(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 117500941 A
(43)申请公布日 2024.02.02

(21)申请号 202280041034.8
(22)申请日 2022.06.07
(30)优先权数据
63/208,506 2021.06.09 US
(85)PCT国际申请进入国家阶段日
2023.12.08
(86)PCT国际申请的申请数据
PCT/US2022/032423 2022.06.07
(87)PCT国际申请的公布数据
WO2022/261039 EN 2022.12.15
(71)申请人 美瑞诺康诊断公司
地址 美国特拉华州
(72)发明人 安德鲁·张 胡海
(74)专利代理机构 北京启坤知识产权代理有限公司 11655
专利代理师 王飞
(51)Int.Cl.
C12Q 1/6886(2006.01)

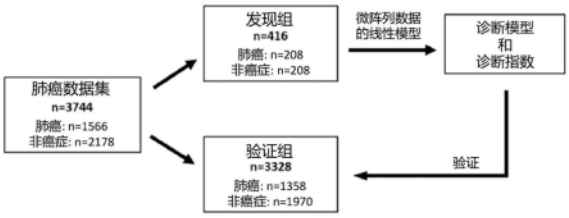
权利要求书8页 说明书26页
序列表16页 附图12页

(54)发明名称

癌症检测方法,试剂盒和系统

(57)摘要

本发明提供了一种方法、试剂盒和系统,能够以高准确度检测一个或多个人类癌症。在基于受试者的液体活检样本确定了一个包含一个或多个miRNAs的miRNA生物标记物组的表达谱之后,计算出一个诊断指数,然后根据该诊断指数可以将所述受试者分类为是否患有癌症。其中,一种包含有4个miRNA生物标记的模型表现出非常高的灵敏度,其针对肺癌和胃癌的灵敏度为99.0-100%,针对胆道、膀胱、结肠直肠、食管、神经胶质瘤、肝癌、胰腺癌和前列腺癌的灵敏度为83.0-99.0%,针对卵巢癌和肉瘤的灵敏度为68.2-72.0%,而特异性能够保持为99.3%。



CN 117500941 A

1.一种用于从受试者获取的生物样本中检测癌症的方法,包括:

确定由所述生物样本中的至少一个miRNA组成的miRNA生物标志物组的表达谱,其中所述miRNA生物标志物组包括hsa-miR-5100;

根据所述miRNA生物标志物组的所述表达谱计算生物样本的诊断指数,其中所述诊断指数是基于以下公式计算的:

$$诊断指数 = \sum_{i=1}^{n} t_i * miRNA_i; \quad (I)$$

其中n是所述miRNA生物标志物组中至少一个miRNA的总数,$miRNA_i$是所述miRNA生物标志物组中第i个miRNA的表达水平,i是大于零且小于或等于n的整数;t_i是第i个miRNA的权重;和

根据计算得到的所述诊断指数将所述受试者分类为有癌或非癌,其中,如果计算得到的所述诊断指数大于或等于预先确定的阈值,则将所述受试者分类为有癌,否则则将所述受试者分类为非癌;

其中,所述方法能够实现具有AUC值大于大约0.780的诊断准确性。

2.根据权利要求1所述的方法,其中,所述miRNA生物标志物组还包含如下的一个或多个miRNA:hsa-miR-1343-3p,hsa-miR-1290,hsa-miR-4787-3p,hsa-miR-6877-5p,hsa-miR-17-3p,hsa-miR-6765-5p,hsa-miR-1268b,hsa-miR-4258,hsa-miR-451a,hsa-miR-1228-5p,hsa-miR-8073,hsa-miR-4454,hsa-miR-187-5p,hsa-miR-4286,hsa-miR-6746-5p,hsa-miR-663b,hsa-miR-6075,hsa-miR-5001-5p,hsa-miR-6789-5p,hsa-miR-4513,hsa-miR-3192-5p,hsa-miR-8060,hsa-miR-668-5p,hsa-miR-1268a,hsa-miR-1273g-3p,hsa-miR-4706,hsa-miR-124-3p,hsa-miR-1260b,hsa-miR-4740-5p,hsa-miR-320b,hsa-miR-7977,hsa-miR-29b-3p,hsa-miR-4708-3p,hsa-miR-4525,hsa-miR-92b-3p,hsa-miR-4257,hsa-miR-4727-3p,hsa-miR-92a-3p,hsa-miR-663a,hsa-miR-6787-5p,hsa-miR-3131,hsa-miR-6802-5p,hsa-miR-654-5p,hsa-miR-6511b-5p,hsa-miR-29b-1-5p,hsa-miR-4417,hsa-miR-4736,hsa-miR-6840-3p,hsa-miR-4710,hsa-miR-4635,hsa-miR-296-3p,hsa-miR-1199-5p,hsa-miR-7975,hsa-miR-4480,hsa-miR-3648,hsa-miR-371a-5p,hsa-miR-4771,hsa-miR-6717-5p,hsa-miR-1254,hsa-miR-1246,hsa-miR-23b-3p,hsa-miR-320a,hsa-miR-4687-5p,hsa-miR-191-5p,hsa-miR-320c,hsa-miR-6131,hsa-miR-4515,hsa-miR-342-5p,hsa-miR-4718,hsa-miR-23a-3p,hsa-miR-4455,hsa-miR-211-3p,hsa-miR-3122,hsa-miR-103a-3p,hsa-miR-4429,hsa-miR-920,hsa-miR-3194-3p,hsa-miR-4754,hsa-miR-1238-5p,hsa-miR-3191-3p,hsa-miR-4755-3p,hsa-miR-3688-5p,hsa-miR-4529-5p,hsa-miR-6861-5p,hsa-miR-1469,hsa-miR-619-5p,hsa-miR-4448,hsa-miR-4658,hsa-miR-22-3p,hsa-miR-4776-5p,hsa-miR-320e,hsa-miR-1225-3p,hsa-miR-6875-5p,hsa-miR-4534,hsa-miR-4652-5p,hsa-miR-648,hsa-miR-4259,hsa-miR-107,以及hsa-miR-650。

3.根据权利要求1所述的方法,其中,所述miRNA生物标志物组还包含如下的一个或多个miRNA:hsa-miR-1343-3p,hsa-miR-1290,hsa-miR-4787-3p,hsa-miR-6877-5p,hsa-miR-17-3p,hsa-miR-6765-5p,hsa-miR-1268b,hsa-miR-4258,hsa-miR-451a,hsa-miR-1228-5p,hsa-miR-8073,hsa-miR-4454,hsa-miR-187-5p,hsa-miR-4286,hsa-miR-6746-5p,hsa-miR-663b,hsa-miR-6075,hsa-miR-5001-5p,hsa-miR-6789-5p,hsa-miR-4513,hsa-miR-

3192-5p,hsa-miR-8060,hsa-miR-668-5p,hsa-miR-1268a,hsa-miR-1273g-3p,hsa-miR-4706,hsa-miR-124-3p,hsa-miR-1260b,hsa-miR-4740-5p,hsa-miR-320b,hsa-miR-7977,hsa-miR-29b-3p,hsa-miR-4708-3p,hsa-miR-4525,hsa-miR-92b-3p,hsa-miR-4257,hsa-miR-4727-3p,hsa-miR-92a-3p,hsa-miR-663a,hsa-miR-6787-5p,hsa-miR-3131,hsa-miR-6802-5p,hsa-miR-654-5p,hsa-miR-6511b-5p,hsa-miR-29b-1-5p,hsa-miR-4417,hsa-miR-4736,hsa-miR-6840-3p,以及hsa-miR-4710。

4.根据权利要求1所述的方法,其中,所述miRNA生物标志物组还包含如下的一个或多个miRNA:hsa-miR-1343-3p,hsa-miR-1290,hsa-miR-4787-3p,hsa-miR-6877-5p,hsa-miR-17-3p,hsa-miR-6765-5p,hsa-miR-1268b,hsa-miR-4258,hsa-miR-451a,hsa-miR-1228-5p,hsa-miR-8073,hsa-miR-4454,hsa-miR-187-5p,hsa-miR-4286,hsa-miR-6746-5p,hsa-miR-663b,hsa-miR-6075,hsa-miR-5001-5p,以及hsa-miR-6789-5p。

5.根据权利要求4所述的方法,其中,所述miRNA生物标志物组由hsa-miR-5100,hsa-miR-1343-3p,hsa-miR-1290,hsa-miR-4787-3p,hsa-miR-6877-5p,hsa-miR-17-3p,hsa-miR-6765-5p,hsa-miR-1268b,hsa-miR-4258,hsa-miR-451a,hsa-miR-1228-5p,hsa-miR-8073,hsa-miR-4454,hsa-miR-187-5p,hsa-miR-4286,hsa-miR-6746-5p,hsa-miR-663b,hsa-miR-6075,hsa-miR-5001-5p,以及hsa-miR-6789-5p组成。

6.根据权利要求1所述的方法,其中,所述miRNA生物标志物组还包含hsa-miR-1343-3p,hsa-miR-1290,以及hsa-miR-4787-3p中的一个或多个。

7.根据权利要求6所述的方法,其中,所述miRNA生物标志物组由hsa-miR-5100,hsa-miR-1343-3p,hsa-miR-1290,以及hsa-miR-4787-3p组成。

8.根据权利要求7所述的方法,其中,所述方法能够实现具有AUC值大于大约0.850的诊断准确性。

9.根据权利要求8所述的方法,其中,所述癌症选自肺癌、胆道癌、膀胱癌、结肠直肠癌、食管癌、胃癌、神经胶质瘤癌、肝癌、胰腺癌、前列腺癌、卵巢癌和肉瘤。

10.根据权利要求8所述的方法,其中,所述方法能够实现具有AUC值大于大约0.950的诊断准确性。

11.根据权利要求10所述的方法,其中,所述癌症选自肺癌、胆道癌、膀胱癌、结肠直肠癌、食管癌、胃癌、神经胶质瘤癌、肝癌、卵巢癌、胰腺癌和前列腺癌。

12.根据权利要求10所述的方法,其中,所述方法能够实现具有AUC值大于大约0.990的诊断准确性。

13.根据权利要求12所述的方法,其中,所述癌症选自肺癌、胆道癌、膀胱癌、食管癌、胃癌、神经胶质瘤癌和前列腺癌。

14.根据权利要求12所述的方法,其中,所述方法能够实现具有AUC值大于大约0.999的诊断准确性。

15.根据权利要求14所述的方法,其中,所述癌症选自肺癌或胃癌。

16.根据权利要求7所述的方法,其中,所述方法能够实现灵敏度高于大约68.0%且特异性高于大约99.0%的诊断准确性。

17.根据权利要求16所述的方法,其中,所述癌症选自肺癌、胆道癌、膀胱癌、结直肠癌、食道癌、胃癌、胶质瘤癌、肝癌、胰腺癌、前列腺癌、卵巢癌和肉瘤。

18.根据权利要求16所述的方法,其中,所述方法能够实现灵敏度高于大约83.0%且特异性高于大约99.0%的诊断准确性。

19.根据权利要求18所述的方法,其中,所述癌症选自肺癌、胆道癌、膀胱癌、结肠直肠癌、食管癌、胃癌、胶质瘤癌、肝癌、胰腺癌、和前列腺癌。

20.根据权利要求18所述的方法,其中所述方法能够实现灵敏度高于大约99.0%,且特异性高于大约99.0%的诊断准确性。

21.根据权利要求20所述的方法,其中,所述癌症选自肺癌和胃癌。

22.根据权利要求1至21中的任何一项权利要求所述的方法,其中,在根据所述miRNA生物标志物组的所述表达谱计算生物样本的诊断指数的步骤中,所述生物样本的诊断指数是通过无加权模型计算的。

23.根据权利要求1至21中的任何一项权利要求所述的方法,其中,在根据所述miRNA生物标记物组的所述表达谱计算生物样本的诊断指数的步骤中,所述诊断指数是通过使用来自线性微阵列数据模型(limma)模型、逻辑回归模型、线性判别分析(LDA)模型、条件逻辑回归模型、套索回归模型、岭回归模型、随机森林、支持向量机和概率回归模型中的其中一种加权模型的权重来计算的。

24.根据权利要求23所述的方法,其中,所述诊断指数是通过使用limma模型的权重来计算。

25.根据权利要求1至24中的任何一项权利要求所述的方法,其中,所述预先确定的阈值为1110,所述方法能够实现具有高于大约0.95的特异性值的诊断准确性。

26.根据权利要求1至24中的任何一项权利要求所述的方法,其中,所述预先确定的阈值为1200,所述方法能够实现具有高于大约0.99的特异性值的诊断准确性。

27.根据权利要求1至26中的任何一项权利要求所述的方法,其中在计算生物样本的诊断指数的步骤之后,并在将受试者分类为有癌或非癌的步骤之前,还包括:

基于计算得到的所述的诊断指数获得归一化诊断指数;

其中:

根据计算得到的所述诊断指数将所述受试者分类为有癌或非癌的步骤包括:

如果所述归一化诊断指数等于或大于预设的阈值,则将所述受试者分类为有癌;

否则将所述受试者分类为非癌。

28.根据权利要求27所述的方法,其中,在基于计算得到的所述诊断指数获得归一化诊断指数的步骤中,所述归一化诊断指数是根据以下公式计算的:

$$归一化诊断指数=\frac{诊断指数-param_{location}}{param_{scale}};\quad (II)$$

其中$param_{location}$和$param_{scale}$分别是位置参数和尺度参数,配置为使归一化诊断指数在不低于第一个预设值且不大于第二个预设值的范围内。

29.根据权利要求28所述的方法,其中,所述诊断指数是通过使用来自limma模型的权重进行计算的,其中,所述第一个预设值为0,所述第二个预设值为10。

30.根据权利要求29所述的方法,其中,所述预设的阈值为5.1,所述方法能够实现特异性值高于大约0.95的诊断准确度。

31.根据权利要求29所述的方法,其中,所述预设的阈值为6.0,所述方法能够实现特异

性值高于大约0.99的诊断准确度。

32.根据权利要求1至31所述的任一方法,其中,所述生物样本选自血液样本、血清样本、血浆样本、尿液样本、唾液样本或痰液样本的其中一种液体活检样本。

33.根据权利要求1至32所述的任一方法,其中,在确定由所述生物样本中的至少一个miRNA组成的miRNA生物标志物组的表达谱的步骤中,所述miRNA生物标志物组的表达谱是通过Northern Blotting、微阵列分析、RNA测序或RNA原位杂交等方法之一获得的。

34.根据权利要求1至32所述的任一方法,其中,在确定由所述生物样本中的至少一个miRNA组成的miRNA生物标志物组的表达谱的步骤中,所述miRNA生物标志物组的表达谱是通过核酸扩增程序获得的,所述核酸扩增程序包括逆转录聚合酶链式反应(RT-PCR)、定量RT-PCR(qRT-PCR)或数字RT-PCR中的至少一种。

35.根据权利要求1至34所述的任一方法,其中还包括:对受试者进行评估,所述评估包括对所述癌症的诊断或所述癌症复发的检测。

36.根据权利要求1至35所述的任一方法,其中还包括:在将受试者分类为癌症患者时,向受试者施行治疗方案。

37.一种用于从来自受试者的生物样本中检测癌症的试剂盒,包括至少一种核酸和至少一条指令,其中:

所述至少一种核酸能够特异性地识别miRNA生物标记物集中的每个miRNA,从而允许从所述生物样本中获取所述miRNA生物标记物集的表达谱,其中所述miRNA生物标记物集包括hsa-miR-5100;

所述至少一条指令包括:

第一指令,包括第一子指令,用于基于所述miRNA生物标记物组的表达谱计算所述生物样本的诊断指数,其中所述诊断指数根据以下公式计算:

$$诊断指数 = \sum_{i=1}^{n} t_i * miRNA_i; \quad (I)$$

其中n是所述miRNA生物标志物组中至少一个miRNA的总数,$miRNA_i$是所述miRNA生物标志物组中第i个miRNA的表达水平,i是大于零且小于或等于n的整数;t_i是第i个miRNA的权重;以及

第二指令,用于将所述受试者分类为是否患有癌症;其中,如果计算得到的所述诊断指数大于或等于预先确定的阈值,则将所述受试者分类为有癌;否则,则将所述受试者分类为非癌。

38.根据权利要求37所述的试剂盒,其中,所述至少一个核酸包含一个多核苷酸,其能够严格地和下面两个多核苷酸之一进行特异杂交:(a)包含具有SEQ ID NO:1序列的多核苷酸,以及其衍生物、具有至少80%相同序列的变体、或者包含15个或更多个连续核苷酸的片段;(b)包含具有和SEQ ID NO:1序列互补的序列的多核苷酸,以及其衍生物、具有至少80%相同序列的变体、或者包含15个或更多个连续核苷酸的片段。

39.根据权利要求37或权利要求38所述的试剂盒,其中,所述miRNA标志物组还包括以下一个或多个miRNA:hsa-miR-1343-3p,hsa-miR-1290,hsa-miR-4787-3p,hsa-miR-6877-5p,hsa-miR-17-3p,hsa-miR-6765-5p,hsa-miR-1268b,hsa-miR-4258,hsa-miR-451a,hsa-miR-1228-5p,hsa-miR-8073,hsa-miR-4454,hsa-miR-187-5p,hsa-miR-4286,hsa-miR-6746-5p,hsa-miR-663b,hsa-miR-6075,hsa-miR-5001-5p,hsa-miR-6789-5p,hsa-miR-

4513,hsa-miR-3192-5p,hsa-miR-8060,hsa-miR-668-5p,hsa-miR-1268a,hsa-miR-1273g-3p,hsa-miR-4706,hsa-miR-124-3p,hsa-miR-1260b,hsa-miR-4740-5p,hsa-miR-320b,hsa-miR-7977,hsa-miR-29b-3p,hsa-miR-4708-3p,hsa-miR-4525,hsa-miR-92b-3p,hsa-miR-4257,hsa-miR-4727-3p,hsa-miR-92a-3p,hsa-miR-663a,hsa-miR-6787-5p,hsa-miR-3131,hsa-miR-6802-5p,hsa-miR-654-5p,hsa-miR-6511b-5p,hsa-miR-29b-1-5p,hsa-miR-4417,hsa-miR-4736,hsa-miR-6840-3p,hsa-miR-4710,hsa-miR-4635,hsa-miR-296-3p,hsa-miR-1199-5p,hsa-miR-7975,hsa-miR-4480,hsa-miR-3648,hsa-miR-371a-5p,hsa-miR-4771,hsa-miR-6717-5p,hsa-miR-1254,hsa-miR-1246,hsa-miR-23b-3p,hsa-miR-320a,hsa-miR-4687-5p,hsa-miR-191-5p,hsa-miR-320c,hsa-miR-6131,hsa-miR-4515,hsa-miR-342-5p,hsa-miR-4718,hsa-miR-23a-3p,hsa-miR-4455,hsa-miR-211-3p,hsa-miR-3122,hsa-miR-103a-3p,hsa-miR-4429,hsa-miR-920,hsa-miR-3194-3p,hsa-miR-4754,hsa-miR-1238-5p,hsa-miR-3191-3p,hsa-miR-4755-3p,hsa-miR-3688-5p,hsa-miR-4529-5p,hsa-miR-6861-5p,hsa-miR-1469,hsa-miR-619-5p,hsa-miR-4448,hsa-miR-4658,hsa-miR-22-3p,hsa-miR-4776-5p,hsa-miR-320e,hsa-miR-1225-3p,hsa-miR-6875-5p,hsa-miR-4534,hsa-miR-4652-5p,hsa-miR-648,hsa-miR-4259,hsa-miR-107,以及hsa-miR-650。

40.根据权利要求39所述的试剂盒,其中,所述至少一个核酸还包含至少一个多核苷酸,每一个能够严格地和下面两个多核苷酸之一进行特异杂交:(a)包含具有SEQ ID NO:2-100之一序列的多核苷酸,以及其衍生物、具有至少80%相同序列的变体、或者包含15个或更多个连续核苷酸的片段;(b)包含具有和SEQ ID NO:2-100之一序列互补的序列的多核苷酸,以及其衍生物、具有至少80%相同序列的变体、或者包含15个或更多个连续核苷酸的片段。

41.根据权利要求37或权利要求38所述的试剂盒,其中,所述miRNA标志物集合还包括以下一个或多个miRNA:hsa-miR-1343-3p,hsa-miR-1290,hsa-miR-4787-3p,hsa-miR-6877-5p,hsa-miR-17-3p,hsa-miR-6765-5p,hsa-miR-1268b,hsa-miR-4258,hsa-miR-451a,hsa-miR-1228-5p,hsa-miR-8073,hsa-miR-4454,hsa-miR-187-5p,hsa-miR-4286,hsa-miR-6746-5p,hsa-miR-663b,hsa-miR-6075,hsa-miR-5001-5p,hsa-miR-6789-5p,hsa-miR-4513,hsa-miR-3192-5p,hsa-miR-8060,hsa-miR-668-5p,hsa-miR-1268a,hsa-miR-1273g-3p,hsa-miR-4706,hsa-miR-124-3p,hsa-miR-1260b,hsa-miR-4740-5p,hsa-miR-320b,hsa-miR-7977,hsa-miR-29b-3p,hsa-miR-4708-3p,hsa-miR-4525,hsa-miR-92b-3p,hsa-miR-4257,hsa-miR-4727-3p,hsa-miR-92a-3p,hsa-miR-663a,hsa-miR-6787-5p,hsa-miR-3131,hsa-miR-6802-5p,hsa-miR-654-5p,hsa-miR-6511b-5p,hsa-miR-29b-1-5p,hsa-miR-4417,hsa-miR-4736,hsa-miR-6840-3p,以及hsa-miR-4710。

42.根据权利要求41所述的试剂盒,其中,所述至少一个核酸还包含至少一个多核苷酸,每一个能够严格地和下面两个多核苷酸之一进行特异杂交:(a)包含具有SEQ ID NO:2-50之一序列的多核苷酸,以及其衍生物、具有至少80%相同序列的变体、或者包含15个或更多个连续核苷酸的片段;(b)包含具有和SEQ ID NO:2-50之一序列互补的序列的多核苷酸,以及其衍生物、具有至少80%相同序列的变体、或者包含15个或更多个连续核苷酸的片段。

43.根据权利要求37或权利要求38所述的试剂盒,其中,所述miRNA标志物集合还包括

以下一个或多个miRNA:hsa-miR-1343-3p,hsa-miR-1290,hsa-miR-4787-3p,hsa-miR-6877-5p,hsa-miR-17-3p,hsa-miR-6765-5p,hsa-miR-1268b,hsa-miR-4258,hsa-miR-451a,hsa-miR-1228-5p,hsa-miR-8073,hsa-miR-4454,hsa-miR-187-5p,hsa-miR-4286,hsa-miR-6746-5p,hsa-miR-663b,hsa-miR-6075,hsa-miR-5001-5p,以及hsa-miR-6789-5p。

