



(12)发明专利申请

(10)申请公布号 CN 110198711 A

(43)申请公布日 2019.09.03

(21)申请号 201780084170.4

(22)申请日 2017.11.22

(30)优先权数据

62/425,549 2016.11.22 US

(85)PCT国际申请进入国家阶段日

2019.07.19

(86)PCT国际申请的申请数据

PCT/US2017/063157 2017.11.22

(87)PCT国际申请的公布数据

W02018/098379 EN 2018.05.31

(71)申请人 第一基因股份有限公司

地址 美国加利福尼亚州

(72)发明人 罗伯特·费尔德曼

梅勒妮·马赫塔尼

(74)专利代理机构 北京安信方达知识产权代理有限公司 11262

代理人 刘晓杰

(51)Int.Cl.

A61K 31/337(2006.01)

A61K 31/704(2006.01)

A61K 45/00(2006.01)

C12N 5/09(2010.01)

C12Q 1/6886(2018.01)

C40B 30/04(2006.01)

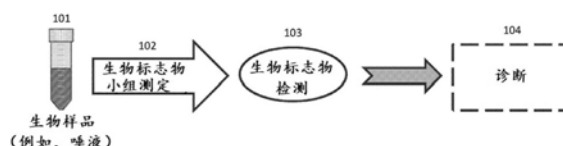
权利要求书3页 说明书45页 附图29页

(54)发明名称

癌症检测方法

(57)摘要

本公开内容提供了用于癌症检测的方法。该方法可包括对来自受试者的生物标志物的非侵入性检测。该方法可以与其他筛查方法结合使用,以使检测的准确度更高。



1. 一种用于确定受试者的健康状态的方法,所述方法包括:
 - a) 提供来自受试者的唾液样品;
 - b) 量化所述唾液样品中生物标志物的样品水平,其中所述生物标志物来自所述唾液样品中的外来体;
 - c) 将所述生物标志物的所述样品水平与所述生物标志物的参考水平进行比较,其中所述参考水平获自患有乳腺癌的受试者;以及
 - d) 基于所述比较,确定所述受试者的乳腺癌风险评分。
2. 如权利要求1所述的方法,进一步包括对所述受试者的乳腺组织进行成像。
3. 如权利要求2所述的方法,其中使用乳房X线照片进行所述成像。
4. 如权利要求1所述的方法,进一步包括基于所述乳房X线照片的结果调整所述受试者的所述风险评分。
5. 如权利要求1所述的方法,进一步包括在步骤b)之前裂解所述外来体以释放所述生物标志物。
6. 如权利要求4所述的方法,进一步包括在所述裂解之前富集所述唾液样品的外来体部分。
7. 如权利要求1所述的方法,进一步包括在所述富集之后稳定所述外来体部分。
8. 如权利要求1所述的方法,其中所述生物标志物是无细胞核酸。
9. 如权利要求8所述的方法,其中所述无细胞核酸是RNA。
10. 如权利要求9所述的方法,其中所述RNA是mRNA或miRNA。
11. 如权利要求10所述的方法,其中所述mRNA是选自LCE2B、HIST1H4K、ABCA1、ABCA2、TNFRSF10A、AK092120、DTYMK、ALKBH1、MCART1、Hs.161434及其任何组合的基因的转录物。
12. 如权利要求9所述的方法,其中所述量化进一步包括将所述RNA逆转录。
13. 如权利要求1所述的方法,其中所述量化包括进行聚合酶链反应(PCR)。
14. 如权利要求13所述的方法,其中所述PCR包括qPCR。
15. 如权利要求1所述的方法,其中所述量化进一步包括进行测序。
16. 如权利要求15所述的方法,其中所述测序包括大规模并行测序。
17. 如权利要求1所述的方法,其中所述确定所述受试者的乳腺癌风险评分以至少90%的准确度进行。
18. 如权利要求1所述的方法,其中所述确定所述受试者的乳腺癌风险评分以至少90%的特异性进行。
19. 如权利要求1所述的方法,其中所述确定所述受试者的乳腺癌风险评分以至少80%的灵敏度进行。
20. 如权利要求1所述的方法,其中所述外来体的来源细胞是乳腺细胞。
21. 如权利要求1所述的方法,其中所述受试者具有致密的乳腺组织。
22. 如权利要求1所述的方法,其中所述受试者的筛查性乳房X线照片的结果不明确。
23. 如权利要求1所述的方法,其中所述受试者的年龄小于50岁。
24. 如权利要求1所述的方法,其中所述生物标志物是与癌症的标志相关的基因的转录物。
25. 如权利要求24所述的方法,其中癌症的所述标志选自:逃避生长抑制因子、避免免

疫破坏、促进复制永生、肿瘤促进性炎症、激活侵袭和转移、诱导血管生成、基因组不稳定性、和突变、抵抗细胞死亡、解除细胞能力学管制、维持增殖信号传导及其任何组合。

26. 如权利要求24所述的方法, 其中与癌症的所述标志相关的所述基因选自: LCE2B、HIST1H4K、ABCA2、TNFRSF10A、AK092120、DTYMK、ALKBH1、MCART1、Hs. 161434及其任何组合。

27. 如权利要求24所述的方法, 其中与癌症的所述标志相关的所述基因选自: ABCA1、ABCA2、TNFRSF10A、DTYMK、ALKBH1及其任何组合。

28. 如权利要求1所述的方法, 其中所述生物标志物是基因的转录物, 所述基因的表达谱类似于与癌症的标志相关的基因。

29. 一种用于减少乳腺癌的假阳性或假阴性结果的数目的方法, 所述方法包括:

- a) 提供受试者的生物样品, 其中所述受试者来自筛查性乳房X线照片结果为阳性、阴性或不明确的受试者群体;
- b) 量化所述受试者的所述生物样品中生物标志物的样品水平;
- c) 将所述生物标志物的所述样品水平与所述生物标志物的参考水平进行比较; 以及
- d) 基于所述比较的结果, 将所述筛查性乳房X线照片的结果鉴定为乳腺癌假阳性或假阴性。

30. 一种用于确定受试者的健康状态的方法, 所述方法包括:

- a) 提供受试者的生物样品;
- b) 量化所述受试者的所述生物样品中至少两种生物标志物的样品水平, 其中所述至少两种生物标志物选自LCE2B、HIST1H4K、ABCA2、TNFRSF10A、AK092120、DTYMK、ALKBH1、MCART1、Hs. 161434及其任何组合;
- c) 将所述至少两种生物标志物的所述样品水平与所述两种生物标志物的参考水平进行比较; 以及
- d) 基于所述比较, 确定所述受试者的健康状态。

31. 一种用于确定受试者的健康状态的方法, 所述方法包括:

- a) 对受试者进行乳房X线照片;
- b) 获得所述受试者的唾液样品;
- c) 量化所述唾液样品中生物标志物的样品水平, 其中所述生物标志物源自外来体, 其中所述生物标志物是选自以下的基因的转录物: LCE2B、HIST1H4K、ABCA2、TNFRSF10A、AK092120、DTYMK、ALKBH1、MCART1、Hs. 161434及其任何组合;
- d) 将所述生物标志物的所述样品水平与所述生物标志物的参考水平进行比较, 其中所述参考水平获自患有乳腺癌的受试者; 以及
- e) 将所述乳房X线照片的结果与所述比较结合, 以确定所述受试者与乳腺癌相关的健康状态。

32. 一种方法, 其包括:

- a) 提供来自受试者的唾液样品;
- b) 量化所述唾液样品中生物标志物的样品水平, 其中所述生物标志物是与癌症的标志相关的基因的转录物;
- c) 将所述生物标志物的所述样品水平与所述生物标志物的参考水平进行比较, 其中所述参考水平获自患有癌症的受试者; 以及

d) 基于所述比较,确定所述受试者的癌症风险评分。

33. 如权利要求32所述的方法,其中癌症的所述标志选自:逃避生长抑制因子、避免免疫破坏、促进复制永生、肿瘤促进性炎症、激活侵袭和转移、诱导血管生成、基因组不稳定性、和突变、抵抗细胞死亡、解除细胞能力学管制、维持增殖信号传导及其任何组合。

癌症检测方法

交叉引用

[0001] 本申请要求于2016年11月22日提交的美国临时申请号62/425,549的权益,其通过引用全文并入于此。

背景技术

[0002] 癌症是一种影响全球数百万人的普遍疾病。在2016年,仅在美国,估计就将有1,685,210例新发癌症被确诊,并且595,690人将死于该疾病。到2020年,将有1820万美国人,大约每19人中就有1人成为癌症患者或癌症幸存者,从2005年的1170万人(每26人中有1人)上升。

[0003] 在美国,大约八分之一的女性将在其一生中患上浸润性乳腺癌。在2012年,乳腺癌占有癌症诊断的近25%。预计在2017年,在美国女性中将诊断出估计252,710例新发浸润性乳腺癌,以及估计63,410例新发非浸润性乳腺癌。预计在2017年,在男性中将诊断出约2,470例新发浸润性乳腺癌。如果癌症诊断发生在早期,可以提高生存率。

援引并入

[0004] 本文中的所有出版物、专利和专利申请均通过引用而并入,其程度如同特别地且单独地指出每个单独的出版物、专利或专利申请通过引用而并入。

发明内容

[0005] 应理解,本发明的不同方面可以单独地、共同地或彼此组合地理解。本文所述的发明的各个方面均可应用于下文阐述的任何特定应用或方法。

[0006] 在一方面,本公开内容提供了一种用于确定受试者的健康状态的方法。该方法可包括:a)提供来自受试者的唾液样品;b)量化唾液样品中生物标志物的样品水平,其中生物标志物来自唾液样品中的外来体;c)将生物标志物的样品水平与该生物标志物的参考水平进行比较,其中参考水平获自患有乳腺癌的受试者;以及d)基于比较,确定受试者的乳腺癌风险评分。在一些实施方案中,该方法进一步包括对受试者的乳腺组织进行成像。在一些实施方案中,使用乳房X线照片(mammogram)进行成像。在一些实施方案中,该方法进一步包括基于乳房X线照片的结果调整来自步骤e的受试者的风险评分。在一些实施方案中,该方法进一步包括在步骤b)之前裂解外来体以释放生物标志物。在一些实施方案中,该方法进一步包括在裂解之前富集唾液样品的外来体部分。在一些实施方案中,该方法进一步包括在富集之后稳定外来体部分。在一些实施方案中,生物标志物是无细胞核酸。在一些实施方案中,无细胞核酸是RNA。在一些实施方案中,RNA是mRNA或miRNA。在一些实施方案中,mRNA是选自LCE2B、HIST1H4K、ABCA1、ABCA2、TNFRSF10A、AK092120、DTYMK、ALKBH1、MCART1、Hs.161434及其任何组合的基因的转录物。在一些实施方案中,量化进一步包括将RNA逆转录。在一些实施方案中,量化进一步包括进行聚合酶链反应(PCR)。在一些实施方案中,PCR包括qPCR。在一些实施方案中,量化进一步包括进行测序。在一些实施方案中,测序包括大规模并行测序。在一些实施方案中,确定受试者的乳腺癌风险评分以至少90%的准确度进

行。在一些实施方案中,确定受试者的乳腺癌风险评分以至少90%的特异性进行。在一些实施方案中,确定受试者的乳腺癌风险评分以至少80%的灵敏度进行。在一些实施方案中,外来体的来源细胞是乳腺细胞。在一些实施方案中,受试者具有致密的乳腺组织。在一些实施方案中,受试者的筛查性乳房X线照片的结果不明确。在一些实施方案中,受试者的年龄范围为18至40岁。在一些实施方案中,生物标志物是与癌症的标志相关的基因的转录物。在一些实施方案中,癌症的标志选自:逃避生长抑制因子、避免免疫破坏、促进复制永生、肿瘤促进性炎症、激活侵袭和转移、诱导血管生成、基因组不稳定性突变、抵抗细胞死亡、解除细胞能力学管制、维持增殖信号传导及其任何组合。在一些实施方案中,与癌症的标志相关的基因选自:LCE2B、HIST1H4K、ABCA2、TNFRSF10A、AK092120、DTYMK、ALKBH1、MCART1、Hs.161434及其任何组合。在一些实施方案中,与癌症标志相关的基因选自:ABCA1、ABCA2、TNFRSF10A、DTYMK、ALKBH1及其任何组合。在一些实施方案中,生物标志物是基因的转录物,该基因的表达谱类似于与癌症标志相关的基因。

[0007] 在一方面,本公开内容提供了一种用于减少乳腺癌的假阳性或假阴性结果的数目的方法。该方法可包括a) 提供受试者的生物样品,其中受试者来自筛查性乳房X线照片结果为阳性、阴性或不明确的受试者群体;b) 量化受试者的生物样品中生物标志物的样品水平;c) 将生物标志物的样品水平与生物标志物的参考水平进行比较;以及d) 基于比较结果,将筛查性乳房X线照片的结果鉴定为乳腺癌假阳性或假阴性。在一些实施方案中,生物标志物是无细胞核酸。在一些实施方案中,无细胞核酸是RNA。在一些实施方案中,RNA是mRNA或miRNA。在一些实施方案中,mRNA是选自LCE2B、HIST1H4K、ABCA2、TNFRSF10A、AK092120、DTYMK、ALKBH1、MCART1、Hs.161434及其任何组合的基因的转录物。在一些实施方案中,生物标志物源自外来体。在一些实施方案中,该方法进一步包括在步骤b) 之前裂解生物样品的外来体部分以释放生物标志物。在一些实施方案中,该方法进一步包括在裂解之前富集生物样品的外来体部分。在一些实施方案中,该方法进一步包括在富集之后稳定外来体部分。在一些实施方案中,生物样品是唾液。在一些实施方案中,以至少90%的准确度进行鉴定。在一些实施方案中,以至少90%的特异性进行鉴定。在一些实施方案中,以至少80%的灵敏度进行鉴定。在一些实施方案中,外来体的来源细胞是乳腺细胞。在一些实施方案中,受试者具有致密的乳腺组织。在一些实施方案中,受试者的乳房X线照片结果不明确。在一些实施方案中,生物标志物是与癌症的标志相关的基因的转录物。在一些实施方案中,癌症的标志可以选自:逃避生长抑制因子、避免免疫破坏、促进复制永生、肿瘤促进性炎症、激活侵袭和转移、诱导血管生成、基因组不稳定性突变、抵抗细胞死亡、解除细胞能力学管制、维持增殖信号传导及其任何组合。在一些实施方案中,与癌症标志相关的基因选自:LCE2B、HIST1H4K、ABCA2、TNFRSF10A、AK092120、DTYMK、ALKBH1、MCART1、Hs.161434及其任何组合。在一些实施方案中,与癌症的标志相关的基因选自:ABCA1、ABCA2、TNFRSF10A、DTYMK、ALKBH1及其任何组合。在一些实施方案中,生物标志物是基因的转录物,该基因的表达谱类似于与癌症标志相关的基因。

[0008] 在一方面,本公开内容提供了一种用于确定受试者的健康状态的方法。该方法可包括a) 提供受试者的生物样品;b) 量化受试者的生物样品中至少两种生物标志物的样品水平,其中该至少两种生物标志物选自LCE2B、HIST1H4K、ABCA2、TNFRSF10A、AK092120、DTYMK、ALKBH1、MCART1、Hs.161434及其任何组合;c) 将至少两种生物标志物的样品水平与两种生

物标志物的参考水平进行比较;以及d)基于比较,确定受试者的健康状态。在一些实施方案中,生物样品是生物流体。在一些实施方案中,生物流体是唾液。在一些实施方案中,至少2种生物标志物之一是HIST1H4K。在一些实施方案中,至少2种生物标志物之一是TNFRSF10A。在一些实施方案中,至少2种生物标志物之一是ALKBH1。在一些实施方案中,至少2种生物标志物之一是ABCA2。在一些实施方案中,至少2种生物标志物之一是DTYMK。在一些实施方案中,量化包括对至少九种生物标志物的样品水平进行量化。在一些实施方案中,该九种生物标志物是LCE2B、HIST1H4K、ABCA2、TNFRSF10A、AK092120、DTYMK、ALKBH1、MCART1和Hs.16143。在一些实施方案中,量化包括对至少两种生物标志物的mRNA转录物进行量化。在一些实施方案中,该方法进一步包括裂解生物样品的外来体部分以释放mRNA。在一些实施方案中,以至少90%的准确度进行生物标志物的样品水平的量化。在一些实施方案中,以至少约80%的灵敏度进行生物标志物的样品水平的量化。在一些实施方案中,以至少90%的特异性进行生物标志物的样品水平的量化。在一些实施方案中,至少2种生物标志物与癌症的标志相关。在一些实施方案中,癌症的标志选自:逃避生长抑制因子、避免免疫破坏、促进复制永生、肿瘤促进性炎症、激活侵袭和转移、诱导血管生成、基因组不稳定性突变、抵抗细胞死亡、解除细胞能力学管制、维持增殖信号传导及其任何组合。在一些实施方案中,该至少两种生物标志物包含类似于与癌症标志相关的基因的表达谱。

[0009] 在一方面,本公开内容提供了一种确定受试者的健康状态的方法。该方法可包括a)对受试者进行乳房X线照片;b)获得受试者的唾液样品;c)量化来自唾液样品的生物标志物的样品水平,其中生物标志物源自外来体,其中生物标志物是选自以下的基因的转录物:LCE2B、HIST1H4K、ABCA2、TNFRSF10A、AK092120、DTYMK、ALKBH1、MCART1、Hs.161434及其任何组合;d)将生物标志物的样品水平与生物标志物的参考水平进行比较,其中参考水平获自患有乳腺癌的受试者;以及e)将乳房X线照片的结果与比较结合,以确定受试者与乳腺癌相关的健康状态。在一些实施方案中,与缺乏步骤e)的组合步骤的方法相比,该方法对于确定受试者与乳腺癌相关的健康状态具有更高的准确度。在一些实施方案中,受试者具有致密的乳腺组织。在一些实施方案中,乳房X线照片为受试者给出的结果不明确。在一些实施方案中,受试者的年龄范围为18至40岁。在一些实施方案中,转录物是mRNA或miRNA。在一些实施方案中,量化包括测序。

[0010] 在一方面,本公开内容提供了一种方法,其包括:a)提供来自受试者的唾液样品;b)量化唾液样品中生物标志物的样品水平,其中生物标志物是与癌症的标志相关的基因的转录物;c)将生物标志物的样品水平与生物标志物的参考水平进行比较,其中参考水平获自患有癌症的受试者;以及d)基于比较,确定受试者的癌症风险评分。在一些实施方案中,癌症的标志选自:逃避生长抑制因子、避免免疫破坏、促进复制永生、肿瘤促进性炎症、激活侵袭和转移、诱导血管生成、基因组不稳定性突变、抵抗细胞死亡、解除细胞能力学管制、维持增殖信号传导及其任何组合。在一些实施方案中,与癌症的标志相关的基因选自:LCE2B、HIST1H4K、ABCA2、TNFRSF10A、AK092120、DTYMK、ALKBH1、MCART1、Hs.161434及其任何组合。在一些实施方案中,与癌症的标志相关的基因选自:ABCA1、ABCA2、TNFRSF10A、DTYMK、ALKBH1及其任何组合。在一些实施方案中,生物标志物获自唾液中的外来体。在一些实施方案中,该方法进一步包括在步骤b之前裂解外来体以从外来体部分释放生物标志物。在一些实施方案中,外来体的来源细胞是乳腺细胞。在一些实施方案中,转录物是RNA。在一些实施

方案中, RNA是mRNA或miRNA。在一些实施方案中, 量化进一步包括将RNA逆转录。在一些实施方案中, 量化进一步包括进行聚合酶链反应(PCR)。在一些实施方案中, PCR是qPCR。在一些实施方案中, 量化进一步包括进行测序。在一些实施方案中, 测序包括大规模并行测序。在一些实施方案中, 确定受试者的癌症风险评分以至少90%的准确度进行。在一些实施方案中, 确定受试者的癌症风险评分以至少90%的特异性进行。在一些实施方案中, 确定受试者的癌症风险评分以至少80%的灵敏度进行。在一些实施方案中, 癌症是乳腺癌。在一些实施方案中, 受试者具有致密的乳腺组织。在一些实施方案中, 受试者的筛查性乳房X线照片结果不明确。在一些实施方案中, 受试者的年龄范围为18至40岁。在一些实施方案中, 该方法进一步包括对受试者的乳腺组织进行成像。在一些实施方案中, 使用乳房X线照片进行成像。在一些实施方案中, 该方法进一步包括基于乳房X线照片的结果调整来自步骤d的受试者风险评分。

附图说明

[0011] 本公开内容的新颖特征在所附权利要求书中具体阐述。通过参考以下对可以利用本公开内容原理的说明性实施方案加以阐述的详细描述及其附图, 将会获得对本公开内容的特征和优点的更好理解, 在附图中:

[0012] 图1示出了将受试者的生物样品(例如, 体液, 如唾液)用于本公开内容的生物标志物测定, 以检测与健康状况(例如, 癌症、乳腺癌)相关的生物标志物。来自生物标志物测定的数据可用于确定受试者的健康状况。

[0013] 图2示出了将本公开内容的生物标志物小组与成像数据(例如, 乳房X线照片)组合使用以用于受试者的癌症(例如, 乳腺癌)检测。将测定与成像数据组合使用可以提供更高的检测准确度。

[0014] 图3描绘了使用唾液样品评估受试者癌症的本公开内容方法的说明性工作流程。

[0015] 图4示出了可作为本公开内容的生物标志物小组的部分的候选基因。

[0016] 图5为框图, 示出了计算机架构系统的示例。

[0017] 图6为示出了计算机网络的图, 该计算机网络具有多个计算机系统、多个蜂窝电话和个人数据助理, 以及NAS设备。

[0018] 图7为使用共享虚拟地址存储空间的多处理器计算机系统的框图。

[0019] 图8示出了从地理位置传输给用户的计算机程序产品。

[0020] 图9示出了用于鉴定乳腺癌生物标志物的研究的结果。显示了来自10名乳腺癌受试者和10名匹配的健康对照的平均连通性值。

[0021] 图10示出了从使用qPCR的9基因测定获得的评分, 取自60位受试者的验证研究。

[0022] 图11示出了连续排序的综合基因表达值。

[0023] 图12示出了生物标志物基因5的二次验证研究结果。

[0024] 图13A、13B、13C和13D显示了候选生物标志物基因的基于RT-qPCR的二次验证研究结果。图13A显示了基因2的基于RT-qPCR的二次验证研究结果。图13B显示了基因3的基于RT-qPCR的二次验证研究结果。图13C显示了基因7的基于RT-qPCR的二次验证研究结果。图13D显示了基因9的基于RT-qPCR的二次验证研究结果。

[0025] 图14显示了对基因2的生物标志物验证研究的参数和结果。

- [0026] 图15显示了对基因3的生物标志物验证研究的参数和结果。
- [0027] 图16显示了对基因7的生物标志物验证研究的参数和结果。
- [0028] 图17显示了对基因9的生物标志物验证研究的参数和结果。
- [0029] 图18A、18B、18C、18D和18E示出了候选生物标志物基因的基于RT-qPCR的二次验证研究结果。图18A显示了基因1的基于RT-qPCR的二次验证研究结果。图18B显示了基因4的基于RT-qPCR的二次验证研究结果。图18C显示了基因5的基于RT-qPCR的二次验证研究结果。图18D显示了基因6的基于RT-qPCR的二次验证研究结果。图18E显示了基因8的基于RT-qPCR的二次验证研究结果。
- [0030] 图19显示了对基因1的生物标志物验证研究的参数和结果。
- [0031] 图20显示了对基因4的生物标志物验证研究的参数和结果。
- [0032] 图21显示了对基因5的生物标志物验证研究的参数和结果。
- [0033] 图22显示了对基因6的生物标志物验证研究的参数和结果。
- [0034] 图23显示了对基因8的生物标志物验证研究的参数和结果。
- [0035] 图24显示了持家基因G-H1的基于RT-qPCR的二次验证研究结果。
- [0036] 图25显示了持家基因G-H2的基于RT-qPCR的二次验证研究结果。
- [0037] 图26示出了唾液生物标志物测试的优化工作流程的示例。
- [0038] 图27描绘了与癌症的标志相关的说明性基因和信号传导系统。
- [0039] 图28描绘了使用本公开内容的方法鉴定的与癌症的一种或多种标志相关的说明性生物标志物。
- [0040] 图29示出了针对乳腺癌基因评估唾液中的基因表达谱的研究的结果。

具体实施方式

[0041] 以下描述和示例详细说明了本公开内容的实施方案。应理解,本公开内容不限于本文所述的特定实施方案,因此可以变化。本领域技术人员将认识到,本公开内容存在许多变化和修改,这些变化和修改均包含在其范围内。

[0042] 诸如乳房X线照片等成像测试可用于筛查和检测乳腺疾病,如乳腺癌(例如,浸润性乳腺癌和原位导管癌)。然而,筛查性乳房X线照片可能无法检测约五分之一的乳腺癌。例如,乳房X线照片的假阳性和假阴性率可以在约7-15%的范围内。假阳性和假阴性率在较年轻的女性(例如,50岁以下的女性)和乳腺致密的女性中可能更高。此外,基于乳房X线照片可能难以区分浸润性乳腺癌与非危及生命的乳腺癌形式。因此,乳房X线照片可能导致患者的过度诊断,者可能导致对非侵袭性癌症的过度治疗。因此,非常需要更准确的乳腺癌检测方法。

[0043] 本文公开了用于检测受试者的癌症的方法。示例性方法可包括以下步骤:(a)获得受试者的生物样品,(b)量化生物样品中生物标志物的样品水平,(c)将生物标志物的样品水平与生物标志物的参考水平进行比较,(d)基于样品水平与参考水平之间的比较确定受试者的癌症风险评分,或其任何组合。例如,生物样品可以是唾液。例如,癌症可以是乳腺癌。例如,生物标志物可以源自外来体。该方法可以另外包括裂解、分离或富集生物样品的特定部分(例如,生物样品中的外来体)的步骤。

[0044] 本公开内容的示例性方法如图1所示。图1示出了将来自受试者的唾液样品用于检

测与例如癌症相关的一种或多种生物标志物。从受试者收集唾液样品(101)。然后处理唾液样品,并使其经受本公开内容的生物标志物小组测定(panel assay,102)以检测生物标志物(103)。使用生物标志物测定的结果来确定受试者是否患有癌症。给予受试者诊断(104)。

[0045] 本公开内容的方法可以与额外的筛查或检测方法组合使用。例如,与单独使用筛查测试相比,本公开内容的生物标志物测定与额外的筛查测试的组合可以提供更高的癌症检测准确度、灵敏度和/或特异性。示例性方法可包括以下步骤:a)对受试者进行筛查测试,以评估受试者出现健康状况的风险,b)获得受试者的生物样品,c)量化受试者的生物样品中生物标志物的样品水平,d)将生物标志物的样品水平与生物标志物的参考水平进行比较,e)将筛查测试的结果与生物标志物比较结合,f)基于筛查测试和生物标志物结果的综合信息确定受试者的健康状态,或其任何组合。额外的筛查测试可包括,例如,对受试者组织或器官的成像测试(例如,使用x射线、声波、放射性粒子,或磁场)。例如,组织可以是乳腺组织。例如,额外的筛查测试可以是乳房X线照片。例如,生物样品可以是唾液。例如,癌症可以是乳腺癌。例如,生物标志物可以源自外来体。例如,生物标志物可以是mRNA。

[0046] 本公开内容的示例性方法如图2所示。图2示出了将本公开内容的基于唾液的生物标志物测定与额外的筛查测试(例如,乳房X线照片)结合使用以用于检测乳腺癌。受试者(201)经历成像测试,如乳房X线照片(202)。受试者还提供用于生物标志物小组测定的样品,如唾液(204)。获得并处理成像数据(203)。使唾液样品经受生物标志物小组测定,以检测生物标志物(205)。成像数据(203)与生物标志物测定结果(205)组合用于诊断受试者的乳腺癌(206)。

[0047] 本文公开了用于减少健康状况的假阳性或假阴性结果的数目的方法。示例性方法可包括以下步骤:a)获得筛查测试的结果为阳性、阴性或不明确的受试者的生物样品,该筛查测试评估受试者出现健康状况的风险,b)量化受试者的生物样品中生物标志物的样品水平,c)将生物标志物的样品水平与该健康状况的生物标志物的参考水平进行比较,d)基于生物标志物比较的结果,将筛查测试的结果鉴定为该健康状况的假阳性或假阴性。筛查测试可包括,例如,对受试者组织或器官的成像测试(例如,使用x射线、声波、放射性粒子,或磁场)。例如,组织可以是乳腺组织。例如,筛查测试可以是乳房X线照片。例如,生物样品可以是唾液。例如,健康状况可以是乳腺癌。例如,生物标志物可以源自外来体。例如,生物标志物可以是mRNA。

[0048] 本公开内容的方法可以提供,例如,用于癌症早期检测的低成本、准确、非侵入性且易于实施的测试。本公开内容的方法可以有助于癌症的早期检测。本公开内容的方法可用于具有致密乳腺组织的受试者。本公开内容的方法可以降低假阳性率和假阴性率,并改善癌症诊断的准确度。在一些实施方案中,本公开内容提供了基于唾液的测试,其包括测量来自受试者唾液样品的源自外来体的mRNA,以确定受试者的乳腺癌风险。

[0049] 在一些实施方案中,本公开内容提供了用于执行本公开内容的方法的设备。该设备可用于分析样品,例如,以生成受试者的生物标志物特征。在一些实施方案中,该设备可用于诊所、医院或乳腺成像中心。

[0050] 本公开内容的方面可以涉及能够改善对有健康状况或疾病的受试者的监测、诊断和/或治疗的方法。健康状况可以是,例如,癌症、神经退行性疾病、炎症病症或药物反应病症。

[0051] 癌症的非限制性示例包括：急性成淋巴细胞性白血病、急性髓样白血病、肾上腺皮质癌、AIDS相关癌症、AIDS相关淋巴瘤、肛门癌、阑尾癌、星形细胞瘤、神经母细胞瘤、基底细胞癌、胆管癌、膀胱癌、骨癌、脑肿瘤（如小脑星形细胞瘤、大脑星形细胞瘤/恶性胶质瘤、室管膜瘤、成神经管细胞瘤、幕上原始神经外胚瘤（supratentorial primitive neuroectodermal tumor）、视通路和下丘脑神经胶质瘤）、乳腺癌、支气管腺瘤、伯基特淋巴瘤（Burkitt lymphoma）、未知原始起源的癌、中枢神经系统淋巴瘤、小脑星形细胞瘤、宫颈癌、儿童期癌症、慢性淋巴细胞白血病、慢性髓性白血病、慢性骨髓增生性疾病、结肠癌、皮肤T细胞淋巴瘤、结缔组织增生性小圆细胞肿瘤、子宫内膜癌、室管膜瘤、食管癌、尤因肉瘤（Ewing's sarcoma）、生殖细胞肿瘤、胆囊癌、胃癌、胃肠道类癌瘤、胃肠道间质瘤、神经胶质瘤、毛细胞白血病、头颈癌、心脏癌、肝细胞（肝）癌、霍奇金淋巴瘤、下咽癌、眼内黑素瘤、胰岛细胞癌、卡波西肉瘤、肾癌、喉癌、唇和口腔癌、脂肪肉瘤、肝癌、肺癌（例如非小细胞和小细胞肺癌）、淋巴瘤、白血病、巨球蛋白血症、骨恶性纤维组织细胞瘤/骨肉瘤、髓母细胞瘤、黑素瘤、间皮瘤、潜伏原发转移性鳞状颈癌、口腔癌、多发性内分泌肿瘤综合征、骨髓增生异常综合征、髓样白血病、鼻腔和副鼻窦癌、鼻咽癌、神经母细胞瘤、非霍奇金淋巴瘤、非小细胞肺癌、口癌（oral cancer）、口咽癌、骨肉瘤/骨恶性纤维组织细胞瘤、卵巢癌、卵巢上皮癌、卵巢生殖细胞瘤、胰腺癌、胰岛细胞胰腺癌、副鼻窦和鼻腔癌、甲状旁腺癌、阴茎癌、咽癌（pharyngeal cancer）、嗜铬细胞瘤、松果体星形细胞瘤、松果体生殖细胞瘤、垂体腺瘤、胸膜肺母细胞瘤、浆细胞瘤、原发性中枢神经系统淋巴瘤、前列腺癌、直肠癌、肾细胞癌、肾盂和输尿管移行细胞癌、视网膜母细胞瘤、横纹肌肉瘤、唾液腺癌、肉瘤、皮肤癌、梅克尔细胞皮肤癌（skin carcinoma merkel cell）、小肠癌、软组织肉瘤、鳞状细胞癌、胃癌、T细胞淋巴瘤、喉癌、胸腺瘤、胸腺癌、甲状腺癌、滋养细胞肿瘤（妊娠期）、未知原发部位的癌症、尿道癌、子宫肉瘤、阴道癌、外阴癌、沃尔登斯特伦巨球蛋白血症（Waldenström macroglobulinemia）以及维尔姆斯瘤（Wilms tumor）。在一些实施方案中，健康状况是癌症。在一些实施方案中，健康状况是乳腺癌。

