



(12)发明专利

(10)授权公告号 CN 104937111 B

(45)授权公告日 2018.05.11

(21)申请号 201380071054.0

(22)申请日 2013.11.26

(65)同一申请的已公布的文献号
申请公布号 CN 104937111 A

(43)申请公布日 2015.09.23

(30)优先权数据
61/730,391 2012.11.27 US
61/775,419 2013.03.08 US

(85)PCT国际申请进入国家阶段日
2015.07.21

(86)PCT国际申请的申请数据
PCT/US2013/071970 2013.11.26

(87)PCT国际申请的公布数据
W02014/085434 EN 2014.06.05

(73)专利权人 智利天主教教皇大学

地址 智利圣地亚哥

(72)发明人 埃尔南·埃乌杰尼奥·冈萨雷斯迪
亚兹
瑟吉奥·瓦尔加斯塞拉斯
约瑟·罗德里戈·沃尔德马尔·马
丁内斯索利斯

(74)专利代理机构 北京英赛嘉华知识产权代理
有限责任公司 11204

代理人 王达佐 洪欣

(51)Int.Cl.
C12Q 1/6886(2018.01)

审查员 张蕊

权利要求书3页 说明书41页
序列表30页 附图10页

(54)发明名称

用于诊断甲状腺肿瘤的组合物和方法

(57)摘要

本发明提供了用于确定生物样本中的甲状腺癌的诊断测定法,包括细针抽吸,以及可用于实施本发明方法的相关组合物和试剂盒。

1. 试剂在制备用于诊断个体的甲状腺癌的试剂盒中的用途,其中所述试剂确定获自所述个体的甲状腺组织样本的不同基因产物的表达水平,其中所述基因产物由以下组成:由CXCR3、CCR3、CXCL10、CK-19、TIMP-1、CLDN-1、CAR、HO-1、XB130和CCR7基因表达的基因产物,并且通过将所确定的表达水平与所述甲状腺组织样本中有或无甲状腺癌相关联来鉴定所述甲状腺组织样本为癌性或良性的。

2. 如权利要求1所述的用途,其中通过将所述基因产物的表达水平与每一种基因产物的正常对照表达水平相比较来进行所述的关联,其中如果所述基因产物的表达水平在所述甲状腺组织样本与所述正常对照表达水平之间存在差异,则将所述甲状腺组织样本鉴定为癌性的。

3. 如权利要求2所述的用途,其中如果所述基因产物中的四种或更多种的表达水平在所述甲状腺组织样本与所述正常对照表达水平之间存在差异,则将所述甲状腺组织样本鉴定为癌性的。

4. 如权利要求2所述的用途,其中所述正常对照表达水平为正常甲状腺组织样本中的表达水平。

5. 如权利要求2所述的用途,其中所述正常对照表达水平为基于多个正常甲状腺组织样本中的表达水平的预定值。

6. 如权利要求2所述的用途,其中如果相比所述正常对照表达水平,所述甲状腺组织样本中的CXCR3、XB130和/或CAR基因中的任何一种或多种的表达水平降低;和/或CCR3、CXCL10、CK-19、TIMP-1、CLDN-1、HO-1和/或CCR7基因中的任一种或多种的表达水平增加,则将所述甲状腺组织样本鉴定为癌性的,其中具有增加或降低表达的基因总数目至少为3种。

7. 如权利要求1所述的用途,其中通过将所述基因产物的表达水平与每一种基因产物的癌症对照表达水平相比较来进行所述的关联,其中如果所述基因产物中的三种或更多种的表达水平在所述甲状腺组织样本与所述癌症对照表达水平之间基本上没有差异,则将所述甲状腺组织样本鉴定为癌性的。

8. 如权利要求7所述的用途,其中如果所述基因产物中的四种或更多种的表达水平在所述甲状腺组织样本与所述癌症对照表达水平之间基本上没有差异,则将所述甲状腺组织样本鉴定为癌性的。

9. 如权利要求7所述的用途,其中所述癌症对照表达水平为癌性甲状腺组织样本中的表达水平。

10. 如权利要求7所述的用途,其中所述癌症对照表达水平为基于多个癌性甲状腺组织样本中的表达水平的预定值。

11. 如权利要求1所述的用途,其中所述的关联包括将所述表达水平与在以下两组生物样本中确定的所述基因产物的基因表达数据相比较:

(i) 多个正常甲状腺组织样本;和

(ii) 多个癌性甲状腺组织样本,

其中如果所述基因产物中的三种或更多种的表达水平在所述甲状腺组织样本与(i)中所述的基因表达数据之间存在差异,或者如果所述基因产物中的一种或多种的基因表达水平在所述甲状腺组织样本与(ii)中所述的基因表达数据之间基本上没有差异,则将所述甲状腺组织样本鉴定为癌性的。

12. 如权利要求1至11中任一项所述的用途,其中使用鉴定不典型CT值和/或非不典型CT值的分类器进行所述的关联。

13. 如权利要求12所述的用途,其中所述分类器使用从多个正常甲状腺组织样本和/或癌性甲状腺组织样本确定的所述基因产物的基因表达数据来生成。

14. 如权利要求1至11中任一项所述的用途,其中所述基因产物为RNA。

15. 如权利要求14所述的用途,其中所述RNA为mRNA。

16. 如权利要求14所述的用途,其中所述RNA表达水平使用微阵列、SAGE、印迹、RT-PCR、定量PCR、分子条形码技术、TRAC、PCR阵列、多重qPCR或qNPA来确定。

17. 如权利要求1至11中任一项所述的用途,其中所述基因产物为蛋白。

18. 如权利要求17所述的用途,其中所述蛋白基因表达使用ELISA、质谱分析、印迹、蛋白质组技术或免疫组化来确定。

19. 如权利要求1至11中任一项所述的用途,其还包括以下步骤:在利用所述试剂盒确定获自所述个体的甲状腺组织样本的不同基因产物的表达水平之前对获自所述个体的甲状腺组织样本进行细胞学分析,以获得初步诊断。

20. 如权利要求19所述的用途,其中利用所述试剂盒对具有中间或不确定的初步诊断的样本进一步分析。

21. 如权利要求20所述的用途,其中经细胞学分析的组织样本和利用所述试剂盒分析中使用的组织样本为同一组织样本。

22. 如权利要求1至11中任一项所述的用途,其中所述甲状腺组织样本以大于或等于95%的灵敏度被诊断为癌性或良性的。

23. 如权利要求1至11中任一项所述的用途,其中所述甲状腺组织样本以大于或等于80%的特异性被诊断为癌性或良性的。

24. 如权利要求1至11中任一项所述的用途,其中所述甲状腺组织样本以大于或等于70%的阳性预测值被诊断为癌性或良性的。

25. 如权利要求1至11中任一项所述的用途,其中所述甲状腺组织样本以大于或等于92%的阴性预测值被诊断为癌性或良性的。

26. 如权利要求1至11中任一项所述的用途,其中所述甲状腺组织样本以大于或等于2的阳性似然比被诊断为癌性或良性的。

27. 如权利要求1至11中任一项所述的用途,其中所述甲状腺组织样本以大于或等于50%的阳性验后概率被诊断为癌性或良性的。

28. 如权利要求1至11中任一项所述的用途,其中所述甲状腺组织样本以小于或等于0.08的阴性似然比被诊断为癌性或良性的。

29. 如权利要求1至11中任一项所述的用途,其中所述甲状腺组织样本以小于或等于7.0%的阴性验后概率被诊断为癌性或良性的。

30. 如权利要求1至11中任一项所述的用途,其还包括从所述个体获取所述甲状腺组织样本。

31. 如权利要求1至11中任一项所述的用途,如果所述甲状腺组织样本被诊断为癌性的,其还包括经手术去除所述个体的甲状腺或其一部分。

32. 诊断甲状腺癌的试剂盒,所述试剂盒包含用于检测不同基因产物的试剂,其中所述

试剂检测基因产物,所述基因产物包含由CXCR3、CCR3、CXCL10、CK-19、TIMP-1、CLDN-1、CAR、HO-1、XB130和CCR7基因表达的基因产物。

33.如权利要求32所述的试剂盒,其中所述试剂为抗体,并且其中每一种抗体特异性结合多肽基因产物。

34.如权利要求32所述的试剂盒,其中所述试剂为寡核苷酸或寡核苷酸组,并且其中每一种寡核苷酸特异性结合核酸基因产物。

35.如权利要求32所述的试剂盒,其中所述试剂各自连接至基质。

36.如权利要求35所述的试剂盒,其中所述试剂共价连接至所述基质。

37.如权利要求35所述的试剂盒,其中所述试剂各自连接至固相基质的不同区域。

38.如权利要求32所述的试剂盒,其中所述试剂为共价结合至固相基质的寡核苷酸或寡核苷酸组。

39.如权利要求38所述的试剂盒,其中所述固相基质为阵列。

40.如权利要求39所述的试剂盒,其中所述阵列为微阵列。

41.如权利要求32所述的试剂盒,其中所述试剂为寡核苷酸组,并且其中所述寡核苷酸组包含DNA。

42.如权利要求32所述的试剂盒,其中所述试剂为寡核苷酸组,并且其中每一组寡核苷酸与所述基因产物中的一种特异性杂交。

43.如权利要求42所述的试剂盒,其中每一组寡核苷酸包含能够PCR扩增所述基因产物中的一种的扩增引物。

44.如权利要求32至43中任一项所述的试剂盒,其中所述试剂被标记。

45.如权利要求32至43中任一项所述的试剂盒,其中所述试剂盒还包含适合使所述试剂与所述基因产物相结合的一种或多种溶液。

46.如权利要求44所述的试剂盒,其中所述试剂为寡核苷酸组,并且所述试剂盒还包含用于进行PCR测定的一种或多种其他试剂。

47.如权利要求46所述的试剂盒,其中所述一种或多种其他试剂选自热稳定的聚合酶、脱氧核苷酸混合物和可检测标记的探针。

48.如权利要求47所述的试剂盒,其中所述可检测标记的探针包含荧光团和淬灭部分。

49.如权利要求47所述的试剂盒,其中所述可检测标记的探针在所述探针被切割时而不是在所述探针完整时发射可检测的信号。

50.如权利要求32至43中任一项所述的试剂盒,其中所述试剂盒还包含用于处理甲状腺组织样本的一种或多种试剂。

51.如权利要求50所述的试剂盒,其中所述甲状腺组织样本的处理包括从所述甲状腺组织样本提取所述基因产物。

52.如权利要求51所述的试剂盒,其中所述基因产物为蛋白或核酸。

53.如权利要求32至43中任一项所述的试剂盒,其还包含一种或多种对照基因产物。

用于诊断甲状腺肿瘤的组合物和方法

[0001] 相关申请的交叉引用

[0002] 本申请要求于2012年11月27日递交的美国临时专利申请系列第61/730,391号和于2013年3月8日递交的美国临时专利申请系列第61/775,419号的优先权。这些申请通过引用整体并入本文。

[0003] 序列表

[0004] 与本申请相关的序列表以文本格式代替纸件提供,并通过引用并入说明书中。包含序列表的文本文件的名称为BMRC_001_02W0_ST25.txt。该文本文件约57KB,于2013年11月26日生成,并通过EFS-Web电子递交。

[0005] 发明背景

技术领域

[0006] 本发明涉及用于诊断甲状腺癌以及评估甲状腺结节以确定其是否为良性或癌性的组合物和方法。

[0007] 相关领域的描述

[0008] 在美国每年进行约350,000份甲状腺结节的细针抽吸(FNA)活组织检查,其中有20%被报道为不确定结节是否为癌性的。在大多数情况下,这些患者因癌症风险接受了手术,其范围仅为15%至30%。这意味着大多数患者不需要手术去除甲状腺。考虑到与甲状腺手术相关的急性和长期风险,以及对患者和医疗体系的成本,迫切需要能够提高甲状腺FNA活组织检查的诊断准确度的工具。

[0009] 最近,已有新的检验投入了美国市场,其在不同程度上改善了甲状腺结节的不确定诊断。这些检验包括 Afirma® 甲状腺FNA分析检验(Veracyte, South San Francisco, CA),其为基于167种基因的基因表达分类器测定法。然而, Afirma® 检验在仅约50%的情况下正确改变手术实施。其他两种检验包括Quest Diagnostics和Asuragen开发的那些,其基于美国甲状腺协会接受的已知生物标志物的突变分析。然而,缺少充足的临床试验验证。不幸的是, Afirma® 检验需要分析大量的生物标志物,所有检验必须在中心实验室进行,因而需要运送用于分析的样本。另外,必须进行第二次FNA以获得足够的样本来进行这些测定。

[0010] 显然,本领域需要改善用于评估甲状腺结节和诊断甲状腺癌的方法和组合物。本发明通过提供用于新且简化的诊断方法而满足了这一需求并提供了其他优势,所述诊断方法用于评估已报导为通过细针抽吸(FNA)活组织检查为不确定的甲状腺结节。

[0011] 发明概述

[0012] 本发明提供了用于确定样本中有或无恶性或良性组织的组合物、方法和试剂盒,所述样本如甲状腺组织或甲状腺结节样本。

[0013] 在某些实施方案中,本发明提供了诊断个体的甲状腺癌的方法,其包括:确定获自所述个体的甲状腺组织样本中的三种或更多种基因产物的表达水平,其中所述三种或更多种基因产物由表1中列出的一种或多种基因表达,并且其中所述基因产物中的至少一种由

CXCR3、CCR3、CXCL10、CAR、XB130、HO-1或CCR7基因表达；以及通过将所确定的表达水平与甲状腺组织样本中有或无甲状腺癌相关联而鉴定所述甲状腺组织样本为癌性或良性的。在不同的实施方案中，将所述方法用于确定所述组织样本是否为癌性或良性的，例如，用于确定所述个体是否患有癌症，例如甲状腺癌。

[0014] 在本发明方法的某些实施方案中，所述关联步骤通过将所述三种或更多种基因产物的表达水平与所述基因产物中的每种的正常对照表达水平相比较而进行，其中如果所述三种或更多种基因产物的表达水平在所述甲状腺组织样本与所述正常对照表达水平之间存在差异，则将所述甲状腺组织样本鉴定为癌性的。在相关的实施方案中，如果四种或更多种基因产物的表达水平在所述甲状腺组织样本与所述正常对照表达水平之间存在差异，则将所述甲状腺组织样本鉴定为癌性的。在某些实施方案中，所述正常对照表达水平为正常甲状腺组织样本中的表达水平，然而在某些实施方案中，所述正常对照表达水平为基于多个正常甲状腺组织样本中的表达水平的预定值。在本发明方法的某些实施方案中，相比正常对照表达水平，如果甲状腺组织样本中CXCR3、CXCL11、SPAG-9、CAR、Nectin-1(粘连蛋白-1)、XB-130和/或CXCL4基因中的任何一种或多种的表达水平降低；和/或CXCR3A、CXCR3B、CXCR4、CCR3、CXCL9、CXCL10、CK-19、TIMP-1、CLDN-1和/或CCR7基因中的任何一种或多种的表达水平增加，则将所述甲状腺组织样本鉴定为良性或癌性的，其中具有增加或降低表达的基因总数目至少为3种。

[0015] 在本发明方法的某些实施方案中，所述关联步骤通过将所述三种或更多种基因产物的表达水平与每一种基因产物的癌症对照表达水平相比较而进行，其中如果所述三种或更多种基因产物的表达水平在所述甲状腺组织样本与所述癌症对照表达水平之间基本上没有差异，则将所述甲状腺组织样本鉴定为癌性的。在特定实施方案中，如果四种或更多种基因产物的表达水平在所述甲状腺组织样本与所述癌症对照表达水平之间基本上没有差异，则将所述甲状腺组织样本鉴定为癌性的。在某些实施方案中，所述癌症对照表达水平为癌性甲状腺组织样本中的表达水平。在特定实施方案中，所述癌症对照表达水平为基于多个癌性甲状腺组织样本中的表达水平的预定值。

[0016] 在本发明方法的某些实施方案中，所述关联步骤包括比较所述表达水平与在以下两组生物样本中确定的所述三种或更多种基因产物的基因表达数据：(i) 多个正常甲状腺组织样本；和(ii) 多个癌性甲状腺组织样本，其中如果所述三种或更多种基因产物的表达水平在所述甲状腺组织样本与(i)中所述基因表达数据之间存在差异，或者如果所述一种或多种基因产物的基因表达水平在所述甲状腺组织样本与(ii)中所述基因表达数据之间基本上没有差异，则将所述甲状腺组织样本鉴定为癌性的。

[0017] 在本发明方法的特定实施方案中，所述关联步骤使用鉴定不典型CT值和/或非不典型CT值的分类器进行。在某些实施方案中，所述分类器使用从多个正常甲状腺组织样本和/或癌性甲状腺组织样本确定的所述三种或更多种基因产物的基因表达数据生成。

[0018] 在本发明方法的特定实施方案中，所述甲状腺组织样本通过针抽吸、细针抽吸、芯针吸活组织检查、真空辅助的活组织检查、大芯针吸活组织检查、切开活组织检查、切除活组织检查或钻取活组织检查来获取。

[0019] 在本发明方法的不同实施方案中，所述基因产物为RNA，例如mRNA、rRNA、tRNA或miRNA。在某些实施方案中，所述RNA表达水平使用微阵列、SAGE、印迹、RT-PCR、定量PCR或

qNPA来确定。在本发明的不同实施方案中,所述基因产物为蛋白。在某些实施方案中,所述蛋白基因表达使用ELISA、质谱分析、印迹、蛋白质组技术或免疫组化来确定。

[0020] 在本发明的方法、组合物和试剂盒的不同实施方案中,所述三种或更多种基因产物包含以下或由以下组成: CXCR3、CCR3、CXCL10、CK19、TIMP-1、CLDN-1、CAR、XB130、HO-1和CCR7基因中的三种或更多种基因产物; CXCR3、CCR3、CXCL10、CK19、TIMP-1、CLDN-1、CAR、HO-1和CCR7基因的基因产物; CCR3、TIMP-1、CAR和XB130基因的基因产物; CXCL10、TIMP-1、CAR和CCR7基因的基因产物; TIMP-1、CAR和CCR7基因的基因产物; 或者CXCL10、TIMP-1、CLDN-1和CCR7基因的基因产物。在与任何这些基因组相关的方法、组合物和试剂盒的特定实施方案中, CXCR3、CXCL11、SPAG-9、CAR、Nectin-1、XB-130和/或CXCL4基因的表达水平降低; 和/或CXCR3A、CXCR3B、CXCR4、CCR3、CXCL9、CXCL10、CK-19、TIMP-1、CLDN-1、HO-1和/或CCR7基因的表达水平增加。因此,在特定实施方案中,所述三种或更多种基因产物包含以下或由以下组成: CXCR3、CCR3、CXCL10、CK19、TIMP-1、CLDN-1、CAR、XB130、HO-1和CCR7基因的基因产物,其中CXCR3、CAR和XB130基因中的一种或多种、两种或更多种或者所有的表达水平降低,和/或CCR3、CXCL10、CK19、TIMP-1、CLDN-1、HO-1和CCR7基因中的一种或多种、两种或更多种、三种或更多种、四种或更多种、五种或更多种、六种或更多种或者所有的表达水平增加。在其他特定实施方案中,所述三种或更多种基因产物包含以下或由以下组成: CXCR3、CCR3、CXCL10、CK19、TIMP-1、CLDN-1、CAR、HO-1和CCR7基因的基因产物,其中CXCR3和CAR基因中的一种或两种的表达水平降低,和/或CCR3、CXCL10、CK19、TIMP-1、CLDN-1、HO-1和CCR7基因中的一种或多种、两种或更多种、三种或更多种、四种或更多种、五种或更多种、六种或更多种或者所有的表达水平增加。在其他特定实施方案中,所述三种或更多种基因产物包含以下或由以下组成: CCR3、TIMP-1、CAR和XB130基因的基因产物,其中CAR和XB130基因中的一种或两种或二者的表达水平降低,和/或CCR3和TIMP-1基因中的一种或多种或者二者的表达水平增加。在其他特定实施方案中,所述三种或更多种基因产物包含以下或由以下组成: CXCL10、TIMP-1、CAR和CCR7基因的基因产物,其中CAR基因的表达水平降低,和/或CXCL10、TIMP-1和CCR7基因中的一种或多种、两种或更多种或者所有的表达水平增加。在其他特定实施方案中,所述三种或更多种基因产物包含以下或由以下组成: TIMP-1、CAR和CCR7基因的基因产物,其中CAR基因的表达水平降低,和/或TIMP-1和CCR7基因中的一种或多种或者二者的表达水平增加。在其他特定实施方案中,所述三种或更多种基因产物包含以下或由以下组成: CXCL10、TIMP-1、CLDN-1和CCR7基因,其中CXCL10、TIMP-1、CLDN-1和CCR7基因中的一种或多种、两种或更多种、三种或更多种或者所有的表达水平增加。

[0021] 在某些实施方案中,本发明的方法还包括以下步骤: 对获自个体的甲状腺组织样本进行细胞学分析以获得初步诊断。在特定实施方案中,根据本发明的方法通过确定基因产物表达水平并将它们与良性或恶性组织相关联,而进一步分析中间或不确定初步诊断的样本。在特定实施方案中,经细胞学分析的组织样本与用于确定基因产物表达水平的组织样本为同一组织样本。在本发明方法的某些实施方案中,组织样本通过细针抽吸获取。

[0022] 在本发明的方法、组合物或试剂盒的特定实施方案中,将甲状腺组织样本以大于或等于92%或者大于或等于97%的灵敏度诊断为癌性或良性的,将甲状腺组织样本以大于或等于60%或者大于或等于90%的特异性诊断为癌性或良性的,将甲状腺组织样本以大于或等于50%或者大于或等于90%的阳性预测值诊断为癌性或良性的,将甲状腺组织样本以

大于或等于92%或者大于或等于94%的阴性预测值诊断为癌性或良性的,将甲状腺组织样本以大于或等于2或者大于或等于10的阳性似然比诊断为癌性或良性的,将甲状腺组织样本以大于或等于50%或者大于或等于80%的阳性验后概率诊断为癌性或良性的,将甲状腺组织样本以小于或等于0.14或者小于或等于0.08的阴性似然比诊断为癌性或良性的,和/或将甲状腺组织样本以小于或等于7.0%或者小于或等于3.0%的阴性验后概率诊断为癌性或良性的。

[0023] 本发明方法的特定实施方案还包括从个体获取甲状腺组织样本。

[0024] 如果甲状腺组织样本被诊断为癌性的,则本发明方法的特定实施方案还包括经手术去除个体的甲状腺或者其一部分。

[0025] 在相关实施方案中,本发明包括用于诊断甲状腺癌的试剂盒,所述试剂盒包含三种或更多种用于检测基因产物的试剂,其中所述三种或更多种试剂中的每一种检测不同的基因产物,其中所述基因产物由表1中列出的一种或多种基因表达,并且其中所述基因产物中的至少一种由CXCR3、CCR3、CXCL10、CAR、XB130、HO-1或CCR7基因表达。在某些实施方案中,试剂为抗体,并且每一种抗体特异性结合多肽基因产物。在某些实施方案中,试剂为寡核苷酸或寡核苷酸组,并且每一种寡核苷酸特异性结合核酸基因产物。在某些实施方案中,试剂各自与基质连接。在特定实施方案中,试剂与基质共价连接。在特定实施方案中,试剂中的每一种与固相基质的不同区域连接。在特定实施方案中,试剂为与固相基质共价结合的寡核苷酸或寡核苷酸组,所述固相基质任选为阵列,并且所述阵列任选为微阵列。在特定实施方案中,试剂为寡核苷酸组,并且所述寡核苷酸组包含DNA。在特定实施方案中,试剂为寡核苷酸组,并且每个寡核苷酸组与基因产物中的一种特异性杂交。在一实施方案中,每组寡核苷酸包含能够PCR扩增一种基因产物的扩增引物。在本发明不同试剂盒的某些实施方案中,基因产物包含以下或由以下组成: CXCR3、CCR3、CXCL10、CK19、TIMP-1、CLDN-1、CAR、XB130、HO-1和CCR7基因中的三种或更多种基因产物; CXCR3、CCR3、CXCL10、CK19、TIMP-1、CLDN-1、CAR、HO-1和CCR7基因的基因产物; CCR3、TIMP-1、CAR和XB130基因的基因产物; CXCL10、TIMP-1、CAR和CCR7基因的基因产物; TIMP-1、CAR和CCR7基因的基因产物; 或者CXCL10、TIMP-1、CLDN-1和CCR7基因的基因产物。

[0026] 在本发明方法、组合物和试剂盒的某些实施方案中,试剂经过标记。在特定实施方案中,本发明的试剂盒还包含一种或多种适合将所述试剂与所述基因产物结合的溶液。在本发明试剂盒的某些实施方案中,试剂为寡核苷酸组,并且所述试剂盒还包含用于进行PCR测定的一种或多种另外的试剂。在特定实施方案中,一种或多种另外的试剂选自热稳定的聚合酶、脱氧核苷酸混合物以及可检测地标记探针。在某些实施方案中,可检测地标记探针包含荧光团和淬灭部分。在特定实施方案中,当所述探针被切割时而不是在所述探针完整时,可检测地标记探针发射可检测的信号。

[0027] 在本发明试剂盒的不同实施方案中,所述试剂盒还包含用于处理甲状腺组织样本的一种或多种试剂。在特定实施方案中,甲状腺组织样本的处理包括从所述甲状腺组织样本提取基因产物,并且在某些实施方案中,基因产物为蛋白或核酸。

[0028] 在不同实施方案中,本发明的试剂盒还包含一种或多种对照基因产物。

[0029] 在本发明试剂盒的特定实施方案中,(存在于试剂盒中时)以下一种或多种存在于单独的容器中:用于检测基因产物的试剂、溶液、任何其他试剂和对照基因产物。

[0030] 在其他相关实施方案中,本发明提供了治疗甲状腺癌的方法,包括:根据本发明的方法或使用本发明的试剂盒来鉴定获自个体的甲状腺组织样本为癌性的;以及经手术去除个体的甲状腺或其一部分,或者对个体进行放疗、化疗或激素治疗。

[0031] 附图的几个视图的简要说明

[0032] 图1提供的图示出了癌症样本(100)和良性结节(56)之间18种基因的差异表达,其通过定量实时PCR确定并通过Pfaffl提议的相对定量模型分析。值0对应于良性结节的基因表达,且柱对应于癌症样本中的基因表达变化(相对倍数变化)。实施例1中鉴定了18种基因。

[0033] 图2提供了实施例1中鉴定的18种单个基因中的每一种的受试者工作特征(ROC)曲线数据。AUC:曲线下面积,FP:假阳性,TP:真阳性。

[0034] 图3提供了用于生成(3A)和使用不同(3B)算法来开发新的分类器的方法的示意图。训练了两种不同的算法;一种为鉴定和分类不典型(离群)CT值,而一种用于分类非不典型CT值(线性和非线性判别分析)(3A)。通过在两个步骤中依序使用所述算法,然后整合开发新分类器(3B)的输出数据,而开发出了新的分类器。

[0035] 图4提供了通过鉴定和分类不典型CT值然后进行线性判别分析(LDA)或非线性判别分析(NLDA)而开发的新分类器的ROC曲线图。AUC:曲线下面积,FP:假阳性,TP:真阳性。

[0036] 图5提供了根据图3中描述的方法开发的新分类器(SV)与最佳单个基因分类器和它们的组合(基因10、11、12)的AUC比较。*对应于 p 值 <0.05 ,显示SV分类器显著优于单个基因或它们的组合。

[0037] 图6提供了最佳单个分类基因的Spearman相关分析,显示它们紧密相关,解释了它们的组合为何不能改善它们作为基因分类器的性能。

[0038] 图7提供了利用训练组(SV、FM72、FM208)、最佳单个基因(基因10、基因11、基因12)、最佳基因的组合(基因(10、11、12))开发的三种新分类器与Veracyte(Affirma)的Affirma®分类器的分类性能的比较表。

[0039] 图8提供了比较使用SV分类器的独立测试组与训练组的性能的ROC曲线图和数据。

[0040] 图9提供的比较表示出了使用SV算法对手术和FNA样本获得的灵敏度、特异性、PPV和NPV值。

[0041] 发明详述

[0042] 本发明部分是基于鉴定一小组基因及其不同亚组,其允许精确分类甲状腺样本为恶性或良性的。相比使用单个基因或先前描述的基因组合,根据本发明使用的基因组合导致令人吃惊的改善了分类甲状腺结节的能力。另外,相比先前可用的基因组和相关的方法,所述基因组提供了更优的诊断和分类结果。例如,相比Affirma®基因分类器,本发明的基因组和方法显示了更佳的可预测性和可靠性。使用本发明的基因组和二相逐步算法避免了过拟合,并且考虑到群体的基因谱表达差异,能够充分地分类具有离群基因谱的患者。此外,本发明的小基因组允许试剂盒可被分发至病理实验室,从而降低成本,以及简化和加快诊断过程。小基因组的另一优势在于其需要较低的组织样本量,从而可能允许从原始FNA样本提取足够的mRNA,由此避免需要使患者进行第二次FNA。

[0043] 在以下描述中,示出了某些具体细节以提供对本发明不同实施方案的完整理解。然而,本领域技术人员将理解本发明可在没有这些细节下实施。

[0044] 定义

[0045] 除非另有定义,本文的所有技术和科学术语具有本发明所属领域技术人员通常理解的相同的含义。为了本发明的目的,以下术语定义如下。

[0046] 词语“一个”和“一种”是指一个(种)或多(种),除非有特别注明。

[0047] “约”是指这样的量、水平、值、数字、频率、百分比、尺寸、大小、数量、重量或长度,其与参照量、水平、值、数字、频率、百分比、尺寸、大小、数量、重量或长度相差多达30%、25%、20%、15%、10%、9%、8%、7%、6%、5%、4%、3%、2%或1%。在联合术语“约”使用的数值的上下文论述的任何实施方案中,明确包括该术语约可以省略。

[0048] “编码序列”是指有助于编码基因的多肽产物的任何多核苷酸序列。相反,术语“非编码序列”是指不促成编码基因的多肽产物的任何多核苷酸序列。

[0049] 除非上下文另有要求,在整个本说明书和权利要求中,词语“包括”及其变型如“包含”和“含有”应理解为开放、包括性含义,即理解为“包括但不限于”。

[0050] “由…组成”是指包括并且限于短语“由…组成”后的任何内容。因此,短语“由…组成”是指所列的元素为必需或强制性的,并且不可以存在其他元素。

[0051] “基本上由…组成”是指包括该短语后所列的任何元素,并且限制于不干扰或有助于所列元素的公开中指明的活性或作用的其他元素。因此,短语“基本上由…组成”是指所列的元素为必需或强制性的,但其他元素是任选的,并且可以存在或不存在,这取决于它们是否影响所列元素的活性或作用。

[0052] “降低的(decreased)”或“减少的(reduced)”或“较少的(lesser)”量通常为“统计上显著的”量,并且可包括降低本文所述量或水平约1.1、1.2、1.3、1.4、1.5、1.6、1.7、1.8、1.9、2、2.5、3、3.5、4、4.5、5、6、7、8、9、10、15、20、30、40或50倍或更多倍(例如,100、500、1000倍)(包括其间的且大于1的所有整数和小数点,例如1.5、1.6、1.7、1.8等)。

[0053] 在整个本说明书中提及“实施方案”或“一(种)实施方案”是指与所述实施方案相关的所述特定特征、结构或特性包含在本发明的至少一种实施方案中。因此,在整个本说明书的不同地方出现的短语“在一(种)实施方案中”或“在实施方案中”未必都指同一实施方案。此外,特定特征、结构或特性可以任何合适的方式组合在一种或多种实施方案中。

[0054] “基因”是指占据染色体上的特定位点的遗传单位,且由转录和/或翻译调控序列和/或编码区域和/或未翻译的序列(即内含子、5'和3'非翻译序列)组成。

[0055] “增加的”或“提高的”量通常为“统计上显著的”量,并且可包括增加本文所述量或水平1.1、1.2、1.3、1.4、1.5、1.6、1.7、1.8、1.9、2、2.5、3、3.5、4、4.5、5、6、7、8、9、10、15、20、30、40或50倍或更多倍(例如,100、500、1000倍)(包括其间的且大于1的所有整数和小数点,例如2.1、2.2、2.3、2.4等)。

