

[19] 中华人民共和国国家知识产权局



[12] 发明专利申请公布说明书

[21] 申请号 200780029465.8

[51] Int. Cl.

C12N 15/11 (2006.01)

C12Q 1/00 (2006.01)

A61B 5/00 (2006.01)

[43] 公开日 2009 年 9 月 16 日

[11] 公开号 CN 101535480A

[22] 申请日 2007.6.7

[21] 申请号 200780029465.8

[30] 优先权

[32] 2006.6.7 [33] US [31] 60/811,407

[86] 国际申请 PCT/NZ2007/000142 2007.6.7

[87] 国际公布 WO2007/142540 英 2007.12.13

[85] 进入国家阶段日期 2009.2.9

[71] 申请人 奥塔戈创新有限公司

地址 新西兰达尼丁

[72] 发明人 斯蒂芬·J·阿辛德

乔-安·斯坦顿

[74] 专利代理机构 北京市柳沈律师事务所

代理人 封新琴

权利要求书 7 页 说明书 64 页 序列表 21 页
附图 6 页

[54] 发明名称

诊断方法和标记物

[57] 摘要

本发明涉及检测、监测和治疗前列腺癌 (PRC) 或前列腺上皮内瘤 (PIN) 或它们的倾向的方法。所述方法中所用的是新型癌标记物 PSPU43、以及生物分析法和试剂盒。

1. 分离的核酸分子，包含 SEQ ID NO:3 的序列或其功能等同片段或变体，或在严谨条件与 SEQ ID NO:3 或其片段或变体杂交的序列。

2. 权利要求 1 的分离的核酸分子，其与 SEQ ID NO:3 具有 70%，优选 75%，优选 80%，优选 90%，优选 95%，优选 99% 的序列同一性。

3. 分离的核酸分子，其包含权利要求 1 的核酸序列优选 SEQ ID NO:3 的至少 10 个核苷酸的片段，或其互补物，该片段或互补物在严谨条件与下述序列杂交：

- (a) 权利要求 1 的核酸序列，优选 SEQ ID NO: 3，或其互补物；
- (b) 与权利要求 1 的核酸序列或其互补物对应的 cDNA 的全长编码序列；
- (c) (a) 或 (b) 的反向互补物。

4. 权利要求 1-3 任一项所述的核酸分子，其长度为至少 20，至少 30，至少 40，至少 50 个核苷酸，至少 60 个核苷酸，至少 70 个核苷酸，至少 80 个核苷酸，至少 90 个核苷酸，或优选至少 100 个核苷酸。

5. 基因构建体，其包含权利要求 1-4 任一项所述核酸分子。

6. 权利要求 5 的基因构建体，其为表达构建体。

7. 载体，其包含权利要求 6 的基因构建体。

8. 宿主细胞，其包含根据权利要求 5-7 任一项所述基因构建体或载体。

9. 分离的多肽，其由权利要求 1-4 任一项所述核酸分子或其功能等同变体或片段编码。

10. 权利要求 9 的分离的多肽，其长度为至少 5 个氨基酸。

11. 分离的多肽，其包含 (a) SEQ ID NO:9, (b) SEQ ID:10, (c) SEQ ID NO:11, (d) SEQ ID NO:12, (e) SEQ ID NO:13, 或 (f) SEQ ID NO:14 的序列；或 (a), (b), (c), (d), (e) 或 (f) 的功能等同变体或片段，或在严谨条件与 (a), (b), (c), (d), (e) 或 (f) 任一项杂交的序列。

12. 权利要求 9-11 任一项所述分离的多肽，其中所述多肽与权利要求 9-11 任一项所述多肽具有至少：70%，优选 75%，优选 80%，优选 85%，优选 90%，优选 95%，优选 99% 的氨基酸同一性。

13. 抗体，其特异性结合权利要求 9-12 任一项所述多肽或所述多肽的功能等同变体或片段。

14. 根据权利要求 13 的抗体, 其为多克隆抗体、单克隆抗体、单链抗体或人源化抗体, 或其免疫活性片段。

15. 根据权利要求 14 的抗体, 其为单克隆抗体。

16. 权利要求 13-15 任一项所述抗体, 其标记了可检测标记物。

17. 重组制备根据权利要求 9-12 任一项所述多肽的方法, 所述方法包括如下步骤:

(a) 培养包含权利要求 5 或 6 的基因构建体的宿主细胞, 其能够表达权利要求 9-12 任一项所述多肽; 和

(b) 选择表达本发明多肽的细胞;

(c) 从所述细胞分离所表达的多肽; 并可选

(d) 纯化所表达的多肽。

18. 权利要求 17 的方法, 其中所述方法包括预先用所述构建体转染所述宿主细胞的步骤。

19. 阵列, 包含结合 PSPU43 (SEQ ID NO:3) 的一或多个核酸序列。

20. 阵列, 其包含一或多个权利要求 1-4 任一项所述核酸序列。

21. 权利要求 19 或 20 的阵列, 其还包含一或多个结合转凝蛋白 1 (SEQ ID NO:7) 的核酸序列。

22. 权利要求 19-21 任一项所述的阵列, 其还包含一或多个结合转凝蛋白 2 (SEQ ID NO:8) 的核酸序列。

23. 权利要求 19-22 任一项所述的阵列, 其还包含一或多个结合 PCA3 (SEQ ID NO:6) 的核酸序列。

24. 权利要求 19-23 任一项所述的阵列, 其还包含一或多个结合前列腺特异性抗原(PSA) (SEQ ID NO:5) 的核酸序列。

25. 权利要求 19-24 任一项所述的阵列, 其中所述核酸序列为 RNA。

26. 权利要求 19-25 任一项所述的阵列, 其中所述核酸序列为 DNA。

27. 筛选改变权利要求 1-4 任一项所述核酸分子的表达的化合物的方法, 所述方法包括如下步骤:

(a) 使表达所述核酸分子的细胞与被测化合物接触;

(b) 测定所述核酸分子的表达水平; 和

(c) 选出化合物, 与在所述被测化合物不存在时的水平相比, 其改变了表达水平。

28. 权利要求 27 的方法，其中所述核酸分子为 PSPU43 (SEQ ID NO:3)。

29. 筛选改变权利要求 1-4 任一项所述核酸分子的活性的化合物的方法，该方法包括：

- (a) 将被测化合物与由所述核酸分子编码的肽接触；
- (b) 检测所述肽的生物活性；和下列任一步骤：
- (c) 选出化合物，与在该化合物不存在时检测的生物活性相比，该化合物改变所述肽的生物活性；或
- (d) 选出结合所述肽的化合物。

30. 权利要求 29 的方法，其中所述核酸分子为 PSPU43 (SEQ ID NO:3)。

31. 改变权利要求 1-4 任一项所述核酸分子的表达或活性的化合物，所述核酸分子优选 PSPU43 (SEQ ID NO:3)，所述化合物由权利要求 27-30 任一项所述筛选方法选出。

32. 权利要求 31 的化合物在制备药物中的用途，所述药物用于治疗或预防前列腺上皮内瘤(PIN)或前列腺癌(PRC)。

33. PIN 或 PRC 表达图谱，其包含包括权利要求 1-4 任一项所述核酸分子的标记物的表达模式。

34. 根据权利要求 33 的图谱，其中所述核酸分子为 PSPU43 (SEQ ID NO:3)。

35. 根据权利要求 33 或 34 的图谱，其还包含选自 PCA3 (SEQ ID NO:6)、转凝蛋白 1 (SEQ ID NO:7)、转凝蛋白 2 (SEQ ID NO:8)和 PSA (SEQ ID NO:5)的一或多种标记物。

36. 根据权利要求 33-35 任一项所述图谱，其包含标记物 PSPU 43 (SEQ ID NO:3)、PCA3 (SEQ ID NO:6)、转凝蛋白 1 (SEQ ID NO:7)、转凝蛋白 2 (SEQ ID NO:8)和 PSA (SEQ ID NO:5)。

37. 治疗或预防患者的 PIN 或 PRC 的方法，所述方法包括改变患者中权利要求 1-4 任一项所述核酸分子的表达水平，或改变权利要求 9-12 任一项所述多肽的活性。

38. 权利要求 37 的方法，其中所述核酸分子为 PSPU43 (SEQ ID NO:3)。

39. 权利要求 37 或 38 的方法，其中通过给患者施用反义组合物、siRNA 组合物、或核酶组合物来抑制表达，所述组合物包含权利要求 1 所述核酸分子的一或多个互补核苷酸序列。

40. 权利要求 39 的方法，其中所述核酸分子为 PSPU43 (SEQ ID NO:3)。

41. 权利要求 39 或 40 的方法，其中所述组合物是疫苗。

42. 权利要求 37 的方法，其中通过施用特异性结合权利要求 9-12 任一项所述多肽的抗体来抑制表达。

43. 权利要求 42 的方法，其中所述多肽由 PSPU43(SEQ ID NO:3)编码。

44. 权利要求 42 或 43 的方法，其中所述抗体是单克隆抗体。

45. 治疗或预防患者的 PIN 或 PRC 的方法，所述方法包括给所述患者施用改变权利要求 9-12 任一项所述多肽的表达或活性的化合物。

46. 权利要求 45 的方法，其中所述多肽由 SEQ ID NO:3 编码。

47. 治疗或预防患者的 PIN 或 PRC 的方法，所述患者中权利要求 1 所述核酸分子是过表达的，所述方法包括给所述患者施用降低由所述核酸分子编码的多肽的表达或活性的化合物。

48. 权利要求 47 的方法，其中所述核酸分子为 PSPU43 (SEQ ID NO:3)。

49. 组合物，其包含药物有效量的，根据权利要求 1-4 任一项所述核酸分子，或根据权利要求 9-12 任一项所述多肽。

50. 组合物，其包含药物有效量的针对权利要求 1-4 任一项所述核酸分子的反义-寡核苷酸、核酶或 siRNA。

51. 权利要求 49 或 50 的组合物，其中所述核酸分子为 PSPU43 (SEQ ID NO:3)。

52. 组合物，其包含药物有效量的特异性结合权利要求 9-12 任一项所述多肽的抗体或其片段。

53. 权利要求 52 的组合物，其中所述多肽由 SEQ ID NO:3 编码。

54. 组合物，其包含药物有效量的由权利要求 27-30 任一项所述筛选方法选出的化合物。

55. 权利要求 49-54 任一项所述组合物，还包含可药用的载体、稀释

剂或赋形剂。

56. 治疗或预防患者的 PIN 或 PRC 的方法，所述方法包括给需要的患者施用有效量的权利要求 31 的化合物或 49-55 任一项所述组合物。

57. PSPU43 (SEQ ID NO:3)或其编码的多肽在制备药物中的用途，所述药物用于治疗或预防患者的 PIN 或 PRC。

58. 权利要求 1-4 任一项所述核酸分子或 9-12 任一项所述多肽在制备药物中的用途，所述药物用于治疗或预防患者的 PIN 或 PRC。

59. 针对权利要求 1 所述核酸分子的反义-寡核苷酸、siRNA 或核酶。

60. 针对 PSPU43 (SEQ ID NO:3)的反义-寡核苷酸、siRNA 或核酶。

61. 分析法，用于检测样本中权利要求 1-4 任一项所述核酸分子的存在，所述核酸分子优选 PSPU43 (SEQ ID NO:3)，该方法包括：

(a) 将所述样本与核苷酸探针接触，该探针与权利要求 1 所述核酸序列在严谨杂交条件杂交；和

(b) 检测样本中杂交复合物的存在。

62. 权利要求 61 的分析法，其中所述探针是标记的探针。

63. 权利要求 62 的分析法，其中所述探针是荧光标记的。

64. 权利要求 59-63 任一项所述分析法，其中所述探针是 SEQ ID NO:3 的互补物。

65. 测定患者样本中权利要求 1 所述核酸分子的表达水平的方法，所述核酸分子优选 PSPU43 (SEQ ID NO:3)，所述方法包括直接或间接测定所述核酸分子。

66. 权利要求 65 的方法，其中所述核酸分子被用于原位杂交或 RT-PCR 分析中。

67. 测定权利要求 1 所述核酸分子在患者样本中的表达水平的方法，所述方法包括：

(a) 扩增所述核酸分子或其互补物的 DNA 序列；或

(b) 扩增所述核酸分子或其互补物的 cDNA 序列；和

(c) 测定所述样本中 DNA、cDNA 或 RNA 之一或多个的水平。

68. 权利要求 67 的方法，其中所述核酸分子为 PSPU43 (SEQ ID NO:3)。

69. 权利要求 67 或 68 的方法，其中所述 DNA 或 cDNA 用 PCR 扩增。

70. 权利要求 65-69 任一项所述方法, 其中 DNA、cDNA 或 RNA 在样本中的水平用电泳测量。

71. 分析法, 其用于检测权利要求 9-12 任一项所述多肽在患者样本中的存在, 所述方法包括:

- (a) 使所述样本接触权利要求 13-16 任一项所述抗体; 和
- (b) 检测样本中结合多肽的存在。

72. 权利要求 71 的分析法, 其中所述抗体带有可检测的标记。

73. 诊断患者的前列腺上皮内瘤(PIN)、前列腺癌(PRC)或发生 PIN 或 PRC 的倾向的方法, 所述方法包括测定权利要求 1 所述核酸分子在患者样本中的表达水平, 其中与该核酸分子对照水平相比表达水平的改变表明该患者患 PIN、PRC、或有发生 PIN 或 PRC 的风险。

74. 权利要求 73 的方法, 其中所述改变是表达水平升高。

75. 根据权利要求 74 的方法, 其中所述表达水平的改变是比正常对照水平高至少 10%。

76. 权利要求 73-75 任一项所述方法, 其中所述对照水平是在正常前列腺样本中测定的。

77. 测试患者的前列腺上皮内瘤(PIN)、前列腺癌(PRC)状态的方法, 所述方法包括测定权利要求 1 所述核酸分子在患者样本中的表达水平, 其中与所述分子的对照水平相比表达水平高, 表明该患者有 PIN 或 PRC 状态、或有发生 PIN 或 PRC 的风险。

78. 监测受试者对 PIN 或 PRC 治疗的反应的方法, 所述方法包括测定权利要求 1 所述核酸分子在患者样本中的表达水平, 并将所述核酸分子的水平与对照水平相比, 其中测定的水平与对照水平相比统计学显著性改变表明对该治疗有反应。

79. 权利要求 73-78 任一项所述方法, 其中所述核酸分子为 PSPU43 (SEQ ID NO:3)。

80. 权利要求 73-79 任一项所述的方法, 其还包括测定 PIN 或 PRC 的一或多种其它标记物的水平, 并将所述水平与对照中的标记物水平比较, 其中所述水平与对照水平相比显著性偏差, 以及权利要求 1-4 任一项所述核酸分子优选 PSPU43 (SEQ ID NO:3)的水平的统计学显著性升高, 两者一起, 表明 PRC 或 PIN, 或可用于监测 PIN 或 PRC 状态。

81. 权利要求 80 的方法,其中所述其它标记物选自由转凝蛋白 1 (SEQ ID NO:7)、前列腺特异性抗原(SEQ ID NO:5)、和 PCA3 (SEQ ID NO:6)组成的组。

82. 权利要求 73-81 任一项所述方法,其中所述样本是尿液、淋巴液、血液、血浆、精液、前列腺按摩液、或前列腺组织样本。

83. 权利要求 73-82 任一项所述方法,其中转凝蛋白 2 (SEQ ID NO:8)被用作参比标记物。

84. 转凝蛋白 2 (SEQ ID NO:8)在权利要求 73-82 任一项所述方法中作为参比标记物的用途。

85. 检测样本中权利要求 1-4 任一项所述核酸分子的存在的试剂盒,所述核酸分子优选 PSPU43 (SEQ ID NO:3),所述试剂盒包含至少一个容器,所述容器中容纳权利要求 1-4 任一项所述核酸分子。

86. 权利要求 85 的试剂盒,还包含一或多种用于检测所述核酸分子的试剂。

87. 试剂盒,其包含一或多种与权利要求 1 所述核酸分子或由所述核酸分子编码的多肽结合的检测试剂,所述核酸分子优选 PSPU43 (SEQ ID NO:3)。

88. 权利要求 85-87 任一项的试剂盒,其还包含下列一或多项:

- (a) 编码转凝蛋白 1 (SEQ ID NO:7)的核酸分子或其互补物;
- (b) 编码转凝蛋白 2 (SEQ ID NO:8)的核酸分子或其互补物;
- (c) 编码 PCA3 (SEQ ID NO:6)的核酸分子或其互补物;和
- (d) 编码 PSA (SEQ ID NO:5)的核酸分子或其互补物。

89. 权利要求 88 的试剂盒,其包含(a), (b), (c)和(d)。

90. 非人动物,其具有的基因组中权利要求 1 的核酸分子被改变、破坏、消除或添加。

91. 权利要求 90 的动物,其为小鼠。

92. 权利要求 90 或 91 的动物,其中所述核酸分子为 PSPU 43 (SEQ ID NO:3)。

诊断方法和标记物

技术领域

本发明涉及检测、诊断、监测和治疗前列腺癌(PRC)、前列腺上皮内瘤(prostatic intraepithelial neoplasia, PIN)或其发病倾向(predisposition to same)的方法;还涉及用于所述方法的标记物。

发明背景

前列腺癌是在欧洲和北美洲男性最常诊断出的癌症。在那些地区,前列腺癌为男性仅次于肺癌的第二大死因(Frankel 等, 2003)。该病在世界其它部分,如日本也逐渐增多,并可表明在东方国家接纳了西方饮食。

前列腺癌是上年纪的男性的疾病。40%的 60 岁男性有局限性的(localised)前列腺肿瘤,且超过 75%的 85 岁以上男性患有前列腺癌。此种癌症是通常作为潜伏性疾病存在而无其它疾病症状,并从诊断到死亡最长可达 10 年。该病通常的进展为,从前列腺内清晰可辨的团块(mass)到破坏和侵入前列腺的侧缘,之后转移至局部淋巴结,和/或转移至骨髓。

前列腺上皮内瘤(PIN)是特定类型的损害,据认为它是前列腺癌的前兆(precursor)(McNeal and Bostwick, 1986)。如果诊断得早,现在用雄激素去势疗法(androgen ablation therapy)治疗患者。去势疗法有不良副作用如性欲和能力丧失。随着该疾病的发展,它变为雄激素非依赖性。在此阶段,手术(根治性前列腺切除术(radical prostatectomy))是主要选择。这可挽救所述患者的性命,但手术的常见后果是失禁(incontinence)、勃起障碍(erection dysfunction)和尿渗漏(urinary leakage)。

仍不清楚前列腺癌为什么发病以及什么决定其进展。另外,对前列腺癌的检测主要限于体格检查、针吸活检(needle biopsy)和骨扫描。现在,循环的前列腺特异性抗原(PSA)的水平升高最常用于预测前列腺癌的存在。这是唯一常见的前列腺癌非侵入性筛选法。

然而,PSA 测试不具有诊断意义。PSA 的水平升高可由其它不相关因子如良性前列腺增生(BPH)和前列腺炎导致。它对疾病状态并不特异并且不

能指出死亡风险(Frankel 等, 2003)。不能仅凭 PSA 筛选提供的信息做出临床决策。PSA 试验不能区分恶性和非恶性形式, 或预测损伤如何发展。并且, 不是所有前列腺癌都导致血清 PSA 浓度上升。事实上, PSA 水平上升且进行根治治疗(前列腺切除术)的男性有 85%如此而无预期益处(Frankel 等, 2003)。

由此, 需要更可靠的方法诊断前列腺癌或其前兆, 并随时监测疾病(动态(active)监测), 特别在早期阶段进行监测, 以便仍有机会选择治疗方案。也需要标记物用于测定患者的状态。

因此, 本发明的一个目的是提供用于测定患者的前列腺癌状态的标记物, 或至少给公众提供了可用的选择。

本发明的另一个目的是提供诊断和/或预测前列腺癌或前列腺上皮内瘤后果或其倾向的方法。

发明概述

第一方面, 本发明提供了分离的核酸分子或其变体或片段, 其包含 SEQ ID NO:3 的序列或其功能等同变体或片段, 或在严谨条件与 SEQ ID NO:3 杂交的序列,。

优选杂交是在严谨条件中进行。

另一方面, 本发明提供了分离的核酸分子, 其包含上述核酸序列优选 SEQ ID NO:3 的至少 10 个核苷酸的片段, 或其互补物, 所述片段在严谨条件与下述杂交:

- (a) 上述核酸序列, 优选 SEQ ID NO: 3 或其互补物;
- (b) 对应于上述核酸序列或其互补物的核酸序列的 cDNA 的全长编码序列;
- (c) 上述(a)或(b)的反向互补物(reverse complement)。

所述核酸分子可以是至少 20, 至少 30, 至少 40, 至少 50 个核苷酸, 至少 60 个核苷酸, 至少 70 个核苷酸, 至少 80 个核苷酸, 至少 90 个核苷酸, 或优选至少 100 个核苷酸。

在另一方面, 本发明提供了包含本发明核酸分子的基因构建体。

优选, 所述构建体是本文所限定的表达构建体。

本发明还提供了包含本发明基因构建体的载体。

本发明也提供了包含本发明基因构建体或载体的宿主细胞。

本发明也提供了由本发明核酸分子编码的分离的多肽。优选，所述多肽长度为至少 5 个氨基酸。

本发明也提供了分离的多肽，其包含(a) SEQ ID NO:9, (b) SEQ ID:10, (c) SEQ ID NO:11, (d) SEQ ID NO:12, (e) SEQ ID NO:13, 或(f) SEQ ID NO:14 的序列；或(a), (b), (c), (d), (e)或(f)的功能等同变体或片段，或在严谨条件与(a), (b), (c), (d), (e)或(f)任一杂交的序列。

在另一方面，本发明提供了重组制备本发明多肽的方法，所述方法包括如下步骤：

(a) 培养包含本发明基因构建体的宿主细胞，所述构建体如本文限定的能够表达本发明多肽的表达构建体；

(b) 选择表达本发明多肽的细胞；

(c) 从所述细胞分离所表达的多肽；并可选(optionally)

(d) 纯化所表达的多肽。

作为预先步骤(pre-step)，所述方法可包括用所述构建体转染所述宿主细胞。

本发明也提供了抗体，其特异性结合本发明核酸分子所编码的多肽或所述多肽的功能等同变体或片段。

优选，所述抗体为多克隆的、单克隆的、单链的抗体或人源化抗体，或其免疫活性片段。

在另一方面，本发明提供了阵列，其包含一或多个结合 PSPU 43 (SEQ ID NO:3)的核酸序列。

本发明也提供了阵列，其包含一或多个本发明的核酸序列。

优选，所述阵列还包含一或多个结合转凝蛋白(transgelin) 1 (SEQ ID NO:7)、转凝蛋白 2 (SEQ ID NO:8)、PCA3 (SEQ ID NO:6)、和 PSA (SEQ ID NO:5)的核酸序列。

本发明也提供了筛选改变本发明核酸分子的表达的化合物的方法，所述核酸分子优选 PSPU 43 (SEQ ID NO:3)，所述方法包括如下步骤：

(a) 使表达所述核酸的被测细胞与被测化合物接触；

(b) 测定所述核酸的表达水平；和

(c) 选出化合物，与在所述被测化合物不存在时相比，其改变了表达

水平。

还提供了筛选改变本发明任一种核酸的活性的化合物的方法，所述核酸优选 PSPU 43 (SEQ ID NO:3)标记物，该方法包括：

- (a) 将被测化合物与由所述核酸分子编码的肽接触；
- (b) 检测所述肽的生物活性；和下列任一步骤：
- (c) 选出化合物，与在该化合物不存在时检测的生物活性相比，该化合物改变所述肽的生物活性；或
- (d) 选出结合所述肽的化合物。

本发明也提供了用本发明的筛选方法选出的改变本发明核酸分子的表达或活性的化合物，所述核酸分子优选 PSPU43 (SEQ ID NO:3)。

另一方面，本发明提供了包含药物有效量的用本发明的筛选方法选出的化合物的组合物。

本发明也涉及本发明化合物在制备用于治疗 PIN 或 PRC 的药物中的用途。

本发明也提供了 PIN 或 PRC 表达图谱(profile)，其包含包括本发明核酸分子的标记物表达的模式，所述核酸分子优选 PSPU 43 (SEQ ID NO:3)。优选，所述图谱还包含选自由转凝蛋白 1 (SEQ ID NO:7)、转凝蛋白 2 (SEQ ID NO:8)、PCA 3 (SEQ ID NO:6)和 PSA (SEQ ID NO:5)组成的组的一或多种标记物。

另一方面，本发明提供了治疗或预防患者的 PIN 或 PRC 的方法，所述方法包括改变患者中本发明核酸分子的表达水平或所述标记物编码的肽的活性，所述核酸分子优选 PSPU 43 (SEQ ID NO:3)。这可以通过促进表达或施用组合物来实现，所述组合物含有所述核酸分子如 PSPU 43 编码的多肽。或者这可以通过抑制表达来实现。促进或抑制表达水平哪个合适，取决于所述核酸分子和 PSPU 43 编码的多肽是过表达还是表达不足 (under-expressed)。不希望受理论所限，此时相信过表达和表达不足都是可能的。

优选，本发明核酸分子所编码的多肽是过表达的，通过给患者施用反义组合物来抑制表达，所述组合物包含本发明核酸分子的一或多个反义核苷酸序列，优选 PSPU 43 (SEQ ID NO:3)的反义序列。

在另一个实施方案中，通过给患者施用 siRNA 组合物来抑制表达。所

述组合物降低本发明核酸分子，优选 PSPU 43 (SEQ ID NO:3)的表达。

在另一个实施方案中，通过给患者施用核酶组合物来抑制表达。

也可以通过施用抗体或活性抗体片段来抑制表达，所述抗体或片段特异性结合本发明核酸分子，优选 PSPU 43 (SEQ ID NO:3)。此活性片段优选免疫活性片段。

在一个实施方案中，所施用的组合物是疫苗。

本发明也涉及反义-寡核苷酸、核酶或 siRNA，其针对本发明核酸分子，优选 PSPU 43 (SEQ ID NO:3)。所述序列可用于上述方法。

本发明也提供了治疗或预防患者的 PIN 或 PRC 的方法，在所述患者中本发明多肽的表达不足，所述多肽如 PSPU 43 编码的多肽，所述方法包括给所述患者施用组合物，该组合物包含本发明核酸分子所编码的表达不足的多肽或该多肽的活性变体或片段，所述核酸分子如 PSPU 43 (SEQ ID NO:3)。

另一方面，本发明提供了治疗或预防患者的 PIN 或 PRC 的方法，所述方法包括给所述患者施用改变本发明多肽的表达或活性的化合物，所述多肽优选 PSPU 43 (SEQ ID NO:3)所编码的多肽。

另一方面，本发明提供了治疗或预防患者的 PIN 或 PRC 的方法，在所述患者中本发明核酸分子是过表达的，所述核酸分子优选 PSPU 43 (SEQ ID NO:3)，所述方法包括给所述患者施用降低本发明多肽的表达或活性的化合物，所述多肽优选由 PSPU 43 编码的多肽。

本发明也提供了组合物，其包含药物有效量的本发明核酸分子，优选 PSPU 43 (SEQ ID NO:3)，或其编码的多肽。

本发明还提供了组合物，其包含药物有效量的针对本发明核酸分子的反义-寡核苷酸、核酶或小干扰(small interfering)RNA，所述核酸分子优选 PSPU 43 (SEQ ID NO:3)。

所述组合物可包含两或多种针对所述核酸分子的反义-寡核苷酸、核酶或 siRNA。

在另一方面，本发明提供了组合物，其包含药物有效量的特异性结合本发明多肽的抗体或其片段，所述多肽优选 PSPU 43 (SEQ ID NO:3)标记物所编码的多肽。

本发明也提供了治疗患者的 PIN 或 PRC 的方法，所述方法包括给需要

的患者施用有效量的本发明化合物或本发明组合物。

本发明也提供了本发明核酸分子，优选 PSPU 43 (SEQ ID NO:3)或其编码的多肽在制备用于治疗或预防患者的 PIN 或 PRC 的药物中的用途。

本发明也提供了分析法(assay)，用于检测样本中本发明核酸分子，优选 PSPU43 (SEQ ID NO:3)的存在，所述方法包括：

(a) 将所述样本与核苷酸探针接触，该探针与本发明核酸序列，优选 PSPU43 (SEQ ID NO:3)在严谨杂交条件杂交；和

(b) 检测样本中杂交复合物的存在。

优选，所述探针是标记的探针，常为荧光标记的探针。在一个实施方案中，所述探针是 SEQ ID NO:3 的互补物。

本发明也提供了测定患者样本中本发明核酸分子的表达水平的方法，所述核酸分子优选 PSPU 43 (SEQ ID NO:3)，所述方法包括直接或间接测定所述核酸分子。

方便地，所述核酸分子被用于原位杂交或 RT-PCR 分析中而被测定。

本发明也涉及测定样本中本发明核酸分子的表达水平的方法，所述核酸分子优选 PSPU 43 (SEQ ID NO:3)，所述方法包括：

(a) 扩增所述核酸分子或其互补物的 DNA 序列；或

(b) 扩增所述核酸分子或其互补物的 cDNA 序列；和

(c) 测定所述样本中 DNA、cDNA 或 RNA 之一或多种的水平

本发明也提供了检测本发明多肽在患者样本中的存在的分析法，所述方法包括：

(a) 使所述样本接触本发明的抗体；和

(b) 检测样本中结合多肽的存在。

优选，所述抗体带有可检测的标记。

在另一方面，本发明涉及诊断患者的前列腺上皮内瘤(PIN)、前列腺癌(PRC)或发生 PIN 或 PRC 的倾向的方法，所述方法包括测定本发明核酸分子，优选 PSPU 43 (SEQ ID NO:3)在患者样本中的表达水平，其中该核酸分子的表达水平相比对照水平的改变表明该患者患 PIN、PRC、或有发生 PIN 或 PRC 的风险。

