



(12) 发明专利

(10) 授权公告号 CN 104011543 B

(45) 授权公告日 2016.06.15

(21) 申请号 201280063658.6

(22) 申请日 2012.10.24

(30) 优先权数据

61/550,617 2011.10.24 US

(85) PCT国际申请进入国家阶段日

2014.06.20

(86) PCT国际申请的申请数据

PCT/US2012/061576 2012.10.24

(87) PCT国际申请的公布数据

W02013/063035 EN 2013.05.02

(73) 专利权人 通用医疗公司

地址 美国马萨诸塞州

(72) 发明人 D. T. 廷 S. 马赫斯瓦兰 D. A. 哈伯

(74) 专利代理机构 北京市柳沈律师事务所

11105

代理人 张文辉

(51) Int. Cl.

C12Q 1/68(2006.01)

(56) 对比文件

EP 1961825 A1, 2007.02.26,

CN 101313220 A, 2008.11.26,

WO 2006/108048 A1, 2006.10.12,

US 2002/0042072 A1, 2002.04.11,

Bin Chen. The HSP90 family of genes in the human genome: Insights into their divergence and evolution. 《Genomics》. 2005, 第 86 卷第 627-637 页.

Wenyue Sun. Identification of differentially expressed genes in human lung squamous cell carcinoma using suppression subtractive hybridization. 《Cancer Letters》. 2004, 第 212 卷

审查员 刘迎鸣

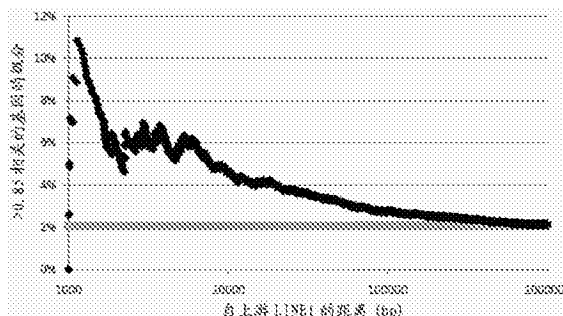
权利要求书2页 说明书20页 附图9页

(54) 发明名称

癌症的生物标记

(57) 摘要

基于检测存在随体相关基因提高的表达水平来诊断癌症并监控治疗效果的方法。



1. 一个或多个随体相关基因在制备用于检测受试者中存在癌症的体外方法中的试剂中的用途,该方法包括:

确定包含来自受试者的测定细胞的样品中的一个或多个随体相关基因的表达水平,以获得测定值,所述随体相关基因是HSP90BB(胞质的热激蛋白90kDa α ,B类成员2,假基因(HSP90AB2P))或HSP90BB和下列任一项或多项:NR_003133(人鸟苷酸结合蛋白1,干扰素-诱导的假基因1(GBP1P1),非编码RNA);BX649144(微管蛋白酪氨酸连接酶(TTL));DERP7(跨膜蛋白45A(TMEM45A));MGC4836(人类似假拟蛋白(L1H_3区域));BC037952cDNA克隆;AK056558cDNA克隆;NM_001001704(FLJ44796假拟的);ODF2L(精子尾外周致密纤维2-样(ODF2L));BC041426(C12orf55染色体12开放阅读框55(C12orf55));REX01L1(RNA外切核酸酶1同系物(酿酒酵母(*S.cerevisiae*))-样1(REX01L1));AK026100(FLJ22447假拟的LOC400221(FLJ22447));AK026825(跨膜蛋白212(TMEM212));KENAE1(人Kenae1mRNA(AB024691));HESRG(ESRG假拟的LOC790952(ESRG));AK095450(LOC285540假拟的LOC285540);FLJ36492(CCR4-NOT转录复合物,亚基1(CNOT1));AK124194(FLJ42200蛋白质);AK096196(假拟的LOC100129434);AK131313(锌指蛋白91假基因(LOC441666));FLJ11292(假拟蛋白FLJ11292);CCDC122(含122(CCDC122)的卷曲螺旋结构域);和BC070093cDNA克隆,和

将所述的测定值与参考值相比较,

其中,测定值明显高于参考值表明所述受试者有癌症,其中所述癌症是胰腺癌。

2. 权利要求1所述的用途,其中所述参考水平是所述随体相关基因在正常细胞中的水平。

3. 权利要求2所述的用途,其中所述正常细胞是相同受试者中与测定细胞相同类型的细胞。

4. 权利要求2所述的用途,其中所述正常细胞是受试者中与测定细胞相同类型的细胞,所述受试者没有癌症。

5. 权利要求1所述的用途,其中所述样品已知或被怀疑包含肿瘤细胞。

6. 权利要求5所述的用途,其中所述样品是已知或被怀疑含有循环肿瘤细胞(CTC)的血液样品,或已知或被怀疑含有肿瘤细胞的活检样品。

7. 一个或多个随体相关基因在制备用于评估受试者中癌症治疗效果的体外方法中的试剂中的用途,该方法包括:

确定来自受试者的第一样品中一个或多个随体相关基因的表达水平以获得第一值,所述随体相关基因是HSP90BB(胞质的热激蛋白90kDa α ,B类成员2,假基因(HSP90AB2P))或HSP90BB和下列任一项或多项:NR_003133;BX649144;DERP7;MGC4836;BC037952;AK056558;NM_001001704;ODF2L;BC041426;REX01L1;AK026100;AK026825;KENAE1;HESRG;AK095450;FLJ36492;AK124194;AK096196;AK131313;FLJ11292;CCDC122;和BC070093;

对受试者施用癌症治疗;

确定在其后由受试者获得的后续样品中所述一个或多个随体相关基因的表达水平,以获得治疗值;和

将所述第一值与所述治疗值相比较,

其中,低于所述第一值的治疗值表明所述治疗是有效的,其中所述癌症是胰腺癌。

8. 权利要求7所述的用途,其中所述第一和后续样品已知或被怀疑包含肿瘤细胞。
9. 权利要求8所述的用途,其中所述样品是已知或被怀疑包含循环肿瘤细胞(CTC)的血液样品,或已知或被怀疑含有肿瘤细胞的活检样品。
10. 权利要求7或8所述的用途,其中所述治疗包括施用手术干涉,化疗,放疗,或其组合。
11. 前述任一项权利要求所述的用途,其中确定一个或多个随体相关基因的水平包括确定转录物的水平。
12. 权利要求11所述的用途,其中确定转录物的水平包括使与所述转录物特异性结合的寡核苷酸探针与所述样品接触。
13. 权利要求12所述的用途,其中所述探针是经标记的。
14. 权利要求12所述的用途,其包括扩增所述转录物。
15. 权利要求11所述的用途,其中确定转录物的水平包含进行RNA测序。
16. 前述任一项权利要求所述的用途,其中所述受试者是人。
17. 权利要求1-7任一项所述的用途,其中所述测定细胞来自受试者的胰腺,且明显高于参考值的测定值表明所述受试者有恶性导管内乳头状粘液癌(IPMN)。
18. 权利要求17所述的用途,其中所述受试者以前诊断有良性IPMN。
19. 前述任一项权利要求所述的用途,其中所述样品为冷冻的,固定的和/或经渗透处理的样品。
20. 前述任一项权利要求所述的用途,其中所述样品为福尔马林固定石蜡包埋(FFPE)的样品。
21. 前述任一项权利要求所述的用途,其中利用Northern印迹, RNA原位杂交(RNA-ISH), RNA表达测定, RT-PCR, RNA测序, 深度测序(deep sequencing)或克隆评估随体相关基因SCG mRNA的水平。
22. 前述任一项权利要求所述的用途,其中利用经修饰的RNA原位杂交技术应用支链DNA测定直接检测并评估样品中生物标记mRNA的水平。
23. 前述任一项权利要求所述的用途,其中通过所述样品与标记的寡核苷酸探针接触评估随体相关基因SCG mRNA的水平。

癌症的生物标记

[0001] 联邦政府资助的研发

[0002] 本发明得到由国家卫生研究院授予的授权号为No.CA129933的政府支持。政府对本发明享有一定权利。

发明领域

[0003] 基于检测存在随体(satellite)相关基因表达水平提高来诊断癌症并监控治疗效果的方法。

[0004] 发明背景

[0005] 全基因组测序手段揭示出越来越多经转录的非编码序列,包括与转录沉默和染色体整合相关联的基因组异染色质区域的“弥漫性转录(pervasive transcription)”(J.Berretta,A.Morillon,EMBO Rep10,973(Sep,2009);A.Jacquier,Nat Rev Genet10,833(Dec,2009))。在鼠中,异染色质由中心(次要的)和两侧(主要的)随体重复(satellite repeat)构成,所述的随体重复是有丝分裂纺锤体复合物形成以及精确的染色体分离所需的(M.Guenatri,D.Bailly,C.Maison,G.Almouzni,J Cell Biol166,493(Aug16,2004)),而人随体重复分为具有类似功能的多种类型(J.Jurka等人,Cytogenet Genome Res110,462(2005))。酵母中随体的双向转录,通过切酶介导RNA-诱导的转录沉默(RITS)以及通过最近鉴定出的不依赖切酶的途径,保持着丝粒DNA的沉默(M.Halic,D.Moazed,Cell1140,504(Feb19)),但是,哺乳动物中的着丝粒随体沉默机制并未得到充分鉴定(A.A.Aravin,G.J.Hannon,J.Brennecke,Science318,761(Nov2,2007))。在鼠和人细胞系中随体转录物的积累是由DICER1中的缺失(C.Kanellopoulou等人,Genes Dev19,489(Feb15,2005);T.Fukagawa等人,Nat Cell Biol6,784(Aug,2004)),DNA去甲基化,热激,或者凋亡诱导(H.Bouzinba-Segard,A.Guais,C.Francastel,Proc Natl Acad Sci U S A103,8709(Jun6,2006);R.Valgardsdottir等人,Nucleic Acids Res36,423(Feb,2008))造成的。培养细胞中的随体应激诱导转录已被和编码RNA聚合酶活动的反转录因子,如LINE-1,的活化相联系(D.Ugarkovic,EMBO Rep6,1035(Nov,2005);D.M.Carone等人,Chromosoma118,113(Feb,2009))。尽管有这些体外模型,但是鉴于针对注释(annotated)编码序列的微阵列平台的偏倚以及重复序列被标准的分析程序特异性排除,原发性肿瘤中重复的ncRNA的整体表达尚未得到分析。

[0006] 发明概述

[0007] 本发明,至少部分地基于随体重复在肿瘤细胞中大规模表达的鉴定,以及对在包括循环肿瘤细胞(CTC)在内的肿瘤细胞中随体相关基因水平提升的鉴定。本申请描述了基于存在随体相关基因水平升高诊断癌症的方法,所述的癌症例如起源于上皮的实体瘤,例如胰腺癌,肺癌,乳腺癌,前列腺癌,肾癌,卵巢癌或者结肠癌。

[0008] 因此,第一方面,本发明提供检测受试者中存在癌症的体外方法。所述的方法包括确定包含来自受试者的测定细胞的样品中一个或以上随体相关基因的表达水平,以获得测定值,所述随体相关基因选自:HSP90BB(热激蛋白90kDa α (胞质的)),B类成员2,假基因

(HSP90AB2P));NR_003133(人(Homo sapiens)鸟苷酸结合蛋白1,干扰素-诱导的假基因1(GBP1P1),非编码RNA);BX649144(微管蛋白酪氨酸连接酶(TTL));DERP7(跨膜蛋白45A(TMEMP45A));MGC4836(人类似假拟蛋白(L1H3区域));BC037952(cDNA克隆);AK056558(cDNA克隆);NM_001001704(FLJ44796假拟的);ODF2L(精子尾外周致密纤维2-样(ODF2L));BC041426(C12orf55染色体12开放阅读框55(C12orf55));)REX01L1(RNA外切核酸酶1同系物(酿酒酵母(S.cerevisiae))-样1(REX01L1));AK026100(FLJ22447假拟的LOC400221(FLJ22447));AK026825(跨膜蛋白212(TMEMP212));KENAE1(人Kenae1mRNA(AB024691));HESRG(ESRG假拟的LOC790952(ESRG));AK095450(LOC285540假拟的LOC285540);FLJ36492(CCR4-NOT转录复合物,亚基1(CNOT1));AK124194(FLJ42200蛋白质);AK096196(假拟的LOC100129434);AK131313(锌指蛋白91假基因(LOC441666));FLJ11292(假拟蛋白FLJ11292);CCDC122(含122(CCDC122)的卷曲-螺旋结构域);和BC070093(cDNA克隆);以及将所述的测定值与参考值相比较。测定值明显高于参考值表明所述受试者患有癌症。

[0009] 在一些具体实施方案中,所述的参照水平是正常细胞中随体相关基因的水平。在一些具体实施方案中,所述的正常细胞是与相同受试者中的测定细胞相同类型的细胞。在一些具体实施方案中,所述的正常细胞是受试者中与测定细胞相同类型的细胞,所述受试者没有癌症。在一些具体实施方案中,所述的细胞在组织样品中。

[0010] 在一些具体实施方案中,所述的样品是已知的或被怀疑包含肿瘤细胞,例如已知的或被怀疑含有循环肿瘤细胞(CTC)的血液样品,或者已知的或被怀疑含有肿瘤细胞的活检样品。

