



(12) 发明专利申请

(10) 申请公布号 CN 105132544 A

(43) 申请公布日 2015. 12. 09

(21) 申请号 201510530592. 2

代理人 郑天松

(22) 申请日 2010. 04. 16

(51) Int. Cl.

(30) 优先权数据

C12Q 1/68(2006. 01)

61/202, 881 2009. 04. 16 US

C12N 15/11(2006. 01)

(62) 分案原申请数据

201080020971. 2 2010. 04. 16

(71) 申请人 加拿大国家研究委员会

地址 加拿大安大略

(72) 发明人 E • 王 李杰 邓迎海

A • EG • 伦费林克

M • D • 奥康纳 - 麦考特

E • 普里西马

(74) 专利代理机构 中国国际贸易促进委员会专

利商标事务所 11038

权利要求书8页 说明书33页

(54) 发明名称

用于肿瘤特征和标记物组鉴定,肿瘤分级的方法和用于癌的标记物组

(57) 摘要

用于鉴定肿瘤特征的方法包括:获得各预测目标特征的3个不同标记物组,自肿瘤细胞获得样品基因表达信号,添加报告子以影响样品中的变化,用于允许评定肿瘤中的目标基因表达信号,将基因表达信号与报告子组合,将提取的基因表达信号与3个不同标记物组关联,根据以下顺序将指定分配给提取的基因表达信号:如果全部3个预测性基因表达信号组的关联预测其具有关注的特征,将其指定为恶性肿瘤;如果全部3个预测性基因表达信号组的关联预测其缺乏关注的特征,将其指定为良性肿瘤;及,如果全部3个预测性基因表达信号组的关联不提供相同的预测的临床结果,将所述肿瘤指定为“中等”;及,输出所述指定。

1. 特异于下列生物标记物组的基因的寡核苷酸：

基因名称	基因描述
组 2 (免疫) NRC-1	
C7	补体组分 7
SELE	选择素 E (内皮粘接分子 1)
CD27	CD27 分子
F3	凝固因子 III (凝血致活酶, 组织因子)
IL23A	白细胞介素 23, α 亚基 p19
CARTPT	CART 前原肽
SPP1	分泌的磷蛋白 1 (骨桥蛋白, 骨涎蛋白 I, 早期 T-淋巴细胞活化 1)
TNNT1	肌钙蛋白 T 1 型 (骨骼, 慢)
CACNB3	钙通道, 电压-依赖性, β 3 亚基
C6	补体组分 6
F13B	凝固因子 XIII, B 多肽
SELP	选择素 P (颗粒剂膜蛋白 140kDa, 抗原 CD62)
POU2F2	POU 2 类同源框 2
STAT3	转录 3 的信号转导物和活化物 (急性-期应答因子)
SERPINA1	丝氨酸蛋白酶抑制剂肽酶抑制剂, 分化体 A (α -1 抗蛋白酶, 抗胰蛋白酶), 成员 1
FGF23	成纤维细胞生长因子 23
MYBPC3	肌球蛋白结合蛋白 C, 心脏
LST1	白细胞特异性转录物 1
LEP	瘦素 (肥胖症同源物, 小鼠)
STAT5A	转录的信号转导物和活化物 5A
AMBP	α -1-微球蛋白/尿抑胰酶素前体
TNNC2	肌钙蛋白 C 2 型 (快)
SCN5A	钠通道, 电压-选通的, V 型, α 亚基
CAV1	小窝蛋白 1, 细胞质膜微囊蛋白, 22kDa
RBM4	RNA 结合基序蛋白 4
BLM	Bloom 综合征
FYN	相关于 SRC, FGR, YES 的 FYN 癌基因
BCL6	B 细胞 CLL/淋巴瘤 6 (锌指蛋白 51)
NMU	神经介素 U
HP	结合珠蛋白
组 11 (细胞周期) NRC-2	
RBL1	成视网膜细胞瘤-样 1 (p107)
AURKA	Aurora 激酶 A
CCNE1	细胞周期蛋白 E1
BIRC5	含有杆状病毒 IAP 重复子的 5 (存活素)
NME1	非-转移性细胞 1, 蛋白 (NM23A) 表达于

ASPM	Asp (异常纺锤体) 同源物, 小头畸形关联的 (果蝇属 (<i>Drosophila</i>))
HELLS	解螺旋酶, 淋巴样-特异性
TXNIP	硫氧还蛋白相互作用蛋白
LIG4	连接酶 IV, DNA, ATP-依赖性
SMC4	染色体的结构维持 4
MCM6	微型染色体维持复合体组分 6
CENPF	着丝粒蛋白 F, 350/400ka (分裂激素)
RASSF1	Ras 关联 (RalGDS/AF-6) 结构域家族 1
GMNN	亨蛋白, DNA 复制抑制剂
DLG7	Discs, 大同源物 7 (果蝇属 (<i>Drosophila</i>))
HRASLS3	HRAS-样抑制剂 3
BIN1	桥接整合器 1
LOH11CR2A	杂合性的损失, 11, 染色体区 2, 基因 A
PPP1CC	蛋白磷酸酶 1, 催化性亚基, γ 亚型
CDK2AP1	CDK2-关联蛋白 1
SPAG5	精子关联的抗原 5
NEK1	NIMA (绝不在有丝分裂基因 a 中)-相关的激酶 1
MAPRE3	微管-关联蛋白, RP/EB 家族, 成员 3
CKS1B	CDC28 蛋白激酶调控亚基 1B
MAPK7	促细胞分裂原-活化的蛋白激酶 7
APPL1	衔接子蛋白, 磷酸酪氨酸相互作用, PH 结构域和亮氨酸拉链含有 1
TLK2	Tousled-样激酶 2
SIRT2	Sirtuin (沉默交配类型信息调节 2 同源物) 2 (酿酒酵母 (<i>S. cerevisiae</i>))
MAPRE1	微管-关联蛋白, RP/EB 家族, 成员 1
RAD17	RAD17 同源物 (粟酒裂殖酵母 (<i>S. pombe</i>))
组 12 (凋亡) NRC-3	
TNFRSF11B	肿瘤坏死因子受体超家族, 成员 11b (骨保护素)
ZBTB16	锌指和 BTB 结构域含有 16
APLP1	淀粉样蛋白 β (A4) 前体-样蛋白 1
PDCD4	程序性细胞死亡 4 (肿瘤形成转化抑制剂)
ARHGEF6	Rac/Cdc42 鸟嘌呤核苷酸交换因子 (GEF) 6
SOCS2	细胞因子信号转导抑制剂 2
BIRC5	含有杆状病毒 IAP 重复子的 5 (存活素)
CDC2	细胞分裂周期 2, G1 至 S 和 G2 至 M
ESPL1	额外纺锤体极体同源物 1 (酿酒酵母 (<i>S. cerevisiae</i>))
PHLDA2	普列克底物蛋白同源性-样结构域, 家族 A, 成员 2
CKAP2	细胞骨架关联的蛋白 2
RABEP1	Rab 接触蛋白, RAB GTP 酶结合效应子蛋白 1
NRAS	成神经细胞瘤 RAS 病毒 (v-ras) 癌基因同源物
NME1	非-转移性细胞 1, 蛋白 (NM23A) 表达于
CLU	聚集素
STAT5A	转录的信号转导物和活化物 5A
BCL2	B 细胞 CLL/淋巴瘤 2
CYCS	细胞色素 c, 体细胞

IL17A	白细胞介素 17A
MGMT	O-6-甲基鸟嘌呤-DNA 甲基转移酶
HTATIP2	HIV-1 Tat 相互作用蛋白 2, 30kDa
LRDD	亮氨酸-富重复及死亡结构域含有
CSE1L	CSE1 染色体分开 1-样(酵母)
LIG4	连接酶 IV, DNA, ATP-依赖性
TRADD	TNFRSF1A-关联的经死亡结构域
BTG1	B 细胞转位基因 1, 抗-增殖
GCLM	谷氨酸-半胱氨酸连接酶, 修饰物亚基
NFKB1	B 细胞内的 κ 轻多肽基因增强子的核因子 1 (p105)
PPP2R1B	蛋白磷酸酶 2 (以前 2A), 调控亚基 A, β 亚型
PRLR	催乳素受体

2. 特异于下列生物标记物组的基因的寡核苷酸：

基因名称	基因描述
组 13 (细胞能动性) NRC-4	
TWSG1	扭曲的原肠胚形成同源物 1 (果蝇属 (Drosophila))
KAL1	Kallmann 综合征 1 序列
ANG	血管生成素, 核糖核酸酶, RNA 酶 A 家族, 5
GAB1	GRB2-关联结合蛋白 1
CCL5	趋化因子 (C-C 基序) 配体 5
SELL	选择素 L (淋巴细胞粘附分子 1)
BLR1	伯基特淋巴瘤受体 1, GTP 结合蛋白(趋化因子 (C-X-C 基序) 受体 5)
IL8	白细胞介素 8
NRG2	神经调节素 2
SOX9	SRY (性别决定区 Y)-盒 9 (躯干发育异常, 常染色体性反转)
NAT8	N-乙酰转移酶 8
CXCR4	趋化因子 (C-X-C 基序) 受体 4
VEZT	Vezatin, 粘着连接跨膜蛋白
CCDC88A	卷曲的-线圈结构域含有 88A
ENPEP	谷氨酰氨肽酶(氨肽酶 A)
LAMB1	层粘连蛋白, β 1
FOXE1	叉头框 E1 (甲状腺转录因子 2)
THBS3	凝血酶敏感蛋白 3
LAMC1	层粘连蛋白, γ 1 (以前 LAMB2)
TGFB3	转化生长因子, β 3
PLXND1	丛蛋白 D1
PPAP2A	磷脂酸性磷酸酶类型 2A
RASA1	RAS p21 蛋白活化物 (GTP 酶活化蛋白) 1
CX3CL1	趋化因子 (C-X3-C 基序) 配体 1
HHEX	造血性地表达的同源框
HAND2	心脏和神经嵴衍生物表达的 2

OR1D2	嗅觉受体, 家族 1, 亚家族 D, 成员 2
MAP3K7	促细胞分裂原-活化的蛋白激酶激酶激酶 7
TFAM	转录因子 A, 线粒体
NDN	抑蛋白同源物(小鼠)
组 14 (细胞增殖) NRC-5	
ANG	血管生成素, 核糖核酸酶, RNA 酶 A 家族, 5
CXCL1	趋化因子 (C-X-C 基序) 配体 1 (黑色素瘤生长刺激活性, α)
GAB1	GRB2-关联结合蛋白 1
IGHM	
TGFB1	转化生长因子, β -诱导的, 68kDa
IL8	白细胞介素 8
CD79A	CD79a 分子, 免疫球蛋白-关联的 α
GDF11	生长分化因子 11
SOX9	SRY (性别决定区 Y)-盒 9 (躯干发育异常, 常染色体性反转)
CDKN2D	细胞周期蛋白-依赖性激酶抑制剂 2D (p19, 抑制 CDK4)
PIM2	Pim-2 癌基因
CFDP1	颅面发育蛋白 1
IL15	白细胞介素 15
PPBP	促-血小板碱性蛋白(趋化因子 (C-X-C 基序) 配体 7)
CUL4A	滞蛋白 4A
PMP22	外周髓磷脂蛋白 22
HDGFRP3	肝细胞瘤-衍生的生长因子, 相关的蛋白 3
LYN	V-yes-1 山口肉瘤病毒相关的癌基因同源物
TRAF5	TNF 受体-关联因子 5
WARS	色氨酸-tRNA 合成酶
IL15RA	白细胞介素 15 受体, α
ENPEP	谷氨酰氨肽酶(氨肽酶 A)
PPP1R8	蛋白磷酸酶 1, 调控 (抑制剂) 亚基 8
CCDC88A	卷曲的-线圈结构域含有 88A
TGFB1I1	转化生长因子 β 1 诱导的转录物 1
TSPO	转位子蛋白(18kDa)
PPAP2A	磷脂酸性磷酸酶类型 2A
PCAF	P300/CBP-关联因子
TIMELESS	Timeless 同源物 (果蝇属 (Drosophila))
CADM1	细胞粘接分子 1
组 15 (性) NRC-6	
CD27	CD27 分子
PSG11	怀孕特异性 β -1-糖蛋白 11
ANG	血管生成素, 核糖核酸酶, RNA 酶 A 家族, 5
SOX9	SRY (性别决定区 Y)-盒 9 (躯干发育异常, 常染色体性反转)
SRD5A1	甾-5- α -还原酶, α 多肽 1 (3-氧代-5 α -甾 δ 4-脱氢酶 α 1)
PER3	时期同源物 3 (果蝇属 (Drosophila))
INDO	吲哚胺-吡咯 2,3 二氧化酶
FOXL2	叉头框 L2

PAX5	配对框 5
PSME4	蛋白酶体 (原体, macropain) 活化物亚基 4
SPIN2A	Spindlin 家族, 成员 2A
CXCR4	趋化因子 (C-X-C 基序) 受体 4
CCL2	趋化因子 (C-C 基序) 配体 2
PVRL1	脊髓灰质炎病毒受体-相关的 1 (疱疹病毒进入介质 C)
IL8	白细胞介素 8
OPRK1	类阿片受体, κ 1
ZNF148	锌指蛋白 148
RRAGA	Ras-相关的 GTP 结合 A
DRD5	多巴胺受体 D5
ITCH	Itchy 同源物 E3 泛素蛋白连接酶(小鼠)
EIF5A	真核翻译起始因子 5A
IDE	胰岛素-降解酶
SMCP	精子线粒体-关联半胱氨酸-富蛋白
DBP	白蛋白启动子的 D 位点 (白蛋白 D-盒) 结合蛋白
HSD17B4	羟类固醇 (17- β) 脱氢酶 4
USP9X	泛素特异性肽酶 9, X-连接的
ASB1	含有锚蛋白重复及 SOCS 框的 1
H2AFX	H2A 组蛋白家族, 成员 X
INHBA	抑制素, β A
RLN2	松弛素 2

3. 权利要求 1 或 2 的寡核苷酸或它们的组合用于筛选癌细胞的用途。
4. 权利要求 3 的用途, 其中所述癌是乳腺癌。
5. 权利要求 4 的用途, 其中所述乳腺癌是 ER+ 乳腺癌。
6. 特异于下列生物标记物组的基因的寡核苷酸:

基因名称	基因描述
组 16 (凋亡) NRC-7	
CDH13	钙粘蛋白 13, H-钙粘蛋白(心脏)
SLAMEF7	SLAM 家族成员 7
ANGPTL4	血管生成素-样 4
SULF1	硫酸酯酶 1
GJA1	缝隙连接蛋白, α 1, 43kDa