通常，所述表达水平的改变是比正常对照水平高至少 10%。方便地，所述对照水平是在正常前列腺中测定的 PSPU 43 的表达水平。

所述样本可包括正常前列腺细胞、PIN 或 PRC 细胞，并优选来自正常前列腺、PIN 或前列腺癌肿瘤的上皮细胞。

本发明也提供了测试患者的前列腺上皮内瘤(PIN)和前列腺癌(PRC)状态的方法，所述方法包括测定本发明核酸分子，优选 PSPU 43 (SEQ ID NO:3) 在患者样本中的表达水平，其中所述分子的表达水平比对照水平高表明该患者患有 PIN 或 PRC 状态、或有发生 PIN 或 PRC 的风险。

在另一方面，本发明涉及监测受试者对 PIN 或 PRC 治疗的反应的方法，所述方法包括测定本发明核酸分子，优选 PSPU 43 (SEQ ID NO:3) 在患者样本中的表达水平，并将所述 PSPU 43 (SEQ ID NO:3) 的水平与对照水平相比，其中测定的水平与对照水平相比有统计学显著性改变表明对该治疗有反应。

优选，这些测定、测试和监测方法还包括测定 PIN 或 PRC 的一或多种其它标记物的水平，并将所述水平与对照中的标记物水平比较，其中所述水平与对照水平相比有显著性偏差，以及本发明核酸分子水平有统计学显著性升高，一起表明 PRC 或 PIN，或可用于监测 PIN 或 PRC 状态，所述核酸分子优选 PSPU 43 (SEQ ID NO:3)。

所述其它标记物可以是选自由转凝蛋白 1 (SEQ ID NO:7)、转凝蛋白 2 (SEQ ID NO:8)、PCA 3 (SEQ ID NO:6) 和 PSA (SEQ ID NO:5) 组成的组的一或多种标记物。

本发明也提供了检测样本中本发明核酸分子的存在的试剂盒，所述核酸分子优选 PSPU 43 (SEQ ID NO:3)，所述试剂盒包含至少一个盛有本发明核酸的容器。

本发明也提供了包含结合本发明核酸分子或本发明多肽的一或多种检测试剂的试剂盒。

在一个实施方案中，所述试剂盒还包含下列之一或多种：

- (a) 编码转凝蛋白 1 (SEQ ID NO:7) 的核酸或其互补物；
- (b) 编码转凝蛋白 2 (SEQ ID NO:8) 的核酸或其互补物；
- (c) 编码 PCA3 (SEQ ID NO:6) 的核酸或其互补物；和
- (d) 编码 PSA (SEQ ID NO:5) 的核酸或其互补物。

在另一方面，本发明涉及本发明的诊断、测试或监测方法，其中转凝蛋白 2 (SEQ ID NO:8) 用作参比标记物。

本发明也涉及转凝蛋白 2 (SEQ ID NO:8)在本发明的诊断、测试和监测方法中作为参比标记物的用途。

也提供了非人动物，其基因组中 PSpU 43 (SEQ ID NO:3)的核酸序列被改变、破坏(disrupt)、消除或添加。

优选所述动物为小鼠。

附图简述

本发明现在参照下面的附图来进行描述：

图 1: 显示 *Pspu43* 的共有(consensus)序列。该共有序列从包含 UniGene 簇 Hs.161160 的 EST 的毗连序列群(contig)产生；

图 2: 显示 SYBR Green qPCR 分析法的解离曲线，该分析法使用针对 *Pspu1*, *Pspu2*, *Pspu8*, *Pspu43*, T1 和 T2 的引物组。所用的 cDNA 模板从 PC3 细胞系产生；

图 3: 显示用于测试前列腺组织中的标记物表达的每种引物和探针/引物组合的 qPCR 效果。标准曲线使用通用的参比 cDNA 作为模板；

图 4: 显示从匹配的组织对产生的 cDNA 模板得到的原始(raw)平均 CT 值。每个误差棒(bar)表示从一式两份的 qPCR 反应计算出的 ± 1 标准偏差；

图 5: 显示校正了基因组 DNA 污染后每份样本中每种标记物的 cDNA 相对量。从平均 CT 和从为每种标记物作的标准曲线计算出的最优拟合线(best fit line)测定相对量。在仅有 RNA 的模板(无 RT 反应)上对 qPCR 进行类似计算，从对 cDNA 模板测定的值中减去此反应的相对量；和

图 6: 显示核苷酸序列 *Pspu43* 的六个读码框的翻译。使用单字母的氨基酸编码。-表示终止密码子。

定义

此说明书及权利要求中所用的术语“包含”指“至少部分由.....组成”，也就是说在解释包含此术语的此说明书及权利要求时，在每次叙述中由该术语打头的特征都需要存在，但其他特征也可存在。相关术语如“含有”和“包括”也以类似方式来解释。

提及本文公开的数值范围(例如 1 到 10)，它也应包括在那个范围内提及的所有相关数值(例如，1, 1.1, 2, 3, 3.9, 4, 5, 6, 6.5, 7, 8, 9 和 10)，并包括在

那个范围内的任何有理数范围(例如 2 到 8, 1.5 到 5.5, 和 3.1 到 4.7), 因此明确公开了所有本文明确公开范围的所有亚范围(sub-range)。这些仅仅是应包括的例子, 并且在列举的最低值和最高值之间的数值的所有可能组合应在此申请中以类似方式明确叙述。

本文所用的术语“标记物”指在染色体上有可鉴定物理位置的 DNA 节段。标记物可以是基因或其它可鉴定的核酸序列, 如开放读码框, 内含子或基因间的基因组 DNA 节段的部分。

本文所用标记物的“对照水平”指在来自正常健康个体的样本中检测的表达水平, 或基于已知没有患 PRC 或 PIN 的个体群测定的水平。所述对照水平可以是来自单一参照群的单一表达类型或可以是多种表达类型。例如, 所述对照水平可以是来自之前测试过的细胞的表达类型数据库。另一个例子可以是参照标记物(如转凝蛋白 2)和本发明标记物之间的比例计尺度(ratiometric measure)。可选地, 所述对照水平可以是早些从同一患者取的一或多个读数或这些读数的均值。

本文所用术语的“多核苷酸”意味着任何长度的单链或双链脱氧核糖核苷酸或核糖核苷酸聚合物, 并包括非限制性例子, 基因的编码和非编码序列, 正义和反义序列, 外显子, 内含子, 基因组 DNA, cDNA, 前体 mRNA(pre-mRNA), mRNA, rRNA, siRNA, miRNA, tRNA, 核酶, 重组多核苷酸, 分离的和纯化的天然存在的 DNA 或 RNA 序列, 合成的 RNA 和 DNA 序列, 核酸探针, 引物, 片段。应类似理解提及的核酸分子。

本文所用的“反义”通常指 DNA 或 RNA 或其组合, 其与至少部分的编码本发明多肽的 mRNA 分子互补, 并能够干扰后转录事件, 如 mRNA 翻译。

本文提供的多核苷酸序列的“片段”是邻接核苷酸的子序列(subsequence), 其能够与目的靶(如长度为至少 10 个核苷酸的序列)特异性杂交。本发明片段包含本发明多核苷酸的邻接核苷酸的 10, 优选 15 个核苷酸, 优选至少 20 个核苷酸, 更优选至少 30 个核苷酸, 更优选至少 40 个核苷酸, 更优选至少 50 个核苷酸并最优选至少 60 个核苷酸。多核苷酸序列的片段可用于反义、基因沉默、三股螺旋(triple helix)或核酶技术, 或作为引物、探针, 包括在微阵列中或用于本发明基于多核苷酸的选择方法。

本文所用的术语“患者”优选哺乳动物患者, 包括人, 和非人哺乳动物如猫、狗、马、母牛、绵羊、鹿、小鼠、负鼠和灵长类动物(包括大猩猩、

猕猴、和黑猩猩)及其它家养家畜或观赏动物(zoo animal)。优选,所述哺乳动物是人。

术语“治疗”和“预防”指治疗和预防(phrophylactic)措施,其停止、逆转或减轻(lessen)前列腺癌或 PIN。所述患者显示出可观察到或可测定(统计学显著性)地使:癌或 PIN 细胞数;肿瘤尺寸;与癌或 PIN 有关的症状的一或多种下降;抑制:肿瘤尺寸;肿瘤生长;转移;提高生活质量。

本文所用的“患者样本”指源自要筛选的患者的生物样本。该生物样本可以是本领域已知的任何适合样本,其中可检测被选标记物的表达。包括从体液或液体获得的个体细胞和细胞群。被测的适合体液的例子是血浆、前列腺按摩液(massage fluid)、血液、精液、淋巴液和尿液。

优选,所述被测样本包括上皮细胞,其源自已知或疑似显示 PIN 或 PRC 的组织,最常为前列腺组织。也可测试来自健康个体的样本。正常健康个体无 PIN 或 PRC 的临床症状或良性前列腺发育不全(hypoplasia)(BPH),并优选在 30 岁以下。可选地,来自前列腺活检的正常区域的正常健康细胞可用作方法的对照。

术语“引物”指通常有游离 3'OH 基团的短多核苷酸,其与模板杂交并用于引导与所述靶互补的多核苷酸的聚合。

术语“探针”指短多核苷酸,其用于在基于杂交的分析法中检测与所述探针互补的多核苷酸序列。该探针可由本文限定的多核苷酸的“片段”组成。

本文所用的术语“多肽”包括任何长度的氨基酸链,但优选至少 5 个氨基酸,优选至少 10,优选至少 15,优选至少 20,优选至少 25,优选至少 30,优选至少 40,优选至少 50,优选至少 60,优选至少 70,优选至少 80,优选至少 90,优选至少 100,优选至少 110,优选至少 120,优选至少 125 个氨基酸,并包括全长蛋白质,其中氨基酸残基通过共价肽键连接。本发明的多肽可以是纯化的天然产物,或可能用重组或合成技术部分或全部地制备。该术语可指多肽,多肽聚集物如二聚体或其它多聚体,融合多肽,多肽片段,多肽变体或其衍生物。

多肽的“片段”是所述多肽的子序列,其执行了生物活性所需的功能和/或提供了所述多肽的三级结构。此术语可指多肽,多肽聚集物如二聚体或其它多聚体,融合多肽,多肽片段,多肽变体或其衍生物。

用于本文公开的多核苷酸或多肽序列的术语“分离的”用于指从它们的

天然细胞环境取出的序列。分离的分子可用任何方法或方法的组合获得，包括生化、重组及合成技术。所述多核苷酸或多肽序列可用至少一步纯化步骤制备。

术语“重组”指多核苷酸序列，其从在天然环境中位于其周围的序列中取出和/或与在其天然环境中不存在的序列重组。

“重组”多肽序列是通过从“重组”多核苷酸序列翻译制备的。

本文所用的术语“变体”指与特异鉴定的序列不同的多核苷酸或多肽序列，其中缺失、取代或添加了一或多个核苷酸或氨基酸残基。变体可以是天然存在的等位变体或非天然存在的变体。变体可来自相同或其它物种，并可包括同源物(homologue)、旁系同源物(paralogue)及直系同源物(orthologue)。在某些实施方案中，本发明多肽和多核苷酸的变体与本发明多肽或多核苷酸具有相同或相似的生物活性。对于多核苷酸和多肽而言的术语“变体”包括本文限定的所有形式的多核苷酸和多肽。

变体多核苷酸序列优选与本发明序列展示了至少 50%，更优选至少 51%，更优选至少 52%，更优选至少 53%，更优选至少 54%，更优选至少 55%，更优选至少 56%，更优选至少 57%，更优选至少 58%，更优选至少 59%，更优选至少 60%，更优选至少 61%，更优选至少 62%，更优选至少 63%，更优选至少 64%，更优选至少 65%，更优选至少 66%，更优选至少 67%，更优选至少 68%，更优选至少 69%，更优选至少 70%，更优选至少 71%，更优选至少 72%，更优选至少 73%，更优选至少 74%，更优选至少 75%，更优选至少 76%，更优选至少 77%，更优选至少 78%，更优选至少 79%，更优选至少 80%，更优选至少 81%，更优选至少 82%，更优选至少 83%，更优选至少 84%，更优选至少 85%，更优选至少 86%，更优选至少 87%，更优选至少 88%，更优选至少 89%，更优选至少 90%，更优选至少 91%，更优选至少 92%，更优选至少 93%，更优选至少 94%，更优选至少 95%，更优选至少 96%，更优选至少 97%，更优选至少 98%，最优选至少 99% 的同一性。与本发明多核苷酸的至少 20 个核苷酸位置，优选至少 50 个核苷酸位置，更优选至少 100 个核苷酸位置，最优选全长的比对窗口可见同一性。

可用总体(global)序列比对程序对候选及受试多核苷酸序列的重叠的全长计算多核苷酸序列同一性(如 Needleman, S. B. and Wunsch, C. D. (1970) J.

Mol. Biol. 48, 443-453)。全面实现 Needleman-Wunsch 总体比对算法见于 EMBOSS 包的 needle 程序(Rice,P. Longden,I. and Bleasby,A. EMBOSS: The European Molecular Biology Open Software Suite, Trends in Genetics June 2000, vol 16, No 6. pp.276-277) , 其可得自 <http://www.hgmp.mrc.ac.uk/Software/EMBOSS/>。欧洲生物信息学研究所(European Bioinformatics Institute) 服务器也提供了在 <http://www.ebi.ac.uk/emboss/align/>对两个序列在线进行 EMBOSS-needle 总体比对的设备。

或者可利用 GAP 程序, 其计算两个序列的最适总体比对, 而不对末端缺口罚分。GAP 在下述论文中进行了描述: Huang, X. (1994) On Global Sequence Alignment. Computer Applications in the Biosciences 10, 227-235。

多核苷酸变体也包括与一或多条特异鉴定的序列显示类似性的那些, 所述特异鉴定的序列可能保留那些序列的功能型等同(functional equivalence), 以及根据合理预测本不会随机出现的那些。此程序在所述序列中发现了类似区域, 并对每个这样的区域报告了“E 值”, 该值是预期能在有固定参照尺寸(reference size)的包含随机序列的数据库中随机看到此匹配的预期次数。在 bl2seq 程序中缺省设定此数据库的尺寸。对于远低于 1 的小 E 值, 该 E 值约为此随机匹配的可能性。

在与任一条特异鉴定的序列相比时, 变体多核苷酸序列优选显示出 E 值不到 1×10^{-5} , 更优选不到 1×10^{-6} , 更优选不到 1×10^{-9} , 更优选不到 1×10^{-12} , 更优选不到 1×10^{-15} , 更优选不到 1×10^{-18} , 最优选不到 1×10^{-21} 。

BLASTN 的应用优选用于测定本发明多核苷酸变体的序列同一性。

多核苷酸序列的同一性可用如下 UNIX 命令行参数测定:

```
bl2seq -i nucleotideseq1 -j nucleotideseq2 -F F -p blastn
```

参数-F F 关闭(turn off)了对低复杂性部分(section)的滤过。参数-p 选择了对序列对合适的算法。bl2seq 程序将序列同一性报告为在一行中同一的核苷酸的数和百分比“Identities =”。

多核苷酸序列同一性和相似性也可用如下方式测定。将受试多核苷酸序列与候选多核苷酸序列用序列比对算法及序列相似性检索工具(如 Genbank, EMBL, Swiss-PROT 及其它数据库中)来比较。Nucleic Acids Res 29:1-10 和 11-16, 2001 提供了在线资源的例子。bl2seq 中的 BLASTN(来自

BLAST 程序套装(suite of programs), 2.2.13 版 March 2007 (Tatiana A. 等, FEMS Microbiol Lett. 174:247-250 (1999), Altschul 等, Nuc.Acids Res 25:3389-3402, (1997)), 其可公开得自 NCBI(<ftp://ftp.ncbi.nih.gov/blast/>)或得自位于 Bethesda, Maryland, USA 的 NCBI。利用 bl2seq 的缺省参数, 只是应关闭对低复杂性部分的滤过。

可选地, 变体多核苷酸与特定的多核苷酸序列, 或其互补物在严谨条件杂交。

术语“在严谨条件杂交”及其语法同义词, 指多核苷酸分子与靶多核苷酸分子(如固定在 DNA 或 RNA 印迹, 如 Southern 印迹或 Northern 印迹上的靶多核苷酸分子)在温度和盐浓度的限定条件杂交的能力。测定在严谨杂交条件杂交的能力可通过, 最初在不太严谨的条件杂交, 然后将严谨度提高到理想的严谨度。

对于长度超过约 100 个碱基的多核苷酸分子, 通常严谨杂交条件是低于天然双链的解链温度(T_m)不超过 25-30°C(例如 10°C)(常见 Sambrook 等, Eds, 1987, Molecular Cloning, A Laboratory Manual, 2nd Ed. Cold Spring Harbor Press; Ausubel 等, 1987, Current Protocols in Molecular Biology, Greene Publishing, incorporated herein by reference)。超过约 100 个碱基的多核苷酸分子的 T_m 可用公式 $T_m = 81.5 + 0.41\%(G + C - \log(Na^+))$ 计算(Sambrook 等, Eds, 1987, Molecular Cloning, A Laboratory Manual, 2nd Ed. Cold Spring Harbor Press; Bolton 和 McCarthy, 1962, PNAS 84:1390)。对于长度超过 100 个碱基的多核苷酸的常用严谨条件会是杂交条件, 如在 6X SSC, 0.2% SDS 的溶液中预洗涤; 在 65°C, 6X SSC, 0.2% SDS 过夜杂交; 随后, 在 1X SSC, 0.1% SDS 于 65°C 各洗涤两次, 每次 30 分钟, 及在 0.2X SSC, 0.1% SDS 于 65°C 各洗涤两次, 每次 30 分钟。

在一个实施方案中, 严谨条件在 42°C 使用 50%甲酰胺、5 x SSC、50 mM 磷酸钠(pH 6.8)、0.1%焦磷酸钠、5x Denhardt's 溶液、超声的鲑鱼精子 DNA (50 $\mu\text{g/ml}$)、0.1% SDS、和 10%硫酸葡聚糖, 然后于 42°C 在 0.2 x SSC 和 50%甲酰胺在 55°C 中多次洗涤, 随后在包含 EDTA 的 0.1 x SSC 中在 55°C 洗涤。

对于长度不足 100 个碱基的多核苷酸分子, 严谨杂交条件举例是低于 T_m 5-10°C。长度不足 100 bp 的多核苷酸分子的 T_m 平均会下降约(500/寡核

苷酸长度) $^{\circ}\text{C}$ 。

对于称为肽核酸(PNAs)的 DNA 模拟物(Nielsen 等, Science. 1991 Dec 6;254(5037):1497-500), 其 T_m 值高于 DNA-DNA 或 DNA-RNA 杂种的那些, 并可用 Giesen 等, Nucleic Acids Res. 1998 Nov 1;26(21):5004-6 所述的公式计算。对长度不足 100 个碱基的 DNA-PNA 杂种的严谨杂交条件举例是低于 T_m 5-10 $^{\circ}\text{C}$ 。

变体多核苷酸也包括与本发明序列不同的多核苷酸, 但遗传密码简并性的结果是编码了多肽, 所述多肽与本发明多核苷酸所编码的多肽具有类似活性。不改变所述多肽的氨基酸序列的序列改变是“沉默变化(silent variation)”。除了 ATG (蛋氨酸)和 TGG (色氨酸)之外, 相同氨基酸的其它密码也可用本领域共知的技术改变, 如在特定宿主生物中使密码表达最适。

本发明也包括了多核苷酸序列改变, 其使得所编码的多肽序列内的一或多个氨基酸被保守取代而不显著改变其生物活性。熟练技术人员了解制备表型沉默的氨基酸取代的方法(见例如 Bowie 等, 1990, Science 247, 1306)。

由于沉默变化以及所编码的多肽序列内的保守取代而得到的变体多核苷酸可用 bl2seq 程序经上述 tblastx 算法来测定。

本文所用的术语“反义-寡核苷酸”包括与靶序列完全互补的核苷酸, 及具有一或多个核苷酸错配的那些, 只要所述反义-寡核苷酸可以特异性地与所述靶序列杂交。例如, 本发明的反义-寡核苷酸包括下述多核苷酸, 其与本发明核酸序列, 优选 PSPU 43 (SEQ ID NO:3)的至少 15、至少 20、至少 30、至少 40、至少 50、优选 75、优选 100、更优选 200 个邻接核苷酸或全长, 具有至少 70%或更高, 优选至少 75%或更高、至少 76%、至少 77%、至少 78%、至少 79%、至少 80%或更高、更优选至少 81%、至少 82%、至少 83%、至少 84%、至少 85%、至少 86%、至少 87%、至少 88%、至少 89%、至少 90%或更高、甚至更优选 91%、92%、93%、94%、95%、96%、97%、或 99%的同一性。上述本领域公知的算法也可用于测定同一性。另外, 所述反义-寡核苷酸的衍生物或经修饰产物也可用作本发明中的反义-寡核苷酸。

对于多肽而言的术语“变体”包括天然存在的、重组及合成制备的多肽。变体多肽序列优选与本发明序列展示至少 50%, 更优选至少 51%, 更优选至少 52%, 更优选至少 53%, 更优选至少 54%, 更优选至少 55%, 更优选至少

56%, 更优选至少 57%, 更优选至少 58%, 更优选至少 59%, 更优选至少 60%, 更优选至少 61%, 更优选至少 62%, 更优选至少 63%, 更优选至少 64%, 更优选至少 65%, 更优选至少 66%, 更优选至少 67%, 更优选至少 68%, 更优选至少 69%, 更优选至少 70%, 更优选至少 71%, 更优选至少 72%, 更优选至少 73%, 更优选至少 74%, 更优选至少 75%, 更优选至少 76%, 更优选至少 77%, 更优选至少 78%, 更优选至少 79%, 更优选至少 80%, 更优选至少 81%, 更优选至少 82%, 更优选至少 83%, 更优选至少 84%, 更优选至少 85%, 更优选至少 86%, 更优选至少 87%, 更优选至少 88%, 更优选至少 89%, 更优选至少 90%, 更优选至少 91%, 更优选至少 92%, 更优选至少 93%, 更优选至少 94%, 更优选至少 95%, 更优选至少 96%, 更优选至少 97%, 更优选至少 98%, 并最优选至少 99% 的同一性。在本发明多肽的至少 20 个氨基酸位置、优选至少 50 个氨基酸位置、更优选至少 100 个氨基酸位置、并最优选全长的比较窗口中可见同一性。

多肽变体也包括与一或多条特异鉴定的序列显示相似性的那些以及根据合理预测本不会随机出现的那些, 所述特异鉴定的序列可能保留那些序列的功能型等同。

多肽序列同一性和相似性可以如下方式测定。比较受试多肽序列与候选多肽序列是用 bl2seq 中的 BLASTP (来自 BLAST 程序套装, 2.2.13 版 May 2007), 其公开得自 NCBI (<ftp://ftp.ncbi.nih.gov/blast/>)。bl2seq 的缺省参数利用, 除了关掉滤过低复杂性的区域。

多肽序列的相似性可用下 UNIX 命令行参数检测:

```
bl2seq -i peptideseq1 -j peptideseq2 -F F -p blastp
```

在与任一条特异鉴定的序列相比时, 变体多核苷酸序列优选显示出 E 值不到 1×10^{-5} , 更优选不到 1×10^{-6} , 更优选不到 1×10^{-9} , 更优选不到 1×10^{-12} , 更优选不到 1×10^{-15} , 更优选不到 1×10^{-18} , 最优选不到 1×10^{-21} 。

参数 -F F 关闭(turn off)了对低复杂性部分(section)的滤过。参数 -p 选择了对序列对合适的算法。此程序在所述序列中发现了类似区域, 并对每个这样的区域报告了“E 值”, 该值是预期能在有固定参照尺寸的包含随机序列的数据库中随机看到此匹配的预期次数。对于远低于 1 的小 E 值, 该 E 值约为此随机匹配的可能性。

也可用总体序列比对程序对在候选和受试多肽序列之间的重叠的全长

计算多肽序列同一性。EMBOSS-needle (自 <http://www.ebi.ac.uk/emboss/align/> 可得)和 GAP(Huang, X. (1994) On Global Sequence Alignment. Computer Applications in the Biosciences 10, 227-235.)如上述, 也适合总体序列比对程序来计算多肽序列同一性。

本文所述的 BLASTP 的用途优选用于本发明多肽变体的测定。

本发明也包括所述多肽序列的一或多个氨基酸的保守取代, 但不显著改变其生物活性。保守取代常包括将一个氨基酸取代为另一个具有类似性质的氨基酸, 如下组内的取代: 缬氨酸、甘氨酸; 甘氨酸、丙氨酸; 缬氨酸、异亮氨酸、亮氨酸; 天冬氨酸、谷氨酸; 天冬酰胺、谷氨酰胺; 丝氨酸、苏氨酸; 赖氨酸、精氨酸; 和苯丙氨酸、酪氨酸。也可用下表 1 内的其它保守取代。

表 1

原始残基	取代举例
Ala (A)	val; leu; ile; gly
Arg (R)	lys; gln; asn
Asn (N)	gln; his; lys; arg
Asp (D)	glu
Cys (C)	ser
Gln (Q)	asn; his
Glu (E)	asp
Gly (G)	pro; ala
His (H)	asn; gln; lys; arg
Ile (I)	leu; val; met; ala; phe; 正亮氨酸
Leu (L)	正亮氨酸; ile; val; met; ala; phe
Lys (K)	arg; gln; asn
Met (M)	leu; phe; ile
Phe (F)	leu; val; ile; ala; tyr
Pro (P)	ala; gly
Ser (S)	thr
Thr (T)	ser
Trp (W)	tyr; phe
Tyr (Y)	trp; phe; thr; ser
Val (V)	ile; leu; met; phe; ala; 正亮氨酸

天然存在的残基基于共有侧链性质而分组:

- (1) 疏水: 正亮氨酸, met, ala, val, leu, ile;
- (2) 中性亲水: cys, ser, thr;
- (3) 酸性: asp, glu;
- (4) 碱性: asn, gln, his, lys, arg;
- (5) 影响链取向(orientation)的残基: gly, pro; 和

(6) 芳香族: trp, tyr, phe.

