



(12) 发明专利申请

(10) 申请公布号 CN 104685067 A

(43) 申请公布日 2015. 06. 03

(21) 申请号 201380031747. 7

G01N 33/50(2006. 01)

(22) 申请日 2013. 04. 26

G01N 33/53(2006. 01)

(30) 优先权数据

G01N 33/68(2006. 01)

1207297. 1 2012. 04. 26 GB

(85) PCT国际申请进入国家阶段日

2014. 12. 16

(86) PCT国际申请的申请数据

PCT/IB2013/053321 2013. 04. 26

(87) PCT国际申请的公布数据

W02013/160882 EN 2013. 10. 31

(71) 申请人 森扎基因有限责任公司

地址 瑞典本克弗卢斯特兰德

(72) 发明人 M. 林德斯泰特 C. 伯莱拜克

H. 乔汉森 A-S. 阿尔布雷克特

(74) 专利代理机构 北京市柳沈律师事务所

11105

代理人 曹立莉

(51) Int. Cl.

C12Q 1/68(2006. 01)

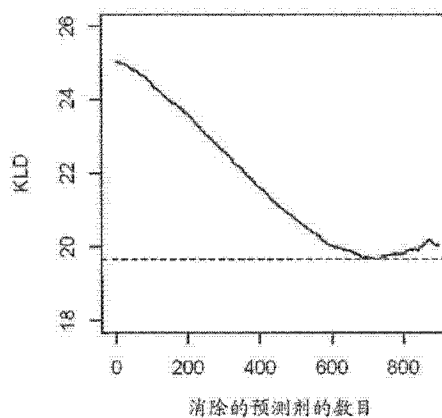
权利要求书5页 说明书58页 附图4页

(54) 发明名称

用于鉴定能够诱导呼吸道致敏的物质的方法
和用于该方法中的阵列和分析试剂盒

(57) 摘要

本发明涉及用于鉴定能够诱导哺乳动物呼吸道致敏的物质的体外方法以及用于所述方法中的阵列和诊断试剂盒。特别地,该方法包括测量表1A、表1B和/或表1C中所列的生物标记物在暴露于测试物质的 MUTZ-3 细胞中的表达。



1. 一种用于鉴定能够诱导哺乳动物呼吸道致敏的物质的方法,其包括以下步骤或由以下步骤组成:

a) 将树突细胞群或树突样细胞群暴露于测试物质;和

b) 测量所述细胞中一种或多种选自表 1A 和 / 或表 1B 中限定的生物标记物的表达,其中步骤 (b) 中测量的细胞中的一种或多种生物标记物的表达表明待测试样品的致敏作用。

2. 根据权利要求 1 的方法,进一步包括:

c) 将分开的树突细胞群或树突样细胞群暴露于一种或多种不是哺乳动物呼吸道致敏剂的阴性对照剂;和

d) 测量细胞中步骤 (b) 中测量的一种或多种生物标记物的表达,

其中在步骤 (d) 中测量的测试样品中的一种或多种生物标记物的存在和 / 或含量不同于步骤 (b) 中测量的对照样品中的一种或多种生物标记物的存在和 / 或含量的情况下,将测试物质鉴定为呼吸道致敏剂。

3. 根据权利要求 1 或 2 的方法,进一步包括:

e) 将分开的树突细胞群或树突样细胞群暴露于一种或多种是哺乳动物呼吸道致敏剂的阳性对照剂;和

f) 测量细胞中步骤 (b) 中测量的一种或多种生物标记物的表达,

其中在步骤 (f) 中测量的测试样品中的一种或多种生物标记物的存在和 / 或含量对应于步骤 (b) 中测量的阳性对照样品中的一种或多种生物标记物的存在和 / 或含量的情况下,将测试物质鉴定为呼吸道致敏剂。

4. 根据前述任一项权利要求的方法,其中步骤 (b) 包括测量一种或多种表 1A 中限定的生物标记物例如至少 2 种表 1A 中限定的生物标记物的表达,或由测量一种或多种表 1A 中限定的生物标记物例如至少 2 种表 1A 中限定的生物标记物的表达组成。

5. 根据前述任一项权利要求的方法,其中步骤 (b) 包括测量 OR5B21 的表达,或由测量 OR5B21 的表达组成。

6. 根据前述任一项权利要求的方法,其中步骤 (b) 包括测量 SLC7A7 的表达,或由测量 SLC7A7 的表达组成。

7. 根据前述任一项权利要求的方法,其中步骤 (b) 包括测量 OR5B21 和 SLC7A7 的表达,或由测量 OR5B21 和 SLC7A7 的表达组成。

8. 根据前述任一项权利要求的方法,其中步骤 (b) 包括测量一种或多种表 1B 中限定的生物标记物例如至少 2、3、4、5、6、7、8、9、10、11、12 或 13 种表 1B 中限定的生物标记物的表达,或由测量一种或多种表 1B 中限定的生物标记物例如至少 2、3、4、5、6、7、8、9、10、11、12 或 13 种表 1B 中限定的生物标记物的表达组成。

9. 根据前述任一项权利要求的方法,其中步骤 (b) 包括测量表 1B 中限定的全部生物标记物的表达,或由测量表 1B 中限定的全部生物标记物的表达组成。

10. 根据前述任一项权利要求的方法,其中步骤 (b) 包括测量一种或多种表 1C 中限定的生物标记物例如 2、3、4、5、6、7、8、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、

75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、114、115、116、117、118、119、120、121、122、123、124、125、126、127、128、129、130、131、132、133、134、135、136、137、138、139、140、141、142、143、144、145、146、147、148、149、150、151、152、153、154、155、156、157、158、159、160、161、162、163、164、165、166、167、168、169、170、171、172、173、174、175、176、177、178、179、180、181、182、183、184、185、186、187、188、189、190、191、192、193、194、195、196、197、198、199、200、201、202、203、204、205、206、207、208、209、210、211、212、213、214、215、216、217、218、219、220、221、222、223、224、225、226、227、228、229、230、231、232、233、234、235、236、237、238、239、240、241、242、243、244、245、246、247、248、249、250、251、252、253、254、255、256、257、258、259、260、261、262、263、264、265、266、267、268、269、270、271、272、273、274、275、276、277、278、279、280、281、282、283、284、285、286 或 287 种表 1C 中限定的生物标记物的表达,或由测量一种或多种表 1C 中限定的生物标记物例如 2、3、4、5、6、7、8、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、114、115、116、117、118、119、120、121、122、123、124、125、126、127、128、129、130、131、132、133、134、135、136、137、138、139、140、141、142、143、144、145、146、147、148、149、150、151、152、153、154、155、156、157、158、159、160、161、162、163、164、165、166、167、168、169、170、171、172、173、174、175、176、177、178、179、180、181、182、183、184、185、186、187、188、189、190、191、192、193、194、195、196、197、198、199、200、201、202、203、204、205、206、207、208、209、210、211、212、213、214、215、216、217、218、219、220、221、222、223、224、225、226、227、228、229、230、231、232、233、234、235、236、237、238、239、240、241、242、243、244、245、246、247、248、249、250、251、252、253、254、255、256、257、258、259、260、261、262、263、264、265、266、267、268、269、270、271、272、273、274、275、276、277、278、279、280、281、282、283、284、285、286 或 287 种表 1C 中限定的生物标记物的表达组成。

11. 根据前述任一项权利要求的方法,其中步骤 (b) 包括测量表 1C 中限定的全部生物标记物的表达,或由测量表 1C 中限定的全部生物标记物的表达组成。

12. 根据前述任一项权利要求的方法,其中步骤 (b) 包括测量表 1 中限定的全部生物标记物的表达,或由测量表 1 中限定的全部生物标记物的表达组成。

13. 根据前述任一项权利要求的方法,其中步骤 (b) 包括测量编码一种或多种生物标记物的核酸分子的表达。

14. 根据权利要求 13 的方法,其中核酸分子是 cDNA 分子或 mRNA 分子。

15. 根据权利要求 13 的方法,其中核酸分子是 mRNA 分子。

16. 根据权利要求 13 的方法,其中核酸分子是 cDNA 分子。

17. 根据权利要求 13 至 16 任一项的方法,其中使用选自 Southern 杂交、Northern 杂交、聚合酶链式反应 (PCR)、逆转录酶 PCR (RT-PCR)、定量实时 PCR (qRT-PCR)、纳米阵列、微阵列、大阵列、放射自显影术和原位杂交的方法来进行步骤 (b) 中测量一种或多种生物标

记物的表达。

18. 根据权利要求 13 至 17 任一项的方法,其中使用 DNA 微阵列确定步骤 (b) 中测量一种或多种生物标记物的表达。

19. 根据前述任一项权利要求的方法,其中使用一种或多种结合部分进行步骤 (b) 中测量一种或多种生物标记物的表达,每种所述结合部分能够选择性地结合编码表 1 中鉴定的生物标记物之一的核酸分子。

20. 根据权利要求 19 的方法,其中一种或多种结合部分各自包括核酸分子或由核酸分子组成。

21. 根据权利要求 20 的方法,其中一种或多种结合部分各自包括 DNA、RNA、PNA、LNA、GNA、TNA 或 PMO 或由 DNA、RNA、PNA、LNA、GNA、TNA 或 PMO 组成。

22. 根据权利要求 19 或 20 的方法,其中一种或多种结合部分各自包括 DNA 或由 DNA 组成。

23. 根据权利要求 20 至 22 任一项的方法,其中一种或多种结合部分的长度是 5 至 100 个核苷酸。

24. 根据权利要求 20 至 23 任一项的方法,其中一种或多种核酸分子的长度是 15 至 35 个核苷酸。

25. 根据权利要求 20 至 24 任一项的方法,其中结合部分包括可检测部分。

26. 根据权利要求 25 的方法,其中可检测部分选自荧光部分;发光部分;化学发光部分;放射性部分(例如,放射性原子);或酶部分。

27. 根据权利要求 26 的方法,其中可检测部分包括放射性原子或由放射性原子组成。

28. 根据权利要求 27 的方法,其中放射性原子选自钨 -99m、碘 -123、碘 -125、碘 -131、钼 -111、氟 -19、碳 -13、氮 -15、氧 -17、磷 -32、硫 -35、氘、氚、铯 -136、铯 -137 和钷 -147。

29. 根据权利要求 26 的方法,其中结合部分的可检测部分是荧光部分。

30. 根据权利要求 1 至 21 任一项的方法,其中步骤 (b) 包括测量步骤 (b) 中限定的一种或多种生物标记物的蛋白质的表达,或由测量步骤 (b) 中限定的一种或多种生物标记物的蛋白质的表达组成。

31. 根据权利要求 30 的方法,其中使用一种或多种结合部分进行步骤 (b) 中测量一种或多种生物标记物的表达,每种所述结合部分能够选择性地结合表 1 中鉴定的生物标记物之一。

32. 根据权利要求 31 的方法,其中一种或多种结合部分包括抗体或其抗原结合片段,或由抗体或其抗原结合片段组成。

33. 根据权利要求 32 的方法,其中抗体或其片段是单克隆抗体或其片段。

34. 根据权利要求 32 或 33 的方法,其中抗体或其抗原结合片段选自完整抗体、Fv 片段(例如,单链 Fv 和二硫化物键合的 Fv)、Fab- 样片段(例如, Fab 片段、Fab' 片段和 F(ab)₂ 片段)、单个可变域(例如, V_H 和 V_L 域)和域抗体(dAb,包括单一和双重形式[即, dAb- 连接物-dAb])。

35. 根据权利要求 34 的方法,其中抗体或抗原结合片段是单链 Fv(scFv)。

36. 根据权利要求 31 的方法,其中一种或多种结合部分包括抗体样结合剂例如亲和体或适体,或由抗体样结合剂例如亲和体或适体组成。

37. 根据权利要求 31 至 36 任一项的方法,其中一种或多种结合部分包括可检测部分。
38. 根据权利要求 37 的方法,其中可检测部分选自荧光部分、发光部分、化学发光部分、放射性部分和酶部分。
39. 根据前述任一项权利要求的方法,其中使用阵列进行步骤 (b)。
40. 根据权利要求 39 的方法,其中阵列是基于珠子的阵列。
41. 根据权利要求 40 的方法,其中阵列是基于表面的阵列。
42. 根据权利要求 39 至 41 任一项的方法,其中阵列选自:大阵列;微阵列;纳米阵列。
43. 用于根据前述任一项权利要求的方法中的阵列,该阵列包括一种或多种权利要求 19 至 29 和 31 至 38 任一项中限定的第一种结合剂。
44. 根据权利要求 43 的阵列,包括能够共同地结合表 1 中限定的全部生物标记物的结合剂。
45. 根据权利要求 43 或 44 的阵列,其中第一种结合剂是固定的。
46. 根据前述任一项权利要求的方法,用于鉴定能够诱导呼吸道超敏反应的物质。
47. 根据前述任一项权利要求的方法,其中超敏反应是体液超敏反应。
48. 根据权利要求 46 或 47 的方法,其中超敏反应是 I 型超敏反应。
49. 根据前述任一项权利要求的方法,用于鉴定能够诱导呼吸道过敏的物质。
50. 根据前述任一项权利要求的方法,其中树突细胞群或树突样细胞群是树突样细胞群。
51. 根据权利要求 50 的方法,其中树突样细胞是骨髓树突样细胞。
52. 根据权利要求 51 的方法,其中骨髓树突样细胞源自骨髓树突细胞。
53. 根据权利要求 52 的方法,其中源自骨髓树突细胞的细胞是骨髓性白血病衍生的细胞。
54. 根据权利要求 53 的方法,其中骨髓性白血病衍生的细胞选自 KG-1、THP-1、U-937、HL-60、Monomac-6、AML-193 和 MUTZ-3。
55. 根据前述任一项权利要求的方法,其中树突样细胞是 MUTZ-3 细胞。
56. 根据前述任一项权利要求的方法,其中步骤 (c) 中提供的一种或多种阴性对照剂选自 1-丁醇、4-氨基苯甲酸、氯苯、二甲基甲酰胺、乙基香草醛、异丙醇、水杨酸甲酯、丙二醇、高锰酸钾、吐温 80™(聚氧乙烯 (20) 山梨聚糖单油酸酯)和硫酸锌。
57. 根据权利要求 56 的方法,其中提供至少 2 种非致敏剂,例如,至少 3、4、5、6、7、8、9、10 或至少 11 种对照非致敏剂。
58. 根据前述任一项权利要求的方法,其中步骤 (e) 中提供的一种或多种阳性对照剂包括一种或多种选自六氯铂酸铵、过硫酸铵、戊二醛、六亚甲基二异氰酸酯、马来酸酐、亚甲基联苯酚二异氰酸酯、邻苯二甲酸酐、甲苯二异氰酸酯和偏苯三酸酐的物质,或由一种或多种选自六氯铂酸铵、过硫酸铵、戊二醛、六亚甲基二异氰酸酯、马来酸酐、亚甲基联苯酚二异氰酸酯、邻苯二甲酸酐、甲苯二异氰酸酯和偏苯三酸酐的物质组成。
59. 根据权利要求 58 的方法,其中提供至少 2 种对照致敏剂,例如,至少 2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19 或至少 20 种对照致敏剂。
60. 根据前述任一项权利要求的方法,其中该方法表示待测试样品的致敏效力。
61. 用于根据前述任一项权利要求的方法中的阵列,所述阵列包括一种或多种如权利

要求 19 至 29 和 31-38 任一项中限定的结合部分。

62. 根据权利要求 61 的阵列,其中结合部分能够结合表 1A 中限定的全部生物标记物。

63. 根据权利要求 61 或 62 的阵列,其中结合部分能够结合表 1B 中限定的全部生物标记物。

64. 根据权利要求 61、62 或 63 的阵列,其中结合部分能够结合表 1C 中限定的全部生物标记物。

65. 根据权利要求 61 至 64 任一项的阵列,其中结合部分能够结合表 1 中限定的全部生物标记物。

66. 根据权利要求 61 至 64 任一项的阵列,其中结合部分是固定的。

67. 选自表 1 中的两种或多种生物标记物组合用于鉴定呼吸道超敏反应致敏剂的用途。

68. 根据权利要求 67 的用途,其中表 1 中限定的全部生物标记物共同地用于鉴定超敏反应致敏剂。

69. 用于根据权利要求 1 至 60 任一项的方法中的分析试剂盒,其包括:

A) 根据权利要求 61 至 66 任一项的阵列;和

B) 用于进行根据权利要求 1 至 60 任一项中限定的方法的说明书(任选)。

70. 根据权利要求 69 的分析试剂盒,进一步包括一种或多种对照样品。

71. 根据权利要求 69 的分析试剂盒,包括一种或多种非致敏剂。

72. 根据权利要求 69、70 或 71 的分析试剂盒,包括一种或多种致敏剂。

73. 基本上如本文中所述的方法或用途。

74. 基本上如本文中所述的阵列或试剂盒。

用于鉴定能够诱导呼吸道致敏的物质的方法和用于该方法中的阵列和分析试剂盒

发明领域

[0001] 本发明涉及用于鉴定能够诱导呼吸道致敏 (respiratory sensitization) 的物质 (agent) 的方法以及用于所述方法中的阵列和分析试剂盒。

背景技术

[0002] 通常,将过敏定义为在对其他方面无害的抗原产生免疫应答后呈现的不良状态。它是称为超敏反应的一类结果中的成员,将超敏反应定义为导致组织损伤的有害免疫应答 (Janeway, C., Travers, P., Hunt, S., Walport, M., 1997. Allergy and hypersensitivity. ImmunoBiology: The Immune System in Health and Disease. Garland Publishing, New York)。所导致的工业毒理学家特别关注的病症包括呼吸道过敏和过敏性接触性皮炎 (ACD)。呼吸道过敏是上和 / 或下呼吸道对异生物体 (xenobiotic) 的超敏反应。这种超敏反应是即时的,具有在异生物体暴露后数分钟至数小时内产生的临床特征,并且可以包括喘息、气促、胸闷、支气管收缩和 / 或鼻充血。在极端情况中,反应可以引发低血压和致命的过敏反应。在一般的人群中,呼吸道过敏最常见是由环境蛋白质诱导的,所述环境蛋白质包括花粉、尘螨排泄物和动物毛屑。然而,在职业环境中,呼吸道过敏可以由工业化合物引起,所述工业化合物包括高分子量 (HMW) 化合物,如蛋白质去污剂,以及低分子量 (LMW) 化学品。由于它们的小尺寸,LMW 化学变应原作为半抗原,其首先与蛋白质反应,形成复合物,然后其能够启动免疫应答。

[0003] 呼吸道对 HMW 和 LMW 化合物的过敏的发展可以归因于职业性哮喘的产生,职业性哮喘的特征在于由于归因于特定工作环境的诱因和条件引起的可变气流限制和 / 或非特异性的支气管超反应性 (Chan-Yeung, M., Malo, J.L., 1994. Aetiological agents in occupational asthma. Eur. Respir. J. 7, 346-371; Karol, M.H., 1994. Animal models of occupational asthma. Eur. Respir. J. 7, 555-568)。重要的是注意到除了这种免疫病因学之外,非免疫原性物质,如刺激物,也在职业性哮喘的发展中起着重要作用。在许多情况中,共同暴露于变应原和刺激物引起了病症。临床研究已经表明了高达 20% 的成人发作哮喘是由职业因素引起的,并且这些病例中的 90% 涉及免疫学机理 (Mapp, C.E., 2005. Genetics and the occupational environment. Curr. Opin. Allergy Clin. Immunol. 5, 113-118)。此外,职业性哮喘是发达国家中最普遍的职业性肺病。因此,具有作为呼吸道变应原的潜能的化合物的鉴定和表征对于工业毒理学家而言是一个重要的研究领域。

[0004] 不是所有引起特定免疫应答的化合物都将具有引起呼吸道超敏性的潜能。较大数量的化合物与皮肤超敏性和 ACD 的发展相关,并且认为对呼吸道不具有致敏作用 (Kimber, I., Dearman, R.J., 2005. What makes a chemical a respiratory sensitizer? Curr. Opin. Allergy Clin. Immunol. 5, 119-124)。与呼吸道过敏不同,ACD 是由细胞介导的免疫应答引起的延迟型超敏反应的实例 (Janeway 等, 1997, 上文)。ACD 是最常见的职业病之一,其中多种化合物成为病原体,因此,这些化合物的主动鉴定和表征也是相当重要

的 (Saary, J., Qureshi, R., Palda, V., DeKoven, J., Pratt, M., Skotnicki-Grant, S., Holness, L., 2005. A systematic review of contact dermatitis treatment and prevention. *J. Am. Acad. Dermatol.* 53, 845)。

[0005] 导致呼吸道过敏或 ACD 的超敏性的发展由两个截然不同的阶段组成。第一个是致敏,其涉及免疫状况的发展,而第二个是激发,其导致过敏的临床呈现 (Briatico-Vangosa, G., Braun, C. L., Cookman, G., Hofmann, T., Kimber, I., Loveless, S. E., Morrow, T., Pauluhn, J., Sorensen, T., Niessen, H. J., 1994. Respiratory allergy:hazard identification and risk assessment. *Fundam. Appl. Toxicol.* 23, 145 - 158)。因此,之前未暴露的(未免疫 (naive))但易感个体没有经历他们第一次暴露于过敏性蛋白质或化学品时的过敏症状。最少,需要两次暴露;然而,在许多情况中,可能需要在数周或数月内的重复暴露。在易感个体初次遭遇过敏性化合物的过程中,化合物被树突细胞(抗原加工和呈递细胞)识别为外来物,呈递给 T 细胞,并且引起特定的原发性免疫应答,这导致致敏。这接着可能是致敏的个体随后暴露于相同化合物时过敏的实际激发。激发通过免疫应答的激活和所得到的细胞信号来介导,所述细胞信号导致炎症反应和过敏症状。过敏反应的性质和严重程度取决于各种因素,包括个体的遗传背景、变应原的特征以及致敏和激发阶段过程中暴露的途径、持续时间和强度 (Arts, J. H., Kuper, C., 2003. Approaches to induce and elicit respiratory allergy:impact of route and intensity of exposure. *Toxicol. Lett.* 140 - 141, 213 - 222 ;Arts, J. H., Mommers, C., de Heer, C., 2006. Dose - response relationships and threshold levels in skin and respiratory allergy. *Crit. Rev. Toxicol.* 36, 219 - 251)。

[0006] 尽管存在一些一般性的相似性,但在目前了解的呼吸道过敏和 ACD 的病因学中存在重要的机理差异。通常,将呼吸道过敏归类为涉及 IgE 的 I 型超敏反应,而 ACD 是由 T 细胞介导的 IV 型超敏反应 (Janeway 等, 1997, 上文)。认为这些超敏性是根据特定的机理发展的,所述机理取决于 T 辅助 (Th) 细胞的功能性亚群 (即, Th1 和 Th2 细胞) 的不同激活。呼吸道致敏和过敏的发展与 T 淋巴细胞的 Th2 群的优先诱导相关。Th2 细胞的特征在于大量白细胞介素 (IL)-4、-10 和 -13 的产生。这些细胞因子的产生有利于体液免疫功能以及 B 细胞的刺激和分化,以产生 IgE (综述于 Dearman, R. J., Betts, C. J., Humphreys, N., Flanagan, B. F., Gilmour, N. J., Basketter, D. A., Kimber, I., 2003. Chemical allergy:considerations for the practical application of cytokine profiling. *Toxicol. Sci.* 71, 137 - 145)。这些抗体结合肥大细胞和嗜碱性粒细胞表面上的受体。在随后暴露于变应原时,这些细胞释放各种炎性介质,包括组胺、白细胞三烯和细胞因子,这导致了呼吸道过敏的即时超敏性。除了促进 IgE 产生, Th2 细胞因子还促进呼吸道过敏中涉及的其他细胞的生长和分化,包括肥大细胞和嗜曙红细胞 (综述于 Kimber, I., 1996. Chemical-induced hypersensitivity. 见 :Smialowicz, R. J., Holsapple, M. P. (Eds.), *Experimental Immunotoxicology*. CRC Press, New York, pp. 391 - 417)。在重复暴露于过敏性化合物和呼吸道过敏激发时,可能发展大范围的气道重塑、粘液积聚和慢性炎症应答,这引起了哮喘症状的发展。

[0007] 相反,接触性致敏和 ACD 的发展主要与 T 淋巴细胞的 Th1 群的诱导相关。这些细胞的特征在于 IL-2、干扰素- γ (IFN- γ) 和肿瘤坏死因子- β (TNF- β) 的产生。研究者已经

表明延迟型接触性超敏性的发展特异地依赖于 Th1 细胞和 IFN- β 的产生 (Diamantstein, T., Eckert, R., Volk, H. D., Kupier-Weglinski, J. W., 1988. Reversal by interferon-gamma of inhibition of delayed-type hypersensitivity induction by anti-CD4 or anti-interleukin 2 receptor (CD25) monoclonal antibodies. Evidence for the physiological role of the CD4+Th1+subset in mice. Eur. J. Immunol. 18, 2101-2103)。致敏应答与记忆 T 细胞的产生相关,在随后遭遇抗原时激活了记忆 T 细胞,导致了超敏性应答。这种反应涉及角质化细胞的激活和促炎细胞因子的释放,以募集非抗原特异性 T 细胞和单核细胞至接触部位,这导致急性炎性应答。有趣的是, Th1 细胞产生的 IFN- γ 也对抗 Th2 细胞应答和 IgE 的产生,而 Th2 细胞产生的 IL-4 对抗 Th1 细胞的发展。此外,已经发现了 IFN- γ 在呼吸道过敏中抑制肥大细胞功能,而 IL-4 抑制 ACD 的激发阶段 (综述于 Kimber, 1996 上文)。因此,不仅每种 Th 细胞类型的细胞因子促进它们谱系的生长和分化以及随后的超敏性应答,它们还对抗其他细胞群的增殖,作为进一步指引免疫应答的方式。

