(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 106414768 B
(45)授权公告日 2020.05.29

(21)申请号 201580016472.9

(22)申请日 2015.03.27

(65)同一申请的已公布的文献号
申请公布号 CN 106414768 A

(43)申请公布日 2017.02.15

(30)优先权数据
61/971,455 2014.03.27 US (续)

(85)PCT国际申请进入国家阶段日
2016.09.26

(86)PCT国际申请的申请数据
PCT/US2015/023197 2015.03.27

(87)PCT国际申请的公布数据
WO2015/149034 EN 2015.10.01

(73)专利权人 生命技术公司
地址 美国加利福尼亚州

(72)发明人 D·罗兹 S·萨迪什
P·韦恩加德 N·哈扎诺夫 (续)

(74)专利代理机构 北京戈程知识产权代理有限公司 11314
代理人 程伟

(51)Int.Cl.
*C12Q 1/6886*(2018.01)

(56)对比文件
WO 2013006195 A1,2013.01.10,
WO 2013006195 A1,2013.01.10,
CN 102575287 A,2012.07.11,
CN 102533945 A,2012.07.04,
CN 103468813 A,2013.12.25,
CN 101466721 A,2009.06.24,
CN 102575287 A,2012.07.11,
Jinkuk Kim 等.NTRK1 Fusion in Glioblastoma Multiforme.《plos one》.2014,第9卷(第3期),第e91940页. (续)

审查员 熊壮壮

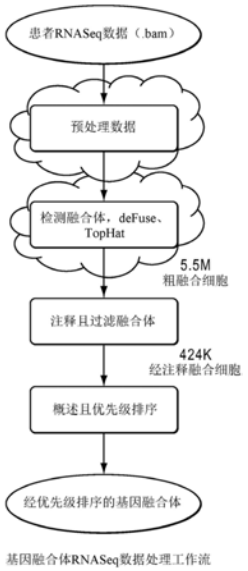
权利要求书1页 说明书202页
序列表77页 附图13页

(54)发明名称

# 与癌症相关的基因融合体和基因变异体

(57)摘要

本发明提供基因融合体、基因变异体和与疾病病况的新颖相关性以及使用其的试剂盒、探针和方法。




基因融合体RNASeq数据处理工作流

CN 106414768 B

[转续页]

[接上页]

(30)优先权数据

61/993,732 2014.05.15 US

62/004,727 2014.05.29 US

62/092,898 2014.12.17 US

(72)发明人 S·班德尔 M·托米略 S·埃迪 E·鲍登 J·李

(56)对比文件

Jinkuk Kim 等.NTRK1 Fusion in Glioblastoma Multiforme.《plos one》.2014,第9卷(第3期),第e91940页.

Lindsey M. Kelly 等.Identification of the transforming STRN-ALK fusion as a potential therapeutic target in the aggressive forms of thyroid cancer.《PNAS》.2014,第111卷(第11期),第4233-4238页.

Lindsey M. Kelly 等.Identification of the transforming STRN-ALK fusion as a potential therapeutic target in the aggressive forms of thyroid cancer.《PNAS》.2014,第111卷(第11期),第4233-4238页.

1.一种试剂盒,其包含一个探针组,其中所述探针组包含:

互补于KANK2基因的第一引物、和

互补于ALK基因的第二引物;

其中所述探针组中的各探针与核酸特异性地杂交,所述核酸包含SEQ ID NO:272所示KANK2/ALK断裂点,并且所述探针组特异性地识别KANK2基因和ALK基因的融合体。

2.根据权利要求1所述的试剂盒,其中所述各探针包含脱氧核糖核酸序列,所述脱氧核糖核酸序列包含至少一个脱氧尿苷残基来代替脱氧胸苷残基。

3.一种组合物,其包含一个探针组,所述探针组包含:

互补于KANK2基因的第一引物、和

互补于ALK基因的第二引物;

其中所述探针组中的各探针与核酸特异性地杂交,所述核酸包含SEQ ID NO:272所示KANK2/ALK断裂点,并且所述探针组特异性地识别KANK2基因和ALK基因的融合体。

4.探针组在制备试剂盒中的用途,其中所述试剂盒用于检测受试者生物样品中的KANK2/ALK基因融合体,其中:

其中所述探针组包含第一引物和第二引物,所述第一引物互补于KANK2基因,并且所述第二引物互补于ALK基因;

其中所述探针组中的各探针与核酸特异性地杂交,所述核酸包含SEQ ID NO:272所示KANK2/ALK断裂点,并且所述探针组特异性地识别KANK2基因和ALK基因的融合体。

5.根据权利要求4所述的用途,其中所述探针组与至少一种包含SEQ ID NO:272所示KANK2/ALK断裂点的核酸特异性地杂交。

6.根据权利要求4所述的用途,其中所述检测通过以下方法进行,所述方法包括步骤:使来自受试者的核酸样品与包含所述第一引物和第二引物的反应混合物接触。

7.根据权利要求4所述的用途,其中可通过由所述探针组所生成的扩增子的存在来检测所述KANK2/ALK基因融合体。

8.根据权利要求4所述的用途,其中所述检测通过以下方法进行,所述方法包括步骤:

从所述生物样品中扩增包含SEQ ID NO:272的核酸;以及

检测包含SEQ ID NO:272的核酸的存在。

9.根据权利要求8所述的用途,其中所述检测包括下一代测序。

10.根据权利要求4所述的用途,其中所述生物样品是肿瘤样品。

11.根据权利要求4所述的用途,其中所述生物样品包含肉瘤细胞。

# 与癌症相关的基因融合体和基因变异体

## 技术领域

[0001] 本发明大体上涉及与癌症相关的基因融合体和基因变异体。

## 背景技术

[0002] 人类癌细胞中经常发现如染色体易位和基因变异体的畸变。染色体易位可能产生表达融合体转录物的嵌合基因，所述嵌合基因接着翻译成融合蛋白，所述融合蛋白会影响正常调节路径且刺激癌细胞生长。基因变异体还可产生会影响正常调节路径的异常蛋白。

[0003] 新融合体基因、已知融合体基因之新变异体和基因变异体或等位基因之鉴别为其他诊断和癌症治疗目标提供机会。

## 发明内容

[0004] 本发明提供新颖基因融合变异体和基因融合体-疾病病况相关性。本文提供的基因融合体与特定癌症相关。本发明进一步提供探针(如扩增引物组和检测探针)以及检测、诊断和治疗的方法和系统以及包括或检测本文所公开的基因融合体的试剂盒。

[0005] 在某些实施例中，与癌症相关的新颖突变和融合体公开于表41-44中。

[0006] 本发明进一步提供包含或用以检测表41-44的新颖突变和融合体事件的反应混合物、探针组、试剂盒、方法、经分离核酸。

[0007] 在一个实施例中，本发明提供一种反应混合物，其包含特异性地识别选自表1-表3、表19和表22的基因融合体的一个探针或一个探针组。所述探针组可例如为一个扩增引物组。在另一实施例中，本文提供一种反应混合物，其包括侧接目标核酸中选自表1-表3、表19和表22的基因融合体的一个引物组。举例来说，所述引物组可以各自在表4-6、20和23中鉴别的融合体断裂点中的一者的相对侧的1000、750、500、250、100、90、80、75、70、65、50或25个核苷酸内结合于人类基因组中的目标序列。此实施例的反应混合物可以进一步包括检测器探针，所述检测器探针结合于选自表1-表3、表19和表22的基因融合体中的断裂点的任一侧，或结合跨越选自表1-表3、表19和表22的基因融合体中的断裂点的结合区，包括其中断裂点鉴别于表4-6、20和23中的特定实施例。在例示性实施例中，检测器探针在表4-6、20和23中鉴别的融合体断裂点中的一者的1000、750、500、250、100、90、80、75、70、60、50或25个核苷酸内结合于人类基因组中的目标序列。包括检测器探针或不包括检测器探针的反应混合物可以进一步包括聚合酶、逆转录酶、dNTP和/或尿嘧啶DNA去糖基化酶(UDG)。聚合酶、逆转录酶和UDG通常不来自人类来源。说明性实施例中的聚合酶是热稳定聚合酶，如Taq聚合酶。在某些实施例中，反应混合物中的dNTP包括dUTP，且在某些实例中，反应混合物可以没有dTTP。此外，反应混合物可以包括扩增子，如DNA扩增子，其包括一个或多个脱氧尿苷(“dU”)残基。在某些实施例中，反应混合物包括DNA扩增子，其包括针对对应人类基因组序列中的每一脱氧胸苷残基的一个或多个dU残基。在某些实施例中，扩增子包括所对应序列并未在人类基因组中发现的区段，如DNA条形码序列。非人类区段的长度可例如为5-10,000、5-5000、5-1000、5-500、5-100、5-50、5-25、5-10、10-10,000、10-5000、10-1000、10-

500、10-100、10-50或10-25个核苷酸。在某些实施例中，扩增子包括对应于人类基因组中跨越内含子的区域的区段，但扩增子并不包括对应于内含子的区段。反应混合物可以进一步包括目标核酸，例如人类目标核酸。人类目标核酸可例如从怀疑患有选自以下各者的癌症的个体的生物样品中分离：BLCA＝膀胱癌、BRCA＝乳癌、CESC＝子宫颈细胞癌、COAD＝结肠腺癌、GBM＝多形性成胶质细胞瘤、HNSC＝头颈鳞状细胞癌、KIRK＝透明细胞肾细胞癌、KIRP＝肾脏肾乳头状细胞癌、LAML＝急性骨髓性白血病、LGG＝脑低级别胶质瘤、LIHC＝肝脏肝细胞癌、LUAD＝肺腺癌、LUSC＝鳞状细胞肺癌、OV＝卵巢浆液性腺癌、PRAD＝前列腺腺癌、READ＝直肠腺癌、SKCM＝皮肤黑素瘤、STAD＝胃腺癌、THCA＝甲状腺癌和UCEC＝子宫体子宫内膜样癌。在某些实施例中，目标核酸来自肿瘤，例如先前句子中列出的癌症类型中的一者的肿瘤。

[0008] 在另一实施例中，提供一种探针组，其特异性地识别包含SEQ ID NO:1-289中的至少一者(基因融合体)的核酸。在另一实施例中，本文提供一种引物组，其特异性地扩增包括SEQ ID NO:1-289中的至少25、30、40、50、75、100、125、150、200个或全部的目标核酸，或扩增SEQ ID NO:1-289中的最多25、30、40、50、75、100、125、150、200个或全部。在另一实施例中，本文提供一种qPCR分析，如TaqMan™分析或MolecularBeacons™分析，其特异性地扩增且检测包括SEQ ID NO:1-289中的至少25、30、40、50、75、100、125、150、200个或全部的目标核酸。

[0009] 本发明还提供一种经分离核酸，其包含选自包括SEQ ID NO:1-289中的至少25、30、40、50、75、100、125、150、200个或全部或包括SEQ ID NO:1-289中的最多25、30、40、50、75、100、125、150、200个或全部的区段的至少一个序列。经分离核酸可以在5'末端上包括第一引物。此外，核酸可能为单链或双链的。在某些实施例中，经分离核酸包括所对应序列并未在人类基因组中发现的区段，如DNA条形码序列。区段的长度可例如为5-10,000、5-5000、5-1000、5-500、5-100、5-50、5-25、5-10、10-10,000、10-5000、10-1000、10-500、10-100、10-50或10-25个核苷酸。

[0010] 在其他实施例中，本发明提供一种试剂盒，其包括一个检测器探针和/或一个探针组，例如一个扩增引物组，其特异性地识别包含选自表1-表3、表19和表22的基因融合体的断裂点的核酸。举例来说，在某些实施例中，所述检测器探针或扩增引物组经设计以扩增和/或检测包括SEQ ID NO:1-289中的至少一者的最多25、30、40、50、75、100、125、150、200个或全部的核酸。所述试剂盒可以进一步在一个或多个各别容器中或在相同容器中包括来自扩增反应混合物的至少一种组分，如聚合酶、dNTP、逆转录酶和/或UDG，通常逆转录酶、聚合酶和UDG不来自人类来源。在某些实施例中，dNTP包括dUTP，且在说明性实例中没有dTTP。说明性实施例中的聚合酶是热稳定聚合酶，如Taq聚合酶。此外，所述试剂盒可以包括对照核酸。举例来说，对照核酸可以包括包含选自表1-表3、表19和表22的基因融合体中的断裂点的序列，如包括SEQ ID NO:1-289中的至少25、30、40、50、75、100、125、150、200个或全部的核酸或包括SEQ ID NO:1-289中的最多25、30、40、50、75、100、125、150、200个或全部的核酸。

[0011] 提供一种检测癌症的方法，其包含扩增跨越选自表1-表3、表19和表22的基因融合体中的断裂点的核酸，例如，所述核酸可以包括选自SEQ ID NO:1-289的序列；以及检测所述核酸的存在，其中所述核酸的存在指示癌症存在于样品中。在另一方法中，本文提供一种

检测选自膀胱癌、结肠癌、乳癌、子宫内膜癌、黑素瘤、卵巢癌、成胶质细胞瘤、神经胶瘤、白血病、肾细胞癌、甲状腺癌和前列腺腺癌的癌症的方法，其包括产生包括选自SEQ ID NO:1-289的序列的扩增子，以及检测所述扩增子的存在，其中所述扩增子的存在指示膀胱癌、结肠癌、黑素瘤、卵巢癌、成胶质细胞瘤、肺癌、神经胶瘤、白血病、肾细胞癌、甲状腺癌、子宫内膜子宫内膜样腺癌、乳癌和前列腺腺癌存在于样品中。所述扩增子通常包括经延伸以形成所述扩增子的引物。所述癌症是选自膀胱尿道上皮癌、乳癌、子宫内膜子宫内膜样腺癌、结肠腺癌、多形性成胶质细胞瘤、透明细胞肾细胞癌、乳头状肾细胞癌、急性骨髓性白血病、脑低级别胶质瘤、肺腺癌、卵巢浆液性囊腺癌、前列腺腺癌、直肠皮肤黑素瘤和甲状腺腺体癌。在特定说明性实施例中，所产生的扩增子是DNA扩增子，其包括dU残基，且在某些实例中不包括dT残基。在此段落中所提供的方法中，扩增子可使用本文提供的反应混合物产生。在某些实施例中，所述方法包括检测跨越选自表1-表3、表19和表22的基因融合体中的断裂点的核酸的表达。用于检测表达的方法通常包括从样品(如肿瘤样品，在说明性实施例中，其可为福尔马林(formalin)固定的样品)分离RNA的步骤。

[0012] 在一个实施例中，反应混合物包括选自SYBR Green、SBYR Greener、荧光素(Fluorescein)、俄勒冈绿(Oregon Green)、FAM、TET、JOE、VIC、雅基马黄(Yakima Yellow)、HEX、Cy3、Bodipy TMR、NED、TAMRA、Cy3.5、ROX、德克萨斯红(Texas Red)、LightCycler Red、Bodipy 630/650、Alexa Fluor 647、Cy5、Alexa Fluor 660或Cy 5.5的染料。在某些实施例中，染料连接于反应混合物中的可检测标记的探针。在其他实施例中，染料直接或经由可检测标记的探针结合于扩增子。

[0013] 提供一种试剂盒，其包含一个探针或一个探针组，例如一个可检测探针或一个扩增引物组，其特异性地识别包含来自表4-6、20和23的断裂点的核酸。所述试剂盒可以在同一容器中或在某些优选实施例中在另一容器中进一步包括来自扩增反应混合物的组分，如聚合酶(通常不来自人类来源)、dNTP和/或UDG。此外，所述试剂盒可以包括对照核酸。举例来说，对照核酸可以包括包含选自表4-6、20和23的断裂点的序列。

[0014] 在另一实施例中，本文提供一种基因融合体，其包括表1-3、19和22中鉴别的基因融合体。在说明性实施例中，基因融合体包括表4-6、20和23中鉴别的断裂点中的一者。因此，本文提供一种经分离基因融合体核酸，其长度在100与10,000个核苷酸之间且在表4-6、20和23中的断裂点中的一者的任一侧上包含至少25个核苷酸。

