



(12) 发明专利申请

(10) 申请公布号 CN 104066851 A

(43) 申请公布日 2014. 09. 24

(21) 申请号 201280058401. 1

(74) 专利代理机构 北京市柳沈律师事务所  
11105

(22) 申请日 2012. 09. 28

代理人 岑晓东

(30) 优先权数据

61/542, 141 2011. 09. 30 US

(51) Int. Cl.

C12Q 1/68(2006. 01)

(85) PCT国际申请进入国家阶段日

2014. 05. 28

(86) PCT国际申请的申请数据

PCT/US2012/057777 2012. 09. 28

(87) PCT国际申请的公布数据

W02013/055530 EN 2013. 04. 18

(71) 申请人 基因泰克公司

地址 美国加利福尼亚州

(72) 发明人 D. 夏姆斯 T. M. 霍尔科姆

K. 沃尔特

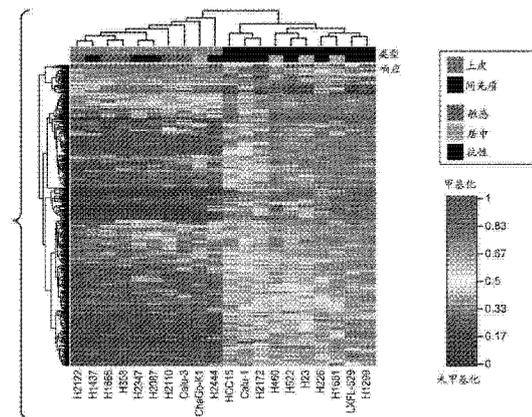
权利要求书1页 说明书70页 附图17页

(54) 发明名称

肿瘤或肿瘤细胞中上皮或间充质表型的诊断性甲基化标志物和对 EGFR 激酶抑制剂的响应

(57) 摘要

本发明提供用于测定肿瘤的上皮和间充质表型以及预测肿瘤生长对使用 EGFR 抑制剂治疗为敏感还是抗性的方法。具体地, 提供在至少一种选自下组的基因中的 CpG 位点处 DNA 甲基化的存在或缺失作为肿瘤细胞间充质表型的标志物, 用于确定肿瘤生长对 EGFR 激酶抑制剂所致抑制的敏感性, 和 / 或鉴定很可能受益于使用 EGFR 抑制剂治疗的癌症患者: CLDN7、HOXC4、CP2L3、TBCD、ESPR1、GRHL2、ERBB2 和 C20orf55。提供在至少一种选自下组的基因中的 CpG 位点处 DNA 甲基化的存在或缺失作为肿瘤细胞上皮表型的标志物: PCDH8、PEX5L、GALR1 和 ZEB2。



1. 一种测定肿瘤细胞是否具有间充质表型的方法,其包括检测至少一种选自下组的基因中 CpG 位点处 DNA 甲基化的存在或缺失:CLDN7、HOXC4、P2L3、TBCD、ESPR1、GRHL2、ERBB2 和 C20orf55,其中所述 CpG 位点处甲基化的存在指示肿瘤细胞具有间充质表型。

2. 一种测定肿瘤生长对 EGFR 激酶抑制剂所致抑制的敏感性的方法,其包括在样品肿瘤细胞中检测至少一种选自下组的基因中 CpG 位点处 DNA 甲基化的存在或缺失:CLDN7、HOXC4、P2L3、TBCD、ESPR1、GRHL2、ERBB2 和 C20orf55,其中所述 CpG 位点处甲基化的存在指示肿瘤生长对于使用 EGFR 抑制剂的抑制为抗性。

3. 一种鉴定很可能受益于使用 EGFR 抑制剂治疗的癌症患者的方法,其包括在来自患者的癌症的样品中检测至少一种选自下组的基因中 CpG 位点处 DNA 甲基化的存在或缺失:CLDN7、HOXC4、P2L3、TBCD、ESPR1、GRHL2 和 C20orf55,其中如果所述 CpG 位点处的 DNA 甲基化检测为缺失,那么将患者鉴定为很可能受益于使用 EGFR 抑制剂治疗。

4. 权利要求 3 的方法,其进一步包括如果患者鉴定为很可能受益于使用 EGFR 抑制剂治疗的患者,那么对所述患者施用治疗有效量的 EGFR 抑制剂。

5. 一种在患者中治疗癌症的方法,其包括对所述患者施用治疗有效量的 EGFR 抑制剂,其中所述患者在施用所述 EGFR 抑制剂之前诊断为患有展现出至少一种选自下组的基因中 CpG 位点处缺失 DNA 甲基化的癌症:CLDN7、HOXC4、P2L3、TBCD、ESPR1、GRHL2 和 C20orf55。

6. 权利要求 2-5 中任一项的方法,其中所述 EGFR 抑制剂是厄洛替尼(erlotinib)、西妥昔单抗(cetuximab)或帕尼单抗(panitumumab)。

7. 一种测定肿瘤细胞是否具有上皮表型的方法,其包括检测至少一种选自下组的基因中 CpG 位点处 DNA 甲基化的存在或缺失:PCDH8、PEX5L、GALR1 和 ZEB2,其中所述 CpG 位点处甲基化的存在指示肿瘤细胞具有上皮表型。

8. 前述权利要求中任一项的方法,其中通过焦磷酸测序检测甲基化的存在或缺失。

9. 前述权利要求中任一项的方法,其中从福尔马林固定、石蜡包埋的(FFPE)组织或新鲜冷冻的组织分离 DNA。

10. 权利要求 9 的方法,其中在焦磷酸测序前预扩增从组织样品分离的 DNA。

11. 权利要求 1 或 2 的方法,其中所述肿瘤细胞是 NSCLC 细胞。

12. 权利要求 3-5 中任一项的方法,其中所述癌症是 NSCLC。

## 肿瘤或肿瘤细胞中上皮或间充质表型的诊断性甲基化标志物和对 EGFR 激酶抑制剂的响应

[0001] 对相关申请的交叉引用

[0002] 本申请要求 2011 年 9 月 30 日提交的美国临时申请 No. 61/542, 141 的权益, 其公开内容通过提述完整并入本文。

### 发明领域

[0003] 本发明提供基于基因甲基化状态来预测对癌症疗法的响应的方法。

[0004] 发明背景

[0005] 本发明针对用于诊断和治疗癌症患者的方法。具体地, 本发明针对用于确定哪些患者将最受益于使用表皮生长因子受体 (EGFR) 激酶抑制剂治疗的方法。

[0006] 癌症是大范围细胞恶性的通用名, 其特征不在于不受调控的生长、缺少分化、和侵入局部组织并转移的能力。这些赘生性恶性以各种患病程度影响体内的每个组织和器官。

[0007] 在过去的数十年中已开发出多种治疗剂来治疗各种类型的癌症。最普遍使用的抗癌剂类型包括: DNA 烷化剂 (例如环磷酰胺 (cyclophosphamide)、异环磷酰胺 (ifosfamide))、抗代谢物 (例如甲氨蝶呤、叶酸拮抗剂和 5- 氟尿嘧啶、嘧啶拮抗剂)、微管破坏物 (例如长春新碱 (vincristine)、长春碱 (vinblastine)、帕利他赛 (paclitaxel))、DNA 嵌入剂 (例如多柔比星 (doxorubicin)、道诺霉素 (daunomycin)、顺铂)、和激素疗法 (例如他莫昔芬 (tamoxifen)、氟他胺 (flutamide))。

[0008] 表皮生长因子受体 (EGFR) 家族包含 4 种紧密相关的受体 (HER1/EGFR、HER2、HER3 和 HER4), 其牵涉细胞应答如分化和增殖。EGFR 激酶或其配体 TGF  $\alpha$  的过表达经常与许多癌症有关, 包括乳腺癌、肺癌、结直肠癌、卵巢癌、肾细胞癌、膀胱癌、头和颈癌、成胶质细胞瘤和星形细胞瘤, 且认为促进这些肿瘤的恶性生长。还发现 EGFR 基因 (EGFRvIII) 中的一种特定的缺失突变增加细胞致瘤性。EGFR 刺激的信号传导途径的活化促进潜在促进癌症的多种过程, 例如增殖、血管发生、细胞运动和侵入、降低的细胞凋亡和药物抗性诱导。增加的 HER1/EGFR 表达经常与晚期疾病、转移和不良预后关联。例如, 在 NSCLC 和胃癌中, 已显示增加的 HER1/EGFR 表达与高转移率、较差的肿瘤分化和增加的肿瘤增殖相关。

[0009] 已在 NSCLC 和成胶质细胞瘤中观察到激活受体的内在蛋白酪氨酸激酶活性和/或增加下游信号传导的突变。然而, 突变作用作为赋予对 EGF 受体抑制剂例如厄洛替尼 (erlotinib) (**TARCEVA**<sup>®</sup>) 或吉非替尼 (gefitinib) (IRESSA<sup>™</sup>) 的敏感性的原理机制一直是有争议的。最近, 已报告全长 EGF 受体的突变体形式预测对 EGF 受体酪氨酸激酶抑制剂吉非替尼的响应性 (Paez, J. G. 等 (2004) Science304:1497-1500 ;Lynch, T. J. 等 (2004) N. Engl. J. Med. 350:2129-2139)。细胞培养研究显示表达 EGF 受体的突变体形式的细胞系 (即 H3255) 对通过 EGF 受体酪氨酸激酶抑制剂吉非替尼的生长抑制更敏感, 且需要高得多浓度的吉非替尼来抑制表达野生型 EGF 受体的肿瘤细胞系。这些观察表明 EGF 受体的特定突变形式可能反映对 EGF 受体抑制剂的更大的敏感性但不鉴定完全不响应的表型。

[0010] 开发用作抗肿瘤剂的直接抑制 EGFR 的激酶活性的化合物, 以及通过阻断 EGFR 活

化来降低 EGFR 激酶活性的抗体是大量研究努力的领域所在 (de Bono J. S. 和 Rowinsky, E. K. (2002) *Trends in Mol. Medicine*8:S19-S26 ;Dancey, J. 和 Sausville, E. A. (2003) *Nature Rev. Drug Discovery*2:92-313)。数项研究已显示、披露或表明一些 EGFR 激酶抑制剂在与某些其他抗癌剂或化疗剂或治疗组合使用时可能改进肿瘤细胞或赘生物杀伤 (例如 Herbst, R. S. 等 (2001) *Expert Opin. Biol. Ther.* 1:719-732 ;Solomon, B. 等 (2003) *Int. J. Radiat. Oncol. Biol. Phys.* 55:713-723 ;Krishnan, S. 等 (2003) *Frontiers in Bioscience*8, e1-13 ;Grunwald, V. 和 Hidalgo, M. (2003) *J. Nat. Cancer Inst.* 95:851-867 ;Seymour L. (2003) *Current Opin. Investig. Drugs*4(6):658-666 ;Khalil, M. Y. 等 (2003) *Expert Rev. Anticancer Ther.* 3:367-380 ;Bulgaru, A. M. 等 (2003) *Expert Rev. Anticancer Ther.* 3:269-279 ;Dancey, J. 和 Sausville, E. A. (2003) *Nature Rev. Drug Discovery*2:92-313 ;Ciardiello, F. 等 (2000) *Clin. Cancer Res.* 6:2053-2063 ;和专利公开文本 No:US2003/0157104)。

[0011] 厄洛替尼 (例如厄洛替尼 HCl, 亦称为 **TARCEVA**<sup>®</sup> 或 OSI-774) 是口服可用的 EGFR 激酶的抑制剂。在体外, 厄洛替尼已在许多人肿瘤细胞系 (包括结直肠癌和乳腺癌中) 显示出针对 EGFR 激酶的实质抑制性活性 (Moyer J. D. 等 (1997) *Cancer Res.* 57:4838), 而且临床前评估显示针对许多表达 EGFR 的人肿瘤异种移植物的活性 (Pollack, V. A. 等 (1999) *J. Pharmacol. Exp. Ther.* 291:739)。最近, 厄洛替尼在许多适应证的 I 期和 II 期试验中显示有前景的活性, 所述适应证包括头和颈癌 (Soulieres, D., 等 (2004) *J. Clin. Oncol.* 22:77)、NSCLC (Perez-Soler R, 等 (2001) *Proc. Am. Soc. Clin. Oncol.* 20:310a, 摘要 1235)、CRC (Oza, M., 等 (2003) *Proc. Am. Soc. Clin. Oncol.* 22:196a, 摘要 785) 和 MBC (Winer, E., 等 (2002) *Breast Cancer Res. Treat.* 76:5115a, 摘要 445)。在 III 期试验中, 厄洛替尼单一疗法在患有晚期治疗不应性 NSCLC 的患者中显著延长存活、延迟疾病进展和延迟肺癌有关症状的恶化 (Shepherd, F. 等 (2004) *J. Clin. Oncology*, 22:14S (July 15 Supplement), 摘要 7022)。尽管许多关于厄洛替尼的临床试验数据涉及其在 NSCLC 中的使用, 但来自 I/II 期研究的初步结果已显示厄洛替尼和卡培他滨 (capecitabine)/厄洛替尼组合疗法在患有大范围的人实体肿瘤类型 (包括 CRC (Oza, M., 等 (2003) *Proc. Am. Soc. Clin. Oncol.* 22:196a, 摘要 785) 和 MBC (Jones, R. J., 等 (2003) *Proc. Am. Soc. Clin. Oncol.* 22:45a, 摘要 180)) 的患者中的有前景的活性。2004 年 11 月美国食品和药物管理局 (FDA) 批准了在至少一种在前的化疗方案失败后, 将厄洛替尼用于治疗患有局部晚期或转移性的非小细胞肺癌 (NSCLC) 的患者。厄洛替尼是表皮生长因子受体 (EGFR) 类中在 III 期临床试验中显示晚期 NSCLC 患者存活增加的唯一药物。

[0012] 抗赘生性药物理想地将选择性杀伤癌细胞, 关于其对非恶性细胞的毒性有较宽的治疗指标。它还会保留它对恶性细胞的功效, 甚至在延长暴露于药物之后。不幸的是, 没有一种目前的化疗具有这类理想概况。相反, 大多数具有非常窄的治疗指标。此外, 暴露于稍微亚致死浓度的化疗剂的癌细胞将非常经常地形成对这类药剂的抗性, 而且还相当经常地形成对几种其他抗赘生性药剂的交叉抗性。另外, 对于任何给定的癌症类型, 甚至在使用较新的基因靶向疗法如 EGFR 激酶抑制剂时也经常不能预测哪个患者很可能应答特定治疗, 如此需要大量试验和错误来发现最有效的疗法, 这对患者经常带来相当大的风险和不适。

[0013] 如此, 需要更有效的用于赘生物形成和其他增殖性病症的治疗, 和更有效的用于

确定哪些肿瘤将响应哪种治疗的手段。用于增强现有药物的治疗功效的策略牵涉其施用时间安排中的变化,以及其与其他抗癌剂或生化调控剂的组合使用。组合疗法公知为是一种相对于使用治疗有关剂量的单独每种药剂,能产生更大功效和减少的副反应的方法。在一些情况中,药物组合的功效是加和性的(组合的功效约等于单独每种药物的效果之和),但在其他情况中,该效果是协同性的(组合的功效大于单独给出的每种药物的效果之和)。

[0014] 靶物特异性的治疗办法如厄洛替尼一般与相比于常规细胞毒性剂的降低的毒性有关,因此,有助于其在组合方案中使用。在厄洛替尼与贝伐单抗(Mininberg, E. D., 等(2003)Proc. Am. Soc. Clin. Oncol. 22:627a, 摘要 2521)和吉西他滨(gemcitabine)(Dragovich, T., (2003)Proc. Am. Soc. Clin. Oncol. 22:223a, 摘要 895)组合的 I/II 期研究中已观察到有前景的结果。在 NSCLC III 期试验中的最近数据显示与标准化疗组合的一线厄洛替尼或吉非替尼未改进存活(Gatzemeier, U., (2004)Proc. Am. Soc. Clin. Oncol. 23:617(摘要 7010);Herbst, R. S., (2004)Proc. Am. Soc. Clin. Oncol. 23:617(摘要 7011);Giaccone, G., 等(2004)J. Clin. Oncol. 22:777;Herbst, R., 等(2004)J. Clin. Oncol. 22:785)。然而,胰腺癌 III 期试验显示一线厄洛替尼组合吉西他滨确实改进存活。

[0015] 几个小组已研究了潜在的生物标志物来预测患者对 EGFR 抑制剂的响应(参见例如,WO 2004/063709、WO 2005/017493、WO 2004/111273、WO2004/071572;US 2005/0019785 和 US 2004/0132097)。一种这类生物标志物是上皮和间充质表型。在大多数癌症转移期间,在肿瘤细胞中发生一项重要的变化,称为上皮-间充质转变(EMT)(Thiery, J. P. (2002)Nat. Rev. Cancer 2:442-454;Savagner, P. (2001)Bioessays 23:912-923;Kang Y. 和 Massague, J. (2004)Cell 118:277-279;Julien-Grille, S., 等 Cancer Research 63:2172-2178;Bates, R. C. 等(2003)Current Biology 13:1721-1727;Lu Z., 等(2003)Cancer Cell. 4(6):499-515)。紧密结合在一起且展现极性的上皮细胞产生间充质细胞,其更宽松地保持在一起,展现出极性的降低且具有行进的能力。这些间充质细胞能传播到原始肿瘤周围的组织中,侵入血管和淋巴管,并行进至新位置,在该处分裂并形成另外的肿瘤。EMT 不发生在健康细胞中,除了在胚胎发生期间以外。在正常情况下,TGF- $\beta$  充当生长抑制剂,然而在癌症转移期间,TGF- $\beta$  开始促进 EMT。

[0016] 上皮和间充质表型与特定的基因表达模式关联。例如,上皮表型在 WO2006101925 中显示为与 E-钙粘蛋白、Brk、 $\gamma$ -连环蛋白(catenin)、 $\alpha$ -连环蛋白、角蛋白 8、角蛋白 18、连接蛋白 31、plakophilin3、stratofin1、核纤层蛋白  $\alpha$  5 和 ST14 的高表达水平有关,而间充质表型与波形蛋白、纤连蛋白、微纤维蛋白-1、微纤维蛋白-2、胶原  $\alpha$  2(IV)、胶原  $\alpha$  2(V)、LOXL1、巢蛋白(nidogen)、C11 或 f9、肌腱蛋白(tenascin)、N-钙粘蛋白、胚 EDB+ 纤连蛋白、微管蛋白  $\alpha$  3 和上皮形成素(epimorphin)的高表达水平有关。

[0017] 表观遗传学(epigenetics)是对由基础 DNA 序列中变化以外的机制导致的基因表达或细胞表型中的可遗传变化的研究,因此名称为 epi-(希腊语:在上、之上、外部)-遗传学。这类变化的例子包括 DNA 甲基化和组蛋白修饰,两者均作用于调控基因表达而不改变相关基因的序列。这些变化可以是体细胞可遗传的,其经由生物体剩余生命中的细胞分裂,而且还可以传递到生物体的后续世代。然而,在生物体的基础 DNA 序列中没有变化;相反,是非遗传因素导致生物体的基因有不同的行为或表达。

[0018] DNA 甲基化是高等生物中正常生物发育和细胞分化中至关重要的部分。DNA 甲基

化稳定地改变细胞中的基因表达模式,从而细胞能“记得其曾在何处”;例如,在胚胎发育期间程序化为胰岛的细胞在生物体的整个生命中在没有告诉它们必须保持胰岛的持续信号的情况下保持为胰岛。另外,DNA 甲基化抑制病毒基因和其他已掺入宿主基因组的有害元件随时间的表达。DNA 甲基化还形成染色质结构的基础,这使得细胞能从 DNA 的单个不可变序列形成多细胞生命所需要的无数特征。DNA 甲基化还在几乎所有类型的癌症的形成中起着至关重要的作用。在胞嘧啶位置 5 的 DNA 甲基化具有降低基因表达的特定效果,且已在检查的每个脊椎动物中发现。在成人细胞组织中,DNA 甲基化通常发生在 CpG 二核苷酸背景中,而非 CpG 甲基化在胚胎干细胞中是普遍的。

[0019] “CpG”是“-C-磷酸-G-”的缩写,即胞嘧啶和鸟嘌呤由仅一个磷酸分开;磷酸将任意两个核苷在 DNA 中连接在一起。“CpG”记法用于将此线性序列与胞嘧啶和鸟嘌呤的 CG 碱基对区分开来。CpG 二核苷酸中的胞嘧啶可经甲基化以形成 5-甲基胞嘧啶(5-mC)。在哺乳动物中,将基因内的胞嘧啶甲基化能关闭该基因。将甲基基团添加到 DNA 的酶称为 DNA 甲基转移酶。在哺乳动物中,70%至 80%的 CpG 胞嘧啶是甲基化的。存在具有更高的 CpG 位点浓度的基因组区域,称为 CpG 岛。这些“CpG 岛”还具有高于预期的 GC 含量(即 >50%)。哺乳动物基因组中的许多基因具有与基因起始关联的 CpG 岛。因此,CpG 岛的存在被用来协助基因的预测和注释。CpG 岛对于甲基化是顽固性的,这可能有助于维持开放的染色质构型。另外,这能产生对转变突变的降低的易受损性,因此导致续存更高平衡密度的 CpG。基因启动子内 CpG 位点的甲基化能导致其沉默,这是在许多人癌症中发现的一种特征(例如肿瘤抑制基因的沉默)。相反,CpG 位点的低甲基化与癌细胞内癌基因的过表达有关。

#### [0020] 发明概述

[0021] 本发明的一个方面提供一种确定肿瘤细胞是否具有上皮表型的方法,包括检测肿瘤细胞中表 2 或表 4 中鉴定的任一个 CpG 位点处 DNA 甲基化的存在或缺失,其中任何 CpG 位点处甲基化的存在指示所述肿瘤细胞具有上皮表型。在某些实施方案中,所述 CpG 位点在 PCDH8、PEX5L、GALR1 或 ZEB2 基因中。在某些实施方案中,所述肿瘤细胞是 NSCLC 细胞。

[0022] 本发明的另一个方面提供一种确定肿瘤细胞是否具有上皮表型的方法,包括检测表 1 或表 3 中鉴定的任一个 CpG 位点处 DNA 甲基化的存在或缺失,其中任何 CpG 位点处甲基化的缺失指示所述肿瘤细胞具有上皮表型。在某些实施方案中,所述 CpG 位点在 CLDN7、HOXC4、P2L3、TBCD、ESPR1、GRHL2、或 C20orf55 基因中。在某些实施方案中,所述肿瘤细胞是 NSCLC 细胞。

[0023] 本发明的另一个方面提供一种测定肿瘤生长对 EGFR 激酶抑制剂所致抑制的敏感性的方法,包括检测样品肿瘤细胞中表 2 或表 4 中鉴定的任一个 CpG 位点处 DNA 甲基化的存在或缺失,其中任一个 CpG 位点处 DNA 甲基化的存在指示所述肿瘤生长对于使用 EGFR 抑制剂的抑制敏感。在一个实施方案中,所述 EGFR 抑制剂是厄洛替尼(erlotinib)、西妥昔单抗(cetuximab)或帕尼单抗(panitumumab)。在某些实施方案中,所述肿瘤细胞是 NSCLC 细胞。

[0024] 本发明的另一个方面提供一种鉴定很可能受益于使用 EGFR 抑制剂治疗的癌症患者的方法,包括在来自所述患者的癌症的样品中检测表 1 或表 3 中鉴定的任一个 CpG 位点处 DNA 甲基化的存在或缺失,其中如果任一个 CpG 位点处的 DNA 甲基化检测为缺失,那么将所述患者鉴定为很可能受益于使用所述 EGFR 抑制剂治疗。在某些实施方案中,所述 CpG 位

点在 CLDN7、HOXC4、P2L3、TBCD、ESPR1、GRHL2、或 C20orf55 基因中。在某些实施方案中，所述 EGFR 抑制剂是厄洛替尼、西妥昔单抗或帕尼单抗。在某些实施方案中，所述癌症是 NSCLC。

[0025] 本发明的又一个方面提供一种鉴定很可能受益于使用 EGFR 抑制剂治疗的癌症患者的方法，包括在来自所述患者的癌症的样品中检测表 2 或表 4 中鉴定的任一个 CpG 位点处 DNA 甲基化的存在或缺失，其中如果任一个 CpG 位点处的 DNA 甲基化检测为存在，那么将所述患者鉴定为很可能受益于使用所述 EGFR 抑制剂治疗。在某些实施方案中，如果所述患者被鉴定为很可能会受益于使用 EGFR 抑制剂治疗的患者，那么对所述患者施用治疗有效量的 EGFR 抑制剂。在某些实施方案中，所述 EGFR 抑制剂是厄洛替尼、西妥昔单抗或帕尼单抗。在某些实施方案中，所述癌症是 NSCLC。

[0026] 本发明的另一个方面提供一种确定肿瘤细胞是否具有间充质表型的方法，包括检测肿瘤细胞中表 2 或表 4 中鉴定的任一个 CpG 位点处 DNA 甲基化的存在或缺失，其中任何 CpG 位点处甲基化的缺失指示所述肿瘤细胞具有间充质表型。在某些实施方案中，所述 CpG 位点在 PCDH8、PEX5L、GALR1 或 ZEB2 基因中。在某些实施方案中，所述肿瘤细胞是 NSCLC 细胞。

[0027] 本发明的另一个方面提供一种确定肿瘤细胞是否具有间充质表型的方法，包括检测表 1 或表 3 中鉴定的任一个 CpG 位点处 DNA 甲基化的存在或缺失，其中任何 CpG 位点处甲基化的存在指示所述肿瘤细胞具有间充质表型。在某些实施方案中，所述 CpG 位点在 CLDN7、HOXC4、P2L3、TBCD、ESPR1、GRHL2、或 C20orf55 基因中。在某些实施方案中，所述肿瘤细胞是 NSCLC 细胞。

[0028] 本发明的又一个方面提供一种测定肿瘤生长对 EGFR 激酶抑制剂所致抑制的敏感性的方法，包括检测样品肿瘤细胞中表 2 或表 4 中鉴定的任一个 CpG 位点处 DNA 甲基化的存在或缺失，其中任一个 CpG 位点处 DNA 甲基化的缺失指示所述肿瘤生长对使用所述 EGFR 抑制剂的抑制有抗性。在某些实施方案中，所述 EGFR 抑制剂是厄洛替尼、西妥昔单抗或帕尼单抗。在某些实施方案中，所述肿瘤细胞是 NSCLC 细胞。

[0029] 本发明的另一个方面提供一种测定肿瘤生长对 EGFR 激酶抑制剂所致抑制的敏感性的方法，包括检测样品肿瘤细胞中表 1 或表 3 中鉴定的任一个 CpG 位点处 DNA 甲基化的存在或缺失，其中任一个 CpG 位点处 DNA 甲基化的存在指示所述肿瘤生长对使用所述 EGFR 抑制剂（例如厄洛替尼、吉非替尼 (gefitinib)、拉帕替尼 (lapatinib)、西妥昔单抗或帕尼单抗）的抑制有抗性。在某些实施方案中，所述 CpG 位点在 CLDN7、HOXC4、P2L3、TBCD、ESPR1、GRHL2、或 C20orf55 基因中。在某些实施方案中，所述 EGFR 抑制剂是厄洛替尼、西妥昔单抗或帕尼单抗。在某些实施方案中，所述肿瘤细胞是 NSCLC 细胞。

[0030] 本发明的另一个方面提供一种在患者中治疗癌症的方法，包括对所述患者施用治疗有效量的 EGFR 抑制剂，其中所述患者在施用所述 EGFR 抑制剂之前诊断为患有展现出在表 2 或表 4 中鉴定的任一个 CpG 位点处存在 DNA 甲基化的癌症。在某些实施方案中，所述 EGFR 抑制剂是厄洛替尼、西妥昔单抗或帕尼单抗。在某些实施方案中，所述癌症是 NSCLC。

[0031] 本发明的另一个方面提供一种在患者中治疗癌症的方法，包括对所述患者施用治疗有效量的 EGFR 抑制剂，其中所述患者在施用所述 EGFR 抑制剂之前诊断为患有展现出在表 1 或表 3 中鉴定的任一个 CpG 位点处缺失 DNA 甲基化的癌症。在某些实施方案中，所述 EGFR 抑制剂是厄洛替尼、西妥昔单抗或帕尼单抗。在某些实施方案中，所述癌症是 NSCLC。

[0032] 本发明的另一个方面提供一种为癌症患者选择疗法的方法,其包括在来自所述患者的癌症的样品中检测表 2 或表 4 中鉴定的一个 CpG 位点处 DNA 甲基化的存在或缺失,且当表 2 或表 4 中鉴定的任一个 CpG 位点处检测为存在甲基化时,选出用于治疗 EGFR 抑制剂的步骤。在一个实施方案中,如果选择 EGFR 疗法,那么对患者施用治疗有效量的 EGFR 抑制剂,如厄洛替尼、西妥昔单抗或帕尼单抗。在某些实施方案中,所述患者患有 NSCLC。

[0033] 本发明的另一个方面提供一种为癌症患者选择疗法的方法,其包括在来自所述患者的癌症的样品中检测表 1 或表 3 中鉴定的一个 CpG 位点处 DNA 甲基化的存在或缺失,且当表 1 或表 3 中鉴定的任一个 CpG 位点处检测为缺失甲基化时,选出用于治疗 EGFR 抑制剂的步骤。在一个实施方案中,如果选择 EGFR 疗法,那么对患者施用治疗有效量的 EGFR 抑制剂,如厄洛替尼、西妥昔单抗或帕尼单抗。在某些实施方案中,所述患者患有 NSCLC。

[0034] 在上文各方面的某些实施方案中,甲基化的存在或缺失通过焦磷酸测序检测。在上文各方面的某些实施方案中,DNA 从福尔马林固定、石蜡包埋的 (FFPE) 组织或新鲜冷冻的组织分离。在一个实施方案中,在焦磷酸测序前将从组织样品分离的 DNA 预扩增。

[0035] 附图简述

[0036] 本专利或申请文件含有至少一幅以彩色制成的图。具有彩图的本专利或专利申请公开文本的拷贝会在请求并支付必要费用后由专利局提供。

[0037] 图 1。依照 Fluidigm EMT 基因表达集分类为上皮和间充质表型的 NSCLC 细胞系。

[0038] 图 2。将细胞系表征为上皮样或间充质样的层次聚类。

[0039] 图 3。分类为对 EGFR 抑制剂厄洛替尼敏感性、居中和抗性的上皮和间充质 NSCLC 细胞系的 DNA 甲基化样式。

[0040] 图 4。选择用于亚硫酸氢钠测序或 qMSP 和焦磷酸测序阵列设计的 DMR 的注释。

[0041] 图 5A。对 CLDN7 启动子区的焦磷酸测序基于上皮样 / 间充质样表型区分 42 种 NSCLC 细胞系。

[0042] 图 5B。CLDN7mRNA 的相对表达,其使用标准的  $\Delta Ct$  方法在 42 种 (n = 20 种上皮样, 19 种间充质样, 3 种居中) 经 DMSO 处理和经 5- 氮杂 -dC 处理的 NSCLC 细胞系中测定。

[0043] 图 6A-H。基于 TaqMan 的甲基化检测测定法,其特异于与基因 (A)MST1R/RON、(C) FAM110A、(E)CP2L3/GRHL2 和 (G)ESRP1 有关的 DMR,以及 (B)RON、(D)FAM110A、(F)GRHL2 和 (H)ESRP1 的接受者操作特征 (Receiver operating characteristic) (ROC) 图。

[0044] 图 7A-M。在厄洛替尼敏感性对厄洛替尼抗性 NSCLC 细胞系中的定量甲基化特异性 PCR 测定法的接受者操作特征 (ROC) 曲线 -PEX5L (A)、PCDH8 (B)、ZEB2 (C)、ME3 (D)、MSTR1 (E)、STX2 (F)、HOXC5 (G)、C20orf55 (H)、ESRP1 (I)、BCAR3 (J)、CLDN7 (K)、NKX6. 2 (L)、CP2L3 (M)。

[0045] 图 8A-B。显示 82 种 NSCLC 细胞系的上皮 (E) 或间充质 (M) 分类和厄洛替尼 IC50 值的表。

[0046] 列表

[0047] 表 1。与间充质表型有关的甲基化胞嘧啶核苷酸 (CpG)。

[0048] 表 2。与上皮表型有关的甲基化胞嘧啶核苷酸 (CpG)。

[0049] 表 3。与间充质表型有关的甲基化胞嘧啶核苷酸 (CpG),通过基因的染色体编号、核苷酸位置和 Entrez ID 标识。

[0050] 表 4。与上皮表型有关的甲基化胞嘧啶核苷酸 (CpG)，通过基因的染色体编号、核苷酸位置和 Entrez ID 标识。

[0051] 发明详述

[0052] 术语“癌症”在动物中指具有致癌细胞典型特征的细胞存在，所述特征如不受控的增殖、永生化、转移潜力、快速生长和增殖速率、和某些特征性形态学特征。癌症细胞经常将以肿瘤的形式，但这类细胞可在动物中单独存在，或可在血液流中作为独立细胞循环，如白血细胞。

[0053] 如本文中使用的，除非另外指示，“异常细胞生长”指独立于正常调控机制的细胞生长（例如失去接触抑制）。这包括以下异常生长：(1) 通过表达突变的酪氨酸激酶或过表达受体酪氨酸激酶来增殖的肿瘤细胞（肿瘤）；(2) 其他增殖性疾病的良性和恶性细胞，其中发生异常酪氨酸激酶活化；(4) 通过受体酪氨酸激酶增殖的任何肿瘤；(5) 通过异常丝氨酸 / 苏氨酸激酶活化增殖的任何肿瘤；和 (6) 其他增殖性疾病的良性和恶性细胞，其中发生异常丝氨酸 / 苏氨酸激酶活化。

[0054] 如本文中使用的，除非另外指示，术语“治疗 / 处理”意指在患者中逆转、减轻、抑制进展、或预防（部分或完全地）肿瘤生长、肿瘤转移或其他致癌或赘生性细胞。如本文中使用的，除非另外指示，术语“治疗 / 处理”指治疗 / 处理的行为。