[0008] 从危险评价和调控观点看,以上呼吸道过敏和 ACD 之间的区别是相当重要的。研究者已经探究了多种动物模型和实验方法来鉴定具有引起呼吸道过敏潜能的化合物,然而,没有一种方法被研究团体或管理机构广泛应用或完全接受 (Arts 等, 2006 上文)。相反,存在多种指导测定,用于检测具有引起接触性致敏和 ACD 潜能的化合物。鉴于呼吸道过敏和 ACD 中致敏应答的一般性相似性,已经表明用于鉴定接触性致敏潜能的模型也可以用于呼吸道过敏潜能的评价。然而,由于已知的机理差异以及归类为呼吸道变应原更严重的健康和调控问题,这些化合物的精确鉴定及其与诱导 ACD 的化合物的区别是至关重要的。需要的是一致的、系统的和公认的用于评价蛋白质和化合物的呼吸道过敏潜能的方法。

[0009] 呼吸道过敏是上呼吸道和下呼吸道对异生物体蛋白质或化学品的 I 型超敏反应,其中临床症状通常包括喘息、气促、支气管收缩和哮喘攻击 (Boverhof DR, Billington R, Gollapudi BB, Hotchkiss JA, Krieger SM 等 (2008) Respiratory sensitization and allergy: current research approaches and needs. Toxicol Appl Pharmacol 226:1-13)。机械地,呼吸道过敏与 Th2 细胞的诱导和提高的 B 细胞的 IgE 产生相关。由效应器粒细胞 (如,肥大细胞和嗜碱性粒细胞) 上 IgE/ 变应原复合物引起的 Fc ϵ R:s 的交联导致促炎分子的释放 (Boverhof 等, 2008, 上文; Banks DE, Tarlo SM (2000) Important issues in occupational asthma. Curr Opin Pulm Med 6:37-42; Sastre J, Vandenplas O, Park HS (2003) Pathogenesis of occupational asthma. Eur Respir J 22:364-373)。通常, I 型超敏反应通过蛋白质变应原来引发,而低分子量化合物具有诱导过敏性接触性皮炎 (ACD) 的倾向, ACD 是主要与 Th1 和 CD8⁺ 效应器细胞的诱导相关的 IV 型超敏反应。然而,各种化学品,如二异氰酸酯 (Zammit-Tabona M, Sherkin M, Kijek K, Chan H, Chan-Yeung M (1983) Asthma caused by diphenylmethane diisocyanate in foundry workers. Clinical, bronchial provocation, and immunologic studies. Am Rev Respir Dis 128:226-230)、酸酐 (Bernstein DI, Patterson R, Zeiss CR (1982) Clinical and immunologic evaluation of trimellitic anhydride- and phthalic anhydride-exposed workers using a questionnaire with comparative analysis of enzyme-linked immunosorbent and radioimmunoassay studies. J Allergy Clin Immunol 69:311-318)、铂盐 (Murdoch RD, Pepys J, Hughes EG (1986) IgE antibody responses to platinum group

metals:a large scale refinery survey.Br J Ind Med 43:37-43)、反应性染料 (Docker A,Wattie JM,Topping MD,Luczynska CM,Newman Taylor AJ, 等 (1987)Clinical and immunological investigations of respiratory disease in workers using reactive dyes.Br J Ind Med 44:534-541) 和氯胺T(Bourne MS,Flindt ML,Walker JM(1979)Asthma due to industrial use of chloramine.Br Med J 2:10-12) 可能诱导呼吸道致敏,结果产生职业性哮喘和鼻炎。与引起接触性皮炎的那些相比,已知较少数的化学品引起呼吸道过敏,然而,健康影响仍然是很糟糕的,因为其可以与致命的结果相关。尽管临床特征与对蛋白质的过敏反应的那些相似,但应答的性质常常仍然是未知的。

[0010] REACH(化学品的注册、评估和授权(Registration,Evaluation,and Authorisation of Chemicals))法规要求欧盟内所有新的和现有的化学品(涉及大约30000种化学品)应当测试有害影响(Johansson H,Lindstedt M,Albrekt AS,Borrebaeck CA:A genomic biomarker signature can predict skin sensitizers using a cell-based in vitro alternative to animal tests.BMC Genomics 2011,12:399)。因为潜在的致敏剂的鉴定目前需要动物测试,REACH立法将对测试需要的动物数量具有巨大影响。此外,2009年生效的化妆品法第7次修正案对于人使用的大部分化妆品成分的动物测试提出了禁止,除了一些在2013年的测试之外。因此,可靠的用于评价化学品致敏能力的实验动物的体外替代方法的研发是迫切的。

[0011] 用于诱导呼吸道致敏的化学品的风险评价的方法的发展是非常不充分的,迄今为止没有可用的得到验证的试验(Verstraelen S,Bloemen K,Nelissen I,Witters H,Schoeters G 等 (2008)Cell types involved in allergic asthma and their use in in vitro models to assess respiratory sensitization.Toxicol In Vitro 22:1419-1431)。为了这一目的设计的主要体内测定是小鼠IgE测试(Dearman RJ,Basketter DA,Kimber I(1992)Variable effects of chemical allergens on serum IgE concentration in mice.Preliminary evaluation of a novel approach to the identification of respiratory sensitizers.J Appl Toxicol 12:317-323;Dearman RJ,Skinner RA,Humphreys NE,Kimber I(2003)Methods for the identification of chemical respiratory allergens in rodents:comparisons of cytokine profiling with induced changes in serum IgE.J Appl Toxicol 23:199-207)。尽管显示出具有前景的初期结果,实验室间的再现性不足以正式地验证该测定,并且目前没有常规地使用。然而,进行了尝试来研发基于细胞的用于呼吸道致敏的测定,使用树突细胞系,如THP-1(Verstraelen S,Nelissen I,Hooyberghs J,Witters H,Schoeters G 等 (2009)Gene profiles of THP-1macrophages after in vitro exposure to respiratory(non-)sensitizing chemicals:identification of discriminating genetic markers and pathway analysis.Toxicol In Vitro 23:1151-1162。)以及上皮细胞系,如BEAS-2B(Verstraelen S,Nelissen I,Hooyberghs J,Witters H,Schoeters G 等 (2009)Gene profiles of a human bronchial epithelial cell line after in vitro exposure to respiratory (non-)sensitizing chemicals:identification of discriminating genetic markers and pathway analysis.Toxicology 255:151-159)和A549(Verstraelen S,Nelissen I,Hooyberghs J,Witters H,Schoeters G,et al. (2009)

Gene profiles of a human alveolar epithelial cell line after in vitro exposure to respiratory (non-)sensitizing chemicals: identification of discriminating genetic markers and pathway analysis. *Toxicol Lett* 185:16-22)。此外,研发了在呼吸道致敏内的化学反应性测定,并且用于 ACD(Lalko JF, Kimber I, Dearman RJ, Gerberick GF, Sarlo K 等 (2011) Chemical reactivity measurements: potential for characterization of respiratory chemical allergens. *Toxicol In Vitro* 25:433-445)。然而,已经显示出肽反应性是皮肤和呼吸道致敏剂的共有特征,这严重地影响了两个组之间的区分。

[0012] 树突细胞 (DC) 通过作为先天性免疫和获得性免疫之间的必要联系的桥梁在免疫应答中起着关键作用。在引发时,它们可以快速地产生产大量介质,其影响炎症部位的其他细胞的迁移和活化,并且通过抗原呈递微调细胞应答来选择性地应答各种病原体和环境因子。因此,探究和利用致敏剂刺激过程中由 DC 引起的免疫决策制定,可以作为有力的用于致敏预测的测试策略。

[0013] 然而,不同 DC 亚群的多层面表型和专门化功能,以及它们广泛和稀有的分布,是复杂的因素,这阻碍了将初级 DC 用作测试平台。因此,对于建立精确且可靠的并且还避开了与获得 DC 中的可变性和困难性相关的问题的体外测定存在实际的需求。

发明内容

[0014] 因此,基于 DC 功能的可预测性的测定的研发应当优选依赖于可替换的细胞类型或体内 DC 的模拟。为此,具有 DC 特征的细胞系将是有利的,因为其构成稳定的、可繁殖的和无限供应的细胞。就 DC 模拟物而言,分化的骨髓单核细胞性 (myelomonocytic) MUTZ-3 细胞是优选的候选物 (Masterson, A. J., C. C. Sombroek, T. D. De Gruijl, Y. M. Graus, H. J. van der Vliet, S. M. Loughheed, A. J. van den Eertwegh, H. M. Pinedo 和 R. J. Scheper. 2002. MUTZ-3, a human cell line model for the cytokine-induced differentiation of dendritic cells from CD34+precursors. *Blood* 100:701-703.)。MUTZ-3 是 CD34⁺DC 祖细胞的无限来源并且其可以在细胞因子刺激时获取与不成熟的 DC 或朗格汉斯样 DC 相似的表型 (Santegoets, S. J., M. W. Schreurs, A. J. Masterson, Y. P. Liu, S. Goletz, H. Baumeister, E. W. Kueter, S. M. Loughheed, A. J. van den Eertwegh, R. J. Scheper, E. Hooijberg 和 T. D. de Gruijl. 2006. In vitro priming of tumor-specific cytotoxic T lymphocytes using allogeneic dendritic cells derived from the human MUTZ-3 cell line. *Cancer Immunol Immunother* 55:1480-1490.)。通过 CD1d、I 和 II 类 MHC 呈递抗原,并且诱导特异性的 T- 细胞增殖 (Masterson, A. J., C. C. Sombroek, T. D. De Gruijl, Y. M. Graus, H. J. van der Vliet, S. M. Loughheed, A. J. van den Eertwegh, H. M. Pinedo 和 R. J. Scheper. 2002. MUTZ-3, a human cell line model for the cytokine-induced differentiation of dendritic cells from CD34+precursors. *Blood* 100:701-703.)。用炎性介质刺激时, MUTZ-3 还呈现出成熟的转录和表型谱 (Larsson K, Lindstedt M 和 Borrebaeck CAK. Functional and transcriptional profiling of MUTZ-3. A myeloid cell line acting as a model for dendritic cells. *Immunology*. 2006 Feb; 117(2):156-66.)。

[0015] 本发明的发明人已经研发了一种用于预测呼吸道致敏剂的新测试原理。令人惊讶

地发现了使用 DC 祖细胞,如 MUTZ-3 细胞,可以精确鉴定 / 预测呼吸道致敏剂,而不需要在过程中进一步分化,由此用一组致敏化学品、非致敏化学品和 / 或其他对照(例如,只包含稀释剂的载体对照,如 DMSO 和 / 或蒸馏水)刺激细胞。发现这实质性地简化和提高了程序的可再现性。

[0016] 因此,本发明的第一个方面提供了一种用于鉴定能够诱导哺乳动物呼吸道致敏的物质的方法,其包括以下步骤或由以下步骤组成:

[0017] a) 将树突细胞群或树突样细胞群暴露于测试物质;和

[0018] b) 测量所述细胞中一种或多种选自表 1 中限定的生物标记物的表达,

[0019] 其中步骤 (b) 中测量的细胞中的一种或多种生物标记物的表达表明测试物质的呼吸道致敏作用。

[0020] “能够诱导呼吸道致敏的物质”表示能够诱导和引发哺乳动物呼吸道中的 I 型即时超敏反应的任何物质。优选,哺乳动物是人。优选,I 型即时超敏反应是 DC- 介导的和 / 或涉及 T 细胞分化成 Th2 细胞。优选,I 型即时超敏反应导致体液免疫和 / 或呼吸道过敏。

[0021] 哺乳动物肺的传导区含有气管、支气管、细支气管和末端细支气管。呼吸区含有呼吸细支气管、肺泡管和肺泡。传导区由气道构成,不与血液进行气体交换,并且由软骨增强,以保持打开气道。传导区湿润吸入的空气并且将其温热至 37°C (99°F)。还通过经由位于所有通路的壁上的纤毛除去颗粒而清洁了空气。呼吸区是与血液进行气体交换的部位。

[0022] 在一个实施方案中,“能够诱导哺乳动物皮肤致敏的物质”是能够在哺乳动物肺上皮部位诱导和引发 I 型即时超敏反应的物质。优选,肺上皮部位是在肺的呼吸区中,但可以替换地或另外地,在肺的传导区中。

[0023] 哺乳动物可以是任何家畜或农场动物。优选,哺乳动物是大鼠、小鼠、豚鼠、猫、狗、马或灵长类。最优选,哺乳动物是人。

[0024] 树突细胞 (DC) 是形成哺乳动物免疫系统一部分的免疫细胞。它们的主要功能是加工抗原物质并且将其在表面上呈递给免疫系统的其他细胞(即,它们作为抗原呈递细胞),作为先天性免疫系统和获得性免疫系统的桥梁。

[0025] 树突细胞存在于接触外部环境的组织中,如皮肤(此处具有称为朗格汉斯细胞的特化树突细胞类型)以及鼻子、肺、胃和肠的内膜。还在血液发现了未成熟状态的树突细胞。一旦激活,树突细胞移动至淋巴结,在此它们与 T 细胞和 B 细胞相互作用,以启动和形成获得性免疫应答。在某些发展阶段,它们生长出支化的突出物,即树突。尽管外观上相似,这些树突与神经元的树突是截然不同的结构。未成熟的树突细胞也称为面纱细胞 (veiled cell),因为它们具有大的细胞质“面纱”,而不是树突。

[0026] “树突样细胞”是指呈现出树突细胞特异性的功能和表型特征的非树突细胞,如形态学特征、共同刺激分子和 II 类 MHC 分子的表达以及吞饮大分子和激活静息 T 细胞的能力。

[0027] 在一个实施方案中,树突样细胞是 CD34⁺树突细胞祖细胞。任选,CD34⁺树突细胞祖细胞可以在细胞因子刺激时获取通过 CD1d、I 和 II 类 MHC 呈递抗原的表型,诱导特异性 T- 细胞增殖和 / 或用炎性介质刺激时呈现出成熟的转录和表型谱(即,与未成熟树突细胞或朗格汉斯-样树突细胞相似的表型)。

[0028] 可以通过功能、通过表型和 / 或通过基因表达模式,特别是通过细胞表面表型,来

识别树突细胞。这些细胞的特征在于它们不同的形态学、高水平的表面 II 类 MHC 表达和呈递抗原至 CD4⁺ 细胞和 / 或 CD8⁺T 细胞,特别是呈递至天然 T 细胞的能力 (Steinman 等, (1991) Ann. Rev. Immunol. 9:271)。

[0029] 树突细胞的细胞表面是与众不同的,具有特征性的面纱样突出物,并且特征在于细胞表面标记物 CD11c 和 II 类 MHC 的表达。大部分 DC 对其他白细胞谱系的标记物是阴性的,包括 T 细胞、B 细胞、单核细胞 / 巨噬细胞和粒细胞。树突细胞的亚群还可以表达其他标记物,包括 33D1、CCR1、CCR2、CCR4、CCR5、CCR6、CCR7、CD1a-d、CD4、CD5、CD8 α 、CD9、CD11b、CD24、CD40、CD48、CD54、CD58、CD80、CD83、CD86、CD91、CD117、CD123 (IL3Ra)、CD134、CD137、CD150、CD153、CD162、CXCR1、CXCR2、CXCR4、DCIR、DC-LAMP、DC-SIGN、DEC205、E- 钙粘蛋白、Langerin、甘露糖受体、MARCO、TLR2、TLR3、TLR4、TLR5、TLR6、TLR9 和几种凝集素。

[0030] 这些细胞表面标记物的表达模式可以连同树突细胞的成熟度、它们的组织来源和 / 或它们的来源物种一起改变。未成熟树突细胞表达低水平的 II 类 MHC,但能够内吞抗原蛋白并且对其进行加工,以在与 II 类 MHC 分子的复合物中呈递。激活的树突细胞表达高水平的 II 类 MHC、ICAM-1 和 CD86,并且能够刺激天然同种异体的 T 细胞的增殖,例如,在混合白细胞反应 (MLR) 中。

[0031] 在功能上,可以通过任一种用于确定抗原呈递的适当测定来鉴定树突细胞和树突样细胞。这样的测定可以包括通过测试抗原呈递、接着确定 T 细胞增殖、IL-2 的释放等来测试刺激抗原引发的和 / 或天然的 T 细胞的能力。

[0032] “表达”是指基因产物 (如, mRNA 或蛋白质) 的水平或含量。

[0033] 检测和 / 或测量蛋白质和 / 或核酸的浓度的方法是本领域技术人员公知的,参见,例如 Sambrook 和 Russell, 2001, Cold Spring Harbor Laboratory Press。

[0034] 用于检测和 / 或测量蛋白质的优选方法包括 Western 印迹、North-Western 印迹、免疫吸附测定 (ELISA)、抗体微阵列、组织微阵列 (TMA)、免疫沉淀、原位杂交和其他免疫组织化学技术、放射性免疫测定 (RIA)、免疫放射性测定 (IRMA) 和免疫酶测定 (IEMA),包括使用单克隆和 / 或多克隆抗体的夹心测定。示例性夹心测定描述于 David 等的美国专利 4,376,110 和 4,486,530 中,将其通过引用的方式并入本申请。载玻片上的细胞的抗体染色可以用于细胞学实验室诊断测试中公知的以及本领域技术人员已知的方法中。

[0035] 通常,ELISA 涉及使用产生有色反应产物的酶,通常在固相测定中。如辣根过氧化物酶和磷酸酶这样的酶已经被广泛使用。扩大磷酸酶反应的一种方式是使用 NADP 作为底物,来产生 NAD,其现在作为用于第二个酶系统的辅酶。因为组织中不存在酶,因此来自大肠杆菌的焦磷酸酶提供了良好的缀合物,其是稳定的并且产生了良好的反应颜色。还可以使用基于酶 (如,萤光素酶) 的化学发光系统。

[0036] 常常使用与维生素生物素的缀合,因为这可以通过其与酶联抗生物素蛋白 (avidin) 或抗生物素蛋白链菌素 (streptavidin) 的反应容易地检测出来,其以很高的特异性和亲和力与酶联抗生物素蛋白或抗生物素蛋白链菌素结合。

[0037] 用于核酸 (例如, mRNA) 的检测和 / 或测量的优选方法包括 southern 印迹、northern 印迹、聚合酶链式反应 (PCR)、逆转录酶 PCR (RT-PCR)、定量实时 PCR (qRT-PCR)、纳米阵列、微阵列、大阵列、放射自显影术和原位杂交。

[0038] 在一个实施方案中,该方法进一步包括以下步骤:

[0039] c) 将分开的树突细胞群或树突样细胞群暴露于一种或多种不是哺乳动物呼吸道致敏剂的阴性对照剂 ;和

[0040] d) 测量细胞中步骤 (b) 中测量的一种或多种生物标记物的表达,

[0041] 其中在步骤 (b) 中测量的测试样品中的一种或多种生物标记物的存在和 / 或含量不同于步骤 (d) 中测量的对照样品中的一种或多种生物标记物的存在和 / 或含量的情况下, 将测试物质鉴定为呼吸道致敏剂。

[0042] “不同于步骤 (b) 中测量的对照样品中的一种或多种蛋白质的存在和 / 或含量”是指测试样品中的存在和 / 或含量以统计学显著的方式不同于一种或多种阴性对照样品中的存在和 / 或含量。优选, 暴露于测试物质的细胞群中的一种或多种生物标记物的表达为:

[0043] 低于或等于暴露于阴性对照剂的细胞群中的表达的 80%, 例如, 不超过暴露于阴性对照剂的细胞群中的表达的 79%、78%、77%、76%、75%、74%、73%、72%、71%、70%、69%、68%、67%、66%、65%、64%、63%、62%、61%、60%、59%、58%、57%、56%、55%、54%、53%、52%、51%、50%、49%、48%、47%、46%、45%、44%、43%、42%、41%、40%、39%、38%、37%、36%、35%、34%、33%、32%、31%、30%、29%、28%、27%、26%、25%、24%、23%、22%、21%、20%、19%、18%、17%、16%、15%、14%、13%、12%、11%、10%、9%、8%、7%、6%、5%、4%、3%、2%、1% 或 0% ;或

[0044] 暴露于阴性对照剂的细胞群中的表达的至少 120%, 例如, 暴露于阴性对照剂的细胞群中的表达的至少 121%、122%、123%、124%、125%、126%、127%、128%、129%、130%、131%、132%、133%、134%、135%、136%、137%、138%、139%、140%、141%、142%、143%、144% , 145%、146%、147%、148%、149%、150%、151%、152%、153%、154% , 155%、156%、157%、158%、159% , 160%、161%、162%、163%、164%、165%、166%、167%、168%、169%、170%、171%、172%、173%、174%、175%、176%、177%、178%、179%、180%、181%、182%、183%、184%、185%、186%、187%、188%、189%、190%、191%、192%、193%、194%、195%、196%、197%、198%、199%、200%、225%、250%、275%、300%、325%、350%、375%、400%、425%、450%、475%, 或至少 500%。

[0045] 一种或多种阴性对照剂可以包括一种或多种选自 1- 丁醇、4- 氨基苯甲酸、氯苯、二甲基甲酰胺、乙基香草醛、异丙醇、水杨酸甲酯、丙二醇、高锰酸钾、吐温 80™ (聚氧乙烯 (20) 山梨聚糖单油酸酯) 和硫酸锌的物质 (即, 表 2 中限定的非致敏剂的组) 或由以上物质组成。

[0046] 阴性对照剂可以是溶剂, 与本发明的测试物质或对照剂一起使用。因此, 阴性对照可以是 DMSO 和 / 或蒸馏水。

[0047] 可替换地或另外地, 将在测试物质暴露之前在步骤 (b) 中测量的树突细胞或树突样细胞中的一种或多种生物标记物的表达用作阴性对照。

[0048] 再一个实施方案包括以下步骤:

[0049] e) 将分开的树突细胞群或树突样细胞群暴露于一种或多种是哺乳动物呼吸道致敏剂的阳性对照剂 ;和

[0050] f) 测量细胞中步骤 (b) 中测量的一种或多种生物标记物的表达,

[0051] 其中在步骤 (f) 中测量的测试样品中的一种或多种生物标记物的存在和 / 或 / 或含量对应于步骤 (b) 中测量的一种或多种阳性对照样品中的一种或多种生物标记物的存

在和 / 或含量的情况下,将测试物质鉴定为呼吸道致敏剂。

[0052] “对应于一种或多种阳性对照样品中的表达”是指暴露于测试物质的细胞群中的一种或多种生物标记物的表达与暴露于一种或多种阳性对照剂的细胞群中的表达的相同或没有显著不同。优选,暴露于测试物质的细胞群中的一种或多种生物标记物的表达为暴露于一种或多种阳性对照剂的细胞群中的表达的 81% 至 119%,例如,高于或等于暴露于一种或多种阳性对照剂的细胞群中的表达的 82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98% 或 99%,并且低于或等于暴露于一种或多种阳性对照剂的细胞群中的表达的 101%、102%、103%、104%、105%、106%、107%、108%、109%、110%、111%、112%、113%、114%、115%、116%、117%、118% 或 119%。

[0053] 因此,根据本发明第一个方面的方法可以包括测量 OR5B21 表达。该方法可以包括测量 SLC7A7 表达。

[0054] 该方法可以包括测量 PIP3-E 表达。该方法可以包括测量 BTNL8 表达。该方法可以包括测量 CLEC4A 表达。该方法可以包括测量 HIST4H4 表达。该方法可以包括测量 YKT-6 表达。该方法可以包括测量 FLJ32679///GOLGA8G///GOLGA8E 表达。该方法可以包括测量 PACSIN3 表达。该方法可以包括测量 PDE1B 表达。该方法可以包括测量 NQO1 表达。该方法可以包括测量 CAMK1D 表达。该方法可以包括测量 MYB 表达。该方法可以包括测量 ENST00000387396 表达。该方法可以包括测量 GRK5 表达。

