

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

C12Q 1/68 (2006.01)

C12P 19/34 (2006.01)

G01N 33/48 (2006.01)



[12] 发明专利申请公布说明书

[21] 申请号 200780038103.5

[43] 公开日 2009 年 9 月 23 日

[11] 公开号 CN 101541976A

[22] 申请日 2007.8.10

[21] 申请号 200780038103.5

[30] 优先权

[32] 2006.8.11 [33] US [31] 60/837,148

[86] 国际申请 PCT/US2007/075713 2007.8.10

[87] 国际公布 WO2008/024642 英 2008.2.28

[85] 进入国家阶段日期 2009.4.13

[71] 申请人 贝勒研究院

地址 美国德克萨斯

共同申请人 德克萨斯州立大学董事会

[72] 发明人 J·F·班彻罗 A·K·帕吕卡

O·拉米罗 D·乔萨贝尔

[74] 专利代理机构 北京泛华伟业知识产权代理有限公司

代理人 刘丹妮 郭广迅

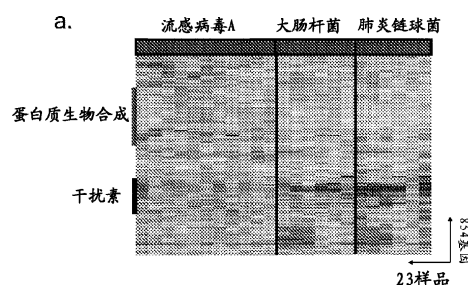
权利要求书 4 页 说明书 61 页 附图 11 页

[54] 发明名称

利用血液白细胞中的基因表达标记物的急性感染的鉴别诊断

[57] 摘要

本发明包括使用基因表达数据对宿主免疫应答的程度、类型和性质以及传染病的性质进行早期检测 and 一致测定的组合物、系统和方法。



1. 一种鉴定疑似患有传染病的人受试者的方法,该方法包括测定生物标记物的表达水平,所述的生物标记物包括以下的一种或多种基因: *cig5*、*DNAPTP6*、*IFI27*、*IFI35*、*IFI44*、*OAS1*、*BST2*、*G1P2*、*LY6E*、*MX1*、*SON*、*TRIM14*、*APOBEC3C*、*C1orf29*、*FLJ20035*、*FLJ38348*、*HSXIAPAF1*、*KIAA0152*、*PHACTR2* 和 *USP18*。

2. 权利要求 1 所述的方法,其中,测定表达水平的步骤是通过测量 mRNA、蛋白质以及它们的组合的量来进行的。

3. 权利要求 1 所述的方法,其中,测定表达水平的步骤是使用在固体支持物上的核酸杂交、寡核苷酸阵列、测序以及它们的组合来进行的。

4. 权利要求 1 所述的方法,其中,测定表达水平的步骤是使用 cDNA 来进行的,所述 cDNA 是使用从人细胞中收集的 mRNA 作为模板来制备的。

5. 权利要求 1 所述的方法,其中,所述生物标记物包括 mRNA 水平,并且所述生物标记物是通过选自以下的方法来进行定量的: 聚合酶链式反应、实时聚合酶链式反应、逆转录酶聚合酶链式反应、杂交、探针杂交、和基因表达阵列。

6. 权利要求 1 所述的方法,其中,测定表达水平的步骤是通过使用至少一种选自以下的技术来实现的: 聚合酶链式反应、异源双链分析、单链构象多态性分析、连接酶链式反应、比较基因组杂交技术、DNA 印迹法、RNA 印迹法、蛋白质印迹法、酶联免疫吸附测定、荧光共振能量转移和测序。

7. 权利要求 1 所述的方法,其中,所述样本包括外周血单核细胞。

8. 一种鉴定疑似患有传染病的人受试者的方法,该方法包括测定生物标记物的表达水平,所述的生物标记物包括以下的一种或多种基因: *EEF1G*、*EIF3S5*、*EIF3S7*、*EIF4B*、*QARS*、*RPL31*、*RPL4*、*PFDN5*、*CD44*、*HADHA*、*PCBP2* 和 *dJ507I15.1*。

9. 权利要求 1 所述的方法,其中,测定表达水平的步骤是通过测量 mRNA、蛋白质以及它们的组合的量来进行的。

10. 权利要求 1 所述的方法，其中，测定表达水平的步骤是使用在固体支持物上的核酸杂交、寡核苷酸阵列、测序以及它们的组合来进行的。

11. 权利要求 1 所述的方法，其中，测定表达水平的步骤是使用 cDNA 来进行的，所述 cDNA 使用从人细胞中收集的 mRNA 作为模板来制备的。

12. 权利要求 1 所述的方法，其中，所述生物标记物包括 mRNA 水平，并且所述的生物标记物通过选自以下的方法来定量的：聚合酶链式反应、实时聚合酶链式反应、逆转录酶聚合酶链式反应、杂交、探针杂交和基因表达阵列。

13. 权利要求 1 所述的方法，其中，测定表达水平的步骤是通过使用至少一种选自以下的技术来实现的：聚合酶链式反应、异源双链分析、单链构象多态性分析、连接酶链式反应、比较基因组杂交技术、DNA 印迹法、RNA 印迹法、蛋白质印迹法、酶联免疫吸附测定、荧光共振能量转移和测序。

14. 权利要求 1 所述的方法，其中，所述样本包括外周血单核细胞。

15. 一种鉴定疑似患有传染病的人受试者的方法，该方法包括通过测定包括一种或多种以下基因的生物标记物的表达水平以区分金黄色葡萄球菌感染和大肠杆菌感染：CXCL1、JAG1、RGS2、GAPD、PIIB、PSMA7、MMP9、p44S10、TRAM2、SEC24C、ACTG1、CGI-96、MGC2963 和 STAU。

16. 权利要求 15 所述的方法，其中，测定表达水平的步骤被用于检测与金黄色葡萄球菌感染相比较的大肠杆菌感染，所述检测是通过测定包括一种或多种以下基因的生物标记物的表达水平：RASA1、SNX4、AF1Q、SMAD2、JUP、PP、MAN1C1、FLJ10287、FLJ20152、LRRN3、SGPP1 和 UBAP2L。

17. 权利要求 1 所述的方法，其中，测定表达水平的步骤是通过测量 mRNA、蛋白质以及它们的组合的量进行的。

18. 权利要求 1 所述的方法，其中，测定表达水平的步骤是通过在固体支持物上进行核酸的杂交、寡核苷酸阵列、测序以及它们的组合进行的。

19. 权利要求 1 所述的方法，其中，测定表达水平的步骤是使用 cDNA

来进行的,所述cDNA是使用从人细胞中收集的 mRNA 作为模板来制备的。

20. 权利要求 1 所述的方法,其中,所述生物标记物包括 mRNA 水平,并且所述的生物标记物是通过选自如下的方法来定量的:聚合酶链式反应、实时聚合酶链式反应、逆转录酶聚合酶链式反应、杂交、探针杂交和基因表达阵列。

21. 权利要求 1 所述的方法,其中,所述测定表达水平的步骤是通过使用至少一种选自如下的技术来实现的:聚合酶链式反应、异源双链分析、单链构象多态性分析、连接酶链式反应、比较基因组杂交技术、DNA 印迹法、RNA 印迹法、蛋白质印迹法、酶联免疫吸附测定、荧光共振能量转移和测序。

22. 权利要求 1 所述的方法,其中,所述样本包括外周血单核细胞。

23. 一种鉴定疑似患有传染病的人受试者的方法,该方法包括测定生物标记物的表达水平以区分细菌感染和病毒感染,所述的生物标记物具有一种或多种以下的基因:EEF1G、EIF3S5、EIF3S7、EIF4B、QARS、RPL31、RPL4、PFDN5、CD44、HADHA、PCBP2 和 dJ507I15.1。

24. 一种用于测定样本表型的计算机实现的方法,该方法包括:

从样本中获得一种或多种探针强度;

根据所述探针强度诊断传染病;

计算所述探针强度与参比探针强度之间的线性相关系数;和

当所述线性相关系数大于阈值时,认为假定的表型为所述样本的表型。

25. 一种计算机可读介质,该介质包括用于进行样本转录组的测定方法的计算机可执行指令,所述方法包括:

获得多个样本的探针强度;

根据针对六种或更多种的基因的样本探针强度来诊断传染病,所述的六种或更多种基因选自表 2、表 3、补充表 1-11 以及它们的组合中所列出的基因;和

计算所述样本探针强度与参比探针强度之间的线性相关系数;并且当所述线性相关系数大于阈值时,认为假定的表型为所述样本的表型。

26. 权利要求 25 所述的系统，其中，所述生物标记物选自 5、6、7、8、9、10、11、12 种或更多种基因。

27. 权利要求 25 所述的系统，其中，所述生物标记物选自补充表 1-11 以及它们的组合中所列的一种或多种基因。

28. 一种基于计算机的用于创建数据集的方法，所述数据集与在个体中传染病的存在相关联，该方法包括以下的计算机执行步骤：

从所述个体获得多个基因的探针强度；

测定六种或更多种选自表 2、表 3、补充表 1-11 以及它们的组合中所列的基因的探针强度；和

针对所述六种或更多种基因中的每一种基因，计算所述样本探针强度与参比探针强度之间的线性相关系数，其中，所述相关性是通过对六种或更多种基因探针强度进行平均，从而计算与所述传染病的存在或缺失相关联的转录组表达矢量。

利用血液白细胞中的基因表达标记物的急性感染的鉴别诊断

技术领域

本发明通常涉及传染病的诊断学领域，更尤其是，本发明涉及用于对急性和慢性传染病进行诊断、预后和跟踪的系统、方法和设备。

长表格

本专利申请包括 11 个补充表格。

发明背景

本发明描述的背景技术与用于传染病的检测、评价、跟踪和预后的诊断方法相关，而不是为了限制本发明的范围。

在世界范围内[1]，尤其是在儿童中，急性传染病是导致死亡的主要原因。与此同时，鉴定传染因子的能力仍然不够充分，尤其是，鉴定在血液中（或者其它可利用的组织中）是否存在生物的能力不够充分。即使由于感染导致了白细胞的增多，但是这不能区分是革兰氏阳性细菌和革兰氏阴性细菌和/或病毒导致的感染。这些诊断上的障碍可能延迟开始适当的治疗，这将导致不必要病态，甚至死亡[2]。此外，最近，由于新出现的病原体导致的疾病发作[1、3] 和生物威胁导致的增加风险促进了传染病的改进的诊断的需要。

不同类别的病原体引发在白细胞上出现差异表达的特定的模式-识别受体 (PRRs) [4、5]。白细胞是天然免疫系统（粒细胞、天然杀伤细胞）、适应性免疫系统(T 和 B 淋巴细胞)、或者它们二者(单核细胞和树枝状细胞)的成分。血液既是这些细胞的储存库也是这些细胞的迁移室，因此，这些细胞可能暴露于传染因子、过敏原、肿瘤、移植反应或者自身免疫反应中。因此，血液中的白细胞构成了临床相关信息的可利用的来源，并且这些细胞的全面的分子表型可以通过基因表达微阵列获得。基因表达技术给癌症的诊断和预后带来了新的前景[6-8]，并且血白细胞中的基因表达信号的分析有助于更好地理解疾病发作以及对治疗的应答的机制[9-11]。

发明概述

本发明包括使用许多可变的基因表达分析方法用于预后以及诊断传染病的分析样本的系统和方法。这些保留的基因表达差异可以被赋予对不匹配的变化具有高度置信值的属性。因此，被鉴定的基因表达差异可以用于，例如，诊断宿主对传染病的应答、鉴定生理状态、鉴定、跟踪和监控免疫细胞的活化、设计药物、和监控治疗。

在一个实施方案中，本发明包括通过测定生物标记物的表达水平鉴定易于受传染因子（如，病毒、细菌、蠕虫、寄生虫、真菌等）的影响的人受试者的免疫应答。

生物标记物的其它的实例包括与传染因子相关的基因或者与由传染因子导致的疾病相关的基因和它们的组合。可以通过定量生物标记物的 mRNA、蛋白质或者 mRNA 和蛋白质的水平以筛选生物标记物。当所述生物标记物为 mRNA 水平时，可以通过选自聚合酶链式反应、实时聚合酶链式反应、逆转录酶聚合酶链式反应、杂交、探针杂交、和基因表达阵列的方法进行定量。所述筛选的方法还可以包括检测生物标记物中的多态性。可供选择地，筛选步骤可以通过使用至少一种选自聚合酶链式反应、异源双链分析、单链构象多态性分析、连接酶链式反应、比较基因组杂交技术、DNA 印迹法、RNA 印迹法、蛋白质印迹法、酶联免疫吸附测定、荧光共振能量转移和测序的技术实现。为了使用本发明，样本可以是许多免疫细胞中的任何细胞，例如，白细胞或者它们的亚成分。

另一个实施方案包括通过组织样本诊断宿主对传染病的应答的方法，该方法包括从免疫组织样本中获得基因表达谱，其中，测量两种或者多于两种以上的以下基因的表达，例如，补充表 1-11 以及它们的组合中所列的基因。与本申请一起提交的长表格的全部内容引入本文作为参考。在本发明的一个实施例中，基因表达谱或者转录组值矢量可以包括表 1、4、5 和补充表 1-11 以及所它们的组合所列的任何基因，这些形成本发明公开内容的一部分，例如，某些基因可以形成用于区分比与细菌感染相关性高于与流感病毒感染相关性的基因的转录组矢量的一部分，例如，那些参与对病毒应答的基因（例如，cig5、DNAPTP6、IFI27、IFI35、IFI44、OAS1）、参与免疫应答的基因（例如，BST2、G1P2、LY6E、MX1）、参与抗凋亡的基因（例如，SON）、参与细胞生长和/或维持的基因（例如，TRIM14）、参与多种功能的基因（例如，APOBEC3C、C1orf29、FLJ20035、FLJ38348、HSXIAPAF1、KIAA0152、

PHACTR2 和 USP18)。为了区分与流感病毒感染相关性更高的与细菌感染相关性的基因,可以考虑参与翻译延长的基因(例如,EEF1G)、参与调节翻译起始的基因(例如,EIF3S5、EIF3S7、EIF4B)、参与蛋白质合成的基因(例如,QARS、RPL31、RPL4)、参与转录调节的基因(例如,PFDN5)、参与细胞粘附的基因(例如,CD44)、参与代谢的基因(例如,HADHA、PCBP2)以及参与多种功能的基因,例如,dJ507I15.1。用于生物标记物例如 RNA 的来源的组织可以为血液。在一个特定的实施方案中,从不同的患者而不是从患者和对照中获得基因图谱,并对它们进行比较。

另一个实施方案包括通过组织样本诊断宿主对特定的传染病的应答的方法,该方法包括从免疫组织样本中获得基因表达图谱或者转录组,其中,可以使用两个或者多于两个的以下基因以区分金黄色葡萄球菌感染和大肠杆菌感染;所述以下基因为,例如参与信号传导的基因(例如,CXCL1、JAG1、G 蛋白信号调节蛋白(RGS2))、参与代谢的基因(例如,GAPD)、PPIB、PSMA7、MMP9、p44S10、参与蛋白质靶向的基因(例如,TRAM2)、参与细胞内蛋白质运输的基因(例如,SEC24C)和参与各种功能的基因(例如,ACTG1、CGI-96、MGC2963 和 STAU)。相反地,可能存在更通常与大肠杆菌感染相关的而不是与金黄色葡萄球菌感染相关的基因,例如,参与细胞内信号传导的基因(例如,RASA1、SNX4)、参与调节翻译起始的基因(例如,AF1Q)、参与转录调节的基因(例如,SMAD2)、参与细胞粘附的基因(例如,JUP)、参与代谢的基因(例如,PP、MAN1C1)以及参与多种功能的基因(例如,FLJ10287、FLJ20152、LRRN3、SGPP1、UBAP2L)。用于生物标记物例如 RNA 的来源的组织可以为血液。所述基因图谱是从不同组的患者之间的比较而不是从患者和对照之间的比较来获得的。

在本发明的方法中,测定表达水平的步骤是通过测量一组基因的 mRNA 的表达量和/或测量一组基因的蛋白质的表达量进行的。测定表达水平的步骤可以通过使用寡核苷酸阵列进行,例如,从样本中分离一种或多种核酸的生物标记物并将它们与位于固体支持物上的已知核酸杂交。测定表达水平的步骤还可以通过使用 cDNA 来进行,cDNA 是使用从人细胞中收集的 mRNA 作为模板制备的。在一些实施方案中,可以使用可检测的标记以标记生物标记物和/或用于确定表达水平能够结合生物标记物的靶(例如,抗体)。筛选的步骤可以通过对生物标记物的 mRNA 水平、蛋白质的水平,或者 mRNA 和蛋白质的水平进行定量实现的。通常,可以检测所述生物标记物的 mRNA

水平,并且可以通过选自聚合酶链式反应、实时聚合酶链式反应、逆转录酶聚合酶链式反应、杂交、探针杂交、和基因表达阵列的方法进行定量。也可以通过检测生物标记物中的多态性进行筛选。用于测定表达水平的其它方式也可以通过使用至少一种选自以下的方法实现:聚合酶链式反应、异源双链分析、单链构象多态性分析、连接酶链式反应、比较基因组杂交技术、DNA印迹法、RNA印迹法、蛋白质印迹法、酶联免疫吸附测定、荧光共振能量转移和测序的技术。所述样本通常为血液,然而,也可以使用任何细胞,例如,白细胞、活组织检查的细胞、流体中的或者分泌物中的细胞等。在一些实施方案中,所述生物标记物可以为从血液中提取的蛋白质。

而本发明的另一种实施方案包括鉴定疑似患有传染病的人受试者的方法,该方法包括测定包括一种或多种以下基因作为所列的靶的生物标记物的表达水平,所述以下基因是细菌感染导致的相对于病毒感染导致的表达的过表达的基因:参与翻译延长的基因、EEF1G;参与调节翻译起始的基因、EIF3S5、EIF3S7、EIF4B;参与蛋白质合成的基因、QARS、RPL31、RPL4;参与转录调节的基因、PFDN5;参与细胞粘附的基因、CD44;参与代谢的基因、HADHA、PCBP2;参与多种功能的基因、dJ507I15.1。所述测定表达水平的步骤是通过检测一组基因的 mRNA 的表达量甚或是通过检测一组基因的蛋白质的表达量来进行的。

而在本发明的另一种鉴定疑似患有传染病的人受试者的方法中,其中,以下基因的过表达表示被金黄色葡萄球菌感染:参与信号传导的基因、CXCL1、JAG1、RGS2;参与代谢的基因、GAPD、PPIB、PSMA7、MMP9、p44S10;参与蛋白质靶向的基因、TRAM2;参与细胞内蛋白质运输的基因、SEC24C;以及参与各种功能的基因、ACTG1、CGI-96、MGC2963、STAU。

而在本发明的另一种鉴定疑似患有传染病的人受试者的方法中,其中,以下基因的过表达表示被大肠杆菌感染:参与细胞内信号传导的基因、RASA1、SNX4;参与调节翻译起始的基因、AF1Q;参与转录调节的基因、SMAD2;参与细胞粘附的基因、JUP;参与代谢的基因、PP、MAN1C1;参与多种功能的基因、FLJ10287、FLJ20152、LRRN3、LRRN3、SGPP1、UBAP2L。

而本发明的另一种方法包括用于测定样本的表型的计算机执行的方法,所述方法通过:获得许多样本的探针强度;根据所述样本探针强度诊断传染病;计算所述样本探针强度与参比探针强度之间的线性相关系数;和当

所述线性相关系数大于阈值时,认为假定的表型为所述样本的表型。在某些实施方案中,所述阈值可以为约 0.7-1 或者更高;而某些阈值包括至少为 0.8、至少为 0.9 和/或至少为 0.95。所述探针的强度可以选自从组织样本中获得的基因表达谱,其中,检测两种或者多于两种的以下基因的表达作为所列的靶:

金黄色葡萄球菌:参与信号传导的基因、CXCL1、JAG1、RGS2;参与代谢的基因、GAPD、PPIB、PSMA7、MMP9、p44S10;参与蛋白质靶向的基因、TRAM2;参与细胞内蛋白质运输的基因基因、SEC24C;和参与各种功能的基因基因、ACTG1、CGI-96、MGC2963、和 STAU;和它们的组合。

大肠杆菌:参与细胞内信号传导的基因、RASA1、SNX4;参与调节翻译起始的基因、AF1Q;参与转录调节的基因、SMAD2;参与细胞粘附的基因、JUP;参与代谢的基因 PP、MAN1C1;以及参与多种功能的基因、FLJ10287、FLJ20152、LRRN3、LRRN3、SGPP1、UBAP2L;和它们的组合。以及

流感病毒:参与对病毒应答的基因、cig5、DNAPTP6、IFI27、IFI35、IFI44、OAS1;参与免疫应答的基因、BST2、G1P2、LY6E、MX1;参与抗凋亡的基因、SON;参与细胞生长和/或维持的基因、TRIM14;参与多种功能的基因、APOBEC3C、C1orf29、FLJ20035、FLJ38348、HSXIAPAF1、KIAA0152、PHACTR2、USP18、ZBP1;和它们的组合。

本发明的另一种实施方案提供了一种计算机可读介质,该介质包括用于执行测定样本的基因型的方法的计算机可执行的指令,所述方法包括:获得许多样本的探针强度;根据对选自表 1、4、5 和/或补充表 1-11 以及它们的组合中所列六种或多于六种的基因的样本探针强度诊断传染病;和计算所述样本探针强度与参比探针强度之间的线性相关系数;和当所述线性相关系数大于阈值时,认为假定的表型为所述样本的表型。

本发明的另一种实施方案提供了鉴定宿主对传染病的免疫应答的系统,该系统包括用于检测基因表达的微阵列,其中,所述微阵列包括选自表 4、表 5 和补充表 1-11 以及它们的组合中所列的四种或多于四种的生物标记物;其中,从所述微阵列中获得的基因表达的数据通过阈值与对传染病的免疫应答相互关联。

本发明的另一种实施方案提供了一种用于诊断传染病的系统,该系统通过从微阵列中获得基因表达的数据,并测定由选自表 1、4 和/或 5 中所列的四种或多于四种的基因所表达的四种或多于四种的生物标记物;其

中,从所述微阵列中获得的基因表达的数据通过阈值至少为 0.8 而与宿主对传染病的免疫应答相互关联。为了使用本发明的系统,所述生物标记物可以选自 5、6、7、8、9、10、11、12 或 13 种基因或者基因模块,和选自一种或多种补充表中的基因,以及它们的组合,它们引入本文作为参考。

另一个实施方案提供了预示性基因阵列,该预示性基因阵列为用户化的基因阵列,其包括代表一个或者多个转录模块的基因的组合,其中,能够与所述用户化的基因阵列发生联系的患者转录组预示患者患有 SLE(系统性红斑狼疮)。所述阵列可以用于监控患者对 SLE 治疗的应答。所述阵列还可以用于区分免疫疾病、病毒感染、细菌感染、癌症和移植排异反应。为了实现某些直接检测的目的,甚至可以将所述阵列组织成两个或者多于两个的转录模块,这些转录模块可以通过视觉扫描,并且表达的程度可以通过视觉分析,例如,通过肉眼和/或图像加工设备分析,例如,可以将所述阵列组织成具有一个或者多种亚模块的三种转录模块,所述亚模块选自 5、6、7、8、9、10、11、12 或 13 种基因或者基因模块,和选自补充表中的基因,以及它们的组合;其中,能够专一性结合到一种或多种基因的探针选自所述三种或者多于三种的模块中,并且这些探针指示患有本文中公开的传染病或者其它病症。

本发明的另一个实施方案包括用于筛选临床试验的患者的方法,该方法包括获得预期患者的转录组;将该转录组与一个或者多个指示疾病或者病症的转录模块进行比较,所述疾病或者病症为将在临床试验中被治疗的;根据一种或者多种在患者转录组中表达的基因的存在、不存在或者表达水平确定患者为合适的候选者的可能性,所述患者转录组处于一个或者多个与临床试验具有很好的相关性的转录模块中。为了使用所述方法,每一个模块可以包括与样本中的转录子的总的比例相关的矢量;一种或多种疾病或者病症与一种或多种矢量相关的矢量;能够与一种或多种位于每一个模块中的基因的表达相关的矢量和/或者能够用于检测、特征化、诊断、预后和/或监控正常个体或者患有传染病、先天性疾病、退行性疾病、获得性疾病或者其它疾病的患者的矢量。

附图的简要说明

为了更全面地理解本发明的特征和优点,参比本发明详细描述的说说明书以及附图,其中:

图 1 表示可以区分被流感病毒 A 感染的患者与被细菌感染的患者。图 1a 表示通过对两组患者进行曼恩-惠特尼 (Mann-Whitney) 等级检验比较 ($p < 0.01$) 获得的 854 基因的等级聚类, 所述两组患者分别为: 被流感病毒 A (流感病毒 A, 11 个样本, 绿色的矩型) 感染的患者和被大肠杆菌 (大肠杆菌, 6 个样本) 或被肺炎链球菌 (肺炎链球菌, 6 个样本) 感染的患者 (被细菌感染的患者) (红色的矩型)。用颜色标准指示改变的表达水平, 红色表示与每个基因在所有的供体中的平均表达值相比其表达水平相对较高, 蓝色表示与每个基因在所有的供体中的平均表达值相比其表达水平相对较低。黑条表示干扰素 (IFN) 可诱导的基因, 而蓝条表示参与蛋白质合成的基因。基因列在补充表 2 中。图 1b 表示使用由监督学习运算法则得到的结果鉴定 35 个代表能最大可能区分两类疾病的基因 (表 1 和补充表 3)。通过对具有 35 个基因的训练组进行弃一交差验证 (Leave-one-out cross-validation) 对样本进行分类, 其准确率为 91%。预测的类用明亮的彩色实心矩形表示 (绿色表示受流感病毒 A 感染, 红色表示受细菌的感染)。有两位被细菌感染的患者被错误分类。图 1c 表示在一组独立的患者中对这样鉴定的一共 35 个分类基因进行的测试 (空心的矩形), 所述一组独立的患者包括 7 例被流感 A 病毒感染的新患者 (绿色), 23 例被大肠杆菌 (红色) 感染的患者, 7 例被肺炎链球菌感染的患者。在本组试验中, 对 37 个样本进行了分类, 其准确率为 95% (预测的类用明亮的彩色矩形表示)。在分类预测中, 一例被错误分类, 一例的分类不确定 (灰色框)。图 1d 表示在 7b 中鉴定出的 35 个分类基因, 在一组独立的患者中对这 35 个分类基因进行了测试 (空心的矩形), 所述一组独立的患者包括 7 例被流感病毒 A (Inf A) 感染的新患者, 31 例被金黄色葡萄球菌感染的患者。对这 38 个样本进行分类, 其准确率为 87%。

图 2 表示能够区分被流感病毒 A 感染的患者与被细菌感染的患者的 35 个分类基因的表达水平。以点对在图 7b 中描绘的 35 个分类基因的成比例的表达水平进行了描绘 (平均区别强度), 这 35 个分类基因能够区分被流感病毒 A 感染的患者 (11 例, 绿色的正方形) 与被细菌感染的患者 (6 例被大肠杆菌感染, 6 例被肺炎链球菌感染, 红色的菱形)。每一个点表示一个样本, 线表示平均表达值。

图 3a-3e 表示可能区分被金黄色葡萄球菌感染的患者和被大肠杆菌感染的患者。图 9a 表示通过对两组患者进行曼恩-惠特 (Mann-Whitney) 等级检验比较 ($p < 0.01$) 获得的 211 个基因的等级聚类, 所述两组患者为被金黄色葡

萄球菌感染的患者（金黄色葡萄球菌, 10 例, 红色的矩形）和被大肠杆菌感染的患者（大肠杆菌, 10 例, 蓝色矩形）。用颜色标准指示改变的表达水平, 红色表示与每个基因在所有的供体中的平均表达值相比其表达水平相对较高, 蓝色表示与每个基因在所有的供体中的平均表达值相比其表达水平相对较低。基因列在补充表 4 中。图 3b 表示使用由监督学习运算法则得到的结果鉴定 30 个代表能最大可能区分两类疾病的基因（还参见补充表 6）。通过对具有 30 个基因的训练组进行弃一交叉验证对样本进行分类, 其准确率为 95%。图 3c 表示在一组独立的患者中对这样鉴定的一共 30 个分类基因进行的测试（空心的矩形），所述一组独立的患者包括 21 例被金黄色葡萄球菌感染的新患者, 19 例被大肠杆菌感染的患者。在本组试验中, 对 40 例患者进行分类, 其准确率为 85%（预测的类用明亮的彩色矩形表示）。在 40 例患者中, 只有 2 例被错误分类, 有 4 例不能确定其类别（空心的矩形）。

图 3d 和 3e 表示通过实时 RT-PCR 验证差异表达的基因。图 3d 表示通过实时 RT-PCR 对从被金黄色葡萄球菌 (Sa)或被大肠杆菌 (Ec) 感染的患者中获得的样本测量了 9 个基因的表达水平（相对于健康对照组的基因表达水平, 基因表达的倍数发生改变, 除了 RGS2、FCAR 和 ALOX 外, 其它基因的测量结果均进行了对数转换）。每一个点表示一个样本, 线表示平均表达值。图 3e 表示通过实时 RT-PCR 分析（横坐标）和微阵列分析（纵坐标-相对于从相同健康的对照中获得的样本的表达水平进行了标准化, 实时 RT-PCR 的数据也相对于从相同健康的对照中获得的样本的表达水平进行了标准化; 对数标准）获得的表达值之间的联系。详细内容请参见补充表 5。

图 4a -4e 表示能够区分被大肠杆菌感染的患者和被金黄色葡萄球菌感染的患者的 30 个分类基因的表达水平。以点对在图 3b 中描绘的 30 个分类基因的成比例的表达水平进行了描绘（平均区别强度），这 30 个分类基因能够区分被大肠杆菌感染的患者（10 例, 蓝色的正方形）与被金黄色葡萄球菌感染的患者（10 例, 红色的菱形）。每一个点表示一个样本, 线表示平均表达值。图 4b-4e 表示本发明可以用于区分被不同细菌感染的患者。图 4b 表示通过对两组患者进行曼恩-惠特（Mann-Whitney）等级检验比较($p < 0.01$)获得的 242 个基因的等级聚类, 所述两组患者为被大肠杆菌感染的患者（11 例）和被肺炎链球菌感染的患者（11 例）。用颜色标准指示改变的表达水平, 红色表示与每个基因在所有的供体中的平均表达值相比其表达水平相对较高, 蓝色表示与每个基因在所有的供体中的平均表达值相比其表达水平相对

较低。基因列在补充表 7 中。图 4c 表示使用由监督学习运算法则得到的结果鉴定代表能最大可能区分两类疾病的基因。通过对具有 45 个预示基因的训练组进行弃一交差验证对样本进行分类,其准确率为 85% (20/22)。分类基因列在补充表 8 中。图 4d 表示来自通过对两组患者进行曼恩-惠特 (Mann-Whitney) 等级检验比较($p<0.01$)获得的 127 个基因的无监督的等级聚类的结果,所述两组患者为被金黄色葡萄球菌的患者 (12 例) 和被肺炎链球菌感染的患者 (11 例)。用颜色标准指示改变的表达水平,红色表示与每个基因在所有的供体中的平均表达值相比其表达水平相对较高,蓝色表示与每个基因在所有的供体中的平均表达值相比其表达水平相对较低。基因列在补充表 9 中。图 4e 表示使用由监督学习运算法则鉴定代表能最大可能区分两类疾病的基因。通过对具有 30 个基因的训练组进行弃一交差验证对样本进行分类,其准确率为 83% (19/23)。分类基因列在补充表 10 中。

图 5 表示从患有急性呼吸性感染的患者中获得的循环中的白细胞的基因表达的独特模式。图 5a 表示使用被发现的能够区分被金黄色葡萄球菌感染的患者和被大肠杆菌感染的患者 (维恩图解 (Venn diagram), 右: 能够区分被金黄色葡萄球菌感染的患者和被大肠杆菌感染的患者的基因; 图 2 和补充表 6) 的 30 种分类基因,以鉴定能够区分被金黄色葡萄球菌感染的患者和被肺炎链球菌感染的患者的 30 种基因 (维恩图解, 左: 能够区分被金黄色葡萄球菌感染的患者和被肺炎链球菌感染的患者的基因; 图 5a 和补充表 10), 以及以够鉴定能够区分被大肠杆菌感染的患者和被肺炎链球菌感染的患者 (维恩图解, 下: 能够区分被大肠杆菌感染的患者和被肺炎链球菌感染的患者的基因; 补充图 5b 和补充表 8) 的 45 种基因。在这些组的任意两组之间一共只有 3 种基因是共享的。在图 5b 中, 将在图 5a 中所示能够区分被不同细菌感染的患者的 3 组基因合并 (102 种独特的基因, 维恩图解, 左), 并且将它们与能够用于区分被流感病毒 A 感染的患者与细菌感染的患者的分类基因 (35 种基因, 维恩图解, 右; 图 5b 和补充表 3) 进行比较。这两组没有共同的基因。图 5c 表示将能够区分被流感病毒 A 感染的患者与细菌感染的患者的 137 种分类基因与三组能够区分被不同细菌感染的患者的分类基因合并, 并用于在 27 例患有呼吸性感染的患者和 7 名健康的志愿者中创建具有辨别力的表达模式。这些值是相对于每一个基因在所有的供体中的平均表达值进行了标准化的。病症的聚类分析将样本分成主要的四类。属于被

流感病毒 A 感染的组的四个样本和被金黄色葡萄球菌感染的组的一个样本形成以混合的符号(*)表示的不同的亚组。

图 6 表示对用于传染病监控的显著性模式的分析。将对每一组患者测量的基因表达水平与从由健康的志愿者组成的对照组中获得的结果进行比较(曼恩惠特尼 (Mann Whitney)U 检验)。将选择的标准应用于由被流感 A 感染的患者(FLU)或者患有系统性红斑狼疮(SLE)的患者产生的 p 值。左列: 过表达基因; 右列: 低表达基因; 上部的箭头: 在 FLU 和 SLE 中均发生显著的变化($p < 0.01$); 中部的箭头: 在 SLE 中发生显著的变化 ($p < 0.01$), 但在 FLU 中没有发生显著的变化 ($p > 0.5$); 底部的箭头: 在 FLU 中发生显著的变化($p < 0.01$), 但在 SLE 中没有发生显著的变化($p > 0.5$)。将基因按照 p 值进行等级聚类排列。颜色标准: 绿色表示低的 p 值; 黄色和白色表示高的 p 值; 系统树图的蓝色分枝表示疾病特异性的符号 (C1-C4; 详细内容请参见补充表 11)。

图 7 表示可以用于在模块水平对转录变化进行作图的基因矢量鉴定疾病特异性模式。

图 8 表示采用微阵列对患有急性感染的患者的疾病严重程度进行评估所获得的分值。

图 9a-9c 概括了通过微阵列平台进行的独立的确认与验证。

发明详述

尽管将在下文详细讨论制备和使用本发明的各种实施方案,但是,应当理解的是本发明提供了多种可适用的创造性概念,这些创造性概念可以广泛包含在具体的上下文中。本文中讨论的具体实施方案仅仅以示例性的具体方式来制备和使用本发明,决不是为了限制本发明的范围。

为了有助于理解本发明,在下文中对许多术语进行了定义。本文中定义的术语具有与本发明相关的领域的普通技术人员所理解的普通含义。术语,例如“一个”、“一种”和“所述的”不仅指单个的实体,还包括概括的一类,在该类中,采用一个具体的实例作为说明。本文中的术语用于描述本发明的具体的实施方案,除了在权利要求书中外,它们的使用并不限定本发明。除非另有定义,本文中所有的科技术语具有本发明术语技术领域的普通技术人员所理解的通常含义。以下的参考文献给本领域的技术人员提供了本发明中所使用的许多术语的一般定义: Singleton, et al., Dictionary Of Microbiology And

Molecular Biology (2d ed. 1994); The Cambridge Dictionary Of Science And Technology (Walker ed., 1988); The Glossary Of Genetics, 5th Ed., R. Rieger et al. (eds.), Springer Verlag (1991); 和 Hale & Marham, The Harper Collins Dictionary Of Biology (1991)。

在本领域中, 许多生物化学和分子生物学的方法是公知的。例如, 分离和纯化核酸的方法详细描述在 WO 97/10365、WO 97/27317 和生物化学和分子生物学的实验室技术的第 3 章: 用核酸探针杂交的第 I 部分 (Chapter 3 of Laboratory Techniques in Biochemistry and Molecular Biology: Hybridization With Nucleic Acid Probes, Part I)、原理和核酸制备 (Theory and Nucleic Acid Preparation, (P. Tijssen, ed.) Elsevier, N.Y. (1993))、生物化学和分子生物学的实验室技术的第 3 章: 用核酸探针杂交的第 I 部分 (Chapter 3 of Laboratory Techniques in Biochemistry and Molecular Biology: Hybridization With Nucleic Acid Probes, Part I)、原理和核酸制备 (Theory and Nucleic Acid Preparation, (P. Tijssen, ed.) Elsevier, N.Y. (1993))、Sambrook 等, 分子克隆: 实验室手册 (Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, N.Y., (1989))、和分子生物学中当前使用的方法 (Current Protocols in Molecular Biology, (Ausubel, F. M. et al., eds.) John Wiley & Sons, Inc., New York (1987-1999)), 包括增刊, 例如增刊 46 (April 1999)。

生物信息学定义

本文中所使用的“目标”是指感兴趣的任何项目或者信息 (一般是按原文的 (generally textual), 包括名词、动词、形容词、副词、短语、句子、符号、数字特征等)。因此, 目标可以是能够形成关系的任何事物以及可以被获得的、鉴定的、和/或从一个源搜索的任何事物。“目标”包括但不限于感兴趣的实体, 例如, 基因、蛋白质、疾病、表型、机制、药物等。在一些方面, 目标可以是如下文将进一步描述的数据。

本文中所使用的“关系”是指在相同的单元 (例如, 短语、句子、两行或者多于两行的文本、段落、网页的部分、页、杂志、报纸、书籍等) 中同时出现的目标。关系可以是文本 (text)、符号、数字及其组合。

本文中所使用的“元数据内容” (“meta data content”) 是指在数据源中组织文本的信息。元数据 (meta data) 可以包括标准的元数据, 例如都柏林核元数据 (Dublin Core metadata); 或者元数据可以是特定收集的。元数据

的格式的实例包括但不限于用于图书馆目录的机读目录 (Machine Readable Catalog (MARC))格式、资源描述格式 (Resource Description Format (RDF))和可扩展标记语言 (Extensible Markup Language (XML))。元目标可以通过人工产生, 或者通过自动信息提取运算法则产生。

本文中所使用的“发动程序”(“engine”)是指为其它的程序执行核心的或者实质的功能的程序。例如, 发动程序可以为操作系统中的中心程序, 或者为协调其它程序全面操作的应用程序。术语“发动程序”还可以指含有可以改变的运算法则的程序。例如, 可以设计知识发现发动程序, 使该程序中包括的鉴定关系的方法可以改变, 以反映鉴定和排列关系的新规则。

本文中所使用的“统计分析”是指一种基于计数每个术语(词、词根、词干、n元文法(n-gram)、短语等)出现次数的技术。在受试者方面的收集, 在不同上下文中所使用的短语可以表示不同的概念。对短语共同出现进行统计分析可以帮助解决词义不明确的问题。“统计分析”可以通过部分的语言分析进一步降低不明确性。本文中所使用的, 一个或多个这种分析更通常被称为“词法分析”。“人工智能(AI)”是指一种方法, 通过该方法, 非人类的设备(例如, 计算机)能够执行人类认为是值得注意的或者“智能的”任务。实例包括鉴定图片、理解口语单词或者书面的文本, 并解决问题。

本文中所使用的术语“数据库”是指未加工的数据或者编辑的数据的储存库, 甚至在数据领域中的各方面的信息都可以在其中被找到。通常, 对数据库进行组织, 以便使其内容能够被获取、管理和更新(例如, 数据库是动态的)。在本发明中, 术语“数据库”和“资源”还可以相互交换使用, 因为数据和信息的第一手来源为数据库。然而, “源数据库”或者“源数据”通常是指被输入系统中以鉴定目标和确定关系的数据, 例如, 未组织的文本和/或组织的数据。源数据库可以是或者可以不是关系数据库。然而, 系统数据库通常包括关系数据库或者一些等价类型的数据库, 在这些关系数据库或者一些等价类型的数据库中存储了与目标之间的关系有关的值。

本文中所使用的“系统数据库”和“关系数据库”可以交换使用, 并且它们是指一次或者多次收集的数据被组织成一组表, 在该表中, 数据被分配到预定的类别中。例如, 数据库表可以包括一个或者多个通过列(例如, 属性)限定的类别, 而数据库的行可以含有通过列限定的类别的独特的目标。因此, 目标(例如, 基因的鉴定)可以以所述基因的存在、不存在和/或表达水平作

为列。一行相关的数据库还可以被称为“集”，并通常由其列的值限定。相关数据库的上下文中的“域”是一个区域（例如，列）可能包括的有效值的范围。

本文中所使用的“知识域”是指研究领域，在该领域中，系统对于，例如，所有的生物医药数据都是可操作的。应当指出的是，对来自几个域中的数据进行合并是有益的，例如，对生物医药数据和工程数据进行合并，因为不同的数据有时能够将对于一个仅仅熟悉一个领域或者研究/学习领域（一个域）的普通人来说不能结合到一起的事物结合到一起。“分布式数据库”是指可以被分散在或者复制在网络中的不同的点中的数据库。

术语“数据”和“信息”通常可以交换使用，“信息”和“知识”也可以交换使用。本文中所使用的“数据”是最基本的单元，它是一个实验测量结果或者一组实验测量结果。对数据进行编辑以构成信息，但信息的基础不依赖于数据。相反，信息来源于感兴趣的事物，例如，可以根据种族、性别、身高和饮食对数据（单元）进行聚集，以找到与患心血管疾病风险相关的各种变量。然而，相同的数据还可以用于形成配方或者创建关于饮食偏好的“信息”，即在超市中的某些产品可能更容易被出售的可能性。

本文中所使用的“信息”是指可以包括数字、字母、数字集、字母集、或者由一组数据得到的或者产生的结果的数据集。因而，“数据”是测量结果或者是统计的和基楚的信息单元。“信息”还可以包括其它类型的数据，例如，字、符号、文本（例如，未组织的自由文本）、编码等。“知识”不严谨地定义为有助于充分理解模拟因果的系统的一组信息。为了扩展以前的实例，关于人口、性别和现有购买的物品的信息可以用于形成对商品销售的区域性市场战略；而关于民族的信息可以被顾客用作进口产品指南。必须指出，数据、信息和知识之间并没有严格的界线，这三个术语有时可以认为是等同的。通常，数据来源于检测；信息来源于相互关联；而知识来源于建模。

本文中所使用的“程序”或者“计算机程序”通常是指语法单元，该语法单元遵循特定的程序设计语言的规则，并且由陈述和声明或者指令组成，可以分成用于解决或者执行某种功能、任务、或者问题的“编码段”。程序设计语言通常是用于表达程序的人工语言。

本文中所使用的“系统”或者“计算机系统”通常是指一个或多个计算机、外围设备、和进行数据处理的软件。“用户”或者“系统操作员”通常是指能够通过“用户设备”（例如，计算机、无线电设备等）进入计算机网络以进行数

据处理和信息交换的人。“计算机”通常是指不需要人为的干涉、能够进行大量计算的功能单元，所述计算包括大量的算术操作和逻辑操作。

本文中所使用的“应用软件”或者“应用程序”通常是指专门用于解决应用问题的软件或程序。“应用问题”通常是指由终端用户提交的，并需要信息处理以获得解决方案的问题。

本文中所使用的“自然语言”是指其规则基于目前的使用而不受特定的规定的语言，例如，英语、西班牙语或者汉语。本文中所使用的“人工语言”是指其规则在使用前已经明确制定的语言，例如，计算机程序设计语言。例如，C++、Java、BASIC、FORTRAN、或 COBOL。

本文中所使用的“统计相关性”是指使用一种或多种排列方案(O/E 比值、强度等)，如果一种相互关系出现的频率显著高于预测的随机出现的频率，则确定的相互关系为统计相关的。

本文中所使用的术语“协同调节的基因”或者“转录模块”，它们可以相互交换使用，并且是指特定基因的、成组的、基因表达图谱（例如，与特定基因序列相关信号值）。可以给一个或多个“协同调节的基因”的组合进行赋值，以提供可以表示为单个值的“转录组值矢量”或者“转录组矢量”。例如，所述值可以以数字、位于辐射图中的点、具有不同强度的点、颜色、数值或者轮廓线（例如，上升图）的形式提供。每一个转录模块可以与一份或多份的数据相关，例如，文献搜索部分与从基因微阵列中获得的实际实验基因表达值数据。被选入转录模块的一组基因是基于基因表达数据的分析进行的（上述的模块抽提算法）。Chaussabel, D. & Sher A.教导了其它步骤（Chaussabel, D. & Sher, A. Mining microarray expression data by literature profiling. *Genome Biol* 3, RESEARCH0055 (2002), (<http://genomebiology.com/2002/3/10/research/0055>))，其相关部分引入本文作为参考，并且表达数据从感兴趣的疾病或者病症中获得，例如从系统红斑狼疮、关节炎、淋巴瘤、癌症、恶性黑素瘤、急性感染、自身免疫疾病、自身炎症疾病等中获得）。

下表中列出了用于形成文献搜索部分或者用于组成转录模块的关键词的实例。本领域的技术人员能够认识到可以容易地选择其它术语用于其它的病症，例如，用于特定的癌症、特定的传染病、移植等。例如，在下文中将描述基因以及用于那些与 T 细胞激活的基因的信号作为模块 ID “M 2.8”，在该模块中，某些关键词（例如，淋巴瘤、T 细胞、CD4、CD8、TCR、胸腺、淋巴样的、IL2）被用于鉴定重要的与 T 细胞相关的基因，例如，T 细

胞表面标记 (CD5、CD6、CD7、CD26、CD28、CD96)；由淋巴细胞表达的分子 (淋巴毒素 β 、IL2-可诱导的 T 细胞激酶、TCF7；和 T 细胞分化蛋白 mal、GATA3、STAT5B)。其次，由来自患者群体的用于这些基因的相关联数据 (不管平台、存在/不存在和/或上调或下调) 形成完整的模块以产生转录模块。在某些情况下，基因谱与用于这些疾病条件的基因的任何特定的聚类以及数据不匹配 (在这时候)，然而，在某些生理途径 (例如，cAMP 信号传导、锌指蛋白、细胞表面标记等) 中发现存在这种“证据不足地说明的”模块。事实上，基因表达数据组可以在进行匹配的关键词搜索前用于提取能够协同表达的基因，即，在参照第二数据组之前，每一数据组都可以是相互关联的。

表 1. 处于不同模块中的基因的实例

模块标志	探针组的数目	关键词选择	评估
M 1.1	76	Ig、免疫球蛋白、骨、骨髓、PreB、IgM、Mu	浆细胞。包括编码免疫球蛋白链 (例如，IGHM、IGJ、IGLL1、IGKC、IGHD) 和浆细胞标记 CD38 的基因。
M 1.2	130	血小板、粘附、聚集、内皮的、血管	血小板。包括编码血小板糖蛋白 (ITGA2B、ITGB3、GP6、GP1A/B) 的基因、编码血小板衍生的免疫调节子如 PPPB (前血小板碱性蛋白) 和 PF4 (血小板因子 4) 的基因
M 1.3	80	免疫受体、BCR、B 细胞、IgG	B 细胞。包括编码 B 细胞表面标记 (CD72、CD79A/B、CD19、CD22) 的基因，和编码其它与 B 细胞相关的分子：早期 B 细胞因子 (EBF)、B 细胞连接蛋白 (BLNK) 和 B 淋巴细胞酪氨酸激酶 (BLK) 的基因。
M 1.4	132	复制、修复、CREB、淋巴的、TNF- α	未确定。这一组包括 cAMP 信号途径 (JUND、ATF4、CREM、PDE4、NR4A2、VIL2) 的调节子和靶，TNF- α 介导的 NF-KB 激活 (CYLD、ASK、TNFAIP3) 的抑制子。
M 1.5	142	单核细胞、树突状的、MHC、协同刺	髓系。包括由髓系细胞表达的分子 (CD86、CD163、FCGR2A) 它们中的一些参与病原体的识别 (CD14、