MUC2	粘蛋白 2, 寡聚体粘液/凝胶-形成
HSPA1B	热休克 70kDa 蛋白 1B
BCL2L14	BCL2-样 14 (凋亡辅助剂)
CASP8AP2	CASP8 关联的蛋白 2
PTK2B	PTK2B 蛋白酪氨酸激酶 2 β
LIG4	连接酶 IV, DNA, ATP-依赖性
GML	GPI 锚定的分子样蛋白
PDCD4	程序性细胞死亡 4 (肿瘤形成转化抑制剂)
MAGEH1	黑色素瘤抗原家族 H, 1
FAS	Fas (TNF 受体超家族, 成员 6)
ANXA5	膜联蛋白 A5
GRM4	谷氨酸受体, 亲代谢型 4
AVEN	凋亡, 胱天蛋白酶活化抑制剂
CASP9	胱天蛋白酶 9, 凋亡-相关的半胱氨酸蛋白酶
CRYAA	眼晶体蛋白, α A
NFKBIA	B 细胞内的 κ 轻多肽基因增强子的核因子抑制剂, α
STK3	丝氨酸/苏氨酸激酶 3 (STE20 同源物, 酵母)
PPP2CB	蛋白磷酸酶 2 (以前 2A), 催化性亚基, β 亚型
CIAPIN1	细胞因子诱导的凋亡抑制剂 1
PEA15	星形胶质细胞中富集的磷蛋白 15
TGFB2	转化生长因子, β 2
E2F2	E2F 转录因子 2
GRIK2	谷氨酸受体, 亲离子型, 红藻氨酸 2
CD74	CD74 分子, 主要组织相容性复合物, II 类不变链
TRAF6	TNF 受体-关联因子 6
组 17 (细胞粘接) NRC-8	
SLAMF7	SLAM 家族成员 7
CDH13	钙粘蛋白 13, H-钙粘蛋白(心脏)
IGSF1	免疫球蛋白超家族, 成员 1
TGFB1	转化生长因子, β -诱导的, 68kDa
HAPLN1	透明质酸糖胺多糖和蛋白聚糖连接蛋白 1
FRAS1	Fraser 综合征 1
CD226	CD226 分子
SUSD5	苏希结构域含有 5
PLEKHC1	普列克底物蛋白同源性结构域含有, 家族 C (有 FERM 结构域) 成员 1
CELSR1	钙粘蛋白, EGF LAG 7-经过 G-类型受体 1 (火烈鸟同源物, 果蝇属 (<i>Drosophila</i>))
NINJ2	Ninjurin 2
ITGAM	整联蛋白, α M (补体组分 3 受体 3 亚基)
GRLF1	糖皮质类固醇受体 DNA 结合因子 1
DCHS2	Dachsous 2 (果蝇属 (<i>Drosophila</i>))
GP1BA	糖蛋白 Ib (血小板), α 多肽
VCL	纽蛋白
SELPLG	选择素 P 配体
PTK2B	PTK2B 蛋白酪氨酸激酶 2 β

NF2	神经纤维瘤蛋白 2 (双侧听神经瘤)
SCARB2	清除剂受体 B 类, 成员 2
ADAM8	ADAM 金属肽酶结构域 8
CYR61	半胱氨酸-富, 血管发生诱导物, 61
NID2	巢蛋白 2 (骨巢蛋白)
VEZT	Vezatin, 粘着连接跨膜蛋白
MUC5AC	粘蛋白 5AC, 寡聚体粘液/凝胶-形成
INPPL1	聚磷酸肌醇磷酸酶-样 1
CXCR3	趋化因子 (C-X-C 基序) 受体 3
NPHP4	肾消耗病 4
SPON1	Spondin 1, 细胞外基质蛋白
PPFIA1	蛋白酪氨酸磷酸酶, 受体类型, f 多肽 (PTPRF), 相互作用蛋白 (LIP 相关蛋白), α 1
组 18 (细胞生长) NRC-9	
CRB1	Crumbs 同源物 1 (果蝇属 (<i>Drosophila</i>))
IGFBP5	胰岛素-样生长因子结合蛋白 5
GAS2	
HTRA1	HtrA 丝氨酸蛋白酶 1
BBC3	BCL2 结合组分 3
GJA1	缝隙连接蛋白, α 1, 43kDa
APOE	载脂蛋白 E
NPR1	促尿钠排泄肽受体 A/鸟苷酸环化酶 A (心房促尿钠排泄肽受体 A)
BAI1	脑-特异性血管发生抑制剂 1
GIN51	GIN5 复合亚基 1 (Psf1 同源物)
EVL	Enah/Vasp-样
SHROOM2	Shroom 家族成员 2
AKT1	V-akt 鼠胸腺瘤病毒癌基因同源物 1
ENO3	烯醇酶 3 (β , 肌肉)
SLC6A3	溶质载体家族 6 (神经递质转运子, 多巴胺), 成员 3
YEATS4	YEATS 结构域含有 4
TBCE	微管蛋白折叠辅因子 E
IGFBP7	胰岛素-样生长因子结合蛋白 7
SOX15	SRY (性别决定区 Y)-盒 15
FGD6	FYVE, RhoGEF 和 PH 结构域含有 6
RB1	成视网膜细胞瘤 1 (包括骨肉瘤)
PLXNA3	丛蛋白 A3
TSG101	肿瘤感受性基因 101
CDA	胞苷脱氨酶
LAMB1	层粘连蛋白, β 1
TGFB2	转化生长因子, β 2
PARD3	Par-3 分隔缺陷 3 同源物 (秀丽隐杆线虫 (<i>C. elegans</i>))
WRN	Werner 综合征
STX2	突触融合蛋白 2
PPP2CB	蛋白磷酸酶 2 (以前 2A), 催化性亚基, β 亚型

7. 权利要求 6 的寡核苷酸用于筛选癌细胞的用途。

8. 权利要求 7 的用途, 其中所述癌是乳腺癌。

9. 权利要求 8 的用途,其中所述乳腺癌是 ER- 乳腺癌。

用于肿瘤特征和标记物组鉴定,肿瘤分级的方法和用于癌的标记物组

[0001] 本申请是 2010 年 4 月 16 日提交的同名发明专利申请 201080020971.2 的分案申请。

【技术领域】

[0002] 本发明涉及癌生物标记物领域,及它们的鉴定及使用方法。

【背景技术】

[0003] 对癌知道越多,可更有效对其进行治疗。例如,多数癌患者进行手术。但是,对于一些患者,用额外的治疗可能得到额外的受益。目前无满意的方法测定哪些患癌患者会受益于手术后额外治疗(例如化学治疗)。可用于预后目的的特异于癌细胞的基因和蛋白的鉴定被认为有帮助。鉴定与用于恢复的差预后关联的肿瘤的这些基因/蛋白,如果仅通过手术处理之后是一般标准的护理,称为差预后生物标记。这些生物标记可用作有价值的工具,其用于预测癌诊断后存活,鉴定复发的风险足够低的患者,在缺失手术后化学治疗和/或放射治疗下或仅一般标准的手术后护理治疗下患者可能同样进展或更佳,及用于向导肿瘤科医生应如何处理癌,以获得最佳结果。

[0004] 类似地,有癌中表达的基因,其在药物应答中起作用。当临床决定时,其会具有对于预测的药物应答有用的信息。

[0005] 为了提供具有临床目标的足够的精确度的筛选工具,优选对于一类癌考虑多标记物。单基因标记物不提供足够的水平的特异性和灵敏度。例如,微阵列技术,其可同时测量多于 25,000 基因,提供发现多-标记物的有用的工具。

[0006] 本发明的目的是提供用于鉴定目标肿瘤特征的一组标记物及它们的鉴定及使用方法。

【发明概述】

[0008] 本发明在一实施方式中教导利用基因表达谱基于几组基因来区别‘良性’和‘恶性’肿瘤。如本文所用当说道预估及患者存活时,术语“良性肿瘤”指可能通过手术和仅一般标准的护理治愈的,无需化学治疗或放射治疗(即便这是一般标准的护理的一部分)的肿瘤。如本文所用,术语“恶性肿瘤”指不可能通过手术和仅一般标准的护理(包括化学治疗或放射治疗)治愈的肿瘤。如本文所用,如果患者在手术后 5 或 10 年内不经历的肿瘤复发(或其转移),则认为肿瘤被“治愈”。

[0009] 可能鉴定表达谱能区别‘良性’和‘恶性’肿瘤的几组基因。现有技术公开 5 种该基因表达信号组,且这些被开发为用于乳腺癌样品的生物标记物。各基因表达信号组源于一组乳腺肿瘤样品。但是,这 5 种生物标记物组不可交叉-使用。特别是,现有技术所谓的“乳腺癌生物标记物”尚未发现当用于另一组乳腺肿瘤样品时可一致地预测预后。用于其他类型的癌的生物标记物具有相同的问题。癌是高度异源性的。常对于一类癌,可发现几个亚型。之前公开的标记物组对于这些亚型不足够通用。

[0010] 为了克服这些问题和数据组（样品）利用度的限制，开发了新方法来发现及使用组生物标记物。

[0011] 在本发明的一实施方式中，从公开的癌数据组产生随机练习（training）数据组，其中包括患者的基因表达谱和临床信息，以发现稳健的生物标记物组。随机练习数据组的基因表达谱与患者存活状态和筛选生物标记物相关。

[0012] 在本发明的一实施方式中提供了鉴定生物标记物的方法，所述方法包括：

[0013] - 自目前可有用的数据组（肿瘤微阵列表征 + 癌患者的临床信息）产生随机练习数据组

[0014] - 针对随机练习数据组筛选基因表达信号组，以鉴定对预后具有预测力的基因表达信号组

[0015] - 基于它们在基因表达信号组中出现的频度排序基因，其具有良好的预测力（经筛选，最后步骤）和由此建造生物标记物组

[0016] - 组合使用利用用于预测的 3 ~ 6 个生物标记物组（即，样品 A 被全部 3 个生物标记物组预测为“良性肿瘤”，我们会说样品 A 是“良性肿瘤”（低 - 风险），如果全部说其是“恶性”的，我们会说其是“恶性”（高 - 风险），另外，我们说其是中等 - 风险）

[0017] - 使用其他独立数据组确证标记物

[0018] “基因表达信号”是基因（例如 mRNA 或蛋白）表达的有形的指示物。

[0019] 在本发明的实施方式中提供了用于鉴定肿瘤特征的方法，所述方法包括以下步骤：

[0020] (1) 获得各预测目标特征的 3 个不同标记物组；

[0021] (2) 自肿瘤细胞提取基因表达信号；

[0022] (3) 将提取的基因表达信号与 3 个不同标记物组关联；

[0023] (4) 根据以下顺序将值分配给提取的基因表达信号：

[0024] (a) 如果全部 3 个预测性基因表达信号组的关联预测其具有关注的特征，将其指定为恶性肿瘤；

[0025] (b) 如果全部 3 个预测性基因表达信号组的关联预测其缺乏关注的特征，将其指定为良性肿瘤；

[0026] (c) 如果全部 3 个预测性基因表达信号组的关联不提供相同的预测的临床结果，将所述肿瘤指定为“中等”。

[0027] 在一些情况中，关注的特征涉及下列之一或多种：转移，发炎，细胞周期，免疫学应答基因，药物抗性基因，及多 - 药物抗性基因。在一些情况中，肿瘤特征负责特定治疗或治疗的组合。

[0028] 在一些情况中，肿瘤特征是导致差患者手术后存活的趋势。

[0029] 在一些情况中，肿瘤特征与患者存活相关，且以上方法的步骤 4 包括根据以下顺序将值分配给提取的基因表达信号：

[0030] (a) 如果全部 3 个预测性基因表达信号组的关联预测其为恶性肿瘤，将其指定为恶性肿瘤，并将建议超过一般标准的护理的更攻击性的治疗；

[0031] (b) 如果全部 3 个预测性基因表达信号组的关联预测其为良性肿瘤，不建议超过护理标准的治疗，且不建议手术后化学治疗或放射治疗；

[0032] (c) 如果全部 3 个预测性基因表达信号组的关联不提供相同的预后, 将所述肿瘤指定为“中等”, 并建议完全一般标准的护理治疗, 包括化学治疗和 / 或放射治疗。

[0033] 当癌具有多于一种亚型时, 其可期望包括预先步骤:

[0034] (a) 鉴定待检查的肿瘤亚型;

[0035] (b) 选择特异于所述肿瘤亚型的标记物组。

[0036] 在一些情况中, 目标肿瘤特征是肿瘤响应特定治疗, 例如化学治疗剂或放射的趋势。在该情况中, 基因表达信号在开发练习组的方法中与肿瘤药物应答相关。需知, 对于特定药物的“良性”肿瘤应答在治疗后会低于 - 平均肿瘤存活, 而“恶性”应答在治疗后会高于平均肿瘤存活。使用此方法, 及依赖于可在用于开发练习组的原肿瘤和临床数据中有用的细节, 开发不仅用于应答个体药物或治疗, 而且治疗的组合的标记物是可能的 (其中原始来源中有足够的数据以允许此)。

[0037] 在本发明的实施方式中提供了测定上述方法中有用的类型的预测性基因表达信号组的方法, 包括以下步骤:

[0038] (1) 获得对于目标癌的已知的肿瘤群的目标特征的基因表达信号信息和患者临床信息;

[0039] (2) 将基因表达信号与有关目标特征的临床患者信息关联, 以鉴定哪些基因具有对于临床结果的预测力;

[0040] (3) 创建至少 30 个自步骤 1 的随机练习数据组;

[0041] (4) 比较步骤 3 的鉴定的基因表达信号与一系列已知的在癌中活跃的基因;

[0042] (5) 选择对应于已知的癌基因列表上的那些的鉴定的基因表达信号;

[0043] (6) 根据它们在生物学过程中的作用将选择的鉴定的基因表达信号分组;

[0044] (7) 自步骤 6 的选择的基因表达信号组产生至少 25 个基因的随机基因表达信号组;

[0045] (8) 将随机基因表达信号组与步骤 3 的随机练习数据组关联;

[0046] (9) 自对于步骤 7 的各基因表达信号组的关联获得用于存活筛选的 P 值;

[0047] (10) 如果对于多于 90 % 的随机练习数据组, 用于基因表达信号组的 P 值小于 0.05, 则保持所述基因表达信号组;

[0048] (11) 基于组中基因出现的频度排序步骤 10 中保持的随机基因表达信号组;

[0049] (12) 选择排在前面的至少 26 个基因作为潜在候选标记物;

[0050] (13) 重复步骤 7 ~ 12, 并产生至少 26 个基因的另一, 独立, 排序组;

[0051] (14) 比较自步骤 12 及步骤 13 的排在前面的基因;

[0052] (15) 如果多于 25 个基因相同, 将所述排在前面的基因保持为标记物组;

[0053] (16) 重复两次步骤 7 ~ 15, 以获得 3 个不同标记物组;

[0054] 在本发明的一实施方式中提供了鉴定相比一般标准的护理需求或多或少攻击性的治疗的患者方法, 所述方法包括:

[0055] ● “基因表达信号”是基因表达 (例如 mRNA) 的有形的指示物 (理论上, 如果技术上可行, 会代之一测量蛋白表达? 或者其他?)。

[0056] 1. 个别研究包含肿瘤和临床患者信息的信息源。将全部报道的细胞内的基因表达信号与患者存活 (5 和 10 年) 关联, 以便鉴定哪些基因在该个体信息源之内具有对于预后

的预测力。鉴定发现与患者存活显著关联的那些基因表达信号用于进一步检查。

[0057] 2. 将步骤 1 中鉴定的基因表达信号与一系列已知的癌基因比较, 并选择对应于已知的基因的已知在癌中具有作用的那些基因表达信号用于进一步分析。(此会通常产生一系列几百到几千个基因表达信号)

[0058] 3. 从步骤 1 的信息源产生至少 30 (一般 30 和 40 之间) 个随机练习数据组。可在多个随机练习数据组中出现相同的个体基因表达信号。

[0059] 4. 将步骤 2 中选择的基因表达信号根据它们在生物学过程中的作用 (例如细胞周期基因, 细胞死亡基因, 免疫学应答基因, 发炎基因等) 来分组, G0 分析

[0060] 5. 产生随机基因表达信号组 (一般约百万个), 各含有约 30 个基因, 自在步骤 3 中产生的单组随机选择。

[0061] 6. 针对各随机练习数据组得到步骤 4 的各随机基因表达信号组的用于存活筛选的 P 值, 你能对此关联进行更详细的描述?