[0015] 在一个相关实施例中，本文提供一种经分离基因融合体核酸，其包含表4-6、20和23中的断裂点中的至少一者。在某些实施例中，经分离基因融合体核酸包含SEQ ID NO:1-289中的至少25、30、40、50、75、100、125、150、200个或全部或包括SEQ ID NO:1-289中的最多25、30、40、50、75、100、125、150、200个或全部的核酸。经分离基因融合体核酸的长度可例如在50与100,000个核苷酸之间、在100与50,000个核苷酸之间、在100与25,000个核苷酸之间、在100与10,000个核苷酸之间、在100与5,000个核苷酸之间、在100与2500个核苷酸之间、在100与1,000个核苷酸之间、在100与500个核苷酸之间、在100与250个核苷酸之间、在100与200个核苷酸之间、在250与10,000个核苷酸之间、在250与5,000个核苷酸之间、在250与1,000个核苷酸之间或在250与500个核苷酸之间。在某些方面中，经分离基因融合体核酸是DNA。在某些说明性实施例中，经分离核酸基因融合体没有内含子序列但跨越包括一个或多个内含子的基因组中的区域。在某些实施例中，经分离基因融合体核酸是cDNA。

[0016] 在另一实施例中，提供经分离基因融合体核酸，其包含表4-6、20和23中的断裂点中的至少一者。

[0017] 在另一实施例中是一种方法，其通过检测选自表1-表3、表19和表22的基因融合体的存在来检测样品中的选自膀胱尿道上皮癌、乳癌、子宫内膜子宫内膜样腺癌、结肠腺癌、多形性成胶质细胞瘤、透明细胞肾细胞癌、乳头状肾细胞癌、急性骨髓性白血病、脑低级别胶质瘤、肺腺癌、卵巢浆液性囊腺癌、前列腺腺癌、直肠皮肤黑素瘤和甲状腺腺体癌的癌症。

[0018] 本发明提供新颖基因变异体和基因变异体-疾病病况相关性。基因变异体可以具有一个或多个产生变异蛋白的突变。本文提供的基因变异体与某些癌症相关。基因变异体产生蛋白变异体。本发明进一步提供探针(如扩增引物组和检测探针)以及检测、诊断和治疗的方法以及包括或检测本文所公开的基因变异体的试剂盒。

[0019] 在一个实施例中，本发明提供一种组合物和一种试剂盒，其包含特异性地识别编码选自表7和/或表11的基因变异体的核苷酸序列的一个探针组。所述探针组可例如为一个扩增引物组。在另一实施例中，本文提供一种组合物，其包括侧接编码表7和/或表11中的一种或多种变异体的基因变异体的一个引物组。此实施例的反应混合物可以进一步包括检测器探针，其结合于包括选自表7和/或表11的基因变异体的核苷酸序列。包括检测器探针或不包括检测器探针的反应混合物可以进一步包括聚合酶、dNTP和/或尿嘧啶DNA去糖基化酶(UDG)。聚合酶和UDG通常不来自人类来源。反应混合物可以进一步包括目标核酸，例如人类目标核酸。人类目标核酸可例如从来自怀疑患有癌症的个体的生物样品中分离。所述癌症可选自:BLCA＝膀胱癌、BRCA＝乳癌、CESC＝子宫颈细胞癌、COAD＝结肠腺癌、GBM＝多形性成胶质细胞瘤、HNSC＝头颈鳞状细胞癌、KIRK＝透明细胞肾细胞癌、KIRP＝肾脏肾乳头状细胞癌、LAML＝急性骨髓性白血病、LGG＝脑低级别胶质瘤、LIHC＝肝脏肝细胞癌、LUAD＝肺腺癌、LUSC＝鳞状细胞肺癌、OV＝卵巢浆液性腺癌、PRAD＝前列腺腺癌、READ＝直肠腺癌、SKCM＝皮肤黑素瘤、STAD＝胃腺癌、THCA＝甲状腺癌和UCEC＝子宫体子宫内膜样癌。

[0020] 编码表7和/或表11中的一种或多种基因变异体的核苷酸序列可为涵盖所述变异的任何大小。举例来说，核苷酸序列可为可容易地使用引物拷贝和/或使用探针检测的任何大小。

[0021] 在另一实施例中，提供一种探针组，其特异性地识别编码选自表7和/或表11(基因变异体)的基因变异体的核酸。在另一实施例中，本文提供一种引物组，其特异性地扩增编码选自表7和/或表11的基因变异体的目标核酸。在另一实施例中，本文提供qPCR分析，如(但不限于)TaqMan™分析、Scorpions分析或MolecularBeacons™分析，其特异性地扩增且检测编码选自表7和/或表11的基因变异体的目标核酸。

[0022] 本发明还提供一种经分离核酸，其包含编码选自表7和/或表11的一种或多种基因变异体的至少一个序列。经分离核酸可以在5'末端上包括第一引物。此外，核酸可能为单链或双链的。

[0023] 在其他实施例中，本发明提供一种试剂盒，其包括一个检测器探针和/或一个探针组，例如一个扩增引物组，其特异性地识别编码选自表7和/或表11的基因变异体的核酸。举例来说，在某些实施例中，检测器探针或扩增引物组经设计以扩增和/或检测编码表7和/或表11中的变异体的核酸。所述试剂盒可以在另一容器中或在同一容器中进一步包括来自扩增反应混合物的组分，如聚合酶(通常不来自人类来源)、dNTP和/或UDG。此外，所述试剂盒

可以包括对照核酸。举例来说，对照核酸可以包括包含选自表7和/或表11的基因变异体的序列。

[0024] 提供一种检测癌症的方法，其包含扩增编码选自表7和/或表11的基因变异体的核酸，例如所述核酸可以包括来自表7和/或表11中的登录号中的一者的序列，除了所述序列含有编码表7和/或表11中的基因变异体的变异体以外；以及检测所述核酸的存在，其中所述核酸的存在指示癌症存在于样品中。在另一方法中，本文提供一种检测癌症的方法，其包括产生包括编码选自表7和/或表11的变异体的序列的扩增子，以及检测所述核酸的存在，其中所述核酸的存在指示癌症或癌细胞存在于样品中。所述扩增子通常包括经延伸以形成所述扩增子的引物。所述癌症是选自膀胱癌、乳癌、子宫颈细胞癌、结肠腺癌、多形性成胶质细胞瘤、头颈鳞状细胞癌、透明细胞肾细胞癌、肾脏肾乳头状细胞癌、急性骨髓性白血病、脑低级别胶质瘤、肝脏肝细胞癌、肺腺癌、鳞状细胞肺癌、卵巢浆液性腺癌、前列腺腺癌、直肠腺癌、皮肤黑素瘤、胃腺癌、甲状腺癌和子宫体子宫内膜样癌。

[0025] 提供一种试剂盒，其包含特异性地识别包含来自表7和/或表11的基因变异体的核酸的一个探针组(例如一个扩增引物组)。所述试剂盒可以在另一容器中或在同一容器中进一步包括来自扩增反应混合物的组分，如聚合酶(通常不来自人类来源)、dNTP和/或UDG。此外，所述试剂盒可以包括对照核酸。举例来说，对照核酸可以包括包含来自表7和/或表11的基因变异体的序列。

[0026] 在某些实施例中，提供一种探针组，其特异性地识别包含来自表7和/或表11的基因变异体的核酸。

[0027] 在另一实施例中，提供一种基因变异体，其包含表7和/或表11中的基因变异体中的至少一者。

[0028] 在另一实施例中是一种方法，其通过检测选自表7和/或表11的基因变异体的存在来检测样品中的选自膀胱癌、乳癌、子宫颈细胞癌、结肠腺癌、多形性成胶质细胞瘤、头颈鳞状细胞癌、透明细胞肾细胞癌、肾脏肾乳头状细胞癌、急性骨髓性白血病、脑低级别胶质瘤、肝脏肝细胞癌、肺腺癌、鳞状细胞肺癌、卵巢浆液性腺癌、前列腺腺癌、直肠腺癌、皮肤黑素瘤、胃腺癌、甲状腺癌和子宫体子宫内膜样癌的癌症。基因变异体可以包括(但不限于)ZNF479变异体R11Q、R295K、R295T、R295I、R345I、R345T、K438T和T466K。

[0029] 在另一实施例中，提供一种向受试者递送药物的方法，其中所述方法包括检测表15中鉴别的基因事件，且用药物治疗受试者，其中所述药物被认为积极地影响具有所述基因事件的患者的临床结果。在说明性实施例中，基因事件与表8中发现的基因相关，且所述药物在表8中列为所述基因的伴侣。在另一实施例中，本文提供一种判定受试者是否接受药物的方法，所述方法包括检测表15中鉴别的基因事件，且接着如果所检测的基因事件在表15中列为与较差的预后相关，那么向所述受试者递送药物，其中所述药物被认为积极地影响具有所述基因事件的患者的临床结果。在说明性实施例中，基因事件与表8中发现的基因相关，且所述药物在表8中列为所述基因的伴侣。

[0030] 在一个实施例中，提供一种试剂盒，其中所述试剂盒包含一个探针组，其中每一探针与包含来自表4-6、20和23的断裂点的核酸特异性地杂交。

[0031] 在一个实施例中，提供一种方法，所述方法包含：扩增来自样品的包含至少一种来自表1-3、19和22的基因融合体的核酸；以及通过以下各项中的至少一者检测所述至少一种

基因融合体的存在:使组合物与至少一种探针接触,其中每一探针与所述核酸特异性地杂交,或观察所述核酸中非天然或非原生化学结构的存在;其中检测所述至少一种基因融合体的存在指示至少一种来自表1-3、19和22的癌症存在于样品中。

[0032] 在一个实施例中,提供一种系统,所述系统包含核酸扩增器,其经配置以扩增来自样品的包含来自表1-3、19和22的至少一种基因融合体的核酸,得到经扩增核酸;检测器,其经配置以通过以下各项中的至少一者检测所述经扩增核酸中的所述至少一种基因融合体的存在:(i)使组合物与至少一种探针接触,其中每一探针与所述核酸特异性地杂交,或(ii)观察所述核酸中非天然或非原生化学结构的存在,且经进一步配置以发送检测指示;以及计算机系统,其经配置以接收所述检测指示且基于所述检测指示来确定来自表1-3、19和22的至少一种癌症存在于所述样品中。

[0033] 在一个实施例中,提供一种非暂时性计算机可读程序存储单元,所述非暂时性计算机可读程序存储单元经指令编码,所述指令当由计算机执行时执行一种方法,所述方法包含:接收包含至少一种癌症类型和事件类型的输入,其中所述癌症类型选自表15,且所述事件类型选自表15;在数据库中查询至少一个包含多个字段的条目,其中所述多个字段包含所述癌症类型和所述事件类型中的至少一者;以及发送包含来自所述至少一个条目的所述多个中的至少一个字段的输出,其中所述至少一个字段包含至少一种基因、至少一种可药化基因、至少一种靶向所述至少一种可药化基因的药物、或预后。

[0034] 在一个实施例中,提供一种方法,其中所述方法包含向具有选自表1-3、19和22中列出的基因融合体的至少一种基因融合体的患者投与选自表8、16-17、21和24中列出的药物的至少一种药物。

[0035] 在一个实施例中,提供一种方法,其中所述方法包含使来自患者的核酸样品与包含与第一基因互补的第一引物和与第二基因互补的第二引物的反应混合物接触,其中所述第一基因和所述第二基因的融合体可通过由所述第一引物和所述第二引物生成的扩增子的存在来检测,其中所述融合体包含选自表4-6、20和23中列出的断裂点的断裂点。

[0036] 在一个实施例中,提供一种非暂时性计算机可读程序存储单元,所述非暂时性计算机可读程序存储单元经指令编码,所述指令当由计算机执行时执行一种方法,所述方法包含:接收来自至少一种癌细胞系的RNA序列数据;对所述序列数据运行至少一种基因融合体调用程序,以鉴别所述经处理数据中的融合基因之间的可能的断裂点;过滤所述可能的断裂点以保留候选断裂点,其中每一候选断裂点是在功能基因区的5'非翻译区(UTR)或编码DNA序列(CDS)中,且每一候选断裂点不出现在内含子中;用适用于针对癌症诊断、癌症预后或癌症治疗中的至少一者测定基因融合体的相关性的至少一种注释对所述候选断裂点进行注释,其中所述基因融合体包含所述候选断裂点。

[0037] 在一个实施例中,提供一种非暂时性计算机可读程序存储单元,所述非暂时性计算机可读程序存储单元经指令编码,所述指令当由计算机执行时执行一种方法,所述方法包含:接收来自至少一种癌细胞系的突变数据;用变异体分类、变异体位置或变异体变化中的至少一者对所述突变数据进行注释,得到经注释的突变数据;过滤所述经注释的突变数据,得到基因区域突变数据;将所述基因区域突变数据归类为热点、有害或其他;以及基于所述基因中的突变相对频率和所述基因中所有基因区域突变的分类,将包含所述基因区域突变的基因命名为获得功能、失去功能或再现的其他基因。

[0038] 在一个实施例中，提供一种方法，所述方法包含检测来自受试者的样品中的一种或多种基因融合体，得到基因融合体检测数据，其中所述基因融合体中的至少一者是选自表1-3、19和22中列出的基因融合体，通过计算机系统接收所述基因融合体检测数据，以及基于所述基因融合体检测数据，通过所述计算机系统确认推荐给所述受试者的至少一种治疗选择。

[0039] 在一个实施例中，提供一种系统，所述系统包含检测器，其经配置以(i)检测来自受试者的样品中的一种或多种基因融合体，得到基因融合体检测数据，其中所述基因融合体中的至少一者选自表1-3、19和22中列出的基因融合体和(ii)发送所述基因融合体检测数据；以及计算机系统，其经配置以接收所述基因融合体检测数据且基于所述基因融合体检测数据来确认推荐给所述受试者的至少一种治疗选择。

[0040] 在另一实施例中，提供新颖TP53WT基因标签以及检测表40中的TP53WT基因标签基因中的一者或多者的表达水平的方法。

## 附图说明

[0041] 图1提供基因融合体RNASeq数据处理的工作流。

[0042] 图2展示获得功能和失去功能的基因的基因变异体的分类流程。

[0043] 图3概述将各种数据类型整合到基因事件数据库(GEDB)中的数据流。

[0044] 图4是展示基因事件的上卷的流程图

[0045] 图5是展示与含有TP53点突变的乳癌样品相比TP53WT表达标签在TP53WT乳癌中显著升高的图式。

[0046] 图6是展示与含有TP53突变的肺癌样品相比TP53WT表达标签在TP53WT肺癌中显著升高的图式。

[0047] 图7是展示与含有TP53突变的卵巢癌样品相比TP53WT表达标签在HP53WT卵巢癌中显著升高的图式。

[0048] 图8A-D为描绘浆液性卵巢癌患者中的PLXNB21和COL7A1的粗RPKM表达值(A-B)对比z分数归一化值(C-D)的曲线。在外显子12、17和23处的PLXNB1表达中的广泛群体突降在归一化数据中被平滑处理。被预测为具有这些基因之间的融合体的样品，红色菱形指示调用程序预测的断裂点外显子。

[0049] 图9A-B是利用氨基酸位置的常见TP53突变的表。显示在泛癌症分析中以在患者中>0.25％的整体频率出现的突变。再现的剪接位点突变在影响T-125的内含子-外显子连接点处鉴别

[0050] 图10A-B是Tp53框内插入和缺失突变的表。所鉴别的最大的检测到的框内插入-缺失为21bp。基因组上的大于99％的非转座子插入缺失<100bp。

## 具体实施方式

[0051] 本发明提供新颖基因融合体和变异体，以及基因融合体和/或基因变异体与某些癌症类型的新颖相关性。进一步提供涉及本文所公开的基因融合体和/或变异体的探针、反应混合物、分析和试剂盒。

[0052] 定义

[0053] 术语“标记物”或“生物标记物”是指细胞中所表达的、癌细胞表面上所表达的或与非癌细胞相比由癌细胞所分泌的分子(通常是蛋白质、核酸、碳水化合物或脂质),且其适用于诊断癌症,用于提供预后,以及用于药物对癌细胞的优选靶向。时常,所述标记物为与非癌细胞相比在癌细胞中过度表达的分子,例如与正常细胞相比1倍过度表达、2倍过度表达、3倍过度表达或更多。另外,标记物可以是在癌细胞中不适当地合成的分子,例如与在正常细胞上表达的分子相比含有缺失、添加或突变的分子。可替代地,此类生物标记物是与非癌细胞相比在癌细胞中表达不足的分子,例如1倍表达不足、2倍表达不足、3倍表达不足或更多。另外,标记物可以是在癌症中不适当地合成的分子,例如与在正常细胞上表达的分子相比含有缺失、添加或突变的分子。