44.根据权利要求43所述的试剂盒,其中,所述至少一个核酸还包括至少一个多核苷酸,每一个能够严格地和下面两个多核苷酸之一进行特异杂交:(a)包含具有SEQ ID NO:2-20之一序列的多核苷酸,以及其衍生物、具有至少80%相同序列的变体、或者包含15个或更多个连续核苷酸的片段;(b)包含具有和SEQ ID NO:2-20之一序列互补的序列的多核苷酸,以及其衍生物、具有至少80%相同序列的变体、或者包含15个或更多个连续核苷酸的片段。

45.根据权利要求43所述的试剂盒,其中,所述miRNA标志物集合包括hsa-miR-5100,hsa-miR-1343-3p,hsa-miR-1290,hsa-miR-4787-3p,hsa-miR-6877-5p,hsa-miR-17-3p,hsa-miR-6765-5p,hsa-miR-1268b,hsa-miR-4258,hsa-miR-451a,hsa-miR-1228-5p,hsa-miR-8073,hsa-miR-4454,hsa-miR-187-5p,hsa-miR-4286,hsa-miR-6746-5p,hsa-miR-663b,hsa-miR-6075,hsa-miR-5001-5p,以及hsa-miR-6789-5p。

46.根据权利要求45所述的试剂盒,其中,所述至少有一个核酸包含总共20个多核苷酸,分别能够严格地和下面两个多核苷酸之一进行特异杂交:(a)包含具有SEQ ID NO:1-20之一序列的多核苷酸,以及其衍生物、或者具有至少80%相同序列的变体、或者包含15个或更多个连续核苷酸的片段;(b)包含具有和SEQ ID NO:1-20之一序列互补的序列的多核苷酸,以及其衍生物、或者具有至少80%相同序列的变体、或者包含15个或更多个连续核苷酸的片段。

47.根据权利要求37或权利要求38所述的试剂盒,其中,所述miRNA标志物集合还包括以下一个或多个miRNA:hsa-miR-1343-3p,hsa-miR-1290,以及hsa-miR-4787-3p。

48.根据权利要求47所述的试剂盒,其中,所述至少一个核酸还包括至少一个多核苷酸,每一个能够严格地和下面两个多核苷酸之一进行特异杂交:(a)包含具有SEQ ID NO:2-4之一序列的多核苷酸,以及其衍生物、或者具有至少80%相同序列的变体、或者包含15个或更多个连续核苷酸的片段;(b)包含具有和SEQ ID NO:2-4之一序列互补的序列的多核苷酸,以及其衍生物、或者具有至少80%相同序列的变体、或者包含15个或更多个连续核苷酸的片段。

49.根据权利要求47所述的试剂盒,其中,所述miRNA标志物集合包括hsa-miR-5100,hsa-miR-1343-3p,hsa-miR-1290,以及hsa-miR-4787-3p。

50.根据权利要求49所述的试剂盒,其中,所述至少一个核酸包含总共4个多核苷酸,分别能够严格地和下面两个多核苷酸之一进行特异杂交:(a)包含具有SEQ ID NO:1-4之一序列的多核苷酸,以及其衍生物、或者具有至少80%相同序列的变体、或者包含15个或更多个连续核苷酸的片段;(b)包含具有和SEQ ID NO:1-4之一序列互补的序列的多核苷酸,以及其衍生物、或者具有至少80%相同序列的变体、或者包含15个或更多个连续核苷酸的片段。

51.根据权利要求37-50中任何一个权利要求所述的试剂盒,其中,在所述第一项指令的所述第一子指令中,所述诊断指数通过无权重模型来计算。

52.根据权利要求37-50中任何一个权利要求所述的试剂盒,其中,在所述第一项指令

的所述第一子指令中,所述诊断指数是通过使用来自从线性模型用于微阵列数据(limma)模型,逻辑回归模型,线性判别分析(LDA)模型,条件逻辑回归模型,套索回归模型,岭回归模型,随机森林,支持向量机,以及概率回归模型的其中一种加权模型的权重来计算的。

53.根据权利要求52所述的试剂盒,其中,所述诊断指数是用来自微阵列数据(limma)模型的加权模型的权重来进行计算的。

54.根据权利要求37-53中任何一个权利要求所述的试剂盒,其中所述预先确定的阈值为1110,并且所述第二项指令还包括分类具有大约0.95以上的特异性值的指令。

55.根据权利要求37-53中任何一个权利要求所述的试剂盒,其中所述预先确定的阈值为1200,并且所述第二项指令还包括分类具有大约0.99以上的特异性值的指令。

56.根据权利要求37-55中任何一个权利要求所述的试剂盒,其中所述第一项指令还包括用于基于根据所述第一条子指令计算的所述诊断指数获取归一化诊断指数的第二条子指令,在所述第二项指令中,如果所述归一化诊断指数大于或等于预设的阈值,则将所述受试者分类为有癌,否则将所述受试者分类为非癌。

57.根据权利要求56所述的试剂盒,其中在所述第二个子指令中,所述归一化诊断指数根据以下公式计算:

$$归一化诊断指数=\frac{诊断指数-para__location}{param_{scale}};\quad (II)$$

其中$param_{location}$和$param_{scale}$分别是位置参数和尺度参数,用于确保所述归一化诊断指数在不低于第一个预设值且不大于第二个预设值的范围内。

58.根据权利要求57所述的试剂盒,其中,在所述第一指令中,所述诊断指数采用来自微阵列数据(limma)模型的权重来计算的,并且所述第一个预设值为0,所述第二个预设值为10。

59.根据权利要求58所述的试剂盒,其中所述预先确定的阈值为5.1,并且所述第二指令还包括分类具有大约0.95以上的特异性值的指令。

60.根据权利要求58所述的试剂盒,其中所述预先确定的阈值为6.0,并且所述第二指令还包括分类具有大约0.95以上的特异性值的指令。

61.根据权利要求37-60中任何一项权利要求所述的试剂盒,其中所述至少一条指令还包括第三条指令,用于对所述受试者进行评估,其中所述评估包括对所述癌症的诊断或对所述癌症复发的检测。

62.根据权利要求37-61中任何一项权利要求所述的试剂盒,其中所述至少一条指令还包括第四条指令,用于在将所述受试者分类为患有癌症时向所述受试者施行治疗方案。

63.根据权利要求37-62中任何一项权利要求所述的试剂盒,其中所述至少一条指令还包括第一条附加指令,用于获取所述miRNA生物标志物组的表达谱,包括通过至少一条核酸进行Northern Blotting、微阵列分析、RNA测序或RNA原位杂交的过程。

64.根据权利要求63所述的试剂盒,其中所述至少一条核酸排列在分子阵列上。

65.根据权利要求37-62中任何一项权利要求所述的试剂盒,还至少包括一组扩增引物,每组能够从所述生物样品中特异性扩增所述miRNA标志物集合中的至少一个miRNA。

66.根据权利要求65所述的试剂盒,其中所述至少一份指令还包括一份用于获取所述miRNA标志物集合的表达谱的第二附加指令,其包括通过利用所述至少一份核酸和所述至

少一组扩增引物执行逆转录聚合酶链反应(RT-PCR)、定量RT-PCR(qRT-PCR)或数字RT-PCR的程序。

67.根据权利要求37-66中任何一项权利要求所述的试剂盒,其中所述生物样本是从以下液体活检样本中选择的,包括血液样本、血清样本、血浆样本、尿液样本、唾液样本和痰样本。

68.一个用于检测受试者癌症的系统,包括:

一个处理器;和

一个非瞬态存储介质,其中包含用于由所述处理器执行的程序,所述程序引导所述处理器执行根据权利要求1至36中任何一项权利要求所述的方法中的步骤。

69.一种非瞬态存储介质,存储有计算机可执行的程序,当这些程序被处理器执行时,会引导处理器执行权利要求1至36中任何一项权利要求所述的方法。

癌症检测方法,试剂盒和系统

[0001] 相关申请的交叉引用

[0002] 该申请要求于2021年6月9日提交的申请号为63/208,506的美国临时申请的各项权益,该临时申请的全部内容通过引用整体并入本申请。

[0003] 已电子方式提交的序列清单参考

[0004] 电子方式提交的序列清单内容,文件名为Top_miRNA_Seq.txt,文件大小为15,063字节,创建日期为2022年5月31日,随附提交,其全部内容通过引用整体并入本申请。

发明领域

[0005] 本发明属于疾病筛查、检测和诊断技术领域,更具体而言,本发明涉及一种用于检测一个或多个人类癌症的方法、试剂盒、系统以及相关的非瞬态存储介质。

[0006] 背景

[0007] 尽管近年来,诊断和治疗技术迅速发展,但癌症仍然是人类面临的一个极具挑战性和潜在致命性的疾病。众所周知,癌症在早期阶段的检测对于降低与癌症相关的死亡率至关重要,因为在早期阶段治疗更有可能成功。因此,迫切需要开发一种能够早期同时检测多种癌症类型的测试,最好是无创的,例如血液测试。这种无创血液检测已成为多癌早期检测(简称为"MCED测试")的基石。MCED测试通常需要非常高的特异性,最好>99%,以确保在筛查高风险人群时出现最少的假阳性。

[0008] 分子如小分子核糖核酸(microRNA,简称miRNA),可以作为MCED的生物标志物。miRNA是在人类基因组中一类由内源基因编码的长度约为22个核苷酸的非编码单链RNA分子。miRNA主要通过与mRNA分子的3'非翻译区(3'UTR)中的互补序列结合来实现基因表达的负后转录调控。miRNA表现出可以调控超过50%的人类基因,miRNA的异常表达也被认为与许多人类癌症有关。很多miRNA也作为细胞外循环分子参与细胞死亡调节或通过外泌体介导的信号传递方式,通过肿瘤细胞释放到血液中结合其在血液和其他体液中的显著稳定性,循环系统中的细胞游离miRNA具有作为癌症筛查和诊断的无创生物标志物的潜力。

发明摘要

[0009] 本发明公开了一种多癌症检测方法(即方法、试剂盒和系统)。这种方法采用一组miRNA生物标志物,一组中至少包含一个miRNA。该方法主要基于miRNA生物标志物组的表达谱,该表达谱可以从人类受试者的生物样本中获得。这种生物样本尤其以液体活检样本最显著,包括血液样本、血清样本、血浆样本、尿液样本、唾液样本或唾液样本,从而实现对癌症的无创或最小侵入式检测。该方法可用于准确可靠地检测人类受试者是否患有肺癌、胆道癌、膀胱癌、结直肠癌、食管癌、胃癌、神经胶质瘤、肝癌、胰腺癌、前列腺癌、卵巢癌和肉瘤等癌症之一。

[0010] 第一个方面,本发明提供了一种从来自受试者的生物样本中检测癌症的方法。该方法主要包括以下三个步骤(1)-(3):

[0011] 步骤(1):从所述生物样本中确定由至少一个miRNA组成的miRNA生物标志物组的

表达谱。在此,miRNA生物标志物组包括hsa-miR-5100。

[0012] 步骤(2):基于miRNA生物标志物组的表达谱计算生物样本的诊断指数。诊断指数的计算公式如下:

[0013] $$诊断指数 = \sum_{i=1}^{n} t_i * miRNA_i; \quad (I)$$

[0014] 其中,n是miRNA生物标志物组中miRNA的总数,$miRNA_i$是miRNA生物标志物组中第i个miRNA的表达水平,其中i是大于零且小于或等于n的整数;t_i是第i个miRNA的权重。

[0015] 步骤(3):根据计算得到的诊断指数的数值,对受试者进行是否患有癌症的分类。如果计算得到的诊断指数大于或等于预定的阈值,将受试者分类为患有癌症;否则,将受试者分类为未患有癌症。

[0016] 该方法经过额外的参数配置,能够实现具有AUC值大于大约0.780的诊断准确性。

[0017] 本文中,miRNA生物标志物组的表达谱基本上是一个包含miRNA生物标志物组中的每个miRNA表达水平数据的数据集。

[0018] 术语“预先确定的阈值”是指可以用来确定在给定特异性/灵敏度下,受试者是否患有癌症类型的诊断指数阈值值。通常,一个已知的数据库包含了病患和非病患的诊断指数数据,由此可以预先计算出一系列诊断指数值。例如,在下面提供的实施例1中,这个预定阈值是miRNA生物标志物组包括前100个miRNA中的任何一个(对应SEQ ID NOS:1-100),且AUC可以达到大于0.780的水平(例如hsa-miR-1238-5p),甚至可以达到约0.999的水平(例如,前4个miRNAs,hsa-miR-5100、hsa-miR-1343-3p、hsa-miR-1290和hsa-miR-4787-3p)(参见表1)。

[0019] 根据该方法的一些实施例,miRNA生物标志物组还包括除hsa-miR-5100(对应SEQ ID NO:1)之外的其他99个miRNAs,如表1中所列,即hsa-miR-1343-3p、hsa-miR-1290、hsa-miR-4787-3p、hsa-miR-6877-5p、hsa-miR-17-3p、hsa-miR-6765-5p、hsa-miR-1268b、hsa-miR-4258、hsa-miR-451a、hsa-miR-1228-5p、hsa-miR-8073、hsa-miR-4454、hsa-miR-187-5p、hsa-miR-4286、hsa-miR-6746-5p、hsa-miR-663b、hsa-miR-6075、hsa-miR-5001-5p、hsa-miR-6789-5p、hsa-miR-4513、hsa-miR-3192-5p、hsa-miR-8060、hsa-miR-668-5p、hsa-miR-1268a、hsa-miR-1273g-3p、hsa-miR-4706、hsa-miR-124-3p、hsa-miR-1260b、hsa-miR-4740-5p、hsa-miR-320b、hsa-miR-7977、hsa-miR-29b-3p、hsa-miR-4708-3p、hsa-miR-4525、hsa-miR-92b-3p、hsa-miR-4257、hsa-miR-4727-3p、hsa-miR-92a-3p、hsa-miR-663a、hsa-miR-6787-5p、hsa-miR-3131、hsa-miR-6802-5p、hsa-miR-654-5p、hsa-miR-6511b-5p、hsa-miR-29b-1-5p、hsa-miR-4417、hsa-miR-4736、hsa-miR-6840-3p、hsa-miR-4710、hsa-miR-4635、hsa-miR-296-3p、hsa-miR-1199-5p、hsa-miR-7975、hsa-miR-4480、hsa-miR-3648、hsa-miR-371a-5p、hsa-miR-4771、hsa-miR-6717-5p、hsa-miR-1254、hsa-miR-1246、hsa-miR-23b-3p、hsa-miR-320a、hsa-miR-4687-5p、hsa-miR-191-5p、hsa-miR-320c、hsa-miR-6131、hsa-miR-4515、hsa-miR-342-5p、hsa-miR-4718、hsa-miR-23a-3p、hsa-miR-4455、hsa-miR-211-3p、hsa-miR-3122、hsa-miR-103a-3p、hsa-miR-4429、hsa-miR-920、hsa-miR-3194-3p、hsa-miR-4754、hsa-miR-1238-5p、hsa-miR-3191-3p、hsa-miR-4755-3p、hsa-miR-3688-5p、hsa-miR-4529-5p、hsa-miR-6861-5p、hsa-miR-1469、hsa-miR-619-5p、hsa-miR-4448、hsa-miR-4658、hsa-miR-22-3p、hsa-miR-4776-5p、hsa-miR-320e、hsa-miR-1225-3p、hsa-miR-6875-5p、hsa-miR-4534、hsa-miR-4652-5p、hsa-miR-648、hsa-miR-

4259、hsa-miR-107和hsa-miR-650,这些miRNA根据调整后的P值排名,分别对应SEQ ID NOS:2-100。

[0020] 根据该方法的其他一些实施例,miRNA生物标志物组还包括除hsa-miR-5100以外的表1中排名前50的一个或多个miRNAs,即hsa-miR-1343-3p、hsa-miR-1290、hsa-miR-4787-3p、hsa-miR-6877-5p、hsa-miR-17-3p、hsa-miR-6765-5p、hsa-miR-1268b、hsa-miR-4258、hsa-miR-451a、hsa-miR-1228-5p、hsa-miR-8073、hsa-miR-4454、hsa-miR-187-5p、hsa-miR-4286、hsa-miR-6746-5p、hsa-miR-663b、hsa-miR-6075、hsa-miR-5001-5p、hsa-miR-6789-5p、hsa-miR-4513、hsa-miR-3192-5p、hsa-miR-8060、hsa-miR-668-5p、hsa-miR-1268a、hsa-miR-1273g-3p、hsa-miR-4706、hsa-miR-124-3p、hsa-miR-1260b、hsa-miR-4740-5p、hsa-miR-320b、hsa-miR-7977、hsa-miR-29b-3p、hsa-miR-4708-3p、hsa-miR-4525、hsa-miR-92b-3p、hsa-miR-4257、hsa-miR-4727-3p、hsa-miR-92a-3p、hsa-miR-663a、hsa-miR-6787-5p、hsa-miR-3131、hsa-miR-6802-5p、hsa-miR-654-5p、hsa-miR-6511b-5p、hsa-miR-29b-1-5p、hsa-miR-4417、hsa-miR-4736、hsa-miR-6840-3p和hsa-miR-4710,这些miRNAs根据调整后的P值排名,分别对应SEQ ID NOS:2-50。

[0021] 根据该方法的其他一些实施例,miRNA生物标志物组还包括除hsa-miR-5100以外的表1中排名前20的一个或多个miRNAs,即hsa-miR-1343-3p、hsa-miR-1290、hsa-miR-4787-3p、hsa-miR-6877-5p、hsa-miR-17-3p、hsa-miR-6765-5p、hsa-miR-1268b、hsa-miR-4258、hsa-miR-451a、hsa-miR-1228-5p、hsa-miR-8073、hsa-miR-4454、hsa-miR-187-5p、hsa-miR-4286、hsa-miR-6746-5p、hsa-miR-663b、hsa-miR-6075、hsa-miR-5001-5p和hsa-miR-6789-5p,这些miRNAs根据调整后的P值排名,分别对应SEQ ID NOS:2-20。在这里还可以选择,miRNA生物标志物组由表1中排名前20的miRNA组成(分别对应SEQ ID NOS:1-20)。

[0022] 根据该方法的其他一些实施例,miRNA生物标志物组还包括除hsa-miR-5100以外的表1中排名前4的一个或多个miRNAs,即hsa-miR-1343-3p、hsa-miR-1290和hsa-miR-4787-3p,这些miRNAs根据调整后的P值排名,分别对应SEQ ID NOS:2-4。在这里还可以选择,miRNA生物标志物组由表1中排名前4的miRNAs组成,即hsa-miR-5100、hsa-miR-1343-3p、hsa-miR-1290和hsa-miR-4787-3p,分别对应SEQ ID NOS:1-4。

[0023] 该方法经进一步的参数配置,能够实现具有更高AUC值的诊断准确程度。

[0024] 根据某些实施例,该方法经过参数配置,能够实现具有AUC值大于大约0.850的诊断准确性。可检测的癌症包括肺癌、胆道癌、膀胱癌、结肠癌、食道癌、胃癌、神经胶质瘤癌、肝癌、胰腺癌、前列腺癌、卵巢癌和肉瘤。

[0025] 根据某些实施例,该方法经过参数配置,能够实现具有AUC值大于大约0.950的诊断准确性。可检测的癌症包括肺癌、胆道癌、膀胱癌、结肠癌、食道癌、胃癌、神经胶质瘤癌、肝癌、卵巢癌、胰腺癌和前列腺癌。

[0026] 根据某些实施例,该方法经过参数配置,能够实现具有AUC值大于大约0.990的诊断准确性。可检测的癌症包括肺癌、胆道癌、膀胱癌、食道癌、胃癌、神经胶质瘤癌和前列腺癌。

[0027] 根据某些实施例,该方法经过参数配置,能够实现具有AUC值大于大约0.999的诊断准确性。在这里可检测的癌症是肺癌或胃癌。

[0028] 根据不同的实际需求,该方法可以通过不同的参数配置,来实现具有不同灵敏度

和特异性水平的诊断准确程度。

[0029] 根据某些实施例,该方法经过参数配置,能够实现具有灵敏度大于大约68.0%和特异性大于大约99.0%的诊断准确性。可检测的癌症包括肺癌、胆道癌、膀胱癌、结肠癌、食道癌、胃癌、神经胶质瘤癌、肝癌、胰腺癌、前列腺癌、卵巢癌和肉瘤。

[0030] 根据某些实施例,该方法经过参数配置,能够实现具有灵敏度大于大约83.0%和特异性大于大约99.0%的诊断准确性。可检测的癌症包括肺癌、胆道癌、膀胱癌、结肠癌、食道癌、胃癌、神经胶质瘤癌、肝癌、胰腺癌和前列腺癌。

[0031] 根据某些实施例,该方法经过参数配置,能够实现具有灵敏度大于大约99.0%和特异性大于大约99.0%的诊断准确性。可检测的癌症是肺癌或胃癌。

[0032] 根据该方法的某些实施例,在第(2)步中,通过miRNA生物标志物组的表达谱计算生物样本的诊断指数,也可以通过无权重模型计算的。

[0033] 根据该方法的其他一些实施例,在第(2)步中,通过miRNA生物标志物组的表达谱计算生物样本的诊断指数,可以通过使用来自以下组中所选的一个的权重的加权模型计算,包括微阵列数据的线性模型(limma)模型、逻辑回归模型、线性判别分析(LDA)模型、条件逻辑回归模型、套索回归模型、岭回归模型、随机森林、支持向量机和概率回归模型。此外,还可以选择使用来自limma模型的权重来计算诊断指数的加权模型。

[0034] 本发明中,术语"未加权模型"和"加权模型"在这里应该被理解为熟悉本领域的专业人士广泛理解的常见定义范围。关于术语"未加权模型",它指的是在计算诊断指数时没有对miRNA生物标志物组中的每个miRNA设置对应的权重的情况。在本发明范围内,参考公式(I),"通过未加权模型计算诊断指数"可以理解为任何miRNA在miRNA生物标志物组中都具有相等的t_i(例如,$t_i=1$)。关于术语"加权模型",它指的是在计算诊断指数时为miRNA生物标志物组中的每个miRNA应用相应的权重的情况。在本发明范围内,参考公式(I),"通过加权模型计算诊断指数"可以理解为对于miRNA生物标志物组中的任何$miRNA_i$,不是所有的t_i都相等(即至少有两个miRNA具有不同的权重)。

[0035] "微阵列数据的线性模型(limma)模型"(Ritchie et al.2015),"逻辑回归模型"(Venable and Ripley 2002),"线性判别分析(LDA)模型"(Venable and Ripley,2002),"条件逻辑回归模型"(Venable and Ripley 2002),"套索回归模型"(Tibshirani,1996),"岭回归模型"(Hoerl and Kennard,1970),"随机森林"(Ripley,1996),"支持向量机"(Ripley,1996)和"概率回归模型"(Venable and Ripley,2002)中的每个术语实质上都是一种由该领域熟练人士通常理解的概率建模统计模型,其详细信息可以参考随后的参考文献。