非保守取代需要将这些类别之一的成员替换为另一类别的成员。

其它变体包括带有影响肽稳定性的修饰的肽。这些类似物可包含例如, 肽序列中的一或多个非肽键(其替代了肽键)。也包括类似物, 其包含天然存在的 L-氨基酸之外的残基如 D-氨基酸, 或非天然存在的合成氨基酸, 如 β 或 γ 氨基酸及环状类似物。

可用本领域公知的诱变方法制备取代、缺失、添加或插入。熟练技术人员会知道制备表型沉默氨基酸替代的方法(见例如 Bowie 等, 1990, Science 247, 1306)。

本发明多肽内也包括在合成之中或之后已经修饰的那些, 所述修饰如通过生物素化、苄基化、糖基化、磷酸化、酰胺化、通过用封闭(blocking)/保护基团衍生化(derivatization)等等。这些修饰可提高所述多肽的稳定性或活性。

术语“基因构建体”指多核苷酸分子, 常为双链 DNA, 其中本可插入其它的多核苷酸分子(插入的多核苷酸分子)如, 但不限于 cDNA 分子。基因构建体可包含允许转录插入的多核苷酸分子并可选翻译转录物为多肽的必要元件。所述插入的多核苷酸分子可衍生自宿主细胞, 或可衍生自不同的细胞或生物体和/或可能是重组的多核苷酸。一旦在宿主细胞内, 所述基因构建体就可能整合到宿主的染色体 DNA 内。所述基因构建体可与载体连接。

术语“载体”指多核苷酸分子, 常为双链 DNA, 其用于将基因构建体转运到宿主细胞中。所述载体可能能够在至少一个其它宿主系统如大肠杆菌中复制。

术语“表达构建体”指包含必要元件的基因构建体, 所述元件允许转录插入的多核苷酸分子, 并可选将转录物翻译为多肽。表达构建体常以 5'到 3'的方向包含:

- a) 在将会用所述构建体转化的宿主细胞中起作用的启动子,
- b) 被表达的多核苷酸, 和
- c) 在将会用所述构建体转化的宿主细胞中起作用的终止子。

术语“编码区”或“开放读码框”(ORF)指基因组 DNA 序列的正义链或 cDNA 序列, 其能受合适的调控序列控制产生转录产物和/或多肽。所述编码序列通过存在 5'翻译起始密码子和 3'翻译终止密码子来被鉴定。在插入

基因构建体时，“编码序列”能够在与启动子和终止子序列和/或其它调控元件可操作地连接时被表达。

“可操作地连接”意味着将表达的序列处于受调控元件控制的位置，所述调控元件包括启动子、转录控制序列、翻译控制序列、复制起点、组织特异性调控元件、时间调控元件、增强子、聚腺苷化(polyadenylation)信号、阻抑物(repressor)和终止子。

术语“非编码区”指在翻译起始位点上游和翻译终止位点下游的未翻译序列。这些序列也分别指 5' UTR 和 3' UTR。这些区域包括转录起始和终止所需以及调控翻译有效性的元件。

终止子是终止转录且发现于在被翻译序列下游的基因的 3' 未翻译末端的序列。终止子是 mRNA 稳定性的重要决定物，且有时已经发现它有空间调控功能。

术语“启动子”指非转录的顺式-调控元件，其位于调控基因转录的编码区的上游。启动子包含顺式-起始元件，其特定针对(specify)转录起始位点及保守盒如 TATA 盒，以及转录因子所结合的基序。

本发明多核苷酸或多肽的术语“要改变的表达”及“已改变的表达”应包括如下情况，其中对应于本发明多核苷酸的基因组 DNA 被修饰，因此使本发明多核苷酸或多肽的表达改变。对基因组 DNA 的修饰可通过本领域公知诱导突变的遗传转化或其它方法。“已改变的表达”可与信使 RNA 和/或所产生的多肽的量的提高或下降有关，也可由于多核苷酸及所产生的多肽的序列改变而导致多肽活性改变。

发明内容

通过利用新的生物信息方法来采集(mine)测序前列腺 cDNA 文库，申请人已经鉴定出对于前列腺癌或 PIN 的新标记物。数据漫游(Data-panning)是一种测定转录物对给定(given)组织具有的特异性程度的技术。此方法利用了 UniGene 数据库(www.ncbi.nlm.nih.gov/UniGene)。评估 UniGene 簇内的 EST 的文库来源，并保留来自特定组织的记录的那些。此记录表示为 UniGene 簇内 EST 总数的百分比。百分比越高，该转录物在特殊指定的组织之外的组织中被检出的几率越低。此方法具有超过其它技术如 cDNA 微阵列的优点(Carlisle 等，2000)，这是由于无偏移的(unbiased)基因选择，以及分辨疾

病状态差异的区分能力更高。之前努力绘制前列腺癌的基因表达图谱所用的方法限于所分析的表达序列标签(expressed sequence tag)数目(Huang 等, 1999)或在基因选择上有偏移(Carlisle 等, 2000)。

申请人已经从此分析鉴定了新标记物,据认为其表达随前列腺癌或 PIN 的进展而改变。此标记物也可能是开发治疗前列腺癌或 PIN 的药物的有潜力的新靶。

所述标记物列于下表 1 中,且序列表中给出全部序列。所述标记物的表达水平在前列腺癌患者中有改变。为了便利,所述标记物在本文被称为前列腺特异性单基因(prostate specific unigene) (PSPU)标记物。所述标记物可以是 DNA 或 RNA 序列,基因或染色体片段。任何由基因编码的对应多肽都被称为 PSPU 多肽或 PSPU 蛋白质。

前列腺癌或 Pin 所涉及的标记物

名称	单基因#	SEQ ID NO:	%富集	组织
PSPU 43	161160	3	83	前列腺, 其它

本发明的或其它本文所述的核酸分子可从组织用多种本领域一般技术人员公知的技术分离。例如,这些多核苷酸可通过利用聚合酶链式反应(PCR)来分离,如 Mullis 等, Eds. 1994 The Polymerase Chain Reaction, Birkhauser 所述。本发明核酸分子可用引物扩增,如本文所限定,衍生自本发明的多核苷酸序列。

分离多核苷酸的其它方法包括利用所有的或部分的本发明多核苷酸,特别是具有 SEQ ID NO:3 所述序列的多核苷酸,作为杂交探针。将标记的多核苷酸探针与固定在固相支持物(如硝酸纤维素滤膜或尼龙膜)上的多核苷酸杂交的技术,可用于筛选基因组或 cDNA 文库。类似地,可将探针与珠子偶联并杂交到靶序列上。可用本领域公知的方案如磁性分离(magnetic separation)进行分离。严谨杂交和洗涤条件举例如上。

多核苷酸片段可用本领域公知的技术,如限制性内切酶消化和寡核苷酸合成来制备。

由此,一方面,本发明提供了分离的核酸,该核酸包含 SEQ ID NO:3,其功能等同变体或片段,或与这些的任一在严谨条件杂交的序列。

另一方面,本发明提供了分离的核酸分子,其由本发明核酸序列的至

少 10 个核苷酸的片段组成, 所述核酸序列优选 SEQ ID NO:3 或其互补物, 其在严谨条件与下述杂交:

- (a) 本发明的核酸序列, 优选 SEQ ID NO: 3 或其互补物;
- (b) 对应于本发明核酸序列或其互补物的 cDNA 的全长编码序列;
- (c) (a)或(b)的反向互补物。

严谨条件如上讨论。

所述核酸分子可以是至少 20 个核苷酸, 至少 30 个核苷酸, 至少 40 个核苷酸, 至少 50 个核苷酸, 至少 60 个核苷酸, 至少 70 个核苷酸, 至少 80 个核苷酸, 至少 90 个核苷酸, 或优选是至少 100 个核苷酸。

部分的多核苷酸序列可在本领域公知方法中用作探针来鉴定样本中的对应全长多核苷酸序列。这些方法包括基于 PCR 的方法, 5'RACE (Methods Enzymol. 218: 340-56 (1993); Sambrook 等, 见上文)及基于杂交的方法、基于计算机/数据库的方法。可检测标记物如放射性同位素、荧光、化学发光和生物发光标记物, 可用于利于检测。反向 PCR 也允许获得未知序列, 位于本文公开的多核苷酸序列侧翼, 从基于已知区域的引物开始(Triglia 等, *Nucleic Acids Res* 16, 8186, (1998))。此方法使用多种限制性酶来产生在基因已知区域内的合适片段。然后通过分子内连接来散布(circularize)此片段, 并用作 PCR 模板。从已知区域设计多样性引物。为了物理上(physically)装配全长克隆, 可用标准分子生物学方法(Sambrook 等, 见上文)。允许本发明多核苷酸扩增的引物和引物对也构成了本发明的其它方面。

可用所述方法鉴定变体(包括直系同源物)。可用基于 PCR-的方法鉴定变体多核苷酸(Mullis 等, Eds. 1994 *The Polymerase Chain Reaction*, Birkhauser)。通常, 引物的多核苷酸序列, 用于通过 PCR 扩增多核苷酸分子的变体, 可基于编码对应氨基酸序列的保守区的序列。

用于鉴定变体多核苷酸的其它方法包括使用所有的或部分的特定多核苷酸作为杂交探针来如上述筛选基因组或 cDNA 文库。可使用通常基于编码对应氨基酸序列的保守区的序列的探针。杂交条件也可不如筛选与所述探针等同的序列时所用的那些条件严谨。

所述变体序列, 包括多核苷酸和多肽变体, 也可用上述讨论的基于计算机的方法来鉴定。

一组相关序列的多重序列比对可用如下进行: CLUSTALW (Thompson,

等, Nucleic Acids Research, 22:4673-4680 (1994), <http://www-igbmc.u-strasbg.fr/BioInfo/ClustalW/Top.html>) 或 T-COFFEE (Cedric Notredame 等, J. Mol. Biol. 302: 205-217 (2000)))或 PILEUP, 其使用递增的成对比对(Feng 等, J. Mol. Evol. 25, 351 (1987))。

存在发现基序或签名序列的类型识别的软件应用。例如, MEME (Multiple Em for Motif Elicitation)在一组序列中发现基序和签名序列, 且 MAST (Motif Alignment and Search Tool)在询问(query)序列中使用这些基序来鉴定类似或相同基序。提供的 MAST 结果为带合适统计数据的系列比对及所发现基序的可视综述(overview)。MEME 和 MAST 是由 University of California, San Diego 开发的。

PROSITE (Bairoch 等, Nucleic Acids Res. 22, 3583 (1994); Hofmann 等, Nucleic Acids Res. 27, 215 (1999))是鉴定从基因组或 cDNA 序列翻译的未表征蛋白质的功能的方法。该 PROSITE 数据库(www.expasy.org/prosite)包含生物显著类型和图谱, 并被设计从而可与合适的计算工具使用, 来将新序列与已知族的蛋白质比对, 或测定在所述序列中存在哪个已知的结构域(Falquet 等, Nucleic Acids Res. 30, 235 (2002))。Prosearch 是能用给定序列类型或签名检索 SWISS-PROT 和 EMBL 数据库的工具。

可根据蛋白质与在相同基因组(旁系同源物)或不同基因组(直系同源物)的其它蛋白质的序列相关性对它们作分类。直系同源(Orthologous)基因是从共同的祖先基因经物种形成而进化的、并通常在进化时保留相同功能的基因。旁系同源基因是在基因组内复制的基因, 并可获得可能与原来有关的新特异性(specificities)或经修饰功能的基因。对系统进化分析(phylogenetic analysis)方法的综述见 Tatusov 等, Science 278, 631-637, 1997)。

除了上述的计算机/数据库方法, 可用物理方法, 例如用通过重组 DNA 技术抗本发明多肽的抗体筛选表达文库来鉴定多肽变体(Sambrook 等, Molecular Cloning: A Laboratory Manual, 2nd Ed. Cold Spring Harbor Press, 1987), 对该重组 DNA 技术的描述见 Sambrook 等, 或通过借助于这些抗体鉴定来自天然来源的多肽。

多肽(包括变体多肽)可用本领域公知的肽合成方法制备, 所述方法如使用固相技术的直接肽合成(如 Merrifield, 1963, in J. Am Chem. Soc. 85, 2149; Stewart 等, 1969, in Solid-Phase Peptide Synthesis, WH Freeman Co, San

Francisco California; Matteucci 等, J. Am. Chem. Soc. 103:3185-3191, 1981)或自动合成, 例如使用 Applied Biosystems(California, USA)的合成仪。所述多肽的突变形式也可用合成方法制备, 所述方法如编码氨基酸序列的 DNA 的定点诱变, 如 Adelman 等, DNA 2, 183(1983)所述。

所述多肽和变体多肽也可用多种本领域公知的技术从天然来源分离或纯化(如 Deutscher, 1990, Ed, Methods in Enzymology, Vol. 182, *Guide to Protein Purification*,)。技术包括但不限于 HPLC、离子交换层析、和免疫层析。

可选地, 可在适合宿主细胞中重组表达所述多肽和变体多肽, 并从下面讨论的细胞中分离。所述多肽和变体用于产生抗体并用于产生配体及其它用途。

因此, 本发明也提供了本发明核酸分子所编码的分离多肽。

本发明的具体多肽包括具有都如随后序列表所述的 SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13 和 SEQ ID NO:14 的氨基酸序列的多肽。也包括本文限定的这些多肽的功能等同性变体和片段, 以及在严谨条件与那些序列杂交的序列。

本文所述的基因构建体可包含本发明的一或多个公开的多核苷酸序列和/或编码所述公开多肽的多核苷酸, 并可用于转化, 例如细菌、真菌、昆虫、哺乳动物或植物生物体。本发明的基因构建体应包括本文限定的表达构建体。包括载体(如 pBR322, pUC18, pU19, Mp18, Mp19, ColE1, PCR1 和 pKRC)、噬菌体(如 lambda gt10)、及 M13 质粒(如 pBR322, pACYC184, pT127, RP4, pIJ101, SV40 和 BPV)、粘粒、YACS、BAC 穿梭载体如 pSA3, PAT28 转座子(如 US 5,792,294 所述)等等。

所述构建体很容易包括选择性基因或可选择性标记物。通常使用抗生素如氨苄青霉素、氨甲蝶呤(methotrexate)或四环素的抗性标记物。

用于此构建体的启动子包括 β 内酰胺酶、碱性磷酸酶、色氨酸及 tac 启动子系统, 它们都是本领域公知的。酵母启动子包括但不限于 3-磷酸甘油激酶、烯醇化酶、己糖激酶、丙酮酸脱羧酶、葡糖激酶、和甘油醛(glyceraldehyde)-3-磷酸(phosphanate)脱氢酶。

可用增强子作用于启动子来增强转录。在本文使用的合适增强子包括 SV40 启动子、巨细胞病毒(cytomegalovirus)早期启动子增强子、珠蛋白、白

蛋白、胰岛素等。

制备并使用基因构建体和载体的方法是本领域公知的，并通常如 Sambrook 等，(见上文)和 Ausubel 等，Current Protocols in Molecular Biology, Greene Publishing, 1987 所述。用所述载体转化被选宿主细胞的方法也是已知的，例如氯化钙处理，如 Cohen, SN; PNAS 69, 2110, 1972 所述。

包含所述基因构建体和载体的宿主细胞可源自原核或真核来源，例如酵母、细菌、真菌、昆虫(如杆状病毒)、动物、哺乳动物或植物生物体。最常用作宿主细胞的原核细胞是大肠杆菌菌株。其它原核生物宿主包括但不限于假单胞菌属、芽孢杆菌属、沙雷氏菌属、克雷伯氏菌属、链霉菌属、李斯特氏菌属、糖酵母属(*Saccharomyces*)、沙门氏菌属和分枝杆菌属。

表达重组蛋白质的真核细胞包括但不限于 Vero 细胞、HeLa、CHO(中国仓鼠卵巢细胞)、293、BHK 细胞、MDCK 细胞和 COS 细胞，以及前列腺癌细胞系，如 PrEC、LNCaP、Du 145 和 RWPE-2。所述细胞得自 ATCC, Virginia, USA。

与本发明核酸分子的表达匹配的原核启动子包括本领域已知的组成型启动子(如 λ 噬菌体的 int 启动子和 pBR322 的 β 内酰胺酶基因序列的 bla 启动子)和可调控启动子(如 lacZ, recA 和 gal)。PSPU 43 编码序列上有的核糖体结合位点也是表达所需的。

包含基因构建体如表达构建体的宿主细胞，可用于重组制备多肽的方法。这些方法是本领域公知的(见例如 Sambrook 等，见上文)。所述方法常涉及在合适培养基中在适合或有助于表达并选择本发明多肽的条件培养宿主细胞。带可选择性标记物的细胞可另外生长在适合选择宿主细胞的培养基中，所述宿主细胞表达本发明的多肽。表达本发明多肽的转化宿主细胞在适合表达所述多肽的条件选择并培养。所表达的重组多肽可用本领域公知方法自培养基分离并纯化，所述方法包括硫酸铵沉淀、离子交换层析、凝胶滤过、亲和层析、电泳等(如 Deutscher, Ed, 1990, Methods in Enzymology, Vol 182, Guide to Protein Purification)。宿主细胞也可用于制备由本发明表达多肽所产生的产物的方法。

本发明也提供了动物模型。有前列腺癌倾向的宿主细胞或动物可用于测试可用于治疗前列腺癌的化合物，或用于鉴定涉及会导致癌症的化合物。动物模型特别用于测试目的。本文限定的非人患者可能适合动物使用。优

选所述动物是啮齿类动物或兔子。优选使用大鼠、特别是小鼠。

动物模型中可掺入编码本发明多肽或其反义或 siRNA 序列的基因，其不天然存在于动物中(外源性)，或不出现在引入了基因的位置，或与引入的基因不出现在相同构象(configuration)中。动物模型也包括对应于本发明核酸分子的内源性基因被改变、破坏或去除的动物。

动物的种系改变可用任何已知的本领域方法实现。例如基因可通过显微注射接合子(zygote)来掺入动物基因组(Brinster 等, PNAS (USA) 82:4438-4442(1985); 通过用逆转录病毒感染卵裂球(blastomer)或囊胚腔(blastocoel)来整合病毒(Jaenuch, R; PNAS (USA)73:1260-1264 (1976), Johner, D 等, Nature 298:623-628(1982); 或通过转化胚胎干细胞(Lovel-Badge, R. H., Tetracarcinomas and Embryonic Stem Cells: A Practical Approach, Robertson, E. J. 等, DRL Press, Oxford, 153-182(1987)。也见 Houdebine, Transgenic Animals – Generation and Use (Harwood Academic, 1997)。

另一方面，本发明提供了诊断和/或预后患者中的前列腺癌、PIN 或发作这些疾病的倾向的方法。

在一个实施方案中，通过测定患者样本中本发明核酸分子如 PSPU 43 (SEQ ID NO:3)的表达水平来进行所述方法。所述分子表达水平相对该分子对照水平的改变表示，该受试者患有或有风险患有 PIN, PRC。所述分子表达水平的改变包括鉴定所述患者样本中存在或缺无所述分子。

在另一实施方案中，本发明提供了测试患者的前列腺上皮内瘤(PIN)、前列腺癌(PRC)状态的方法，该方法包括测定 PSPU 43 (SEQ ID NO:3)或其它本发明核酸分子在患者样本的表达水平，其中所述分子的表达水平比对照水平升高表明该患者有 PIN、PRC 状态或有风险患有 PIN 或 PRC。

如果本发明分子的表达水平与对照水平相差统计学显著性的量，可以认为其表达水平发生改变，包括上升。通常与对照水平相差 5%以上，10%以上，20%以上，30%以上，40%以上，优选 50%以上，或更多。统计学显著性可以可选计算为 $P < 0.05$ 。在其它可选方案中，可以借助于分析法参考限值(assay reference limit)或参考区间(reference interval)来测定偏差。可以从直观评估(intuitive assessment)或非参数方法计算这些。总体而言，这些方法计算 0.025, 和 0.975 分位数(fractile)为 $0.025 \cdot (n+1)$ 和 $0.975 \cdot (n+1)$ 。这些方法是本领域公知的。见例如 the Immunoassay Handbook, 3rd edition, ed. David

Wild. Elsevier Ltd, 2005; 和 Solber H. Approved Recommendation (1987) Collected reference values. Determination of reference limits. Journal of Clinical Chemistry and Clinical Biochemistry 1987, 25:645-656。

存在对照中不存在的标记物，或不存在对照中存在的标记物，也包括在表达水平的改变中。

在样本中存在标记物及它们的表达水平可根据本领域公知方法测定，如 Southern Blotting、Northern Blotting、FISH 或定量 PCR 来定量 mRNA 的转录[(Thomas, Pro. NAH, Acad. Sci. USA 77: 5201-5205 1980), (Jain KK, Med Device Technol. 2004 May; 15(4):14-7)]、点印迹、(DNA 分析)或体外杂交，其基于本文提供的标记物序列使用合适标记的探针。

因此，本发明也提供了检测样本中存在的本发明核酸分子，优选 PSPU 43 (SEQ ID NO:3)的分析法，所述方法包括：

- (a) 使所述样本与多核苷酸探针接触，所述探针在严谨杂交条件与所述核酸序列杂交；和
- (b) 检测样本中杂交复合物的存在。

优选所述杂交探针是标记的探针。标记物的例子包括荧光的、化学发光的、放射性酶的和生物素-亲和素的标记物。用如上述那些的本领域公知方法来标记并观察标记的探针。

为了方便，所述核酸探针可固定在固相支持物上，包括树脂(如聚丙烯酰胺)、碳水化合物(如琼脂糖(sepharose))、塑料(如聚碳酸酯(polycarbonate))及胶乳(latex)珠子。

如上讨论，所述核酸分子探针可以是 RNA 或 DNA 分子。优选的探针包括

Pspu43	正向	5'-AACAAATATAAAGTACCAGACACTCCA -3' (SEQ ID NO:15)
	反向	5'-ATCTCCAGATCTTCCTTCTAGCC -3' (SEQ ID NO:16)

所述核酸标记物的表达水平可用本领域公知技术测定，如 RT-PCR 和电泳技术，包括 SDS-PAGE。用这些技术，扩增患者样本中的本发明核酸分子，及 PSPU 43 (SEQ ID NO:3)的 DNA 或 cDNA 序列，并测定 DNA 或 cDNA 或 RNA 的水平。

在另一方法中，可直接测定样本中的 DNA、cDNA 或 RNA 水平而不需

扩增。

现在优选的方法是 Northern 印迹杂交分析。用于 Northern 印迹杂交分析的探针可以基于本文鉴定的标记物序列来制备。探针优选包括参照序列的至少 10, 至少 15, 至少 20, 至少 30, 至少 40, 至少 50, 优选 75, 优选 100, 或更优选 200 或更多个邻接核苷酸。

可选地, 可用基于逆转录的 PCR (RT-PCR) 分析法用对所述核酸序列特异的引物来测量表达水平。如果需要, 将样本中所述标记物的表达水平相比表达不依赖于被测参数或条件的对照核酸分子作比较。对照核酸分子指其表达在 PIN/PRC 状态和健康状态无差别的分子。对照分子的表达水平可用于使所比较群的表达水平正态化。此对照分子的例子是转凝蛋白 2。所述标记物会改变疾病的表达水平。

可选地, 对于肽标记物, 可用抗体, 其能识别特异性双链体, 包括 DNA 双链体, RNA 双链体, 及 DNA-RNA 杂种双链体或 DNA-蛋白质双链体。反过来, 可标记所述抗体, 并可进行所述分析法, 此时所述双链体结合到表面上, 从而在表面上形成双链体时, 可检测结合到所述双链体上的抗体的存在。

由此, 另一方面本发明提供了检测患者样本中本发明核酸分子所编码的多肽或其功能等同变体或片段的存在的分析法, 该方法包括使所述样本与本发明抗体在形成免疫复合物的条件接触, 并检测样本中结合多肽的存在。

检测样本中的本发明抗体的反向(reverse)测试也是可行的。此时, 所述样本与本发明的肽在适合形成免疫复合物的条件接触, 并检测结合抗体的存在。

为此目的本领域常可用的免疫分析法包括放射性免疫分析(RIA)、酶免疫吸附分析(ELISA)等等(Lutz 等, Exp. Cell. Res. 175: 109-124 (1988))。

可以任选通过免疫方法如免疫组织化学染色细胞或组织切片, 和细胞培养物或体液的分析法来直接定量表达水平, 来测定标记物表达。用于免疫组织化学染色和/或用于样本液体的分析法的抗体优选是单克隆或多克隆的, 并如下以更多细节讨论。可方便地制备抗体, 所述抗体针对本发明多肽或针对合成的肽基于本文公开的 DNA 序列, 或针对与本发明核酸分子(特别是 PSPU43)的 DNA 融合的外源序列, 并编码特定抗体表位。

用本发明生物标记物从血细胞监测前列腺健康,可使用下述这些技术进行,例如 Fehm 等, *Clinical Cancer Research*, 8:2073-2084 (2002) “Cytogenetic evidence that circulating epithelial cells in patients with carcinoma are malignant”所述。

尿取样也是可行的。尿道经过前列腺使得前列腺细胞被排到尿中。对尿中核酸 PCA3 标记物的前列腺试验已经由 Bostwick Laboratories (<http://bostwicklaboratories.com/patientservices/uPM3.html>)开发。可对 PSPU 43 进行类似试验。

患者样本中一或多种 PSPU 标记物相对正常对照水平的表达改变,表明该患者患有或有风险患上 PIN 或 PRC。该改变是升高还是下降取决于疾病的阶段。通常,已经显示 PSPU 43 在前列腺癌患者是过表达的。然而,表达不足,例如在前列腺癌的晚期(advanced stages)也是可能的。

其它标记物也可与本发明的 PSPU 标记物联用。可用的标记物包括已知的前列腺癌标记物如 PCA3, PSA。也可使用已经显示在前列腺癌患者表达不足的转凝蛋白 1。也可包括不随疾病状态而改变的基准(benchmark)或参照标记物。转凝蛋白 2 可用于此目的。将 PSPU 标记物的水平与其它标记物相关可提高预测性诊断本发明 PSPU 标记物的监测值。PSPU 43 与已知前列腺癌标记物使用可提高患者结果的预测或诊断值。

对数个肽标记物的分析可用单一被测样本同时或分别进行。优选同时的或二位模式分析法(two site format assays)。微阵列或生物芯片分析时特别有用的。所述分析法或芯片可有数个审慎的可寻址定位(addressable location),其包含针对一或多种标记物包括本发明 PSPU 标记物的抗体。US2005/0064511 提供了对本发明所用芯片和技术的描述。

因此,在另一个实施方案中,本发明提供了监测受试者对治疗 PIN 或 PRC 的反应的方法,所述方法包含测定本发明核酸分子,优选 PSPU 43 (SEQ ID NO:3)在患者样本中的表达水平,并将所述分子的水平与对照水平比较,其中测定水平与对照水平的统计学显著性改变表示对治疗有反应。

PSPU 分子(特别是 PSPU 43)的统计学显著性升高表明了 PIN 或 PRC,或者所述结果可与非 PSPU 43 标记物的改变(如包括上述讨论的那些)相关。这些标记物水平与对照的改变,与 PSPU 分子相对对照的升高偶联,更可以表明 PRC 或 PIN。

在将要监测受试者时，可随时取数种生物样本。连续取样允许标记物水平有改变，特别是被测的 PSPU 43 有随时的改变。取样可为癌症出现、癌症严重程度提供信息，该治疗方案可以是合适的，使用了对治疗方案的反应，以及长期预后。在现场护理(points of care)如在大夫办公室、在临床表现(clinical presentation)、在住院期间、在门诊病人、或在常规健康筛选时可进行分析。

本发明方法也可与一或多种风险因子的分析联用进行，例如但不限于年龄、家族病史及伦理背景。

本文的方法可用作疗法的指导。例如，用什么疗法起始并且在什么时候。

在另一方面，本发明提供了包含一或多种检测试剂的试剂盒，所述检测试剂特异性结合本发明的 PSPU 核酸标记物分子或由该核酸序列编码的多肽。优选，该试剂盒包括 PSPU43 (SEQ ID NO:3)。所述检测试剂可以是与 PSPU 标记物部分相互补的寡核苷酸序列，可以被设计成已知位于 PSPU 标记物侧翼的核酸或肽序列或与 PSPU 标记物编码的多肽结合的抗体。所述试剂可以与如上讨论的固相基质结合，或与将它们结合到基质的试剂一起包装。固相基质或底物可以是珠子、板、管(tube)、浸量尺(dip stick)、条(strip)或生物芯片的形式。带可寻址定位并审慎的微滴板的生物芯片或板是特别有用的。

检测试剂包括洗涤试剂以及能够检测结合抗体(如标记的二级抗体)的试剂，或能够与标记抗体反应的试剂。

所述试剂盒将会有利地包括对照试剂(阳性和/或阴性)和/或检测所述核酸或抗体的工具。所述试剂盒内也可包括所用的说明。最通常，所述试剂盒将会被设计成(format)用于本领域公知的分析法，更常用于 PCR、Northern 杂交或 Southern ELISA 分析法，这些是本领域公知的。

也可通过使用本发明核酸分子用于筛选方法如 FISH 来设计试剂盒，所述筛选方法检测与疾病和疾病进展有关的染色体重排。所述试剂盒还可包含针对核酸的检测试剂及对照。

所述试剂盒也可包含一或多种其它前列腺癌的标记物或对照，包括转凝蛋白 1，转凝蛋白 2, PCA 3 及 PSA。在一个实施方案中，在试剂盒中包含所有所述的标记物。

该试剂盒会包含一或多个容器，并也可包含收集容器，例如瓶、袋(如静脉内输液袋(intravenous fluids bag))、小瓶、注射器和试管。至少一个容器盛有可有效用于诊断、监测或治疗 PIN 或 PRC 的组合物。在组合物中的活性剂通常是本发明的化合物、多肽或抗体。在优选实施方案中，在所述容器上或连接的说明书或标签(label)表明所述组合物用于诊断、监测或治疗 PIN 或 PRC。其它组分可包括针(needle)、稀释剂和缓冲液。通常所述试剂盒可包括至少一个容器，该容器包含可药用的缓冲液如磷酸盐缓冲盐水、Ringer's 溶液和右旋糖溶液。

用于分析法和试剂盒的抗体可以是单克隆或多克隆的，并可以在任何哺乳动物中制备。它们优选针对本发明 PSPU 核酸序列编码或标示的天然肽，或基于相同序列的合成肽来制备，或可针对与编码本发明 PSPU 肽的核酸序列融合的外源序列产生。

抗体结合研究可用任何已知分析方法进行，如竞争性结合分析法、非竞争性分析法、直接和间接夹心分析法、荧光免疫分析法、免疫放射性分析法(immunoradiometric assay)、发光分析法(luminescence assay)、化学发光分析法、酶联免疫荧光分析法(ELIFA)及免疫沉淀分析法。Zola, *Monoclonal Antibodies: A Manual of Techniques*, pp.147-158 (CRC Press, Inc., 1987); Harlow and Lome (1998) *Antibodies, A Laboratory Manual*, Cold Spring Harbour Publications, New York; US 5,221,685; US 5,310,687; US 5,480,792; US 5,525,524; US 5,679,526; US 5,824,799; US 5,851,776; US 5,885,527; US 5,922,615; US 5,939,272; US 5,647,124; US 5,985,579; US 6,019,944; US 6,113,855; US 6,143,576; US 5,955,371; US 5,631,171 和 US 2005/0064511。

例如，一类夹心分析法是 ELISA 分析法，此时可检测的部分是酶。ELISA 特别可用于预测、检测或监测 PIN 或 PRC。

可替代的本文可用的分析技术包括质谱分析如表面增强激光解吸离子化(surface-enhanced laser desorption and ionization, SELDI)、电喷射离子化(electrospray ionization, ESI)和基质辅助激光解吸离子化(matrix assisted laser-desorption ionization, MALDI)。

对于免疫组织化学，例如，组织样本可能是新鲜的或冷冻的，或可能包埋在石蜡中，并用防腐剂如甲醛固定。

在一个试剂盒实施方案中，将 PSPU 检测试剂固定在固相基质如有孔

条上来形成至少一个 PSPU 检测位点。有孔条的测量或检测区域可包含多个检测位点,这些检测位点包含 PSPU 检测试剂。所述位点可以棒、十字(cross)或点或其它排列方式来排列。测试条也可包含阴性和/或阳性对照的位点。所述对照位点可任选在不同的条上。不同的检测位点可包含不同量的固定的核酸,如在第一个检测位点的量较大而在随后的位点量较小。在加入被测生物样本后,展示可检测信号的位点数目提供了对在样本中存在的 PSPU 的量的定量指示(indication)。

在另一方面,本发明提供了分析法,其包含与 PSPU 43 的一或多个 PSPU 核酸序列结合的一或多个核酸序列。大范围的正义和反义探针及引物可以从本文 PSPU 的核酸序列来设计。所述 PSPU 序列的表达水平用上述讨论的本领域已知技术来鉴定。所述阵列可以是固相底物如“芯片”,如 US Patent No. 5,744,305 所述,或硝酸纤维素膜。

本文 PSPU 标记物所表达的蛋白质也可用于分析法中,并将结果与在正常样本中表达的相同蛋白质的表达水平相比较。蛋白质的存在及量可用本领域公知并且在本文讨论的分析法模式来评估。

在另一方面,本发明提供了筛选改变本发明核酸分子特别是 PSPU 43 (SEQ ID NO:3)的表达的化合物的方法。在广义术语中,被测化合物与本发明核酸分子(标记物)所编码的肽接触,评估所述肽的生物活性,并选出在所述化合物不存在时改变所述分子生物活性的化合物,或结合所述肽的化合物。在可选的实施方案中,使表达所述分子的被测细胞接触被测化合物,并选出在所述化合物不存在时改变所述标记物表达水平的化合物。这些化合物包括激动或拮抗所述核酸分子表达的分子。

更具体地,设计药物候选物的筛选分析法来鉴定化合物,其与本文鉴定的核酸分子(标记物)所编码的多肽或其生物活性片段结合、优选特异性结合、或复合,或者干扰所编码肽与其它细胞蛋白质的相互作用。这些筛选分析法包括适合高通量筛选化学文库的分析法,使它们特别适合鉴定小分子药物候选物。小分子通常具有低于 500 Daltons 的分子量,包括合成的有机或无机化合物包括肽,优选可溶肽,(多)肽-免疫球蛋白融合物,并特别是抗体,包括但不限于多克隆的和单克隆抗体及抗体片段、单链抗体、抗独特型抗体、和嵌合的或人源化的这些抗体或片段、以及人抗体(human antibodies)和抗体片段。