[0056] “分离的”是指这样的物质,在其天然状态下基本上或实质上没有通常伴随其的组分。例如,如本文所用的“分离的多核苷酸”包括从其天然存在的状态下位于其两侧的序列纯化而来的多核苷酸,例如,从通常紧邻DNA片段的序列去除的DNA片段。可选地,如本文所用的“分离的肽”或“分离的多肽”等,包括从其天然细胞环境和与细胞其他组分的结合体外分离和/或纯化的肽或多肽分子;即,其不与体内物质明显结合。

[0057] 如本文所用的术语“mRNA”或有时提及为“mRNA转录物”,包括但不限于前mRNA转录物、转录物加工中间物、准备好翻译的成熟mRNA和一种或多种基因的转录物,或者来源于

mRNA转录物的核酸。转录物加工可包括剪接、编辑和降解。如本文所用,来源于mRNA转录物的核酸是指其合成最终将mRNA转录物或其子序列用作模板的核酸。从mRNA逆转录的cDNA、从该cDNA转录的RNA、从cDNA扩增的DNA、从扩增的DNA转录的RNA等均来源于mRNA转录物,并且检测这样来源的产物指示样本中原始转录物的存在和/或丰度。因此,mRNA来源的样本包括但不限于一种或多种基因的mRNA转录物、从mRNA逆转录的cDNA、从cDNA转录的cRNA、从所述基因扩增的DNA、从扩增的DNA转录的RNA等。

[0058] “获自”是指样本如组织、多核苷酸或多肽分离自或来源于特定来源,如所需生物体(例如,个体)或所需生物体中的特定组织。例如,组织样本可获自个体,或者多核苷酸或多肽可获自直接从个体分离的组织或生物流体。“来源于”或“获自”也可指组织或多肽或多核苷酸序列的来源。

[0059] 如本文所用的叙述“多核苷酸”或“核酸”是指mRNA、RNA、cRNA、rRNA、cDNA或DNA。该术语通常是指至少10个碱基长度的核苷酸的聚合形式,所述核酸为核糖核苷酸或脱氧核苷酸或每种类型的核苷酸的修饰形式。该术语包括单链和双链形式的DNA和RNA。如本领域技术人员能够理解的,在不同的实施方案中,本发明的多核苷酸序列可包括基因组序列、基因组外的和质粒编码的序列,以及较小的经改造的表达或可适合于表达蛋白、多肽、肽等的基因区段。这样的区段可为天然分离的,或通过人工合成改良的。本发明的多核苷酸,无论编码序列本身的长度,均可与其他DNA序列如启动子、聚腺苷酸化信号、另外的限制酶位点、多克隆位点、其他编码区段等相组合,使得它们的整个长度可变化非常大。因此,包括可使用几乎任何长度的多核苷酸片段,其总长度优选受到易于制备和用于预期的重组DNA方案的限制。

[0060] 术语“多核苷酸变体”是指显示与参照多核苷酸序列大量序列同一性的多核苷酸或者能够在下文定义的严格条件下与参照序列杂交的多核苷酸。该术语还包括通过至少一个核苷酸的添加、缺失或取代而不同于参照多核苷酸的多核苷酸。因此,术语“多核苷酸变体”包括这样的多核苷酸,其中添加或缺失或用不同核苷酸取代了一个或多个核苷酸。就这一点而言,本领域完全理解可对参照多核苷酸进行某些改变,包括突变、添加、缺失和取代,由此改变的多核苷酸保留有参照多核苷酸的生物功能或活性,或者具有相对于参照多核苷酸增加的活性(即,优化的)。多核苷酸变体包括,例如与本文所述的参照多核苷酸序列具有至少50%(和至少51%至至少99%,以及其间的所有整数百分比,例如90%、95%或98%)序列同一性的多核苷酸。术语“多核苷酸变体”和“变体”还包括编码这些酶的天然存在的等位基因变体和直系同源物。

[0061] 叙述“序列同一性”或例如本文所用的包含“与...具有50%同一性的序列”,是指基于核苷酸-核苷酸或氨基酸-氨基酸,序列在比较窗上的同一性程度。因此,可通过如下方法计算“序列同一性百分比”:在比较窗上比较两条最佳比对的序列,确定在两条序列上存在相同的核酸碱基(例如,A、T、C、G、I)或相同的氨基酸残基(例如,Ala、Pro、Ser、Thr、Gly、Val、Leu、Ile、Phe、Tyr、Trp、Lys、Arg、His、Asp、Glu、Asn、Gln、Cys和Met)的位点数,从而生成匹配的位点数,将匹配的位点数除以比较窗中的位点总数(即,窗大小),并将结果乘以100,从而生成序列同一性百分比。

[0062] 用于描述两条或更多条多核苷酸或多肽之间的序列相关性的术语包括“参照序列”、“比较窗”、“序列同一性”、“序列同一性百分比”和“基本上相同”。“参照序列”的长度为

至少12个,但通常为15-18个,且常常为至少25个单体单位,包含核苷酸和氨基酸残基。因为两条多核苷酸可各自包含(1)两个多核苷酸之间相似的序列(即,仅完整多核苷酸序列的一部分),和(2)两个多核苷酸之间不同的序列,通常通过在“比较窗”上比较两个多核苷酸的序列来进行两个(或更多个)多核苷酸之间的序列比较,以确定和比较序列相似的局部区域。“比较窗”是指至少6个,通常约50个至约100个,更通常约100个至约150个连续位点的概念区段,其中在将两条序列最佳比对后,将序列与参照序列的相同数目连续位点进行比较。对于两条序列的最佳比对,相比参照序列(其不包含添加或缺失),比较窗可包含约20%或更少的添加或缺失(即,缺口)。可通过计算机执行的算法(Wisconsin Genetics Software Package Release 7.0中的GAP、BESTFIT、FASTA和TFASTA,Genetics Computer Group, 575Science Drive Madison,WI,USA)或者通过目测和使用选择的不同方法中的任何一种产生的最佳比对(即,在比较窗产生最高百分比的同源性),进行用于比对比较窗的序列最佳比对。还可参照例如Altschul等(1997,Nucl.Acids Res.25:3389)公开的程序的BLAST家族。序列分析的详细论述可见于Ausubel等(“Current Protocols in Molecular Biology”,John Wiley&Sons Inc,1994-1998)第15章第19.3单元。

[0063] 术语“多肽”和“蛋白”在本文中可互换使用,是指氨基酸残基的聚合物及其变体和合成及天然存在的类似物。因此,这些术语适用于其中一个或多个氨基酸残基为合成的非天然存在的氨基酸的氨基酸聚合物,如相应的天然存在的氨基酸的化学类似物,以及天然存在的氨基酸聚合物及其天然存在的化学衍生物。

[0064] 如本文所用的“个体”包括显示症状或存在显示症状风险的任何动物,所述症状可根据本发明进行治疗或诊断。还包括适合描述本发明的基因产物水平的个体,以用于诊断或其他目的。合适的个体(患者)包括实验室动物(如小鼠、大鼠、兔或豚鼠)、农场动物和家畜或宠物(如猫或狗)。包括哺乳动物,包括非人的灵长类和人类。

[0065] 如本文所用的“治疗(treatment/treating)”包括对疾病或病症如甲状腺癌的症狀或病理的任何有利的影响,并且可包括接受治疗的疾病或病症的一种或多种可测量的标志物的甚至最小的改变或改善。“治疗”未必指完全根除或治愈该疾病或病症或其相关症状。接受这种治疗的个体为有需要的任何个体。临床改善的示例性标志物对本领域技术人员是显而易见的。

[0066] 如本文所用的术语“野生型”是指这样的微生物(例如,细菌物种或菌株)、基因或基因产物,其在从天然存在的来源分离时具有该微生物(例如,细菌物种或菌株)、基因或基因产物的特性。野生型基因或基因产物(例如,多肽)为在群体中最常观察到并因而被任意地称为基因的“正常”或“野生型”形式的基因或基因产物。

[0067] 除非另外明确相反的指出,本发明的实施将采用本领域技术范围内的常规分子生物学方法和重组DNA技术,其中许多为了说明的目的在下文有描述。这样的技术在文献中已得到充分解释。参见,例如Sambrook,et al.,Molecular Cloning:A Laboratory Manual (3rd Edition,2000);DNA Cloning:A Practical Approach,vol.I&II(D.Glover,ed.);Oligonucleotide Synthesis(N.Gait,ed.,1984);Oligonucleotide Synthesis:Methods and Applications(P.Herdewijn,ed.,2004);Nucleic Acid Hybridization(B.Hames&S.Higgins,eds.,1985);Nucleic Acid Hybridization:Modern Applications(Buzdin and Lukyanov,eds.,2009);Transcription and Translation(B.Hames&S.Higgins,eds.,

1984);Animal Cell Culture(R.Freshney,ed.,1986);Freshney,R.I.(2005)Culture of Animal Cells,a Manual of Basic Technique,5th Ed.Hoboken NJ,John Wiley&Sons; B.Perbal,A Practical Guide to Molecular Cloning(3rd Edition 2010);Farrell,R., RNA Methodologies:A Laboratory Guide for Isolation and Characterization(3rd Edition 2005),Methods of Enzymology:DNA Structure Part A:Synthesis and Physical Analysis of DNA Methods in Enzymology,Academic Press;Using Antibodies:A Laboratory Manual:Portable Protocol NO.I,Edward Harlow,David Lane,Ed Harlow(1999,Cold Spring Harbor Laboratory Press,ISBN 0-87969-544-7); Antibodies:A Laboratory Manual by Ed Harlow(Editor),David Lane(Editor)(1988, Cold Spring Harbor Laboratory Press,ISBN 0-87969-3,4-2),1855.Handbook of Drug Screening,编辑Ramakrishna Seethala,Prabhavathi B.Fernandes(2001,New York, N.Y.,Marcel Dekker,ISBN 0-8247-0562-9);以及Lab Ref:A Handbook of Recipes, Reagents,and Other Reference Tools for Use at the Bench,Edited Jane Roskams and Linda Rodgers,(2002,Cold Spring Harbor Laboratory,ISBN 0-87969-630-3)。

[0068] 某些实施方案可采用常规的生物学方法、软件和系统以用于本发明的诊断目的。本发明的计算机软件产品通常包括具有用于执行本发明方法的逻辑步骤的计算机可执行指令的计算机可读介质。合适的计算机可读介质包括软盘、CD-ROM/DVD/DVD-ROM、硬盘驱动器、闪存、ROM/RAM、磁带等。可以合适的计算机语言或几种语言的组合书写计算机可执行的指令。基本的计算机生物学方法描述于,例如,Setubal和Meidanis et al.,Introduction to Computational Biology Methods(PWS Publishing Company,Boston,1997); Salzberg,Searles,Kasif,(Ed.),Computational Methods in Molecular Biology,(Elsevier,Amsterdam,1998);Rashidi和Buehler,Bioinformatics Basics:Application in Biological Science and Medicine(CRC Press,London,2000)及Ouelette and Bzevanis Bioinformatics:A Practical Guide for Analysis of Gene and Proteins (Wiley&Sons,Inc.,2nd ed.,2001)中。参见美国专利号6,420,108。

[0069] 某些实施方案可采用不同的计算机程序产品和软件,以用于各种目的,如探针设计、数据管理、分析和仪器操作。参见美国专利号5,593,839、5,795,716、5,733,729、5,974,164、6,066,454、6,090,555、6,185,561、6,188,783、6,223,127、6,229,911和6,308,170。

[0070] 诊断测定法

[0071] 本发明部分是基于甲状腺癌的新生物标志物的鉴定,其允许将甲状腺肿瘤或结节分类为良性或癌性的,即恶性的。因此,本发明提供了用于分析获自个体的生物样本以确定所述个体是否患有甲状腺癌的诊断测定法和相关的试剂盒。在不同的实施方案中,将本发明的方法和试剂盒用于诊断或检测个体中有或无甲状腺癌,例如通过确定获自所述个体的生物样本中有或无甲状腺癌细胞。在特定实施方案中,将本发明的方法和试剂盒用于诊断先前例如通过细胞学分析诊断为不确定的个体中有或无甲状腺癌。

[0072] 甲状腺的异常生长可引起结节形成,其可为良性或癌性的(即恶性的)。甲状腺癌包括至少四种不同类型的甲状腺恶性肿瘤:乳头状的、滤泡性的、髓样的和未分化的;恶性的亚型包括例如,滤泡性癌(FC)、乳头状甲状腺癌(PTC)、乳头状癌的滤泡性变体(FVPTC)、髓样甲状腺癌(MTC)、Hurthle细胞癌(HC)和未分化甲状腺癌(ATC)。良性(非癌性的)甲状腺

肿瘤或结节的实例包括,例如滤泡性腺瘤(FA)、结节状增生(NHP)、淋巴细胞性甲状腺炎(LCT)和Hurthle细胞腺癌(HA)。在本发明的几方面,甲状腺癌为侵入性癌症或具有转移潜能,例如,具有转移潜能的侵入性髓样或滤泡性甲状腺癌或髓样或滤泡性甲状腺癌。在本发明的特定实施方案中,甲状腺癌为未分化甲状腺癌(ATC)。“转移潜能”是指癌细胞从身体中的初始部位(即甲状腺)移动至其他部位的能力或可能性。本领域技术人员将理解,通过使用一组合适的参照对照样本,可将本发明的方法容易地用于诊断或检测有或无甲状腺的这些癌性肿瘤或非癌性病症中的任何一种。

[0073] 术语“诊断”或“诊断性的”或“诊断的”包括鉴定病理状况如甲状腺癌的存在或性质,表征发展这样的病症的风险,和/或测量病理状况响应治疗的改变(或无改变)。诊断方法的灵敏度和特异性可以不同。在某些实施方案中,诊断测定法的“灵敏度”是指检测呈阳性的患病细胞、组织或个体的百分比(“真阳性”百分比)。通过测定未检测到的患病细胞、组织或个体通常称为“假阳性”。未患病且在测定中检测呈阴性的细胞、组织或个体可称为“真阴性”。在某些实施方案中,诊断测定法的“特异性”可定义为一(1)减去假阳性率,其中所述“假阳性”率定义为未患病且检测呈阳性的那些样本或个体的比例。虽然特定的诊断方法不能够提供病症的确诊,但如果该方法提供了有助于诊断的积极指示就合格。

[0074] 在本发明方法和试剂盒的特定实施方案中,甲状腺组织或结节样本以大于或等于90%、大于或等于91%、大于或等于92%、大于或等于93%、大于或等于94%、大于或等于95%、大于或等于96%、大于或等于97%、大于或等于98%或大于或等于99%的灵敏度被诊断为癌性或良性的。

[0075] 在本发明方法和试剂盒的特定实施方案中,甲状腺组织或结节样本以大于或等于50%、大于或等于60%、大于或等于70%、大于或等于80%、大于或等于90%或大于或等于95%的特异性被诊断为癌性或良性的。

[0076] 在本发明方法和试剂盒的特定实施方案中,甲状腺组织或结节样本以大于或等于50%、大于或等于60%、大于或等于70%、大于或等于80%、大于或等于90%或大于或等于95%的阳性预测值被诊断为癌性的或良性的。

[0077] 在本发明方法和试剂盒的特定实施方案中,甲状腺组织或结节样本以大于或等于90%、大于或等于91%、大于或等于92%、大于或等于93%、大于或等于94%、大于或等于95%、大于或等于96%、大于或等于97%、大于或等于98%或大于或等于99%的阴性预测值被诊断为癌性或良性的。

[0078] 在本发明方法和试剂盒的特定实施方案中,甲状腺组织或结节样本以大于或等于2、大于或等于3、大于或等于4、大于或等于5、大于或等于6、大于或等于7、大于或等于8、大于或等于10、大于或等于15、大于或等于20或大于或等于25的阳性似然比被诊断为癌性或良性的。

[0079] 在本发明方法和试剂盒的特定实施方案中,甲状腺组织或结节样本以大于或等于50%、大于或等于60%、大于或等于70%、大于或等于80%、大于或等于90%或大于或等于95%的阳性验后概率被诊断为癌性或良性的。

[0080] 在本发明方法和试剂盒的特定实施方案中,甲状腺组织或结节样本以小于或等于0.20、小于或等于0.18、小于或等于0.16、小于或等于0.14、小于或等于0.12、小于或等于0.10、小于或等于0.08或小于或等于0.06的阴性似然比被诊断为癌性或良性的。

[0081] 在本发明方法和试剂盒的特定实施方案中,甲状腺组织或结节样本以小于或等于10.0%、小于或等于9.0%、小于或等于8.0%、小于或等于7.0%、小于或等于6.0%、小于或等于5.0%、小于或等于4.0%或小于或等于3.0%的阴性验后概率被诊断为癌性或良性的。

[0082] 在本发明方法和试剂盒的特定实施方案中,甲状腺组织或结节样本以大于或等于92%或者大于或等于97%的灵敏度和大于或等于60%或者大于或等于90%的特异性被诊断为癌性或良性的。在特定实施方案中,AUC大于0.97,灵敏度和特异性值均分别大于或等于92%和60%。在特定实施方案中,AUC大于0.97,灵敏度和特异性值均分别大于或等于92%和90%。在特定实施方案中,AUC大于0.97,灵敏度和特异性值均分别大于或等于97%和90%。

[0083] 在一些实施方案中,本发明提供了诊断、鉴定或分类癌症如甲状腺癌的方法,其包括如下步骤:获得生物样本如甲状腺组织样本的一种或多种基因产物的表达水平;以及鉴定所述生物样本为良性的,其中所述基因产物表达水平指示生物样本中没有癌症。在其他实施方案中,本发明提供了诊断、鉴定、分类或诊断癌症如甲状腺癌的方法,其包括如下步骤:获得生物样本的一种或多种基因产物的表达水平;以及鉴定所述生物样本为恶性或疑似的,其中所述基因产物表达水平指示生物样本中的癌症。例如,这可通过将生物样本中基因产物的表达水平与对照样本中同一基因产物的表达水平或参照值相关联而进行,以鉴定(或排除)生物样本中甲状腺癌的存在。

[0084] 在特定实施方案中,本发明提供了诊断、鉴定或分类个体的癌症如甲状腺癌的方法,包括如下步骤:进行测定以确定生物样本如甲状腺组织样本的一种或多种基因产物的表达水平;以及鉴定所述生物样本为良性的,其中所述基因产物表达水平指示所述生物样本中没有癌症,或者鉴定所述生物样本为恶性或疑似的,其中所述基因产物表达水平指示生物样本中的癌症。在特定实施方案中,所述方法包括确定甲状腺组织样本中两种或更多种或者三种或更多种基因产物的表达水平,其中所述两种或更多种或者三种或更多种基因产物由表1中列出的一种或多种基因表达,并且其中至少一种基因产物由CXCR3、CCR3、CXCL10、CAR、XB130、HO-1或CCR7基因表达。在某些实施方案中,如果所述生物样本确定为癌性或恶性的,则所述方法还包括对所述个体实施手术,例如甲状腺切除术。在特定实施方案中,所述基因产物为RNA,并且所述测定包括PCR、RT-PCR或定量PCR,或者用于测量RNA量或表达水平的任何其他测定,包括本文中描述的那些测定中的任何一种。在特定实施方案中,所述基因产物为多肽,并且所述测定包括用于测量多肽量或表达水平的免疫组化测定或任何其他测定,包括本文中描述的那些中的任何一种测定。

[0085] 在特定实施方案中,本发明提供了诊断、鉴定或分类个体的癌症如甲状腺癌的方法,包括如下步骤:从个体获取生物样本如甲状腺组织样本;进行测定以确定所述生物样本中一种或多种基因产物的表达水平;以及鉴定所述生物样本为良性的,其中所述基因产物表达水平指示所述生物样本中没有癌症,或者鉴定所述生物样本为恶性或疑似的,其中所述基因产物表达水平指示生物样本中的癌症。在特定实施方案中,所述方法包括确定甲状腺组织样本中两种或更多种或者三种或更多种基因产物的表达水平,其中所述两种或更多种或者三种或更多种基因产物由表1中列出的一种或多种基因表达,并且其中至少一种基因产物由CXCR3、CCR3、CXCL10、CAR、XB130、HO-1或CCR7基因表达。在某些实施方案中,如果所述生物样本确定为癌性或恶性的,则所述方法还包括对所述个体实施手术,例如甲状腺

切除术。在特定实施方案中,所述基因产物为RNA,并且所述测定包括PCR、RT-PCR或定量PCR,或者用于测量RNA量或表达水平的任何其他测定,包括本文中描述的那些中的任何一种测定。在特定实施方案中,所述基因产物为多肽,并且所述测定包括免疫组化测定或用于测量多肽量或表达水平的任何其他测定,包括本文中描述的那些中的任何一种测定。

[0086] 如本文中进一步描述,在特定实施方案中,所述生物样本获自个体,例如疑似患有癌症或存在患癌风险的个体。确定表达的基因产物包括本文中描述的那些,并且可包含两种或更多种基因产物,其也可称为“基因产物组”。本文中描述的可用于确定有或无癌症如甲状腺癌的基因产物也可称为“生物标志物”。

[0087] 在特定实施方案中,本发明提供了检测或诊断个体中有或无甲状腺癌的方法,包括确定获自所述个体的甲状腺组织样本中两种或更多种或者三种或更多种基因产物的表达水平,其中所述两种或更多种或者三种或更多种基因产物由表1中列出的一种或多种基因表达,并且其中所述基因产物中的至少一种由CXCR3、CCR3、CXCL10、CAR、XB130、HO-1或CCR7基因表达;以及通过将确定的来自所述个体的生物样本的表达水平与有或无甲状腺癌相关联而鉴定所述甲状腺组织样本为癌性或良性的。

[0088] 本发明包括治疗有需要的个体的方法,包括如果个体已确定为患有甲状腺癌,则对所述个体实施手术,例如手术去除个体的甲状腺或其一部分(例如,甲状腺切除术),其中所述确定通过本发明诊断方法中的任何一种进行。在特定实施方案中,治疗有需要的个体的方法包括如果个体确定为患有甲状腺癌时,则通过包括以下步骤的方法对所述个体实施手术,例如甲状腺切除术:进行测定以确定生物样本如甲状腺组织样本的一种或多种基因产物的表达水平;以及其中所述基因产物表达水平指示所述生物样本中没有癌症时,鉴定所述生物样本为良性,或者其中所述基因产物表达水平指示所述生物样本中的癌症时,鉴定生物样本为恶性或疑似的。在特定实施方案中,所述方法包括确定甲状腺组织样本中两种或更多种或者三种或更多种基因产物的表达水平,其中所述两种或更多种或者三种或更多种基因产物由表1中列出的一个或多个基因表达,并且其中所述基因产物中的至少一种由CXCR3、CCR3、CXCL10、CAR、XB130、HO-1或CCR7基因表达。在特定实施方案中,所述方法还包括对例如通过针吸活组织检查或细针抽吸所述个体获得的生物样本进行细胞学或组织化学分析来鉴定所述个体存在患甲状腺癌的风险。

[0089] 在相关实施方案中,本发明包括治疗有需要的个体的方法,其包括:通过本发明诊断方法中的任何一种来确定所述个体是否患有癌症,如甲状腺癌;以及如果所述个体确定为患有癌症如甲状腺癌,则实施手术,如手术去除个体的肿瘤或其一部分(例如,甲状腺切除术)。在特定实施方案中,治疗有需要的个体的方法包括:(i) 通过包括如下步骤的方法来确定所述个体是否患有甲状腺癌:进行测定以确定生物样本如甲状腺组织样本的一种或多种基因产物的表达水平;以及其中所述基因产物表达水平指示生物样本中没有癌症时,鉴定所述生物样本为良性,或者其中所述基因产物表达水平指示生物样本中的癌症时,确定所述生物样本为恶性或疑似的;以及(ii) 如果步骤(i)的结果指示所述个体患有或可能患有癌症如甲状腺癌,则对所述个体进行手术如甲状腺切除术。在特定实施方案中,所述方法包括确定甲状腺组织样本中两种或更多种,或者三种或更多种基因产物的表达水平,其中所述两种或更多种,或者三种或更多种基因产物由表1中列出的一个或多个基因表达,并且其中所述基因产物中的至少一种由CXCR3、CCR3、CXCL10、CAR、XB130、HO-1或CCR7基因表达。

在特定实施方案中,所述方法还包括对例如通过针吸活组织检查或细针抽吸所述个体获得的生物样本进行细胞学或组织化学分析来鉴定所述个体存在患甲状腺癌的风险。

[0090] 在某些实施方案中,本发明包括治疗有需要的个体的方法,包括:(i) 索要或获取本文中描述的对获自所述个体的生物样本如甲状腺样本进行的诊断测定的结果;以及(ii) 如果所述诊断测定的结果指示所述个体患有癌症如甲状腺癌,则进行手术,如手术去除所述个体的肿瘤或其一部分(例如,甲状腺切除术)。在特定实施方案中,所述诊断测定包括:进行测定以确定生物样本如甲状腺组织样本的一种或多种基因产物的表达水平;以及其中所述基因产物表达水平指示所述生物样本中没有癌症时,鉴定所述生物样本为良性的,或者其中所述基因产物表达水平指示生物样本中的癌症时,确定所述生物样本为恶性或疑似的。在特定实施方案中,所述方法包括确定甲状腺组织样本中两种或更多种,或者三种或更多种基因产物的表达水平,其中所述两种或更多种,或者三种或更多种基因产物由表1中列出的一个或多个基因表达,并且其中所述基因产物中的至少一种由CXCR3、CCR3、CXCL10、CAR、XB130、HO-1或CCR7基因表达。在特定实施方案中,所述方法还包括通过索要或获取例如通过针吸活组织检查或细针抽吸所述个体获得的生物样本的细胞学或组织化学分析结果来鉴定所述个体存在患甲状腺癌的风险。

[0091] 在某些实施方案中,本发明包括确定个体是否患有癌症如甲状腺癌的方法,其中对获自所述个体的生物样本如甲状腺组织的FNA进行初步检测为不确定的,所述方法包括进行、索要或获取对获自所述个体的生物样本如甲状腺样本进行本文所述的诊断测定的结果。在特定实施方案中,所述诊断测定包括:进行测定以确定生物样本如甲状腺组织样本的一种或多种基因产物的表达水平;以及其中所述基因产物表达水平指示所述生物样本中没有癌症时,鉴定所述生物样本为良性的,或者其中所述基因产物表达水平指示所述生物样本中的癌症时,鉴定所述生物样本为恶性的或癌症。在特定实施方案中,所述方法包括确定甲状腺组织样本中两种或更多种,或者三种或更多种基因产物的表达水平,其中所述两种或更多种,或者三种或更多种基因产物由表1中列出的一种或多种基因表达,并且其中所述基因产物中的至少一种由CXCR3、CCR3、CXCL10、CAR、XB130、HO-1或CCR7基因表达。

[0092] 在本发明任何方法的特定实施方案中,通过将获自个体的样本中基因产物的表达水平与检测的每一种基因产物的对照或参照表达水平相比较来进行所述关联。如果甲状腺组织样本与正常对照或参照表达水平之间的基因产物表达水平存在显著差异,则将所述甲状腺组织样本鉴定为癌性的。在某些实施方案中,如果甲状腺组织样本与正常对照或参照表达水平之间的两种或更多种、三种或更多种、或者四种或更多种基因产物的表达水平存在显著差异,则将所述甲状腺组织样本鉴定为癌性的。同样地,如果甲状腺组织样本与正常对照或参照表达水平之间的基因产物表达水平没有显著差异(即,存在很大的相似性),则将所述甲状腺组织样本鉴定为良性的。在某些实施方案中,如果甲状腺组织样本与正常对照或参照表达水平之间的两种或更多种、三种或更多种、或者四种或更多种基因产物的表达水平没有显著差异,则将所述甲状腺组织样本鉴定为良性的。

[0093] 在本文所述任何方法的某些实施方案中,如果相比正常对照表达水平,甲状腺组织样本中基因1、8、9、13、14、15和/或18中的任何一种或多种的表达水平降低;和/或基因2、3、4、5、6、7、10、11、12、16和/或17中的任何一种或多种的表达水平增加,则将所述甲状腺组织样本鉴定为癌性的,其中增加或降低表达的基因总数目至少为3种。每种基因的名称如

下: CXCR3 (基因1)、CXCR3A (基因2)、CXCR3B (基因3)、CXCR4 (基因4)、CCR3 (基因5)、CXCL9 (基因6)、CXCL10 (基因7)、CXCL11 (基因8)、SPAG-9 (基因9)、CK-19 (基因10)、TIMP-1 (基因11)、CLDN-1 (基因12)、CAR (基因13)、Nectin-1 (基因14)、XB-130 (基因15)、HO-1 (基因16)、CCR7 (基因17) 和 CXCL4 (基因18)。在本文所述任何方法的其他实施方案中, 如果本文所述的任何基因亚组中的一种或多种、两种或更多种、三种或更多种、或者四种或更多种基因产物的表达水平如上文所述变化, 则将所述甲状腺组织样本鉴定为癌性的。在特定实施方案中, 所述基因组包括 CXCR3、CCR3、CXCL10、CAR、XB130、HO-1 或 CCR7 中的一种或多种。

[0094] 在特定实施方案中, 通过使用算法比较获自个体的样本中基因产物的表达水平与参照水平来进行所述关联。所述参照水平可包括从多种癌性和/或非癌性的生物样本中预先确定的每一种基因产物的表达水平。

[0095] 在某些实施方案中, 所述关联包括将所述表达水平与以下两组生物样本的基因产物确定的基因表达水平相比较:

[0096] 多种正常甲状腺组织样本; 和

[0097] 多种癌性甲状腺组织样本,

[0098] 其中如果甲状腺组织样本的基因产物表达水平与多种正常甲状腺组织样本的基因表达水平之间存在显著差异, 或者如果甲状腺组织样本基因产物的基因表达水平与所述多种癌性甲状腺组织样本的基因表达水平之间基本上没有差异, 则将所述甲状腺组织样本鉴定为癌性的。

[0099] 在本发明任何方法和试剂盒的特定实施方案中, 确定表达水平的两种或更多种或者三种或更多种基因产物包含以下或由以下组成:

[0100] CXCR3、CCR3、CXCL10、CK19、TIMP-1、CLDN-1、CAR、XB130、HO-1 和 CCR7 基因中的两种或更多种或者三种或更多种基因产物;

[0101] CXCR3、CCR3、CXCL10、CK19、TIMP-1、CLDN-1、CAR、HO-1 和 CCR7 基因的基因产物;

[0102] CCR3、TIMP-1、CAR 和 XB130 基因的基因产物;

[0103] CXCL10、TIMP-1、CAR 和 CCR7 基因的基因产物;

[0104] TIMP-1、CAR 和 CCR7 基因的基因产物; 或者

[0105] CXCL10、TIMP-1、CLDN-1 和 CCR7 基因的基因产物。

[0106] 在任何上述基因产物组的特定实施方案中, 所述两种或更多种或者三种或更多种基因产物包含 CXCR3、CCR3、CXCL10、CAR、XB130、HO-1 或 CCR7 中的一种或多种。在特定实施方案中, 所述基因产物包含表1中列出的一种或多种、两种或更多种、或者三种或更多种基因, 并且所述基因产物中的至少一种由 CXCR3 基因表达。在特定实施方案中, 所述基因产物包含表1中列出的基因中的一种或多种、两种或更多种、或者三种或更多种, 并且所述基因产物中的至少一种由 CCR3 基因表达。在特定实施方案中, 所述基因产物包含表1中列出的基因中的一种或多种、两种或更多种、或者三种或更多种, 并且所述基因产物中的至少一种由 CXCL10 基因表达。在特定实施方案中, 所述基因产物包含表1中列出的基因中的一种或多种、两种或更多种、或者三种或更多种, 并且所述基因产物中的至少一种由 CAR 基因表达。在特定实施方案中, 所述基因产物包含表1中列出的基因中的一种或多种、两种或更多种、或者三种或更多种, 并且所述基因产物中的至少一种由 XB130 基因表达。在特定实施方案中, 所述基因产物包含表1中列出的基因中的一种或多种、两种或更多种、或者三种或更多种,

并且所述基因产物中的至少一种由HO-1基因表达。在特定实施方案中,所述基因产物包含表1中列出的基因中的一种或多种、两种或更多种,或者三种或更多种,并且所述基因产物中的至少一种由CCR7基因表达。

[0107] 在不同实施方案中,本发明的方法还包括如下步骤:对获自个体的生物样本如甲状腺组织样本进行细胞学或组织学分析,例如以获得初步诊断。可在基于如本文所述的基因产物的表达进行分析之前、同时或之后进行细胞学或组织学分析。在某些实施方案中,通过本发明的方法进一步分析了具有中间或不确定的初步诊断的样本。