[0052] 本公开内容的方法可包括检测生物标志物的存在。生物标志物可以是健康状况（例如，癌症）的可测量指标。生物标志物可以由肿瘤分泌，或由于例如对癌症存在的生理反应而产生。例如，生物标志物可以是遗传的、表观遗传的、蛋白质组学的、糖组学的或成像的生物标志物。生物标志物可用于诊断、预后或流行病学。可以在侵入式收集的样品（如组织活检物）中测定生物标志物。生物标志物可以在非侵入式收集的样品（如体液，例如，唾液）中测定。

[0053] 多种生物标志物适合与本公开内容的方法一起使用。生物标志物可以是，例如，核酸如DNA或RNA、肽、蛋白质、脂质、抗原、抗体、碳水化合物、蛋白多糖，或其任何组合。生物标志物可以是无细胞核酸，如无细胞DNA或无细胞RNA。生物标志物可以是选自以下的RNA：mRNA、小RNA、miRNA、snoRNA、snRNA、rRNA、tRNA、siRNA、hnRNA、shRNA及其组合。在一些实施方案中，生物标志物可以是RNA。在一些实施方案中，生物标志物可以是mRNA。在一些实施方案中，生物标志物可以是miRNA。

[0054] 生物标志物可以是基因的产物（例如，表达产物）。生物标志物可以测量基因的活性。可以在转录组水平（例如，RNA、mRNA、miRNA）、蛋白质组水平（例如，蛋白质、多肽）或其组合下测量生物标志物基因的表达。

[0055] 生物标志物基因可以差异表达(例如,过度表达或表达不足),例如,与健康状况的参考水平或对照相比。例如,与健康状况的参考水平相比,生物标志物的表达水平可以变化至少2倍、3倍、4倍、5倍、6倍、7倍、10倍、15倍、20倍、50倍或100倍。在一些实施方案中,基因表达水平的差异为至少10%、15%、20%、25%、30%、35%、40%、45%或50%或更多。可以从一个或多个受试者获得参考水平。本公开内容的方法可包括确定生物标志物基因与对照相比的差异表达。

[0056] 本公开内容的方法可包括检测一种以上的生物标志物。例如,方法可以评估至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90、100、150、200、300、400、500、600、700、800、900或1000种生物标志物。例如,方法可以评估1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90、100、150、200、300、400、500、600、700、800、900或1000种生物标志物。方法可包括检测至少2种生物标志物。方法可包括检测至少3种生物标志物。方法可包括检测至少4种生物标志物。方法可包括检测至少5种生物标志物。方法可包括检测至少6种生物标志物。方法可包括检测至少7种生物标志物。方法可包括检测至少8种生物标志物。方法可包括检测至少9种生物标志物。

[0057] 生物标志物的检测或分析可包括确定生物标志物的表达水平、存在、缺失、突变、拷贝数变异、截短、复制、插入、修饰、序列变化、分子缔合或其组合。

[0058] 在一些实施方案中,可以分析基因共表达网络以发现生物标志物。也参见例如,美国专利公开20120010823,其通过引用整体并入于此用于所有目的。基因共表达网络的分析可以基于细胞对变化的条件的转录反应。由于基因的协调共表达可以编码相互作用的蛋白质,因此研究共表达模式可以提供对潜在细胞过程的深入了解。

[0059] 可以在Pearson相关系数上设置阈值以达成基因共表达网络,其可以被称为“相关性”网络。在这些网络中,节点可以对应于给定基因的基因表达谱。例如,如果节点具有显著的成对表达谱,则可以连接节点。在一些实施方案中,Pearson相关的绝对值可以用作基因表达聚类分析的标准。在一些实施方案中,Pearson相关系数可以用作共表达量度。

[0060] 本公开内容的方法可包括基因表达模块的分析。聚类程序可用于鉴定在其基因表达值之间具有高相关性(例如,大于0.95)的相连节点的模块。然后可以分析这些模块之间的平均连通性。平均连通性可以是跨越所有模块的 k_i 的平均值。模块 i 的连通性可以定义为以例如大于约0.95的相关性与模块 i 相关联的 k_i 模块:

$$k_i = \sum_{i \neq j} a_{ij}$$

$$i \neq j$$

其中 a_{ij} 可以是与第 i 个模块具有大于0.95的相关性的模块。

[0061] 在一些实施方案中,可以对基因表达值进行加权。在一些实施方案中,可以将新的和/或额外的基因添加至生物标志物小组。在一些实施方案中,基因表达值和/或额外基因的加权可以改善受试者的评分,这可以使生物标志物检测的准确度更高。改善的评分可以导致增加的灵敏度,例如,大于90%。在一些情况下,可不使用加权方案。

[0062] 鉴定的生物标志物基因可以在有健康状况(乳腺癌)的受试者与健康受试者之间表现出连通性或共表达的差异。例如,当从在发现阶段研究中检查来自基因芯片的基因表达输出转至使用qPCR检查基因表达输出(具有更大的动态范围和灵敏度)时,这可以发生。基因子网络内的平均连通性的量度可用于将qPCR结果评分,并指示乳腺癌受试者与健康受

试者之间的差异。比较基因子网络内的平均连通性可以提供允许将基因表达值加权的数据,或增加新基因,以改善分数,从而获得更高的测试准确度。

[0063] 图4和表1显示了使用本公开内容的方法鉴定的说明性生物标志物。这些生物标志物中的一种或多种可以是生物标志物小组的一部分。“生物标志物小组”、“生物标志物基因小组”或“生物标志物测定小组”可以指可在生物样品中分析以确定受试者的健康状态或健康状况(乳腺癌)风险的一组生物标志物。在一些情况下,可以使用小组的子集或变体。

表1

探针 ID	基因符号	GenBank 登录号	备注
238574_at	<i>MCART1</i>	BF724944	
207710_at	<i>LCE2B</i>	NM_014357	
243218_at	<i>Hs.161434</i>	AI424847	
205621_at	<i>ALKBH1</i>	NM_006020	
208580_x_at	<i>HIST1H4A</i> /// <i>HIST1H4B</i> /// <i>HIST1H4C</i> /// <i>HIST1H4D</i> /// <i>HIST1H4E</i> /// <i>HIST1H4F</i> /// <i>HIST1H4H</i> /// <i>HIST1H4I</i> /// <i>HIST1H4J</i> /// <i>HIST1H4K</i> /// <i>HIST1H4L</i> /// <i>HIST2H4A</i> /// <i>HIST2H4B</i> /// <i>HIST4H4</i>	NM_021968	
212772_s_at	<i>ABCA2</i>	AL162060	
241371_at	<i>Hs.57851</i>	AW451259	TNFRSF10A
1566840_at	<i>LOC283674</i>	AK092120	
1565694_at	<i>DTYMK</i> /// <i>LOC727761</i>	AK022132	
标准化基因			
212686_at	<i>PPMIH</i>	AB032983	
	<i>TFRC</i>		
	<i>ACTB</i>		

[0064] 生物标志物小组可包含至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90、100、150、200、300、400、500、600、700、800、900或1000种生物标志物。生物标志物小组可包含1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90、100、150、200、300、400、500、600、700、800、900或1000种生物标志物。生物标志物小组可包含至少2种生物标志物。生物标志物小组可包含至少3种生物标志物。生物标志物小组可包含至少4种生物标志物。生物标志物小组可包含至少5种生物标志物。生物标志物小组可包含至少6种生物标志物。生物标志物小组可包含至少7种生物标

志物。生物标志物小组可包含至少8种生物标志物。生物标志物小组可包含至少9种生物标志物。生物标志物小组可包含至少10种生物标志物。

[0065] 生物标志物可以是,例如,癌症相关或癌症关联的基因、乳腺癌途径中的基因、癌基因、与癌症标志相关或涉及癌症标志的基因,或其组合。生物标志物可选自:MCART1、LCE2B、HIST1H4K、ABCA1、ABCA2、ABCA12、TNFRSF10A、AK092120、DTYMK、Hs.161434、ALKBH1,及其同源物、变体、衍生物、产物及组合。生物标志物可以是本文公开的基因的同源物、变体、衍生物或产物。应理解,本公开内容涵盖本文公开的基因的其他名称和别名。

[0066] 在一些实施方案中,生物标志物可以是MCART1。在一些实施方案中,生物标志物可以是LCE2B。在一些实施方案中,生物标志物可以是HIST1H4K。在一些实施方案中,生物标志物可以是ABCA1。在一些实施方案中,生物标志物可以是ABCA2。在一些实施方案中,生物标志物可以是TNFRSF10A。在一些实施方案中,生物标志物可以是AK092120。在一些实施方案中,生物标志物可以是DTYMK。在一些实施方案中,生物标志物可以是Hs.161434。在一些实施方案中,生物标志物可以是ALKBH1。

[0067] 在一些实施方案中,生物标志物小组包含至少2种生物标志物,例如,HIST1H4K和TNFRSF10A。在一些实施方案中,生物标志物小组包含至少2种生物标志物,例如,HIST1H4K和TNFRSF10A。在一些实施方案中,生物标志物小组可包含至少2、3、4、5、6、7、8或9种生物标志物,该生物标志物选自MCART1、LCE2B、HIST1H4K、ABCA1、ABCA2、ABCA12、TNFRSF10A、AK092120、DTYMK、ALKBH1、Hs.161434及其变体。

[0068] 生物标志物可以是溶质载体 (SLC) 基因家族的基因或基因产物,例如,MCART1。MCART1也可称为SLC25A51、CG7943或MGC14836。MCART1可以在9号染色体上找到,染色体位置(bp)为37879400-37904353。MCART1可以在癌症(例如,乳腺癌)中差异表达。突变,例如,乳腺癌中9号染色体区域(例如,9p13.3-p13.2)中的扩增,可能与MCART1的过度表达有关。SLC25是嵌入线粒体内膜和其他细胞器膜的大型核编码转运蛋白家族。SLC25超家族的成员可以参与多种代谢途径和细胞功能。SLC25家族成员可以通过其序列特征识别,如三联结构、六个跨膜 α -螺旋和3倍重复的特征基序。SLC25成员在其所转运的底物的性质和大小、转运方式(即,单向转运、同向转运或反向转运)和驱动力方面差别较大。SLC25基因的突变可能与多种病症有关,该病症涉及,例如,肉碱/酰基肉碱载体缺乏症、高鸟氨酸血症-高氨血症-同型瓜氨酸尿症、天冬氨酸/谷氨酸亚型1和2缺乏症、先天性Amish小头、神经病伴双侧纹状体坏死、先天性铁粒幼细胞贫血、新生儿癫痫性脑病和柠檬酸载体缺乏症。

[0069] 生物标志物可以是晚期角质化包膜2B (LCE2B) 或其产物。LCE2B也可称为小富脯氨酸样表皮分化复合蛋白1B (SPRL1B)、皮肤特异性蛋白Xp5 (XP5) 和晚期包膜蛋白10 (LEP10)。LCE2B可位于染色体带1q21上。LCE2B可参与表皮分化。与LCE2B相关的途径可以是,例如,角质化、细胞因子炎症和对细菌的宿主响应。也可用作生物标志物的LCE2B基因的种内同源基因是LCE2C。

[0070] 生物标志物可以是组蛋白簇1H4家族中的基因或基因产物,例如,组蛋白簇1H4成员K (HIST1H4K),其也可以称为H4组蛋白家族成员D、组蛋白簇1-H4k、H4/D、H4FD、组蛋白H4或DJ160A22.1。组蛋白可以是基础核蛋白,其负责染色体纤维的核小体结构,以及癌症中基因的转录激活。四种核心组蛋白 (H2A、H2B、H3和H4) 中每一种的两个分子可形成八聚体,约146bp的DNA可以以重复单元的形式包裹在八聚体周围,这称为核小体。接头组蛋白H1可与

核小体之间的接头DNA相互作用,并在染色质压缩成高级结构时发挥作用。HIST1H4K可以是无内含子的,并且可以编码复制依赖性组蛋白,该组蛋白是组蛋白H4家族的成员,组蛋白H4。HIST1H4K的转录物可能缺乏polyA尾,并可能包含回文终止元件。HIST1H4K可以在染色体6p22-p21.3上的组蛋白基因簇中找到。

[0071] 生物标志物可以是ATP结合盒(ABC)家族的基因或基因产物,例如,ATP结合盒亚家族A成员2(ABCA2),其也可称为ATP结合盒转运蛋白2、ATP结合盒2、ABC2、EC 3.6.3.41、KIAA1062和EC 3.6.3。ABC蛋白可以跨越细胞外膜和细胞内膜转运多种分子。ABC亚家族编码的蛋白质可以在,例如,脑组织中高度表达,并且可以在巨噬细胞脂质代谢和神经发育中发挥作用。ABC基因可分为7个亚家族:ABC1、MDR/TAP、MRP、ALD、OABP、GCN20和White。生物标志物可以是,例如,ABCA1、ABCA2、ABCA3、ABCA4、ABCA7、ABCA12或ABCA13。ABCA2可以是ABC1亚家族的成员。ABCA2可以编码,例如,两种转录物变体。ABC转运蛋白的过度表达可为肿瘤细胞逃避细胞毒性剂积累提供适应性优势。例如,可以在神经系统和造血系统的细胞中高度表达的ABCA2可能与包括肿瘤干细胞在内的癌细胞中的脂质转运和耐药性相关。

[0072] 生物标志物可以是肿瘤坏死因子受体超家族的基因或基因产物,例如,肿瘤坏死因子受体超家族成员10A(TNFRSF10A),也可以称为TNF相关的凋亡诱导配体受体1、死亡受体4、TRAIL受体1(TRAILR-1)、APO2、DR4和CD261抗原。TNF受体可以被可转导细胞死亡信号并诱导细胞凋亡的TNF相关的凋亡诱导配体(TNFSF10/TRAIL)激活。具有死亡结构域FADD(一种含有衔接蛋白的死亡结构域)的Fas相关蛋白可能是TNF受体蛋白介导的凋亡所需的。衔接分子FADD可以将胱天蛋白酶-8募集到活化的受体上。产生的死亡诱导信号传导复合物(DISC)可以执行胱天蛋白酶-8蛋白水解激活,这可以启动介导凋亡的胱天蛋白酶(例如,天冬氨酸特异性半胱氨酸蛋白酶)的后续级联。TNFRSF10A可以促进NF- κ -B的激活。与TNFRSF10A相关的疾病可包括后巩膜炎和咽结膜热。TNFRSF10A可与TRAF途径、凋亡和自噬相关。TNFRSF10A的种内同源基因可以是TNFRSF10B,其也可以用作本文的生物标志物。

[0073] 生物标志物可以是AK092120。AK092120可以与LOC283674相关。AK092120可以是,例如,miRNA或转录结合位点。AK092120可以与和癌症标志相关的基因有关、相关、作为其替代或表现相似(例如,类似地表达)。AK092120可以是癌症基因的标志。

[0074] 生物标志物可以是脱氧胸苷酸激酶(DTYMK),其也可称为胸苷酸激酶、CDC8、TMPK、TYMK、EC 2.7.4.9和PP3731。嘧啶脱氧核糖核苷酸从头生物合成和嘌呤代谢(KEGG途径)的超途径可以是DTYMK的相关途径之一。DTYMK可以参与,例如,激酶活性和胸苷酸激酶活性。DTYMK编码的蛋白质可以催化脱氧胸苷单磷酸(dTMP)向脱氧胸苷二磷酸(dTDP)的转化。DTYMK缺乏可能与癌细胞的生长和致死率降低有关。

[0075] 生物标志物可以是Hs.161434。Hs.161434可以是,例如,miRNA或转录结合位点。Hs.161434可以与癌症基因的标志有关、相关、作为其替代或表现相似(例如,类似地表达)。Hs.161434可以是癌症基因的标志。

[0076] 生物标志物可以是A1kB家族的基因或基因产物,例如,A1kB同系物1(ALKBH1),其属于2-氧化戊二酸和Fe²⁺依赖性羟化酶家族。ALKBH1是一种组蛋白双加氧酶,其可以去除组蛋白H2A的甲基基团。ALKBH1可以是与癌症标志相关的基因。它可以作用于核酸,如DNA、RNA、tRNA。它可以作为翻译起始和延伸的调节剂,例如,响应于葡萄糖剥夺。ALKBH1可以是DNA N6-甲基腺嘌呤(N6-mA)的脱甲基酶(表观遗传修饰),并可以与胚胎干细胞的核心转录

多能性网络相互作用。ALKBH1在人间充质干细胞(MSC)中的表达可在干细胞诱导中上调。ALKBH1的耗尽可导致N6-mA在激活转录因子4(ATF4)的启动子区域上积累,这可沉默ATF4转录。ALKBH1可参与tRNA的可逆甲基化,该可逆甲基化可作为转录后基因表达调控的机制。

[0077] 生物标志物可以是与癌症标志相关的基因(参见例如,Hanahan D和Weinberg RA (2000年1月)Cell.100(1):57-70以及Hanahan,D和Weinberg,R.A. (2011年)Cell.144(5):646-674,其全部内容通过引用并入本文用于所有目的)。本文公开的基因,如图4或表1所示,可以是癌症基因的标志。癌症基因的标志或标志基因可以是与癌症的标志相关的基因。癌症可以具有标志,该标志可以控制正常细胞向恶性或肿瘤细胞的转化。这些特征或标志可以是,例如,生长信号的自给自足、对抗生长信号不敏感、逃避凋亡、无限制的复制潜力、持续的血管生成、组织侵袭、转移、异常代谢途径、逃避免疫系统、基因组不稳定性炎症。癌症微环境可能需要信号传导系统,该系统将标志基因用于肿瘤生长。图27示出了癌症基因的示例性标志和信号传导系统。图28示出了使用本公开内容的方法鉴定的生物标志物,如DTYMK、TNFRSF10A、ABCA1/2和ALKBH1,它们与癌症的一种或多种标志相关。一种方法可包括确定与癌症标志相关的基因的差异表达。一种方法可包括确定基因的差异表达,该基因是与癌症标志相关的基因的替代(例如,与其相关或具有相似的表达谱)。

[0078] 生物标志物可以是与癌症基因的标志有关、相关、作为其替代或取代,或表现相似(例如,类似地表达,相关)的基因。例如,本文公开的基因,如图4或表1所示,可以与癌症基因的标志相关或具有类似的表达谱,并因此可以用作标志基因的取代物或代理物(proxy)。

[0079] 生物标志物可以是与乳腺癌途径相关的基因。此类基因的非限制性示例包括ABL1、AHR、AKT1、ANXA1、AR、ARAF、ATF1、ATM、ATR、BACH1、BAD、BAK1、BARD1、BAX、CCND1、BCL2、BID、BLM、BMPR1A、BMPR2、BRCA1、BRAF、BRCA2、CASP3、CASP8、CASP9、CDC25A、CDC25B、CDC42、CDH1、CDK2、CDK4、CDK7、CHEK1、CHUK、PLK3、CREB1、CSNK1D、CTNNB1、CYP19A1、DAG1、GADD45A、E2F1、EGFR、EP300、ESR1、FAU、FER、FOXO1、MTOR、GDI1、GRN、GSK3A、MSH6、HDAC1、HMGCR、IMPA1、IRS1、JAK1、JUN、KRAS、SMAD1、SMAD2、SMAD4、SMAD6、SMAD7、MAX、MDM2、MMP1、MRE11、MSH2、MYC、MYT1、NAB1、NF1、NFKB1、ODC1、PAK1、PHB、PIGR、PIK3R2、PLK1、PML PKIA、MAPK1、PTEN、RAC1、RAD51、RALA、RAP1A、RB1、RHEB、RHO、RRAS、SMARCA4、SP1、STAT1、AURKA、STK11、TFPI、TGFB1、TGFB2、TP53、TPR、TSC1、TSC2、VEGFA、WEE1、XRCC3、FOSL1、NCOA3、RAD54L、PIAS1、TRADD、FADD、ALKBH1、MAP3K13、USP15、RAD50、TAB1、RPP38、USP16、NOXA1、EDAR、CHEK2、MYCBP2、SIRT1、ZMYND8、RASGRP3、ERAL1、USP21、FILIP1、HIPK2、LGALS13、DHTKD1、PPP4R3A、MAP3K7CL、ZMIZ1、PPP4R3B、CCNB1IP1、APOBEC3G、CERK、ZNF655、DCAKD、NUP85、ITPKC、USP38、UBE2F、JAKMIP1、RASGEF1A、RALGAP1、MIR1281、GRIK1-AS2或其任何组合。

[0080] 一组生物标志物可基于,例如,特定乳腺癌亚型、疾病严重性、可预测疾病治疗类型或形式的基因或其组合来定制。生物标志物小组可包括,例如,至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90、100、150、200、300、400、500、600、700、800、900或1000种生物标志物。在一些实施方案中,生物标志物小组可包含至少9种生物标志物。

[0081] 本公开内容的方法可包括从受试者获得或提供生物样品。生物样品可以是含有或假定含有生物标志物的任何物质。生物样品可以是含有或假定含有核酸或蛋白质的任何物质。

[0082] 生物样品可以是液体样品。生物样品可以是体液。生物样品可以是包含外来体的样品。生物流体可以是基本上无细胞的液体样品,例如唾液、血浆、血清、汗液、尿液和泪液。在其他实施方案中,生物样品可以是固体生物样品,例如,粪便或组织活检物,例如,肿瘤活检物。样品还可包含体外细胞培养成分,包括但不限于由细胞在细胞培养基中生长而产生的条件培养基、重组细胞和细胞组分。

[0083] 生物样品可选自:血液、血清、血浆、尿液、汗液、泪液、唾液、痰液、其组分及其任何组合。在一些实施方案中,生物样品可以是唾液。在一些实施方案中,生物样品可以是血液。

[0084] 生物样品的非限制性示例包括唾液、全血、外周血、血浆、血清、腹水、脑脊液、汗液、尿液、泪液、口腔样品、腔冲洗液、痰液、器官冲洗液、骨髓、滑液、房水、羊水、耳垢、母乳、支气管肺泡灌洗液、精液(包括前列腺液)、考珀液或射精前液、女性射出液、汗液、粪便物质、头发、眼泪、囊液、胸膜液和腹膜液、心包液、淋巴、食糜、乳糜、胆汁、间质液、月经、脓液、皮脂、呕吐物、阴道分泌物、粘膜分泌物、稀便、胰液、来自窦腔的灌洗液、支气管肺抽吸物或其他灌洗液。生物样品还可包括囊胚腔、脐带血或母体循环,它们可以源自胎儿或母体。生物样品也可以是可从中获得外来体的组织样品或活检物。例如,如果样品是固体样品,则可以培养来自样品的细胞,并诱导或取回外来体产物。

[0085] 生物样品的收集可以在任何合适的环境中进行,例如,医院、家庭、诊所、药房、乳腺成像诊所和诊断实验室。生物样品可以通过邮寄或通讯员运送到中心诊所进行分析。在分析之前,生物样品可以在合适的条件下储存。

[0086] 本公开内容的方法可用于检测来自外来体的生物标志物。例如,来自生物样品的外来体部分的生物标志物或源自外来体的生物标志物。

[0087] 外来体可以是小膜结合囊泡,其可以从多种不同的细胞释放到细胞外环境中,该细胞例如但不限于,源自或衍生自外胚层、内胚层或中胚层的细胞,包括已经历遗传、环境和/或任何其他变化或改变的任何此类细胞(例如,细菌/病毒感染的细胞、肿瘤细胞或具有基因突变的细胞)。当一段细胞膜自发地内陷并最终被胞吐时,外来体可在细胞内产生。

[0088] 外来体可具有约30-1000nm、约30-800nm、约30-200nm、约30-100nm、约20nm至约100nm、约30nm至约150nm、约30nm至约120nm、约50nm至约150nm或约50nm至约120nm的直径。

[0089] 外来体也可称为微泡、纳米泡、小泡、dexosome、疱、大疱、前列腺小体、微粒、内腔小泡、内体样小泡或胞吐载体。外来体还可包括源自质膜或内膜的任何脱落的膜结合颗粒。外来体还可包括由脂质双层膜界定的细胞衍生结构,该脂质双层膜产生于突出的外凸(起泡)分离和部分质膜的密封,或产生于任何含有多种肿瘤来源的膜相关蛋白的细胞内膜界定小泡结构的输出,该肿瘤来源的膜相关蛋白包括源自宿主循环的、与肿瘤衍生的蛋白质以及包含在外来体腔中的分子(包括但不限于肿瘤衍生的微RNA或细胞内蛋白质)选择性地结合的表面结合分子。

[0090] 外来体可以是生物标志物的来源。外来体可以存在于,例如,生物流体中,如唾液、血液、尿液、脑脊液和母乳中。外来体可包含蛋白质和核酸。培养的所有细胞类型都可以分泌外来体。外来体可以参与细胞间信号传导。外来体可含有其来源细胞(即外来体起源的细胞)的分子成分。在从生物样品(例如,唾液)获得的外来体与从组织(例如,乳腺癌组织)获得的外来体之间可以存在相关性。外来体内的生物标志物可以与在受试者的致癌组织中发现的生物标志物相同。

[0091] 外来体可以是来源细胞特异性外来体。外来体可以源自肿瘤或癌细胞。外来体的来源细胞可以是,例如,肺、胰腺、胃、肠、膀胱、肾、卵巢、睾丸、皮肤、结直肠、乳腺、前列腺、脑、食道、肝、胎盘或胎儿的细胞。在一些实施方案中,外来体的来源细胞是乳腺组织。

[0092] 本公开内容的方法可包括测定从外来体释放的生物标志物。在一些实施方案中,可以从生物样品直接测定外来体生物标志物,使得在无需预先从生物样品中分离、纯化或浓缩外来体的情况下分析外来体的一种或多种生物标志物。在一些实施方案中,外来体可以从生物样品分离,并在生物标志物分析之前富集。

[0093] 外来体可以在分析之前纯化或浓缩。外来体的分析可包括定量生物样品的一个或多个外来体群体的量。例如,可以定量外来体的异源群体,或可以从外来体的异源群体分离并定量外来体的同源群体,如具有特定生物标志物谱的外来体群体,或源自特定细胞类型的外来体群体(来源细胞特异性外来体)。外来体的分析还可包括定量或定性地检测外来体的特定生物标志物谱或生物特征。富集的外来体群体可以从衍生自能够产生并释放外来体至体液中的任何细胞的生物样品获得。

[0094] 可以使用,例如,尺寸排阻色谱、密度梯度离心、差速离心、纳米膜超滤、免疫吸附捕获、亲和纯化、微流体分离、蛋白质纯化试剂盒或其组合,从生物样品浓缩或分离外来体。

[0095] 尺寸排阻色谱如凝胶渗透柱、离心或密度梯度离心和过滤方法可用于外来体分离。例如,可通过差速离心、阴离子交换色谱和/或凝胶渗透色谱、蔗糖密度梯度、细胞器电泳、磁性激活细胞分选(MACS)或纳米膜超滤浓缩器来分离外来体。可以使用分离或浓缩方法的各种组合。

[0096] 高丰度蛋白质,如白蛋白和免疫球蛋白,可能阻碍外来体从生物样品分离。例如,可以使用多种抗体的系统从生物样品分离外来体,该抗体对在该生物样品中发现的最丰富的蛋白质具有特异性。这种系统可以一次去除多达几种蛋白质,从而显露出较低丰度的物质,如来源细胞特异性外来体。外来体从生物样品的分离还可通过高丰度蛋白质去除方法增强。外来体从生物样品的分离可以通过使用糖肽捕获来去除血清蛋白而增强。此外,可以通过差速离心,然后与针对细胞质表位或抗细胞质表位的抗体接触而分离来自生物样品的外来体。蛋白质试剂盒可用于外来体分离。

[0097] 外来体的存在可以通过经由蛋白质印迹或其他检测手段检测已知的外来体标志物来确认,该标志物例如但不限于MHC I类蛋白、LAMP1、CD9、CD63和CD81。透射电子显微术(TEM)、蛋白质浓度和Nano-Sight LM-10HS分析也可用于分析所分离的外来体的存在和纯度。

[0098] 生物标志物从外来体的释放可以例如通过裂解外来体进行。外来体的裂解可以直接在生物样品中进行。外来体的裂解可以在富集外来体部分之后进行。可以使生物样品经受裂解条件,例如,以便裂解外来体部分。裂解可以通过例如超声进行。裂解方法的非限制性示例包括试剂辅助的裂解方法(例如,使用洗涤剂)、无试剂裂解方法、化学裂解法、机械裂解法(例如,使用破碎、研磨、超声)、热裂解法(例如,使用加热)和电裂解法(例如,靶颗粒的脂质双层的不可逆电穿孔)。

[0099] 可以在生物标志物分析之前处理生物样品以去除细胞(例如,全部的完整细胞)。可使没有细胞的样品经受外来体分离和富集。包含外来体的样品可以在生物标志物分析之前保存和/或储存。

[0100] 本公开内容的方法可以使用核酸扩增。扩增的核酸可以使用例如大规模并行测序(例如,新一代测序方法)或杂交平台来分析。合适的扩增反应可以是指数或等温的,并且可包括任何DNA扩增反应,包括但不限于PCR、链置换扩增(SDA)、连接酶链反应(LCR)、线性扩增、多重置换扩增(MDA)、滚环扩增(RCA)或其组合。

[0101] 本公开内容的方法可包括生物标志物检测和分析。来自生物标志物分析的结果可用于生成受试者的生物标志物特征。图3示出了用于从唾液样品进行生物标志物分析以评估受试者的癌症的示例方法的工作流程。唾液可以从受试者(301)收集,例如,通过将唾液吐入收集管。然后将唾液样品运送到实验室进行处理和储存(302)。样品可以运送到实验室进行测定。样品可以离心。测定试剂可以添加至样品中。可以分离和/或稳定包含RNA的唾液样品的外来体部分(303)。可以用合适的技术例如磁珠测定系统从样品分离RNA。可以将分离的RNA样品储存(例如,在-80℃),用于后续处理和分析。在一些实施方案中,分离的RNA样品可以进一步处理而不储存。RNA可以逆转录(例如,使用RT-PCR)生成cDNA,并且可以进行预扩增步骤(304)。在一些实施方案中,RNA可以在一步式反应中逆转录和预扩增。在一些实施方案中,逆转录和预扩增可以分步进行。在一些实施方案中,可以不进行预扩增。可以扩增cDNA。例如,可以处理cDNA以增加稳定性。可以储存cDNA用于后续处理。在一些实施方案中,cDNA可以被处理而不储存。可以对cDNA进行qPCR(305)。在一些实施方案中,qPCR可以在一步式反应中进行。可以分析来自qPCR的数据,以检测候选生物标志物基因的表达水平。可以在工作流程中的任何步骤之前、之后或期间加入纯化步骤。在一些实施方案中,RNA测序可用于分析RNA。在一些实施方案中,靶向RNA测序可用于分析RNA。在一些实施方案中,miRNA或小RNA测序可用于分析RNA。