可从多种已知化合物、从天然来源如植物获得的未知化合物、提取物和微生物获得本发明的被测化合物，或用本领域公知的组合文库的多种方法中的任一种。见例如 Lam Anticancer Drug Des. 12: 1 145 (1997) and DeWitt 等, PNAS 90:6909 (1993)。

可以多种模式进行分析法，包括蛋白质-蛋白质结合分析、生化筛选分析、免疫分析及基于细胞的分析法，它们在本领域都是表征良好的。所有分析法都是常用的，因为它们都需要使药物候选物接触由本文鉴定的 PSPU 核酸分子所编码的肽，接触的条件和时间足以使这两种组分相互作用。

如果候选化合物与本文鉴定的标记物所编码的特定肽相互作用但不结合，其与该肽的相互作用可用公知检测蛋白质-蛋白质相互作用的方法来分析。这些分析法包括传统方法，如交联、共免疫沉淀、及经梯度或层析柱的共纯化。另外，可用基于酵母的遗传系统监测蛋白质-蛋白质相互作用，见例如 Fields 及其合作者所述 [Chevray 等, PNAS 89: 5789-5793 (1991)。Clontech, California, USA 提供了试剂盒 (MATCHMAKER™) 用双-杂交 (two-hybrid) 技术鉴定两种特定蛋白质之间的此种蛋白质-蛋白质相互作用。此系统也可扩展到定位 (map) 特定蛋白质相互作用所涉及的蛋白质结构域，以及精确定位 (pinpoint) 对这些相互作用关键的氨基酸残基。

为测试被测化合物抑制结合的能力，制备反应混合物并在被测化合物不存在和存在时应用 (run)。该反应混合物通常含有本文所述的正 PSU 多肽，被测化合物，及与所述标记物多肽相互作用的组分。也可应用阳性对照。如上监测被测化合物以及与所述标记物多肽相互作用的组分之间的结合 (复合物形成)。在对照反应物中而不是在含被测化合物的反应混合物中形成复合物表明，该被测化合物干扰被测化合物及其反应配偶体 (partner) 的相互作用。

用这些筛选分析法，可鉴定改变 PSU 标记物优选 PSU 43 的活性的化合物。活化 PSU 标记物的功能的化合物是激动剂。类似地，抑制 PSU 标记物功能的化合物是拮抗剂。用本发明筛选方法鉴定的这些化合物也形成部分本发明。

当被检测的生物活性是细胞增殖、不依赖支持物 (anchorage-independent) 的生长、侵入和迁移，对该活性的检测可以例如，通过制备表达一或多种 PSU 肽的细胞，在存在被测化合物时培养所述细胞，并测定细

胞增殖的速度，测量细胞周期、和/或软琼脂中的集落形成活性、改良的 Boyden 侵入分析法和迁移分析法。

由本发明 PSPU 核酸序列所编码的一或多种肽的结合活性或生物活性比该筛选方法所检测的标记物的正常对照水平低，表明该被测化合物是 PSPU 标记物的抑制剂或拮抗剂。相反，所述 PSPU 标记物的结合活性或生物活性比正常对照水平高，表明该被测化合物是该标记物的强化剂(enhancer)或激动剂。

肽、非肽化合物、合成的小分子(micromolecular)化合物和天然化合物可用于本发明的筛选方法。

对本发明核酸分子的激动剂和拮抗剂的计算机制模也可用公知程序如 AUTODOCK (Dunbrack 等, 1997, Folding and Design 2:R27-42) CHARMm 和 QUANTA 程序(Polygen Corporation, Massachusetts, USA)成为可能。

本发明也提供了 PIN 或 PRC 参照表达图谱。这包括了本发明核酸分子，优选 PSPU 43 的标记物表达的类型。通常，所述表达图谱包括一或多种选自 PCA3、转凝蛋白 1、转凝蛋白 2、和 PSA 的其它标记物。在一个实施方案中，所述标记物是 PCA3 和 PSA。在另一个实施方案中，包括了所有其它的标记物。用上述讨论的表达技术，可产生所述图谱并用作在诊断 PIN 或 PRC 或其倾向中比较新患者样本的点。所述图谱也可用于监测治疗 PIN 或 PRC 的过程，并作为鉴定为患有 PIN 或 PRC 的患者的预后工具。

由此，本发明的另一种方面提供了治疗或预防患者的 PIN 或 PRC 的方法，其中过表达了本发明 PSPU 分子。所述方法包括改变所述 PSPU 标记物的表达或其编码的肽的活性。可通过施用用上述筛选方法获得的一或多种化合物来实现抑制。可选地，可用本领域已知方法抑制表达，如施用抑制或拮抗所述标记物表达的核酸。破坏标记物表达的反义寡核苷酸、siRNA、细胞内抗体和核酶都能用于抑制表达。

对应于本文 PSPU 分子、优选 PSPU 43 的反义-寡核苷酸可用于在需要的情况降低所述 PSPU 分子的表达水平。本发明反义-寡核苷酸可通过结合本发明 PSPU 核酸分子编码的多肽起作用，或对应的 DNA 或 mRNA，从而抑制所述标记物的转录或翻译，提高所述 mRNA 的降解，和/或抑制所述 PSPU 核酸分子编码的蛋白质的表达，并最后抑制所述蛋白质的功能。

抑制过表达基因的一或多种基因产物的核酸也包括小干扰 RNAs

(siRNA), 其包含编码所述 PSPU 标记物的核苷酸序列的正义链核酸和反义链核酸的组合。术语“siRNA”指预防靶 mRNA 翻译的双链 RNA 分子。将本发明 siRNA 引入所述细胞的标准技术可用于治疗或预防 PIN 或 PRC, 包括其中 DNA 是转录出 RNA 的模板的那些。可这样构建 siRNA, 从而单个转录物具有来自靶基因的正义和互补的反义序列, 如发夹。

所述方法用于阻抑(suppress)上调表达本发明 PSPU 分子的细胞的基因表达。靶细胞中 siRNA 与 PIN 或 PRC 标记物转录物的结合导致由细胞产生的 PIN 或 PRC 蛋白质下降。所述寡核苷酸的长度为至少 10 个核苷酸 并可能与天然存在的转录物等长。优选, 所述寡核苷酸长度不足 100, 不足 75, 不足 50 或不足 25 个核苷酸。优选, 所述寡核苷酸长度为 19-25 个核苷酸。

siRNA 的核苷酸序列可用 siRNA 设计计算机程序来设计, 所述程序得自 Ambion 的网址(http://www.ambion.com/techlib/misc/siRNA_finder.html)并如 Yuan 等, Nucleic Acids Research 2004 vol 32, W130-W134 所述。用如下方案由计算机程序选择 siRNA 的核苷酸序列:

siRNA 靶位点的选择:

1. 从转录物的 AUG 起始密码开始, 扫描 AA 双核苷酸序列的下游。记录出现的每个 AA 和 3'邻近的 19 个核苷酸作为潜在的 siRNA 靶位点。Harborth 等, (2003)推荐设计针对 5'和 3'末翻译区(UTRs)以及邻近起始密码子的区(在 75 个碱基内)的 siRNA, 所述区可能较富含调控蛋白质结合位点。设计针对这些区域的内切核酸酶和 siRNA 的复合体可能会干扰 UTR-结合蛋白和/或翻译起始复合物的结合。

2. 与人基因组数据库比较潜在的靶位点, 并从考虑中去除与其它编码序列有显著同源性的任何靶序列。可用 BLAST 如上所述进行同源性检索, 其可发现于在www.ncbi.nlm.nih.gov/BLAST/的 NCBI 服务器中。

3. 选择合格的靶序列用于合成。在 Ambion 网址, 可选择多项优选靶序列以及基因的长度作评估。

所述 siRNA 可抑制 PSPU 分子的表达, 由此可用于阻抑所述蛋白质的生物活性。因此, 包含 siRNA 的组合物可用于治疗或预防 PIN 或 PRC, 其中涉及到 PSPU 分子的过表达。

抑制过表达基因的一或多种基因产物的核酸也包括抗所述过表达标记物的核酶。核酶通常是具有以类似于 DNA 限制性内切酶的方式切割其它单

链 RNA 能力的 RNA 分子。

设计和构建核酶的方法是本领域公知的(见例如 Koizumi 等, FEBS Lett. 228: 225 (1998); Kikuchi 等, NAR 19: 6751 (1992)), 并且可基于编码 PIN 或 PRC 蛋白质的核苷酸序列的序列信息, 根据制备核酶的常用方法, 构建抑制过表达 PIN 或 PRC 蛋白质表达的核酶。因此, 包含所述核酶的组合物可用于治疗或预防 PIN 或 PRC。

可选地, 任何过表达基因的一或多种基因产物的功能可通过施用化合物来被抑制, 所述化合物结合或者抑制所述基因产物的功能。例如, 结合过表达标记物的一或多种产物的抗体可用于 PIN/PRC 治疗以及诊断和预后分析法。

本发明也涉及抗体或所述抗体的片段的用途。本文所用的术语“抗体”指具有特定结构的免疫球蛋白分子, 其与包含用于合成所述抗体的抗原的分子或与所述抗体密切相关的抗原特异性相互作用(结合)。抗体特异性结合本发明 PSPU 多肽, 如果它不结合非-PSPU 多肽。通常, 该抗体对 PSPU 抗原会有不超过 10^{-7} M, 优选低于约 10^{-8} M, 优选低于约 10^{-9} M 的结合亲和力(解离常数(Kd)值)。结合亲和力可用表面等离子体共振(surface plasma resonance)来估测。

结合本文 PSPU 标记物多肽的抗体可以是任何形式, 如单克隆或多克隆抗体, 并包括通过用所述多肽免疫接种动物如小鼠、大鼠或兔子获得的抗血清, 所有类型的多克隆的、单克隆的、人的抗体, 以及通过遗传重组制备的人源化的和细胞内抗体。

另外, 用于治疗或预防 PIN 或 PRC 的方法的抗体可以是抗体或经修饰的抗体的片段, 只要它结合本文所述标记基因所编码的一或多种蛋白质。所述片段通常包含抗原结合区和/或其互补决定区。所述抗体片段可以是 Fab、F(ab')₂ 及 Fc 或 Fv 或单链 Fv (scFv), 其中 H 和 L 链的 Fv 片段通过适当接头而连接(Huston 等, Proc. Natl. Acad. Sci. USA 85: 5879-83 (1988))。

制备抗体的方法是本领域公知的(见例如 Antibodies: A Laboratory Manual, CSH press, eds, Harlow and Lane (1988))。最常用的抗体是通过免疫接种上述讨论的合适宿主动物制备的。与 PSPU 蛋白质的融合蛋白也可用作免疫原。

可通过与多种分子(如聚乙二醇(PEG))偶联来修饰抗体。经修饰的抗体

可通过化学修饰抗体来获得。这些修饰方法是本领域常用的。

任选地，获得的抗体可以是源自非人抗体的可变区与源自人抗体的恒定区的嵌合抗体，或是人源化抗体，后者包含源自非人抗体的互补决定区(CDR)、源自人抗体的框架区(FR)以及恒定区。这些抗体可用本领域已知方法制备。

简言之，制备多克隆抗体的方法是本领域技术人员已知的。可在哺乳动物产生多克隆抗体，例如通过一或多次注射免疫剂，以及佐剂(如果需要的话)。通常，可在哺乳动物内通过多次皮下或腹腔内注射来注射免疫剂和/或佐剂。所述免疫剂可包括 PSPU 多肽或其融合蛋白。它可用于将免疫剂与被免疫的哺乳动物内已知为免疫原性的蛋白质偶联。这些免疫原性的蛋白的例子包括但不限于匙孔血蓝蛋白(keyhole limpet hemocyanin)、血清白蛋白、牛甲状腺球蛋白、及大豆胰蛋白酶抑制剂。可用的佐剂的例子包括弗氏完全佐剂和 MPL TDM 佐剂(单磷酸脂类(moriophosphoryl Lipid) A、合成的海藻糖二霉菌酸酯(dicorynomycolate))。可由本领域熟练技术人员不需不当试验就可选择免疫方案。

细胞内抗体通常是本文的单链抗体，它们会包含特异性结合 PSPU 多肽的单链抗体。它们可用于基因疗法，所述疗法通过将编码所述抗体的序列掺入重组载体并施用到过表达 PSPU 多肽的细胞来结合并抑制其功能。制备这些抗体的方法是本领域已知的。(见例如 Tanaka 等, *Nucleic Acids Research* 31(5):e23 (2003)。

可用本领域也公知的杂交瘤方法制备单克隆抗体。见例如 Kohler and Milstein, *Nature*, 256:495 (1975)。可在适合培养基中培养杂交瘤细胞，任选地，所述杂交瘤细胞可作为哺乳动物内的腹水体内存养。优选的无限增殖细胞系是鼠骨髓瘤细胞系，如 MPC-11 和 MOPC-21，它们可得自例如来自美国典型培养物保藏中心(American Type Culture Collection), Virginia, USA。可用免疫分析法筛选分泌目的抗体的无限增殖的细胞系。由本文 PSPU 标记物或其变体或片段编码的多肽可用于筛选。

由此，本文也包括杂交瘤，其为能分泌 PSPU 肽特异性单克隆抗体的无限增殖的细胞系。

建立由杂交瘤细胞产生的单克隆抗体的结合特异性的公知手段包括免疫沉淀、放射免疫分析法(radio-linked immunoassay) (RIA)、酶联免疫吸附分

析法(ELISA)及 Western 印迹。(Lutz 等, Exp. Cell. Res. 175:109-124 (1988))。可类似筛选来自免疫动物的抗病毒是否存在多克隆抗体。

为了利于检测,本文的抗体和片段可用允许直接测定抗体结合的可检测标记物,如荧光的、生物发光的、和化学发光的化合物,以及放射性同位素、磁珠、和亲和标记(如生物素和亲和素)来标记。允许间接测定结合的标记物的例子包括酶,其中底物可提供有色荧光产物,适合的酶包括辣根过氧化物酶、碱性磷酸酶、苹果酸脱氢酶等。荧光色素(Fluorochrome)(如德克萨斯红(Texas Red)、荧光素(fluorescein)、藻胆蛋白(phycobiliprotein)、和藻红蛋白(phycoerythrin))可与荧光激活细胞分选仪(fluorescence activated cell sorter)应用。标记技术是本领域公知的。

可以通过常用的免疫球蛋白纯化方法如,例如蛋白质 A-Sepharose、羧基磷灰石层析、凝胶电泳、透析、或亲和层析将所述细胞分泌的单克隆抗体从培养基或腹水分离或纯化。

单克隆抗体或片段也可通过重组 DNA 工具制备(见例如 U.S. Patent No. 4,816,567)。DNA 修饰如用人的重链和轻链恒定区的编码序列替代同源的鼠序列(U.S. Patent No. 4,816,567; 见上文)也是可行的。本文也包括制备嵌合的二价抗体。

所述抗体可以是单价抗体。制备单价抗体的方法是本领域公知的。

本发明的抗-PSPU 抗体还可包含人源化抗体或人抗体。这些人源化抗体优选作治疗用。人源化抗体包括人免疫球蛋白,其中受者互补决定区(CDR)的残基被非人物种 CDR 的残基替代。从非人来源如兔子、大鼠和小鼠制备人源化抗体是公知的。(Verhoeyen 等, Science, 239:1534-1536 (1988); Jones 等, Nature, 321:522-525 (1986); Riechmann 等, Nature 332:323-329 (1988)。

人抗体也可用本领域已知的多项技术,包括噬菌体展示技术(Hoogenboom and Winter, J. Mol. Biol. 227:381 (1991)); 及转基因方法,见例如 Nature Biotechnology 14, 826 (1996); 和 Vaughan *et al*, Nature Biotechnology 16:535-539 (1998)来制备。

双特异性抗体是单克隆的,优选人的或人源化的抗体,其对至少两种不同抗原具有结合特异性。本文包括双特异性抗体,其中一种结合特异性是针对所述 PSPU 标记物,另一种是针对任何其它抗原,并优选针对细胞表面蛋白或受体或受体亚基(subunit)。

制备双特异性抗体的方法是本领域已知的。见例如 Milstein 和 Cuello, *Nature*, 305:537-539 (1983)和 Suresh 等, *Method in Enzymology*, 121:210 (1986), Brennan 等, *Science* 229:81 (1985)。

双特异性抗体可结合本文给定 PSPU 多肽上的两个不同表位。任选, 它们可结合抗-PSPU 及表位, 所述表位结合在表达 PSPU 的细胞中参与细胞防御的一或多个分子。例如, 白细胞 T-细胞受体分子, 和针对 IgG 的 Fc 受体。还任选, 双特异性抗体可包含结合细胞毒性剂的表位, 所述细胞毒性剂如蓖麻毒蛋白(ricin)A 链、皂草素(saporin)、或氮甲蝶呤或放射性核素螯合剂, 如 EOTUBE 或 DOTA。

本文也包括具有超过两种特异性的抗体, 如三特异性抗体(trispecific antibodies)。

本文也包括由两条共价连接的抗体组成的异偶联(Heteroconjugate)抗体。已经提示这些抗体用于将免疫系统细胞靶向不需要的细胞(US Patent No. 4,676,980)。所述抗体可用本领域已知的交联技术体外产生。

可提高所述抗体的有效性。例如, 通过向 Fe 区引入一或多个半胱氨酸残基, 从而允许在此区域形成链间二硫键来产生同二聚体(Homodimeric)抗体。同二聚体抗体可用本领域已知的交联剂产生, 如 Wolff 等, *Cancer Research*, 53: 2560-2565 (1993)所述。

抗独特型(Antiidiotypic)抗体也可用于本文讨论的疗法, 来诱导对表达 PSPU 蛋白质的细胞的免疫应答。制备这些抗体也是公知的(见例如 Wagner 等, *Hybridoma* 16:33-40 (1997))。

本发明抗体可固定在固相支持物上: 合适的支持物包括对于核酸序列上述讨论的那些。将抗体与固相支持物结合可通过本领域公知的技术实现。见例如 *Handbook of Experimental Immunology*, 4th Edition, Blackwell Scientific Publications, Oxford (1986)。结合的抗体用于本文讨论的分析法。

本发明提供了在有需要的患者中治疗或预防 PIN 或 PRC 的方法, 其使用抗 PSPU 多肽的抗体。根据所述方法, 给所述患者施用药物有效量的抗 PIN 或 PRC 多肽的抗体。以足以降低 PIN 或 PRC 多肽活性的剂量施用, 在 PIN 或 PRC 中涉及本发明 PSPU 分子特别是 PSPU43 的过表达。任选地, 与对肿瘤细胞特异的细胞表面标记物结合的抗体可用作药物递送的工具。因此, 例如, 与细胞毒性剂(如美登醇(maytansinoid)、氟尿嘧啶、紫杉醇(taxol)、

蓖麻毒蛋白 A 链、相思豆毒蛋白 A 链、白喉毒素、柔红霉素(doxorubicin)、氮甲蝶呤、依诺霉素(neomycin)、白树毒素(gelonin)、放射性核素如 ^{186}Re , ^{212}Bi , p^{32} , I^{125} 和 ^{131}I 偶联的抗 PSPU 多肽的抗体, 可以足以损伤或杀灭肿瘤细胞的剂量施用。治疗方法可涉及施用一或多种抗体。用于制备在这些方法中使用的免疫偶联物的方法例如见 Vitetta 等, Science, 238: 1098 (1987) 所述。

本发明也涉及治疗或预防患者的 PIN 或 PRC 的方法, 该方法是通过施用改变本发明 PSPU 多肽的表达或活性的化合物。在过表达时, 施用降低所述 PSPU 多肽的表达或活性的化合物。所述化合物或组合物可以是疫苗, 其包含本发明 PSPU 多肽或所述多肽的免疫活性片段, 或编码所述多肽或其片段的多核苷酸。施用所述多肽可诱导受试者内的抗肿瘤免疫。所述多肽或其免疫活性片段也可用作抗 PIN 或 PRC 的疫苗。可施用包含本文的一或多种 PSPU 多肽的疫苗, 如同施用多种包含单一 PSPU 多肽的疫苗。在受试者内经诱导抗肿瘤免疫可治疗或预防良性肿瘤。有时候, 可以施用所述蛋白质或其片段的形式是结合到 T 细胞受体(TCR)或递呈到抗原递呈细胞(APC), 特别是树突细胞(DC)上。

在本发明中, 术语 PIN 或 PRC 疫苗指经接种诱导抗肿瘤免疫或经免疫系统起作用来阻抑 PSPU 的物质。通常, 抗肿瘤免疫包括免疫应答、诱导抗肿瘤的细胞毒性淋巴细胞, 诱导识别肿瘤的抗体, 并诱导产生抗肿瘤的细胞因子。

可通过观察宿主动物内抗所述蛋白质的免疫系统应答来检测诱导抗肿瘤免疫。检测应答的系统是本领域公知的。

诱导针对肿瘤细胞的细胞毒性 T 淋巴细胞的多肽, 如同所诱导的细胞毒性 T 淋巴细胞, 可用于抗 PIN 或 PRC 的疫苗。能诱导针对 PIN 或 PRC 的细胞毒性 T 淋巴细胞的抗原递呈细胞也用于针对 PIN 或 PRC 的疫苗。用结构不同的蛋白质/肽的组合可提高对细胞毒性 T 淋巴细胞的诱导。这些组合可包括用于本文讨论的免疫疗法。

也可通过测定抗肿瘤抗体的产生来评估多肽的抗肿瘤免疫。如果生长, 肿瘤细胞的增殖或转移就被抗体阻抑, 用于制备所述抗体的多肽明显具有诱导抗肿瘤免疫的能力。

施用本发明疫苗, 从而允许通过诱导抗肿瘤免疫来治疗和/或预防 PIN 或 PRC。治疗性和预防性治疗 PIN 或 PRC 可包括任何对肿瘤细胞生长的抑

制,以及对出现肿瘤细胞的阻抑,PIN或PRC标记物在血中水平的改变,陪同PIN或PRC的可检测症状减轻,并降低患者死亡率。这些治疗性和预防性作用优选是统计学显著性的。例如,与对照相比的显著水平在5%或更高,优选10%或更高。

在配制本发明疫苗时,具有免疫活性的多肽或编码所述多肽的多核苷酸或载体可与佐剂组合。佐剂指在与具有免疫活性的蛋白质一起(或连续(successively))施用增强抗该蛋白质的免疫应答的化合物。佐剂的例子包括但不限于霍乱菌(cholera)毒素、沙门菌(salmonella)毒素、及明矾(alum)。本发明疫苗可与可药用载体如无菌水、生理盐水及磷酸盐缓冲液组合。另外,如果必要,所述疫苗可含稳定剂、混悬剂(suspension)、防腐剂(preservative)、表面活性剂等。所述疫苗可在全身或局部单一或多重施用。在用作免疫原时,所述多肽可与载体如KLH、BSA或其它本领域公知蛋白质偶联。其它治疗组合物见下讨论。

本发明也提供了治疗或预防受试者的PIN或PRC的方法,通过施用改变本发明PSPU核酸分子或PSPU多肽的表达或生物活性的化合物。

在此方法的一个实施方案中,所述治疗化合物包括表达不足的标记物的多肽产物,或其生物活性片段,编码位于表达控制元件下游的表达不足的基因的核酸,所述表达控制元件允许所述基因在PIN或PRC细胞中表达,提高在PIN或PRC细胞中内源性存在的标记物的表达水平的化合物。可用本文所述的筛选方法获得这些化合物。为向细胞递送丢失的基因或蛋白质,可用逆转录病毒系统。这些系统是本领域公知的。见例如US 5,082,670及DNA cloning: A Practical Approach, Volume 3, DRL Press, Washington (1987)中的“Retroviral Vector”。如上讨论,将载体掺入细胞可通过技术如显微注射、转染、转导和电穿孔及其它(Sambrook等,见上文)。可用基因疗法来抑制不当的过表达或增强PSPU 43分子或多肽的表达。

本发明也提供了治疗或预防PIN或PRC的组合物,其包含药物有效量的:本发明方法鉴定的化合物;或结合本发明PSPU多肽的抗体或其片段。所述组合物可包括两或多种这些化合物、抗体和多肽或其组合。所述组合物中也包括可药用的载体、赋形剂或稀释剂。也提供了药物组合物,其包含有效量的至少一个PSPU反义序列、siRNA、核酶或多肽,以及可药用的载体、赋形剂或稀释剂。

含本发明化合物、反义序列、siRNA、核酶、多肽或抗体的药物组合物，可以通过将有效量的活性分子与可选的可药用载体、赋形剂或稳定剂混合来制备(*Remington's Pharmaceutical Sciences* 16th edition, Osol, A. Ed. [1980])。

本文所用的有效量指任何本文公开的活性物质(active)包括多肽、抗体、小分子、siRNA、反义序列、核酶、激动剂或拮抗剂为足以实现所述目的的量。熟练工人可根据经验用常规方法确定所述量。类似地，“药物有效量”或“治疗有效量”指本文公开的活性物质的量，其可有效预防或治疗上述 PIN 或 PRC (见“治疗”的定义)。

所述组合物可配制为了口服施用(如胶囊、片剂、锭剂(lozenge)、粉剂、糖浆剂等)、肠胃外施用(如静脉内溶液、皮下、肌肉或栓剂配制剂)、局部施用(如乳膏、凝胶)、吸入(如鼻内、肺内)或其它这些本领域公知的施用形式。

可接受的载体、赋形剂或稳定剂是本领域公知的。它们在所用剂量和浓度必须对受者无害，并包括缓冲液(如磷酸和柠檬酸(citric acid))、水、油、特别是橄榄油、芝麻油、椰子油、矿物油和植物油；碳水化合物，包括乳糖、葡萄糖、甘露糖或糊精；螯合剂如 EDTA；糖，如蔗糖、甘露醇、海藻糖或山梨醇；成盐反离子(counter-ion)如钠；金属复合物(如 Zn-蛋白复合物)；非离子型表面活性剂如 TWEENTM，或聚乙二醇(PEG))。

对于片剂，稀释剂如碳酸盐(如钠盐和钙盐)、磷酸盐(如磷酸钙)或乳糖，常与抗氧化剂、成粒剂(granulating)和崩解剂(disintegrating agent)(如玉米淀粉)、粘合剂(binding agent)如淀粉、及润滑剂如硬脂酸和硬脂酸镁共用。可包被片剂来摄入(ingestion)、稳定性或崩解(disintegration)。

可注射的组合物通常用润湿剂(wetting agent) (如聚氧乙烯硬脂酸酯(polyoxyethylene stearate)、卵磷脂和聚氧乙烯失水山梨醇单油酸酯(polyoxyethylene sorbitol monooleate))、助悬剂(suspending agent) (如甲基纤维素、海藻酸钠(sodium alginate)和西黄蓍胶(gum tragacanth))以及稀释剂制备。

如果必要，所述组合物也可包括添加剂如着色剂、抗氧化剂(如抗坏血酸)、甜味剂(sweetener)、增稠剂(thickening agent) (如石蜡、蜂蜡)、调味剂(flavouring agent)和防腐剂(如烷基对羟基苯甲酸酯(alkyl parabens)、苯酚、间苯二酚和苯扎氯铵(benzalkonium chloride))。

可用任何常用技术制备片剂、局部和静脉内配制剂、糖浆剂、水包油乳化剂、吸入剂等(Remington's 见上文)。

也可用脂质体将活性物质递送到细胞。当使用抗体片段时,优选特异性结合靶蛋白结合区的最小抑制片段。可如上讨论或如本领域公知来通过化学合成重组制备肽。参见例如 PNAS USA 90,7889-7893 (1993)。

治疗组合物也可包含一或多种其它活性剂。其它所选的活性物质不应对上述讨论的主要活性剂有明显副作用。其它活性剂的例子有细胞毒性剂、细胞因子、化疗剂如 Taxol®以及顺铂、和/或生长抑制剂。所述活性物质以治疗有效量组合存在。所述活性物质可配制为治疗组合物的部分,或分开来与所述治疗组合物同时或顺序(sequential)使用。

所述活性剂也可配制为微胶囊(microcapsule)或水性混悬液(aqueous suspension)例如,与助悬剂如海藻酸钠、甲基纤维素类(如甲基纤维素,羧甲基纤维素,羟丙甲基(hydroxylpropylmethyl)纤维素)或大分子乳液(macroemulsion)。这些技术公开在 *Remington's Pharmaceutical Sciences* 16th edition, Osol, A. Ed. (1980)中。

要用于体内施用的组合物必须是无菌的。这可通过用无菌过滤膜或其他已知本领域技术过滤来容易地实现。

可制备缓释(Sustained-release)制剂。缓释制剂的合适例子如上述讨论的微胶囊。缓释基质的例子包括聚酯及水凝胶。

在免疫辅助(immunoadjuvant)疗法中,施用本发明的蛋白质、抗体或化合物可与化疗、化学雄激素去势疗法、或放疗,或与分开(separate)、同时或顺序施用其它抗癌剂联用。药剂的制备和按剂量给药(dosing)的方案可根据生产商的说明书使用或由熟练操作人员确定。化疗剂的制备和按剂量给药方案参见例如 *Chemotherapy Service* Ed., M.C. Perry, Williams & Wilkins, Baltimore, MD (1992)中。为了治疗或降低 PIN 或 PRC 严重程度或其症状,本发明活性物质的合适剂量会依赖于患者的年龄、所治疗疾病的类型和严重程度(无论该药剂(agent)是为了预防性还是治疗性目的施用)、之前的或其它共同时存在的疗法、施用途径、患者病史、以及根据药物的公知原理对活性物质的反应。所述化合物可以给患者只施用一次、或持续地或重复地施用。例如每日、每周、每月、每日若干次,并且施用可以是规律的、间歇的或以隔开的(spaced)间隔。

根据疾病的种类及严重程度, 约 1 $\mu\text{g/kg}$ 到 15 mg/kg (如 0.005 到 20 mg/kg , 优选 0.1 到 1 mg/kg)的包括本发明的化合物、组合物、核酸分子、多肽或抗体的本发明活性物质是给患者的起始候选剂量, 以单一或分开的剂量或通过持续输注施用。通常的日剂量可能从约 1 $\mu\text{g/kg}$ 到 100 mg/kg , 更为 1 mg -75 mg/kg , 或更依赖于上述强调的因素。可进行治疗, 直到已经消除(abate)了疾病或其症状, 或直至做出了终止的决定。可用上面讨论的分析法或其它常用监测技术来监测所述治疗方案。

