



(12) 发明专利

(10) 授权公告号 CN 108957006 B

(45) 授权公告日 2021.07.27

(21) 申请号 201810800026.2

(22) 申请日 2014.03.10

(65) 同一申请的已公布的文献号
申请公布号 CN 108957006 A

(43) 申请公布日 2018.12.07

(30) 优先权数据
61/787,967 2013.03.15 US

(62) 分案原申请数据
201480022424.6 2014.03.10

(73) 专利权人 私募蛋白质体公司
地址 美国科罗拉多州

(72) 发明人 马尔蒂·尼克瑞德
斯图尔特·G·菲尔德
斯蒂芬·阿拉里克·威廉斯

(74) 专利代理机构 中科专利商标代理有限责任
公司 11021
代理人 吴小明

(51) Int.Cl.
G01N 33/68 (2006.01)

(56) 对比文件

CN 102640001 A,2012.08.15

US 2008311593 A1,2008.12.18

WO 2010000835 A1,2010.01.07

US 2007111933 A1,2007.05.17

Jang Yel Shin等.Serum sex hormone-binding globulin levels are independently associated with nonalcoholic fatty liver disease in people with type 2 diabetes.《Diabetes Research and Clinical Practice》.2011,第94卷第156-162页.

G-X. Tian等.Oestradiol is a protective factor for non-alcoholic fatty liver disease in healthy men.《obesity》.2012,第13卷第381-387页.

Lauren N. Bell等.Serum Proteomics and Biomarker Discovery Across the Spectrum of Nonalcoholic Fatty Liver Disease.《HEPATOLOGY》.2010,第51卷(第1期),

审查员 郭雅文

权利要求书2页 说明书46页 附图23页

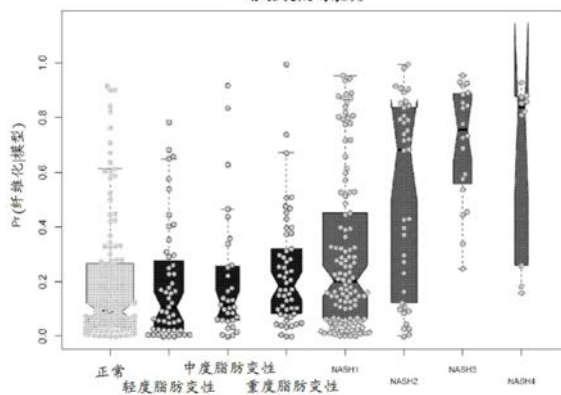
(54) 发明名称

非酒精性脂肪肝疾病 (NAFLD) 和非酒精性脂肪性肝炎 (NASH) 生物标记及其用途

(57) 摘要

提供用于确定受试者是否患有非酒精性脂肪肝疾病 (NAFLD) 的方法、组合物和试剂盒。也提供用于确定受试者是否患有非酒精性脂肪变性的方法、组合物和试剂盒。也提供用于确定受试者是否患有非酒精性脂肪性肝炎 (NASH) 的方法、组合物和试剂盒。

4种标记的随机森林分类器:
根据类别的表决



1. 一组生物标记捕获剂在制备用于确定受试者是否患有非酒精性脂肪性肝炎 (NASH) 的试剂盒中的用途, 其中所述生物标记捕获剂用于检测来自所述受试者的样品中选自自由 ACY1, THBS2, LGALS3BP, KYNU, COLEC11, CTSZ, IL19, POR, INS, SHBG, GPT, GNS, RET, AFM, SELE, CD163, ENPP7, IGFBP7, GSN, SIGLEC7, LAMA1, LAMB1, LAMC1, IL18R1, FN1, TGFBI, 和 AKR1A1 组成的组的 N 种生物标记蛋白质各自的水平, 其中所述 N 种生物标记蛋白质中的至少一种是 COLEC11, 并且其中所述样品是血液样品、血清样品或血浆样品。

2. 如权利要求 1 所述的用途, 其中 N 是 1 至 10, 或 N 是 2 至 10, 或 N 是 3 至 10, 或 N 是 4 至 10, 或 N 是 5 至 10, 或 N 是 1 至 5, 或 N 是 2 至 5, 或 N 是 3 至 5。

3. 如权利要求 1 或权利要求 2 所述的用途, 其中所述组的生物标记捕获剂用于检测 COLEC11 的水平以及 C7、PPID、IGFBP3、SIGLEC14、AIMP1、TOP1、CA6、STX1A、GPC5、HIPK3、IGFBP7 和 IL3RA 中的一种或多种的水平。

4. 一组生物标记捕获剂在制备用于确定患有非酒精性脂肪变性的受试者是否患有非酒精性脂肪性肝炎的试剂盒中的用途, 其中所述生物标记捕获剂用于检测来自所述受试者的样品中选自自由 C7, COLEC11, PPID, IGFBP3, GPC5, HIPK3, IGFBP7, 和 IL3RA 组成的组的 N 种生物标记蛋白质各自的水平, 其中所述 N 种生物标记蛋白质中的至少一种是 COLEC11。

5. 如权利要求 4 所述的用途, 其中 N 是 1 至 8, 或 N 是 2 至 8, 或 N 是 3 至 8, 或 N 是 4 至 8, 或 N 是 5 至 8, 或 N 是 1 至 5, 或 N 是 2 至 5, 或 N 是 3 至 5。

6. 用于检测来自受试者的样品中生物标记蛋白质的水平的一组生物标记捕获剂在制备用于确定患有非酒精性脂肪变性的受试者是否患有非酒精性脂肪性肝炎或监测患有非酒精性脂肪变性的受试者的 NASH 发展的试剂盒中的用途, 其中所述组的生物标记捕获剂用于检测来自所述受试者的样品中 COLEC11 的水平以及任选 C7、PPID、IGFBP3、SIGLEC14、AIMP1、TOP1、CA6 和 STX1A 中的一种或多种的水平, 其中 C7、COLEC11 或 SIGLEC14 的水平高于相应生物标记的对照水平, 和/或 IGFBP3、AIMP1、TOP1、CA6、PPID 或 STX1A 的水平低于相应生物标记的对照水平指示所述受试者患有 NASH。

7. 如权利要求 4 至 6 中任一项所述的用途, 其中所述受试者患有轻度、中度或重度脂肪变性。

8. 如权利要求 1 至 2 和 4 至 6 中任一项所述的用途, 其中所述 NASH 是 1、2、3 或 4 期 NASH。

9. 如权利要求 4 至 6 中任一项所述的用途, 其中所述组的生物标记捕获剂用于检测 COLEC11 以及至少一种、至少两种、或三种选自 C7、PPID 和 IGFBP3 的生物标记的水平; 或检测 COLEC11 以及至少一种、至少两种、至少三种、至少四种、或五种选自 SIGLEC14、AIMP1、TOP1、CA6 和 STX1A 的生物标记的水平。

10. 如权利要求 1 至 2 和 4 至 6 中任一项所述的用途, 其中所述受试者处于发展 NAFLD、脂肪变性或 NASH 的风险下。

11. 如权利要求 10 所述的用途, 其中所述受试者患有选自肥胖症、腹部肥胖、代谢综合征、心血管疾病和糖尿病的 NAFLD 共患病。

12. 如权利要求 1 至 2 和 4 至 6 中任一项所述的用途, 其中所述组的生物标记捕获剂用于与来自所述受试者的所述样品的生物标记接触, 其中所述组的生物标记捕获剂中的各生物标记捕获剂特异性结合所检测的不同生物标记。

13. 如权利要求 12 所述的用途, 其中各生物标记捕获剂是抗体或适体。

14. 如权利要求13所述的用途,其中各生物标记捕获剂是适体,并且其中至少一种适体是缓慢解离速率适体。

15. 如权利要求14所述的用途,其中至少一种缓慢解离速率适体包含至少一个、至少两个、至少三个、至少四个、至少五个、至少六个、至少七个、至少八个、至少九个、或至少十个具有修饰的核苷酸。

非酒精性脂肪肝疾病 (NAFLD) 和非酒精性脂肪性肝炎 (NASH) 生物标记及其用途

[0001] 本申请是中国发明专利申请号201480022424.6的分案申请。对应的初始申请的发明名称为“非酒精性脂肪肝疾病 (NAFLD) 和非酒精性脂肪性肝炎 (NASH) 生物标记及其用途”，申请日为2014年3月10日，优先权日期为2013年3月15日。

发明领域

[0002] 本申请大体上涉及检测生物标记以及表征非酒精性脂肪肝疾病 (NAFLD)，例如以鉴定患有脂肪变性和非酒精性脂肪性肝炎 (NASH) 的受试者。在各种实施方案中，本发明涉及一种或多种用于表征个体的NAFLD和NASH的生物标记、方法、装置、试剂、系统和试剂盒。

[0003] 背景

[0004] 以下描述提供信息概述，并且并非认可任何在本文中提供的信息或参照的出版物是本申请的先前技术。

[0005] 非酒精性脂肪肝疾病 (NAFLD) 定义为在不存在酒精史下，存在伴有或不伴有炎症和纤维化的肝性脂肪变性。NAFLD被再分成非酒精性脂肪肝 (NAFL) 和非酒精性脂肪性肝炎 (NASH)。在NAFL 中，肝性脂肪变性在无显著炎症的迹象下存在，而在NASH中，肝性脂肪变性伴有可在组织学上不可与酒精性脂肪性肝炎区分的肝性炎症。

[0006] NAFLD已变为一种全世界流行病，并且由于肥胖症的快速递增发病而成为北美主要肝病病因。然而，关于NAFL和NASH的发病率的基于人口的准确数据是不足的，部分地因为诊断需要组织病理学文件编制。NAFLD的主要风险因素是向心性肥胖症、2型糖尿病、高血液甘油三酯 (脂肪) 水平和高血压。在美国，NAFLD存在于 20-40% 的人口中，并且NASH存在于约25%的肥胖人口中。10至29%的NASH患者发展肝硬化，并且那些患者中的4-27%发展肝癌。

[0007] 大多数患有NASH的人员不具有症状。一些人员可具有右上象限疼痛、肝肿大或非特异性症状，如腹部不舒适、虚弱、疲劳或不适。医生或护士可根据常规血液测试的结果怀疑存在NASH。在NAFLD 中，肝酶天冬氨酸转氨酶 (AST) 和丙氨酸转氨酶 (ALT) 常常较高。

[0008] 当前用以确认NASH的金标准是组织学评估肝活检体，此举是昂贵的，侵袭性的，并且可导致疼痛、出血或甚至死亡。

[0009] 将鉴定和区分各种时期的NAFLD与NASH两者 (并且由此减少对肝活检的需要) 的简单血液测试将为高度合乎需要的。

[0010] 概述

[0011] 在一些实施方案中，提供确定受试者是否患有非酒精性脂肪肝疾病 (NAFLD) 的方法。在一些实施方案中，提供鉴定患有脂肪变性的受试者的方法。在一些实施方案中，提供确定脂肪变性的严重性的方法。

[0012] 在一些实施方案中，提供确定受试者是否患有非酒精性脂肪性肝炎 (NASH) 的方法。在一些实施方案中，提供鉴定患有NASH的受试者的方法。在一些实施方案中，提供区分患有NASH的受试者与患有脂肪变性的受试者的方法。在一些实施方案中，提供确定NASH 的

严重性的方法。

[0013] 在一些实施方案中,提供确定受试者是否患有非酒精性脂肪肝疾病 (NAFLD) 的方法。在一些实施方案中,方法包括检测来自受试者的样品中至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、至少八种、至少九种、至少十种、或十一种选自 ACY、SHBG、CTS Z、MET、GSN、LGALS3BP、PLAT、CHL1、SERPINC1、SIGLEC7和SIGLEC14的生物标记的水平。在一些实施方案中,方法包括检测至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、或八种选自ACY、SHBG、CTS Z、MET、GSN、LGALS3BP、CHL1、SERPINC1的生物标记的水平。在一些实施方案中,方法包括检测来自受试者的样品中ACY以及任选SHBG、CTS Z、MET、GSN、LGALS3BP、PLAT、CHL1、SERPINC1、SIGLEC7和SIGLEC14中的一个或多个的水平。在一些实施方案中,方法包括检测来自受试者的样品中SHBG以及任选ACY、CTS Z、MET、GSN、LGALS3BP、PLAT、CHL1、SERPINC1、SIGLEC7和SIGLEC14中的一个或多个的水平。在一些实施方案中,方法包括检测来自受试者的样品中ACY和SHBG以及任选CTS Z、MET、GSN、LGALS3BP、PLAT、CHL1、SERPINC1、SIGLEC7和SIGLEC14中的一个或多个的水平。在一些实施方案中,至少一种选自ACY、CTS Z、LGALS3BP、SIGLEC7、SIGLEC14和PLAT的生物标记的水平高于相应生物标记的对照水平指示受试者患有NAFLD。在一些实施方案中,至少一种选自SHBG、MET、GSN、CHL1和SERPINC1 的生物标记的水平低于相应生物标记的对照水平指示受试者患有 NAFLD。在一些实施方案中,方法包括确定受试者是否患有脂肪变性。在一些实施方案中,脂肪变性是轻度、中度或重度脂肪变性。在一些实施方案中,方法包括检测至少一种、至少两种、或三种选自 ACY、SHBG和SIGLEC14的生物标记。在一些实施方案中,方法包括确定受试者是否患有非酒精性脂肪性肝炎 (NASH)。在一些实施方案中,NASH是1、2、3或4期NASH。在一些实施方案中,方法包括检测至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、或七种选自ACY、SHBG、CTS Z、MET、GSN、LGALS3BP 和SIGLEC7的生物标记。在一些实施方案中,方法包括检测至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、至少八种、或九种选自ACY、SHBG、CTS Z、MET、GSN、LGALS3BP、PLAT、CHL1和SERPINC1的生物标记。

[0014] 在一些实施方案中,提供确定患有非酒精性脂肪变性的受试者是否患有非酒精性脂肪性肝炎 (NASH) 的方法。在一些实施方案中,方法包括检测来自受试者的样品中至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、至少八种、至少九种、或至少十种选自C7、PPID、IGFBP3、SIGLEC14、AIMP1、TOP1、COLEC11、CA6、STX1A、GPC5、HIPK3、IGFBP7和IL3RA的生物标记的水平。在一些实施方案中,方法包括检测来自受试者的样品中至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、至少八种、或九种选自C7、PPID、IGFBP3、SIGLEC14、AIMP1、TOP1、COLEC11、CA6和STX1A的生物标记的水平。在一些实施方案中,方法包括检测来自受试者的样品中至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、或八种选自C7、PPID、IGFBP3、COLEC11、GPC5、HIPK3、IGFBP7和IL3RA的生物标记的水平。在一些实施方案中,方法包括检测来自受试者的样品中COLEC11以及任选C7、PPID、IGFBP3、SIGLEC14、AIMP1、TOP1、CA6、STX1A、GPC5、HIPK3、IGFBP7和IL3RA 中的一个或多个的水平。在一些实施方案中,方法包括检测来自受试者的样品中C7以及任选COLEC11、PPID、IGFBP3、SIGLEC14、AIMP1、TOP1、CA6、STX1A、GPC5、HIPK3、IGFBP7和IL3RA 中的一个或多个的水平。在一些实施方案中,方法包括检测来

自受试者的样品中PPID以及任选C7、COLEC11、IGFBP3、SIGLEC14、AIMP1、TOP1、CA6、STX1A、GPC5、HIPK3、IGFBP7和IL3RA 中的一个或多个的水平。在一些实施方案中,方法包括检测来自受试者的样品中IGFBP3以及任选COLEC11、PPID、C7、SIGLEC14、AIMP1、TOP1、CA6、STX1A、GPC5、HIPK3、IGFBP7和IL3RA 中的一个或多个的水平。在一些实施方案中,方法包括检测来自受试者的样品中C7、COLEC11、PPID和IGFBP3以及任选SIGLEC14、AIMP1、TOP1、CA6、STX1A、GPC5、HIPK3、IGFBP7和IL3RA 中的一个或多个的水平。在一些实施方案中,至少一种选自C7、COLEC11、HIPK3、IGFBP7、IL3RA和SIGLEC14的生物标记的水平高于相应生物标记的对照水平指示受试者患有NASH。在一些实施方案中,至少一种选自IGFBP3、AIMP1、TOP1、CA6、PPID、GPC3 和STX1A的生物标记的水平低于相应生物标记的对照水平指示受试者患有NASH。在一些实施方案中,受试者患有轻度、中度或重度脂肪变性。在一些实施方案中,NASH是1、2、3或4期NASH。在一些实施方案中,NASH是2、3或4期NASH。在一些实施方案中,方法包括检测至少一种、至少两种、至少三种、或四种选自C7、COLEC11、PPID和IGFBP3的生物标记的水平。在一些实施方案中,方法包括检测至少一种、至少两种、至少三种、至少四种、至少五种、或六种选自SIGLEC14、AIMP1、TOP1、COLEC11、CA6和STX1A 的生物标记的水平。

[0015] 在一些实施方案中,提供监测患有非酒精性脂肪变性的受试者的 NASH发展的方法。在一些实施方案中,方法包括检测来自受试者的样品中至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、至少八种、至少九种、或至少十种选自C7、PPID、IGFBP3、SIGLEC14、AIMP1、TOP1、COLEC11、CA6、STX1A、GPC5、HIPK3、IGFBP7和IL3RA的生物标记的水平。在一些实施方案中,方法包括检测来自受试者的样品中至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、至少八种、或九种选自C7、PPID、IGFBP3、SIGLEC14、AIMP1、TOP1、COLEC11、CA6和STX1A的生物标记的水平。在一些实施方案中,方法包括检测来自受试者的样品中至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、或八种选自C7、PPID、IGFBP3、COLEC11、GPC5、HIPK3、IGFBP7和IL3RA的生物标记的水平。在一些实施方案中,至少一种选自C7、COLEC11和SIGLEC14的生物标记的水平高于相应生物标记的对照水平指示受试者患有NASH。在一些实施方案中,至少一种选自IGFBP3、AIMP1、TOP1、CA6、PPID和STX1A的生物标记的水平低于相应生物标记的对照水平指示受试者患有NASH。在一些实施方案中,受试者患有轻度、中度或重度脂肪变性。在一些实施方案中,NASH是1、2、3或4期NASH。在一些实施方案中,NASH是2、3或4期NASH。在一些实施方案中,方法包括检测至少一种、至少两种、至少三种、或四种选自C7、COLEC11、PPID和IGFBP3的生物标记的水平。在一些实施方案中,方法包括检测至少一种、至少两种、至少三种、至少四种、至少五种、或六种选自SIGLEC14、AIMP1、TOP1、COLEC11、CA6和STX1A 的生物标记的水平。

[0016] 在一些实施方案中,提供确定受试者是否患有NAFLD的方法,其中所述方法包括检测至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、至少八种、至少九种、或至少十种选自表7中的生物标记的生物标记的水平。在一些实施方案中,方法包括形成具有N种来自表7中所列的生物标记蛋白质的生物标记蛋白质的生物标记组,以及检测来自受试者的样品中所述组的所述N 种生物标记蛋白质各自的水平。在一些实施方案中,N是1至25。在一些实施方案中,N是2至25。在一些实施方案中,N是3至25。在一些实施方案中,N是4至25。在一些实施方案中,N是5至25。在一些实施方案中,N是1至10。在一些实施

方案中,N是2至10。在一些实施方案中,N是3至10。在一些实施方案中,N是4至10。在一些实施方案中,N是5至10。在一些实施方案中,N种生物标记蛋白质中的至少一种选自ACY和SHBG。在一些实施方案中,N种生物标记蛋白质中的两个是ACY和SHBG。在一些实施方案中,表7中除SHBG和GSN以外的生物标记的水平高于相应生物标记的对照水平指示受试者患有NAFLD。在一些实施方案中,至少一种选自SHBG和GSN的生物标记的水平低于相应生物标记的对照水平指示受试者患有NAFLD。在一些实施方案中,方法包括确定受试者是否患有NASH。

[0017] 在一些实施方案中,提供确定受试者是否患有NASH的方法,其中所述方法包括检测至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、或八种选自表9中的生物标记的生物标记的水平。在一些实施方案中,方法包括形成具有N种来自表9中所列的生物标记蛋白质的生物标记蛋白质的生物标记组,以及检测来自受试者的样品中所述组的所述N种生物标记蛋白质各自的水平。在一些实施方案中,N是1至8。在一些实施方案中,N是2至8。在一些实施方案中,N是3至8。在一些实施方案中,N是4至8。在一些实施方案中,N是5至8。在一些实施方案中,N种生物标记蛋白质中的至少一种是COLEC11。在一些实施方案中,N种生物标记蛋白质中的至少一种选自C7、COLEC11、PPID和IGFBP3。在一些实施方案中,表9中除IGFBP3和GPC3以外的生物标记的水平高于相应生物标记的对照水平指示受试者患有NASH。在一些实施方案中,至少一种选自IGFBP3和GPC3的生物标记的水平低于相应生物标记的对照水平指示受试者患有NASH。

[0018] 在任何本文所述的实施方案中,受试者可处于发展NAFLD的风险下。在任何本文所述的实施方案中,受试者可处于发展脂肪变性的风险下。在任何本文所述的实施方案中,受试者可处于发展NASH 的风险下。在任何本文所述的实施方案中,受试者可患有选自肥胖症、腹部肥胖、代谢综合征、心血管疾病和糖尿病的NAFLD共患病。在任何本文所述的实施方案中,受试者可为肥胖的。

[0019] 在任何本文所述的实施方案中,至少一种生物标记可为蛋白质生物标记。在任何本文所述的实施方案中,各生物标记可为蛋白质生物标记。在一些实施方案中,方法包括使来自受试者的样品的生物标记与一组生物标记捕获试剂接触,其中所述组的生物标记捕获试剂中的各生物标记捕获试剂特异性结合所检测的不同生物标记。在一些实施方案中,各生物标记捕获试剂是抗体或适体。在一些实施方案中,各生物标记捕获试剂是适体。在一些实施方案中,至少一种适体是缓慢解离速率适体。在一些实施方案中,至少一种缓慢解离速率适体包含至少一个、至少两个、至少三个、至少四个、至少五个、至少六个、至少七个、至少八个、至少九个、或至少10个具有修饰的核苷酸。在一些实施方案中,各缓慢解离速率适体以 ≥ 30 分钟、 ≥ 60 分钟、 ≥ 90 分钟、 ≥ 120 分钟、 ≥ 150 分钟、 ≥ 180 分钟、 ≥ 210 分钟或 ≥ 240 分钟的解离速率($t_{1/2}$)结合它的靶标蛋白质。

[0020] 在任何本文所述的实施方案中,样品可为血液样品。在任何本文所述的实施方案中,样品可选自血清样品和血浆样品。

[0021] 在任何本文所述的实施方案中,如果受试者患有NAFLD或 NASH,那么受试者可被推荐选自减轻重量、控制血糖和避免酒精的方案。在任何本文所述的实施方案中,如果受试者患有NAFLD或 NASH,那么受试者可被推荐胃绕道手术。在任何本文所述的实施方案中,如果受试者患有NAFLD或NASH,那么受试者可被指定至少一种选自匹格列酮(pioglitazone)、

维生素E和二甲双胍(metformin) 的治疗剂。

[0022] 在一些实施方案中,本文所述的方法用于确定医疗保险费或人寿保险费的目的。在一些实施方案中,方法进一步包括确定医疗保险费或人寿保险费。在一些实施方案中,本文所述的方法进一步包括使用由方法产生的信息来预测和/或管理对医疗资源的利用。

[0023] 在一些实施方案中,提供试剂盒。在一些实施方案中,试剂盒包括至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、至少八种、至少九种、至少十种、或十一种特异性结合选自ACY、SHBG、CTSΖ、MET、GSN、LGALS3BP、PLAT、CHL1、SERPINC1、SIGLEC7和SIGLEC14的靶标蛋白质的适体。在一些实施方案中,试剂盒包括至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、或八种特异性结合选自ACY、SHBG、CTSΖ、MET、GSN、LGALS3BP、CHL1、SERPINC1的靶标蛋白质的适体。在一些实施方案中,试剂盒包括特异性结合ACY的适体以及任选一种或多种特异性结合一种或多种选自SHBG、CTSΖ、MET、GSN、LGALS3BP、PLAT、CHL1、SERPINC1、SIGLEC7和SIGLEC14 的靶标蛋白质的适体。在一些实施方案中,试剂盒包括特异性结合 SHBG的适体以及任选一种或多种特异性结合一种或多种选自ACY、CTSΖ、MET、GSN、LGALS3BP、PLAT、CHL1、SERPINC1、SIGLEC7 和SIGLEC14的靶标蛋白质的适体。在一些实施方案中,试剂盒包括特异性结合ACY的适体和特异性结合SHBG的适体以及任选一种或多种特异性结合一种或多种选自CTSΖ、MET、GSN、LGALS3BP、 PLAT、CHL1、SERPINC1、SIGLEC7和SIGLEC14的靶标蛋白质的适体。

[0024] 在一些实施方案中,试剂盒包括至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、至少八种、至少九种、或至少十种特异性结合选自表7中的蛋白质的靶标蛋白质的适体。在一些实施方案中,试剂盒包括N种适体,其中各适体特异性结合表7中所列的生物标记蛋白质。在一些实施方案中,N是1至25。在一些实施方案中,N是2至25。在一些实施方案中,N是3至25。在一些实施方案中,N是4至25。在一些实施方案中,N是5至25。在一些实施方案中,N是1至10。在一些实施方案中,N是2至10。在一些实施方案中,N是3至10。在一些实施方案中,N是4至10。在一些实施方案中,N是5至10。在一些实施方案中,N种适体中的至少一种特异性结合选自ACY和SHBG的生物标记蛋白质。在一些实施方案中,N种适体中的一种特异性结合ACY,并且N种适体中的一种特异性结合SHBG。在一些实施方案中,各适体特异性结合不同靶标蛋白质。

[0025] 在一些实施方案中,试剂盒包括至少一种、至少两种、或三种特异性结合选自ACY、SHBG和SIGLEC14的靶标蛋白质的适体。在一些实施方案中,试剂盒包括至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、或七种特异性结合选自ACY、SHBG、CTSΖ、MET、GSN、LGALS3BP和SIGLEC7的靶标蛋白质的适体。在一些实施方案中,试剂盒包括至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、至少八种、或九种特异性结合选自ACY、SHBG、CTSΖ、MET、GSN、LGALS3BP、PLAT、CHL1和SERPINC1的靶标蛋白质的适体。

[0026] 在一些实施方案中,提供一种试剂盒,其中所述试剂盒包括至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、至少八种、至少九种、或至少十种特异性结合选自C7、PPID、IGFBP3、SIGLEC14、AIMP1、TOP1、COLEC11、CA6、STX1A、GPC5、HIPK3、IGFBP7和IL3RA的靶标蛋白质的适体。在一些实施方案中,提供一种试剂盒,其中所述试剂盒包括至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、至少八

种、或九种特异性结合选自C7、PPID、IGFBP3、SIGLEC14、AIMP1、TOP1、COLEC11、CA6和STX1A的靶标蛋白质的适体。在一些实施方案中,提供一种试剂盒,其中所述试剂盒包括至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、或八种特异性结合选自C7、PPID、IGFBP3、COLEC11、GPC5、HIPK3、IGFBP7和IL3RA的靶标蛋白质的适体。在一些实施方案中,各适体结合不同靶标蛋白质。

[0027] 在一些实施方案中,试剂盒包括至少一种、至少两种、至少三种、或四种特异性结合选自C7、COLEC11、PPID和IGFBP3的靶标蛋白质的适体。在一些实施方案中,试剂盒包括至少一种、至少两种、至少三种、至少四种、至少五种、或六种特异性结合选自SIGLEC14、AIMP1、TOP1、COLEC11、CA6和STX1A的靶标蛋白质的适体。

[0028] 在一些实施方案中,试剂盒包括至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、或八种特异性结合选自表9中的蛋白质的靶标蛋白质的适体。在一些实施方案中,试剂盒包括N种适体,其中各适体特异性结合表9中所列的生物标记蛋白质。在一些实施方案中,N是1至8。在一些实施方案中,N是2至8。在一些实施方案中,N是3至8。在一些实施方案中,N是4至8。在一些实施方案中,N是5至8。在一些实施方案中,N种适体中的至少一种特异性结合COLEC11。在一些实施方案中,N种适体中的至少一种特异性结合选自C7、COLEC11、PPID和IGFBP3的生物标记蛋白质。在一些实施方案中,N种适体中的四种各自特异性结合选自C7、COLEC11、PPID和IGFBP3的不同生物标记蛋白质。在一些实施方案中,各适体特异性结合不同靶标蛋白质。

[0029] 在任何本文所述的实施方案中,至少一种适体可为缓慢解离速率适体。在任何本文所述的实施方案中,各适体可为缓慢解离速率适体。在一些实施方案中,至少一种缓慢解离速率适体包含至少一个、至少两个、至少三个、至少四个、至少五个、至少六个、至少七个、至少八个、至少九个、或至少10个具有疏水性修饰的核苷酸。在一些实施方案中,各缓慢解离速率适体以 ≥ 30 分钟、 ≥ 60 分钟、 ≥ 90 分钟、 ≥ 120 分钟、 ≥ 150 分钟、 ≥ 180 分钟、 ≥ 210 分钟或 ≥ 240 分钟的解离速率($t_{1/2}$)结合它的靶标蛋白质。

[0030] 在一些实施方案中,提供组合物。在一些此类实施方案中,组合物包含来自受试者的样品的蛋白质以及至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、至少八种、至少九种、至少十种、或十一种特异性结合选自ACY、SHBG、CTS、MET、GSN、LGALS3BP、PLAT、CHL1、SERPINC1、SIGLEC7和SIGLEC14的靶标蛋白质的适体。在一些此类实施方案中,组合物包含来自受试者的样品的蛋白质以及至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、或八种特异性结合选自ACY、SHBG、CTS、MET、GSN、LGALS3BP、CHL1、SERPINC1的靶标蛋白质的适体。在一些实施方案中,组合物包含来自受试者的样品的蛋白质以及至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、至少八种、至少九种、或至少十种特异性结合选自表7中的蛋白质的靶标蛋白质的适体。在一些实施方案中,组合物包含来自受试者的样品的蛋白质以及N种适体,其中各适体特异性结合表7中所列的生物标记蛋白质。在一些实施方案中,N是1至25。在一些实施方案中,N是2至25。在一些实施方案中,N是3至25。在一些实施方案中,N是4至25。在一些实施方案中,N是5至25。在一些实施方案中,N是1至10。在一些实施方案中,N是2至10。在一些实施方案中,N是3至10。在一些实施方案中,N是4至10。在一些实施方案中,N是5至10。在一些实施方案中,N种适体中的至少一种特异性结合选自ACY和SHBG的生物标记蛋白质。在一些