		激、TLR4、MYD88	TLR2、MYD88)。这一组还包括 TNF 家族成员 (TNFR2、BAFF)。
M 1.6	141	锌指、P53、RAS	未确定。这一组包括编码信号传递的分子的基因, 例如, 含有激活的 STAT (PIAS1 和 PIAS2) 抑制剂的锌指, 或激活的 T 细胞 NFATC3 的核因子。
M 1.7	129	核糖体、翻译的、40S、60S、HLA	MHC/核糖体蛋白。几乎仅仅由编码 MHC 类 I 分子 (HLA-A、B、C、G、E) + β 2-微球蛋白 (B2M)或核糖体蛋白(RPLs、RPSs)的基因形成
M 1.8	154	代谢、生物合成、复制、蜗牛酶抗体	未确定。包括编码代谢酶 (GLS、NSF1、NAT1) 的基因以及编码涉及 DNA 复制 (PURA、TERF2、EIF2S1) 的因子的基因。
M 2.1	95	NK、杀伤细胞的、溶细胞的、CD8、细胞介导的、T 细胞、CTL、IFN-g	细胞毒性细胞。包括细胞毒性 T 细胞和 N 细胞表面标记 (CD8A、CD2、CD160、NKG7、KLRs)、溶细胞分子 (颗粒酶、穿孔素 (perforin)、颗粒溶素 (granulysin)、细胞因子 (CCL5、XCL1) 和与 CTL/NK 细胞相关的分子 (CTSW)。
M 2.2	49	粒细胞、嗜中性的、防御、髓的、骨髓的 (Marrow)	嗜中性的。这一组包括在嗜中性粒细胞中发现的天然分子 (乳运铁蛋白 (LTF)、防御素 (DEAF1)、细菌通透性增加蛋白 (BPI)、阳离子抗菌蛋白 (CAMP))
M 2.3	148	红细胞、红色、贫血、球蛋白、血红蛋白	红细胞。包括血红蛋白基因 (HGBs)和其它与红细胞相关的基因 (红细胞 ankirin (ANK1)、血型糖蛋白 C (GYPC)、羟甲基胆素合成酶 (HMBS)、网织红细胞关联因子 (erythroid associated factor): (ERAF)。
M 2.4	133	核糖核蛋白、60S、细胞核、组装、延长	核糖体蛋白。包括编码核糖体蛋白的基因 (RPLs、RPSs)、真核翻译延长因子家族成员 (EEFs)和核蛋白 (NPM1、NOAL2、NAP1L1)。
M 2.5	315	腺瘤、间质的、间冲质 (Mesenchyme)、树突、运动	未确定。该模块包括编码与免疫相关的基因 (CD40、CD80、CXCL12、IFNA5、IL4R), 以及编码与骨架相关的分子的基因 (肌球蛋白、胞质分裂作用因子 (Dedicator of Cytokinesis)、多配体聚糖 (Syndecan) 2、Plexin C1、短棒菌素 (Distrobrevin))。

M 2.6	165	粒细胞、单核细胞、髓的、ERK、坏死	髓系。包括在髓系中表达的基因 (IGTB2/CD18、淋巴毒素 β 受体、与髓相关的蛋白 8/14 甲酸基肽受体 1), 例如单核细胞和嗜中性白细胞。
M 2.7	71	没有提出的关键词	未确定。这一模块主要由功能不清楚的转录因子组成。只有 20 基因与文献有联系, 包括大量的趋化因子样因子超家族 (CKLFSF8)。
M 2.8	141	淋巴瘤、T-细胞 CD4、CD8、TCR、胸腺、淋巴样的、IL2	T 细胞。包括 T 细胞表面标记 (CD5、CD6、CD7、CD26、CD28、CD96) 和由淋巴瘤细胞表达的分子 (淋巴毒素 β 、IL2-可诱导的 T 细胞激酶、TCF7、T 细胞分化蛋白 mal、GATA3、STAT5B)。
M 2.9	159	ERK、转录激活、细胞骨架、MAPK、JNK	未确定。包括编码与细胞骨架 (与蛋白 2/3 的肌动蛋白、MAPK1、MAP3K1、RAB5A) 相关的分子的基因。其它存在的为 T 细胞表达的基因 (FAS、ITGA4/CD49D、ZNF1A1)。
M 2.10	106	髓的、巨噬细胞树突的、炎症的、白介素	未确定。包括编码与免疫相关的细胞的表面分子 (CD36、CD86、LILRB) 的基因、编码细胞因子 (IL15) 和涉及信号传导途径的分子 (FYB、TICAM2-Toll-样受体途径) 的基因
M 2.11	176	复制、抑制、RAS、自动磷酸化、致癌的	未确定。包括激酶 (UHMK1、CSNK1G1、CDK6、WNK1、TAOK1、CALM2、PRKCI、ITPKB、SRPK2、STK17B、DYRK2、PIK3R1、STK4、CLK4、PKN2) 和 RAS 家族成员 (G3BP、RAB14、RASA2、RAP2A、KRAS)。
M 3.1	122	ISRE、流行性感、抗病毒的、IFN- γ 、IFN- α 、干扰素	干扰素可诱导的。这一组包括干扰素可诱导的基因: 抗病毒分子 (OAS1/2/3/L、GBP1、G1P2、EIF2AK2/PKR、MX1、PML)、细胞因子 (CXCL10/IP-10)、信号传导分子 (STAT1、STAT2、IRF7、ISGF3G)。
M 3.2	322	TGF- β 、TNF、炎症的、凋亡的、脂多糖	炎症 I。包括编码涉及炎症过程的分子 (例如、IL8、ICAM1、C5R1、CD44、PLAUR、IL1A、CXCL16) 的基因, 和编码凋亡的调节子 (MCL1、FOXO3A、RARA、BCL3/6/2A1、GADD45B)。的基因

M 3.3	276	炎症、防御、溶酶体、氧化应激、LPS	炎症 II。包括能被炎症诱导或者可诱导的分子 (IL18, ALOX5、ANPEP、AOAH、HMOX1、SERPINB1)、以及溶酶体酶 (PPT1、CTSB/S、NEU1、ASAHI、LAMP2、CAST)。
M 3.4	325	连接酶、激酶、KIP1、泛素、伴侣蛋白	未确定。包括蛋白质磷酸酯酶 (PPP1R12A、PTPRC、PPP1CB、PPM1B) 和肌醇磷脂 3-激酶 (PI3K) 家族成员 (PIK3CA、PIK32A、PIP5K3)。
M 3.5	22	没有提取的关键词	未确定。只由少量的转录子组成。包括血红蛋白基因 (HBA1、HBA2、HBB)。
M 3.6	288	核糖体的、T 细胞、 β - 连 锁 蛋 白 (β -catenin)	未确定。这一组包括线粒体核蛋白 (MRPLs、MRPs)、线粒体延长因子 (GFM1/2)、筛选连接素 (Sortin Nexins) (SN1/6/14) 以及溶酶体 ATPases (ATP6V1C/D)。
M 3.7	301	剪接体、甲基化、泛素	未确定。包括编码蛋白酶体亚基 (PSMA2/5、PSMB5/8)、泛素蛋白连接酶 HIP2、STUB1、以及泛素连接酶复合物 (SUGT1) 的成分的基因。
M 3.8	284	CDC、TCR、CREB、糖基化酶	未确定。包括编码以下酶的基因：氨甲基转移酶、精氨酰基转移酶、天冬酰胺合成酶、甘油二酯激酶、肌醇磷酸镁、甲基转移酶、蜗牛酶。
M 3.9	260	染色质、检测点、复制、转录激活	未确定。包括编码激酶 (IBTK、PRKRIR、PRKDC、PRKCI) 和磷酸酯酶 (例如, PTPLB、PPP2CB/3CB、PTPRC、MTM1、MTMR2) 的基因

生物学定义

本文中所使用的术语“阵列”是指具有一种或多种连接到支持物上的肽或者核酸探针的固体支持物或者底物。阵列通常具有结合到底物表面不同的已知的位置上的一种或多种不同的核酸探针或者肽探针。这些阵列也描述为“微阵列”、“基因芯片”或 DNA 芯片，它们可以具有基于已知基因组（例如，人基因组）的 10,000、20,000、30,000、或 40,000 不同的可鉴定的基因。这些泛阵列被用于检测在样本中表达或者发现的基因的完整的“转录组”或者转录库，例如，表达为 RNA、mRNA 的核酸以及将被进行 RT 和/或 RT-PCR 以制备 DNA 复制子的互补组的核酸。可以使用机械合成的方法、光导合成

方法、以及组合了非平板印刷方法和/或照相平板印刷方法以及固相合成的方法的方法产生阵列。可以使用包括连接到3微米的珠的50-mer的寡核苷酸探针的珠阵列，该珠阵列为，例如被固定到微孔中的载玻片的表面，或者为液相悬浮阵列（例如，Luminex或Illumina）的一部分，所述液相悬浮阵列为位于液相中的数字珠阵列，并使用“条形码”以进行检测和鉴定。

已经描述了用于合成这些核酸阵列的各种技术，例如，被制备在事实上任何形状的表面或者甚至在多种表面上。阵列可以是位于珠、凝胶、聚合物表面、纤维（例如，光导纤维）、玻璃或者任何其它合适的表面上的肽或者核酸。阵列可以被包装成用于诊断或者用于包含一切的设备（all inclusive device）的其它操作，实例请参见美国专利 No. 6,955,788，其相关内容引入本文作为参考。

本文中所使用的术语“疾病”是指其细胞处于任何异常的生物状态的生物的生理状态。疾病包括但不限于细胞的、组织的、肌体的功能、系统的或者器管的功能被中断、停止或者失调，这些中断、停止或者失调可以是固有的、遗传的、由感染导致的、异常细胞功能导致的、细胞异常分裂等导致的。能够导致“疾病状态”的疾病通常对于生物系统（即，疾病的宿主）有害。根据本发明，与疾病或者失调相关的任何生物状态（例如，感染的（例如，病毒感染、细菌感染、真菌感染、蠕虫感染等）、炎症的、自身炎症的、自身免疫的、过敏性反应的、过敏症的、癌前期的、癌症的、外科的、移植的、生理的等生物状态）均被认为是疾病状态。病理状态通常等价于疾病状态。

疾病状态还可以分为不同级别的疾病状态。本文中所使用的疾病或疾病状态的级别是能够反映疾病或者疾病状态的进程以及反映在治疗时、治疗中以及治疗后的生理应答的任何度量标准。通常，疾病或者疾病状态向高级的级别或者阶段发展，其中，疾病的侵害变得越来越严重。疾病状态的级别可能受样本中的细胞的生理状态的影响。

本文中所使用的术语“治疗”或者“治疗方案”是指采取的用于减轻或者改变疾病状态的医学步骤，例如，使用药理学的、外科的、善食的和/或其它技术以减轻或者缓解疾病的侵害或症状。治疗方案可以包括开出一定剂量的一种或多种药物，或者进行外科手术。通常，治疗是有益的，能够减轻疾病的状态，但在许多情况下，治疗的作用包括不需要的作用或者副作用。治疗的作用还受宿主的生理状态影响，例如，宿主的年龄、性别、遗传、重量、其它疾病状况等。

本文中所使用的术语“药理学状态”或者“药理学状况”是指用一种或多种药物、外科手术等处理样本，以影响样本中的一种或多种核酸的药理学状态，例如，由于药理学的干预，一种或多种核酸进行了新的转录、被稳定和/或去稳定。样本的药理学状态涉及在药物处理之前、之中和/或之后生物状态的变化，并且可以作为本文所述的诊断或者预示的功能。在药物治疗或者外科手术后出现的一些变化可能与疾病的状态有关和/或可能与治疗的副作用不相关。药理学的状态的变化可能是由于治疗持续的时间、所开的药物的类型和剂量、患者在给定治疗时期依从的程度、和/或摄取的未规定的药物所导致的。

本文中所使用的术语“生物学状态”是指分离和纯化的用于分析表达的变化的细胞样本的转录组（转录组为全部收集的 RNA 转录子）的状态。通过测量细胞中组成的丰度和/或活性，根据形态学上的表型或者用于检测转录子的方法的组合进行表征，生物学状态能够反应样本中细胞的生理学状态。

本文中所使用的术语“表达谱”是指 RNA、DNA 或者蛋白质的相对丰度或者活性水平。所述表达谱可以通过任意种方法和使用以下任何方法对例如转录状态或者翻译状态所获得的测量结果，所述以下方法为：基因芯片、基因阵列、珠、多元 PCR、多重 PCR、定量 PCR、连缀分析、RNA 印迹分析、蛋白质印迹分析、蛋白质表达、荧光激活细胞分类 (FACS)、酶联免疫吸附分析 (ELISA)、化学发光研究、酶分析、增殖研究或者其它任何方法、用于检测和/或分析基因表达的设备 and 系统，它们易于通过商购获得。

本文中所使用的术语样本的“转录状态”包括不同 RNA，尤其是 mRNAs 在所述样本中的同一性和相对丰度。样本的完整转录状态，即 RNA 的同一性和丰度的组合，在本文中也称为转录组。通常，测量样本中整套不同种 RNA 的所有的相关组成中的实质部分。

本文中所使用的术语“转录矢量”、“表达矢量”和“基因组矢量”（可以交换使用）是指能够反应“差异表达的基因的比例”的转录表达数据，例如，对于每一模块，所述“差异表达的基因的比例”为至少两组样本（例如，健康受试者与患者）之间的差异表达的转录子的比例。这种矢量来自两组样本的比较。分析的第一步是用于选择处于每一模块中的特定疾病组的转录子。其次存在“表达水平”。对于给定疾病的组之间的比较可以提供对于每一模块的差异表达转录子的列表。发现不同的疾病形成不同亚组的模块转录子。然后，根据表达水平可以通过被鉴定为差异表达的基因的疾病特异性亚组的表达平均值计算单个样本对每一模块的矢量。这种方法可以产生对于单个样本的

模块表达矢量的图谱，例如，在本文所公开的模块图谱中所描述的那些。这种矢量模块图谱代表对于可以来自每一个样本的每一个模块的平均表达水平（而不是差异表达基因的比例）。这种复合的“表达矢量”可以通过连续几轮以下的筛选形成：1) 对全部研究组中的显著变化的模块的筛选和 2) 对处于全部研究组中的显著变化的模块的基因的筛选。随后通过对获得的用于形成每一矢量的转录子亚组的值进行平均获得表达水平。然后，可以通过在一个图中（例如，雷达测绘图）描绘获得的对于每一个矢量的表达水平来表示患者的特征。因此，从两轮选择得到一组矢量，首先在模块水平，然后在基因水平。通过按照它们源自形成矢量的转录子的平均表达值的方式进行构建，复合成矢量表达值。

通过使用本发明，不仅可以在模块水平，并且可以在基因水平进行鉴定和区分疾病；即两种疾病可以具有相同的矢量（差异表达转录子的相同比例，相同的“极性”），但是表达矢量的基因组成可以仍然是疾病特异性的。这种疾病特异性的用户化可以使使用者通过增加其特异性优化给定组标记的性能。

使用模块作为基础建立含有最小量噪音的相关功能和转录单元的表达矢量。此外，本发明利用了复合的转录标记。本文中所使用的术语“复合的转录标记”是指多基因（模块的亚组）的平均表达值，而不是指使用单基因作为标记（这些标记的组成可以是疾病特异性的）。这种复合转录的标记的方法是独特的，因为，使用者能够利用多变量微阵列得分以评价例如，被病毒、细菌感染的患者或者患有其它传染病的患者的严重程度，或者得到本文公开的表达矢量。表达矢量是复合的（即，通过组合转录子形成）事实还能助于这些标记的稳定性。更为重要地，发现使用本发明的这种复合的模块转录标记，本文在微阵列平台上所获得的结果是可再现的，从而，对调节许可（regulatory approval）提供更强的可靠性。事实上，矢量表达值被证明具有显著的强大功能，这正如通过从微阵列平台获得的优异的再现性所示，并且在独立的一组儿科狼疮患者中也获得了验证的结果。由于微阵列数据的可靠性的提高是该技术广泛应用于临床实践的先决条件，因此，这些结果是重要的（参见，例如，FDA MAQC 程序，该程序旨在通过阵列平台创建再现性）。

用于本发明的基因表达监控系统可以包括具有有限的和/或基本数量的基因的用户化的基因阵列，所述基因对一种或多种靶疾病是特异性的或者用户化的。与习惯使用中的普通的、泛基因组阵列不同，本发明不仅能够在不

需要特定的平台的条件下使用这些普通的泛阵列进行回顾性基因和基因组分析；更为重要的是，本发明还能够发展用户化的阵列，该用户化的阵列能够提供优化的一组基因用于分析而不需要成千上万的其它不相关的基因。与现有技术相比，本发明的优化的阵列和模块的一个突出的优点是降低了财务成本（例如，每一次分析的成本、材料、设备、时间、职员、培训等），更重要的是降低了制备泛阵列的环境成本，其中，在泛阵列中，绝大多数的数据是不相关的。本发明的模块能够第一次设计简单的、用户化的阵列，该阵列能够通过最少量的探针提供具有最大信噪比的最佳的数据。由于不需要对所有基因进行分析，可以，例如，在制备能够提供大量的不相关数据的泛基因芯片过程中，不需要制备成千上万的用于照相平板技术的昂贵的铂掩光膜。如果本发明的有限量的探针组与以下技术一起使用，使用本发明可能完全不需要微阵列，所述以下技术为，例如，数字光学化学阵列、球珠阵列（ball bead arrays）、珠（例如，Luminex）、多重 PCR、定量 PCR、连缀分析、RNA 印迹分析、或甚至用于蛋白质印迹的分析、2-维和 3-维凝胶蛋白质表达、MALDI、MALDI-TOF、荧光激活细胞分类(FACS)(细胞表面或者细胞内部)、荧光激活细胞分类 (FACS)、酶联免疫吸附分析 (ELISA)、化学发光研究、酶分析、增殖研究或者其它任何方法、用于检测和/或分析基因表达的设备 and 系统，它们易于通过商购获得。

本发明的“分子指纹系统”可以有利于对以下样本中的表达进行比较分析：不同的细胞或组织、相同细胞或组织的不同亚群、处于不同生理状态的相同细胞或组织、处于不同的发展阶段的相同细胞或组织、与患有其它疾病的患者和/或正常细胞对照具有的相同组织的不同细胞群体。在一些情况下，所述正常的或者野生的表达数据可以来自在同时或者几乎同时分析的样本中，或者为从现有的基因阵列表达数据库中获得的或者精选的表达数据，例如，从公共数据库，例如 NCBI 基因表达综合数据库（NCBI Gene Expression Omnibus database）中获得或者精选的表达数据。

本文使用的术语“差异表达的”是指对存在于两种或者多于两种样本（例如，疾病样本和正常样本）中的发生变化的细胞组成（例如，核酸、蛋白质、酶活性等）的测量结果。所述细胞的成分可以存在或不存在，相对于参比上调或者下调。为了使用基因芯片或者基因阵列，可以使用核酸（例如，mRNA 或其它 RNAs（miRNA、siRNA、hnRNA、rRNA、tRNA 等））的差异基因表达区别不同的细胞类型或者核酸。更通常地，细胞的转录状态的测量可以

通过定量逆转录酶(RT) 和/或定量逆转录酶-聚合酶链式反应(RT-PCR)、基因组表达分析、翻译后分析、基因组 DNA 的修饰、易位、原位杂交等实现。

对于一些疾病阶段,尤其在疾病处于早期时,可以鉴定细胞或者形态的差异。本发明通过考虑细胞本身的基因模块,或者更重要的是免疫效应细胞中基因的细胞的 RNA 表达的模块而不需要鉴定那些特定的突变或一种或多种基因,所述免疫效应细胞在生理调节情况下起作用,即在免疫激活中、免疫耐受性或甚至免疫无能中起作用。当基因突变可能导致一组基因的表达水平发生显著的变化时,生物系统通常通过改变其它基因的表达以补偿所述的变化。这种内部的补偿应答所导致的后果是,许多扰动可能对系统的可以观察到的表型具有很小的影响,但是对细胞组成的组合具有深刻的影响。同样地,基因转录子的实际拷贝数可能没有增减,但是,所述转录子的半衰期或者寿命可能受到影响,从而导致蛋白质的产量显著增加。在一个实施方案中,本发明通过考虑效应细胞(例如,白细胞、淋巴细胞和/或它们的亚群)而不是单个信息和/或突变来消除检测真实信息的需要。

本领域的技术人员易于理解的是所述样本可以来自各种来源,包括,例如单个的细胞、细胞群体、组织、细胞培养物等。在某些情况下,甚至可以从例如,尿、血液、唾液、组织或者活组织检查样本等得到的细胞中分离足够的 RNA。在某些情况下,可以从以下样本中获得足够的细胞和/或 RNA: 黏膜分泌物、粪、眼泪、血浆、腹膜的流体、空隙的流体、硬脑膜内的流体、脑脊髓液、汗液或其它体液。来自例如组织或细胞源的核酸源可以包括活组织检查样本、一个或多个分类的细胞群体、细胞培养、细胞克隆、转化的细胞、活检(biopsies)或单个的细胞。所述组织来源可以包括,例如,脑、肝脏、心脏、肾脏、肺、脾脏、视网膜、骨、神经系统的、淋巴结、内分泌腺体、生殖器官、血液、神经、血管组织、和嗅觉上皮。

本发明包括以下基本的组成,这些组成可以单独使用或者组合使用,即:一种或多种数据挖掘运算法则;一种或多种模块-级的分析方法;血白细胞转录模块的特征;在多变量分析中使用聚集的模块数据以对人类疾病进行分子的诊断/预示;和/或检验模块-级的数据和结果。使用本发明,还可以开发和分析复合的转录标记,这些转录标记可以进一步聚集成单一的多变量得分。

本发明的发明者认识到目前基于微阵列的研究正面临重大的挑战,它们是分析众人皆知的“噪音”,所述“噪音”即为难以解释的、且在不同的实验室

和平台之间不能很好地进行比较数据。广泛公认的用于分析微阵列的数据的方法首先鉴定研究组之间差异表达的基因的亚组。其次，使用者随后设法使用模式发现运算法则和现有的科学知识了解得到的基因列表的意义。

本发明的发明者不是解决对平台的极大的可变性，而是开发了强调在分析的早期选择生物相关的基因的策略。简单地说，所述方法包括对表征给定生物系统的转录成分进行鉴定，对于该给定的生物系统，开发改进的数据挖掘预算法则以从大量收集的数据中分析和提取协同表达的基因组、或者转录模块。

本文所描述的生物标记物发现策略特别是很好地适于全面利用（on global scale）获得的微阵列数据。从约 44,000 转录子开始，定义了由约 5000 个转录子组成的含有 28 个模块的组。然后形成疾病特异性的复合表达矢量的组。矢量表达值（表达矢量）能够提供显著的强大的功能，这正如通过从微阵列平台获得的优异的再现性所示。这一发现是值得注意的，因为微阵列数据可靠性的提高是该技术广泛应用于临床实践的先决条件。最后，表达矢量可以依次结合，以获得独特的多变量得分，因此，产生的结果适合主流临床实践。令人感兴趣的是，多变量得分概括的是变化的总模式，而不是单个标记的变化。这种发展的“总的生物标记物”可以用于诊断和药物基因组学领域。

在一个实例中，从 239 个血白细胞转录谱中获得重组了 4742 个探针组的 28 个转录模块。通过文献概括证实了功能集中在形成这些模块的基因中。第二步包括研究转录系统对模块基础的干扰。为了说明这一观点，从健康的志愿者和患者中获得白细胞，并对它们进行比较和分析。通过对公布的微阵列数据集的分析，进一步验证了这种基因指纹策略。值得注意的是，使用现有的数据的本发明的模块转录仪器、系统和方法在两种商购的微阵列平台上表现出很高的再现性。

本发明包括实施设计的用于转录系统模块分析的广泛可应用的、两步微阵列数据挖掘策略。这种新的方法被用于表征血白细胞的转录信号，这种转录信号构成了临床相关信息的主要可利用来源。

如本文中所说明的那样，可以根据两种失量对两种疾病进行确定、区分和/或区别，即使所述失量对于两种疾病来说是相同的(+/+)(例如，对于 SLE 和 FLU，M1.3 = 53%下降)，这是因为每一种失量的组成仍然可以用于区分两种疾病。例如，虽然所述两种疾病对于 M1.3 来说差异表达的转录子的比

例和极性都相同，但是基因的组合仍然可以是疾病特异性的。将对基因-水平和模块-水平的分析组合可以显著地增强辨别力。此外，可以使用 2、3、4、5、10、15、20、25、28 或更多的模块来区分疾病。

术语“基因”是指包括对于产生多肽（例如，）、前体、或者 RNA（例如，mRNA）所必需的编码序列的核酸（例如，DNA）序列。所述多肽可以由全长的编码序列所编码或者由部分的编码序列所编码，只要所述全长序列或者片段保留了所需要的活性或者功能特性（例如，酶活性、配体结合、信号转导、免疫原性等）。所述术语还包括结构基因的编码区和长度为约 2 kb 的序列，该序列接近于所述编码区并分别位于 5' 和 3' 两个末端，或者更通常地位于任意一个末端，这样使所述基因对应于全长 mRNA 的长度以及影响基因的转录特性的 5' 调节序列。位于编码区的 5' 端并存在 mRNA 中的序列称为 5'-非翻译序列。所述 5'-非翻译序列通常含有所述调节序列。位于所述编码区的 3' 端或者下游并存在 mRNA 中的序列称为 3'-非翻译序列。术语“基因”包括 cDNA 和基因组形式的基因。基因的基因组形式或者克隆包括被非编码序列中断的编码区，所述非编码序列的术语为“内含子”或者“干涉区”或者“干涉序列”。内含子是能够转录到核 RNA (hnRNA) 中的基因的片段；内含子可以含有调节元件，例如增强子。内含子被除出或者“剪切出”核或者初始的转录子。因此，内含子不存在于信使 RNA (mRNA) 转录子中。mRNA 的功能是在翻译过程中使初始生成的多肽的氨基酸具有特定的序列或者顺序。

