、CA12、CEACAM5、CFTR、CXCL17、CYP4F8、FABP7、FOXI1、GGTLC1、GP2、IL20、ITIH6、LDLRAD1、LEMD1、LMX1B、MMP7、NKAIN1、NKX2-1、ROPN1、ROS1、SCGB1D2、SCGB2A2、SFTA2、SFTA3、SLC34A2、SOX9、STK32A、STMND1、TFAP2A、TFAP2B、TFF1、TRPV6、VGLL1及VTCN1。下表14提供了多个信息丰富的癌症生物标记物的多个示例。

表 14				
[0176]	ADIPOQ	CXCL17	LDLRAD1	SCGB1D2
	AGR3	CYP4F8	LEMD1	SCGB2A2
				TFAP2A
				TFAP2B

[0177]

ANKRD30A	FABP7	LMX1B	SFTA2	TFF1
AQP4	FOXI1	MMP7	SFTA3	TRPV6
BPIFA1	GGTLC1	NKAIN1	SLC34A2	VGLL1
CA12	GP2	NKX2-1	SOX9	VTCN1
CEACAM5	IL20	ROPN1	STK32A	
CFTR	ITIH6	ROS1	STMND1	

[0178] 在一些实施例中,一个或多个靶cfRNA分子衍生自选自表15中所列的所述多个基因的一个或多个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表15中的至少2、3、4、5、10、25、50、100、150、200、300或400个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表15中的至少5个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表15中的至少25个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表15中的至少100个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表15中的至少200个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表15中的至少300个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表15中的所述多个基因的所有。在多个实施例中,测量的所述多个靶cfRNA分子来自少于500个基因(例如少于400、300、200、100或50个基因)。

[0179]

表 15					
RNU1-1	ESYT3	DXO	SFTA1P	NKX2-1	FBN3
PADI3	CLSTN2	C4B	MKX	NKX2-1-AS1	CASP14
ACTL8	AGTR1	CYP21A2	ANKRD30A	FOXA1	CYP4F23P
PAX7	GPR87	ATF6B	LINC00993	TTC6	CYP4F8
NCMAP	ARL14	FKBPL	VN1R53P	SIX4	B3GNT3
EXTL1	LRRC31	AGER	RET	GPX2	PDE4C
NKAIN1	PIK3CA	GPSM3	ANXA8	CHGA	GDF15
GJB5	SOX2	TAP2	PLA2G12B	PRIMA1	TMEM59L
TMEM125	ADIPOQ	PSMB8	SFTPA2	SERPINA11	ZNF737
CYP4Z2P	FGFR3	RXRB	SFTPA1	DISP2	UPK1A
DMRTA2	FGFBP1	PFDN6	DYDC2	PPP1R14D	MIA

[0180]

LDLRAD1	SLC34A2	KIFC1	SFRP5	RHOV	CEACAM5
CLCA2	SHISA3	IP6K3	ADRB1	PLA2G4E	CXCL17
SLC6A17	GABRG1	LINC01016	GFRA1	CKMT1B	FUT2
CHIA	UGT2B15	SPDEF	FGFR2	CKMT1A	KLK5
FAM19A3	CSN1S1	CLPSL2	NKX1-2	DUOXA1	KLK8
VTCN1	ODAM	CLPSL1	MUC6	GCNT3	OSCAR
ITGA10	CSN3	PGC	MUC5B	C2CD4A	NDUFA3
ANKRD35	SMR3A	ADGRF1	CCKBR	CA12	CNOT3
CCT8P1	AFP	TFAP2B	SYT9	CT62	LENG1
IVL	CDKL2	BMP5	SPON1	TMC3	MBOAT7
SPRR2D	ART3	CGA	CALCA	LINC00052	TSEN34
S100A7A	NKX6-1	SRSF12	KCNJ11	RGMA	TTYH1
S100A7	ADH7	FAXC	ABCC8	SYNM	LILRB1
S100A1	ARHGEF38	POPDC3	SAA2	MSLNL	SMIM17
MEX3A	PITX2	LAMA4	SAA1	CLDN6	PRND
CRABP2	NDNF	ROS1	NELL1	SMIM22	LRRN4
NTRK1	PPP1R14BP3	FABP7	MUC15	SHISA9	FLRT3
ETV3L	UCP1	TCF21	ELF5	GP2	OVOL2
PCP4L1	TNRC18P1	ESR1	TRIM48	SCNN1G	NKX2-2
BRINP2	HHIP	AGR2	SCGB2A1	ERN2	LINC00261
BRINP3	GRIA2	AGR3	SCGB1D2	SEZ6L2	FOXA2
LEMD1	IRX4	SP8	SCGB2A2	DOC2A	CST9
SLC26A9	IRX2	PRR15	SCGB1A1	FBXL19-AS1	CST1
CTSE	IRX1	SUN3	CHRM1	PRSS8	GGTLC1
EIF2D	C5orf49	VSTM2A	FOLR1	PYDC1	TSPY26P
IL20	CCNO	EGFR	DNAJB13	ABCC11	BPIFB2
MIR205HG	CCDC125	FZD9	B3GNT6	TOX3	BPIFA2
KCNK2	GCNT4	GNAT3	CWC15	IRX5	BPIFA1
WNT3A	HAPLN1	GJC3	PGR	RRAD	NNAT
GNG4	GIN1	AZGP1P1	MMP7	CDH3	KCNK15
KCNK3	PIIP5K2	SLC26A3	MMP10	CLEC3A	WFDC2
ALK	C5orf30	MET	MMP12	SLC22A31	WFDC10B
GKN2	CSF2	CFTR	ALG9	TRARG1	MOCS3
EMX1	SOWAHA	PTPRZ1	CRYAB	RTN4RL1	RBBP8NL
SFTPFB	SLC25A48	FEZF1-AS1	DIXDC1	GGT6	NKAIN4
CNGA3	STK32A	LEP	TTC36	KRT16P2	SIM2
EN1	C5orf46	OPN1SW	SLC37A4	SEZ6	DSCAM-AS1
SCTR	ATP10B	CEP41	BARX2	TAF15	TFF1
CYP27C1	FOXI1	COPG2	ADAMTS8	EPOP	ERVH48-1
RAB6C	STC2	AKR1B10	ADAMTS15	STAC2	USP41
POTEKP	MSX2	AKR1B15	DSTNP2	ERBB2	SEC14L6
LINC01087	FGFR4	ATP6V0A4	LMO3	KRT23	GAL3ST1
GPR39	FOXQ1	TRPV6	KRAS	KRT15	RASD2

[0181]

KCNJ3	FOXC1	PIP	LALBA	KRT14	MB
ITGB6	TFAP2A	SHH	KRT83	FKBP10	ELFN2
SCN7A	STMND1	FGFR1	KRT6B	MAPT	RABL2B
DLX1	TRIM27	SFRP1	HOXC13	PRR15L	CD99
TTC30B	ZFP57	ESRP1	HOTAIR	HOXB13	GPR143
FZD7	ZNRD1	GRHL2	HOXC11	IGF2BP1	PTCHD1
ERBB4	PPP1R11	AARD	HOXC10	DLX3	SUV39H1
ABCA12	TRIM31	KLHL38	HOXC6	EPN3	SYP
TMEM198	TRIM39	COL22A1	HOXC9	TBX4	ITIH6
INHA	GNL1	LY6D	HOXC8	MILR1	ATP7A
ALPP	DHX16	RHPN1-AS1	MUCL1	CACNG4	TMSB15A
ALPG	PPP1R18	INSL4	KCNC2	CACNG1	DCX
ECEL1	SFTA2	YBX1P10	EPYC	SLC16A6	AGTR2
SCARNA5	MUCL3	NXNL2	PLA2G1B	SOX9	SLC6A14
FAM198A	MUC21	C9orf152	GPR12	LINC00511	IGSF1
SPINK8	POU5F1	LMX1B	STOML3	FOXJ1	CCDC160
SEMA3B	MICA	OBP2B	FAM216B	CIDEA	RTL8B
SNTN	AIF1	SURF6	SLITRK6	ANKRD30B	SLC9A6
ROPN1	CSNK2B	C9orf116	SOX21	AQP4	VGLL1
ROPN1B	CLIC1	PAEP	CLDN10	CHST9	SLITRK4
TRH	VAR5	UCN3	HS6ST3	SERPINB5	MTM1
COL6A5	SLC44A4	CALML5	F7	SERPINB4	MAGEA6
COL6A6	EHMT2	LINC00707	LINC00641	SERPINB7	MAGEA3
CLDN18	ZBTB12	GATA3-AS1	SFTA3	FUT3	

[0182] 在多个实施例中,一个或多个靶cfRNA分子衍生自选自表8或11-14中的一者或多者的一个或多个基因(例如2、3、5个或多个基因)与选自表1-6中的一者或多者的一个或多个基因(例如2、3、5个或多个基因)的组合。在多个实施例中,一个或多个靶cfRNA分子衍生自选自表8或11-14中的一者或多者的一个或多个基因(例如2、3、5个或多个基因)与选自表7中的一个或多个基因(例如2、3、5个或多个基因)的组合。在多个实施例中,选自表8或11-14的所述表为表11。在多个实施例中,选自表8或11-14的所述表为表12。在多个实施例中,选自表8或11-14的所述表为表13。在多个实施例中,选自表8或11-14的所述表为表14。在多个实施例中,选自表8或11-14的所述表为表13。在多个实施例中,选自表8或11-14的所述表为表8。在多个实施例中,从第一表和第二表中选择多个基因包括在所述第一表和第二表的两者中选择一个或多个基因。在多个实施例中,从第一表和第二表中选择多个基因包括从所述第一表中选择不在于所述第二表中的一个或多个基因,以及从所述第二表中选择不在于所述第一表中的一个或多个基因。在多个实施例中,测量的所述多个靶cfRNA分子来自至少500个基因(例如少于400、300、200、100或50个基因)。

[0183] 在多个实施例中,所述癌症是肺癌,并且检测到高于一阈值的所述多个靶cfRNA分子选自表2、5或12中的一者或多者(例如2、3、5个或多个基因)的多个转录物。在多个实施例中,一个或多个靶cfRNA分子衍生自选自表2、5和12中每一者的一个或多个基因(例如2、3、5个或多个基因)。在多个实施例中,从第一表和第二表中选择多个基因包括在所述第一表和第二表的两者中选择一个或多个基因。在多个实施例中,从第一表和第二表中选择多个基因包括从所述第一表中选择不在于所述第二表中的一个或多个基因,以及从所述第二

表中选择不在于所述第一表中的一个或多个基因。在多个实施例中,测量的所述多个靶cfRNA分子来自少于500个基因(例如少于400、300、200、100或50个基因)。

[0184] 在多个实施例中,所述癌症是乳腺癌,并且检测到高于一阈值的所述多个靶cfRNA分子选自表3、4、6或13中的一者或多者的多个基因(例如2、3、5个或多个基因)的多个转录物。在多个实施例中,一个或多个靶cfRNA分子衍生自选自表3、4、6和13中每一者的一个或多个基因(例如2、3、5个或多个基因)。在多个实施例中,从第一表和第二表中选择多个基因包括在所述第一表和第二表的两者中选择一个或多个基因。在多个实施例中,从第一表和第二表中选择多个基因包括从所述第一表中选择不在于所述第二表中的一个或多个基因,以及从所述第二表中选择不在于所述第一表中的一个或多个基因。在多个实施例中,测量的所述多个靶cfRNA分子来自少于500个基因(例如少于400、300、200、100或50个基因)。

[0185] 在多个实施例中,一个或多个靶cfRNA分子衍生自选自表11中的一个或多个基因(例如2、3、5个或多个基因)与(a)选自表5或表6中的一个或多个基因(例如2、3、5个或多个基因)、及/或(b)选自表7中的一个或多个基因(例如2、3、5个或多个基因)的组合。在多个实施例中,从第一表和第二表中选择多个基因包括在所述第一表和第二表的两者中选择一个或多个基因。在多个实施例中,从第一表和第二表中选择多个基因包括从所述第一表中选择不在于所述第二表中的一个或多个基因,以及从所述第二表中选择不在于所述第一表中的一个或多个基因。在多个实施例中,测量的所述多个靶cfRNA分子来自少于500个基因(例如少于400、300、200、100或50个基因)。

[0186] 在多个实施例中,一个或多个靶cfRNA分子衍生自选自表12中的一个或多个基因(例如2、3、5个或多个基因)与(a)选自表5中的一个或多个基因(例如2、3、5个或多个基因)、及/或(b)选自表7中的一个或多个基因(例如2、3、5个或多个基因)的组合。在多个实施例中,从第一表和第二表中选择多个基因包括在所述第一表和第二表的两者中选择一个或多个基因。在多个实施例中,从第一表和第二表中选择多个基因包括从所述第一表中选择不在于所述第二表中的一个或多个基因,以及从所述第二表中选择不在于所述第一表中的一个或多个基因。在多个实施例中,测量的所述多个靶cfRNA分子来自少于500个基因(例如少于400、300、200、100或50个基因)。

[0187] 在多个实施例中,一个或多个靶cfRNA分子衍生自选自表13中的一个或多个基因(例如2、3、5个或多个基因)与(a)选自表4中的一个或多个基因(例如2、3、5个或多个基因)、(b)选自表6中的一个或多个基因(例如2、3、5个或多个基因)、及/或(c)选自表7中的一个或多个基因(例如2、3、5个或多个基因)的组合。在多个实施例中,从第一表和第二表中选择多个基因包括在所述第一表和第二表的两者中选择一个或多个基因。在多个实施例中,从第一表和第二表中选择多个基因包括从所述第一表中选择不在于所述第二表中的一个或多个基因,以及从所述第二表中选择不在于所述第一表中的一个或多个基因。在多个实施例中,测量的所述多个靶cfRNA分子来自少于500个基因(例如少于400、300、200、100或50个基因)。

[0188] 在多个实施例中,一个或多个靶cfRNA分子衍生自选自表4中的一个或多个基因(例如2、3、5个或多个基因)与(a)选自表3中的一个或多个基因(例如2、3、5个或多个基因)、(b)选自表6中的一个或多个基因(例如2、3、5个或多个基因)、及/或(c)选自表7中的一个或多个基因(例如2、3、5个或多个基因)的组合。在多个实施例中,从第一表和第二表中选择多

个基因包括在所述第一表和第二表的两者中选择一个或多个基因。在多个实施例中,从第一表和第二表中选择多个基因包括从所述第一表中选择不在所述第二表中的一个或多个基因,以及从所述第二表中选择不在所述第一表中的一个或多个基因。在多个实施例中,测量的所述多个靶cfRNA分子来自少于500个基因(例如少于400、300、200、100或50个基因)。

[0189] 在一些实施例中,一个或多个靶cfRNA分子衍生自选自表8中所列的所述多个基因的一个或多个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表8中的至少2、3、4、5、10、15、20或30个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表8中的至少5个基因(例如所述前5个基因,CEACAM5、RH0V、SFTA2、SCGB1D2和IGF2BP1)。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表8中的至少10个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表8中的至少25个基因。在多个实施例中,所述一个或多个靶cfRNA分子包括来自表8中的所述多个基因的所有。在多个实施例中,测量的所述多个靶cfRNA分子来自少于500个基因(例如少于400、300、200、100或50个基因)。在多个实施例中,检测到高于一阈值的所述多个靶cfRNA分子是衍生自选自以下多个基因所组成的群组的多个cfRNA分子:CEACAM5、RH0V、SFTA2、SCGB1D2、IGF2BP1、SFTPA1、CA12、SFTPb、CDH3、MUC6、SLC6A14、HOXC9、AGR3、TMEM125、TFAP2B、IRX2、POTekP、ARHGEF38、GPR87、LMX1B、ATP10B、NELL1、MUC21、SOX9、LINC00993、STMND1、ERVH48-1、SCTR、MAGEA3、MB、LEMD1、SIX4及NXNL2。下表8提供了多个信息丰富的癌症生物标记物的多个示例。

[0190] 表8:

[0191]

CEACAM5	RH0V	SFTA2	SCGB1D2	IGF2BP1
SFTPA1	CA12	SFTPb	CDH3	MUC6
SLC6A14	HOXC9	AGR3	TMEM125	TFAP2B
IRX2	POTekP	ARHGEF38	GPR87	LMX1B
ATP10B	NELL1	MUC21	SOX9	LINC00993
STMND1	ERVH48-1	SCTR	MAGEA3	MB
LEMD1	SIX4	NXNL2		

[0192] 在多个实施例中,检测高于一阈值水平的所述多个靶cfRNA分子的一个或多个的步骤包括(i)检测、(ii)高于背景检测、或(iii)在一水平检测,所述水平比未患有所述病况的多位受试者中的所述多个靶cfRNA分子的一水平高。在多个实施例中,检测高于一阈值的步骤包括检测。在多个实施例中,检测高于一阈值的步骤包括检测高于一阈值。在多个实施例中,检测高于一阈值的步骤包括在一水平检测,所述水平比未患有所述病况的多位受试者中的所述多个靶cfRNA分子的一水平高。

[0193] 在多个实施例中,检测高于一阈值水平的所述多个靶cfRNA分子的一个或多个的步骤包括在一水平检测所述一个或多个靶cfRNA分子,所述水平比未患有所述病况的多位受试者中的一水平高至少大约或正好10倍(例如高15、20、50、100或更多倍)。在多个实施例中,高于一阈值检测包括在一水平检测所述一个或多个靶cfRNA分子,所述水平比未患有所述病况的多位受试者中的一水平高至少大约或正好25倍。在多个实施例中,高于一阈值检测包括在一水平检测所述一个或多个靶cfRNA分子,所述水平未患有所述病况的

多位受试者中的一水平高至少约或正好50倍。

[0194] 在多个实施例中,检测高于一阈值水平的所述多个靶cfRNA分子的一个或多个的步骤包括高于每百万个中0.5至5个读数(RPM)的一阈值检测,例如大约或正好1、1.5、2、2.5、3、3.5、4或大约或正好4.5RPM。在多个实施例中,检测高于一阈值包括高于1RPM检测。在多个实施例中,检测高于一阈值包括高于1RPM检测。在多个实施例中,检测高于一阈值包括高于2RPM检测。在多个实施例中,检测高于一阈值包括高于5RPM检测。

[0195] 多个疾病和多个病症

[0196] 根据本公开的多个实施例的多种方法可用于检测多种疾病或病况的任一者的所述存在或不存在,包括但不限于心血管疾病、肝病或癌症。在一些实施例中,所述多种方法涉及确定一癌症阶段。在一些实施例中,所述癌症阶段是第I期癌症、第II期癌症、第III期癌症或第IV期癌症。

[0197] 在一些实施例中,所述多种方法涉及检测一上皮癌(carcinoma)、一恶性肉瘤(sarcoma)、一骨髓瘤(myeloma)、一白血病(leukemia)、一淋巴瘤(lymphoma)、一胚细胞瘤(blastoma)、一生殖细胞瘤(germ cell tumor)或其任意组合的所述存在或不存在、确定一上皮癌、一恶性肉瘤、一骨髓瘤、一白血病、一淋巴瘤、一胚细胞瘤、一生殖细胞瘤或其任意组合的所述阶段、监测一上皮癌、一恶性肉瘤、一骨髓瘤、一白血病、一淋巴瘤、一胚细胞瘤、一生殖细胞瘤或其任意组合的所述进展及/或分类一上皮癌、一恶性肉瘤、一骨髓瘤、一白血病、一淋巴瘤、一胚细胞瘤、一生殖细胞瘤或其任意组合的。在一些实施例中,所述上皮癌可以是一腺癌(adenocarcinoma)。在其他多个实施例中,所述上皮癌可以是一鳞状细胞癌(squamous cell carcinoma)。在还有其他多个实施例中,所述上皮癌是选自以下所组成的群组:小细胞肺癌(small cell lung cancer)、非小细胞肺癌(non-small-cell lung cancer)、鼻咽(nasopharyngeal)癌、结直肠(colorectal)癌、肛门(anal)癌、肝(liver)癌、膀胱(urinary bladder)癌、子宫颈(cervical)癌、睾丸(testicular)癌、卵巢(ovarian)癌、胃(gastric)癌、食道(esophageal)癌、头颈(head-and-neck)癌、胰脏(pancreatic)癌、前列腺(prostate)癌、肾(renal)癌、甲状腺(thyroid)癌、黑色素瘤(melanoma)和乳癌(breast carcinoma)。在一些实施例中,所述乳癌为激素受体阴性乳癌(hormone receptor negative breast carcinoma)或三阴性乳癌(triple negative breast carcinoma)。

[0198] 在一些实施例中,所述多种方法涉及检测一恶性肉瘤的所述存在或不存在、确定一恶性肉瘤的所述阶段、监测一恶性肉瘤的所述进展及/或分类一恶性肉瘤。在多个实施例中,所述恶性肉瘤可选自以下所组成的群组:骨肉瘤(osteosarcoma)、软骨肉瘤(chondrosarcoma)、平滑肌肉瘤(leiomyosarcoma)、横纹肌肉瘤(rhabdomyosarcoma)、间皮肉瘤(mesothelial sarcoma)(间皮瘤(mesothelioma))、纤维肉瘤(fibrosarcoma)、血管肉瘤(angiosarcoma)、脂肪肉瘤(liposarcoma)、神经胶质瘤(glioma)和星形细胞瘤(astrocytoma)。在还有其他多个实施例中,所述多种方法涉及检测白血病的所述存在或不存在、确定白血病的所述阶段、监测白血病的所述进展及/或分类白血病。在各种实施例中,所述白血病可选自以下所组成的群组:骨髓性(myelogenous)白血病、粒细胞性(granulocytic)白血病、淋巴性(lymphatic)白血病、淋巴球性(lymphocytic)白血病和淋巴胚细胞性(lymphoblastic)白血病。在还有其他多个实施例中,所述多种方法涉及检测一淋巴瘤的所述存在或不存在、确定一淋巴瘤的所述阶段、监测一淋巴瘤的所述进展及/或分

类一淋巴瘤。在各种实施例中,所述淋巴瘤可选自以下所组成的群组:霍奇金淋巴瘤(Hodgkin's lymphoma)和非霍奇金淋巴瘤(Non-Hodgkin's lymphoma)。

[0199] 本公开的多个方面包括多种用于确定一疾病的一起源组织的方法,其中所述起源组织选自以下所组成的群组:胰组织、肝胆(hepatobiliary)组织、肝组织、肺组织、脑组织、神经内分泌组织、子宫组织、肾组织、尿路上皮组织、肾组织、子宫颈组织、乳房组织、脂肪、结肠组织、直肠组织、心脏组织、骨骼肌组织、前列腺组织和甲状腺组织。

[0200] 本公开的多个方面包括多种用于确定一癌细胞类型的方法,其中所述癌细胞类型选自以下所组成的群组:膀胱癌、乳癌、子宫颈癌、结直肠癌、子宫内膜癌(endometrial cancer)、食道癌、胃癌、头/颈癌、肝胆癌、血液癌(hematological cancer)、肝癌、肺癌、一淋巴瘤、一黑色素瘤、多发性骨髓瘤(multiple myeloma)、卵巢癌、胰脏癌、前列腺癌、肾癌、甲状腺癌、尿道癌(urethral cancer)和子宫癌(uterine cancer)。

[0201] 治疗多种病况

[0202] 本文公开的多种方法可用于制定多个治疗决策、指导和监测,以及多种癌症疗法的开发和多个临床试验。例如,可以通过比较在使用多种特定疗法,例如多种分子靶标疗法(多种单克隆药物)、多种化学疗法药物、多个放射方案等或这些的多种组合之前、期间和之后的多个样本中的患者cfRNA来监测治疗功效。在一些实施例中,监测cfRNA以查看某些癌症生物标记物在治疗后是否增加或减少,这可以让一医师在比监测追踪多个传统患者症状的多种方法所提供的时间短得多的时间内改变一治疗(例如继续、停止或改变治疗)。在一些实施例中,一方法还包括基于所述多个源自RNA序列诊断一受试者的所述步骤,例如诊断受试者患有与一检测到的cfRNA生物标记物相关的一特定阶段或癌症类型,或报告所述患者患有或会发展成此癌症的一可能性。在多个实施例中,本文公开的多种方法进一步包括基于检测到的所述病况选择一治疗。在多个实施例中,将所述选择的治疗施用于所述受试者。当所述病况是癌症或一特定癌症类型及/或阶段时,可以选择一合适的抗癌疗法。多种抗癌疗法的多个非限制性示例包括放射疗法、手术切除、施用一抗癌剂(例如一免疫疗法剂、一化学疗法剂等)或这些中的一者或多者的一组合。

[0203] 分类模型

[0204] 本公开的多个方面是有关多个分类模型。例如,一机器学习或深度学习模型(例如一疾病分类器)可用于根据从一个或多个RNA分子或序列读数(衍生自一个或多个cfRNA分子)确定的一个或多个特征的多个值来确定一疾病状态。在各种实施例中,所述机器学习或深度学习模型的所述输出是一疾病状态的一预测评分或概率(例如一预测癌症评分)。因此,所述机器学习或深度学习模型根据所述预测评分或概率生成一疾病状态分类。

[0205] 在一些实施例中,所述机器学习模型包括一逻辑回归分类器。在其他多个实施例中,所述机器学习或深度学习模型可以是一决策树、一集成演算法(ensemble)(例如拔靴整合法(bagging)、提升演算法(boosting)、随机森林)、梯度提升机(gradient boosting machine)、离子、朴素贝叶斯(Naïve Bayes)、支援向量机(support vector machine)或一神经网络之一。所述疾病状态模型包括在训练期间调整的所述多个特征的多个学习权重。此处一般使用所述术语多个权重来表示与一模型的任何给定特征相关联的所述学习量,而不管使用哪种特定的机器学习技术。在一些实施例中,一癌症指标评分是通过将衍生自一个或多个RNA序列(或其多个DNA序列读数)的多个特征的多个值输入到一机器学习或深度

学习模型来确定的。

[0206] 在训练期间,处理训练数据以生成用于训练所述疾病状态模型的所述多个权重的多个特征的多个值。作为一个示例,训练数据可以包括从多个训练样本中获得的cfRNA数据及/或WBC RNA数据,以及一输出标签。例如,所述输出标签可以指示所述个体是否已知患有一特定疾病(例如已知患有癌症)或已知是健康的(即,没有一疾病)。在其他多个实施例中,所述模型可用于确定一疾病类型或起源组织(例如癌症起源组织),或所述疾病的一严重程度的一指示(例如癌症阶段)并为其生成一输出标签。取决于所述实施例,所述疾病状态模型接收从用于检测和量化一cfRNA分子或由其衍生的序列的一RNA测定确定的所述多个特征的一个或多个的所述多个值,以及与要训练的所述模型相关的多个计算分析。在一个实施例中,所述一个或多个特征包括一定量的一个或多个cfRNA分子或由其衍生的多个序列读数。取决于所述训练中的模型输出的所述多个评分与所述训练数据的所述多个输出标签之间的所述多个差异,优化所述预测癌症模型的所述多个权重,使所述疾病状态模型能够做出更准确的多个预测。在各种实施例中,一疾病状态模型可以是一非参数模型(例如k-近邻(k-nearest neighbors)),因此,可以训练所述预测癌症模型以更准确地做出多个预测而无需优化多个参数。

[0207] 所述训练的疾病状态模型可以储存在一电脑可读介质中,并且随后在需要时检索,例如,在所述模型的部署期间。

[0208] 在一些实施例中,所述多种方法涉及通过将一基因表达矩阵(G)与一组织特异性矩阵(TS)相乘,将所述基因表达矩阵(G)转换为一组织评分矩阵(S)。 $G_{m,n}$ 是样本m中基因n的所述表达水平。 $TS_{n,j}$ 是基因n对组织j的所述组织特异性。如果基因n对组织j不特异,则 $TS_{n,j}=0$ 。在一些实施例中,使用所述组织RNA-seq数据库(GTEX)计算所述组织特异性矩阵。所述多个组织评分可作为多个特征来建立多个模型以进行分类,例如癌症样本与非癌症样本。在一个非限制性实施例中,使用从多个肺癌样本中识别的所述多个暗通道基因(SFTPA2、SLC39A4、NKX2_1、SFTPA1、BPIFA1、SLC34A2、CXCL17、SFTA3、MUC1、AGR2、WFDC2、ABCA12、VSIG10、CRABP2)以建立用于区分肺癌和非癌症cfRNA样本的一决策树分类器。此分析的所述多个结果显示于图10中。

[0209] 测序和生物资讯

[0210] 本公开的多个方面包括多个核酸分子的测序,以产生多个序列读数,以及所述多个序列读数的生物资讯操纵,以执行所述多种主题方法。

[0211] 在各种实施例中,从一受试者中收集一样本,然后对感兴趣的多个遗传区域或多个遗传片段进行富集。例如,在一些实施例中,可以通过与包含多个癌症相关基因或感兴趣的多个基因片段的一核苷酸阵列杂交来富集一样本。在一些实施例中,可以使用本领域已知的其他多种方法,例如杂交捕获,来富集一样本中感兴趣的多个基因(例如多个癌症相关基因)。参见例如,Lapidus(美国第7,666,593号专利),其内容通过引用其整体并入本文中。在一种杂交捕获方法中,使用一基于溶液的杂交方法,所述方法包括使用多个生物素化的寡核苷酸和多个链霉亲和素包覆的磁珠。参见例如,Duncavage等人,J Mol Diagn.13(3):325-333(2011);和Newman等人,Nat Med.20(5):548-554(2014)。根据本公开的所述多种方法从一样本中分离核酸可以根据本领域已知的任何方法进行。

[0212] 可以通过本领域已知的任何方法或多种方法的组合测序。例如,多种已知的核酸

测序技术包括,但不限于使用标签的多个终止子或引子及板或毛细管中的凝胶分离的多种经典的双脱氧测序反应(dideoxy sequencing reaction)(Sanger方法)、使用可逆终止的多个标签的核苷酸的合成式测序、焦磷酸测序、454测序、与多个标签的寡核苷酸探针的一文库的等位基因特异性杂交、使用与多个标签的克隆的一文库的等位基因特异性杂交的合成式测序,接着进行连接、实时监测一聚合步骤期间多个标签的核苷酸的所述掺入、Polony测序和SOLiD测序。多个分离分子的测序最近已通过使用多个聚合酶或多个连接酶的顺序或单次延伸反应以及通过使用多个探针的多个文库的单次或顺序差异杂交得到证实。