[0055] 短语“一种治疗方法”或其等同体，在应用于例如癌症时指设计为减少或消除动物中的癌症细胞数，或减轻癌症症状的作用规程或过程。“一种治疗”癌症或另一种增殖性病症的“方法”不必然指事实上将消除癌细胞或其他病症，事实上将减少细胞数或病症，或事实上将减轻癌症或其他病症的症状。

[0056] 术语“治疗有效的药剂”意指将引发研究者、兽医、医学医师或其他临床医师寻求的组织、系统、动物或人的生物学或医学应答的组合物。

[0057] 术语“治疗有效量”或“有效量”意指将引发研究者、兽医、医学医师或其他临床医师寻求的组织、系统、动物或人的生物学或医学应答的受试化合物或组合的量。

[0058] 术语“ErbB1”、“HER1”、“表皮生长因子受体”和“EGFR”和“EGFR 激酶”在本文中可交换使用且指 EGFR，如例如在 Carpenter 等 *Ann. Rev. Biochem.* 56:881-914(1987) 中披露的，包括其天然存在的突变体形式（例如一种缺失突变体 EGFR，如 Humphrey 等 *PNAS(USA)* 87:4207-4211(1990) 中的）。erbB1 指编码 EGFR 蛋白产物的基因。

[0059] 如本文中使用的，术语“EGFR 激酶抑制剂”和“EGFR 抑制剂”指本领域中目前已知或未来将鉴定的任何 EGFR 激酶抑制剂，其包括在对患者施用导致与患者中 EGF 受体活化有关的生物学活性（包括另外从 EGFR 天然配体对 EGFR 的结合得到的任何下游生物学效应）受抑制的任何化学实体。这类 EGFR 激酶抑制剂包括能阻断 EGFR 活化或 EGFR 活化的任何下游生物学效应（与患者中的癌症治疗有关）的任何药剂。这类抑制剂能通过直接结合受体的胞内域并抑制其激酶活性起作用。或者，这类抑制剂能通过占据 EGF 受体的配体结合位点或其部分起作用，由此使得受体为其天然配体不可达的，从而阻止或降低其正常生物学活性。或者，这类抑制剂能通过调控 EGFR 多肽的二聚体化，或 EGFR 多肽与其他蛋白质的相互作用，或增强 EGFR 的遍在蛋白化和胞吞降解起作用。EGFR 激酶抑制剂包括但不限于低分子量抑制剂、抗体或抗体片段、反义构建体、小抑制 RNA（即通过 dsRNA 的 RNA 干扰；RNAi）和核酶。在一个优选的实施方法中，所述 EGFR 激酶抑制剂是特异性结合人 EGFR 的小

有机分子或抗体。

[0060] EGF 受体功能的抑制剂已显示临床效用,而且描述最可能受益于治疗的患者亚组的关键 EGF 受体信号传导途径的定义已成为一个重要的研究领域。已在 NSCLC 和成胶质细胞瘤中观察到活化受体的内在蛋白质酪氨酸激酶活性和 / 或增加下游信号传导的突变。体外和临床研究已显示 wt EGF 受体细胞系和肿瘤之间在其对 EGF 受体抑制的细胞应答中的相当大的变异性,这部分已显示为源自磷脂酰肌醇 3- 激酶途径的 EGF 受体独立性的活化,导致抗细胞凋亡丝氨酸 - 苏氨酸激酶 Akt 的持续的磷酸化。PI3 激酶活化的备选路径和由此的 EGF 受体抑制剂不敏感性的分子决定物是一个积极的研究领域。例如强烈活化 PI3 激酶途径的胰岛素样生长因子 1 受体 (IGF-1 受体) 牵连对 EGF 抑制剂的细胞抗性。在介导对选择性 EGF 受体抑制的不敏感性中经由 PI3 激酶途径也能施加存活信号的细胞 - 细胞和细胞 - 粘着网络的作用不那么清楚且将假定为影响对 EGF 受体阻断物的细胞敏感性。在缺少对胞外基质的粘附或细胞 - 细胞接触的情况下肿瘤细胞维持生长和存活信号的能力是重要的,不仅在细胞迁移和转移的背景中,而且还在维持伤口样肿瘤环境中的细胞增殖和存活中,其中胞外基质被重塑且细胞接触抑制减少。

[0061] 先前已开发了与 NSCLC 细胞系对厄洛替尼的体外敏感性相关的 EMT 基因表达签名。(Yauch 等, 2005, Clin Cancer Res 11, 8686-8698)。基于该项工作开发了与上皮和间充质表型有关的基于 fluidigm 的 EMT 表达签名 (图 1)。

[0062] 本发明部分基于使用整合的基因组办法,其将基因表达分析与全基因组甲基化概况分析组合以显示甲基化生物标志物能对癌症 (如 NSCLC) 中的上皮和间充质表型分类,从而证明 DNA 甲基化模式中的基因组范围内的差异与癌症的独特的生物学和临床相关子集有关。使用本文中描述的方法将甲基化模式用于癌症的表型子集分类是有利的,因为它需要比更传统的基于 DNA 和 RNA 的分析方法更少量的测试组织。该特征在组织有限的情况下分析临床样品时特别有用。

[0063] 在开发预测性生物标志物中的一项主要挑战是需要建立用于前景评估的有力的“截点 (cut-point)”。这对于基于蛋白质的测定法如免疫组织化学尤其成问题。尽管受到广泛使用,但免疫组织化学受制于许多限制其在预测性生物标志物开发中使用的技术挑战。这些限制包括抗体特异性和灵敏性、表位可获性和稳定性、以及不同病理学家对数据解译的内在主观性 (24、25)。能利用 PCR 的动态范围和特异性的分子测定法是更期望的。然而,基于 PCR 的测定法也存在限制:RNA 是高度不稳定的且需要预先设定的截留点。突变检测测定法,尽管是潜在二元性的,但受到高流行性突变热点和靶序列的可获性的限制。如实施例中显示的,基于 PCR 的甲基化测定法潜在地解决了这些问题中的许多个,因为它们具有突变测定法的许多特性,包括广泛的动态范围和基本二元性的读出 (具有与突变测定法类似的灵敏性),但由于 CpG 甲基化状态的局部相关行为,用于测定法设计的靶区域可以是相当大的。更重要的是, DNA 甲基化可用于以差不多与过去使用基因表达相同的方式来推断肿瘤的生物学状态。

[0064] 本文实施例中呈现的数据显示,在细胞培养中或在体内生长的含有野生型 EGFR 的肿瘤细胞如 NSCLC 或胰腺癌细胞,根据其是否经历上皮到间充质转变 (EMT),显示对 EGFR 激酶抑制剂的抑制的一定范围的敏感性。在 EMT 之前,肿瘤细胞对于 EGFR 激酶抑制剂如厄洛替尼 HCl (Tarceva<sup>®</sup>) 的抑制非常敏感,而经历 EMT 的肿瘤细胞对于这类化合物的抑制

要不敏感得多。数据指示 EMT 可能是决定肿瘤对 EGFR 激酶抑制剂的敏感性水平的一般性生物学转换。本文中显示,肿瘤对 EGFR 激酶抑制剂的敏感性水平可通过测定由肿瘤细胞表达的、EMT 事件之前或之后的细胞特征性的生物标志物的水平来评估。例如,上皮生物标志物如 E-钙粘蛋白的高肿瘤细胞表达水平(指示细胞尚未经历 EMT)与对 EGFR 激酶抑制剂的高敏感性相关。相反,间充质生物标志物如波形蛋白或纤连蛋白的高肿瘤细胞表达水平(指示细胞已经历 EMT)与对 EGFR 激酶抑制剂的低敏感性相关。如此,这些观察能形成用于预测 EGFR 激酶抑制剂对肿瘤生长的影响的诊断方法的基础,并给予肿瘤学家一项协助他们来选择针对其患者的最适宜治疗的工具。

[0065] 如实施例中描述的,基于 DNA 甲基化模式,癌症可区分成为上皮样(EL)和间充质样(ML)肿瘤。间充质表型(或已经历 EMT 的肿瘤细胞)与表 1 和表 3 中显示的特定基因的甲基化有关。因此,本发明提供一种确定肿瘤细胞是否具有间充质表型的方法,包括检测肿瘤细胞中表 1 或表 3 中鉴定的任一个 CpG 位点处 DNA 甲基化的存在或缺失,其中任何 CpG 位点处的甲基化指示所述肿瘤细胞具有间充质表型。相反,表 1 或表 3 中鉴定的任一个 CpG 位点处 DNA 甲基化的缺失指示所述肿瘤具有上皮表型。

[0066] 在一个具体的实施方案中,确定肿瘤细胞是否具有间充质表型的方法包括检测以下一种或多种基因中 CpG 位点处甲基化的存在或缺失:CLDN7(密蛋白(claudin)-7)、HOXC4(同源框 C4)、CP2L3(grainyhead 样 3)、STX2(突触融合蛋白 2)、RON(巨噬细胞刺激性 1 受体)、TBCD(微管蛋白特异性分子伴侣 D)、ESRP1(上皮剪接调控蛋白 1)、GRHL2(grainyhead 样 2)、ERBB2 和 C20orf55(第 20 号染色体开读框 55)基因,其中任一个 CpG 位点处甲基化的存在指示所述肿瘤具有间充质表型。相反,任一个 CpG 位点处 DNA 甲基化的缺失指示所述肿瘤具有上皮表型。在一个具体的实施方案中,所述方法包括在以下一种或多种基因中检测 CpG 位点处的甲基化:CLDN7、HOXC4、CP2L3、STX2、RON、TBCD、ESRP1、GRHL2、ERBB2 和 C20orf55 基因,其中任一个 CpG 位点处甲基化的存在指示所述肿瘤具有间充质表型。在一个具体的实施方案中,检测出表 1 或表 3 中的 2 种基因中 CpG 位点处甲基化的存在指示所述肿瘤具有间充质表型。在一个具体的实施方案中,检测出表 1 或表 3 中的 3 种基因中 CpG 位点处甲基化的存在指示所述肿瘤具有间充质表型。在一个具体的实施方案中,检测出表 1 或表 3 中的 4 种基因中 CpG 位点处甲基化的存在指示所述肿瘤具有间充质表型。在一个具体的实施方案中,检测出表 1 或表 3 中的 5 种基因中 CpG 位点处甲基化的存在指示所述肿瘤具有间充质表型。在一个具体的实施方案中,检测出 CLDN7、HOXC4、CP2L3、STX2、RON、TBCD、ESRP1、GRHL2 和 C20orf55 基因中的 2、3 或 4、5、6、7、8 或所有 9 种基因中 CpG 位点处甲基化的存在指示所述肿瘤具有间充质表型。在另一个实施方案中,检测出 CLDN7、RON、ESRP1 和 GRHL2 中的 2、3 或 4 种中 CpG 位点处甲基化的存在指示所述肿瘤具有间充质表型。在另一个实施方案中,检测出 CLDN7、RON、ESRP1 和 GRHL2 的所有 4 种基因中 CpG 位点处甲基化的存在指示所述肿瘤具有间充质表型。

[0067] 另外,本发明提供一种预测肿瘤生长对通过 EGFR 抑制剂的抑制的敏感性的方法,包括在取自肿瘤样品细胞中检测表 1 或表 3 中鉴定的任一个 CpG 位点处 DNA 甲基化的存在或缺失,其中在任一个 CpG 位点处 DNA 甲基化的存在指示肿瘤生长对于使用 EGFR 抑制剂的抑制是抗性的。相反,在任一个 CpG 位点处 DNA 甲基化的缺失指示肿瘤生长对于通过 EGFR 抑制剂的抑制是敏感的(即响应的)。在一个具体的实施方案中,检测出表 1 或表 3 中

的 2 种基因中 CpG 位点处甲基化的存在指示所述肿瘤生长对于使用 EGFR 抑制剂的抑制是抗性的。在一个具体的实施方案中,检测出表 1 或表 3 中的 3 种基因中 CpG 位点处甲基化的存在指示所述肿瘤生长对于使用 EGFR 抑制剂的抑制是抗性的。在一个具体的实施方案中,检测出表 1 或表 3 中的 4 种基因中 CpG 位点处甲基化的存在指示所述肿瘤生长对于使用 EGFR 抑制剂的抑制是抗性的。在一个具体的实施方案中,检测出表 1 或表 3 中的 5 种基因中 CpG 位点处甲基化的存在指示所述肿瘤生长对于使用 EGFR 抑制剂的抑制是抗性的。在一个具体的实施方案中,检测出 CLDN7、HOXC4、CP2L3、STX2、RON、TBCD、ESRP1、GRHL2、ERBB2 和 C20orf55 基因中的 2、3 或 4、5、6、7、8 或所有 9 种基因中 CpG 位点处甲基化的存在指示所述肿瘤生长对于使用 EGFR 抑制剂的抑制为抗性的。在另一个实施方案中,检测出 CLDN7、RON、ESRP1 和 GRHL2 中的 2、3 或 4 种中 CpG 位点处甲基化的存在指示所述肿瘤生长对于使用 EGFR 抑制剂的抑制为抗性的。在另一个实施方案中,检测出 CLDN7、RON、ESRP1 和 GRHL2 的所有 4 种基因中 CpG 位点处甲基化的存在指示所述肿瘤生长对于使用 EGFR 抑制剂的抑制为抗性的。

[0068] 本发明的另一个方面提供一种鉴定很可能受益于使用 EGFR 抑制剂治疗的癌症患者的方法,包括在来自所述患者的癌症的样品中检测表 1 或表 3 中鉴定的任一个 CpG 位点处 DNA 甲基化的存在或缺失,其中如果任一个 CpG 位点处的 DNA 甲基化检测为缺失,那么将所述患者鉴定为很可能受益于使用 EGFR 抑制剂治疗。相反,在任一个 CpG 位点处 DNA 甲基化的存在指示患者不太可能受益于使用 EGFR 抑制剂治疗。在一个具体的实施方案中,检测表 1 或表 3 中的 2 种基因中 CpG 位点处的甲基化为缺失指示患者很可能受益于使用 EGFR 抑制剂治疗。在一个具体的实施方案中,检测表 1 或表 3 中的 3 种基因中 CpG 位点处的甲基化为缺失指示所述患者很可能受益于使用 EGFR 抑制剂治疗。在一个具体的实施方案中,检测表 1 或表 3 中的 4 种基因中 CpG 位点处的甲基化为缺失指示所述患者很可能受益于使用 EGFR 抑制剂治疗。在一个具体的实施方案中,检测表 1 或表 3 中的 5 种基因中 CpG 位点处的甲基化为缺失指示所述患者很可能受益于使用 EGFR 抑制剂治疗。在一个具体的实施方案中,检测 CLDN7、HOXC4、CP2L3、STX2、RON、TBCD、ESRP1、GRHL2、ERBB2 和 C20orf55 基因中的 2、3 或 4、5、6、7、8 或所有 9 种基因中 CpG 位点处甲基化为缺失指示所述患者很可能受益于使用 EGFR 抑制剂治疗。在另一个实施方案中,检测 CLDN7、RON、ESRP1 和 GRHL2 中的 2、3 或 4 种中 CpG 位点处甲基化为缺失指示所述患者很可能受益于使用 EGFR 抑制剂治疗。在另一个实施方案中,检测 CLDN7、RON、ESRP1 和 GRHL2 的所有 4 种中 CpG 位点处甲基化为缺失指示所述患者很可能受益于使用 EGFR 抑制剂治疗。在某些实施方案中,对视为很可能受益于使用 EGFR 抑制剂治疗的患者施用治疗有效量的 EGFR 抑制剂。

[0069] 如实施例中描述的,肿瘤细胞中的上皮表型与表 2 和表 4 中显示的特定基因的甲基化有关。因此,本发明提供一种确定肿瘤细胞是否具有上皮表型的方法,包括在肿瘤细胞中检测表 2 或表 4 中鉴定的任一个胞嘧啶核苷酸 (CpG 位点) 处 DNA 甲基化的存在或缺失,其中任一个胞嘧啶核苷酸 (CpG 位点) 处甲基化的存在指示所述肿瘤细胞具有上皮表型。相反,本发明还提供一种确定肿瘤细胞是否具有上皮表型的方法,包括在肿瘤细胞中检测表 2 或表 4 中鉴定的任一个 CpG 位点处 DNA 甲基化的存在或缺失,其中 CpG 位点处甲基化的缺失指示所述肿瘤具有间充质表型。

[0070] 在一个具体的实施方案中,所述方法包括在以下一种或多种基因中的 CpG 位点处

检测甲基化的存在或缺失:PCDH8(原钙粘蛋白(proto-cadherin)8)、PEX5L(过氧化物酶体生源因子5样)、GALR1(神经节肽(galanin)受体1)、ZEB2(锌指E盒结合同源框2)和ME3(苹果酸酶3)基因,其中CpG位点处甲基化的存在指示肿瘤具有上皮表型。在一个具体的实施方案中,所述方法包括检测ZEB2基因中的CpG位点处甲基化的存在或缺失,其中CpG位点处甲基化的存在指示所述肿瘤具有上皮表型。在一个具体的实施方案中,检测表2或表4中的2种基因中CpG位点处的甲基化为存在指示所述肿瘤具有上皮表型。在一个具体的实施方案中,检测表2或表4中的3种基因中CpG位点处的甲基化为存在指示所述肿瘤具有上皮表型。在一个具体的实施方案中,检测表2或表4中的4种基因中CpG位点处的甲基化为存在指示所述肿瘤具有上皮表型。在一个具体的实施方案中,检测表2或表4中的5种基因中CpG位点处的甲基化为存在指示所述肿瘤具有上皮表型。在一个具体的实施方案中,检测PCDH8、PEX5L、GALR1或ZEB2基因的每一种中CpG位点处的甲基化为存在指示所述肿瘤具有上皮表型。

[0071] 另外,本发明提供一种预测肿瘤生长对通过EGFR抑制剂的抑制的敏感性的方法,包括在取自肿瘤样品细胞中检测表2或表4中鉴定的任一个CpG位点处DNA甲基化的存在或缺失,其中在任一个CpG位点处DNA甲基化的存在指示肿瘤生长对于使用EGFR抑制剂的抑制是敏感的。相反,在任一个CpG位点处DNA甲基化的缺失指示肿瘤生长对于通过EGFR抑制剂的抑制是抗性的。在一个具体的实施方案中,所述方法包括检测PCDH8、PEX5L、GALR1或ZEB2基因的一种或多种中CpG位点的甲基化,其中在任一个CpG位点处甲基化的存在指示肿瘤生长对于使用EGFR抑制剂的抑制是敏感的。在一个具体的实施方案中,所述方法包括检测ZEB2基因中CpG位点处的甲基化,其中ZEB2基因中CpG位点处甲基化的存在指示肿瘤生长对于使用EGFR抑制剂的抑制是敏感的。在一个具体的实施方案中,检测表2或表4中的2种基因中CpG位点处的甲基化为存在指示所述肿瘤生长对于使用EGFR抑制剂的抑制是敏感的。在一个具体的实施方案中,检测表2或表4中的3种基因中CpG位点处的甲基化为存在指示所述肿瘤生长对于使用EGFR抑制剂的抑制是敏感的。在一个具体的实施方案中,检测表2或表4中的4种基因中CpG位点处的甲基化为存在指示所述肿瘤生长对于使用EGFR抑制剂的抑制是敏感的。在一个具体的实施方案中,检测表2或表4中的5种基因中CpG位点处的甲基化为存在指示所述肿瘤生长对于使用EGFR抑制剂的抑制是敏感的。在一个具体的实施方案中,检测PCDH8、PEX5L、GALR1或ZEB2基因的每一种中CpG位点处的甲基化为存在指示所述肿瘤生长对于使用EGFR抑制剂的抑制是敏感的。

[0072] 本发明的另一个方面提供一种鉴定很可能受益于使用EGFR抑制剂治疗的癌症患者的方法,包括在来自所述患者的癌症样品中检测表2或表4中鉴定的任一个CpG位点处DNA甲基化的存在或缺失,其中如果任一个CpG位点处的DNA甲基化检测为存在,那么将所述患者鉴定为很可能受益于使用EGFR抑制剂治疗。相反,任一个CpG位点处的DNA甲基化的缺失指示患者不太可能受益于使用EGFR抑制剂治疗。在一个具体的实施方案中,检测表2或表4中的2种基因中CpG位点处的甲基化为存在指示所述患者很可能受益于使用EGFR抑制剂治疗。在一个具体的实施方案中,检测表2或表4中的3种基因中CpG位点处的甲基化为存在指示所述患者很可能受益于使用EGFR抑制剂治疗。在一个具体的实施方案中,检测表2或表4中的4种基因中CpG位点处的甲基化为存在指示所述患者很可能受益于使用EGFR抑制剂治疗。在一个具体的实施方案中,检测表2或表4中的5种基因中CpG

位点处的甲基化为存在指示所述患者很可能受益于使用 EGFR 抑制剂治疗。在一个具体的实施方案中,检测 PCDH8、PEX5L、GALR1 或 ZEB2 中的 2、3 或 4 种中 CpG 位点处的甲基化为存在指示所述患者很可能受益于使用 EGFR 抑制剂治疗。在另一个实施方案中,检测 ZEB2 中 CpG 位点处的甲基化为存在指示所述患者很可能受益于使用 EGFR 抑制剂治疗。在某些实施方案中,对视为很可能受益于使用 EGFR 抑制剂治疗的患者施用治疗有效量的 EGFR 抑制剂。

[0073] 本发明的另一个方面提供一种治疗癌症患者的方法,使用本文中描述的 DNA 甲基化概况分析先前已鉴定所述癌症患者是很可能受益于使用 EGFR 抑制剂治疗的患者,所述方法包括对该患者施用治疗有效量的 EGFR 抑制剂。

[0074] 本发明的另一个方面提供一种基于本文中描述的 DNA 甲基化概况分析方法为癌症患者选择疗法的方法。在一个实施方案中,所述方法包括在来自所述患者的癌症的样品中检测表 2 或表 4 中鉴定的 CpG 位点之一处甲基化的存在或缺失,并且当表 2 或表 4 中鉴定的 CpG 位点之一处检测为存在甲基化时,选出 EGFR 抑制剂用于疗法。在另一个实施方案中,所述方法包括在来自所述患者的癌症的样品中检测表 1 或表 3 中鉴定的 CpG 位点之一处 DNA 甲基化的存在或缺失,并且当表 1 或表 3 中鉴定的 CpG 位点之一处检测为缺失甲基化时,选出 EGFR 抑制剂用于疗法。在某些实施方案中,如果选择 EGFR 抑制剂疗法,那么对患者施用治疗有效量的 EGFR 抑制剂,如厄洛替尼、西妥昔单抗或帕尼单抗。

[0075] 医学领域,特别是关于诊断性测试的应用和治疗学治疗的医学领域中的技术人员将认可生物学系统可展现出变异性且并非总是完全可预测的,如此许多较好的诊断性测试或治疗学有时是无效的。如此,最终由主治医师基于测试结果、患者状况和病史、及其自身经验的判断来确定个体患者的最适宜的治疗过程。甚至有以下情况,例如,当基于来自诊断测试或来自其他标准的数据肿瘤未预测为对 EGFR 激酶抑制剂特别敏感时,医师会选择用 EGFR 抑制剂来治疗患者,尤其是在如果所有或大多数其他明显的治疗选项已经失败,或者预期在与另一种治疗一起施用时有某些协同性的情况下。与癌症治疗中使用的许多其他抗癌药物如更传统的化疗剂或细胞毒性剂相比,EGFR 抑制剂作为一类相对耐受较好的药物的事实使其成为更可行的选择。

[0076] 因此,本发明提供一种预测肿瘤细胞生长对 EGFR 激酶抑制剂所致抑制的敏感性的方法,包括:评估肿瘤细胞中一种或多种(或一组)上皮生物标志物的 DNA 甲基化水平;和预测肿瘤细胞生长对 EGFR 抑制剂所致抑制的敏感性,其中所有肿瘤细胞上皮生物标志物的同时较高的 DNA 甲基化水平与对通过 EGFR 抑制剂的抑制的高度敏感性相关。在该方法的一个具体的实施方案中,所述上皮生物标志物包括基因 PCDH8、PEX5L、GALR1、ZEB2 和 ME3,其中两种肿瘤细胞上皮生物标志物的同时的高表达水平与对 EGFR 激酶抑制剂所致抑制的高度敏感性相关。

[0077] 本发明还提供一种预测肿瘤细胞生长对 EGFR 激酶抑制剂所致抑制的敏感性的方法,包括:评估肿瘤细胞中一种或多种(或一组)间充质生物标志物的水平;和预测肿瘤细胞生长对 EGFR 抑制剂所致抑制的敏感性,其中所有肿瘤细胞间充质生物标志物的同时的高水平与对通过 EGFR 抑制剂的抑制的抗性相关。在该方法的一个具体的实施方案中,所述间充质生物标志物包括基因 CLDN7、HOXC4、CP2L3、TBCD、ESRP1、GRHL2 和 C20orf55,其中至少两种肿瘤细胞间充质生物标志物的同时的高 DNA 甲基化水平与对通过 EGFR 抑制剂的抑制的抗性相关。

[0078] 本发明还提供一种预测癌症患者是否患有会对使用 EGFR 激酶抑制剂的治疗有效响应的肿瘤的方法,包括:评估肿瘤细胞中一种或多种(或一组)上皮生物标志物 PCDH8、PEX5L、GALR1、ZEB2 和 ME3 的 DNA 甲基化水平;和预测肿瘤是否会有有效响应使用 EGFR 抑制剂的治疗,其中所有肿瘤细胞上皮生物标志物的同时的高表达水平与会有有效响应使用 EGFR 抑制剂的治疗的肿瘤相关。

[0079] 本发明还提供一种预测癌症患者是否患有会对使用 EGFR 激酶抑制剂的治疗有效响应的肿瘤的方法,包括:评估肿瘤细胞中一种或多种(或一组)间充质生物标志物 CLDN7、HOXC4、CP2L3、TBCD、ESRP1、GRHL2 和 C20orf55 的水平;和预测肿瘤是否会有有效响应使用 EGFR 抑制剂的治疗,其中所有这类肿瘤细胞间充质生物标志物的高 DNA 甲基化水平与对使用 EGFR 抑制剂的治疗有抗性的肿瘤相关。

[0080] 在本文所述方法中,肿瘤细胞通常将来自诊断为患有癌症、癌症前期疾患、或异常细胞生长的另一种形式,并且需要治疗的患者。所述癌症可以是肺癌(例如非小细胞肺癌(NSCLC))、胰腺癌、头和颈癌、胃癌、乳腺癌、结肠癌、卵巢癌或下文中描述的多种其他癌症中的任一种。所述癌症是已知潜在可用 EGFR 抑制剂治疗的癌症。肿瘤细胞可获自患者痰、唾液、血液、尿、粪、脑脊髓液,或直接获自肿瘤,例如通过细针穿刺(fine needle aspirate)。

[0081] 可通过许多方法学来分析样品中各种生物标志物的存在和/或水平/量,其中许多是本领域中已知且熟练技术人员理解的,包括但不限于免疫组织化学(“IHC”)、Western 印迹分析、免疫沉淀、分子结合测定法、ELISA、ELIFA、荧光激活细胞分选(“FACS”)、MassARRAY、蛋白组学、基于血液的定量测定法(如例如血清 ELISA)、生化酶活性测定法、原位杂交、Southern 分析、Northern 分析、全基因组测序、聚合酶链式反应(“PCR”)(包括定量实时 PCR(“qRT-PCR”)和其他扩增类型检测方法,如例如分支的 DNA、SISBA、TMA 等)、RNA-Seq、FISH、微阵列分析、基因表达概况分析、和/或基因表达系列分析(“SAGE”),以及可通过蛋白质、基因和/或组织阵列分析实施的许多种测定法中的任一种。用于评估基因和基因产物状态的典型方案见于例如 Ausubel 等编,1995, Current Protocols In Molecular Biology, Units2(Northern Blotting), 4(Southern Blotting), 15(Immunoblotting) 和 18(PCR Analysis)。还可使用多重免疫测定法如那些可从 Rules Based Medicine or Meso Scale Discovery(“MSD”)获得的。

[0082] 用于评估 DNA 甲基化的方法是公知的。例如, Laird(2010)Nature Reviews Genetics11:191-203 提供关于 DNA 甲基化分析的综述。在一些实施方案中,评估甲基化的方法包括将基因组 DNA 随机剪切或随机片段化,用甲基化依赖性 or 甲基化敏感性限制酶切割 DNA,随后选择性鉴定和/或分析切割或未切割的 DNA。选择性鉴定可包括例如,分离切割和未切割的 DNA(例如通过大小)并量化感兴趣的切割或未切割的序列。参见例如,美国专利 No. 7, 186, 512。在一些实施方案中,所述方法可涵盖在限制酶消化后扩增完整的 DNA,由此仅扩增在扩增区域中未被限制酶切割的 DNA。参见例如,美国专利申请 No. 10/971, 986; 11/071, 013; 和 10/971, 339。在一些实施方案中,扩增可使用基因特异性的引物进行。或者,可将接头添加到随机片段化的 DNA 的端,所述 DNA 可用甲基化依赖性 or 甲基化敏感性限制酶消化,完整 DNA 可使用与接头序列杂交的引物扩增。在一些实施方案中,可实施第二步骤来测定 DNA 的扩增池中特定基因的存在、缺失或量。在一些实施方案中,使用实时定量 PCR 来扩增 DNA。

[0083] 在一些实施方案中,所述方法包括量化基因组 DNA 群体内的靶序列中的平均甲基化密度。在一些实施方案中,所述方法包括使基因组 DNA 与甲基化依赖性限制酶或甲基化敏感性限制酶在允许基因座中至少一些潜在的限制酶切割位点的拷贝保持未切割的条件下接触;量化基因座的完整拷贝;并将扩增产物的量与代表对照 DNA 甲基化的量的对照值比较,由此相比于对照 DNA 的甲基化密度量化基因座中的平均甲基化密度。

[0084] DNA 基因座的甲基化的量可通过以下来测定:提供包含该基因座的基因组 DNA 样品,将 DNA 用甲基化敏感或甲基化依赖性的限制酶切割,然后在感兴趣的 DNA 基因座处量化完整 DNA 的量或量化切割 DNA 的量。完整或切割 DNA 的量将依赖于含有该基因座的基因组 DNA 的起始量,基因座中甲基化的量,和在基因组 DNA 中甲基化的基因座中核苷酸的数目(即部分)。DNA 基因座中的甲基化量可通过将完整 DNA 或切割 DNA 的量与在类似处理的 DNA 样品中代表完整 DNA 或切割 DNA 的量的对照值比较来测定。对照值能代表已知或预测数目的甲基化核苷酸。或者,对照值能代表来自另一种细胞(例如正常、未患病的)中的同一基因座或第二基因座的完整或切割 DNA 的量。

[0085] 通过在允许基因座中至少一些潜在的限制酶切割位点的拷贝保持未切割的条件下使用甲基化敏感性或甲基化依赖性限制酶,随后量化剩余的完整拷贝,并将该量与对照比较,能测定基因座的平均甲基化密度。如果使甲基化敏感性限制酶与 DNA 基因座的拷贝在允许基因座中至少一些潜在的限制酶切割位点的拷贝保持未切割的条件下接触,那么剩余的完整 DNA 将与甲基化密度直接成比例,如此可与对照比较来测定样品中基因座的相对甲基化密度。类似地,如果使甲基化依赖性限制酶与 DNA 基因座的拷贝在允许基因座中至少一些潜在的限制酶切割位点的拷贝保持未切割的条件下接触,那么剩余的完整 DNA 将与甲基化密度成反比,如此可与对照比较来测定样品中基因座的相对甲基化密度。这类测定法披露在例如美国专利申请 Ser. No. 10/971,986 中。

[0086] 在一些实施方案中,可使用定量扩增方法(例如定量 PCR 或定量线性扩增)来量化限制性消化后由扩增引物侧接的基因座中完整 DNA 的量。定量扩增的方法披露于例如美国专利 No. 6,180,349;6,033,854;和 5,972,602,以及例如 Gibson 等,Genome Research6:995-1001(1996);DeGraves 等,Biotechniques34(1):106-10,112-5(2003);Deiman B 等,Mol Biotechnol. 20(2):163-79(2002)。

[0087] 用于检测 DNA 甲基化的其他方法可牵涉在用亚硫酸氢盐处理 DNA 之前和之后的基因组测序。参见例如,Frommer 等,Proc. Natl. Acad. Sci. USA89:1827-1831(1992)。当将亚硫酸氢钠与 DNA 接触时,未甲基化的胞嘧啶被转化成尿嘧啶,而甲基化的胞嘧啶不被修饰。

[0088] 在一些实施方案中,将从亚硫酸氢盐转化后的 DNA 扩增的 PCR 产物的限制酶消化用于检测 DNA 甲基化。参见例如,Sadri&Hornsby,Nucl. Acids Res. 24:5058-5059(1996);Xiong&Laird,Nucleic Acids Res. 25:2532-2534(1997)。

[0089] 在一些实施方案中,单独或与其他方法组合地使用 MethyLight 测定法来检测 DNA 甲基化(参见 Eads 等,Cancer Res. 59:2302-2306(1999))。简言之,在 MethyLight 方法中,基因组 DNA 在亚硫酸氢钠反应中转化(亚硫酸氢盐方法将未甲基化的胞嘧啶残基转化成尿嘧啶)。然后,实施感兴趣 DNA 序列的扩增,其使用与 CpG 二核苷酸杂交的 PCR 引物。通过使用仅与得自未甲基化 DNA 的亚硫酸氢盐转化的序列(或者与未转化的甲基化序列)杂交的引物,扩增可指示引物杂交之处序列的甲基化状态。类似地,扩增产物可使用一种探针检

测,该探针特异性结合得自未甲基化(或甲基化)DNA的亚硫酸氢盐处理的序列。如果期望的话,可将两种引物和探针均用于检测甲基化状态。如此,用于与MethyLight使用的试剂盒可包括亚硫酸氢钠以及在经亚硫酸氢盐处理的甲基化和未甲基化DNA之间区分的引物或可检测标记的探针(包括但不限于Taqman或分子信标探针)。其他试剂盒组分可包括,例如DNA扩增所需的试剂,包括但不限于PCR缓冲液、脱氧核苷酸;和热稳定的聚合酶。