本文中所使用的术语“核酸”是指包括分子（包括但不限于 DNA、cDNA 和 RNA）的任何核酸。尤其是将在下文中发现的术语“在表 X 中的基因”是指在特定表中所列的基因的至少一部分或者全长序列。所述基因甚至可以被发现或被检测为基因组形式的，即，所述基因包括一个或多个内含子。基因组形式的基因可以含有位于编码序列的 5' 和 3' 两个末端的序列，该序列存在于 RNA 转录子中。这些序列被称为“侧翼”序列或者区。5' 侧翼区可以含有调节序列，例如，控制或者影响基因的转录的。3' 侧翼区可以含有影响转录终止、转录后切割、mRNA 稳定性和多腺苷化的序列。

本文中所使用的术语“野生型”是指从天然存在来源中分离得到的基因或者基因产物。野生型基因是通常在群体中观察到的，并且是任意设计的“正常的”或者“野生”型的基因。与此相反，术语“修饰的”或者“突变的”是指与野生型的基因或者基因产物相比，其序列和/或功能发生改变（即，改变的特征）

的基因或者基因产物。必需指出的是，可以分离到天然存在的突变体；这些突变体可以通过它们具有与野生型基因或者基因产物相比改变的特征（包括改变的核酸序列）而确定。

本文中所使用的术语“多态性”是在单个杂交群体中两个或者多于两个的等位基因有规律地并且同时出现，其中，稀少等位基因出现的频率大于通过经常性的单独的突变所能解释的频率（通常大于1%）。

本文中所使用的术语“编码的核酸分子”、“编码的DNA序列”和“编码的DNA”是指沿着脱氧核糖核苷酸的一条链的脱氧核糖核苷酸的顺序或者序列。这些脱氧核糖核苷酸的顺序确定了沿着多肽蛋白质链的氨基酸序列。因此，所述DNA序列编码氨基酸的序列。

本文中所使用的术语“互补的”或者“互补”是指通过碱基配对规律联系的多聚核苷酸（即，核苷酸序列）。例如，序列“A-G-T”与序列“T-C-A”互补。互补可以是“部分的”，在这种情况下，只有一些核酸的碱基按照碱基配对规则匹配。或者互补可以是核酸之间“完全的”或者“全部的”互补。核酸链之间互补的程度对于核酸链之间的杂交的效率和强度具有显著的影响。在扩增反应以及依赖于核酸之间的结合的检测方法中这是尤为重要的。

本文中所使用的术语“杂交”是指互补核酸的配对。杂交和杂交的强度（即，核酸之间结合的程度）受以下因素的影响：核酸之间互补的程度、涉及的条件的严格性、形成的杂交体的 T_m 和核酸中 G:C 比例。在其结构中含有配对的互补的核酸的单个分子被称为是“自体杂交”。

本文中所使用的术语“严格的”是指温度、离子强度、和存在的其它化合物（例如，有机溶剂）的条件使得核酸的杂交能够被引导地进行。在“低的严格条件下”，感兴趣的核酸序列能够杂交到它精确的互补物、具有单个碱基错配的序列、密切相关的序列（例如，序列具有90%或者更高的同源性），以及只具有部分的同源性的序列（例如，序列具有50-90%同源性）上。在“中等严格的条件下”，感兴趣的核酸序列能够杂交到它精确的互补物、具有单个碱基错配的序列和密切相关的序列（例如，序列具有90%或者更高的同源性）上。在“高的严格的条件下”，感兴趣的核酸序列只能杂交到它精确的互补物、和（取决于温度条件）具有单个碱基错配的序列上。换句话说，在高的严格条件下，可以提高温度以排除与具有单个碱基错配的序列的杂交。

本文中所使用的术语“探针”是指天然存在的纯化的限制性消化物或者合成制备的、重组或者通过PCR扩增的寡核苷酸（即，核苷酸序列），它

能够杂交到另一种感兴趣的寡核苷酸上。探针可以是单链的或者双链的。在检测、鉴定和分离特定基因序列中，探针是有用的。用于本发明任何探针可以用任何的“报告分子”标记，以便使之在任何检测系统中是可检测的，所述检测系统包括但不限于酶（例如，ELISA，以及基于酶的组织分析）的、荧光的、放射性的、发光的系统等。本文的意图不是要将本发明限制使用任何特定的检测系统或者标记。

本文中所使用的术语“靶”是指被引物结合的核酸区域。因此，“靶”是被寻找的待从其它核酸序列中挑出来的。“片段”被定义为处于靶序列中的核酸区域。

本文中所使用的术语“DNA 印迹法”是指在琼脂凝胶或者丙烯酰胺凝胶上将 DNA 按照大小分离后，再将 DNA 从所述凝胶上转移到固体支持物（例如，硝酸纤维或者尼龙膜）上的 DNA 分析。然后用标记的探针探查固定的 DNA 以检测与使用的探针互补的 DNA 种类。在电泳前，可以使用限制性酶切割 DNA。在电泳后，将 DNA 转移到所述固体支持物之前或之中，可以对 DNA 进行部分地脱嘌呤和变性。DNA 印迹法是分子生物学家标准的工具 (Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, NY, pp 9.31-9.58, 1989)。

本文中所使用的术语“RNA 杂交”是指在琼脂凝胶上将 RNA 按照大小分离后，再将 RNA 从所述凝胶上转移到固体支持物（例如，硝酸纤维或者尼龙膜）上的 RNA 分析。然后用标记的探针探查固定的 RNA 以检测与使用的探针互补的 RNA 种类。RNA 印迹法是分子生物学家标准的工具 (Sambrook, et al., *supra*, pp 7.39-7.52, 1989)。

本文中所使用的术语“蛋白质印迹法”是指固定到支持物（例如，硝酸纤维或者尼龙膜）上的蛋白质分析。蛋白质在丙烯酰胺凝胶上电泳以分离蛋白质，随后将蛋白质从所述凝胶上转移到固体支持物（例如，硝酸纤维或者尼龙膜）上。然后，将固定的蛋白质暴露于对感兴趣的抗原具有反应活性的抗体中。可以通过包括使用放射性标记的抗体在内的各种方法检测所述抗体的结合。

本文中所使用的术语“聚合酶链式反应”(“PCR”)是指 K. B. Mullis 所提出的方法(美国专利 Nos. 4,683,195、4,683,202、和 4,965,188，引入本文作为参考)，这些文献描述了在不进行克隆或者纯化的条件下增加靶序列的片段在基因组 DNA 的混合中的浓度的方法。用于扩增靶序列的方法包括向含有

所需要的靶序列的 DNA 混合物中引入大量过剩的两种寡核苷酸引物,随后在 DNA 聚合酶存在的情况下进行精确顺序的热循环。所述两种引物与双链靶序列的两条链分别互补。为了实现扩增,使混合物变性,并随后使所述引物退火到所述靶分子中的互补序列。在退火后,所述引物在聚合酶作用下延长,以形成新的一对互补链。可以重复多次变性、引物退火和聚合酶延长的步骤(即,变性、退火和延长构成一个“循环”;可以有許多次“循环”)以获得高浓度的所需靶序列的扩增片段。通过引物之间相对的位置确定所需靶序列的扩增片段的长度,因此,该长度是一个可控制的参数。由于该方法具有重复的特征,因此该方法被称为“聚合酶链式反应”(下文以“PCR”表示)。由于靶序列的所需要的扩增的片段在混合物中为优势序列(根据浓度),它们被称为“PCR 扩增”。

本文中所使用的术语“PCR 产物”、“PCR 片段”和“扩增产物”是指在完成两轮或者多于两轮的变性、退火和延长的 PCR 步骤后得到的化合物的混合物。这些术语包括一种或多种靶序列的一种或多种片段扩增的情况。

本文中所使用的术语“实时 PCR”是指各种 PCR 扩增,其中,在反应过程中而不是在反应结束后检测扩增。适合用于本发明的实时 PCR 实施方案中的试剂包括但不限于 TaqMan 探针、分子信号(molecular beacons)、蝎引物或者双链 DNA 结合染料。

本文中所使用的术语“转录上调”、“过表达”和“过表达的”是指使用 DNA 模板通过 RNA 聚合酶合成的 RNA 量出现增加。例如,当在本发明的方法中使用,术语“转录上调”是指相对于在来源于易患 SLE 的个体中的样本中检测到的感兴趣基因的 mRNA 的量比相对于在来源于不易患 SLE 的个体中的样本中检测到的感兴趣基因的 mRNA 的量增加约 1 倍、2 倍、2-3 倍、3-10 倍、甚至大于 10 倍。然而,所述系统和评价是足够特异性的,以在待检测的表达方面要求低于 2 倍的变化。此外,可以在细胞水平(在单个的细胞或者细胞群体中的表达的变化)或者甚至可以在组织水平评价所述表达的变化,其中,在所述组织中,表达所述基因的细胞数目发生变化。在分析组织的情况下,基因表达的变化可能是由于基因活性发生的调节或者细胞组成发生相对变化导致的。尤其有用的区别是那些在统计上具有显著性的区别。

与此相反,术语“转录下调”、“低表达”和“低表达的”可以交换使用,它们是指使用 DNA 模板通过 RNA 聚合酶合成的 RNA 量出现降低。例如,当在本发明的方法中使用,术语“转录下调”是指相对于在来源于易患 SLE 的

个体中的样本中检测到的感兴趣基因的 mRNA 的量比相对于在来源于不易患 SLE 的个体中的样本中检测到的感兴趣基因的 mRNA 的量，或者比相对于在来源于野生型和/或正常对照（例如，纤维肌痛（fibromyalgia））中的样本中检测到的感兴趣基因的 mRNA 的量，降低至少 1 倍、2 倍、2-3 倍、3-10 倍、甚至大于 10 倍。再次，所述系统和评价是足够特异性的，以在待检测的表达方面要求低于 2 倍的变化。此外，可以在细胞水平（在单个的细胞或者细胞群体中的表达的变化）或者甚至可以在组织水平评价所述表达的变化，其中，在所述组织中，表达所述基因的细胞数目发生变化。在分析组织的情况下，基因表达的变化可能是由于基因活性发生的调节或者在细胞组成中相对变化导致的。尤其有用的区别是那些在统计上具有显著性的区别。

还可以通过检测相对于感兴趣的基因的转录产物或者蛋白质水平而间接地监控转录的“上调/过表达”和转录的“下调/低表达”。本发明不限于与转录的上调或者下调相关的任何给定的机制。

本文中所使用的术语“真核细胞”是指具有膜被的、结构上分离的核以及其它发展完好的亚细胞器的细胞或者生物。真核细胞包括除病毒、细菌核蓝绿藻之外的所有生物。

本文中所使用的术语“体外转录”是指包括含有启动子的纯化的 DNA 模板、核糖核苷三磷酸、含有还原剂和阳离子（例如二硫苏糖醇（DTT）和镁离子）的缓冲系统、适当的 RNA 聚合酶的转录反应，该转录反应在活细胞或者生物的体外进行的。

本文中所使用的术语“扩增试剂”是指除引物、核酸模板和扩增酶之外的扩增中必需的那些试剂（脱氧核糖核苷三磷酸、环冲剂等）。通常，扩增试剂与其它反应成分一起置于和包含在反应容器中（试管、微孔板（microwell）等）。

本文中所使用的术语“诊断”是指确定病例中疾病的性质。在本发明的一些实施方案中，提供了用于产生诊断的方法，该方法能够确定传染因子或者为传染病来源的因子。在某些实施方案中，本发明的分析方法可以与共同-未决的专利 60,748,884、11,446,825 和 _____（这些共同-未决的专利的相关部分引入本文作为参考）的一种或多种模块组合使用，以确定疾病病症的性质，例如，自身免疫疾病、自身炎症疾病、癌症、移植排斥、病毒感染、细菌感染、蠕虫或寄生虫感染等。

本发明可以单独使用或者与疾病治疗的方法结合使用，以监控疾病的进程和/或患者管理。例如，可以对患者进行一次或多次检测，以确定最佳的疗程，以确定治疗是否具有预计的医料效果，以确定患者是否是用于特定治疗的候选者，或者它们的组合。本领域的技术人员能够认识到，如果疾病是急性或者慢性的，一种或多种表达失量可以指示一种或多种疾病，并且可能受其它条件的影响。

本文中所使用的术语“药物遗传学测试”是指意在研究涉及例如，药物吸收和倾向（药物代谢动力学）或者药物作用（药物动力学）的 DNA 序列中的个体之间的变化，药物遗传学测试可以包括在编码具有以下功能的蛋白质的一种或多种基因中的多态性的变化，所述具有以下功能的蛋白质为，例如转运蛋白、代谢酶、受体和其它蛋白质。

本文中所使用的术语“药物基因组测试”是指用于在整个基因组或者候选基因（例如，单核苷酸多态性 (SNP) 图谱或者单倍型标记）中研究个体之间的变化，以及可能与药理学功能和治疗应答相关的基因表达的改变或失活的分析方法。

本文中所使用的“表达谱”是指细胞的多种组成的相对丰度的测量结果。这些测量结果可以包括，例如 RNA 或蛋白质的丰度或者活性水平。所述表达谱可以是，例如转录状态或者翻译状态的测量结果。参见美国专利 Nos. 6,040,138、5,800,992、6,020,135、6,033,860，其相关部分引入本文作为参考。所述基因监控系统包括核酸探针阵列、膜印迹（例如，在杂交分析如 DNA 印迹、RNA 印迹、斑点印迹等）、或者微孔板、样本管、凝胶、珠或者纤维（或者包括结合的核酸的任何固体支持物）。参见，例如，美国专利 Nos. 5,770,722、5,874,219、5,744,305、5,677,195 和 5,445,934，这些专利的相关部分引入本文作为参考。所述基因表达监控系统还可以包括溶液中的核酸探针。

可以使用本发明的基因表达监控系统以促进下述样本中表达的比较分析：不同的细胞或组织、相同细胞或组织的不同亚群、处于不同生理状态的相同细胞或组织、处于不同的发展阶段的相同细胞或组织、相同组织的不同细胞群体。

本文中所使用的术语“差异表达”是指在两种或者多于两种的样本中细胞组成的测量结果发生了变化。所述细胞的组成可以是测试样本中的表达相对于参考发生上调，或者是测试样本中的表达相对于一个或者多个参考发生

下调。差异的基因表达还可以用于区分细胞类型或者核酸。参见美国专利 No. 5,800,992, 该专利的相关部分引入本文作为参考。

治疗或者治疗方案: 为了缓解或者改变疾病的状态, 通常采取治疗或者治疗方案。本文中所使用的治疗或者治疗方案是指为了减轻或者缓解疾病的侵袭或者症状而进行的疗程。治疗方案通常包括但不限于开出的一种或多种药物的剂量或者外科手术。理想情况下, 治疗有益于并能减轻疾病的状态, 但在许多情况下, 治疗的效果还包括不需要的效果。所述治疗的效果还受样本的生理状态的影响。

本文中所使用的术语“药理学状态”或者“药理学状况”是指用一种或多种药物、外科手术等处理的那些样本, 所述处理可能影响样本中的一种或多种核酸的药理学状态, 例如由于药理学的干预, 一种或多种核酸进行了新的转录、被稳定和/或去稳定。样本的药理学状态涉及在药物处理之前、之中和/或之后生物状态的变化, 并且可以作为本文所述的诊断或者预示的功能。在药物治疗或者外科手术后出现的一些变化可能与疾病的状态有关和/或可能与治疗的副作用不相关。药理学的状态的变化可能是由于治疗持续的时间、所开的药物的类型和剂量、患者在给定治疗时期依从的程度、和/或摄取的未规定的药物所导致的。

由于每一种病原体代表与特定的模式识别受体(PRRs)相互作用的病原体分子模式(PAMPs)的独特组合, 本发明的发明者认为如果从患有急性感染的患者的外周血中分离的白细胞具有独特的转录信号, 则这种转录信号本身又能够区别病原体。为了验证这一假想, 检测了来自患者的血中的白细胞的基因表达模式, 所述患者患有被四种常见的人类病原体导致的急性感染: (i) 流感病毒 A, 一种 RNA 病毒; (ii)金黄色葡萄球菌; 和(iii)肺链球菌, 两种革兰氏阳性菌; 和(iv) 大肠杆菌, 革兰氏阴性菌。

表 2. 141 位患有急性传染病的患者和 7 位健康的对照的特征。

患者	年龄	种族	性别	临床疾病	细菌对病毒	大肠杆菌	抗菌剂治疗
						对金黄色葡萄球菌	
第 1 组: 大肠杆菌 (n=29) 平均年龄 2 月(2 周-16 年)							
12	5 个月	黑人	男性	菌血症	训练	训练	头孢曲松钠
13	5 个月	白人	女性	UTI	训练	训练	头孢曲松钠

		西班牙					
31	3 个月	裔美国人	女性	UTI、菌血症	训练	训练	庆大霉素
34	16 岁	白人	女性	肾盂肾炎	测试 1	训练	庆大霉素
48	2 个月	白人	男性	UTI	测试 1	测试 3	氨苄青霉素、头孢曲松钠
57	3 个月	黑人	女性	UTI、菌血症	测试 1	训练	庆大霉素
		西班牙					
74	4 个月	裔美国人	女性	UTI、菌血症	训练	训练	头孢曲松钠
		西班牙					
82	2 个月	裔美国人	男性	UTI	测试 1	训练	氨苄青霉素、头孢曲松钠
		西班牙					
86	3 个月	裔美国人	男性	UTI	训练	训练	头孢曲松钠
118	1.5 个月	白人	男性	UTI	测试 1	测试 3	测试 1 & 2
		西班牙					
120	1.5 个月	裔美国人	男性	UTI	测试 1	测试 3	氨苄青霉素、头孢曲松钠
		西班牙					
133	2 个月	裔美国人	男性	UTI	测试 1	测试 3	头孢曲松钠
		西班牙					
139	1 个月	裔美国人	男性	UTI	测试 1	测试 3	氨苄青霉素、头孢曲松钠
		西班牙					
148	8 岁	裔美国人	女性	UTI	测试 1	训练	头孢曲松钠
		西班牙					
151	1.5 个月	裔美国人	男性	UTI	测试 1	测试 3	氨苄青霉素、庆大霉素
		西班牙					
152	2.5 个月	黑人	男性	菌血症、脑膜	训练	训练	头孢曲松钠、庆大霉素

炎							
154	2 个月	西班牙 裔美国 人	男性	UTI	测试 1	测试 3	头孢曲松钠
161	1.7 个月	西班牙 裔美国 人	男性	UTI	测试 1	测试 3	氨苄青霉素、头孢曲松钠
168	3 个月	白人	女性	UTI	测试 1	测试 3	头孢曲松钠
171	3 个月	西班牙 裔美国 人	女性	UTI	测试 1	测试 3	头孢曲松钠
175	0.5 个月	西班牙 裔美国 人	女性	UTI、菌血症	测试 1	测试 3	头孢曲松钠
180	1 个月	西班牙 裔美国 人	男性	UTI	测试 1	测试 3	氨苄青霉素、庆大霉素
183	1.5 个月	西班牙 裔美国 人	男性	UTI	测试 1	测试 3	氨苄青霉素、庆大霉素, 头孢曲松钠
184	0.5 个月	白人	女性	UTI、菌血症	测试 1	测试 3	氨苄青霉素、庆大霉素
188	1.5 个月	白人	男性	UTI	测试 1	测试 3	氨苄青霉素、庆大霉素, 头孢曲松钠
197	1.25 个月	白人	男性	UTI	测试 1	测试 3	氨苄青霉素、庆大霉素
219	5 个月	白人	女性	UTI、菌血症	测试 1	测试 3	头孢曲松钠
222	3 个月	西班牙 裔美国 人	女性	UTI、菌血症	测试 1	测试 3	头孢曲松钠、庆大霉素
229	4 个月	西班牙 裔美国 人	女性	UTI、菌血症	测试 1	测试 3	头孢曲松钠

第 1 组: 金黄色葡萄球菌 (n=32) 平均年龄 7 岁(3 月-18 岁)

5	10 岁	西班牙裔 美国人	男性	骨髓炎	测试 2	训练	头孢唑啉
24	3 岁	黑人	男性	骨髓炎	测试 2	测试 3	万古霉素、利福平
30	15 岁	黑人	男性	菌血症	测试 2	测试 3	万古霉素
40	12 岁	白人	男性	骨髓炎、菌血症	测试 2	测试 3	头孢唑啉
43	7 岁	黑人	男性	臀部脓肿、菌血症	测试 2	测试 3	万古霉素、利福平
62	2 岁	白人	男性	骨髓炎	测试 2	训练	克林霉素
66	3 个月	黑人	女性	肺炎	测试 2	训练	万古霉素、庆大霉素
67	7 岁	白人	女性	骨髓炎、菌血症	测试 2	训练	万古霉素、利福平
69	9 个月	西班牙裔 美国人	男性	肺部脓肿	测试 2	训练	万古霉素、头孢唑啉
70	15 个月	白人	女性	脓肿	测试 2	训练	万古霉素
84	18 岁	黑人	女性	脓肿	测试 2	测试 3	头孢唑啉
88	11 个月	西班牙裔 美国人	男性	骨髓炎、菌血症	测试 2	训练	万古霉素
89	4 个月	黑人	女性	脓肿	测试 2	训练	克林霉素
90	8 个月	黑人	男性	脓毒性关节炎	测试 2	训练	Oxacillin (苯唑青霉素)
150	9 岁	黑人	女性	骨髓炎、菌血症	测试 2	测试 3	万古霉素、利福平
179	12 岁	白人	男性	内心膜炎、菌血症	测试 2	测试 3	苯唑青霉素、庆大霉素、 利福平
205	7 岁	西班牙裔 美国人	男性	肺炎、菌血症	测试 2	测试 3	万古霉素
206	1 岁	西班牙裔 美国人	女性	脓肿	测试 2	测试 3	克林霉素
208	10 岁	白人	女性	骨髓炎、菌血症、肺炎	测试 2	测试 3	万古霉素、克林霉素、利 福平
216	10 岁	西班牙裔	女性	骨髓炎、菌血症	测试 2	训练	万古霉素、利福平

		美国人		症			
220	11 岁	西班牙裔 美国人	男性	骨髓炎、菌血症	测试 2	测试 3	头孢唑啉、利福平
221	6 岁	黑人	女性	骨髓炎、菌血症	测试 2	测试 3	万古霉素、利福平
224	10 岁	白人	男性	骨髓炎、菌血症	测试 2	测试 3	苯唑青霉素、利福平
241	10 个月	黑人	女性	肺炎、菌血症	测试 2	测试 3	万古霉素、利福平
242	13 个月	黑人	男性	脓肿、菌血症	测试 2	测试 3	克林霉素
258	8 岁	白人	女性	骨髓炎、菌血症	测试 2	测试 3	头孢唑啉
262	13 岁	西班牙裔 美国人	男性	脓肿、菌血症	测试 2	测试 3	克林霉素
264	13 岁	黑人	男性	脓毒性关节炎	测试 2	测试 3	万古霉素、头孢唑啉、庆大霉素
271	13 岁	黑人	男性	骨髓炎	测试 2	测试 3	克林霉素
281	3 岁	白人	女性	骨髓炎	测试 2	测试 3	克林霉素
315	3 岁	西班牙裔 美国人	女性	蜂窝织炎	测试 2	测试 3	万古霉素
374	21 个月	黑人	男性	脓毒性关节炎 菌血症			万古霉素、克林霉素

第 1 组: 肺炎链球菌 (n=16) 平均年龄 1.5 岁(2 月-16 岁)

9	4 个月	白人	男性	脓肿	训练	不适用	头孢唑啉
25	2 个月	西班牙裔 美国人	男性	脑膜炎	训练	不适用	氨苄青霉素、头孢曲松钠
41	23 个月	白人	女性	肺炎、 积脓症	训练	不适用	头孢曲松钠
64	10 个月	白人	女性	脑膜炎、菌血症	测试 1	不适用	头孢曲松钠、万古霉素
96	16 个月	西班牙裔	男性	肺炎、	训练	不适用	头孢曲松钠、阿奇霉素