[0102] 生物标志物检测可包括使用,例如,微阵列分析、聚合酶链反应(PCR)(包括基于PCR的方法,如RT-PCR和定量PCR(qPCR))、与等位基因特异性探针杂交、酶突变检测、连接酶链反应(LCR)、寡核苷酸连接测定(OLA)、流式细胞术异源双链分析、错配的化学切割、质谱、核酸测序、单链构象多态性(SSCP)、变性梯度凝胶电泳(DGGE)、温度梯度凝胶电泳(TGGE)、限制性片段多态性、基因表达系列分析(SAGE)、免疫印迹、免疫沉淀、酶联免疫吸附测定(ELISA)、放射免疫测定(RIA)、流式细胞术、电子显微术、使用G带核型分析的基因测试、脆性X检测、染色体微阵列(CMA,也称为比较基因组杂交(CGH))(例如,用于测试亚微观的基因缺失和/或重复)、基于阵列的比较基因组杂交、用阵列检测单核苷酸多态性(SNP)、亚端粒荧光原位杂交(ST-FISH)(例如,用于检测亚微观拷贝数变体(CNV))、表达谱分析、DNA微阵列、RNA微阵列、mRNA微阵列、miRNA微阵列、高密度寡核苷酸微阵列、全基因组RNA表达阵列、肽微阵列、酶联免疫吸附测定(ELISA)、基因组测序、DNA测序、RNA测序、miRNA测序、从头测序、454测序(Roche)、焦磷酸测序、Helicos真正单分子测序、SOLiD™测序(Applied Biosystems, Life Technologies)、SOLEXA测序(Illumina测序)、纳米测序、化学敏感的场效应晶体管(chemFET)阵列测序(Ion Torrent)、离子半导体测序(Ion Torrent)、DNA纳米球测序、纳米孔测序、Pacific Biosciences SMRT测序、Genia Technologies纳米孔单分子DNA测序、Oxford纳米孔单分子DNA测序、聚合酶克隆测序、拷贝数变异(CNV)分析测序、小核苷酸多态性(SNP)分析、免疫组织化学(IHC)、免疫细胞化学(ICC)、质谱、串联质谱、基质辅助激光解吸电离飞行时间质谱(MALDI-TOF MS)、原位杂交、荧光原位杂交(FISH)、显色原位杂交(CISH)、银原位杂交(SISH)、聚合酶链反应(PCR)、数字PCR(dPCR)、逆转录PCR、定量PCR

(Q-PCR)、单标记qPCR、实时PCR、nCounter分析(Nanostring technology)、蛋白质印迹、DNA印迹、SDS-PAGE、凝胶电泳和RNA印迹,或其任何组合。

[0103] 本公开内容的方法可包括将基因的表达量化。基因的表达可以在转录组水平(例如, RNA、mRNA、miRNA)、蛋白质组水平(例如, 蛋白质、多肽)或其组合下量化。该基因可以是癌症相关基因。该基因可以是乳腺癌途径中的基因。该基因可以是癌基因。该基因可以与癌症的标志有关。

[0104] 表达分析可以在例如从外来体提取的RNA上进行。RNA可以是, 例如, 总RNA、mRNA、miRNA和tRNA。在一些实施方案中, 外来体可以是来源细胞特异性外来体。从这些外来体产生的表达模式可以指示给定的疾病状态、疾病阶段、治疗相关的特征或生理状况。一旦分离出总RNA, 就可以生成互补DNA(cDNA)。然后可以进行特定mRNA靶标的qRT-PCR测定。在一些实施方案中, 可以进行表达微阵列, 以检测和鉴定高度复用的表达标志物组。用于建立基因表达谱的方法可包括确定由可以编码蛋白质或肽的基因所产生的RNA的量。这可以通过定量逆转录酶PCR(qRT-PCR)、竞争RT-PCR、实时RT-PCR、差异显示RT-PCR、RNA印迹分析、测序或其他测试来实现。尽管可以使用单个的PCR反应执行这些技术, 但也可以对mRNA产生的互补DNA(cDNA)进行扩增并分析。

[0105] qPCR或实时PCR可以指这样的PCR方法, 其中用PCR的每个循环监测一定量的可检测信号。可以确定其中可检测信号达到可检测水平的循环阈值(Ct)。Ct值越低, 所询问的等位基因的浓度越高。可以在PCR的指数生长(对数)阶段期间收集数据, 其中PCR产物的量与模板核酸的量成正比。用于实时PCR的系统可包括ABI 7700和7900HT序列检测系统。在PCR的指数阶段中信号的增加可以提供对含有突变等位基因的模板的量的定量测量。

[0106] 生物标志物可以通过等位基因特异性PCR测定, 等位基因特异性PCR可包括特异性引物, 用于同时扩增和区分基因的两个等位基因。在一些实施方案中, 可以测定生物标志物以检测单链构象多态性(SSCP), 这涉及基于序列差异和DNA和RNA适体的单链核酸电泳分离。DNA和RNA适体可以是短寡核苷酸序列, 该短寡核苷酸序列可以根据其以高亲和力结合特定分子的能力从随机库中选择。

[0107] 在一些实施方案中, 可以通过分析RNA来确定生物标志物的差异表达。该方法可包括产生相应的cDNA, 然后分析所得的DNA。

[0108] 在一些实施方案中, 该方法可包括RNA测序。例如, 该方法可包括以下的一个或多个: RNA提取、片段化、cDNA生成、测序文库制备和高通量测序(例如, 新一代测序、大规模并行测序)。在一些实施方案中, 该方法可包括将靶特异性探针用于本文公开的生物标志物。在一些实施方案中, 该方法可包括使用特异性微阵列(例如, 用于miRNA、mRNA)。

[0109] 在一些实施方案中, 小RNA测序或miRNA测序可用于RNA分析。miRNA测序可包括产生从RNA(例如, 从唾液获得)制备的含有miRNA和其他小RNA的RNA文库。

[0110] 生物标志物分析可包括, 例如, 确定突变的不存在(例如, 野生型)或一种或多种突变的存在(例如, 从头突变、无义突变、错义突变、沉默突变、移码突变、插入、置换、点突变、单核苷酸多态性(SNP)、单核苷酸变体、从头单核苷酸变体、缺失、重排、扩增、染色体易位、中间缺失、染色体倒位、杂合性缺失、功能丧失、功能获得、显性失活突变或致死突变); 核酸修饰(例如, 甲基化); 或存在或不存在蛋白质翻译后修饰(例如, 乙酰化、烷基化、酰胺化、生物素化、谷氨酰化、糖基化、糖化、甘氨酸化、羟化、碘化、异戊二烯化、脂化、异戊烯化、豆蔻

酰化、法尼基化、牻牛儿基牻牛儿基化、ADP-核糖基化、氧化、棕榈酰化、聚乙二醇化、磷脂酰肌醇加成、磷酸泛酰巯基乙胺基化、磷酸化、聚唾液酸化、焦谷氨酸形成、精氨酰化、硫酸化或硒化)。

[0111] 本文描述的方法可以使用一种或多种新一代测序或高通量测序,诸如但不限于美国专利号7,335,762;7,323,305;7,264,929;7,244,559;7,211,390;7,361,488;7,300,788;和7,280,922中所述的那些方法。

[0112] 新一代测序技术可包括,例如,Helicos真正单分子测序(tSMS)(Harris T.D.等人(2008)Science 320:106-109);454测序(Roche)(Margulies,M.等人2005,Nature,437,376-380);SOLiD技术(Applied Biosystems);SOLEXA测序(Illumina);Pacific Biosciences的单分子实时(SMRTTM)技术;纳米孔测序(Soni GV和Meller A.(2007)Clin Chem 53:1996-2001);半导体测序(Ion Torrent;Personal Genome Machine);DNA纳米球测序;使用来自Dover Systems(Polonator)的技术测序,以及在测序前不需要扩增或以其他方式转化天然DNA的技术(例如,Pacific Biosciences和Helicos),如基于纳米孔的策略(例如,Oxford Nanopore、Genia Technologies和Nabsys)。

[0113] 在一些实施方案中,新一代测序技术可以是454测序(Roche)(参见例如,Margulies,M等人(2005)Nature 437:376-380)。454测序可以涉及两个步骤。在第一步中,可以将DNA剪切成约300-800个碱基对的片段,并且这些片段可以是平端的。然后可以将寡核苷酸衔接子连接至片段的末端。衔接子可以作为杂交引物的位点,以用于片段的扩增和测序。可以使用例如衔接子B将片段附接至DNA捕获珠子,例如,链霉亲和素包被的珠子,衔接子B可以含有5'-生物素标签。可以通过杂交将片段附接至DNA捕获珠子。每个珠子可以捕获单个片段。附接至珠子的片段可以在油-水乳液的液滴内被PCR扩增。结果可以是在每个珠子上多个拷贝的克隆扩增DNA片段。可以破坏乳液,同时使扩增的片段仍结合至其特定的珠子。在第二步中,可以在孔(皮升大小;PicoTiterPlate (PTP) 设备)中捕获珠子。表面可以设计成使得每孔仅有一个珠子。PTP设备可以加载到用于测序的仪器中。可以对每个DNA片段并行地进行焦磷酸测序。添加一个或多个核苷酸可以产生光信号,该光信号可以通过测序仪器中的CCD相机记录。信号强度可以与所掺入的核苷酸数目成比例。

[0114] 焦磷酸测序可以利用可在添加核苷酸时释放的焦磷酸(PPi)。在腺苷5'磷酸硫酸存在下,PPi可以通过ATP硫酸化酶转化为ATP。荧光素酶可以使用ATP将荧光素转化为氧化荧光素,并且该反应可以产生光,该光可以被检测和分析。使用的454测序系统可以是GS FLX+系统或GS Junior系统。

[0115] 新一代测序技术可以是SOLiD技术(Applied Biosystems;Life Technologies)。在SOLiD测序中,可以将基因组DNA剪切成片段,并且可以将衔接子附接至片段的5'和3'端,以生成片段文库。或者,内部衔接子的引入可以通过将衔接子连接至片段的5'和3'端,使片段环化,消化环化的片段以产生内部衔接子,并将衔接子附接至所得片段的5'和3'端而实现,从而生成配对文库。接下来,可以在含有珠子、引物、模板和PCR组分的微反应器中制备克隆珠子群体。在PCR之后,可以使模板变性,并可以富集珠子,使珠子与延伸的模板分离。所选珠子上的模板可以经受3'修饰,该3'修饰允许键合至载玻片。测序引物可以与衔接子序列结合。一组四个荧光标记的双碱基探针可以竞争连接至测序引物。通过在每个连接反应中询问每个第一和第二碱基,可以实现双碱基探针的特异性。模板的序列可以通过顺序

杂交,以及部分随机寡核苷酸与可以通过特定荧光团鉴定的确定碱基(或碱基对)连接来确定。记录颜色后,可以切割并去除连接的寡核苷酸,然后重复该过程。在一系列连接循环后,可以去除延伸产物,并可以用与n-1位置互补的引物重置模板,以进行第二轮连接循环。每个序列标签可以完成五轮引物重置。通过引物重置过程,可以通过两种不同的引物,在两个独立的连接反应中询问大多数碱基。使用多碱基编码方案,通过额外引物进行测序可以实现高达99.99%的准确度。

[0116] 新一代测序技术可以是SOLEXA测序(ILLUMINA测序)。ILLUMINA测序可以基于使用折回PCR和锚定引物在固体表面上的DNA扩增。ILLUMINA测序可涉及文库制备步骤。可以将基因组DNA片段化,并且可以将剪切的末端修复和腺苷酸化。可以将衔接子添加到片段的5'和3'端。可以对片段进行大小选择和纯化。ILLUMINA序列可包括簇生成步骤。DNA片段可以通过与附接于流动池通道表面的一片寡核苷酸杂交而附接于流动池通道的表面。可以通过桥式扩增将片段延伸和克隆扩增以产生独特的簇。片段变成双链,并且可以使双链分子变性。固相扩增随后变性的多个循环可在流动池的每个通道中产生由具有相同模板的单链DNA分子的大约1,000个拷贝组成的数百万个簇。反向链可以被切割并冲洗掉。末端可以被阻断,并且引物可以与DNA模板杂交。ILLUMINA测序可包括测序步骤。可以同时为数亿个簇进行测序。引物、DNA聚合酶和四种荧光团标记的可逆终止核苷酸可用于顺序测序。所有四种碱基可以彼此竞争该模板。核苷酸掺入后,可使用激光激发荧光团,并捕获图像并记录第一碱基的身份。去除来自每个掺入的碱基的3'终止剂和荧光团,并重复掺入、检测和鉴定步骤。每个循环可以读取单一碱基。在一些实施方案中,使用HiSeq系统(例如,HiSeq 2500、HiSeq 1500、HiSeq 2000或HiSeq1000)进行测序。在一些实施方案中,使用MiSeq个人测序仪。在一些实施方案中,使用Genome Analyzer IIx。

[0117] 新一代测序技术可包括Pacific Biosciences的实时(SMRT™)技术。在SMRT中,四种DNA碱基中的每一种可以附接至四种不同荧光染料中的一种。这些染料可以被磷酸连接。单个DNA聚合酶可以在零模波导(ZMW)的底部用单分子的模板单链DNA固定。ZMW可以是限制结构,使得DNA聚合酶能够在荧光核苷酸背景下观察单个核苷酸被DNA聚合酶的掺入,荧光核苷酸可以快速扩散进出ZMW(在数微秒内)。将核苷酸掺入生长链中可能需要几毫秒。在此期间,荧光标记可被激发并产生荧光信号,并且可以切割荧光标签。ZMW可以从下方照射。来自激发束的衰减光可以穿透每个ZMW的下部20-30nm。可创建检测限为20仄升(zeptoliter)(10^{-21} 升)的显微镜。微小的检测体积可以使背景噪声的降低提高1000倍。染料的相应荧光的检测可以指示掺入了哪种碱基。该过程可以重复。

[0118] 新一代测序方法可包括纳米孔测序(参见例如,Soni GV和Meller A. (2007) Clin Chem 53:1996-2001)。纳米孔可以是直径约为1纳米级别的小孔。将纳米孔浸入导电流体中并在其上施加电势可由于离子通过纳米孔的传导而产生轻微电流。流过的电流的量可对纳米孔的大小敏感。当DNA分子穿过纳米孔时,DNA分子上的每个核苷酸可以在不同程度上阻塞纳米孔。因此,当DNA分子通过纳米孔时,通过纳米孔的电流的变化可以表示DNA序列的读取。纳米孔测序技术可以来自Oxford Nanopore Technologies;例如,GridION系统。单个纳米孔可以插入跨微孔顶部的聚合物膜中。每个微孔可以具有用于单独感测的电极。微孔可以制成阵列芯片,每个芯片具有100,000或更多个微孔(例如,超过约200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000或1,000,000个)。仪器(或节点)可

用于分析芯片。数据可以实时分析。一次可以操作一个或多个仪器。纳米孔可以是蛋白质纳米孔,例如,蛋白质 α -溶血素纳米孔——一种七聚体蛋白质孔。纳米孔可以是固态纳米孔制成的,例如,在合成膜(例如, SiNx或SiO₂)中形成的纳米大小的孔。纳米孔可以是杂化孔(例如,蛋白质孔整合到固态膜中)。纳米孔可以是具有集成传感器(例如,隧道电极检测器、电容检测器或基于石墨烯的纳米间隙或边缘状态检测器(参见例如, Garaj等人(2010) Nature 第67卷, doi:10.1038/nature09379))的纳米孔。可以将纳米孔功能化,以分析特定类型的分子(例如, DNA、RNA或蛋白质)。纳米孔测序可包括“链测序”,其中完整的DNA聚合物可以通过蛋白质纳米孔,随着DNA在孔中移位而实时测序。酶可以分离双链DNA的链,并将链投入纳米孔。DNA的一端可以有发夹,并且系统可以读取两条链。在一些实施方案中,纳米孔测序是“核酸外切酶测序”,其中各个核苷酸可以通过进行性外切核酸酶从DNA链切割,并且核苷酸可以通过蛋白质纳米孔。核苷酸可以瞬时结合孔中的分子(例如,环糊精)。电流的特征性中断可用于鉴定碱基。

[0119] 可以使用来自GENIA的纳米孔测序技术。工程化蛋白质孔可以嵌入脂质双层膜中。“主动控制”技术可用于实现有效的纳米孔-膜组装并控制DNA通过通道的移动。在一些实施方案中,纳米孔测序技术来自NABsys。基因组DNA可以片段化成平均长度约100kb的链。100kb片段可以被制成单链,并随后与6-mer探针杂交。具有探针的基因组片段可以被驱动通过纳米孔,这可以产生电流-时间轨迹。电流轨迹可以提供每个基因组片段上探针的位置。基因组片段可以被排列起来,以产生基因组的探针图。对于探针文库,该过程可以并行完成。可以生成每个探针的基因组长度探针图谱。错误可以通过称为“移动窗口杂交测序(mwSBH)”的过程来修复。在一些实施方案中,纳米孔测序技术来自IBM/Roche。电子束可用于在微芯片中制造纳米孔大小的开口。电场可用于将DNA牵引过或穿过纳米孔。纳米孔中的DNA晶体管器件可包括交替的纳米大小的金属层和介电层。DNA骨架中的离散电荷可被DNA纳米孔内的电场捕获。关闭和打开门电压可允许DNA序列被读取。

[0120] 新一代测序方法可包括离子半导体测序(例如,使用来自Life Technologies (Ion Torrent)的技术)。离子半导体测序可以利用以下事实的优点:当将核苷酸掺入DNA链中时,可以释放离子。为进行离子半导体测序,可以形成高密度的微机械加工孔阵列。每个孔可以容纳单个DNA模板。孔的下方可以是离子敏感层,而在离子敏感层的下方可以是离子传感器。当将核苷酸添加至DNA时,可以释放H⁺,其可以作为pH的变化来测量。H⁺离子可以转换为电压并由半导体传感器记录。阵列芯片可以依次用一个接一个的核苷酸充满。无需扫描、光或相机。在一些实施方案中,IONPROTON™测序仪用于将核酸测序。在一些实施方案中,使用IONPGM™测序仪。

[0121] 新一代测序可包括DNA纳米球测序(例如,通过Complete Genomics进行的DNA纳米球测序;参见例如, Drmanac等人(2010) Science 327:78-81)。DNA可以被分离、片段化并选择大小。例如, DNA可以被片段化(例如,通过超声)至约500bp的平均长度。衔接子(Ad1)可以附接至片段的末端。衔接子可用于与用于测序反应的锚定物杂交。每个末端结合有衔接子的DNA可以被PCR扩增。可以修饰衔接子序列,使得互补的单链末端彼此结合,从而形成环状DNA。可以将DNA甲基化,以保护其免于被在后续步骤中使用IIS型限制酶切割。衔接子(例如,右衔接子)可以具有限制识别位点,并且限制识别位点可以保持非甲基化。衔接子中的非甲基化限制识别位点可以被限制酶(例如, AclI)识别,并且DNA可以被AclI在右衔接子的

右侧13bp处切割,以形成线性双链DNA。第二轮左右衔接子(Ad2)可以连接至线性DNA的任一端,并且所有结合有两个衔接子的DNA都可以被PCR扩增(例如,通过PCR)。可以修饰Ad2序列,以使其彼此结合并形成环状DNA。可将DNA甲基化,但限制酶识别位点可在左Ad1衔接子上保持非甲基化。可以应用限制酶(例如,AclI),并DNA可以在Ad1左侧13bp处被切割,以形成线性DNA片段。第三轮左右衔接子(Ad3)可以连接至线性DNA的左右翼,并且得到的片段可以被PCR扩增。可以修饰衔接子,以使其彼此结合并形成环状DNA。可添加III型限制酶(例如,EcoP15);EcoP15可以将DNA在Ad3左侧26bp和Ad2右侧26bp处切割。这种切割可以去除大段DNA,并再次使DNA线性化。第四轮左右衔接子(Ad4)可以连接至DNA,DNA可以被扩增(例如,通过PCR),并且被修饰以使其彼此结合并形成完整的环状DNA模板。滚环复制(例如,使用Phi 29 DNA聚合酶)可用于扩增DNA的小片段。四个衔接子序列可包含可以杂交的回文序列,并且单链可以折叠到其自身上以形成DNA纳米球(DNB™),该纳米球的平均直径可以为约200-300纳米。DNA纳米球可以附接(例如,通过吸附)到微阵列(测序流动池)。流动池可以是涂有二氧化硅、钛和六甲基二硅氮烷(HMDS)以及光刻胶材料的硅晶片。通过将荧光探针连接至DNA,可以通过非链式测序进行测序。询问位置的荧光颜色可以通过高分辨率相机可视化。可以确定衔接子序列之间的核苷酸序列的同一性。

[0122] 新一代测序技术可以是Helicos真正单分子测序(tSMS)(参见例如,Harris T.D.等人(2008)Science 320:106-109)。在tSMS技术中,可以将DNA样品切割成约100至200个核苷酸的链,并且可以将polyA序列添加至每条DNA链的3'端。可以通过添加荧光标记的腺苷核苷酸将每条链标记。然后可以将DNA链杂交至流动池,该流动池可包含数百万个固定在流动池表面的寡核苷酸-T捕获位点。模板的密度可以为约1亿个模板/cm²。然后可以将流动池加载到仪器例如HELISCOPE™测序仪中,并且激光可以照射流动池的表面,从而显露出每个模板的位置。CCD相机可以映射模板在流动池表面上的位置。然后模板荧光标记可以被切割并洗掉。可以通过引入DNA聚合酶和荧光标记的核苷酸开始测序反应。寡核苷酸-T核酸可作为引物。DNA聚合酶可以以模板定向的方式将标记的核苷酸掺入至引物。可以去除DNA聚合酶和未掺入的核苷酸。可以通过对流动池表面成像来检测已经定向掺入荧光标记核苷酸的模板。成像后,切割步骤可以去除荧光标记,并可以用其他荧光标记核苷酸重复该过程,直到达到所需的读取长度。可以通过每个核苷酸添加步骤收集序列信息。测序可以是异步的。测序可包括每天或每小时至少10亿个碱基。

[0123] 测序技术可包括配对末端测序,其中正向和反向模板链都可被测序。在一些实施方案中,测序技术可包括配对文库测序。在配对文库测序中,DNA可以是片段,并且可以对2-5kb片段进行端修复(例如,用生物素标记的dNTP)。可以将DNA片段环化,并且可以通过消化去除未环化的DNA。可以将环状DNA片段化并纯化(例如,使用生物素标记)。纯化的片段可以进行端修复并连接至测序衔接子。

[0124] 序列读取可以为约、大于约、小于约或至少约10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、114、115、116、117、118、119、120、121、122、123、124、125、126、127、128、129、130、131、132、133、

134、135、136、137、138、139、140、141、142、143、144、145、146、147、148、149、150、151、152、153、154、155、156、157、158、159、160、161、162、163、164、165、166、167、168、169、170、171、172、173、174、175、176、177、178、179、180、181、182、183、184、185、186、187、188、189、190、191、192、193、194、195、196、197、198、199、200、201、202、203、204、205、206、207、208、209、210、211、212、213、214、215、216、217、218、219、220、221、222、223、224、225、226、227、228、229、230、231、232、233、234、235、236、237、238、239、240、241、242、243、244、245、246、247、248、249、250、251、252、253、254、255、256、257、258、259、260、261、262、263、264、265、266、267、268、269、270、271、272、273、274、275、276、277、278、279、280、281、282、283、284、285、286、287、288、289、290、291、292、293、294、295、296、297、298、299、300、301、302、303、304、305、306、307、308、309、310、311、312、313、314、315、316、317、318、319、320、321、322、323、324、325、326、327、328、329、330、331、332、333、334、335、336、337、338、339、340、341、342、343、344、345、346、347、348、349、350、351、352、353、354、355、356、357、358、359、360、361、362、363、364、365、366、367、368、369、370、371、372、373、374、375、376、377、378、379、380、381、382、383、384、385、386、387、388、389、390、391、392、393、394、395、396、397、398、399、400、401、402、403、404、405、406、407、408、409、410、411、412、413、414、415、416、417、418、419、420、421、422、423、424、425、426、427、428、429、430、431、432、433、434、435、436、437、438、439、440、441、442、443、444、445、446、447、448、449、450、451、452、453、454、455、456、457、458、459、460、461、462、463、464、465、466、467、468、469、470、471、472、473、474、475、476、477、478、479、480、481、482、483、484、485、486、487、488、489、490、491、492、493、494、495、496、497、498、499、500、525、550、575、600、625、650、675、700、725、750、775、800、825、850、875、900、925、950、975、1000、1100、1200、1300、1400、1500、1600、1700、1800、1900、2000、2100、2200、2300、2400、2500、2600、2700、2800、2900或3000个碱基。在一些实施方案中，序列读取为约10至约50个碱基、约10至约100个碱基、约10至约200个碱基、约10至约300个碱基、约10至约400个碱基、约10至约500个碱基、约10至约600个碱基、约10至约700个碱基、约10至约800个碱基、约10至约900个碱基、约10至约1000个碱基、约10至约1500个碱基、约10至约2000个碱基、约50至约100个碱基、约50至约150个碱基、约50至约200个碱基、约50至约500个碱基、约50至约1000个碱基、约100至约200个碱基、约100至约300个碱基、约100至约400个碱基、约100至约500个碱基、约100至约600个碱基、约100至约700个碱基、约100至约800个碱基、约100至约900个碱基或约100至约1000个碱基。

[0125] 来自样品的序列读取数目可以为约、大于约、小于约或至少约100、1000、5,000、10,000、20,000、30,000、40,000、50,000、60,000、70,000、80,000、90,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、1,000,000、2,000,000、3,000,000、4,000,000、5,000,000、6,000,000、7,000,000、8,000,000、9,000,000、或10,000,000个。

[0126] 样品的测序深度可以为约、大于约、小于约或至少约1x、2x、3x、4x、5x、6x、7x、8x、9x、10x、11x、12x、13x、14x、15x、16x、17x、18x、19x、20x、21x、22x、23x、24x、25x、26x、27x、28x、29x、30x、31x、32x、33x、34x、35x、36x、37x、38x、39x、40x、41x、42x、43x、44x、45x、46x、47x、48x、49x、50x、51x、52x、53x、54x、55x、56x、57x、58x、59x、60x、61x、62x、63x、64x、65x、66x、67x、68x、69x、70x、71x、72x、73x、74x、75x、76x、77x、78x、79x、80x、81x、82x、83x、84x、

85x、86x、87x、88x、89x、90x、91x、92x、93x、94x、95x、96x、97x、98x、99x、100x、110x (H0x)、120x、130x、140x、150x、160x、170x、180x、190x、200x、300x、400x、500x、600x、700x、800x、900x、1000x、1500x、2000x、2500x、3000x、3500x、4000x、4500x、5000x、5500x、6000x、6500x、7000x、7500x、8000x、8500x、9000x、9500x或10,000x。样品的测序深度可为约1x至约5x、约1x至约10x、约1x至约20x、约5x至约10x、约5x至约20x、约5x至约30x、约10x至约20x、约10x至约25x、约10x至约30x、约10x至约40x、约30x至约100x、约100x至约200x、约100x至约500x、约500x至约1000x、约1000x至约2000x、约1000x至约5000x或约5000x至约10,000x。测序深度可以是序列(例如,基因组)被测序的次数。在一些实施方案中,使用Lander/Waterman方程计算覆盖度。一般方程式可以是: $C=LN/G$,其中C=覆盖度;G=单倍体基因组长度;L=读取长度;而N=读取数目。

[0127] 在一些实施方案中,不同的条形码可以添加到不同样品中的多核苷酸(例如,通过使用引物或衔接子),并且不同的样品可以被合并,并在多重测定中分析。条形码可以允许确定多核苷酸来源的样品。

[0128] 在一些实施方案中,一种方法可包括使用生物标志物分析以及对健康状况的额外筛查测试。在一些实施方案中,一种方法可包括对额外筛查测试的结果不明确、为阳性或阴性的受试者进行生物标志物分析。额外筛查测试可以是对健康状况的预筛查测试。额外筛查测试可以是评估受试者出现健康状况的风险的测试。额外筛查方法可以在生物标志物分析之前、之后或与之结合进行。这种包含两种或更多种筛查方法的组合方法可以提高检测的准确度、灵敏度和/或特异性。此外,组合方法可用于增加早期癌症检测、引导对高风险受试者或具有致密乳腺组织和/或在如乳房X线照片等筛查测试中结果不明确的受试者进行额外筛查选择。

[0129] 多种额外筛查测试或方法适合与本公开内容的方法一起使用。这样的筛查测试的非限制性示例包括成像方法(使用例如,X射线、声波、放射性粒子,或磁场)、乳房X线照相术、乳腺核素成像(scintimammography)、乳腺检查(例如,临床和自身)、基因筛查(例如,BRCA测试)、超声、磁共振成像(MRI)、分子乳腺成像、活检、超声波扫描术、非侵入性诊断方法(例如,包括量化循环无细胞核酸,如与健康状况相关的DNA(例如,cfdDNA)或RNA(例如,cfrRNA))及其任何组合。在一些实施方案中,额外筛查测试是乳房X线照片。

[0130] 在一些实施方案中,额外的方法可以是活检。在一些实施方案中,额外筛查测试可以是基因筛查(例如,BRCA测试)。在一些实施方案中,额外筛查方法可以是非侵入性诊断方法,例如,包括量化循环无细胞核酸,如与健康状况相关的DNA(例如,cfdDNA)或RNA(例如,cfrRNA)。在一些实施方案中,循环无细胞核酸从生物流体生物样品中量化。在一些实施方案中,样品可以是,例如,血液、血浆、血清、尿液或粪便。在一些实施方案中,量化可以通过对无细胞核酸的高通量测序来实现。

[0131] 在一些实施方案中,额外筛查方法是乳腺癌的预筛查测试,如成像测试,例如,乳房X线照相术。例如,生物标志物分析可以与使用例如乳房X线照相术对乳腺癌高风险受试者进行的年度乳腺癌筛查或测试结合使用。在一些实施方案中,本公开内容的方法可包括基于唾液的生物标志物测定与乳房X线照片、计算机断层扫描(CT)、乳腺磁共振成像(MRI)扫描或其组合的结合,用于乳腺癌检测。

[0132] 可以调整从乳房X线照片获得的评分,或使用本公开内容的方法生成新评分。乳房

X线照片结果可表示为乳腺成像报告和数据系统 (BI-RADS) 评估类别 (即, BI-RADS 评分), 其范围可以从 0 (不完全) 至 6 (已知活检—证实为恶性)。乳房 X 线照片可以按 1-5 的等级评分 (1 = 正常, 2 = 良性, 3 = 不确定, 4 = 疑似恶性, 5 = 恶性)。例如, 基于生物标志物分析的结果, 乳房 X 线照片评分为 3 的受试者可以重新分类为 1。

[0133] 例如, 包括生物标志物分析和额外筛查测试或额外筛查测试结果的方法的检测灵敏度和/或特异性, 与用单独的筛查测试所获得的检测灵敏度和/或特异性相比, 可以增加。在一些实施方案中, 通过将受试者正确地鉴定为健康状况阴性, 可以增加或最大化特异性。例如, 通过对“回访者 (call-backs)” (例如, 正常但乳房 X 线照片不明确的患者) 使用本公开内容的组合方法, 将受试者正确地鉴定为乳腺癌阴性。在一些实施方案中, 通过将受试者正确地鉴定为健康状况阳性, 可以增加或最大化灵敏度。例如, 通过对具有高癌症风险或高密度乳腺组织的受试者使用本公开内容的组合方法, 将受试者正确地鉴定为乳腺癌阳性。

[0134] 本公开内容的方法可包括为受试者的健康状况生成风险评分。风险评分可以指示受试者出现健康状况的风险。风险评分可以基于生物标志物测定的结果计算。风险评分可以基于额外筛查测试的数据来计算、组合和/或调整。风险评分可以结合乳房 X 线照片结果提供, 并且组合信息可以用于确定, 例如, 患者患癌症的概率。

