



(12) 发明专利申请

(10) 申请公布号 CN 114402084 A

(43) 申请公布日 2022.04.26

(21) 申请号 202080060499.9

(22) 申请日 2020.06.26

(30) 优先权数据

62/867,853 2019.06.27 US

62/882,402 2019.08.02 US

62/965,486 2020.01.24 US

(85) PCT国际申请进入国家阶段日

2022.02.25

(86) PCT国际申请的申请数据

PCT/US2020/039991 2020.06.26

(87) PCT国际申请的公布数据

W02020/264426 EN 2020.12.30

(71) 申请人 赛福医药公司

地址 美国马萨诸塞州

(72) 发明人 苏珊·吉亚西安

西奥多·R·梅勒斯

马克·桑托里尼 阿瑟·阿梅利

南茜·肖恩布伦纳

维亚切斯拉夫·R·埃克麦弗

基斯·J·约翰逊

(74) 专利代理机构 北京安信方达知识产权代理

有限公司 11262

代理人 贺淑东

(51) Int.Cl.

G12Q 1/6883 (2018.01)

G01N 33/68 (2006.01)

权利要求书8页 说明书79页 附图25页

(54) 发明名称

开发用于对患者分层的分类器

(57) 摘要

本文呈现了用于开发分类器的系统和方法，所述分类器可用于预测对特定治疗的应答。例如，在一些实施方案中，本公开提供了一种治疗患有自身免疫性病症的受试者的方法，所述方法包括以下步骤：向已通过分类器确定为有应答的受试者施用抗TNF疗法，所述分类器被建立用于区分已接受所述抗TNF疗法的队列中有应答和无应答的先前受试者。

1. 一种治疗患有自身免疫性病症的受试者的方法,所述方法包括以下步骤:
 向已通过分类器确定为有应答的受试者施用抗TNF疗法,所述分类器被建立用于区分已接受所述抗TNF疗法的队列中有应答和无应答的先前受试者;
 其中所述分类器通过评估以下各项进行开发:
 表达水平与临床应答性或无应答性显著相关的一种或多种基因;
 以下各项中的至少一个:
 所述一种或多种基因的表达序列中一个或多个单核苷酸多态性(SNP)的存在;或
 所述有应答和无应答的先前受试者的至少一种临床特征;并且
 其中所述分类器通过不是已接受所述抗TNF疗法的所述队列的一独立队列进行验证。
2. 根据权利要求1所述的方法,其中所述分类器在至少170名受试者的群体中以至少90%的准确度预测受试者的应答性。
3. 根据权利要求1或2所述的方法,其中当映射到人类相互作用组图谱上时,所述一种或多种基因通过其拓扑性质进行表征。
4. 根据权利要求1-3中任一项所述的方法,其中参考人类基因组鉴定所述SNP。
5. 根据权利要求1-4中任一项所述的方法,其中所述分类器通过评估以下各项进行开发:
 表达水平与临床应答性或无应答性显著相关的所述一种或多种基因;
 所述一个或多个SNP的存在;以及
 所述至少一种临床特征。
6. 根据权利要求1-5中任一项所述的方法,其中所述一种或多种基因选自:

ARPC1A	LOC100506083
ATAT1	MORN2
ATRAID	NGF
CD27	RHOBTB1
CHMP7	RNGTT
COMMD5	SETD9
CORO2B	SNX8
CYSLTR2	SPINT2
DNAJC7	SPON2
GOLGA1	SSNA1
HPCA	STOML2
IMPDH2	SUOX
KAT8	TMEM258
KLHDC3	UBL7-AS1

7. 根据权利要求1-6中任一项所述的方法,其中所述至少一种临床特征选自:体重指数(BMI)、性别、年龄、种族、先前疗法治疗、疾病持续时间、C反应蛋白水平、抗环瓜氨酸肽的存在、类风湿因子的存在、患者整体评估、治疗应答率(例如,ACR20、ACR50、ACR70)以及其组合。

8. 根据权利要求1-6中任一项所述的方法,其中所述分类器以至少90%的准确度鉴定群体中无应答的至少40%的受试者,其中所述群体包括至少170名受试者。

9. 根据权利要求1-8中任一项所述的方法,其中所述抗TNF疗法包括施用英夫利昔单抗、阿达木单抗、依那西普、赛妥珠单抗、goliluma或其生物仿制药。

10. 根据权利要求1-9中任一项所述的方法,其中所述疾病、病症或病况选自类风湿性关节炎、银屑病性关节炎、强直性脊柱炎、克罗恩病、溃疡性结肠炎、慢性银屑病、化脓性汗腺炎、多发性硬化症和幼年特发性关节炎。

11. 根据权利要求1-10中任一项所述的方法,其中使用源自所述有应答和无应答的先前受试者的微阵列分析建立所述分类器。

12. 根据权利要求1-11中任一项所述的方法,其中使用源自所述独立队列的RNAseq数据验证所述分类器。

13. 根据权利要求1-12中任一项所述的方法,其中所述SNP选自表5。

14. 一种向患有自身免疫性疾病的受试者施用疗法的方法,所述方法包括:

向已通过分类器分类为有应答或无应答的受试者施用所述疗法,所述分类器被建立用于区分已接受所述抗TNF疗法的队列中有应答和无应答的先前受试者;

其中,如果所述受试者被所述分类器分类为有应答,则施用抗TNF疗法;并且如果所述受试者被所述分类器分类为无应答,则施用抗TNF疗法的替代方案;并且

所述分类器通过评估以下各项进行开发:

表达水平与临床应答性或无应答性显著相关的一种或多种基因;

以下各项中的至少一个:

所述一种或多种基因的表达序列中一个或多个单核苷酸多态性(SNP)的存在;或

所述有应答和无应答的先前受试者的至少一种临床特征;并且

所述分类器通过不是已接受所述抗TNF疗法的所述队列的一独立队列进行验证。

15. 根据权利要求14所述的方法,其中所述分类器在至少170名受试者的群体中以至少90%的准确度预测受试者的应答性。

16. 根据权利要求15所述的方法,其中所述分类器通过评估以下各项进行开发:

表达水平与临床应答性或无应答性显著相关的所述一种或多种基因;

所述一个或多个SNP的存在;以及

所述至少一种临床特征。

17. 根据权利要求14-16中任一项所述的方法,其中所述至少一种临床特征选自:体重指数(BMI)、性别、年龄、种族、先前疗法治疗、疾病持续时间、C反应蛋白水平、抗环瓜氨酸肽的存在、类风湿因子的存在、患者整体评估、治疗应答率(例如,ACR20、ACR50、ACR70)以及其组合。

18. 根据权利要求14-17中任一项所述的方法,其中所述抗TNF疗法包括施用英夫利昔单抗、阿达木单抗、依那西普、赛妥珠单抗、goliluma或其生物仿制药。

19. 根据权利要求14-18中任一项所述的方法,其中抗TNF疗法的替代方案选自利妥昔单抗、沙利姆单抗、枸橼酸托法替尼、来氟米特、维多珠单抗、托珠单抗、阿那白滞素和阿巴西普。

20. 根据权利要求19所述的方法,其中所述疾病、病症或病况选自类风湿性关节炎、银

屑病性关节炎、强直性脊柱炎、克罗恩病、溃疡性结肠炎、慢性银屑病、化脓性汗腺炎、多发性硬化症和幼年特发性关节炎。

21. 一种开发分类器以用于针对一个或多个治疗属性对受试者分层的方法，其包括以下步骤：

分析在针对至少一个所述治疗属性代表至少两个不同类别的受试者中表达的RNA的序列数据；

由所述序列数据评估一个或多个单核苷酸多态性 (SNP) 的存在；

确定所述一个或多个SNP的存在与所述至少一个治疗属性相关；以及

将所述一个或多个SNP包括在所述分类器中。

22. 根据权利要求21所述的方法，其中所述一个或多个治疗属性选自以下各项组成的组：发展特定疾病、病症或病况的风险；特定疾病、病症或病况的特定结果的可能性；对特定疗法产生应答的可能性。

23. 根据权利要求21或22所述的方法，其中所述一个或多个治疗属性是对特定疗法产生应答的可能性。

24. 根据权利要求21-23中任一项所述的方法，其中所述特定疗法为抗TNF疗法。

25. 开发分类器的方法，所述分类器用于通过分析在对于至少一个治疗属性代表至少两个不同类别的受试者中表达的RNA的序列数据，针对所述一个或多个治疗属性对受试者分层，所述方法的改进包括：

由所述序列数据评估一个或多个单核苷酸多态性 (SNP) 的存在；以及

确定所述一个或多个SNP的存在与所述至少一个治疗属性相关；以及

将所述一个或多个SNP包括在所述分类器中。

26. 根据权利要求25所述的方法，其中所述一个或多个治疗属性选自以下各项组成的组：发展特定疾病、病症或病况的风险；特定疾病、病症或病况的特定结果的可能性；对特定疗法产生应答的可能性。

27. 根据权利要求25或26所述的方法，其中所述一个或多个治疗属性是对特定疗法产生应答的可能性。

28. 根据权利要求25-27中任一项所述的方法，其中所述特定疗法为抗TNF疗法。

29. 根据权利要求25-28中任一项所述的方法，其中评估所述一个或多个SNP的存在包括将RNA的所述序列数据与参考人类基因组进行比较。

30. 一种治疗患有自身免疫性疾病的受试者的方法，所述方法包括以下步骤：

(a) 通过计算设备的处理器接收与所述受试者的一种或多种基因中的每一种的表达水平相对应的数据，其中所述基因选自表1；

(b) 使用在步骤 (a) 中接收的所述数据，由所述处理器自动确定所述受试者为对抗TNF疗法有应答或无应答的分类；以及，任选地，

(c) 如果所述受试者被分类为对所述抗TNF疗法无应答，则向所述受试者施用抗TNF疗法的替代方案以治疗所述自身免疫性疾病；或者，如果所述受试者被分类为对所述抗TNF疗法有应答，则施用所述抗TNF疗法。

31. 根据权利要求30所述的方法，其中步骤 (a) 中的所述处理器进一步接收与以下各项中的至少一个相对应的数据：

一种或多种临床特征;或
一个或多个单核苷酸多态性(SNP)。

32. 根据权利要求30或31所述的方法,其中所述一种或多种临床特征选自:体重指数(BMI)、性别、年龄、种族、先前疗法治疗、疾病持续时间、C反应蛋白(CRP)水平、抗环瓜氨酸肽的存在、类风湿因子的存在、患者整体评估、治疗应答率(例如,ACR20、ACR50、ACR70)以及其组合。

33. 根据权利要求30-32中任一项所述的方法,其中所述自身免疫性疾病选自类风湿性关节炎、银屑病性关节炎、强直性脊柱炎、克罗恩病、溃疡性结肠炎、慢性银屑病、化脓性汗腺炎和幼年特发性关节炎。

34. 根据权利要求30-33中任一项所述的方法,其中所述自身免疫性疾病为类风湿性关节炎或溃疡性结肠炎。

35. 根据权利要求30-34中任一项所述的方法,其中抗TNF疗法的替代方案选自利妥昔单抗、沙利姆单抗、枸橼酸托法替尼、来氟米特、维多珠单抗、托珠单抗、阿那白滞素和阿巴西普。

36. 根据权利要求30-34中任一项所述的方法,其中所述抗TNF疗法选自英夫利昔单抗、阿达木单抗、依那西普、赛妥珠单抗、goliluma或其生物仿制药。

37. 根据权利要求30-36中任一项所述的方法,其中步骤(b)包括使用机器学习模型自动确定所述分类。

38. 根据权利要求37所述的方法,其中所述机器学习模型是随机森林模型。

39. 根据权利要求30-38中任一项所述的方法,其中步骤(b)包括在不使用临床协变量分类器(CC)的情况下自动确定所述分类。

40. 根据权利要求30-38中任一项所述的方法,其中步骤(b)包括在不使用组合基因组-临床分类器(GCC)的情况下自动确定所述分类。

41. 根据权利要求30-38中任一项所述的方法,其中步骤(b)包括在不使用临床协变量分类器(CC)和不使用组合基因组-临床分类器(GCC)的情况下自动确定所述分类。

42. 一种方法,其包括以下步骤:

(a) 通过计算设备的处理器接收与患有疾病(例如,自身免疫性疾病,例如,类风湿性关节炎)的受试者的一种或多种基因中的每一种的表达水平相对应的数据,所述一种或多种基因包括选自由CORO2B、STOML2、CHMP7、SPINT2、CYSLTR2、SSNA1、IMPDH2、GOLGA1、SUOX、ARPC1A和KLHDC3组成的组的至少一个成员(例如,一个、两个、三个、四个、五个、六个、七个、八个、九个、十个或全部十一个成员);

(b) 使用在步骤(a)中接收的所述数据由所述处理器自动确定所述受试者为对第一疗法(例如,抗TNF疗法)无应答的分类;以及,任选地,

(c) 向所述受试者开处方和/或施用第二疗法(例如,所述第一疗法的替代方案,例如,抗TNF疗法的替代方案)以治疗所述疾病,从而避免向所述受试者开处方和/或施用所述第一疗法。

43. 一种用抗TNF疗法治疗受试者的方法,所述方法包括以下步骤:

向已通过分类器确定为有应答的受试者施用所述抗TNF疗法,所述分类器被建立用于区分已接受所述抗TNF疗法的有应答和无应答的先前受试者;

其中所述分类器通过评估以下各项进行开发：

表达水平与临床应答性或无应答性显著相关的一种或多种基因；以及

以下各项中的至少一个：

所述一种或多种基因的表达序列中一个或多个单核苷酸多态性 (SNP) 的存在；或

所述有应答和无应答的先前受试者的至少一种临床特征。

44. 一种在对患有自身免疫性疾病的受试者进行抗TNF疗法的任何施用之前将所述受试者分类为对所述抗TNF疗法可能有应答或可能无应答的方法，所述方法包括以下步骤：

(a) 通过计算设备的处理器接收与选自由以下各项组成的组的一种或多种基因中的每一种的所述受试者的表达水平相对应的数据：

ARPC1A	LOC100506083
ATAT1	MORN2
ATRAID	NGF
CD27	RHOBTB1
CHMP7	RNGTT
COMMD5	SETD9
CORO2B	SNX8
CYSLTR2	SPINT2
DNAJC7	SPON2
GOLGA1	SSNA1
HPCA	STOML2
IMPDH2	SUOX
KAT8	TMEM258
KLHDC3	UBL7-AS1

以及

(b) 使用在步骤 (a) 中接收的所述数据，由所述处理器自动确定所述受试者为对所述抗TNF疗法有应答或无应答的分类。

45. 根据权利要求44所述的方法，其还包括：

(c) 如果所述受试者在步骤 (b) 中被分类为对所述抗TNF疗法无应答，则为所述受试者开处方和/或施用抗TNF疗法的替代方案以治疗所述自身免疫性疾病；或者，如果所述受试者在步骤 (b) 中被分类为对所述抗TNF疗法有应答，则向所述受试者施用所述抗TNF疗法。

46. 根据权利要求44或45所述的方法，其中步骤 (a) 中的所述处理器进一步接收与如下 (i) 和 (ii) 中至少一个相对应的数据：

(i) 所述受试者的一种或多种临床特征；

(ii) 所述受试者的一个或多个单核苷酸多态性 (SNP)。

47. 根据权利要求44至46中任一项所述的方法，其中步骤 (a) 中的所述处理器接收与所述受试者的一种或多种临床特征相对应的数据，所述临床特征选自由以下各项组成的组：体重指数 (BMI)、性别、年龄、种族、先前疗法治疗、疾病持续时间、C反应蛋白 (CRP) 水平、抗环瓜氨酸肽的存在、类风湿因子的存在、患者整体评估以及治疗应答率 (例如，ACR20、ACR50、ACR70)。

48. 根据权利要求44至47中任一项所述的方法,其中步骤(a)中的所述处理器接收与表5中列出的一个或多个SNP相对应的数据。

49. 根据权利要求44至48中任一项所述的方法,其中所述自身免疫性疾病是选自由类风湿性关节炎、银屑病性关节炎、强直性脊柱炎、克罗恩病、溃疡性结肠炎、慢性银屑病、化脓性汗腺炎和幼年特发性关节炎组成的组的成员。

50. 根据权利要求44至49中任一项所述的方法,其中所述自身免疫性疾病为类风湿性关节炎或溃疡性结肠炎。

51. 根据权利要求44至50中任一项所述的方法,其中抗TNF疗法的替代方案包括选自由利妥昔单抗、沙利姆单抗、枸橼酸托法替尼、来氟米特、维多珠单抗、托珠单抗、阿那白滞素和阿巴西普组成的组的至少一个成员。

52. 根据权利要求44至50中任一项所述的方法,其中所述抗TNF疗法包括选自由英夫利昔单抗、阿达木单抗、依那西普、赛妥珠单抗、goliluma以及前述中任一个的生物仿制药组成的组的至少一个成员。

53. 根据权利要求44至52中任一项所述的方法,其中抗TNF疗法的替代方案不包括选自由英夫利昔单抗、阿达木单抗、依那西普、赛妥珠单抗、goliluma以及前述中任一个的任何生物仿制药组成的组的任何成员。

54. 根据权利要求44至53中任一项所述的方法,其中步骤(b)包括使用机器学习模型自动确定所述分类。

55. 根据权利要求54所述的方法,其中所述机器学习模型是随机森林模型。

56. 根据权利要求44至55中任一项所述的方法,其中步骤(b)包括在不使用临床协变量分类器(CC)的情况下自动确定所述分类。

57. 根据权利要求44至55中任一项所述的方法,其中步骤(b)包括在不使用组合基因组-临床分类器(GCC)的情况下自动确定所述分类。

58. 根据权利要求44至55中任一项所述的方法,其中步骤(b)包括在不使用临床协变量分类器(CC)和不使用组合基因组-临床分类器(GCC)的情况下自动确定所述分类。

59. 一种在对患有自身免疫性疾病的受试者进行抗TNF疗法的任何施用之前将所述受试者分类为对所述抗TNF疗法可能有应答或可能无应答的方法,所述方法包括以下步骤:

(a) 通过计算设备的处理器接收与患有所述自身免疫性疾病的受试者的一种或多种基因中的每一种的表达水平相对应的数据,所述一种或多种基因包括选自由CORO2B、STOML2、CHMP7、SPINT2、CYSLTR2、SSNA1、IMPDH2、GOLGA1、SUOX、ARPC1A和KLHDC3组成的组的至少一个成员;

(b) 使用在步骤(a)中接收的所述数据,由所述处理器自动确定所述受试者为对所述抗TNF疗法可能有应答或可能无应答的分类。

60. 根据权利要求59所述的方法,其还包括:

(c) 如果所述受试者在步骤(b)中被分类为对所述抗TNF疗法无应答,则为所述受试者开处方和/或施用抗TNF疗法的替代方案以治疗所述自身免疫性疾病;或者,如果所述受试者在步骤(b)中被分类为对所述抗TNF疗法有应答,则向所述受试者施用所述抗TNF疗法。

61. 根据权利要求59或60所述的方法,其中所述自身免疫性疾病为类风湿性关节炎。

62. 根据权利要求59至61中任一项所述的方法,其中步骤(a)包括接收与选自由

CORO2B、STOML2、CHMP7、SPINT2、CYSLTR2、SSNA1、IMPDH2、GOLGA1、SUOX、ARPC1A和KLHDC3组成的组的至少两种基因中的每一种的表达水平相对应的数据。

63. 根据权利要求59至61中任一项所述的方法,其中步骤(a)包括接收与选自由CORO2B、STOML2、CHMP7、SPINT2、CYSLTR2、SSNA1、IMPDH2、GOLGA1、SUOX、ARPC1A和KLHDC3组成的组的至少三种基因中的每一种的表达水平相对应的数据。

64. 根据权利要求59至61中任一项所述的方法,其中步骤(a)包括接收与选自由CORO2B、STOML2、CHMP7、SPINT2、CYSLTR2、SSNA1、IMPDH2、GOLGA1、SUOX、ARPC1A和KLHDC3组成的组的至少五种基因中的每一种的表达水平相对应的数据。

65. 根据权利要求59至64中任一项所述的方法,其中步骤(a)中的所述处理器进一步接收与如下(i)和(ii)中至少一个相对应的数据:

(i) 所述受试者的一种或多种临床特征;

(ii) 所述受试者的一个或多个单核苷酸多态性(SNP)。

66. 根据权利要求59至65中任一项所述的方法,其中步骤(a)中的所述处理器接收与受试者的一种或多种临床特征相对应的数据,所述临床特征选自由以下各项组成的组:体重指数(BMI)、性别、年龄、种族、先前疗法治疗、疾病持续时间、C反应蛋白(CRP)水平、抗环瓜氨酸的存在、类风湿因子的存在、患者整体评估以及治疗应答率(例如,ACR20、ACR50、ACR70)。

67. 根据权利要求59至66中任一项所述的方法,其中步骤(a)中的所述处理器接收与表5中列出的一个或多个SNP相对应的数据。

68. 一种在对患有自身免疫性疾病的受试者进行抗TNF疗法的任何施用之前将所述受试者分类为对所述抗TNF疗法可能有应答或可能无应答的系统,所述系统包括:

处理器;以及

其上具有指令的存储器,所述指令在由所述处理器执行时,使所述处理器:

(a) 接收数据集,所述数据集包括选自由以下各项组成的组的一种或多种基因中的每一种的所述受试者的表达水平:

ARPC1A	LOC100506083
ATAT1	MORN2
ATRAID	NGF
CD27	RHOBTB1
CHMP7	RNGTT

COMMD5	SETD9
CORO2B	SNX8
CYSLTR2	SPINT2
DNAJC7	SPON2
GOLGA1	SSNA1
HPCA	STOML2
IMPDH2	SUOX
KAT8	TMEM258
KLHDC3	UBL7-AS1

以及

(b) 使用所述数据集,自动确定所述受试者为对所述抗TNF疗法有应答或无应答的分类。

69. 根据权利要求68所述的系统,其中所述数据集还包括与如下(i)和(ii)中至少一个相对应的数据:

(i) 所述受试者的一种或多种临床特征;

(ii) 所述受试者的一个或多个单核苷酸多态性(SNP)。

70. 根据权利要求69所述的系统,其中所述数据集包括与所述受试者的一种或多种临床特征相对应的数据,所述临床特征选自由以下各项组成的组:体重指数(BMI)、性别、年龄、种族、先前疗法治疗、疾病持续时间、C反应蛋白(CRP)水平、抗环瓜氨酸肽的存在、类风湿因子的存在、患者整体评估以及治疗应答率(例如,ACR20、ACR50、ACR70)。

71. 根据权利要求69或70所述的系统,其中所述数据集包括表5中列出的一个或多个SNP。

72. 根据权利要求69至71中任一项所述的系统,其中所述自身免疫性疾病是选自由类风湿性关节炎、银屑病性关节炎、强直性脊柱炎、克罗恩病、溃疡性结肠炎、慢性银屑病、化脓性汗腺炎和幼年特发性关节炎组成的组的成员。

73. 根据权利要求69至71中任一项所述的系统,其中所述自身免疫性疾病为类风湿性关节炎或溃疡性结肠炎。

74. 根据权利要求69至71中任一项所述的系统,其中所述自身免疫性疾病为类风湿性关节炎。

开发用于对患者分层的分类器

相关申请的交叉引用

[0001] 本申请要求2020年1月24日提交的美国临时申请第62/965,486号、2019年6月27日提交的美国临时申请第62/867,853号以及2019年8月2日提交的美国临时申请第62/882,402号的优先权,其各自特此通过引用整体并入。

[0002] 本专利申请的主题与2019年3月15日提交的标题为“用于预测对抗TNF疗法的应答的方法和系统(Methods and Systems for Predicting Response to Anti-TNF Therapies)”的待决国际专利申请第PCT/US19/22588号有关,所述申请要求于2018年3月16日提交的美国临时申请第62/644,070号的优先权和权益,其各自的内容特此通过引用整体并入。

背景技术

[0003] 诸如类风湿性关节炎(RA)的自身免疫性疾病影响数百万患者,并且其治疗是整体医疗支出的重要组成部分。自身免疫性疾病可以分为两组——器官特异性和系统性自身免疫。包括RA在内的类风湿性疾病属于系统性自身免疫性疾病,其主要表现在滑膜关节并最终导致肌腱、软骨和骨的不可逆破坏。虽然目前尚无治愈RA的方法,但主要通过用于中和这种细胞因子的促炎性信号传导的抗TNF(肿瘤坏死因子)剂的开发在管理这些患者的治疗方面取得了重大改善。此类生物疗法(例如,Humira、Enbrel、Remicade、Simponi和Cymzia)显著改善了一些RA患者的治疗结果。

[0004] 大约34%的RA患者(低百分比)表现出对抗TNF疗法的临床应答,达到低疾病活动度(LDA),有时达到缓解。这些所谓的“应答者”患者的疾病进展可能是由不适当的TNF驱动的促炎性应答引起的。对于对抗TNF无应答的患者,存在可用的替代性批准疗法,诸如抗CD20、共刺激阻断、JAK和抗IL6疗法。然而,患者通常只有在首先循环使用不同的抗TNF药物(这可能需要一年多的时间)后才转为使用这种替代疗法,而症状持续存在并且疾病进一步发展,使其更难达到治疗目标。

[0005] 除了治疗延迟的问题外,与抗TNF疗法相关的严重感染和恶性肿瘤的已知风险非常大,以致于产品批准通常需要在标签上包括所谓的“黑盒子警告”。这种疗法的其他潜在副作用包括,例如,充血性心力衰竭、脱髓鞘疾病和其他系统性副作用。

[0006] 此外,精准医疗依赖于使用专有算法解释基因组和/或多组学数据的能力,所述专有算法是通过使用组学数据和临床结果利用机器学习、人工智能和网络科学方法开发的。然而,一个主要挑战是分析组学数据的软件正在不断改进和升级。考虑到需要在临床实验室环境中锁定和验证系统(包括软件组件),此类升级极难及时实施。

发明内容

[0007] 抗TNF疗法的一个重要的已知问题是应答率不一致。事实上,最近的国际会议旨在汇集免疫学和风湿病学领域的顶尖科学家和临床医生,以确定这些领域中未满足的需求,几乎普遍认为应答率的不确定性是一个持续的挑战。例如,第19届年度国际靶向疗法会议,

该会议举行了与多种疾病治疗中的挑战相关的分会讨论,包括类风湿性关节炎、银屑病性关节炎、中轴型脊柱关节炎、系统性红斑狼疮和结缔组织疾病(例如,干燥综合征、系统性硬化症、血管炎,包括白塞病和IgG4相关疾病),确定了所有这些疾病的某些常见问题,具体地,“需要更好地了解每种疾病内的异质性……因此可以开发治疗应答的预测工具。参见 Winthrop 等人,“The unmet need in rheumatology: Reports from the targeted therapies meeting 2017,” Clin. Immunol. pii: S1521-6616(17) 30543-0, 2017年8月12日。类似地,关于用抗TNF疗法治疗克罗恩病的大量文献一直抱怨应答率不稳定并且无法预测哪些患者将受益。参见,例如, M.T. Abreu, “Anti-TNF Failures in Crohn’s Disease,” Gastroenterol Hepatol (N.Y.), 7(1): 37-39 (2011年1月); 另外参见 Ding 等人, “Systematic review: predicting and optimising response to anti-TNF therapy in Crohn’s disease-algorithm for practical management,” Aliment Pharmacol. Ther., 43(1): 30-51 (2016年1月) (报告“抗TNF治疗的原发性无应答影响13-40%的患者。”)。

[0008] 所提供的技术,除其他外,允许护理提供者在受试者类别之间或之中进行区分——例如,可能受益于特定疗法(例如,抗TNF疗法)的受试者与不可受益于所述疗法的受试者、更可能实现或遭受特定结果或副作用的受试者等。在一些实施方案中,所提供的此类技术因此降低了患者的风险、增加了无应答患者群体的护理时间和质量、提高了药物开发的效率,和/或避免了与对无应答患者施用无效疗法或与治疗此类患者在接受相关疗法(例如,抗TNF疗法)后所经历的副作用相关的费用。

[0009] 在一些实施方案中,本公开提供了一种用特定疗法(例如,抗TNF疗法)治疗受试者的方法,所述方法包括以下步骤:向已通过分类器确定为有应答的受试者施用所述疗法,所述分类器被建立用于区分预期对所述疗法有应答的受试者与无应答的受试者。

[0010] 除其他事项外,本文进一步描述了可用于确定受试者对某些疗法有应答还是无应答和/或是否实现或遭受特定结果或副作用的分类器的实施方案,以及制备这种分类器的方法。

[0011] 除其他事项外,所提供的技术体现和/或产生于某些见解,包括,例如,用某些常规方法确定问题的根源,以定义应答者与无应答者群体,和/或代表特别有用的策略以用于定义区分此类群体的分类器。

[0012] 例如,如本文所述,本公开利用了W0 2019/178546提供的见解,即,通过考虑群体中的基因表达差异来定义应答者与无应答者群体的许多传统策略的问题的一个根源是,他们通常优先考虑或以其他方式关注最高倍数(即,显著性)变化;如W0 2019/178546中所述,这种方法忽略了与疾病生物学相关的细微但有意义的差异。此外,本公开还利用了也在W0 2019/178546中描述的见解,即,将表达水平改变的基因映射到人类相互作用组图谱上(特别是在表示细胞组分之间的实验支持的物理相互作用的人类相互作用组图谱上,在一些实施方案中,其明确排除已提出但未经实验验证的任何理论、计算得到的或其他相互作用),可以提供有用且有效的分类器,以用于定义对某些疗法,尤其是抗TNF疗法的有应答者与无应答者。在一些实施方案中,包括在这种分类器中的基因代表人类相互作用组上的连接模块。

[0013] 本公开还提供了这样的见解,即,通过使用跨平台数据开发特别有用的治疗分类器。例如,本公开教导,将差异基因表达分析与表达序列的序列分析(例如,单核苷酸多态性

(SNP)分析)相结合在治疗分类器的开发中特别有用(即,当/如果施用相关疗法时,预测一个或多个受试者将产生应答和/或将实现或遭受一个或多个特定结果或副作用等的可能性的分类器)。

[0014] 各种疗法(例如,抗TNF疗法)的一个重要的已知问题是应答率不一致。事实上,最近的国际会议旨在汇集免疫学和风湿病学领域的顶尖科学家和临床医生,以确定这些领域中未满足的需求,几乎普遍认为应答率的不确定性是一个持续的挑战。例如,第19届年度国际靶向治疗会议,该会议举行了与多种疾病治疗中的挑战相关的分会讨论,包括类风湿性关节炎、银屑病性关节炎、中轴型脊柱关节炎、系统性红斑狼疮和结缔组织疾病(例如,干燥综合征、系统性硬化症、血管炎,包括白塞病和IgG4相关疾病),确定了所有这些疾病的某些常见问题,具体地,“需要更好地了解每种疾病内的异质性……因此可以开发治疗应答的预测工具。参见Winthrop等人,“The unmet need in rheumatology: Reports from the targeted therapies meeting 2017,”*Clin. Immunol. pii: S1521-6616 (17) 30543-0*, 2017年8月12日。类似地,关于用抗TNF疗法治疗克罗恩病的大量文献一直抱怨应答率不稳定并且无法预测哪些患者将受益。参见,例如, M. T. Abreu, “Anti-TNF Failures in Crohn’s Disease,” *Gastroenterol Hepatol (N.Y.)*, 7 (1): 37-39 (2011年1月); 另外参见Ding等人, “Systematic review: predicting and optimising response to anti-TNF therapy in Crohn’s disease-algorithm for practical management,” *Aliment Pharmacol Ther.*, 43 (1): 30-51 (2016年1月) (报告“抗TNF治疗的原发性无应答影响13.40%的患者。”)。

[0015] 因此,目前正在施用抗TNF疗法的相当一部分患者并未从治疗中获益,甚至可能受到伤害。与抗TNF疗法相关的严重感染和恶性肿瘤的已知风险非常大,以致于产品批准通常需要在标签上包括所谓的“黑盒子警告”。这种疗法的其他潜在副作用包括,例如,充血性心力衰竭、脱髓鞘疾病和其他系统性副作用。此外,鉴于在将患者鉴定为对抗TNF疗法无应答(即,是抗TNF疗法的无应答者)之前需要数周至数月的治疗,由于目前无法鉴定应答与无应答受试者,此类患者的适当治疗可能会显著延迟。参见,例如, Roda等人, “Loss of Response to Anti-TNFs: Definition, Epidemiology, and Management,” *Clin. Tranl. Gastroenterol.*, 7 (1): e135 (2016年1月) (引用Hanauer等人, “ACCENT I Study group. Maintenance Infliximab for Crohn’s disease: the ACCENT I randomized trial,” *Lancet* 59: 1541-1549 (2002); Sands等人, “Infliximab maintenance therapy for fistulizing Crohn’s disease,” *N. Engl. J. Med.* 350: 876-885 (2004))。

[0016] 因此,在一些实施方案中,本公开提供了一种用抗TNF疗法治疗受试者的方法,所述方法包括以下步骤:向已通过分类器确定为有应答的受试者施用抗TNF疗法,所述分类器被建立用于区分已接受抗TNF疗法的有应答和无应答的先前受试者,其中所述分类器通过评估以下各项进行开发:表达水平与临床应答性或无应答性显著相关(例如,以线性和/或非线性方式)的一种或多种基因;以及以下各项中的至少一个:一种或多种基因的表达序列中一个或多个单核苷酸多态性(SNP)的存在;或有应答和无应答的先前受试者的至少一种临床特征。

[0017] 在一些实施方案中,本公开提供了一种治疗炎症性疾病、病症或病况的方法,其包括以下步骤:向通过应用分类器分类为有应答的受试者施用抗TNF疗法,所述分类器被确定为在至少170名受试者的群体中以至少90%的准确度预测受试者的应答性。

[0018] 在一些实施方案中,本公开提供了一种开发分类器以用于针对一个或多个治疗属性对受试者分层的方法,其包括以下步骤:分析针对至少一个治疗属性代表至少两个不同类别的受试者中表达的RNA的序列数据;由序列数据评估一个或多个单核苷酸多态性(SNP)的存在;确定一个或多个SNP的存在与至少一个治疗属性相关;以及将一个或多个SNP包括在分类器中。

[0019] 在一些实施方案中,本公开提供了开发分类器以用于通过分析针对至少一个治疗属性代表至少两个不同类别的受试者中表达的RNA的序列数据来针对一个或多个治疗属性对受试者分层的方法的改进,其包括:由序列数据评估一个或多个单核苷酸多态性(SNP)的存在;确定一个或多个SNP的存在与至少一个治疗属性相关;以及将一个或多个SNP包括在分类器中。

[0020] 在一些实施方案中,本公开提供了一种治疗患有自身免疫性病症的受试者的方法,所述方法包括以下步骤:向已通过分类器确定为有应答的受试者施用抗TNF疗法,所述分类器被建立用于区分已接受抗TNF疗法的队列(cohort)中有应答和无应答的先前受试者,其中所述分类器通过评估以下各项进行开发:表达水平与临床应答性或无应答性显著相关(例如,以线性和/或非线性方式)的一种或多种基因;以下各项中的至少一个:一种或多种基因的表达序列中一个或多个单核苷酸多态性(SNP)的存在;或有应答和无应答的先前受试者的至少一种临床特征;并且其中分类器通过不是已接受所述抗TNF疗法的队列的独立队列进行验证。

[0021] 在一些实施方案中,分类器在至少170名受试者的群体中以至少90%的准确度预测受试者的应答性。在一些实施方案中,当映射到人类相互作用组图谱上时,一种或多种基因通过其拓扑性质进行表征。在一些实施方案中,参考人类基因组鉴定SNP。在一些实施方案中,通过评估以下各项开发分类器:表达水平与临床应答性或无应答性显著相关(例如,以线性和/或非线性方式)的一种或多种基因;一个或多个SNP的存在;以及至少一种临床特征。

[0022] 在一些实施方案中,一种或多种基因选自:

ARPC1A	LOC100506083
ATAT1	MORN2
ATRAID	NGF
CD27	RHOBTB1
CHMP7	RNGTT
COMMD5	SETD9
CORO2B	SNX8
CYSLTR2	SPINT2
DNAJC7	SPON2
GOLGA1	SSNA1
HPCA	STOML2
IMPDH2	SUOX
KAT8	TMEM258
KLHDC3	UBL7-AS1

[0023] 在一些实施方案中,至少一种临床特征选自:体重指数(BMI)、性别、年龄、种族、先前疗法治疗、疾病持续时间、C反应蛋白水平、抗环瓜氨酸肽的存在、类风湿因子的存在、患者整体评估、治疗应答率(例如,ACR20、ACR50、ACR70)以及其组合。在一些实施方案中,分类器以至少90%的准确度鉴定群体中无应答的至少40%的受试者,其中群体包括至少170名受试者。

[0024] 在一些实施方案中,抗TNF疗法包括施用英夫利昔单抗、阿达木单抗、依那西普、赛妥珠单抗、goliluma或其生物仿制药。在一些实施方案中,所述疾病、病症或病况选自类风湿性关节炎、银屑病性关节炎、强直性脊柱炎、克罗恩病、溃疡性结肠炎、慢性银屑病、化脓性汗腺炎、多发性硬化症和幼年特发性关节炎。在一些实施方案中,使用源自有应答和无应答的先前受试者的微阵列分析建立分类器。在一些实施方案中,使用源自独立队列的RNAseq数据验证分类器。在一些实施方案中,SNP选自表5。

[0025] 在一些实施方案中,本公开提供了一种向患有自身免疫性疾病的受试者施用疗法的方法,所述方法包括向已通过分类器分类为有应答或无应答的受试者施用疗法,所述分类器被建立用于区分已接受抗TNF疗法的队列中有应答和无应答的先前受试者;其中,如果受试者被分类器分类为有应答,则施用抗TNF疗法;并且如果受试者被分类器分类为无应答,则施用抗TNF疗法的替代方案;并且分类器通过评估以下各项进行开发:表达水平与临床应答性或无应答性显著相关(例如,以线性和/或非线性方式)的一种或多种基因;以下各项中的至少一个:一种或多种基因的表达序列中一个或多个单核苷酸多态性(SNP)的存在;或有应答和无应答的先前受试者的至少一种临床特征;并且分类器通过不是已接受抗TNF疗法的队列的独立队列进行验证。

[0026] 在一些实施方案中,分类器在至少170名受试者的群体中以至少90%的准确度预测受试者的应答性。在一些实施方案中,通过评估以下各项开发分类器:表达水平与临床应答性或无应答性显著相关(例如,以线性和/或非线性方式)的一种或多种基因;一个或多个SNP的存在;以及至少一种临床特征。

[0027] 在一些实施方案中,至少一种临床特征选自:体重指数(BMI)、性别、年龄、种族、先前疗法治疗、疾病持续时间、C反应蛋白水平、抗环瓜氨酸肽的存在、类风湿因子的存在、患者整体评估、治疗应答率(例如,ACR20、ACR50、ACR70)以及其组合。在一些实施方案中,抗TNF疗法包括施用英夫利昔单抗、阿达木单抗、依那西普、赛妥珠单抗、goliluma或其生物仿制药。

[0028] 在一些实施方案中,抗TNF疗法的替代方案选自利妥昔单抗、沙利姆单抗(sarilumab)、枸橼酸托法替尼(tofacitinib citrate)、来氟米特、维多珠单抗(vedolizumab)、托珠单抗、阿那白滞素和阿巴西普。