	月	美国人		积脓症			
113	7 个月	西班牙裔 美国人	女性	脓毒性关节炎	训练	不适用	头孢曲松钠、克林霉素
155	3 个月	西班牙裔 美国人	男性	脑膜炎	训练	不适用	头孢曲松钠、万古霉素
261	13 岁	白人	男性	脑膜炎	测试 1	不适用	头孢曲松钠、万古霉素
268	3 岁	西班牙裔 美国人	男性	积脓症	测试 1	不适用	头孢曲松钠、克林霉素
265	2 岁	白人	女性	积脓症	测试 1	不适用	头孢曲松钠、万古霉素
277	16 岁	白人	男性	积脓症	测试 1	不适用	头孢曲松钠、万古霉素
287	3 岁	白人	女性	肺炎、菌血症	测试 1	不适用	头孢曲松钠、万古霉素
289	2 岁	西班牙裔 美国人	男性	积脓症	测试 1	不适用	头孢曲松钠、万古霉素
338	12 个月	白人	男性	脑膜炎			头孢曲松钠、万古霉素
339	2.5 岁	白人	男性	乳般突起炎			头孢曲松钠
388	6 个月	白人	男性	脑膜炎			头孢曲松钠、万古霉素、 利福平

第 1 组：流感病毒 A (n=18) 平均年龄 14 月 (3 周-36 岁)

55	11 个月	西班牙裔 美国人	男性	呼吸性痛苦	训练	不适用	头孢呋辛
87	19 个月	白人	女性	发烧、组织缺氧	训练	不适用	头孢呋辛
92	1 个月	西班牙裔 美国人	女性	发烧	训练	不适用	氨苄青霉素、头孢曲松钠
95	4 岁	西班牙裔 美国人	男性	发烧	测试 1 & 2	不适用	没有
101	4 个月	西班牙裔 美国人	男性	发烧、URI	训练	不适用	头孢呋辛、奥塞米韦 (Oseltamivir)
104	17 个月	西班牙裔 美国人	男性	突然发作、发烧、呼吸障碍	训练	不适用	头孢曲松钠
105	4 岁	西班牙裔	女性	发烧、脑病	测试 1 &	不适用	头孢曲松钠、阿昔洛韦

		美国人			2		(Aciclovir)、奥塞米韦
107	1.5 个月	亚洲人	男性	发烧、嗜眠症	训练	不适用	氨苄青霉素、头孢曲松钠
108	5 个月	西班牙裔 美国人	男性	发烧	训练	不适用	头孢曲松钠
112	1 个月	西班牙裔 美国人	男性	发烧、URI	测试 1 & 2	不适用	氨苄青霉素、庆大霉素
114	18 个月	黑人	女性	呼吸性痛苦、发烧	训练	不适用	头孢呋辛、奥塞米韦
115	20 个月	白人	男性	突然发作	训练	不适用	阿莫西林
116	2 岁	白人	男性	发烧、URI	测试 1 & 2	不适用	头孢呋辛、克林霉素
117	24 岁	白人	女性	发烧	测试 1 & 2	不适用	没有
128	11 个月	西班牙裔 美国人	女性	发烧、组织缺氧	训练	不适用	头孢呋辛
132	6 个月	白人	男性	呼吸性痛苦、发烧	训练	不适用	苯唑青霉素、妥布霉素
259	3 个月	西班牙裔 美国人	女性	肺炎	测试 1 & 2	不适用	没有
266	36 岁	白人	女性	发烧、咳嗽	测试 1 & 2	不适用	没有
患者	年龄	种族	性别	临床疾病	分析	平台	抗菌剂治疗
第 2 组: 流感病毒(n= 18) 平均年龄= 11 月 (2 周-13 岁)							
311	0.1 岁	西班牙裔 美国人	男性	流感病毒 B 发烧、URI	图 6c	Illumina Sentrix Hu6	氨苄青霉素 + 头孢曲松钠
320	0.04 岁	西班牙裔 美国人	女性	流感病毒 B 发烧、URI	图 6c	Illumina Sentrix Hu6	氨苄青霉素+ 庆大霉素
517	0.5 岁	西班牙裔 美国人	女性	流感病毒 A 肺炎	图 6a 图 6b	Affymetrix U133plus2	没有

519	0.13 岁	西班牙裔 美国人	女性	流感病毒 A 发烧	图 6c	Illumina Sentrix Hu6	没有
524	6 岁	西班牙裔 美国人	男性	流感病毒 A 发烧	图 6a	Affymetrix U133plus2	没有
527	0.13 岁	黑人	男性	流感病毒 A 发烧	图 6c	Illumina Sentrix Hu6	氨苄青霉素 + 头孢曲 松钠
530	0.38 岁	西班牙裔 美国人	男性	流感病毒 A 发烧、突然发 作	图 6c	Illumina Sentrix Hu6	没有
532	0.08 岁	西班牙裔 美国人	女性	流感病毒 A 发烧、咳嗽	图 6a 图 6b	Affymetrix U133plus2	氨苄青霉素 + 庆大霉 素
533	11 岁	高加索人	男性	流感病毒 B 发烧、咳嗽	图 6a 图 6b	Affymetrix U133plus2	没有
536	2 岁	西班牙裔 美国人	女性	流感病毒 A 发烧、咳嗽	图 6a 图 6b	Affymetrix U133plus2	没有
540	0.08 岁	西班牙裔 美国人	男性	流感病毒 A 发烧、咳嗽	图 6a 图 6b	Affymetrix U133plus2	氨苄青霉素+ 庆大霉素
542	0.04 岁	西班牙裔 美国人	女性	流感病毒 A 发烧	图 6c	Illumina Sentrix Hu6	氨苄青霉素+ 庆大霉素
547	1.33 岁	黑人	女性	流感病毒 A 脑炎	图 6a	Affymetrix U133plus2	头孢曲松钠 + 奥塞米 韦
549	13 岁	西班牙裔 美国人	女性	流感病毒 B 发烧、昏厥	图 6a	Affymetrix U133plus2	头孢曲松钠 + 万古霉 素 + 奥塞米韦
553	1.5 岁	高加索人	女性	流感病毒 A 发烧、URI	图 6a 图 6b	Affymetrix U133plus2	奥塞米韦
556	3.5 岁	高加索人	女性	流感病毒 A 发烧、突然发	图 6c	Illumina Sentrix Hu6	头孢曲松钠 + 奥塞米 韦

作							
560	10 岁	黑人	女性	流感病毒 B 脑炎	图 6c	Illumina Sentrix Hu6	阿昔洛韦
567	2 岁	西班牙裔 美国人	女性	流感病毒 B 发烧、URI	图 6a 图 6b	Affymetrix U133plus2	没有
第 2 组: 金黄色葡萄球菌 (n=19) 平均年龄=7.5 岁 (0.08-14 岁)							
305	4.5 岁	西班牙裔 美国人	女性	MSSA 菌血症、化脓 性关节炎、骨 髓炎	图 6a	Affymetrix U133plus2	头孢唑啉
308	12 岁	黑人	女性	MSSA 散布性肺炎	图 6a 图 6b	Affymetrix U133plus2	苯唑青霉素 + 克林霉 素
369	14 岁	黑人	男性	MRSA 散布的	图 6a	Affymetrix U133plus2	万古霉素、利福平
372	14 岁	高加索人	男性	MRSA 菌血症、骨髓 炎	图 6a	Affymetrix U133plus2	万古霉素、利福平
374	1.75	黑人	男性	MRSA 菌血症、化脓 性关节炎	图 6a	Affymetrix U133plus2	万古霉素
380	7.5 岁	黑人	男性	MRSA 骨髓炎、化脓 性关节炎	图 6a	Affymetrix U133plus2	克林霉素
458	12 岁	黑人	男性	MRSA 散布的	图 6c	Illumina Sentrix Hu6	万古霉素 + 利福平 + 利奈唑胺
459	10 岁	高加索人	女性	MSSA 骨髓炎、化脓 性关节炎	图 6c	Illumina Sentrix Hu6	苯唑青霉素 + 利福平
465	13 岁	高加索人	男性	MRSA 骨髓炎、化脓 性关节炎、菌	图 6c	Illumina Sentrix Hu6	万古霉素

				血症			
466	0.5 岁	黑人	男性	MRSA SST 脓肿	图 6c	Illumina Sentrix Hu6	克林霉素
472	0.08 岁	高加索人	男性	MSSA SST 脓肿	图 6c	Illumina Sentrix Hu6	头孢唑啉
475	1.33 岁	黑人	男性	MSSA 化脓性关节炎	图 6c	Illumina Sentrix Hu6	萘夫西林 (Nafcillin)
477	6 岁	黑人	男性	MRSA 菌血症、化脓性 性关节炎	图 6c	Illumina Sentrix Hu6	克林霉素 + 利福平
480	12 岁	高加索人	男性	MSSA 菌血症	图 6c	Illumina Sentrix Hu6	克林霉素 + 强力霉素
489	1.08 岁	高加索人	男性	MRSA SST 脓肿	图 6c	Illumina Sentrix Hu6	克林霉素
522	9.5 岁	黑人	女性	MRSA 菌血症、骨髓炎	图 6c	Illumina Sentrix Hu6	万古霉素 + 利福平
529	1.75	黑人	男性	MRSA 菌血症、肺炎	图 6c	Illumina Sentrix Hu6	万古霉素 + 利福平
535	0.58 岁	其它	女性	MSSA 化脓性关节炎	图 6c	Illumina Sentrix Hu6	头孢唑啉
537	9 岁	黑人	女性	MSSA 菌血症、骨髓炎、化脓性关 节炎	图 6c	Illumina Sentrix Hu6	苯唑青霉素

第 2 组:肺炎链球菌 (n=9) 平均年龄=2.5 岁 (1.3-16 岁)

					图 6a	Affymetrix	
96	1.33 岁	西班牙裔 美国人	男性	肺炎、积脓症	图 6b	U133plus2	头孢曲松钠 + 阿奇霉素
265	2.2 岁	高加索人	女性	肺炎、 积脓症	图 6a 图 6b	Affymetrix U133plus2	头孢曲松钠 + 万古霉 素

268	3 岁	西班牙裔 美国人	男性	肺炎、积脓症	图 6a 图 6b	Affymetrix U133plus2	头孢曲松钠
277	16 岁	高加索人	男性	肺炎、 积脓症	图 6a 图 6b	Affymetrix U133plus2	头孢曲松钠 + 克林霉素
287	3.2 岁	高加索人	女性	肺炎、菌血症	图 6a 图 6b	Affymetrix U133plus2	头孢曲松钠
289	2.5 岁	西班牙裔 美国人	男性	肺炎、积脓症	图 6a 图 6b	Affymetrix U133plus2	头孢曲松钠
471	2 岁	高加索人	女性	菌血症、脑膜 炎	图 6c	Illumina Sentrix Hu6	万古霉素+ 头孢曲松钠
473	2.5 岁	西班牙裔 美国人	男性	菌血症、 肺炎	图 6c	Illumina Sentrix Hu6	头孢曲松钠
523	3 岁	西班牙裔 美国人	男性	化脓性关节炎	图 6c	Illumina Sentrix Hu6	头孢唑啉

表 3. SLE 患者统计

SLE 患者(n=11) 13 岁 (9-17 岁)				
SLE 87	11 岁	白人	女性	不适用
SLE 85	16 岁	黑人	女性	不适用
SLE 79	10 岁	西班牙裔美 国人	女性	不适用
SLE 76	15 岁	黑人	女性	不适用
SLE 66	17 岁	西班牙裔美 国人	女性	不适用
SLE 57	12 岁	黑人	男性	不适用
SLE 48	14 岁	白人	女性	不适用
SLE 45	9 岁	黑人	女性	不适用
SLE 107	14 岁	黑人	女性	不适用
SLE 27	13 岁	黑人	女性	不适用
SLE 19	9 岁	黑人	男性	不适用

健康对照 (n=17) 4.5 岁 (4 月-17 岁)

INF 20N	11 个月	西班牙裔美 国人	男性	健康的
INF 19N	4 个月	西班牙裔美 国人	男性	健康的
INF 27N	10 个月	白人	男性	健康的
INF 25N	11 个月	西班牙裔美 国人	女性	健康的
INF 204	2 岁	白人	男性	健康的
INF 7N	19 个月	黑人	女性	健康的
INF 391	18 个月	白人	女性	健康的
INF 392	10 个月	黑人	男性	健康的
H 162	15 岁	黑人	女性	健康的
HJM	13 岁	西班牙裔美 国人	女性	健康的
HJC	8 岁	西班牙裔美 国人	男性	健康的
HBW	14 岁	白人	女性	健康的
H45	12 岁	黑人	女性	健康的
H42	9 岁	西班牙裔美 国人	男性	健康的
H36	7 岁	白人	女性	健康的
H28	11 岁	西班牙裔美 国人	女性	健康的
H37	8 岁	白人	女性	健康的

表 4. 35 个能够区分流感病毒 A 感染和细菌感染的分类基因的列表。如下所示, 根据存在和重要性程度对基因进行功能性分组。全部的详细资料可以从补充表 3 获得。

流感病毒>细菌		细菌>流感病毒	
对病毒的应答		翻译延长	
cig5	1.46E-05	EEF1G	4.52E-06
DNAPTP6	4.52E-06	EEF1G	2.35E-06
IFI27	4.52E-06	对翻译起始的调节	
IFI35	0.00033	EIF3S5	9.34E-08
IFI44	0.00023	EIF3S7	2.35E-07
IFI44	0.00015	EIF4B	1.16E-06
OAS1	6.52E-05	蛋白质合成	
免疫应答		QARS	5.41E-07
BST2	4.08E-05	RPL31	4.52E-06
G1P2	0.000101	RPL4	2.35E-07
LY6E	8.28E-06	对转录的调节	
MX1	6.52E-05	PFDN5	5.41E-07
抗凋亡		细胞黏附	
SON	0.00067	CD44	2.35E-07
细胞生长和/或维持		代谢	
TRIM14	4.08E-05	HADHA	4.08E-05
多种功能		PCBP2	9.34E-08
APOBEC3C	2.35E-07	多种功能	
C1orf29	0.00015	dJ507I15.1	6.52E-05
FLJ20035	4.08E-05		
FLJ38348	0.00128		
HSXIAPAF1	4.52E-06		
KIAA0152	2.48E-05		
PHACTR2	9.34E-08		

USP18	1.46E-05
ZBP1	5.41E-07

表 5. 30 个能够区别金黄色葡萄球菌和大肠杆菌感染的分类基因的列表。如下所示，根据存在和重要性程度对基因进行功能性分组。全部的资料可以从补充表 6 获得。

金黄色葡萄球菌 > 大肠杆菌		大肠杆菌 > 金黄色葡萄球菌	
信号传导		细胞内信号传导	
CXCL1	0.00106	RASA1	1.20E-05
JAG1	0.00158	SNX4	4.92E-05
RGS2	0.00027	对翻译起始的调节	
代谢		AF1Q	0.00106
GAPD	0.00044	对转录的调节	
PPIB	0.00044	SMAD2	0.00044
PSMA7	0.00106	细胞粘附	
MMP9	0.00837	JUP	0.00158
p44S10	0.00158	代谢	
蛋白质靶向		PP	4.92E-05
TRAM2	0.00384	MAN1C1	0.00016
细胞内蛋白质运输		多种功能	
SEC24C	4.92E-05	FLJ10287	4.92E-05
多种功能		FLJ20152	0.00622
ACTG1	0.00622	LRRN3	1.20E-05
CGI-96	0.00454	LRRN3	0.00027
MGC2963	0.00158	SGPP1	0.00158
STAU	4.92E-05	UBAP2L	2.12E-06
STAU	4.92E-05		

患者特征。PBMCs 来自 29 例被大肠杆菌感染的患者、51 例被金黄色葡萄球菌感染的患者、25 例被肺炎链球菌感染的患者和 36 例被流感病毒 A 感染的患者。由于年轻的患者的伴发病和治疗比年纪较大的成年人的伴发病和治疗少，因此，选择年轻的患者。排除患有潜在的免疫抑制的、接受免疫调节的治疗的（包括皮质类固醇治疗）、或具有明显的慢性医学问题的患者。在取血时，住院的平均时间（范围）为 3 天(0-9 天)，症状持续的平均时间（范围）为 6 天(2-22 天)。临床诊断包括急性呼吸性感染、菌血症、局部脓肿、骨和关节感染、尿道感染和脑膜炎 (表 1)。按照标准的医院治疗方案治疗患者，并且，在急诊室立刻开始抗菌素的治疗。

逐步数据分析策略。为了确定从患有急性感染的患者中分离得到的血液中的白细胞是否具有能够区分病原体类型的基因表达信号，进行了以下的逐步数据分析：(1)对组进行统计的比较：采用非参数曼恩-惠特尼验证 (Mann-Whitney test)、以成对比较的方式鉴定差异表达的基因。按照它们的表达水平排列的等级聚类的基因揭示了两组之间表达的倒数式的模式。(2)样本分类：通过比较患者组鉴定能够区分两组患者的基因，即分类基因，所述患者的年龄范围是可以比较的，并且采用相同的抗菌剂进行治疗（训练组）。然后，在同一组患者中采用弃一交差验证方案对这些基因进行评价。(3)对分类基因进行独立的验证：检测相同基因对独立组的患者（测试组）分类的能力。仔细筛选包括在所述训练组中使用的以鉴定分类基因的患者，以避免潜在的混淆因素。在仔细筛选后，在新的一组不同种类的患者中评估所述分类基因（也描述为转录标记），因而更能代表真实的临床情况（测试组）。(4)通过微阵列平台进行独立的验证：然后使用不同的微阵列平台(Illumina BeadChips)在另一组独立的患者中对结果进行进一步验证。

转录信号能够区分被流感病毒 A 感染的患者和被细菌感染的患者。为了确定来自被流感病毒感染的患者或者被细菌感染的患者的样本之间的差异表达的基因，以相似的年龄和相似的抗菌素治疗为基准，选择 11 例被流感病毒 A 感染的患者和 12 例被大肠杆菌感染或者被肺炎链球菌感染的患者作为训练组。被流感病毒 A 感染的训练组和被细菌感染的训练组在平均年龄（年龄范围）上没有明显的区别(分别为 11 个月[1-20 个月]和 4 个月[2 个月-23 个月]； $P=0.22$ ；)，或者在收集样本前住院的天数（范围）没有明显区别(分别为 2 天[1-2 天] 和 2.5 天[2-5 天]， $P=0.06$)。所有被流感病毒 A 感染的 11 例患者接受内酰胺抗生素的治疗，12 例被细菌感染的患者组中 10 例也是接

受内酰胺抗生素的治疗($P=0.16$)。在两组 PBMCs 中,嗜中性粒细胞、淋巴细胞和单核细胞的相对比例没有统计上的显著差异(补充表 1)。

通过对被流感病毒 A 感染的患者组和被细菌感染的患者组进行统计比较,获得 854 个差异表达的基因($P<0.01$) (补充表 2), 其中, 394 个基因在被流感病毒 A 感染的患者中相对地过表达, 而 460 个基因在被细菌感染的患者中过表达。被流感病毒 A 感染的患者表现出明显的 I 型干扰素 (IFN) 的信号 (图 1a), 包括编码以下物质的基因: 抗病毒 (例如, 抗黏病毒) 的分子的基因 (MX1、MX2)、2'-5'-寡腺苷酸合成酶 (OAS1、OAS2)、GBP1 (鸟苷酸结合蛋白 1)、和 CIG5 (蝰蛇毒素 (viperin)、病毒抑制蛋白、内质网相关的干扰素诱导的)。调节转录和翻译的基因表示了高达 25% 的 460 个探针组, 这些基因在细菌感染组中以更高水平表达的。

k-NN 运算法则鉴定 35 个能够区分被流感病毒急性感染的患者和被细菌急性感染的患者的基因 (图 2、表 2、和补充表 3)。对训练组进行的弃一交叉验证正确地将 23 个样本的 21 个样本 (91% 准确性) 分为流感病毒 A 感染的患者组或者被细菌感染的患者的组 (图 1b)。

然后, 在独立的样本组 (测试组) 中验证被鉴定的能够区分流感病毒 A 感染和细菌感染的区分能力。第一测试组的患者包括 7 例新的被流感病毒 A 感染的患者, 和 30 例被细菌感染的患者 (7 例被肺炎链球菌感染和 23 例大肠杆菌感染)。在测试组中选择患者时, 不考虑患者的年龄或者采用抗生素治疗所用的抗生素的类型 (年龄 [范围]; 流感病毒 A, 4 岁 [3 周-36 岁]; 大肠杆菌, 2 个月 [2 月-16 岁])。预示性的基因将 37 个样本中的 35 个样本正确地进行了分类 (95% 的准确性) (图 1c)。一个样本 (INF48) 被错误地分类和一个样本的分类不明确 (INF120)。

然后在第 2 测试组中评价了 35 个分类基因, 第 2 测试组包括 7 例被流感病毒 A 感染的患者和 31 例被金黄色葡萄球菌感染的患者, 区分的准确率为 87% (图 1d)。不考虑患者的年龄或者采用抗生素治疗时所使用的抗生素的类型, 重新选择测试组 (年龄 [范围]; 流感病毒 A, 4 月 [3 周-36 岁]; 金黄色葡萄球菌, 7 岁 [3 个月-15 岁])。5 例金黄色葡萄球菌样本被错误分类 (INF62、INF70、INF89、INF221 和 INF242)。

约三分之一被细菌感染的患者表现出干扰素相关的基因的表达水平升高。然而, 这一信号对于分类的结果影响有限, 因为从被细菌感染的患者中获得的样本的表达信号特征与流感病毒感染 (与细菌感染相比, 在流感病毒

感染中为低表达的基因)的表达信号特征不互为倒数,并且还部分由于在细菌感染的情况下干扰素可诱导的基因的表达水平较低(图 1c)。干扰素可诱导的基因的表达水平的升高可能归因于对备有证明文件的细菌的感染的本身的应答[12],或者归因于对未确诊的或者前述的病毒的感染的应答。因此,宿主对流感病毒的感染和细菌的感染应答的转录信号能够被鉴定。这些信号能够区分这些致病因子。

转录信号能够区分被大肠杆菌感染的患者和被金黄色葡萄球菌感染的患者。为了鉴定在被大肠杆菌感染的患者中和在被金黄色葡萄球菌感染的患者中差异表达的基因,选择每组 10 个患者为训练组。在被大肠杆菌感染的组和被金黄色葡萄球菌感染的组中,平均年龄[范围](分别为 2 个月[3.5 个月-16 个月] 和 12 个月[4 个月-10 岁]; $P=0.06$)没有显著的区别。每一组包括 6 例用 β -内酰胺抗生素治疗的患者和 4 例用其它抗生素治疗的患者。外周白细胞计数和外周中血细胞类型的相对比例在两组中没有显著的差异(补充表 1)。在收集样本前,对于被大肠杆菌感染的组,患者住院的平均天数为 2 天;而对于被金黄色葡萄球菌感染的组,患者住院的平均天数为 4 天($P=0.01$),出现显著差异的原因可能是由于为了获得确定的微生物诊断结果所需要的时间不同。

通过对组之间进行统计比较,产生 211 个表达水平差异显著的($p<0.01$)基因(补充表 4 和图 3a)。通过实时 PCR 独立确定所选择的基因的表达水平(图 3d 和 3e)。与被大肠杆菌感染的患者相比,在被金黄色葡萄球菌感染的患者中存在的大量的过表达基因与嗜中性白细胞的活性有关,这些过表达的基因包括编码化学引诱物分子(例如, CXCL1 (CXC 趋化因子配位体 1、GRO-1)) 和 PPIB (亲环素 B)的基因[13、14]。此外,基质金属蛋白 9 (MMP9)在嗜中性外渗和迁移中起了很重要的作用[15]; PRG1 (分泌的颗粒蛋白聚糖 1) 参与了颗粒蛋白在人体中的嗜中性白细胞中的包装[16]; 和 ALOX5AP 激活了花生四烯酸酯 5-脂氧化酶 (arachidonate 5-lipoxygenase) 和延长嗜中性白细胞合成白三烯的能力[17]。最后,最近,嗜中性白细胞被鉴定为在金黄色葡萄球菌感染模型中的 S100A8 和 S100A9 (钙粒蛋白 A 和 B, 别名为 MRP 8 和 14)的主要来源 [18]。这些结果可以部分地解释从被大肠杆菌感染的患者和被金黄色葡萄球菌感染的患者中获得的样本中基因表达水平的差异。以前的对患有系统红斑狼疮(SLE)的患者的研究表明,类似的信号可以追溯到存在低密度的未成熟的嗜中性白细胞,在密度梯度离心中,这些未成熟的嗜中性

白细胞与单核细胞共同纯化[9]。令人感兴趣的是,从患有 SLE (观察结果没有出版)的患者的全血中得到的表达谱中也可以观察到这种相应于更快速的嗜中性白细胞更新速度的“粒细胞生成信号”。