[0011] 在一些具体实施方案中,所述的方法还包括基于存在测定值明显高于参考值对患有癌症的受试者进行诊断的方法;基于存在测定值明显高于参考值鉴定受试者患有癌症的方法;基于存在测定值明显高于参考值选择受试者进行治疗的方法;基于存在测定值明显高于参考值对受试者进行癌症治疗(例如对受试者施用癌症治疗方案)的方法;或者基于存在测定值明显高于参考值选择进行进一步的诊断试验(例如,成像或活组织检查)的患者的方法。

[0012] 另一方面,本发明提供评估对受试者的癌症治疗效果的体外方法。所述方法包括确定来自受试者的一个或以上随体相关基因的表达水平以获得第一值,所述随体相关基因选自:HSP90BB(热激蛋白90kDa α (胞质的),B类成员2,假基因(HSP90AB2P));NR_003133(人鸟苷酸结合蛋白1,干扰素-诱导的假基因1(GBP1P1),非编码RNA);BX649144(微管蛋白酪氨酸连接酶(TTL));DERP7(跨膜蛋白45A(TMEMP45A));MGC4836(人类似假拟蛋白(L1H3区));BC037952(cDNA克隆);AK056558(cDNA克隆);NM_001001704(FLJ44796假拟的);ODF2L(精子尾外周致密纤维2-样(ODF2L));BC041426(C12orf55染色体12开放阅读框55(C12orf55));)REX01L1(RNA外切核酸酶1同系物(酿酒酵母)-样1(REX01L1));AK026100(FLJ22447假拟的LOC400221(FLJ22447));AK026825(跨膜蛋白212(TMEMP212));KENAE1(人Kenae1mRNA(AB024691));HESRG(ESRG假拟的LOC790952(ESRG));AK095450(LOC285540假拟的LOC285540);FLJ36492(CCR4-NOT转录复合物,亚基1(CNOT1));AK124194(FLJ42200蛋白);AK096196(假拟的LOC100129434);AK131313(锌指蛋白91假基因(LOC441666));FLJ11292(假拟蛋白FLJ11292);CCDC122(含122的卷曲螺旋结构域(CCDC122));和BC070093(cDNA克隆);对受试者实施癌症治疗;确定在其后由受试者获得的后续样品中,所述的一个或以上

(例如相同的)随体相关基因的表达水平;以获得治疗值;和将所述的第一值与所述治疗值相比较。低于所述第一值的治疗值表明所述治疗是有效的(无改变,或下降,说明所述治疗是无效的)。

[0013] 在一些具体实施方案中,所述的第一和第二样品是已知的或被怀疑包含肿瘤细胞,例如已知的或被怀疑包含循环肿瘤细胞(CTC)的血液样品,或已知的或被怀疑含有肿瘤细胞的活检样品。

[0014] 在一些具体实施方案中,所述的治疗包括实施手术干涉,化疗,放疗,或其组合。

[0015] 在所描述的方法的一些具体实施方案中,所述的受试者是人。

[0016] 在所描述的方法的一些具体实施方案中,所述的癌症是起源于上皮的实体瘤,例如胰腺癌,肺癌,乳腺癌,前列腺癌,肾癌,卵巢癌或结肠癌。

[0017] 在所描述的方法的一些具体实施方案中,所述的方法包括确定AK056558; BC037952; HSP90BB; 和/或AK096196中一个或以上,例如两个,三个或四个的水平。在所描述的方法的一些具体实施方案中,所述的方法包括确定HSP90BB和/或AK056558中的一个或两者的水平。在一些具体实施方案中,所述的方法包括确定HSP90BB的水平。在一些具体实施方案中,所述的方法包括确定AK096196的水平。在一些具体实施方案中,所述的方法包括确定AK056558的水平。在一些具体实施方案中,所述的方法包括确定BC037952的水平。

[0018] 在一些具体实施方案中,确定一个或以上随体相关基因的水平包括确定转录物的水平。在一些具体实施方案中,确定转录物的水平包括使与所述转录物特异性结合的寡核苷酸探针与所述样品接触。在一些具体实施方案中,所述的探针是经标记的。

[0019] 在一些具体实施方案中,“确定……的水平”包括检测存在或不存在,例如存在高于所使用的测定的检测限的水平。

[0020] 在一些具体实施方案中,本发明的方法可用于确定受试者患有癌症的可能性。

[0021] 除非另有定义,本申请所用的全部技术和科学术语具有本发明所属领域普通技术人员通常所理解的含义。本申请中描述了用于本发明的材料和方法。其他本领域已知的合适的材料和方法也可以用于本发明。全部的出版物,专利申请,专利,序列,数据条目以及本申请所提及的其他参考文献通过引用全文并入本申请。此外,PCT/US2011/055108特别地通过引用全文并入本申请,在一些具体实施方案中,本申请所描述的方法可以与该申请中所描述的方法一起使用。存在分歧时,包括定义在内以本说明书为准。

[0022] 本发明的其它特征和有益效果在下述的详细描述和附图以及权利要求中显而易见。

[0023] 附图简述

[0024] 图1A是在不同的肿瘤、细胞系和组织中全部基因组比对读长(reads)的主要随体水平百分比。原发肿瘤的基因型和细胞系显示于各肿瘤型和细胞系之下。(Kras = KrasG12D; Tp53, SMAD4, APC代表缺失的基因)

[0025] 图1B是在全部的原发肿瘤,癌细胞系和正常组织当中,主要随体的序列读长分布图示。

[0026] 图2A示出三个KrasG12D, Tp53lox/+胰腺的原发肿瘤(肿瘤1-3)以及肿瘤3衍生的稳定细胞系(CL3)的Northern印迹分析结果。

[0027] 图2B显示CL3在经DNA甲基化不足试剂(hypomethylating agent)5-氮胞苷(AZA)

治疗前(0)和治疗后(+)的Northern印迹分析结果。

[0028] 图2C显示来自多个成年和胎鼠组织的总RNA Northern印迹分析结果。全部的Northern印迹结果。全部的Northern印迹曝光约30分钟。

[0029] 图2D是一对显微照片,显示正常胰腺(左)和原发性胰腺管状腺癌(右),与1kb主要随体重复探针杂交的RNA原位杂交结果。

[0030] 图2E是一组三张显微照片,显示肿瘤发生前的PanIN(P)病变,邻近PDAC(T)和正常胰腺(N)的ISH分析结果,显示出PanIN中染色阳性,满布的癌中增多的表达。PanIN(左)和PDAC(右)病变的更高放大率(40x)。

[0031] 图2F是一组三张显微照片,显示在PDAC细胞转移到肝中随体的标记表达,肝脏本身不表达随体(左)。大的腺性转移瘤沉积物可容易地通过标准的组织学评估和随体染色鉴定(中间)。随体ISH足够灵敏以检测通过标准组织学分析(右,箭头所指)所不易发现的肝实质中的微转移。所有图像20x放大(比例尺=100 μ m)。

[0032] 图3A是柱状图,显示经DGE定量的、在人胰腺管状腺癌(PDAC),正常胰腺,其它癌症(L-肺的,K-肾,O-卵巢,P-前列腺),和其他正常人组织(1-胎脑,2-脑,3-结肠,4-胎肝,5-肝,6-肺,7-肾,8-胎盘,9-前列腺,和10-子宫)中的总随体表达。随体表达以每百万与人基因组比对的转录物的形式表示。

[0033] 图3B是柱状图,在经测序的人PDAC(黑色,n=15)和正常人组织(白色,n=12)中以总随体百分比的形式显示随体重复类别的分类(breakdown)。随体按照在肿瘤中最高绝对差异到在正常组织中的最高绝对差异排列(从左到右)。误差棒代表平均值的标准误差。前三位的癌症(左,黑色柱)的倍数差异(Fold differential)和正常的(右,白色柱)组织随体类别示出(柱图,中间)。

[0034] 图3C显示在带有邻近的非癌性间质组织(N)(图像顶部)的肿瘤发生前PanIN(P)病变,PDAC(T)细针抽吸物活检样品,和正常的邻近白细胞(N)中的人随体HSATII RNA原位杂交(RNA-ISH)。

[0035] 图4A显示如热图所描述的,主要随体相对于全部鼠肿瘤和正常组织当中的其它细胞转录物的多重线性相关分析结果。X-轴是根据主要随体表达排列的样品,y-轴是根据主要随体表达线性相关排列的基因。浅灰(高)和深灰(低)颜色是 \log_2 (读长/百万)。作为基因组比对读长百分比(y-轴)的主要随体表达的等级根据随体读长(x-轴)排列,附有排在前列的基因的扩展视图,所述基因具有随体水平最高线性关系($R \geq 0.85$)。

[0036] 图4B是点状图,显示根据与随体表达线性关系(深灰;左侧为最高线性关系)排列或随机(浅灰)排列的所有基因的转录起始位点到LINE-1元件的距离中间值。

[0037] 图4C是点状图,显示排在前列具有最高线性关系($R > 0.85$)的基因,其被定义为随体相关基因或SCG,用与这些基因的预期发生率(frequency)(浅灰)相比,这些基因相对于转录起始位点到LINE-1元件的距离的发生率(深灰)来描绘。

[0038] 图4D是一组四张显微照片,显示鼠PDAC(KrasG12D,Tp53lox/+)神经内分泌标记嗜铬粒蛋白A免疫组织化学结果。肿瘤被描绘成起到使得嗜铬粒蛋白A染色加重(深灰)的作用,各肿瘤主要随体表达相对水平注释在每一图像底部(全部转录物的百分比)。

[0039] 图5是显微照片,显示在人原发性PDAC FFPE样品上应用SinglePlex ViewRNA显色试验进行RNA-ISH的结果。图示中HSP90BB阳性(原本是红色,导管周围黑色区域,其中的一

些用箭头标示)上皮导管癌细胞(导管被示出)覆盖于所述组织的明视野图像上。

[0040] 发明详述

[0041] 本发明至少部分基于对在人和鼠癌症中大规模产生随体RNA以及对一些随体相关基因的鉴定。因此,本发明的方法对于癌症的早期检测有用,并且可用于预测临床效果。

[0042] 利用随体相关基因作为生物标记诊断癌症

[0043] 本申请所描述的方法可用于诊断受试者中的癌症,例如起源于上皮的实体瘤,如胰腺,肺,乳腺,前列腺,肾,卵巢或结肠的癌症的存在,并监控治疗的效果。

[0044] 本申请所用的术语“高增殖性(hyperproliferative)”指具有自主生长能力的细胞,即以快速增殖的细胞生长为特征的异常状态或条件。高增殖性疾病状态可归类于病理的,即以疾病状态为特征或构成疾病状态,或者也可以归类为非-病理的,即由正常态衍生而来与疾病状态无关。所述术语的意思包括全部的癌性生长或致癌过程,瘤转移组织或恶性转化的细胞,组织或器官,无论何种组织病理类型或侵染阶段。“肿瘤”是高增殖性细胞的非正常生长。“癌症”指病理疾病状态,例如以恶性肿瘤生长为特征。

[0045] 如本申请所展示的,癌症的存在,例如由ICD-0(国际疾病分类-肿瘤学)规范(第3版),章节(8010-8790)所定义的、例如源于上皮的实体瘤,如癌症的早期阶段,与存在高随体水平相关,这是由于胰腺的癌细胞中随体重复的转录和加工增多,以及在循环肿瘤细胞中SCG表达水平的升高。因此,本申请所述的方法可以包括样品中随体重复表达水平的检测,所述样品包含已知或被怀疑是肿瘤细胞的细胞,例如来自源于上皮的实体瘤的细胞,例如胰腺癌,肺癌,乳腺癌,前列腺癌,肾癌,卵巢癌或结肠癌的细胞。作为替代或附加地,所述的方法可以包括应用例如本申请所描述的微观流体装置检测在样品,例如已知或被怀疑包含肿瘤细胞,例如循环肿瘤细胞(CTC),的样品中,升高的SCG水平。

[0046] 源于上皮的癌症可包括胰腺癌(例如胰腺腺癌或导管内乳头状粘液癌(IPMN,胰腺团块(mass)),肺癌(例如非小细胞肺癌),前列腺癌,乳腺癌,肾癌,卵巢癌,或结肠癌。例如本发明的方法可用于区分以监测作为标准治疗方式的良性IPMN和需要实施切除术的恶性IPMN,所述的切除术是与明显的病态相关的手术,具有很小但确实存在的死亡可能。在一些具体实施方案中,诊断为患有IPMN的受试者中,本发明的方法可用于监测/监督所述患者,例如在选定的时间间隔(例如每3,6,12或24个月)重复所述方法确定是否良性的IPMN已经转化为有必要进行手术干涉的恶性IPMN。此外,在一些具体实施方案中,所述方法可用于在怀疑患有非-小细胞肺癌的受试者的样品中,将细支气管肺泡癌与反应性过程(例如肺炎后反应性过程)区分开。在一些具体实施方案中,在被怀疑患有乳腺癌的受试者的样品中,应用本发明的方法使得导管增生与非典型的导管增生和导管原位癌(DCIS)区分开。后两者需要接受切除术/辐射;而前者无需干预。在一些具体实施方案中,在被怀疑患有前列腺癌的受试者中,所述方法可用于区分非典型的小腺泡增殖和恶性的癌症。在一些具体实施方案中,在被怀疑患有膀胱癌的受试者中,所述方法可用于在例如尿样本中,检测例如移行细胞癌(TCC)。在一些具体实施方案中,在被诊断患有Barrett's Esophagus(Sharma, N Engl J Med. 2009, 24; 361(26): 2548-56. Erratum in: N Engl J Med. 2010 Apr 15; 362(15): 1450)的受试者中,所述方法可用于区分Barrett's esophagus中的发育异常与反应性过程。这对临床的影响是显著的,因为发育异常需要治疗性介入。其他的具体实施方案包括但不限于对下述病症的诊断,充分分化的肝细胞癌,壶腹和胆管癌,胶质瘤vs. 反应性胶质增生,黑色素瘤

vs. 真皮痣, 低级肉瘤, 和胰腺的内分泌肿瘤, 以及其他。

[0047] 因此, 本申请包括癌症, 例如受试者中的源于上皮的肿瘤, 如胰腺癌, 肺癌, 乳腺癌, 前列腺癌, 肾癌, 卵巢癌或结肠癌的诊断方法。在一些具体实施方案中, 所述的方法包括, 从受试者获取样品, 评估所述样品中SCG和/或随体的存在和/或水平, 将所述的存在和/或水平与一个或以上的参照相比较, 例如代表正常SCG或随体水平的对照参照, 如来自未受影响的受试者或来自相同个体的正常细胞中的水平, 和/或代表与癌症相关的SCG或随体水平的疾病参照, 例如患有胰腺癌, 肺癌, 乳腺癌, 前列腺癌, 肾癌, 卵巢癌或结肠癌的受试者中的水平相比较。