[0135] 本公开内容的方法可包括单独基于受试者的生物标志物特征 (例如, 从生物标志物分析的结果获得) 或基于受试者的生物标志物特征与额外筛查测试的结果的组合, 将受试者分类为两组或更多组。受试者可被分类为健康状况的阳性组 (例如, 乳腺癌阳性) 或阴性组 (乳腺癌阴性)。受试者可被分类为健康状况的高风险类别、低风险类别和中等风险类别。在一个示例中, 生物标志物特征可用于确定受试者患乳腺癌的风险较低, 并且可能不需要每年进行乳房 X 线照片筛查。在另一个示例中, 可以基于患者的生物标志物特征和预筛查测试, 将患者分类为乳腺癌高风险, 并可建议患者增加监测以进行癌症检测。

[0136] 本公开内容的方法可以提供可指示受试者当前实时状态的风险。实时状态可以与给定的疾病状态、疾病阶段、治疗相关特征或生理状况相关。由于风险可以反映受试者的当前状态, 因此可以在患者的一生中重复执行本公开内容的方法, 例如每年、每半年或每季度。例如, 高风险患者可以每季度执行本公开内容的方法。本公开内容的方法可与基因测试不同, 基因测试可能在受试者的一生中进行一次。基因测试 (例如, 乳腺癌基因测试, 如针对 BRCA1 或 BRCA2) 可以使用来自受试者的任何细胞进行, 并且可以代表终身风险。基因测试可能无法指示受试者当前的健康状态, 而本公开内容的方法可以在测试时确定风险。

[0137] 本公开内容的方法可以具有低假阳性率。在一些实施方案中, 本公开内容的方法的假阳性率可以为, 例如, 小于约 1%、小于约 2%、小于约 3%、小于约 4%、小于约 5%、小于约 6%、约 7%、小于约 8%、小于约 9%、小于约 10%、小于约 11%、小于约 12%、小于约 13%、小于约 14%、小于约 15%、小于约 16%、小于约 17%、小于约 18%、小于约 19% 或小于约 20%。

[0138] 本公开内容的方法的灵敏度可以为, 例如, 约 75%、约 80%、约 83%、约 85%、约 87%、约 90%、约 91%、约 92%、约 93%、约 94%、约 95%、约 96%、约 97%、约 98%、约 99%、约 99.5% 或 100%。本公开内容的方法的灵敏度可以为, 例如, 至少 75%、至少 80%、至少 83%、至少 85%、至少 87%、至少 90%、至少 93%、至少 95%、至少 96%、至少 97%、至少 98%、至少 99% 或至少 99.5%。本公开内容的方法的灵敏度可以为, 例如, 大于 75%、大于 80%、大于

83%、大于85%、大于87%、大于90%、大于93%、大于95%、大于96%、大于97%、大于98%、大于99%或大于99.5%。在一些实施方案中,本公开内容的方法的灵敏度为约83%。在一些实施方案中,本公开内容的方法的灵敏度大于83%。

[0139] 本公开内容的方法的特异性可以为,例如,约75%、约80%、约83%、约85%、约87%、约90%、约91%、约92%、约93%、约94%、约95%、约96%、约97%、约98%、约99%、约99.5%或100%。本公开内容的方法的特异性可以为,例如,至少75%、至少80%、至少83%、至少85%、至少87%、至少90%、至少93%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%。本公开内容的方法的特异性可以为,例如,大于75%、大于80%、大于83%、大于85%、大于87%、大于90%、大于93%、大于95%、大于96%、大于97%、大于98%、大于99%或大于99.5%。在一些实施方案中,本公开内容的方法的特异性为约97%。在一些实施方案中,本公开内容的方法的特异性大于97%。

[0140] 本公开内容的方法的准确度可以为,例如,约75%、约80%、约83%、约85%、约87%、约90%、约91%、约92%、约93%、约94%、约95%、约96%、约97%、约98%、约99%、约99.5%或100%。本公开内容的方法的准确度可以为,例如,至少75%、至少80%、至少83%、至少85%、至少87%、至少90%、至少93%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%。本公开内容的方法的准确度可以为,例如,大于75%、大于80%、大于83%、大于85%、大于87%、大于90%、大于93%、大于95%、大于96%、大于97%、大于98%、大于99%或大于99.5%。在一些实施方案中,本公开内容的方法的准确度为约90%。在一些实施方案中,本公开内容的方法的准确度大于97%。

[0141] 在一些实施方案中,组合的一组基因给予大于70%、75%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或99.5%的特异性或灵敏度,和/或至少70%、75%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、95.5%、96%、96.5%、97%、97.5%、98%、98.5%、99%、99.5%或更高的准确度。

[0142] 本公开内容的方法可以具有高信噪比,这可以有助于区分肿瘤谱。

[0143] 受试者可以是人、患者、非人灵长类动物,如黑猩猩,及其他猿和猴类;农场动物,如牛、马、绵羊、山羊、猪;家养动物,如兔、狗和猫;实验动物,包括啮齿动物,如大鼠、小鼠和豚鼠等。受试者可以是任何年龄。受试者可以是,例如,男性、女性、老年人、成年人、青少年、青春期前儿童、儿童、幼儿、婴儿。

[0144] 受试者可以是,例如,10至90岁。受试者可以是,例如,10至60岁、18至25岁、18至30岁、18至35岁、18至40岁、18至45岁、18至50岁、18至55岁、18至60岁、18至65岁、18至70岁、20至25岁、20至30岁、20至35岁、20至40岁、20至45岁、20至50岁、20至55岁、20至60岁、20至65岁、20至70岁、25至30岁、25至35岁、25至40岁、25至45岁、25至50岁、25至55岁、25至60岁、25至65岁、25至70岁、30至35岁、30至40岁、30至45岁、30至50岁、30至55岁、30至60岁、30至65岁、30至70岁、35至40岁、35至45岁、35至50岁、35至55岁、35至60岁、35至65岁、35至70岁、40至45岁、40至50岁、40至55岁、40至60岁、40至65岁、40至70岁、45至50岁、45至55岁、45至60岁、45至65岁、45至70岁、50至55岁、50至60岁、50至65岁、50至70岁、55至60岁、55至65岁、55至70岁、60至65岁、60至70岁或65至70岁。在一些实施方案中,受试者可以在18至40岁之间。在一些实施方案中,受试者可以小于40岁。在一些实施方案中,受试者可以小于35岁。在一

些实施方案中,受试者可以小于50岁。在一些实施方案中,受试者可以小于60岁。在一些实施方案中,受试者可以小于70岁。

[0145] 受试者可能患有先前存在的疾病或病况,如癌症。或者,受试者可能没有任何已知的先前存在的病况。受试者也可能对现有或过去的治疗(如癌症治疗)无反应。受试者可能正在经历癌症治疗,例如化疗。

[0146] 在一些实施方案中,受试者可具有高密度乳腺组织或致密乳腺组织。在一些实施方案中,受试者可以是高风险受试者,例如BRCA1和/或BRCA2携带者。对于健康状况的预筛查测试,受试者可以具有阳性、阴性或不明确的结果。受试者可具有阳性、阴性或不明确的乳房X线照片结果。受试者可具有不明确的乳房X线照片结果和致密的乳腺组织。

[0147] 由乳腺成像报告和数据库系统或BI-RADS分类的受试者的乳腺密度可以是主要为脂肪、分散密度、一致密度或极度致密。主要为脂肪的分类可以指示乳腺主要由脂肪组成,并包含很少的纤维和腺体组织。这可能意味着乳房X线照片可以显示出任何异常。分散密度的分类可以指示乳腺具有许多脂肪,但也有一些纤维和腺体组织区域。一致密度的分类可以指示乳腺具有许多纤维和腺体组织区域,这些区域均匀分布在乳房中。这可能使在乳腺中很难看到小肿块。极度致密的类别可以指示乳房具有大量纤维和腺体组织。这可能使乳房X线照片很难看到癌症,因为癌症可以与正常组织融合。在一些实施方案中,受试者可具有极度致密的乳腺。

[0148] 受试者数据(例如,与年龄、性别、种族、身体状况、乳腺癌类型或阶段以及乳腺组织密度相关的数据)的组合可以与本公开内容的方法一起使用。

[0149] 多种计算机架构适用于本公开内容。图5为示出计算机架构系统(500)示例的框图。计算机系统500可结合本公开内容的示例实施方案使用。如图5所示,示例计算机系统可包括用于处理指令的处理器(502)。处理器的非限制性示例包括: Intel Core i7™、Intel Core i5™、Intel Core i3™、Intel Xeon™、AMD Opteron™、Samsung 32-bit RISC ARM 1176JZ(F)-S v1.0™、ARM Cortex-A8 Samsung S5PC100™、ARM Cortex-A8 Apple A4™、Marvell PXA 930™或功能相当的处理器。可使用多个执行线程来并行处理。在一些实施方案中,可以使用多个处理器或多核的处理器。在一些实施方案中,多个处理器或多核的处理器可用于单个计算机系统中、集群中或通过网络跨系统分布。在一些实施方案中,多个处理器或多核的处理器可以通过包括多个计算机、蜂窝电话和/或个人数据助理设备的网络跨系统分布。

a. 数据采集、处理和存储

[0150] 高速缓存(501)可以连接至或并入处理器(502),以便为处理器(502)最近或经常使用的指令或数据提供高速存储。处理器(502)通过处理器总线(505)连接至北桥(506)。北桥(506)通过存储器总线(504)连接至随机存取存储器(RAM)(503),并管理处理器(502)对RAM(503)的访问。北桥(506)还通过芯片组总线(507)连接至南桥(508)。南桥(508)转而连接至外围总线(509)。外围总线可以是,例如,PCI、PCI-X、PCI Express或另一种外围总线。北桥和南桥,通常称为处理器芯片组,管理处理器、RAM和外围总线(509)上的外围组件之间的数据传输。在一些计算机架构系统中,北桥的功能可以合并至处理器中,而不使用单独的北桥芯片。

[0151] 在一些实施方案中,计算机架构系统(500)可包括加速器卡(512)。在一些实施方

案中,计算机架构系统(500)可包括附接至外围总线(509)的加速器卡。在一些实施方案中,加速器卡(512)可包括现场可编程门阵列(FPGA)或用于加速处理的其他硬件。

b. 软件接口

[0152] 软件和数据存储在外部存储模块(513)中,并且可以加载到RAM(503)和/或缓存(501)中,以供处理器使用。计算机架构系统(2300)可包括用于管理系统资源的操作系统。操作系统的非限制性示例包括:Linux、Windows™、MACOS™、BlackBerry OS™、iOS™和其他功能等同的操作系统。在一些实施方案中,操作系统可以是在操作系统上运行的应用软件。

[0153] 在图5中,计算机架构系统(500)进一步包括网络接口卡(NIC)(510和511),它们连接至外围总线以向外部存储器提供网络接口。在一些实施方案中,网络接口卡是网络附加存储(NAS)设备或可用于分布式并行处理的另一计算机系统。

c. 计算机网络

[0154] 图6为显示了计算机网络(600)的图,其具有多个计算机系统(602a和602b)、多个蜂窝电话和个人数据助理(602c)以及NAS设备(601a和601b)。在一些实施方案中,系统602a、602b和602c可以管理数据存储并为NAS设备(601a和602b)上存储的数据优化数据访问。数学模型可用于使用跨计算机系统(602a和602b)以及蜂窝电话和个人数据助理系统(602c)的分布式并行处理来评估数据。计算机系统(602a和602b)以及蜂窝电话和个人数据助理系统(602c)还可以为存储在NAS设备(601a和601b)上的数据的自适应数据重构提供并行处理。

[0155] 图6仅示出了一个示例,并且多种其他计算机架构和系统可以结合本公开内容的各种实施方案使用。例如,刀片服务器可用于提供并行处理。处理器刀片可以通过背板连接,以提供并行处理。存储也可以通过单独的网络接口连接至背板或NAS设备。

[0156] 在一些实施方案中,处理器可以维持单独的存储空间,并通过网络接口、背板或其他连接器传输数据,以供其他处理器并行处理。在一些实施方案中,一些或所有处理器可以使用共享虚拟地址存储空间。

d. 虚拟系统

[0157] 图7为使用共享虚拟地址存储空间的多处理器计算机系统的框图。该系统包括可以访问共享存储器子系统(702)的多个处理器(701a-701f)。该系统在存储器子系统(702)中并入多个可编程硬件存储器算法处理器(MAP)(703a-703f)。每个MAP(703a-703f)可包括存储卡(704a-704f)和一个或多个现场可编程门阵列(FPGA)(705a-705f)。MAP提供可配置的功能单元。算法或算法的部分可以提供给FPGA(705a-705f),以便与相应的处理器密切协作进行处理。在一些实施方案中,所有处理器可全局访问每个MAP。在一些实施方案中,每个MAP可以使用直接存储器存取(DMA)来访问相关联的存储卡(704a-704f),从而允许其相对于相应的微处理器(701a-701f)独立且异步地执行任务。在一些这种配置中,MAP可以将结果直接提供给另一个MAP,以进行算法的流水线操作和并行执行。

[0158] 上述计算机架构和系统仅为示例,并且多种其他的计算机、蜂窝电话和个人数据助理架构和系统可以结合示例实施方案使用。在一些实施方案中,本公开内容的系统可以使用通用处理器、协处理器、FPGA和其他可编程逻辑器件、片上系统(SOC)、专用集成电路(ASIC)以及其他处理和逻辑元件的任何组合。许多种数据存储介质可以结合示例实施方案使用,包括RAM、硬盘驱动器、闪存、磁带驱动器、磁盘阵列、NAS设备和其他本地或分布式数

据存储设备和系统。

[0159] 在一些实施方案中,计算机系统可以使用在上述的任何计算机架构和系统上执行的软件模块来实现。在一些实施方案中,系统的功能可以部分或完全地在固件或可编程逻辑器件(例如,FPGA)(如图7所示)、片上系统(SOC)、专用集成电路(ASIC)或其他处理和逻辑元件中实现。例如,可以使用硬件加速卡,如图5所示的加速器卡(512),通过硬件加速来实现集处理器和优化器。

[0160] 本文描述的本公开内容的任何实施方案均可以例如由在同一地理位置的用户产生和传输。本公开内容的产品可以例如从一个国家的地理位置产生和/或传输,而本公开内容的用户可以在另一个国家。在一些实施方案中,由本公开内容的系统访问的数据可以是计算机程序产品,其可以从多个地理位置(801)之一传输到用户(802)。图8示出了从地理位置传输到用户的计算机程序产品。本公开内容的计算机程序产品所生成的数据可以在多个地理位置之间来回传输。在一些实施方案中,本公开内容的计算机程序产品所生成的数据可以通过网络连接、安全网络连接、不安全网络连接、互联网连接或内联网连接来传输。在一些实施方案中,本文的系统在实体产品和有形产品上编码。

实施例

[0161] 通过参考以下实施例进一步详细描述本发明。提供这些实施例仅用于说明目的,除非另有说明,否则并非旨在是限制性的。因此,本发明不应被解释为限于以下实施例,而应被解释为包括由于本文提供的教导而变得明显的任何和所有变化。

实施例1:鉴定生物标志物

[0162] 使用公开可用的数据分析基因共表达网络。在许多条件下发现了18种可能的生物标志物。发现了针对包括乳腺癌、结肠癌、肺癌、神经退行性疾病和炎症病症在内的病况的生物标志物。

实施例2:基因表达水平分析

[0163] 使用获自10名乳腺癌受试者和10名匹配的健康对照的唾液样品来分析基因表达水平。该研究使用来自10名患者和10个对照样品的发现数据集的微阵列数据,为基于唾液的乳腺癌检测提供了概念验证。样品包括例如种族、BRCA和非BRCA、致密样品和非致密样品的混合群体。

[0164] 生物标志物发现阶段分析研究鉴定了约8800个相关基因,其可用于确定微阵列(例如,Affymetrix HG-U133Plus 2.0基因芯片)上模块间的平均连通性。检查源自这些基因的模块的平均连通性,以确定平均连通性是否产生具有高灵敏度、高特异性和高统计学显著性的生物标志物特征。该研究产生的结果的准确度为约90%。

[0165] 图9示出了源自10名乳腺癌受试者和10名匹配的健康对照的平均连通性值。衍生出这些值的基因表达微阵列数据从NCBI基因表达综合数据库GSE 20266获得。通过t-检验在乳腺癌受试者与对照受试者之间进行比较,得到约0.002的p值。两组之间的虚线以约90%的准确度将受试者在两个方向上分开。

[0166] 然后检查对结果贡献最大的基因表达模块,并对基因表达模块具有最高显著性的个体基因进行分析。对基因表达模块具有最高显著性的个体基因进行分析以产生最有效的子网络,从而在研究中产生对照组与乳腺癌组的分离。最有效的子网络包括含有9个重要基因的4个模块,如图4所示。图4示出了在产生平均连通性的分离中所涉及的主要子网络,该

平均连通性源自鉴定了约8800个基因的微阵列数据。模块1 (401) 包括SLC25A51 (也称为MCART1) 和LCE2B; 模块2 (402) 包括HIST1H4K和ABCA2; 模块3 (403) 包括TNFRSF10A、AK092120和DTYMK; 而模块4 (404) 包括Hs.161434和ALKBH1。在说明性实施例中, “9基因生物标志物测定”或“9基因生物标志物小组”可包含在该实施例中鉴定并在图4中示出的生物标志物基因中的一种或多种。

[0167] 与将网络视作整体相比, 该子网络内的相关性可以在更高程度上反映表型差异。随着生物标志物小组减少至9个基因, 可以使用例如qPCR检查基因表达。与例如使用微阵列 (例如, Affymetrix基因芯片) 相比, 采用例如qPCR进行基因表达检测可以更便宜且更可扩展。

实施例3: 验证所鉴定的生物标志物

[0168] 针对9基因生物标志物小组 (例如, 图4中的基因), 在60个患者样品上进行了实施例2中鉴定的生物标志物的初步验证。样品包括例如种族、BRCA和非BRCA、致密样品和非致密样品的混合群体。当采用乳房X线照片存在困难时, 如存在致密乳腺组织时, 使用来自9基因测定的数据指导进一步筛查的需要, 从而大大提高检测率。

[0169] 图10示出了从使用qPCR进行的9基因测定获得的评分, 取自60例受试者的验证研究。验证研究包括30名经鉴定患有浸润性导管癌 (IDC) 的乳腺癌受试者和30名健康对照受试者。图10所示的结果验证了实施例2所使用的方法。该测定具有约83%的灵敏度。该测定的特异性为约97%, 相比之下, 乳房X线照片的特异性水平为90%。

[0170] 来自60名患者的研究的数据表明, 基于唾液的乳腺癌检测的9基因测定具有约90%的总体准确度, 约83%的灵敏度和约97%的特异性。基于这些结果, 9基因测定能够检测出约83%的患癌女性。

[0171] 该初始验证研究的结果显示, 生物标志物值超越了技术平台 (例如, qRT-PCR和微阵列)。检测显示了在诊断测试范围内的灵敏度和特异性水平。

[0172] 图11示出了连续排序的综合基因表达值。数据呈现了乳腺癌受试者与对照受试者 (30名患者的复制组) 的良好分离。该数据获自60名患者的队列, 针对9基因生物标志物测定连续排序。

[0173] 在包括120名患者和120名对照的大型队列组上进行二次验证研究。样品包括例如种族、BRCA阳性受试者、BRCA阴性受试者、致密样品和非致密样品的混合群体。针对9种生物标志物基因和2种管家基因中的每一种, 在图12-25示出该研究的数据。

[0174] 图12示出了生物标志物基因5的二次验证研究的结果。该数据来自包括120名患者和120个对照样品的大型队列研究。结果显示出癌症患者与对照患者的明显分离。对于来自9基因生物标志物小组的9种生物标志物基因中的5种获得了类似的结果。

[0175] 图13A-D至图18示出了对于9种说明性生物标志物基因的基于RT-qPCR的二次验证研究的结果。该数据来自包括120名患者和120个对照样品的大型队列研究。

[0176] 图13A显示了基因2的基于RT-qPCR的二次验证研究的结果。基因2是唾液样品中乳腺癌的最大基因贡献者之一。数据显示出癌症患者与对照患者的良好分离, 表现出的特异性为84.2%, p值小于0.0001。图14显示了基因2的生物标志物验证研究的参数和结果。

[0177] 图13B显示了基因3的基于RT-qPCR的二次验证研究的结果。基因3是唾液样品中乳腺癌的最大基因贡献者之一。数据显示出癌症患者与对照患者的良好分离, 表现出的p值小

于0.0001。图15显示了基因3的生物标志物验证研究的参数和结果。

[0178] 图13C显示了基因7的基于RT-qPCR的二次验证研究的结果。基因7是唾液样品中乳腺癌的最大基因贡献者之一。数据显示出癌症患者与对照患者的良好分离,表现出的灵敏度为60.8%,特异性为94.2%,p值小于0.0001。图16显示了基因7的生物标志物验证研究的参数和结果。

图13D显示了基因9的基于RT-qPCR的二次验证研究的结果。基因9是唾液样品中乳腺癌的最大基因贡献者之一。数据显示出癌症患者与对照患者的良好分离,表现出的灵敏度为72.5%,特异性为85%,p值小于0.0001。图17显示了基因9的生物标志物验证研究的参数和结果。

[0179] 图18A显示了基因1的基于RT-qPCR的二次验证研究的结果。基因1不是唾液样品中乳腺癌的最大基因贡献者之一。数据显示出癌症患者与对照患者的良好分离,p值为0.0167。图19显示了基因1的生物标志物验证研究的参数和结果。

[0180] 图18B显示了基因4的基于RT-qPCR的二次验证研究的结果。基因4不是唾液样品中乳腺癌的最大基因贡献者之一。数据显示出癌症患者与对照患者的良好分离,表现出的灵敏度水平为81.7%,特异性水平为41.7%,p值小于0.0001。图20显示了基因4的生物标志物验证研究的参数和结果。

[0181] 图18C显示了基因5的基于RT-qPCR的二次验证研究的结果。基因5不是唾液样品中乳腺癌的最大基因贡献者之一。数据显示出癌症患者与对照患者的良好分离,表现出的灵敏度水平为50.8%,特异性水平为74.2%,p值为0.0014。图21显示了基因5的生物标志物验证研究的参数和结果。

[0182] 图18D显示了基因6的基于RT-qPCR的二次验证研究的结果。基因6不是唾液样品中乳腺癌的最大基因贡献者之一。数据显示出癌症患者与对照患者的良好分离,表现出的灵敏度水平为63.3%,特异性为63.3%,p值为0.0001。图22显示了基因6的生物标志物验证研究的参数和结果。

[0183] 图18E显示了基因8的基于RT-qPCR的二次验证研究的结果。基因8不是唾液样品中乳腺癌的最大基因贡献者之一。数据显示出癌症患者与对照患者的良好分离,表现出的灵敏度为约85%,特异性为58.5%,p值小于0.0001。图23显示了基因8的生物标志物验证研究的参数和结果。

[0184] 图24显示了管家基因G-H1的基于RT-qPCR的二次验证研究的结果。数据显示出癌症患者与对照患者的良好分离,表现出的灵敏度为96.7%,特异性为25.8%,p值为0.1551。

[0185] 图25显示了管家基因G-H2的基于RT-qPCR的二次验证研究的结果。数据显示出癌症患者与对照患者的良好分离,表现出的灵敏度为84.2%,特异性为30.8%,p值为0.0355。

[0186] 表2显示了测定9种生物标志物基因和2种管家基因的引物。数据表明,多个基因在大型队列研究中显示出个体显著性。当生物标志物与其他生物标志物组合使用时,初始分析显示具有显著性。数据显示,9基因生物标志物小组的基因2、基因7和基因9对测试的特异性贡献最大,例如通过正确地排除癌症或正确地鉴定阴性样品为正常。9基因生物标志物小组的基因4和基因7对灵敏度贡献最大,例如通过正确地排除正常样品或正确地鉴定阳性样品为癌症。使用Medcalc软件计算测试的灵敏度和特异性。通过将测试与乳房X线照片伴随进行,可以进一步提高方法的灵敏度和特异性。

表2

		5' - 3'	长度	开始	停止	Tim	GC %	自身互补性
基因 1	F1	CCTCCTAGAGGGTTGTAT TGCC	22	230	251	59.9	54.55	6
AI424847	R1	AGGGAGGCTTGGAAGAG AGAA	21	331	311	60.2	52.38	4
243218_at	F2	CAGTCGGAACCACATCCT CC	20	177	196	60.11	60	3
	R2	CTGTCTGACTGGTGGTCA CAT	21	385	365	59.65	52.38	5
基因 2	F1	CTTGTCATAGCCAAGCAC GC	20	63	82	59.9	55	6
AW451259	R1	TGTCCCATGGAGGTAGGG AG	20	288	269	60.03	60	6
241371_at	F2	GCTTCTTGATTGACCTG GC	20	84	103	59.19	55	3
	R2	GCCTCGTGGTTCAATCCT CC	20	253	234	60.74	60	4
基因 3	F1	GGGGAACCTAACCGAGTC CT	20	302	321	60.62	60	3

BF724944	R1	GATTACTTGGGGCACGCT GT	20	480	461	60.68	55	2
23874_at	F2	TCCTGAGAGTGTAAGCCA GC	20	318	337	59.1	55	3
	R2	GACAGGATGATCCAGCCC TT	20	569	550	59.16	55	7
基因 4	F1	GTGGAGGTTTCAGGACCAA GG	20	1231	1250	59.96	60	5
AL162060	R1	CTCTCGGCATCAGCCTTC AT	20	1520	1501	59.89	55	3
212772 x at	F2	CGCCCCTAATTGTGCCAA AG	20	1307	1326	59.83	55	4
	R2	TCTCGGCATCAGCCTTCA TC	20	1519	1500	59.89	55	3
基因 5	F1	GCATCTCCGGCCTCATCT AC	20	137	156	60.04	60	4
NM_021968	R1	AGAAAGGGACGCTCAACC AC	20	324	305	60.25	55	3
208580 x at	F2	CGCCGTGACCTATACAGA GC	20	207	226	60.32	60	4
	R2	CACCGAAACCGTAGAGGG TG	20	307	288	60.39	60	4
基因 6	F1	CCGACTGCTGTGAGAGTG AA	20	320	339	59.68	55	4
NM_014357	R1	AGGTGCTCCATCAAGTGC AA	20	545	526	59.89	50	4
207710_at	F2	ACAGCCTGATGCTTAACC CTT	21	434	454	59.64	47.62	5
	R2	GTGCTCCATCAAGTGCAA AGT	21	543	523	59.39	47.62	4
基因 7	F1	AAACGGAGACCCCGAAGT TT	20	481	500	59.53	50	5
NM_006020	R1	CCACCCAGGAGAAAGATG GC	20	782	763	60.39	60	3
205621_at	F2	ACCTTTCCCTTCTGACCTG G	20	579	598	58.93	55	3
	R2	AGCTGAATGACAGCAAGG GT	20	751	732	59.6	50	4
基因 8	F1	TGGAGGTCTTTGCCACCA AT	20	1949	1968	59.52	50	4
AK092120	R1	AATTGCCTCTTCTGCTCCC C	20	2094	2075	60.03	55	4
1566840_at	F2	CACCAATGGGAGATGAGC CA	20	1962	1981	59.74	55	5
	R2	GCTAAATGGCCCTCTCCT CC	20	2072	2053	59.89	60	4
基因 9	F1	TGTAAACAGCCAGACGTG GG	20	374	393	60.25	55	4
AK022132	R1	CATAGCGCTCATGGCCAA AC	20	479	460	59.97	55	8
1565694_at	F2	GTGTAAACAGCCAGACGT GG	20	373	392	59.13	55	4

	R2	CTGGAAAGCCCCGTCTC AT	20	496	477	60.04	55	3
管家基因								
基因 1-1	F1	ACCGGCACCATCAAGCT	17			54.9	59	
TFRC	R1	TGATCACGCCAGACTTTG C	19			57.5	53	
	F2	CTCGTGAGGCTGGATCTC AAA	21	744	764	59.79	52.38	4
	R2	TCACGCCAGACTTTGCTG AG	20	834	815	60.6	55	3
基因 1-2	F1	CCATCATGAAGTGTGACG TGG	21	926	946	61.2	52	
ATCB	R1	GTCCGCCTAGAAGCATTT GCG	21	1198	1218	63.2	57	
	F2	TTGCCGACAGGATGCAGA AGGA	23	1010	1031	64.2	55	
	R2	AGGTGGACAGCGAGGCC AGGAT	23	1117	1138	67.9	64	
基因 1-3	F1	AGGCTGGGGCATACAAAA CCA						
AB032983	R1	CGCCATCCTCTGTCCTTCA G	20	1877	1858	60.18	60	3
212686_at	F2	GCCCGGGTAATGGCAACT AT	20	1552	1571	60.18	55	6
	R2	CGTCCCAGAGTCCATCAG TG	20	1738	1719	59.83	60	3

实施例4：将生物标志物小组测定中基因的转录水平与肿瘤学的已知参与相关联

[0187] 生物标志物水平 (例如, 来自实施例2的9种生物标志物基因的转录水平) 与已知参与乳腺癌形成和进展的那些基因相关联。在癌症与健康受试者以及与年龄、种族、身体状况、乳腺癌类型或阶段和乳腺组织密度相关的亚类之间检查与生物标志物相关的倍数变化差异。该信息用于指导针对已知参与乳腺癌形成和进展的基因和相关途径的基因本体论搜索。基于该信息, 确定生物标志物表达水平的排名和权重, 以提高测试的灵敏度。

[0188] 获得30名患者 (15名癌症患者和15名对照受试者) 的患者信息和唾液样品。唾液样品在Oragene RE-100杯中运输, 该杯含有RNA酶抑制剂, 以使唾液中的RNA在室温下稳定60天。从唾液中提取RNA。qPCR在每个样品上一式两份进行。基于倍数变化差异和p值 (t-检验), 在健康受试者与癌症受试者之间检查基因小组 (例如, 实施例2中鉴定的9种生物标志物基因) 的基因表达水平的差异。从基因小组获得的数据用作基线。

[0189] 然后分析与患者信息相关的亚类中基因表达的变化。亚类的基因表达的变化用于指导基因本体论搜索。例如, 如果年龄从基线产生最大的差值, 则分析9种生物标志物基因与涉及与乳腺癌有关的年龄相关途径的基因的关系。使用基因本体论工具 (例如, NCBI的AmiG02和Gene) 进行基因本体论搜索。对健康受试者-癌症受试者基线的三个最大差值进行基因本体论搜索程序。基于三个指导的基因本体论搜索的结果, 计算了三种排名和加权方案。然后, 对30个样品的未加权评分测试三种加权方案的准确度、灵敏度和特异性, 并比较结果。加权方案将灵敏度提高到超过90%, 改善了测定的总体准确度, 并将特异性水平保持

在97%或以上。

实施例5:检查乳腺癌相关外来体的mRNA内容物

[0190] 检查了在有和没有标准照护化疗剂的培养物中生长的永生化乳腺癌细胞系(例如,MDA-MB-231和MCF7)所释放的外来体的mRNA内容物。将从外来体的mRNA内容物获得的数据用于进一步改进基因表达值的加权并改善测试结果量度。

[0191] MDA-MB-231和MCF7永生化乳腺癌细胞系可以将含有mRNA的外来体样小泡释放到生长培养基中。在用和不用多柔比星(即,标准照护化疗剂)培养的永生化乳腺癌细胞系(例如,MDA-MB-231和MCF7)所释放的外来体中检查生物标志物小组中基因(例如,实施例2中鉴定的9种生物标志物基因)的转录水平。使用标准实验室技术如qPCR分析样品。分析表达水平的差异,例如,如实施例4中所讨论的。基于该分析,得出改进的排名和加权方案。使用实施例4中来自30个样品的数据,相对于实施例4的加权评分方案和未加权评分方案测试改进加权方案的准确度、灵敏度和特异性。然后比较结果。