[0036] 方便起见,根据某些实施例,在第(2)步之后和第(3)步之前,该方法可以进一步包括一个归一化步骤,即:获取基于计算得到的诊断指数的归一化诊断指数。相应地,第(3)步包括:如果归一化诊断指数等于或大于预设的阈值,则将受试者分类为患有癌症;否则,将受试者分类为不患有癌症。

[0037] 在这里,进行归一化步骤可以采用不同的方式。根据某些实施例,归一化诊断指数是基于公式(II)计算的:

[0038] $$归一化诊断指数=\frac{诊断指数-par\quad location}{param_{scale}};\quad (II)$$

[0039] 在这里，$param_{location}$和$param_{scale}$分别是配置的位置参数和尺度参数，以使归一化诊断指数在不小于第一个预设值且不大于第二个预设值的范围内。

[0040] 更具体地说，$param_{location}$实质上是一个位置参数，配置为将归一化诊断指数的最小值移至第一个预设值，而$param_{scale}$实质上是一个尺度参数，配置为将归一化诊断指数的最大值缩放到第二个值。因此，第一个预设值和第二个预设值分别是已从已知患有癌症和已知不患有癌症的现有人群中获得和计算的归一化诊断指数值范围内的最小值和最大值，同时排除异常值。

[0041] 可选择应用多种设置。例如，在下面的例1中，现有数据集中的诊断指数值被确定为在600到1600之间的，剔除异常值，为了将范围移至在0(即第一个预设值)和10(即第二个预设值)之间，$param_{location}$和$param_{scale}$可以分别设置为600和100，以便最终的归一化诊断指数不小于0且不大于10。需要注意的是，在下面的例1中采用了这种归一化方案。

[0042] 另一种方法是，$param_{location}$和$param_{scale}$可以分别设置为600和1000，以便将最终的归一化诊断指数设置为不小于0且不大于1。或者，$param_{location}$和$param_{scale}$可以分别设置为600和10，以便将最终的归一化诊断指数设置为不小于0且不大于100。也可以换一种方法，$param_{location}$和$param_{scale}$可以分别设置为350和250，以便将最终的归一化诊断指数设置为不小于1且不大于5。

[0043] 在将归一化诊断指数归一化为在0和10之间的实施例中，预设的阈值可以选择设置为5.1，从而允许该方法具有大约0.95的特异性，或者可以选择设置为6.0，从而允许该方法具有大约0.99的特异性。

[0044] 在上述任何方法的任何实施例中，所述生物样本是从血液样本、血清样本、血浆样本、尿液样本(Yun等，2012)、唾液样本(Park等，2009)和痰液样本中选择的液体活检样本。

[0045] 在上述所述方法的任何实施例中，在步骤(1)中，可以选择通过Northern Blotting、微阵列分析、RNA测序或RNA原位杂交等手段获取miRNA生物标记物集合的表达谱，或者可以通过核酸扩增程序获取，包括逆转录PCR(RT-PCR)、定量RT-PCR(qRT-PCR)或数字RT-PCR。

[0046] 在本发明中，上述每种miRNA检测方法都应理解为在该领域的普通技术人员中普遍理解的常见定义。下面将提供有关实施这些方法以确定miRNA生物标记物集合的表达谱的更多详细信息。

[0047] 在上述所述方法的任何实施例中，本方法可以进一步包括对受试者评估癌症的诊断或癌症的复发。

[0048] 在此，“癌症的诊断”指的是在先前已知没有癌症的受试者中检测到癌症，而“癌症的复发”指的是在已经接受治疗去除癌症后的患者中再次检测到癌症。

[0049] 在上述所述方法的任何实施例中，当将受试者分类为患有癌症时，该方法可以进一步用于管理受试者的治疗方案。这些治疗方案可以涵盖多种已知的治疗方案，包括手术、放射治疗、化疗、激素疗法、靶向疗法、免疫疗法或其组合。这些治疗方案对本发明提到的不同癌症已经有了明确的治疗效果。

[0050] 在上述所述方法的任何实施例中，当将受试者分类为患有癌症时，该方法可以进一步包括对受试者进行诊断程序的步骤。在这里，诊断程序可以包括体格检查、从受试者中取的活检的病理检查、免疫组织化学检查，或成像检查，例如X线、计算机断层扫描(CT)、超

声检查和/或磁共振成像等。

[0051] 第二方面,本发明还公开了一种用于从患者生物样本中检测癌症的试剂盒,该试剂盒主要用于实施本发明第一方面描述的方法。

[0052] 如本申请所述,以及在本发明的其他地方,“试剂盒”是指一组物品和/或指导说明。试剂盒中的物品可以是物理实体或其组件。如本发明所述,试剂盒中可以包括的物品实施例包括一个或多个核酸(例如多聚核苷酸),或一个或多个设备、装置或设备(例如包括一个或多个核酸的分子阵列或微阵列)。试剂盒中包含的说明可以是要执行的特定步骤的描述(例如手册),可以打印在物理介质上(例如纸张、卡片等),也可以存储在计算机可读的存储介质上(例如硬盘、光盘或CD、闪存驱动器等),甚至存储在互联网上(例如在可访问的云空间中等)。

[0053] 试剂盒可以包括以下至少两个组成部分(1)和(2)(即物品和/或说明):

[0054] 组成部分(1)(以下简称组分(1)):至少一种能够特异性识别miRNA生物标记集中的每个miRNA的核酸,从而允许从生物样本中获得miRNA生物标记集的表达谱。在这里,miRNA生物标记集包括hsa-miR-5100(SEQ ID NO:1)。

[0055] 组成部分(2)(以下简称组分(2)):至少一个说明,包括第一说明和第二说明。第一说明包括用于根据miRNA生物标记集的表达谱计算生物样本的诊断指数的第一子说明,其中诊断指数是根据以下公式计算的:

[0056] $$诊断指数=\sum_{i=1}^{n} t_i * miRNA_i; \quad (I)$$

[0057] 其中,n是miRNA生物标志物组中miRNA的总数,$miRNA_i$是miRNA生物标志物组中第i个miRNA的表达水平,其中i是大于零且小于或等于n的整数;t_i是第i个miRNA的权重。第二说明用于将被测试者分类为是否患有癌症,如果计算得到的诊断指数大于或等于预定的阈值,那么被测试者被分类为患有癌症,否则被分类为没有癌症。

[0058] 在该试剂盒的组分(1)中,至少一个核酸可以能够严格地和下面两个多核苷酸之一进行特异杂交:(a)包含SEQ ID NO:1序列的多核苷酸,以及其衍生物或者变体,也就是说具有至少80%相同的序列,或者包含15个或更多个相同的连续核苷酸的片段;(b)包含和SEQ ID NO:1序列互补的多核苷酸,以及其衍生物或者变体,也就是说具有至少80%相同的序列,或者包含15个或更多个相同的连续核苷酸的片段。

[0059] 在试剂盒的一些实施方式中,miRNA生物标记集还包括除hsa-miR-5100之外的另外99种miRNA。相应地,在试剂盒的组分(1)中,至少一个核酸可以能够严格地和下面两个多核苷酸之一进行特异杂交:(a)包含SEQ ID NO:2-100序列的多核苷酸,以及其衍生物或者变体,也就是说具有至少80%相同的序列,或者包含15个或更多个相同的连续核苷酸的片段;(b)包含和SEQ ID NO:2-100序列互补的多核苷酸,以及其衍生物或者变体,也就是说具有至少80%相同的序列,或者包含15个或更多个相同的连续核苷酸的片段。

[0060] 在试剂盒的一些实施例中,miRNA生物标记集还包括除hsa-miR-5100之外的其他前50种miRNA。相应地,在试剂盒的组分(1)中,至少一个核酸可以能够严格地和下面两个多核苷酸之一进行特异杂交:(a)包含SEQ ID NO:2-50序列的多核苷酸,以及其衍生物或者变体,也就是说具有至少80%相同的序列,或者包含15个或更多个相同的连续核苷酸的片段;(b)包含和SEQ ID NO:2-50序列互补的多核苷酸,以及其衍生物或者变体,也就是说具有至少80%相同的序列,或者包含15个或更多个相同的连续核苷酸的片段。

[0061] 在试剂盒的一些实施例中,miRNA生物标记集还包括除hsa-miR-5100之外的其他前20个miRNA。相应地,在试剂盒的组分(1)中,至少一个核酸可以能够严格地和下面两个多核苷酸之一进行特异杂交:(a)包含SEQ ID NO:2-20序列的多核苷酸,以及其衍生物或者变体,也就是说具有至少80%相同的序列,或者包含15个或更多个相同的连续核苷酸的片段;(b)包含和SEQ ID NO:2-20序列互补的多核苷酸,以及其衍生物或者变体,也就是说具有至少80%相同的序列,或者包含15个或更多个相同的连续核苷酸的片段。

[0062] 更进一步的,miRNA生物标记集由表1中的前20个miRNA组成,相应地,在试剂盒的组分(1)中,至少一个核酸包含总共20个多核苷酸,分别能够严格地和下面两个多核苷酸之一进行特异杂交:(a)包含SEQ ID NO:1-20序列的多核苷酸,以及其衍生物或者变体,也就是说具有至少80%相同的序列,或者包含15个或更多个相同的连续核苷酸的片段;(b)包含和SEQ ID NO:1-20序列互补的多核苷酸,以及其衍生物或者变体,也就是说具有至少80%相同的序列,或者包含15个或更多个相同的连续核苷酸的片段。

[0063] 在试剂盒的一些实施例中,miRNA生物标记集包括表1中一个或者多个除了hsa-miR-5100以外排名前4的miRNA。因此,在试剂盒的组分(1)中,至少一个核酸可以能够严格地和下面两个多核苷酸之一进行特异杂交:(a)包含SEQ ID NO:2-4序列的多核苷酸,以及其衍生物或者变体,也就是说具有至少80%相同的序列,或者包含15个或更多个相同的连续核苷酸的片段;(b)包含和SEQ ID NO:2-4序列互补的多核苷酸,以及其衍生物或者变体,也就是说具有至少80%相同的序列,或者包含15个或更多个相同的连续核苷酸的片段。

[0064] 更进一步的,miRNA生物标记集由表1中的前4个miRNA组成,即hsa-miR-5100、hsa-miR-1343-3p、hsa-miR-1290和hsa-miR-4787-3p。因此,在试剂盒的组分(1)中,至少一个核酸由总共4个多核苷酸组成,分别能够严格地和下面两个多核苷酸之一进行特异杂交:(a)包含SEQ ID NO:1-4序列的多核苷酸,以及其衍生物或者变体,也就是说具有至少80%相同的序列,或者包含15个或更多个相同的连续核苷酸的片段;(b)包含和SEQ ID NO:1-4序列互补的多核苷酸,以及其衍生物或者变体,也就是说具有至少80%相同的序列,或者包含15个或更多个相同的连续核苷酸的片段。

[0065] 在试剂盒中,根据组分(2)中的第一指令的第一子指令,可以通过无权重模型来计算诊断指数,或者可以通过使用上述第一方案中提供的概率建模统计模型之一的权重的加权模型来计算诊断指数。根据试剂盒的一些实施例,诊断指数是通过使用limma模型中的权重来进行加权模型计算的。

[0066] 根据试剂盒的一些实施例,预定阈值可以设置为1110,并且第二指令进一步指令使用1110作为预定阈值的分类具有特异性大于0.95。根据试剂盒的一些其他实施例,预定阈值可以设置为1200,第二指令进一步指令,使用1200作为预定阈值的分类具有特异性大于0.99。

[0067] 根据试剂盒的一些实施例,第一指令进一步包括第二个子指令,用于根据第一指令计算的诊断指数获取归一化的诊断指数,而在第二指令中,如果归一化的诊断指数大于或等于预设的分界值,则将受试者分类为有癌,否则将其分类为非癌。归一化过程与上述第一方法方面中提到的归一化过程基本相同,因此在此跳过描述。

[0068] 根据一些试剂盒的实施例,归一化的诊断指数可以通过使用来自limma模型的权重进行加权模型计算,第一个预设值为0,第二个预设值为10。此外,预设的分界值可以选择

设置为5.1或6.0,从而使使用预设的分界值进行分类的特异性分别为>0.95或>0.99。

[0069] 根据不同的实施例,试剂盒中的部分(2)中的至少一个说明可以进一步包括第三个说明,用于对受试者进行评估,其中该评估包括对癌症的诊断或对癌症复发的检测;或者可以进一步包括第四个说明,用于在将受试者分类为癌症患者时向其施行治疗方案。

[0070] 在一些实施例中,试剂盒中组分(2)中的至少一个说明还可以包括一个额外的第一个说明,这个说明是关于如何获取miRNA生物标志物组的表达谱,可以包括Northern Blotting、微阵列分析、RNA测序或RNA原位杂交等过程。在这里,至少一个核酸包含在分子阵列上。

[0071] 在一些实施例中,试剂盒可以进一步包括至少一套扩增引物,每套扩增引物能够从生物样本中特异性地扩增miRNA生物标志物组中的至少一个miRNA。因此,试剂盒中组分(2)的至少一个说明进一步包括一个额外的第二说明,用于通过至少一个核酸和至少一套扩增引物执行逆转录PCR(RT-PCR)、定量RT-PCR(qRT-PCR)或数字RT-PCR等过程来获取miRNA生物标志物组的表达谱。

[0072] 在上述所描述的试剂盒的任何实施例中,生物样本可以是从以下一组中选取的液体活检样本,该组包括血样、血清样、血浆样、尿样、唾液样和痰样。

[0073] 第三方面,本发明提供了一个系统用于检测癌症。在这里,该系统主要包括一个计算机化系统,包括一系列硬件(例如处理器、内存、I/O接口、存储介质等)和软件(即计算机程序,包括操作系统软件和特定程序软件等),这些硬件和软件被配置为协同工作,以共同实施上述第一方面中所描述的方法的所有或一些步骤。根据某些实施例,该系统包括处理器和非瞬态存储介质。非瞬态存储介质被配置为包含用于由处理器执行的软件(即程序),并且程序被配置为指导处理器执行上述第一方面中描述的各种不同方法,以及各种不同实施例中的各种步骤。

[0074] 第四方面,本发明还提供了一种非瞬态存储介质,配置为存储可由处理器执行的计算机可执行程序,当处理器运行时,上述第一方面中描述的各种不同实施例的方法会被执行。

[0075] 根据上面提到的各种系统和非瞬态存储介质可以有各种不同的配置方案,包括miRNA生物标记物集中包含哪些miRNA成分,是否以及如何对诊断指数进行归一化,如何将受试者分类为癌或非癌,可以使用哪种样本作为生物样本,以及需要实现何种检测准确度水平等。关于这些不同方案的具体细节,可以参考第一方面中描述的方法的各种实施例,为简洁起见,在此不再详细叙述。

[0076] 除非在其他地方定义过的,本发明中使用的术语定义如下:

[0077] 一般而言,"受试者"指的是哺乳动物,例如灵长类动物,包括人类和黑猩猩,宠物动物,包括狗和猫,家畜动物,包括牛、马、绵羊和山羊,以及啮齿动物,包括老鼠和大鼠。术语"健康受试者"也指代没有检出待检测癌症的哺乳动物。值得注意的是,整个公开说明更具体地涉及人类受试者,但也可以选择性地应用于其他非人哺乳动物。

[0078] 除非另有指令或定义,如“核酸”、“核苷酸”、“多核苷酸”、“DNA”、“RNA”和“miRNA”等术语或缩写遵循通常用法。

[0079] 如本发明所使用,“多核苷酸”与“核酸”是可以互换使用的,指的是包括RNA、DNA和RNA/DNA(嵌合体)在内的核酸。DNA包括cDNA、基因组DNA和合成DNA。RNA包括总RNA、mRNA、

rRNA、miRNA、siRNA、snoRNA、snRNA、非编码RNA和合成RNA。

[0080] 根据本发明的用法,“片段”是指具有多核苷酸的核苷酸序列,通常具有15个或更多核苷酸,例如15、16、17、18、19个核苷酸等。

[0081] 根据本发明的用法,“基因”包括RNA、双链DNA以及构成双链的单链DNA,包括正链(或正义链)和互补链(或反义链)。基因的长度没有特定限制。在本发明中,“基因”包括双链DNA,包括人类基因组DNA,包括cDNA的单链DNA,包括与正链互补的序列的单链DNA(互补链),miRNA(miRNA)以及它们的片段和转录产物,除非另有规定。"基因"不仅包括由特定核苷酸序列(或SEQ ID NO)表示的"基因",还包括编码具有与基因编码的RNA等效的生物学功能的RNA的"核酸",例如同源物(同源基因或同源基因),变体(例如,遗传多态性)和衍生物。这样的"核酸"的具体实施例可以包括在下述严格条件下杂交的具有与任何SEQ ID NOs:1到100之一所表示的核苷酸序列或该核苷酸序列的变体的互补序列的"核酸",或者是通过用核苷酸"U"(或"u")替换核苷酸"T"(或"t")来自该核苷酸序列的核苷酸序列。"基因"的功能区域没有特定限制,可以包括表达控制区域、编码区域、外显子或内含子。"基因"可以存在于细胞中,也可以在释放到细胞外后独立存在。或者,"基因"可以以包含在称为外泌体的囊泡内的状态存在。

[0082] 根据本发明的用法,"microRNA(miRNA)"是指一个15到25个核苷酸的非编码RNA,作为具有类似发夹结构的RNA前体转录,经由具有RNase III切割活性的双链RNA切割酶裁剪,整合到一个称为RISC的蛋白质复合物中,并参与抑制mRNA的翻译,除非另有规定。在本发明中,"miRNA"这个术语不仅包括由特定核苷酸序列(或SEQ ID NO)表示的"miRNA",还包括"miRNA"的前体(pre-miRNA或pri-miRNA)以及具有等效生物功能的miRNA,例如同源物(同源基因或同源物)、变体(例如,遗传多态性)和衍生物。这种前体、同源物、变体或衍生物可以通过使用miRBase Release 20(Kozomara和Griffiths-Jones,2010)进行特异性识别,其中的实施例可以包括根据后面描述的严格条件下杂交到由SEQ ID NOs:1到100之一所表示的任何特定核苷酸序列的互补序列的"miRNA"。根据本发明的用法,"miRNA"可能是一个miRNA基因的基因产物。这种基因产物包括成熟的miRNA(例如,如上所述参与抑制mRNA翻译的15到25个核苷酸或19到25个核苷酸的非编码RNA)或miRNA前体(例如,pre-miRNA或pri-miRNA)。

[0083] 在本发明中,“探针”包括用于特异检测基因表达所产生的RNA或源自RNA的多核苷酸,和/或与其互补的多核苷酸。

[0084] 在本发明中,“引物”或“扩增引物”包括用于特异识别和扩增基因表达所产生的RNA或源自RNA的多核苷酸,和/或与其互补的多核苷酸。

[0085] 在这种情况下,互补多核苷酸(互补链或反向链)指的是一种基于A:T(U)和G:C碱基对的完全序列与由任何SEQ ID NO:1到100中的任何核苷酸序列或源自该核苷酸序列的核苷酸序列相互关联的多核苷酸,或者它的部分序列(这里,出于方便起见,该完整或部分序列称为正链)。然而,这样的互补链不限于与目标正链的核苷酸序列完全互补,而可以具有在严格条件下与目标正链发生杂交的程度的互补关系。

[0086] 在本发明中,“严格条件”指的是核酸探针与其靶标序列杂交的条件,使其杂交程度较其他序列更大(例如,测量值等于或大于背景测量值的平均值+背景测量值的标准差×2)。严格条件取决于序列,并且根据进行杂交的环境而有所不同。通过控制杂交和/或洗涤

条件的严格性,可以识别与核酸探针100％互补的靶标序列。下面将提到“严格条件”的具体实施例。

[0087] 在本发明中,“变体”是指核酸的情况下,其是与多态性、突变等相关的自然变体;包含删除、替代、添加或插入的变体,其包含任何一个SEQ ID NO:1至100或源自该核苷酸序列的核苷酸序列的替代和插入的1个、2个或3个或更多核苷酸的核苷酸序列,或者其部分序列;包含在一种与任何一个SEQ ID NO:1至100或源自该核苷酸序列的替代和插入的1个或2个或更多核苷酸的核苷酸序列,或其部分序列;其对这些核苷酸序列或其部分序列的％同源性约为90％或更高,约为95％或更高,约为97％或更高,约为98％或更高,约为99％或更高;或者在上面定义的严格条件下与包括这些核苷酸序列或其部分序列的多聚核苷酸或寡聚核苷酸杂交的核酸。变体可以通过使用众所周知的技术,如定点突变或基于PCR的突变法来制备。

[0088] “相同序列百分比(％)”可以在引入或不引入序列空隙的情况下,使用上述基于BLAST或FASTA的蛋白质或基因搜索系统来确定(Zhang et al.,2000;Altschul et al.1990;Pearson et al.1988)。

[0089] ″衍生物″指的是包括经修改的核酸,例如,带有荧光团标记的衍生物,或者包含经修改的核苷酸的衍生物(例如,核苷酸含有卤素、烷基如甲基、烷氧基如甲氧基、硫代基或羧甲基等官能团,以及已经经历碱基重排、双键饱和、脱氨基、氧分子替换为硫原子等改变的核苷酸)。此外,还包括PNA(肽核酸;Nielsen et al.1991)和LNA(锁定核酸;Obika et al.1998)等,没有限制。

[0090] 能够与上述miRNA中选择的多聚核苷酸特异结合的“核酸”是合成或制备的核酸,特别包括“核酸探针”或“引物”。这种“核酸”可直接或间接用于检测受试者体内癌症的存在或缺失,用于诊断癌症的严重程度、改善程度或治疗灵敏度,或用于筛选用于预防、改善或治疗癌症的候选物质。这种“核酸”包括核苷酸、寡核苷酸和能够特异识别和结合到由任何SEQ ID NOs:1到100中的任何一个所代表的转录本,或其合成的cDNA核酸。特别是,在与癌症发展相关的样本(例如血液或尿液等体液)中,这种核苷酸、寡核苷酸和多聚核苷酸可以作为探针,基于上述性质有效用于检测体内、组织、细胞等中表达的上述基因,或作为扩增体内表达的上述基因的引物。

[0091] 在这里,“检测”可以与“考察”、“测量”或“检测或决策支持”等互换使用。在这里,“评估”是指基于检查结果或测量结果进行的诊断或评估支持。

[0092] 在本发明中,“P值”、“准确性”、“AUC”、“灵敏度”和“特异性”应理解为专业人士广泛认可的常见定义,并具体定义如下:

[0093] “P值”或“P”可互换地写作“p-value”或“p”,指的是在零假设下,观察到比实际从数据中计算的统计量更极端的概率。因此,较小的“P”或“P值”意味着要比较的对象之间具有更显著的差异。.