[0048] 本发明的方法还可以用于确定癌症的阶段, 例如样品是否包含来自癌前期病变, 肿瘤早期, 或肿瘤晚期的细胞。例如, 本发明的方法可用于确定受试者是否具有癌前期胰腺的, 乳腺的, 或前列腺的病变。当所用的标记是SCG转录物或者编码的蛋白质时, 提高的水平与更高级的阶段相关。

[0049] 样品

[0050] 在本发明的方法的一些具体实施方案中, 所述的样品是, 或包括血液, 血清, 和/或血浆, 或其部分或亚级分, 例如血清的游离RNA或血液中的外体内RNA。在一些具体实施方案中, 所述的样品包含(或怀疑包含)CTC。在一些具体实施方案中, 所述的样品是, 或包括尿或其部分或亚级分。在一些具体实施方案中, 所述的样品包括已知的或被怀疑的肿瘤细胞, 例如是活检样品, 例如细针抽吸物(FNA), 内窥镜活检, 或针吸活检。在一些具体实施方案中, 所述的样品包含来自受试者的胰腺, 肺, 乳腺, 前列腺, 肾, 卵巢或结肠的细胞。在一些具体实施方案中, 所述的样品包含来自痰样品的肺细胞或通过下述方式由受试者的肺获得的肺细胞, 即通过刷拭、清洗、支气管镜活检、经支气管活检或FNA, 如支气管镜检, 荧光镜检或CT-向导FNA(此种方法也可以用于获得来自组织的样品)获得。在一些具体实施方案中, 所述的样品是冷冻的, 固定的和/或经渗透处理的, 例如是福尔马林固定石蜡包埋(FFPE)的样品。

[0051] 检测的方法

[0052] 可以使用本领域任何已知的方法来检测和/或定量本申请所述的生物标记的水平。例如可以利用本领域已知的方法评估SCG mRNA(转录物)的水平, 例如Northern印迹, RNA原位杂交(RNA-ISH), RNA表达测定, 例如微阵列分析, RT-PCR, RNA测序(例如利用随机引物或寡T引物), 深度测序(deep sequencing), 克隆, Northern印迹, 和扩增转录物, 例如利用定量实时聚合酶链式反应(qRT-PCR)。确定RNA表达的分析技术是本领域已知的。参见, 例如Sambrook等人, Molecular Cloning: A Laboratory Manual, 3rd Ed., Cold Spring Harbor Press, Cold Spring Harbor, NY(2001)。

[0053] 在一些具体实施方案中, 若所述的SCG是编码转录物(参见表6), 则检测SCG-编码的蛋白质的水平。可利用本领域已知的方法评估蛋白质的存在和/或水平, 例如利用定量免疫分析方法, 如酶联免疫吸附试验(ELISAs), 免疫沉淀, 免疫荧光, 免疫组织化学, 酶免疫测定(EIA), 放射免疫测定(RIA)和Western印迹分析。

[0054] 在一些具体实施方案中, 所述的方法包括使选择性地与生物标记结合的试剂与样品接触, 以评估样品中所述生物标记, 例如SCG转录物/mRNA或蛋白质(例如寡核苷酸探针, 抗体或其抗原结合部分)的水平。在一些具体实施方案中, 所述的试剂带有可检测的标记。

关于经标记的试剂,术语“经标记的”涵盖了通过将所述试剂与可检测的物质相偶联(即物理连接)的直接标记,以及通过所述试剂与可检测的物质的反应性进行间接标记。可检测的物质的示例是本领域已知的,包括化学发光的,荧光的,放射性的或显色的标记。例如可检测的物质包括各种酶,辅基,荧光材料,发光材料,生物发光材料和放射性材料。合适的酶的例子包括辣根过氧化物酶,碱性磷酸酶, β -半乳糖苷酶,或乙酰胆碱酯酶;合适的辅基复合物的例子包括链霉抗生物素蛋白/生物素,以及亲和素/生物素;合适的荧光物质的例子包括伞形酮,荧光素,异硫氰酸荧光素,罗丹明),二氯三嗪氨荧光素(dichlorotriazinylamine fluorescein),丹磺酰氯,量子点(quantum dots)或藻红蛋白;发光材料的示例包括鲁米诺;生物发光材料的示例包括荧光素酶,荧光素和水母发光蛋白;合适的放射性物质的示例包括 ^{125}I , ^{131}I , ^{35}S or ^3H 。总体而言,可利用抗体检测蛋白质。抗体可以是多克隆抗体,更优选单克隆抗体。可应用完整抗体或其抗原结合片段(例如Fab或F(ab')₂)。

[0055] 在一些具体实施方案中,可利用高通量方法,例如本领域已知的蛋白质或基因芯片(参见,例如Ch.12,“Genomics,”in Griffiths等人,Eds.Modern genetic Analysis,1999,W.H.Freeman and Company;Ekins and Chu,Trends in Biotechnology,1999;17:217-218;MacBeath and Schreiber,Science2000,289(5485):1760-1763;Simpson,Proteins and Proteomics:A Laboratory Manual,Cold Spring Harbor Laboratory Press;2002;Hardiman,Microarrays Methods and Applications:Nuts&Bolts,DNA Press,2003),来检测随体或SCG的存在和/或水平。

[0056] 在一些具体实施方案中,所述的方法包括利用经修饰的RNA原位杂交技术应用支链DNA测定来直接检测并评估样品中生物标记mRNA的水平(参见,例如Luo等人,US Pat.No.7,803,541B2,2010;Canales等人,Nature Biotechnology24(9):1115-1122(2006);Nguyen等人,Single Molecule in situ Detection and Direct Quantification of miRNA in Cells and FFPE Tissues,poster available at panomics.com/index.php?id=product_87)。用于实施这一测定的试剂盒可通过商业途径由Affymetrix(ViewRNA)购置。

[0057] CTC中SCG转录物的检测

[0058] 在一些具体实施方案中,微管流体(例如“芯片实验室(lab-on-a-chip)”)设备可用于本发明的方法。此种设备已经成功地用于微管流体流式细胞术,连续的基于大小的分离,以及层析分离。总体而言,在循环肿瘤细胞(CTC)中检测SCG转录物表达的方法可用于癌症的早期检测,例如源于上皮的肿瘤的早期检测,如胰腺癌,肺癌,乳腺癌,前列腺癌,肾癌,卵巢癌或者结肠癌。

[0059] 所述的设备可用于从细胞混合物中分离CTC,或制备经富集的CTC群。具体而言,此类设备可用于从如全血之类的复合混合物分离CTC。

[0060] 多种方法可用于从异源样品中分离CTC。例如设备可包含在封闭屏障上游微流体通道内的、以六边形组装方式排布多柱阵列。所述的柱和封闭屏障可被设计成具有不同的结合组成成分。例如,所述的柱可被设计成带有抗-EPCAM抗体以捕获循环肿瘤细胞(CTC);参见,例如Nagrath等人,Nature450:1235-1239(2007),任选地,所述设备还具有被设计成用以捕获SCG核酸或蛋白质或随体的下游封闭屏障,参见例如(13-15),所述的申请和参考

文献列于本申请。

[0061] 从样品中富集特定颗粒的处理过程通常基于依次进行的处理步骤,其中的每一步均使得混合物中不需要的细胞/颗粒减少,但是在有些具体实施方案中,一个处理步骤已经足够。实施各种处理步骤的设备可以是相互分离的,或者是整合到一个微流体系统中。所述的设备包括用于细胞/颗粒结合的设备,用于细胞裂解的设备,使细胞形成阵列的设备,和例如基于大小、形状,和/或变形性,或是其他标准进行颗粒分离的设备。在一些特定的具体实施方案中,处理步骤用于在将细胞引入到设备或系统之前减少细胞的数量。在一些具体实施方案中,与起始样品混合物相比,所述的设备保留至少75%,例如80%,90%,95%,98%,或99%所需的细胞,而相对于一个或以上非所需细胞类型,将所需细胞群富集至少100倍,例如1000,10,000,100,000,甚或1,000,000倍。

[0062] 一些颗粒分离设备依赖于基于大小的分离,带有或没有同步进行的细胞结合。一些基于大小的分离设备包括一个或以上的障碍物,其使得CTC和流体的其他组分向侧面移动,由此提供富集机制或另外处理此种组分。所述的根据大小分离颗粒的障碍物阵列通常定义一种缺口网络,其中流体通过缺口被不均衡地分配到接下来的缺口中。滤筛和阵列形式的、基于大小的分离设备均可包括之前有关细胞结合设备所描述的、选择性的可渗透障碍物。

[0063] 包括形成缺口网络的障碍物阵列的设备可包括,例如相交错的二维障碍物阵列,如使得每一连续的行在不到一半的时间段内偏离于之前的行。所述的障碍物也可以以不同的方式排布。可能的障碍物形状和方式在W02004/029221中有更详细的描述。

[0064] 在一些具体实施方案中,所述的设备可利用基于阵列的大小分离方法,如U.S.Pat.Pub.No.2007/0026413中所描述的,提供CTC的分离和/或富集。总体而言,所述的设备提供一个或以上使得诸如CTC之类的大颗粒以及悬浮于流体样品中的其他组分向侧面移动的选择性可渗透障碍物阵列,由此提供富集机制或另外处理此种组分,同时也提供与进入到构成障碍物的纳米管稠密基质中的空隙里的、其他更小颗粒选择性结合的可能。应用此种选择性可渗透障碍物,基于大小、形状,和/或变形性富集颗粒的设备包括过滤器和滤筛,以及富集或分离设备,在国际专利公开号Nos.2004/029221和2004/113877,Huang等人Science304:987-990(2004),U.S.Publication No.2004/0144651,U.S.Pat.Nos.5,837,115和6,692,952,以及U.S.专利申请号60/703,833,60/704,067,和11/227,904中有描述;对亲和捕获有用的设备,例如那些在国际专利公开号2004/029221和U.S.专利申请Ser.No.11/071,679中所公开的;对样品中细胞优先裂解有用的设备,例如那些在国际专利公开号2004/029221,U.S.Pat.No.5,641,628,和U.S.专利申请号60/668,415中所描述的;对使细胞形成阵列有用的设备,例如那些在国际专利公开号2004/029221,U.S.Pat.No.6,692,952,和U.S.专利申请Ser.Nos.10/778,831和11/146,581所描述的;对流体递送有用的设备,例如那些在U.S.专利申请号11/071,270和11/227,469中所描述的。两个或以上的设备可以串连结合在一起,例如在国际专利公开号W02004/029221所描述的。以上文献均作为参考并入本申请。

[0065] 在一些具体实施方案中,设备可包含障碍物,所述障碍物包括结合组成成分,例如单克隆抗-EpCAM抗体或其片段,其选择性地与特定的细胞类型结合,例如源于上皮的细胞,例如肿瘤细胞。所述设备的所有障碍物均可包含这些结合组成成分;可供替代地,仅所述障

碍物的亚组包括它们。设备还可以包括附加的模块,例如细胞计数模块或检测模块,它们在与微观流体通道设备的通讯中。例如所述的检测模块可设置用于使所述设备的输出样品形象化。

[0066] 在一个示例中,检测模块在与分离或富集设备的通讯中。所述的检测模块可利用任何本申请所公开的检测方法,或其他本领域已知的方法操作。例如所述的检测模块包括显微镜,细胞计数器,磁体,生物腔激光(biocavity laser)(参见,例如Gourley等人, J.Phys.D:Appl.Phys.,36:R228-R239(2003)),质谱仪,PCR设备,RT-PCR设备,微阵列,进行RNA原位杂交的设备,或者超光谱(hyperspectral)成像系统(参见,例如Vo-Dinh等人, IEEE Eng.Med.Biol.Mag.,23:40-49(2004))。在一些具体实施方案中,可将计算机终端与检测模块相连接。例如检测模块可以检测与所关注的细胞、蛋白质或核酸选择性结合的标记物,例如SCG转录物或所编码的蛋白质。

