



[12] 发明专利申请公布说明书

[21] 申请号 200580012425.3

[43] 公开日 2007年4月18日

[11] 公开号 CN 1950701A

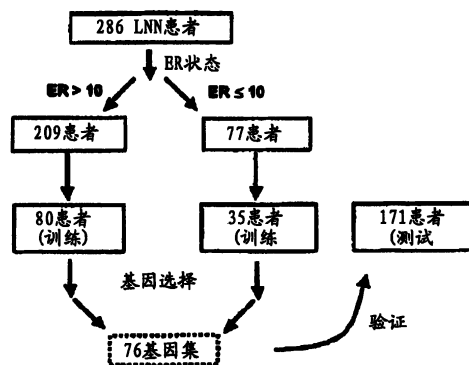
[22] 申请日 2005.2.18
 [21] 申请号 200580012425.3
 [30] 优先权
 [32] 2004.2.20 [33] US [31] 10/783,271
 [32] 2004.12.8 [33] US [31] 60/634,430
 [86] 国际申请 PCT/US2005/005711 2005.2.18
 [87] 国际公布 WO2005/083429 英 2005.9.9
 [85] 进入国家阶段日期 2006.10.20
 [71] 申请人 维里德克斯有限责任公司
 地址 美国新泽西州
 [72] 发明人 Y·王

[74] 专利代理机构 中国专利代理(香港)有限公司
 代理人 程 淼 黄可峻

权利要求书9页 说明书74页 附图11页
 序列表(电子公布)

[54] 发明名称
 乳腺癌预后

[57] 摘要
 通过分析一组基因的表达进行了提供乳腺癌预后的方法。包括了在各种介质例如微阵列中的基因表达分布型，还包括了含有它们的试剂盒。



1. 评定乳腺癌状态的方法，包括步骤
 - a. 从乳腺癌患者获得生物样品；和
 - b. 测量样品中基因的表达水平，所述基因选自由编码 mRNA 的那些构成的组，所述 mRNA：
 - i. 相应于 SEQ ID NO: 1-111；或
 - ii. 由探针集识别，所述探针集选自由表 10 中描述的相应于 SEQ ID NO: 1-111 的 psid 构成的组

其中所述基因表达水平高于或低于预定的截断水平是乳腺癌状态的指示。

2. 给乳腺癌患者分期的方法，包括步骤
 - a. 从乳腺癌患者获得生物样品；和
 - b. 测量样品中基因的表达水平，所述基因选自由编码 mRNA 的那些构成的组，所述 mRNA：
 - i. 相应于 SEQ ID NO: 1-111；或
 - ii. 由探针集识别，所述探针集选自由表 10 中描述的相应于 SEQ ID NO: 1-111 的 psid 构成的组

其中所述基因表达水平高于或低于预定的截断水平是乳腺癌分期的指示。

3. 权利要求 2 的方法，其中所述分期相应于 TNM 系统的分类。
4. 权利要求 2 的方法，其中所述分期相应于具有类似基因表达分布型的患者。

5. 确定乳腺癌患者治疗方案的方法，包括步骤
 - a. 从乳腺癌患者获得生物样品；和
 - b. 测量样品中基因的表达水平，所述基因选自由编码 mRNA 的那些构成的组，所述 mRNA：
 - i. 相应于 SEQ ID NO: 1-111；或
 - ii. 由探针集识别，所述探针集选自由表 10 中描述的相应于 SEQ ID NO: 1-111 的 psid 构成的组

其中所述基因表达水平高于或低于预定的截断水平是复发的风险的充分指示，以允许医师确定推荐以预防复发的治疗的程度和种类。

6. 治疗乳腺癌患者的方法，包括步骤：

- a. 从乳腺癌患者获得生物样品；和
- b. 测量样品中基因的表达水平，所述基因选自由编码 mRNA 的那些构成的组，所述 mRNA：
 - i. 相应于 SEQ ID NO: 1-111；或
 - ii. 由探针集识别，所述探针集选自由表 10 中描述的相应于 SEQ ID NO: 1-111 的 psid 构成的组

其中所述基因表达水平高于或低于预定的截断水平是复发的高风险的指示，和；

- c. 如果他们是高风险患者，用辅助治疗来治疗患者。
7. 权利要求 1、2、5 或 6 的方法，其中所述 SEQ ID NO 是 1-35。
8. 权利要求 1、2、5 或 6 的方法，其中所述 SEQ ID NO 是 36-95。
9. 权利要求 8 的方法，用于为 ER 阳性患者提供预后。
10. 权利要求 1、2、5 或 6 的方法，其中所述 SEQ ID NO 是 96-111。
11. 权利要求 10 的方法，用于为 ER 阴性患者提供预后。
12. 权利要求 1、2、5 或 6 的方法，其中所述 SEQ ID NO 是 36-111。
13. 权利要求 1、2、5 或 6 的方法，其中所述样品通过选自由大块组织制备和激光俘获显微解剖构成的组的方法来制备。
14. 权利要求 13 的方法，其中大块组织制备物从活组织检查或外科手术样本获得。
15. 权利要求 1、2、5 或 6 的方法，进一步包含测量编码 mRNA 的至少一个基因的表达水平，所述 mRNA：
 - i. 相应于 SEQ ID NO: 112-132；或
 - ii. 由探针集识别，所述探针集选自由表 10 中描述的相应于 SEQ ID NO: 112-132 的 psid 构成的组。
16. 权利要求 1、2、5 或 6 的方法，进一步包含测量在样品中组成性表达的至少一个基因的表达水平。
17. 权利要求 1、2、5 或 6 的方法，进一步包含确定样品的雌激素受体 (ER) 状态。
18. 权利要求 17 的方法，其中通过测量指示 ER 状态的至少一个

基因的表达水平确定 ER 状态。

19. 权利要求 17 的方法，其中通过测量样品中 ER 的存在确定 ER 状态。

20. 权利要求 19 的方法，其中免疫组织化学地测量 ER 的存在。

21. 权利要求 1、2、5 或 6 的方法，其中所述样品是从原发性肿瘤获得的。

22. 权利要求 1、2、5 或 6 的方法，其中特异性是至少约 40%。

23. 权利要求 1、2、5 或 6 的方法，其中敏感性是至少约 90%。

24. 权利要求 1、2、5 或 6 的方法，其中将基因的表达模式与指示复发患者的表达模式相比较。

25. 其中权利要求 24 的方法，其中表达模式的比较通过模式识别方法来进行。

26. 权利要求 25 的方法，其中所述模式识别方法包括使用 Cox's 比例危险分析。

27. 权利要求 1、2、5 或 6 的方法，其中相对于良性细胞或正常组织，所述预定的截断水平是在样品中至少 1.5 倍过量或减量表达。

28. 权利要求 1、2、5 或 6 的方法，其中相对于良性细胞或正常组织，在具有转移性细胞的样品中，所述预定的截断水平具有至少统计学上显著的 p 值过量表达。

29. 权利要求 28 的方法，其中所述 p 值小于 0.05。

30. 权利要求 1、2、5 或 6 的方法，其中在微阵列或基因芯片上测量基因表达。

31. 权利要求 21 的方法，其中所述微阵列是 cDNA 阵列或寡核苷酸阵列。

32. 权利要求 21 的方法，其中所述微阵列或基因芯片进一步包括一个或多个内部对照试剂。

33. 权利要求 1、2、5 或 6 的方法，其中通过对提取自所述样品的 RNA 的聚合酶链式反应 (PCR) 进行的核酸扩增来测定基因表达。

34. 权利要求 33 的方法，其中所述 PCR 是逆转录聚合酶链式反应 (RT-PCR)。

35. 权利要求 34 的方法，其中所述 RT-PCR 进一步包含一个或多个内部对照试剂。

36. 权利要求 1、2、5 或 6 的方法，其中通过测量或检测由所述基因编码的蛋白来检测基因表达。

37. 权利要求 36 的方法，其中由特异于所述蛋白的抗体来检测所述蛋白。

38. 权利要求 1、2、5 或 6 的方法，其中通过测量基因的特征来检测基因表达。

39. 权利要求 38 的方法，其中所述测量的特征选自由 DNA 扩增、甲基化、突变和等位变异构成的组。

40. 为乳腺癌患者交叉验证预后基因表达分布型的方法，包括步骤：

- a. 从统计学上显著数目的患者生物样品获得基因表达数据；
- b. 将样品顺序随机化；
- c. 撇开约 10%-50% 的样品的数据；
- d. 对剩余样品，对所有变量计算感兴趣的因数并选择满足 p 值截断 (p) 的变量；
- e. 使用前向搜索来选择满足预测模型的变量，并评估训练误差直到它命中预定的误差率；
- f. 对留出的 10-50% 的样品测试所述预测模型；
- g. 除去一组新的样品重复步骤 c-g；和
- h. 继续步骤 c) -g) 直到已经测试了 100% 的样品并记录了分类表现。

41. 根据权利要求 40 的方法，其中步骤 h 中获得的基因表达数据由选自编码 mRNA 的那些构成的组的基因来表示，所述 mRNA：

- i. 相应于 SEQ ID NO: 3、4、10、18、37、40、42、43、45、55、58、64、67、72-74、76、81、85-86、89、97、100-101、110-111、125 和 132-442，或 2、3、5、12、20、25、36、37、39、40、41、42、43、45、46、51、52、53、54、55、57、58、59、60、61、62、63、64、66、67、69、70、72、73、74、76、80、81、83、84、85、86、87、88、89、90、94、97、98、101、102、104、107、110、111、132、139、142、150、151、153、154、158、161、163、167、170、171、173、175、181、182、183、186、188、190、192、204、206、207、212、215、218、221、223、225、228、231、232、236、238、239、

240、241、242、243、246、248、249、255、257、267、269、270、271、273、280、281、282、288、290、291、299、301、304、306、311、314、315、318、327、328、338、342、346、348、354、366、368、371、375、385、388、391、395、397、402、405、409、422、424、428、429、435、436、440、651 和 443-650; 或

ii. 由探针集识别,所述探针集选自由表 10 中描述的相应于 SEQ ID NO: 3、4、10、18、37、40、42、43、45、55、58、64、67、72-74、76、81、85-86、89、97、100-101、110-111、125 和 132-442, 或 2、3、5、12、20、25、36、37、39、40、41、42、43、45、46、51、52、53、54、55、57、58、59、60、61、62、63、64、66、67、69、70、72、73、74、76、80、81、83、84、85、86、87、88、89、90、94、97、98、101、102、104、107、110、111、132、139、142、150、151、153、154、158、161、163、167、170、171、173、175、181、182、183、186、188、190、192、204、206、207、212、215、218、221、223、225、228、231、232、236、238、239、240、241、242、243、246、248、249、255、257、267、269、270、271、273、280、281、282、288、290、291、299、301、304、306、311、314、315、318、327、328、338、342、346、348、354、366、368、371、375、385、388、391、395、397、402、405、409、422、424、428、429、435、436、440、651 和 443-650 的 psid 构成的组。

42. 为乳腺癌患者独立验证预后基因表达分布型的方法, 包括步骤:

- a. 从统计学上显著数目的患者生物样品获得基因表达数据;
- b. 将基因表达数据中的来源变异性标准化;
- c. 对早先选择的所有变量计算感兴趣的因数; 和
- d. 对所述样品测试预测模型并记录分类表现。

43. 根据权利要求 42 的方法, 其中步骤 d. 中获得的基因表达数据由选自编码 mRNA 的那些构成的组的基因来表示, 所述 mRNA:

- i. 相应于 SEQ ID No: 3、4、10、18、37、40、42、43、45、55、58、64、67、72-74、76、81、85-86、89、97、100-101、110-111、125 和 132-442, 或 2、3、5、12、20、25、36、37、39、40、41、42、43、45、46、51、52、53、54、55、57、58、59、60、61、62、63、

64、66、67、69、70、72、73、74、76、80、81、83、84、85、86、87、88、89、90、94、97、98、101、102、104、107、110、111、132、139、142、150、151、153、154、158、161、163、167、170、171、173、175、181、182、183、186、188、190、192、204、206、207、212、215、218、221、223、225、228、231、232、236、238、239、240、241、242、243、246、248、249、255、257、267、269、270、271、273、280、281、282、288、290、291、299、301、304、306、311、314、315、318、327、328、338、342、346、348、354、366、368、371、375、385、388、391、395、397、402、405、409、422、424、428、429、435、436、440、651 和 443-650；或

ii. 由探针集识别,所述探针集选自由表 10 中描述的相应于 SEQ ID NO: 3、4、10、18、37、40、42、43、45、55、58、64、67、72-74、76、81、85-86、89、97、100-101、110-111、125 和 132-442, 或 2、3、5、12、20、25、36、37、39、40、41、42、43、45、46、51、52、53、54、55、57、58、59、60、61、62、63、64、66、67、69、70、72、73、74、76、80、81、83、84、85、86、87、88、89、90、94、97、98、101、102、104、107、110、111、132、139、142、150、151、153、154、158、161、163、167、170、171、173、175、181、182、183、186、188、190、192、204、206、207、212、215、218、221、223、225、228、231、232、236、238、239、240、241、242、243、246、248、249、255、257、267、269、270、271、273、280、281、282、288、290、291、299、301、304、306、311、314、315、318、327、328、338、342、346、348、354、366、368、371、375、385、388、391、395、397、402、405、409、422、424、428、429、435、436、440、651 和 443-650 的 psid 构成的组。

44. 由根据权利要求 42 或 43 的方法获得的基因分布型。

45. 产生复发危险分值使得可进行乳腺癌患者的预后的方法, 包括步骤:

- a. 从统计学上显著数目的患者生物样品获得基因表达数据;
- b. 对所述数据应用单变量 Cox's 回归分析来获得选择的基因;
- c. 用标准 Cox's 系数对所选择的基因应用加权的表达水平, 以获得可以作为复发危险分值来应用的预测模型。

46. 权利要求 38 的方法，其中所述复发危险分值通过以下公式获得：

$$\text{复发分值} = A \cdot I + \sum_{i=1}^{60} I \cdot w_i x_i + B \cdot (1 - I) + \sum_{j=1}^{16} (1 - I) \cdot w_j x_j$$

其中

$$I = \begin{cases} 1 & \text{如果 X 基因水平} > 10 \text{ fmol 每 mg 蛋白} \\ 0 & \text{如果 X 基因水平} \leq 10 \text{ fmol 每 mg 蛋白} \end{cases}$$

A 和 B 是常数

w_i 是 X 基因 + 标记物的标准化的 Cox's 回归系数

x_i 是按 log2 尺度的 ER + 标记物的表达值

w_j 是 X 基因 - 标记物的标准化的 Cox's 回归系数

x_j 是按 log2 尺度的 ER - 标记物的表达值

X 基因选自由编码 mRNA 的那些构成的组，所述 mRNA：

- i. 相应于 SEQ ID NO: 1-111；或
- ii. 由探针集识别，所述探针集选自由表 10 中描述的相应于 SEQ ID NO: 1-111 的 psid 构成的组。

47. 产生乳腺癌预后患者报告的方法，包括步骤：

- a. 从所述患者获得生物样品；
- b. 测量所述样品的基因表达；
- c. 对步骤 b. 的结果应用复发危险分值；和
- d. 使用步骤 c. 中获得的结果来产生报告。

48. 权利要求 47 的方法，其中所述报告含有相对于患者群体的患者结局和/或风险概率的评定。

49. 由根据权利要求 47 的方法产生的患者报告。

50. 一种组合物，包含至少一个探针集，所述探针集选自以下构成的组：SEQ ID NO: 1-111；或表 10 中描述相应于 SEQ ID NO: 1-111 的 psid。

51. 用于进行分析以确定生物样品中的乳腺癌预后的试剂盒，包含：用于检测基因组合的分离的核酸序列、它们的互补物、或其部分

的材料，所述基因选自由编码 mRNA 的那些构成的组，所述 mRNA：

i. 相应于 SEQ ID NO: 1-111；或

ii. 由探针集识别，所述探针集选自由表 10 中描述的相应于 SEQ ID NO: 1-111 的 psid 构成的组。

52. 权利要求 51 的试剂盒，其中所述 SEQ ID NO 是 36-95。

53. 权利要求 51 的试剂盒，其中所述 SEQ ID NO 是 96-111。

54. 权利要求 51 的试剂盒，其中所述 SEQ ID NO 是 36-111。

55. 权利要求 51 的试剂盒，进一步包含用于进行微阵列分析的试剂。

56. 权利要求 51 的试剂盒，进一步包含介质，通过所述介质分析所述核酸序列、它们的互补物或其部分。

57. 用于评定乳腺癌状态的物品，包含：用于检测基因组合的分离的核酸序列、它们的互补物、或其部分的材料，所述基因选自由编码 mRNA 的那些构成的组，所述 mRNA：

i. 相应于 SEQ ID NO: 1-111；或

ii. 由探针集识别，所述探针集选自由表 10 中描述的相应于 SEQ ID NO: 1-111 的 psid 构成的组。

58. 权利要求 57 的物品，其中所述 SEQ ID NO 是 36-95。

59. 权利要求 57 的物品，其中所述 SEQ ID NO 是 96-111。

60. 权利要求 57 的物品，其中所述 SEQ ID NO 是 36-111。

61. 权利要求 57 的物品，进一步包含用于进行微阵列分析的试剂。

62. 权利要求 57 的物品，进一步包含介质，通过所述介质分析所述核酸序列、它们的互补物或其部分。

63. 用于进行权利要求 1、2、5 或 6 的方法的微阵列或基因芯片。

64. 权利要求 63 的微阵列，包括基因组合的分离的核酸序列、它们的互补物或其部分，所述基因选自由编码 mRNA 的那些构成的组，所述 mRNA：

i. 相应于 SEQ ID NO: 1-111；或

ii. 由探针集识别，所述探针集选自由表 10 中描述的相应于 SEQ ID NO: 1-111 的 psid 构成的组

其中所述组合足以表征生物样品中的乳腺癌状态或复发风险。

65. 权利要求 63 的微阵列，其中所述测量或表征是至少 1.5 倍过量或减量表达。

66. 权利要求 63 的微阵列，其中所述测量提供了统计学上显著的 p 值过量或减量表达。

67. 权利要求 63 的微阵列，其中所述 p 值小于 0.05。

68. 权利要求 63 的微阵列，包含 cDNA 阵列或寡核苷酸阵列。

69. 权利要求 63 的微阵列，进一步包含一个或多个内部对照试剂。

70. 一种诊断/预后档案，包含基因组合的分离的核酸序列、它们的互补物、或其部分，所述基因选自由编码 mRNA 的那些构成的组，所述 mRNA：

i. 相应于 SEQ ID NO: 1-111；或

ii. 由探针集识别，所述探针集选自由表 10 中描述的相应于 SEQ ID NO: 1-111 的 psid 构成的组

其中所述组合足以表征生物样品中的乳腺癌状态或复发风险。

71. 权利要求 70 的档案，其中所述测量或表征是至少 1.5 倍过量或减量表达。

72. 权利要求 70 的档案，其中所述测量提供了统计学上显著的 p 值过量或减量表达。

73. 权利要求 70 的档案，其中所述 p 值小于 0.05。

乳腺癌预后

发明领域

本发明涉及根据患者生物样品的基因表达分布型的乳腺癌患者预后。

背景技术

约 60-70% 的淋巴结阴性 (LNN) 乳腺癌患者通过单独的局部或区域治疗被治愈。Lancet (1998a); 以及 Lancet (1998b)。已经发展出指导原则来帮助临床医师选择应接受辅助治疗的患者。最广泛使用的建议是 St. Gallen 指标 (Goldhirsch et al. (2003)) 和国立卫生研究院 (NIH) 的公认指标。Eifel et al. (2001)。这些指导原则指出 85% -90% 的 LNN 患者是辅助性全身治疗的候选者。

需要具体地鉴定患者疾病复发的风险以确保她接受合适的治疗。当前, 只有少数诊断工具能鉴定有风险 (at-risk) 患者。已经使用基因表达模式来将乳腺肿瘤分成不同的临床上的相关亚型。Perou et al. (2000); Sørlie et al. (2001); Sørlie et al. (2003); Gruvberger et al. (2001); van't Veer et al. (2002); van de Vijver et al. (2002); Ahr et al. (2002); Huang et al. (2003); Sotiriou et al. (2003); Woelfle et al. (2003); Ma et al. (2003); Ramaswamy et al. (2003); Chang et al. (2003); Sotiriou et al. (2003); 和 Hedenfalk et al. (2001)。

当前在 LNN 患者中, 在外科手术切除原发肿瘤之后应用辅助治疗、以及应用何种类型 (内分泌的和/或化疗) 的决定, 主要取决于患者的年龄、绝经状态、肿瘤大小、肿瘤级别和甾类激素受体状态。在指导原则例如 St. Gallen 指标和国立卫生研究院 (NIH) 公认指标中这些因素都被考虑了。根据这些指标, 超过 85% -90% 的 LNN 患者将成为接受辅助性全身治疗的候选者。为了指导治疗选项的选择, 明显地需要鉴定更好的预后因素。

发明概述

认识到疾病发展的复杂性, 我们在此报道了基因表达的全面的基

基因组规模的评定来鉴定可广泛应用的预后标记物。Ntzani et al. (2003); 和 Wang et al. (2004)。

本发明包括评定乳腺癌状态的方法，通过从乳腺癌患者获取生物样品；和测量样品中基因的表达水平来进行，所述基因选自编码 mRNA 的那些，所述 mRNA 相应于 SEQ ID NO: 1-111；或由探针集识别，所述探针集选自由表 10 中描述的相应于 SEQ ID NO: 1-111 的 psid 构成的组，其中所述基因表达水平高于或低于预定的截断水平是乳腺癌状态的指示。

本发明包括为乳腺癌分期 (staging breast cancer) 的方法，通过从乳腺癌患者获取生物样品；和测量样品中基因的表达水平来进行，所述基因选自编码 mRNA 的那些，所述 mRNA 相应于 SEQ ID NO: 1-111；或由探针集识别，所述探针集选自由表 10 中描述的相应于 SEQ ID NO: 1-111 的 psid 构成的组，其中所述基因表达水平高于或低于预定的截断水平是乳腺癌分期 (stage) 的指示。

本发明包括确定乳腺癌患者治疗方案的方法，通过从乳腺癌患者获取生物样品；和测量样品中基因的表达水平来进行，所述基因选自编码 mRNA 的那些，所述 mRNA 相应于 SEQ ID NO: 1-111；或由探针集识别，所述探针集选自表 10 中描述的相应于 SEQ ID NO: 1-111 的 psid，其中所述基因表达水平高于或低于预定的截断水平是复发风险的充分指示，以允许医师确定推荐的治疗的程度和类型以防止复发。

本发明包括治疗乳腺癌患者的方法，通过从乳腺癌患者获取生物样品；和测量样品中基因的表达水平来进行，所述基因选自编码 mRNA 的那些，所述 mRNA 相应于 SEQ ID NO: 1-111；或由探针集识别，所述探针集选自表 10 中描述的相应于 SEQ ID NO: 1-111 的 psid，其中所述基因表达水平高于或低于预定的截断水平是表明复发的高风险；如果她们是高风险患者则用辅助治疗来治疗患者。

本发明包括交叉验证乳腺癌患者的预后基因表达分布型的方法，通过从统计学上显著数量的患者生物样品获得基因表达数据；使样品顺序随机化；撇开约 10%-50% 的样品的数据；对余下的样品，对所有变量计算感兴趣的因数并选择满足 p 值截断 (p) 的变量；使用前向搜索来选择符合预测模型的变量，评估训练误差 (training error)

直到它命中预定的误差率；对留出的 10-50% 的样品测试预测模型；除去一套新的样品重复步骤 c-g；继续步骤 c)-g) 直到已经测试了 100% 的样品并记录了分类表现。

本发明包括独立验证乳腺患者的预后基因表达分布型的方法，通过以下方式进行：从统计学上显著数量的患者生物样品获得基因表达数据；使基因表达数据中的来源变异性标准化；对预先选择的所有变量计算感兴趣的因数；对样品测试预测模型并记录分类表现。

本发明包括产生复发危险分值 (Relapse Hazard Score) 以允许进行乳腺癌患者预后的方法，通过从统计学上显著数量的患者生物样品获得基因表达数据；对该数据应用单变量 Cox's 回归分析来获得选定的基因；对具有标准 Cox's 系数的选定基因应用加权表达水平来获得可以用作复发危险分值的预测模型来进行。

本发明包括产生乳腺癌预后患者报告的方法，通过以下来进行：从患者获得生物样品；测量样品的基因表达；对步骤 b 的结果应用复发危险分值；使用步骤 c 中获得的结果来产生报告，从而产生患者报告。

本发明包括至少一个探针集的组合；所述探针集选自：SEQ ID NO: 1-111；或表 10 中描述的相应于 SEQ ID NO: 1-111 的 psid。

本发明包括用于进行分析以确定生物样品中的乳腺癌预后的试剂盒，含有：用于检测基因组合的分离的核酸序列、它们的互补物或其部分的材料，所述基因选自编码 mRNA 的那些，所述 mRNA 相应于 SEQ ID NO: 1-111；或由探针集识别，所述探针集选自表 10 中描述的相应于 SEQ ID NO: 1-111 的 psid。

本发明包括用于评定乳腺癌状态的物品，含有：用于检测基因组合的分离的核酸序列、它们的互补物或其部分的材料，所述基因选自编码 mRNA 的那些，所述 mRNA 相应于 SEQ ID NO: 1-111；或由探针集识别，所述探针集选自表 10 中描述的相应于 SEQ ID NO: 1-111 的 psid。

本发明包括诊断/预后档案 (portfolio)，含有：基因组合的分离的核酸序列、它们的互补物或其部分，所述基因选自编码 mRNA 的那些，所述 mRNA 相应于 SEQ ID NO: 1-111；或由探针集识别，所述探针集选自表 10 中描述的相应于 SEQ ID NO: 1-111 的 psid，其中