[0062] 7. 对于多于 90% 的随机数据组, 如果 P 值小于 0.05, 则保持随机基因组

[0063] 8. 基于它们中基因出现的频度排序自步骤 7 保持的随机基因表达信号组

[0064] 9. 选择具有最高预测值 (如在步骤 8 中测定) 的排在前面的 30 个基因 (步骤 8 中排序的) 作为潜在候选体。

[0065] 10. 步骤 5 ~ 9 重复, 开始于自在步骤 3 中形成的各组产生随机基因表达信号组, 及产生是潜在候选体的排在前面的 30 个基因的另一, 独立, 排序的组。

[0066] 11. 将步骤 10 中产生的排在前面的 30 个基因与自步骤 9 的排在前面的 30 个基因比较。如果 30 个中 25 或更多相同, 称其为 “稳定的标记”, 且在筛选患者样品中 useful。如果少于 25/30 相同, 除去该数据 (从两组潜在候选体)。需要至少 25 个, 由此可使用第 1 或第 2 组潜在候选体。

[0067] 12. 再重复两次步骤 5 ~ 11, 用于其他 2 组 (步骤 3 的) 基因表达信号。由此, 会有 3 组稳定的标记, 各涉及自步骤 3 的不同组。

[0068] 13. 检查自患者的癌细胞, 以评定它们的基因表达活性及其与 3 个稳定的标记中的基因表达信号的关联。一般而言, 稳定的标记会指示转移的似然性, 因此匹配该标记的高患者表达会指示 “恶性” 肿瘤。但是也可能稳定的标记可指示表达的保护性基因, 例如凋亡基因, 在该情况中, 对于该标记, 那些基因表达标记的高患者表达会指示 “良性” 肿瘤。在任意情况中, 将各稳定的标记与患者样品比较, 及通过各稳定的标记个别预测 “良性” 或 “恶性” 肿瘤。自单稳定的标记指示 “恶性” 或 “良性” 的阈值是什么? Eg. 是否样品中超过 50% 的见于标记的基因表达, 就是 “恶性”? 是否超过 50% 的见于标记的基因表达在对应的非 - 癌性组织中的正常基础水平以上, 就是 “恶性”?

[0069] 14. 组合各 3 组基因表达信号的预测作为患者样品考虑, 及如下将值分配给肿瘤: (a) 如果全部 3 基因表达信号组 (标记) 预测其为恶性肿瘤, 将其指定为恶性肿瘤, 且患者应提供超过一般标准的护理的更攻击性的治疗; (b) 如果全部 3 数据组预测其为良性肿瘤, 患者不应接收超过标准护理的治疗, 且不应使经历手术后化学治疗或放射治疗; (c) 如果全部 3 组基因表达产物不提供相同的预后, 将所述肿瘤指定为 “中等”, 且患者应接收完全一般标准的护理治疗, 包括化学治疗和 / 或放射治疗。

[0070] 在一些情况中, 为此方法, 期望根据它们在生物学过程中的作用使用基因本体论

分析将选择的鉴定的基因表达信号分组。

[0071] 优选创建 30 和 50 之间个随机练习组。更优选地,创建 30 和 40 之间个练习组。

[0072] 有时期望选择已知在癌中活跃的基因,其选自:负责转移,细胞增殖,肿瘤血管化及药物应答的基因。

[0073] 在本发明的一些实施方式中涉及上述方法,在步骤 7 中,产生约 750,000 和 1,250,000 之间,或约 900,000 和 1,100,000 之间或约百万个随机基因表达信号组。

[0074] 在本发明的一些实施方式中,如在以上描述的方法中,在步骤 7 中,产生的随机基因表达信号组含有约 25 和 50 之间,或 28 ~ 32 或约 30 个基因。

[0075] 在本发明的实施方式中,如以上描述的方法中,在步骤 12 中,选择排在前面的 26 ~ 50,或 28 ~ 32 或约 30 个基因。

[0076] 在一些情况中,当考虑涉及患者存活的肿瘤特征时,期望采用选自基本上由以下构成的列表的至少一种癌生物标记物组:NRC-1,NRC-2,NRC-3,NRC-4,NRC-5,NRC-6,NRC-7,NRC-8 和 NRC-9。

[0077] 在本发明的实施方式中提供了试剂盒,其包含至少 3 个标记物组,及用于进行上述方法的使用说明,以便鉴定目标肿瘤特征。在一些情况中,试剂盒包含列于表 1A 或 1B 的至少 10 个基因表达信号。在一些情况中,试剂盒包含根据上述方法鉴定的至少 30 个核酸生物标记物。

[0078] 在本发明的实施方式中提供了表 1A 或 1B 中的任何基因表达信号用于鉴定一种或更多目标肿瘤特征的用途。在一些情况中,至少不同 3 个标记物组在一些情况中用于至少 1,2 或 3 个标记物组,包括见于表 1A 或 1B 的至少 1,5,10,20 或 25 个基因表达信号。在一些情况中,各标记物组含有见于表 1A 或 1B 的至少 1,5,10,20 或 25 个基因表达信号。

[0079] 在本发明的实施方式中,癌生物标记物是乳腺癌生物标记物,且样品的第 1 亚型是 ER+ 样品。

[0080] 在本发明的实施方式中,在上述方法中,随机练习组如下产生:随机挑选样品,同时维持与选出它们的组相同的“良性”和“恶性”肿瘤比。

[0081] 在一些情况中,目标肿瘤特征会涉及患者存活(例如,手术和一般标准的护理后),且在该情况中,方法可用于鉴定相比一般标准的护理需求或多或少攻击性的治疗的患者。(化学治疗和放射治疗对本人是危险的。由此,最好避免将该治疗提供给不需要它们的患者)。

[0082] 在一些情况中,期望通过提取基因表达信号(例如 mRNA,蛋白)及使用特异于目标基因表达信号的报告子检定目标基因表达信号的存在(及在一些情况中,水平)来研究患者的肿瘤组织。此可以允许基本上同时检查多基因表达信号的微-阵列形式进行。报告子可为结合目标核酸序列,目标抗体特异于蛋白,或任何其他该类物质的探针(本领域中已知及常规使用许多该报告子)。报告子实现样品中的变化,允许评定目标基因表达信号。在一些情况中,实现的变化可为样品光学方面的变化,在其他情况中,变化可为样品的另一可检定的方面,例如其放射性或荧光性质的变化。

[0083] 在特定类型的癌具有多于一种亚型的情况中(例如 ER+ 和 ER- 乳腺癌),可优选起初通过亚型分类患者的癌,然后使用关于该亚型开发的标记物。

[0084] 在一些情况中,目标肿瘤特征会涉及对特定治疗的肿瘤应答,并在该情况中,方法

可用于鉴定对于患肿瘤的患者有前途的治疗方法（一种或更多化学治疗剂或治疗的组合）。

[0085] 如本文所用，“肿瘤”包括期望在患者中破坏或中和的任何癌细胞。例如，其可包括见于实体瘤，骨髓瘤，淋巴瘤和白血病的癌细胞。

[0086] 肿瘤通常是哺乳动物或鸟肿瘤，且可为下列之肿瘤：人，猿，猫，狗，猪，牛，绵羊，山羊，兔，小鼠，大鼠，豚鼠，仓鼠，沙鼠，鸡，鸭或鹅。

[0087] 显而易见的是，3个独立组基因表达信号的组合的使用不限于根据本文所述的方法产生的基因表达信号，也可应用于市售的或文献中报道的癌生物标记物数据组。（尽管最终筛选结果的可靠性会某种程度依赖于使用的组的稳健性，因此建议使用稳健的癌生物标记物数据组）。在一些实例中，期望选择包含涉及不同生物学过程的基因的癌生物标记物数据组（例如一种数据组可涉及发炎，另一种涉及细胞周期，而第3种涉及转移）。

[0088] 方法是通用的，且可将其应用于任何类型的癌。例如其关于列于表4的那些癌类型有用。

[0089] 在本发明的实施方式中，将处理应用于测定手术后应如何攻击性地治疗乳腺癌患者。

[0090] 以下提供方法的一实施方式，平行于实施例1的描述：

[0091] - 步骤1：使用肿瘤患者的癌细胞基因微阵列数据和存活信息开发自动存活筛选方法（作为非限制性例，由JM01细胞系的微阵列数据（小鼠乳腺癌细胞系，室内细胞系和数据）鉴定表面及分泌的蛋白），以筛选公共乳腺癌数据组（295个样品，Chang等人，PNAS 102:3738, 2005）。术语“存活筛选”定义为通过进行的Kaplan-Meier分析通过实施Cox-Mantel时间等级测试的各单基因表达值和患者存活状态（“良性”或“恶性”）之间的关联的统计学意义的检查（Cui et al., Molecular Systems Biology, 3:152, 2007）。由此筛选，得到7种蛋白，其可个别区别‘良性’和‘恶性’肿瘤。例如，在实施例1的一部分，选择蛋白（MMP9），在原细胞系中以实验方式确认。当将MMP9抗体应用于细胞系时，癌进展中的上皮到间叶转变被阻断。此结果指示，方法适于发现转移相关的基因。

[0092] - 步骤2：进行表达值与乳腺癌患者存活相关的基因的基因组-范围的存活筛选（在实施例1中，使用2个练习数据组，定义为数据组1（78个样品，van't Veer等人，Nature, 2002）和数据组2（286个样品，Wang等人，Lancet, 365:671, 2005））。得到的基因表达信号列表分别称为S1及S2。这2个列表的总基因称为St基因表达信号列（ $St = S1 + S2$ ）。

[0093] - 步骤3：当目标癌具有多于一种亚-类型时，产生用于第1亚-类型的标记物。（例如，在实施例1中，产生ER+和ER-标记物）。在实施例1中，通过自以上数据组提取全部ER+样品产生ER+肿瘤标记物，并分别定义为S1-ER+（提取自数据组1）和S2-ER+组（提取自数据组2）。通过自S2-ER+组随机挑选N样品（ $N = 60$ ）来产生35个随机-练习-组。“良性”和“恶性”肿瘤比保存与S2-ER+组基本上相同。通过将S1-ER+添加到上述35个随机-练习-组来得到36个练习-组。

[0094] - 步骤4：通过基因组-范围的存活筛选获得基因表达信号列（在实施例1中，St-ER+基因表达信号列表），其涉及重复步骤2，但对于第1肿瘤亚型使用子集，例如数据组，在实施例1中的S1-ER+和S2-ER+组。使用St-ER+基因表达信号列表，进行基因本

体论 (GO) 分析 (使用 GO 注释软件, David, <http://david.abcc.ncifcrf.gov/>), 仅将已知与癌相关的属于 GO 术语的基因, 例如细胞周期, 细胞死亡等用于进一步标记物筛选。

[0095] - 步骤 5: 通过从一个 GO 术语标注的基因随机挑选 30 个基因来从各选择的 GO 术语标注的基因 (正常每 GO 术语约 60-80 基因) 产生 1 百万个不同随机 - 基因 - 组 (各随机 - 基因 - 组含有 30 个基因)。

[0096] - 步骤 6 和 7: 优选使用针对全部第 1 肿瘤亚型练习组 (例如, 在实施例 1 中, 36 个 ER+ 练习组) 的 1 百万随机 - 基因 - 组 (在步骤 3 中产生的) 进行进一步存活筛选。对于各练习组, 检查各随机 - 基因 - 组 (30 个基因) 的表达值和患者存活状态 (“良性”或“恶性”) 之间的关联的统计学意义, 例如通过进行的 Kaplan-Meier 分析通过实施 Cox-Mantel 时间等级测试。如果对于使用针对一个练习组的一个随机 - 基因 - 组的存活筛选, P 值小于 0.05, 表示, 该随机 - 基因 - 组通过该练习组。

[0097] 步骤 7: 当全部第 1 亚型 (例如 36 个 ER+) 练习组具有多于 2,000 个通过的随机 - 基因 - 组时, 或对于多于 90% 的随机练习数据组得到大于 0.05 的 P 值, 则保持这些通过的随机 - 基因 - 组。

[0098] 步骤 8: 基于通过的随机 - 基因 - 组中出现的频度将步骤 7 的保持的随机 - 基因 - 组中的基因排序。

[0099] 步骤 9: 选择排在前面的 30 个基因 (定义为潜在标记物组) 作为潜在 - 标记物 - 组。需知, 当优选 30 个基因时, 可使用 20 和 40 之间个, 更优选 25 和 35 之间个或更优选 27 ~ 33 个。在一些实例中, 用于筛选目的使用的各组中期望 25 ~ 30 个体基因表达信号, 由此, 各输入数可用于产生此输出。

[0100] 步骤 10: 使用在步骤 5 中起初使用的相同的 GO 术语重复步骤 5, 并产生另外 1 百万个不同的随机 - 基因 - 组, 其用于重复步骤 6 和 7。

[0101] 步骤 11: 如果对于排在前面的 30 个的基因成员与潜在 - 标记物 - 组中的那些 (步骤 9) 基本上相同, 其是指潜在 - 标记物 - 组稳定, 并可用作真实癌生物标记物组。此潜在 - 标记物 - 组指定为标记物组 (此现在可用于患者), 如果对于 2 个潜在标记物组的基因表达信号不是基本上相同, 这指示, 这些 GO 术语基因对于发现生物标记物组不合适, 并丢弃所述潜在标记物组而不进一步分析。在一些情况中, 期望在指定为标记物组之前, 在 2 个潜在标记物组中具有 30 个基因表达信号中的至少 25 个相同。在一些情况中, 在 2 个潜在标记物组中期望具有 26, 27, 28, 29 或 30 个基因表达信号相同。

[0102] 步骤 12: 对于 2 个其他组 (步骤 3) 的基因表达信号再重复两次步骤 5 ~ 11。由此, 会有 3 组稳定的标记, 各涉及自步骤 3 的不同组。

[0103] 在实施例 1 中, 得到 3 组标记物 (分别称为 NRC-1, -2 和 -3, 各组含有 30 个基因, 见表 1), 并在 ER+ 练习组 (S1-ER+ 和 S2-ER+) 中测试。阐释测试过程。各练习组中的样品可分为 3 组: 低 - 风险, 中等 - 风险和高 - 风险组。

[0104] 任选的步骤 12b: 作为任选的步骤, 其在实施例 1 中进行, 其可用于进一步分析生物标记物组, 以进一步将高 - 风险组分级。此步骤涉及自高 - 风险组取样品 (其在实施例 1 中通过练习组 S2-ER+ 的 NRC-1, -2 和 -3 分级), 并重复步骤 3, 4, 5, 6, 7 和 8。

[0105] 在实施例 1 中, 分别得到另外 3 组标记物 (称为 NRC-4, -5 和 -6)。各组含有 30 个基因 (见表 1)。这些组靶向通过 NRC-1, -2 和 -3 分级的高 - 风险组。