[0094] “AUC”表示接收者操作特征曲线下的面积。“准确性”表示(真正例数+真负例数)/(总病例数)的值。准确性指令正确识别的样本占所有样本的比率,并用作评估检测性能的主要指标。

[0095] 在此,“灵敏度”表示(真正例数)/(真正例数+假负例数)的值。高灵敏度可以检测到癌症,从而进行临床治疗干预。

[0096] 在此,“特异性”表示(真负例数)/(真负例数+假正例数)的值。高特异性可以防止将被错误判断为癌症患者的健康受试者进行不必要的额外检查,减轻了患者的负担并降低了医疗费用。

[0097] 除非另有定义,以下是可用于确定miRNA生物标志物组的表达谱的技术摘要:

[0098] 需要注意的是,确定miRNA生物标志物组的表达谱实质上包括确定miRNA生物标志物组中每个miRNA的表达水平。最好,在一个经过良好控制的单个实验中可以同时确定miRNA生物标志物组中所有miRNA的表达水平。然而,也可以选择通过不止一个实验和不同的实验程序来确定这些miRNA的表达水平。

[0099] 在此使用的测量或检测miRNA生物标志物组中任何miRNA的表达,包括测量或检测与该miRNA相对应的任何核酸转录物。

[0100] 通常,可以基于miRNA或相应的反转录cDNA水平来检测或测量表达。可以使用任何测量RNA水平或cDNA水平的定量或定性方法。用于检测或测量miRNA或cDNA水平的合适方法包括Northern Blotting、微阵列分析、RNA测序、RNA原位杂交或核酸扩增程序,例如反转录聚合酶链反应(RT-PCR)或实时RT-PCR,也被称为定量RT-PCR(qRT-PCR)或数字RT-PCR。这些方法在技术领域中是众所周知的(参见例如,Green和Sambrook等人2012)。其他技术包括基因表达的数字、多重分析,如nCounter®(NanoString Technologies,Seattle,WA)基因表达分析,这些在US20100112710和US20100047924中有进一步描述。

[0101] 检测感兴趣的核酸通常涉及目标(例如miRNA或cDNA)与探针之间的杂交。已知在各种癌症基因表达谱中使用的miRNA的序列。因此,技术领域中的专业人士可以很容易地设计用于检测这些miRNA的杂交探针(参见例如,Green and Sambrook et al.2012)。例如,可以使用miRNA或cDNA靶标的核酸序列自身通过常规技术(例如PCR或合成)创建特异性结合到本发明描述的miRNA转录本(或其合成的cDNA)的多核苷酸探针。在本发明中,“探针”表示包括大约10个或更多相邻核苷酸,大约15个或更多相邻核苷酸,大约20个或更多相邻核苷酸的多核苷酸序列的一部分或部分。在某些实施例中,多核苷酸探针将包括10个或更多核酸、15个或更多核酸或20个或更多核酸。为了提供足够的特异性,探针可以与目标序列的互补物具有约90%或更高的序列同一性,例如约95%或更高(例如约98%或更高或约99%或更高),如使用著名的BLAST算法(可通过国家生物技术信息中心(NCBI)提供,位于马里兰州贝塞斯达)进行确定。

[0102] 每个探针可能在很大程度上特异于其靶标,以避免任何交叉杂交和假阳性。使用特异性探针的替代方法是在从转录本中提取材料时使用特异性试剂(例如,在cDNA制备期间,或在扩增期间使用特异性引物)。在这两种情况下,可以通过与正在分析的miRNA组内在很大程度上独特的目标部分的杂交来实现特异性,例如,与多个miRNA的聚A尾的杂交不会提供特异性。如果一个靶标具有多个剪接变体,可以设计一个能够识别每个变体共有区域的杂交试剂,或者使用多个试剂,每个试剂可以识别一个或多个变体。

[0103] 杂交反应的严格程度通常由在这门艺术领域具有普通技能的人员轻松确定,通常是一种依赖于探针长度、洗涤温度和盐浓度的经验计算。一般来说,较长的探针可能需要更高的温度以进行适当的退火,而较短的探针可能需要较低的温度。杂交通常取决于变性的核酸序列在其熔化温度以下的环境中,当存在互补链时,它们的能力重新结合。所需同源性的程度越高,可以使用的相对温度就越高。因此,更高的相对温度往往会使反应条件更加严

格,而更低的温度则相对较不严格。

[0104] 在这里定义的“严格条件”或“高严格条件”是指但不限于以下条件:(1)在洗涤时使用低离子强度和高温,例如0.015M氯化钠/0.0015M柠檬酸钠/0.1%十二烷基硫酸钠,温度为50℃;(2)在杂交过程中使用变性剂,例如50%甲酰胺,加入0.1%牛血清白蛋白/0.1%菲考尔/0.1%聚乙烯吡咯烷酮/50mM磷酸钠缓冲液(pH 6.5),含750mM氯化钠、75mM柠檬酸钠,温度为42℃;或者(3)使用50%甲酰胺、5×SSC(0.75M氯化钠、0.075M柠檬酸钠)、50mM磷酸钠缓冲液(pH 6.8)、0.1%焦磷酸钠、5×Denhardt's溶液、超声处理的鲑鱼精子DNA(50μg/ml)、0.1%十二烷基硫酸钠和10%右旋糖等,温度为42℃,随后在55℃的0.2×SSC(氯化钠/柠檬酸钠)和50%甲酰胺中进行洗涤,然后在55℃含有EDTA的0.1×SSC中进行高温洗涤。“适度严格条件”如Sambrook等人在1989年所描述,包括使用较不严格的洗涤液和杂交条件(例如温度、离子强度和SDS的百分比)。适度严格条件的一个例子是在含有以下成分的溶液中,以37℃进行过夜孵育:20%甲酰胺、5×SSC(150mM氯化钠、15mM柠檬酸三钠)、50mM磷酸钠(pH 7.6)、5×Denhardt's溶液、10%右旋糖、和20mg/mL变性的切碎的鲑鱼精子DNA,随后在约37-50℃的1×SSC中洗涤滤纸。熟练的技术人员将会了解如何根据需要调整温度、离子强度等因素,以适应探针长度等因素。

[0105] 在某些实施例中,可以使用微阵列分析、Northern印迹、RNA原位杂交或基于PCR的方法。在这方面,测定上述miRNA在生物样品中的表达可以包括,例如,将含有或被怀疑含有癌细胞的样品与特异于感兴趣miRNA的多聚核苷酸探针或设计用于扩增感兴趣miRNA一部分的引物接触,然后检测探针与核酸靶物的结合或核酸的扩增。设计PCR引物的详细方案在技术领域中是已知的(见例如Green和Sambrook等人在2012年的文献)。在某些实施例中,从样本中获得的miRNA可能会经受qRT-PCR。逆转录可以通过技术领域已知的任何方法进行,例如通过使用Omniscript RT试剂盒(Qiagen)。然后可以使用技术领域已知的任何扩增技术扩增所得的cDNA。然后,可以通过使用如下所述的对照样品来分析miRNA的表达。如本发明所述,与对照组相对的miRNA的过度表达或欠表达可以测量以确定个体生物样品的miRNA表达谱。类似地,技术领域已知并在本发明中描述了准备和使用微阵列分析miRNA表达的详细方案。

[0106] 如本发明所述,RNA测序(RNA-seq),也称为全转录组Shotgun测序,指的是用于实时检测RNA转录本存在和数量的各种高通量测序技术之一。参见Wang,Z.,M.Gerstein,and M.Snyder,RNA-Seq:a revolutionary tool for transcriptomics,NAT REV GENET,2009.10(1):p.57-63.RNA测序可用于显示样品miRNA的某一时刻来自基因组的快照。在某些实施例中,在测序之前,miRNA通过逆转录转化为cDNA片段,而在某些实施例中,miRNA可以直接测序而无需转化为cDNA。连接器可以连接到miRNA的5'端和/或3'端,miRNA或cDNA可以选择性地扩增,例如通过PCR。然后使用高通量测序技术对片段进行测序,例如Roche(例如,454平台)、Illumina公司和Applied Biosystem(例如,SOLiD系统)提供的技术。

[0107] 附图简要描述

[0108] 图1A-1C显示了肺癌数据集的案例流程图(图1A,分为发现和验证组),以及卵巢、肝脏和膀胱癌数据集的案例流程图(图1B,在删除冗余样本后合并为单一验证数据集),并总结了肺癌、膀胱癌、卵巢癌和肝癌的患者和肿瘤特征以及相应对照组的人口统计信息(图1C);

[0109] 图2A-2G显示了肺癌数据集中4-miRNA诊断模型的开发和验证,图2A显示通过在发现组中进行10倍交叉验证确定诊断模型的最佳miRNA数量(虚线);图2B显示在发现组中的ROC分析;图2C显示在发现组中的归一化诊断指数分布;图2D显示在验证组中的ROC分析;图2E显示在验证组中的归一化诊断指数分布;图2F显示了180名肺癌患者配对血清样本(术前vs术后)的归一化诊断指数的比较;图2G显示了验证组中的临床子集中归一化诊断指数的分布。虚线水平线代表了我们模型的归一化诊断指数的阈值。图中显示的百分比是每个癌症亚组的灵敏度。

[0110] 图3A和3B显示了4-miRNA诊断模型在其他癌症数据集中的性能,图3A显示ROC分析,图3B显示了4-miRNA模型的归一化诊断指数分布。图中显示的百分比是每种癌症类型的灵敏度和非癌症对照的特异性;

[0111] 图4A和4B显示了在肺癌数据集中的年龄和性别组中的ROC分析和归一化诊断指数分布。

[0112] 发明详细描述

[0113] 本发明提供了一种方法、试剂盒和计算机化系统,能够基于从受试者获得的生物样本中确定的至少一个miRNA的表达谱,准确可靠地检测一个或多个人类癌症。

[0114] 在本节的第一个方面,提供了一种具有AUC值大于大约0.780的诊断准确性的检测方法,主要包括以下三个步骤:

[0115] 步骤(1):确定miRNA生物标志物组的表达谱;

[0116] 步骤(2):根据miRNA生物标志物组的表达谱计算生物样本的诊断指数。诊断指数的计算基于:

[0117] $$诊断指数 = \sum_{i=1}^{n} t_i * miRNA_i; \quad (I)$$

[0118] 其中,n是miRNA生物标志物组中miRNA的总数,$miRNA_i$是miRNA生物标志物组中第i个miRNA的表达水平,其中i是大于零且小于或等于n的整数;t_i是第i个miRNA的权重;和

[0119] 步骤(3):基于计算得到的诊断指数的值,将受试者分类为有癌或非癌。如果计算得到的诊断指数大于或等于预定的阈值,那么将受试者分类为有癌;否则,将受试者分类为非癌。

[0120] 在此,miRNA生物标志物组包括hsa-miR-5100,并且可以选择性地进一步包括表1中列出的任意一个或多个miRNA的组合(见实施例1)。根据不同的实施方式,除了hsa-miR-5100之外,miRNA生物标志物组还可以进一步包括来自表1中前2-100个miRNAs的miRNA,或者可以进一步包括来自表1中前2-50个miRNAs的miRNA(s),或者可以进一步包括来自表1中前2-20个miRNAs的miRNA(s),或者可以进一步包括来自表1中前2-4个miRNAs的miRNAs。

[0121] 优选的,miRNA生物标志物组由前4个miRNAs组成(即hsa-miR-5100、hsa-miR-1343-3p、hsa-miR-1290和hsa-miR-4787-3p)。在此,根据不同的实施方式,可以有不同的AUC截止水平(例如0.780、0.850、0.950、0.990和0.999),或不同的灵敏度-特异性水平(例如68%-99%、68%-99%、83%-99%和99%-99%),在至少在这些水平上,该方法能够准确检测特定的癌症类型。例如,该方法可以在AUC>0.999,并/或在灵敏度>99.0%和特异性>99.0%时准确检测肺癌和胃癌。

[0122] 基于公式(I)可以有不同的方式来计算诊断指数。计算可以基于未加权模型,或者加权模型。在后一种情况下,可以选择应用不同的模型(例如limma模型、logistic回归模型

等)来获取miRNA生物标志物组中miRNAs的权重。

[0123] 优选的,诊断指数是通过使用limma模型的权重进行计算的。在该方法的步骤(3)中,可以将预定的阈值设置为1110,从而允许该方法具有>0.95的特异性;或者可选择将预定的阈值设置为1200,以使该方法具有>0.99的特异性。

[0124] 另一种方案是,在第(2)步中计算的诊断指数可以进一步经历归一化处理,第(3)步可以基于归一化后的诊断指数是否小于或大于预设的截点来确定癌症分类。

[0125] 需要注意的是,归一化的选择是人工定义的。根据某些实施例,归一化过程可以基于以下公式:

[0126] $$归一化诊断指数=\frac{诊断指数-param_{location}}{param_{scale}};\ (II)$$

[0127] 其中,$param_{location}$和$param_{scale}$分别是位置参数和尺度参数,被配置为使得归一化诊断指数在不低于第一个预设值和不高于第二个预设值的范围内。

[0128] 在此,可以选择$param_{location}$和$param_{scale}$分别为600和1000,从而使得归一化诊断指数介于0和10之间,根据这种归一化,可以将预设阈值设置为5.1以获得特异性>0.95或设置为6.0以获得特异性>0.99。

[0129] 在该方法中,有利的生物样本是液体活检样本,例如血液样本、血清样本、血浆样本、尿液样本、唾液样本或痰液样本等。可以通过多种基于探针的方法来确定miRNA生物标志物组的表达谱,包括Northern Blotting、微阵列分析、RNA测序或RNA原位杂交,也可以通过多种依赖扩增的方法来实现,包括逆转录聚合酶链式反应(RT-PCR)、定量逆转录聚合酶链式反应(qRT-PCR)或数字RT-PCR。

[0130] 该方法还可以增加对受试者进行评估的步骤,以确定受试者是否被诊断出患有癌症(如果受试者以前没有癌症),或者受试者是否癌症复发(如果受试者以前接受治疗以去除或摆脱癌症)。为此,评估可以进一步包括体格检查、来自受试者的活检的病理检查、免疫组织化学检查或包括X线、计算机断层扫描(CT)、超声波、磁共振成像等的影像学检查。.

[0131] 当将受试者分类为患有癌症时,该方法还可以包括向受试者施行治疗方案的步骤,如手术、放疗、化疗、激素疗法、靶向疗法、免疫疗法或其组合。

[0132] 在本节的第二个方面中,根据第一个方面中所描述的各种不同实施例的方法,该发明会具体实施的各个步骤来提供试剂盒。

[0133] 该试剂盒主要包括一些物品(如,组成(1),包括一种或多种可以特异性识别miRNA生物标记集中的每个miRNA的核酸,可以有一种或多种扩增引物),可用于确定miRNA生物标记集的表达谱的核酸和引物,还包括一些指令(即组成(2)),用于计算诊断指数和癌症分类。

[0134] 根据miRNA生物标记集中包括的miRNA,部件(1)中的每个核酸可以包括一个多聚核苷酸,能够在苛刻条件下与以下情况之一发生特异杂交:(a)包含或由SEQ ID NOS中的核苷酸序列组成或由它们构成的核酸序列,其编号为1-100、1-50、1-20或1-4,其衍生物、至少具有80%序列同一性的其变体,或包含15个或更多连续核苷酸的其片段;或(b)包含或由SEQ ID NOS中的核苷酸序列相互补的核酸序列组成或由它们构成,其编号为1-100、1-50、1-20或1-4,其衍生物、至少具有80%序列同一性的其变体,或包含15个或更多连续核苷酸的其片段。

[0135] 该试剂盒有不同的实施方式,涉及以下元素/特征,包括:miRNA生物标记集中包括哪些miRNA成分;是否以及如何对诊断指数进行归一化;如何对患者是否患有癌症进行分类;可以用什么样本作为生物样本;以及需要达到什么样的检测准确度水平等。有关这些不同实施方式的具体细节可以参考上述所述方法的各种实施方式,这里不再详细介绍。

[0136] 在本节的第三个方面,还提供了一种计算机化解决方案,主要是以计算机化和自动化的方式实施上述第一方面中所描述的各种方法步骤。

[0137] 这种计算机化解决方案可以应用于需要通过在计算机中运行包含程序指令的软件程序来自动执行上述方法步骤(1)-(3)的情况,从而带来高效率和便利等优势。

[0138] 具体来说,这种计算机化解决方案可以包括一个计算机系统或计算机,该系统包括一个处理器(即控制器)和与处理器通信耦合的计算机可读的非瞬态存储介质。计算机可读的非瞬态存储介质被配置为存储可以由处理器执行的程序指令,从而使处理器执行上述方法中的各个不同步骤,包括:

[0139] 步骤(1):确定miRNA标志物集的表达谱;

[0140] 步骤(2):根据miRNA标志物集的表达谱并按照公式(I)计算生物样本的诊断指数;以及

[0141] 步骤(3):基于计算的诊断指数的值将受试者分类为是否有癌。

[0142] 在此,“处理器”可理解为“中央控制器”或“中央计算单元(CPU)”,并可视为单核或多核处理器,或者用于并行处理的多个处理器。在这里使用的“非临时”一词用于描述除传播的电磁信号之外的有形计算机可读存储介质,但并不打算以其他方式限制该短语所包含的物理计算机可读存储设备的类型。例如,可以包括任何有形或非临时存储介质或存储介质,如电子、磁性或光学介质(例如,磁盘或CD/DVD-ROM)或非瞬态存储器存储(例如,“闪存”存储器)等。

[0143] 如图5所示,系统100除了包括处理器10和计算机可读非临时存储介质20之外,还包括总线30、存储器40、I/O接口50和通信接口60。处理器10、存储介质20、存储器40、I/O接口50和通信接口60都通过总线30进行通信连接。

[0144] 存储介质20存储着计算机可执行的程序指令,当被处理器10执行时,引导处理器10执行上述所述方法的步骤(1)-(3)。存储器40被配置为短暂地存储从存储介质20获得的程序指令,处理器10被配置为短暂地执行存储在存储器40中的程序指令。I/O接口50允许系统100与用户之间进行输入/输出,实现对系统100的控制。通信接口60可以允许系统100与另一个计算设备建立通信连接,以交换数据。需要注意的是,这些计算机硬件组件可以在本地配置,也可以通过网络,如局域网、互联网或云,进行远程配置。

[0145] 下文提供了一个实施例来说明上述发明的各个方面。

实施例1

[0146] 在这个实施例中,利用四个标准化微阵列平台的大型miRNA微阵列数据集,开发和验证了基于血液中的miRNA表达的诊断标志。

[0147] 2.材料和方法

[0148] 2.1.研究设计

[0149] 本研究包括了四个微阵列数据集,共计7536名独立参与者,包括3604名癌症患者

和3932名非癌症对照者。这些数据集均来自于一个名为“体液中miRNA检测技术开发与诊断”的日本全国性研究项目,该项目旨在使用标准化的微阵列平台对13种癌症类型的50000多名参与者进行血清miRNA的特征化研究(Asakura et al.2020;Yokoi et al.2018;Usuba et al.2019,Yamamoto et al.2020)。这四个数据集最初是为了开发肺癌(GSE137140)、卵巢癌(GSE106817)、肝癌(GSE113740)和膀胱癌(GSE113486)的诊断标志而组建的。

[0150] 肺癌数据集是单一癌症类型中样本量最大的(n=1566),同时还有非癌症对照组(n=2178)。原始的肺癌研究建立了一个包含2个miRNA的诊断模型(在本研究中称为“原始2-miRNA模型”),用于检测肺癌,该模型在灵敏度和特异性方面表现出色(Asakura et al.2020)。本研究最初的目标是使用这一数据集开发和验证一个新的诊断模型,可能会在肺癌检测方面胜过原始的2-miRNA模型。随着发现了其他癌症类型的数据集,新模型的性能也被评估,以用于其他癌症的检测。

[0151] 2.2.参与者和血清样本

[0152] 血清样本的收集已在原始出版物中(Asakura et al.2020;Yokoi et al.2018;Usuba et al.2019,Yamamoto et al.2020)进行了详细描述。简而言之,从2008年至2016年之间被转诊或入院接受手术治疗的癌症患者收集血清样本,并在存放在-20℃之前在4℃下存放一周。那些在采集血清样本之前接受术前化疗和放射治疗的癌症患者被排除在外。没有癌症病史并且在前三个月内没有住院的非癌症对照组的血清样本是从三个来源的门诊部收集的:国立癌症中心医院(NCCH)、国立老年和老年病研究中心(NCGG)生物样本库和横滨美濑诊所(YMC)。从NCCH收集的血清样本的储存方式与癌症患者相同,而来自NCGG和YMC的血清样本则存放在-80℃直到使用。原始研究获得了NCCH机构审查委员会、NCGG伦理和利益冲突委员会以及医学公司信德会YMC的研究伦理委员会的批准。从每位参与者那里获取了书面知情同意。.