[0108] 在本发明方法的特定实施方案中,所述方法还包括从个体获取生物样本。

[0109] 在本发明方法的某些实施方案中,如果患者被诊断为患有甲状腺癌,所述方法还包括治疗个体的甲状腺癌。在某些实施方案中,所述治疗包括手术去除个体的甲状腺或其一部分。

[0110] 本发明还包括用于表征甲状腺癌的方法和试剂盒。如本文所用,术语个体的“表征甲状腺癌”是指鉴定个体的癌症如特定类型的甲状腺癌样本的一种或多种特性,并且还可包括确定个体的预后或存活。可通过鉴定一种或多种标志物(包括但不限于本文中公开的基因产物)的表达来表征癌症。技术人员将理解,本文所述的一般方法可容易地适于确定甲状腺癌类型,例如通过将所述基因产物的表达水平与确定的不同类型的甲状腺癌的那些表达水平进行比较。基于确定甲状腺癌类型,可例如基于历史数据或结果确定或估计预后、存活和/或转移可能性。

[0111] 生物样本

[0112] 在某些实施方案中,本发明的方法利用获自个体的生物样本,并且某些方法包括从个体获取生物样本。生物样本可为任何物质,其包含待检测个体的组织、细胞、核酸、基因、基因片段、表达产物、基因产物(例如,mRNA或蛋白),或者基因产物片段。提供了确定样本适用性和/或充分性的方法。样本可包含但不限于来自细胞或来源于个体细胞的组织、细胞或生物材料。样本可为异源或同源的细胞或组织群。在某些实施方案中,生物样本为组织样本,例如获自个体的甲状腺或甲状腺结节的样本。甲状腺结节为甲状腺的生长。在特定实施方案中,生物样本包含基因产物,例如核酸,如mRNA和/或蛋白。

[0113] 在不同的实施方案中,个体为动物(例如,哺乳动物),包括但不限于人类、非人的灵长类、啮齿类、狗、猫、猪、鱼等。在特定实施方案中,本方法和组合物适用于来自人的生物样本。在一些实施方案中,所述人为儿童、青少年或成人。在特定实施方案中,个体已确定为存在患有或疑似患有甲状腺肿瘤的风险。

[0114] 术语“疑似患有甲状腺癌的个体”是指显示指示甲状腺癌(例如,明显的肿块或团块)的一种或多种症状或者进行癌症筛查(例如,在常规体检中)的个体。例如,个体可能已被确定为具有增大的甲状腺和/或一个或多个甲状腺结节。疑似患有甲状腺癌的个体也可具有一种或多种风险因子。疑似患有甲状腺癌的个体包括已接受初步诊断但癌症阶段未知的个体。该术语还包括曾经患过癌症的人(例如,缓解期的个体)。另外,某些个体可能已预先测试了甲状腺肿瘤但结果为非决定性或者不确定的。

[0115] 如本文所用,术语“存在甲状腺癌风险的个体”是指具有发展甲状腺癌,特别是侵入性或转移性甲状腺癌,更特别地ATC的一种或多种风险因子的个体。风险因子包括但不限于性别、年龄、遗传易感性、环境暴露、先前的癌症发生、已存在的非癌症疾病以及生活方

式。

[0116] 可使用本领域已知的能够提供适于本文所述分析方法的样本的任何方法获取生物样本。从个体获取生物样本的方法包括但不限于活组织检查法,包括细针抽吸(FNA)、针抽吸、芯针吸活组织检查、真空辅助活组织检查、切开活组织检查、切除活组织检查或钻取活组织检查。在特定实施方案中,本发明的方法和试剂盒利用通过FNA获取的生物样本。获取合适的甲状腺样本的方法为本领域已知,并且在甲状腺结节管理的ATA指导(ATA Guidelines for thyroid nodule management)(Cooper et al. Thyroid Vol.16No.2 2006)中有进一步的描述,其通过引用整体并入本文。获取生物样本的一般方法在本领域中也有描述,并且在例如Ramzy(Ibrahim Clinical Cytopathology and Aspiration Biopsy 2001)中有进一步描述,其通过引用整体并入本文中。在一实施方案中,样本为细针抽吸的甲状腺、甲状腺结节或疑似甲状腺肿瘤。在一些情况下,可通过使用超声、X-射线或其他成像设备来引导细针抽吸取样过程。

[0117] 在一些情况下,例如在相同或不同的时间,可获取多个生物样本如多个甲状腺样本用于通过本发明方法进行诊断。在一些情况下,在相同或不同时间获取的一个或多个样本或样本被保存和/或通过不同的方法进行分析。例如,可获取样本并通过细胞学分析(常规染色)来进行分析。在一些情况下,可在相同时间或晚些时候,例如,基于细胞学分析的结果,从个体获取其他样本。所述其他样本可用于本发明的方法,例如,当细胞学分析对于有或无癌症为不确定时。在本发明方法的其他实施方案中,可获取单个样本,并通过细胞学分析对一部分样本进行分析,而通过本发明的方法对另一部分样本进行分析。

[0118] 在某些实施方案中,生物样本由医学专业人员,例如,在医院、医疗诊所、测试中心或实验室从个体获得。在某些实施方案中,可使用试剂盒获取生物样本,所述试剂盒可包含用于获取如本文所述样本的工具(means)、保存样本的工具和使用试剂盒的说明书。在一些情况下,所述试剂盒由分子表征服务(molecular profiling service)提供,其也可对生物样本进行诊断测定。

[0119] 在特定实施方案中,在获取样本之后且在通过本发明的一种或多种方法分析样本之前,将生物样本保存一段时间如数秒、数分钟、数小时、几天、几周、几个月、几年或更长的时间。在一些情况下,在保存步骤或进一步分析之前,对获自个体的样本进行细分,以便对样本的不同部分实施不同的下游方法或处理,包括但不限于保存、细胞学分析、充分性测试、核酸提取、蛋白提取、分子表征或上述的组合。在一些情况下,将一部分样本保存,而对所述样本的另一部分进行进一步的操作。这样的操作可包括但不限于:分子表征;细胞学染色;核酸(例如,mRNA)提取、检测或定量;蛋白提取、检测或定量;固定(例如,福尔马林固定的石蜡包埋的样本);和/或检查。可在通过本领域已知的任何方法如使用戊二醛、甲醛或甲醇保存之前或期间对所述样本进行固定。在其他情况下,获取并保存所述样本,并在保存以用于进一步分析的步骤后进行细分,以便对样本的不同部分实施不同的下游方法或处理,包括但不限于保存、细胞学分析、充分性测试、核酸提取、多肽提取、分子表征、确定一种或多种基因产物的表达或者上述的组合。在一些情况下,获取样本并通过例如细胞学分析进行分析,并通过本发明的一种或多种方法对进一步分析所得到的样本材料,包括例如通过分子表征确定本文所述的基因产物的表达水平。在这样的情况下,可在细胞学分析步骤和例如通过分子表征确定基因产物表达水平的步骤之间保存样本。在某些实施方案中,将样

本(例如,在约0℃、-1℃、-2℃、-3℃、-4℃、-5℃、-6℃、-7℃、-8℃、-9℃、-10℃、-12℃、-14℃、-15℃、-16℃、-20℃、-22℃、-25℃、-28℃、-30℃、-35℃、-40℃、-45℃、-50℃、-60℃、-70℃、-80℃、-100℃、-120℃、-140℃、-180℃、-190℃或约-200℃中的任何一个),或在降低温度(例如,约20℃至约0℃)下冷冻保存于合适的介质、赋形剂或溶液中,如适合于保存用于随后细胞学分析的细胞的商业制剂,例如但不限于Cytoc ThinPrep、SurePath或Monoprep。

[0120] 可在样本获取期间或之后,包括在保存样本的步骤之前或之后,评估生物样本在本发明的方法和组合物中使用的适合性。由于许多因素可将样本确定为足够或不足用于进一步分析,所述许多因素包括但不限于:细胞不够、遗传材料不够、蛋白、DNA或RNA不够、用于所示测试的细胞不合适或者用于所示测试的材料不合适、样本年龄、获取样本的方式,或者保存或运输样本的方式。可使用本领域已知的多种方法如细胞染色方法、细胞数目或组织量的测量、总蛋白的测量、核酸的测量、目测检查、显微镜检查或者温度或pH测定来确定充分性。在一实施方案中,由进行基因产物水平分析实验的结果来确定样本充分性。

[0121] 确定存在足够数目的特定类型细胞的方法的实例包括PCR、定量PCR、RT-PCR、免疫组化分析、细胞学分析、显微镜检查和/或目测分析。

[0122] 可使用本领域已知的多种方法,如紫外吸光度,包括但不限于使用分光光度计测得260纳米处的吸光度,通过从生物样本提取后确定核酸含量来对样本进行分析。在某些实施方案中,使用NanoDrop分光光度计在纳克至微克范围测量来自给定样本的RNA量或产率。在一些实施方案中,使用Agilent 2100生物分析仪测量RNA量,并通过计算的RNA完整指数(RIN,1-10)进行表征。RNA完整性指数(RIN)是用于分配完整值至RNA测量的算法。RIN算法可适用于电泳RNA测量,并且基于贡献关于RNA完整性信息的不同特征的组合来提供更强的通用测量。该算法分配1-10个RIN分数,其中水平为10的RNA质量最高。在一方面,本发明提供了分析来自具有等于或小于6.0、等于或小于5或者等于或小于4的RNA RIN值的样本的基因表达的方法。在一些实施方案中,使用本发明的目标方法和算法分析了包含具有约1.0、2.0、3.0、4.0、5.0或6.0中的任何一个RIN数值的RNA的样本。

[0123] 在一些实施方案中,使用本领域已知的多种方法中的任何一种来测量生物样本中的蛋白含量,包括但不限于:280纳米处的紫外吸光度、利用例如考马斯蓝或二喹啉甲酸(bichichonic acid)的细胞染色或蛋白染色。在一些情况下,在测量样本之前从生物样本提取蛋白。

[0124] 在特定实施方案中,通过本领域已知且可利用的例如从生物样本分离基因产物的任何方法来处理样本。在某些实施方案中,所述基因产物选自核酸(例如,RNA,包括mRNA)和蛋白。

[0125] 细胞学分析

[0126] 可例如通过细胞染色结合显微镜检查生物样本中的细胞进行生物样本的细胞学分析。可通过本领域已知的多种方法和合适的试剂进行细胞染色或细胞学检查,包括但不限于:EA染色、苏木精染色、细胞染色、帕帕尼科拉乌氏染色、伊红、尼氏染色、甲苯胺蓝、银染、偶氮胭脂红、中性红或杰纳斯绿。在一些情况下,例如使用甲醇、乙醇、戊二醛或甲醛,在染色程序之前或期间将细胞固定和/或渗透,而在其他情况下,其未被固定和/或渗透。例如使用染色程序,如利用溴化乙锭、苏木精、尼氏染色或本领域已知的任何核酸染色,来进行

或不进行核酸含量。

[0127] 在一些实施方案中,可通过本领域熟知的细胞学检查的标准方法将细胞涂在载片上。在其他情况下,可使用基于脂质的细胞学(LBC)方法。在LBC方法中,从个体将生物样本转移至容器或小瓶中,其含有液体细胞制备溶液,如Cytoc ThinPrep、SurePath或Monoprep,或者本领域已知的任何其他基于液体的细胞制备溶液。另外,可利用液体细胞制备溶液将样本从收集装置冲洗至容器或小瓶中,以确保基本上定量转移所述样本。然后可保存含有基于液体的细胞制备溶液中的生物样本的溶液,和/或通过仪器或由本领域技术人员进行处理,以在玻璃载片上制备细胞层。可在显微镜下以与常规细胞学制备相同的方式对样本进一步染色和检查。

[0128] 在一些实施方案中,可通过免疫组化染色分析样本。免疫组化染色提供了通过使用生物样本(例如,细胞或组织)中的抗体,分析特定分子或抗原的存在、定位或分布的方法。抗原可为小分子、蛋白、肽、核酸或能够被抗体特异性识别的任何其他分子。可通过有或无预先固定和/或渗透步骤的免疫组化方法来分析样本。在一些情况下,可通过将样本与抗原特异性的抗体相接触来检测目标抗原,然后可通过一次或多次洗涤去除非特异性结合。接着可通过抗体检测试剂如标记的二抗或标记的亲合素/链霉亲和素检测特异性结合的抗体。在一些情况下,可相反直接标记抗原特异性的抗体。适合用于免疫组化的标志物包括但不限于荧光团如荧光素和若丹明、酶如碱性磷酸酶和辣根过氧化物酶,以及放射性核素如 ^{32}P 和 ^{125}I 。可通过免疫组化染色检测的基因产物标志物的实例包括但不限于Her2/Neu、Ras、Rho、EGFR、VEGFR、UbcH10、RET/PTC1、细胞角蛋白20、降钙素、GAL-3、甲状腺过氧化物酶和甲状腺球蛋白。

[0129] 常规细胞学检查的结果可指示生物样本为阴性的(无癌症)、模糊或疑似的(暗示存在癌症)、诊断的(癌症的阳性诊断)或者未诊断的或不确定的(提供关于有或无癌症的信息不充分)。诊断结果可分类为恶性或良性的。在一些情况下,诊断结果可指示特定类型的癌症或病况,如本文中描述的疾病或病况中的任何一种。

[0130] 基因和基因产物

[0131] 在不同实施方案中,本发明的方法包括例如通过确定本文中鉴定的任何基因或基因组表达的基因产物的表达水平,对生物样本进行分子表征。基因产物包括但不限于mRNA和由该基因表达的蛋白。根据本发明的方法确定、测量或分析了表达水平的基因产物(又称为“基因表达产物”)包含以下或由以下组成:由表1中示出的一种或多种、两种或更多种、三种或更多种或者四种或更多种基因,以及上述的变体或同源物(即,来自其他物种的相应基因)表达的基因产物。确定了表达的基因产物可仅为表1中的基因(或者其变体或同源物)表达的基因产物,或者其可包含一种或多种另外的基因产物,包括任何先前与甲状腺癌有关的以及例如PCT专利申请W02011/032296、W02011/143361或W02010/056374中描述的那些。表1提供了人基因的示例性序列的名称和登录号。容易获得来自其他物种的相应基因。

[0132] 表1. 基因

[0133]

基因	全称	其他名称	登录号 RefSeq	
			多核苷酸	多肽
CXCR3 (基因 1)	趋化因子(C-X-C 基序)受体 3	G 蛋白-偶联受体 9 (GPR9)和 CD183	NM_001504.1 (SEQ ID NO:1)	NP_001495.1 (SEQ ID NO:2)
CCR3 (基因 5)	趋化因子(C-C 基 序)受体 3	CD193 (分化簇 193)	NM_001837.3 (SEQ ID NO:3)	NP_001828.1 (SEQ ID NO:4)
CXCL10 (基因 7)	趋化因子(C-X-C 基序)配体 10	干扰素 γ -诱导的蛋白 10 (IP-10)或小型-可 诱导的细胞因子 B10	NM_001565 (SEQ ID NO:5)	NP_001556.2 (SEQ ID NO:6)
CK19 (基因 10)	细胞角蛋白 19	角蛋白, I 型细胞骨架 19 或角蛋白-19 (K19)	NM_002276.4 (SEQ ID NO:7)	NP_002267.2 (SEQ ID NO:8)
TIMP-1 (基因 11)	金属蛋白酶-1 的 组织抑制剂	TIMP 金属肽酶抑制 剂 1	NM_003254 (SEQ ID NO:9)	NP_003245.1 (SEQ ID NO:10)
CLDN-1 (基因 12)	紧密连接蛋白 1 (claudin 1)		NM_021101.4 (SEQ ID NO:11)	NP_066924.1 SEQ ID NO:12)
CAR (基因 13)	柯 萨 奇 (coxsackie) 病毒 和腺病毒受体	CXADR	NM_001207066.1 (SEQ ID NO:13)	NP_001193995.1 SEQ ID NO:14)
XB-130 (基因 15)	肌动蛋白丝结合 蛋白 1-样 2	AFAP1L2	NM_001001936.1 (SEQ ID NO:15)	NP_001001936.1 (SEQ ID NO:16)
HO-1 (基因 16)	血红素氧化酶 (脱环) 1	HMOX1 (血红素氧化 酶(脱环) 1)	NM_002133.2 (SEQ ID NO: 17)	NP_002124.1 SEQ ID NO:18)
CCR7 (基因 17)	趋化因子(C-C 基 序)受体 7	C-C 趋化因子受体类 型 7 或 CD197	NM_001838.3 (SEQ ID NO: 19)	NP_001829.1 (SEQ ID NO:20)

[0134] 在特定实施方案中,本发明的方法包括确定生物样本如甲状腺组织样本中的表1中示出的两种或更多种、三种或更多种、四种或更多种或者五种或更多种基因的基因产物的表达水平。在某些实施方案中,所述方法包括确定选自CXCR3、CCR3、CXCL10、CAR、XB130、HO-1和CCR7基因的至少一种、至少两种或至少三种基因的基因产物的表达水平。在一实施方案中,确定了包含以下或由以下组成的基因产物表达水平: CXCR3、CCR3、CXCL10、CK19、TIMP-1、CLDN-1、CAR、HO-1和CCR7基因的基因产物。在其他实施方案中,确定了包含以下或由以下组成的基因产物表达水平: CCR3、TIMP-1、CAR和XB130基因的基因产物; CXCL10、TIMP-1、CAR和CCR7基因的基因产物; TIMP-1、CAR和CCR7基因的基因产物; 或者CXCL10、TIMP-1、CLDN-1和CCR7基因的基因产物。确定了表达的基因产物的特定集合可称为基因产物“集”,并且确定了表达的特定基因可称为“基因集”。在某些实施方案中,根据本发明的方法,确定了使用的每一个基因集的同类型的基因产物如mRNA或蛋白的表达水平。因此,在特定实施方案中,在对特定组织样本实施本发明的方法时,确定了仅一种类型的基因产物如mRNA或蛋白的表达水平。然而,预期可以实施本发明的方法,其中确定了被使用的基因集中每一个基因的两种或更多种不同类型的基因产物如mRNA和蛋白的表达水平。

[0135] 测量基因产物的水平

[0136] 可通过本领域可用的任何合适的方法,并且主要取决于待测量的基因产物的类型,确定基因产物的表达水平。例如,用于检测基因产物的试剂可测量RNA表达、蛋白表达、

蛋白活性或由本文所述的基因编码的蛋白的下游生物学功能。因此,本发明包括用于检测这样的基因或其基因产物的试剂,包括核酸、DNA探针或结合所编码蛋白的抗体等。

[0137] 可例如通过免疫组化,使用特异性结合蛋白基因产物的抗体来确定蛋白表达水平。在某些实施方案中,使用多肽阵列或抗体阵列来确定蛋白表达水平。某些实施方案可采用标准方法和检测器如采用了成像装置的蛋白印迹和免疫沉淀、酶联免疫吸附测定(ELISA)、流式细胞术和免疫吸附测定(IFA)。这些熟知的方法通常采用特异性结合本发明选定的靶多肽基因产物或者该多肽的独特区域的一种或多种单克隆或多克隆抗体或者其片段,并且通常不显著结合其他多肽。如果抗体或其抗原结合片段以可检测的水平(在例如ELISA测定中)与多肽反应,并且在相似的条件不与非相关的多肽以统计上显著的可检测方式反应,则抗体或其抗原结合片段被认为是与本发明的多肽“特异性结合”、“免疫结合”,和/或“免疫反应”。

[0138] 某些实施方案可采用“阵列”,如“微阵列”。在某些实施方案中,“微阵列”也可称为具有基质结合集合或多种多肽的“肽微阵列”或“蛋白微阵列”,与所述多种结合多肽中的每一种的结合可单独检测。可选地,肽微阵列可具有多种结合剂,包括但不限于单克隆抗体、多克隆抗体、噬菌体展示结合剂、酵母2杂合结合剂以及适配子,其可特异性检测本文所述的多肽基因产物的结合。所述阵列可基于例如Robinson等(Nature Medicine 8(3):295-301(2002))所述多肽的自身抗体检测。肽阵列的实例可见于WO 02/31463、WO 02/25288、WO 01/94946、WO 01/88162、WO 01/68671、WO 01/57259、WO 00/61806、WO 00/54046、WO 00/47774、WO 99/40434、WO 99/39210和WO 97/42507及美国专利6,268,210、5,766,960和5,143,854,其中每一份文献均通过引用并入本文。

[0139] 某些实施方案可采用质谱法(MS)或其他基于分子量的方法来诊断性检测多肽基因产物。MS通常是指用于确定样本或分子的元素组成的分析技术。MS也可用于确定分子如肽和其他化合物的化学结构。示例性的MS仪器具有3个组件:离子源,其将气相样本分子转化为离子(或者,在电喷射离子化的情况下,将溶液中存在的离子移至气相中);质量分析器,其通过施加电磁场按质量分选离子;以及检测器,其测量指示剂量的值并由此提供用于计算存在的每一种离子的丰度的数据。MS技术具有定性和定量应用,包括定量样本中多肽基因产物的量。包括气相色谱-质谱(GC/MS或GC-MS),液相色谱质谱(LC/MS或LC-MS),以及离子迁移色谱/质谱(IMS/MS或IMMS)。因此,可根据本文提供的任何方法使用MS技术,以测量样本中多肽基因产物的水平,以及任选地,将那些水平与对照样本或预定值进行比较。

[0140] 某些实施方案可采用细胞分选或细胞可视化或成像装置/技术来检测或定量样本中基因产物的存在或水平。实例包括流式细胞术或FACS、免疫荧光分析(IFA)和原位杂交技术,如荧光原位杂交(FISH)。

[0141] 在某些实施方案中,本发明的方法包括检测多肽基因产物或者确定或测量多肽基因产物的量的步骤,包括在足够这样的结合发生的条件和时间下,将所述生物样本与结合多肽基因产物的一种或多种探针(例如,多肽或抗体)相接触,然后检测与探针结合的多肽基因产物的量,例如探针和多肽基因产物的复合物的量。可通过本领域已知的多种方法中的任何一种进行检测,并且可采用使用可检测的标志物如免疫荧光部分。在某些实施方案中,探针被可检测地标记。在某些实施方案中,检测在溶液中或在固相载体上结合的多肽基因产物。

[0142] 本发明的方法可包括,例如在确定样本中的特定基因产物的量之前或期间,从生物样本分离一种或多种基因产物的步骤。在某些实施方案中,被检测的基因产物为多肽,并且从生物样本纯化或部分纯化多肽。在某些实施方案中,被检测的基因产物为mRNA,并且从生物样本纯化或部分纯化多核苷酸或mRNA。从生物样本纯化或部分纯化多肽或多核苷酸的方法为本领域已知的。

[0143] 可使用多种不同的测定来确定核酸表达水平,包括但不限于扩增测定和杂交测定。可用于本发明的扩增测定包括但不限于聚合酶链式反应(PCR)测定,包括逆转录酶-PCR(RT-PCR)和实时PCR,以及等温扩增法。杂交测定包括但不限于Northern印迹、定量或定性聚合酶链式反应(PCR)、定量或定性逆转录酶PCR(RT-PCR)、微阵列、斑点或狭缝印迹,以及原位杂交如荧光原位杂交(FISH)。

[0144] 某些实施方案可采用杂交法来测量多核苷酸基因产物如mRNA的表达。用于进行多核苷酸杂交测定的方法在本领域已得到良好开发。杂交测定方案和条件根据应用而不同,并且根据已知的常规结合方法进行选择,包括如下文献中提及的那些方法:Maniatis et al. *Molecular Cloning: A Laboratory Manual* (2nd Ed. Cold Spring Harbor, N.Y., 1989); Berger and Kimmel *Methods in Enzymology*, Vol. 152, *Guide to Molecular Cloning Techniques* (Academic Press, Inc., San Diego, Calif., 1987); Young and Davis, *PNAS*. 80:1194 (1983)。用于进行重复和受控杂交反应的方法和装置在美国专利5,871,928、5,874,219、6,045,996和6,386,749、6,391,623中已有描述,其中每一份文献均通过引用并入本文。

[0145] 某些实施方案可采用分子条形码技术,如可从NanoString Technologies Inc (Seattle, WA, USA) 商购的商标名为nCounter基因表达测定的技术。这一技术允许从复杂的样本如组织溶解物直接测定靶标表达水平,而不需要扩增转录物。在这一方法中,生物素化的捕获探针和有条形码的报告探针(有色码)直接与目的基因特异性杂交。在去除过量未结合的探针后,将探针/靶标复合物固定,并在nCounter盒上比对,然后获取并定量数字图像,每一个条形码计数代表相应靶基因的转录物拷贝。重要的是,该条形码方法的使用允许以非常高的重复性在单一样本反应中多重检测超过500个基因。

[0146] 某些实施方案可使用转录物分析和亲和捕获(TRAC)法来测量RNA表达,其中可直接对细胞溶解物或纯化的总RNA样本进行表达的RNA种类的多重检测,而不需要任何逆转录或扩增。在这一新的方案中,生物素化的寡-dT和双-荧光团-标记的基因特异性探针(每一个探针具有明显不同的大小)与给定样本中的mRNA杂交。将杂交的物质与链霉亲和素磁珠相结合,洗涤,释放,然后通过毛细管电泳溶解。因此,通过分析探针/靶标大小和荧光图谱而同时实现了靶标鉴定和定量。

[0147] 某些实施方案可采用定量核酸酶保护测定(qNPA)或qNPA微阵列,如可从高通量基因组分子诊断公司(High Throughput Genomics Molecular Diagnostics Inc) (Tucson, AZ, USA) 商购的那些,其中将基因特异性探针与表达的RNA杂交来使转录物/探针复合物免受单链特异性S1核糖核酸酶消化。在S1处理和碱基介导的探针/靶标复合物解离后,剩下的探针与其相应的靶标序列以1:1的化学计量存在,并因而通过在微阵列表面上捕获而定量探针,允许推断相应的靶标序列表达水平。

[0148] 某些实施方案可采用核酸扩增方法来检测基因产物。术语“扩增”或“核酸扩增”是

指产生靶核酸序列的至少一部分的多个拷贝。所述多个拷贝可被称为扩增子或扩增产物。如本文所用的“选择性扩增”或“特异性扩增”是指根据本发明靶核酸序列的扩增,其中靶序列的可检测扩增基本上限于来源于测试的目标核酸样本的靶序列扩增,而非来源于一些其他样本来源的靶核酸序列,例如扩增反应中使用的试剂中或者进行扩增反应的环境中存在的污染物。

[0149] 术语“扩增条件”是指根据本发明允许核酸扩增的条件。用于本发明的扩增反应的寡核苷酸与其预期的靶标在扩增条件下杂交。本领域技术人员根据采用的特定扩增方法,可容易地确定根据本发明进行核酸扩增的可接受条件。

[0150] 许多熟知的核酸扩增方法需要热循环以交替变性双链核酸和杂交引物;然而,其他熟知的核酸扩增方法是等温的。聚合酶链式反应(美国专利4,683,195;4,683,202;4,800,159;4,965,188),通常称为PCR,使用多个循环的变性、退火引物对至相反链,以及引物延伸,从而指数增加靶序列的拷贝数。在称为RT-PCR的变型中,逆转录酶(RT)被用于由mRNA产生互补DNA(cDNA),然后通过PCR扩增cDNA以产生多个拷贝的DNA。

[0151] 如上文所示,术语“PCR”是指选择性扩增靶核酸种类的多个扩增循环。包括定量PCR(qPCR)、实时PCR、逆转录PCR(RT-PCR)和定量逆转录PCR(qRT-PCR),其在本领域已得到很好的描述。术语“qPCR”是指定量聚合酶链式反应,而术语“qRT-PCR”是指定量逆转录聚合酶链式反应。qPCR和qRT-PCR可用于扩增和同时定量靶cDNA分子。其使得能够检测和定量cDNA库中的具体序列。然后可通过多种方法,例如通过染色直接在凝胶中成像扩增产物,所述产物可通过利用可检测的探针杂交,和/或通过使用二代测序进行检测。

[0152] “实时PCR”可使用DNA结合染料来结合PCR反应混合物中的双链(ds)DNA,产生染料荧光。因此,PCR期间DNA产物的增加引起荧光强度的增加,并且在每个循环进行测量,从而允许定量DNA浓度。dsDNA染料如SYBR Green将结合所有dsDNA PCR产物。某些实施方案可使用Taqman探针,其在5'端用荧光团标记,而在3'端用淬灭剂标记,且被设计为在所需扩增子的正向和反向引物结合位点之间退火。只要淬灭剂保持与荧光团紧密靠近,就不发射荧光。然而,在PCR循环中,当聚合酶延伸引物时,结合探针的荧光团和淬灭剂通过聚合酶的5'-3'核酸外切酶活性的方式被切割,并发射荧光。因此,样本中存在的转录物的量与检测的荧光量成正比,并且PCR循环后转录物数目的增加引起发射荧光相应增加。在实时PCR热循环仪中检测和测量荧光,并且其对应于产物指数增加的几何增加被用于确定每个反应中的阈值循环(“Ct”)。

[0153] 术语“Ct分数”是指阈值循环数,其为PCR扩增超过阈值水平时的循环。如果样本中特定基因存在较高数量的mRNA,其将比低表达的基因更早跨过阈值,因为存在更多的起始RNA用于扩增。因此,低Ct分数指示样本中的高基因表达,并且高Ct分数指示低基因表达。不典型CT值为大于给定基因的平均CT值的两个标准差的值。

[0154] 在某些实施方案中,所述扩增方法可采用基于多重PCR的技术,在本文中定义为在单一反应中检测两种或更多种靶序列的方法。所述扩增方法的其他实施方案可采用微流体控制热循环过程,或者精确控制添加的试剂量和时间,或者在其他实施方案中,允许将所述扩增方法调整为代替个性化诊断的适合所述方法临床应用的便携式试剂盒。

[0155] 某些实施方案可以阵列形式采用基于扩增的检测方法,例如可通过商购获得的Qiagen(Hilden,Germany)的商标名RT² Profiler PCR Array,或者Lonza(Basel,

Switzerland) 的商标名为StellArray, 或者Life Technologies (Carlsbad, CA, USA) 的商标名为OpenArray。在这些方法中, 可在高密度96孔板、384孔板、100孔盘或48孔芯片中同时运行独立的qPCR反应, 从而允许以qPCR的所有优势和高通量的附加优势定量标准或定制的基因组。

[0156] 某些实施方案可采用连接酶链式反应(Weiss, Science. 254:1292, 1991), 通常称为LCR, 其使用与靶核酸的邻近区域杂交的两组互补DNA寡核苷酸。DNA寡核苷酸通过DNA连接酶在热变性、杂交和连接的重复循环中共价连接, 从而产生可检测的双链连接的寡核苷酸产物。

[0157] 另一种方法为链置换扩增(Walker, G. et al., 1992, Proc. Natl. Acad. Sci. USA 89:392-396; 美国专利第5,270,184和5,455,166号), 通常称为SDA, 其使用以下循环: 将引物序列对与靶序列的互补链退火、引物在dNTPs存在下延伸以产生双链半硫代磷酸化的引物延伸产物、核酸内切酶介导切开半修饰的限制性核酸内切酶识别位点, 以及聚合酶介导的引物从切口3'端延伸以取代存在的链并产生用于下一轮引物退火、切开和链置换的链, 导致产物几何扩增。耐热SDA (tSDA) 在基本上相同的方法中在较高的温度下使用耐热核酸内切酶和聚合酶(欧洲专利第0684315号)。

[0158] 其他扩增方法包括例如: 基于核酸序列的扩增(美国专利5,130,238), 通常称为NASBA; 其使用RNA复制酶扩增探针分子本身(Lizardi, P. et al., 1988, BioTechnol. 6: 1197-1202), 通常称为Q β 复制酶; 基于转录的扩增方法(Kwoh, D. et al., 1989, Proc. Natl. Acad. Sci. USA 86:1173-1177); 自我维持的序列复制(Guatelli, J. et al., 1990, Proc. Natl. Acad. Sci. USA 87:1874-1878); 以及转录介导的扩增(美国专利5,480,784和5,399,491), 通常称为TMA。已知扩增方法的进一步讨论参见Persing, David H., 1993, "In Vitro Nucleic Acid Amplification Techniques", Diagnostic Medical Microbiology: Principles and Applications (Persing et al., Eds.), pp.51-87 (American Society for Microbiology, Washington, DC)。