[0054] 熟练的业内人士应了解,标记物可与其他标记物或测试组合用于本文中所公开的用途(例如癌症的预测、诊断或预后)中的任一个。

[0055] “生物样品”包括组织切片,如活检和尸检样品,以及为了组织学目的获取的冷冻切片。举例来说,生物样品可以包括新鲜冷冻的石蜡包埋的(FFPE)样品。可替代地,生物样品可以包括血液和血液级分或产物(例如血清、血浆、血小板、红细胞等)、痰液、支气管肺泡灌洗液、经培养细胞(例如原代培养物、外植体和转化细胞)、粪便、尿液等。生物样品通常从真核生物体获得,最优选地是哺乳动物,如灵长类动物,例如黑猩猩或人类;牛;犬;猫;啮齿动物,例如豚鼠、大鼠、小鼠;兔;或鸟;爬行动物;或鱼。

[0056] “活检”是指移出组织样品以便诊断或预后评估的过程,以及是指组织样本自身。所属领域中已知的任何活检技术均可应用于本发明的诊断和预后方法。所应用的活检技术将取决于待评估的组织类型(例如肺等)、肿瘤的大小和类型以及其他因素。代表性活检技术包括(但不限于)切除活检、切取活检、针活检、手术活检和骨髓活检。“切除活检”是指移出整个肿瘤块体,围绕所述肿瘤块体有较小边缘的正常组织。“切取活检”是指从肿瘤内移出楔形组织。通过内窥镜检查或放射照相引导进行的诊断或预后可能需要“核心针活检”或“细针抽吸活检”,其通常从目标组织内获得细胞悬浮液。活检技术例如论述在《哈里森内科学原理》(Harrison’s Principles of Internal Medicine),卡斯珀(Kasper)等人编,第16版,2005,第70章,和整个第V部分中。

[0057] 术语“过度表达(overexpress/overexpression)”或“过度表达的”可互换地指与正常细胞相比通常在癌细胞中以可检测地更大水平翻译或转录的蛋白质或核酸(RNA)。所述术语包括与正常细胞相比因转录、转录后加工、翻译、翻译后加工、细胞定位(例如细胞器、细胞质、细胞核、细胞表面)以及RNA和蛋白质稳定性所致的过度表达。过度表达可以使用检测mRNA(即,RT-PCR、PCR杂交)或蛋白质(即,ELISA、免疫组织化学技术)的常规技术检测。过度表达可以是与正常细胞相比10%、20%、30%、40%、50%、60%、70%、80%、90%或更多。在某些情况下,过度表达是与正常细胞相比1倍、2倍、3倍、4倍或更高水平的转录或翻译。

[0058] 术语“表达不足(underexpress/underexpression)”或“表达不足的”或“下调”可互换地指与正常细胞相比在癌细胞中以可检测地较低水平翻译或转录的蛋白质或核酸。所述术语包括与对照相比因转录、转录后加工、翻译、翻译后加工、细胞定位(例如细胞器、细胞质、细胞核、细胞表面)以及RNA和蛋白质稳定性所致的表达不足。表达不足可以使用检测mRNA(即,RT-PCR、PCR杂交)或蛋白质(即,ELISA、免疫组织化学技术)的常规技术检测。表达

不足可以是与对照相比10％、20％、30％、40％、50％、60％、70％、80％、90％或更小。在某些情况下，表达不足是与对照相比1倍、2倍、3倍、4倍或更低水平的转录或翻译。

[0059] 术语“有差异地表达的”或“有差异地调节的”通常是指与至少一种其他样品相比在一种样品中(在本发明的上下文中通常是与非癌性组织的样品相比在癌症患者中)过度表达(上调)或表达不足(下调)的蛋白质或核酸。

[0060] 阐述一组组分的术语“系统”(真实或抽象)包含一个整体，其中每一组分与整体内的至少一个其他组分相互作用或与其相关。

[0061] 术语“多肽”、“肽”和“蛋白质”在本文中可互换使用以指氨基酸残基的聚合物。这些术语适用于氨基酸聚合物，其中一个或多个氨基酸残基是相应天然存在的氨基酸的人造化学模拟剂，以及适用于天然存在的氨基酸聚合物和非天然存在的氨基酸聚合物。

[0062] 术语“氨基酸”是指天然存在的和合成氨基酸，以及以与天然存在的氨基酸类似的一种方式起作用的氨基酸类似物和氨基酸模拟物。天然存在的氨基酸是由遗传密码编码的氨基酸、以及后来被修饰的那些氨基酸，例如羟基脯氨酸、γ-羧基谷氨酸以及O-磷酸丝氨酸。氨基酸类似物是指具有与天然存在氨基酸相同的基本化学结构(即与氢、羧基、氨基以及R基团结合的α碳)的化合物，例如高丝氨酸、正亮氨酸、蛋氨酸亚砜、蛋氨酸甲基锍。此类类似物具有被修饰的R基团(例如正亮氨酸)或被修饰的肽主链，但保持与天然存在的氨基酸相同的基本化学结构。氨基酸模拟物是指具有与氨基酸的一般化学结构不同的结构但以与天然存在的氨基酸类似的方式起作用的化合物。

[0063] 氨基酸在本文中可以由其通常已知的三字母符号或由IUPAC-IUB生物化学命名法委员会(Biochemical Nomenclature Commission)所推荐的单字母符号来提及。同样地，核苷酸可以由其通常接受的单字母密码来提及。

[0064] 关于氨基酸序列，技术人员将认识到改变、添加或缺失编码序列中的单个氨基酸或较小百分比的氨基酸的核酸、肽、多肽或蛋白质序列的个别取代、缺失或添加是“经保守修饰的变异体”，其中所述变化引起氨基酸经化学上类似的氨基酸取代。提供功能上类似的氨基酸的保守取代表在所属领域中众所周知。所述经保守修饰的变异体另外为且不排除本发明的多晶型变异体、种间同系物和等位基因。

[0065] 以下八组各自含有彼此是保守取代的氨基酸：1)丙氨酸(A)、甘氨酸(G)；2)天冬氨酸(D)、谷氨酸(E)；3)天冬酰胺(N)、谷氨酰胺(Q)；4)精氨酸(R)、赖氨酸(K)；5)异亮氨酸(I)、亮氨酸(L)、蛋氨酸(M)、缬氨酸(V)；6)苯丙氨酸(F)、酪氨酸(Y)、色氨酸(W)；7)丝氨酸(S)、苏氨酸(T)；和8)半胱氨酸(C)、蛋氨酸(M)。参见例如，克赖顿(Creighton)，《蛋白质》(Proteins)(1984)。

[0066] 当提及蛋白质、核酸、抗体或小分子化合物时，短语“特异性地(或选择性地)结合”是指通常在蛋白质或核酸与其他生物制剂的异质群体中决定蛋白质或核酸(如本发明的差异表达基因)的存在的结合反应。在抗体的情况下，在指定的免疫分析条件下，指定抗体可以至少两倍背景且更通常超过10到100倍背景结合于特定蛋白质。在此类条件下特定结合于抗体需要针对其对特定蛋白质的特异性选择的抗体。举例来说，多克隆抗体可以经选择以仅获得与所选抗原而不与其他蛋白质特异性免疫反应的那些多克隆抗体。这一选择可以通过去掉与其他分子交叉反应的抗体来实现。多种免疫分析形式可以用于选择与特定蛋白质特异性免疫反应的抗体。举例来说，固相ELISA免疫分析常规地用于选择与蛋白质特异性

免疫反应的抗体(参见例如,哈洛(Harlow)和莱恩(Lane),《抗体实验指南》(Antibodies,A Laboratory Manual)(1988)关于可以用于测定特异性免疫反应的免疫分析形式和条件的描述)。

[0067] 在用于调节标记物蛋白的测试化合物的分析的情形下,短语“功能效应”包括测定间接或直接受到本发明的生物标记物的影响的参数,例如化学物或表型。因此,功能效应尤其包括配体结合活性、转录活化或抑制、细胞增殖的能力、迁移的能力。“功能效应”包括体外、体内和离体活性。

[0068] “测定功能效应”意指针对增大或减小间接或直接受到本发明的生物标记物的影响的参数的化合物的分析,例如测量物理和化学或表型效应。此类功能效应可以通过所属领域的技术人员已知的任何手段测量,例如光谱特征的变化(例如荧光、吸光度、折射率);水动力(例如形状)、色谱;或蛋白质的溶解度特性;配体结合分析,例如结合于抗体;测量诱导型标记物或标记物的转录活化;测量酶活性的变化;增加或减少细胞增殖、细胞凋亡、细胞周期停滞的能力,测量细胞表面标记物的变化。功能效应可以通过所属领域的技术人员已知的许多手段评估,例如显微镜用于定量或定性测量形态特征的变化,测量在胎盘组织中表达的其他基因的RNA或蛋白质水平的变化,测量RNA稳定性,鉴别下游或报导基因表达(CAT、荧光素酶、β-gal、GFP等),例如经由化学发光、荧光、比色反应、抗体结合、诱导型标记物等。

[0069] 标记物的“抑制剂”、“活化剂”和“调节剂”用于指使用癌症生物标记物的体外和体内分析所鉴别的活化性、抑制性或调节性分子。抑制剂是例如结合以部分或完全阻断活性、降低、阻止、延时活化、灭活、钝化或下调癌症生物标记物的活性或表达的化合物。“活化剂”是增加、打开、启动、促进、增加活化、敏化、激动或上调癌症生物标记物的活性的化合物,例如激动剂。抑制剂、活化剂或调节剂还包括癌症生物标记物的经基因修饰型式(例如具有改变活性的型式)以及天然存在和合成配体、拮抗剂、激动剂、抗体、肽、环肽、核酸、反义分子、核酶、RNAi和siRNA分子、小有机分子等。针对抑制剂和活化剂的此类分析包括例如在体外、在细胞或细胞提取物中表达癌症生物标记物,施加推定的调节剂化合物,且接着测定对活性的功能效应,如上文所述。

[0070] “探针(probe或probes)”是指长度为至少八(8)个核苷酸的聚核苷酸,且其因探针中的至少一个序列与目标区域中的序列的互补性而与目标序列形成杂合结构。聚核苷酸可由DNA和/或RNA构成。在某些实施例中,探针经可检测地标记,如本文中更详细地论述。探针的大小可显著变化。一般来说,探针的长度是例如至少8到15个核苷酸。其他探针是例如至少20、30或40个核苷酸长。其他探针稍微更长,为至少例如50、60、70、80、90个核苷酸长。其他探针还更长,且是至少例如100、150、200个或更多个核苷酸长。探针也可以是落在前述范围内的任何特定长度。优选地,探针不含与在聚合酶链反应期间用于目标序列的引物的序列互补的序列。

[0071] 术语“互补”或“互补性”用于提及通过碱基对规则相关的聚核苷酸(即,核苷酸序列)。举例来说,序列“A-G-T”与序列“T-C-A”互补。互补可以“部分”,其中仅一些核酸的碱基根据碱基配对规则匹配。可替代地,可能存在核酸之间的“完整”或“完全”互补性。核酸链之间的互补性程度对核酸链之间杂交的效率和强度具有显著影响。

[0072] 术语“寡核苷酸”或“聚核苷酸”是指任何长度的核苷酸的聚合形式,为脱氧核糖核

苷酸或核糖核苷酸。这些术语包括(但不限于)单链、双链或三链DNA、基因组DNA、cDNA、RNA、DNA-RNA杂混物或包含嘌呤和嘧啶碱基或其他天然化学、生物化学修饰的非天然或衍生核苷酸碱基的聚合物。

[0073] “扩增检测分析”是指定义扩增子的引物对和匹配的探针,其中引物对侧接目标核酸区(通常为目标基因),且其中探针结合于扩增子。

[0074] 术语“基因变异体”和“核苷酸变异体”在本文中可互换使用以指在特定基因座处相对于参考人类基因或cDNA序列的变化或改变,包括(但不限于)编码区和非编码区中的核苷酸碱基缺失、插入、倒位和取代。缺失可为单个核苷酸碱基、基因的核苷酸序列的一部分或一个区域、或整个基因序列的缺失。插入可为一个或多个核苷酸碱基的插入。“基因变异体”或“核苷酸变异体”可出现在转录调节区、mRNA的非翻译区、外显子、内含子或外显子/内含子连接点中。“基因变异体”或“核苷酸变异体”可能产生或可能不产生终止密码子、框移、氨基酸缺失、改变的基因转录物剪接形式或改变的氨基酸序列。

[0075] 术语“基因”是指聚核苷酸(例如DNA区段),其编码多肽且包括编码区之前和之后的区域以及个别编码区段(外显子)之间的插入序列(内含子)。亲本基因或蛋白质序列以Entrez Gene ID或登录号形式呈现。举例来说,ZNF479Entrez Gene ID是90827。如果已对Entrez中的Gene ID中的序列作出任何改变,那么在所述Gene ID之后用一个小数点和改变数量指示(例如90827.1)。另外,举例来说,TPM1具有登录号NM_004304。

[0076] 术语“等位基因”或“基因等位基因”在本文中用以大体上指具有参考序列的天然存在的基因或含有特定核苷酸变异体的基因。

[0077] 如本文所用,“单倍型”是在个体中发现的染色体上的mRNA或基因组DNA的区域中的基因(核苷酸)变异体的组合。因此,单倍型包括通常作为一个单元一起遗传的许多基因连接的多态变异体。

[0078] 如本文所用,术语“氨基酸变异体”用于指由编码参考蛋白质的参考人类基因的“基因变异体”或“核苷酸变异体”产生的相对于参考人类蛋白质序列的氨基酸变化。术语“氨基酸变异体”不仅打算涵盖单个氨基酸取代,而且打算涵盖氨基酸缺失、插入和参考蛋白质中的氨基酸序列的其他显著变化。本发明的变异体通过以下命名法描述:[原始氨基酸残基/位置/取代的氨基酸残基]。举例来说,在位置76处的亮氨酸取代精氨酸表示为R76L。

[0079] 如本文所用的术语“基因型”意指在基因(或特定染色体区域)的任一等位基因或两个等位基因中的特定核苷酸变异体标记物(或基因座)处的核苷酸字符。就相关基因的特定核苷酸位置来说,在一个或两个等位基因中的所述基因座或其等效物处的核苷酸形成基因在所述基因座处的基因型。基因型可为纯合或杂合的。因此,“基因分型”意指测定基因型,即,在特定基因座处的核苷酸。基因分型也可通过测定在蛋白质的特定位置处的氨基酸变异体来进行,其可以用于推导对应的核苷酸变异体。

[0080] 一个探针组通常是指一个引物组,通常是引物对,和/或用于检测目标遗传变异的可检测标记的探针。引物对用于扩增反应以定义跨越上述基因中的每一种的目标遗传变异区域的扩增子。所述扩增子组通过一组匹配的探针检测。在一个例示性实施例中,本发明是一组TaqMan™(加利福尼亚州普莱森顿的罗氏分子系统公司(Roche Molecular Systems,Pleasanton,CA)分析,其用于检测用于本发明的方法的一组目标遗传变异。

[0081] 在一个实施例中,所述探针组是一组用于生成扩增子的引物,所述扩增子通过核

酸测序反应(如下一代测序反应)检测。举例来说,在这些实施例中,可使用AmpliSEQ™(加利福尼亚州卡尔斯巴德的生命技术/离子激流公司(Life Technologies/Ion Torrent,Carlsbad,CA))或TruSEQ™(加利福尼亚州圣地亚哥的伊路米那公司(Illumina,San Diego,CA))技术。在其他实施例中,两个或更多个探针是引物对。

[0082] 经修饰的核糖核苷酸或脱氧核糖核苷酸是指可以代替核酸中天然存在的碱基使用的分子,且包括(但不限于)经修饰嘌呤和嘧啶;稀有碱基;可转化核苷;嘌呤和嘧啶的结构类似物;经标记、经衍生和经修饰的核苷和核苷酸;结合的核苷和核苷酸;序列修饰剂;末端修饰剂;间隔子修饰剂;和具有主链修饰的核苷酸,包括(但不限于)核糖经修饰的核苷酸、氨基磷酸酯、硫代磷酸酯、膦酰胺酸酯、甲基膦酸酯、甲基亚磷酰胺、甲基膦酰胺酸酯、5'-β-氰基乙基亚磷酰胺、亚甲基膦酸酯、二硫代磷酸酯、肽核酸、非手性和中性核苷酸间键。

[0083] "杂交(Hybridize/hybridization)"是指核酸之间的结合。杂交的条件可以根据待结合的核酸的序列同源性改变。因此,如果受试核酸之间的序列同源性较高,那么使用严格条件。如果序列同源性较低,那么使用温和条件。当杂交条件严格时,杂交特异性增加,且杂交特异性的这一增加引起非特异性杂交产物的产率的降低。然而,在温和杂交条件下,杂交特异性减小,且杂交特异性的这一减小引起非特异性杂交产物的产率的增加。