[0055] 该方法可以包括测量 CD86 表达。该方法可以包括测量 CD1A 表达。该方法可以包括测量 WWOX 表达。该方法可以包括测量 IKZF2 表达。该方法可以包括测量 FUCA1 表达。该方法可以包括测量 C10orf76 表达。该方法可以包括测量 AMICA1 表达。该方法可以包括测量 PDPK2///PDPK1 表达。该方法可以包括测量 AZU1 表达。该方法可以包括测量 ACN9 表达。该方法可以包括测量 PDPN 表达。该方法可以包括测量 LOC642587 表达。该方法可以包括测量 SEC61A2 表达。该方法可以包括测量 ELA2 表达。该方法可以包括测量 BMP2K 表达。该方法可以包括测量 HCCS 表达。该方法可以包括测量 CXorf26 表达。该方法可以包括测量 TYSND1 表达。该方法可以包括测量 CARS 表达。该方法可以包括测量 NECAP1 表达。该方法可以包括测量 CDH26 表达。该方法可以包括测量 SERPINB1 表达。该方法可以包括测量 STEAP4 表达。该方法可以包括测量 TXNIP 表达。该方法可以包括测量 ENST00000386628 表达。该方法可以包括测量 C12orf35 表达。该方法可以包括测量 HMGA2 表达。该方法可以包括测量 KRT16 表达。该方法可以包括测量 GGTL2 表达。该方法可以包括测量 ENST00000386437 表达。该方法可以包括测量 OSBPL11 表达。该方法可以包括测量 FAM71F1 表达。该方法可以包括测量 ATP6V1B2 表达。该方法可以包括测量 LOC128102 表达。该方法可以包括测量 TBX19 表达。该方法可以包括测量 NID1 表达。该方法可以包括测量 LPXN 表达。该方法可以包括测量 C15orf45 表达。该方法可以包括测量 RNF111 表达。该方法可以包括测量 ENST00000386861 表达。该方法可以包括测量 CD33 表达。该方法可以包括测量 TANK 表达。该方法可以包括测量 ANKRD44 表达。该方法可以包括测量 WDFY1 表达。该方法可以包括测量 SDC4 表达。该方法可以包括测量 TMPRSS11B 表达。该方法可以包括测量 AFF4 表达。该方法可以包括测量 HBEGF 表达。该方法可以包括测量 XK 表达。该方法可以包括测量 SLAMF7 表达。该方法可以包括测量 S100A4 表达。该方法可以包括测量 MPZL3 表达。该方法可以包括

测量 GENSCAN00000044853 表达。该方法可以包括测量 TRAV8-3 表达。该方法可以包括测量 LOC100131497 表达。该方法可以包括测量 KIAA1468 表达。该方法可以包括测量 SPHK2 表达。该方法可以包括测量 ENST00000309260 表达。该方法可以包括测量 CCR6 表达。该方法可以包括测量 GSTA3 表达。该方法可以包括测量 RALA 表达。该方法可以包括测量 C7orf53 表达。该方法可以包括测量 AF480566 表达。该方法可以包括测量 CERCAM 表达。该方法可以包括测量 hsa-mir-147 表达。该方法可以包括测量 NFYC 表达。该方法可以包括测量 CD53 表达。该方法可以包括测量 PSEN2 表达。该方法可以包括测量 CISD1 表达。该方法可以包括测量 SCD 表达。该方法可以包括测量 MED19 表达。该方法可以包括测量 SYT17 表达。该方法可以包括测量 KRT16///LOC400578///MGC102966 表达。该方法可以包括测量 C18orf51 表达。该方法可以包括测量 CD79A 表达。该方法可以包括测量 C19orf56 表达。该方法可以包括测量 AGFG1 表达。该方法可以包括测量 FOXP1 表达。该方法可以包括测量 TLR6 表达。该方法可以包括测量 SUS3 表达。该方法可以包括测量 ENST00000387842 表达。该方法可以包括测量 ENST00000387842 表达。该方法可以包括测量 GPA33 表达。该方法可以包括测量 CDC123 表达。该方法可以包括测量 C10orf11 表达。该方法可以包括测量 ENST00000322493 表达。该方法可以包括测量 PTMAP7 表达。该方法可以包括测量 ARRDC4 表达。该方法可以包括测量 ENST00000388199 表达。该方法可以包括测量 ENST00000388437 表达。该方法可以包括测量 KRT9 表达。该方法可以包括测量 ENST00000379371 表达。该方法可以包括测量 HDAC4 表达。该方法可以包括测量 CD200 表达。该方法可以包括测量 PAPSS1 表达。该方法可以包括测量 ORAI2 表达。该方法可以包括测量 AK124536 表达。该方法可以包括测量 ZBTB10 表达。该方法可以包括测量 ENST00000387422 表达。该方法可以包括测量 RAB9A 表达。该方法可以包括测量 7895613 表达。该方法可以包括测量 DRD5 表达。该方法可以包括测量 CNR2 表达。该方法可以包括测量 OIT3 表达。该方法可以包括测量 ENST00000386981 表达。该方法可以包括测量 C10orf90 表达。该方法可以包括测量 OR52D1 表达。该方法可以包括测量 ZNF214 表达。该方法可以包括测量 ENST00000386959 表达。该方法可以包括测量 ART4 表达。该方法可以包括测量 RCBTB2 表达。该方法可以包括测量 HOMER2 表达。该方法可以包括测量 WWP2 表达。该方法可以包括测量 WDR24 表达。该方法可以包括测量 MED31 表达。该方法可以包括测量 CALM2 表达。该方法可以包括测量 DLX2 表达。该方法可以包括测量 BTBD2 表达。该方法可以包括测量 ENST00000339367 表达。该方法可以包括测量 TBCA 表达。该方法可以包括测量 GIN1 表达。该方法可以包括测量 NOL7 表达。该方法可以包括测量 ENST00000402365 表达。该方法可以包括测量 C7orf28B///C7orf28A 表达。该方法可以包括测量 DPP7 表达。该方法可以包括测量 hCG_1749005 表达。该方法可以包括测量 PNPLA4 表达。该方法可以包括测量 USP51 表达。该方法可以包括测量 HLA-DQA1///HLA-DRA 表达。该方法可以包括测量 FAAH 表达。该方法可以包括测量 GDAP2 表达。该方法可以包括测量 CD48 表达。该方法可以包括测量 PTPRJ 表达。该方法可以包括测量 EXPH5 表达。该方法可以包括测量 RPS26///LOC728937///RPS26L///hCG_2033311 表达。该方法可以包括测量 ALDH2 表达。该方法可以包括测量 CALM1 表达。该方法可以包括测量 NOX5///SPESP1 表达。该方法可以包括测量 RHBDL1 表达。该方法可以包括测量 CYLD 表达。该方法可以包括测量 OSBPL1A 表达。该方法可以包括测量 GYPC 表达。该方法可以包括测量 RQCD1 表达。该方法可以包括测量 RBM44 表达。该方法可以包括测量 ENST00000384680 表达。该

方法可以包括测量 C3orf58 表达。该方法可以包括测量 MFSD1 表达。该方法可以包括测量 HAC11 表达。该方法可以包括测量 SATB1 表达。该方法可以包括测量 USP4 表达。该方法可以包括测量 ENST00000410125 表达。该方法可以包括测量 ENST00000384055 表达。该方法可以包括测量 L7R 表达。该方法可以包括测量 ENST00000364497 表达。该方法可以包括测量 FAM135A 表达。该方法可以包括测量 CD164 表达。该方法可以包括测量 DYNLT1 表达。该方法可以包括测量 NRCAM 表达。该方法可以包括测量 ZNF596 表达。该方法可以包括测量 ENST00000332418 表达。该方法可以包括测量 TCEAL3///TCEAL6 表达。该方法可以包括测量 SNAPIN 表达。该方法可以包括测量 DENND2D 表达。该方法可以包括测量 SAMD8 表达。该方法可以包括测量 LHPP 表达。该方法可以包括测量 SLC37A2 表达。该方法可以包括测量 FLI1///EWSR1 表达。该方法可以包括测量 OR9G4 表达。该方法可以包括测量 LOC338799 表达。该方法可以包括测量 HEXDC 表达。该方法可以包括测量 NOTUM 表达。该方法可以包括测量 MCOLN1 表达。该方法可以包括测量 PRKACA 表达。该方法可以包括测量 CRIM1 表达。该方法可以包括测量 CECR5 表达。该方法可以包括测量 RNF13 表达。该方法可以包括测量 40969 表达。该方法可以包括测量 ZNF366 表达。该方法可以包括测量 ENST00000410754 表达。该方法可以包括测量 GIMAP5 表达。该方法可以包括测量 ENST00000362484 表达。该方法可以包括测量 TFE3 表达。该方法可以包括测量 RHOU 表达。该方法可以包括测量 MED8 表达。该方法可以包括测量 CASQ2 表达。该方法可以包括测量 NUDT5 表达。该方法可以包括测量 C11orf73 表达。该方法可以包括测量 PAK1 表达。该方法可以包括测量 PRSS21 表达。该方法可以包括测量 ENST00000332418 表达。该方法可以包括测量 BTBD12 表达。该方法可以包括测量 DHRS13 表达。该方法可以包括测量 CCDC102B 表达。该方法可以包括测量 BCL2 表达。该方法可以包括测量 ZNF211///ZNF134 表达。该方法可以包括测量 NDUFV2 表达。该方法可以包括测量 MYCN 表达。该方法可以包括测量 ENST00000385528 表达。该方法可以包括测量 ENST00000264275 表达。该方法可以包括测量 CASP8 表达。该方法可以包括测量 RTN4 表达。该方法可以包括测量 PLCG1 表达。该方法可以包括测量 MGC42105 表达。该方法可以包括测量 EMB 表达。该方法可以包括测量 ENST00000386433 表达。该方法可以包括测量 COL21A1 表达。该方法可以包括测量 LRP12 表达。该方法可以包括测量 LMNA 表达。该方法可以包括测量 ENST00000385567 表达。该方法可以包括测量 ENST00000362863 表达。该方法可以包括测量 ZNF503 表达。该方法可以包括测量 NLRX1 表达。该方法可以包括测量 ENST00000391173 表达。该方法可以包括测量 NDRG2 表达。该方法可以包括测量 TRAF7 表达。该方法可以包括测量 KRT40 表达。该方法可以包括测量 KRT40 表达。该方法可以包括测量 DRD5 表达。该方法可以包括测量 ZC3H8 表达。该方法可以包括测量 MMP9 表达。该方法可以包括测量 PLTP 表达。该方法可以包括测量 ENST00000362686 表达。该方法可以包括测量 SPEF2 表达。该方法可以包括测量 LRRC16A 表达。该方法可以包括测量 FBX09 表达。该方法可以包括测量 EEPD1 表达。该方法可以包括测量 FCN1 表达。该方法可以包括测量 EFNA3 表达。该方法可以包括测量 ENST00000314893 表达。该方法可以包括测量 TMEM19 表达。该方法可以包括测量 PLXNC1 表达。该方法可以包括测量 NHLRC3 表达。该方法可以包括测量 MBNL2 表达。该方法可以包括测量 EIF5 表达。该方法可以包括测量 PLEKHG4 表达。该方法可以包括测量 COPS3 表达。该方法可以包括测量 FAM171A2 表达。该方法可以包括测量 LOC653653///AP1S2 表达。该方法可以包括测量 VAPA 表达。该方法可以包括测

量 MATK 表达。该方法可以包括测量 ACTR2 表达。该方法可以包括测量 BPI 表达。该方法可以包括测量 ERG 表达。该方法可以包括测量 LAMB2 表达。该方法可以包括测量 BC090058 表达。该方法可以包括测量 PHTF2 表达。该方法可以包括测量 ENST00000333261 表达。该方法可以包括测量 C8orf55 表达。该方法可以包括测量 PDE7A 表达。该方法可以包括测量 NAPRT1 表达。该方法可以包括测量 HLA-DRA 表达。该方法可以包括测量 SLC22A15 表达。该方法可以包括测量 FCGR1A//FCGR1B//FCGR1C 表达。该方法可以包括测量 SLC27A3 表达。该方法可以包括测量 ID3 表达。该方法可以包括测量 TBCEL 表达。该方法可以包括测量 FAM138D 表达。该方法可以包括测量 POMP 表达。该方法可以包括测量 SNN 表达。该方法可以包括测量 MED13 表达。该方法可以包括测量 ZFP36L2 表达。该方法可以包括测量 UXS1 表达。该方法可以包括测量 CD40 表达。该方法可以包括测量 ENST00000362620 表达。该方法可以包括测量 GGT5 表达。该方法可以包括测量 BC035666 表达。该方法可以包括测量 G6PD 表达。该方法可以包括测量 ENST00000384272 表达。该方法可以包括测量 CLCC1 表达。该方法可以包括测量 SCGB2A1 表达。该方法可以包括测量 GAA 表达。该方法可以包括测量 SERPINB2 表达。该方法可以包括测量 GPI 表达。该方法可以包括测量 LASS6 表达。该方法可以包括测量 EIF4A2 表达。该方法可以包括测量 HLA-DRA 表达。该方法可以包括测量 ENST00000385586 表达。该方法可以包括测量 ANXA2P2 表达。该方法可以包括测量 FANCG 表达。该方法可以包括测量 FAM53B 表达。该方法可以包括测量 RFXAP 表达。该方法可以包括测量 UBR1 表达。该方法可以包括测量 TBC1D2B 表达。该方法可以包括测量 SERPINB10 表达。该方法可以包括测量 SEC23B 表达。该方法可以包括测量 MN1 表达。该方法可以包括测量 CRTAP 表达。

[0056] 该方法可以包括在步骤 (b) 中测量表 1A 中限定的一种或多种生物标记物例如至少 2 种表 1A 中限定的生物标记物的表达,或由在步骤 (b) 中测量表 1A 中限定的一种或多种生物标记物例如至少 2 种表 1A 中限定的生物标记物的表达组成。因此,该方法可以包括测量 OR5B21 的表达。该方法可以包括测量 SLC7A7 的表达。在一个优选的实施方案中,该方法包括在步骤 (b) 中测量 OR5B21 和 SLC7A7 的表达,或由在步骤 (b) 中测量 OR5B21 和 SLC7A7 的表达组成。

[0057] 该方法可以另外地或替换地包括在步骤 (b) 中测量表 1B 中限定的一种或多种生物标记物例如 2、3、4、5、6、7、8、10、11、12 或 13 种表 1B 中限定的生物标记物的表达,或由在步骤 (b) 中测量表 1B 中限定的一种或多种生物标记物例如 2、3、4、5、6、7、8、10、11、12 或 13 种表 1B 中限定的生物标记物的表达组成。

[0058] 该方法可以另外地或替换地包括在步骤 (b) 中测量表 1C 中限定的一种或多种生物标记物例如 2、3、4、5、6、7、8、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、114、115、116、117、118、119、120、121、122、123、124、125、126、127、128、129、130、131、132、133、134、135、136、137、138、139、140、141、142、143、144、145、146、147、148、149、150、151、152、153、154、155、156、157、158、159、160、161、162、163、164、165、166、167、168、169、170、171、172、173、174、175、176、177、

178、179、180、181、182、183、184、185、186、187、188、189、190、191、192、193、194、195、196、197、198、199、200、201、202、203、204、205、206、207、208、209、210、211、212、213、214、215、216、217、218、219、220、221、222、223、224、225、226、227、228、229、230、231、232、233、234、235、236、237、238、239、240、241、242、243、244、245、246、247、248、249、250、251、252、253、254、255、256、257、258、259、260、261、262、263、264、265、266、267、268、269、270、271、272、273、274、275、276、277、278、279、280、281、282、283、284、285、286 或 287 种表 1C 中限定的生物标记物的表达,或由在步骤 (b) 中测量表 1C 中限定的一种或多种生物标记物例如 2、3、4、5、6、7、8、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、114、115、116、117、118、119、120、121、122、123、124、125、126、127、128、129、130、131、132、133、134、135、136、137、138、139、140、141、142、143、144、145、146、147、148、149、150、151、152、153、154、155、156、157、158、159、160、161、162、163、164、165、166、167、168、169、170、171、172、173、174、175、176、177、178、179、180、181、182、183、184、185、186、187、188、189、190、191、192、193、194、195、196、197、198、199、200、201、202、203、204、205、206、207、208、209、210、211、212、213、214、215、216、217、218、219、220、221、222、223、224、225、226、227、228、229、230、231、232、233、234、235、236、237、238、239、240、241、242、243、244、245、246、247、248、249、250、251、252、253、254、255、256、257、258、259、260、261、262、263、264、265、266、267、268、269、270、271、272、273、274、275、276、277、278、279、280、281、282、283、284、285、286 或 287 种表 1C 中限定的生物标记物的表达组成。

[0059] 因此,在步骤 (b) 中可以测量表 1A 中限定的所有生物标记物的表达和 / 或表 1B 中限定的所有生物标记物的表达和 / 或表 1C 中限定的所有生物标记物的表达。因此,该方法可以包括在步骤 (b) 中测量表 1 中限定的所有生物标记物,或由在步骤 (b) 中测量表 1 中限定的所有生物标记物组成。

[0060] 在优选的实施方案中,步骤 (b) 包括测量编码一种或多种生物标记物的核酸分子的表达,或由测量编码一种或多种生物标记物的核酸分子的表达组成。核酸分子可以是 cDNA 分子或 mRNA 分子。优选,核酸分子是 mRNA 分子。然而,核酸分子可以是 cDNA 分子。

[0061] 在一个实施方案中,使用选自 Southern 杂交、Northern 杂交、聚合酶链式反应 (PCR)、逆转录酶 PCR (RT-PCR)、定量实时 PCR (qRT-PCR)、纳米阵列、微阵列、大阵列、放射自显影术和原位杂交来进行步骤 (b) 中的一种或多种生物标记物的表达。优选,使用 DNA 微阵列测量一种或多种生物标记物的表达。

[0062] 该方法可以包括使用一种或多种结合部分在步骤 (b) 中测量一种或多种生物标记物的表达,每种所述结合部分能够选择性地结合编码表 1 中鉴定的生物标记物之一的核酸分子。在一个实施方案中,一种或多种结合部分各自包括核酸分子或由核酸分子组成。在进一步的实施方案中,一种或多种结合部分各自包括 DNA、RNA、PNA、LNA、GNA、TNA 或 PMO,或由 DNA、RNA、PNA、LNA、GNA、TNA 或 PMO 组成。优选,一种或多种结合部分各自包括 DNA 或由 DNA 组成。在一个实施方案中,一种或多种结合部分的长度为 5 至 100 个核苷酸。然而,

在可替换的实施方案中,它们的长度是 15 至 35 个核苷酸。

[0063] 如以下讨论的,基于其结合给定的核酸、蛋白质或氨基酸基序的能力,从文库选择或筛选合适的结合剂(也称为结合分子)。

[0064] 在优选的实施方案中,结合部分包括可检测部分(detectable moiety)。

[0065] “可检测部分”包括允许直接或间接测定其存在和/或相对含量和/或位置(例如,在阵列上的位置)的部分。

[0066] 合适的可检测部分是本领域公知的。

[0067] 例如,可检测部分可以是荧光和/或发光和/或化学发光部分,其暴露于特定条件时,可以检测到。这样的荧光部分可能需要暴露于特定波长和强度的照射(即,光),以引起荧光部分的激发,由此使其发射出可以检测到的特定波长的可检测荧光。

[0068] 或者,可检测部分可以是酶,其能够将(优选不可检测的)底物转化成可观察和/或检测到的可检测产物。合适的酶的实例在以下涉及例如 ELISA 测定中更详细地讨论。

[0069] 因此,可检测部分可以选自:荧光部分;发光部分;化学发光部分;放射性部分(例如,放射性原子);或酶部分。优选,可检测部分包括放射性原子或由放射性原子组成。放射性原子可以选自钨-99m、碘-123、碘-125、碘-131、铟-111、氟-19、碳-13、氮-15、氧-17、磷-32、硫-35、氘、氚、铯-137、铯-138 和钪-45。

[0070] 明显地,待检测物质(如,例如,本文中所述的测试样品和/或对照样品中的一种或多种生物标记物和/或用于检测所选蛋白的抗体分子)必须具有足够的合适的原子同位素,使得能容易地检测到可检测部分。

[0071] 在可替换的优选实施方案中,结合部分的可检测部分是荧光部分。

[0072] 可以以已知的方式将放射性或其他标记结合至本发明方法的样品中存在的生物标记物中和/或本发明的结合部分中。例如,如果结合剂是多肽,其可以是生物合成的或可以使用适当的氨基酸前体通过化学氨基酸合成来合成,其中涉及例如用氟-19 替代氢。例如,可以经由结合部分中半胱氨酸残基连接标记物,如 ^{99m}Tc 、 ^{123}I 、 ^{186}Rh 、 ^{188}Rh 和 ^{111}In 。可以经由赖氨酸残基连接钪-45。IODOGEN 方法(Fraker 等(1978) *Biochem. Biophys. Res. Comm.* 80, 49-57)可以用于结合 ^{125}I 。参考文献(“*Monoclonal Antibodies in Immunoscintigraphy*”, J-F Chatal, CRC Press, 1989)详细描述了其他方法。用于将其他可检测部分(如酶、荧光、发光、化学发光或放射性部分)缀合蛋白质的方法是本领域公知的。

[0073] 本领域技术人员将认识到可以用间接帮助确定所述蛋白质的存在、含量和/或位置的部分来标记待测试样品中的生物标记物。因此,该部分可以构成多组分可检测部分中的一种组分。例如,可以用生物素标记待测试样品中的生物标记物,这允许随后使用与可检测标记融合或以其它方式结合的抗生物素蛋白链菌素来检测。

[0074] 本发明第一个方面中提供的方法可以包括在步骤(b)中测定表1中限定的一种或多种生物标记物的蛋白质的表达,或由在步骤(b)中测定表1中限定的一种或多种生物标记物的蛋白质的表达组成。该方法可以包括使用一种或多种结合部分在步骤(b)中测量一种或多种生物标记物的表达,所述结合部分各自能够选择性地结合表1中鉴定的生物标记物之一。一种或多种结合部分可以包括抗体或其抗原结合片段如单克隆抗体或其片段,或由抗体或其抗原结合片段如单克隆抗体或其片段组成。

[0075] 术语“抗体”包括任何合成的抗体、重组抗体或抗体杂合物,如但不限于,由免疫球

蛋白轻链和 / 或重链可变区和 / 或恒定区的噬菌体展示产生的单链抗体分子,或在本领域技术人员已知的免疫测定形式中能够结合抗原的其他免疫相互作用分子。

[0076] 本发明还包括使用抗体样结合剂,如亲和体 (affibody) 和适体。

[0077] 涉及保留了其特异性结合位点的抗体片段的合成的技术的综述可以在 Winter&Milstein(1991)Nature 349,293-299 中找到。

[0078] 另外地,或替换地,第一个结合分子中的一个或多个可以是适体 (参见 Collett 等,2005,Methods 37:4-15)。

[0079] 分子文库,如抗体文库 (Clackson 等,1991,Nature 352,624-628 ;Marks 等,1991, J Mol Biol 222(3):581-97)、肽文库 (Smith,1985, Science 228(4705):1315-7)、表达的 cDNA 文库 (Santi 等,(2000)J Mol Biol 296(2):497-508)、抗体框架以外的其他支架 (如亲和体) 上的文库 (Gunneriusson 等,1999,Appl Environ Microbiol 65(9):4134-40) 或基于适体的文库 (Kenan 等,1999,Methods Mol Biol 118,217-31),可以用作下述来源,从所述来源选择用于本发明方法中的对于给定基序特异性的结合分子。

[0080] 分子文库可以在原核细胞 (Clackson 等,1991,op.cit. ;Marks 等,1991,op.cit.) 或真核细胞 (Kieckhefer 等,1999,Proc Natl Acad Sci USA,96(10):5651-6) 中体内表达或可以在体外表达,而没有细胞的涉及 (Hanes&Pluckthun,1997,Proc Natl Acad Sci USA 94(10):4937-42 ;He&Taussig,1997, Nucleic Acids Res25(24):5132-4 ;Nemoto 等,1997, FEBS Lett,414(2):405-8)。

[0081] 在使用基于蛋白质的文库时的情况中,常常将编码潜在结合分子的文库的基因包装在病毒中,并且在病毒表面展示潜在的结合分子 (Clackson 等,1991, supra ;Marks 等, 1991, supra ;Smith,1985, supra)。

[0082] 可能,最常使用的展示系统是在其表面上展示抗体片段的丝状噬菌体,所述抗体片段作为与噬菌体的次要外壳蛋白的融合表达 (Clackson 等,1991, supra ;Marks 等,1991, supra)。然而,用于展示的其他合适系统包括使用其他病毒 (EP 39678)、细菌 (Gunneriusson 等,1999, supra ;Daugherty 等,1998, Protein Eng 11(9):825-32 ; Daugherty 等,1999, Protein Eng 12(7):613-21) 和酵母 (Shusta 等,1999, J Mol Biol 292(5):949-56)。

[0083] 此外,已经利用所谓核糖体展示系统中的多肽产物与其编码 mRNA 的连接 (Hanes&Pluckthun,1997, supra ;He&Taussig,1997, supra ;Nemoto 等,1997, supra),或替换地,多肽产物与编码 DNA 的连接 (参见美国专利 5,856,090 和 WO 98/37186),研发了展示系统。

[0084] 抗体的可变重链 (V_H) 和可变轻链 (V_L) 域涉及抗原识别,这是首先在早先蛋白酶消化实验中认识到的事实。通过啮齿动物抗体的“人源化”发现了更多证实。啮齿动物来源的可变域可以与人来源的恒定域融合,使得所得到的抗体保留啮齿动物亲本抗体的抗原特异性 (Morrison 等,(1984)Proc.Natl.Acad.Sci.USA 81,6851-6855)。