[0192] 在用和不用多柔比星培养的细胞中,观察到生物标志物小组(例如,实施例2的9基因测定)中基因的表达水平的差异。基因表达水平的差异可以为涉及唾液中乳腺癌检测的外来体机制提供进一步的证据。改进的加权方案可以将灵敏度提高到90%以上,并且在不显著影响测定特异性的情况下改善生物标志物测定的总体准确度。

实施例6:使用盲的患者样品(n=30)测试生物标志物小组的预测能力

[0193] 使用实施例4和实施例5中收集的任何新的加权优化,如实施例4所示对盲的唾液样品(例如,30个癌症、对照信息未知的样品)进行分析和评分。

实施例7:基于唾液的诊断测定的工作流程

[0194] 图26示出了唾液基因测试的优化工作流程。在30分钟内将50mL唾液收集至50 mL收集管中,并将该管运送至诊断实验室(2601)。样品在4℃以2600 g离心15分钟。收集上清液。每mL唾液上清液添加5μL(即,100个单位)的超级酶抑制剂,并储存样品(2602)。然后从唾液样品中分离RNA(2603)。解冻唾液上清液样品。将200μL解冻的样品直接转移至样品管中。根据标准MagNA方案分离总RNA。将RNA样品储存在-80℃下。使用表3和表4所示的实验参数,在一步式反应中逆转录并预扩增RNA(2604)。

表3

	每管	管数	总计
StarScript II RT Mix	1	96	96
引物3管家基因+9对(各100μM)	1.44	96	138.24
2x反应混合物	10	96	960
mRNA	4	96	
H ₂ O	3.56	96	341.76
总体积	20		

表4

时间	温度	循环	
1 min	60	1	RT
30 min	50		
2 min	95		
15 sec	95	15	预扩增
30 sec	50		
10 sec	60		
10 sec	72		
10 min	72	1	延伸
保持	4	1	

[0195] 扩增后,使用ExoSAP-IT处理纯化扩增产物,例如,以去除残留在PCR产物混合物中的未消耗的dNTP和引物,它们可能干扰下游应用(例如,qPCR和测序)。纯化后,将cDNA稀释约40倍。使用表5和表6所示的实验条件,在一步式反应中进行qPCR (2605)。

表5

	每管	管数	总计
内引物混合物 (100 μ M/每个)	0.18	96	17.28
2x qPCR混合物	5	96	480
水	0.82	96	78.72
cDNA	4		
总体积	10		

表6

时间	温度	循环	
10 min	95	1	热启动
15 sec	95	40	qPCR
30 sec	60		
15 sec	95	1	解离
1 min	60		
15 sec	95		
15 sec	60		

实施例8:针对乳腺癌相关基因评估唾液中的基因表达谱

[0196] 使用从10名乳腺癌患者和10名正常患者的唾液中收集的RNA,计算与癌症和正常表型的相关性。这用于确定癌症样品与正常样品之间差异表达的基因。图29以热图的形式示出了研究结果。在图29中,前10列显示了来自癌症患者唾液的数据,后10列显示了来自正常样品唾液的数据。每行中的框显示了20名患者的基因表达。蓝框表示基因表达下降。红框表示基因表达上升。

[0197] 使用P值截止值为0.05的检验Kolmogorov-Smirnoff统计学,分析确定为差异表达的基因,以确定其在癌症标志(例如,如GO本体论注释的)中的富集。图28示出了发现为差异表达并对应于癌症标志的基因。这些包括TNFRSF10A、ABCA1/2、DTYMK和ALKBH1,它们在实施例2中独立地鉴定为候选生物标志物。因此,本研究证实,在实施例2中鉴定并在图4显示的

候选基因可用作从唾液检测乳腺癌的生物标志物。

实施例9:乳腺癌诊断试验

[0198] 使女性受试者经历乳房X线照片。受试者被告知她有致密的乳腺组织。乳房X线照片显示癌症的阴性指征。由于受试者具有致密的乳腺组织,因此医疗保健提供者建议进行乳腺癌生物标志物测定(例如,实施例4中描述的9基因测定)。

[0199] 根据医疗保健提供者的建议,受试者将唾液吐入杯中。使用本公开内容的方法分析唾液样品,以确定来自生物标志物小组的基因的转录水平。基于对来自生物标志物测定的数据的分析,给予受试者诊断。

实施例10:伴随诊断

[0200] 图2示出了使用基于唾液的生物标志物测定结合乳房X线照片成像进行准确的癌症诊断。使女性受试者(201)经历乳房X线照片(202)并向医疗保健提供者提交唾液样品(204)。分析乳房X线照片(203)以检测癌症。同时,使用本公开内容的方法分析唾液样品(205)以确定来自生物标志物小组的基因的转录水平(例如,实施例4中描述的9基因测定)。基于对乳房X线照片和来自生物标志物测定的数据的分析给予受试者诊断(206)。

实施例11:乳腺癌诊断

[0201] 受试者希望每年进行乳腺癌筛查。受试者在成像前通过邮件或亲自提供唾液样品。针对生物标志物小组分析受试者的样品。基于结果(例如,通过邮件或亲自交流),生物标志物将受试者划分到风险类别中。这些类别可用于指明癌症风险和随诊频率。如果唾液结果指示很高的风险,该结果还可以提供通过乳房X线照片、MRI和/或更密切的监测来进行额外筛查的建议。

实施例12:乳腺癌诊断

[0202] 受试者来到医疗保健提供者处进行年度筛查性乳房X线照片,并同时提供唾液样品。对这两项测试分别进行了分析。将结果组合以产生癌症的单个组合概率评分,该评分可以是比任一单独测试更可靠的乳腺癌风险估计。

实施例13:乳腺癌诊断

[0203] 受试者来到医疗保健提供者处进行年度筛查,并获得具有“不明确结果”读取的乳房X线照片。大约1/7的乳房X线照片可能是不明确的。受试者提供唾液样品,使用本公开内容的生物标志物测定分析该唾液样品。将生物标志物测定和乳房X线照片的结果组合,以使受试者优先考虑持续的随诊,如重复进行乳房X线照片、MRI、活检或增加通过唾液样品测试或乳房X线照片或两者进行监测的频率。

实施例14:癌症筛查

[0204] 受试者前来筛查并提供样品。分析样品,并且测试结果鉴定患者体内的癌症。受试者经历后续测试,如本公开内容的生物标志物测定,以将癌症定位到特定的身体组织,如乳腺。

实施例15:避免乳房X线照片

[0205] 受试者希望避免乳房X线照片。乳房X线照片对于例如乳腺致密的受试者,或年轻的受试者(例如,年龄范围在18至40岁、或低于40岁、或低于35岁),或乳腺癌高风险受试者可能具有较高的假阴性和假阳性率。年轻的受试者具有致密乳腺的频率可能更高。该受试者34岁并具有致密的乳腺组织。使该受试者经历本公开内容的基于唾液的生物标志物测

定。向该受试者给予乳腺癌的风险评分,进行额外测试的建议以及未来筛查的频率。

实施方案

[0206] 以下非限制性实施方案提供了本发明的说明性示例,但不限制本发明的范围。

[0207] 实施方案1.一种方法,其包括:

a) 对受试者进行筛查测试,其中所述筛查测试包括评估所述受试者发生健康状况的风险;

b) 获得所述受试者的生物样品;

c) 量化所述受试者的所述生物样品中生物标志物的样品水平;

d) 将所述生物标志物的所述样品水平与所述生物标志物的参考水平进行比较;

e) 将所述筛查测试的结果与所述比较组合;以及

f) 基于所述组合确定所述受试者的健康状态。

[0208] 实施方案2.如实施方案1所述的方法,其中所述筛查测试包括对所述受试者的乳腺组织进行成像。

[0209] 实施方案3.如实施方案1-2中任一项所述的方法,其中使用乳房X线照片进行所述成像。

[0210] 实施方案4.如实施方案1-3中任一项所述的方法,其中所述筛查测试包括量化所述受试者中无细胞核酸的样品水平。

[0211] 实施方案5.如实施方案1-4中任一项所述的方法,其中所述无细胞核酸是无细胞RNA。

[0212] 实施方案6.如实施方案1-4中任一项所述的方法,其中所述无细胞核酸是无细胞DNA。

[0213] 实施方案7.如实施方案1-6中任一项所述的方法,其中所述无细胞核酸对所述受试者的组织是特异的。

[0214] 实施方案8.如实施方案1-7中任一项所述的方法,其中所述组织是乳腺组织。

[0215] 实施方案9.如实施方案1-7中任一项所述的方法,其中所述无细胞核酸来自生物流体。

[0216] 实施方案10.如实施方案1-7和9中任一项所述的方法,其中所述生物流体选自:血液、血成分、血清、血浆、唾液、痰液、尿液、精液、经阴道液体、脑脊液、汗液、胆汁、囊液、泪液、乳腺吸出液和乳腺液。

[0217] 实施方案11.如实施方案1-10中任一项所述的方法,其中所述筛查测试包括基因测试。

[0218] 实施方案12.如实施方案1-11中任一项所述的方法,其中所述基因测试包括测试乳腺癌易感基因中的突变。

[0219] 实施方案13.如实施方案1-12中任一项所述的方法,其中所述基因测试包括测试选自以下的基因中的突变:ATM、BARD1、BRCA1、BRCA2、BRIP1、CASP8、CDH1、CHEK2、CTLA4、CYP19A1、FGFR2、H19、LSP1、MAP3K1、MRE11、NBN、PALB2、PTEN、RAD51、RAD51C、STK11、TERT、TOX3、TP53、XRCC2、XRCC3及其任何组合。

[0220] 实施方案14.如实施方案1-13中任一项所述的方法,其中所述健康状况是癌症。

[0221] 实施方案15.如实施方案1-14中任一项所述的方法,其中所述癌症是乳腺癌。

- [0222] 实施方案16.如实施方案1-15中任一项所述的方法,其中所述生物样品是生物流体。
- [0223] 实施方案17.如实施方案1-16中任一项所述的方法,其中所述生物流体是唾液。
- [0224] 实施方案18.如实施方案1-16中任一项所述的方法,其中所述生物流体是血液。
- [0225] 实施方案19.如实施方案1-18中任一项所述的方法,其中所述生物流体选自:血液、血成分、血清、血浆、唾液、痰液、尿液、精液、经阴道液体、脑脊液、汗液、胆汁、囊液、泪液、乳腺吸出液和乳腺液。
- [0226] 实施方案20.如实施方案1-19中任一项所述的方法,其中所述生物标志物选自:核酸、肽、蛋白质、脂质、抗原、碳水化合物和蛋白多糖。
- [0227] 实施方案21.如实施方案1-20中任一项所述的方法,其中所述生物标志物是核酸,其中所述核酸是DNA或RNA。
- [0228] 实施方案22.如实施方案1-21中任一项所述的方法,其中所述核酸是RNA,其中所述RNA选自:mRNA、miRNA、snoRNA、snRNA、rRNA、tRNA、siRNA、hnRNA和shRNA。
- [0229] 实施方案23.如实施方案1-22中任一项所述的方法,其中所述RNA是mRNA。
- [0230] 实施方案24.如实施方案1-22中任一项所述的方法,其中所述RNA是miRNA。
- [0231] 实施方案25.如实施方案1-21中任一项所述的方法,其中所述生物标志物是核酸,其中所述核酸是DNA,其中所述DNA选自:双链DNA、单链DNA、互补DNA和非编码DNA。
- [0232] 实施方案26.如实施方案1-21中任一项所述的方法,其中所述生物标志物是无细胞核酸。
- [0233] 实施方案27.如实施方案1-24中任一项所述的方法,其中所述无细胞核酸是无细胞RNA。
- [0234] 实施方案28.如实施方案1-21中任一项所述的方法,其中所述无细胞RNA是无细胞mRNA或无细胞miRNA。
- [0235] 实施方案29.如实施方案1-20中任一项所述的方法,其中所述生物标志物是无细胞DNA。
- [0236] 实施方案30.如实施方案1-20中任一项所述的方法,其中所述生物标志物源自外来体。
- [0237] 实施方案31.如实施方案1-20中任一项所述的方法,其中所述生物标志物是蛋白质。
- [0238] 实施方案32.如实施方案1-20中任一项所述的方法,其中所述生物标志物是乳腺癌途径中的基因。
- [0239] 实施方案33.如实施方案1-20和30-32中任一项所述的方法,其中所述生物标志物选自:MCART1、LCE2B、HIST1H4K、ABCA2、TNFRSF10A、AK092120、DTYMK、Hs.161434、ALKBH1及其任何组合。
- [0240] 实施方案34.如实施方案1-20和30-33中任一项所述的方法,其中所述生物标志物是HIST1H4K。
- [0241] 实施方案35.如实施方案1-20和30-33中任一项所述的方法,其中所述生物标志物是TNFRSF10A。
- [0242] 实施方案36.如实施方案1-20和30-35中任一项所述的方法,其中量化所述生物标

志物的样品水平包括量化至少两种生物标志物,其中所述至少两种生物标志物选自LCE2B、HIST1H4K、ABCA2、TNFRSF10A、AK092120、DTYMK、ALKBH1、MCART1和Hs.161434。

[0243] 实施方案37.如实施方案1-20和30-36中任一项所述的方法,其中量化所述生物标志物的样品水平包括量化两种生物标志物,其中所述两种生物标志物是HIST1H4K和TNFRSF10A。

[0244] 实施方案38.如实施方案1-37中任一项所述的方法,其中确定所述健康状态包括确定所述受试者的所述组织的健康状态。

[0245] 实施方案39.如实施方案1-38中任一项所述的方法,其中所述组织是乳腺组织。

[0246] 实施方案40.如实施方案1-39中任一项所述的方法,进一步包括实验性地裂解所述生物样品的外来体部分,以从所述外来体部分释放所述生物标志物。

[0247] 实施方案41.如实施方案1-40中任一项所述的方法,其中所述参考水平获自患有乳腺癌的受试者。

[0248] 实施方案42.如实施方案1-41中任一项所述的方法,其中所述量化进一步包括将所述RNA实验性地逆转录。

[0249] 实施方案43.如实施方案1-42中任一项所述的方法,其中所述量化进一步包括进行聚合酶链反应。

[0250] 实施方案44.如实施方案1-43中任一项所述的方法,其中所述PCR是定量PCR。

[0251] 实施方案45.如实施方案1-44中任一项所述的方法,其中所述量化进一步包括进行测序,其中所述测序包括大规模并行测序。

[0252] 实施方案46.如实施方案1-45中任一项所述的方法,其中所述量化生物标志物的样品水平以至少90%的准确度进行。

[0253] 实施方案47.如实施方案1-46中任一项所述的方法,其中所述量化生物标志物的样品水平以至少约90%、91%、92%、93%、94%、95%、95%、97%、98%或99%的准确度进行。

[0254] 实施方案48.如实施方案1-47中任一项所述的方法,其中所述量化生物标志物的样品水平以至少约80%的灵敏度进行。

[0255] 实施方案49.如实施方案1-48中任一项所述的方法,其中所述量化生物标志物的样品水平以至少80%、85%、90%、91%、92%、93%、94%、95%、95%、97%、98%或99%的灵敏度进行。

[0256] 实施方案50.如实施方案1-49中任一项所述的方法,其中所述量化生物标志物的样品水平以至少90%的特异性进行。

[0257] 实施方案51.如实施方案1-50中任一项所述的方法,其中所述量化生物标志物的样品水平以至少90%、91%、92%、93%、94%、95%、95%、97%、98%或99%的特异性进行。

[0258] 实施方案52.一种方法,其包括:

a) 获得受试者的唾液样品;

b) 实验性地裂解所述唾液样品的外来体部分以释放生物标志物;

c) 量化所述生物标志物的样品水平;以及

d) 将所述生物标志物的所述样品水平与所述生物标志物的参考水平进行比较,其中所述参考水平获自患有乳腺癌的受试者。

[0259] 实施方案53.如实施方案52的方法,进一步包括额外测试,其中所述额外测试包括评估所述受试者发生乳腺癌的风险。

[0260] 实施方案54.如实施方案52-53中任一项所述的方法,进一步包括将所述额外测试的所述结果与步骤d中的所述比较组合。

[0261] 实施方案55.如实施方案52-54中任一项所述的方法,进一步包括基于所述组合确定所述受试者的乳腺癌状态。

[0262] 实施方案56.如实施方案52-55中任一项所述的方法,其中所述额外测试包括对所述受试者的乳腺组织进行成像。

[0263] 实施方案57.如实施方案52-56中任一项所述的方法,其中使用乳房X线照片进行所述成像。

[0264] 实施方案58.如实施方案52-57中任一项所述的方法,其中所述额外测试包括量化所述受试者中无细胞核酸的样品水平。

[0265] 实施方案59.如实施方案52-58中任一项所述的方法,其中所述无细胞核酸是无细胞RNA或无细胞DNA。

[0266] 实施方案60.如实施方案52-59中任一项所述的方法,其中所述无细胞核酸对所述受试者的组织具有是特异的。

[0267] 实施方案61.如实施方案52-60中任一项所述的方法,其中所述组织是乳腺组织。

[0268] 实施方案62.如实施方案52-60中任一项所述的方法,其中所述无细胞核酸来自生物流体。

[0269] 实施方案63.如实施方案52-60和62中任一项所述的方法,其中所述生物流体选自:血液、血成分、血清、血浆、唾液、痰液、尿液、精液、经阴道液体、脑脊液、汗液、胆汁、囊液、泪液、乳腺吸出液和乳腺液。

[0270] 实施方案64.如实施方案52-63中任一项所述的方法,其中所述生物标志物选自:核酸、肽、蛋白质、脂质、抗原、碳水化合物和蛋白多糖。

[0271] 实施方案65.如实施方案52-64中任一项所述的方法,其中所述生物标志物是核酸。

[0272] 实施方案66.如实施方案52-65中任一项所述的方法,其中所述核酸是RNA。

[0273] 实施方案67.如实施方案52-66中任一项所述的方法,其中所述RNA是mRNA。

[0274] 实施方案68.如实施方案52-66中任一项所述的方法,其中所述RNA是miRNA。

[0275] 实施方案69.如实施方案52-65中任一项所述的方法,其中所述核酸是DNA。

[0276] 实施方案70.如实施方案52-69中任一项所述的方法,其中所述生物标志物是乳腺癌途径中的基因。

[0277] 实施方案71.如实施方案52-70中任一项所述的方法,其中所述生物标志物选自:LCE2B、HIST1H4K、ABCA2、TNFRSF10A、AK092120、DTYMK、Hs.161434、ALKBH1、MCART1和其任何组合。

[0278] 实施方案72.如实施方案52-71中任一项所述的方法,其中所述生物标志物是HIST1H4K。

[0279] 实施方案73.如实施方案52-71中任一项所述的方法,其中所述生物标志物是TNFRSF10A。

[0280] 实施方案74.如实施方案52-73中任一项所述的方法,其中所述量化所述生物标志物的样品水平包括量化至少两种生物标志物,其中所述至少两种生物标志物选自:LCE2B、HIST1H4K、ABCA2、TNFRSF10A、AK092120、DTYMK、ALKBH1、MCART1和Hs.161434。

[0281] 实施方案75.如实施方案52-74中任一项所述的方法,其中所述量化所述生物标志物的样品水平包括量化两种生物标志物,其中所述两种生物标志物是HIST1H4K和TNFRSF10A。

[0282] 实施方案76.如实施方案52-75中任一项所述的方法,进一步包括在步骤b之前实验性地富集所述唾液样品的所述外来体部分。

[0283] 实施方案77.如实施方案52-76中任一项所述的方法,进一步包括在实验性地富集之后稳定所述外来体部分。

[0284] 实施方案78.如实施方案52-68和70-77中任一项所述的方法,其中所述生物标志物是RNA,其中所述量化进一步包括将所述RNA实验性地逆转录。

[0285] 实施方案79.如实施方案52-78中任一项所述的方法,其中所述量化进一步包括进行聚合酶链反应。

[0286] 实施方案80.如实施方案52-79中任一项所述的方法,其中所述PCR是定量PCR。

[0287] 实施方案81.如实施方案52-80中任一项所述的方法,其中所述量化进一步包括进行测序,其中所述测序包括大规模并行测序。

[0288] 实施方案82.如实施方案52-81中任一项所述的方法,其中所述量化生物标志物的样品水平以至少90%的准确度进行。

[0289] 实施方案83.如实施方案52-82中任一项所述的方法,其中所述量化生物标志物的样品水平以至少约90%、91%、92%、93%、94%、95%、95%、97%、98%或99%的准确度进行。

[0290] 实施方案84.如实施方案52-83中任一项所述的方法,其中所述量化生物标志物的样品水平以至少约80%的灵敏度进行。

[0291] 实施方案85.如实施方案52-84中任一项所述的方法,其中所述量化生物标志物的样品水平以至少80%、85%、90%、91%、92%、93%、94%、95%、95%、97%、98%或99%的灵敏度进行。

[0292] 实施方案86.如实施方案52-85中任一项所述的方法,其中所述量化生物标志物的所述样品水平以至少90%的特异性进行。

[0293] 实施方案87.如实施方案52-86中任一项所述的方法,其中所述量化生物标志物的所述样品水平以至少90%、91%、92%、93%、94%、95%、95%、97%、98%或99%的特异性进行。

[0294] 实施方案88.如实施方案52-87中任一项所述的方法,进一步包括基因测试。

[0295] 实施方案89.如实施方案52-88中任一项所述的方法,其中所述基因测试包括测试乳腺癌易感基因中的突变。

[0296] 实施方案90.如实施方案52-89中任一项所述的方法,其中所述基因测试包括测试选自以下的基因中的突变:ATM、BARD1、BRCA1、BRCA2、BRIP1、CASP8、CDH1、CHEK2、CTLA4、CYP19A1、FGFR2、H19、LSP1、MAP3K1、MRE11、NBN、PALB2、PTEN、RAD51、RAD51C、STK11、TERT、TOX3、TP53、XRCC2、XRCC3及其任何组合。

[0297] 实施方案91.一种方法,其包括:

- a) 对受试者进行乳房X线照片;
- b) 获得所述受试者的唾液样品;
- c) 量化所述受试者的所述唾液样品中生物标志物的样品水平,其中所述生物标志物源自外来体;
- d) 将所述生物标志物的所述样品水平与所述生物标志物的参考水平进行比较,其中所述参考水平获自患有乳腺癌的受试者;以及
- e) 将所述乳房X线照片的结果与所述比较组合以确定所述受试者的健康状态。

[0298] 实施方案92.如实施方案91所述的方法,其中所述乳房X线照片结果对于所述受试者的乳腺癌为阴性。

[0299] 实施方案93.如实施方案91-92中任一项所述的方法,进一步包括基于步骤e中的所述组合将来自所述乳房X线照片的所述阴性结果鉴定为假阴性。

[0300] 实施方案94.如实施方案91所述的方法,其中所述乳房X线照片结果对于所述受试者的乳腺癌为阳性。

[0301] 实施方案95.如实施方案91和94中任一项所述的方法,进一步包括基于步骤e中的所述组合将来自所述乳房X线照片的阳性结果鉴定为假阳性结果。

[0302] 实施方案96.如实施方案91所述的方法,其中所述乳房X线照片结果对于所述受试者的乳腺癌是不明确的。

[0303] 实施方案97.如实施方案91-96中任一项所述的方法,其中所述生物标志物选自:核酸、肽、蛋白质、脂质、抗原、碳水化合物和蛋白多糖。

[0304] 实施方案98.如实施方案91-97中任一项所述的方法,其中所述生物标志物是核酸。

[0305] 实施方案99.如实施方案91-98中任一项所述的方法,其中所述核酸是RNA。

[0306] 实施方案100.如实施方案91-99中任一项所述的方法,其中所述RNA是mRNA。

[0307] 实施方案101.如实施方案91-99中任一项所述的方法,其中所述RNA是miRNA。

[0308] 实施方案102.如实施方案91-98中任一项所述的方法,其中所述核酸是DNA。

[0309] 实施方案103.如实施方案91-102中任一项所述的方法,其中所述生物标志物是乳腺癌途径中的基因。

[0310] 实施方案104.如实施方案91-103中任一项所述的方法,其中所述生物标志物选自:LCE2B、HIST1H4K、ABCA2、TNFRSF10A、AK092120、DTYMK、Hs.161434、ALKBH1、MCART1和任何组合。

[0311] 实施方案105.如实施方案91-104中任一项所述的方法,其中所述生物标志物是HIST1H4K。

[0312] 实施方案106.如实施方案91-104中任一项所述的方法,其中所述生物标志物是TNFRSF10A。

[0313] 实施方案107.如实施方案91-106中任一项所述的方法,其中所述量化所述生物标志物的样品水平包括量化至少两种生物标志物,其中所述至少两种生物标志物选自LCE2B、HIST1H4K、ABCA2、TNFRSF10A、AK092120、DTYMK、ALKBH1、MCART1和Hs.161434。

[0314] 实施方案108.如实施方案91-107中任一项所述的方法,其中所述量化所述生物标

志物的样品水平包括量化两种生物标志物,其中所述两种生物标志物是HIST1H4K和TNFRSF10A。

[0315] 实施方案109.如实施方案91-108中任一项所述的方法,进一步包括裂解所述唾液样品的的外来体部分。

[0316] 实施方案110.如实施方案91-109中任一项所述的方法,进一步包括在裂解之前实验性地富集所述唾液样品的的所述外来体部分。

[0317] 实施方案111.如实施方案91-110中任一项所述的方法,进一步包括在实验性地富集之后稳定所述外来体部分。

[0318] 实施方案112.如实施方案91-101和103-111中任一项所述的方法,其中所述生物标志物是RNA,其中所述量化进一步包括将所述RNA实验性地逆转录。

[0319] 实施方案113.如实施方案91-101和103-112中任一项所述的方法,其中所述量化进一步包括进行聚合酶链反应。

[0320] 实施方案114.如实施方案91-101和103-113中任一项所述的方法,其中所述PCR是定量PCR。

[0321] 实施方案115.如实施方案91-114中任一项所述的方法,其中所述量化进一步包括进行测序,其中所述测序包括大规模并行测序。

[0322] 实施方案116.如实施方案91-115中任一项所述的方法,其中所述量化生物标志物的样品水平以至少90%的准确度进行。

[0323] 实施方案117.如实施方案91-116中任一项所述的方法,其中所述量化生物标志物的样品水平以至少约90%、91%、92%、93%、94%、95%、95%、97%、98%或99%的准确度进行。

[0324] 实施方案118.如实施方案91-117中任一项所述的方法,其中所述量化生物标志物的样品水平以至少约80%的灵敏度进行。

[0325] 实施方案119.如实施方案91-118中任一项所述的方法,其中所述量化生物标志物的所述样品水平以至少80%、85%、90%、91%、92%、93%、94%、95%、95%、97%、98%或99%的灵敏度进行。

[0326] 实施方案120.如实施方案91-119中任一项所述的方法,其中所述量化生物标志物的所述样品水平以至少90%的特异性进行。

[0327] 实施方案121.如实施方案91-120中任一项所述的方法,其中所述量化生物标志物的所述样品水平以至少90%、91%、92%、93%、94%、95%、95%、97%、98%或99%的特异性进行。

[0328] 实施方案122.如实施方案91-121中任一项所述的方法,进一步包括基因测试。

[0329] 实施方案123.如实施方案91-122中任一项所述的方法,其中所述基因测试包括测试乳腺癌易感基因中的突变。

[0330] 实施方案124.如实施方案91-123中任一项所述的方法,其中所述基因测试包括测试选自以下的基因中的突变:ATM、BARD1、BRCA1、BRCA2、BRIP1、CASP8、CDH1、CHEK2、CTLA4、CYP19A1、FGFR2、H19、LSP1、MAP3K1、MRE11、NBN、PALB2、PTEN、RAD51、RAD51C、STK11、TERT、TOX3、TP53、XRCC2、XRCC3及其任何组合。

[0331] 实施方案125.如实施方案91-125中任一项所述的方法,其中所述受试者具有致密

的乳腺组织。

[0332] 实施方案126.一种用于减少健康状况的假阳性或假阴性结果的数目的方法,所述方法包括:

a) 对受试者进行筛查测试,其中所述筛查测试包括评估所述受试者发生健康状况的风险;

b) 获得所述受试者的生物样品,其中所述受试者来自所述筛查测试的结果为阳性、阴性或不明确的受试者群体;

c) 量化所述受试者的所述生物样品中生物标志物的样品水平,其中所述生物标志物与所述健康状况相关;

d) 对于所述健康状况,将所述生物标志物的所述样品水平与所述生物标志物的参考水平进行比较;以及

e) 基于所述比较的结果将所述筛查测试的结果鉴定为所述健康状况的假阳性或假阴性。

[0333] 实施方案127.如实施方案126所述的方法,其中所述健康状况是癌症。

[0334] 实施方案128.如实施方案126-127中任一项所述的方法,其中所述癌症是乳腺癌。

[0335] 实施方案129.如实施方案126-128中任一项所述的方法,其中所述筛查测试包括对所述受试者的乳腺组织进行成像。

[0336] 实施方案130.如实施方案126-129中任一项所述的方法,其中使用乳房X线照片进行所述成像。

[0337] 实施方案131.如实施方案126-130中任一项所述的方法,其中所述筛查测试包括量化所述受试者中无细胞核酸的样品水平。

[0338] 实施方案132.如实施方案126-131中任一项所述的方法,其中所述无细胞核酸是无细胞RNA或无细胞DNA。

[0339] 实施方案133.如实施方案126-132中任一项所述的方法,其中所述无细胞核酸获自生物流体。

[0340] 实施方案134.如实施方案121-133中任一项所述的方法,其中所述生物流体选自:血液、血成分、血清、血浆、唾液、痰液、尿液、精液、经阴道液体、脑脊液、汗液、胆汁、囊液、泪液、乳腺吸出液和乳腺液。

[0341] 实施方案135.如实施方案126-134中任一项所述的方法,其中所述生物样品是生物流体。

[0342] 实施方案136.如实施方案126-135中任一项所述的方法,其中所述生物流体是唾液。

[0343] 实施方案137.如实施方案126-135中任一项所述的方法,其中所述生物流体是血液。

[0344] 实施方案138.如实施方案126-137中任一项所述的方法,其中所述生物流体选自:血液、血成分、血清、血浆、唾液、痰液、尿液、精液、经阴道液体、脑脊液、汗液、胆汁、囊液、泪液、乳腺吸出液和乳腺液。

[0345] 实施方案139.如实施方案126-138中任一项所述的方法,其中所述生物标志物选自:核酸、肽、蛋白质、脂质、抗原、碳水化合物和蛋白多糖。

[0346] 实施方案140.如实施方案126-139中任一项所述的方法,其中所述生物标志物是核酸,其中所述核酸是DNA或RNA。

[0347] 实施方案141.如实施方案126-140中任一项所述的方法,其中所述核酸是RNA,其中所述RNA选自:mRNA、miRNA、snoRNA、snRNA、rRNA、tRNA、siRNA、hnRNA和shRNA。

[0348] 实施方案142.如实施方案126-141中任一项所述的方法,其中所述RNA是mRNA。

[0349] 实施方案143.如实施方案126-141中任一项所述的方法,其中所述RNA是miRNA。

[0350] 实施方案144.如实施方案126-140中任一项所述的方法,其中所述生物标志物是核酸,其中所述核酸是DNA,其中所述DNA选自:双链DNA、单链DNA、互补DNA和非编码DNA。

[0351] 实施方案145.如实施方案126-144中任一项所述的方法,其中所述生物标志物是无细胞核酸。

[0352] 实施方案146.如实施方案126-143中任一项所述的方法,其中所述无细胞核酸是无细胞RNA。

[0353] 实施方案147:如实施方案126-143和145-146中任一项所述的方法,其中所述无细胞RNA是无细胞mRNA或无细胞miRNA。

[0354] 实施方案148.如实施方案126-147中任一项所述的方法,其中所述生物标志物源自外来体。

[0355] 实施方案149.如实施方案126-148中任一项所述的方法,其中所述生物标志物是乳腺癌途径中的基因。

[0356] 实施方案150.如实施方案126-149中任一项所述的方法,其中所述生物标志物选自:LCE2B、HIST1H4K、ABCA2、TNFRSF10A、AK092120、DTYMK、Hs.161434、ALKBH1、MCART1和任何组合。