本说明书参照了专利说明书、其他外部文件(external document)、或其它信息来源, 这常常是为了提供讨论本发明特征的上下文的目的。除非特别额外指出, 参照这些外部文件不应视为认可(admission)这些文件; 或者这些信息来源, 在任何法律视角中(in any jurisdiction), 都是现有技术, 或形成本领域的一部分公知常识。

现在以非限制性方式参照如下实施例来阐述本发明:

实施例 1

鉴定第 8 号染色体, 前列腺-富集序列: *Pspu1*, *Pspu2*, *Pspu8* 和 *Pspu43* 引言

我们开发了自动数据采集(datamining), 我们称为数据清查(data-panning)的系统。此系统始于从组织样本捕捉转录物(mRNAs), 将它们转变为 cDNA 文库形式的稳定产物(cDNAs)。cDNA 文库经仔细测序, 结果在 GenBank 中保藏了大量表达序列标签(EST)。这些表达序列是 UniGene 数据库中 EST 的来源(Schuler 等, 1996)。目前在人 UniGene 数据库中有约 410 万的人 EST/mRNA。

UniGene 将从 GenBank 引入的 EST 隔成基因定位簇(gene-oriented cluster)的非冗余组(non-redundant set), 而每个 UniGene 簇名义上含有代表单个基因所得转录物的序列(Schuler 等, 1996)。UniGene 的关键特征是给 EST 指定 dbEST 文库 ID。因 dbEST 文库 ID 识别用于生成 dbEST 文库的组织, 故此 ID 是 UniGene 簇中该 EST 的组织来源在计算机中唯一的标记。

dbEST 文库源自广泛多种器官和组织。如果这些文库从整体上代表该机体, 则单个文库的转录组(transcriptome)聚集起来将反映整体的转录组。同时, 单个文库将反映转录体组中可归因于器官和组织的区域性(regional)

差别。含有高比例的来自单一组织的 EST 的 UniGene 簇可在整体 UniGene 背景中被识别出来。

我们用此计算机方法鉴定了前列腺组织中的富集基因表达图谱 (enriched gene expression profile)。这无需事先知晓这些基因的功能或最可能的分布。用我们的方法鉴定出位于第 8 号染色体上的四个转录物, 详述于此。我们称这些转录物为前列腺特异性单基因(unigene)1(*Pspu1*)、前列腺特异性单基因 2 (*Pspu2*)、前列腺特异性单基因 8 (*Pspu8*)和前列腺特异性单基因 43 (*Pspu43*)。EST 详情如下:

表 1: 构成每个第 8 号染色体候选物的 EST 序列列表

标记物	GenBank 登录号	全长 EST 序列
<i>Pspu1</i>	AA635472 AI659328 AI659339 BX112742 AW014583	
<i>Pspu2</i>	CB050448 BX113278 AI420913 AI927409 AW293795	
<i>Pspu8</i>	BX283231 CF139278 BM043676 BU543602	
<i>Pspu43</i>	AI611685 AI418055 BF446403	ctttcttttttttttgctctatctccagatcttcccttcta gccaaactcctttgcacccaaaaagcagcctttgctttc ttgagatgaaagaacattcatgaaaatcatccctctact ggagtcctctagcaattcctgtgatttccacttacctga ctatgtacacaagcccagataacctggccttagtgtgggga cagagcagagtgaccaagagtccagacctagagcctgct tgcctgggttcaaatctcatctctaccactcagtaaact ctgtcccaactttcctcatctgaaaaatgggcataacaat agtcccttatctacagg (SEQ ID NO: 28) ttttttttttttttgctctatctccagatcttcccttcta gccaaactcctttgcacccaaaaagcagcctttgctttc ttgagatgaaagaacattcatgaaaatcatccctctact ggagtcctctagcaattcctgtgatttccacttacctga ctatgtacacaagcccagataacctggccttagtgtgggga cagagcagagtgaccaagagtccagacctagagcctgct tgcctgggttcaaatctcatctctaccactcagtaaact ctgtcccaactttcctcatctgaaaaatgggcataacaat (SEQ ID NO: 29) ttttttttttttttgctctatctccagatcttcccttct agccaaactcctttgcacccaaaaagcagcctttgcttt cttgagatgaaagaacattcatgaaaatcatccctctac

		tggagtcctctagcaattcctgtgatttccacttacctg actatgtacacaagcccagataacctggccttagtggtggg acagagcaaagtgaccaagagtccaaacctagagcctgc ttgcctgggttcaaattctcatctctaccactcagtaaac tctgtcccactttcctcatctgaaaaatgggcataacaa tagtcccttatctcaca (SEQ ID NO: 30)
	BF222603	ctttcttttttttggctctatctccagatcttccttcta gccaaactcctttgcacccaaaaagcagcctttgctttc ttgagatgaaagaacattcatgaaaatcatccctctact ggagtcctctagcaattcctgtgatttccacttacctga ctatgtacacaagcccagataacctggccttagtggtggga cagagcagagtgaccaagagtccagacctagagcctgct tgcctgggttcaaattctcatctctaccactcagtaaac ctgtcccactttcctcatctgaaaaatgggcataacaat agtcccttatctcacaggtttttagtaaaattaaatgag ttaatttaattttctaagcact (SEQ ID NO: 31)
	BX109457	tttaattaacaaatataaagtaccagacactccaagtgt tagaaaaattaaattaactcatttaattttactaaaaacc tgtgagataagggaactattgttatgcccattttcagat gaggaagtgggacagagtttactgagtggtagagatga gatttgaaccaggcaagcaggctctaggctctggactct tggtcactctgtctgtgtcccccacactaagccaggatct gggcttgtgtacatagtcaggtaagtggaaatcacagga attgcttagaggactccagtagagggatgattttcatgaa tgttctttcatctcaagaaagcaaaggctgctttttggg tgcaaaggagtttggctagaaggaagatctggagataga gcaaaaaaaaaagaaagaaaaaaaaaaaaaaaaaaaa (SEQ ID NO: 32)
	AA658380	ttttttttttggctctatctccagatcttccttctagcca aactcctttgcacccaaaaagcagcctttgctttcttga gatgaaagaacattcatggaaatcatccctctactggag tccctctagcaattcctgtgatttccacttacctgactat gtacacaagcccagataacctggccttagtggtggggacaga gcagagtgaccaagagtccagacctagagcctgcttggc tgggttcaaattctcatctctaccactcagtaaacctctgt cccactttcctcatctgggtcgac (SEQ ID NO: 33)

系统和方法

记录在每个 UniGene 簇中的数据提供了针对下述组织或细胞类型产生富集基因表达图谱的方法，所述组织或细胞已测序 cDNA 文库并分配了 dbEST 文库 ID。经 UniGene 算法归类的每组 EST 都分配了一个 UniGene 编号。在 UniGene 数据库中，凭此编号进入所述簇，有生成所述簇的基因的任何已知信息，该基因的任何 STS 标记物，可能的蛋白质相似性，染色体定位等。在 UniGene 记录内称为‘scount’的领域显示出形成 UniGene 簇的序列数。每个 UniGene 记录的最后部分列出了所有形成该簇的序列。此列表

的格式包括每个序列的登录号, 如果该 EST 作为 cDNA 文库的一部分测序的话, 还包括 dbEST 文库 ID 号。所述 dbEST ID 号起标记给定序列的生物来源的作用。

从此信息产生基因表达图谱依赖大量随机测序的 cDNA 文库以及 dbEST 文库编号系统。如果一个基因只被一种组织表达, 那么只有从那种组织构建的文库具有来自那个基因的代表性序列。通过测定每个 UniGene 簇的 dbEST 文库 ID, 在所有 EST 都源自只从一种组织构建的文库时才显示出组织特异性基因表达。

大多数基因不是完全对于一种组织特异的, 而是显示分布在一个范围的组织内。在随机测序的 cDNA 文库中, 高丰度表达的基因将会比那些低水平表达的基因测序得更频繁。这会通过 UniGene 反映出来。用来自组织的 dbEST 文库 ID 给序列做标签, 比起以低水平表达所述基因的组织, 在该组织中所述基因更常高表达。这意味着序列用来自 UniGene 簇内的目标组织的 dbEST ID 号作标记的次数可表明该组织内基因表达的水平。可用占构成(contributed to)该 UniGene 簇的 EST 总数的百分比来表示。我们已经用此方法来获得对前列腺特异的生物标记物。

算法

从 <ftp://ftp.ncbi.nlm.nih.gov/repository/UniGene> 下载 UniGene 数据文件(Hs.data.gz)。用利用 Bioperl 工具试剂盒(toolkit)的三个 Perl 脚本(script)(Stajich 等, 2002)编辑这些文件。特定地, 这些脚本引用了(call)Bio::Cluster::Unigene 和 Bio::ClusterIO 模块(module)。第一个脚本“lib_extract”自动浏览每个 UniGene 簇中的 EST 序列行数, 并舍弃构成的(contributing)序列数超过特定阈水平的那些 UniGene。

通过限定变量“\$threshold”等于 4 来设定此阈水平, 并将它与每个 UniGene 簇的 scount 相比。Scount 是构成簇的序列总数。舍弃任何构成该簇的序列数等于或低于阈值的记录。将高于序列行数阈值的那些带入(parse into)内部(in-house)数据库做随后的数据漫游(见下)。

从原始 Hs.data 文件带入 Scount, 并且包括那些无区分 dbEST 文库 ID 号的行。因为随后的计算是基于 dbEST 文库 ID, 舍弃了显示出具有低于规定的(stipulated)阈值数的 EST 的一些簇。一旦出现这种情况, 较低值会反映

出 UniGene 簇中有 dbEST 文库 ID 的 EST 数。

从 NIH 下载的 UniGene 文件大(对于人为 400Mb)。创造一系列编辑的内部数据库极有利于随后的计算, 所述数据库仅保留了 UniGene 簇号且为每个构成的 EST 保留了 dbEST ID。此方法利用“lib_extract”脚本, 将人的数据文件降到 40Mb。

在 UniGene 的网址有带一些描述符(descriptor)的文库目录(Hs.lib.info.gz)。在 UniGene Library Browser (<http://ncbi.nlm.nih.gov/UniGene/>) 中有关于文库构建的更多细节。所有 UniGene 文库都有 dbEST 文库 ID。

用“lib_percentage”脚本进行数据漫游。这需要研究者指定的一组 UniGene 文库, 然后查询(interrogate)所编辑的内部数据库。这些内部数据库不再含有序列数低于阈值的 UniGene 簇。所述脚本测定了源自指定组的文库的每个 UniGene 簇内的 EST 序列行数。这些表示为占 UniGene 簇内的 EST 行总数的百分比。

执行

从网址 <http://ftp.ncbi.nih.gov/repository/dbEST> 下载 Human dbEST 文库列表。在程序 BBedit 中打开该列表, 鉴定了前列腺文库, 并用 GREP 功能制备编辑的列表。鉴定 290 个文库是从前列腺组织构建的。用于此分析来鉴定人前列腺特异性序列的 dbEST 文库为 689, 787, 792, 876, 910, 924, 925, 926, 928, 934, 935, 940, 994, 1016, 1017, 1053, 1054, 1055, 1333, 1410, 1498, 1654, 1655, 1668, 1670, 1671, 1672, 1673, 4267, 4268, 4711, 4712, 4713, 4714, 4715, 4716, 4717, 4718, 4719, 4720, 6013, 6014, 6015, 6016, 6017, 6018, 6019, 6020, 6021, 6022, 6023, 6024, 6025, 6026, 6027, 6028, 6029, 6030, 6031, 6032, 6033, 6034, 6035, 6036, 6037, 6038, 6039, 6040, 6041, 6042, 6043, 6044, 6045, 6046, 6047, 6048, 6049, 6050, 6051, 6052, 6053, 6054, 6055, 6056, 6057, 6058, 6059, 6060, 6061, 6062, 6063, 6064, 6065, 6066, 6067, 6068, 6069, 6070, 6071, 6072, 6073, 6074, 6075, 6076, 6077, 6078, 6079, 6080, 6081, 6082, 6083, 6084, 6085, 6086, 6087, 6088, 6089, 6090, 6091, 6092, 6093, 6094, 6095, 6096, 6097, 6098, 6099, 6100, 6101, 6102, 6103, 6104, 6105, 6106, 6107, 6308, 6309, 6310, 6311, 6312, 6313, 6314, 6315, 6316, 6317, 6318, 6319, 6320, 6321, 6322, 6323, 6324, 6325, 6326, 6327, 6328, 6329, 6330, 6331, 6332, 6333, 6334, 6335, 6336,

6337, 6338, 6339, 6340, 6341, 6342, 6343, 6344, 6345, 6346, 6347, 6348, 6349, 6350, 6351, 6352, 6353, 6354, 6355, 6356, 6357, 6358, 6359, 6360, 6361, 6362, 6363, 6364, 6365, 6366, 6367, 6368, 6369, 6370, 6371, 6372, 6373, 6374, 6375, 6376, 6377, 6378, 6379, 6380, 6381, 6382, 6383, 6384, 6385, 6386, 6387, 6388, 6389, 6390, 6391, 6392, 6393, 6394, 6395, 6396, 6397, 6398, 6399, 6400, 6401, 6402, 6601, 6602, 6603, 6763, 6831, 7180, 7181, 8480, 8481, 8482, 8483, 8484, 8485, 8486, 8487, 8488, 8489, 8490, 8491, 8492, 8493, 8494, 8495, 8496, 8497, 8498, 8499, 8500, 8501, 8502, 8503, 8504, 8505, 8506, 8507, 8508, 8509, 8510, 8511, 8512, 8513, 8514, 8515, 8585, 8834, 9134, 9135, 9136, 9137, 9138, 10161, 10549, 11034, 11037, 14129, 14130, 14131。这些文库都是从正常或患病的前列腺材料构建的。在 2004 年 3 月 4 日构建的人 UniGene 上进行计算。将结果输入 Microsoft Excel 作分选。

结果和讨论

几项研究提示从第 8 号染色体短臂(8p)失去基因序列是早期分子事件 (Cher 等, 1994; Macoska 等, 1994; 1995; 2000; Haggman 等, 1997), 它出现在几乎所有前列腺癌中, 并显著出现在前列腺癌最可能的前兆前列腺上皮内瘤(PIN)中(Bostwick, 1996)。用数据漫游算法鉴定为 100%、75%和 80%富集水平的 3 个转录物(下表 2)位于 8p。两个, 在硅片上(in silico)显示与在正常和患病组织之间失去表达相一致的类型。另一个对前列腺表达有 83%富集的转录物位于 8q。我们称这些转录物为前列腺特异性单基因 1(Pspu1)、前列腺特异性单基因 2(Pspu2)、前列腺特异性单基因 8(Pspu8)和前列腺特异性单基因 43(Pspu43)。之前没有描述过这些转录物。Pspu1 位于 8p12, 发现 Pspu2 在 8p21, Pspu8 位于 8p22-23 且 Pspu43 位于 8q23。

表 2: 对 Pspu1, Pspu2, Pspu8 和 Pspu43 的富集结果

UniGene 号	<i>Pspu</i> ID 号	前列腺文库的 EST	其它文库的 EST	UniGene 簇中的 EST 总数	百分比 (%)
197095	Pspu1	4	1	5	80
444680	Pspu2	3	1	4	75
458397	Pspu8	4	0	4	100
161160	Pspu43	5	1	6	83

对于 *Pspu1* 和 *Pspu2* 的硅片上基因表达图谱用 Stanton and Green(2001)所述的荟萃分析(meta-analysis)系统测定。简言之,从 NCBI (<http://ncbi.nlm.nih.gov>)下载测序的前列腺 cDNA 文库。将每个文库置于由制备它的组织所决定的目录。这意味着从 PIN 组织制备的所有文库被分组到一起,而来自正常组织的所有 cDNA 文库形成另一组。每个文库中的 EST 基于 UniGene 而聚类。这样得到落入每个目录的转录单元列表。通过记录在每个目录中给定 UniGene 的 EST 数,产生硅片上的基因表达图谱,其表示每个组织类型所表达的特定转录物的水平。此数据用于产生下表 3。*Pspu8* 和 *Pspu43* 不包括在内,因为产生它们的文库由于它们的构建性质被排除在荟萃分析之外。

表 3: 正常前列腺上皮细胞和前列腺癌进展阶段中的基因的表达图谱例子。给正常前列腺作正态化丰度的评分,而患病组织中的水平为正常表达的百分比。在括号中给出的显著性水平用拟合度(goodness of fit)的卡方检验(Chi squared test)测定为 2 类,ns=无显著性差异

描述	正常前列腺	前列腺上皮内瘤	侵入性癌	转移的损伤
前列腺特异性抗原	1580	188 % (0.001)	103 % (ns)	50 % (0.001)
前列腺酸性磷酸酶	790	169 % (0.001)	83 % (ns)	33 % (0.001)
前列腺特异性单基因 1	20	0 % (0.001)	0 % (0.001)	0 % (0.001)
前列腺特异性单基因 2	10	0 % (0.001)	0 % (0.001)	0 % (0.001)

已知若干基因在前列腺癌进展中的表达改变,对这些基因作的比较与此荟萃分析一致。例如,数字表达图谱显示前列腺特异性抗原(PSA)在前列腺上皮内瘤的表达上升与普遍认为的一致(表 3)。另外,荟萃分析显示前列腺酸性磷酸酶下调(表 3)。前列腺酸性磷酸酶被用于在 PSA 试验前诊断前列腺癌(Bostwick, 1996),现在认为它与前列腺肿瘤的失去雄激素反应性有关(Meng 等, 2000)。

从独立文库多次取样 *Pspu1*, *Pspu2*, *Pspu8* 和 *Pspu43* 的序列,由此确定它们代表了合法转录物。5 个 EST 构成了 UniGene 簇 Hs.197095,我们称此序列毗连序列群为 *Pspu1*。这些来自 3 个 cDNA 文库 NCI_CGAP_Pr28 (dbEST 文库 ID.1410)、NCI_CGAP_Pr22(dbEST 文库 ID.910)和 NCI_CGAP_Sub2

(dbEST 文库 ID.2359)。文库 1410 和 910 是从正常前列腺组织构建的。文库 2359 是从为了鉴定乳腺特异性基因而设立的减差(subtracted)cDNA 文库产生(Bonaldo 等, 1996)。

4 个 EST 构成了 UniGene Hs.444680 或 *Pspu2*。在 cDNA 文库 NCI_CGAP_Pr28(dbEST 文库 ID.1410)、NCI_CGAP_Pr22 (dbEST 文库 ID.910) 和 NCI_CGAP_Sub4(dbEST 文库 ID.2721)中鉴定了这些 EST。这些文库中的 2 个是从正常前列腺制备的(1410 和 910)。第三个文库是另一个为发现前列腺特异性基因而设立的减差文库(2721) (Bonaldo 等, 1996)。执行我们的数据漫游算法不是为了鉴定减差文库 2721 是从前列腺构建的, 并因此给予此簇的富集水平只有 75%。事实上此 UniGene 可代表仅限于前列腺的转录物。

Pspu8 由 4 条 EST 序列组成, 它们取自包含 UniGene Hs.458397 的 3 个克隆。这些克隆自 2 个文库分离, 它们两个都是从前列腺癌细胞系构建的。这些文库是 dbEST 文库 14129 和文库 8834。

6 个 EST 组成 UniGene Hs.161160。从这些 EST 序列形成的毗连序列群被称为 *Pspu43*。这些 EST 来自在 2 个 cDNA 文库中发现的 5 个克隆。这些文库是 NCI_CGAP_Pr28(dbEST 文库 ID.1410)和 NCI_CGAP_Pr2 (dbEST 文库 ID. 574)。此数据漫游算法指出此转录物在前列腺中只有 83%富集。然而, 文库 574 没有掺入前列腺特异性文库的列表, 所以来自此文库的 EST 没有被标记为前列腺源性。像 *Pspu2* 一样, 此转录物可能仅仅是前列腺源性的。

对组成 *Pspu1*, *Pspu2*, *Pspu8* 和 *Pspu43* 的 EST 做比对, 以得到对每个候选物最佳的共有序列。图 1 中给出共有序列, 下表 4 给出相对在 NCBI 的非冗余的 GenBank 数据库获得的 BLASTN (Altschul 等, 1990)结果。

表 4: 对 *Pspu1*, *Pspu2*, *Pspu8* 和 *Pspu43* 前列腺特异性候选物的 BLASTN 分析汇总

毗连序列群序列	与 GenBank 候选物比对的毗连序列群碱基对	最佳 GenBank 比对的登录号	意见
Pspu1	1-353	AC044849.12	基因组 DNA
Pspu2	1-379	AC090786.6	基因组 DNA
Pspu8	1-1320	NM_0540281	cDNA

Pspu43	1-392	AP001207/AP000426	基因组 DNA
--------	-------	-------------------	---------

Pspu1 和 *Pspu2* 共有序列长度分别为 368bp 和 394bp，然而它们都在事实上长度不同的 polyA 序列片段(stretch) (分别为 18bp 和 15bp)终止。两端序列定位在人基因组上，但不与已知的表达基因产生高得分匹配。所述 *Pspu43* 共有序列长为 392bp，并也定位在人基因组上但不是表达的序列。*Pspu8* 长为 1320bp 并定位在人酰基-丙二酰基(Acyl-malonyl)缩合酶的表达序列上。

结论

我们已经用 UniGene 数据采集算法鉴定了 4 条前列腺特异性核酸序列 *Pspu1*, *Pspu2*, *Pspu8* 和 *Pspu43*。 *Pspu1*, *Pspu2* 和 *Pspu8* 分别定位在 8p12/21 边缘、8p21 和 8p22-p23。已知从这些位点缺失是前列腺癌的早期事件 (MacGrogan 等, 1994)。在疾病中失去这些序列中的两条得到正常前列腺和前列腺癌的基因表达荟萃分析的支持。*Pspu43* 定位在第 8 号染色体长臂的 8q23。最近，邻近区域 8q24 已经与前列腺癌易感性遗传连锁(Amundadottir 等, 2006)。我们设想 *Pspu1*, *Pspu2*, *Pspu8* 和 *Pspu43* 会是作为发病前列腺癌的早期事件而出现的第 8 号染色体改变的有用标记物。

实施例 2: **Pspu43: 表征第 8 号染色体长臂上的前列腺疾病标记物。**

引言

通过用上面实施例 1 所述的数据漫游算法对 cDNA 组织文库作数据采集，鉴定出 *Pspu43* 对前列腺有重要性。用数据漫游方法，总共鉴定了定位在第 8 号染色体上的 4 条序列。3 条位于第 8 号染色体的短臂上，而 *Pspu43* 位于该染色体的长臂上。*Pspu43* 紧邻第 8 号染色体上在前列腺肿瘤中常改变的区域，并已经与前列腺癌易感性遗传连锁(Amundadottir 等, 2006)。

此实施例汇总了我们对 *Pspu43* 的发现，包括确认基因组序列、细胞培养系统中的表达图谱以及组织特异性数据。

系统和方法

PCR 引物设计

从由 EST 簇 Hs.161160 产生的毗连序列群设计对 *Pspu43* 的 PCR 引物。

该毗连序列群经 BLASTN (Altschul 等, 1990)分析确认该毗连序列群未掺入克隆载体的序列。将此经编辑的序列带入(load into)PCR 引物设计程序 (Primer 3, Rozen and Skaletsky, 2000)。选出产生最长扩增子的最适引物对, 并在 NCBI 用 BLASTN 与非冗余的基因序列数据库进行比较。用 AMPLIFY 程序(William Engles, Genetics Department, University of Wisconsin)进行毗连序列群和引物序列的模拟 PCR, 来确保匹配的忠诚度, 避免引物二聚体形成并测试可能的引物二级结构。通过 Invitrogen (USA)合成引物。表 5 给出引物序列。

表 5: PCR 引物序列

<i>Pspu43</i>	正向	5'-AACAAATATAAAGTACCAGACACTCCA -3'
	反向	5'-ATCTCCAGATCTTCCTTCTAGCC -3'
转凝蛋白 2	正向	5'- CTTCCAGAACTGGCTCAAGG-3'
	反向	5'- GAGAAGAGCCCATCATCTCG-3'
PSA	正向	5'- CACTGCATCAGGAACAAAAGCGT-3'
	反向	5'-CATCACCTGGCCTGAGGAATC-3'

提取并扩增 RNA 和基因组 DNA

正常前列腺上皮和基质细胞(PrEC 和 PrSC; Clonetics, San Diego CA)生长并保持在专用培养基(PrEGM Bullet Kit; Clonetics, San Diego CA)中, 而前列腺癌细胞系 LNCaP (ATCC CRL 1740, Manassas, VA)、DU145 (ATCC HTB-81, Manassas, VA)和 RWPE-2 (ATCC CRL-11610, Manassas, VA)生长并保持在补充了 10%胎牛血清的 RPMI 1640 培养基中。

RWPE-2 是人乳头瘤(papillary)病毒无限增殖的前列腺上皮细胞系 (RWPE-1)被 v-Ki-ras 转化的衍生物。它是雄激素反应性的、侵入性的 (invasive)且致肿瘤的(tumorigenic) (Bello 等, 1997; Webber 等, 1997a)。LNCaP 是转移的前列腺癌损伤的衍生物, 其对雄激素是反应性的(Webber 等, 1997b)。DU-145 衍生自转移的前列腺癌损伤, 它是雄激素无反应性的, 高度侵入性的且致肿瘤的(Webber 等, 1997b)。

分别以 2500 细胞.cm⁻²或 4000 细胞.cm⁻²的密度接种 PrEC 和 LNCaP 细胞, 并在 5% CO₂ 的潮湿环境于 37°C 在 25 cm² 排气式锥形瓶(vented flask)中培养到 70%汇合。通过胰蛋白酶消化收获细胞, 在无胰蛋白酶的培养基

中洗涤,并在 500 g 浓集。用 TriZol (Invitrogen, Carlsbad, USA) 根据生产商手册从细胞粒(pellet)提取基因组 DNA(gDNA)和 RNA。用 Superscript II (Invitrogen, Carlsbad, USA)根据生产商说明书将 RNA 转变为 cDNA。在 160ng 基因组 DNA 或 37.5ng cDNA 上用上述引物进行 PCR 扩增。用 Amplitaq Gold™ master mix (Applied Biosystems, NJ, USA)进行 PCR。使 PCR 条件最适并确立了 65°C 的有效退火温度。比较来自所有前列腺细胞系的样本。

来自一系列组织的 RNA 购自 Clontech laboratories Inc. (Mountain View, CA, USA)或得自其它研究计划的捐助。这些样本源自乳腺、卵巢、睾丸、肾和血。这些被如上转化为 cDNA,并在 PCR 分析法中以 37.5ng RNA 当量的浓度使用。

测序 PCR 产物

用 QIAGEN PCR 纯化系统(QIAGEN GmbH, Hilden, Germany)对 PCR 产物进行凝胶纯化。从 1%琼脂糖凝胶用 QX1 缓冲液根据生产商说明书提取 DNA。从纯化柱在无菌 milliQ 水中洗脱 DNA。

使用 BigDye Terminator v3.1 chemistry, 用正向和反向 PCR 引物(10μM)一起测序纯化的 PCR 扩增子。

结果

表 6 显示了从 3 个前列腺细胞系的基因组 DNA 和合成 cDNA 获得的 PCR 结果。这些是 LNCaP、PrEC 和 PrSC 细胞系。PrEC 和 PrSC 分别源自正常前列腺上皮和基质细胞。LNCaP 源自患前列腺癌的患者淋巴结转移损伤。在从所有 3 个细胞系分离的基因组 DNA 中检测到 *Pspu43* 序列。这表明 *Pspu43* 是人基因组的一部分,并完全不从 LNCaP 细胞系中丢失。

表 6: 从 PrEC、PrSC 和 LNCaP 细胞系分离的基因组和 cDNA 的 PCR 分析法结果汇总。

	PrEC	PrSC	LNCaP
基因组 DNA	+	+	+
cDNA	+	-	+

+ = 预期 PCR 产物/- = 无 PCR 产物

用 RT-PCR 从自每个前列腺细胞系合成的 cDNA 模板获得基因表达结果。所有 5 个细胞系的结果汇总在表 7 中。在至少 2 个模板上重复 RT-PCR 最少 3 次。在 5 个细胞系中的 4 个中表达 *Pspu43*。也就是说,它存在于 PrEC、LNCaP、DU145 和 RWPE2 样本上,而不是在 PrSC 样本上。这里包括了转凝蛋白 2 和 PSA 的 RT-PCR 结果作比较。

表 7: 细胞系的 RT-PCR 结果

细胞系	PSPU 43	转凝蛋白 2	PSA
PrSC	-	+	+
PrEC	+	+	-
LNCaP	+	+	+
DU145	+	+	+
RWPE2	+	+	+

+ = PCR 试验阳性, - = PCR 试验阴性

用 RT-PCR 在一些分离自卵巢、肾、乳腺、睾丸和血的不同 RNA 样本上检测对于 *Pspu43* 的组织特异性。在乳腺和肾而不是其它被测组织检测 *Pspu43* 序列(表 8)。

表 8: 分离自 5 种组织的 RNA 的 PCR 分析法结果汇总。

卵巢	睾丸	乳腺	肾	血
-	-	+	+	-

+ = 预期 PCR 产物 / - = 无 PCR 产物

从首先鉴定 Hs.161160 为前列腺生物学目的的 UniGene 簇开始,它已经 142 个 EST、7 条 mRNA 序列一起分组,并称为 TFCEP2L3(Grainyhead-样 2(Drosophila))。组成毗连序列群 *Pspu43* 的原来 6 个 EST 仍位于此 UniGene 簇中。然而,毗连序列群 *Pspu43* 并不定位在现在与 TFCEP2L3 有关的任一条 mRNA 序列上。这用毗连序列群 *Pspu43* 和所有 7 个 mRNA 序列的 2-向 (way)BLAST 来显示。所述 *Pspu43* 毗连序列群与非冗余序列数据库的 BLAST 比对显示,它仅与 2 个基因组克隆 AP000426 和 AP001207 完全匹配。这些克隆是未经注释的大 DNA 序列,它们分别为 239,116 和 153,936 个核酸。