[0213] 一种常规的执行测序的方法是通过链终止和凝胶分离,如Sanger等人,Proc Natl.Acad.Sci.U S A,74(12):5463-67(1977),其内容通过引用其整体并入本文中。另一种常规测序方法涉及多个核酸片段的化学降解。参见Maxam等人,Proc.Natl.Acad.Sci.,74:560-564(1977),其内容通过引用其整体并入本文中。还开发了基于杂交测序的多种方法。参见例如,Harris等人,(美国第2009/0156412号专利申请案),其内容通过引用其整体并入本文中。

[0214] 可以在所述提供的公开的所述多种方法中使用的一测序技术包括,例如,Helicos True Single Molecule Sequencing(tSMS)(Harris T.D.等人(2008)Science 320:106-109),其内容通过引用其整体并入本文中。对tSMS的进一步描述显示于,例如Lapidus等人中(美国第7,169,560号专利),其内容通过引用其整体并入本文中、Lapidus等人(美国第2009/0191565号专利申请公开案,其内容通过引用其整体并入本文中)、Quake等人(美国第6,818,395号专利,其内容通过引用整体并入本文中)、Harris(美国第7,282,337号专利,其内容通过引用其整体并入本文中)、Quake等人(美国第2002/0164629号专利申请公开案,其内容通过引用其整体并入本文中)、以及Braslavsky等人,PNAS(USA),100:3960-3964(2003),其内容通过引用其整体并入本文中。

[0215] 可以在所述提供的公开的所述多种方法中使用的一核酸测序技术的另一个示例是454测序(Roche)(Margulies,M等人2005,Nature,437,376-380,其内容通过引用其整体并入本文中)。可以在所述提供的公开的所述多种方法中使用的一DNA测序技术的另一示例是SOLiD技术(Applied Biosystems)。可以在所述提供的公开的所述多种方法中使用的一DNA测序技术的另一示例是Ion Torrent测序(美国第2009/0026082号、第2009/0127589号、第2010/0035252号、第2010/0137143号、第2010/0188073号、第2010/0197507号、第2010/0282617号、第2010/0300559号、第2010/0300895号、第2010/0301398号及第2010/0304982号专利申请公开案,其内容通过引用其整体并入本文中)。

[0216] 在一些实施例中,所述测序技术为Illumina测序。Illumina测序基于使用折返PCR(fold-back PCR)和多个锚定引子在一固体表面上扩增DNA。可以片段化基因组DNA,或者在cfDNA的情况下,由于多个片段已经很短,因此不需要片段化。多个转接子连接到所述多个片段的所述5'和3'端。附着在多个流通池通道(follow cell channel)的所述表面的多个DNA片段被延伸并桥接扩增。所述多个片段变成双股,并且所述多个双股分子变性。所述固相扩增后变性的多个循环可以在所述流通池的每个通道中创建数百万个丛集,每个丛集具有大约1,000个拷贝的所述相同模板的多个单股DNA分子。多个引子、DNA聚合酶和四个荧光团标签的可逆终止核苷酸用于执行顺序测序。核苷酸掺入后,使用一激光激发所述多个荧光团,捕获一图像并记录所述第一个碱基的所述身份。去除每个掺入碱基的所述多个3'终

止子和多个荧光团,并重复所述掺入、检测和识别步骤。

[0217] 可以在所述提供的公开的所述多种方法中使用的一测序技术的另一示例包括 Pacific Biosciences 的所述单分子、实时 (single molecule, real-time, SMRT) 技术。可以在所述提供的公开的所述多种方法中使用的一测序技术的又一示例是纳米孔测序 (Soni G V 和 Meller A. (2007) Clin Chem 53:1996-2001, 其内容通过引用其整体并入本文中)。可以在所述提供的公开的所述多种方法中使用的一测序技术的另一示例涉及使用一化学敏感场效应晶体管 (chemical-sensitive field effect transistor, chemFET) 阵列对 DNA 测序 (例如, 如美国第 20090026082 号专利申请公开案中所述, 其内容通过引用其整体并入本文中)。可以在所述提供的公开的所述多种方法中使用的一测序技术的另一示例涉及使用一电子显微镜 (Moudrianakis E.N. 和 Beer M. Proc Natl Acad Sci USA. 1965 March; 53:564-71, 其内容通过引用其整体并入本文中)。

[0218] 如果所述样本中的所述核酸被降解或只能从所述样本中获得一最少量的核酸, 则可以对所述核酸进行 PCR 以获得一足够量的核酸用于测序 (参见, 例如 Mullis 等人的美国第 4,683,195 号专利, 其内容通过引用其整体并入本文中)。

[0219] 多种电脑系统和多种装置

[0220] 本文描述的本公开的多个方面可以使用任何类型的计算装置来执行, 例如一电脑, 其包括一处理器, 例如一中央处理单元, 或多种计算装置的任意组合, 其中每个装置执行所述过程或方法的至少一部分。在一些实施例中, 本文描述的多种系统和多种方法可以用一手持装置执行, 例如一智慧型平板电脑或一智慧型手机, 或为所述系统生产的一专用装置。

[0221] 本公开的多种方法可以使用软件、硬件、固件、硬连线或这些的任一者的多种组合来实现。实现多个功能的多个特征也可以在物理上位于各种位置, 包括分布使得所述多个功能的多个部分在多个不同的物理位置实现 (例如一个空间的成像设备和另一个空间的主机工作站, 或在多个单独的建物中, 例如在具有无线或有线连接的多个单独的建物中)。

[0222] 适于执行多个电脑程式的多个处理器包括, 例如, 通用和专用微处理器, 以及任何种类的数位电脑的任何一个或多个处理器。一般而言, 一处理器将从一唯读存储器 (read-only memory) 或一随机存取存储器 (random-access memory) 或两者接收多个指令和数据。一电脑的所述多个基本要素是用于执行多个指令的一处理器及用于储存多个指令和数据的一个或多个存储器装置。一般而言, 一电脑还将包括或可操作地耦合以从一个或多个用于储存数据的大容量储存装置, 例如多个磁盘、多个磁光盘或多个光盘接收数据, 或将数据传输到一个或多个用于储存数据的大容量储存装置, 例如多个磁盘、多个磁光盘或多个光盘, 或两者兼具。适用于体现多个电脑程式指令和数据的多个信息载体包括非易失性存储器的所有形式, 包括例如多种半导体存储器装置 (例如可抹除可程式唯读存储器 (EPROM)、电子可抹除可程式唯读存储器 (EEPROM)、固态硬盘 (solid state drive, SSD) 和多种快闪存储器装置); 多种磁盘 (例如多种内部硬盘或多种可移磁盘); 多种磁光盘; 及多种光盘 (例如多种 CD 和 DVD 光盘)。所述处理器和所述存储器可以由专用逻辑电路补充或合并到专用逻辑电路中。

[0223] 为了提供与一用户的互动, 本文描述的所述主题可以在具有一 I/O 装置的一电脑上实现, 例如一 CRT、LCD、LED 或用于向所述用户显示信息的投影装置和一输入或输出装置,

例如一键盘和一定点设备 (pointing device) (例如一鼠标或一轨迹球 (trackball)), 所述用户可以通过它们向所述电脑提供输入。也可以使用其他多个种类的装置来提供与一用户的互动。例如, 提供给所述用户的反馈可以是感官反馈的任何形式 (例如视觉反馈、听觉反馈或触觉反馈), 并且可以以任何形式接收来自所述用户的输入, 包括声音、语音或触觉输入。

[0224] 本文描述的所述主题可以在包括一后端组件 (例如一数据伺服器)、一中间件组件 (例如一应用伺服器) 或一前端组件 (例如具有一图形用户界面或一网络浏览器的一客户端电脑, 一用户可以通过它与本文描述的所述主题的一实现进行互动), 或此类后端、中间件和前端组件的任意组合。所述系统的所述多个组件可以通过数位数据通信的任何形式或介质的一网络, 例如一通信网络互连。例如, 一参考数据集可以储存在一远程位置并且一电脑可以通过一网络通信, 以存取所述参考数据集以用于多个比较目的。然而, 在其他多个实施例中, 一参考数据集可以局部地储存在所述电脑内, 并且所述电脑存取所述CPU内的所述参考数据集以用于多个比较目的。多种通信网络的多个示例包括, 但不限于多种蜂巢式网络 (cell network) (例如3G或4G)、一局部区域网络 (local area network, LAN) 和一广域网络 (wide area network, WAN), 例如所述Internet。

[0225] 本文描述的所述主题可以实现为一个或多个电脑程式产品, 例如有形地体现在一信息载体中 (例如在一非暂时性电脑可读介质中) 的一个或多个电脑程式, 用于由一数据处理设备 (例如一可编程处理器、一电脑或多台电脑) 执行或控制一数据处理设备 (例如一可编程处理器、一电脑或多台电脑) 的所述操作。一电脑程式 (也称为一程式、软件、软件应用、應用程式 (app)、巨集或代码) 可以用编程语言的任何形式编写, 包括编译或解释语言 (例如C、C++、Perl), 并且它可以以任何形式部署, 包括作为一独立程式或作为一模块、组件、子程序或其他适合在一计算环境中使用的单元。本公开的多种系统和多种方法可以包括以本领域已知的任何合适的编程语言编写的多个指令, 包括但不限于C、C++、Perl、Java、ActiveX、HTML5、Visual Basic或JavaScript。

[0226] 一电脑程式不一定对应于一文件。一程式可以储存在一文件或一文件的一部分中, 所述文件或文件的所述部分包含其他多个程式或数据、储存在专用于有问题程式的一单个文件中、或储存在多个协调文件中 (例如储存一个或多个模块的多个文件、多个子程式或代码的多个部分)。一电脑程式可以部署为在一电脑上或在一个站点的多台电脑上执行, 或者分布在多个站点并通过一通信网络互连。

[0227] 一文件可以是一数位文件, 例如, 储存在一硬盘、SSD、CD或其他有形的非暂时性介质上。一文件可以通过一网络从一装置发送到另一装置 (例如作为多个数据包从一伺服器发送到一客户端, 例如, 通过一网络界面卡、数据机、无线网卡等)。

[0228] 根据本公开写入一文件涉及例如, 通过添加、移除或重新排列多个粒子转换一有形的、非暂时性电脑可读介质 (例如通过多个读/写头将一净电荷或偶极矩转换为磁化的多个模式), 然后这些模式代表了关于所述用户期望和有利的客观物理现象的信息的多个新搭配。在一些实施例中, 写入涉及多个有形的、非暂时性电脑可读介质中材料的一物理转换 (例如具有某些光学特性, 以便多个光学读/写装置可以读取所述新的且有利的信息搭配, 例如刻录一光碟唯读存储器 (CD-ROM))。在一些实施例中, 写入一文件包括转换一物理快闪存储设备, 例如NAND快闪存储装置和通过转换由多个浮栅晶体管 (floating-gate

transistor) 制成的多个储存单元的一阵列中的多个物理元素来储存信息。写入一文件的多种方法是本领域所周知的,例如,可以通过一程式手动或自动调用,或者通过来自软件的一保存命令或来自一编程语言的一写入命令调用。

[0229] 多种合适的计算装置通常包括大容量存储器、至少一个图形用户界面、至少一个显示装置,并且通常包括多个装置之间的通信。所述大容量存储器说明一类型的多个电脑可读介质,即多个电脑储存介质。多个电脑储存介质可以包括以用于存储诸如多个电脑可读指令、多个数据结构、多个程式模块或其他数据的信息的任何方法或技术实现的易失性、非易失性、可移动和不可移动介质。多个电脑储存介质的多个示例包括随机存取存储器(RAM)、唯读存储器(ROM)、EEPROM、快闪存储器或其他存储器技术、CD-ROM、多个数位多功能光碟(digital versatile disk,DVD)或其他光储存、多个磁胶片盒(magnetic cassette)、磁带(magnetic tape)、磁盘储存或其他多个磁储存装置、多个无限射频识别(Radiofrequency Identification,RFID)标志或芯片、或可用于储存所述需要信息并可由一计算装置存取的任何其他介质。

[0230] 本文描述的多个功能可以使用软件、硬件、固件、硬连线或这些的任一者的多种组合来实现。所述软件的任一者都可以在物理上位于各种位置,包括分布使得所述多个功能的多个部分在多个不同的物理位置实现。

[0231] 如本领域技术人员将认识到的,对于执行本公开的所述多种方法是必要的或最适合的,用于实施所述多种描述的发明方法的一些或所有的一电脑系统可以包括一个或多个处理器(例如一中央处理单元(central processing unit,CPU)、一图形处理单元(graphics processing unit,GPU)、或两者)、主存储器和静态存储器,它们通过一总线相互通信。

[0232] 一处理器通常将包括一芯片,例如一单核心或多核心芯片,以提供一中央处理单元(CPU)。一处理可以由Intel或AMD的一芯片提供。

[0233] 存储器可以包括一个或多个机器可读装置,在其上储存一组或多组的多个指令(例如软件),当由所述公开的多个电脑的任何一者的所述处理器执行时,这些指令可以完成此处描述的所述多种方法或多种功能的一些或所有。在所述电脑系统执行所述软件期间,所述软件还可以完全或至少部分地驻留在所述主存储器内及/或所述处理器内。优选地,每台电脑包括一非暂时性存储器,例如一固态硬盘、快闪硬盘(flash drive)、磁盘硬盘(disk drive)、硬盘(hard drive)等。

[0234] 虽然所述多个机器可读装置在一示例性实施例中可以是一单个介质,但所述术语“机器可读装置”应被理解为包括储存一组或多组的多个指令及/或数据的一单个介质或多种介质(例如一集中式或分布式数据库,及/或关联的多个快取(cache)和多个伺服器)。这些术语还应当被理解为包括能够储存、编码或保存一组的多个指令以供所述机器执行并使所述机器执行本公开的所述多种方法的任何一种或多种的任何介质或多种介质。因此,这些术语应被视为包括,但不限于一个或多个固态存储器(例如用户身份模块(subscriber identity module,SIM)卡、安全数位卡(secure digital card,SD card)、微型SD卡或固态硬盘(SSD))、多种光学和磁性介质、及/或任何其他有形储存介质或多种介质。

[0235] 本公开的一电脑将通常包括一个或多个I/O装置,例如一视频显示单元(例如一液晶显示器(liquid crystal display,LCD)或一阴极射线管(cathode ray tube,CRT))的一

者或多者、一字母数字输入装置(例如一键盘)、一游标控制装置(例如一鼠标)、一磁盘硬盘单元、一信号生成装置(例如一扬声器)、一触摸屏、一加速度计、一麦克风、一蜂巢式射频天线(cellular radio frequency antenna)和一网络界面装置,例如可以是一网络界面卡(network interface card,NIC)、Wi-Fi卡或蜂巢式数据机。

[0236] 所述软件的任一者都可以在物理上位于各种位置,包括分布使得所述多个功能的多个部分在多个不同的物理位置实现。

[0237] 另外地,可以提供本公开的多种系统以包括参考数据。可以储存任何合适的基因组数据以供在所述系统内使用。多个示例包括,但不限于:来自癌症基因体图谱计划(The Cancer Genome Atlas,TCGA)的癌症的多个主要类型和多个亚型的所述多个关键基因组变化的多个全面、多维图谱;来自国际癌症基因组联盟(The International Cancer Genome Consortium,ICGC)的多个基因组异常的一目录;来自COSMIC的癌症中的多个体细胞突变的一目录;所述人类基因组和其他多个流行模式生物的最新构建;来自dbSNP的多个最新参考SNP;来自所述千人基因组计划(1000Genomes Project)和所述博德研究所(Broad Institute)的多个金标准插入缺失;来自Illumina、Agilent、Nimblegen和Ion Torrent的多个外显体(exome)捕获试剂盒注解;转录物注解;用于多个管道的试验的小测试数据(例如对于多个新用户)。

[0238] 在一些实施例中,数据在一系统中包括的一数据库的所述上下文中可用。可以使用任何合适的数据库结构,包括多个关系数据库(relational database)、多个物件导向数据库(object-oriented database)等。在一些实施例中,参考数据储存在诸如一“非关联式(not-only SQL)”(NoSQL)数据库的一关系数据库中。在各种实施例中,一图形数据库(graph database)被包括在本公开的多种系统内。还应理解,此处使用的所述术语“数据库”不限于一个单一的数据库;相反,多个数据库可以包括在一系统中。例如,根据本公开的多个实施例,一数据库可以包括两个、三个、四个、五个、六个、七个、八个、九个、十个、十五个、二十个或更多个单独的数据库,包括其中的多个数据库的任何整数。例如,一个数据库可以包含公共参考数据,一第二数据库可以包含来自一患者的测试数据,一第三数据库可以包含来自多位健康受试者的数据,以及一第四数据库可以包含来自患有一已知病况或病症的多位患病受试者的数据。应当理解,此处描述的所述多种方法还考虑了关于其中包含的所述数据的多个数据库的任何其他配置。

[0239] 多个示例性实施例(A)

[0240] 本发明提供以下多个实施例:

[0241] 实施例P1.一种检测一受试者中癌症的方法,所述方法包括以下步骤:

[0242] (a) 测量所述受试者的一样本中的多个靶血浆游离RNA(cfRNA)分子,其中所述多个靶cfRNA分子选自表11,并且可选地选自表12-15中的一个或多个转录物;及

[0243] (b) 检测所述癌症,其中检测所述癌症包括检测高于一阈值水平的所述多个靶cfRNA分子中的一个或多个。

[0244] 实施例P2.如实施例P1所述的方法,其中所述多个靶cfRNA分子选自来自表11-14中的至少5、10、15或20个转录物。

[0245] 实施例P3.如实施例P1或P2所述的方法,其中所述多个靶cfRNA分子包括来自以下的多个转录物:(i) 表11;(ii) 表2、5和12中的每一者;(iii) 表3、4、6和13中的每一者;或

(iv) 表14。

[0246] 实施例P4.如实施例P1-P3中任一者所述的方法,其中所述多个靶cfRNA分子包括来自表11-15中的一者或多者所述多个转录物的所有。

[0247] 实施例P5.如实施例P1-P4中任一者所述的方法,其中所述多个靶cfRNA分子是选自表14中的多个转录物。

[0248] 实施例P6.如实施例P1-P5中任一者所述的方法,其中检测到高于一阈值的所述多个靶cfRNA分子是衍生自选自以下多个基因所组成的群组的多个cfRNA分子:ADIPOQ、AGR3、ANKRD30A、AQP4、BPIFA1、CA12、CEACAM5、CFTR、CXCL17、CYP4F8、FABP7、FOXI1、GGTLC1、GP2、IL20、ITIH6、LDLRAD1、LEMD1、LMX1B、MMP7、NKAIN1、NKX2-1、ROPN1、ROS1、SCGB1D2、SCGB2A2、SFTA2、SFTA3、SLC34A2、SOX9、STK32A、STMND1、TFAP2A、TFAP2B、TFF1、TRPV6、VGLL1、VTCN1。

[0249] 实施例P7.如实施例P1-P6中任一者所述的方法,其中所述多个靶cfRNA分子包括表11-14中的一者或多者的多个转录物以及表1-6中的一个或多个转录物。

[0250] 实施例P8.如实施例P1-P7中任一者所述的方法,其中所述多个靶cfRNA分子包括表11-14中的一者或多者的多个转录物以及表7中的一个或多个转录物。

[0251] 实施例P9.如实施例P1-P4中任一者所述的方法,其中(i)所述癌症是肺癌,及(ii)检测到高于一阈值的所述多个靶cfRNA分子是选自表2、5或12中的一者或多者的多个转录物。

[0252] 实施例P10.如实施例P1-P4中任一者所述的方法,其中(i)所述癌症是乳癌,及(ii)检测到高于一阈值的所述多个靶cfRNA分子是选自表3、4、6或13中的一者或多者的多个转录物。

[0253] 实施例P11.如实施例P1-P10中任一者所述的方法,其中所述测量包括测序、微阵列分析、反转录PCR、实时PCR、定量实时PCR、数位PCR、数位微滴PCR、数位乳化PCR、多重PCR、杂交捕获、多种寡核苷酸连接测定或其任意组合。

[0254] 实施例P12.如实施例P1-P11中任一者所述的方法,其中所述测量的步骤包括对多个cfRNA分子测序,以产生多个cfRNA序列读数。

[0255] 实施例P13.如实施例P12所述的方法,其中对所述多个cfRNA分子测序的步骤包括全转录组测序。

[0256] 实施例P14.如实施例P12或P13所述的方法,其中对所述多个cfRNA分子测序的步骤包括反转录以产生多个cDNA分子,以及对所述多个cDNA分子测序以产生所述多个cfRNA序列读数。

[0257] 实施例P15.如实施例P12所述的方法,其中对所述多个cfRNA分子测序的步骤包括富集所述多个靶cfRNA分子或其多个cDNA分子。

[0258] 实施例P16.如实施例P1-P15中任一者所述的方法,其中所述样本包括一生物流体。

[0259] 实施例P17.如实施例P16所述的方法,其中所述生物流体包括血液、血浆、血清、尿液、唾液、脑膜液、心包液、脑脊液(CSF)、腹膜液或其任意组合。

[0260] 实施例P18.如实施例P16所述的方法,其中所述生物流体包括所述受试者的血液、一血液级分、血浆或血清。

[0261] 实施例P19.如实施例P1-P18中任一者所述的方法,其中检测高于一阈值水平的所述多个靶cfRNA分子的一个或多个的步骤包括(i)检测、(ii)高于背景检测、或(iii)在一水平检测,所述水平比未患有所述病况的多位受试者中的所述多个靶cfRNA分子的一水平高。

[0262] 实施例P20.如实施例P1-P18中任一者所述的方法,其中检测高于一阈值水平的所述多个靶cfRNA分子中的一个或多个的步骤包括在一水平检测所述一个或多个靶cfRNA分子,所述水平比未患有所述病况的多位受试者的一水平高至少约10倍。

[0263] 实施例P21.如实施例P12-P18中任一者所述的方法,其中检测高于一阈值水平的所述多个靶cfRNA分子中的一个或多个的步骤包括检测高于每百万个中0.5至5个读数(RPM)的一阈值。

[0264] 实施例P22.如实施例P1-P18中任一者所述的方法,其中检测高于一阈值水平的所述多个靶cfRNA分子中的一个或多个包括以下步骤:

[0265] (a)通过将所述多个靶cfRNA分子的每一者的所述表达水平与一RNA组织评分矩阵进行比较,确定每一个靶cfRNA分子的一指标评分;

[0266] (b)汇总每一个靶cfRNA分子的所述多个指标评分;及

[0267] (c)当所述指标评分超过一阈值时检测所述癌症。

[0268] 实施例P23.如实施例P12-P22中任一者所述的方法,其中检测高于一阈值水平的所述多个靶cfRNA分子中的一个或多个的步骤包括将所述多个序列读数输入到一机器学习或深度学习模型中。

[0269] 实施例P24.如实施例P23所述的方法,其中所述机器学习或深度学习模型包括逻辑回归、随机森林、梯度提升机、朴素贝叶斯、神经网络或多项回归。

[0270] 实施例P25.如实施例P23所述的方法,其中所述机器学习或深度学习模型通过包括多个学习权重的一函数将所述一个或多个特征的所述多个值转换为所述受试者的所述疾病状态预测。

[0271] 实施例P26.如实施例P1-P25中任一者所述的方法,其中所述癌症包括:

[0272] (i)一上皮癌、一恶性肿瘤、一骨髓瘤、一白血病、一淋巴瘤、一胚胎瘤、一生殖细胞瘤或其任意组合;

[0273] (ii)选自以下所组成的群组的一上皮癌:腺癌、鳞状细胞癌、小细胞肺癌、非小细胞肺癌、鼻咽癌、结直肠癌、肛门癌、肝癌、膀胱癌、睾丸癌、子宫颈癌、卵巢癌、胃癌、食道癌、头颈癌、胰脏癌、前列腺癌、肾癌、甲状腺癌、黑色素瘤和乳癌;

[0274] (iii)激素受体阴性乳癌或三阴性乳癌;

[0275] (iv)选自以下所组成的群组的一恶性肿瘤:骨肉瘤、软骨肉瘤、平滑肌肉瘤、横纹肌肉瘤、间皮肉瘤(间皮瘤)、纤维肉瘤、血管肉瘤、脂肪肉瘤、神经胶质瘤和星形细胞瘤;

[0276] (v)选自以下所组成的群组的一白血病:骨髓性白血病、粒细胞性白血病、淋巴性白血病、淋巴球性白血病和淋巴胚细胞性白血病;或者

[0277] (vi)选自以下所组成的群组的一淋巴瘤:霍奇金淋巴瘤和非霍奇金淋巴瘤。

[0278] 实施例P27.如实施例P1-P26中任一者所述的方法,其中检测所述癌症的步骤包括确定一癌症阶段、确定癌症进展、确定一癌症类型、确定癌症起源组织或其一组合。

[0279] 实施例P28.如实施例P1-P27中任一者所述的方法,其中所述方法进一步包括根据

所述检测到的癌症选择一治疗。

[0280] 实施例P29.如实施例P28所述的方法,其中所述治疗包括手术切除、放射疗法或施用一抗癌剂。

[0281] 实施例P30.如实施例P28或P29所述的方法,其中所述方法进一步包括用所述选择的治疗治疗所述受试者。

[0282] 实施例P31.一种用于执行实施例P1-P30中任一者所述的方法中的一个或多个步骤的电脑系统。

[0283] 实施例P32.一种非暂时性电脑可读介质,其上储存有用于执行实施例P1-P30中任一者所述的方法中的一个或多个步骤的多个电脑可读指令。

[0284] 多个示例性实施例(B)

[0285] 本发明提供以下多个实施例:

[0286] 实施例1.一种检测一受试者中癌症的方法,所述方法包括以下步骤:

[0287] (a) 测量所述受试者的一样本中的多个标靶血浆游离RNA(cfRNA)分子,其中所述多个标靶cfRNA分子选自表11中的一个或多个转录物;及

[0288] (b) 检测所述癌症,其中检测所述癌症包括检测高于一阈值水平的所述多个标靶cfRNA分子中的一个或多个。

[0289] 实施例2.如实施例1所述的方法,其中所述多个标靶cfRNA分子选自表8或12-15中的一者或多者。

[0290] 实施例3.如实施例1所述的方法,其中所述多个标靶cfRNA分子选自来自表8或11-14中的一者或多者的至少5、10、15或20个基因的多个转录物。

[0291] 实施例4.如实施例1或3所述的方法,其中所述多个标靶cfRNA分子包括来自以下的多个转录物:(i)表11;(ii)表2、5和12中的每一者;(iii)表3、4、6和13中的每一者;(iv)表14;或(v)表8。

[0292] 实施例5.如实施例1-4中任一者所述的方法,其中所述多个标靶cfRNA分子包括来自表8或11-15中的一者或多者的至少30个基因的多个转录物。

[0293] 实施例6.如实施例1-5中任一者所述的方法,其中所述多个标靶cfRNA分子是选自表14中的多个转录物。

[0294] 实施例7.如实施例1-6中任一者所述的方法,其中检测到高于一阈值的所述多个标靶cfRNA分子是衍生自选自以下多个基因所组成的群组的多个cfRNA分子:ADIPOQ、AGR3、ANKRD30A、AQP4、BPIFA1、CA12、CEACAM5、CFTR、CXCL17、CYP4F8、FABP7、FOXI1、GGTLC1、GP2、IL20、ITIH6、LDLRAD1、LEMD1、LMX1B、MMP7、NKAIN1、NKX2-1、ROPN1、ROS1、SCGB1D2、SCGB2A2、SFTA2、SFTA3、SLC34A2、SOX9、STK32A、STMND1、TFAP2A、TFAP2B、TFF1、TRPV6、VGLL1及VTCN1。

[0295] 实施例8.如实施例1-5中任一者所述的方法,其中所述多个标靶cfRNA分子是选自表8中的多个转录物。

[0296] 实施例9.如实施例1-5中任一者所述的方法,其中检测到高于一阈值的所述多个标靶cfRNA分子是衍生自选自以下多个基因所组成的群组的多个cfRNA分子:CEACAM5、RHOF、SFTA2、SCGB1D2、IGF2BP1、SFTPA1、CA12、SFTPB、CDH3、MUC6、SLC6A14、HOXC9、AGR3、TMEM125、TFAP2B、IRX2、POTKP、ARHGEF38、GPR87、LMX1B、ATP10B、NELL1、MUC21、SOX9、

LINC00993、STMND1、ERVH48-1、SCTR、MAGEA3、MB、LEMD1、SIX4及NXNL2。

[0297] 实施例10.如实施例1-9中任一者所述的方法,其中所述多个靶cfRNA分子包括(a)表8或11-14中的一者或多者的多个转录物、及(b)表1-6中的一个或多个转录物。

[0298] 实施例11.如实施例1-10中任一者所述的方法,其中所述多个靶cfRNA分子包括(a)表8或11-14中的一者或多者的多个转录物、及(b)表7中的一个或多个转录物。

[0299] 实施例12.如实施例1-5中任一者所述的方法,其中(i)所述癌症是肺癌,及(ii)检测到高于一阈值的所述多个靶cfRNA分子是选自表2、5或12中的一者或多者的多个转录物。

[0300] 实施例13.如实施例1-5中任一者所述的方法,其中(i)所述癌症是乳癌,及(ii)检测到高于一阈值的所述多个靶cfRNA分子是选自表3、4、6或13中的一者或多者的多个转录物。

[0301] 实施例14.如实施例1-13中任一者所述的方法,其中所述测量包括测序、微阵列分析、反转录PCR、实时PCR、定量实时PCR、数位PCR、数位微滴PCR、数位乳化PCR、多重PCR、杂交捕获、多种寡核苷酸连接测定或其任意组合。