[0084] "严格条件"是指探针将与其目标子序列(通常在核酸的复杂混合物中)但不与其他序列杂交的条件。严格条件与序列相关,并且会随情况不同而不同。更长序列在更高温度下特异性地杂交。有关核酸杂交的详尽指导可见于迪杰森(Tijssen),《生物化学和分子生物学技术--核酸探针杂交》(Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes),"杂交原理和核酸分析策略综述(Overview of principles of hybridization and the strategy of nucleic acid assays)"(1993)中。通常,选择严格条件比所定义的离子强度pH值下的特定序列的热熔点($T_m$)低约5℃-10℃。$T_m$是50%的与目标互补的探针与目标序列在平衡(当目标序列过量存在时,在$T_m$下,在平衡下占据50%探针)下杂交的温度(在所定义的强度、pH值和核浓度下)。严格条件也可以用添加去稳定化剂(如甲酰胺)来实现。对于选择性或特异性杂交,正信号是至少两倍背景、优选地10倍背景杂交。例示性严格杂交条件可以如下:50%甲酰胺、5×SSC和1%SDS,在42℃下孵育,或5×SSC、1%SDS,在65℃下孵育,在0.2×SSC和0.1%SDS中在65℃下洗涤。

[0085] 如果核酸所编码的多肽实质上相同,那么在严格条件下并不彼此杂交的所述核酸仍实质上相同。这在例如使用由遗传密码准许的最大密码简并形成核酸拷贝时出现。在此类情况下,核酸通常在中等严格杂交条件下杂交。例示性"中等严格杂交条件"包括在40%甲酰胺、1M NaCl、1%SDS的缓冲液中在37℃下杂交,且在1×SSC中在45℃下洗涤。阳性杂交是至少两倍背景。所属领域的技术人员将易于认识到替代杂交和洗涤条件可以用于提供类似严格度的条件。测定杂交参数的其他指导原则提供于许多参考文献例如和《最新分子生物学实验方法汇编》(Current Protocols in Molecular Biology)版中。

[0086] 核酸之间的杂交可以在DNA分子与DNA分子之间进行,在DNA分子与RNA分子之间杂交,以及在RNA分子与RNA分子之间杂交。

[0087] "突变蛋白质"或"变异体"是指分别通过一个或多个核苷酸或氨基酸的交换、缺失或插入而相对于个体群体中的野生型或最流行形式不同的聚核苷酸或多肽。所交换、缺失

或插入的核苷酸或氨基酸的数量可以是1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20个或更多个，如25、30、35、40、45或50个。术语突变蛋白质还可涵盖易位，例如由ALK和TPM1基因编码的多肽的融合体(TPM1/ALK)。

[0088] “基因融合体”是指由第一基因的至少一部分与第二基因的一部分的融合产生的嵌合基因组DNA。融合体中来自第一基因的序列与融合体中来自第二基因的序列之间的过渡点被称为“断裂点”或“融合点”。

[0089] 基因融合体的转录产生嵌合mRNA。

[0090] “单核苷酸多态性”或“SNP”是指当生物物种成员或人类的成对染色体之间的基因组中的单核苷酸(A、T、G或C)不同时出现的DNA序列变异。

[0091] “突变”在本文中定义为基因组位置处的特定变化，即：染色体、启动、终止、参考碱基、替代碱基、变异体类型(SNP、INS、DEL)等。

[0092] “注释”在本文中定义为一组转录物特有的特性，其描述突变的效应，即：基因、转录物、变异体分类、变异体变化、变异体密码子位置等。

[0093] “引物”或“引物序列”是指与目标核酸序列(例如待扩增的DNA模板)杂交以引发核酸合成反应的寡核苷酸。引物可以是DNA寡核苷酸、RNA寡核苷酸或嵌合序列。引物可含有天然、合成或经修饰核苷酸。引物长度的上限和下限凭经验确定。引物长度的下限是在杂交时与目标核酸在核酸扩增反应条件下形成稳定双螺旋所需的最低长度。极短引物(通常是小于3-4个核苷酸长)并不与目标核酸在此类杂交条件下形成热力学上稳定的双螺旋。上限通常是由目标核酸中除预定核酸序列以外的区域中形成双螺旋的可能性来确定的。一般来说，适合的引物长度在约10到约40个核苷酸长范围内。在某些实施例中，举例来说，引物可以是10-40个、15-30个或10-20个核苷酸长。当置放在适当条件下时，引物能够充当聚核苷酸序列上合成的起始点。

[0094] 引物将与待拷贝的目标聚核苷酸序列的区域完全或实质上互补。因此，在有助于杂交的条件下，引物将粘接到目标序列的互补区域上。在添加适合的反应物(包括(但不限于)聚合酶、核苷酸三磷酸酯等)后，引物通过聚合试剂延长以形成目标序列的拷贝。引物可以是单链或可替代地可以是部分地双链。

[0095] “检测”、“可检测”和其语法等效者是指测定目标核酸序列的存在和/或量和/或身份的方式。在一些实施例中，检测进行，扩增目标核酸序列。在其他实施例中，目标核酸的测序的特征可以为“检测”目标核酸。连接于探针的标记可以包括所属领域中已知的可以通过例如化学或物理手段检测的多种不同标记中的任一个。可以连接于探针的标记可以包括例如荧光和发光材料。

[0096] “扩增(Amplifying/amplification)”和其语法等效者是指目标核酸序列中的至少一部分以模板依赖性方式复制的任何方法，包括(但不限于)线性或指数扩增核酸序列的广泛范围的技术。执行扩增步骤的例示性方式包括连接酶链式反应(LCR)、连接酶检测反应(LDR)、接合继之以Q-复制酶扩增、PCR、引物延伸、链置换扩增(SDA)、超支化链置换扩增、多重置换扩增(MDA)、基于核酸链的扩增(NASBA)、两步多重扩增、滚环扩增(RCA)、重组酶-聚合酶扩增(RPA)(英国剑桥的推斯特克斯公司(TwistDx,Cambridg,UK))和自维持序列复制(3SR)，包括多重型式或其组合，例如(但不限于)OLA/PCR、PCR/OLA、LDR/PCR、PCR/PCR/LDR、PCR/LDR、LCR/PCR、PCR/LCR(也被称作组合链式反应-CCR)等。此类技术的描述可以见于萨

姆布鲁克(Sambrook)等人《分子克隆》(Molecular Cloning),第3版;奥苏贝尔(Ausbel)等人;《PCR引物:实验室手册》(PCR Primer:A Laboratory Manual),迪芬巴赫(Diffenbach)编,冷泉港出版社(Cold Spring Harbor Press)(1995);《电子方案图书》(The Electronic Protocol Book),常生物科学(Chang Bioscience)(2002),苏(Msuih)等人,《临床微生物学》(J.Clin.Micro.)34:501-07(1996);《核酸方案手册》(The Nucleic Acid Protocols Handbook),R.瑞雷(R.Rapley)编,纽约州托托瓦的胡马纳出版社(Humana Press,Totowa,N.J.)(2002)以及其他地方。

[0097] 核酸标记物的分析可以使用所属领域中已知的技术执行,包括(但不限于)序列分析和电泳分析。序列分析的非限制性实例包括马克塞姆-吉尔伯特测序(Maxam-Gilbert sequencing)、桑格测序(Sanger sequencing)、毛细管阵列DNA测序、热循环测序(西尔(Sears)等人,《生物技术》(Biotechniques),13:626-633(1992))、固相测序(齐默曼(Zimmerman)等人,《分子细胞生物学方法》(Methods Mol Cell Biol),3:39-42(1992))、测序伴以质谱如基质辅助激光解吸/电离飞行时间质谱(MALDI-TOF/MS;付(Fu)等人,《自然·生物技术》(Nat.Biotechnol),16:381-384(1998))和杂交测序。奇(Chee)等人,《科学》(Science),274:610-614(1996);德尔马纳茨(Drmanac)等人,《科学》(Science),260:1649-1652(1993);德尔马纳茨等人,《自然·生物技术》,16:54-58(1998)。电泳分析的非限制性实例包括平板凝胶电泳(如琼脂糖或聚丙烯酰胺凝胶电泳)、毛细管电泳和变性梯度凝胶电泳。另外,下一代测序方法可以使用商业上可获得的试剂盒和仪器(来自公司如生命技术/离子激流公司PGM或Proton、伊路米那公司HiSEQ或MiSEQ和罗氏公司(Roche)/454下一代测序系统)执行。

[0098] 在一些实施例中,回应于激发光给出荧光信号的探针的量通常与扩增反应中产生的核酸的量有关。因此,在一些实施例中,荧光信号的量与扩增反应中形成的产物的量有关。在所述实施例中,可以因此通过测量来自荧光指示剂的荧光信号的强度来测量扩增产物的量。

[0099] “可检测标记的探针”或“检测器探针”是指用于扩增反应的分子,通常用于定量或实时PCR分析以及终点分析。所述检测器探针可以用于监测目标核酸序列的扩增。在一些实施例中,扩增反应中所存在的检测器探针适用于监测随时间产生的扩增子的量。此类检测器探针包括(但不限于)5'-核酸外切酶分析(本文所述的TAQMAN®探针(也参见美国专利号5,538,848)各种茎环分子信标(参见例如美国专利号6,103,476和5,925,517以及亚吉(Tyagi)和克拉默(Kramer),1996,《自然·生物技术》14:303-308)、无茎或线性信标(参见例如WO 99/21881)、PNA Molecular Beacons™(参见例如美国专利号6,355,421和6,593,091)、线性PNA信标(参见例如库比斯塔(Kubista)等人,2001,SPIE 4264:53-58)、非FRET探针(参见例如美国专利号6,150,097)、Sunrise®/Amplifluor™探针(美国专利号6,548,250)、茎环和双螺旋蝎型探针(索利纳斯(Solinas)等人,2001,《核酸研究》(Nucleic Acids Research)29:E96和美国专利号6,589,743)、凸环探针(美国专利号6,590,091)、假结探针(美国专利号6,589,250)、环化子(cyclicon)(美国专利号6,383,752)、MGB Eclipse™探针(爱博克生物科学公司(Epoch Biosciences))、发夹探针(美国专利号6,596,490)、肽核酸(PNA)光探针、自组装纳米粒子探针和二茂铁修饰探针,例如描述在美国专利号6,485,901;马哈郎加(Mhlanga)等人,2001,《方法》(Methods)25:463-471;惠特科姆(Whitcombe)等人,

1999,《自然·生物技术》.17:804-807;伊萨克森(Isacsson)等人,2000,《分子细胞探针》(Molecular Cell Probes).14:321-328;斯万维克(Svanvik)等人,2000,《分析生物化学》(Anal Biochem.)281:26-35;沃尔夫(Wolffs)等人,2001,《生物技术》(Biotechniques)766:769-771;特索卡斯(Tsourkas)等人,2002,《核酸研究》(Nucleic Acids Research).30:4208-4215;里切利(Riccelli)等人,2002,《核酸研究》30:4088-4093;张(Zhang)等人,2002上海(Shanghai).34:329-332;麦克斯维尔(Maxwell)等人,2002,《美国化学学会杂志》(J.Am.Chem.Soc.)124:9606-9612;布劳德(Broude)等人,2002,《生物技术趋势》(Trends Biotechnol.)20:249-56;黄(Huang)等人,2002,《毒理学化学研究》(Chem.Res.Toxicol.)15:118-126;和余(Yu)等人,2001,《美国化学学会杂志》14:11155-11161。

[0100] 检测器探针还可以包括淬灭剂,包括(但不限于)黑洞淬灭剂(生物谷猎头(Biosearch))、爱荷华黑(Iowa Black)(IDT)、QSY淬灭剂(分子探针公司(Molecular Probes))以及二甲氨基偶氮苯甲酰(Dabsyl)和Dabcel磺酸酯/甲酸酯淬灭剂(爱博克公司(Epoch))。

[0101] 检测器探针还可以包括两个探针,其中例如荧光剂在一个探针上,并且淬灭剂在另一个探针上,其中两个探针在目标上杂交在一起淬灭信号,或其中在目标上杂交通过改变荧光改变了信号特征。检测器探针还可以包含具有$SO_3$而非羧酸酯基团的荧光素染料的磺酸酯衍生物、荧光素的亚磷酰胺形式、CY 5的亚磷酰胺形式(商业上可例如从安玛西亚公司(Amersham)获得)。在一些实施例中,使用嵌入螯合剂标记,如溴化乙锭、SYBR®Green I(分子探针公司)和PicoGreen®(分子探针公司),由此允许在不存在检测器探针的情况下实时或终点观测扩增产物。在一些实施例中,实时观测可以包含插入检测器探针并且可以采用基于序列的检测器探针。在一些实施例中,检测器探针在扩增反应中未杂交到互补序列时被至少部分淬灭,并且在扩增反应中杂交到互补序列时至少部分未淬灭。在一些实施例中,本发明教示的检测器探针的Tm是63-69℃,但是应了解,通过本发明教示的引导,常规实验可以产生具有其他Tm的检测器探针。在一些实施例中,探针可以进一步包含各种修饰,如小沟结合物(参见例如美国专利号6,486,308),用于进一步提供所要热力学特征。

[0102] 在一些实施例中,检测可以基于不同被分析物质之间的迁移速率差异经由多种运动性相关分析技术中的任一种进行。例示性运动性相关分析技术包括电泳、色谱、质谱、沉降(例如梯度离心)、场流分级、多级提取技术等。在一些实施例中,运动性探针可以与扩增产物杂交,且目标核酸序列的身份经由洗脱运动性探针的迁移率依赖性分析技术测定,如例如公开的P.C.T.申请案WO04/46344(罗森布拉姆(Rosenblum)等人)和WO01/92579(文茨(Wenz)等人)所述。在一些实施例中,检测可以通过各种微阵列和相关软件实现,尤其如应用生物系统公司(Applied Biosystems)阵列系统与应用生物系统公司1700化学发光微阵列分析仪和其他商业上可获得的可自昂飞公司(Affymetrix)、安捷伦公司(Agilent)、伊路米那(Illumina)和安玛西亚生物科学公司(Amersham Biosciences)获得阵列系统(也参见格里(Gerry)等人,《分子生物学杂志》(J.Mol.Biol.)292:251-62,1999;德百里斯(De Bellis)等人,《密涅瓦生物技术》(Minerva Biotec)14:247-52,2002;和斯蒂尔斯(Stears)等人,自然·医学《Nat.Med.》9:14045,包括增刊,2003)。还应了解,检测可包含报告基因,其结合到反应产物中,作为经标记引物的一部分或归因于扩增期间经标记dNTP的结合,或例如(但不限于)经由包含报告基因的杂交标签互补序列或经由整体的或连接于反应产物

的连接子臂连接于反应产物。未标记反应产物例如使用质谱的检测也在当前教示内容的范围内。

[0103] “畸变”意指DNA的基因组结构性变异或变化。实例包括:过度表达/表达不足;拷贝数扩增/缺失;突变;基因融合体;等。

[0104] “驱动事件”意指基因组畸变,代表获得功能(GoF)的突变、融合体或拷贝数峰值。

[0105] “再现的”意指事件在3个或更多个肿瘤样品中出现。

[0106] “Mitelman”意指从文献中人工管理的癌症中的染色体畸变和基因融合体的数据库。http://goo.gl/PnXMT

[0107] 基因融合体

[0108] 表1:基因融合体

[0109]

| 癌症类型 | 基因 A 符号 | 基因 B 符号 | 定向 | 可药化基因 |
|---|---|---|---|---|
| 膀胱尿道上皮癌 | ALK | TPM1 | TPM1/ALK | ALK |
| 结肠腺癌 | ALK | PRKAR1A | PRKAR1A/ALK | ALK |
| 皮肤黑素瘤 | ALK | NCOA1 | NCOA1/ALK | ALK |
| 卵巢浆液性囊腺癌 | CASR | LPP | LPP/CASR | CASR |
| 成胶质细胞瘤 | EGFR | MDM2 | MDM2/EGFR | EGFR |
| 低级别胶质瘤 | ELAVL3 | FGFR3 | FGFR3/ELAVL3 | FGFR3 |
| 急性骨髓性白血病 | B2M | GNAS | B2M/GNAS | GNAS |
| 透明细胞肾细胞癌 | DOCK8 | JAK2 | DOCK8/JAK2 | JAK2 |
| 乳头状肾细胞癌 | HNF1B | NOTCH1 | HNF1B/NOTCH1 | NOTCH1 |
| 成胶质细胞瘤 | NFASC | NTRK1 | NFASC/NTRK1 | NTRK1 |
| 甲状腺腺体癌 | NTRK1 | SSBP2 | SSBP2/NTRK1 | NTRK1 |
| 甲状腺腺体癌 | NTRK1 | SQSTM1 | SQSTM1/NTRK1 | NTRK1 |