[0029] 在一些实施方案中,所述疾病、病症或病况选自类风湿性关节炎、银屑病性关节炎、强直性脊柱炎、克罗恩病、溃疡性结肠炎、慢性银屑病、化脓性汗腺炎、多发性硬化症和幼年特发性关节炎。

[0030] 在一些实施方案中,本公开提供了一种开发分类器以用于针对一个或多个治疗属性对受试者分层的方法,其包括以下步骤:分析针对至少一个治疗属性代表至少两个不同类别的受试者中表达的RNA的序列数据;由序列数据评估一个或多个单核苷酸多态性(SNP)的存在;确定一个或多个SNP的存在与至少一个治疗属性相关;以及将一个或多个SNP包括

在分类器中。

[0031] 在一些实施方案中,本公开提供了开发分类器的方法,所述分类器用于通过分析针对至少一个治疗属性代表至少两个不同类别的受试者中表达的RNA的序列数据,来针对一个或多个治疗属性对受试者分层的方法,该方法的改进包括:由序列数据评估一个或多个单核苷酸多态性(SNP)的存在;以及确定一个或多个SNP的存在与至少一个治疗属性相关;以及将一个或多个SNP包括在分类器中。

[0032] 在一些实施方案中,一个或多个治疗属性选自以下各项组成的组:发展特定疾病、病症或病况的风险;特定疾病、病症或病况的特定结果的可能性;对特定疗法产生应答的可能性。在一些实施方案中,一个或多个治疗属性是对特定疗法产生应答的可能性。在一些实施方案中,特定疗法为抗TNF疗法。在一些实施方案中,评估一个或多个SNP的存在包括将RNA的序列数据与参考人类基因组进行比较。

[0033] 在一些实施方案中,本公开提供了一种治疗患有自身免疫性疾病的受试者的方法,所述方法包括以下步骤:(a)通过计算设备的处理器接收与受试者的一种或多种基因中的每一种的表达水平相对应的数据,其中所述基因选自表1;(b)使用在步骤(a)中接收的数据由处理器自动确定受试者对抗TNF疗法有应答或无应答的分类;以及,任选地,(c)如果受试者被分类为对抗TNF疗法无应答,则向受试者施用抗TNF疗法的替代方案以治疗自身免疫性疾病;或者,如果受试者被分类为对抗TNF疗法有应答,则施用抗TNF疗法。

[0034] 在一些实施方案中,步骤(a)中的处理器进一步接收与以下各项中的至少一个相对应的数据:一种或多种临床特征;或一个或多个单核苷酸多态性(SNP)。

[0035] 在一些实施方案中,一种或多种临床特征选自:体重指数(BMI)、性别、年龄、种族、先前疗法治疗、疾病持续时间、C反应蛋白(CRP)水平、抗环瓜氨酸肽的存在、类风湿因子的存在、患者整体评估、治疗应答率(例如,ACR20、ACR50、ACR70)以及其组合。在一些实施方案中,自身免疫性疾病选自类风湿性关节炎、银屑病性关节炎、强直性脊柱炎、克罗恩病、溃疡性结肠炎、慢性银屑病、化脓性汗腺炎和幼年特发性关节炎。在一些实施方案中,自身免疫性疾病为类风湿性关节炎或溃疡性结肠炎。

[0036] 在一些实施方案中,抗TNF疗法的替代方案选自利妥昔单抗、沙利姆单抗、枸橼酸托法替尼、来氟米特、维多珠单抗、托珠单抗、阿那白滞素和阿巴西普。在一些实施方案中,抗TNF疗法选自英夫利昔单抗、阿达木单抗、依那西普、赛妥珠单抗、goliluma或其生物仿制药。

[0037] 在一些实施方案中,步骤(b)包括使用机器学习模型自动确定所述分类。在一些实施方案中,机器学习模型是随机森林模型。

[0038] 在一些实施方案中,步骤(b)包括在不使用临床协变量分类器(CC)的情况下自动确定所述分类。在一些实施方案中,步骤(b)包括在不使用组合基因组-临床分类器(GCC)的情况下自动确定所述分类。在一些实施方案中,步骤(b)包括在不使用临床协变量分类器(CC)和不使用组合基因组-临床分类器(GCC)的情况下自动确定所述分类。

[0039] 在一些实施方案中,本公开提供了一种方法,其包括以下步骤:(a)通过计算设备的处理器接收与患有自身免疫性疾病(例如,类风湿性关节炎)的受试者的一种或多种基因中的每一种的表达水平相对应的数据,所述一种或多种基因包括选自COR02B、STOML2、CHMP7、SPINT2、CYSLTR2、SSNA1、IMPDH2、GOLGA1、SUOX、ARPC1A和KLHDC3组成的组的至少一

个成员(例如,一个、两个、三个、四个、五个、六个、七个、八个、九个、十个或全部十一个成员);(b)使用在步骤(a)中接收的数据由处理器自动确定受试者为对抗TNF疗法无应答的分类;以及,任选地,(c)为受试者开处方和/或施用第二疗法(例如,第一疗法的替代方案,例如,抗TNF疗法的替代方案)以治疗疾病,从而避免为受试者开处方和/或施用第一疗法。

[0040] 在一些实施方案中,本公开提供了一种方法,其包括以下步骤:(a)通过计算设备的处理器接收与患有疾病(例如,自身免疫性疾病,例如,类风湿性关节炎)的受试者的一种或多种基因中的每一种的表达水平相对应的数据,所述一种或多种基因包括选自自由CORO2B、STOML2、CHMP7、SPINT2、CYSLTR2、SSNA1、IMPDH2、GOLGA1、SUOX、ARPC1A和KLHDC3组成的组的至少一个成员(例如,一个、两个、三个、四个、五个、六个、七个、八个、九个、十个或全部十一个成员);(b)使用在步骤(a)中接收的数据由处理器自动确定受试者为对第一疗法(例如,抗TNF疗法)有应答的分类;以及,任选地,(c)为受试者开处方和/或施用第一疗法以治疗疾病。

[0041] 在一些实施方案中,步骤(b)包括使用机器学习模型(例如,随机森林模型)自动确定所述分类。在一些实施方案中,步骤(b)包括在不使用临床协变量分类器(CC)和不使用组合基因组-临床分类器(GCC)的情况下自动确定所述分类。

[0042] 在一些实施方案中,本公开提供了一种用于分析受试者的基因组数据(例如,下一代RNA-seq数据)(例如,并用于基于所述基因组数据确定所述受试者的分类)的管道(例如,计算机架构管道),所述管道包括多个模块,每个模块能够在所述模块升级后独立验证。

[0043] 在一些实施方案中,所述多个模块包括一个或多个机器学习模型。

[0044] 在一些实施方案中,所述多个模块包括一个或多个已知的生物信息学模块(例如,RSEM和/或STAR)和一个或多个专有分类模块。

[0045] 在一些实施方案中,本公开提供了一种在对患有自身免疫性疾病的受试者进行抗TNF疗法的任何施用之前将所述受试者分类为对所述抗TNF疗法可能有应答或可能无应答的方法,所述方法包括以下步骤:(a)通过计算设备的处理器接收与选自以下各项组成的组的一种或多种基因中的每一种的受试者的表达水平相对应的数据:

ARPC1A	LOC100506083
ATAT1	MORN2
ATRAID	NGF
CD27	RHOBTB1
CHMP7	RNGTT
COMMD5	SETD9
CORO2B	SNX8
CYSLTR2	SPINT2
DNAJC7	SPON2
GOLGA1	SSNA1
HPCA	STOML2
IMPDH2	SUOX
KAT8	TMEM258
KLHDC3	UBL7-AS1

以及(b)使用在步骤(a)中接收到的数据由处理器自动确定受试者为对抗TNF疗法有应答或无应答的分类。

[0046] 在一些实施方案中,一种在对患有自身免疫性疾病的受试者进行抗TNF疗法的任何施用之前将所述受试者分类为对抗TNF疗法可能有应答或可能无应答的方法进一步包括:(c)如果受试者在步骤(b)中被分类为对抗TNF疗法无应答,则为受试者开处方和/或施用抗TNF疗法的替代方案以治疗自身免疫性疾病;或者,如果受试者在步骤(b)中被分类为对抗TNF疗法有应答,则向受试者施用抗TNF疗法。

[0047] 在一些实施方案中,步骤(a)中的处理器进一步接收与如下(i)和(ii)中至少一个相对应的数据:(i)受试者的一种或多种临床特征;(ii)受试者的一个或多个单核苷酸多态性(SNP)。

[0048] 在一些实施方案中,步骤(a)中的处理器接收与受试者的一种或多种临床特征相对应的数据,所述临床特征选自由以下各项组成的组:体重指数(BMI)、性别、年龄、种族、先前治疗、疾病持续时间、C反应蛋白(CRP)水平、抗环瓜氨酸肽的存在、类风湿因子的存在、患者整体评估以及治疗应答率(例如,ACR20、ACR50、ACR70)。

[0049] 在一些实施方案中,步骤(a)中的处理器接收与表5中列出的一个或多个SNP相对应的数据。

[0050] 在一些实施方案中,自身免疫性疾病是选自由类风湿性关节炎、银屑病性关节炎、强直性脊柱炎、克罗恩病、溃疡性结肠炎、慢性银屑病、化脓性汗腺炎和幼年特发性关节炎组成的组的成员。在一些实施方案中,自身免疫性疾病为类风湿性关节炎或溃疡性结肠炎。

[0051] 在一些实施方案中,抗TNF疗法的替代方案包括选自由利妥昔单抗、沙利姆单抗、枸橼酸托法替尼、来氟米特、维多珠单抗、托珠单抗、阿那白滞素和阿巴西普组成的组的至少一个成员。

[0052] 在一些实施方案中,抗TNF疗法包括选自由英夫利昔单抗、阿达木单抗、依那西普、赛妥珠单抗、goliluma以及前述中任一个的生物仿制药组成的组的至少一个成员。

[0053] 在一些实施方案中,抗TNF疗法的替代方案不包括选自由英夫利昔单抗、阿达木单抗、依那西普、赛妥珠单抗、goliluma以及前述中任一个的任何生物仿制药组成的组的任何成员。

[0054] 在一些实施方案中,步骤(b)包括使用机器学习模型自动确定所述分类。在一些实施方案中,机器学习模型是随机森林模型。

[0055] 在一些实施方案中,步骤(b)包括在不使用临床协变量分类器(CC)的情况下自动确定所述分类。在一些实施方案中,步骤(b)包括在不使用组合基因组-临床分类器(GCC)的情况下自动确定所述分类。在一些实施方案中,步骤(b)包括在不使用临床协变量分类器(CC)和不使用组合基因组-临床分类器(GCC)的情况下自动确定所述分类。

[0056] 在一些实施方案中,本公开提供了一种在对患有自身免疫性疾病的受试者进行抗TNF疗法的任何施用之前将所述受试者分类为对抗TNF疗法可能有应答或可能无应答的方法,所述方法包括以下步骤:(a)通过计算设备的处理器接收与患有自身免疫性疾病的受试者的一种或多种基因中的每一种的表达水平相对应的数据,所述一种或多种基因包括选自由COR02B、STOML2、CHMP7、SPINT2、CYSLTR2、SSNA1、IMPDH2、GOLGA1、SUOX、ARPC1A和KLHDC3组成的组的至少一个成员;(b)使用在步骤(a)中接收到的数据由处理器自动确定受试者为对

抗TNF疗法可能有应答或可能无应答的分类。

[0057] 在一些实施方案中,一种在对患有自身免疫性疾病的受试者进行抗TNF疗法的任何施用之前将所述受试者分类为对所述抗TNF疗法可能有应答或可能无应答的方法进一步包括:(c)如果受试者在步骤(b)中被分类为对抗TNF疗法无应答,则为受试者开处方和/或施用抗TNF疗法的替代方案以治疗自身免疫性疾病;或者,如果受试者在步骤(b)中被分类为对抗TNF疗法有应答,则向受试者施用抗TNF疗法。

[0058] 在一些实施方案中,自身免疫性疾病为类风湿性关节炎。

[0059] 在一些实施方案中,步骤(a)包括接收与选自由CORO2B、STOML2、CHMP7、SPINT2、CYSLTR2、SSNA1、IMPDH2、GOLGA1、SUOX、ARPC1A和KLHDC3组成的组的至少两种基因中的每一种的表达水平相对应的数据。

[0060] 在一些实施方案中,步骤(a)包括接收与选自由CORO2B、STOML2、CHMP7、SPINT2、CYSLTR2、SSNA1、IMPDH2、GOLGA1、SUOX、ARPC1A和KLHDC3组成的组的至少三种基因中的每一种的表达水平相对应的数据。

[0061] 在一些实施方案中,步骤(a)包括接收与选自由CORO2B、STOML2、CHMP7、SPINT2、CYSLTR2、SSNA1、IMPDH2、GOLGA1、SUOX、ARPC1A和KLHDC3组成的组的至少五种基因中的每一种的表达水平相对应的数据。

[0062] 在一些实施方案中,步骤(a)中的处理器进一步接收与如下(i)和(ii)中至少一个相对应的数据:(i)受试者的一种或多种临床特征;(ii)受试者的一个或多个单核苷酸多态性(SNP)。

[0063] 在一些实施方案中,步骤(a)中的处理器接收与受试者的一种或多种临床特征相对应的数据,所述临床特征选自由以下各项组成的组:体重指数(BMI)、性别、年龄、种族、先前疗法治疗、疾病持续时间、C反应蛋白(CRP)水平、抗环瓜氨酸肽的存在、类风湿因子的存在、患者整体评估以及治疗应答率(例如,ACR20、ACR50、ACR70)。

[0064] 在一些实施方案中,步骤(a)中的处理器接收与表5中列出的一个或多个SNP相对应的数据。

[0065] 在一些实施方案中,本公开提供了一种在对患有自身免疫性疾病的受试者进行抗TNF疗法的任何施用之前将所述受试者分类为对抗TNF疗法可能有应答或可能无应答的系统,所述系统包括:处理器;以及在其上具有指令的存储器,所述指令在由处理器执行时,使处理器:(a)接收数据集,所述数据集包括选自由以下各项组成的组的一种或多种基因中的每一种的受试者的表达水平:

ARPC1A	LOC100506083
ATAT1	MORN2
ATRAID	NGF
CD27	RHOBTB1
CHMP7	RNGTT
COMMD5	SETD9
CORO2B	SNX8
CYSLTR2	SPINT2
DNAJC7	SPON2

GOLGA1	SSNA1
HPCA	STOML2
IMPDH2	SUOX
KAT8	TMEM258
KLHDC3	UBL7-AS1

以及 (b) 使用数据集自动确定受试者为对抗TNF疗法有应答或无应答的分类。

[0066] 在一些实施方案中,数据集还包括与如下 (i) 和 (ii) 中至少一个相对应的数据:
(i) 受试者的一种或多种临床特征; (ii) 受试者的一个或多个单核苷酸多态性 (SNP)。

[0067] 在一些实施方案中,数据集包括与受试者的一种或多种临床特征相对应的数据,所述临床特征选自由以下各项组成的组:体重指数 (BMI)、性别、年龄、种族、先前疗法治疗、疾病持续时间、C反应蛋白 (CRP) 水平、抗环瓜氨酸肽的存在、类风湿因子的存在、患者整体评估以及治疗应答率 (例如, ACR20、ACR50、ACR70)。

[0068] 在一些实施方案中,数据集包括表5中列出的一个或多个 SNP。

[0069] 在一些实施方案中,自身免疫性疾病是选自由类风湿性关节炎、银屑病性关节炎、强直性脊柱炎、克罗恩病、溃疡性结肠炎、慢性银屑病、化脓性汗腺炎和幼年特发性关节炎组成的组的成员。在一些实施方案中,自身免疫性疾病为类风湿性关节炎或溃疡性结肠炎。在一些实施方案中,自身免疫性疾病为类风湿性关节炎。

定义

[0070] 施用:如本文所用,术语“施用”通常是指将组合物施用到受试者或系统,例如,以实现作为组合物或包含在组合物中或以其他方式通过组合物递送的药剂的递送。

[0071] 药剂:如本文所用,术语“药剂”是指实体 (例如,脂质、金属、核酸、多肽、多糖、小分子等,或其复合物、组合、混合物或系统 [例如,细胞、组织、生物体]) 或现象 (例如,热、电流或电场、磁力或磁场等)。

[0072] 氨基酸:如本文所用,术语“氨基酸”是指可以例如通过形成一个或多个肽键而掺入多肽链的任何化合物和/或物质。在一些实施方案中,氨基酸具有一般结构 $H_2N-C(H)(R)-COOH$ 。在一些实施方案中,氨基酸是天然存在的氨基酸。在一些实施方案中,氨基酸是非天然氨基酸;在一些实施方案中,氨基酸是D-氨基酸;在一些实施方案中,氨基酸是L-氨基酸。如本文所用,术语“标准氨基酸”是指在天然存在的肽中常见的二十种L-氨基酸中的任一种。“非标准氨基酸”是指除标准氨基酸以外的任何氨基酸,无论其是否可以存在于天然来源中。在一些实施方案中,与以上一般结构相比,包括多肽中的羧基和/或氨基末端氨基酸在内的氨基酸可以包含结构修饰。例如,在一些实施方案中,与一般结构相比,氨基酸可通过甲基化、酰胺化、乙酰化、聚乙二醇化、糖基化、磷酸化和/或取代 (例如,氨基、羧基团、一个或多个质子和/或羟基) 进行修饰。在一些实施方案中,与含有原本相同的未修饰氨基酸的多肽相比,此类修饰可例如改变含有修饰氨基酸的多肽的稳定性或循环半衰期。在一些实施方案中,与含有原本相同的未修饰氨基酸的多肽相比,此类修饰不显著改变含有修饰氨基酸的多肽的相关活性。如从上下文中可以清楚地看到,在一些实施方案中,术语“氨基酸”可用于指代游离氨基酸;在一些实施方案中,其可用于指代多肽的氨基酸残基,例如,多肽内的氨基酸残基。

[0073] 类似物:如本文所用,术语“类似物”是指与参考物质共享一个或多个特定结构特

征、元素、组分或部分的物质。通常，“类似物”显示出与参考物质的显著结构相似性，例如共享核或共有结构，但也以某些离散方式不同。在一些实施方案中，类似物是可以从参考物质生成的物质，例如通过对参考物质的化学操纵。在一些实施方案中，类似物是可以通过执行与生成参考物质的合成过程基本上类似（例如，共享多个步骤）的合成过程生成的物质。在一些实施方案中，类似物是或可以通过执行不同于用于生成参考物质的合成过程生成的。

[0074] 拮抗剂：如本文所用，术语“拮抗剂”可指其存在、水平、程度、类型或形式与靶标的水平或活性降低相关联的药剂或状况。拮抗剂可包括任何化学类别的药剂，包括例如小分子、多肽、核酸、碳水化合物、脂质、金属和/或显示出相关抑制活性的任何其他实体。在一些实施方案中，拮抗剂可以是“直接拮抗剂”，因为其与其靶标直接结合；在一些实施方案中，拮抗剂可以是“间接拮抗剂”，因为其通过除直接结合其靶标之外的方式施加其影响；例如，通过与靶标的调节剂相互作用，从而改变靶标的水平或活性）。在一些实施方案中，“拮抗剂”可称为“抑制剂”。

[0075] 抗体：如本文所用，术语“抗体”是指包括足以赋予与特定靶抗原的特异性结合的经典免疫球蛋白序列元件的多肽。如本领域所知，在自然界中产生的完整抗体是大约150kD的四聚体剂，由两个相同的重链多肽（每个约50kD）和两个相同的轻链多肽（每个约25kD）组成，它们相互缔合成通常称为“Y形”结构的物质。每条重链由至少四个结构域（每个长约110个氨基酸）组成——一个氨基末端可变（VH）结构域（位于Y结构的顶端），接着是三个恒定结构域：CH1、CH2和羧基末端CH3（位于Y的茎的底部）。称为“开关”的短区域连接重链可变区和恒定区。“铰链”将CH2和CH3结构域连接到抗体的其余部分。该铰链区中的两个二硫键将两个重链多肽连接到完整抗体中的另一个。每个轻链由两个结构域组成——氨基末端可变（VL）结构域，接着是羧基末端恒定（CL）结构域，由另一个“开关”彼此分开。完整的抗体四聚体由两个重链-轻链二聚体组成，其中重链和轻链通过单个二硫键相互连接；另外两个二硫键将重链铰链区相互连接，从而使二聚体相互连接并形成四聚体。天然产生的抗体也被糖基化，通常在CH2结构域上。天然抗体中的每个结构域都具有这样一种结构：其特征是由在压缩的反平行β桶（beta barrel）中彼此紧密堆积的两个β折叠（例如，3、4或5链折叠）形成的“免疫球蛋白折叠”。每个可变结构域包含三个称为“互补决定区”的高变环（CDR1、CDR2和CDR3）和四个在一定程度上不变的“框架”区（FR1、FR2、FR3和FR4）。当天然抗体折叠时，FR区形成为结构域提供结构框架的β折叠，并且使得来自重链和轻链的CDR环区在三维空间中聚集在一起，使得它们形成位于Y结构顶端的单个高变抗原结合位点。天然存在的抗体的Fc区结合到补体系统的元件，并且还结合到效应细胞上的受体，包括例如介导细胞毒性的效应细胞。如本领域所知，Fc区对Fc受体的亲和力和/或其他结合属性可以通过糖基化或其他修饰进行调节。在一些实施方案中，根据本发明产生和/或利用的抗体包括糖基化Fc结构域，包括具有修饰或工程化的此类糖基化的Fc结构域。出于本发明的目的，在某些实施方案中，包括天然抗体中发现的足够免疫球蛋白结构域序列的任何多肽或多肽复合物可以被称为和/或用作“抗体”，无论所述多肽是天然产生的（例如，由生物体对抗原反应产生的），还是通过重组工程化、化学合成或其他人工系统或方法产生的。在一些实施方案中，抗体是多克隆的；在一些实施方案中，抗体是单克隆的。在一些实施方案中，抗体具有小鼠、兔、灵长类动物或人类抗体所特有的恒定区序列。在一些实施方案中，如本领域所知，抗体序列元件为人源化的、灵长类化的、嵌合的等。此外，如本文所用，术语“抗体”可以在适当的实施方案中

(除非另有说明或由上下文清楚)指用于在替代呈现中利用抗体结构和功能特征的任何现有技术已知或开发的构建体或形式。例如,在实施方案中,根据本发明使用的抗体的形式选自但不限于完整IgA、IgG、IgE或IgM抗体;双特异性或多特异性抗体(例如,Zybodies[®]等);抗体片段,诸如Fab片段、Fab'片段、F(ab')₂片段、Fd'片段、Fd片段和分离的CDR或其集合;单链Fv;多肽-Fc融合体;单结构域抗体(例如,鲨鱼单结构域抗体,诸如IgNAR或其片段);骆驼样抗体;隐蔽抗体(例如,Probodies[®]);小型模块化免疫药物(Small Modular ImmunoPharmaceuticals,“SMIPsTM”);单链或串联双体(TandAb[®]);VHH;Anticalins[®];Nanobodies[®]微型抗体;BiTE[®];锚蛋白重复蛋白或DARPINS[®];Avimers[®];DART;TCR样抗体;Adnectins[®];Afflins[®];Trans-bodies[®];Affibodies[®];TrimerX[®];微蛋白(MicroProteins);Fynomers[®], Centyrins[®];以及KALBITOR[®]。在一些实施方案中,抗体可能缺少如果其自然产生将具有的共价修饰(例如,附接聚糖)。在一些实施方案中,抗体可包含共价修饰(例如,附接聚糖、有效载荷[例如,可检测部分、治疗部分、催化部分等]或其他侧基[例如,聚乙二醇等])。

[0076] 关联:如果一个事件或实体的存在、水平、程度、类型和/或形式与另一个相关,则两个事件或实体彼此“关联”,如所述术语在本文所使用的。例如,如果特定实体(例如,多肽、基因签名(genetic signature)、代谢物、微生物等)的存在、水平和/或形式与疾病、病症或病况的发生率和/或易感性(例如,在相关群体中)相关,则认为所述实体与特定疾病、病症或病况相关联。在一些实施方案中,如果两个或更多个实体直接或间接地相互作用,使得它们彼此物理接近和/或保持彼此物理接近,则它们彼此在物理上“关联”。在一些实施方案中,彼此物理关联的两个或更多个实体彼此共价连接;在一些实施方案中,彼此物理关联的两个或更多个实体不是彼此共价连接,而是非共价缔合,例如借助于氢键、范德华相互作用、疏水相互作用、磁性及其组合。

[0077] 生物样本;如本文所用,术语“生物样本”通常是指如本文所述从感兴趣的生物来源(例如,组织或生物体或细胞培养物)获得或衍生的样本。在一些实施方案中,感兴趣的来源包括生物体,诸如动物或人类。在一些实施方案中,生物样本是或包括生物组织或流体。在一些实施方案中,生物样本可以是或包括骨髓;血液;血细胞;腹水;组织或细针活检样本;含有细胞的体液;游离漂浮核酸;痰;唾液;尿;脑脊液、腹腔液;胸膜液;粪便;淋巴;妇科流体;皮肤拭子;阴道拭子;口腔拭子;鼻拭子;冲洗液或灌洗液,诸如导管灌洗液或支气管肺泡灌洗液;抽吸物;刮屑;骨髓标本;组织活检标本;外科标本;粪便、其他体液、分泌物和/或排泄物;和/或其中的细胞等。在一些实施方案中,生物样本是或包括从个体获得的细胞。在一些实施方案中,获得的细胞是或包括来自获得样本的个体的细胞。在一些实施方案中,样本是通过任何适当的手段直接从感兴趣的来源获得的“原始样本”。例如,在一些实施方案中,原始生物样本通过选自由以下各项组成的组的方法获得:活检(例如,细针穿刺或组织活检)、手术、收集体液(例如,血液、淋巴、粪便等)等。在一些实施方案中,如从上下文中可以清楚地看出,术语“样本”是指通过处理原始样本(例如,通过去除其一种或多种组分和/或通过向其中添加一种或多种剂)而获得的制备物。例如,使用半渗透膜过滤。这样的“处理样本”可包括,例如,从样本中提取或通过使原始样本经受诸如mRNA的扩增或逆转录、某些组分的分离和/或纯化等技术而获得的核酸或蛋白质。

[0078] 联合疗法;如本文所用,术语“联合疗法”是指受试者同时暴露于两种或更多种治疗方案(例如,两种或更多种治疗剂)的临床干预。在一些实施方案中,可同时施用两种或更多种治疗方案。在一些实施方案中,可顺序施用两种或更多种治疗方案(例如,在施用任何剂量的第二方案之前施用第一方案)。在一些实施方案中,以重叠给药方案施用两种或更多种治疗方案。在一些实施方案中,组合疗法的施用可涉及向接受另外一种或多种药剂或模式的受试者施用一种或多种治疗剂或模式。在一些实施方案中,组合疗法不一定要求单独药剂在单一组合物中(或甚至不一定同时)一起施用。在一些实施方案中,通过单独的施用途径(例如,一种药剂口服施用,并且另一种药剂静脉内施用)和/或在不同的时间点,单独(例如,以单独的组合物形式)向受试者施用组合疗法的两种或更多种治疗剂或模式。在一些实施方案中,可通过相同的施用途径和/或同时在组合组合物中或甚至在组合化合物中(例如,作为单一化学复合物或共价实体的一部分)一起施用两种或更多种治疗剂。

[0079] 相当的;如本文所用,术语“相当的”是指两种或更多种药剂、实体、情况、条件集等可能彼此不相同,但足够相似以允许相互进行比较,使得本领域技术人员将理解,可基于观察到的差异或相似性合理地得出结论。在一些实施方案中,相当的条件集、环境、个体或群体的特征在于多个基本上相同的特征和一个或少量的变化特征。本领域的普通技术人员将理解,在上下文中,在任何给定情况下,两种或更多种这样的药剂、实体、情况、条件集等被认为是相当的所需要的同一性程度。例如,本领域普通技术人员将理解,当通过足够数量和类型的基本上相同的特征表征以保证以下合理结论时,环境、个体或群体的集合是彼此相当的:在不同的环境、个体或群体集合下或使用所述不同的环境、个体或群体集合获得的结果或观察到的现象中的差异是由那些发生变化的特征的变化引起的或指示所述特征的变化。

[0080] 对应于;如本文所用,短语“对应于”是指两个实体、事件或现象之间的关系,这两个实体、事件或现象共享足以可合理比较的特征,使得“对应”属性是明显的。例如,在一些实施方案中,该术语可参考化合物或组合物使用,以通过与适当的参考化合物或组合物进行比较来指定化合物或组合物中结构元件的位置和/或身份。例如,在一些实施方案中,聚合物中的单体残基(例如,多肽中的氨基酸残基或多核苷酸中的核酸残基)可被鉴定为“对应于”适当的参考聚合物中的残基。例如,普通技术人员将理解,为了简单起见,通常基于参考相关多肽使用规范编号系统来指定多肽中的残基,使得例如,“对应于”位置190处的残基的氨基酸实际上不需要是特定氨基酸链中的第190个氨基酸,而是对应于参考多肽中190处的残基;本领域普通技术人员容易理解如何鉴定“对应”氨基酸。例如,本领域技术人员将了解各种序列比对策略,包括软件程序,例如BLAST、CS-BLAST、CUSASW++、DIAMOND、FASTA、GGSEARCH/GLSEARCH、Genoogle、HMMER、HHpred/HHsearch、IDF、Infernal、KLAST、USEARCH、parasail、PSI-BLAST、PSI-Search、ScalaBLAST、Sequilab、SAM、SSEARCH、SWAPHI、SWAPHI-LS、SWIMM或SWIPE,其可以根据本公开用于例如鉴定多肽和/或核酸中的“对应”残基。

[0081] 给药方案;如本文所用,术语“给药方案”是指单独施用给受试者的通常按时间段分开的一组单位剂量(通常多于一次)。在一些实施方案中,给定治疗剂具有推荐的给药方案,其可涉及一次或多次剂量。在一些实施方案中,给药方案包括多次剂量,其各自在时间上与其他剂量分开。在一些实施方案中,单次剂量彼此间隔相同长度的时间段;在一些实施方案中,给药方案包括多次剂量和分开单次剂量的至少两个不同的时间段。在一些实施方

案中,给药方案内的所有剂量均具有相同的单位剂量总量。在一些实施方案中,给药方案内的不同剂量具有不同的量。在一些实施方案中,给药方案包括第一剂量总量的第一剂量,然后是与第一剂量总量不同的第二剂量总量的一次或多次另外剂量。在一些实施方案中,给药方案包括第一剂量总量的第一剂量,然后是与第一剂量总量相同的第二剂量总量的一次或多次另外剂量。在一些实施方案中,当在相关群体中施用,给药方案与期望的或有益的结果相关(即,是治疗性给药方案)。

[0082] 改善、增加或减少;如本文所用,术语“改善”、“增加”或“减少”,或其语法上相当的可比较性术语,指示相对于相当的参考测量结果的值。例如,在一些实施方案中,使用感兴趣的药剂获得的评估值相对于使用相当的参考药剂获得的评估值可以是“改善的”。替代地或附加地,在一些实施方案中,相对于在不同条件下(例如,在诸如施用感兴趣的药剂的事件之前或之后)的相同受试者或系统中,或在不同的相当受试者(例如,在存在感兴趣的特定疾病、病症或病况的一个或多个指标的情况下,或在之前暴露于病况或药剂等,与感兴趣的受试者或系统不同的相当的受试者或系统)中获得的评估值,在感兴趣的受试者或系统中获得的评估值可以是“改善的”。

[0083] 药物组合物:如本文所用,术语“药物组合物”是指与一种或多种药学上可接受的载体一起配制的活性剂。在一些实施方案中,活性剂以适合在治疗向相关受试者施用的单位剂量总量存在(例如,已证明在施用显示达到预定治疗效果的统计显著概率的量),或在不同的相当受试者(例如,在存在感兴趣的特定疾病、病症或病况的一个或多个指标的情况下,或在暴露于病况或药剂之前等,与感兴趣的受试者或系统不同的相当的受试者或系统)中存在。在一些实施方案中,比较性术语是指统计学上相关的差异(例如,具有足以实现统计相关性的普遍性和/或幅度)。本领域技术人员将意识到,或将容易地能够确定在给定的上下文中实现这种统计学显著性所需或足够的差异程度和/或普遍性。

[0084] 药学上可接受的:如本文所用,短语“药学上可接受的”是指在合理医学判断的范围内适合用于与人类和动物的组织接触而没有过度的毒性、刺激、过敏反应或其他问题或并发症,与合理的收益/风险比相称的那些化合物、材料、组合物和/或剂型。

[0085] 参考:如本文所用,术语“参考”描述了与其进行比较的标准或对照。例如,在一些实施方案中,将感兴趣的药剂、动物、个体、群体、样本、序列或值与参考或对照药剂、动物、个体、群体、样本、序列或值进行比较。在一些实施方案中,与感兴趣的测试或确定基本上同时地测试和/或确定参考或对照。在一些实施方案中,参考或对照是历史参考或对照,任选地体现在有形介质中。通常,如本领域技术人员所理解的,在与所评估的那些相当的条件或情况下确定或表征参考或对照。本领域技术人员将认识到存在足够的相似性来证明对特定的可能参考或对照的依赖和/或比较的情况。

[0086] 治疗有效量;如本文所用,术语“治疗有效量”是指当作为治疗方案的一部分施用引起期望的生物应答的物质(例如,治疗剂、组合物和/或制剂)的量。在一些实施方案中,物质的治疗有效量是当向患有或易感疾病、病症和/或病况的受试者施用足以治疗、诊断、预防和/或延迟疾病、病症和/或病况的发作的量。如本领域普通技术人员将理解的,物质的有效量可根据诸如期望的生物终点、待递送的物质、靶细胞或组织等因素而变化。例如,用于治疗疾病、病症和/或病况的制剂中化合物的有效量是减轻、改善、缓解、抑制、预防、延迟疾病、病症和/或病况的一种或多种症状或特征的发作、降低其严重程度和/或降低

其发生率的量。在一些实施方案中,以单剂量施用治疗有效量;在一些实施方案中,需要多个单位剂量来递送治疗有效量。

[0087] 变体:如本文所用,术语“变体”是指表现出与参考实体的显著结构同一性,但与参考实体相比,在一个或多个化学部分的存在或水平上与参考实体在结构上不同的实体。在许多实施方案中,变体在功能上也不同于其参考实体。通常,特定实体是否被恰当地视为参考实体的“变体”取决于其与参考实体的结构同一性程度。如本领域技术人员将理解的,任何生物或化学参考实体具有某些特征结构元件。根据定义,变体是共享一个或多个此类特征结构元件的独特化学实体。仅举几个例子,小分子可具有特征核结构元件(例如,大环核)和/或一个或多个特征侧基部分,使得小分子的变体是共享核结构元件和特征侧基部分但在其他侧基部分和/或核内存在的键类型上不同的变体(单对双、E对Z等),多肽可具有由在线性或三维空间中相对于彼此具有指定位置和/或有助于特定生物功能的多个氨基酸组成的特征序列元件,核酸可具有由在线性或三维空间中相对于彼此具有指定位置的多个核苷酸残基组成的特征序列元件。例如,由于氨基酸序列中的一个或多个差异和/或共价连接到多肽主链的化学部分(例如,碳水化合物、脂质等)中的一个或多个差异,变体多肽可不同于参考多肽。在一些实施方案中,变体多肽显示出与参考多肽具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%或99%的总体序列同一性。替代地或附加地,在一些实施方案中,变体多肽不与参考多肽共享至少一个特征序列元件。在一些实施方案中,参考多肽具有一种或多种生物活性。在一些实施方案中,变体多肽共享参考多肽的一种或多种生物活性。在一些实施方案中,变体多肽缺乏参考多肽的一种或多种生物活性。在一些实施方案中,与参考多肽相比,变体多肽表现出降低水平的一种或多种生物活性。在许多实施方案中,如果感兴趣的多肽具有与亲本相同但在特定位置有少量序列改变的氨基酸序列,则将感兴趣的多肽视为亲本或参考多肽的“变体”。通常,与亲本相比,变体中少于20%、15%、10%、9%、8%、7%、6%、5%、4%、3%、2%的残基被取代。在一些实施方案中,与亲本相比,变体具有10、9、8、7、6、5、4、3、2或1个取代残基。通常,变体具有非常少量(例如,少于5、4、3、2或1)的取代功能残基(即,参与特定生物活性的残基)。此外,与亲本相比,变体通常具有不超过5、4、3、2或1个添加或缺失,并且通常没有添加或缺失。此外,任何添加或缺失通常少于约25、约20、约19、约18、约17、约16、约15、约14、约13、约10、约9、约8、约7、约6,并且通常少于约5、约4、约3或约2个残基。在一些实施方案中,亲本或参考多肽是在自然界中发现的多肽。

附图说明

[0088] 图1是用于将受试者分类为抗TNF疗法的应答者的特定基因表达水平的潜在相对预测值的图表。

[0089] 图2是示例性实施方案的例示,所述示例性实施方案结合了(在红色框中)RNAseq生物信息学管道(bioinformatics pipeline)和(在蓝色框中)示例性专有算法(例如,在本实例中,PrismRA算法)。

[0090] 图3是用于分类的RNA-seq数据的模块化处理的例示性管道(例如,诊断、鉴定对特定治疗的应答性,以及其他应用)。

[0091] 图4是用于在各种实施方案中使用的示例性网络环境和计算设备。

[0092] 图5示出可以用于实施本公开中描述的技术的计算设备500和移动计算设备550的实例。

[0093] 图6A是用于特征选择 ($n=75$)、模型训练 ($n=185$) 和模型验证 ($n=46$) 的队列 (cohort) 的流程图。

[0094] 图6B是描述图6A的队列1中的特征选择过程的过程。POC数据集是指图6A的队列1数据集。80%的数据用于使用U检验来训练分类器,以鉴定最具区分性的基因,并且80%的数据使用10倍交叉验证分成10组,以提供最终确定的特征集;最终确定的特征集用于对其余20%的数据进行模型训练和评估。该过程重复至少100次。

[0095] 图6C是应用于队列1的特征选择方法的流程图。

[0096] 图7A-图7D涉及Abcon队列 (Affymetrix) 中区分抗TNF疗法的应答者和无应答者之间的基因的选择。图7A是前100个中选择的特征相对于交叉验证轮数的图。图7B是接受者操作特征曲线,其示出在100轮交叉验证中,聚合20%的保留测试集的平均曲线下面积 (AUC)。图7C例示了在100轮交叉验证中,按照每个随机森林模型要求的,聚合20%的保留测试集样本的预测类别概率。图7D例示了在,考虑到在100轮交叉验证中的30轮交叉验证中排序前100位的38个基因时,抗TNF疗法的应答者 (红色) 和无应答者 (黑色) 的主成分分析。

[0097] 图8A-图8B涉及终点和纳入标准的优化。图8A是由具有血清阳性和没有血清阳性的患者建立的模型的曲线下面积分布。图8B是利用仅针对RF和CCP为血清阳性的患者建立的模型的曲线下面积分布。血清阳性样本中的基线RNAseq数据是男性和女性患者6个月内ACR50结果的最佳预测。