[0090] 在一些实施方案中,单独或与其他方法组合地使用Ms-SNuPE(甲基化敏感性单核苷酸引物延伸,Methylation-sensitive Single Nucleotide Primer Extension)来检测DNA甲基化(参见Gonzalgo&Jones *Nucleic Acids Res.* 25:2529-2531(1997))。Ms-SNuPE技术是用于基于DNA的亚硫酸氢盐处理来评估特定CpG位点处的甲基化差异,接着是单核苷酸引物延伸的定量方法。简言之,使基因组DNA与亚硫酸氢钠反应来将未甲基化的胞嘧啶转化成尿嘧啶,而留下5-甲基胞嘧啶不变。然后使用特异于经亚硫酸氢盐转化的DNA的PCR引物来实施对期望靶序列的扩增,将所得产物分离并用作在感兴趣的CpG位点处甲基化分析的模板。

[0091] 在一些实施方案中,单独或与其他方法组合地使用甲基化特异性PCR(“MSP”)反应来检测DNA甲基化。MSP测定法必须首先通过亚硫酸氢钠来修饰DNA,将所有未甲基化的而非甲基化的胞嘧啶转化成尿嘧啶,随后使用特异于甲基化相对于未甲基化DNA的引物进行扩增。参见,Herman等,*Proc. Natl. Acad. Sci. USA*93:9821-9826,(1996);美国专利No. 5,786,146。在一些实施方案中,DNA甲基化通过QIAGEN PyroMark CpG Assay预设计的焦磷酸测序DNA甲基化测定法检测。

[0092] 在一些实施方案中,使用高通量DNA甲基化分析来测定细胞甲基化状态以确定对EGFR抑制剂的敏感性。简言之,从细胞或组织样品(例如肿瘤样品或血液样品)分离基因组DNA并在亚硫酸氢钠反应中转化(亚硫酸氢盐方法将未甲基化的胞嘧啶残基转化成尿嘧啶),其使用本领域中的标准测定法。将经亚硫酸氢盐转化的DNA产物扩增、片段化并与含有来自整个基因组的CpG位点的阵列杂交,其使用本领域中的标准测定法。杂交后,将阵列成像,并使用本领域中的标准测定法处理用于分析DNA甲基化状态。在一些实施方案中,所述组织样品是福尔马林固定的石蜡包埋的(FFPE)组织。在一些实施方案中,所述组织样品是新鲜冷冻的组织。在一些实施方案中,在亚硫酸氢盐转化前将从组织样品分离的DNA预扩增。在一些实施方案中,在亚硫酸氢盐转化前将从组织样品分离的DNA预扩增,其通过使用Invitrogen Superscript III One-Step RT-PCR系统和Platinum Taq。在一些实施方案中,在亚硫酸氢盐转化前使用基于Taqman的测定法将从组织样品分离的DNA预扩增。在一些实施方案中,所述亚硫酸氢钠反应使用Zymo EZ DNA甲基化试剂盒进行。在一些实施方案中,扩增经亚硫酸氢盐转化的DNA并与阵列杂交,其使用Illumina Infinium HumanMethylation450Beadchip试剂盒。在一些实施方案中,将阵列在Illumina iScan Reader上成像。在一些实施方案中,将图像用GenomeStudio软件甲基化模块处理。在一些实施方案中,使用Bioconductor lumi软件程序包来分析甲基化数据。参见Du等,*Bioinformatics*, 24(13):1547-1548(2008)。

[0093] 在一些实施方案中,使用亚硫酸氢盐测序PCR(BSP)鉴定DNA甲基化位点以确定对EGFR抑制剂的敏感性。简言之,从细胞或组织样品(例如肿瘤样品或血液样品)分离基因组DNA并在亚硫酸氢钠反应中转化(该亚硫酸氢盐方法将未甲基化的胞嘧啶残基转化成尿

嘧啶),其使用本领域中的标准测定法。将经亚硫酸氢盐转化的 DNA 产物用设计为特异于经亚硫酸氢盐转化的 DNA 的引物(例如亚硫酸氢盐特异性引物)扩增并连接到载体中用于转化到宿主细胞中,其使用本领域中的标准测定法。在选出含有 PCR 扩增的、经亚硫酸氢盐转化的、感兴趣的 DNA 产物的宿主细胞之后,分离 DNA 产物并测序以确定甲基化的位点,其使用本领域中的标准测定法。在一些实施方案中,所述组织样品是福尔马林固定的石蜡包埋的 (FFPE) 组织。在一些实施方案中,所述组织样品是 FFPE 组织,其已经过处理用于 IHC 分析;例如针对基因表达。在一些实施方案中,所述组织样品是通过 IHC 显示极少或无基因表达的 FFPE 组织。在一些实施方案中,所述组织样品是新鲜冷冻的组织。在一些实施方案中,在亚硫酸氢盐转化前将从组织样品分离的 DNA 预扩增。在一些实施方案中,在亚硫酸氢盐转化前将从组织样品分离的 DNA 预扩增,其使用 Invitrogen Superscript III One-Step RT-PCR 系统和 Platinum Taq。在一些实施方案中,在亚硫酸氢盐转化前使用基于 Taqman 的测定法将从组织样品分离的 DNA 预扩增。在一些实施方案中,所述亚硫酸氢钠反应使用 Zymo EZ DNA 甲基化金试剂盒进行。在一些实施方案中,设计为特异于经亚硫酸氢盐转化的 DNA 的引物是使用 Applied Biosystems Methyl Primer Express 软件设计的。在一些实施方案中,所述亚硫酸氢盐转化的 DNA 产物是使用 Invitrogen Superscript III One-Step RT-PCR System 和 Platinum Taq PCR 扩增的。在进一步的实施方案中,PCR 扩增的、经亚硫酸氢盐转化的 DNA 产物使用 Invitrogen TOPO TA 克隆试剂盒连接到载体中。在一些实施方案中,宿主细胞是细菌。在一些实施方案中,分离的 PCR 扩增的、经亚硫酸氢盐转化的、感兴趣的 DNA 产物使用 Applied Biosystems 3730x1 DNA Analyzer 测序。在一些实施方案中,设计为特异于经亚硫酸氢盐转化的 DNA 的引物是使用 Qiagen PyroMark Assay Design 软件设计的。在一些实施方案中,经亚硫酸氢盐转化的 DNA 产物是使用 Invitrogen Superscript III One-Step RT-PCR System 和 Platinum Taq PCR 扩增的。在进一步的实施方案中,PCR 扩增的、经亚硫酸氢盐转化的 DNA 产物使用 Qiagen Pyromark Q24 测序并用 Qiagen PyroMark 软件分析。

[0094] 在一些实施方案中,使用定量甲基化特异性 PCR (QMSP) 来鉴定 DNA 甲基化位点以确定对 EGFR 抑制剂的敏感性。简言之,从细胞或组织样品分离基因组 DNA 并在亚硫酸氢钠反应中转化(该亚硫酸氢盐方法将未甲基化的胞嘧啶残基转化成尿嘧啶),其使用本领域中的标准测定法。在一些实施方案中,所述组织样品是福尔马林固定的石蜡包埋的 (FFPE) 组织。在一些实施方案中,所述组织样品是 FFPE 组织,其已经过处理用于 IHC 分析。在一些实施方案中,所述组织样品是通过 IHC 显示极少或无基因表达的 FFPE 组织。在一些实施方案中,所述组织样品是新鲜冷冻的组织。经亚硫酸氢盐转化的 DNA 产物是使用设计为特异于经亚硫酸氢盐转化的 DNA 的引物(例如定量甲基化特异性 PCR 引物)扩增的。将经亚硫酸氢盐转化的 DNA 产物使用定量甲基化特异性 PCR 引物扩增并分析甲基化,其使用本领域中的标准测定法。在一些实施方案中,所述组织样品是福尔马林固定的石蜡包埋的 (FFPE) 组织。在一些实施方案中,所述组织样品是新鲜冷冻的组织。在一些实施方案中,在亚硫酸氢盐转化前将从组织样品分离的 DNA 预扩增,其使用 Invitrogen Superscript III One-Step RT-PCR 系统和 Platinum Taq。在一些实施方案中,在亚硫酸氢盐转化前将从组织样品分离的 DNA 预扩增。在一些实施方案中,在亚硫酸氢盐转化前使用基于 Taqman 的测定法将从组织样品分离的 DNA 预扩增。在一些实施方案中,所述亚硫酸氢钠反应使用商品化

的试剂盒进行。在一些实施方案中,所述亚硫酸氢钠反应使用 Zymo EZ DNA 甲基化金试剂盒进行。在一些实施方案中,设计为特异于经亚硫酸氢盐转化的 DNA 的引物是使用 Applied Biosystems Methyl Primer Express 软件设计的。在一些实施方案中,经亚硫酸氢盐转化的 DNA 是使用基于 Taqman 的测定法扩增的。在一些实施方案中,在 Applied Biosystems 7900HT 上扩增经亚硫酸氢盐转化的 DNA 并使用 Applied Biosystems SDS 软件分析。

[0095] 在一些实施方案中,本发明提供测定甲基化的方法,其通过 1) 对肿瘤样品的 IHC 分析,接着是 2) 对从步骤 1 的 IHC 分析中使用的肿瘤组织提取的 DNA 的定量甲基化特异性 PCR。简言之,通过以下两种方法之一从 IHC 载玻片移去盖玻片:将载玻片置于冷冻器中达至少 15 分钟,然后使用剃刀片从显微镜载玻片撬开盖玻片。然后,将载玻片在二甲苯中室温温育以溶解封固剂。或者,将载玻片浸泡在二甲苯中直至盖玻片脱落。这可能占用数天。所有载玻片都通过 5 分钟二甲苯 (x3) 和 5 分钟 100% 乙醇 (x2) 的脱石蜡化规程。将组织用剃刀片从载玻片刮下,置于含有蛋白酶 K 的组织裂解缓冲液中,并在 56°C 过夜温育。在其中组织在温育后仍然存在的情况下,可再添加 10  $\mu$ l 蛋白酶 K,并将组织再温育 30 分钟。继续 DNA 提取;例如通过使用 QIAamp DNA FFPE 组织试剂盒。对从 IHC 载玻片直接提取的 DNA 进行 QMSP 分析,其使用如上文描述的 QMSP3 引物和探针。

[0096] 在一些实施方案中,通过深度测序来对经亚硫酸氢盐转化的 DNA 测序。深度测序是其中多次读序列的方法,如直接焦磷酸测序。深度测序可用于检测稀有事件如稀有突变。对有限数目基因座的超深度测序可通过对 PCR 产物的直接焦磷酸测序和通过在单次运行中对超过 100 种 PCR 产物测序来实现。在对经亚硫酸氢盐转化的 DNA 测序中的挑战源于其在胞嘧啶残基到胸腺嘧啶(尿嘧啶)残基的亚硫酸氢盐转化后的低序列复杂性。可引入降低代表性亚硫酸氢盐测序 (Reduced representation bisulphite sequencing, RRBS) 来减少序列冗余,其通过仅选择基因组的一些区域来通过 DNA 片段的大小-分选分离 (size-fractionation) 测序 (Laird, PW Nature Reviews 11:195-203 (2010))。靶向可通过在测序前阵列捕捉或挂锁捕捉 (padlock capture) 来完成。例如,在固定阵列上或通过溶液杂交选择的靶向性捕捉能富集 DNA 或 RNA 寡核苷酸库靶向的序列并且可在亚硫酸氢盐转化之前或之后实施。或者,挂锁捕捉提供改进的富集效率,其通过组合两种系留探针的增加的退火特异性,随后使用通用引物来进行扩增,其允许比使用基因座特异性引物扩增更均一的代表性。

[0097] 其他甲基化检测方法包括但不限于,甲基化的 CpG 岛扩增(参见 Toyota 等, Cancer Res. 59:2307-12 (1999)) 和记载于例如美国专利公开文本 2005/0069879; Rein 等, Nucleic Acids Res. 26(10):2255-64 (1998); Olek 等, Nat Genet. 17(3):275-6 (1997); Laird, PW Nature Reviews 11:195-203 (2010); 和 PCT 公开文本 No. W000/70090) 的那些。

[0098] DNA 甲基化水平可由甲基化指标表示为甲基化的 DNA 拷贝数(循环次数)对参照基因的循环次数(其等同扩增甲基化和非甲基化靶物)的比率。高水平的 DNA 甲基化可通过比较非赘生性细胞的样品(特别是相同组织类型的或来自外周血单个核细胞的)中的 DNA 甲基化水平来测定。在一个具体的实施方案中,特定基因中的高水平 DNA 甲基化可在相比于正常细胞中更高的水平检测到。在另一个具体的实施方案中,高水平的 DNA 甲基化相比于正常细胞中为约 2 倍或更高。在一个具体的实施方案中,高水平的 DNA 甲基化相比于正常细胞中为约 3 倍或更高。在一个具体的实施方案中,高水平的 DNA 甲基化相比于正常

细胞中为约 4 倍或更高。在一个具体的实施方案中,高水平的 DNA 甲基化相比于正常细胞中为约 5 倍或更高。在一个具体的实施方案中,高水平的 DNA 甲基化相比于正常细胞中为约 6 倍或更高。在一个具体的实施方案中,高水平的 DNA 甲基化相比于正常细胞中为约 7 倍或更高。在一个具体的实施方案中,高水平的 DNA 甲基化相比于正常细胞中为约 8 倍或更高。在一个具体的实施方案中,高水平的 DNA 甲基化相比于正常细胞中为约 9 倍或更高。在一个具体的实施方案中,高水平的 DNA 甲基化相比于正常细胞中为约 10 倍或更高。

[0099] “低甲基化”意指可能甲基化的 CpG 位点中多数是未甲基化的。在某些实施方案中,低甲基化意指基因的一部分中低于 50%、低于 45%、低于 40%、低于 35%、低于 30%、低于 25%、低于 20%、低于 15%、低于 10%、低于 5%、或低于 1%的可能的甲基化位点是甲基化的。在再一个实施方案中,低甲基化意指相比于在正常水平例如在非肿瘤细胞中表达的基因,更少的可能甲基化的位点是甲基化的。在另一个实施方案中,低甲基化意指没有一个 CpG 位点是甲基化的。

[0100] “高甲基化”意指可能甲基化的 CpG 位点中多数是甲基化的。在某些实施方案中,高甲基化意指基因的一部分中高于 50%、高于 55%、高于 60%、高于 65%、高于 70%、高于 75%、高于 80%、高于 85%、高于 90%、高于 95%、或高于 99%的可能的甲基化位点是甲基化的。在再一个实施方案中,高甲基化意指相比于在正常水平例如在非肿瘤细胞中表达的基因,更多的可能甲基化的位点是甲基化的。在另一个实施方案中,高甲基化意指所有的 CpG 位点是甲基化的。

[0101] 在一些实施方案中,通过评估细胞中的 mRNA 来测定细胞中的生物标志物表达。用于评估细胞中 mRNA 的方法是公知的,且包括例如使用互补 DNA 探针的杂交测定法(如使用特异于一种或多种基因的经标记 riboprobe 的原位杂交、Northern 印迹和相关技术)和各种核酸扩增测定法(如 RT-PCR,其使用特异于一种或多种基因的互补引物,和其他扩增类型检测方法如例如分支的 DNA、SISBA、TMA 等)。在一些实施方案中,比较生物标志物在测试样品与参照样样品中的表达。例如,测试样品可以是肿瘤组织样品,参照样样品可以是来自正常组织或细胞如 PBMC。

[0102] 来自哺乳动物的样品可使用 Northern、点印记或 PCR 分析方便地测定。另外,这类方法可包括允许测定生物学样品中靶 mRNA 的水平的一个或多个步骤(例如通过同时检查“看家”基因如肌动蛋白家族成员的比较性对照 mRNA 序列的水平)。任选地,可测定扩增的靶 cDNA 的序列。

[0103] 本发明的任选方法包括在组织或细胞样品中通过微阵列技术来检查或检测 mRNA 如靶 mRNA 的方案。使用核酸微阵列,将来自测试和对照组织样品的测试和对照 mRNA 样品逆转录并标记以生成 cDNA 探针。然后,将探针与固定化在固体支持物上的核酸阵列杂交。阵列配置为使得阵列的每个成员的序列和位置都是已知的。例如,可将其表达与抗血管生成疗法的增加或降低的临床益处相关的基因选择排列在固体支持物上。经标记探针与特定阵列成员的杂交指示该探针来源的样品表达该基因。

[0104] 依照一些实施方案,通过观察前述基因的蛋白质表达水平来测量存在和/或水平/量。在某些实施方案中,所述方法包括使生物学样品与针对本文中描述的生物标志物的抗体在允许生物标志物结合的条件下接触,并检测抗体与生物标志物之间是否形成复合物。这类方法可以是体外或体内方法。

[0105] 在某些实施方案中,使用 IHC 和染色方案来检查样品中生物标志蛋白的存在和 / 或水平 / 量。已显示组织切片的 IHC 染色是测定或检测样品中蛋白质存在的可靠方法。在一个方面,使用包括以下的方法测定生物标志物的水平,包括:(a) 使用抗体实施对样品(如受试者癌症样品)的 IHC 分析;并 b) 测定样品中生物标志物的水平。在一些实施方案中,相对于参照值测定 IHC 染色强度。

[0106] IHC 可与另外的技术如形态学染色和 / 或荧光原位杂交组合实施。IHC 有两种一般性方法;直接和间接测定法。依照第一种测定法,直接测定抗体对靶抗原的结合。该直接测定法使用经标记的试剂,如荧光标签或酶标记的一抗,其可以在没有别的抗体相互作用的情况下可视化。在典型的间接测定法中,未缀合的一抗结合抗原,然后经标记的二抗结合该一抗。在二抗缀合于酶标记物的情况下,添加生色或产荧光底物以提供抗原的可视化。发生信号扩增,因为几个二抗可以与一抗上的不同表位反应。

[0107] 用于 IHC 的一抗和 / 或二抗通常用可检测模块标记。存在大量标记,其一般可分组成以下类别:(a) 放射性同位素,如  $^{35}\text{S}$ 、 $^{14}\text{C}$ 、 $^{125}\text{I}$ 、 $^3\text{H}$  和  $^{131}\text{I}$ ;(b) 胶体金颗粒;(c) 荧光标记物,包括但不限于稀土螯合物(钨螯合物)、TexasRed、罗丹明、荧光素、丹酰、丽丝胺(Lissamine)、伞形酮(umbelliferone)、藻红蛋白(phycocrytherin)、藻蓝蛋白、或商品化的荧光团如 SPECTRUMORANGE7 和 SPECTRUM GREEN7 和 / 或上述任意一种或多种的衍生物;(d) 存在各种酶-底物标记物,且美国专利 No. 4, 275, 149 提供了对其中一些的综述。酶标记物的例子包括萤光素酶(例如萤火虫萤光素酶和细菌萤光素酶;美国专利 No. 4, 737, 456)、萤光素、2, 3-二氢酞嗪二酮(dihydrophthalazinedione)、苹果酸脱氢酶、脲酶、过氧化物酶如辣根过氧化物酶(HRPO)、碱性磷酸酶、 $\beta$ -半乳糖苷酶、葡糖淀粉酶、溶菌酶、糖氧化酶(例如葡萄糖氧化酶、半乳糖氧化酶、和葡萄糖-6-磷酸脱氢酶)、杂环氧化酶(如尿酸酶和黄嘌呤氧化酶)、乳过氧化物酶(lactoperoxidase)、微过氧化物酶(micropoxidase)等。

[0108] 酶-底物组合的例子包括,例如辣根过氧化物酶(HRPO)和作为底物的过氧化氢酶;碱性磷酸酶(AP)和作为生色底物的对硝基苯基磷酸;和  $\beta$ -D-半乳糖苷酶( $\beta$ -D-Gal)与生色底物(例如对硝基苯基- $\beta$ -D-半乳糖苷酶)或产荧光底物(例如 4-甲基伞形基(methylumbelliferyl)- $\beta$ -D-半乳糖苷酶)。对于这些的一般性综述,参见美国专利 No. 4, 275, 149 和 4, 318, 980。

[0109] 可将如此制备的标本封固并盖上盖玻片。然后确定载玻片评估,例如使用显微镜,并可采用本领域中常规使用的染色强度标准。在一些实施方案中,约 1+ 或更高的染色模式得分是诊断性和 / 或预后性的。在某些实施方案中,IHC 测定法中约 2+ 或更高的染色模式得分是诊断性和 / 或预后性的。在其他实施方案中,约 3 或更高的染色模式得分是诊断性和 / 或预后性的。在一个实施方案中,理解当使用 IHC 检查来自肿瘤或结肠腺瘤的细胞和 / 或组织时,通常在肿瘤细胞和 / 或组织中测定或评估染色(相对于可能存在于样品中的基质或周围组织)。

[0110] 在备选的方法中,可使样品与特异于生物标志物的抗体在足以形成抗体-生物标志物复合物的条件下接触,然后检测该复合物。可以许多途径来检测生物标志物的存在,如通过 Western 印迹和 ELISA 规程,其用于测定很多种组织和样品,包括血浆或血清。有大量使用这类测定形式的免疫测定技术,参见例如美国专利 No. 4, 016, 043、4, 424, 279 和 4, 018, 653。这些包括非竞争型的单位点和双位点二者或“三明治式”测定法,以及传统的

竞争性结合测定法。这些测定法还包括经标记抗体对靶生物标志物的直接结合。

[0111] 选定的生物标志物在组织或细胞样品中的存在和 / 或水平 / 量还可经由功能性或基于活性的测定法来检查。例如, 如果生物标志物是酶, 可以进行本领域中已知的测定法来测定或检测给定的酶活性在组织或细胞样品中的存在。

[0112] 在某些实施方案中, 针对测定的生物标志物的量中的差异和使用的样品质量中的可变性, 以及测定轮数之间的变异性两者将样品标准化。这类标准化可通过检测并纳入特定标准化生物标志物 (包括公知的看家基因如 ACTB) 水平来实现。或者, 标准化可基于所有测定基因或其较大子集的均值或中值信号 (全局标准化办法)。在一个基因接一个基因的基础上, 将受试者肿瘤 mRNA 或蛋白质的测量的经标准化的量与在参照集中发现的量比较。每种 mRNA 或蛋白质每份测试肿瘤每位受试者的经标准化表达水平可表示为在参照集中测量的表达水平的百分数。在要分析的特定受试者样品中测量的存在和 / 或表达水平 / 量将落在该范围内的某个百分数处, 这可通过本领域中公知的方法来测定。

[0113] 在某些实施方案中, 如下测定基因的相对表达水平:

[0114] 相对表达基因 1 样品 1 =  $2^{\exp(\text{Ct 看家基因} - \text{Ct 基因 1})}$ , Ct 在样品中测定。

[0115] 相对表达基因 1 参照 RNA =  $2^{\exp(\text{Ct 看家基因} - \text{Ct 基因 1})}$ , Ct 在参照样品中测定。

[0116] 标准化的相对表达基因 1 样品 1 = (相对表达基因 1 样品 1 / 相对表达基因 1 参照 RNA) x 100

[0117] Ct 是阈值循环。Ct 是反应中生成的荧光与阈值线相交的循环数。

[0118] 将所有实验均对参照 RNA 标准化, 参照 RNA 是来自各种组织来源的 RNA 的广泛混合物 (例如参照 RNA#636538, 来自 Clontech, Mountain View, CA)。将等同的参照 RNA 包含在每次 qRT-PCR 运行中, 从而允许在不同的实验轮之间比较结果。

[0119] 在一个实施方案中, 所述样品是临床样品。在另一个实施方案中, 所述样品用在诊断性测定法中。在一些实施方案中, 所述样品从原发性或转移性肿瘤获得。经常使用组织活检 (biopsy) 来获得肿瘤组织的代表性的片 / 块。或者, 可以以已知或认为含有感兴趣肿瘤细胞的组织或流体的形式间接获得肿瘤细胞。例如, 可通过切除、支气管镜检、细针抽吸、支气管刷检、或从痰、胸膜液或血液获得肺癌损伤的样品。在一些实施方案中, 所述样品包括循环的肿瘤细胞; 例如血液、尿液或痰中的循环癌细胞。可从癌症或肿瘤组织或其他身体样品如尿液、痰、血清或血浆检测基因或基因产物。上文论述的用于检测癌性样品中靶基因或基因产物的相同技术可应用于其他身体样品。癌细胞可能从癌损伤脱落并出现在这类身体样品中。通过筛选这类身体样品, 可实现对这些癌症的简单的早期诊断。另外, 通过测试这类身体样品中的靶基因或基因产物能更容易地监测治疗的进展。

[0120] 在某些实施方案中, 参照样品、参照细胞、参照组织、对照样品、对照细胞或对照组织是来自同一受试者或个体的单一样品或组合的多重样品, 其在不同于获得测试样品时的一个或多个时间点获得。例如, 参照样品、参照细胞、参照组织、对照样品、对照细胞或对照组织在早于获得测试样品时的时间点从同一受试者或个体获得。如果参照样品在癌症的初始诊断期间获得而测试样品在癌症变成转移性时的更晚时候获得, 那么这类参照样品、参照细胞、参照组织、对照样品、对照细胞或对照组织可以是有用的。

[0121] 在某些实施方案中, 参照样品、参照细胞、参照组织、对照样品、对照细胞或对照组

织是来自一个或多个并非该受试者或个体的健康个体的组合的多重样品。在某些实施方案中,参照样品、参照细胞、参照组织、对样品、对照细胞或对照组织是来自患有疾病或病症(例如癌症)的、并非该受试者或个体的一个或多个个体的组合的多重样品。在某些实施方案中,参照样品、参照细胞、参照组织、对样品、对照细胞或对照组织是来自并非该受试者或个体的一个或多个个体的正常组织的合并 RNA 样品或合并的血浆或血清样品。在某些实施方案中,参照样品、参照细胞、参照组织、对样品、对照细胞或对照组织是来自患有疾病或病症(例如癌症)的、并非该受试者或个体的一个或多个个体的肿瘤组织的合并 RNA 样品或合并的血浆或血清样品。

[0122] 在本发明的方法中,组织样品可以是体液或排泄物如血液、尿液、唾液、粪、胸膜液、淋巴液、痰、腹水、前列腺液、脑脊髓液(CSF)、或任何其他的身体分泌物或其衍生物。血液意为包括全血液、血浆、血清或血液的任何衍生物。在其中侵入性取样方法不适宜或不方便的情况下,有时可优选在这类体液或排泄物中评估肿瘤上皮或间充质生物标志物。

[0123] 在本发明的方法中,肿瘤细胞可以是肺癌肿瘤细胞(例如非小细胞肺癌(NSCLC))、胰腺癌肿瘤细胞、乳腺癌肿瘤细胞、头和颈癌肿瘤细胞、胃癌肿瘤细胞、结肠癌肿瘤细胞、卵巢癌肿瘤细胞、或来自如下文中描述的多种其他癌症中任一种的肿瘤细胞。所述肿瘤细胞优选为已知或预期表达 EGFR 的类型,来自实体肿瘤的所有肿瘤细胞均如此。EGFR 激酶可以是野生型或突变体形式。

[0124] 在本发明的方法中,肿瘤可以是肺癌肿瘤(例如非小细胞肺癌(NSCLC))、胰腺癌肿瘤、乳腺癌肿瘤、头和颈癌肿瘤、胃癌肿瘤、结肠癌肿瘤、卵巢癌肿瘤、或来自如下文中描述的多种其他癌症中任一种的肿瘤。所述肿瘤优选为其细胞已知或预期表达 EGFR 的类型,如所有实体肿瘤均如此。EGFR 可以是野生型或突变体形式。

[0125] 抑制剂和药物组合物

[0126] 适用于本发明的例示性 EGFR 激酶抑制剂包括,例如喹唑啉 EGFR 激酶抑制剂、吡啶并嘧啶 EGFR 激酶抑制剂、嘧啶并嘧啶 EGFR 激酶抑制剂、吡咯并嘧啶 EGFR 激酶抑制剂、吡唑并嘧啶 EGFR 激酶抑制剂、苯基氨基-嘧啶 EGFR 激酶抑制剂、羟吡啶 EGFR 激酶抑制剂、indolocarbazole EGFR 激酶抑制剂、酞嗪 EGFR 激酶抑制剂、异黄酮 EGFR 激酶抑制剂、quinalone EGFR 激酶抑制剂、和 tyrphostin EGFR 激酶抑制剂,如记载于以下专利公开文本的那些,和 EGFR 激酶抑制剂的所有药学可接受的盐和溶剂合物:国际专利公开文本 No. WO 96/33980、WO 96/30347、WO 97/30034、WO 97/30044、WO 97/38994、WO 97/49688、WO 98/02434、WO 97/38983、WO 95/19774、WO 95/19970、WO 97/13771、WO 98/02437、WO 98/02438、WO 97/32881、WO 98/33798、WO 97/32880、WO 97/3288、WO 97/02266、WO 97/27199、WO 98/07726、WO 97/34895、WO 96/31510、WO 98/14449、WO 98/14450、WO 98/14451、WO 95/09847、WO 97/19065、WO 98/17662、WO 99/35146、WO 99/35132、WO 99/07701 和 WO 92/20642;欧洲专利申请 No. EP 520722、EP 566226、EP787772、EP 837063 和 EP 682027;美国专利 No. 5,747,498、5,789,427、5,650,415 和 5,656,643;和德国专利申请 No. DE 19629652。低分子量 EGFR 激酶抑制剂的其他非限制性例子包括 Traxler, P., 1998, Exp. Opin. Ther. Patents8(12):1599-1625 中描述的任何 EGFR 激酶抑制剂。

[0127] 可依照本发明使用的低分子量 EGFR 激酶抑制剂的特定优选的例子包括 [6,7-二(2-甲氧基乙氧基)-4-喹唑啉-4-基]-(3-乙炔基苯基)胺(亦称为 OSI-774、厄洛

替尼或 TARCEVA™(厄洛替尼 HCl); OSI Pharmaceuticals/Genentech/Roche)(美国专利 No. 5,747,498; 国际专利公开文本 No. WO 01/34574 和 Moyer, J. D. 等 (1997) *Cancer Res.* 57:4838-4848); CI-1033(之前已知为 PD183805; Pfizer)(Sherwood 等, 1999, *Proc. Am. Assoc. Cancer Res.* 40:723); PD-158780(Pfizer); AG-1478(University of California); CGP-59326(Novartis); PKI-166(Novartis); EKB-569(Wyeth); GW-2016(亦称为 GW-572016 或二甲苯磺酸拉帕替尼 (lapatinib ditosylate); GSK); 和吉非替尼(亦称为 ZD1839 或 IRESSA™; Astrazeneca)(Woodburn 等, 1997, *Proc. Am. Assoc. Cancer Res.* 38:633)。可依照本发明使用的一种特别优选的低分子量 EGFR 激酶抑制剂是 [6,7-二(2-甲氧基乙氧基)-4-喹唑啉-4-基]-(3-乙炔基苯基)胺(即厄洛替尼)、其盐酸盐(即厄洛替尼 HCl, TARCEVA™)、或其他盐形式(例如甲磺酸 (mesylate) 厄洛替尼)。

[0128] 基于抗体的 EGFR 激酶抑制剂包括任何抗 EGFR 抗体或抗体片段, 其能部分或完全地阻断 EGFR 通过其天然配体的活化。基于抗体的 EGFR 激酶抑制剂的非限制性例子包括记载于 Modjtahedi, H., 等, 1993, *Br. J. Cancer* 67:247-253; Teramoto, T., 等, 1996, *Cancer* 77:639-645; Goldstein 等, 1995, *Clin. Cancer Res.* 1:1311-1318; Huang, S. M., 等, 1999, *Cancer Res.* 15:59(8):1935-40; 和 Yang, X., 等, 1999, *Cancer Res.* 59:1236-1243 的那些。如此, EGFR 激酶抑制剂可以是单克隆抗体 Mab E7.6.3(Yang, X. D. 等 (1999) *Cancer Res.* 59:1236-43)、或 Mab C225(ATCC 登录号 HB-8508), 或具有其结合特异性的抗体或抗体片段。适宜的单克隆抗体 EGFR 激酶抑制剂包括但不限于 IMC-C225(亦称为西妥昔单抗或 ERBITUX™; Imclone Systems)、ABX-EGF(Abgenix)、EMD 72000(Merck KgaA, Darmstadt)、RH3(York Medical Bioscience Inc.) 和 MDX-447(Medarex/Merck KgaA)。

[0129] 本发明的方法可扩展至抑制 EGFR 和其他靶物的那些化合物。这些化合物在本文中称为“双特异性抑制剂”。在一个实施方案中, 所述双特异性抑制剂是双特异性 HER3/EGFR、EGFR/HER2、EGFR/HER4 或 EGFR c-Met 抑制剂。在一个实施方案中, 所述双特异性抑制剂是双特异性抗体。在一个实施方案中, 所述双特异性抑制剂是包含特异性结合 EGFR 和第二靶物的抗原结合域的双特异性抗体。在一个实施方案中, 所述双特异性抑制剂是包含特异性结合 HER3 和 EGFR 的抗原结合域的双特异性抗体。在一个实施方案中, 所述双特异性 HER3/EGFR 抑制剂是包含两个等同抗原结合域的双特异性抗体。这类抗体记载于 US 8,193,321、20080069820、W02010108127、US20100255010 和 Schaefer 等, *Cancer Cell*, 20:472-486(2011)。在一个实施方案中, 所述双特异性 HER2/EGFR 是拉帕替尼 / GW572016。

[0130] 可依照已知方法产生另外的基于抗体的抑制剂, 其通过将适宜的抗原或表位施用给选定的宿主动物, 例如猪、牛、马、家兔、山羊、绵羊和小鼠等。可将本领域中已知的各种佐剂用于增强抗体生成。

[0131] 尽管可用于实践本发明的抗体可以是多克隆的, 但优选单克隆抗体。可使用通过培养连续细胞系提供抗体分子生成的任何技术来制备并分离单克隆抗体。用于生成和分离的技术包括但不限于, 最初由 Kohler 和 Milstein 描述的杂交瘤技术 (*Nature*, 1975, 256:495-497); 人 B-细胞杂交瘤技术 (Kosbor 等, 1983, *Immunology Today* 4:72; Cote 等, 1983, *Proc. Natl. Acad. Sci. USA* 80:2026-2030); 和 EBV-杂交瘤技术 (Cole 等, 1985, *Monoclonal Antibodies and Cancer Therapy*, Alan