所述组合足以表征生物样品中的乳腺癌状态或复发风险。

附图的简要说明

附图 1 描述了样品选择的分布型 (profile), 用于患有淋巴结阴性的 (LNN) 乳腺癌的 286 患者的基因表达数据的分析和无监督聚类分析。(A)。选择用于分析的患者样品的流程图。ER 状态被用于确定患者亚组。为了选择标记物然后对每个亚组单独地分析。亚组中的患者被分配成训练集或测试集。将从每个亚组选出的标记物组合以形成单个标志, 来预测测试集中所有患者总体上的肿瘤复发。(B) 286 位 LNN 乳腺癌患者的基因表达数据的无监督聚类分析。左边的画面是 17,819 个信息基因的视图。红色表示高的相对表达; 绿色表示相对低的表达。每一列是一种样品, 每一行是一种基因。右边的画面是从左侧视图鉴定的、在两个患者亚组之间具有强烈差异表达的两个显性基因簇的放大的视图。在 ER 阳性亚组中上部的基因簇具有一组 282 个减量调节的基因。在 ER 阳性亚组中底部基因簇被表示为一组 339 个增量调节的基因。每个树状图底部的标记条表示由常规分析测量的患者 ER 状态。

附图 2. 对无远处转移 (distant-metastasis-free, DMFS) 和总体存活率 (OS) 建立 76 基因分布型和 Kaplan-Meier 分析。A. 选择用于预后标志的基因。使用 ROC 分析从两个训练集选择 ER 阳性和 ER 阴性组的基因标记物。60 个基因来自 ER 阳性组, 16 个基因来自 ER 阴性组(左), 来自测试集中 171 位患者的分析的、76-基因标志的 ROC 曲线(右)。B. 171 位 LNN 患者的验证集中的 DMFS 分析(左)和 OS 分析(右)。根据 76-基因标志评定每个患者的失败风险, 通过训练集确定阈值。log rank test 被用于测试差异。

附图 3. LNN 患者的亚组中的 DMFS 和 OS 分析。84 位绝经前患者 (A)、87 位绝经后患者 (B) 和 79 位具有 10-20mm 大小范围的肿瘤的患者 (C) 中的 DMFS (左) 和 OS (右)。结果来自验证集中的独立的患者。根据 76-基因标志评定每个患者的复发风险, 通过训练集确定阈值。Log rank 检验被用于测试差异。

附图 4 描述了根据 5121 个基因的分级聚类。

附图 5 是描述 21 个对照基因的表达水平的柱形图 (表 7)。

附图 6 描述了数据分析工作流程。

附图 7 描述了用过滤的基因集的 PCA 分析。

附图 8 是描述用于分配患者亚组的 ER 状态的饼分图。通过 Student T-测验来定义 LCM 和大块组织 (bulk tissue) 样品中 ER 阳性和 ER 阴性的子聚类之间差异表达的基因。

附图 9 是一系列柱形图，描述了通过 Gene Ontology 对 LCM 样品中、仅在大块组织中，和对于 LCM 和大块组织是相同的那些样品中仅与 ER 相关的基因的途径分析的结果。

附图 10 显示了独立的验证的结果。附图 10 中的流程图显示了 132 位患者是在四个不同的来源收集的。根据 76-基因标志的表达水平对每个患者计算复发危险分值。将患者分成良好和不良结局组。

附图 11 描述了 A.用于验证研究的患者和它们的肿瘤的临床和病理特征和 B.验证研究中 76-基因标志的 ROC 曲线。

附图 12 描述了在验证研究中 132 位患者的分类结果。预测良好和不良结局组的 Kaplan-Meier 存活曲线和 log rank 检验在图表中显示。

附图 13 描述了途径分析的结果。附图 13 中的流程图显示了用来选择用于统计分析的样品的方法。A. 将 286 位患者随机分成 115 位患者的训练集以及测试集。B. 将 286 位患者随机分成 80% 患者的训练集和 20% 患者的测试集。训练集被用于选择基因标记物和构建预后标志。测试集用于进行验证。两个过程都重复 10 次。

附图 14 描述了 GO 本体论分析的概要。每个柱形代表 76-基因标志中显著地过度呈现的途径 ($p < 0.05$)。根据可选择标志的结果计算标准偏差。A. 使用 115 位患者的训练集的 10 个标志的结果。B. 使用 80% 患者的训练集的 10 个标志的结果。

详细说明

本发明包括评定乳腺癌状态的方法，通过从乳腺癌患者获取生物样品；和测量样品中基因的表达水平来进行，所述基因选自编码 mRNA 的那些，所述 mRNA 相应于 SEQ ID NO: 1-111；或由探针集识别，所述探针集选自由表 10 中描述的相应于 SEQ ID NO: 1-111 的 psid 构成的组，其中所述基因表达水平高于或低于预定的截断水平 (cut-off level) 指示乳腺癌状态。

本发明包括为乳腺癌分期的方法，通过从乳腺癌患者获取生物样品；和测量样品中基因的表达水平来进行，所述基因选自编码 mRNA 的那些，所述 mRNA 相应于 SEQ ID NO: 1-111；或由探针集识别，所述探针集选自由表 10 中描述的相应于 SEQ ID NO: 1-111 的 psid 构成的组，其中所述基因表达水平高于或低于预定的截断水平指示乳腺癌分期。该方法利用了任何本领域已知的分类法，包括 TNM 系统美国癌症联合委员会 (American Joint Committee on Cancer www.cancerstaging.org)，以及与相应于具有类似基因表达分布型的患者的阶段的比较。

本发明包括确定乳腺癌患者治疗方案的方法，通过从乳腺癌患者获取生物样品；和测量样品中基因的表达水平来进行，所述基因选自编码 mRNA 的那些，所述 mRNA 相应于 SEQ ID NO: 1-111；或由探针集识别，所述探针集选自表 10 中描述的相应于 SEQ ID NO: 1-111 的 psid，其中所述基因表达水平高于或低于预定的截断水平是复发风险的充分指示，以允许医师确定推荐的治疗的程度和类型以防止复发。

本发明包括治疗乳腺癌患者的方法，通过从乳腺癌患者获取生物样品；和测量样品中基因的表达水平来进行，所述基因选自编码 mRNA 的那些，所述 mRNA 相应于 SEQ ID NO: 1-111；或由探针集识别，所述探针集选自表 10 中描述的相应于 SEQ ID NO: 1-111 的 psid，其中所述基因表达水平高于或低于预定的截断水平是表明复发的高风险；如果她们是高风险患者则用辅助治疗来治疗患者。

在任何上述方法中，SEQ ID NO 可以是 1-35、36-95、96-111 或 36-111。优选的，SEQ ID NO 对于雌激素受体 (ER) 患者是 36-95，对于 ER 阴性患者是 96-111。

样品可以通过本领域已知的任何方法来制备，包括大块组织制备和激光俘获显微解剖。大块组织制备物可以从活组织检查或外科样本获得。

上述的方法可以进一步包括测量至少一个编码 mRNA 的基因的表达水平，所述 mRNA 相应于 SEQ ID NO: 112-132；或被探针集识别，所述探针集选自表 10 中描述的相应于 SEQ ID NO: 112-132 的 psid。它们还可以包括测量在样品中组成性表达的至少一个基因的表

达水平。

上述的方法可以进一步包括确定样品的雌激素受体 (ER) 状态。ER 状态可以通过本领域已知的任何方法来测量, 包括测量指示 ER 状态的至少一个基因的表达水平, 测量样品中 ER 的存在和免疫组织化学测量。

上述的方法可以用于来自任何生物学来源的样品。优选的, 所述样品是从原发性肿瘤获得。

上述方法优选的具有至少 40% 的特异性和至少 90% 的灵敏度。

当基因的表达模式与指示复发患者的表达模式相比较时可以使用上述方法。所述比较可以通过本领域已知的任何方法, 包括表达模式的比较通过模式识别方法来进行。模式识别方法可以是本领域已知的任何方法, 包括 Cox's 比例危害分析。

相对于良性细胞或正常组织, 当预定的截断水平在样品中是至少 1.5 倍过量或减量表达时, 可以使用上述方法。优选的, 相对于良性细胞或正常组织, 在具有转移性细胞的样品中预定的截断水平至少具有统计学上显著的 p 值过量表达。更优选的, 所述 p 值低于 0.05。

当在微阵列或基因芯片上测量基因表达时可以使用上述方法。适合于在此使用的基因芯片和微阵列也包括在本发明中。所述微阵列可以是 cDNA 阵列或寡核苷酸阵列, 可以进一步含有一种或多种内部对照试剂。

当对从样品提取的 RNA 通过聚合酶链式反应 (PCR) 进行核酸扩增来测定基因表达时, 可以使用上述方法。所述 PCR 可以是逆转录聚合酶链式反应 (RT-PCR)。所述 RT-PCR 可以进一步含有一种或多种内部对照试剂。

当通过测量或检测由基因编码的蛋白来检测基因表达时, 可以使用上述方法。可以使用本领域已知的任何方法, 包括通过特异于所述蛋白的抗体来检测和测量所述基因的特征。适合的特征非限制性地包括 DNA 扩增、甲基化、突变和等位变异。

本发明包括交叉验证乳腺癌患者的预后基因表达分布型的方法, 通过下述方式来进行: 从统计学上显著数量的患者生物样品获得基因表达数据; 使样品顺序随机化; 撇开约 10% -50% 的样品的数据; 对余下的样品, 对所有变量计算感兴趣的因数并选择满足 p 值截断 (p)

的变量；使用前向搜索来选择符合预测模型的变量，评估训练误差直到它命中预定的误差率；对留出的 10-50% 的样品测试预测模型；除去一套新的样品重复步骤 c-g；继续步骤 c) -g) 直到已经测试了 100% 的样品并记录了分类结果。

可以使用所述交叉验证方法，其中步骤 h 中获得的基因表达数据由选自编码 mRNA 的那些基因的基因来表示，所述 mRNA：相应于 SEQ ID No: 3、4、10、18、37、40、42、43、45、55、58、64、67、72-74、76、81、85-86、89、97、100-101、110-111、125 和 132-442，或 2、3、5、12、20、25、36、37、39、40、41、42、43、45、46、51、52、53、54、55、57、58、59、60、61、62、63、64、66、67、69、70、72、73、74、76、80、81、83、84、85、86、87、88、89、90、94、97、98、101、102、104、107、110、111、132、139、142、150、151、153、154、158、161、163、167、170、171、173、175、181、182、183、186、188、190、192、204、206、207、212、215、218、221、223、225、228、231、232、236、238、239、240、241、242、243、246、248、249、255、257、267、269、270、271、273、280、281、282、288、290、291、299、301、304、306、311、314、315、318、327、328、338、342、346、348、354、366、368、371、375、385、388、391、395、397、402、405、409、422、424、428、429、435、436、440、651 和 443-650；或由探针集识别，所述探针集选自由表 10 中描述的相应于 SEQ ID NO: 3、4、10、18、37、40、42、43、45、55、58、64、67、72-74、76、81、85-86、89、97、100-101、110-111、125 和 32-442，或 2、3、5、12、20、25、36、37、39、40、41、42、43、45、46、51、52、53、54、55、57、58、59、60、61、62、63、64、66、67、69、70、72、73、74、76、80、81、83、84、85、86、87、88、89、90、94、97、98、101、102、104、107、110、111、132、139、142、150、151、153、154、158、161、163、167、170、171、173、175、181、182、183、186、188、190、192、204、206、207、212、215、218、221、223、225、228、231、232、236、238、239、240、241、242、243、246、248、249、255、257、267、269、270、271、273、280、281、282、288、290、291、299、301、304、306、311、314、315、318、327、328、338、342、346、348、354、366、

368、371、375、385、388、391、395、397、402、405、409、422、424、428、429、435、436、440、651 和 443-650 的 psid 构成的组。

本发明包括独立验证乳腺患者的预后基因表达分布型 (gene expression profile) 的方法, 通过从统计学上显著数量的患者生物样品获得基因表达数据; 使基因表达数据中的来源变异性标准化; 对预先选择的所有变量计算感兴趣的因数; 对样品测试预测模型并记录分类表现来进行。

可以使用所述独立验证方法, 其中步骤 d. 中获得的基因表达数据由选自编码 mRNA 的那些基因的基因来表示, 所述 mRNA: 相应于 SEQ ID NO: 3、4、10、18、37、40、42、43、45、55、58、64、67、72-74、76、81、85-86、89、97、100-101、110-111、125 和 132-442, 或 2、3、5、12、20、25、36、37、39、40、41、42、43、45、46、51、52、53、54、55、57、58、59、60、61、62、63、64、66、67、69、70、72、73、74、76、80、81、83、84、85、86、87、88、89、90、94、97、98、101、102、104、107、110、111、132、139、142、150、151、153、154、158、161、163、167、170、171、173、175、181、182、183、186、188、190、192、204、206、207、212、215、218、221、223、225、228、231、232、236、238、239、240、241、242、243、246、248、249、255、257、267、269、270、271、273、280、281、282、288、290、291、299、301、304、306、311、314、315、318、327、328、338、342、346、348、354、366、368、371、375、385、388、391、395、397、402、405、409、422、424、428、429、435、436、440、651 和 443-650; 或由探针集识别, 所述探针集选自表 10 中描述的相应于 SEQ ID NO: 3、4、10、18、37、40、42、43、45、55、58、64、67、72-74、76、81、85-86、89、97、100-101、110-111、125 和 132-442, 或 2、3、5、12、20、25、36、37、39、40、41、42、43、45、46、51、52、53、54、55、57、58、59、60、61、62、63、64、66、67、69、70、72、73、74、76、80、81、83、84、85、86、87、88、89、90、94、97、98、101、102、104、107、110、111、132、139、142、150、151、153、154、158、161、163、167、170、171、173、175、181、182、183、186、188、190、192、204、206、207、212、215、218、221、223、225、228、231、232、236、238、239、

240、241、242、243、246、248、249、255、257、267、269、270、271、273、280、281、282、288、290、291、299、301、304、306、311、314、315、318、327、328、338、342、346、348、354、366、368、371、375、385、388、391、395、397、402、405、409、422、424、428、429、435、436、440、651 和 443-650 的 psid 构成的组。

本发明包括产生复发危险分值 (Relapse Hazard Score) 以允许乳腺癌患者预后的方法, 通过从统计学上显著数量的患者生物样品获得基因表达数据; 对该数据应用单变量 Cox's 回归分析来获得选定的基因; 对具有标准 Cox's 系数的选定基因应用加权表达水平来获得可以用作复发危险分值的预测模型来进行。

复发危险分值可以通过以下公式获得:

$$\text{复发分值} = A \cdot I + \sum_{i=1}^{60} I \cdot w_i x_i + B \cdot (1-I) + \sum_{j=1}^{16} (1-I) \cdot w_j x_j$$

其中

$$I = \begin{cases} 1 & \text{如果X基因水平} > 10 \\ 0 & \text{如果X基因水平} \leq 10 \end{cases}$$

A 和 B 是常数

w_i 是 X 基因 + 标记物的标准化的 Cox's 回归系数

x_i 是按 log2 尺度的 ER + 标记物的表达值

w_j 是 X 基因 - 标记物的标准化的 Cox's 回归系数

x_j 是按 log2 尺度的 ER - 标记物的表达值

X 基因选自自由编码 mRNA 的那些构成的组, 所述 mRNA:

- i. 相应于 SEQ ID NO: 1-111; 或
- ii. 由探针集识别, 所述探针集选自表 10 中描述的相应于 SEQ ID NO: 1-111 的 psid 构成的组。

本发明包括产生乳腺癌预后患者报告的方法, 通过以下方式进行: 从患者获得生物样品; 测量样品的基因表达; 对步骤 b 的结果应用复发危险分值; 使用步骤 c 中获得的结果来产生报告, 从而产生患

者报告。所述报告可以含有相对于患者群体而言的患者结局和/或风险概率的评定。

本发明包括至少一个探针集的组合，所述探针集选自：SEQ ID NO: 1-111；或表 10 中描述的相应于 SEQ ID NO: 1-111 的 psid。所述组合物可以含有至少一个选自 SEQ ID NO: 36-95；96-111；或 36-111 的探针集。所述组合物可以进一步含有用于进行微阵列分析的试剂，和介质，通过所述介质分析所述核酸序列、它们的互补物或其部分。

本发明包括用于进行分析以确定生物样品中的乳腺癌预后的试剂盒，含有：用于检测基因组合的分离的核酸序列、它们的互补物或其部分的材料，所述基因选自编码 mRNA 的那些，所述 mRNA 相应于 SEQ ID NO: 1-111；或由探针集识别，所述探针集选自表 10 中描述的相应于 SEQ ID NO: 1-111 的 psid。所述试剂盒可以进一步含有用于进行微阵列分析的试剂，和介质，通过所述介质分析所述核酸序列、它们的互补物或其部分。

本发明包括用于评定乳腺癌状态的物品，含有：用于检测基因组合的分离的核酸序列、它们的互补物或其部分的材料，所述基因选自编码 mRNA 的那些，所述 mRNA 相应于 SEQ ID NO: 1-111；或由探针集识别，所述探针集选自表 10 中描述的相应于 SEQ ID NO: 1-111 的 psid。所述物品可以进一步含有用于进行微阵列分析的试剂，和介质，通过所述介质分析所述核酸序列、它们的互补物或其部分。

在此提供的微阵列可以含有：基因组合的分离的核酸序列、它们的互补物或其部分，所述基因选自编码 mRNA 的那些，所述 mRNA 相应于 SEQ ID NO: 1-111；或由探针集识别，所述探针集选自表 10 中描述的相应于 SEQ ID NO: 1-111 的 psid，其中所述组合足以表征生物样品中的乳腺癌状态或复发风险。

所述微阵列测量或表征优选的是至少 1.5 倍过量或减量表达。所述微阵列可以检测预定的截断水平，相对于良性细胞或正常组织，所述预定的截断水平是在样品中至少 1.5 倍过量或减量表达。优选的，相对于良性细胞或正常组织，在具有转移性细胞的样品中预定的截断水平至少具有统计学上显著的 p 值过量表达。更优选的，所述 p 值低于 0.05。

本发明包括诊断/预后档案 (portfolio), 含有: 基因组合的分离的核酸序列、它们的互补物或其部分, 所述基因选自编码 mRNA 的那些, 所述 mRNA 相应于 SEQ ID NO: 1-111; 或由探针集识别, 所述探针集选自表 10 中描述的相应于 SEQ ID NO: 1-111 的 psid, 其中所述组合足以表征生物样品中的乳腺癌状态或复发风险。所述档案可以检测预定的截断水平, 相对于良性细胞或正常组织, 所述预定的截断水平是在样品中至少 1.5 倍过量或减量表达。优选的, 相对于良性细胞或正常组织, 在具有转移性细胞的样品中预定的截断水平至少具有统计学上显著的 p 值过量表达。更优选的, 所述 p 值低于 0.05。

SEQ ID NO: 1-650 在表 10 中概述。在 SEQ ID NO: 1-650 的每一个中, 标记物通过 psid 或 Affymetrix 基因芯片名称来标识, 代表相应于给定 SEQ ID NO、编码任何变体、等位基因等的基因。标记物还被定义为编码 mRNA 的基因, 所述 mRNA 由相应于给定 psid 的探针识别。以下更详细地描述一些标记物。

M83, SEQ ID NO: 45 被美国专利申请公开号 20020110547; 20040009491; 和 20030236392, 和 PCT 公开号 WO0149716; WO0170979; WO03025138; WO03042661; WO03016475; WO2004063355; WO2004047728 和 WO2004030615 提及。

ABLIM-s, SEQ ID NO: 43 被美国专利申请公开号 20020131971; 20030087818; 20030166064; 20030236392; 和 20040005560 和 PCT 公开号 WO0036107; WO0159063; WO0160860; WO0174405; WO0177288; WO02059271; WO0212328; WO0224956; WO0229086; WO03025138; WO2004024097; WO2004037996; WO2004039956; WO2004043361 和 WO2004047728 提及。

ATAD2, SEQ ID NO: 54 被美国专利申请公开号 20020064872; 20020085998; 20020150581; 20020156011; 20020192678; 20030069180; 20030073623; 20030104366; 20040005560; 20040013663; 和 20040058340 和 PCT 公开号 WO0060076; WO0142467; WO0155322; WO0160860; WO0170979; WO0179286; WO0196388; WO02057414; WO02059377; WO02060317; WO02086443; WO0231198; WO0250279; WO03004989; WO03025138; WO03042661; WO03062379; WO2004047728;

WO2004048938; WO2004048938; WO2004063355 和 WO9938972 提及。

C11ORF9, SEQ ID NO: 48 被美国专利申请公开号 20030065157; 和 20030073623 和 PCT 公开号 WO0056880; WO0058473; WO0159063; WO0174405; WO02059260; WO0229086; WO03101283; WO2004031413 和 WO2004060270 提及。

C3, SEQ ID NO: 46 被美国专利申请公开号 20040005560 和 20040058340 和 PCT 公开号 WO9101999; WO0055350; WO0160860; WO0175067; WO0188088; WO0224956; WO0294629; WO03078572; WO2004028479; WO2004030615; WO2004041170 和 WO2004060270 提及。

CGI-41, SEQ ID NO: 65 被美国专利申请公开号 20030073623 和 20030096982 和 PCT 公开号 WO9514772; WO9940100; WO0154472; WO0157188; WO03046152; WO2004060270 提及。

CLN8, SEQ ID NO: 36 被 EP 1 104 808 B1 美国专利申请号 20030073623 和 20040009491 和 PCT 公开号 WO9845437; WO9935158; WO0142451; WO0190304; WO02090526; WO02095010; WO02102993; WO02102994; WO03004622; WO04024892 和 WO04041170 提及。

CNK1, SEQ ID NO: 60 被美国专利申请公开号 20040058340 和 PCT 公开号 WO9938972 和 WO2004037996 提及。

DUSP4, SEQ ID NO: 57 被美国专利申请公开号 20030073623; 20030194704; 20040058340 和 20040110194 和 PCT 公开号 WO0100828; WO0204514; WO02059377; WO02103320; WO0224956; WO03004989; WO2004018641; WO2004023973; WO2004037996; WO2004039956; WO2004043361; WO2004047728; WO2004048938 和 WO9423039 提及。

FEN1, SEQ ID NO: 62 被美国专利申请公开号 5874283, 美国专利申请号 20020156011; 20030069180; 20030073623 和 20030194704 和 PCT 公开号 WO0196388; WO03016475; WO03016500; WO03072035; WO03101283; WO2004022059; WO2004030615; WO2004039956; WO2004043361; WO2004047728; WO2004048938;

WO2004060270 和 WO2004079014 提及。

FKBP2, SEQ ID NO: 63 被美国专利申请公开号 20030073623 和 PCT 公开号 WO02059377 ; WO02070737 ; WO02081731 ; WO02083898 ; WO0224956 ; WO03016475 ; WO03054152 ; WO2004039956; WO2004043361 和 WO9219745 提及。

GP73, SEQ ID NO: 37 被美国专利申请公开号 20030022239; 20030073623; 20030236392; 20040009478; 20040034196 和 20040058340 和 PCT 公开号 WO0012708; WO0061610; WO0078961; WO0140466; WO0160860; WO0177168; WO0194629; WO02059260; WO02083876; WO0224956 ; WO03004989 ; WO2004024892 ; WO2004031414 和 WO2004053081 提及。

H4FH, SEQ ID NO: 61 被美国专利申请公开号 20020151681 和 PCT 公开号 WO0055174; WO0175067; WO0224956; WO03016476; WO2004039943 提及。

IL-18, SEQ ID NO: 39 被美国专利号 6060283, 美国专利申请公开号 20040115636 和 PCT 公开号 WO0194629; WO02057414; WO0224956; WO0229086; WO03025138; WO03042661 和 WO9724441 提及。

KIAA0748, SEQ ID NO: 56 被 PCT 公开号 WO0021991; WO02094988 和 WO2004003162。

KPNA2, SEQ ID NO: 64 被美国专利申请公开号 20020151681; 20020156011 ; 20030069180 和 20040058340 和 PCT 公开号 WO9712967; WO9964576; WO0055174; WO0146697; WO0170979; WO0196388 ; WO02057414 ; WO02059377 ; WO02086443 ; WO02102235 ; WO0224956 ; WO03010336 ; WO03025138 ; WO03027285 ; WO03042661 ; WO2004024097 ; WO2004024892 ; WO2004028479; WO2004037996; WO2004039956; WO2004043361; WO2004047728 和 WO2004063355 提及。

MGC11335 , SEQ ID NO : 66 被美国专利申请公开号 20030073623 ; 20030219744 和 20030236392 和 PCT 公开号 WO0144448 ; WO0153312 ; WO0157182 ; WO0188088 和 WO2004030615 提及。

OR12D2, SEQ ID NO: 52 被 PCT 公开号 WO0127158; WO02068652; WO02081498; WO02081517; WO030000735 和 WO03091388 提及。

ORC3, SEQ ID NO: 53 被美国专利申请公开号 20040013663 和 PCT 公开号 WO9810067; WO9958642; WO0060078; WO0175067; WO02059377; WO02070737; WO02102235 和 WO03042661 提及。

PLK1, SEQ ID NO: 59 被 DE4329177, 美国专利申请公开号 20020081659 和 20020151681 和 PCT 公开号 WO0055174; WO0055320; WO02070737; WO02086443; WO0224956; WO03003906; WO03016476; WO03018807; WO03025138; WO03042661; WO2004030615; WO2004037996; WO2004042022; WO2004047728; WO2004048938; WO2004063355 和 WO2004070062 提及。

PPP1CC, SEQ ID NO: 42 被美国专利申请公开号 20030073623 和 20040013663 和 PCT 公开号 WO03016476; WO0063438; WO0170979; WO0202623; WO02036766; WO02074237; WO0228999; WO03016475 和 WO2004030615 提及。