[0067] 在一些具体实施方案中,所述的微流体系统包括(i)用于分离或富集CTC的设备;(ii)用于裂解所富集的CTC的设备;和(iii)检测SCG转录物或所编码的蛋白质的设备。

[0068] 在一些具体实施方案中,用本申请所描述的微观流体设备制备的CTC群体用于通过例如在上述文献和Sambrook,Molecular Cloning:A Laboratory Manual,Third Edition (Cold Spring Harbor Laboratory Press;3rd edition(January 15,2001));以及Short Protocols in Molecular Biology,Ausubel等人,eds.(Current Protocols;52edition (November5,2002))所描述的已知的分子生物学技术,分析SCG转录物或蛋白质的表达。

[0069] 总体而言,在本申请中描述了用于检测和/或定量在所富集的CTC群体中SCG转录物或编码的蛋白质的表达的设备,所述设备可以用于癌症的早期检测,例如源于上皮的肿瘤,例如胰腺癌,肺癌,乳腺癌,前列腺癌,肾癌,卵巢癌或者结肠癌的早期检测。

[0070] 监控疾病发展或治疗效果的方法

[0071] 在一些具体实施方案中,一旦确定了某人患有癌症,或具有提高的患上癌症的风险,则可以实施例如本领域已知的那些治疗。可以利用本发明的方法监控治疗的效果;在治疗之后(或治疗期间)对附加的样品进行评估,例如在施用了一个或以上的治疗剂量之后,SCG转录物或编码的蛋白质的表达水平或是表达SCG转录物或蛋白质的细胞数量下降,将指示所述治疗有效;没有改变,或表达SCG转录物或蛋白质的细胞水平提高将指示所述治疗无效。所述方法可以在治疗过程中和/或治疗结束后的多个时间重复,例如监控疾病复发的可能性。

[0072] 在一些具体实施方案中,例如对于已被诊断为处于为可能导致癌症的良性状态的受试者,已经对其癌症进行了成功治疗的受试者,或者具有增加的癌症风险的受试者,例如由于遗传倾向或暴露于致癌试剂的环境的受试者可在选定的时间间隔,例如间隔3,6,12或24个月应用本发明的方法监控受试者中的疾病,对发展为恶性肿瘤或患上癌症进行早期检测。

实施例

[0073] 在下述的实施例中对本发明进行进一步的描述,所述实施例不限制权利要求中所描述的本发明的范围。

[0074] 实施例1.与细胞系和正常组织相比在肿瘤组织中主要随体水平大幅度升高

[0075] 应用Helicos BioSciences的新一代数字基因表达(DGE)(D.Lipson等人,Nat Biotechnol27,652(Jul,2009))来比较原发性癌症和它们所衍生的瘤转移前体中的肿瘤标记表达。我们先确定通过活化Kras靶向组织的表达和Tp53(Kras^{G12D},Tp53^{lox/+})缺失(N.Bardeesy等人,Proc Natl Acad Sci U S A103,5947(Apr11,2006))生成的原发性鼠胰腺管状腺癌(PDAC)的DGE谱。这些肿瘤是人PDAC组织病理的模拟物,其显示几乎普遍存在的突变体KRAS(>90%的情况)以及TP53丧失(50-60%)。

[0076] 在Bardeesy实验室,如之前描述的饲养患有不同基因型的胰腺癌的鼠(Bardeesy等人,Proc Natl Acad Sci U S A103,5947(2006))。正常的野生型鼠购自Jackson实验室。根据动物实验指南对所述动物实施安乐死。无菌提取胰腺肿瘤和正常组织,然后用液氮急速冷冻。将组织保存于-80°C。如之前描述的新鲜生成动物AH367和AH368的细胞系(Aguirre等人,Genes Dev17,3112(2003)),并将所建立的细胞系在RPMI-1640+10%FBS+1%Pen/Strep(Gibco/Invitrogen)中培养。来自结肠和肺的另外的鼠肿瘤由Kevin Haigis(Massachusetts General Hospital)和Kwok-Kin Wong(Dana Farber Cancer Institute)惠赠。

[0077] 在放置在干冰上的微离心管中,用无菌杵将新鲜冷冻的组织磨成粉。在提取核酸前,培养细胞系并在液氮中新鲜冷冻。来自细胞系的RNA和DNA,以及新鲜冷冻的肿瘤和正常组织以相同的方式进行处理。用TRIzol[®] Reagent(Invitrogen)依照制造商所附的说明书提取RNA。利用QIAamp Mini Kit依照制造商所附的说明书从组织和细胞系中提取DNA(QIAGEN)。

[0078] 经纯化的RNA用于准备数字基因表达(DGE)样品并在Helicos BioSciences的HeliScope[™]单模块测序仪上进行分析。所述方法之前已有描述(Lipson等人,Nat Biotechnol27,652(2009))。简要地说,应用dTU25V引物和Superscript III cDNA合成试剂盒(Invitrogen)由RNA反转录单链cDNA。消化RNA,由固相翻转固定(SPRI)技术,利用Agencourt[®] AMPure[®]磁珠纯化单链cDNA。变性单链cDNA,利用末端转移酶(New England Biolabs)在3'末端加上poly-A尾。

[0079] 经纯化的DNA用于如前所述的Helicos DNA测序样品准备方案(Pushkarev, N.F.Neff, S.R.Quake, Nat Biotech27,847(2009))。简要地说,用Covaris S2自然超声发生器(acoustic sonicator)剪切基因组DNA,产生平均200bps的、在100-500bps范围的片段。剪切的DNA用SPRI清洁。使DNA变性并利用末端转移酶在3'末端加上poly-A尾。

[0080] 而后,加尾的cDNA或DNA与测序流动池(sequencing flow cell)杂交,接下来进行“填充并锁定(Fill and Lock)”以及单分子测序。接下来,利用DGE程序将基因表达序列读长与已知的人或鼠的转录组文库比对。将基因组DNA序列读长与鼠基因组比对并计数,以确定主要鼠随体(CNV)的拷贝数。

[0081] 第一个被分析的鼠胰腺肿瘤,AH284,引人注目,这是由于DGE序列与注释的鼠转录组展示出48-52%的不一致,相比之下,来自相同的鼠的正常肝脏转录物仅呈现3-4%的不同。几乎所有不一致的序列作图定位(mapped to)于两侧的(主要的)鼠随体重复。所述的随体转录物占肿瘤中全部细胞转录物的~49%(495,421tpm),相比之下在正常胰腺或肝脏中占0.02-0.4%(196-4,115tpm)(表1)。

[0082] 表1

[0083] 总基因组比对读长,主要随体分解读长和转录组读长。括号中是总基因组比对读长百分比

[0084]

	总计读长	主要随体读长	转录组
胰腺肿瘤	18,063,363	8,460,135(47%)	1,726,768(10%)
正常肝脏	2,270,669	8,973(0.4%)	1,718,489(75%)
正常胰腺	492,301	2,026(0.4%)	63,160(13%)

[0085] 随体序列读长在正义和反义方向均有发现,在poly-A纯化的RNA中没有发现。肿瘤AH284因此包含大量的非-多腺苷酸化的dsRNA元件,定量地确定为,高于来自相同动物的正常组织中存在量>100-倍的升高。通过比较,肿瘤组织中随体转录物水平比高丰度mRNA Gapdh高出约8,000-倍。来自相同鼠的第二个独立胰腺肿瘤结节表现出较低的,但也是大幅升高的随体转录物水平(总细胞转录物的4.5%)。

[0086] 对来自($Kras^{G12D}$, $Tp53^{1ox/+}$)鼠和具有可供替代的胰腺致肿瘤基因型($Kras^{G12D}$, $SMAD4^{1ox/1ox}$)的4只鼠的另外4个胰腺肿瘤的分析揭示出,在6/8另外的肿瘤(全部细胞转录物的1-15%的范围)中随体表达升高。2/3鼠结肠癌肿瘤($Kras^{G12D}$, $APC^{1ox/+}$)和2/2肺癌($Kras^{G12D}$, $Tp53^{1ox/1ox}$)中,随体表达水平占全部细胞转录物的2-16%。总计,与正常鼠组织(图1A,表2)相比,12/15(80%)独立的鼠肿瘤具有显著升高的随体表达水平。

[0087] 表2

[0088] 总基因组读长以及并列出的转录组和多个鼠肿瘤,细胞系和正常组织中的主要随体读长百分比

[0089]

鼠ID	组织类型	基因型	总计基因组读长	%转录组读长	%主要随体读长
AH284Rep1	胰腺癌	<i>Kras</i> ^{G12D} , <i>Tp53</i> ^{lox/+}	18,063,363	9.56%	46.84%
AH284Rep2	胰腺癌	<i>Kras</i> ^{G12D} , <i>Tp53</i> ^{lox/+}	16,948,693	10.15%	49.54%
AH284-2*	胰腺癌	<i>Kras</i> ^{G12D} , <i>Tp53</i> ^{lox/+}	1,613,592	48.67%	4.78%
AH287	胰腺癌	<i>Kras</i> ^{G12D} , <i>Tp53</i> ^{lox/+}	2,227,850	54.70%	0.07%
AH288	胰腺癌	<i>Kras</i> ^{G12D} , <i>Tp53</i> ^{lox/Δ}	6,780,821	26.57%	14.79%
AH291	胰腺癌	<i>Kras</i> ^{G12D} , <i>Tp53</i> ^{lox/+}	1,388,906	43.12%	1.22%
AH294	胰腺癌	<i>Kras</i> ^{G12D} , <i>Tp53</i> ^{lox/+}	969,896	37.20%	3.73%
AH323	胰腺癌	<i>Kras</i> ^{G12D} , <i>SMAD4</i> ^{lox/lox}	1,887,663	72.73%	0.29%
AH346	胰腺癌	<i>Kras</i> ^{G12D} , <i>SMAD4</i> ^{lox/lox}	1,291,648	32.92%	6.07%
AH347	胰腺癌	<i>Kras</i> ^{G12D} , <i>SMAD4</i> ^{lox/lox}	1,634,314	38.94%	8.59%
AH348	胰腺癌	<i>Kras</i> ^{G12D} , <i>SMAD4</i> ^{lox/lox}	2,030,197	45.84%	5.61%
结肠的1	结肠癌-1	<i>Kras</i> ^{G12D} , <i>APC</i> ^{lox/lox}	2,954,930	77.49%	0.07%
结肠的1	结肠癌-2	<i>Kras</i> ^{G12D} , <i>APC</i> ^{lox/lox}	985,510	53.13%	6.27%
结肠的1	结肠癌-3	<i>Kras</i> ^{G12D} , <i>APC</i> ^{lox/lox}	1,017,319	30.71%	16.02%
KN2128	肺癌	<i>Kras</i> ^{G12D} , <i>Tp53</i> ^{lox/lox}	2,233,183	60.78%	2.66%
KN2199	肺癌	<i>Kras</i> ^{G12D} , <i>Tp53</i> ^{lox/lox}	1,653,948	43.21%	5.37%

[0090]

AH323	PDAC 细胞系	<i>Kras</i> ^{G12D} , <i>SMAD4</i> ^{lox/lox}	1,958,108	83.13%	0.02%
AH324	PDAC 细胞系	<i>Kras</i> ^{G12D} , <i>Tp53</i> ^{lox/+}	3,301,108	86.32%	0.04%
NB490	PDAC 细胞系	<i>Kras</i> ^{G12D} , <i>Tp53</i> ^{lox/lox}	15,378,82	76.85%	0.03%
AH284Rep1	匹配的 正常肝脏	<i>Kras</i> ^{G12D} , <i>Tp53</i> ^{lox/+}	2,270,669	75.68%	0.40%
AH284Rep2	匹配的 正常肝脏	<i>Kras</i> ^{G12D} , <i>Tp53</i> ^{lox/+}	1,627,749	56.59%	0.34%
AH284-2*	匹配的 正常肝脏	<i>Kras</i> ^{G12D} , <i>Tp53</i> ^{lox/+}	644,316	41.10%	0.31%
结肠的1	匹配的 正常肝脏	<i>Kras</i> ^{G12D} , <i>APC</i> ^{lox/lox}	1,536,346	86.53%	0.02%
正常1	正常胰腺	WT	247,582	14.49%	0.41%
正常2	正常胰腺	WT	244,719	11.15%	0.41%

[0091] *AH284-2是来自胰腺肿瘤和肝脏不同部分的RNA提取物。

[0092] 可以注意到,在编码的、核糖体的和其他非编码转录物中全部RNA读长的综合分

布,在原发肿瘤和正常组织之间显示出显著的变化(图1A),其表明整体细胞转录机制受原发肿瘤中随体大规模表达的影响。由3个原发性胰腺肿瘤建立的无限增殖化的细胞系显示出极小的随体重复表达,表明体外增殖中的负选择压力或者是体外培养条件下稳定随体沉默机制重建(图1A)。在随体过表达的原发肿瘤中,在编码的,核糖体的和其他非编码转录物中全部RNA读长的综合分布,与正常组织中的分布情况表现出显著的差异(图1B),其表明细胞转录机制受到这些肿瘤中随体转录物大规模表达的影响。