R. Liss, Inc., pp. 77-96)。

[0132] 或者,可将针对单链抗体生产所描述的技术(参见例如美国专利 No. 4, 946, 778)改为适应于产生具有期望特异性的单链抗体。可用于实践本发明的基于抗体的抑制剂还包括抗体片段,包括但不限于 F(ab')<sub>2</sub> 片段,其可通过对完整抗体分子的胃蛋白酶消化生成,和 Fab 片段,其可通过还原 F(ab')<sub>2</sub> 片段的二硫桥生成。或者,可构建 Fab 和/或 scFv 表达库(参见例如 Huse 等, 1989, Science246:1275-1281)以允许快速鉴定出具有期望特异性的片段。

[0133] 用于生产和分离单克隆抗体和抗体片段的技术是本领域中公知的,且记载于 Harlow 和 Lane, 1988, Antibodies:A Laboratory Manual, Cold Spring Harbor Laboratory 和 J. W. Goding, 1986, Monoclonal Antibodies:Principles and Practice, Academic Press, London。人源化的抗 EGFR 抗体和抗体片段还可依照已知的技术来制备,如记载于 Vaughn, T. J. 等, 1998, Nature Biotech. 16:535-539 的那些和其中引用的参考文献,而且这类抗体或其片段也可用于实践本发明。

[0134] 或者,用于本发明的抑制剂可基于反义寡核苷酸构建体。反义寡核苷酸,包括反义 RNA 分子和反义 DNA 分子,将作用于通过对其结合来直接阻断靶 mRNA 的翻译,并如此预防蛋白质翻译或提高 mRNA 降解,如此降低细胞中靶蛋白的水平,如此降低活性。例如,可合成至少约 15 个碱基且与编码 EGFR 或 HER2 的 mRNA 转录本序列的独特区域互补的反义寡核苷酸,例如通过常规的磷酸二酯技术,并通过例如静脉内注射或输注施用。用于使用反义技术来特异性抑制其序列已知的基因的基因表达的方法是本领域中公知的(例如参见美国专利 No. 6, 566, 135 ;6, 566, 131 ;6, 365, 354 ;6, 410, 323 ;6, 107, 091 ;6, 046, 321 ;和 5, 981, 732)。

[0135] 小抑制性 RNA (siRNA) 还可充当用于本发明的抑制剂。可通过以下来降低靶基因表达:使肿瘤、受试者或细胞与小双链 RNA (dsRNA) 或导致小双链 RNA 产生的载体或构建体接触,从而使靶基因的表达受到特异性抑制(即 RNA 干扰或 RNAi)。针对其序列已知的基因,用于选择适宜 dsRNA 或 dsRNA 编码载体的方法是本领域中公知的(例如参见 Tuschl, T., 等 (1999) Genes Dev. 13(24):3191-3197 ;Elbashir, S. M. 等 (2001) Nature411:494-498 ; Hannon, G. J. (2002) Nature418:244-251 ;McManus, M. T. 和 Sharp, P. A. (2002) Nature Reviews Genetics3:737-747 ;Bremmelkamp, T. R. 等 (2002) Science296:550-553 ;美国专利 No. 6, 573, 099 和 6, 506, 559 ;和国际专利公开文本 No. W001/36646, W099/32619 和 W001/68836)。

[0136] 核酶也可充当用于本发明的抑制剂。核酶是能够催化 RNA 的特异性切割的酶性 RNA 分子。核酶作用机制牵涉核酶分子对互补性靶 RNA 的序列特异性杂交,继之以内切核苷酸切割。由此,特异且有效地催化 mRNA 序列的内切核苷酸切割的工程化的发夹或锤头状基序核酶分子在本发明的范围内可用。初始通过扫描靶分子中的核酶切割位点来鉴定任何潜在 RNA 靶物内的特异性核酶切割位点,所述位点通常包括以下序列:GUA、GUU 和 GUC。一旦鉴定出来,即可评估对应于含切割位点的靶基因区域的约 15 至 20 个核糖核苷酸的短 RNA 序列中的使得该寡核苷酸序列不适宜的预测性结构特征,如二级结构。还可通过测试候选靶物对与互补性寡核苷酸杂交的可达性来评估其适宜性,使用例如核糖核酸酶保护测定法。

[0137] 可通过已知方法来制备可用作抑制剂的反义寡核苷酸和核酶两者。这些包括用于

化学合成的技术,如例如通过固相亚磷酰胺化学合成。或者,反义 RNA 分子可通过编码该 RNA 分子的 DNA 序列的体外或体内转录生成。这类 DNA 序列可掺入很多种载体中,所述载体掺有适宜的 RNA 聚合酶启动子如 T7 或 SP6 聚合酶启动子。可引入对本发明寡核苷酸的各种修饰作为增加胞内稳定性和半衰期的手段。可能的修饰包括但不限于向分子的 5' 和 / 或 3' 端添加核糖核苷酸或脱氧核糖核苷酸的侧翼序列,或在寡核苷酸主链内使用硫代磷酸酯或 2' -O- 甲基而非磷酸二酯连接。

[0138] 在本发明治疗方法的背景中,将抑制剂(如 EGFR 抑制剂)作为包含药学可接受载体和非毒性治疗有效量的 EGFR 激酶抑制剂化合物(包括其药学可接受的盐)的组合物使用。

[0139] 术语“药学可接受的盐”指从药学可接受的非毒性碱或酸制备的盐。当本发明的化合物是酸性时,其相应的盐可从药学可接受的非毒性碱方便地制备,包括无机碱或有机碱。从这类无机碱衍生的盐包括铝盐、铵盐、钙盐、铜盐(二价铜和一价铜)、铁盐、亚铁盐、锂盐、镁盐、锰盐(三价锰和二价锰)、钾盐、钠盐、锌盐等盐。特别优选的是铵盐、钙盐、镁盐、钾盐和钠盐。源自药学可接受的有机非毒性碱的盐包括伯胺、仲胺或叔胺,以及环胺和经取代胺如天然存在的和合成的经取代胺的盐。其他可从其形成盐的药学可接受的有机非毒性碱包括离子交换树脂如例如精氨酸、甜菜碱、咖啡因、胆碱、N', N' - 联苄基乙撑二胺、二乙胺、2- 二乙基氨基乙醇、2- 二甲基氨基乙醇、乙醇胺、乙撑二胺、N- 乙基吗啉、N- 乙基哌啶、葡糖胺(glucamine)、葡糖胺(glucosamine)、组氨酸、hydrabamine、异丙胺、赖氨酸、甲基葡糖胺、吗啉、哌嗪、哌啶、多胺树脂、普鲁卡因、嘌呤、可可碱(theobromine)、三乙胺(triethylamine)、三甲胺、三丙胺、氨丁三醇(tromethamine)等。

[0140] 当用于本发明的化合物是碱性时,其相应的盐可从药学可接受的非毒性酸方便地制备,包括无机或有机酸。这类酸包括例如乙酸、苯磺酸、苯甲酸、樟脑磺酸(camphorsulfonic)、柠檬酸、乙磺酸、富马酸、葡糖酸、谷氨酸、氢溴酸、盐酸、羟基乙磺酸(isethionic)、乳酸、马来酸、苹果酸、扁桃酸、甲磺酸、粘酸、硝酸、双羟萘酸(pamoic)、泛酸、磷酸、琥珀酸、硫酸、酒石酸、对甲苯磺酸等。特别优选的是柠檬酸、氢溴酸、盐酸、马来酸、磷酸、硫酸和酒石酸。

[0141] 包含作为活性成分的抑制剂化合物(包括其药学可接受的盐)的用于本发明的药物组合物可包含药学可接受的载体和任选地其他治疗成分或佐剂。其他治疗剂可包括那些细胞毒性剂、化疗剂或抗癌剂,或增强这类药剂的效果的药剂,如上文所列的。所述组合物包括适用于口服、直肠、局部和胃肠外(包括皮下、肌内和静脉内)施用的组合物,尽管在任意给定情况中的最适宜的路径将依赖于具体的宿主、和施用活性成分所针对的状况的性质和严重性。药物组合物可以单位剂量形式方便地呈现并通过药学领域中公知的任何方法来制备。

[0142] 在实践中,本发明的抑制剂化合物(包括其药学可接受的盐)可作为活性成分与药学载体依照常规的药物复合技术组合到密切的混合物中。根据期望用于施用(例如口服或胃肠外(包括静脉内))的制备物的形式,载体可采用很多种形式。如此,本发明的药物组合物可呈现为适用于口服施用的离散单位,如胶囊、扁囊剂(cachet)或片剂,其各自含有预确定量的活性成分。另外,所述组合物可呈现为粉剂、颗粒剂、溶液、水性液体中的混悬剂、非水性液体、水包油乳剂、或油包水液体乳剂。除了上文列出的常见剂量形式以外,抑制

剂化合物（包括其每种组分的药学可接受的盐）还可通过受控释放手段和 / 或投递装置施用。所述组合组合物可通过任何药学方法来制备。一般地，这类方法包括使活性成分与构成一种或多种必要成分的载体产生联合的步骤。一般地，所述组合物通过将活性成分与液体载体或精细分开的固体载体或两者均一旦密切混合来制备。然后，产物可方便地成形为期望的呈现物。

[0143] 本发明中使用的抑制剂化合物（包括其药学可接受的盐）还可以与一种或多种其他治疗活性化合物组合纳入药学组合物中。其他治疗活性化合物可包括那些细胞毒性剂、化疗剂或抗癌剂，或增强这类药剂效果的药剂，如上文列出的。

[0144] 如此，在本发明的一个实施方案中，所述药物组合物可包含与抗癌剂组合的抑制剂化合物，其中所述抗癌剂是选自下组的成员：烷化药物、抗代谢物、微管抑制剂、鬼臼毒素 (podophyllotoxin)、抗生素、亚硝基脲、激素疗法、激酶抑制剂、肿瘤细胞凋亡的激活剂、和抗血管生成剂。

[0145] 采用的药学载体可以是例如固体、液体或气体。固体载体的例子包括乳糖、石膏粉 (terra alba)、蔗糖、滑石、明胶、琼脂、果胶、阿拉伯树胶 (acacia)、硬脂酸镁和硬脂酸。液体载体的例子是糖浆剂、花生油、橄榄油和水。气态载体的例子包括二氧化碳和氮气。

[0146] 在针对口服剂量形式的组合物的制备中，可采用任何方便的药学介质。例如，水、二醇、油、醇、增味剂、防腐剂、着色剂等可用于形成口服液体制备物如混悬剂、酏剂和溶液；而载体如淀粉、糖、微晶纤维素、稀释剂、颗粒化剂、滑润剂、结合剂、分解剂等可用于形成口服固体制备物如粉剂、胶囊和片剂。由于其易于施用，片剂和胶囊是优选的口服剂量单位，由此采用固体药学载体。任选地，片剂可通过标准的水性或非水性技术包被。

[0147] 含有用于本发明的组合物的片剂可通过压缩或模塑制备，任选地连同一种或多种附属成分或佐剂。压缩的片剂可通过在合适的机器中压缩来制备，活性成分处于自由流动形式如粉剂或颗粒剂，任选地与结合剂、滑润剂、惰性稀释剂、表面活性剂或分散剂混合。模塑的片剂可通过在合适的机器中模塑来制备，将成粉状的化合物的混合物用惰性液体稀释剂加湿。每种片剂优选含有约 0.05mg 至约 5g 的活性成分，每种扁囊剂或胶囊优选含有约 0.05mg 至约 5g 的活性成分。

[0148] 例如，意图口服施用给人的配制剂可含有约 0.5mg 至约 5g 的活性剂，其与适宜且方便量的载体材料复合，所述载体材料的量可在总组合物的约 5 至约 95 个百分数变化。单位剂量形式一般将含有约 1mg 至约 2g 活性成分，通常为 25mg、50mg、100mg、200mg、300mg、400mg、500mg、600mg、800mg、或 1000mg。

[0149] 用在本发明中的适用于胃肠外施用的药物组合物可制备为活性化合物在水中的溶液或混悬剂。可纳入合适的表面活性剂，如例如羟丙基纤维素。分散体还可在甘油、液体聚乙二醇及其混合物中，在油中制备。另外，可纳入防腐剂以阻止微生物的有害生长。

[0150] 在本发明中使用的适合注射使用的药物组合物包括无菌水性溶液或分散体。此外，组合物可以以供临场制备这类无菌可注射溶液或分散体用的无菌粉剂的形式。在所有情况下，最终可注射形式必须是无菌的，且必须为有效流动的从而得到容易的可注射性。所述药物组合物在制造和存储条件下必须是稳定的；如此，优选应保存为免于微生物如细菌和真菌的污染作用。载体可为含有例如水、乙醇、多元醇（例如甘油、丙二醇、和液体聚乙二醇）、植物油、及其适当混合物的溶剂或分散介质。

[0151] 本发明的药物组合物可以以适于局部使用的形式,如例如气雾剂、乳膏、软膏剂、洗剂、粉化散剂(dusting powder)等。另外,所述组合物可以以适用于经皮装置中的形式。这些制剂可利用抑制剂化合物(包括其药学可接受的盐)经由常规处理方法制备。例如,乳膏或软膏剂通过将亲水性材料和水与约5wt%至约10wt%的化合物混合在一起以产生具有期望稠度的乳膏或软膏剂来制备。

[0152] 本发明的药物组合物可以以适用于直肠施用的形式,其中载体是固体。优选混合物形成单位剂量栓剂。合适的载体包括可可脂和本领域中普遍使用的其他材料。可通过以下便利地形成栓剂,首先将组合物与软化或溶解的载体混合,接着在模具中冷却并成形。

[0153] 除了前述载体成分外,上文描述的药物制剂在适宜时还可包含一种或多种其他载体成分如稀释剂、缓冲剂、增味剂、结合剂、表面活性剂、增稠剂、滑润剂、防腐剂(包括抗氧化物)等。此外,可纳入其他佐剂以使得所述制剂与意图受体的血液等渗。还可以粉剂或液体浓缩物形式制备含有抑制剂化合物(包括其药学可接受的盐)的组合物。

[0154] 用于实践本发明的化合物的剂量水平将大概如本文中描述的,或如本领域中对这些化合物描述的。然而,理解对任何具体患者的特定剂量水平将依赖于多种因素,包括年龄、体重、一般健康、性别、饮食、施用时间、施用路径、排泄速率、药物组合和进行治疗的具体疾病的严重性。

[0155] 本领域中已知的许多备选的实验方法可成功地取代在实践本发明时本文中具体描述的那些,如例如记载于本发明相关技术领域中可获的许多优秀的手册和教科书中的(例如 Using Antibodies, A Laboratory Manual, Harlow, E. 和 Lane, D. 编, 1999, Cold Spring Harbor Laboratory Press, (例如 ISBN0-87969-544-7); Roe B. A. 等 1996, DNA Isolation and Sequencing (Essential Techniques Series), John Wiley&Sons. (例如 ISBN0-471-97324-0); Methods in Enzymology: Chimeric Genes and Proteins”, 2000, ed. J. Abelson, M. Simon, S. Emr, J. Thorner. Academic Press; Molecular Cloning: a Laboratory Manual, 2001, 3rd Edition, 由 Joseph Sambrook 和 Peter MacCallum, (以前的 Maniatis Cloning manual) (例如 ISBN0-87969-577-3); Current Protocols in Molecular Biology, Ed. Fred M. Ausubel, 等 John Wiley&Sons (例如 ISBN0-471-50338-X); Current Protocols in Protein Science, Ed. John E. Coligan, John Wiley&Sons (例如 ISBN0-471-11184-8); 和 Methods in Enzymology: Guide to protein Purification, 1990, Vol. 182, Ed. Deutscher, M. P., Academic Press, Inc. (例如 ISBN0-12-213585-7)), 或如致力于描述分子生物学中的实验方法的许多大学和商业性网站上描述的。

[0156] 医学领域中的普通技术人员将领会,在诊断患者对抑制剂的可能的响应性之后,对患者施用治疗有效量的如本文中描述的抑制剂(例如 EGFR 激酶抑制剂、双特异性 EGFR 激酶抑制剂或 HER2 抑制剂)的确切方式将由主治医师的判断而定。施用的模式,包括剂量、与其他抗癌剂的组合、施用时机和频率等可能受到患者对抑制剂的可能响应性的诊断,以及患者的状况和病史的影响。如此,甚至诊断为患有预测对抑制剂类型相对不敏感的肿瘤的患者仍可能受益于使用这类抑制剂治疗,特别是在与其他抗癌剂或可能改变对抑制剂的肿瘤敏感性的药剂组合时。

[0157] 就本发明而言,抑制剂与其他抗癌剂“的共施用”和将抑制剂与其他抗癌剂“共施

用”(两种组分在后文中均称为“两种活性药剂”)指两种活性药剂的分开或一起的任何施用,其中两种活性药剂作为设计为获得组合疗法益处的适宜剂量方案的部分施用。如此,两种活性成分可作为同一药物组合物的部分施用,或在分别的药物组合物中施用。所述其他药剂可在抑制剂施用之前、同时或之后施用,或以其一些组合施用。当以重复间期对患者施用抑制剂时,例如在标准的治疗过程期间,所述其他药剂可在每次抑制剂施用之前、同时或之后,或以其一些组合,或相对于抑制剂治疗以不同的间期,或以单一剂量在使用抑制剂的治疗过程之前、期间任何时间、或之后施用。

[0158] 所述抑制剂通常将以对所治疗患者提供最有效癌症治疗(从功效和安全性角度两者)的剂量方案施用给患者,如本领域中已知的,和如披露在例如国际专利公开文本 No. WO 01/34574 中的。在进行本发明的治疗方法中,所述抑制剂可以本领域中已知的任何有效方式施用,如通过口服、局部、静脉内、腹膜内、肌内、关节内、皮下、鼻内、眼内、阴道、直肠或皮内路径,根据所治疗的癌症类型、所使用的抑制剂类型(例如,小分子、抗体、RNAi、核酶或反义构建体)、和处方医师的医学判断,如例如基于公开的临床研究结果。

[0159] 施用的抑制剂的量和抑制剂施用的时机将取决于所治疗患者的类型(种、性别、年龄、重量等)和状况、所治疗疾病或状况的严重性、和施用路径。例如,小分子抑制剂可以范围从 0.001 至 100mg/kg 体重每天或每周的剂量以单一或分开的剂量或通过连续输注施用给患者(参见例如,国际专利公开文本 No. WO 01/34574)。具体地,厄洛替尼 HCl 可以范围从 5-200mg 每天或 100-1600mg 每周的剂量以单一或分开的剂量或通过连续输注施用给患者。优选的剂量是 150mg/天。基于抗体的抑制剂或反义、RNAi 或核酶构建体可以范围从 0.1 至 100mg/kg 体重每天或每周的剂量以单一或分开的剂量或通过连续输注施用给患者。在一些情况中,在前述范围下限之下的剂量水平可以是更适宜的,而在其他情况中,可采用仍更大的剂量而不导致任何有害的副作用,只要这类更大的剂量首先分成几个小剂量以供在整天施用。

[0160] 所述抑制剂和其他另外的药剂可通过相同或不同路径分开或一起施用,且以很多种不同的剂量形式施用。例如,所述抑制剂优选口服或胃肠外施用。当抑制剂是厄洛替尼 HCl (TARCEVA™) 时,优选口服施用。抑制剂和其他另外的药剂均可以单一或多剂量施用。

[0161] 所述抑制剂可与各种药学可接受的惰性载体以片剂、胶囊、锭剂(lozenge)、圆形锭剂(troche)、硬糖(hard candies)、粉剂、喷雾剂、乳膏、软膏(salve)、栓剂、凝胶剂(jellies)、凝胶(gel)、糊剂、洗剂、软膏剂、酞剂、糖浆剂等的形式施用。这类剂量形式的施用可以单剂量或多剂量进行。载体包括固体稀释剂或填充剂、无菌水性介质和各种无毒有机溶剂等。口服药物组合物可适宜地加甜和/或增味。

[0162] 所述抑制剂可以与各种药学可接受的惰性载体组合在一起以喷雾剂、乳膏、软膏(salve)、栓剂、凝胶剂(jellies)、凝胶(gel)、糊剂、洗剂、软膏剂等的形式。这类剂量形式的施用可以单剂量或多剂量实施。载体包括固体稀释剂或填充剂、无菌水性介质和各种无毒有机溶剂等。

[0163] 应选择所有包含蛋白质性抑制剂的制剂以避免抑制剂生物学活性的变性和/或降解和损失。

[0164] 制备包含抑制剂的药学组合物的方法是本领域中已知的,且记载于例如国际专利公开文本 No. W001/34574。就本发明的教导而言,制备包含抑制剂的药学组合物的方

法从上文引用的公开出版物和其他已知的参考文献,如 Remington's Pharmaceutical Sciences, Mack Publishing Company, Easton, Pa., 18<sup>th</sup> edition (1990), 将是显而易见的。

[0165] 对于抑制剂的口服施用,将含有一种或两种活性药剂的片剂与多种赋形剂中的一种例如,微晶纤维素、柠檬酸钠、碳酸钙、磷酸二钙和甘氨酸,连同各种分解剂如淀粉(优选玉米、马铃薯或木薯淀粉)、海藻酸和某些复合硅酸盐,与颗粒化结合剂如聚乙烯吡咯烷酮、蔗糖、明胶和阿拉伯树胶一起组合。另外,滑润剂如硬脂酸镁、月桂醇硫酸钠和滑石对于形成片剂的目的经常极为有用。还可以采用类似类型的固体组合物作为凝胶胶囊中的填充剂;在这一点上,优选的材料还包括乳糖(lactose)或乳糖(milk sugar)以及高分子量聚乙二醇。当期望将水性混悬剂和/或酞剂用于口服施用时,抑制剂可与各种增甜剂或增味剂、着色物质或染料和(若期望的)乳化剂和/或助悬剂,以及这类稀释剂如水、乙醇、聚乙二醇、甘油和其各种类似组合一起组合。

[0166] 对于一种或两种活性剂的胃肠外施用,可采用在芝麻油或花生油或水性丙二醇中的溶液,以及无菌水性溶液,其包含活性剂或其相应的水溶性盐。这类无菌水性溶液优选是经过适宜缓冲的,且还优选例如用充足的盐水或葡萄糖使之成为等渗的。这些具体的水性溶液尤其适用于静脉内、肌内、皮下和腹膜内注射目的。油性溶液适用于关节内、肌内和皮下注射目的。所有这些溶液在无菌条件下的制备是通过本领域技术人员公知的标准药学技术容易实现的。任何选定用于施用蛋白质性抑制剂的胃肠外制剂应选择为避免所述抑制剂生物学活性的变性和损失。

[0167] 另外,依照标准的药学实践,可能局部施用一种或两种所述活性成分,通过例如乳膏、洗剂、凝胶剂(jellies)、凝胶(gel)、糊剂、软膏剂、软膏(salve)等。例如,可以制备包含约0.1% (w/v) 至约5% (w/v) 浓度的抑制剂的局部制剂。

[0168] 对于兽医用目的,可使用上文所述任何形式并通过任何路径将所述活性成分分开或一起地施用给动物。在一个优选的实施方案中,所述抑制剂以胶囊、推注、片剂、大剂量药液(liquid drench)的形式通过注射或作为植入物施用。一个备选方案是,所述抑制剂可与动物饲料一起施用,而出于此目的可制备浓缩的饲料添加剂或预混合物用于正常的动物给料。这类制剂以常规方式依照标准的兽医实践制备。

[0169] 医学领域,特别是关于诊断性测试的应用和治疗学治疗的医学领域中的技术人员将认可生物学系统可展现出变异性且并非总是完全可预测的,如此许多较好的诊断性测试或治疗学有时是无效的。如此,最终由主治医师基于测试结果、患者状况和病史、及其自身经验的判断来确定个体患者的最适宜的治疗过程。甚至有以下情况,例如,当基于来自诊断测试或来自其他标准的数据肿瘤未预测为对EGFR激酶抑制剂特别敏感时,医师会选择用EGFR抑制剂来治疗患者,尤其是在如果所有或大多数其他明显的治疗选项已经失败,或者预期在与另一种治疗一起施用时有某些协同性的情况下。与癌症治疗中使用的许多其他抗癌药物如更传统的化疗剂或细胞毒性剂相比,EGFR抑制剂作为一类相对耐受较好的药物的事实使其成为更可行的选择。

[0170] 做广告的方法

[0171] 本文中的发明还涵盖一种用于为EGFR抑制剂或其药学可接受的组合物做广告的方法,其包括给目标受众宣传抑制剂或其药物组合物治疗患有特征为指示上皮样肿瘤的甲基化模式的癌症类型的患者群体的用途,或给目标受众宣传抑制剂或其药物组合物不用来

治疗患有特征为指示间充质样肿瘤的甲基化模式的癌症类型的患者群体的用途。

[0172] 做广告为经由非个人媒体进行的、通常付费的通讯，其中发起人受到鉴别且信息受到控制。为本文目的的做广告包括宣传、公共关系、产品布置、赞助、保险 (underwriting)、和促销。该术语还包括出现在任何印刷传播媒体中的赞助的信息公告，其设计用于引起大众的兴趣以劝说、通知、宣传、激发、或以其它方式改变行为，向购买、支持、或认可本文中发明的有利方式发展。

[0173] 本文中诊断方法的广告和宣传可通过任何手段实现。用于传递这些信息的广告媒体的例子包括电视、电台、电影、杂志、报纸、因特网、和告示板，包括商业性的，即出现在广播媒体中的信息。广告还包括那些在食品货车座位上、机场通道墙壁上、和公共汽车侧面上的广告、或电话等候信息或店内布告 (PA) 系统中听到的广告、或可放置视觉或听觉通讯的任何地方的广告。

[0174] 宣传或做广告手段的更特定的例子包括电视、电台、电影、因特网 (诸如网络传播和网络研讨会 (webinar))、意图到达同步用户的交互式计算机网络、固定的或电子的告示板和其它公共标牌、海报、传统的或电子的文献 (诸如杂志和报纸)、其它媒体渠道、讲座或个体接触，例如通过电子邮件、电话、即时信息、邮寄、快递、大众 (mass)、或承运人邮件 (carrier mail)、亲自访问等进行。

[0175] 所使用的做广告的类型会取决于许多因素，例如待传达的目标受众的性质，例如医院、保险公司、诊所、医生、护士、和患者，以及成本考虑和管理药物和诊断剂做广告的相关管辖权法律和条例。可根据由服务相互作用和 / 或其它数据 (诸如用户人口统计状况和地理学定位) 限定的用户表征使做广告个性化或用户化。

[0176] 表

[0177] 表 1 与间充质表型有关的甲基化胞嘧啶核苷酸

[0178]

基因	染色体	位置	基因	染色体	位置
PON2	7	94888497	TBCD	17	78440559
	1	113544125	TBCD	17	78440498
BET1	7	93459766	TBCD	17	78440426
	X	48900705	MYST1, PRSS8	16	31050024
	X	48900845	ARHGEF38	4	106693255
	X	48900694		1	27023897
SCNN1A	12	6353969	LIMA1	12	48882614
SCNN1A	12	6354033		7	80389667
SCNN1A	12	6354000	KIAA0182	16	84236385
ELMO3	16	65791484		19	49971610
ELMO3	16	65791362		19	49971605
NRBP1, KRTCAP3	2	27519047	ITGB6	2	160822102
NRBP1, KRTCAP3	2	27519011	LOC643008, RECQL5	17	71147845
KRTCAP3	2	27519215	LOC643008, RECQL5	17	71147779
KRTCAP3	2	27519142	CCDC57	17	77655395
NRBP1, KRTCAP3	2	27518810		7	155407896
NRBP1, KRTCAP3	2	27518521		7	155407740
NRBP1, KRTCAP3	2	27518632		7	155407629
NRBP1, KRTCAP3	2	27518654		16	86850497
NRBP1, KRTCAP3	2	27518645		16	86850474
NRBP1, KRTCAP3	2	27518643		16	29204205
NRBP1, KRTCAP3	2	27518583		16	29204115
MST1R	3	49914707		16	29204298
SLC9A7	X	46499386		16	29204194
LYN	8	57066177		7	2447019
ACAP2	3	196640585		7	2447061
TBC1D14	4	7008013	TMEM79	1	154520773
PITPNM3	17	6396092	LOC254559	15	87723993
	10	11963508	CCDC19	1	158137355
	1	41738700	CCDC19	1	158137539
ARHGAP39	8	145777560		4	129368833
ARHGAP39	8	145777354		1	24156244
COX10	17	14050396		3	135552584
	7	27744012	CAMK2G	10	75302072
COL18A1, SLC19A1	21	45757802		2	74006703
RAB25	1	154297806		2	74006825
CGN	1	149753111		2	74006594
TBCD	17	78440835		2	74006721
TBCD	17	78440951	PPP1R13L	19	50595498
TBCD	17	78440786		3	49919159

[0179] 表1(续上)与间充质表型有关的甲基化胞嘧啶核苷酸

[0180]

基因	染色体	位置	基因	染色体	位置
VPS37C	11	60682632	FURIN	15	89213122
NA, RCC1	1	28726284	BRE	2	28261767
CTNND1	11	57305264		11	67106174
EPHB2	1	23025895		11	67106217
	6	134742250		11	67106166
FRMD6	14	51101498		11	67105913
GRHL2	8	102576558		11	67105885
P2RY6	11	72658514	SLC44A2	19	10596482
VTIIA	10	114516704	SLC44A2	19	10596548
SI00A14	1	151855406	SLC44A2	19	10596578
SI00A14	1	151855551	SLC44A2	19	10596594
PRSS8	16	31054183	RNF144A	2	7089748
THSD4	15	69416262		1	201755387
	2	189266325	QSOX1	1	178404541
SIPA1L1	14	71183551	CCDC85C	14	99114910
ARL13B, STX19	3	95230100	PLA2G4F	15	40236307
ARL13B, STX19	3	95230218	PLA2G4F	15	40236078
PVRL4	1	159326278	FTO	16	52372627
PVRL4	1	159326159		1	234153824
PVRL4	1	159326053	PPFIBP2	11	7578132
PVRL4	1	159326082	NINJ2	12	587066
ARHGAP32	11	128399061		2	30294748
ARHGAP32	11	128399150		4	189558130
	8	125219633		4	189558238
	15	76213522		1	206105036
PNKD	2	218868246	KRT8	12	51586560
CD44	11	35152089		4	185905387
ANKRD22	10	90600502	LIMK2	22	30001702
CEACAM19	19	49866511	BOLA2, GDPD3	16	30023636
CEACAM19	19	49866752		3	129636934
CEACAM19	19	49866521		4	154136934
	11	71134595		9	131184926
	11	71134808		19	1855554
SCYL3	1	168127429		8	102519033
CPA4	7	129749798		1	100204254
CLUAP1	16	3499552	IMMP2L	7	110988180
CLUAP1	16	3499688		19	60699327
CLUAP1	16	3499569	PLEKHG6	12	6292029
	8	28514725	PLEKHG6	12	6292067
ASAP2	2	9458210	JMJD7-PLA2G4B	15	39918027

[0181] 表 1 (续上) 与间充质表型有关的甲基化胞嘧啶核苷酸

[0182]

基因	染色体	位置	基因	染色体	位置
JMJD7-PLA2G4B	15	39917942		2	216504192
JMJD7-PLA2G4B	15	39917954	PLEKHF1	19	34854611
PNPLA8	7	107955947	PLEKHF1	19	34854406
PNPLA8	7	107955918		10	6202160
PNPLA8	7	107955957		10	6202194
HIVEP3	1	41753450		10	6202124
RAI1	17	17572988	SH3KBP1	X	19812050
DIXDC1	11	111337086		10	28996094
BOLA2, TBX6	16	30009181		11	73806817
SEMA3A	7	83655334		8	82806235
	2	27838190		11	354805
TNFAIP8	5	118637864		11	354809
SNX8	7	2267181		11	354623
JARID2	6	15564181		11	354752
AHRR	5	439883		10	31428642
CDH5	16	64970341		17	52465903
CDH5	16	64970599		12	15846291
	6	8381465	TESK2	1	45587219
SLC35B3	6	8381262		1	2456135
	6	8381295	TSEN54	17	71032445
DAGLA	11	61221452	TSEN54	17	71032354
	19	2105603	ACOT2	14	73109663
SVOPL	7	137999314	PDGFRA, LNX1	4	54153003
	17	8252319	PDGFRA, LNX1	4	54152866
	17	8252561	SLC40A1	2	190154739
	17	8252360	ATL1	14	50069808
	17	8252425	ZNF398	7	148472457
IGF1R	15	97074397		17	37862949
WDR82	3	52277292		17	37862906
WDR82	3	52277190		4	40328026
FBXO34	14	54834400		2	41940393
RAB11FIP1	8	37868570	AFF1	4	88113322
VPS37B	12	121944095	INPP5A	10	134254904
NAV2	11	19732081	INPP5A	10	134254935
C4orf36	4	88031692	MST1R	3	49913008
PLXNB2	22	49062415	PHGDH	1	120075342
PLXNB2	22	49062595	GLI2	2	121266304
C19orf46	19	41191166	GLI2	2	121266336
	2	70222288	GLI2	2	121266195
VTG1A	10	114492308	C2orf54	2	241484135

[0183] 表 1 (续上) 与间充质表型有关的甲基化胞嘧啶核苷酸

[0184]