[0085] 从涉及均含有一个或多个可变域的抗体片段的细菌表达的实验已知,通过可变域给予抗原特异性并且与恒定域无关。这些分子包括 Fab- 样分子 (Better 等,(1988)Science 240,1041) ;Fv 分子 (Skerra 等,(1988)Science 240,1038) ;单链 Fv (ScFv) 分子,其中 V_H 和 V_L 亲本域经由弹性寡核苷酸连接 (Bird 等 (1988)Science 242,423 ;Huston 等 (1988)

Proc. Natl. Acad. Sci. USA 85,5879) 和包含分离的 V 域的单域抗体 (dAb) (Ward 等 (1989) Nature 341,544)。涉及保留其特异性结合位点的抗体片段的合成的技术的综述可以在 Winter&Milstein(1991)Nature 349,293-299 中找到。

[0086] 抗体或抗原结合片段可以选自完整抗体、Fv 片段 (例如,单链 Fv 和二硫化物键合的 Fv)、Fab- 样片段 (例如,Fab 片段、Fab' 片段和 F(ab)₂片段)、单个可变域 (例如,VH 和 VL 域) 和域抗体 (dAb,包括单一和双重形式 [即, dAb- 连接物 -dAb])。优选,抗体或抗原结合片段是单链 Fv(scFv)。

[0087] 一种或多种结合部分可以替换地包含抗体样结合剂例如亲和体或适体,或由抗体样结合剂例如亲和体或适体组成。

[0088] “scFv 分子”是指其中 V_H和 V_L配偶体域通过弹性寡肽连接的分子。

[0089] 使用抗体片段,而不是完整抗体的优点是几倍的。片段的较小尺寸可以导致提高的药理学特性,如更好地渗透实体组织。除去了完整抗体的效应物功能,如补体结合。Fab、Fv、ScFv 和 dAb 抗体片段全部可以在大肠杆菌中表达并由其分泌出来,因此允许容易地生产大量的所述片段。

[0090] 完整抗体和 F(ab')₂片段是“二价的”。“二价”是指所述抗体和 F(ab')₂片段具有两个抗原结合位点。相反,Fab、Fv、ScFv 和 dAb 片段是单价的,只具有一个抗原结合位点。

[0091] 抗体可以是单克隆或多克隆的。合适的单克隆抗体可以通过已知技术来制备,例如,“Monoclonal Antibodies:A manual of techniques”,H Zola(CRC Press,1988) 和 “Monoclonal Hybridoma Antibodies:Techniques and applications”,J G R Hurrell(CRC Press,1982) 中公开的那些,通过引用的方式将这两篇并入本申请。

[0092] 当潜在结合分子选自文库时,通常使用具有限定基序的一个或多个选择剂肽。在用于选择剂肽的基序的设计中,可以使用提供结构从而降低肽中灵活性的氨基酸残基或允许与结合分子相互作用的带电的、极性或疏水性侧链。例如:

[0093] (i) 脯氨酸可以稳定肽结构,因为其侧链对 α 碳以及氮都能结合;

[0094] (ii) 苯丙氨酸、酪氨酸和色氨酸具有芳香族侧链并且是高度疏水性的,而亮氨酸和异亮氨酸具有脂肪族侧链并且也是疏水性的;

[0095] (iii) 赖氨酸、精氨酸和组氨酸具有碱性侧链,并且在中性 pH 下带正电,而天冬氨酸盐和谷氨酸盐具有酸性侧链并且在中性 pH 下是带负电的;

[0096] (iv) 天冬氨酸和谷氨酸在中性 pH 下是中性的,但含有参与氢键的酰胺基团;

[0097] (v) 丝氨酸、苏氨酸和酪氨酸侧链含有羟基基团,其可以参与氢键。

[0098] 通常,结合分子的选择可以涉及阵列技术和系统的使用,以分析与对应于结合分子类型的点的结合。

[0099] 一种或多种蛋白结合部分可以包括可检测部分。可检测部分可以选自荧光部分、发光部分、化学发光部分、放射性部分和酶部分。

[0100] 在本发明方法进一步的实施方案中,可以使用包括能够结合一种或多种蛋白质的第二种结合剂的测定来进行步骤 (b),所述第二种结合剂还可以包括可检测部分。相对于以上的第一种结合剂详细描述了合适的第二种结合剂。

[0101] 因此,可以首先使用第一种结合剂分离和 / 或固定待测试样品中的目标蛋白质,此后,可以使用第二种结合剂来测定所述生物标记物的存在和 / 或相对含量。

[0102] 在一个实施方案中,第二种结合剂是抗体或其抗原结合片段;通常是重组抗体或其片段。方便地,抗体或其片段选自:scFv;Fab;免疫球蛋白分子的结合域。以上详细描述了合适的抗体和片段,及其制备方法。

[0103] 或者,第二种结合剂可以是抗体样结合剂,如亲和体或适体。

[0104] 或者,在待测试样品中的蛋白质上的可检测部分包括特定结合对的成员(例如,生物素)或由特定结合对的成员(例如,生物素)组成的情况中,第二种结合剂可以包括特定结合对的互补成员(例如,抗生物素蛋白链菌素)或由特定结合对的互补成员(例如,抗生物素蛋白链菌素)组成。

[0105] 在使用检测测定的情况中,优选可检测部分选自:荧光部分;发光部分;化学发光部分;放射性部分;酶部分。以上描述了用于本发明方法中的合适的可检测部分的实例。

[0106] 用于检测血清或血浆蛋白的优选测定包括酶联免疫吸附测定(ELISA)、放射性免疫测定(RIA)、免疫放射性测定测定(IRMA)和免疫酶测定(IEMA),包括使用单克隆和/或多克隆抗体的夹心测定。示例性夹心测定描述于David等的美国专利4,376,110和4,486,530中,通过引用的方式并入本申请。载玻片上的细胞的抗体染色可以用于细胞学实验室诊断测试中公知的以及本领域技术人员已知的方法中。

[0107] 因此,在一个实施方案中,测定是ELISA(酶联免疫吸附测定),其通常涉及使用产生有色反应产物的酶,通常在固相测定中。如辣根过氧化物酶和磷酸酶这样的酶已经被广泛使用。扩大磷酸酶反应的一种方式是使用NADP作为底物,来产生NAD,其现在作为用于第二个酶系统的辅酶。因为组织中不存在酶,因此来自大肠杆菌的焦磷酸酶提供了良好的缀合物,其是稳定的并且产生了良好的反应颜色。还可以使用基于酶(如,萤光素酶)的化学发光系统。

[0108] 常常使用与维生素生物素的缀合,因为这可以通过其与酶联抗生物素蛋白或抗生物素蛋白链菌素的反应容易地检测出来,其以很高的特异性和亲和力与酶联抗生物素蛋白或抗生物素蛋白链菌素结合。

[0109] 在可替换的实施方案中,用于蛋白质检测的测定方便地是荧光测定。因此,第二种结合剂的可检测部分可以是荧光部分,如Alexa荧光团(例如,Alexa-647)。

[0110] 优选,使用阵列进行步骤(b)。阵列可以是基于珠子的阵列或基于表面的阵列。优选,阵列选自:大阵列;微阵列;纳米阵列。

[0111] 在一个实施方案中,方法是用于鉴定能够诱导呼吸道超敏反应的物质。优选,超敏反应是体液超敏反应,例如,I型超敏反应。优选,方法是用于鉴定能够诱导呼吸道过敏的物质。

[0112] 在一个实施方案中,树突细胞群或树突样细胞群是树突细胞群。优选,树突细胞是初级树突细胞(primary dendritic cell)。优选,树突细胞是骨髓树突细胞(myeloid dendritic cell)。

[0113] 树突细胞或树突样细胞群优选是哺乳动物来源的。优选,哺乳动物是大鼠、小鼠、豚鼠、猫、狗、马或灵长类。最优选,哺乳动物是人。

[0114] 在一个实施方案中,树突细胞群或树突样细胞群是树突样细胞群,优选骨髓树突样细胞。

[0115] 在一个实施方案中,树突样细胞表达至少一种选自CD54、CD86、CD80、HLA-DR、

CD14、CD34 和 CD1a 的标记物,例如,2、3、4、5 或 7 种标记物。在进一步的实施方案中,树突样细胞表达标记物 CD54、CD86、HLA-DR、CD14、CD34 和 CD1a。

[0116] 在进一步的实施方案中,树突样细胞可以源自骨髓树突细胞。优选,树突样细胞是骨髓性白血病 (myeloid leukaemia) 衍生的细胞。优选,骨髓性白血病衍生的细胞选自 KG-1、THP-1、U-937、HL-60、Monomac-6、AML-193 和 MUTZ-3。最优选,树突样细胞是 MUTZ-3 细胞。MUTZ-3 细胞是人急性骨髓性白血病细胞,其在 1995 年 5 月 15 日依据保藏号 ACC195 由 Deutsche Sammlung für Mikroorganismen und Zellkulturen GmbH(DSMZ), Inhoffenstraße 7B, Braunschweig, Germany (www.dsmz.de) 可得。

[0117] 在一个实施方案中,树突样细胞,用细胞因子刺激后,通过 CD1d、I 和 II 类 MHC 呈递抗原和 / 或诱导特异性 T- 细胞增殖。

[0118] 在一个实施方案中,一种或多种阴性对照剂包括一种或多种选自 1- 丁醇、4- 氨基苯甲酸、氯苯、二甲基甲酰胺、乙基香草醛、异丙醇、水杨酸甲酯、丙二醇、高锰酸钾、吐温 80™(聚氧乙烯 (20) 山梨聚糖单油酸酯) 和硫酸锌的物质 (即,表 2 中限定的非致敏剂的基团) 或由以上物质组成。因此,步骤 (c) 可以包括将分离的树突细胞群或树突样细胞群暴露于表 2 中限定的每种阴性对照剂。

[0119] 该方法可以包括使用至少 2 种阴性对照剂 (即,非致敏剂),例如,至少 3、4、5、6、7、8、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99 或至少 100 种阴性对照剂;或由使用至少 2 种阴性对照剂 (即,非致敏剂),例如,至少 3、4、5、6、7、8、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99 或至少 100 种阴性对照剂组成。

[0120] 在另一个实施方案中,一种或多种阳性对照剂包括一种或多种选自六氯铂酸铵、过硫酸铵、戊二醛、六亚甲基二异氰酸酯、马来酸酐、亚甲基联苯酚二异氰酸酯、邻苯二甲酸酐、甲苯二异氰酸酯和偏苯三酸酐 (即,表 2 中限定的致敏剂组) 的物质,或由一种或多种选自六氯铂酸铵、过硫酸铵、戊二醛、六亚甲基二异氰酸酯、马来酸酐、亚甲基联苯酚二异氰酸酯、邻苯二甲酸酐、甲苯二异氰酸酯和偏苯三酸酐 (即,表 2 中限定的致敏剂组) 的物质组成。因此,步骤 (d) 可以包括将分离的树突细胞群或树突样细胞群暴露于表 2 中限定的每种阳性对照剂,或由将分离的树突细胞群或树突样细胞群暴露于表 2 中限定的每种阳性对照剂组成。

[0121] 该方法可以包括使用至少 2 种阳性对照剂 (即,致敏剂),例如,至少 3、4、5、6、7、8、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99 或至少 100 种阳性对照剂;或由使用至少 2 种阳性对照剂 (即,致敏剂),例如,至少 3、4、5、6、7、8、10、11、12、13、14、15、16、17、18、

19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99 或至少 100 种阳性对照剂组成。

[0122] 因此,在一个实施方案中,该方法表示测试物质是否是呼吸道致敏剂。在替换的或另外的实施方案中,该方法表示待测试样品的呼吸道致敏效力。

[0123] 因此,在一个实施方案中,该方法表示测试物质的致敏效力(即,测试物质是非致敏剂、弱致敏剂、中等致敏剂、强致敏剂或极端致敏剂)。PCA 的判定值和距离与致敏效力相关。

[0124] 替换地或另外地,测试物质效力可以在步骤(e)中通过提供以下物质来确定:

[0125] (i) 一种或多种极端呼吸道致敏剂阳性对照剂;

[0126] (ii) 一种或多种强呼吸道致敏剂阳性对照剂;

[0127] (iii) 一种或多种中等呼吸道致敏剂阳性对照剂;和/或

[0128] (iv) 一种或多种弱呼吸道致敏剂阳性对照剂,

[0129] 其中在步骤(b)中测量的测试样品中的一种或多种生物标记物的存在和/或含量对应于步骤(f)中测量的极端阳性对照样品(在存在的情况下)中的一种或多种生物标记物的存在和/或含量的情况下;和/或不同于步骤(f)中测量的强、中等、弱和/或阴性对照样品(在存在的情况下)中的存在和/或含量的情况下,将测试物质鉴定为极端呼吸道致敏剂,

[0130] 其中在步骤(b)中测量的测试样品中的一种或多种生物标记物的存在和/或含量对应于步骤(f)中测量的强阳性对照样品(在存在的情况下)中的一种或多种生物标记物的存在和/或含量的情况下;和/或不同于步骤(f)中测量的极端、中等、弱和/或阴性对照样品(在存在的情况下)中的存在和/或含量的情况下,将测试物质鉴定为强呼吸道致敏剂,

[0131] 其中在步骤(b)中测量的测试样品中的一种或多种生物标记物的存在和/或含量对应于步骤(f)中测量的中等阳性对照样品(在存在的情况下)中的一种或多种生物标记物的存在和/或含量的情况下;和/或不同于步骤(f)中测量的极端、强、弱和/或阴性对照样品(在存在的情况下)中的存在和/或含量的情况下,将测试物质鉴定为中等呼吸道致敏剂,

[0132] 其中在步骤(b)中测量的测试样品中的一种或多种生物标记物的存在和/或含量对应于步骤(f)中测量的弱阳性对照样品(在存在的情况下)中的一种或多种生物标记物的存在和/或含量的情况下;和/或不同于步骤(f)中测量的极端、强、中等和/或阴性对照样品(在存在的情况下)中的存在和/或含量的情况下,将测试物质鉴定为弱呼吸道致敏剂。

[0133] 因此,步骤(e)可以包括提供以下类别的呼吸道致敏剂阳性对照或由提供以下类别的呼吸道致敏剂阳性对照组成:

[0134] (a) 极端、强、中等和弱;

[0135] (b) 强、中等和弱;

[0136] (c) 极端、中等和弱;

[0137] (d) 极端、强和中等；

[0138] (e) 极端和强；

[0139] (f) 强和中等；

[0140] (g) 中等和弱；

[0141] (h) 强和弱；

[0142] (i) 极端和中等；

[0143] (j) 极端和弱；

[0144] (k) 极端；

[0145] (l) 强；

[0146] (m) 中等；

[0147] (n) 弱。

[0148] 基于对人的临床观察,阴性对照和阳性对照可以归类为呼吸道非致敏剂或呼吸道致敏剂。

[0149] 替换地或另外地,该方法可以包括将步骤 (b) 中测量的一种或多种生物标记物的表达与表示步骤 (c) 和 / 或步骤 (e) 中测量的一种或多种生物标记物的表达的一个或多个预定参照值进行比较。

[0150] 通常,用至少 0.55 的 ROC AUC 来确定呼吸道致敏剂,例如,用至少 0.60、0.65、0.70、0.75、0.80、0.85、0.90、0.95、0.96、0.97、0.98、0.99 的 ROC AUC 或用 1.00 的 ROC AUC 来确定呼吸道致敏剂。优选,用至少 0.85 的 ROC AUC,并且最优选用 1 的 ROC AUC 来确定皮肤致敏剂。

[0151] 通常,使用支持向量仪 (SVM, support vector machine),如可从 <http://cran.r-project.org/web/packages/e1071/index.html> (例如, e10711.5-24) 获得的那些,来鉴定能够诱导呼吸道致敏的物质。然而,也可以使用任何其他合适的方式。SVM 也可以用于测定包括如本文中限定的一种或多种表 1 生物标记物或由本文中限定的一种或多种表 1 生物标记物组成的生物标记物标签 (biomarker signature) 的 ROC AUC。

[0152] 支持向量仪 (SVM) 是用于分类和回归的一组相关管理学习方法。给定一组训练实例,每个标记为属于两类中的一类,SVM 训练算法构建预测新的实例是否落入一类或另一类的模型。直观地,SVM 模型是作为空间中的点的实例表示,作图,使得分开类别的实例通过明显的尽可能宽的间隙分开。然后将新的实例作图至相同的空间中,并且基于它们落入间隙的哪一侧来预测属于哪一类。

[0153] 更规范地,支持向量仪在高或无穷的维度空间中构建了超平面或超平面组,其可以用于分类、回归或其他任务。直观地,通过对任何类别的最接近训练数据点具有最大距离的超平面来实现良好的分离 (称为功能边缘),因为通常边缘越大,分类器的泛化误差越低。关于 SVM 的更多信息,参见,例如, Burges, 1998, Data Mining and Knowledge Discovery, 2:121-167。

[0154] 在本发明的一个实施方案中,在使用已知物质 (即,已知致敏剂或非致敏剂) 的生物标记物谱进行本发明的方法之前,对 SVM “进行训练”。通过运行这样的训练样品,SVM 能够知道何种生物标记物谱与能够诱导致敏的物质相关。一旦完成训练过程,然后 SVM 能够确定测试的生物标记物样品是否来自致敏剂或非致敏剂。

[0155] 这允许将测试物质归类为致敏或非致敏。此外,通过用已知效力的致敏剂(即,非致敏、弱、中等、强或极端致敏剂)训练 SVM,还可以比较地鉴定测试物质的效力。

[0156] 然而,可以通过用必需的训练参数对 SVM 进行预先编程来绕过这种训练程序。例如,基于表 1 中所列的所有生物标记物的测量,使用表 3 中详述的 SVM 算法,根据已知的 SVM 参数来鉴定能够诱导致敏的物质。

[0157] 本领域技术人员将认识到可以通过用合适的数据选择(即,来自暴露于已知致敏和/或非致敏剂的细胞的生物标记物测量)训练 SVM 仪器,针对表 1 中所列的生物标记物的任意组合,来确定合适的 SVM 参数。或者,根据本领域已知的任何其他合适的统计学方法,表 1 生物标记物可以用于鉴定能够诱导呼吸道致敏的物质。

[0158] 或者,根据本领域已知的任何其他合适的统计学方法(例如,AVONA、ANCOVA、MANOVA、MANCOVA、多变量回归分析、主成分分析(PCA)、因子分子、典型相关分析、典型相关分析、多余性分析、对应分析(CA;互相平均)、多维定标、判别分析、线性判别分析(LDA)、群集系统、递归分类和人工神经网络),表 1 数据可以用于鉴定能够诱导呼吸道致敏的物质。

[0159] 优选,本发明的方法具有至少 65% 的准确性,例如,66%,67%,68%,69%,70%,71%,72%,73%,74%,75%,76%,77%,78%,79%,80%,81%,82%,83%,84%,85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%,99% 或 100% 的准确性。

[0160] 优选,本发明的方法具有至少 65% 的灵敏度,例如,66%,67%,68%,69%,70%,71%,72%,73%,74%,75%,76%,77%,78%,79%,80%,81%,82%,83%,84%,85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%,99% 或 100% 的灵敏度。

[0161] 优选,本发明的方法具有至少 65% 的特异性,例如,66%,67%,68%,69%,70%,71%,72%,73%,74%,75%,76%,77%,78%,79%,80%,81%,82%,83%,84%,85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%,99% 或 100% 的特异性。

[0162] “准确性”是指方法的正确结果的比例,“灵敏度”是指正确归类为阳性的所有阳性化学物质的比例,和“特异性”是指正确归类为阴性的所有阴性化学物质的比例。

[0163] 在一个实施方案中,本发明的第一个方面的方法包括同时或连续地进行用于鉴定能够诱导哺乳动物皮肤致敏的物质的方法,如 PCT 公开 W02012/056236 中所述的,将其按引用并入本文中。优选,用于鉴定能够诱导哺乳动物皮肤致敏的物质的方法与本发明的第一个方面的方法同时进行(即,通过测量暴露于测试物质的相同细胞样品中的相关标记物表达来确定测试化合物是否是皮肤和/或呼吸道致敏剂)。

[0164] 本发明的第二个方面提供了用于本发明第一个方面的方法中的阵列(或其任何实施方案或实施方案的组合),所述阵列包括一种或多种如上限定的结合部分。在一个实施方案中,结合部分(共同地)能够结合表 1A 中限定的全部生物标记物。在进一步的实施方案中,结合部分(共同地)能够结合表 3B 中限定的全部生物标记物。在再进一步的实施方案中,结合部分(共同地)能够结合表 3B 中限定的全部生物标记物。优选,结合部分(共同地)能够结合表 1 中限定的全部生物标记物。

[0165] 结合部分可以是固定的。

[0166] 阵列本身是本领域公知的。通常，它们由具有隔开（即，分离）区域（“点”）的线性或二维结构形成，每个具有有限的面积，在固体支持物的表面上形成。阵列还可以是珠子结构，其中每个珠子通过分子编码或颜色编码来识别或在连续流动中识别。在样品通过一系列点的情况下，还可以按序进行分析，每个点从溶液中吸附所述类别的分子。固体支持物通常是玻璃或聚合物，最常用的聚合物是纤维素、聚丙烯酰胺、尼龙、聚苯乙烯、聚氯乙烯或聚丙烯。固体支持物可以是管子、珠子、圆盘、硅芯片、微板、聚乙二烯二氟化物（PVDF）膜、硝基纤维素膜、尼龙膜、其他多孔膜、无孔膜（例如，塑料、聚合物、塑胶、硅等）、多个聚合针，或多个微滴定孔，或任何其他适用于固定蛋白质、多核苷酸和其他合适分子和 / 或进行免疫测定的表面的形式。结合过程也是本领域公知的，并且通常由交联共价结合或物理吸附蛋白质分子、多核苷酸等至固体支持物上组成。或者，可以使用经由亲和力和标签（tag）或相似构建体的探针的亲和力偶联。通过使用公知技术，如接触或非接触印刷、掩蔽或照相平版印刷术，可以限定每个点的位置。对于综述，参见，Jenkins, R. E., Pennington, S. R. (2001, *Proteomics*, 2, 13-29) 和 Lal 等 (2002, *Drug Discov Today* 15 ; 7 (18Suppl) : S143-9)。

[0167] 通常，阵列是微阵列。“微阵列”包括具有至少约 $100/\text{cm}^2$ 的离散区域密度的区域的阵列的含义，并且优选至少 $1000/\text{cm}^2$ 。微阵列中的区域具有典型的尺寸，例如，直径，在约 $10\text{--}250\ \mu\text{m}$ 的范围中，并且与阵列中的其他区域隔开约相同的距离。阵列可替换地是大阵列或纳米阵列。

[0168] 一旦已经鉴定并分离合适的结合分子（以上讨论的），本领域技术人员可以使用分子生物学领域公知的方法来制造阵列；参见以下的实施例。

[0169] 本发明的第三个方面提供了一种或多种（优选两种或多种）选自表 1A、表 1B 和 / 或表 1C 中的生物标记物组合用于鉴定超敏反应致敏剂的用途。优选，表 1A 和表 1B 中限定的全部生物标记物共同地用于鉴定超敏应答致敏剂。优选，所述用途与本发明的第一个方面中所述的方法以及其中所述的实施方案相一致。

[0170] 本发明的第四个方面提供了用于根据本发明的第一个方面的方法中的分析试剂盒，其包括：

[0171] A) 根据本发明第二个方面的阵列；和

[0172] B) 用于进行根据本发明第一个方面的方法的说明书（任选）。

[0173] 分析试剂盒可以包括一种或多种对照剂。优选，分析试剂盒包括以上特征以及一种或多种阴性对照剂和 / 或一种或多种阳性对照剂，或由以上特征以及一种或多种阴性对照剂和 / 或一种或多种阳性对照剂组成。

[0174] 本发明的第五个方面提供了一种治疗或预防患者呼吸道 I 型超敏反应（如哮喘）的方法，其包括以下步骤：

[0175] (a) 提供一种或多种暴露或已经暴露于患者的测试物质；

[0176] (b) 使用本发明第一个方面中提供的方法确定步骤 (a) 中提供的一种或多种测试物质是否是呼吸道致敏剂；和

[0177] (c) 在一种或多种测试物质鉴定为呼吸道致敏剂的情况下，降低或防止患者暴露于一种或多种鉴定为呼吸道致敏剂的测试物质。

[0178] 优选，一种或多种暴露或已经暴露于患者的测试物质是目前暴露于患者至少一个月一次的物质，例如，至少每两周一次，至少每周一次，或至少每天一次。

[0179] 现在将参照以下附图描述实施本发明特定方面的优选的非限制性实施例。

[0180] 图 1: 呼吸道致敏的潜在生物标记物的后向消除 (backward elimination)。将通过 p - 值分选选择的 1029 个基因用作输入信息。消除 727 个基因后, 观察到 KLD 中的局部极小值。因此, 剩余的 302 个基因共同地包含了分离呼吸道致敏剂和非致敏剂相关的大部分信息。将这种生物标记物标签称为 GARD 呼吸道预测标签。

[0181] 图 2: 基于通过 p - 值过滤和后向消除选择的 302 个转录产物的主成分分析。观察到用呼吸道致敏剂 (蓝色) 和非致敏剂 (绿色) 刺激的样品之间的完全分离。