[0357] 实施方案151.如实施方案126-150中任一项所述的方法,其中所述生物标志物是HIST1H4K。

[0358] 实施方案152.如实施方案126-150中任一项所述的方法,其中所述生物标志物是TNFRSF10A。

[0359] 实施方案153.如实施方案126-152中任一项所述的方法,其中量化所述生物标志物的样品水平包括量化至少两种生物标志物,其中所述至少两种生物标志物选自LCE2B、HIST1H4K、ABCA2、TNFRSF10A、AK092120、DTYMK、ALKBH1、MCART1和Hs.161434。

[0360] 实施方案154.如实施方案126-153中任一项所述的方法,其中量化所述生物标志物的样品水平包括量化两种生物标志物,其中所述两种生物标志物是HIST1H4K和TNFRSF10A。

[0361] 实施方案155.如实施方案126-154中任一项所述的方法,其中所述受试者具有致密的乳腺组织。

[0362] 实施方案156.如实施方案126-155中任一项所述的方法,进一步包括实验性地裂解所述生物样品的外来体部分,以从所述外来体部分释放所述生物标志物。

[0363] 实施方案157.如实施方案126-156中任一项所述的方法,其中所述量化进一步包括将所述RNA实验性地逆转录。

[0364] 实施方案158.如实施方案126-157中任一项所述的方法,其中所述量化进一步包括进行聚合酶链反应。

[0365] 实施方案159.如实施方案126-158中任一项所述的方法,其中所述PCR是定量PCR。

[0366] 实施方案160.如实施方案126-159中任一项所述的方法,其中所述量化进一步包括进行测序,其中所述测序包括大规模并行测序。

[0367] 实施方案161.如实施方案126-160中任一项所述的方法,其中所述量化生物标志物的样品水平以至少90%的准确度进行。

[0368] 实施方案162.如实施方案126-161中任一项所述的方法,其中所述量化生物标志物的样品水平以至少约90%、91%、92%、93%、94%、95%、95%、97%、98%、99%或99.5%的准确度进行。

[0369] 实施方案163.如实施方案126-162中任一项所述的方法,其中所述量化生物标志物的样品水平以至少约80%的灵敏度进行。

[0370] 实施方案164.如实施方案163中任一项所述的方法,其中所述量化生物标志物的样品水平以至少约80%、85%、90%、91%、92%、93%、94%、95%、95%、97%、98%、99%或99.5%的灵敏度进行。

[0371] 实施方案165.如实施方案126-164中任一项所述的方法,其中所述量化生物标志物的样品水平以至少90%的特异性进行。

[0372] 实施方案166.如实施方案126-165中任一项所述的方法,其中所述量化生物标志物的样品水平以至少约90%、91%、92%、93%、94%、95%、95%、97%、98%、99%或99.5%的特异性进行。

[0373] 实施方案167.如实施方案126-166中任一项所述的方法,其中所述量化所述生物标志物的样品水平包括量化至少2、3、4、5、6、7、8、9或10种生物标志物。