基因	染色体	位置	基因	染色体	位置
C2orf54	2	241484343	TBCD	17	78426815
	6	112413022	TBCD	17	78426927
	4	100956161	TBCD	17	78427378
CCNY	10	35716880	TBCD	17	78427517
MLPH	2	238063229		2	64687610
CDKAL1	6	21131464		16	52025113
GPR81	12	121777086	PPARD	6	35417845
	17	41697531		8	144892671
F11R	1	159258976		8	144892697
F11R	1	159258982		8	144892896
CDC42SE2	5	130692428		8	144892814
FTO	16	52472915	LRP5	11	67866681
	10	73752495	XAB2	19	7590468
MYO18A	17	24529971	RAP1GAP2	17	2815637
MYO18A	17	24529383	SLC37A1	21	42809566
DGATI	8	145518703		13	109313445
SDCBP2	20	1258000		12	13179838
SDCBP2	20	1257800	OFCC1	6	10271555
SDCBP2	20	1257722	PTK7	6	43172438
TRAK1	3	42147101	TEAD3	6	35562412
SCNN1A	12	6354480	TEAD3	6	35562047
SCNN1A	12	6354974	TEAD3	6	35561916
SCNN1A	12	6354868	C16orf72	16	9097893
SCNN1A	12	6354990	ARID1A	1	26953185
SCNN1A	12	6354782	SGK223	8	8276184
ZCCHC14	16	86078911	GNA12	7	2739598
ZCCHC14	16	86078864	GNA12	7	2739653
GLIS1	1	53831204	GNA12	7	2739536
TSPAN1	1	46418811	PWWP2B	10	134072208
TSPAN1	1	46418555	PWWP2B	10	134072043
TSPAN1	1	46418745	SMARCD2	17	59270462
ST3GAL2	16	68973602	GPR56	16	56211203
ST3GAL2	16	68973365	GPR56	16	56211170
C10orf95	10	104201478	GPR56	16	56211418
C10orf95	10	104201378	GPR56	16	56211405
C10orf95	10	104201309	GPR110	6	47117696
C10orf95	10	104201286	GPR110	6	47118136
C10orf95	10	104201414	GPR110	6	47118050
C10orf95	10	104201318	EHF	11	34599461
TBCD	17	78426682		21	38521991

[0185] 表 1 (续上) 与间充质表型有关的甲基化胞嘧啶核苷酸

[0186]

基因	染色体	位置	基因	染色体	位置
	14	64792338	CDS1	4	85724598
NSMCE2	8	126223268	GNAI3	1	109914827
PPCDC	15	73115889	NCOA2	8	71402682
WISP1	8	134293893		12	103938284
WISP1	8	134294072	CPEB3	10	93872825
WISP1	8	134293996	TACC2	10	123744125
	17	36932205		1	227296135
	8	144727414		6	7477665
CHD2	15	91266091		19	50356091
	1	8053252	LLGL2	17	71057739
DDR1	6	30959396	ANKFY1	17	4098222
DDR1	6	30958847	CLDN7	17	7106144
DDR1	6	30958892		1	59052878
DDR1	6	30958855		17	75403317
DDR1	6	30959065		17	75403479
DDR1	6	30959030		16	66828383
DDR1	6	30959048	ESRP2	16	66826542
DDR1	6	30958956	ESRP2	16	66826796
BAIAP2	17	76626137	OVOL1	11	65310618
BAIAP2	17	76625735		8	95720275
BAIAP2	17	76625947	FAM110A	20	770788
BAIAP2	17	76625872	SPINT1	15	38924311
MANF	3	51401417	GRHL2	8	102575162
PVRL4	1	159325891	SH3YL1	2	253559
PVRL4	1	159325951	SH3YL1	2	253656
RHOBTB3	5	95089583	TMEM159, DNAH3	16	21078740
	2	70221961	TMEM159, DNAH3	16	21078585
GPR56	16	56211848	TMEM159, DNAH3	16	21078568
RAB25	1	154297433	TMEM159, DNAH3	16	21078598
RAB25	1	154297468	Clorf210	1	43524150
	3	53164930	Clorf210	1	43523857
RAB24	5	176661226	Clorf210	1	43524084
SPINT1	15	38925452	Clorf210	1	43523963
RAB24	5	176661618	Clorf210	1	43523950
	8	8356184	Clorf210	1	43524056
	20	36661934	Clorf210	1	43524091
	1	113106832	Clorf210	1	43523957
CHD3	17	7732607	CLDN7	17	7105979
ABCF1	6	30667066	CLDN7	17	7105734
	16	83945057	CLDN7	17	7106573

[0187] 表 1 (续上) 与间充质表型有关的甲基化胞嘧啶核苷酸

[0188]

基因	染色体	位置	基因	染色体	位置
CLDN7	17	7106633		1	1088914
CLDN7	17	7106571		1	1088855
CLDN7	17	7106564	AGAP3	7	150443215
CLDN7	17	7106566	ARHGEF1	19	47084177
CLDN7	17	7106555		4	100955681
GRHL2	8	102575727	ARHGAP39	8	145777081
GRHL2	8	102575565	STX2	12	129869431
GRHL2	8	102575811	STX2	12	129869200
GRHL2	8	102574732	STX2	12	129869047
GRHL2	8	102574469	STX2	12	129869147
GRHL2	8	102574689	STX2	12	129868969
TMEM30B	14	60817996		22	35136360
TMEM30B	14	60818107		22	35136601
TMEM30B	14	60818193		22	35136526
TMEM30B	14	60818089		22	35136389
PDGFRA, LNX1	4	54152685	CLDN15	7	100662856
PDGFRA, LNX1	4	54152402	E2F4, ELMO3	16	65790422
PDGFRA, LNX1	4	54152494	E2F4, ELMO3	16	65790778
PDGFRA, LNX1	4	54152599	ELMO3	16	65790933
PDGFRA, LNX1	4	54152503	PTPRF	1	43788610
GRHL2	8	102573922	PTPRF	1	43788636
GRHL2	8	102574035	PTPRF	1	43788601
GRHL2	8	102573658	PWWP2B	10	134071493
GRHL2	8	102573623	PWWP2B	10	134071845
GRHL2	8	102573655	PWWP2B	10	134071623
GRHL2	8	102573797		14	64239711
GRHL2	8	102573842		14	64239802
GRHL2	8	102573677	ETV6	12	11922571
GRHL2	8	102573740	SH3BP5	3	15344685
	4	124687980	GAS8	16	88638299
	4	124687986	SULT2B1	19	53747224
	4	124688290	SULT2B1	19	53747250
	1	117976728	SULT2B1	19	53747255
	1	1088243	SULT2B1	19	53747202
	1	1089514	SULT2B1	19	53747244
	1	1089426	LAMA3	18	19707129
	1	1089493	LAMA3	18	19706893
	1	1089446	LAMA3	18	19706728
	1	1088763	LAMA3	18	19706786
	1	1089029	LAMA3	18	19706817

[0189] 表 1 (续上) 与间充质表型有关的甲基化胞嘧啶核苷酸

[0190]

基因	染色体	位置	基因	染色体	位置
LAMA3	18	19706827	APBB1	11	6375542
LAMA3	18	19706842	ABCA7	19	1016712
NCRNA00093, DNMBP	10	101680658	ABCA7	19	1016728
C1orf106	1	199130846	ABCA7	19	1016688
C1orf106	1	199130930		11	66580163
	12	6942075		11	66580192
	12	6941973		11	66580260
	12	6943440	ANK3	10	62162307
	12	6943501	ANK3	10	62161917
	12	6943503	ANK3	10	62162163
	12	6943508	ABLIM1	10	116269176
	12	6943525		14	53642609
	12	6942988	XDH	2	31491126
	12	6943026	XDH	2	31491353
	12	6943152	DAPP1	4	100957034
	12	6942957	DAPP1	4	100956844
TALDOI	11	753339	DAPP1	4	100956853
TALDOI	11	753485	TNS4	17	35911401
CNKSRI	1	26376210	TNS4	17	35911460
CNKSRI	1	26376363	TNS4	17	35911441
CNKSRI	1	26376365	TNS4	17	35911475
CNKSRI	1	26376449	PARD3	10	34756309
CNKSRI	1	26376445	RGL2	6	33373111
CNKSRI	1	26376434	RGL2	6	33373221
CNKSRI	1	26376520	RGL2	6	33373245
CNKSRI	1	26376566		19	17763242
CNKSRI	1	26376606		1	150076158
CNKSRI	1	26376578	PCCA	13	99941258
	3	37200270	RAP1GAP2	17	2855119
MERTK	2	112421048	EPHB3	3	185766002
RGS3	9	115383006	TNFRSF10C	8	23019312
PLXNB2	22	49062679	MICAL2	11	12226862
PLXNB2	22	49062940	SGSM2	17	2197812
	16	86381426	RABGAPIL	1	173111020
	10	75306867	RABGAPIL	1	173111113
FAM83A	8	124264314	ARHGEF10L	1	17750038
FAM83A	8	124264583	TBC1D1	4	37666838
FAM83A	8	124264373	CGN	1	149751930
TAF1B	2	9955012	ELF3	1	200243703
ERI3	1	44566745	PROM2	2	95304202

[0191] 表 1 (续上) 与间充质表型有关的甲基化胞嘧啶核苷酸

[0192]

基因	染色体	位置	基因	染色体	位置
PROM2	2	95304432	EPN3	17	45966053
PROM2	2	95303758		2	128333798
PROM2	2	95303838	GJB3	1	35020553
PHEX	X	22046472	C10orf91	10	134111645
ADAP1	7	952365	C10orf91	10	134111403
ADAP1	7	952156	C10orf91	10	134111470
ADAP1	7	952310		20	30796655
ADAP1	7	952245	DLEU1	13	49829837
ADAP1	7	952140		8	101497819
VCL	10	75485630		22	28307949
	11	67206458		22	28308158
	11	67206243		16	86536340
	14	51288831	UNC5A	5	176181830
	14	51288704		4	154076297
	21	36592419		4	154075997
	14	34872148		4	154075953
PLA2G4F	15	40236052	USP43	17	9491021
	1	201096568	USP43	17	9490981
FAM46B	1	27207475	USP43	17	9490898
OPA3	19	50723356	USP43	17	9490862
	11	3454830	CXCL16	17	4588805
	6	36205670	CXCL16	17	4588796
CST6	11	65535543		7	139750442
FGGY	1	59989219		7	139750014
	15	72463284		7	139750140
FUT3	19	5802616		7	139750233
FUT3	19	5802465		7	139750195
FUT3	19	5802504		7	139750252
PLS3	X	114734137		7	139750206
WWC1	5	167725172		7	139750225
	8	15408729	CLDN4	7	72883980
RASA3	13	113862226	ARAP1, STARD10	11	72169819
ST3GAL4	11	125781207		9	131185398
ST3GAL4	11	125781216	CDKN1A	6	36758711
	12	104024772	MBP	18	72930014
IL17RE, CIDEA	3	9919512	ERBB2	17	35115639
IL17RE, CIDEA	3	9919537	C14orf43	14	73281541
SIGIRR, ANO9	11	407907	MED16	19	834879
SYT8	11	1812460		2	101234948
SYT8	11	1812427		2	101234788

[0193] 表 1 (续上) 与间充质表型有关的甲基化胞嘧啶核苷酸

[0194]

基因	染色体	位置	基因	染色体	位置
IL10RB	21	33563377	MACC1	7	20223703
ESRP2	16	66825963	MACC1	7	20223521
ESRP2	16	66825753	MACC1	7	20223687
SPINT2	19	43448222		1	27160055
CCDC120	X	48803602	ST14	11	129535669
CCDC120	X	48803499	ST14	11	129535471
C19orf46, ALKBH6	19	41191679	SPINT1	15	38923139
C19orf46, ALKBH6	19	41191561	SPINT1	15	38923085
C19orf46, ALKBH6	19	41191506	SPINT1	15	38923161
CLDN7	17	7105010	SPINT1	15	38923192
PRSS8	16	31054518	C1orf172	1	27159869
PRSS8	16	31054678		2	74064398
PRSS8	16	31054500		2	74064468
PRSS8	16	31054545		2	74064365
PRSS8	16	31054555		8	102573146
PRSS8	16	31054700		8	102573120
	2	238165128		8	102573068
ANKRD22	10	90601891	POU6F2	7	39022925
ANKRD22	10	90601835	LAMB3	1	207892566
ITGB6	2	160764766	LAMB3	1	207892295
ITGB6	2	160764885	LAMB3	1	207892301
ITGB6	2	160764846	LAMB3	1	207892354
BOK	2	242150379	LAMB3	1	207892472
TMC8, TMC6	17	73640271	LAMB3	1	207892370
TMC8, TMC6	17	73640278	LAMB3	1	207892479
CRB3	19	6415885		3	129911695
EPS8L1	19	60279005		16	2999774
EPS8L1	19	60278851	BMF	15	38186296
	12	88144203	BMF	15	38186393
	7	64096077	BMF	15	38186423
KIAA0247	14	69194460	GALNT3	2	166357860
	14	64239962		8	144893629
	5	74369044		8	144893700
	16	11613936	C20orf151	20	60435990
NEURLIB	5	172048817	C20orf151	20	60436252
CLDN4	7	72882009	C20orf151	20	60436261
PAK4	19	44350154	C20orf151	20	60436106
P2RY2	11	72616798	C20orf151	20	60436134
	4	69806346	C20orf151	20	60436052
MACC1	7	20223945			

[0195] 表 1 (续上) 与间充质表型有关的甲基化胞嘧啶核苷酸

[0196]

基因	染色体位置
ADAMTS16	hg19 hr5:5139160-5139859
ANKRD34A	hg19 chr1:145472863-145473562
ARID5A	hg19 chr2:97215439-97216138
APC2	hg19 chr19:1467602-1468301
BMP4	hg19 chr14:54422575-54423274
CA12	hg19 chr15:63673688-63674360
CCK	hg19 chr3:42306174-42306873
CCNA1	hg19 chr13:37005581-37006453
CDH4	hg19 chr20:59826862-59827561
CLDN7	hg19 chr17:7165943-7166642
DKK1	hg19 chr10:54072931-54073630
SEPTIN9	hg19 chr17:75404213-75404912
DLX1	hg19 chr2:172950047-172950746
ERBB4	hg19 chr2:213402181-213402880
ESRP1	hg19 chr8:95651545-95652244
FGFR1	hg19 chr8:38279279-38279921
FOXA1	hg19 chr14:38061638-38062337
GATA2	hg19 chr3:128202381-128203080
GNE	hg19 chr10:54072931-54073630
GRHL2	hg19 chr8:102504509-102505208
GLI3	hg19 chr7:42267369-42268068
HDAC4	hg19 chr2:240113948-240114647
HOXA10	hg19 chr7:27213776-27214475
HS3TS3B1	hg19 chr17:14202839-14203538
ID2	hg19 chr2:8823406-8824105
ITIH4	hg19 chr3:52854493-52855192
LAMA1	hg19 chr18:7013604-7014303
LAD1	hg19 chr1:201368681-201369380
LHX9	hg19 chr1:197889343-197890042
MAP6	hg19 chr11:75378150-75378849
MEOX1	hg19 chr17:41738845-41739544
MGC45800	hg19 chr4:183061951-183062650
MSX1	hg19 chr4:4859635-4860334
MTMR7	hg19 chr8:17270755-17271454
PARD3	hg19 chr10:35104748-35105447
PAX6	hg19 chr11:31833994-31834693
PCDHGA8	hg19 chr5:140807001-140807700
PI3KR5	hg19 chr17:8798216-8798915
RNF220	hg19 chr1:44883347-44884046
RNLS	hg19 chr10:90342854-90343553
RPS6KA2	hg19 chr6:167177930-167178629
SFRP1	hg19 chr8:41167914-41168613

[0197] 表 1 (续上) 与间充质表型有关的甲基化胞嘧啶核苷酸

[0198]

基因	染色体位置
WNT5B	hg19 chr12:1739567-1740266
MEOX2	hg19 chr7:15727091-15727790
TP73	hg19 chr1:3569053-3569719
RASGRF1	hg19 chr15:79381517-79382216
TWIST	hg19 chr7:19157773-19158472
AGAP3	hg18 chr7:150442790-150443639
ANKRD33B	hg18 chr5:10617913-10618612
ARHGEF1	hg18 chr19:47083827-47084526
C10orf91	hg18 chr10:134111053-134111752
CHD3	hg18 chr17:7732182-7733031
CXCL16	hg18 chr17:4588455-4589154
ESRP2	hg18 chr16:66828033-66828732
KIAA1688	hg18 chr8:145777004-145777703
TBC1D1	hg18 chr4:37654711-37655410
SERPINB5	hg18 chr18:59295387-59296621
STX2	hg18 chr12:129868969-129869727
miR200C	hg18 chr12:6942800-6943200
MST1R	hg18 chr3:49916155-49916617
MACC1	hg18 chr7:20223293-20224058
HOXC4/HOXC5	hg18 chr12:52712961-52713967
CP2L3	hg19 chr8:102504509-102505208
RON	hg18 chr3:49916155-49916617
TBCD	hg18 chr17:78440426-78440951
C20orf55	hg18 chr20:770741-770860
ERBB2	hg19 chr17:37861100-37863650

[0199] 表 2 与上皮表型有关的甲基化胞嘧啶核苷酸

[0200]

基因	染色体	位置	基因	染色体	位置
ALDH3B2	11	67204971	COLEC10	8	120175608
	2	62409583		5	147237412
AMICA1	11	117590946		5	147237649
TMPRSS13	11	117294776		5	147237518
	1	20440804		1	167067800
	20	1420914	DLG2	11	84558496
	1	20374821	RAB19	7	139760606
DAPP1	4	101009384	PRR5-ARHGAP8, ARHGAP8	22	43564823
AMICA1	11	117590130		2	230797885
	19	47895208	CHMP4C	8	82834065
MYO1D	17	28170353		7	21037502
AFAP1	4	7945103		7	21031624
SPINK5	5	147423445		20	36533851
SPINK5	5	147423477		14	74735436
SPINK5	5	147423260	MYCBPAP	17	45964331
ANO3, MUC15	11	26538572	TMEM30B	14	60814048
	3	183702446		9	84869514
	3	183767276		10	100127021
SYT16	14	61608563		6	80178446
SYT16	14	61532507		3	106815813
TC2N	14	91391048	CNGA1	4	47710738
TC2N	14	91375355	SLAMF9	1	158190485
CEACAM6	19	46966764	CD180	5	66513564
KIAA0040	1	173395004	ESR1	6	152166508
KIAA0040	1	173396807		12	72730512
SYK	9	92692359	MRVII	11	10559098
SYK	9	92659288	CYP4B1	1	47037188
SEMA6D	15	45522047	MFSD4	1	203816771
ERP27	12	14982225	PLA2G2F	1	20338719
IVL	1	151148554	CYP4B1	1	47057214
IVL	1	151148439	CYP4A22	1	47375597
KRTAP3-3	17	36403692		1	47036300
KRTAP3-3	17	36403856	SDR16C5	8	57375347
	5	55990316		5	39796557
DHRS9	2	169653716	SAMD12	8	119525751
	4	55490421		1	190775347
SPAMI	7	123353161	TAT	16	70168544
	8	127777938	SALL3	18	74858829
	8	120206280		11	128964767
COLEC10	8	120187865	PKHDI	6	51787037

[0201] 表 2 (续上) 与上皮表型有关的甲基化胞嘧啶核苷酸

[0202]

基因	染色体	位置	基因	染色体	位置
	11	2178652	ZC4H2	X	64171340
IRF6	1	208029710	TRAM2	6	52549027
UBXN10	1	20391491	BVES	6	105690909
	7	7359082	BVES	6	105690842
SCEL	13	77066219	MLLT11	1	149299631
TMCI	9	74639567	MLLT11	1	149299586
	8	127457153	MLLT11	1	149299347
	4	55742586		2	42128114
PHLDB2	3	113112565		2	42128123
HMHB1	5	143180348		1	113301372
	7	19927552		12	95407994
	16	68155671	TENC1	12	51729752
LAMA2	6	129245818	TENC1	12	51729851
LAMA2	6	129245899		3	42088628
	11	65020493	SPRY4	5	141675957
SHANK2	11	70217854	SPRY4	5	141679764
SHANK2	11	70350904		19	13808229
NFIC	19	3312374		19	13808284
NFIC	19	3312154		19	13808262
FLNB	3	58020380		19	13808473
TEAD4	12	2978486		19	13808469
ABCC3	17	46113650	DGAT1	8	145510701
TMEM120B	12	120670919	NRM	6	30764049
SCNN1A	12	6347262	NRM	6	30764073
	8	103890347	NRM	6	30764003
SAMD11, NOC2L	1	869821	FLOT1	6	30817584
KIRREL	1	156231153	FLOT1	6	30817649
MYADM	19	59061828		19	52793249
INPP5B	1	38185106	LAMB3	1	207868102
INPP5B	1	38185405	LAMB3	1	207867974
INPP5B	1	38185271	AP1M2	19	10544470
INPP5B	1	38185275	MAP3K14	17	40747948
INPP5B	1	38185298	MAP3K14	17	40748115
INPP5B	1	38185331	ELOVL7	5	60094877
PDE4D	5	58457316	ADAPI	7	913183
	11	65013559		17	17470224
CADPS2	7	122024033	PTK2B	8	27325072
ITGA5	12	53098352		1	19211638
ZC4H2	X	64171392		17	54761551
ZC4H2	X	64171381	ITGB3	17	42685877

[0203] 表 2 (续上) 与上皮表型有关的甲基化胞嘧啶核苷酸

[0204]

基因	染色体	位置	基因	染色体位置
ITGB3	17	42686081	INPP5B	hg18 chr1:38184921-38185620
ITGB3	17	42685928	BVES	hg18 chr6:105690492-105691191
ITGB3	17	42685861	ITGA5	hg18 chr12:53098002-53098701
ITGB3	17	42686060	ITGB3	hg18 chr17:42685578-42686277
C11orf70	11	101423920	JAKMIP2	hg18 chr5:147142066-147142765
EPN3	17	45974036	MLLT11	hg18 chr1:149299281-149299980
	1	20672639	NFIC	hg18 chr19:3311804-3312503
LIX1L	1	144189643	NTNG2	hg18 chr9:134026339-134027038
SIGIRR	11	403594	ZEB2	hg18 chr2:144989568-144989952
	17	73861330	PCDH8	hg18 chr13:52321009-52321560
	11	32068730	PEX5L	hg18 chr3:181236933-181237780
KLF16	19	1810341	GALR1	hg18 chr18:73090412-73090797
	1	28457982		
	10	129592716		
LY6G6C	6	31795616		
CDS1	4	85777370		
MRVII	11	10562607		
	10	17309649		
	10	17309781		
	17	23722410		
	16	67990001		
ZEB2	2	144994583		
	4	40953131		
ANK3	10	62002852		
	5	10618263		
	5	66600290		
NTNG2	9	134026689		
JAKMIP2	5	147142568		
JAKMIP2	5	147142416		
JAKMIP2	5	147142654		
JAKMIP2	5	147142625		
	10	30178594		
TBC1D1	4	37655153		
TBC1D1	4	37655061		
TBC1D1	4	37655126		

[0205] 表 3 与间充质表型有关的甲基化胞嘧啶核苷酸

[0206]

基因	染色体	位置	EntrezID	CpG_岛
TSPAN14	10	82209390	81619	
LFNG	7	2530782	3955	CpG_35
PRKCH	14	61062573	5583	
SDC4	20	43406736	6385	
SCYL3	1	168127429	57147	
TNXB	6	32162072	7148	
ARHGAP39	8	145777560	80728	CpG_52
SPINT1	15	38937072	6692	
SLC9A7	X	46499386	84679	
	3	49919159		
TBCD	17	78440786	6904	
VTH1A	10	114492308	143187	
LDLRAP1	1	25767066	26119	
PLEKHG6	12	6291928	55200	
PNPLA8	7	107955918	50640	
PNPLA8	7	107955957	50640	
ARID1A	1	26953185	8289	
ABTB2	11	34241186	25841	
SLC9A3R1	17	70267242	9368	
	7	2447061		
GALNTL2	3	16220636	117248	
ZNF321	19	58139084	399669	
DIP2B	12	49261110	57609	
	3	178803691		
	2	242481901		
	7	6491550		
WDR82	3	52277292	80335	
TRAF5	1	209569842	7188	
PPARD	6	35417906	5467	CpG_65
LYN	8	57066177	4067	
LOC254559	15	87723993	254559	CpG_155
LOC254559	15	87723796	254559	CpG_155
	7	27744012		
TMEM79	1	154520773	84283	
	8	102520036		
JMJD7-PLA2G4B	15	39918027	8681	

[0207] 表 3 (续上) 与间充质表型有关的甲基化胞嘧啶核苷酸

[0208]

基因	染色体	位置	EntrezID	CpG_岛
FTO	16	52372627	79068	
	15	78857906		CpG_157
BAIAP2	17	76625735	10458	
	8	102520234		
	8	102520167		
NRBPI, KRTCAP3	2	27518521	29959, 200634	CpG_42
PVRL2	19	50073777	5819	
	7	6491523		
CSK	15	72868625	1445	
PITPNM3	17	6396092	83394	
GRHL2	8	102575811	79977	CpG_104
PVRL4	1	159325891	81607	
LAMA3	18	19706786	3909	
	8	144892697		
	8	144892671		
STX2	12	129869200	2054	CpG_56
STX2	12	129869147	2054	CpG_56
OBSCN	1	226625610	84033	CpG_30
GNA13	17	60466557	10672	
ACAP2	3	196640585	23527	
WDR82	3	52277190	80335	
NSMCE2	8	126223268	286053	
	10	73752495		
RAB24	5	176661226	53917	
ETV6	12	11922571	2120	
ENDOD1	11	94481032	23052	
	7	155407740		
LIMA1	12	48882614	51474	
TBCD	17	78426682	6904	
TBCD	17	78426927	6904	
TBCD	17	78426815	6904	
C10orf91	10	134111645	170393	
	2	64687934		
	2	64687784		
	2	64687610		
SPIRE1	18	12636025	56907	
STX2	12	129869047	2054	CpG_56
LRP5	11	67866681	4041	
OBSCN	1	226625713	84033	CpG_30

[0209] 表 3 (续上) 与间充质表型有关的甲基化胞嘧啶核苷酸

[0210]

基因	染色体	位置	EntrezID	CpG_岛
OBSCN	1	226625944	84033	CpG_30
OBSCN	1	226625706	84033	CpG_30
OBSCN	1	226625779	84033	CpG_30
CGN	1	149753111	57530	
	12	6943501		
RAB25	1	154297806	57111	
	12	6943503		
	12	6943508		
TBCD	17	78440951	6904	
MYST1, PRSS8	16	31050024	84148, 5652	
TBCD	17	78440835	6904	
TBCD	17	78440498	6904	
TBCD	17	78440559	6904	
TBCD	17	78440426	6904	
GRHL2	8	102574469	79977	CpG_104
GRHL2	8	102575727	79977	CpG_104
GPR110	6	47118050	266977	
	6	7477665		
THSD4	15	69416262	79875	
	3	53164930		
C20orf151	20	60435990	140893	
PWWP2B	10	134072208	170394	
	7	2447019		
	2	70221961		
LAMA3	18	19706842	3909	
LAMA3	18	19706817	3909	
RHOBTB3	5	95089583	22836	
GPR56	16	56211848	9289	
RAB25	1	154297468	57111	
RAB25	1	154297433	57111	
TMEM159, DNAH3	16	21078585	57146, 55567	
C1orf210	1	43524091	149466	
CCDC19	1	158136950	25790	
C1orf210	1	43524084	149466	
CLDN7	17	7105979	1366	CpG_159
GRHL2	8	102574035	79977	CpG_31
SPINT1	15	38925452	6692	

[0211] 表 3 (续上) 与间充质表型有关的甲基化胞嘧啶核苷酸

[0212]

基因	染色体	位置	EntrezID	CpG_岛
ADAPI	7	952140	11033	
	12	6943440		
	12	6943525		
RAP1GAP2	17	2815637	23108	
VPS37C	11	60682632	55048	
IGFIR	15	97074397	3480	
BOLA2, GDPD3	16	30023636	552900, 79153	
	22	28307742		
	22	28308158		
NA, RCC1	1	28726284	751867, 1104	
CTNND1	11	57305264	1500	
	2	101234788		
MPRIP	17	16907618	23164	
FRMD6	14	51101498	122786	
	16	86381426		
ARHGAP39	8	145777354	80728	CpG_52
MAPK13	6	36207101	5603	
	10	5583926		
	10	5583949		
	13	109313445		
F11R	1	159258982	50848	
SDCBP2	20	1257722	27111	
F11R	1	159258976	50848	
EHF	11	34599461	26298	
ABLIM1	10	116269176	3983	
MCCC2	5	70933152	64087	
COX10	17	14050396	1352	
SLC37A1	21	42809566	54020	
MYO18A	17	24529971	399687	
IL17RE, CIDEA	3	9919537	132014, 63924	
S100A14	1	151855406	57402	
IL17RE, CIDEA	3	9919512	132014, 63924	
TALDO1	11	753485	6888	
PHGDH	1	120075342	26227	
SIPAIL1	14	71183551	26037	
	2	189266325		

[0213] 表 3 (续上) 与间充质表型有关的甲基化胞嘧啶核苷酸

[0214]

基因	染色体	位置	EntrezID	CpG_岛
TMEM159, DNAH3	16	21078740	57146, 55567	
PPCDC	15	73115889	60490	
GPR56	16	56211418	9289	
LLGL2	17	71057739	3993	
SPINT1	15	38923139	6692	CpG_135
CLDN15	7	100662856	24146	CpG_54
CNKSR1	1	26376445	10256	
GRB7	17	35149701	2886	
NRBPI, KRTCAP3	2	27519047	29959, 200634	CpG_42
KRTCAP3	2	27519215	200634	CpG_42
	16	83945057		
GPR56	16	56211405	9289	
TACC2	10	123744125	10579	
ADAT3, SCAMP4	19	1858677	113179, 113178	CpG_34
CHD2	15	91266091	1106	
GRHL2	8	102575565	79977	CpG_104
	7	139750195		
	8	102573120		
	1	227296135		
PDGFRA, LNX1	4	54152503	5156, 84708	
PDGFRA, LNX1	4	54152494	5156, 84708	
	11	3454830		
ITGB6	2	160764885	3694	
PDGFRA, LNX1	4	54152866	5156, 84708	
	20	36661934		
	1	1088243		CpG_183
ST14	11	129535669	6768	CpG_64
	7	139750206		
C20orf151	20	60436134	140893	
	7	139750140		
LOC643008, RECQL5	17	71147779	643008, 9400	
GRB7	17	35147553	2886	
GRB7	17	35147540	2886	
C1orf210	1	43523857	149466	

[0215] 表 3 (续上) 与间充质表型有关的甲基化胞嘧啶核苷酸

[0216]

基因	染色体	位置	EntrezID	CpG_岛
CNKSR1	1	26376606	10256	
CNKSR1	1	26376566	10256	
CLDN7	17	7106571	1366	CpG_159
CLDN7	17	7106564	1366	CpG_159
CLDN7	17	7106566	1366	CpG_159
C1orf210	1	43523950	149466	
C1orf210	1	43523957	149466	
CLDN4	7	72883688	1364	CpG_46
CLDN7	17	7105734	1366	CpG_159
C1orf210	1	43523963	149466	
CLDN7	17	7106573	1366	CpG_159
KRTCAP3	2	27519142	200634	CpG_42
MST1R	3	49914707	4486	CpG_23
MST1R	3	49915923	4486	CpG_53
XAB2	19	7590468	56949	
KIAA0182	16	84236385	23199	
PWWP2B	10	134072043	170394	
CCDC57	17	77655395	284001	
NRBP1, KRTCAP3	2	27518810	29959, 200634	CpG_42
NRBP1, KRTCAP3	2	27518583	29959, 200634	CpG_42
NRBP1, KRTCAP3	2	27518645	29959, 200634	CpG_42
MOCOS	18	32022494	55034	CpG_141
PWWP2B	10	134071493	170394	
LAMA3	18	19706893	3909	
	12	6943152		
	12	6942988		
	12	6943026		
	12	6942957		
	14	64239711		
PRSS8	16	31054518	5652	
	17	75403479		CpG_427
C20orf151	20	60436252	140893	
GRHL2	8	102574732	79977	CpG_104
C20orf151	20	60436106	140893	

[0217] 表 3 (续上) 与间充质表型有关的甲基化胞嘧啶核苷酸

[0218]

基因	染色体	位置	EntrezID	CpG_岛
SULT2B1	19	53747255	6820	
SULT2B1	19	53747244	6820	
SULT2B1	19	53747224	6820	
SULT2B1	19	53747250	6820	
CBLC	19	49973124	23624	
NRBP1, KRTCAP3	2	27519011	29959, 200634	CpG_42
NRBP1, KRTCAP3	2	27518654	29959, 200634	CpG_42
GRHL2	8	102573658	79977	CpG_31
DOK7	4	3457234	285489	
FAM110A	20	770788	83541	CpG_71
NRBP1, KRTCAP3	2	27518643	29959, 200634	CpG_42
PWWP2B	10	134071623	170394	
TALDO1	11	753339	6888	
OVOL1	11	65310618	5017	CpG_204
SH3YL1	2	253656	26751	CpG_176
	7	139750225		
LAD1	1	199635571	3898	CpG_54
TMEM159, DNAH3	16	21078568	57146, 55567	
GRHL2	8	102573922	79977	CpG_31
PDGFRA, LNX1	4	54152402	5156, 84708	
LAD1	1	199635569	3898	CpG_54
LAD1	1	199635537	3898	CpG_54
KRT8	12	51586560	3856	
	3	135552584		
	19	49971605		
ITGB6	2	160822102	3694	
ADAP1	7	952310	11033	
ADAP1	7	952245	11033	
PROM2	2	95304202	150696	
PROM2	2	95304432	150696	
PROM2	2	95303758	150696	
SYT8	11	1811862	90019	
	16	70401148		
	17	15737821		
QSOX1	1	178404541	5768	
CCDC85C	14	99114910	317762	

[0219] 表 3 (续上) 与间充质表型有关的甲基化胞嘧啶核苷酸

[0220]

基因	染色体	位置	EntrezID	CpG_岛
C1orf116	1	205273070	79098	
GRHL2	8	102576558	79977	
C19orf46	19	41191166	163183	
CBLC	19	49973366	23624	
CAMK2G	10	75302072	818	
SCNN1A	12	6354990	6337	
SCNN1A	12	6354868	6337	
JUP	17	37182909	3728	
	19	60699327		
VCL	10	75485630	7414	
BOLA2, TBX6	16	30009181	552900, 6911	
IMMP2L	7	110988180	83943	
SLC44A2	19	10596548	57153	CpG_46
	8	144726627		
RAI1	17	17572988	10743	
SYT1	12	78333487	6857	
	8	28514725		
	6	134742250		
GPR56	16	56211203	9289	
EPN3	17	45967146	55040	
GPR56	16	56211170	9289	
C4orf36	4	88031692	132989	
ARL13B, STX19	3	95230218	200894, 415117	
	2	70222288		CpG_118
PVRL4	1	159326053	81607	
	1	27066922		
GPR110	6	47117696	266977	
EPHB2	1	23025895	2048	
ANKRD22	10	90601891	118932	
ZNF398	7	148472457	57541	
PWWP2B	10	134071845	170394	
ARHGAP32	11	128399061	9743	
	7	80389667		
	4	154136934		
	1	27023897		
	19	1855554		
BAIAP2	17	76626137	10458	
PLXNB2	22	49062595	23654	