实施方案中,N种适体中的一种特异性结合ACY,并且N种适体中的一种特异性结合SHBG。在一些实施方案中,各适体特异性结合不同靶标蛋白质。

[0031] 在一些实施方案中,组合物包含来自受试者的样品的蛋白质以及至少一种、至少两种、或三种特异性结合选自ACY、SHBG和 SIGLEC14的靶标蛋白质的适体。在一些实施方案中,组合物包含来自受试者的样品的蛋白质以及至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、或七种特异性结合选自ACY、SHBG、 CTSZ、MET、GSN、LGALS3BP和 SIGLEC7的靶标蛋白质的适体。在一些实施方案中,组合物包含来自受试者的样品的蛋白质以及至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、至少八种、或九种特异性结合选自ACY、SHBG、CTSZ、MET、 GSN、LGALS3BP、PLAT、CHL1和SERPINC1的靶标蛋白质的适体。

[0032] 在一些实施方案中,提供一种组合物,其包含来自受试者的样品的蛋白质以及至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、至少八种、至少九种、或至少十种特异性结合选自C7、PPID、IGFBP3、SIGLEC14、AIMP1、TOP1、COLEC11、 CA6、STX1A、GPC5、HIPK3、IGFBP7和IL3RA的靶标蛋白质的适体。在一些实施方案中,提供一种组合物,其包含来自受试者的样品的蛋白质以及至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、至少八种、或九种特异性结合选自C7、 PPID、IGFBP3、SIGLEC14、AIMP1、TOP1、COLEC11、CA6和STX1A 的靶标蛋白质的适体。在一些实施方案中,提供一种组合物,其包含来自受试者的样品的蛋白质以及至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、至少八种、或八种特异性结合选自 C7、PPID、IGFBP3、COLEC11、GPC5、HIPK3、IGFBP7和IL3RA 的靶标蛋白质的适体。在一些实施方案中,各适体结合不同靶标蛋白质。

[0033] 在一些实施方案中,提供一种组合物,其包含来自受试者的样品的蛋白质以及至少一种、至少两种、至少三种、或四种特异性结合选自C7、COLEC11、PPID和IGFBP3的靶标蛋白质的适体。在一些实施方案中,提供一种组合物,其包含来自受试者的样品的蛋白质以及至少一种、至少两种、至少三种、至少四种、至少五种、或六种特异性结合选自SIGLEC14、AIMP1、TOP1、COLEC11、CA6和STX1A 的靶标蛋白质的适体。

[0034] 在一些实施方案中,提供一种组合物,其包含来自受试者的样品的蛋白质以及至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、或八种特异性结合选自表9中的蛋白质的靶标蛋白质的适体。在一些实施方案中,提供一种组合物,其包含来自受试者的样品的蛋白质以及N种适体,其中各适体特异性结合表9中所列的生物标记蛋白质。在一些实施方案中,N是1至8。在一些实施方案中,N是2至8。在一些实施方案中,N是3至8。在一些实施方案中,N是4至8。在一些实施方案中,N是5至8。在一些实施方案中,N种适体中的至少一种特异性结合COLEC11。在一些实施方案中,N种适体中的至少一种特异性结合选自C7、COLEC11、 PPID和IGFBP3的生物标记蛋白质。在一些实施方案中,N种适体中的四种各自特异性结合选自C7、COLEC11、PPID和IGFBP3的不同生物标记蛋白质。在一些实施方案中,各适体特异性结合不同靶标蛋白质。

[0035] 在任何本文所述的实施方案中,样品可为血液样品。在任何本文所述的实施方案中,样品可选自血清样品和血浆样品。

[0036] 在任何本文所述的实施方案中,至少一种适体可为缓慢解离速率适体。在任何本

文所述的实施方案中,各适体可为缓慢解离速率适体。在一些实施方案中,至少一种缓慢解离速率适体包含至少一个、至少两个、至少三个、至少四个、至少五个、至少六个、至少七个、至少八个、至少九个、或至少10个具有疏水性修饰的核苷酸。在一些实施方案中,各缓慢解离速率适体以 ≥ 30 分钟、 ≥ 60 分钟、 ≥ 90 分钟、 ≥ 120 分钟、 ≥ 150 分钟、 ≥ 180 分钟、 ≥ 210 分钟或 ≥ 240 分钟的解离速率($t_{1/2}$)结合它的靶标蛋白质。

[0037] 在任何本文所述的实施方案中,各生物标记可为蛋白质生物标记。在任何本文所述的实施方案中,方法可包括使来自受试者的样品的生物标记与一组生物标记捕获试剂接触,其中所述组的生物标记捕获试剂中的各生物标记捕获试剂特异性结合所检测的生物标记。在一些实施方案中,生物标记捕获试剂组中的各生物标记捕获试剂特异性结合所检测的不同生物标记。在任何本文所述的实施方案中,各生物标记捕获试剂可为抗体或适体。在任何本文所述的实施方案中,各生物标记捕获试剂可为适体。在任何本文所述的实施方案中,至少一种适体可为缓慢解离速率适体。在任何本文所述的实施方案中,至少一种缓慢解离速率适体可包含至少一个、至少两个、至少三个、至少四个、至少五个、至少六个、至少七个、至少八个、至少九个、或至少10个具有修饰的核苷酸。在一些实施方案中,修饰是疏水性修饰。在一些实施方案中,修饰是疏水性碱基修饰。在一些实施方案中,一个或多个修饰可选自图11中所示的修饰。在一些实施方案中,各缓慢解离速率适体以 ≥ 30 分钟、 ≥ 60 分钟、 ≥ 90 分钟、 ≥ 120 分钟、 ≥ 150 分钟、 ≥ 180 分钟、 ≥ 210 分钟或 ≥ 240 分钟的解离速率($t_{1/2}$)结合它的靶标蛋白质。

[0038] 在任何本文所述的实施方案中,样品可为血液样品。在一些实施方案中,血液样品选自血清样品和血浆样品。

[0039] 在任何本文所述的实施方案中,组合物中的样品可为血液样品。在一些实施方案中,血液样品选自血清样品和血浆样品。

[0040] 在任何本文所述的实施方案中,试剂盒或组合物可包含至少一种是缓慢解离速率适体的适体。在任何本文所述的实施方案中,试剂盒或组合物的各适体可为缓慢解离速率适体。在一些实施方案中,至少一种缓慢解离速率适体包含至少一个、至少两个、至少三个、至少四个、至少五个、至少六个、至少七个、至少八个、至少九个、或至少10个具有修饰的核苷酸。在一些实施方案中,至少一个具有修饰的核苷酸是具有疏水性碱基修饰的核苷酸。在一些实施方案中,各具有修饰的核苷酸是具有疏水性碱基修饰的核苷酸。在一些实施方案中,各疏水性碱基修饰独立地选自图11中的修饰。在一些实施方案中,试剂盒中的各缓慢解离速率适体以 ≥ 30 分钟、 ≥ 60 分钟、 ≥ 90 分钟、 ≥ 120 分钟、 ≥ 150 分钟、 ≥ 180 分钟、 ≥ 210 分钟或 ≥ 240 分钟的解离速率($t_{1/2}$)结合它的靶标蛋白质。

[0041] 附图简述

[0042] 图1显示如实施例2中所述的脂肪变性分类器的稳定性选择路径。

[0043] 图2显示如实施例2中所述的针对脂肪变性的9种生物标记的分类器的ROC曲线。

[0044] 图3显示如实施例2中所述的针对脂肪变性的9种标记的随机森林分类器的根据类别的表决。

[0045] 图4显示如实施例2中所述的针对脂肪变性的9种标记的分类器中的各生物标记的累积分布函数。

[0046] 图5显示如实施例2中所述的NASH(纤维化)分类器的稳定性选择路径。

[0047] 图6显示如实施例2中所述的针对NASH(纤维化)的4种生物标记的分类器的ROC曲线。

[0048] 图7显示如实施例2中所述的各受试者组中针对NASH(纤维化)的4种标记的分类器的箱线图。

[0049] 图8显示如实施例2中所述的针对NASH(纤维化)的4种标记的分类器中的各生物标记的累积分布函数。

[0050] 图9说明供与本文所述的各种计算机实施方法一起使用的非限制性示例性计算机系统。

[0051] 图10说明可用于检测生物样品中的一种或多种生物标记的非限制性示例性适体测定。

[0052] 图11显示可并入如缓慢解离速率适体的适体中的某些示例性修饰的嘧啶。

[0053] 图12显示如实施例5中所述的各受试者组中针对脂肪变性的8种标记的分类器的箱线图。

[0054] 图13显示如实施例5中所述的各受试者组中针对纤维化的8种标记的分类器的箱线图。

[0055] 图14显示如实施例5中所述的针对NASH(纤维化)的4种标记的分类器中的各生物标记的累积分布函数。

[0056] 图15显示如实施例6中所述的在破盲之后盲化样品的根据它们的实际样品组的针对脂肪变性的8种标记的分类器的箱线图。

[0057] 图16显示如实施例6中所述的发现组和盲化验证组的8种标记的脂肪变性分类器性能的ROC曲线。

[0058] 图17显示如实施例6中所述的在破盲之后盲化样品的根据它们的实际样品组的针对纤维化的8种标记的分类器的箱线图。

[0059] 图18显示如实施例6中所述的发现组和两个盲化验证组的8种标记的纤维化分类器性能的ROC曲线。

[0060] 图19A和B显示如实施例6中所述的使用(A)8种标记的脂肪变性分类器和(B)8种标记的纤维化分类器,20%保留验证组的2500次引导迭代的灵敏性和特异性分布。

[0061] 图20显示如实施例7中所述的各儿科样品中根据它们的实际样品组的8种标记的脂肪变性分类器的箱线图。

[0062] 详述

[0063] 尽管本发明将连同某些代表性实施方案一起加以描述,但应了解本发明由权利要求限定,并且不限于那些实施方案。

[0064] 本领域技术人员将认识到与本文所述的方法和材料类似或等效的许多方法和材料可用于实施本发明。本发明决不限于所述方法和材料。

[0065] 除非另外定义,否则本文所用的技术和科学术语具有由本发明所属领域中的普通技术人员通常理解的含义。尽管与本文所述的方法、装置和材料类似或等效的任何方法、装置和材料都可用于实施本发明,但本文描述某些方法、装置和材料。

[0066] 本文引用的所有出版物、公开的专利文件和专利申请都据此以引用的方式并入本文,所述引用的程度就好像各个别出版物、公开的专利文件或专利申请被特定地以及个别

地指示为以引用的方式并入一样。

[0067] 除非上下文另外明确规定,否则如包括随附权利要求的本申请中所用,单数形式“一个(种)”和“所述”包括复数形式,并且可与“至少一个(种)”和“一个(种)或多个(种)”互换使用。因此,提及“一种适体”包括适体的混合物,提及“一种探针”包括探针的混合物等。

[0068] 如本文所用,术语“包含”、“包括”、“含有”及其任何变化形式意图涵盖非排它性内含物,以使包含、包括或含有某一要素或要素列表的工艺、方法、通过工艺获得的产物、或物质组合物可包括未明确列出的其它要素。

[0069] 本申请包括用于确定受试者是否患有NAFLD的生物标记、方法、装置、试剂、系统和试剂盒。本申请也包括用于确定受试者是否患有 NASH的生物标记、方法、装置、试剂、系统和试剂盒。在一些实施方案中,提供用于确定患有NAFLD的受试者是否患有NASH的生物标记、方法、装置、试剂、系统和试剂盒。

[0070] 在一些实施方案中,提供一种或多种单独或以各种组合用于确定受试者是否患有NAFLD的生物标记。如以下所详述,示例性实施方案包括表3、4、6、7、8和9中提供的使用多种基于适体的测定鉴定的生物标记。表3列出9种适用于区分自正常肥胖个体获得的样品与来自患有NAFLD的个体的样品的生物标记。表8列出8种适用于区分自正常肥胖个体获得的样品与来自患有NAFLD的个体的样品的生物标记。表4列出4种适用于区分自患有脂肪变性的个体获得的样品与来自患有2、3和4期NASH的个体的样品的生物标记。表9列出 8种适用于区分自患有脂肪变性的个体获得的样品与来自患有2、3 和4期NASH的个体的样品的生物标记。表6和7列出可彼此和/或与来自表3、4、8和/或9的生物标记以任何组合使用的其它生物标记。在一些实施方案中,将一子组来自表3、4、6、7、8和9的生物标记组合成表5中所示的组。

[0071] 在一些实施方案中,提供一种或多种单独或以各种组合用于确定受试者是否患有脂肪变性的生物标记。在一些实施方案中,受试者是肥胖的。如以下所详述,示例性实施方案包括表3和/或表8中提供的使用多种基于适体的测定鉴定的生物标记。表3列出4种适用于区分自肥胖个体获得的样品与来自患有脂肪变性的个体的样品的生物标记。表8列出8种适用于区分自肥胖个体获得的样品与来自患有脂肪变性的个体的样品的生物标记。此外,表3和/或表8中的一种或多种生物标记可与一种或多种来自表4和/或表6和/或表7和/或表8 和/或表9的生物标记组合,与或不与一种或多种未列于表3、4、6、7、8或9中的任一个中的生物标记组合用于本文所述的方法中。

[0072] 在一些实施方案中,提供一种或多种单独或以各种组合用于确定受试者是否患有任何时期的NASH的生物标记。在一些实施方案中,提供一种或多种单独或以各种组合用于确定受试者是否患有2、3或 4期NASH的生物标记。在一些实施方案中,已知受试者患有脂肪变性。如以下所详述,示例性实施方案包括表4和/或表9中提供的使用多种基于适体的测定鉴定的生物标记。表4列出4种适用于区分自患有脂肪变性的个体获得的样品与来自患有NASH的个体的样品的生物标记。表9列出8种适用于区分自患有脂肪变性的个体获得的样品与来自患有NASH的个体的样品的生物标记。此外,表4和/或表 9中的一种或多种生物标记可与一种或多种来自表3和/或表6和/或表7和/或表8的生物标记组合,与或不与一种或多种未列于表3、4、6、7、8或9中的任一个中的生物标记组合用于本文所述的方法中。

[0073] 在一些实施方案中,基于生物标记值的特定组合的灵敏性和特异性来选择某一组

中的生物标记的数目和身份。术语“灵敏性”和“特异性”在本文中关于基于在生物样品中检测的一个或多个生物标记水平将个体正确分类为患有疾病或不患有疾病的能力来使用。在一些实施方案中,术语“灵敏性”和“特异性”可在本文中关于基于在生物样品中检测的一个或多个生物标记水平将个体正确分类为患有脂肪变性或不患有脂肪变性的能力来使用。在所述实施方案中,“灵敏性”指示生物标记关于正确分类患有脂肪变性的个体的性能。“特异性”指示生物标记关于正确分类不患有脂肪变性的个体的性能。举例来说,用于测试一组对照样品(如来自已知不患有脂肪变性的健康个体或受试者的样品)和测试样品(如来自患有脂肪变性的个体的样品)的一组生物标记的85%特异性和90%灵敏性指示85%的对照样品由所述组正确分类为对照样品,并且90%的测试样品由所述组正确分类为测试样品。

[0074] 在一些实施方案中,术语“灵敏性”和“特异性”可在本文中关于基于在生物样品中检测的一个或多个生物标记水平将个体正确分类为患有NASH(或2、3或4期NASH)或患有脂肪变性的能力来使用。“灵敏性”指示生物标记关于正确分类患有NASH(或2、3或4期 NASH)的个体的性能。“特异性”指示生物标记关于正确分类不患有 NASH(或不患有2、3或4期NASH)的个体的性能。举例来说,用于测试一组对照样品(如来自患有脂肪变性的个体的样品)和测试样品(如来自患有NASH或2、3或4期NASH的个体的样品)的一组生物标记的85%特异性和90%灵敏性指示85%的对照样品由所述组正确分类为对照样品,并且90%的测试样品由所述组正确分类为测试样品。

[0075] 在一些实施方案中,一组一种或多种生物标记的总体性能由曲线下面积(AUC)值表示。AUC值由本文例示的接受者操作特征(ROC)曲线获得。ROC曲线是测试的真阳性率(灵敏性)相对于测试的假阳性率(1-特异性)的图。术语“曲线下面积”或“AUC”是指接受者操作特征(ROC)曲线的曲线下面积,其两者均在本领域中是熟知的。AUC量度适用于比较分类器跨越完全数据范围的准确度。具有较大 AUC的分类器具有较大的用在两个目标组(例如正常个体和患有 NAFLD的个体、或患有脂肪变性的个体和患有NASH的个体)之间正确分类未知物的能力。ROC曲线适用于绘制特定特征(例如任何本文所述的生物标记和/或任何其他生物医学信息条目)在区分两个群体方面的性能。通常,基于单一特征的值以上升顺序分类跨越整个群体的特征数据。接着,对于那个特征的值,计算数据的真阳性率和假阳性率。通过计数高于那个特征的值的事例的数目,接着除以事例的总数来确定真阳性率。通过计数高于那个特征的值的事例的数目,接着除以对照的总数来确定假阳性率。尽管这个定义涉及其中相较于对照,事例中的特征升高的情况,但这个定义也适用于其中相较于对照,事例中的特征较低的情况(在所述情况下,将计数低于那个特征的值的事例)。可产生单一特征以及其它单一输出的ROC曲线,例如可以数学方式组合(例如加、减、乘等)两种或更多种特征的组合以提供单一总和值,并且这个单一总和值可绘制在ROC曲线中。另外,可在ROC曲线中绘制多种特征的任何组合,其中组合导出单一输出值。

[0076] 在一些实施方案中,方法包括检测来自受试者的样品中至少一种列于表3、4、6、7、8和/或9中的生物标记的水平以确定受试者是否患有NAFLD。在一些此类实施方案中,方法包括使来自受试者的样品或样品的一部分与至少一种捕获试剂接触,其中各捕获试剂特异性结合其水平被检测的生物标记。在一些实施方案中,方法包括使样品或来自样品的蛋白质与至少一种适体接触,其中各适体特异性结合其水平被检测的生物标记。

[0077] 在一些实施方案中,方法包括检测来自受试者的样品中至少一种、至少两种、至少

三种、至少四种、至少五种、至少六种、至少七种、至少八种、至少九种、至少十种、或十一种选自ACY、SHBG、CTSZ、MET、GSN、LGALS3BP、PLAT、CHL1、SERPINC1、SIGLEC7 和SIGLEC14的生物标记的水平,其中至少一种选自ACY、CTSZ、LGALS3BP、SIGLEC7、SIGLEC14和PLAT的生物标记的水平高于相应生物标记的对照水平,和/或至少一种选自SHBG、MET、GSN、CHL1和SERPINC1的生物标记的水平低于相应生物标记的对照水平指示所述受试者患有NAFLD。在一些实施方案中,方法包括检测来自受试者的样品中至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、或八种选自ACY、SHBG、CTSZ、MET、GSN、LGALS3BP、CHL1、SERPINC1的生物标记的水平,其中至少一种选自ACY、CTSZ和LGALS3BP的生物标记的水平高于相应生物标记的对照水平,和/或至少一种选自SHBG、MET、GSN、CHL1 和SERPINC1的生物标记的水平低于相应生物标记的对照水平指示所述受试者患有NAFLD。在一些实施方案中,受试者是肥胖受试者。在一些实施方案中,方法包括确定受试者是否患有脂肪变性,和/或确定脂肪变性是轻度、中度或重度。在一些此类实施方案中,方法包括检测至少一种、至少两种、或三种选自ACY、SHBG和SIGLEC14 的生物标记的水平。在一些实施方案中,方法包括确定受试者是否患有NASH,如1、2、3或4期NASH。在一些此类实施方案中,方法包括检测至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、或七种选自ACY、SHBG、CTSZ、MET、GSN、LGALS3BP 和SIGLEC7的生物标记的水平。在一些实施方案中,方法包括确定受试者是否患有可为脂肪变性或NASH的NAFLD。在一些此类实施方案中,方法包括检测至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、至少八种、或九种选自ACY、SHBG、CTSZ、MET、GSN、LGALS3BP、PLAT、CHL1和SERPINC1的生物标记的水平。

[0078] 在一些实施方案中,方法包括检测至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、至少八种、至少九种、或至少十种选自表7中的生物标记的生物标记的水平。在一些实施方案中,表7中除SHBG和GSN以外的生物标记的水平高于相应生物标记的对照水平,和/或至少一种选自SHBG和GSN的生物标记的水平低于相应生物标记的对照水平指示受试者患有NAFLD。

[0079] 本文鉴定的生物标记提供对可用于有效鉴定NAFLD的生物标记的子组或组的许多选择。对所述生物标记的适当数目的选择可取决于所选生物标记的特定组合。此外,在任何本文所述的方法中,除其中明确指示之外,一组生物标记可包括未显示于表3、4、6、7、8或9中的其它生物标记。

[0080] 在一些实施方案中,方法包括检测来自受试者的样品中至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、至少八种、至少九种、或至少十种选自C7、PPID、IGFBP3、SIGLEC14、AIMP1、TOP1、COLEC11、CA6、STX1A、GPC5、HIPK3、IGFBP7和IL3RA的生物标记的水平,其中至少一种选自C7、COLEC11、HIPK3、IGFBP7、IL3RA和SIGLEC14的生物标记的水平高于相应生物标记的对照水平,和/或至少一种选自IGFBP3、AIMP1、TOP1、CA6、PPID、GPC3、IGFBP3和STX1A的生物标记的水平低于相应生物标记的对照水平指示所述受试者患有NASH。在一些实施方案中,方法包括检测来自受试者的样品中至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、至少八种、或九种选自C7、PPID、IGFBP3、SIGLEC14、AIMP1、TOP1、COLEC11、CA6和STX1A的生物标记的水平,其中至少一种选自C7、COLEC11和SIGLEC14的生物标记的水平高于相应生物标记的对照水平,和/或至少一种选自IGFBP3、AIMP1、TOP1、CA6、PPID 和STX1A的生物标记的水平低于相应生物标记的对

照水平指示所述受试者患有NASH。在一些实施方案中,方法包括检测来自受试者的样品中至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、或八种选自C7、PPID、IGFBP3、COLEC11、GPC5、HIPK3、IGFBP7和IL3RA的生物标记的水平,其中至少一种选自 C7、COLEC11、HIPK3、IGFBP7和IL3RA的生物标记的水平高于相应生物标记的对照水平指示所述受试者患有NASH,和/或至少一种选自IGFBP3、PPID和GPC3的生物标记的水平低于相应生物标记的对照水平指示所述受试者患有NASH。在一些实施方案中,受试者是肥胖的。在一些实施方案中,受试者已被确定患有可为轻度、中度或重度脂肪变性的脂肪变性。在一些实施方案中,NASH是1、2、3或 4期NASH。在一些实施方案中,NASH是2、3或4期NASH。在一些此类实施方案中,方法包括检测至少一种、至少两种、至少三种、或四种选自C7、COLEC11、PPID和IGFBP3的生物标记的水平。

[0081] 在一些实施方案中,方法包括检测来自受试者的样品中至少一种列于表3、4、6、7、8和/或9中的生物标记的水平以确定受试者是否患有NASH或2、3或4期NASH。在一些此类实施方案中,方法包括使来自受试者的样品或样品的一部分与至少一种捕获试剂接触,其中各捕获试剂特异性结合其水平被检测的生物标记。在一些实施方案中,方法包括使样品或来自样品的蛋白质与至少一种适体接触,其中各适体特异性结合其水平被检测的生物标记。

[0082] 本文鉴定的生物标记提供对可用于有效鉴定NASH或2、3或4 期NASH的生物标记的子组或组的许多选择。对所述生物标记的适当数目的选择可取决于所选生物标记的特定组合。此外,在任何本文所述的方法中,除其中明确指示之外,一组生物标记可包括未显示于表3、4、6、7、8或9中的其它生物标记。

[0083] 如本文所用,“非酒精性脂肪肝疾病”或“NAFLD”是指其中在不存在过度酒精使用下,脂肪沉积在肝中的伴有或不伴有炎症和纤维化的病状(肝性脂肪变性)。

[0084] 如本文所用,“脂肪变性”和“非酒精性脂肪变性”可互换使用,并且包括在不存在过度酒精使用下,不伴有炎症或纤维化的轻度、中度和重度脂肪变性。表1显示轻度、中度和重度脂肪变性的示例性分类。

[0085] 如本文所用,“非酒精性脂肪性肝炎”或“NASH”是指其中在肝中存在炎症和/或纤维化的NAFLD。NASH可被分成四个时期。确定 NASH的时期的示例性方法例如描述于Kleiner等,2005,Hepatology, 41 (6):1313-1321;以及Brunt等,2007,Modern Pathol.,20:S40-S48中。表1显示1、2、3和4期NASH的示例性分类。

[0086] 如本文所用,关于受试者的“肥胖”是指受试者具有30或更多的 BMI。

[0087] “生物样品”、“样品”和“测试样品”在本文中可互换用于指代得自或另外源于个体的任何材料、生物流体、组织或细胞。这包括血液(包括全血、白细胞、外周血液单核细胞、白细胞层、血浆和血清)、痰、泪、粘液、鼻洗涤物、鼻吸出物、尿、唾液、腹膜洗涤物、腹水、囊液、腺液、淋巴液、支气管吸出物、滑液、关节吸出物、器官分泌物、细胞、细胞提取物和脑脊髓液。这也包括所有先前各物的实验分离部分。举例来说,血液样品可被分级分离成血清、血浆或含有特定类型的血细胞,如红血细胞或白血细胞(白细胞)的部分。在一些实施方案中,样品可为来自个体的样品的组合,如组织和流体样品的组合。术语“生物样品”也包括如来自例如粪便样品、组织样品或组织活检体的含有均质化固体物质的材料。术语“生物样品”也包括由组织培养或细胞培养获得的材料。可采用任何适于获得生物样品的方法;示例性

方法包括例如放血、擦拭(例如经颊擦拭)和细针吸出活检程序。易经受细针吸出的示例性组织包括淋巴结、肺、甲状腺、乳房、胰腺和肝。也可例如通过显微解剖(例如激光捕获显微解剖(LCM)或激光显微解剖(LMD))、膀胱洗涤、涂抹(例如PAP涂抹)或导管灌洗来收集样品。得自或源于个体的“生物样品”包括在自所述个体获得之后已用任何适合方式处理的任何所述样品。

[0088] 此外,在一些实施方案中,可通过自许多个体获取生物样品并将它们汇合,或汇合各个体的生物样品的等分试样来获得生物样品。可如本文对于来自单一个体的样品所述处理汇合的样品,并且例如如果在汇合的样品中确定不良预后,那么可再测试各个别生物样品以确定哪些个体患有脂肪变性和/或NASH。

[0089] “靶标”、“靶标分子”和“分析物”在本文中可互换用于指代可存在于生物样品中的任何目标分子。“目标分子”包括特定分子的任何微小变化形式,如在蛋白质的情况下,例如在氨基酸序列、二硫键形成、糖基化、脂质化、乙酰化、磷酸化或不实质上改变分子的身份的任何其它操作或修饰(如与标记组分缀合)方面的微小变化形式。“靶标分子”、“靶标”或“分析物”是指一种类型或种类的分子或多分子结构的一组拷贝。“靶标分子”、“靶标”和“分析物”是指超过一种类型或种类的分子或多分子结构。示例性靶标分子包括蛋白质、多肽、核酸、碳水化合物、脂质、多糖、糖蛋白、激素、受体、抗原、抗体、亲和体、抗体模拟物、病毒、病原体、毒性物质、底物、代谢物、过渡态类似物、辅因子、抑制剂、药物、染料、营养物、生长因子、细胞、组织以及任何前述各物的任何片段或部分。在一些实施方案中,靶标分子是蛋白质,在所述情况下,靶标分子可被称为“靶标蛋白质”。

[0090] 如本文所用,“捕获剂”或“捕获试剂”是指能够特异性结合生物标记的分子。“靶标蛋白质捕获试剂”是指能够特异性结合靶标蛋白质的分子。非限制性示例性捕获试剂包括适体、抗体、纤维连接蛋白、锚蛋白、其它抗体模拟物和其它蛋白质骨架、自体抗体、嵌合体、小分子、核酸、凝集素、配体结合受体、印迹聚合物、高亲和性多聚体、肽模拟物、激素受体、细胞因子受体、合成受体以及任何以上提及的捕获试剂的修饰形式和片段。在一些实施方案中,捕获试剂选自适体和抗体。

[0091] 术语“抗体”是指任何物种的全长抗体以及所述抗体的片段和衍生物,包括Fab片段、 $F(ab')_2$ 片段、单链抗体、Fv片段和单链Fv片段。术语“抗体”也指合成获得的抗体,如噬菌体展示获得的抗体和片段、亲和体、纳米体等。

[0092] 如本文所用,“标记”和“生物标记”可互换用于指代指示以下或是以下的征象的靶标分子:个体中正常或异常过程或个体的疾病或其它状况。更具体来说,“标记”或“生物标记”是与存在特定正常或异常,并且如果异常,那么慢性或急性生理状态或过程相关的解剖学、生理学、生物化学或分子参数。生物标记可通过多种方法,包括实验室测定和医学成像来检测和测量。在一些实施方案中,生物标记是靶标蛋白质。

[0093] 如本文所用,“生物标记水平”和“水平”是指使用用于检测生物样品中的生物标记的任何分析方法获得,以及指示生物样品中的生物标记的或对应于生物样品中的生物标记的存在、不存在、绝对量或浓度、相对量或浓度、滴度、水平、表达水平、测量水平的比率等的测量结果。“水平”的精确性质取决于用于检测生物标记的特定分析方法的特定设计和组成部分。

[0094] 靶标分子的“对照水平”是指来自未患有疾病或病状的个体或来自未怀疑患有疾

病或病状的个体的相同样品类型中的靶标分子的水平。靶标分子的“对照水平”无需每次在执行本发明方法时加以测定,并且可为先前测定的用作用以确定特定样品中的水平高于或低于正常水平的参照或阈值的水平。在一些实施方案中,本文所述的方法中的对照水平是已在一个或多个不患有NAFLD的受试者中观察的水平。在一些实施方案中,本文所述的方法中的对照水平是已在一个或多个患有NAFLD而非NASH的受试者中观察的水平。在一些实施方案中,本文所述的方法中的对照水平是已在多个正常受试者或患有 NAFLD而非NASH的受试者中观察的平均或均值水平,任选加上或减去统计变化。

[0095] 如本文所用,“个体”和“受试者”可互换用于指代测试受试者或患者。个体可为哺乳动物或非哺乳动物。在各种实施方案中,个体是哺乳动物。哺乳动物个体可为人或非人。在各种实施方案中,个体是人。健康或正常个体是其中不可通过常规诊断方法检测目标疾病或病状(如NASH)的个体。