[0182] 图 3: 使用交叉验证的 GARD 呼吸道预测标签的预测力的估算。使用 70% 的随机选择的数据 (训练集) 构建了标签。随后将验证生物标记物标签用于将剩余的 30% 数据 (测试集) 中的样品分类。A) 测试集中样品的 SVM 预测后的 ROC AUC 分布。B) 用主成分分析说明的预测性能的代表性图示。

[0183] 图 4: 化学刺激后 MUTZ-3 细胞的 CD86 表达。显示的数据是 (化学刺激, $n = 3$, DMSO 和未刺激细胞 ($n = 6$)) 的平均值, 误差棒显示了标准偏差。通过 student's t - 检验确定了统计学显著性, 将每次刺激与其相应的媒介物进行比较, 通过 * 来表示 $p < 0.05$ 。

[0184] 图 5: 预测性生物标记物标签的确立。A) 基于通过 p - 值过滤选择的 1029 个转录产物的主成分分析。B) 基于通过后向消除选择的 302 个转录产物的主成分分析。将样品着色为呼吸道致敏剂 (蓝色, $n = 27$) 或非致敏剂 (绿色, $n = 47$)。表示了由 74 个样品组成的所有数据, 包括所有复制物。C) 根据它们的机械子域, 将呼吸道致敏剂着色。

[0185] 图 6: 图 3. 使用外部测试集和交叉验证估算了 GPRS 的预测力。A) 将由三份非致敏剂组成的外部数据集作图于由 GPRS 构成的 PCA 空间中。只允许训练数据影响主成分。表示了样品的所有复制物 (训练数据 $n = 74$, 测试数据 $n = 48$)。B) 使用 70% 的随机选择的数据 (训练集) 构建了验证生物标记物标签。将训练集用于构件 PCA 空间, 使用验证生物标记物标签作为可变输入信息。将剩余的 30% 数据 (测试集) 作图于这个空间中, 没有允许其影响主成分。C) 将数据随机分成训练集和测试集重复 20 次。用盒形图报道 ROC AUC 分布, 其中跳动的数据点覆盖。中值 ROC AUC 为 0.84。

实施例

[0186] 介绍

[0187] 对低分子量化合物的呼吸道致敏是职业性哮喘的主要诱因, 这已经与致命的结果相关。为了防止呼吸道化学致敏剂的出现和最小化工作环境中的风险, 需要努力来研发预测化合物诱导呼吸道致敏的能力的测定。然而, 迄今为止, 不存在在体外或体内可靠地完成将化学物质精确地归类成呼吸道致敏剂的有效体外或体内方法。最近, 提出了一种新的用于评价皮肤致敏剂的体外测定, 称为 GARD (Johansson 等, 2011, BMC Genomics, 12:339)。发明人使用包括 302 个与导致树突细胞成熟的免疫事件相关的基因的新基因组生物标记物标签集, 已经将 GARD 的适用性延伸到也能够将呼吸道致敏剂分类。因此, 发明人提出了具有能够同时预测测定化合物的皮肤和呼吸道致敏能力的组合能力的测定。

[0188] 材料和方法

[0189] 化学物质

[0190] 一组 20 种化合物, 由 9 种呼吸道致敏剂和 11 种非致敏剂组成, 将其用于细胞刺

激。致敏剂是戊二醛、过硫酸铵、邻苯二甲酸酐、亚甲基联苯酚二异氰酸酯、六氯铂酸铵、偏苯三酸酐、六亚甲基二异氰酸酯、马来酸酐和甲苯二异氰酸酯。非致敏剂是氯苯、硫酸锌、4-氨基苯甲酸、水杨酸甲酯、乙基香草醛、异丙醇、二甲基甲酰胺、1-丁醇、高锰酸钾、丙二醇和吐温 80(表 2)。所有化学物质来自 Sigma-Aldrich, St. Louis, MO, USA。将化合物溶解于二甲亚砜(DMSO)或蒸馏水中。在刺激之前,使用制造商提供的实验方案,使用碘化丙啶(PI)(BD Biosciences, San Diego, CA),监测了所有化合物的细胞毒性。如下计算刺激的细胞的相对存活力:

[0191]

$$\text{相对存活力} = \frac{\text{存活的受刺激细胞的分数}}{\text{存活的未受刺激细胞的分数}} \cdot 100$$

[0192] 对于有毒的化合物,使用产生 90%相对存活力(Rv90)的浓度。对于无毒的化合物,使用了 500 μM 的浓度。对于在介质中以 500 μM 的浓度不溶的无毒化合物,使用最高可溶浓度。对于溶解于 DMSO 中的化合物,每个孔中的 DMSO 的终浓度为 0.1%。对于任何给定的化学物质使用的浓度称为“GARD 输入浓度”,并且列于表 2 中。

[0193] 细胞的化学物质暴露

[0194] 按照所述的(Johansson H,Lindstedt M,Albrekt AS,Borrebaeck CA:A genomic biomarker signature can predict skin sensitizers using a cell-based in vitro alternative to animal tests.BMC Genomics 2011,12:399;Rasaiyaah J,Yong K,Katz DR,Kellam P,Chain BM:Dendritic cells and myeloid leukaemias:plasticity and commitment in cell differentiation.Br J Haematol 2007,138(3):281-290),将人骨髓性白血病衍生的细胞系 MUTZ-3(DSMZ, Braunschweig, 德国)保持于补充有 20%(体积/体积)胎牛血清(Invitrogen, Carlsbad, CA)和 40ng/ml rhGM-CSF(Bayer HealthCare Pharmaceuticals, Seattle, WA)的 α -MEM(Thermo Scientific Hyclone, Logan, UT)中。扩大过程中,以 200,000 个细胞/ml 来维持培养物,每 3-4 天更换培养基。没有进行分化步骤,而是将增殖的祖细胞 MUTZ-3 用于刺激。在每个实验之前,使用流式细胞术将细胞进行免疫表型定型,作为质量控制。将细胞以 200,000 个细胞/ml 接种于 6-孔板中。在 DMSO 或蒸馏水中制备了每种化合物的储备溶液,并且随后稀释,使得孔内的浓度对应于 GARD 输入浓度,并且 DMSO 的孔内浓度为 0.1%。将细胞在 37°C 和 5% CO₂ 孵育 24h。此后,收集细胞并且通过流式细胞术分析。平行地,将收集的细胞在 TRIzol 试剂(Invitrogen)中裂解,并且储存在 -20°C,直至 RNA 提取。在三个独立实验中进行了用化学物质的刺激,使得获得了重复三份的样品。

[0195] 使用流式细胞术的表型分析

[0196] 所有细胞表面染色和洗涤步骤在含有 1% BSA(w/v) 的 PBS 中进行。用特定的小鼠 mAb 在 4°C 将细胞孵育 15min。将以下的 mAb 用于流式细胞术:FITC-缀合的 CD1a(DakoCytomation, Glostrup, 丹麦)、CD34、CD86 和 HLA-DR(BD Biosciences)、PE-缀合的 CD14(DakoCytomation)、CD54 和 CD80(BD Biosciences)。将缀合 FITC 或 PE 的小鼠 IgG1 用作同种型对照(BD Biosciences),并将 PI 用于评价细胞存活力。将 FACSDiva 软件用于数据获取,使用 FACSCanto II 仪器(BD Bioscience)。获取了 10,000 个事件,并且基

于光散射特性设定了门控,以排除碎片和不能存活的细胞。使用 FCS Express V3(De Novo Software, Los Angeles, CA) 进行了进一步的数据分析。

[0197] 表型分析、化学暴露、细胞收集和 RNA 分离

[0198] 按照之前所述的 (Johansson H, Albrekt AS, Borrebaeck CA, Lindstedt M (2012) The GARD assay for assessment of chemical skin sensitizers. *Toxicol in Vitro*), 进行了 MUTZ-3 的维持和化学刺激以及所有随后的 RNA 分离和 cDNA 的制备。简而言之,在化学刺激之前,进行了 MUTZ-3 的表型控制。收集经刺激的细胞并且分离 RNA。通过 CD86 的流式细胞术分析进行了细胞的成熟状态的控制。根据制造商 (Affymetrix) 提供的标准实验方案,进行了 cDNA 的制备和杂交,洗涤和 Human Gene 1.0ST 阵列 (Affymetrix, Santa Clara, CA, USA) 的扫描。

[0199] 微阵列数据分析和统计学方法

[0200] 之前已经描述了通过其确定预测标签的方法 (Johansson H, Lindstedt M, Albrekt AS, Borrebaeck CA (2011) A genomic biomarker signature can predict skin sensitizers using a cell-based in vitro alternative to animal tests. *BMC Genomics* 12:399)。简而言之,将微阵列数据标准化并且用 RMA 算法检查质量,其使用 Affymetrix Expression Console (Affymetrix)。通过来自 ANOVA 的 p- 值选择了头 1029 个预测剂,比较呼吸道致敏剂和非致敏剂。对头 1029 个预测剂应用了用于后向消除的算法 (Johansson 等, 2011, 上文; Carlsson A, Wingren C, Kristensson M, Rose C, Ferno M 等, (2011) Molecular serum portraits in patients with primary breast cancer predict the development of distant metastases. *Proc Natl Acad Sci U S A* 108:14252-14257), 以进一步减少生物标记物标签大小。对后向消除算法进行修改,以最小化 Kullback-Leibler 误差 (Kullback S, Leibler RA (1951) On Information and Sufficiency. *Annals of Mathematical Statistics* 22:79 - 86), 而不是最大化接收者操作特征下面积 (ROC AUC) (Lasko TA, Bhagwat JG, Zou KH, Ohno-Machado L (2005) The use of receiver operating characteristic curves in biomedical informatics. *J Biomed Inform* 38:404-415), 使得能够在其中 ROC AUC 达到 1.0 的情况下持续标签优化。将选择的头 302 个预测剂共同地称为“GARD 呼吸道预测标签”(GPRS)。将用于后向消除的脚本在 R 中编程 (R Development Core Team (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria), 其使用另外的软件包 e1071 (Weingart SN, Iezzoni LI, Davis RB, Palmer RH, Cahalane M 等 (2000) Use of administrative data to find substandard care: validation of the complications screening program. *Med Care* 38:796-806)。使用 QluCore Omics Explorer 2.3 (QluCore AB, Lund, 瑞典) 进行了 ANOVA 分析和使用主成分分析 (PCA) 的结果的观察 (Ringner M (2008) What is principal component analysis? *Nat Biotechnol* 26:303-304)。按照所述的 (Johansson 等, 2011, 上文), 使用由阴性化学刺激组成的外部数据集以及基于支持向量仪 (SVM) 的交叉验证方法 (Noble WS (2006) What is a support vector machine? *Nat Biotechnol* 24:1565-1567), 估算了 GPRS 的预测性能。使用 Ingenuity Pathways Analysis (IPA) (Ingenuity Systems, Inc. Mountain View, USA), 通过进行“核心分析 (Core Analysis)”, 探索了 GPRS 的生物相关性。将头 1029 个基因连同倍数变化值用作 IPA 输入

信息。通过探测与输入分子相关的标准途径,确定了生物相关性。阵列数据已经依据登记号 E-MEXP-3773 上传至 ArrayExpress(<http://www.ebi.ac.uk/arrayexpress/>)。

[0201] 用于鉴定预测标签的方法的询问

[0202] 将数据集分成训练集和测试集,分别由 70%和 30%的化合物组成。随机进行分离,同时将每个子集中致敏剂和非致敏剂的比例保持在与完整数据集中的相同比例。按照以上所述的,使用 ANOVA 过滤和后向消除,在训练集中鉴定出测试生物标记物标签。使用该训练集,将这个测试标签用于训练支持向量仪 (SVM) (Noble WS:What is a support vector machine? Nat Biotechnol 2006, 24(12):1565-1567),此后将该训练集用于预测测试集的样品。将该过程重复 20 次,并且将接收者操作特征下面积 (ROC AUC) 的分布 (Lasko TA, Bhagwat JG, Zou KH, Ohno-Machado L:The use of receiver operating characteristic curves in biomedical informatics. J Biomed Inform 2005, 38(5):404-415) 用作模型性能的测量。

[0203] 结果

[0204] 化学刺激的 MUTZ-3 细胞中的转录谱的分析

[0205] 用一组参照化学物质刺激 24h 后,将来自 MUTZ-3 的 mRNA 收集用于转录谱测定。刺激包括 9 种不同的化学呼吸道致敏剂和 11 种不同的非致敏剂,所有样品在生物学上重复三次,除了 4-氨基苯甲酸和高锰酸钾之外,由于内部对照,所述 4-氨基苯甲酸重复了 6 次取样,以及由于有缺陷的阵列,所述高锰酸钾只重复了 2 次取样。此外,DMSO 和蒸馏水各自重复 6 次取样,作为媒介物对照。总的来说,准备用于分析的数据集由 74 个阵列组成,其各自具有 29141 个转录产物的测量。

[0206] 分析的第一个步骤涉及基因的 p- 值过滤,其取决于它们分离呼吸道致敏剂与非致敏剂的能力,如通过比较两组的 ANOVA 确定的。基于之前的经验,大约 1000 个基因是合适量的潜在预测剂,以用作用于后向消除的算法中的输入信息 (Johansson H, Lindstedt M, Albrekt AS, Borrebaeck CA:A genomic biomarker signature can predict skin sensitizers using a cell-based in vitro alternative to animal tests. BMC Genomics 2011, 12:399)。使用 0.0067 的 p- 值截留 (FDR 19%),鉴定了 1029 个基因。应用了后向消除算法,除去了以重复方式引起最小信息的预测剂。当消除了 727 个预测剂时,观察到 Kullback-Liebler Divergence (KLD) 中的局部最小值 (图 1)。将剩余的 302 个基因总称为“GARD 呼吸道预测标签”,并且图 2 中说明了它们区分呼吸道化学致敏剂和非致敏剂的能力。

[0207] 用于鉴定预测标签的分析的询问

[0208] 为了验证本申请标签的预测力,发明人使用了称为支持向量仪 (SVM) 的机器学习方法 (Noble, 2006, 上文),其将来自训练集的数据作图于空间中,使得由致敏和非致敏化学物质引起的基因表达的分离最大化。作为训练集,随机选择 70%的数据集,并且重复生物标记物选择的整个过程。用 29,141 个转录产物开始,如上所述,使用 ANOVA 过滤和后向消除,将标签降至与 GARD 呼吸道预测标签相等大小的基因列表,即,302 个转录产物,称为“验证生物标记物标签”。使用验证生物标记物标签,在训练数据上训练 SVM。然后将训练过的 SVM 用于将剩余 30%数据 (即,测试集) 中的每个样品归类为呼吸道致敏剂或非致敏剂。用接受者操作特征下面积 (ROC AUC) 来评价分类的性能。将这个完整的交叉验证重复

20 次,每次产生不同的训练集和测试集,其中每个训练集产生不同的验证生物标记物标签。图 3 中说明了这些交叉验证的结果。发现 ROC AUC 的中值为 0.84,具有 0.66 至 0.96 的范围。预测性能中的大变化意味着随机排除 30% 的数据很大程度上影响了验证生物标记物标签的组成。然而,实现高达 0.96 的 ROC AUC 的能力是在所有可用数据上训练模型时精确的分类实际上是可能的强有力的证据。这种交叉验证证明了 GARD 呼吸道预测生物标记物能够精确地预测未知样品的呼吸道致敏特性。

[0209] 未刺激和刺激细胞中的 MUTZ-3 表型

[0210] 在化学挑战之前,通过使用流式细胞术测量常见骨髓和树突细胞标记物的细胞表达,对细胞进行质量控制。这些标记物包括 CD1a、CD14、CD34、CD54、CD80、CD86 和 HLA-DR。发现没有背离之前公开的数据 (Johansson 等,2011,上文),这确保未刺激的细胞成功地维持在未成熟状态。化学刺激后,再次控制细胞的一般成熟状态,如通过共同刺激标记物 CD86 的表达确定的,其结果呈现于图 4 中。多种化学刺激后,CD86 的上调是明显的,然而,由于大的标准偏差,只有戊二醛和六亚甲基二异氰酸酯导致了统计学上显著的 CD86 上调。此外,尽管不是统计学上显著的,在多种对照刺激后,CD86 的上调也是明显的。因此,发明人推断 CD86 对于呼吸道化学致敏剂不是合适的生物标记物。然而,在这个研究中用于刺激的许多化合物在细胞培养基中的溶解性差,并且不能以足以诱导细胞毒性的浓度来使用。为此,CD86 表达的提高可以作为补充工具来确保化学刺激的生物可用性。

[0211] 化学刺激的 MUYZ-3 细胞中的转录谱的分析

[0212] 24h 刺激后,使用一组参照化学物质,收集了来自 MUTZ-3 的 mRNA,用于转录谱测定。刺激包括 9 种不同的化学呼吸道致敏剂和 11 种不同的非致敏剂 (阴性对照),所有分析在生物学中重复三份,除了 4-氨基苯甲酸以及高锰酸钾之外,由于内部对照,所述 4-氨基苯甲酸重复了 6 次取样,以及由于有缺陷的阵列,所述高锰酸钾只重复了 2 次取样。此外,DMSO 和蒸馏水每个重复 6 次取样,作为媒介物对照。总的来说,准备用于分析的数据集由 74 个阵列组成,每个具有 29,141 个转录产物的测量。

[0213] 分析的第一个步骤涉及基因的 p- 值过滤,其取决于它们区分呼吸道致敏剂与非致敏剂的能力,如通过比较两组的 ANOVA 确定的。由于计算的限制,大约 1000 个基因是合适的潜在预测剂的量,用作后向消除算法中的输入信息。在本发明数据集中,这种预测剂候选物的预选获得了 1029 个基因,具有 0.0067 或更低的 p- 值,具有 19% 的假发现率 (FDR) (Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B 57:289 - 300)。共同地,这些基因能够分离呼吸道致敏剂和非致敏剂。然而,没有获得明显分离,如使用 3D 主成分分析 (PCA) 所示的 (图 5A)。通过由其 p- 值给定的等级,进一步减少预测剂的数量,其没有获得明显的分离,即使数据含有 p- 值降至 10^{-10} 的预测剂候选物。然后应用后向消除算法,除去引起最小信息的预测剂 (基因)。当消除了 727 个预测剂后,观察到 Kullback-Liebler Divergence (KLD) 中的局部最小值 (数据未显示)。将剩余的 302 个基因总称为“GARD 呼吸道预测标签 (GPRS)”,并且它们区分呼吸道化学致敏剂和非致敏剂的能力显示于图 5B 中。基因的鉴定列于表 1 中。

[0214] 值得注意的是,与非致敏剂组相比,呼吸道致敏剂组内存在显著较大的转录谱的变化。研究皮肤致敏剂时,也观察到了相似的现象,这与致敏剂的效力以及不同化学物质诱

导不同信号传导途径的倾向有关 (Johansson 等, 2011, 上文)。然而, 明确限定的致敏效率不适用于这些呼吸道化学致敏剂 (Basketter DA, Kimber I (2011) Assessing the potency of respiratory allergens: uncertainties and challenges. Regul Toxicol Pharmacol 61:365-372)。相反, 发明人旨在描述涉及每种化学致敏剂的机械子域的转录谱的差异 (Enoch SJ, Roberts DW, Cronin MT (2010) Mechanistic Category Formation for the Prediction of Respiratory Sensitization. Chem Res Toxicol)。图 5C 显示了与图 5B 中相同的 PCA 曲线, 根据其机械子域将致敏剂着色, 如表 2 中所列。铵盐趋于放置在更远离非致敏剂簇 (cluster) 的位置, 这表明就转录谱而言与非致敏剂的最大不同。然而, 二异氰酸酯和酸酐簇紧密关联在一起, 因此在这点上没有可能得出这两组之间存在任何不同的任何结论。据发明人所知, 戊二醛没有分配给机械子域, 尽管其分组与酸酐和二异氰酸酯都接近, 并且因此将这些样品称为“子域未知”(图 5C)。

[0215] 预测标签的预测精确性的评价

[0216] 以两种方式评价了 GPRS 的预测性能。首先, 基于 GPRS, 将由非致敏剂组成的外部测试集用于证实它们在 PCA 曲线中的位置。其次, 发明人使用了交叉验证方法, 其将数据随机分成训练和测试集, 然后将其用于训练和评价支持向量分类。

[0217] 由于另外的对照化学物质集的可利用性, 第一种方法可以进行, 其在之前的试验组中运行, 在该试验中首先考虑 GARD (Johansson 等, 2011, 上文)。这个测试集中的化合物是苯甲醛、氯苯、酞酸二乙酯、甘油、乳酸、辛酸、苯酚、水杨酸和十二烷基硫酸钠, 所有都是生物学重复三份取样。此外, 测试集分别含有九种 DMSO 样品和未刺激对照。图 6A 显示了与图 5B 相同的 PCA 曲线, 其中已经基于样品的转录谱对测试集进行了作图, 同时没有允许影响主要成分。测试集的所有样品正确地与训练集的非致敏剂归在一组。这个训练集缺乏呼吸道致敏剂是由于在进行用于建立 GPRS 的分析中, 发明人不愿放弃这些样品中的任一种。由于过度拟合 (over fitting) 的风险, 这个分析中包括的任一种样品不适于包括在测试集中。

[0218] 为了克服在真实的测试集中没有呼吸道致敏剂的问题, 发明人使用了交叉验证的方法。作为训练集, 随机选择 70% 的数据集, 并且重复生物标记物选择的整个过程。如上所述, 用 29, 141 个转录产物开始, 使用 p- 值过滤和后向消除, 将标签减少至与 GPRS 相等大小的基因列表, 即, 302 个转录产物, 称为“验证生物标记物标签”。使用验证生物标记物标签, 在训练数据集上训练支持向量机 (SVM) (Noble WS (2006) What is a support vector machine? Nat Biotechnol 24:1565-1567)。然后将训练过的 SVM 用于将剩余 30% 数据 (即, 测试集) 中的每种样品归类为呼吸道致敏剂或非致敏剂。用接受者操作特征下面积 (ROC AUC) 来评价分类的性能。将这个完整的交叉验证重复 20 次, 每次产生不同的训练集和测试集, 每个训练集产生不同的验证生物标记物标签。这些交叉验证的结果显示于图 6B-C 中。发现中值 ROC AUC 为 0.84, 具有 0.66 至 0.96 的范围。此外, 针对 GPRS 中的每个基因的验证召集频率 (Validation Call Frequency) 列于表 1 中。VCF 描述了特定基因包括在 20 个验证生物标记物标签的任一个中的频率, 由此提供第二个测量, 通过其可以将预测剂分级。

[0219] 与 GARD 呼吸道预测标签相关的标准途径

[0220] 目的在于研究由呼吸道化学致敏剂在 MUTZ-3 细胞中启动的生物学应答, 使用

Ingenuity Pathway Analysis (IPA) 分析了数据。将使用 p - 值过滤选择的头 1029 个基因连同每个基因的倍数变化值用作 IPA 的输入信息。在 1029 个基因中, IPA 能够将 933 个作图于唯一的 ID。考虑了重复两次, 准备用于 IPA 分子的数据集由 901 个分子组成。主要的目的是说明哪个标准途径鉴定的分子与其相关。结果列于表 1 中, 按照根据 IPA 的统计学显著性的顺序。

[0221] 明显这些鉴定的和显著相关途径中的大部分主要是由有限的分子集驱动的。这些途径包括 TREM1 信号传导、类风湿性关节炎中改变的 T 细胞和 B 细胞信号传导、获得性和先天性免疫细胞之间的信息传递、B 细胞发育、芳基羟受体信号传导、树突细胞成熟、T- 辅助细胞中的 CD28 信号传导、通过 CD1 的脂质抗原呈递、细胞毒性 T 细胞介导的靶细胞凋亡和自身免疫甲状腺疾病信号传导。值得注意的是, 所有这些途径的中心都是先天性和获得性免疫之间的桥梁, 以及由外源物质识别启动的先天性免疫应答的衔接, 其导致树突细胞成熟。通过 GPRS 很好地监测的这一过程的关键方面包括先天性受体如 TLR 和 AHR 的上调、抗原呈递相关分子如 HLA 和 CD1 的上调、共同刺激分子如 CD86 和 CD40 的上调以及促炎效应分子如 IL-8 和 IL-1B 的上调。

[0222] 讨论

[0223] 各种化学物质不仅诱导皮肤的过敏性致敏, 而且还诱导呼吸道致敏, 这产生职业性哮喘和其他症状 (Kimber I, Dearman RJ(1997) Cell and molecular biology of chemical allergy. Clin Rev Allergy Immunol 15:145-168)。尽管没有与导致过敏性接触性皮炎的诱导皮肤致敏的化学物质那样普遍, 但呼吸道化学致敏剂的鉴定和危害评价同等重要, 尤其是鉴于严重的症状, 以及可能的致命结果 (Chester DA, Hanna EA, Pickelman BG, Rosenman KD(2005) Asthma death after spraying polyurethane truck bedliner. Am J Ind Med 48:78-84; Kimber I, Wilks MF(1995) Chemical respiratory allergy. Toxicological and occupational health issues. Hum Exp Toxicol 14:735-736)。