[0221] 表3(续上)与间充质表型有关的甲基化胞嘧啶核苷酸

[0222]

基因	染色体	位置	EntrezID	CpG_岛
ACAA1	3	38150460	30	
DNAJC17	15	38867650	55192	
	7	72795287		
COL18A1, SLC19A1 LOC643008, RECQL5	21	45757802	80781, 6573	
MANF	17	71147845	643008, 9400	
TRAK1	3	51401417	7873	
TRAK1	3	42147101	22906	
GRB7	17	35147329	2886	
C1orf210	1	43524150	149466	
RNF144A	2	7089548	9781	
GRB7	17	35147290	2886	
	19	58230499		
	1	234153824		
PPFIBP2	11	7578132	8495	
GPR81	12	121777086	27198	
	19	58230695		
	8	101497819		
CPEB3	10	93872825	22849	
RABGAP1L	1	173111113	9910	
RABGAP1L	1	173111020	9910	
RNF207	1	6202430	388591	
MUC1	1	153429495	4582	
	1	2456135		
PLEKHG6	12	6292029	55200	
PLEKHG6	12	6292067	55200	
PNPLA8	7	107955947	50640	
RASA3	13	113862226	22821	
ARL13B, STX19	3	95230100	200894, 415117	
VTIL1A	10	114516704	143187	
COL21A1	6	56342813	81578	
	2	74064468		CpG_113
SDCBP2	20	1258000	27111	
FAM167A	8	11340393	83648	
S100A14	1	151855551	57402	
PRSS8	16	31054183	5652	
HIVEP3	1	41753450	59269	

[0223] 表 3 (续上) 与间充质表型有关的甲基化胞嘧啶核苷酸

[0224]

基因	染色体	位置	EntrezID	CpG_岛
PRSS8	16	31054700	5652	
SULT2B1	19	53747202	6820	
C19orf46, ALKBH6	19	41191679	163183, 84964	CpG_49
C19orf46, ALKBH6	19	41191506	163183, 84964	
C19orf46, ALKBH6	19	41191561	163183, 84964	CpG_49
	17	52465903		
RAP1GAP2	17	2855119	23108	
C10orf91	10	134110971	170393	
	8	144892814		
	9	131184926		CpG_71
BMF	15	38186423	90427	
RGS3	9	115383006	5998	
	19	17763242		
	19	50356091		
DLEU1	13	49829837	10301	
MBP	18	72930014	4155	
	1	150076158		
JMJD7-PLA2G4B	15	39917942	8681	
PARD3	10	34756309	56288	
MICAL2	11	12226862	9645	
ANKFY1	17	4098222	51479	
CDKN1A	6	36758711	1026	
	19	49971610		
JARID2	6	15564181	3720	
SGSM2	17	2197812	9905	
SMARCD2	17	59270462	6603	
PNKD	2	218868246	25953	
EVPLL	17	18221746	645027	
EVPLL	17	18221574	645027	
MED16	19	834879	10025	
RAB24	5	176661618	53917	
	7	155407629		
ERBB2	17	35115639	2064	
CGN	1	149751930	57530	
	8	8356184		

[0225] 表 3 (续上) 与间充质表型有关的甲基化胞嘧啶核苷酸

[0226]

基因	染色体	位置	EntrezID	CpG_岛
GNAI3	1	109914827	2773	
	8	37880723		
ANKRD22	10	90601835	118932	
	15	81670543		
PAK4	19	44350154	10298	
PRR15L	17	43390182	79170	
RAB17	2	238164820	64284	
P2RY2	11	72616798	5029	
	22	28307949		
	8	144893700		CpG_78
SPINT1	15	38923085	6692	CpG_135
PVRL4	1	159326159	81607	
	6	13981646		CpG_39
Clorf210	1	43524056	149466	
	7	139750233		
TBC1D1	4	37666838	23216	
	7	72795153		
	2	238165064		
ARHGAP32	11	128399150	9743	
	12	88144203		
TMC8	17	73650109	147138	
ABCF1	6	30667066	23	
ST3GAL4	11	125781216	6484	
ST3GAL4	11	125781207	6484	
STAP2	19	4289769	55620	
STAP2	19	4289932	55620	
LAMA3	18	19706827	3909	
GSDMC	1	201096568	56169	CpG_80
	8	130868275		
AFF1	4	88113322	4299	
	17	71380179		
	14	34872148		
ASB13	10	5742089	79754	
CLDN7	17	7106144	1366	CpG_159
CDC42BPG	11	64367663	55561	
FAM46B	1	27207475	115572	
EPS8L1	19	60278851	54869	
	16	70401060		CpG_91

[0227] 表3(续上)与间充质表型有关的甲基化胞嘧啶核苷酸

[0228]

基因	染色体	位置	EntrezID	CpG_岛
ESRP2	16	66825963	80004	
IL10RB	21	33563377	3588	
C14orf43	14	73281541	91748	
CCDC120	X	48803602	90060	
CCDC120	X	48803499	90060	
ESRP2	16	66825753	80004	
CNKSR1	1	26377135	10256	
CLDN7	17	7107442	1366	CpG_159
SCNN1A	12	6354974	6337	
MUC1	1	153429380	4582	
PRSS8	16	31054500	5652	
SLC35B3	6	8381262	51000	CpG_68
	12	13179838		
EPS8L1	19	60279005	54869	
GPR110	6	47118136	266977	
LAMA3	18	19706728	3909	
PVRL4	1	159325951	81607	
PVRL4	1	159326082	81607	
RIPK4	21	42058454	54101	
NEURL1B	5	172048817	54492	
PROM2	2	95303838	150696	
FAM167A	8	11340449	83648	
CLDN4	7	72882009	1364	
	8	102573068		
CANTI	17	74513111	124583	
PRR15L	17	43390296	79170	
MICALL2	7	1461837	79778	
NCOA2	8	71402682	10499	
ITGB6	2	160764766	3694	
ITGB6	2	160764846	3694	
	14	64792338		
	8	102573146		
NRBP1, KRTCAP3 TMEM159, DNAH3	2	27518632	29959, 200634	CpG_42
DNAH3	16	21078598	57146, 55567	
ADAP1	7	952156	11033	

[0229] 表 3 (续上) 与间充质表型有关的甲基化胞嘧啶核苷酸

[0230]

基因	染色体	位置	EntrezID	CpG_岛
TMEM159, DNAH3	16	21078428	57146, 55567	
SH3YL1	2	253559	26751	CpG_176
	7	139750252		
PRSS22	16	2848212	64063	
PRSS22	16	2848220	64063	
SDCBP2	20	1257800	27111	
LAMA3	18	19707129	3909	
	2	74064398		CpG_113
	2	74064365		CpG_113
DAPP1	4	100956844	27071	
DAPP1	4	100956853	27071	
DAPP1	4	100957034	27071	
	1	999308		
ATG9B	7	150352451	285973	
CLDN7	17	7107017	1366	CpG_159
	9	131185398		CpG_71
STX2	12	129868969	2054	CpG_56
CNKSR1	1	26376578	10256	
E2F4, ELMO3	16	65790778	1874, 79767	
E2F4, ELMO3	16	65790422	1874, 79767	
CNKSR1	1	26376365	10256	
CNKSR1	1	26376363	10256	
ARAP1, STARD10	11	72169819	116985, 10809	CpG_41
CNKSR1	1	26376520	10256	
CNKSR1	1	26376434	10256	
CNKSR1	1	26376449	10256	
MUC1	1	153429376	4582	
PRSS8	16	31054545	5652	
PRSS8	16	31054555	5652	
	7	72795319		
PDGFRA, LNX1	4	54152685	5156, 84708	
C20orf151	20	60436261	140893	
LAD1	1	199635654	3898	CpG_54
PDGFRA, LNX1	4	54152599	5156, 84708	
	12	50912694		CpG_79

[0231] 表 3 (续上) 与间充质表型有关的甲基化胞嘧啶核苷酸

[0232]

基因	染色体	位置	EntrezID	CpG_岛
GRHL2	8	102573655	79977	CpG_31
GRHL2	8	102573677	79977	CpG_31
GRHL2	8	102574689	79977	CpG_104
GRHL2	8	102573797	79977	CpG_31
GRHL2	8	102573623	79977	CpG_31
RNF144A	2	7089414	9781	
NCRNA00093, DNMBP	10	101680658	100188954, 23268	
PRKCA	17	62088295	5578	
KIAA0247	14	69194460	9766	
ELF3	1	200246387	1999	
ELF3	1	200246469	1999	
ELF3	1	200246561	1999	
GAS8	16	88638299	2622	
HSH2D	19	16115489	84941	
C10orf91	10	134111403	170393	
	12	88143460		
SYT8	11	1812078	90019	
SYT8	11	1812322	90019	
	10	126879805		
	4	8587017		
ERGIC1	5	172264397	57222	
	12	50911753		
SYT8	11	1812236	90019	
	8	144727414		
	16	11613936		
CLDN7	17	7106555	1366	CpG_159
	5	74369044		
BAIAP2	17	76625947	10458	
BAIAP2	17	76625872	10458	
OPA3	19	50723356	80207	
GRHL2	8	102573740	79977	CpG_31
GRHL2	8	102573842	79977	CpG_31
	8	102085088		
CLDN7	17	7106633	1366	CpG_159
CLDN7	17	7107214	1366	CpG_159
ERBB3	12	54761038	2065	CpG_116

[0233] 表 3 (续上) 与间充质表型有关的甲基化胞嘧啶核苷酸

[0234]

基因	染色体	位置	EntrezID	CpG_岛
CLDN7	17	7105010	1366	CpG_159
	16	2999774		
	15	72610634		
	11	66579429		
ANKRD22	10	90601762	118932	
	14	64239802		
	14	64239962		
SLC44A2	8	144893629	57153	CpG_78
	19	10596578		CpG_46

[0235] 表 4 与上皮表型有关的甲基化胞嘧啶核苷酸

[0236]

基因	染色体	位置	EntrezID	CpG_岛
HBQ1	16	170343	3049	CpG_150
HBQ1	16	170341	3049	CpG_150
IGF2BP1	10	118912877	10642	CpG_110
	17	44427906		CpG_255
	17	44430879		CpG_255
TC2N	4	25120404	123036	
	14	91391048		
ALDH3B2	11	67204971	222	
MYO1D	17	28170353	4642	
SYK	9	92692359	6850	
SYK	9	92659288	6850	
AMICA1	11	117590130	120425	
MAL2	8	120326244	114569	
MACROD2	20	14267035	140733	
OVOL2	20	17972215	58495	
CAPN13	2	30821257	92291	
PLG	6	161094476	5340	

[0237] 表 4 (续上) 与上皮表型有关的甲基化胞嘧啶核苷酸

[0238]

基因	染色体	位置	EntrezID	CpG_岛
NCALD	8	102871791	83988	
	6	147353266		
	14	100245858		CpG_79
	14	100245905		CpG_79
	14	100246063		CpG_79
TRIM9	14	50630238	114088	CpG_199
KIAA0040	1	173395004	9674	
KIAA0040	1	173396807	9674	
	7	50601267		
	8	127777938		
	7	50601390		
	7	50601219		
SYDE1	19	15079713	85360	CpG_56
	11	65013559		
NUAK1	12	104998452	9891	
MMP2	16	54071026	4313	CpG_42
ZNF521	18	21184882	25925	
ZNF521	18	21185001	25925	
IRF6	1	208029710	3664	
SRD5A2	2	31656355	6716	
MMP2	16	54070981	4313	CpG_42
IGF2BP1	17	44430854	10642	CpG_255
ZC4H2	X	64171686	55906	CpG_71
IGF2BP1	12	72730512	10642	CpG_255
	17	44430757		
PAX7	1	18830954	5081	CpG_205
	17	44427856		
MLLT11	1	149299347	10962	CpG_53
MLLT11	1	149299586	10962	
	6	114284192		
	6	114284228		
	6	114284034		
	6	114284022		
	10	118912831		CpG_110
	10	118912483		CpG_110
	10	118912726		CpG_110
	X	64171940		CpG_71

[0239] 本发明将从以下实施例得到更好的理解。然而，本领域技术人员将容易地领会，论述的特定方法和结果仅为本发明例示性的，如在随后权利要求中更完全描述的，且不视为以任何方式对其限定的。

## 实施例

[0240] 实施例 1- 材料和方法

[0241] Fluidigm 表达分析: 使用 BioMark96×96 基因表达平台 (Fluidigm) 和 20 基因 EMT 表达集 (补充表 S1 和方法), 在 82 种 NSCLC 细胞系上进行 EMT 基因表达分析。将  $\Delta Ct$  值用于使用 Cluster v. 3.0 和 Treeview v. 1.60 软件依照 EMT 基因表达水平将细胞系聚类。

[0242] Illumina Infinium 分析: 如下文描述使用 Illumina Human Methylation450Bead Chip (Illumina, San Diego, CA), 在 Expression Analysis, Inc. (Durham, N. C.) 收集微阵列数据。分析阵列数据, 并使用“留一 (leave-one-out)”交叉验证策略建立甲基化分类器 (classifier) (在下文和参考文献 25、26 中描述)。阵列数据已提交到 Gene Expression Omnibus 数据库 (登录号 GSE36216)。

[0243] 细胞系: 所有 NSCLC 细胞系购自美国典型培养物保藏中心 (ATCC) 或由 UT Southwestern 的 Adi Gazdar 和 John Minna 提供。永生化的支气管上皮 (gBEC) 和小气道 (gSAC) 细胞系在 Genentech 使用含有 cdk4、hTERT 和 G418 作为选择标志物的三顺反子载体创建。该三顺反子载体从含有 hTERT 的 pQCXIN 主链工程化而来。永生化的过程基于先前公开的方案, 有一些修改 (Ramirez, Sheridan 等 2004; Sato, Vaughan 等 2006)。gBEC 和 gSAC 具有二倍体核型而且是非致瘤性的。如描述的实施工 5- 氮杂 dC、厄洛替尼或 TGF  $\beta$  1 对细胞系的处理。

[0244] NSCLC 正常肺组织, 原发性肿瘤和活组织检查组织: 代表早期、手术可切除肿瘤的 31 份 NSCLC 新鲜冷冻的原发性肿瘤组织 (N = 28 份腺癌, 3 份鳞状细胞癌) 和来自继续在前线化疗失败的患者 60 份福尔马林固定的、石蜡包埋 (FFPE) NSCLC 活组织检查。35 份新鲜冷冻的正常肺组织 (匹配原发性肿瘤组织的 31 份也是此收集的一部分)。所有样品均在有知情同意书的情况下以 IRB 批准的方案获得。所有样品均由病理学家评估组织质量和肿瘤阶段、级别和肿瘤内容物。在 Genentech 临床处从健康志愿者获得外周血单核细胞 (N = 20)。

[0245] 5- 氮杂 dC 治疗和 TGF  $\beta$  1 治疗: 细胞在补充有 10% 胎牛血清和 2mM L- 谷氨酰胺的 RPMI1640 中生长。在第 0 天, 将细胞以 4000-9000 个细胞 /cm<sup>2</sup> 接种, 并在第 1、3 和 5 天用 1  $\mu$ M 5- 氮杂 -2'- 脱氧胞苷 (5- 氮杂 -dC) (SIGMA-ALDRICH 货号 A3656) 或 DMSO 对照 (货号 D2650) 给药。在第 6 天, 将细胞在冷的 PBS 中清洗 1 次, 通过刮擦到 Trizol (Invitrogen, 货号 15596018) 中收获, 并提取 RNA 或速冻用于其后的 RNA 提取。对于 EMT 的诱导, 将细胞以 20000-50000 个细胞 /10cm<sup>2</sup> 铺板于完全培养基并补充有 2ng/mL 的人转化生长因子 beta1 (TGF  $\beta$  1) (R&D Systems, 货号 100-B/CF) 或 PBS 对照。每 3 天替换培养基和 TGF  $\beta$  1, 并在 EMT 的 TGF  $\beta$  1 诱导后 4-5 周提取 RNA。使用 Taqman 测定法对 20 基因 EMT 集评估基因表达变化 (图 1)。

[0246] 厄洛替尼处理: 对于厄洛替尼 IC<sub>50</sub> 测定, 将细胞一式四份地以 3x10<sup>2</sup> 个细胞每孔铺板于 384 孔板中的含 0.5% FBS 的 RPMI 中 (测定培养基), 并过夜温育。24 小时后, 将细胞用含 3nM TGF  $\alpha$  和剂量范围为 10  $\mu$ M-1pM 终浓度的厄洛替尼的测定培养基处理。72 小时后, 使用 Celltiter-Glo Luminescent Cell Viability Assay (Promega) 测量细胞存活力。从 4 参数曲线分析计算导致细胞存活力 50% 抑制的厄洛替尼浓度并从最少 2 个实验测定。展现出厄洛替尼 IC<sub>50</sub>  $\leq$  2.0  $\mu$ M 的细胞系定义为敏感性的, 2.0-8.0  $\mu$ M 的为居中的, 而

$\geq 8.0 \mu\text{M}$  为抗性的。

[0247] Fluidigm 基因表达分析: 使用 Superscript III/Platinum Taq (Invitrogen) 和预扩增反应混合物 (Invitrogen), 在单次反应中将  $2 \mu\text{l}$  的总 RNA 逆转录成 cDNA 并预扩增。将选择用于 EMT 表达集 (图 1) 的 20 种 Taqman 引物 / 探针集纳入以 0.05x 初始 Taqman 测定浓度 (Applied Biosystems) 最终稀释的预扩增反应中。热循环条件如下:  $50^\circ\text{C}$  15 分钟 1 个循环,  $70^\circ\text{C}$  2 分钟 1 个循环, 然后  $95^\circ\text{C}$  15 秒和  $60^\circ\text{C}$  4 分钟的 14 个循环。

[0248] 将预扩增的 cDNA 稀释 1.94 倍, 然后使用 Taqman Universal PCR Master Mix (Applied Biosystems) 在 BioMark BMK-M-96.96 平台 (Fluidigm) 上依照制造商用法说明书扩增。所有样品均一式三份地测定。将先前评估其在多种细胞系 (新鲜冷冻的组织样品和 FFPE 组织样品, AL-1377271 和 VPS-33B) 中的表达稳定性的 2 种定制设计的参照基因纳入该表达集。对于每种样品计算两种参照基因的 Ct 值的均值, 并使用 delta Ct (dCt) 方法如下测定 EMT 靶基因的表达水平: 均值 Ct (靶基因) - 均值 Ct (参照基因)。

[0249] Illumina Infinium 分析: 使用 Illumina Human Methylation 450 BeadChip (Illumina), 在 Expression Analysis, Inc. (Durham, NC; www.expressionanalysis.com) 收集微阵列数据。这些阵列含有用于约 450,000 个 CpG 基因座位点的探针。靶物依照 “Illumina Infinium HD Methylation Assay, Manual Protocol” (Illumina Part#15019522 Rev. A) 制备并杂交。

[0250] 亚硫酸氢盐转化: 使用 500ng 的基因组 DNA 依照针对 Zymo EZ DNA 甲基化试剂盒 (Zymo Research) 的制造商方案采用亚硫酸盐转化反应。将 DNA 添加到 Zymo M 稀释缓冲液并在  $37^\circ\text{C}$  温育 15 分钟。然后添加 CT 转化试剂, 并通过加热至  $95^\circ\text{C}$  达 30 秒, 接着在  $50^\circ\text{C}$  温育 1 小时将混合物变性。将此变性 / 温育循环重复总共 16 小时。在亚硫酸氢盐转化后, DNA 结合 Zymo 旋转柱并使用脱磺基试剂按照制造商方案在柱上脱磺基。将经亚硫酸氢盐转化的 DNA 在  $10 \mu\text{l}$  洗脱缓冲液中从柱洗脱。

[0251] Infinium 甲基化测定法: 将  $4 \mu\text{l}$  经亚硫酸氢盐转化的产物转移至具有等量 0.1N NaOH 和 20ul MA1 试剂 (Illumina) 的新板, 然后允许在 RT 温育 10 分钟。紧接温育后, 添加 68ul 的 MA2 试剂和 75ul 的 MSM 试剂 (均来自 Illumina), 并将板在  $37^\circ\text{C}$  过夜温育以扩增。在扩增后, 将 DNA 酶法片段化, 沉淀并重悬于 RA1 杂交缓冲液。杂交和扫描: 将片段化的 DNA 分配到多通道 Human Methylation BeadChip, 并在 Illumina 杂交炉中实施杂交达 20 小时。依照制造商方案清洗 BeadChip, 延伸引物, 并染色。将 BeadChip 包被, 然后在 Illumina iScan Reader 上成像, 将图像用 GenomeStudio 软件甲基化模块处理 (版本 1.8 或以上)。

[0252] Infinium 分析: 甲基化数据使用 Bioconductor lumi 软件程序包处理 (Du, Kibbe 等 2008)。Infinium 450K 平台包括在同一张阵列上的 Infinium I 和 II 测定。Infinium I 测定采用两种珠类型每 CpG 基因座, 在一些情况中甲基化状态由红色染料报告, 在其他情况中由绿色染料 (与先前的 Infinium 27K 平台相同)。Infinium II 测定使用一种珠类型且总是用同一种染料报告甲基化状态, 从而涉及到染料偏差。将两阶段标准化规程应用于阵列: 首先, 对于每张阵列, 从 Infinium I 数据使用平滑的分位数标准化方法估测颜色 - 偏差校正曲线; 然后, 将该校正曲线应用于来自该阵列的所有数据。其次, 通过对所有颜色校正过的信号应用标准的分位数标准化来将阵列彼此标准化。在预处理后, 对于每份样品计算甲基化 M 值 (甲基化对未甲基化探针的  $\log_2$  比) 和值 (M 值经由对数转化尺度改变为 0 至 1 范围) (Du, Zhang

等 2010)。对于可视化,使用完全连接和欧几里得距离 (complete linkage and Euclidean distance) 来进行值的聚集性层次聚类 (agglomerative hierarchical clustering)。

[0253] 甲基化分类器:将 10x10 倍交叉验证策略用于选择一组差异甲基化的 CpG 位点 (DMR) 并同时评估基于甲基化的 EL 对 ML 分类器的准确性。将细胞系分成 10 个均衡大小的组。使用 9/10 的系 (训练集),通过以下来鉴定候选的 DMR:首先计算每种细胞系 M 值的移动平均 (以查询的 CpG 位点为中心的 500bp 窗口);然后,利用 t 检验来对比上皮样对间充质样训练系相关的窗口得分。调整 DMR p 值来控制错误发现率 (Benjamini 和 Hochberg, 1995) 并与 0.01 的截留比较。为了富集更生物学相关的现象,需要候选物具有以下平均窗口得分 (i) 在上皮和间充质系之间至少相差 1 个单位,且 (ii) 在两种细胞系的集中具有相反的标志 (sign)。该过程得到间充质有关 (阳性信号) 和上皮有关 (阴性信号) 的候选 DMR。为了评估性能,将扣留用于测试的 1/10 的系通过汇总其对阳性 DMR 的信号并减去对阴性 DMR 的信号,然后除以 DMR 总数来打分。将测试系的已知的上皮对间充质标记物与结果的标志比较。最后,将每 1/10 作为测试集,重复交叉验证过程。最终,交叉验证过程本身再重复 9 次,且总体准确性评估是 100 种不同测试集准确率率的平均。为了构建最终的 DMR 集,我们仅保留了在 100% 的交叉验证分组中鉴定为相关的候选物。将符合此标准的连续 DMR 合并成单一 DMR,如果它们相隔少于 2kb 的话。

[0254] 基于表达的 EMT 得分:观察到我们的 20 基因 Fluidigm 表达集中一些基因的行为在细胞系和肿瘤样品之间不同。为了鉴定该集的更有力的子集用于 EL 对 ML 分类目的,我们采用 CDH1 表达作为 EMT 锚定物,然后选出其与 CDH1 的相关性在细胞系和肿瘤样品二者中显示相同标志的基因 (总共 13 种)。为了向肿瘤样品分配 EMT 表达得分,首先中心化 13 种基因中每一种的 -dCT 值以具有均值 0,并调整比例以具有标准偏差 1。接着翻转那些显示与 CDH1 负相关的基因的标志。最后,通过平均经标准化和标志调整的结果来计算各肿瘤样品得分。

[0255] 亚硫酸氢盐测序和分析:将基因组 DNA 使用 EZ DNA 甲基化金试剂盒 (Zymo Research) 亚硫酸氢盐转化。特异于经转化 DNA 的引物使用 Methyl Primer Express 软件 v1.0 (Applied Biosystems) 设计 (序列在需要时可获得)。PCR 扩增使用 1  $\mu$ l 经亚硫酸氢盐转化的 DNA 在 25- $\mu$ l 反应中使用 Platinum PCR supermix (Invitrogen) 实施。PCR 热循环条件如下:95 $^{\circ}$ C 10 分钟的 1 个初始变性循环,接着是 94 $^{\circ}$ C 30 秒、65 $^{\circ}$ C 1 分钟并每个循环降低 1 $^{\circ}$ C 和 72 $^{\circ}$ C 1 分钟的 10 个循环,然后是 94 $^{\circ}$ C 30 秒、55 $^{\circ}$ C 1.5 分钟和 72 $^{\circ}$ C 1 分钟的 30 个循环,接着是 72 $^{\circ}$ C 15 分钟的最终延伸。PCR 产物通过使用含溴化乙锭 (Invitrogen) 的 2% 琼脂糖 E- 凝胶的电泳来解析并使用 FluorChem8900 照相机 (Alpha Innotech) 可见。

[0256] 将 PCR 产物使用 TOPO TA 克隆试剂盒 (Invitrogen) 依照制造商用法说明书连接到 pCR4-TOPO 载体中。将 2  $\mu$ l 连接的质粒 DNA 转化到 TOP10 感受态细菌 (Invitrogen) 中,并将 100  $\mu$ l 经转化的细菌铺板于含 50  $\mu$ g/ml 羧苄西林 (carbenicillin) (Teknova) 的 LB- 琼脂板上,在 37 $^{\circ}$ C 过夜温育。将针对每个候选基因座的 12 个菌落每细胞系接种到 1ml 含 50  $\mu$ g/ml 羧苄西林的 LB 中,并在摇动培养箱中于 37 $^{\circ}$ C 过夜生长。质粒 DNA 使用 Qiaprep miniprep 试剂盒以 96 孔形式 (Qiagen) 分离并在 3730x1DNA Analyzer (Applied Biosystems) 上测序。

[0257] 亚硫酸氢盐测序分析:测序数据使用 Sequencher v4.5 软件和 BiQ Analyzer 软件分析 (Bock, Reither 等 2005)。

[0258] **焦磷酸测序**:亚硫酸氢盐特异性 PCR(BSP) 引物使用 Methyl Primer Express 软件 v1.0(Applied Biosystems) 或 PyroMark Assay Design 软件 v2.0(Qiagen) 设计。PCR 引物合成为在前向或反向引物上具有 5' 生物素标记物以促进 PCR 产物对链霉亲和素琼脂糖珠的结合。测序引物使用 PyroMark Assay Design 软件 v2.0(Qiagen) 在 5' - 生物素标记的 PCR 引物的相反方向设计。引物序列在需要时可获得。将 1  $\mu$ l 亚硫酸氢盐修饰的 DNA 在 25  $\mu$ l 反应中使用 Platinum PCR Supermix(Invitrogen) 扩增并将 20  $\mu$ l PCR 产物用于在 PyromarkQ24(Qiagen) 上测序。将 PCR 产物与链霉亲和素琼脂糖珠温育 10 分钟,之后用 70%乙醇、Pyromark 变性溶液和 Pyromark 清洗缓冲液清洗。然后,将变性的 PCR 产物使用 0.3  $\mu$ M 测序引物来测序。使 pyrogram 可见并评估序列质量,使用 PyroMark 软件版本 2.0.4(Qiagen) 测定在各个 CpG 位点处的百分比甲基化。

[0259] **定量甲基化特异性 PCR**:设计靶向通过 Infinium 概况分析鉴定的 DMR 的定量甲基化特异性 PCR(qMSP) 测定法。将经亚硫酸氢钠转化的 DNA 用各种 20X Custom Taqman Assay 扩增,其使用 **TaqMan®** Universal PCR Master Mix、No **AmpErase®** UNG(Applied Biosystems),循环条件为 95°C 10 分钟,然后 50 个循环的 95°C 15 秒和 60°C 1 分钟。扩增在 7900HT 上进行并使用 SDS 软件(Applied Biosystems) 分析。DNA 含量使用 meRNaseP Taqman 测定法标准化。使用预扩增规程实施 FFPE 材料的 qMSP。

[0260] **FFPE 肿瘤材料的预扩增**:如下开发出一种用于对从福尔马林固定石蜡包埋的(FFPE) 组织提取的 pg 量 DNA 的甲基化分析的预扩增方法。将 2  $\mu$ l(等同于 100pg-1ng) 经亚硫酸氢盐转化的 DNA 首先在 20  $\mu$ l 反应中以 0.1X qMSP 引物-探针浓度扩增,其使用 **TaqMan®** Universal PCR Master Mix、No **AmpErase®** UNG(Applied Biosystems,货号 4324018),循环条件为 95°C 10 分钟,然后 14 个循环的 95°C 15 秒和 60°C 1 分钟。然后,将 1  $\mu$ l 预扩增的材料在第二 PCR 反应中扩增,循环条件为 95°C 10 分钟,然后 50 个循环的 95°C 15 秒和 60°C 1 分钟。DNA 含量使用参照 meRNaseP Taqman 测定法的预扩增来确认,且在 qMSP 反应的进一步分析中仅纳入对于 meRNaseP 阳性的样品。所有反应均一式两份地进行。

[0261] 实施例 2- 上皮样和间充质样表达签名与体外厄洛替尼敏感性相关

[0262] 先前定义了与 NSCLC 细胞系对厄洛替尼的体外敏感性相关的基因表达签名(11)。该基因集对于涉及 EMT 的基因是高度富集的。开发了在 Fluidigm 纳米流控平台上的基于定量逆转录酶 PCR 的 EMT 表达集(图 1)。来自 Yauch 等的研究(11)的 100 探针集与 Yauch 等的研究中概况分析的 42 种系的 20 基因 EMT Fluidigm 集的比较显示该 20 基因表达集是 EMT 的代表性分类器(参考文献 11)。

[0263] 为了进一步评估 20 基因集是否是 EMT 有关的表型变化代表性的,将 2 种细胞系用 TGF  $\beta$  1 处理。该研究的结果显示, TGF  $\beta$  1 诱导与 EMT 有关的形态学变化。在这些细胞系中,与上皮表型有关的基因下调,而与间充质表型有关的基因上调。

[0264] 为了确定 DNA 甲基化概况分析是否可用于将 NSCLC 细胞系分类到上皮样和间充质样组中,20 基因表达集被用于将上皮样对间充质样状态分配给 82 种细胞系。在此研究中使用的 NSCLC 细胞系包括在 Yauch 等的研究(11)中概况分析的大多数系和另外的 52 种系,其包括具有 EGFR 突变的 6 种系。在 82 种细胞系中,基于其中这些标志物的表达,36 种被

分类为上皮样,而 34 种被分类为间充质样(图 2)。将表达数据标准化并中值中心化(样品和基因)。绿色指示指定基因的低水平或无 mRNA 表达;红色指示高表达。12 种系(指示在图 2 的底部类中)被分类为上皮样,但表达上皮和间充质标志物的组合,指示这些系代表称为中间体的独特的生物学。如此,在 82 种 NSCLC 系中,89%可清楚地分类为上皮或间充质。对于大部分而言,该上皮样对间充质样表达表型是互相排斥的,可能反映了独特的根本性的生物学,这可能与独特的 DNA 甲基化概况有关联。包括组织学在内的细胞系描述的汇总显示于图 8A-B。

[0265] 实施例 3-NSCLC 细胞系中基因组范围内的甲基化概况与基于 Fluidigm 的 EMT 签名相关

[0266] 将 Illumina Infinium450K 阵列作为平台分析高通量甲基化概况分析,其通过比较 52 个探针的  $\beta$  值和细胞系子集( $N = 12$ )上的亚硫酸氢钠测序数据。观察到 Infinium 阵列的甲基化访问(call)与直接亚硫酸氢盐测序之间的高显著性的、强烈正相关性( $r = 0.926$ )。

[0267] 为了鉴定在上皮样和间充质样细胞系之间区分的 DMR,使用同时构建基于甲基化的分类器的交叉验证策略,并评估其预测准确性,如实施例 1 中描述的。当应用于 69 细胞系训练集时,该分析得到 549 种 DMR,其代表选择为限定上皮样对间充质样 NSCLC 细胞系的 915 个分别的 CpG 位点,在 100%的交叉验证迭代中经错误发现率调整的 P 值低于 0.01。基于甲基化的分类器的交叉验证估测的准确性为 88.0% ( $\pm 2.4\%$ , 95%置信区间)。

[0268] 接着,将包括在我们的基于甲基化的 EMT 分类器中的 CpG 位点用于对 69 种 NSCLC 细胞系(包括 6 个 EGFR 突变的厄洛替尼敏感系)和 2 种主要正常肺细胞菌株及其永生化对应体聚类。该分析揭示,上皮样、间充质样和正常系的惊人的分离(图 3)。在该测定法中,使用 Illumina Infinium450K 甲基化阵列平台将 72 种 NSCLC 细胞系和正常肺上皮细胞概况分析。使用 915 种探针进行监督层次聚类,所述探针在上皮样和间充质样细胞系之间显著差异甲基化(错误发现率 = 0.01;实施例 1)。列出了用于该聚类分析的注释过的探针集。每行代表在 Infinium450K 阵列上的单个探针,每列代表细胞系。在热图中呈蓝色的区域代表未甲基化区域,呈红色的区域代表甲基化区域。顶部色条显示代表每种细胞系的上皮样或间充质样状态(如通过 Fluidigm EMT 基因表达分析测定的)的柱。绿色指示上皮样,黑色指示间充质样细胞系。底部色条指示每种细胞系的厄洛替尼响应表型。红色指示厄洛替尼敏感性系;黑色指示厄洛替尼抗性系;灰色指示具有对厄洛替尼的居中敏感性的系。将欧几里得距离计量用于无中心化的聚类;配色方案代表绝对甲基化差异。