[0106] - 步骤 12c :作为任选的步骤,在实验 1 中进行的,以获得用于肿瘤的第 2 亚 - 类型的生物标记物 (在实施例 1 中, ER- 肿瘤),提取数据组 1 和 2 中的全部第 2 亚型样品 (例如分别自数据组 1 和 2 的 ER- 样品,及分别定义为 S1-ER- (提取自数据组 1) 和 S2-ER- (提取自数据组 2) 组)。通过自数据组 2,亚型 2 组 (例如 S2-ER- 组) 随机挑选 N 个样品 (N = 40) 产生 35 个随机 - 练习 - 组。维持如总体数据组 2,亚型 2 组 (S2-ER- 组) 中的“良性”和“恶性”肿瘤比。通过将数据组 1,2 型 (例如 S1-ER-) 添加到上述 35 个随机 - 练习 - 组来得到练习 - 组 (36 个,在实施例 1 中)。使用数据组 1,亚型 2 (例如 S1-ER-) 和数据组 2,亚型 2 (例如 S2-ER-) 组重复步骤 4 来获得组合的数据组,亚型 2 (例如 St-ER-) 基因表达信号列表,然后进行 GO 分析。然后重复步骤 5,6,7 和 8。

[0107] 在实施例 1 中,得到另外 3 组标记物 (分别称为 NRC-7, -8 和 -9。各组含有 30 个基因,见表 1)。这些组用于 ER- 样品。

[0108] 【测试过程】

[0109] 【概观,实施例 1 :】

[0110] 在实施例 1 中,对于各标记物组,采用最近皱缩的形心分级和略一法 (leave-one-out method)。然后我们一起组合使用了 3 个标记物组,用于预测各样品的复发。

[0111] 对于给定的数据组,其含有 n 个样品,用于实施例 1 的测试过程如下 (逐步) :

[0112] 步骤 13 :对于靶标的测试样品,我们提取了标记物组的基因表达谱。对于各基因表达值,我们倍增其标记物 - 因子及获得测试样品的修饰的基因表达谱。我们使用 PAM 方法由对于标记物组的 n-1 个样品计算了对于“良性”和“恶性”类二者的标准化的形心 (Tibshirani 等人,PNAS,99:6567,2002)。倍增各基因的标记物 - 因子至类形心 (class centroid),并获得标记物组的修饰的类形心。

[0113] 为了使用标记物组预测靶标的测试样品的复发 :我们比较样品的修饰的基因表达谱与各这些修饰的类形心。最接近的形心的类,在平方的距离中,是对于该样品的预测的类。如果样品被预测为“良性”肿瘤,其被表示为 0,否则,其被表示为 1。

[0114] 步骤 14 :对于 ER+ 样品,如果对于全部 3 个标记物组样品预测为 0,我们将其分配在低 - 风险组中 ;如果对于全部 3 个标记物组样品预测为 1,我们将其分配到高 - 风险组 ;如果样品未分配在低 - 风险组中,也未分配在高 - 风险组中,我们将其分配到中等 - 风险组。对于 ER- 样品,对于全部 3 个标记物组样品预测为 0,我们将其分配到低 - 风险组,否则,我们将其分配到高 - 风险组。这是将两可样品分配到中等组的通常实践的修饰。在高度攻击性的癌亚型的情况中,可期望分类不是明显如高风险低 - 风险的全部癌,并攻击性地治疗它们,超过护理的普通标准物。

[0115] 【3 个测试数据组中的标记物组的有效性】

[0116] 为了测试标记物组的稳健性及预测精确度,我们测试了自这些出版物的 3 个独立乳腺癌数据组中的标记物组 (Koe et al.,Cancer Cell,2006 ;Chang et al.,PNAS 102:3738,2005and Sotiriou C,et al.,J.Natl Cancer Inst,98:262,2006),测试总共 644 个样品。

[0117] 对于 ER+ 样品,在各数据组中,我们首先使用了 NRC-1,-2 和 -3 标记物组 (自以上提及的 3 个乳腺癌数据组) 来将样品分级为低 (LG),中等 (MG) 和高 (HG) - 风险组。如果

高-风险组具有少于 10 个样品,我们合并了 MG 和 HG 组并将其称为中等-风险组。另外,我们使用了 NRC-4, -5 和 -6 标记物组来将 HG 组分级为 3 个新组:低 (NLG), 中等 (NMG) 和高 (NHG)-风险组。我们合并了 NLG 和 MG,并将其称为中等-风险组,且合并 NMG 和 NHG 并将其称为高-风险组。LG 是低-风险组。我们获得了对于低-风险组具有高预测性精确度(对于非-复发患者~90%)的非常良好的结果,并在全部 3 个测试数据组中良好地分了 3 组(见表 2)。

[0118] 对于 ER- 样品,在各数据组中,我们使用了 NRC-7, -8 和 -9 标记物组来将样品分级为低 (LG-) 和高 (HG-)-风险组。我们也获得了对于低-风险组具有高预测精确度(对于非-复发患者~92~100%)的非常良好的结果,并在全部 3 个测试数据组中良好地分了 2 组(见表 2)。

[0119] 【标记物组的组合利用改善预测精确度】

[0120] 对于 ER+ 样品,当 NRC-1, NRC-2 和 NRC-3 全部一致地预测样品为“良性”肿瘤时,精确度相比使用单标记物组(例如 NRC-1, NRC-2 或 NRC-3)显著改善(表 3)。当对于 ER- 样品, NRC-7, NRC-8 和 NRC-9 全部一致预测样品为“良性”肿瘤时,得到相同的结果(表 3)。一般而言,发现 3 个标记物组的综合利用相比使用单组改善预测性精确度。在本发明的一实施方式中,精确度自约 70%改善到约 90%。在本发明的一实施方式中,精确度是至少 90%。在另一实施方式中其为至少 95%。

[0121] 由此,本文提供了稳健的生物标记物组及其用途。

[0122] 应理解,依赖于癌的类型,及患者条件,不同基因表征可认为是“恶性”的。转移通常认为是决定如何治疗患癌患者的重要因素,而生物标记物组物组(例如本文公开的那些)对于该目的有用。此外,生物标记物组可用于鉴定可能良好(或差地)响应一种或更多特定药物的癌细胞类型。无论确切的因素被认为是“良性”或“恶性”,通常期望以含有“良性”和“恶性”基因二者的练习组 S1 和 S2 开始方法。当鉴定良好药物靶标时可考虑基因表达水平,由于高度-表达的靶常是良好的药物靶标。

[0123] 一般而言,低-风险组(具有“良好预后标记”)不会经历治疗,但高-风险组(具有“差预后标记”)应接收除手术之外的治疗。一般而言,中等-风险组也会同样;但是,此会依赖于对于该肿瘤类型的一般标准的护理。

[0124] 本文个别公开了在预测对额外的治疗的需求中有用的各生物标记物组,通过使用多生物标记物组可显著改善总体预测精确度。

[0125] 例如,如果针对 NRC_1, NRC_2 和 NRC_3 筛选患者样品,且全部 3 组指示“良性”预后,患者被认为是低风险。如果全部指示“恶性”预后,样品被认为是高风险。如果 1 或 2 组说“恶性”而其他说是“良性”,癌被认为是中等风险。

[0126] 在本发明的实施方式中,为了测定关于任何 1 个生物标记物组(例如 NRC_1)患者样品是否是“良性”或“恶性”,将生物标记物组用于独立地筛选自大量的患者的 2 库癌细胞代表样品。第 1 库代表“良性”癌细胞(具有已知的不呈现关注的行为或特征,例如转移的临床历史),而第 2 库代表“恶性”癌细胞(具有已知的呈现关注的行为或特征的临床历史)。各“良性”和“恶性”库会对于各生物标记物组分别产生基因表达标记(对于“良性”和“恶性”肿瘤的标准“良性”和“恶性”基因表达标记)。对于患者样品,将患者样品的生物标记物组的基因表达标记与生物标记物组的标准“良性”和“恶性”基因表达标记比较。

最接近地模拟生物标记物组的标准“恶性”标记的那些患者样品被认为是“恶性”，而最接近地模拟生物标记物组的标准“良性”标记的那些被认为是“良好”。

[0127] 在一些情况中方法可涉及一种或更多以下癌生物标记物组的组合使用：NRC-1，NRC-2，NRC-3，NRC-4，NRC-5，NRC-6，NRC-7，NRC-8，NRC-9。

[0128] 当已鉴定亚型（对于此例，ER+/ER-）时使用所述方法的一种可能的方法例：

[0129] - 对于癌细胞的肿瘤样品测定 ER 状态（此常常在临床环境下进行）。

[0130] - 对于 ER+ 样品，如果对于全部 3 个标记物组（NRC-1，-2 和 -3），样品预测为“良性”，将其分配到低 - 风险组；如果对于全部 3 个标记物组，样品预测为“恶性”，将其分配到高 - 风险组；如果样品未被分配到低 - 风险组或高 - 风险组，将其分配到中等 - 风险组。

[0131] - 对于 ER+ 高 - 风险组，其通过标记物组（NRC-1，-2 和 -3）定义，使用标记物组（NRC-4，-5 和 -6）再次预测。如果对于全部 3 个标记物组样品预测为“恶性”，将其分配到高 - 风险组。另外，将其分配到中等 - 风险组，其通过 NRC-1，-2 和 -3 定义。

[0132] - 对于 ER- 样品，对于全部 3 个标记物组（NRC-7，-8 和 -9）样品预测为“良性”，将其分配到低 - 风险组，另外，将其分配到高 - 风险组。

[0133] 在本发明的实施方式中提供了评定患者受益于手术之外的额外的癌治疗的似然性的方法，所述方法包括：

[0134] - 将标记物组的基因探针打印到微阵列基因芯片上

[0135] - 自肿瘤样品提取信使 RNA。

[0136] - 将信使 RNA 杂交到微阵列基因芯片。

[0137] - 扫描杂交的微阵列芯片，以获得对于样品的标记物基因的全部读数。

[0138] - 标准化读数

[0139] - 对于样品，构建各标记物组的基因表达谱

[0140] - 将各标记物组的基因表达谱与标准（已知为“良性”和“恶性”）肿瘤样品的那些关联，以进行预测。

[0141] 对于制造微阵列基因芯片，扫描及标准化阵列数据的详细信息可见于在以公共渠道可得到的文献中，Agilent 公司网站：<http://www.chem.agilent.com/en-US/products/instruments/dnamicroarrays/pages/default.aspx>。

[0142] 表 1A. 用于 ER+ 和 ER- 乳腺癌患者的 NRC 生物标记物基因标记的列表：

[0143]

EntrezGene ID	基因名称	描述
NRC_1 (免疫)		
730	C7	补体组分 7
6401	SELE	选择素 E (内皮粘分子 1)
939	CD27	CD27 分子
2152	F3	凝固因子 III (凝血致活酶, 组织因子)
51561	IL23A	白细胞介素 23, α 亚基 p19
9607	CARTPT	CART 前原肽
6696	SPP1	分泌的磷蛋白 1 (骨桥蛋白, 骨涎蛋白 I, 早期 T-淋巴细胞活化 1)
7138	TNNT1	肌钙蛋白 T 1 型 (骨骼, 慢)
784	CACNB3	钙通道, 电压-依赖性, β 3 亚基
729	C6	补体组分 6
2165	F13B	凝固因子 XIII, B 多肽
6403	SELP	选择素 P (颗粒剂膜蛋白 140kDa, 抗原 CD62)
5452	POU2F2	POU 2 类同源框 2
6774	STAT3	转录 3 的信号转导物和活化物 (急性-期应答因子)
5265	SERPINA1	丝氨酸蛋白酶抑制剂肽酶抑制剂, 分化体 A (α -1 抗蛋白酶, 抗胰蛋白酶), 成员 1
8074	FGF23	成纤维细胞生长因子 23
4607	MYBPC3	肌球蛋白结合蛋白 C, 心脏
7940	LST1	白细胞特异性转录物 1
3952	LEP	瘦素 (肥胖症同源物, 小鼠)
6776	STAT5A	转录的信号转导物和活化物 5A
259	AMBP	α -1-微球蛋白/尿抑胰酶素前体
7125	TNNC2	肌钙蛋白 C 2 型 (快)
6331	SCN5A	钠通道, 电压-选通的, V 型, α 亚基
857	CAV1	小窝蛋白 1, 细胞质膜微囊蛋白, 22kDa
5936	RBM4	RNA 结合基序蛋白 4
641	BLM	Bloom 综合征
2534	FYN	相关于 SRC, FGR, YES 的 FYN 癌基因
604	BCL6	B 细胞 CLL/淋巴瘤 6 (锌指蛋白 51)
10874	NMU	神经介素 U
3240	HP	结合珠蛋白
NRC_2 (细胞周期)		
5933	RBL1	成视网膜细胞瘤-样 1 (p107)
6790	AURKA	Aurora 激酶 A
898	CCNE1	细胞周期蛋白 E1
332	BIRC5	含有杆状病毒 IAP 重复子的 5 (存活素)
4830	NME1	非-转移性细胞 1, 蛋白 (NM23A) 表达于
259266	ASPM	Asp (异常纺锤体) 同源物, 小头畸形关联的 (果蝇属 (<i>Drosophila</i>))
3070	HELLS	解螺旋酶, 淋巴样-特异性

[0144]

10628	TXNIP	硫氧还蛋白相互作用蛋白
3981	LIG4	连接酶 IV, DNA, ATP-依赖性
10051	SMC4	染色体的结构维持 4
4175	MCM6	微型染色体维持复合体组分 6
1063	CENPF	着丝粒蛋白 F, 350/400ka (分裂激素)
11186	RASSF1	Ras 关联 (RalGDS/AF-6) 结构域家族 1
51053	GMNN	亨蛋白, DNA 复制抑制剂
9787	DLG7	Discs, 大同源物 7 (果蝇属 (<i>Drosophila</i>))
11145	HRASLS3	HRAS-样抑制剂 3
274	BIN1	桥接整合器 1
4013	LOH11CR2A	杂合性的损失, 11, 染色体区 2, 基因 A
5501	PPP1CC	蛋白磷酸酶 1, 催化性亚基, γ 亚型
8099	CDK2AP1	CDK2-关联蛋白 1
10615	SPAG5	精子关联的抗原 5
4750	NEK1	NIMA (绝不在有丝分裂基因 a 中)-相关的激酶 1
22924	MAPRE3	微管-关联蛋白, RP/EB 家族, 成员 3
1163	CKS1B	CDC28 蛋白激酶调控亚基 1B
5598	MAPK7	促细胞分裂原-活化的蛋白激酶 7
26060	APPL1	衔接子蛋白, 磷酸酪氨酸相互作用, PH 结构域和亮氨酸拉链含有 1
11011	TLK2	Tousled-样激酶 2
22933	SIRT2	Sirtuin (沉默交配类型信息调节 2 同源物) 2 (酿酒酵母 (<i>S. cerevisiae</i>))
22919	MAPRE1	微管-关联蛋白, RP/EB 家族, 成员 1
5884	RAD17	RAD17 同源物 (粟酒裂殖酵母 (<i>S. pombe</i>))
NRC_3 (凋亡)		
4982	TNFRSF11B	肿瘤坏死因子受体超家族, 成员 11b (骨保护素)
7704	ZBTB16	锌指和 BTB 结构域含有 16
333	APLP1	淀粉样蛋白 β (A4) 前体-样蛋白 1
27250	PDCD4	程序性细胞死亡 4 (肿瘤形成转化抑制剂)
9459	ARHGEF6	Rac/Cdc42 鸟嘌呤核苷酸交换因子 (GEF) 6
8835	SOCS2	细胞因子信号转导抑制剂 2
332	BIRC5	含有杆状病毒 IAP 重复子的 5 (存活素)
983	CDC2	细胞分裂周期 2, G1 至 S 和 G2 至 M
9700	ESPL1	额外纺锤体极体同源物 1 (酿酒酵母 (<i>S. cerevisiae</i>))
7262	PHLDA2	普列克底物蛋白同源性-样结构域, 家族 A, 成员 2
26586	CKAP2	细胞骨架关联的蛋白 2
9135	RABEP1	Rab 接触蛋白, RAB GTP 酶结合效应子蛋白 1
4893	NRAS	成神经细胞瘤 RAS 病毒 (v-ras) 癌基因同源物
4830	NME1	非-转移性细胞 1, 蛋白 (NM23A) 表达于
1191	CLU	聚集素
6776	STAT5A	转录的信号转导物和活化物 5A
596	BCL2	B 细胞 CLL/淋巴瘤 2
54205	CYCS	细胞色素 c, 体细胞
3605	IL17A	白细胞介素 17A
4255	MGMT	O-6-甲基鸟嘌呤-DNA 甲基转移酶
10553	HTATIP2	HIV-1 Tat 相互作用蛋白 2, 30kDa
55367	LRDD	亮氨酸-富重复及死亡结构域含有