[0159] 可根据本发明使用的基于示例性转录的扩增系统包括转录介导的扩增(TMA), 其采用RNA聚合酶产生靶区域的多种RNA转录物(美国专利5,480,784和5,399,491)。TMA使用在逆转录酶和RNA聚合酶存在下与靶核酸杂交的“启动子-引物”来形成双链启动子, RNA聚合酶可从中产生RNA转录物。在能够与RNA转录物杂交的第二引物存在下, 这些转录物可成为其他轮TMA的模板。不同于PCR、LCR或其他需要热变性的方法, TMA为等温方法, 其利用RNase H活性消化RNA:DNA杂合体的RNA链, 从而使得DNA链可用于与引物或启动子-引物杂交。通常, 使用提供用于扩增的与逆转录酶结合的RNase H活性。

[0160] 在某些实施方案中, 其他技术可用于确定多核苷酸基因产物的表达, 包括微阵列分析(Han, M., et al., Nat Biotechnol, 19:631-635, 2001; Bao, P., et al., Anal Chem, 74: 1792-1797, 2002; Schena et al., Proc. Natl. Acad. Sci. USA 93:10614-19, 1996; 和Heller et al., Proc. Natl. Acad. Sci. USA 94:2150-55, 1997) 以及SAGE(基因表达的系列分析)。如同MPSS, SAGE为数字并且可生成大量的标签序列(参见例如Velculescu, V.E., et al., Trends Genet, 16:423-425., 2000; Tuteja R. and Tuteja N. Bioessays. 2004 Aug; 26(8): 916-22), 尽管数量级比可从诸如MPSS的技术获得的小。

[0161] 在某些实施方案中, 术语“微阵列”包括具有与基质结合的多种核酸的“核酸微阵

列”，可分别检测与所述多种结合核酸中的每一种的杂交。基质可为固相或多孔的、平面或非平面的、一致或分散的。核酸微阵列包括Schena (ed.), DNA Microarrays: A Practical Approach (Practical Approach Series), Oxford University Press (1999); Nature Genet. 21 (1) (suppl.): 1-60 (1999); Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing Company/BioTechniques Books Division (2000) 中所述的所有装置。核酸微阵列可包括与基质结合的多种核酸，其中所述多种核酸被置于多个珠子上，而非一致的平面基质上如，例如Brenner et al., Proc. Natl. Acad. Sci. USA 97 (4): 1665-1670 (2000) 所述。核酸微阵列的实例可见于美国专利6,391,623、6,383,754、6,383,749、6,380,377、6,379,897、6,376,191、6,372,431、6,351,712、6,344,316、6,316,193、6,312,906、6,309,828、6,309,824、6,306,643、6,300,063、6,287,850、6,284,497、6,284,465、6,280,954、6,262,216、6,251,601、6,245,518、6,263,287、6,251,601、6,238,866、6,228,575、6,214,587、6,203,989、6,171,797、6,103,474、6,083,726、6,054,274、6,040,138、6,083,726、6,004,755、6,001,309、5,958,342、5,952,180、5,936,731、5,843,655、5,814,454、5,837,196、5,436,327、5,412,087和5,405,783中，其公开内容通过引用并入本文。

[0162] 另外的实例包括可从Affymetrix (Santa Clara, Calif.) 商购获得，商标名为GENECHIP™或Illumina (San Diego, CA) 的核酸阵列。制备和使用阵列的其他示例性方法在例如美国专利7,028,629; 7,011,949; 7,011,945; 6,936,419; 6,927,032; 6,924,103; 6,921,642; 和6,818,394中提供。

[0163] 本发明涉及阵列和微阵列时，还预期用于附着至固体基质的聚合物的许多应用。示例性的基因表达监测和谱图分析的方法以及在基因表达监测和谱图分析中有用的方法描述于美国专利5,800,992、6,013,449、6,020,135、6,033,860、6,040,138、6,177,248和6,309,822中。基因分型及其应用在美国系列号10/442,021、10/013,598 (美国申请2003/0036069)，以及美国专利5,925,525、6,268,141、5,856,092、6,267,152、6,300,063、6,525,185、6,632,611、5,858,659、6,284,460、6,361,947、6,368,799、6,673,579和6,333,179中示出。可用于与本文公开的方法相组合的其他核酸扩增、标记和分析方法描述于美国专利5,871,928、5,902,723、6,045,996、5,541,061和6,197,506中。

[0164] 某些实施方案可采用RNA测序，如全转录组鸟枪法测序 (WTSS)，通常称为RNA-Seq，来分析RNA表达，其中可通过使用深度测序技术组装具有单核苷酸分辨率的表达谱的转录组图。在一实施方案中，RNA样本如mRNA被转换为片段化的cDNA，然后连接至测序适配子。高通量测序允许产生数百万的短读序列，其可以从头组装或者针对已知基因组或参照序列比对。然后可将单个读序记录读序总数的比率用于生成表达谱。

[0165] 在某些实施方案中，本发明的方法包括检测多核苷酸基因产物 (例如，mRNA) 或确定或测量多核苷酸基因产物的量的步骤，包括在足够这样的结合发生的条件和时间下，使生物样本与一个或多个结合多核苷酸基因产物的探针 (例如，引物或多核苷酸) 相接触，然后检测与探针结合的多核苷酸基因产物的量，例如探针和多核苷酸基因产物的复合物的量。可通过本领域已知的多种方法中的任何一种进行检测，并且可使用可检测的标记，如免疫荧光部分。在某些是实施方案中，探针被可检测地标记。在某些实施方案中，在溶液中或在固相载体上检测结合的多核苷酸基因产物。

[0166] 用于检测基因产物的试剂

[0167] 如对本领域技术人员显而易见的,某些实施方案可采用检测基因产物的试剂,如抗体和寡核苷酸,包括本文所述的引物或探针。在特定实施方案中,用于检测基因产物的试剂特异性结合或特异性杂交靶基因产物,而非不相关的基因产物,例如由不同的不相关基因表达的基因产物。制备特异性结合或特异性杂交靶多肽或核酸序列的试剂如抗体和寡核苷酸的方法为本领域已知。例如,可通过本领域技术人员已知的多种技术中的任何一种来制备抗体。参见,例如Harlow和Lane (Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, 1988)。可以例如,使用Kohler和Milstein (Eur. J. Immunol. 6: 511-519, 1976)的技术及其改进技术来制备目标多肽特异性的单克隆抗体。

[0168] 关于单链核酸,特别是寡核苷酸,术语“特异性杂交”是指足够互补序列的两个单链核苷酸分子之间的结合,从而允许发生在本领域通常使用的预定条件下这样的杂交(有时称为“基本上互补”)。特别地,在一实施方案中,该术语是指寡核苷酸与单链DNA或RNA分子中包含的基本上互补序列的杂交,基本上排除寡核苷酸与非互补序列的单链核酸的杂交。例如,特异性杂交可指在合适条件下与基因产物杂交的序列,使不同互补性的单链核酸分子能够特异性杂交的合适条件是为本领域熟知的。

[0169] 在一实施方案中,通过测量RNA转录的相对比率,如通过制备相应的cDNA然后使用探针(如从表1中鉴定的基因序列开发的那些)来分析所得到的DNA,确定了生物样本中基因产物的表达水平。因此,通过使用逆转录酶利用生物样本的RNA产生的cDNA水平产生相当量的cDNA,其随后可使用聚合酶链式反应或一些其他方法来进行扩增,从而确定所得到的cDNA相对水平,并由此确定基因表达的相对水平。

[0170] 检测核酸基因产物的示例性试剂包括核酸,并且尤其包括寡核苷酸。核酸可为DNA或RNA,并且可为单链或双链。在一实施方案中,寡核苷酸为DNA探针,或者为用于扩增由靶基因产生的核酸的引物。在一实施方案中,本发明的寡核苷酸能够与本文所述的基因产物特异性杂交(例如,在中度或严格杂交条件下)。寡核苷酸可为天然存在或合成的,但通常通过合成方法制备。如本文所述的寡核苷酸可包括DNA片段或其互补物。寡核苷酸的准确大小取决于不同因素和特定应用以及寡核苷酸的使用。寡核苷酸包括探针和引物,其可为靶基因产物(如从表1中提供的基因表达的那些)的3个核苷酸至全长的任意长度,并且明确地包括靶基因产物的3至全长的每一个可能的连续核酸数目。因此,寡核苷酸可为5-100个连续的碱基,并且范围通常为5、10、12、13、14、15、16、17、18、19、20、21、22、23、24或25个核苷酸至30、35、40、45、50、55、60、65、70、75、80、85、90、95或100个核苷酸。5-10、5-20、10-20、12-30、15-30、10-50、20-50或20-100个碱基长度的寡核苷酸也很常见。

[0171] 本发明的寡核苷酸可为RNA、DNA或二者的衍生物。这样的寡核苷酸的最小为在本发明的核酸分子(例如,表达的基因产物或由扩增所产生的cDNA或其拷贝)上形成寡核苷酸与互补序列之间的稳定杂交所需的大小。本发明包括这样的寡核苷酸,其可用作例如鉴定核酸分子的探针(例如,DNA探针)或扩增核酸分子的引物。

[0172] 在一实施方案中,寡核苷酸可为探针,其指例如纯化的限制酶消化物中天然存在的或合成制备的核苷酸、多核苷酸或核酸,即RNA或DNA,其能够与核酸退火或特异性杂交,所述核酸具有与探针互补的序列。探针可为单链或双链。探针的准确长度取决于多种因素,包括温度、探针来源和使用方法。例如,对于诊断应用,根据靶序列的复杂性,寡核苷酸探针

通常包含15-25个或更多个核苷酸,尽管其可包含更少的核苷酸。在某些实施方案中,探针可为5-100个连续碱基,并且通常为约5、10、12、13、14、15、16、17、18、19、20、21、22、23、24或25个核苷酸长度,或者可为约30、35、40、45、50、55、60、65、70、75、80、85、90、95或100个核苷酸长度。本文的探针经选择与特定靶核酸序列的不同链互补。这意味着探针必须足够互补,以便能够在预定条件下与其各自的靶链特异性杂交或退火。因此,探针序列不需要反映靶标的准确互补序列。例如,非互补核苷酸片段可与探针的5'或3'端结合,探针序列的剩下部分与靶链互补。可选地,可将非互补碱基或较长的序列散布在探针中,只要所述探针序列与靶核酸序列具有足够的互补性,从而与其特异性退火。

[0173] 在一实施方案中,寡核苷酸可为引物,其是指寡核苷酸,即单链或双链RNA或DNA,来源于生物系统,通过限制酶消化产生或合成制备,当置于合适的环境中时,能够在功能上用作依赖于模板的核酸合成的起始物。当提供合适的核酸模板、核酸合适的核苷三磷酸前体、聚合酶、合适的辅因子和条件如合适的温度和pH时,可通过聚合酶或类似活性作用在其3'端添加核苷酸来延长引物,从而产生引物延伸产物。引物长度可根据特定条件和应用要求而不同。例如,在某些应用中,寡核苷酸引物为约15-25个或更多个核苷酸长度,但在某些实施方案中,可为5-100个连续碱基,并且通常为约5、10、12、13、14、15、16、17、18、19、20、21、22、23、24或25个核苷酸长度,或者在某些实施方案中,可为约30、35、40、45、50、55、60、65、70、75、80、85、90、95或100个核苷酸长度。引物必须与所需的模板具有足够的互补性,从而引发所需的延伸产物的合成,即,能够以这样的方式与所需的模板链退火,其足以以合适的并置提供引物的3'羟基部分,用于起始聚合酶或类似酶的合成。不需要引物序列代表所需模板的准确互补。例如,非互补核苷酸序列可与另外互补引物的5'端结合。可选地,非互补碱基可散布于寡核苷酸引物序列中,前提是引物序列与所需模板链的序列具有足够的互补性,从而在功能上提供用于合成延伸产物的模板-引物复合物。

[0174] 在某些实施方案中,用于确定基因产物表达的试剂包含两种、三种或更多种寡核苷酸的组,其中该组的寡核苷酸中的每一种与本文所述的基因产物或所述基因产物的互补链杂交。因而,例如,每条寡核苷酸可与mRNA基因产物和/或由mRNA产生的cDNA的一条或多条链杂交。在一实施方案中,一组寡核苷酸包含DNA探针。在某些实施方案中,一组寡核苷酸包含至少两条扩增引物或PCR引物,其一起能够扩增靶核酸序列的至少一部分,例如mRNA基因产物或产生的cDNA。在另一实施方案中,所述寡核苷酸或DNA探针组可提供于阵列如固相阵列、染色体/DNA微阵列或微珠阵列上。阵列技术为本领域所熟知且在本文中所述。

[0175] 可通过本领域技术人员已知的技术,如通过化学或生物化学合成,以及通过来自重组核酸分子如细菌或病毒载体的体外或体内表达,产生具有确定序列和化学结构的寡核苷酸。在某些实施方案中,寡核苷酸并非仅仅由野生型染色体DNA或其体内转录产物组成。

[0176] 在某些实施方案中,本发明提供了分离的多核苷酸,例如引物或探针,其包括与本文所述的基因或多核苷酸基因产物如mRNA相同或互补的序列的不同长度的连续延伸。

[0177] 可以任何方式修饰寡核苷酸、探针或引物,只要给出的修饰与给定寡核苷酸的所需功能兼容。本领域技术人员可容易地确定给定的修饰是否为本发明任何给定寡核苷酸所适合或需要。

[0178] 虽然寡核苷酸的设计和序列依赖于其如本文所述的功能,通常要考虑到几个变量。最为相关的为:长度、解链温度(T_m)、特异性、与系统中其他寡核苷酸的互补性、G/C含

量、聚嘧啶 (T,C) 或聚嘌呤 (A,G) 延伸, 以及3' 端序列。这些和其他变量的控制是标准化的并且为寡核苷酸设计的熟知方面, 并且可容易获取多种计算机程序来筛选大量潜在寡核苷酸中的最优化寡核苷酸。

[0179] 如技术人员所认可的, 多核苷酸可为单链 (编码链或反义链) 或者双链的, 并且可为DNA (基因组、cDNA或合成的) 或者RNA分子。另外的编码或非编码序列可以但不必存在于本发明的多核苷酸中, 并且多核苷酸可以但不必与其他分子和/或支撑物质相连接。

[0180] 本发明包括使用能在适合进行本文所述的测定中的任何一种的条件下特异性结合靶基因产物的寡核苷酸, 包括但不限于PCR、RT-PCR和实时PCT。合适的条件为本领域已知, 并且包括用于根据PCR仪和试剂零售商提供的指导进行的标准反应的那些条件。在某些实施方案中, 本发明包括多核苷酸, 包括在下文所述的严格条件下与参照核苷酸序列 (例如, 靶基因产物) 或其互补物杂交的寡核苷酸。如本文所用, 术语“在低严格、中度严格、高严格或非常高严格条件下杂交”描述杂交和洗涤条件。进行杂交反应的指导可见于Ausubel et al., (1998, 同上) 第6.3.1-6.3.6部分。水性和非水性方法在该参考文献中有描述, 并且均可使用。

[0181] 本文中提及低严格条件包括和包含: 对于42℃下杂交, 至少约1% v/v至至少约15% v/v甲酰胺, 和至少约1M至至少约2M盐, 以及对于42℃下洗涤, 至少约1M至至少约2M盐。低严格条件还可包括对于65℃下杂交, 1%牛血清白蛋白 (BSA), 1mM EDTA, 0.5M NaHPO₄ (pH 7.2), 7% SDS, 以及对于室温下洗涤, (i) 2×SSC, 0.1% SDS; 或 (ii) 0.5% BSA, 1mM EDTA, 40mM NaHPO₄ (pH 7.2), 5% SDS。低严格条件的一个实施方案包括约45℃下在6×氯化钠/柠檬酸钠 (SSC) 中杂交, 然后在至少约50℃ (对于低严格条件, 洗涤温度可增加至55℃) 下于0.2×SSC, 0.1% SDS中洗涤两次。

[0182] 中度严格条件包括和包含: 对于42℃下杂交, 至少约16% v/v至至少约30% v/v甲酰胺, 和至少约0.5M至至少约0.9M盐, 以及对于55℃下洗涤, 至少约0.1M至至少约0.2M盐。中度严格条件还可包括对于65℃下杂交, 1%牛血清白蛋白 (BSA), 1mM EDTA, 0.5M NaHPO₄ (pH 7.2), 7% SDS, 以及对于60-65℃下洗涤, (i) 2×SSC, 0.1% SDS; 或 (ii) 0.5% BSA, 1mM EDTA, 40mM NaHPO₄ (pH 7.2), 5% SDS。中度严格条件的一个实施方案包括约45℃下于6×SSC中杂交, 然后60℃下在0.2×SSC, 0.1% SDS中进行一次或多次洗涤。高严格条件包括和包含: 对于42℃下杂交, 至少约31% v/v至至少约50% v/v甲酰胺, 和约0.01M至约0.15M盐, 以及对于55℃下洗涤, 约0.01M至约0.02M盐。

[0183] 高严格条件也可包括对于65℃下杂交, 1% BSA, 1mM EDTA, 0.5M NaHPO₄ (pH 7.2), 7% SDS, 以及对于超过65℃的温度下的洗涤, (i) 0.2×SSC, 0.1% SDS; 或者 (ii) 0.5% BSA, 1mM EDTA, 40mM NaHPO₄ (pH 7.2), 1% SDS。高严格条件的一个实施方案包括约45℃下于6×SSC中杂交, 然后65℃下在0.2×SSC, 0.1% SDS中进行一次或多次洗涤。非常高严格条件的一个实施方案包括65℃下在0.5M磷酸钠, 7% SDS中杂交, 然后65℃下在0.2×SSC, 1% SDS中进行一次或多次洗涤。

[0184] 其他严格条件为本领域熟知, 且技术人员认可可操作不同的因素来优化杂交特异性。最终洗涤的严格性的优化可用于确保高度杂交。对于详细的实例, 参见Ausubel等 (同上文, 2.10.1-2.10.16页) 以及Sambrook等 (同上文, 1.101-1.104节, 1989)。

[0185] 虽然通常在约42℃-68℃的温度下进行严格洗涤, 本领域技术人员理解其他温度

可适合于严格条件。形成DNA-DNA杂合物的最大杂交率通常发生在约 T_m 20℃-25℃以下。本领域熟知 T_m 为解链温度,或者两条互补多核苷酸序列解离的温度。估算 T_m 的方法为本领域所熟知的(参见Ausubel等,同上文,2.10.8页)。

[0186] 一般而言,可通过下式近似预测DNA的完全匹配双链的 T_m : $T_m = 81.5 + 16.6 (\log 10M) + 0.41 (\%G+C) - 0.63 (\% \text{甲酰胺}) - (600/\text{长度})$, 其中: M 为 Na^+ 浓度,优选在0.01摩尔至0.4摩尔范围内; $\%G+C$ 为鸟苷和胞嘧啶碱基的总和占碱基总数的百分比,范围为30%至75% $G+C$; $\% \text{甲酰胺}$ 为按体积计的百分比甲酰胺浓度;长度为DNA双链中的碱基对数目。随机错配的碱基对数目每增加1%,双链DNA的 T_m 降低约1℃。对于高严格性,通常在 $T_m-15^\circ C$ 下进行洗涤,或者对于中度严格性在 $T_m-30^\circ C$ 下进行洗涤。

[0187] 本发明的寡核苷酸、引物和探针可包含这样的区域或由这样的区域组成其为靶基因或mRNA序列或者其互补序列的一部分的多核苷酸变体。在特定实施方案中,寡核苷酸可包含这样的序列或由这样的序列组成,其与靶基因或mRNA序列或其互补序列区域具有至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或100%的同源性。

[0188] 可使用本领域已知的软件来选择用作引物或探针的寡核苷酸。例如,OLIGO 4.06软件可用于选择每条多达100个核苷酸的PCR引物对,以及用于从输入的多达32千碱基的多核苷酸序列分析寡核苷酸和较大的多达5,000个核苷酸的多核苷酸。类似的引物选择程序并入了另外的特征用于扩充能力。例如,PrimOU引物选择程序(公众可从德克萨斯州达拉斯的德克萨斯大学西南医疗中心的基因组中心获取)能够从兆碱基序列中选择特定引物,并因此可用于在基因组范围设计引物。Primer3引物选择程序(公众可从怀特黑德研究所/MIT基因组中心(Cambridge Mass.)获取)允许用户输入“错误引发文库”,其中避免作为引物结合位点的序列为用户指定的。Primer3尤其可用于选择微阵列的寡核苷酸。(最后两个引物选择程序的源代码也可从它们各自的来源获取并经调整以满足用户的特定要求)。PrimeGen程序(公众可从UK人类基因组作图工程资源中心(Cambridge UK)获取)基于多重序列比对设计引物,从而允许选择与比对核酸序列的最保守或最不保守区域杂交的引物。因此,这一程序可用于确定独特和保守的寡核苷酸和多核苷酸片段。寡核苷酸选择的方法不限于本文所述的那些方法。

[0189] 在某些实施方案中,可通过逐步固相合成,采用上文引用的参考文献中详细描述以及下文关于具有混合物或不带电的和阳离子骨架连接的寡核苷酸的合成方法,来制备寡核苷酸。在一些情况下,添加另外的化学部分至寡核苷酸,例如以增加药代动力学或促进化合物的捕获或检测可能是可取的。可根据标准合成方法将这样的部分共价连接,通常连接至寡聚物的末端。例如,添加聚乙二醇部分或其他亲水聚合物,如具有10-100个单体亚基的聚合物,可用于增加溶解性。一个或多个带电基团,例如带负电荷的基团如有机酸,可增加细胞摄取。

[0190] 多种可检测的分子可用于使寡核苷酸或蛋白成为可检测的(即,可检测地标记的),如放射性同位素、荧光染料、染料、酶、纳米粒子、化学发光标记物、生物素或本领域已知的可直接(例如,通过发光)或间接(例如,通过荧光标记抗体的结合)检测的其他单体。

[0191] 放射性同位素提供了可用于本发明的某些实施方案的可检测的分子的实例。几种放射性同位素可用作标记核苷酸或蛋白的可检测分子,包括例如 ^{32}P 、 ^{33}P 、 ^{35}S 、 3H 和 ^{125}I 。这些放射性同位素具有不同的半衰期、衰变类型和能量水平,其可被调整以匹配特定方案

的需要。例如,3H为产生低背景水平的低能发射物,然而该低能量也引起长时间段的放射自显影。放射性标记的核糖核苷酸、脱氧核糖核苷酸和氨基酸可商购获得。可获得在第一个或 α 磷酸基,或者第三个或 γ 磷酸基处放射性标记的核苷酸。例如,[α -32P]dATP和[γ -32P]dATP均可商购获得。另外,具有不同特异性活性的放射性标记的核苷酸的也可商购获得,且可被调整用于不同的方案。

[0192] 可用于检测寡核苷酸的可检测分子的其他实例包括荧光团。几种荧光团可用于标记核苷酸,包括例如荧光素、四甲基若丹明、德克萨斯红和多种其他物质(例如,Haugland, Handbook of Fluorescent Probes-9th Ed.,2002,Molec.Probes,Inc.,Eugene OR; Haugland, The Handbook: A Guide to Fluorescent Probes and Labeling Technologies-10th Ed.,2005,Invitrogen,Carlsbad,CA)。

[0193] 作为一个实例,可在化学合成期间将寡核苷酸荧光标记,因为核苷酸合成期间掺入胺或硫醇允许添加荧光团。荧光标记的核苷酸可商购获得。例如,可获得与覆盖光谱的10种不同荧光团缀合的尿苷和脱氧尿苷三磷酸酯。可直接结合核苷酸的荧光染料也可用作可检测的分子。例如,FAM、JOE、TAMRA和ROX为胺反应性荧光染料,其已与核苷酸结合并被用于自动化DNA测序。这些荧光标记的核苷酸如ROX-ddATP、ROX-ddCTP、ROX-ddGTP和ROX-ddUTP,可商购获得。

[0194] 非可放射性和非可荧光检测的分子也可适用。如上文所示,可将生物素直接与核苷酸连接并通过与亲和素或链霉亲和素的特异性和高亲和力结合来进行检测,所述亲和素或链霉亲和素与催化显色反应的酶(如磷酸酶、荧光素酶或过氧化物酶)相结合。地高辛标记的核苷酸也可类似地用于核酸的非同位素检测。生物素化的和地高辛标记的核苷酸可商购获得。

[0195] 非常小的粒子,称为纳米粒子,也可用于标记寡核苷酸探针。这些粒子大小范围为1-1000nm,并且包括不同的化学结构,如金和银颗粒以及量子点。当用成角度的入射白光辐照时,40-120nm范围的银或金纳米粒子将以高强度散射单色光。散射光的波长取决于颗粒的大小。紧密靠近的4-5种不同的粒子将各自散射单色光,其在叠加时将给出特定独特的颜色。所述粒子由公司如Genicon Sciences (Carlsbad,CA) 制造。可将衍生的银或金颗粒附着于大范围的分子阵列,包括蛋白、抗体、小分子、受体配体和核酸。例如,可对粒子表面化学衍生以允许附着至核苷酸。

[0196] 可用于检测可检测的分子的其他类型的纳米粒子包括量子点。量子点为发荧光的晶体,直径1-5nm,易被大范围波长的光激发。在通过具有合适波长的光激发后,这些晶体发射光,如单色光,波长取决于其化学组成和大小。量子点如CdSe、ZnSe、InP或InAs,具有独特的光学性能;这些和类似的量子点可从众多商业来源(例如,NN-Labs,Fayetteville,AR; Ocean Nanotech,Fayetteville,AR; Nanoco Technologies,Manchester,UK; Sigma-Aldrich,St.Louis,MO) 获得。

[0197] 在某些实施方案中,可用一种或多种发射光的或者可检测的染料标记寡核苷酸引物或探针。染料发射的光可为可见光或不可见光,如紫外光或红外光。在示例性实施方案中,所述染料可为荧光共振能量转移(FRET)染料;咕吨染料,如荧光素和若丹明;在 α 或 β 位置具有氨基的染料(如萘胺染料、1-二甲基氨基萘基-5-磺酸盐、1-苯胺基-8-萘磺酸盐和2-对甲苯胺基-6-萘磺酸盐);染料,其具有3-苯基-7-异氰酰香豆素;吡啶,如9-异硫代氰酰吡

啉和吡啶橙;芘、苯恶二唑和芘;染料,其具有3-(ε-羧基戊基)-3'-乙基-5,5'-二甲氧羰花青(CYA);6-羧基荧光素(FAM);5&6-羧基若丹明-110(R110);6-羧基若丹明-6G(R6G);N,N,N',N'-四甲基-6-羧基若丹明(TAMRA);6-羧基-X-若丹明(ROX);6-羧基-4',5'-二氯-2',7'-二甲氧基荧光素(JOE);ALEXA FLUOR™;Cy2;德克萨斯红和若丹明红;6-羧基-2',4,7,7'-四氯荧光素(TET);6-羧基-2',4,4',5',7,7'-六氯荧光素(HEX);5-羧基-2',4',5',7'-四氯荧光素(ZOE);NAN;NED;Cy3;Cy3.5;Cy5;Cy5.5;Cy7;和Cy7.5;IR800CW、ICG、Alexa Fluor 350;Alexa Fluor 488;Alexa Fluor 532;Alexa Fluor 546;Alexa Fluor 568;Alexa Fluor 594;Alexa Fluor 647;Alexa Fluor 680或Alexa Fluor 750。

[0198] 因此,某些实施方案包括检测样本中的靶多核苷酸基因产物的方法,包括:a)将所述样本与包含与所述样本中的靶多核苷酸基因产物互补的序列的探针杂交,并且其中探针与所述多核苷酸基因产物在能够在所述探针与所述靶多核苷酸基因产物或其片段之间形成杂交复合物的条件下特异性杂交,以及b)检测有或无所述杂交复合物,以及当存在时,其量。还包括检测样本中的靶多核苷酸基因产物的方法,包括a)扩增所述靶多核苷酸基因产物或其片段,以及b)检测有或无所述扩增的靶多核苷酸基因产物或其片段,以及当存在时,其量。在特定实施方案中,所述探针被可检测地标记。

[0199] 基因产物表达水平的分析

[0200] 根据某些实施方案,基于表1中示出的两种或更多种基因表达的基因产物(包括但不限于本文所述的基因产物的任何特定组合)的表达水平,本发明的方法和试剂盒可用于确定个体是否患有甲状腺癌或者甲状腺肿瘤或结节是否为良性的。通常,当相比获自正常或非癌性甲状腺组织的生物样本中观察到的表达模式,确定表达模式与获自癌性的甲状腺组织的生物样本中观察到的表达模式更紧密相关时,进行甲状腺癌确定。类似地,当相比获自癌性的甲状腺组织的生物样本中观察到的表达模式,确定表达模式与获自正常或非癌性甲状腺组织的生物样本中观察到的表达模式更紧密相关时,通常进行甲状腺组织样本是否为良性的确定。

[0201] 在某些实施方案中,通过将本文所述的一种或多种基因产物的表达水平与合适的对照比较来诊断甲状腺癌的存在。“合适的对照”或“适当的对照”包括参照值,如确定显示例如正常性状,如没有病况如甲状腺癌的组织或生物体,如对照或正常细胞、组织或生物体的细胞或其他生物样本的表达水平、特征、特性或性质。在某些实施方案中,“合适的对照”或“适当的对照”为预定值、水平、特征、特性或性能。

[0202] 在某些实施方案中,将生物样本中一种或多种基因产物的表达与所述基因产物的参照表达水平比较,在某些实施方案中,所述参照表达水平可以为预定的或预先确定的值。可基于一份或多份合适的对照样本中基因产物的表达水平来确定参照表达水平,所述对照样本可为正常组织样本或肿瘤组织样本,如甲状腺肿瘤。例如,在某些实施方案中,基于获自正常非癌性组织如正常甲状腺组织的1、2、3、5、10、20、50或更多份生物样本中基因产物的表达水平,确定与正常组织相关的参照表达水平。在其他实施方案中,基于获自癌症组织如甲状腺癌组织的1、2、3、5、10、20、50或更多份生物样本中基因产物的表达水平,确定与癌症组织如甲状腺癌相关的参照表达水平。在某些实施方案中,确定了在正常和癌症组织的任一或二者中由表1中的基因表达的一种或多种、两种或更多种或者三种或更多种基因产物,或者选自CXCR3、CCR3、CXCL10、CAR、XB130、HO-1和CCR7基因的至少一种、至少两种或至

少三种基因的参照表达水平。在相关实施方案中,确定了正常和癌症组织中的任一或二者中的本文所述的基因产物组中的任何一种,如基因组包含或由以下组成:CCR3、TIMP-1、CAR和XB130基因的基因产物;CXCL10、TIMP-1、CAR和CCR7基因的基因产物;TIMP-1、CAR和CCR7基因的基因产物;或者CXCL10、TIMP-1、CLDN-1和CCR7基因的基因产物。

[0203] 在特定实施方案中,相比正常(非癌性的)对照样本或对照参照表达,生物样本中一种或多种、两种或更多种或者三种或更多种基因产物的差异表达指示癌症。相反,相比正常(非癌性的)对照样本或者对照参照表达,生物样本中一种或多种、两种或更多种或者三种或更多种基因产物的基本上相似的表达指示良性组织,而相比癌症对照样本或者癌症对照参照表达,生物样本中一种、两种或更多种基因产物的基本上相似的表达指示癌症。

[0204] 差异表达包括相比合适对照中相同基因产物的表达水平,一种或多种基因产物的表达水平在统计上显著差异。统计上显著差异可以指如通过RNA水平、蛋白水平、蛋白功能或基因表达的任何其他相关测量如本文中描述的那些测得的表达水平的增加或降低。如果不可能偶然发生,则结果通常称为统计上显著的。检验或结果的显著水平传统上是指概率统计假设检验概念。在简单的情况下,统计显著性可确定为当零假设实际上为正确时决定拒绝零假设的可能性(称为I型错误决定或者“假阳性决定”)。通常使用p-值作出该决定:如果p-值小于显著水平,则拒绝零假设。p-值越小,结果越显著。贝叶斯因子也可用于确定统计显著性(参见,例如Goodman S., Ann Intern Med 130:1005-13, 1999)。在某些情况下,检验或结果的显著水平可反映这样的分析,其中当零假设实际上正确时决定拒绝零假设的可能性不大于所述的可能性。这类分析允许那些应用,其中决定拒绝的可能性可比零假设中包括的一些假设集合的显著水平低得多。