[0269] 特别地,来自这些 CpG 位点的甲基化信号将上皮样和间充质样细胞系聚类成其相应的上皮样和间充质样组,仅有 6 个例外:间充质样系 H1435、HCC4017、H647、H2228、H1755 和 HCC15 聚类为上皮样组。有意思的是,这 6 种系中有 5 种通过 EMT 基因表达分析一起紧密聚类到间充质样系的一个独特子集中,表明该基因表达表型与有些独特的根本性甲基化签名相关。重要的是,间充质样表型携带比上皮表型更大比例的高甲基化位点。这表明可能需要甲基化中的变化来稳定化 NSCLC 中 EMT 期间所获得的表型改变。

[0270] EGFR 突变体 NSCLC 通常在周围肺中呈现为分化较高的腺癌。基于其上皮样表达表型及其特征性组织学,EGFR 突变体细胞系的行为相比于间充质样系更多地类似于上皮样系。注意到基于对厄洛替尼的体外敏感性的细胞系分离模式(图 3,由中间的敏感性指示)。

几乎所有的厄洛替尼敏感性系都与上皮样表型相关,而几乎所有的间充质样系都对厄洛替尼抗性。然而,并非所有的上皮样系对于厄洛替尼都是敏感的。10种厄洛替尼抗性系聚类为上皮样系,而4种厄洛替尼敏感性系 H838、H2030、RERF-LC-MS 和 SK-MES-1 聚类为间充质样系。特别地,在使用我们先前定义的 EMT 表达签名通过基因表达聚类时 (11), H838 和 SK-MES-1 就厄洛替尼敏感性而言行为表现为外部者 (outlier)。一些其他外部者就厄洛替尼敏感性而言具有解释其明显抗性的突变。例如,上皮样系 H1975 携带 EGFR 中的 T790M 突变, H1993 携带 MET 扩增。这些遗传变化特异性赋予对厄洛替尼的抗性,表明观察到的表观遗传签名本身是细胞系生物学状态而非对厄洛替尼敏感性的替代品。

[0271] 实施例 4- 选定的 DMR 的亚硫酸氢钠测序验证 Infinium 甲基化概况分析

[0272] 检查由 Infinium 鉴定的与基因空间有关 (在 5' CpG 岛中或基因内) 的 17 种 DMR (图 4) 的甲基化状态,其通过对经亚硫酸氢钠转化的 DNA 的克隆片段的直接测序。选择 5 种上皮样系,4 种间充质样系,和 1 种中间系来测序验证。对于 10 个基因座的约 10 个克隆每细胞系的亚硫酸氢盐测序揭示,近乎所有这些标志物在至少 4 种间充质样细胞系和中间系 H522 中几乎完全甲基化。相比之下,这些基因座在所有 5 种上皮样系中完全未甲基化。在间充质样系中甲基化的 10 种标志物中的 4 种, ESRP1 和 CP2L3/GRHL2、miR200C 和 MST1R/RON, 牵涉上皮分化 (2、27、28)。ESRP1 是可变剪接的上皮特异性调控物,其在间充质细胞中下调,而 CP2L3/GRHL2 是顶端连接复合物的转录调控物 (27、28); miR200C 是 EMT 诱导物 ZEB1 的已知负向调控物 (29)。ESRP1 和 GRHL2 表达在相对于所有上皮样系的更大的间充质样系的集中下调,与间充质细胞中已知的 ESRP 蛋白缺少和这些蛋白调控上皮转录本 (转换 EMT 期间的剪接) 的能力一致。焦磷酸测序分析指示 GRHL2 还在相对于上皮样系更广的间充质样系的集中高甲基化。

[0273] 实施例 5-DMR 的生物学相关性

[0274] 为了评估甲基化在调控与选定 DMR 有关的基因表达中的作用,在一组 34 种经 5-氮杂-2'-脱氧胞苷 (5-氮杂-dC) 和二甲亚砷处理的 NSCLC 细胞系中实施定量 PCR。并非所有的 DMR 在 5-氮杂-dC 处理后与明显的基因表达变化有关,但注意到间充质样对上皮样系中显著的 GRHL2、ESRP1 和 CLDN7 转录本的诱导。从该组基因中选择 CLDN7 为 EMT 的代表性标志物且其甲基化状态在扩展的 42 种细胞系的集中通过焦磷酸测序定量。近乎所有的间充质样系在 CLDN7 启动子区处甲基化且展现出响应 5-氮杂-dC 处理的急剧的 CLDN7 表达诱导 (>10 倍) (图 5A 和 B)。相比之下, CLDN7 在多数的上皮样细胞系中表达,且并未受到 5-氮杂-dC 处理的进一步诱导。这些数据显示, NSCLC 细胞系中上皮样和间充质样状态有关的基因子集中基因座特异性的 DNA 高甲基化与转录沉默的直接联系。

[0275] 在图 5A 中,在 7 个 CpG 位点处通过 Pyromark 分析软件测定定量甲基化,其使用等式: %甲基化 = (C 峰高度 x100 / (C 峰高度 + T 峰高度))。数据呈现为 7 个 CpG 位点处甲基化百分比的均值 ± SD。在图 5B 中,使用标准的  $\Delta Ct$  方法在 42 种 (n = 20 上皮样, 19 间充质样, 3 中间体) 经 DMSO 处理和经 5-氮杂-dC 处理的 NSCLC 细胞系中测定 CLDN7 mRNA 的相对表达。表达值计算为经 5-氮杂-dC 处理的相对于经 DMSO 处理的对照细胞的倍数变化。将数据标准化至看家基因 GAPDH 并表示为 2 个重复的均值。DMSO, 二甲亚砷; GAPDH, 甘油醛-3-磷酸脱氢酶。

[0276] 实施例 6- 定量 MSP 将 NSCLC 细胞系分类成上皮和间充质亚型并预测厄洛替尼敏

## 感性

[0277] 在通过直接测序分析独立验证 17 种标志物的甲基化状态后,分析 70 个 NSCLC 细胞系来确定这些标志物是否能正确分类上皮样和间充质样表型。基于亚硫酸氢钠测序分析,选择最好地将上皮样系与间充质样系区分的甲基化区域,并基于 TaqMan 技术设计定量甲基化特异性 PCR(qMSP)测定法。使用 qMSP 作为测定平台,因为其已显示具有在从患有癌症的患者获得的标本中检测肿瘤特异性启动子高甲基化的用途。该方法对于量化甲基化等位基因是高度灵敏度和特异性的,且易于适应于高通量形式,从而使其适于临床应用(30-33)。对于 MSP, TaqMan 技术优于基于 SYBR 的设计,原因是不充当引物的荧光探针赋予增加的测定法特异性。为了标准化样品用于 DNA 输入,设计经亚硫酸氢盐修饰的 RNase P 参照测定法来独立于其甲基化状态扩增输入 DNA。使用对照甲基化 DNA,源自外周血单核细胞的 DNA ( $N = 20$ ),和来自已知每个 DMR 甲基化状态的细胞系的 DNA 建立滴度曲线。值得注意的是,近乎所有开发的测定法对于甲基化的存在或缺失均产生基本二元性输出,这避免了定义截留点的需要。

[0278] 测试上皮 (E) 或间充质 (M) 状态的 13 种候选标志物以确定其是否基于 EMT 基因表达分类区分上皮样和间充质样细胞系,所述分类包括 RON/MST1R (M)、STX2 (M)、HOXC5 (M)、PEX5L (E)、FAM110A (M)、ZEB2 (E)、ESRP1 (M)、BCAR3 (E)、CLDN7 (M)、PCDH8 (E)、NKX6.2 (M)、ME3 (E) 和 GRHL2 (M)。使用  $P < 0.05$  的截留值,13 种标志物中有 10 种与上皮样或间充质样状态显著相关(图 6)。在此测定法中,利用 qMSP 测定来确定上皮样 ( $n = 36$ ) 和间充质样 ( $n = 34$ ) NSCLC 细胞系中的甲基化。总输入 DNA 使用亚硫酸氢盐特异性 RNase P TaqMan 探针标准化。在图 6 中,对于 y 轴上的每份样品,甲基化水平绘制为  $-\Delta C_t$  (指示靶基因 -RNase P)。增加的  $-\Delta C_t$  值指示增加的甲基化。细胞系通过 x 轴上的上皮样/间充质样状态分组。P 值使用双侧不配对 Student t 检验测定。呈现 (B) RON、(D) FAM110A、(F) GRHL2 和 (H) ESRP1 的接受者操作特征 (ROC) 图。P 值使用 Wilcoxon 秩和检验测定。

[0279] 检查这些相同的标志物以确定其是否预测体外厄洛替尼敏感性。13 个 DMR 中有 7 个强烈预测厄洛替尼抗性(分别的  $P < 0.005$ ; 图 7),而 13 个 DMR 中有 3 个, PEX5L、ME3 和 ZEB2,与上皮表型显著相关,但不统计学预测厄洛替尼敏感性。在该测定中,使用指示的 qMSP 测定法实施 58 种 NSCLC 细胞系 DNA 样品的 qMSP 扩增。使用 R 统计学软件生成厄洛替尼敏感性对厄洛替尼抗性细胞系的 ROC 曲线。P 值使用 Student t 检验测定。图 7A-M 和图 8A-B。

[0280] 1. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. *CA Cancer J Clin* 2010 ;60:277 - 300.

[0281] 2. Singh A, Greninger P, Rhodes D, Koopman L, Violette S, Bardeesy N, et al. A gene expression signature associated with "K-Ras addiction" reveals regulators of EMT and tumor cell survival. *Cancer Cell* 2009 ;15:489 - 500.

[0282] 3. Herbst RS, Heymach JV, Lippman SM. Lung cancer. *N Engl J Med* 2008 ;359:1367 - 80.

[0283] 4. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, et al. International association for the study of lung cancer/American Thoracic Society/European Respiratory Society: international multidisciplinary

classification of lung adenocarcinoma. *J Thorac Oncol*2011 ;6:244 - 85.

[0284] 5. Sato M, Shames DS, Gazdar AF, Minna JD. A translational view of the molecular pathogenesis of lung cancer. *J Thorac Oncol*2007 ;2:327 - 43.

[0285] 6. Patel NV, Acarregui MJ, Snyder JM, Klein JM, Sliwkowski MX, Kern JA. Neuregulin-1 and human epidermal growth factor receptors 2 and 3 play a role in human lung development in vitro. *Am J Respir Cell Mol Biol*2000 ;22:432 - 40.

[0286] 7. Pao W, Chmielecki J. Rational, biologically based treatment of EGFR mutant non-small-cell lung cancer. *Nat Rev Cancer*2010 ;10:760 - 74.

[0287] 8. O'Byrne KJ, Gatzemeier U, Bondarenko I, Barrios C, Eschbach C, Martens UM, et al. Molecular biomarkers in non-small-cell lung cancer: a retrospective analysis of data from the phase 3 FLEX study. *Lancet Oncol*2011 ;12:795 - 805.

[0288] 9. Wu J-Y, Wu S-G, Yang C-H, Chang Y-L, Chang Y-C, Hsu Y-C, et al. Comparison of gefitinib and erlotinib in advanced NSCLC and the effect of EGFR mutations. *Lung Cancer*2011 ;72:205 - 12.

[0289] 10. Shepherd FA, Rodrigues Pereira J, Ciuleanu T, Tan EH, Hirsh V, Thongprasert S, et al. Erlotinib in previously treated non-small-cell

[0290] lung cancer. *N Engl J Med*2005 ;353:123 - 32.

[0291] 11. Yauch RL, Januario T, Eberhard DA, Cavet G, Zhu W, Fu L, et al. Epithelial versus mesenchymal phenotype determines in vitro sensitivity and predicts clinical activity of erlotinib in lung cancer patients. *Clin Cancer Res*2005 ;11:8686 - 98.

[0292] 12. Suda K, Tomizawa K, Fujii M, Murakami H, Osada H, Maehara Y, et al. Epithelial to mesenchymal transition in an epidermal growth factor receptor-mutant lung cancer cell line with acquired resistance to erlotinib. *J Thorac Oncol*2011 ;6:1152 - 61.

[0293] 13. Sequist LV, Waltman BA, Dias-Santagata D, Digumarthy S, Turke AB, Fidias P, et al. Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. *Sci Transl Med*2011 ;3:75ra26.

[0294] 14. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*2011 ;144:646 - 74.

[0295] 15. Singh A, Settleman J. EMT, cancer stem cells and drug resistance: an emerging axis of evil in the war on cancer. *Oncogene*2010 ;29:4741 - 51.

[0296] 16. Kalluri R, Weinberg RA. The basics of epithelial-mesenchymal transition. *J Clin Invest*2009 ;119:1420 - 8.

[0297] 17. Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY, et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell*2008 ;133:704 - 15.

[0298] 18. Davidson NE, Sukumar S. Of Snail, mice, and women. *Cancer Cell*2005 ;8:173 - 4.

[0299] 19. Kang Y, Massague J. Epithelial-mesenchymal transitions: twist in

development and metastasis. *Cell*2004 ;118:277 - 9.

[0300] 20. Dumont N, Wilson MB, Crawford YG, Reynolds PA, Sigaroudinia M, Tlsty TD. Sustained induction of epithelial to mesenchymal transition activates DNA methylation of genes silenced in basal-like breast cancers. *Proc Natl Acad Sci U S A*2008 ;105:14867 - 72.

[0301] 21. Wiklund ED, Bramsen JB, Hulf T, Dyrskjot L, Ramanathan R, Hanse TB, et al. Coordinated epigenetic repression of the miR-200family and miR-205in invasive bladder cancer. *Int J Cancer*2011 ;128:1327 - 34.

[0302] 22. Stinson S, Lackner MR, Adai AT, Yu N, Kim HJ, O'Brien C, et al. miR-221/222targeting of trichorhinophalangeal1 (TRPS1)promotes epithelial-to-mesenchymal transition in breast cancer. *Sci Signal*2011 ;4:pt5.

[0303] 23. Stinson S, Lackner MR, Adai AT, Yu N, Kim HJ, O'Brien C, et al. TRPS1targeting by miR-221/222promotes the epithelial-to-mesenchymal transition in breast cancer. *Sci Signal*2011 ;4:ra41.

[0304] 24. McDonald OG, Wu H, Timp W, Doi A, Feinberg AP. Genome-scale epigenetic reprogramming during epithelial-to-mesenchymal transition. *Nat Struct Mol Biol*2011 ;18:867 - 74.

[0305] 25. Du P, Kibbe WA, Lin SM. lumi:a pipeline for processing Illumina microarray. *Bioinformatics*2008 ;24:1547 - 8.

[0306] 26. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*2010 ;11:587.

[0307] 27. Warzecha CC, Jiang P, Amirikian K, Dittmar KA, Lu H, Shen S, et al. An ESRP-regulated splicing programme is abrogated during the epithelial-mesenchymal transition. *EMBO J*2010 ;29:3286 - 300.

[0308] 28. Werth M, Walentin K, Aue A, Schonheit J, Wuebken A, Pode-Shakked N, et al. The transcription factor grainyhead-like2regulates the molecular composition of the epithelial apical junctional complex. *Development*2010 ;137:3835 - 45.

[0309] 29. Tryndyak VP, Beland FA, Pogribny IP. E-cadherin transcriptional down-regulation by epigenetic and microRNA-200family alterations is related to mesenchymal and drug-resistant phenotypes in human breast cancer cells. *Int J Cancer*2010 ;126:2575 - 83.

[0310] 30. Belinsky SA, Liechty KC, Gentry FD, Wolf HJ, Rogers J, Vu K, et al. Promoter hypermethylation of multiple genes in sputum precedes lung cancer incidence in a high-risk cohort. *Cancer Res*2006 ;66:3338 - 44.

[0311] 31. Brock MV, Hooker CM, Ota-Machida E, Han Y, Guo M, Ames S, et al. DNAmethylation markers and early recurrence in stage I lung cancer. *N Engl J Med*2008 ;358:1118 - 28.

[0312] 32. Shivapurkar N, Stastny V, Suzuki M, Wistuba II, Li L, Zheng Y, et

al. Application of a methylation gene panel by quantitative PCR for lung cancers. *Cancer Lett*2007 ;247:56 - 71.

[0313] 33. Shames DS, Girard L, Gao B, Sato M, Lewis CM, Shivapurkar N, et al. A genome-wide screen for promoter methylation in lung cancer identifies novel methylation markers for multiple malignancies. *PLoS Med*2006 ; 3:e486.

[0314] 34. Hirsch FR, Varella-Garcia M, Cappuzzo F. Predictive value of EGFR and HER2 overexpression in advanced non-small-cell lung cancer. *Oncogene*2009 ; 28Suppl1:S32 - 7.

[0315] 35. Moulder SL, Yakes FM, Muthuswamy SK, Bianco R, Simpson JF, Arteaga CL. Epidermal growth factor receptor (HER1) tyrosine kinase inhibitor ZD1839 (Iressa) inhibits HER2/neu (erbB2)-overexpressing breast cancer cells in vitro and in vivo. *Cancer Res*2001 ;61:8887 - 95.

[0316] 36. Hoeflich KP, O'Brien C, Boyd Z, Cavet G, Guerrero S, Jung K, et al. In vivo antitumor activity of MEK and phosphatidylinositol 3-kinase inhibitors in basal-like breast cancer models. *Clin Cancer Res*2009 ;15:4649 - 64.

[0317] 37. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell*2006 ;10:515 - 27.

[0318] 38. Noshmehr H, Weisenberger DJ, Diefes K, Phillips HS, Pujara K, Berman BP, et al. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell*2010 ;17:510 - 22.

[0319] 39. Fang F, Turcan S, Rimmer A, Kaufman A, Giri D, Morris LG, et al. Breast cancer methylomes establish an epigenomic foundation for metastasis. *Sci Transl Med*2011 ; 3:75ra25.

[0320] 40. Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature*1983 ;301:89 - 92.

[0321] 41. Baylin SB, Jones PA. A decade of exploring the cancer epigenome—biological and translational implications. *Nat Rev Cancer*2011 ;11:726 - 34.

[0322] 42. Sharma SV, Lee DY, Li B, Quinlan MP, Takahashi F, Maheswaran S, et al. Achromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell*2010 ;141:69 - 80.

[0323] 43. Benjamini, Y. and Y. Hochberg (1995). "Controlling the false discovery rate: a practical and powerful approach to multiple testing." *J Royal Statist Soc B*57(1) :289-300.

[0324] 44. Du, P., W. A. Kibbe, et al. (2008). "lumi: a pipeline for processing Illumina microarray." *Bioinformatics*24(13) :1547-1548.

[0325] 45. Du, P., X. Zhang, et al. (2010). "Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis." *BMC*

Bioinformatics11:587.

[0326] 46. Ramirez, R. D., S. Sheridan, et al. (2004). "Immortalization of human bronchial epithelial cells in the absence of viral oncoproteins." *Cancer Res*64(24):9027-9034.

[0327] 47. Sato, M., M. B. Vaughan, et al. (2006). "Multiple oncogenic changes (K-RAS(V12), p53knockdown, mutant EGFRs, p16bypass, telomerase) are not sufficient to confer a full malignant phenotype on human bronchial epithelial cells." *Cancer Res*66(4):2116-2128.

[0328] 48. Walter, K., et al, (2012). "DNA Methylation Profiling Defines Clinically Relevant Biological Subsets of Non - Small Cell Lung Cancer." *Clin Cancer Res*18:2360.

[0329] 提述并入

[0330] 本文中公开的所有专利、公开的专利申请和其他参考文献通过提述完整明确地并入本文。

[0331] 等同

[0332] 本领域技术人员将认可或能够确定,仅使用常规实验的本文中特定描述的本发明特定实施方案的许多等同体。这类等同体意图涵盖在所附权利要求的范围内。术语“包含/包括”用于本文是非限制性的,且包括指定的要素而不限制其他要素的纳入。

基于 Fluidigm 的 EMT 基因表达集

基因标志符	ABI 测定 ID	基因描述	上皮 / 间充质
CDH1	Hs01023894_m1	E-钙粘蛋白	E
CDH3	Hs00354998_m1	P-钙粘蛋白	E
CLDN7	Hs00600772_m1	密蛋白 7	E
ERBB3	Hs00176538_m1	V-erb-b2 成红细胞白血病病毒基因同源物 3	E
GJB3	Hs02378125_s1	间隙连接蛋白, $\beta 3$	E
JUP	Hs00158408_m1	连接盘状球蛋白 (Junction plakoglobin)	E
KRT7	Hs00559840_m1	角蛋白 7	E
LAD1	Hs00194326_m1	Ladinin 1	E
LAMB3	Hs00165078_m1	层粘连蛋白, $\beta 3$	E
PTK6	Hs00178742_m1	蛋白酪氨酸激酶 6	E
FGFR1	Hs00241111_m1	成纤维细胞生长因子受体 1	M
HOXC6	Hs00171690_m1	同源框 C6	M
MAGED4	Hs00602656_g1	黑素瘤抗原家族 D, 4	M
STX2	Hs00181827_m1	突触融合蛋白 2	M
TGFB1	Hs00998133_m1	转化生长因子, $\beta 1$	M
TUBB	Hs00962420_g1	微管蛋白, $\beta$	M
UCHL1	Hs00985157_m1	泛素羧基末端酯酶 L1	M
EMP3	Hs00171319_m1	上皮膜蛋白 3	M
VIM	Hs00958116_m1	波形蛋白	M
ZEB1	Hs00232783_m1	锌指 E 盒结合同源框 1	M

图 1

图2  
图2A  
图2B

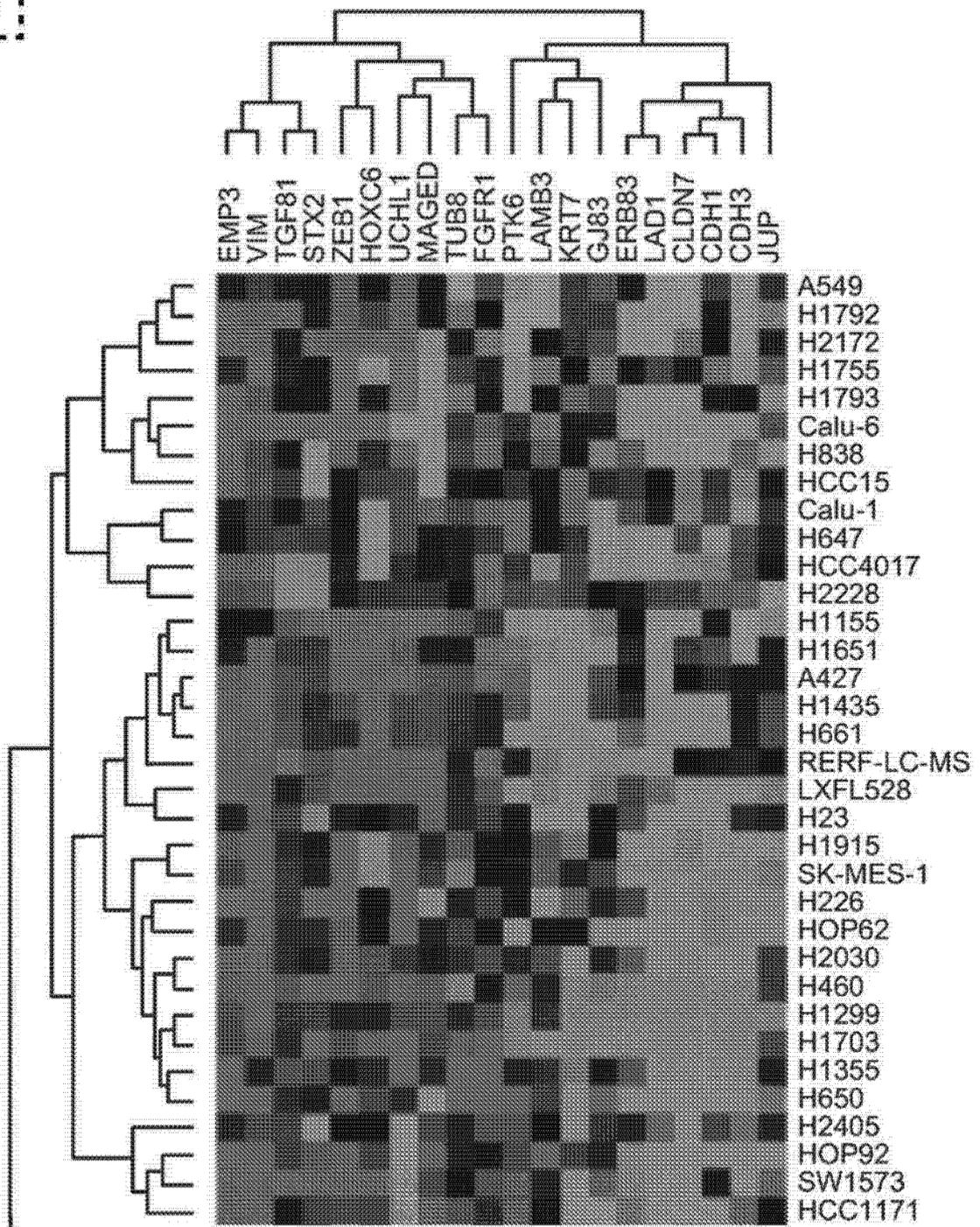


图 2A

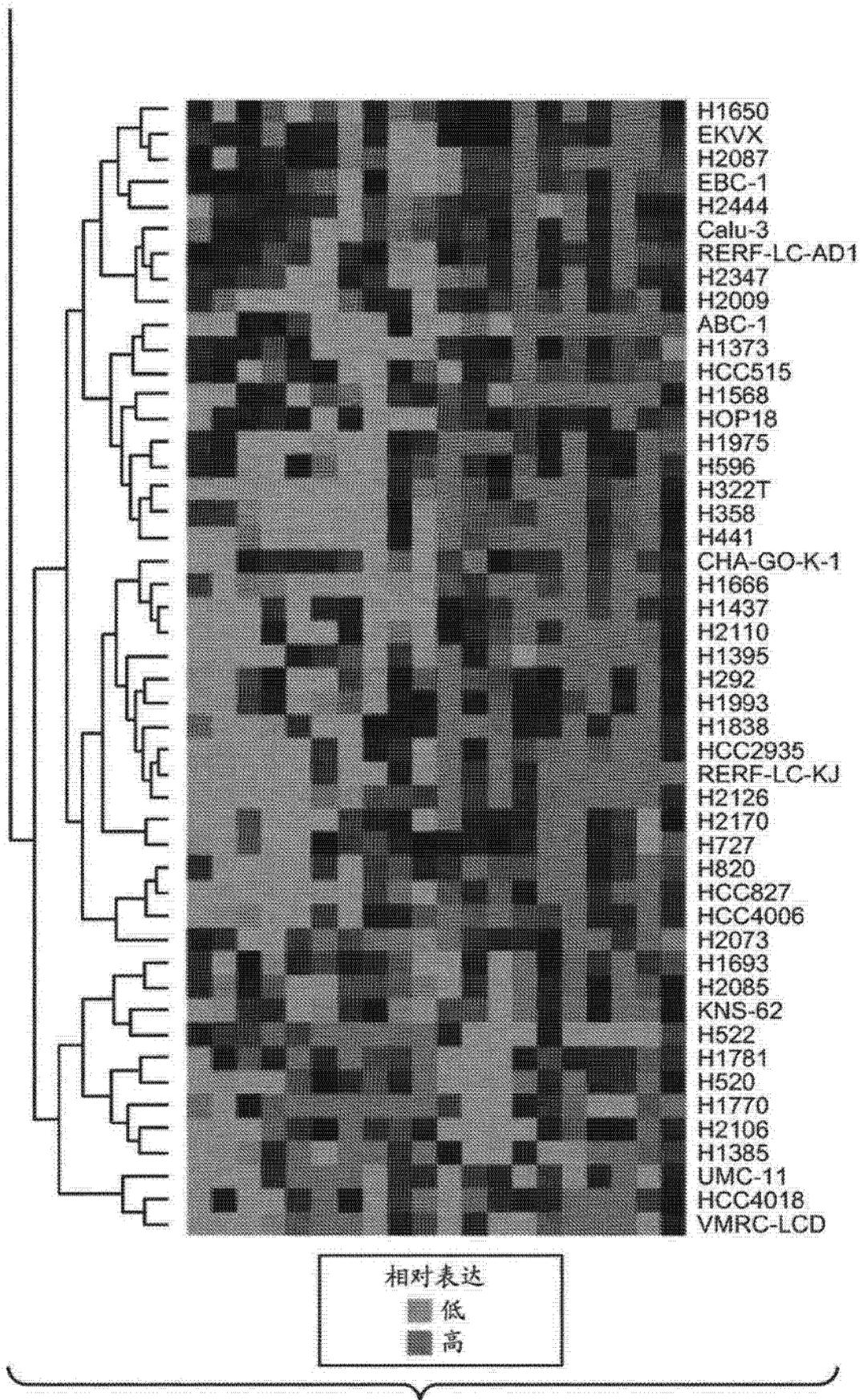


图 2B

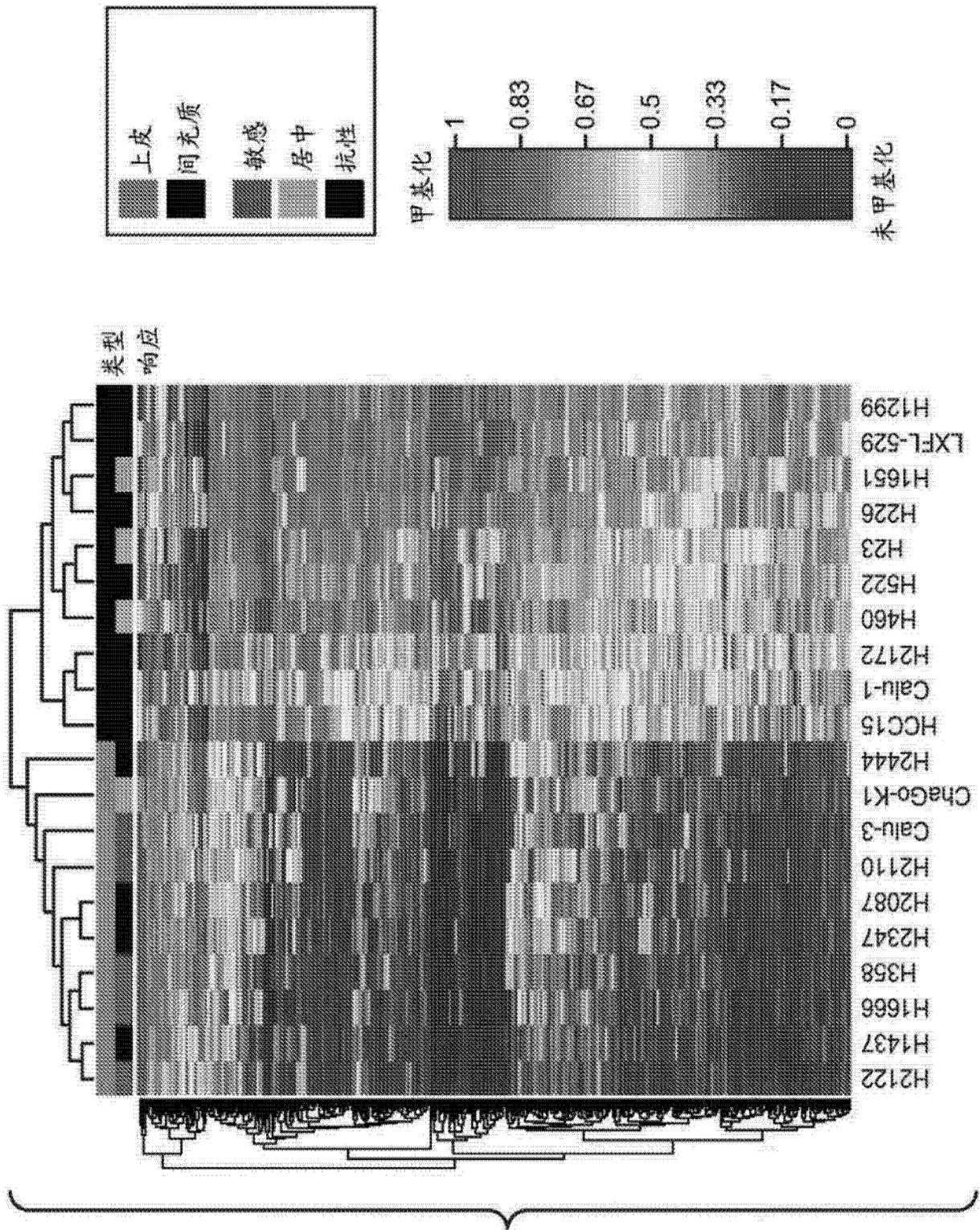


图 3

图 4



基因 标识符	基因名称	染色体位置	调控元件	5-羧杂- dC 诱导	PBMC	正常肺	上皮/间充质	P/E 对 M)	P/S 对 R)
ZEB2	锌指 E 盒结合同源框 2	chr2: 144,989,352-144,989,168	保守调控潜力	是	未甲基化	否	E	0.0012	0.3432
NKX6.2	同源框蛋白 Nkx-6.2	chr10: 134,448,826-134,449,879	CpG 岛	NE	部分甲基化	否	E	0.0568	0.8353
PEX5L	过氧化氢酶体生源 因子 5 样	chr3: 181,236,933-181,237,780	CpG 岛	否	未甲基化	否	E	0.0084	0.6566
GALR1	神经节肽受体 1	chr18: 73,090,412-73,090,797	CpG 岛	NE	未甲基化	否	E	NE	NE
PTPRM	蛋白质酪氨酸磷酸酶, 受体类型, M	chr18: 7,932,674-7,933,993	保守调控潜力	是	甲基化	是	E	NE	NE
ME3	NADP 依赖性苹果酸酶 3	chr11: 86,060,344-86,061,158	CpG 岛	一些	未甲基化	否	E	0.0295	0.5771
SYK	脾酪氨酸激酶	chr9: 92,631,210-92,632,740	未限定	是	甲基化	是	E	NE	NE
PCDH8	原钙粘蛋白 8	chr13: 52,321,012-52,321,485	CpG 岛	是	未甲基化	否	E	0.0656	0.5107
HOXC5	同源框 C5	chr12: 52,712,688-52,713,529	CpG 岛	NE	未甲基化	否	M	0.7493	0.0008
miR200C	microRNA 200c	chr12: 6,942,800-6,943,200	未限定	NE	甲基化	否	M	NE	NE