[0145]

1434	CSE1L	CSE1 染色体分开 1-样(酵母)
3981	LIG4	连接酶 IV, DNA, ATP-依赖性
8717	TRADD	TNFRSF1A-关联的经死亡结构域
694	BTG1	B 细胞转位基因 1, 抗-增殖
2730	GCLM	谷氨酸-半胱氨酸连接酶, 修饰物亚基
4790	NFKB1	B 细胞内的 κ 轻多肽基因增强子的核因子 1 (p105)
5519	PPP2R1B	蛋白磷酸酶 2 (以前 2A), 调控亚基 A, β 亚型
5618	PRLR	催乳素受体
NRC_4 (细胞能动性)		
57045	TWSG1	扭曲的原肠胚形成同源物 1 (果蝇属 (<i>Drosophila</i>))
3730	KAL1	Kallmann 综合征 1 序列
283	ANG	血管生成素, 核糖核酸酶, RNA 酶 A 家族, 5
2549	GAB1	GRB2-关联结合蛋白 1
6352	CCL5	趋化因子 (C-C 基序) 配体 5
6402	SELL	选择素 L (淋巴细胞粘接分子 1)
643	BLR1	伯基特淋巴瘤受体 1, GTP 结合蛋白(趋化因子 (C-X-C 基序) 受体 5)
3576	IL8	白细胞介素 8
9542	NRG2	神经调节素 2
6662	SOX9	SRY (性别决定区 Y)-盒 9 (躯干发育异常, 常染色体性反转)
9027	NAT8	N-乙酰转移酶 8
7852	CXCR4	趋化因子 (C-X-C 基序) 受体 4
55591	VEZT	Vezatin, 粘着连接跨膜蛋白
55704	CCDC88A	卷曲的-线圈结构域含有 88A
2028	ENPEP	谷氨酰氨肽酶(氨肽酶 A)
3912	LAMB1	层粘连蛋白, β 1
2304	FOXE1	叉头框 E1 (甲状腺转录因子 2)
7059	THBS3	凝血酶敏感蛋白 3
3915	LAMC1	层粘连蛋白, γ 1 (以前 LAMB2)
7043	TGFB3	转化生长因子, β 3
23129	PLXND1	丛蛋白 D1
8611	PPAP2A	磷脂酸性磷酸酶类型 2A
5921	RASA1	RAS p21 蛋白活化物 (GTP 酶活化蛋白) 1
6376	CX3CL1	趋化因子 (C-X3-C 基序) 配体 1
3087	HHEX	造血性地表达的同源框
9464	HAND2	心脏和神经嵴衍生物表达的 2
4991	OR1D2	嗅觉受体, 家族 1, 亚家族 D, 成员 2
6885	MAP3K7	促细胞分裂原-活化的蛋白激酶激酶激酶 7
7019	TFAM	转录因子 A, 线粒体
4692	NDN	抑蛋白同源物(小鼠)
NRC_5 (细胞增殖)		
283	ANG	血管生成素, 核糖核酸酶, RNA 酶 A 家族, 5
2919	CXCL1	趋化因子 (C-X-C 基序) 配体 1 (黑色素瘤生长刺激活性, α)
2549	GAB1	GRB2-关联结合蛋白 1
3507	IGHM	
7045	TGFB1	转化生长因子, β -诱导的, 68kDa
3576	IL8	白细胞介素 8
973	CD79A	CD79a 分子, 免疫球蛋白-关联的 α

[0146]

10220	GDF11	生长分化因子 11
6662	SOX9	SRY (性别决定区 Y)-盒 9 (躯干发育异常, 常染色体性反转)
1032	CDKN2D	细胞周期蛋白-依赖性激酶抑制剂 2D (p19, 抑制 CDK4)
11040	PIM2	Pim-2 癌基因
10428	CFDP1	颅面发育蛋白 1
3600	IL15	白细胞介素 15
5473	PPBP	促-血小板碱性蛋白(趋化因子 (C-X-C 基序) 配体 7)
8451	CUL4A	滞蛋白 4A
5376	PMP22	外周髓磷脂蛋白 22
50810	HDGFRP3	肝细胞瘤-衍生的生长因子, 相关的蛋白 3
4067	LYN	V-yes-1 山口肉瘤病毒相关的癌基因同源物
7188	TRAF5	TNF 受体-关联因子 5
7453	WARS	色氨酸-tRNA 合成酶
3601	IL15RA	白细胞介素 15 受体, α
2028	ENPEP	谷氨酰氨肽酶(氨肽酶 A)
5511	PPP1R8	蛋白磷酸酶 1, 调控 (抑制剂) 亚基 8
55704	CCDC88A	卷曲的-线圈结构域含有 88A
7041	TGFB1I1	转化生长因子 β 1 诱导的转录物 1
706	TSPO	转位子蛋白 (18kDa)
8611	PPAP2A	磷脂酸性磷酸酶类型 2A
8850	PCAF	P300/CBP-关联因子
8914	TIMELESS	Timeless 同源物 (果蝇属 (Drosophila))
23705	CADM1	细胞粘分子 1
NRC_6 (性)		
939	CD27	CD27 分子
5680	PSG11	怀孕特异性 β -1-糖蛋白 11
283	ANG	血管生成素, 核糖核酸酶, RNA 酶 A 家族, 5
6662	SOX9	SRY (性别决定区 Y)-盒 9 (躯干发育异常, 常染色体性反转)
6715	SRD5A1	甾-5- α -还原酶, α 多肽 1 (3-氧代-5 α -甾 δ 4-脱氢酶 α 1)
8863	PER3	时期同源物 3 (果蝇属 (Drosophila))
3620	INDO	吲哚胺-吡咯 2,3 二氧化酶
668	FOXL2	叉头框 L2
5079	PAX5	配对框 5
23198	PSME4	蛋白酶体 (原体, macropain) 活化物亚基 4
54466	SPIN2A	Spindlin 家族, 成员 2A
7852	CXCR4	趋化因子 (C-X-C 基序) 受体 4
6347	CCL2	趋化因子 (C-C 基序) 配体 2
5818	PVRL1	脊髓灰质炎病毒受体-相关的 1 (疱疹病毒进入介质 C)
3576	IL8	白细胞介素 8
4986	OPRK1	类阿片受体, κ 1
7707	ZNF148	锌指蛋白 148
10670	RRAGA	Ras-相关的 GTP 结合 A
1816	DRD5	多巴胺受体 D5
83737	ITCH	Itchy 同源物 E3 泛素蛋白连接酶(小鼠)
1984	EIF5A	真核翻译起始因子 5A
3416	IDE	胰岛素-降解酶
4184	SMCP	精子线粒体-关联半胱氨酸-富蛋白
1628	DBP	白蛋白启动子的 D 位点 (白蛋白 D-盒) 结合蛋白

[0147]

3295	HSD17B4	羟类固醇 (17- β) 脱氢酶 4
8239	USP9X	泛素特异性肽酶 9, X-连接的
51665	ASB1	含有锚蛋白重复及 SOCS 框的 1
3014	H2AFX	H2A 组蛋白家族, 成员 X
3624	INHBA	抑制素, β A
6019	RLN2	松弛素 2
NRC_7 (凋亡)		
1012	CDH13	钙粘蛋白 13, H-钙粘蛋白(心脏)
57823	SLAMF7	SLAM 家族成员 7
51129	ANGPTL4	血管生成素-样 4
23213	SULF1	硫酸酯酶 1
2697	GJA1	缝隙连接蛋白, α 1, 43kDa
4583	MUC2	粘蛋白 2, 寡聚体粘液/凝胶-形成
3304	HSPA1B	热休克 70kDa 蛋白 1B
79370	BCL2L14	BCL2-样 14 (凋亡辅助剂)
9994	CASP8AP2	CASP8 关联的蛋白 2
2185	PTK2B	PTK2B 蛋白酪氨酸激酶 2 β
3981	LIG4	连接酶 IV, DNA, ATP-依赖性
2765	GML	GPI 锚定的分子样蛋白
27250	PDCD4	程序性细胞死亡 4 (肿瘤形成转化抑制剂)
28986	MAGEH1	黑色素瘤抗原家族 H, 1
355	FAS	Fas (TNF 受体超家族, 成员 6)
308	ANXA5	膜联蛋白 A5
2914	GRM4	谷氨酸受体, 亲代谢型 4
57099	AVEN	凋亡, 胱天蛋白酶活化抑制剂
842	CASP9	胱天蛋白酶 9, 凋亡-相关的半胱氨酸肽酶
1409	CRYAA	眼晶体蛋白, α A
4792	NFKBIA	B 细胞内的 κ 轻多肽基因增强子的核因子抑制剂, α
6788	STK3	丝氨酸/苏氨酸激酶 3 (STE20 同源物, 酵母)
5516	PPP2CB	蛋白磷酸酶 2 (以前 2A), 催化性亚基, β 亚型
57019	CIAPIN1	细胞因子诱导的凋亡抑制剂 1
8682	PEA15	星形胶质细胞中富集的磷蛋白 15
7042	TGFB2	转化生长因子, β 2
1870	E2F2	E2F 转录因子 2
2898	GRIK2	谷氨酸受体, 亲离子型, 红藻氨酸 2
972	CD74	CD74 分子, 主要组织相容性复合物, II 类不变链
7189	TRAF6	TNF 受体-关联因子 6
NRC_8 (细胞粘接)		
57823	SLAMF7	SLAM 家族成员 7
1012	CDH13	钙粘蛋白 13, H-钙粘蛋白(心脏)
3547	IGSF1	免疫球蛋白超家族, 成员 1
7045	TGFB1	转化生长因子, β -诱导的, 68kDa
1404	HAPLN1	透明质酸糖胺多糖和蛋白聚糖连接蛋白 1
80144	FRAS1	Fraser 综合征 1
10666	CD226	CD226 分子
26032	SUSD5	苏希结构域含有 5
10979	PLEKHC1	普列克底物蛋白同源性结构域含有, 家族 C (有 FERM 结构域) 成员 1

[0148]

9620	CELSR1	钙粘蛋白, EGF LAG 7-经过 G-类型受体 1 (火烈鸟同源物, 果蝇属 (Drosophila))
4815	NINJ2	Ninjurin 2
3684	ITGAM	整联蛋白, α M (补体组分 3 受体 3 亚基)
2909	GRLF1	糖皮质类固醇受体 DNA 结合因子 1
54798	DCHS2	Dachsous 2 (果蝇属 (Drosophila))
2811	GP1BA	糖蛋白 Ib (血小板), α 多肽
7414	VCL	纽蛋白
6404	SELPLG	选择素 P 配体
2185	PTK2B	PTK2B 蛋白酪氨酸激酶 2 β
4771	NF2	神经纤维瘤蛋白 2 (双侧听神经瘤)
950	SCARB2	清除剂受体 B 类, 成员 2
101	ADAM8	ADAM 金属肽酶结构域 8
3491	CYR61	半胱氨酸-富, 血管发生诱导物, 61
22795	NID2	巢蛋白 2 (骨巢蛋白)
55591	VEZT	Vezatin, 粘着连接跨膜蛋白
4586	MUC5AC	粘蛋白 5AC, 寡聚体粘液/凝胶-形成
3636	INPPL1	聚磷酸肌醇磷酸酶-样 1
2833	CXCR3	趋化因子 (C-X-C 基序) 受体 3
261734	NPHP4	肾消耗病 4
10418	SPON1	Spondin 1, 细胞外基质蛋白
8500	PPF1A1	蛋白酪氨酸磷酸酶, 受体类型, f 多肽 (PTPRF), 相互作用蛋白 (LIP 相关蛋白), α 1
NRC_9 (细胞生长)		
23418	CRB1	Crumbs 同源物 1 (果蝇属 (Drosophila))
3488	IGFBP5	胰岛素-样生长因子结合蛋白 5
2620	GAS2	
5654	HTRA1	HtrA 丝氨酸蛋白酶 1
27113	BBC3	BCL2 结合组分 3
2697	GJA1	缝隙连接蛋白, α 1, 43kDa
348	APOE	载脂蛋白 E
4881	NPR1	促尿钠排泄肽受体 A/鸟苷酸环化酶 A (心房促尿钠排泄肽受体 A)
575	BAI1	脑-特异性血管发生抑制剂 1
9837	GIN51	GIN5 复合亚基 1 (Psf1 同源物)
51466	EVL	Enah/Vasp-样
357	SHROOM2	Shroom 家族成员 2
207	AKT1	V-akt 鼠胸腺瘤病毒癌基因同源物 1
2027	ENO3	烯醇酶 3 (β , 肌肉)
6531	SLC6A3	溶质载体家族 6 (神经递质转运子, 多巴胺), 成员 3
8089	YEATS4	YEATS 结构域含有 4
6905	TBCE	微管蛋白折叠辅因子 E
3490	IGFBP7	胰岛素-样生长因子结合蛋白 7
6665	SOX15	SRV (性别决定区 Y)-盒 15
55785	FGD6	FYVE, RhoGEF 和 PH 结构域含有 6
5925	RB1	成视网膜细胞瘤 1 (包括骨肉瘤)
55558	PLXNA3	丛蛋白 A3
7251	TSG101	肿瘤感受性基因 101
978	CDA	胞苷脱氨酶

[0149]

3912	LAMB1	层粘连蛋白, β 1
7042	TGFB2	转化生长因子, β 2
56288	PARD3	Par-3 分隔缺陷 3 同源物 (秀丽隐杆线虫 (<i>C. elegans</i>))
7486	WRN	Werner 综合征
2054	STX2	突触融合蛋白 2
5516	PPP2CB	蛋白磷酸酶 2 (以前 2A), 催化性亚基, β 亚型

[0150] 注:用于列于此表的各基因的信使 RNA 序列已附在此文档结尾处。用于表 1 中的各基因的全部信使 RNA 序列摘自美国生物技术信息中心 (NCBI), 公共数据库。

[0151] 序列格式是 FASTA 格式。FASTA 格式中的序列开始于单 - 行描述, 之后是序列数据行。描述行通过第 1 栏中的大于 (“>”) 符号与序列数据相区别。

[0152] FASTA 格式的例序列:

[0153] >6019|NM_005059

[0154] ATGCCTCGCCTGTTTTTTTTCCACCTGCTAGGAGTCTGTTACTACTGAACCAATTTTCCAGAGCAGTC
GCGGACTCATGGATGGAGGAAGTTATTAATATGCGGCCGGAATTAGTTCGCGCGCAGATTGCCATTTGCGGCAT
GAGCACCTGGAGCAAAAGGTCTCTGAGCCAGGAAGATGCTCCTCAGACACCTAGACCAGTGGCAGGTGATTTTATTC
AAACAGTCTCACTGGGAATCTCACCGGACGGAGGGAAAGCACTGAGAACAGGAAGCTGCTTACCCGAGAGTTCTT
GGTGCCCTTTCCAAATTGTGCCATCCTTCATCAACAAAGATACAGAAACCATAAATATGATGTCAGAATTTGTTGCT
AATTTGCCACAGGAGCTGAAGTTAACCTGTCTGAGATGCAGCCAGCATTACCACAGCTACAACAACATGTACCTGT
ATTTAAAGATTCCAGTCTTCTCTTTGAAGAATTTAAGAACTTATTCGCAATAGACAAAGTGAAGCCGCAGACAGCA
GTCCTTCAGAATTTAAATACTTAGGCTTGATACTCATTCTCGAAAAAAGAGACAACCTCTACAGTGCATTGGCTAAT
AAATGTTGCCATGTTGGTTGTACCAAAAGATCTCTTGCTAGATTTTGCTGAGATGAAGCTAATTGTGCACATCTCGT
ATAATATTCACACATATTCTTAATGACATTTCACTGATGCTTCTATCAGGTCCCATCAATTCTTAGAATATCTAAGA
ATCTTTGTTAGATATTAGGTCCCATCAATTCTTAGAATATCTAAACATCTTTGTTGATGTTTAGATTTTTTTATTTG
ATGTGTAAGAAAATGTTCTTTGTGTGATTAAATGACACATTTTTTTTGCTG

[0155] 在描述行中, 第 1 项, 6019 是 NCBI EntrezGene ID, 其为表 1 的第 1 栏中的 ID; 符号 (“|”) 后的另一项是 NCBI 参考信息 RNA 序列 ID。应注意, 一个 EntrezGene ID 可有几个参考信息 RNA 序列。在此情况中, 列出了对于 1 个 EntrezGene ID 的全部信使 RNA 序列。各序列代表 1 个参考信息 RNA 序列。

[0156] 表 1B. NRC 基因标记的基因表达信号列表

[0157]

NRC-1 (细胞周期)		
基因名称	EntrezGene ID	基因描述
RBL1	5933	成视网膜细胞瘤-样 1 (p107)
CCNF	899	细胞周期蛋白 F
NME1	4830	非-转移性细胞 1, 蛋白 (NM23A) 表达于
CDK2AP1	8099	CDK2-关联蛋白 1
BIRC5	332	含有杆状病毒 IAP 重复子的 5 (存活素)
TLK2	11011	Tousled-样激酶 2
SMC4	10051	染色体的结构维持 4
CCNE1	898	细胞周期蛋白 E1
APPL1	26060	衔接子蛋白, 磷酸酪氨酸相互作用, PH 结构域和亮氨酸拉链含有

[0158]

		1
LOH11CR2A	4013	杂合性的损失, 11, 染色体区 2, 基因 A
MAPRE1	22919	微管-关联蛋白, RP/EB 家族, 成员 1
HRASLS3	11145	HRAS-样抑制剂 3
GADD45A	1647	生长停止及 DNA-损伤-诱导性, α
HELLS	3070	解螺旋酶, 淋巴样-特异性
PPP1CC	5501	蛋白磷酸酶 1, 催化性亚基, γ 亚型
GMNN	51053	亭蛋白, DNA 复制抑制剂
EPHB2	2048	EPH 受体 B2
RAD17	5884	RAD17 同源物 (粟酒裂殖酵母 (<i>S. pombe</i>))
AURKA	6790	Aurora 激酶 A
NEK1	4750	NIMA (绝不在有丝分裂基因 a 中)-相关的激酶 1
RASSF1	11186	Ras 关联 (RalGDS/AF-6) 结构域家族 1
VASH1	22846	Vasohibin 1
MAPRE3	22924	微管-关联蛋白, RP/EB 家族, 成员 3
CDC48	55143	细胞分裂周期关联的 8
CDC73	79577	细胞分裂周期 73, Paf1/RNA 聚合酶 II 复合体组分, 同源物 (酿酒酵母 (<i>S. cerevisiae</i>))
SIRT2	22933	Sirtuin (沉默交配类型信息调节 2 同源物) 2 (酿酒酵母 (<i>S. cerevisiae</i>))
MAPK7	5598	促细胞分裂原-活化的蛋白激酶 7
MKI67	4288	抗原鉴定的通过单克隆抗体 Ki-67
TFDP1	7027	转录因子 Dp-1
DMBT1	1755	恶性脑肿瘤中删除的 1
NRC-2(免疫)		
C7	730	补体组分 7
SELE	6401	选择素 E (内皮粘附分子 1)
CD27	939	CD27 分子
F3	2152	凝固因子 III (凝血致活酶, 组织因子)
IL23A	51561	白细胞介素 23, α 亚基 p19
CARTPT	9607	CART 前原肽
SPPI	6696	分泌的磷蛋白 1 (骨桥蛋白, 骨涎蛋白 I, 早期 T-淋巴细胞活化 1)
TNNT1	7138	肌钙蛋白 T 1 型 (骨骼, 慢)
CACNB3	784	钙通道, 电压-依赖性, β 3 亚基
C6	729	补体组分 6
F13B	2165	凝固因子 XIII, B 多肽
SELP	6403	选择素 P (颗粒粘附蛋白 140kDa, 抗原 CD62)
POU2F2	5452	POU 2 类同源框 2
STAT3	6774	转录 3 的信号转导物和活化物 (急性-期应答因子)
SERPINA1	5265	丝氨酸蛋白酶抑制剂肽酶抑制剂, 分化体 A (α -1 抗蛋白酶, 抗胰蛋白酶), 成员 1
FGF23	8074	成纤维细胞生长因子 23
MYBPC3	4607	肌球蛋白结合蛋白 C, 心脏
LST1	7940	白细胞特异性转录物 1
LEP	3952	瘦素 (肥胖症同源物, 小鼠)
STAT5A	6776	转录的信号转导物和活化物 5A
AMBP	259	α -1-微球蛋白/尿抑胰酶素前体
TNNC2	7125	肌钙蛋白 C 2 型 (快)
SCN5A	6331	钠通道, 电压-选通的, V 型, α 亚基
CAV1	857	小窝蛋白 1, 细胞质膜微囊蛋白, 22kDa
RBM4	5936	RNA 结合基序蛋白 4

[0159]

BLM	641	Bloom 综合征
FYN	2534	相关于 SRC, FGR, YES 的 FYN 癌基因
BCL6	604	B 细胞 CLL/淋巴瘤 6 (锌指蛋白 51)
NMU	10874	神经介素 U
HP	3240	结合珠蛋白
NRC-3 (凋亡)		
ZBTB16	7704	锌指和 BTB 结构域含有 16
ARHGEF6	9459	Rac/Cdc42 鸟嘌呤核苷酸交换因子 (GEF) 6
PHLDA2	7262	普列克底物蛋白同源性-样结构域, 家族 A, 成员 2
TNFRSF11B	4982	肿瘤坏死因子受体超家族, 成员 11b (骨保护素)
CYCS	54205	细胞色素 c, 体细胞
TRADD	8717	TNFRSF1A-关联的经死亡结构域
BIRC5	332	含有杆状病毒 IAP 重复子的 5 (存活素)
PDCD4	27250	程序性细胞死亡 4 (肿瘤形成转化抑制剂)
SOCS2	8835	细胞因子信号转导抑制剂 2
PPP2R1B	5519	蛋白磷酸酶 2 (以前 2A), 调控亚基 A, β 亚型
MGMT	4255	O-6-甲基鸟嘌呤-DNA 甲基转移酶
IKBKKG	8517	B 细胞中的 κ 轻多肽基因增强子的抑制剂, 激酶 γ
BTG1	694	B 细胞转位基因 1, 抗-增殖
NRAS	4893	成神经细胞瘤 RAS 病毒 (v-ras) 癌基因同源物
ESPL1	9700	额外纺锤体极体同源物 1 (酿酒酵母 (<i>S. cerevisiae</i>))
CDC2	983	细胞分裂周期 2, G1 至 S 和 G2 至 M
APLP1	333	淀粉样蛋白 β (A4) 前体-样蛋白 1
TCTN3	26123	Tectonic 家族成员 3
NME1	4830	非-转移性细胞 1, 蛋白 (NM23A) 表达于
STAT5A	6776	转录的信号转导物和活化物 5A
CLU	1191	聚集素
BCL2	596	B 细胞 CLL/淋巴瘤 2
HTATIP2	10553	HIV-1 Tat 相互作用蛋白 2, 30kDa
EEF1A2	1917	真核翻译延伸因子 1 α 2
INHA	3623	抑制素, α
TNFSF9	8744	肿瘤坏死因子 (配体) 超家族, 成员 9
LRDD	55367	亮氨酸-富重复及死亡结构域含有
FADD	8772	Fas (TNFRSF6)-关联的经死亡结构域
IL19	29949	白细胞介素 19
KIAA0367	23273	
NRC-4 (细胞粘接)		
CHL1	10752	与 L1CAM 具有同源性的细胞粘接分子 (L1 的近同源物)
COL15A1	1306	胶原, XV 型, α 1
CRNN	49860	Cornulin
KAL1	3730	Kallmann 综合征 1 序列
SOX9	6662	SRY (性别决定区 Y)-盒 9 (躯干发育异常, 常染色体性反转)
PTPRF	5792	蛋白酪氨酸磷酸酶, 受体类型, F
ITGA7	3679	整联蛋白, α 7
MFAP4	4239	微原纤维-关联蛋白 4
EDG1	1901	内皮分化, 鞘脂 G-蛋白-偶联的受体, 1
ZEB2	9839	锌指 E-盒结合同源框 2
PDZD2	23037	PDZ 结构域含有 2
ROBO1	6091	Roundabout, 轴突引导受体, 同源物 1 (果蝇属 (<i>Drosophila</i>))
FBN2	2201	原纤维蛋白 2 (先天性挛缩性蜘蛛指)

[0160]

POSTN	10631	骨膜素, 成骨细胞特定因子
CDH5	1003	钙粘蛋白 5, 2 型, VE-钙粘蛋白(血管上皮)
PKD1	5310	多囊性肾疾病 1 (常染色体显性)
TGFB1I1	7041	转化生长因子 β 1 诱导的转录物 1
ITGA5	3678	整联蛋白, α 5 (纤维蛋白受体, α 多肽)
RASA1	5921	RAS p21 蛋白活化物 (GTP 酶活化蛋白) 1
COL11A2	1302	胶原, XI 型, α 2
VEZT	55591	Vezatin, 粘着连接跨膜蛋白
CLDN4	1364	密蛋白 4
BCL6	604	B 细胞 CLL/淋巴瘤 6 (锌指蛋白 51)
AMIGO2	347902	具有 Ig-样结构域的粘接分子 2
ECM2	1842	细胞外基质蛋白 2, 雌性器官和脂细胞特定
FAF1	11124	Fas (TNFRSF6) 关联的因子 1
ITGB8	3696	整联蛋白, β 8
PRPH2	5961	外周蛋白 2 (视黄醛变性, 慢)
CEACAM1	634	癌胚抗原-相关的细胞粘接分子 1 (胆道糖蛋白)
THY1	7070	Thy-1 细胞表面抗原
NRC_5 (细胞周期)		
NDN	4692	抑蛋白同源物(小鼠)
CDCA8	55143	细胞分裂周期关联的 8
CHEK2	11200	CHK2 关卡同源物 (粟酒裂殖酵母 (<i>S. pombe</i>))
CDC45L	8318	CDC45 细胞分裂循环 45-样 (酿酒酵母 (<i>S. cerevisiae</i>))
STRN3	29966	隆纹菌素, 钙调素结合蛋白 3
PYCARD	29108	PYD 和 CARD 结构域含有
HERC5	51191	Heet 结构域和 RLD 5
MN1	4330	脑膜瘤 (在平衡的转位中破裂) 1
XRCC2	7516	中国仓鼠细胞中的 X-射线修复互补缺陷修复 2
NOLC1	9221	核仁和螺旋小体磷蛋白 1
CHFR	55743	具有叉头及环指结构域的关卡
NHP2L1	4809	NHP2 非-组蛋白染色体蛋白 2-样 1 (酿酒酵母 (<i>S. cerevisiae</i>))
MCM7	4176	微型染色体维持复合体组分 7
PIM2	11040	Pim-2 癌基因
INHBA	3624	抑制素, β A
ACPP	55	酸磷酸酶, 前列腺
CETN3	1070	中心粒蛋白, EF-手蛋白, 3 (CDC31 同源物, 酵母)
MIS12	79003	MIS12, MIND 着丝粒复合体组分, 同源物(酵母)
PCAF	8850	P300/CBP-关联因子
PTMA	5757	前胸腺素, α (基因序列 28)
AXL	558	AXL 受体酪氨酸激酶
Sep-11	55752	隔蛋白 11
LTBP2	4053	潜伏转化生长因子 β 结合蛋白 2
SUPT5H	6829	Ty 5 同源物抑制剂 (酿酒酵母 (<i>S. cerevisiae</i>))
TOB2	10766	ERBB2 转导物, 2
CDK5R1	8851	细胞周期蛋白-依赖性激酶 5, 调控亚基 1 (p35)
ILF3	3609	白细胞介素增强子结合因子 3, 90kDa
POLD1	5424	聚合酶 (DNA 导向的), δ 1, 催化性亚基 125kDa
GADD45B	4616	生长停止及 DNA-损伤-诱导性, β
CDT1	81620	染色质许可及 DNA 复制因子 1
NRC_6 (细胞能动性)		
KAL1	3730	Kallmann 综合征 1 序列

[0161]