[0153] 2.3.miRNA微阵列表达分析

[0154] 有关微阵列分析的详细信息已在原始出版物(Asakura et al.2020;Yokoi et al.2018;Usuba et al.2019,Yamamoto et al.2020)中进行了描述。简而言之,从300μL血清中提取总RNA,使用3DGene®miRNA Labeling kit进行标记,然后杂交到3D-Gene® Human miRNA Oligo Chip(Toray Industries,日本神奈川)上,该芯片旨在研究miRBase release21中注册的2588个miRNA序列。排除了以下低质量样本:负对照探针的变异系数>0.15;由3D-Gene®Scanner识别为“不均匀斑点图像”的标志探针数量>10。当信号强度大于负对照信号的均值加上两倍标准偏差时,认为存在miRNA,并在使用负对照信号时,将信号强度按排名的最高和最低5%移除。背景减法是通过从miRNA信号中减去负对照信号的均值(在根据信号强度排名删除最高和最低5%后)来执行的。通过根据三个预选的内部对照miRNA(miR-149-3p、miR-2861和miR-4463)进行校准,可以实现跨微阵列的标准化。

[0155] 2.4.诊断模型的开发

[0156] 在肺癌数据集中,受试者被分为与原始出版物相同的发现和验证组(如图1A所示)(Asakura et al.2020),因为(1)原始作者选择了发现组,以在年龄、性别和吸烟史方面在癌症和非癌症之间保持平衡;(2)发现组中50%的非癌受试者来自NCCH,与癌症患者具有相同的血清储存条件,以最小化miRNA候选者选择中的潜在偏差;(3)使用相同的发现和验证组允许直接比较新的诊断模型与原始的2-miRNA模型的性能。由于诊断模型是从肺癌发现

组中开发的,因此在肺癌验证组中验证后,我们进一步测试了其作为多癌症诊断模型的能力,这些癌症类型未在模型开发中使用。

[0157] 研究人员使用了Microarray数据的线性模型(limma)来评估发现组中肺癌与非癌症之间的miRNA差异表达的统计显着性(Ritchie et al.2015)。在发现组中,基于受试者操作特征(ROC)曲线分析下曲线下面积(AUC),进行了十倍交叉验证,以确定最佳诊断模型的miRNA数量。诊断指数是通过使用limma统计加权的miRNA表达水平的线性总和来计算的。诊断指数的阈值被选择,以确保在发现组中不会对非癌症对照进行错误分类,以最小化假阳性,因为诊断模型可能被用作处于风险的普通大众的筛查测试。

[0158] 2.5.统计分析

[0159] 识别癌症与非癌症的诊断性能是通过ROC曲线分析的AUC、灵敏度和特异性来确定的。使用pROC软件包中的roc.test函数和自举法进行了两个ROC曲线的AUC比较。对于肺癌临床子组的配对灵敏度的比较,使用McNemar检验进行。limma分析使用Bioconductor软件包limma进行(Bioconductor开源生物信息学软件(于2020年8月27日访问))。所有统计分析均使用R版本4.0.5进行(R项目统计计算(于2020年7月15日访问))。

[0160] 3.结果

[0161] 3.1.参与者和数据集

[0162] 肺癌数据集包含了1566名肺癌患者和2178名非癌症对照(图1A)(Asakura et al.2020)。卵巢癌数据集包括333名卵巢癌患者和2759名非癌症对照,以及乳腺、结肠直肠、食管、胃、肝脏、肺、胰腺和肉瘤癌症患者(图1B)(Yokoi et al.2018)。肝癌和膀胱癌数据集分别包括345名肝癌患者/1033名非癌症对照和392名膀胱癌患者/100名非癌症对照,以及胆道、乳腺、结肠直肠、食管、胃、胶质瘤、肺、卵巢、胰腺、前列腺和肉瘤癌症患者(图1B)(Usuba et al.2019,Yamamoto et al.2020)。在保留肺癌数据集的同时,其他三个数据集中显示彼此之间或与肺癌数据集中的样本之间的相关性大于0.99的冗余样本被删除。然后,来自卵巢、肝脏和膀胱癌数据集的唯一样本被合并成一个总共3792个样本的非肺癌数据集,其中包括12种癌症类型的2038名癌症患者和1754名非癌症对照(图1B)。

[0163] 肺癌数据集被分为与原始研究相同的发现组(n=416)和验证组(n=3328)(图1A)。发现组包括208名肺癌患者和208名非癌症对照,按年龄、性别和吸烟状况匹配(Asakura et al.2020)。验证组包括1358名肺癌患者和1970名非癌症对照。肺癌患者中有57%为男性,62%为过去或现在吸烟者,78%为腺癌,14%为鳞状细胞癌,72%为I期,15%为II期,13%为III期(图1C)。

[0164] 这392名膀胱癌患者的平均年龄为68岁,男性占72%,5%有转移,12%淋巴结为阳性,77%病理分期T2或以下,80%为高级别(图1C)。这333名卵巢癌患者的平均年龄为57岁,25%为I期,10%为II期,55%为浆液性,19%为透明细胞型,13%为子宫内膜样组织学(图1C)。这348名肝癌患者的平均年龄为68岁,男性占78%,37%为I期,33%为II期(图1C)。原始研究没有提供其他癌症的详细人口统计学信息和肿瘤特征。

[0165] 表1.前100个在肺癌发现组中表达不同的miRNAs

[0166]

miRNA 名称	miRBase 编号	对数倍变化	*调整后的P值*	ROC 的 AUC	序列号(SEQ ID NO.)
hsa-miR-5100	MIMAT0022259	3.931	9.99E-176	0.9988	UUCAGAUCCCAGCGGUGCCUCU (SEQ ID NO. 1)
hsa-miR-1343-3p	MIMAT0019776	2.609	5.83E-94	0.9690	CUCCUGGGGCCCGCACUCUCGC (SEQ ID NO. 2)

[0167]

hsa-miR-1290	MIMAT0005880	6.538	2.22E-87	0.9979	UGGAUUUUUGGAUCAGGGA (SEQ ID NO. 3)
hsa-miR-4787-3p	MIMAT0019957	1.854	6.81E-67	0.9352	GAUGCGCCGCCCACUGCCCCGCGC (SEQ ID NO. 4)
hsa-miR-6877-5p	MIMAT0027654	1.364	1.37E-63	0.9490	AGGGCCGAAGGGUGGAAGCUGC (SEQ ID NO. 5)
hsa-miR-17-3p	MIMAT0000071	4.088	1.66E-62	0.9346	ACUGCAGUGAAGGCACUUGUAG (SEQ ID NO. 6)
hsa-miR-6765-5p	MIMAT0027430	-0.688	5.09E-62	0.9254	GUGAGGCGGGGCCAGGAGGGUGUGU (SEQ ID NO. 7)
hsa-miR-1268b	MIMAT0018925	-0.618	1.45E-61	0.9374	CGGGCGUGGUGGUGGGGGUG (SEQ ID NO. 8)
hsa-miR-4258	MIMAT0016879	1.777	6.45E-59	0.8983	CCCCGCCACCGCCUUGG (SEQ ID NO. 9)
hsa-miR-451a	MIMAT0001631	5.008	3.71E-58	0.9384	AAACCGUUACCAUUACUGAGUU (SEQ ID NO. 10)
hsa-miR-1228-5p	MIMAT0005582	-0.780	1.01E-57	0.9175	GUGGGCGGGGGCAGGUGUGUG (SEQ ID NO. 11)
hsa-miR-8073	MIMAT0031000	2.087	6.42E-55	0.9058	ACCUGGCAGCAGGGAGCGUCGU (SEQ ID NO. 12)
hsa-miR-4454	MIMAT0018976	1.661	3.91E-51	0.8982	GGAUCCGAGUCACGGCACCA (SEQ ID NO. 13)
hsa-miR-187-5p	MIMAT0004561	1.044	1.90E-50	0.9126	GGCUACAACACAGGACCCGGGC (SEQ ID NO. 14)
hsa-miR-4286	MIMAT0016916	1.590	1.05E-49	0.8849	ACCCCACUCCUGGUACC (SEQ ID NO. 15)
hsa-miR-6746-5p	MIMAT0027392	1.346	1.53E-49	0.8734	CCGGGAGAAGGAGGUGGCCUGG (SEQ ID NO. 16)
hsa-miR-663b	MIMAT0005867	1.201	9.31E-49	0.8872	GGUGGCCCGGCCGUGCCUGAGG (SEQ ID NO. 17)
hsa-miR-6075	MIMAT0023700	0.794	2.77E-47	0.8913	ACGGCCCAGGCGGCAUUGGUG (SEQ ID NO. 18)
hsa-miR-5001-5p	MIMAT0021021	0.796	3.16E-46	0.8841	AGGGCUGGACUCAGCGGCGGAGCU (SEQ ID NO. 19)
hsa-miR-6789-5p	MIMAT0027478	0.683	6.98E-46	0.8925	GUAGGGGCGUCCCGGGCGCGCGGG (SEQ ID NO. 20)
hsa-miR-4513	MIMAT0019050	1.063	1.19E-45	0.8946	AGACUGACGGCUGGAGGCCCAU (SEQ ID NO. 21)
hsa-miR-3192-5p	MIMAT0015076	4.111	1.76E-45	0.8605	UCUGGGAGGUUGUAGCAGUGGAA (SEQ ID NO. 22)
hsa-miR-8060	MIMAT0030987	3.502	1.77E-45	0.8779	CCAUGAAGCAGUGGGUAGGAGGAC (SEQ ID NO. 23)
hsa-miR-668-5p	MIMAT0026636	2.748	2.02E-45	0.8934	UGCGCCUCGGGUGAGCAUG (SEQ ID NO. 24)
hsa-miR-1268a	MIMAT0005922	-0.610	2.40E-45	0.8739	CGGGCGUGGUGGUGGGGG (SEQ ID NO. 25)
hsa-miR-1273g-3p	MIMAT0022742	1.448	2.67E-45	0.8620	ACCACUGCACUCCAGCCUGAG (SEQ ID NO. 26)
hsa-miR-4706	MIMAT0019806	1.063	5.43E-45	0.8509	AGCGGGGAGGAAGUGGGCGCUGCUU (SEQ ID NO. 27)
hsa-miR-124-3p	MIMAT0000422	3.734	5.43E-45	0.8903	UAAGGCACGCGGUGAAUGCCAA (SEQ ID NO. 28)
hsa-miR-1260b	MIMAT0015041	1.278	9.38E-45	0.8641	AUCCCACCACUGCCACCAU (SEQ ID NO. 29)

[0168]

hsa-miR-4740-5p	MIMAT0019869	3.165	9.50E-45	0.8884	AGGACUGAUCCUCUCGGGCAGG (SEQ ID NO. 30)
hsa-miR-320b	MIMAT0005792	2.317	1.08E-44	0.8868	AAAAGCUGGGUUGAGAGGGCAA (SEQ ID NO. 31)
hsa-miR-7977	MIMAT0031180	1.267	4.78E-43	0.8679	UUCCCAGCCAACGCACCA (SEQ ID NO. 32)
hsa-miR-29b-3p	MIMAT0000100	4.104	1.07E-42	0.8607	UAGCACCAUUUGAAAUCAGUGUU (SEQ ID NO. 33)
hsa-miR-4708-3p	MIMAT0019810	2.780	2.73E-42	0.8571	AGCAAGGCGGCAUCUCUCUGAU (SEQ ID NO. 34)
hsa-miR-4525	MIMAT0019064	2.389	3.12E-42	0.8480	GGGGGGAUGUGCAUGCUGGUU (SEQ ID NO. 35)
hsa-miR-92b-3p	MIMAT0003218	2.494	3.43E-42	0.8677	UAUUGCACUCGUCCCGGCCUCC (SEQ ID NO. 36)
hsa-miR-4257	MIMAT0016878	1.007	4.69E-42	0.8588	CCAGAGGUGGGGACUGAG (SEQ ID NO. 37)
hsa-miR-4727-3p	MIMAT0019848	2.681	7.55E-42	0.8641	AUAGUGGGAAGCUGGCAGAUUC (SEQ ID NO. 38)
hsa-miR-92a-3p	MIMAT0000092	2.012	9.49E-42	0.8628	UAUUGCACUUGUCCCGGCCUGU (SEQ ID NO. 39)
hsa-miR-663a	MIMAT0003326	1.077	1.02E-41	0.8429	AGGCGGGGCGCCGCGGGACCGC (SEQ ID NO. 40)
hsa-miR-6787-5p	MIMAT0027474	1.234	5.33E-41	0.8343	UGGCGGGGGUAGAGCUGGCUGC (SEQ ID NO. 41)
hsa-miR-3131	MIMAT0014996	1.186	7.21E-41	0.8529	UCGAGGACUGGUGGAAGGGCCUU (SEQ ID NO. 42)
hsa-miR-6802-5p	MIMAT0027504	0.851	2.03E-40	0.8382	CUAGGUGGGGGGCUUGAAGC (SEQ ID NO. 43)
hsa-miR-654-5p	MIMAT0003330	2.538	3.90E-40	0.8698	UGGUGGGCCGCAGAACAUGUGC (SEQ ID NO. 44)
hsa-miR-6511b-5p	MIMAT0025847	1.931	1.70E-39	0.8943	CUGCAGGCAGAAGUGGGGCUGACA (SEQ ID NO. 45)
hsa-miR-29b-1-5p	MIMAT0004514	4.291	1.38E-38	0.8268	GCUGGUUUCAUAUGGUGGUUUAGA (SEQ ID NO. 46)
hsa-miR-4417	MIMAT0018929	0.424	1.66E-38	0.8815	GGUGGGCUUCCCGGAGGG (SEQ ID NO. 47)
hsa-miR-4736	MIMAT0019862	1.509	2.07E-38	0.8664	AGGCAGGUUAUCUGGGCUG (SEQ ID NO. 48)
hsa-miR-6840-3p	MIMAT0027583	0.913	3.82E-38	0.8361	GCCCAGGACUUUGUGCGGGGUG (SEQ ID NO. 49)
hsa-miR-4710	MIMAT0019815	2.579	4.97E-38	0.8454	GGGUGAGGGCAGGUGGUU (SEQ ID NO. 50)
hsa-miR-4635	MIMAT0019692	2.883	6.15E-38	0.8521	UCUUGAAGUCAGAACCCGCAA (SEQ ID NO. 51)
hsa-miR-296-3p	MIMAT0004679	1.513	1.29E-37	0.8258	GAGGGUUGGGUGGAGGCUCUCC (SEQ ID NO. 52)
hsa-miR-1199-5p	MIMAT0031119	1.674	1.86E-37	0.9206	CCUGAGCCCGGGCCGCGCAG (SEQ ID NO. 53)
hsa-miR-7975	MIMAT0031178	1.390	2.19E-37	0.8393	AUCCUAGUCACGGCACCA (SEQ ID NO. 54)
hsa-miR-4480	MIMAT0019014	3.982	4.89E-37	0.8496	AGCCAAGUGGAAGUUACUUUA (SEQ ID NO. 55)
hsa-miR-3648	MIMAT0018068	0.970	5.72E-37	0.8367	AGCCGCGGGGAUCGCCGAGGG (SEQ ID NO. 56)

[0169]

hsa-miR-371a-5p	MIMAT0004687	0.870	6.10E-37	0.8597	ACUCAAACUGUGGGGGCACU (SEQ ID NO. 57)
hsa-miR-4771	MIMAT0019925	3.676	8.98E-37	0.8670	AGCAGACUUGACCUACAAUUA (SEQ ID NO. 58)
hsa-miR-6717-5p	MIMAT0025846	1.864	1.57E-36	0.8297	AGGCGAUGUGGGGAUGUAGAGA (SEQ ID NO. 59)
hsa-miR-1254	MIMAT0005905	1.180	1.74E-36	0.8502	AGCCUGGAAGCUGGAGCCUGCAGU (SEQ ID NO. 60)
hsa-miR-1246	MIMAT0005898	4.358	2.88E-36	0.8329	AAUGGAUUUUUGGAGCAGG (SEQ ID NO. 61)
hsa-miR-23b-3p	MIMAT0000418	3.275	4.55E-36	0.8473	AUCACAUUGCCAGGGAUUACCAC (SEQ ID NO. 62)
hsa-miR-320a	MIMAT0000510	1.560	6.90E-36	0.8407	AAAAGCUGGGUUGAGAGGGCGA (SEQ ID NO. 63)
hsa-miR-4687-5p	MIMAT0019774	1.025	1.02E-35	0.8424	CAGCCCUCCUCCCGCACCCAAA (SEQ ID NO. 64)
hsa-miR-191-5p	MIMAT0000440	3.613	2.26E-35	0.8409	CAACGGAAUCCCAAAAGCAGCUG (SEQ ID NO. 65)
hsa-miR-320c	MIMAT0005793	2.483	2.27E-35	0.8688	AAAAGCUGGGUUGAGAGGGU (SEQ ID NO. 66)
hsa-miR-6131	MIMAT0024615	3.119	4.64E-35	0.7915	GGCUGGUCAGAUGGGAGUG (SEQ ID NO. 67)
hsa-miR-4515	MIMAT0019052	2.382	4.69E-35	0.8161	AGGACUGGACUCCCGGCAGCCC (SEQ ID NO. 68)
hsa-miR-342-5p	MIMAT0004694	1.848	4.72E-35	0.8405	AGGGGUGCUAUCUGUGAUUGA (SEQ ID NO. 69)
hsa-miR-4718	MIMAT0019831	3.469	4.73E-35	0.8508	AGCUGUACCUGAAACCAAGCA (SEQ ID NO. 70)
hsa-miR-23a-3p	MIMAT0000078	3.010	5.66E-35	0.8385	AUCACAUUGCCAGGGAUUUCC (SEQ ID NO. 71)
hsa-miR-4455	MIMAT0018977	2.615	6.45E-35	0.8547	AGGGUGUGUGUGUUUUU (SEQ ID NO. 72)
hsa-miR-211-3p	MIMAT0022694	1.267	1.43E-34	0.8042	GCAGGGACAGCAAAGGGGUGC (SEQ ID NO. 73)
hsa-miR-3122	MIMAT0014984	1.673	2.63E-34	0.8801	GUUGGGACAAGAGGACGGUCUU (SEQ ID NO. 74)
hsa-miR-103a-3p	MIMAT0000101	3.839	4.23E-34	0.8246	AGCAGCAUUGUACAGGGCUAUGA (SEQ ID NO. 75)
hsa-miR-4429	MIMAT0018944	1.427	1.36E-33	0.8235	AAAAGCUGGGCUGAGAGGCG (SEQ ID NO. 76)
hsa-miR-920	MIMAT0004970	2.319	1.79E-33	0.8328	GGGGAGCUGUGGAAGCAGUA (SEQ ID NO. 77)
hsa-miR-3194-3p	MIMAT0019218	3.177	1.93E-33	0.8231	AGCUCUGCUGCUCACUGGCAGU (SEQ ID NO. 78)
hsa-miR-4754	MIMAT0019894	3.293	2.21E-33	0.8155	AUGCGGACCUGGGUUAGCGGAGU (SEQ ID NO. 79)
hsa-miR-1238-5p	MIMAT0022947	1.377	2.33E-33	0.7838	GUGAGUGGGAGCCCCAGUGUGUG (SEQ ID NO. 80)
hsa-miR-3191-3p	MIMAT0015075	1.600	2.38E-33	0.8673	UGGGGACGUAGCUGGCCAGACAG (SEQ ID NO. 81)
hsa-miR-4755-3p	MIMAT0019896	3.798	3.45E-33	0.8298	AGCCAGGCUCUGAAGGGAAAGU (SEQ ID NO. 82)
hsa-miR-3688-5p	MIMAT0019223	3.645	7.47E-33	0.8113	AGUGGCAAAGUCUUUCCAUAU (SEQ ID NO. 83)

[0170]

hsa-miR-4529-5p	MIMAT0019236	3.453	1.07E-32	0.8208	AGGCCAUCAGCAGUCCAA UGAA (SEQ ID NO. 84)
hsa-miR-6861-5p	MIMAT0027623	0.818	1.20E-32	0.8007	ACUGGGUAGGUGGGGCUC CAGG (SEQ ID NO. 85)
hsa-miR-1469	MIMAT0007347	0.758	2.45E-32	0.8228	CUCGGCGCGGGGCGCGGGC UCC (SEQ ID NO. 86)
hsa-miR-619-5p	MIMAT0026622	1.750	2.88E-32	0.8413	GCUGGGAUUACAGGCAUG AGCC (SEQ ID NO. 87)
hsa-miR-4448	MIMAT0018967	2.410	3.95E-32	0.8064	GGCUCCUUGGUCUAGGGG UA (SEQ ID NO. 88)
hsa-miR-4658	MIMAT0019725	2.788	4.02E-32	0.8321	GUGAGUGUGGAUCCUGGA GGAAU (SEQ ID NO. 89)
hsa-miR-22-3p	MIMAT0000077	2.815	5.70E-32	0.8327	AAGCUGCCAGUUGAAGAA CUGU (SEQ ID NO. 90)
hsa-miR-4776-5p	MIMAT0019932	2.510	6.41E-32	0.8355	GUGGACCAGGAUGGCAAG GGCU (SEQ ID NO. 91)
hsa-miR-320e	MIMAT0015072	3.365	1.05E-31	0.8191	AAAGCUGGGUUGAGAAGG (SEQ ID NO. 92)
hsa-miR-1225-3p	MIMAT0005573	0.741	1.99E-31	0.8297	UGAGCCCCUGUGCCGCCCC CAG (SEQ ID NO. 93)
hsa-miR-6875-5p	MIMAT0027650	-0.840	2.62E-31	0.8291	UGAGGGACCCAGGACAGG AGA (SEQ ID NO. 94)
hsa-miR-4534	MIMAT0019073	1.324	3.60E-31	0.8167	GGAUGGAGGAGGGGUCU (SEQ ID NO. 95)
hsa-miR-4652-5p	MIMAT0019716	3.280	3.60E-31	0.8156	AGGGGACUGGUUAAUAGA ACUA (SEQ ID NO. 96)
hsa-miR-648	MIMAT0003318	3.145	4.13E-31	0.8014	AAGUGUGCAGGGCACUGG U (SEQ ID NO. 97)
hsa-miR-4259	MIMAT0016880	2.262	4.13E-31	0.8147	CAGUUGGGUCUAGGGGUC AGGA (SEQ ID NO. 98)
hsa-miR-107	MIMAT0000104	3.642	6.38E-31	0.8167	AGCAGCAUUGUACAGGGC UAUCA (SEQ ID NO. 99)
hsa-miR-650	MIMAT0003320	2.399	7.82E-31	0.8214	AGGAGGCAGCGCUCUCAG GAC (SEQ ID NO. 100)

[0171] 3.2.诊断模型的开发

[0172] 在肺癌数据集的发现组中进行了诊断模型的开发,该数据集包括208名肺癌患者和208名非癌症对照(图1A)。使用limma分析来评估肺癌患者与非癌症对照之间miRNA表达差异的统计学显著性。表1列出了前100个差异表达的miRNAs。十折交叉验证表明,基于经过调整的p值排名的前4个miRNAs(hsa-miR-5100、hsa-miR-1343-3p、hsa-miR-1290和hsa-miR-4787-3p)的诊断模型将在ROC曲线分析中获得最佳AUC(图2A)。由4个miRNA表达水平的加权和计算的诊断指数,并归一化为零到十的范围,显示了接近完美的AUC值为0.999(图2B),在数值上优于原始出版物中原始2-miRNA模型的AUC为0.993(Asakura et al.2020)(p=0.16)。为了确保在发现组中不会错误分类非癌症对照而最小化假阳性,选择了6.0的阈值,这可以推出了98%的灵敏度和100%的特异性(图2C),而原始2-miRNA模型(Asakura et al.2020)的灵敏度和特异性都为99%。

[0173] 3.3.肺癌验证组的诊断模型验证

[0174] 4-miRNA模型在肺癌验证组中的性能表现得到评估,该验证组包括1358名肺癌患者和1970名非癌症对照。4-miRNA模型取得了0.999的AUC值为(见图2D),明显优于原始2-miRNA模型的AUC值0.996(Asakura et al.2020)(p=0.01)。新模型还表现出99%的灵敏度和99%的特异性(见图2E),而原始的2-miRNA模型显示出95%的灵敏度和99%的特异性(Asakura et al.2020)。

[0175] 此外,还对验证组的临床小组进行了4-miRNA模型的性能评估,这些小组是根据临床分期、T分期、N分期、M分期和组织学定义的。在所有临床小组中,4-miRNA模型显示出灵敏度约为99%或更高(见图2G,表2),优于原始2-miRNA模型的灵敏度(表2)。特别是对于早期肺癌,例如对于I期肺癌患者和T1肿瘤患者,4-miRNA模型表现出>99%的灵敏度(见图2G,表2),而2-miRNA模型的灵敏度分别为95.4%和95.9%(表2)。在常见的腺癌和鳞状细胞癌组织学类型中,4-miRNA模型也表现出更好的性能(见图2G,表2),相比原始的2-miRNA模型(表2)。

[0176] 表2.在肺癌验证组的临床小组中,原始的2-miRNA模型和新的4-miRNA模型的灵敏度比较,同时保持特异性>99%

[0177]

Clinical Subsets		N	Original 2-miRNA model	New 4-miRNA model	P value*
Clinical Stage	IA	686	96.1%	99.6%	<0.001
	IB	285	93.7%	99.6%	<0.001
	IIA	146	97.3%	97.9%	0.99
	IIB	61	96.7%	98.4%	0.99
	IIIA	164	90.2%	99.4%	<0.001
	IIIB	6	83.3%	100.0%	0.99
	IV	8	100.0%	100.0%	1.00
T Stage	T1a	466	96.1%	99.6%	<0.001
	T1b	297	95.6%	99.3%	0.003
	T2a	435	93.6%	99.1%	<0.001
	T2b	52	92.3%	100.0%	0.134
	T3	89	94.4%	98.9%	0.221
	T4	17	94.1%	100.0%	0.99
N Stage	N0	1047	95.5%	99.5%	<0.001
	N1	166	95.8%	98.2%	0.289
	N2	142	90.1%	99.3%	<0.001
M Stage	M0	1348	94.7%	99.3%	<0.001
	M1a	8	100.0%	100.0%	1.00
Histology	ADC	1038	95.1%	99.2%	<0.001
	SqCC	205	94.2%	99.5%	0.006
	LCC	34	97.1%	100.0%	0.99
	SCLC	22	90.9%	100.0%	0.480
	Others	57	96.5%	100.0%	0.480

[0178] *p values calculated by McNemar Test.