鉴定了 30 种能够区分被大肠杆菌感染的患者和被金黄色葡萄球菌感染的患者的训练组的分类基因(图 4 和表 3 以及补充表 6)。通过弃一交差验证,20 个样本中有 19 个样本被正确地分类(95%的准确率) (还可以参见图 3b)。一例被金黄色葡萄球菌感染的患者(INF 89)被错误分类。用独立组的被金黄色葡萄球菌感染的患者(n=21)和被大肠杆菌感染的患者(n=19)对所述分类基因进行验证,独立组的患者是重新选择的,并且在选择时不考虑患者的年龄或者采用抗生素治疗时所使用的抗生素类型(被金黄色葡萄球菌感染的患者: 9 岁[10 个月-18 岁];被大肠杆菌感染的患者: 2 个月 [2 周-5 个月])。这 30 个基因对 40 个样本中的 34 个样本进行了正确地分类(85%的准确率;图 3c)。2 个样本(INF175 和 INF206) 被错误地分类,并且 4 个样本的分类不明确(INF168、INF220、INF281 和 INF315)。尽管没有特定模式的错误分类是显然的,但是被金黄色葡萄球菌感染的患者所表现出的临床疾病和严重程度的更大差异性可能对该组患者较低预测准确性产生影响。

因此,这些结果表明血液中的白细胞的转录信号能够区分被金黄色葡萄球菌急性感染的患者或者被大肠杆菌急性感染的患者的发病病因。此外,在具有辨别能力的信号中鉴定出了显著的功能集中的信号:在被流感病毒 A 感染的患者的过表达的基因中发现了干扰素可诱导的基因;而与被大肠杆菌感染的患者组相比,在被金黄色葡萄球菌感染的患者组中,与嗜中性白细胞相关的基因以较高的水平表达。

能够区分来自被流感病毒 A、大肠杆菌、金黄色葡萄球菌或者肺炎链球菌急性感染的患者的样本的基因重叠程度最低。本发明的发明者已经定义了能够区分被流感病毒 A 感染和被细菌感染的患者的分类基因的组,和能够区分被大肠杆菌感染和被金黄色葡萄球菌感染的患者的分类基因的组。为了完善分类基因组,对能够区分被肺炎链球菌感染的患者的基因的组进行了额外的成对比较和鉴定。通过对被大肠杆菌 (n=11)感染的组和被肺炎链球菌 (n=11)感染的组进行比较,产生 264 个差异表达水平显著($P<0.01$)的基因,和 45 个分类基因(图 4b 和 4c, 以及补充表 7 和 8);通过对训练组进行弃一交差验证,对 22 个样本中的 20 个样本进行了准确地分类(91% 的准确率)。通过对被金黄色葡萄球菌(n=12)感染的组和被肺炎链球菌(n=11)感染的组进行

比较, 产生 127 个差异表达的基因($P < 0.01$)和 34 个分类基因 (图 4d 和 4e, 以及补充表 9 & 10)。通过对训练组进行弃一交差验证, 对 23 个样本中的 19 个样本进行了准确地分类(83%的准确率)。

对获得的用于每一个成对分析的分类基因的组进行系统的比较, 发现这些分类基因绝大多数都是相互排斥的(图 5a)。此外, 102 个能够区分不同细菌感染的基因中, 没有一个基因是区分被流感病毒 A 感染和被细菌感染所必需的(图 5b)。因此, 使用独立组的转录信号可以区分多重感染致病病因。

由不同的病原体导致的患有急性呼吸性感染的患者具有截然不同的表达模式。在一群具有相同临床表现的患者中检测了基因表达模式。将在整个研究(图 5a 和 5b)中鉴定的分类基因的组混合, 然后用于产生患有较下的呼吸道感染患者亚组的表达模式(表 1 中所列的 27 个样本)。从健康的志愿者中收集的 7 个样本用作对照(表 1)。基因和样本的等级聚类鉴定了四种原型表达信号: 根据 PBMC 表达谱, 健康的对照能够明显区别于所有传染病组。这一发现本身就是值得注意的, 因为, 用于产生分类基因的训练组中没有一个包括来自健康的志愿者的样本。第二信号与来自被流感病毒 A 感染的患者的样本相关(包括干扰素可诱导的基因), 并且与第三信号显著不同, 第三信号的特征在于被金黄色葡萄球菌和肺炎链球菌感染(包括与嗜中性白细胞相关的基因)。由能够区分三大类样本的明显优势的信号可以将这两种革兰氏阳性细菌的区别降低到最小。

令人感兴趣的是, 有四种样本属于被流感病毒 A 感染的组和一种属于被金黄色葡萄球菌感染的组可以通过第四信号表征, 第四信号组合了前述信号的成分(干扰素可诱导的基因和与嗜中性白细胞相关的基因: 图 5c, 由星号表示)。这一发现暗示了至少两种可能情况中的一种: 1)混合的信号是由于共同感染导致的, 这种共同感染不能通过常规的诊断方法检测出来; 或者 2)对 PBMC 转录信号的分析可以揭示存在的不同的患者亚组。需要一大群患者以研究这些可能的情况并鉴定潜在的临床含义。进一步研究了 5 位具有混合信号的患者病历卡, 鉴定了 3 位被流感病毒 (#101、#128 和 #132)感染的患者, 这三位患者的放射性检查的结果为肺炎, 并且白细胞计数不同, 分别属于 11%、16% 和 28% 的组。这些证据表明, 这 3 例患者可能存在共同感染。这些证据清楚地表明了表现相似的症状的患者中可以获得有差别的血液中的白细胞转录模式。

在新的独立组的样本中和微阵列平台中可以再现结果。研究设计包括用于鉴定分类基因的训练组(图 1b: 流感病毒对细菌; n=23 样本)和用于独立地验证这些发现的测试组(图 1c : 流感病毒对细菌; n=37 样本; 和图 1d : 另外 31 位被金黄色葡萄球菌感染的患者)。获得自总计 91 位患者的这些数据是使用 Affymetrix U133A 和 U133B GeneChips 产生的。进一步对数据的有效性进行了验证以进一步确定这些发现, 并且在另一组患者中采用不同的微阵列平台进行了类似的分析。征集了由 22 位被流感病毒/细菌急性感染的患者组成的新的一群, 并用最新的 Affymetrix GeneChips 版本 (U133 plus 2.0) 分析了 PBMC 样本。

图 9a-9c 概括了用微阵列平台进行的独立的确证和验证。图 9a 表示使用 Affymetrix U133 和 2.0 GeneChips 对来自新的一组被流感病毒 (n=10)或被细菌(金黄色葡萄球菌: n= 6; 肺炎链球菌: n=6) 急性感染的患者分析的结果。使用能够区分被流感病毒 A 感染和细菌感染的分类基因(35 基因, 维恩图 (Venn diagram), 右; 图 1 和补充表 3)对所述新的一组样本进行聚类。在图 9b 中, 使用来自图 5 中的 137 转录子的列表从包含在图 9a 中的患有急性呼吸性感染的患者中聚类出 14 个样本的亚组。图 9c 表示使用 Illumina Sentrix Hu6 whole genome BeadChips 对另一独立组的样本(这些样本中没有一个被用于上述任意一个分析中)进行分析所获得的结果, 所述样本从新的一组被流感病毒(n=8)或者被细菌(金黄色葡萄球菌: n=13; 肺炎链球菌: n=3)急性感染的患者中获得的。使用能够区分被流感病毒 A 感染和被细菌感染的分类基因(35 基因, 维恩图 (Venn diagram), 右; 图 1 补充表 3)对所述新的一组样本进行聚类。用颜色标准指示改变的表达水平, 红色表示与每个基因在所有的供体中的平均表达值相比其表达水平相对较高, 蓝色表示与每个基因在所有的供体中的平均表达值相比其表达水平相对较低。

本发明几乎能够完美地区分由金黄色葡萄球菌或者肺炎链球菌导致的感染与由流感病毒导致的感染 (图 9a: 一个流感病毒样本被归类到细菌感染组中), 并且能够在患有急性呼吸性感染的患者中获得具有区别性能的信号 (图 9b)。众所周知, 微阵列数据很难对完全不同的平台进行比较[19、24-26], 然而, 本发明使用 Illumina's whole genome Sentrix Hu6 BeadChips 对新的一组 24 个样本进行分析时, 能够再一次再现初始结果 (图 9c: 一个来自细菌感染组的样本被聚类到流感样本中)。在这一群体中, 只有两位属于被金黄色葡萄球菌或肺炎链球菌感染的组的患者表现急性呼吸性感染。

同样地,进行了 148 个微阵列分析,包括 141 个微阵列分析是针对于从患有急性感染的患者中收集的样本。除了通过实时 PCR 进行确证外(图 3d),还通过微阵列平台对独立数据的有效性进行验证,结果证实了这些发现的可靠性(robustness)。

不同的转录信号能够区分被急性感染的患者和患有自身免疫疾病的患者。在被流感病毒急性感染的患者中发现干扰素可诱导基因为过表达的。以前在系统红斑狼疮的患者的血液中的白细胞中也鉴定出了干扰素信号[9]。其次,检测了在血液中的白细胞的表达模式是否还能区分被流感病毒感染和 SLE。将来自 SLE 患者的样本与他们各自健康的对照组进行比较。类似地,将来自各种传染病组的患者与适当的健康志愿者组进行比较(每组有 11 位患者:流感病毒 A、大肠杆菌、金黄色葡萄球菌、肺炎链球菌;对照组为 9 位健康的志愿者)。对通过每一次比较获得的 P 值(总共 5 组患者和它们各自的对照组)进行编辑,并进行共同地分析。这一方法概括了从许多研究和大量的样本中观察到的变化,该方法尤其能够很好地适合于所有的潜在的混淆因素不能解释的情况(例如,在女性中, SLE 发病率明显较高)。对显著性模式进行了分析,以评价从被流感病毒感染的组和 SLE 组中获得的基因表达信号之间的重叠(图 6)。使用过滤标准以在两组患者组中筛选出与所述两组患者组的各自的对照组相比过表达或者低表达的转录子(图 6, 上部)。发现在过表达的转录子中,包括干扰素可诱导的基因的一类(图 6: 上部-IFN; 补充表 11)在被流感病毒感染组和 SLE 组中均发生显著的变化,但在被细菌感染的患者中没有发生显著的变化。相反,还可能鉴定出在一组中相对于健康的对照发生显著变化(FLU 或 SLE, $p < 0.01$)而在另一组中没有发生显著变化($p > 0.5$; 图 6, 中部和下部)的基因。该方法揭示了疾病特异性的信号(数据未示出)。在图 6 和补充表 1-11 中可以发现几类能够独特地表征被流感病毒 A 感染的患者的信号。这些结果进一步表明血液中的白细胞转录谱的混乱是疾病特异性的。

对宿主-病原体微阵列数据集(包括 32 次研究)进行的概略性比较分析鉴定了普通宿主对感染的转录应答和病原体特异的信号[27]。存在广泛的相似性,例如,在真菌感染、细菌感染或者病毒感染中观察到参与免疫细胞激活和征募中细胞因子和趋化因子的动态的级联 30-34。然而,有两个因素有助于对感染产生专一性转录应答: 1) 参与病原体识别的分子机制的多样性; 和 2) 宿主应答被病原体改变。一旦激活, Toll 样受体(TLR)家族成员引发信

号传导途径,在该途径中,共用相同的成分,但是保留了导致专一性转录应答的独特特征。因此,观察到在对革兰氏阳性和革兰氏阴性的细菌的应答的定性和定量的区别分别通过 TLR2 和 TLR4 识别。此外,还发现对暴露于流感病毒(通过 TLR3)、大肠杆菌(通过 TLR4)和念珠菌属(通过 TLR2/TLR4)的树枝状细胞检测的应答显著不同。由病原体导致的宿主应答程序的改变还明显有助于对感染进行多样性转录应答。如微阵列检测所示,分枝杆菌的产物为,例如,在巨噬细胞中能够抑制干扰素 γ 诱导的基因调节[28]。类似地,微阵列研究还表明疱疹病毒、假狂犬病毒、丙型肝炎病毒、水痘-带状疱疹病毒或者鼻病毒能够限制宿主通过各种机制进行有效的抗病毒应答的能力。过去几年积累的大量的实验数据暗示宿主能够对感染发起病原体专一性的转录应答。

大量的研究表明当在体外将免疫细胞暴露于各种病原体时,可能引发不同的转录程序[19-22]。在此处,表明在血液中的白细胞的基因表达模式可以用于区分由以下四种不同病原体导致的急性感染:流感病毒 A、革兰氏阴性细菌、大肠杆菌和革兰氏阳性细菌、金黄色葡萄球菌和肺炎链球菌,这些感染属于导致儿童住院的最普遍的感染。

两个参数可以解释观察到的在血液的白细胞的基因表达水平的差异: 1) 转录活性的变化(例如,干扰素可诱导基因的上调)和/或 2) 血样的细胞组成的变化(例如,嗜中性白细胞信号)。由于这些参数中的一个或者两个所导致的表达的变化可以通过病原体产生的分子或者由所述宿主释放的第二因子(例如,细胞因子)的作用直接进行调节。观察到从不同患者组中获得的血样本所细胞组成存在显著的差异。事实上,在临床实践中被大家所接受的常规的白细胞记数以及不同的记数不能区分病毒感染和细胞感染,更不能区分由革兰氏阳性细菌和革兰氏阴性细菌导致的感染。然而,本发明的发明者发现微小的差异可能能够解释观察到的转录信号,如由系统红斑狼疮中的嗜中性白细胞的转录信号所例证,系统红斑狼疮是由于 PBMC 制剂中存在的低密度的嗜中性白细胞的增加的流出物所导致的。疾病所涉及的位点可能也影响在血液中的白细胞中所观察到的表达谱,并且反应某些种类的病原体对不同感染位点的偏好。例如,大肠杆菌更可能导致尿道感染,而金黄色葡萄球菌最普遍的临床表现是皮肤/软组织感染和骨髓炎。然而,本发明的研究的结果暗示在单一疾病表现的情况下,可以发现独特的表达信号。事实上,当分析来自

患有较低呼吸性感染的患者的样本时，观察到由不同的病原体导致的感染中存在明显的区别，这证实了存在与病原体相关的转录信号。

在许多临床情况下，鉴定出导致急性传染病的致病因子的能力仍然是相当低的，而血白细胞转录谱分析可能能够改变对传染病的诊断。与微生物培养相反地是，通过血清学分析或者甚至基于 PCR 的测试，可以迅速且可靠地得到白细胞基因表达谱的结果而不需考虑疾病所涉及的位点。这些信息能够迅速地启动适当的抗感染治疗方案并制定适当的控制感染的措施。而且，血白细胞转录分析能够提供关于患者的信息，这些信息可以用于疾病的诊断，并作为用于疾病进展和预后潜在的标记。与通常用作疾病发展的指示剂的炎症标记（例如，白细胞记数、红细胞沉降速度和 C-反应蛋白）相比，基因表达阵列提供了全面的分子图，该分子图不仅能够反映组织的相对细胞组成，还能够反映由正在进行的免疫反应和/或病原体暴露导致的基因调节。

这些结果证明了血液白细胞中转录信号分析作为一种诊断传染病的辅助方法的价值，该方法不仅能用于单个的现场分析，而且也能用于对急性传染病、慢性传染病或者两者的结果进行检测、确定、评价、预后、诊断和预测。必需进行大量的多通道研究，以收集并独立地评价大量的样本，最终产生接近常规的临床诊断应用的血液中的白细胞基因表达谱。

患者信息。从 29 位被大肠杆菌感染的患者(平均年龄：2 个月；年龄范围：2 周- 16 岁)，31 位被金黄色葡萄球菌感染的患者(平均年龄：7 岁；年龄范围：3 个月- 18 岁)，13 位被肺炎链球菌感染的患者(平均年龄：2 岁；年龄范围：2 个月- 16 岁)，1 位被流感病毒 A 感染的患者(平均年龄：1.5 岁；年龄范围：3 周- 36 岁)，和 7 位健康的对照(平均年龄：11 个月；年龄范围：3 个月- 22 个月)获得血样。按照年龄和抗生素治疗将患者分为训练组和测试组(表 1)。所有患有急性感染的受试者以及它们的对照是从达拉斯的儿童医药中心（Children's Medical Center Dallas (CMC)）征募的。而 SLE 患者以及他们各自的对照从德克萨斯州苏格兰教会医院（Texas Scottish Rite Hospital）中征募的。该研究获得了伦理委员会（Institutional Review Boards）的批准，并获得所有患者的知情同意。通过相关组织标本或者血液的标准细菌培养方法，以及直接的荧光抗原测试和病毒培养确定微生物诊断。调研者根据来自微生物实验室数据和住院患者供认的记录的日常基楚鉴定潜在的符合条件的患者。根据病史、临床发现、细菌和病毒培养，以及免疫荧光试验进行第二步，以确定患者是否符合条件。排除怀疑患有（通过临床发现）或确定患

有（通过微生物试验）多微生物感染的患者，怀疑具有（通过临床发现）或确定具有（通过微生物试验）免疫缺陷病史的患者，患有慢性疾病或接受类固醇或者其它免疫调节试剂的治疗的患者。患者一旦被征募，就制定确定的微生物诊断方案。在研究开始后进行系统的检测，以确定是否存在伴随的病毒感染，并且对被细菌感染的 73 位患者中的 60 位(82%)进行呼吸性病毒培养。对照样本从预定要进行任选的外科手术的健康个体和健康的门诊病人中获得。

对血样的处理。在儿童医药中心（Children's Medical Center Dallas (CMC)）或者德克萨斯州苏格兰教会医院（Texas Scottish Rite Hospital, Dallas, TX），所有的血液收集在酸性柠檬酸盐葡萄糖试管(BD Vacutainer)中，随后在室温下立刻送到贝勒免疫研究院（Baylor Institute for Immunology Research, Dallas, TX），以进行处理。通过聚蔗糖梯度离心（Ficoll gradient）从 3-4 ml 血液中分离外周血单核细胞(PBMCs)，并随后将分离的单核细胞裂解在含有β巯基乙醇(BME)的 RLT 试剂(Qiagen, Valencia, CA) 中，然后由同一组人员将样本保存在在同一实验室的-80°C（在取血后 4-6 内）下，以使质量符合标准，和进行 RNA 样本的处理。

微阵列分析。按照厂商用法说明，使用 RNeasy kit (Qiagen, Valencia, CA) 分离总 RNA，并使用 Agilent 2100 Bioanalyzer (Agilent, Palo Alto, CA) 评价 RNA 的完整性。从 2-5 微克的总 RNA 中产生双链 cDNA，随后使用 Affymetrix RNA transcript labeling kits (Affymetrix Inc, Santa Clara, CA)在体外用生物素标记的核苷进行单轮转录。按照厂商的标准方法，使用 Sample Cleanup Module (Affymetrix)对生物素化的 cRNA 靶进行纯化，并随后被杂交到 Affymetrix HGU133A GeneChips (该基因芯片含有 22,283 探针组)上。使用 Affymetrix 共聚焦激光扫描仪对阵列进行扫描。通过实时 PCR 确定一组基因的表达结果。

实时 RT-PCR 分析。使用 TURBO DNA-free kit (Ambion Inc., Austin, TX) 对总的 RNAs 进行第二 DNA 酶处理。使用 Two-Cycle cDNA Synthesis kit (Affymetrix) 合成 cDNA，随后在体外进行转录(MEGAscript T7 kit, Ambion, Inc., Austin, TX)。按照厂商的用法说明，使用 Applied Biosystems TaqMan Assays 根据所需要的探针和引物组进行两步 RT-PCR。使用 High Capacity cDNA Archive Kit (Applied Biosystems) 进行逆转录 6（Reverse 6 transcription）。使用 ABI Prism 7700 Sequence Detection System 进行实时 PCR。

从一组含有 10 个人内源性对照中选择人 β -葡萄糖醛酸糖苷酶(GUSB)作为样本中的组成型表达最典型的对照,从而用于标准化的参比基因。按照厂商的用法说明,使用可比较的循环时间(CT)方法计算 mRNA 相对表达值。在 CT 方法中,通过以健康的供体的表达值为基准,将给定的患有传染病的患者的表达结果计算为标准化的差别,所述健康的供体的表达值接近所有健康供体的平均值($\Delta\Delta CT$)。

Illumina BeadChips: 这些微阵列由连接到 $3\mu m$ 珠上的 50 mer 寡核苷酸探针组成,所述珠被固定在微孔中的载玻片的表面。对样本进行处理,并随后通过 Illumina Inc(San Diego, CA)获取数据。使用 Illumina RNA amplification kit (Ambion, Austin, TX)制备靶。将 cRNA 靶杂交到 Sentrix Hu6 BeadChips (>46,000 探针)上, Sentrix Hu6 BeadChips 可以用 Illumina BeadStation 500 进行扫描。使用 Illumina's Beadstudio 软件评价荧光杂交信号。

将从所有 148 个分析的样本中获得的原始数据保存到 GEO (www.ncbi.nlm.nih.gov/geo/) (登记号 _____)。

微阵列数据分析。使用 Microarray Suite, 5.0 版本(MAS 5.0; Affymetrix)软件评价荧光杂交信号,标准化信号,和评价信号检测。通过在 MAS 5.0 中进行的运算法则分析对于每一个探针组的原始信号强度的值。为了最大程度地降低技术上的可变性,在每一天的运行中,最多对八个样本进行随机的杂交和染色。

使用 MAS 5.0 全面的方法对每个基因芯片的 500 的靶强度按比例缩放对每一个芯片的信号值进行标准化。仅对在至少一个被评价的患者类中的在至少 75% 的基因芯片中获得 P (存在) 信号的探针组 (质量控制探针) 进行分析。使用一种基因表达分析软件程序 (GeneSpring, Version 7.1 (Agilent)) 进行统计的分析、样本的等级聚类 and 分类。根据基因区分预先确定组的患者的能力,使用非参数单变量检验(曼恩-惠特尼 (Mann-Whitney) U 或费希尔精确检验 (Fishers exact test))对基因进行分级。采用 K 近邻 (K-Nearest Neighbors) (kNN) 方法确定对预先确定类的病原体具有最高区别能力的基因 (即, 分类基因) [23]。

K 近邻(kNN)方法: (1) 该运算法则通过基因预测能力的强度对基因进行分级,所述强度为通过非参数检验所确定的 p 值的负自然对数值;和(2)使用弃一交差验证通过从已知样本中系统性除去一个供体以用作测试样本来评价预测错误率 (或准确率)。反复进行该过程直到所有的供体都被“检测”。

将通过曼恩-惠特尼(Mann-Whitney)U或费希尔精确检验(Fishers exact test)获得的具有区别能力的基因列表合并,并用于区别不同的样本类别。(3)为了对样本进行分类,根据标准化的表达强度的欧几里得距离,所述运算法则通过测试与未知类样本最接近的已知类的编号得出类别,并计算p值。将具有最低p值的类归入到未知样本中。在所有的区别分析中,p值的截止率(ratio cut-off)为0.5。如果预测类的p值至少低于其它类的p值的2倍,则一个样本属于该类别(例如,流感病毒A类的p值/细菌类的p值)。

显著性模式分析。对每一组患者和他们各自的健康对照组之间进行统计比较(曼恩-惠特尼(Mann-Whitney)级别检验)。将发生显著变化的基因($p < 0.01$)分为两组:与两个提及的组(FLU和SLE)的对照相比,过表达组和低表达组。通过使用对这些组所选择的标准(例如,对于FLU组, $P < 0.01$;对于SLE组, $P > 0.5$)鉴定基因组。获得在其它疾病情况下的这些基因的P值(比较组:金黄色葡萄球菌、肺炎链球菌、大肠杆菌)。当基因表达的变化与所提及组的基因表达的变化相反时,将比较组的P值设定为1。使用GeneSpring(7.2版本)(Agilent)对p值数据进行处理,GeneSpring(7.2版本)是根据显著性模式用于对基因进行等级聚类 and 分组。

转录信号能够区分被流感病毒A感染的患者和被细菌感染的患者。使用标准的基因-水平分析,发现可以使用微阵列分析区分病毒感染(流感病毒A感染)和细菌感染(大肠杆菌感染和肺炎链球菌感染),如图1C的图解所示。图1C表示能够区分被流感病毒A感染和被细菌感染的基因表达信号。然后,对在训练组中能够准确区分(准确率为91%(21/23))被流感病毒A感染的患者和被细菌感染(大肠杆菌或金黄色葡萄球菌)的患者的35个分类基因在测试组($n=37$)中进行独立地的验证,准确率为95%(35/37)。