PRSS3	5646	蛋白酶, 丝氨酸, 3 (mesotrypsin)
CHL1	10752	与 L1CAM 具有同源性的细胞粘接分子(L1 的近同源物)
ROBO1	6091	Roundabout, 轴突引导受体, 同源物 1 (果蝇属 (<i>Drosophila</i>))
ZEB2	9839	锌指 E-盒结合同源框 2
EDG1	1901	内皮分化, 鞘脂 G-蛋白-偶联的受体, 1
CDA	978	胞苷脱氨酶
ATP1A3	478	ATP 酶, Na ⁺ /K ⁺ 运输, α 3 多肽
IGFBP7	3490	胰岛素-样生长因子结合蛋白 7
INHBA	3624	抑制素, β A
CSPG4	1464	硫酸软骨素蛋白聚糖 4
WFDC1	58189	WAP 4-二硫化物核心结构域 1
PF4	5196	血小板因子 4 (趋化因子 (C-X-C 基序) 配体 4)
ALOX12	239	花生四烯酸 12-脂氧合酶
NDN	4692	抑蛋白同源物(小鼠)
CCDC88A	55704	卷曲的-线圈结构域含有 88A
CEACAM1	634	癌胚抗原-相关的细胞粘接分子 1 (胆道糖蛋白)
ARPC3	10094	肌动蛋白相关的蛋白 2/3 复合物, 亚基 3, 21kDa
BCL6	604	B 细胞 CLL/淋巴瘤 6 (锌指蛋白 51)
PPAP2B	8613	磷脂酸性磷酸酶类型 2B
LAMB1	3912	层粘连蛋白, β 1
DNAH2	146754	动力蛋白, 轴丝, 重链 2
SLIT3	6586	Slit 同源物 3 (果蝇属 (<i>Drosophila</i>))
CDK5R1	8851	细胞周期蛋白-依赖性激酶 5, 调控亚基 1 (p35)
ADRA2A	150	肾上腺素能, α -2A-, 受体
AMOT	154796	Angiomotin
ACTG1	71	肌动蛋白, γ 1
TGFB3	7043	转化生长因子, β 3
KDR	3791	激酶插入结构域受体(III 型受体酪氨酸激酶)
ABI3	51225	ABI 基因家族, 成员 3
NRC-7 (凋亡)		
CDH13	1012	钙粘蛋白 13, H-钙粘蛋白(心脏)
SLAMF7	57823	SLAM 家族成员 7
ANGPTL4	51129	血管生成素-样 4
SULF1	23213	硫酸酯酶 1
GJA1	2697	缝隙连接蛋白, α 1, 43kDa
MUC2	4583	粘蛋白 2, 寡聚体粘液/凝胶-形成
INPP5D	3635	聚磷酸肌醇-5-磷酸酶, 145kDa
BCL2L14	79370	BCL2-样 14 (凋亡辅助剂)
CASP8AP2	9994	CASP8 关联的蛋白 2
PTK2B	2185	PTK2B 蛋白酪氨酸激酶 2 β
LIG4	3981	连接酶 IV, DNA, ATP-依赖性
GML	2765	GPI 锚定的分子样蛋白
PDCD4	27250	程序性细胞死亡 4 (肿瘤形成转化抑制剂)
MAGEH1	28986	黑色素瘤抗原家族 H, 1
FAS	355	Fas (TNF 受体超家族, 成员 6)
ANXA5	308	膜联蛋白 A5
GRM4	2914	谷氨酸受体, 亲代谢型 4
AVEN	57099	凋亡, 胱天蛋白酶活化抑制剂
CASP9	842	胱天蛋白酶 9, 凋亡-相关的半胱氨酸蛋白酶
CRYAA	1409	眼晶体蛋白, α A

[0162]

NFKBIA	4792	B 细胞内的 κ 轻多肽基因增强子的核因子抑制剂, α
STK3	6788	丝氨酸/苏氨酸激酶 3 (STE20 同源物, 酵母)
PPP2CB	5516	蛋白磷酸酶 2 (以前 2A), 催化性亚基, β 亚型
CIAPIN1	57019	细胞因子诱导的凋亡抑制剂 1
PEA15	8682	星形胶质细胞中富集的磷蛋白 15
TGFB2	7042	转化生长因子, β 2
OLFR@	4972	嗅觉受体簇
MGC29506	51237	假想蛋白 MGC29506
CD74	972	CD74 分子, 主要组织相容性复合物, II 类不变链
TRAF6	7189	TNF 受体-关联因子 6
NRC-8 (细胞粘接)		
SLAMF7	57823	SLAM 家族成员 7
CDH13	1012	钙粘蛋白 13, H-钙粘蛋白(心脏)
IGSF1	3547	免疫球蛋白超家族, 成员 1
TGFB1	7045	转化生长因子, β -诱导的, 68kDa
HAPLN1	1404	透明质酸糖胺多糖和蛋白聚糖连接蛋白 1
FRAS1	80144	Fraser 综合征 1
PLEKHC1	10979	普列克底物蛋白同源性结构域含有, 家族 C (有 FERM 结构域) 成员 1
CD226	10666	CD226 分子
SUSD5	26032	苏希结构域含有 5
CELSR1	9620	钙粘蛋白, EGF LAG 7-经过 G-类型受体 1(火烈鸟同源物, 果蝇属 (<i>Drosophila</i>))
GRLF1	2909	糖皮质类固醇受体 DNA 结合因子 1
NID2	22795	巢蛋白 2 (骨巢蛋白)
DDR1	780	网柄菌凝素结构域受体家族, 成员 1
NINJ2	4815	Ninjurin 2
DCHS2	54798	Dachsous 2 (果蝇属 (<i>Drosophila</i>))
ITGAM	3684	整联蛋白, α M (补体组分 3 受体 3 亚基)
SCARB2	950	清除剂受体 B 类, 成员 2
CYR61	3491	半胱氨酸-富, 血管发生诱导物, 61
PVRL2	5819	脊髓灰质炎病毒受体-相关的 2 (疱疹病毒进入介质 B)
PTK2B	2185	PTK2B 蛋白酪氨酸激酶 2 β
SELPLG	6404	选择素 P 配体
GP1BA	2811	糖蛋白 Ib (血小板), α 多肽
VCL	7414	纽蛋白
CXCR3	2833	趋化因子 (C-X-C 基序) 受体 3
WFDC1	58189	WAP 4-二硫化物核心结构域 1
DLG1	1739	Discs, 大同源物 1 (果蝇属 (<i>Drosophila</i>))
ENTPD1	953	外核苷酸三磷酸二磷酸水解酶 1
CTNNA3	29119	连环蛋白 (钙粘蛋白-关联蛋白), α 3
PPFIA1	8500	蛋白酪氨酸磷酸酶, 受体类型, f 多肽 (PTPRF), 相互作用蛋白 (LIP 相关蛋白), α 1
NF2	4771	神经纤维瘤蛋白 2 (双侧听神经瘤)
NRC-9 (细胞生长)		
WFDC1	58189	WAP 4-二硫化物核心结构域 1
CDH13	1012	钙粘蛋白 13, H-钙粘蛋白(心脏)
ETV4	2118	Ets 变体基因 4 (E1A 增强子结合蛋白, E1AF)
DDR1	780	网柄菌凝素结构域受体家族, 成员 1
PLEKHC1	10979	普列克底物蛋白同源性结构域含有, 家族 C (有 FERM 结构域) 成员 1

[0163]

SELPLG	6404	选择素 P 配体
CYR61	3491	半胱氨酸-富, 血管发生诱导物, 61
TKT	7086	转酮酶(Wernicke-Korsakoff 综合征)
VAX2	25806	腹前同源框 2
RAI1	10743	视黄酸诱导的 1
SEMA6A	57556	Sema 结构域, 跨膜结构域(TM), 及细胞质结构域, (semaphorin) 6A
DLG1	1739	Discs, 大同源物 1 (果蝇属 (Drosophila))
BTG1	694	B 细胞转位基因 1, 抗-增殖
PTCH1	5727	Patched 同源物 1 (果蝇属 (Drosophila))
FGF20	26281	成纤维细胞生长因子 20
OGFR	11054	类阿片生长因子受体
NINJ2	4815	Ninjurin 2
MORF4L2	9643	死亡率因子 4 如 2
VCL	7414	纽蛋白
ESR2	2100	雌激素受体 2 (ER β)
OPHN1	4983	Oligophrenin 1
NTRK3	4916	神经营养酪氨酸激酶, 受体, 类型 3
CDKN2C	1031	细胞周期蛋白-依赖性激酶抑制剂 2C (p18, 抑制 CDK4)
CDK5R1	8851	细胞周期蛋白-依赖性激酶 5, 调控亚基 1 (p35)
TOP2B	7155	拓扑异构酶 (DNA) II β 180kDa
PPT1	5538	棕榈酰-蛋白硫酯酶 1 (蜡样脂-脂褐质沉积症, 神经元 1, 婴儿)
GDF2	2658	生长分化因子 2
GFRA3	2676	GDNF 家族受体 $\alpha 3$
GP1BA	2811	糖蛋白 Ib (血小板), α 多肽
PPP2CB	5516	蛋白磷酸酶 2 (以前 2A), 催化性亚基, β 亚型

[0164] 表 2.3 个测试数据组中标记物组的有效性的表现

[0165]

ER+样品			
组	测试组 1 (173 个样品)*	测试组 2 (74 个样品)	测试组 3 (201 个样品)
低-风险	N=99, R=57.2%, R1=93.9%	N=22, R=29.7%, R1=90.9%	N=87, R=43.3%, R1=86.8%
中等	N=34, R=19.6%, R1=82.4%	N=52, R=70.3%, R1=79.7%	N=78, R=38.8%, R1=69.2%
高-风险	N=40, R=23.1%, R2=42.5%	---	N=36, R=17.9%, R2=33.3%
ER-样品			
组	测试组 1 (46 个样品)*	测试组 2 (43 个样品)	测试组 3 (31 个样品)
低-风险	N=9, R=19.6%, R1=100%	N=13, R=30.2%, R1=92.3%	N=14, R=45.2%, R1=100%
高-风险	N=37, R=80.4%, R2=51.4%	N=30, R=69.8%, R2=40%	N=17, R=54.8%, R2=35.3%

[0166] 注:* 原测试组 1 中有 295 个样品。但是, 其包括 76 个样品, 其来自 van' t Veer 等人, Nature, 415:530, 2002。因为我们使用 van' t Veer 数据组 (van' t Veer 等人, Nature, 415:530, 2002) 作为练习组, 然后我们自 295 个样品去除这 76 个样品。因此, 测试组 1 含有 219 个样品。

[0167] 1. N 代表样品数

[0168] 2. R 代表组中样品数与测试组的总样品数的比

[0169] 3. R1 代表具有非 - 复发的样品的百分率 (精确度)

[0170] 4. R2 代表具有复发的样品的百分率 (精确度)

[0171] 5. 测试组 1 来自 Chang 等人, PNAS, 2005

[0172] 6. 测试组 2 来自 Koe 等人, Cancer Cell, 2006

[0173] 7. 测试组 3 来自 Sotiriou 等人, J. Natl Cancer Inst, 98:262, 2006

[0174] 表 3. 标记物组的组合利用和用于预测低 - 风险组样品的各个体标记物组的比较

[0175]

标记物组	精确度 (在低 - 风险组中)
	测试组 1 (173 个样品)
NRC-1	92.80%
NRC-2	91.80%
NRC-3	92.20%
NRC-1, 2, 3	94%
	测试组 2 (74 个样品)
NRC-1	86.80%
NRC-2	88.90%
NRC-3	78.30%
NRC-1, 2, 3	91%
	测试组 3 (201 个样品)
NRC-1	83.10%
NRC-2	80.50%
NRC-3	79.50%
NRC-1, 2, 3	87%
ER- 样品	
	测试组 1 (46 个样品)*
NRC-7	76%
NRC-8	72.70%
NRC-9	56.50%

NRC-7, 8, 9	100%
	测试组 2(43 个样品)
NRC-7	85%
NRC-8	84. 20%
NRC-9	73. 10%
NRC-7, 8, 9	92. 30%
	测试组 3(31 个样品)
NRC-7	91%
NRC-8	100%
NRC-9	86. 40%
NRC-7, 8, 9	100%

[0176] 注意：使用的数据组与表 2 中的那些相同。

[0177] 【表 4】癌列表

[0178] 急性成淋巴细胞白血病，成年

[0179] 急性成淋巴细胞白血病，儿童

[0180] 急性骨髓样白血病，成年

[0181] 急性骨髓样白血病，儿童

[0182] 肾上腺皮质癌

[0183] 肾上腺皮质癌，儿童

[0184] AIDS- 相关的癌

[0185] AIDS- 相关的淋巴瘤

[0186] 肛门癌

[0187] 附录癌

[0188] 星形细胞瘤，儿童)

[0189] 非典型畸胎样 / 杆状肿瘤，儿童，中枢神经系统

[0190] 基础细胞癌，见皮肤癌（非黑色素瘤）

[0191] 胆管癌，肝外

[0192] 膀胱癌

[0193] 膀胱癌，儿童

[0194] 骨癌，骨肉瘤和恶性纤维状组织细胞瘤

[0195] 脑茎神经胶质瘤，儿童

[0196] 脑肿瘤，成年

- [0197] 脑肿瘤,脑茎神经胶质瘤,儿童
- [0198] 脑肿瘤,中枢神经系统非典型畸胎样 / 杆状肿瘤,儿童
- [0199] 脑肿瘤,中枢神经系统胚性肿瘤,儿童
- [0200] 脑肿瘤,颅咽管瘤,儿童
- [0201] 脑肿瘤,成室管膜细胞瘤,儿童
- [0202] 脑肿瘤,室管膜瘤,儿童
- [0203] 脑肿瘤,成髓细胞瘤,儿童
- [0204] 脑肿瘤,髓上皮瘤,儿童
- [0205] 脑肿瘤,中等分化的松果体实质肿瘤,儿童)
- [0206] 脑肿瘤,幕上原始神经外胚层肿瘤和成松果体细胞瘤,儿童
- [0207] 脑和脊髓肿瘤,儿童(其他)
- [0208] 乳腺癌
- [0209] 乳腺癌和怀孕
- [0210] 乳腺癌,儿童
- [0211] 乳腺癌,雄性
- [0212] 支气管肿瘤,儿童
- [0213] 伯基特淋巴瘤
- [0214] 类癌肿瘤,儿童
- [0215] 类癌肿瘤,消化道
- [0216] 未知的原发性癌
- [0217] 中枢神经系统非典型畸胎样 / 杆状肿瘤,儿童
- [0218] 中枢神经系统胚性肿瘤,儿童
- [0219] 中枢神经系统淋巴瘤,原发性
- [0220] 子宫颈癌
- [0221] 子宫颈癌,儿童
- [0222] 儿童癌
- [0223] 脊索瘤,儿童
- [0224] 慢性淋巴细胞白血病
- [0225] 慢性髓性白血病
- [0226] 慢性骨髓增生障碍
- [0227] 结肠癌
- [0228] 结直肠癌,儿童
- [0229] 颅咽管瘤,儿童
- [0230] 皮肤 T- 细胞淋巴瘤,见蕈样真菌病和塞扎里综合征
- [0231] 胚性肿瘤,中枢神经系统,儿童
- [0232] 子宫内膜癌
- [0233] 成室管膜细胞瘤,儿童
- [0234] 室管膜瘤,儿童
- [0235] 食管癌

- [0236] 食管癌, 儿童
- [0237] 肿瘤的尤因肉瘤家族
- [0238] 颅外生殖细胞肿瘤, 儿童
- [0239] 性腺外生殖细胞肿瘤
- [0240] 肝外胆管癌
- [0241] 眼癌, 眼球内黑色素瘤
- [0242] 眼癌, 成视网膜细胞瘤
- [0243] 胆囊癌
- [0244] 胃癌
- [0245] 胃癌, 儿童
- [0246] 消化道类癌肿瘤
- [0247] 消化道间质肿瘤 (GIST)
- [0248] 消化道间质细胞肿瘤, 儿童
- [0249] 生殖细胞肿瘤, 颅外, 儿童
- [0250] 生殖细胞肿瘤, 性腺外
- [0251] 生殖细胞肿瘤, 卵巢
- [0252] 妊娠滋养层肿瘤
- [0253] 神经胶质瘤, 成年
- [0254] 神经胶质瘤, 儿童脑茎
- [0255] 多毛细胞白血病
- [0256] 头和颈癌
- [0257] 肝细胞 (肝) 癌, 成年 (原发性)
- [0258] 肝细胞 (肝) 癌, 儿童 (原发性)
- [0259] 组织细胞增多症, 朗格汉斯细胞
- [0260] 霍奇金淋巴瘤, 成年
- [0261] 霍奇金淋巴瘤, 儿童
- [0262] 咽下癌
- [0263] 眼球内黑色素瘤
- [0264] 岛细胞肿瘤 (内分泌腺)
- [0265] 卡波西肉瘤
- [0266] 肾 (肾细胞) 癌
- [0267] 肾癌, 儿童
- [0268] 朗格汉斯细胞组织细胞增多症
- [0269] 喉癌
- [0270] 喉癌, 儿童
- [0271] 白血病, 急性成淋巴细胞, 成年
- [0272] 白血病, 急性成淋巴细胞, 儿童
- [0273] 白血病, 急性骨髓样, 成年
- [0274] 白血病, 急性骨髓样, 儿童