[0205] 在某些示例性的实施方案中,统计上显著的差异表达可包括这样的情况,其中相比合适的对照,给定基因产物的表达水平为生物样本中至少约1.2X、1.3X、1.4X、1.5X、1.6X、1.7X、1.8X、1.9X、2.0X、2.2X、2.4X、2.6X、2.8X、3.0X、4.0X、5.0X、6.0X、7.0X、8.0X、9.0X、10.0X、15.0X、20.0X、50.0X、100.0X或更大的表达差异(即,可能更高或更低表达的差异表达),包括其间的所有整数和小数点(例如,1.24X、1.25X、2.1X、2.5X、60.0X、75.0X等)。在某些实施方案中,统计上显著的差异表达可包括这样的情况,其中相比合适的对照,给定基因产物的表达水平为生物样本中至少约4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、45、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000百分比(%)或更大的表达差异(即,可能更高或更低的差异表达),包括其间的所有整数和小数点。

[0206] 另一个实例,差异表达也可通过进行Z-检验,即计算如本领域已知的绝对Z分数来确定。Z-检验通常用于确定样本均值与总体均值之间的显著差异。例如,相比标准正常表(例如,对照组织),在95%置信区间(即,在5%显著水平),绝对值大于1.96的Z-分数指示非随机性。对于99%置信区间,如果绝对Z大于2.58,其指 $p < 0.01$,并且差异甚至更显著——以更大的置信度拒绝零假设。在这些和相关实施方案中,1.96、2、2.58、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20或更大(包括其间的所有小数点(例如10.1、10.6、11.2等))的绝对Z-分数可提供更强的统计显著性测量。在某些实施方案中,大于6的绝对Z-分数可提供异常高的统计显著性。

[0207] 大量的相似性通常是指生物样本与参照对照之间的表达水平没有统计上显著的

差异。大量相似的表达水平的实例可包括这样的情况,其中相比参照样本,给定基因产物的表达水平提供生物样本中小于约.05X、0.1X、0.2X、0.3X、0.4X、0.5X、0.6X、0.7X、0.8X、0.9X、1.0X、1.1X、1.2X、1.3X或1.4X的表达差异(即,可能更高或更低表达的差异表达),包括其间的所有小数点(例如,.15X、0.25X、0.35X等)。在某些实施方案中,差异表达可包括这样的情况,其中相比参照样本,给定基因产物的表达水平提供生物样本中小于约0.25、0.5、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50百分比(%)的表达差异(即,可能更高或更低的差异表达),包括其间的所有小数点。

[0208] 差异表达可指给定基因产物表达相比对照样本或值的增加或降低。在本文所述方法的特定实施方案中,相比正常对照样本的表达水平或值,生物样本中CCR3、CXCL10、CK19、TIMP1、CLDN1、CCR7和HO-1基因中的一种或多种、两种或更多种、三种或更多种或者四种或更多种的基因产物表达的增加,和/或CXCR3、CAR和XB130基因中的一种或多种、两种或更多种或者三种或更多种的基因产物表达的降低,或者本文所述基因产物的不同亚组中的任何一种的相应改变,指示存在肿瘤或癌症。所述增加或降低可对应于以上示出的改变中的任何一种,如增加约10%、20%、50%、100%、200%、500%或1000%,或者降低约10%、20%、50%或90%。

[0209] 如上文所示,在特定实施方案中,确定生物样本是否包含癌症组织包括将一种或多种基因产物的表达水平与有或无癌症相关联。这可通过将所述生物样本中基因产物的表达水平与一个或多个对照样本,如获自己知患有癌症或已知未患癌症的个体的样本中的表达水平相比较来完成。在某些实施方案中,所述关联或比较可使用更早的时间获自对照样本的基因表达数据,如预定或预先确定的参照值进行。

[0210] 如本文所述,本发明的方法可使用本领域已知的多种方法中的任何一种,定性或定量确定一种或多种基因产物的表达水平。在一些情况下,可检测探针的杂交程度与生物样本中基因产物的量直接相关。可使用软件来提取、标准化、概括和分析来自探针的信号。在一些实施方案中,可针对参照组比较确定的生物样本的给定探针的信号强度,从而确定所述样本中是否发生差异表达。例如,在微阵列的上下文中,对应于所表达的基因产物的阵列上的位置处的相对强度的增加或降低分别指示生物样本中相应基因表达的增加或降低。

[0211] 可使用分类器或算法,如设计用于标准化和/或改善数据可靠性的算法,进行基因表达分析并与癌症或良性表型相关联。可以使用的算法包括本文所述的那些方法中的任何一种,例如实施例2和3中描述的分类器或算法。在某些实施方案中,基因产物为mRNA,其可使用如实施例1中所述的实时PCR测量。在某些实施方案中,可使用每一种基因的CT值来确定不同基因的分类潜力,从而基于获自恶性和良性甲状腺结节的甲状腺组织样本中不同基因的表达来生成ROC曲线,如实施例1所示。也可确定每一种基因的灵敏度和特异性。

[0212] 在特定实施方案中,所述算法为多基因分类测试。可使用两步法开发这样的分类器,其中第一步鉴定和分类具有不典型CT值的生物样本组,且第二步产生例如用于分类在第一步骤后剩下的未分类生物样本的算法,如实施例2中所述。可整合来自两个步骤的合并数据,并在ROC曲线中作图。

[0213] 在某些实施方案中,所述第一步包括两个阶段。第一阶段鉴定对给定基因组具有不典型CT值的生物样本。确定不典型CT的标准可确定为平均CT值 $\pm$ 两个标准差,从而代表

任何给定基因的仅5%的CT值。对于每一种基因,可计算归入癌症组但事实上属于良性组的不典型CT的错误可能性(EP)。考虑到较小的EP反映较大的分类能力,可基于EP计算每一种基因的分值。在第二阶段,可将具有最佳分值的基因获得的合成分值用于分类所述样本为癌症或良性的。

[0214] 在第二步的某些实施方案中,未在第一步中分类的样本可用作剩下的训练组以产生下游算法,从而完成所述数据的分类,如图3中所示。例如,这可使用两种方法,即线性判别分析(LDA)和非线性判别分析(NLDA)来进行。对于LDA分析,可使用逐步LDA方法,因为如果变量和样本满足LDA所需的条件,则其可为未知的。可以选择所述逐步方法,因为其可同时鉴定给出最大分类确定性且满足LDA条件的变量组合。对于NLDA,可以使用基于遗传算法的方法来包括一组数学函数,从而产生两种或更多种特征的线性或非线性组合(Melo et al, Protein Science, 2007)。该方法产生了多达4种基因的组合作的非线性转换,从而产生单一的合成分值。

[0215] 为了整合两个步骤,可将数值分配至在第一步中获得的数据,从而给予其与在第二步或阶段中获得的值相似的输出同一性,如图3所示。

[0216] 在某些实施方案中,选择分类器或算法(或其所基于的基因)来使灵敏度/特异性关系最大化,使阳性预测值(PPV)和阴性预测值(NPV)均达到大于90%,和/或实现AUC>90。

[0217] 在特定实施方案中,基于本文所述的一种或多种、两种或更多种或者三种或更多种基因产物的表达,包括任何以下基因产物组表达的基因产物,使用分类器或算法如两步骤分类器来确定获自个体的生物样本是否为癌性的或良性的关联和/或确定:

[0218] CXCR3、CCR3、CXCL10、CK19、TIMP-1、CLDN-1、CAR、XB130、HO-1和CCR7基因中的两种或更多种或者三种或更多种的基因产物;

[0219] CXCR3、CCR3、CXCL10、CK19、TIMP-1、CLDN-1、CAR、HO-1和CCR7基因的基因产物;

[0220] CCR3、TIMP-1、CAR和XB130基因的基因产物;

[0221] CXCL10、TIMP-1、CAR和CCR7基因的基因产物;

[0222] TIMP-1、CAR和CCR7基因的基因产物;或者

[0223] CXCL10、TIMP-1、CLDN-1和CCR7基因的基因产物。

[0224] 在某些实施方案中,本发明的方法可使用“机器学习算法”进行,其是指基于计算机的预测方法,在本领域又被技术人员称为“分类器”,其可用于表征基因表达谱。通常将对应于某些表达水平的信号(其通过例如基于微阵列的杂交测定获取),应用于算法以分类表达谱。监督式学习通常包括“训练”分类器来识别类别之间(如癌症相对于良性甲状腺组织)的区别,然后“测试”分类器对独立测试组的准确度。对于新的未知样本,所述分类器可用于确定所述样本所属的类别。在特定实施方案中,使用算法进行关联,包括基于本文所述任何基因产物组的表达水平。

[0225] 适于样本分类算法类型的实例包括但不限于k-最近邻算法、支持向量算法、朴素贝叶斯算法、神经网络算法、隐马尔可夫模型算法、遗传算法或以上算法的任意组合。在本发明的一些实施方案中,提供了对角线性判别分析、k-最近邻算法、支持向量机(SVM)算法、线性支持向量机、随机森林算法或基于概率模型的方法或者以上算法的组合用于分类微阵列数据。在一些实施方案中,基于目标类别之间表达水平差异的统计显著性来选择鉴定的标志物,其能够区分样本(例如,良性相比恶性、正常相比恶性)或区分亚型(例如,PTC相比

FVPTC)。在一些情况下,通过应用Benjamini Hochberg或另外校正错误发现率(FDR)而调整统计显著性。

[0226] 在某些实施方案中,本发明的算法可包括一种或多种其他的计算机分析特征。例如,分析基因产物表达水平的方法可包括使用特征选择算法,其可通过使用LIMMA软件包(Smyth,G.K.(2005).Limma:linear models for microarray data.Bioinformatics and Computational Biology Solutions using R and Bioconductor,R.Gentleman,V.Carey,S.Dudoit,R.Irizarry,W.Huber(eds.),Springer,New York,pages 397-420)而提供。分析基因产物表达水平的方法可包括使用预分类器算法。例如,算法可使用细胞特异性分子指纹来根据其组分对所述样本预分类,然后应用校正/标准化因子。然后可将该数据/信息输入最终分类算法,其并入该信息以有助于最终的诊断。

[0227] 可使用分类器算法将选择的特征分类。示例性的算法包括但不限于减少变量数目的方法,如主成分分析算法、偏最小二乘法和独立成分分析算法。示例性的算法还包括但不限于直接处理大量变量的方法,如统计方法和基于机器学习技术的方法。统计方法可包括判罚逻辑回归、微阵列预测分析(PAM)、基于收缩质心(shrunken centroid)的方法、支持向量机分析和系统化的线性判别分析。机器学习技术包括装袋(bagging)法、提升(boosting)法、随机森林法以及以上方法的组合。Cancer Inform(2008;6:77-97)提供了以上提供的用于分析微阵列强度数据的分类技术的综述。

[0228] 在一些情况下,根据本发明的算法可补充有如Fishel和Kaufman等(2007Bioinformatics 23(13):1599-606)所述的meta分析方法。在一些情况下,分类器算法可补充有meta分析方法如重复性分析。在一些情况下,所述重复性分析选择在至少一个预测性表达产物标志物组中出现的标志物。

[0229] 在某些实施方案中,可使用特征选择技术来分析基因产物表达水平,例如所得到的在样本中测得的每一种基因产物的强度值,诸如但不限于通过观察数据的固有特性评估特征相关性的滤除技术、在特征子集搜索中嵌入模型假设的包装(wrapper)方法,以及其中最佳特征集合的搜索被构建至分类器算法中的嵌入技术。

[0230] 可用于本发明方法的滤除技术的实例包括:(1)参数方法,如使用双样本t-检验、ANOVA分析、贝叶斯框架和 γ 分布模型;(2)无模型方法,如使用威尔科克森秩和检验、类群间-内平方和检验、秩积(rank product)方法、随机置换方法或TnoM,其包括设定两个数据集之间表达的倍数变化差异的阈值点,然后检测每一种基因中使错误分类的数目最小化的阈值点;以及(3)多变量方法,如二变量方法、基于相关性的特征选择方法(CFS)、最小冗余最大相关性方法(MRMR)、马尔可夫毯滤除方法和不相关的收缩质心方法。用于本发明方法的包装方法包括顺序搜索方法、遗传算法和分布估计算法。用于本发明方法的嵌入(Embedded)方法包括随机森林算法、支持向量机算法的权向量以及逻辑回归权重算法。Bioinformatics(2007Oct 1;23(19):2507-17)提供了以上提供的用于强度数据分析的滤除技术相对优劣性的综述。

[0231] 在某些实施方案中,可将样本中基因产物的表达水平与两种或更多种不同的生物标志物组的基因表达数据比较,每一组生物标志物的基因表达数据包含与一种或多种组织类型如甲状腺肿瘤组织的存在相关的一种或多种参照基因的表达水平,其中所述表达水平以顺序形式与两种或更多种生物标志物的基因表达数据比较。表达水平与生物标志物组的

基因表达数据的比较可包括对所述基因表达数据应用分类器,或顺序应用不同的分类器(包括本文中描述的那些)。顺序分析可包括应用获自癌组织样本的基因表达分析的分类器,然后应用获自不同生物样本的混合物的分析的分类器,这样的样本中的一些包含癌组织并且其他样本包含良性组织。

[0232] 在某些实施方案中,用于顺序分析早期的分类器可用于划入或划出样本为良性或疑似的。在一些实施方案中,这样的顺序分析以将“主要”分类器应用于来自未被前述分类器划出的样本的数据结束,其中所述主要的分类器获自多种组织类型的基因表达水平的数据分析,并且其中所述主要的分类器能够指示样本为良性或疑似的(或恶性的)。

[0233] 例如,在某些实施方案中,使用基因或生物标志物组的谱型分析可用于表征甲状腺组织为良性、疑似和/或恶性的。组可来源于包含如下亚型的集群的基因表达水平的分析:良性(非癌性的)甲状腺亚型,包括滤泡性腺瘤(FA)、结节性增生(NHP)、淋巴细胞性甲状腺炎(LCT)和Hurthle细胞腺瘤(HA);恶性亚型,包括滤泡性癌(FC)、乳头状甲状腺癌(PTC)、乳头状癌的滤泡性变体(FVPTC)、髓质甲状腺癌(MTC)、Hurthle细胞癌(HC)和退行性甲状腺癌(ATC)。这样的小组(panel)也可来源于非甲状腺亚型,包括肾癌(RCC)、乳腺癌(BCA)、黑色素瘤(MMN)、B细胞淋巴瘤(BCL)和甲状旁腺(PTA)。与正常甲状腺组织(NML)相关的生物标志物组也可用于本文提供的方法和组合物。示例性的生物标志物提供于表1中,并且在本文中描述了具体的生物标志物组。

[0234] 如上文所述,为鉴定、分类、诊断或定征生物样本的目的,本发明的方法和试剂盒可涉及使用特定基因或基因产物组如“生物标志物组”。本发明还使用生物标志物组,在本文中描述为“分类组”。通常,确定组中的生物标志物(也称为标签)的基因表达水平的常见模式,然后例如通过测量所述样本标签与参照标签之间的相似性来评估生物样本中的相同生物标志物组的标签。在一些实施方案中,所述方法包括测量(或获取)生物标志物组和/或分类组中的两种或更多种基因产物的水平。例如,在一些实施方案中,生物标志物组或分类组可包含至少1、2、3、4、5、6、7、8、9、10、15、20、25、30、33、35、38、40、43、45、48、50、53、58、63、65、68、100、120、140、142、145、147、150、152、157、160、162、167、175、180、185、190、195、200或300种生物标志物。在一些实施方案中,生物标志物组或分类组包含不多于1、2、3、4、5、6、7、8、9、10、15、20、25、30、33、35、38、40、43、45、48、50、53、58、63、65、68、100、120、140、142、145、147、150、152、157、160、162、167、175、180、185、190、195、200或300种生物标志物。在一些实施方案中,分类组包含至少1、2、3、4、5、6、7、8、9、10、15、20或25种不同的生物标志物组。在其他实施方案中,分类小组包含不多于1、2、3、4、5、6、7、8、9、10、15、20、25种不同的生物标志物组。

[0235] 可选择生物标志物组以容纳从非良性或疑似的表达谱充分分离的良性物。可对众多生物样本,如至少50、100、200、300、400、500、600、700、800、900、1000、1500、2000、2500、3000、3500或4000份生物样本(例如,甲状腺样本),进行分类器即算法的训练。总样本群体可由获自FNA的样本组成,或者所述样本群体可为通过FNA及通过其他方法获得的样本如术后组织的混合物。在一些实施方案中,许多训练/测试组被用于开发初步算法。总算法错误率可示出为良性相比非良性样本基因数目的函数。在一些实施方案中,可使用其他性能度量,如这样的性能度量,其为亚型或者良性相比恶性(B相比M)基因数目的函数。这样的性能度量可使用CV或本领域已知的其他方法获取。所有结果均可使用支持向量机模型获取,其

在交叉验证的模型中对所述样本进行训练和测试。

[0236] 分子谱型分析结果的统计学评估可提供指示以下中的一种或多种的定量值：诊断准确的可能性；癌症可能性；或者特定癌症亚型的可能性。可将数据以其最有用的形式直接提供给医师，从而指导患者护理。

[0237] 可使用本领域已知的多种方法来统计评估分子谱型分析的结果，包括但不限于：学生T检验、双侧T检验、皮尔森秩和分析、隐马尔可夫模型分析、q-q图分析、主成分分析、单因素ANOVA、双因素ANOVA、LIMMA等。

[0238] 在本发明的一些实施方案中，本发明的方法单独或组合细胞学分析时，可提供分类、鉴定或诊断例如癌症或良性的，其为约85%准确度和约99%或约100%准确度。

[0239] 基于本文所述的每一种生物标志物或分类组的算法可使用在算法训练期间确定的基因产物表达水平信息，以划入或划出给定样本为“良性”、“疑似”或者为包含或不包含一种或多种组织类型（例如，NML、FA、NHP、LCT、HA、FC、PTC、FVPTC、MTC、HC、ATC、RCC、BCA、MMN、BCL和PTA）。每一种生物标志物或分类组算法可使用简单的判定规则来滤除进入的样本，如果满足判定规则（例如，关于其中包含一种或多种组织类型的同一性或状态而表征样本），则从随后的评估有效去除任何标记的样本。本文提供的生物标志物组和分类组可用于分类、表征、鉴定和/或诊断甲状腺癌或其他甲状腺病况（包括诊断甲状腺为正常或良性的）。

[0240] 基因表达水平的分析可包括将本文所述的不同分类器或算法依序应用于基因表达数据。在某些实施方案中，这样的顺序分析可包括应用获自癌性甲状腺组织的多个样本的基因表达分析的分类器，然后应用获自甲状腺组织的不同样本的混合物分析的分类器，一些样本包含癌性甲状腺组织而其他样本包含良性甲状腺组织。在一些实施方案中，所述分类器获自良性组织、正常组织和/或非甲状腺组织（例如，甲状旁腺组织）的基因表达模式的分析。在一些实施方案中，患病组织为HA和/或HC组织。

[0241] 在一些实施方案中，所述分类方法在每一个分类小组从生物样本接收作为输入生物标志物表达水平（例如，总结的微阵列强度值、qPCR或测序数据）时开始。然后评估分类小组中指定的生物标志物和表达水平。如果来自给定样本的数据与分类小组中指定的规则匹配（或者与分类小组的标签相关），则其数据输出标记所述样本，并阻止其通过主要（下游）分类器进一步评估和评分。当分类小组标记样本时，系统自动返回该样本的“疑似”调用。当分类小组不标记样本时，评估向下游继续至下一分类小组，并且其可被标记或不标记。在一些情况下，所述分类小组以特定的顺序施用；在其他情况下，施用顺序可为任何顺序。

[0242] 在某些实施方案中，所述分类方法以例如通过来自个体的样本（例如，甲状腺组织样本）的一种或多种基因产物的基因表达分析、表达水平确定开始。单独地，可分析一组或多组参照或训练样本来确定至少两组不同的生物标志物的基因表达数据，每一组生物标志物组的基因表达数据包含与一种或多种组织类型的存在相关的一种或多种基因表达水平。第一组生物标志物的基因表达数据可用于训练第一分类器；第二组的基因表达数据可用于训练第二分类器；并且对2、3、4、5、6、7、8、9、10、11、12、13、14、15、16或更多组生物标志物以及任选相应的分类器依次进行。用于分析每一组生物标志物的参照或训练样本组可为重叠的或非重叠的。在一些实施方案中，所述参照或训练样本包含HA和/或HC组织。在实例分类方法接下来的步骤中，进行所述样本的基因表达水平与第一组生物标志物或第一分类器之

间的第一次比较。如果该第一次比较的结果匹配,则分类过程以某一结果结束,如指示所述样本为疑似、癌性或包含特定组织类型(例如HA或HC)。如果比较的结果不匹配,则在第二轮的比较中将所述样本的基因表达水平与第二组生物标志物或第二分类器比较。如果该第二次比较的结果匹配,则所述分类过程以某一结果结束,如指示所述样本为疑似、癌性或者包含特定组织类型(如HA或HC)。如果比较的结果不匹配,则该过程以类似的逐步比较过程继续,直至出现匹配,或者直到分类方法中包括的所有生物标志物或分类器组用作比较的基础。如果在样本的基因表达水平与分类方法中使用的生物标志物或分类器的任何组之间未出现匹配,则所述样本可指示为“良性的”。在一些实施方案中,分类过程中的最终比较发生于样本的基因表达水平与如本文所述的主要分类器之间。

[0243] 在本发明的一些实施方案中,由于处理的大量单个数据点,数据分析或关联需要计算机或其他设备、仪器或装置,以应用本文所述的不同算法。

[0244] 试剂盒

[0245] 在特定实施方案中,本发明提供了用于诊断或预测个体的癌症如甲状腺癌的试剂盒或诊断测试。本文所述的诊断测试可为体外诊断测试。诊断测试包括但不限于FDA批准或者允许的体外诊断(IVD)、实验室开发的测试(LDT)或者直接面对消费者(DTC)的测试,其可用于测定生物样本并检测或指示有或无癌症如甲状腺癌。在一实施方案中,可将诊断测试或试剂盒用于实验室或其他医疗卫生专业环境。在另一实施方案中,诊断测试或试剂盒可供消费者在家使用。

[0246] 诊断测试和试剂盒包含一种或多种用于检测本文所述的基因产物的试剂,并且可包含旨在用于体外诊断癌症如甲状腺癌的其他试剂、仪器和系统,以治愈、缓解、治疗或预防疾病或其后遗症。在一实施方案中,本文所述的试剂盒或诊断测试可旨在用于收集、制备和检查采自人体的标本。在某些实施方案中,试剂盒、诊断测试和产品可包括一种或多种实验室测试。如本文所用,术语“实验室测试”是指包括测试获自个体的生物样本的一项或多项医学或实验室程序。

[0247] 本发明的试剂盒和诊断测试包含一种或多种试剂,以用于检测本文所述的基因产物,如由表1中列出的那些基因表达。就此而言,用于检测的试剂可包括技术人员已知的用于检测基因产物的任何试剂,包括但不限于抗体和寡核苷酸。在某些实施方案中,试剂盒或诊断测定还可包含使用所述试剂盒如何进行如本文所述的测定以确定基因产物的表达水平的书面指导。

[0248] 在某些实施方案中,本发明的试剂盒或诊断测定包含两种或更多种、三种或更多种或者四种或更多种试剂如探针,以用于检测本文所述的基因产物。在特定实施方案中,基因产物为蛋白或核酸,如mRNA。在某些实施方案中,所述试剂为抗体或寡核苷酸,包括本文所述的任何一种。在某些实施方案中,每一种试剂例如寡核苷酸组,其中每组包含两种寡核苷酸或由两种寡核苷酸组成,所述两种寡核苷酸一起能够通过PCR扩增靶多核苷酸基因产物。在某些实施方案中,所述试剂被可检测地标记。在一实施方案中,试剂盒或诊断测定包含两种或更多种、三种或更多种、或者四种或更多种检测基因产物的试剂,其中所述试剂盒或诊断测定包含两种或更多种、三种或更多种、或者四种或更多种引物组,每组能够扩增至少一部分靶基因产物。在某些实施方案中,所述试剂盒或诊断测定还包含两种或更多种、三种或更多种或者四种或更多种可检测地标记的探针,每一种探针特异性结合靶基因产物

或其互补物中的一种。因此,在某些实施方案中,每一组引物用于扩增靶基因产物,然后每一种探针用于检测扩增产物并测量每一种基因产物的表达水平。在特定实施方案中,每种试剂如每组引物,检测不同的基因产物。然而,应理解某些实施方案可包含两种或更多种能够扩增相同基因产物的试剂。例如,试剂盒或诊断测定可包含两种试剂,一种为一组扩增引物,且另一种为探针,其中每一种特异性结合和/或检测相同的基因产物。另外,试剂盒或诊断测定可包含各自特异性结合或检测相同基因产物的两种或更多种试剂的多种组合,例如其中每一种组合特异性结合或监测多不同的基因产物,从而允许例如同时扩增和检测多种基因产物,。

[0249] 在特定实施方案中,通过试剂盒或诊断测定的试剂检测的基因产物由表1中列出的一种或多种基因表达,其中所述基因产物中的至少一种由CXCR3、CCR3、CXCL10、CAR、XB130、HO-1或CCR7基因表达。在某些实施方案中,所述基因产物包含以下或由以下组成: CXCR3、CCR3、CXCL10、CK19、TIMP-1、CLDN-1、CAR、XB130、HO-1和CCR7基因中的三种或更多种的基因产物; CXCR3、CCR3、CXCL10、CK19、TIMP-1、CLDN-1、CAR、HO-1和CCR7基因的基因产物; CCR3、TIMP-1、CAR和XB130基因的基因产物; CXCL10、TIMP-1、CAR和CCR7基因的基因产物; TIMP-1、CAR和CCR7基因的基因产物;或者CXCL10、TIMP-1、CLDN-1和CCR7基因的基因产物。

[0250] 在某些实施方案中,试剂盒或诊断测定在单独的容器中包含一种试剂。在其他实施方案中,每一种试剂提供于相同的容器中。在特定实施方案中,所述试剂中的每一种与基质如阵列连接。在特定实施方案中,所述试剂中的每一种与固体基质的不同区域连接。因此,在一实施方案中,所述试剂为与固体基质共价结合的寡核苷酸或寡核苷酸组,其中所述固体基质任选为阵列,并且其中所述阵列任选为微阵列。在某些实施方案中,所述试剂为寡核苷酸组如引物组,并且所述寡核苷酸组包含DNA。

[0251] 本发明的试剂盒或诊断测定还可包含一种或多种适合将所述试剂与所述基因产物结合的溶液,和/或一种或多种用于实施本发明方法的溶液或试剂,以确定基因产物的表达水平。例如,在特定实施方案中,包含PCR引物组的试剂盒或诊断测定还可包含一种或多种另外的试剂用于进行PCR测定,如热稳定的聚合酶、脱氧核苷酸混合物和/或可检测地标记的探针。在特定实施方案中,可检测地标记的探针包含荧光团和淬灭部分,并且在所述探针被切割时而不是在所述探针完整时,所述探针可发射可检测的信号。

[0252] 本发明的试剂盒或诊断测定还可包含一种或多种试剂用于例如处理和/或保存生物样本,其中甲状腺组织样本的处理包括从该生物样本提取基因产物。

[0253] 本发明的试剂盒或诊断测定还可包含一种或多种对照基因产物,如包含样本的基因产物的阳性对照和/或不包含样本的阴性对照,以确保进行的方法能够成功地特异性鉴定和/或测量基因产物的表达和/或存在。

[0254] 在某些实施方案中,试剂盒或诊断测定包含数据或信息,例如,对应于阳性和/或阴性对照样本中基因产物的基因表达水平或者指示有或无癌症如甲状腺癌的基因表达的预定阈值水平的数据或信息。在相关实施方案中,试剂盒或诊断测定包含用于将生物样本中基因的表达水平与有或无癌症如甲状腺癌相关联的算法。在特定实施方案中,试剂盒或诊断测定包含计算机可读介质,其含有数据和/或算法或者含有数据或算法的代码。

[0255] 在特定实施方案中,用于检测基因产物的试剂、溶液、另外的试剂和对照基因产物中的一种或多种分装于不同的容器中。

实施例

[0256] 实施例1

[0257] 用于诊断测定的基因选择

[0258] 将基于其与甲状腺癌的关系而预先选择的18种基因用于开发能够准确地分类不确定的甲状腺结节为良性或癌症的改进的诊断测定。这些基因包括CXCR3 (基因1)、CXCR3A (基因2)、CXCR3B (基因3)、CXCR4 (基因4)、CCR3 (基因5)、CXCL9 (基因6)、CXCL10 (基因7)、CXCL11 (基因8)、SPAG-9 (基因9)、CK-19 (基因10)、TIMP-1 (基因11)、CLDN-1 (基因12)、CAR (基因13)、Nectin-1 (基因14)、XB-130 (基因15)、HO-1 (基因16)、CCR7 (基因17) 和CXCL4 (基因18)。

[0259] 作为训练组,在操作室前瞻性地收集了来自恶性和良性甲状腺结节的新鲜快速冷冻的甲状腺组织样本(n=156),包括100份甲状腺癌和56份良性甲状腺结节。甲状腺癌的类别包括乳头状甲状腺癌(PTC)普通型(56份)、滤泡性变体(22份)、扩散性硬化(8份)和滤泡性癌(FC)(14份)。良性结节包括滤泡性增生(26份)、甲状腺炎(14份)和滤泡性腺瘤(16份)。对所有手术样本的最终活组织检查报告进行了评估,并由两名独立的专业病理学家验证。将所述样本立即置于4℃RNA保存溶液(RNAlater, Ambion)中,然后使用RNAeasy Plus小型试剂盒(Qiagen)均质化,并保存于-80℃。使用Picodrop 100分光光度计确定RNA浓度。使用琼脂糖-甲醛电泳验证RNA完整性。

[0260] 通过实时PCR分析基因差异表达以确定每一个基因单独和不同组合时的分类潜力。使用ImProm-II逆转录系统(Promega)进行互补DNA(cDNA)的合成。利用来自Qiagen的RotorGene Q循环仪,使用Brilliant II SYBR Green master mix(Agilent)试剂盒,按照厂商指导,以一式双份的样本进行实时PCR。通过RotorGene软件应用以确定每一种基因的最佳扩增条件的方式分析标准曲线。获得的效率值范围为95%-109%。获得的线性回归系数值(Rsq)的范围为0.990-0.999。对于初始比较,将基因表达以18s和β-肌动蛋白标准化,并通过Pfaffl提议的相对定量模型进行分析。结果在图1中以癌症相对于良性甲状腺结节的相对倍数变化示出。

[0261] 使用每个基因的CT值,生成了受试者工作特征(ROC)曲线以比较每一个单个基因分类甲状腺样本的能力。确定了曲线下面积(AUC)和最佳灵敏度和特异性,且在图2中提供。独立地,具有最佳AUC和灵敏度/特异性值的基因为基因11(AUC:0.87和灵敏度/特异性:96%/78%)、基因12(AUC:0.85和灵敏度/特异性:85%/73%)以及基因10(AUC:0.84和灵敏度/特异性:81%/79%)。基因17和5分别显示0.74和0.70的AUC。所有其他基因均显示较差的分类特性,如图2中所示。

[0262] 实施例2

[0263] 诊断测定的开发

[0264] 使用了多基因分类测试(基因标签),其使灵敏度/特异性关系最大化,使得阳性预测值(PPV)和阴性预测值(NPV)均达到大于90%。分类器中选择最佳基因组的标准为:AUC大于0.97,灵敏度和特异性值分别大于92%和90%;基因分类器应当稳健且耐受不典型基因谱变化而无过拟合。另外,标签应当使用较小的基因组,从而允许在即时诊断环境如病理实验室中使用简单的试剂盒。为了满足这些标准,训练了单独的算法:第一种鉴定和分类具有

不典型(离群)CT值的样本,且第二种分类具有非不典型CT值的样本(图3A)。为了整合来自两种算法的输出数据,对获自第一种算法的数据进行数学转换以给出与获自第二种算法(图3A)的数据相似的输出同一性。