[0179] 其中,分析180名受试者的配对血清样本数据(术前与术后)后发现,4-miRNA模型的诊断指数对于术后血清样本降至正常水平,低于诊断指数的截止值(图2F)。

[0180] 3.4.诊断模型在其他癌症中的应用

[0181] 4-miRNA模型的性能进一步在3792名受试者的综合数据集中进行了评估,其中包括来自12种不同癌症类型的2038名癌症患者和1754名非癌症对照。膀胱癌、肝癌和卵巢癌样本规模最大,每种癌症都有300多名患者。除了乳腺癌表现不佳外,4-miRNA模型在胆道、膀胱、结直肠、食管、胃、胶质瘤、肝癌、卵巢癌、胰腺癌和前列腺癌等癌症中均表现出很强的性能,AUC值均>0.95,并在肉瘤中的AUC为0.876(图3A)。因此,4-miRNA模型在胆道、膀胱、结直肠、食管、胃、胶质瘤、肝癌、胰腺癌和前列腺癌的灵敏度范围从83.2%到100%,而在卵巢癌和肉瘤中的灵敏度分别为68.2%和72.0%(图3B)。此外,对于1754名非肺癌数据集中的

非癌症对照,4-miRNA模型保持了99.3%的高特异性。

[0182] 使用将特异性降低到95%的备用诊断指数截断值为5.1的进一步灵敏度分析得出了所有11种癌症类型的灵敏度增加,显示出在除了肉瘤的10种癌症类型中的灵敏度均>90%(表3)。

[0183] 表3.使用默认阈值与特异性达到95%的备用阈值,额外癌症数据集中4-miRNA诊断模型的灵敏度比较

[0184]

	Default cut-point based on 99% specificity	Alternative cut-point based on 95% specificity
Biliary Tract Cancer	97.5%	100.0%
Bladder Cancer	98.2%	99.2%
Colorectal Cancer	85.8%	91.6%
Esophageal Cancer	84.7%	95.2%
Gastric Cancer	100.0%	100.0%
Glioma	87.5%	97.5%
Liver Cancer	84.2%	92.5%
Ovarian Cancer	68.2%	90.1%
Pancreatic Cancer	83.2%	95.3%
Prostate Cancer	92.5%	97.5%
Sarcoma	72.0%	76.5%

[0185] 4.讨论

[0186] 在这个实施例中,我们报告了一个多癌症早期检测的4-miRNA诊断模型的开发和性能评估。我们证明在一个包括3396名癌症患者和3724名非癌症个体的大型独立验证组中,这个4-miRNA模型可以同时高灵敏度(80%-100%对于十种癌症类型,大约70%对于两种癌症类型)地检测12种癌症类型(胆道、膀胱、结直肠、食道、胃、胶质瘤、肝、肺、卵巢、胰腺、前列腺和肉瘤),同时仍然保持着通常需要在风险人群中使用的筛查测试非常高的特异性,即99%。据我们所知,这是首个基于血液miRNA的MCED诊断模型。值得一提的是,肺癌患者的诊断指数在肿瘤切除后降至非癌症对照组的水平,这表明这个诊断模型可能具有监测肿瘤复发的潜力。

[0187] 分析血液中的核酸和/或蛋白质的非侵入性筛查测试已成为MCED运动的推动力,最近取得了显著进展。几乎所有正在开发的MCED测试都是基于评估血液中的细胞游离肿瘤DNA,而且大多数利用下一代亚硫酸盐测序技术来评估这些肿瘤DNA的甲基化模式。其中,有两种测试,Galleri和PanSeer,是作为甲基化基因表观遗传学标志物开发的。在Circulating Cell-free Genome Atlas(CCGA)的病例对照研究中,Galleri调查了100,000多个甲基化区域,并显示了对12种预定癌症(肛门、膀胱、结肠/直肠、食道、头颈、肝/胆道、肺、淋巴瘤、卵巢、胰腺、浆细胞肿瘤、胃)的灵敏度在I至III期疾病患者(n=874)中为67.6%,当包括IV期癌症时,灵敏度提高至76.3%(n=1346),并达到了99.3%的特异性,基于1254名非癌症对照(Klein et al.2021)。另一方面,PanSeer检测仅针对477个甲基化基因组区域进行回顾性分析,研究了一组无症状个体的血浆样本,这些个体已在长期癌症监测研究中入组,并在抽血后四年内被诊断出患有五种癌症之一(胃癌、食道癌、结直肠癌、肺

癌和肝癌)。这些个体的灵敏度高达95％,但对207名非癌对照组的特异性较低,为96％(Chen et al.2020)。然而,令人困惑的是,当PanSeer在113名诊断后的血浆样本中进行评估时,该测试仅显示出较低的88％灵敏度(Chen et al.2020)。另一个名为DELFI的测试,它基于下一代测序技术对细胞外DNA的全基因组分析,实现了对七种癌症(n＝208,乳腺、胆道、结直肠、胃、肺、卵巢和胰腺)的73％灵敏度和98％特异性(n＝215)(Cristiano et al.2019)最后,CancerSEEK是一项将9个蛋白质生物标志物的测量与血液中的细胞游离DNA中16个基因突变的检测相结合的测试,经过十倍交叉验证,在八种癌症中(卵巢、肝、胃、胰腺、食道、结直肠、肺、乳腺)实现了中位数70％的灵敏度(n＝1005)和99％的特异性(n＝812)(Cohen et al.2018))。总的来说,目前正在开发的MCED测试通常在99％的特异性要求下显示出60-70％的灵敏度范围。与这些测试相比,我们的诊断模型要简单得多,只包括4种miRNA,但在超过7000名参与者的大样本中,对于12种癌症中的10种,它表现出明显更高的灵敏度范围,为80-100％。值得注意的是,一个简单的诊断模型不仅成本明显较低,还可以使用传统技术平台(例如RT-PCR)开发成体外诊断(IVD)测试,这与通常作为实验室开发的测试(LDT)实施的基于NGS的测试相比具有优势。这些特点对于推动广泛采用MCED测试是重要的,因为它们旨在针对高风险或潜在风险的人群做癌症筛查。.

[0188] 在本研究中检测的13种癌症中,唯独乳腺癌未能被4种miRNA诊断模型成功检测到。尽管尚不清楚造成这种性能不佳的原因,但它可能表明乳腺癌具有不同的miRNA表达谱和/或不同的miRNA进入血液流的释放模式。有趣的是,Galleri和CancerSEEK在乳腺癌中也表现出较低的灵敏度,分别为30.5％和33％(Klein et al.2021;Cohen et al.2018)。然而,乳腺癌的性能不佳可能在临床上不重要,因为乳腺X线摄影筛查在检测早期乳腺癌和降低乳腺癌死亡率方面非常有效(Nelson et al.2016)。

[0189] 这些多癌早期检测(MCED)测试的最终诊断性能和临床价值必须在包括无症状个体的大型前瞻性筛查试验中建立。在涉及超过10,000名无症状女性的DETECT-A试验中,共发现了96例癌症,跨足10种癌症类型,CancerSEEK显示出27％的灵敏度,当结合标准护理筛查测试检测出的癌症时,灵敏度提高到52％(Lennon et al.2020)。此外,CancerSEEK与PET-CT扫描结合使用,显示出99.6％的特异性和40.6％的阳性预测值(PPV)。另一方面,在Galleri测试的前瞻性PATHFINDER研究的4033名参与者的中期分析中,有40名参与者测试结果呈阳性,其中18名被证实患有癌症,这推出PPV为45％(Beer et al.2021)。对于我们的4-miRNA诊断模型,假设1％的癌症发病率和平均灵敏度为85％以及99.3％的特异性,当筛查无症状个体时,我们的模型将提供55％的PPV。这明显高于四种USPSTF推荐的单一癌症筛查的PPV,这些筛查的PPV在3.7％到4.4％之间(Lehman et al.2017;U.S.Food and Drug Administration Cologuard Summary of Safety and Effectiveness Data,2014;and National Lung Screening Trial Research Team,2013)。.

[0190] 5.结论

[0191] 总之,我们的研究为一种简单、经济的基于血液的多癌症诊断测试提供了概念验证数据。在本研究中检测到的这12种癌症类型占2021年美国估计的癌症死亡人数的近380,000例(约62％)。尽管早期发现这些癌症应该可以减少与癌症相关的死亡人数,但最终的临床性能和临床效用的确定需要在大规模前瞻性研究中进行评估,该研究对象是来自预期使用人群中的无症状个体。

[0192] 尽管上述提供的实施例和数据仅涵盖了12种癌症,尤其是4-miRNA标志物组在检测癌症方面表现出卓越的准确性,但miRNA标志物组不仅限于适用于这些癌症类型。因此,本发明的范围将被解释为涵盖其他癌症类型。本发明提供的模型适用于研究的13种癌症中的12种,这提示该方法很可能适用于几乎所有癌症类型。

[0193] 参考文献

[0194] Ritchie,ME;et al.(2015).limma powers differential expression analyses for RNA-sequencing and microarray studies.Nucleic Acids Research 43(7),e47.

[0195] Venables,WN and Ripley,BD(2002)Modern Applied Statistics with S.Fourth edition.Springer.

[0196] Tibshirani,R(1996)."Regression Shrinkage and Selection via the lasso".Journal of the Royal Statistical Society.Series B(methodological).Wiley.58(1):267-88.

[0197] Hoerl,AE and Kennard,RW(1970)."Ridge Regression:Biased Estimation for Nonorthogonal problems".Technometrics.12(1):55-67.

[0198] Ripley,BD(1996)Pattern Recognition and Neural Networks.Cambridge University Press.

[0199] Kozomara,A and Griffiths-Jones,S(2010)."MiRBase:integrating microRNA annotation and deep-sequencing data".Nucleic Acids Research.39(Database issue):D152-7.

[0200] miRBase:the microRNA database:http://www.mirbase.org/

[0201] The Bioconductor Open Source Software For Bioinformatics:http://www.bioconductor.org

[0202] The R Project for Statistical Computing:https://www.r-project.org/

[0203] Asakura,K;et al.(2020).A MiRNA-Based Diagnostic Model Predicts Resectable Lung Cancer in Humans with High Accuracy.Commun.Biol.3,134.

[0204] Yokoi,A;et al.(2018).Integrated Extracellular MicroRNA Profiling for Ovarian Cancer Screening.Nat.Commun.9,4319.

[0205] Usuba,W;et al.(2019).Circulating MiRNA Panels for Specific and Early Detection in Bladder Cancer.Cancer Sci.110,408-419.

[0206] Yamamoto,Y;et al.(2020).Highly Sensitive Circulating MicroRNA Panel for Accurate Detection of Hepatocellular Carcinoma in Patients With Liver Disease.Hepatol.Commun.4,284-297.

[0207] Klein,EA;et al.(2021).Clinical Validation of a Targeted Methylation-Based Multi-Cancer Early Detection Test Using an Independent Validation Set.Ann.Oncol.:Off.J.Eur.Soc.Med.Oncol.32,1167-1177.

[0208] Cohen,JD;et al.(2018).Detection and Localization of Surgically Resectable Cancers with a Multi-Analyte Blood Test.Science.359,926-930.

[0209] Chen,X;et al.(2020).Non-Invasive Early Detection of Cancer Four Years before Conventional Diagnosis Using a Blood Test.Nat.Commun.11,3475.

[0210] Cristiano,S;et al.(2019).Genome-Wide Cell-Free DNA Fragmentation in Patients with Cancer.Nature.570,385-389.

[0211] Nelson,HD;et al.(2016).Effectiveness of Breast Cancer Screening: Systematic Review and Meta-Analysis to Update the 2009 U.S.Preventive Services Task Force Recommendation.Ann.Intern.Med.164,244-255.

[0212] Lennon,AM;et al.(2020).Feasibility of Blood Testing Combined with PET-CT to Screen for Cancer and Guide Intervention.Science.369,eabb9601.

[0213] Beer,T;et al.(2021).Interim Results of PATHFINDER,a Clinical Use Study Using a Methylation-Based Multi-Cancer Early Detection Test.J.Clin.Oncol.39,3010.

[0214] Lehman,CD;et al.(2017).National Performance Benchmarks for Modern Screening Digital Mammography:Update from the Breast Cancer Surveillance Consortium.Radiology.283,49-58.U.S.Food and Drug Administration Cologuard Summary of Safety and Effectiveness Data(Premarket Approval Application P130017);2014.

[0215] National Lung Screening Trial Research Team;Church,TR;et al.(2013).Results of Initial Low-Dose Computed Tomographic Screening for Lung Cancer.New Engl.J.Med.2013,368,1980-1991.

[0216] Nielsen,PE;et al.(1991).Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide.Science.254,p.1497-500.

[0217] Obika,S;et al.(1998).Stability and structural features of the duplexes containing nucleoside analogues with a fixed N-type conformation,2'-O,4'-C-methyleneribonucleosides.Tetrahedron Lett..39,p.5401-5404.

[0218] Green,MR and Sambrook,J.(2012).Molecular Cloning:A Laboratory Manual, 4th Ed.,Cold Spring Harbor Press,Cold Spring Harbor,N.Y.

[0219] Sambrook,J;et al.(1989).Molecular Cloning:A Laboratory Manual,New York:Cold Spring Harbor Press.

[0220] Zhang,Z;et al.(2000).A greedy algorithm for aligning DNA sequences.J.Comput.Biol.7,p.203-214.

[0221] Altschul,SF;et al.(1990).Basic local alignment search tool.Journal of Molecular Biology,Vol.215,p.403-410.

[0222] Pearson,WR et al.(1988).Improved tools for biological sequence comparison.Proc.Natl.Acad.Sci.U.S.A.,Vol.85,p.2444-2448.

[0223] Yun,SJ;et al.(2012).Cell-free microRNAs in urine as diagnostic and prognostic biomarkers of bladder cancer.Int J Oncol.2012 Nov;41(5):1871-8.

[0224] Park,NJ;et al.(2009).Salivary microRNA:discovery,characterization,and clinical utility for oral cancer detection.Clin Cancer Res.2009 Sep 1;15(17):5473-7.

SEQUENCE LISTING

<110> 美瑞诺康诊断公司
安德鲁·张
胡海
<120> 癌症检测方法,试剂盒和系统
<130> ZH.001
<150> US 63/208,506
<151> 2021-06-09
<160> 100
<170> PatentIn version 3.5
<210> 1
<211> 22
<212> RNA
<213> 人
<400> 1
uucagauccc agcggugccu cu 22
<210> 2
<211> 22
<212> RNA
<213> 人
<400> 2
cuccuggggc ccgcacucuc gc 22
<210> 3
<211> 19
<212> RNA
<213> 人
<400> 3
uggauuuuug gaucaggga 19
<210> 4
<211> 24
<212> RNA
<213> 人
<400> 4
gaugcgccgc ccacugcccc gcgc 24
<210> 5
<211> 22
<212> RNA
<213> 人

<400> 5
agggccgaag gguggaagcu gc 22
<210> 6
<211> 22
<212> RNA
<213> 人
<400> 6
acugcaguga aggcacuugu ag 22
<210> 7
<211> 25
<212> RNA
<213> 人
<400> 7
gugaggcggg gccaggaggg ugugu 25
<210> 8
<211> 20
<212> RNA
<213> 人
<400> 8
cgggcguggu gguggggggug 20
<210> 9
<211> 17
<212> RNA
<213> 人
<400> 9
ccccgccacc gccuugg 17
<210> 10
<211> 22
<212> RNA
<213> 人
<400> 10
aaaccguuac cauuacugag uu 22
<210> 11
<211> 21
<212> RNA
<213> 人
<400> 11
gugggcgggg gcaggugugu g 21
<210> 12

<211> 22
<212> RNA
<213> 人
<400> 12
accuggcagc agggagcguc gu 22
<210> 13
<211> 20
<212> RNA
<213> 人
<400> 13
ggauccgagu cacggcacca 20
<210> 14
<211> 22
<212> RNA
<213> 人
<400> 14
ggcuacaaca caggacccgg gc 22
<210> 15
<211> 17
<212> RNA
<213> 人
<400> 15
accccacucc ugguacc 17
<210> 16
<211> 22
<212> RNA
<213> 人
<400> 16
ccgggagaag gagguggccu gg 22
<210> 17
<211> 22
<212> RNA
<213> 人
<400> 17
gguggcccgg ccgugccuga gg 22
<210> 18
<211> 21
<212> RNA
<213> 人

<400> 18
acggcccagg cggcauuggu g 21
<210> 19
<211> 24
<212> RNA
<213> 人
<400> 19
agggcuggac ucagcggcgg agcu 24
<210> 20
<211> 24
<212> RNA
<213> 人
<400> 20
guaggggcgu cccgggcgcg cggg 24
<210> 21
<211> 22
<212> RNA
<213> 人
<400> 21
agacugacgg cuggaggccc au 22
<210> 22
<211> 23
<212> RNA
<213> 人
<400> 22
ucugggaggu uguagcagug gaa 23
<210> 23
<211> 24
<212> RNA
<213> 人
<400> 23
ccaugaagca guggguagga ggac 24
<210> 24
<211> 19
<212> RNA
<213> 人
<400> 24
ugcgccucgg gugagcaug 19
<210> 25

<211> 18
<212> RNA
<213> 人
<400> 25
cgggcguggu gguggggg 18
<210> 26
<211> 21
<212> RNA
<213> 人
<400> 26
accacugcac uccagccuga g 21
<210> 27
<211> 25
<212> RNA
<213> 人
<400> 27
agcggggagg aagugggcgc ugcuu 25
<210> 28
<211> 22
<212> RNA
<213> 人
<400> 28
uaaggcacgc ggugaaugcc aa 22
<210> 29
<211> 19
<212> RNA
<213> 人
<400> 29
aucccaccac ugccaccau 19
<210> 30
<211> 22
<212> RNA
<213> 人
<400> 30
aggacugauc cucucgggca gg 22
<210> 31
<211> 22
<212> RNA
<213> 人

<400> 31
aaaagcuggg uugagagggc aa 22
<210> 32
<211> 18
<212> RNA
<213> 人
<400> 32
uucccagcca acgcacca 18
<210> 33
<211> 23
<212> RNA
<213> 人
<400> 33
uagcaccauu ugaaaucagu guu 23
<210> 34
<211> 22
<212> RNA
<213> 人
<400> 34
agcaaggcgg caucucucug au 22
<210> 35
<211> 21
<212> RNA
<213> 人
<400> 35
ggggggaugu gcaugcuggu u 21
<210> 36
<211> 22
<212> RNA
<213> 人
<400> 36
uauugcacuc gucccggccu cc 22
<210> 37
<211> 18
<212> RNA
<213> 人
<400> 37
ccagaggugg ggacugag 18
<210> 38

<210> 22
<211> 22
<212> RNA
<213> 人
<400> 38
auagugggaa gcuggcagau uc 22
<210> 39
<211> 22
<212> RNA
<213> 人
<400> 39
uauugcacuu gucccggccu gu 22
<210> 40
<211> 22
<212> RNA
<213> 人
<400> 40
aggcggggcg ccgcgggacc gc 22
<210> 41
<211> 22
<212> RNA
<213> 人
<400> 41
uggcgggggu agagcuggcu gc 22
<210> 42
<211> 23
<212> RNA
<213> 人
<400> 42
ucgaggacug guggaagggc cuu 23
<210> 43
<211> 20
<212> RNA
<213> 人
<400> 43
cuaggugggg ggcuugaagc 20
<210> 44
<211> 22
<212> RNA
<213> 人

<400> 44
uggugggccg cagaacaugu gc 22
<210> 45
<211> 24
<212> RNA
<213> 人
<400> 45
cugcaggcag aaguggggcu gaca 24
<210> 46
<211> 24
<212> RNA
<213> 人
<400> 46
gcugguuuca uaugugguuu uaga 24
<210> 47
<211> 18
<212> RNA
<213> 人
<400> 47
ggugggcuuc ccggaggg 18
<210> 48
<211> 19
<212> RNA
<213> 人
<400> 48
aggcagguua ucugggcug 19
<210> 49
<211> 22
<212> RNA
<213> 人
<400> 49
gcccaggacu uugugcgggg ug 22
<210> 50
<211> 18
<212> RNA
<213> 人
<400> 50
gggugagggc aggugguu 18
<210> 51

<211> 21
<212> RNA
<213> 人
<400> 51
ucuugaaguc agaacccgca a 21
<210> 52
<211> 22
<212> RNA
<213> 人
<400> 52
gaggguuggg uggaggcucu cc 22
<210> 53
<211> 20
<212> RNA
<213> 人
<400> 53
ccugagcccg ggccgcgcag 20
<210> 54
<211> 18
<212> RNA
<213> 人
<400> 54
auccuaguca cggcacca 18
<210> 55
<211> 21
<212> RNA
<213> 人
<400> 55
agccaagugg aaguuacuuu a 21
<210> 56
<211> 21
<212> RNA
<213> 人
<400> 56
agccgcgggg aucgccgagg g 21
<210> 57
<211> 20
<212> RNA
<213> 人

<400> 57
acucaaacug ugggggcacu 20
<210> 58
<211> 21
<212> RNA
<213> 人
<400> 58
agcagacuug accuacaauu a 21
<210> 59
<211> 22
<212> RNA
<213> 人
<400> 59
aggcgaugug gggauguaga ga 22
<210> 60
<211> 24
<212> RNA
<213> 人
<400> 60
agccuggaag cuggagccug cagu 24
<210> 61
<211> 19
<212> RNA
<213> 人
<400> 61
aauggauuuu uggagcagg 19
<210> 62
<211> 23
<212> RNA
<213> 人
<400> 62
aucacauugc cagggauuac cac 23
<210> 63
<211> 22
<212> RNA
<213> 人
<400> 63
aaaagcuggg uugagagggc ga 22
<210> 64