[0224] 最近, 发明人提出了用于皮肤致敏剂的基于细胞的体外测试方法, 称为 GARD, 其能够以高精确性将化学物质分类 (Johansson 等, 2011, 上文; Johansson H, Albrekt AS, Borrebaeck CAK, Lindstedt M(2012) The GARD assay for assessment of chemical skin sensitizers. Toxicol in Vitro)。该测定依赖于化合物刺激后 MUTZ-3 细胞的转录谱, 其使用预定的生物标记物标签作为读出信息。因为这些生物标记物的测量是基于表达阵列技术, 所以存在很大的机会来扩大这一测定的应用领域。在本发明研究中, 发明人提出了进一步的 GARD 研发, 使用称为 GARD 呼吸道预测标签 (GPRS) 的分离的生物标记物标签, 允许鉴定呼吸道化学致敏剂。使用已知是呼吸道致敏剂或非致敏剂的一组参照化学物质, 并且通过 ANOVA p - 值过滤以及接着用于后向消除的特征选择算法来鉴定这两组中差异表达的基因。所获得的 GPRS 的意在用途因此将在组合体外测定中, 其中用待归类的未知化合物刺激 MUTZ-3 细胞。使用两种截然不同的生物标记物标签, 可以将化合物归类为皮肤致敏剂、呼吸道致敏剂或非致敏剂。能够同时诱导呼吸道和皮肤致敏的化学物质也将同样明确地归类。

[0225] 通过两种形式的验证评估了所述测定在归类呼吸道化学变应原中的预测性能。首先, 由重复三份的 9 种阴性刺激组成的外部测试集成功归类, 如图 6A 中所示的。其次, 应用了全面的交叉验证方法, 其中随机反复排除 30% 的数据, 以形成测试集, 其之后用在剩余的 70% 数据上训练的 SVM 模型归类。这种交叉验证的结果呈现为 ROC AUC (图 6B 和 6C), 具有

0.66 至 0.96 范围中的 0.84 的中值。预测性能的大的变化意味着随机排除 30% 的数据很大程度上影响了验证生物标记物标签的组成。实际上,不同验证生物标记物标签之间的变化比由之前经验所预期的更大,而 VCF:s 比由之前经验所预期的更小 (Johansson 等,2011,上文)。已经研究了每种验证生物标记物标签的组成的影响,并且考虑了多种因素之间的相关性,如训练集中特定机械域的存在以及从每个训练集除去的每种刺激的重复数量。没有显示出可以解释预测性能变化的明显模式。然而,获得高达 0.96 的 ROC AUC 的能力显示出有力的证据:当在所有可利用的数据上训练模型时,精确的归类实际上是可能的。

[0226] 目前不存在经过验证的或甚至广泛接受的用于对诱导呼吸道致敏的化学物质进行危害评估的方法,这很大一部分原因是由于缺乏对引起化学呼吸道致敏发生的免疫生物学机理的了解 (. Isola D, Kimber I, Sarlo K, Lalko J, Sipes IG(2008) Chemical respiratory allergy and occupational asthma:what are the key areas of uncertainty? J Appl Toxicol 28:249-253)。具体地,尚未解决的最难懂的问题之一是 IgE 抗体在呼吸道对化学物质的过敏性致敏中的作用,以及是否存在与 IgE 抗体无关的可以获得这种致敏的机理 (Kimber I, Dearman RJ(2002) Chemical respiratory allergy:role of IgE antibody and relevance of route of exposure. Toxicology 181-182:311-315)。对于多种化学变应原,例如,对于酸酐,在 IgE 抗体水平和临床症状之间存在真正的关联。相反,少于一半的对二异氰酸酯敏感的患者证明了血清中特定的 IgE 抗体。此外,一致的观点是 IgE 抗体和化学呼吸道过敏之间的关联是强烈的 (Kimber I, Basketter DA, Gerberick GF, Ryan CA, Dearman RJ(2011) Chemical allergy:translating biology into hazard characterization. Toxicol Sci 120Suppl 1:S238-268)。最令人信服的论证是在设计成功检测对化学半抗原特异的 IgE 抗体的探针中存在技术困难。另外,相对于最后暴露时间的变应原特异性 IgE 的血液取样时间可能影响此类测定的结果。

[0227] 为了监控和比较呼吸道化学变应原的不同亚型的转录谱,图 1C 显示了基于 GPRS 基因的 PCA,其中根据机械域将化学物质着色。在该图中,在二异氰酸酯和酸酐之间没有可检测的明显差异,因为这两个组的簇紧挨在一起。尽管这没有解决在体内、IgE 相关或 IgE 不相关的致敏中可能是不同的机械途径问题,但其证实了这些组的化学物质在 MUTZ-3 中诱导相似的转录变化。替代地,最极端的转录变化由铵盐如六氯铂酸铵和过硫酸铵引发。然而,这两种化合物的转录谱的主要差异沿着第一种主要成分的轴是可检测的,即,以致敏剂与非致敏剂分离的相同的向量方向是可检测的。因此,发明人推断 GPRS 能够从各种机械子域精确地将变应原归类。

[0228] 为了进一步探究致敏化学物质对 MUTZ-3 的生物作用,进行了 IPA 分析。为了在数据中获得足够的显著性,将来自 p- 值过滤的头 1029 个基因用作 IPA 软件中的输入信息,而不是 GPRS 的头 302 个基因。呈现于表 1 中的 IPA 输出信息列出了标准信号传导途径,其中头 1029 个基因与其最显著相关。这些途径中的大部分主要是由核心的分子集来驱动的,包括 CD86、CD40、TLR1、TLR、各种 HLA-DR 分子和 CD1 分子。因此,呼吸道化学致敏剂诱导 MUTZ-3 中提高的抗原呈递和共同刺激分子的上调,可辩证地应答各种模式识别受体 (PRR) 和胞内氧化应激的连接,如通过芳基烃受体 (AHR) 信号传导和谷胱甘肽代谢的显著性所示的。

[0229] 总之,MUTZ-3 对化学呼吸道变应原的生物应答是由先天性免疫应答信号传导途径

控制的,其最终导致这种树突细胞模型的细胞成熟,且具有增强的抗原呈递以及与其他免疫细胞的相互作用作为最终结果。此外,之前已经与对蛋白质变应原的呼吸道致敏相关的信号传导途径的用途的新发现能够解释应答化学变应原而导致呼吸道致敏的生物学过程。因此,GPRS在免疫学机械观点中确实是相关的,并且提供了对监控导致呼吸道致敏的生物学事件的转录产物的测量。

[0230] 总之,发明人提出了补充之前所述的用于皮肤致敏剂评估的 GARD 测定的用于 MUTZ-3 细胞中的呼吸道化学致敏剂的预测生物标记物标签。在相同样品中测试两种不同终点的能力提供了有吸引力的且迄今为止唯一的用于安全地评估体外环境中的化学物质的测定。

[0231] 参考文献

[0232] 1. Boverhof DR, Billington R, Gollapudi BB, Hotchkiss JA, Krieger SM 等 (2008) Respiratory sensitization and allergy: current research approaches and needs. *Toxicol Appl Pharmacol* 226:1-13.

[0233] 2. Banks DE, Tarlo SM (2000) Important issues in occupational asthma. *Curr Opin Pulm Med* 6:37-42.

[0234] 3. Sastre J, Vandenplas O, Park HS (2003) Pathogenesis of occupational asthma. *Eur Respir J* 22:364-373.

[0235] 4. Zammit-Tabona M, Sherkin M, Kijek K, Chan H, Chan-Yeung M (1983) Asthma caused by diphenylmethane diisocyanate in foundry workers. Clinical, bronchial provocation, and immunologic studies. *Am Rev Respir Dis* 128:226-230.

[0236] 5. Bernstein DI, Patterson R, Zeiss CR (1982) Clinical and immunologic evaluation of trimellitic anhydride- and phthalic anhydride-exposed workers using a questionnaire with comparative analysis of enzyme-linked immunosorbent and radioimmunoassay studies. *J Allergy Clin Immunol* 69:311-318.

[0237] 6. Murdoch RD, Pepys J, Hughes EG (1986) IgE antibody responses to platinum group metals: a large scale refinery survey. *Br J Ind Med* 43:37-43.

[0238] 7. Docker A, Wattie JM, Topping MD, Luczynska CM, Newman Taylor AJ 等 (1987) Clinical and immunological investigations of respiratory disease in workers using reactive dyes. *Br J Ind Med* 44:534-541.

[0239] 8. Bourne MS, Flindt ML, Walker JM (1979) Asthma due to industrial use of chloramine. *Br Med J* 2:10-12.

[0240] 9. Verstraelen S, Bloemen K, Nelissen I, Witters H, Schoeters G 等 (2008) Cell types involved in allergic asthma and their use in in vitro models to assess respiratory sensitization. *Toxicol In Vitro* 22:1419-1431.

[0241] 10. Dearman RJ, Basketter DA, Kimber I (1992) Variable effects of chemical allergens on serum IgE concentration in mice. Preliminary evaluation of a novel approach to the identification of respiratory sensitizers. *J Appl Toxicol* 12:317-323.

[0242] 11. Dearman RJ, Skinner RA, Humphreys NE, Kimber I (2003) Methods for the

identification of chemical respiratory allergens in rodents:comparisons of cytokine profiling with induced changes in serum IgE. J Appl Toxicol 23:199-207.

[0243] 12.Verstraelen S,Nelissen I,Hooyberghs J,Witters H,Schoeters G等(2009) Gene profiles of THP-1macrophages after in vitro exposure to respiratory(non-)sensitizing chemicals:identification of discriminating genetic markers and pathway analysis. Toxicol In Vitro 23:1151-1162.

[0244] 13.Verstraelen S,Nelissen I,Hooyberghs J,Witters H,Schoeters G 等(2009)Gene profiles of a human bronchial epithelial cell line after in vitro exposure to respiratory(non-)sensitizing chemicals:identification of discriminating genetic markers and pathway analysis. Toxicology 255:151-159.

[0245] 14.Verstraelen S,Nelissen I,Hooyberghs J,Witters H,Schoeters G 等(2009)Gene profiles of a human alveolar epithelial cell line after in vitro exposure to respiratory(non-)sensitizing chemicals:identification of discriminating genetic markers and pathway analysis. Toxicol Lett 185:16-22.

[0246] 15.Lalko JF,Kimber I,Dearman RJ,Gerberick GF,Sarlo K等(2011)Chemical reactivity measurements:potential for characterization of respiratory chemical allergens. Toxicol In Vitro 25:433-445.

[0247] 16.Johansson H,Lindstedt M,Albrekt AS,Borrebaeck CA(2011)A genomic biomarker signature can predict skin sensitizers using a cell-based in vitro alternative to animal tests. BMC Genomics 12:399.

[0248] 17.Johansson H,Albrekt AS,Borrebaeck CAK,Lindstedt M(2012)The GARD assay for assessment of chemical skin sensitizers. Toxicol in Vitro.

[0249] 18.Santegoets SJ,Masterson AJ,van der Sluis PC,Lougheed SM,Fluitsma DM 等(2006)A CD34(+)human cell line model of myeloid dendritic cell differentiation:evidence for a CD14(+)CD11b(+)Langerhans cell precursor. J Leukoc Biol 80:1337-1344.

[0250] 19.Masterson AJ,Sombroek CC,De Gruijl TD,Graus YM,van der Vliet HJ 等(2002)MUTZ-3,a human cell line model for the cytokine-induced differentiation of dendritic cells from CD34+precursors. Blood 100:701-703.

[0251] 20.Larsson K,Lindstedt M,Borrebaeck CA(2006)Functional and transcriptional profiling of MUTZ-3,a myeloid cell line acting as a model for dendritic cells. Immunology 117:156-166.

[0252] 21.Carlsson A,Wingren C,Kristensson M,Rose C,Ferno M等(2011)Molecular serum portraits in patients with primary breast cancer predict the development of distant metastases. Proc Natl Acad Sci U S A 108:14252-14257.

[0253] 22.Kullback S,Leibler RA(1951)On Information and Sufficiency. Annals of Mathematical Statistics 22:79 - 86.

[0254] 23.Lasko TA,Bhagwat JG,Zou KH,Ohno-Machado L(2005)The use of receiver operating characteristic curves in biomedical informatics. J Biomed Inform

38:404-415.

[0255] 24. R Development Core Team(2008)R:A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

[0256] 25. Weingart SN, Iezzoni LI, Davis RB, Palmer RH, Cahalane M等 (2000) Use of administrative data to find substandard care: validation of the complications screening program. Med Care 38:796-806.

[0257] 26. Ringner M(2008)What is principal component analysis? Nat Biotechnol 26:303-304.

[0258] 27. Noble WS(2006)What is a support vector machine ? Nat Biotechnol 24:1565-1567.

[0259] 28. Benjamini Y, Hochberg Y(1995)Controlling the false discovery rate:a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B 57:289 - 300.

[0260] 29. Basketter DA, Kimber I(2011)Assessing the potency of respiratory allergens:uncertainties and challenges. Regul Toxicol Pharmacol 61:365-372.

[0261] 30. Enoch SJ, Roberts DW, Cronin MT(2010)Mechanistic Category Formation for the Prediction of Respiratory Sensitization. Chem Res Toxicol.

[0262] 31. Kimber I, Dearman RJ(1997)Cell and molecular biology of chemical allergy. Clin Rev Allergy Immunol 15:145-168.

[0263] 32. Chester DA, Hanna EA, Pickelman BG, Rosenman KD(2005)Asthma death after spraying polyurethane truck bedliner. Am J Ind Med 48:78-84.

[0264] 33. Kimber I, Wilks MF(1995)Chemical respiratory allergy. Toxicological and occupational health issues. Hum Exp Toxicol 14:735-736.

[0265] 34. Isola D, Kimber I, Sarlo K, Lalko J, Sipes IG(2008)Chemical respiratory allergy and occupational asthma:what are the key areas of uncertainty? J Appl Toxicol 28:249-253.

[0266] 35. Kimber I, Dearman RJ(2002)Chemical respiratory allergy:role of IgE antibody and relevance of route of exposure. Toxicology 181-182:311-315.

[0267] 36. Kimber I, Basketter DA, Gerberick GF, Ryan CA, Dearman RJ(2011)Chemical allergy:translating biology into hazard characterization. Toxicol Sci 120Suppl 1:S238-268.

[0268] 表 1. 来自 GARD 呼吸道预测标签的“核心”、“优选”和“任选”生物标记物

[0269]

基因符号	Entrez Gene ID	Affymetrix 探针组 ID	验证呼叫频率
(A) 核心生物标记物			
1. OR5B21	ENST00000360374	7948330	100
2. SLC7A7	ENST00000404278	7977786	95
(B) 优选生物标记物			
3. PIP3-E	ENST00000265198	8130408	85
4. BTNL8	ENST00000400706	8116537	85
5. CLEC4A	ENST00000360500	7953723	90
6. HIST4H4	ENST00000358064	7961483	80
7. YKT6	ENST00000223369	8132580	80
8. FLJ32679 ///	ENST00000327271	7981895	85
GOLGA8G ///			
GOLGA8E			
9. PACSIN3	ENST00000298838	7947801	90
10. PDE1B	ENST00000243052	7955943	80
11. NQO1	ENST00000320623	8002303	80
12. CAMK1D	ENST00000378845	7926223	95
13. MYB	ENST00000341911	8122202	95
14. ---	ENST00000387396	8065752	80
15. GRK5	ENST00000369106	7930894	90

[0270]

(C)任选的生物标记物

16.	CD86	ENST00000330540	8082035	100
17.	CD1A	ENST00000289429	7906339	85
18.	WWOX	ENST00000355860	7997352	85
19.	IKZF2	ENST00000374319	8058670	85
20.	FUCA1	ENST00000374479	7913694	80
21.	C10orf76	ENST00000370033	7935951	80
22.	AMICA1	ENST00000356289	7952022	80
23.	PDPK2 /// PDPK1	ENST00000382326	7998825	80
24.	AZU1	ENST00000334630	8024038	80
25.	ACN9	ENST00000360382	8134415	80
26.	PDPN	ENST00000400804	7898057	75
27.	LOC642587	NM_001104548	7909422	75
28.	SEC61A2	ENST00000379051	7926189	75
29.	ELA2	ENST00000263621	8024056	75
30.	BMP2K	ENST00000335016	8096004	75
31.	HCCS	ENST00000321143	8165995	75
32.	CXorf26	ENST00000373358	8168447	75
33.	TYSND1	ENST00000287078	7934114	70
34.	CARS	ENST00000380525	7945803	70
35.	NECAP1	ENST00000339754	7953715	70
36.	CDH26	ENST00000348616	8063761	70
37.	SERPINB1	ENST00000380739	8123598	70
38.	STEAP4	ENST00000301959	8140840	70
39.	TXNIP	ENST00000369317	7904726	65
40.	---	ENST00000386628	7925821	65
41.	C12orf35	ENST00000312561	7954711	65
42.	HMGA2	ENST00000393578	7956867	65
43.	KRT16	ENST00000301653	8015376	65
44.	GGTLC2	ENST00000215938	8071662	65
45.	---	ENST00000386437	8089926	65
46.	OSBPL11	ENST00000393455	8090277	65
47.	FAM71F1	ENST00000315184	8135945	65
48.	ATP6V1B2	ENST00000276390	8144931	65

[0271]

49.	LOC128102	AF252254	7904429	60
50.	TBX19	ENST00000367821	7907146	60
51.	NID1	ENST00000264187	7925320	60
52.	LPXN	ENST00000263845	7948332	60
53.	C15orf45	AK057017	7982375	60
54.	RNF111	ENST00000380504	7983953	60
55.	---	ENST00000386861	7993183	60
56.	CD33	ENST00000262262	8030804	60
57.	TANK	ENST00000259075	8045933	60
58.	ANKRD44	ENST00000282272	8057990	60
59.	WDFY1	ENST00000233055	8059361	60
60.	SDC4	ENST00000372733	8066513	60
61.	TMPRSS11B	ENST00000332644	8100701	60
62.	AFF4	ENST00000265343	8114083	60
63.	HBEGF	ENST00000230990	8114572	60
64.	XK	ENST00000378616	8166723	60
65.	SLAMF7	ENST00000368043	7906613	55
66.	S100A4	ENST00000368715	7920271	55
67.	MPZL3	ENST00000278949	7952036	55
68.	---	GENSCAN00000044853	7967586	55
69.	TRAV8-3	ENST00000390435	7973298	55
70.	LOC100131497	GENSCAN00000046821	7980481	55
71.	KIAA1468	ENST00000299783	8021496	55
72.	SPHK2	ENST00000245222	8030078	55
73.	---	ENST00000309260	8096554	55
74.	CCR6	ENST00000283506	8123364	55
75.	GSTA3	ENST00000370968	8127087	55
76.	RALA	ENST00000005257	8132406	55
77.	C7orf53	ENST00000312849	8135532	55
78.	---	AF480566	8141421	55
79.	CERCAM	ENST00000372842	8158250	55
80.	---	hsa-mir-147	8163729	55
81.	NFYC	ENST00000372655	7900468	50
82.	CD53	ENST00000271324	7903893	50

[0272]

83.	PSEN2	ENST00000366783	7910146	50
84.	CISD1	ENST00000333926	7927649	50
85.	SCD	ENST00000370355	7929816	50
86.	MED19	ENST00000337672	7948293	50
87.	SYT17	ENST00000396244	7993624	50
88.	KRT16 /// LOC400578 /// MGC102966	ENST00000399124	8013465	50
89.	C18orf51	ENST00000400291	8023864	50
90.	CD79A	ENST00000221972	8029136	50
91.	C19orf56	ENST00000222190	8034448	50
92.	AGFG1	ENST00000409979	8048847	50
93.	FOXP1	ENST00000318796	8088776	50
94.	TLR6	ENST00000381950	8099841	50
95.	SUSD3	ENST00000375472	8156393	50
96.	---	ENST00000387842	8176921	50
97.	---	ENST00000387842	8177424	50
98.	GPA33	ENST00000367868	7922029	45
99.	CDC123	ENST00000281141	7926207	45
100.	C10orf11	ENST00000354343	7928534	45
101.	---	ENST00000322493	7937971	45
102.	PTMAP7	AF170294	7976239	45
103.	ARRDC4	ENST00000268042	7986350	45
104.	---	ENST00000388199	7997738	45
105.	---	ENST00000388437	8009299	45
106.	KRT9	ENST00000246662	8015357	45
107.	---	ENST00000379371	8035868	45
108.	HDAC4	ENST00000345617	8060030	45
109.	CD200	ENST00000315711	8081657	45
110.	PAPSS1	ENST00000265174	8102214	45
111.	ORAI2	ENST00000356387	8135172	45
112.	---	AK124536	8144569	45
113.	ZBTB10	ENST00000379091	8147040	45
114.	---	ENST00000387422	8159963	45
115.	RAB9A	ENST00000243325	8166098	45

[0273]

116.	---	---	7895613	40
117.	DRD5	ENST00000304374	7905025	40
118.	CNR2	ENST00000374472	7913705	40
119.	OIT3	ENST00000334011	7928330	40
120.	---	ENST00000386981	7933008	40
121.	C10orf90	ENST00000356858	7936996	40
122.	OR52D1	ENST00000322641	7938008	40
123.	ZNF214	ENST00000278314	7946288	40
124.	---	ENST00000386959	7954690	40
125.	ART4	ENST00000228936	7961507	40
126.	RCBTB2	ENST00000344532	7971573	40
127.	HOMER2	ENST00000304231	7991034	40
128.	WWP2	ENST00000359154	7996976	40
129.	WDR24	ENST00000248142	7998280	40
130.	MED31	ENST00000225728	8011968	40
131.	CALM2	ENST00000272298	8052010	40
132.	DLX2	ENST00000234198	8056784	40
133.	BTBD3	ENST00000399006	8060988	40
134.	---	ENST00000339367	8075817	40
135.	TBCA	ENST00000380377	8112767	40
136.	GIN1	ENST00000399004	8113403	40
137.	NOL7	ENST00000259969	8116969	40
138.	---	ENST00000402365	8117628	40
139.	C7orf28B /// C7orf28A	ENST00000325974	8138128	40
140.	DPP7	ENST00000371579	8165438	40
141.	hCG_1749005	NR_003933	8167640	40
142.	PNPLA4	ENST00000381042	8171229	40
143.	USP51	ENST00000330856	8173174	40
144.	HLA-DQA1 ///	ENST00000383127	8178193	40
	HLA-DRA			
145.	FAAH	ENST00000243167	7901229	35
146.	GDAP2	ENST00000369443	7918955	35
147.	CD48	ENST00000368046	7921667	35
148.	PTPRJ	ENST00000278456	7939839	35

[0274]

149.	EXPH5	ENST00000265843	7951545	35
150.	RPS26 /// LOC728937 /// RPS26L /// hCG_2033311	ENST00000393490	7956114	35
151.	ALDH2	ENST00000261733	7958784	35
152.	CALM1	ENST00000356978	7976200	35
153.	NOX5 /// SPESP1	ENST00000395421	7984488	35
154.	RHBDL1	ENST00000352681	7992010	35
155.	CYLD	ENST00000311559	7995552	35
156.	OSBPL1A	ENST00000357041	8022572	35
157.	GYPC	ENST00000259254	8045009	35
158.	RQCD1	ENST00000295701	8048340	35
159.	RBM44	ENST00000316997	8049552	35
160.	---	ENST00000384680	8051862	35
161.	C3orf58	ENST00000315691	8083223	35
162.	MFSD1	ENST00000264266	8083656	35
163.	HACL1	ENST00000321169	8085608	35
164.	SATB1	ENST00000338745	8085716	35
165.	USP4	ENST00000351842	8087380	35
166.	---	ENST00000410125	8089928	35
167.	---	ENST00000384055	8097445	35
168.	IL7R	ENST00000303115	8104901	35
169.	---	ENST00000364497	8117018	35
170.	FAM135A	ENST00000370479	8120552	35
171.	CD164	ENST00000310786	8128716	35
172.	DYNLT1	ENST00000367088	8130499	35
173.	NRCAM	ENST00000379027	8142270	35
174.	ZNF596	ENST00000308811	8144230	35
175.	---	ENST00000332418	8170322	35
176.	TCEAL3 /// TCEAL6	ENST00000372774	8174134	35
177.	SNAPIN	ENST00000368685	7905598	30
178.	DENND2D	ENST00000369752	7918487	30
179.	SAMD8	ENST00000372690	7928516	30
180.	LHPP	ENST00000368842	7931204	30

[0275]

181.	SLC37A2	ENST00000298280	7944931	30
182.	FLI1 /// EWSR1	ENST00000344954	7945132	30
183.	OR9G4	ENST00000395180	7948157	30
184.	LOC338799	ENST00000391388	7967210	30
185.	HEXDC	ENST00000337014	8010787	30
186.	NOTUM	ENST00000409678	8019334	30
187.	MCOLN1	ENST00000394321	8025183	30
188.	PRKACA	ENST00000350356	8034762	30
189.	CRIM1	ENST00000280527	8041447	30
190.	CECR5	ENST00000336737	8074227	30
191.	RNF13	ENST00000392894	8083310	30
192.	40969	ENST00000339875	8103508	30
193.	ZNF366	ENST00000318442	8112584	30
194.	---	ENST00000410754	8120979	30
195.	GIMAP5	ENST00000358647	8137257	30
196.	---	ENST00000362484	8147242	30
197.	TFE3	ENST00000315869	8172520	30
198.	RHOU	ENST00000366691	7910387	25
199.	MED8	ENST00000290663	7915516	25
200.	CASQ2	ENST00000261448	7918878	25
201.	NUDT5	ENST00000378940	7932069	25
202.	C11orf73	ENST00000278483	7942932	25
203.	PAK1	ENST00000356341	7950578	25
204.	PRSS21	ENST00000005995	7992722	25
205.	---	ENST00000332418	7997907	25
206.	BTBD12	ENST00000294008	7999008	25
207.	DHRS13	ENST00000394901	8013804	25
208.	CCDC102B	ENST00000319445	8021685	25
209.	BCL2	ENST00000398117	8023646	25
210.	ZNF211 /// ZNF134	ENST00000396161	8031784	25
211.	NDUFV2	ENST00000340013	8039068	25
212.	MYCN	ENST00000281043	8040419	25
213.	---	ENST00000385528	8045561	25
214.	---	ENST00000362957	8046522	25