[0265] 为了开发分类器,在两步法中顺序使用了所述算法。分类器的第一个步骤包括两个阶段:第一阶段鉴定样本具有不典型CT值(离群值),其定义为大于与每一个基因的平均CT值的两个标准差的值(图3A和B)。如果样本满足基因的不典型CT的标准,则其进入第二阶段,该阶段计算归入了癌症组但事实上应归入良性组的不典型CT值的错误概率(EP)。计算所选择基因的EP表示为单个分数,其被用于产生分类样本为癌症或良性的合成分数(图3A和B)。第一步骤鉴定56份良性样本中的21份以及100份癌症样本中的36份为不典型,且以100%的准确度将它们分类。

[0266] 不满足第一阶段的不典型CT值标准或者未在第二阶段中分类的样本根据分类方法进入第二步骤(图3B)。第二步骤使用通过两种方法,即线性判别分析(LDA)和非线性判别分析(NLDA)训练的算法(图3B)。使用SPSS 15.0软件以逐步法进行LDA分析,因为不确定变量和样本是否满足LDA需要的条件。选择了逐步法,因为其同时鉴定给出最大分类确定性且满足LDA所需条件的变量组合。对于NLDA,使用了基于遗传算法的方法来推出一组数学函数,产生使用两种或更多种特征的线性或非线性组合(Melo et al, Protein Science, 2007)。该方法形成了产生单一合成分数的多达4种基因组合的非线性转换。

[0267] 实施例3

[0268] 诊断基因组的选择

[0269] 基于ROC曲线参数,包括曲线下面积(AUC)、灵敏度和特异性,选择了由如实施例2中所述的LDA和NLDA策略生成的新的基因分类器。通过LDA(SV)和NLDA(FM72和FM208)获得的代表性算法的结果在图4中示出。所有分类器均显示卓越的性能,具有AUC大于0.98,灵敏度范围为94-97.8%,且特异性为92-99%。最重要的是,具有最佳性能(SV;图4)的算法分别显示95.8%和96.1%的阳性预测值和阴性预测值(图7)。

[0270] 尽管先前已显示具有最佳单个性能的3种基因(CK19、TIMP-1、CLDN1)为甲状腺癌的强生物标志物,其单独的组不能解释这些新的算法的性能。为了证明这一点,使用实施例2所述的相同的两步法策略,将这些基因单独或组合在一起整合,与本文鉴定的新的基因组合(包括SV基因组合)进行比较。显著地,相比单独基因,3种基因的组合未改变其性能,如图5所示。此外,Spearman相关分析显示其紧密相关($p < 0.001$) (图6),这解释了为何这些基因的组合未改变其单独的性能。此外,本文鉴定的特定基因组合(包括SV基因组合)的性能,在统计上优于单独或组合的CK19、TIMP-1和CLDN1(图5)。不受理论的束缚,相信本文所述的基因组合的优势是基于“良好”生物标志物与“不良”生物标志物的组合,从而实现分类能力最大化。这可能是因为本文所述的特定组合中,大多数基因彼此不相关这一事实,因此,分类器的改善性能通过每一种基因鉴定并正确分类不同的样本组这一事实来实现。具体而言,CK19、TIMP-1和CLDN1正确分类80%的癌症样本,而用于诊断本发明的基因组合的其他基因分类大多数的良性样本。

[0271] 如本文所述开发的分类器中使用的3种基因组(SV、FM72和FM208)的性能相比单独或组合的基因10、基因11和基因12的比较概要图在图7中示出。SV包括以下基因: CXCR3、CCR3、CXCL10、CK19、TIMP-1、CLDN-1、CAR、HO-1和CCR7。FM72包括以下基因: CCR3、TIMP-1、CAR和

XB130.FM208包括以下基因: CXCL10、TIMP-1、CLDN-1和CCR7。本文所述的不同基因组和分类器的性能与Afirma®甲状腺FNA分析测试(Veracyte, South San Francisco, CA)的比较在图7中示出。

[0272] 实施例4

[0273] 使用独立测试组的基因分类器的测试

[0274] 针对独立的数据集确定了所选择的诊断基因标志物在检测癌症或良性组织中的准确度。使用了包含20份癌症(PTC普通型(15份)、PTC滤泡性变体(4份)和滤泡性癌(1份))及39份良性(滤泡性增生(28份)、甲状腺炎(7份)和滤泡性腺瘤(4份))组织样本的新的样本组。在这种情况下,包括了较小数目的癌症样本以代表在不确定的甲状腺结节中细针抽吸活组织检查见到的癌症的相似普遍性。

[0275] 该使用SV分类器的独立组的分析显示卓越的性能,具有AUC为0.95,灵敏度95%,特异性为90%,PPV 83%且NPV 98%(图8)。考虑到PPV与癌症的普遍性有关,预测阳性预测值减少至83%。相似地,增加的阴性预测值取决于良性病况的普遍性,并因此相比训练组,在该测试组中增加至98%(图8)。这些结果证实用于本文所述的分类器中的标志物产生未过拟合的准确结果,并提供可靠地新的诊断测定。

[0276] 实施例5

[0277] 使用获自细针抽吸样本的第二独立测试组的基因分类器测试

[0278] 针对获自细针抽吸(FNA)样本的独立组确定了所选择的诊断基因标志物在检测癌症或良性甲状腺结节中的准确度。这些相当于进行测定的真实临床环境。因为FNA对应于非常小的样本,有可能减少的细胞性能降低测定的性能。因此,测试了本发明的标记物组预测FNA样本中的甲状腺结节性质的能力。新的样本组包含26份乳头状甲状腺癌和74份良性甲状腺结节。包括了较少数目的癌症样本以代表在不确定的甲状腺结节的细针抽吸活组织检查中见到的癌症的相似普遍性。将FNA样本的细胞病理学诊断用作黄金标准来比较测定的分子分类结果。

[0279] 该使用SV分类器的独立组的分析显示出卓越的性能,具有灵敏度为96%,特异性为89%,PPV为75%且NPV为98%(图9)。这些结果证实本文所述的分类器中使用的标志物产生了未过拟合的准确结果,且提供了本发明的测定在临床环境常规使用的FNA样本中具有卓越的性能的可靠证明。

[0280] 本说明书中提及的所有美国专利、美国专利申请公开、美国专利申请、外国专利、外国专利申请和非专利出版物均通过引用整体并入本文中,其程度直至与当前描述不一致。

[0281] 从上文应理解,尽管为了说明的目的已在本文中描述了本发明的具体实施方案,可进行不同的修改而不脱离本发明的精神和范围。因此,除了所附权利要求外,本发明不受限制。

[0001]

序列表

<110> 智利天主教教皇大学

<120> 用于诊断甲状腺肿瘤的组合物和方法

<130> BMRC-001/02W0

<150> US 61/730,391

<151> 2012-11-27

<150> US 61/775,419

<151> 2013-03-08

<160> 20

<170> PatentIn version 3.5

<210> 1

<211> 1670

<212> DNA

<213> 智人

<400> 1

ccaaccacaa gcaccaaagc agaggggcag gcagcacacc acccagcage cagagcacca	60
gcccagceat ggtecttgag gtgagtgage accaagtget aaatgacgcc gaggttgccg	120
ccctcctgga gaacttcagc tcttctatg actatggaga aaacgagagt gaetctgget	180
gtacctcccc gccctgceca caggacttea gcctgaactt cgaccgggcc ttcttgccag	240
ccctctacag cctcctcttt ctgctggggc tgctgggcaa cggcgcggtg geagccgtgc	300
tgttgagccg gcggacagcc ctgagcagea ccgacacctt cctgctccac ctagnetgtag	360
cagacacgct gctgggtgtg acactgccgc tctgggcagt ggacgctgcc gtccagtggg	420
tctttggctc tggectctgc aaagtggcag gtgccctctt caacatcaac ttctacgcag	480
gagccctect gctggcctgc atcagctttg accgetaect gaacatagtt catgccacce	540
agetctaacc cggggggccc ccggcccggc tgacctcac ctgcctggct gtctgggggc	600
tctgcctgct tttegcctc ccagaacttea tcttctgtc ggcccaccac gacgagcgcc	660
tcaacgccac ccaactgcaa tacaacttec cacaggtggg ccgcacggct ctgcgggtgc	720
tgcagctggt ggctggettt ctgctgcccc tgctggatcat ggctactgc tatgcccaca	780
tcctggccgt gctgctggtt tccaggggcc agcggcgccct gcgggccatg cgctggtgg	840
tggtggctgt ggtggccttt gccctctgct ggacccccta tcacctggtg gtgctggtgg	900
acatcctcat ggaactgggc gctttggccc gcaactgtgg ccgagaaagc agggtagacg	960
tggccaaagtc ggteacctca ggctgggct acatgeactg ctgcctcaac ccgtgctct	1020

[0002]

```

atgcctttgt aggggtcaag ttccgggagc ggatgtggat gctgctcttg cgccctgggt 1080
gccccaaacca gagagggctc cagaggcagc catcgtcttc ccgccgggat tcctcctggg 1140
ctgagaacctc agaggectcc tactcgggct tgtgaggecg gaatccgggc tccccctteg 1200
cccacagtct gacttccccg cattccaggc tcttccctcc ctctgccggc tctggtctct 1260
cccaatatacc tcgttccccg gactcactgg cagccccagc accaccaggt ctccccggaa 1320
gccacctctc cagctctgag gactgcacca ttgtgtctcc ttagctgcca agccccatcc 1380
tgccgcccga ggtggctgcc tggagcccca ctgcccttct catttggaac ctaaaacttc 1440
atcttcccca agtgcgggga gtacaaggca tggcgtagag ggtgctgccc catgaagcca 1500
cagcccaggc ctccagctca gcagtactg tggccatggt cccaagacc tctatatttg 1560
ctctttttatt tttatgtcta aaatcctgct taaaactttt caataaacia gatcgtcagg 1620
accaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1670

```

<210> 2
 <211> 368
 <212> PRT
 <213> 智人

<400> 2

```

Met Val Leu Glu Val Ser Asp His Gln Val Leu Asn Asp Ala Glu Val
1          5          10          15

```

```

Ala Ala Leu Leu Glu Asn Phe Ser Ser Ser Tyr Asp Tyr Gly Glu Asn
20          25          30

```

```

Glu Ser Asp Ser Cys Cys Thr Ser Pro Pro Cys Pro Gln Asp Phe Ser
35          40          45

```

```

Leu Asn Phe Asp Arg Ala Phe Leu Pro Ala Leu Tyr Ser Leu Leu Phe
50          55          60

```

```

Leu Leu Gly Leu Leu Gly Asn Gly Ala Val Ala Ala Val Leu Leu Ser
65          70          75          80

```

```

Arg Arg Thr Ala Leu Ser Ser Thr Asp Thr Phe Leu Leu His Leu Ala
85          90          95

```

```

Val Ala Asp Thr Leu Leu Val Leu Thr Leu Pro Leu Trp Ala Val Asp
100          105          110

```

Ala Ala Val Gln Trp Val Phe Gly Ser Gly Leu Cys Lys Val Ala Gly
115 120 125

Ala Leu Phe Asn Ile Asn Phe Tyr Ala Gly Ala Leu Leu Leu Ala Cys
130 135 140

Ile Ser Phe Asp Arg Tyr Leu Asn Ile Val His Ala Thr Gln Leu Tyr
145 150 155 160

Arg Arg Gly Pro Pro Ala Arg Val Thr Leu Thr Cys Leu Ala Val Trp
165 170 175

Gly Leu Cys Leu Leu Phe Ala Leu Pro Asp Phe Ile Phe Leu Ser Ala
180 185 190

His His Asp Glu Arg Leu Asn Ala Thr His Cys Gln Tyr Asn Phe Pro
195 200 205

Gln Val Gly Arg Thr Ala Leu Arg Val Leu Gln Leu Val Ala Gly Phe
210 215 220

[0003]

Leu Leu Pro Leu Leu Val Met Ala Tyr Cys Tyr Ala His Ile Leu Ala
225 230 235 240

Val Leu Leu Val Ser Arg Gly Gln Arg Arg Leu Arg Ala Met Arg Leu
245 250 255

Val Val Val Val Val Val Ala Phe Ala Leu Cys Trp Thr Pro Tyr His
260 265 270

Leu Val Val Leu Val Asp Ile Leu Met Asp Leu Gly Ala Leu Ala Arg
275 280 285

Asn Cys Gly Arg Glu Ser Arg Val Asp Val Ala Lys Ser Val Thr Ser
290 295 300

Gly Leu Gly Tyr Met His Cys Cys Leu Asn Pro Leu Leu Tyr Ala Phe
305 310 315 320

Val Gly Val Lys Phe Arg Glu Arg Met Trp Met Leu Leu Leu Arg Leu
325 330 335

[0004]

Gly Cys Pro Asn Gln Arg Gly Leu Gln Arg Gln Pro Ser Ser Ser Arg
 340 345 350

Arg Asp Ser Ser Trp Ser Glu Thr Ser Glu Ala Ser Tyr Ser Gly Leu
 355 360 365

<210> 3
 <211> 1796
 <212> DNA
 <213> 智人

<400> 3
 ctgatggtat ctctgtttca ggagtgggtga cgcctaagct atcactggac atatcaagga 60
 ettcactaaa ttagcaggta ccaactggtct tcttgtgctt atcegggcaa gaacttateg 120
 aaatacaata gaagtitttta cttagaagag attttcagca gatgagaagc tggtaacaga 180
 gaccaaata gtttggagac taaagaatca ttgcacattt cactgctgag ttgtattgga 240
 gaagtgaaat gacaacctca ctagatacag ttgagacctt tggtaccaca tectactatg 300
 atgacgtggg cctgctctgt gaaaaagctg ataccagagc actgatggcc cagtttgtgc 360
 ccccgctgta ctccctgggtg ttcactgtgg gcctcttggg caatgtggtg gtggtgatga 420
 tectcataaa atacaggagg cteegaatta tgaccaacat ctacctgctc aacctggcca 480
 ttteggacct gctcttctc gtcaccttc cattctggat ccactatgtc agggggcata 540
 actgggtttt tggccatggc atgtgtaage tcctctcagg gttttatcac acaggcttgt 600
 acagcgagat ctttttcata atcctgctga caatcgacag gtacctggcc attgtccatg 660
 ctgtgtttgc ccttegagcc cggactgtca cttttgggtg cateaccage atcgtcacct 720
 ggggcctggc agtgctagca gctcttcctg aatttatcct ctatgagact gaagagttgt 780
 ttgaagagac tctttgcagt gctctttacc cagaggatac agtatatagc tggaggcatt 840
 tccacactct gagaatgacc atcttctgtc tegtctctcc tetgctcgtt atggecatct 900
 gctacacagg aatcatcaaa acgtgctga ggtgccccag taaaaaaaaag tacaaggcca 960
 tccgctcat tttgtcatc atggcggtgt ttttcatttt ctggacaccc tacaatgtgg 1020
 ctatccttct ctcttctat caatccatct tatitggaaa tgactgtgag cggagcaagc 1080
 atctggacct ggteatgctg gtgacagagg tgategccta ctccactgc tgcataacc 1140
 cggatgateta cgcctttgtt ggagagaggt tccggaagta cctgegccac ttcttcaca 1200
 ggcacttget catgcacctg ggcagataca tccattcct tctagttag aagctggaaa 1260

[0005]

```

gaaccagctc tgtctctcca tccacagcag agccggaact ctctattgtg ttttaggtca      1320
gatgcagaaa attgcctaaa gaggaaggac caaggagatg aagcaaacac attaagcctt      1380
ccacactcac ctctaaaaca gtcttcaaaa ettceagtgc aacactgaag ctcttgaaga      1440
cactgaaata tacacacagc agtagcagta gatgcatgta ccctaaggtc attaccacag      1500
gccaggggct gggcagcgta ctcatcatca accctaaaaa gcagagcttt gcttctctct      1560
ctaaaatgag ttacctacat tttaatgcac ctgaatgta gatagttact atatgccgct      1620
acaaaaaggt aaaacttttt atattttata cattaacttc agccagctat tgatataaat      1680
aaaacatttt cacacaatac aataagttaa ctattttatt ttetaatgtg cctagttctt      1740
tcctgctta atgaaaagct tgttttttca gtgtgaataa ataategtaa gcaaca      1796

```

<210> 4
 <211> 355
 <212> PRT
 <213> 智人

<400> 4

```

Met Thr Thr Ser Leu Asp Thr Val Glu Thr Phe Gly Thr Thr Ser Tyr
1           5           10           15

```

```

Tyr Asp Asp Val Gly Leu Leu Cys Glu Lys Ala Asp Thr Arg Ala Leu
          20           25           30

```

```

Met Ala Gln Phe Val Pro Pro Leu Tyr Ser Leu Val Phe Thr Val Gly
          35           40           45

```

```

Leu Leu Gly Asn Val Val Val Val Met Ile Leu Ile Lys Tyr Arg Arg
          50           55           60

```

```

Leu Arg Ile Met Thr Asn Ile Tyr Leu Leu Asn Leu Ala Ile Ser Asp
65           70           75           80

```

```

Leu Leu Phe Leu Val Thr Leu Pro Phe Trp Ile His Tyr Val Arg Gly
          85           90           95

```

```

His Asn Trp Val Phe Gly His Gly Met Cys Lys Leu Leu Ser Gly Phe
100           105           110

```

```

Tyr His Thr Gly Leu Tyr Ser Glu Ile Phe Phe Ile Ile Leu Leu Thr
115           120           125

```

Ile Asp Arg Tyr Leu Ala Ile Val His Ala Val Phe Ala Leu Arg Ala
130 135 140

Arg Thr Val Thr Phe Gly Val Ile Thr Ser Ile Val Thr Trp Gly Leu
145 150 155 160

Ala Val Leu Ala Ala Leu Pro Glu Phe Ile Phe Tyr Glu Thr Glu Glu
165 170 175

Leu Phe Glu Glu Thr Leu Cys Ser Ala Leu Tyr Pro Glu Asp Thr Val
180 185 190

Tyr Ser Trp Arg His Phe His Thr Leu Arg Met Thr Ile Phe Cys Leu
195 200 205

Val Leu Pro Leu Leu Val Met Ala Ile Cys Tyr Thr Gly Ile Ile Lys
210 215 220

Thr Leu Leu Arg Cys Pro Ser Lys Lys Lys Tyr Lys Ala Ile Arg Leu
225 230 235 240

[0006]

Ile Phe Val Ile Met Ala Val Phe Phe Ile Phe Trp Thr Pro Tyr Asn
245 250 255

Val Ala Ile Leu Leu Ser Ser Tyr Gln Ser Ile Leu Phe Gly Asn Asp
260 265 270

Cys Glu Arg Ser Lys His Leu Asp Leu Val Met Leu Val Thr Glu Val
275 280 285

Ile Ala Tyr Ser His Cys Cys Met Asn Pro Val Ile Tyr Ala Phe Val
290 295 300

Gly Glu Arg Phe Arg Lys Tyr Leu Arg His Phe Phe His Arg His Leu
305 310 315 320

Leu Met His Leu Gly Arg Tyr Ile Pro Phe Leu Pro Ser Glu Lys Leu
325 330 335

Glu Arg Thr Ser Ser Val Ser Pro Ser Thr Ala Glu Pro Glu Leu Ser
340 345 350

[0007]

Ile Val Phe
355<210> 5
<211> 1227
<212> DNA
<213> 智人

<400> 5
 ctttgcagat aaatatggca cactagcccc acgttttctg agacattcct caattgctta 60
 gacatattct gagectacag cagaggaacc tccagtctca gcaccatgaa tcaaactgcc 120
 attctgattt gctgecttat ctttctgact ctaagtggea ttcaaggagt acctctctct 180
 agaactgtac gctgtacctg catcagcatt agtaateaac ctgttaatcc aaggctctta 240
 gaaaaacttg aaattattcc tgcaagccaa ttttgtecac gtgttgagat cattgctaca 300
 atgaaaaaga agggtgagaa gagatgtctg aatccagaat cgaaggccat caagaattta 360
 ctgaaagcag ttagcaagga aaggtctaaa agatctcctt aaaaccagag gggagcaaaa 420
 tcgatgcagt gcttccaagg atggaccaca cagaggetgc ctctcccatc acttcctac 480
 atggagtata tgtcaagcca taattgttct tagtttgcag ttacactaaa aggtgaccaa 540
 tgatggtcac caaatcagct gctactactc ctgtaggaag gttaatgttc atcatcctaa 600
 gctattcagt aataactcta ccttggaact ataatgtaag ctctactgag gtgctatgtt 660
 cttagtggat gttctgaccc tgettcaaat atttccctca ctttcccat cttccaaggg 720
 tactaaggaa tctttctgct ttggggttta tcagaattct cagaatctca aataactaaa 780
 aggtatgcaa tcaaatctgc tttttaaaga atgctcttta ctcatggac ttccactgcc 840
 atctctccaa ggggccc aaa tttttcagt ggtacctac atacaattcc aaacacatac 900
 aggaaggtag aaatatctga aaatgtatgt gtaagtatc ttatttaatg aaagactgta 960
 caaagtagaa gtcttagatg tatatatctt ctatattgtt ttcagtgtac atggaataac 1020
 atgtaattaa gtactatgta tcaatgagta acaggaaaat tttaaaaata cagatagata 1080
 tatgctctgc atgttacata agataaatgt gctgaatggt tttcaaaaata aaaatgaggt 1140
 actctcctgg aaatattaag aaagactatc taaatgttga aagatcaaaa ggttaataaa 1200
 gtaattataa ctaagaaaaa aaaaaaa 1227

<210> 6
<211> 98
<212> PRT
<213> 智人

[0008]

<400> 6

Met Asn Gln Thr Ala Ile Leu Ile Cys Cys Leu Ile Phe Leu Thr Leu
1 5 10 15

Ser Gly Ile Gln Gly Val Pro Leu Ser Arg Thr Val Arg Cys Thr Cys
20 25 30

Ile Ser Ile Ser Asn Gln Pro Val Asn Pro Arg Ser Leu Glu Lys Leu
35 40 45

Glu Ile Ile Pro Ala Ser Gln Phe Cys Pro Arg Val Glu Ile Ile Ala
50 55 60

Thr Met Lys Lys Lys Gly Glu Lys Arg Cys Leu Asn Pro Glu Ser Lys
65 70 75 80

Ala Ile Lys Asn Leu Leu Lys Ala Val Ser Lys Glu Arg Ser Lys Arg
85 90 95

Ser Pro

<210> 7

<211> 1490

<212> DNA

<213> 智人

<400> 7

```

agatatccgc ccctgacacc attcctccct tccccctcc aecggccgcg ggcataaaag      60
gcgccagggtg agggcctcgc cgctcctccc gcgaatcgca gcttctgaga ccagggttgc      120
tccgtccgtg ctccgcctcg ccatgacttc ctacagctat cgccagtcgt cggccacgtc      180
gtccttcgga ggctgggcg gcggtccgt gcgttttggg ccgggggtcg cctttcgcgc      240
gcccagcatt cacggggggt cggcgggcgc cggcgtatcc gtgtcctcgc cccgctttgt      300
gtcctcgtec tctcggggg cctacggcgg cggtacggc ggcgtcctga ccgcgtccga      360
cgggctgctg gcgggcaacg agaagctaac catgcagaac ctcaacgacc gcctggcctc      420
ctacctggac aagggtgcgc cctggaggc ggccaacggc gagctagagg tgaagatccg      480
cgactggtae cagaagcagg ggctgggccc ctcccgcgac tacagccact actacacgac      540
catccaggac ctgcgggaca agattcttgg tgccaccatt gagaactcca ggattgtcct      600

```

[0009]

```

gcagatcgac aatgeecgtc tggctgcaga tgacttccga accaagtttg agacggaaca      660
ggctctgcgc atgagcgtgg aggccgacat caacggcctg cgcagggtgc tggatgagct      720
gaccttggcc aggaccgacc tggagatgea gategaagge ctgaaggaag agctggceta      780
cctgaagaag aacctatgagg aggaaatcag tacgctgagg ggccaagtgg gaggccaggt      840
cagtgtggag gtggattccg ctccgggcac cgatctcgcc aagatcctga gtgacatgcg      900
aagccaatat gaggtcatgg ccgagcagaa ccggaaggat gctgaagcct ggttcaccag      960
ccggactgaa gaattgaacc gggaggtcgc tggccacacg gacgagctcc agatgagcag     1020
gtccgagggtt actgacctgc ggcgaccctt tcagggtctt gagattgagc tgcagtcaca     1080
gctgagcatg aaagetgcct tggaagacac actggcagaa accgaggcgc gctttggagc     1140
ccagctggcg catatccagg cgtgatcag cggattgaa gccagctgg gegatgtgcg     1200
agctgatagt gagcggcaga atcaggagta ccagcggctc atggacatca agtcgaggct     1260
ggagcaggag attgccacct accgcagcct gctcgaggga caggaagatc actacaacaa     1320
tttgtctgcc tccaaggtcc tctgaggcag caggctctgg ggtttctgct gtcctttgga     1380
gggtgtcttc tgggtagagg gatgggaagg aagggaccct tacccecggc tcttctcttg     1440
acctgccaat aaaaatttat ggtccaaggg aaaaaaaaaa aaaaaaaaaa     1490

```

<210> 8
 <211> 400
 <212> PRT
 <213> 智人

<400> 8

```

Met Thr Ser Tyr Ser Tyr Arg Gln Ser Ser Ala Thr Ser Ser Phe Gly
1           5           10           15

```

```

Gly Leu Gly Gly Gly Ser Val Arg Phe Gly Pro Gly Val Ala Phe Arg
20           25           30

```

```

Ala Pro Ser Ile His Gly Gly Ser Gly Gly Arg Gly Val Ser Val Ser
35           40           45

```

```

Ser Ala Arg Phe Val Ser Ser Ser Ser Ser Gly Ala Tyr Gly Gly Gly
50           55           60

```

```

Tyr Gly Gly Val Leu Thr Ala Ser Asp Gly Leu Leu Ala Gly Asn Glu
65           70           75           80

```

Lys Leu Thr Met Gln Asn Leu Asn Asp Arg Leu Ala Ser Tyr Leu Asp
85 90 95

Lys Val Arg Ala Leu Glu Ala Ala Asn Gly Glu Leu Glu Val Lys Ile
100 105 110

Arg Asp Trp Tyr Gln Lys Gln Gly Pro Gly Pro Ser Arg Asp Tyr Ser
115 120 125

His Tyr Tyr Thr Thr Ile Gln Asp Leu Arg Asp Lys Ile Leu Gly Ala
130 135 140

Thr Ile Glu Asn Ser Arg Ile Val Leu Gln Ile Asp Asn Ala Arg Leu
145 150 155 160

Ala Ala Asp Asp Phe Arg Thr Lys Phe Glu Thr Glu Gln Ala Leu Arg
165 170 175

Met Ser Val Glu Ala Asp Ile Asn Gly Leu Arg Arg Val Leu Asp Glu
180 185 190

[0010]

Leu Thr Leu Ala Arg Thr Asp Leu Glu Met Gln Ile Glu Gly Leu Lys
195 200 205

Glu Glu Leu Ala Tyr Leu Lys Lys Asn His Glu Glu Glu Ile Ser Thr
210 215 220

Leu Arg Gly Gln Val Gly Gly Gln Val Ser Val Glu Val Asp Ser Ala
225 230 235 240

Pro Gly Thr Asp Leu Ala Lys Ile Leu Ser Asp Met Arg Ser Gln Tyr
245 250 255

Glu Val Met Ala Glu Gln Asn Arg Lys Asp Ala Glu Ala Trp Phe Thr
260 265 270

Ser Arg Thr Glu Glu Leu Asn Arg Glu Val Ala Gly His Thr Glu Gln
275 280 285

Leu Gln Met Ser Arg Ser Glu Val Thr Asp Leu Arg Arg Thr Leu Gln
290 295 300

[0011]

Gly Leu Glu Ile Glu Leu Gln Ser Gln Leu Ser Met Lys Ala Ala Leu
305 310 315 320

Glu Asp Thr Leu Ala Glu Thr Glu Ala Arg Phe Gly Ala Gln Leu Ala
325 330 335

His Ile Gln Ala Leu Ile Ser Gly Ile Glu Ala Gln Leu Gly Asp Val
340 345 350

Arg Ala Asp Ser Glu Arg Gln Asn Gln Glu Tyr Gln Arg Leu Met Asp
355 360 365

Ile Lys Ser Arg Leu Glu Gln Glu Ile Ala Thr Tyr Arg Ser Leu Leu
370 375 380

Glu Gly Gln Glu Asp His Tyr Asn Asn Leu Ser Ala Ser Lys Val Leu
385 390 395 400

<210> 9
<211> 931
<212> DNA
<213> 智人

<400> 9
tttcgtcgge ccgccccttg gcttctgcac tgatggtggg tggatgagta atgcatccag 60
gaagcctgga ggcctgtggt ttccgcaccc gctgccaccc ccgcccctag cgtggacatt 120
tatectctag cgctcaggec ctgccgccat cgccgcagat ccagcgcaca gagagacacc 180
agagaaccca ccattgcccc ctttgagccc ctggttcttg gcatcctggt gtgtctgtgg 240
ctgatagccc ccagcagggc ctgcacctgt gtccacccc acccacagac ggccttctgc 300
aattccgacc tcgtcctcag ggccaagtgc gtggggacac cagaagtcaa ccagaccacc 360
ttataccagc gttatgagat caagatgacc aagatgtata aagggttcca agccttaggg 420
gatgccgttg acatccggtt cgtctacacc cccgccatgg agagtgtctg cggatacttc 480
cacaggtecc acaaccgcag cgaggagttt ctcatctgtg gaaaactgca ggatggactc 540
ttgcacatca ctacctgcag ttttgtggct ccctggaaca gcctgagctt agctcagegc 600
cggggcttca ccaagacctc cactgttggc tgtgaggaat gcacagtgtt tccctgttta 660
tccatccctt gcaaactgca gactggcact cattgcttgt ggacggacca gctcctccaa 720
ggctctgaaa agggcttcca gtcccgtcac cttgctgtgc tgcctcggga gccagggtg 780
tgcacctgga agtccctgcg gtcccagata gctgaatcc tgcctggagt ggaagctgaa 840

[0012]

gcctgcacag tgtccacctt gtcccactc ccatttttct tccggacaat gaaataaaga 900

gttaccaccc agcagaaaaa aaaaaaaaaa a 931

<210> 10

<211> 207

<212> PRT

<213> 智人

<400> 10

Met Ala Pro Phe Glu Pro Leu Ala Ser Gly Ile Leu Leu Leu Leu Trp
1 5 10 15

Leu Ile Ala Pro Ser Arg Ala Cys Thr Cys Val Pro Pro His Pro Gln
20 25 30

Thr Ala Phe Cys Asn Ser Asp Leu Val Ile Arg Ala Lys Phe Val Gly
35 40 45

Thr Pro Glu Val Asn Gln Thr Thr Leu Tyr Gln Arg Tyr Glu Ile Lys
50 55 60

Met Thr Lys Met Tyr Lys Gly Phe Gln Ala Leu Gly Asp Ala Ala Asp
65 70 75 80

Ile Arg Phe Val Tyr Thr Pro Ala Met Glu Ser Val Cys Gly Tyr Phe
85 90 95

His Arg Ser His Asn Arg Ser Glu Glu Phe Leu Ile Ala Gly Lys Leu
100 105 110

Gln Asp Gly Leu Leu His Ile Thr Thr Cys Ser Phe Val Ala Pro Trp
115 120 125

Asn Ser Leu Ser Leu Ala Gln Arg Arg Gly Phe Thr Lys Thr Tyr Thr
130 135 140

Val Gly Cys Glu Glu Cys Thr Val Phe Pro Cys Leu Ser Ile Pro Cys
145 150 155 160

Lys Leu Gln Ser Gly Thr His Cys Leu Trp Thr Asp Gln Leu Leu Gln
165 170 175

[0013]