<211> 22
<212> RNA
<213> 人
<400> 64
cagcccuccu cccgcaccca aa 22
<210> 65
<211> 23
<212> RNA
<213> 人
<400> 65
caacggaauc ccaaaagcag cug 23
<210> 66
<211> 20
<212> RNA
<213> 人
<400> 66
aaaagcuggg uugagagggu 20
<210> 67
<211> 19
<212> RNA
<213> 人
<400> 67
ggcuggucag augggagug 19
<210> 68
<211> 22
<212> RNA
<213> 人
<400> 68
aggacuggac ucccggcagc cc 22
<210> 69
<211> 21
<212> RNA
<213> 人
<400> 69
aggggugcua ucugugauug a 21
<210> 70
<211> 21
<212> RNA
<213> 人

<400> 70
agcuguaccu gaaaccaagc a 21
<210> 71
<211> 21
<212> RNA
<213> 人
<400> 71
aucacauugc cagggauuuc c 21
<210> 72
<211> 17
<212> RNA
<213> 人
<400> 72
agggugugug uguuuuu 17
<210> 73
<211> 21
<212> RNA
<213> 人
<400> 73
gcagggacag caaaggggug c 21
<210> 74
<211> 22
<212> RNA
<213> 人
<400> 74
guugggacaa gaggacgguc uu 22
<210> 75
<211> 23
<212> RNA
<213> 人
<400> 75
agcagcauug uacagggcua uga 23
<210> 76
<211> 20
<212> RNA
<213> 人
<400> 76
aaaagcuggg cugagaggcg 20
<210> 77

<210> 20
<211> 20
<212> RNA
<213> 人
<400> 77
ggggagcugu ggaagcagua 20
<210> 78
<211> 22
<212> RNA
<213> 人
<400> 78
agcucugcug cucacuggca gu 22
<210> 79
<211> 23
<212> RNA
<213> 人
<400> 79
augcggaccu ggguuagcgg agu 23
<210> 80
<211> 23
<212> RNA
<213> 人
<400> 80
gugaguggga gccccagugu gug 23
<210> 81
<211> 23
<212> RNA
<213> 人
<400> 81
uggggacgua gcuggccaga cag 23
<210> 82
<211> 22
<212> RNA
<213> 人
<400> 82
agccaggcuc ugaagggaaa gu 22
<210> 83
<211> 21
<212> RNA
<213> 人

<400> 83
aguggcaaag ucuuuccaua u 21
<210> 84
<211> 22
<212> RNA
<213> 人
<400> 84
aggccaucag caguccaaug aa 22
<210> 85
<211> 22
<212> RNA
<213> 人
<400> 85
acuggguagg uggggcucca gg 22
<210> 86
<211> 22
<212> RNA
<213> 人
<400> 86
cucggcgcgg ggcgcgggcu cc 22
<210> 87
<211> 22
<212> RNA
<213> 人
<400> 87
gcugggauua caggcaugag cc 22
<210> 88
<211> 20
<212> RNA
<213> 人
<400> 88
ggcuccuugg ucuaggggua 20
<210> 89
<211> 23
<212> RNA
<213> 人
<400> 89
gugagugugg auccuggagg aau 23
<210> 90

<210> 22
<211> 22
<212> RNA
<213> 人
<400> 90
aagcugccag uugaagaacu gu 22
<210> 91
<211> 22
<212> RNA
<213> 人
<400> 91
guggaccagg auggcaaggg cu 22
<210> 92
<211> 18
<212> RNA
<213> 人
<400> 92
aaagcugggu ugagaagg 18
<210> 93
<211> 22
<212> RNA
<213> 人
<400> 93
ugagccccug ugccgccccc ag 22
<210> 94
<211> 21
<212> RNA
<213> 人
<400> 94
ugagggaccc aggacaggag a 21
<210> 95
<211> 17
<212> RNA
<213> 人
<400> 95
ggauggagga ggggucu 17
<210> 96
<211> 22
<212> RNA
<213> 人

<400> 96
aggggacugg uuaauagaac ua 22
<210> 97
<211> 19
<212> RNA
<213> 人
<400> 97
aagugugcag ggcacuggu 19
<210> 98
<211> 22
<212> RNA
<213> 人
<400> 98
caguuggguc uaggggucag ga 22
<210> 99
<211> 23
<212> RNA
<213> 人
<400> 99
agcagcauug uacagggcua uca 23
<210> 100
<211> 21
<212> RNA
<213> 人
<400> 100
aggaggcagc gcucucagga c 21

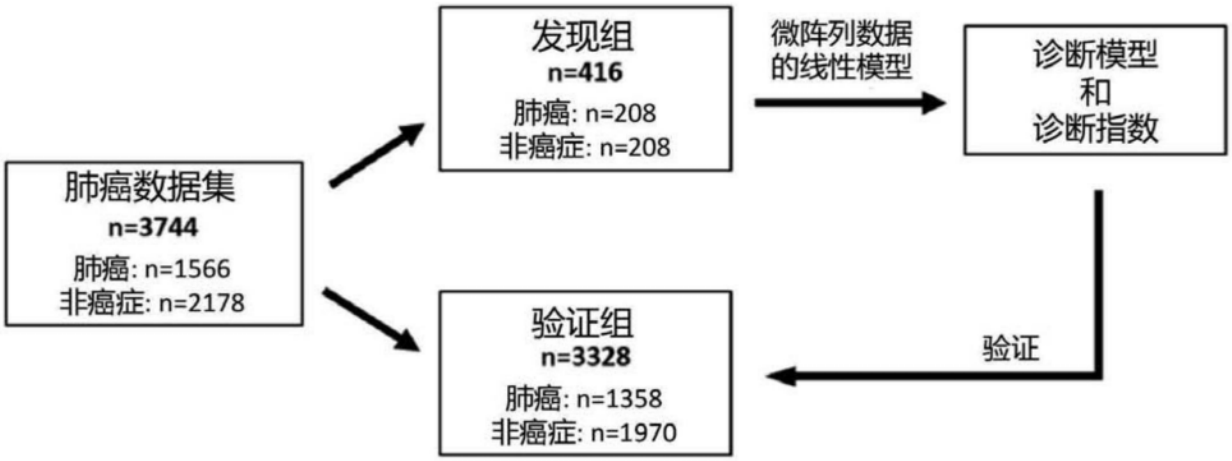

肺癌数据集
n=3744
肺癌: n=1566
非癌症: n=2178
发现组
n=416
肺癌: n=208
非癌症: n=208
验证组
n=3328
肺癌: n=1358
非癌症: n=1970
微阵列数据
的线性模型
诊断模型
和
诊断指数
验证

图.1A

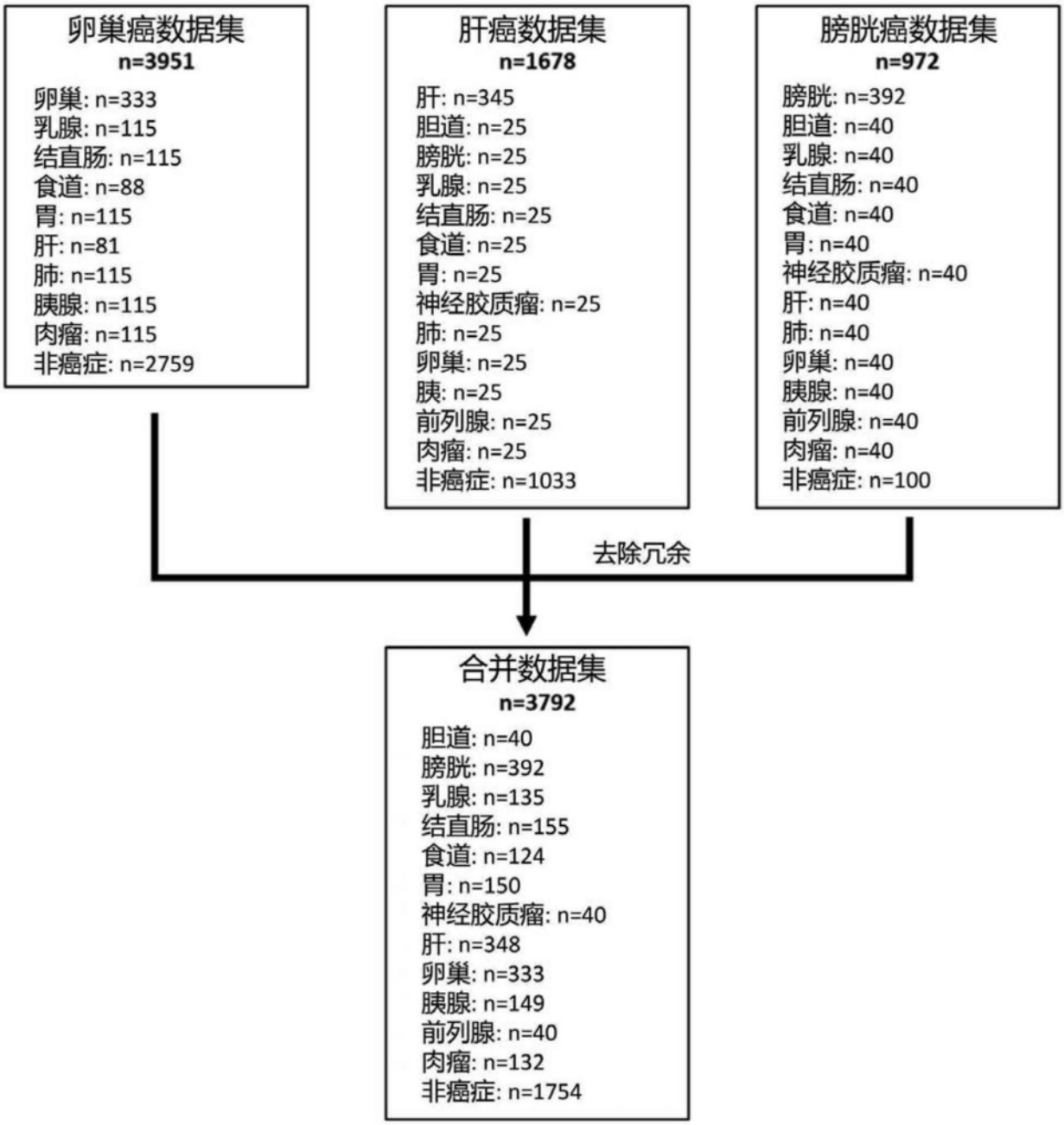

卵巢癌数据集
n=3951
卵巢: n=333
乳腺: n=115
结直肠: n=115
食道: n=88
胃: n=115
肝: n=81
肺: n=115
胰腺: n=115
肉瘤: n=115
非癌症: n=2759
肝癌数据集
n=1678
肝: n=345
胆道: n=25
膀胱: n=25
乳腺: n=25
结直肠: n=25
食道: n=25
胃: n=25
神经胶质瘤: n=25
肺: n=25
卵巢: n=25
胰: n=25
前列腺: n=25
肉瘤: n=25
非癌症: n=1033
膀胱癌数据集
n=972
膀胱: n=392
胆道: n=40
乳腺: n=40
结直肠: n=40
食道: n=40
胃: n=40
神经胶质瘤: n=40
肝: n=40
肺: n=40
卵巢: n=40
胰腺: n=40
前列腺: n=40
肉瘤: n=40
非癌症: n=100
去除冗余
合并数据集
n=3792
胆道: n=40
膀胱: n=392
乳腺: n=135
结直肠: n=155
食道: n=124
胃: n=150
神经胶质瘤: n=40
肝: n=348
卵巢: n=333
胰腺: n=149
前列腺: n=40
肉瘤: n=132
非癌症: n=1754

图.1B

肺癌 n=1566		膀胱癌 n=392		卵巢癌 n=333		肝癌 n=348	
年龄 (均值，方差)	65 (10)	年龄 (均值，标准差)	68 (11)	年龄 (均值，标准差)	57 (12)	年龄 (均值，方差)	68 (9)
性别		性别		分期		性别	
男性	895 (57%)	男性	283 (72%)	I	82 (25%)	男性	268 (78%)
女性	671 (43%)	女性	109 (28%)	II	33 (10%)	女性	77 (22%)
吸烟史		泌尿细胞学		III-IV	218 (65%)	未知	3
以前/现在	972 (62%)	I类	36 (10%)	组织学		分期	
从不	594 (38%)	II类	115 (31%)	浆液性的	182 (55%)	I	123 (37%)
组织学		III类	73 (19%)	透明细胞	64 (19%)	II	108 (33%)
腺癌	1217 (78%)	IV类	50 (13%)	子宫内膜样	43 (13%)	III	80 (24%)
鳞状	221 (14%)	V类	103 (27%)	黏液性	14 (4%)	IV	19 (6%)
腺鳞状	18 (1%)	未知	15	其他上皮细胞	17 (5%)	未知	18
小细胞	23 (1%)	T分期		非上皮细胞	13 (4%)	Child-Pugh	
其他	87 (6%)	<pT2	300 (77%)	非癌症 (N=2759)		A	303 (88%)
分期		≥pT2	90 (23%)	信息无资料		B	40 (12%)
I	1126 (72%)	未知	2			未知	5
II	233 (15%)	肿瘤分级				病毒	
III-IV	203 (13%)	低	77 (20%)			HBsAg+	57 (16%)
0	4 (0%)	高	315 (80%)			HCVAb+	141 (41%)
		淋巴结分期				non-B non-C	147 (43%)
		N+	42 (12%)			未知	3
		N0	320 (88%)				
		未知	30				
非癌症 (N=2178)		M分期				非癌症 (N=1033)	
年龄 (均值，标准差)	51 (11)	M1	17 (5%)			年龄 (均值，方差)	65 (10)
性别		M0	347 (95%)			性别	
男性	1129 (52%)	未知	28			男性	239 (23%)
女性	1049 (48%)					女性	794 (77%)
吸烟史							
以前/现在	482 (22%)						
从不	1696 (78%)						
		非癌症 (N=100)					
		年龄 (均值，标准差)	64 (16)				
		性别					
		男性	48 (48%)				
		女性	52 (52%)				

* 分别改编自参考文献 17, 19, 18, 20。

图.1C

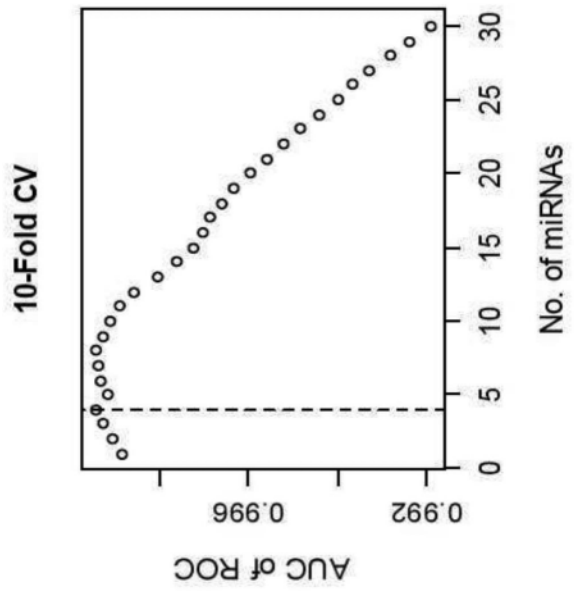

10-Fold CV
No. of miRNAs
0
5
10
15
20
25
30
AUC of ROC
0.992
0.996

图.2A

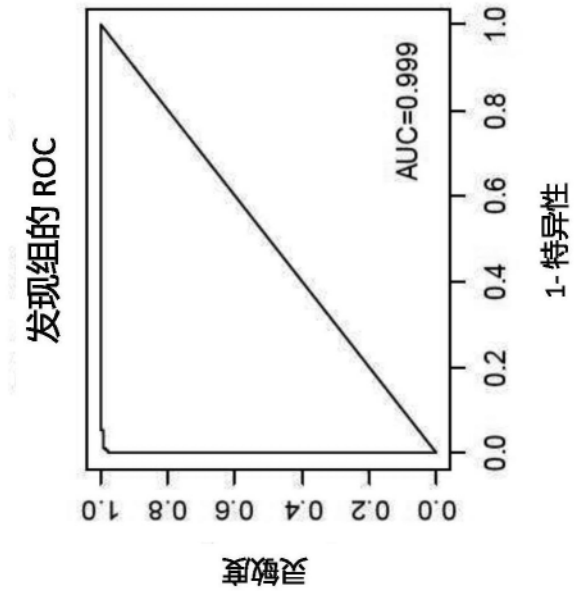

发现组的 ROC
AUC=0.999
灵敏度
1-特异性
0.0
0.2
0.4
0.6
0.8
1.0

图.2B

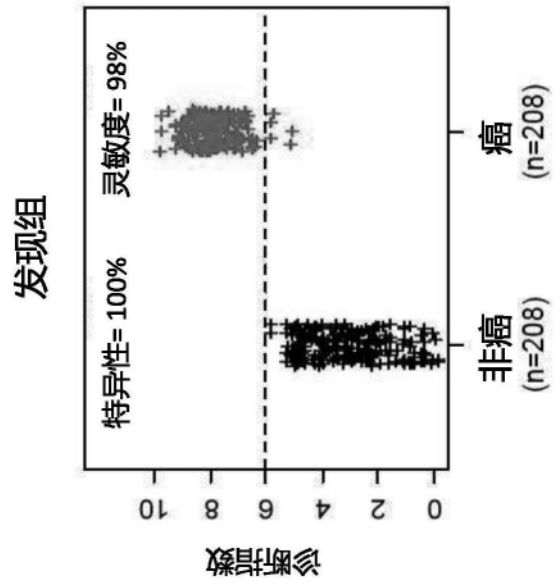

发现组
特异性= 100%
灵敏度= 98%
癌
(n=208)
非癌
(n=208)
诊断指数
0
2
4
6
8
10

图.2C

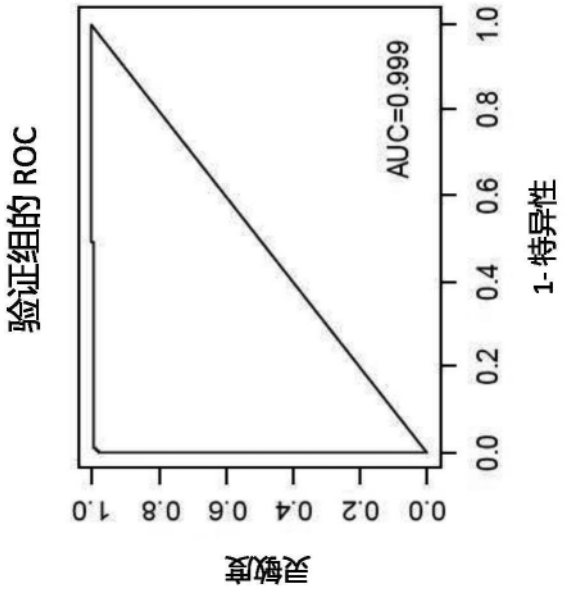

验证组的 ROC
AUC=0.999
灵敏度
1-特异性
0.0
0.2
0.4
0.6
0.8
1.0

图.2D

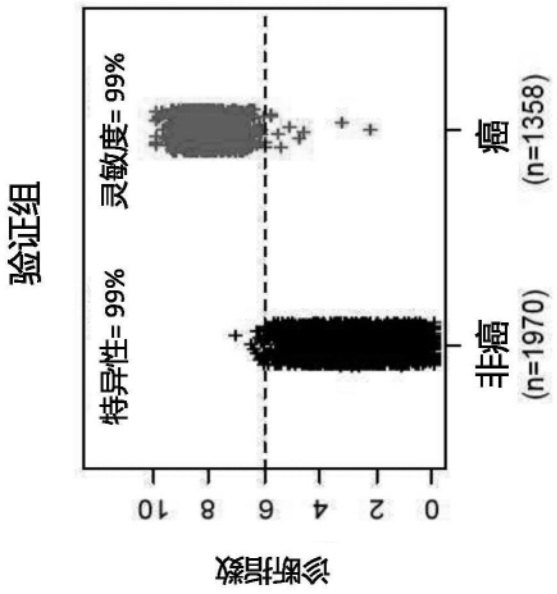

验证组
特异性= 99%
灵敏度= 99%
非癌
(n=1970)
癌
(n=1358)
诊断指数
0
2
4
6
8
10

图.2E

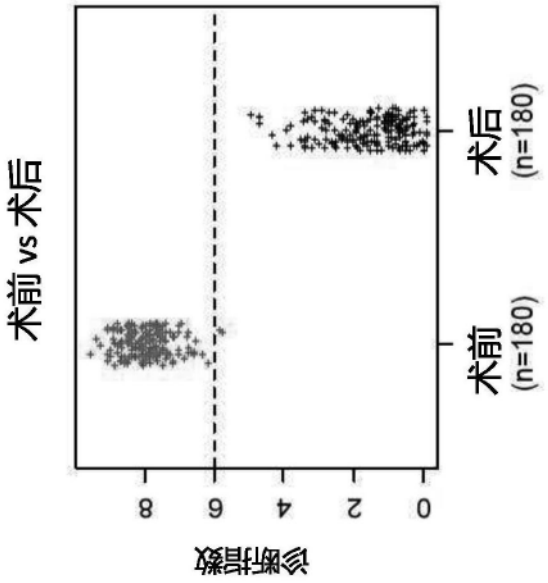

术前 vs 术后
术前
(n=180)
术后
(n=180)
诊断指数
0
2
4
6
8

图.2F

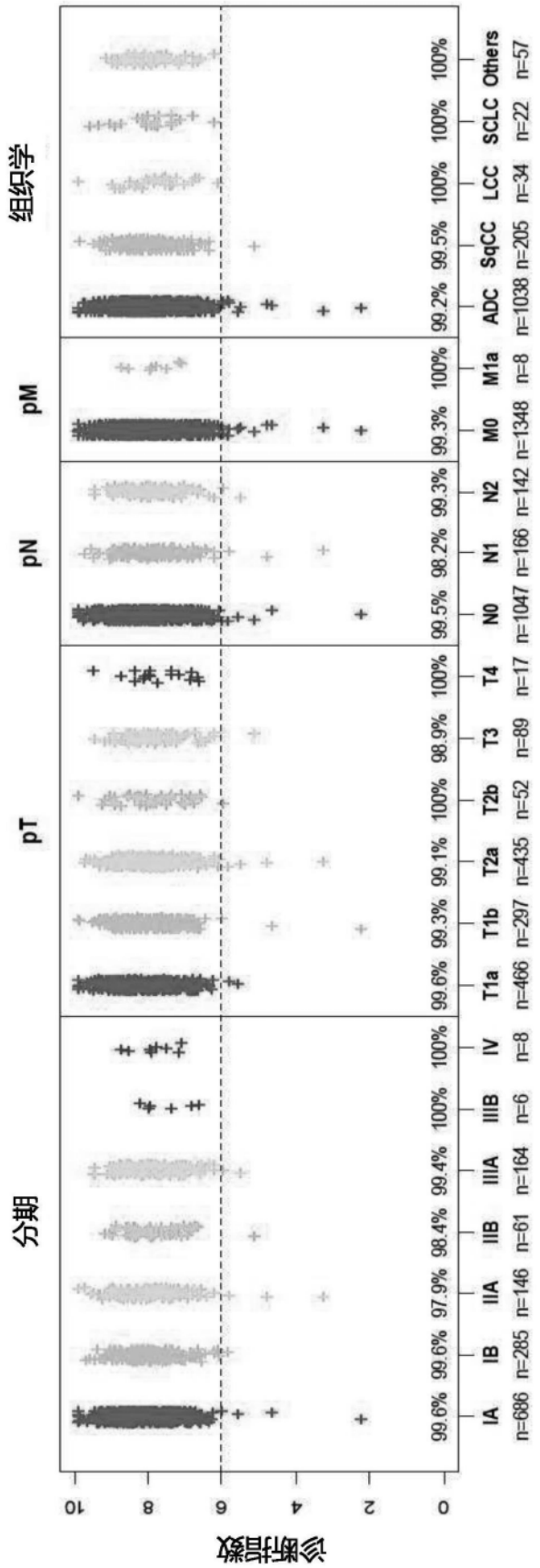

诊断指数
0
2
4
6
8
10
分期
IA IB IIA IIB IIIA IIIB IV
n=686 n=285 n=146 n=61 n=164 n=6 n=8
99.6% 99.6% 97.9% 98.4% 99.4% 100% 100%
pT
T1a T1b T2a T2b T3 T4
n=466 n=297 n=435 n=52 n=89 n=17
99.6% 99.3% 99.1% 100% 98.9% 100%
pN
N0 N1 N2
n=1047 n=166 n=142
99.5% 98.2% 99.3%
pM
M0 M1a
n=1348 n=8
99.3% 100%
组织学
ADC SqCC LCC SCLC Others
n=1038 n=205 n=34 n=22 n=57
99.2% 99.5% 100% 100% 100%