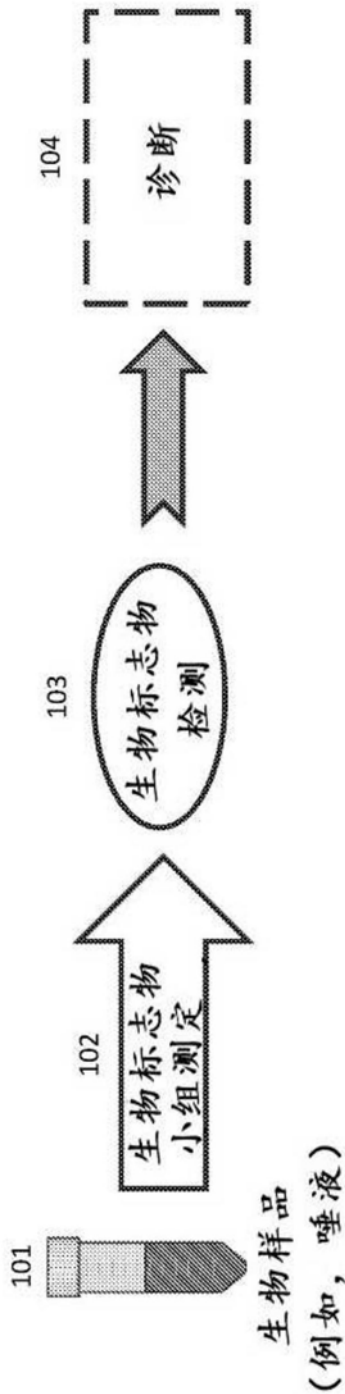


图1

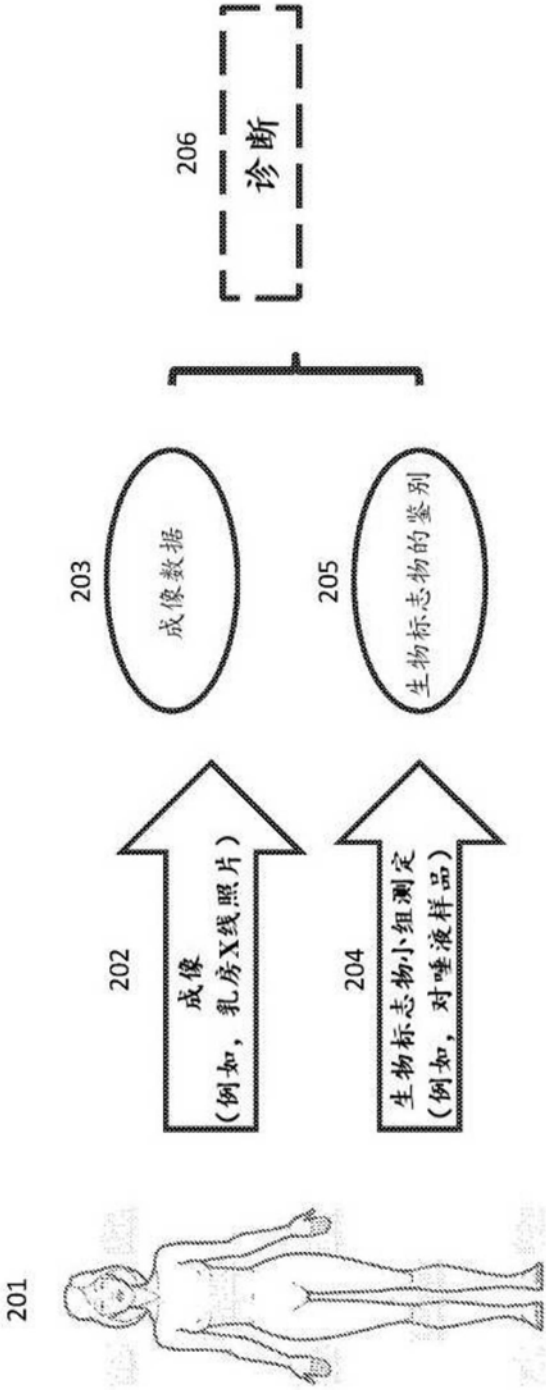


图2

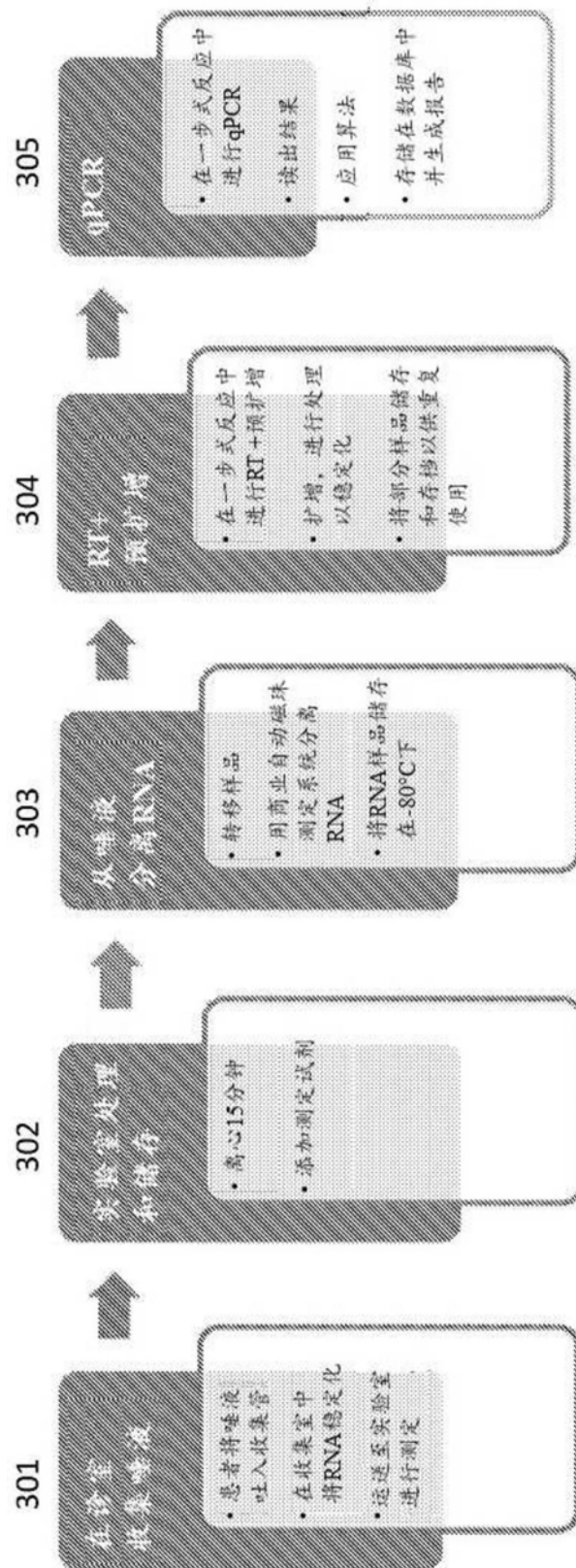


图3

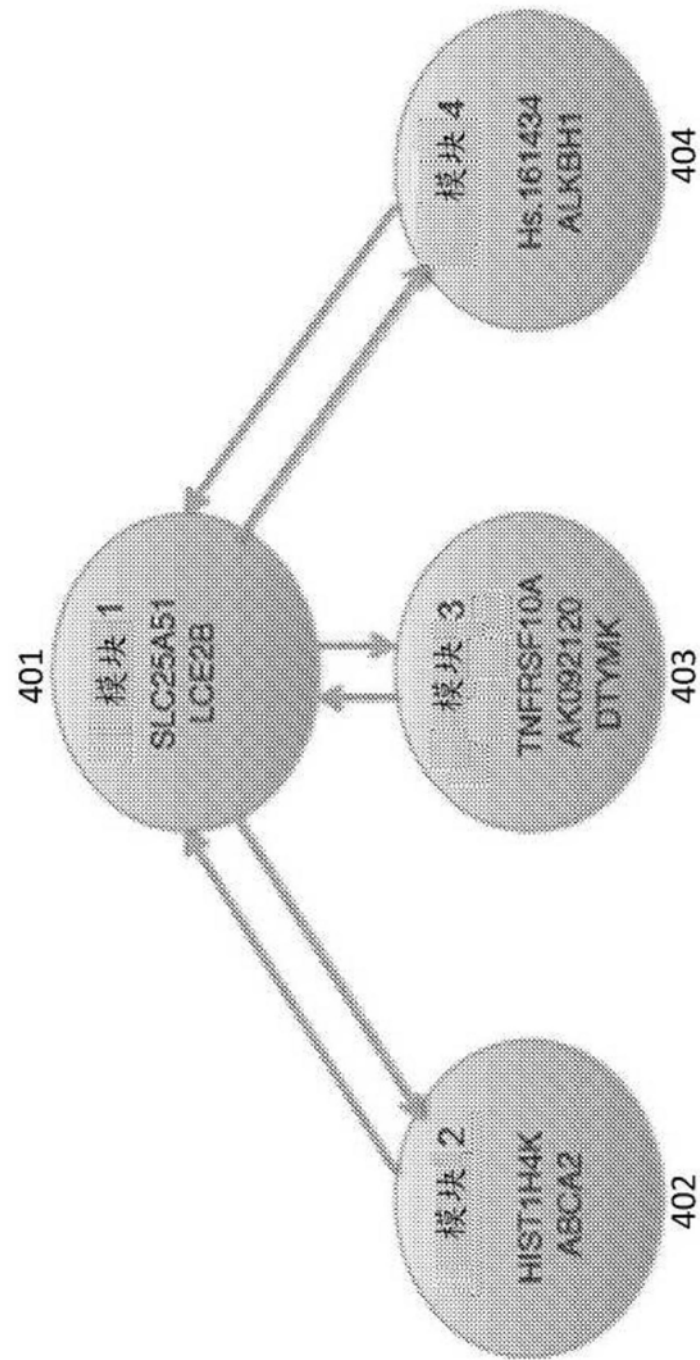


图4

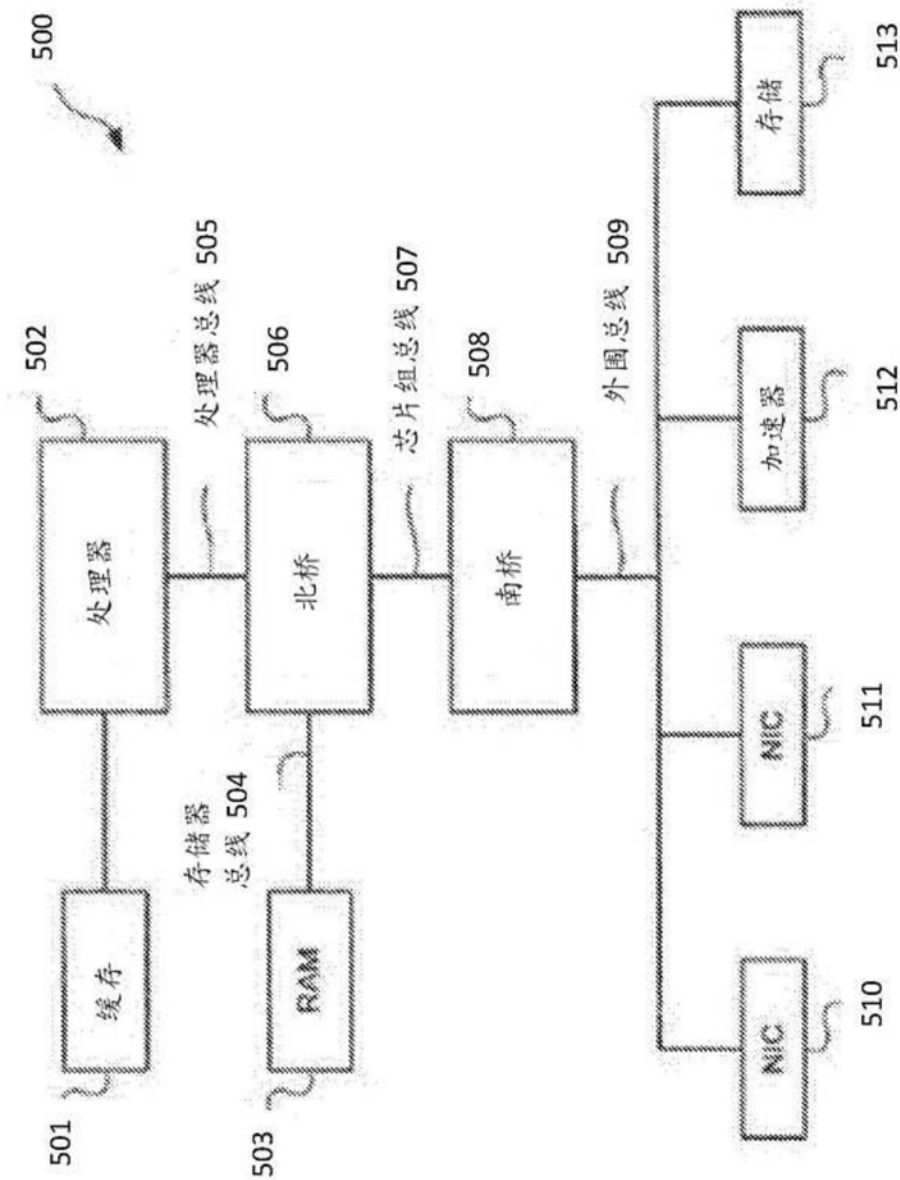


图5

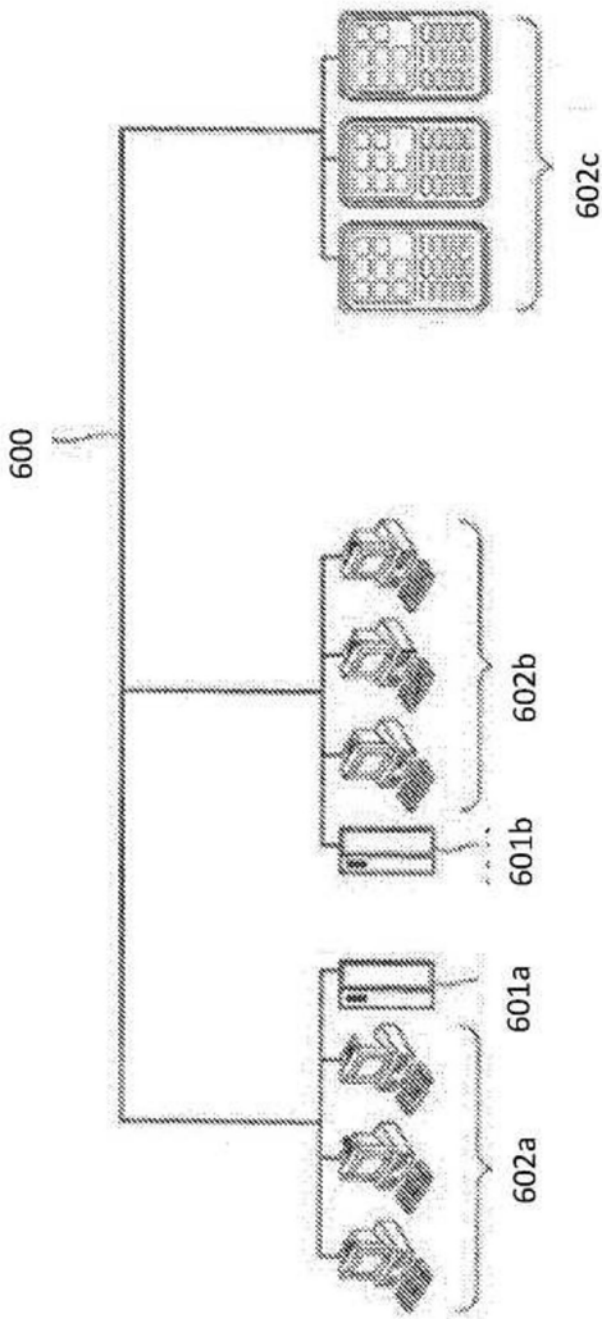


图6

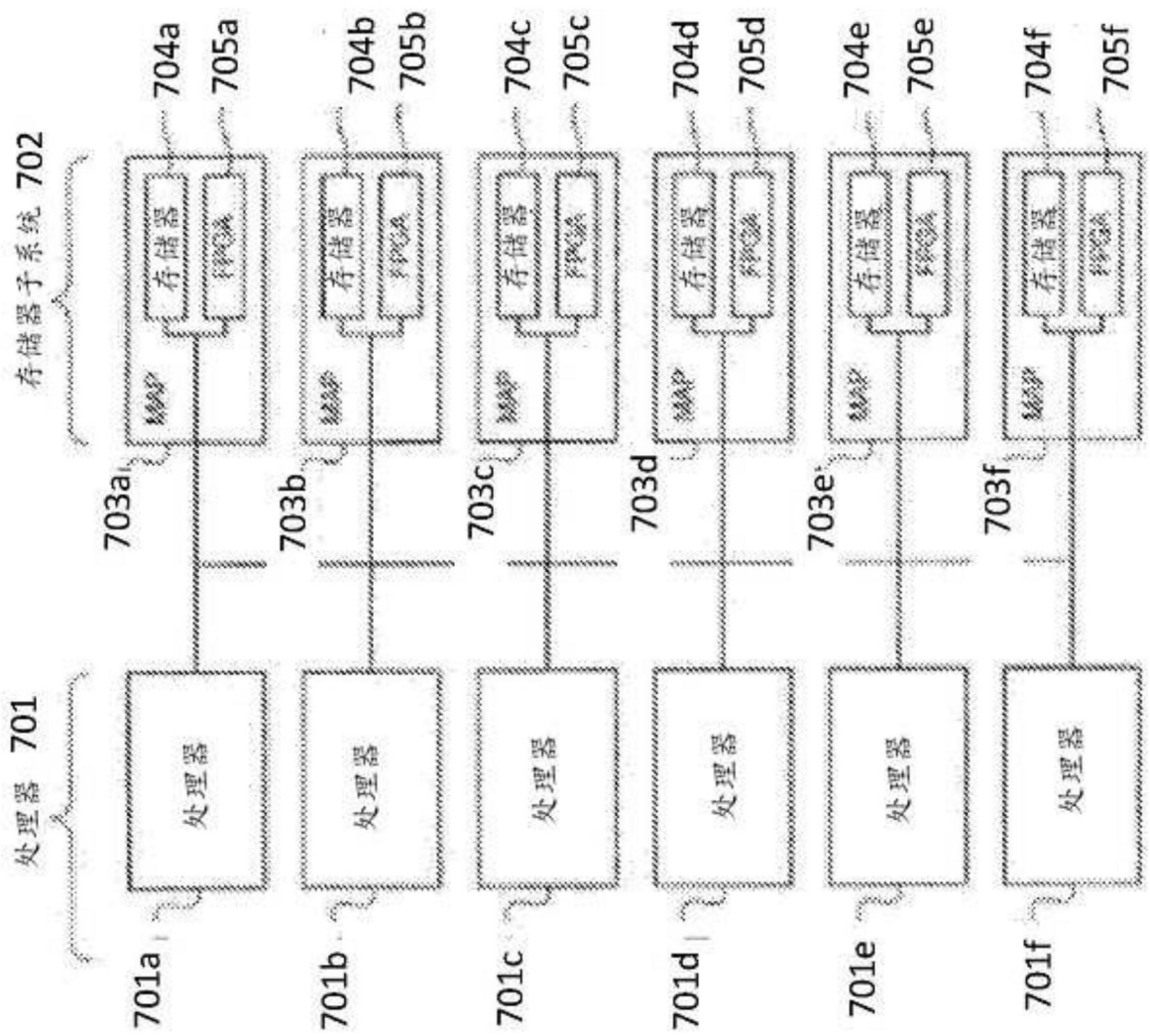


图7

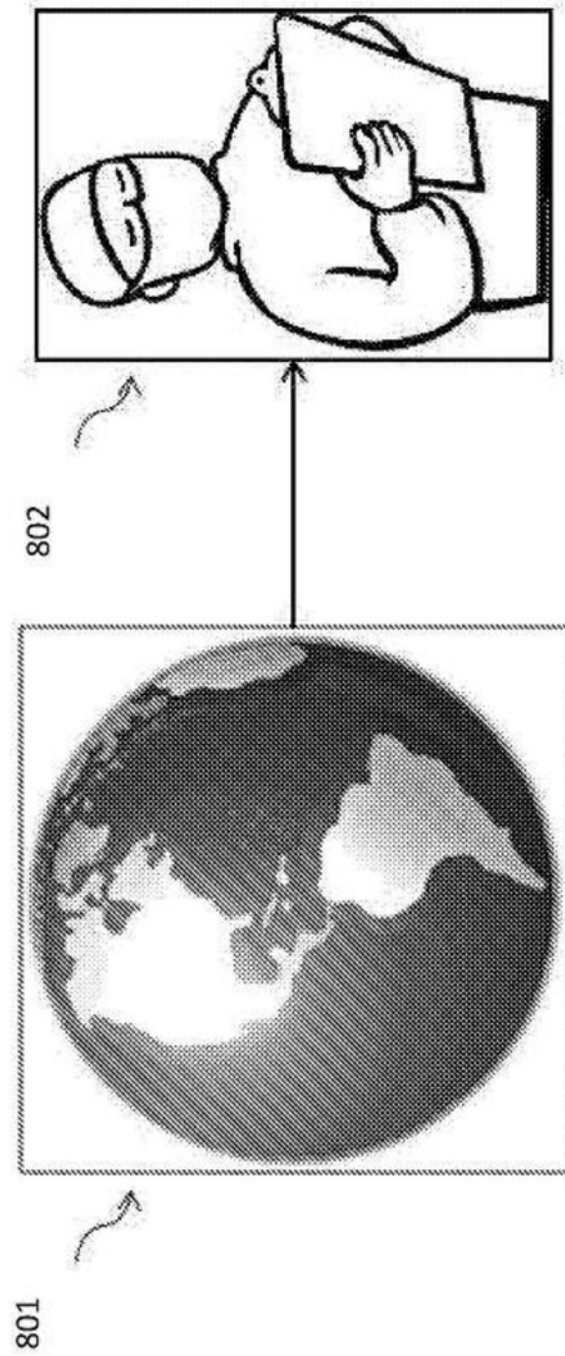


图8

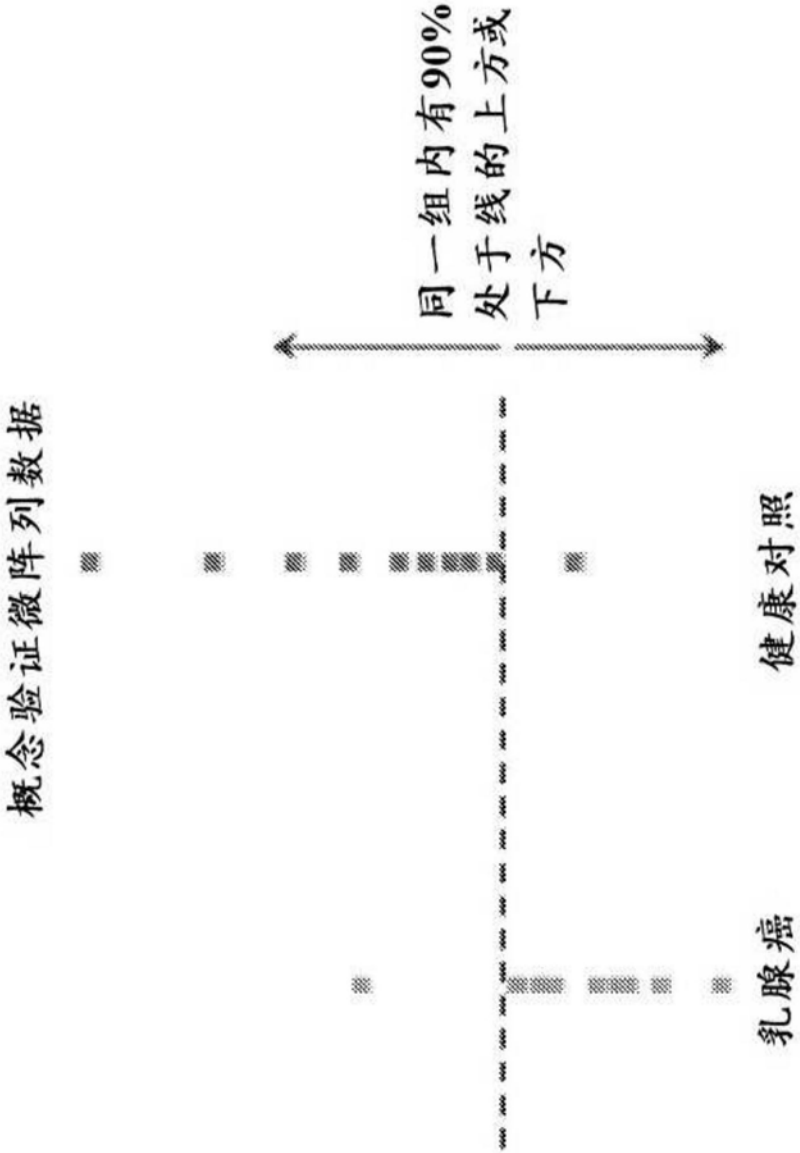


图9

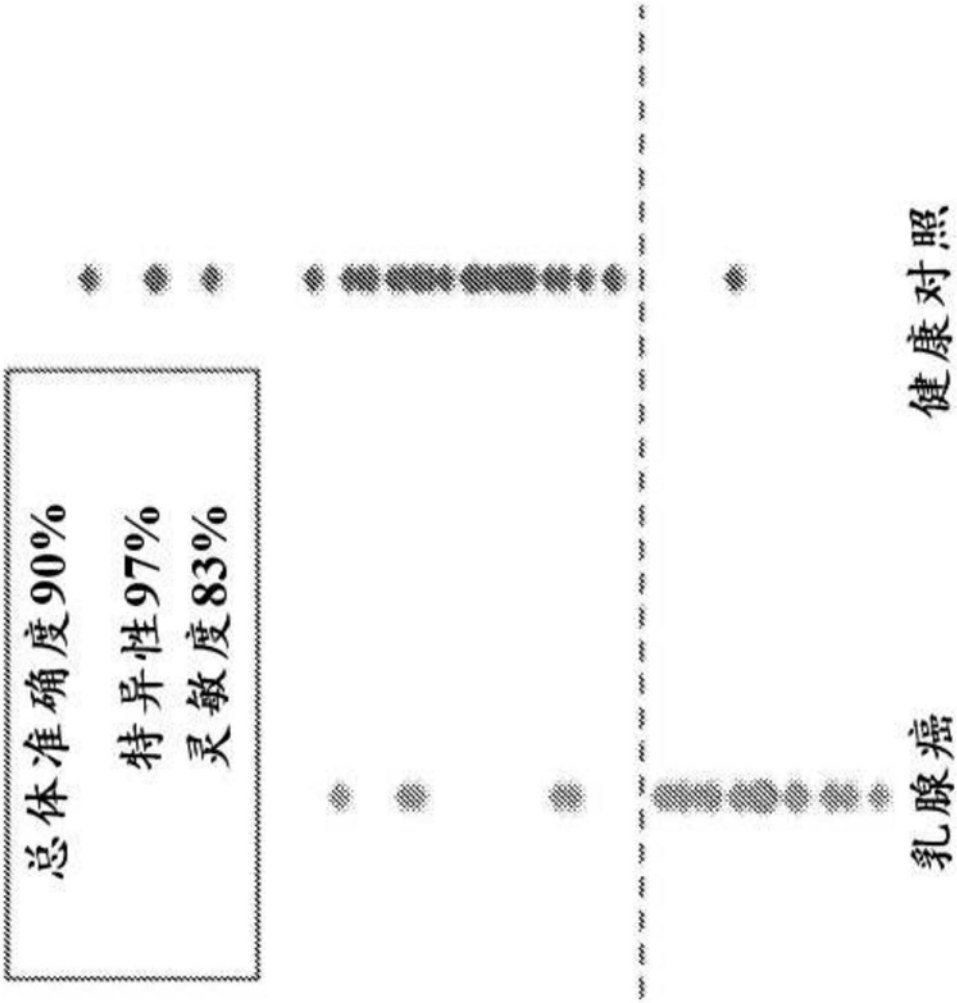


图10

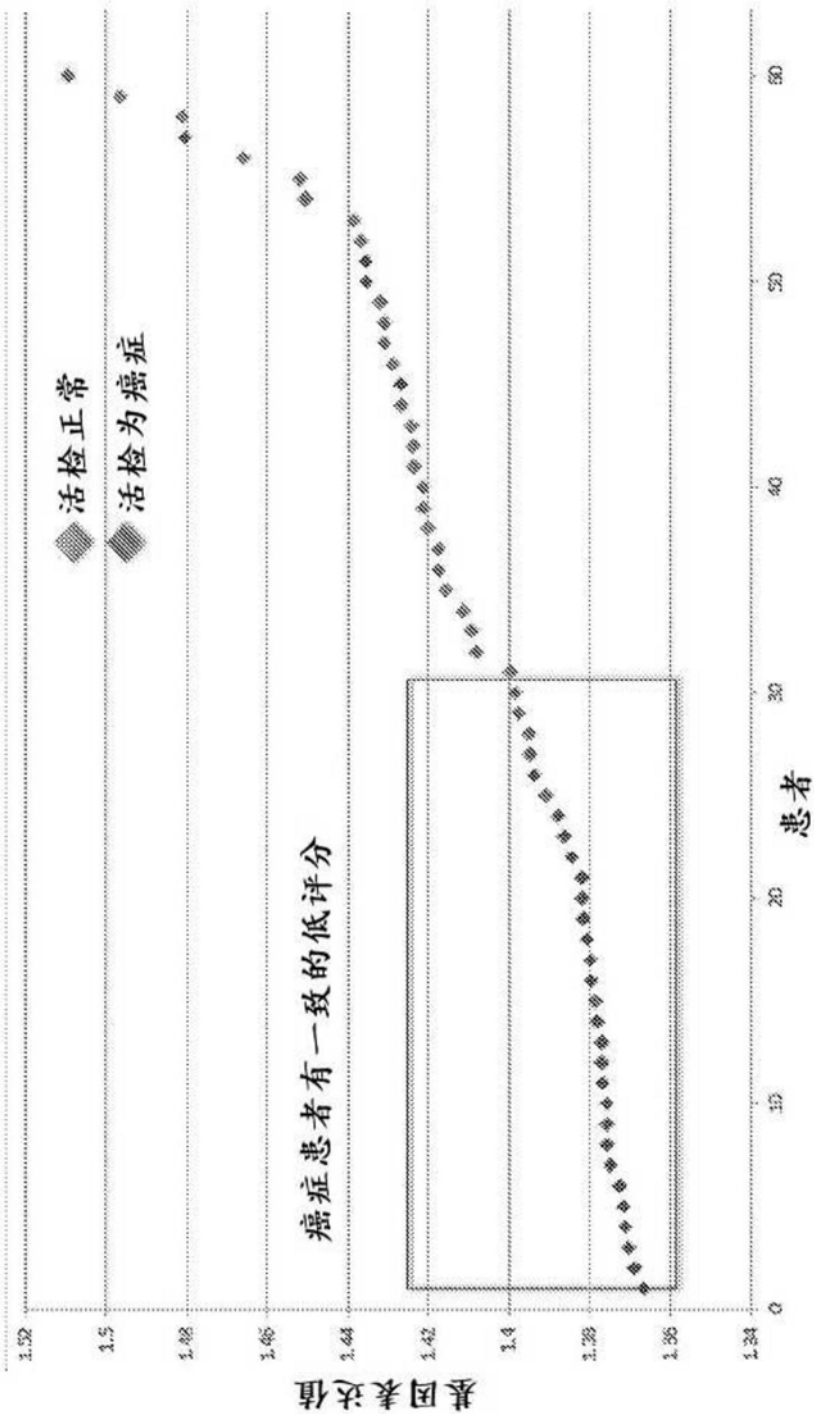


图11

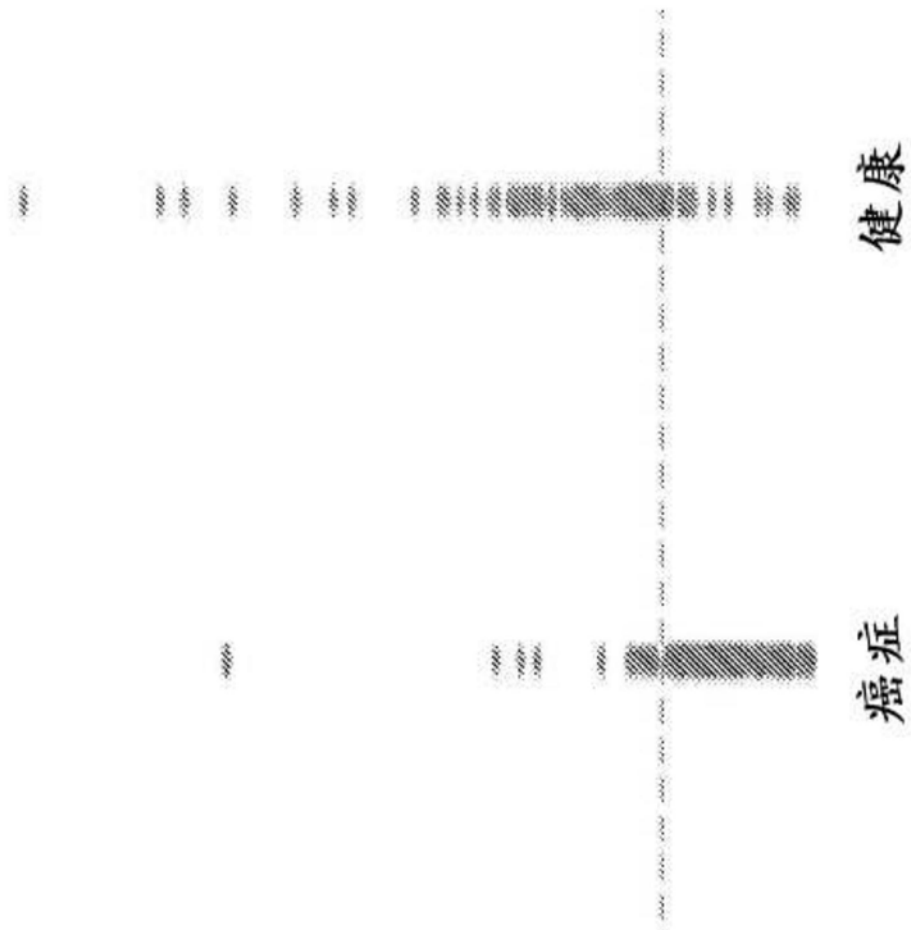


图12

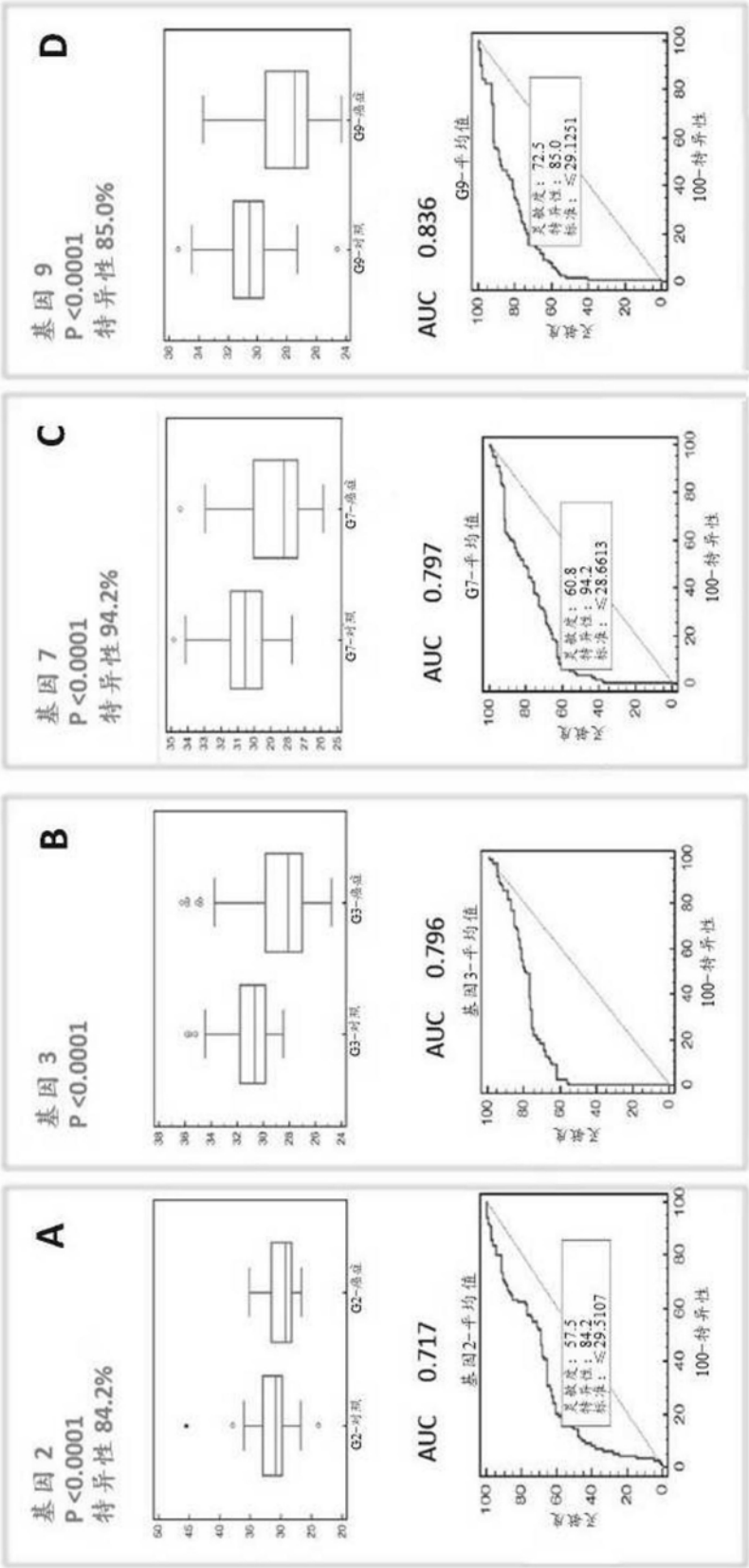


图13

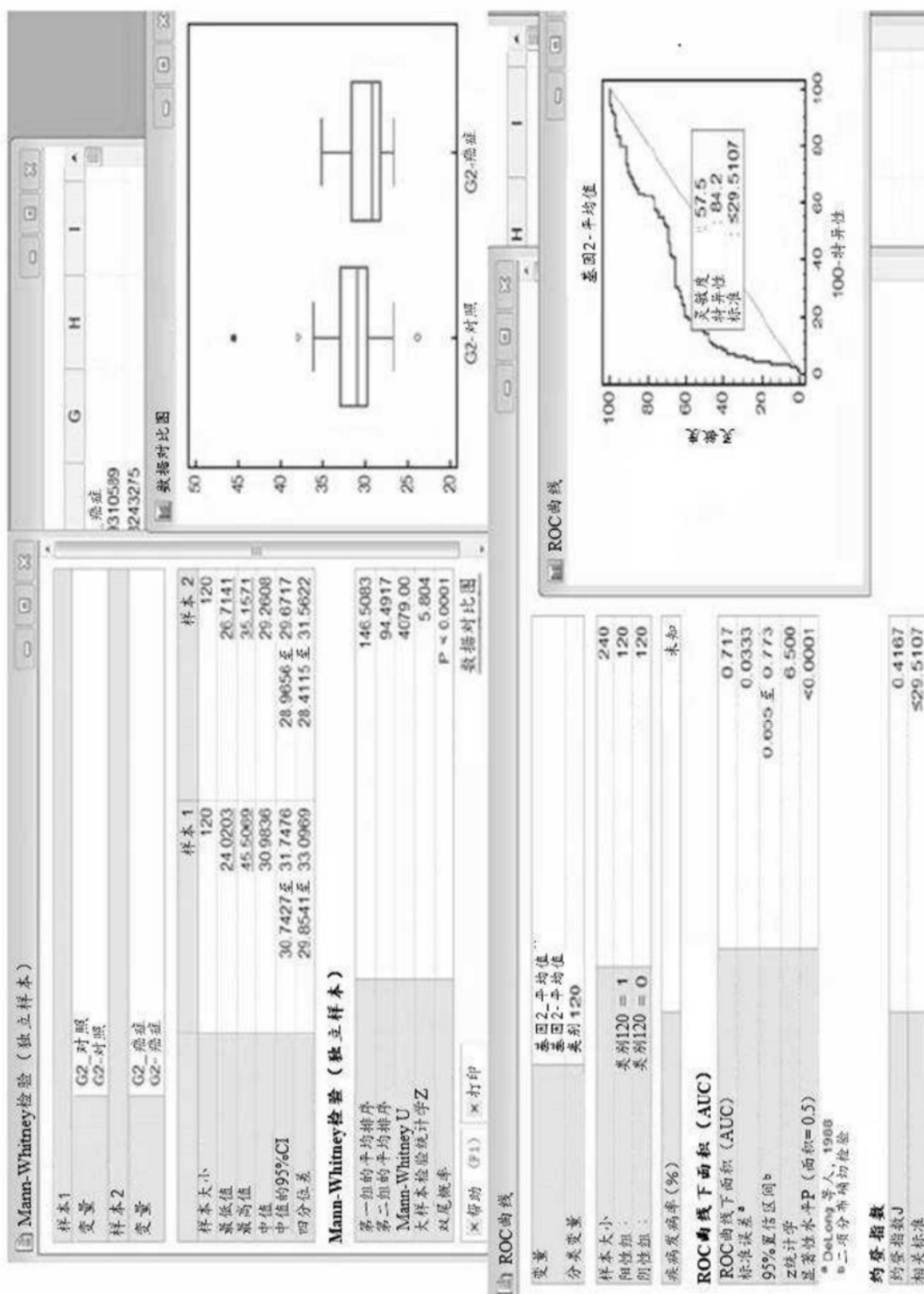


图14