[0096] “诊断(Diagnose/diagnosing/diagnosis)”及其变化形式是指基于一种或多种关于个体的征象、症状、数据或其它信息来检测、确定或识别那个个体的健康状态或状况。个体的健康状态可被诊断为健康/正常(即诊断不存在疾病或病状)或诊断为生病/异常(即诊断存在疾病或病状或评估有疾病或病状的特征)。术语“诊断”等关于特定疾病或病状涵盖初始检测疾病;表征或分类疾病;检测疾病的进展、缓解或复发;以及在向个体施用治疗或疗法之后检测疾病应答。诊断 NAFLD包括区分患有NAFLD的个体与不患有NAFLD的个体。诊断NASH包括区分患有NASH的个体与患有肝脂肪变性而非NASH 的个体以及与不患有肝病的个体。

[0097] “预测/预后(Prognose/prognosing/prognosis)”及其变化形式是指预测患有疾病或病状的个体的所述疾病或病状的未来病程(例如预测患者存活期),并且所述术语涵盖在向所述个体施用治疗或疗法之后评估疾病应答。

[0098] “评估(Evaluate/evaluating/evaluation)”及其变化形式涵盖“诊断”与“预测”两者,并且也涵盖关于未患有疾病的个体的疾病或病状的未来病程进行确定或预测,以及关于疾病或病状将重现于表面上已治愈疾病的个体中的可能性进行确定或预测。术语“评估”也涵盖估计个体对疗法的应答,例如像预测个体可能对治疗剂有利地应答或不可能对治疗剂应答(或将经受例如毒性或其它不合需要的副作用),选择供向个体施用的治疗剂,或监测或测定个体对已向个体施用的疗法的应答。因此,“评估”NAFLD可包括例如以下中的任一个:预测个体的 NAFLD的未来病程;预测NAFLD是否将进展成NASH;预测特定时期的NASH是否将进展成更高时期的NASH;等。

[0099] 如本文所用,关于生物标记水平进行的“检测”或“测定”包括使用用于观察以及记录对应于生物标记水平的信号的仪器与为产生那个信号所需的材料两者。在各种实施方案中,使用任何适合方法检测水平,所述方法包括荧光、化学发光、表面等离子体共振、表面声波、质谱测定法、红外光谱法、拉曼光谱法(Raman spectroscopy)、原子力显微术、扫描隧穿显微术、电化学检测方法、核磁共振、量子点等。

[0100] 如本文所用,“患有NAFLD的受试者”是指已被诊断有NAFLD 的受试者。在一些实施方案中,NAFLD在常规检查、监测代谢综合征和肥胖症、或监测药物(例如胆固醇降低剂或类固醇)的可能副作用期间被怀疑。在某一情况下,如AST和ALT的肝酶较高。在一些实施方案中,受试者在腹部或胸部成像、肝超声或磁共振成像之后被确诊。在一些实施方案中,在

NAFLD诊断之前,如过量酒精消耗、丙型肝炎和威尔森氏病(Wilson's disease)的其它病状已被排除。在一些实施方案中,在肝活检之后,受试者已被确诊。

[0101] 如本文所用,“患有脂肪变性的受试者”和“患有非酒精性脂肪变性的受试者”可互换使用,并且是指已被诊断有脂肪变性的受试者。在一些实施方案中,通过以上大体上对于NAFLD所述的方法来诊断脂肪变性。

[0102] 如本文所用,“患有NASH的受试者”是指已被诊断有NASH的受试者。在一些实施方案中,通过以上大体上对于NAFLD所述的方法来诊断NASH。在一些实施方案中,例如根据Gambino R等Annals of Medicine 2011;43(8):617-49,在患有NAFLD的患者中诊断出晚期纤维化。

[0103] 如本文所用,“处于发展NAFLD的风险下的受试者”是指患有的一种或多种NAFLD共患病,如肥胖症、腹部肥胖、代谢综合征、心血管疾病和糖尿病的受试者。

[0104] 如本文所用,“处于发展脂肪变性的风险下的受试者”是指尚未被诊断为患有脂肪变性,但患有的一种或多种NAFLD共患病,如肥胖症、腹部肥胖、代谢综合征、心血管疾病和糖尿病的受试者。

[0105] 如本文所用,“处于发展NASH的风险下的受试者”是指延续患有的一种或多种NAFLD共患病,如肥胖症、腹部肥胖、代谢综合征、心血管疾病和糖尿病的患有脂肪变性的受试者。

[0106] “固体载体”在本文中是指具有分子可通过共价或非共价键直接或间接与其连接的表面的任何基底。“固体载体”可具有多种实体形式,其可包括例如膜;芯片(例如蛋白质芯片);载片(例如玻璃载片或盖玻片);管柱;中空、固体、半固体、含有孔或空腔的粒子,例如像珠粒;凝胶;纤维,包括光学纤维材料;基质;和样品贮器。示例性样品贮器包括样品孔、管、毛细管、小瓶以及能够固持样品的任何其它容器、沟槽或凹槽。样品贮器可含于多样品平台,如微量滴定板、载片、微流体装置等上。载体可由天然或合成材料、有机或无机材料组成。其上连接捕获试剂的固体载体的组成通常取决于连接方法(例如共价连接)。其它示例性贮器包括在其内部可进行测定和相关操作的微滴和微流体控制或散装油/水乳液。适合固体载体包括例如塑料、树脂、多糖、二氧化硅或基于二氧化硅的材料、官能化玻璃、改性硅、碳、金属、无机玻璃、膜、尼龙、天然纤维(例如像丝、羊毛和棉花)、聚合物等。组成固体载体的材料可包括用于连接捕获试剂的反应性基团,例如像羧基、氨基或羟基。聚合固体载体可包括例如聚苯乙烯、聚对苯二甲酸乙二醇酯、聚乙酸乙烯酯、聚氯乙烯、聚乙烯吡咯烷酮、聚丙烯腈、聚甲基丙烯酸甲酯、聚四氟乙烯、丁基橡胶、苯乙烯丁二烯橡胶、天然橡胶、聚乙烯、聚丙烯、(聚)四氟乙烯、(聚)偏二氟乙烯、聚碳酸酯和聚甲基戊烯。可使用的适合固体载体粒子包括例如编码的粒子(如Luminex[®]型编码的粒子)、磁性粒子和玻璃粒子。

[0107] 生物标记的示例性用途

[0108] 在各种示例性实施方案中,提供用于确定受试者是否患有NAFLD的方法。在各种实施方案中,提供用于确定受试者是否患有可为轻度、中度或重度脂肪变性的脂肪变性的方法。在各种实施方案中,提供用于确定受试者是否患有可为1、2、3或4期NASH,或可为2、3或4期NASH的NASH的方法。在一些实施方案中,提供用于确定患有脂肪变性的受试者是否患有可为1、2、3或4期NASH,或可为2、3或4期NASH的NASH的方法。方法包括通过许多分析方法,包括任何本文所述的分析方法检测一个或多个对应于一种或多种存在于个体的循环中,如血清或血浆中的生物标记的生物标记水平。相较于正常个体(其中正常个体可为肥胖个

体),这些生物标记例如在不同水平下存在于患有NAFLD的个体中。在一些实施方案中,相较于正常个体(其中正常个体可为肥胖个体),生物标记在不同水平下存在于患有NASH(如1、2、3或4期NASH或2、3或4期 NASH)的个体中。在一些实施方案中,相较于患有可为轻度、中度或重度脂肪变性的脂肪变性的受试者,生物标记在不同水平下存在于患有NASH(如1、2、3或4期NASH或2、3或4期NASH)的个体中。

[0109] 对个体中生物标记的差异性水平的检测可例如用于允许确定个体是否患有NAFLD(其可为脂肪变性或NASH),或患有脂肪变性的个体是否已发展NASH。在一些实施方案中,任何本文所述的生物标记都可用于监测个体(如肥胖个体)的NAFLD发展,或可用于监测患有脂肪变性的个体的NASH发展。

[0110] 作为任何本文所述的生物标记可用于确定受试者是否患有 NAFLD所依的方式的实例,尚未被诊断有NAFLD,但患有的一种或多种NAFLD共患病的个体中的一种或多种所述生物标记的水平可在早于将使用侵袭性测试(如肝活检)进行确定的时期指示所述个体已发展NAFLD。因为本发明方法是非侵袭性的,所以它们可用于监测处于发展NAFLD的风险下的个体(例如像肥胖个体)。通过在较早时期检测NAFLD,医学干预可更有效。所述医学干预可包括但不限于减轻重量、控制血糖和避免酒精。在一些实施方案中,可使用治疗剂,如匹格列酮、维生素E和/或二甲双胍。参见例如Sanyal等,2010, NEJM,362:1675-1685。在一些情况下,所述早期干预可延迟或防止肝衰竭和对肝移植的需要。

[0111] 类似地,作为本文所述的生物标记可用于确定患有脂肪变性的受试者是否正发展NASH所依的方式的另一实例,患有脂肪变性的个体中的一种或多种所述生物标记的水平可指示个体正发展NASH。因为本发明方法是非侵袭性的,所以可监测患有脂肪变性的个体的NASH发展。通过在较早时期检测NASH,医学干预可更有效。所述医学干预可包括但不限于减轻重量、控制血糖和避免酒精。在一些实施方案中,可使用治疗剂,如匹格列酮、维生素E和/或二甲双胍。参见例如Sanyal等,2010,NEJM,362:1675-1685。在一些情况下,所述早期干预可延迟或防止肝衰竭和对肝移植的需要。

[0112] 此外,在一些实施方案中,个体中的一种或多种生物标记随时间的差异性表达水平可指示个体对特定治疗方案的应答。在一些实施方案中,在追踪监测期间,一种或多种生物标记的表达变化可指示特定疗法有效,或可表明治疗方案应以某一方式改变,如通过更积极控制血糖、更积极贯彻重量减轻等。在一些实施方案中,个体中的一种或多种生物标记随时间的恒定表达水平可指示个体的脂肪变性未恶化,或未发展成NASH。

[0113] 除测试生物标记水平以作为独立诊断测试之外,生物标记水平也可连同确定指示疾病的易感性风险增加的单核苷酸多态性(SNP)或其它遗传病变或变异性一起进行。(参见例如Amos等,Nature Genetics 40,616-622(2009))。

[0114] 除测试生物标记水平以作为独立诊断测试之外,生物标记水平也可连同其它NAFLD筛选方法,如检测肝增大、血液测试(例如以检测如ALT和/或AST的某些肝酶的升高)、腹部超声和肝活检一起进行。在一些情况下,使用本文所述的生物标记的方法可有助于对实施更积极NAFLD或NASH治疗、更频繁追踪筛选等的医学和经济论证。如果诊断测试指示处于发展NAFLD的风险下,但尚未被诊断有脂肪变性的个体可能发展疾病,那么生物标记也可用于开始在他们中进行的治疗。

[0115] 除连同其它NAFLD诊断方法一起测试生物标记水平之外,也可连同其它类型的数

据,特别是指示个体的NAFLD风险的数据一起评估关于生物标记的信息。这些各种数据可通过可包括在计算机或其它器具/装置中的自动化方法,如计算机程序/软件来估计。

[0116] 检测和测定生物标记和生物标记水平

[0117] 可使用多种已知分析方法中的任一个检测本文所述的生物标记的生物标记水平。在一个实施方案中,使用捕获试剂检测生物标记水平。在各种实施方案中,可在溶液中使捕获试剂暴露于生物标记,或可在使捕获试剂固定在固体载体上的情况下,使捕获试剂暴露于生物标记。在其它实施方案中,捕获试剂含有与固体载体上的第二特征具有反应性的特征。在这些实施方案中,可在溶液中使捕获试剂暴露于生物标记,接着捕获试剂上的特征可连同固体载体上的第二特征一起用于使生物标记固定在固体载体上。基于待进行的分析的类型选择捕获试剂。捕获试剂包括但不限于适体、抗体、纤维连接蛋白、锚蛋白、其它抗体模拟物和其它蛋白质骨架、自体抗体、嵌合体、小分子、F(ab')₂片段、单链抗体片段、Fv片段、单链Fv片段、核酸、凝集素、配体结合受体、亲和体、纳米体、印迹聚合物、高亲和性多聚体、肽模拟物、激素受体、细胞因子受体和合成受体以及这些捕获试剂的修饰形式和片段。

[0118] 在一些实施方案中,使用生物标记/捕获试剂复合物检测生物标记水平。

[0119] 在一些实施方案中,生物标记水平由生物标记/捕获试剂复合物获得,并且例如像由于在生物标记/捕获试剂相互作用之后,但取决于生物标记/捕获试剂复合物的形成的反应而间接检测。

[0120] 在一些实施方案中,生物标记水平直接自生物样品中的生物标记检测。

[0121] 在一些实施方案中,使用允许同时检测生物样品中的两种或更多种生物标记的多路形式检测生物标记。在多路形式的一些实施方案中,使捕获试剂直接或间接、共价或非共价固定于固体载体上的离散位置。在一些实施方案中,多路形式使用离散固体载体,其中各固体载体具有与那个固体载体缔合的独特捕获试剂,例如像量子点。在一些实施方案中,个别装置用于检测待在生物样品中检测的多种生物标记中的每一个。个别装置可被配置来允许生物样品中的各生物标记被同时处理。举例来说,可使用微量滴定板以使板中的各孔用于分析待在生物样品中检测的多种生物标记中的一个或多个。

[0122] 在一个或多个前述实施方案中,荧光标签可用于标记生物标记/捕获试剂复合物的组分以使得能够检测生物标记水平。在各种实施方案中,可使用已知技术使荧光标记缀合于对任何本文所述的生物标记具有特异性的捕获试剂,并且荧光标记可接着用于检测相应生物标记水平。适合荧光标记包括稀土螯合物、荧光素和它的衍生物、若丹明(rhodamine)和它的衍生物、丹磺酰、别藻蓝蛋白、PBXL-3、Qdot 605、丽丝胺、藻红素、得克萨斯红和其它所述化合物。

[0123] 在一些实施方案中,荧光标记是荧光染料分子。在一些实施方案中,荧光染料分子包括至少一个取代的吡啶鎓环系统,其中吡啶鎓环的3-碳上的取代基含有化学反应性基团或缀合的物质。在一些实施方案中,染料分子包括AlexaFluor分子,例如像AlexaFluor 488、AlexaFluor 532、AlexaFluor 647、AlexaFluor 680或AlexaFluor 700。在一些实施方案中,染料分子包括第一类型和第二类型的染料分子,例如像两种不同AlexaFluor分子。在一些实施方案中,染料分子包括第一类型和第二类型的染料分子,并且两种染料分子具有不同发射光谱。

[0124] 荧光可用可与广泛范围的测定形式相容的多种仪器测量。举例来说,分光荧光计

已被设计以分析微量滴定板、显微镜载片、印制的阵列芯片、比色皿等。参见由 J.R.Lakowicz 所著的 *Principles of Fluorescence Spectroscopy*, Springer Science+Business Media, Inc., 2004。参见 *Bioluminescence&Chemiluminescence: Progress&Current Applications*; Philip E. Stanley 和 Larry J. Kricka 编, World Scientific Publishing Company, 2002 年 1 月。

[0125] 在一个或多个实施方案中,化学发光标签可任选用于标记生物标记/捕获复合物的组分以使得能够检测生物标记水平。适合化学发光物质包括以下中的任一个:草酰氯、若丹明 6G、Ru(联吡啶)₃²⁺、TMAE(四(二甲基氨基)乙烯)、连苯三酚(1,2,3-三羟基苯)、光泽精(Lucigenin)、过氧草酸酯、草酸芳酯、吡啶鎓酯、二氧杂环丁烷等。

[0126] 在一些实施方案中,检测方法包括产生对应于生物标记水平的可检测信号的酶/底物组合。通常,酶催化可使用各种技术(包括分光光度测定法、荧光和化学发光)测量的显色底物的化学改变。适合酶包括例如荧光素酶、虫荧光素、苹果酸脱氢酶、脲酶、辣根过氧化物酶(HRPO)、碱性磷酸酶、β-半乳糖苷酶、葡糖淀粉酶、溶菌酶、葡萄糖氧化酶、半乳糖氧化酶和葡萄糖-6-磷酸脱氢酶、尿酸酶、黄嘌呤氧化酶、乳过氧化物酶、微过氧化物酶等。

[0127] 在一些实施方案中,检测方法可为产生可测量信号的荧光、化学发光、放射性核素或酶/底物组合的组合。在一些实施方案中,多模态信号传导可在生物标记测定形式中具有独特和有利的特征。

[0128] 在一些实施方案中,可使用任何分析方法检测本文所述的生物标记的生物标记水平,所述方法包括单路适体测定、多路适体测定、单路或多路免疫测定、mRNA 表达谱分析、miRNA 表达谱分析、质谱分析、组织学/细胞学方法等,如以下所讨论。

[0129] 使用基于适体的测定来测定生物标记水平

[0130] 涉及检测和定量生物样品和其它样品中的生理重要分子的测定是科学研究中以及健康护理领域中的重要工具。所述测定的一个类别涉及使用包括一种或多种固定在固体载体上的适体的微阵列芯片。适体各自能够以高度特异性方式以及以极高亲和力结合靶标分子。参见例如题为“Nucleic Acid Ligands”的美国专利号 5,475,096;也参见例如各自题为“Nucleic Acid Ligand Diagnostic Biochip”的美国专利号 6,242,246、美国专利号 6,458,543 和美国专利号 6,503,715。一旦使微阵列芯片与样品接触,适体即结合它们的存在于样品中的相应靶标分子,并且由此使得能够测定对应于生物标记的生物标记水平。

[0131] 如本文所用,“适体”是指对靶标分子具有特异性结合亲和力的核酸。应认识到亲和力相互作用是程度问题;然而,在这个情形下,适体对它的靶标的“特异性结合亲和力”意指适体结合它的靶标的亲和力程度通常比它结合测试样品中的其它组分高得多。“适体”是一种类型或种类的具有特定核苷酸序列的核酸分子的一组拷贝。适体可包括任何适合数目的核苷酸,包括任何数目的化学修饰的核苷酸。“各适体”是指超过一组所述分子。不同适体可具有相同或不同数目的核苷酸。适体可为 DNA 或 RNA 或化学修饰的核酸,并且可为单链、双链,或含有双链区域,并且可包括更高级结构。适体也可作为光适体,其中光反应性或化学反应性官能团被包括在适体中以允许它共价连接于它的相应靶标。任何本文公开的适体方法都可包括使用两种或更多种特异性结合相同靶标分子的适体。如以下进一步所述,适体可包括标签。如果适体包括标签,那么适体的所有拷贝无需具有相同标签。此外,如果不同适体各自包括标签,那么这些不同适体可具有相同标签或不同标签。

[0132] 可使用任何已知方法,包括SELEX方法来鉴定适体。一旦鉴定,即可根据任何已知方法,包括化学合成方法和酶促合成方法来制备或合成适体。

[0133] 术语“SELEX”和“SELEX方法”在本文中可互换用于大体上指代 (1) 选择以合乎需要的方式(例如以高亲和力与蛋白质结合)与靶标分子相互作用的适体,与(2)扩增那些所选核酸的组合。SELEX方法可用于鉴定对特定靶标或生物标记具有高亲和力的适体。

[0134] SELEX通常包括制备核酸的候选混合物,使候选混合物与所需靶标分子结合以形成亲和复合物,使亲和复合物与未结合的候选核酸分离,自亲和复合物分离和离析核酸,纯化核酸,以及鉴定特定适体序列。方法可包括多轮以进一步细化所选适体的亲和力。方法可包括在方法中的一个或多个时点进行扩增步骤。参见例如题为“Nucleic Acid Ligands”的美国专利号5,475,096。SELEX方法可用于产生共价结合它的靶标的适体以及非共价结合它的靶标的适体。参见例如题为“Systematic Evolution of Nucleic Acid Ligands by Exponential Enrichment:Chemi-SELEX”的美国专利号5,705,337。

[0135] SELEX方法可用于鉴定含有赋予适体改进的特征(例如像改进的体内稳定性或改进的递送特征)的修饰的核苷酸的高亲和力适体。所述修饰的实例包括在核糖和/或磷酸和/或碱基位置处的化学取代。SELEX方法鉴定的含有修饰的核苷酸的适体描述于题为“High Affinity Nucleic Acid Ligands Containing Modified Nucleotides”的美国专利号5,660,985中,所述专利描述含有在嘧啶的5'位置和2'位置处加以化学修饰的核苷酸衍生物的寡核苷酸。美国专利号5,580,737(参见上文)描述含有一个或多个用2'-氨基(2'-NH₂)、2'-氟代(2'-F)和/或2'-O-甲基(2'-OMe)修饰的核苷酸的高度特异性适体。也参见题为“SELEX and PHOTOSELEX”的美国专利申请公布号2009/0098549,其描述具有扩展的物理和化学性质的核酸文库和它们在SELEX和photoSELEX中的用途。

[0136] SELEX也可用于鉴定具有合乎需要的解离速率特征的适体。参见题为“Method for Generating Aptamers with Improved Off-Rates”的美国公布号US 2009/0004667,其描述用于产生可结合靶标分子的适体的改进SELEX方法。描述了用于产生自它们的相应靶标分子的解离速率较慢的适体和光适体的方法。方法涉及使候选混合物与靶标分子接触,使核酸-靶标复合物的形成发生,以及进行缓慢解离速率富集方法,其中具有快速解离速率的核酸-靶标复合物将解离并且不再形成,而具有缓慢解离速率的复合物将保持完整。另外,方法包括在候选核酸混合物的产生中使用修饰的核苷酸以产生具有改进的解离速率性能的适体。非限制性示例性修饰的核苷酸包括例如图11中所示的修饰的嘧啶。在一些实施方案中,适体包含至少一个具有修饰,如碱基修饰的核苷酸。在一些实施方案中,适体包含至少一个具有疏水性修饰,如疏水性碱基修饰,从而允许与靶标蛋白质进行疏水性接触的核苷酸。在一些实施方案中,所述疏水性接触促进由适体进行较大亲和力和/或较缓慢解离速率结合。具有疏水性修饰的非限制性示例性核苷酸显示于图11中。在一些实施方案中,适体包含至少两个、至少三个、至少四个、至少五个、至少六个、至少七个、至少八个、至少九个、或至少10个具有疏水性修饰的核苷酸,其中各疏水性修饰可相同或不同于其它疏水性修饰。在一些实施方案中,适体中的至少一个、至少两个、至少三个、至少四个、至少五个、至少六个、至少七个、至少八个、至少九个、或至少10个疏水性修饰可独立地选自图11中所示的疏水性修饰。

[0137] 在一些实施方案中,缓慢解离速率适体(包括包含至少一个具有疏水性修饰的核

昔酸的适体)具有 ≥ 30 分钟、 ≥ 60 分钟、 ≥ 90 分钟、 ≥ 120 分钟、 ≥ 150 分钟、 ≥ 180 分钟、 ≥ 210 分钟或 ≥ 240 分钟的解离速率($t_{1/2}$)。

[0138] 在一些实施方案中,测定采用包括使得适体能够共价结合或“光交联”它们的靶标分子的光反应性官能团的适体。参见例如题为“Nucleic Acid Ligand Diagnostic Biochip”的美国专利号6,544,776。这些光反应性适体也被称为光适体。参见例如其各自题为“Systematic Evolution of Nucleic Acid Ligands by Exponential Enrichment: Photoselection of Nucleic Acid Ligands and Solution SELEX”的美国专利号5,763,177、美国专利号6,001,577和美国专利号6,291,184;也参见例如题为“Photoselection of Nucleic Acid Ligands”的美国专利号 6,458,539。在使微阵列芯片与样品接触,并且光适体已具有机会来结合它们的靶标分子之后,使光适体光活化,并且洗涤固体载体以移除任何非特异性结合的分子。可使用严苛洗涤条件,因为由于由光适体上的光活化的官能团产生的共价键,所以结合于光适体的靶标分子通常不被移除。以这个方式,测定使得能够检测对应于测试样品中的生物标记的生物标记水平。

[0139] 在一些测定形式中,适体在与样品接触之前被固定在固体载体上。然而,在某些情况下,在与样品接触之前固定适体可能不提供最优测定。举例来说,预先固定适体可导致适体与靶标分子在固体载体的表面上低效混合,从而也许导致反应时间冗长,并且因此延长用以允许适体高效结合它们的靶标分子的孵育时期。此外,当光适体用于测定中时以及视用作固体载体的材料而定,固体载体可倾向于散射或吸收用于实现在光适体与它们的靶标分子之间形成共价键的光。此外,视所用方法而定,对结合于它们的适体的靶标分子的检测可遭受不精确,因为固体载体的表面也可暴露于使用的任何标记试剂,并且受所述标记试剂影响。最后,使适体固定在固体载体上通常涉及在使适体暴露于样品之前的适体制备步骤(即固定化),并且这个制备步骤可影响适体的活性或功能性。

[0140] 也已描述允许适体在溶液中捕获它的靶标,接着采用被设计以在检测之前移除适体-靶标混合物的特定组分的分离步骤的适体测定(参见题为“Multiplexed Analyses of Test Samples”的美国公布号 2009/0042206)。所述适体测定方法使得能够通过检测和定量核酸(即适体)来检测和定量测试样品中的非核酸靶标(例如蛋白质靶标)。所述方法产生用于检测和定量非核酸靶标的核酸替代物(即适体),由此允许广泛多种核酸技术(包括扩增)应用于更广泛范围的所需靶标,包括蛋白质靶标。

[0141] 适体可被构建以有助于自适体生物标记复合物(或光适体生物标记共价复合物)分离测定组分,并且允许离析适体供检测和/或定量。在一个实施方案中,这些构建体可在适体序列内包括可裂解或可释放元件。在其它实施方案中,其它功能性可被引入适体中,例如标记的组分或可检测组分、间隔子组分、或特定结合标签或固定元件。举例来说,适体可包括通过可裂解部分连接于适体的标签、标记、分隔所述标记的间隔子组分、以及所述可裂解部分。在一个实施方案中,可裂解元件是光可裂解连接子。光可裂解连接子可被连接于生物素部分和间隔子区段,可包括用于使胺衍生的NHS基团,并且可用于将生物素基团引入适体中,由此允许在测定方法中稍后释放适体。

[0142] 以所有测定组分都在溶液中所进行的均质测定不需要在检测信号之前分离样品和试剂。这些方法是快速的,并且易于使用。这些方法基于分子捕获或结合试剂与它的特定靶标反应来产生信号。在本文所述的方法的一些实施方案中,分子捕获试剂包括适体或抗

体等,并且特定靶标可为表3、4、6、7、8和/或9中所示的生物标记。

[0143] 在一些实施方案中,一种用于产生信号的方法利用归因于荧光团标记的捕获试剂与它的特定生物标记靶标的相互作用的各向异性信号变化。当标记的捕获剂与它的靶标反应时,增加的分子量导致连接于复合物的荧光团的旋转运动变得缓慢得多,从而改变各向异性值。通过监测各向异性变化,结合事件可用于定量测量溶液中的生物标记。其它方法包括荧光偏振测定、分子信标方法、时间分辨的荧光淬灭、化学发光、荧光共振能量转移等。

[0144] 可用于检测生物样品中的生物标记水平的示例性溶液基适体测定包括以下:(a) 通过使生物样品与包括第一标签,并且对生物标记具有特定亲和力的适体接触来制备混合物,其中当生物标记存在于样品中时形成适体亲和复合物;(b) 使混合物暴露于包括第一捕获成分的第一固体载体,以及使第一标签与第一捕获成分缔合;(c) 移除混合物的未与第一固体载体缔合的任何组分;(d) 使第二标签连接于适体亲和复合物的生物标记组分;(e) 自第一固体载体释放适体亲和复合物;(f) 使释放的适体亲和复合物暴露于包括第二捕获成分的第二固体载体,以及使第二标签与第二捕获成分缔合;(g) 通过自适体亲和复合物划分非复合适体来自混合物移除任何非复合适体;(h) 自固体载体洗脱适体;以及(i) 通过检测适体亲和复合物的适体组分来检测生物标记。

[0145] 使用适体检测生物样品中的生物标记的非限制性示例性方法描述于实施例7中。也参见Kraemer等,PLoS One 6(10):e26332。

[0146] 使用免疫测定来测定生物标记水平

[0147] 免疫测定方法基于抗体与它的相应靶标或分析物的反应,并且可视特定测定形式而定来检测样品中的分析物。为改进基于免疫反应性的测定方法的特异性和灵敏性,单克隆抗体及其片段由于它们的特异性表位识别而常加以使用。多克隆抗体由于它们相较于单克隆抗体对靶标的亲和力增加而也已成功用于各种免疫测定中。免疫测定已被设计供与广泛范围的生物样品基质一起使用。免疫测定形式已被设计以提供定性、半定量和定量结果。

[0148] 通过使用用已知浓度的待检测的特定分析物创建的标准曲线产生定量结果。将来自未知样品的应答或信号绘制成标准曲线,并且确定对应于未知样品中的靶标的量或水平。

[0149] 已设计众多免疫测定形式。ELISA或EIA可定量用于检测分析物。这个方法依赖于使标记连接于分析物或抗体,并且标记组分直接或间接包括酶。ELISA测试可被形式化以直接、间接、竞争性或夹心式检测分析物。其它方法依赖于例如像放射性同位素(I^{125})或荧光的标记。其它技术包括例如凝集、浊度测定法、比浊法、蛋白质印迹、免疫沉淀、免疫细胞化学分析、免疫组织化学分析、流式细胞术、Luminex测定等(参见ImmunoAssay:A Practical Guide,由Brian Law 编辑,由Taylor&Francis,Ltd.出版,2005版)。

[0150] 示例性测定形式包括酶联免疫吸附测定(ELISA)、放射免疫测定、荧光、化学发光和荧光共振能量转移(FRET)或时间分辨的FRET(TR-FRET)免疫测定。用于检测生物标记的程序的实例包括生物标记免疫沉淀继之以允许尺寸和肽水平辨别的计量方法,如凝胶电泳、毛细管电泳、平面电色谱法等。

[0151] 检测和/或定量可检测标记或信号产生物质的方法取决于标记的性质。由适当酶催化的反应(其中可检测标记是酶;参见上文)的产物可具有不限于荧光性、发光性或放射性,或它们可吸收可见光或紫外光。适于检测所述可检测标记的检测器的实例包括不限于x

射线胶片、放射性计数器、闪烁计数器、分光光度计、色度计、荧光计、光度计和比重计。

[0152] 任何检测方法都可以允许对反应进行任何适合制备、处理和分析的任何形式来执行。这可例如呈多孔测定板(例如96孔或386孔)形式,或使用任何适合阵列芯片或微阵列芯片。可手动或用机器人制备各种试剂的储备溶液,并且所有随后吸移、稀释、混合、分配、洗涤、孵育、样品读出、数据收集和分析都可使用可商购获得的分析软件、机器人和能够检测可检测标记的检测仪器,用机器人进行。

[0153] 使用基因表达谱分析来测定生物标记水平

[0154] 在一些实施方案中,测量生物样品中的mRNA可用作检测生物样品中的相应蛋白质的水平的替代。因此,在一些实施方案中,本文所述的生物标记或生物标记组可通过检测适当RNA来检测。