SMC4, SEQ ID NO: 55 被美国专利申请公开号 20020123619; 20020168637 和 20040058340 和 PCT 公开号 WO9938972; WO9964576; WO0175067; WO0190154; WO0204514; WO02059377; WO02074237; WO02086443; WO02102235; WO03003906; WO03016475; WO03025138; WO03042661; WO2003094848; WO2004024892; WO2004047728; WO2004048938; WO2004063355 和 WO2004074301 提及。

TRDL-1, SEQ ID NO: 44 被美国专利号 6171787 和 6440694, 美国专利申请公开号 20020055474; 20020072089; 20020081659; 20030065157; 20030073623; 和 20030219767 和 PCT 公开号 WO9733902; WO9912965; WO9926976; WO9928462; WO9933980; WO9935170; WO9950416; WO9954460; WO0026244; WO0032776; WO0055320; WO0125256; WO0177291; WO2004020593; WO2004024892; WO2004033637 和 04037996 提及。

Yif1p, SEQ ID NO: 38 被美国专利申请公开号 20020131971;

20030166064；和 20040037842 和 PCT 公开号 WO9933981；WO9947540；WO0173027；WO02060317；WO02070737；WO02076488；WO02095010；WO0224956；WO03004622；WO03038063；WO03042661；WO03054152；WO03064589；WO2004024892；WO2004037996；WO2004047728 和 WO2004048938 提及。

利用来自没有接受辅助性全身治疗的 286 位淋巴结阴性 (LNN) 乳腺癌患者的冷冻肿瘤样品的总 RNA，使用 Affymetrix Human U133a 基因芯片分析了 22,000 个转录产物的表达。我们发现，基因表达的基因组范围的测量可以鉴定基因活性的模式，其将肿瘤再分类，并提供了在患有淋巴结阴性乳腺癌的患者中改进的个体风险评估方法。

在 115 个肿瘤的训练集中，我们鉴定了包含 ER 阳性患者的 60 个基因和 ER 阴性患者的 16 个基因的 76 个基因标志 (signature)。在随后的 171 位 LNN 患者的独立测试集中，这种标志显示了 93% 的敏感性和 48% 的特异性。在鉴定 5 年内发生远处转移的患者时，即使对多变量分析中传统的预后因素进行了修正 (HR: 5.55, CI: 2.46–12.5)，基因分布型仍是高度信息性的 (危险比例, HR: 5.67, 95% 置信区间, CI: 2.59–12.4)。76-基因分布型还代表了在 84 位绝经前患者 (HR: 9.60, $p = 1.6 \times 10^{-4}$)、87 位绝经后患者 (HR: 4.04, $p = 1.7 \times 10^{-3}$) 以及疾病预后特别困难的 79 位具有 10 到 20mm 大小的肿瘤的患者 (HR: 14.1, $p = 2.3 \times 10^{-6}$) 的亚组中发生转移的强预后因素。

在这项研究中，我们提供了来自所有年龄和肿瘤尺寸组的 286 位 LNN 乳腺癌患者的原发肿瘤分析结果。患者没有接受辅助性全身治疗，使得这成为第一种多基因预后评定，不含有由与全身性治疗有关的预测因素带来的潜在混淆性贡献，通过独立的测试集适当地验证，对肿瘤大小和年龄没有限制。所描述的基于基因表达的算法可以高置信度地在 LNN 患者中预测发生远处复发的概率。

在组织样品中单纯地存在或缺乏特定的核酸序列仅极少地被发现有诊断或预后价值。另一方面，关于各种蛋白、肽或 mRNA 的表达的信息被越来越认为是重要的。在基因组内单纯地存在具有表达蛋白、肽或 mRNA 的潜力的核酸序列 (这种序列被称为基因) 本身对于蛋白、肽或 mRNA 是否在给定细胞中表达不是决定性的。能表达

蛋白、肽或 mRNA 的给定基因是否进行表达、或如果表达的话这种表达发生到什么程度，是由各种复杂的因素决定的。不考虑在了解和评定这些因素方面的难度，分析基因表达可以提供关于发生重要事件的有用信息，所述事件例如肿瘤发生、转移、细胞凋亡和其他临床上相关的现象。在基因表达分布型中可以发现基因的活性或无活性所达到的程度的相关指示。本发明的基因表达分布型被用于提供预后和治疗乳腺癌患者。

样品制备需要收集患者样品。用于该发明性方法的患者样品是被怀疑含有患病细胞如上皮细胞、取自乳腺样品中原发肿瘤的那些。取自手术边缘 (surgical margins) 的样品也是优选的。然而，最优选的，样品取自乳腺癌手术获得的淋巴结。激光俘获显微解剖 (LCM) 技术是选择要研究的细胞、最小化由细胞类型异质性引起的变异的一种方法。因而，可以容易地检测正常细胞和癌细胞之间基因表达方面的中度或微小的变化。样品也可以包含取自外周血的循环上皮细胞。可以根据许多方法来获得这些，但最优选的方法是美国专利 6,136,182 中描述的磁分离技术。一旦获得了含有目标细胞的样品，提取 RNA 并进行扩增，对于在适合的档案中的基因，优选的通过微阵列来获得基因表达分布型。

确定基因表达分布型的优选的方法包括测定由可编码蛋白或肽的基因产生的 RNA 的数量。这可以通过逆转录酶 PCR (RT-PCR)、竞争性 RT-PCR、实时 RT-PCR、差异展示 RT-PCR、Northern 印迹分析和其他相关的测试来实现。虽然可能使用单独的 PCR 反应来进行这些技术，但最好是扩增由 mRNA 产生的互补 DNA (cDNA) 或互补 RNA (cRNA)，并通过微阵列来分析。许多不同的阵列结构和它们的产生方法是本领域的技术人员已知的，已在美国专利中描述，例如：5,445,934；5,532,128；5,556,752；5,242,974；5,384,261；5,405,783；5,412,087；5,424,186；5,429,807；5,436,327；5,472,672；5,527,681；5,529,756；5,545,531；5,554,501；5,561,071；5,571,639；5,593,839；5,599,695；5,624,711；5,658,734 和 5,700,637。

微阵列技术允许同时测量数千个基因的稳态 mRNA 水平，从而提供了鉴定不受控细胞增殖的影响，例如发作、延滞或调节的有效方法。两种微阵列技术是当前广泛使用的。第一种是 cDNA 阵列，第二

种是寡核苷酸阵列。尽管在这些芯片的结构方面存在差别，基本上所有的下游数据分析和输出是相同的。这些分析的结果一般是来自标记的探针的信号强度的度量，所述探针被用于检测样品中与微阵列的已知位置上的核酸序列杂交的 cDNA 序列。一般地，信号的强度与 cDNA 的数量成比例，因而与样品细胞中表达的 mRNA 成比例。许多这样的技术是可用的和有用的。测定基因表达的优选的方法可以在美国专利 6,271,002; 6,218,122; 6,218,114; 和 6,004,755 中找到。

通过比较信号强度进行表达水平的分析。通过生成测试样品中的基因相对于对照样品中的基因的表达强度的比例矩阵 (ratio matrix)，来最好地进行。例如，可以将来自患病组织的基因表达强度与产自相同类型的正常组织的表达强度相比较（例如，患病的乳腺组织样品相对于正常乳腺组织样品）。这些表达强度的比例表明了测试样品和对照样品之间基因表达方面的倍数变化。

也可以按许多种方式展现基因表达分布型。最普遍的方法是将未加工的荧光强度或比率矩阵编排到图形化树形图中，其中各个列代表测试样品，各个行代表基因。排列数据使得具有类似的表达分布型的基因相互接近。每个基因的表达比例用一种颜色显现。例如，小于一的比例（表明减量调节）可以在谱图的蓝色部分中显现，而大于一的比例（表明增量调节）可以作为一种颜色在谱图的红色部分显现。商业上可获得的计算机程序是可用的，来显示这样的数据，包括来自 Agilent Technologies 的 GeneSpring 和来自 Partek® 的 Partek Discover™ 和 Partek Infer™ 软件。

在实施例中描述了在本发明的方法中使用的调节基因。相对于没有复发的患者，在有乳腺癌复发的患者中差别表达的基因不是增量调节就是减量调节的。增量调节和减量调节是相对的术语，意思是相对于某些基准线在基因的表达数量中存在的可检测的差异（超过用于测量它的系统中噪声的贡献）。在这种情况下，基准线是测量的非复发患者的基因表达。使用相同的测量方法，相对于所述基准线，在患病的细胞（来自复发患者）中的目标基因是增量或减量调节的。在此，患病的是指身体状态的改变，其阻断或干扰、或有可能干扰身体的正常机能，如同细胞的不受控增殖那样。当一个人的基因型或表型的某个方面与疾病的存在相一致时，他被诊断为患有疾病。然而，进行诊

断或预后的行动包括疾病/状态问题的确定，例如确定复发的可能性和治疗监视。在治疗监视中，通过比较一段时间的基因表达来确定基因表达分布型是否已经改变或正在改变成更为符合正常组织的模式，来针对给定的治疗过程的效果进行临床判断。

优选的，根据杂交的微阵列探针的强度测定的倍数变化对增量或减量调节的水平进行辨别。对于进行这样的辨别，2.0 倍的差异是优选的（或 p 值小于 0.05）。也就是说，在将基因说成是在患病/复发细胞与正常/非复发细胞中是差别表达的之前，与标准细胞相比，发现患病细胞产生至少 2 倍高或低至 1/2 的强度。倍数差异越大，该基因作为诊断或预后工具的用途越是优选的。为本发明的基因表达分布型选择的基因具有引起信号产生的表达水平，通过超出所使用的临床实验室检测设备的背景的一定量，该信号与正常或未调节基因的信号是可区分的。

统计数值可以用来确信地将调节的基因与未调节基因和噪声辨别开来。统计试验找出在各组样品之间最显著不同的基因。Student's t-检验是可以用于找出两个组之间的显著差异的强有力的统计检验的实例。p 值越低，在不同的组之间基因显示出差异的证据就越强。尽管如此，由于微阵列同时测量超过一个的基因，可以一次进行数万个统计检验。因此，人们不太可能仅是偶然地看见小的 p 值，可以利用 Sidak 校正对这进行调整以及进行随机化/置换实验。t-检验的 p 值低于 0.05 是该基因显著不同的证据。更强的证据是在考虑 Sidak 校正之后 p 值小于 0.05。对于每个组中的许多样品，在随机化/置换试验之后 p 值小于 0.05 是显著差异的最强证据。

可以用来选择基因的另一个参数是使用绝对信号差异的检测，所述基因产生的信号大于未调节基因的信号或噪音。优选的，由调节的基因表达产生的信号至少 20% 不同于正常的或未调节基因的信号（在绝对值的基础上）。再更优选的是，这种基因产生的表达模式至少 30% 不同于正常的或未调节基因的表达模式。

基因可以被分组，从而获得的关于该组中基因集的信息提供了用于进行临床上相关的判断，例如诊断、预后或治疗选择的可靠的基础。这些基因集（sets of genes）构成了本发明的档案。在这种情况下，由所述档案支持的判断包括乳腺癌和它复发的机会。同最具诊断性的标

记物一样，常常希望的是使用足以进行正确医学判断的最少数量的标记物。这防止了在等待进一步分析的治疗中的延迟以及不适当的时间和资源的耗费。

优选的，建立档案，从而在档案中基因的组合相对于单独的基因或随机选择的基因组合展现出改善的敏感性和特异性。在本发明的上下文中，档案的敏感性可以反映在由患病状态中基因表达所展现的、相对正常状态的倍数差异中。特异性可以反映在对基因表达的信号与感兴趣状况的相关性的统计度量中。例如，标准误差可以用作这种度量。在考虑用于包含在档案中的一组基因时，在表达度量中的小的标准差与更高的特异性相关。也可以使用变异的其他度量，例如相关系数。

建立基因表达档案的一种方法是通过使用优化的算法，例如在建立股票档案 (stock portfolios) 中广泛使用的平均变异算法。在美国专利公开号 20030194734 中详细地描述了该方法。基本上，该方法要求建立将会优化人们为使用它而接收到的返回 (例如，产生的信号) 的一组输入 (金融应用中的股票，在此为由强度来度量的表达)，而最小化返回的变异性。对于进行这种操作，许多商业的软件程序是可获得的。“Wagner Associates Mean-Variance Optimization Application”，在本说明书中称为“Wagner Software”，是优选的。该软件使用来自“Wagner Associates Mean-Variance Optimization Library”的函数来确定效率界限 (efficient frontier)，在 Markowitz 意义上的最佳档案是优选的。使用这类软件需要转化微阵列数据，从而可以将其作为输入来处理，如同该软件被用于预定的金融分析目的时使用股票返回 (stock return) 和风险度量的方式。

选择档案的过程也可以包括试探规则的应用。优选的，根据生物学和对用来产生临床结果的技术的理解来制定这种规则。更优选的，将它们应用于来自优化方法的输出。例如，档案选择的平均方差法可以应用于微阵列数据，用于在患有乳腺癌的受试者中差异表达的许多基因。该方法的输出将是优化的基因集，其将包括在外周血中以及在患病组织中表达的某些基因。如果用于测试方法的样品是从外周血获得的，并且在乳腺癌的情形中差异表达的某些基因是在外周血中差异表达的，则可以应用试探规则，其中档案选自排除了外周血中差异表

达的那些的效率界限。当然地，例如，通过在数据预选期间应用规则，可以在形成效率界限之前应用规则。

可以使用不一定与讨论的生物学相关的其他试探规则。例如，人们可以应用一种规则，仅档案的指定百分比可以由特定的基因或一组基因表示。商业上可获得的软件例如 Wagner Software 容易地提供了这些类型的试探。例如，当不同于准确度和精确度（例如，预期的许可费用）的因素对包括一个或多个基因的需要性（desirability）具有影响时，这是有用的。

本发明的一个方法包括比较各种的基因（或档案）的基因表达分布型来归类（ascribe）预后。将构成档案的每个基因的基因表达分布型固定在媒介，例如计算机可读媒介上。这可以采用许多形式。例如，可以建立表格，指示疾病的信号范围（例如，强度度量）被输入到该表格中。因而实际的患者数据可以与表格中的数值比较，来确定患者样品是正常还是患病的。在更完善的实施方式中，表达信号（例如，荧光强度）的模式是数字地或图形地记录的。

然后来自与患者样品结合使用的基因档案的基因表达模式可以与表达模式相比较。然后可以使用模式比较软件来确定患者样品是否具有指示疾病复发的模式。当然地，这些比较也可以用来确定患者是否不可能经历疾病复发。然后样品的表达分布型可以与对照细胞的档案相比较。如果样品表达模式与乳腺癌的复发的表达模式相符，则（不存在抵销性的医学考虑）该患者像治疗复发患者一样被治疗。如果样品表达模式与来自正常/对照细胞的表达模式相符，则患者被诊断为乳腺癌阴性。

本发明的优选的分布型是由 SEQ ID NO: 1-35 组成的 35 基因档案，由 SEQ ID NO: 36-95 组成的 60 基因档案，其最好地用于预测 ER 阳性患者，和由 SEQ ID NO: 96-111 组成的 16 基因档案，其最好地用于预测 ER 阴性患者。最优选的，所述档案由 SEQ ID NO: 36-111 组成。这个最优选的档案不管 ER 状态如何，最好地将高复发风险的乳腺癌患者从其他患者中分离出来。一旦鉴定了高风险患者，则他们可以用辅助治疗来治疗。

在本发明中，用于分析患者的基因表达模式来确定乳腺癌的预后的最优选的方法是使用 Cox's 危险分析程序。最优选的，使用 S-Plus

软件（可从 Insightful Corporation 商业地获得）进行分析。使用这种方法，将基因表达分布型与确信地代表复发的分布型（即，在分布型中基因组合的表达水平是复发的指示）进行比较。具有建立的阈值的 Cox's 危险模型被用于比较两个分布型（已知的复发对比患者）的相似性，然后确定患者分布型是否超出阈值。如果是，则该患者被分类为将会复发的患者，给予治疗，例如辅助治疗。如果患者分布型没有超过阈值，则他们被分类为不复发患者。也可以使用其他分析工具来回答相同的问题，例如，线性区别分析，逻辑回归和神经网络方法。

模式识别的许多的其他公知的方法是可用的。以下参考文献提供了一些实例：

Weighted Voting: Golub et al. (1999) .

Support Vector Machines: Su et al. (2001); and Ramaswamy et al. (2001) .

K-nearest Neighbors: Ramaswamy (2001) .

Correlation Coefficients: van 't Veer et al. (2002) .

本发明的基因表达分布型也可以连同在癌症诊断、预后或治疗监视中有用的其他非基因性的诊断方法来使用。例如，在某些情况下，有益的是将基于如上所述方法的基因表达的诊断能力与来自常规标记物，例如血清蛋白标记物（例如，癌症抗原 27.29 (“CA 27.29”））的数据组合。现有大量的这种标记物，包括例如 CA 27.29 的被分析物。在一个这样的方法中，从治疗的患者周期性地采取血液，然后进行针对如上所述血清标记物之一的酶免疫测定。当标记物的浓度暗示肿瘤的回归或治疗失败时，采取可经受基因表达分析的样品源。在存在怀疑的肿块处，采取微针吸出物（FNA），然后如上所述分析来自肿块的细胞的基因表达分布型。做为选择，组织样品可以来自邻近于早先除去肿瘤的组织区域。当其他测试产生不明确的结果时这种方法是特别有用的。

本发明的物品包括对于治疗、诊断、预后以及评定疾病有用的基因表达分布型的呈现（representations）。这些分布型呈现被化为可以通过机器自动读取的媒介，例如计算机可读媒介（磁的、光的，等等）。所述物品还可以包括在这种媒介中用于评定基因表达分布型的说明书。例如，所述物品可以包括具有计算机指令的 CD ROM，所述计算

机指令用于比较如上所述的基因档案的基因表达分布型。所述物品还可以具有数字性记录在其中的基因表达分布型，从而它们可以与来自患者样品的基因表达数据相比较。做为选择，所述分布型可以以不同的表现形式记录。图形的记录是一种这样的形式。聚类算法，例如整合在上述来自 Partek® 的 Partek Discover™ 和 Partek Infer™ 软件中的那些，可以最好的帮助这种数据的可视化。

根据本发明的制造物品的不同类型是用于揭示基因表达分布型的介质或格式化的分析。这些可以包括，例如，微阵列，在其中互补物或探针附着到基质上，指示目标基因的序列组合到其上产生它们之存在的可读决定簇。做为选择，根据本发明的物品可以制成试剂盒，用于进行杂交、扩增和信号产生，指示用于检测乳腺癌的目标基因的表达水平。

根据本发明制造的试剂盒包括用于确定基因表达分布型的格式化的分析。这些可以包括进行分析所需的所有或某些材料，例如试剂和说明书。

通过以下非限制性的实施例进一步说明本发明。在此引用的所有参考文献通过引用合并在此。

实施例：根据本发明分析的基因一般与编码蛋白或肽的产生的全长核酸序列有关。本领域的技术人员将认识到，从分析的角度来看，标明全长序列不是必需的。也就是说，可以根据公知的原则选择序列的部分或 EST，对它们可以设计探针来评定相应基因的基因表达。

实施例 1

样品处理和微阵列工作

来自在 1980-1995 年间治疗的、但未用全身性的最新辅助治疗进行治疗的 LNN 患者的冷冻肿瘤样本，选自我们在 Erasmus Medical Center (Rotterdam, Netherlands) 的肿瘤库。所有肿瘤样品被提交给我们来自 25 个地方医院的参考实验室，进行甾类激素受体测量。初级处理的指导原则对于所有医院都是相似的。以避免偏见的方式选择肿瘤。假定 5 年内 25-30%，以及由于质量控制原因肿瘤的实质上的损失，处理了 436 个侵入性肿瘤样品。包括了具有不良、中间和良好临床结局的患者。根据不充足的肿瘤内容 (53 个)、不良的 RNA

质量 (77 个) 或不良的芯片质量 (20 个) 排除了样品, 留下适合于进一步分析的 286 个样品。

研究方案得到医学伦理委员会机构的批准 (MEC no. 02.953)。在手术时患者的中间年龄是 52 岁 (范围, 26-83 岁) (乳腺保留手术: 219 位患者; 修改的根治性乳房切除术: 67 位患者)。根据我们的设立方案给予 248 位患者放疗 (87%)。经历乳腺保留治疗和放疗的患者的比例对于 LNN 疾病是正常的。不考虑放疗状态来包括患者, 因为这项研究目的不是研究特定类型的手术或辅助放疗的潜在效果。此外, 研究显示, 放疗对于远处疾病复发没有明确的影响。早期乳腺癌实验列表 (1995)。淋巴结阴性是基于地方病理学家的病理检查。Foekens et al. (1989a)。

在被包括进来之前, 证实所有 286 个肿瘤样品具有足够的 (>70%) 肿瘤和在 H & E 染色的 5 μ m 冷冻切片中均匀的包含肿瘤。通过配体结合分析或酶免疫测定 (EIA) (Foekens et al. (1989b))、或通过免疫组织化学 (9 个肿瘤) 来测定 ER (和 PgR) 水平。用来将患者分类为 ER 和 PR 的阳性或阴性的截断值是 10fmol/mg 蛋白或 10% 阳性肿瘤细胞。手术后的跟踪包括在第一个 2 年内每 3 个月、第 3 年到第 5 年每 6 个月、和第 5 年开始每 12 个月进行检查。诊断转移的日期被定义为在患者报告症状、检测临床征象之后或在常规的跟踪时确认转移的日期。存活患者 (n = 198) 的中值跟踪周期是 101 个月 (范围, 20-171)。在包括的 286 位患者中, 93 位 (33%) 显示了 5 年内的远处转移的证据, 被当做在无远处转移存活 (DMFS) 分析中的失败。五位患者 (2%) 没有疾病证据而死亡, 在最后的跟踪中审查。83 位患者 (29%) 在先前的复发之后死亡。因此, 在总体存活 (OS) 分析中总共 88 位患者 (31%) 失败。

实施例 2

实施例 1 获得的数据的基因表达分析

用 RNazol B (Campro Scientific, Veenendaal, Netherlands) 从 30 μ m 厚的 (50-100 mg) 20 到 40 个低温恒温切片中分离总 RNA。使用公开的方法 (Affymetrix, CA, Lipshutz et al. (1999)) 制备生物素化的靶点, 与 Affymetrix 寡核苷酸微阵列 U133a 基因芯片杂交。

使用标准的 Affymetrix 方案扫描阵列。每个探针集作为独立的基因处理。使用 Affymetrix GeneChip 分析软件 MAS 5.0 计算表达值。如果平均强度 <40 或如果背景信号 >100, 排除芯片。为标准化芯片信号, 将探针集换算 (scaled) 到 600 的目标强度, 不选择 scale mask files。

实施例 3

实施例 2 中鉴定的基因的统计分析

过滤基因表达数据以包括在两个或多个样品中称为“存在”的基因。17,819 个基因通过了这个过滤, 用于分级聚类。在聚类之前, 每个基因的表达水平除以患者中它的中值表达水平。这个标准化步骤限制了基因表达的数量影响, 在聚类分析中将具有类似表达模式的基因集中在一起。为了鉴定患者亚组, 我们使用 GeneSpring 6.0 对基因和样品都进行了平均联系 (average linkage) 分级聚类。

为了鉴定将发展出远处转移的患者与保持无转移 5 年的患者区别开来的基因, 我们使用了两种指导性的分类预测方法。在第一种方法中, 将 286 位患者分别随机分配到 80 和 206 位患者的训练和测试集。对两个集检查 Kaplan-Meier 存活曲线 (Kaplan et al. (1958)), 以确保没有显著差异和没有偏见被通过随机选择训练和测试集导入。在第二种方法中, 患者被分配到由 ER 状态分开的两个亚组 (附图 1) 之一中。

为了选择标记物对每个亚组单独地分析。在 ER 阳性亚组中的患者被分别随机分配到 80 和 129 位患者的训练和测试集。在 ER 阴性亚组中的患者被分别随机分入 35 和 42 位患者的训练和测试集。为每个亚组训练集选择的标记物被组合以形成单个标志 (signature), 来在随后的独立验证中为 ER 阳性和 ER 阴性患者预测肿瘤转移。

训练集的样品大小由再取样方法确定, 以确保它的统计可信水平。简要地, 在训练集中的患者数目从 15 位患者开始, 每步增加 5 位患者。对于给定的样本大小, 产生随机选择患者的 10 个训练集。从每个训练集构建基因标志, 然后通过分析在作为定义点的 5 年内远处转移的接受者操作特性 (ROC) 曲线、在指定的患者测试集中测试。对于给定的样本大小计算曲线下区域 (AUC) 的平均数和变异系数 (CV)。在平均 AUC 达到平台期和 10 个 AUC 的 CV 低于 5% 的点,

选择训练集所需的患者最小数目。

如下选择基因。首先，使用单变量 Cox's 比例危险回归来鉴定基因，所述基因的表达（在 log2 尺度上）与 DMFS 的长度相关。为了降低多次测试的影响和测试选择的基因的强大程度（robustness），用训练集中患者的自展（bootstrapping）构建 Cox's 模型。Efron et al. (1981)。简要地，构建训练集的 400 个自展样品，每一个具有随机置换选择的 80 位患者。对每一个自展样品（bootstrap sample）运行 Cox's 模型。通过除去顶部和底部 5% p 值，然后将余下的自展 p 值的倒数平均，来为每个基因产生自展分值。这个分值被用于排列基因。为了构建多基因标志，通过根据所述排列的顺序一次添加一个基因，测试基因标记物的组合。进行利用 5 年内的远处转移作为定义点的 ROC 分析，来计算随着基因数目的提高、每个标志的 AUC 下的区域，直到达到最大 AUC 值。

使用复发分值（RS）来计算每个患者的远处转移的风险。该分值被定义为用标准化的 Cox's 回归系数作为加权的、加权的表达信号的线性组合。

$$\text{复发分值} = A \cdot I + \sum_{i=1}^{60} I \cdot w_i x_i + B \cdot (1 - I) + \sum_{j=1}^{16} (1 - I) \cdot w_j x_j$$