[0110]

| 癌症类型 | 基因 A 符号 | 基因 B 符号 | 定向 | 可药化基因 |
|---|---|---|---|---|
| 前列腺腺癌 | PIK3CA | TBL1XR1 | TBL1XR1/PIK3CA | PIK3CA |
| 甲状腺腺体癌 | AKAP13 | RET | AKAP13/RET | RET |
| 甲状腺腺体癌 | FKBP15 | RET | FKBP15/RET | RET |
| 甲状腺腺体癌 | RET | TBL1XR1 | TBL1XR1/RET | RET |
| 成胶质细胞瘤 | CEP85L | ROS1 | CEP85L/ROS1 | ROS1 |
| 甲状腺腺体癌 | ALK | GTF2IRD1 | GTF2IRD1/ALK | ALK |
| 卵巢浆液性囊腺癌 | BRS3 | HTATSF1 | HTATSF1/BRS3 | BRS3 |
| 侵袭性乳癌 | CCDC132 | CDH1 | CDH1/CCDC132; CCDC132/CDH1 | CDH1 |
| 侵袭性乳癌 | ERBB2 | SLC29A3 | ERBB2/SLC29A3 | ERBB2 |
| 甲状腺腺体癌 | MET | TFG | MET/TFG; TFG/MET | MET |
| 卵巢浆液性囊腺癌 | MNDA | NOTCH2 | NOTCH2/MNDA | NOTCH2 |
| 甲状腺腺体癌 | IRF2BP2 | NTRK1 | IRF2BP2/NTRK1 | NTRK1 |
| 卵巢浆液性囊腺癌 | EIF2C2 | PTK2 | EIF2C2/PTK2 | PTK2 |
| 侵袭性乳癌 | HOXB3 | RARA | RARA/HOXB3 | RARA |
| 前列腺腺癌 | ETV4 | STAT3 | STAT3/ETV4 | STAT3 |
| 侵袭性乳癌 | C17orf64 | TOP1 | TOP1/C17orf64 | TOP1 |
| 前列腺腺癌 | KIAA0753 | TP53 | TP53/KIAA0753 | TP53 |
| 成胶质细胞瘤 | GFAP | VIM | GFAP/VIM; VIM/GFAP | VIM |
| 甲状腺腺体癌 | LTK | UACA | UACA/LTK | LTK |
| 乳头状肾细胞癌 | ALK | STRN | STRN/ALK | ALK |
| 甲状腺腺体癌 | ALK | STRN | STRN/ALK | ALK |
| 皮肤黑素瘤 | BRAF | CDC27 | CDC27/BRAF | BRAF |
| 甲状腺腺体癌 | BRAF | MACF1 | MACF1/BRAF | BRAF |
| 甲状腺腺体癌 | BRAF | MKRN1 | MKRN1/BRAF | BRAF |
| 皮肤黑素瘤 | BRAF | TAX1BP1 | TAX1BP1/BRAF | BRAF |
| 前列腺腺癌 | BRAF | JHDM1D | JHDM1D/BRAF | BRAF |

[0111] 表2:基因融合体

[0112]

| 癌症类型 | 基因 A 符号 | 基因 B 符号 | 定向 | 可药化基因 |
|---|---|---|---|---|
| 皮肤黑素瘤 | CLCN6 | RAF1 | CLCN6/RAF1 | RAF1 |
| 皮肤黑素瘤 | TRAK1 | RAF1 | TRAK/RAF1 | RAF1 |
| 结肠腺癌 | AKT1 | PRKACA | PRKACA/AKT1 | AKT1 |
| 子宫内膜子宫内膜样腺癌 | AKT1 | PRKACA | PRKACA/AKT1 | AKT1 |
| 结肠腺癌 | AKT2 | PRKACA | PRKACA/AKT2 | AKT2 |
| 肺腺癌 | FYN | MLL | MLL/FYN | FYN |
| 肺腺癌 | ECHD1 | FYN | ECHD1/FYN | FYN |
| 侵袭性乳癌 | JAK2 | TTC13 | TTC13/JAK2 | JAK2 |
| 胃腺癌 | CAB39 | ERBB2 | CAB39/ERBB2 | ERBB2 |
| 子宫内膜子宫内膜样腺癌 | BRAF | EXOC4 | EXOC4/BRAF | BRAF |
| 侵袭性乳癌 | HOOK3 | IKBKB | HOOK3/IKBKB | IKBKB |
| 侵袭性乳癌 | CDK6 | KRIT1 | KRIT1/CDK6 | CDK6 |
| 胃腺癌 | CAPZA2 | MET | CAPZA2/MET | MET |
| 侵袭性乳癌 | ACE | MLLT6 | MLLT6/ACE | ACE |
| 子宫内膜子宫内膜样腺癌 | HLA-C | MUC16 | HLA-C/MUC16 | MUC16 |
| 头颈鳞状细胞癌 | LYN | NTRK3 | LYN/NTRK3 | LYN, NTRK3 |
| 卵巢浆液性囊腺癌 | MUC16 | OR7G2 | MUC16/OR7G2 | MUC16 |
| 卵巢浆液性囊腺癌 | MDK | RAB11B | RAB11B/MDK | MDK |
| 鳞状细胞肺癌 | GADD45GIP1 | RB1 | RB1/GADD45GIP1 | RB1 |

[0113]

| 癌症类型 | 基因 A 符号 | 基因 B 符号 | 定向 | 可药化基因 |
|---|---|---|---|---|
| 胃腺癌 | PRKAR2A | RHOA | PRKAR2A/RHOA | RHOA |
| 皮肤黑素瘤 | MAPK1 | SHANK3 | SHANK3/MAPK1 | MAPK1 |
| 甲状腺腺体癌 | RET | SPECC1L | SPECC1L/RET | RET |
| 卵巢浆液性囊腺癌 | IGFBP2 | SPP1 | IGFBP2/SPP1 | IGFBP2, SPP1 |
| 侵袭性乳癌 | PAPD7 | SRD5A1 | PAPD7/SRD5A1; SRD5A1/ PAPD7 | SRD5A1 |
| 成胶质细胞瘤 | RARA | TAOK1 | TAOK1/RARA | RARA |
| 胃腺癌 | CDK12 | THRA | THRA/CDK12 | THRA |
| 侵袭性乳癌 | NARS2 | TOP1 | NARS2/TOP1 | TOP1 |
| 胃腺癌 | PTK2 | TRAPPC9 | PTK2/TRAPPC9; TRAPPC9/PTK2 | PTK2 |
| 侵袭性乳癌 | CBL | UBE4A | CBL/UBE4A | CBL |
| 低级别胶质瘤 | GFAP | VIM | GFAP/VIM; VIM/GFAP | VIM |
| 侵袭性乳癌 | ADAM9 | WRN | WRN/ADAM9 | ADAM9 |
| 结肠和直肠腺癌 | MAP2K2 | YWHAE | YWHAE/MAP2K2 | MAP2K2 |
| 头颈鳞状细胞癌 | ALK | CLIP4 | CLIP4/ALK | ALK |
| 鳞状细胞肺癌 | ALK | CLIP4 | CLIP4/ALK | ALK |
| 甲状腺腺体癌 | ALK | MEMO1 | MEMO1/ALK | ALK |
| 甲状腺腺体癌 | BRAF | SND1 | BRAF/SND1; SND1/BRAF | BRAF |
| 甲状腺腺体癌 | BRAF | ZC3HAV1 | ZC3HAV1/BRAF | BRAF |

[0114] 表3:基因融合体

[0115]

| 癌症类型 | 基因 A 符号 | 基因 B 符号 | 定向 | 可药化基因 | 癌症类型先例 |
|---|---|---|---|---|---|
| 甲状腺腺体癌 | NOTCH1 | SEC16A | SEC16A-NOTCH1 | NOTCH1 | 乳癌 |
| 侵袭性乳癌 | ERC1 | RET | ERC1-RET | RET | 甲状腺癌 |
| 卵巢浆液性囊腺癌 | CCDC170 | ESR1 | ESR1/CCDC170 | ESR1 | 侵袭性乳癌 |
| 头颈鳞状细胞癌 | RPS6KB1 | VMP1 | RPS6KB1/VMP1; VMP1/RPS6KB1 | RPS6KB1 | 侵袭性乳癌 |
| 肺腺癌 | RPS6KB1 | VMP1 | RPS6KB1/VMP1 | RPS6KB1 | 侵袭性乳癌 |
| 鳞状细胞肺癌 | RPS6KB1 | VMP1 | RPS6KB1/VMP1 | RPS6KB1 | 侵袭性乳癌 |
| 卵巢浆液性囊腺癌 | RPS6KB1 | VMP1 | RPS6KB1/VMP1 | RPS6KB1 | 侵袭性乳癌 |
| 皮肤黑素瘤 | RPS6KB1 | VMP1 | RPS6KB1/VMP1 | RPS6KB1 | 侵袭性乳癌 |
| 胃腺癌 | RPS6KB1 | VMP1 | RPS6KB1/VMP1 | RPS6KB1 | 侵袭性乳癌 |

[0116]

表 4：表 1 的断裂点序列

| 表 4 融合体名称 | 5'基因染色体 | 5' 基因符号 | 5'登录号 | 5'基因断裂点 | 3' 基因染色体 | 3' 基因符号 | 3'登录号 | 3'基因断裂点 | 断裂点序列 | SEQ ID NO: |
|---|---|---|---|---|---|---|---|---|---|---|
| TPM1/ALK | chr15 | TPM1 | NM_000366 | 63,354,844 | chr2 | ALK | NM_004304 | 29446394 | TGCGGAGAGGTCAGTAACTAAATTGGAGAAAAGCATTGATGACTTAGAAG\|TGTACCGCCGGAAGCACCAGGAGCTGCAAGCCATGCAGATGGAGCTGC | 1 |
| PRKAR1/ALK | chr17 | PRKAR1A | NM_002734 | 66,511,717 | chr2 | ALK | NM_004304 | 29446263 | CTGAGAGACCCATGGCATTCCTCAGGGAATACTTTGAGAGGTTGGAGAAG\|ACCTCCTCCATCAGTGACCTGAAGGAGGTGCCGCGGAAAAACATCACC CT | 2 |
| NCOA1/ALK | chr2 | NCOA1 | NM_003743 | 24,991,142 | chr2 | ALK | NM_004304 | 30143047 | GTGCAACAGGTTCAGGTGTTTGCTGACGTCCAGTGTACAGTGAATCTGGT\|AGGCGGCTGTGGGGCTGCTCCAGTTCAATCTCAGCGAGCTGTTCAGTTGG | 3 |
| LPP/CASR | chr3 | LPP | NM_005578 | 188,202,492 | chr3 | CASR | NM_000388 | 121972795 | GAAACTTTCCTCCTCCACCACCTCTTGATGAAGAGGCTTTCAAAGTACAG\|AAGGCATCACAGGAGGCCTCTGCATGATGTGGCTTCCAAAGACTCAAGG A | 4 |
| MDM2/EGFR | chr12 | MDM2 | NM_002392 | 69,203,072 | chr7 | EGFR | NM_005228 | 55231426 | GATGGTGCTGTAACCACCTCACAGATTCCAGCTTCGGAACAAGAGACCCT\|GTGAGCCAAGGGAGTTTGTGGAGAACTCTGAGTGCATACAGTGCCACCC A | 5 |
| FGFR3/ELAVL3 | chr4 | FGFR3 | NM_000142 | 1,808,638 | chr19 | ELAVL3 | NM_001420 | 11577572 | GCCCTCCCAGAGGCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGT\|GTCCTTGGTACAAATGGAGCCACTGACGACAGCAAGACCAACCTCATCG TC | 6 |
| B2M/GNAS | chr15 | B2M | NM_004048 | 45,003,811 | chr20 | GNAS | NM_000516 | 57470667 | TAGCTGTGCTCGCGCTACTCTCTCTTTCTGGCCTGGAGGCTATCCAGCGT\|GCTGGAGAATCTGGTAAAAGCACCATTGTGAAGCAGATGAGGATCCTGC A | 7 |
| DOCK8/JAK2 | chr9 | DOCK8 | NM_203447 | 340,321 | chr9 | JAK2 | NM_004972 | 5050686 | GAGATTTTGGAATTTCCAACACGAGAAGTATATGTCCCTCACACTGTGTA\|CAGTGGCGGCATGATTTTGTGCACGGATGGATAAAAGTACCTGTGACTCA | 8 |
| HNF1B/NOTCH1 | chr17 | HNF1B | NM_000458 | 36,099,431 | chr9 | NOTCH1 | NM_017617 | 139396940 | TGCCGCTCTGTACACCTGGTACGTCAGAAAGCAACGAGAGATCCTCCGAC\|GTGAGACCGTGGAGCCGCCCCGCCGGCGCAGCTGCACTTCATGTACGT G | 9 |
| NFASC/NTRK1 | chr1 | NFASC | NM_015090 | 204,951,148 | chr1 | NTRK1 | NM_002529 | 156844363 | GGGAAGGGCCCTGAGCCAGAGTCCGTCATCGGTTACTCCGGAGAAGATT A\|CACTAACAGCACATCTGGAGACCCGGTGGAGAAGGAGGACGAAACAC CTT | 10 |
| SSBP2/NTRK1 | chr5 | SSBP2 | NM_012446 | 80,742,687 | chr1 | NTRK1 | NM_002529 | 156845312 | TCCAGGAGGTGGAGGGCCACCAGGAACACCCATCATGCCTAGTCCAGCA G\|GCCCGGCTGTGCTGGCTCCAGAGGATGGGCTGGCCATGTCCCTGCATTT C | 11 |
| SQSTM1/NTRK1 | chr5 | SQSTM1 | NM_003900 | 179,252,226 | chr1 | NTRK1 | NM_002529 | 156844363 | TTTCCTGAAGAACGTTGGGGAGAGTGTGGCAGCTGCCCTTAGCCCTCTGG \|ACACTAACAGCACATCTGGAGACCCGGTGGAGAAGGAGGACGAAACAC CT | 12 |
| TBL1XR1/PIK3CA | chr3 | TBL1XR1 | NM_024665 | 176,914,909 | chr3 | PIK3CA | NM_006218 | 178916538 | CATATAAAACTACTTTAAGGAATTAGATGTATGGTTGTCCCAAAGCAGAA\|ACCTGGAAACGGTGGCCTCCAACGCCGCTCCCCCCTCCGGGAATGGAG G | 13 |
| AKAP13/RET | chr15 | AKAP13 | NM_006738 | 86,286,839 | chr10 | RET | NM_020630 | 43612067 | CGCCATCTGCACCTTCCATAGCCAAATCAGGGTCATTGGACTCAGAACTT\|GGTTCTTGGAAAAACTCTAGGAGAAGGCGAATTTGGAAAAGTGGTCAAG | 14 |