[0155] 在一些实施方案中,通过反转录定量聚合酶链反应(RT-PCR,继之以qPCR)测量mRNA表达水平。RT-PCR用于自mRNA产生cDNA。cDNA可在qPCR测定中用于随DNA扩增过程进展而产生荧光。通过与标准曲线比较,qPCR可产生绝对测量结果,如每个细胞的mRNA的拷贝数。RNA印迹、微阵列芯片、Invader测定以及RT-PCR与毛细管电泳组合已全部用于测量样品中的mRNA的表达水平。参见Gene Expression Profiling:Methods and Protocols, Richard A. Shimkets编,Humana Press,2004。

[0156] 使用体内分子成像技术来检测生物标记

[0157] 在一些实施方案中,本文所述的生物标记可用于分子成像测试中。举例来说,可使成像剂偶联于可用于在体内检测生物标记的捕获试剂。

[0158] 体内成像技术提供用于确定个体的体内的特定疾病的状态的非侵袭性方法。举例来说,身体的整个部分或甚至整个身体可被视为三维图像,由此提供关于体内形态和结构的有价值信息。所述技术可与本文所述的生物标记的检测组合以提供关于体内生物标记的信息。

[0159] 由于各种技术进步,对体内分子成像技术的使用正在扩大。这些进步包括开发出可在体内提供强烈信号的新型造影剂或标记,如放射性标记和/或荧光标记;以及开发出可以足以提供有用信息的灵敏性和准确度在体外检测和分析这些信号的强力新型成像技术。可在适当成像系统中观察造影剂,由此提供身体的一个或多个其中定位有造影剂的部分的图像。可使造影剂与捕获试剂(例如像适体或抗体),和/或与肽或蛋白质、或寡核苷酸(例如用于检测基因表达)、或含有这些中的任一个与一种或多种大分子和/或其它颗粒形式的复合物结合或缔合。

[0160] 造影剂的特征也可适用于成像的放射性原子。适合放射性原子包括用于闪烁照相研究的镓-99m或碘-123。其它可易于检测的部分包括例如用于磁共振成像(MRI)的自旋标记,例如像再次碘-123、碘-131、铟-111、氟-19、碳-13、氮-15、氧-17、钆、锰或铁。所述标记在本领域中是熟知的,并且可由本领域普通技术人员容易地选择。

[0161] 标准成像技术包括但不限于磁共振成像、计算机断层摄影术扫描、正电子发射断层摄影术(PET)、单光子发射计算机断层摄影术(SPECT)等。对于诊断性体内成像,可用的检测仪器的类型是选择给定造影剂(如给定放射性核素)和用于靶向的特定生物标记(蛋白质、mRNA等)的主要因素。所选放射性核素通常具有某一类型的衰减,其可通过给定类型的仪器检测。此外,当选择用于体内诊断的放射性核素时,它的半衰期应足够长以使得能够在

由靶标组织进行最大摄取时检测,但应足够短暂以使对宿主的有害辐射得以最小。

[0162] 示例性成像技术包括但不限于PET和SPECT,其是其中向个体综合或局部施用放射性核素的成像技术。随时间测量放射性示踪剂的随后摄取,并且用于获得关于靶向的组织 and 生物标记的信息。由于所用特定同位素的高能(γ 射线)发射以及用于检测它们的仪器的灵敏性和精密化,所以可在体外推断放射性的二维分布。

[0163] PET中通常使用的正电子发射核素包括例如碳-11、氮-13、氧-15 和氟-18。通过电子捕获和/或 γ 发射来衰减的同位素用于SPECT中,并且包括例如碘-123和锝-99m。一种用于用锝-99m标记氨基酸的示例性方法是在螯合前体存在下还原高锝酸根离子以形成不稳定锝-99m-前体复合物,其转而与双官能修饰的趋化性肽的金属结合基团反应以形成锝-99m-趋化性肽缀合物。

[0164] 抗体常用于所述体内成像诊断方法。用于体内诊断的抗体的制备和使用在本领域中是熟知的。类似地,适体可用于所述体内成像诊断方法。举例来说,用于鉴定本文所述的特定生物标记的适体可被适当标记,并且注射至个体中以在体内检测生物标记。如先前所述,将根据待使用的成像模态选择所用标记。关于组织渗透、组织分布、动力学、消除、效价和选择性,适体定向成像剂相较于其它成像剂可具有独特和有利特征。

[0165] 所述技术也可任选用例如用于通过用反义寡核苷酸进行成像来检测基因表达的标记的寡核苷酸来执行。这些方法用于例如与作为标记的荧光分子或放射性核素的原位杂交。用于检测基因表达的其它方法包括例如检测报道体基因的活性。

[0166] 另一一般类型的成像技术是光学成像,其中受试者内的荧光信号由在受试者外部的光学装置检测。这些信号可归因于实际荧光和/或生物发光。光学检测装置的灵敏性的改进已使光学成像用于体内诊断测定的适用性增加。

[0167] 对于其它技术的综述,参见N.Blow,Nature Methods,6,465-469, 2009。

[0168] 使用组织学/细胞学方法来测定生物标记

[0169] 在一些实施方案中,可使用组织学或细胞学方法检测多种组织样品中的本文所述的生物标记。举例来说,支气管内和经支气管活检、细针吸出、切割针和核心活检可用于组织学分析。支气管洗涤和刷拂、胸膜吸出和痰可用于细胞学分析。本文鉴定的任何生物标记都可用于染色作为疾病的指示的试样。

[0170] 在一些实施方案中,一种或多种对相应生物标记具有特异性的捕获试剂用于对样品的细胞学评估,并且可包括以下中的一个或多个:收集细胞样品,固定所述细胞样品,脱水,清洁,固定所述细胞样品于显微镜载片上,使所述细胞样品透化,处理以达成分析物摄取,染色,脱色,洗涤,阻断,以及在缓冲溶液中与一种或多种捕获试剂反应。在另一实施方案中,细胞样品由细胞块产生。

[0171] 在一些实施方案中,一种或多种对相应生物标记具有特异性的捕获试剂用于对组织样品的组织学评估,并且可包括以下中的一个或多个:收集组织试样,固定组织样品,脱水,清洁,固定组织样品于显微镜载片上,使组织样品透化,处理以达成分析物摄取,染色,脱色,洗涤,阻断,再水合,以及在缓冲溶液中与捕获试剂反应。在另一实施方案中,固定和脱水被冷冻替换。

[0172] 在另一实施方案中,使一种或多种对相应生物标记具有特异性的适体与组织学或细胞学样品反应,并且可充当核酸扩增方法中的核酸靶标。适合核酸扩增方法包括例如

PCR、q-β复制酶、滚环扩增、链置换、解旋酶依赖性扩增、环介导的等温扩增、连接酶链反应、以及限制和环化辅助的滚环扩增。

[0173] 在一个实施方案中,在可包括以下中的任一个的缓冲溶液中混合一种或多种用于组织学或细胞学评估的对相应生物标记具有特异性的捕获试剂:阻断物质、竞争剂、去污剂、稳定剂、载体核酸、聚阴离子物质等。

[0174] “细胞学方案”通常包括样品收集、样品固定、样品固定化和染色。“细胞制备”可包括在样品收集之后的若干处理步骤,包括使用一种或多种用于染色制备的细胞的适体。

[0175] 使用质谱测定方法来测定生物标记水平

[0176] 多种配置的质谱仪可用于检测生物标记水平。若干类型的质谱仪是可用的,或可以各种配置来产生。一般来说,质谱仪具有以下主要组件:样品进口、离子源、质量分析器、检测器、真空系统、以及仪器控制系统和数据系统。样品进口、离子源和质量分析器的差异通常确定仪器的类型和它的性能。举例来说,进口可为毛细管柱液相色谱源,或可为如用于基质辅助的激光解吸中的直接探针或阶段。常见离子源是例如电喷雾(包括纳米喷雾和微米喷雾)或基质辅助的激光解吸。常见质量分析器包括四极滤质器、离子阱质量分析器和飞行时间质量分析器。其它质谱测定方法在本领域中是熟知的(参见 Burlingame等 Anal.Chem.70:647R-716R(1998);Kinter和Sherman, New York(2000))。

[0177] 可通过以下中的任一个来检测和测量蛋白质生物标记和生物标记水平:电喷雾离子化质谱测定法(ESI-MS)、ESI-MS/MS、ESI-MS/(MS)_n、基质辅助的激光解吸离子化飞行时间质谱测定法(MALDI-TOF-MS)、表面增强的激光解吸/离子化飞行时间质谱测定法(SELDI-TOF-MS)、硅上解吸/离子化(DIOS)、次级离子质谱测定法(SIMS)、四极飞行时间(Q-TOF)、称为ultraflex III TOF/TOF的串联飞行时间(TOF/TOF)技术、大气压化学离子化质谱测定法(APCI-MS)、APCI-MS/MS、APCI-(MS)^N、大气压光致电离质谱测定法(APPI-MS)、APPI-MS/MS和APPI-(MS)^N、四极质谱测定法、傅里叶变换质谱测定法(FTMS)、定量质谱测定法和离子阱质谱测定法。

[0178] 样品制备策略用于在质谱表征蛋白质生物标记以及测定生物标记水平之前标记和富集样品。标记方法包括但不限于用于相对和绝对定量的同量异序标签(iTRAQ)和在细胞培养物中用氨基酸进行稳定同位素标记(SILAC)。用于在质谱分析之前选择性富集样品中的候选生物标记蛋白质的捕获试剂包括但不限于适体、抗体、核酸探针、嵌合体、小分子、F(ab')₂片段、单链抗体片段、Fv片段、单链Fv片段、核酸、凝集素、配体结合受体、亲和体、纳米体、锚蛋白、结构域抗体、替代性抗体骨架(例如双链抗体等)、印迹聚合物、高亲和性多聚体、肽模拟物、类肽、肽核酸、苏糖核酸、激素受体、细胞因子受体和合成受体以及这些捕获试剂的修饰形式和片段。

[0179] 前述测定使得能够检测适用于本文所述的方法中的生物标记水平,其中所述方法包括检测来自个体的生物样品中的至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、至少八种或至少九种选自表3、4、6、7、8和9中的生物标记的生物标记。在各种实施方案中,方法包括检测一种或多种选自本文所述的生物标记组(如表5中所示的组以及表3、4、6、7、8和9中所示的生物标记子组)中的任一个的生物标记的水平。因此,尽管一些所述生物标记可单独适用于检测NAFLD和/或NASH,但本文也描述用于将多种生物标记以及生物标记子组的编组以形成具有两种或更多种生物标记的组的方法。根据任何本文所述

的方法,可个别地检测和分类生物标记水平,或可集合检测和分类它们,如例如在多路测定形式中。

[0180] 分类生物标记以及计算疾病计分

[0181] 在一些实施方案中,给定诊断测试的生物标记“标志”含有一组生物标记,各生物标记在目标群体中具有特征水平。在一些实施方案中,特征水平可指特定组中个体的生物标记水平的均值或平均。在一些实施方案中,本文所述的诊断方法可用于将来自个体的未知样品指定至两个组中的一个 (NAFLD或正常) 中。在一些实施方案中,本文所述的诊断方法可用于将来自个体的未知样品指定至两个组中的一个 (NASH或NAFLD) 中。在一些实施方案中,本文所述的诊断方法可用于将来自个体的未知样品指定至以下三个组中的一个中:正常、NAFLD无NASH、以及NASH。

[0182] 将样品指定至两个或更多个组中的一个中被称为分类,并且用于实现这个指定的程序被称为分类器或分类方法。分类方法也可被称为计分方法。存在可用于由一组生物标记水平构建诊断分类器的许多分类方法。在一些情况下,使用监督学习技术来执行分类方法,其中使用自希望区分的两个(或更多个(对于多个分类状态))不同组内的个体获得的样品收集数据集。因为事先已知各样品的由各样品所属的类别(组或群体),所以分类方法可被训练以给与所需分类应答。也有可能使用未监督学习技术以产生诊断分类器。

[0183] 用于产生诊断分类器的常用方法包括决策树;装袋+提升+森林;基于规则推断的学习;帕曾窗口(Parzen Window);线性模型;逻辑;神经网络方法;未监督群集;K-均值;分层上升/下降;半监督学习;原型方法;最近邻;核密度估计;支持向量机;隐藏马尔可夫模型(hidden Markov model);波耳兹曼(Boltzmann)学习;并且分类器可简单地或以使特定目标函数最小的方式组合。对于综述,参见例如 Pattern Classification,R.O.Duda,等编,John Wiley&Sons,第2版,2001;也参见The Elements of Statistical Learning-Data Mining, Inference,and Prediction,T.Hastie,等编,Springer Science+Business Media,LLC,第2版,2009。

[0184] 为使用监督学习技术产生分类器,获得一组称为训练数据的样品。在诊断测试的情形下,训练数据包括来自未知样品将稍后被指定的不同组(类别)的样品。举例来说,自对照群体中的个体和特定疾病群体中的个体收集的样品可构成训练数据以产生可将未知样品(或更具体来说,自其获得样品的个体)分类为患有疾病或无疾病的分类器。由训练数据产生分类器被称为训练分类器。关于分类器训练的特定细节取决于监督学习技术的性质。训练首次用于实验的贝叶斯(Bayesian)分类器是所述监督学习技术的实例(参见例如 Pattern Classification,R.O.Duda,等编,John Wiley&Sons,第2版,2001;也参见The Elements of Statistical Learning-Data Mining,Inference,and Prediction,T.Hastie等编,Springer Science+Business Media,LLC,第2版,2009)。首次用于实验的贝叶斯分类器的训练例如描述于美国公布号:2012/0101002和2012/0077695中。

[0185] 因为通常存在许多比训练集中的样品更潜在的生物标记水平,所以必须小心以避免过度拟合。当统计模型描述随机误差或噪声而非潜伏关系时,发生过度拟合。可以多种方式避免过度拟合,包括例如通过限制在产生分类器时使用的生物标记的数目,通过假定生物标记应答不依赖于彼此,通过限制所用潜伏统计模型的复杂性,以及通过确保潜伏统计模型符合数据。

[0186] 使用一组生物标记来开发诊断测试的说明性实例包括在严格独立处理生物标记下应用首次用于实验的贝叶斯分类器,即一种基于贝叶斯定理的简单概率性分类器。各生物标记通过各类别中测量的RFU 值或log RFU(相对荧光单位)值的类别依赖性概率密度函数(pdf) 来描述。一个类别中的生物标记组的联合pdf被假定为各生物标记的个别类别依赖性pdf的乘积。在这个情形下训练首次用于实验的贝叶斯分类器相当于指定参数(“参数化”)以表征类别依赖性pdf。可使用类别依赖性pdf的任何潜伏模型,但模型应大体上符合在训练集中观察的数据。

[0187] 首次用于实验的贝叶斯分类器的性能取决于用于构建和训练分类器的生物标记的数目和质量。单一生物标记将根据它的KS距离(柯尔莫哥罗夫-斯米尔诺夫(Kolmogorov-Smirnov))来表现。如果随后添加的生物标记不依赖第一生物标记,那么添加具有良好KS距离(例如 >0.3)的随后生物标记将通常改进分类性能。使用灵敏性加特异性作为分类器计分,可用贪心算法的变化形式产生许多高分分类器。(贪心算法是在各阶段遵循进行局部最优选择的问题解决性元启发式算法,希望发现整体最优解的任何算法。)

[0188] 用以描绘分类器性能的另一方式通过接受者操作特征(ROC),或简称为ROC曲线或ROC图来实现。ROC是二元分类器系统当它的辨别阈值改变时的灵敏性或真阳性率相对于假阳性率(1-特异性或1-真阴性率)的图示图。ROC也可通过相对于阴性中的假阳性的分数(FPR=假阳性率)绘制阳性中的真阳性的分数(TPR=真阳性率)来等效表示。也称为相对操作特征曲线,因为它是当准则变化时两种操作特征(TPR和FPR)的比较。ROC曲线下面积(AUC)通常用作诊断准确度的概述量度。它可采用0.0至1.0的值。AUC具有重要统计性质:分类器的AUC等效于分类器将使随机选择的阳性情况分级为高于随机选择的阴性情况的概率(Fawcett T, 2006. An introduction to ROC analysis. Pattern Recognition Letters. 27: 861-874)。这等效于威尔卡森(Wilcoxon)秩检验(Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143, 29-36.)。

[0189] 示例性实施方案以各种组合使用表3、4、6、7、8和9中所列的任何数目的生物标记以产生用于鉴定患有NAFLD的个体的诊断测试。表3、4、6、7、8和9中所列的生物标记可以许多方式组合以产生分类器。在一些实施方案中,视所选特定诊断性能准则而定,生物标记组包含不同组生物标记。举例来说,生物标记的某些组合可产生比其它组合更灵敏(或更具特异性)的测试。在一些实施方案中,用于鉴定患有NAFLD的个体的生物标记组选自表5中的组。

[0190] 示例性实施方案以各种组合使用表3、4、6、7、8和9中所列的任何数目的生物标记以产生用于鉴定患有脂肪变性的个体的诊断测试。表3、4、6、7、8和9中所列的生物标记可以许多方式组合以产生分类器。在一些实施方案中,视所选特定诊断性能准则而定,生物标记组包含不同组生物标记。举例来说,生物标记的某些组合可产生比其它组合更灵敏(或更具特异性)的测试。在一些实施方案中,用于鉴定患有脂肪变性的个体的生物标记组选自表5中的组。在一些实施方案中,用于鉴定患有脂肪变性的个体的生物标记组包含表3中的生物标记。在一些实施方案中,用于鉴定患有脂肪变性的个体的生物标记组包含表8中的生物标记。

[0191] 示例性实施方案以各种组合使用表3、4、6、7、8和9中所列的任何数目的生物标记

以产生用于鉴定患有NASH的个体的诊断测试。表3、4、6、7、8和9中所列的生物标记可以许多方式组合以产生分类器。在一些实施方案中,视所选特定诊断性能准则而定,生物标记组包含不同组生物标记。举例来说,生物标记的某些组合可产生比其它组合更灵敏(或更具特异性)的测试。在一些实施方案中,用于鉴定患有NASH的个体的生物标记组选自表5中的组。在一些实施方案中,用于鉴定患有NASH的个体的生物标记组包含表4中的生物标记。在一些实施方案中,用于鉴定患有NASH的个体的生物标记组包含表9中的生物标记。

[0192] 示例性实施方案以各种组合使用表3、4、6、7、8和9中所列的任何数目的生物标记以产生用于鉴定患有NAFLD、脂肪变性和/或 NASH的个体的诊断测试。表3、4、6、7、8和9中所列的生物标记可以许多方式组合以产生分类器。在一些实施方案中,视所选特定诊断性能准则而定,生物标记组包含不同组生物标记。举例来说,生物标记的某些组合可产生比其它组合更灵敏(或更具特异性)的测试。在一些实施方案中,用于鉴定患有NAFLD、脂肪变性和/或NASH的个体的生物标记组选自表5中的组。

[0193] 在一些实施方案中,一旦确定某一组包括来自表3、4、6、7、8 和9的一组特定生物标记,并且由一组训练数据构建分类器,诊断测试参数即完全。在一些实施方案中,在一个或多个测定中操作生物样品以产生用于分类的相关定量生物标记水平。测量的生物标记水平用作输出分类和样品的反映类别指定的置信度的任选计分的分类方法的输入。

[0194] 在一些实施方案中,任选稀释生物样品,并且在多路适体测定中操作,且如下估计数据。首先,任选标准化和校正来自测定的数据,并且所得生物标记水平用作贝叶斯分类流程的输入。其次,计算各个别测量的生物标记的对数似然比,接着求和以产生最终分类计分,其也被称为诊断计分。可报告所得指定以及总体分类计分。在一些实施方案中,也可报告对各生物标记水平计算的个别对数似然风险因数。

[0195] 试剂盒

[0196] 可使用如用于执行本文公开的方法的适合试剂盒来检测本文所述的生物标记的任何组合。此外,任何试剂盒都可含有一种或多种如本文所述的可检测标记,如荧光部分等。

[0197] 在一些实施方案中,试剂盒包括(a)一种或多种用于检测生物样品中的一种或多种生物标记的捕获试剂(例如像至少一种适体或抗体),以及任选(b)一种或多种用于预测自其获得生物样品的个体是否患有NAFLD、脂肪变性和/或NASH(如1、2、3或4期NASH或 2、3或4期NASH)的软件或计算机程序产品。或者,并非一种或多种计算机程序产品,可提供一种或多种供人手动进行以上步骤的说明书。

[0198] 在一些实施方案中,试剂盒包括固体载体、捕获试剂和信号产生物质。试剂盒也可包括用于使用装置和试剂、处理样品、以及分析数据的说明书。此外,试剂盒可与计算机系统或软件一起用于分析和报告生物样品的分析结果。

[0199] 试剂盒也可含有一种或多种用于处理生物样品的试剂(例如溶解缓冲剂、去污剂、洗涤剂或缓冲剂)。任何本文所述的试剂盒都也可包括例如缓冲剂、阻断剂、质谱测定法基质材料、抗体捕获剂、阳性对照样品、阴性对照样品、软件和信息,如方案、指导和参照数据。

[0200] 在一些实施方案中,提供用于分析NAFLD和/或NASH的试剂盒,其中所述试剂盒包括针对一种或多种本文所述的生物标记的PCR 引物。在一些实施方案中,试剂盒可进一步包括使用以及关联生物标记与NAFLD和/或NASH预后的说明书。在一些实施方案中,试剂盒

可包括用于扩增或离析样品DNA的含有一种或多种本文所述的生物标记的互补序列的DNA阵列芯片、试剂和/或酶。试剂盒可包括用于实时PCR的试剂,例如TaqMan探针和/或引物以及酶。

[0201] 举例来说,试剂盒可包括(a)试剂,其包括至少一种用于测定测试样品中的一种或多种生物标记的水平捕获试剂,以及任选(b)一种或多种用于进行将测试样品中定量的各生物标记的量与一种或多种预定截断值进行比较的步骤的算法或计算机程序。在一些实施方案中,算法或计算机程序基于所述比较对定量的各生物标记指定计分,并且在一些实施方案中,组合定量的各生物标记的指定计分以获得总计分。此外,在一些实施方案中,算法或计算机程序将总计分与预定计分进行比较,并且使用比较来确定个体是否患有NAFLD、脂肪变性和/或NASH。或者,并非一种或多种算法或计算机程序,可提供一种或多种供人手动进行以上步骤的说明书。

[0202] 计算机方法和软件

[0203] 一旦选择生物标记或生物标记组,一种用于估计个体的NAFLD的方法可包括以下:1)收集或另外获得生物样品;2)执行分析方法以检测和测量所述生物样品中的所述生物标记或所述组中的生物标记;以及3)报告生物标记水平的结果。在一些实施方案中,生物标记水平的结果被定性而非定量报告,例如像提出的诊断(“NAFLD”、“脂肪变性”、“NASH”、“2、3或4期NASH”等)或简单地是其中定义“阳性”和“阴性”的阳性/阴性结果。在一些实施方案中,一种用于估计个体的NAFLD的方法可包括以下:1)收集或另外获得生物样品;2)执行分析方法以检测和测量所述生物样品中的所述生物标记或所述组中的生物标记;3)进行任何数据规范化或标准化;4)计算各生物标记水平;以及5)报告生物标记水平的结果。在一些实施方案中,以某一方式组合生物标记水平,并且报告组合的生物标记水平的单一值。在这个方法中,在一些实施方案中,报告的值可为由所有生物标记计算结果的总和确定的单一数值,其与是存在或不存在疾病的指示的预设置阈值进行比较。或者,诊断计分可为一系列各自代表生物标记值的棒条,并且可将应答的样式与预设置样式进行比较以确定存在或不存在疾病。

[0204] 本文所述的方法的至少一些实施方案可使用计算机来实施。计算机系统100的实例显示于图9中。参照图9,显示系统100,其包括通过总线108电耦合的硬件元件,包括处理器101、输入装置102、输出装置103、存储装置104、计算机可读存储介质读取器105a、通信系统106、处理加速(例如DSP或特殊目的处理器)107和存储器109。计算机可读存储介质读取器105a被进一步耦合于计算机可读存储介质105b,所述组合全面地代表远程、本地、固定和/或可移动存储装置加用于暂时和/或更永久含有计算机可读信息的存储介质、存储器等,其可包括存储装置104、存储器109和/或任何其它所述可访问系统100资源。系统100也包括软件元件(显示为当前位于工作存储器191内),包括操作系统192和其它代码193,如程序、数据等。

[0205] 关于图9,系统100具有广泛灵活性和可配置性。因此,举例来说,单一构造可能用于实施一个或多个可进一步根据当前合乎需要的方案、方案变化形式、延伸形式等配置的服务器。然而,将为本领域技术人员显而易见的是实施方案很可能根据更特定应用要求加以利用。举例来说,一个或多个系统元件可能以系统100组件内(例如通信系统106内)的子元件形式被实施。定制硬件也可能加以利用和/或特定元件可能以硬件、软件或两者形式

被实施。此外,尽管可采用与其它计算装置,如网络输入/输出装置(未显示)的连接,但应了解也可能利用与其它计算装置的有线、无线、调制解调器和/或其它一种或多种连接。

[0206] 在一个方面,系统可包括含有表征NAFLD和/或NASH的生物标记的特征的数据库。生物标记数据(或生物标记信息)可用作向计算机的输入以用作计算机实施的方法的一部分。生物标记数据可包括如本文所述的数据。

[0207] 在一个方面,系统进一步包括一个或多个用于向一个或多个处理器提供输入数据的装置。

[0208] 系统进一步包括用于存储分级数据元的数据集的存储器。

[0209] 在另一方面,用于提供输入数据的装置包括用于检测数据元的特征的检测器,例如像质谱仪或基因芯片读取器。

[0210] 系统另外可包括数据库管理系统。用户请求或查询可以由处理查询以自训练集的数据库提取相关信息的数据库管理系统理解的适当语言加以格式化。

[0211] 系统可连接于网络服务器和一个或多个客户端与其连接的网络。网络可为如本领域中所周知的局域网(LAN)或广域网(WAN)。优选地,服务器包括为运行计算机程序产品(例如软件)以访问数据库数据来处理用户请求所必需的硬件。

[0212] 系统可包括用于执行来自数据库管理系统的指令的操作系统(例如UNIX[®]或Linux)。在一个方面,操作系统可在全球通信网络(如因特网)上操作,并且利用全球通信网络服务器来连接于所述网络。

[0213] 系统可包括一个或多个包括图形显示器界面的装置,所述界面包括如常规见于本领域中已知的图形用户界面中的界面元件,如按钮、下拉菜单、滚动条、用于输入文本的区域等。在用户界面上输入的请求可被传输至系统中的应用程序中以进行格式化来在一个或多个系统数据库中搜索相关信息。由用户输入的请求或查询可以任何适合数据库语言来构建。

[0214] 图形用户界面可通过图形用户界面代码来产生作为操作系统的一部分,并且可用于输入数据和/或显示输入数据。处理的数据的结果可显示在界面中,在与系统通信的打印机上打印,保存在存储装置中,和/或经网络传输,或可以计算机可读介质的形式提供。

[0215] 系统可与用于向系统提供关于数据元的数据(例如表达值)的输入装置通信。在一个方面,输入装置可包括基因表达谱分析系统,包括例如质谱仪、基因芯片或阵列芯片读取器等。

[0216] 根据各种实施方案的用于分析生物标记信息的方法和器具可以任何适合方式,例如使用在计算机系统上操作的计算机程序来实施。可使用包括处理器和随机存取存储器的常规计算机系统,如可远程访问的应用服务器、网络服务器、个人计算机或工作站。其它计算机系统组件可包括存储装置或信息存储系统(如大容量存储系统)和用户接口(例如常规监视器、键盘和追踪装置)。计算机系统可为独立系统,或可为包括服务器和一个或多个数据库的计算机网络的一部分。

[0217] 生物标记分析系统可提供用以完成数据分析的功能和操作,如数据采集、处理、分析、报告和/或诊断。举例来说,在一个实施方案中,计算机系统可执行可接收、存储、搜索、分析和报告与生物标记相关的信息的计算机程序。计算机程序可包括执行各种功能或操作的多个模块,如用于处理原始数据和产生补充性数据的处理模块以及用于分析原始数据和

补充性数据以产生疾病状态和/或诊断的分析模块。鉴定NAFLD、脂肪变性和/或NASH可包括产生或收集任何其它信息(包括其它生物医学信息),关于疾病考虑个体的状况,鉴定进一步测试是否可为合乎需要的,或另外评估个体的健康状态。

[0218] 本文所述的一些实施方案可被实施以包括计算机程序产品。计算机程序产品可包括具有计算机可读程序代码的计算机可读介质,所述代码包括在所述介质中,用于使应用程序在具有数据库的计算机上执行。

[0219] 如本文所用,“计算机程序产品”是指呈自然或程序设计语言语句形式的一组组织化指令,其含于任何性质(例如书面、电子、磁性、光学或另外性质)的物理介质上,并且可与计算机或其它自动数据处理系统一起使用。所述程序设计语言语句在由计算机或数据处理系统执行时使所述计算机或数据处理系统根据语句的特定内容起作用。计算机程序产品包括但不限于:嵌入计算机可读介质中的呈源代码和目标代码和/或测试或数据文库形式的程序。此外,使得计算机系统或数据处理设备装置能够以预先选择的方式起作用的计算机程序产品可以许多形式提供,所述形式包括但不限于原始源代码、汇编代码、目标代码、机器语言、前述各物的加密或压缩形式以及任何和所有等效物。

[0220] 在一个方面,提供用于指示个体是否患有NAFLD,个体是否患有脂肪变性,和/或个体是否患有NASH(如1、2、3或4期NASH 或2、3或4期NASH)的计算机程序产品。计算机程序产品包括计算机可读介质,其包括可由计算装置或系统的处理器执行的程序代码,所述程序代码包括:撷取归于来自个体的生物样品的数据的代码,其中所述数据包括对应于一种或多种本文所述的生物标记的生物标记水平;以及执行指示个体的随生物标记水平而变的NAFLD、脂肪变性和/或NASH状态的分类方法的代码。

[0221] 尽管各种实施方案已描述为方法或器具,但应了解实施方案可通过与计算机耦合的代码来实施,所述代码例如驻留在计算机上或可由计算机访问的代码。举例来说,软件和数据库可用于实施许多以上讨论的方法。因此,除通过硬件实现的实施方案之外,也应注意这些实施方案可通过使用包括其中包括有计算机可读程序代码的计算机可用介质的制品来实现,所述代码使得能够实现本描述中公开的功能。因此,需要的是实施方案也被视为在它们的程序代码部件方面也受本专利保护。此外,实施方案可以存储在实际上任何种类的计算机可读存储器中的代码形式来体现,所述存储器包括但不限于RAM、ROM、磁性介质、光学介质或磁光介质。甚至更通常地,实施方案可以软件或硬件或其任何组合形式来实施,包括但不限于在通用处理器上运行的软件、微代码、可编程逻辑阵列(PLA)或专用集成电路(ASIC)。