其中

$$I = \begin{cases} 1 & \text{如果 ER 水平} > 10 \text{ fmol 每 mg 蛋白} \\ 0 & \text{如果 ER 水平} \leq 10 \text{ fmol 每 mg 蛋白} \end{cases}$$

A 和 B 是常数

w_i 是 ER+ 标记物的标准化的 Cox's 回归系数

x_i 是按 log2 尺度的 ER+ 标记物的表达值

w_j 是 ER- 标记物的标准化的 Cox's 回归系数

x_j 是按 log2 尺度的 ER- 标记物的表达值

从训练集的 ROC 曲线测定阈值以确保 100% 敏感性和最高的特异性。选择常数 A 的值 313.5 和常数 B 的值 280，以对于 ER 阳性和

ER 阴性患者将 RS 的阈值都中心化到零。具有正 RS 分值的患者被分类到不良预后组中，具有负 RS 分值的患者被分类到良好预后组中。在测试集中验证基因标志和截断。使用 Kaplan-Meier 存活率标绘图和 log-rank 检验来评定所预测的高风险和低风险组的远处转移的时间差异。几率比 (OR) 被计算为被预测复发的患者和被预测保持无复发的患者之间远处转移几率的比例。

对有或没有基因标志的单独的临床变量进行使用 Cox's 比例危险回归的单变量和多变量分析。HR 和它的 95% 置信区间 (CI) 来自于这些结果。使用 S-Plus 6.1 软件 (Insightful, VA) 进行所有的统计分析。

实施例 4

实施例 3 中鉴定的基因的途径分析

将功能性分类分配给预后标志基因中的每个基因。使用 Ingenuity 1.0 软件 (Ingenuity Systems, CA) 进行途径分析。Affymetrix 探针被用作输入来查找该软件构建的生物学网络。通过 GO 本体论分类法在一般性功能分类的环境中评定由该程序鉴定的生物学网络。选择和评估预后标志中两个或多个基因的途径。

实施例 5

实施例 1-4 的结果

患者和肿瘤特征

286 位患者的临床和病理的特征在表 1 中概述。

表 1. 患者和他们的肿瘤的临床和病理特征

特征	所有患者 (%)	ER 阳性训练集 (%)	ER 阴性训练集 (%)	验证集 (%)
数目	286	80	35	171
年龄 (平均值 ± SD)	54±12	54±13	54±13	54±12
<40 岁	36 (13)	12 (15)	3 (9)	21 (12)
41-55 岁	129 (45)	30 (38)	17 (49)	82 (48)
56-70 岁	89 (31)	28 (35)	11 (31)	50 (29)
>70 岁	32 (11)	10 (13)	4 (11)	18 (11)
绝经状态				
绝经前	139 (49)	39 (49)	16 (46)	84 (49)
绝经后	147 (51)	41 (51)	19 (54)	87 (51)
T 阶段				
T1	146 (51)	38 (48)	14 (40)	94 (55)
T2	132 (46)	41 (51)	19 (54)	72 (42)
T3/4	8 (3)	1 (1)	2 (6)	5 (3)
级别				
不良	148 (52)	37 (46)	24 (69)	87 (51)
中度	42 (15)	12 (15)	3 (9)	27 (16)
良好	7 (2)	2 (3)	2 (6)	3 (2)
未知	89 (31)	29 (36)	6 (17)	54 (32)
ER*				
阳性	209 (73)	80 (100)	0 (0)	129 (75)
阴性	77 (27)	0 (0)	35 (100)	42 (25)
PgR*				
阳性	165 (58)	59 (74)	5 (14)	101 (59)
阴性	111 (39)	19 (24)	29 (83)	63 (37)
未知	10 (3)	2 (2)	1 (3)	7 (4)
转移 <5 年				
是	93 (33)	24 (30)	13 (37)	56 (33)
否	183 (64)	51 (64)	17 (49)	115 (67)
如果 <5 年, 检查的	10 (3)	5 (6)	5 (14)	0 (0)

*ER 阳性和 PgR 阳性: >10 fmol/mg 蛋白或 >10% 阳性肿瘤细胞。

在年龄或绝经状态中没有差异。正如所料, 与 ER 阳性训练集相比, ER 阴性的训练组具有稍微更高比例的更大的肿瘤, 以及更多不良分级的肿瘤。对于任何患者或肿瘤特征, 171 位患者的验证组 (129 位 ER 阳性, 42 位 ER 阴性) 没有不同于 286 位患者的总组。

使用两种方法来鉴定预测疾病复发的标记物。首先, 我们将所有 286 位患者 (ER 阳性和 ER 阴性组合起来) 分到训练集和测试集中。从训练集的 80 位患者中选出 35 个基因, 构建 Cox's 模型来预测远处

转移的发生。观察到中度的预后值。表 2. 无监督的聚类分析显示了与肿瘤 ER 状态高度相关的两个不同的亚组 (卡方检验 $p < 0.0001$)。附图 1B, 其支持了第二种方法, 在其中患者首先根据 ER 状态被放置到亚组中。

表 2

SEQ ID NO:	Cox's 系数	p 值
1	4.008	0.00006
2	-3.649	0.00026
3	4.005	0.00006
4	-3.885	0.00010
5	-3.508	0.00045
6	-3.176	0.00150
7	3.781	0.00016
8	3.727	0.00019
9	-3.570	0.00036
10	-3.477	0.00051
11	3.555	0.00038
12	-3.238	0.00120
13	-3.238	0.00120
14	3.405	0.00066
15	3.590	0.00033
16	-3.157	0.00160
17	-3.622	0.00029
18	-3.698	0.00022
19	3.323	0.00089
20	-3.556	0.00038
21	-3.317	0.00091
22	-2.903	0.00370
23	-3.338	0.00085
24	-3.339	0.00084
25	-3.355	0.00079
26	3.713	0.00021
27	-3.325	0.00088
28	-2.984	0.00284
29	3.527	0.00042
30	-3.249	0.00116
31	-2.912	0.00360
32	3.118	0.00182
33	3.435	0.00059
34	-2.971	0.00297
35	3.282	0.00103

为了选择标记物对每个亚组进行分析。从训练集的患者中选出 76 个基因 (60 个来自 ER 阳性组, 16 个来自 ER 阴性组) (附图 2A, 左侧)。将选择的基因和 ER 状态合起来, 对所有 LNN 患者构建预测癌症复发的 Cox's 模型。171 位患者的测试集中 76 基因预测物的验证产生 0.694 的 AUC 值、93% 的敏感性 (52/56) 和 48% 的特异性 (55/115) 的 ROC (附图 2A, 右侧)。具有高于预后标志的阈值的复发分值的患者具有 11.9 倍的 OR (95% CI: 4.04-35.1; $p < 0.0001$) 在 5 年内发展出远处转移。作为对照, 产生随机选择的 76 基因集。在测试组中这些产生了 0.515 的平均 AUC 值、91% 的敏感性和 12% 的特异性的 ROC。通过这种基因集分级的患者具有 1.3 的几率比 (0.50-3.90; $p = 0.8$) 发展出转移, 指示了随机的分类。此外, 作为 76 基因标志的函数的无远处转移存活率 (DMFS) 和总体存活率 (OS) 的 Kaplan-Meier 分析, 显示了预测具有良好和不良预后的组之间转移在时间上的高显著性的差异。(附图 2)。在 60 和 80 个月, 在具有预测的良好和不良预后的组之间在 DMFS 中各自的绝对差异分别为 40% (93% vs. 53%) 和 39% (88% vs. 49%), 在 OS 中分别为 27% (97% vs. 70%) 和 32% (95% vs. 63%)。

76-基因分布型还代表了在 84 位绝经前患者 (HR: 9.60)、87 位绝经后患者 (HR: 4.04) 以及 79 位具有 10 到 20mm 大小的肿瘤的患者 (HR: 14.1) 的亚组中发生远处转移的强预后因素 (附图 3)。

在表 3 中概述了单变量和多变量 Cox's 回归分析。

表 3: 在 171 位 LNN 患者的测试集中 DMFS 的单变量和多变量分析

	单变量分析			多变量分析*		
	HR†	(95% CI) †	p 值	HR†	(95% CI) †	p 值
年龄‡						
年龄 2vs.年龄 1	1.16	(0.51 – 2.65)	0.7180	1.14	(0.45 – 2.91)	0.7809
年龄 3vs.年龄 1	1.32	(0.56 – 3.10)	0.5280	0.87	(0.26 – 2.93)	0.8232
年龄 4vs.年龄 1	0.95	(0.32 – 2.82)	0.9225	0.61	(0.15 – 2.60)	0.5072
绝经状态§	1.24	(0.76 – 2.03)	0.3909	1.53	(0.68 – 3.44)	0.3056
阶段	1.08	(0.66 – 1.77)	0.7619	2.57	(0.23 – 29.4)	0.4468
区别¶	0.38	(0.16 – 0.90)	0.0281	0.60	(0.24 – 1.46)	0.2590
肿瘤大小**	1.06	(0.65 – 1.74)	0.8158	0.34	(0.03 – 3.90)	0.3849
ER††	1.09	(0.61 – 1.98)	0.7649	1.05	(0.54 – 2.04)	0.8935
PR††	0.83	(0.51 – 1.38)	0.4777	0.85	(0.47 – 1.53)	0.5882
76-基因标志	5.67	(2.59 – 12.4)	1.5 x 10 ⁻⁵	5.55	(2.46 – 12.5)	3.6 x 10 ⁻⁵

*由于 9 位患者中缺失数值，多变量模型包括 162 位患者。

†危险比例和 95% 置信区间

‡年龄 1 是 <40 岁，年龄 2 是 41 到 55 岁，年龄 3 是 56 到 70 岁，年龄 4 是 >70 岁。

§绝经后 vs.绝经前

|阶段: II & III vs. I

¶级别: 中度/好 vs.不良，未知级别作为独立的组来包括。

**肿瘤大小: > 20 mm vs. < 20 mm

††阳性 vs.阴性

不同于该 76 基因标志，在单变量分析中仅级别是显著的，中度和良好的区别与有利的 DMFS 相关。5 年内发生肿瘤转移的 HR 的多变量回归估计为 5.55 ($p < 0.0001$)，表明 76 基因集代表了与肿瘤转移的更高风险强烈相关的独立预后标志。还为 ER 阳性和 ER 阴性患者单独地进行单变量和多变量分析，在由 ER 状态分级的亚组中 76 基因标志也是独立的预后变量。

分析预后标志中 76 个基因的功能 (表 4) 以将基因与生物学途径相关联。

表 4

ER 状态	SEQ ID NO:	Std. Cox's 系数	Cox's p 值
+	36	-3.83	0.00005
+	37	-3.865	0.00001
+	38	3.63	0.00002
+	39	-3.471	0.00016
+	40	3.506	0.00008
+	41	-3.476	0.00001
+	42	3.392	0.00006
+	43	-3.353	0.00080
+	44	-3.301	0.00038
+	45	3.101	0.00033
+	46	-3.174	0.00128
+	47	3.083	0.00020
+	48	3.336	0.00005
+	49	-3.054	0.00063
+	50	-3.025	0.00332
+	51	3.095	0.00044
+	52	-3.175	0.00031
+	53	-3.082	0.00086
+	54	3.058	0.00016
+	55	3.085	0.00009

+	56	-2.992	0.00040
+	57	-2.791	0.00020
+	58	-2.948	0.00039
+	59	2.931	0.00020
+	60	-2.896	0.00052
+	61	2.924	0.00050
+	62	2.915	0.00055
+	63	-2.968	0.00099
+	64	2.824	0.00086
+	65	-2.777	0.00398
+	66	-2.635	0.00160
+	67	-2.854	0.00053
+	68	2.842	0.00051
+	69	-2.835	0.00033
+	70	2.777	0.00164
+	71	-2.759	0.00222
+	72	-2.745	0.00086
+	73	2.79	0.00049
+	74	2.883	0.00031
+	75	-2.794	0.00139
+	76	-2.743	0.00088
+	77	-2.761	0.00164
+	78	-2.831	0.00535
+	79	2.659	0.00073
+	80	-2.715	0.00376
+	81	2.836	0.00029
+	82	-2.687	0.00438
+	83	-2.631	0.00226
+	84	-2.716	0.00089
+	85	2.703	0.00232
+	86	-2.641	0.00537
+	87	-2.686	0.00479
+	88	-2.654	0.00363
+	89	2.695	0.00095
+	90	-2.758	0.00222
+	91	2.702	0.00084
+	92	-2.694	0.00518
+	93	2.711	0.00049
+	94	-2.771	0.00156
+	95	2.604	0.00285
-	96	-3.495	0.00011
-	97	3.224	0.00036
-	98	-3.225	0.00041
-	99	-3.145	0.00057
-	100	-3.055	0.00075
-	101	-3.037	0.00091

-	102	-3.066	0.00072
-	103	3.06	0.00077
-	104	-2.985	0.00081
-	105	-2.983	0.00104
-	106	-3.022	0.00095
-	107	-3.054	0.00082
-	108	-3.006	0.00098
-	109	-2.917	0.00134
-	110	-2.924	0.00149
-	111	-2.882	0.0017

尽管 76 个基因中的 18 个具有未知的功能，鉴定了很好地表现的几种途径或生化活性，例如细胞死亡、细胞周期和增殖、DNA 复制和修复，和免疫反应（表 5）。

表 5. 来自预后标志的 76 个基因的途径分析

功能分类	76-基因标志
细胞死亡	TNFSF10, TNFSF13, MAP4, CD44, IL18, GAS2, NEFL, EEF1A2, BCLG, C3
细胞周期	CCNE2, CD44, MAP4, SMC4L1, TNFSF10, AP2A2, FEN1, KPNA2, ORC3L, PLK1
增殖	CD44, IL18, TNFSF10, TNFSF13, PPP1CC, CAPN2, PLK1, SAT
DNA 复制、重组	TNFSF10, SMC4L1, FEN1, ORC3L, KPNA2, SUPT16H,
修复	POLQ, ADPRTL1
免疫反应	TNFSF10, CD44, IL18, TNFSF13, ARHGDIB, C3
生长	PPP1CC, CD44, IL18, TNFSF10, SAT, HDGFRP3
细胞装配和构造	MAP4, NEFL, TNFSF10, PLK1, AP2A2, SMC4L1
转录	KPNA2, DUSP4, SUPT16H, DKFZP434E2220, PHF11, ETV2
细胞-细胞信号传递和相互作用	CD44, IL18, TNFSF10, TNFSF13, C3
存活	TNFSF10, TNFSF13, CD44, NEFL
发育	IL18, TNFSF10, COL2A1
细胞形态	CAPN2, CD44, TACC2
蛋白质合成	IL18, TNFSF10, EEF1A2
ATP 结合	PRO2000, URKL1, ACACB
DNA 结合	HIST1H4H, DKFZP434E2220, PHF11
集落形成	CD44, TNFSF10
粘附	CD44, TMEM8
神经发生	CLN8, NEURL
高尔基体	GOLPH2, BICD1
激酶活性	CNK1, URKL1
转移酶活性	FUT3, ADPRTL1

所发现的与疾病发展有关的基因包括 calpain2、起点识别蛋白、

双特异性磷酸酶、Rho-GDP 分离抑制物、TNF 超家族蛋白、互补成分 3、微管相关蛋白、蛋白磷酸酶 1 和细胞凋亡调节物 BCL-G。此外，早先表征的预后基因，例如 cyclin E2 (Keyomarsi et al. (2002)) 和 CD44 (Herrera-Gayol et al. (1999)) 存在于该基因标志中。

实施例 6

讨论实施例 1-5

我们提供了来自所有年龄组和肿瘤大小的 286 位淋巴结阴性乳腺癌患者的原发肿瘤分析结果。患者没有接受辅助性全身治疗，从而预后的多基因评定不受到与全身性治疗有关的预测因素的潜在混淆性贡献的影响。

这些研究显示了精确地预测远处肿瘤复发的 76 基因标志。这个标志独立于年龄、肿瘤大小和级别和 ER 状态适合于所有 LNN 乳腺癌患者。在 DMFS 的 Cox's 多变量分析中，76 基因标志是唯一重要的变量，取代了临床的变量，包括级别。5 年后，具有良好的和不良 76 基因标志的患者之间 DMFS 和 OS 中的绝对差异分别是 40% 和 27%。对于具有良好预后标志的患者，5 年内 7% 发展了远处转移，3% 死亡。如果进一步验证，假定 LNN 患者中 25% 比例的疾病复发，这个预后标志将产生 37% 的正预测值和 95% 的负预测值。特别地，这个标志对于定义升高比例的 T1 肿瘤 (<2 cm) 的复发风险是很有价值的。与 St Gallen 和 NIH 指导原则比较是有益的。虽然确保相同数量的高风险患者将接受必需的治疗，我们的 76 基因标志将仅对 52% 的低风险患者建议全身性佐剂化疗，与 St. Gallen 和 NIH 指导原则分别是 90% 和 89% 相对比 (表 6)。

表 6. 比较 76 基因标志和当前治疗乳腺癌的常规共识

方法	在测试集中被指导接受佐剂化疗的患者	
	5 年内的转移性疾病 (%)	5 年内的无转移性疾病 (%)
St Gallen	52/55 (95)	104/115 (90)
NIH	52/55 (95)	101/114 (89)
76 基因标志	52/56 (93)	60/115 (52)

常规的公认指标。St. Gallen: 肿瘤 > 2cm, ER-阴性, 级别 2-3, 患者 <35 岁 (这些指标的任何一个); NIH: 肿瘤 >1cm.

因而该 76 基因标志可以引起被推荐进行不必要的辅助性全身治疗的低风险 LNN 患者数目的降低。

预后标志中的 76 个基因属于许多功能分类, 表明不同的途径可能引起疾病发展。标志包括良好表征的基因和 18 个未知基因。这个发现可以解释与其他预后因素相比这种标志的优越性能。虽然在由 ER 状态分级的患者组中都发现了涉及细胞死亡、细胞增殖和转录调节的基因, 为 ER 阳性组所选的 60 个基因和为 ER 阴性组所选的 16 个基因没有重叠。这个结果支持了这样的思想, 疾病发展的异质性的程度和根本的机制对于乳腺癌患者的两种基于 ER 的亚组可能是不同的。

因为在患者、技术和使用的材料上的差异, 将我们的结果与 van de Vijver et al. (2002) 的研究相比较是困难的。van de Vijver et al. 包括了淋巴结阴性和淋巴结阳性患者, 它们有或没有接受过辅助性全身治疗, 并仅包括小于 53 岁的女性。此外, 在研究中使用的微阵列平台是不同的, Affymetrix 对比 Agilent。对于 van't Veer (2002) 研究的 70 个基因, 仅 48 个存在于 Affymetrix U133a 阵列上, 而我们的 76 个基因仅 38 个存在于 Agilent 阵列上。在两种标志之间有 3 个基因重叠 (cyclin E2、起点识别复合物和 TNF 超家族蛋白)。尽管具有显著差异, 两种标志都包括一些基因, 所述基因鉴定了几种可能涉及肿瘤复发的常见途径。这种发现支持了这样的思想, 虽然在基因成员中可能有冗余, 可能需要有效的标志来包括特定途径的呈现。

与 van de Vijver et al. (2002) 的研究相比, 我们的研究的优点在于更大数目的未治疗的 LNN 患者 (286 vs. 141), 以及我们的 76 基因

标记物对于年龄、绝经状态和肿瘤大小的独立性。与其他的报道的 90% 相比，我们的患者的验证集与训练集完全没有重叠。Ransohoff (2004)。总之，由于仅大约 30-40% 的未治疗 LNN 患者发生了肿瘤复发，该预后标志可以提供有力的工具来鉴定那些低风险的患者，防止在相当数量的患者中的过度治疗。在患有原发性乳腺癌的患者中推荐辅助性全身治疗将来可以通过这种预后标志来指导。可以在辅助情形或在患有转移性疾病的患者中测试我们的基因标志对于不同的全身性治疗方式有效性的预测值。

实施例 7

比较产自阶段 I/II 乳腺癌激光俘获显微解剖和大块组织的乳腺癌肿瘤基因分布型

已经证明了基因表达分布型是对各种癌症类型的强大的诊断和预后工具。几乎在所有情况下大块肿瘤 RNA 被用于在芯片上杂交。在激素依赖性肿瘤的发展和生长中，雌激素扮演了重要的角色。

约 75% 的乳腺癌表达雌激素受体 (ER)，其是 (辅助性) 它莫西芬 (tamoxifen) 治疗的指示物，并与患者结局相关。

为了看清在乳腺上皮细胞中由雌激素触发的机制以及它们与肿瘤发生的相关性，使用激光俘获显微解剖 (LCM) 来从 29 个早期原发性乳腺癌取得组织学上同质的肿瘤细胞群体，与基因芯片表达分析相结合。这 29 位患者中，根据对肿瘤胞质液的定量配体结合或酶免疫测定，11 位是 ER 阴性，17 位是 ER 阳性。为了对比，还使用分离自相同的 29 位患者的组的大块组织 RNA 获得了基因表达分布型。

从 29 位淋巴结阴性乳腺癌患者 (肿瘤特征，表 2) 收集新鲜的冷冻组织样品，所述患者已经外科手术治疗了乳腺肿瘤，没有接受新的辅助性全身治疗。对于每个患者组织样品，我们首先使用 H & E 载玻片来评估细胞形态。通过在低温恒温切片上进行 LCM (PALM)，和从整个低温恒温切片，即相同肿瘤的大块组织获得肿瘤细胞，从两种肿瘤细胞分离 RNA。通过 Agilent BioAnalyzer 分析 RNA 样品质量。将 RNA 样品与含有大约 22,000 个探针集的 Affymetrix 人类 U133A 芯片杂交。对荧光进行定量，对强度进行标准化。使用聚类分析和主要成分分析来将患者按相似的基因表达分布型分组。选择在 ER 阳性

和 ER 阴性样品之间差别表达的基因。

在目标制备中，分离自 LCM 取得的乳腺癌细胞的总 RNA 经历两轮基于 T7 的扩增，相比之下大块组织 RNA 经历一轮扩增。比较 LCM 数据集和大块组织集之间的 21 个对照基因的表达水平（表 7），来证明线性扩增的精确度。

表 7: 对照基因列表

SEQ ID NO:	名称
112	蛋白磷酸酶 2, 调节亚基 B (B56), delta 同种型
113	CCCTC 结合因子 (锌指蛋白)
114	溶质载体家族 4 (阴离子交换剂), 成员 1, 衔接蛋白
115	核糖核酸酶 P
116	假定的蛋白 FLJ20188
117	KIAA0323 蛋白
118	cDNA FLJ12469
119	翻译起始因子 eIF-2b delta 亚基
120	核不均一核糖核蛋白 K
121	羟基甲基胆色烷合酶
122	cDNA DKFZp586O0222
123	染色体 20 开放阅读框 4
124	甲状腺激素受体互作因子 (interactor) 4
125	次黄嘌呤磷酸核糖转移酶 1 (Lesch-Nyhan 综合征)
126	DnaJ (Hsp40) 同源物, 亚族 C, 成员 8
127	双特异性磷酸酶 11 (RNA/RNP 复合物 1-相互作用)
128	钙结合特异性相关自身抗原 1
129	基质细胞衍生因子 2
130	Ewing 肉瘤断点区域 1
131	CCR4-NOT 转录复合物, 亚基 2
132	F-盒唯一蛋白 7 (F-box only protein 7)

获得的结果在表 8 和附图 4-9 中描述。

表 8 患者的临床特征

特征		患者数 (%)
年龄		
	< 40	1 (3)
	40-44	5 (17)
	45-49	8 (28)
	> 50	15 (52)
肿瘤直径 mm		
	< 20	11 (40)
	> 20	17 (59)
组织级别		
	II (中度)	5 (17)
	III (不良)	12 (41)
雌激素受体状态		
	阴性	11 (40)
	阳性	17 (59)
手术		
	乳腺保留治疗	26 (90)
	乳腺切除术	3 (10)
化疗		
	无	29 (100)
激素治疗		
	无	29 (100)
按月的无疾病存活		
	< 48	13 (45)
	> 48	16 (55)

附图 4 是根据 5121 个基因的分级聚类，显示了 LCM 和大块组织样品根据全局 RNA 表达分布型是完全独立的。附图 5 是描述在分离

自 LCM 样品和大块组织的 RNA 中 21 个对照基因 (表 6) 表达水平的柱形图, 显示了用于 LCM 所获 RNA 的另一轮线性扩增没有引起对照基因的差异表达。附图 6 是数据分析途径, 附图 7 显示了使用过滤的基因集的 PCA 分析。附图 8 是用来分配患者亚组的 ER 状态的饼分图。通过 Student t-测验来定义 LCM 和大块组织样品中 ER 阳性和 ER 阴性的亚簇 (sub-clusters) 之间差异表达的基因。附图 9 是一系列柱形图, 显示了通过 Gene Ontology 对 LCM 样品中、仅在大块组织中, 和对于 LCM 和大块组织是共有的那些样品中仅与 ER 相关的基因的途径分析的结果。

总的来说, 获得的结果显示几个重要的结论。首先, 在大块组织数据集和 LCM 数据集中都看到了 ER-/ER+ 患者中与细胞增殖和能量代谢相关的基因的差异表达。第二, 由于通过 LCM 对乳腺癌细胞的富集, 发现涉及细胞表面受体相关信号转导、RAS 信号转导、JAK-STAT 信号转导和细胞凋亡的基因与 ER 状态相关。在大块数据集中没有鉴定出这些基因。第三, 显微解剖提供了研究上皮细胞瘤的敏感方法, 以及对与雌激素受体相关的信号途径的了解。因此, 明显的是, 对 LCM 分离的肿瘤细胞应用在此描述的基因表达分布型与在异质大块组织中获得的结果相当。