[0302] 实施例15.如实施例1-14中任一者所述的方法,其中所述测量的步骤包括对多个cfRNA分子测序,以产生多个cfRNA序列读数。

[0303] 实施例16.如实施例15所述的方法,其中对所述多个cfRNA分子测序的步骤包括全转录组测序。

[0304] 实施例17.如实施例15或16所述的方法,其中对所述多个cfRNA分子测序的步骤包括反转录以产生多个cDNA分子,以及对所述多个cDNA分子测序以产生所述多个cfRNA序列读数。

[0305] 实施例18.如实施例15所述的方法,其中对所述多个cfRNA分子测序的步骤包括富集所述多个靶cfRNA分子或其多个cDNA分子。

[0306] 实施例19.如实施例1-18中任一者所述的方法,其中所述样本包括一生物流体。

[0307] 实施例20.如实施例19所述的方法,其中所述生物流体包括血液、血浆、血清、尿液、唾液、肋膜液、心包液、脑脊液(CSF)、腹膜液或其任意组合。

[0308] 实施例21.如实施例19所述的方法,其中所述生物流体包括所述受试者的血液、一血液级分、血浆或血清。

[0309] 实施例22.如实施例1-21中任一者所述的方法,其中检测高于一阈值水平的所述多个靶cfRNA分子的一个或多个的步骤包括(i)检测、(ii)高于背景检测、或(iii)在一水平检测,所述水平比未患有所述病况的多位受试者中的所述多个靶cfRNA分子的一水平高。

[0310] 实施例23.如实施例1-21中任一者所述的方法,其中检测高于一阈值水平的所述多个靶cfRNA分子中的一个或多个的步骤包括在一水平检测所述一个或多个靶cfRNA分子,所述水平比未患有所述病况的多位受试者的一水平高至少10倍。

[0311] 实施例24.如实施例15-21中任一者所述的方法,其中检测高于一阈值水平的所述多个靶cfRNA分子中的一个或多个的步骤包括检测高于每百万个中0.5至5个读数(RPM)的一阈值。

[0312] 实施例25.如实施例1-21中任一者所述的方法,其中检测高于一阈值水平的所述

多个靶cfRNA分子中的一个或多个包括以下步骤：

[0313] (a) 通过将所述多个靶cfRNA分子的每一者的所述表达水平与一RNA组织评分矩阵进行比较，确定每一个靶cfRNA分子的一指标评分；

[0314] (b) 汇总每一个靶cfRNA分子的所述多个指标评分；及

[0315] (c) 当所述指标评分超过一阈值时检测所述癌症。

[0316] 实施例26. 如实施例15-25中任一者所述的方法，其中检测高于一阈值水平的所述多个靶cfRNA分子中的一个或多个的步骤包括将所述多个序列读数输入到一机器学习或深度学习模型中。

[0317] 实施例27. 如实施例26所述的方法，其中所述机器学习或深度学习模型包括逻辑回归、随机森林、梯度提升机、朴素贝叶斯、神经网络或多项回归。

[0318] 实施例28. 如实施例26所述的方法，其中所述机器学习或深度学习模型通过包括多个学习权重的一函数将所述一个或多个特征的所述多个值转换为所述受试者的所述疾病状态预测。

[0319] 实施例29. 如实施例1-28中任一者所述的方法，其中所述癌症包括：

[0320] (i) 一上皮癌、一恶性肿瘤、一骨髓瘤、一白血病、一淋巴瘤、一胚胎瘤、一生殖细胞瘤或其任意组合；

[0321] (ii) 选自以下所组成的群组的一上皮癌：腺癌、鳞状细胞癌、小细胞肺癌、非小细胞肺癌、鼻咽癌、结直肠癌、肛门癌、肝癌、膀胱癌、睾丸癌、子宫颈癌、卵巢癌、胃癌、食道癌、头颈癌、胰脏癌、前列腺癌、肾癌、甲状腺癌、黑色素瘤和乳癌；

[0322] (iii) 激素受体阴性乳癌或三阴性乳癌；

[0323] (iv) 选自以下所组成的群组的一恶性肿瘤：骨肉瘤、软骨肉瘤、平滑肌肉瘤、横纹肌肉瘤、间皮肉瘤(间皮瘤)、纤维肉瘤、血管肉瘤、脂肪肉瘤、神经胶质瘤和星形细胞瘤；

[0324] (v) 选自以下所组成的群组的一白血病：骨髓性白血病、粒细胞性白血病、淋巴性白血病、淋巴球性白血病和淋巴胚细胞性白血病；或者

[0325] (vi) 选自以下所组成的群组的一淋巴瘤：霍奇金淋巴瘤和非霍奇金淋巴瘤。

[0326] 实施例30. 如实施例1-29中任一者所述的方法，其中检测所述癌症的步骤包括确定一癌症阶段、确定癌症进展、确定一癌症类型、确定癌症起源组织或其一组合。

[0327] 实施例31. 如实施例1-30中任一者所述的方法，其中所述方法进一步包括根据所述检测到的癌症选择一治疗。

[0328] 实施例32. 如实施例31所述的方法，其中所述治疗包括手术切除、放射疗法或施用一抗癌剂。

[0329] 实施例33. 如实施例31或32所述的方法，其中所述方法进一步包括用所述选择的治疗治疗所述受试者。

[0330] 实施例34. 一种测量一样本中多个靶血浆游离RNA (cfRNA) 分子的方法，所述方法包括以下步骤：

[0331] (a) 富集所述多个靶cfRNA分子或其多个cDNA分子，以产生多个多核苷酸的一富集样本；及

[0332] (b) 对所述富集样本的所述多个多核苷酸或其多个扩增产物测序；

[0333] 其中，所述多个靶cfRNA分子选自表11中的一个或多个转录物。

[0334] 实施例35.如实施例34所述的方法,其中所述多个靶cfRNA分子选自表8或12-15中的一者或多者。

[0335] 实施例36.如实施例34所述的方法,其中所述多个靶cfRNA分子选自表8。

[0336] 实施例37.如实施例34所述的方法,其中所述多个靶cfRNA分子选自表14。

[0337] 实施例38.如实施例34所述的方法,其中所述多个靶cfRNA分子选自来自表8或11-14中的一者或多者的至少5、10、15或20个基因的多个转录物。

[0338] 实施例39.一种用于执行实施例1-38中任一项所述方法中的一个或多个步骤的电脑系统。

[0339] 实施例40.一种非暂时性电脑可读介质,其上储存有用于执行实施例1-38中任一项所述方法中的一个或多个步骤的多个电脑可读指令。

[0340] 多个示例

[0341] 应当理解,本文描述的所述多个示例和所述多个实施例仅用于说明目的,并且将向本领域技术人员建议根据其进行的各种修改或变化,并且将包括在本申请的精神和范围内以及所附权利要求的范围内。

[0342] 示例1:检测多位癌症患者所述血浆中的组织特异性RNA

[0343] 血浆游离RNA(cfRNA)是一种很有前途的癌症检测分析物,但缺乏对cfRNA的全面评估。为了表征血浆中源自肿瘤的RNA,我们对一循环血浆游离基因组图谱(Circulating Cell-free Genome Atlas,CCGA)子研究进行了一探索性分析,以检查患有和未患有癌症的多位参与者的cfRNA表达。此分析重点关注乳癌、肺癌和结直肠癌,因为它们在所述普通人群和CCGA中的发病率很高。

[0344] 我们从所述CCGA训练集中选择了210名参与者(Klein等人,ASCO,2018)。共有98名参与者在抽血时被诊断出患有第III期癌症(乳癌(47名患者)、肺癌(32名患者)、结直肠癌(15名患者)和肛门直肠癌(4名患者))。选择多个第III期样本以最大化所述血液中的信号,并避免来自潜在的多个继发性转移的混淆信号。还包括112名与所述癌症组频率年龄匹配的非癌症参与者。对于每位参与者,生成了来自多个肿瘤组织活检的肤色血球层、cfRNA和FFPE的多个全转录组文库。

[0345] 从参与者血浆中提取多个核酸,对所述多个样本进行DNase处理以去除血浆游离DNA(cfDNA)和基因组DNA,并使用多个随机六聚体引子进行反转录,以捕获每个研究参与者的所述全转录组。所述得到的cDNA被转化为多个DNA文库,被扩增,并去除了由核糖体转录物、粒线体和血液相关的转录物,例如多个球蛋白(globin)产生的丰富序列。所述产生的多个全转录组RNA-seq文库在每个样本约750M双端读数的一深度进行测序,并使用一定制的生物信息管道进行分析,所述管道在一逐个样本的基础上为每个基因生成多个UMI折叠计数。当可用时,使用此相同的程序从匹配的肤色血球层和组织RNA创建和分析多个RNA-seq文库。由于残留DNA污染的所述存在,所有下游分析都依赖于使用多个严格的RNA读数,在此示例中定义为多个读数对,其中至少一个读数与一外显子-外显子连接处重叠。如图11显示了所述端到端工作流程的一摘要。表9提供了多个参与者样本的一摘要:

表 9

[0346]

疾病状态	通过的 QC	cfRNA	WBC	组织
乳癌	没通过	1	0	0
肺癌	没通过	2	1	0
非癌症	没通过	4	0	0
肛门直肠癌	通过	4	1	4
乳癌	通过	46	32	40
结直肠癌	通过	15	11	10
肺癌	通过	30	26	12
非癌症	通过	89	93	0
年轻健康	通过	19	19	0
总计	不适用	210	183	66

[0347] 我们将我们的数据与来自TCGA的多个RNA样本进行了比较(图12A)。当我们将CCGA肿瘤组织RNA-seq数据投影到衍生自TCGA肿瘤组织RNA-seq数据的所述多个主要组分上时,所述多个CCGA肿瘤组织样本可按癌症类型分离(图12B)。这些结果显示,尽管样本收集/处理/文库制备存在多个差异,但多个CCGA和TCGA肿瘤的所述多个表达谱非常相似,并验证了所述分析方法。将来自所述CCGA队列的多个癌症cfRNA样本投影到衍生自TCGA肿瘤组织RNA-seq数据的所述多个主要组分上,显示所述样本没有按癌症类型分离(图12C),这意味着癌症类型不是cfRNA变异的所述主要来源。

[0348] 血浆中的大部分cfRNA被认为源自多个健康的免疫细胞。因此,我们将这些转录物视为背景噪音,并将源自肿瘤的cfRNA作为癌症信号的一来源。我们的分析识别了cfRNA数据中的两类基因:“多个暗通道”和“多个暗通道生物标记物”。多个暗通道是未在多位非癌症参与者的所述cfRNA中检测到的多个基因。在57,783个注解的基因中,39,564个(68%)被识别为多个暗通道。多个暗通道生物标记物(DCB)基因符合三个标准:1)所述非癌症队列中所述基因的中位表达为零,2)在所述癌症队列中不止一名参与者检测到基因表达,以及3)基因表达在所述癌症组中上调。

[0349] 针对肺癌识别了14个DCB基因:SLC34A2、GABRG1、ROS1、AGR2、GNAT3、SFTPA2、MUC5B、SFTA3、SMIM22、CXCL17、BPIFA1、WFDC2、NKX2-1和GGTLC1(参见表2)。针对乳癌识别了10个DCB基因:RNU1-1、CSN1S1、FABP7、OPN1SW、SCGB2A2、LALBA、CASP14、KLK5、WFDC2和VTCN1(参见表3)。没有发现结直肠癌的DCB基因。

[0350] 多个DCB基因表达出几个不同的特征。首先,针对多个组织特异性基因富集了多个DCB基因(图13)。在所述57,783个注解的基因中,0.3%是肺特异性的,0.2%是乳房特异性的。相比之下,50%的所述多个肺DCB基因是肺特异性的,44%的所述多个乳房DCB基因是乳

房特异性的(由所述蛋白质图谱数据库定义(Uhlén等人,Science,2015))。

[0351] 此外,一些DCB基因是仅在某些癌症亚型中检测到的多个亚型特异性生物标记物(图14A和14B)。FABP7仅在多个三阴性乳癌(TNBC)样本中检测到。相反地,在TNBC中未检测到SCGB2A2,但在HER2+和HR+/HER-乳癌样本中检测到。在多个肺腺癌患者样本的cfRNA中检测到SLC34A2、ROS1、SFTPA2和CXCL17基因,但在多个鳞状细胞癌患者样本中未检测到。与源自相同器官的其他多个癌症亚型相比,这些亚型特异性基因在肿瘤组织中的表达也更高。

[0352] 为了确定所述血液中多个肿瘤相关转录物的所述来源,评估了多个暗通道生物标记物基因的cfRNA和肿瘤组织RNA之间的一致性。观察到cfRNA和肿瘤组织表达之间的高度一致性(图15A)。在所述肿瘤组织中未检测到的多个基因不太可能在所述匹配的cfRNA样本中检测到,而在所述肿瘤组织中检测到的多个基因更有可能在所述匹配的cfRNA样本中检测到。另外地,作为一给定患者cfDNA肿瘤分数与匹配肿瘤组织中所述基因表达的所述乘积测量的肿瘤含量是多位乳癌患者所述cfRNA中的一DCB基因的所述可检测性的一强预测因子(图15B)。

[0353] 多个暗通道生物标记物(DCB),在多位非癌症受试者的cfRNA中未发现的多个转录物,在多位癌症患者中表达出高信噪比的所述潜力。DCB信号与肿瘤含量相关(作为所述血液中肿瘤分数和所述组织中RNA表达的所述乘积测量的)。cfRNA的多个DCB在多位癌症参与者中以一组织特异性和亚型特异性的方式被识别出。我们观察了高肿瘤组织表达导致DCB信号放大并能够检测多位低cfDNA肿瘤分数患者的癌症的多个病例。综上所述,这些数据显示多个组织特异性转录物具有用于基于血液的多癌症检测的潜力。示例2:识别多个异质性样本中的多个生物标记物

[0354] 我们使用标准差异表达(differential expression,DE)分析在多个异质性样本的生物标记物发现中观察到多个假阳性的两种常见来源。首先,由于对照组和癌症组中的遗传异质性或基因扩增缺失,所述基因表达遵循双峰分布(bimodal distribution)。其次,一单个有影响力的离群值(outlier)夸大了所述广义线性模型(generalized linear model,GLM)的所述斜率和p值。

[0355] 开发了一种方法来识别多个高度异质性样本中多个差异表达的基因,例如基于组织表达的cfRNA,称为heteroDE。所述heteroDE模型使用一负二项式广义线性模型(negative binomial generalized linear model,NB-GLM)。为了减少所述多个假阳性,heteroDE包括两个额外的功能:(1)检查所述非癌症组中的所述基因表达是否由于遗传异质性或基因扩增缺失而遵循双峰分布;(2)检查是否只有一个离群值样本影响所述NB-GLM的所述p值。使用库克距离(Cook's distance)识别所述离群值样本。所述NB-GLM是在没有具有最大库克距离的所述样本的情况下一第二次执行的。

[0356] 与多种现有差异表达(DE)方法相反,heteroDE使用所述肿瘤含量作为所述NB-GLM中的一共变量。所述多个非癌症样本的所述肿瘤含量设为零。一cfRNA肿瘤生物标记物基因的所述假设是,所述基因在所述组织中的表达越高,所述cfDNA中的所述肿瘤分数越大,就越有可能在cfRNA中检测到所述基因。当我们将这种方法应用于多个乳癌样本时,我们识别了9个cfRNA生物标记物:TRGV10、SCGB2A2、CASP14、FABP7、CRABP2、VGLL1、SERPINB5、TFF1和AC007563.5(参见表4)。这些生物标记物中的三个(FABP7、SCGB2A2、CASP14)与识别为多个DCB基因的所述多个基因重叠。

[0357] 图19中显示了说明根据heteroDE的所述样本处理和参数确定的一示例工作流程。由于缺乏组织样本,多位非癌症受试者的肿瘤含量被限制为零。所述工作流程的一示例实现如下:

[0358] $K_{i,j}$:患者j的所述cfRNA中基因i的多个读数计数;

[0359] $\mu_{i,j}$:患者j的所述cfRNA中基因i的多个平均读数计数;

[0360] α_i :基因i的离散度;

[0361] γ_i :基因i血浆中无肿瘤含量时的所述平均读数计数;

[0362] $x_{i,j}$:多个肿瘤含量, $\log_{10}$ (匹配的cfDNA中的肿瘤分数*匹配的肿瘤组织中的基因表达)

[0363] β_i :多个肿瘤含量的所述系数;

[0364] $K_{i,j} \sim \text{NB}(\mu_{i,j}, \alpha_i)$

[0365] $\log(\mu_{i,j}) = (\gamma_i + x_{i,j}\beta_i)$

[0366] 还测试了使用一信息增益方法的特征选择。信息增益是一种在所述二值化(binanzied)的cfRNA基因表达与所述癌症/非癌症标签之间选择具有高互信息的多个基因的方法。所述基因表达RPM矩阵被转换为一二元矩阵。如果所述基因的一RPM>0,则将其转换为1。如果基因的一RPM=0,则将其设为0。给定所述癌症类型(例如肺癌)和使用所述二进制表达值的非癌症标签计算每个基因的所述信息增益。所述乳腺癌组的所述非癌症组与性别平衡一仅选择所述非癌症组中的所述多位女性受试者。选择信息增益最高的前100个基因作为建模的所述特征。每个基因的所述值在所述建模过程中被转换为二进制值。对乳腺癌与非癌症以及结直肠癌与非癌症重复这些程序。肺癌信息增益最高的前30个基因如表5所示,乳腺癌信息增益最高的前30个基因如表6所示。

[0367] 在另一实施例中,从多个癌症组织样本中进行特征选择,以识别在多个癌症组织样本中表达但在多位非癌症参与者中不表达的多个基因。如上文示例1中所述制备多个文库并测序。对于每个癌症组织样本,我们从多个暗通道中识别出在癌症组织(组织RPM>10)中多个表达水平相对较高的多个基因。这些基因被归类为“多个组织亮通道基因(tissue bright channel gene)”。识别出的前15个组织亮通道基因如表7所示。

[0368] 示例3:在一单独的队列中验证DCB

[0369] 我们着手验证在一从商业供应商(Discovery Life Sciences)获得的乳腺癌(38)和肺癌(18)癌症样本的一正交集(orthogonal set)中在我们的CCGA队列中识别的所述多个DCB。选择第I-IV期患者以评估疾病进展过程中多个DCB的所述患病率,并纳入38个年龄匹配的非癌症样本作为多位非癌症患者DCB表达的多个对照。为了提高灵敏度并降低多个测序要求,我们开发了一种标靶富集方法来选择在我们的CCGA队列中识别的23个DCB。我们还富集了通常存在于非癌症血浆中的33个阳性对照基因。这些转录物在所述富集步骤中充当载体材料,因为大多数非癌症样本不包含DCB转录物。对所述多个产生的标靶RNA-seq文库进行测序和子采样,一深度为每个样本100M双端读数,并对标靶基因和脱靶基因的多个严格RNA读数的所述数量定量。与所述全转录组测定相比,我们发现所述标靶方法将多个标靶cfRNA转录物的转化效率提高了2至3倍。

[0370] 在我们的CCGA队列中识别的所述23个DCB中,除一个(CRABP2)外,所有DCB在所述非癌症组中均具有0的一中位表达(以RPM计)。我们套组中的19个DCB在所述验证队列的至

少1个癌症样本 (≥ 2 个独特片段) 中表达, 其中这些DCB中的16个在至少一个癌症类型中与多个非癌症样本相比有差异表达。随着测定效率和阶段的提高, 我们注意到一些组织特异性标记物存在于乳癌和肺癌中, 尽管它们在所述两组之间仍然存在差异表达。还有一些DCB仅在一个癌症类型中表达, 如乳癌中的SCGB2A2和肺癌中的ROS1、SFTA3和SFTPA2。对于在此验证队列中观察到的所述多个DCB的所有, 多个癌症样本中DCB表达的所述水平随着阶段的增加而增加, 我们队列中第IV期的多个样本的表达最高, 支持这些特征作为癌症的多个组织特异性标记物的所述有效性。尽管有这种趋势, 我们还在我们的队列中观察到多个早期癌症中的DCB表达, 这显示有机会使用富含多个DCB的一方法检测多个早期癌症。多个说明性的结果显示在图16A至图16D中, 所述y轴为多个读数计数的所述数量。

[0371] 示例4: 多个分类结果

[0372] 我们应用留一者 (leave-one-out, L00) 交叉验证和5倍 (5-fold) 交叉验证分类, 使用不同的特征选择方法, 包括多个暗通道生物标记物 (DCB)、heteroDE和信息增益 (information gain, IG)。说明性的工作流程显示于图17A至17B中。因为heteroDE使用匹配的肿瘤组织, 由于多个肺组织样本数量有限, 这种特征选择方法不适用于肺癌/非癌症分类。总体而言, 与乳癌/非癌症分类中的5倍交叉验证相比, L00在L00中的分类表达明显更好, 这意味着由于每个训练集中的样本量较小, 所述乳癌分类器在5倍分类中训练不足。DCB对肺癌/非癌症分类器具有最佳表达 (98%特异性下的灵敏度: 0.2 ± 0.037), 而heteroDE对乳癌/非癌症分类器具有最佳表达 (98%特异性下的灵敏度: 0.303 ± 0.046) (表10)。

[0373] 表10:

[0374]	癌症类型	特征选择	交叉验证	95% 特异性下的灵敏度
	肺癌	DCB	L00	0.3 ± 0.042
	肺癌	IG	L00	0.333 ± 0.043
	乳癌	heteroDE	L00	0.394 ± 0.049
	乳癌	DCB	L00	0.212 ± 0.041
	乳癌	IG	L00	0.303 ± 0.046
	肺癌	DCB	5 倍	0.261 ± 0.146
	乳癌	heteroDE	5 倍	0.177 ± 0.142

[0375] 多个说明性的结果也绘制于图18A至图18C中, 其是使用留一者交叉验证生成的。图18A显示了使用所述heteroDE特征选择方法和一随机森林分类器的乳癌与非癌症的留一者 (L00) 交叉验证分类的一接受者操作特征 (receiver operating characteristic, ROC) 图和一变量重要性图。所述输入数据是每个基因的多个计数, 使用所述DESeq2 R包中的大小因子规一化 (使用所述estimateSizeFactors) 函数对其进行规一化。如表10所示, 95%下

的所述灵敏度为0.394+/-0.049。

[0376] 图18B显示了使用所述暗通道特征选择方法和一随机森林分类器的肺癌与非癌症标签的留一者 (LOO) 交叉验证分类的一ROC图。所述输入数据是每个基因的多个规一化计数,以每百万个读数 (rpm) 为单位。如表10所示,95%特异性下的所述灵敏度为0.3+/-0.042。

[0377] 图18C显示了来自使用所述暗通道特征选择方法和一随机森林分类器的乳癌与非癌症标签的留一者 (LOO) 交叉验证分类的一ROC图和变量重要性图。所述输入数据是每个基因的多个规一化计数,以每百万个读数 (rpm) 为单位。如表10所示,95%特异性下的所述灵敏度为0.212+/-0.041。

[0378] 示例5:多种材料和多种方法

[0379] 测序数据处理

[0380] 使用STAR版本2.5.3a将多个原始读数与所有转录物的gencode v19主要组装比对。根据基因组比对位置和多个非随机UMI序列检测并去除多个重复序列读数。大多数双端读数的多个UMI序列与多个预期序列正好匹配。多个读数的一子集包含所述UMI序列中的多个错误,并且应用了一探索式错误更正。如果所述UMI与一预期UMI的一汉明距离 (hamming distance) 在1以内,则将其分配给所述UMI序列。在汉明距离超过1的情况下,或者多个已知序列于一汉明距离在1以内的情况下,丢弃具有所述UMI错误的所述读数。通过多个成员读数的多序列比对,共享比对位置和多个校正UMI的多个读数的多个集被错误更正,并生成一单个共识序列/比对。将多个读数比对与gencode v19中的多个注解转录物进行比较。只有跨越带多个注解的外显子-外显子连接的多个读数才被计算在内,以去除因多个DNA污染读数而导致的多个错误计数。

[0381] 样本收集

[0382] 全血收集在多个Streck血浆游离DNA BCT管 (Streck Cell-free DNA BCT tube) 中,在血浆分离之前将其运输并储存在环境温度下。全血于4℃下在一旋翼式转子 (swing-bucket rotor) 中以1600g离心10分钟以分离血浆。将血浆层转移到一单独的管中,并于4℃下以15000g离心12分钟,以进一步去除多个细胞污染物。双旋转血浆储存在-80℃并在提取前在室温下解冻以避免形成多个冷冻沉淀物。

[0383] 样品选择标准

[0384] 我们从循环血浆游离基因组图谱研究 (CCGA, NCT02889978) 中选择了第III期乳癌、肺癌和结直肠癌样本的一子集。我们要求所述多位选择的患者至少有两管未处理的1-2级血浆 (无溶血),每位患者6-8mL血浆。我们进一步要求多位选择的患者具有与多个先前研究中的cfDNA测序数据相匹配的数据。一旦选择了所述多位癌症患者,我们就选择了与所述多个癌症样本在年龄、性别和种族方面匹配的一相同数量的多个非癌症样本。基于此标准,我们选择了210个样本。使用R中的一随机化函数将这些样本随机分为14个批次,以确保每批次中多个癌症类型 (癌症和非癌症样本) 的一随机混合。

[0385] 样本处理

[0386] 使用来自所述QIAamp循环核酸试剂盒 (QIAamp Circulating Nucleic Acids kit; Qiagen, 55114) 的所述循环miRNA方案,从多达8mL的冷冻血浆中提取多个血浆游离核酸。根据所述制造商的说明,使用不含RNase的DNase套组 (RNase-free DNase Set; Qiagen,

79254) 对所述提取的材料进行DNase处理,并使用所述高灵敏度RNA片段分析仪试剂盒(High Sensitivity RNA Fragment Analyzer kit;Agilent,DNF-472)进行定量。使用所述TruSeq RNA外显体试剂盒(TruSeq RNA Exome kit;Illumina,20020189)进行反转录和转接子连接。使用用于人类rRNA和粒线体的AnyDeplete试剂盒(AnyDeplete for Human rRNA and Mitochondrial Kit;Tecan,9132)补充了一组定制的多个消耗目标,从而消耗了所述多个产生的文库中的丰富序列。

[0387] 多个测序样本被筛选并且表达出低质量控制指标的那些被排除在后续分析之外。一个测定指标和三个管道指标被选为“红旗”,用于排除指标不佳的多个样本。所述测定指标测量多个样本是否有足够的材料用于测序,所述管道指标是测序深度、RNA纯度和交叉样本污染。

[0388] 基因表达量化

[0389] 尽管在文库制备期间有所述DNase消化步骤,但对所述数据的初步检查揭示了多个cfRNA样本中多个不同水平的残留DNA。污染的所述程度极低(每个样本<6个单倍体基因组当量),并且与消化前的cfDNA的所述量或多个批次特定问题无关。相反,它似乎是随机的,与多个先前的报告一致。

[0390] 一QC度规(QC metric),“分位数95链特异性(quantile 95strand specificity)”,定义为处于或低于表达的第95分位数的多个基因的所述链特异性,用于评估每个样本中DNA污染的所述水平。多个UHR阳性对照样本表达出高分位数95链特异性(>0.85)。多个cfRNA分位数95链特异性值分布在一很宽的范围内(0.52-0.89)。作为参考,多个cfDNA样本的一分位数95链特异性约为0.5,这表示一些cfRNA样本主要受来自残留DNA的信号支配。所述多个读数链颜色显示NC67中多个有义(sense)和反义(anti-sense)读数的分布均匀,而NC3中只有多个有义读数。另外地,NC67中的多个内含子和多个外显子都有丰富的覆盖范围,正如DNA存在所预期的。DNA污染程度高的多个样本中片段长度的所述分布显示它们模仿cfDNA的所述长度分布(中位数~160),强烈表示未消化的cfDNA是主要污染物。

[0391] 具有低于0.84的分位数95链特异性的多个样本被标记并从随后的分析中移除。为了进一步防止DNA污染导致的多个RNA计数的所述膨胀,此处显示的所述多个基因计数是使用多个严格计数生成的,定义为多个读数对,其中所述两个读数中的至少一个映射到一外显子-外显子连接处。使用多个不同水平的cfDNA掺入一cfRNA样本进行的一实验显示,使用多个严格计数对多个RNA水平的所述估计保持不变,支持在所述多个先期研究样本中使用多个严格计数来定量和比较基因表达。

[0392] 多个暗信道特征的选择

[0393] 所述多个暗通道基因通过以下标准识别:1)此基因在所述非癌症组中的所述中位表达(以RPM计)为0,并且此基因的所述标准偏差小于0.1RPM。使用以下标准识别每个癌症类型的所述多个暗通道生物标记物(DCB):1)在所述特定癌症组中至少有两个样本表达了所述基因、2)所述第二高表达样本的所述RPM大于0.1、以及3)与所述非癌症组相比,所述基因在所述特定癌症组中差异表达(肺癌的p值<2e-02及乳癌的p值<2e-01)。两组差异表达的所述p值由所述edgeR包计算。肺癌组与非癌症组之间FDR<0.05的基因有816个。乳癌组与非癌症组之间FDR<0.05的基因有28个。结直肠癌组与非癌症组之间FDR<0.05的基因有4个。对

于所述箱形图和热图,我们仅显示了最显著的多个差异表达基因(肺癌和乳癌的 $FDR < 2e-06$,结直肠癌的 $FDR < 2e-02$)。

[0394] 如下进行多个组织特异性基因的注解。肺癌、乳癌和结肠癌的所述组织特异性基因文件可从人类蛋白质图谱网站(www.proteinatlas.org/)下载。多个组织特异性基因分为三类:1)组织富集的:与所有其他组织相比,一特定组织中的mRNA水平至少高4倍、2)组富集的:一组2-5个组织中的mRNA水平至少高4倍、3)组织增强的:与所有组织中的平均水平相比,一特定组织中的mRNA水平至少高出4倍。所有这三个类别都包含在我们对多个组织特异性基因的定义中。