- [0275] 白血病,慢性淋巴细胞
- [0276] 白血病,慢性髓性
- [0277] 白血病,多毛细胞
- [0278] 唇及口服腔癌
- [0279] 肝癌,成年(原发性)
- [0280] 肝癌,儿童(原发性)
- [0281] 肺癌,非-小细胞
- [0282] 肺癌,小细胞
- [0283] 淋巴瘤,AIDS-相关的
- [0284] 淋巴瘤,伯基特
- [0285] 淋巴瘤,皮肤 T- 细胞,见蕈样真菌病和塞扎里综合征
- [0286] 淋巴瘤,霍奇金,成年
- [0287] 淋巴瘤,霍奇金,儿童
- [0288] 淋巴瘤,非-霍奇金,成年
- [0289] 淋巴瘤,非-霍奇金,儿童
- [0290] 淋巴瘤,原发性中枢神经系统
- [0291] 巨球蛋白血症,瓦尔登斯特伦
- [0292] 骨的恶性纤维状组织细胞瘤和骨肉瘤
- [0293] 成髓细胞瘤,儿童
- [0294] 髓上皮瘤,儿童
- [0295] 黑色素瘤
- [0296] 黑色素瘤,眼球内(眼)
- [0297] 默克尔细胞癌
- [0298] 间皮瘤,成年恶性
- [0299] 间皮瘤,儿童
- [0300] 隐性原发性转移性鳞状颈癌
- [0301] 口癌
- [0302] 多内分泌瘤形成综合征,儿童
- [0303] 多骨髓瘤/血浆细胞肿瘤
- [0304] 蕈样真菌病
- [0305] 骨髓增生异常综合征
- [0306] 骨髓增生异常/骨髓增生肿瘤
- [0307] 髓性白血病,慢性
- [0308] 骨髓样白血病,成年急性
- [0309] 骨髓样白血病,儿童急性
- [0310] 骨髓瘤,多
- [0311] 骨髓增生障碍,慢性
- [0312] 鼻腔和鼻旁窦癌
- [0313] 鼻咽癌

- [0314] 鼻咽癌, 儿童
- [0315] 成神经细胞瘤
- [0316] 非 - 霍奇金淋巴瘤, 成年
- [0317] 非 - 霍奇金淋巴瘤, 儿童
- [0318] 非 - 小细胞肺癌
- [0319] 口服癌, 儿童
- [0320] 口服腔癌, 唇及
- [0321] 口咽癌
- [0322] 骨肉瘤和骨的恶性纤维状组织细胞瘤
- [0323] 卵巢癌, 儿童
- [0324] 卵巢上皮癌
- [0325] 卵巢生殖细胞肿瘤
- [0326] 卵巢低恶性潜在肿瘤
- [0327] 胰癌
- [0328] 胰癌, 儿童
- [0329] 胰癌, 岛细胞肿瘤
- [0330] 乳头状瘤病, 儿童
- [0331] 鼻旁窦和鼻腔癌
- [0332] 甲状旁腺癌
- [0333] 阴茎癌
- [0334] 咽癌
- [0335] 中等分化的松果体实质肿瘤, 儿童
- [0336] 成松果体细胞瘤及幕上原始神经外胚层肿瘤, 儿童
- [0337] 垂体肿瘤
- [0338] 血浆细胞肿瘤 / 多骨髓瘤
- [0339] 胸膜肺母细胞瘤
- [0340] 怀孕及乳腺癌
- [0341] 原发性中枢神经系统淋巴瘤
- [0342] 前列腺癌
- [0343] 直肠癌
- [0344] 肾细胞 (肾) 癌
- [0345] 肾细胞 (肾) 癌, 儿童
- [0346] 肾盂和输尿管, 移行细胞癌
- [0347] 涉及染色体 15 上的 NUT 基因的呼吸道癌
- [0348] 成视网膜细胞瘤
- [0349] 横纹肌肉瘤, 儿童
- [0350] 唾液腺癌
- [0351] 唾液腺癌, 儿童
- [0352] 肉瘤, 肿瘤的尤因肉瘤家族

- [0353] 肉瘤,卡波西
- [0354] 肉瘤,软组织,成年
- [0355] 肉瘤,软组织,儿童
- [0356] 肉瘤,子宫
- [0357] 塞扎里综合征
- [0358] 皮肤癌(非黑色素瘤)
- [0359] 皮肤癌,儿童
- [0360] 皮肤癌(黑色素瘤)
- [0361] 皮肤癌,默克尔细胞
- [0362] 小细胞肺癌
- [0363] 小肠癌
- [0364] 软组织肉瘤,成年
- [0365] 软组织肉瘤,儿童
- [0366] 鳞状细胞癌,见皮肤癌(非黑色素瘤)
- [0367] 隐性原发性鳞状颈癌,转移性
- [0368] 胃癌
- [0369] 胃癌,儿童
- [0370] 幕上原始神经外胚层肿瘤,儿童
- [0371] T-细胞淋巴瘤,皮肤,
- [0372] 睾丸癌
- [0373] 喉癌
- [0374] 胸腺瘤及胸腺癌
- [0375] 胸腺瘤及胸腺癌,儿童
- [0376] 甲状腺癌
- [0377] 甲状腺癌,儿童
- [0378] 肾盂和输尿管的移行细胞癌
- [0379] 滋养层肿瘤,妊娠
- [0380] 输尿管及肾盂,移行细胞癌
- [0381] 尿道癌
- [0382] 子宫癌,子宫内膜
- [0383] 子宫肉瘤
- [0384] 阴道癌
- [0385] 阴道癌,儿童
- [0386] 外阴癌
- [0387] 瓦尔登斯特伦巨球蛋白血症
- [0388] 维尔姆斯肿瘤
- [0389] 本说明书还包括下列内容:
- [0390] 1. 用于鉴定肿瘤特征的方法,所述方法包括以下步骤:
- [0391] (1) 获得各预测目标特征的 3 个不同标记物组;

- [0392] (2) 自肿瘤细胞获得样品基因表达信号；
- [0393] (3) 添加报告子以影响样品中的变化,用于允许评定肿瘤中的目标基因表达信号；
- [0394] (4) 将基因表达信号与报告子组合；
- [0395] (5) 将提取的基因表达信号与 3 个不同标记物组关联；
- [0396] (6) 根据以下顺序将指定分配给提取的基因表达信号：
- [0397] (a) 如果全部 3 个预测性基因表达信号组的关联预测其具有关注的特征,将其指定为恶性肿瘤；
- [0398] (b) 如果全部 3 个预测性基因表达信号组的关联预测其缺乏关注的特征,将其指定为良性肿瘤；
- [0399] (c) 如果全部 3 个预测性基因表达信号组的关联不提供相同的预测的临床结果,将所述肿瘤指定为“中等”；
- [0400] (7) 输出所述指定。
- [0401] 2. 实施方式 1 的方法,其中关注的特征涉及下列之一或多种:转移,发炎,细胞周期,免疫学应答基因,药物抗性基因,及多-药物抗性基因。
- [0402] 3. 实施方式 1 的方法,其中肿瘤特征是导致差患者手术后存活的趋势。
- [0403] 4. 实施方式 3 的方法,其中步骤 4 包括根据以下顺序将值分配给提取的基因表达信号：
- [0404] (a) 如果全部 3 个预测性基因表达信号组的关联预测其为恶性肿瘤,将其指定为恶性肿瘤,并将建议超过一般标准的护理的更攻击性的治疗；
- [0405] (b) 如果全部 3 个预测性基因表达信号组的关联预测其为良性肿瘤,不建议超过护理标准的治疗,且不建议手术后化学治疗或放射治疗；
- [0406] (c) 如果全部 3 个预测性基因表达信号组的关联不提供相同的预后,将所述肿瘤指定为“中等”,并建议完全一般标准的护理治疗,包括化学治疗和 / 或放射治疗。
- [0407] 5. 实施方式 1 的方法,其在步骤 1 之前包括预先步骤：
- [0408] (a) 鉴定待检查的肿瘤亚型
- [0409] (b) 选择特异于所述肿瘤亚型的标记物组。
- [0410] 6. 用于测定实施方式 1 中使用的类型的预测性基因表达信号组的方法,包括以下步骤：
- [0411] (1) 获得对于目标癌的已知的肿瘤群的目标特征的基因表达信号信息和患者临床信息；
- [0412] (2) 将基因表达信号与有关目标特征的临床患者信息关联,以鉴定哪些基因具有对于临床结果的预测力；
- [0413] (3) 自鉴定的基因表达信号创建至少 30 个随机练习数据组；
- [0414] (4) 比较步骤 1 的鉴定的基因表达信号与一系列已知的在癌中活跃的基因；
- [0415] (5) 选择对应于已知的癌基因列表上的那些的鉴定的基因表达信号；
- [0416] (6) 根据它们在生物学过程中的作用将选择的鉴定的基因表达信号分组；
- [0417] (7) 自步骤 6 的选择的基因表达信号组产生至少 25 个基因的随机基因表达信号组；

- [0418] (8) 将所述随机基因表达信号组与步骤 3 中获得的随机练习数据组关联；
- [0419] (9) 自对于步骤 7 的各基因表达信号组的关联获得用于存活筛选的 P 值；
- [0420] (10) 如果对于多于 90 % 的随机练习数据组,用于基因表达信号组的 P 值小于 0.05,则保持所述基因表达信号组；
- [0421] (11) 基于组中基因出现的频度排序步骤 10 中保持的随机基因表达信号组；
- [0422] (12) 选择排在前面的至少 26 个基因作为潜在候选标记物；
- [0423] (13) 重复步骤 7 ~ 12,并产生至少 26 个基因的另一,独立,排序组；
- [0424] (14) 比较自步骤 12 及步骤 13 的排在前面的基因；
- [0425] (15) 如果多于 25 个基因相同,将所述排在前面的基因保持为标记物组；
- [0426] (16) 重复两次步骤 7 ~ 15,以获得 3 个不同标记物组；
- [0427] (17) 输出所述 3 个不同标记物组。
- [0428] 7. 实施方式 6 的方法,其中根据它们在生物学过程中的作用的选择的鉴定的基因表达信号的分组使用基因本体论分析进行。
- [0429] 8. 实施方式 6 的方法,其中在步骤 3 中,创建 30 和 50 之间个随机练习组。
- [0430] 9. 实施方式 8 的方法,其中创建 30 和 40 之间个练习组。
- [0431] 10. 实施方式 6 的方法,其中在步骤 4 中,已知在癌中活跃的基因选自负责转移,细胞增殖,肿瘤血管化及药物应答的基因。
- [0432] 11. 实施方式 6 的方法,其中在步骤 7 中,产生约 750,000 和 1,250,000 之间个随机基因表达信号组。
- [0433] 12. 实施方式 6 的方法,其中在步骤 7 中,产生约 900,000 和 1,100,000 之间个随机基因表达信号组。
- [0434] 13. 实施方式 6 的方法,其中在步骤 7 中,产生约 1,000,000 随机基因表达信号组。
- [0435] 14. 实施方式 6 的方法,其中在步骤 7 中,产生的随机基因表达信号组含有约 25 和 50 之间个基因。
- [0436] 15. 实施方式 6 的方法,其中在步骤 7 中,产生的随机基因表达信号组含有约 28 和 32 之间个基因。
- [0437] 16. 实施方式 6 的方法,其中在步骤 12 中,选择排在前面的 26 ~ 50 个基因。
- [0438] 17. 实施方式 6 的方法,其中在步骤 12 中,选择排在前面的 28 ~ 32 个基因。
- [0439] 18. 实施方式 1 的方法,其中肿瘤是哺乳动物肿瘤。
- [0440] 19. 实施方式 18 的方法,其中肿瘤是下列之一的肿瘤:人,猿,猫,狗,猪,牛,绵羊,山羊,兔,小鼠,大鼠,豚鼠,仓鼠或沙鼠。
- [0441] 20. 实施方式 4 的方法,其中至少一种癌生物标记物组选自基本上由下列构成的列表:NRC-1, NRC-2, NRC-3, NRC-4, NRC-5, NRC-6, NRC-7, NRC-8 和 NRC-9。
- [0442] 21. 试剂盒,其包含至少 3 个标记物组及用于实施实施方式 1 的方法的使用说明。
- [0443] 22. 实施方式 21 的试剂盒,所述试剂盒包含列于表 1A 或 1B 的至少 10 个基因表达信号。
- [0444] 23. 实施方式 21 的试剂盒,其含有根据实施方式 6 的方法鉴定的至少 30 种核酸生物标记物。
- [0445] 24. 表 1A 或 1B 中的任何序列用于鉴定一种或更多目标肿瘤特征的用途。

- [0446] 25. 实施方式 23 的用途, 其中使用至少 3 个不同标记物组。
- [0447] 26. 实施方式 5 的方法, 其中癌生物标记物是乳腺癌生物标记物, 且样品的第 1 亚型是 ER+ 样品。
- [0448] 27. 实施方式 5 的方法, 其中随机练习组如下产生: 随机挑选样品, 同时维持与选出它们的其他组相同的“良性”和“恶性”肿瘤比。
- [0449] 28. 实施方式 1 的方法, 其中将指定为恶性肿瘤的全部基因表达值分组, 并进行以下步骤:
- [0450] (1) 自鉴定的基因表达信号创建至少 30 个随机练习数据组;
- [0451] (2) 比较新组的鉴定的基因表达信号与一系列已知的在癌中活跃的基因;
- [0452] (3) 选择对应于已知的癌基因列表上的那些的鉴定的基因表达信号;
- [0453] (4) 根据它们在生物学过程中的作用将选择的鉴定的基因表达信号分组;
- [0454] (5) 自步骤 4 的选择的基因表达信号组产生至少 25 个基因的随机基因表达信号组;
- [0455] (6) 将所述随机基因表达信号组与步骤 1 中获得的随机练习数据组关联;
- [0456] (7) 自对于步骤 6 的各基因表达信号组的关联获得用于存活筛选的 P 值;
- [0457] (8) 如果对于多于 90% 的随机练习数据组, 用于基因表达信号组的 P 值小于 0.05, 则保持所述基因表达信号组;
- [0458] (9) 基于组中基因出现的频度排序步骤 8 中保持的随机基因表达信号组;
- [0459] (10) 选择排在前面的至少 26 个基因作为潜在候选标记物;
- [0460] (11) 重复步骤 5 ~ 10, 并产生至少 26 个基因的另一, 独立, 排序组;
- [0461] (12) 比较自步骤 10 及步骤 11 的排在前面的基因;
- [0462] (13) 如果多于 25 个基因相同, 将所述排在前面的基因保持为标记物组;
- [0463] (14) 重复两次步骤 5 ~ 13, 以获得 3 个新的和不同的标记物组;
- [0464] (15) 输出所述 3 个不同的新标记物组。