实施例 8

在乳腺癌中 76 基因预后标志的验证和途径分析

这个实施例报告了验证研究的结果, 在所述验证研究中, 76 基因标志被用于预测来自 4 个独立来源的 132 位患者的结局。

此外, 为了评估这些基因标志的强大程度, 这个实施例进一步提供了鉴定所述标志的可替代成分, 并描述了这样的替换如何产生有效标志中关键途径的鉴定。

新鲜的冷冻组织样品收集自 132 位患者, 所述患者已经手术治疗了乳腺肿瘤, 并且没有接受辅助性全身治疗。使用的患者样品收集自 1980 和 1996 年之间 (附图 10)。对于每个患者组织样品, 使用 H & E 载玻片来评估细胞形态。然后制备总 RNA 样品, 通过 Agilent BioAnalyzer 分析样品质量。通过微阵列分析来分析 RNA 样品。对荧光进行定量, 对强度进行标准化。根据 76 个基因标志的表达水平对

每个患者计算复发危险分值。将患者分成良好和不良结局组。附图 11 和 12。

为了评估这个基因标志的强大程度，设计和使用了两种统计分析。附图 13。首先，重复了我们用来发现该 76 基因标志的基因选择和标志构建过程。如表 8 所示，从总共 286 位患者中随机选择 115 位患者的十个训练集。余下的患者充当测试集。

第二，将训练集中的患者数目增加到 286 位患者的 80%，使用剩余的 20% 患者作为测试集。这个选择过程也重复 10 次。在两种过程中，使用 Kaplan-Meier 存活曲线来确保训练和测试对之间无疾病存活方面没有显著差异。选择基因，使用 Cox's 比例-危险回归从每个训练集构建标志。在相应的测试集中验证每个标志。此外，使用 GO 本体论分类法将 76 基因预后标志分类到功能组中。选择覆盖了标志中显著数量的基因的途径 (p 值 <0.05 以及 >2 次命中)。还在来自不同训练集的所有预后标志中评估了所选择的途径。

在表 9 中，A 含有使用 115 位患者的训练集的 10 个标志的结果，B 含有使用 80% 患者的训练集的 10 个标志的结果。

A		B	
ROC 的 AUC	0.62 (0.55-0.70)	ROC 的 AUC	0.62 (0.53-0.72)
敏感性	86% (0.84-0.88)	敏感性	83% (0.81-0.85)
特异性	34% (0.21-0.56)	特异性	46% (0.28-0.62)
复发频率	33%	复发频率	33%
PPV	40% (0.35-0.49)	PPV	47% (0.32-0.58)
NPV	81% (0.75-0.89)	NPV	82% (0.78-0.89)
几率比	3.5 (1.7-7.9)	几率比	5.6 (1.7-15)

这个实施例中获得的结果显示了：

- 在 132 位独立患者中成功地验证了 76 基因标志，在来自 4 个独立来源的 132 位 LNN 乳腺癌患者中得出 0.757 的 AUC 值。该标志显示了 88% 的敏感性和 41% 的特异性。

- 取代标志的平均 AUC 是 0.64 (95% CI : 0.53-0.72)。这个结果与 76 基因预测物的相一致 (AUC 0.69)。在 76 基因标志中过量

呈现的 21 个途径也在所有其他预后标志中发现，表明了肿瘤复发中涉及共同的生物学途径。

- 这些结果表明，基因表达分布型提供了进行患者结局的风险评估的强大方法。该数据突出了为患者提供肿瘤复发的定量测定的分子性预后分析的可行性。

尽管已经通过例示和举例为了清楚和理解的目的较详细地描述了上述发明，这些说明和实施例不应被看作限制本发明的范围。

表 10.
序列标识

SEQ ID NO:	Psid	基因名称	登记 #	基因描述
1	213165_at	CDABP0086	AI041204	
2	217432_s_at		AF179281	艾杜糖醛酸 2-硫酸酯酶 (Hunter 综合症)
3	221500_s_at		BE782754	syntaxin 16/
4	208452_x_at	MYO9B	NM_004145	肌球蛋白 IXB
5	220234_at	CA8	NM_004056	碳酸酐酶 VIII
6	207865_s_at	BMP8	NM_001720	骨骼形态发生蛋白 8 (成骨蛋白 2)
7	201769_at	KIAA0171	NM_014666	KIAA0171 基因产物
8	218940_at	FLJ13920	NM_024558	假定的蛋白 FLJ13920
9	209018_s_at	BRPK	BF432478	蛋白激酶 BRPK
10	216647_at	DKFZp586L1 824	AL117663	来自克隆 DKFZp586L1824
11	213405_at	DKFZp564E1 22	N95443	来自克隆 DKFZp564E122
12	202921_s_at	ANK2	NM_001148	锚蛋白 2, 神经元的, 转录产物变体 1
13	208401_s_at		U01157	具有 CA 二核苷酸重复的胰高血糖素样肽-1 受体
14	218090_s_at	WDR11	NM_018117	WD40 重复结构域 11 蛋白
15	218139_s_at	FLJ10813	NM_018229	假定的蛋白 FLJ10813
16	202485_s_at	MBD2	NM_003927	甲基-CpG 结合结构域蛋白 2, 转录产物变体 1
17	201357_s_at	SF3A1	NM_005877	剪接因子 3a, 亚基 1, 120kD
18	214616_at	H3FD	NM_003532	H3 组蛋白家族, 成员 D
19	207719_x_at	KIAA0470	NM_014812	KIAA0470 基因产物
20	202734_at	TRIP10	NM_004240	甲状腺激素受体互作因子 (interactor) 10
21	202175_at	FLJ22678	NM_024536	假定的蛋白 FLJ22678
22	213870_at		AL031228	染色体 6p21.2-21.31 上的克隆 1033B10

23	208967_s_at	adk2	U39945	腺苷酸激酶 2
24	204312_x_at		AI655737	cAMP 应答元件结合蛋白 1
25	203815_at	GSTT1	NM_000853	谷胱甘肽 S-转移酶 ζ 1
26	207996_s_at	C18ORF1	NM_004338	染色体 18 开放阅读框 1
27	221435_x_at	HT036	NM_031207	假定的蛋白 HT036
28	219987_at	FLJ12684	NM_024534	假定的蛋白 FLJ12684
29	221559_s_at	MGC:2488	BC000229	克隆 MGC:2488
30	207007_at	NR1I3	NM_005122	核受体亚族 1, 组 1, 成员 3
31	219265_at	FLJ13204	NM_024761	假定的蛋白 FLJ13204
32	40420_at		AB015718	蛋白激酶的 lok mRNA
33	202266_at	AD022	NM_016614	TRAF 和 TNF 受体相关蛋白
34	219522_at	FJX1	NM_014344	fjx1 的推定的分泌配体同源物
35	212334_at	AKAP350C	BE880245	AKAP350C, 选择性剪接的
36	219340_s_at	CLN8	AF123759	推定的跨膜蛋白
37	217771_at	GP73	NM_016548	Golgi 膜蛋白 (LOC51280)
38	202418_at	Yif1p	NM_020470	推定的跨膜蛋白; 酵母 Golgi 膜蛋白的同源物
39	206295_at	IL-18	NM_001562	白细胞介素 18
40	201091_s_at		BE748755	异染色质样蛋白
41	204015_s_at	DUSP4	BC002671	双特异性磷酸酶 4
42	200726_at	PPP1CC	NM_002710	蛋白磷酸酶 1, 催化亚单位, γ 同种型
43	200965_s_at	ABLIM-s	NM_006720	肌动蛋白结合 LIM 蛋白 1, 转录产物变体
44	210314_x_at	TRDL-1	AF114013	肿瘤坏死因子相关死亡配体 1 γ
45	221882_s_at	M83	AI636233	五跨度跨膜蛋白
46	217767_at	C3	NM_000064	互补物成分 3
47	219588_s_at	FLJ20311	NM_017760	假定的蛋白
48	204073_s_at	C11ORF9	NM_013279	染色体 11 开放阅读框 9
49	212567_s_at		AL523310	推定的翻译起始因子
50	211382_s_at	TACC2	AF220152	
51	201663_s_at	CAP-C	NM_005496	染色体相关多肽 C

52	221344_at	OR12D2	NM_013936	嗅感受蛋白, 家族 12, 亚族 D, 成员 2
53	210028_s_at	ORC3	AF125507	起点识别复合物亚基 3
54	218782_s_at	PRO2000	NM_014109	PRO2000 蛋白
55	201664_at	SMC4	AL136877	(酵母染色体 4 的结构维持) 样
56	219724_s_at	KIAA0748	NM_014796	KIAA0748 基因产物
57	204014_at	DUSP4	NM_001394	双特异性磷酸酶 4
58	212014_x_at	CD44	AI493245	CD44
59	202240_at	PLK1	NM_005030	Polo (果蝇) 样激酶 1
60	204740_at	CNK1	NM_006314	KSR 样的连接物增强子 (ras 的果蝇激酶抑制物)
61	208180_s_at	H4FH	NM_003543	H4 组蛋白家族, 成员 H
62	204768_s_at	FEN1	NM_004111	Flap 结构特异性核酸内切酶
63	203391_at	FKBP2	NM_004470	FK506 结合蛋白 2
64	211762_s_at	KPNA2	BC005978	Karyopherin α 2 (RAG cohort 1, importin α 1)
65	218914_at	CGI-41	NM_015997	CGI-41 蛋白
66	221028_s_at	MGC11335	NM_030819	假定的蛋白 MGC11335
67	211779_x_at	MGC13188	BC006155	克隆 MGC:13188
68	218883_s_at	FLJ23468	NM_024629	假定的蛋白 FLJ23468
69	204888_s_at		AA772093	Neuralized (果蝇) 样
70	217815_at	FACTP140	NM_007192	染色质特异性转录延伸因子, 140 kD 亚基
71	201368_at	Tis11d	U07802	
72	201288_at	ARHGDI B	NM_001175	Rho GDP 分离抑制物 (GDI) β
73	201068_s_at	PSMC2	NM_002803	蛋白酶体 (prosome, macropain) 26S 亚基, ATPase, 2
74	218478_s_at	DKFZP434E220	NM_017612	假定的蛋白 DKFZP434E2220
75	214919_s_at	KIAA1085	R39094	
76	209835_x_at		BC004372	类似于 CD44
77	217471_at		AL117652	

78	203306_s_at	SLC35A1	NM_006416	溶质载体家族 35 (CMP 唾液酸转运蛋白), 成员 1
79	205034_at	CCNE2	NM_004702	Cyclin E2
80	221816_s_at		BF055474	推定的锌指蛋白 NY-REN-34 抗原
81	219510_at	POLQ	NM_006596	聚合酶 (DNA 指导的) ζ
82	217102_at		AF041410	恶性肿瘤相关蛋白
83	208683_at	CANP	M23254	Ca ²⁺ 活化的中性蛋白酶大亚基
84	215510_at		AV693985	ets 变体基因 2
85	218533_s_at	FLJ20517	NM_017859	假定的蛋白 FLJ20517
86	215633_x_at	LST-1N	AV713720	LST-1N 蛋白的 mRNA
87	221928_at		AI057637	Hs234898 EST, 微弱地类似于 2109260A B 细胞生长因子
88	214806_at	BICD	U90030	Bicaudal-D
89	204540_at	EEF1A2	NM_001958	真核的翻译延伸因子 1 α 2
90	221916_at		BF055311	假定的蛋白
91	216693_x_at	DKFZp434C1 722	AL133102	
92	209500_x_at		AF114012	肿瘤坏死因子相关死亡配体-1 β
93	209534_at	FLJ10418	AK001280	中度类似于肝细胞瘤衍生的生长因子
94	207118_s_at	MMP23A	NM_004659	基质金属蛋白酶 23A
95	211040_x_at		BC006325	G-2 和 S 期表达的 1
96	218430_s_at	FLJ12994	NM_022841	假定的蛋白 FLJ12994
97	217404_s_at		X16468	α -1 II 类胶原蛋白
98	205848_at	GAS2	NM_005256	生长延滞特异性 2
99	214915_at	FLJ11780	AK021842	克隆 HEMBA1005931, 微弱地类似于锌 指蛋白 83
100	216010_x_at		D89324	α (1,31,4) 岩藻糖基转移酶
101	204631_at	MYH2	NM_017534	肌球蛋白重肽 2 骨骼肌 成年人
102	202687_s_at		U57059	Apo-2 配体 mRNA
103	221634_at		BC000596	类似于核糖体蛋白 L23a, 克隆 MGC:2597
104	220886_at	GABRQ	NM_018558	γ -氨基丁酸 (GABA) 受体, ζ

105	202237_at	ADPRTL1	NM_006437	ADP-核糖基转移酶 (NAD ⁺ ; 聚 (ADP-核糖) 聚合酶) 样 1
106	204218_at	DKFZP564M082	NM_014042	蛋白 DKFZP564M082
107	221241_s_at	BCLG	NM_030766	细胞凋亡调节物 BCL-G
108	209862_s_at		BC001233	类似于 KIAA0092 基因产物, 克隆 MGC:4896
109	217019_at	RPS4X	AL137162	含有新的基因和类似 X 连接的核糖体蛋白 4 的新蛋白基因的 5 部分
110	210593_at		M55580	亚精胺精胺 N1 转乙酰酶
111	216103_at	KIAA0707	AB014607	KIAA0707
112	202513_s_at	PPP2R5D	NM_006245	蛋白磷酸酶 2, 调节亚基 B (B56), delta 同种型
113	202521_at	CTCF	NM_006565	CCCTC 结合因子 (锌指蛋白)
114	218682_s_at	SLC4A1AP	NM_018158	溶质载体家族 4 (阴离子交换剂), 成员 1, 衔接蛋白
115	203436_at	RPP30	NM_006413	核糖核酸酶 P
116	220127_s_at	FLJ20188	NM_017703	假定的蛋白 FLJ20188
117	212355_at	KIAA0323	AI075450	KIAA0323 蛋白
118	215158_s_at	FLJ12469	AK022531	cDNA FLJ12469
119	209429_x_at		AF112207	翻译起始因子 eIF-2b delta 亚基
120	200097_s_at		AI701949	核不均一核糖核蛋白 K
121	203040_s_at	HMBS	NM_000190	羟基甲基胆色烷合酶
122	221647_s_at	DKFZp586O0222	AL136935	克隆 DKFZp586O0222
123	218089_at	C20orf4	NM_015511	染色体 20 开放阅读框 4 DKFZP564N1363
124	203732_at	TRIP4	NM_016213	甲状腺激素受体互作因子 (interactor) 4
125	202854_at	HPRT1	NM_000194	次黄嘌呤磷酸核糖转移酶 1 (Lesch-Nyhan 综合征)

126	205545_x_at	DNAJC8	NM_014280	DnaJ (Hsp40) 同源物, 亚族 C, 成员 8
127	202703_at	DUSP11	NM_003584	双特异性磷酸酶 11 (RNA/RNP 复合物 1-相互作用)
128	216903_s_at	CBARA1	AK022697	钙结合特异性相关自身抗原 1
129	203090_at	SDF2	NM_006923	基质细胞衍生因子 2
130	209214_s_at	EWSR1	BC004817	Ewing 肉瘤断点区域 1
131	217798_at	CNOT2	AI123426	CCR4-NOT 转录复合物, 亚基 2
132	201178_at	FBXO7	NM_012179	F-box 唯一的蛋白 7
133	212160_at		AI984005	输出蛋白, tRNA (tRNA 的核输出受体)
134	201111_at	CSE1	AF053641	脑细胞的细胞凋亡敏感性蛋白
135	201112_s_at	CSE1L	NM_001316	染色体分离 1 (酵母同源物) 样
136	204817_at	KIAA0165	NM_012291	额外的纺锤极, <i>S. cerevisiae</i> , (KIAA0165) 的同源物
137	215623_x_at	FLJ11338	AK002200	高度类似于染色体相关多肽 C mRNA
138	38158_at	KIAA0165	D79987	
139	201076_at	NHP2L1	NM_005008	非组蛋白染色体蛋白 2 (<i>S. cerevisiae</i>) 样 1
140	201947_s_at	CCT2	NM_006431	含有 chaperonin 的 TCP1, 亚基 2 (beta)
141	202647_s_at	NRAS	NM_002524	成神经细胞瘤 RAS 病毒 (v-ras) 癌基因同源物
142	202705_at	CCNB2	NM_004701	cyclin B2
143	204009_s_at		NM_004985	v-Ki-ras2 Kirsten 大鼠肉瘤 2 病毒癌基因同源物
144	204566_at	PPM1D	NM_003620	蛋白磷酸酶 1D 镁依赖性, delta 同种型
145	214710_s_at		BE407516	
146	202095_s_at	BIRC5	NM_001168	baculoviral IAP 含重复 5 (存活蛋白)
147	204900_x_at	SAP30	NM_003864	sin3 相关多肽, 30kD
148	201986_at	KIAA0593	AB011165	KIAA0593 蛋白 mRNA, cds 部分
149	201987_at		AI984051	甲状腺激素受体相关蛋白, 240 kDa 亚基

150	203605_at	SRP54	NM_003136	信号识别粒子 54kD
151	213226_at		AI346350	多肌炎硬皮病自身抗原 (75kd)
152	205757_at	ENTPD5	NM_001249	外核苷三磷酸盐二磷酸水解酶 5
153	212062_at	KIAA0611	AB014511	KIAA0611 蛋白的 mRNA
154	213007_at		W74442	聚合酶 (DNA 指导的), γ
155	203362_s_at	MAD2L1	NM_002358	MAD2 (有丝分裂延滞缺陷, 酵母, 同源物) 样 1
156	204641_at	NEK2	NM_002497	NIMA (不在有丝分裂基因 a 中) -rel. 激酶 2
157	206983_at	CCR6	NM_004367	趋化因子 (C-C 基序) 受体 6
158	210375_at	EP3a2	X83858	前列腺素 E 受体
159	213933_at	DKFZp586M0723	AW242315	来自克隆 DKFZp586M0723
160	201756_at	RPA2	NM_002946	复制蛋白 A2 (32kD)
161	208688_x_at	eIF3	U78525	真核的翻译起始因子
162	212655_at	KIAA0579	AB011151	KIAA0579 蛋白
163	213124_at		BG538800	DKFZP434N043 蛋白
164	213520_at	RECQL4	NM_004260	RecQ 蛋白样 4
165	218277_s_at	FLJ22060	NM_024612	假定的蛋白 FLJ22060
166	201938_at	DOC1	NM_004642	口腔癌中删除的 (小鼠, 同源物) 1
167	202692_s_at	UBTF	NM_014233	上游结合转录因子, RNA 聚合酶 I
168	203616_at	POLB	NM_002690	聚合酶 (DNA 指导的), β
169	204407_at		AF080255	北极星蛋白
170	206188_at	KIAA0628	NM_014789	KIAA0628 基因产物
171	209831_x_at		AB004574	脱氧核糖核酸酶 II, 溶酶体
172	211980_at		AI922605	胶原蛋白, IV 型, $\alpha 1$
173	214853_s_at		AI091079	(含有 Src 同源性 2 结构域) 转化蛋白 1
174	215888_at	FLJ23236	AK026889	克隆 COL00725
175	216037_x_at		AA664011	转录因子 7 样 2 (T 细胞特异性, HMG-盒)
176	202666_s_at	BAF53A	NM_004301	BAF53

177	204146_at		BE966146	RAD51 相互作用蛋白
178	203920_at	NR1H3	NM_005693	核受体亚族 1, 组 H, 成员 3
179	205322_s_at		AW182367	金属调节转录因子 1
180	206644_at	NR0B1	NM_000475	核受体亚族 0, 组 B, 成员 1
181	201558_at	RAE1	NM_003610	(RNA 输出 1, <i>S. pombe</i>) 同源物
182	209448_at		BC002439	Tat 相互作用蛋白 (30kD)
183	220960_x_at	RPL22	NM_000983	核糖体蛋白 L22
184	207320_x_at	STAU	NM_004602	staufen (果蝇, RNA 结合蛋白) 转录产物变体 T4
185	208948_s_at	MGC:4921	BC000830	MGC:4921
186	213037_x_at		AJ132258	staufen 蛋白, 部分果蝇, RNA 结合蛋白
187	200725_x_at	RPL10	NM_006013	核糖体蛋白 L10
188	200937_s_at	RPL5	NM_000969	核糖体蛋白 L5
189	208696_at	PNAS-102	AF275798	含有 PNAS-102 mRNA chaperonin 的 TCP1, 亚基 5 (epsilon)
190	209593_s_at	FKSG18	AF317129	FKSG18
191	209619_at		K01144	MHC II 抗原 γ 链
192	218336_at	PFDN2	NM_012394	prefoldin 2
193	219390_at	FLJ20731	NM_017946	假定的蛋白 FLJ20731
194	206976_s_at	HSP105B	NM_006644	热激 105kD
195	204444_at	KNSL1	NM_004523	驱动蛋白样 1
196	206364_at	KIAA0042	NM_014875	KIAA0042 基因产物
197	209408_at		U63743	有丝分裂着丝点相关驱动蛋白
198	202629_at	APPBP2	AV681579	淀粉样蛋白 β 前体蛋白 (细胞质尾部) 结合蛋白 2
199	202630_at	APPBP2	AA046411	淀粉样蛋白 β 前体蛋白 (细胞质尾部) 结合蛋白 2
200	202631_s_at	APPBP2	NM_006380	淀粉样蛋白 β 前体蛋白 (细胞质尾部) 结合蛋白 2
201	210629_x_at		AF000425	cLST1A 剪接变体

202	204670_x_at	HLA-DRB5	NM_002125	MHC, II类, DR β 5
203	208306_x_at	HLA-DRB4	NM_021983	MHC, II类, DR β 4
204	206654_s_at	RPC32	NM_006467	聚合酶 (RNA) III (DNA 指导的) (32kD)
205	218360_at	RAB22A	NM_020673	RAB22A, 成员 RAS 癌基因家族
206	209380_s_at	CFTRMRP	AF146074	ABC 蛋白, TP 结合盒, 亚族 C
207	201114_x_at	PSMA7	NM_002792	proteasome (prosome, macropain) 亚基, α 型, 7
208	202243_s_at	PSMB4	NM_002796	proteasome (prosome, macropain) 亚基, β 型, 4
209	202244_at	PSMB4	NM_002796	proteasome (prosome, macropain) 亚基, β 型, 4
210	203878_s_at	MMP11	NM_005940	基质金属蛋白酶 11 (stromelysin 3)
211	216474_x_at		AF206667	肥大细胞 β I 类胰蛋白酶, 选择性剪接的
212	217009_at		AL121974	染色体 6p12-21.3 上来自克隆 RP3-417L20 的 DNA 序列
213	202968_s_at	Dyrk2	Y09216	蛋白激酶的 mRNA, Dyrk2
214	202969_at		AI216690	双特异性酪氨酸 (Y) 磷酸化调节激酶 2
215	204092_s_at	STK15	NM_003600	丝氨酸苏氨酸激酶 15
216	204171_at	RPS6KB1	NM_003161	核糖体蛋白 S6 激酶, 70kD, 多肽 1
217	204825_at	KIAA0175	NM_014791	KIAA0175 基因产物
218	208079_s_at	STK6	NM_003158	丝氨酸苏氨酸激酶 6
219	219148_at	TOPK	NM_018492	PDZ 结合激酶; T 细胞来源的蛋白激酶
220	219813_at	LATS1	NM_004690	LATS (大肿瘤抑制物, 果蝇) 同源物 1
221	202779_s_at	E2-EPF	NM_014501	遍在蛋白载体蛋白
222	217978_s_at	HSA243666	NM_017582	NICE-5 蛋白
223	210413_x_at	SCCA2	U19557	鳞状细胞癌抗原 2
224	219478_at	WFDC1	NM_021197	WAP 四二硫化物核心结构域 1
225	204319_s_at	RGS10	NM_002925	G 蛋白信号调节物 10
226	204017_at	KDEL3	NM_006855	KDEL (Lys-Asp-Glu-Leu) 内质网蛋白保留受体 3

227	206150_at	TNFRSF7	NM_001242	肿瘤坏死因子受体超家族, 成员 7
228	205926_at	WSX-1	NM_004843	I 类细胞因子受体
229	205400_at	WAS	NM_000377	Wiskott-Aldrich 综合症 (湿疹-血小板减少)
230	209539_at	KIAA0006	D25304	KIAA0006 基因的 mRNA
231	221922_at		AW195581	KIAA0761 蛋白
232	200614_at	CLTC	NM_004859	内涵蛋白, 重的多肽 (Hc)
233	202550_s_at	VAPB	NM_004738	VAMP (囊相关膜蛋白) 相关蛋白 B 和 C
234	212159_x_at		AI125280	衔接头相关蛋白复合物 2, $\alpha 2$ 亚基
235	202733_at	P4HA2	NM_004199	前胶原-脯氨酸 2-酮戊二酸 4-加双氧酶 (脯氨酸 4-羟化酶) α 多肽 II
236	208905_at		BC005299	细胞色素 C, 克隆 MGC:12367
237	32137_at	JAG2	AF029778	Jagged2
238	201088_at	KPNA2	NM_002266	亲核蛋白 $\alpha 2$ (RAG 群组 1, 输入蛋白 $\alpha 1$)
239	202824_s_at	TCEB1	NM_005648	转录延伸因子 B (SIII), 多肽 1 (15kD, elongin C)
240	201584_s_at	DDXL	NM_005804	核 RNA 解旋酶, DEAD 盒家族的 DECD 变体
241	218461_at	LOC51184	NM_016301	蛋白 x 0004 (LOC51184)
242	204489_s_at	CD44	NM_000610	CD44
243	204490_s_at	CDw44	M24915	CDw44 抗原
244	207165_at	HMMR	NM_012485	透明质酸介导的运动受体 (RHAMM)
245	210916_s_at	CD44	AF098641	CD44 同种型 RC
246	212063_at	CMPX1	BE903880	锌指蛋白 6
247	204470_at	GRO1	NM_001511	GRO1 癌基因 (黑素瘤生长刺激活性, α)
248	207430_s_at	MSMB	NM_002443	microseminoprotein, β
249	210297_s_at		U22178	前列腺分泌蛋白 57
250	213009_s_at	FLJ12639	AK022701	cDNA FLJ12639