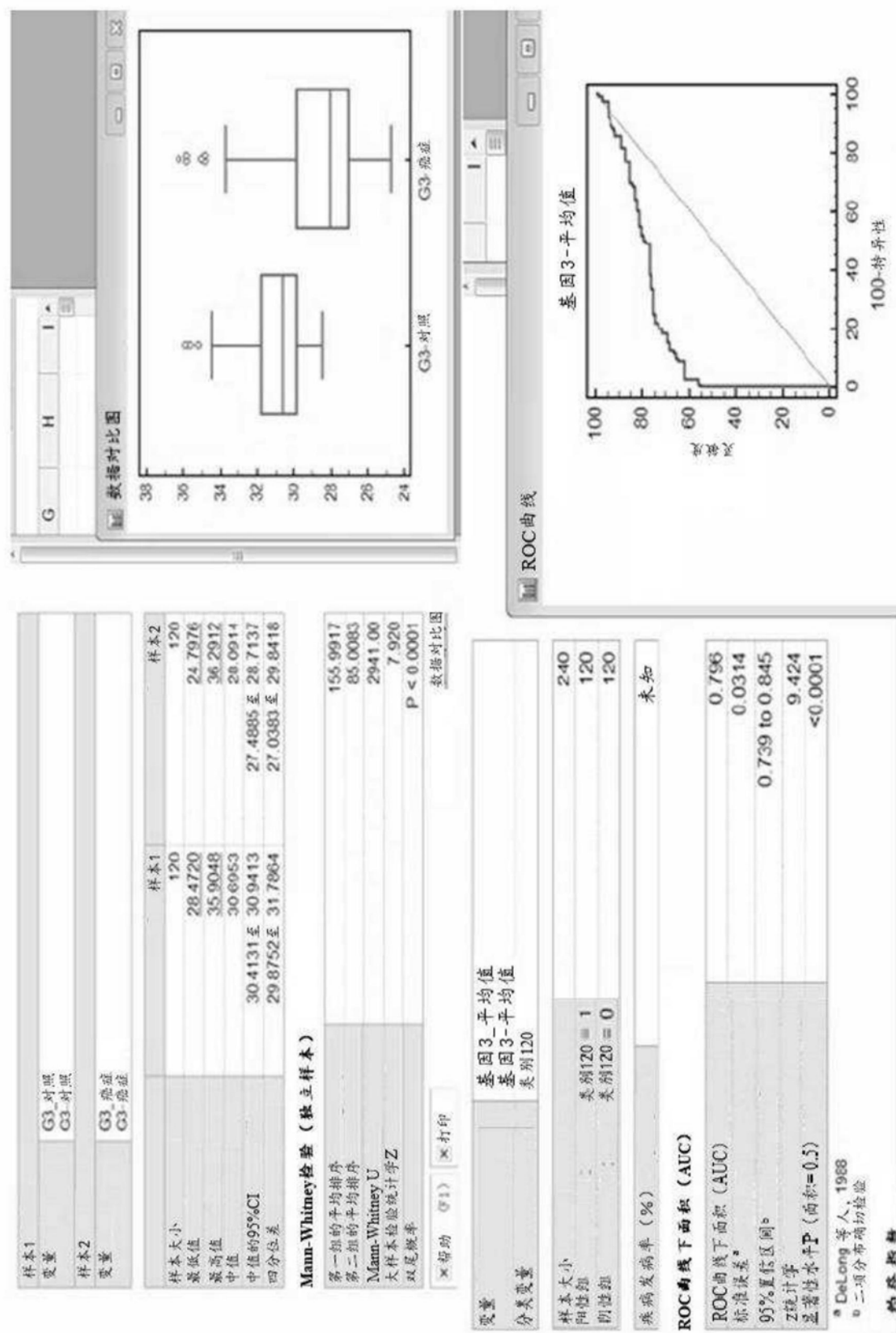


图15

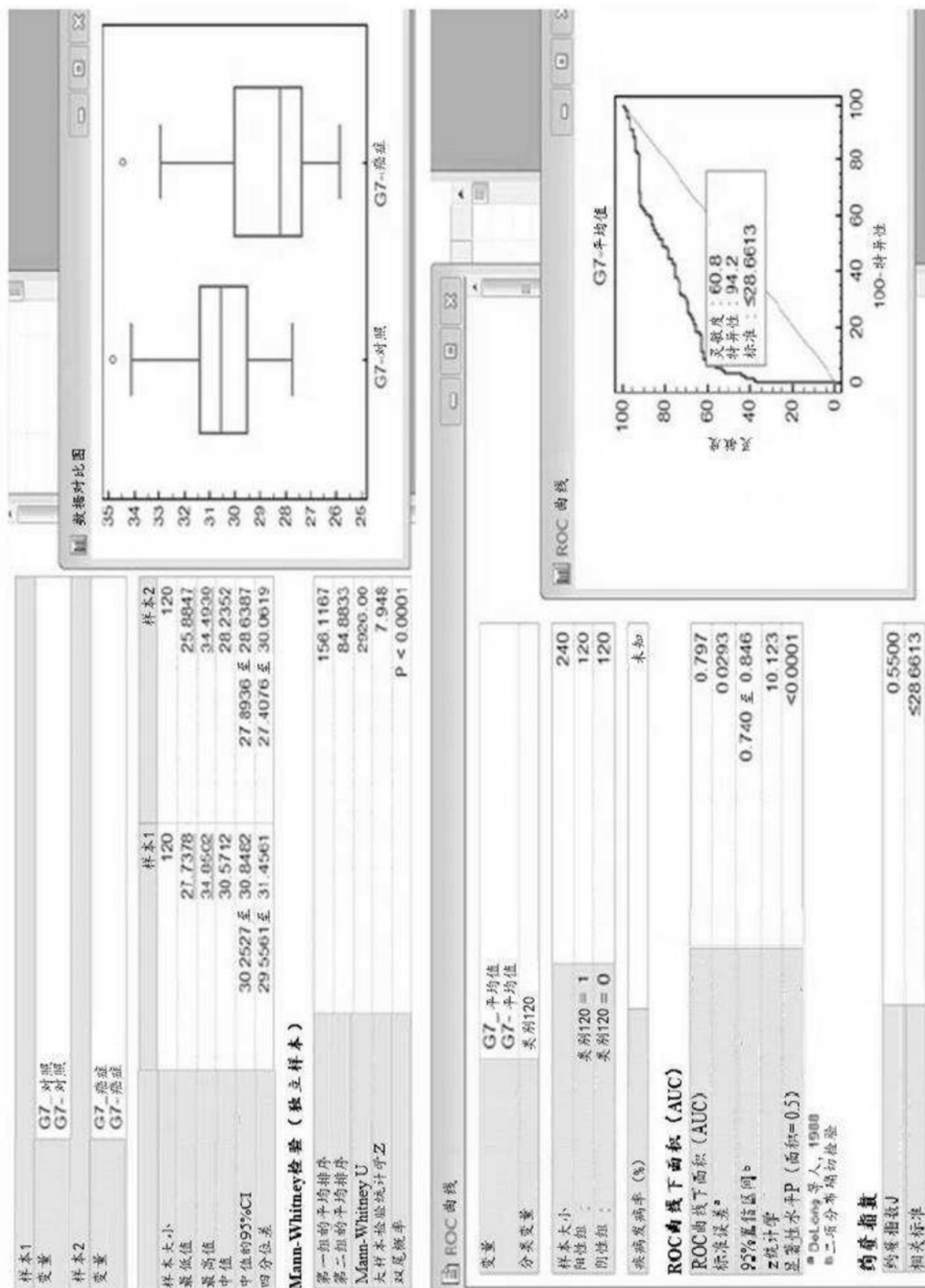


图16

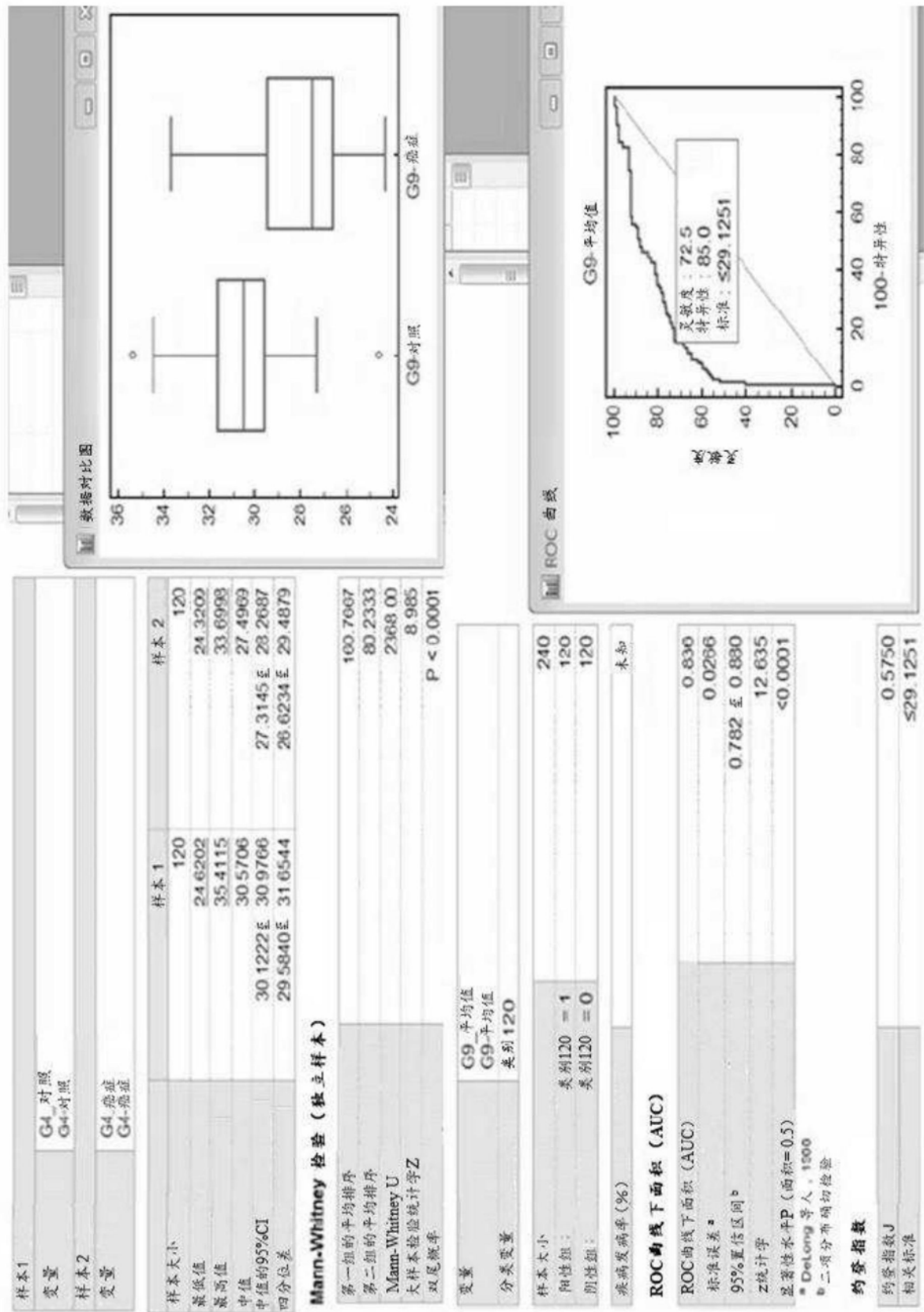


图17

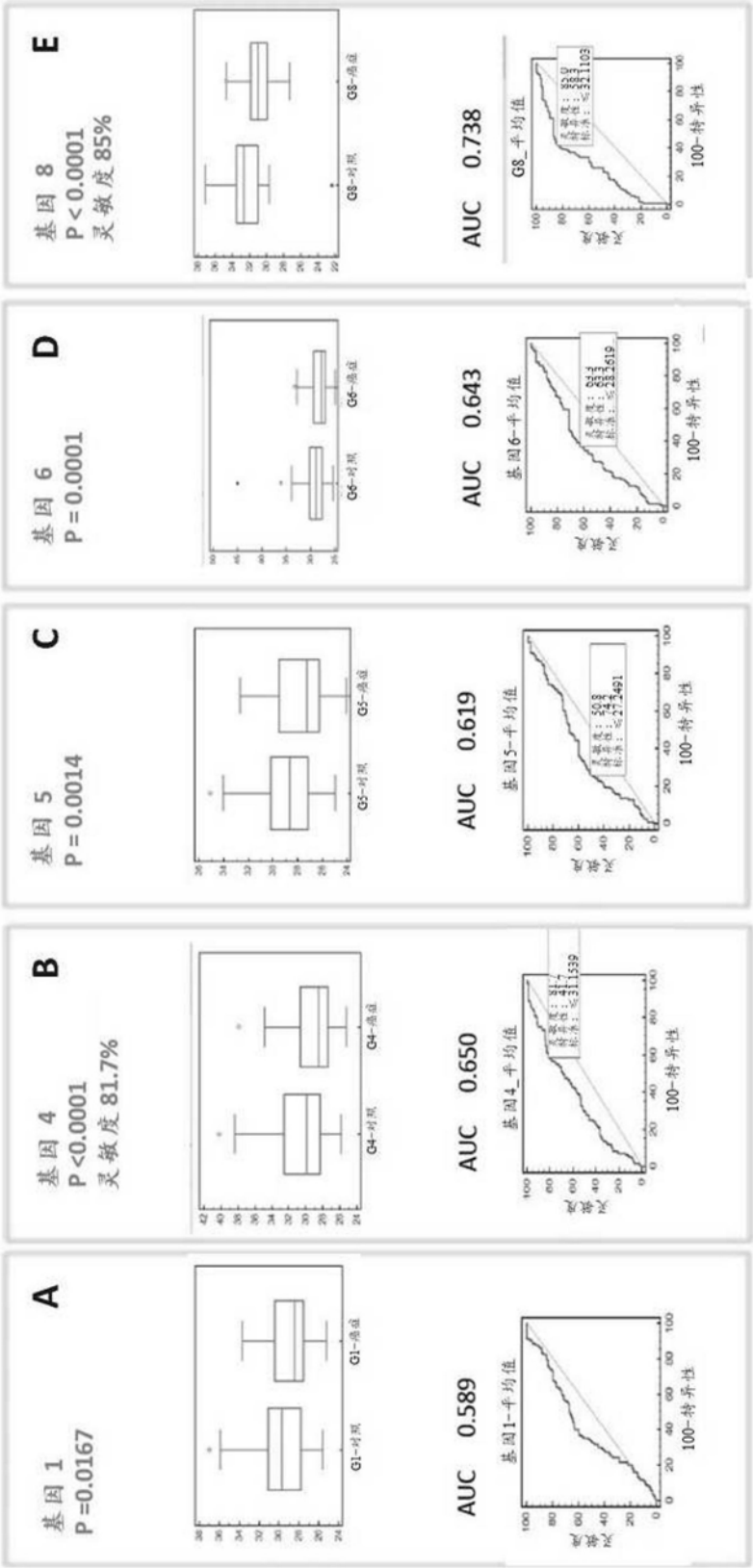


图18

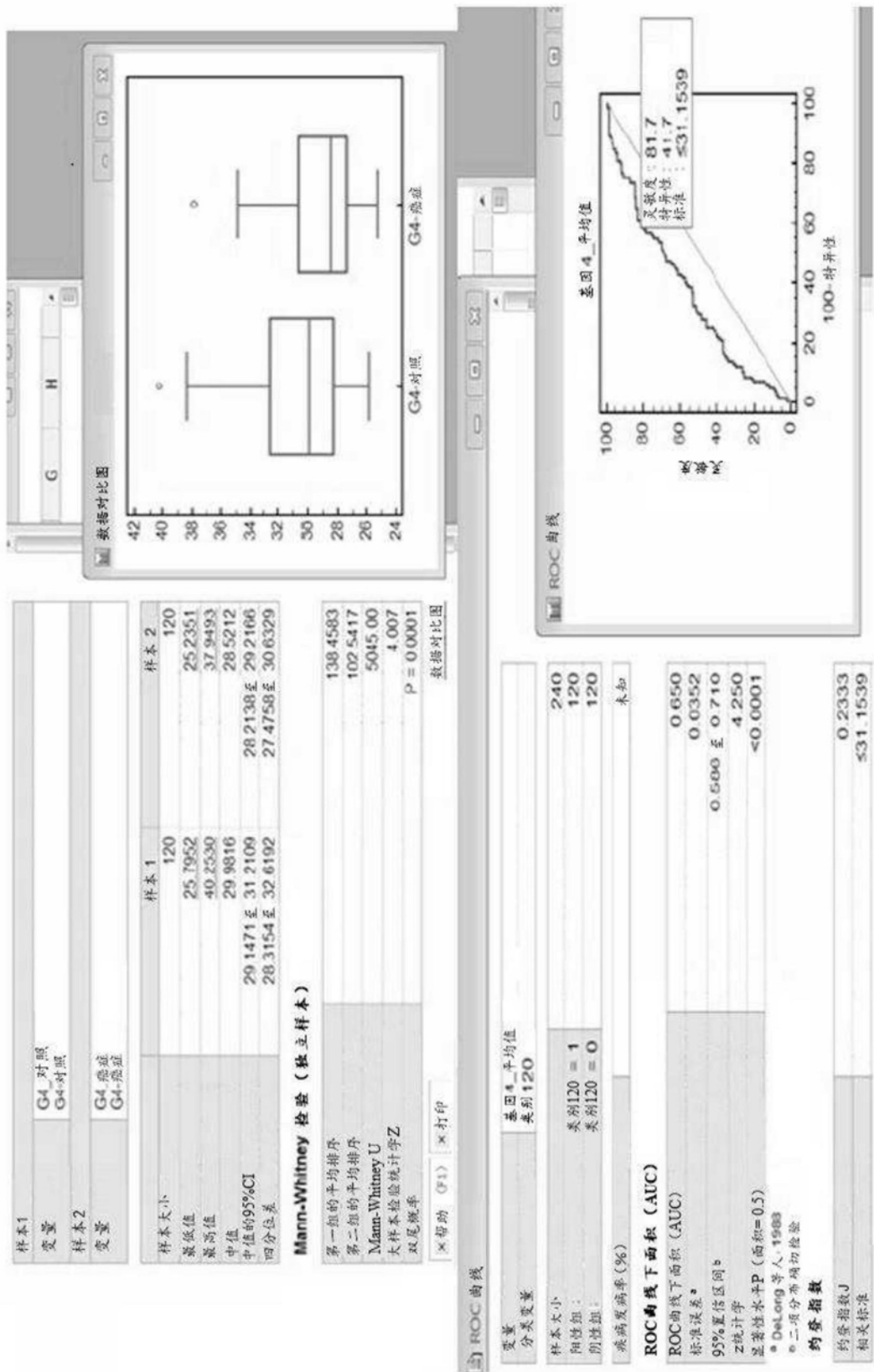


图20

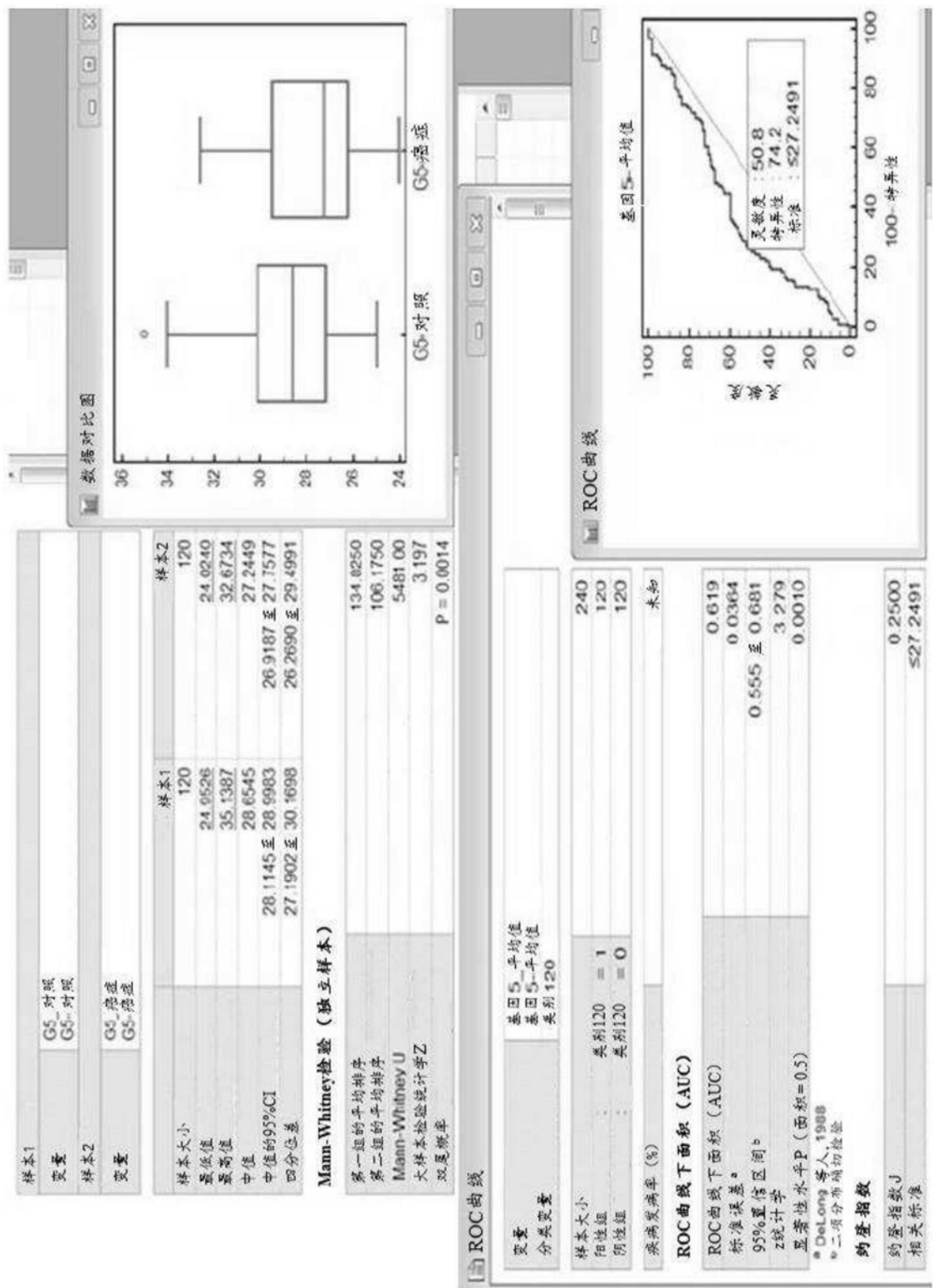


图21

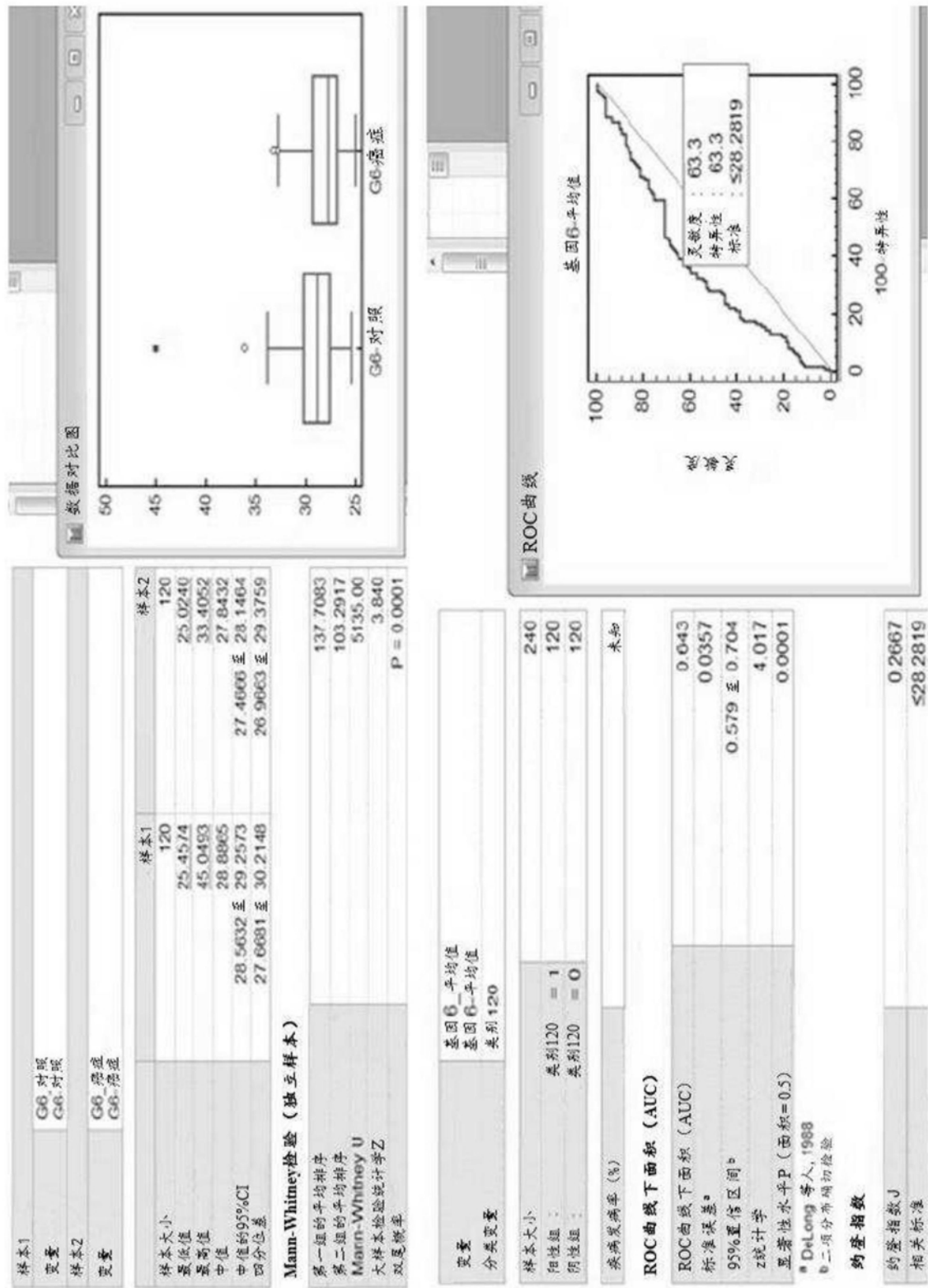


图22

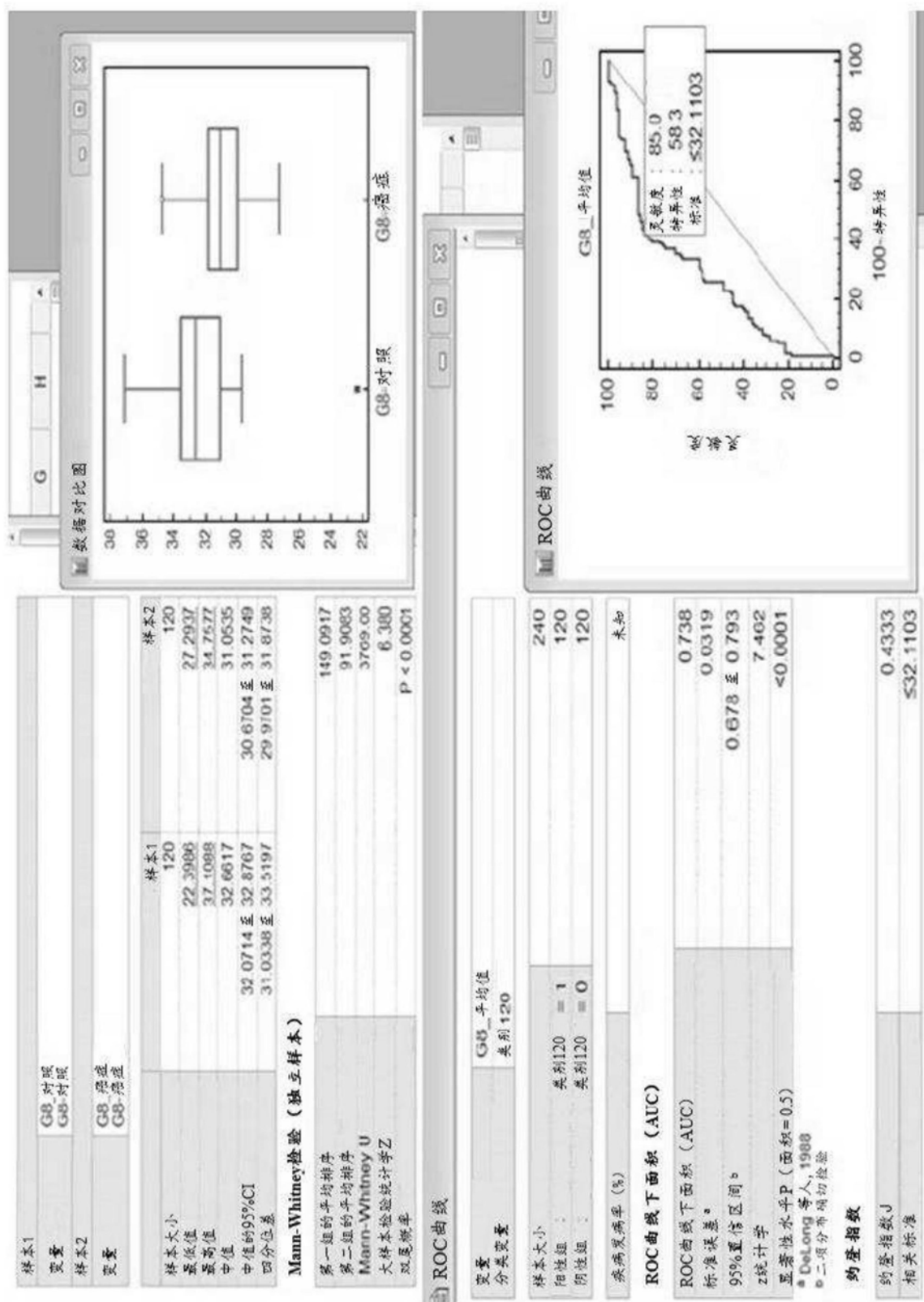


图23

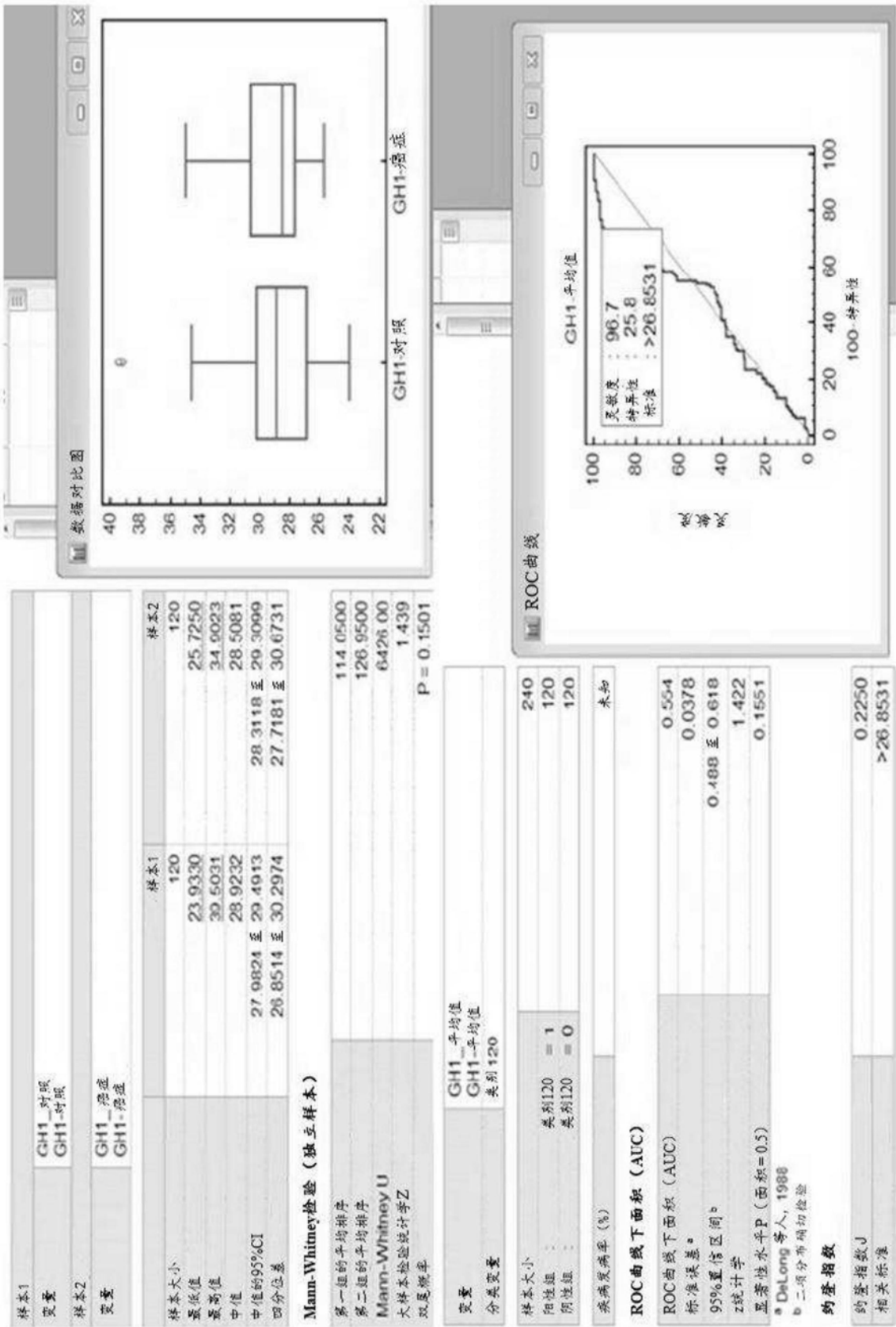


图24

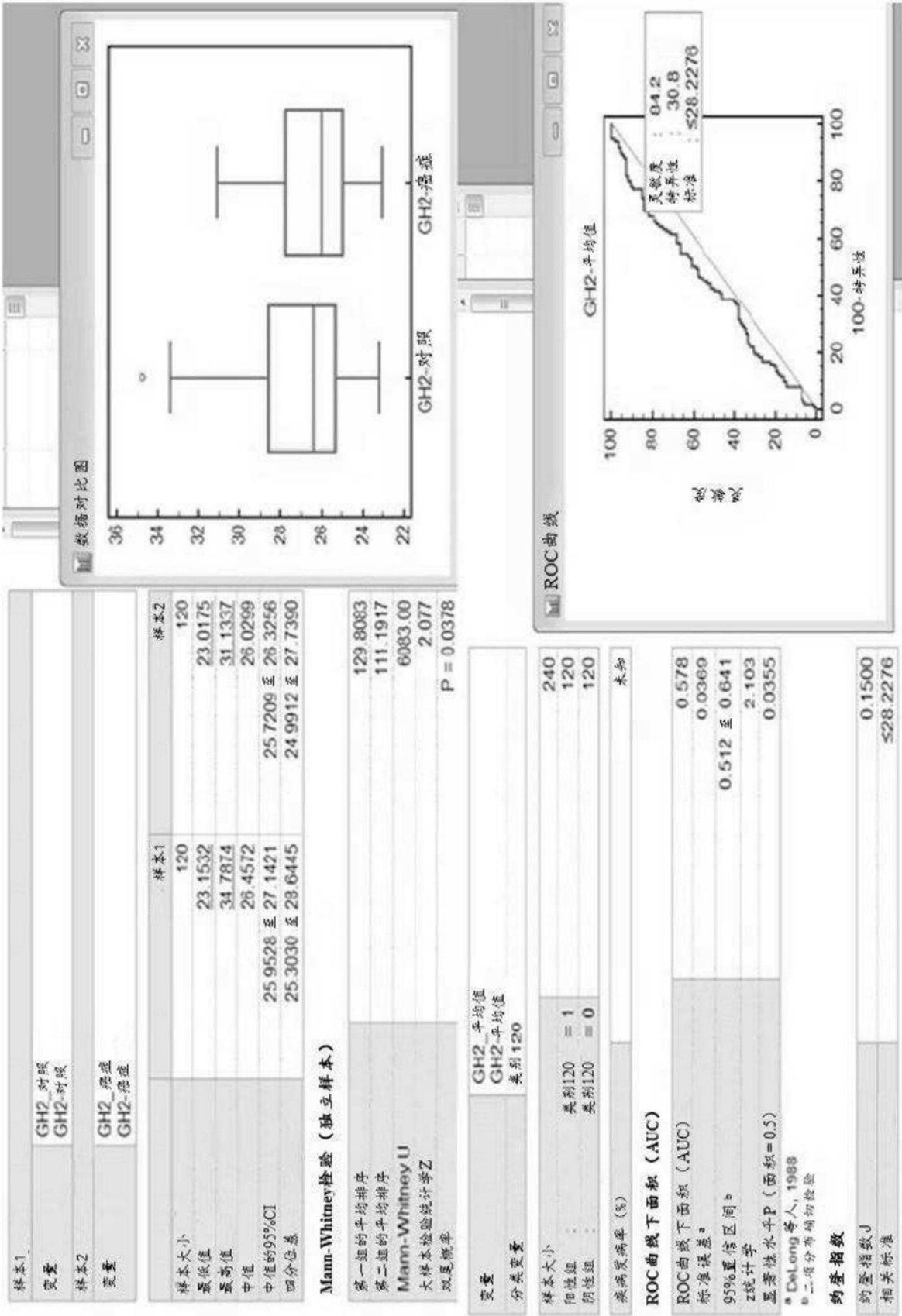


图25

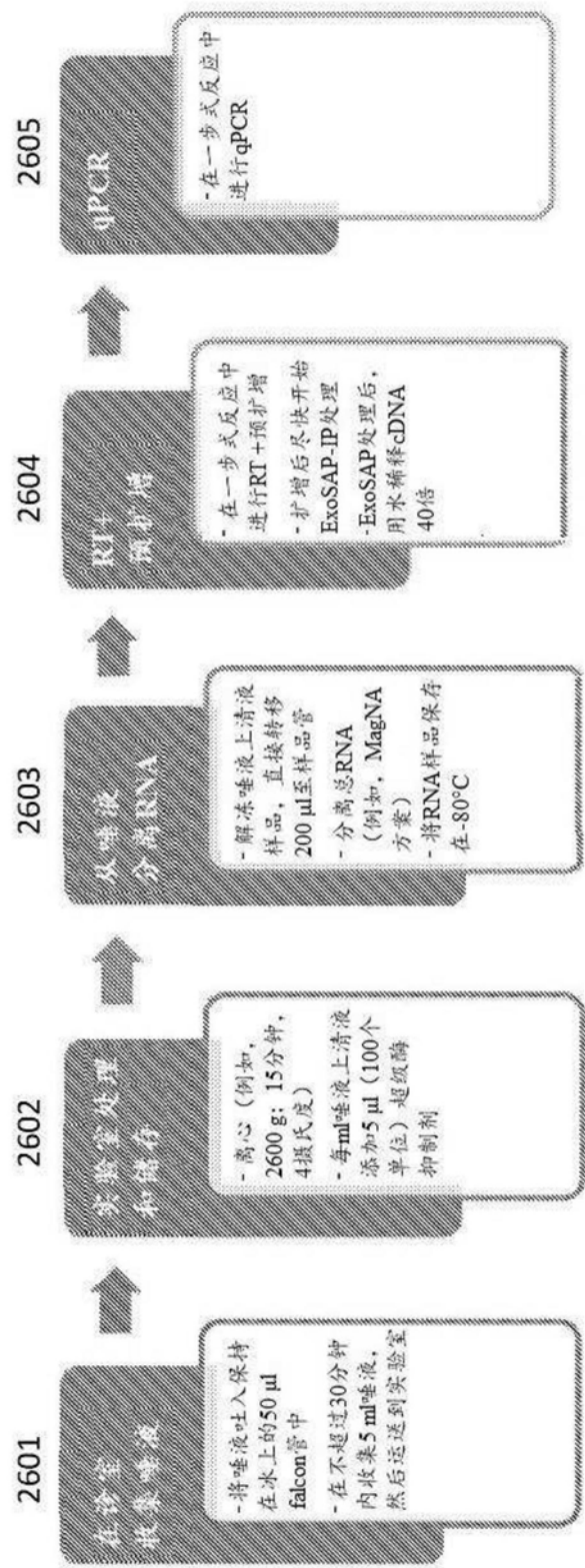


图26

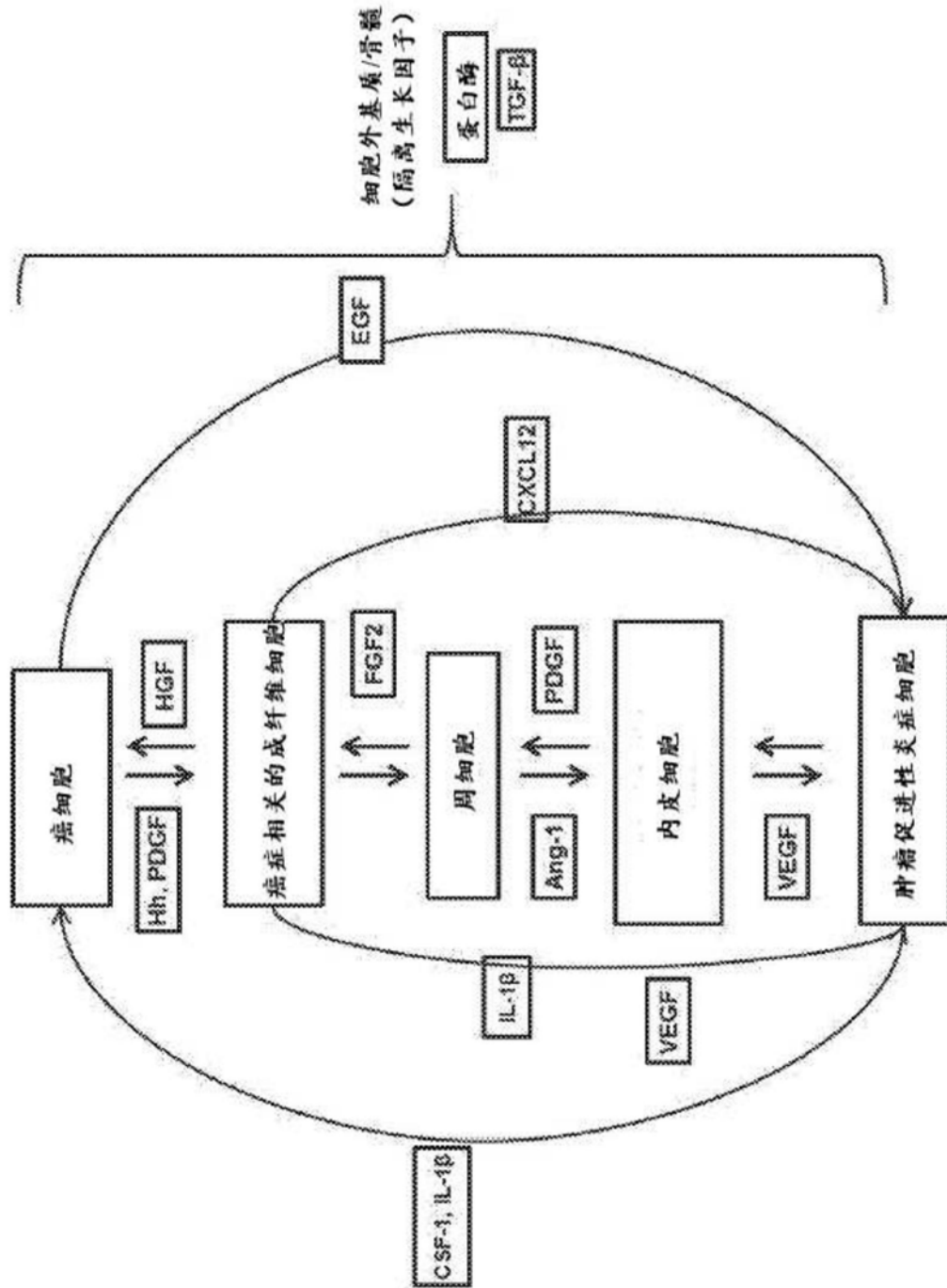


图27

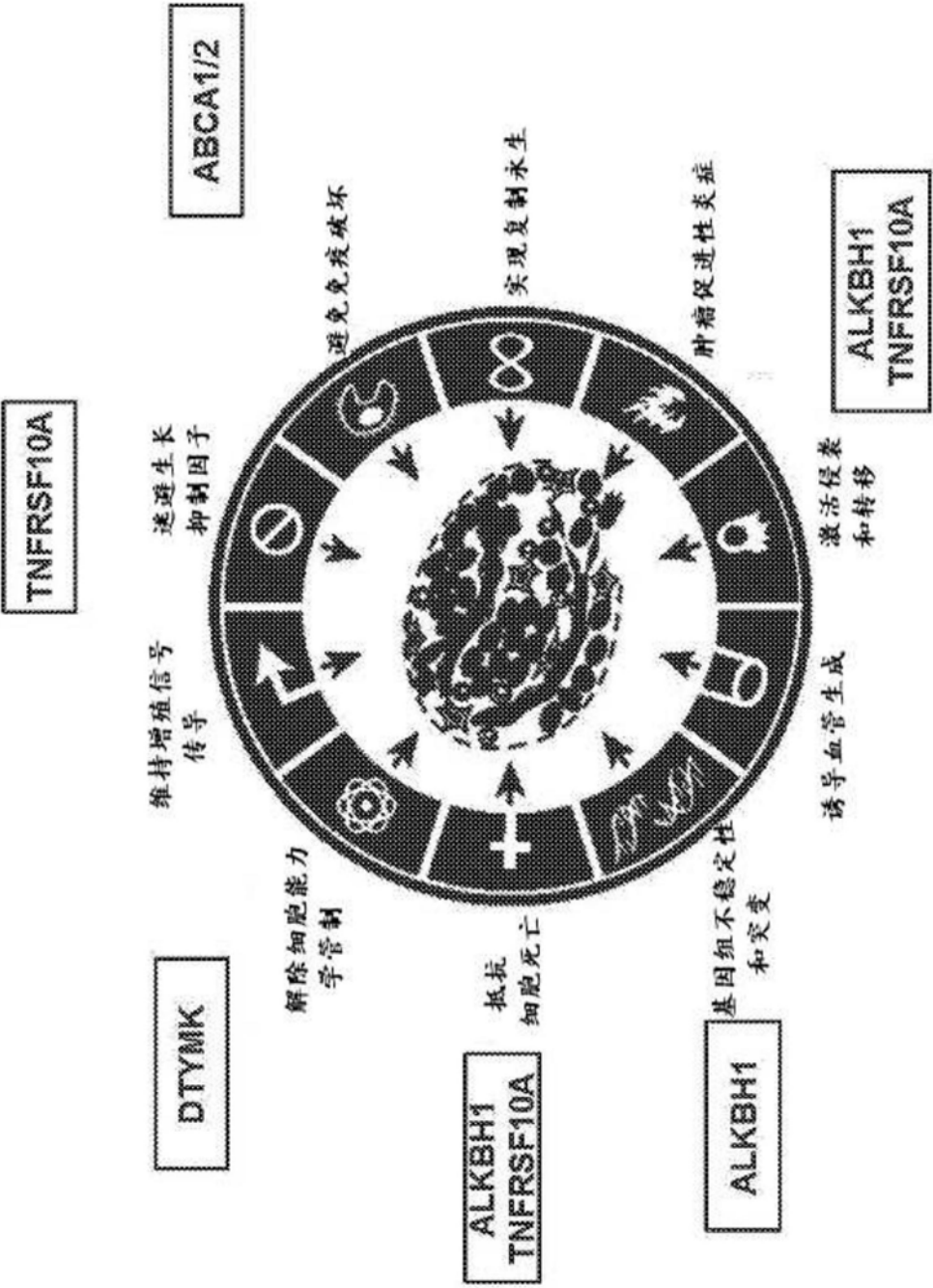


图28

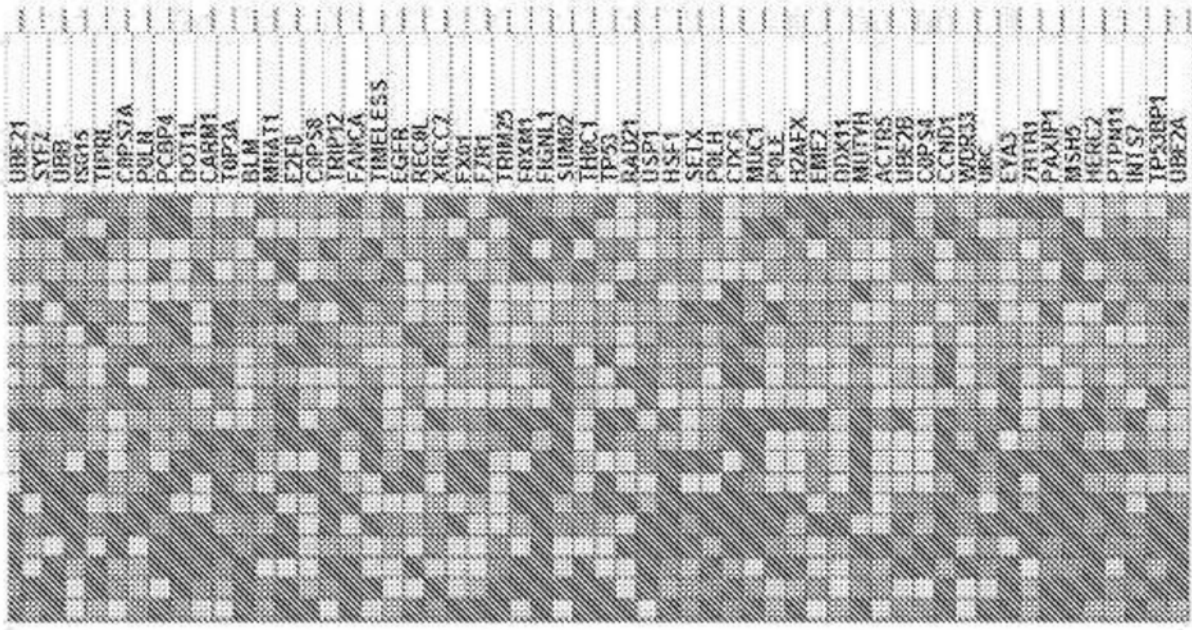


图29