[0117]

| 表 4 融合体名称 | 5'基因染色体 | 5' 基因符号 | 5'登录号 | 5'基因断裂点 | 3' 基因染色体 | 3' 基因符号 | 3'登录号 | 3'基因断裂点 | 断裂点序列 | SEQ ID NO: |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | G | |
| FKBP15/RET | chr9 | FKBP15 | NM_015258 | 115,932,802 | chr10 | RET | NM_020630 | 43612032 | AATCTTACAATGGCAGGACCATTCTGGGAACCATCATGAATACGATCAAG\|GAGGATCCAAAGTGGGAATTCCCTCGGAAGAACTTGGTTCTTGGAAAAAC | 15 |
| TBL1XR1/RET | chr3 | TBL1XR1 | NM_024665 | 176,765,103 | chr10 | RET | NM_020630 | 43610136 | GCCCTATATTTGCATTAAAATGGAATAAGAAAGGAAATTTCATCCTAAGT\|GCTGGACTCCATGGAGAACCAGGTCTCCGTGGATGCCTTCAAGATCCTGG | 16 |
| CEP85L/ROS1 | chr6 | CEP85L | 387119 | 118,802,942 | chr6 | ROS1 | NM_002944 | 117641193 | TTAATATGCCAGAAAAAGAAAGGAGTTAGTAACTACCGTTCAGAG\|TACTCTTCCAACCCAAGAGGAGATTGAAAATCTTCCTGCCTTCCCTCGG | 17 |
| CCDC132/CDH1 | CCDC132 | NM_017667 | chr7 | 92,940,584 | CDH1 | NM_004360 | chr16 | 68,857,494 | GAATGCACCTATCTTAACAAATACAACATTGAACGTCATAAGACTTGTTG\|TTCTGGGGATTCTTGGAGGAATTCTTGCTTTGCTAATTCTGATTCTGCTG | 30 |
| CDH1/CCDC132 | CDH1 | NM_004360 | chr16 | 68,857,529 | CCDC132 | NM_017667 | chr7 | 92,952,923 | AACATCAAAGGCAATTGGCTTAAGAATGTTCATCATCTGCATATATTTTC\|TTAGCAAAGCAAGAATTCCTCCAAGAATCCCCAGAATGGCAGGAATTTGC | 31 |
| CDH1/CCDC132 | CDH1 | NM_004360 | chr16 | 68,857,529 | CCDC132 | NM_017667 | chr7 | 92,952,923 | GCAAATTCCTGCCATTCTGGGGATTCTTGGAGGAATTCTTGCTTTGCTAA\|GAAAATATATGCAGATGATGAACATTCTTAAGCCAATTGCCTTTGATGTT | 32 |
| EIF2C2/PTK2 | EIF2C2 | NM_012154 | chr8 | 141,645,584 | PTK2 | NM_005607 | chr8 | 141,685,598 | GCTGCAGGATCTGGTTTACCCACAGGCTGATATATATGTTGGTTTCCAAT\|CGGGGCCGGCTCCCGAGTACATGGTGGCGCCGCCGAGGGGCTCCGGGGCC | 33 |
| EIF2C2/PTK2 | EIF2C2 | NM_012154 | chr8 | 141,645,584 | PTK2 | NM_005607 | chr8 | 141,685,598 | GGCCCCGGAGCCCCTCGGCGGCGCCACCATGTACTCGGGAGCCGGCCCC\|ATTGGAAACCAACATATATATCAGCCTGTGGGTAAACCAGATCCTGCAG | 34 |
| EIF2C2/PTK2 | EIF2C2 | NM_012154 | chr8 | 141,645,584 | PTK2 | NM_005607 | chr8 | 141,712,806 | CCCCGGAGCCCCTCGGCGGCGCCACCATGTACTCGGGAGCCGGCCCCGG\|TTCTGGCTACCCTGGTTCACATGGAATCACAGCCATGGCTGGCAGCATC | 35 |
| EIF2C2/PTK2 | EIF2C2 | NM_012154 | chr8 | 141,645,584 | PTK2 | NM_005607 | chr8 | 141,762,415 | CGAAGTACAGTTTTTACATGTTTTAATTGCAACCGCCAAAGCTGGATTCT\|CCGGGGCCGGCTCCCGAGTACATGGTGGCGCCGCCGAGGGGCTCCGGGG | 36 |
| EIF2C2/PTK2 | EIF2C2 | NM_012154 | chr8 | 141,645,584 | PTK2 | NM_005607 | chr8 | 141,675,096 | GGCCCCGGAGCCCCTCGGCGGCGCCACCATGTACTCGGGAGCCGGCCCC\|GAAGTCGGCTTGGCCCTGAGGACATTATTGGCCACTGTGGATGAGACCA | 37 |
| ERBB2/SLC29A3 | ERBB2 | NM_004448 | chr17 | 37,883,211 | SLC29A3 | NM_018344 | chr10 | 73,115,986 | ACACATGGGCCGCAAGAACAGGCCTCATGTAGTACCTGGCATACTCCAGC\|GCCCGGGGCAGGGTCTGGACAGAAGAAGCCCTGCTGGGGTACCAGATAC | 38 |
| ERBB2/SLC29A3 | ERBB2 | NM_004448 | chr17 | 37,883,548 | SLC29A3 | NM_018344 | chr10 | 73,121,774 | GGGCGCTGGGGGGCATGGTCCACCACAGGCACCGCAGCTCATCTACCAGG\|CTCCCTCAGTGCCCCTTCGGTGGCCTCCAGATTCATTGATTCCCACACA | 39 |
| ERBB2/SLC29A3 | ERBB2 | NM_004448 | chr17 | 37,883,598 | SLC29A3 | NM_018344 | chr10 | 73,121,726 | GTGGCGGTGGGGACCTGACACTAGGGCTGGAGCCCTCTGAAGAGGAGGGC\|TGTTCTTGCGGCCCATGTGTTTTCTGGTGAAGAGGAGCTTCCCCAGGAC | 40 |

[0118]

| 表 4 融合体名称 | 5'基因染色体 | 5' 基因符号 | 5'登录号 | 5'基因断裂点 | 3' 基因染色体 | 3' 基因符号 | 3'登录号 | 3'基因断裂点 | 断裂点序列 | SEQ ID NO: |
|---|---|---|---|---|---|---|---|---|---|---|
| ERBB2/SLC29A3 | ERBB2 | NM_004448 | chr17 | 37,883,205 | SLC29A3 | NM_018344 | chr10 | 73,115,911 | CTGAGGAGTATCTGGTACCCCAGCAGGCTTCTTCTGTCCAGACCCTGCC\|CAGCGCCCTGGCCTTCTTCCTGACGGCCACTGTCTTCCTCGTGCTCTGCA | 41 |
| ERBB2/SLC29A3 | ERBB2 | NM_004448 | chr17 | 37,882,078 | SLC29A3 | NM_018344 | chr10 | 73,115,911 | TGCAGAGCACGAGGAAGACAGTGGCCGTCAGGAAGAAGGCCAGGGCGC\|TG\|GTGCAGATGGGGGGCTGGGGCAGCCGCTCCCCCTTTTCCAGCAGGT CAG | 42 |
| GFAP/VIM | GFAP | NM_002055 | chr17 | 42,987,987 | VIM | NM_003380 | chr10 | 17,277,377 | AGGAGAACCGGATCACCATTCCCGTGCAGACCTTCTCCAACCTGCAGATT\|CGAGGAGAGCAGGATTCTCTGCCTCTTCCAAACTTTTCCTCCCTGAACCT | 43 |
| GFAP/VIM | GFAP | NM_002055 | chr17 | 42,988,732 | VIM | NM_003380 | chr10 | 17,277,285 | ACGTGCGGGAGGCGGCCAGTTATCAGGAGGCGCTGGCGCGGCTGGAGGA A\|ATGGCTCGTCACCTTCGTGAATACCAAGACCTGCTCAATGTTAAGATGG C | 44 |
| GFAP/VIM | GFAP | NM_002055 | chr17 | 42,987,987 | VIM | NM_003380 | chr10 | 17,277,377 | AGGAGAACCGGATCACCATTCCCGTGCAGACCTTCTCCAACCTGCAGATT\|CGAGGAGAGCAGGATTTCTCTGCCTCTTCCAAACTTTTCCTCCCTGAACC | 45 |
| GFAP/VIM | GFAP | NM_002055 | chr17 | 42,988,622 | VIM | NM_003380 | chr10 | 17,277,371 | AATGTCAAGCTGGCCCTGGACATCGAGATCGCCACCTACAGGAAGCTGCT \|GGAAGGCGAGGAGAGCAGGATTTCTCTGCCTCTTCCAAACTTTTCCTCCC | 46 |
| GFAP/VIM | GFAP | NM_002055 | chr17 | 42,985,511 | VIM | NM_003380 | chr10 | 17,277,237 | ATCACCATTCCCGTGCAGACCTTCTCCAACCTGCAGATTCGAGAAACCAG\|GACACTATTGGCCGCCTGCAGGATGAGATTCAGAATATGAAGGAGGAAAT | 47 |
| GFAP/VIM | GFAP | NM_001131019 | chr17 | 42,987,602 | VIM | NM_003380 | chr10 | 17,277,286 | CTTCTCCAACCTGCAGATTCGAGGGGGCAAAAGCACCAAAGACGGGGAA A\|TGGCTCGTCACCTTCGTGAATACCAAGACCTGCTCAATGTTAAGATGGC C | 48 |
| GFAP/VIM | GFAP | NM_002055 | chr17 | 42,987,983 | VIM | NM_003380 | chr10 | 17,278,298 | GAACCGGATCACCATTCCCGTGCAGACCTTCTCCAACCTGCAGATTCGAG\|AATCTGGATTCACTCCCTCTGGTTGATACCCCACTCAAAAAGGACACTTCT | 49 |
| GFAP/VIM | GFAP | NM_002055 | chr17 | 42,992,594 | VIM | NM_003380 | chr10 | 17,271,785 | CAGAGATGATGGAGCTCAATGACCGCTTTGCCAGCTACATCGAGAAGGTT\|CGCTTCCTGGAGCAGCAGAATAAGATCCTGCTGGCCGAGCTCGAGCAGC T | 50 |
| GFAP/VIM | GFAP | NM_002055 | chr17 | 42,985,469 | VIM | NM_003380 | chr10 | 17,277,285 | GAAACCAGCCTGGACACCAAGTCTGTGTCAGAAGGCCACCTCAAGAGGA A\|ATGGCTCGTCACCTTCGTGAATACCAAGACCTGCTCAATGTTAAGATGG C | 51 |
| GFAP/VIM | GFAP | NM_002055 | chr17 | 42,988,779 | VIM | NM_003380 | chr10 | 17,277,168 | CACGAACGAGTCCCTGGAGAGGCAGATGCGCGAGCAGGAGGAGCGGCA CG\|AATGAGTCCCTGGAACGCCAGATGCGTGAAATGGAAGAGAACTTTGC CGT | 52 |
| GFAP/VIM | GFAP | NM_002055 | chr17 | 42,988,637 | VIM | NM_003380 | chr10 | 17,277,351 | GGCAGAGAAATCCTGCTCTCCTCGCCTTCCAGCAGCTTCCTGTAGGTGGC\| GTGGCGATCTCGATGTCCAGGGCCAGCTTGACATTGAGCAGGTCCTGGTA | 53 |
| GFAP/VIM | GFAP | NM_002055 | chr17 | 42,992,627 | VIM | NM_003380 | chr10 | 17,271,752 | CTGGCTTCAAGGAGACCCGGCCAGTGAGCGGGCAGAGATGATGGAGCT C\|AATGACCGCTTCGCCAACTACATCGACAAGGTGCGCTTCCTGGAGCAG CA | 54 |
| GFAP/VIM | GFAP | NM_002055 | chr17 | 42,988,742 | VIM | NM_003380 | chr10 | 17,277,351 | GAGGAGCGGCACGTGCGGGAGGCGGCCAGTTATCAGGAGGCGCTGGCG CG\|GCCCACCTACAGGAAGCTGCTGGAAGGCGAGGAGAGCAGGATTTCTCT GCC | 55 |
| GFAP/VIM | GFAP | NM_002055 | chr17 | 42,988,025 | VIM | NM_003380 | chr10 | 17,276,771 | CATCGAGATCGCCACCTACAGGAAGCTGCTAGAGGGCGAGGAGAACCGG A\|GACAGGTGCAGTCCCTCACCTGTGAAGTGGATGCCCTTAAGGAACCA | 56 |

[0119]

| 表4 融合体名称 | 5'基因染色体 | 5' 基因符号 | 5'登录号 | 5'基因断裂点 | 3' 基因染色体 | 3' 基因符号 | 3'登录号 | 3'基因断裂点 | 断裂点序列 | SEQ ID NO: |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | AT | |
| GFAP/VIM | GFAP | NM_002055 | chr17 | 42,988,742 | VIM | NM_003380 | chr10 | 17,277,367 | GAGGAGCGGCACGTGCGGGAGGCGGCCAGTTATCAGGAGGCGCTGGCG CG\|TGCTGGAAGGCGAGGAGAGCAGGATTTCTCTGCCTCTTCCAAACTTT TCC | 57 |
| GFAP/VIM | GFAP | NM_002055 | chr17 | 42,988,642 | VIM | NM_003380 | chr10 | 17,277,351 | GGCAGAGAAATCCTGCTCTCCTCGCCTTCCAGCAGCTTCCTGTAGGTGGC\| GATCTCGATGTCCAGGGCCAGCTTGACATTGAGCAGGTCCTGGTACTCCT | 58 |
| GFAP/VIM | GFAP | NM_002055 | chr17 | 42,988,642 | VIM | NM_003380 | chr10 | 17,277,351 | AGGAGTACCAGGACCTGCTCAATGTCAAGCTGGCCCTGGACATCGAGATC \|GCCACCTACAGGAAGCTGCTGGAAGGCGAGGAGCAGGATTTCTCTGC C | 59 |
| GFAP/VIM | GFAP | NM_002055 | chr17 | 42,992,612 | VIM | NM_003380 | chr10 | 17,271,824 | CCCGGGCCAGTGAGCGGGCAGAGATGATGGAGCTCAATGACCGCTTTGC C\|CTCGAGCAGCTCAAGGGCCAAGGCAAGTCGCGCCTGGGGGACCTCTAC GA | 60 |
| GFAP/VIM | GFAP | NM_002055 | chr17 | 42,992,610 | VIM | NM_003380 | chr10 | 17,271,769 | CGGGCCAGTGAGCGGGCAGAGATGATGGAGCTCAATGACCGCTTTGCCA G\|CTACATCGACAAGGTGCGCTTCCTGGAGCAGCAGAATAAGATCCTGCT GG | 61 |
| GTF2IRD1/ALK | GTF2IRD1 | NM_005685 | chr7 | 73,935,627 | ALK | NM_004304 | chr2 | 29,446,394 | ACGTCCATGCCTCCAAGCGCATTCTCTTCTCCATCGTCCATGACAAGTCA\| GTGTACCGCCGGAAGCACCAGGAGCTGCAAGCCATGCAGATGGAGCTGC A | 62 |
| HTATSF1/BRS3 | HTATSF1 | NM_014500 | chrX | 135,586,622 | BRS3 | NM_001727 | chrX | 135,572,292 | CCATGAGCGAGTTGTCATCATCAAGAATATGTTTCATCCTATGGATTTTG\|A GATACAAGGCAGTTGTGAAGCCACTTGAGCGACAGCCCTCCAATGCCAT | 63 |
| IRF2BP2/NTRK1 | IRF2BP2 | NM_182972 | chr1 | 234,744,241 | NTRK1 | NM_002529 | chr1 | 156,844,363 | CTCGGGGCCCTTCGAGAGCAAGTTTAAGAAGGAGCCGGCCCTGACTGCA G\|ACACTAACAGCACATCTGGAGACCCGGTGGAGAAGGAGGACGAAACA CCT | 64 |
| IRF2BP2/NTRK1 | IRF2BP2 | NM_182972 | chr1 | 234,744,241 | NTRK1 | NM_002529 | chr1 | 156,844,363 | AGGTGTTTCGTCCTTCTTCTCCACCGGGTCTCCAGATGTGCTGTTAGTGT\|C TGCAGTCAGGGCCGGCTCCTTCTTAAACTTGCTCTCGAAGGGCCCCGAG | 65 |
| MET/TFG | MET | NM_000245 | chr7 | 116,412,043 | TFG | NM_006070 | chr3 | 100,455,420 | AGAAATGGTTTCAAATGAATCTGTAGACTACCGAGCTACTTTTCCAGAAG\| GGCCACCCAGTGCTCCTGCAGAAGATCGTTCAGGAACACCCGACAGCAT T | 66 |
| MET/TFG | MET | NM_000245 | chr7 | 116,412,013 | TFG | NM_006070 | chr3 | 100,455,435 | TGTAAGTGCCCGAAGTGTAAGCCCAACTACAGAAATGGTTTCAAATGAAT\| CTGCAGAAGATCGTTCAGGAACACCCGACAGCATTGCTTCCTCCTCCTCA | 67 |
| MET/TFG | MET | NM_000245 | chr7 | 116,414,937 | TFG | NM_006070 | chr3 | 100,455,447 | AATGGTTTCAAATGAATCTGTAGACTACCGAGCTACTTTTCCAGAAGATC\| GTTCAGGAACACCCGACAGCATTGCTTCCTCCTCCTCAGCAGCTCACCCA | 68 |
| MET/TFG | MET | NM_000245 | chr7 | 116,415,078 | TFG | NM_006070 | chr3 | 100,455,435 | TATATCCAGTCCATTACTGCAAAATACTGTCCACATTGACCTCAGTGCTC\|C TGCAGAAGATCGTTCAGGAACACCCGACAGCATTGCTTCCTCCTCCTCA | 69 |
| NOTCH2/MNDA | NOTCH2 | NM_024408 | chr1 | 120,478,095 | MNDA | NM_002432 | chr1 | 158,815,377 | TATTGACCTTGTGAACCATTTCAAGTGCTCTTGCCCACCAGGCACTCGGG\| AATCAGGAAACCCAGGCCCAACGGCAGGTGGATGCAAGAAGAAATGTTC C | 70 |
| NOTCH2/MNDA | NOTCH2 | NM_024408 | chr1 | 120,478,095 | MNDA | NM_002432 | chr1 | 158,815,377 | GTATTGACCTTGTGAACCATTTCAAGTGCTCTTGCCCACCAGGCACTCGG\| GAATCAGGAAACCCAGGCCCAACGGCAGGTGGATGCAAGAAGAAATGTT | 71 |