[0222] 也设想实施方案可以包括在载波中的计算机信号以及通过传输介质传播的信号(例如电学和光学)形式来实现。因此,以上讨论的各种类型的信息可以某一结构(如数据结构)加以格式化,并且通过传输介质以电信号形式传输或存储在计算机可读介质上。

[0223] 治疗方法

[0224] 在一些实施方案中,在确定受试者患有NAFLD、脂肪变性或 NASH之后,所述受试者经受治疗方案以延迟或防止疾病恶化。针对 NAFLD、脂肪变性和/或NASH的非限制性示例性治疗方案包括减轻重量、控制血糖和避免酒精。在一些实施方案中,给与受试者治疗剂,如匹格列酮、维生素E和/或二甲双胍。参见例如Sanyal等,2010, NEJM, 362:1675-1685。在一些实施方案中,受试者经受胃绕道(或类似)手术例如以加速重量减轻。

[0225] 在一些实施方案中,提供监测NAFLD的方法。在一些实施方案中,在时间0时执行确定受试者是否患有NAFLD的本发明方法。在一些实施方案中,在时间1,以及任选时间2,以及任选时间3等时再次执行方法以监测受试者的NAFLD的进展。在一些实施方案中,视个体的疾病的当前状态而定和/或视疾病据信或被预测进展所处的速率而定,在不同时间点使用不同生物标记。

[0226] 其它方法

[0227] 在一些实施方案中,本文所述的生物标记和方法用于确定医疗保险费和/或人寿保险费。在一些实施方案中,本文所述的方法的结果用于确定医疗保险费和/或人寿保险费。在一些此类情况下,提供医疗保险或人寿保险的机构要求或另外获得关于受试者的NAFLD或 NASH状态的信息,并且使用那个信息来确定所述受试者的适当医疗保险费或人寿保险费。在一些实施方案中,测试由提供医疗保险或人寿保险的机构要求并付款。

[0228] 在一些实施方案中,本文所述的生物标记和方法用于预测和/或管理对医疗资源的利用。在一些此类实施方案中,方法并非用于所述预测的目的而执行,而是由方法获得的信息用于所述预测和/或管理对医疗资源的利用。举例来说,测试机构或医院可汇编许多受试者的由本发明方法获得的信息以预测和/或管理在特定机构或特定地理区域对医疗资源的利用。

实施例

[0229] 以下实施例仅出于说明目的而提供,并且不意图限制如由随附权利要求限定的本申请的范围。以下实施例中所述的常规分子生物学技术可如标准实验室手册如Sambrook等,Molecular Cloning:A Laboratory Manual,第3版,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.,(2001)中所述来执行。

[0230] 实施例1.NAFLD研究受试者

[0231] 用于鉴定生物标记的样品来自Geisinger Health。收集血清样品,并且在576名肥胖患者为减轻重量而经受减肥手术之前对他们进行肝活检。

[0232] 将样品收集在红顶血清管中,并且根据方案加以处理;简要说,使样品在室温下凝结30分钟,接着在1300×g下离心10分钟,并且移除顶层并储存在-80°C下。解冻样品,一次用于等分,并且一次用于测定。

[0233] 为鉴定肝中将区分患有NAFLD的受试者与正常肥胖受试者的生物标记以及将区分患有NASH的受试者与患有脂肪变性的受试者的生物标记,根据肝活检结果,使用Brunt分类方法(Brunt等,2007, Modern Pathol.,20:S40-S48)将研究受试者分成正常、三个程度的肝脂肪变性(轻度、中度和重度脂肪变性)、以及四个时期的NASH。如表1中所示将各组再分类。

[0234] 表1:脂肪变性和NASH时期组的再分类

[0235]

组	脂肪变性	炎症	气球般膨胀	纤维化
肥胖对照	0	0	0	0
轻度脂肪变性	1	0或1	0	0
中度脂肪变性	2	0、1或2	0	0
重度脂肪变性	3	0、1或2	0	0

1期NASH	1、2或3	0、1或2	1	0或1
2期NASH	1、2或3	0、1或2	2	2
3期NASH (桥接)	1、2或3	0、1或2	2	3
4期NASH (肝硬化)	1、2或3	0、1或2	2	4

[0236] 受试者人口统计

[0237] 以上讨论的各组中的个体的某些特征显示于表2中。

[0238] 表2:受试者人口统计

组	女性数	男性数	总计	盲化数	在活检时的年龄, 均值岁数	均值 BMI, kg/m ²	均值 LDL, mg/dl	患有糖尿病%
对照	111	14	125	41	44.59	46.8	107.86	28 (22.4%)
轻度脂肪变性	44	7	51	16	45.75	46.6	112.12	15 (29.4%)
中度脂肪变性	25	7	32	10	44.5	44.77	103.29	12 (37.5%)
重度脂肪变性	45	9	54	17	47.11	47.75	101.11	23 (42.6%)
1期 NASH	93	20	113	37	46.69	47.29	103.37	58(51.3%)
2期 NASH	27	11	38	12	48.52	48.9	109.74	22 (57.9%)
3期 NASH	12	8	20	0	50.05	46.27	101.06	17(85%)
4期 NASH	6	4	10	0	48.7	47.22	101.9	5 (50%)
总计	363	80	443					

[0240] 如表2中所示,确定受试者的年龄、体重指数和LDL水平,并且被认为在所有组之间是平衡的。

[0241] 实施例2.用于生物标记鉴定的多路适体测定

[0242] 通过比较生物标记在病例和对照中的分布来估计以上提及的组中的正常、NAFLD和NASH样品的样品质量,所述分布与样品处理(如剪切力)、细胞溶解和补体活化相关。样品质量是良好的,并且不存在病例对照偏差。

[0243] 多路适体测定用于分析样品和对照以鉴定预测NAFLD和NASH的生物标记。用于这个实验中的多路分析包括适体以在检测下限(1 pM中值)、~7log的动态范围和~5%中值变异系数下检测小样品体积的血液(~65μl血清或血浆)中的1129种蛋白质。多路适体测定例如描述于Gold等(2010)Aptamer-Based Multiplexed Proteomic Technology for Biomarker Discovery.PLoS ONE 5(12):e15004;以及美国公布号:2012/0101002和2012/0077695中。

[0244] 稳定性选择采用许多多数数据子组,并且使用作为正则化逻辑回归模型的套索分类器进行生物标记选择。参见例如Meinshausen等,2010,J.Royal Statistical Soc: Series B(Statistical Methodology),72: 417-473。单一生物标记的选择路径是分摊通过套索模型历经一定范围的λ选择那个生物标记所针对的这些子组。λ是决定多少生物标记

通过套索来选择的调谐参数。历经一定范围的 λ 值的最大选择概率是用于选择一组生物标记的最终量度。

[0245] 通过稳定性选择来鉴定候选生物标记,其接着用于产生随机森林分类器模型。参见例如Shi等,J.Comput.Graph.Stat.15(1):118-138(2006)。简要说,随机森林预测器是个别分类树预测器的系综。参见例如Breiman,Machine Learning,45(1):5-32(2001)。对于各观察结果,各个别树表决一个类别,而森林预测具有多个表决的类别。用户指定待在各结点搜索最佳分割的随机选择的变量的数目(mtry)。Gini 指数用作分割准则。参见例如Breiman等,Classification and Regression Trees,Chapman and Hall,New York,1984。使最大可能的树生长,并且不修剪。森林中的各树的根结点含有来自原始数据的引导样本作为训练集。不在训练集中的观察结果(大致是原始数据集的1/3)被称为袋外(out-of-bag, OOB)观察结果。可如下获得OOB预测:对于原始数据中的情况,通过仅涉及在它们的相应引导样本中不含有所述情况的那些树的多个表决来预测结果。通过对比这些OOB预测与训练集结果,可获得被称为OOB误差率的预测误差率的估计值。

[0246] 使用非参数柯尔莫哥罗夫-斯米尔诺夫检验(KS统计学)进行单变量分析,所述检验定量各适体关于指定为病例(轻度、中度和重度脂肪变性和/或NASH 1-4)和对照(正常肥胖)的两个参照分布的累积分布函数之间的距离。随机森林分类器的性能取决于用于构建和训练分类器的生物标记的数目和质量。单一生物标记将根据如本文例示的它的KS距离和它的PCA(主成分分析)值来表现。如果分类器性能量度定义为灵敏性(真阳性的分数, fTP)和特异性(1减去假阳性的分数, 1-fFP)的总和,那么完美分类器将具有计分2,并且随机分类器平均来说将具有计分1。使用KS距离的定义,使cdf(累积分布函数)函数的差异最大的那个值 x^* 可通过针对导致 $p(x^*|c) = p(x^*|d)$ 的 x (即当类别依赖性pdf(概率密度函数)交叉时存在的KS距离)求解

$$[0247] \quad \frac{\partial KS}{\partial x} = \frac{\partial(cdf_c(x) - cdf_d(x))}{\partial x} = 0$$

[0248] 来得到。将 x^* 的这个值代入KS距离的表达式来产生对KS的以下定义

$$KS = cdf_c(x^*) - cdf_d(x^*)$$

$$= \int_{-\infty}^{x^*} p(x|c)dx - \int_{-\infty}^{x^*} p(x|d)dx$$

[0249]

$$= 1 - \int_{x^*}^{\infty} p(x|c)dx - \int_{-\infty}^{x^*} p(x|d)dx$$

$$= 1 - f_{FP} - f_{FN},$$

[0250] KS距离是1减去使用截断值在 x^* 下的检验(基本上是单分析物贝叶斯分类器)的总误差分数。因为我们定义计分灵敏性+特异性=2-fFP-f_{FN},所以组合KS距离的以上定义,我们见到灵敏性+特异性=1+KS。我们选择统计量固有地适于构造分类器的生物标记。

[0251] 如果随后添加的生物标记不依赖第一生物标记,那么添加具有良好KS距离(例如 >0.3)的随后生物标记将通常改进分类性能。使用灵敏性加特异性作为分类器计分,可产生许多高计分分类器。

[0252] A. 脂肪变性分类器

[0253] 基于受试者分类,我们假定所有脂肪变性组以及1-4期NASH都在肝细胞中具有脂肪。通过将肥胖正常受试者与所有NAFLD受试者进行比较来产生分类器(脂肪变性或肝中有脂肪)。

[0254] 向随机森林算法供给通过稳定性选择(参见图1)来选择的标记以产生模型。提供所得ROC曲线(参见下文)。

[0255] 针对NAFLD(脂肪变性)的9种标记的分类器的ROC曲线显示于图2中。曲线下面积(AUC)是 0.90 ± 0.03 。灵敏性是92%,并且特异性是63%,其中截断值是0.5。

[0256] 将跨越所有组的各个体的由各分类器的模型得到的概率计分(即 Prob(脂肪变性))绘图以估计除它构建于其上的二元判定之外,它是否也可用作严重性/监测模型(图3)。图显示在无脂肪变性与脂肪变性之间有明确辨别,并且概率表决随脂肪变性的程度而增加。在所有时期的NASH受试者都患有重度脂肪变性。

[0257] 图4显示分类器中9种生物标记的累积分布函数(CDF)。

[0258] 表3显示9种标记的分类器中的生物标记。表3也提供某些生物标记的替代性名称,各生物标记的基因名称和UniProt登记号,以及相较于正常群体,生物标记在较高或较低水平下存在于NAFLD群体中。

[0259] 表3:针对NAFLD的9种生物标记的分类器

生物标记/别名	基因名称	UniProt	在所有 NAFLD 群体中较高/较低的生物标记水平
氨基酰化酶-1	ACY1	Q03154	较高
性激素结合球蛋白	SHBG	P04278	较低
组织蛋白酶 Z、组织蛋白酶 P、组织蛋白酶 X	CTSZ	Q9UBR2	较高
c-met、肝细胞生长因子受体、Met 原致癌基因酪氨酸激酶	MET	P08581	较低
凝溶胶蛋白	GSN	P06396	较低
半乳糖凝集素-3 结合蛋白、凝集素半乳糖苷结合可溶性 3 结合蛋白	LGALS3BP	Q08380	较高
组织型纤维蛋白溶酶原活化因子、tPA	PLAT	P00750	较高
神经细胞粘着分子 L1 样蛋白	CHL1	O00533	较低
抗凝血酶 III	SERPINC1	P01008	较低

[0261] 图3显示各受试者组(自左至右:正常、轻度脂肪变性、中度脂肪变性、重度脂肪变性、NASH1、NASH2、NASH3、NASH4)中9种生物标记分类器的箱线图。各箱内的黑线代表数据点的中值(或第50百分位数),并且箱自身代表四分位间范围(IQR),即涵盖自第25百分位数至第75百分位数的数据点的区域。须触线延伸以涵盖箱的顶部和底部的 $1.5 \times IQR$ 内的数据点。

[0262] B. NASH(纤维化)分类器

[0263] 所有患有NASH的受试者都具有与纤维化相伴的某种形式的炎症和气球般膨胀。因此,我们将所有脂肪变性组与2、3和4期NASH进行比较。为确保鉴定真实纤维化生物标记,排除1期NASH组。

[0264] 向随机森林算法供给通过稳定性选择(图5)来选择的标记以产生模型。以下提供所得ROC。

[0265] 针对2、3和4期NASH(纤维化)的4种标记的分类器的ROC曲线显示于图6中。曲线下面积(AUC)是 0.82 ± 0.07 ,其中灵敏性是62%,并且特异性是92%,处于截断值0.5下。

[0266] 将跨越所有组的各个体的由各分类器的模型得到的概率计分(即 Prob(脂肪变性))绘图以估计除它构建于其上的二元判定之外,它是否也可用作严重性/监测模型(图7)。图显示在无脂肪变性与脂肪变性之间有明确辨别,并且概率表决随脂肪变性的程度而增加。在所有时期的NASH受试者都患有重度脂肪变性。

[0267] 表4显示4种标记的分类器中的生物标记。表4也提供某些生物标记的替代性名称,各生物标记的基因名称和UniProt登记号,以及相较于所有NAFLD群体,生物标记在较高或

较低水平下存在于 NASH 2、3和4群体中。

[0268] 表4:2、3和4期NASH相对于脂肪变性 (NAFLD) 的4种生物标记的分类器

生物标记/别名	基因名称	UniProt	在纤维化 (NASH) 群体中较高/较低的生物标记水平
补体 C7	C7	P10643	较高
[0269] 肾胶原凝集素 1	COLEC11	Q9BWP8	较高
肽基脯氨酰基异构酶 D	PPID	Q08752	较低
胰岛素样生长因子结合蛋白 3	IGFBP3	P17936	较低

[0270] 图7显示各受试者组 (自左至右:正常、轻度脂肪变性、中度脂肪变性、重度脂肪变性、NASH1、NASH2、NASH3、NASH4) 中4 种生物标记的分类器的箱线图。各箱内的黑线代表数据点的中值 (或第50百分位数), 并且箱自身代表四分位间范围 (IQR), 即涵盖自第25百分位数至第75百分位数的数据点的区域。须触线延伸以涵盖箱的顶部和底部的 $1.5 \times IQR$ 内的数据点。

[0271] 图8显示分类器中4种生物标记的CDF。

[0272] 实施例3. 针对NAFLD和/或NASH的其它生物标记和分类器

[0273] 稳定性选择采用许多半数数据子组, 并且使用作为正则化逻辑回归模型的套索分类器进行生物标记选择。参见例如Meinshausen等, 2010, J. Royal Statistical Soc: Series B (Statistical Methodology), 72: 417-473。单一生物标记的选择路径是分摊通过套索模型历经一定范围的 λ 选择那个生物标记所针对的这些子组。 λ 是决定多少生物标记通过套索来选择的调谐参数。历经一定范围的 λ 值的最大选择概率是用于选择一组生物标记的最终量度。

[0274] 使用稳定性选择方法, 确定其它分类器以区分各个个体组。分类器 (包括以上讨论的分类器) 显示于表5中。由比较2和5得到的标记用于构造如上讨论的针对脂肪变性 (NAFLD) 和纤维化 (NASH) 的随机森林分类器。

[0275] 表5. 使用稳定性选择获得的分类器

	比较	灵敏性	特异性	标记
	1 对照相对于 1 至 4 期 NASH	0.8641304	0.8196721	ACY、SHBG、LGALS3BP、SIGLEC7、CTSZ、MET、GSN
[0276]	2 对照相对于 (所有脂肪变性+1-4 期 NASH)	0.8635015	0.7452830	ACY、SHBG、CTSZ、MET、GSN、LGALS3BP、PLAT、CHL1、SERPINC1
	3 对照相对于所有脂肪变性	0.7664234	0.7440000	ACY、SHBG、SIGLEC14
	4 所有脂肪变性相对于 NASH 1	0.4901961	0.5743243	TOP1、SIGLEC14

[0277]	5	所有脂肪变性相对于 2-4 期 NASH	0.7543860	0.8310811	C7、COLEC11、PPID、IGFBP3
	6	所有脂肪变性相对于 1-4 期 NASH	0.6734694	0.5983607	SIGLEC14、AIMP1、TOP1、COLEC11、CA6、STX1A

[0278] *所有脂肪变性:轻度、中度和重度脂肪变性

[0279] 表5中的比较1显示以86.4%灵敏性和82%特异性区分对照受试者与1至4期NASH的7种标记的分类器。比较3显示以76.6%灵敏性和74.4%特异性相对于所有脂肪变性(轻度、中度和重度)来区分对照受试者的3种标记分类器。

[0280] 关于表5中所列的不在上表3和4中的生物标记的其它信息显示于表6中。

[0281] 表6:针对NAFLD和/或NASH的其它生物标记

生物标记/别名	基因名称	UniProt	相对于肥胖对照, 相对于脂肪变性, 在 NASH 中较高/较低的生物标记水平
唾液酸结合 Ig 样凝集素 7	SIGLEC 7	Q9Y286	较高
Siglec-14	SIGLEC14	Q08ET2	较高
拓扑异构酶 I	TOP1	P11387	较低
内皮-单核细胞活化多肽 2 (EMAP-2)	AIMP1	Q12904	较低
碳酸酐酶 6	CA6	Q16623	较低
突触融合蛋白 1A	STX1A	Q9Y286	较低

[0283] 根据对照组相对于1至4期NASH的单变量KS距离的前25种生物标记显示于表7中。这些生物标记以及这些生物标记的组合可用于区分对照受试者(如肥胖受试者)与患有NASH的受试者和/或区分对照受试者与患有脂肪变性的受试者。

[0284] 表7:前25种生物标记

生物标记	有正负之分的ks距离	UniProt	相对于肥胖对照, 在NASH中较高/较低的生物标记水平
ACY1	0.61843	Q03154	较高
THBS2	0.48071	P35442	较高
LGALS3BP	0.44946	Q08380	较高
KYNU	0.44146	Q16719	较高
COLEC11	0.43708	Q9BWP8	较高
CTSZ	0.42166	Q9UBR2	较高
IL19	0.41688	Q9UHD0	较高
POR	0.41021	P16435	较高
INS	0.40888	P01308	较高
SHBG	-0.39253	P04278	较低
GPT	0.37459	Q9H3H5	较高
GNS	0.37136	P15586	较高
RET	0.37043	P07949	较高
AFM	0.36986	P43652	较高
SELE	0.36239	P16581	较高
CD163	0.36146	Q86VB7	较高
ENPP7	0.35518	Q6UWV6	较高
IGFBP7	0.35063	Q16270	较高
GSN	-0.34851	P06396	较低
SIGLEC7	0.34564	Q9Y286	较高
LAMA1.LAMB1.LAMC1	0.33401	P25391 P07942 P11047	较高
IL18R1	0.31973	Q13478	较高
FNI	0.31898	P02751	较高
TGFBI	0.31823	Q15582	较高
AKR1A1	0.31421	P14550	较高

[0286] 实施例4:使用适体进行的示例性生物标记检测

[0287] 检测样品中的一种或多种生物标记的示例性方法描述于例如 Kraemer等, PLoS One 6(10):e26332中,并且在以下描述。描述三种不同定量方法:基于微阵列芯片的杂交、基于Luminex珠粒的方法和 qPCR。

[0288] 试剂

[0289] 可例如自Fisher Biosciences购买HEPES、NaCl、KCl、EDTA、EGTA、MgCl₂和吐温-20。可例如自AIC购买标称分子量8000的硫酸葡聚糖钠盐(DxS04),并且在1次交换下相对于去离子水透析至少20小时。可例如自VWR购买KOD EX DNA聚合酶。可例如自Sigma-Aldrich购买氯化四甲铵和CAPSO,并且可例如自Moss Inc购买抗生蛋白链菌素-藻红素(SAPE)。可例如自Gold Biotechnology购买4-(2-氨基乙基)-苯磺酰氟盐酸盐(AEBSF)。可例如自Thermo Scientific购买抗生蛋白链菌素涂布的96孔板(Pierce抗生蛋白链菌素涂布的板HBC,透明,96孔,产品号15500或15501)。可例如自Thermo Scientific购买NHS-PE04-生物素(EZ-Link NHS-PE04-生物素,产品号21329),溶解于无水DMSO中,并且可以单次使用等分试样形式冷冻储存。可例如自R&D Systems购买IL-8、MIP-4、载脂蛋白-2、RANTES、MMP-7和

MMP-9。可例如自PeproTech购买抵抗素和MCP-1,并且可例如自VWR购买tPA。

[0290] 核酸

[0291] 可例如自Integrated DNA Technologies (IDT) 购买常规(包括胺和生物素取代的)寡脱氧核苷酸。Z-区块是具有序列 5'- (AC-BnBn) 7-AC-3' 的单链寡脱氧核苷酸,其中Bn指示苯甲基取代的脱氧尿苷残基。Z-区块可使用常规亚磷酰胺化学来合成。适体捕获试剂也可通过常规亚磷酰胺化学来合成,并且可例如使用例如 timberline TL-600或TL-150加热器和用以洗脱产物的一定梯度的碳酸氢三乙铵 (TEAB) /ACN, 在于80°C下在Waters Autopurification 2767 系统(或Waters 600系列半自动化系统)上操作的21.5×75mm PRP-3 柱上纯化。在260nm下进行检测,并且跨越主峰收集洗脱份,随后汇合最佳洗脱份。

[0292] 缓冲液

[0293] 缓冲液SB18由40mM HEPES、101mM NaCl、5mM KCl、5mM MgCl₂和0.05% (v/v) 吐温20组成,用NaOH调整至pH 7.5。缓冲液SB17是补充有1mM EDTA三钠的SB18。缓冲液PB1由10mM HEPES、101mM NaCl、5mM KCl、5mM MgCl₂、1mM EDTA三钠和0.05% (v/v) 吐温-20组成,用NaOH调整至pH 7.5。CAPSO 洗脱缓冲液由100mM CAPSO (pH 10.0) 和1M NaCl组成。中和缓冲液含有500mM HEPES、500mM HCl和0.05% (v/v) 吐温-20。Agilent杂交缓冲液是作为试剂盒(寡aCGH/ChIP-芯片上杂交试剂盒)的一部分供给的专有制剂。Agilent洗涤缓冲液1是专有制剂(寡 aCGH/ChIP-芯片上洗涤缓冲液1,Agilent)。Agilent洗涤缓冲液2是专有制剂(寡aCGH/ChIP-芯片上洗涤缓冲液2,Agilent)。TMAC杂交溶液由4.5M氯化四甲铵、6mM EDTA三钠、75mM Tris-HCl (pH 8.0) 和0.15% (v/v) 十二烷基肌氨酸钠组成。KOD缓冲液(10倍浓缩)由1200mM Tris-HCl、15mM MgSO₄、100mM KCl、60mM (NH₄)₂SO₄、1% v/v曲通(Triton)-X 100和1mg/mL BSA组成。

[0294] 样品制备

[0295] 在25°C水浴中使血清(以100μL等分试样形式储存在-80°C下)解冻10分钟,接着储存在冰上,随后进行样品稀释。通过温和涡旋 8秒来混合样品。通过稀释至补充有0.6mM MgCl₂、1mM EGTA三钠、0.8mM AEBSF和2μM Z-区块的0.94×SB17中来制备6%血清样品溶液。一部分6%血清储备溶液于SB17中稀释10倍以产生0.6%血清储备物。在一些实施方案中,6%和0.6%储备物分别用于检测高丰度和低丰度分析物。

[0296] 捕获试剂(适体)和抗生蛋白链菌素板制备

[0297] 适体根据它们的同源分析物(或生物标记)的相对丰度被编组成 2种混合物。各适体的储备物浓度是4nM,并且各适体的最终浓度是 0.5nM。适体储备混合物于SB17缓冲液中稀释4倍,加热至95°C,持续5分钟,并且历经15分钟时期冷却至37°C,随后进行使用。这个变性-复性循环意图使适体构象异构体分布标准化,并且因此确保尽管来历可变,但适体活性可重现。抗生蛋白链菌素板在使用之前用 150μL缓冲液PB1洗涤两次。

[0298] 平衡和板捕获

[0299] 使加热-冷却的2×适体混合物(55μL)与相等体积的6%或0.6%血清稀释液组合,从而产生含有3%和0.3%血清的平衡混合物。各板用硅酮密封垫(Axymat硅酮密封垫,VWR)密封,并且在37°C下孵育1.5小时。接着将平衡混合物转移至洗涤的96孔抗生蛋白链菌素板的孔中,并且在设置在37°C下的Eppendorf热混合器上在800rpm 下振荡下再孵育2小时。

[0300] 手动测定

[0301] 除非另外规定, 否则通过倾卸来移除液体, 随后于分层的纸巾上轻叩2次。洗涤体积是150 μ L, 并且所有振荡孵育都在设置在25 $^{\circ}$ C、800rpm下的Eppendorf热混合器上进行。通过吸移来移除平衡混合物, 并且各板用补充有1mM硫酸葡聚糖和500 μ M生物素的缓冲液PB1洗涤两次持续1分钟, 接着用缓冲液PB1洗涤4次持续15秒。添加1mM NHS-PEO4-生物素于缓冲液PB1中的新鲜制备溶液(150 μ L/孔), 并且在振荡下孵育各板5分钟。移除NHS-生物素溶液, 并且各板用补充有20mM甘氨酸的缓冲液PB1洗涤3次, 并且用缓冲液PB1洗涤3次。接着添加85 μ L补充有1mM D_xS₀4的缓冲液PB1 至各孔中, 并且在BlackRay紫外灯(标称波长365nm)下在5cm 的距离下在振荡下将各板照射20分钟。将样品转移至未用过的洗涤的抗生素链菌素涂布的板、或现有洗涤的抗生素链菌素板的未使用孔中, 从而将高样品稀释度和低样品稀释度混合物组合至单一孔中。在室温下在振荡下孵育样品10分钟。移除未吸附的物质, 并且各板用补充有30%甘油的缓冲液PB1洗涤8次, 各次持续15秒。各板接着用缓冲液PB1洗涤1次。在室温下用100 μ L CAPSO洗脱缓冲液洗脱适体5分钟。将90 μ L洗脱物转移至96孔HybAid板中, 并且添加10 μ L中和缓冲液。

[0302] 半自动化测定

[0303] 将携带吸附的平衡混合物的抗生素链菌素板放置在被编程以进行以下步骤的BioTek EL406板洗涤器的平台上: 通过吸出来移除未吸附的物质, 并且各孔用300 μ L补充有1mM硫酸葡聚糖和500 μ M 生物素的缓冲液PB1洗涤4次。各孔接着用300 μ L缓冲液PB1洗涤 3次。添加150 μ L的1mM NHS-PEO4-生物素于缓冲液PB1中的新鲜制备(自于DMSO中的100mM储备物) 溶液。在振荡下孵育各板5分钟。吸出液体, 并且各孔用300 μ L补充有10mM甘氨酸的缓冲液PB1洗涤8次。添加100 μ L补充有1mM硫酸葡聚糖的缓冲液 PB1。在这些自动化步骤之后, 自板洗涤器移除各板, 并且放置在以 5cm的距离安放在紫外光源(BlackRay, 标称波长365nm) 下的热振荡器上20分钟。热振荡器被设置在800rpm和25 $^{\circ}$ C下。在20分钟照射之后, 将样品手动转移至未用过的洗涤的抗生素链菌素板(或现有洗涤的板的未使用孔) 中。此时将高丰度(3%血清+3%适体混合物) 和低丰度反应混合物(0.3%血清+0.3%适体混合物) 组合至单一孔中。将这个“两者汇集(Catch-2)”板放置在被编程以进行以下步骤的BioTek EL406板洗涤器的平台上: 在振荡下将板孵育10分钟。吸出液体, 并且各孔用300 μ L补充有30%甘油的缓冲液PB1洗涤21 次。各孔用300 μ L缓冲液PB1洗涤5次, 并且吸出最终洗涤物。添加100 μ L CAPSO洗脱缓冲液, 并且在振荡下洗脱适体5分钟。在这些自动化步骤之后, 接着将板自板洗涤器的平台移除, 并且将样品的 90 μ L等分试样手动转移至含有10 μ L中和缓冲液的HybAid 96孔板的各孔中。

[0304] 与定制Agilent 8 \times 15k微阵列芯片杂交

[0305] 将24 μ L中和的洗脱物转移至新的96孔板中, 并且添加6 μ L含有一组由10种Cy3适体组成的杂交对照的10 \times Agilent阻断剂(寡 aCGH/ChIP-芯片上杂交试剂盒, 大体积, Agilent 5188-5380) 至各孔中。添加30 μ L 2 \times Agilent杂交缓冲液至各样品中并混合。将40 μ L 所得杂交溶液手动吸移至杂交垫圈载片(杂交垫圈载片, 每个载片形式8个微阵列, Agilent)的各“孔”中。根据制造商方案将每个阵列携带10种互补于具有20 \times dT连接子的各适体的40核苷酸随机区域的探针的定制Agilent微阵列载片放置于垫圈载片上。钳夹装配体(杂交室试剂盒-由SureHyb实现, Agilent), 并且在20rpm下旋转下在 60 $^{\circ}$ C下孵育19小时。

[0306] 杂交后洗涤

[0307] 将约400mL Agilent洗涤缓冲液1放入两个单独玻璃染色碟中的每一个中。当浸没在洗涤缓冲液1中时,拆散并分开载片(一次至多两个),接着转移至也含有洗涤缓冲液1的第二染色碟中的载片架。在搅拌下将载片于洗涤缓冲液1中再孵育5分钟。将载片转移至预平衡至37℃的洗涤缓冲液2中,并且在搅拌下孵育5分钟。将载片转移至含有乙腈的第四染色碟中,并且在搅拌下孵育5分钟。

[0308] 微阵列成像

[0309] 在100%PMT设置和在0.05下实现的XRD选项下,在5μm分辨率下,使用Cy3通道,用Agilent G2565CA微阵列扫描器系统使微阵列载片成像。使用10.5.1.1版Agilent特征提取软件,以 GE1_105_Dec08方案处理所得TIFF图像。