TFCP2L3 (Grainyhead-样 2 (*Drosophila*)) EST 在前列腺是高度代表性的 (表达图谱产生单位(expression profiler) – NCBI), 并通过 Northern Blotting 显示(Peters 等, 2002)。然而, 显示 TFCP2L3 没掺入 *Pspu43* 序列。TFCP2L3 是已经与导致听觉丧失的突变有关的转录因子(Peters 等, 2002)。

结论

Pspu43 是表达的 RNA 序列, 据鉴定它在从正常和癌性前列腺组织制备的 cDNA 文库中专一存在。它可以在前列腺上皮细胞中正常表达。因此, *Pspu43* 表达可能是前列腺健康的标记物。

实施例 3: 尿分析

引言

我们希望检测 *Pspu43* 在经过了前列腺疾病临床检查的男性的尿中是否可检测。从 Department of Urology, Dunedin Hospital, Dunedin, New Zealand 收集尿样本。从这些尿样本提取 RNA, 并进行 RT-PCR 来检测 *Pspu43*、转凝蛋白 2 和 PSA。给这些分析法计分。只有在获得 RT-PCR 结果之后才可诊断患者。

方法

此研究中的所有参与者都出具了书面同意, 且该项目得到了 Lower South Regional Ethics Committee (“Development of non-invasive, Diagnosis and prognostic tests of Prostate cancer” LRS/05/05/016)的伦理学认可。男性接受了作为常规检查方法的一部分的前列腺操作, 来确定他们前列腺的物理状态。前列腺操作包括手指触诊(digital palpation)右和左叶及通过用食指从顶端到底部在每侧抚过 3 次。之后, 收集 20-30 ml 的尿样本。将等体积的磷酸盐缓冲液(pH7.0)加入尿样本。将此样本于 4°C 储存过夜。通过在 2500g 于 4°C 离心 15 分钟来收获细胞, 弃上清, 并在 800μl TriZol (Invitrogen, Carlsbad, USA)中重悬细胞粒。加入糖原(Invitrogen, Carlsbad, USA)得到 250μg/ml 的终浓度。根据生产商方案提取 RNA。在 16.5μl H₂O 中重悬 RNA 粒。用 Dnase I (Roche, Switzerland)根据生产商说明书处理 8μl 所述样本。用 Superscript II (Invitrogen, Carlsbad, USA)根据生产商说明书将一半用 Dnase I

处理的样本转变为 cDNA。在 1 到 2.5 μ l 的 cDNA 和等体积的 Dnase I 处理的 RNA 间中用下述引物(表 9)进行 PCR 扩增。用 Amplitaq Gold™ master mix (Applied Biosystems, NJ, USA)进行 PCR。使 PCR 条件最适并确立 65°C 的有效退火温度。

表 9: *Pspu43*、转凝蛋白 2 和 PSA 的 PCR 引物

<i>Pspu43</i>	正向	5'-AACAAATATAAAGTACCAGACACTCCA-3'
	反向	5'-ATCTCCAGATCTTCCTTCTAGCC-3'
转凝蛋白 2	正向	5'- CTTCCAGAACTGGCTCAAGG-3'
	反向	5'- GAGAAGAGCCCATCATCTCG-3'
PSA	正向	5'- CACTGCATCAGGAACAAAAGCGT-3'
	反向	5'-CATCACCTGGCCTGAGGAATC-3'

结果和讨论

我们从 8 个男性提供的尿样本获得了可靠 RT-PCR 结果, 这些男性经过了前列腺检查排除疑似疾病, 以及来自 1 个男性的尿样本, 该男性无前列腺疾病症状。如果仅有 RNA 的样本不产生 PCR 产物, 那么认为 PCR 结果是可靠的。我们的 PCR 系统中所用的酶不能用 RNA 作为模板。因此, 从仅有 RNA 的反应物产生的 PCR 产物表明样本中存在基因组 DNA。此时, 不可能将基因表达与基因组污染区分。这些结果汇总在表 10 中。

表 10: 9 个男性的尿样本

样本 ID	PSPU 43	转凝蛋白 2	PSA	Gleason 级别
U "S"	-	+	+	正常
U1	-	+	+	无肿瘤
U3			+	7
U6	+	+	+	6
U7	+	+	+	无肿瘤
U12	-	+(+RT-)	+	7
U14	+	+	+	良性前列腺增生
U20	-	+	+	良性前列腺增生
U22	+	+	+	7

+ = PCR 试验阳性, - = PCR 试验阴性

对于无路径的细胞无可靠结果。

这些试验使用非定量的 RT-PCR 分析法, 并且这些患者无长期随访数据。样本 U “S”来自无明显疾病的男性。没有试图表征这些尿样本中的细胞群。它证明从尿样本提取始终一致的高品质 RNA 是有挑战性的, 并且获得的 RNA 量是变化的。结果, 很多样本不适合源于建立所述技术的试验变量。

从这些结果清楚知道 *Pspu43*、转凝蛋白 2 和 PSA 的转录物在尿中可检出。在 2 个已知癌症患者(U6 & U22)、1 个无肿瘤者(U7)及 1 个良性者(U14)中检出 *Pspu43*。然而, 我们没有这些患者的长期随访数据。如果 *Pspu43* 是早期疾病进展指标, 患者 U7 和 U14 可能已被误诊并确定在疾病的很早阶段。在正常的(U “S”)、无肿瘤(U1)或良性(U20)样本中不能检出 *Pspu43*, 并且在 1 个癌症患者(U12)中不能检出它。

在所有样本检出转凝蛋白 2, 尽管疑似 U12 的读数为也从非 RT 对照产生的产物。这也是不对 *Pspu43* 产生阳性反应的唯一已知肿瘤样本。试图从此样本扩增 2 个不相关的产物 3 次, 每次都得到不一致的结果(数据未示)。仅试了 PSA 分析法一次。对从此样本产生的结果最有利的解释是, 该 RT-PCR 反应不可靠, 这是由于从此样本分离的 RNA 品质低或浓度低。

不管疾病状态如何, 在所有样本中都检出 PSA。这会表明 PSA 是否存在于尿样本中不能是诊断性的, 因为无论前列腺健康或个人的病理如何, 都会检出它。

结论

Pspu43 可以在男性的尿中检出。它在随后从前列腺活检诊断出患有前列腺肿瘤的患者中比在无肿瘤或有良性前列腺增生(BPH)的男性中更常检出。在所有样本中检出 PSA 和转凝蛋白 2。因此, PSA 或 Transgelin 2 均不合适确认存在或不存在标记物的简单诊断试验。从另一方面, *Pspu43* 能够从患肿瘤的患者检出。这支持了 *Pspu43* 作为前列腺癌标记物的用途。PSA 升高作为前列腺癌进展的预测物的问题在于, 它不能将前列腺癌与其它病理区分。BPH 和前列腺炎都可提高血 PSA 得分。因 *Pspu43* 可区分 BPH 及前列腺癌, 它具有潜在作为癌存在标记物的较高灵敏性。

实施例 4: 前列腺肿瘤组织中的 *Pspu43* 表达

引言

此试验的目的是检查在患病状态的前列腺中的 Pspu43 表达改变。

此研究中使用来自单个个体的 10 个匹配对的正常的和损伤的活检。这些样本收集自经过对前列腺腺癌的前列腺切除术的男性, 且都是 Gleason 级别 6 及以上的。从这些样本提取 RNA, 并进行定量 PCR(qPCR)处理, 来确定 Pspu43 是否在前列腺中表达, 并且比较肿瘤对比正常组织的相对表达水平。匹配的活检确保不同个体之间的潜在遗传变异不会混淆结果。我们使用转凝蛋白 1, 显示出它在其它癌症中下调(Chang 等, 2001, Shields 等, 2002), 作为与 Pspu43 表达的比较。

方法

组织

从 Cancer Society Tissue Bank (Christchurch, New Zealand)获得组织活检。从向组织银行捐献材料的所有患者获得书面同意, 并从 Cancer Society Tissue Bank Governing Board and The Lower South Regional Ethics Committee (“Development of non-invasive, Diagnosis and prognostic tests of Prostate cancer” LRS/05/05/016)获得对此项目的特别认可。以冷冻组织块(block)提供组织, 在从患者取出后, 所述组织块已经在 15-30 分钟内于液氮中速冻。表 11 和 12 中给出了组织和相应的患者细节。所有肿瘤都是前列腺腺癌组织型的, 并且都显示出神经周侵入。

表 11: 组织和患者细节

ID	年龄	N 或 T	Gleason 级别	肿瘤最大尺寸 mm	肿瘤的% vol	淋巴/血管侵入	坏死	结节中 Mets *
5 F6	60	正常				无		
5 F8	60	肿瘤	6		25	无		
9 B3	70	正常						
9 B4	70	肿瘤	6		30			
9 F5	70	正常						
9 F6	70	肿瘤	6		25	有		
14 B9	71	正常						1/5
14 C1	71	肿瘤	9		85	有		1/5
14 D1	49	正常						
14 D2	49	肿瘤	7		80	+/-		
37 i9	67	正常					无	0/12
38 A1	67	肿瘤	7		20	无	无	0/12
38 A3	63	正常		20			有	0/5
38 A4	63	肿瘤	8	20	10	有	有	0/5
38 F1	66	正常					无	0/14
38 F2	66	肿瘤	7		30	无	无	0/14
42 B1	63	正常		20			无	0/11
42 B2	63	肿瘤	7	20		无	无	0/11

43 C3	62	正常		25			无	0/2
43 C4	62	肿瘤	7	25	20	+/-	无	0/2

*淋巴结转移

表 12: 病理学详情

ID	病理学详情
5 F6/5 F8	产自左叶的高中度(well-moderately)分化的前列腺腺癌，其邻近从组织银行取的新鲜材料。 Gleason 级别 2+4 (得分=6)
9 B3/9 B4	1 级水平囊状(capsular)侵入，边缘阴性。该肿瘤涉及约 30%的腺体体积，并涉及在右侧有尿道周分布的左叶和右叶。
9 F5/9 F6	前列腺腺癌，Gleason 得分 6，2 级水平囊状侵入，涉及精囊。肿瘤涉及 25%的腺体体积。存在淋巴血管侵入
14 B9/14 C1	在切除边缘存在
14 D1/14 D2	2 级水平囊状】散布(spread)。双叶大部分被肿瘤浸润
37 i9/38 A1	无描述
38 A3/38 A4	不受限的(non-confined)3 级水平，在病灶(focal)
38 F1/38 F2	受限的 2 级水平
42 B1/42 B2	2 级水平，受限的
43 C3/43 C4	不受限的，确立 3 级水平。癌是部分乳头瘤性的。面积提示但不诊断血管侵入。PSA=21

RNA 分离

将所述前列腺组织块的每一块封固(mount)在 OCT(Lab Tek products, Tennessee, USA)的 Cryomould 中。然后切片组织来制备 RNA。所取的第一份和最后一份切片由 8μm 的切片组成，它们被封固在玻片上。此玻片被储存在-80℃，并用作组织学参照，如果需要的话。在第一份和最后一份切片之间，切割 10 片 60μm 切片，并置于半加仑瓶(pottle)中。

将 4μl 的 TriZol (Invitrogen, Carlsbad, USA)加入半加仑瓶中，并立即将所述切片匀质化 30 秒。将匀浆物(homogenate)转移到 15 ml 离心管(falcon tube)中，然后加入 1.5ml 氯仿。将匀浆物涡旋震荡 15 秒，并立即置于冰上。然后将这些匀浆物在 4000rpm 于 4℃ 离心 15 分钟。将顶层转移到干净的管中，再用 1 ml 氯仿重提取，在 4000rpm 于 4℃ 重复离心 15 分钟。再收集顶层并转移到新的管中。在剧烈涡旋震荡所述样本时加入 0.53 体积的 100%乙醇。然后将全部核酸/乙醇混合物转移到与多头抽真空装置(vacuum manifold)连接的 Rneasy 柱(Qiagen, Germany)中，并应用真空装置。然后用 700μl RW1 洗涤缓冲液洗涤带有结合核酸的该柱，之后再次应用真空装置。然后在真空从柱中抽出 500μl 的洗涤缓冲液 RPE。拆下该柱，然后通过 8000rpm 离心 15-30 秒来干燥该柱。将该柱置于新的 1.5ml 离心管中，向该柱加入 30μl 水，然后在 8000rpm 再离心 15-30 秒。将 30μl 流出(flow through)体积再加

回柱顶端,然后在 8000rpm 再离心该单位 15-30 秒。此柱洗脱物含有从每个组织块获得的每组 10 片 60 μ m 切片所提取的 RNA。通过用 Nanodrop 分光计测定光密度及 260/280 比率, 并用 Experion Bio-analyzer 芯片系统 (Bio-Rad, California, USA)电泳,来测试回收 RNA 的品质。在-80 $^{\circ}$ C 储存 RNA 备用。

cDNA 合成

将 1 μ l 的 5 μ g/ μ l Random Hexamers (Roche, Switzerland)加入 600ng 的 RNA。将此混合物加热到 95 $^{\circ}$ C 5 分钟, 之后在 25 $^{\circ}$ C 5 分钟。然后将样本转移到冰上。将 4 μ l First 链缓冲液(Invitrogen, Carlsbad, USA)、每种 2.5mM 的 4 μ l dNTP、2 μ l 0.1M DTT、0.5 μ l 40U/ μ l Rnase 抑制剂(Invitrogen, Carlsbad, USA)及 1 μ l 200U/ μ l Superscript II (Invitrogen, Carlsbad, USA)的分层混合物 (cocktail)加入所述 RNA,并通过吸排(pipetting)混合。在 42 $^{\circ}$ C 温育 120 分钟, 之后在 70 $^{\circ}$ C 温育 10 分钟, 再在 80 $^{\circ}$ C 温育 1 分钟。

用 Qiagen PCR cleanup 柱(Qiagen, Germany)清洗所述 cDNA。将 80 μ l 水和 500 μ l 的 PB 缓冲液加入 20 μ l 的 cDNA 合成反应物(reaction)。通过 Qiagen cleanup 柱在 12000rpm 离心它 1 分钟。弃流出物, 并通过在 12000rpm 离心 750 μ l PE 缓冲液 1 分钟来洗涤含有结合 cDNA 的柱。然后通过 12000rpm 离心 1 分钟来干燥该柱。将该柱转移到新的洗脱物收集器, 并加入 40 μ l 水。通过在 8000rpm 离心 1 分钟来从该柱洗脱 cDNA。将第二份 40 μ l 的等份体积(aliquot)加入所述柱, 并重复离心步骤。将干净的洗脱的 cDNA 样本储存在-80 $^{\circ}$ C 备用。

qPCR

将 cDNA 稀释到 7.5ng/ μ l 的浓度, 并将 2 $\times$ 5 μ l 等份体积的每种样本吸量到 96 opti-孔板的孔中。也将 RNA 样本稀释到 7.5ng/ μ l, 并将 1 份 5 μ l 等份体积吸量到 96 孔板的孔中。做无模板对照来检测 PCR 污染, 并将复制的标准曲线 cDNA 加入每个 96 孔板。向每个样本加入合适的探针/引物混合物, 或加入 SYBR green (Applied Biosystems, Foster City, CA, USA)和 qPCR master mix (Applied Biosystems, Foster City, CA, USA)。密封、混合然后简单离心所述板, 来确保内容物已收集到每个孔的底部。在 ABI7000 机器

(Applied Biosystems, Foster City, CA, USA)上进行 qPCR。

表 13 中给出了引物和探针组。用两个系统来测量 RNA 转录物的水平。在通过解离分析检测单一产物时, 用 SYBR green 作为非特异性的中间螯合的(inter-chelating)染料来检测 DNA 扩增。在解离分析中存在多条带时, 用双标记的 Taqman 探针来提供扩增子区分。根据生产商的方案以 96 孔模式使用每个系统。用来自 Applied Biosystems, California, USA 的 SDS 软件包(1.2.3 版)分析结果。

表 13: 用于 qPCR 的引物和探针

靶	正向	探针	反向
<i>Pspu43</i>	5'GGCTCTAGGTC TGGACTCTTGCT3'	(5'FAM)TGCTCTGTCCCCACA CTAAGCCAGG(3'DABCYL)	5'CCTGACTATGTA CACAAGCCCAGAT3'
<i>Pspu8</i>	5'GGCTGGGCCTG CTTCTGT3'	(5'FAM)CTCAACGTCCTCAGC ATTGGATGTGC(3'DABCYL)	5'GCGGAGCATACG GTGGAA3'
<i>Pspu2</i>	5'CCCTGTATGAA ATACTAAGAGGAG TCCTT3'	(5'FAM)CGGACATGAAAGGA CACTAGACAAATCCACA (3'DABCYL)	5'CTATCGTTTATA TTGCCTATGTAG TTACTTCAC3'
<i>Pspu1</i>	5'TGGCTGTTACC TGCTCTTTCAC3'	(5'FAM)AGCTATCTTGCCACTG CAGACTCAGCAGT(3'DABCYL)	5'CAGGAGGGCTGA GGTACTGTGT3'
转凝蛋白 1 (T1)	5'AAGAATGATG GGCACTACCG 3'	SYBR Green	5'ACTGATGATCTG CCGAGGTC3'
转凝蛋白 2 (T2)	5'CTTCCAGAACT GGCTCAAGG3'	SYBR Green	5'GAGAAGAGCCC ATCATCTCG3'

结果和讨论

从每对匹配的来自前列腺的肿瘤和正常样本提取 RNA, 所述前列腺来自进行了前列腺切除术的个体。从每份提取物获得的 RNA 的平均量为 586ng/μl, 260/280 比率在 1.77 到 2 之间。

最初, 用来自 PC3 前列腺癌细胞系(ATCC CRL 1435, Virginia, USA)的 cDNA 使针对每个引物/探针组合的 qPCR 反应物达到最适。如解离峰分析测定, 用 SYBR green 技术的分析法对 T1 和 T2 合适, 但不适合任何 Pspu 转录物分析法(图 2)。因此, 要为 Pspu 候选物设计与 TaqMan qPCR 系统一起使用的引物/探针对。

用绝对定量方法来测定每份样本中的转录物量。从通用标准的多个稳定细胞系产生的标准曲线, 对每个引物/探针对显示了 0.9998 到 0.9932 的 R² 值(图 3)。图 4 给出了对每个样本对的未经处理的总 CT 值。计算一式两份 qPCR 反应物的 CT 的标准偏差, 且图 4 对每份样本给出了 +/-1 标准偏差。然后从平均 CT 计算转录物 cDNA 的相对浓度, 并从标准曲线计算对应的

cDNA 值。然后校正此值的基因组 DNA 污染,这是通过从仅有 RNA 的 qPCR 反应物测定相对 cDNA 浓度,并从用转录 cDNA 的 qPCR 获得的值减去该值。图 5 给出了来自每个组织对的每个标记物的经校正的相对 cDNA 定量。表 14 给出了数据汇总,以及每个患者的肿瘤的 Gleason 级别。

表 14: 前列腺肿瘤样本: 来自 10 个个体的正常和患病组织的匹配成对样本(qPCR)

患者	T1	T2	Pspu1	Pspu2	Pspu8	Pspu43	Gleason
5 F6/5 F8	<	< ns	<	<	< ns	<	9
9 B3/9 B4	<	<	>	< ns	ns	< ns	7
9 F5/9 F6	<	> ns	< ns	> ns	> ns	>	7
14 B9/14 C1	<	>	>	<	>	>	8
14 D1/14 D2	<	< ns	< ns	>	<	>	7
37 i9/38 A1	<	<	<	>	<	>	7
38 A3/38 A4	<	<	> ns	>	> ns	>	7
38 F1/38 F2	> ns	> ns	> ns	> ns	>	> ns	6
42 B1/42 B2	<	< ns	>	>	> ns	>	6
43 C3/43 C4	<	> ns	< ns	< ns	<	< ns	6
得分	1/9	4/6	5/5	6/4	5/5	7/3	

< =肿瘤相对正常的下降

> =肿瘤相对正常的上升

ns =正常和肿瘤的无统计学显著性的差异

得分=肿瘤上升/肿瘤下降。

从这些结果显示,所有标记物在正常和肿瘤组织中都作为 RNA 转录物来表达。通常,较少的 T1 转录物存在于肿瘤组织中,而 Pspu43 转录物增加(在肿瘤比正常组织中 Pspu43 升高的显著相关性($P=0.0055$)为 $\times 3$ 列联表(contingency table),费希氏(Fisher's)精确检验)。T1 丢失与在其它癌症的发现一致(Chang 等, 2001, Shields 等, 2002),并与细胞支架完整性的丧失和转移相对应。

Pspu43 在 Gleason 级别在 6 到 8 之间的所有肿瘤相对正常样本上调,其中所有样本取自每份样本之间的差异超过标准误差的每个个体。在最严重的损伤(14B9/C1, Gleason 9)中,在前列腺的正常部分中有相对较多的

Pspu43 标记物。令人质疑的是，反映前列腺作用正常的此‘正常的’样本给出了此特定器官(85%涉及该腺体)内疾病的范围，且这可代表该疾病的晚期性质。很可能 Gleason 级别 9 的肿瘤的转录体特征显著不同于该疾病不太严重的形式。

在标记物 Pspu1、Pspu2、Pspu8 或 T2 的表达中没有见到公开一致的类型。因此，尽管我们的生物信息算法鉴定了 Pspu 标记物 1, 2 和 8，我们不能证明它们能够在此试验中区分肿瘤和正常的前列腺。

结论

此试验显示，无论疾病状态如何，转凝蛋白 1 和 2 以及 Pspu 1、2、8 和 43 都表达在前列腺组织中。标记物 T2、Pspu 1、Pspu2 或 Pspu8 没有显示公开的癌症分化表达类型。然而，T1 显示在肿瘤组织中下调，这与他人对于肺癌、乳腺癌和结肠癌的发现相一致(Chang 等，2001, Shields 等，2002)。反过来，Pspu43 显示在肿瘤组织中比正常样本上调。这是显著的。我们知道第 8 号染色体的此区域在很多男性的早期疾病过程中发生改变。这些结果表明 Pspu43 升高会提示患有前列腺癌。

参考文献

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* 215:403-410

Amundadottir LT, Sulem P, Gudmundsson J et al. (2006) A common variant associated with prostate cancer in European and African populations. *Nat. Genet.* (doi:10.1038/ng1808)

Beheshti B, Park PC, Sweet JM, Trachtenberg J, Jewett MA, Squire JA (2001) Evidence of chromosomal instability in prostate cancer determined by spectral karyotyping (SKY) and interphase fish analysis. *Neoplasia* 3: 62-69

Bello et al. (1997). Androgen responsive adult human prostatic epithelial cell lines immortalised by human papillomavirus 18. *Carcinogenesis.*, 18, 1215-1223.

Bonaldo, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery. *Genome Research* 6, 791-806.

Bostwick DG (1996). Prospective origins of prostate carcinoma. *Cancer* 78: 330-336.

Chang JW, Jeon HB, Lee JH, Yoo JS, Chin JS, Kim JH, Yoo YJ (2001). Augmented expression of peroxidoxin I in lung cancer /*Biochem Biophys Res Com*/ 289: 507-512

Cher ML, MacGrogan D, Bookstein R, Brown JA, Jenkins RB, Jensen RH (1994). Comparative genomic hybridization, allelic imbalance, and fluorescence in situ hybridization on chromosome 8 in prostate cancer. *Genes, Chromosomes & Cancer* 11: 153-162.

Fehm et al (2002). Cytogenetic evidence that circulating epithelial cells in Patients with Carcinomas are malignant. *Clinical Cancer Research* 8 : 2073-2084.

Häggman MJ, Wojno KJ, Pearsall CP, Macoska JA (1997). Allelic loss of 8p sequences in prostatic intraepithelial neoplasia and carcinoma. *Urology* 50: 643-647.

Jefford CE, Irminger-Finger I (2006). Mechanisms of chromosome instability in cancers. *Critical Reviews in Oncology/Hematology* 59: 1-14

MacGrogan D, Levy A, Bostwick D, Wagner M, Wells D, Bookstein R, Loss of chromosome arm 8p loci in prostate cancer: Mapping by quantitative allelic imbalance (1994) *Genes, Chromosomes & Cancer* 10: 151-159.

Macoska JA, Trybus TM, Sakr WA et al., (1994). Fluorescence in situ hybridisation analysis of 8p allelic loss and chromosome 8 instability in human prostate cancer. *Cancer Research* 54: 5390-5395.

Macoska JA, Trybus TM, Benson PD et al., (1995). Evidence for three tumor suppressor gene loci on chromosome 8p in human prostate cancer. *Cancer Research* 55: 5390-5395.

Macoska JA, Trybus TM, Wojno KJ (2000). 8p22 loss concurrent with 8c gain is associated with poor outcome in prostate cancer. *Urology* 55: 776-782.

Meng TC, Lee MS, Lin MF (2000) Interaction between protein tyrosine phosphatase and protein tyrosine kinase is involved in androgen-promoted growth of human prostate cancer cells. *Oncogene* 19:2664-77.

Peters LM, Anderson DW, Griffith AJ, Grundfast KM, San Agustin TB, Madeo AC, Friedman TB, Morell RJ (2002) Mutation of a transcription factor, TFCP2L3, causes

progressive autosomal dominant hearing loss, DFNA28. *Hum. Mol. Genet.* 11: 2877 – 2885.

Rozen S, and Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp 365-386

Schuler, G.D., Boguski, M.S., Stewart, E.A., Stein, L.D., Gyapay, G., Rice, K., White, R.E., Rodriguez-Tome, P., Aggarwal, A., Bajorek, E., Bentolila, S., Birren, B.B., Butler, A., Castle, A.B., Chiannikulchai, N., Chu, A., Clee, C., Cowles, S., Day, P.J., Dibling, T., Drouot, N., Dunham, I., Duprat, S., East, C., Hudson, T.J., et al. (1996) A gene map of the human genome, *Science* 274, 540-546.

Shields JM, Rogers-Graham K, Der CJ (2002). Loss of 转凝蛋白 in breast and colon tumours and in RIE-I cells by ras deregulation of gene expression through raf independent pathways. *J Biol Chem* 277, 9790-9799

Stajich, J.E., Block, D., Boulez, K., Brenner, S.E., Chervitz, S.A., Dagdigian, C., Fuellen, G., Gilbert, J.G., Korf, I., Lapp, H., Lehvaslaiho, H., Matsalla, C., Mungall, C.J., Osborne, B.I., Pocock, M.R., Schattner, P., Senger, M., Stein, L.D., Stupka, E., Wilkinson, M.D., Birney, E. (2002) The Bioperl toolkit: Perl modules for the life sciences, *Genome Res.* 12, 1611-1618.

Stanton, JL, Green DPL (2001). Meta-analysis of gene expression in mouse preimplantation embryo development. *Molecular Human Reproduction*, 7, 545-552.

Webber M et al. (1997a). Acinar differentiation by non-malignant immortalized human prostatic epithelial cells and its loss by malignant cells. *Carcinogenesis*, 18, 1225-1231.

Webber M. et al. (1997). Immortalised and tumorigenic adult human prostatic epithelial cell lines: Characteristics and applications part 2. Tumorigenic cell lines. *Prostate*, 30, 58-64.