模块-水平微阵列数据分析。该策略是基于从大的微阵列基因表达数据组(8种疾病,接近250样本 $\times$ 44,000转录子)中最初提取的28组协同表达的基因(重编接近5000转录子),或者转录模块。随后,将这些模块用作分析用的结构单元,这些结构单元是基于模块-模块实现的:首先通过文献分析进行功能的解释,然后对从健康的受试者和患有急性感染的患者中获得的样本进行组间的比较。

图7表示可以用于在模块-水平描绘转录变化的基因失量鉴定疾病特异性的模式。根据模块-模块(module-by-module)对患者和未感染的个体之间进行组间的比较。斑点表示在模块中显著过表达(红色)或低表达(蓝色)

的百分比。该信息表示在栅格上，其坐标对应于 28 模块标志中的一个(例如，模块 M3.1 是第三行与第一列的交叉点)。

基因失量和作图方法可以减少噪音水平，并有助于数据的判读。在达拉斯的组也表明通过微阵列平台可以再现模块转录数据。

对在患者血液中的诊断性标记的鉴定。在模块-水平对总的转录的变化进行作图有助于解释患者 PBMC 转录谱。这还能揭示模块变化的疾病专一性组合。如在图 7 的图解所示，在模块 M3.1 (干扰素，绿色的圆-分别用红色和蓝色的斑点表示差异过表达或者低表达的转录子的百分比) 中比较了由革兰氏阳性细菌和革兰氏阴性细菌导致的变化。在两种革兰氏阳性菌中也发现了区别 (图 7，橙色的圆)。最令人感兴趣的是，尽管被流感病毒和被 RSV (呼吸道合胞病毒) 感染的临床症状类似，但是发现在这些病毒感染的患者的模块图谱中存在显著的差异。在被 RSV 感染的患者中，完全不存在干扰素可诱导的转录子，这与被流感病毒感染的情况显著不同，在被流感病毒感染的患者中，存在强大的干扰素应答 (M3.1)。还观察到其它的差别，最为显著的为模块 M1.4 和 M1.7 (在图 7 中以紫色突出表示)。

疾病严重性的标记的鉴定：目前可用于传染病的诊断的工具依赖于对病原体的直接检测(例如，通过培养、染色或 PCR)。与这些方法形成对照的是，监测患者免疫细胞的基因表达的变化可以预测疾病的严重性。事实上，模块表达水平(转录子的平均值)与疾病严重性的临床指示相关。在根据 U-统计进行多变量分析后，将与严重性相关 (正或负) 的模块合并成单一得分(产生“U-得分”-金黄色葡萄球菌和流感病毒的结果如图 8 所示)。

图 8 表示对患有急性感染的患者进行的疾病严重性评价的微阵列得分。根据 U-统计通过多变量分析将模块-水平微阵列表达数据合并为单一得分。这样获得的微阵列得分与通过疾病炎症性相关指示(例如，发烧、低血压、酸中毒)构成的临床得分相关。在训练组中鉴定标记，并在独立患者组 (测试组) 中技能型验证。这样基于血液的微阵列分析产生的临床信息可以用于：(1)确定致病因子；和(2)评价患有急性感染的患者的疾病的严重性。图 9a- 9c 概括了通过微阵列平台进行的独立的确定和验证。

可以理解的是本文描述的特定的实施方案用于例证说明的目的，而不是为了限制本发明。本发明的主要特征可以被用在不同的实施方案中而不超出本发明的范围。本领域的技术人员仅仅使用例行试验将会认识到，或者能够

确定许多等价于本文所述的特定方法。这些等价的技术方案被认为属于本发明的范围，并且包含在本发明的权利要求中。

在本说明书中所提到的所有的出版物和专利申请可表示本发明所述的技术领域的技术人员的水平。所有的出版物和专利申请引入本文作为参考，引入的程度犹如每一篇出版物或者专利申请具体地且单个地被指出引入本文作为参考。

所有在本文中公开的和要求保护的组合物和/或方法可以根据本发明公开内容在没有进行不适当的实验的情况下被制备和被实施。尽管根据优选的实施方案描述了本发明的组合物和方法，但是各种改变可以用于本文所述的组合物和/或方法中，用于本文所述方法的步骤或步骤的序列中，而不背离本发明的观念、精神实质和范围，这对于本领域的技术人员来说是显而易见的。更具体地说，某些化学和生理相关的因子可以取代本文所述的因子，这是很显然的。所有这种类似的取代和改变被认为属于由附加的权利要求所限定的精神实质、范围和观念中，这对于本领域的技术人员来说是显而易见的。

参考文献

1. Fauci, A.S. 2005. The global challenge of infectious diseases: the evolving role of the National Institutes of Health in basic and clinical research. *Nat Immunol* 6:743-747.

2. Relman, D.A. 2002. New technologies, human-microbe interactions, and the search for previously unrecognized pathogens. *J Infect Dis* 186 Suppl 2:S254-258.

3. Fauci, A.S. 2004. Emerging infectious diseases: a clear and present danger to humanity. *Jama* 292:1887-1888.

4. Medzhitov, R., and C.A. Janeway, Jr. 1997. Innate immunity: the virtues of a nonclonal system of recognition. *Cell* 91:295-298.

5. Medzhitov, R., and C. Janeway, Jr. 2000. Innate immune recognition: mechanisms and pathways [In Process Citation]. *Immunol Rev* 173:89-97.

6. Alizadeh, A.A., M.B. Eisen, R.E. Davis, C. Ma, I.S. Lossos, A. Rosenwald, J.C. Boldrick, H. Sabet, T. Tran, X. Yu, J.I. Powell, L. Yang, G.E. Marti, T. Moore, J. Hudson, Jr., L. Lu, D.B. Lewis, R. Tibshirani, G. Sherlock, W.C. Chan, T.C. Greiner, D.D. Weisenburger, J.O. Armitage, R. Warnke, R.

Levy, W. Wilson, M.R. Grever, J.C. Byrd, D. Botstein, P.O. Brown, and L.M. Staudt. 2000. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403:503-511.

7. Golub, T.R., D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J.P. Mesirov, H. Coller, M.L. Loh, J.R. Downing, M.A. Caligiuri, C.D. Bloomfield, and E.S. Lander. 1999. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286:531-537.

8. van de Vijver, M.J., Y.D. He, L.J. van't Veer, H. Dai, A.A. Hart, D.W. Voskuil, G.J. Schreiber, J.L. Peterse, C. Roberts, M.J. Marton, M. Parrish, D. Atsma, A. Witteveen, A. Glas, L. Delahaye, T. van der Velde, H. Bartelink, S. Rodenhuis, E.T. Rutgers, S.H. Friend, and R. Bernards. 2002. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347:1999-2009.

9. Bennett, L., A.K. Palucka, E. Arce, V. Cantrell, J. Borvak, J. Banchereau, and V. Pascual. 2003. Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. *J Exp Med* 197:711-723.

10. Rubins, K.H., L.E. Hensley, P.B. Jahrling, A.R. Whitney, T.W. Geisbert, J.W. Huggins, A. Owen, J.W. Leduc, P.O. Brown, and D.A. Relman. 2004. The host response to smallpox: analysis of the gene expression program in peripheral blood cells in a nonhuman primate model. *Proc Natl Acad Sci U S A* 101:15190-15195.

11. Baechler, E.C., F.M. Batliwalla, G. Karypis, P.M. Gaffney, W.A. Ortmann, K.J. Espe, K.B. Shark, W.J. Grande, K.M. Hughes, V. Kapur, P.K. Gregersen, and T.W. Behrens. 2003. Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc Natl Acad Sci U S A* 100:2610-2615.

12. Hoebe, K., and B. Beutler. 2004. LPS, dsRNA and the interferon bridge to adaptive immune responses: Trif, Tram, and other TIR adaptor proteins. *J Endotoxin Res* 10:130-136.

13. Yurchenko, V., M. O'Connor, W.W. Dai, H. Guo, B. Toole, B. Sherry, and M. Bukrinsky. 2001. CD147 is a signaling receptor for cyclophilin B. *Biochem Biophys Res Commun* 288:786-788.

14. Geiser, T., B. Dewald, M.U. Ehrengruber, I. Clark-Lewis, and M. Baggiolini. 1993. The interleukin-8-related chemotactic cytokines GRO alpha, GRO beta, and GRO gamma activate human neutrophil and basophil leukocytes. *J Biol Chem* 268:15419-15424.
15. Nagaoka, I., and S. Hirota. 2000. Increased expression of matrix metalloproteinase-9 in neutrophils in glycogen-induced peritoneal inflammation of guinea pigs. *Inflamm Res* 49:55-62.
16. Niemann, C.U., J.B. Cowland, P. Klausen, J. Askaa, J. Calafat, and N. Borregaard. 2004. Localization of serglycin in human neutrophil granulocytes and their precursors. *J Leukoc Biol* 76:406-415.
17. Pouliot, M., P.P. McDonald, P. Borgeat, and S.R. McColl. 1994. Granulocyte/macrophage colony-stimulating factor stimulates the expression of the 5-lipoxygenase-activating protein (FLAP) in human neutrophils. *J Exp Med* 179:1225-1232.
18. Herndon, B.L., S. Abbasi, D. Bennett, and D. Bamberger. 2003. Calcium-binding proteins MRP 8 and 14 in a *Staphylococcus aureus* infection model: role of therapy, inflammation, and infection persistence. *J Lab Clin Med* 141:110-120.
19. Chaussabel, D., R.T. Semnani, M.A. McDowell, D. Sacks, A. Sher, and T.B. Nutman. 2003. Unique gene expression profiles of human macrophages and dendritic cells to phylogenetically distinct parasites. *Blood* 102:672-681.
20. Nau, G.J., J.F. Richmond, A. Schlesinger, E.G. Jennings, E.S. Lander, and R.A. Young. 2002. Human macrophage activation programs induced by bacterial pathogens. *Proc Natl Acad Sci U S A* 99:1503-1508.
21. Boldrick, J.C., A.A. Alizadeh, M. Diehn, S. Dudoit, C.L. Liu, C.E. Belcher, D. Botstein, L.M. Staudt, P.O. Brown, and D.A. Relman. 2002. Stereotyped and specific gene expression programs in human innate immune responses to bacteria. *Proc Natl Acad Sci U S A* 99:972-977.
22. Huang, Q., D. Liu, P. Majewski, L.C. Schulte, J.M. Korn, R.A. Young, E.S. Lander, and N. Hacohen. 2001. The plasticity of dendritic cell responses to pathogens and their components. *Science* 294:870-875.
23. Silicon Genetics Inc. 2002. Class prediction.

-
24. Bammler T, Beyer RP, Bhattacharya S, et al. Standardizing global gene expression analysis between laboratories and across platforms. *Nat Methods*. 2005;2:351-356.
 25. Irizarry RA, Warren D, Spencer F, et al. Multiple-laboratory comparison of microarray platforms. *Nat Methods*. 2005;2:345-350.
 26. Larkin JE, Frank BC, Gavras H, Sultana R, Quackenbush J. Independence and reproducibility across microarray platforms. *Nat Methods*. 2005;2:337-344.
 27. Jenner RG, Young RA. Insights into host responses against pathogens from transcriptional profiling. *Nat Rev Microbiol*. 2005;3:281-294.
 28. Pai RK, Pennini ME, Tobian AA, Canaday DH, Boom WH, Harding CV. Prolonged toll-like receptor signaling by *Mycobacterium tuberculosis* and its 19-kilodalton lipoprotein inhibits gamma interferon-induced regulation of selected genes in macrophages. *Infect Immun*. 2004;72:6603-6614.