[0395] 为了测试所述多个组织特异性基因的富集。1)Fisher精确检定(Fisher's exact test)用于检验所有所述注解人类基因的多个肺DCB与多个肺特异性基因之间的所述独立性。2)Fisher精确检定用于检验所有所述注解人类基因的多个乳房DCB与多个乳房特异性基因之间的所述独立性。

[0396] 示例6:多个cfRNA癌症生物标记物的群组

[0397] 设计了一项研究,以从不同于一正常非癌症队列的所述全转录组中识别多个肺癌和乳癌特异性cfRNA生物标记物,并从多个癌症样本中识别cfRNA中特定代表的多个生物信号,这些信号可能对癌症二元检测和从血浆中识别起源组织(TOO)有用。我们的工作重点是识别与多个早期阶段难以检测的多个癌症亚型相关的多个基因特征,即肺腺癌和HR+及三阴性(TNBC)乳癌。

[0398] 用于执行此分析的数据包括1)从CCGA和一商业供应商测序的全转录组血浆数据、2)来自TCGA的全转录组组织数据、以及3)来自所述人类蛋白质图谱的多个基因注解(Uhlén等人,Science 2015)。从所述循环血浆游离基因组图谱研究(CCGA,NCT02889978)中选择并测序了第III期乳癌和肺癌样本的一子集。选择多个第III期样本以最大化所述血液中的信号,同时避免来自潜在的多个继发性转移的混淆信号。我们总共分析了来自CCGA的47个乳癌、14个肺腺癌和93个非癌症的血浆样本。另外地,我们包括了另一组来自一商业供应商(Conversant)的多个全转录组样本。这包括一组14个第IV期乳癌血浆样本,用于捕获所述血液中多个生物标记物的多个晚期信号。这些源自血浆的数据用于定义哪些基因在健康血浆中表达,哪些基因在癌症血浆中差异表达,这可能对这些亚型中癌症的二元检测有价值。我们将每个样本的所述基因表达编译成一RPM(每百万个读数)规一化基因特征矩阵,其中每个样本是一行,每一列是一基因特征。

[0399] 此研究还包括来自所述TCGA联盟的乳癌(BRCA)和肺腺癌(LUAD)组织全转录组数据,从所述GDC入口网站下载。总共包括第I-IV期的533个肺腺癌和1102个乳癌样本。这些数据用于识别用于二元检测的多个高表达源自肿瘤的基因特征。另外地,这种高维数据用于识别可用于TOO的多个组织特异性基因特征。我们将每个样本的所述基因表达编译成一RPM(每百万个读数)规一化基因特征矩阵,其中每个样本是一行,每一列是一基因特征。

[0400] 最后,我们查询了在所述人类蛋白质图谱中的所有所述基因特征,人类蛋白质图谱是来自多个癌症肿瘤样本和健康组织的各种组学(omics)技术(转录组学(transcriptomic)和基于抗体的)的一开放式汇编,并提供组织腔隔和疾病注解。我们使用这些注解来根据诊断时肿瘤中的多个表达水平和多位患者的总体存活率,捕获所述基因是否是癌症类型富集的/增强的,以及对疾病预后有利/不利。

[0401] 为了建立一组用于二元检测和T00分类的多个目标,我们首先评估了我们是否可能使用从所述GDC数据入口网站下载的TCGA组织表达数据来选择可能的多个生物标记物。对于每个基因,我们计算了两个队列中多个样本的所述平均基因表达,并计算了所述队列中的所述皮尔逊相关分析(Pearson correlation)。一般而言,我们发现TCGA组织中的高平均基因表达与CCGA血浆中的高平均基因表达大致相关(多个乳癌的斯皮尔曼次序相关系数(Spearman's rho)为0.568,多个肺癌为0.509)。因此,我们推断TCGA组织数据可以为特征选择提供信息。我们优先考虑平均TCGA组织表达大于1RPM的多个基因特征,因为它们可能在源自癌症的血浆中检测到,并且可能为二元癌症检测或起源组织检测提供信息。在过滤这些可能常见的多个人工制品诱导转录物(映射到HLA、IGH、IGL和核糖体基因的多个转录物)之后,这导致了2898个潜在的基因特征。

[0402] 然而,即使这些基因特征在所述TCGA组织中高度表达,也不确定这些基因特征在所述血浆中的表达有多普遍。图22(乳癌)和图23(肺癌)中显示了与血浆相比组织中平均RPM的散点图。图21提供了在多个癌组织样本中以多个高水平表达的多个基因的多个示例结果,在血浆中几乎没有可检测到的转录物。还利用从CCGA的所述血浆中的表达获得的信息进行了基因特征选择。我们将所述多个CCGA血浆样本中检测到或未检测到的多个基因表达特征二值化,检测到的表达等于或高于每百万个中0.005个读数(RPM)。然后,我们根据从所有癌症血浆到所有非癌症血浆的多个观察结果计算了每个基因的所述血浆对数可能性比率(LOR)。这定量了一基因在一癌症样本中出现的所述可能性超过所述基因在一非癌症样本中出现的所述可能性。一 $LOR > 0$ 表示与多个非癌症病例相比,在多个癌症病例中检测到一基因的可能性更大,而 $LOR < 0$ 表示与多个癌症病例相比,在多个非癌症病例中检测到一基因的一可能性更大。我们选择了所述血浆中信息量最大的多个基因,一 $LOR > 0.1$,产生了281个基因特征。多个cfRNA生物标记物的LOR的一示例图如图24所示。

[0403] 此外,我们着手评估哪些基因特征特别适用于T00分类。由于cfRNA的所述CCGA数据集限于<200个样本,我们决定使用所述TCGA肿瘤基因矩阵并执行一递归特征消除算法(recursive feature elimination algorithm)来识别对于区分肺腺癌、乳癌HR+和乳癌TNBC很重要的多个基因特征。一随机森林多类模型用于递归地选择前K个基因,对所有基因特征进行10倍交叉验证。通过优化跨多个折叠的准确度,跨多个迭代消除了多个特征。当使用750个基因特征时,所述交叉验证模型以96.7%的准确度对所述多个TCGA样本进行分类,因此我们将这些前750个生物标记物识别为对所述组织中的亚型分类很重要。

[0404] 所述人类蛋白质图谱汇集了来自多个癌症肿瘤样本以及多个健康组织样本的TCGA转录组学和基于抗体的蛋白质数据,以提供我们用来为二元检测和T00确定多个基因特征优先级的两个特定图谱。所述组织图谱包括基于正常组织中mRNA和蛋白质水平的组织富集的(与其他多个组织相比在组织中升高)和组织增强的(在低特异性的组织中表达)的多个基因的多个注解。另外地,所述病理学图谱包括癌症类型富集的(与其他多个肿瘤相比在肿瘤类型中升高)或增强的(在低特异性的肿瘤类型中表达)的多个基因的多个注解,以及根据诊断时肿瘤中的多个表达水平和多位患者的总体存活率,对疾病预后有利/不利。我们将多个基因标记为具有乳癌和肺癌的这些注解的多个潜在生物标记物(3028个基因特征)。

[0405] 在所述血浆中发现的大部分转录物被认为衍生自多个健康的免疫细胞。为了选择

多个健康白血球细胞中不存在的多个生物标记物,这会混淆癌症检测,我们过滤了多个基因特征,使其在来自所述CCGA队列的多个健康个体的血浆中具有低表达(中位数RPM<1,标准偏差RPM<0.1)。这些产生的41391个基因特征被称为“多个暗通道”。我们通过整合上述识别二元癌症检测和多个T00生物标记物的多种方法进一步过滤了这些暗通道。所述多个暗通道被过滤,以便所述基因二值化LOR>0.1用于多个癌症相关基因特征,或者所述基因包含在所述随机森林模型选择的所述750个基因中。这些基因被进一步过滤,以便它们要么被所述人类蛋白质图谱注解,要么在一TCGA队列中的所述平均表达大于5RPM。来自示例1-4的多个额外阳性对照和DCB基因被添加到这个更新的生物标记物组,使多个cfRNA生物标记物的所述总数达到467,列于表15(其中的一个子集在表11中提供)。表14的所述多个基因代表了特别有用的多个cfRNA生物标记物的一子集。乳癌和肺癌的多个选择的生物标记物的多个示例结果分别在图10A和20B中示出。示例7:多个癌症样本中多个cfRNA生物标记物的评估

[0406] 测试了表15中列出的所述467个cfRNA生物标记物在难以检测的具有低肿瘤分数的乳癌和肺癌中识别癌症和区分多个非癌症的所述能力。根据在所述样本中的任何基因中观察到的最高证据对所有样本进行评分。我们选择了所有在多个高信号癌症中具有一些信号证据的基因。对于每个样本,我们识别了在所述样本中比在所有其他非癌症中具有更多证据的所有基因,并使用以下标准按每个样本中所述证据最多的基因对多个样本进行排序:(1)在任何非癌症中观察到的多个最大计数(越低越好)、(2)在任何高信号癌症中观察到的多个最大计数(越高越好)、以及(3)在所述样本中观察到的多个计数。一留一者分类器在训练和留出充足的集合中使用这些生物标记物进行评估。结果如图7所示。如所述星号所示,所述验证队列特异性相对于所述训练队列有一显著降低($p=.0.02$)。不希望受理论的束缚,这可能表示此特定实验中的潜在过度拟合。

[0407] 基于所述多个cfRNA生物标记物的所述留一者分类器应用于具有低或高信号的一DNA甲基化癌症生物标记物的多个癌症样本。多个样本包括多个肺癌和乳癌样本。如图8A至8C所示,所述分类器展示了高特异性表达。

[0408] 数个基因被证明是特别有用的多个cfRNA癌症生物标记物,一些基因对乳癌或肺癌具有特异性,一些基因在乳癌和肺癌中都升高。这33个基因列于上表8中。在图26A至26D中以图形方式呈现多个严格读数计数的所述多个结果。关于这33个基因的多个结果的多个额外细节在下表16中提供。

[0409] 表16:

[0410]

基因符号	最大高信号癌症计数	最大非癌症计数	检测到的乳癌数量* (n=206)	检测到的肺癌数量* (n=81)
CEACAM5	1125	3	4	8
RHOV	725	5	2	6
SFTA2	589	12	0	7
SCGB1D2	381	6	5	0
IGF2BP1	335	4	2	3
SFTPA1	305	6	1	5
CA12	226	7	3	4

[0411]	SFTPB	197	8	1	11
	CDH3	195	18	0	7
	MUC6	146	1	3	2
	SLC6A14	132	3	2	6
	HOXC9	106	2	2	2
	AGR3	101	6	3	5
	TMEM125	84	6	2	8
	TFAP2B	65	1	6	1
	IRX2	41	1	5	7
	POTEKP	38	1	2	1
	ARHGEF38	36	3	3	7
	GPR87	25	1	0	6
	LMX1B	24	2	6	0
	ATP10B	24	2	1	4
	NELL1	22	2	3	3
	MUC21	20	1	0	4
	SOX9	17	4	5	6
	LINC00993	17	1	3	0
	STMND1	14	1	3	1
	ERVH48-1	12	1	2	1
	SCTR	12	2	0	6
	MAGEA3	10	0	0	3
	MB	8	1	5	2
	LEMD1	8	2	3	4
	SIX4	8	2	1	2
	NXNL2	7	2	2	4

[0412] *如果严格的RNA计数高于所述最大非癌症计数或2,以较高者为准,则称检测到的多个基因。

[0413] 参考文献

[0414] Klein et al.开发用于早期检测多个肿瘤类型的一全面血浆游离DNA(cfDNA)测定:所述循环血浆游离基因组图谱(CCGA)研究(Development of a comprehensive cell-free DNA(cfDNA) assay for early detection of multiple tumor types:The Circulating Cell-free Genome Atlas(CCGA) study).ASCO(2018).

[0415] Uhlén et al.基于组织的所述人类蛋白质组图谱(Tissue-based map of the human proteome)(www.proteinatlas.org).Science doi:10.1126/science.1260419(2015).

[0416] A.M.Newman,et al.,一种超灵敏的定量循环肿瘤DNA的方法,具有广泛的患者覆盖率(An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage).Nat.Med.20,548-554(2014).

[0417] E.Kirkizlar,et al.,使用一大规模多重PCR方法检测乳腺癌患者血浆游离DNA中的克隆和亚克隆拷贝数变异(Detection of Clonal and Subclonal Copy-Number Variants in Cell-Free DNA from Patients with Breast Cancer Using a Massively Multiplexed PCR Methodology).Transl.Oncol.8,407-416(2015).

[0418] S.Y.Shen,et al.,使用多个血浆游离DNA甲基化组进行灵敏的肿瘤检测和分类(Sensitive tumour detection and classification using plasma cell-free DNA

methyomes). *Nature* 563,579-583 (2018) .

[0419] C.Bettegowda, et al., 检测早期和晚期人类恶性肿瘤中的循环肿瘤DNA (Detection of circulating tumor DNA in early-and late-stage human malignancies). *Sci.Transl.Med.* 6,224ra24 (2014) .

[0420] K.C.A.Chan, et al., 通过血浆DNA亚硫酸氢盐测序癌症相关全基因组低甲基化和拷贝数畸变的无创检测 (Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing). *Proc.Natl.Acad.Sci.U.S.A.* 110,18761-18768 (2013) .

[0421] I.S.Haque, O.Elemento, 使用ctDNA实现癌症早期检测的多个挑战 (Challenges in Using ctDNA to Achieve Early Detection of Cancer). *bioRxiv*, 237578 (2017) .

[0422] K.C.A.Chan, et al., 血浆中的癌症基因组扫描: 通过大规模并行测序检测肿瘤相关的拷贝数畸变、单核苷酸变异和肿瘤异质性 (Cancer genome scanning in plasma: detection of tumor-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing). *Clin.Chem.* 59,211-224 (2013) .

[0423] C.Abbosh, et al., 系统发育ctDNA分析描绘了早期肺癌的演变 (Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution). *Nature* 545,446-451 (2017) .

[0424] K.-W.Lo, et al., 鼻咽癌患者血浆中游离爱泼斯坦-巴尔病毒相关RNA的分析 (Analysis of Cell-free Epstein-Barr Virus-associated RNA in the Plasma of Patients with Nasopharyngeal Carcinoma). *Clin.Chem.* 45,1292-1294 (1999) .

[0425] M.S.Kopreski, F.A.Benko, L.W.Kwak, C.D.Gocke, 恶性黑色素瘤患者血清中肿瘤信使RNA的检测 (Detection of tumor messenger RNA in the serum of patients with malignant melanoma). *Clin.Cancer Res.Off.J.Am.Assoc.Cancer Res.* 5,1961-1965 (1999) .

[0426] J.D.Arroyo, et al., Argonaute2复合物携带一群独立于人血浆中囊泡的循环小分子RNA (Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma). *Proc.Natl.Acad.Sci.U.S.A.* 108,5003-5008 (2011) .

[0427] P.M.Godoy, et al., 人类生物流体之间小RNA组成的巨大差异 (Large Differences in Small RNA Composition Between Human Biofluids). *Cell Rep.* 25, 1346-1358 (2018) .

[0428] M.F.de Souza, et al., 循环mRNA和miRNA作为前列腺癌诊断和预后的候选标记物 (Circulating mRNAs and miRNAs as candidate markers for the diagnosis and prognosis of prostate cancer). *PLoS ONE* 12 (2017) .

[0429] G.Y.F.Ho, et al., 根据结直肠肿瘤的严重程度循环microRNA的差异表达 (Differential expression of circulating microRNAs according to severity of colorectal neoplasia). *Transl.Res.* 166,225-232 (2015) .

[0430] I.Lee, D.Baxter, M.Y.Lee, K.Scherler, K.Wang, 标准化对分析循环RNA的重要性

(The importance of standardization on analyzing circulating RNA).Mol.Diagn.Ther.21,259-268(2017) .

[0431] X.Q.Chen,et al.,端粒酶RNA作为乳癌患者血清中的检测标记物(Telomerase RNA as a detection marker in the serum of breast cancer patients).Clin.Cancer Res.Off.J.Am.Assoc.Cancer Res.6,3823-3826(2000) .17.R.C.Kamm,A.G.Smith,人血浆中的核糖核酸酶活性(Ribonuclease activity in human plasma).Clin.Biochem.5,198-200(1972) .

[0432] T.El-Hefnawy,et al.,血浆中可扩增循环RNA的表征及其作为癌症诊断工具的潜力(Characterization of amplifiable,circulating RNA in plasma and its potential as a tool for cancer diagnostics).Clin.Chem.50,564-573(2004) .N.B.Y.Tsui,E.K.O.Ng,Y.M.D.Lo,血液样品、血清和血浆中内源性和添加的RNA的稳定性(Stability of endogenous and added RNA in blood specimens,serum,and plasma).Clin.Chem.48,1647-1653(2002) .

[0433] J.D.Arroyo,et al.,Argonaute2复合物携带一群独立于人血浆中囊泡的循环小分子RNA(Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma).Proc.Natl.Acad.Sci.U.S.A.108,5003-5008(2011) .

[0434] G.J.S.Talhouarne,J.G.Gall,脊椎动物红血球细胞中的7SL RNA(7SL RNA in vertebrate red blood cells).RNA 24,908-914(2018) .

[0435] L.A.Hancock,et al.,Muc5b过度表达导致黏膜纤毛功能障碍并增强小鼠肺纤维化(Muc5b overexpression causes mucociliary dysfunction and enhances lung fibrosis in mice).Nat.Comm.9,1-10(2018) .

[0436] T.Handa,et al.,Caspase14表达与乳癌患者的三阴性表型和癌症干细胞标记物表达相关(Caspase14 expression is associated with triple negative phenotypes and cancer stem cell marker expression in breast cancer patients).J.Surg.Oncol.116,706-715(2017) .

[0437] R.Hrstka,et al.,促转移蛋白前梯度2预测他莫昔芬治疗的乳癌预后不良(The pro-metastatic protein anterior gradient-2predicts poor prognosis in tamoxifen-treated breast cancers).Oncogene 29,4838-4847(2010) .

[0438] M.Pizzi,et al.,肺腺癌前梯度2过度表达(Anterior gradient 2overexpression in lung adenocarcinoma).Appl.Immunohistochem.Mol.Morphol.AIMM 20,31-36(2012) .

[0439] H.Cho,A.B.Mariotto,L.M.Schwartz,J.Luo,S.Woloshin,癌症生存率的变化何时意味着进步?来自人口发病率和死亡率的见解(When do changes in cancer survival mean progress?The insight from population incidence and mortality).J.Natl.Cancer Inst.Monogr.2014,187-197(2014) .

[0440] Y.M.Lo,et al.,从母体血浆中快速清除胎儿DNA(Rapid clearance of fetal DNA from maternal plasma).Am.J.Hum.Genet.64,218-224(1999) .

[0441] M.A.Watson,T.P.Fleming,乳腺球蛋白是子宫球蛋白基因家族的一乳腺特异性成

员,在人类乳癌中过度表达(Mammaglobin,a mammary-specific member of the uteroglobin gene family,is overexpressed in human breast cancer).Cancer Res.56,860-865(1996) .

[0442] G.H.Lewis,et al.,浸润性乳癌分子亚型与囊肿液蛋白15、乳腺球蛋白表达的关系(Relationship between molecular subtype of invasive breast carcinoma and expression of gross cystic disease fluid protein 15and mammaglobin) .Am.J.Clin.Pathol.135,587-591(2011) .

[0443] R.-Z.Liu,et al.,脂肪酸结合蛋白7/RXR β 路径增强三阴性乳癌的存活和增殖(A fatty acid-binding protein 7/RXR β pathway enhances survival and proliferation in triple-negative breast cancer) .J.Pathol.228,310-321(2012) .A.Cordero,et al.,FABP7是HER2+乳癌脑转移的关键代谢调节因子(FABP7is a key metabolic regulator in HER2+breast cancer brain metastasis) .Oncogene38,6445-6460(2019) .

[0444] H.Zhang,et al.,蛋白质FABP7和OATP2与人类乳癌的基础表型和患者结果相关(The proteins FABP7 and OATP2 are associated with the basal phenotype and patient outcome in human breast cancer) .Breast Cancer Res.Treat.121,41-51(2010) .

[0445] J.Xiao,et al.,区分肺腺癌和鳞状细胞癌的八个潜在生物标记物(Eight potential biomarkers for distinguishing between lung adenocarcinoma and squamous cell carcinoma) .Oncotarget 8,71759-71771(2017) .

[0446] M.Grageda,P.Silveyra,N.J.Thomas,S.L.DiAngelo,J.Floros,肺癌中表面活性蛋白A2基因的DNA甲基化谱和表达(DNA methylation profile and expression of surfactant protein A2 gene in lung cancer) .Exp.Lung Res.41,93-102(2015) .

[0447] Z.Zhang,et al.,SLC34A2的高表达是肺腺癌患者的有利预后标记物(High expression of SLC34A2 is a favorable prognostic marker in lung adenocarcinoma patients) .Tumour Biol.J.Int.Soc.Oncodevelopmental Biol.Med.39,1010428317720212(2017) .

[0448] F.Diehl,et al.,循环突变DNA以评估肿瘤动力学(Circulating mutant DNA to assess tumor dynamics) .Nat.Med.14,985-990(2008) .

[0449] Liu M.C.et al.,使用血浆游离DNA中的甲基化特征进行灵敏和特异性的多癌症检测和定位(Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA) .Ann Oncol.31(6) ,745-59(2020) .

[0450] 在整个本公开中已经对其他文件进行了参考和引用,例如专利、专利申请案、专利出版物、期刊、书籍、论文、网络内容。出于所有目的,所有此类文件均特此通过引用其整体并入本文中。

[0451] 除了本文所示和描述的那些之外,根据本文件的全部内容,本发明的各种修改及其许多进一步的实施例对于本领域技术人员而言将变得显而易见,包括对本文引用的科学和专利文献的参考。本文的主题包含重要的信息、示例和指导,它们可以适用于本发明的各种实施例及其均等物的实践。整个说明书中引用的所有参考文献均通过引用明确并入本文中。

[0452] 以上对多个实施例的详细描述参考了多个附图,附图说明了本公开的多个具体实施例。具有不同结构和操作的其他多个实施例不脱离本公开的所述范围。所述术语“本发明”等是参考本说明书中阐述的所述申请人发明的许多替代方面或实施例的某些特定示例而使用的,其使用或不使用均不旨在限制所述申请人的发明范围或所述权利要求的范围。为了方便读者,本说明书分为几个部分。多个标题不应被解释为限制本发明的所述范围。这些定义旨在作为本发明描述的一部分。应当理解,在不脱离本发明所述范围的情况下可以改变本发明的各种细节。此外,前述描述仅用于说明的目的,而不是用于限制的目的。

[0453] 虽然本发明已经参考其多个具体实施例进行了描述,但是本领域技术人员应该理解,在不脱离本发明的真实精神和范围的情况下,可以进行各种改变并且可以用多个均等物代替。此外,为了适应本发明的目的、精神和范围可以进行许多修改以适应一特定情况、材料、物质组成、过程、过程步骤或多个步骤。所有这些修改都旨在落入所附权利要求的范围内。