[0276]

215.	CASP8	ENST00000264275	8047419	25
216.	RTN4	ENST00000394611	8052204	25
217.	PLCG1	ENST00000244007	8062623	25
218.	MGC42105	ENST00000326035	8105146	25
219.	EMB	ENST00000303221	8112007	25
220.	---	ENST00000386433	8121249	25
221.	COL21A1	ENST00000370817	8127201	25
222.	LRP12	ENST00000276654	8152280	25
223.	LMNA	ENST00000368301	7906085	20
224.	---	ENST00000385567	7907535	20
225.	---	ENST00000362863	7926805	20
226.	ZNF503	ENST00000372524	7934553	20
227.	NLRX1	ENST00000397884	7944463	20
228.	---	ENST00000391173	7954775	20
229.	NDRG2	ENST00000298687	7977621	20
230.	TRAF7	ENST00000326181	7992529	20
231.	KRT40	ENST00000400879	8015152	20
232.	KRT40	ENST00000400879	8019604	20
233.	DRD5	ENST00000304374	8053725	20
234.	ZC3H8	ENST00000409573	8054664	20
235.	MMP9	ENST00000372330	8063115	20
236.	PLTP	ENST00000372420	8066619	20
237.	---	ENST00000362686	8100476	20
238.	SPEF2	ENST00000282469	8104856	20
239.	LRRC16A	ENST00000332168	8117243	20
240.	FBXO9	AK095315	8120269	20
241.	EEPD1	ENST00000242108	8132305	20
242.	FCN1	ENST00000371807	8165011	20
243.	EFNA3	ENST00000368408	7905918	15
244.	---	ENST00000314893	7910385	15
245.	TMEM19	ENST00000266673	7957167	15
246.	PLXNC1	ENST00000258526	7957570	15
247.	NHLRC3	ENST00000379599	7968703	15
248.	MBNL2	ENST00000397601	7969677	15

[0277]

249.	EIF5	ENST00000216554	7977058	15
250.	PLEKHG4	ENST00000360461	7996516	15
251.	COPS3	ENST00000268717	8013094	15
252.	FAM171A2	ENST00000398346	8016033	15
253.	LOC653653 /// AP1S2	ENST00000380291	8017210	15
254.	VAPA	ENST00000340541	8020129	15
255.	MATK	ENST00000395040	8032682	15
256.	ACTR2	ENST00000377982	8042337	15
257.	BPI	ENST00000262865	8062444	15
258.	ERG	ENST00000398905	8070297	15
259.	LAMB2	ENST00000305544	8087337	15
260.	---	BC090058	8133752	15
261.	PHTF2	ENST00000248550	8133818	15
262.	---	ENST00000333261	8133902	15
263.	C8orf55	ENST00000336138	8148559	15
264.	PDE7A	ENST00000379419	8151074	15
265.	NAPRT1	ENST00000340490	8153430	15
266.	HLA-DRA	ENST00000383127	8179481	15
267.	SLC22A15	ENST00000369503	7904226	10
268.	FCGR1A /// FCGR1B /// FCGR1C	ENST00000369384	7905047	10
269.	SLC27A3	ENST00000271857	7905664	10
270.	ID3	ENST00000374561	7913655	10
271.	TBCEL	ENST00000284259	7944623	10
272.	FAM138D	ENST00000355746	7960172	10
273.	POMP	ENST00000380842	7968297	10
274.	SNN	ENST00000329565	7993259	10
275.	MED13	ENST00000262436	8017312	10
276.	ZFP36L2	ENST00000282388	8051814	10
277.	UXS1	ENST00000409501	8054395	10
278.	CD40	ENST00000279061	8063156	10
279.	---	ENST00000362620	8066960	10
280.	GGT5	ENST00000327365	8074991	10
281.	---	BC035666	8103023	10

[0278]

282.	G6PD	ENST00000393562	8176133	10
283.	---	ENST00000384272	7902365	5
284.	CLCC1	ENST00000369971	7918255	5
285.	SCGB2A1	ENST00000244930	7940626	5
286.	GAA	ENST00000302262	8010354	5
287.	SERPINB2	ENST00000404622	8021635	5
288.	GPI	ENST00000356487	8027621	5
289.	LASS6	ENST00000392687	8046086	5
290.	EIF4A2	AB209021	8084704	5
291.	HLA-DRA	ENST00000383127	8118548	5
292.	---	ENST00000385586	8136889	5
293.	ANXA2P2	M62898 /// NR_003573	8154836	5
294.	FANCG	ENST00000378643	8160935	5
295.	FAM53B	ENST00000337318	7936884	0
296.	RFXAP	ENST00000255476	7968653	0
297.	UBR1	ENST00000382177	7987981	0
298.	TBC1D2B	ENST00000409931	7990657	0
299.	SERPINB10	ENST00000397996	8021645	0
300.	SEC23B	ENST00000377481	8061186	0
301.	MN1	ENST00000302326	8075126	0
302.	CRTAP	ENST00000320954	8078450	0

[0279] 通过 ANOVA 和后向消除鉴定的用于呼吸道化学致敏的潜在预测剂基因的列表。基因用 Entrez Gene ID 来注释,其中可以在 (www.ncbi.nlm.nih.gov/gene) 找到。提供了用于 Human ST 1.0 阵列的 Affymetrix 探针组 ID。验证呼叫频率(%)是在交叉验证过程中获得的 20 个验证生物标记物标签中每个基因的出现频率。

[0280] 表 2. 用于每种参照化学物质的浓度和媒介物。

[0281]

化合物	缩写	媒介物	最大溶解度 (μM)	Rv90 (μM)	GARD 输入 浓度(μM)
呼吸道致敏剂					
六氯铂酸铵	AH	水	35	-	35
过硫酸铵	AP	DMSO	-	-	500
戊二醛	GA	水	-	10	10

[0282]

六亚甲基二异氰酸酯	HDI	DMSO	100	-	100
马来酸酐	MA	DMSO	-	-	500
亚甲基联苯酚二异氰酸酯	MDI	DMSO	50	-	50
邻苯二甲酸酐	PA	DMSO	200	-	200
甲苯二异氰酸酯	TDI	DMSO	40	-	40
偏苯三酸酐	TMA	DMSO	150	-	150

非致敏剂

1-丁醇	BUT	DMSO	-	-	500
4-氨基苯甲酸	PABA	DMSO	-	-	500
氯苯	CB	DMSO	98	-	98
二甲基甲酰胺	DMF	水	-	-	500
乙基香草醛	EV	DMSO	-	-	500
异丙醇	IP	水	-	-	500
水杨酸甲酯	MS	DMSO	-	-	500
丙二醇	PG	水	-	-	500
高锰酸钾	PP	水	38	-	38
吐温 80	T80	DMSO	-	-	500
硫酸锌	ZS	水	126	-	126

[0283] 用于测定研发的每种参照化学物质的浓度和媒介物的列表。

[0284] 通过人的临床观察,将参照化学物质归类为呼吸道致敏剂或非致敏剂。

[0285] 表 3. 支持向量仪 (SVM) 算法

[0286] 1. 未知数据的 SVM 预测的 R- 脚本

[0287]

#The submitted script reads traindata.txt, testdata.txt and predictionsignature.txt, with example files provided.

```
source("NaiveBayesian")  
library(e1071)
```

#PART 1. USER INPUT

```
filnamnTraining<-"traindata.txt" #Provide the correct filename for traindata  
filnamnTest<-"testdata.txt" #Provide the correct filename for testdata
```

```
lista <- read.delim("predictionsignature.txt",header=FALSE) ##Provide the correct filename for  
the prediction signature  
lista <- as.character(lista[[1]])
```

[0288]

```

group1<- "pos" #Provide the correct label of sample class 1
group2<- "neg" #Provide the correct label of sample class 2

#PART 2. READ DATA

rawfile <- read.delim(filnamnTraining, header=FALSE)
rawfile <- t(rawfile)
samplenames <- as.character(rawfile[-1,1])
groupsTraining <- rawfile[-1,2]
dataTraining <- t(rawfile[-1,-c(1,2)])
dimdataTraining <- dim(dataTraining)
dataTraining <- as.numeric(dataTraining)
dim(dataTraining) <- dimdataTraining
ProteinNames <- as.character(rawfile[1,-c(1,2)])
rownames(dataTraining) <- ProteinNames
colnames(dataTraining) <- samplenames
logdataTraining <- dataTraining

listaBoolean <- is.element(ProteinNames, lista)
logdataTraining <- logdataTraining[listaBoolean,]

rawfile <- read.delim(filnamnTest, header=FALSE)
rawfile <- t(rawfile)
samplenames <- as.character(rawfile[-1,1])
groupsTest <- rawfile[-1,2]
dataTest <- t(rawfile[-1,-c(1,2)])
dimdataTest <- dim(dataTest)
dataTest <- as.numeric(dataTest)
dim(dataTest) <- dimdataTest
ProteinNames <- as.character(rawfile[1,-c(1,2)])
rownames(dataTest) <- ProteinNames
colnames(dataTest) <- samplenames
logdataTest <- dataTest

logdataTest <- logdataTest[listaBoolean,]

# PART 3. TRAIN THE SVM AND USE IT TO PREDICT SAMPLE CLASS OF TEST SET

svmfacTraining<- factor(rep('rest',ncol(logdataTraining)),levels=c(group1, group2, 'rest'))
subset1Training<- is.element(groupsTraining , strsplit(group1,"")[[1]])
subset2Training<- is.element(groupsTraining , strsplit(group2,"")[[1]])
svmfacTraining[subset1Training] <- group1
svmfacTraining[subset2Training] <- group2
facTraining <- factor(as.character(svmfacTraining [subset1Training|subset2Training]),levels=c(group1,group2))

svmfacTest<- factor(rep('rest',ncol(logdataTest)),levels=c(group1, group2, 'rest'))
subset1Test<- is.element(groupsTest , strsplit(group1,"")[[1]])
subset2Test<- is.element(groupsTest , strsplit(group2,"")[[1]])
svmfacTest[subset1Test] <- group1
svmfacTest[subset2Test] <- group2
facTest <- factor(as.character(svmfacTest [subset1Test|subset2Test]),levels=c(group1,group2))
n1 <- sum(facTest ==levels(facTest )[[1]])
n2 <- sum(facTest ==levels(facTest )[[2]])
nsamples <- n1+n2
SampleInformation <- paste(levels(facTest )[[1]], " ",n1," ", " ",levels(facTest )[[2]], " ",n2,sep="")

svmtrain <- svm(t(logdataTraining) , facTraining , kernel="linear" )
pred<-predict(svmtrain , t(logdataTest) , decision.values=TRUE)
res<-attr(pred, "decision.values")
names <- colnames(logdataTest, do.NULL=FALSE)

```

[0289]

```
orden <- order(res , decreasing=TRUE)
Samples <- data.frame(names[orden],res[orden],facTest[orden])
ROCdata <- myROC(res,facTest)
SenSpe <- SensitivitySpecificity(res,facTest)

# PART 4. IF SAMPLE CLASSES OF TEST DATA ARE KNOWN, PRINT ROC
ROCplot(list(SampleInformation=SampleInformation,ROCArea=ROCdata[1],p.value=ROCdata[2],SenSpe <- SenSpe,samples=Samples), sensspecnumber=4)

# PART 5. FOR UNKNOWN SAMPLES, PRINT DECISION VALUES
write.table(res, file="Predicted_resultsp1206.txt", sep= "\t", row.names = TRUE)
```

[0290] 2. 使用后向消除建立预测标签的 R- 脚本

[0291]

```
filnamn <- "inputdata.txt" #Provide correct filename for inputdata. Correct Format should be
t(traindata).
group1 <- "pos" #Provide label of sample class 1
group2 <- "neg" #Provide label of sample class 2

# Include
source("NaiveBayesian")
library(e1071)

# Hämta data
rawfile <- read.delim(filnamn)

# Läs in grupper
groups <- rawfile[,2]

# Hämta provnamn i datafilen
samplernames <- as.character(rawfile[,1])

# Skapa dataset ur råfilen
data <- t(rawfile[, -c(1,2)])

# Log
# data <- log(data)/log(2)

# antal prover
nsamples <- ncol(data)

# Skapa antikroppsnamnlista ur NYA datafilen
ProteinNames <- read.delim(filnamn,header=FALSE)
ProteinNames <- as.character(as.matrix(ProteinNames)[1,])
ProteinNames <- ProteinNames[-(1:2)]

# Kolla antal Ab i nya datasetet
antal <- length(ProteinNames)

# Ge rätt prov- och Ab-namn
rownames(data) <- ProteinNames
colnames(data) <- samplernames

# Skapa subsets
subset1 <- is.element(groups , strsplit(group1,"")[[1]])
subset2 <- is.element(groups , strsplit(group2,"")[[1]])

# Skapa factorlista
svmfac <- factor(rep('rest',ncol(data)),levels=c(group1,group2,'rest'))
svmfac[subset1] <- group1
svmfac[subset2] <- group2
```

[0292]

```

svmfac <- svmfac[subset1|subset2]

# Skapa vektor för K-L felen där det minsta för varje signaturlängd sparas
smallestErrorPerLength <- rep(NA,antal)

# Beräkna medelvärde för varje Ab över alla prov som är med
averages <- apply(data, 1, mean)

# Skapa vektor för Ab-ordningen efter K-L felen som erhållits när
# respektive antikropp var satt till medelvärde.
abOrder <- rep(NA,antal)

# Skapa ett dataset att eliminera i
elimData <- data[,subset1|subset2]

# Lista att förvara SVM-modellerna i
models <- numeric(nsamples)

# Skapa variabel för att hålla reda på hur många Ab som tagits bort
borttagna <- 0

#####
# BEGIN BACKELIM #####
#####

print(Sys.time())

# Kör tills bara två analyser återstår
for(j in 1:(antal-1))
{
  # Check if groups are given in correct order
  control <- as.numeric(svmfac)
  if(sum(control[subset1]) > sum(control[subset2]))
  {
    print("ERROR: Change order of your group1 and group2!!!")
    break
  }
  # För varje signaturlängd, där alla är med från början, träna en modell för
  # varje N-1 kombination av prover med den data som finns i elimData
  for (i in 1:nsamples)
  {
    # Modellerna sparas i en array av listor kallad models
    models[i] <- list(svm(t(t(elimData[-i])), svmfac[-i], kernel="linear"))
  }

  # Nu är alla modeller som behövs för LOO tränade och ska testas på elimData.
  # I elimData sätts först en analyt till medelvärde, sen testas var och en av
  # modellerna med det prov som var borttaget när den tränades.
  # När alla modellerna är testade en gång beräknas KL-fel som sparas i errors.
  # Nu sätts nästa analyt till medelvärde och testprocessen görs om, tills alla
  # analyser varit medelvärdeseleminerade en gång. Resultatet blir en KL-fel
  # lista lika lång som antalet analyser som är kvar i datasetet.

  # Skapa en lista med K-L fel en viss signaturlängd (antal + 1 - j lång)
  # där areorna för varje körning där en Ab i taget har satts till medelvärde
  errors <- testModels(models, elimData, averages)

  # Lägg namnet på Ab med sämst inverkan på felet i abOrder
  abOrder[j] <- getWorstAb(errors, row.names(elimData))
}

```

[0293]

```
# Lägger till värdet på det minsta felet
smallestErrorPerLength[j] <- getSmallestError(errors)

# Tar bort sämsta Ab ur medelvärdeslistan
averages <- getNewAverages(errors, averages)

# Tar bort sämsta Ab ur elimData
elimData <- getNewElimData(errors, elimData)

# Noterar att en Ab tagits bort
borttagna <- borttagna + 1

# Ange hur många analyter som eliminerats, samt vad klockan är.
print(paste(j, "analytes eliminated @", Sys.time()), sep="")
}

# Lägg till namnet på sista analyten, som aldrig blen eliminerad
abOrder[length(abOrder)] <- setdiff(ProteinNames, abOrder)

# Spara resultatet till fil
filename <- paste("Backward elimination result(", rnorm(1)+1, ").txt", sep="")
write.table(cbind(smallestErrorPerLength, abOrder), file=filename, sep="\t", quote =
F, row.names = F)
```

[0294] 3. 由脚本 1 和 2 要求的各种 R- 函数

[0295]

```
# getWorstAb: Rapporterar namnet på antikroppen som kommer tas bort
# (den där ROC-arean var som störst)
getWorstAb <- function(errors, abNames)
{
  return(abNames[order(errors, decreasing = F)[1]])
}

# testModels: testar alla modeller som finns i 'models' med alla
# analyser sätta till medelvärde en gång
testModels <- function(models, elimData, averages)
{
  nsamples <- ncol(elimData)
  d <- as.numeric(svmfac)-1
  y <- numeric(nsamples)
  E <- numeric(nsamples)
  analytes <- nrow(elimData)
  errors <- numeric(nrow(elimData))

  for(k in 1:analytes)
  {
    # Sätt analyt k till medelvärde i elimData
    # Men spara först analytens originalvärde
    backup <- elimData[k,]
    elimData[k,] <- averages[k]

    # Gör LOO loop för datasetet med de redan färdiga modellerna
    for (i in 1:nsamples)
    {
      pred <- predict(models[[i]], t(elimData[,i]), decision.values=TRUE)
      #spara decision values
      y[i] <- as.numeric(attributes(pred)$decision.values)
    }
  }
}
```

[0296]

```

# Beräkna "sannolikheterna"
y = 1-(1/(1 + exp(-y)))

# Beräkna KL-fel när aktuell analyt är eliminerad
for (i in 1:nsamples)
{
  E[i] <- -(d[i]*log(y[i])+(1-d[i])*log(1-y[i]))
}

# Spara felet
errors[k] <- sum(E)

# Lägg tillbaka analyten
elimData[k,] <- backup
}
return( errors )
}

# getNewElimData: Väljer vilken antikropp som ska tas bort ur träningsdatan och tar bort den
getNewElimData <- function(errors, elimData)
{
  # Positionen för det minsta felet
  tasBort <- order(errors,decreasing = F)[1]
  return(elimData[-tasBort,])
}

# getSmallestError: Rapporterar minsta K-L felet
getSmallestError <- function(errors)
{
  return(min(errors))
}

# getNewAverages: skapar en ny lista med medelvärden efter att en analyt
# eliminerats.
getNewAverages <- function(errors, averages)
{
  # Positionen för det minsta felet
  tasBort <- order(errors, decreasing = F)[1]
  return(averages[-tasBort])
}

# getRemovedAb: tar fram ID på analyt som eliminerats
getRemovedAb <- function(errors, abNames)
{
  return(abNames[order(errors, decreasing = T)[1]])
}

NBtrainer <- function(data, fac){
  MeanVariancePval <- function(vec , fac){
    vec1 <- vec[fac==levels(fac)[1]]
    vec2 <- vec[fac==levels(fac)[2]]
    if (sum(!is.na(vec1))<=2 | sum(!is.na(vec2))<=2){
      return(c(NA,NA,NA,NA,NA))
    }
    mean1 <- mean(vec1 , na.rm=TRUE)

```

[0297]

```

var1 <- var(vec1 , na.rm=TRUE)
mean2 <- mean(vec2 , na.rm=TRUE)
var2 <- var(vec2 , na.rm=TRUE)
if (var1==0 | var2==0){return(c(NA,NA,NA,NA,NA))}
pval <- t.test(vec1,vec2,var.equal=TRUE)$p.value
return(c(mean1,var1,mean2,var2,pval))
}
return(t(apply(data , 1 , MeanVariancePval , fac)))
}

NBpredicter <- function(testdata , NBtrained , topnumber=Inf , logfoldcut=0 , pcut =1){
  if (topnumber==Inf){
    indices <- !is.na(NBtrained[,5]) & NBtrained[,5]<=pcut & abs(NBtrained[,1]-
NBtrained[,3])>=logfoldcut
  }else{
    preindices <- !is.na(NBtrained[,5]) & NBtrained[,5]<=pcut
    abs(NBtrained[preindices,1]-NBtrained[preindices,3]) -> foldchange
    cutfold <- sort(foldchange , decreasing=TRUE )[min(topnumber,length(foldchange))]
    indices <- preindices & (abs(NBtrained[,1]-NBtrained[,3]) >= cutfold)
  }
  NBtrainedred <- matrix(NBtrained[indices,],ncol=ncol(NBtrained))
  testdatared <- matrix(testdata[indices,], ncol=ncol(testdata))
  singlegene <- function( genepred ){
    l1 <- -((genepred[6] - genepred[1])^2)/(2*genepred[2])-0.5*log(2*pi*genepred[2])
    l2 <- -((genepred[6] - genepred[3])^2)/(2*genepred[4])-0.5*log(2*pi*genepred[4])
    #print(genepred)
    return(l1-l2)
  }
  NBvectorpredicter <- function(vec){
    combined <- cbind(NBtrainedred , vec)
    combined <- matrix(combined[!is.na(vec),], ncol=6)
    return(sum(apply( combined , 1 , singlegene)))
  }
  return(apply(testdatared , 2 , NBvectorpredicter))
}

myROC <- function(numbers , fac){
  n1 <- sum(fac==levels(fac)[1])
  n2 <- sum(fac==levels(fac)[2])
  wilcoxresult <- wilcox.test(numbers~fac , alternative="greater")
  ROCarea <- as.numeric(wilcoxresult$statistic)/(n1*n2)
  pval <- wilcoxresult$p.value
  return(c(ROCarea,pval))
}

SensitivitySpecificity <- function(numbers , fac){
  n1 <- sum(fac==levels(fac)[1])
  n2 <- sum(fac==levels(fac)[2])
  un <- sort(unique(numbers), decreasing=TRUE)
  SenSpe <- function(x){
    sen <- sum(numbers>=x & fac==levels(fac)[1])/n1
    spe <- 1 - sum(numbers>=x & fac==levels(fac)[2])/n2
    return(list(Sensitivity=sen,Specificity=spe))
  }
  return(t(sapply(un , SenSpe)))
}

NBlooppreparer <- function(data , fac){
  nsamples <- ncol(data)
  ngenes <- nrow(data)
  NBtrainedarray <- array(NA , dim=c(ngenes,5,nsamples))

```

[0298]

```

for (i in 1:nsamples){ print(i)
  NBtrainedarray[,i] <- NBtrainer(matrix(data[,i],ncol=nsamples-1),fac[-i])
}
return(NBtrainedarray)
}

NBleaveoneout <- function(NBtrainedarray , data , fac , topnumber=Inf , logfoldcut=0 , pcut=1){
  nsamples <- ncol(data)
  loglikelihoods <- rep(NA , nsamples)
  for (i in 1:nsamples){
    loglikelihoods[i]<-
NBpredicter(matrix(data[,i],ncol=1),NBtrainedarray[,i],topnumber,logfoldcut,pcut)
  }
  return(loglikelihoods)
}

NBloocv <- function(NBtrainedarray , data , fac , topnumber=Inf , logfoldcut=0 , pcut=1){
  n1 <- sum(fac==levels(fac)[1])
  n2 <- sum(fac==levels(fac)[2])
  SampleInformation <- paste(levels(fac)[1]," ",n1," ",",",levels(fac)[2]," ",n2,sep="")
  loglikelihoods <- NBleaveoneout(NBtrainedarray , data,fac,topnumber,logfoldcut,pcut)
  names <- colnames(data , do.NULL=FALSE)
  orden <- order(loglikelihoods , decreasing=TRUE)
  Samples <- data.frame(names[orden],loglikelihoods[orden],fac[orden])
  ROCdata <- myROC(loglikelihoods,fac)
  SenSpe <- SensitivitySpecificity(loglikelihoods,fac)

  return(list(SampleInformation=SampleInformation,ROCArea=ROCdata[1],p.value=ROCdata[2]
],
          topnumber=topnumber,pcut=pcut,SenSpe <- SenSpe,samples=Samples))
}

NBtwooutpreparer <- function(data , fac){
  nsamples <- ncol(data)
  ngenes <- nrow(data)
  NBdoublearray <- array(NA , dim=c(ngenes,5,nsamples*(nsamples-1)/2))
  for (i in 2:nsamples){
    for (j in 1:(i-1)){ print(paste(i," ",j));
      NBdoublearray[,,(i-1)*(i-2)/2+j] <- NBtrainer(matrix(data[,c(i,j)],ncol=nsamples-2),fac[-c(i,j)])
    }
  }
  return(NBdoublearray)
}

NBmaximizer <- function(NBtrainedarray , data , fac){
  functomaximize <- function(pcut , topnumber){
    NBloocv(NBtrainedarray , data , fac , topnumber=top , pcut=pcut)$ROCArea
  }
  rocmx <- 0
  pcutmax <- numeric(0)
  topmax <- numeric(0)
  pcutset <- c(1,0.05,0.01,0.005,0.001 , 0.0003 , 0.0005,0.0001)
  topset <- c(1,2,5,10,20,50,100)
  for (pcut in pcutset){
    for (top in topset){
      currentroc <- functomaximize(pcut,top); # print(paste(pcut," ",top," ",currentroc))
      if (currentroc >= rocmx){
        rocmx <- currentroc
        pcutmax <- pcut
        topmax <- top
      }
    }
  }
}