[0310] Luminex探针设计

[0311] 固定于珠粒的探针具有40个互补于靶标适体的40个核苷酸随机区域的3'末端的脱氧核苷酸。通过携带5'氨基末端的六乙二醇(HEG) 连接子使适体互补区域偶联于Luminex微球体。生物素化检测脱氧寡核苷酸包含17-21个互补于靶标适体的5'引物区域的脱氧核苷酸。生物素部分附接于检测寡聚物的3'末端。

[0312] 使探针偶联于Luminex微球体

[0313] 基本上根据制造商说明书,但在以下修改下使探针偶联于 Luminex Microplex微球体:氨基末端寡核苷酸量是每 2.5×10^6 个微球体0.08nMol,并且第二次EDC添加量是在10mg/mL下5μL。在设置在25℃和600rpm下的Eppendorf热振荡器中进行偶联反应。

[0314] 微球体杂交

[0315] 涡旋微球体储备溶液(约40000个微球体/μL),并且在Health Sonics超声清洗器(型号:T1.9C)中超声处理60秒以混悬微球体。于 $1.5 \times$ TMAC杂交溶液中将混悬的微球体稀释成每个反应2000个微球体,并且通过涡旋和超声处理来混合。将每个反应33μL的珠粒混合物转移至96孔HybAid板中。添加7μL的于 $1 \times$ TE缓冲液中的15 nM生物素化检测寡核苷酸储备物至各反应中并混合。添加10μL中和的测定样品,并且用硅盖垫密封件将板密封。首先在96℃下将板孵育5分钟,并且在常规杂交烘箱中在50℃下在不搅拌下孵育过夜。过滤板(Dura pore, Millipore部件号MSBVN1250,1.2μm孔尺寸)用75μL补充有0.5%(w/v)BSA的 $1 \times$ TMAC杂交溶液预先湿润。将来自杂交反应的整个样品体积转移至过滤板中。杂交板用75μL含有0.5%BSA的 $1 \times$ TMAC杂交溶液冲洗,并且将任何剩余物质都转移至过滤板中。在缓慢真空下过滤样品,其中历经约8秒抽空150μL 缓冲液。过滤板用75μL含有0.5%BSA的 $1 \times$ TMAC杂交溶液洗涤1次,并且将过滤板中的微球体再混悬于75μL含有0.5%BSA的 $1 \times$ TMAC杂交溶液中。保护过滤板免遭光照,并且在Eppendorf热混合器R上在1000rpm下孵育5分钟。过滤板接着用75μL含有0.5%BSA的 $1 \times$ TMAC杂交溶液洗涤1次。添加75μL的于 $1 \times$ TMAC杂交溶液中的10μg/mL 抗生蛋白链菌素藻红素(SAPE-100, MOSS, Inc.) 至各反应中,并且在Eppendorf热混合器R上在25℃下,在1000rpm 下孵育60分钟。过滤板用75μL含有0.5%BSA的 $1 \times$ TMAC杂交溶液洗涤两次,并且将过滤板中的微球体再混悬于75μL含有0.5%BSA的 $1 \times$ TMAC杂交溶液中。接着在Eppendorf热混合器R上在1000rpm 下将过滤板避光孵育5分钟。过滤板接着用75μL含有0.5%BSA的 $1 \times$ TMAC杂交溶液洗涤1次。将微球体再混悬于75μL补充有0.5% BSA的 $1 \times$ TMAC杂交溶液中,并且在运行XPonent 3.0软件的 Luminex 100仪器

上分析。在高PMT校正和7500至18000的双联体辨别器设置下,对每种珠粒类型计数至少100个微球体。

[0316] QPCR读出

[0317] 于水中,在108至102个拷贝的范围内,在10倍稀释以及无模板对照下制取qPCR的标准曲线。将中和的测定样品40倍稀释至 diH₂O中。在2×最终浓度下制备qPCR主要混合物(2×KOD缓冲液、400μM dNTP混合物、400nM正向和反向引物混合物、2×SYBR Green I和0.5U KOD EX)。添加10μL 2×qPCR主要混合物至10μL 稀释的测定样品中。以96℃下2分钟继之以96℃5秒和72℃30秒的40个循环,在BioRad MyIQ iCycler上操作qPCR。

[0318] 实施例5. 进一步NAFLD和NASH分析

[0319] 使用逻辑回归模型,使来自实施例2中所述的对实施例1中所述的样品的多路适体测定的数据经受进一步分析。

[0320] 稳定性选择采用许多半数数据子组,并且使用作为正则化逻辑回归模型的套索分类器进行生物标记选择。参见例如Meinshausen等, 2010, J. Royal Statistical Soc: Series B (Statistical Methodology), 72: 417-473。单一生物标记的选择路径是分摊通过套索模型历经一定范围的λ选择那个生物标记所针对的这些子组。λ是决定多少生物标记通过套索来选择的调谐参数。历经一定范围的λ值的最大选择概率是用于选择一组生物标记的最终量度。

[0321] C. 脂肪变性分类器

[0322] 基于受试者分类,我们假定所有脂肪变性组以及1-4期NASH都在肝细胞中具有脂肪。通过将肥胖正常受试者与所有NAFLD和 NASH受试者进行比较来产生分类器(脂肪变性或肝中有脂肪)。

[0323] 用逻辑回归模型拟合标记以产生分类器。

[0324] 表8显示8种标记的脂肪变性 (NAFLD) 分类器中的生物标记。表8也显示模型的逻辑回归系数以及系数估计值的95%置信区间。8种标记的分类器含有表3中所示分类的9种标记中除PLAT之外的所有标记。

[0325] 表8: 针对NAFLD的8种生物标记的分类器

生物标记	在所有 NAFLD 群体 中较高/较低的 生物标记水平	系数	低 95%置信 区间	高 95%置信 区间
截距		23.682	0.0205235	48.665007
ACY1	较高	2.205	1.479981	3.028281
SHBG	较低	-1.219	-1.96066	-0.52774
CTSZ	较高	3.070	1.607459	4.636245
MET	较低	-1.900	-3.84342	-0.01765
GSN	较低	-2.014	-4.08462	0.024845
LGALS3BP	较高	1.814	0.58276	3.132223
CHL1	较低	-2.491	-4.46361	-0.59133
SERPINC1	较低	-1.502	-2.96705	-0.13184

[0326]

[0327] 图12显示各组脂肪变性和NASH受试者的8种标记的分类器的箱线图。各箱内的黑线代表数据点的中值(或第50百分位数),并且箱自身代表四分位间范围(IQR),即涵盖自第25百分位数至第75百分位数的数据点的区域。须触线延伸以涵盖箱的顶部和底部的 $1.5 \times IQR$ 内的数据点。图显示在无脂肪变性与脂肪变性之间有明确辨别,并且概率表决随脂肪变性的程度而增加。在所有时期的NASH受试者都患有重度脂肪变性。

[0328] D. NASH(纤维化)分类器

[0329] 所有患有NASH的受试者都具有与纤维化相伴的某种形式的炎症和气球般膨胀。对于这个分析,我们将轻度和中度脂肪变性组与2、3和4期NASH进行比较。为确保鉴定真实纤维化生物标记,排除1期NASH组。在这个分析中省略患有重度脂肪变性的受试者,因为在那些受试者中进行的活检体取样可已遗漏一些纤维化。

[0330] 用逻辑回归模型拟合标记以产生用以产生模型分类器。

[0331] 表9显示8种标记的纤维化(NASH)分类器中的生物标记。表9也显示模型的逻辑回归系数以及系数估计值的95%置信区间。8种标记的分类器含有表4中所示分类的4种标记中的所有标记加4种其它标记。

[0332] 表9:2、3和4期NASH相对于轻度至中度脂肪变性(NAFLD)的8种生物标记的分类器

生物标记/别名	在纤维化(NASH)群体中较高/较低的生物标记水平	系数	低 95%置信区间	高 95%置信区间
截距		-58.307	-117.01013	-8.491203
C7	较高	2.711339	0.016014	5.569299
[0333] COLEC11	较高	0.578295	-1.01116	2.184399
PPID	较低	-1.47222	-3.09367	0.024429
IGFBP3	较低	-1.88591	-4.25566	0.245891
GPC5	较低	-2.24248	-3.99262	-0.74934
HIPK3	较高	1.176491	0.286007	2.806758
IGFBP7	较高	4.363192	1.274937	8.250915
IL3RA	较高	3.106253	0.455772	6.208888

[0334] 图13显示各受试者组(自左至右:正常、轻度脂肪变性、中度脂肪变性、重度脂肪变性、NASH1、NASH2、NASH3、NASH4)中8种生物标记的纤维化分类器的箱线图。各箱内的黑线代表数据点的中值(或第50百分位数),并且箱自身代表四分位间范围(IQR),即涵盖自第25百分位数至第75百分位数的数据点的区域。须触线延伸以涵盖箱的顶部和底部的 $1.5 \times IQR$ 内的数据点。

[0335] 表9中的纤维化分类器中的4种其它生物标记连同各生物标记的替代性名称、基因名称和UniProt登记号一起显示于表10中。

[0336] 表10:纤维化8种生物标记的分类器中的其它生物标记

生物标记/别名	基因名称	UniProt
磷脂酰肌醇蛋白聚糖-5	GPC5	<u>Q8N158</u>
[0337] 同源结构域相互作用蛋白质激酶 3	HIPK3	<u>Q9H422</u>
胰岛素样生长因子结合蛋白 7	IGFBP7	<u>Q16270</u>
白介素-3 受体 α	IL3RA	<u>P26951</u>

[0338] 图14显示针对纤维化的8种标记的分类器中的4种其它标记的累积分布函数。

[0339] 实施例6. 盲化样品的破盲

[0340] 如表2中所示,使对照、轻度脂肪变性、中度脂肪变性、重度脂肪变性、1期NASH和2期NASH中的每一个中的一组样品盲化。使用表8中所示的8种标记的分类器来确定各盲化样品来自患有脂肪变性的受试者的概率。图15显示在破盲之后盲化样品的根据它们的实际样品组的箱线图。

[0341] 图16显示发现组和盲化验证组的8种标记的脂肪变性分类器性能。上部曲线是发现组的ROC曲线,其具有 0.927 ± 0.03 的曲线下面积(AUC)以及92.8%的灵敏性和73.3%的特异性,处于截断值0.5下。下部曲线是盲化验证组的ROC曲线,其具有 0.889 ± 0.06 的AUC以及88%的灵敏性和65.8%的特异性,处于截断值0.5下。

[0342] 表10显示盲化验证组的8种标记的脂肪变性分类器用于将受试者鉴定为脂肪变性阳性或阴性的性能。

[0343] 表10:盲化验证组的8种标记的脂肪变性分类器的性能

真实/测试	阴性	阳性	总计
阴性	27	14	41
阳性	11	81	92
总计	38	95	133

[0344] 使用表9中所示的8种标记的分类器来确定各盲化样品来自患有纤维化的受试者的概率。图17显示在破盲之后盲化样品的根据它们的实际样品组的箱线图。

[0345] 图18显示发现组和两个不同盲化验证组的8种标记的纤维化分类器性能。第二验证组包括第一验证组加正常组(疾病阴性个体)。因为模型未用正常组训练,所以它对那些样品来说是首次用于实验的,此可揭示假阳性。上部曲线(红色)是发现组(对照(轻度+中度脂肪变性)相对于2、3和4级NASH)的ROC曲线,其具有 0.929 ± 0.04 的曲线下面积(AUC)以及78.8%的灵敏性和89.9%的特异性,处于截断值0.5下。中部曲线(蓝色)是第一盲化验证组的ROC曲线,其具有 0.885 ± 0.11 的AUC以及66.7%的灵敏性和84.6%的特异性,处于截断值0.5下。下部曲线(青色)是第二盲化验证组的ROC曲线,其具有 0.891 ± 0.08 的AUC以及66.7%的灵敏性和89.5%的特异性,处于截断值0.5下。

[0346] 表11和12显示两个盲化验证组的8种标记的纤维化分类器用于将受试者鉴定为纤维化阳性或阴性的性能。

[0347] 表11:第一盲化验证组的8种标记的纤维化分类器的性能

[0349]	真实/测试	阴性	阳性	总计
	阴性	22	4	26
	阳性	4	8	12
	总计	26	12	38

[0350] 表12: 第二盲化验证组的8种标记的纤维化分类器的性能

[0351]	真实/测试	阴性	阳性	总计
	阴性	60	7	67
[0352]	阳性	4	8	12
	总计	64	15	79

[0353] 图19A和图19B分别显示使用8种标记的脂肪变性分类器和8种标记的纤维化分类器, 20%保留验证组的2500次引导迭代的灵敏性和特异性分布。显示经验性95%置信区间, 以及训练和验证估计值, 其两者均在95%置信区间内。

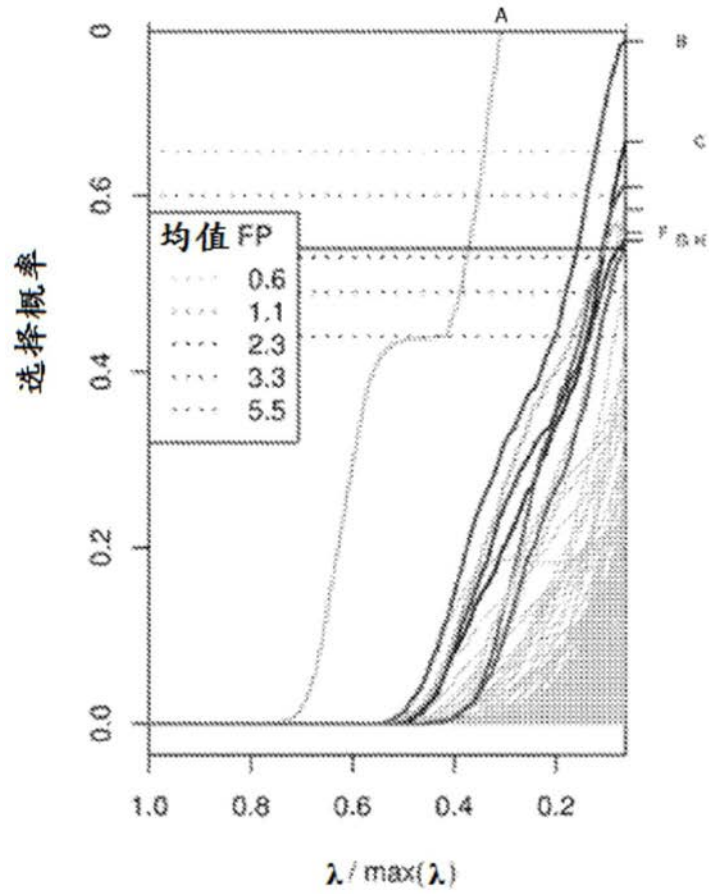
[0354] 实施例7. 分类器在儿科受试者中的性能

[0355] 自威斯康星医学院的儿科胃肠病学和营养学系 (Pediatric Gastroenterology and Nutrition Department of the Medical College of Wisconsin) 获得来自儿童的90个血清样品。捐献样品的受试者属于4个组: 非肥胖对照 (N=45), 肝功能测试正常的肥胖 (N=20), 肝功能测试升高的肥胖 (N=7), 以及被诊断有NASH的受试者 (N=18)。使用表8中所示的8种标记的脂肪变性分类器和表9中所示的8种标记的纤维化分类器测试样品。

[0356] 使用表8中所示的8种标记的分类器来确定各儿科样品来自患有脂肪变性的受试者的概率。图20显示儿科样品的根据它们的实际样品组的箱线图。脂肪变性分类器准确鉴定患有NASH的儿科受试者。

[0357] 不同于脂肪变性分类器, 8种标记的纤维化分类器在这个实验中不有效鉴定患有纤维化的儿科受试者。(数据未显示。)

[0358] 前述实施方案和实施例仅意图为实例。无特定实施方案、实施例、或特定实施方案或实施例的要素应被解释为是任何权利要求的关键、所需或必需要素或特征。可在不脱离本申请的由随附权利要求限定的范围下对公开的实施方案进行各种改变、修改、替代和其它变化。包括附图和实施例的说明书应以说明方式而非限制方式被考虑, 并且所有所述修改和替代都意图包括在本申请的范围。任何方法或过程权利要求中叙述的步骤都可以任何可行顺序来执行, 并且不限于任何实施方案、实施例或权利要求中呈现的顺序。此外, 在任何以上提及的方法中, 一种或多种明确列出的生物标记可被明确排除为个别生物标记或来自任何组的生物标记。



	靶标
A	ACY1
B	SHBG
C	CTSZ
D	MET
E	GSN
F	LGALS3BP
G	PLAT
H	CHL1
I	SERPINC1

图1

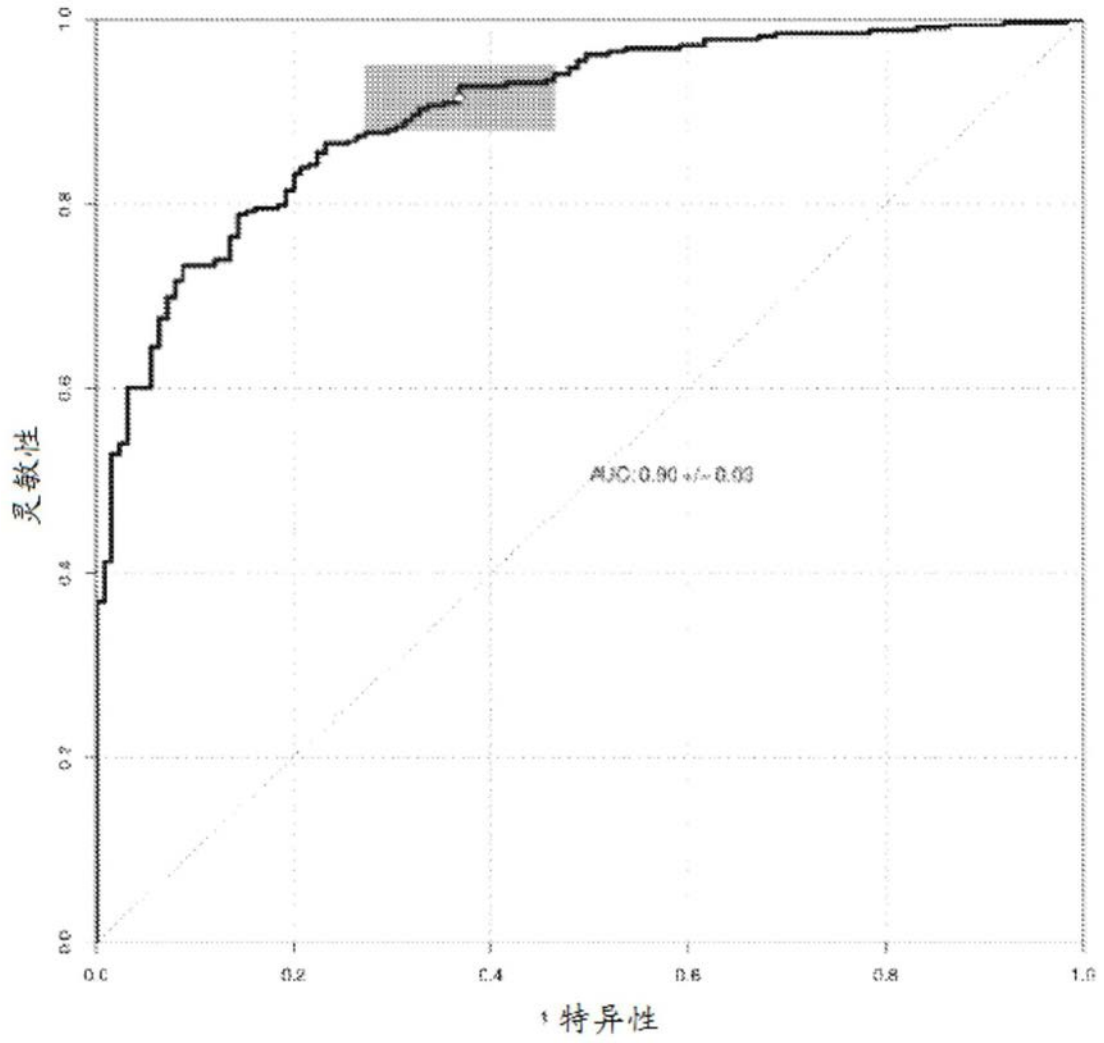


图2

9种标记的随机森林分类器：
根据类别的表决

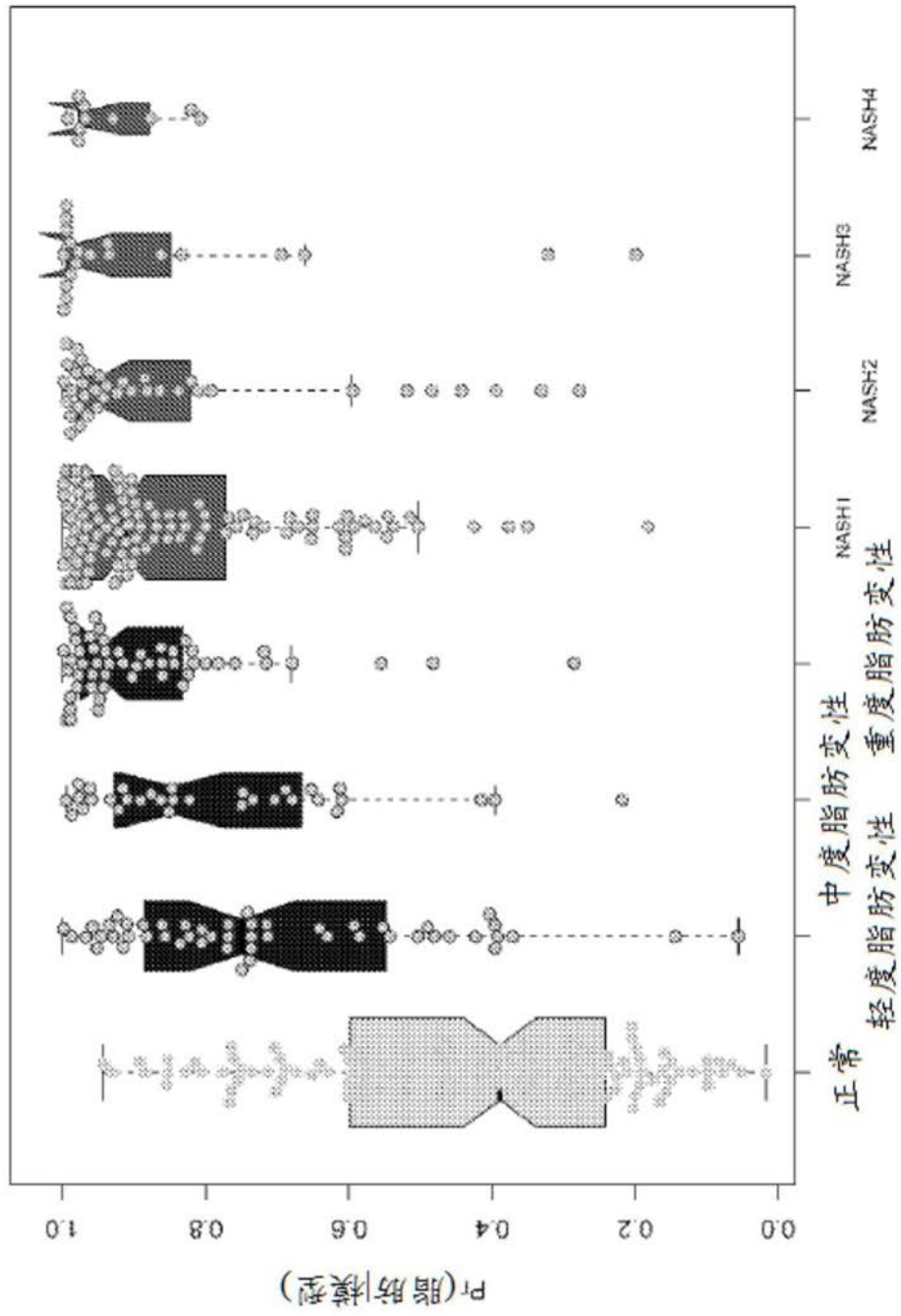


图3

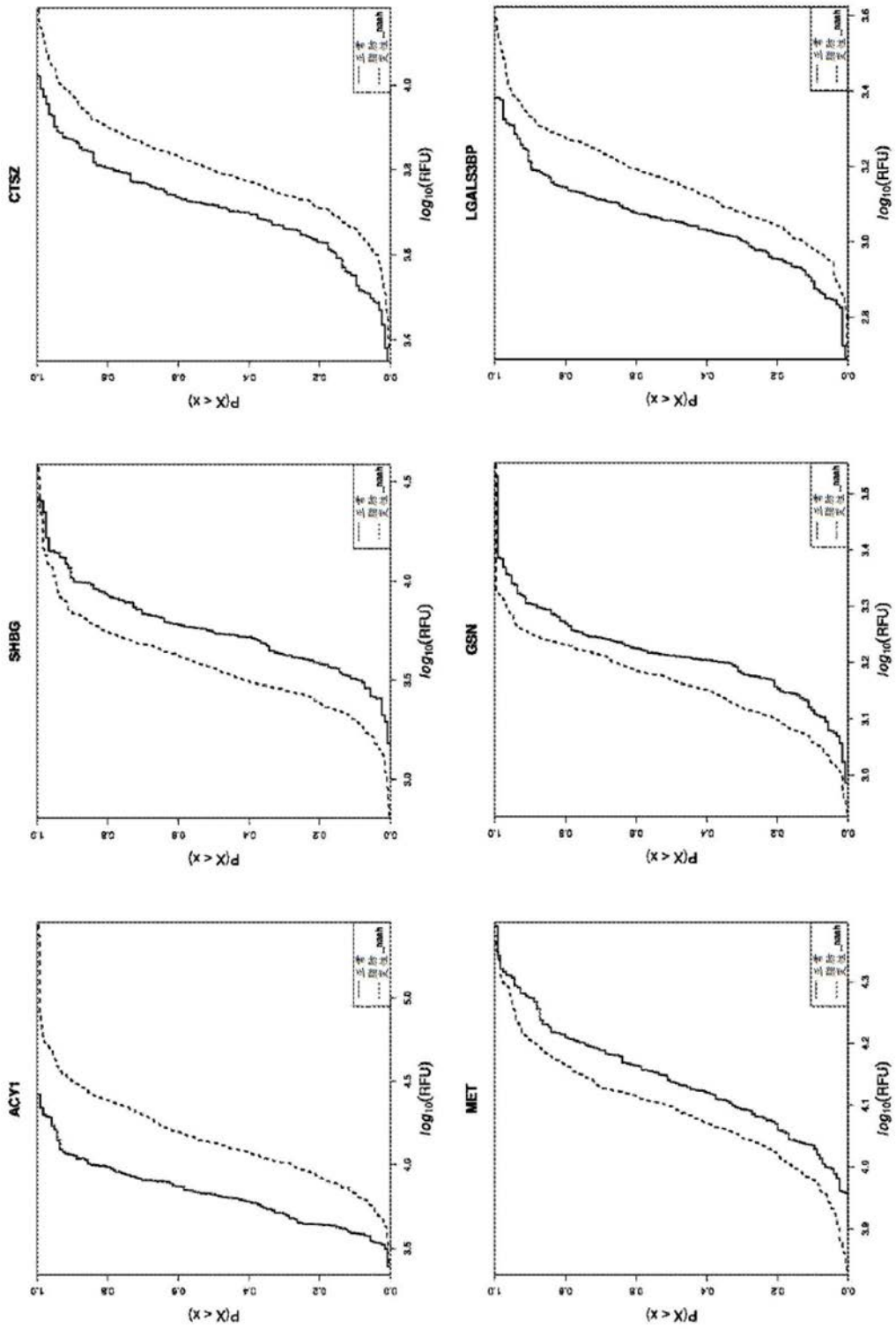


图4

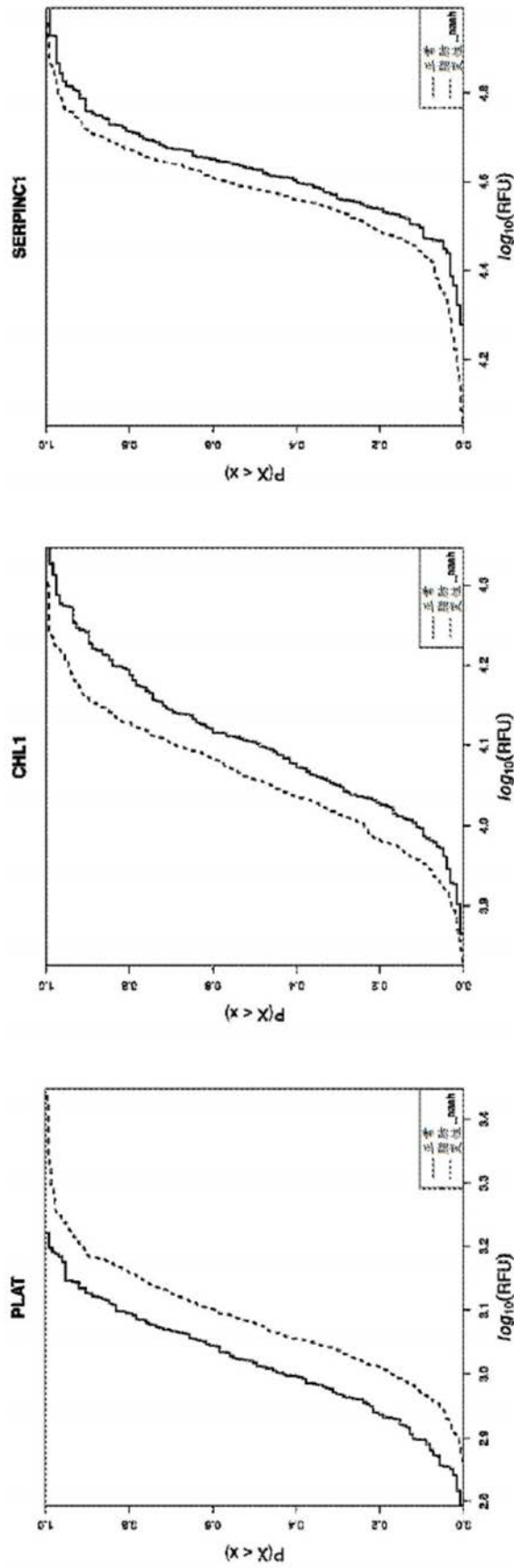
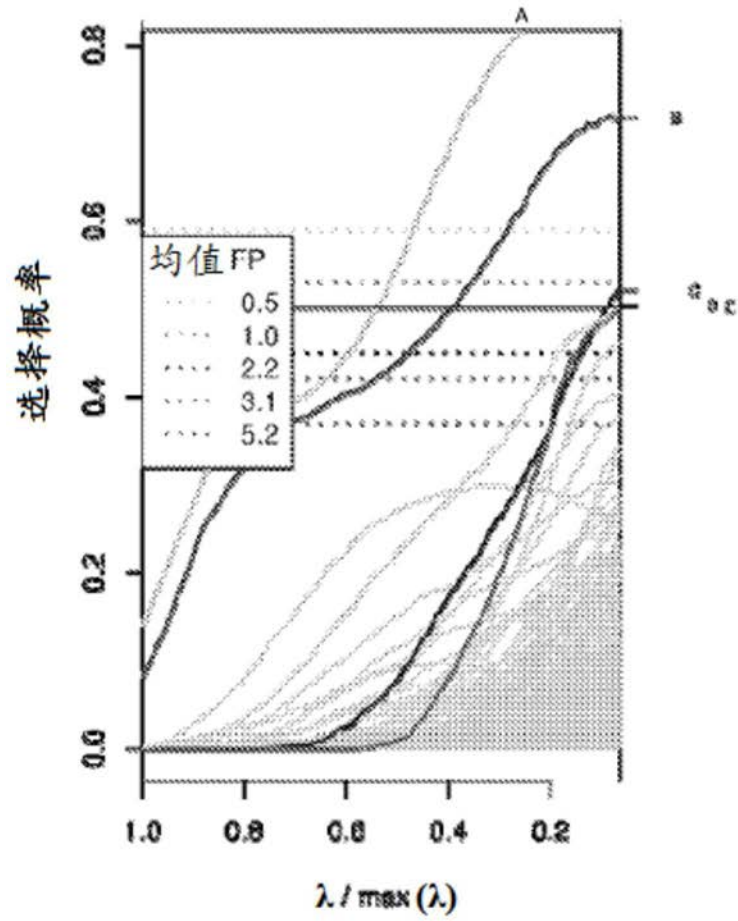