<110> ASSINDER, Stephen J
STANTON, Jo-Ann L

<120> 诊断方法和标记物

<130> 516938 TVG

<150> 60/811,407

<151> 2006-06-09

<160> 35

<170> PatentIn version 3.3

<210> 1

<211> 368

<212> DNA

<213> Pspu1

<400> 1

aagaattcgg cacgaggaat atttaagggt aaaatttttc tactttttaa gcttaaaaaa 60

atgttttttt actactgtaa aagtaatgca gagaaatgtt cacttaccac acacatacct 120

ttglaaaaat caccacttaa agtttggttc taaagatttt aggacaccaa gatgcaaata 180

atattttttg cgtttaccig ctttttact actgctgagt ctgcagtggc aagatagcta 240

cacagtacct cagccctcct gctcagtttt taacatctat tgataatact aattacaaga 300

aaatttaaaa tgtctttttg caaaaaata ccataagcag tcaaaacaca attaaaaaaa 360

aaaaaaaaa 368

<210> 2

<211> 394

<212> DNA

<213> Pspu2

<220>

<221> misc_feature

<222> (143).. (143)

<223> "N"可以是"A"或什么也不是

<400> 2

gcaaaattag gacattccca gataaacaaa taaagagact tcatcactaa taaacatgcc 60

ctgtatgaaa tactaagagg agtccttggg gcggacatga aaggacacta gacaaatcca 120

catgaagacg taaagattgc cangtgaagt aactacatag gcaaataata acgatagtat 180

aaatgtattt ttgtttgtga cttttttttt tgtgctgatt taaaagacag ctccatagaa 240

aaaaattata aatctgtgtt gatgggcaca cagtgcacaa agatgtaata ttaatacaat 300

aacagcacag gaatggagag ggaacagagg aatacaaaat ctaactagac ttataacaag 360

caaatggatt gatttaatta aaaaaaaaaa aaaa 394

<210> 3

<211> 392

<212> DNA

<213> Pspu43

<400> 3
 tttattaaca aatataaagt accagacact ccaagtgcctt agaaaaattaa attaaactcat 60
 ttaatttttac taaaaacctg tgagataagg gactattggt atgcccattt ttcagatgag 120
 gaaagtggga cagagtttac tgagtggtag agatgagatt tgaaccagg caagcaggct 180
 ctaggtctgg actcttggtc actctgctct gtccccacac taagccagg atctgggctt 240
 gtgtacatag tcaggtaagt ggaaatcaca ggaattgcta gaggactcca gtagagggat 300
 gattttcatg aatgttcttt catctcaaga aagcaaaggc tgctttttgg gtgcaaagga 360
 gtttggctag aaggaagatc tggagataga gc 392

<210> 4
 <211> 1320
 <212> DNA
 <213> Pspu8

<220>
 <221> misc-feature
 <222> (1021)..(1021)
 <223> s 可以是 C 或 G, w 可以是 A 或 T, 而 k 可以是 T 或 G

<400> 4
 cacgaggcga gaaacaccca agaatgatca attaaaaaaaa aaaaagaaag aaagaaagaa 60
 agaaaagaag cagagttaga ctacagcgga aaataggcgt gctaccacct caggagagtt 120
 ccagggaaga accccacccg cactccaatg aggtcacaaat ggctggagct ctgaggggcc 180
 caggctccct gagccaggag gagaggagaa agtccaagga aagatggctg gcagtcaccc 240
 ctacttcaac ctgcctgact ccacacaccc atcgccgcc tccgctccac ccagcctccg 300
 ctggcaccag cgctgccagc cctctggtgc caccaatggc ctgctggtgg cctctgctggg 360
 tgggggcctg cctgctggct tcgtggggcc ctttctcgt atggcttacc agggttccaa 420
 cctgccctcg ctggagctgc tcattctgtc atgcctcttc cacctcccta ttgccctgct 480
 acttaaacctg cgtggcgacc ccttctggtg acctcctgac atccgaggct gggcctgctt 540
 ctgtgccctg ctcaacgtcc tcagcattgg atgtgcctac agtgcagttc aggtggtgcc 600
 cgctggcaac gctgccactg ttcgcaaagg ttcttccacc gtaigtctcc ctgtcctcac 660
 cctactgcct tgagagccag ggtctcgggt gctacgagtg gtgtggactg ttgggcagca 720
 tcctaggact aatcatcatt ctgggacctg gactctggac actaccagga ggggaccaca 780
 ggtgtctaca ccacctggg ctatgtgcag gctttcctgg gaggcctggc gctgtccctg 840
 gggciaatct ggtctatcgt atctctgcac ttcccatcc tgcctccaa cagtggcctt 900
 cctatctggc ttggatgggg ctgctgggct stgtgccagg cctctttgtg ctgcagaccc 960
 ccgtgttgcc cagtgcctc ctgagttaga gttgtgtggg gggcagagg gatcctcgcc 1020
 swtkggtctc ccttcacatg ggtgtgggct atgcgggtca ccaaggcccc acacctgcc 1080
 ctgggtgctg ctgtcctgca ttccgagggt gttgtggccc ttatactgca gtattatatg 1140
 ctccatgaga ctgtggcact ttctgacatc atggggggcag gggttgtgct gggcagcatt 1200

```

gccatcatta cagcccggaa cctcagctgt gagaggacag ggaagggtga ggagtgaat 1260
agaacttggg agcccggggg ttgggagggg cagggataaa taaagacaaa gactgaagac 1320

<210> 5
<211> 1906
<212> DNA
<213> PSA

<400> 5
agccccaagc ttaccacctg caccgggaga gctgtgtcac catgtgggtc ccggttgtct 60
tcctcacctt gtccgtgacg tggattggig ctgcacccct catcctgtct cggatttgtg 120
gaggctggga gtgcgagaag cattcccaac cctggcaggt gcttgtggcc tctcgtggca 180
gggcagtctg cggcgggtgt ctggtgcacc ccagttgggt cctcacagct gccactgca 240
tcaggaacaa aagcgtgacg ttgctgggtc ggcacagcct gtttcacct gaagacacag 300
gccaggtatt tcaggtcagc cacagcttcc cacaccgct ctacgatatg agcctcctga 360
agaatcgatt cctcaggcca ggtgatgact ccagccacga cctcatgtct ctcgcctgt 420
cagagcctgc cgagctcacg gatgtgtga aggtcatgga cctgccacc caggagccag 480
cactggggac caccigtctc gcctcagggt ggggcagcat tgaaccagag gatttcttga 540
cccaaagaa acttcagtgt gtggacctcc atgttatct caatgacgtg tgtgcgcaag 600
ttcacctca gaaggtagc aagttcatgc tgtgtgtcgg acgtggaca gggggcaaaa 660
gcacctgctc ggggtcatt ctgatcacc aactgacct gccagccctg ccgatgttcc 720
tccatggctc cctagtgcct tggagaggag gtgtctagtc agagagtagt cctggaaggt 780
ggcctctgtg aggagccacg gggacagcat cctgcagatg gtcttggccc ttgtcccacc 840
gacctgtcta caaggactgt cctcgtggac cctccccct gcacaggagc tggacctga 900
agtcccttcc ccaccggcca ggactggagc ccctaccct ctgttggat ccctgccac 960
cttctcttg aagtggctc tggagacatt tctctctct tccaaagctg ggaactgcta 1020
tctgttatct gcctgtccag gtctgaaaga taggatgcc caggcagaaa ctgggactga 1080
cctatctcac tctctccctg cttttacct tagggtgatt ctggggccc acttgtctgt 1140
aatggtgtgc ttcaaggtat cacgtcatgg ggcagtgaa catgtgccct gcccgaaagg 1200
ccttccctgt acaccaaggt ggtgcattac cggaagtga tcaaggacac catcgtggcc 1260
aaccctgag caccctatc aacccttat ttagtaaac ttggaacctt ggaaatgacc 1320
aggccaagac tcaagcctcc ccagttctac tgaccttgt ccttaggtgt gaggtccagg 1380
gttgctagga aaagaaatca gcagacacag gtgtagacca gagtgttct taaatggtgt 1440
aatttgtcc tctctgtgtc ctggggaata ctggccatgc ctggagacat atcactcaat 1500
ttctctgagg acacagatag gatgggggt ctgtgttatt tgtgggtac agagatgaaa 1560
gaggggtggg atccacactg agagagtgga gattgacatg tgctggacac tgtccatgaa 1620
gcactgagca gaagctggag gcacaacgca ccagacactc acagcaagga tggagctgaa 1680

```

```

aacataaccc actctgtcct ggaggcactg ggaagcctag agaaggctgt gagccaagga 1740
gggagggtct tccitttgca tgggatgggg atgaagtaag gagagggact ggacccccctg 1800
gaagctgatt cactatgggg ggagggtgtat tgaagtcctc cagacaaccc tcagatttga 1860
tgatttccta gtagaactca cagaaataaa gagctgttat actgtg 1906

<210> 6
<211> 3922
<212> DNA
<213> PCA3/DD3

<400> 6
acagaagaaa tagcaagtgc cgagaagctg gcatcagaaa aacagagggg agatttgtgt 60
ggctgcagcc gagggagacc aggaagatct gcatggtggg aaggacctga tgatacagag 120
gaattacaac acataacttt agtgtttcaa tgaacaccaa gataaataag tgaagagcta 180
gtccgctgtg agtctcctca gtgacacagg gctggatcac catcgacggc actttctgag 240
tactcagtgc agcaaagaaa gactacagac atctcaatgg caggggtgag aaataagaaa 300
ggctgctgac ttaccatctt gaggccacac atctgctgaa atggagataa ttaacatcac 360
tagaaacagc aagatgacaa tataatgtct aagtagtgac atgtttttgc acatttccag 420
ccccittaaa tatccacaca cacaggaagc acaaaaggaa gcacagagat ccttgggaga 480
aatgcccggc cgccatcttg ggcatcgat gagcctcgcc ctgtgccctg tcccgttgt 540
gaggaagga cattagaaaa tgaattgatg tgttccttaa aggatgggca ggaaaacaga 600
tcctgttgtg gatatttatt tgaacgggat tacagatttg aaatgaagtc acaaagttag 660
cattaccaat gagaggaaaa cagacgagaa aatcttgatg gcttcacaag acatgcaaca 720
aacaaaatgg aatactgtga tgacatgagg cagccaagct ggggaggaga taaccacggg 780
gcagagggtc aggattcttg cctgctgcc taaactgtgc gttcataacc aaatcatttc 840
atatttctaa cctcaaaac aaagctgttg taatatctga tctctacggt tccttctggg 900
cccaacattc tccatataac cagccacact catttttaat atttagttcc cagatctgta 960
ctgtgacctt tctacactgt agaataacat tactcatttg ttcaaagacc cttcgtgttg 1020
ctgcctaata ttagctgac tgtttttcct aaggagtgtt ctggcccagg ggatctgtga 1080
acaggctggg aagcatctca agatctttcc agggttatac ttactagcac acagcatgat 1140
cattacggag tgaattatct aatcaacatc atcctcagtg tctttgccc tactgaaatt 1200
catttccac ttttggtccc attctcaaga cctcaaaatg tcattccatt aatatcacag 1260
gattaaactt ttttttaac ctggaagaat tcaatgttac atgcagctat gggaatttaa 1320
ttacatatit tgttttccag tgcaaagatg actaagtcct ttatccctcc ctttgtttg 1380
atttttttc cagtataaag ttaaaatgct tagccttgta ctgaggctgt atacagcaca 1440
gcctctcccc atccctccag ccttatctgt catcaccatc aaccctccc ataccaccta 1500
aacaaaatct aacttgtaat tccitgaaca tgtcaggaca tacattattc cttctgcctg 1560
agaagctcct ccttgcctct taaatctaga atgatgtaaa gttttgaata agttgactat 1620

```

cttacttcat	gcaaagaagg	gacacatag	agattcatca	tcacatgaga	cagcaaatac	1680
taaaagtgtg	atttgattat	aagagtttag	ataaatatat	gaaatgcaag	agccacagag	1740
ggaatgttta	tggggcacgt	tigtgaacct	gggatgtgaa	gcaaaggcag	ggaacctcat	1800
agtatcttat	ataatatact	tcatttctct	atctctatca	caataacca	caagcttttc	1860
acagaattca	tgcagtgcaa	atccccaag	gtaaccttta	tccatttcat	ggtgagtgcg	1920
ctttagaatt	tggcaaatc	atactggtea	cttatctcaa	ctttgagatg	tgtttgtcct	1980
tgtagttaat	tgaagaaaat	agggcactct	tgtgagccac	tttagggttc	actcctggca	2040
ataaagaatt	tacaaagagc	tactcaggac	cagttgttaa	gagctctgtg	tgtgtgtgtg	2100
tgtgtgtgtg	agtgatcatg	ccaaagtgtg	cctctctctc	ttagccatt	atttcagact	2160
taaaacaagc	atgttttcaa	atggcactat	gagctgccaa	tgatgtatca	ccaccatatac	2220
tcattattct	ccagtaaatg	tgataataat	gtcatctgtt	aacataaaaa	aagtttgact	2280
tcacaaaagc	agctggaaat	ggacaaccac	aatatgcata	aacttaactc	ctaccatcag	2340
ctacacactg	cttgacatat	attgttagaa	gcacctcgca	tltgtgggtt	ctcttaagca	2400
aaatacttgc	attaggtctc	agctggggct	gtgcatcagg	cggtttgaga	aatattcaat	2460
tctcagcaga	agccagaatt	tgaattccct	catcttttag	gaatcattta	ccaggtttgg	2520
agaggattca	gacagctcag	gtgctttcac	taatgtctct	gaacttctgt	ccctctttgt	2580
gttcatggat	agtccaataa	ataatgttat	ctttgaactg	atgctcatag	gagagaatat	2640
aagaactctg	agtataatca	acattaggga	ttcaaagaaa	tattagattt	aagctcacac	2700
tggtaaaaag	gaaccaagat	acaaagaact	ctgagctgtc	atcgccccca	tctctgtgag	2760
ccacaaccaa	cagcaggacc	caacgcatgt	ctgagatcct	taaatcaagg	aaaccagtgt	2820
catgagttag	attctcciat	tatggatgct	agcttctggc	catctctggc	tctcctcttg	2880
acacatatata	gttcttagcc	tttgcttcca	cgacttttat	cttttctcca	acacatcgct	2940
taccaatcct	ctctctgtct	tgttgctttg	gacttcccca	caagaatttc	aacgactctc	3000
aagtcctttc	tccatcccc	accactaacc	tgaatgccta	gacccttatt	tttatttaatt	3060
tccaatagat	gtgcctatg	ggctatatig	ctttagatga	acattagata	tttaaagctc	3120
aagaggttca	aaatccaact	cattatcttc	tctttctttc	acctccctgc	tectctccct	3180
atattactga	tigcactgaa	cagcatggtc	cccaatgtag	ccatgcaa	gagaaaccca	3240
gtggctcctt	gtggtacatg	catgcaagac	tgtgaagcc	agaaggatga	ctgattacgc	3300
ctcatgggtg	gaggggacca	ctcctgggcc	ttcgtgattg	tcaggagcaa	gacctgagat	3360
gtccctgcc	ttcagtgtcc	tctgcatctc	ccctttctaa	tgaagatcca	tagaatttgc	3420
tacatttgag	aattccaatt	aggaactcac	atgttttatac	tgccctatca	attttttaaa	3480
cttgctgaaa	attaagtttt	ttcaaaatct	gtccttgtaa	attacttttt	cttacagtgt	3540
cttggcatac	tatatcaact	ttagattcttt	gttacaactt	ttcttactct	tttatcacca	3600
aagtggtctt	tattctcttt	attattatta	ttttctttta	ctactatatt	acgttggtat	3660

tatittgttc tctatagtat caatttatit gatttagitt caatttatit ttattgctga 3720
 cttttaaaat aagtgattcg gggggtggga gaacagggga gggagagcat taggacaaat 3780
 acciaatgca igtgggactt aaaacctaga tgatgggttg atagggtgcag caaaccacta 3840
 tggcacacgt atacctgtgt aacaaaccta cacattctgc acatgtatcc cagaacgtaa 3900
 agtaaaattt aaaaaaagt ga 3922

<210> 7
 <211> 1574
 <212> DNA
 <213> 转凝蛋白 1

<400> 7
 tcaccacggc ggcagccctt taaacccctc acccagccag cgccccatcc tgtctgtccg 60
 aaccagaca caagtcttca ctcttccctg cgagccctga ggaagccttg tgagtgcatt 120
 ggctggggct tggaggggaag tgggctgga gctggacagg agcagtggtt gcatttcagg 180
 caggctctcc tgaggtccca ggcgcagct ccagctccct ggctaggga acccaccctc 240
 tcagtcagca tgggggcccc agctccaggc aggggtgggt ggatcactag cgtcctggat 300
 ctctctcaga ctgggcagcc ccgggctcat tgaaatgccc cggatgactt ggctagtgc 360
 gaggaattga tggaaaccac cggggtgaga gggaggctcc ccatctcagc cagccacatc 420
 cacaagggtg gtgtaagggt gcaggcgcgg gccggttagg ccaaggctct actgtctgtt 480
 gccctccag gagaacttcc aaggagcttt cccagacat ggccaacaag ggtccttcct 540
 atggcatgag ccgcgaagtg cagtccaaaa tcgagaagaa giatgacgag gagctggagg 600
 agcggctggt ggagtggatc atagtgcagt gtggccctga tgtgggccgc ccagaccgtg 660
 ggcgcttggg ctccaggctc tggctgaaga atggcgtgat tctgagcaag ctggtgaaca 720
 gcctgtaccc tgatggctcc aagccggtga aggtgccga gaaccaccc tccatggtct 780
 tcaagcagat ggagcaggtg gctcagttcc tgaaggcggc tgaggactat ggggtcatca 840
 agactgacat gtccagact gttgacctct ttgaaggcaa agacatggca gcagtgcaga 900
 ggaccctgat ggctttgggc agcttggcag tgaccaagaa tgatgggcac taccgtggag 960
 atcccaactg gtttatgaag aaagcgcagg agcataagag ggaattcaca gagagccagc 1020
 tgcaggaggg aaagcatgtc attggccttc agatgggcag caacagaggg gcctcccagg 1080
 ccggcatgac aggctacgga cgacctcggc agatcatcag ttagagcgga gagggctagc 1140
 cctgagcccc gccctcccc agctccttgg ctgcagccat cccgcttagc ctgcctcacc 1200
 cacacccgtg tggtaacctc agccctggcc aagctttgag gctctgtcac tgagcaatgg 1260
 taactgcacc tgggcagctc ctccctgtgc cccagcctc agcccaacti cttaccgaa 1320
 agcatcactg ccttggcccc tccctcccgg ctgccccat caccctact gtctcctccc 1380
 tgggctaagc aggggagaag cgggctgggg gtagcctgga tgtgggcaa gtccactgtc 1440
 ctcttggcg gcaaaagccc attgaagaag aaccagccca gcctgcccc tatcttgtcc 1500

tggaatatTT ttgggttgg aactcaaaaa aaaaaaaaaa aaatcaatct tttctcaaaa 1560
 aaaaaaaaaa aaaa 1574

<210> 8
 <211> 1360
 <212> DNA
 <213> 转凝蛋白 2

<400> 8
 gcccttgcct tgagtcagtg cgctgctctc cagcccgtt gaacgctccc cgcagccacc 60
 gccacccatt ggaatggcca acaggggacc tgcaiatggc ctgagccggg aggtgcagca 120
 gaagattgag aaacaatatg atgcagatct ggagcagatc ctgatccagt ggaicaccac 180
 ccagtgccga aaggatgttg gccggcccca gcctggacgc gagaacttcc agaactggct 240
 caaggatggc acgggtgctat gtgagctcat taatgcactg taccccagg ggcaggcccc 300
 agtaaagaag atccaggcct ccacatggc cttcaagcag atggagcaga tctctcagtt 360
 cctgcaagca gctgagcgt atggcattaa caccactgac atcttccaaa ctgtggacct 420
 ctgggaagga aagaacatgg cctgtgtgca gcggacgtg atgaatctgg gtgggctggc 480
 agtagccga gatgatgggc tcttctctgg ggatcccaac tggttcccta agaaatccaa 540
 ggagaatcct cggaacttct cagataacca gctgcaagag ggcaagaacg tgatcgggtt 600
 acagatgggc accaaccgcg gggcgtctca ggcaggcatg actggctacg ggatgccacg 660
 ccagatcctc tgatcccacc ccaggccttg cccctgccct cccacgaatg gttaatatat 720
 atgtagatat atattttagc agtgacattc ccagagagcc ccagagctct caagctcctt 780
 tcgtcaggg tgggggggtc agcctgtcct gtcacctctg aggtgcctgc tggcatcctc 840
 tcccccatgc ttactaatac attcccttcc ccatagccat caaaactgga ccaactggcc 900
 tcttcccttc ccttgggacc aaaatttagg ggctcagtc cctcaccgcc atgccctggc 960
 ctattctgtc tctccttctt ccccttggcc tgttctgtct ctgagctctg tgtcctccgt 1020
 tcattccatg gctgggagtc actgatgctg cctctgcctt ctgatgctgg actggccttg 1080
 ctcttacaag tatgtctctc ccacagctgt ggctgcagga acttaattta tagggaggag 1140
 cctgtggcag ctgctgcccc agccacagct gcactgactg tgctcaccac acatctgggg 1200
 cagccttccc tggcaggggc cctcgtggct tctcatcttc cattcccttc actgtggcta 1260
 aggggtgggg tgaggggatg gagaggagg gctgcctacc atggctctgg gcttgaggaa 1320
 gatgagtttg ttgatitaaa taaagaattt gtcatttttg 1360

<210> 9
 <211> 130
 <212> PRT
 <213> 正向读诌 1

<220>
 <221> MISC_FEATURE
 <222> (14)..(14)
 <223> Xaa 表示终止密码子

<220>
 <221> MISC_FEATURE
 <222> (50).. (50)
 <223> Xaa 表示终止密码子

<220>
 <221> MISC_FEATURE
 <222> (52).. (52)
 <223> Xaa 表示终止密码子

<220>
 <221> MISC_FEATURE
 <222> (86).. (86)
 <223> Xaa 表示终止密码子

<220>
 <221> MISC_FEATURE
 <222> (98).. (98)
 <223> Xaa 表示终止密码子

<400> 9

Phe Ile Asn Lys Tyr Lys Val Pro Asp Thr Pro Ser Ala Xaa Lys Ile
 1 5 10 15

Lys Leu Thr His Leu Ile Leu His Lys Thr Cys Glu Ile Arg Asp Tyr
 20 25 30

Cys Tyr Ala His Phe Ser Asp Glu Glu Ser Gly Thr Glu Phe Thr Glu
 35 40 45

Trp Xaa Arg Xaa Asp Leu Asn Pro Gly Lys Gln Ala Leu Gly Leu Asp
 50 55 60

Ser Trp Ser Leu Cys Cys Val Pro Thr Leu Ser Gln Val Ser Gly Leu
 65 70 75 80

Val Tyr Ile Val Arg Xaa Val Glu Cys Thr Gly Ile Ala Arg Gly Leu
 85 90 95

Gln Xaa Arg Asp Asp Phe His Glu Cys Ser Phe Ile Ser Arg Lys Gln
 100 105 110

Arg Leu Leu Phe Gly Cys Lys Gly Val Trp Leu Glu Gly Arg Ser Gly
 115 120 125

Asp Arg
 130

<210> 10
 <211> 130
 <212> PRT
 <213> 正向读诌 2

<220>
 <221> MISC_FEATURE
 <222> (18).. (18)

<223> Xaa 表示终止密码子

<220>

<221> MISC_FEATURE

<222> (21).. (21)

<223> Xaa 表示终止密码子

<220>

<221> MISC_FEATURE

<222> (24).. (24)

<223> Xaa 表示终止密码子

<220>

<221> MISC_FEATURE

<222> (29).. (29)

<223> Xaa 表示终止密码子

<220>

<221> MISC_FEATURE

<222> (54).. (54)

<223> Xaa 表示终止密码子

<220>

<221> MISC_FEATURE

<222> (61).. (61)

<223> Xaa 表示终止密码子

<220>

<221> MISC_FEATURE

<222> (74).. (74)

<223> Xaa 表示终止密码子

<220>

<221> MISC_FEATURE

<222> (83).. (83)

<223> Xaa 表示终止密码子

<220>

<221> MISC_FEATURE

<222> (123).. (123)

<223> Xaa 表示终止密码子

<400> 10

Leu Leu Thr Asn Ile Lys Tyr Gln Thr Leu Gln Val Leu Arg Lys Leu
1 5 10 15

Asn Xaa Leu Ile Xaa Phe Tyr Xaa Lys Pro Val Arg Xaa Gly Thr Ile
20 25 30

Val Met Pro Ile Phe Gln Met Arg Lys Val Gly Gln Ser Leu Leu Ser
35 40 45

Gly Arg Asp Glu Ile Xaa Thr Gln Ala Ser Arg Leu Xaa Val Trp Thr
50 55 60

Leu Gly His Ser Ala Leu Ser Pro His Xaa Ala Arg Tyr Leu Gly Leu
65 70 75 80

Cys Thr Xaa Ser Gly Lys Trp Lys Ser Gln Glu Leu Leu Glu Asp Ser
85 90 95

Ser Arg Gly Met Ile Phe Met Asn Val Leu Ser Ser Gln Glu Ser Lys
 100 105 110

Gly Cys Phe Leu Gly Ala Lys Glu Phe Gly Xaa Lys Glu Asp Leu Glu
 115 120 125

Ile Glu
 130

<210> 11
 <211> 130
 <212> PRT
 <213> 正向读框 3

<220>
 <221> MISC_FEATURE
 <222> (2).. (2)
 <223> Xaa 表示终止密码子

<220>
 <221> MISC_FEATURE
 <222> (5).. (5)
 <223> Xaa 表示终止密码子

<220>
 <221> MISC_FEATURE
 <222> (16).. (16)
 <223> Xaa 表示终止密码子

<220>
 <221> MISC_FEATURE
 <222> (27).. (27)
 <223> Xaa 表示终止密码子

<220>
 <221> MISC_FEATURE
 <222> (39).. (39)
 <223> Xaa 表示终止密码子

<220>
 <221> MISC_FEATURE
 <222> (47).. (47)
 <223> Xaa 表示终止密码子

<220>
 <221> MISC_FEATURE
 <222> (93).. (93)
 <223> Xaa 表示终止密码子

<220>
 <221> MISC_FEATURE
 <222> (100).. (100)
 <223> Xaa 表示终止密码子

<220>
 <221> MISC_FEATURE
 <222> (103).. (103)
 <223> Xaa 表示终止密码子

<220>
 <221> MISC_FEATURE
 <222> (129).. (129)
 <223> Xaa 表示终止密码子

<400> 11

Tyr Xaa Gln Ile Xaa Ser Thr Arg His Ser Lys Cys Leu Glu Asn Xaa
 1 5 10 15

Ile Asn Ser Phe Asn Phe Thr Lys Asn Leu Xaa Asp Lys Gly Leu Leu
 20 25 30

Leu Cys Pro Phe Phe Arg Xaa Gly Lys Trp Asp Arg Val Tyr Xaa Val
 35 40 45

Val Glu Met Arg Phe Glu Pro Arg Gln Ala Gly Ser Arg Ser Gly Leu
 50 55 60

Leu Val Thr Leu Leu Cys Pro His Thr Lys Pro Gly Ile Trp Ala Cys
 65 70 75 80

Val His Ser Gln Val Ser Gly Asn His Arg Asn Cys Xaa Arg Thr Pro
 85 90 95

Val Glu Gly Xaa Phe Ser Xaa Met Phe Phe His Leu Lys Lys Ala Lys
 100 105 110

Ala Ala Phe Trp Val Gln Arg Ser Leu Ala Arg Arg Lys Ile Trp Arg
 115 120 125

Xaa Ser
 130

<210> 12

<211> 130

<212> PRT

<213> 反向读诨 1

<220>

<221> MISC_FEATURE

<222> (29)..(29)

<223> Xaa 表示终止密码子

<220>

<221> MISC_FEATURE

<222> (98)..(98)

<223> Xaa 表示终止密码子

<220>

<221> MISC_FEATURE

<222> (100)..(100)

<223> Xaa 表示终止密码子

<220>

<221> MISC_FEATURE

<222> (112)..(112)

<223> Xaa 表示终止密码子

<220>

<221> MISC_FEATURE

<222> (118)..(118)

<223> Xaa 表示终止密码子

<400> 12

Ala Leu Ser Pro Asp Leu Pro Ser Ser Gln Thr Pro Leu His Pro Lys
1 5 10 15

Ser Ser Leu Cys Phe Leu Glu Met Lys Glu His Ser Xaa Lys Ser Ser
20 25 30

Leu Tyr Trp Ser Pro Leu Ala Ile Pro Val Ile Ser Thr Tyr Leu Thr
35 40 45

Met Tyr Thr Ser Pro Asp Thr Trp Leu Ser Val Gly Thr Glu Gln Ser
50 55 60

Asp Gln Glu Ser Arg Pro Arg Ala Cys Leu Pro Gly Phe Lys Ser His
65 70 75 80

Leu Tyr His Ser Val Asn Ser Val Pro Leu Ser Ser Ser Glu Lys Trp
85 90 95

Ala Xaa Gln Xaa Ser Leu Ile Ser Gln Val Phe Ser Lys Ile Lys Xaa
100 105 110

Val Asn Leu Ile Phe Xaa Ala Leu Gly Val Ser Gly Thr Leu Tyr Leu
115 120 125

Leu Ile
130

<210> 13

<211> 130

<212> PRT

<213> 反向读诌 2

<220>

<221> MISC_FEATURE

<222> (24).. (24)

<223> Xaa 表示终止密码子

<220>

<221> MISC_FEATURE

<222> (38).. (38)

<223> Xaa 表示终止密码子

<220>

<221> MISC_FEATURE

<222> (42).. (42)

<223> Xaa 表示终止密码子

<220>

<221> MISC_FEATURE

<222> (47).. (47)

<223> Xaa 表示终止密码子

<220>

<221> MISC_FEATURE

<222> (85).. (85)

<223> Xaa 表示终止密码子

<220>

<221> MISC_FEATURE

<222> (115).. (115)

<223> Xaa 表示终止密码子

<220>

<221> MISC_FEATURE

<222> (129).. (129)

<223> Xaa 表示终止密码子

<220>

<221> MISC_FEATURE

<222> (130).. (130)

<223> Xaa 表示终止密码子

<400> 13

Leu Tyr Leu Gln Ile Phe Leu Leu Ala Lys Leu Leu Cys Thr Gln Lys
1 5 10 15

Ala Ala Phe Ala Phe Leu Arg Xaa Lys Asn Ile His Glu Asn His Pro
20 25 30

Ser Thr Gly Val Leu Xaa Gln Phe Leu Xaa Phe Pro Leu Thr Xaa Leu
35 40 45

Cys Thr Gln Ala Gln Ile Pro Gly Leu Val Trp Gly Gln Ser Arg Val
50 55 60

Thr Lys Ser Pro Asp Leu Glu Pro Ala Cys Leu Gly Ser Asn Leu Ile
65 70 75 80

Ser Thr Thr Gln Xaa Thr Leu Ser His Phe Pro His Leu Lys Asn Gly
85 90 95

His Asn Asn Ser Pro Leu Ser His Arg Phe Leu Val Lys Leu Asn Glu
100 105 110

Leu Ile Xaa Phe Ser Lys His Leu Glu Cys Leu Val Leu Tyr Ile Cys
115 120 125

Xaa Xaa
130

<210> 14

<211> 130

<212> PRT

<213> 反向读诌 3

<220>

<221> MISC_FEATURE

<222> (8).. (8)