[0093] 实施例2.主要随体转录物在不同的组织类型中表现为多种大小,表达水平与基因组甲基化和扩增相关

[0094] 依照下述进行鼠原发胰腺肿瘤Northern印迹分析。利用Northern Max-Gly Kit (Ambion)进行Northern印迹。总RNA(10ug)与等体积的Glyoxal Load Dye(Ambion)混合,在50°C下温育30min。经1%琼脂糖凝胶电泳后,将RNA转移到BrightStar-Plus膜(Ambion)上,并用紫外光进行交联。所述的膜在ULTRAhyb缓冲液中(Ambion)于68°C下预杂交30min。利用MAXIscript Kit(Ambion)制备鼠RNA探针(1100bp),用BrightStar Psoralen-Biotin Kit (Ambion)根据制造商附带的说明书进行非同位素标记。利用0.1nM探针,使所述的膜在ULTRAhyb缓冲液(Ambion)中于68°C下杂交2小时。用低强度洗涤液于室温洗膜10min,用高强度洗涤液于68°C下洗膜15min,两次。对于非同位素化学发光检测,根据制造商附带的说明书利用BrightStar BioDetect Kit进行。

[0095] 结果说明,主要随体-衍生的转录物在100bp到2.5kb之间的范围(图2A)与预计的、包含多个串连重复的大初级转录物被切酶1切割的一致(C.Kanellopoulou等人,Genes Dev19,489(Feb15,2005);T.Fukagawa等人,Nat Cell Biol6,784(Aug,2004);H.Bouzinba-Segard,A.Guais,C.Francastel,Proc Natl Acad Sci U S A103,8709(Jun6,2006)),其在表达随体的鼠胰腺肿瘤中的表达高出平均值2.6-倍($p=0.0006$,t-试验)。由高表达随体的原发肿瘤衍生的、经建立的胰腺癌细胞系具有很少的随体表达确认了我们的测序结果(T3和CL3;图2A)。用5-氮杂胞苷治疗CL3导致大规模的随体转录物再表达,支持DNA甲基化作为体外随体沉默的机制(图2B)。大多数的正常成年鼠组织,除了肺之外,显示出极少的随体重复表达(图2B)。但是在胚胎组织中明显有非切割的5kb随体转录物的表达(图2C)。因此,原发胰腺肿瘤中的随体重复异常表达并非简单地概括胚胎细胞的命运,也反映出改变的初级5kb随体转录物的加工。单个分子测序平台对小的重复ncRNA片段的定量特别敏感,每个ncRNA片段被给出独特的读长得分。鼠主要随体的高水平表达在原发肿瘤内的所有细胞中都很明显(图2D),如RNA原位杂交(ISH)所示。引人注目的是在早期的肿瘤发生前病变、胰腺上皮内瘤形成(PanIN)中,表达已经升高,并且在转换为全面的胰腺腺癌时,进一步升高(图2E)。很明确地,到肝部(liver ware)的瘤转移病变在RNA ISH中表现为强阳性,而肝实质中的单个PDAC细胞,其在其他的组织病理分析中未曾被检测到(图2F)。在整铺片(whole mount)胚胎分析中所示,在肝和肺中明显低水平扩散表达,但是与在肿瘤组织中明显表达相比,在正常的成年或胚胎组织没有显示随体表达。

[0096] 为了确定是否随体重复的基因组扩增也是造成鼠胰腺肿瘤中额外丰富的这些转录物的原因,如以上对基因组DNA测序的描述,用新一代DNA数字拷贝数变异(CNV)分析对指标性的AH284肿瘤进行了分析。

[0097] 表3所示的结果显示随体DNA占该肿瘤中全部基因组比对读长的18.8%,相比之

下,匹配的正常肝脏中只占基因组序列的2.3%。已经估算过,所述的主要随体重复占正常鼠基因组的大约3%(J.H.Martens等人,EMBO J24,800(Feb23,2005))。因此,在随体重复的表达升高>100-倍的这些肿瘤中,所述重复的大约8-倍的基因扩增可能导致它们的异常表达。

[0098] 表3

[0099] 来自鼠AH284指标性胰腺肿瘤和正常肝脏的CNV分析。主要随体读长占全部基因组比对读长(最后一栏)的百分比。

[0100]

	主要随体读长	总计基因组读长
AH284 肝	183,327 (2.3%)	7,995,538
AH284 PDAC	2,283,436 (18.8%)	12,124,201

[0101] 实施例3. 随体转录物在人胰腺癌和其它上皮癌中的过表达

[0102] 为了测定是否人肿瘤也过表达随体ncRNAs,我们将所述的DGE分析扩展到人胰腺癌样本。人胰腺肿瘤组织是根据IRB议定书由Massachusetts综合医院作为过量丢弃的人材料获得。在提取核酸之前,全部的(Gross)肿瘤被切除并新鲜冷冻在液氮中。正常胰腺RNA由两个提供商,Clontech和Ambion获得。所述样品的制备和分析与上述实施例1中描述的相同。

[0103] 对15个PDAC的分析表明,与正常胰腺相比,随体转录物表达平均升高21-倍。非小细胞肺癌,肾细胞癌,卵巢癌,和前列腺癌之类也具有随体和HSATII随体水平的显著提高。其它的正常人组织,包括胎脑,脑,结肠,胎肝,肝,肺,肾,胎盘,前列腺和子宫具有稍高的总随体表达水平(表4,图3A)。

[0104] 表4

[0105]

样品ID	基因组	总随体(tpm)	ALR(tpm)	HSATII(tpm)
PDAC1	4,472,810	25,209	14,688	3,589
PDAC2	1,668,281	22,001	12,653	3,295
PDAC3	5,211,399	27,366	15,921	5,057
PDAC4	1,649,041	23,556	13,428	3,167
PDAC5	239,483	15,095	8,259	509
PDAC6	1,520,470	374	195	14
PDAC7	1,449,321	7,738	4,400	750
PDAC8	1,950,197	574	316	9
PDAC9	3,853,773	19,572	12,563	1,731
PDAC10	2,748,850	28,225	18,767	2,489
PDAC11	2,848,599	23,163	14,634	2,589
PDAC12	3,723,326	21,243	12,940	2,122
PDAC13	1,834,743	24,549	15,342	3,150

PDAC14	2,481,332	25,650	18,016	2,564
PDAC15	1,752,081	38,514	25,899	5,210
正常胰腺1	1,196,372	908	284	0
正常胰腺2	975,676	1,043	303	0
肺癌1	1,549,237	28,658	18,751	4,417
肺癌2	13,829,845	33,030	26,143	2,555
肾癌1	2,104,859	10,814	6,505	1,501
肾癌2	4,753,409	5,025	2,739	625
卵巢癌1	12,596,542	26,658	14,513	3,074
卵巢癌2	7,290,000	4,089	2,058	403
前列腺癌1	3,376,849	43,730	22,244	9,793
前列腺癌2	12,052,244	23,947	14,201	3,209
前列腺癌3	3,631,148	21,411	12,390	2,804
正常胎脑	384,453	2,843	1,516	3
正常脑	371,161	5,184	2,573	3
正常结肠	183,855	13,059	7,229	5
正常胎肝	147,977	11,218	5,879	7
正常肝脏	117,976	7,968	3,730	25
正常肺	208,089	15,027	7,857	5
正常肾	144,173	15,218	8,094	7
正常胎盘	207,929	13,990	7,815	0
正常前列腺	263,406	8,409	2,228	19
正常子宫	477,480	2,702	1,395	2

[0106] 在多个类别当中,人随体的细分揭示出肿瘤和全部正常组织的主要差异。鼠随体大体上进一步分为主要的和次要的随体,而人随体进行了更深入地分类。在全部的人随体中,周围着丝粒随体HSATII(平均2,416tpm;10.3%of随体reads)明显具有最高的表达倍数差异,其在正常人胰腺中未检测到(图3B)。相比之下,正常组织呈现更高的GSATII, β 随体(BSR),和TAR1类别(分别占全部随体读长的21.1%,17.3%,和2.1%),而这些类别在胰腺癌中构成随体读长的一小部分。

[0107] 正常表达的人随体的最高丰度类别, α (ALR)(Okada等人,Ce11131,1287(Dec28,2007))在正常人胰腺中以294tpm表达,但是在人胰腺癌中平均包含12,535tpm(43-倍的表达差异;随体读长的60.3%)。因而,人ALR重复的过表达堪比鼠主要随体重复的情况,次要丰度的HSATII(高于GAPDH49-倍),对人PDAC表现出显著的特异性。在人胰腺肿瘤中LINE-1与随体转录物的共表达也很突出,平均16,089tpm(358-38,419的范围)。

[0108] ALR重复以外,正常胰腺和PDAC的随体表达谱显著不同;例如正常胰腺组织呈现相当高的GSATII,TAR1和SST1类别(占全部随体读长的26.4%,10.6%,和8.6%),而它们仅占胰腺癌中随体读长的一小部分。相比之下,癌症中有高水平的HSATII随体表达(4,000/106转录物;随体读长的15%),这一亚型在正常胰腺中未检测到(图3B)。鼠相对于人胰腺癌中的随体转录的定量比较显示,鼠主要随体表达平均高于高丰度Gapdh mRNA466-倍,而人ALR

和HSATII随体表达分别高于GAPDH180-倍和47-倍。

[0109] 实施例4. 为干细胞和神经元件其与组蛋白去甲基化酶和RNA加工酶相关联的富集与升高的随体水平线性相关的细胞转录物

[0110] 不同组织学和遗传背景的25个不同鼠组织的综合DGE谱的生成,使得在宽泛的定量范围将细胞转录物与随体转录物的表达相关联成为可能。为鉴定这种共调节基因,由DGE定量的全部注释转录物进行线性回归分析,将与随体表达具有最高相关系数的转录物排列次序。

[0111] 全部的鼠样品读长与订制的鼠主要随体文库(序列来自于UCSC基因组浏览器)相比对。人样品与订制的全部随体重复参考文库相比对,LINE-1变体由Repbase文库(Pushkarev等人,Nat Biotech27,847(2009))生成。此外,全部的样品由DGE程序进行转录组分析。将读长规范化为每106全部样品的基因组比对读长。

[0112] 根据主要随体水平将鼠主要随体与转录组,全部组织和细胞系的线性相关性排列次序。将全部注释的基因在全部的组织范围进行线性回归分析。根据线性回归的Pearson系数将基因排列次序,并用Matlab绘图。

[0113] 对具有最高线性相关性($R>0.85$)的一组297个基因的分析,揭示出190注释的细胞mRNAs和转座元件的亚组(图4A)。

[0114] 细胞mRNA的亚组显示出,其与多种鼠肿瘤中随体重复表达水平高度相关(本申请中指“随体相关基因(SCGs)”)。具有 $R>0.85$ 线性相关的基因用DAVID程序(Dennis,Jr.等人,Genome Biol4,P3(2003);Huang等人,Nat Protoc4,44(2009))作图。用功能注释簇(Functional Annotation clustering)程序和UP_TISSUE数据库对这些基因进行分析,以便对作图基因中的每一个进行分类。包括胚原基/干细胞基因在内的基因在睾丸,卵,滋养层和神经干细胞中高表达。包括神经基因在内的基因在脑,脊髓和包括嗅觉,听觉和视觉在内的专门的感觉神经中高表达。HOX和锌指蛋白应用INTERPRO数据库归类。

[0115] 利用DAVID基因本体论程序对190个注释转录物的分析鉴定出这些转录物中的120个(63%)与神经细胞的命运相关,50个(26%)与胚原基/肝细胞途径相关(表5)。

[0116] 表5

[0117]

	胚原基/干细胞	神经	HOX区域	锌指结构域
总计	50	120	10	16
%作图的(190)	26%	63%	5%	8%

[0118] 此外,转录调节蛋白明显富集,包括HOX相关的(9,5%)和锌指蛋白(16,8%)。这些基因小组不能与GSEA数据库(Subramanian等人,Proc Natl Acad Sci U S A102,15545(Oct25,2005))中任何已知的特征相匹配,但是本体论分析指向神经内分泌表型。神经内分泌分化已在多种上皮恶性肿瘤,包括胰腺癌(Tezel等人,Cancer89,2230(Dec1,2000))中有描述,并在前列腺癌中进行了最好的表征,其与更具侵害性的疾病相关(Cindolo等人,Urol Int79,287(2007))。观察到特征性神经内分泌标记嗜铬粒蛋白A染色的癌细胞数量显著升高,在鼠PDAC中更高随体表达的函数(图4D),支持整体改变的ncRNA表达和特定细胞分化程序之间的关联。

[0119] 利用ALR在人胰腺癌和正常组织中进行平行分析,最高丰度人随体,收获总计539

个SCG。其中206个可通过DAVID基因本体论程序作图,具有类似的胚原基/干细胞富集和神经细胞命运(表6)。将这些观察结合起来表明,如在鼠遗传模型中,随体-衍生的重复的肿瘤相关抑制去除与细胞mRNA亚类的表达升高高度相关。

[0120] 表6

[0121]

	胚原基/干细胞	神经	锌指结构域
总数	101	63	35
%作图的(206)	49.0%	30.6%	17.0%

[0122] 进一步通过采取人SCG在癌症和正常组织之间最低20倍的差异以及每百万500读长的最低表达,对作为生物标记用途的SCG基因列表进行提炼。结果示于表7。

[0123]