图4(接续)



	靶标
A	C7
B	COLEC11
C	PPID
D	IGFBP3

图5

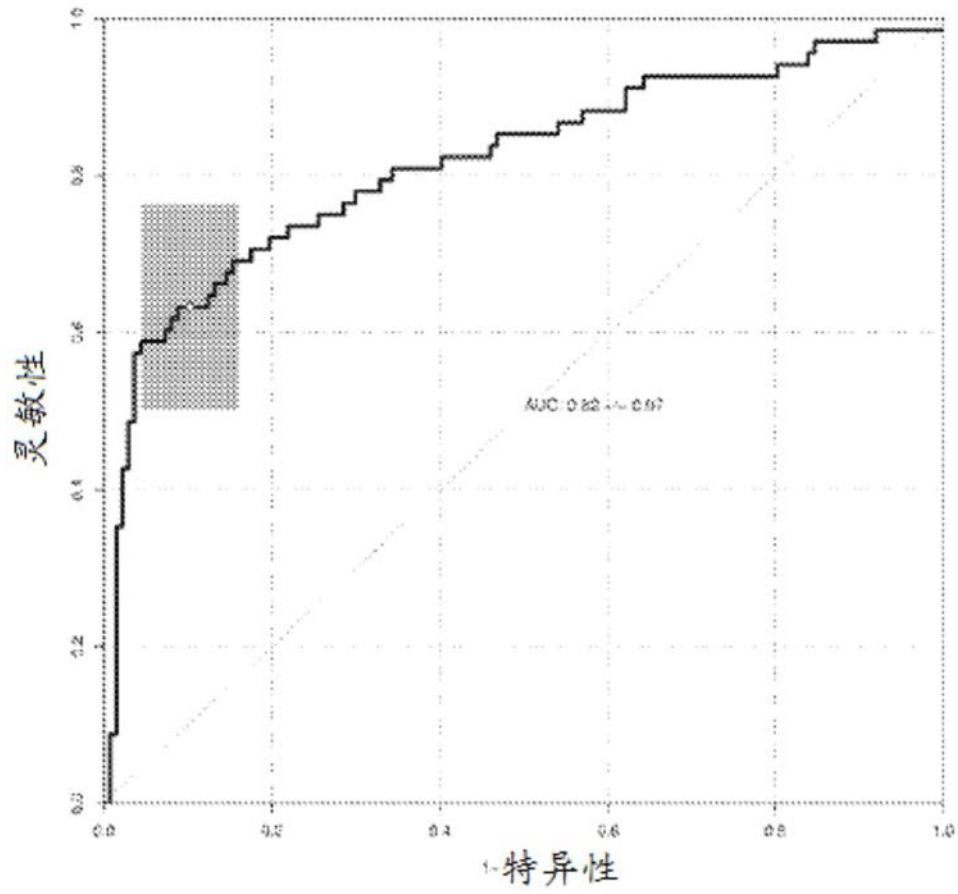


图6

4种标记的随机森林分类器：
根据类别的表决

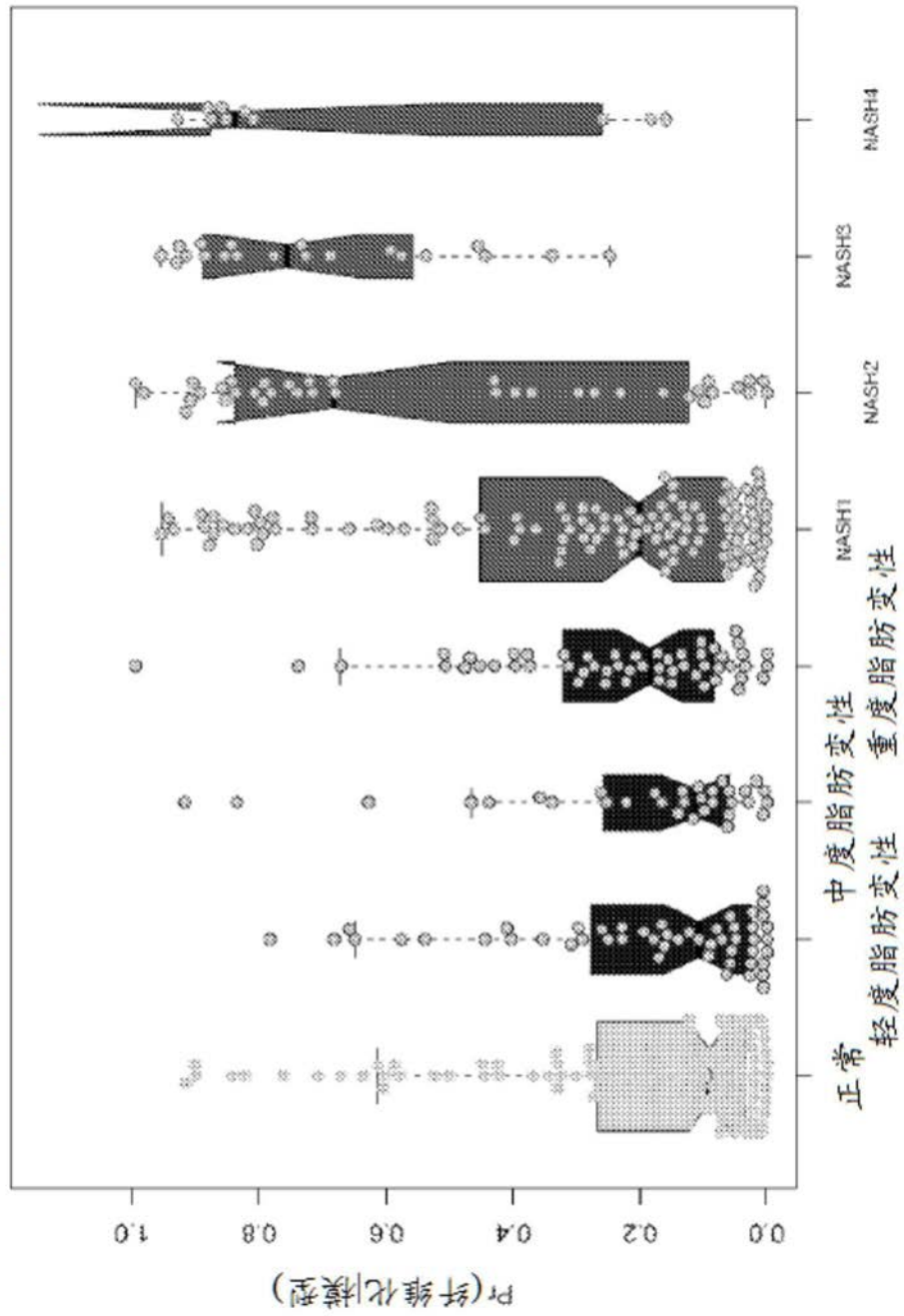


图7

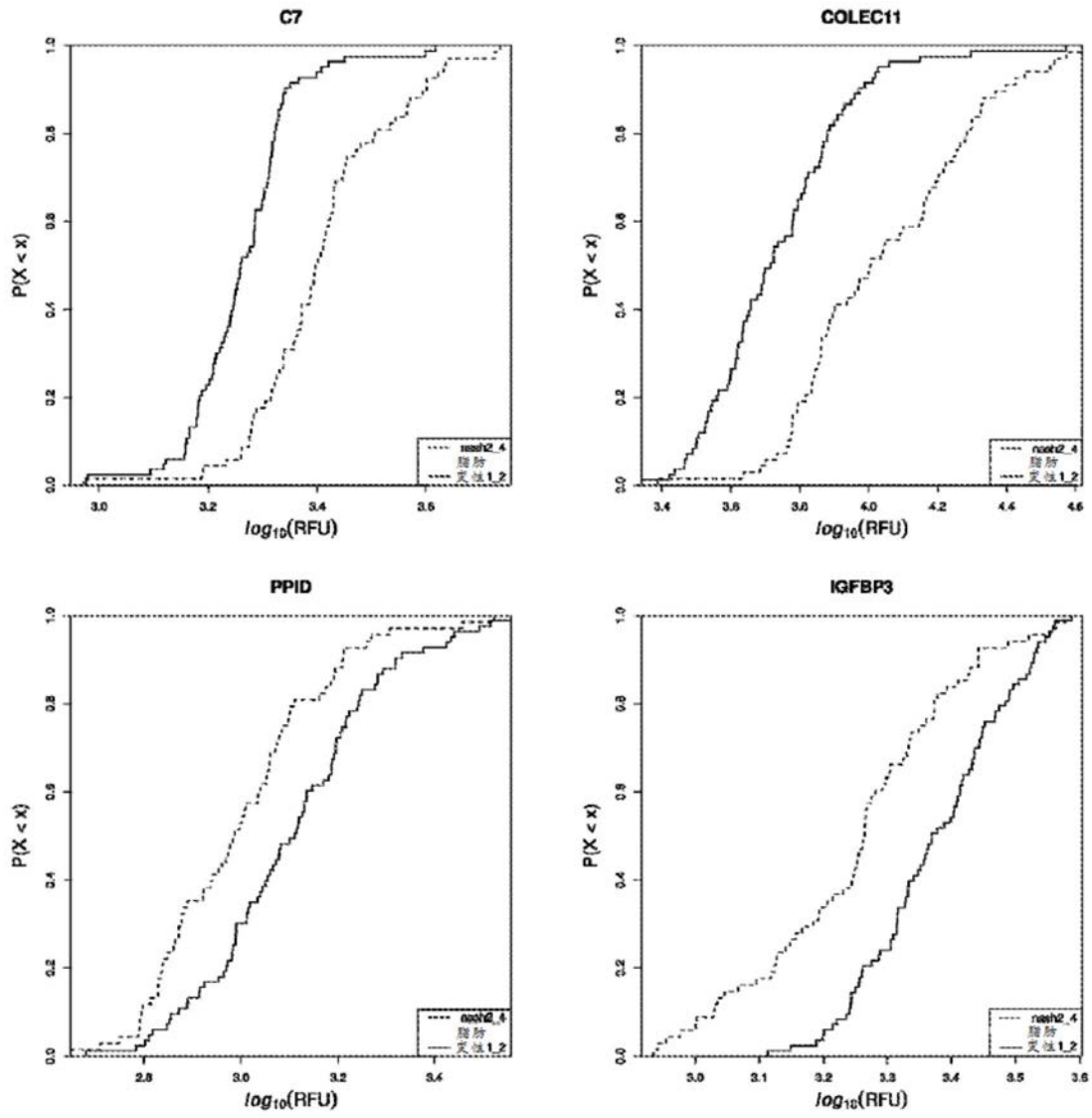


图8

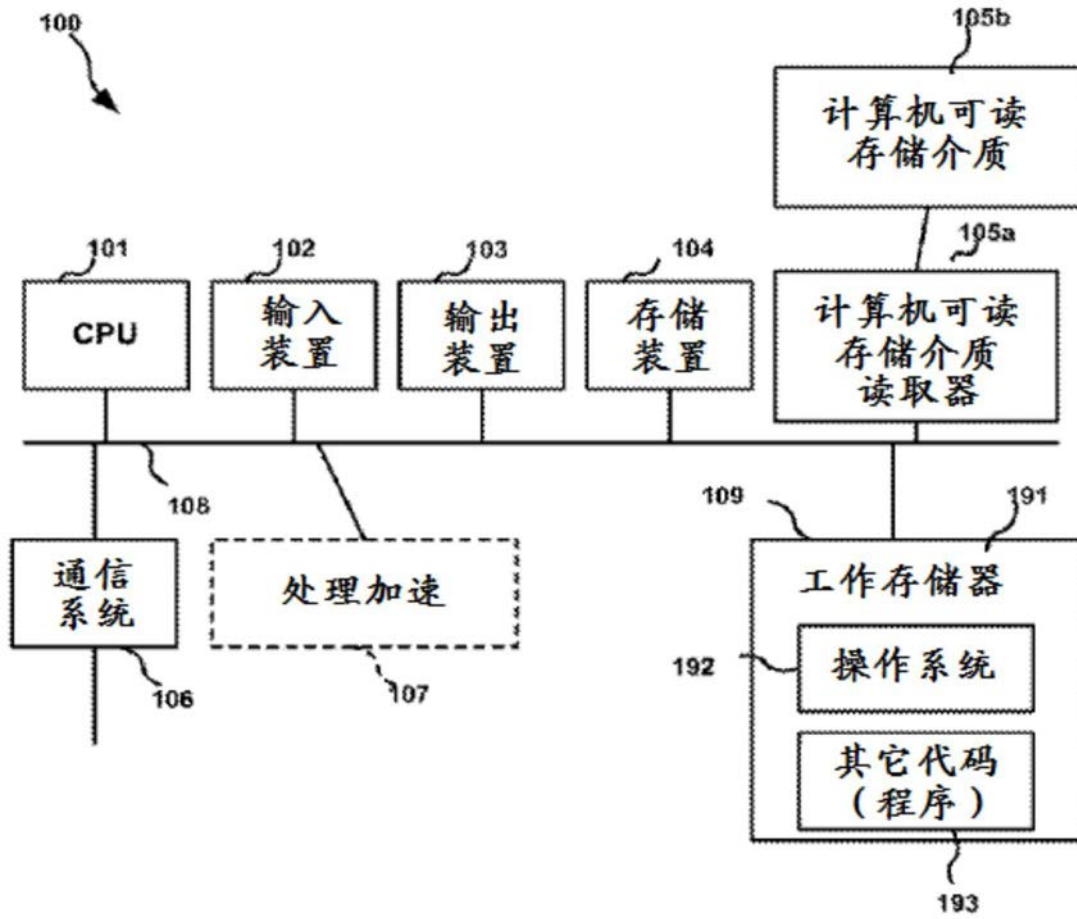


图9

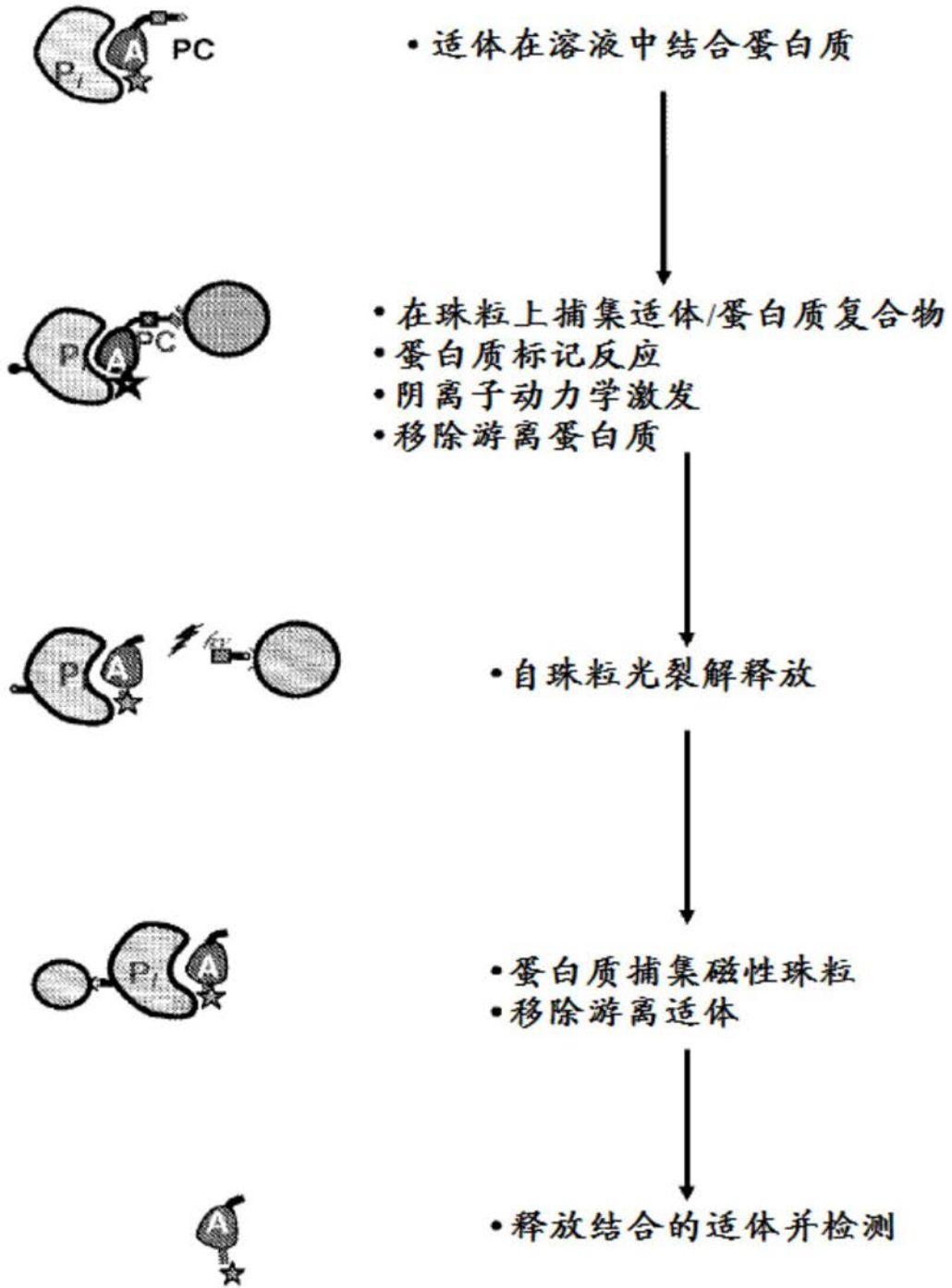
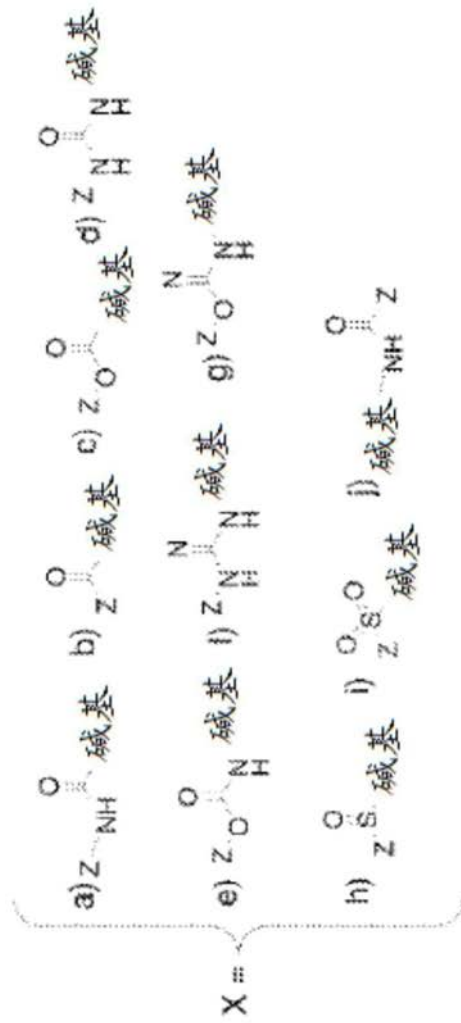


图10



碱基 = 尿苷 (U) 或胞苷 (C) (连接是对于5位的)

Z = R 加 $(\text{CH}_2)_n$ 连接基团, 其中 $n=0-3$

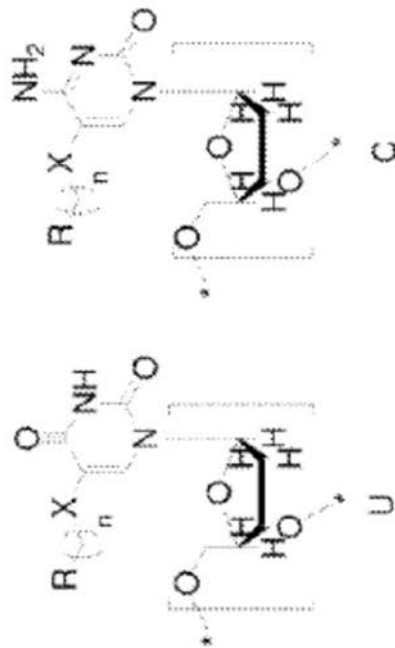


图11

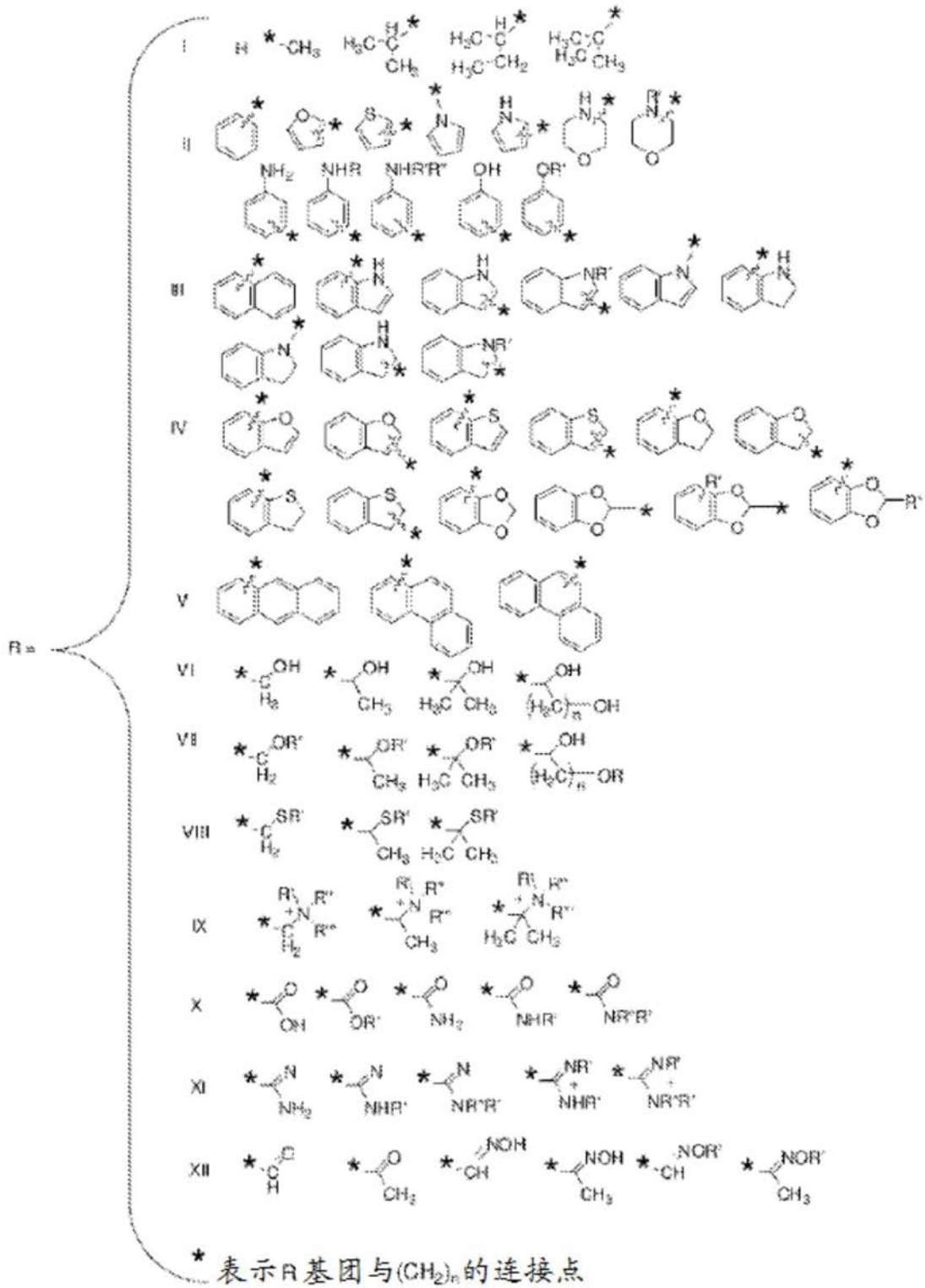


图11(接续)