[0098] 图9A-图9D涉及代表性模型性能。中值AUC模型是在考虑到男性和女性血清阳性患者并使用治疗后6个月的ACR50作为临床终点时产生的。图9A例示了接受者操作特征曲线,其示出重复10次的聚合10%的保留交叉验证的平均曲线下面积 (AUC)。图9B是重复10次的聚合10%保留交叉验证的阴性预测值相对于真阴性率的图。图9C是按照随机森林模型要求的应答者和无应答者之间聚合验证集样本的预测概率的图。图9D是聚合验证集样本的混淆矩阵。

[0099] 图10A例示了使用仅协变量、仅RNA以及协变量与RNA的模型性能。

[0100] 图10B例示了协变量的选择频率。

[0101] 图11是在10%交叉验证的50次重复中,前15个的基因选择的频率图。

[0102] 图12A-图12D涉及用于预测对抗TNF疗法的应答的模型的验证。图12A例示了接受者操作特征曲线,其示出验证集样本的曲线下面积 (AUC)。图12B是验证集样本的阴性预测值相对于真阴性率的图。图12C是按照随机森林模型要求的每个验证集样本的预测类别概率的图。图12D是与用于预测对抗TNF疗法的应答的模型的验证相关的混淆矩阵。

[0103] 图13A-图13D涉及仅血清阳性患者 ($n=23$) 中用于预测对抗TNF疗法的应答的模型的验证。图13A例示了接受者操作特征曲线,其示出验证集样本的曲线下面积 (AUC)。图13B是验证集样本的阴性预测值相对于真阴性率的图。图13C是按照随机森林模型要求的每个验证集样本的预测类别概率的图。图13D是与仅血清阳性患者 ($n=23$) 中用于预测对抗TNF疗法的应答的模型的验证相关的混淆矩阵。

[0104] 图14是用于开发分类器的示例性工作流程。

[0105] 图15是对38个基因的RNA表达数据的层次聚类分析,例示了两个主要分组,一个主

要为无应答者,另一个为应答者,从而证实了这些基因在抗TNF应答预测中的区分性质。热图以任意单位表示相对RNA表达水平。

[0106] 图16A是人类相互作用组蛋白质-蛋白质网络子集的可视化。网络上包括的蛋白质用灰色圆圈表示。用红色标出的那些代表由含SNP的RNA编码的蛋白质并且用蓝色标出的那些代表由应答区分性基因编码的蛋白质。

[0107] 图16B是对分子药物靶标、与RA相关的蛋白质(疾病模块蛋白)、PrismRA预测分类算法的开发中包括的所有分子特征、由含SNP的RNA(SNP)编码的蛋白质和由区分性基因(DG)编码的蛋白质的邻近性的定量分析。最终预测分类算法中包括的分子特征为黄色。

[0108] 图17是一系列条形图,描绘了在至少10名患者中检测到所选SNP的训练队列中的有应答和无应答RA患者的百分比。

[0109] 图18是包含RA疾病模块、RA药物靶标和分类器特征(SNP和DG)的人类相互作用组子网络。

[0110] 图19是描述在RA中开发抗TNF药物应答算法的流程图。区分抗TNF疗法的应答者和无应答者的基因表达选自可公开获得的微阵列数据集。在跨平台分析中,将这些特征与网络疾病模块相关SNP和临床因素相结合,然后用于使用RNAseq数据训练机器学习算法。在一项独立验证试验中验证预测药物应答算法的性能。

具体实施方式

[0111] 本文呈现了用于自动预测受试者对抗TNF疗法的应答的系统和方法。本文还呈现了用于自动解释基因组和/或多组学数据的模块化系统。

(a) 提供的分类器(一个或多个)

[0112] 本公开提供了一种分类器以及这种分类器的开发,所述分类器可以鉴定(即,预测)哪些患者将对特定疗法产生或不产生应答。在一些实施方案中,分类器被建立用于区分已接受抗TNF疗法(例如,特定抗TNF剂和/或方案)的有应答和无应答的先前受试者。

[0113] 除其他事项外,本公开涵盖以下见解:特定基因集的表达水平,单独或彼此组合,任选地与特定临床特征和/或存在或不存在特定单核苷酸多态性相结合,可用于预测对抗TNF疗法的应答(例如,应答的一个或多个特征)。

[0114] 在一些实施方案中,本公开提供了一种分类器,其为或包括此类基因表达水平、临床特征和/或SNP,并证明已被建立用于区分对抗TNF疗法作出应答和不作出应答的受试者。在一些实施方案中,所提供的分类器被建立用于通过对接受抗TNF疗法且其应答性已知(例如,先前确定)的历史(即,先前)受试者群体的回顾性分析在对抗TNF疗法有应答或无应答的受试者(例如,抗TNF疗法原初(**naive**)受试者)之间进行区分。在一些实施方案中,当应用于此类历史(即,先前)群体时,以至少70%的准确度鉴定队列中至少50%的无应答者的分类器被视为“经验证”。在一些实施方案中,当应用于此类历史(即,先前)群体时,以至少70%的准确度鉴定队列中至少60%的无应答者的分类器被视为“经验证”。在一些实施方案中,当应用于此类历史(即,先前)群体时,以至少70%的准确度鉴定队列中至少70%的无应答者的分类器被视为“经验证”。在一些实施方案中,当应用于此类历史(即,先前)群体时,以至少70%的准确度鉴定队列中至少80%的无应答者的分类器被视为“经验证”。在一些实施方案中,当应用于此类历史(即,先前)群体时,以至少70%的准确度鉴定队列中至少90%

的无应答者的分类器被视为“经验证”。在一些实施方案中,当应用于此类历史(即,先前)群体时,以至少70%的准确度鉴定队列中至少99%的无应答者的分类器被视为“经验证”。

[0115] 在一些实施方案中,当应用于此类历史(即,先前)群体时,以至少80%的准确度鉴定队列中至少50%的无应答者的分类器被视为“经验证”。在一些实施方案中,当应用于此类历史(即,先前)群体时,以至少90%的准确度鉴定队列中至少50%的无应答者的分类器被视为“经验证”。在一些实施方案中,当应用于此类历史(即,先前)群体时,以至少99%的准确度鉴定队列中至少50%的无应答者的分类器被视为“经验证”。

[0116] 在一些实施方案中,本公开提供了治疗患有疾病、病症或病况的受试者的方法,其包括向已通过应用所提供的分类器确定可能对这种抗TNF疗法产生应答的一个或多个受试者施用抗TNF疗法;替代地或另外地,在一些实施方案中,本公开提供了治疗患有疾病、病症或病况的受试者的方法,其包括停止抗TNF疗法,和/或对通过应用提供的分类器确定不太可能对这种抗TNF疗法产生应答的一个或多个受试者施用抗TNF疗法的替代方案。

[0117] 在一些实施方案中,所提供的分类器可以是或包括一种或多种基因的基因表达信息。替代地或另外地,在一些实施方案中,所提供的分类器可以是或包括一个或多个单核苷酸多态性(SNP)的存在或不存在和/或相关受试者的一种或多种临床特征或特性。

[0118] 在一些实施方案中,通过评估以下各项开发分类器:表达水平与临床应答性或无应答性显著相关(例如,以线性和/或非线性方式)的一种或多种基因;一个或多个SNP的存在;以及至少一种临床特征。

[0119] 在一些实施方案中,如本文所述,分类器是通过对自己接受抗TNF疗法并已确定有应答(即,应答者)或无应答(即,无应答者)的患者(例如,先前受试者)的生物样本的一种或多种特征(例如,基因表达水平、存在或不存在一个或多个SNP等)进行回顾性分析来开发的;替代地或另外地,在一些实施方案中,分类器是通过对此类患者的一种或多种临床特征的回顾性分析来开发的,其可能涉及或可能不涉及任何生物样本的评估(并且可以例如通过参考医疗记录来完成)。在一些实施方案中,所有此类患者已接受相同的抗TNF疗法(任选地持续相同或不同的时间段);替代地或另外地,在一些实施方案中,所有此类患者已被诊断患有相同的疾病、病症或病况。在一些实施方案中,其生物样本在回顾性分析中被分析的患者已经接受了不同的抗TNF疗法(例如,使用不同的抗TNF剂和/或根据不同的方案);替代地或另外地,在一些实施方案中,其生物样本在回顾性分析中被分析的患者已被诊断为患有不同的疾病、病症或病况。

1. 基因表达

[0120] 通常,如本文所述的分类器的基因表达方面通过评估以下各项来确定:表达水平与临床应答性或无应答性显著相关(例如,以线性和/或非线性方式)的一种或多种基因;以及以下各项中的至少一个:一种或多种基因的表达序列中一个或多个单核苷酸多态性(SNP)的存在;或有应答和无应答的先前受试者的至少一种临床特征。基因应答签名中可包括表达水平在应答者和无应答者群体之间表现出统计学显著性差异的基因。

[0121] 在一些实施方案中,本公开体现了这样一种见解,即在有应答与无应答受试者之间鉴定或提供分类器的某些先前努力的问题根源是应答者和无应答者群体中的基因表达水平的比较强调和/或集中在(通常仅集中在)表现出群体间表达水平的最大差异(例如,大于2倍的变化)的基因上。本公开认识到,即使那些表达水平差异相对较小(例如,表达变化

小于2倍)的基因也提供有用的信息,并且有价值地包括在本文所述实施方案中的分类器中。

[0122] 此外,在一些实施方案中,本公开体现了这样一种见解,即如本文所述对表达水平在应答者与无应答者群体之间显示出统计学显著性差异(任选地包括小的差异)的基因的相互作用模式的分析提供新的且有价值的信息,这从实质上提高分类器的质量和预测能力。

[0123] 在一些实施方案中,所提供的分类器是或包括可以用于确定受试者是否将对特定疗法(例如,抗TNF疗法)作出应答(例如,表达水平与其相关)的基因或基因集。在一些实施方案中,分类器通过评估以下各项进行开发:表达水平与临床应答性或无应答性显著相关(例如,以线性和/或非线性方式)的一种或多种基因;以及以下各项中的至少一个:一个或多个单核苷酸多态性(SNP)的存在;以及有应答和无应答的先前受试者的至少一种临床特征。

[0124] 在一些实施方案中,用于在分类器中使用和/或用于测量基因表达的一种或多种基因选自表1中的基因:

表1

ARPC1A	LOC100506083
ATAT1	MORN2
ATRAID	NGF
CD27	RHOBTB1
CHMP7	RNGTT
COMMD5	SETD9
CORO2B	SNX8
CYSLTR2	SPINT2
DNAJC7	SPON2
GOLGA1	SSNA1
HPCA	STOML2
IMPDH2	SUOX
KAT8	TMEM258
KLHDC3	UBL7-AS1

[0125] 在某些实施方案中,使用上文鉴定的11种鉴定基因(和/或图1所示的17种另外的基因)的各种子集的基因表达水平(例如,有或没有加权因子)。例如,在某些实施方案中,分类模型可基于表达水平,可使用11种基因的集合(CORO2B、STOML2、CHMP7、SPINT2、CYSLTR2、SSNA1、IMPDH2、GOLGA1、SUOX、ARPC1A和KLHDC3)的三个、四个、五个、六个、七个、八个、九个

或十个成员的子集,或者,在某些实施方案中,分类模型可基于图1所示的28种基因的集合的子集的表达水平。在某些实施方案中,在模型中使用分类器中的每种基因的RNA测序(RNA-seq)数据读段计数。在某些实施方案中,RNA-seq读段计数作为每百万个读段每千碱基外显子的片段数(fragments per kilobase of exon per million reads,FPKM)提供。

[0126] 此外,令人惊讶的是,发现基于11种基因的集合(COR02B、STOML2、CHMP7、SPINT2、CYSLTR2、SSNA1、IMPDH2、GOLGA1、SUOX、ARPC1A和KLHDC3)的基因表达水平的模型与临床协变量分类器(CC)或组合基因组-临床分类器(GCC)相比是对抗TNF疗法的应答/无应答的更好预测器。

[0127] 在一些实施方案中,测量了11种基因的集合(COR02B、StGL2、CHMP7、SPIT2、CYSLT2、SnAA1、IMPDH2、GALGA1、SUOX、ARPC1A和KLHDC3)的一种、两种、三种、四种、五种、六种、七种、八种、九种、十种或全部十一种基因,以预测抗TNF应答。在一些实施方案中,测量一种基因。在一些实施方案中,选择两种基因。在一些实施方案中,选择三种基因。在一些实施方案中,选择四种基因。在一些实施方案中,选择五种基因。在一些实施方案中,选择六种基因。在一些实施方案中,选择七种基因。在一些实施方案中,选择八种基因。在一些实施方案中,选择九种基因。在一些实施方案中,选择十种基因。在一些实施方案中,选择十一种基因。在一些实施方案中,选择多于八种基因。在一些实施方案中,选择八、九、十或十一种基因。

[0128] 在一些实施方案中,使用选自COR02B、STOML2、CHMP7、SPINT2、CYSLTR2、SSNA1、IMPDH2、GOLGA1、SUOX、ARPC1A和KLHDC3的一种或多种基因的表达水平来确定作为应答性或无应答性的受试者的分类(即,确定或预测)。在一些实施方案中,使用机器学习模型确定分类。在一些实施方案中,机器学习模型是随机森林模型。

[0129] 在一些实施方案中,所提供的方法在不使用临床协变量分类器(CC)的情况下将受试者的分类确定为应答者或无应答者。在一些实施方案中,所提供的方法在不使用组合基因组-临床分类器(GCC)的情况下将受试者的分类确定为应答者或无应答者。在一些实施方案中,所提供的方法在不使用临床协变量分类器(CC)和不使用组合基因组-临床分类器(GCC)的情况下将受试者的分类确定为应答者或无应答者。

[0130] 在一些实施方案中,用于分类器的一种或多种基因选自由以下各项组成的组:COR02B、STOML2、CHMP7、SPINT2、CYSLTR2、SSNA1、IMPDH2、GOLGA1、SUOX、ARPC1A和KLHDC3。

[0131] 在一些实施方案中,分类器中的基因表达可以使用mRNA和/或蛋白质表达数据集来鉴定,例如,可以是或已经从经验证的生物数据(例如,来源于诸如基因表达综合数据库(Gene Expression Omnibus,“GEO”)的可公开获得的数据库的生物数据)制备。在一些实施方案中,可通过比较特定疗法(例如,抗TNF疗法)的已知有应答和已知无应答的先前受试者的基因表达水平来衍生分类器。在一些实施方案中,某些基因(即,签名基因)选自待用于开发分类器的该基因表达数据队列。

[0132] 在一些实施方案中,通过与通过引用并入本文的Santolini,“A personalized, multiomics approach identifies genes involved in cardiac hypertrophy and heart failure”, Systems Biology and Applications, (2018) 4:12; doi:10.1038/s41540-018-0046-3中报告的那些类似的方法鉴定签名基因。在一些实施方案中,通过比较已知有应答和无应答的先前受试者的基因表达水平并鉴定两组之间的显著

变化来鉴定签名基因,其中显著变化可以是表达上的巨大差异(例如,大于2倍的变化)、表达上的微小差异(例如,小于2倍的变化),或两者。在一些实施方案中,通过表达差异的显著性对基因进行排序。在一些实施方案中,通过基因表达与应答结果之间的Pearson相关性来测量显著性。在一些实施方案中,通过表达差异的显著性从排序中选择签名基因。在一些实施方案中,所选择的签名基因的数量小于所分析基因的总数。在一些实施方案中,选择200种或更少的签名基因。在一些实施方案中,选择100种或更少的基因。

[0133] 在一些实施方案中,结合其在人类相互作用组(HI)(蛋白质-蛋白质相互作用的图谱)上的位置选择签名基因。以这种方式使用HI包括认识到mRNA活性是动态的,并确定对理解某些疾病至关重要的蛋白质的实际上过度和不足的表达。在一些实施方案中,与对某些疗法(即,抗TNF疗法)的应答相关联的基因可聚类在HI图谱上的离散模块中(即,形成基因的聚类)。此类聚类的存在与基本的潜在疾病生物学的存在相关联。在一些实施方案中,分类器源自从HI图谱上的基因聚类中选择的签名基因。因此,在一些实施方案中,分类器源自与人类相互作用组图谱上对抗TNF疗法的应答相关联的基因聚类。

[0134] 在一些实施方案中,当映射到人类相互作用组图谱上时,与对某些疗法的应答相关联的基因表现出某些拓扑性质。例如,在一些实施方案中,多个基因与抗TNF疗法的应答相关联,并通过其在人类相互作用组图谱上的位置(即拓扑性质,例如,它们的彼此邻近性)进行表征。

[0135] 在一些实施方案中,与对某些疗法(即,抗TNF疗法)的应答相关联的基因可在HI图谱上彼此非常接近地存在。所述邻近基因不一定需要共享基本的潜在疾病生物学。也就是说,在一些实施方案中,邻近基因不共享显著的蛋白质相互作用。因此,在一些实施方案中,分类器源自在人类相互作用组图谱上邻近的基因。在一些实施方案中,分类器源自人类相互作用组图谱上的某些其他拓扑特征。

[0136] 在一些实施方案中,当与HI图谱结合使用时,可通过扩散状态距离(Diffusion State Distance,DSD)(参见Cao等人,PLoS One,8(10):e76339(Oct.23,2013))确定与对某些疗法(即,抗TNF疗法)的应答相关联的基因。

[0137] 在一些实施方案中,通过(1)基于与已知应答者和已知无应答者相比的基因表达差异的显著性对基因进行排序;(2)从排序的基因中选择基因并将所选基因映射到人类相互作用组图谱上;以及(3)从映射到人类相互作用组图谱上的基因中选择签名基因来选择签名基因。

[0138] 在一些实施方案中,将签名基因(例如,从Santolini方法中选择,或使用各种网络拓扑特性,包括但不限于基于聚类、接近性和扩散的方法选择)提供给概率神经网络,从而提供(即,“训练”)分类器。在一些实施方案中,概率神经网络实施D.F.Specht在“Probabilistic Neural Networks,”Neural Networks,3(1):109-118(1990)(其通过引用并入本文)中提出的算法。在一些实施方案中,概率神经网络以R统计语言编写,并且已知通过定量变量的向量描述的一组观察结果,将观察结果分类为给定数量的组(例如,应答者和无应答者)。该算法使用获自己知应答者和无应答者的签名基因的数据集进行训练,并猜测所提供的新观察结果。在一些实施方案中,概率神经网络是源自<https://CRAN.R-project.org/package=pnn>的概率神经网络。在一些实施方案中,根据随机森林模型分析签名基因以提供分类器。

2. 单核苷酸多态性

[0139] 本公开还包括这样的见解,即可以通过RNA序列数据鉴定单核苷酸多态性(SNP)。也就是说,通过将RNA序列数据与参考人类基因组进行比较,例如,通过将RNA序列数据映射到GRCh38人类基因组。在不受理论约束的情况下,认为与分类器中使用的RNA序列相关的SNP的存在可有助于鉴定对某些疗法(例如,抗TNF疗法)有应答或无应答的受试者亚群。也就是说,可以使用网络医学和通路富集分析来分析区分性基因和含SNP的RNA的蛋白质产物。由包括在分类器中的区分性基因和含SNP的RNA编码的蛋白质可以在例如人类相互作用组的图谱上重叠,以通过鉴定某些区分性基因集合来帮助鉴定受试者的某些亚群。

[0140] 在一些实施方案中,所提供的分类器和使用此类分类器的方法结合与单核苷酸多态性(SNP)相关的评估。在一些实施方案中,本公开提供了一种开发分类器以用于针对一个或多个治疗属性对受试者分层的方法,其包括以下步骤:分析针对至少一个治疗属性代表至少两个不同类别的受试者中表达的RNA的序列数据;由序列数据评估一个或多个单核苷酸多态性(SNP)的存在;确定一个或多个SNP的存在与至少一个治疗属性相关;以及将一个或多个SNP包括在分类器中。

[0141] 在一些实施方案中,本公开提供了开发分类器以用于通过分析针对至少一个治疗属性代表至少两个不同类别的受试者中表达的RNA的序列数据来针对一个或多个治疗属性对受试者分层的方法的改进,其包括:由序列数据评估一个或多个单核苷酸多态性(SNP)的存在;以及确定一个或多个SNP的存在与至少一个治疗属性相关;以及将一个或多个SNP的存在包括在分类器中。

[0142] 在一些实施方案中,一个或多个SNP选自:

SNP
chr1.161644258
chr1.2523811

SNP
chr11.107967350
chr17.38031857
chr7.128580042
rs10774624
rs10985070
rs11889341
rs1571878
rs1633360
rs17668708
rs1877030
rs1893592
rs1980422
rs2228145
rs2233424
rs2236668
rs2301888
rs2476601
rs3087243

SNP
rs3218251
rs331463
rs34536443
rs34695944
rs4239702
rs4272
rs45475795
rs508970
rs5987194
rs657075
rs6715284
rs706778
rs72634030
rs73013527
rs73194058
rs773125
rs7752903
rs8083786

SNP
rs9653442

3. 临床特征

[0143] 分类器还可以结合附加信息,以便进一步提高分类器在应答者与无应答者之间鉴定的预测能力。例如,在一些实施方案中,分类器通过评估以下各项进行开发:表达水平与临床应答性或无应答性显著相关(例如,以线性和/或非线性方式)的一种或多种基因;以及以下各项中的至少一个:一种或多种基因的表达序列中一个或多个单核苷酸多态性(SNP)的存在;或有应答和无应答的先前受试者的至少一种临床特征。也就是说,在一些实施方案中,通过评估表达水平与临床应答性或无应答性显著相关(例如,以线性和/或非线性方式)的一种或多种基因以及一种或多种基因的表达序列中一个或多个单核苷酸多态性(SNP)的存在来开发分类器。在一些实施方案中,通过评估表达水平与临床应答性或无应答性显著相关(例如,以线性和/或非线性方式)的一种或多种基因以及有应答和无应答的先前受试者的至少一种临床特征来开发分类器。

[0144] 本公开还包括这样的见解,即某些临床特征(例如,BMI、性别、年龄等)可以结合到本文提供的分类器中。在一些实施方案中,所提供的分类器和使用此类分类器的方法结合了与临床特征相关的评估。在一些实施方案中,本公开提供了一种开发分类器以用于针对一个或多个治疗属性对受试者分层的方法,其包括以下步骤:分析针对至少一个治疗属性代表至少两个不同类别的受试者中表达的RNA的序列数据;评估一种或多种临床特征的存在;确定与所述临床特征相关的表达与至少一个治疗属性相关;以及将一种或多种临床特征包括在分类器中。

[0145] 在一些实施方案中,至少一种临床特征选自:体重指数(BMI)、性别、年龄、种族、先前疗法治疗、疾病持续时间、C反应蛋白(CRP)水平、抗环瓜氨酸肽的存在、类风湿因子的存在、患者整体评估、治疗应答率(例如,ACR20、ACR50、ACR70)以及其组合。

[0146] 在一些实施方案中,临床特征选自:

临床特征
年龄
出生性别
疾病持续时间（以年为单位）
种族（包括白人、亚洲人、黑人、混血人种、美洲原住民、太平洋岛民和其他种族）
纤维肌痛史
慢性血管疾病史（包括急性冠脉综合征、冠状动脉疾病、充血性心力衰竭、高血压、心肌梗死、外周动脉疾病、中风、不稳定型心绞痛、心脏骤停、血运重建术和室性心律失常）
导致住院的严重感染史（包括粘液囊或关节感染、蜂窝组织炎、鼻窦炎、憩室炎、败血症、肺炎、支气管炎、胃炎、脑膜炎、尿路感染、上呼吸道感染和结核病）
癌症史（包括乳腺癌、肺癌、皮肤癌、淋巴瘤，但不包括非黑色素瘤皮肤）
BMI
吸烟状态（包括从不吸烟、以前吸烟或目前吸烟）
强的松剂量
DMARD 剂量

临床特征
基线时的 C 反应蛋白水平
基线时的 DAS28-CRP
基线时的肿胀的 28 关节计数
基线时的压痛的 28 关节计数
基线时的患者整体评估
基线时的医师整体评估
基线时的 CDAI
基线时的经过修改的健康评估问卷得分
基线时的患者疼痛评估
使用 DAS28-CRP 获得的基线时的 EULAR 应答（包括差、中或好）
抗 CCP 状态（阳性或阴性）
基线时的抗 CCP 滴度
类风湿因子状态（阳性或阴性）
基线时的类风湿因子滴度

4. 验证分类器

[0147] 替代地或另外地，在一些实施方案中，可以使用已知应答者和无应答者的队列，使用留一法交叉和/或k折交叉验证，在概率神经网络中训练分类器。在一些实施方案中，这样的过程在分析中留一个样本（即留一法），并且仅基于其余样本训练分类器。在一些实施方案中，随后使用升级的分类器预测留下样本的应答概率。在一些实施方案中，例如，可以迭代地重复这样的过程，直到所有样本被留一次。在一些实施方案中，这样的过程将已知应答者和无应答者的队列随机分配为k个大小相等的组。在k个组中，保留一个组作为验证数据，以用于测试模型，并且其余组用作训练数据。这样的过程可以重复k次，其中k个组中的每一个只使用一次作为验证数据。在一些实施方案中，结果是训练集中的每个样本的概率得分。可以将这种概率得分与实际应答结果相关。递归操作曲线 (Recursive Operating Curves, ROC) 可以用于估计分类器的性能。在一些实施方案中，约0.6或更高的曲线下面积 (AUC) 反

映了合适的经验验证分类器。在一些实施方案中,0.9的阴性预测值(NPV)反映了合适的经验验证分类器。在一些实施方案中,可以在完全独立(即,盲的)队列中测试分类器,以例如确认适合性(即,使用留一法和/或k折交叉验证)。因此,在一些实施方案中,所提供的方法还包括验证分类器的一个或多个步骤,例如,通过将应答概率分配给一组已知应答者和无应答者;并且针对盲组应答者和无应答者检查分类器。这些过程的输出是经过训练的分类器,其可用于确定受试者是否将对特定疗法(例如,抗TNF疗法)产生应答。

[0148] 因此,在一些实施方案中,分类器被建立用于区分已接受一种类型的疗法(例如,抗TNF疗法)的有应答和无应答的先前受试者。该分类器可以预测受试者是否将对给定疗法产生应答。在一些实施方案中,有应答和无应答的先前受试者患有相同的疾病、病症或病况。

[0149] 在一些实施方案中,通过微阵列、RNA测序、实时定量逆转录PCR(qRT-PCR)、珠粒阵列、ELISA和蛋白质表达中的至少一种测量受试者的基因。

[0150] 在一些实施方案中,使用先前已使用抗TNF疗法治疗但独立于用于制备分类器的受试者队列的受试者队列来验证分类器。在一些实施方案中,使用基因表达数据、SNP数据或临床特征升级分类器。在一些实施方案中,当在验证队列中以60%或更高的准确度预测90%或更多的无应答受试者时,分类器被视为“经验证”。

[0151] 在一些实施方案中,分类器以至少60%的准确度预测受试者的应答性,从而在至少100名受试者的群体中预测应答性。在一些实施方案中,分类器在至少150名受试者的群体中以至少60%的准确度预测受试者的应答性。在一些实施方案中,分类器在至少170名受试者的群体中以至少60%的准确度预测受试者的应答性。在一些实施方案中,分类器在至少200名或更多受试者的群体中以至少60%的准确度预测受试者的应答性。

[0152] 在一些实施方案中,分类器在至少100名受试者的群体中以至少80%的准确度预测受试者的应答性。在一些实施方案中,分类器在至少150名受试者的群体中以至少80%的准确度预测受试者的应答性。在一些实施方案中,分类器在至少170名受试者的群体中以至少80%的准确度预测受试者的应答性。在一些实施方案中,分类器在至少200名或更多受试者的群体中以至少80%的准确度预测受试者的应答性。

[0153] 在一些实施方案中,分类器在至少100名受试者的群体中以至少90%的准确度预测受试者的应答性。在一些实施方案中,分类器在至少150名受试者的群体中以至少90%的准确度预测受试者的应答性。在一些实施方案中,分类器在至少170名受试者的群体中以至少90%的准确度预测受试者的应答性。在一些实施方案中,分类器在至少200名或更多受试者的群体中以至少90%的准确度预测受试者的应答性。

B. 检测一个或多个基因签名和/或SNP

[0154] 使用经过训练的分类器检测受试者中的基因签名是本领域技术人员的常规事项。换句话说,通过首先定义(来自分类器的)基因签名,可以使用多种方法确定一个受试者或一组受试者是否表达已建立的基因签名。例如,在一些实施方案中,从业者可以在施用疗法之前从受试者获得血液或组织样本,并从所述血液或组织样本中提取和分析mRNA图谱(mRNA profile)。mRNA图谱的分析可以通过本领域技术人员已知的任何方法执行,包括但不限于基因阵列、RNA测序、纳米串(nanostring)测序、实时定量逆转录PCR(qRT-PCR)、珠粒阵列或酶联免疫吸附测定(ELISA)。因此,在一些实施方案中,本公开提供了确定将受试者

分类为应答者还是无应答者的方法,其包括通过微阵列、RNA测序、实时定量逆转录PCR (qRT-PCR)、珠粒阵列和ELISA中的至少一种测量基因表达。在一些实施方案中,本公开提供了确定将受试者分类为应答者还是无应答者的方法,其包括通过RNA测序(即,RNAseq)测量受试者的基因表达。

[0155] 本公开还包括这样的见解,即可以通过RNA序列数据鉴定单核苷酸多态性(SNP)。也就是说,通过将RNA序列数据与参考人类基因组进行比较,例如,通过将RNA序列数据映射到GRCh38人类基因组。在不受理论约束的情况下,认为与分类器中使用的RNA序列相关的SNP的存在可有助于鉴定对某些疗法(例如,抗TNF疗法)有应答或无应答的受试者亚群。也就是说,可以使用网络医学和通路富集分析来分析区分性基因和含SNP的RNA的蛋白质产物。由包括在分类器中的区分性基因和含SNP的RNA编码的蛋白质可以在例如人类相互作用组的图谱上重叠,以通过鉴定某些区分性基因集来帮助鉴定受试者的某些亚群。

[0156] 在一些实施方案中,通过减去背景数据、校正批次效应并除以看家基因的平均表达来测量基因表达。参见Eisenberg&Levanon,“Human housekeeping genes, revisited,” *Trends in Genetics*, 29 (10) :569-574 (2013年10月)。在微阵列数据分析的上下文中,背景差减是指从每个探针特征的荧光信号强度中减去芯片上与任何mRNA序列不互补的探针特征产生的平均荧光信号,即由非特异性结合产生的信号。背景差减可以使用不同的软件包执行,诸如Affymetrix基因表达控制台(Gene Expression Console)。看家基因参与基本的细胞维持,并且因此,预期在所有细胞和条件下保持恒定的表达水平。感兴趣的基因(即应答签名中的那些基因)的表达水平可以通过将表达水平除以一组选定看家基因的平均表达水平来归一化。该看家基因归一化程序针对实验变异性校准基因表达水平。此外,归一化方法,诸如针对不同批次微阵列的变异性进行校正的鲁棒多阵列平均(“RMA”)可以在Illumina和/或Affymetrix平台推荐的R软件包中获得。对归一化数据进行对数变换,并去除样本中检测率低的探针。此外,从分析中去除没有可用基因符号或Entrez ID的探针。

[0157] 在一些实施方案中,本公开提供了一种试剂盒,其包括被建立用于区分已接受抗TNF疗法的有应答和无应答的先前受试者的分类器。

C. 使用分类器

1. 患者分层(Patient Stratification)

[0158] 除其他外,本公开提供了用于预测对抗TNF疗法的应答性的技术。在一些实施方案中,所提供的技术表现出优于先前方法的跨队列的一致性和/或准确度。

[0159] 因此,本公开提供了用于患者分层、定义和/或区分应答者和无应答者群体的技术。例如,在一些实施方案中,本公开提供了用抗TNF疗法治疗受试者的方法,在一些实施方案中,所述方法包括以下步骤:向已通过分类器确定为有应答的受试者施用抗TNF疗法,所述分类器被建立用于区分已接受抗TNF疗法的有应答和无应答的先前受试者。

[0160] 在一些实施方案中,本公开提供了一种开发分类器以用于针对一个或多个治疗属性对受试者分层的方法,其包括以下步骤:分析针对至少一个治疗属性代表至少两个不同类别的受试者中表达的RNA的序列数据;由序列数据评估一个或多个单核苷酸多态性(SNP)的存在;确定一个或多个SNP的存在与至少一个治疗属性相关;以及将一个或多个SNP包括在分类器中。

2. 疗法监测

[0161] 此外,本公开提供了用于监测用于给定受试者或受试者队列的疗法的技术。由于受试者的基因表达水平可随时间变化,因此在一些情况下,可能有必要或希望在一个或多个时间点(例如,以指定的和/或周期性的间隔)评估受试者。

[0162] 在一些实施方案中,在一定时间内进行重复监测允许或实现受试者的基因表达谱或特征中可能影响正在进行治疗方案的一个或多个变化的检测。在一些实施方案中,响应于对受试者施用的特定疗法的持续、改变或暂停而检测变化。在一些实施方案中,疗法可以例如通过增加或减少已经用来治疗受试者的一种或多种药剂或治疗的施用频率和/或量来改变。替代地或另外地,在一些实施方案中,疗法可通过添加具有一种或多种新药剂或治疗的疗法来改变。在一些实施方案中,疗法可通过暂停或停止一种或多种特定药剂或治疗来改变。

[0163] 仅举一个实例,如果受试者最初被归类为有应答(因为受试者的基因表达通过分类器被确定为与疾病、病症或病况相关联),则可以施用给定的抗TNF疗法。以给定的间隔(例如,每六个月、每年等),可以再次对受试者进行测试,以确保他们仍然满足对给定抗TNF疗法的“有应答”的标准。在给定受试者的基因表达水平随时间变化,并且受试者不再表达与疾病、病症或病况相关的基因,或者现在表达与无应答性相关的基因的情况下,可以改变受试者的疗法以适应基因表达的改变。

[0164] 因此,在一些实施方案中,本公开提供了向先前通过分类器确定为对抗TNF疗法有应答的受试者施用疗法的方法。

[0165] 在一些实施方案中,本公开提供了方法,所述方法进一步包括在施用之前通过分类器确定受试者不是应答者;以及施用抗TNF疗法的替代疗法。

[0166] 在一些实施方案中,通过微阵列、RNA测序、实时定量逆转录PCR(qRT-PCR)、珠粒阵列、ELISA和蛋白质表达中的至少一种测量受试者的基因。

[0167] 在一些实施方案中,受试者患有选自以下各项的疾病、病症或病况:类风湿性关节炎、银屑病性关节炎、强直性脊柱炎、克罗恩病、溃疡性结肠炎、慢性银屑病、化脓性汗腺炎、多发性硬化症和幼年特发性关节炎。

[0168] 在一些实施方案中,抗TNF疗法是或包括施用英夫利昔单抗、阿达木单抗、依那西普、赛妥珠单抗、goliluma或其生物仿制药。在一些实施方案中,抗TNF疗法是或包括施用英夫利昔单抗或阿达木单抗。

[0169] 在一些实施方案中,有应答和无应答的先前受试者患有相同的疾病、病症或病况。

[0170] 在一些实施方案中,施用抗TNF疗法的受试者患有与先前有应答和无应答的先前受试者相同的疾病、病症或病况。

[0171] 在一些实施方案中,所述疾病、病症或病况选自类风湿性关节炎、银屑病性关节炎、强直性脊柱炎、克罗恩病、溃疡性结肠炎、慢性银屑病、化脓性汗腺炎、多发性硬化症和幼年特发性关节炎。

[0172] 在一些实施方案中,所述疾病、病症或病况为类风湿性关节炎。

[0173] 在一些实施方案中,所述疾病、病症或病况为溃疡性结肠炎。

D. 治疗方法

[0174] 在一些实施方案中,针对其施用抗TNF疗法或停止抗TNF疗法(和/或施用替代疗法)的受试者或群体是被确定为表现出一种或多种基因(并且通常多种基因)的特定表达水

平的受试者或群体。在一些实施方案中,一种或多种基因被确定为具有低于特定阈值的表达水平;替代地或另外地,在一些实施方案中,一种或多种基因被确定为具有低于特定阈值的表达水平。在一些实施方案中,特定的基因集被确定为具有表达模式,其中相对于特定阈值评估每个基因(并且例如,被确定为高于、低于此类阈值或与其相当)。

[0175] 在一些实施方案中,本公开提供了一种治疗患有疾病、病症或病况的受试者的方法,其包括向已被确定为表现出低于一种或多种基因的特定表达水平的受试者施用抗TNF疗法的替代方案。

[0176] 因此,在一些实施方案中,本公开提供了一种方法,其包括以下步骤:

(a) 通过计算设备的处理器接收与患有炎症性疾病(例如,自身免疫性疾病、类风湿性关节炎)的受试者的一种或多种基因中的每一种的表达水平相对应的数据,所述一种或多种基因包括选自由CORO2B、STOML2、CHMP7、SPINT2、CYSLTR2、SSNA1、IMPDH2、GOLGA1、SUOX、ARPC1A和KLHDC3组成的组的至少一个成员(例如,一个、两个、三个、四个、五个、六个、七个、八个、九个、十个或全部十一个成员);

(b) 使用在步骤(a)中接收的数据由处理器自动确定受试者为对抗TNF疗法无应答的分类;以及,任选地,

(c) 向受试者开处方和/或施用第二疗法(例如,抗TNF疗法的替代方案)以治疗疾病,从而避免向受试者开处方和/或施用第一疗法。

[0177] 在一些实施方案中,本公开提供了一种方法,其包括以下步骤:(a) 通过计算设备的处理器接收与患有疾病(例如,自身免疫性疾病,例如,类风湿性关节炎)的受试者的一种或多种基因中的每一种的表达水平相对应的数据,所述一种或多种基因包括选自由CORO2B、STOML2、CHMP7、SPINT2、CYSLTR2、SSNA1、IMPDH2、GOLGA1、SUOX、ARPC1A和KLHDC3组成的组的至少一个成员(例如,一个、两个、三个、四个、五个、六个、七个、八个、九个、十个或全部十一个成员);

(b) 使用在步骤(a)中接收的数据由处理器自动确定受试者为对第一疗法(例如,抗TNF疗法)有应答的分类;以及,任选地,

(c) 为受试者开处方和/或施用第一疗法以治疗疾病。

[0178] 在一些实施方案中,本公开提供了向已通过分类器确定为有应答的受试者施用抗TNF疗法的方法,所述分类器被建立用于区分已接受抗TNF疗法的有应答和无应答的先前受试者(即,其中已通过回顾性分析建立分类器,以在对其所接受的抗TNF疗法产生应答的患者与不产生应答的那些患者之间进行区分);其中所述分类器通过评估以下各项进行开发:表达水平与临床应答性 or 无应答性显著相关(例如,以线性和/或非线性方式)的一种或多种基因;以及以下各项中的至少一个:表达序列中一个或多个单核苷酸多态性(SNP)的存在;以及有应答和无应答的先前受试者的至少一种临床特征。

[0179] TNF介导的病症目前通过抑制TNF,特别是通过施用抗TNF剂(即,通过抗TNF疗法)进行治疗。在美国批准使用的抗TNF剂的实例包括单克隆抗体,诸如阿达木单抗(Humira®)、赛妥珠单抗(Cimiza®)、英夫利昔单抗(Remicade®)以及诱骗循环受体融合蛋白,诸如依那西普(Enbrel®)。这些药剂目前被批准用于根据下文在表2中列出的给药方案治疗适应症:

表 2

适应症	阿达木单抗 ¹	赛妥珠单抗 ¹	英夫利昔单抗 ²	依那西普 ¹	戈利木单抗 ¹	戈利木单抗 ²
幼年特发性关节炎	<ul style="list-style-type: none"> • 10 kg (22 lbs) 至 <15 kg (33 lbs): 每隔一周 10 mg • 15 kg (33 lbs) 至 <30 kg (66 lbs): 每隔一周 20 mg • ≥ 30 kg (66 lbs): 	N/A	N/A	每周 0.8 mg/kg, 每周最多 50 mg	N/A	N/A

适应症	阿达木单抗 ¹	赛妥珠单抗 ¹	英夫利昔单抗 ²	依那西普 ¹	戈利木单抗 ¹	戈利木单抗 ²
	每隔一周 40 mg					
银屑病性关节炎	每隔一周 40 mg	最初以及在第 2 周和第 4 周 400 mg, 然后每隔一周 200 mg; 对于维持给药, 每周 400 mg	在第 0、2 和 6 周, 然后在第 8 周, 5 mg/kg	每周一次 50 mg, 含或不含甲氨蝶呤	通过皮下注射施用 50 mg, 每月一次	N/A
类风湿性关节炎	每隔一周 40 mg	最初以及在第 2 周和第 4 周 400 mg, 然后每隔一周 200 mg; 对于维持给药, 每周	与甲氨蝶呤联合, 在第 0、2 和 6 周, 然后在第 8 周, 3 mg/kg	每周一次 50 mg, 含或不含甲氨蝶呤	50 mg, 每月一次	在第 0 周和第 4 周, 在 30 分钟内静脉注射 2 mg/kg,

适应症	阿达木单抗 ¹	赛妥珠单抗 ¹	英夫利昔单抗 ²	依那西普 ¹	戈利木单抗 ¹	戈利木单抗 ²
		4周 400 mg				然后每8周
强直性脊柱炎	每隔一周 40 mg	最初以及在第2周和第4周 400 mg (作为2次皮下注射给予, 每次 200 mg), 然后每隔一周 200 mg 或每4周 400 mg	在第0、2和6周、然后每6周, 5 mg/kg	50 mg, 每周一次	通过皮下注射施用 50 mg, 每月一次	N/A
成人克罗恩病	<ul style="list-style-type: none"> 初始剂量 (第1天): 160 	<ul style="list-style-type: none"> 最初以及第2周和第4 	在第0、2和6周、然后每8周,	N/A	N/A	N/A

适应症	阿达木单抗 ¹	赛妥珠单抗 ¹	英夫利昔单抗 ²	依那西普 ¹	戈利木单抗 ¹	戈利木单抗 ²
小儿克罗恩病	mg • 两周后(第15天)第二次剂量: 80 mg • 两周后(第29天): 开始每隔一周40 mg 的维持剂量	周 400 mg • 每四周以400 mg 继续	5 mg/kg			
	17 kg (37 lbs) 至 < 40 kg (88 lbs): • 初始剂量 (第1天): 80 mg		在第0、2和6周、然后每8周, 5 mg/kg。			

适应症	阿达木单抗 ¹	赛妥珠单抗 ¹	英夫利单抗 ²	依那普 ¹	戈利木单抗 ¹	戈利木单抗 ²
	<ul style="list-style-type: none"> ● 两周后(第 15 天)第二次剂量: 40 mg ● 两周后(第 29 天): 开始每周隔一周 20 mg 的维持剂量 <p>≥ 40 kg (88 lbs):</p> <ul style="list-style-type: none"> ● 初始剂量 (第 1 天): 160 mg 					

适应症	阿达木单抗 ¹	赛妥珠单抗 ¹	英夫利昔单抗 ²	依那西普 ¹	戈利木单抗 ¹	戈利木单抗 ²
	<ul style="list-style-type: none"> ● 两周后(第15天)第二次剂量: 80 mg ● 两周后(第29天): 开始每周隔一周40 mg的维持剂量 					
溃疡性结肠炎	<ul style="list-style-type: none"> ● 初始剂量(第1天): 160 mg ● 两周后(第15天)第 	N/A	在第0、2和6周、然后每8周, 5 mg/kg。	N/A	N/A	N/A

适应症	阿达木单抗 ¹	赛妥珠单抗 ¹	英夫利昔单抗 ²	依那西普 ¹	戈利木单抗 ¹	戈利木单抗 ²
	二剂量： 80 mg • 两周后(第29天)：开始每周隔一周的40 mg的维持剂量					
斑块型银屑病	初始剂量 80 mg；初始剂量后一周开始每隔一周40 mg	N/A	N/A	50 mg, 每周两次, 持续3个月, 然后 50 mg, 每周一次	N/A	N/A
化脓性	• 初始剂量	N/A	N/A	N/A	N/A	N/A

适应症	阿达木单抗 ¹	赛妥珠单抗 ¹	英夫利昔单抗 ²	依那西普 ¹	戈利木单抗 ¹	戈利木单抗 ²
汗腺炎	(第 1 天): 160 mg • 两周后(第 15 天) 第二剂量: 80 mg • 第三剂量 (第 29 天) 和后续剂量: 每周 40 mg					
葡萄膜炎	初始剂量 80 mg; 初始剂量后一周开始, 每隔一	N/A	N/A	N/A	N/A	N/A

适应症	阿达木单抗 ¹	赛妥珠单抗 ¹	英夫利昔单抗 ²	依那普 ¹	戈利木单抗 ¹	戈利木单抗 ²
	周 40 mg					

1. 通过皮下注射施用。
2. 通过静脉内输注施用。

[0180] 本公开提供了与抗TNF疗法相关的技术,包括表2中所述的那些治疗方案。在一些实施方案中,抗TNF疗法是或包括施用英夫利昔单抗(Remicade[®])、阿达木单抗

(Humira[®])、赛妥珠单抗(Cimiza[®])、依那西普(Enbrel[®])或其生物仿制药。在一些实施方案中,抗TNF疗法是或包括施用英夫利昔单抗(Remicade[®])或阿达木单抗(Humira[®])。在一些实施方案中,抗TNF疗法是或包括施用英夫利昔单抗(Remicade[®])。在一些实施方案中,抗TNF疗法是或包括施用阿达木单抗(Humira[®])。

[0181] 在一些实施方案中,抗TNF疗法是或包括施用生物仿制药抗TNF剂。在一些实施方案中,抗TNF剂选自英夫利昔单抗原生物仿制药,诸如CT-P13、BOW015、SB2、Incoctra、Renflexis和Ixifi、阿达木单抗原生物仿制药,诸如ABP 501 (AMGEVITA[™])、Adfrar和Hulio[™],以及依那西普生物仿制药,诸如HD203、SB4(Benepali[®])、GP2015、Erlezi和Intacept。

[0182] 在一些实施方案中,本公开提供了一种治疗患有自身免疫性病症的受试者的方法,所述方法包括以下步骤:向已通过分类器确定为有应答的受试者施用抗TNF疗法,所述分类器被建立用于区分已接受抗TNF疗法的队列中有应答和无应答的先前受试者,其中分类器通过评估以下各项进行开发:表达水平与临床应答性或无应答性显著相关(例如,以线性和/或非线性方式)的一种或多种基因;以下各项中的至少一个:一种或多种基因的表达序列中一个或多个单核苷酸多态性(SNP)的存在;或有应答和无应答的先前受试者的至少一种临床特征;并且其中,分类器通过不是已接受抗TNF疗法的队列的独立队列进行验证。

[0183] 在一些实施方案中,从已接受抗TNF疗法的队列中的受试者得到的数据是一种类型(例如,微阵列、RNAseq等),并且独立队列中用于验证分类器的数据源自不同类型(例如,微阵列、RNAseq)。因此,在一些实施方案中,使用源自有应答和无应答的先前受试者的微阵列分析建立分类器。在一些实施方案中,使用源自独立队列的RNAseq数据验证分类器。

[0184] 如本文所提供,某些基因组合的基因表达水平(任选地与某些临床特征结合)可以用于确定患者/受试者是否对特定疗法(例如,抗TNF疗法)产生应答。例如,在一些实施方案中,基于11种基因的集合(COR02B、STOML2、CHMP7、SPINT2、CYSLTR2、SSNA1、IMPDH2、GOLGA1、SUOX、ARPC1A和KLHDC3)的基因表达水平的经验证的随机森林模型获得了作为抗TNF疗法应答者的受试者分类中0.72的曲线下面积(AUC)、0.85的阴性预测值(NPV)和0.61的真阴性率(TNR)。在不受理论约束的情况下,这些基因似乎与炎症性过程、RA、自身免疫和抗TNF剂的作用机制高度相关。获得的高NPV对于鉴定无应答者特别有价值,这在药物副作用可能很严重并且实现LDA或缓解可能需要很长时间的RA治疗中非常重要。例如,风湿病专家可以能够准确预测患有RA的给定受试者将不对抗TNF疗法产生应答,并且因此能够开出替代疗法的处方,从而使受试者避免抗TNF药物的副作用风险和由无效疾病治疗造成的延迟。

[0185] 在一些实施方案中,受试者患有选自以下各项的疾病、病症或病况:类风湿性关节炎、银屑病性关节炎、强直性脊柱炎、克罗恩病、溃疡性结肠炎、慢性银屑病、化脓性汗腺炎和幼年特发性关节炎。

[0186] 在一些实施方案中,抗TNF疗法是或包括施用英夫利昔单抗、阿达木单抗、依那西普、赛妥珠单抗、goliluma或其生物仿制药。在一些实施方案中,抗TNF疗法是或包括施用英夫利昔单抗或阿达木单抗。

[0187] 在一些实施方案中,抗TNF疗法的替代疗法选自利妥昔单抗(Rituxan[®])、沙利姆单抗(Kevzara[®])、枸橼酸托法替尼(Xeljanz[®])、来氟米特(Arava[®])、维多珠单抗

(Entyvio[®])、托珠单抗(Actemra[®])、阿那白滞素(Kineret[®])和阿巴西普(Orencia[®])。

[0188] 通常,所提供的公开在预期或实施抗TNF疗法的施用的任何情况下都是有用的。在一些实施方案中,所提供的技术可用于诊断和/或治疗患有与异常(例如,升高的)TNF表达和/或活性相关联的疾病、病症或病况的受试者。在一些实施方案中,所提供的技术可用于监测正在接受或已经接受抗TNF疗法的受试者。在一些实施方案中,所提供的技术鉴定受试者是否将对给定的抗TNF疗法产生应答。在一些实施方案中,所提供的技术鉴定受试者是否将对给定的抗TNF疗法形成耐药性。

E. 疾病、病症或病况

[0189] 通常,所提供的公开在预期或实施抗TNF疗法的施用的任何情况下都是有用的。在一些实施方案中,所提供的技术可用于诊断和/或治疗患有与异常(例如,升高的)TNF表达和/或活性相关联的疾病、病症或病况的受试者。在一些实施方案中,所提供的技术可用于监测正在接受或已经接受抗TNF疗法的受试者。在一些实施方案中,所提供的技术鉴定受试者是否将对给定的抗TNF疗法产生应答。在一些实施方案中,所提供的技术鉴定受试者是否将对给定的抗TNF疗法形成耐药性。

[0190] 因此,本公开提供了与TNF相关的各种病症的治疗相关的技术,包括表2中列出的那些。在一些实施方案中,受试者患有选自以下各项的疾病、病症或病况:类风湿性关节炎、银屑病性关节炎、强直性脊柱炎、克罗恩病(成人或小儿)、溃疡性结肠炎、炎症性肠病、慢性银屑病、斑块型银屑病、化脓性汗腺炎、哮喘、葡萄膜炎和幼年特发性关节炎。在一些实施方案中,所述疾病、病症或病况为类风湿性关节炎。在一些实施方案中,所述疾病、病症或病况为银屑病关节炎。在一些实施方案中,所述疾病、病症或病况为强直性脊柱炎。在一些实施方案中,所述疾病、病症或病况为克罗恩病。在一些实施方案中,所述疾病、病症或病况为成人克罗恩病。在一些实施方案中,所述疾病、病症或病况为小儿克罗恩病。在一些实施方案中,所述疾病、病症或病况为炎症性肠病。在一些实施方案中,所述疾病、病症或病况为溃疡性结肠炎。在一些实施方案中,所述疾病、病症或病况为慢性银屑病。在一些实施方案中,所述疾病、病症或病况为斑块型银屑病。在一些实施方案中,所述疾病、病症或病况为化脓性汗腺炎。在一些实施方案中,所述疾病、病症或病况为哮喘。在一些实施方案中,所述疾病、病症或病况为葡萄膜炎。在一些实施方案中,所述疾病、病症或病况为幼年特发性关节炎。

[0191] 在一些实施方案中,所述疾病、病症或病况为环状肉芽肿、类脂质渐进性坏死、化脓性汗腺炎、坏疽性脓皮病、斯威特氏综合征、角层下脓疱性皮肤病、系统性红斑狼疮、硬皮病、皮炎、白塞病、急性/慢性移植物抗宿主病、毛发红糠疹、干燥综合征、韦格纳肉芽肿、风湿性多肌痛、皮炎和坏疽性脓皮病。

[0192] 此外,如前所述,本公开提供了允许从业者可靠且一致地预测受试者队列中的应答的技术。特别地,例如,在给定的受试者队列中,一些抗TNF疗法的应答率低于35%。所提供的技术允许预测受试者队列的应答率(即,特定受试者是否将对给定治疗产生应答)的准确度大于65%。在一些实施方案中,本文所述的方法和系统预测65%或更高的给定队列中为应答者(即,将对抗TNF疗法产生应答)的受试者。在一些实施方案中,本文所述的方法和系统预测70%或更高的给定队列中为应答者的受试者。在一些实施方案中,本文所述的方法和系统预测80%或更高的给定队列中为应答者的受试者。在一些实施方案中,本文所述

的方法和系统预测90%或更高的给定队列中为应答者的受试者。在一些实施方案中,本文所述的方法和系统预测100%的给定队列中为应答者的受试者。在一些实施方案中,本文所述的方法和系统预测65%或更高的给定队列中为无应答者(即,将不对抗TNF疗法产生应答)的受试者。在一些实施方案中,本文所述的方法和系统预测70%或更高的给定队列中为无应答者的受试者。在一些实施方案中,本文所述的方法和系统预测80%或更高的给定队列中为无应答者的受试者。在一些实施方案中,本文所述的方法和系统预测90%或更高的给定队列中为无应答者的受试者。在一些实施方案中,本文所述的方法和系统预测100%的给定队列中为无应答者的受试者。

F. 管道分析(Pipeline Analysis)

[0193] 本文还描述了一种用于基因组数据分析的管道,其具有允许临床软件所需的可追溯性和V&V测试/文档记录,但灵活以允许随时间升级组件的模块化架构。DNA和RNA的下一代测序(NGS)(RNAseq)是快速发展的领域,随着生物信息学算法和组件的定期改进而变得可用。这与传统上锁定并验证的标准IVD软件相冲突。对软件的任何修改都需要进行广泛的再确认和再验证。在某些实施方案中,本文描述的管道允许使用现有技术的生物信息学模块,例如RSEM、STAR,它们是用R编写的算法,其将与一个或多个专有算法一起组装在管道中。每个模块都“封装”在包装器中。管道的每个组件都可以独立验证。将对整个管道进行验证,并使用湿实验室(wet-lab)生成的数据对系统进行验证。如果将来要升级管道的任何组件,例如RSEM、STAR或一个或多个专有算法,则可以通过验证升级的组件、使用升级的组件验证系统并且重新分析先前生成的湿实验室数据来完成简单的V&V。

[0194] 图2描绘了RNA-seq生物信息学管道的例示性实施方案,因为它可以用于产生用于专有分类算法的输入——在本实例中,产生R(特定疗法的应答者)或NR(疗法的无应答者)的分类结果的专有算法。可执行其他种类的分类。

[0195] 如图2的示例性实施方案中所示:

1. RNAseq生物信息学管道

a. 该应用程序读取fastq文件和患者标识符,并返回结构化文本报告,所述报告的特征为相同的患者标识符以及一系列属性值(格式可以是例如,XML、JSON或其他选项)。

b. 输入文件是fastq文本数据文件(平均约30M读段,每个文件的文件大小为约1.4G)。

c. 从fastq输入文件开始到最终报告的过程包括:

i. fastq短读段记录与参考基因组组装(人类基因组组装GRCh37或GRCh38)的比对,以及与参考(STAR)比对器的剪接转录本比对[Dobin等人,2013]

ii. 通过期望最大化(RSEM)软件包用RNA-Seq对转录本丰度定量(Li和Wewey, 2011)。

iii. 通过自定义方法进行任选归一化

2. PrismRA算法(例示性专有算法)

a. 一组二进制结果的建模w.r.t.前一步使用r包生成的输入数据矩阵。这个R包的特征为对其他包的依赖性,包括glmnet、beanlots、pROC和samr。

3. QC跟踪和触发器

a. 除了最终文本报告外,该应用程序还可以编写和读取中间QC报告,所述报告的

值触发软件的不同行为。

b. QC报告是利用FastQC包由fastq文件(其需要三个Picard BAM/SAM库)和利用RNA SeQC包由BAM文件生成的。

4. 数据可追溯性、安全性、隐私性、合规性

a. 结果完全可追溯到所有输入和过程。

b. 数据隐私性和安全性:例如,21CFR第11部分,HIPAA和ISO27001安全合规性;GDPR合规性。

c. 稳定性:

i. 软件不受自动软件包升级的影响。

ii. 软件是便携式的,例如,处于可以部署在内部或私有云上的容器中

[0196] 如图4所示,示出并描述了用于提供如本文所述的系统、方法和架构的网络环境400的实施方式。在简要概述中,现在参考图4,示出并描述了示例性云计算环境400的框图。云计算环境400可包括一个或多个资源提供者402a、402b、402c(统称402)。每个资源提供者402可包括计算资源。在一些实施方式中,计算资源可包括用于处理数据的任何硬件和/或软件。例如,计算资源可包括能够执行算法、计算机程序和/或计算机应用程序的硬件和/或软件。在一些实施方式中,示例性计算资源可包括具有存储和检索能力的应用服务器和/或数据库。每个资源提供者402可连接至云计算环境400中的任何其他资源提供者402。在一些实施方式中,资源提供者402可通过计算机网络408连接。每个资源提供者402可通过计算机网络408连接至一个或多个计算设备404a、404b、404c(统称为404)。

[0197] 云计算环境400可包括资源管理器406。资源管理器406可通过计算机网络408连接至资源提供者402和计算设备404。在一些实施方式中,资源管理器406可促进由一个或多个资源提供者402向一个或多个计算设备404提供计算资源。资源管理器406可从特定计算设备404接收对计算资源的请求。资源管理器406可识别能够提供计算设备404所请求的计算资源的一个或多个资源提供者402。资源管理器406可选择资源提供者402来提供计算资源。资源管理器406可促进资源提供者402与特定计算设备404之间的连接。在一些实施方式中,资源管理器406可在特定资源提供者402与特定计算设备404之间建立连接。在一些实施方式中,资源管理器406可使用所请求的计算资源将特定计算设备404重定向至特定资源提供者402。

[0198] 图5示出可以用于实施本公开中描述的技术的计算设备500和移动计算设备550的实例。计算设备500旨在表示各种形式的数字计算机,诸如膝上型计算机、台式机、工作站、个人数字助理、服务器、刀片式服务器、大型机和其他适当的计算机。移动计算设备550旨在表示各种形式的移动设备,诸如个人数字助理、蜂窝电话、智能电话以及其他类似计算设备。此处所示的组件、它们的连接和关系以及它们的功能仅是示例性的,而不是限制性的。

[0199] 计算设备500包括处理器502、存储器504、存储设备506、连接至存储器504和多个高速扩展端口510的高速接口508,以及连接至低速扩展端口514和存储设备506的低速接口512。处理器502、存储器504、存储设备506、高速接口508、高速扩展端口510和低速接口512中的每一个使用各种总线相互连接,并且可在适当时安装在公共主板上或以其他方式安装。处理器502可以处理用于在计算设备500内执行的指令,包括存储在存储器504中或存储设备506上的指令,以在外部输入/输出设备(诸如耦合到高速接口508的显示器516)上显示

GUI的图形信息。在其他实施方式中,可在适当时使用多个处理器和/或多条总线以及多个存储器和存储器类型。此外,可连接多个计算设备,每个设备提供必要操作的一部分(例如,作为服务器库(server bank)、刀片式服务器组或多处理器系统)。因此,如该术语在本文所使用的,在多个功能被描述为由“处理器”执行的情况下,这包括其中多个功能由任何数量的计算设备(一个或多个)的任何数量的处理器(一个或多个)执行的实施方案。此外,在功能被描述为由“处理器”执行的情况下,这包括其中功能由任何数量的计算设备(一个或多个)(例如,在分布式计算系统中)的任何数量的处理器(一个或多个)执行的实施方案。

[0200] 存储器504将信息存储在计算设备500内。在一些实施方式中,存储器504是一个或多个易失性存储器单元。在一些实施方式中,存储器504是一个或多个非易失性存储器单元。存储器504还可以是另一种形式的计算机可读介质,诸如磁盘或光盘。

[0201] 存储设备506能够为计算设备500提供大容量存储。在一些实施方式中,存储设备506可以是或包含计算机可读介质,诸如软盘设备、硬盘设备、光盘设备或带式设备、闪存或其他类似固态存储器设备,或包括存储区域网络或其他配置中的设备的设备阵列。指令可以存储在信息载体中。当指令由一个或多个处理设备(例如,处理器502)执行时,执行一种或多种方法,诸如上文所述的那些方法。指令还可以由一个或多个存储设备存储,诸如计算机或机器可读介质(例如,存储器504、存储设备506或处理器502上的存储器)。

[0202] 高速接口508管理计算设备500的带宽密集型操作,而低速接口512管理较低带宽密集型操作。这种功能分配只是一个实例。在一些实施方式中,高速接口508耦合到存储器504、显示器516(例如,通过图形处理器或加速器)和高速扩展端口510,所述高速扩展端口510可接受各种扩展卡(未示出)。在该实施方式中,低速接口512耦合到存储设备506和低速扩展端口514。可包括各种通信端口(例如,USB、Bluetooth[®]、以太网、无线以太网)的低速扩展端口514可例如通过网络适配器耦合到一个或多个输入/输出设备,诸如键盘、定点设备、扫描仪或网络设备,诸如交换机或路由器。

[0203] 计算设备500可以多种不同的形式实施,如图所示。例如,它可以实施为标准服务器520,或者在一组这样的服务器中多次实施。此外,它可在诸如膝上型计算机522的个人计算机中实施。它也可以实施为机架式服务器系统524的一部分。替代地,来自计算设备500的组件可与移动设备(未示出)中的其他组件组合,诸如移动计算设备550。此类设备中的每一个可包含计算设备500和移动计算设备550中的一个或多个,并且整个系统可由相互通信的多个计算设备组成。

[0204] 移动计算设备550包括处理器552、存储器564、诸如显示器554的输入/输出设备、通信接口566和收发器568以及其他组件。移动计算设备550还可设置有存储设备,诸如微驱动器或其他设备,以提供附加存储。处理器552、存储器564、显示器554、通信接口566和收发器568中的每一个使用各种总线相互连接,并且所述组件中的几个可安装在公共主板上或以其他适当方式安装。

[0205] 处理器552可以执行移动计算设备550内的指令,包括存储在存储器564中的指令。处理器552可实施为包括独立和多个模拟和数字处理器的芯片的芯片组。处理器552可提供例如移动计算设备550的其他组件的协调,诸如用户接口的控制、由移动计算设备550运行的应用以及由移动计算设备550进行的无线通信。

[0206] 处理器552可通过控制接口558和耦合到显示器554的显示器接口556与用户通信。

显示器554可以是例如TFT(薄膜晶体管液晶显示器)显示器或OLED(有机发光二极管)显示器,或其他适当的显示技术。显示接口556可包括用于驱动显示器554向用户呈现图形和其他信息的适当电路。控制接口558可从用户接收命令,并将其转换以提交给处理器552。此外,外部接口562可提供与处理器552的通信,以便使移动计算设备550能够与其他设备进行邻近区域通信。外部接口562可在一些实施方式中提供例如有线通信或在其他实施方式中提供无线通信,并且还可使用多个接口。

[0207] 存储器564在移动计算设备550内存储信息。存储器564可以实施为一个或多个计算机可读介质、一个或多个易失性存储器单元或一个或多个非易失性存储器单元中的一个或多个。还可提供扩展存储器574,并通过扩展接口572连接到移动计算设备550,所述扩展接口572可包括例如SIMM(单列直插存储器模块(Single In Line Memory Module))卡接口。扩展存储器574可为移动计算设备550提供额外的存储空间,或者还可存储移动计算设备550的应用程序或其他信息。具体地,扩展存储器574可包括执行或补充上述过程的指令,并且还可包括安全信息。因此,例如,扩展存储器574可作为移动计算设备550的安全模块提供,并且可使用允许安全使用移动计算设备550的指令进行编程。此外,可通过SIMM卡提供安全应用程序,以及附加信息,诸如以不可破解的方式将识别信息放置在SIMM卡上。

[0208] 存储器可包括例如,闪存和/或NVRAM存储器(非易失性随机存取存储器),如下所述。在一些实施方式中,指令存储在信息载体中。当指令由一个或多个处理设备(例如,处理器552)执行时,执行一种或多种方法,诸如上文所述的那些。指令还可以由一个或多个存储设备存储,诸如一个或多个计算机或机器可读介质(例如,存储器564、扩展存储器574或处理器552上的存储器)。在一些实施方式中,可以例如通过收发器568或外部接口562在传播信号中接收指令。

[0209] 移动计算设备550可通过通信接口566进行无线通信,必要时,所述通信接口566可包括数字信号处理电路。通信接口566可提供各种模式或协议下的通信,诸如GSM语音呼叫(全球移动通信系统)、SMS(短消息服务)、EMS(增强型消息服务)或MMS消息(多媒体消息服务)、CDMA(码分多址)、TDMA(时分多址)、PDC(个人数字蜂窝)、WCDMA(宽带码分多址)、CDMA2000或GPRS(通用分组无线业务)等。例如,可使用射频通过收发器568进行这种通信。此外,可进行短距离通信,诸如使用Bluetooth[®]、Wi-Fi[™]或其他此类收发器(未示出)。此外,GPS(全球定位系统)接收器模块570可向移动计算设备550提供另外的导航和位置相关无线数据,移动计算设备550上运行的应用程序可在适当时使用这些数据。

[0210] 移动计算设备550还可使用音频编解码器560进行音频通信,所述音频编解码器560可从用户接收语音信息并将其转换为可用的数字信息。音频编解码器560同样可以诸如通过扬声器(例如,在移动计算设备550的手持设备(handset)中)为用户生成可听声音。这种声音可包括来自语音电话呼叫的声音,可包括记录的声音(例如,语音消息、音乐文件等),并且还可包括由在移动计算设备550上操作的应用程序生成的声音。

[0211] 移动计算设备550可以多种不同的形式实施,如图所示。例如,它可实施为蜂窝电话580。它还可实施为智能电话582、个人数字助理或其他类似移动设备的一部分。

[0212] 这里描述的系统和技术各种实施方式可以在数字电子电路系统、集成电路系统、专门设计的ASIC(专用集成电路)、计算机硬件、固件、软件和/或其组合中实现。这些不同的实施方式可以包括在可编程系统上可执行和/或可解释的一个或多个计算机程序中的

实施方式,所述可编程系统包括至少一个可编程处理器、至少一个输入设备和至少一个输出设备,所述可编程处理器可以是专用或通用的,耦合以从存储系统接收数据和指令,并向所述存储系统传输数据和指令。

[0213] 这些计算机程序(也称为程序、软件、软件应用程序或代码)包括用于可编程处理器的机器指令,并且可以高阶程序和/或面向对象编程语言和/或汇编/机器语言实施。如本文所用,术语机器可读介质和计算机可读介质是指用于向可编程处理器提供机器指令和/或数据的任何计算机程序产品、设备和/或装置(例如,磁盘、光盘、存储器、可编程逻辑器件(PLD)),包括机器可读介质,其接收作为机器可读信号的机器指令。术语机器可读信号是指用于向可编程处理器提供机器指令和/或数据的任何信号。

[0214] 为了提供与用户的交互,这里描述的系统和技术可以在具有用于向用户显示信息的显示设备(例如,CRT(阴极射线管)或LCD(液晶显示器)监视器)以及用户可以用来向计算机提供输入的键盘和定点设备(例如,鼠标或轨迹球)的计算机上实施。其他种类的设备也可以用于提供与用户的交互;例如,提供给用户的反馈可以是任何形式的感官反馈(例如,视觉反馈、听觉反馈或触觉反馈);并且可以任何形式接收来自用户的输入,包括声音、语音或触觉输入。

[0215] 这里描述的系统和技术可以在计算系统中实施,所述计算系统包括后端组件(例如,作为数据服务器)或包括中间件组件(例如,应用服务器),或包括前端组件(例如,具有图形用户界面或Web浏览器的客户端计算机,用户可以通过其与此处描述的系统和技术的方式进行交互),或此类后端、中间件或前端组件的任何组合。系统的组件可以通过任何形式的数字数据通信或数字数据通信的任何媒介(例如,通信网络)相互连接。通信网络的实例包括局域网(LAN)、广域网(WAN)和互联网。

[0216] 计算系统可以包括客户端和服务端。客户端和服务端通常彼此远离,并且通常通过通信网络进行交互。客户端和服务端的关系是由于在各自的计算机上运行的计算机程序而产生的,并且彼此之间具有客户端-服务端关系。

[0217] 在一些实施方式中,本文描述的模块可以分离、组合或并入单个或组合模块中。图中所示的模块不旨在将本文所述的系统限制为其中所示的软件架构。

[0218] 本文描述的不同实施方式的元件可组合以形成上文未具体阐述的其他实施方式。可将元件从本文所述的过程、计算机程序、数据库等中排除,而不对其操作产生不利影响。此外,图中描绘的逻辑流不需要所示的特定顺序或相继次序来实现期望的结果。各种独立元件可组合成一个或多个单独元件以执行本文所述的功能。鉴于这里描述的系统和方法的结构、功能和装置,在一些实施方式中。

[0219] 可以设想,所要求保护的发明的系统、架构、设备、方法和过程包括使用来自本文所描述的实施方案的信息开发的变型和改编。如本说明书所设想的,可执行本文所述的系统、架构、设备、方法和过程的改编和/或修改。

[0220] 在整个描述中,其中制品、设备、系统和架构被描述为具有、包括或包含特定组件,或其中过程和方法被描述为具有、包括或包含特定步骤,预期另外还存在基本上由所述组件组成或由所述组件组成的本发明的制品、设备、系统和架构,并且存在基本上由所述处理步骤组成或由所述处理步骤组成的根据本发明的过程和方法。

[0221] 应当理解,只要本发明保持可操作,步骤顺序或执行特定动作的顺序就无关紧要。

此外,可同时进行两个或更多个步骤或动作。

[0222] 本文例如在背景技术部分中对任何出版物的提及并不表示承认该出版物就本文中提出的任何权利要求而言是现有技术。背景技术部分是为了清楚起见而提供的,并不意味着就任何权利要求而言对现有技术的描述。

[0223] 标题是为了方便读者而提供的——标题的存在和/或放置并不旨在限制本文所述主题的范围。

G. 示例性实施方案

[0224] 以下编号的实施方案虽然是非限制性的,但是也是本公开的某些方面的示例性实施方案:

实施方案1. 一种治疗患有自身免疫性病症的受试者的方法,所述方法包括以下步骤:

向已通过分类器确定为有应答的受试者施用抗TNF疗法,所述分类器被建立用于区分已接受所述抗TNF疗法的队列中有应答和无应答的先前受试者;

其中所述分类器通过评估以下各项进行开发:

表达水平与临床应答性或无应答性显著相关(例如,以线性和/或非线性方式)的一种或多种基因;

以下各项中的至少一个:

所述一种或多种基因的表达序列中一个或多个单核苷酸多态性(SNP)的存在;或

所述有应答和无应答的先前受试者的至少一种临床特征;并且其中所述分类器通过不是已接受所述抗TNF疗法的所述队列的独立队列进行验证。

实施方案2. 根据实施方案1所述的方法,其中所述分类器在至少170名受试者的群体中以至少90%的准确度预测受试者的应答性。

实施方案3. 根据实施方案1或2所述的方法,其中当映射到人类相互作用组图谱上时,所述一种或多种基因通过其拓扑性质进行表征。

实施方案4. 根据实施方案1-3中任一项所述的方法,其中所述SNP参考人类基因组进行鉴定。

实施方案5. 根据实施方案1.4中任一项所述的方法,其中所述分类器通过评估以下各项进行开发:

表达水平与临床应答性或无应答性显著相关(例如,以线性和/或非线性方式)的所述一种或多种基因;

所述一个或多个SNP的存在;以及

所述至少一种临床特征。

实施方案6. 根据实施方案1-5中任一项所述的方法,其中所述一种或多种基因选自:

ARPC1A	LOC100506083
ATAT1	MORN2
ATRAID	NGF
CD27	RHOBTB1
CHMP7	RNGTT
COMMD5	SETD9
CORO2B	SNX8
CYSLTR2	SPINT2
DNAJC7	SPON2
GOLGA1	SSNA1
HPCA	STOML2
IMPDH2	SUOX
KAT8	TMEM258
KLHDC3	UBL7-AS1

实施方案7. 根据实施方案1-6中任一项所述的方法, 其中所述至少一种临床特征选自: 体重指数 (BMI)、性别、年龄、种族、先前疗法治疗、疾病持续时间、C反应蛋白水平、抗环瓜氨酸肽的存在、类风湿因子的存在、患者整体评估、治疗应答率 (例如, ACR20、ACR50、ACR70) 以及其组合。

实施方案8. 根据实施方案1-6中任一项所述的方法, 其中所述分类器以至少90%的准确度鉴定群体中无应答的至少40%的受试者, 其中所述群体包括至少170名受试者。

实施方案9. 根据实施方案1-8中任一项所述的方法, 其中所述抗TNF疗法包括施用英夫利昔单抗、阿达木单抗、依那西普、赛妥珠单抗、goliluma或其生物仿制药。

实施方案10. 根据实施方案1-9中任一项所述的方法, 其中所述疾病、病症或病况选自类风湿性关节炎、银屑病性关节炎、强直性脊柱炎、克罗恩病、溃疡性结肠炎、慢性银屑病、化脓性汗腺炎、多发性硬化症和幼年特发性关节炎。

实施方案11. 根据实施方案1-10中任一项所述的方法, 其中使用源自所述有应答和无应答的先前受试者的微阵列分析建立所述分类器。

实施方案12. 根据实施方案1-11中任一项所述的方法, 其中使用源自所述独立队列的RNAseq数据验证所述分类器。

实施方案13. 根据实施方案1-12中任一项所述的方法, 其中所述SNP选自表5。

实施方案14. 一种向患有自身免疫性疾病的受试者施用疗法的方法, 所述方法包括

向已通过分类器分类为有应答或无应答的受试者施用所述疗法, 所述分类器被建

立用于区分已接受所述抗TNF疗法的队列中有应答和无应答的先前受试者；

其中，如果所述受试者被所述分类器分类为有应答，则施用抗TNF疗法；并且如果所述受试者被所述分类器分类为无应答，则施用抗TNF疗法的替代方案；并且

所述分类器通过评估以下各项进行开发：

表达水平与临床应答性或无应答性显著相关（例如，以线性和/或非线性方式）的一种或多种基因；

以下各项中的至少一个：

所述一种或多种基因的表达序列中一个或多个单核苷酸多态性（SNP）的存在；或

所述有应答和无应答的先前受试者的至少一种临床特征；并且所述分类器通过不是已接受所述抗TNF疗法的所述队列的独立队列进行验证。

实施方案15. 根据实施方案14所述的方法，其中所述分类器在至少170名受试者的群体中以至少90%的准确度预测受试者的应答性。

实施方案16. 根据实施方案15所述的方法，其中所述分类器通过评估以下各项进行开发：

表达水平与临床应答性或无应答性显著相关（例如，以线性和/或非线性方式）的所述一种或多种基因；

所述一个或多个SNP的存在；以及

所述至少一种临床特征。

实施方案17. 根据实施方案14-16中任一项所述的方法，其中所述至少一种临床特征选自：体重指数（BMI）、性别、年龄、种族、先前疗法治疗、疾病持续时间、C反应蛋白水平、抗环瓜氨酸肽的存在、类风湿因子的存在、患者整体评估、治疗应答率（例如，ACR20、ACR50、ACR70）以及其组合。

实施方案18. 根据实施方案14-17中任一项所述的方法，其中所述抗TNF疗法包括施用英夫利昔单抗、阿达木单抗、依那西普、赛妥珠单抗、goliluma或其生物仿制药。

实施方案19. 根据实施方案14-18中任一项所述的方法，其中抗TNF疗法的替代方案选自利妥昔单抗、沙利姆单抗、枸橼酸托法替尼、来氟米特、维多珠单抗、托珠单抗、阿那白滞素和阿巴西普。

实施方案20. 根据实施方案19所述的方法，其中所述疾病、病症或病况选自类风湿性关节炎、银屑病性关节炎、强直性脊柱炎、克罗恩病、溃疡性结肠炎、慢性银屑病、化脓性汗腺炎、多发性硬化症和幼年特发性关节炎。

实施方案21. 一种开发分类器以用于针对一个或多个治疗属性对受试者分层的方法，其包括以下步骤：

分析在针对至少一个所述治疗属性代表至少两个不同类别的受试者中表达的RNA的序列数据；

由所述序列数据评估一个或多个单核苷酸多态性（SNP）的存在；确定所述一个或多个SNP的存在与所述至少一个治疗属性相关；以及

将所述一个或多个SNP包括在所述分类器中。

实施方案22. 根据实施方案21所述的方法，其中所述一个或多个治疗属性选自由以下各项组成的组：发展特定疾病、病症或病况的风险；特定疾病、病症或病况的特定结果

的可能性;对特定疗法产生应答的可能性。

实施方案23.根据实施方案21或22所述的方法,其中所述一个或多个治疗属性是对特定疗法产生应答的可能性。

实施方案24.根据实施方案21-23中任一项所述的方法,其中所述特定疗法为抗TNF疗法。

实施方案25.开发分类器,以用于通过分析针对至少一个治疗属性代表至少两个不同类别的受试者中表达的RNA的序列数据来针对所述一个或多个治疗属性对受试者分层的方法的改进,其包括:

由所述序列数据评估一个或多个单核苷酸多态性(SNP)的存在;以及
确定所述一个或多个SNP的存在与所述至少一个治疗属性相关;以及
将所述一个或多个SNP包括在所述分类器中。

实施方案26.根据实施方案25所述的方法,其中所述一个或多个治疗属性选自以下各项组成的组:发展特定疾病、病症或病况的风险;特定疾病、病症或病况的特定结果的可能性;对特定疗法产生应答的可能性。