人基因名	读长/百万		比例	名称	GenBank 登录号.	
	癌症	正常			mRNA	蛋白质
HSP90BB	5589.1	123.6	45.2	热激蛋白90kDa α (胞质的), B类成员 2, 假基因 (HSP90AB2P)	NR_003132.1	
NR_003133	30208.2	357.1	84.6	人鸟苷酸结合蛋白 1, 干扰素-诱导的假 基因1(GBP1P1), 非 编码RNA	NR_003133.2	
BX649144	12117.1	153.4	79.0	微管蛋白酪氨酸连 接酶(TTL)	NM_153712.4	NP_714923.1
DERP7	7428.7	101.5	73.2	跨膜蛋白 45A(TMEM45A)	NM_018004.1	NP_060474.1
MGC4836	5461.5	38.3	142.7	人类似假拟蛋白	BC036758.2	AAH36758.1

[0124]

				(LIH 3 区域)		
BC037952	2188.6	36.4	60.1	cDNA克隆	BC037952.1	
AK056558	1960.9	18.6	105.2	cDNA克隆	AK056558.1	
NM_001001704	1703.9	68.4	24.9	FLJ44796 假拟的	NM_001001704.1	NP_001001704.1
ODF2L	1213.9	22.0	55.2	精子尾外周致密纤维2-样(ODF2L)	NM_020729.2 NM_001007022.2 NM_001184765.1 NM_001184766.1	NP_065780.2 NP_001007023.2 NP_001171694.1 NP_001171695.1
BC041426	1020.8	43.0	23.7	C12orf55 染色体12 开放阅读框	XM_001715090.3	XP_001715142.3
REXO1L1	1013.4	9.9	102.2	RNA 外切核酸酶1 同系物 (S. cerevisiae)-样 1 (REXO1L1)	NM_172239.4	NP_758439.4
AK026100	977.9	35.8	27.3	FLJ22447 假拟的 LOC400221 (FLJ22447)	NR_039985.1	
AK026825	823.9	13.5	61.1	跨膜蛋白 212(TMEM212)	NM_001164436.1	NP_001157908.1
KENAE1	793.7	36.9	21.5	人Kenae1 mRNA(AB024691)	AB024691.1	BAC57450.1
HESRG	790.4	7.6	104.5	ESRG 假拟的 LOC790952(ESRG)	NR_027122.1	
AK095450	764.2	35.5	21.5	LOC285540 假拟的 的 LOC285540	NR_037934.1	
FLJ36492	733.9	20.9	35.1	CCR4-NOT 转录复合物, 亚基 1(CNOT1)	NM_016284.3 NM_206999.1	NP_057368.3 NP_996882.1
AK124194	688.2	13.4	51.2	FLJ42200 蛋白质	AK124194.1	BAC85800.1
AK096196	619.2	11.0	56.0	假拟的 LOC100129434	XR_109938.1	

[0125]

AK131313	580.8	11.7	49.5	锌指蛋白91假基因 (LOC441666)	NR_024380.1	
FLJ11292	547.0	22.8	24.0	假拟的蛋白质 FLJ11292	NM_018382.1	NP_060852.1
CCDC122	541.3	13.1	41.5	含122的卷曲螺旋结 构域(CCDC122)	NM_144974.3	NP_659411.2
BC070093	522.8	4.3	121.1	cDNA克隆	BC070093.1	

[0126] 实施例5.利用寡核苷酸分支DNA杂交测定进行的验证

[0127] 由上述的Helicos RNA测序规则确定的候选SCGs(列于表7)用Affymetrix QUANTIGENE探针进一步评估。来自4个原发性胰腺管状腺癌(PDAC)的总RNA用QUANTIGENE Plex RNA测定进行分析。结果如下述的(表8)所示。

[0128]

人基因名	PDAC AVG 信号	细胞系 AVG 信号	PDAC/ 细胞系 细胞比例
HSP90BB	36.17	6.22	5.81
KENAE1	0.19	0.16	1.16
AK056558	20.12	5.19	3.88
MGC4836	1.01	0.90	1.11
BC037952	7.78	2.22	3.50
BC070093	0.14	0.01	18.40
BC041426	0.04	0.00	23.22
CCDC122	0.06	0.06	0.97
FLJ36492	1.22	1.14	1.07
HESRG	1.22	0.11	11.16
FLJ11292	0.04	0.00	78.08
AK026100	0.04	0.00	9.04
AK124194	0.03	0.00	9.28
NM_001001704	1.36	0.24	5.61

[0129]

AK096196	10.87	0.32	33.89
AK095450	0.32	0.01	24.64
AK131313	1.04	0.58	1.81
NR_003133	0.21	0.01	32.70
ODF2L	0.55	0.12	4.52
REXO1L1	0.14	0.00	88.12
AK026825	0.05	0.01	4.37
DERP7	0.12	0.10	1.25
BX649144	0.22	0.19	1.20

[0130] 根据这些数据,开发出Affymetrix VIEWRNA探针,用于在福尔马林固定石蜡包埋的(FFPE)原发肿瘤样本中检测P90BB和AK096196。在MGH,于FFPE材料上利用RNA原位杂交(RNA-ISH)对这些探针进行测定。在利用结肠癌细胞系HCT-116,在Nu/Nu鼠中制备的人癌症皮下异种移植物上,观察到阳性染色。在来自MGH的原发性人PDAC样本中对HSP90BB进行了进一步的测定;所述结果如图5所示,结果显示出上皮导管癌细胞的丰度和特异性染色。

[0131] 其他具体实施方案

[0132] 应当理解,尽管本发明结合本申请中的详细描述进行表述,上述的描述仅旨在举例说明,而非限定本发明的范围,本发明由后附的权利要求的范围限定。其他的方面、有益之处和修改也包含在后附的权利要求的范围之内。