251	203536_s_at	CIAO1	NM_004804	WD40 蛋白 Ciao1
252	204026_s_at	ZWINT	NM_007057	ZW10 互作因子
253	204435_at	KIAA0410	NM_014778	KIAA0410 基因产物
254	209271_at		AB032251	溴结构域 PHD 指转录因子的 BPTF mRNA
255	212074_at	KIAA0810	BE972774	KIAA0810 蛋白
256	218009_s_at	PRC1	NM_003981	胞质分裂 1 的蛋白调节物
257	218768_at	NUP107	NM_020401	核孔复合体蛋白
258	M33197_3_at	GAPDH	M33197	甘油醛-3-磷酸脱氢酶
259	203524_s_at	MPST	NM_021126	巯基丙酮酸硫酸转移酶
260	206335_at	GALNS	NM_000512	半乳糖胺 (N-乙酰基) -6-硫酸盐硫酸酯酶 (Morquio 综合症, 粘多糖病 IVA 型)
261	203503_s_at	PEX14	NM_004565	过氧化物酶体发生因子 14
262	202673_at	DPM1	NM_003859	长-磷酸盐甘露糖基转移酶多肽 1, 催化亚基 1
263	207543_s_at	P4HA1	NM_000917	前胶原-脯氨酸, 2-酮戊二酸 4-加双氧酶 (脯氨酸 4 羟化酶), α 多肽 I
264	204192_at	CD37	NM_001774	CD37 抗原
265	204960_at	PTPRCAP	NM_005608	蛋白酪氨酸磷酸酶, 受体型, C 相关蛋白
266	211991_s_at		M27487	MHC II 类 DPw3- α -1 链
267	200822_x_at	TPI1	NM_000365	磷酸丙糖异构酶 1
268	219502_at	FLJ10858	NM_018248	假定的蛋白 FLJ10858
269	219499_at	FLJ10578	NM_018144	假定的蛋白 FLJ10578
270	209238_at		BE966922	syntaxin 3A
271	212593_s_at		N92498	
272	201598_s_at	INPPL1	NM_001567	环己六醇多磷酸磷酸酶样 1
273	201760_s_at	LOC55884	NM_018639	含 CS 盒的 WD 蛋白
274	222077_s_at		AU153848	GTPase 活化蛋白
275	203764_at	KIAA0008	NM_014750	KIAA0008 基因产物
276	59705_at		AA911739	

277	204070_at	RARRES3	NM_004585	视黄酸受体效应器 (tazarotene 诱导的) 3
278	212149_at		AW470003	
279	214039_s_at	DKFZp586E1 124	T15777	DKFZp586E1124
280	217956_s_at	MASA	NM_021204	E-1 酶
281	200994_at		BG291787	RAN 结合蛋白 7
282	200995_at		AI741392	
283	205106_at	MTCP1	NM_014221	成熟 T 细胞增殖 1
284	209035_at	hMK-1	M69148	Midkine (生长促进因子 2)
285	210311_at	FGF5	AF171928	成纤维细胞生长因子 5 短变体
286	211251_x_at	NFY-C	U78774	NFY-C mRNA
287	220406_at	TGFB2	NM_003238	转化生长因子, β 2
288	206967_at	CCNT1	NM_001240	cyclin T1
289	204252_at		M68520	cdc2 相关蛋白激酶 mRNA
290	205955_at	FLJ11136	NM_018336	假定的蛋白 FLJ11136
291	203258_at	DRAP1	NM_006442	DR1-assoc. 蛋白 1 (阴性辅助因子 2 α)
292	204022_at		AI668780	
293	207660_at		NM_004019	营养不良蛋白 (肌营养不良, Duchenne 和 Becker 型),
294	215050_x_at		BG325734	促分裂素活化蛋白激酶-活化的蛋白激酶 2
295	204337_at		AL514445	G 蛋白信号调节物 4
296	217687_at		AA224446	
297	221546_at		BC000794	类似于 <i>S. cerevisiae</i> Prp18 的前 mRNA 剪接因子
298	206546_at	SYCP2	NM_014258	联会复合体蛋白 2
299	206278_at		D10202	血小板活化因子受体的 mRNA
300	206429_at	F2RL1	NM_005242	凝结因子 II (凝血酶) 受体样 1
301	216408_at		AJ302584	嗅感受蛋白的基因, 细胞系 BM28.7
302	221306_at	GPR27	NM_018971	G 蛋白联结受体 27
303	221442_at	MC3R	NM_019888	melanocortin 3 受体

304	201446_s_at		BF692742	
305	205018_s_at		NM_005757	C3H型锌指蛋白;类似于D. melanogaster muscleblind B 蛋白 (MBLL)
306	214379_at		AI954458	
307	214698_at		AW190873	
308	219336_s_at		NM_015947	CGI-18 蛋白 (LOC51008)
309	220760_x_at	FLJ14345	NM_024733	假定的蛋白 FLJ14345
310	221480_at		BG180941	
311	221615_at	PPIE	AF104013	肽酰-脯氨酰顺-反异构酶 E
312	221923_s_at		AA191576	nucleophosmin (核仁磷蛋白 B23, numatrin)
313	201211_s_at	DDX3	AF061337	DEAD 盒 RNA 解旋酶 DDX3
314	205638_at	BAI3	NM_001704	脑特异性血管生成抑制物 3
315	205881_at	ZNF74	NM_003426	锌指蛋白 74 (Cos52)
316	206179_s_at		NM_007030	脑特异性蛋白质 p25 α (p25)
317	206308_at		AJ223333	推定的 DNA 甲基转移酶的 mRNA
318	207361_at	HBP1	NM_012257	含 HMG 盒的蛋白 1
319	208902_s_at		BF431363	
320	209603_at		AI796169	
321	214174_s_at		BE043700	
322	215747_s_at		X06130	细胞周期基因 RCC1 的 mRNA
323	216480_x_at	AF10CALM	AF060927	I 型 AF10CALM 融合蛋白
324	216711_s_at		M73444	CCG1p mRNA
325	222115_x_at		BC003693	类似于 RIKEN cDNA 3930401K13 基因
326	221686_s_at	DKFZp434J0450	AL136869	DKFZp434J0450
327	210533_at	MSH4	AF104243	减数分裂特异性 MutS 同源物
328	217485_x_at	hPMS3	D38435	hPMS3 mRNA
329	202162_s_at		AI769416	CCR4-NOT 转录复合物, 亚基 8
330	202401_s_at	SRF	NM_003131	血清响应因子 (c-fos 血清响应元件结合转录因子)

331	206067_s_at		NM_024426	Wilms 瘤 1 (WT1), 转录产物变体 D
332	206127_at	ELK3	NM_005230	ETS 结构域蛋白 (SRF 附属蛋白 2)
333	207402_at	ZNF132	NM_003433	锌指蛋白 132
334	207768_at	EGR4	NM_001965	早期生长反应 4
335	208414_s_at	HOXB3	NM_002146	相似盒 B3
336	214879_x_at		AY007087	克隆 TCCCIA00046, 上游转录因子 2, c-fos 相互作用
337	219314_s_at	ZNF219	NM_016423	锌指蛋白 219
338	219779_at	FLJ20980	NM_024721	假定的蛋白 FLJ20980
339	220653_at	ZIM2	NM_015363	锌指, 印记 2
340	219778_at	FOG2	NM_012082	GATA2 的朋友
341	203947_at	CSTF3	NM_001326	裂解刺激因子, 3 前 RNA, 亚基 3, 77kD
342	220096_at	FLJ20378	NM_017795	假定的蛋白 FLJ20378
343	201326_at		BE737030	chaperonin cont ' ing TCP1, 亚基 6A (zeta 1)
344	206769_at	TMSB4Y	NM_004202	胸腺素, β 4, Y 染色体
345	211197_s_at	KIAA0653	AL355690	来自克隆 34465 的 EST, 完全插入物
346	204994_at	MX2	NM_002463	粘病毒 (流感) 抗性 2, 鼠同源物
347	201662_s_at		D89053	Acyl-CoA 合成酶 3
348	206141_at	MOCS3	NM_014484	molybdopterin 合酶硫酸化酶
349	209992_at		AB044805	6-phosphofructo-2-激酶心脏同种型
350	210160_at		BC000398	血小板活化因子乙酰基水解酶, 同种型 Ib, β 亚基 (30kD),
351	218016_s_at	FLJ10509	NM_018119	假定的蛋白 FLJ10509
352	220582_at	FLJ12190	NM_025071	假定的蛋白 FLJ12190
353	222294_s_at		AW971415	
354	202239_at	ADPRTL1	NM_006437	ADP-核糖基转移酶 (NAD ⁺ ; 聚 (ADP-核糖) 聚合酶) 样 1
355	205342_s_at		AF026303	磺基转移酶家族, cytosolic, 1C, 成员 1
356	202294_at		AI126490	
357	201597_at	COX7A2	NM_001865	Cyt C 氧化酶亚基 VIIa 多肽 2 (肝脏)

358	206353_at	COX6A2	NM_005205	Cyt C 氧化酶亚基 VIa 多肽 2
359	218739_at	LOC51099	NM_016006	CGI-58 蛋白
360	217557_s_at		AV710357	
361	202413_s_at	USP1	NM_003368	遍在蛋白特异性蛋白酶 1
362	213661_at		AI671186	DKFZP586H2123 蛋白
363	212729_at		AI916274	KIAA1232 蛋白
364	202951_at		BE048506	丝氨酸苏氨酸蛋白激酶
365	207667_s_at	MAP2K3	NM_002756	促分裂素活化蛋白激酶 3
366	212565_at		BE302191	KIAA0965 蛋白
367	212740_at		BF740111	磷酸肌醇-3-激酶, 调节亚基 4, p150
368	213490_s_at		AI762811	促分裂素活化蛋白激酶 2
369	213595_s_at	KIAA0451	AA127643	KIAA0451 基因产物
370	220640_at	CSNK1G1	NM_022048	酪蛋白激酶 1, γ 1
371	207569_at	ROS1	NM_002944	v-ros 禽类 UR2 肉瘤病毒癌基因同源物 1
372	209041_s_at		BG395660	遍在蛋白共轭物酶 E2G 2
373	207214_at	PEC-60	NM_014471	胃肠肽
374	214425_at		AV645756	α -1-microglobulinbikunin 前体
375	203650_at	EPCR	NM_006404	蛋白 C 受体, 内皮的
376	210733_at	PRO1292	AF130055	FLB4941 PRO1292, 转位链相关膜蛋白
377	220056_at	IL22R	NM_021258	白细胞介素 22 受体
378	205049_s_at	CD79A	NM_001783	CD79A 抗原 (免疫球蛋白相关 α) 转录产物变体 1
379	211245_x_at		AF002256	杀伤细胞抑制的受体同源物 cl-9 mRNA, FEAn 样受体, 两个结构域, 长细胞质尾部, 4
380	212128_s_at		AW411370	dystroglycan 1 (dystrophin 相关糖蛋白 1)
381	205019_s_at	VIPR1	NM_004624	血管活性肠肽受体 1
382	206001_at	NPY	NM_000905	神经肽 Y
383	216289_at	MAMMA100 2427	AU148039	MAMMA1002427
384	208250_s_at	DMBT1	NM_004406	在恶性脑肿瘤中的删除 1

385	202091_at		BC003087	Arl Two 的粘合物, 克隆 MGC:1121
386	205068_s_at		BE671084	与病灶粘附激酶 pp125 相关的 GTPase 调节物 (FAK)
387	221136_at	GDF2	NM_016204	生长变异因子 2
388	202688_at	TNFSF10	NM_003810	肿瘤坏死因子 (配体) 超家族, 成员 10
389	214336_s_at	TNFSF10	AI621079	coatomer 蛋白质复合物, 亚基 α , TNF (配体) 超家族, 成员 10
390	213055_at		BF693956	CD47 抗原 (Rh 相关抗原, 整联蛋白相关信号传感器)
391	201719_s_at	EPB41L2	NM_001431	红血球膜蛋白带 4.1 样 2
392	208353_x_at	ANK1	NM_020480	锚蛋白 1, 红血球的转录产物变体 7
393	215717_s_at		X62009	肌原纤维蛋白 5 的部分 mRNA
394	206826_at	PMP2	NM_002677	周围的 = 髓鞘质蛋白 2
395	207542_s_at	AQP1	NM_000385	aquaporin 1 (通道形成整合蛋白, 28kD)
396	207596_at	PRO2176	NM_018515	假定的蛋白 PRO2176
397	208297_s_at	EVIS	NM_005665	亲嗜性病毒整合位点 5
398	217289_s_at	G6PT	AF097831	葡糖-6-磷酸转运蛋白
399	205972_at		NM_006841	转运蛋白; 系统 N1 联结 Na^+ 和 H^+ 的谷氨酰胺转运蛋白 (G17)
400	218835_at	SFTPA2	NM_006926	表面活性剂, 肺部相关蛋白 A2
401	219716_at	APOL6	NM_030641	载脂蛋白 L, 6
402	218928_s_at	SLC37A1	NM_018964	溶质载体家族 37 (甘油-3-磷酸盐转运蛋白), 成员 1
403	214205_x_at	FLJ12069	AK022131	FLJ12069
404	205008_s_at	KIP2	NM_006383	DNA 依赖性蛋白激酶催化亚单位相互作用蛋白 2
405	211272_s_at	DAGK1	AF064771	克隆 24 甘油二酯激酶 α
406	206316_s_at	KIAA0166	NM_014708	KIAA0166 基因产物
407	209737_at	KIAA0705	AB014605	KIAA0705 蛋白
408	213117_at	FLJ10262	AW138594	假定的蛋白 FLJ10262
409	217161_x_at		X17406	软骨特异性蛋白多糖

410	209436_at	KIAA0762	AB018305	KIAA0762 蛋白
411	213993_at		AI885290	spondin 1, (f-spondin) 细胞外基质蛋白
412	214354_x_at		T91506	N-酰基鞘氨醇氨基水解酶 (酸性神经酰胺酶) 样
413	205799_s_at		M95548	氨基酸转运蛋白
414	213522_s_at		AA527578	溶质载体家族 16 (一元羧酸转运蛋白), 成员 3
415	212393_at		AL096767	染色体 22 上来自克隆 579N16 的 DNA 序列
416	219179_at	LOC51339	NM_016651	肝细胞癌新基因-3 蛋白
417	201502_s_at		AI078167	B 细胞抑制物中 kappa 轻多肽基因增强子的核因子, α
418	201512_s_at		BC003633	外线粒体膜 70 的移位酶 (酵母) 同源物 A
419	201576_s_at	GLB1	NM_000404	半乳糖苷酶, β 1
420	201765_s_at		AL523158	
421	215155_at	HEXA	J04178	异常 β -氨基己糖苷酶 α 链
422	203518_at	CHS1	NM_000081	Chediak-Higashi 综合症 1
423	220801_s_at	HAO2	NM_016527	醇酸氧化酶 2 (长链)
424	204690_at	STX8	NM_004853	syntaxin 8
425	201126_s_at	MGAT1	NM_002406	甘露糖基 (α -3-)-糖蛋白 β -2-N-乙酰基 glucosaminyl 转移酶
426	206925_at	SIAT8D	NM_005668	唾液酸转移酶 8 (α -2 8-聚唾液酸转移酶) D
427	205319_at	PSCA	NM_005672	前列腺干细胞抗原
428	206199_at	CEACAM7	NM_006890	癌胚抗原 rel. 细胞粘附分子 7
429	207695_s_at	IGSF1	NM_001555	免疫球蛋白超家族, 成员 1
430	219249_s_at	FLJ22041	NM_021939	假定的蛋白 FLJ22041
431	213413_at	FLJ13555	BG434174	FLJ13555
432	219793_at	SNX16	NM_022133	sorting nexin 16
433	212035_s_at	KIAA1067	AI817079	KIAA1067 蛋白

434	202730_s_at	PDCD4	NM_014456	程序化细胞死亡 4
435	202130_at	sudD	AA725102	bimD6 的抑制物, <i>Aspergillus nidulans</i> 同源物
436	206623_at	PDE6A	NM_000440	磷酸二酯酶 6A, cGMP 特异性, rod, α
437	222201_s_at	KIAA1315	AB037736	KIAA1315 蛋白
438	221752_at	KIAA1298	AL041728	KIAA1298 蛋白
439	205003_at	KIAA0716	NM_014705	KIAA0716 基因产物
441	205006_s_at	NMT2	NM_004808	N-肉豆蔻酰转移酶 2
442	205677_s_at	DLEU1	NM_005887	淋巴细胞白血症中的删除, 1
443	204947_at	E2F1	NM_005225	E2F 转录因子 1
444	207505_at	PRKG2	NM_006259	蛋白激酶, cGMP 依赖性, II 型
445	211036_x_at		BC006301	后期促进复合物亚基 5, 克隆 MGC:13295,
446	211269_s_at		K03122	白细胞介素-2 受体 mRNA (短形式)
447	202610_s_at	TRAP170	AF135802	甲状腺激素受体相关蛋白质复合物成分
448	204762_s_at		BE670563	鸟苷酸结合蛋白 (G 蛋白), α 活化活性多肽 O
449	201313_at	ENO2	NM_001975	烯醇酶 2, (γ , 神经元的)
450	203225_s_at	FLJ11149	NM_018339	假定的蛋白 FLJ11149
451	221567_at	NOP	AF064599	nucleolar 蛋白 Nop30
452	201518_at	CBX1	NM_006807	chromobox 同源物 1 (果蝇 HP1 β)
453	204970_s_at		NM_002359	v-maf 肌腱膜的纤维肉瘤 (禽类) 癌基因家族, 蛋白 G (MAFG),
454	214615_at	P2Y10	NM_014499	推定的 purinergic 受体
455	221386_at	OR3A2	NM_002551	嗅感受蛋白家族 3 亚族 A, 成员 2
456	200014_s_at	HNRPC	NM_004500	核不均一核糖核蛋白 C (C1C2)
457	200053_at	SPAG7	NM_004890	精子相关抗原 7
458	203462_x_at	EIF3S9	NM_003751	翻译起始因子 3 sub 9 (eta 116kD)
459	205917_at	ZNF264	NM_003417	锌指蛋白 264
460	207753_at	ZNF304	NM_020657	锌指蛋白 304

461	209751_s_at		AF291676	MBP-1 相互作用蛋白-2A
462	217403_s_at	BC228680	AC074331	染色体 19, BAC CIT-HSPC_204F22
463	219571_s_at	GIOT-3	NM_016265	促性腺激素可诱导转录阻遏物-3 的 GIOT-3
464	203119_at	MGC2574	NM_024098	假定的蛋白 MGC2574
465	203721_s_at	LOC51096	NM_016001	CGI-48 蛋白
466	204880_at	MGMT	NM_002412	O-6-甲基鸟嘌呤-DNA 甲基转移酶
467	205664_at	KIN	NM_012311	recA 蛋白的抗原决定簇 (小鼠) 同源物
468	212597_s_at		AL079310	定位到染色体 22 的新的类基因
469	212759_s_at		AI703074	转录因子 7 样 2 (T 细胞特异性, HMG- 盒)
470	213824_at		AF221520	蛋白激酶 C 结合蛋白 2
471	218137_s_at	FLJ13159	NM_021940	假定的蛋白 FLJ13159
472	220445_s_at	TRAG3	NM_004909	紫杉酚抗性相关基因 3
473	203720_s_at		NM_001983	切除修复交叉互补啮齿动物修复缺陷, 互补群 1
474	201046_s_at	RAD23A	NM_005053	RAD23 (S. cerevisiae) 同源物 A
475	202344_at	HSF1	NM_005526	热激转录因子 1
476	202580_x_at	FOXO1	NM_021953	叉头盒 M1
477	205690_s_at	G10	NM_003910	母系的 G10 转录产物
478	206307_s_at	FOXO1	NM_004472	叉头盒 D1
479	213090_s_at		AI744029	TATA 盒结合蛋白 (TBP) 相关因子
480	218215_s_at	NR1H2	NM_007121	核受体亚族 1, 组 H, 成员 2
481	207469_s_at	PIR	NM_003662	Pirin
482	209062_x_at	RAC3	AF010227	受体相关共活化物 3
483	200823_x_at	RPL29	NM_000992	核糖体蛋白 L29
484	202868_s_at	POP4	NM_006627	POP4 (前体的加工, S. cerevisiae) 同 源物
485	217747_s_at	RPS9	NM_001013	核糖体蛋白 S9
486	32723_at		L02547	(克隆 pZ50-19) 裂解刺激因子 50kDa 亚基

487	212105_s_at		BE910323	DEADH (Asp-Glu-Ala-AspHis) 盒多肽 9
488	207320_x_at		NM_004602	staufen (果蝇, RNA 结合蛋白) (STAU), 转录产物 v. T4,
489	200081_s_at		BE741754	核糖体蛋白 S6
490	217559_at		AI001784	高度类似于 A42735 核糖体蛋白 L10, cytosolic
491	217907_at	HSPC071	NM_014161	HSPC071 蛋白
492	213504_at		W63732	COP9 亚基 6 (MOV34 同源物, 34 kD)
493	204102_s_at	EEF2	NM_001961	真核的翻译延伸因子 2
494	208696_at		AF275798	PNAS-102
495	209275_s_at	CLN3	AF015593	CLN3 蛋白
496	214591_at		BF215673	果蝇 Kelch 样蛋白
497	201642_at	IFNGR2	NM_005534	干扰素 γ 受体 2 (干扰素 γ 转导物 1)
498	219080_s_at	CTPS2	NM_019857	CTP 合成酶 II
499	202282_at	HADH2	NM_004493	羟基酰基辅酶 A 脱氢酶 II 型
500	201453_x_at	RHEB2	NM_005614	在脑中富集的 Ras 同源物 2
501	208733_at		NM_002865	RAB2, 成员 RAS 癌基因家族
502	211004_s_at		BC002553	类似于醛脱氢酶 7
503	201400_at	PSMB3	NM_002795	蛋白酶体 (prosome, macropain) sub β 3 型
504	208799_at		BC004146	蛋白酶体 (prosome, macropain) sub β 5 型
505	216088_s_at		AL078633	染色体 20 上来自 RP5-1005f21 的 DNA 序列
506	213912_at	KIAA0984	AW134976	KIAA0984 蛋白
507	213913_s_at	KIAA0984	AW134976	KIAA0984 蛋白
508	205356_at	USP13	NM_003940	遍在蛋白 spec 蛋白酶 13 (异肽酶 T-3)
509	208166_at	MMP16	NM_022564	基质金属蛋白酶 16 (膜-插入的) 转录产 物 v. 2,
510	202018_s_at	LTF	NM_002343	乳运铁蛋白

511	213829_x_at	FLJ20478	AK000485	FLJ20478
512	212766_s_at		AW294587	Similar to 假定的蛋白 FLJ12484,
513	202969_at		Y09216	双特异性酪氨酸 (Y) 磷酸化调节激酶 2
514	203218_at		W37431	促分裂素活化蛋白激酶 9
515	204171_at	RPS6KB1	NM_003161	核糖体蛋白 S6 激酶,
516	204906_at		BC002363	核糖体蛋白 S6 激酶, 90kD, 多肽 2
517	205126_at	VRK2	NM_006296	牛痘相关激酶 2
518	214716_at	DKFZp434P0116	AW504018	假定的蛋白 DKFZp434P0116
519	37170_at	HRIHFB2017	AB015331	HRIHFB2017
520	211474_s_at		BC004948	MGC:10846,
521	204237_at	CED-6	NM_016315	CED-6 蛋白
522	203005_at	LTBR	NM_002342	淋巴细胞毒素 β 受体 (TNFR 超家族, 成员 3)
523	211000_s_at	gp130-RAPS	AB015706	带有类风湿性关节炎抗原肽的 gp130, 可溶形式
524	202679_at	NPC1	NM_000271	Niemann-Pick 疾病, C1 型
525	205282_at	LRP8	NM_004631	低密度脂蛋白受体相关蛋白 8, 载脂蛋白 e 受体
526	207270_x_at	CMRF35	NM_006678	CMRF35 白细胞免疫球蛋白样受体
527	211846_s_at		AF110314	疱疹病毒免疫球蛋白样受体 HIgR
528	201810_s_at		AL562152	SH3 结构域结合蛋白 5 (BTK-assoc)
529	213684_s_at		BF671400	LIM 蛋白 (类似于大鼠蛋白激酶 C 结合 enigma)
530	206636_at	RASA2	NM_006506	RAS p21 蛋白活化物 2
531	219907_at	SNT-2	NM_006653	suc1 相关神经营养因子靶点 2 (FGFR 信号连接物)
532	201281_at	GP110	NM_007002	细胞膜糖蛋白, 110000M (r) (表面抗原)
533	209462_at		U48437	淀粉样蛋白前体样蛋白 1
534	210880_s_at		AB001467	Efs2,

535	212567_s_at		AL523310	CS0DC001YN06 (3 prime)
536	202222_s_at	DES	NM_001927	结蛋白
537	219615_s_at	KCNK5	NM_003740	钾通道, 亚族 K, 成员 5 (TASK-2)
538	214210_at		AL049764	染色体 22q13.1-13.31 上的克隆 RP3-362J20
539	52078_at		AI828080	
540	203765_at	GCL	NM_012198	grancalcin
541	213308_at		AB028945	cortactin SH3 结构域结合蛋白
542	203082_at	KIAA0187	NM_014753	KIAA0187 基因产物
543	212543_at	AIM1	U83115	非晶状 β γ -晶状体蛋白样蛋白
544	209709_s_at	RHAMM	U29343	hyaluronan 受体
545	205807_s_at	TUFT1	NM_020127	tuftelin 1
546	213895_at		BF445047	上皮膜蛋白 1
547	201565_s_at	ID2	NM_002166	DNA 结合 2 的抑制物, 主要阴性螺旋-环-螺旋蛋白
548	202639_s_at		AI689052	RAN 结合蛋白 3
549	209230_s_at	COM1	AF135266	p8 蛋白同源物
550	218626_at	4E-T	NM_019843	eIF4E 转运蛋白
551	221506_s_at		BG258639	亲核蛋白 β 2b, 转运蛋白
552	221521_s_at		BC003186	HSPC037 蛋白, 克隆 MGC:673,
553	202069_s_at		AI826060	异柠檬酸脱氢酶 3 (NAD ⁺) α
554	203524_s_at	MPST	NM_021126	巯基丙酮酸硫酸转移酶
555	201916_s_at	SEC63L	NM_007214	SEC63, 内质网转位子成分 (S. cerevisiae) 样
556	217776_at	ARSDR1	AF167438	雄激素调节短链 dehydrogenasereductase 1
557	218623_at	LOC51617	NM_015980	HMP19 蛋白
558	208777_s_at		AF001212	26S 蛋白酶体亚基 9
559	203971_at	SLC31A1	NM_001859	溶质载体家族 31 (铜转运蛋白), 成员 1
560	215243_s_at		AF099730	连接蛋白 31 (GJB3) 基因,
561	201591_s_at		NM_007184	咪唑啉受体候选物 (I-1),