图 4A

SERPINS5	丝氨酸 (或半胱氨酸) 蛋白酶抑制剂, 分枝 B, 成员 5; 乳腺丝抑蛋白	chr18: 59,294,906-59,295,319	保守调控潜力	是	甲基化	是	M	NE	NE
BCAR3	乳腺癌抗雌激素 抗性 3	chr1: 93,852,868-93,853,418	保守调控潜力	一些	甲基化	是	M	0.001	0.2071
FAM110A	具有序列相似性的 家族 110, 成员 A	chr20: 822,480-822,120	未限定	一些	甲基化	否	M	<0.0001	0.0007
CLDN7	密蛋白 7	chr17: 7,103,446-7,106,446	CpG 岛	是	部分甲基化	否	M	<0.0001	0.0011
ESRP1	上皮剪接调控蛋白	chr8: 95,653,500-95,654,240	是	是	未甲基化	否	M	<0.0001	0.0043
GRHL2	Grainyhead 样 2	chr8: 102,575,373-102,575,793	CpG 岛	是	未甲基化	否	M	<0.0001	0.0004
RON	巨噬细胞刺激性 1 受体	chr3: 49,916,089-49,916,545	CpG 岛	是	未甲基化	一些	M	0.0009	0.0028
STX2	突触融合蛋白 2	chr12: 129,868,924-129,869,427	CpG 岛	否	部分甲基化	是	M	0.0001	0.0023
TBCD	微管蛋白特异性 分子伴侣 D	chr17: 78,440,425-78,440,951	未限定	NE	甲基化	是	M	NE	NE
ERBB2	v-erb-b2 成红细胞 白血病病毒基因 同源物 2	chr17: 37,861,100-37,863,650	推定的增强子	否	甲基化	否	M	<0.0001	0.0004

图 4B

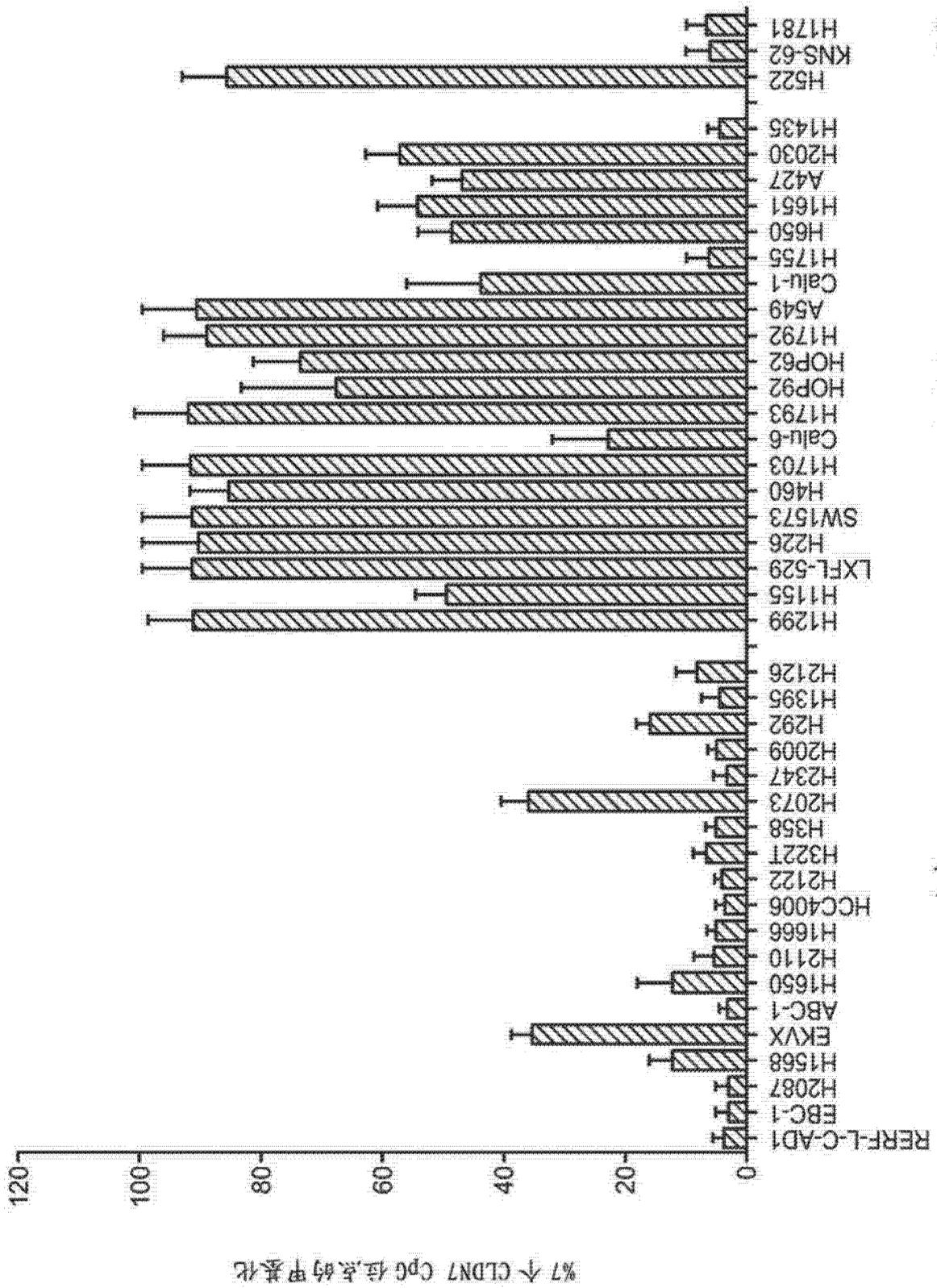


图 5A

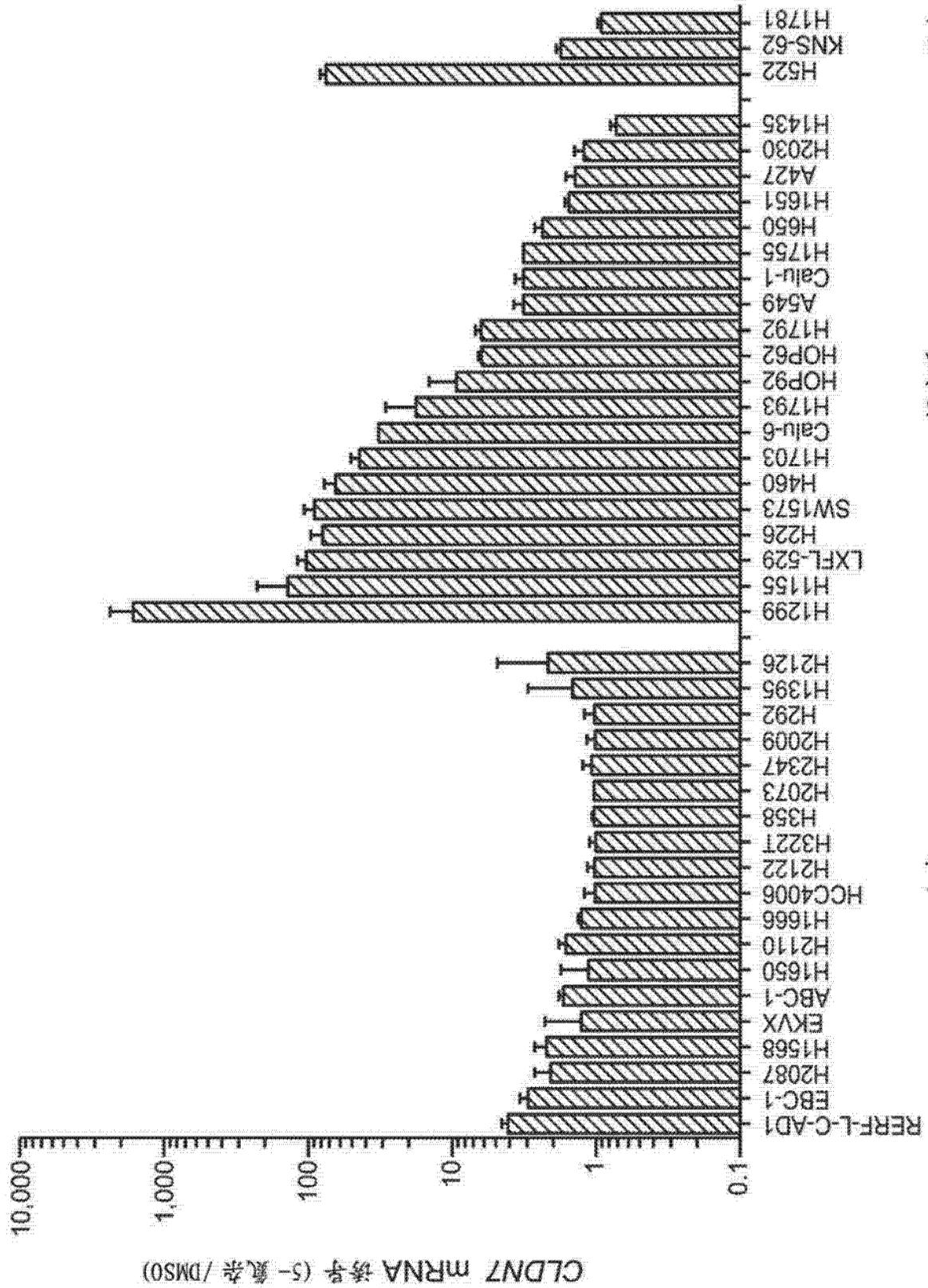


图 5B

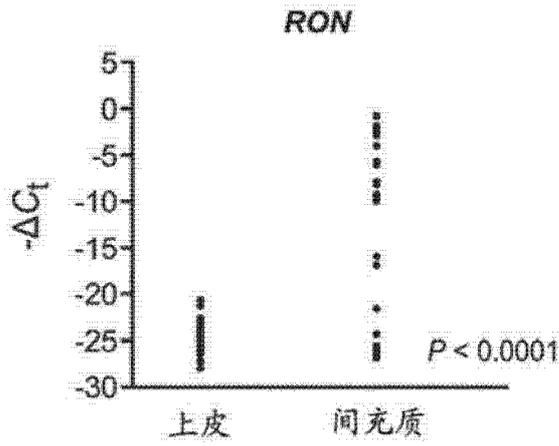


图 6A

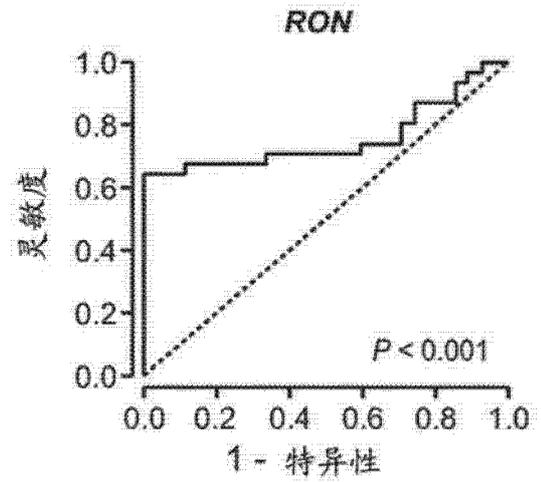


图 6B

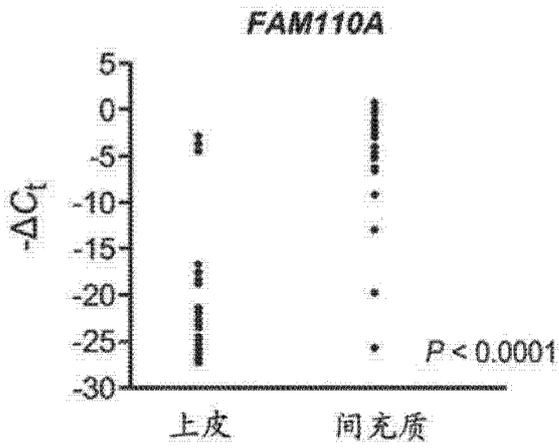


图 6C

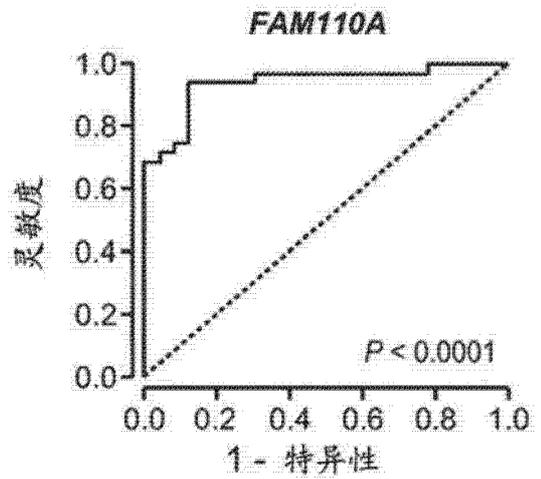


图 6D

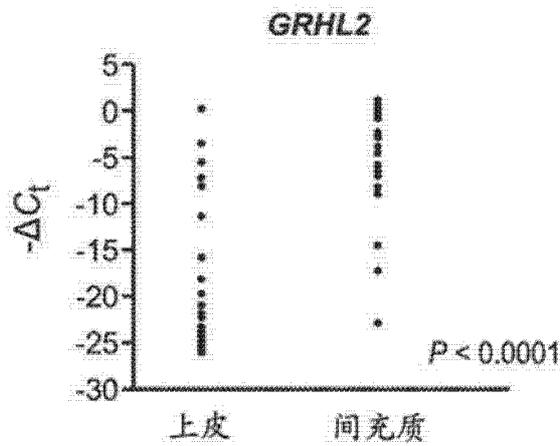


图 6E

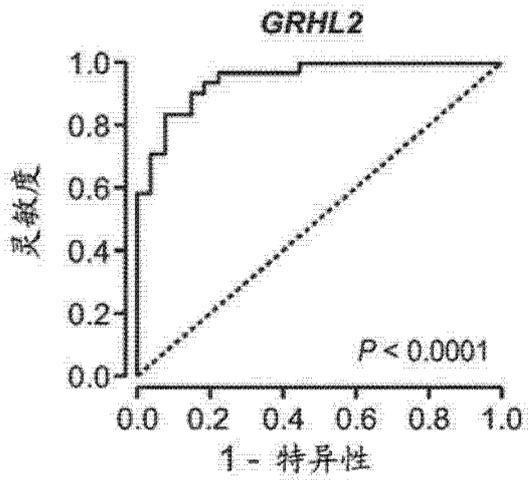


图 6F

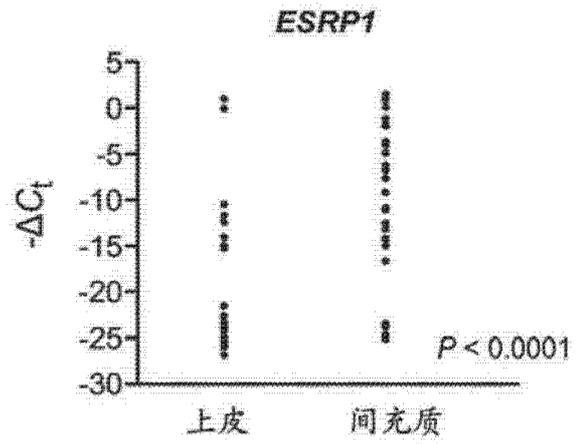


图 6G

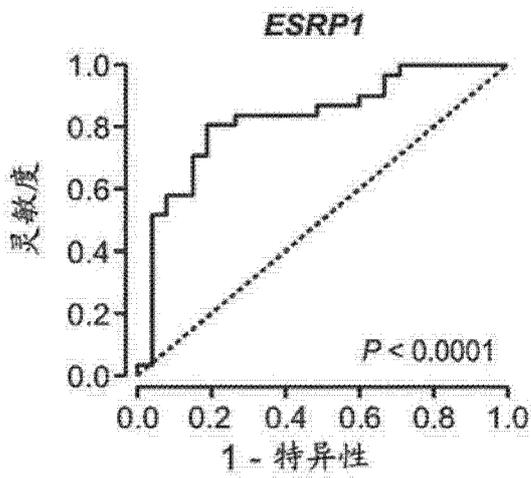


图 6H

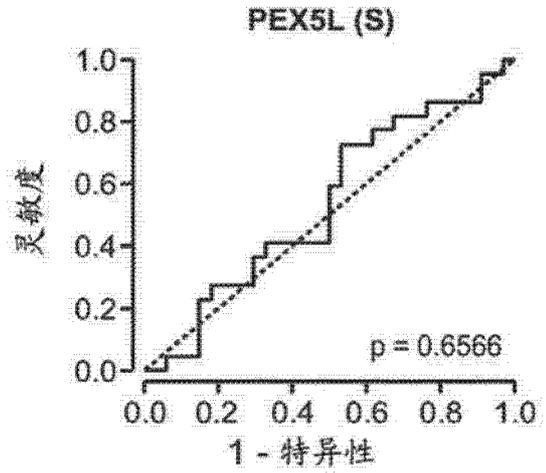


图 7A

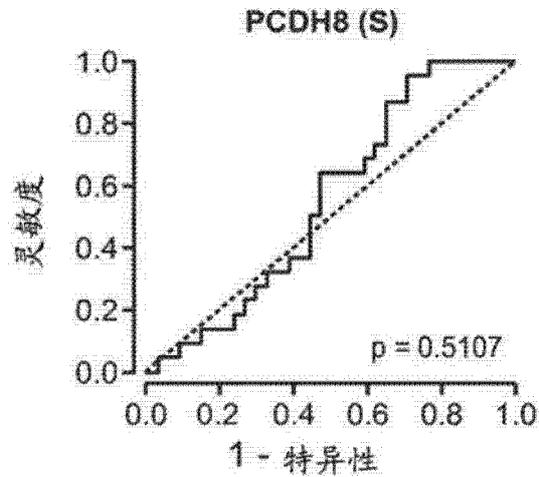


图 7B

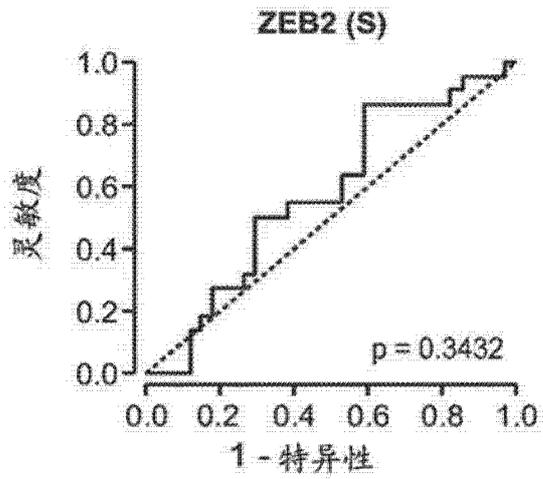


图 7C

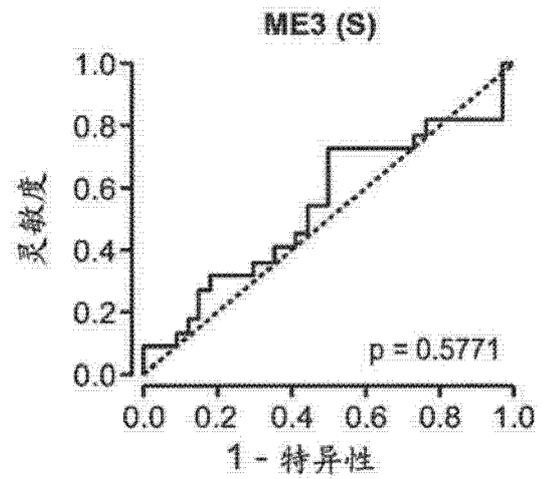


图 7D

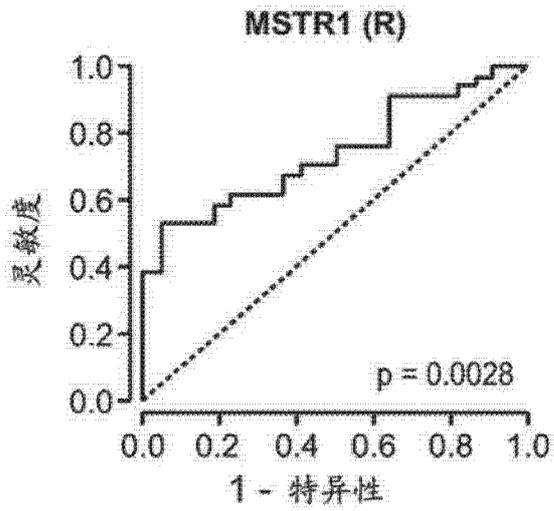


图 7E

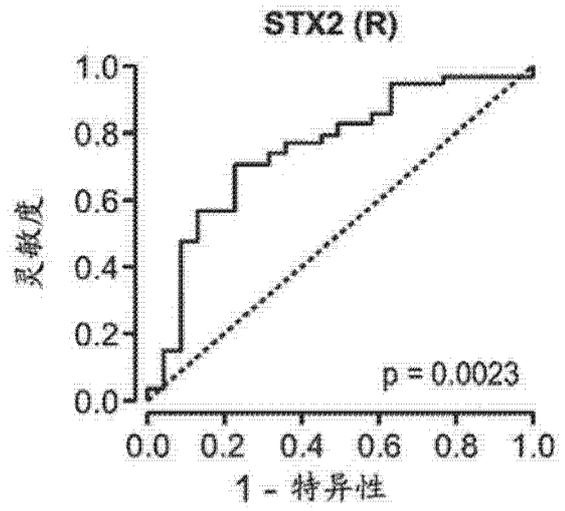


图 7F

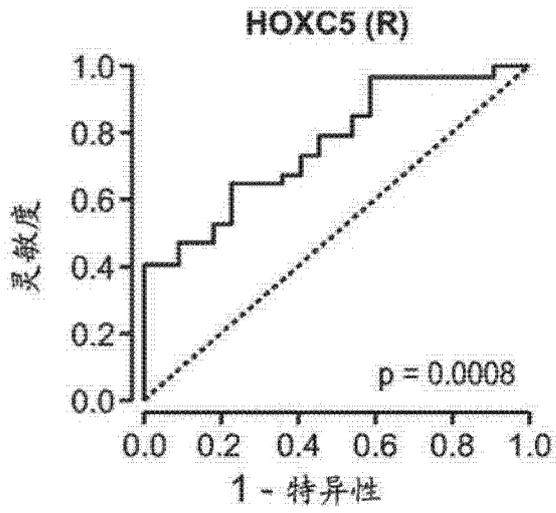


图 7G

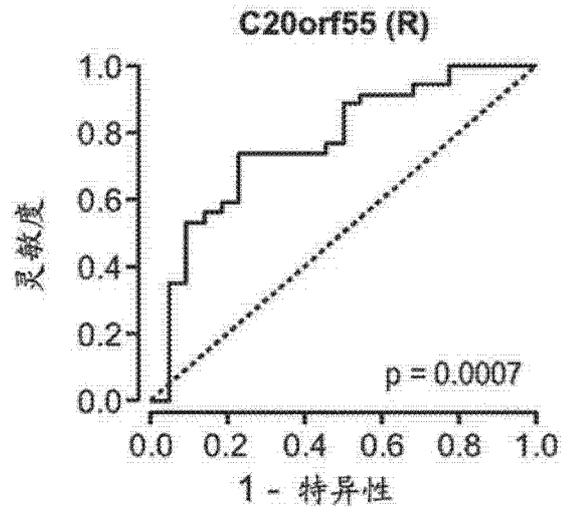


图 7H

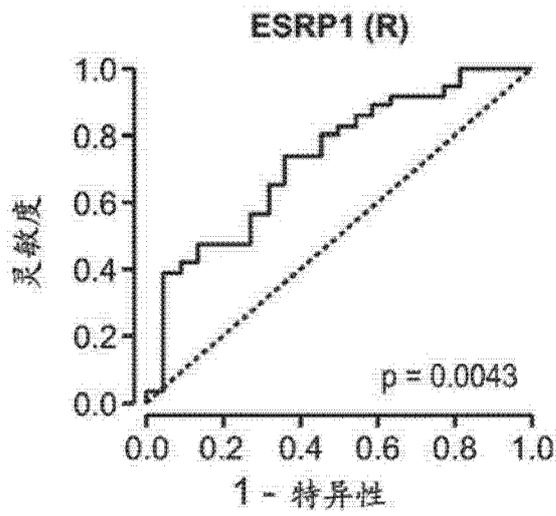


图 7I

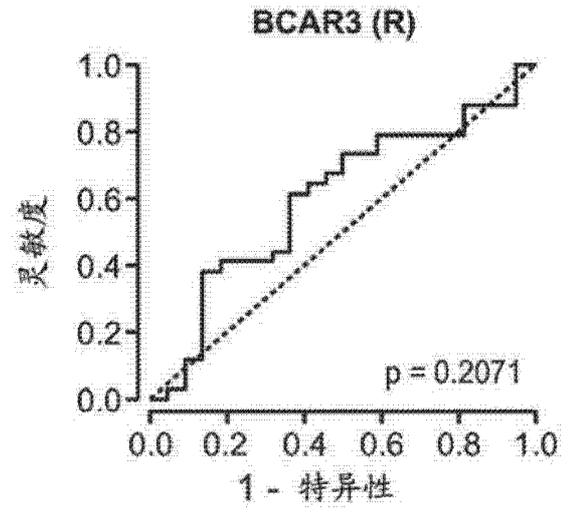


图 7J

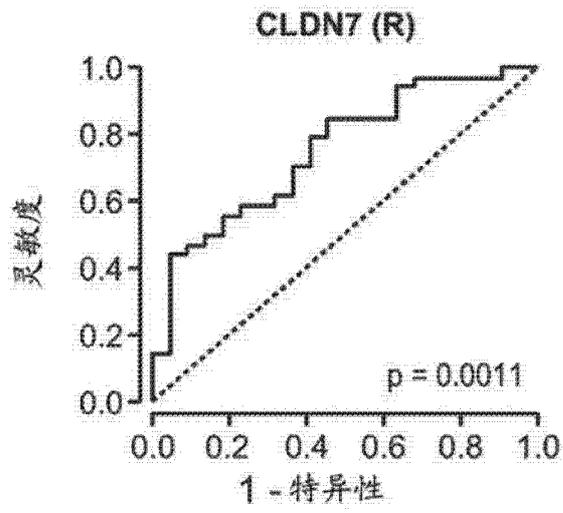


图 7K

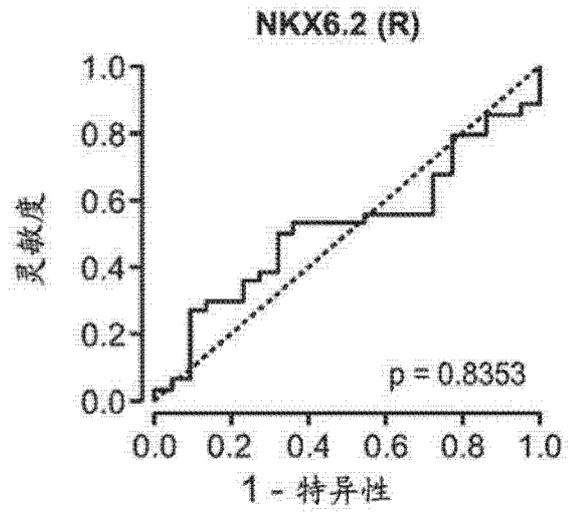


图 7L

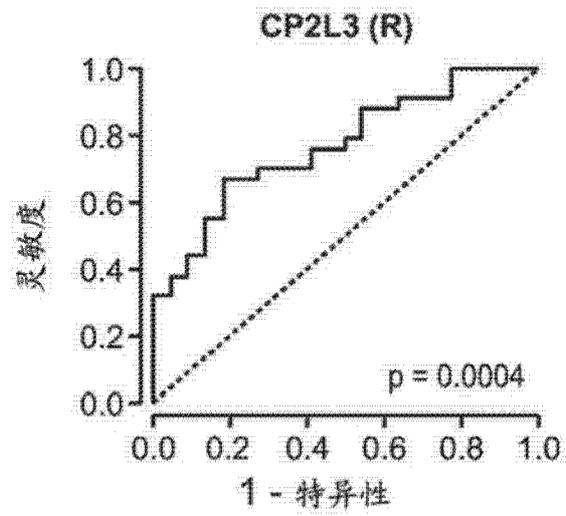
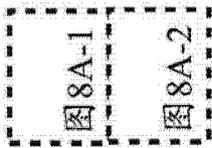


图 7M

图 8A



细胞系	组织来源	组织学	EIM	厄洛替尼 IC50 (μM) + (TGFα)	EGFR
H2122	PE	腺	E	0.7	WT
H358	肺	BAC	E	1.9	WT
H1437	PE	腺	E	10.0	WT
H1666	PE	腺	E	0.8	WT
H2110	PE	NSCLC (NOS)	E	0.9	WT
H2347	肺	腺	E	10.0	WT
H2087	LN	腺	E	10.0	WT
Calu-3	PE	腺	E	1.0	WT
H2444	肺	NSCLC (NOS)	E	10.0	WT
ChaGo-K-1	N/A	支气管癌	E	4.7	WT
H1650	肺	腺	E	0.5	[c.2235_2249del15; p.E746_A750del] (Sanger), 2235-2249 del (E746-A750 del) (Gazdar)
EKVX	肺	腺	E	1.6	WT
EBC-1	N/A	N/A	E	10	N/A
RERF-LC-Ad1	N/A	N/A	E	1.51	N/A
H2009	肺	腺	E	1.5	WT
ABC-1	N/A	N/A	E	10	N/A
H1373	肺	腺	E	N/A	WT
HCC515	肺	腺	E	8.3	WT
H1568	肺	腺	E	1.7	WT

图 8A-1

HOP18	N/A	N/A	E	0.5	N/A
H1975	肺	腺	E	10	[(c.2369C>T; p.T790M), [c.2573>G; p.L858R] (Sanger), T790M, L858R (Gazdar)]
H596	肺	腺鳞状	E	1.1	WT
H322T	N/A	N/A	E	0.7	N/A
H441	肺	腺	E	2.8	WT
CHA-GO-K1	N/A	支气管癌	E	4.65	WT
H1395	肺	腺	E	1.7	WT
H292	肺	NSCLC (NOS)	E	0.8	WT
H1993	肺	腺	E	10	WT
H1838	肺	腺	E	1.6	WT
HCC2935	肺	腺	E	0.11	2235-2252 del, 2254T>A, 2255C>T (E746-S752 del, I ins) (Gazdar)
RERF-LC-KJ	N/A	N/A	E	0.34	N/A
H2126	肺	腺	E	0.9	WT
H2170	肺	SCC	E	1.928	WT
H727	肺	类癌	E	10	WT
H820	肺	腺	E	1	2235-2246 del, (E746-E749 del), T790M (Gazdar)
HCC827	肺	腺	E	0.01	2236-2250 del, (E746-A750 del) (Gazdar)
HCC4006	肺	腺	E	0.11	2239-2247 del, 2248G>C (L747-A750 del, P ins) (Gazdar)
H2073	肺	腺	E	0.63	WT

图 8A-2

图8B

图8B-1

图8B-2

细胞系	组织来源	组织学	E/M	厄洛替尼 IC50 (μM) + (TGFα)	EGFR
A549	肺	腺	M	3.9	WT
H1792	肺	腺	M	5	WT
H1755	肺	腺	M	10	WT
H1793	肺	腺	M	N/A	WT
Calu-6	肺	腺	M	10	WT
H838	肺	腺	M	1.2	WT
H647	肺	腺	M	6.3	WT
HCC4017	肺	腺鳞状	M	10	WT
H2228	肺	大细胞	M	10	WT
H1155	肺	腺	M	10	WT
A427	肺	大细胞神经内分泌	M	4.9	WT
H1435	肺	N/A	M	0.9	WT
H661	肺	腺	M	6	WT
RERF-LC-MS	N/A	LCC	M	0.47	WT
H1915	肺	N/A	M	8.65	N/A
SK-MES-1	N/A	LLC	M	0.3	N/A
HOP62	肺	N/A	M	3.6	WT
H2030	肺	腺	M	2	WT
H1703	肺	腺	M	8.4	WT
H1355	肺	腺	M	10	WT

图 8B-1

H650	肺	腺	M	10	WT
H2405	肺	腺	M	5.6	WT
HOP92	肺	LCC	M	8.3	WT
SW1573	N/A	N/A	M	7.1	N/A
HCC1171	肺	NSCLC (NOS)	M	10	WT
H23	肺	腺	M	6.1	WT
H1299	LN	LCC	M	9.6	WT
Calu-1	PE	SCC	M	10.0	WT
H522	肺	腺	M	10.0	WT
H2172	肺	NSCLC (NOS)	M	10	WT
HCC15	液 FNA	SCC	M	10	WT
H226	PE	表皮状	M	10	WT
H1651	肺	腺	M	6.2	WT
LXFL529	N/A	LCC	M	10	WT
H460	PE	LCC	M	6.7	WT
KNS-62	N/A	N/A	居中	10	N/A
H1781	肺	腺	居中	10	WT
H520	肺	SCC	居中	10	WT
H1770	肺	神经内分泌	居中	10	WT
H2106	肺	大细胞神经内分泌	居中	10	N/A
H1385	肺	鳞状	居中	10	WT
UMC-11	肺	类癌-内分泌	居中	10	WT
HCC4018	肺	腺	居中	10	WT
VMRC-LCD	N/A	N/A	居中	10	N/A
H1693	肺	腺	居中	2.24	WT

图 8B-2