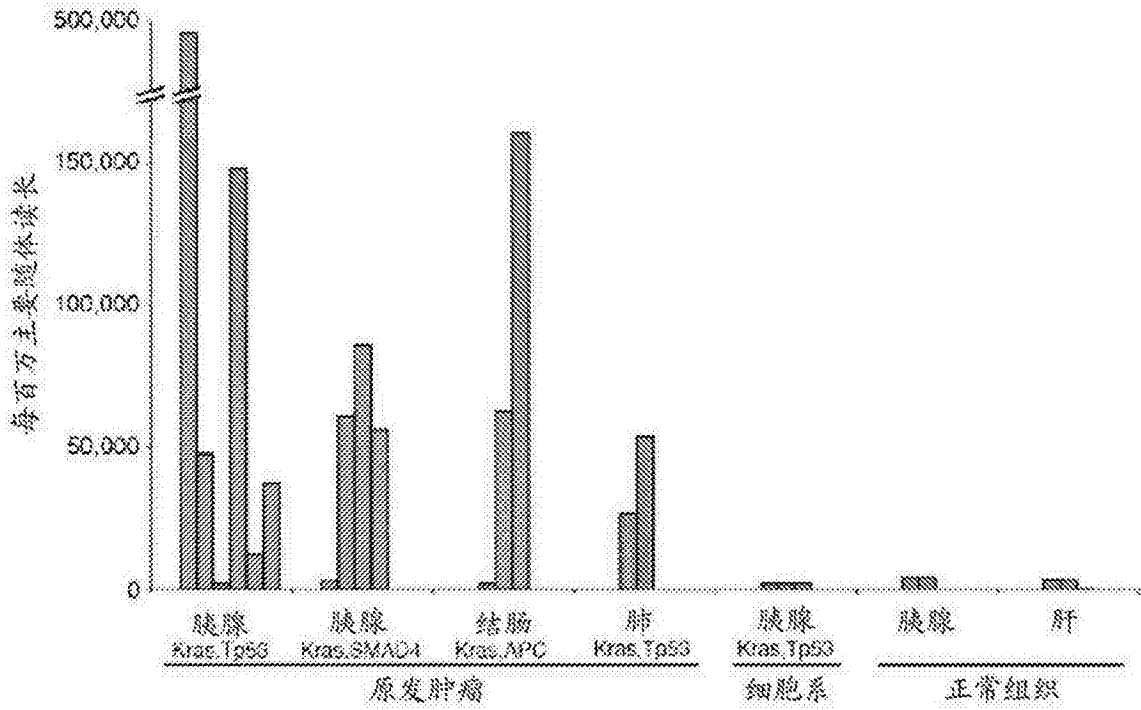


图1A

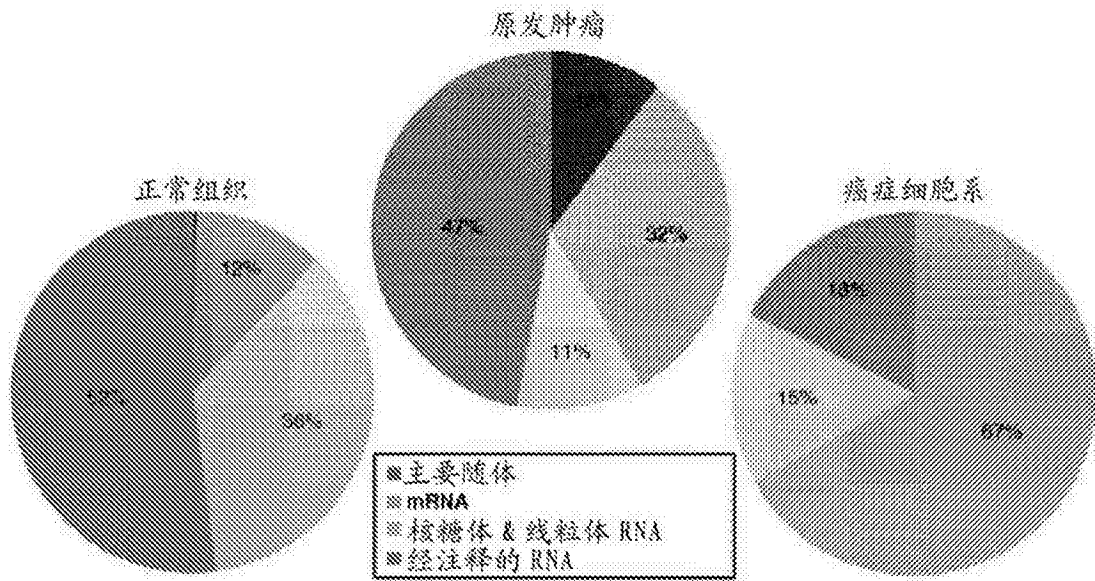


图1B

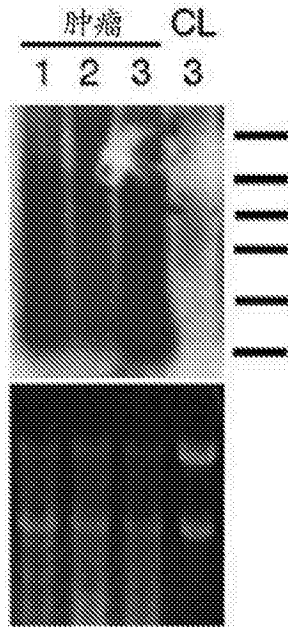


图2A

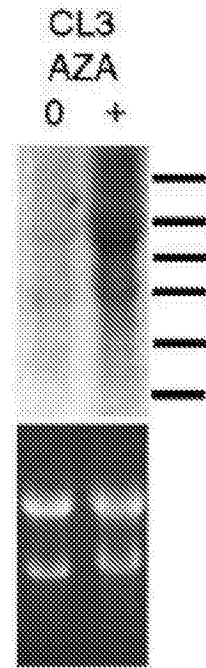


图2B

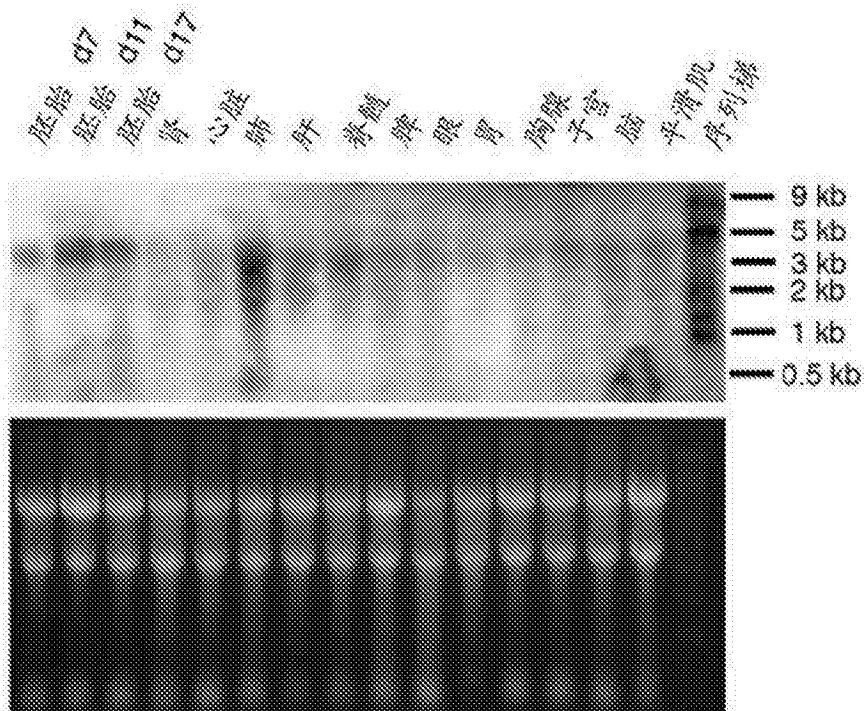


图2C

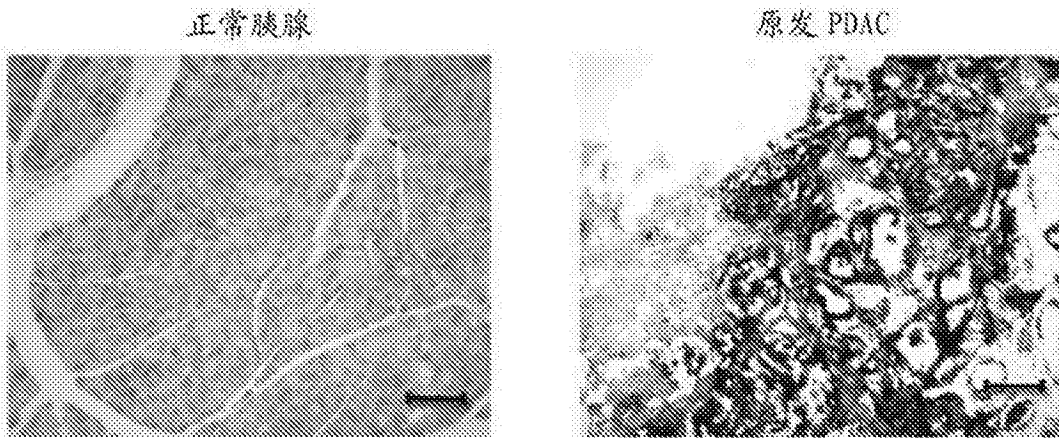


图2D

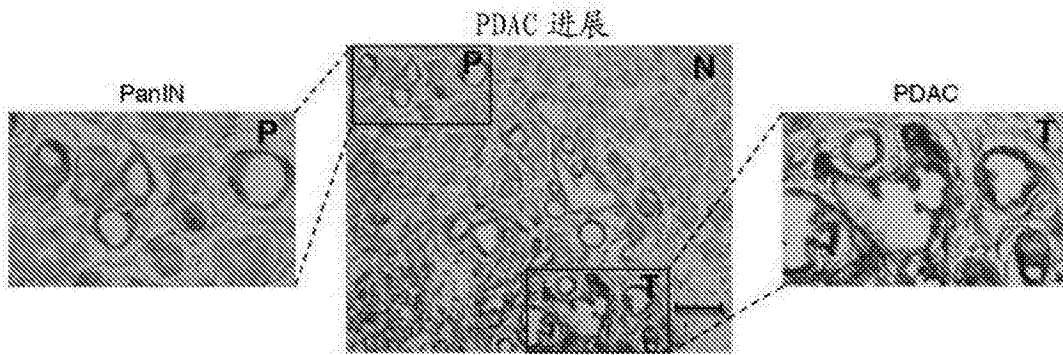


图2E

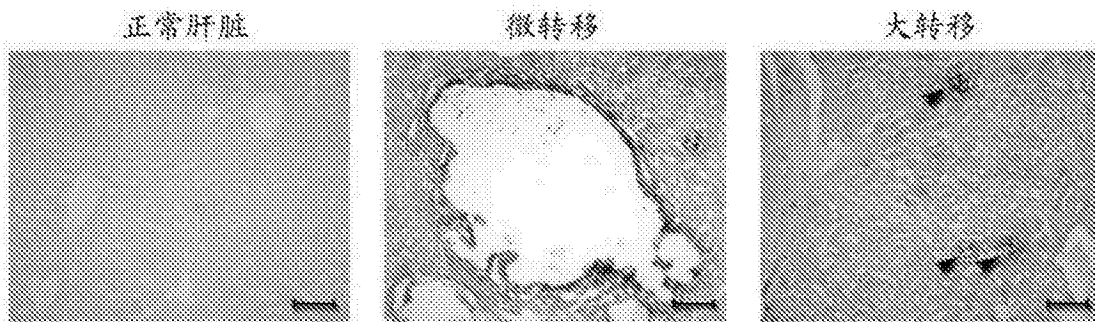


图2F

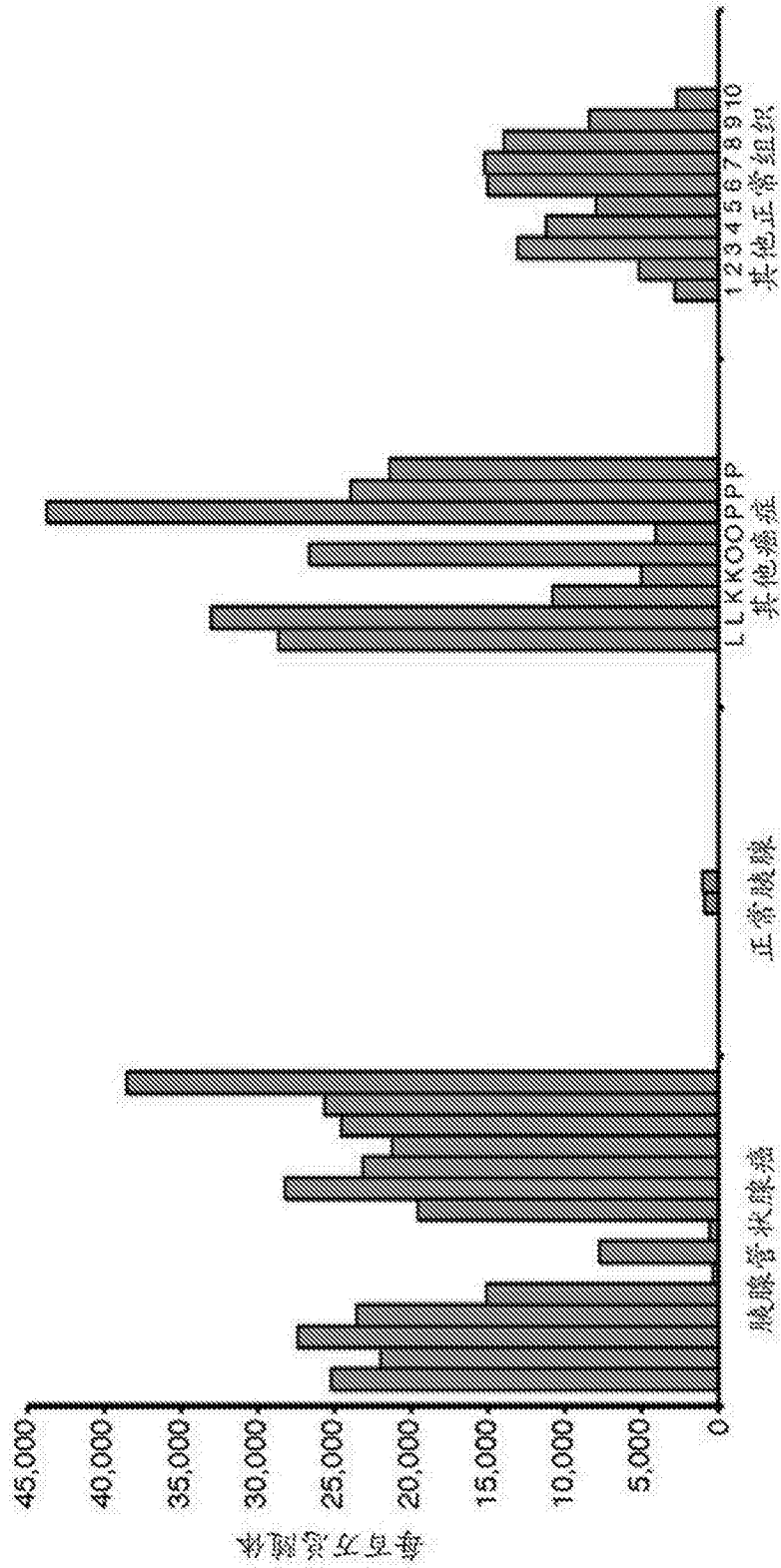


图3A

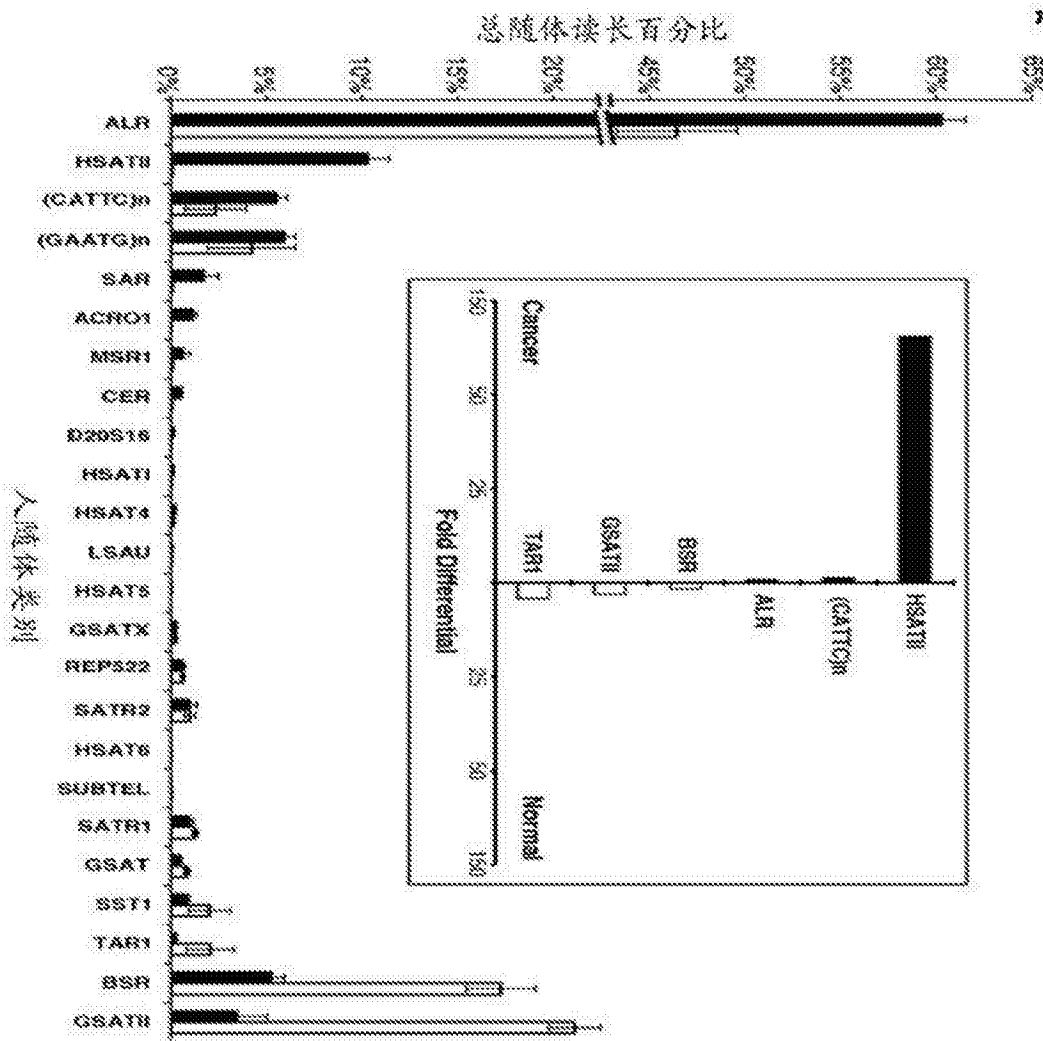
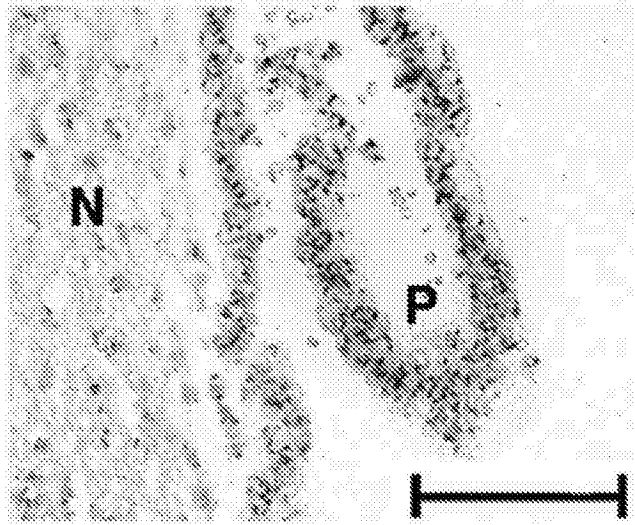