[0120]

| 表4 融合体名称 | 5'基因染色体 | 5' 基因符号 | 5'登录号 | 5'基因断裂点 | 3' 基因染色体 | 3' 基因符号 | 3'登录号 | 3'基因断裂点 | 断裂点序列 | SEQ ID NO: |
|---|---|---|---|---|---|---|---|---|---|---|
| DA | | | | | | | | | C | |
| RARA/HOXB3 | RARA | NM_000964 | chr17 | 38,508,759 | HOXB3 | NM_002146 | chr17 | 46,632,980 | CCATCGCCGACCAGATCACCCTCCTCAAGGCTGCCTGCCTGGACATCCTG\|GAGGGGAGATTTGTCGCCTGCCGCTCGCTCTGGGGCTCGATGTGAAGATA | 72 |
| STAT3/ETV4 | STAT3 | NM_003150 | chr17 | 40,468,807 | ETV4 | NM_001986 | chr17 | 41,611,353 | GTTTGGAAATAATGGTGAAGGTGCTGAACCCTCAGCAGGAGGGCAGTTT G\|TAGCTTTCCACAGCCCCACCACCAGGATCAAGAAGGAGCCCCAGAGTC CC | 73 |
| STAT3/ETV4 | STAT3 | NM_003150 | chr17 | 40,468,860 | ETV4 | NM_001986 | chr17 | 41,613,825 | AGCAATACCATTGACCTGCCGATGTCCCCCCGCACTTTAGATTCATTGAT\|G CAGTTTGTTCCTGATTTCCATTCAGAAAACCTAGCTTTCCACAGCCCCA | 74 |
| STAT3/ETV4 | STAT3 | NM_003150 | chr17 | 40,468,846 | ETV4 | NM_001986 | chr17 | 41,610,042 | CCTGCCGATGTCCCCCCGCACTTTAGATTCATTGATGCAGTTTGGAAATA\|G ATGTCACCGGGTGCGCATCAATGTACCTCCACACAGAGGCTTCTCTGG | 75 |
| TFG/MET | TFG | NM_006070 | chr3 | 100,451,516 | MET | NM_000245 | chr7 | 116,414,935 | ATCAATAAAAAATGTTATGTCAGCGTTTGGCTTAACAGATGATCAGGTTTC\|A GATCAGTTTCCTAATTCATCTCCAGAACGGTTCATGCCGACAAGTGCAGT | 76 |
| TOP1/C17orf64 | TOP1 | NM_003286 | chr20 | 39,729,993 | C17orf64 | NM_181707 | chr17 | 58,503,144 | CATCCAAGGTTCCATTAAATACATCATGCTTAACCCTAGTTCACGAATCA\|A GGTGACAAATGTGTCATGCCTGGAGACAAGCTCCAGCGCCAGCCCTGCT | 77 |
| TOP1/C17orf64 | TOP1 | NM_003286 | chr20 | 39,729,993 | C17orf64 | NM_181707 | chr17 | 58,503,144 | CCAAGGTTCCATTAAATACATCATGCTTAACCCTAGTTCACGAATCAAGGT GACAAATGTGTCATGCCTGGAGACAAGCTCCAGCGCCAGCCCTGCTAGA | 78 |
| TOP1/C17orf64 | TOP1 | NM_003286 | chr20 | 39,728,797 | C17orf64 | NM_181707 | chr17 | 58,503,167 | TGGCATGGCGCATGAGCGAGTCTCTAGCAGGGCTGGCGCTGGAGCTTGTC \|TCCAGGAGGCTCTATCTTGAAGTTAGCAATCCTCTCTCTTTGTGGTTATCCA | 79 |
| TP53/IAA0753 | TP53 | NM_000546 | chr17 | 7,590,695 | KIAA0753 | NM_014804 | chr17 | 6,498,373 | TCAGCATATGCGATTTTATTATTATCTTTGACGAACAGACTCCTGGTATTT\|CC AATCCAGGGAAGCGTGTCACCGTCGTGGAAAGCACGCTCCCAGCCCGA | 80 |
| TP53/IAA0753 | TP53 | NM_000546 | chr17 | 7,579,529 | KIAA0753 | NM_014804 | chr17 | 6,493,323 | TCCCAAGCAATGGATGATTTGATGCTGTCCCCGGACGATATTGAACAATG\| TTCCCTGGATGAAAGTGTGGGAACAGAGGAAGGATCAGAGAAAAGAGA GG | 81 |
| TP53/IAA0753 | TP53 | NM_000546 | chr17 | 7,590,695 | KIAA0753 | NM_014804 | chr17 | 6,498,373 | TTCGGGCTGGGAGCGTGCTTTCCACGACGGTGACACGCTTCCCTGGATTG \|GAAATACCAGGAGTCTGTTCGTCAAAGATATATAAAATCGCATATGCTG | 82 |
| VIM/GFAP | VIM | NM_003380 | chr10 | 17,277,255 | GFAP | NM_002055 | chr17 | 42,987,987 | GAACTTTGCCGTTGAAGCTGCTAACTACCAAGACACTATTGGCCGCCTGC\| TCGAGAAACCAGCCTGGACACCAAGTCTGTGTCAGAAGGCCACCTCAAG A | 83 |
| VIM/GFAP | VIM | NM_003380 | chr10 | 17,277,325 | GFAP | NM_002055 | chr17 | 42,988,666 | AAGGAGGAAATGGCTCGTCACCTTCGTGAATACCAAGACCTGCTCAATGT \|CAAGCTGGCCCTGGACATCGAGATCGCCACCTACAGGAAGCTGCTAGAG G | 84 |
| VIM/GFAP | VIM | NM_003380 | chr10 | 17,277,255 | GFAP | NM_002055 | chr17 | 42,987,987 | TTGAAGCTGCTAACTACCAAGACACTATTGGCCGCCTGC\|TCGAGAAACC AGCCTGGACACCAAGTCTGTGTCAGAAGGCCACCTCAAGA | 85 |
| VIM/GFAP | VIM | NM_003380 | chr10 | 17,277,370 | GFAP | NM_002055 | chr17 | 42,988,621 | AATGTTAAGATGGCCCTTGACATTGAGATTGCCACCTACAGGAAGCTGCT\| AGAGGGCGAGGAGAACCGGATCACCATTCCCGTGCAGACCTTCTCCAAC C | 86 |
| VIM/GFAP | VIM | NM_003380 | chr10 | 17,271,830 | GFAP | NM_002055 | chr17 | 42,992,688 | GGTGCGCTTCCTGGAGCAGCAGAATAAGATCCTGCTGGCCGAGCTCGAG C\|GGCACTCAATGCTGGCTTCAAGGAGACCCGGGCCAGTGAGCGGGCA GAG | 87 |

[0121]

| 表 4 融合体名称 | 5'基因染色体 | 5' 基因符号 | 5'登录号 | 5'基因断裂点 | 5'基因符号 | 5'登录号 | 5' 基因断裂点 | 3' 基因染色体 | 3' 基因符号 | 3'登录号 | 3'基因断裂点 | 断裂点序列 | SEQ ID NO: |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| VIM/GFAP | VIM | NM_003380 | chr10 | 17,277,350 | GFAP | NM_002055 | chr17 | 42,988,641 | GTGAATACCAAGACCTGCTCAATGTTAAGATGGCCCTTGACATTGAGATT\|GCCACCTACAGGAAGCTGCTAGAGGGCGAGGAGAACCGGATCACCATTCC | 88 |
| VIM/GFAP | VIM | NM_003380 | chr10 | 17,277,877 | GFAP | NM_002055 | chr17 | 42,988,655 | GAAGGCGAGGAGAGCAGGATTTCTCTGCCTCTTCCAAACTTTTCCTCCCT\|TGGACATCGAGATCGCCACCTACAGGAAGCTGCTAGAGGGCGAGGAGAA C | 89 |
| VIM/GFAP | VIM | NM_003380 | chr10 | 17,277,877 | GFAP | NM_002055 | chr17 | 42,988,655 | GGAAGGCGAGGAGAGCAGGATTCTCTGCCTCTTCCAAACTTTTCCTCCCT\|TGGACATCGAGATCGCCACCTACAGGAAGCTGCTAGAGGGCGAGGAGAA C | 90 |
| UACA/LTK | chr15 | UACA | NM_018003 | 70,957,001 | chr15 | LTK | NM_002344 | 41799372 | TGATTGACACTCTGCAGCACCAAGTGAAATCTCTGGAGCAACAGCTGGC C\|GTGGGGCTTGGCCCGGCCCAGTCCTGGCCTCTGCCACCAGGTGTCACC GA | 184 |
| STRN/ALK | chr2 | STRN | NM_003162 | 37,143,221 | chr2 | ALK | NM_004304 | 29446394 | TACGGGACAGAATTGAATCAGGGAGATATGAAGCCTCCAAGCTATGATTC\|TGTGTACCGCCGGAAGCACCAGGAGCTGCAAGCCATGCAGATGGAGCTG C | 185 |
| STRN/ALK | chr2 | STRN | NM_003162 | 37,143,221 | chr2 | ALK | NM_004304 | 29446394 | TACGGGACAGAATTGAATCAGGGAGATATGAAGCCTCCAAGCTATGATTC\|TGTGTACCGCCGGAAGCACCAGGAGCTGCAAGCCATGCAGATGGAGCTG C | 186 |
| JHDM1D/BRAF | chr7 | JHDM1D | NM_030647 | 139,810,895 | chr7 | BRAF | NM_004333 | 140481493 | TAGACCTGGACACCTTATTAAAGAACTTTCTAAAGTAATTCGAGCAATAG\|AGAAAACACTTGGTAGACGGGACTCGAGTGATGATTGGGAGATTCCTGAT | 187 |
| JHDM1D/BRAF | chr7 | JHDM1D | NM_030647 | 139,810,895 | chr7 | BRAF | NM_004333 | 140481493 | GACCTGGACACCTTATTAAAGAACTTTCTAAAGTAATTCGAGCAATAGAG\|AAAACACTTGGTAGACGGGACTCGAGTGATGATTGGGAGATTCCTGATGG | 188 |
| TAX1BP1/BRAF | chr7 | TAX1BP1 | NM_006024 | 27,827,222 | chr7 | BRAF | NM_004333 | 140481493 | CTGAAAAGGAAAATCTGCAAAGAACTTTCCTGCTTACAACCTCAAGTAA A\|AAAACACTTGGTAGACGGGACTCGAGTGATGATTGGGAGATTCCTGAT GG | 189 |
| MKRN1/BRAF | chr7 | MKRN1 | NM_013446 | 140,158,807 | chr7 | BRAF | NM_004333 | 140487384 | TGCAGGTCCTGCATCCAATGGATGCTGCCCAGAGATCGCAGCATATCAAA\|GACTTGATTAGAGACCAAGGATTTCGTGGTGATGGAGGATCAACCACAGG | 190 |
| MACF1/BRAF | chr1 | MACF1 | NM_012090 | 39,896,580 | chr7 | BRAF | NM_004333 | 140487384 | TTGGACAAAGGGTGGATGAAATTGATGCTGCTATTCAGAGATCACAACAG\|GACTTGATTAGAGACCAAGGATTTCGTGGTGATGGAGGATCAACCACAGG | 191 |
| CDC27/BRAF | chr17 | CDC27 | NM_001256 | 45,206,816 | chr7 | BRAF | NM_004333 | 140487365 | CAGAGAAGGCTTTGGATACCCTAAACAAAGCCATTGTCATTGATCCCAAG\|GATTTCGTGGTGATGGAGGATCAACCACAGGTTTGTCTGCTACCCCCCCT | 192 |

表 5：表 2 的断裂点序列

| 表 5 癌症 | 融合体名称 | 5'基因染色体 | 5'基因符号 | 5'登录号 | 5' 基因断裂点 | 3' 基因染色体 | 3'基因 | 3'登录号 | 3'基因断裂点 | 断裂点序列 |
|---|---|---|---|---|---|---|---|---|---|---|
| 黑素瘤 | CLCN6/ RAF1 | chr1 | CLCN6 | NM_001286 | 11867247 | chr3 | RAF1 | NM_002880 | 126419114 | GAGAAACACAGGAGAGGAGGATGAGATTCTTCCAAGGAAAGACT ATGAG\|GATGCAATTCGAAGTCACAGCGAATCAGCCTCACCTTCAGC |

[0122]

| 表5 癌症 | 融合体名称 | 5'基因染色体 | 5'基因符号 | 5'登录号 | 5'基因断裂点 | 3'基因染色体 | 3'基因 | 3'登录号 | 3'基因断裂点 | 断裂点序列 |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | CCTGTCCAG SEQ ID NO: 18 |
| 黑素瘤 | TRAK1/ RAF1 | chr3 | TRAK1 | NM_014965 | 42235390 | chr3 | RAF1 | NM_002880 | 12641914 | TCCAGCATCTGGGGGCTGCTAAGGATGCCCAGCGGCAGCTCACAGC CGAG\|GATGCAATTCGAAGTCACAGCGAATCAGCCTCACCTTCAGC CCTGTCCAG SEQ ID NO: 19 |
| 结肠腺癌 | PRKACA/AKT1 | chr19 | PRKACA | NM_002730 | 14208406 | chr14 | AKT1 | NM_005163 | 1.05E+08 | AGGGCCGCACTTGGACCTTGTGCGGCACCCCTGAGTACCTGGCCCC TGAG\|GTGCTGGAGGACAATGACTACGGCCGTGCAGTGGACTGGTG GGGGCTGGG SEQ ID NO: 20 |
| 结肠腺癌 | PRKACA/AKT1 | chr19 | PRKACA | NM_002730 | 14208406 | chr14 | AKT1 | NM_005163 | 1.05E+08 | AGGGCCGCACTTGGACCTTGTGCGGCACCCCTGAGTACCTGGCCCC TGAG\|GTGCTGGAGGACAATGACTACGGCCGTGCAGTGGACTGGTG GGGGCTGGG SEQ ID NO: 21 |
| 结肠腺癌 | PRKACA/AKT1 | chr19 | PRKACA | NM_002730 | 14208406 | chr14 | AKT1 | NM_005163 | 1.05E+08 | AGGGCCGCACTTGGACCTTGTGCGGCACCCCTGAGTACCTGGCCCC TGAG\|GTGCTGGAGGACAATGACTACGGCCGTGCAGTGGACTGGTG GGGGCTGGG SEQ ID NO: 22 |
| 子宫内膜异位症 | PRKACA/AKT1 | chr19 | PRKACA | NM_002730 | 14208406 | chr14 | AKT1 | NM_005163 | 1.05E+08 | AGGGCCGCACTTGGACCTTGTGCGGCACCCCTGAGTACCTGGCCCC TGAG\|GTGCTGGAGGACAATGACTACGGCCGTGCAGTGGACTGGTG GGGGCTGGG SEQ ID NO: 23 |
| 结肠腺癌 | PRKACA/AKT2 | chr19 | PRKACA | NM_002730 | 14208406 | chr19 | AKT2 | NM_001626 | 40742011 | AGGGCCGCACTTGGACCTTGTGCGGCACCCCTGAGTACCTGGCCCC TGAG\|GTGCTGGAGGACAATGACTATGGCCGGGCCGTGGACTGGTG GGGGCTGGG SEQ ID NO: 24 |
| 肺腺癌 | MLL/FYN | chr11 | MLL | NM_005933 | 1.18E+08 | chr6 | FYN | NM_002037 | 1.12E+08 | CCAGGAAGCTCGATCAAATGCCCGCCTAAAGCAGCTCTCATTTGCA GGTG\|GTACTTTGGAAAACTTGGCCGAAAAGATGCTGAGCGACAGC TATTGTCCT SEQ ID NO: 25 |
| 肺腺癌 | ECHDC1/FYN | chr6 | ECHDC1 | NM_001002030 | 1.28E+08 | chr6 | FYN | NM_002037 | 1.12E+08 | CAAGGTTGGGCATTGGGTGGAGGAGCAGAATTTACTACAGCATGTG ATTT\|CAGGGAAGGAGATTGGTGGGAAGCCCGCTCCTTGACAACTG GAGAGACAG SEQ ID NO: 26 |
| 乳癌 | TTC13/JAK2 | chr1 | TTC13 | NM_024525 | 2.31E+08 | chr9 | JAK2 | NM_004972 | 5055786 | CTTCATATCAGAGGACTATGCAACAGCCCATGAAGACTTTCAGCAG TCCT\|CTGGAAATTGAACTTAGCTCATTAAGGGAAGCTTTGTCTTTC GTGTCATT SEQ ID NO: 27 |
| 胃腺癌 | CAB39/ERBB2 | chr2 | CAB39 | NM_016289 | 231,577,945 | chr17 | ERBB2 | NM_004448 | 37,863,243 | GGGGACAGCGACGACGCGGAGGCAGAGAAGGGAACGCCCGGCCC AGCCCC\|TGTGCACCGGCACAGACATGAAGCTGCGGCTCCCCTGCCA GTCCCGAGACC SEQ ID NO: 91 |
| 胃腺癌 | CAPZA2/MET | chr7 | CAPZA2 | NM_006136 | 116,502,704 | chr7 | MET | NM_000245 | 116,435,709 | CCAGAAGGAAGATGGCGGATCTGGAGGAGCAGTTGTCTGATGAAG AGAAG\|TGGTCCTTTGGCGTGCTCCTCTGGGAGCTGATGACAAGAG GAGCCCCACC SEQ ID NO: 92 |
| 侵袭性乳癌 | CBL/UBE4A | chr11 | CBL | NM_005188 | 119,158,656 | chr11 | UBE4A | NM_004788 | 118,261,372 | CAAAATCAAACCTTCCTCATCTGCCAATGCCATTTATTCTCTGGCTG CCA\|GGGATGAGGAGAATTTCTGTGCCACTGTGCCCAAGGATGGAC GTTCCTAT SEQ ID NO: 93 |
| 子宫内膜子宫内膜样腺癌 | EXOC4/BRAF | chr7 | EXOC4 | NM_021807 | 133,164,892 | chr7 | BRAF | NM_004333 | 140,434,570 | TCTGCGAGAACAGAGAAGGGAGCTCTATAGTCGGAGTGGAGAACT GCAAG\|ATTCTCGCCTCTATTGAGCTGCTGGCCCGCTCATTGCCAAA AATTCACCG SEQ ID NO: 94 |
| 低级别胶质瘤 | VIM/GFAP | chr10 | VIM | NM_003 | 17,271, | chr17 | GFAP | NM_00205 | 42,992,7 | CCTGCTGGCCGAGCTCGAGCAGCTCAAGGGCCAAGGCAAGTCGCG |