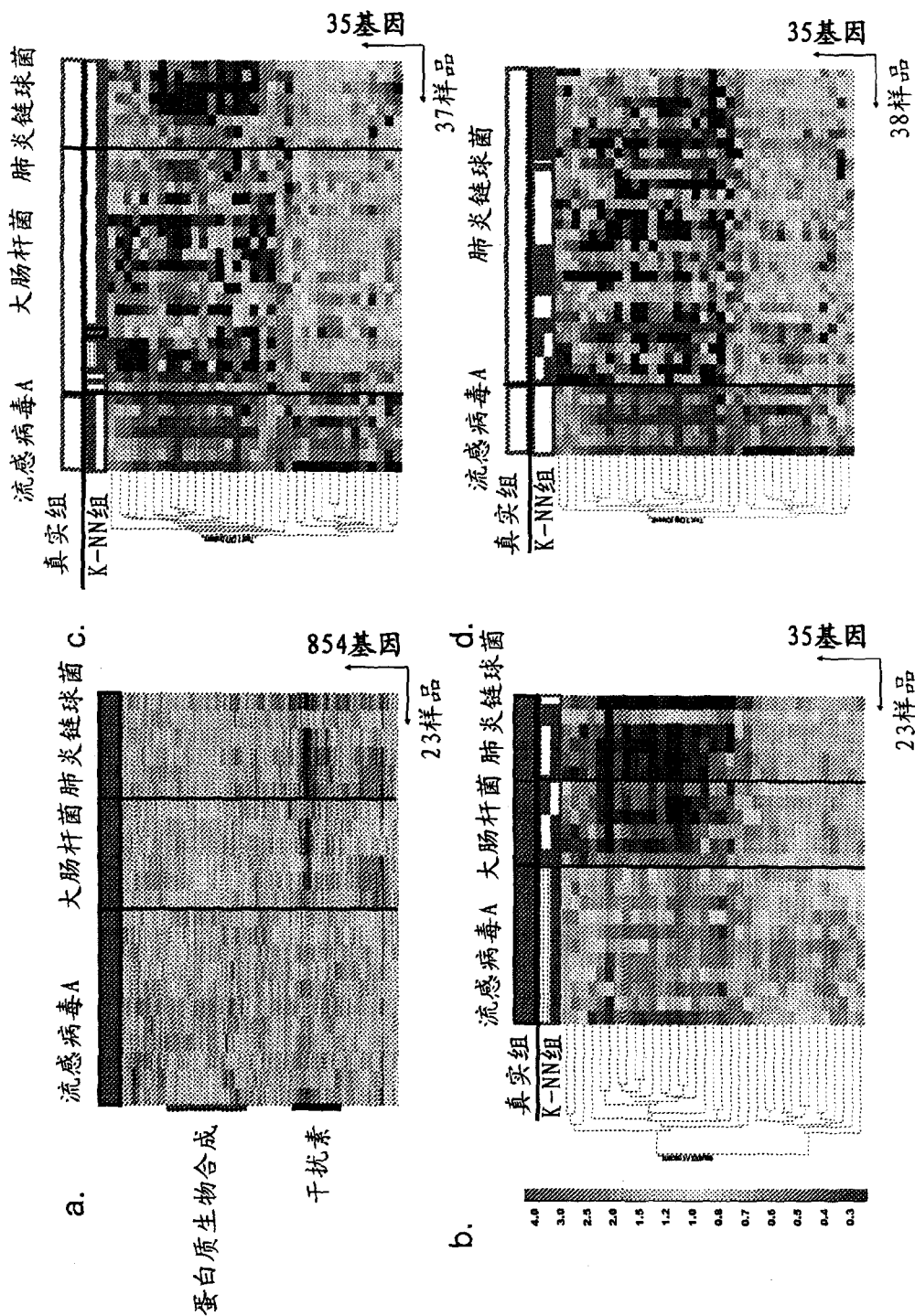


图1

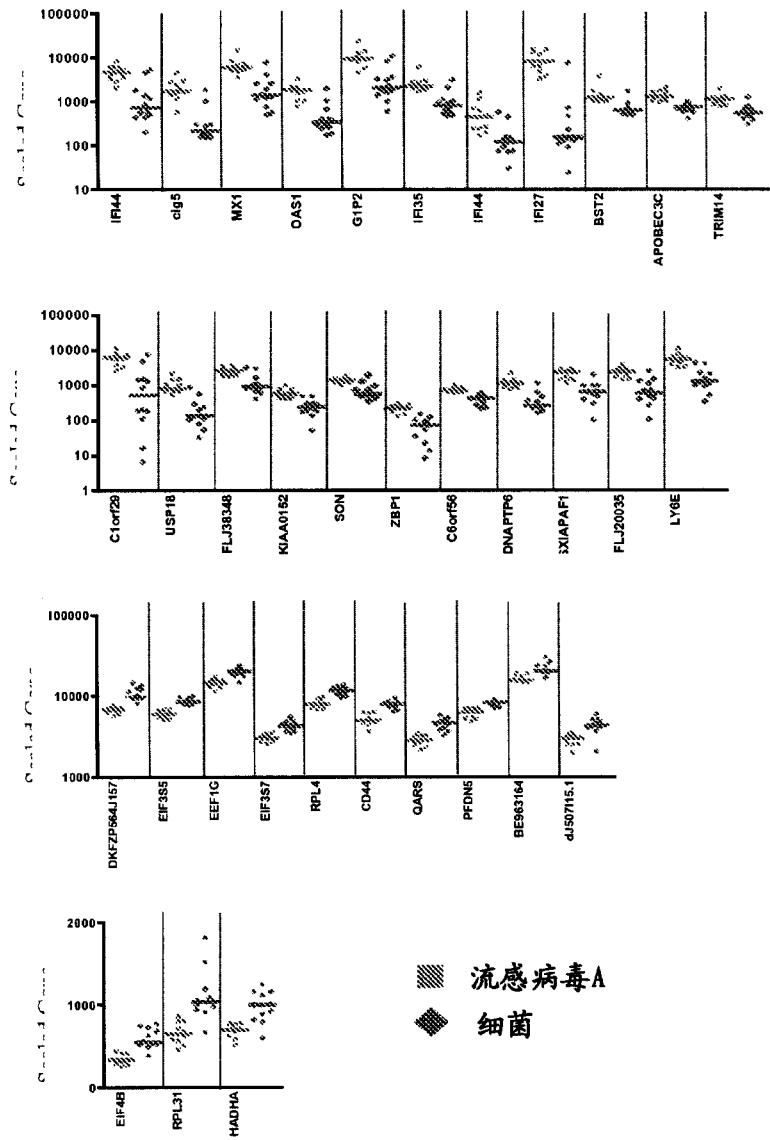


图2

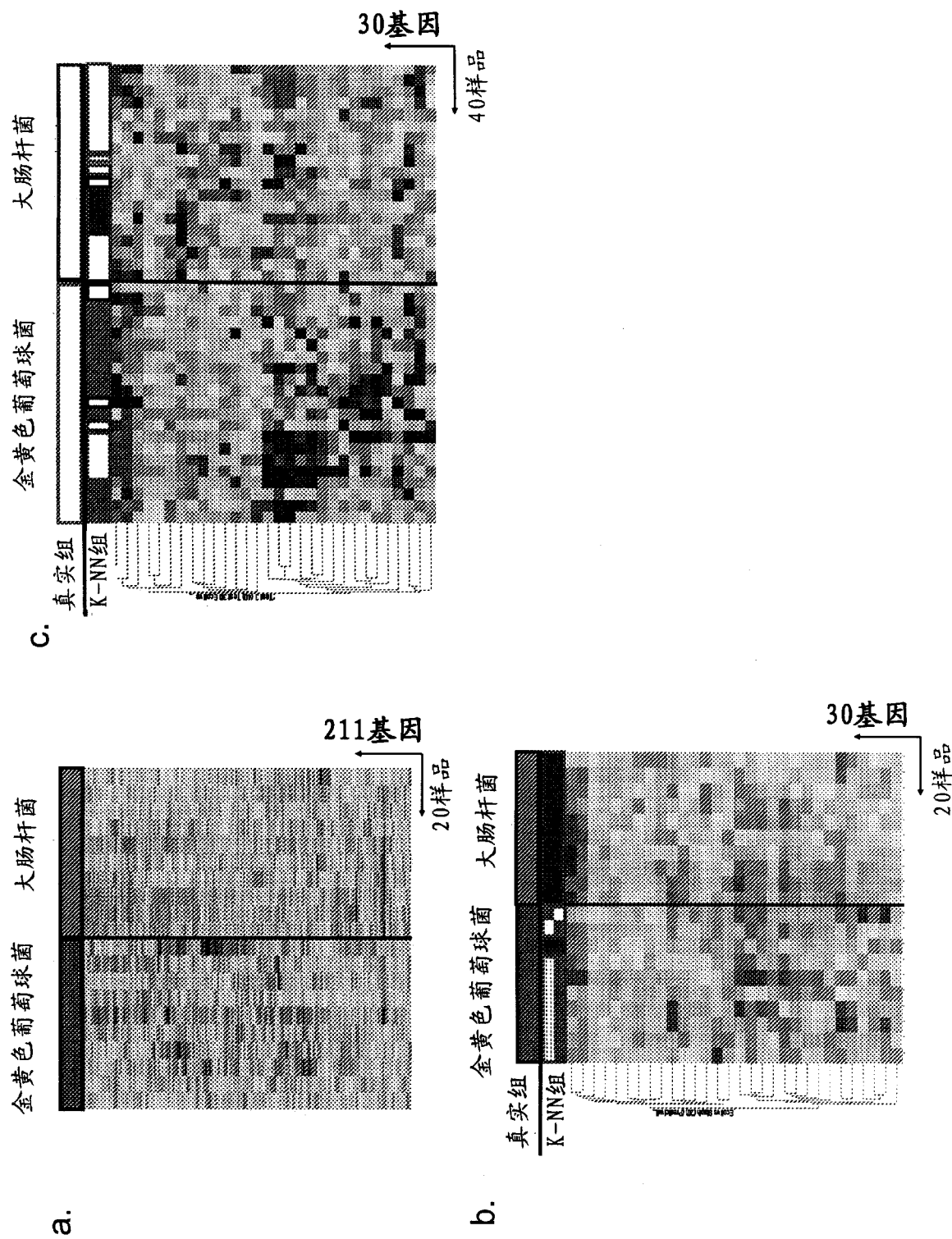


图3

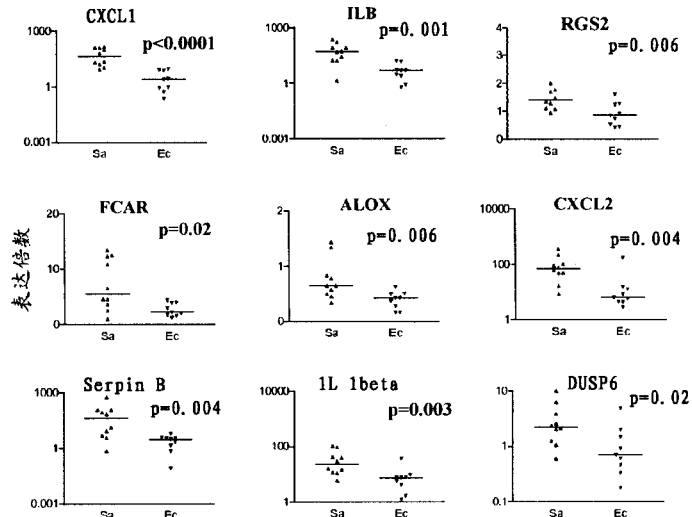


图 3d

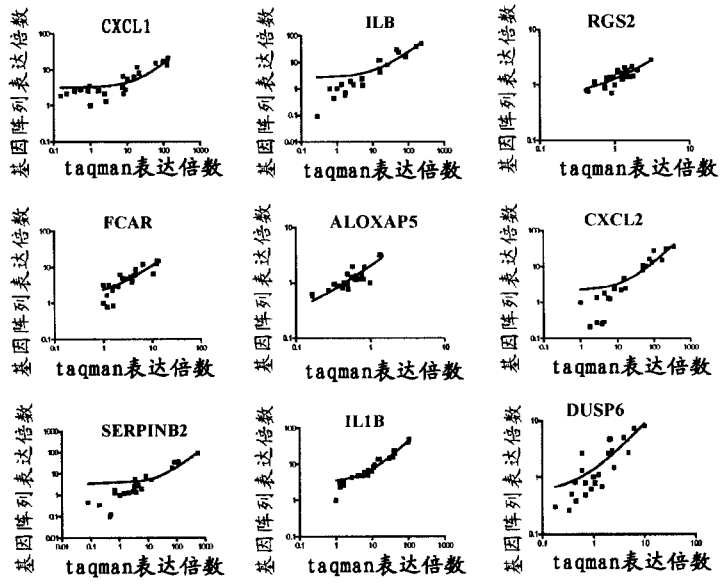


图 3e

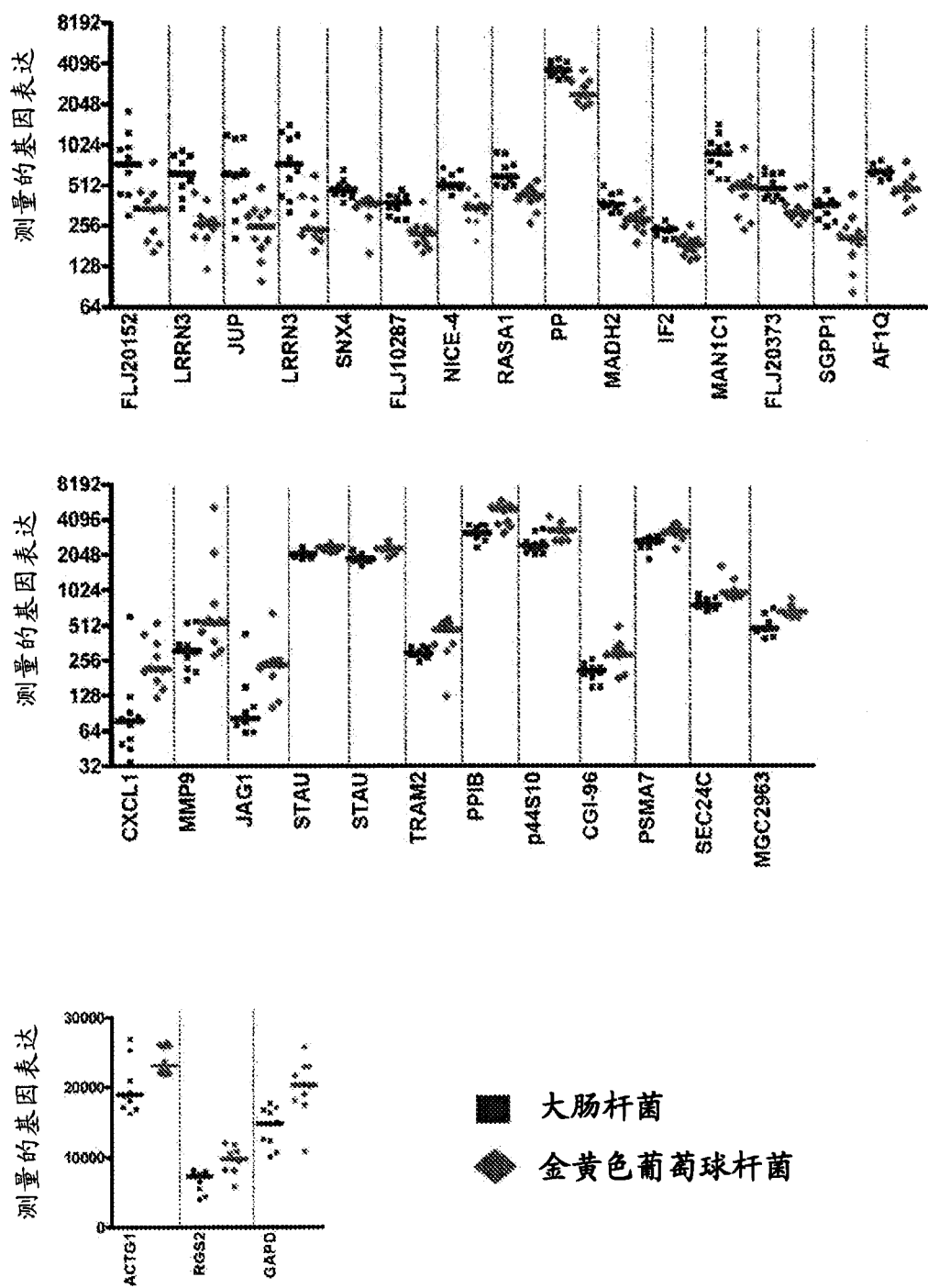
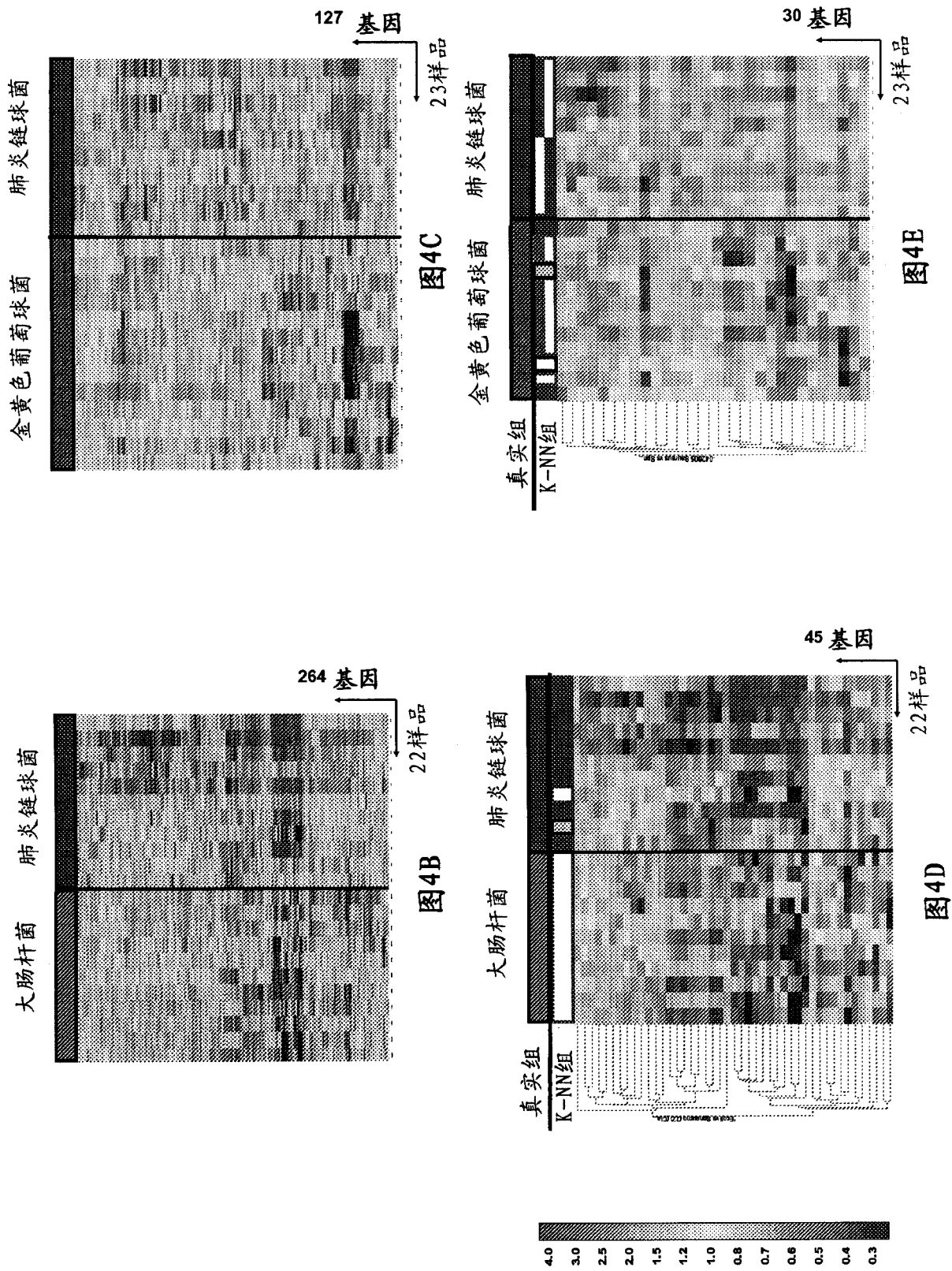
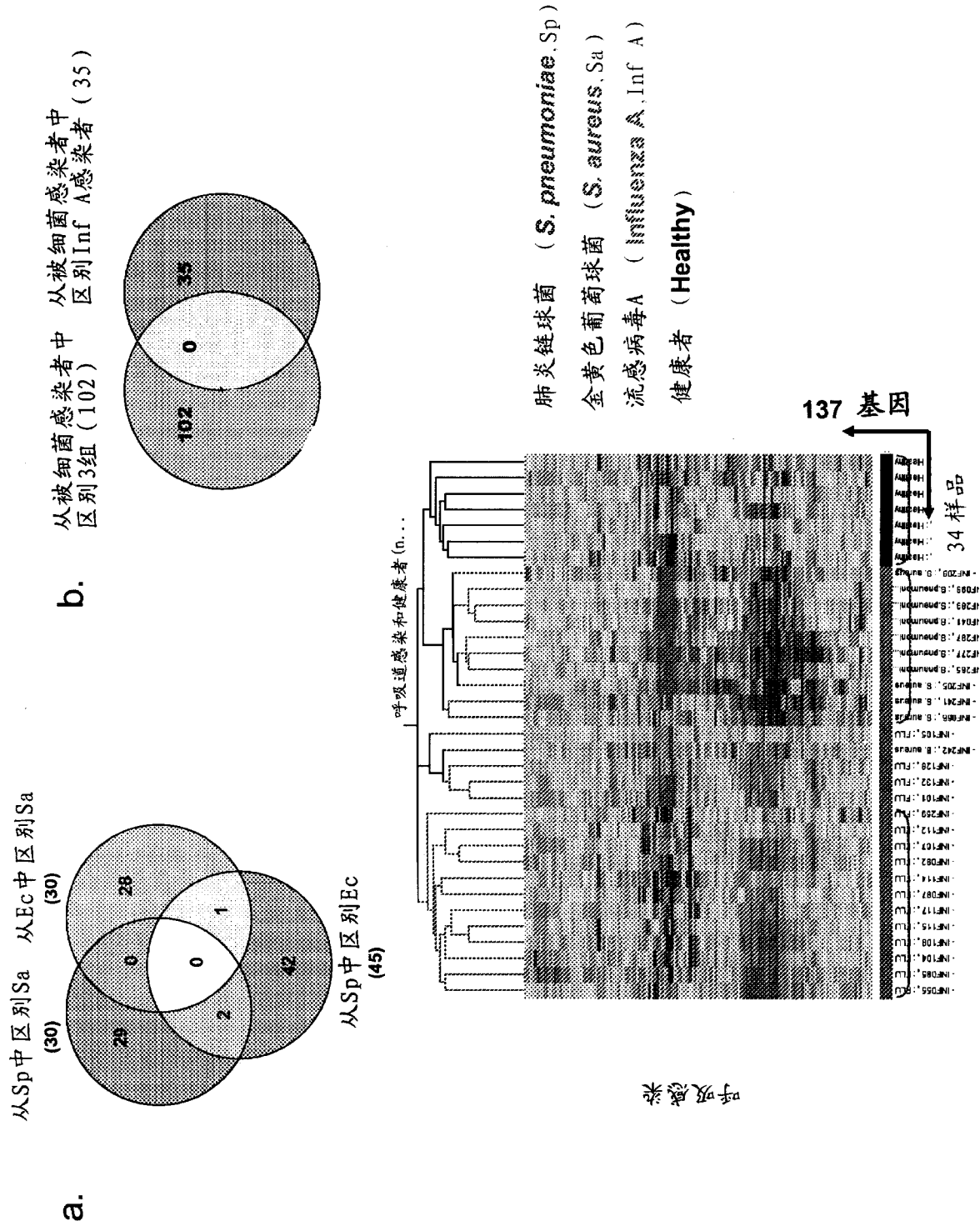


图 4A





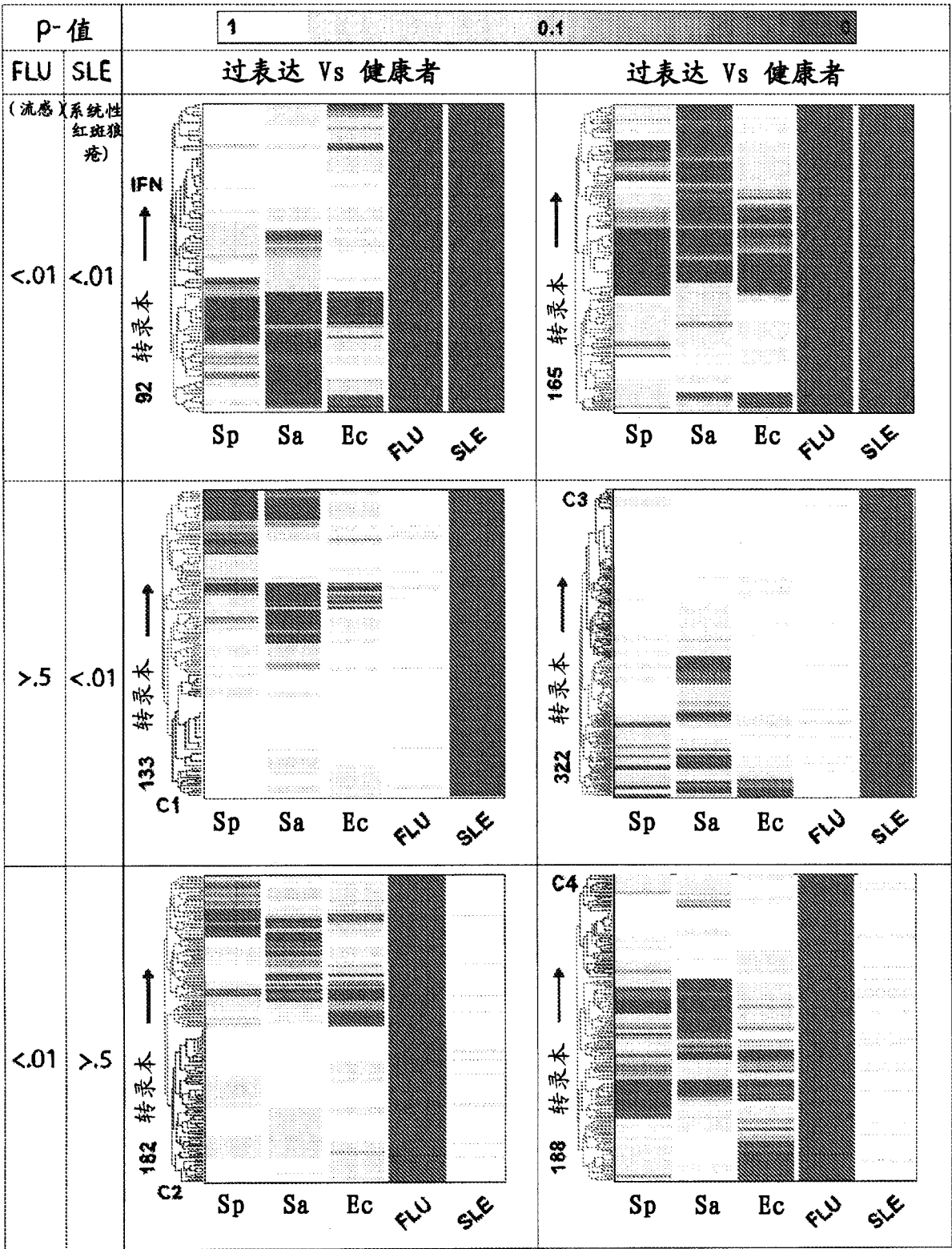
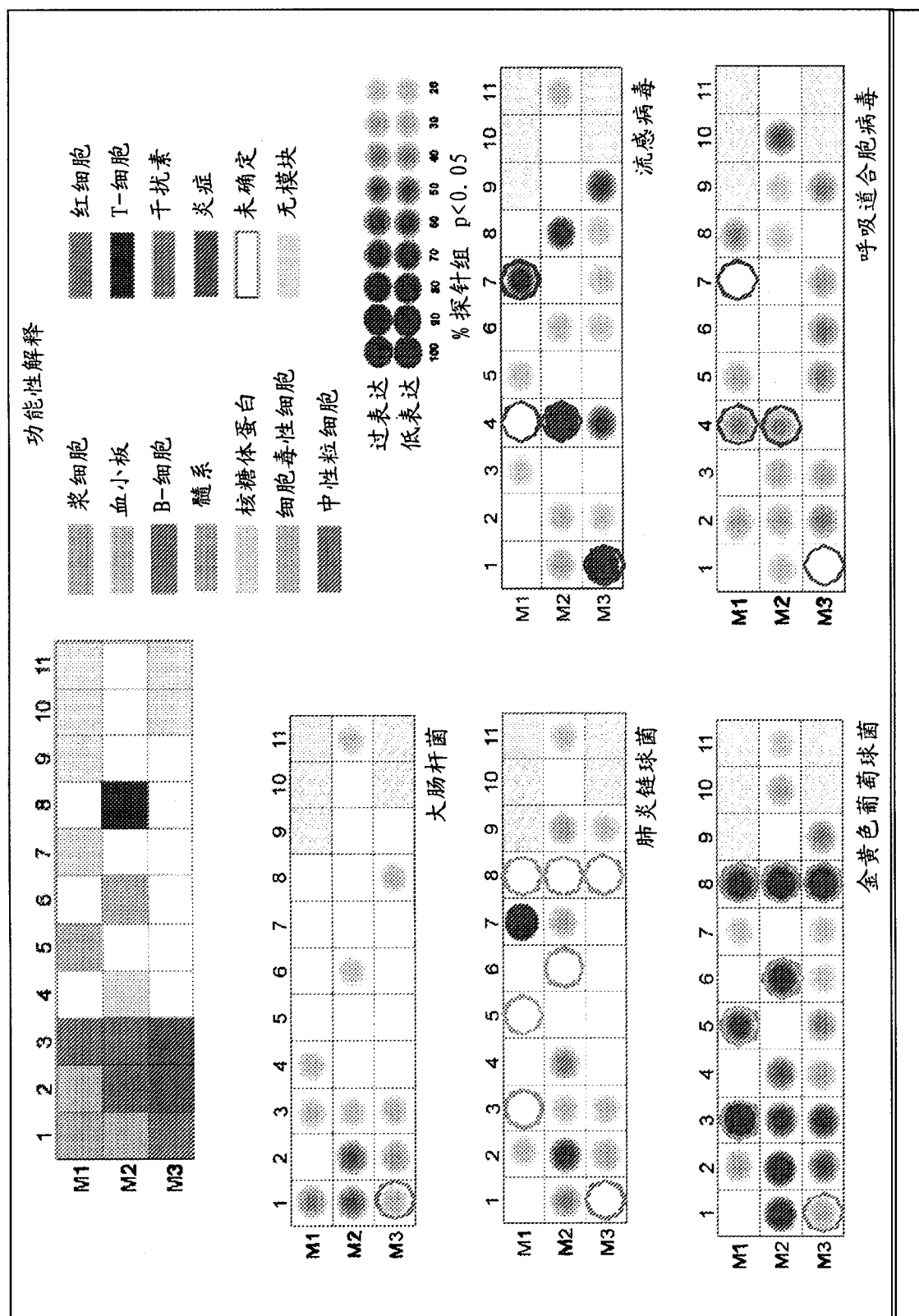


图 6



7

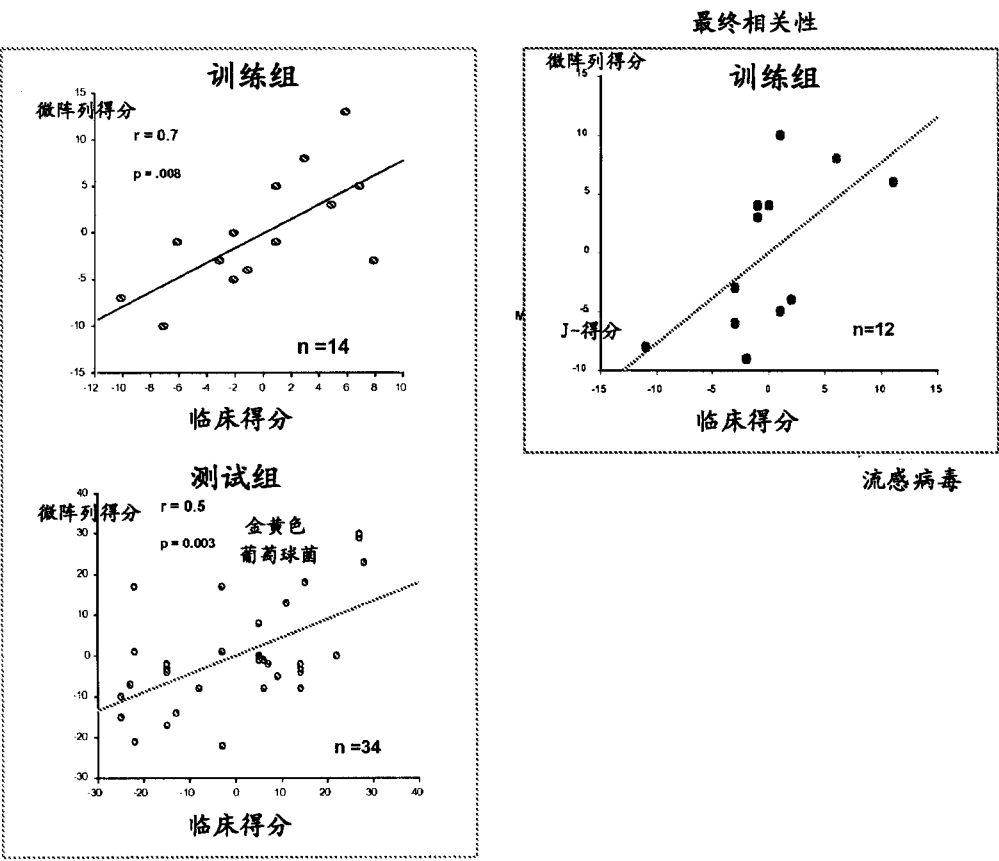


图 8

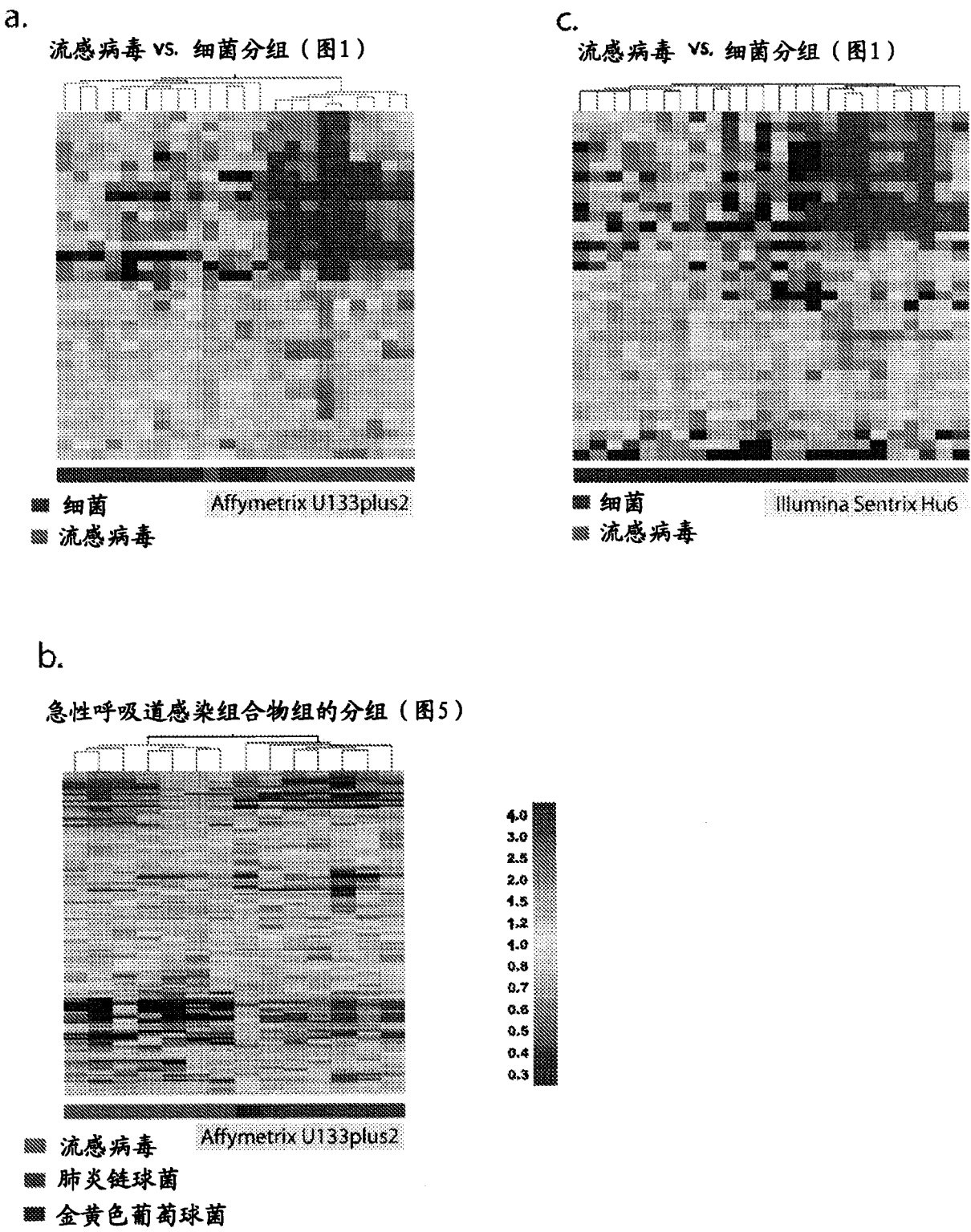


图 9