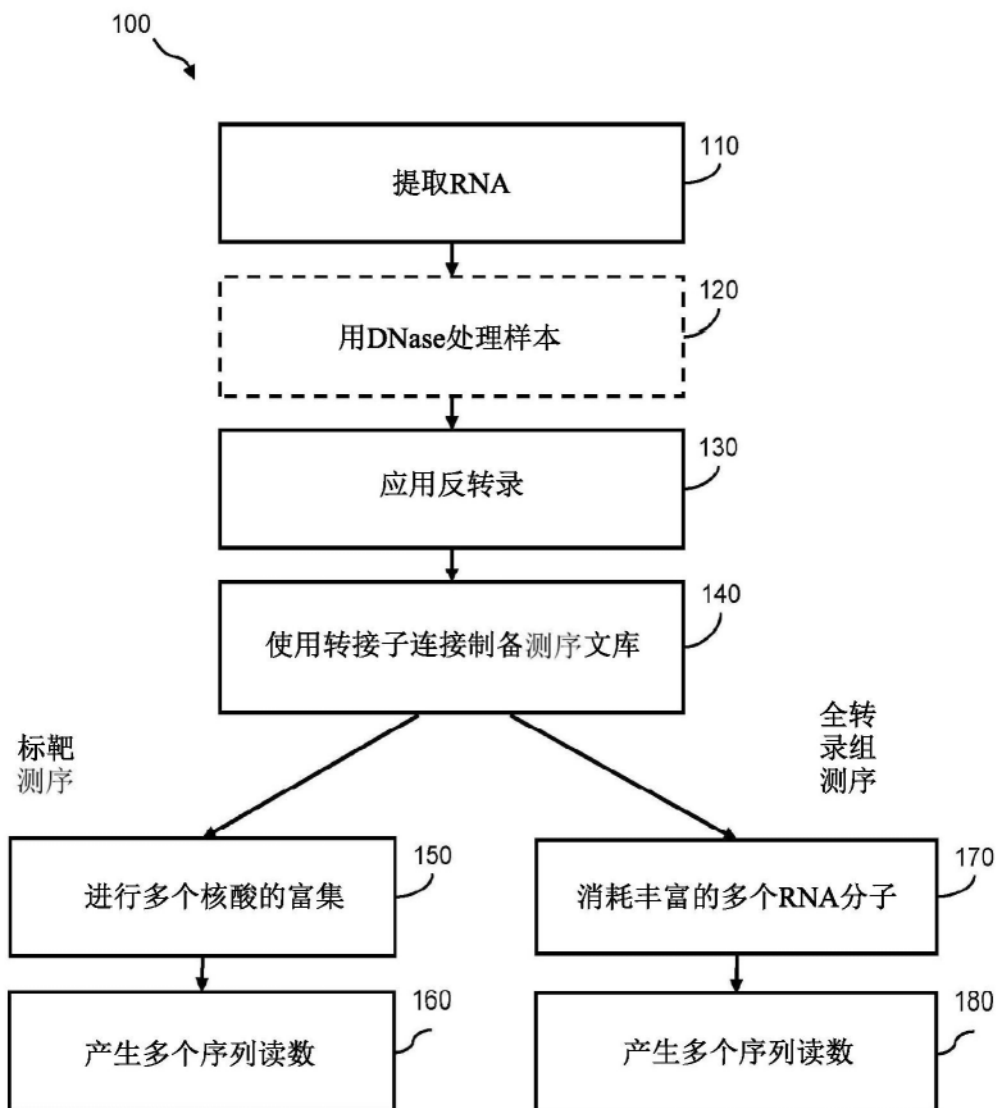


图1

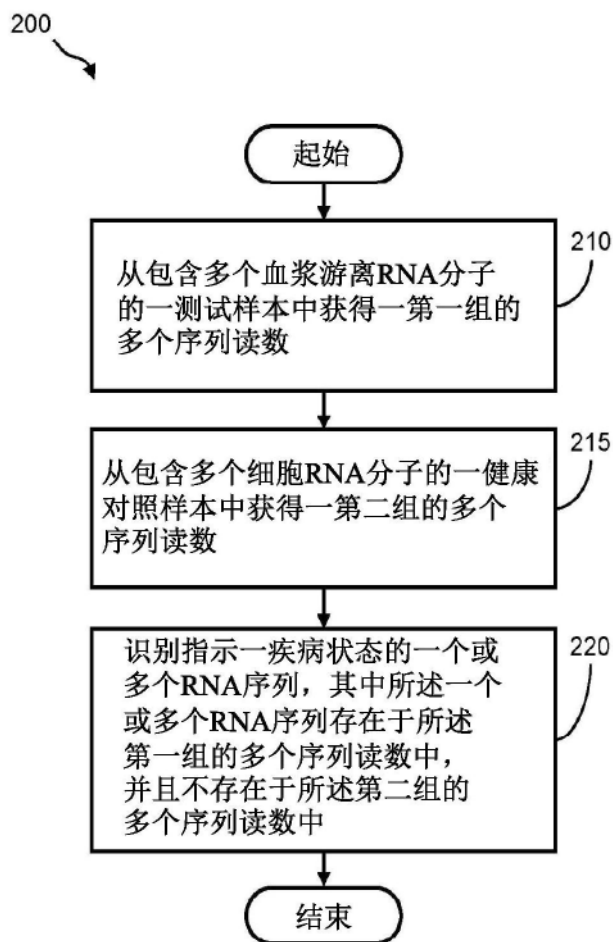


图2

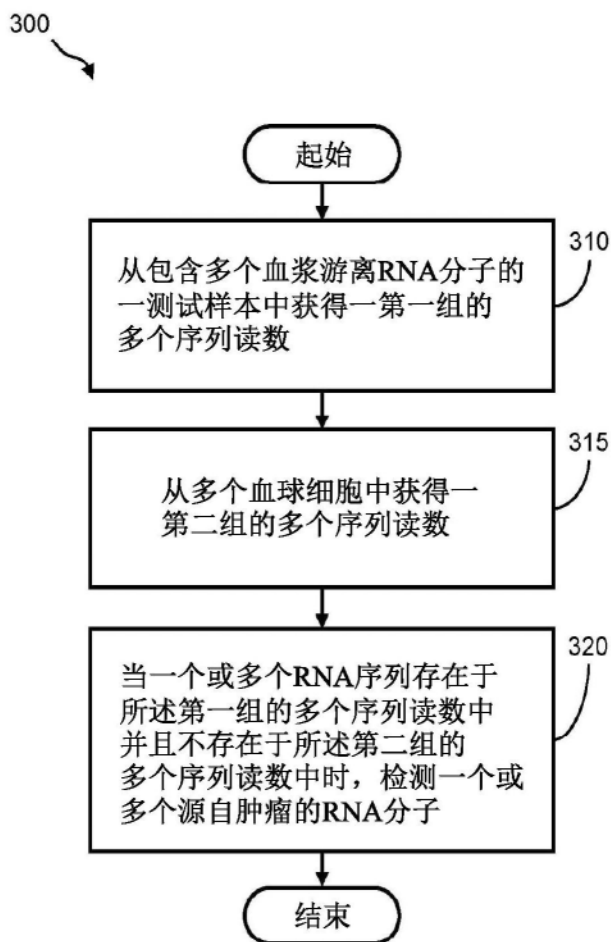


图3

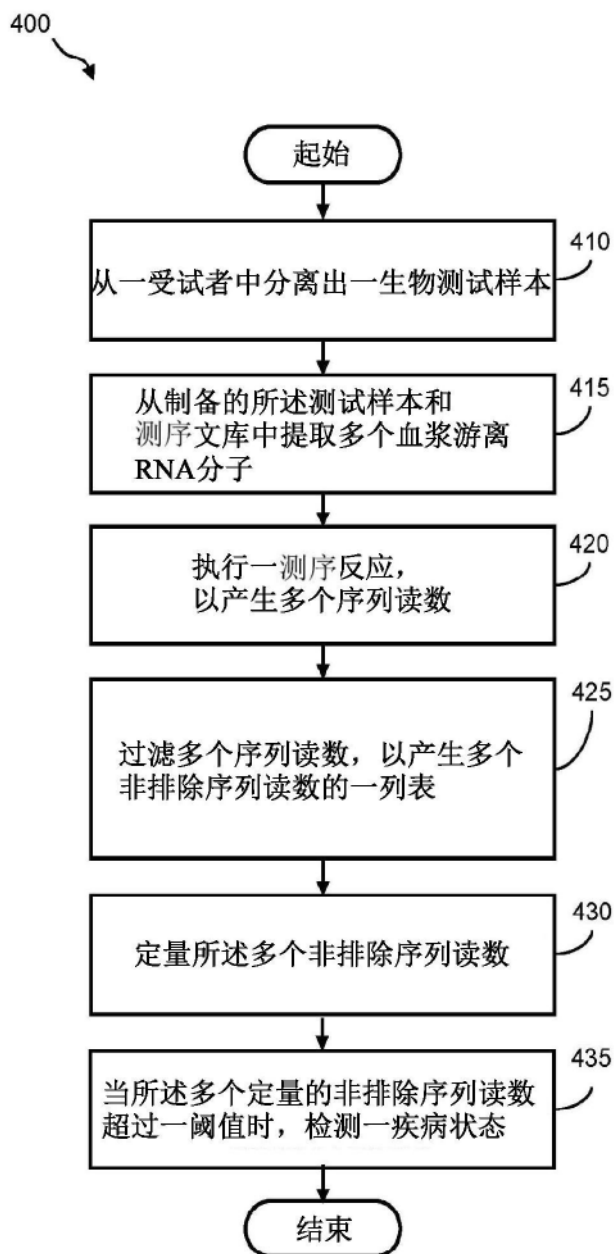


图4

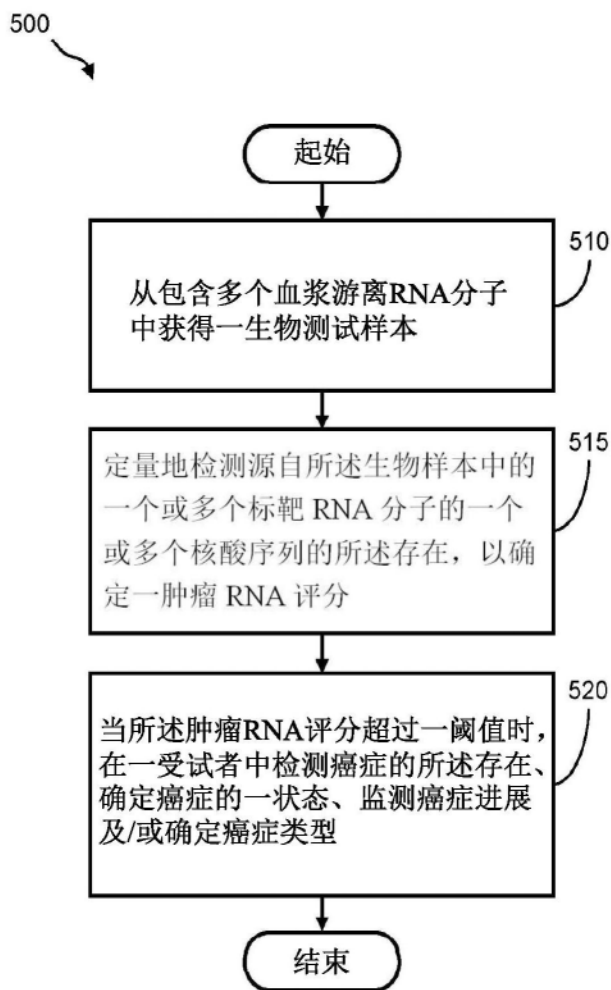


图5

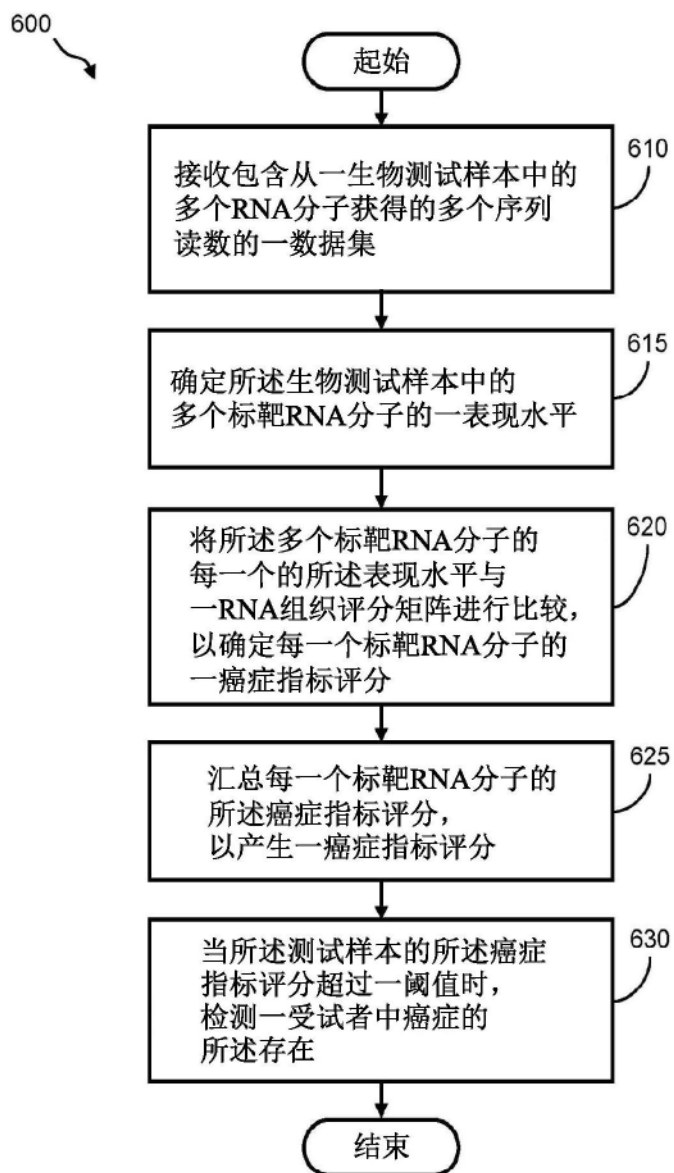


图6

训练 (119 乳房 , 32 肺 , 194 非癌症):

特异性	灵敏度, 肺	灵敏度, 乳房
.98 [.95 - .99]	.16 [.05 - .33]	.14 [.09 - .22]

保留 (25 乳房 , 21 肺 , 49 非癌症):

特异性, 标靶	特异性, 观察	灵敏度, 肺	灵敏度, 乳房
.98	.92 [.80 - .98]*	.19 [.05 - .42]	.12 [.03 - .31]

图7

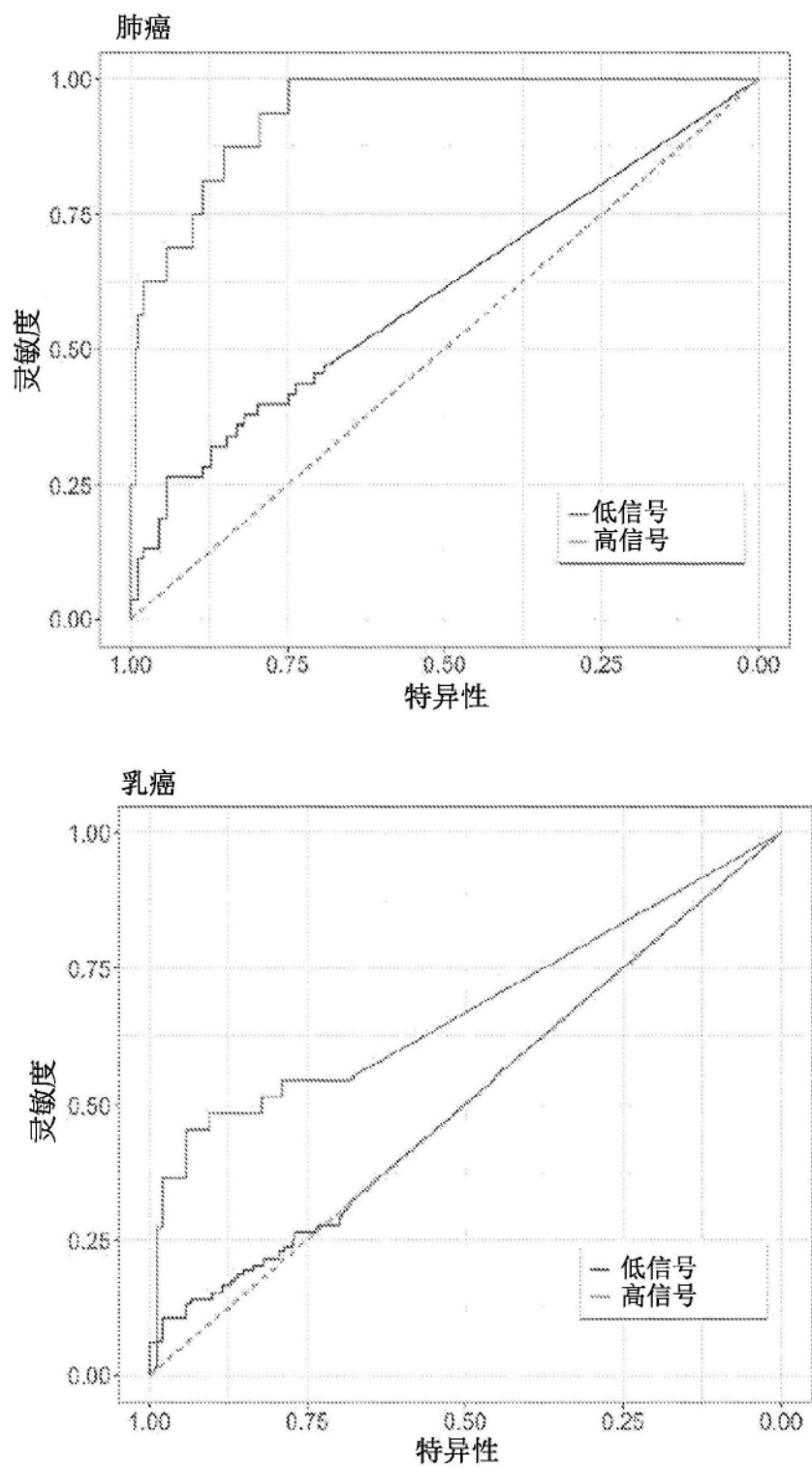


图8A

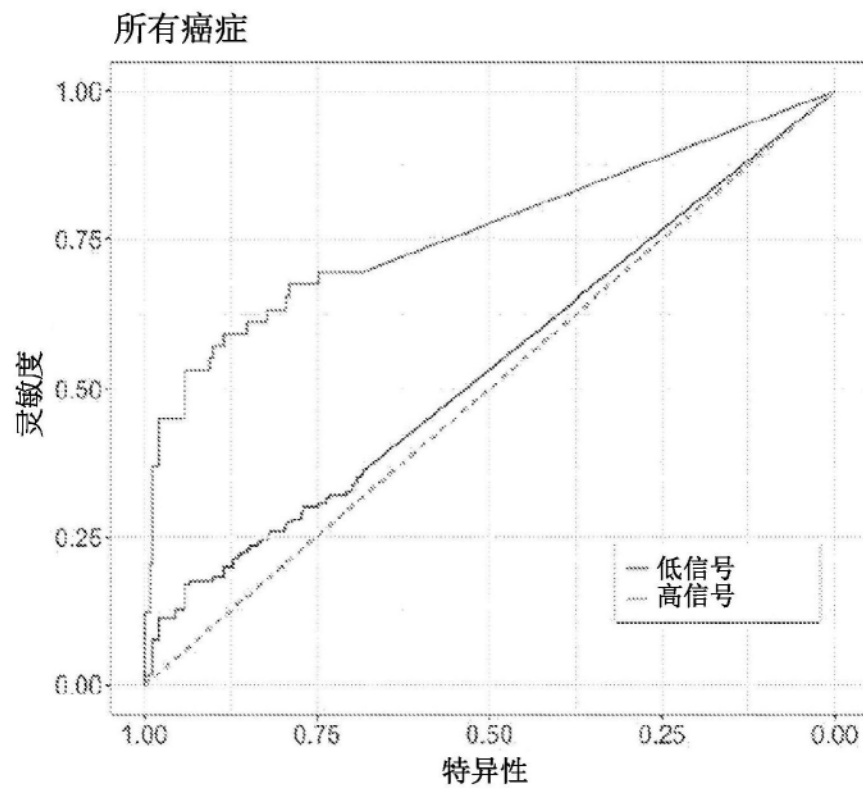


图8B

特异性 .98 [.95 - .99]	特异性, 肺		特异性, 乳房		特异性, 所有	
	低信号	高信号	低信号	高信号	低信号	高信号
	0.13 [.05 - .25]	0.63 [.35 - .85]	0.10 [.05 - .16]	0.36 [.20 - .55]	0.11 [.07 - .16]	0.45 [.31 - .60]

图8C

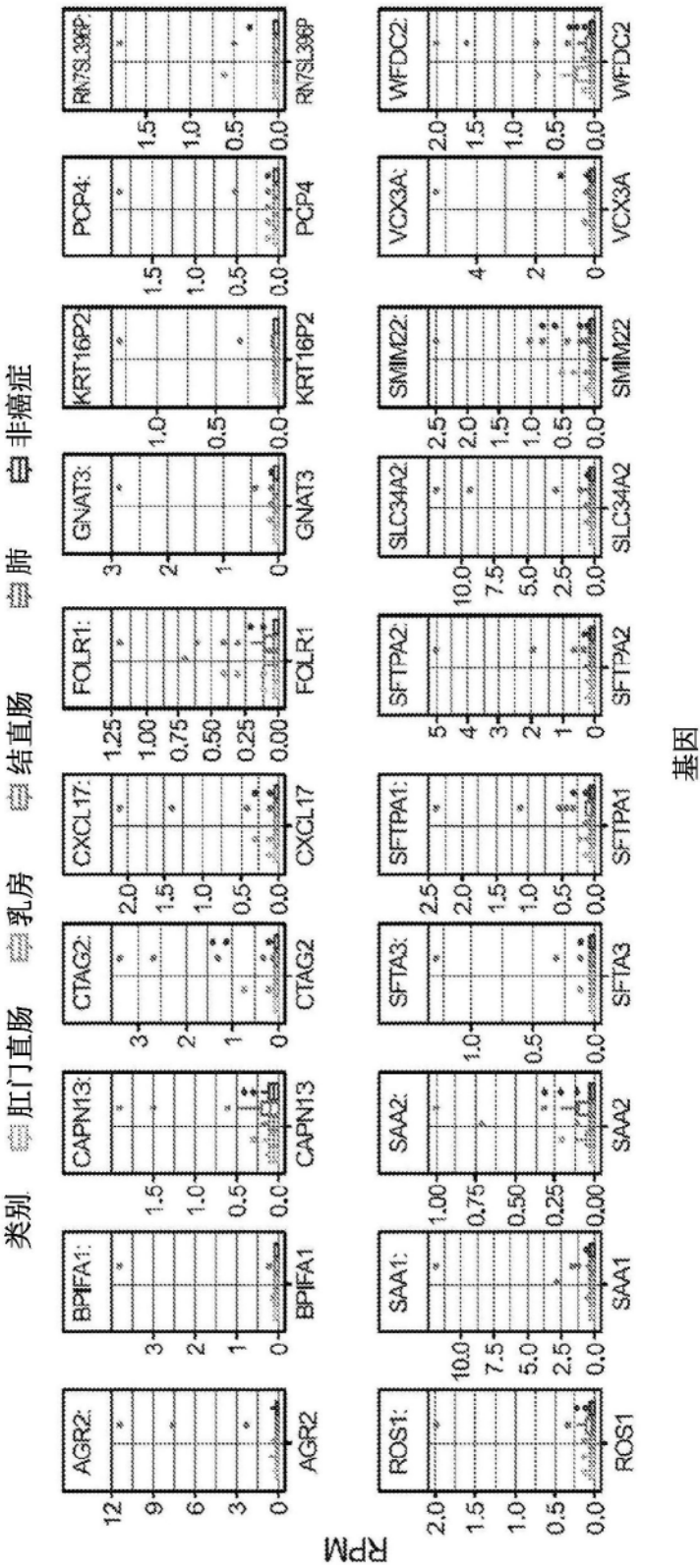


图9

$$S = G \cdot TS$$

$$\begin{bmatrix} S_{1,1} & S_{1,2} & \cdots & S_{1,n} \\ S_{2,1} & S_{2,2} & \cdots & S_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ S_{m,1} & S_{m,2} & \cdots & S_{m,n} \end{bmatrix} = \begin{bmatrix} G_{1,1} & S_{1,2} & \cdots & S_{1,n} \\ G_{2,1} & S_{2,2} & \cdots & S_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ G_{m,1} & S_{m,2} & \cdots & S_{m,n} \end{bmatrix} \begin{bmatrix} TS_{1,1} & TS_{1,2} & \cdots & TS_{1,j} \\ TS_{2,1} & TS_{2,2} & \cdots & TS_{2,j} \\ \vdots & \vdots & \ddots & \vdots \\ TS_{n,1} & TS_{n,2} & \cdots & TS_{n,j} \end{bmatrix}$$