图3B

PanIN



FNA 活检 PDAC

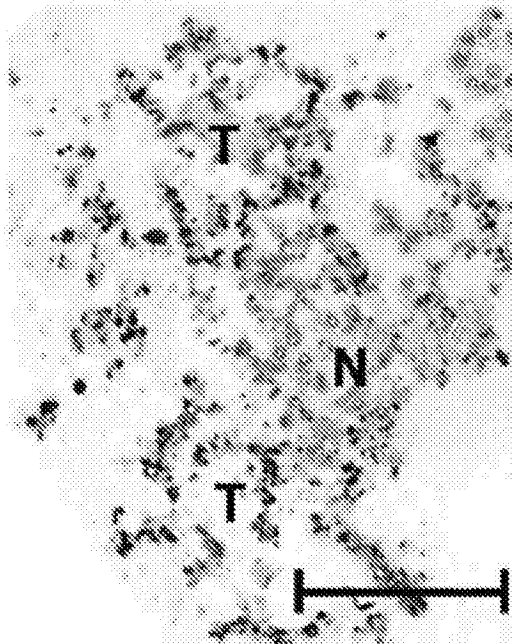


图3C

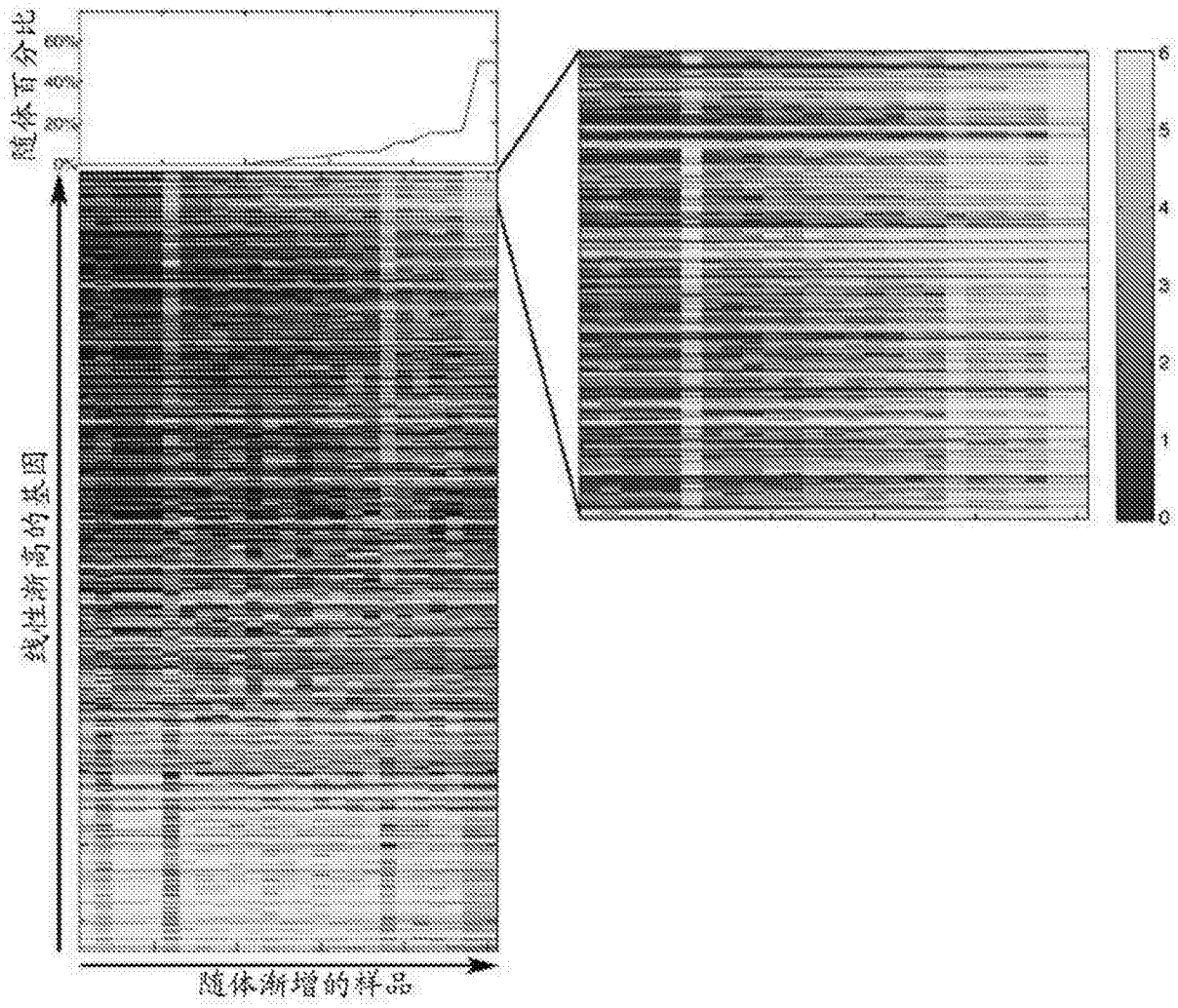


图4A

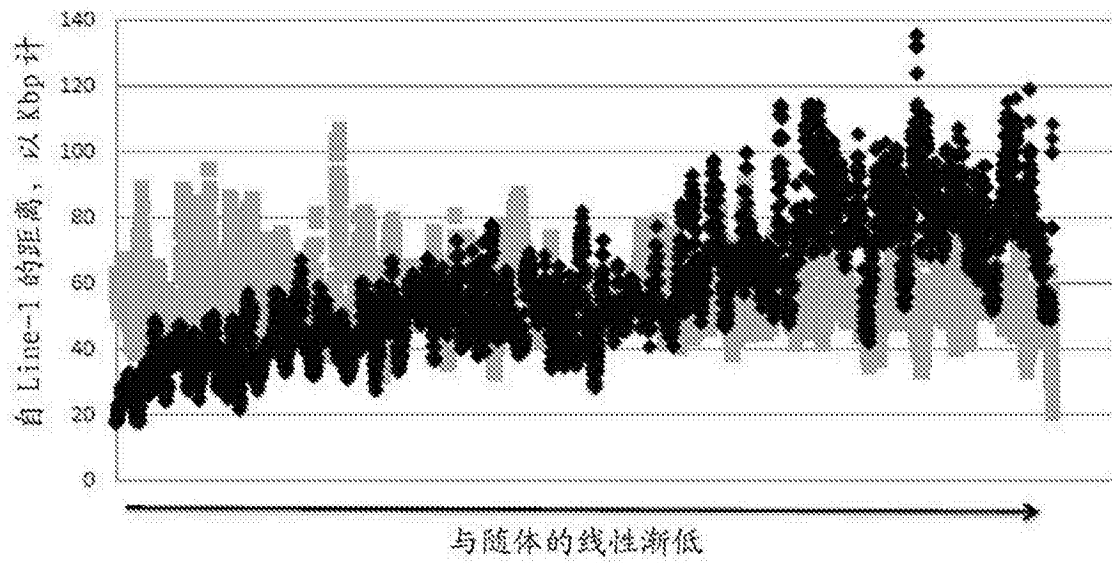


图4B

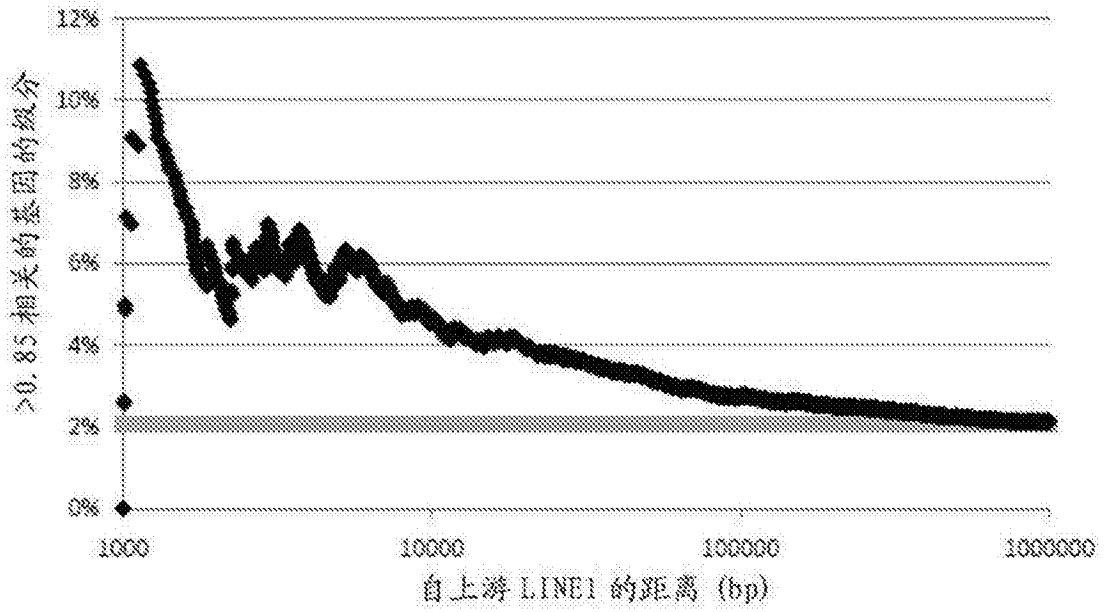


图4C

嗜铬粒蛋白 A

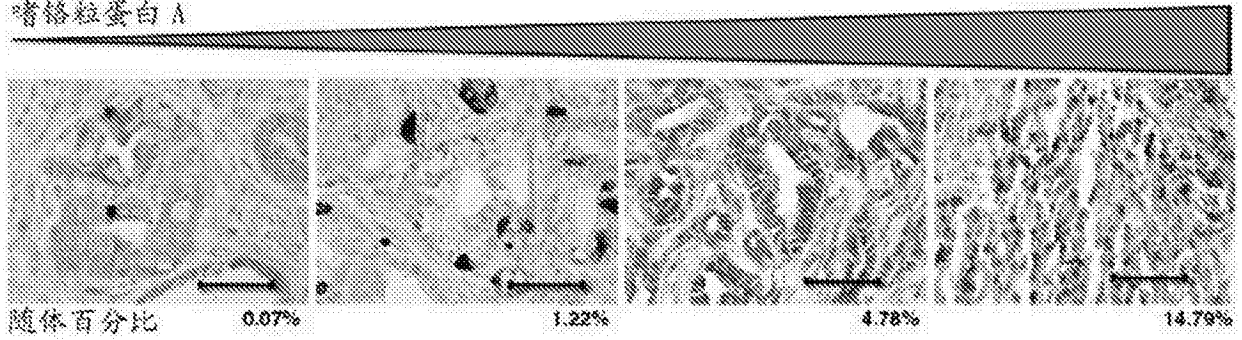


图4D

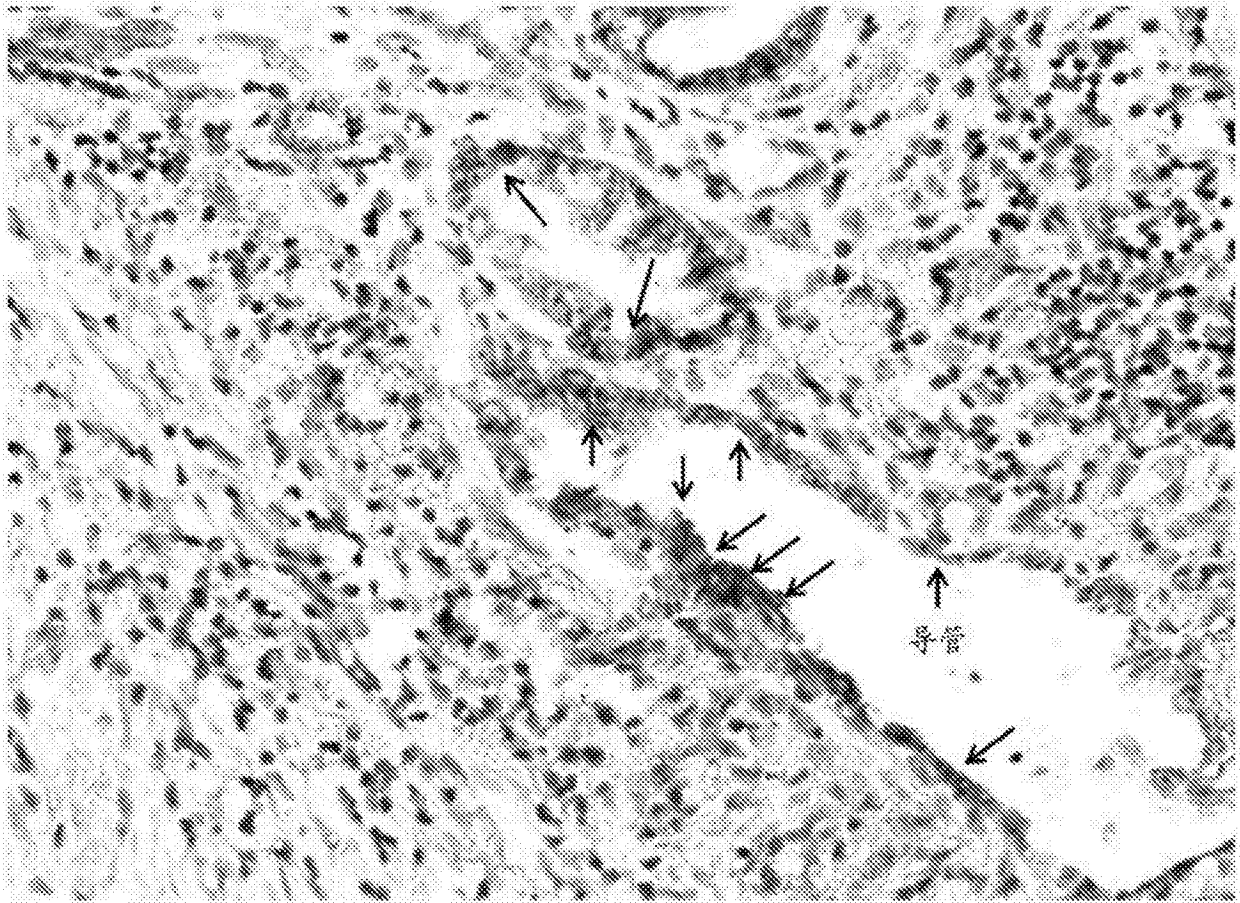


图5