```

[0299]

```

    }
  }
  print(paste("Result ",pcutmax," ",topmax," ",rocmax))
  return(c(pcutmax,topmax))
}

NBtotalvalidation <- function(NBdoublearray, NBtrainedarray, data, fac){
  n1 <- sum(fac==levels(fac)[1])
  n2 <- sum(fac==levels(fac)[2])
  nsamples <- n1+n2
  ngenes <- nrow(data)
  SampleInformation <- paste(levels(fac)[1]," ",n1," ",levels(fac)[2]," ",n2,sep="")
  maxarray <- matrix(NA, nrow=nsamples, ncol=2)
  colnames(maxarray) <- c('pcut','topnumber')
  NormScore <- numeric(nsamples)
  loglikelihoods <- numeric(nsamples)
  for (i in 1:nsamples){
    NBtemptrainedarray <- array(NA, dim=c(ngenes,5,nsamples-1))
    if (i>1){
      for (j in 1:(i-1)){
        NBtemptrainedarray[,,j] <- NBdoublearray[,,(i-1)*(i-2)/2+j]
      }
    }
    if (i < nsamples){
      for (j in (i+1):nsamples){
        NBtemptrainedarray[,,j-1] <- NBdoublearray[,,(j-1)*(j-2)/2+i]
      }
    }
    maxarray[i,] <- NBmaximizer( NBtemptrainedarray, data[,~i], fac[-i])
    temploglikelihoods <- NBpredicter( data, NBtrainedarray[,i], pcut = maxarray[i,1],
    topnumber=maxarray[i,2])
    loglikelihoods[i] <- temploglikelihoods[i]
    meanll <- mean(temploglikelihoods[-i])
    sdll <- sd(temploglikelihoods[-i])
    NormScore[i] <- (temploglikelihoods[i] - meanll)/sdll
  }
  names <- colnames(data, do.NULL=FALSE)
  orden <- order( NormScore, decreasing=TRUE)
  Samples <- data.frame(names[orden], NormScore[orden], loglikelihoods[orden], fac[orden],
  maxarray[orden,])
  ROCdata <- myROC(NormScore, fac)
  SenSpe <- SensitivitySpecificity(NormScore, fac)

  return(list(SampleInformation=SampleInformation, ROCarea=ROCdata[1], p.value=ROCdata[2],
  SenSpe <- SenSpe, samples=Samples))
}

ROCplot <- function(clasRes, sensspecnumber=6){
  Sensitivity <- as.numeric(clasRes[[sensspecnumber]][,1])
  Specificity <- as.numeric(clasRes[[sensspecnumber]][,2])
  OneMinusSpecificity <- 1- Specificity
  ROCarea <- round(clasRes$ROC,digits=2)
  plot(OneMinusSpecificity, Sensitivity, type="l", xlab="1-specificity", ylab="sensitivity")
  title(paste("ROC area = ",ROCarea), font.main=1)
}

ROCplotReverse <- function(clasRes){
  Sensitivity <- rev(as.numeric(clasRes[[4]][,2]))
  Specificity <- rev(as.numeric(clasRes[[4]][,1]))
  OneMinusSpecificity <- 1- Specificity

```

[0300]

```

ROCarea <- round(clasRes$ROC,digits=2)
plot(OneMinusSpecificity , Sensitivity , type="l" , xlab="1-specificity" , ylab="sensitivity")
title(paste("ROC area = ",ROCarea),font.main=1)
}

```

[0301] 4. 训练数据、测试数据和预测标签的实例文件

[0302] 4.1 训练数据。表应当保存为制表符分隔的 .txt- 文件

[0303]

样品	样品 1	样品 2	样品 3	样品 4	样品 5	样品 6	样品 7	样品 8	样品 9	样品 10
样品类别	阳性	阳性	阳性	阳性	阳性	阴性	阴性	阴性	阴性	阴性
预测剂 1	10	7	4	10	4	4	6	1	9	1
预测剂 2	5	9	2	6	2	9	3	5	4	10
预测剂 3	8	3	9	1	9	2	5	1	6	3
预测剂 4	4	8	7	7	5	6	8	2	2	3
预测剂 5	9	2	2	6	3	4	7	8	9	8
预测剂 6	5	4	7	10	4	2	1	9	1	10
预测剂 7	6	4	5	5	10	1	5	7	10	1
预测剂 8	5	4	1	10	1	6	2	6	8	4
预测剂 9	7	1	3	10	3	1	2	10	2	2
预测剂 10	10	8	2	8	2	6	3	4	6	10

[0304] 4.2 测试数据。表应当保存为制表符分隔的 .txt- 文件

[0305]

样品	样品 11	样品 12	样品 13	样品 14	样品 15	样品 16	样品 17	样品 18	样品 19	样品 20
样品类别	阳性	阳性	阳性	阳性	阳性	阴性	阴性	阴性	阴性	阴性
预测剂 1	8	3	10	7	6	8	4	6	3	3
预测剂 2	6	4	8	9	5	9	7	5	3	9
预测剂 3	4	10	1	8	9	2	2	6	6	2
预测剂 4	5	9	10	10	8	4	4	9	4	1

预测剂 5	1	10	1	6	1	10	1	3	8	5
预测剂 6	6	4	1	3	2	5	9	9	10	10
预测剂 7	5	1	3	3	3	5	6	3	3	6
预测剂 8	8	2	2	7	2	2	10	4	10	10
预测剂 9	4	10	8	6	5	10	9	9	4	1
预测剂 10	3	4	8	3	2	3	8	10	7	1

[0306] 4.3 预测标签。表应当保存为制表符分隔的 .txt- 文件

[0307]

预测剂 3
预测剂 5
预测剂 9

[0308] 表 4. 与 GPRS 相关的标准途径

[0309]

标准途径	$-\log(p\text{-值})$	受调控的分子 ¹
TREM1 信号传导	5.4	CASP1, CCL2, CCL3, CD40 , CD86 , FCGR2B, IL8, IL1B, MPO, PLCG1, SIGIRR, TLR1, TLR6
类风湿性关节炎中改变的 T 细胞和 B 细胞信号传导	3.7	CD40 , CD86 , CD79A, FAS, FCER1G, HLA-DQA1, HLA-DRA , IL1B, IL1RN, PRTN3, SPPI, TLR1, TLR6
烟酸和烟酰胺代谢	3.6	CD38, CDK6, DFFB, ENPP2, GRK5 , MAP2K1, MAPK6, NADK, NAPRT1, NNT, PAK1 , PPM1F, PTPRJ , PTPRO, SGK1
获得性和先天性免疫细胞之	2.9	CCL3, CD40 , CD86 , FCER1G, HLA-DRA ,

[0310]

间的信号传递		IFNA5, IL8, IL1B, IL1RN, TLR1, TLR6
B 细胞发育	2.9	CD40 , CD86 , CD79A, HLA-DQA1, HLA-DRA, IL7R
神经鞘脂类代谢	2.6	ASAH2, CERK, CERS6, FUT4, KDSR, NAAA, PPM1F, PTPRJ , PTPRO, SPHK2, SPTLC2
染色体复制的细胞周期控制	2.6	CDK6, CDT1, MCM2, MCM4, MCM6, MCM7
核黄素代谢	2.6	ACPP, ENPP2, PPM1F, PTPRJ , PTPRO
谷胱甘肽代谢	2.5	G6PD , GGT5, GGTLC2 , GLRX, GSTA3, H6PD, IDH2, MGST1,
芳基烃受体信号传导	2.4	AHR, ALDH1A1, CDK6, CDKN1A, CYP1B1, FAS, GSTA3, IL1B, JUN, MCM7, MGST1, NCOA3, NQO1 , NQO2, RB1
移植物对宿主防御信号传导	2.3	CD86 , FAS, FCER1G, HLA-DQA1, HLA-DRA, IL1B, IL1RN
树突细胞成熟	2.3	CD40 , CD86 , CD1A , CD1B, CD1C, CREB3L4, FCER1G, FCGR2A, FCGR2B, HLA-DQA1, HLA-DRA, IFNA5, IL1B, IL1RN, MAPK12, PIK3CD, PLCG1
T 辅助细胞中的 CD28 信号传导	2.3	ACTR2 , CALM1, CD86 , FCER1G, HLA-DQA1, HLA-DRA, JUN, MAP2K1, MAPK12, PAK1 , PDPK1, PIK3CD, PLCG1
CD1 的脂质抗原呈递	2.3	CD1A , CD1B, CD1C, FCER1G
细胞毒性T细胞介导的靶细胞凋亡	2.2	BCL2 , CASP8 , DFFB, FAS, FCER1G, HLA-DQA1, HLA-DRA
脂肪酸生物合成	2.1	ACACA, FASN, SLC27A3
自体免疫甲状腺疾病信号传导	2.0	CD40 , CD86 , FAS, FCER1G, HLA-DQA1, HLA-DRA

[0311] ¹⁾ 粗体表示的分子存在于 GARD 呼吸道预测标签中。红色分子为上调。

[0312] 表 4 说明。与头 1029 个预测剂相关的上端标准途径能够分离呼吸道化学致敏剂与非致敏剂。粗体表示的分子存在于 GPRS 中。红色分子在化学呼吸道致敏剂中是上调的，而绿色分子在化学呼吸道致敏剂中是下调的。

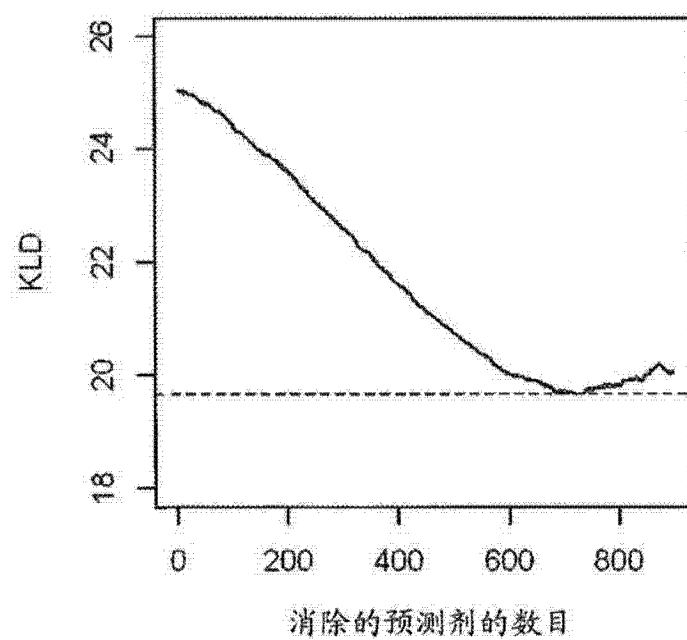


图 1

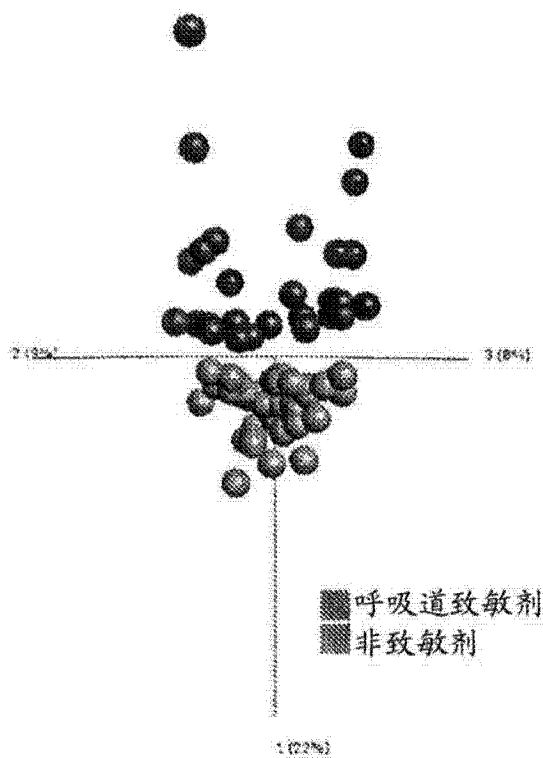


图 2

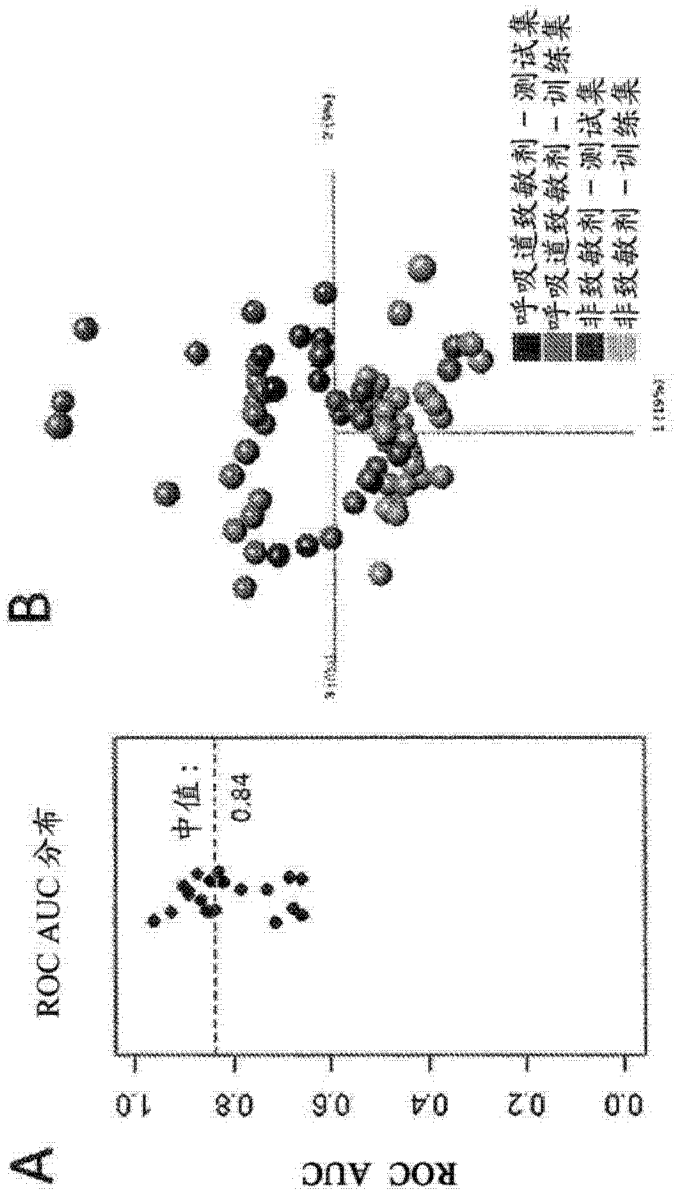


图 3

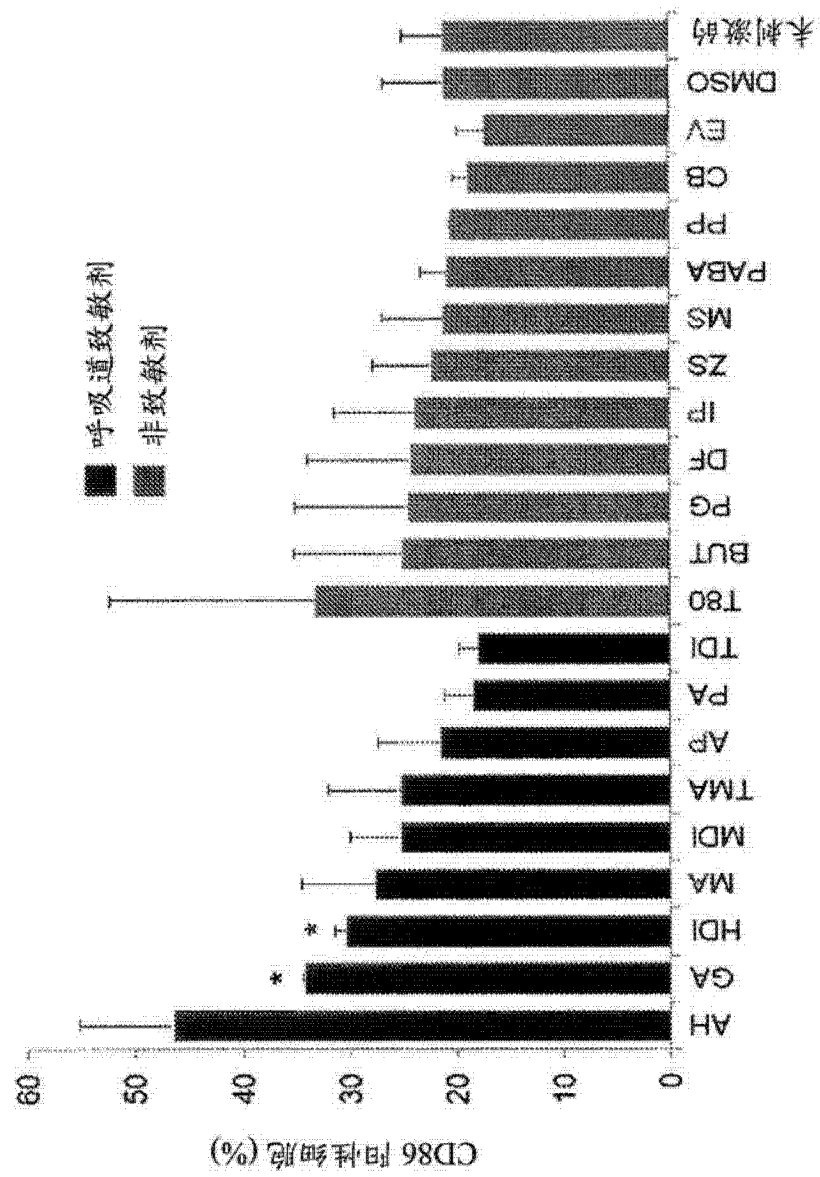


图 4

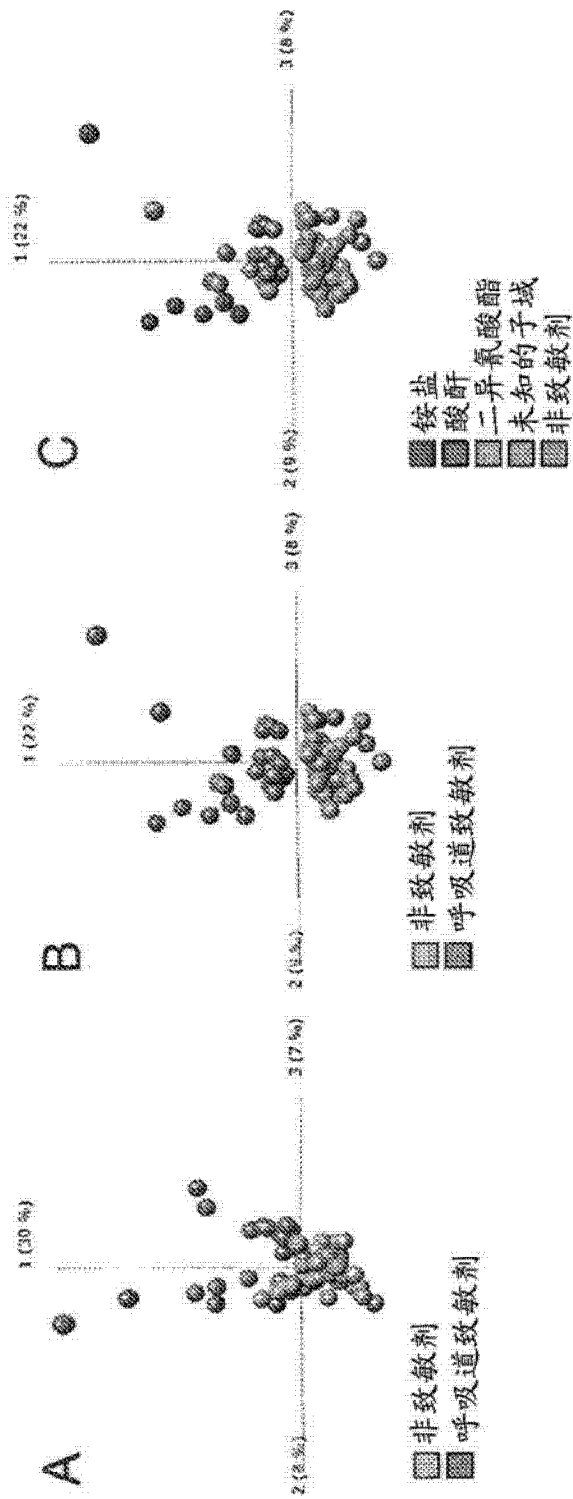


图 5

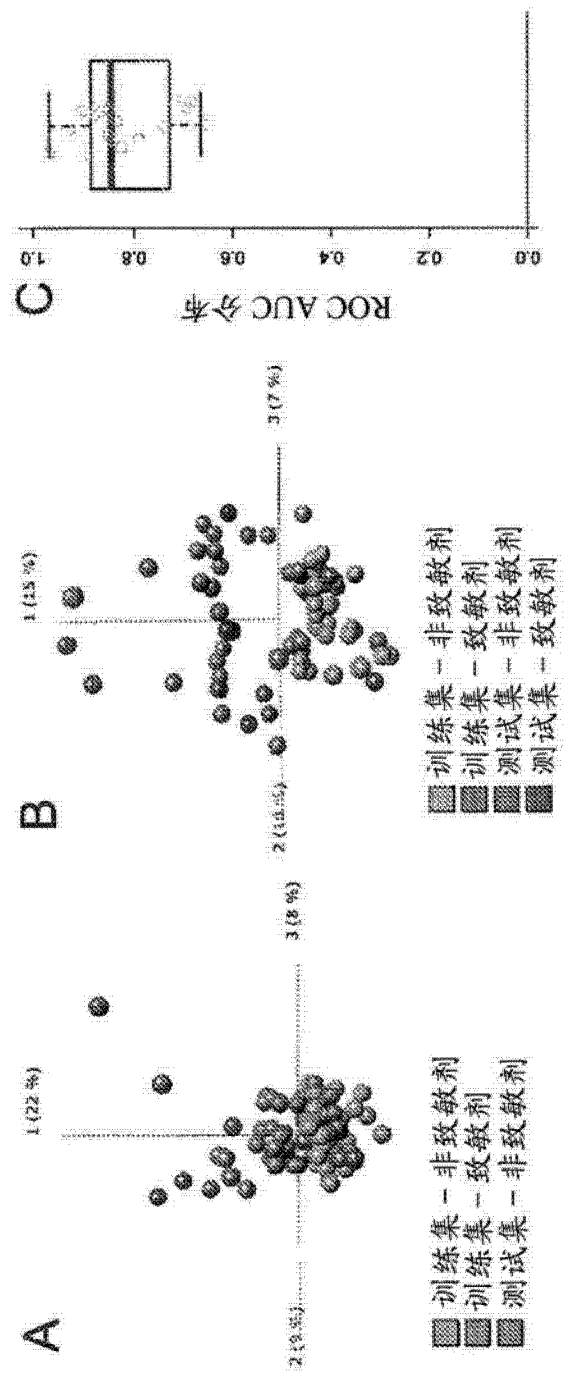


图 6

专利名称(译)	用于鉴定能够诱导呼吸道致敏的物质的方法和用于该方法中的阵列和分析试剂盒		
公开(公告)号	CN104685067A	公开(公告)日	2015-06-03
申请号	CN201380031747.7	申请日	2013-04-26
[标]申请(专利权)人(译)	森扎基因有限责任公司		
申请(专利权)人(译)	森扎基因有限责任公司		
当前申请(专利权)人(译)	森扎基因有限责任公司		
[标]发明人	M 林德斯泰特 C 伯莱拜克 H 乔汉森 A S 阿尔布雷克特		
发明人	M.林德斯泰特 C.伯莱拜克 H.乔汉森 A-S.阿尔布雷克特		
IPC分类号	C12Q1/68 G01N33/50 G01N33/53 G01N33/68		
CPC分类号	C12Q1/6883 C12Q2600/106 C12Q2600/158 G01N33/5023 G01N33/5047 G01N33/6893 G01N2800/24 G01N2800/52 C12Q1/6888 G01N33/68 G01N2800/122 G01N2800/14		
优先权	2012007297 2012-04-26 GB		
外部链接	Espacenet SIPO		

摘要(译)

本发明涉及用于鉴定能够诱导哺乳动物呼吸道致敏的物质的体外方法以及用于所述方法中的阵列和诊断试剂盒。特别地，该方法包括测量表1A、表1B和/或表1C中所列的生物标记物在暴露于测试物质的MUTZ-3细胞中的表达。