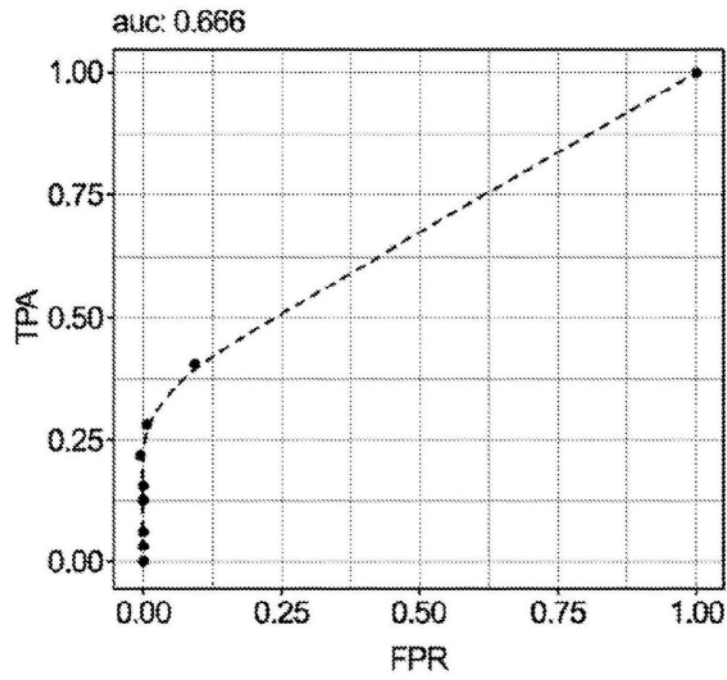


图10



图11

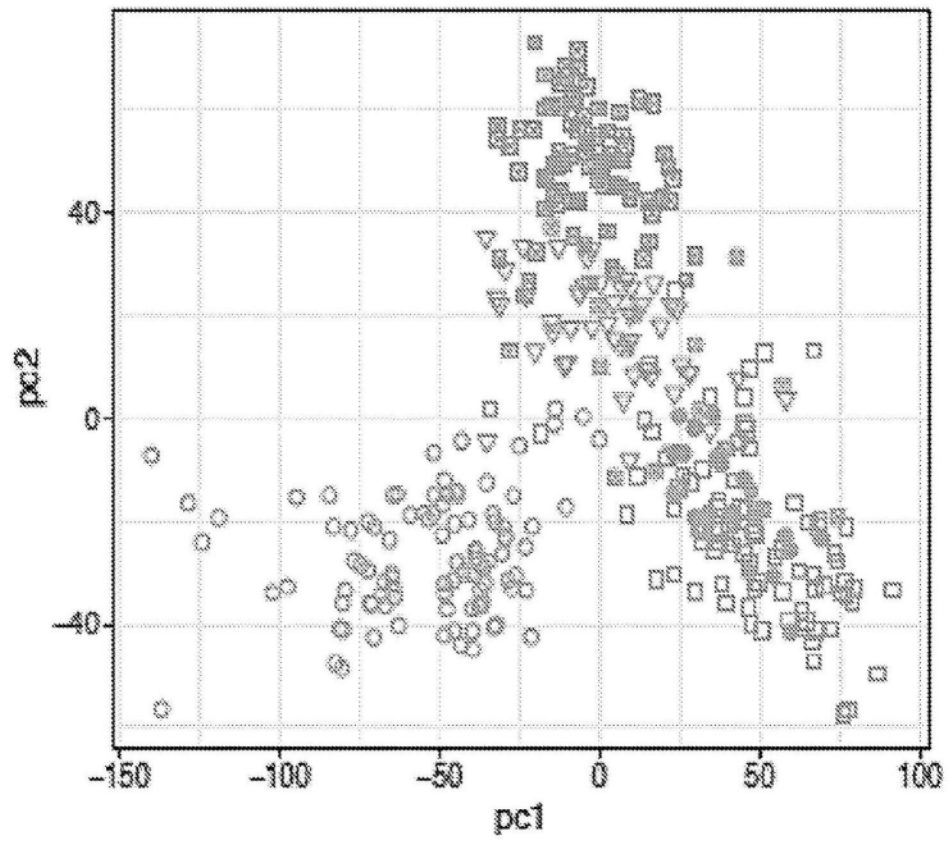


图12A

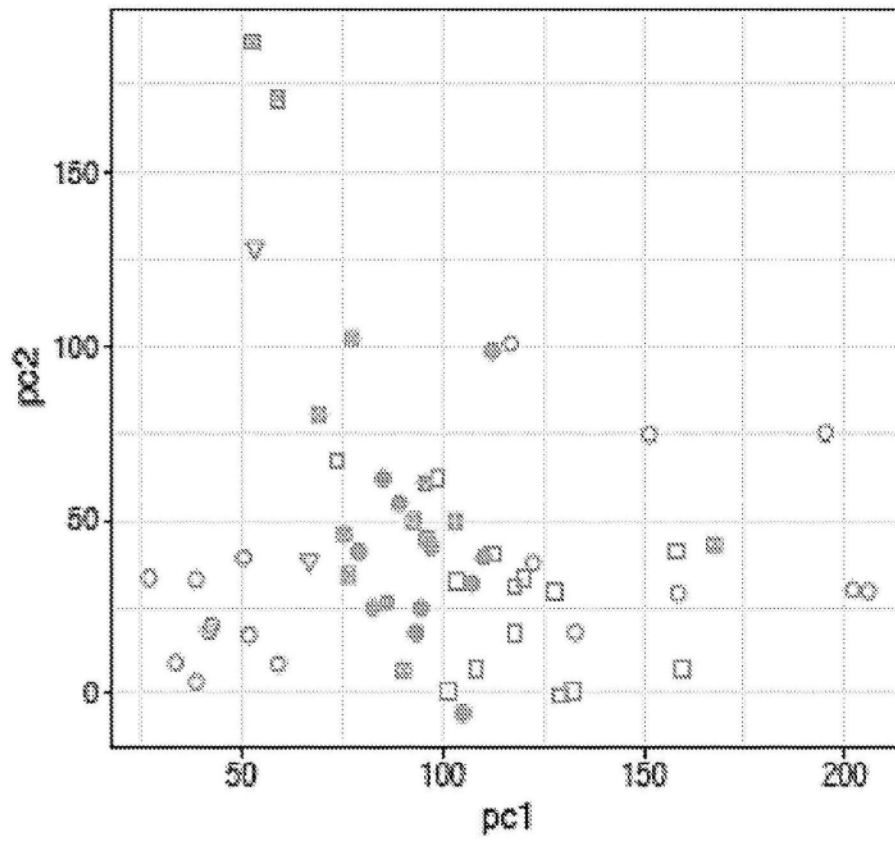


图12B

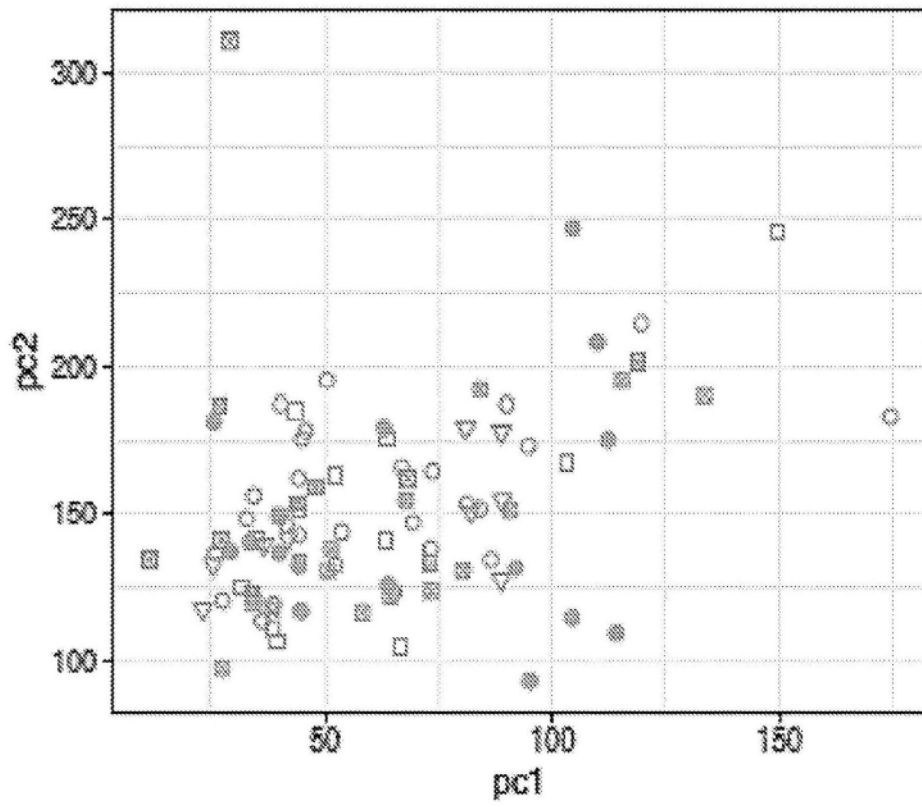


图12C

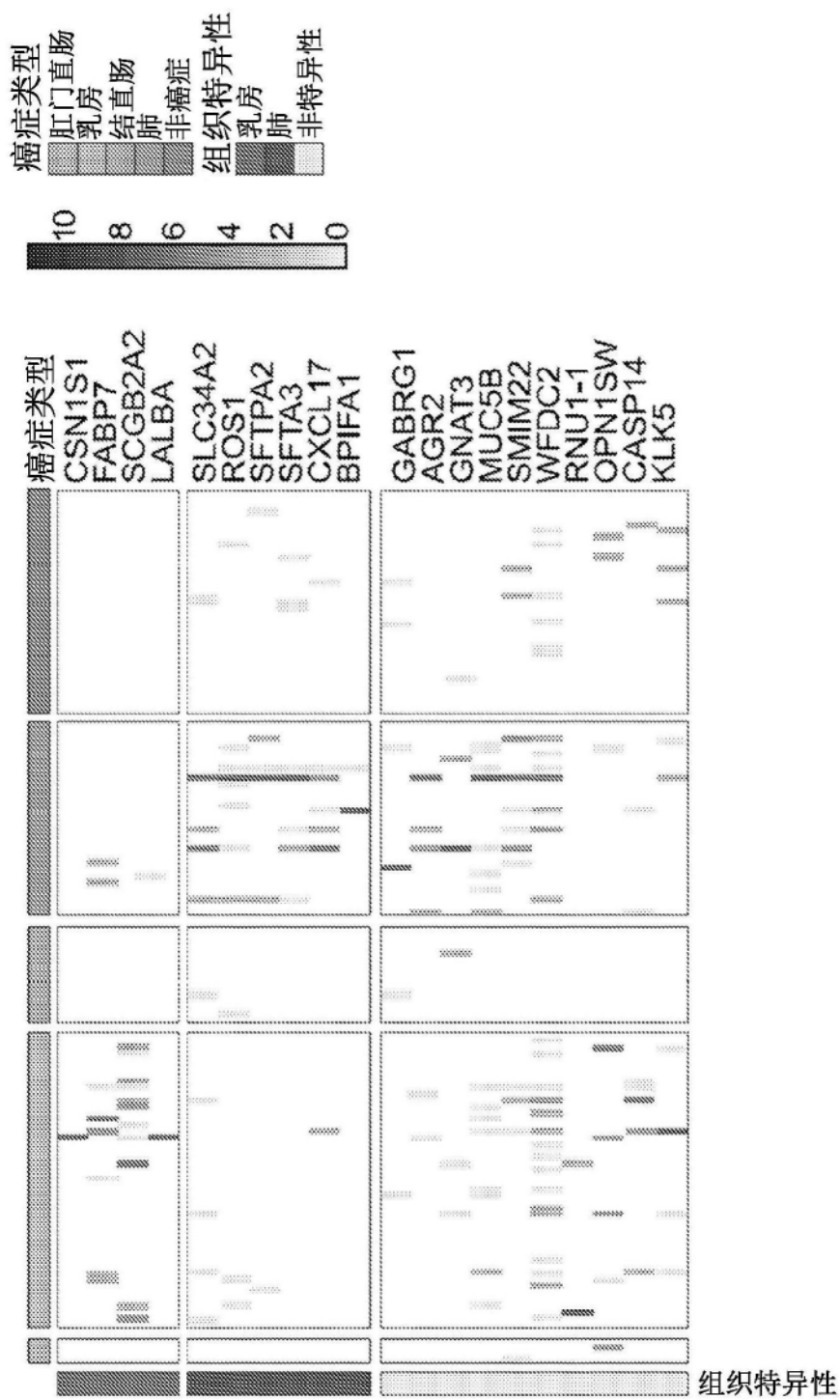


图13

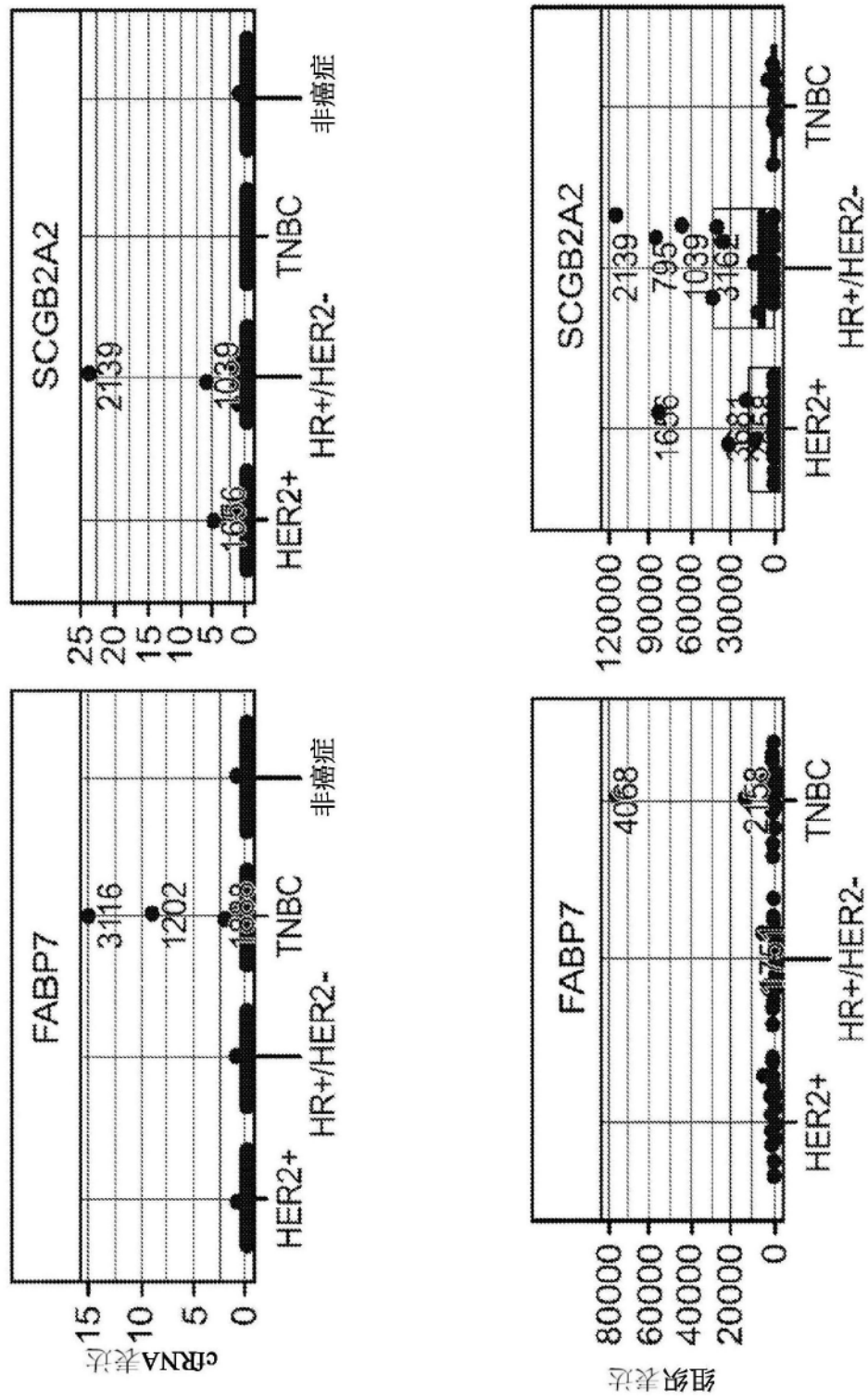


图14A

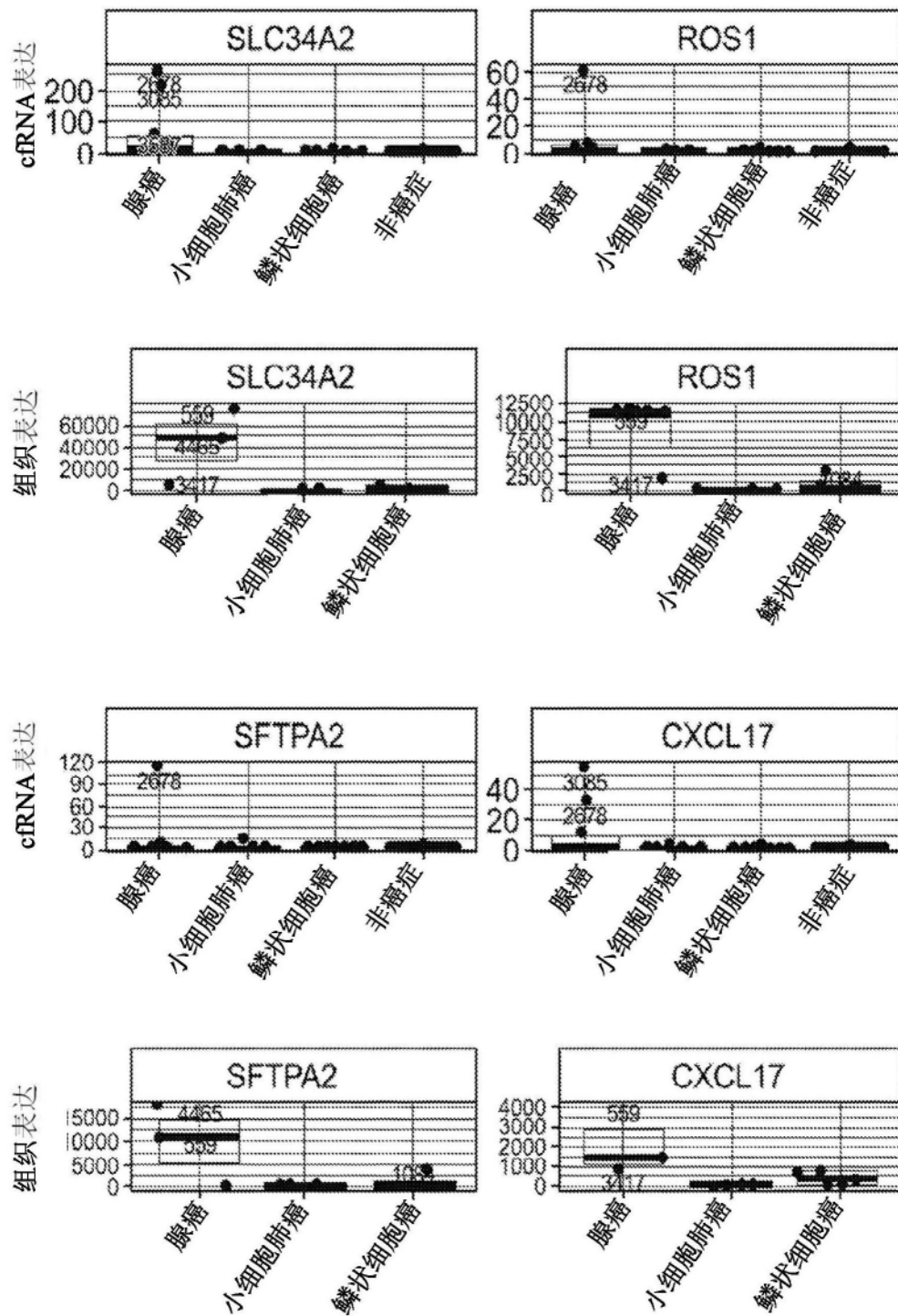


图14B

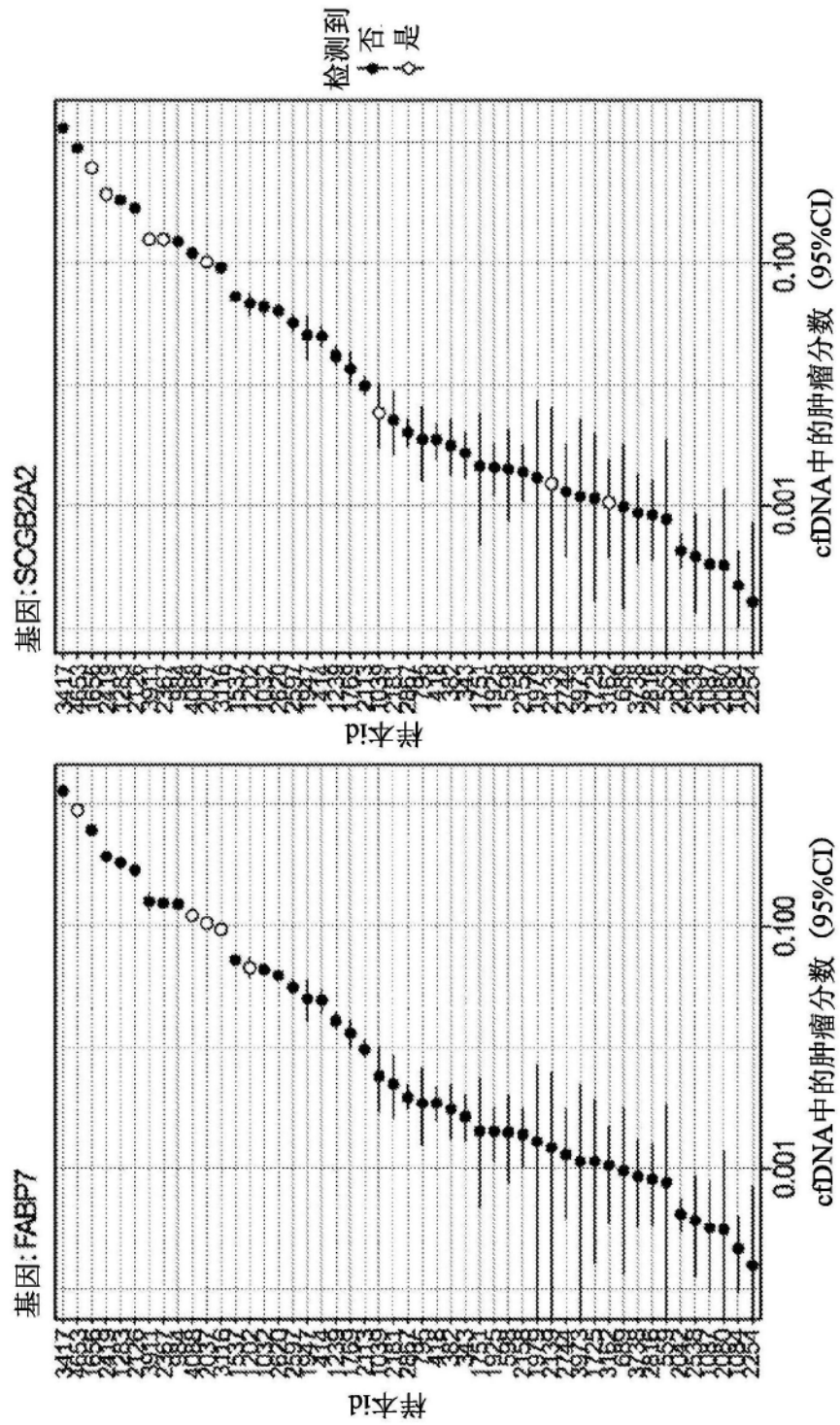


图15A

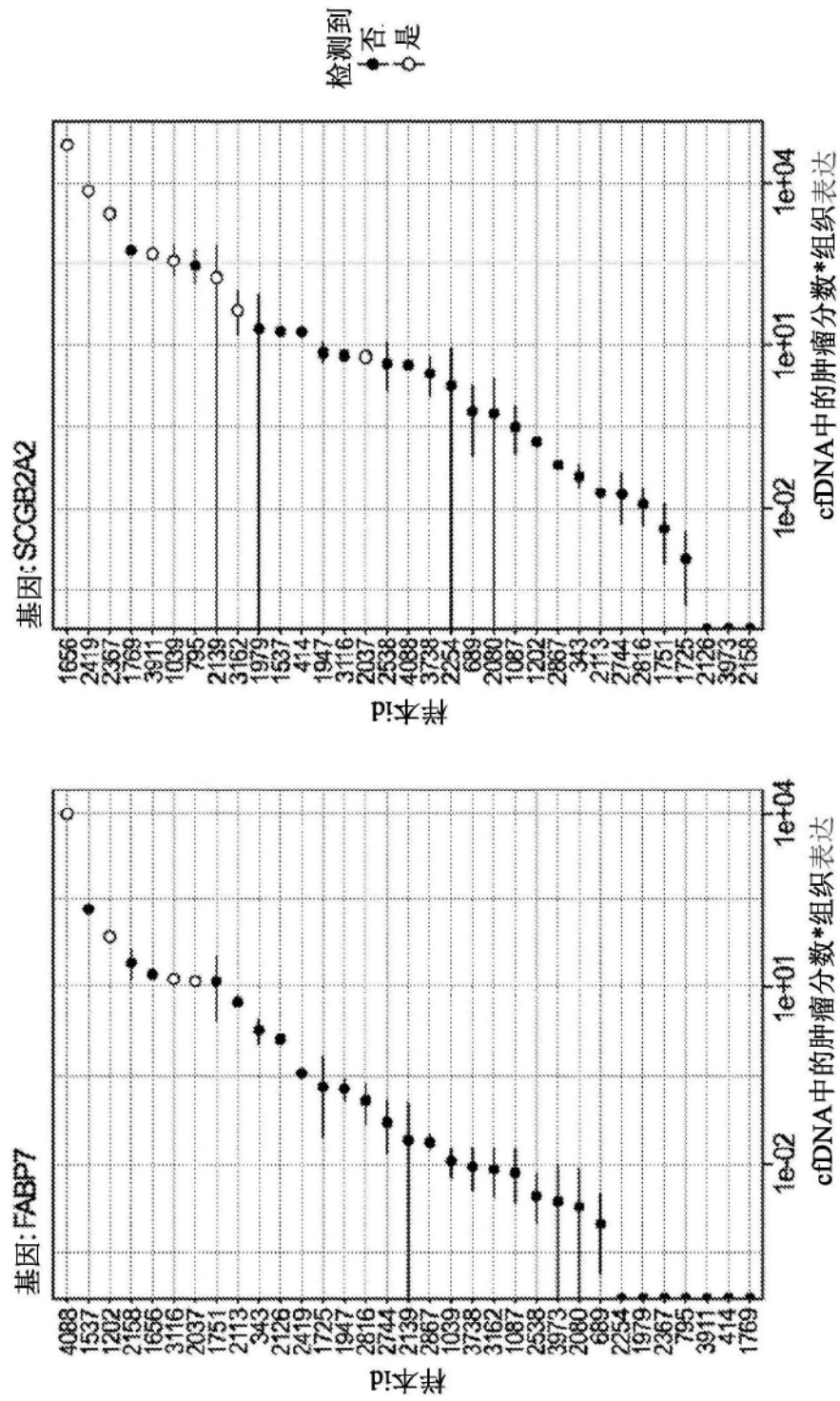


图15B

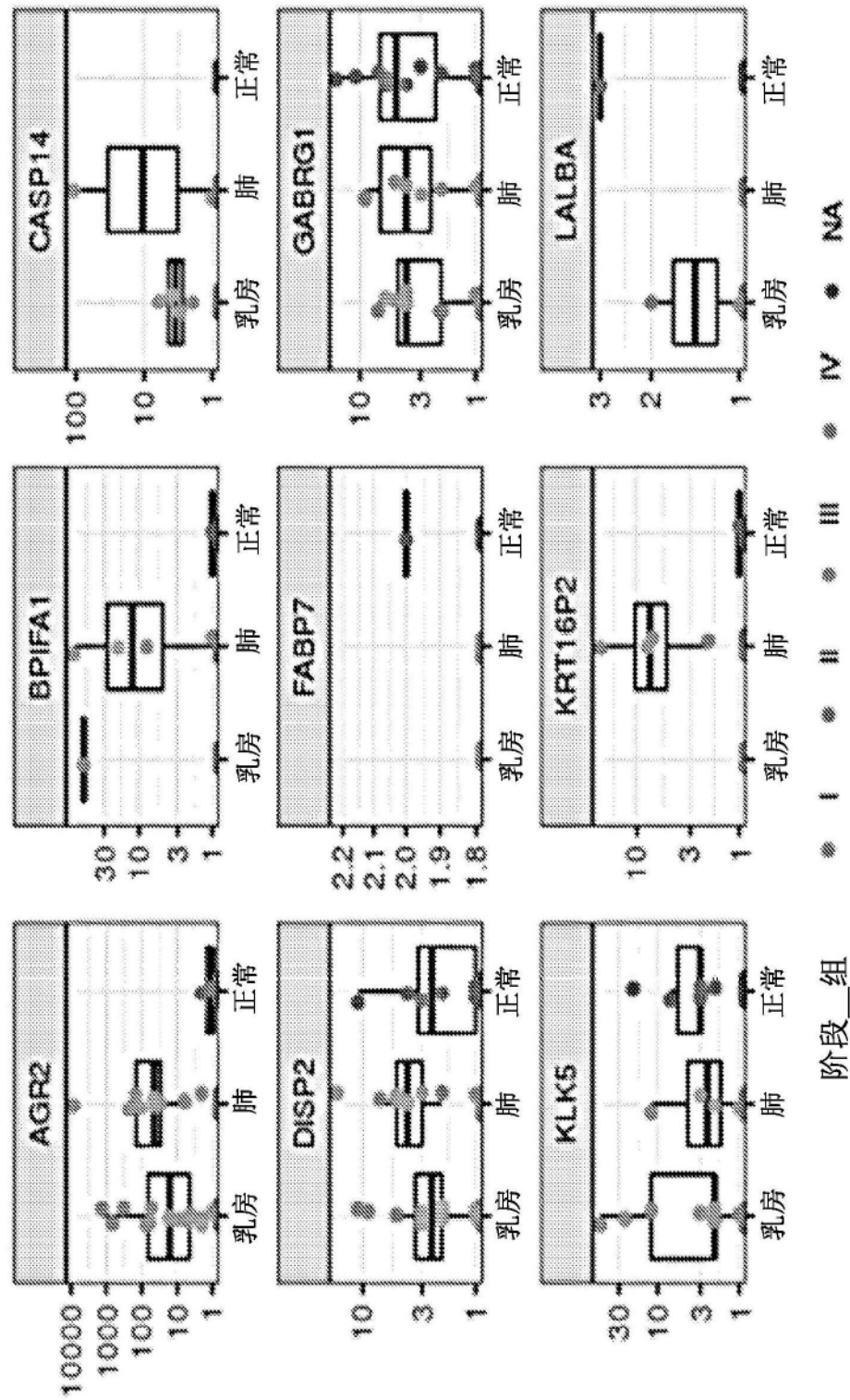