实施方案27.根据实施方案25或26所述的方法,其中所述一个或多个治疗属性是对特定疗法产生应答的可能性。

实施方案28.根据实施方案25-27中任一项所述的方法,其中所述特定疗法为抗TNF疗法。

实施方案29.根据实施方案25-28中任一项所述的方法,其中评估所述一个或多个SNP的存在包括将RNA的所述序列数据与参考人类基因组进行比较。

实施方案30.一种治疗患有自身免疫性疾病的受试者的方法,所述方法包括以下步骤:

(a)通过计算设备的处理器接收与所述受试者的一种或多种基因中的每一种的表达水平相对应的数据,其中所述基因选自表1;

(b)使用在步骤(a)中接收的所述数据由所述处理器自动确定所述受试者为对抗TNF疗法有应答或无应答的分类;以及,任选地,

(c)如果所述受试者被分类为对所述抗TNF疗法无应答,则向所述受试者施用抗TNF疗法的替代方案以治疗所述自身免疫性疾病;或者,如果所述受试者被分类为对所述抗TNF疗法有应答,则施用所述抗TNF疗法。

实施方案31.根据实施方案30所述的方法,其中步骤(a)中的所述处理器进一步接收与以下各项中的至少一个相对应的数据:一种或多种临床特征;或
一个或多个单核苷酸多态性(SNP)。

实施方案32.根据实施方案30或31所述的方法,其中所述一种或多种临床特征选自:体重指数(BMI)、性别、年龄、种族、先前疗法治疗、疾病持续时间、C反应蛋白(CRP)水平、抗环瓜氨酸肽的存在、类风湿因子的存在、患者整体评估、治疗应答率(例如,ACR20、ACR50、ACR70)以及其组合。

实施方案33.根据实施方案30-32中任一项所述的方法,其中所述自身免疫性疾病选自类风湿性关节炎、银屑病性关节炎、强直性脊柱炎、克罗恩病、溃疡性结肠炎、慢性银屑病、化脓性汗腺炎和幼年特发性关节炎。

实施方案34.根据实施方案30-33中任一项所述的方法,其中所述自身免疫性疾病为类风湿性关节炎或溃疡性结肠炎。

实施方案35.根据实施方案30-34中任一项所述的方法,其中抗TNF疗法的替代方案选自利妥昔单抗、沙利姆单抗、枸橼酸托法替尼、来氟米特、维多珠单抗、托珠单抗、阿那白滞素和阿巴西普。

实施方案36.根据实施方案30-34中任一项所述的方法,其中所述抗TNF疗法选自英夫利昔单抗、阿达木单抗、依那西普、赛妥珠单抗、goliluma或其生物仿制药。

实施方案37.根据实施方案30-36中任一项所述的方法,其中步骤(b)包括使用机器学习模型自动确定所述分类。

实施方案38.根据实施方案37所述的方法,其中所述机器学习模型是随机森林模型。

实施方案39.根据实施方案30-38中任一项所述的方法,其中步骤(b)包括在不使用临床协变量分类器(CC)的情况下自动确定所述分类。

实施方案40.根据实施方案30-38中任一项所述的方法,其中步骤(b)包括在不使用组合基因组-临床分类器(GCC)的情况下自动确定所述分类。

实施方案41.根据实施方案30-38中任一项所述的方法,其中步骤(b)包括在不使用临床协变量分类器(CC)和不使用组合基因组-临床分类器(GCC)的情况下自动确定所述分类。

实施方案42.一种方法,其包括以下步骤:

(a)通过计算设备的处理器接收与患有自身免疫性疾病(例如,类风湿性关节炎)的受试者的一种或多种基因中的每一种的表达水平相对应的数据,所述一种或多种基因包括选自由COR02B、STOML2、CHMP7、SPINT2、CYSLTR2、SSNA1、IMPDH2、GOLGA1、SUOX、ARPC1A和KLHDC3组成的组的至少一个成员(例如,一个、两个、三个、四个、五个、六个、七个、八个、九个、十个或全部十一个成员);

(b)使用在步骤(a)中接收的所述数据由所述处理器自动确定所述受试者为对抗TNF疗法无应答的分类;以及,任选地,

(c)向所述受试者开处方和/或施用第二疗法(例如,所述第一疗法的替代方案,例如,抗TNF疗法的替代方案)以治疗所述疾病,从而避免向所述受试者开处方和/或施用所述第一疗法。实施方案43.一种方法,其包括以下步骤:

(a)通过计算设备的处理器接收与患有疾病(例如,自身免疫性疾病,例如,类风湿性关节炎)的受试者的一种或多种基因中的每一种的表达水平相对应的数据,所述一种或多种基因包括选自由COR02B、STOML2、CHMP7、SPINT2、CYSLTR2、SSNA1、IMPDH2、GOLGA1、SUOX、ARPC1A和KLHDC3组成的组的至少一个成员(例如,一个、两个、三个、四个、五个、六个、七个、八个、九个、十个或全部十一个成员);

(b)使用在步骤(a)中接收的所述数据由所述处理器自动确定所述受试者为对第一疗法(例如,抗TNF疗法)有应答的分类;以及,任选地,

(c)为所述受试者开处方和/或施用所述第一疗法以治疗所述疾病。

实施方案44.根据实施方案42或43所述的方法,其中步骤(b)包括使用机器学习模型(例如,随机森林模型)自动确定所述分类。

实施方案45. 根据实施方案43至44中任一项所述的方法, 其中步骤(b) 包括在不使用临床协变量分类器(CC) 和不使用组合基因组-临床分类器(GCC) 的情况下自动确定所述分类。

实施方案46. 一种用于分析受试者的基因组数据(例如, 下一代RNA-seq数据)(例如, 并用于基于所述基因组数据确定所述受试者的分类)的管道(例如, 计算机架构管道), 所述管道包括多个模块, 每个模块能够在所述模块升级后独立验证。

实施方案47. 根据实施方案46所述的管道, 其中所述多个模块包括一个或多个机器学习模型。

实施方案48. 根据实施方案46或47所述的管道, 其中所述多个模块包括一个或多个已知的生物信息学模块(例如, RSEM和/或STAR) 和一个或多个专有分类模块。

实施方案49. 一种在对患有自身免疫性疾病的受试者进行抗TNF疗法的任何施用之前将所述受试者分类为对所述抗TNF疗法可能有应答或可能无应答的方法, 所述方法包括以下步骤:

(a) 通过计算设备的处理器接收与从选自由以下各项组成的组的一种或多种基因中的每一种的所述受试者的表达水平相对应的数据:

ARPC1A	LOC100506083
ATAT1	MORN2
ATRAID	NGF
CD27	RHOBTB1
CHMP7	RNGTT
COMMD5	SETD9
CORO2B	SNX8
CYSLTR2	SPINT2
DNAJC7	SPON2
GOLGA1	SSNA1
HPCA	STOML2
IMPDH2	SUOX
KAT8	TMEM258
KLHDC3	UBL7-AS1

以及

(b) 使用在步骤(a) 中接收的所述数据由所述处理器自动确定所述受试者为对所述抗TNF疗法有应答或无应答的分类。

实施方案50. 根据实施方案49所述的方法, 其还包括:

(c) 如果所述受试者在步骤 (b) 中被分类为对抗TNF疗法无应答,则为所述受试者开处方和/或施用所述抗TNF疗法的替代方案以治疗所述自身免疫性疾病;或者,如果所述受试者在步骤 (b) 中被分类为对所述抗TNF疗法有应答,则向所述受试者施用所述抗TNF疗法。

实施方案51. 根据实施方案49或50所述的方法,其中步骤 (a) 中的所述处理器进一步接收与如下 (i) 和 (ii) 中至少一个相对应的数据:

(i) 所述受试者的一种或多种临床特征;

(ii) 所述受试者的一个或多个单核苷酸多态性 (SNP)。

实施方案52. 根据实施方案49至51中任一项所述的方法,其中步骤 (a) 中的所述处理器接收与所述受试者的一种或多种临床特征相对应的数据,所述临床特征选自以下各项组成的组: 体重指数 (BMI)、性别、年龄、种族、先前疗法治疗、疾病持续时间、C反应蛋白 (CRP) 水平、抗环瓜氨酸肽的存在、类风湿因子的存在、患者整体评估以及治疗应答率 (例如, ACR20、ACR50、ACR70)。

实施方案53. 根据实施方案49至52中任一项所述的方法,其中步骤 (a) 中的所述处理器接收与表5中列出的一个或多个SNP相对应的数据。

实施方案54. 根据实施方案49至53中任一项所述的方法,其中所述自身免疫性疾病是选自自由类风湿性关节炎、银屑病性关节炎、强直性脊柱炎、克罗恩病、溃疡性结肠炎、慢性银屑病、化脓性汗腺炎和幼年特发性关节炎组成的组的成员。

实施方案55. 根据实施方案49至54中任一项所述的方法,其中所述自身免疫性疾病为类风湿性关节炎或溃疡性结肠炎。

实施方案56. 根据实施方案49至55中任一项所述的方法,其中抗TNF疗法的替代方案包括选自自由利妥昔单抗、沙利姆单抗、枸橼酸托法替尼、来氟米特、维多珠单抗、托珠单抗、阿那白滞素和阿巴西普组成的组的至少一个成员。

实施方案57. 根据实施方案49至55中任一项所述的方法,其中所述抗TNF疗法包括选自自由英夫利昔单抗、阿达木单抗、依那西普、赛妥珠单抗、goliluma以及前述中任一个的生物仿制药组成的组的至少一个成员。

实施方案58. 根据实施方案44至52中任一项所述的方法,其中抗TNF疗法的替代方案不包括由英夫利昔单抗、阿达木单抗、依那西普、赛妥珠单抗、goliluma以及前述中任一个的任何生物仿制药组成的组的任何成员。

实施方案59. 根据实施方案49至58中任一项所述的方法,其中步骤 (b) 包括使用机器学习模型自动确定所述分类。

实施方案60. 根据实施方案59所述的方法,其中所述机器学习模型是随机森林模型。

实施方案61. 根据实施方案49至60中任一项所述的方法,其中步骤 (b) 包括在不使用临床协变量分类器 (CC) 的情况下自动确定所述分类。

实施方案62. 根据实施方案49至60中任一项所述的方法,其中步骤 (b) 包括在不使用组合基因组-临床分类器 (GCC) 的情况下自动确定所述分类。

实施方案63. 根据实施方案49至60中任一项所述的方法,其中步骤 (b) 包括在不使用临床协变量分类器 (CC) 和不使用组合基因组-临床分类器 (GCC) 的情况下自动确定所述

分类。

实施方案64. 一种在对患有自身免疫性疾病的受试者进行抗TNF疗法的任何施用之前将所述受试者分类为对所述抗TNF疗法可能有应答或可能无应答的方法,所述方法包括以下步骤:

(a) 通过计算设备的处理器接收与患有所述自身免疫性疾病的受试者的一种或多种基因中的每一种的表达水平相对应的数据,所述一种或多种基因包括选自由COR02B、STOML2、CHMP7、SPINT2、CYSLTR2、SSNA1、IMPDH2、GOLGA1、SUOX、ARPC1A和KLHDC3组成的组的至少一个成员;

(b) 使用在步骤(a)中接收的所述数据由所述处理器自动确定所述受试者为对所述抗TNF疗法可能有应答或可能无应答的分类。

实施方案65. 根据实施方案64所述的方法,其还包括:

(c) 如果所述受试者在步骤(b)中被分类为对所述抗TNF疗法无应答,则为所述受试者开处方和/或施用所述抗TNF疗法的替代方案以治疗所述自身免疫性疾病;或者,如果所述受试者在步骤(b)中被分类为对所述抗TNF疗法有应答,则向所述受试者施用所述抗TNF疗法。

实施方案66. 根据实施方案64或65所述的方法,其中所述自身免疫性疾病为类风湿性关节炎。

实施方案67. 根据实施方案64至66中任一项所述的方法,其中步骤(a)包括接收与选自由COR02B、STOML2、CHMP7、SPINT2、CYSLTR2、SSNA1、IMPDH2、GOLGA1、SUOX、ARPC1A和KLHDC3组成的组的至少两种基因中的每一种的表达水平相对应的数据。

实施方案68. 根据实施方案64至66中任一项所述的方法,其中步骤(a)包括接收与选自由COR02B、STOML2、CHMP7、SPINT2、CYSLTR2、SSNA1、IMPDH2、GOLGA1、SUOX、ARPC1A和KLHDC3组成的组的至少三种基因中的每一种的表达水平相对应的数据。

实施方案69. 根据实施方案64至66中任一项所述的方法,其中步骤(a)包括接收与选自由COR02B、STOML2、CHMP7、SPINT2、CYSLTR2、SSNA1、IMPDH2、GOLGA1、SUOX、ARPC1A和KLHDC3组成的组的至少五种基因中的每一种的表达水平相对应的数据。

实施方案70. 根据实施方案64至69中任一项所述的方法,其中步骤(a)中的所述处理器进一步接收与如下(i)和(ii)中至少一个相对应的数据:

(i) 所述受试者的一种或多种临床特征;

(ii) 所述受试者的一个或多个单核苷酸多态性(SNP)。

实施方案71. 根据实施方案64至70中任一项所述的方法,其中步骤(a)中的所述处理器接收与所述受试者的一种或多种临床特征相对应的数据,所述临床特征选自由以下各项组成的组:体重指数(BMI)、性别、年龄、种族、先前疗法治疗、疾病持续时间、C反应蛋白(CRP)水平、抗环瓜氨酸肽的存在、类风湿因子的存在、患者整体评估以及治疗应答率(例如,ACR20、ACR50、ACR70)。

实施方案72. 根据实施方案64至71中任一项所述的方法,其中步骤(a)中的所述处理器接收与表5中列出的一个或多个SNP相对应的数据。

实施方案73. 一种在对患有自身免疫性疾病的受试者进行抗TNF疗法的任何施用之前将所述受试者分类为对所述抗TNF疗法可能有应答或可能无应答的系统,所述系统包

括：

处理器；以及

其上具有指令的存储器，所述指令在由所述处理器执行时，使所述处理器：

(a) 接收数据集，所述数据集包括选自由以下各项组成的组中的一种或多种基因中的每一种的所述受试者的表达水平：

ARPCIA	LOC100506083
ATAT1	MORN2
ATRAID	NGF
CD27	RHOBTB1
CHMP7	RNGTT
COMMD5	SETD9
CORO2B	SNX8
CYSLTR2	SPINT2
DNAJC7	SPON2
GOLGA1	SSNA1
HPCA	STOML2
IMPDH2	SUOX
KAT8	TMEM258
KLHDC3	UBL7-AS1

以及

(b) 使用所述数据集自动确定所述受试者为对所述抗TNF疗法有应答或无应答的分类。

实施方案74. 根据实施方案73所述的系统，其中所述数据集还包括与如下(i)和(ii)中至少一个相对应的数据：

(i) 所述受试者的一种或多种临床特征；

(ii) 所述受试者的一个或多个单核苷酸多态性(SNP)。

实施方案75. 根据实施方案74所述的系统，其中所述数据集包括与所述受试者的一种或多种临床特征相对应的数据，所述临床特征选自由以下各项组成的组：体重指数(BMI)、性别、年龄、种族、先前疗法治疗、疾病持续时间、C反应蛋白(CRP)水平、抗环瓜氨酸肽的存在、类风湿因子的存在、患者整体评估以及治疗应答率(例如，ACR20、ACR50、ACR70)。

实施方案76. 根据实施方案74或75所述的系统，其中所述数据集包括表5中列出的一个或多个SNP。

实施方案77. 根据实施方案74至76中任一项所述的系统，其中

所述自身免疫性疾病是选自由类风湿性关节炎、银屑病性关节炎、强直性脊柱炎、克罗恩病、溃疡性结肠炎、慢性银屑病、化脓性汗腺炎和幼年特发性关节炎组成的组的成员。

实施方案78. 根据实施方案74至76中任一项所述的系统，其中所述自身免疫性疾病为类风湿性关节炎或溃疡性结肠炎。

实施方案79. 根据实施方案74至76中任一项所述的系统，其中所述自身免疫性疾

病为类风湿性关节炎。

范例

实施例1

概述

[0225] 诸如RA的自身免疫性疾病共同影响数百万患者,其治疗是整体医疗支出的重要组成部分并且因此需要关注以改善患者结果。

[0226] 自身免疫性疾病可以分为两组:器官特异性和系统性自身免疫。包括类风湿性关节炎(RA)在内的类风湿性疾病属于系统性自身免疫性疾病,其主要表现在滑膜关节并最终导致肌腱、软骨和骨的不可逆破坏。虽然目前尚无治愈RA的方法,但主要通过用于中和这种细胞因子的促炎性信号传导的抗TNF剂的开发在管理这些患者的治疗方面取得了重大改善。此类疗法(Humira、Enbrel、Remicade、Simponi、Cymzia)显著改善了一些RA患者的治疗结果。

[0227] 大约三分之一的RA患者表现出对抗TNF疗法的临床应答,有时达到缓解。这些所谓的“应答者”患者的疾病进展可能是由不适当的TNF驱动的促炎性应答引起的。

[0228] 治疗RA最广泛使用的生物疗法是抗TNF。这类药物的引入通过使一些患者实现LDA或缓解而改变了RA的治疗。然而,达到这些治疗目标的患者比例很低:在多个临床研究中为大约34%。对于无法对抗TNF产生应答的患者,存在可用的替代性获批疗法,诸如抗CD20、共刺激阻断、JAK和抗IL6疗法。然而,患者只有在首先循环使用不同的抗TNF药物(这可能需要一年多的时间)后才转为使用这种替代疗法,而症状持续存在并且疾病进一步发展,使其更难达到治疗目标。

[0229] 靶向治疗、使患者尽快达到LDA或缓解以阻止或最小化损伤进展的概念已被广泛接受,但在许多患者中往往难以实现。基于风湿病专家和患者的临床评估,使用疾病严重程度量表对RA进行评估。轻度、中度或重度疾病活动在RA的情况下使用以下确定:肿胀和压痛关节数量的评估;患者和医生报告的疼痛、功能和整体健康评估;以及血清生物标志物水平,当组合起来时,其产生疾病活动度评分(DAS)。中度至重度疾病活动表明患者的症状不再得到有效管理。使用DAS的变化监测对疗法的应答,并治疗以达到与LDA或缓解水平相对应的国际通用的DAS28评分。

[0230] 由样本处理、文库制备和相对于滑膜RA来说血液中较低的生物信号(如大多数复杂疾病一样)导致的交叉队列不可重复性是具有共同症状终点的不同疾病的集合并且因此需要个性化的医疗方法,以确保患者获得针对其个体疾病生物学最佳的靶向疗法。当对任何给定疗法的应答率低,每种疗法都有相关的严重副作用,并且可以为患者提供替代有效疗法时,复杂疾病的个性化健康护理尤其有意义。所有这些因素都适用于RA。通过以个体生物学为目标的补充疗法,医生、患者和付款人将最有效地实现治疗目标。此外,应根据其独特的疾病生物学,开发专门针对现有疗法的无应答者的新疗法。当考虑到与对当前护理标准疗法(standard-of-care therapy)无应答的患者相关的高额药物、一般医疗和社会成本时,这一点甚至更为重要。

[0231] 在开始疗法之前预测抗TNF无应答的测试将使得患者能够比当前护理标准更快地转向替代疗法,从而使更多患者达到LDA或缓解的治疗目标。这种预测应答测试及其对患者护理的影响将为患者提供实质性的临床益处,并为医疗系统节约成本。

[0232] 在分子水平上将给定患者的活动性疾病机制与治疗剂的作用机制相匹配将显著提高应答率,从而使更多患者能够达到低疾病活动度(LDA)或缓解的治疗目标。与在考虑替代疗法之前等待疾病症状明显恶化的常规做法相比,实现治疗目标在早期阶段对RA患者具有显著的临床益处。

[0233] 图6A是用于所选特征(n=75)、模型训练(n=185)和模型验证(n=46)的队列的流程图。图6B是描述图6A的队列1中的特征选择过程的过程。POC数据集是指图6A的队列1数据集。80%的数据用于使用U检验来训练分类器,以鉴定最具区分性的基因,并且80%的数据使用10倍交叉验证分成10组,以提供最终确定的特征集;最终确定的特征集用于在其余20%的数据上进行模型训练和评估。该过程重复至少100次。图6C是应用于队列1的特征选择方法的流程图。

结果

[0234] 鉴定预测对抗TNF疗法的应答的生物标志物。在本研究中,在抗TNF疗法的应答者与无应答者之间进行区分的预测性生物标志物选自Bienkowska等人公布的可公开获得的数据集(队列1)并在从CORRONA购买的231名类风湿性关节炎患者的队列(队列2)中进一步评估,其中185个样本用于模型训练(80%),并且46个(20%)用于验证。队列描述参见材料和方法,并且特征选择、模型训练和模型验证所用样本的流程图参见图6A。

[0235] 队列1样本中用于特征选择的方法描绘于图6C中。在数据集中的21,818个总基因转录本中,38个在使用随机森林的100轮20%保留交叉验证中的30轮中被选择为区分性(图7A)。通过将100轮交叉验证中所有保留验证集样本的得分聚合,对从中选择预测基因的模型进行评估。在这里,与聚合随机森林模型输出的无应答者得分相比,发现应答者具有显著更高的得分($p < 0.0001$) (图7B)。聚合得分得出曲线下面积(AUC)为60% (图7C),表明从中选择预测基因的模型可以显著区分应答者和无应答者。

[0236] 当考虑38个选定基因时,通过主成分分析(PCA)获得的样本之间的方差在图7D中示出。在这里,数据中解释的大部分方差(22%)沿着第一主成分(PC1),并且在很大程度上与应答状态(应答者与无应答者)相关联,表明这38个选定基因的基因表达谱在应答者与无应答者之间是不同的。

[0237] 跨平台模型优化与训练。从队列1中选择的38个预测基因在从CORRONA收集的231名类风湿性关节炎患者(队列2)中进行进一步评估。重要的是,来自这两个队列的数据是通过两个不同的平台生成的:微阵列(队列1)和RNASeq(队列2),需要在新的RNASeq平台上使用选定的特征对模型进行再训练。在队列1中鉴定的38个预测基因中,有28个可在队列2训练中检测到,并对最终确定的模型开发进行评估。在队列2的231名患者中,80%(n=185)用于优化终点、纳入标准、最终确定的特征集和模型训练,而其余的20%(n=46)留出用于保留验证。

[0238] 对于纳入标准优化,在所有训练集患者中(图8A)以及仅对RF和CCP测试血清阳性的训练组患者中(图8B)建立模型。作为模型训练的纳入标准对血清阳性进行研究,因为观察到血清阴性与纤维肌痛密切相关并可能导致较不准确的应答评估,这可能导致使用错误标记的样本进行训练。在这两个群体中,在使用不同的临床终点(ACR50 3m、ACR50 6m、CDAI 3m、CDAI 6m、EULAR 3m和EULAR 6m)时对模型进行评估。有关每个终点如何定义应答者和无应答者的信息,请参见材料和方法。

[0239] 利用使用ACR50 6m作为临床终点在男性和女性血清阳性患者 (n=90, 参见图6A) 中建立的模型实现了最高的模型性能。在此, 在重复10次的10%保留交叉验证的50次迭代中实现了72%的中值AUC。该模型的性能显著高于使用任何其他终点或纳入标准构建的模型 ($p < 0.0001$)。由于其高性能, 选择该模型进行进一步评估和保留验证。

[0240] 使用ACR50 6m作为临床终点在男性和女性血清阳性患者中建立的中值模型详述于图9A-图9D中。在10轮10%保留交叉验证后, 将验证集样本的得分聚合, 导致与无应答者相比, 应答者中的得分显著较高 ($p < 0.00001$) (图9C), 并且与72%的AUC相关联 (图9A)。

[0241] 图9B示出基于聚合的验证集样本计算的阴性预测值 (NPV) 与真阴性率 (TNR)。在这里, 0.89的NPV可以在0.55的TNR下获得。示出在该选定的最佳得分截止值 (0.29) 处预测分类与真实分类之间的一致性的混淆矩阵在图9D中示出。

[0242] 通过将分类准确度与使用临床基线协变量 (诸如年龄、BMI、患者疼痛以及压痛和肿胀关节的数量) 建立的模型进行比较, 进一步评估所选模型的性能。图10A-图10B中示出在血清阳性患者中仅使用RNA、RNA与协变量以及单独的协变量并使用ACR50 6m作为临床终点建立的模型的箱线图。使用仅使用RNA构建的模型 (中值AUC为0.72) 获得最高的模型AUC, 发现其显著高于 ($p < 0.00001$) 仅使用协变量构建的模型 (中值AUC为0.59) 以及使用与协变量组合的RNA构建的模型 (中值AUC为0.70, $p < 0.005$)。

[0243] 在50轮10%交叉验证中, 通过测量前15个 (28个中的) 特征的选择频率来评估特征鲁棒性 (图11)。通过模拟, 随机确定预期选择数为2,679 (54%)。在5000个可能的选择中的大于4000个 (80%) 中选择11种基因 (COR02B、STOML2、CHMP7、SPINT2、CYSLTR2、SSNA1、IMPDH2、GOLGA1、SUOX、ARPC1A和KLHDC3), 并将其认为是所评估的28个基因候选者中最鲁棒的子集, 因为它们的选择频率高于随机预期。观察到在前11个之后选择频率大幅下降, 并且在5000个可能的选择中, 下一个最常被选择的基因仅以54%被选择。其余的这17种基因没有被考虑用于最终确定的模型开发和验证, 因为它们的选择频率没有超过随机预期。

[0244] 用于预测对抗TNF疗法的应答的优化模型的验证。分类器性能通过作为同一CORRONA研究队列中的一部分的另外46名患者进一步评估, 但随后对其RNA进行测序以用于最终模型验证。在为模型验证保留的46个样本中, 35个在疗法开始后6个月进行随访, 并用于模型验证。全套90个训练样本用于训练最终确定的随机森林模型, 其中使用11种选定基因作为输入 (COR02B、STOML2、CHMP7、SPINT2、CYSLTR2、SSNA1、IMPDH2、GOLGA1、SUOX、ARPC1A和KLHDC3)。模型验证结果示出于图11中, 其实现了0.72的AUC, 0.85的NPV以及0.61的TNR。根据训练数据确定的最佳模型截止值 (0.29) (参见图9C) 用于混淆矩阵生成 (图9D)。仅血清阳性患者 (n=23) 的模型验证结果在图13A-图13D中示出。

讨论

[0245] 据我们所知, 这是首次在跨队列、跨平台分析中选择、优化和验证在类风湿性关节炎的背景下预测对抗TNF疗法的应答的生物标志物。在这里, 从可公开获得的数据集中选择28种预测基因, 其中基因表达通过微阵列 (Affymatrix) 进行量化。在RNASeq平台上进行模型优化后, 使用28种基因中的子集 (11) 构建模型, 所述模型表现为验证AUC为72%, NPV为0.85, 并且TNR为0.61。这些结果表明, 基因表达谱可以用于可重复性地预测对抗TNF疗法的无应答, 并且所述谱可潜在地用于筛选无应答者, 以探索替代疗法, 并随后提高针对用于治疗RA的那些所开处方的抗TNF疗法的应答率。

[0246] 目前缺乏预测抗TNF应答的临床有用分类器并非由于研究界缺乏努力。类风湿因子(RF)和抗环瓜氨酸肽(CCP)是用于诊断RA的生物标志物,并且在约80%的患者中发现了它们。已将这些循环生物标志物的水平与对利妥昔单抗的应答相关,但是趋势是温和的并且没有足够的实用性用于临床。已经开展了许多遗传关联研究,以评估针对使用抗TNF和替代疗法进行治疗的应答的遗传风险基因座的预测价值。发现了关联,但影响大小适中。一项研究估计,在接受托珠单抗治疗的患者中,DAS28评分的变化中只有2%可以解释为由于每个风险等位基因。在一项独立研究中,测试了八种不同的预测抗TNF应答的基因表达签名,大多数签名未能证明足以具有临床实用性的预测价值。有几项研究讨论了在独立的患者队列中,各种分类器无法满足所要求的性能标准。用于预测对RA托珠单抗和利妥昔单抗的替代性批准疗法的应答的分类器也已公布,但这些分类器也未能在独立的患者队列中得到验证。许多这些研究的共同点是,所使用的分析框架仅寻找遗传和临床结果数据集之间的数学相关性。因此,没有捕捉生物学背景,这就解释了为什么许多这些分类器表现不佳,或者为什么不能在独立的患者队列中复制结果。

[0247] 来自血液或组织的RNA计数包含描述单个患者的疾病生物学所需的信息。然而,考虑到通过全转录组分析(参见方法)产生的大量特征(RNA)和相对较少的样本数量,挑战始终是鉴定为患者的疾病生物学原因的较小的致病基因集。增加样本量的一种方法是利用从可公开获得的数据集中收集的信息,但是研究设计的差异,特别是用于RNA定量的平台,对在队列之间直接转化造成问题。基因表达数据集的跨平台分析是在包括结核病的领域中进行诊断开发的广泛研究主题。在本研究中,从由微阵列生成的数据集中鉴定生物标志物,并在RNASeq数据集中进一步评估。尽管在这两个平台上生成的数据已显示为高度相关,但动态范围的差异对在一个平台上进行诊断模型开发并在另一个平台上进行验证提出了挑战。在本研究中,使用微阵列数据集将高维基因表达数据降到与应答相关的较少数量的基因,使得在我们的RNASeq数据集中可以采取更具针对性的方法。在任一平台上构建的交叉验证模型都表现为具有一致的性能。

[0248] 由于对抗TNF疗法的应答有许多公认的临床定义,并且没有金标准,我们选择评估所有有数据可用的终点,并确定训练集样本中患者的基线基因表达谱最能反映哪个终点。发现对模型性能影响最大的是用于类别标签的临床终点。某些临床终点(诸如CDAI 3m、CDAI 6m和EULAR 6m)不能表现得显著高于随机水平(AUC 50%),而其他终点(诸如ACR50 6m和EULAR 6m)达到高达72%的中值AUC。由于这些终点彼此高度相关,这一发现表明,小百分比的错误标记样本可能对整体模型性能产生很大影响,反过来,一些终点更好地反映在患者转录谱中。未来的研究应进一步调查临床终点与患者转录组之间的关系,以了解哪些终点可能最具生物学相关性。

[0249] 由于以下事实将血清阳性作为纳入标准进行了调查:血清阳性患者的疾病进展相似,并且更可能患有较严重的RA,而血清阴性RA患者被假设患有不同形式的炎症性关节炎,诸如银屑病相关关节炎或脊椎关节病。因此,选择血清阳性患者进行诊断模型开发可被认为是异质性低于组合血清阳性和血清阴性患者二者的RA群体。将血清阴性患者从我们的模型训练中剔除后,模型AUC的中值和标准偏差均较高。

[0250] 协变量分析。为了进一步证明我们基于RNA的基因表达测试的鲁棒性,与基线时使用临床协变量建立的模型进行了比较。仅使用RNA构建的模型比使用临床协变量构建的模

型表现得要好得多,这一事实表明,从转录组获得的信息比在临床就诊时可在基线获取的其他简单测量具有价值。

[0251] TNF- α 是炎症和微生物免疫以及许多稳态生理功能的有效和中心介质。它是抗TNF- α 药物治疗自身免疫性疾病(包括RA)的关键靶标。TNF- α 作为跨膜蛋白表达并且其基因表达受三种转录因子调控:核因子 κ B(NF- κ B)、c-Jun激活蛋白-1(AP1)和与激活T细胞相关的核因子(NFAT)。TNF- α 与其同源受体的结合启动了复杂多样的信号传导通路活动,其可最终导致细胞死亡或细胞激活。TNF- α 受体激活导致促炎性细胞因子、趋化因子和前列腺素的合成。此外,TNF- α 调节白细胞迁移所必需的细胞骨架重组。TNF- α 激活的这种多效性作用反映在由分类器中包括的基因编码的11种蛋白质中的9种的功能中。两种参与促炎性机制,五种参与细胞骨架重塑和功能,一种参与NF- κ B调节并且两种参与T细胞激活。例如,半胱氨酰白三烯受体2(CYSLTR2)是G蛋白偶联受体(GPCR),由白三烯激活,所述白三烯是由激活的天然免疫细胞(诸如粒细胞、巨噬细胞和肥大细胞)快速生成的促炎性脂质介质(PMID:25572555)。这类GPCR触发多种白细胞功能,包括趋化性、脱粒和超氧阴离子的产生。

[0252] 细胞骨架重塑的作用在该分类器中显著可见。细胞骨架是三维的细胞内蛋白质网络,负责细胞形状、运动、内部组织以及细胞器和囊泡的移动。细胞骨架组分与RA疾病生物学的相互作用可能是多方面的。动态肌动蛋白重塑对于免疫细胞有效归巢和外渗到炎症部位至关重要。细胞因子的时间和极化释放取决于细胞骨架重塑。在我们的分类器中,细胞骨架组分的普遍性表明,细胞骨架可能在RA疾病生物学中起着比以前所认识到的更重要的作用,并且可能是预测对抗TNF- α 疗法无应答的关键。

[0253] 自身抗原是建立对自身免疫性疾病(诸如RA)最重要的不适当免疫应答的关键组分,因此,在我们的分类器中强调自身抗原并不奇怪。RA的免疫学标志是血液中存在自身抗体,包括类风湿因子(RF)——一个针对免疫球蛋白(Ig)G的Fc部分的自身抗体家族。我们的分析鉴定了两种与RA相关的另外的自身抗原基因:GOLGA1和NA14。GOLGA1编码高尔基体蛋白-97,其为将运输囊泡定位到高尔基体的分子系链(5)。此外,高尔基体蛋白-97的消除激活NF- κ B。NA14是攻击产生眼泪和唾液的腺体的另一种被称为干燥综合症的自身免疫性疾病的标志性自身抗原。此外,针对细胞骨架组分肌动蛋白和肌球蛋白的自身抗体在一些RA患者血清中富集(PMID:12447636)。

[0254] 与免疫介导疾病(诸如RA)有关的遗传和环境要素。患RA的风险在患者的一级亲属中增加并且同卵双胞胎的自身免疫的一致性相对较高(40%),表明免疫介导疾病的病因中存在遗传组分。在过去的十年中,随着人类和小鼠基因组的单核苷酸多态性(SNP)和单倍型图谱的可用,结合对来自大量良好表征的患者队列的DNA的分析,我们观察到被发现与人类自身免疫性疾病相关的基因数量激增。然而,在单卵双胞胎中,免疫介导疾病的一致性虽然显著,但小于40%,这表明环境因素和表观遗传要素也在此类疾病的发展中发挥作用。此外,目前的研究表明,与免疫基因中的风险相关的等位基因变异并未进化为增加自身免疫的易感性,而是由于诸如饮食、微生物群、维生素、吸烟等环境要素而被选择。总之,自身免疫性疾病(特别是RA)的复杂性使用不同的工具和机制通过许多不同的组进行了证实。这些研究证明了RA在分子水平上的异质性,并解释了为什么没有单一的靶向治疗可以对所有患者有效,这进一步放大了我们分类器的重要性。在这项研究中,我们已经能够使用基于基因的分类器在RA患者中预测对抗TNF疗法的无应答。分类器中的基因与这些众所周知的遗传

风险因素中的一些相关,此外还有其他新的、未被充分研究的基因,包括那些可能受环境因素和表观遗传要素影响的基因。

方法

[0255] 研究群体:队列1-从Bienkowska等人.,*Genomics*,94:423-432(2009)中公布的可公开获得的数据集GSE15258中鉴定预测基因,所述文献的全部内容通过引用并入本文。

[0256] 队列2:

- 生物学原初患者。
- 自2001年6月5日以来,北美风湿病学研究人员联盟(Consortium of Rheumatology Researchers of North America, CORRONA)一直在收集来自美国38,000多名RA患者及其医生的“现实(real world)”纵向数据。在每次登记访问时,通过机构审查委员会/伦理委员会的批准收集数据,包括疾病严重程度和活动度、药物、不良事件、生存质量、实验室和成像结果,以及社会人口统计信息65。

- CORRONA注册中心启动了CERTAIN试验,以收集和比较来自开始使用(生物原初)或更换生物药剂的具有中度至重度疾病活动的RA患者的数据,以比较生物药物的有效性和安全性,所述生物药物诸如抗TNF疗法(Humira、依那西普、英夫利昔单抗、Cimzia和Simponi),以及替代作用机制药物阿巴西普、利妥昔单抗和托珠单抗。主要的比较有效性终点(primary comparative effectiveness endpoint)是在登记前暴露于至少一种抗TNF剂的患者在一年内达到低疾病活动度。要求每三个月进行一次患者就诊和血液检查。

- CERTAIN试验对所有开始使用或更换生物药剂的具有中等水平疾病活动度(如临床疾病活动度评分(CDAI) > 10所定义)的患者开放。此外,在本研究之前,在基线时从CERTAIN患者收集PAXgene™血液RNA管,并将其储存在-80℃下。

- 对于这项回顾性验证研究,PAXgene全血RNA样本选自在生物(b)DMARD开始前全部使用伴随MTX治疗并且在三个月的随访中仍使用MTX且没有>5mg的剂量变化的患者。在三个月的随访后,患者可以停止生物制剂。在基线时或三个月的随访中,允许患者同时使用羟氯喹(Plaquenil),但不使用其他csDMARD。在三个月的随访中允许患者使用强的松(≤10mg/天)。在六个月的随访中,允许使用更高剂量(>10mg/天)的强的松。临床评估包括压痛和肿胀关节计数、医生和患者整体疾病活动度评分、患者疼痛和疲劳以及生存质量调查。实验室研究包括全血计数、C反应蛋白、RF和抗CCP滴度。如果患者需要改变MTX或羟氯喹的剂量,或在基线或三个月的随访中使用其他csDMARD疗法进行治疗,则将其排除在研究之外。强的松剂量>10mg/天的患者也被排除在研究之外。

[0257] 对抗TNF疗法的应答的评估。申请人将应答定义为达到通过CDAI < 10、DAS28CRP < 2.9或相当于ACR50或EULAR良好应答的应答定义的低疾病活动度。在抗TNF疗法开始后3个月和6个月时评估患者对抗TNF疗法的应答。

[0258] RNA分离和QC。根据制造商的说明,使用PAXgene血液miRNA试剂盒(PreAnalytix)从PAXgene血液RNA管中收集的血液中分离总RNA,包括miRNA。纯化的RNA储存在-80℃下。使用Agilent生物分析仪评估RNA质量,并使用NanoDrop ND-8000分光光度计通过分光光度法对样品进行定量。具有大于5的RIN和至少1μg的RNA的样本继续进行GlobinClear处理。

[0259] 珠蛋白耗竭和cDNA文库制备。根据制造商的说明,使用GlobinClear-人试剂盒(Thermo Fisher)从血液来源的总RNA样本中耗竭珠蛋白mRNA。使用NanoDrop ND-8000分光

光度计通过分光光度法对最终的珠蛋白mRNA耗竭的RNA样本进行定量。

[0260] 使用Illumina-TruSeq Stranded总RNA方法创建测序文库。根据制造商的说明,在使用RiboZero Magnetic Gold试剂盒生成文库之前,优先去除核糖体RNA。通过PCR扩增最终文库,并进行定量、归一化和汇集以进行测序。文库在Nextseq 500测序仪上运行75个循环,目标是每个样本至少30M读段。