Gly Ser Glu Lys Gly Phe Gln Ser Arg His Leu Ala Cys Leu Pro Arg
 180 185 190

Glu Pro Gly Leu Cys Thr Trp Gln Ser Leu Arg Ser Gln Ile Ala
 195 200 205

<210> 11

<211> 3452

<212> DNA

<213> 智人

<400> 11

```

gtctcagttc ccgagcctgg gagcaaccgc agctttctagt atccagactc cagcgcgcgc 60
ccgggcgcgg accccaaccc cgaccagag cttctccage ggcggcgcag cgagcagggc 120
tccccgcett aatttcctcc ggggggcccc gccaccttcg ggagtcgggg ttgcccacct 180
gcaaactctc cgccttctgc acctgccacc cctgagccag cgcgggccc cgagcgagtc 240
atggccaacg cggggctgca gctgttgggc ttctattctc ctttcctggg atggatcggc 300
gccatcgtea gcactgccct gcccagtgagg aggttttact cctatgccgg cgacaacatc 360
gtgaccgccc aggccatgta cgaggggctg tggatgtctt gcgtgtcgca gagcaccggg 420
cagatccagt gcaaagtctt tgactccttg ctgaatctga gcagcacatt gcaagcaacc 480
cgtgccttga tgggtggttg cactctcttg ggagtgatag caatctttgt ggccaccgtt 540
ggcatgaagt gtatgaagtg cttggaagac gatgaggtgc agaagatgag gatggctgtc 600
attgggggtg cgatatttct tcttgagggt ctggctattt tagttgccac agcatgggat 660
ggcaatagaa tcgttcaaga attctatgac cctatgacce cagtcaatgc caggtacgaa 720
tttggtcagg ctctcttcac tggctgggct gctgcttctc tctgccttct gggaggtgcc 780
ctactttgct gttcctgtcc ccgaaaaaca acctcttacc caacaccaag gccctatcca 840
aaacctgcac cttccagcgg gaaagactac gtgtgacaca gaggcaaaag gagaaaatca 900
tgttgaaaca aaccgaaaat ggacattgag atactatcat taacattagg accttagaat 960
tttgggtatt gtaatctgaa gtatggtatt acaaaacaaa caaacaacaa aaaaacccat 1020
gtgttaaaat actcagtgtc aaacatggct taatcttatt ttatcttctt tctcaatat 1080
aggagggaag atttttccat ttgtattact gcttcccatt gagtaatcat actcaactgg 1140
gggaaggggt gctccttaaa tatatataga tatgtatata tacatgtttt tctattaaaa 1200
atagacagta aaatactatt ctcatatgt tgatactage ataactaaaa tatctctaaa 1260
ataggtaaat gtatttaatt ccatattgat gaagatgttt attggtatat ttcttttttc 1320

```

[0014]

gtctatatat acatatgtaa cagtcaaata tcatttactc ttcttcatta gctttgggtg	1380
cctttgccac aagacctagc ctaatttacc aaggatgaat tctttcaatt cttcatgcgt	1440
gccccittca tatacittatt ttatitttta ccataatett atagcaettg catcgttatt	1500
aagcccttat ttgttttggg ttccattggg ctctatctcc tgaatctaac acatttcata	1560
gcctacattt tagttttctaa agccaagaag aattttattac aaatcagaac ttggaggca	1620
aatctttctg catgacaaaa gtgataaatt cctgttgacc ttcccacaca atccctgtac	1680
tctgacccat agcaactctg ttgtctttga aaatatttgt ccaattgagt agctgcatgc	1740
tgttccccca ggtgttgtaa cacaacttta ttgattgaat ttttaageta cttattcata	1800
gttttatatc cccetaaact acctttttgt tcccattcc ttaattgtat tgttttccca	1860
agtgtaatat tcatgcgttt tatactctcc taataagggtg tggctctgtt gtctgaacaa	1920
agtgtagac ttcttgaggat gataatctgg tgacaaatat tctctctgta gctgtaagca	1980
agtcacttaa tctttctacc tcttttttct atctgccaaa ttgagataat gatacttaac	2040
cagttagaag aggtagtgtg aatattaatt agtttatatt actctcatte ttgaacatg	2100
aactatgcct atgtagtgtc ttattttgtc cagctggctg agacactgaa gaagtcactg	2160
aacaaaacct acacaegtac ctteatgtga tteactgect tectctctct accagtctat	2220
ttccactgaa caaaacctac acacatacct tcatgtgggt cagtgccttc ctctctctac	2280
cagtctatit ccactgaaca aaacctacgc acataccttc atgtggctca gtgccttcct	2340
ctctctacca gtctatttcc attctttcag ctgtgtctga catgtttgtg ctctgttcca	2400
ttttaacaac tgctcttact ttccagttc gtacagaatg ctatttcact tgagcaagat	2460
gatgtaattg aaaggggtgt ggcatgtgtg tctggagacc tggatttgag tcttgggtgt	2520
atcaatcacc gtctgtgttt gagcaaggca ttiggtgtgt gtaagcttat tgcctcatct	2580
gtaagcgggt gtttgtaatt cctgatcttc ccacctcaca gtgatgttgt ggggatccag	2640
tgagatagaa tacatgtaag tgtggttttg taatttaaaa agtgcctatac taagggaag	2700
aattgaggaa ttaactgcat acgttttggg gtgtcttttc aaatgtttga aaacaaaaaa	2760
aatgttaaga aatgggtttc ttgccttaac cagtctctca agtgatgaga cagtgaagta	2820
aaattgagtg cactaaacaa ataagattct gaggaagtct tatctcttgc agtgagtatg	2880
gcccgatgct ttctgtggct aaacagatgt aatgggaaga aataaaagcc tacgtgttgg	2940
taaatccaac agcaagggtg atttttgaat cataataact cataaggtgc tatctgttca	3000

[0015]

```

gtgatgcct cagagetctt gctgttagct ggcagctgac gctgctagga tagttagttt    3060
ggaaatggta cttcataata aactacacaa ggaaagtcag ccactgtgtc ttatgaggaa    3120
ttggacctaa taaatttttag tgtgccttcc aaacctgaga atatattgctt ttggaagtta    3180
aaatttfaat ggctttttgcc acatacatag atcttcatga tgtgtgagtg taattccatg    3240
tggatatcag ttaccaaaca ttacaaaaaa attttatggc ccaaaatgac caacgaaatt    3300
gttacaatag aatttatecca attttgatct ttttatattc ttctaccaca cctggaaaca    3360
gaccaataga cattttggggg ttttataata ggaatttgta taaagcatta ctctttttca    3420
ataaattggt ttttaattta aaaaaaggat ta                                3452

```

<210> 12
 <211> 211
 <212> PRT
 <213> 智人

<400> 12

Met Ala Asn Ala Gly Leu Gln Leu Leu Gly Phe Ile Leu Ala Phe Leu
 1 5 10 15

Gly Trp Ile Gly Ala Ile Val Ser Thr Ala Leu Pro Gln Trp Arg Ile
 20 25 30

Tyr Ser Tyr Ala Gly Asp Asn Ile Val Thr Ala Gln Ala Met Tyr Glu
 35 40 45

Gly Leu Trp Met Ser Cys Val Ser Gln Ser Thr Gly Gln Ile Gln Cys
 50 55 60

Lys Val Phe Asp Ser Leu Leu Asn Leu Ser Ser Thr Leu Gln Ala Thr
 65 70 75 80

Arg Ala Leu Met Val Val Gly Ile Leu Leu Gly Val Ile Ala Ile Phe
 85 90 95

Val Ala Thr Val Gly Met Lys Cys Met Lys Cys Leu Glu Asp Asp Glu
 100 105 110

Val Gln Lys Met Arg Met Ala Val Ile Gly Gly Ala Ile Phe Leu Leu
 115 120 125

Ala Gly Leu Ala Ile Leu Val Ala Thr Ala Trp Tyr Gly Asn Arg Ile

[0016]

Val	Gln	Glu	Phe	Tyr	Asp	Pro	Met	Thr	Pro	Val	Asn	Ala	Arg	Tyr	Glu
145					150					155					160
Phe	Gly	Gln	Ala	Leu	Phe	Thr	Gly	Trp	Ala	Ala	Ala	Ser	Leu	Cys	Leu
				165					170					175	
Leu	Gly	Gly	Ala	Leu	Leu	Cys	Cys	Ser	Cys	Pro	Arg	Lys	Thr	Thr	Ser
			180					185					190		
Tyr	Pro	Thr	Pro	Arg	Pro	Tyr	Pro	Lys	Pro	Ala	Pro	Ser	Ser	Gly	Lys
		195					200					205			
Asp	Tyr	Val													
	210														

<210> 13
 <211> 1669
 <212> DNA
 <213> 智人

<400> 13
 ggcaacecca cgcggetgga gaagccggcg ctgcagccc ggeccgggce getgccggaa 60
 gtgacgcgag ttcacctgcc gagcgggggc tgggaggagg ggcggagggt gcagaggtgc 120
 cgccgccgcc gcgagccagt cgggagcgcg cgaggcgcgg ggagcctggg accaggagcg 180
 agagccgcct acctgcagcc gccgcccacg gcacggcagc caccatggcg ctctgtctgt 240
 gcttcgtget cctgtgcgga gtagtggatt tcgccagaag tttagtate actactctg 300
 aagagatgat tgaaaaagcc aaaggggaaa ctgcctatct gccatgcaa tttacgctta 360
 gtcccgaaga ccagggaccg ctggacateg agtggctgat atcaccagct gataatcaga 420
 aggtggatca agtgattatt ttatattctg gagacaaaat ttatgatgac tactatccag 480
 atctgaaagg ccgagtacat tttacgagta atgattcaa atctggtgat gcatcaataa 540
 atgtaacgaa ttacaactg tcagatattg gcacatatca gtgcaaagtg aaaaaagctc 600
 ctgggtgttg aaataagaag attcatctgg tagttcttgt taagccttca ggtgcgagat 660
 gttacgttga tggatctgaa gaaattggaa gtgactttaa gataaaatgt gaacccaaag 720
 aaggttcact tccattacag tatgagtggc aaaaattgtc tgactcacag aaaatgccca 780
 cttcatgggt agcagaaatg acttcatctg ttatatctgt aaaaaatgcc tcttctgagt 840

[0017]

```

actctgggac atacagctgt acagtcagaa acagagtggg ctctgatcag tgccgtttgc      900
gtctaaacgt tgccctcct tcaaataaag ctggactaat tgcaggagcc attataggaa      960
ctttgcttgc tctagegetc attggtctta tcattttttg ctgtcgtaaa aagegcagag    1020
aagaaaaata tgaaaaggaa gttcatcacg atatcaggga agatgtgeca cctccaaaga    1080
gccgtacgtc cactgccaga agctacatcg gcagtaatca ttcateccctg gggtcctatgt    1140
ctccttccaa catggaagga tattccaaga ctcagtataa ccaagtacca agtgaagact    1200
ttgaacgcac tcctcagagt ccgactctcc cacctgctaa gttcaagtac ccttacaaga    1260
ctgatggaat tacagttgta taaatatgga ctactgaaga atctgaagta ttgtattatt    1320
tgacttttatt ttaggectct agtaaagact taaatgtttt ttaaaaaaag cacaaggcac    1380
agagattaga gcagctgtaa gaacacatct actttatgca atggcattag acatgtaagt    1440
cagatgtcat gtcaaaatta gtacgagcca aattctttgt taaaaaaccc tatgtatagt    1500
gacactgata gttaaaagat gttttattat attttcaata actaccacta acaaattttt    1560
aacttttcat atgcatatte tgatatgtgg tcttttagga aaagtatggg taatagttga    1620
tttttcaaag gaaattttta aattcttaag ttctgtttta tgtttttgc      1669

```

<210> 14
 <211> 352
 <212> PRT
 <213> 智人

<400> 14

```

Met Ala Leu Leu Leu Cys Phe Val Leu Leu Cys Gly Val Val Asp Phe
1           5           10           15

```

```

Ala Arg Ser Leu Ser Ile Thr Thr Pro Glu Glu Met Ile Glu Lys Ala
          20           25           30

```

```

Lys Gly Glu Thr Ala Tyr Leu Pro Cys Lys Phe Thr Leu Ser Pro Glu
          35           40           45

```

```

Asp Gln Gly Pro Leu Asp Ile Glu Trp Leu Ile Ser Pro Ala Asp Asn
50           55           60

```

```

Gln Lys Val Asp Gln Val Ile Ile Leu Tyr Ser Gly Asp Lys Ile Tyr
65           70           75           80

```

```

Asp Asp Tyr Tyr Pro Asp Leu Lys Gly Arg Val His Phe Thr Ser Asn

```

	85	90	95
	Asp Leu Lys Ser Gly Asp Ala Ser Ile Asn Val Thr Asn Leu Gln Leu 100	105	110
	Ser Asp Ile Gly Thr Tyr Gln Cys Lys Val Lys Lys Ala Pro Gly Val 115	120	125
	Ala Asn Lys Lys Ile His Leu Val Val Leu Val Lys Pro Ser Gly Ala 130	135	140
	Arg Cys Tyr Val Asp Gly Ser Glu Glu Ile Gly Ser Asp Phe Lys Ile 145	150	155
	Lys Cys Glu Pro Lys Glu Gly Ser Leu Pro Leu Gln Tyr Glu Trp Gln 165	170	175
	Lys Leu Ser Asp Ser Gln Lys Met Pro Thr Ser Trp Leu Ala Glu Met 180	185	190
[0018]	Thr Ser Ser Val Ile Ser Val Lys Asn Ala Ser Ser Glu Tyr Ser Gly 195	200	205
	Thr Tyr Ser Cys Thr Val Arg Asn Arg Val Gly Ser Asp Gln Cys Leu 210	215	220
	Leu Arg Leu Asn Val Val Pro Pro Ser Asn Lys Ala Gly Leu Ile Ala 225	230	235
	Gly Ala Ile Ile Gly Thr Leu Leu Ala Leu Ala Leu Ile Gly Leu Ile 245	250	255
	Ile Phe Cys Cys Arg Lys Lys Arg Arg Glu Glu Lys Tyr Glu Lys Glu 260	265	270
	Val His His Asp Ile Arg Glu Asp Val Pro Pro Pro Lys Ser Arg Thr 275	280	285
	Ser Thr Ala Arg Ser Tyr Ile Gly Ser Asn His Ser Ser Leu Gly Ser 290	295	300
	Met Ser Pro Ser Asn Met Glu Gly Tyr Ser Lys Thr Gln Tyr Asn Gln 305	310	315
			320

[0019]

Val Pro Ser Glu Asp Phe Glu Arg Thr Pro Gln Ser Pro Thr Leu Pro
 325 330 335

Pro Ala Lys Phe Lys Tyr Pro Tyr Lys Thr Asp Gly Ile Thr Val Val
 340 345 350

<210> 15
 <211> 3997
 <212> DNA
 <213> 智人

<400> 15

```

gtgccaaggc ggctgggggc ggcgagcggg gccgcgggcg cgcaccgact caagagccga      60
ctgtcagcct cggcggggccg gagttctccg gcgctgggac aggggcgctg ggacaggggc      120
gctggggggcg agccctggcg ggggccaggt ccgaggacct tgggcgcggc ggccccgcca      180
ggaggtccgg ccgcgagcgt gacctcacgg ggagggggcca gcgcggcggc ctgggcgctg      240
agccgagcgc cgggagagca gcgcagaagc cgagccgcga ggagcgcact ccgtggcccc      300
gatggagcgg tacaaagccc tggaacagct gctgacagag ttggatgact tectcaagat      360
tcttgaccag gagaacctga gcagcacagc actggtgaag aagagctgcc tggcggagct      420
cctccggett tacaccaaaa gcagcagetc tgatgaggag tacatttata tgaacaaagt      480
gaccatcaac aagcaacaga atgcagagtc tcaaggcaaa gcgcctgagg agcagggcct      540
gctacceaat ggggagccca gccagcactc ctcggcccct cagaagagcc ttccagacct      600
cccgccacce aagatgattc cagaacggaa acagcttgcc atcccaaaga cggagtctcc      660
agagggttac tatgaagagg ctgagccata tgacacatcc ctcaatgagg acggagaggc      720
tgtgagcagc tcttacgagt cctacgatga agaggacggc agcaagggca agtcggcccc      780
ttaccagtgg ccctcgccgg aggcgggcat cgagctgatg cgtgacgccc gcatctgcgc      840
cttctgtggt cgcaagaagt ggctgggaca gtgggccaag cagctctgtg tcatcaagga      900
caacaggctt ctgtgttaca aatcctccaa ggaccacagc cctcagctgg acgtgaacct      960
actgggcagc agcgtcattc acaaggagaa gcaagtgcgg aagaaggagc acaagctgaa     1020
gatcacaccg atgaatgccg atgtgattgt gctgggcctg cagagcaagg accaggtga      1080
gcagtggctc agggtcattc aggaagtgag cggcctgcct tccgaaggag catctgaagg     1140
aaaccagtac accccggatg cccagcgctt taactgccag aaaccagata tagctgagaa     1200
gtacctgtcg gcttcagagt atgggagetc cgtggatggc caccctgagg tcccagaaac     1260

```


[0020]

caaagacgtc aagaagaaat gttctgctgg cctcaaactg agcaacctaa tgaatctggg	1320
caggaagaaa tccacctcac tggagcctgt ggagagggtc ctcgagacat ccagttacct	1380
gaacgtgctg gtgaacagcc agtggaagtc tcgctgggtc tctgtcaggg acaatcacct	1440
gcatttctac caggaccgga accggagcaa ggtggeccag caaccctca gcctgggtggg	1500
ctgcgagggtg gtcccagacc ccagccccga ccacctctac tcttccgca tctccacaa	1560
gggcgaggag ctggccaagc ttgaggccaa gtcttccgag gaaatgggce actggctggg	1620
tctctgtctc tctgagtcag gctccaagac agaccagaa gagttcacct acgactatgt	1680
ggatgccgat agggctctct gtattgtgag tgcggccaaa aactctctct tactgatgca	1740
gagaaagttc tcagagccca acacttacat cgatggcctg cctagccagg accgccagga	1800
ggagctgtat gacgacgtgg acctgtcaga gctcacagct gcggtggage ctaccgagga	1860
agccaccctt gttgcagatg acccaaatga gagagaatct gaccgagtgt acctggacct	1920
cacacctgtc aagtcctttc tgcattggcc cagcagtga caggcccagg cctctctccc	1980
gacgttgtec tgccctggaca atgcaactga ggccctcccg gcagactcag gccaggtcc	2040
caccccagat gagccctgca taaagtgtcc agagaacctg ggagaacagc agctggagag	2100
tttgagacca gaggatcctt ccttgagaat caccacgtc aaaatccaga cggaacagca	2160
gagaatctcc ttcccaccga gctgcccgga tgcctgtgtg gccacccac ctggtgccag	2220
cccacctgtg aaggacaggt tgcgctgac cagtgcagag atcaagcttg gcaagaatcg	2280
gacagaagct gaggtgaagc ggtacacaga ggagaaggag aggcttgaaa agaagaagga	2340
agaaatccgg gggcacctgg ctcagctccg gaaagagaaa cgggagctaa aggaaacct	2400
actgaaatgc acagacaagg aagtcttggc gagcctggag cagaagctga aggaaattga	2460
cgaggagtgc cggggcgagg agagcaggcg cgtggacctg gagctcagca tcatggaggt	2520
gaaggacaac ctgaagaagg ctgaggcagg gcctgtgacg ttaggcacca cctggacac	2580
caccacctg gagaatgtga gcccgcgcc caaagctgtc acacctgect ctgcccaga	2640
ctgtacecca gtcaactctg caaccacact caagaacagg cctctctcgg tegtgtgac	2700
aggcaaaggc actgtactcc agaaagccaa ggaatgggag aagaaaggag caagttagaa	2760
aacaagcttc atctaaagac tctcatgtca atgtggacct tggtgacaat cctgtttgt	2820
taaagcaaaa actatgcgaa aggggtgagtc tgtttagaag aaaaagcaaa gactgaggta	2880
ctgtgaatgg agagcttcag ctaagaggag gctctgtccc ttttcagagc caaaggaaat	2940

[0021]

```

aatacaacaa aaaggagget tctttggaga cctaagteta ttggatgtaa acaagacgtt      3000
gtatttaggg atgttctgtg tttctttctt ttttgaagtt gtcatacaatt gctttactaa      3060
gatttttaaa tagtgaaaac ctectgttta gactttgggtg gaagatgaat caaggaagca      3120
gggccctgtc ttatgggtca cgtgtctttg gtgagtgaga agacctaaac tectggccat      3180
catctcttat ccaatactta gcagttgggg attaaaceat ccttgccttc agttctctcc      3240
aatattacca ggcccaactc agtcttcagt gatttttaaac agcatigaca tcactgttaa      3300
aaccatcatc tgtaaaacca tctatgacat gagttttgag aaacaataat ggggaaaata      3360
tttgggacca agctgaagca ctaatcccac taagttaaag acttctttcc agtccaaggc      3420
aggcctgaat caactgtctt taaataaaaat tttaagtgat gctgtattat atataggaaa      3480
aaatgcttaa aatcctgtca tttagaacag tgaaaagtat cttttgagat taaagtgact      3540
ctttactgta ggaaaaatat tactctgtgt ttacagattc attgctgtgg tcaggccatt      3600
tttaagggaa gagttattta atataaatag tctctgattt taagttctgt ttaatgttca      3660
ttctccttcc aagaacaaag tggtgatttt tggttagggt gatcgccctc ttaaaattgg      3720
cagtgtgttt ccttgtgtg cccctgtctt ttctctgat ggcatttttt tttttttttt      3780
tttaacacag gttgaaacat ttcatctatt atctctgcct cattcttgga gggttgtgta      3840
tcagttctct aacacttggt cctgagaact aaatgtcttt ttattcttta ttctctctct      3900
cataaacatt tggtgacctt ttaccaagtg gtgagttagt taggtttttt aaaataaaaat      3960
gttcattgta ttgaaaaaa aaaaaaaaaa aaaaaaa      3997

```

<210> 16
 <211> 818
 <212> PRT
 <213> 智人

<400> 16

Met Glu Arg Tyr Lys Ala Leu Glu Gln Leu Leu Thr Glu Leu Asp Asp
 1 5 10 15

Phe Leu Lys Ile Leu Asp Gln Glu Asn Leu Ser Ser Thr Ala Leu Val
 20 25 30

Lys Lys Ser Cys Leu Ala Glu Leu Leu Arg Leu Tyr Thr Lys Ser Ser
 35 40 45

Ser Ser Asp Glu Glu Tyr Ile Tyr Met Asn Lys Val Thr Ile Asn Lys

67

Arg Phe Asn Cys Gln Lys Pro Asp Ile Ala Glu Lys Tyr Leu Ser Ala
290 295 300

Ser Glu Tyr Gly Ser Ser Val Asp Gly His Pro Glu Val Pro Glu Thr
305 310 315 320

Lys Asp Val Lys Lys Lys Cys Ser Ala Gly Leu Lys Leu Ser Asn Leu
325 330 335

Met Asn Leu Gly Arg Lys Lys Ser Thr Ser Leu Glu Pro Val Glu Arg
340 345 350

Ser Leu Glu Thr Ser Ser Tyr Leu Asn Val Leu Val Asn Ser Gln Trp
355 360 365

Lys Ser Arg Trp Cys Ser Val Arg Asp Asn His Leu His Phe Tyr Gln
370 375 380

Asp Arg Asn Arg Ser Lys Val Ala Gln Gln Pro Leu Ser Leu Val Gly
385 390 395 400

[0023]

Cys Glu Val Val Pro Asp Pro Ser Pro Asp His Leu Tyr Ser Phe Arg
405 410 415

Ile Leu His Lys Gly Glu Glu Leu Ala Lys Leu Glu Ala Lys Ser Ser
420 425 430

Glu Glu Met Gly His Trp Leu Gly Leu Leu Leu Ser Glu Ser Gly Ser
435 440 445

Lys Thr Asp Pro Glu Glu Phe Thr Tyr Asp Tyr Val Asp Ala Asp Arg
450 455 460

Val Ser Cys Ile Val Ser Ala Ala Lys Asn Ser Leu Leu Leu Met Gln
465 470 475 480

Arg Lys Phe Ser Glu Pro Asn Thr Tyr Ile Asp Gly Leu Pro Ser Gln
485 490 495

Asp Arg Gln Glu Glu Leu Tyr Asp Asp Val Asp Leu Ser Glu Leu Thr
500 505 510

Ala Ala Val Glu Pro Thr Glu Glu Ala Thr Pro Val Ala Asp Asp Pro
515 520 525

Asn Glu Arg Glu Ser Asp Arg Val Tyr Leu Asp Leu Thr Pro Val Lys
530 535 540

Ser Phe Leu His Gly Pro Ser Ser Ala Gln Ala Gln Ala Ser Ser Pro
545 550 555 560

Thr Leu Ser Cys Leu Asp Asn Ala Thr Glu Ala Leu Pro Ala Asp Ser
565 570 575

Gly Pro Gly Pro Thr Pro Asp Glu Pro Cys Ile Lys Cys Pro Glu Asn
580 585 590

Leu Gly Glu Gln Gln Leu Glu Ser Leu Glu Pro Glu Asp Pro Ser Leu
595 600 605

Arg Ile Thr Thr Val Lys Ile Gln Thr Glu Gln Gln Arg Ile Ser Phe
610 615 620

[0024]

Pro Pro Ser Cys Pro Asp Ala Val Val Ala Thr Pro Pro Gly Ala Ser
625 630 635 640

Pro Pro Val Lys Asp Arg Leu Arg Val Thr Ser Ala Glu Ile Lys Leu
645 650 655

Gly Lys Asn Arg Thr Glu Ala Glu Val Lys Arg Tyr Thr Glu Glu Lys
660 665 670

Glu Arg Leu Glu Lys Lys Lys Glu Glu Ile Arg Gly His Leu Ala Gln
675 680 685

Leu Arg Lys Glu Lys Arg Glu Leu Lys Glu Thr Leu Leu Lys Cys Thr
690 695 700

Asp Lys Glu Val Leu Ala Ser Leu Glu Gln Lys Leu Lys Glu Ile Asp
705 710 715 720

Glu Glu Cys Arg Gly Glu Glu Ser Arg Arg Val Asp Leu Glu Leu Ser
725 730 735

[0025]

Ile Met Glu Val Lys Asp Asn Leu Lys Lys Ala Glu Ala Gly Pro Val
 740 745 750

Thr Leu Gly Thr Thr Val Asp Thr Thr His Leu Glu Asn Val Ser Pro
 755 760 765

Arg Pro Lys Ala Val Thr Pro Ala Ser Ala Pro Asp Cys Thr Pro Val
 770 775 780

Asn Ser Ala Thr Thr Leu Lys Asn Arg Pro Leu Ser Val Val Val Thr
 785 790 795 800

Gly Lys Gly Thr Val Leu Gln Lys Ala Lys Glu Trp Glu Lys Lys Gly
 805 810 815

Ala Ser

<210> 17
 <211> 1606
 <212> DNA
 <213> 智人

<400> 17

```

aatgtgacc ggccggggt ceggcagtea acgcctgect cctctegage gtcttcageg      60
cagccgcgc ccggggagcc agcacgaacg agcccageac cgcccgatg gagcgtccgc      120
aaccgcacag catgccccag gatttgtagc aggcctgaa ggaggccacc aaggaggtgc      180
acaccagge agagaatgct gagttcatga ggaactttca gaagggccag gtgaccagag      240
acggcttcaa gctggtagat gcctccctgt accacateta tgtggccctg gaggaggaga      300
ttgagcgcaa caaggagagc ccagtcttcg cccctgteta ctccccagaa gagetgcacc      360
gcaaggctgc cctggagcag gacctggect tctggtacgg gcccgcctgg caggaggtca      420
tcccctaac accagccatg cagcgctatg tgaagcggt ccaagaggtg gggcgcacag      480
agcccgagct gctgggtgcc cagccctaca cccgtacct gggtagactg tctgggggcc      540
aggtgctcaa aaagattgcc cagaaagccc tggacctgcc cagctctggc gagggcctgg      600
ccttcttca cttcccacac attgccagtg ccaccaagtt caagcagetc taccgctccc      660
gcatgaactc cctggagatg actcccgtag tcaggcagag ggtgatagaa gaggccaaga      720
ctgegtteet getcaacatc cagctctttg aggagttgea ggagctgctg acctatgaca      780
ccaaggacca gagccctca cgggcaccag ggtcttgcca gcgggccagc aacaaagtgc      840

```

[0026]

```

aagatttctgc ccccgctggag actcccagag ggaagecccc actcaacacc cgctcccagg      900
ctccgctttct ccgatgggtc cttacactca gctttctggg ggcgacagtt gctgtagggc      960
tttatgccat gtgaatgcag gcatgctgga tcccagggcc atgaactttg tccggtggaa    1020
ggcctttcttt ctagagaggg aattctcttg gctggtctcc ttaccgtggg cactgaagge    1080
tttcagggcc tccagccctc tcaactgtgc cctctctctg gaaaggagga aggagcctat    1140
ggcatcttcc ccaacgaaaa gcacatccag gcaatggcct aaacttcaga gggggcgaag    1200
ggatcagccc tgccttccag catcctcagt tcctgcagca gagcctggaa gacaccctaa    1260
tgtggcagct gtctcaaacc tccaaaagcc ctgagtttca agtatccttg ttgacaeggc    1320
catgaccact ttcccgtgg gccatggcaa tttttacaca aacctgaaaa gatgttgtgt    1380
cttggtgttt tgtcttattt ttgttggage caetctgttc ctggtcagc ctcaaagca    1440
gtatttttgt tgtgttctgt tgtttttata gcagggttgg ggtggttttt gagccatgeg    1500
tgggtgggga gggaggtggt taacggcact gtggccttgg tctaaacttt gtgtgaaata    1560
ataaacaaca ttgtctgata gtagcttgaa aaaaaaaaaa aaaaaa      1606

```

<210> 18
 <211> 288
 <212> PRT
 <213> 智人

<400> 18

```

Met Glu Arg Pro Gln Pro Asp Ser Met Pro Gln Asp Leu Ser Glu Ala
1           5           10           15

```

```

Leu Lys Glu Ala Thr Lys Glu Val His Thr Gln Ala Glu Asn Ala Glu
20           25           30

```

```

Phe Met Arg Asn Phe Gln Lys Gly Gln Val Thr Arg Asp Gly Phe Lys
35           40           45

```

```

Leu Val Met Ala Ser Leu Tyr His Ile Tyr Val Ala Leu Glu Glu Glu
50           55           60

```

```

Ile Glu Arg Asn Lys Glu Ser Pro Val Phe Ala Pro Val Tyr Phe Pro
65           70           75           80

```

```

Glu Glu Leu His Arg Lys Ala Ala Leu Glu Gln Asp Leu Ala Phe Trp
85           90           95

```

[0027]