图16A

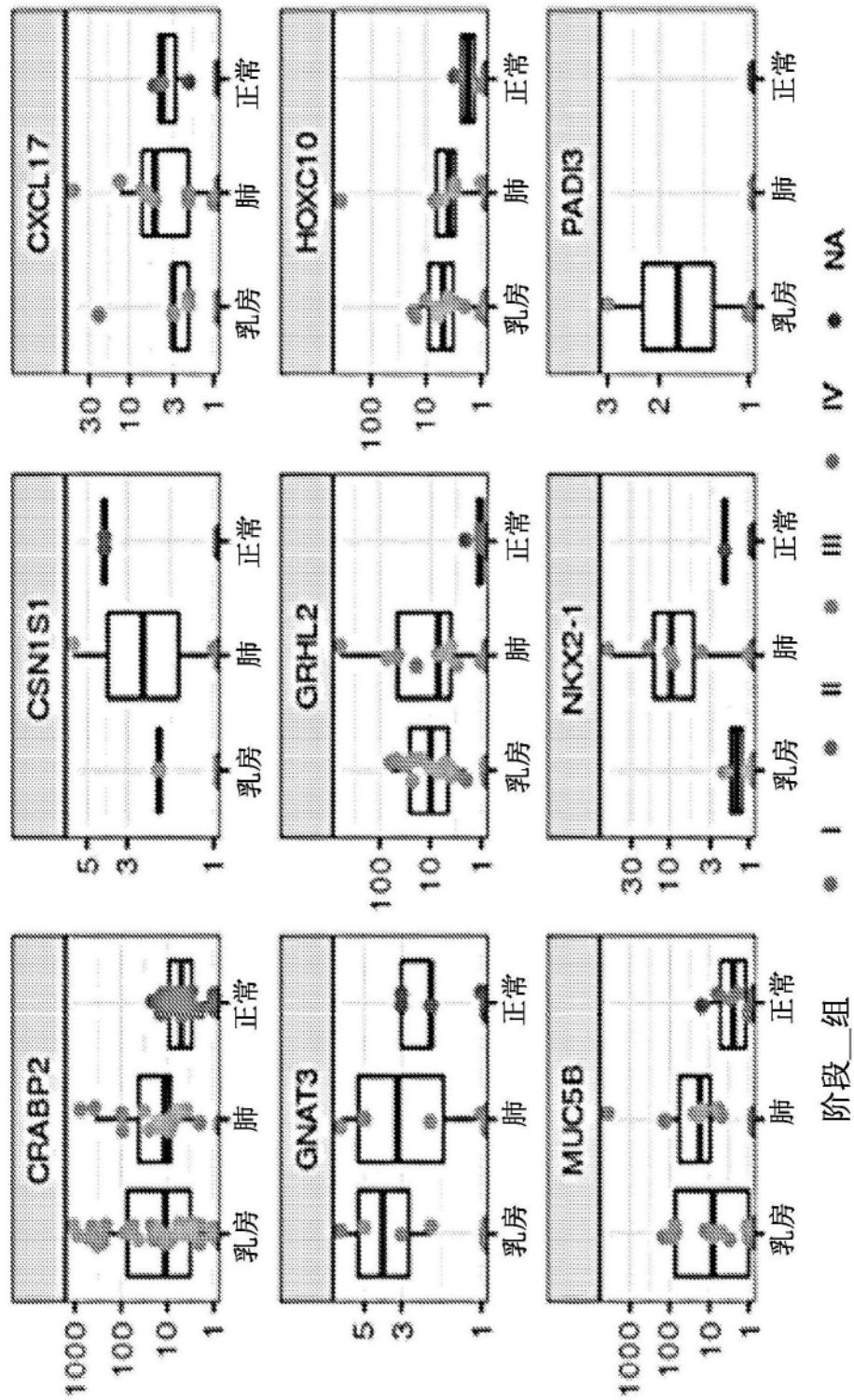


图16B

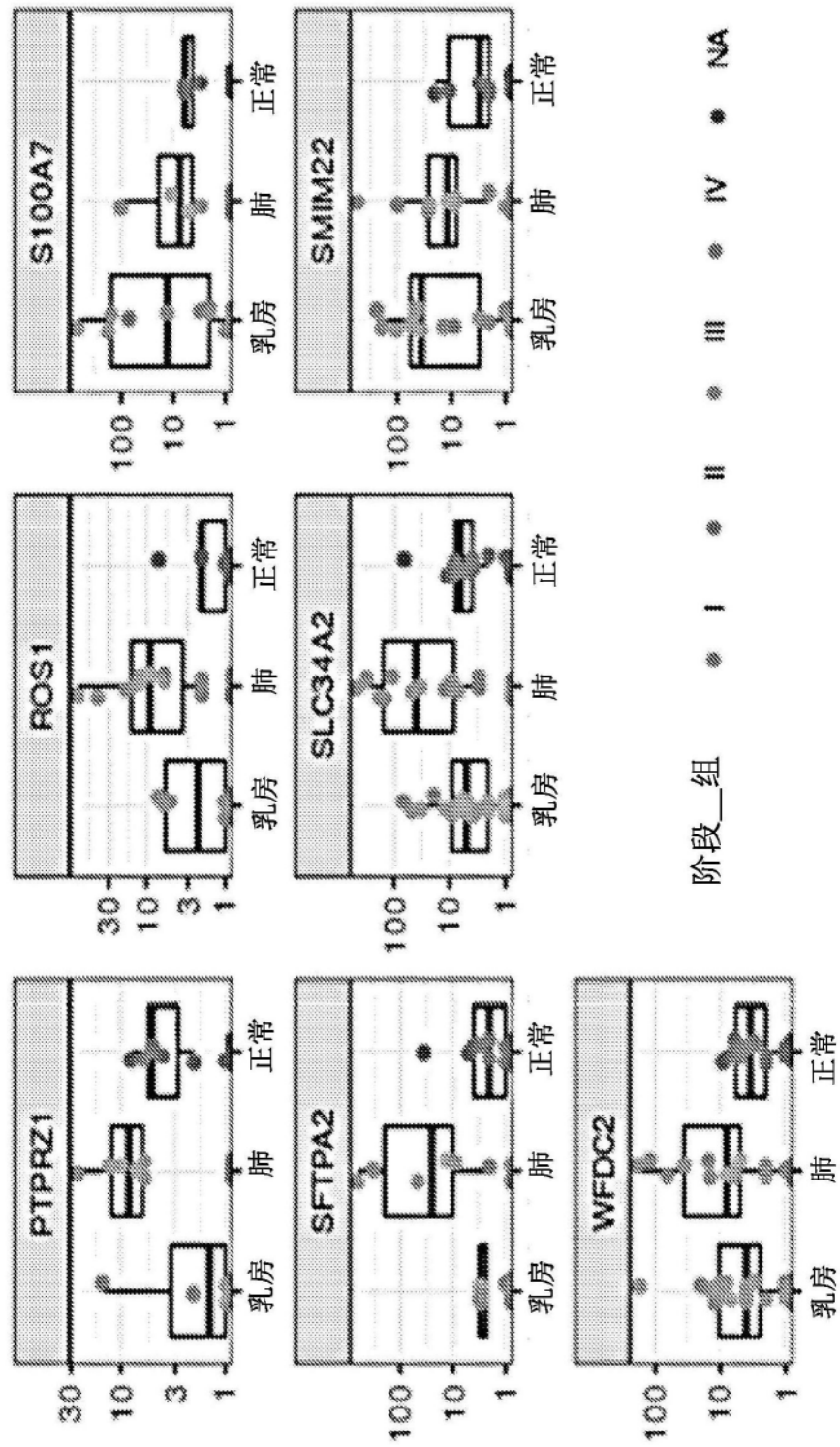


图16C

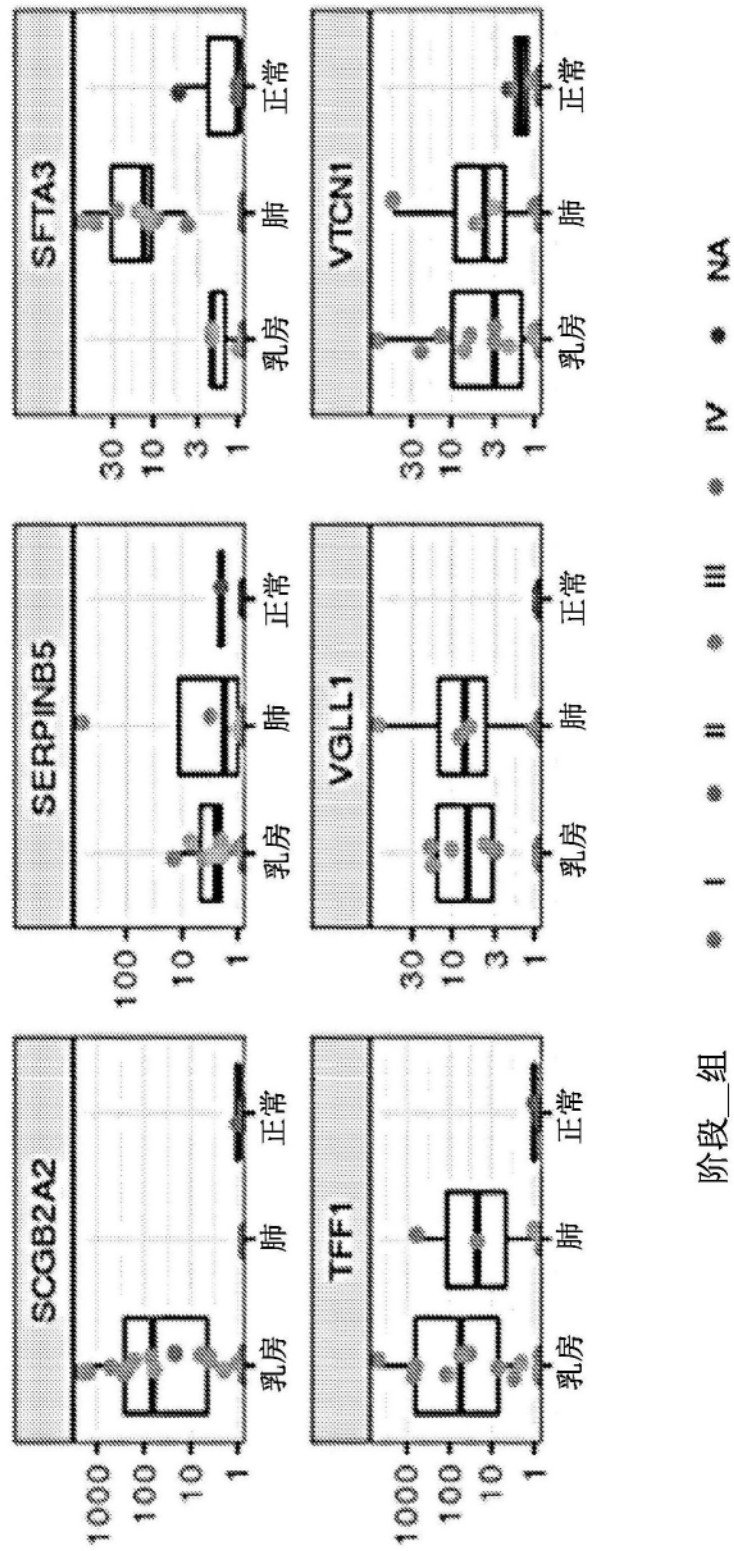


图16D

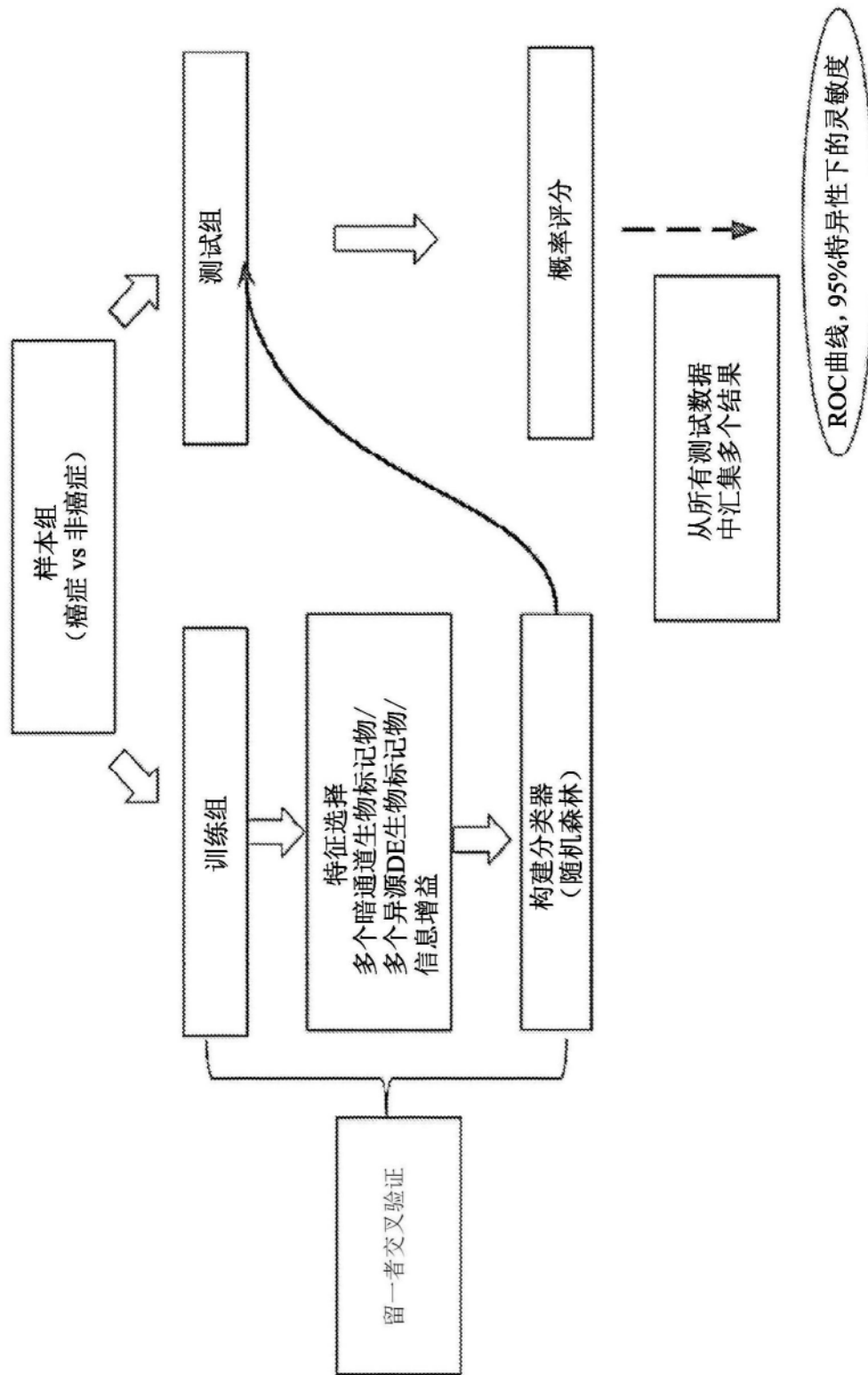


图17A

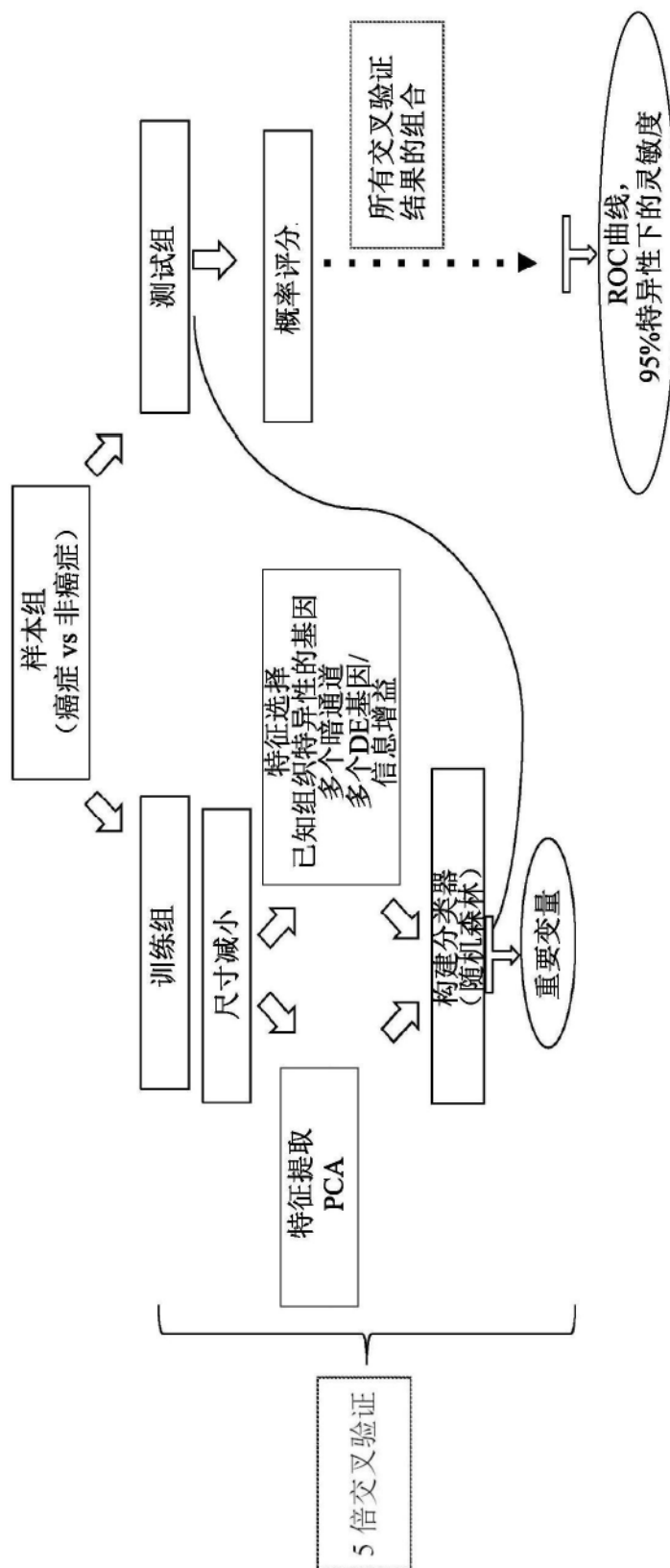


图17B

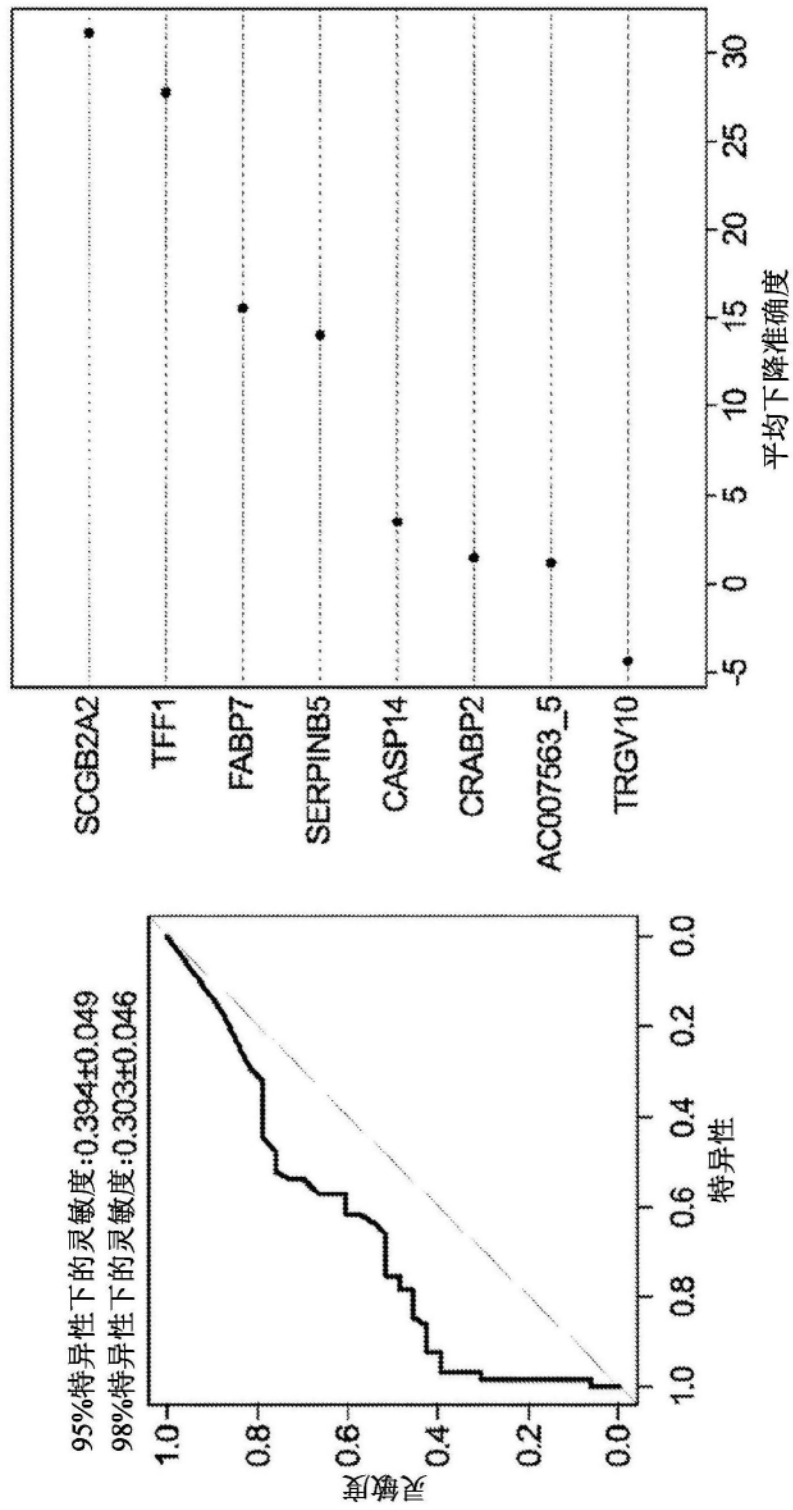


图18A

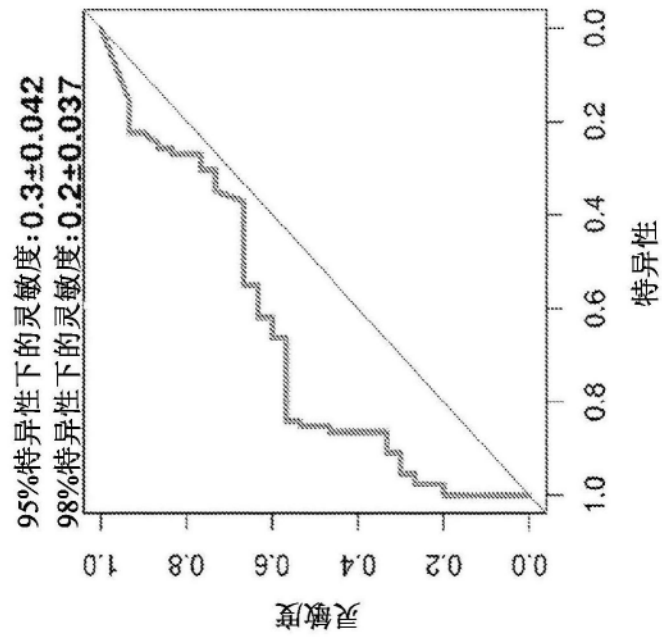


图18B

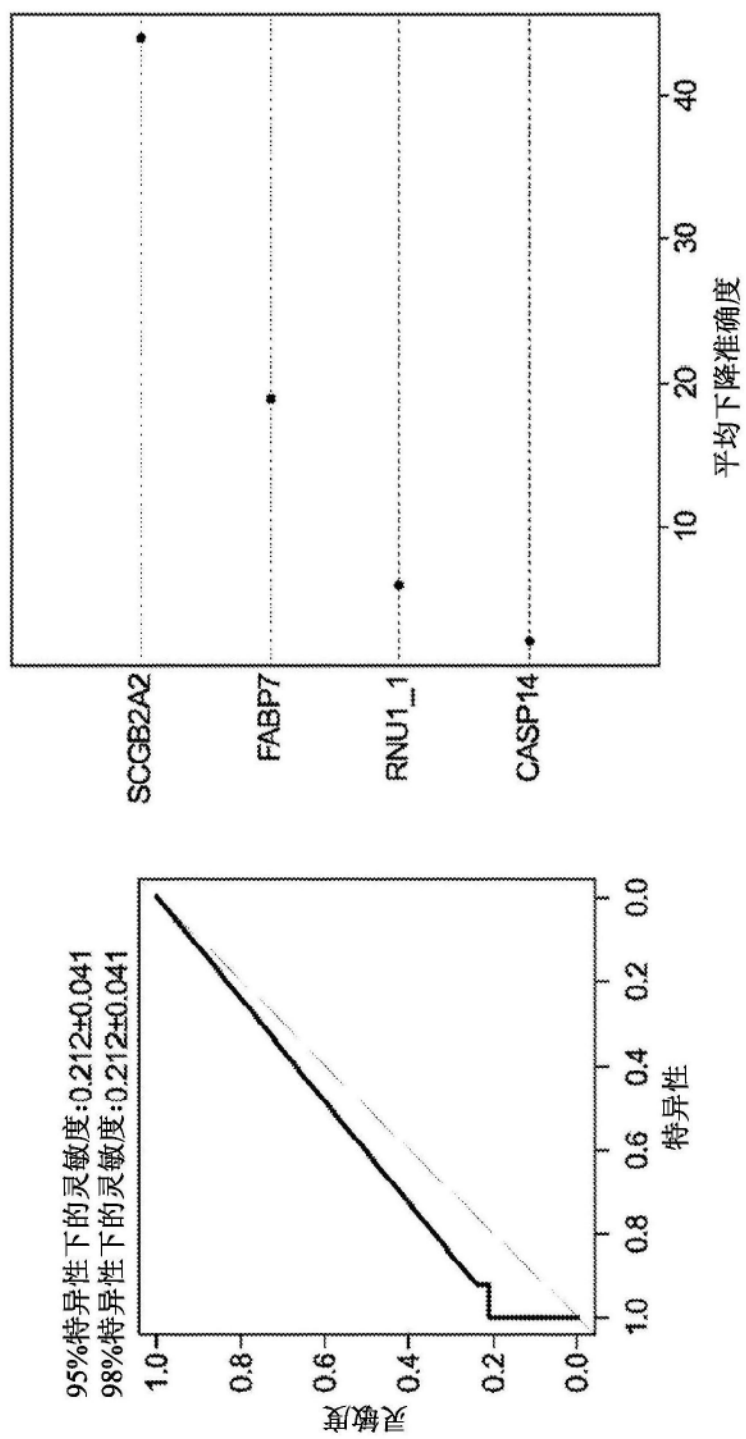


图18C

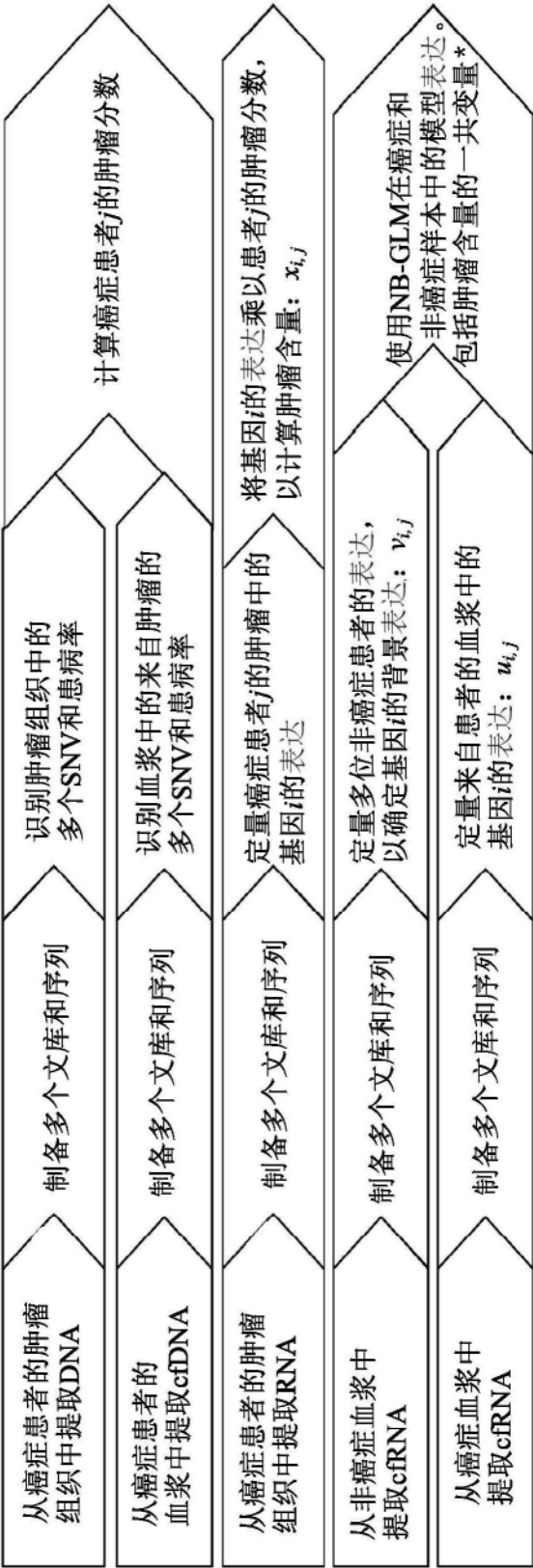


图19

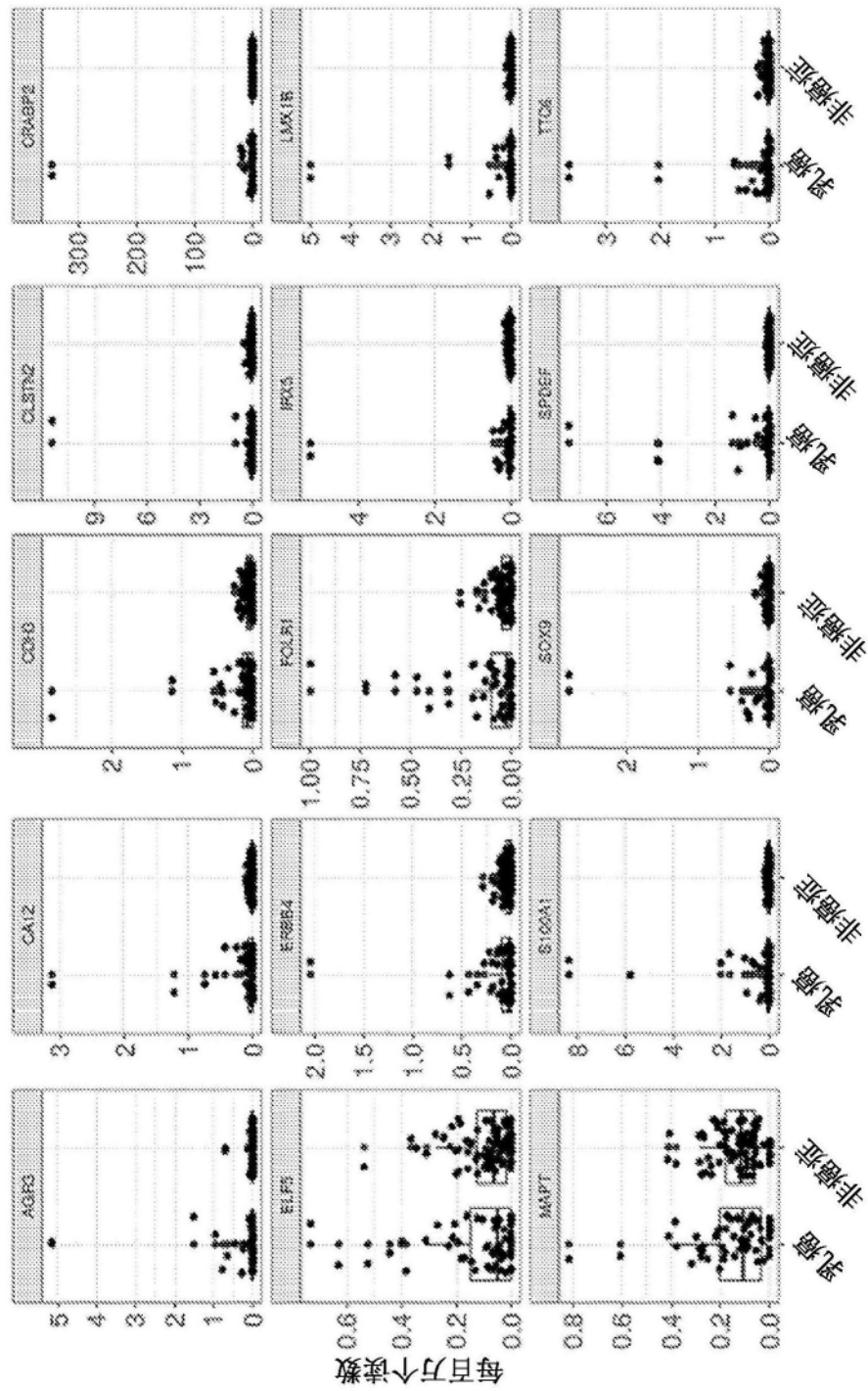