[0261] RNA-seq分析和基因表达预处理。每位患者捕获平均约40.2M读段,范围为33.7至47.5M。每个基因和每个转录本的初始值由以下公式得出:(i)通过将剪接转录物与参考(STAR)比对器比对,将短的50-mer读段映射到GRCh37版本的人类基因组装配[Dobin等人,2013],然后通过期望最大化(RSEM)软件包使用RNA-Seq对转录物丰度进行定量(Li和Wewey,2011)。在分析中使用原始计数。

[0262] 统计分析:所有统计分析均使用Python 2.7.15进行。图6A详细描绘了用于特征选择、模型训练和模型验证的队列的流程图,以及用于最终模型开发的纳入标准。图6B详细描绘了队列1中用于预测基因鉴定的方法的流程图。在100轮交叉验证中,随机选择80%的样本进行特征选择和模型训练。在80%的训练样本中,使用Mann-Whitney U检验消除应答者与无应答者之间表达无显著差异的任何特征($p > 0.05$)。随后,通过10倍交叉验证对其余特征进行进一步评估,其中使用随机森林根据其平均不纯度减少(mean decrease impurity)对每个特征进行排序。在5/10交叉验证折叠中排名前100位的特征用于构建随机森林模型,所述模型使用20%的保留测试样本进行测试。该过程重复100次并且在100次迭代中的30次中选择的特征被认为是用于进一步评估的预测性生物标志物。

[0263] 在从北美风湿病学研究人员联盟(CORRONA)收集的231名患者的独立队列中进一步评估抗TNF疗法应答的预测性生物标志物(参见第3.1节,研究群体)。185个样本(队列的80%)用于最终确定的特征选择、临床终点评估和模型训练。最佳模型参数和临床终点通过10次重复的10%保留交叉验证(重复50次)进行评估。在每一轮交叉验证中,使用通过随机森林平均不纯度减少排序前15位的特征构建模型,并在保留的10%样本上进行测试。对于最终确定的特征选择,如果特征在5000个可能选择中的4000多个排序在前15位,则选择所述特征。最终确定的随机森林模型随后使用这些选定特征进行训练,并在队列中的其余46个样本(20%)中进行验证。

实施例2

[0264] 本实施例描述了分类算法的跨队列、跨平台开发和验证,所述算法在治疗开始前在类风湿性关节炎(RA)患者中预测在基线时对抗TNF疗法无应答。此外,本实施例描述了可以应用于开发其他分类器(例如,预测其他特征(诸如,实现或遭受特定结果或副作用的可能性等)和/或与其他疗法相关)的某些技术。

1. 方法

(i) 研究群体

(a) 发现队列

[0265] 患者微阵列数据(登录号GSE15258)获自基因表达综合数据库(GEO)。样本收集和队列信息的详细信息先前已有报告。简而言之,登记对于抗TNF疗法为原初的RA患者,并将血液样本收集在PAXgene管中。根据DAS28-CRP EULAR应答定义,在开始治疗14周后评估治疗应答。使用58个女性患者样本。

(b) 训练队列和验证队列

[0266] 北美风湿病学研究人员联盟 (CORRONA) 在CERTAIN试验中前瞻性地收集了RA患者全血样本和临床测量结果。CERTAIN研究设计为前瞻性比较有效性研究,涉及43个地点和117名风湿病专家。在样本收集和参与研究之前,获得了机构审查委员会和伦理委员会的批准。本研究选择的样本在样本收集时是生物学原初的(即,患者没有经历以下疗法:阿达木单抗、依那西普、英夫利昔单抗、赛妥珠单抗、戈利木单抗、托珠单抗、利妥昔单抗和阿巴西普)。根据治疗医师的判断,使用阿达木单抗、依那西普、赛妥珠单抗、戈利木单抗或英夫利昔单抗治疗患者,并接着进行至少6个月的纵向随访。除病史外,在疗法开始后零个月和六个月时收集的临床评估包括压痛和肿胀关节计数、医生和患者整体疾病活动度评分、cSDMARD剂量、患者疼痛评估和生存质量调查。在中心实验室进行的实验室研究包括全血计数、C反应蛋白(CRP)水平、类风湿因子滴度和抗环瓜氨酸蛋白(抗CCP)滴度。患者人口统计特征和临床变量详见表3。训练(n=143)和验证试验(n=175)患者队列在应答率、年龄和性别方面是平衡的。如果患者的视觉模拟量表疼痛评分在100分的最高评分中为至少15分,则将其包括在独立验证试验中。与CERTAIN研究的纳入标准一致,验证试验中所有患者的CDAI均大于10。

表3

特征	训练队列 (n=143)	验证试验队列 (n=175)
年龄, 岁(平均值 ± SD)	55.0 ± 12.9	53.8 ± 11.9
女性, %	72.7	73.1
疾病持续时间, 年(平均值 ± SD)	4.7 ± 6.7	5.0 ± 7.5
抗环瓜氨酸肽阳性, %	66.9	60.9
类风湿因子阳性, %	73.9	70.9
种族		
白人	88.1	84.6

特征	训练队列 (n=143)	验证试验队列 (n=175)
黑人	6.3	6.3
其他	5.6	9.1
csDMARD 使用, %		
甲氨蝶呤	65.0	60.0
羟氯喹	3.5	4.0
>=2 csDMARDS	13.3	13.7
无	15.4	16.0
伴随强的松, %	34.3	22.9
抗 TNF 使用, %		
阿达木单抗	38.5	38.9
依那西普	33.6	30.9
英夫利昔单抗	16.8	20.0
赛妥珠单抗	8.4	7.4
戈利木单抗	2.8	2.9
CDAI (平均值 ± SD)	28.5 ± 13.5	31.0 ± 12.6
DAS28-CRP(平均值 ± SD)	4.9 ± 1.1	5.0 ± 1.0
肿胀关节计数 (平均值 ± SD)	7.2 ± 6.0	8.1 ± 5.5
压痛关节计数 (平均值 ± SD)	10.8 ± 7.3	12.0 ± 7.3

特征	训练队列 (n=143)	验证试验队列 (n=175)
SD)		
ACR50 应答者, %	30.8	30.3

(ii) 对抗TNF疗法的临床应答的评估

[0267] 在CORRONA研究样本中,抗TNF疗法开始后6个月的应答通过ACR50定义(ACR50是指由美国风湿病学会定义的患者治疗目标,即患者实现缓解或低疾病活动度,其中数字(例如50)是指一组标准量度的改善百分比,包括肿胀和压痛关节数量、患者和医生报告评估以及疼痛和整体健康,以及血清生物标志物)。应答者是指表现出28个压痛关节计数改善 $>50\%$ 、28个肿胀关节计数改善 $>50\%$ 并且5个临床变量中的至少3个(疾病指数、患者疼痛、患者整体评估、医生整体评估和CRP水平)改善 $>50\%$ 的个体。

(iii) RNA分离与QC

[0268] 根据制造商的说明,使用PAXgene血液miRNA试剂盒(PreAnalytix)从PAXgene血液RNA管中收集的血液中分离总RNA。使用Agilent生物分析仪评估RNA质量,并使用NanoDrop Nd-8000分光光度计对样品进行定量。

(iv) RNA测序(RNAseq)分析与基因表达预处理

[0269] 根据制造商的说明,使用GLOBINclear(Thermo Fisher)、Ribo Zero Magnetic Gold(流行病学(Epidemiology))和TruSeq Stranded总RNA(Illumina)试剂盒处理RNA。文库在NextSeq 550 DX或NovaSeq 6000测序仪上处理75个循环。每个患者捕获平均42.4M读段,范围为33.7-58.6M。使用STAR将50-nt读段映射到GRCh37人类基因组。使用RSEM软件包计算每百万个映射读段每千碱基转录物的片段(FPKM)的每个基因丰度。对RNA完整性评分(RIN)大于400万或大于700万蛋白质编码读段的样本进行分析。基于主成分分析,在NextSeq与NovaSeq处理文库之间未观察到可检测的批次效应。

(v) 单核苷酸多态性(SNP)分析

[0270] 除其他事项外,本公开提供了这样的见解,即SNP分析提供了特别有用的信息以包括在如本文所述的分类器中——例如,使用RNAseq数据和/或考虑到基因表达水平的微小但显著差异开发,如本文所述。为了获得SNP数据,使用STAR将样本与GRCh38人类基因组比对。SNP是使用在RNAseq数据上调用的SNP和indel的Genome Analysis ToolKit Best Practices工作流的修改版本调用的。评估了39个RA相关SNP。

(vi) 机器学习与药物应答算法开发

[0271] 对58名患者的发现队列中的21,818个转录本特征进行评估。在随机选取的80%样本中,Mann-Whitney U检验用于消除在应答者与无应答者之间无显著区分性的任何基因表达($p>0.05$)。使用随机森林机器学习算法,根据其平均不纯度减少对每个其余特征进行排序。使用前100个特征构建随机森林模型,所述模型使用保留数据的20%子集进行测试。该过程重复100次,并进一步考虑在 $>30\%$ 的迭代中选择的特征。在训练队列中,通过RNAseq可检测到37个选定基因中26个的基因表达。由于低表达,消除另外两个特征。在最终模型开发过程中,使用训练队列数据,通过如上所述的随机森林评估70个区分性分子和临床特征(表4-6)。通过将每个特征从特征列表中去除并在保留数据的20%子集上评估模型性能来评估所述特征的重要性。前50%最重要的特征用于使用整个训练集构建随机森林预测分类模型。使用接受者操作曲线下面积、阴性预测值和真阴性率评估模型性能。所有统计分析均使用Python 2.7.15进行(www.python.org)。

[0272] 表4提供了在类风湿性关节炎患者中预测对抗TNF无应答的分类算法的开发过程中评估的区分性基因集。

表4

基因名称
基因名称
ARPC1A
ATAT1
ATRAID
CD27
CHMP7
COMMD5
CORO2B
CYSLTR2
DNAJC7
GOLGA1
HPCA
IMPDH2
KAT8
KLHDC3
LOC100506083
MORN2
NGF
RHOBTB1
RNGTT

基因名称
SETD9
SNX8
SPINT2
SPON2
SSNA1
STOML2
SUOX
TMEM258
UBL7-AS1

[0273] 表5提供了在类风湿性关节炎患者中预测对抗TNF无应答的分类算法的开发过程中评估的SNP。

表5

SNP
chr1.161644258
chr1.2523811
chr11.107967350
chr17.38031857
chr7.128580042
rs10774624

SNP
rs10985070
rs11889341
rs1571878
rs1633360
rs17668708
rs1877030
rs1893592
rs1980422
rs2228145
rs2233424
rs2236668
rs2301888
rs2476601
rs3087243
rs3218251
rs331463
rs34536443
rs34695944
rs4239702

SNP
rs4272
rs45475795
rs508970
rs5987194
rs657075
rs6715284
rs706778
rs72634030
rs73013527
rs73194058
rs773125
rs7752903
rs8083786
rs9653442

[0274] 表6提供了来自CERTAIN队列的本研究中可用的临床因素。

表6

临床特征
年龄
出生性别

临床特征
疾病持续时间（以年为单位）
种族（包括白人、亚洲人、黑人、混血人种、美洲原住民、太平洋岛民和其他种族）
纤维肌痛史
慢性血管疾病史（包括急性冠脉综合征、冠状动脉疾病、充血性心力衰竭、高血压、心肌梗死、外周动脉疾病、中风、不稳定型心绞痛、心脏骤停、血运重建术和室性心律失常）
导致住院的严重感染史（包括粘液囊或关节感染、蜂窝组织炎、鼻窦炎、憩室炎、败血症、肺炎、支气管炎、胃炎、脑膜炎、尿路感染、上呼吸道感染和结核病）
癌症史（包括乳腺癌、肺癌、皮肤癌、淋巴瘤，但不包括非黑色素瘤皮肤）
BMI
吸烟状态（包括从不吸烟、以前吸烟或目前吸烟）
强的松剂量
DMARD 剂量
基线时的 C 反应蛋白水平
基线时的 DAS28-CRP
基线时的肿胀的 28 关节计数
基线时的压痛的 28 关节计数
基线时的患者整体评估

临床特征
基线时的医师整体评估
基线时的 CDAI
基线时的经过修改的健康评估问卷得分
基线时的患者疼痛评估
使用 DAS28-CRP 获得的基线时的 EULAR 应答 (包括差、中或好)
抗 CCP 状态 (阳性或阴性)
基线时的抗 CCP 滴度
类风湿因子状态 (阳性或阴性)
基线时的类风湿因子滴度

(vii) 构建人类相互作用组和RA疾病模块,并执行分子特征的网络医学分析

[0275] 如先前所述,人类相互作用组由21个公共数据库(表5)组装而成,所述数据库包含不同类型的通过实验得到的蛋白质-蛋白质相互作用(PPI)数据:1)二元PPI,来源于高通量酵母双杂交(Y2H)实验(HI-Union)、三维(3D)蛋白质结构(Interactome3D、Instruct、Insider)或文献管理(PINA、MINT、LitBM17、Interactome3D、Instruct、Insider、BioGrid、HINT、HIPPIE、APID、InWeb);2)通过亲和纯化然后质谱鉴定的PPI,存在于BioPlex2、QUBIC、CoFrac、HINT、HIPPIE、APID、LitBM17和InWeb中;3)来自KinomeNetworkX和PhosphoSitePlus的激酶-底物相互作用;4)来自Signalink和InnateDB的信号传导相互作用;以及5)由ENCODE联盟得到的调控相互作用。Alonso-López等人提供的PSI-MI ID的整理列表用于区分文献整理数据库中存在的几种实验方法之间的二元相互作用。特别是对于InWeb,未考虑管理评分<0.175(第75百分位)的相互作用。所有蛋白质都定位在其相应的Entrez ID(NCBI)中并去除无法定位的蛋白质。所得到的人类相互作用组包括18,505种蛋白质和327,924种相互作用。

[0276] 使用DIAMOnD方法生成RA疾病模块。用作疾病模块种子的蛋白质通过五个数据库中的至少两个与RA相连:(GWAS Catalog、HuGE Navigator Phenopedia、ClinVar、OMIM和MalaCards)。DIAMOnD鉴定了与疾病相关蛋白相同的Gene Ontology生物学过程而言显著富集的蛋白质。

[0277] 如先前所述,计算了人类相互作用组图谱上分子特征彼此之间的邻近性。简而言之,最近距离定义为集合中各蛋白质与下一个最近蛋白质之间的平均最短路径长度。将观察到的最近距离的显著性与从10,000个相同大小的随机蛋白质集中确定的预期最近距离

进行比较。如先前所述进行随机化。

(viii) 路径富集分析

[0278] KEGG、Biocarta、Reactome和信号转导(Signal Transduction,ST)通路注释获自分子签名数据库(Molecular Signatures Database,MSigDB),版本6.2。使用Fisher精确检验鉴定生物通路。Bonferroni校正的 p 值 <0.05 的通路被认为是富集的。IL10、POMC、JAK1、ICOSLG、TNF、TNFSF11、NR3C1、P2RY12 (NCT02874092)、PTGER4、GGPS1、FDPS、TNFRSF13B (NCT03016013)、IL6、ESR1、ESR2、ITK (NCT02919475)、BTK、TLR4 (NCT03241108)、IRAK4、JAK2、JAK3、HDAC1 (NCT02965599)、PSMB5、ADORA3、ITGA9 (NCT02698657、NCT03257852)、IFNB1 (NCT02727764:NCT03445715)、CX3CL1是RA中批准的药物靶标。

2. 结果

构建人类相互作用组和RA疾病生物学图谱

[0279] 为了开始开发评估人类疾病生物学所需的网络医学工具,创建了细胞组分及其物理相互作用的图谱。通过合并总共18,505种蛋白质之间327,924种蛋白质-蛋白质相互作用的可公开获得的数据(表7),创建了称为人类相互作用组的生物学综合图谱(参见方法)。

表7

数据库	相互作用类型	证据类型	链接
APID	二元/共复合 (Co-complex)	文献管理	http://apid.dep.usal.es/
BioGrid	二元	文献管理	https://thebiogrid.org/
BioPlex2	共复合	高通量	https://bioplex.hms.harvard.edu/
CoFrac	共复合	高通量	http://human.med.utoronto.ca
ENCODE	调节性	高通量	http://encodenets.gersteinlab.org/enets2.Proximal_filtered.txt
HI-Union	二元	高通量	http://www.interactome-atlas.org/download
HINT	二元/共复合	文献管理	http://hint.yulab.org/
HIPPIE	二元/共复合	文献管理	http://cbdm-01.zdv.uni-mainz.de/~mschaefer/hippie/

数据库	相互作用类型	证据类型	链接
InnateDB	信号传导	文献管理	https://www.innatedb.com/
知情人	3D / 二元	文献管理	http://interactomeinsider.yulab.org/
指导	3D / 二元	文献管理	http://instruct.yulab.org/
IntAct	二元	文献管理	https://www.ebi.ac.uk/intact/
Interactome3d	3D / 二元	文献管理	https://interactome3d.irbbarcelona.org/
InWeb (仅得分 >0.175)	二元/共复合	文献管理	https://www.intomics.com/inbio/map.html#downloads
KinomeNetworkX	激酶-底物	文献管理	https://bioinfo.uth.edu/kinomenetworkX/
LitBM17	二元/共复合	文献管理	http://www.interactome-atlas.org/download
MINT	二元	文献管理	https://mint.bio.uniroma2.it
PhosphoSitePlus	激酶-底物	文献管理	https://www.phosphosite.org
PINA	二元	文献管理	https://omics.bjcancer.org/pina/

数据库	相互作用类型	证据类型	链接
QUBIC	共复合	高通量	表 S2 在 Hein MY 等人中 <i>Cell</i> 2015; 163 (3): 712-23.
SignaLink	信号传导	文献管理	http://signalink.org/

[0280] 疾病相关蛋白倾向于在称为疾病模块的人类相互作用组的子网络中相互作用。使用根据其已知疾病相关蛋白的邻近性聚合潜在疾病相关蛋白的DIAMOnD方法,生成了包含大约200种蛋白质的RA疾病模块。其中,66%与全基因组关联研究数据库中的RA相关,并且DIAMOnD鉴定了与疾病相关蛋白相同的Gene Ontology生物学过程而言显著富集的其余蛋白质。

[0281] 利用人类相互作用组和RA疾病模块,目前的应用试图创建基于血液的分类算法,所述算法整合了临床变量和分子特征,以预测哪些RA患者不对抗TNF疗法产生应答(图11)。简而言之,区分抗TNF疗法的应答者和无应答者的分子特征选自可公开获得的微阵列数据集。在跨平台分析中,将这些特征与RA疾病模块相关SNP和临床因素相组合。使用RNAseq数据训练机器学习算法。最后,在独立验证试验中验证预测药物应答算法的性能。

(i) 在RA患者中预测对抗TNF疗法的应答不足的全血中的区分性基因表达的跨平台鉴定

[0282] 为了最大限度地发挥预测对疗法无应答的测试的临床效用,不需要专门的标本采集程序的常规无创或微创样本源是理想的。因此,对源自全血的基因表达数据进行分析。使用随机森林机器学习算法,从58名生物学原初RA患者的可公开获得的微阵列发现队列数据集中选择在被视为抗TNF疗法的应答者和无应答者的患者之间具有区分性的基因表达。在发现数据集中评估基因表达的21,818个基因中,37个在抗TNF疗法的应答者(n=24)与无应答者(n=51)之间具有区分性(图9;表3)。通过微阵列和RNAseq进行的转录谱分析在动态范围内变化,并且在观察到的差异基因表达的数量和程度上表现出一些不一致性。尽管如此,在微阵列数据中被鉴定为对抗TNF药物应答具有区分性的大部分转录本在RNAseq数据中也在应答者与无应答者之间进行区分。

(ii) 从RNAseq数据评估疾病相关SNP

[0283] RNAseq提供了微阵列分析中缺少的核苷酸序列的信息。除了基因表达之外,RNA序列的变化可在RA患者中预测对抗TNF疗法无应答。为此,由从CORRONA CERTAIN研究中的143名RA患者获得的临床数据和全血RNAseq数据生成训练数据集。表3汇总了患者群体的特征和人口统计资料。尽管传统上SNP分析是在全基因组测序数据上进行的,但大部分基因组是转录的。因此,大多数SNP变体可以在核糖体RNA缺失的RNAseq数据中检测到。通过表达数量性状基因座(eQTL)分析,与类风湿性关节炎相关的几个SNP在功能上与外周血单核细胞的基因表达变化相关(表4)。与所选SNP相关的遗传基因座与RA疾病模块有显著重叠(图8B)。在患者RNAseq数据中,22个这样的SNP高于检测限,并被纳入进一步的分析(图9)。

(iii) 整合SNP、基因表达数据和临床变量以开发多因素预测药物应答算法

[0284] 使用指示药物应答的基因表达(表4)、RA相关SNP(表5)和临床因素(表6)来训练和开发预测抗TNF疗法无应答的药物应答算法。以6个月时的ACR50为基准,训练队列群体对抗TNF疗法的应答率为30.8%。这是一般群体的代表并且反映了CORRONA CERTAIN研究的现实前瞻性收集方法。随机森林用于使用80%的RNAseq训练数据集,使用来自区分性基因表达集、SNP以及临床因素或其组合的特征生成预测模型。在其余20%的训练数据集上进行的交叉验证使用接受者操作特征曲线下面积评估模型性能。该分析评估假阳性与真阳性之间的关系,以衡量模型能够在应答者与无应答者之间进行区分的程度。最能预测抗TNF疗法无应答的模型是使用所有三种特征类型的组合生成的。模型交叉验证性能的接受者操作曲线下面积(AUC)为0.66。仅利用临床因素,AUC为0.56,并且仅利用分子特征,AUC为0.64(表8)。交叉验证性能数据证实了转录本和SNP特征在分类器中的关键重要性。表8提供了来自使用完整特征集及其子集的模型的交叉验证结果。

表8

用于构建随机森林模型的特征	曲线下面积
区分性基因、SNP和临床因素	0.66
区分性基因和SNP	0.64
仅临床因素	0.56

(iv) PrismRA™(在RA患者中预测抗TNF疗法无应答的生物标志物小组)的盲法独立验证试验

[0285] 许多研究假设抗TNF疗法无应答的生物学反映在全血的转录组中。然而,没有一个被转化为临床,这可能反映了RA疾病生物学的复杂性和用于算法开发的不同方法。此外,到目前为止,有限的样本量和基因表达数据分析的复杂性阻碍了可在患者队列和更广泛的患者群体中泛化的算法的开发。为了确认由12个SNP、8个转录本、2个实验室测试和3个基本人口统计学因素(性别、BMI、患者疾病评估)组成的PrismRA™生物标志物小组是可泛化的,使用前瞻性收集样本的独立组(n=175)进行盲法验证试验。验证队列中包括的样本未用于算法开发的任何阶段,并且算法不具有从这些患者的基因表达数据或临床结果中得到的信息。

[0286] PrismRA™生物标志物小组的盲法独立验证将验证队列分层为预测的无应答者和应答者,其中在相应子组中为应答者的高度统计学显著性优势比为0.152(CI 0.064、0.364)。PrismRA™以89.7%(阴性预测值)的准确度鉴定50.0%(特异性)的无应答患者(表9)。预测为无应答者的患者在抗TNF情况下观察到的ACR50应答率为10.3%,显著低于30.3%的总体应答率。将这些个体重新定向到替代疗法将大大增加他们实现有意义的临床改变的可能性。相反,预测应答者的观察到的ACR50应答率为43.0%,其比未分层患者群体的ACR50应答率提高41.8%。表9提供了PrismRA™预测性生物标志物小组验证性能。

表9

	ACR50 (R)	ACR50 (NR)	总和
PrismRA™ (R)	46	61	107
PrismRA™ (NR)	7	61	68
总和	53	122	

(v) 区分抗TNF疗法的应答者和无应答者的基因产物的生物学解释

[0287] 所有已知的蛋白质-蛋白质相互作用的人类相互作用组图谱为评估抗TNF疗法无应答的潜在生物学提供了独特的优势。为了表征PrismRA™预测性生物标志物小组对RA疾病生物学的适用性,利用人类相互作用组和通路富集分析对区分性基因和SNP eQTL的蛋白质产物进行分析。PrismRA™小组中包括的区分性基因和SNP eQTL编码的蛋白质在人类相互作用组图谱上重叠(图8A)。总共有42种蛋白质映射到人类相互作用组上:24种由区分性基因贡献,18种由SNP eQTL贡献。这些分子特征在图谱上交织在一起,突出显示了小而凝聚的生物网络,它将不对抗TNF疗法产生应答的RA患者的亚群统一起来。这种邻近性的量化(参见方法)表明这些不同的分子特征彼此非常接近(z 评分=-2.18)。此外,RA疾病模块(z 评分=-4.09)和RA药物靶标诸如JAK和TNF(z 评分=-3.98)共同接近SNP和DG(图8B)。

[0288] 进行通路富集分析以深入了解参与抗TNF疗法应答的分子通路。在所查询的通路分析数据库中,T细胞信号传导被鉴定为最富集的通路。T细胞信号传导与抗TNF疗法应答和RA疾病生物学的相关性和重要性已得到充分证实。

3. 讨论

[0289] 通过将微阵列基因表达数据、RNA下一代测序、生物网络分析和机器学习结合到大型患者队列中,本研究表明,使用全血基因表达数据预测抗TNF疗法无应答的药物应答算法可以在不同的RA患者队列中执行。用于PrismRA™的盲法独立验证试验的前瞻性收集的RNAseq数据集证明,药物应答算法可以90%的准确度在RA患者的独立群体中预测对抗TNF疗法无应答。抗TNF疗法未能帮助近70%的未分层患者群体达到治疗目标。将PrismRA结合到临床护理中将使这一数字减少一半以上。

[0290] 预测药物应答算法源自微阵列数据集中鉴定的区分性分子特征,并结合来自源自RNAseq数据的第二数据集的临床因素进一步评估。这两个基因表达分析平台在RNA检测方法和统计工具方面有所不同以确定归一化的基因表达值。尽管在技术上存在这些差异,但在本研究中鉴定的分子特征的跨平台和跨队列普遍性突出了在抗TNF药物应答的生物学基础上存在鲁棒且可泛化的分子过程。

[0291] 单一的大规模高通量分析方法尚未获得RA疾病生物学的完整分子架构。SNP可以影响细胞生物学的许多方面,包括调控元件与其同源蛋白质配偶体相互作用的倾向、从基因位点产生的替代性剪接变体的比率或身份、转录水平和蛋白质序列。因此,疾病相关SNP的功能读数有助于个体发展疾病的倾向以及环境因素影响病理生物学的倾向。许多不编码蛋白质的调控元件和基因组区域被转录,诸如以增强子RNA和启动子相关转录物的形式。因

此,许多影响转录的时空特异性变化的SNP可以从RNAseq数据评估。总之,SNP和基因表达分析可以捕获表型变异和可能以其他方式被掩盖的通路关联。将核苷酸变体和区分性基因表达整合到单一预测算法中产生了对疾病发病机制和药物应答的可执行见解。

[0292] 定制治疗方案以匹配每个患者的个体化疾病生物学是现代医学的目标。这种个性化的医学方法被用于肿瘤学,在其中对具有特定基因组标志物的患者开出特定疗法的处方。诸如RA的自身免疫性疾病的复杂性和异质性需要新的多方面的方法来开发临床有用的工具。对鉴定不对抗TNF疗法产生应答的患者的分子通路的检查表明了T细胞信号传导与RA疾病生物学之间的联系。滑膜炎症是由白细胞浸润并滞留在滑膜室中,以及慢性炎症性细胞的凋亡不足导致。该滑膜浸润物包括自然杀伤(NK)细胞、CD4⁺和CD8⁺T细胞。此外,在RA患者的关节中可以检测到大量激活的T调节细胞。其余与T细胞信号传导无关的区分性基因可能代表RA疾病的不同方面,所述方面在那些将对或不对抗TNF疗法产生应答的患者之间不同。与RA疾病生物学的联系证明了PrismRATM成为用于鉴定抗TNF无应答者的有力临床工具的可靠性和适用性。

[0293] 对于预测对抗TNF疗法应答不足的患者,替代性生物制剂和靶向合成疗法通过快速且持续地改善疾病活动度而提供更高的安全性和有效性,其中在替代性治疗开始后6个月,ACR50应答率为30.40%。FDA和RA患者治疗指南支持若干此类替代疗法作为一线疗法。将这种分层预测分类算法作为血液测试结合到RA的临床护理中将导致更快地实现更多患者的低疾病活动度。这将节省患者数月甚至数年的时间,防止潜在的疾病进展,避免暴露于与抗TNF疗法的循环相关联的潜在严重副作用,并减少无效治疗和不良事件管理的医疗支出。

[0294] 利用这种机器学习和网络医学方法开发和验证预测对靶向疗法的无应答性的药物应答算法,不仅对RA,而且对其他复杂的自身免疫性疾病,都在推进精准医疗方面显示出极大的前景。

[0295] 前述内容是对本文所述主题的某些非限制性实施方案的描述。因此,应当理解,本说明书中描述的实施方案仅仅是例示本说明书中报告的主题。参考所例示实施方案的细节并不旨在限制权利要求的范围,所述权利要求本身列举了被视为基本的那些特征。

[0296] 可以设想,所要求保护的题目的系统和方法包括使用来自本说明书中描述的实施方案的信息开发的变型和改编。本说明书中所述的系统和方法的改编、修改或两者均可由相关领域的普通技术人员执行。

[0297] 在整个说明书中,在系统被描述为具有、包括或包含特定组件的情况下,或者在方法被描述为具有、包括或包含特定步骤的情况下,可以设想,另外,存在基本上由或由所列举的组件组成的本发明主题所涵盖的系统,并且存在基本上由或由所列举的处理步骤组成的本发明主题所涵盖的方法。

[0298] 应当理解,只要本说明书中所描述的主题的任何实施方案保持可操作,步骤的顺序或用于执行特定动作的顺序就无关紧要。此外,可同时进行两个或更多个步骤或动作。