Tyr Gly Pro Arg Trp Gln Glu Val Ile Pro Tyr Thr Pro Ala Met Gln
 100 105 110

Arg Tyr Val Lys Arg Leu His Glu Val Gly Arg Thr Glu Pro Glu Leu
 115 120 125

Leu Val Ala His Ala Tyr Thr Arg Tyr Leu Gly Asp Leu Ser Gly Gly
 130 135 140

Gln Val Leu Lys Lys Ile Ala Gln Lys Ala Leu Asp Leu Pro Ser Ser
 145 150 155 160

Gly Glu Gly Leu Ala Phe Phe Thr Phe Pro Asn Ile Ala Ser Ala Thr
 165 170 175

Lys Phe Lys Gln Leu Tyr Arg Ser Arg Met Asn Ser Leu Glu Met Thr
 180 185 190

Pro Ala Val Arg Gln Arg Val Ile Glu Glu Ala Lys Thr Ala Phe Leu
 195 200 205

Leu Asn Ile Gln Leu Phe Glu Glu Leu Gln Glu Leu Leu Thr His Asp
 210 215 220

Thr Lys Asp Gln Ser Pro Ser Arg Ala Pro Gly Leu Arg Gln Arg Ala
 225 230 235 240

Ser Asn Lys Val Gln Asp Ser Ala Pro Val Glu Thr Pro Arg Gly Lys
 245 250 255

Pro Pro Leu Asn Thr Arg Ser Gln Ala Pro Leu Leu Arg Trp Val Leu
 260 265 270

Thr Leu Ser Phe Leu Val Ala Thr Val Ala Val Gly Leu Tyr Ala Met
 275 280 285

<210> 19

<211> 2207

<212> DNA

<213> 智人

<400> 19

cacttctctcc ccagacaggg gtagtgegag gccgggcaca gccttctctgt gtggttttac

60

[0028]

cgcccagaga ggcgtcatgga cctggggaaa ccaatgaaaa gcgtgctggt ggtggctctc	120
cttgtcattt tccaggtatg cctgtgtcaa gatgaggta cggacgatta catcggagac	180
aacaccacag tggactacac ttgtttcgag tctttgtgct ccaagaagga cgtgeggaac	240
tttaaagcct gggtccctccc tatcatgtac tccatcattt gtttcgtggg cctactgggc	300
aatgggctgg tcgtgttgac ctatatctat ttcaagaggc tcaagacat gaccgatacc	360
tacctgctca acctggcggt ggcagacatc ctcttctctc tgaccttcc cttctgggcc	420
tacagecggg ccaagtcctg ggtcttcggt gtccactttt geaagctcat ctttgcctc	480
tacaagatga gcttcttcag tggcatgctc ctacttcttt geatcagcat tgaccgctac	540
gtggccatcg tccaggctgt ctacagctac cgccaccgtg cccgcgtctt tctcatcagc	600
aagctgtctt gtgtgggcat ctggatacta gccacagtgc tctccatccc agagctcctg	660
tacagtgacc tccagaggag cagcagttag caagcgatgc gatgctctct catcacagag	720
catgtggagg cctttatcac catccagggt gcccagatgg tgatcggctt tctggteccc	780
ctgctggcca tgagcttctg ttaccttgte atcatccgca cctgtctcca ggcacgcaac	840
tttgagcgca acaaggecat caaggtgac atcgctgtgg tcgtggcttt catagtctt	900
cagctgecct acaatggggt ggtcttggec cagacgggtg ccaacttcaa catcaccagt	960
agcacctgtg agctcagtaa gcaactcaac atcgctacg acgtcaccta cagctggcc	1020
tgcgtccgt gctgcgtcaa cctttcttg tacgccttca tcggcgtcaa gttccgcaac	1080
gatctcttca agctcttcaa ggacctgggc tgcctcagcc aggagcagct ccggcagtgg	1140
tcttctgtc ggcacatccg gcgctcctcc atgagtgtgg aggccgagac caccaccacc	1200
ttctcccat aggegactct tctgcctgga ctagaggga cttctccagg gtccctgggg	1260
tggggatagg gagcagatgc aatgactcag gacatcccc cgccaaaage tgcacaggga	1320
aaagcagctc tccctcaga gtgcaagccc ctgctccaga agatagctt accccaatcc	1380
cagctacctc aaccaatgcc aaaaaagac agggctgata agctaacacc agacagacaa	1440
cactgggaaa cagaggctat tgtcccttaa accaaaaact gaaagtgaaa gtccagaaac	1500
tgttcccacc tgctggagtg aaggggcca ggagggtgag tgcaaggggc gtgggagtgg	1560
cctgaagagt cctctgaatg aaccttctgg cctcccacag actcaaatgc tcagaccagc	1620
tcttccgaaa accaggcctt atctccaaga ccagagatag tggggagact tcttgcttg	1680
gtgaggaaaa gcggacatca gctggtcaaa caactctct gaacctctc ctccatcgtt	1740

[0029]

```

tttttcaactg tectccaagc cagcgggaat ggcagctgcc acgcegecct aaaagcacac      1800
tcattcccttc acttgccgcg tegeccctcc aggcctctcaa caggggagag tgtggtgttt      1860
cctgcaggcc aggccagctg cctccgcgtg atcaaagcca cactctgggc tccagagtgg      1920
ggatgacatg cactcagctc ttggctccac tgggatggga ggagaggaca agggaaatgt      1980
caggggcggg gaggtgaca gtggccgcc aagggccacg agcttgttct ttgttctttg      2040
tcacagggac tgaaaacctc tctcatgtt ctgctttcga ttcgtaaga gagcaacatt      2100
ttaccacacac acagataaag ttttcccttg aggaaacaac agctttaaaa gaaaaagaaa      2160
aaaaaagtct ttggtaaatg gcaaaaaaaaa aaaaaaaaaa aaaaaaaa      2207

```

<210> 20
 <211> 378
 <212> PRT
 <213> 智人

<400> 20

```

Met Asp Leu Gly Lys Pro Met Lys Ser Val Leu Val Val Ala Leu Leu
1          5          10          15

```

```

Val Ile Phe Gln Val Cys Leu Cys Gln Asp Glu Val Thr Asp Asp Tyr
          20          25          30

```

```

Ile Gly Asp Asn Thr Thr Val Asp Tyr Thr Leu Phe Glu Ser Leu Cys
          35          40          45

```

```

Ser Lys Lys Asp Val Arg Asn Phe Lys Ala Trp Phe Leu Pro Ile Met
          50          55          60

```

```

Tyr Ser Ile Ile Cys Phe Val Gly Leu Leu Gly Asn Gly Leu Val Val
65          70          75          80

```

```

Leu Thr Tyr Ile Tyr Phe Lys Arg Leu Lys Thr Met Thr Asp Thr Tyr
          85          90          95

```

```

Leu Leu Asn Leu Ala Val Ala Asp Ile Leu Phe Leu Leu Thr Leu Pro
          100          105          110

```

```

Phe Trp Ala Tyr Ser Ala Ala Lys Ser Trp Val Phe Gly Val His Phe
          115          120          125

```

```

Cys Lys Leu Ile Phe Ala Ile Tyr Lys Met Ser Phe Phe Ser Gly Met

```

	130	135	140	
	Leu Leu Leu Leu Cys	Ile Ser Ile Asp Arg	Tyr Val Ala Ile Val Gln	
	145	150	155	160
	Ala Val Ser Ala His	Arg His Arg Ala Arg	Val Leu Leu Ile Ser Lys	
		165	170	175
	Leu Ser Cys Val Gly	Ile Trp Ile Leu Ala Thr	Val Leu Ser Ile Pro	
		180	185	190
	Glu Leu Leu Tyr Ser	Asp Leu Gln Arg Ser Ser	Ser Glu Gln Ala Met	
		195	200	205
	Arg Cys Ser Leu Ile	Thr Glu His Val Glu Ala	Phe Ile Thr Ile Gln	
		210	215	220
	Val Ala Gln Met Val	Ile Gly Phe Leu Val	Pro Leu Leu Ala Met Ser	
		225	230	235
[0030]	Phe Cys Tyr Leu Val	Ile Ile Arg Thr Leu Leu	Gln Ala Arg Asn Phe	
		245	250	255
	Glu Arg Asn Lys Ala	Ile Lys Val Ile Ile	Ala Val Val Val Val Phe	
		260	265	270
	Ile Val Phe Gln Leu	Pro Tyr Asn Gly Val	Val Leu Ala Gln Thr Val	
		275	280	285
	Ala Asn Phe Asn Ile	Thr Ser Ser Thr Cys	Glu Leu Ser Lys Gln Leu	
		290	295	300
	Asn Ile Ala Tyr Asp	Val Thr Tyr Ser Leu	Ala Cys Val Arg Cys Cys	
		305	310	315
	Val Asn Pro Phe Leu	Tyr Ala Phe Ile Gly	Val Lys Phe Arg Asn Asp	
		325	330	335
	Leu Phe Lys Leu Phe	Lys Asp Leu Gly Cys	Leu Ser Gln Glu Gln Leu	
		340	345	350
	Arg Gln Trp Ser Ser	Cys Arg His Ile Arg	Arg Ser Ser Met Ser Val	
		355	360	365
[0031]	Glu Ala Glu Thr Thr	Thr Thr Phe Ser Pro		
		370	375	

Pfaffl法

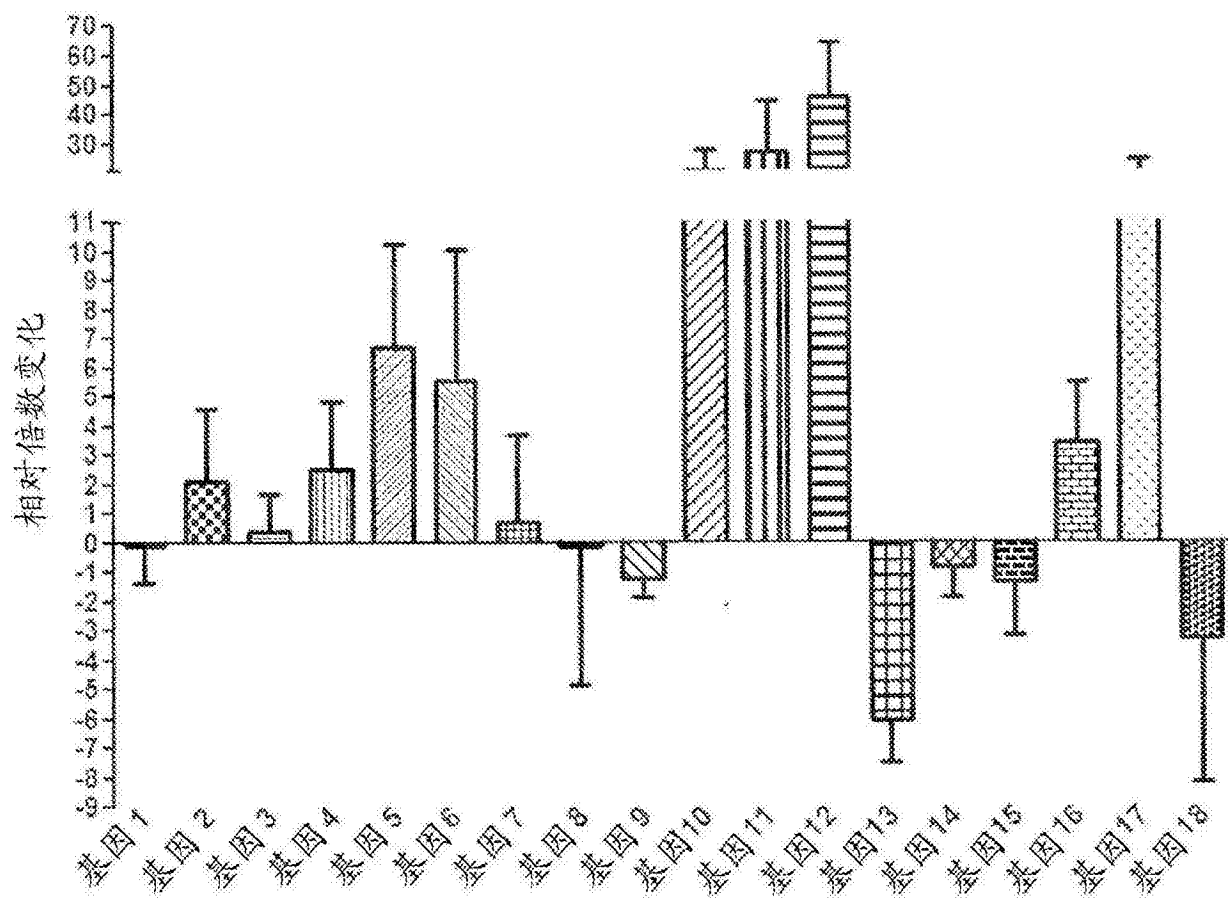


图1

分类器	名称	AUC	FP	TP
基因 1	CXCR3	0,559	0,796	0,969
基因 2	CXC3A	0,614	0,648	0,898
基因 3	CXCRB	0,611	0,759	0,918
基因 4	CXCR4	0,646	0,888	0,979
基因 5	CCR3	0,705	0,629	0,887
基因 6	CXCL9	0,605	0,759	0,949
基因 7	CXCL10	0,568	0,740	0,979
基因 8	CXCL11	0,574	0,759	0,949
基因 9	SPAG9	0,516	0,814	0,959
基因 10	CK19	0,841	0,204	0,846
基因 11	TIMP1	0,865	0,092	0,785
基因 12	CLDN1	0,851	0,203	0,785
基因 13	CAR	0,672	0,796	0,959
基因 14	NEC-1	0,549	0,963	1,000
基因 15	XB130	0,505	0,814	0,928
基因 16	HO-1	0,646	0,777	0,989
基因 17	CCR7	0,747	0,500	0,877
基因 18	CXCL4	0,543	0,870	0,938

图2

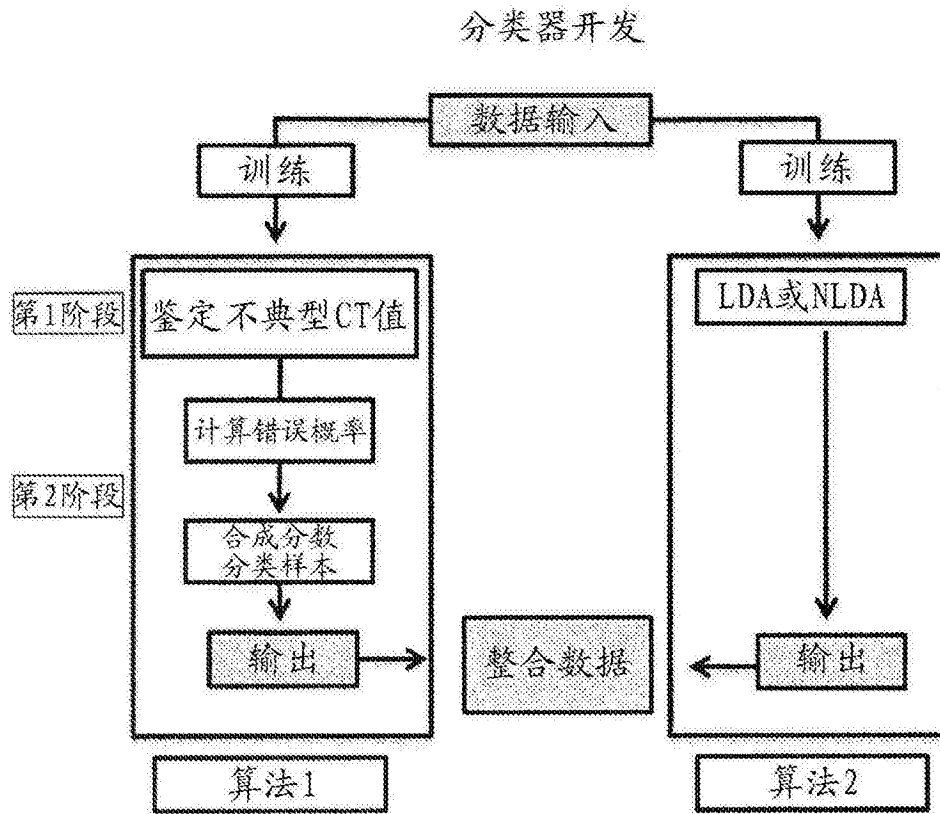


图3A

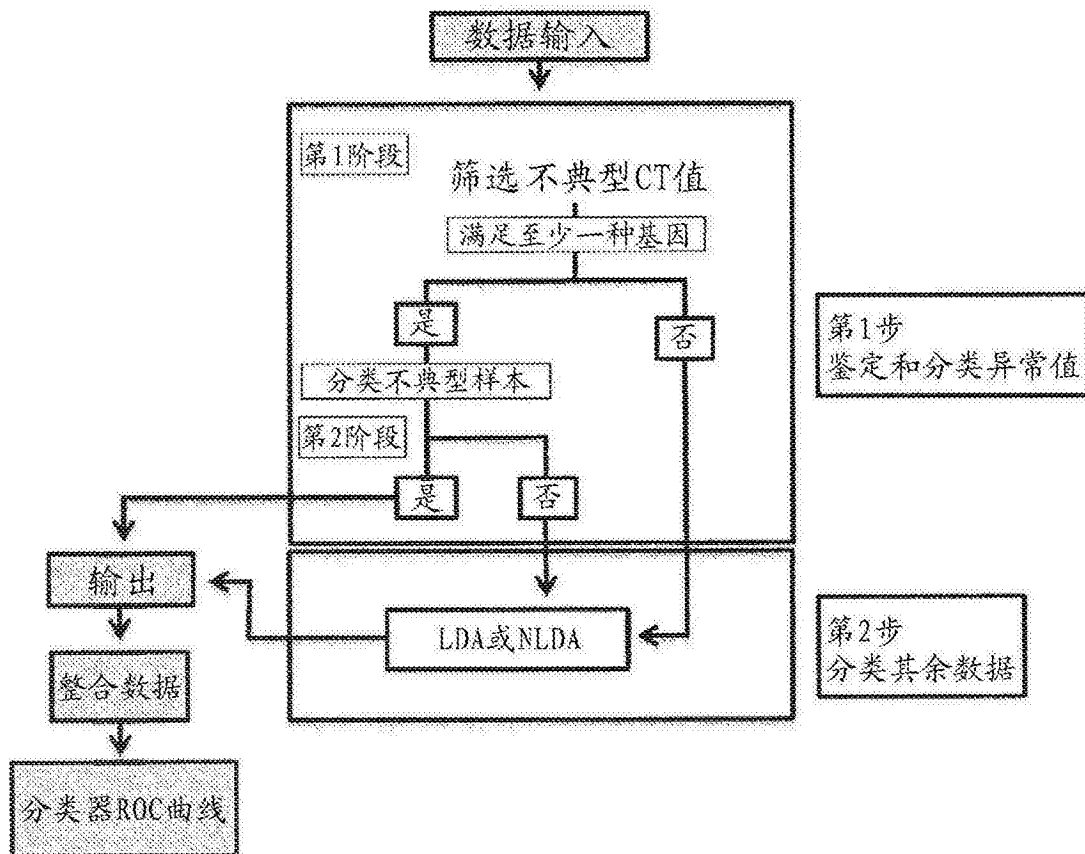


图3B

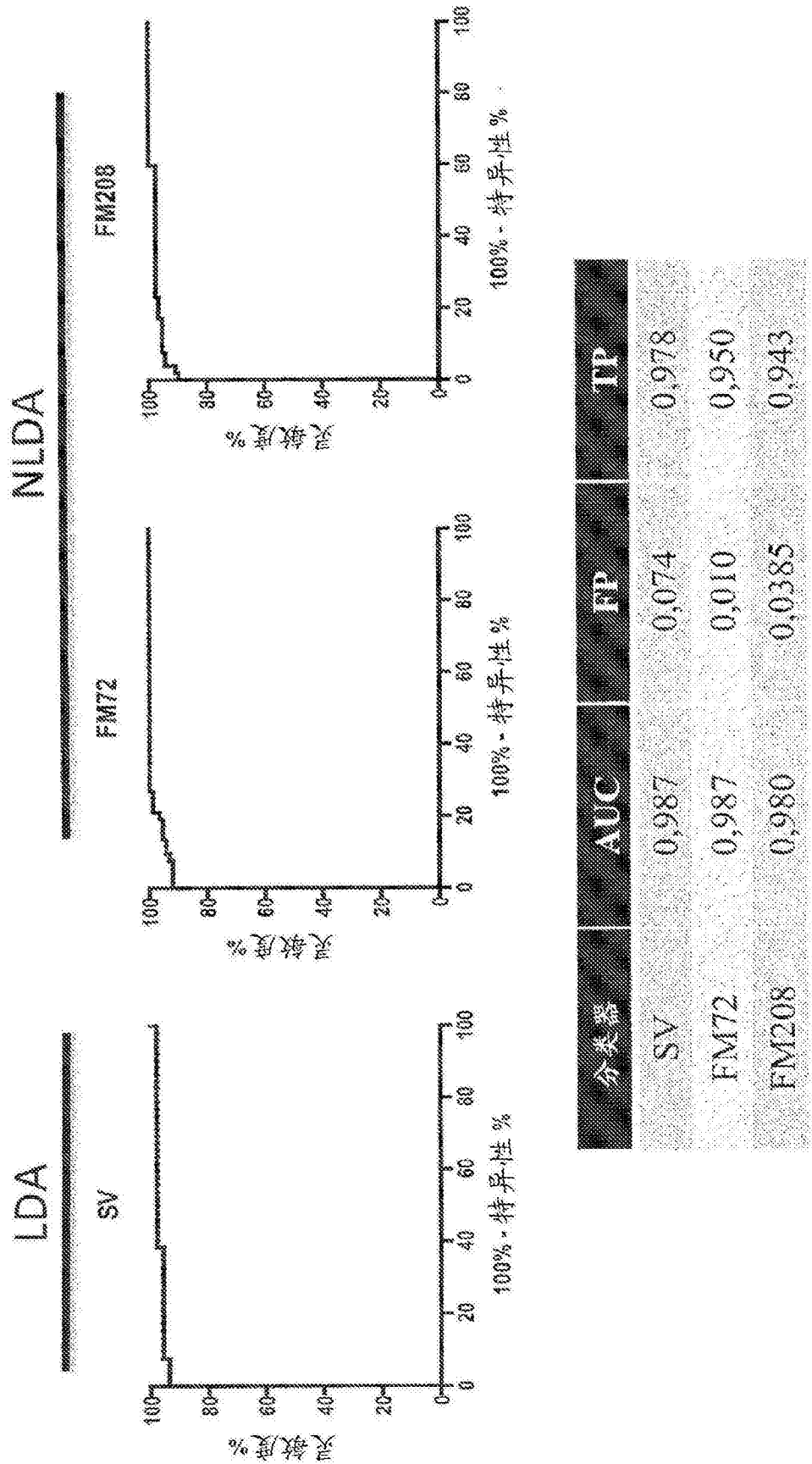
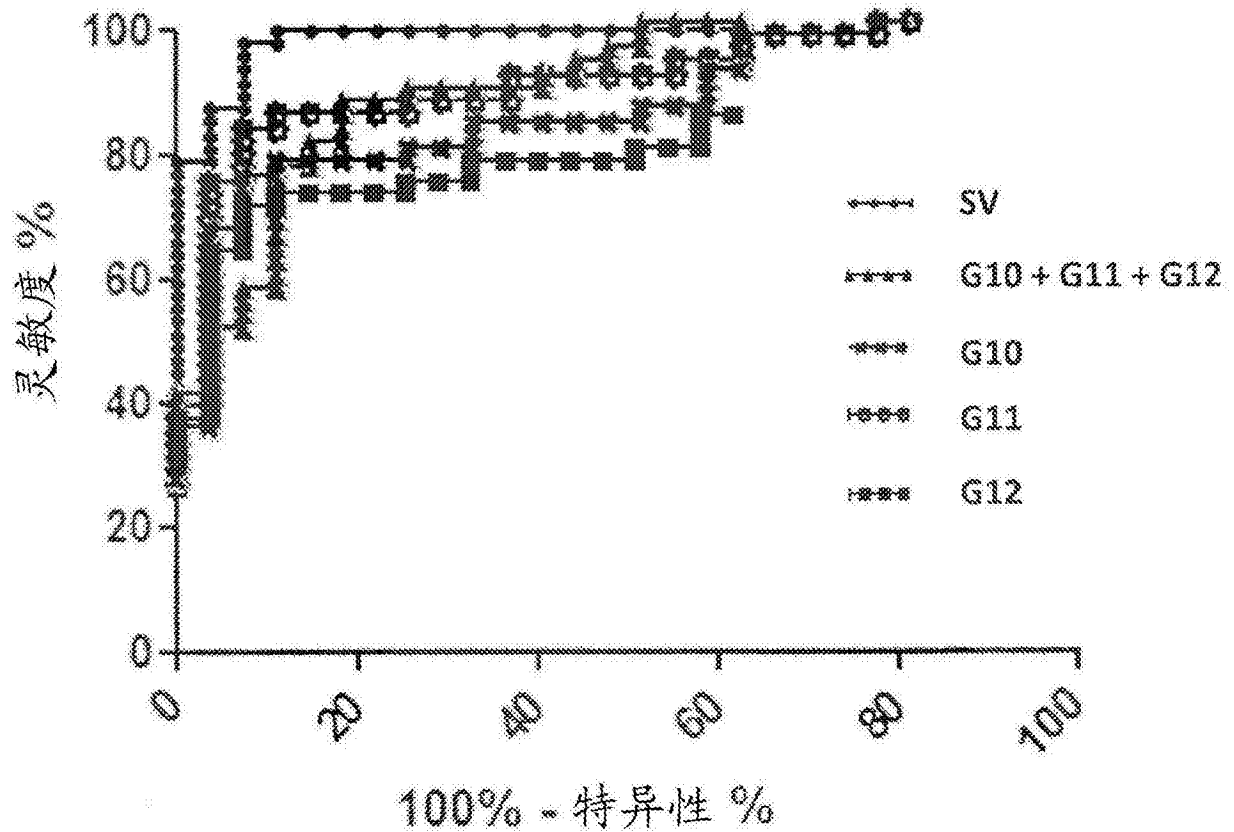


图4



	SV	基因 11	基因 (10,11,12)	基因 10	基因 12
SV	-	0,07565	0,08511	0,09062	0,10166
基因 11	0,02056*	-	0,00946	0,01497	0,026
基因 (10,11,12)	0,01623*	0,15067	-	0,00552	0,01655
基因 10	0,00871*	0,63883	0,87176	-	0,01103
基因 12	0,00534*	0,38495	0,605575	0,70021	-

显著水平 (α): 0.05

上三角: AUC 差异

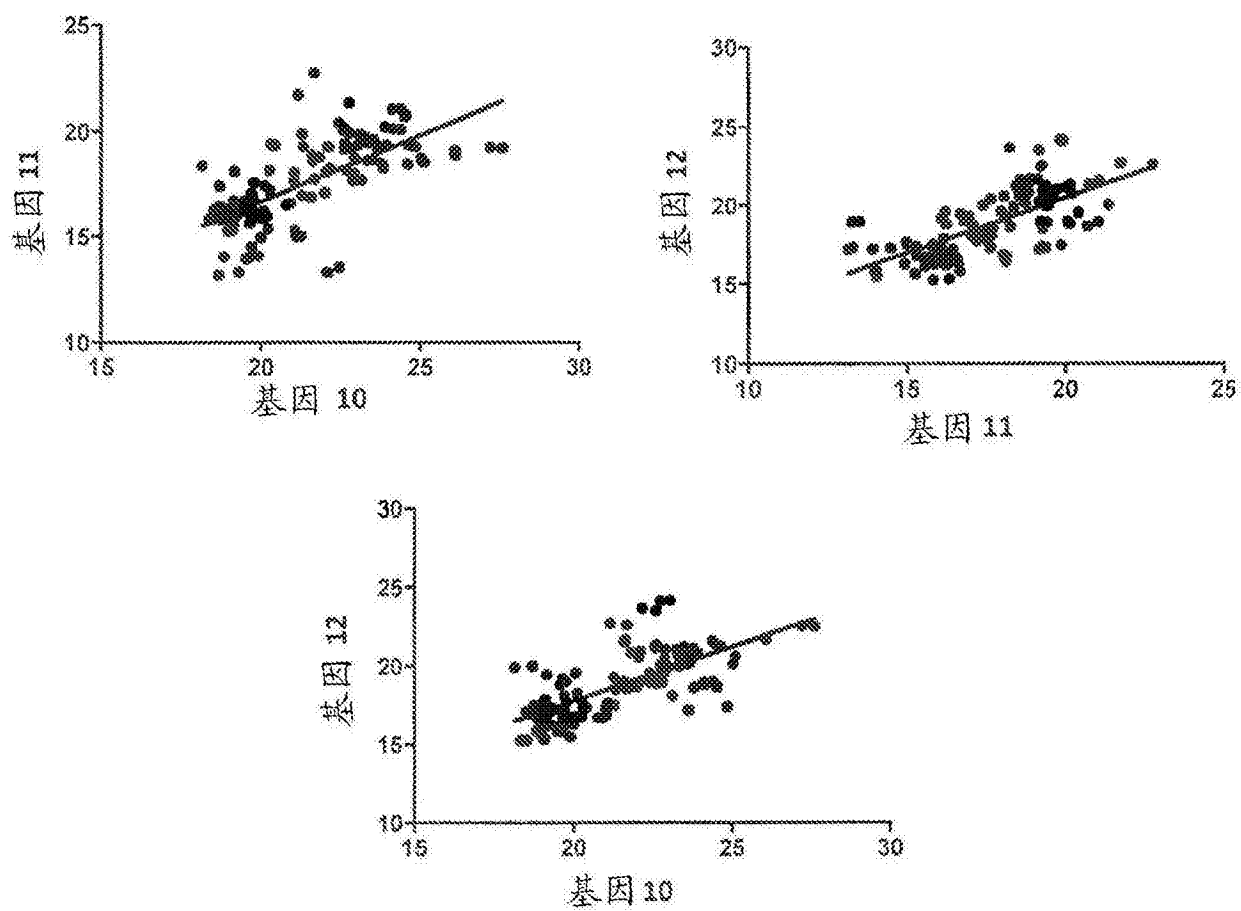
下三角: p-值

*p-值 < 0.05

总p-值: N. D. (矩阵LSL T为单数)

图5

Spearman相关分析



Spearman系数

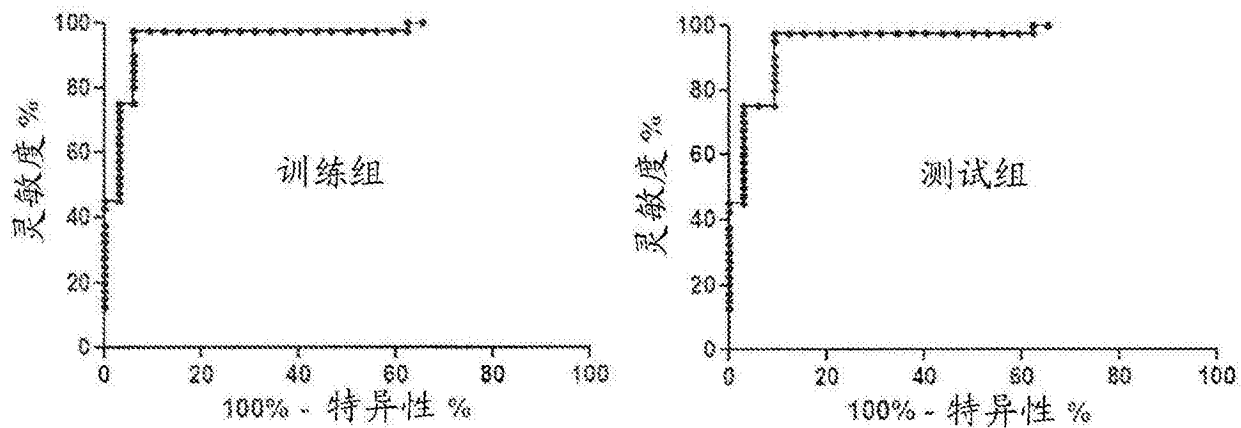
基因 10/11: 0.79 $p < 0.05$ 基因 11/12: 0.78 $p < 0.05$ 基因 10/12: 0.79 $p < 0.05$

图6

参数	SV	FM72	FM208	基因 10	基因 11	基因 12	基因 (10,11,12)	Afima
AUC	0.987	0.987	0.980	0.840	0.866	0.851	0.890	0.900
95% CI	0.967 - 1.000	0.911 - 0.999	0.948 - 1.000	0.809 - 0.929	0.824 - 0.946	0.809 - 0.921	0.85 - 0.95	
灵敏度 (%)	97.87	95.05	94.32	84.69	78.57	78.57	87.23	92.00
95% CI	89.0 - 100	88.71 - 99.95	85.46 - 99.48	77.51 - 92.43	78.76 - 93.23	71.44 - 88.24	78.76 - 93.23	
特异性 (%)	92.60	99.00	96.15	75.93	91.74	79.63	72.72	58.00
95% CI	75.7 - 99.0	61.92 - 93.70	66.27 - 95.81	64.40 - 87.96	58.36 - 83.54	66.47 - 89.37	58.35 - 83.54	
阳性预测值	95.84	99.40	97.71	83.79	93.32	85.00	84.53	68.66
阴性预测值	96.15	91.99	90.68	77.15	74.46	71.67	76.47	87.88
测试前(癌症发病率)	63.51	63.51	63.51	59.49	59.49	59.49	63.51	50.00
似然比 +	13.23	95.05	24.50	3.52	9.51	3.86	3.14	2.19
测试后 +	95.84	99.40	97.71	83.79	93.32	85.00	84.53	68.66
测试前(癌症发病率)	63.51	63.51	63.51	59.49	59.49	59.49	59.49	50.00
似然比 -	0.02	0.05	0.06	0.20	0.23	0.27	0.18	0.14
测试后 -	3.85	8.01	9.32	22.85	25.54	28.33	20.62	12.12

图7

SV算法



	训练组	测试组
AUC	0,98	0,95
灵敏度	97,8 %	95,0 %
特异性	92,6 %	90,6 %
阳性预测值	96,1 %	83,5 %
阴性预测值	95,8 %	98,0 %
癌症发病率	63,5 %	33,0 %

图8

SV算法

样本	手术	手术	FNA
	训练组	测试组1	测试组2
灵敏度	98%	95%	96%
特异性	93%	90%	89%
PPV	95 %	83 %	75%
NPV	96 %	98 %	98%
癌症发病率	63 %	33%	26%

图9