<223> Xaa 表示终止密码子

<220>

<221> MISC_FEATURE

<222> (22)..(22)
 <223> Xaa 表示终止密码子

<220>
 <221> MISC_FEATURE
 <222> (57)..(57)
 <223> Xaa 表示终止密码子

<220>
 <221> MISC_FEATURE
 <222> (64)..(64)
 <223> Xaa 表示终止密码子

<220>
 <221> MISC_FEATURE
 <222> (70)..(70)
 <223> Xaa 表示终止密码子

<220>
 <221> MISC_FEATURE
 <222> (93)..(93)
 <223> Xaa 表示终止密码子

<220>
 <221> MISC_FEATURE
 <222> (107)..(107)
 <223> Xaa 表示终止密码子

<220>
 <221> MISC_FEATURE
 <222> (108)..(108)
 <223> Xaa 表示终止密码子

<220>
 <221> MISC_FEATURE
 <222> (110)..(110)
 <223> Xaa 表示终止密码子

<220>
 <221> MISC_FEATURE
 <222> (113)..(113)
 <223> Xaa 表示终止密码子

<400> 14

Ser Ile Ser Arg Ser Ser Phe Xaa Pro Asn Ser Phe Ala Pro Lys Lys
 1 5 10 15

Gln Pro Leu Leu Ser Xaa Asp Glu Arg Thr Phe Met Lys Ile Ile Pro
 20 25 30

Leu Leu Glu Ser Ser Ser Asn Ser Cys Asp Phe His Leu Pro Asp Tyr
 35 40 45

Val His Lys Pro Arg Tyr Leu Ala Xaa Cys Gly Asp Arg Ala Glu Xaa
 50 55 60

Pro Arg Val Gln Thr Xaa Ser Leu Leu Ala Trp Val Gln Ile Ser Ser
 65 70 75 80

Leu Pro Leu Ser Lys Leu Cys Pro Thr Phe Leu Ile Xaa Lys Met Gly
 85 90 95

Ile Thr Ile Val Pro Tyr Leu Thr Gly Phe Xaa Xaa Asn Xaa Met Asn
100 105 110

Xaa Phe Asn Phe Leu Ser Thr Trp Ser Val Trp Tyr Phe Ile Phe Val
115 120 125

Asn Lys
130

<210> 15
<211> 29
<212> DNA
<213> Pspu1 正向引物

<400> 15
attgtgtttt gactgcttat ggtattttt 29

<210> 16
<211> 28
<212> DNA
<213> Pspu1 反向引物

<400> 16
cagagaaatg ttcacttacc aaacacat 28

<210> 17
<211> 30
<212> DNA
<213> Pspu2 正向引物

<400> 17
aatcaatcca ttigtctgtt ataagtctag 30

<210> 18
<211> 29
<212> DNA
<213> Pspu2 反向引物

<400> 18
aaattaggac attcccagat aaacaaata 29

<210> 19
<211> 21
<212> DNA
<213> Beta-actin 正向引物

<400> 19
aaccgtgaga agatgaccga g 21

<210> 20
<211> 21
<212> DNA
<213> Beta-actin 反向引物

<400> 20
gtgaggatct tcatgaggta g 21

<210> 21

<211> 19	
<212> DNA	
<213> Pspu8 正向引物	
<400> 21	
ccctttctcg tatggctta	19
<210> 22	
<211> 20	
<212> DNA	
<213> Pspu8 反向引物	
<400> 22	
gctgtgagag gacaggggaag	20
<210> 23	
<211> 27	
<212> DNA	
<213> Pspu43 正向引物	
<400> 23	
aacaaatata aagtaccaga cactcca	27
<210> 24	
<211> 23	
<212> DNA	
<213> Pspu43 反向引物	
<400> 24	
atctccagat cttccttcta gcc	23
<210> 25	
<211> 238	
<212> PRT	
<213> PSA 氨基酸序列	
<400> 25	
Met Trp Val Pro Val Val Phe Leu Thr Leu Ser Val Thr Trp Ile Gly	
1 5 10 15	
Ala Ala Pro Leu Ile Leu Ser Arg Ile Val Gly Gly Trp Glu Cys Glu	
20 25 30	
Lys His Ser Gln Pro Trp Gln Val Leu Val Ala Ser Arg Gly Arg Ala	
35 40 45	
Val Cys Gly Gly Val Leu Val His Pro Gln Trp Val Leu Thr Ala Ala	
50 55 60	
His Cys Ile Arg Asn Lys Ser Val Ile Leu Leu Gly Arg His Ser Leu	
65 70 75 80	
Phe His Pro Glu Asp Thr Gly Gln Val Phe Gln Val Ser His Ser Phe	
85 90 95	
Pro His Pro Leu Tyr Asp Met Ser Leu Leu Lys Asn Arg Phe Leu Arg	
100 105 110	

Pro Gly Asp Asp Ser Ser His Asp Leu Met Leu Leu Arg Leu Ser Glu
115 120 125

Pro Ala Glu Leu Thr Asp Ala Val Lys Val Met Asp Leu Pro Thr Gln
130 135 140

Glu Pro Ala Leu Gly Thr Thr Cys Tyr Ala Ser Gly Trp Gly Ser Ile
145 150 155 160

Glu Pro Glu Glu Phe Leu Thr Pro Lys Lys Leu Gln Cys Val Asp Leu
165 170 175

His Val Ile Ser Asn Asp Val Cys Ala Gln Val His Pro Gln Lys Val
180 185 190

Thr Lys Phe Met Leu Cys Ala Gly Arg Trp Thr Gly Gly Lys Ser Thr
195 200 205

Cys Ser Trp Val Ile Leu Ile Thr Glu Leu Thr Met Pro Ala Leu Pro
210 215 220

Met Val Leu His Gly Ser Leu Val Pro Trp Arg Gly Gly Val
225 230 235

<210> 26

<211> 201

<212> PRT

<213> 转凝蛋白1氨基酸序列

<400> 26

Met Ala Asn Lys Gly Pro Ser Tyr Gly Met Ser Arg Glu Val Gln Ser
1 5 10 15

Lys Ile Glu Lys Lys Tyr Asp Glu Glu Leu Glu Glu Arg Leu Val Glu
20 25 30

Trp Ile Ile Val Gln Cys Gly Pro Asp Val Gly Arg Pro Asp Arg Gly
35 40 45

Arg Leu Gly Phe Gln Val Trp Leu Lys Asn Gly Val Ile Leu Ser Lys
50 55 60

Leu Val Asn Ser Leu Tyr Pro Asp Gly Ser Lys Pro Val Lys Val Pro
65 70 75 80

Glu Asn Pro Pro Ser Met Val Phe Lys Gln Met Glu Gln Val Ala Gln
85 90 95

Phe Leu Lys Ala Ala Glu Asp Tyr Gly Val Ile Lys Thr Asp Met Phe
100 105 110

Gln Thr Val Asp Leu Phe Glu Gly Lys Asp Met Ala Ala Val Gln Arg
115 120 125

Thr Leu Met Ala Leu Gly Ser Leu Ala Val Thr Lys Asn Asp Gly His
130 135 140

Tyr Arg Gly Asp Pro Asn Trp Phe Met Lys Lys Ala Gln Glu His Lys
145 150 155 160

Arg Glu Phe Thr Glu Ser Gln Leu Gln Glu Gly Lys His Val Ile Gly
165 170 175

Leu Gln Met Gly Ser Asn Arg Gly Ala Ser Gln Ala Gly Met Thr Gly
180 185 190

Tyr Gly Arg Pro Arg Gln Ile Ile Ser
195 200

<210> 27

<211> 199

<212> PRT

<213> 转凝蛋白 2 氨基酸序列

<400> 27

Met Ala Asn Arg Gly Pro Ala Tyr Gly Leu Ser Arg Glu Val Gln Gln
1 5 10 15

Lys Ile Glu Lys Gln Tyr Asp Ala Asp Leu Glu Gln Ile Leu Ile Gln
20 25 30

Trp Ile Thr Thr Gln Cys Arg Lys Asp Val Gly Arg Pro Gln Pro Gly
35 40 45

Arg Glu Asn Phe Gln Asn Trp Leu Lys Asp Gly Thr Val Leu Cys Glu
50 55 60

Leu Ile Asn Ala Leu Tyr Pro Glu Gly Gln Ala Pro Val Lys Lys Ile
65 70 75 80

Gln Ala Ser Thr Met Ala Phe Lys Gln Met Glu Gln Ile Ser Gln Phe
85 90 95

Leu Gln Ala Ala Glu Arg Tyr Gly Ile Asn Thr Thr Asp Ile Phe Gln
100 105 110

Thr Val Asp Leu Trp Glu Gly Lys Asn Met Ala Cys Val Gln Arg Thr
115 120 125

Leu Met Asn Leu Gly Gly Leu Ala Val Ala Arg Asp Asp Gly Leu Phe
130 135 140

Ser Gly Asp Pro Asn Trp Phe Pro Lys Lys Ser Lys Glu Asn Pro Arg
145 150 155 160

Asn Phe Ser Asp Asn Gln Leu Gln Glu Gly Lys Asn Val Ile Gly Leu
165 170 175

Gln Met Gly Thr Asn Arg Gly Ala Ser Gln Ala Gly Met Thr Gly Tyr
180 185 190

Gly Met Pro Arg Gln Ile Leu
195

<210> 28

<211> 329

<212> DNA

<213> EST 1

<400> 28

```
ctttcttttt ttttgcctta tctccagatc ttccttctag ccaaactcct ttgcacccaa    60
aaagcagcct ttgctttctt gagatgaaag aacattcatg aaaatcatcc ctctactgga    120
gtcctctagc aattcctgtg atttccactt acctgactat gtacacaagc ccagatacct    180
ggcttagtgt ggggacagag cagagtgacc aagagtcagc acctagagcc tgcctgcctg    240
ggttcaaadc tcattctctac cactcagtaa actctgtccc actttcctca tctgaaaaat    300
gggcataaca atagtccttt atctacagg                                     329
```

<210> 29

<211> 312

<212> DNA

<213> EST 2

<400> 29

```
tttttttttt ttttgcctta tctccagatc ttccttctag ccaaactcct ttgcacccaa    60
aaagcagcct ttgctttctt gagatgaaag aacattcatg aaaatcatcc ctctactgga    120
gtcctctagc aattcctgtg atttccactt acctgactat gtacacaagc ccagatacct    180
ggcttagtgt ggggacagag cagagtgacc aagagtcagc acctagagcc tgcctgcctg    240
ggttcaaadc tcattctctac cactcagtaa actctgtccc actttcctca tctgaaaaat    300
gggcataaca at                                     312
```

<210> 30

<211> 329

<212> DNA

<213> EST 3

<400> 30

```
tttttttttt ttttgcctct atctccagat ctctcttcta gccaaactcc ttgcacccaa    60
aaaagcagcc ttgctttctt tgagatgaaa gaacattcat gaaaatcatc ctctactgg    120
agtcctctag caattcctgt gatitccact taccigacta tgtacacaag cccagatacc    180
tggcttagtg tggggacaga gcaaagtac caagagtcca aacctagagc ctgctgcct    240
```

```

gggttcaaat ctcatctcta ccactcagta aactctgtcc cactttcctc atctgaaaaa 300
tgggcataac aatagtccct tatctcaca 329

```

```

<210> 31
<211> 373
<212> DNA
<213> EST 4

```

```

<400> 31
ctttcttttt ttttgctcta tctccagatc ttcttcttag ccaaactcct ttgcacccaa 60
aaagcagcct ttgctttctt gagatgaaag aacattcatg aaaatcatcc ctctactgga 120
gtcctctagc aattcctgtg atttccactt acctgactat gtacacaagc ccagatacct 180
ggcttagtgt ggggacagag cagagtgacc aagagtccag acctagagcc tgcttgcctg 240
ggttcaaate tcatctctac cactcagtaa actctgtccc actttcctca tctgaaaaat 300
gggcataaca atagtccctt atctcacagg ttttagtaa aattaaatga gttaatttaa 360
ttttctaagc act 373

```

```

<210> 32
<211> 421
<212> DNA
<213> EST 5

```

```

<400> 32
tttattaaca aatataaagt accagacact ccaagtgtt agaaaattaa attaactcat 60
ttaattttac taaaacacgt tgagataagg gactattgtt atgccattt ttcagatgag 120
gaaagtggga cagagttaac tgagtggtag agatgagatt tgaaccagg caagcaggct 180
ctaggctctgg actcttggtc actctgctct gtccccacac taagccagg atctgggctt 240
gtgtacatag tcaggtaagt ggaaatcaca ggaattgcta gaggactcca gtagagggat 300
gattttcatg aatgttcttt catctcaaga aagcaaaggc tgctttttgg gtgcaaagga 360
gtttggctag aaggaagaic tggagataga gcaaaaaaaaa agaaagaaaa aaaaaaaaaa 420
a 421

```

```

<210> 33
<211> 296
<212> DNA
<213> EST 6

```

```

<400> 33
ttttttttt gtctatctc cagatcttcc ttctagccaa actcctttgc acccaaaaag 60
cagcctttgc ttcttgaga tgaaagaaca ttcatggaaa tcatccctct actggagtcc 120
tctagcaatt cctgtgattt ccacttacct gactatgtac acaagcccag atacctggct 180
tagtgtgggg acagagcaga gtgaccaaga gtccagacct agagcctgct tgccitgggt 240
caaatctcat ctctaccact cagtaaactc tgtccactt tctcatctg gtcgac 296

```

```

<210> 34
<211> 51

```

<212> PRT

<213> PCA 3A 氨基酸序列

<400> 34

Met Phe Leu His Ile Ser Ser Pro Phe Lys Tyr Pro His Thr Gln Glu
1 5 10 15

Ala Gln Lys Glu Ala Gln Arg Ser Leu Gly Glu Met Pro Gly Arg His
20 25 30

Leu Gly Ser Ser Met Ser Leu Ala Leu Cys Leu Val Pro Leu Val Arg
35 40 45

Gln Gly His
50

<210> 35

<211> 51

<212> PRT

<213> PCA 3B 氨基酸序列

<400> 35

Met Phe Leu His Ile Ser Ser Pro Phe Lys Tyr Pro His Thr Gln Glu
1 5 10 15

Ala Gln Lys Glu Ala Gln Arg Ser Leu Gly Glu Met Pro Gly Arg His
20 25 30

Leu Gly Ser Ser Met Ser Leu Ala Leu Cys Leu Val Pro Leu Val Arg
35 40 45

Glu Gly His
50

Pspu43

TTTATTAAACAAATATAAAGTACCAGACACTCCAAGTGCTTAGAAAATTAAATTAACTCATTTAATTTTACTAAAAACCTGT
GAGATAAGGGACTATTGTTATGCCCATTTTTCAGATGAGGAAAGTGGGACAGAGTTTACTGAGTGGTAGAGATGAGATTG
AACCAGGCAAGCAGGCTCTAGGTCTGGACTCTTGCTCACTCTGCTCTGTCCCCACCTAAGCCAGGTATCTGGGCTTGTG
TACATAGTCAGGTAAGTGGAATCACAGGAATTGCTAGAGGACTCCAGTAGAGGGATGATTTTCATGAATGTTCTTTCATC
TCAAGAAAGCAAAGGCTGCTTTTGGGTGCAAGGAGTTTGGCTAGAAGGAAGATCTGGAGATAGAGC

图 1

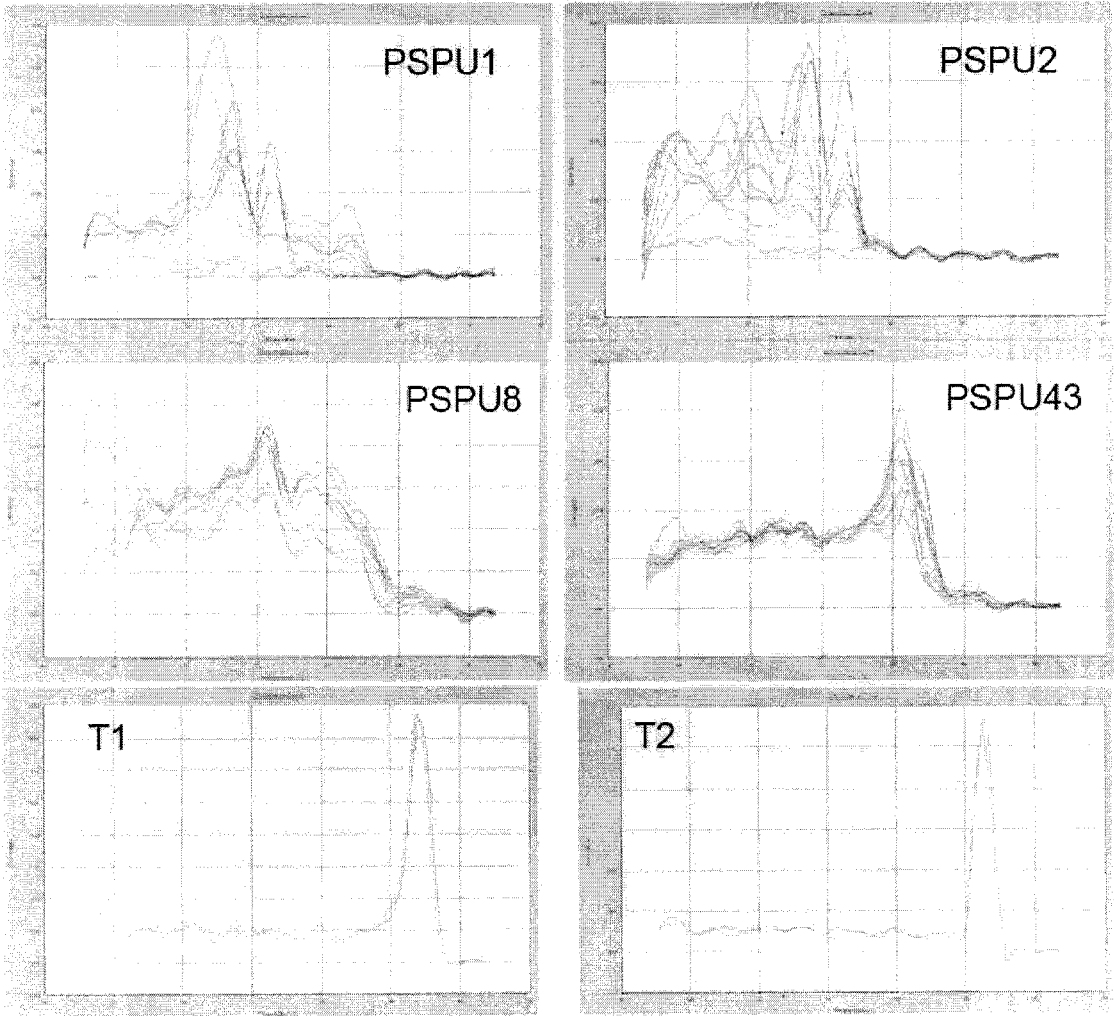


图 2

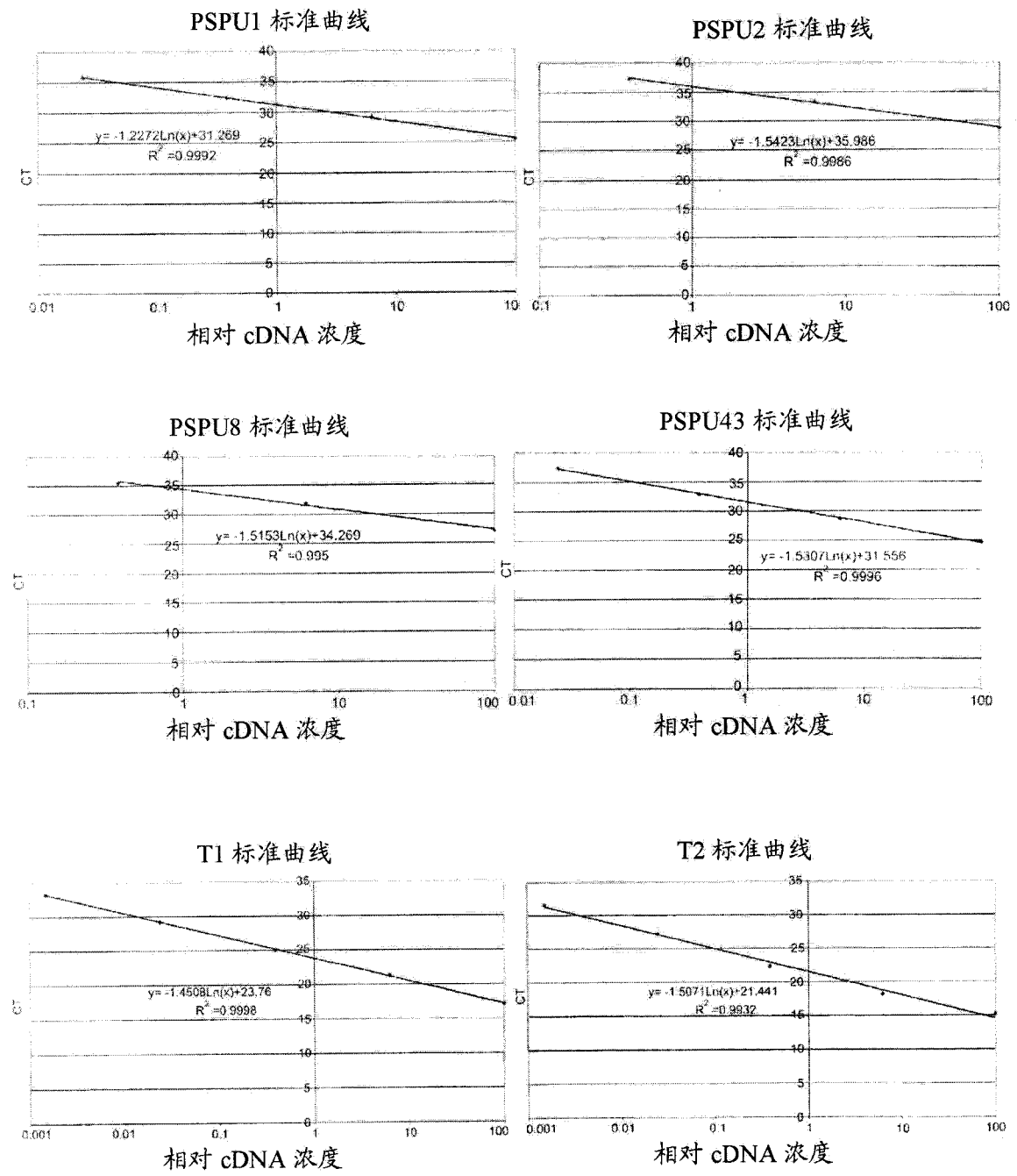


图 3

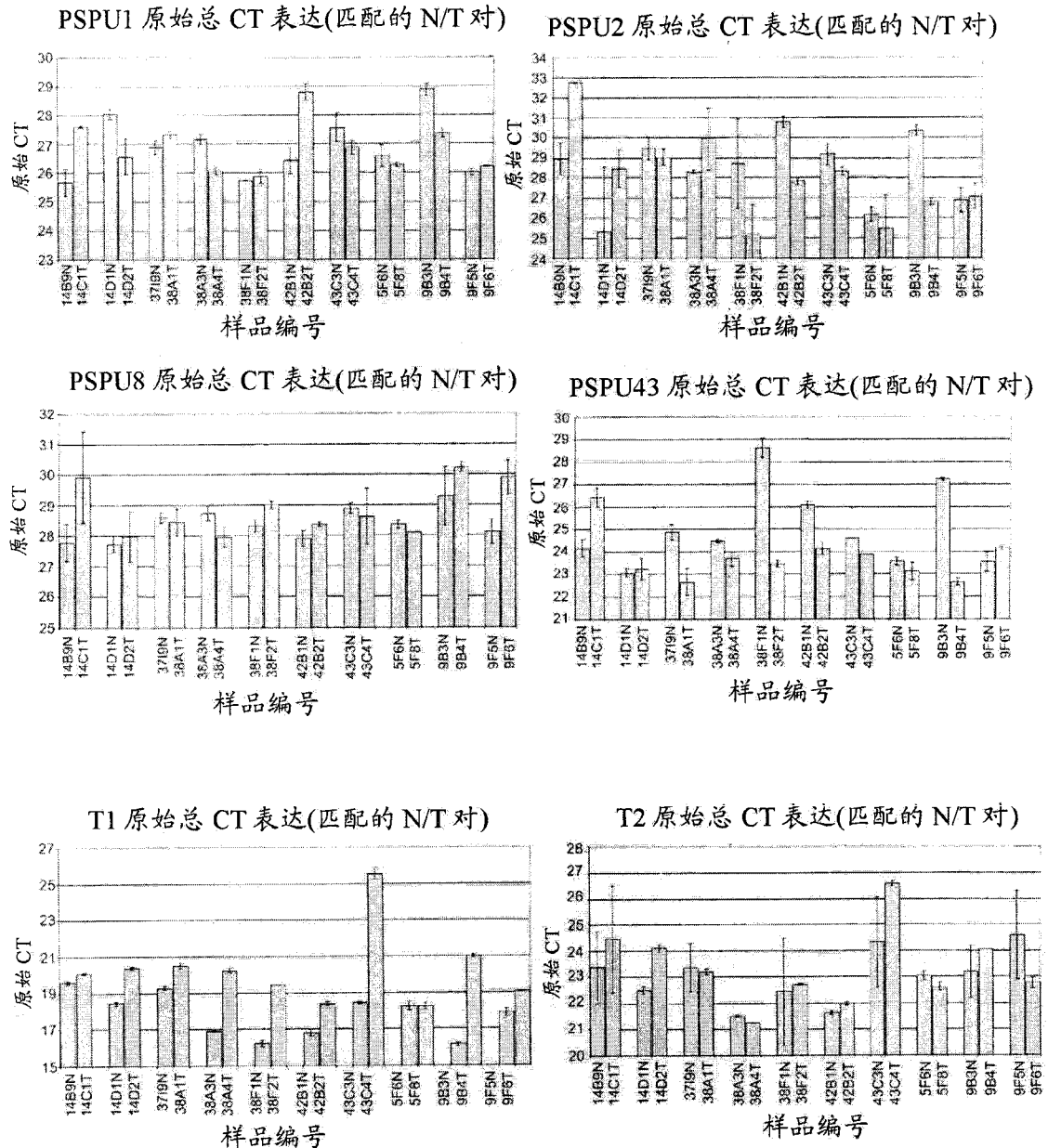


图 4

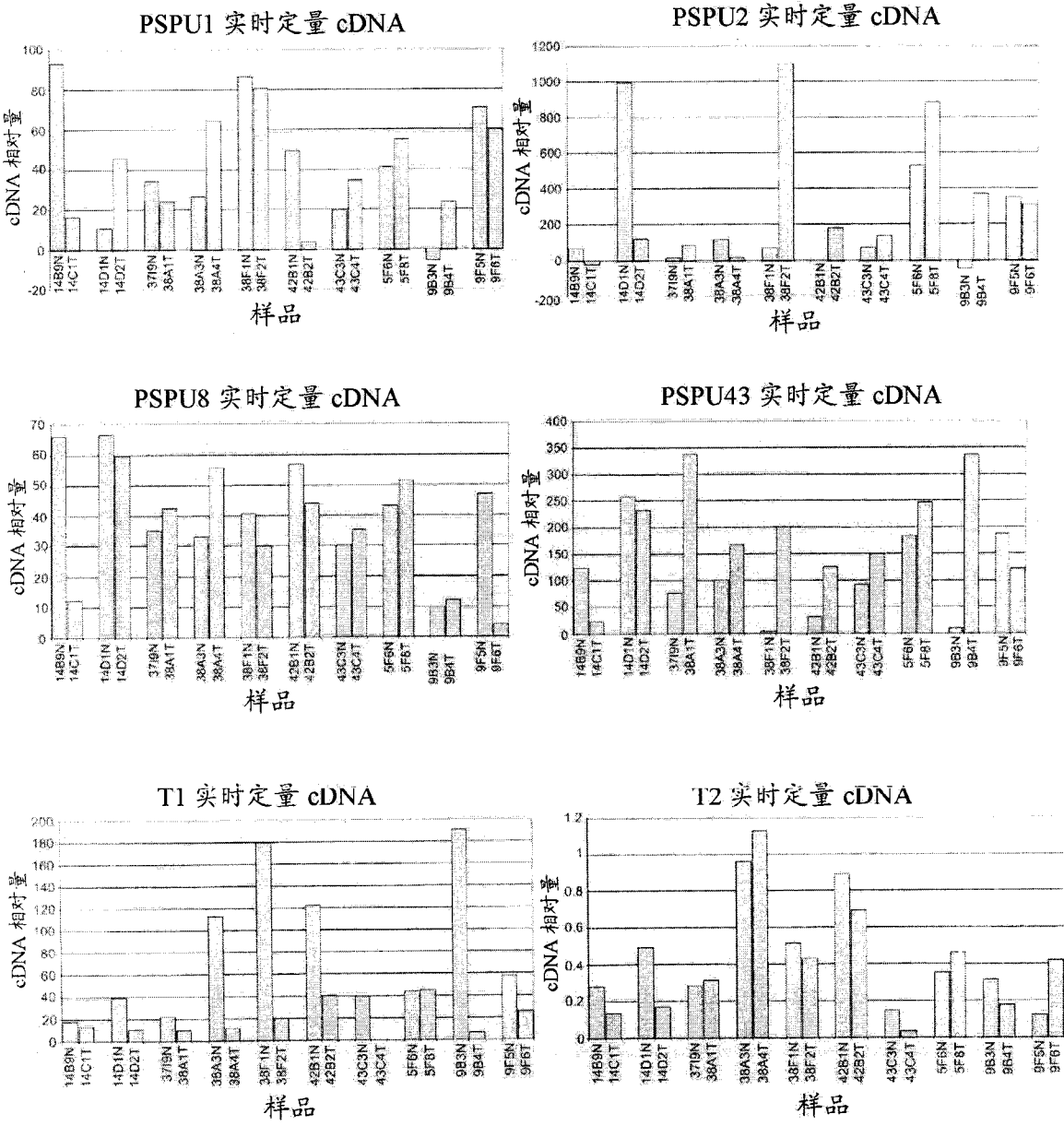


图 5

正向读框 1:

FINKYKVPDTPSA-KIKLTHLILHKTCEIRDYCYAHFSDEESGTEFTW-R-
DLNPGKQALGLDSWSLCCVPTLSQVSGLVYIVR-VECTGIARGLQ-RDDFHECSFISRKQRLLFGCKGVWLEGRSGDR

正向读框 2:

LLTNIKYQTLQVLRKLN-LI-FY-KPVR-GTIVMPIFQMRKVGQSLLSGRDEI-TQASRL-VWTLGHSALSPH-
ARYLGLCT-SGKWSQELLEDDSSRGMI FMNVLSSQESKGCFLGAKEFG-KEDLEIE

正向读框 3:

Y-QI-STRHSKCLN-INSFNFTKNL-DKGLLLCPFFR-GKWDRVY-
VVEMRFEPRQAGSRGGLLVTLCPHTKPGIWACVHSQVSGNHRNC-RTPVEG-FS-
MFFHLKKAKAAPVQVRS LARRKIWR-S

反向读框 1:

ALSPDLPSSTPLHPKSSLCFLEMKEHS-
KSSLYWSPLAIPVISTYLTMYTSPDTWLSVGTEQSDQESRPRACLPFGKSHLYHSVNSVPLSSSEKWA-Q-
SLISQVFSKIK-VNLI F-ALGVSGTLYLLI

反向读框 2:

LYLQIFLLAKLLCTQKAFAFLR-KNIHENHPSTGVL-QFL-FPLT-
LCTQAQIPGLVWGQSRVTKSPDLEPACLGSNLISTTQ-TLSHFPHLKNGHNNSPLSHRFLVKLNELI-
FSKHLECLVLYIC--

反向读框 3:

SISRSSF-PNSFAPKKQPLLS-DERTFMKIIPLESSSNSCDFHLPDYVHKPRYLA-CGDRAE-PRVQT-
SLLAWQISSLPLSKLCPTFLI-KMGITIVPYLTGF--N-MN-FNFLSTWSVWYFIFVNK

图 6