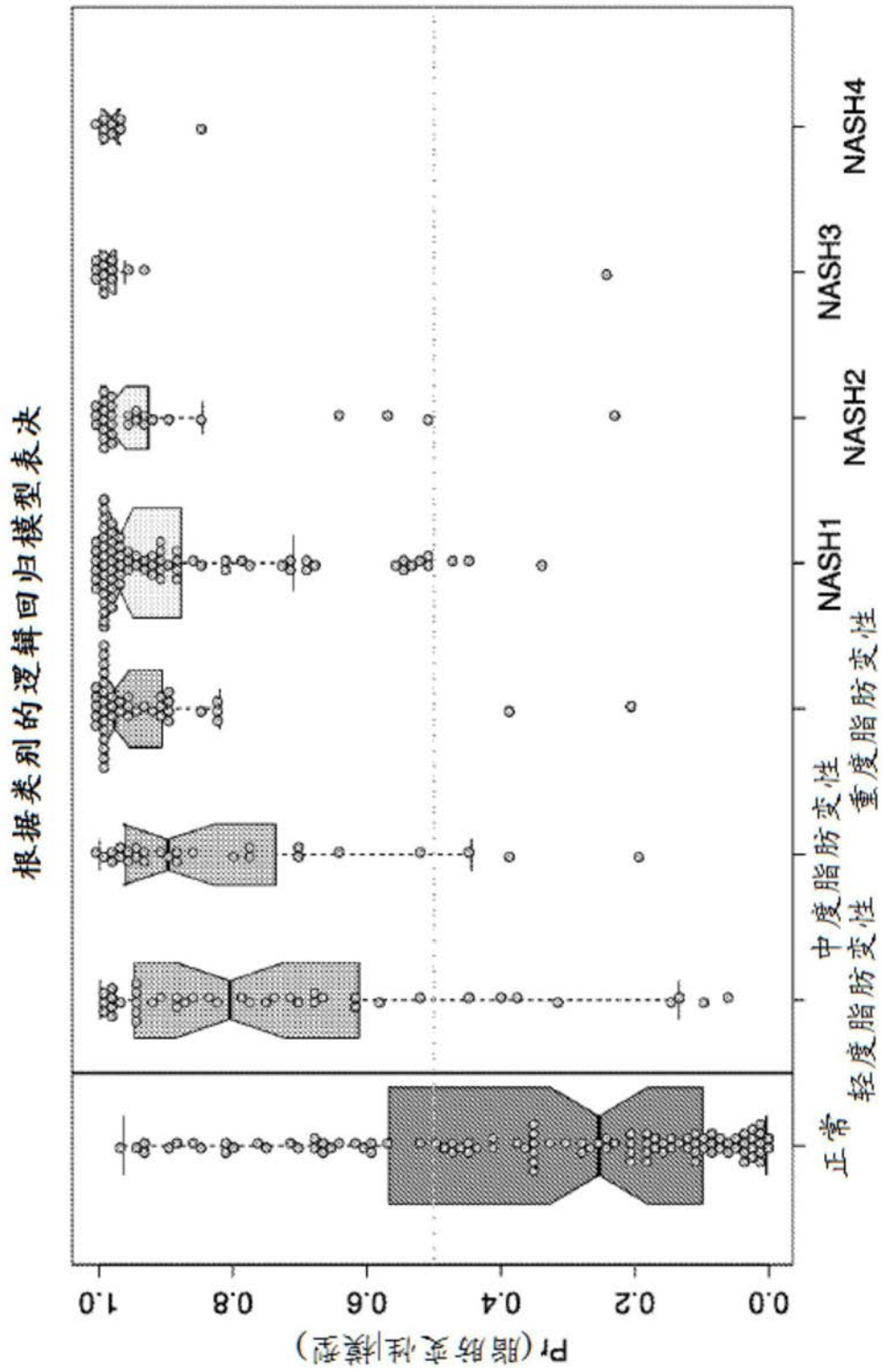


图12

根据类别的逻辑回归模型表决

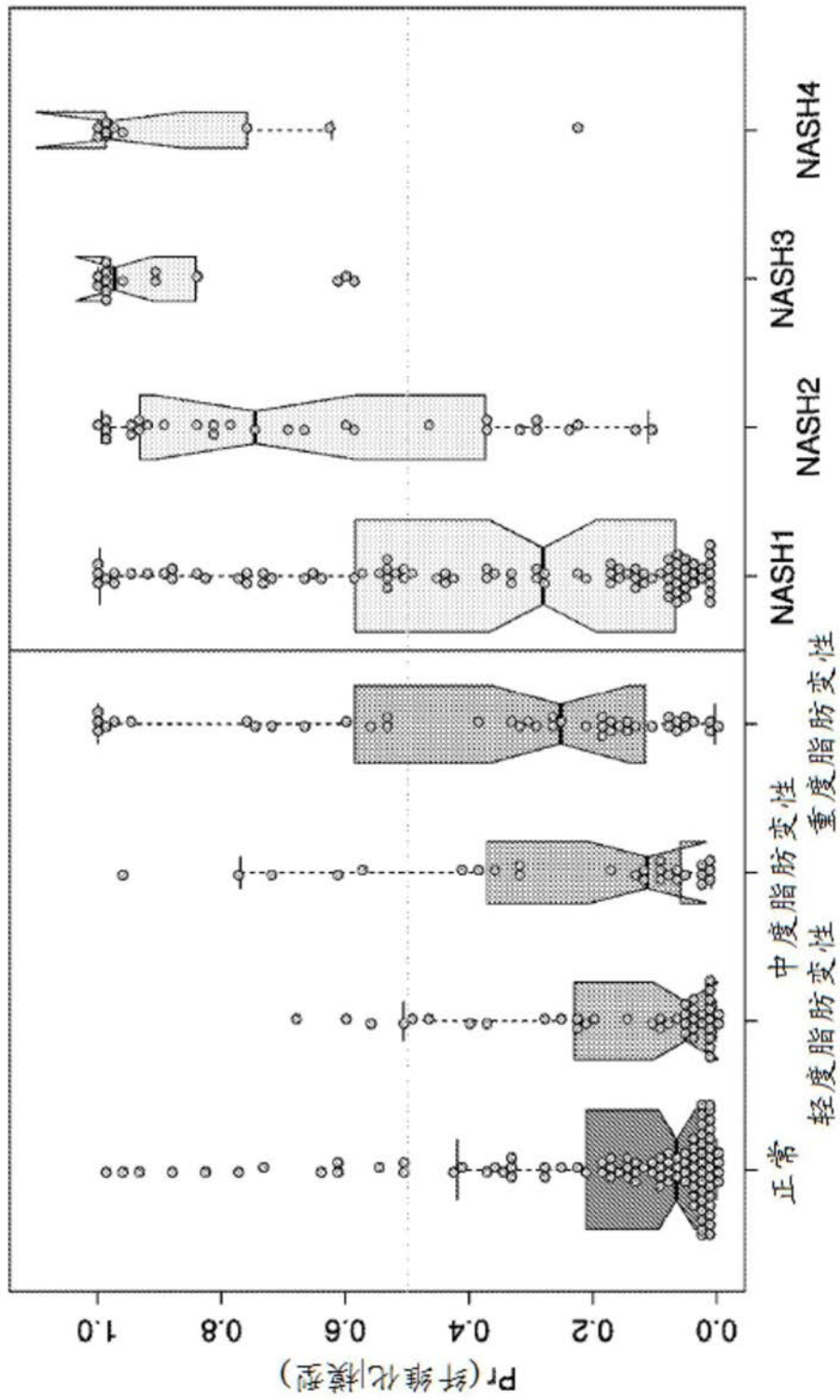


图13

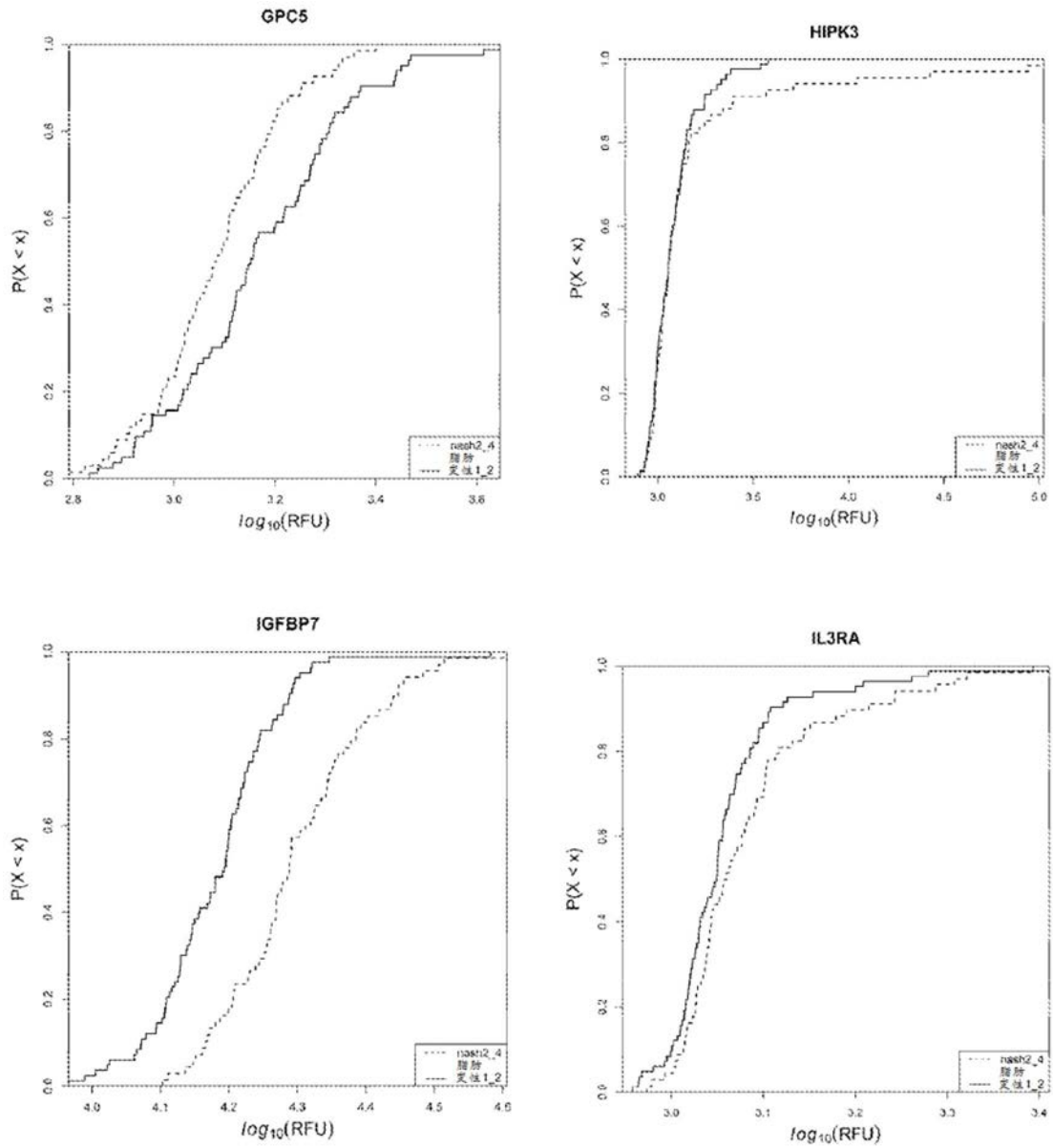


图14

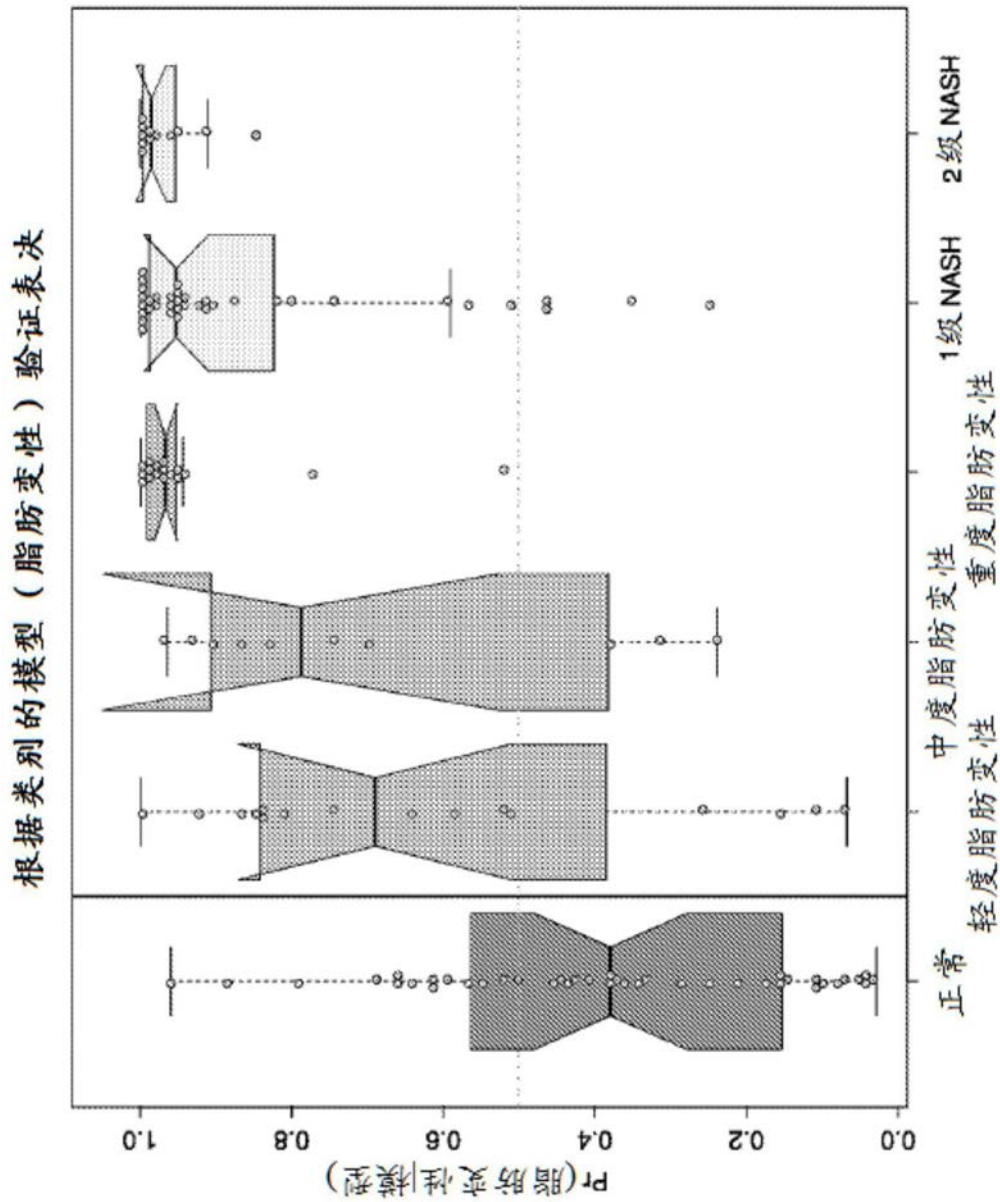


图15

脂肪变性模型性能

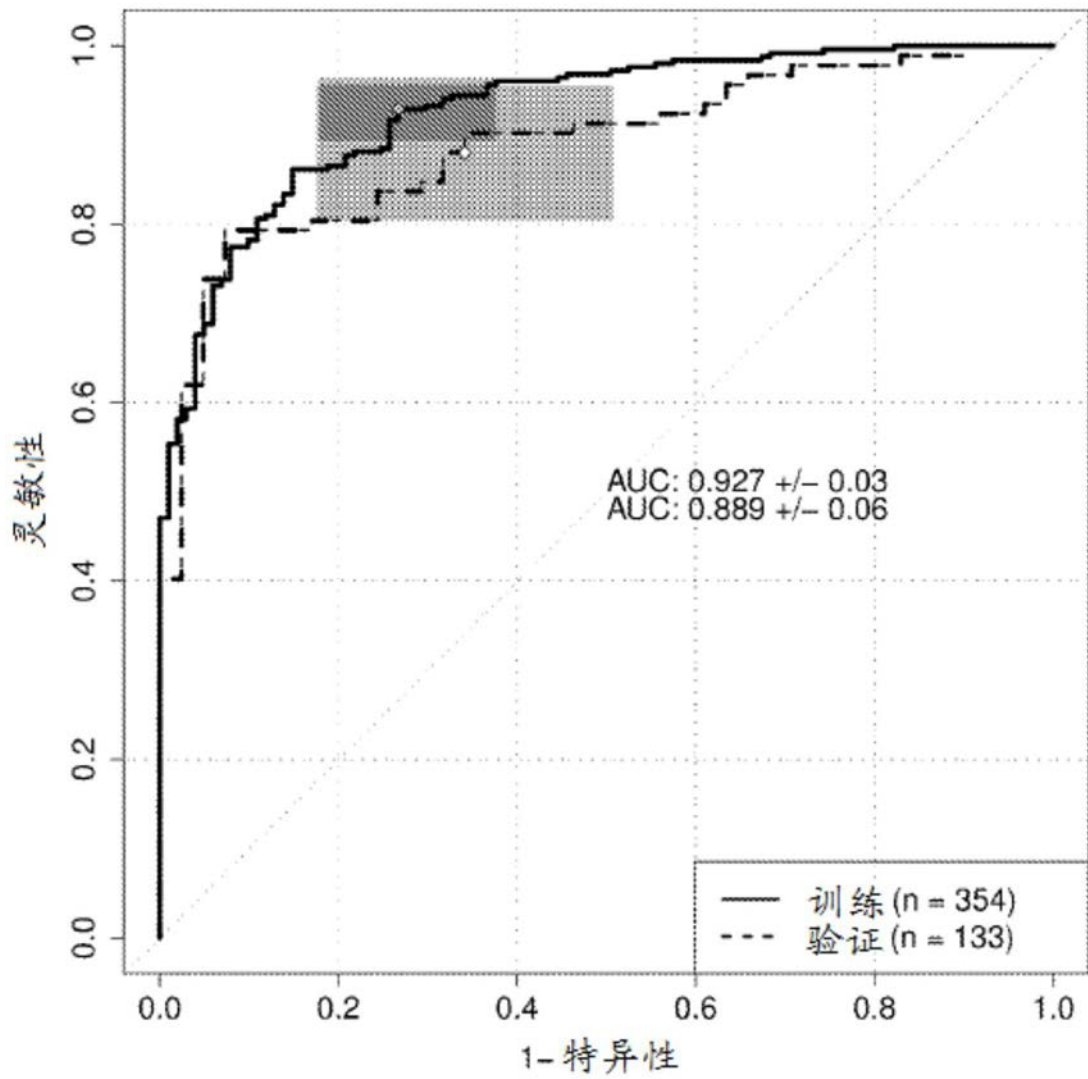


图16

根据类别的模型 (纤维化) 验证表决

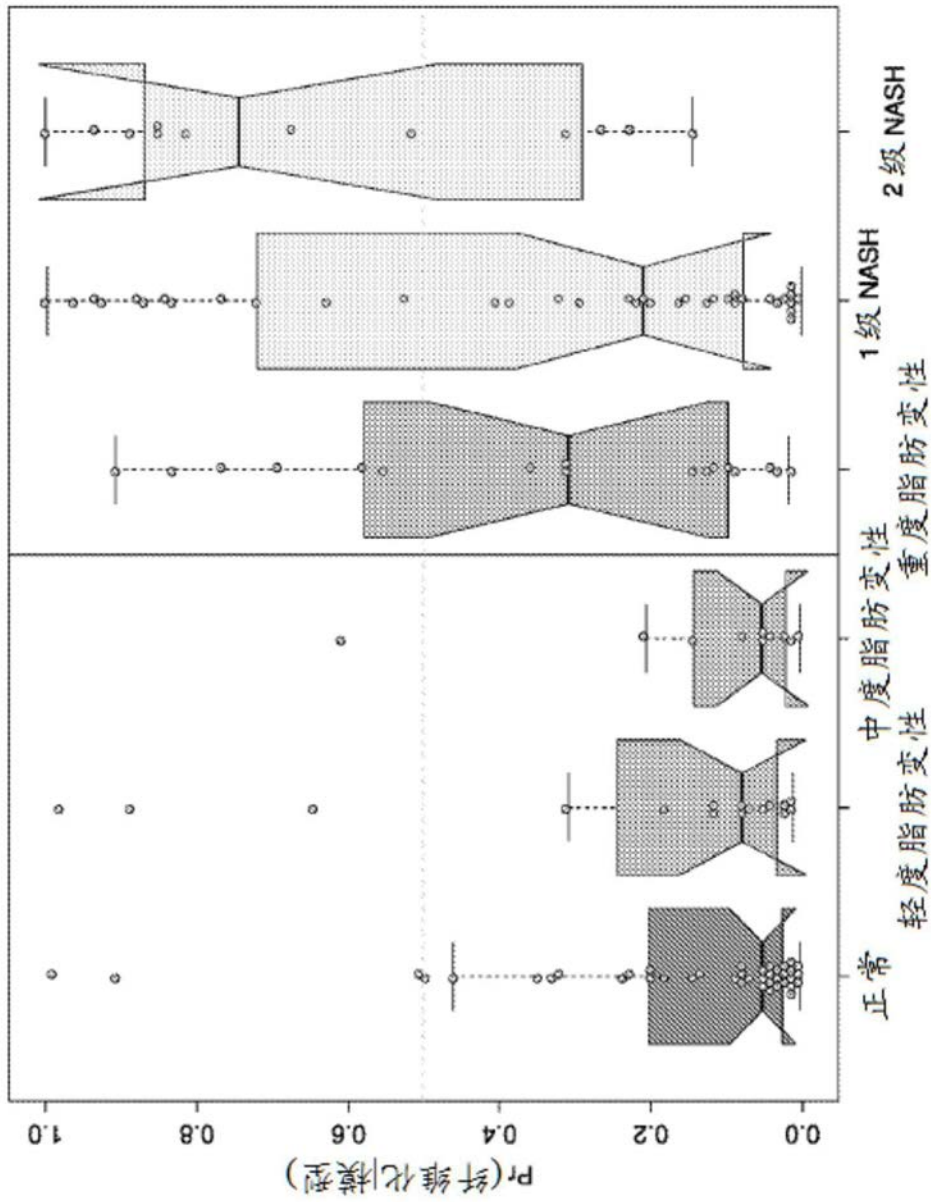


图17

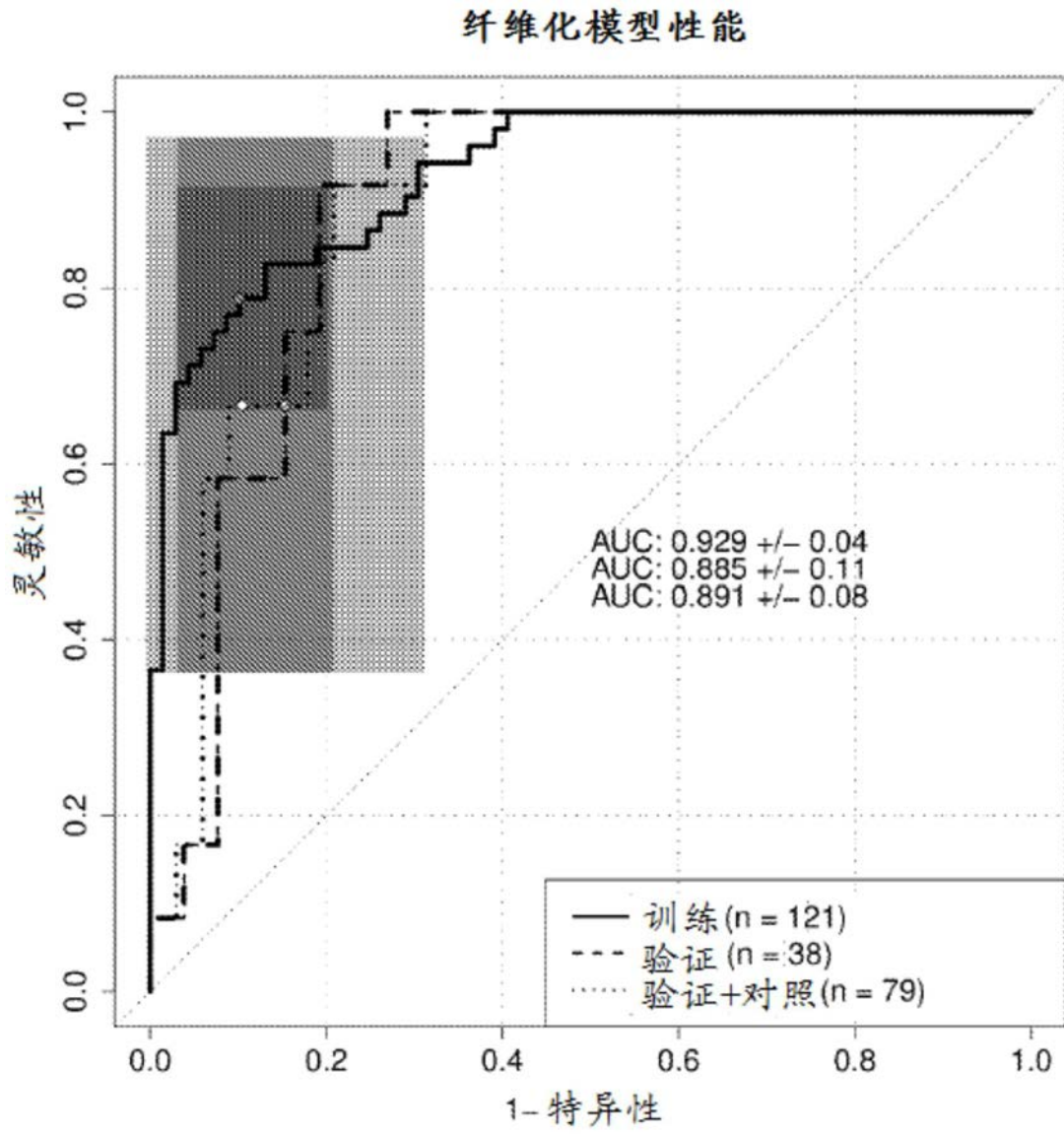


图18

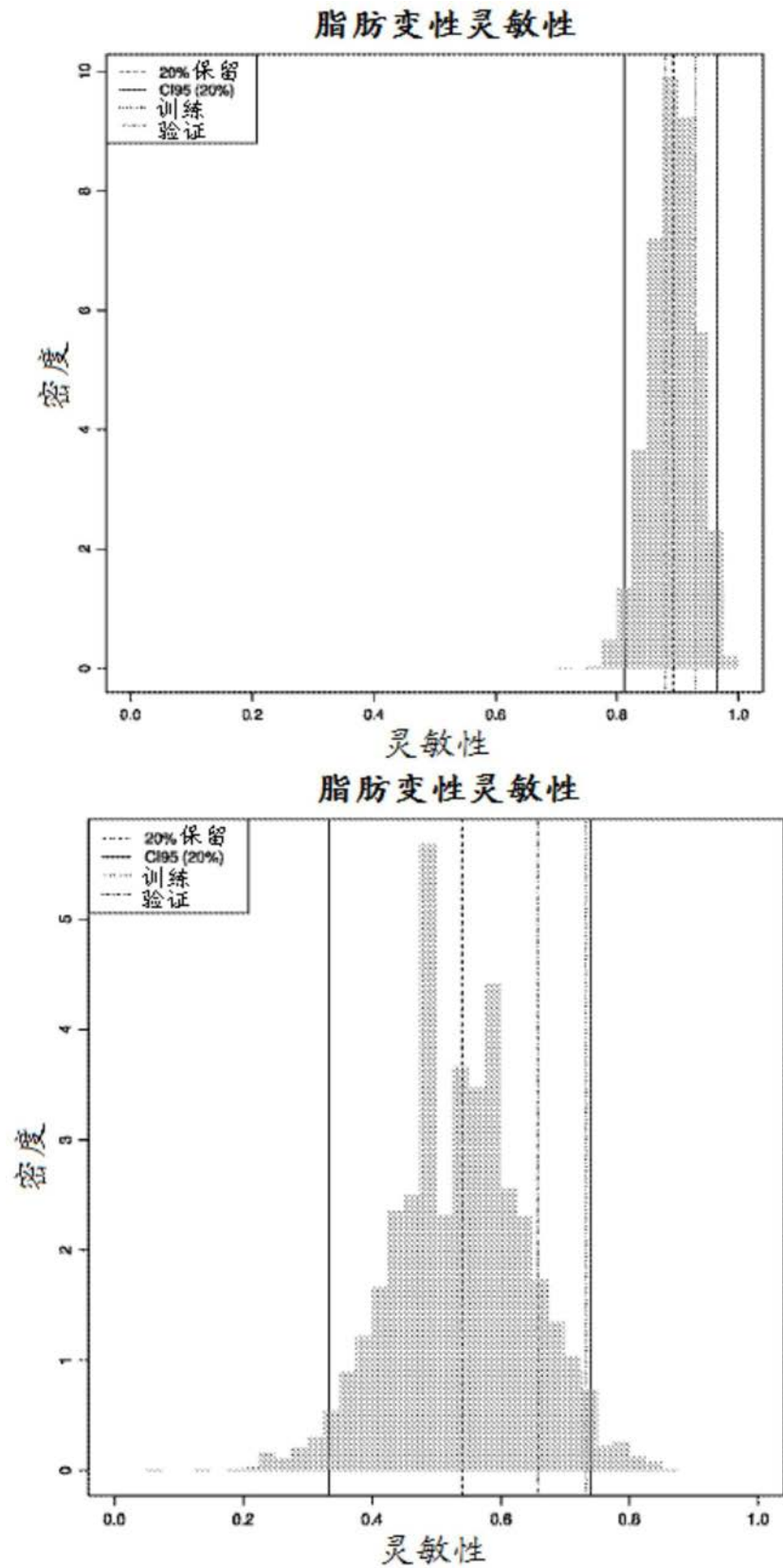


图19A

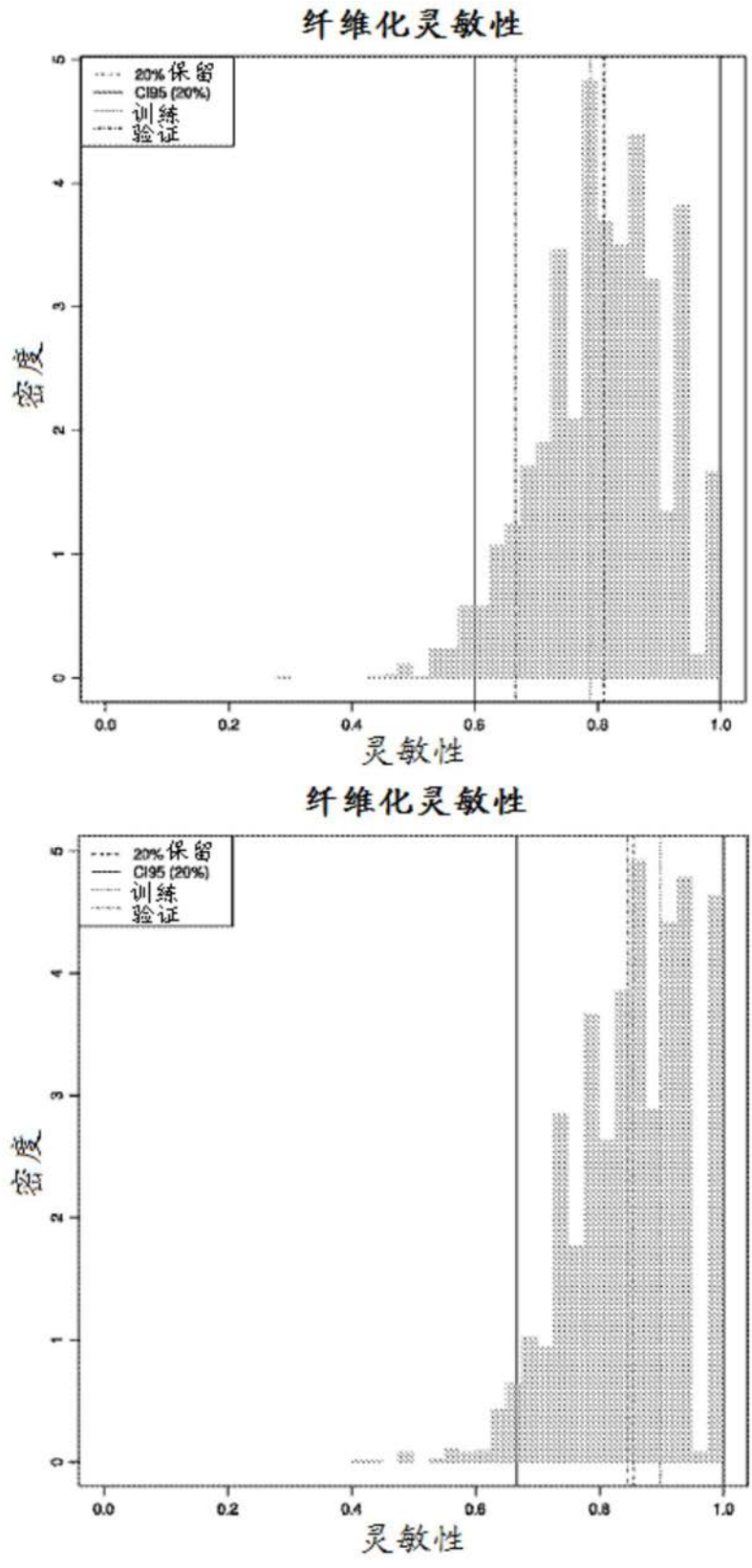


图19B

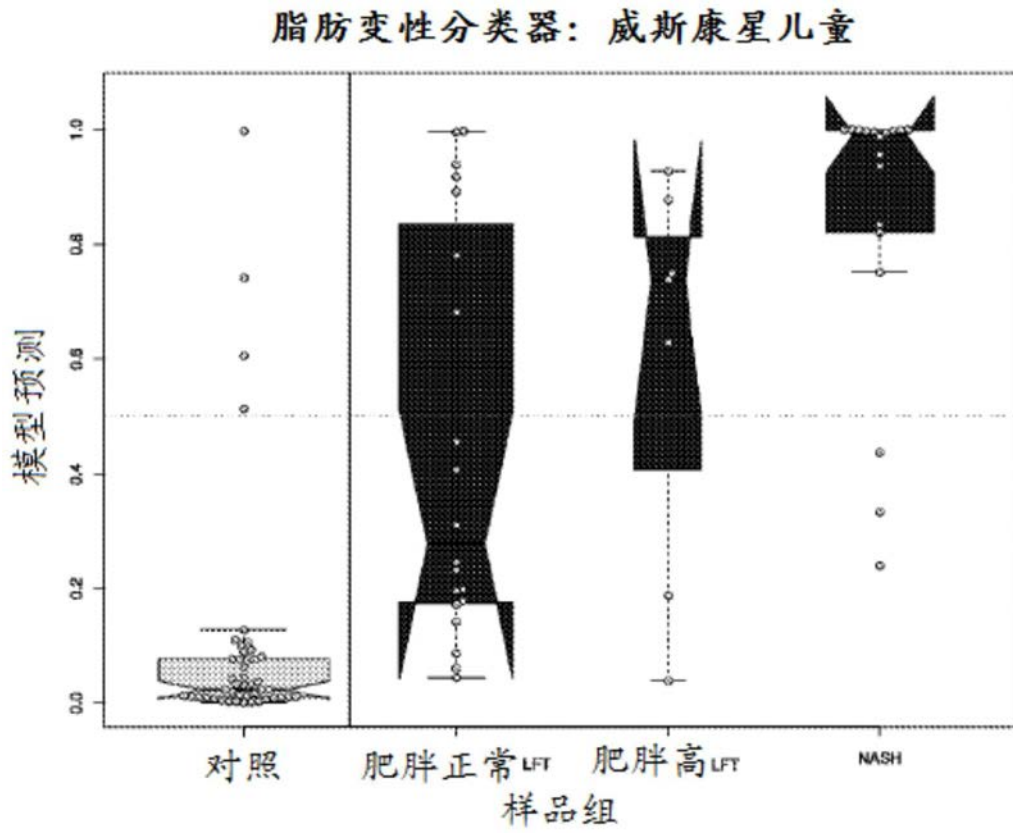


图20