图.2G

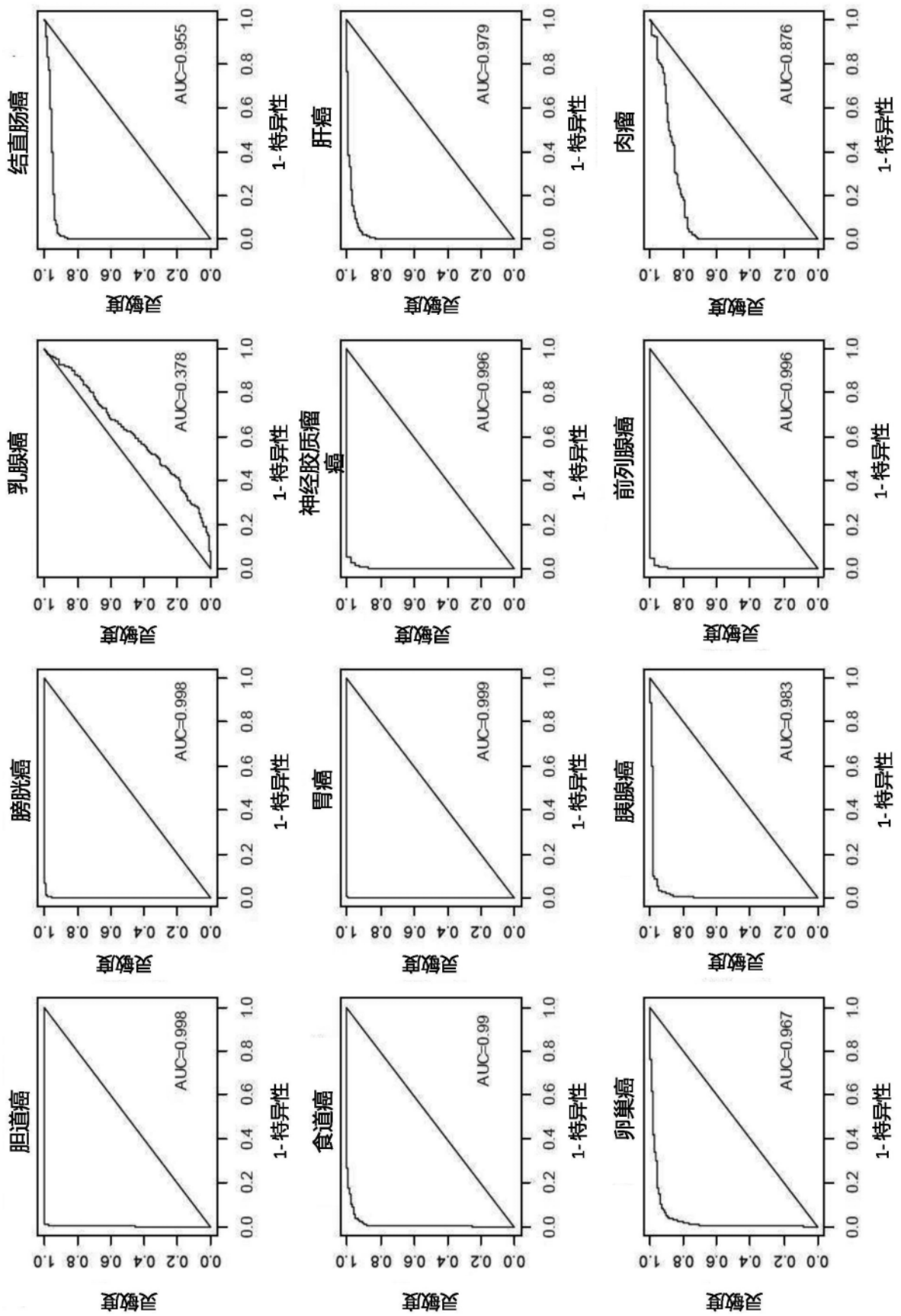

结直肠癌
AUC=0.955
肝癌
AUC=0.979
肉瘤
AUC=0.876
乳腺癌
AUC=0.378
神经胶质瘤
AUC=0.996
前列腺癌
AUC=0.996
膀胱癌
AUC=0.998
胃癌
AUC=0.999
胰腺癌
AUC=0.983
胆道癌
AUC=0.998
食道癌
AUC=0.99
卵巢癌
AUC=0.967
灵敏度
1-特异性
0.0 0.2 0.4 0.6 0.8 1.0

图.3A

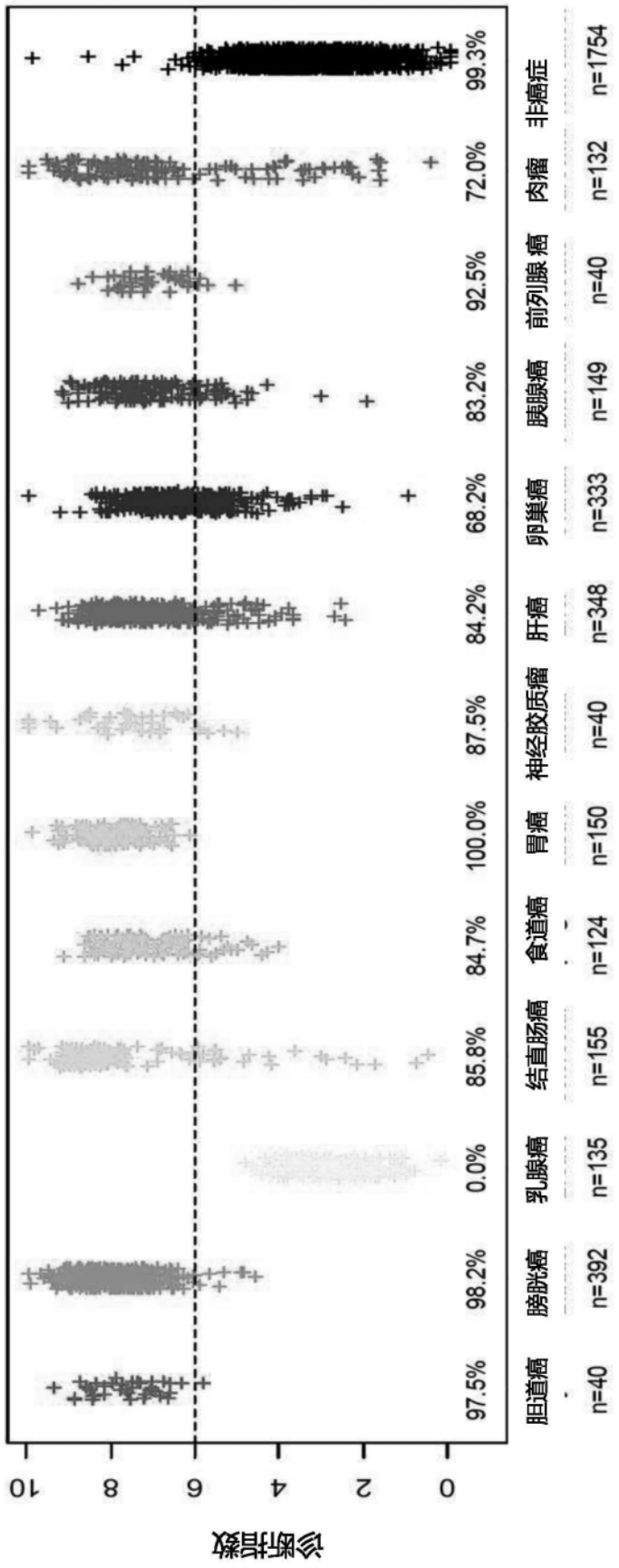

诊断指数
0
2
4
6
8
10
胆道癌 膀胱癌 乳腺癌 结直肠癌 食道癌 胃癌 神经胶质瘤 肝癌 卵巢癌 胰腺癌 前列腺癌 肉瘤 非癌症
n=40 n=392 n=135 n=155 n=124 n=150 n=40 n=348 n=333 n=149 n=40 n=132 n=1754
97.5% 98.2% 0.0% 85.8% 84.7% 100.0% 87.5% 84.2% 68.2% 83.2% 92.5% 72.0% 99.3%

图.3B

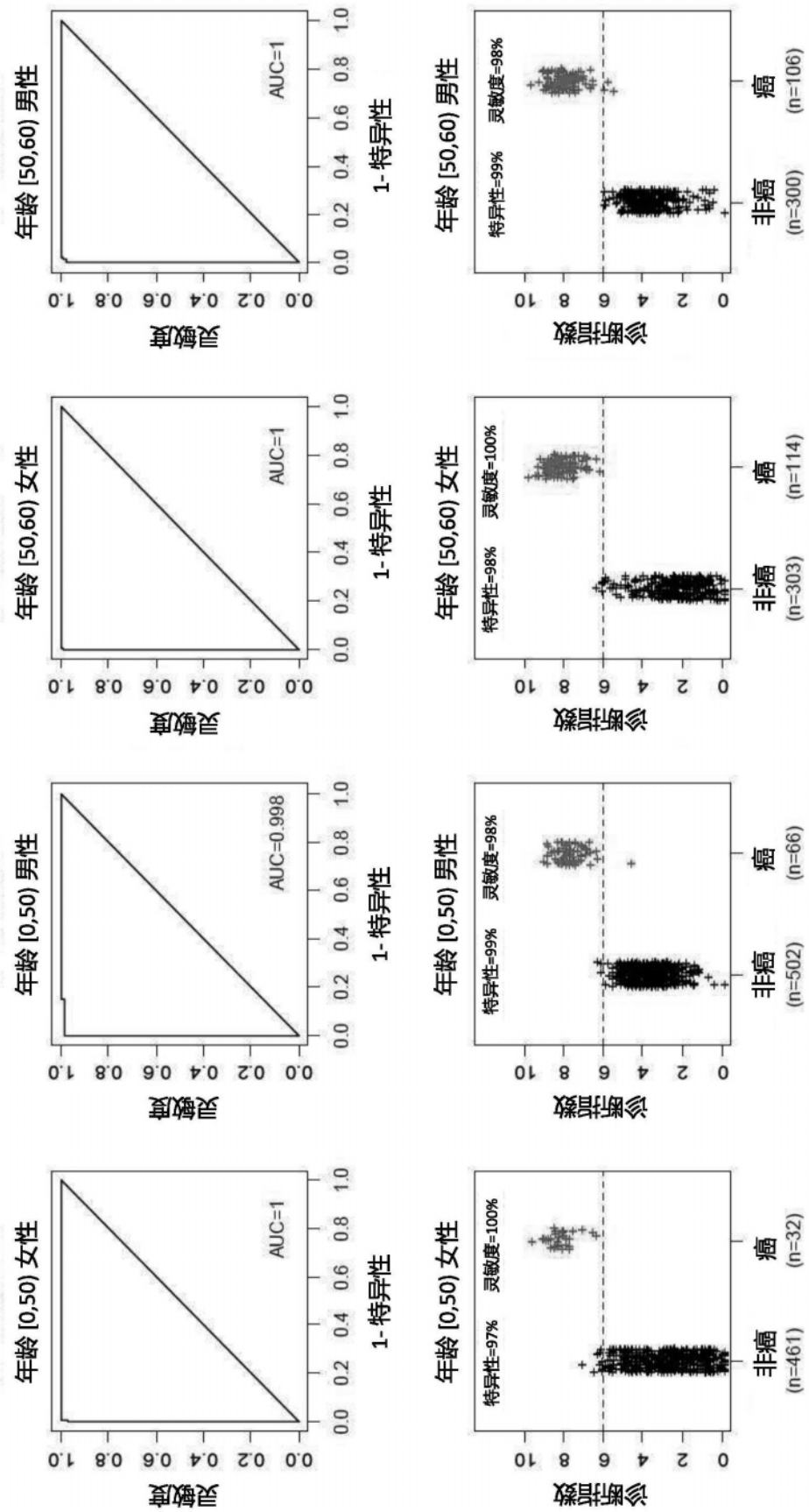

年龄 [0,50) 女性
AUC=1
灵敏度
1- 特异性
年龄 [0,50) 男性
AUC=0.998
灵敏度
1- 特异性
年龄 [50,60) 女性
AUC=1
灵敏度
1- 特异性
年龄 [50,60) 男性
AUC=1
灵敏度
1- 特异性
年龄 [0,50) 女性
特异性=97% 灵敏度=100%
诊断指数
非癌 (n=461)
癌 (n=32)
年龄 [0,50) 男性
特异性=99% 灵敏度=98%
诊断指数
非癌 (n=502)
癌 (n=66)
年龄 [50,60) 女性
特异性=98% 灵敏度=100%
诊断指数
非癌 (n=303)
癌 (n=114)
年龄 [50,60) 男性
特异性=99% 灵敏度=98%
诊断指数
非癌 (n=300)
癌 (n=106)

图.4A

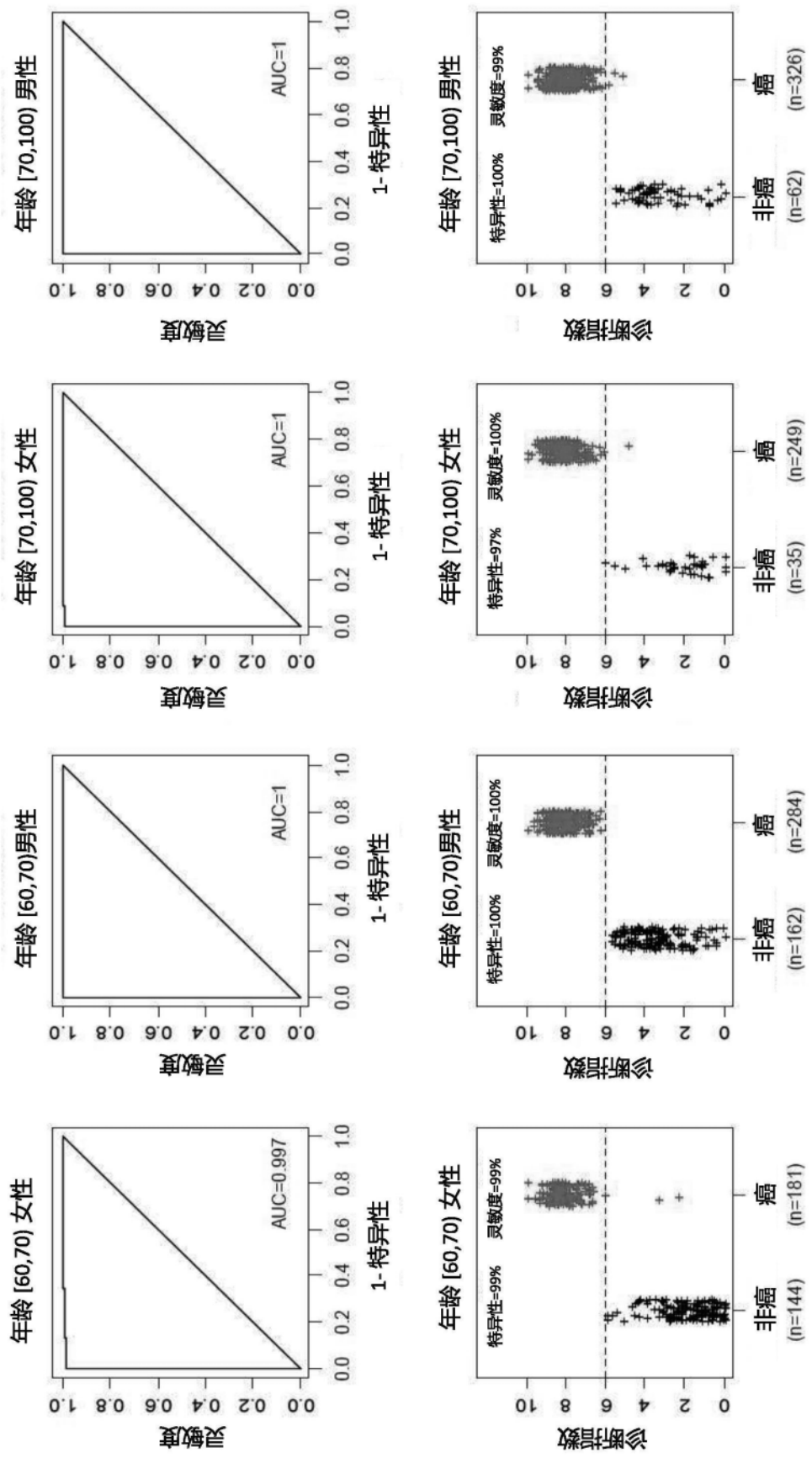

年龄 [70,100) 男性
AUC=1
灵敏度
1- 特异性
年龄 [70,100) 男性
特异性=100% 灵敏度=99%
诊断指数
非癌 (n=62)
癌 (n=326)
年龄 [70,100) 女性
AUC=1
灵敏度
1- 特异性
年龄 [70,100) 女性
特异性=97% 灵敏度=100%
诊断指数
非癌 (n=35)
癌 (n=249)
年龄 [60,70)男性
AUC=1
灵敏度
1- 特异性
年龄 [60,70)男性
特异性=100% 灵敏度=100%
诊断指数
非癌 (n=162)
癌 (n=284)
年龄 [60,70) 女性
AUC=0.997
灵敏度
1- 特异性
年龄 [60,70) 女性
特异性=99% 灵敏度=99%
诊断指数
非癌 (n=144)
癌 (n=181)

图.4B

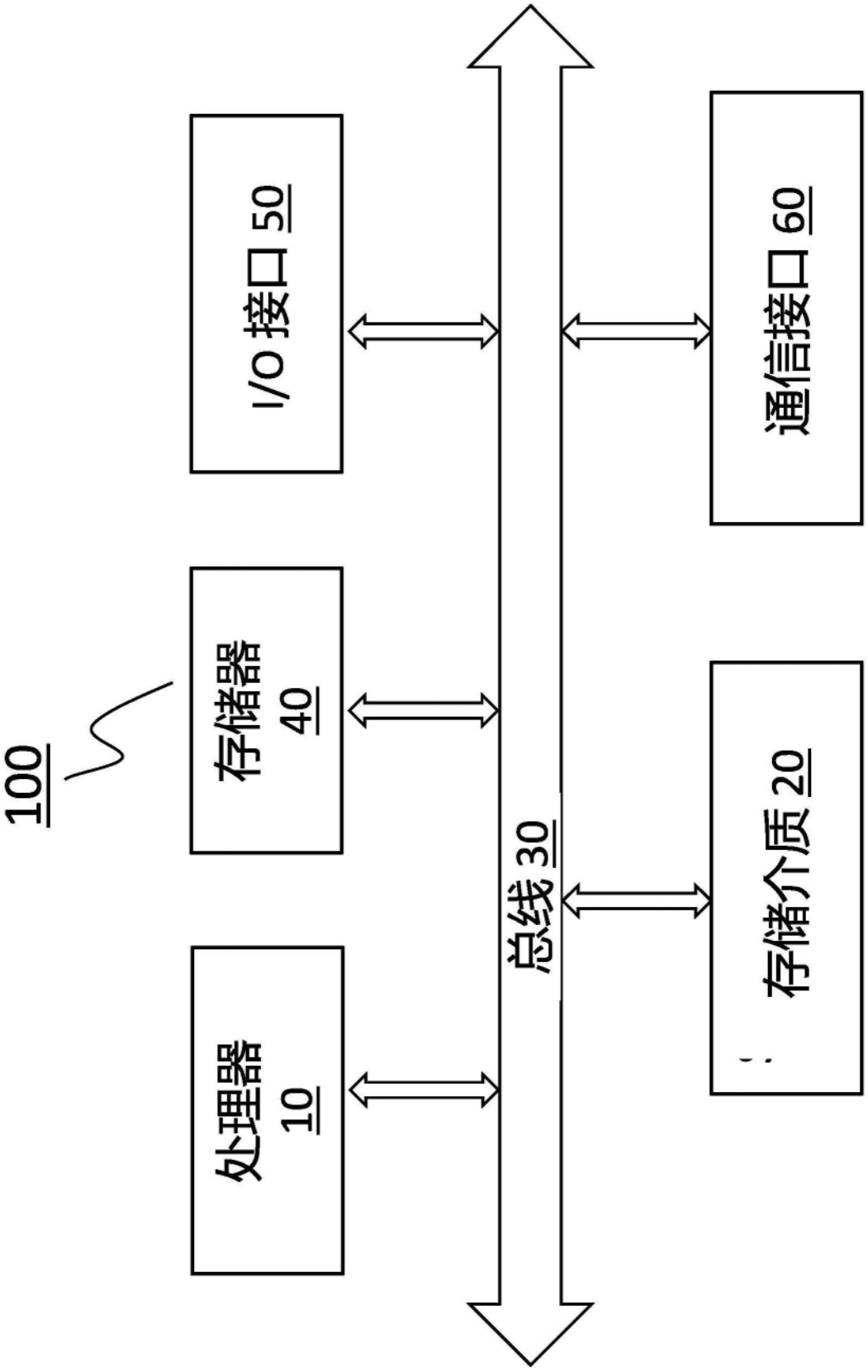


图.5