图20A

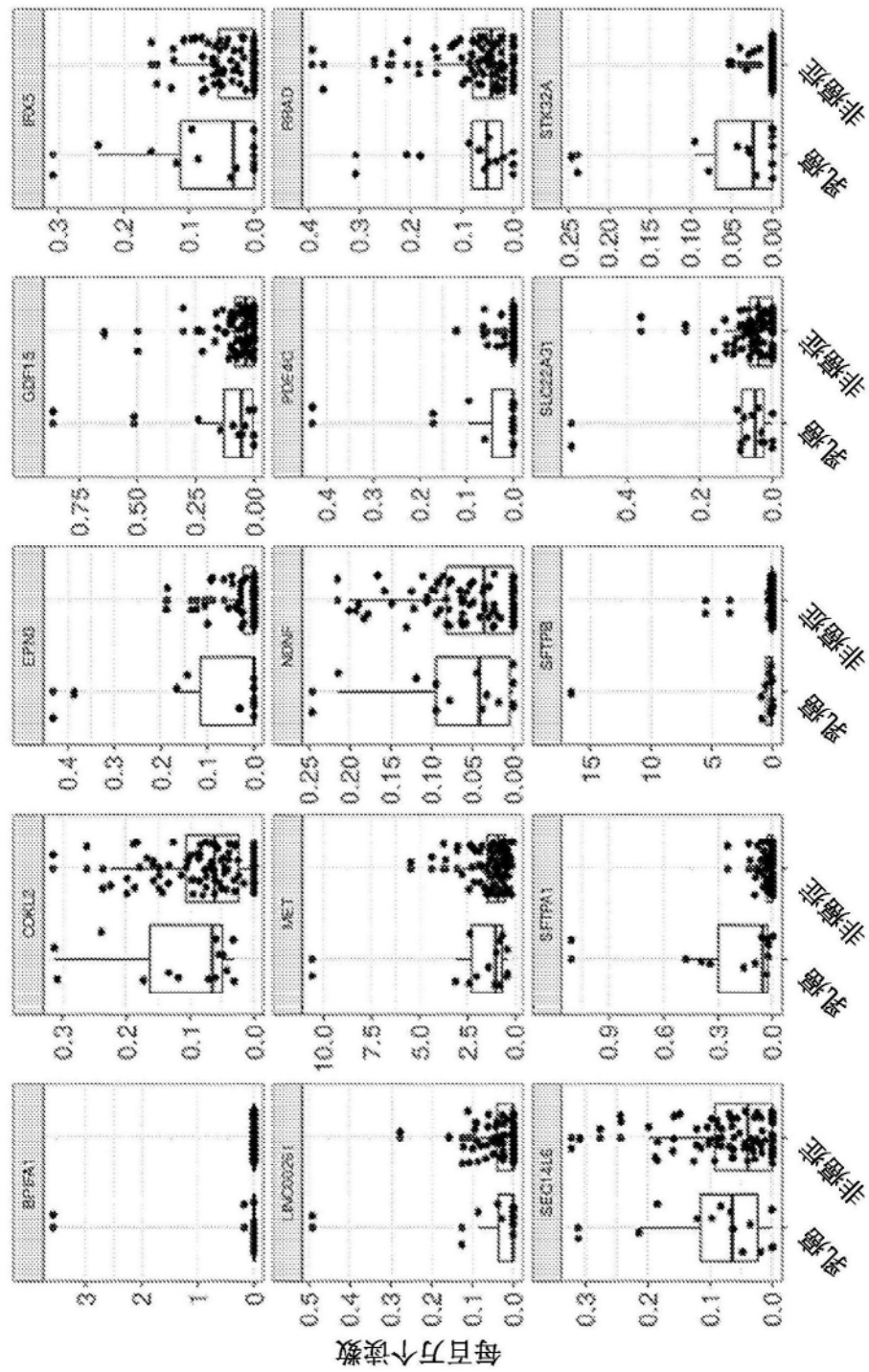


图20B

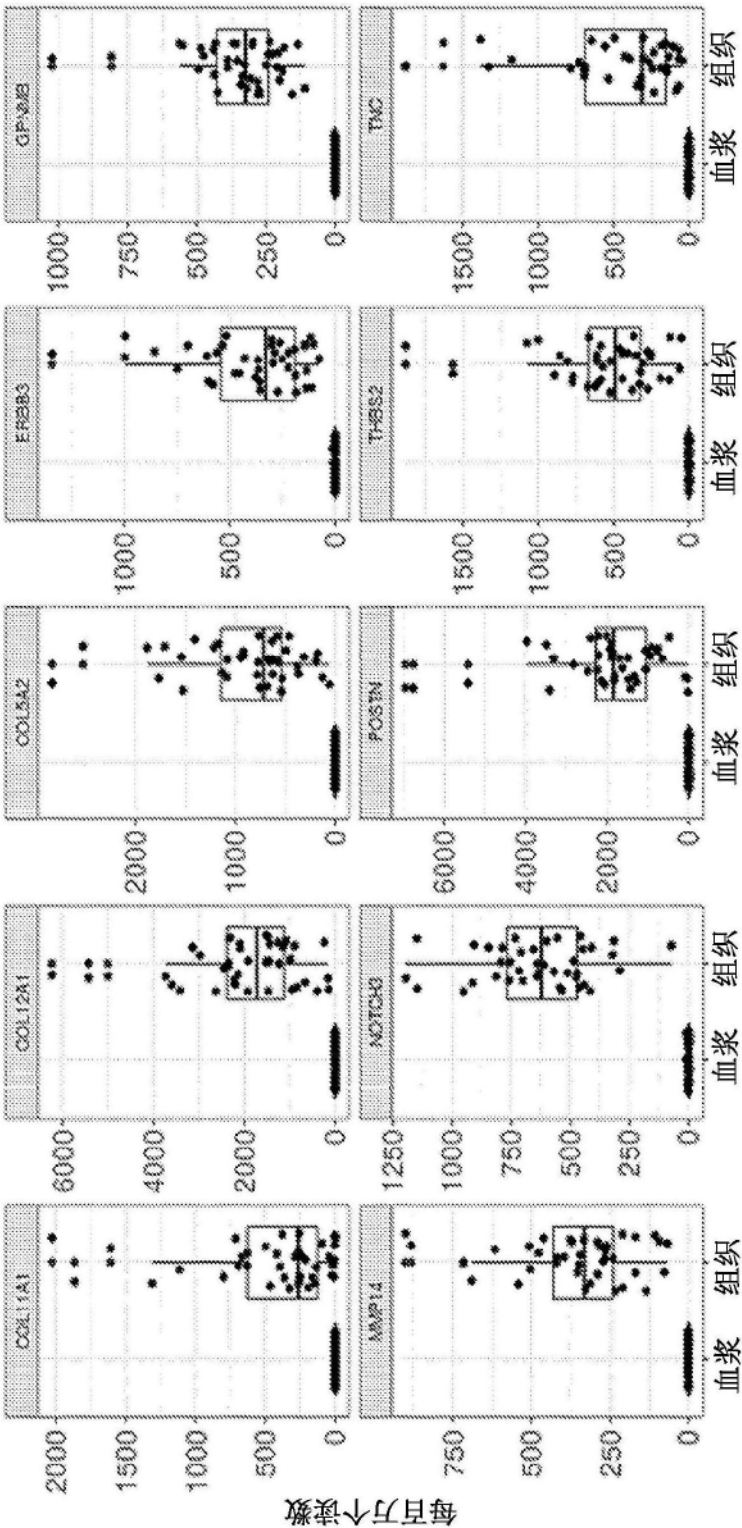


图21

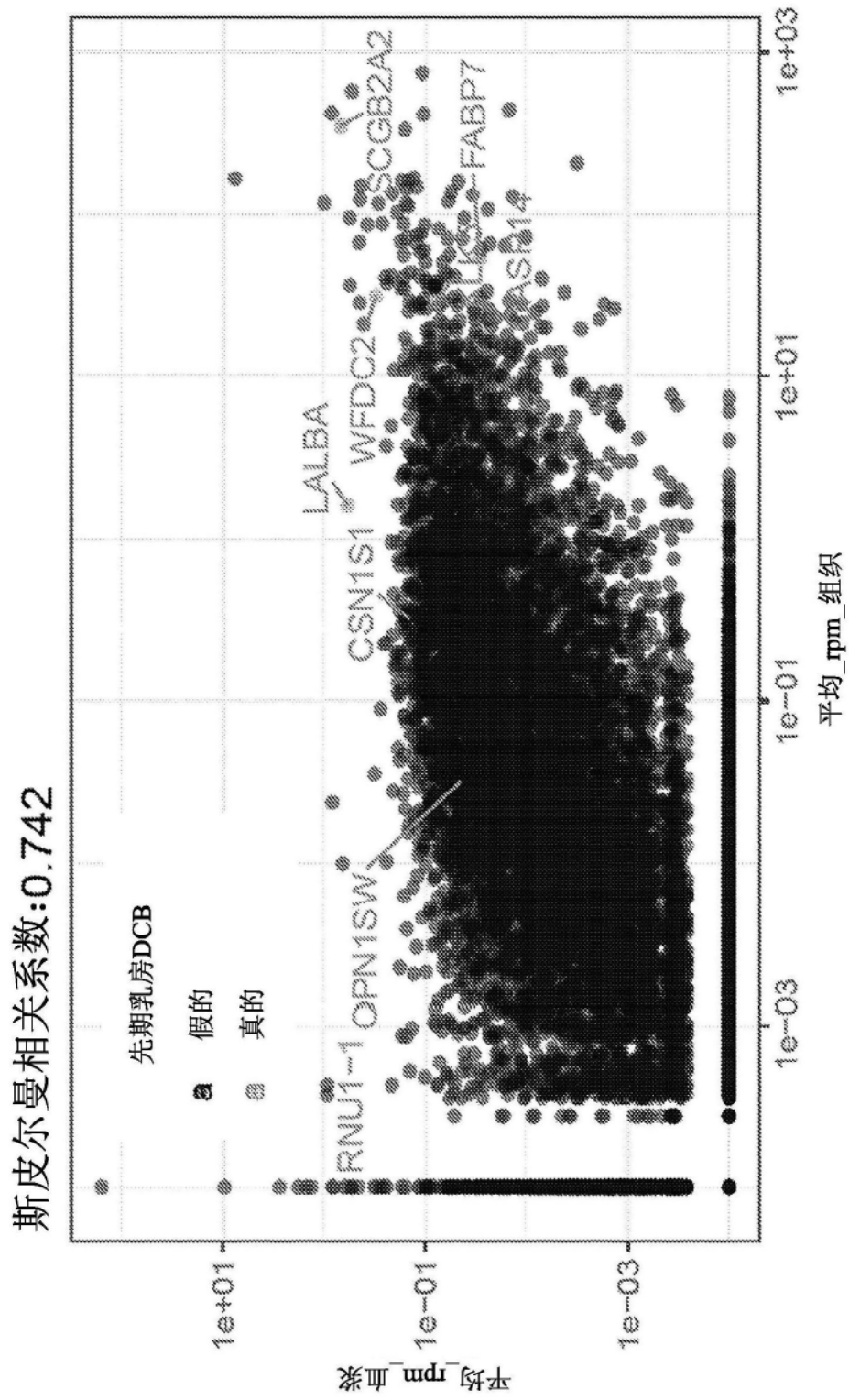


图22

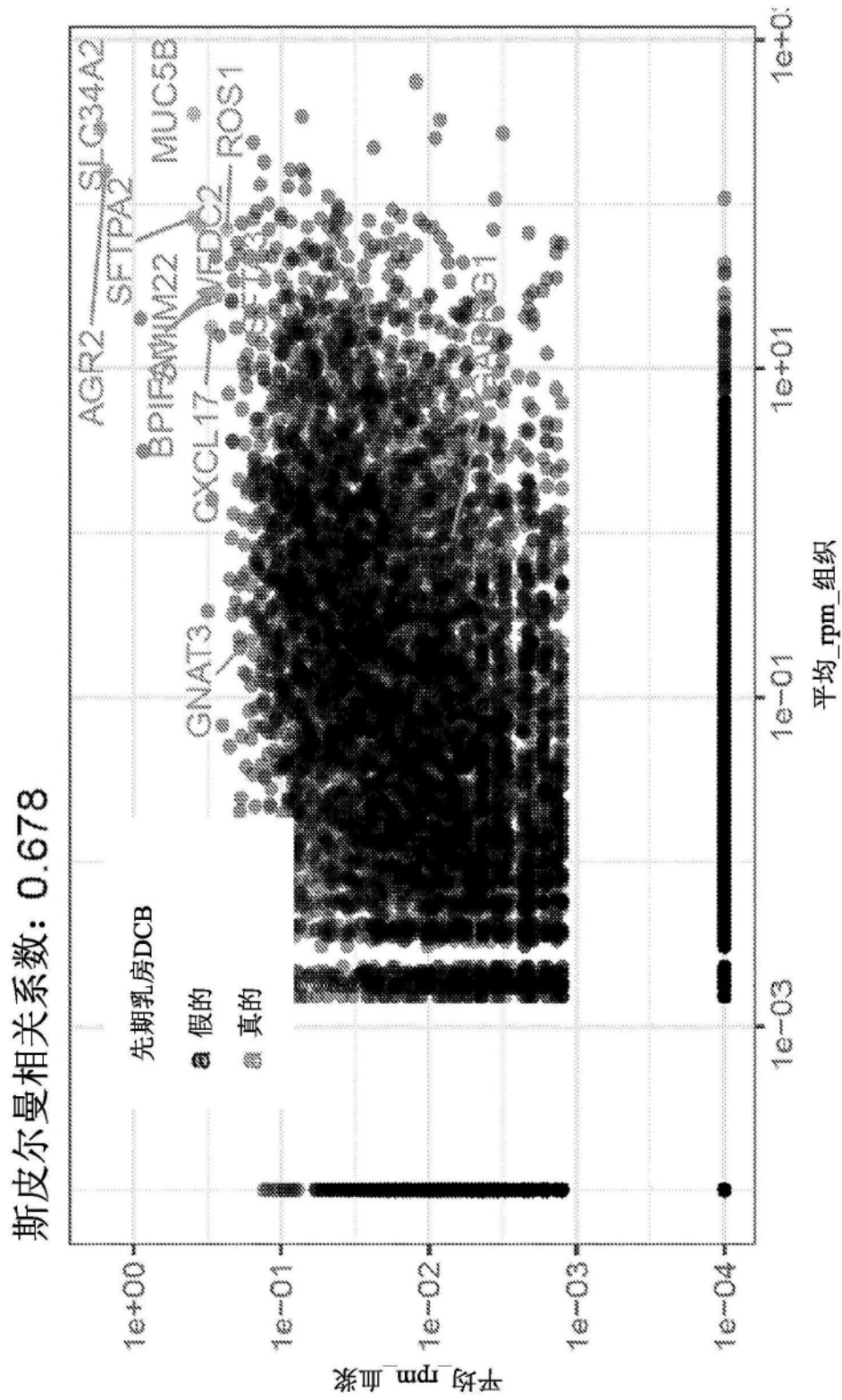


图23

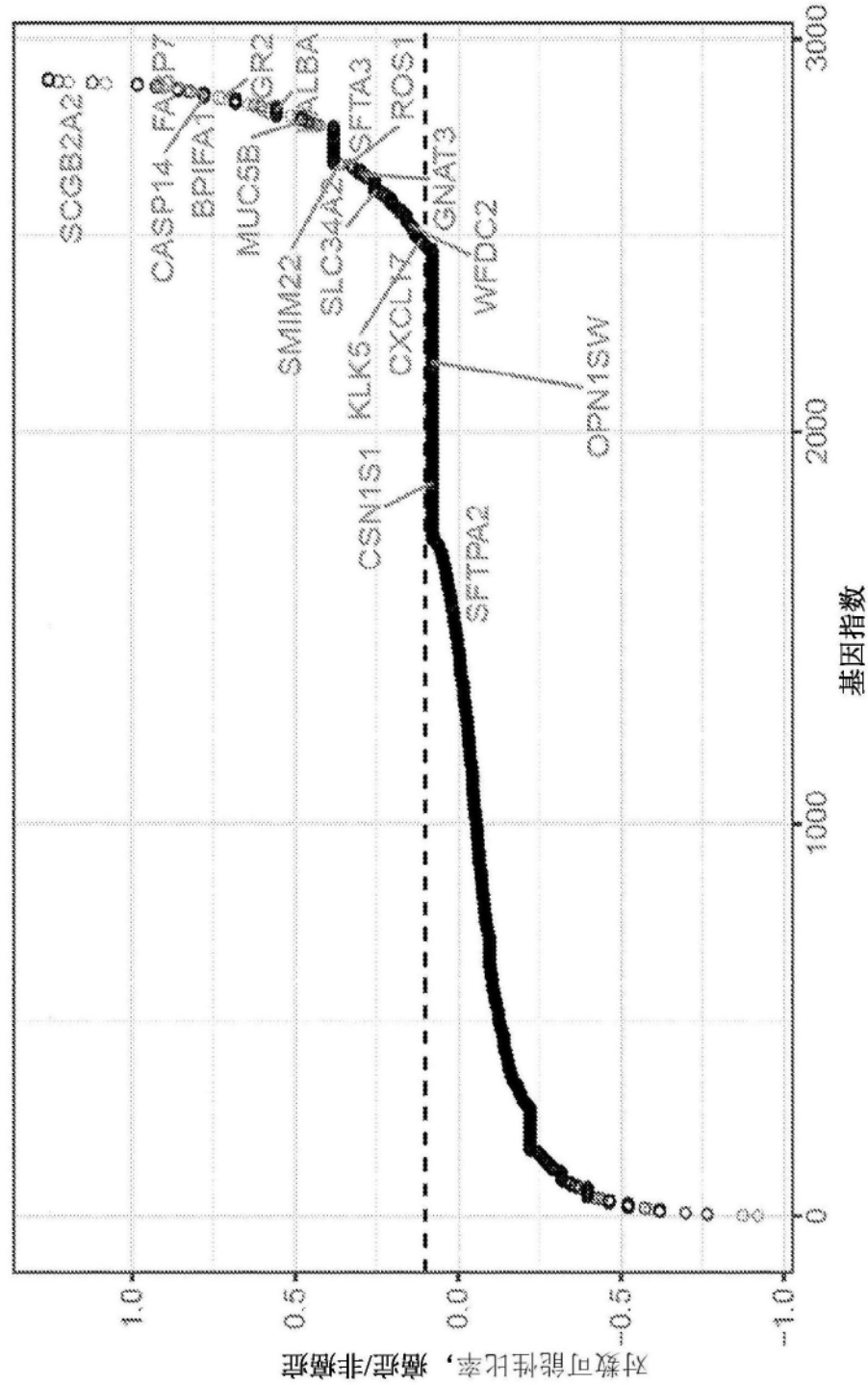


图24

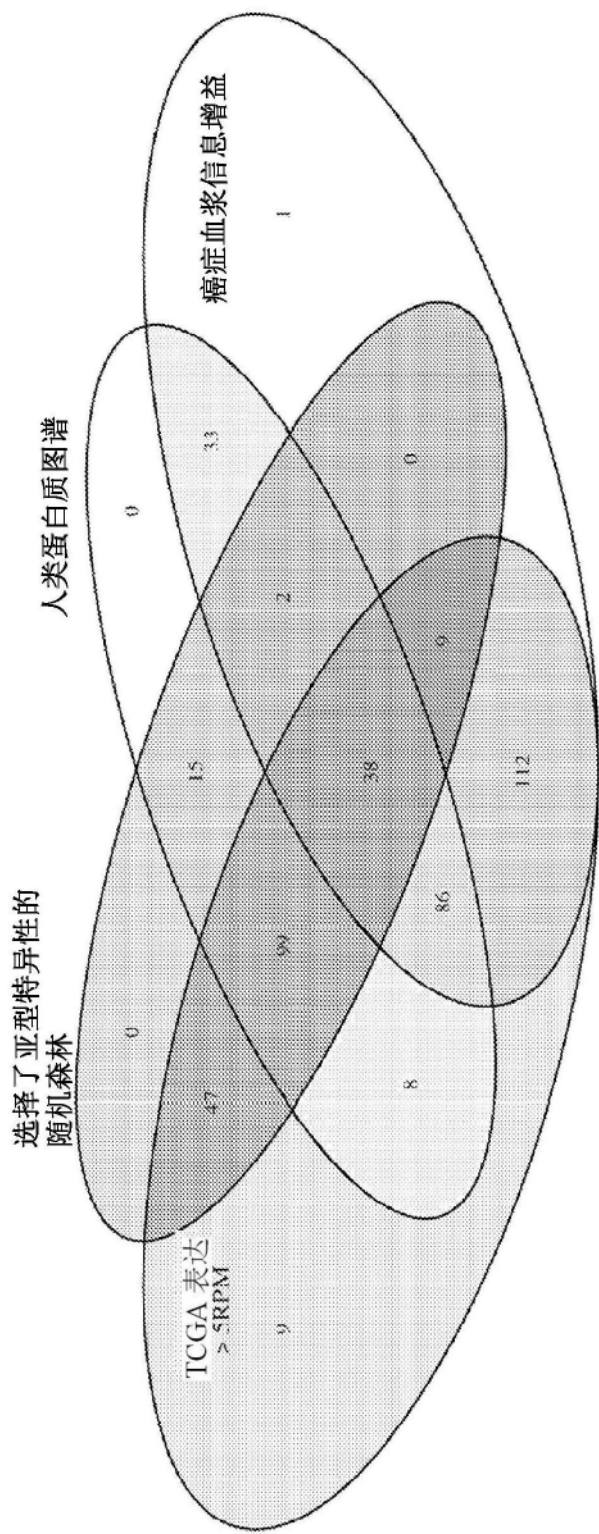


图25

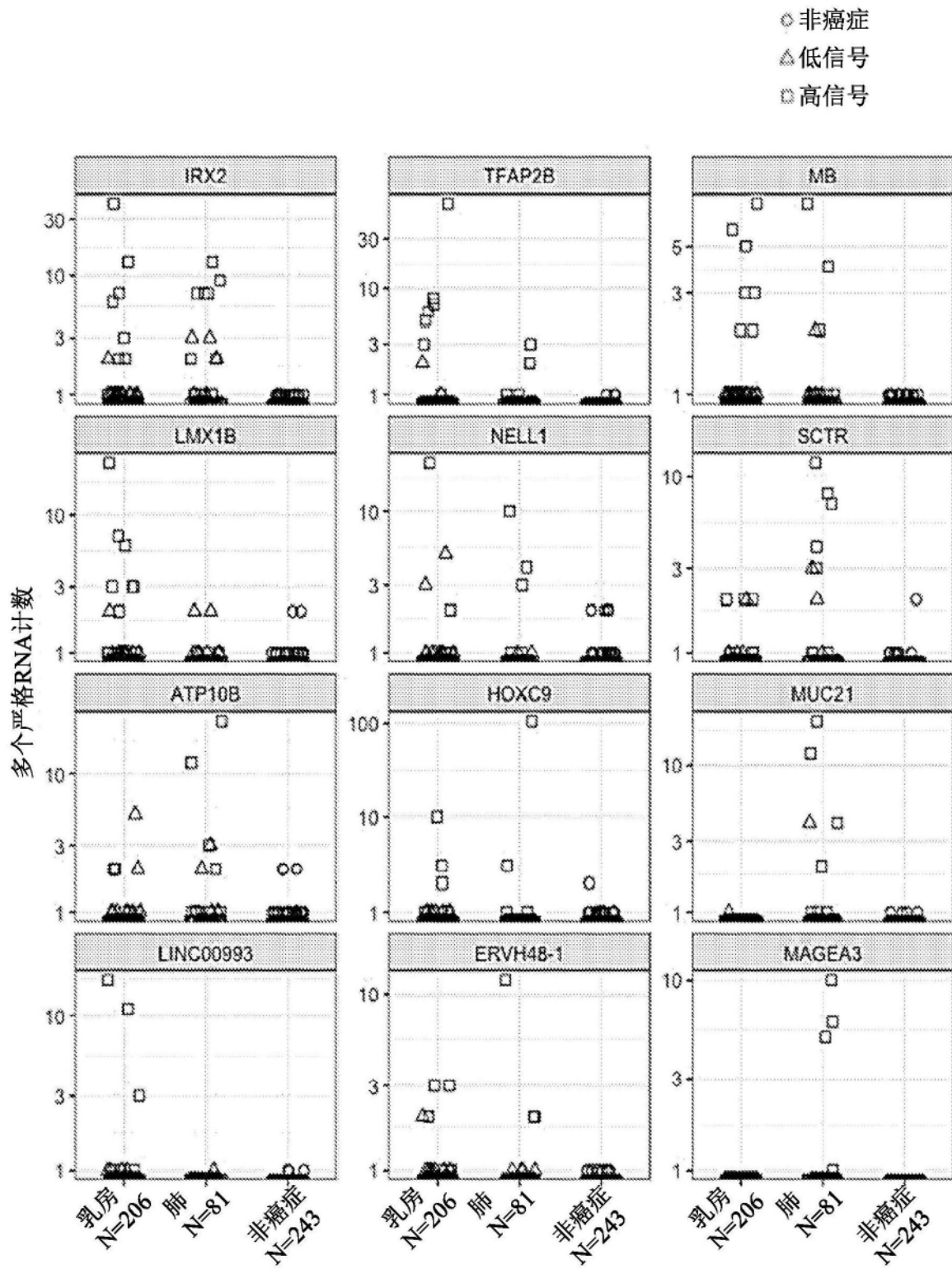


图26A

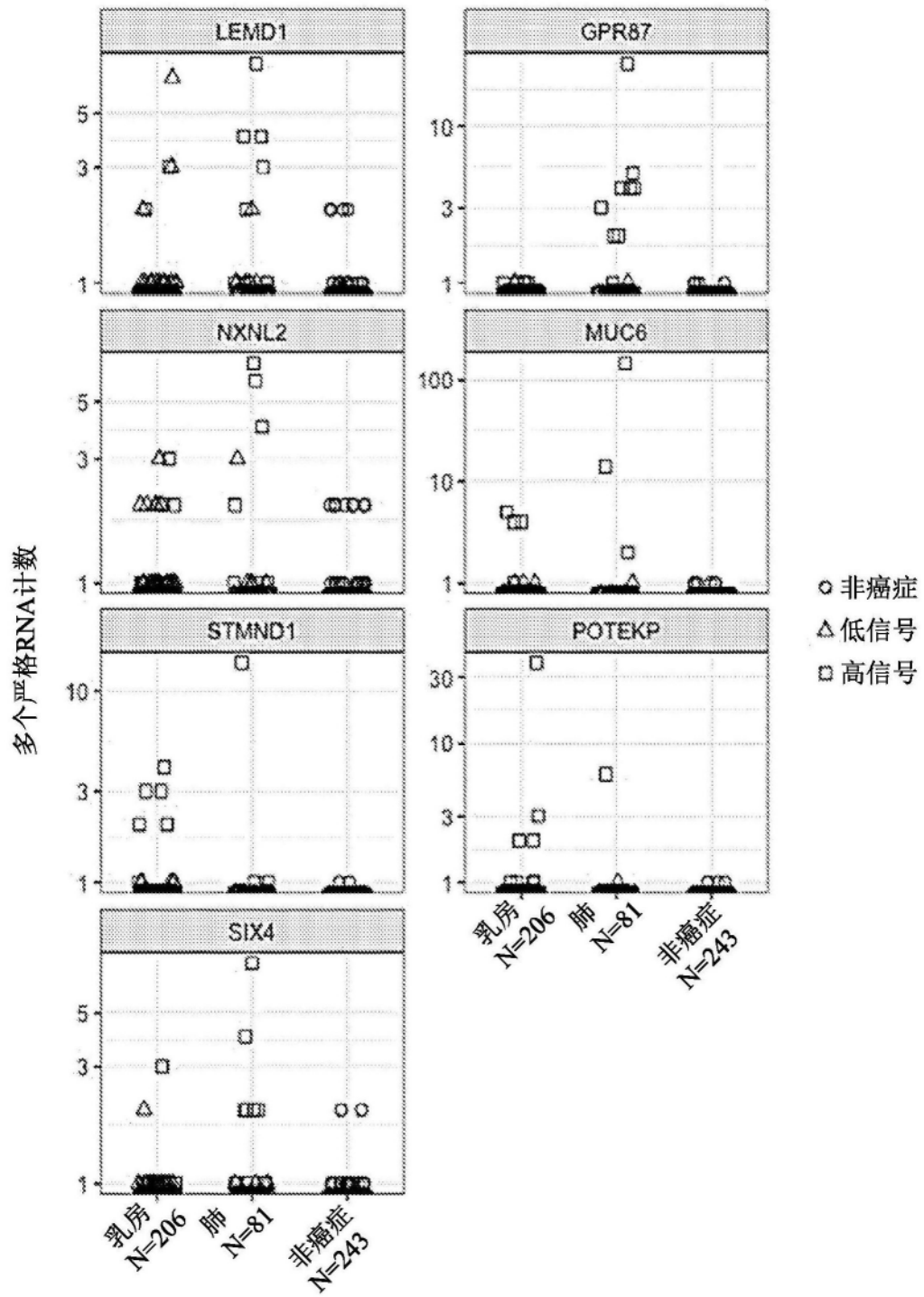


图26B

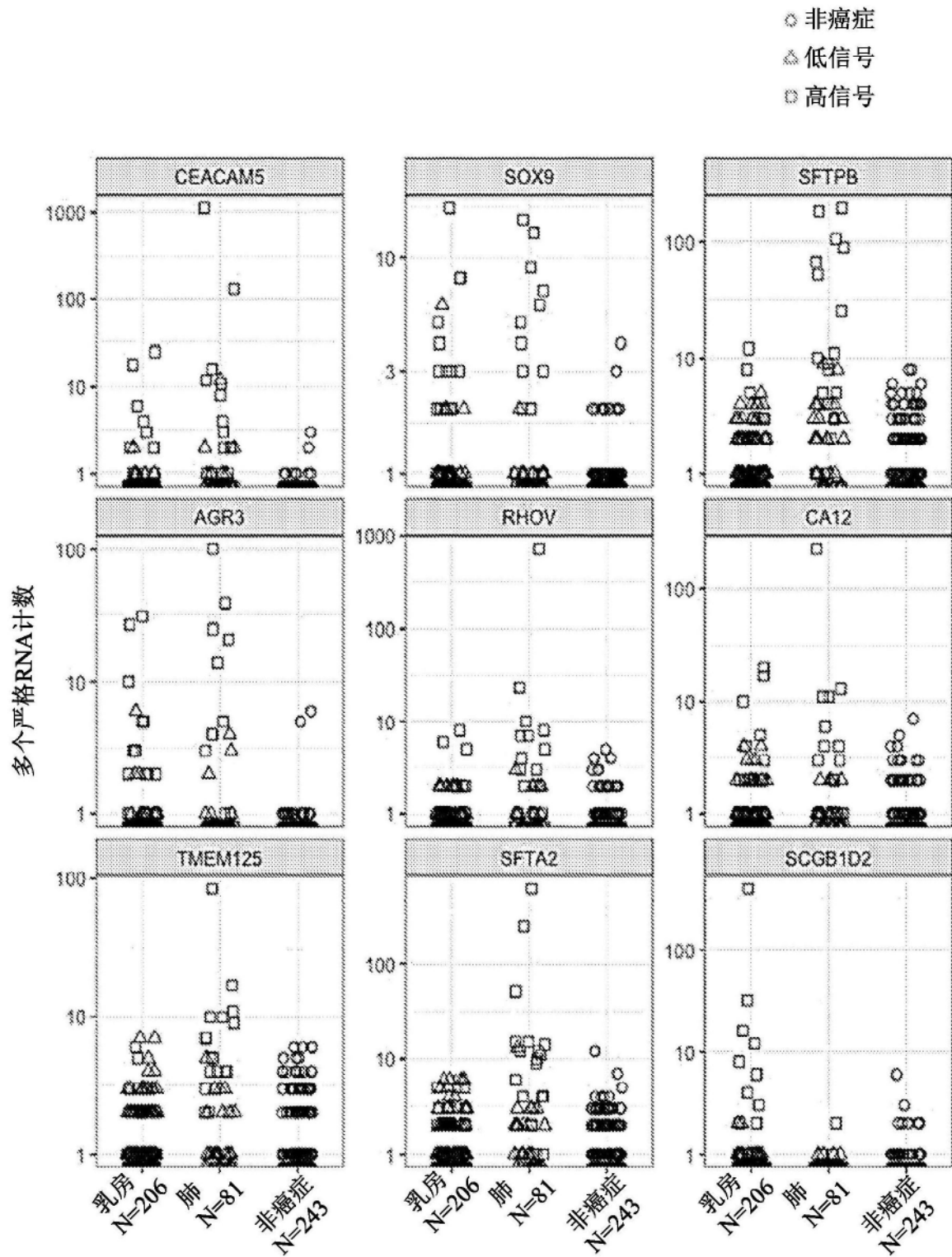


图26C

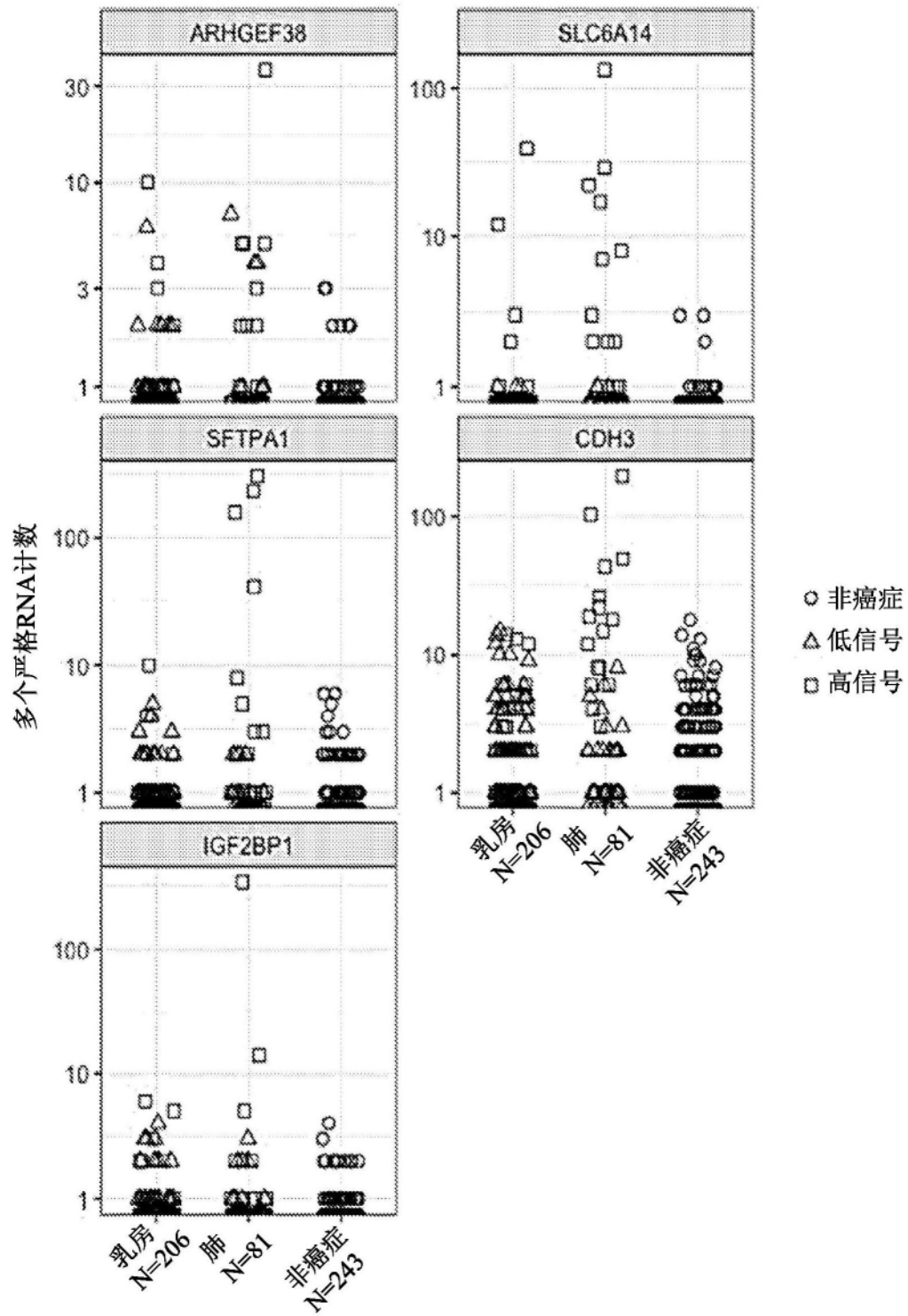


图26D