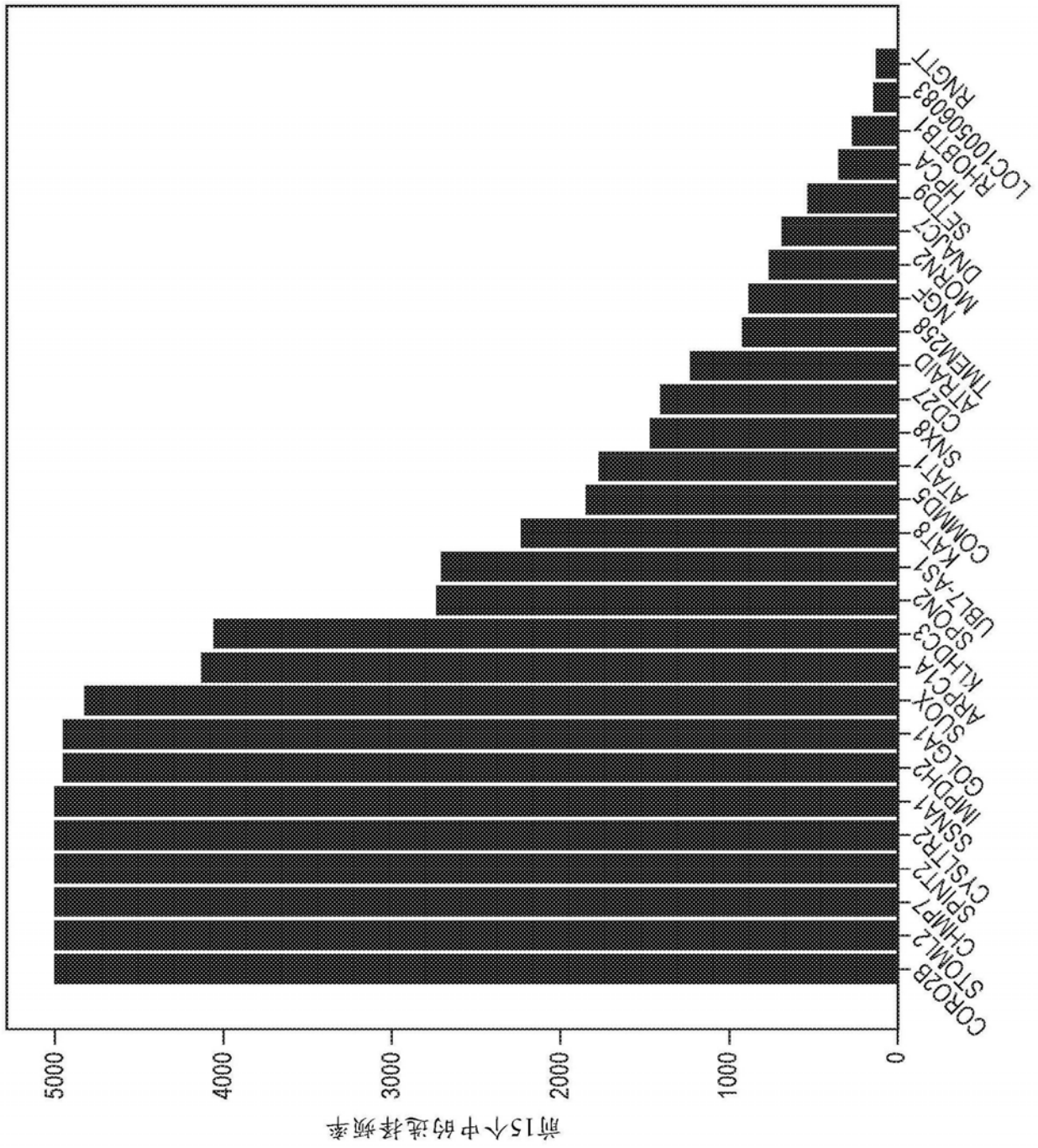


图1

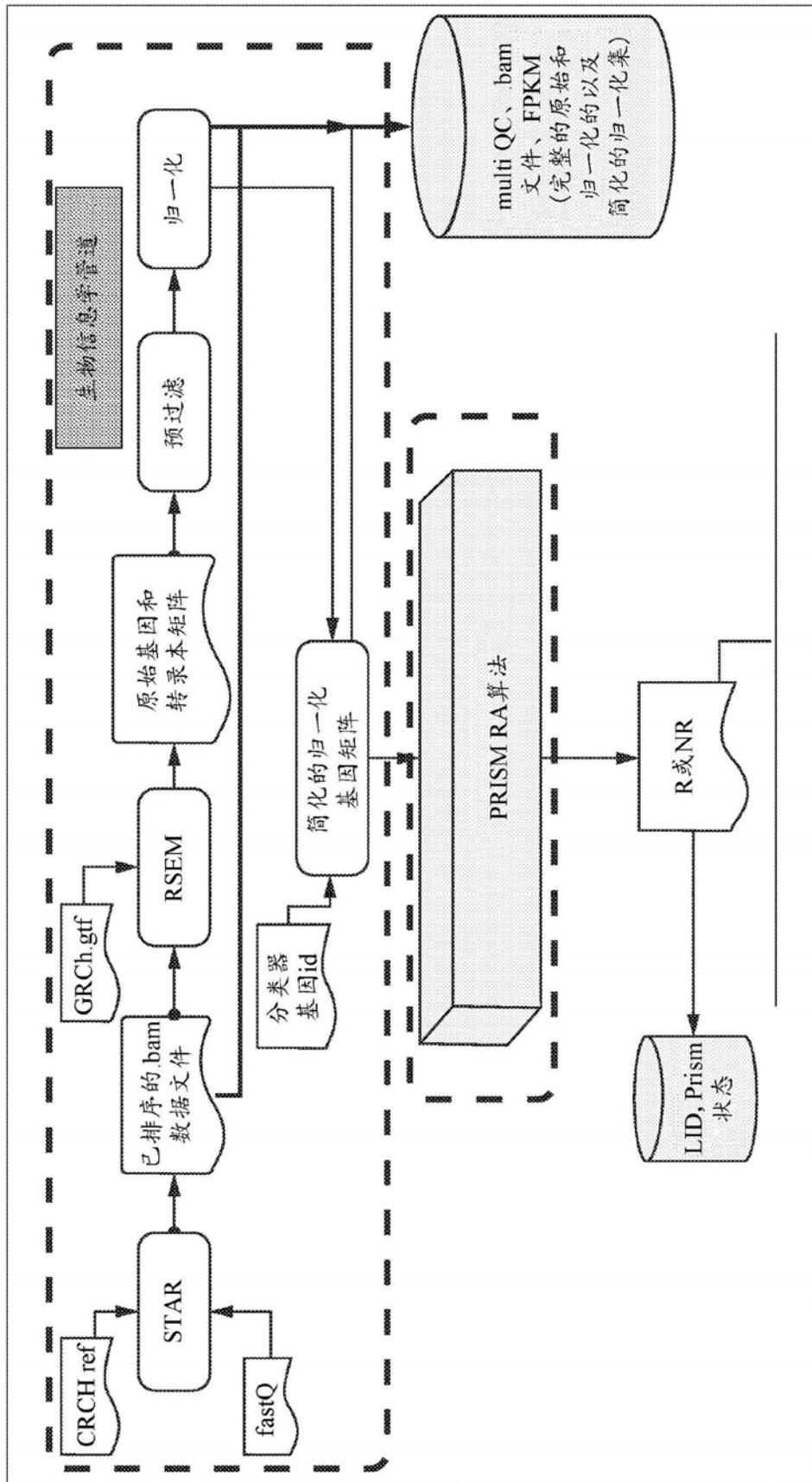


图2

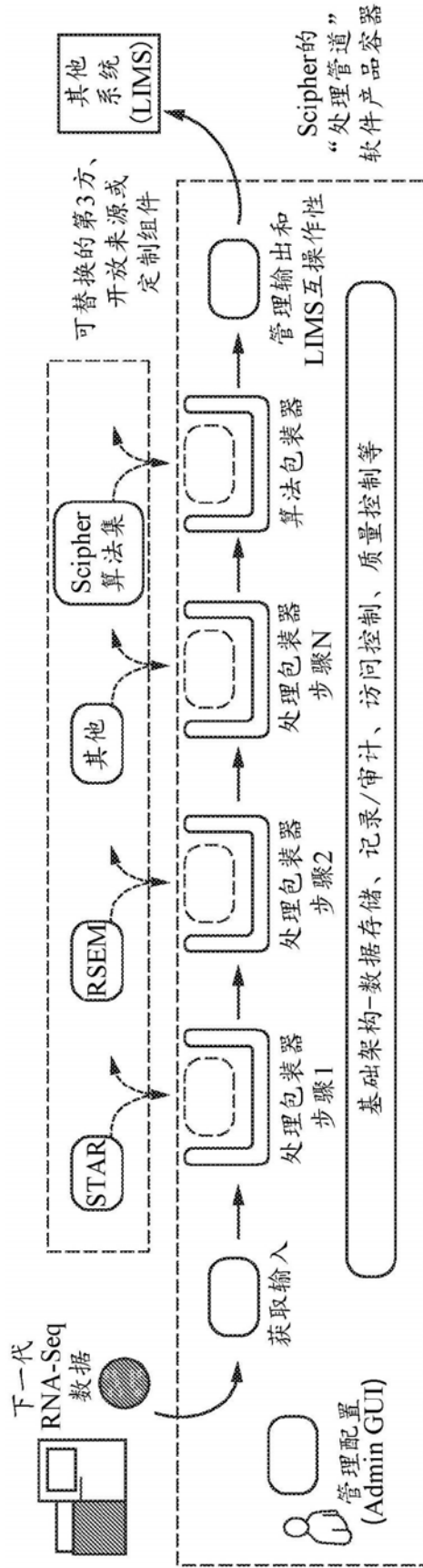


图3

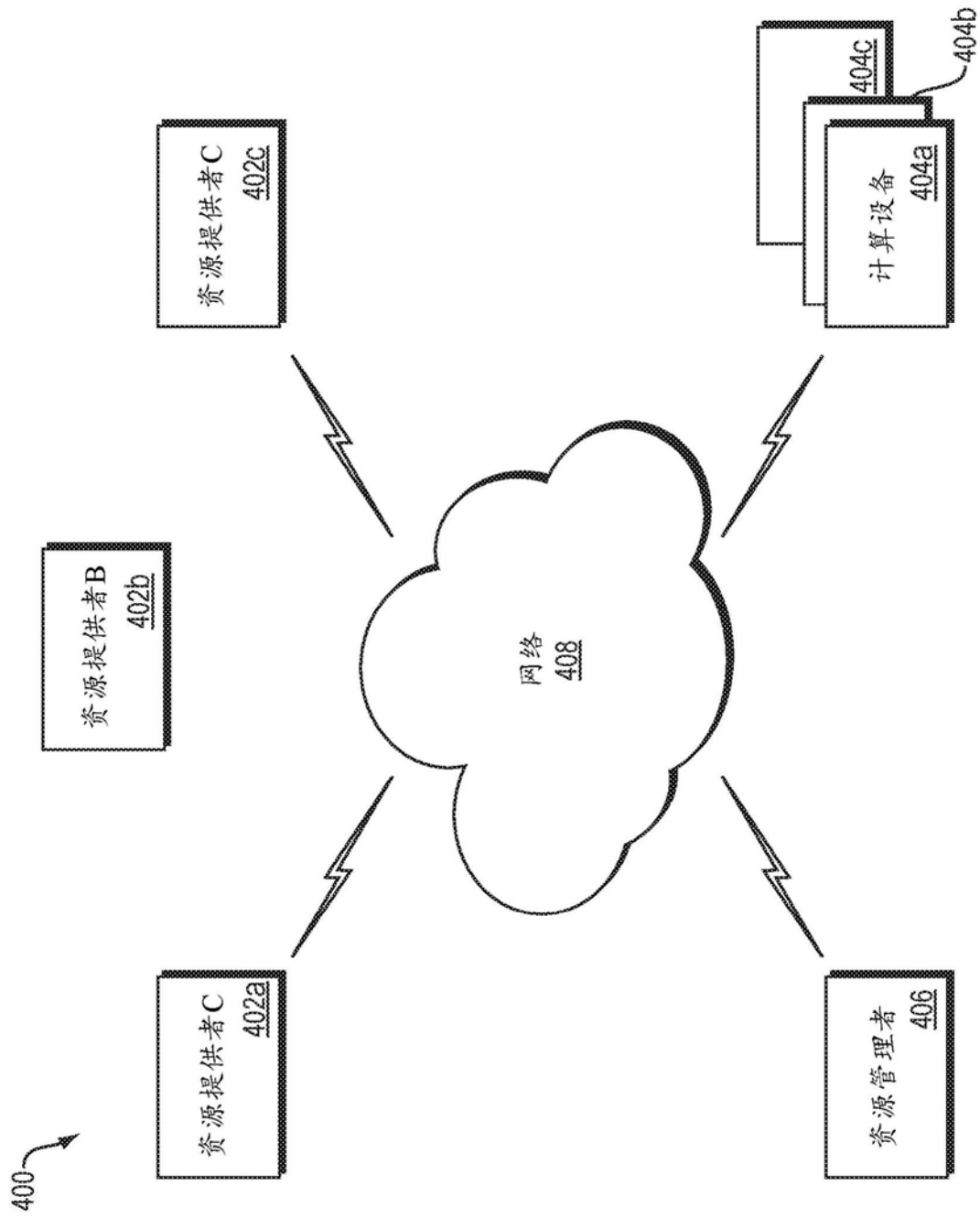


图4

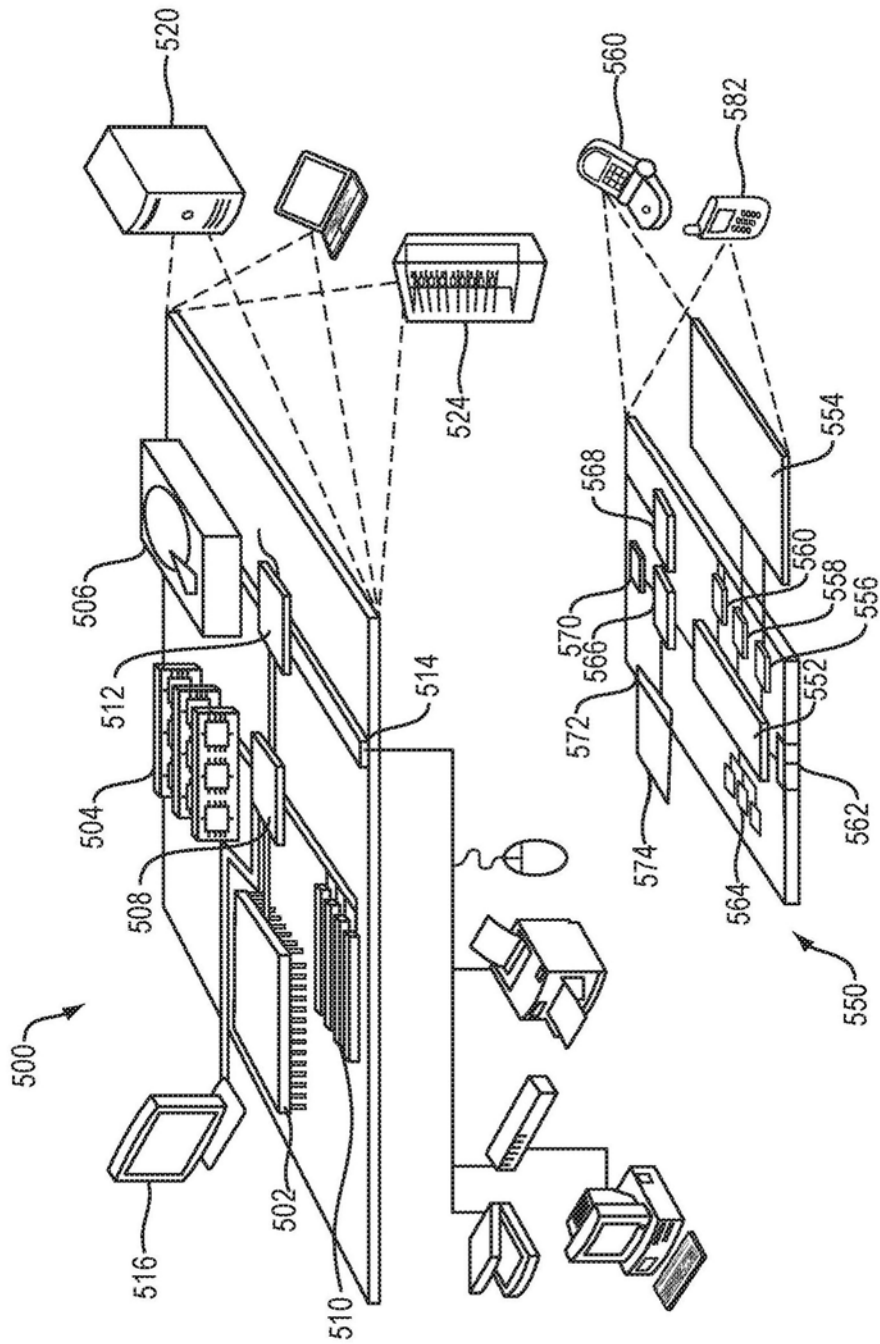


图5

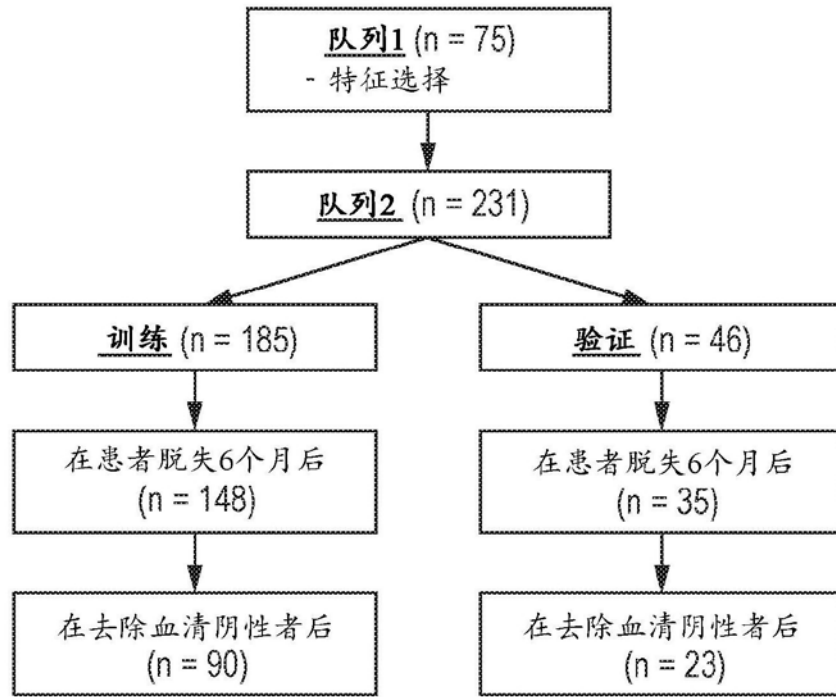
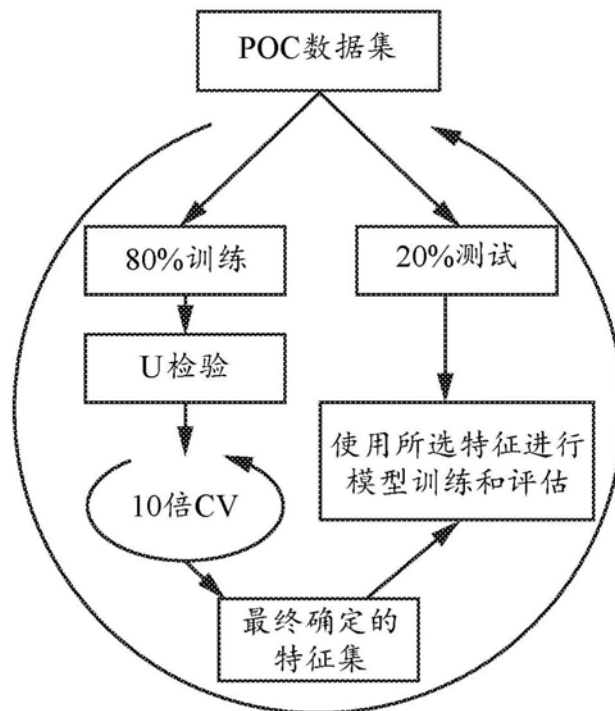