562	205466_s_at	HS3ST1	NM_005114	硫酸乙酰肝素(葡糖胺)3-O-磺基转移酶1
563	207829_s_at	BNIP1	NM_013978	Bcl-2 腺病毒 E1B 19kD 相互作用蛋白 1 转录产物变体 BNIP1-a,
564	205742_at	TNNI3	NM_000363	肌钙蛋白 I, 心脏的
565	211178_s_at		AF038602	CD2 结合蛋白 1 短形式
566	219113_x_at	LOC51171	NM_016246	视网膜短链 dehydrogenasereductase retSDR3
567	217807_s_at	GLTSCR2	NM_015710	胶质瘤抑制物候选物区域基因 2
568	206133_at	HSXIAPAF1	NM_017523	XIAP 相关因子-1
569	201379_s_at	TPD52L2	NM_003288	肿瘤蛋白 D52 样 2
570	209373_at		BC003179	MGC:4419,
571	210142_x_at		AF117234	flotillin
572	218942_at	FLJ22055	NM_024779	假定的蛋白 FLJ22055
573	216716_at		U15197	组血型 ABO 蛋白, 部分 3 UTR 序列
574	218185_s_at	FLJ10511	NM_018120	假定的蛋白 FLJ10511
575	219215_s_at	FLJ20327	NM_017767	假定的蛋白 FLJ20327
576	214683_s_at		AI251890	CDC 样激酶 1
577	202612_s_at	CRSP2	NM_004229	Sp1 转录活化所需的辅助因子, 亚基 2 (150kD)
578	208932_at		BC001416	蛋白磷酸酶 4 (以前的 X), 催化亚单位,
579	206766_at	ITGA10	AF112345	整联蛋白 α 10 亚基
580	216213_at		AF155113	NY-REN-55 抗原
581	205609_at	ANGPT1	NM_001146	angiopoietin 1
582	204644_at	tNOX	AF207881	肿瘤相关对苯二酚 (NADH) 氧化酶
583	206812_at	ADRB3	NM_000025	肾上腺素的, β -3-, 受体
584	202176_at	ERCC3	NM_000122	切除修复交叉-互补啮齿动物修复缺陷, 互补群 3 (着色性干皮病 B 组互补)
585	219816_s_at	FLJ10482	NM_018107	假定的蛋白 FLJ10482
586	200934_at	DEK	NM_003472	DEK 癌基因 (DNA 结合)
587	206098_at	ZID	NM_006626	具有相互作用结构域的锌指蛋白

588	206987_x_at	FGF18	NM_003862	成纤维细胞生长因子 18
589	208392_x_at	IFI75	NM_004510	干扰素诱导蛋白 75, 52kD
590	213048_s_at		W26593	SET 转位 (骨髓性白血病相关)
591	220861_at		AF118067	PRO1578
592	222115_x_at		BC003693	类似于 RIKEN cDNA 3930401K13 基因,
593	222236_s_at	FLJ20246	AK000253	FLJ20246
594	41644_at	KIAA0790	AB018333	KIAA0790 蛋白,
595	204107_at		NM_002505	核转录因子 Y, α
596	209706_at		AF247704	间源盒蛋白 NKX3.1
597	210174_at		AF228413	肝细胞转录因子
598	213844_at	HOXA5	NM_019102	相似盒 A5
599	214108_at		AI346181	MAX 蛋白
600	212064_x_at		AI471665	MYC-相关锌指蛋白 (嘌呤结合转录因子)
601	214316_x_at		AI348935	calreticulin
602	201514_s_at	G3BP	NM_005754	Ras-GTPase-活化蛋白 SH3 结构域结合蛋白
603	203092_at	hTIM44	AF026030	推定的线粒体内膜蛋白输入受体
604	208725_at		AL031668	染色体 20q11.21-11.23 上来自克隆 RP1-64K7 的 DNA 序列
605	219819_s_at	HSPC007	NM_014018	HSPC007 蛋白
606	200724_at		BC003358	核糖体蛋白 L10,
607	214350_at		AI762021	Syntrophin, β 2 (dystrophin 相关蛋白 A1, 59kD, 基本成分 2)
608	210197_at		BC003622	类似于 1,3,4-三磷酸盐 56 激酶,
609	218150_at	ARL5	NM_012097	ADP-核糖基化因子样 5
610	214241_at		AA723057	NADH 脱氢酶 (泛醌) 1 β 子复合物, 8 (19kD, ASH1)
611	206253_at	DLG2	NM_001364	discs large (Dros) 同源物 2 (chapsyn-110)

612	205435_s_at	KIAA1048	NM_014911	KIAA1048 蛋白
613	215154_at	DKFZp434G043	AL080134	DKFZp434G043
614	215771_x_at		X15786	ret-II 基因
615	213795_s_at		BF740139	克隆 RP4-534B8 染色体 20
616	209042_s_at		BC001738	类似于遍在蛋白共轭酶 E2G 2 (与酵母 UBC7 同源),
617	201602_s_at		NM_002480	肌球蛋白磷酸酶, 靶点亚基 1
618	211133_x_at		AF009643	克隆 6 免疫球蛋白样转录产物 5 蛋白
619	211245_x_at		AF002256	杀伤细胞抑制受体同源物 cl-9
620	210713_at		U61166	含有 SH3 结构域的蛋白 SH3P17
621	210840_s_at	KIAA0051	D29640	KIAA0051
622	218157_x_at	SPEC1	NM_020239	Cdc42 的小蛋白效应物 1
623	208292_at	BMP10	NM_014482	骨骼形态发生蛋白 10
624	210518_at		AB035305	cadherin-8 的 CDH8 mRNA,
625	205064_at	SPRR1B	NM_003125	富含小脯氨酸蛋白 1B (cornifin)
626	206990_at	TNR	NM_003285	tenascin R (restrictin, janusin)
627	211304_x_at		D50134	内调整 K 通道,
628	203950_s_at	CLC-6a	NM_001286	氯化物通道 6 (CLCN6), 转录产物变体
629	214211_at		AA083483	铁蛋白, 重多肽 1
630	211056_s_at		BC006373	克隆 MGC:12762
631	212185_x_at	MT2A	NM_005953	金属硫蛋白 2A
632	203415_at	PDCD6	NM_013232	程序化细胞死亡 6
633	205388_at	TNNC2	NM_003279	肌钙蛋白 C2, 快
634	208871_at		D31840	DRPLA
635	218840_s_at	FLJ10631	NM_018161	假定的蛋白 FLJ10631
636	206400_at	LGALS7	NM_002307	凝集素半乳糖苷结合可溶的 7 (galectin 7)
637	207733_x_at	PSG9	NM_002784	妊娠特异性 β -1-糖蛋白 9
638	207850_at	GRO3	NM_002090	GRO3 癌基因
639	210884_s_at		AF168619	HE2 gamma1

640	220162_s_at	LOC64170	NM_022352	caspase 补充结构域 9
641	209749_s_at		AI623989	血管紧张素 I 转化酶 (肽酰-二肽酶 A) 1
642	216010_x_at		D89324	α (1,31,4) 岩藻糖基转移酶的 DNA,
643	203970_s_at	PEX3	NM_003630	过氧化物酶体发生因子 3
644	203972_s_at		AB035307	Pex3p 的 mRNA,
645	217749_at	LOC51137	NM_016128	外被蛋白 γ -cop
646	207022_s_at	LDHC1	NM_002301	乳酸脱氢酶 C, 转录产物变体 1
647	209905_at		AI246769	相似盒 A9
648	209099_x_at	HJ1	U73936	Jagged 1
649	222201_s_at	KIAA1315	AB037736	KIAA1315 蛋白,
650	201335_s_at	KIAA0382	NM_015313	KIAA0382 蛋白;
651	221748_s_at	tensin	AL046979	tensin
652	205139_s_at	UST	NM_005715	uronyl 2-磺基转移酶
653	210678_s_at		U56418	溶血磷脂酸酰基转移酶- β
654	215293_s_at	DKFZp564E0 53	AL049261	DKFZp564E053
655	41858_at	DKFZp564E0 53	AL049261	DKFZp564E053
656	218047_at	FLJ12492	NM_024586	假定的蛋白 FLJ12492
657	215069_at	FLJ21412	AK025065	FLJ21412
658	201095_at	DAP	NM_004394	死亡相关蛋白
659	220197_at	ATP6N1B	NM_020632	ATPase, H (+) 转运, 溶酶体, 非催化 附属蛋白 1B
660	213351_s_at	KIAA0779	AI934469	KIAA0779
661	203285_s_at	HS2ST1	NM_012262	硫酸乙酰肝素 2-O-磺基转移酶
662	204326_x_at	MT1L	NM_002450	金属硫蛋白 1L
663	208581_x_at	MT1X	NM_005952	金属硫蛋白 1X
650	212690_at	KIAA0725	AB018268	KIAA0725 蛋白,

引用的参考文献

5,242,974	20030069180	WO01053312	WO02094988	WO04030615
5,384,261	20030073623	WO01054472	WO02095010	WO04031413
5,405,783	20030087818	WO01055322	WO02102235	WO04031414
5,412,087	20030096982	WO01057182	WO02102993	WO04033637
5,424,186	20030104366	WO01057188	WO02102994	WO04037996
5,429,807	20030124128	WO01059063	WO03000735	WO04039943
5,436,327	20030124148	WO01060860	WO03003906	WO04039956
5,445,934	20030128884	WO01070979	WO03004622	WO04041170
5,472,672	20030166064	WO01073027	WO03004989	WO04042022
5,527,681	20030194704	WO01075067	WO03010336	WO04043361
5,529,756	20030219744	WO01077168	WO03016475	WO04047728
5,532,128	20030219767	WO01077288	WO03016476	WO04048938
5,545,531	20030236392	WO01077291	WO03016500	WO04053081
5,554,501	20040005560	WO01088088	WO03018807	WO04060270
5,556,752	20040009478	WO01090154	WO03025138	WO04063355
5,561,071	20040009491	WO01090304	WO03027258	WO04070062
5,571,639	20040013663	WO01094629	WO03038063	WO04072265
5,593,839	20040034196	WO02002623	WO03042661	WO04074301
5,599,695	20040037842	WO02004514	WO03042661	WO04076613
5,624,711	20040058340	WO02012235	WO03042661	WO04079014
5,658,734	20040110194	WO02012328	WO03046152	WO91001999
5,700,637	20040115636	WO02024956	WO03054152	WO92019745
5874283	DE 432977	WO02028999	WO03062379	WO94023039
6,004,755	EP 1104808	WO02029086	WO03064589	WO95014772
6060283	WO00021991	WO02031198	WO03072035	WO97012967
6,136,182	WO00026244	WO02036766	WO03078572	WO97024441
6171787	WO00032776	WO02050279	WO03081201	WO97033902
6,218,114	WO00036107	WO02057414	WO03091388	WO98010067
6,218,122	WO00055320	WO02059260	WO03094848	WO98045437
6,271,002	WO00055350	WO02059271	WO03101283	WO99026976

6440694	WO00056880	WO02059377	WO04003162	W990033980
200200131971	WO00058473	WO02060317	WO04018641	WO99033981
20020055474	WO00060078	WO02068652	WO04020497	WO99035158
20020064872	WO00061610	WO02070737	WO04020593	WO99035170
20030065157	WO00063438	WO02074237	WO04022059	WO99038972
20020072089	WO00078961	WO02076488	WO04023973	WO99047540
20020081659	WO01000828	WO02081498	WO04024097	WO99050416
20020110547	WO01025256	WO02081517	WO04024892	WO99054460
20020123619	WO01027158	WO02081731	WO04028479	WO99058642
20020150581	WO01040466	WO02083876		WO99064576
20020151681	WO01042451	WO02083898		
20020168637	WO01042467	WO02086443		
20020192678	WO01044448	WO02090526		
20030194734	WO01046697	WO02094629		
20030022239	WO01049716			

Ahr et al.(2002)"Identification of high risk breast-cancer patients by gene-expression profiling" Lancet 359:131-132

Chang et al. (2003) "Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer" Lancet 362:362-9

Early Breast Cancer Trialists' Collaborative Group (1995) "Effects of radiotherapy and surgery in early breast cancer. An overview of the randomized trials" N Engl J Med 333:1444-1455

Early Breast Cancer Trialists' Collaborative Group (1998a) "Polychemotherapy for early breast cancer: an overview of the randomized trials" Lancet 352:930-942

Early Breast Cancer Trialists' Collaborative Group (1998b) "Tamoxifen for early breast cancer: an overview of randomized trials" Lancet 351:1451-1467

Efron (1981) "Censored data and the bootstrap" J Am Stat Assoc 76:312-319

Eifel et al. (2001) "National Institutes of Health Consensus Development Conference Statement: adjuvant therapy for breast cancer, November 1–3, 2000" J Natl Cancer Inst 93:979–989

Foekens et al. (1989b) "Prognostic value of estrogen and progesterone receptors measured by enzyme immunoassays in human breast tumor cytosols" Cancer Res 49:5823–5828

Foekens et al. (1989a) "Prognostic value of receptors for insulin-like growth factor 1, somatostatin, and epidermal growth factor in human breast cancer" Cancer Res 49:7002–7009

Goldhirsch et al. (2003) "Meeting highlights: Updated International Expert Consensus on the Primary Therapy of Early Breast Cancer" J Clin Oncol 21:3357–3365

Golub et al. (1999) "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring" Science 286:531-537

Gruvberger et al. (2001) "Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns" Cancer Res 61:5979–5984

Hedenfalk et al. (2001) "Gene-expression profiles in hereditary breast cancer" N Engl J Med 344:539–548

Herrera-Gayol et al. (1999) "Adhesion proteins in the biology of breast cancer: contribution of CD44" Exp Mol Pathol 66:149–156

Huang et al. (2003) "Gene expression predictors of breast cancer outcomes" Lancet 361:1590–1596

Kaplan et al. (1958) "Non-parametric estimation of incomplete observations" J Am Stat Assoc 53:457–481

Keyomarsi et al. (2002) "Cyclin E and survival in patients with breast cancer" N Engl J Med 347:1566–1575

Lipshutz et al. (1999) "High density synthetic oligonucleotide arrays" Nat Genet 21:20-24

Ma et al. (2003) "Gene expression profiles of human breast cancer progression" Proc Natl Acad Sci USA 100:5974–5979

Ntzani et al. (2003) "Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment" Lancet 362:1439-1444

Perou et al. (2000) "Molecular portraits of human breast tumors" Nature 406:747-752

Ramaswamy et al. (2001) "Multiclass cancer diagnosis using tumor gene expression signatures" Proc Natl Acad Sci USA 98:15149-15154

Ramaswamy et al. (2003) "A molecular signature of metastasis in primary solid tumors" Nat Genet 33:1-6

Ransohoff(2004) "Rules of evidence for cancer molecular-marker discovery and validation" Nat Rev Cancer 4:309-314

Sørliet al.(2001) "Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications" Proc Natl Acad Sci USA 98:10869-10874

Sørliet al. (2003) "Repeated observation of breast tumor subtypes in independent gene expression data sets" Proc Natl Acad Sci USA 100:8418-8423

Sotiriou et al. (2003) "Gene expression profiles derived from fine needle aspiration correlate with response to systemic chemotherapy in breast cancer" Breast Cancer Res 4:R3

Sotiriou et al. (2003) "Breast cancer classification and prognosis based on gene expression profiles from a population-based study" Proc Natl Acad Sci USA 100:10393-10398

Su et al. (2001) "Molecular classification of human carcinomas by use of gene expression signatures" Cancer Res 61:7388-7393

van de Vijver et al. (2002) "A gene expression signature as a predictor of survival in breast cancer" N Engl J Med 347:1999-2009

van't Veer et al.(2002) "Gene expression profiling predicts clinical outcome of breast cancer" Nature 415:530-536

Wang et al. (2004) "Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer" J Clin

Oncol 22:1564-1571

Woelfle et al. (2003) "Molecular signature associated with bone marrow micrometastasis in human breast cancer" Cancer Res 63:5679-5684

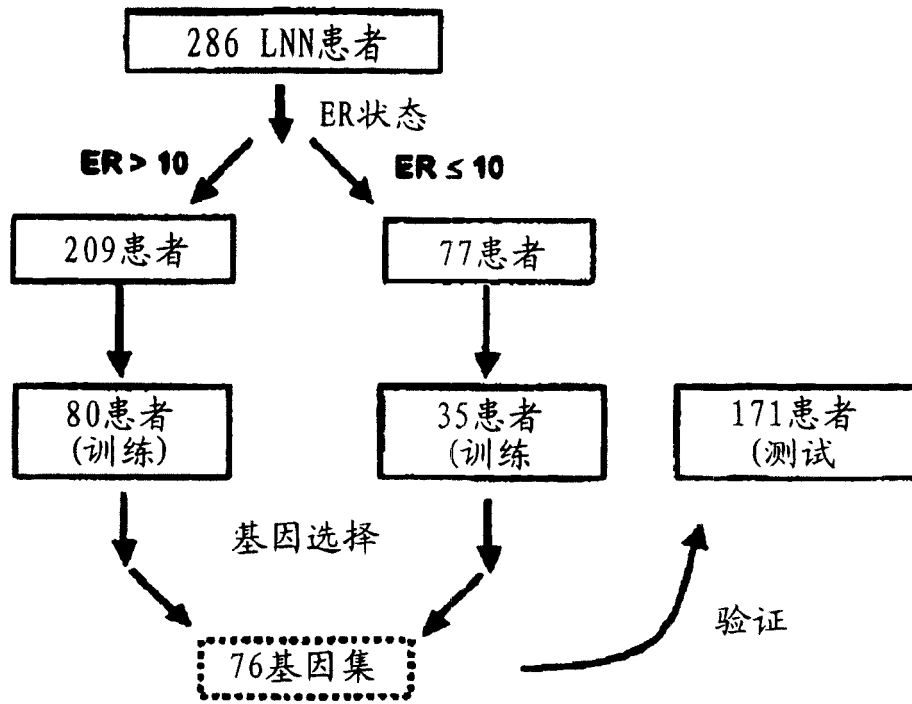


图 1A

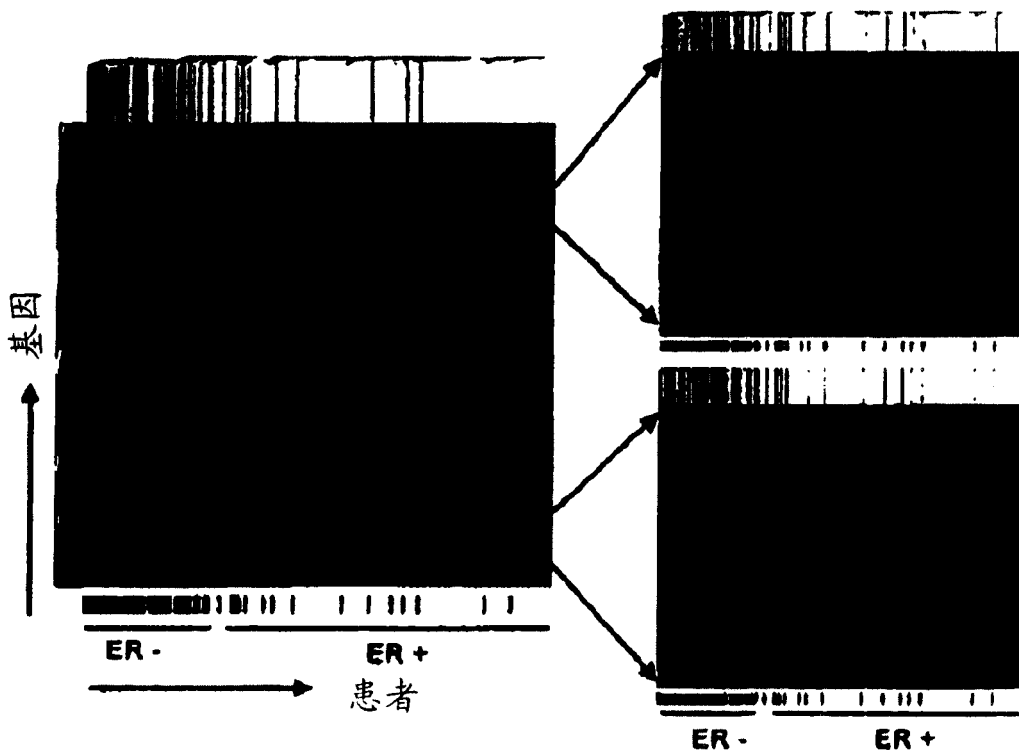
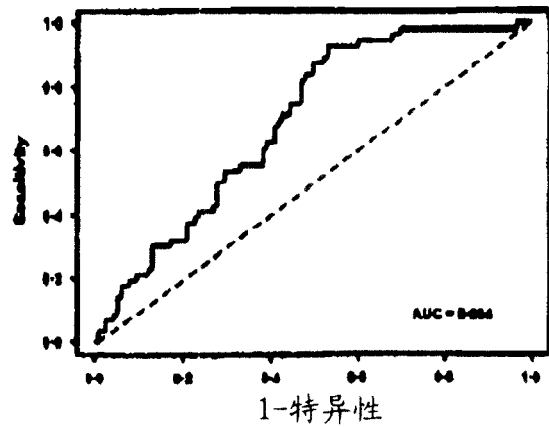
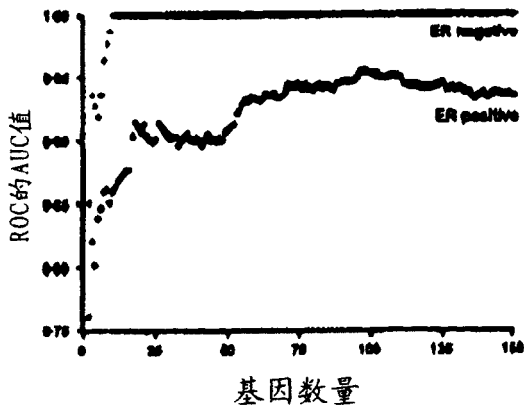


图 1B

A. 基因的选择



B. 验证集171 LNN患者

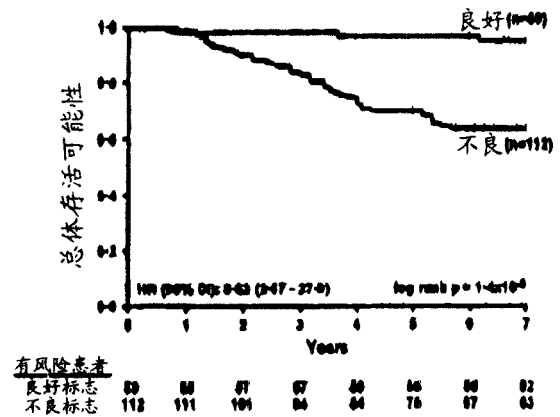
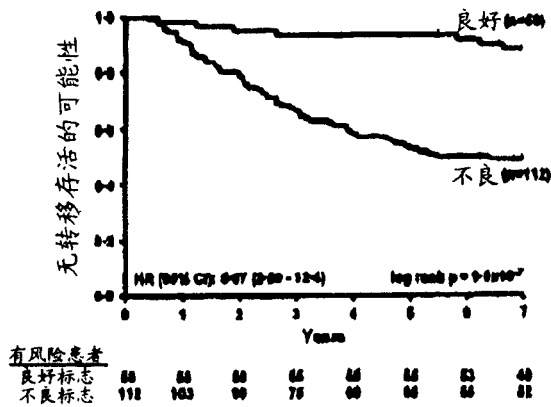
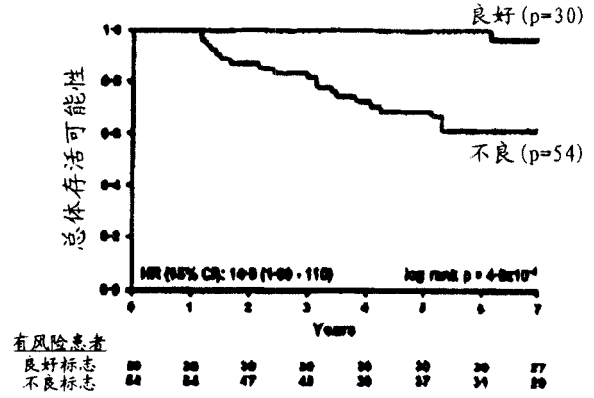
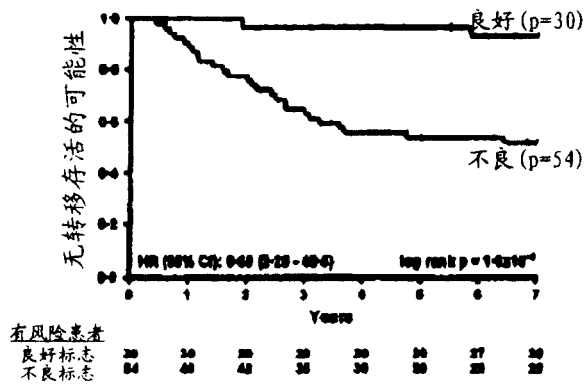
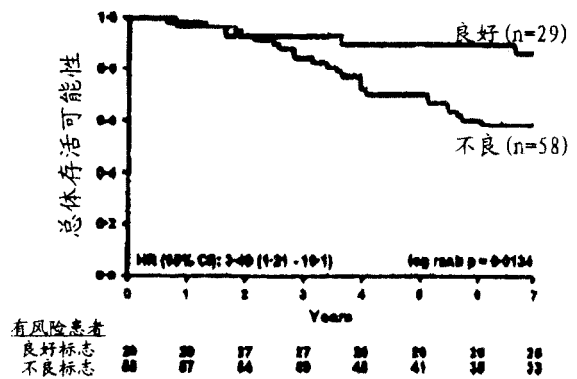
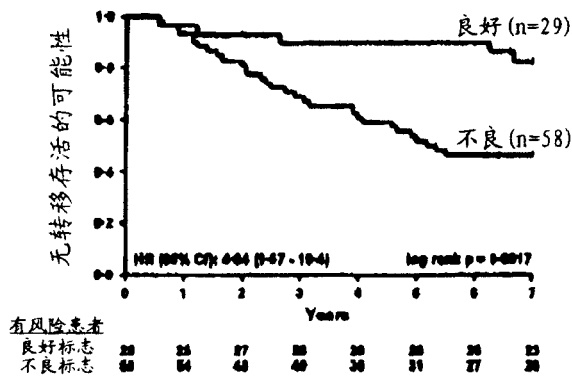


图 2

A. 绝经前



B. 绝经后



C. 肿瘤10-20mm

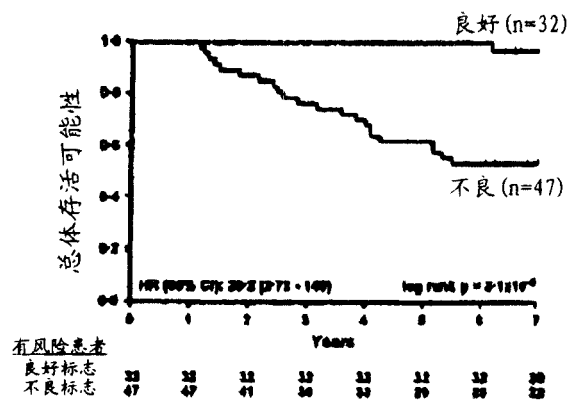
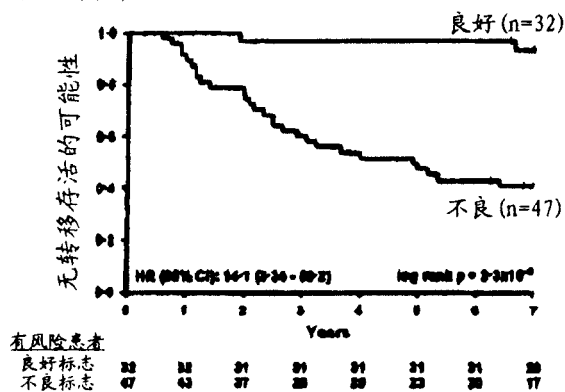


图 3

一轮和两轮RNA扩增之间对照基因表达强度的比较

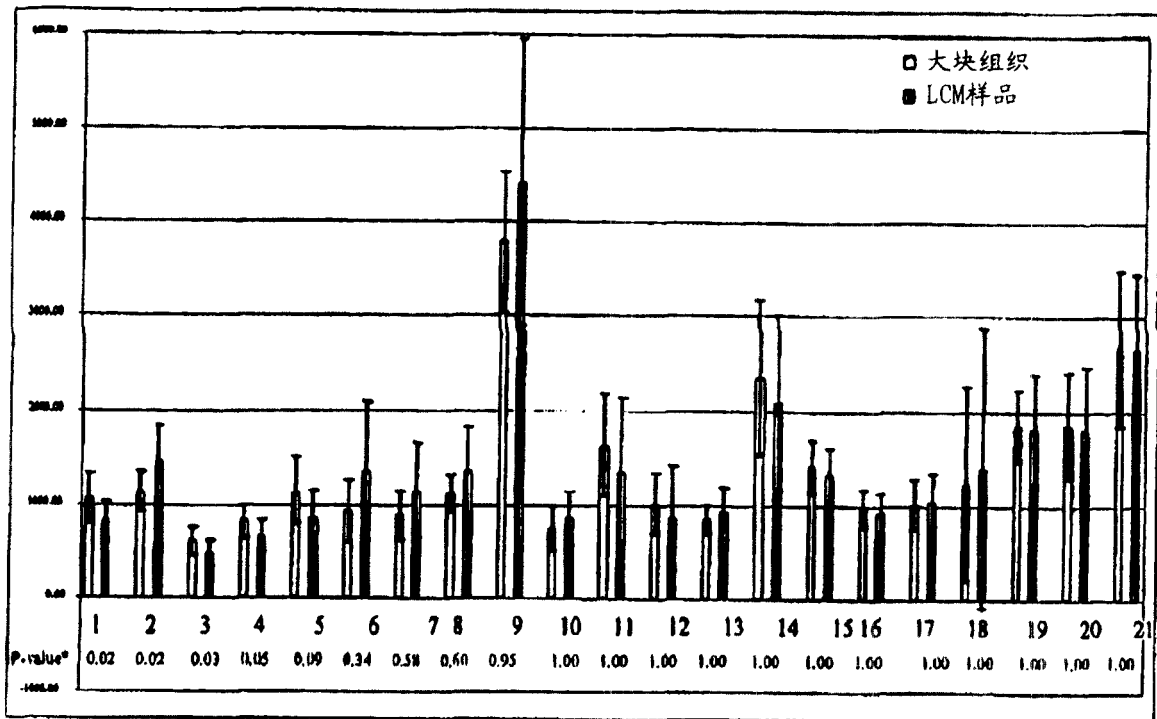
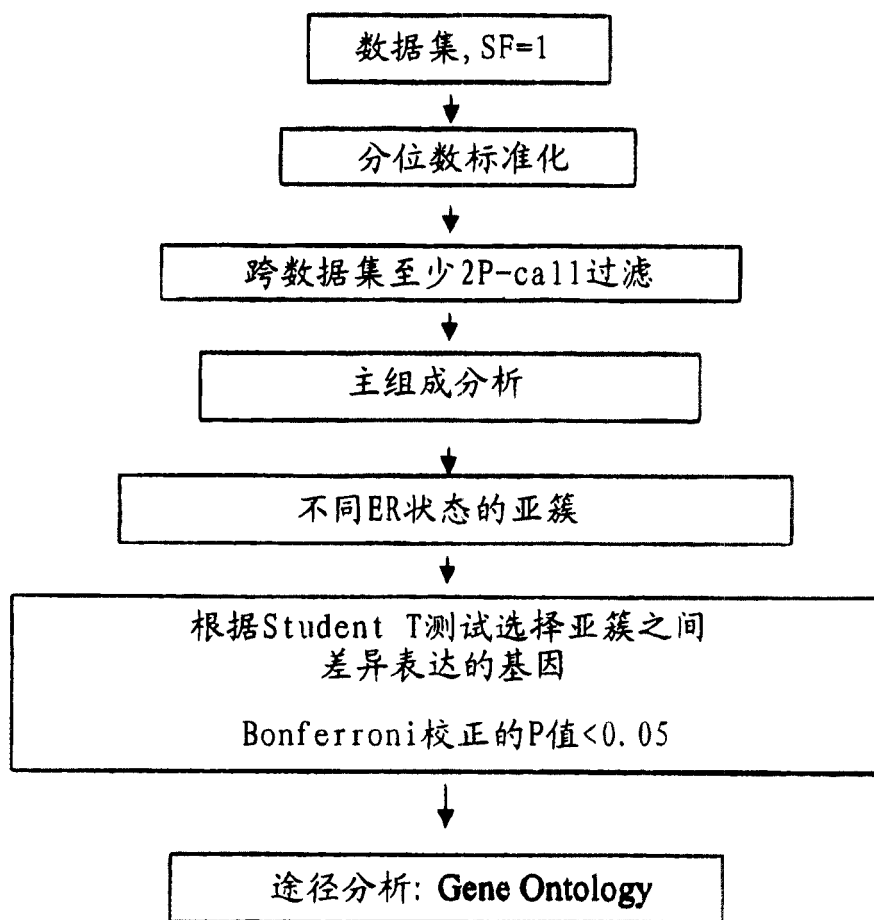


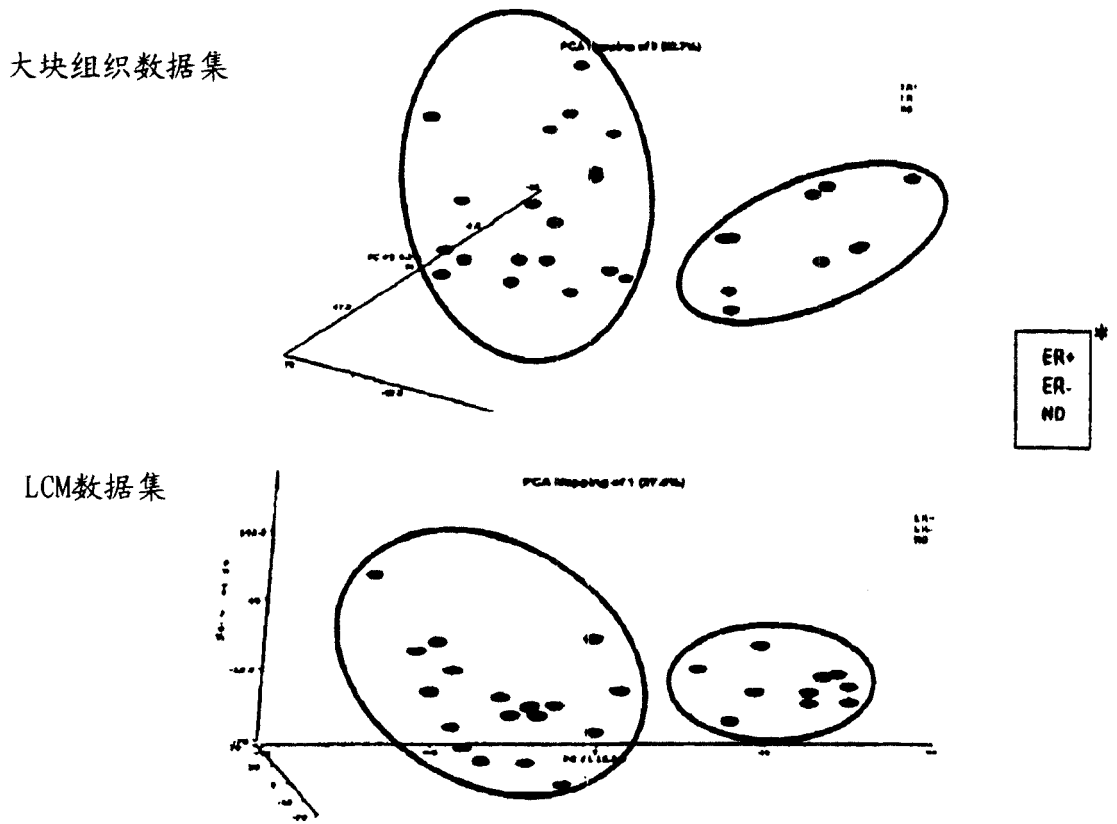
图 5



数据分析工作流程

图 6

过滤的基因集的主要成分分析



*ER分类基于酶免疫分析。使用10fmol/mg做为截断

图 7

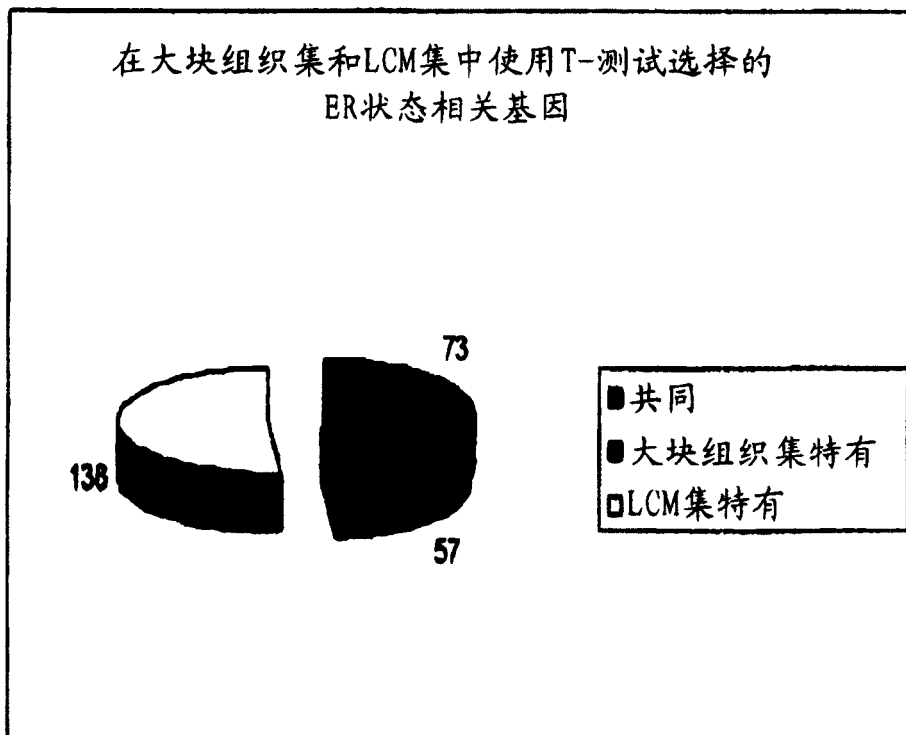


图 8

途径分析结果

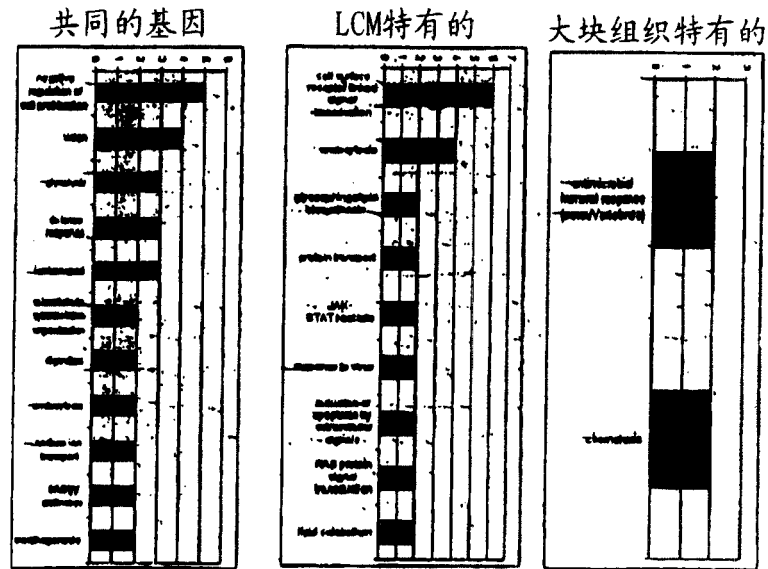


图 9

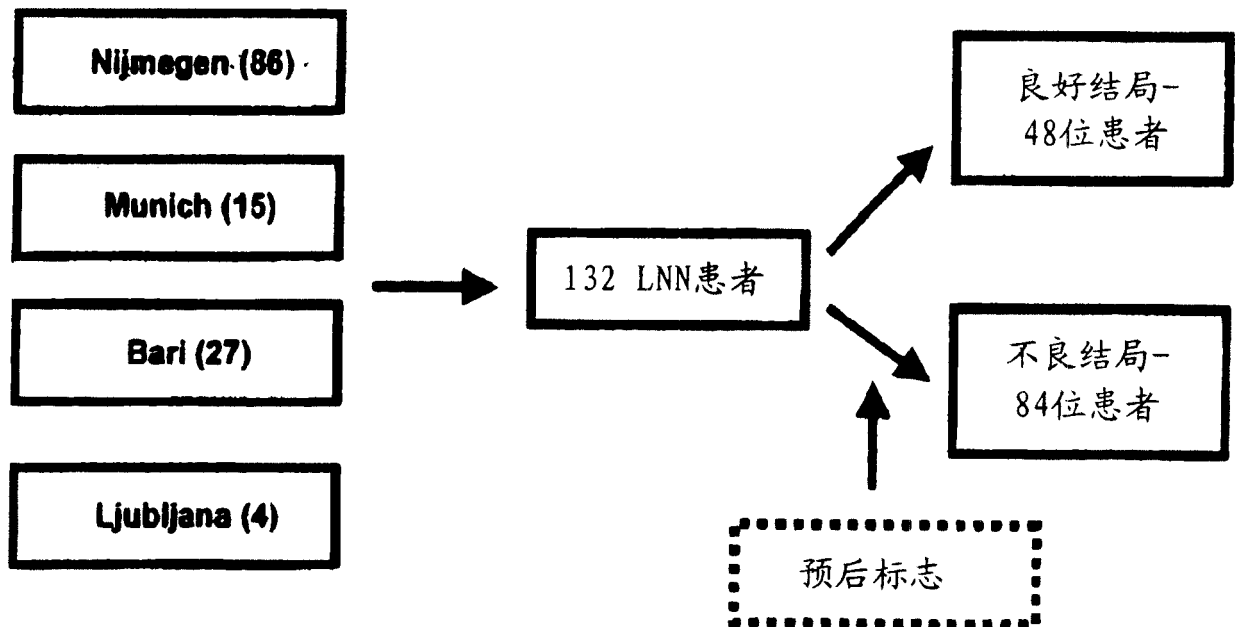


图 10

A.

Characteristic	Value (n, %)
Number	132
Age (mean±SD)	66±11
1-65 yr	9 (7%)
65-70 yr	41 (32%)
70-75 yr (mean 72.1)	62 (47%)
>75 yr	17 (13%)
Unknown	1
Menopausal status	
Premenopausal	31 (23%)
Postmenopausal	101 (77%)
T stage	
T1	77 (58%)
T2	54 (41%)
T3-4	1
Grade	
Poor	40 (30%)
Moderate	80 (61%)
Good	16 (12%)
Unknown	21 (16%)
Metastasis <5 years	
Yes	24 (18%)
No	108 (82%)

B.

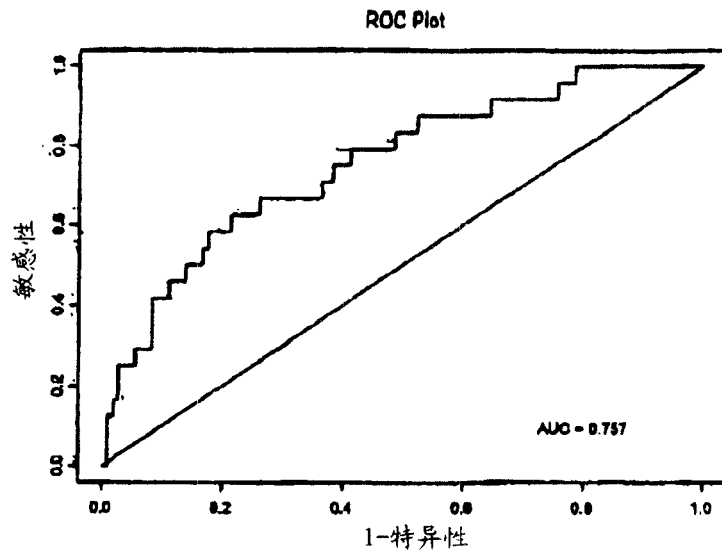


图 11

132位患者		
	复发	未复发
阳性:	21	63
阴性:	3	44
敏感性:	89% (0.69 - 0.96)	
特异性:	41% (0.32 - 0.51)	
复发频率:	18%	
PPV:	25%	
NPV:	84%	
几率比:	4.9 (1.4-17)	

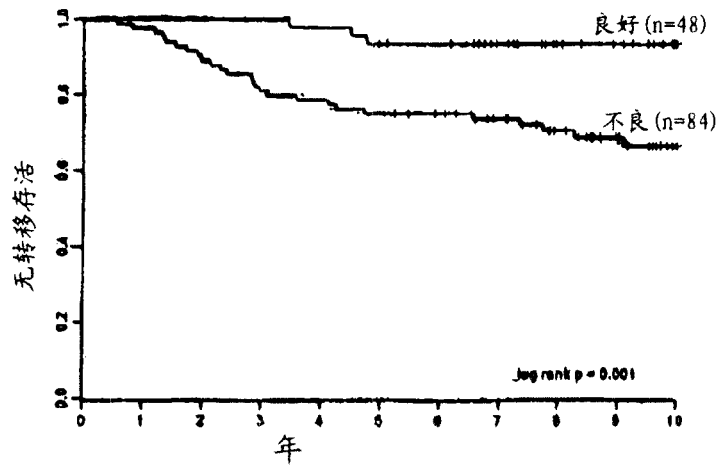


图 12

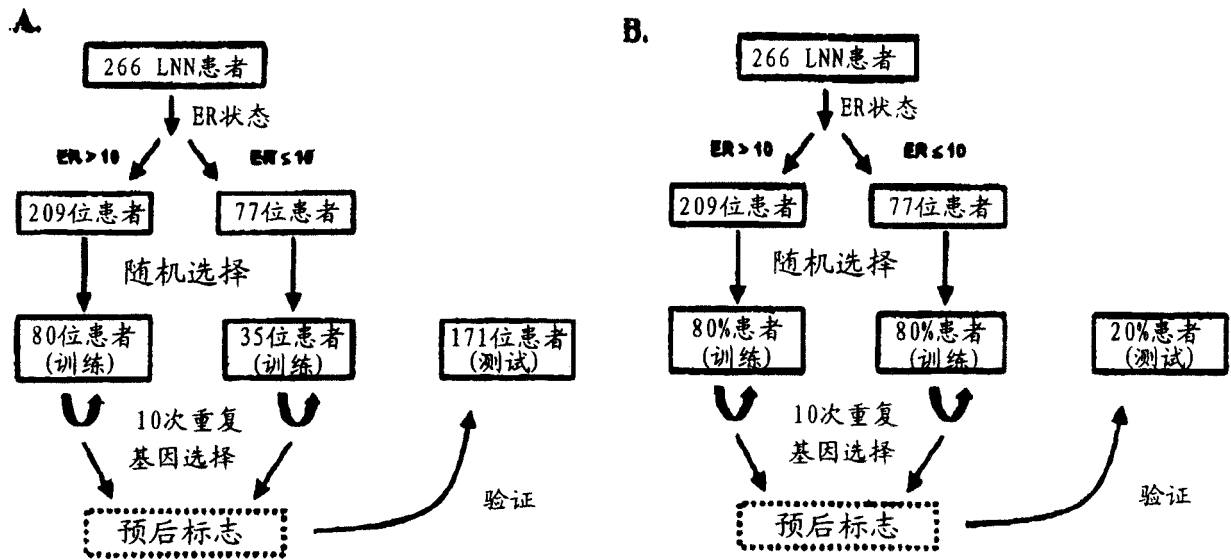


图 13

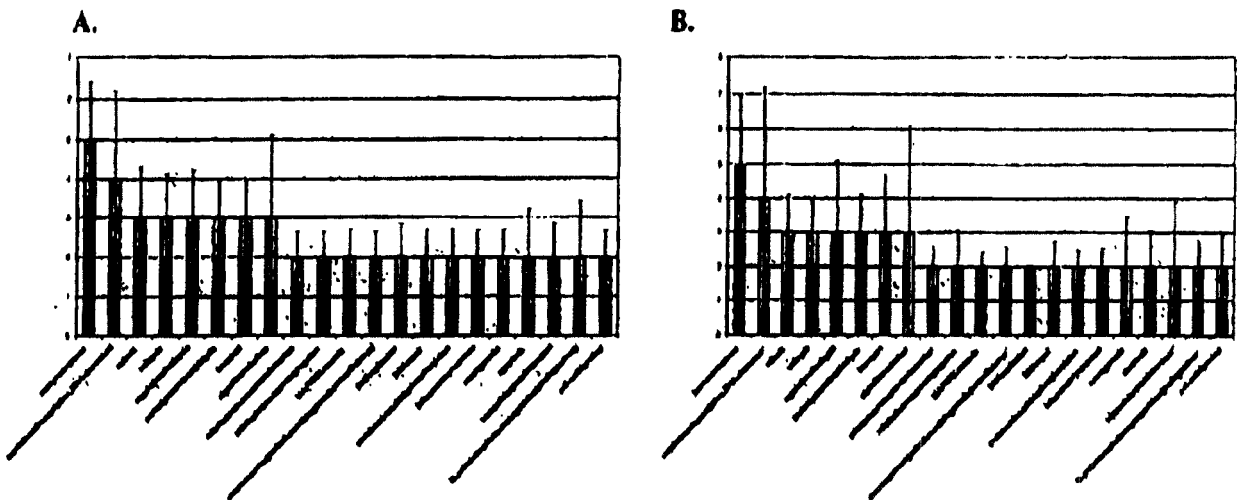


图 14

专利名称(译)	乳腺癌预后		
公开(公告)号	CN1950701A	公开(公告)日	2007-04-18
申请号	CN200580012425.3	申请日	2005-02-18
[标]申请(专利权)人(译)	维里德克斯有限责任公司		
申请(专利权)人(译)	维里德克斯有限责任公司		
当前申请(专利权)人(译)	维里德克斯有限责任公司		
[标]发明人	Y王		
发明人	Y·王		
IPC分类号	G01N33/53 C12Q1/68		
CPC分类号	C12Q2600/112 C12Q1/6886 C12Q2600/158 C12Q2600/106 C12Q2600/118		
代理人(译)	程淼		
优先权	10/783271 2004-02-20 US 60/634430 2004-12-08 US		
其他公开文献	CN1950701B		
外部链接	Espacenet SIPO		

摘要(译)

通过分析一组基因的表达进行了提供乳腺癌预后的方法。包括了在各种介质例如微阵列中的基因表达分布型，还包括了含有它们的试剂盒。