图6A



100次整体迭代

图6B

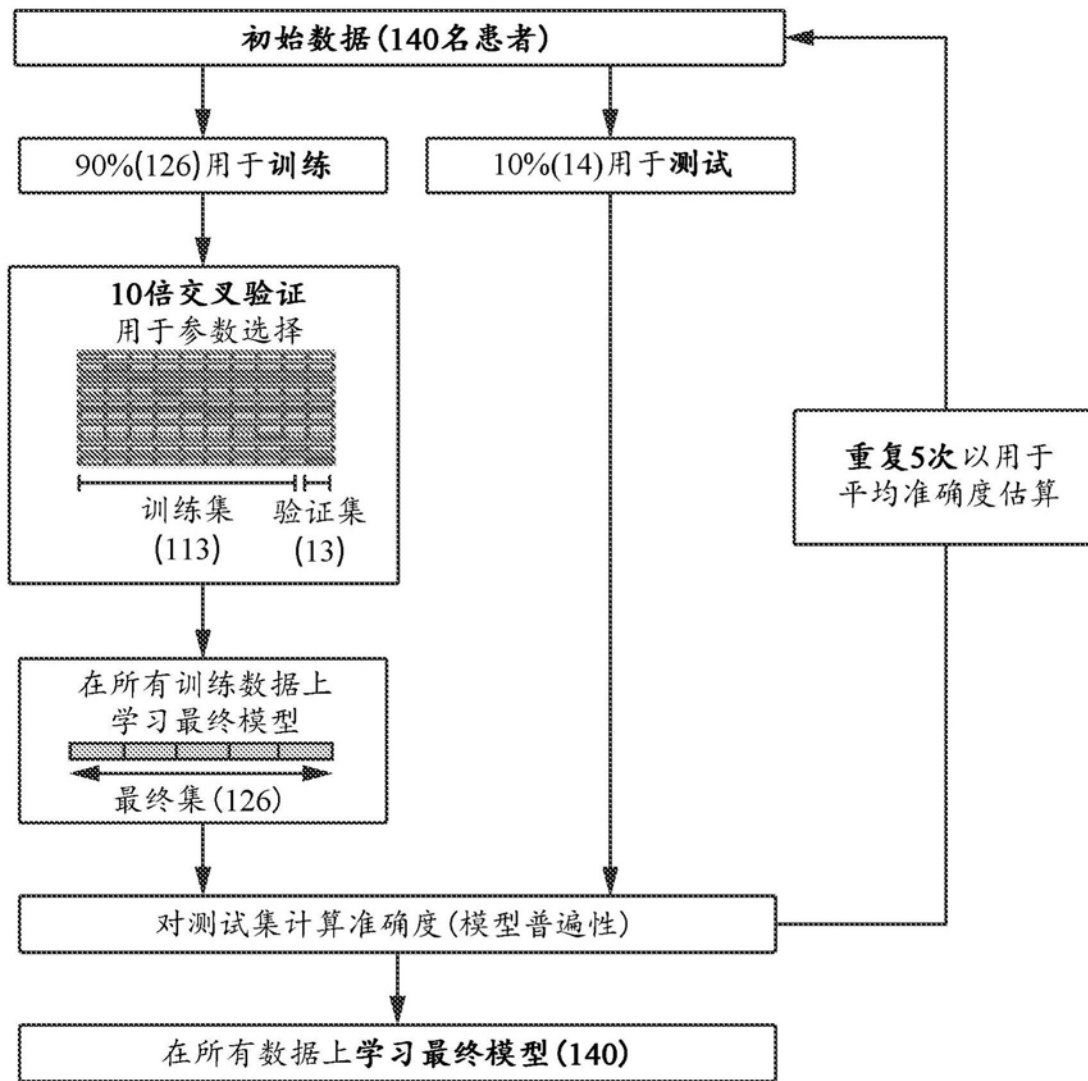


图6C

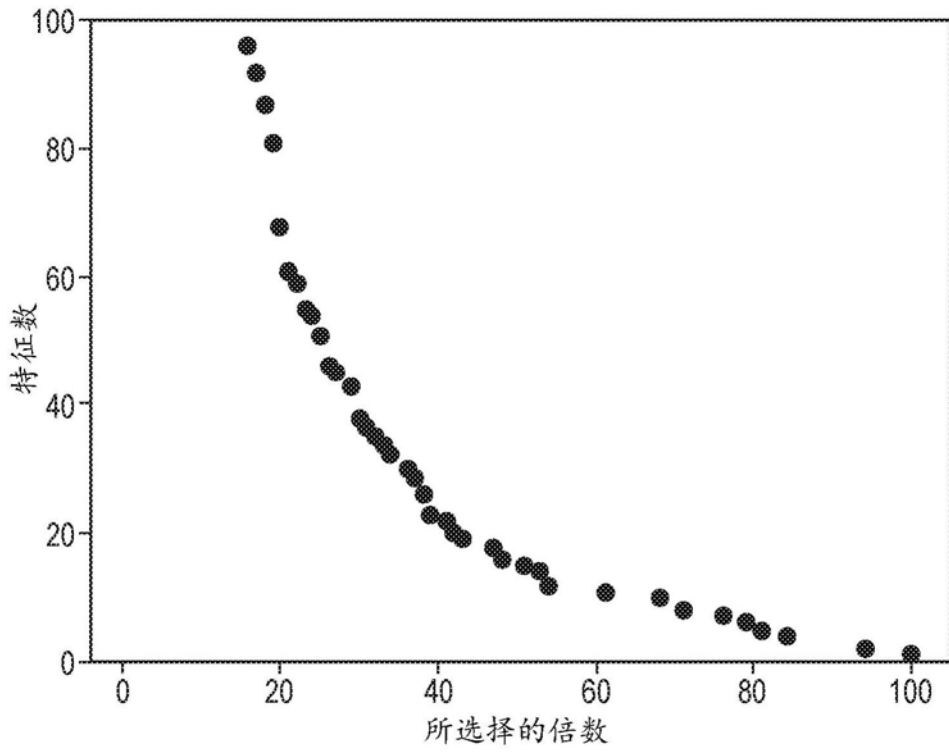


图7A

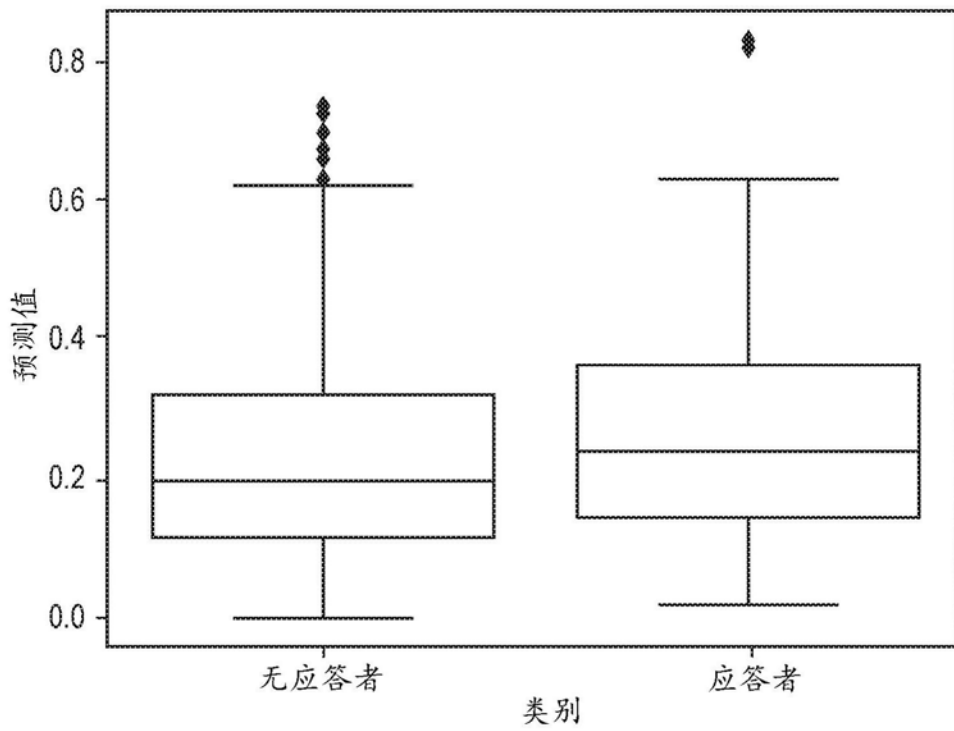


图7B

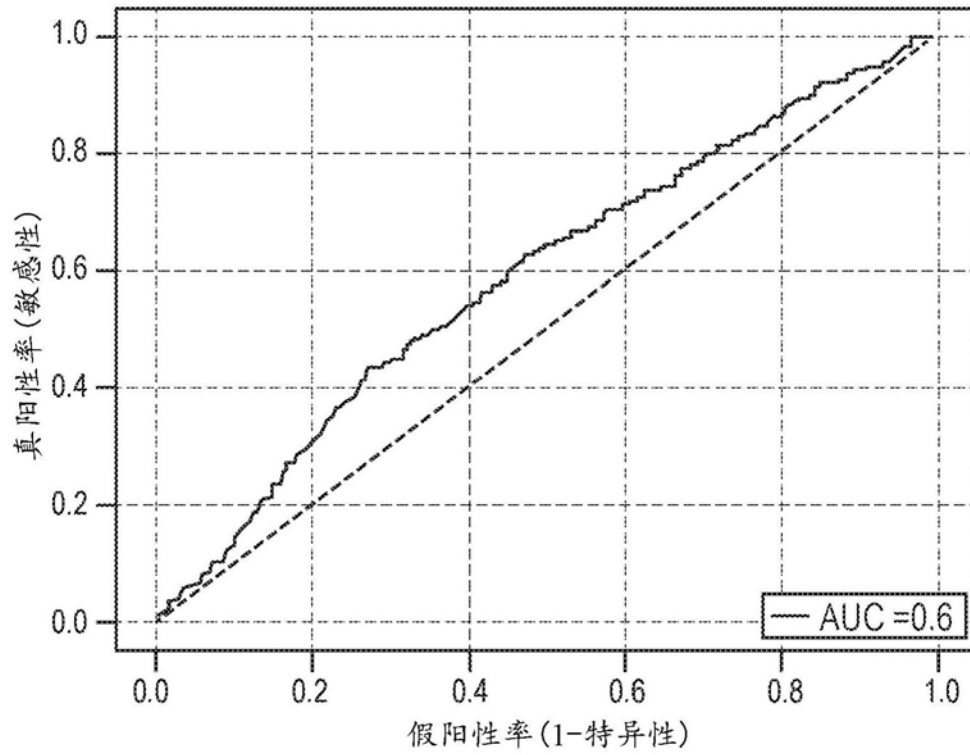


图7C

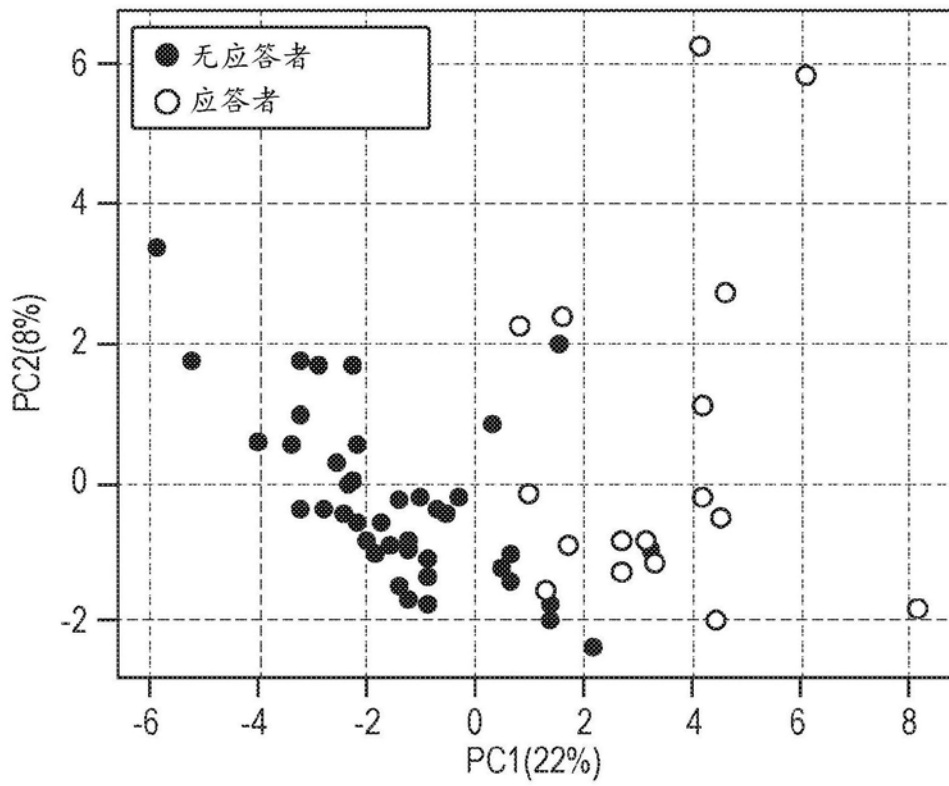


图7D

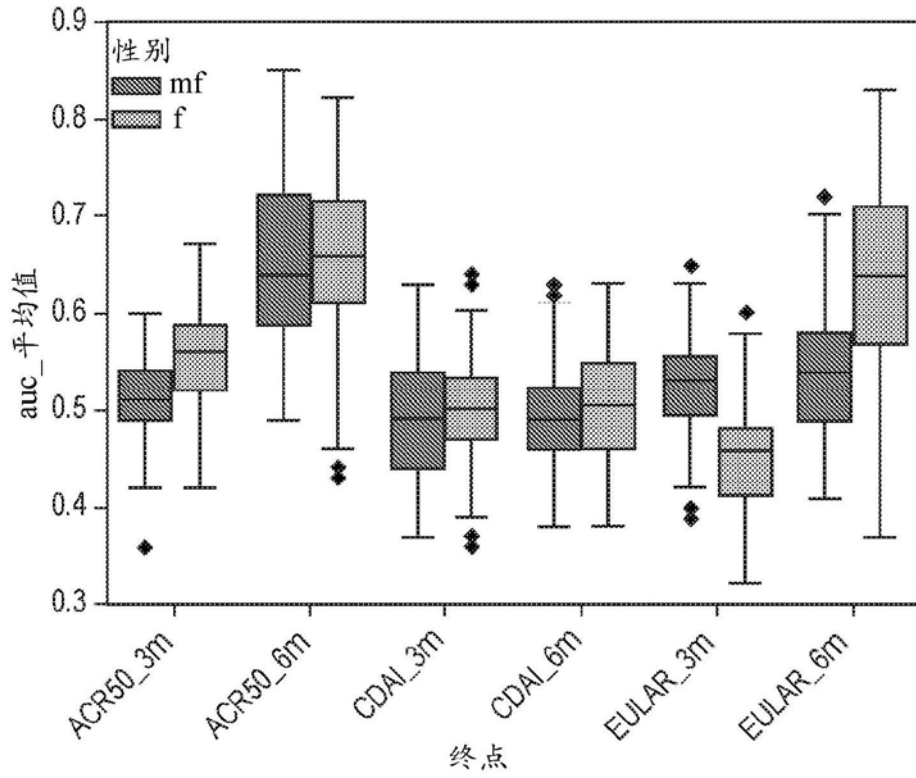


图8A

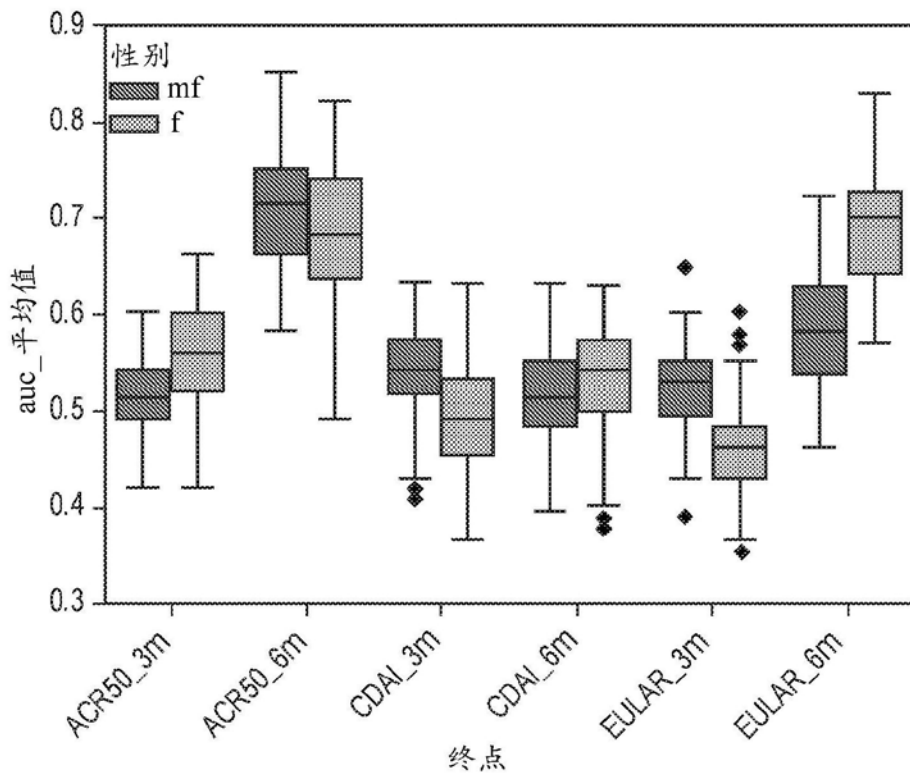


图8B

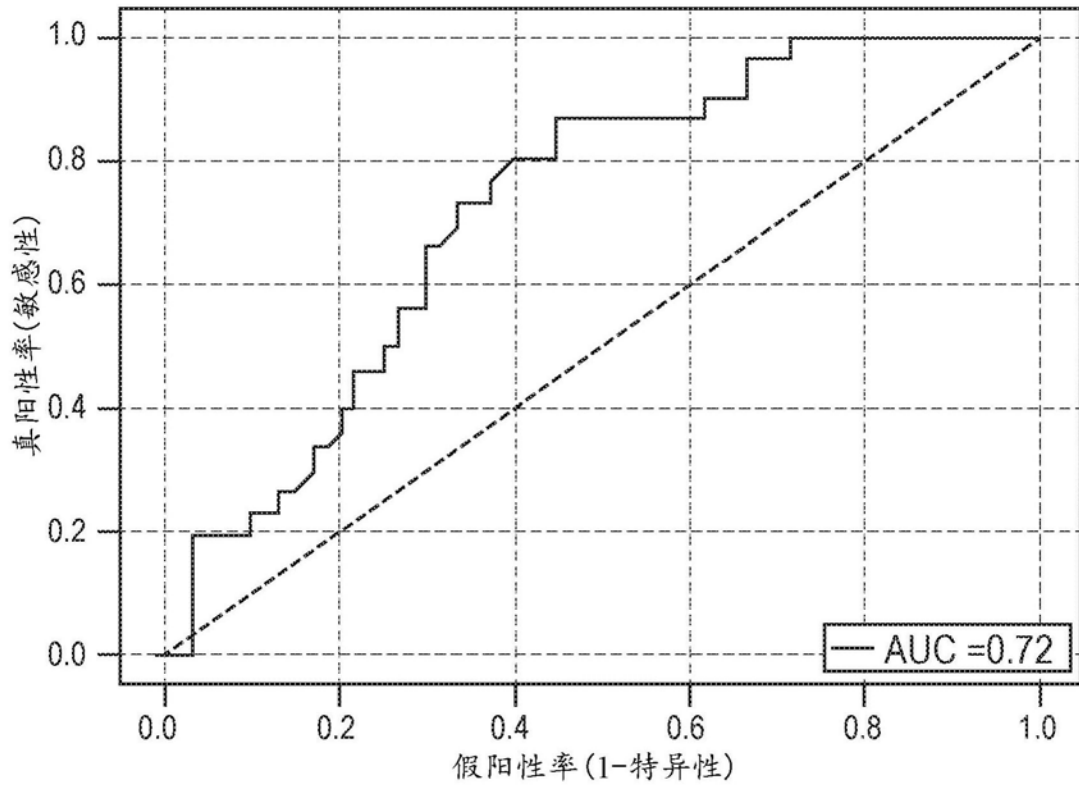


图9A

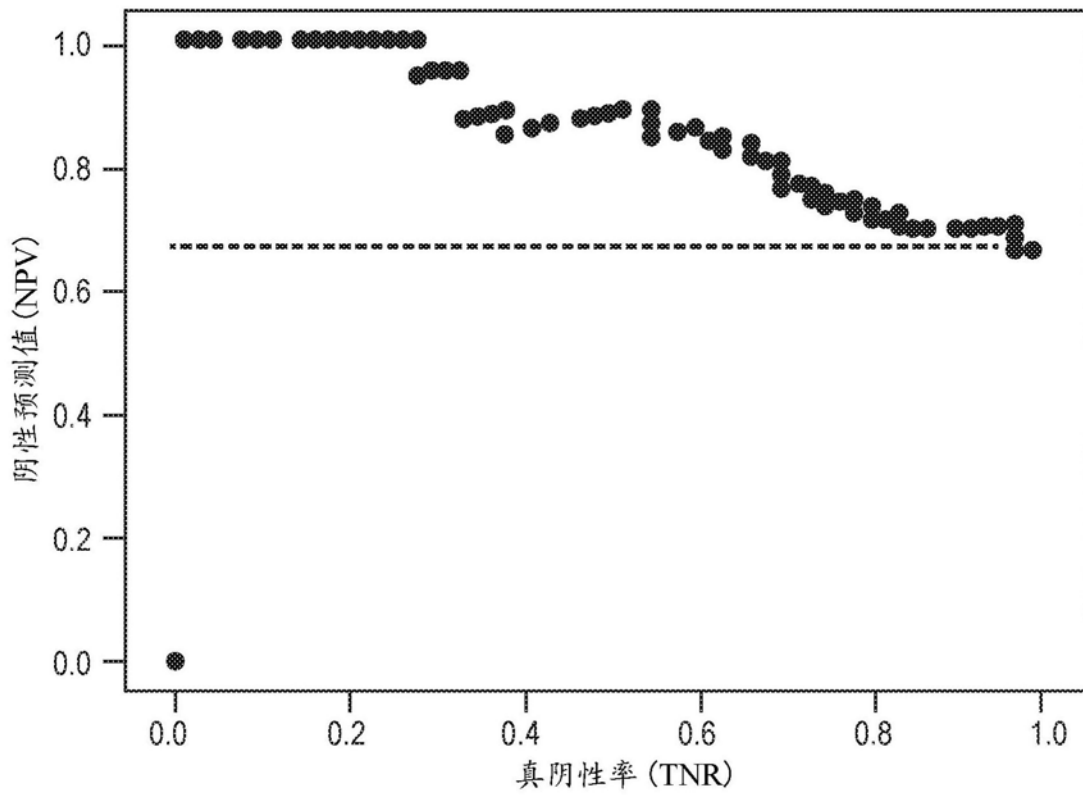


图9B

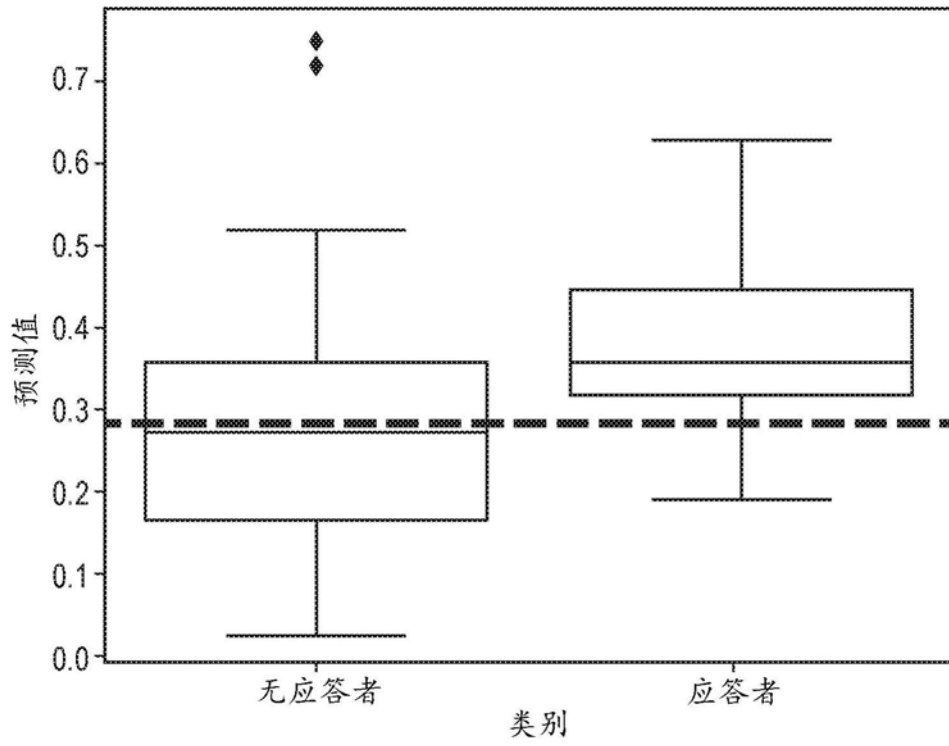


图9C

	PrismRA (+)	PrismRA (-)
真(+)	26	4
真(-)	27	33

图9D

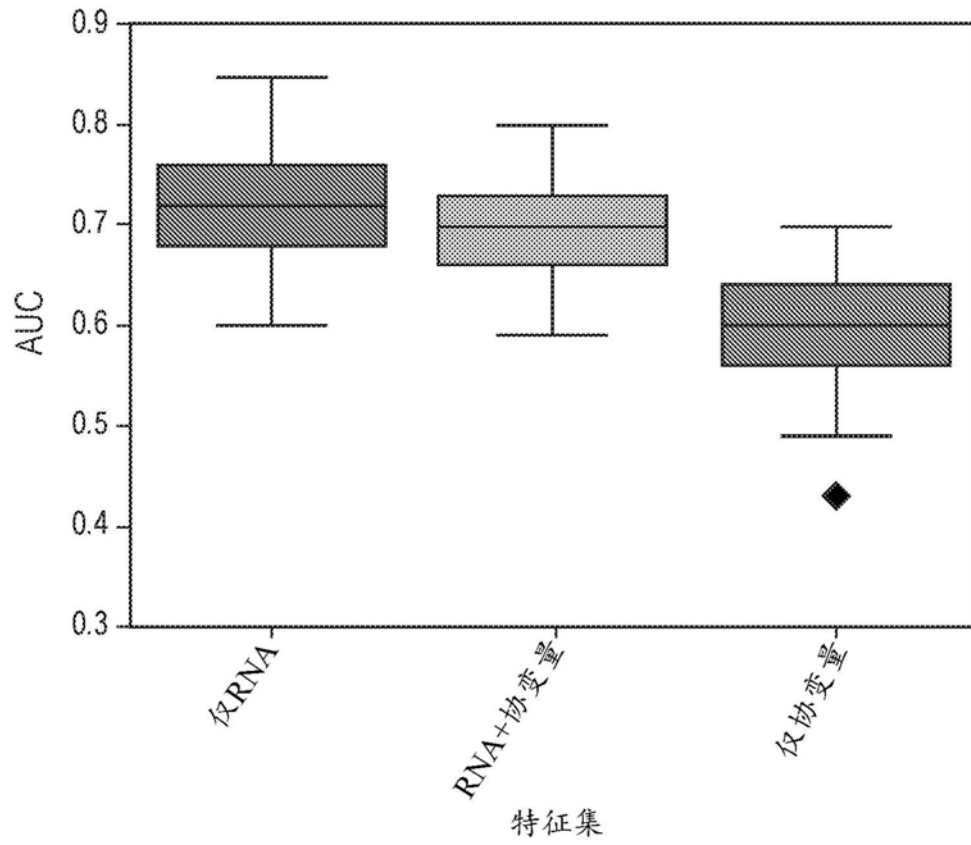


图10A

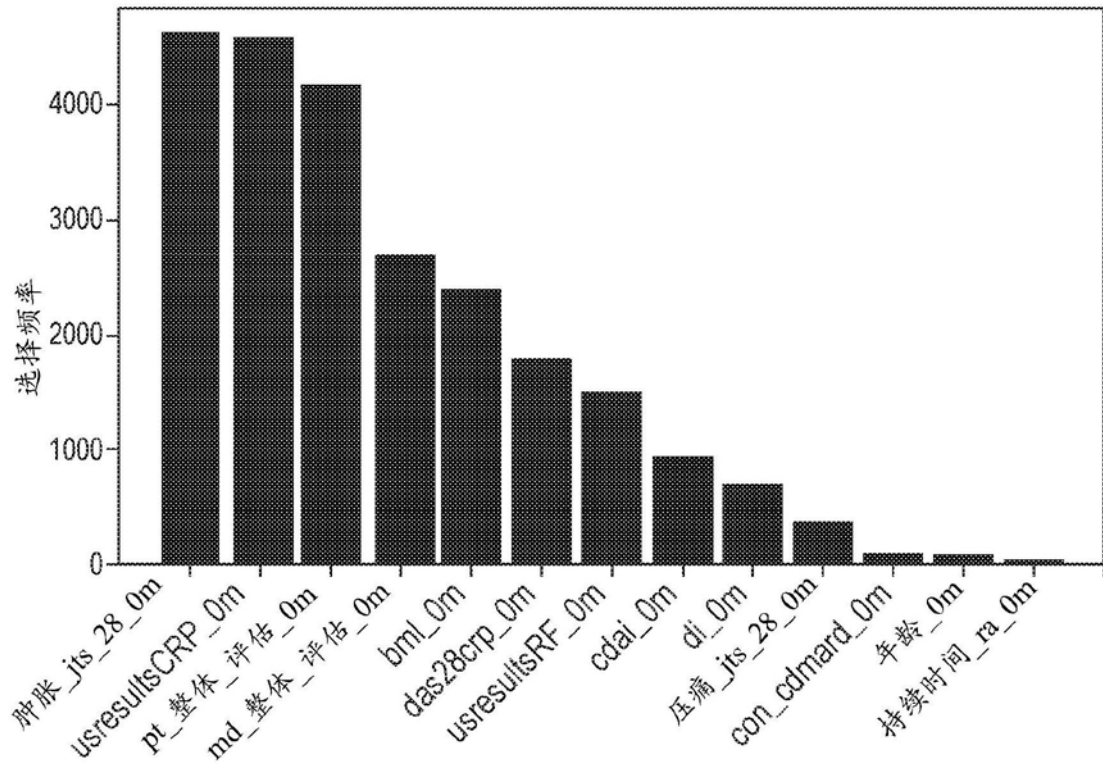


图10B

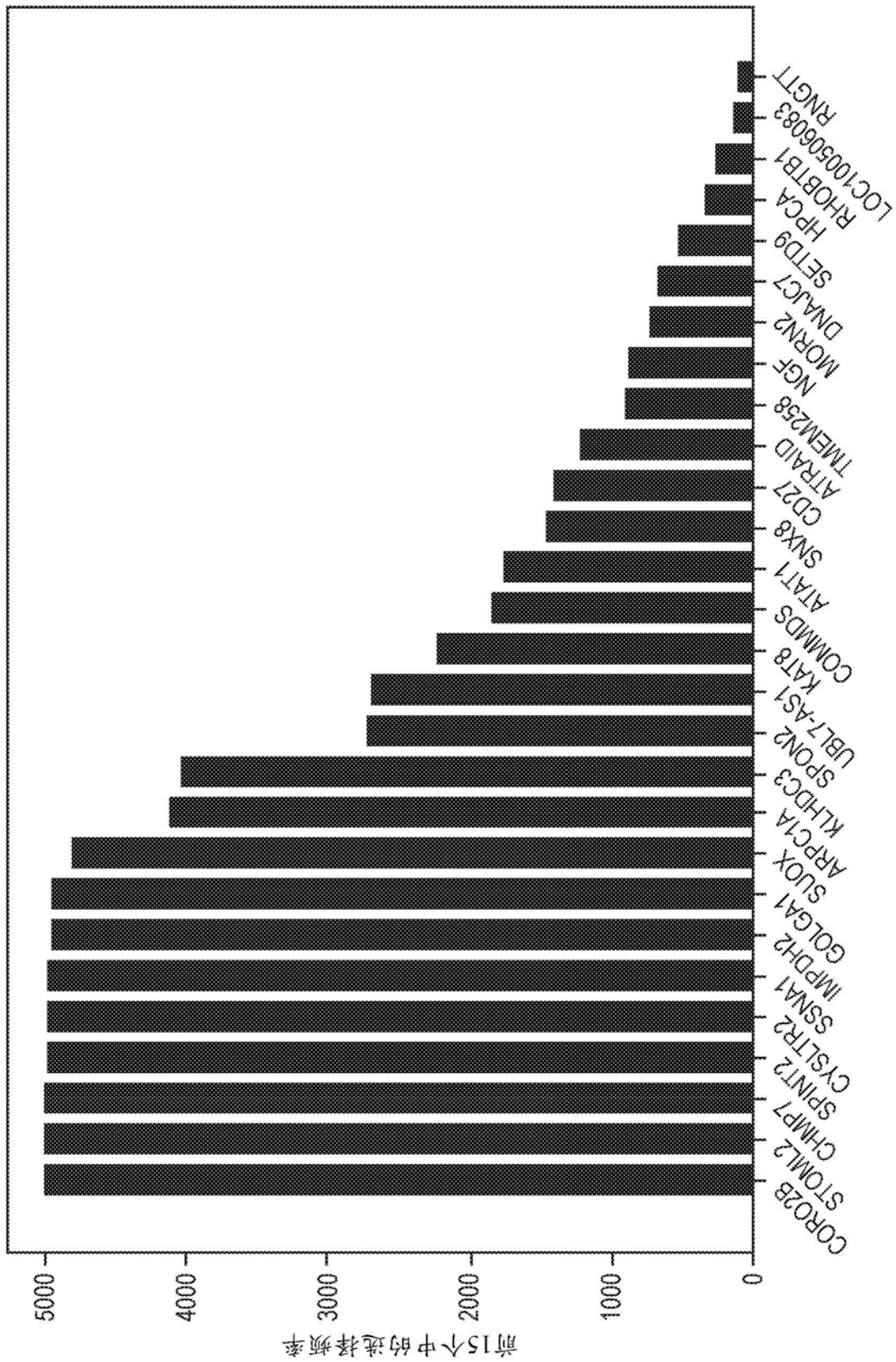


图11

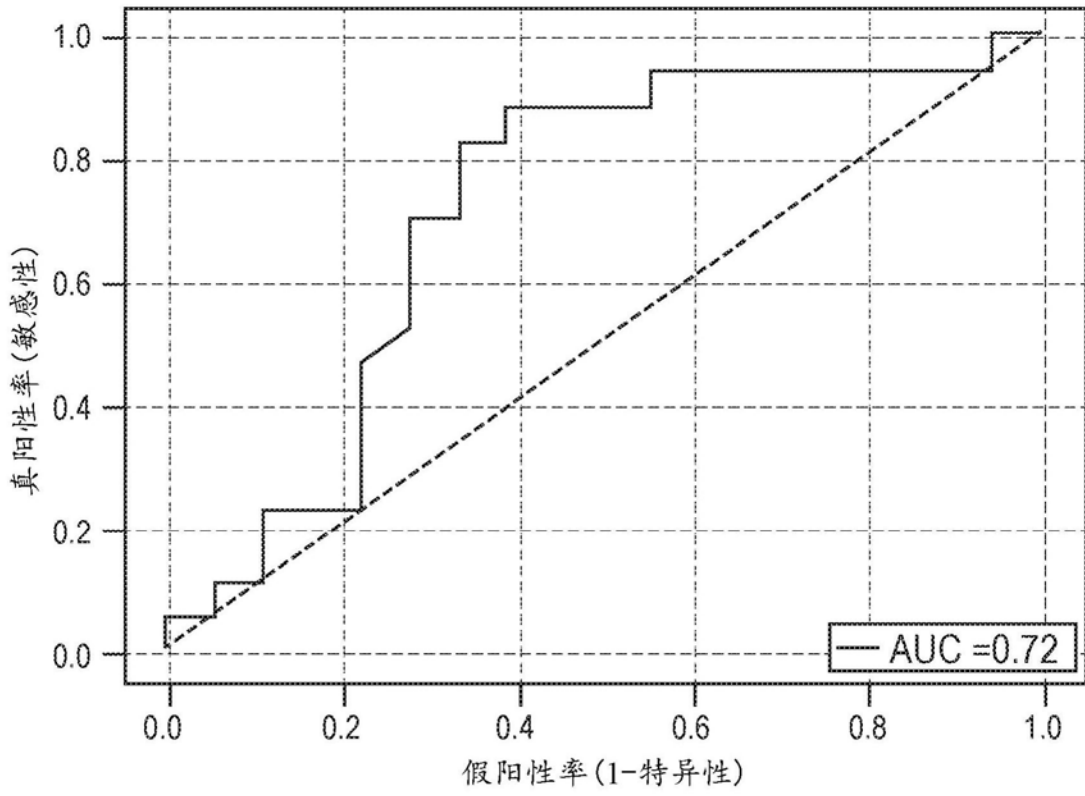


图12A

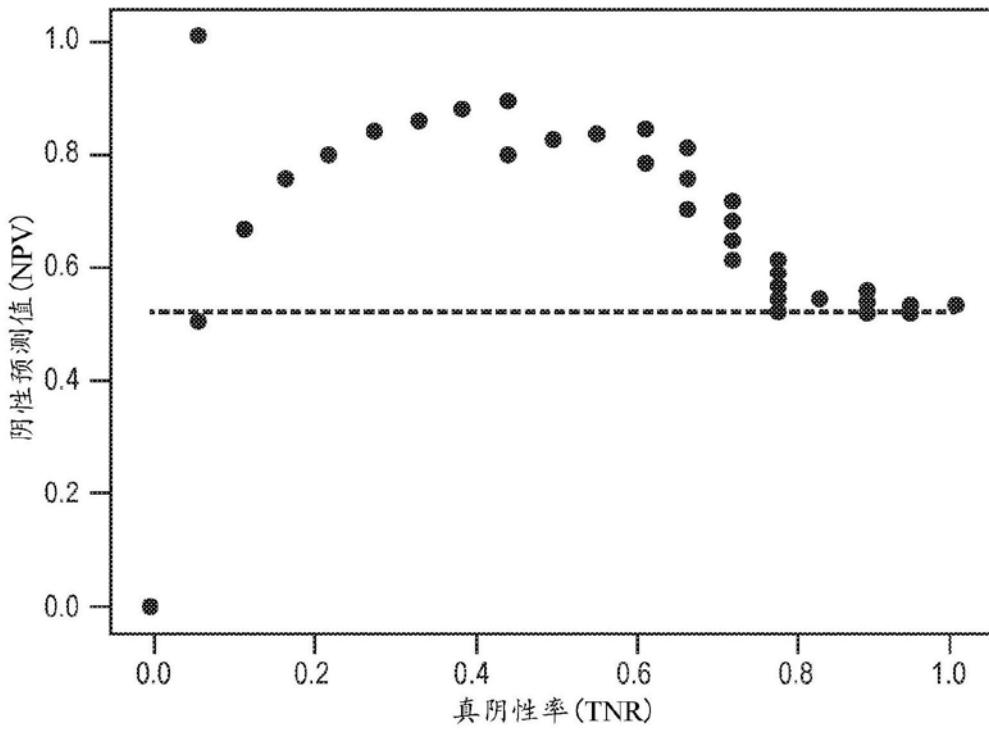


图12B

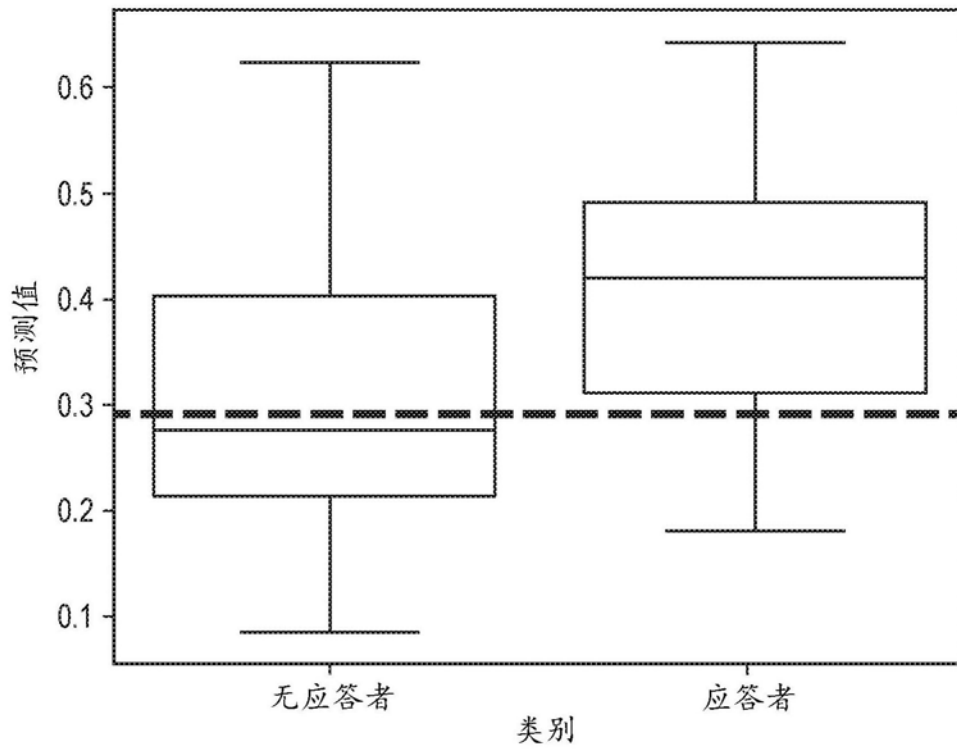


图12C

	PrismRA (+)	PrismRA (-)
真 (+)	15	2
真 (-)	7	11

图12D

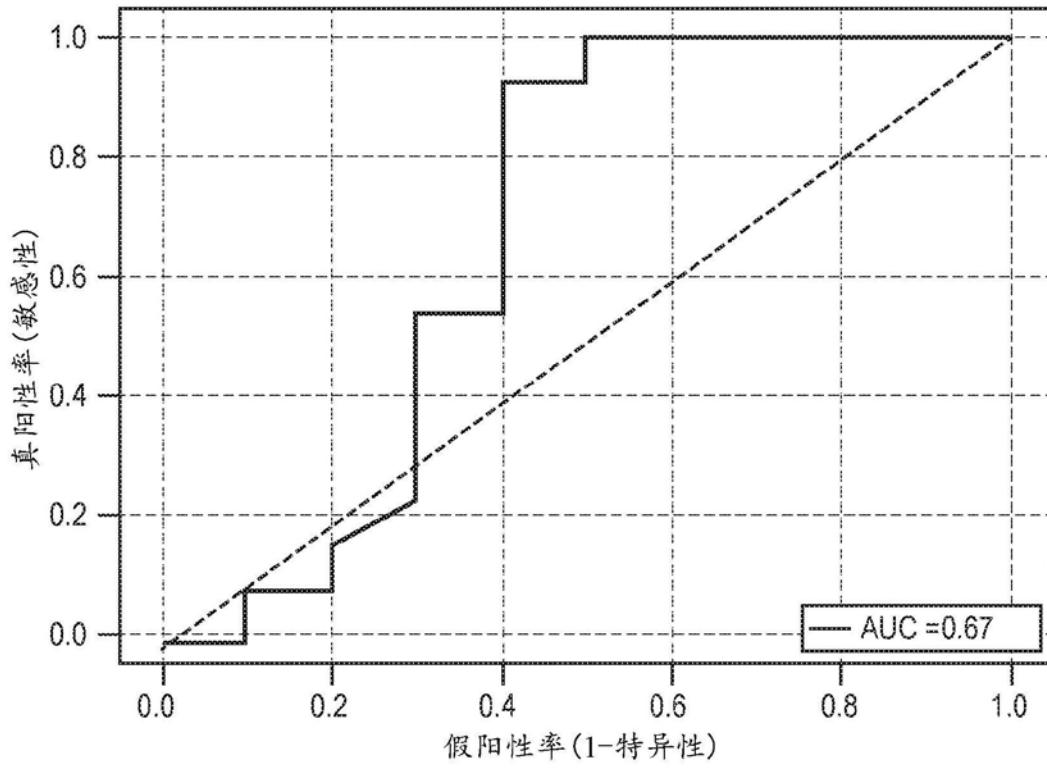


图13A

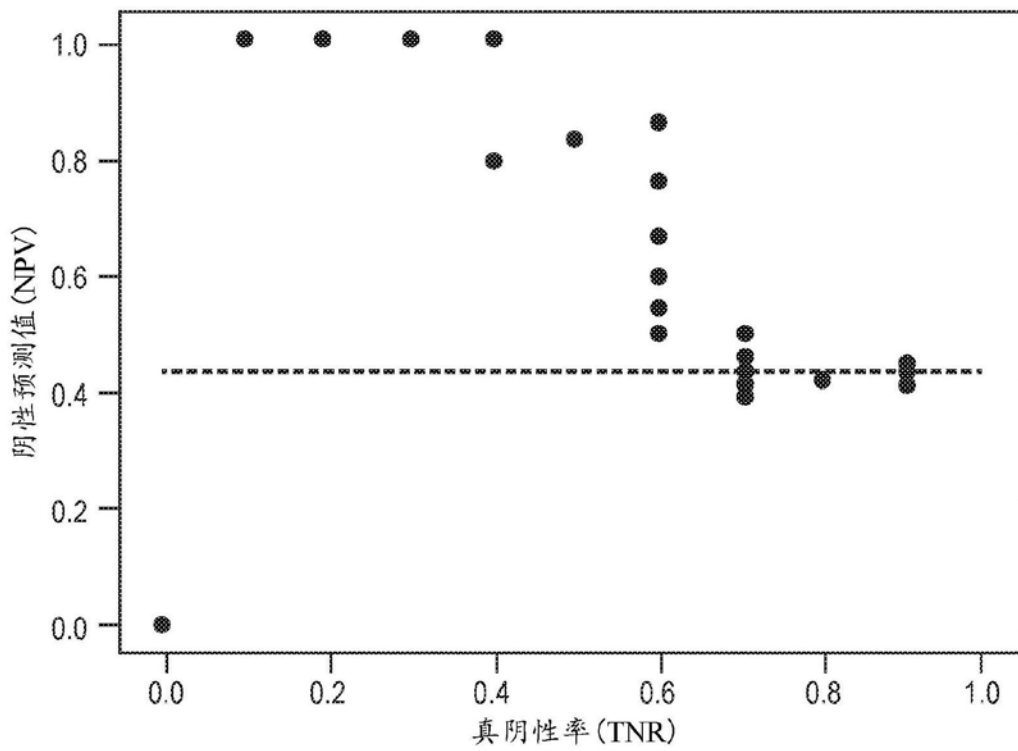


图13B

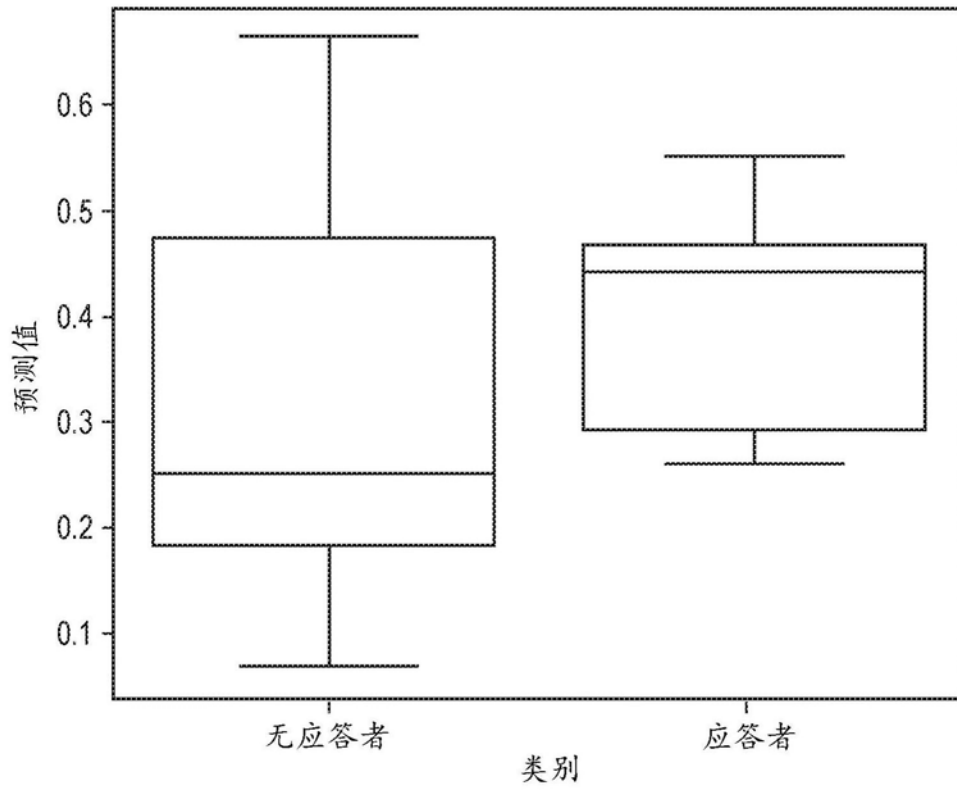


图13C

	PrismRA (+)	PrismRA (-)
真(+)	12	1
真(-)	3	6

图13D

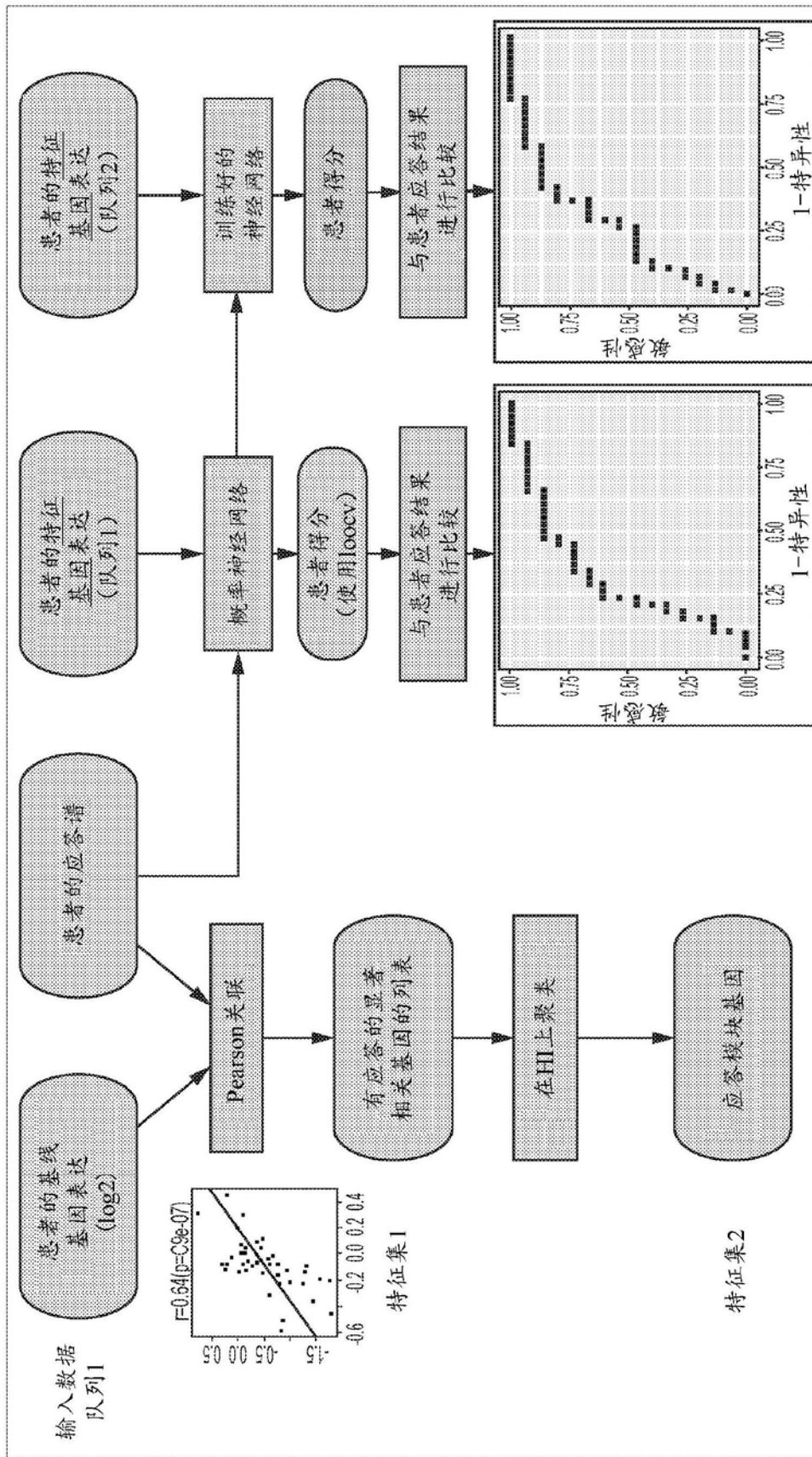


图14

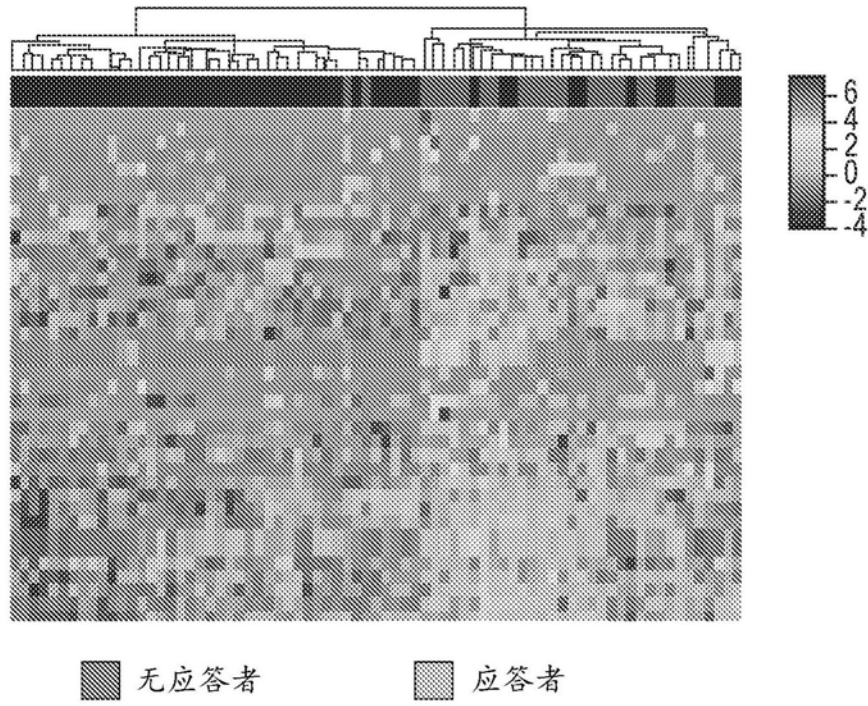


图15

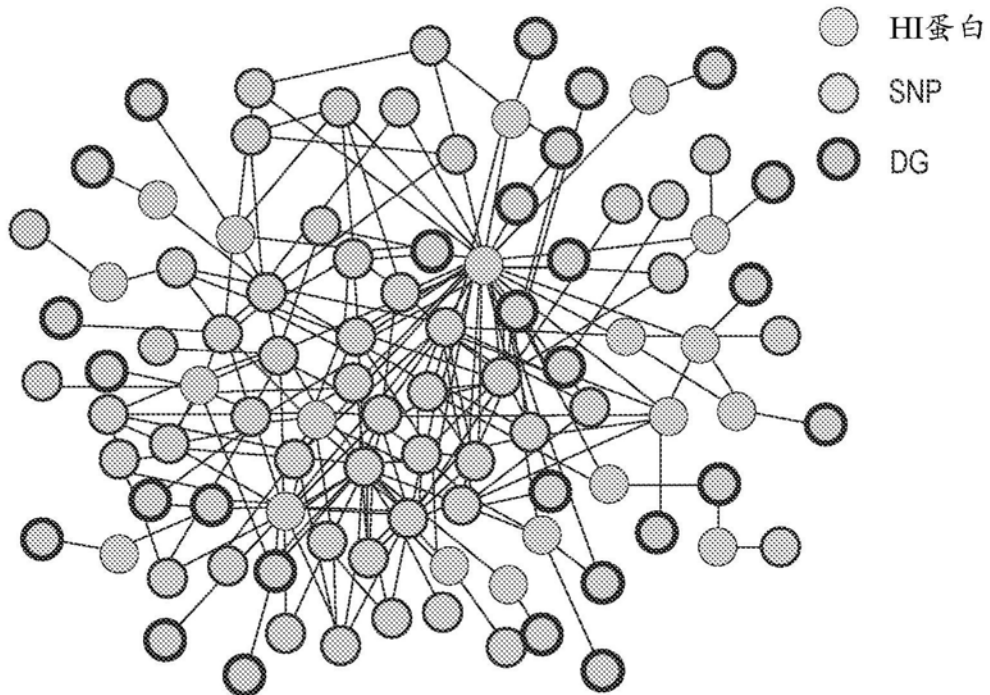


图16A

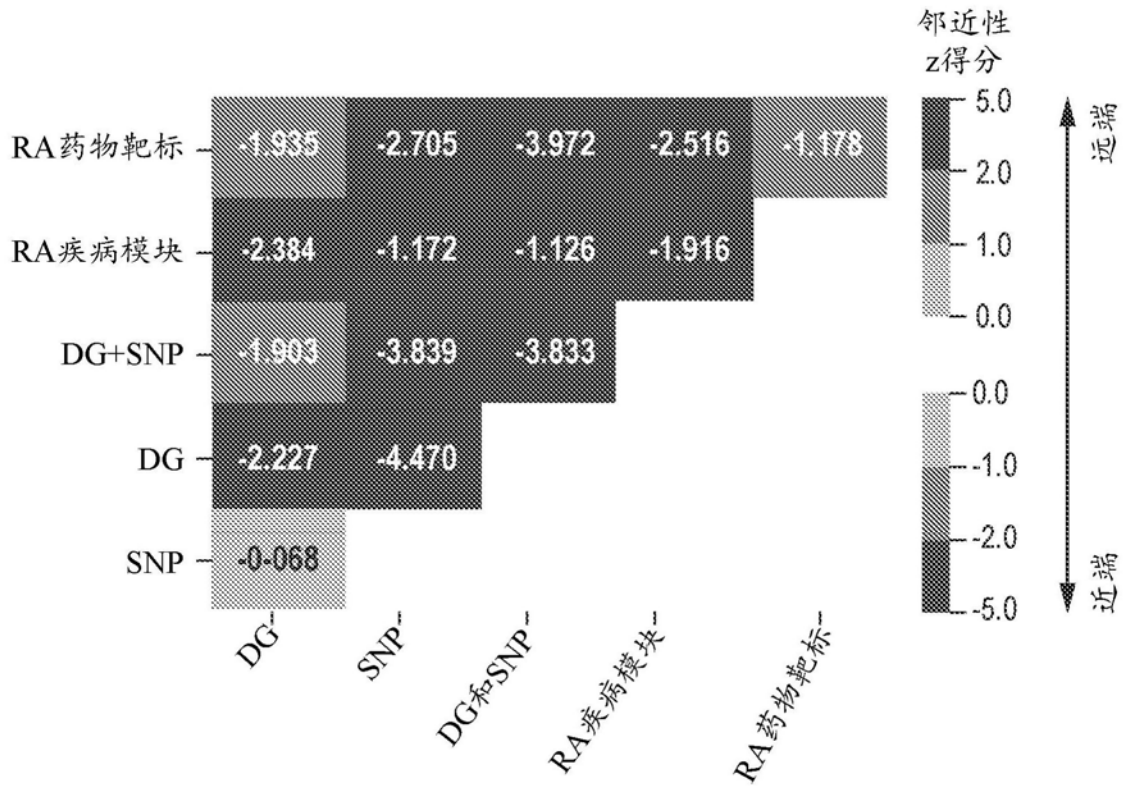


图16B

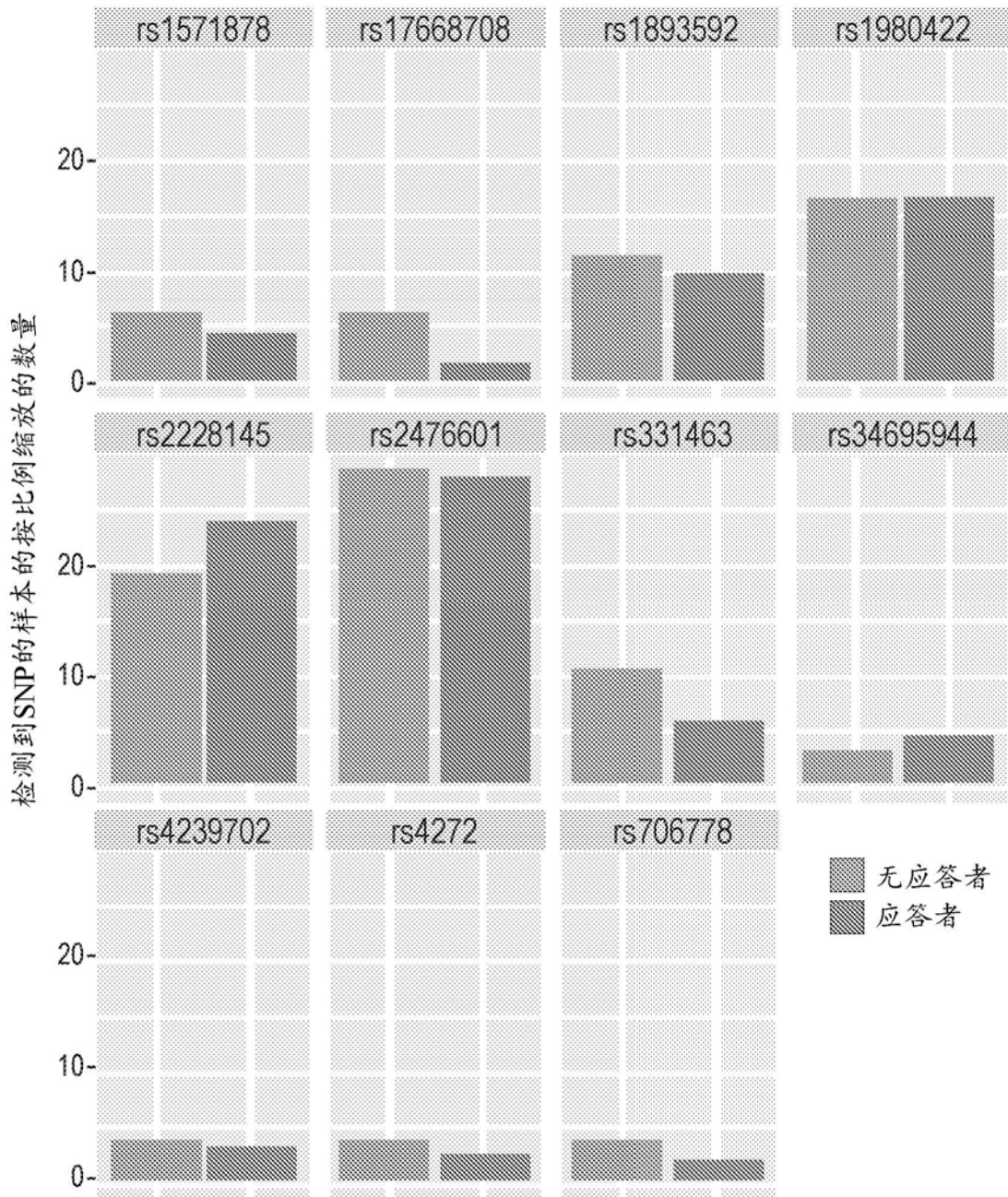


图17

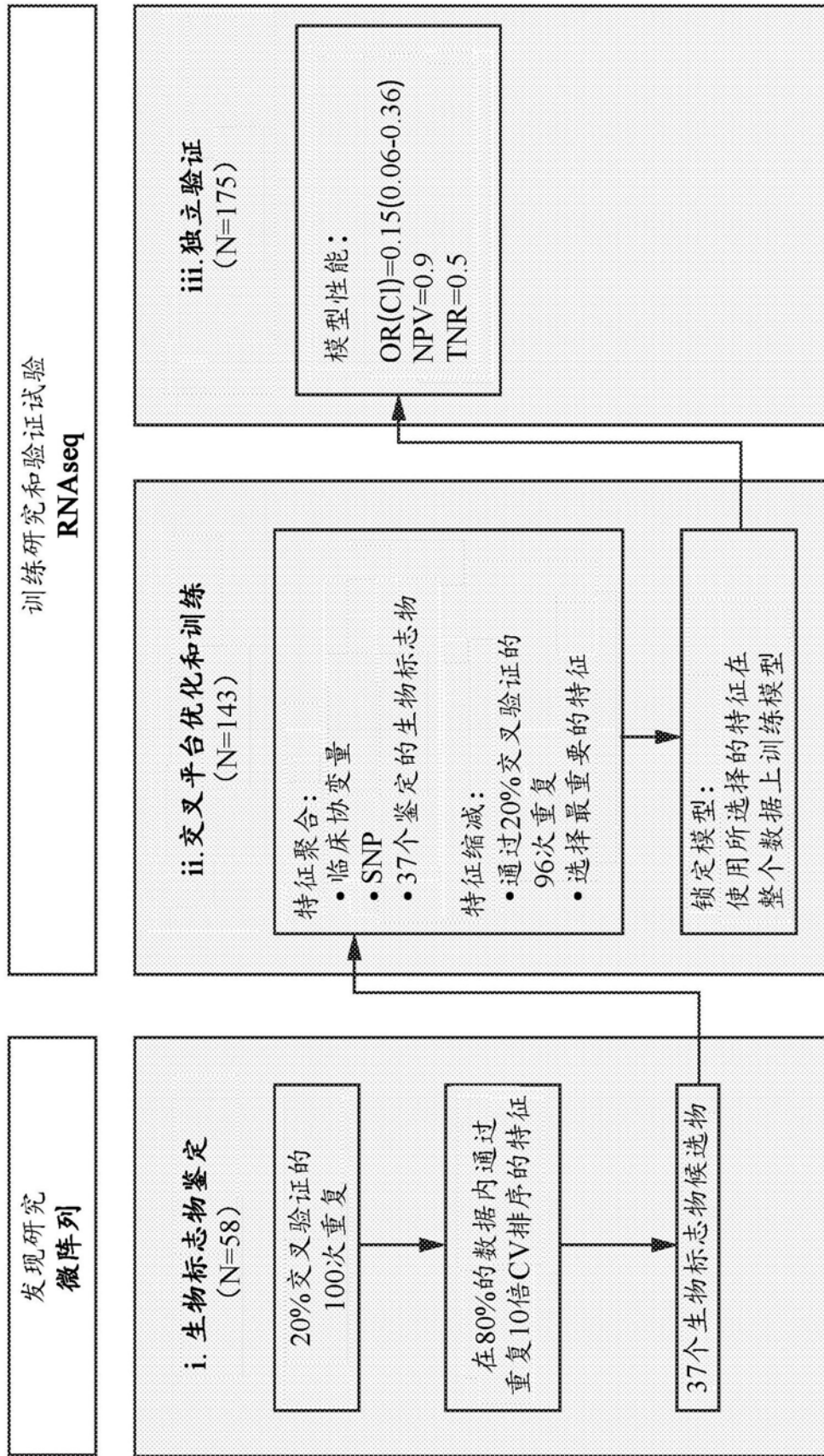


图19