[0123]

| 基_5癌症 | 融合体名称 | 5'基因染色体 | 5'基因符号 | 5'登录号 | 5' 基因断裂点 | 3' 基因染色体 | 3'基因 | 3'登录号 | 3'基因断裂点 | 断裂点序列 |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 380 | 860 | | | 5 | 78 | CCTGG\|CTCCTGGCCGCCGTCTGGGTCCTGGCACCCGCCTCTCCCTG GCTCGAATG SEQ ID NO: 95 |
| 低级别胶质瘤 | VIM/GFAP | chr10 | VIM | NM_003380 | 17,276,745 | chr17 | GFAP | NM_002055 | 42,988,692 | CTGACCTCTCTGAGGCTGCCAACCGGAACAATGACGCCCTGCGCC AGGCA\|CAGGAGTACCAGGACCTGCTCAATGTCAAGCTGGCCCTGG ACATCGAGAT SEQ ID NO: 96 |
| 低级别胶质瘤 | VIM/GFAP | chr10 | VIM | NM_003380 | 17,276,789 | chr17 | GFAP | NM_002055 | 42,990,649 | CAGGCAAAGCAGGAGTCCACTGAGTACCGGAGACAGGTGCAGTCC CTCAC\|GTACCGCTCCAAGTTTGCAGACCTGACAGACGCTGCTGCC CGCAACGCGGSEQ ID NO: 97 |
| 低级别胶质瘤 | VIM/GFAP | chr10 | VIM | NM_003380 | 17,276,817 | chr17 | GFAP | NM_002055 | 42,988,824 | TTTGCCGTTGAAGCTGCTAACTACCAAGACACTATTGGCCGCCTGC AGGA\|GTACCAGGACCTGCTCAATGTCAAGCTGGCCCTGGACATCG AGATCGCCA SEQ ID NO: 98 |
| 低级别胶质瘤 | VIM/GFAP | chr10 | VIM | NM_003380 | 17,276,817 | chr17 | GFAP | NM_002055 | 42,988,824 | CATTGAGATTGCCACCTACAGGAAGCTGCTGGAAGGCGAGGAGAG CAGGA\|GTACCAGGACCTGCTCAATGTCAAGCTGGCCCTGGACATC GAGATCGCCA SEQ ID NO: 99 |
| 低级别胶质瘤 | VIM/GFAP | chr10 | VIM | NM_003380 | 17,277,255 | chr17 | GFAP | NM_002055 | 42,987,988 | GAACTTTGCCGTTGAAGCTGCTAACTACCAAGACACTATTGGCCGC CTGC\|TTCGAGAAACCAGCCTGGACACCAAGTCTGTGTCAGAAGGC CACCTCAAG SEQ ID NO: 100 |
| 低级别胶质瘤 | VIM/GFAP | chr10 | VIM | NM_003380 | 17,277,259 | chr17 | GFAP | NM_002055 | 42,988,687 | TTTGCCGTTGAAGCTGCTAACTACCAAGACACTATTGGCCGCCTGC AGGA\|GTACCAGGACCTGCTCAATGTCAAGCTGGCCCTGGACATCG AGATCGCCA SEQ ID NO: 101 |
| 低级别胶质瘤 | VIM/GFAP | chr10 | VIM | NM_003380 | 17,277,259 | chr17 | GFAP | NM_002055 | 42,988,687 | TTTGCCGTTGAAGCTGCTAACTACCAAGACACTATTGGCCGCCTGC AGGA\|GTACCAGGACCTGCTCAATGTCAAGCTGGCCCTGGACATCG AGATCGCCA SEQ ID NO: 102 |
| 低级别胶质瘤 | VIM/GFAP | chr10 | VIM | NM_003380 | 17,277,323 | chr17 | GFAP | NM_002055 | 42,988,623 | TGAAGGAGGAAATGGCTCGTCACCTTCGTGAATACCAAGACCTGCT CAAT\|CTAGAGGGCGAGGAGAACCGGATCACCATTCCCGTGCAGAC CTTCTCCAA SEQ ID NO: 103 |
| 低级别胶质瘤 | VIM/GFAP | chr10 | VIM | NM_003380 | 17,277,325 | chr17 | GFAP | NM_002055 | 42,988,666 | AAGGAGGAAATGGCTCGTCACCTTCGTGAATACCAAGACCTGCTCA ATGT\|CAAGCTGGCCCTGGACATCGAGATCGCCACCTACAGGAAGC TGCTAGAGG SEQ ID NO: 104 |
| 低级别胶质瘤 | VIM/GFAP | chr10 | VIM | NM_003380 | 17,277,370 | chr17 | GFAP | NM_002055 | 42,988,621 | AATGTTAAGATGGCCCTTGACATTGAGATTGCCACCTACAGGAAGC TGCT\|AGAGGGCGAGGAGAACCGGATCACCATTCCCGTGCAGACCT TCTCCAACC SEQ ID NO: 105 |
| 低级别胶质瘤 | VIM/GFAP | chr10 | VIM | NM_003380 | 17,277,370 | chr17 | GFAP | NM_002055 | 42,988,621 | AATGTTAAGATGGCCCTTGACATTGAGATTGCCACCTACAGGAAGC TGCT\|AGAGGGCGAGGAGAACCGGATCACCATTCCCGTGCAGACCT TCTCCAACC SEQ ID NO: 106 |
| 低级别胶质瘤 | VIM/GFAP | chr10 | VIM | NM_003380 | 17,277,370 | chr17 | GFAP | NM_002055 | 42,988,621 | AATGTTAAGATGGCCCTTGACATTGAGATTGCCACCTACAGGAAGC TGCT\|AGAGGGCGAGGAGAACCGGATCACCATTCCCGTGCAGACCT TCTCCAACC SEQ ID NO: 107 |
| 低级别胶质瘤 | VIM/GFAP | chr10 | VIM | NM_003380 | 17,277,370 | chr17 | GFAP | NM_002055 | 42,988,621 | AATGTTAAGATGGCCCTTGACATTGAGATTGCCACCTACAGGAAGC TGCT\|AGAGGGCGAGGAGAACCGGATCACCATTCCCGTGCAGACCT TCTCCAACC SEQ ID NO: 108 |

[0124]

| 表 5 癌症 | 融合体名称 | 5'基因染色体 | 5'基因符号 | 5'登录号 | 5' 基因断裂点 | 3' 基因染色体 | 3'基因 | 3'登录号 | 3'基因断裂点 | 断裂点序列 |
|---|---|---|---|---|---|---|---|---|---|---|
| 低级别胶质瘤 | VIM/GFAP | chr10 | VIM | NM_003380 | 17,277,375 | chr17 | GFAP | NM_002055 | 42,988,777 | TAAGATGGCCCTTGACATTGAGATTGCCACCTACAGGAAGCTGCTG GAAG\|GCGGGAGGCGGCCAGTTATCAGGAGGCGCTGGCGCGGCTG GAGGAAGAGG SEQ ID NO: 109 |
| 低级别胶质瘤 | VIM/GFAP | chr10 | VIM | NM_003380 | 17,277,877 | chr17 | GFAP | NM_002055 | 42,988,655 | GAAGGCGAGGAGAGCAGGATTTCTCTGCCTCTTCCAAACTTTTCCT CCCT\|TGGACATCGAGATCGCCACCTACAGGAAGCTGCTAGAGGGC GAGGAGAAC SEQ ID NO: 110 |
| 低级别胶质瘤 | VIM/GFAP | chr10 | VIM | NM_003380 | 17,277,877 | chr17 | GFAP | NM_002055 | 42,988,655 | GAAGGCGAGGAGAGCAGGATTTCTCTGCCTCTTCCAAACTTTTCCT CCCT\|TGGACATCGAGATCGCCACCTACAGGAAGCTGCTAGAGGGC GAGGAGAAC SEQ ID NO: 111 |
| 低级别胶质瘤 | GFAP/VIM | chr17 | GFAP | NM_002055 | 42,984,756 | chr10 | VIM | NM_003380 | 17,278,322 | CCTCAAGAGGAACATCGTGGTGAAGACCGTGGAGATGCGGGATGG AGAGG\|GATACCCACTCAAAAAGGACACTTCTGATTAAGACGGTTG AAACTAGAGA SEQ ID NO: 112 |
| 低级别胶质瘤 | GFAP/VIM | chr17 | GFAP | NM_002055 | 42,985,436 | chr10 | VIM | NM_003380 | 17,277,187 | GGCCACCTCAAGAGGAACATCGTGGTGAAGACCGTGGAGATGCGG GATGG\|AGATGCGTGAAATGGAAGAACTTTGCCGTTGAAGCTGC TAACTACCAA SEQ ID NO: 113 |
| 低级别胶质瘤 | GFAP/VIM | chr17 | GFAP | NM_002055 | 42,985,438 | chr10 | VIM | NM_003380 | 17,277,380 | AAGGCCACCTCAAGAGGAACATCGTGGTGAAGACCGTGGAGATGC GGGAT\|GGAGAGCAGGATTTCTCTGCCTCTTCCAAACTTTTCCTCCC TGAACCTGA SEQ ID NO: 114 |
| 低级别胶质瘤 | GFAP/VIM | chr17 | GFAP | NM_002055 | 42,985,452 | chr10 | VIM | NM_003380 | 17,277,278 | CAAGTCTGTGTCAGAAGGCCACCTCAAGAGGAACATCGTGGTGAA GACCG\|GGAGGAAATGGCTCGTCACCTTCGTGAATACCAAGACCTG CTCAATGTTA SEQ ID NO: 115 |
| 低级别胶质瘤 | GFAP/VIM | chr17 | GFAP | NM_001131019 | 42,987,510 | chr10 | VIM | NM_003380 | 17,277,303 | TTATACCAATACAGGCTCACCAGATTGTAAATGGAACGCCGCCGGC TCGC\|GAATACCAAGACCTGCTCAATGTTAAGATGGCCCTTGACATT GAGATTGC SEQ ID NO: 116 |
| 低级别胶质瘤 | GFAP/VIM | chr17 | GFAP | NM_002055 | 42,987,987 | chr10 | VIM | NM_003380 | 17,277,377 | AGGAGAACCGGATCACCATTCCCGTGCAGACCTTCTCCAACCTGCA GATT\|CGAGGAGAGCAGGATTTCTCTGCCTCTTCCAAACTTTTCCTC CCTGAACC SEQ ID NO: 117 |
| 低级别胶质瘤 | GFAP/VIM | chr17 | GFAP | NM_002055 | 42,987,987 | chr10 | VIM | NM_003380 | 17,277,377 | AGGAGAACCGGATCACCATTCCCGTGCAGACCTTCTCCAACCTGCA GATT\|CGAGGAGAGCAGGATTTCTCTGCCTCTTCCAAACTTTTCCTC CCTGAACC SEQ ID NO: 118 |
| 低级别胶质瘤 | GFAP/VIM | chr17 | GFAP | NM_002055 | 42,988,642 | chr10 | VIM | NM_003380 | 17,277,351 | AGGAGTACCAGGACCTGCTCAATGTCAAGCTGGCCCTGGACATCGA GATC\|GCCACCTACAGGAAGCTGCTGGAAGGCGAGGAGAGCAGGAT TTCTCTGCC SEQ ID NO: 119 |
| 低级别胶质瘤 | GFAP/VIM | chr17 | GFAP | NM_002055 | 42,988,655 | chr10 | VIM | NM_003380 | 17,277,336 | GCCCGCCACTTGCAGGAGTACCAGGACCTGCTCAATGTCAAGCTGG CCCT\|CTTGACATTGAGATTGCCACCTACAGGAAGCTGCTGGAAGGC GAGGAGAG SEQ ID NO: 120 |
| 子宫内膜子宫内膜样腺癌 | HLA-C/MUC16 | chr6 | HLA-C | NM_002117 | 31,237,270 | chr19 | MUC16 | NM_024690 | 8,959,665 | GCATTTTCTTCCCACAGGTGGAAAAGGAGGAGCTGCTCTCAGGCT GCGT\|CCAGCAACAGTGCCCAGGCTACTACCAGTCACACCTAGACC TGGAGGATC SEQ ID NO: 121 |
| 子宫内膜子宫内膜样腺癌 | HLA-C/MUC16 | chr6 | HLA-C | NM_002117 | 31,237,270 | chr19 | MUC16 | NM_024690 | 8,959,665 | GCATTTTCTTCCCACAGGTGGAAAAGGAGGAGCTGCTCTCAGGCT GCGT\|CCAGCAACAGTGCCCAGGCTACTACCAGTCACACCTAGACC TGGAGGATC SEQ ID NO: 122 |

[0125]

| 表5 癌症 | 融合体名称 | 5'基因染色体 | 5'基因符号 | 5'登录号 | 5' 基因断裂点 | 3' 基因染色体 | 3'基因 | 3'登录号 | 3'基因断裂点 | 断裂点序列 |
|---|---|---|---|---|---|---|---|---|---|---|
| 子宫内膜子宫内膜样腺癌 | HLA-C/MUC16 | chr6 | HLA-C | NM_002117 | 31,237,270 | chr19 | MUC16 | NM_024690 | 8,959,665 | GCATTTTCTTCCCACAGGTGGAAAAGGAGGGAGCTGCTCTCAGGCTGCGTCCAGCAACAGTGCCCAGGCTACTACCAGTCACACCTAGACCTGGAGGATC SEQ ID NO: 123 |
| 子宫内膜子宫内膜样腺癌 | HLA-C/MUC16 | chr6 | HLA-C | NM_002117 | 31,237,270 | chr19 | MUC16 | NM_024690 | 8,959,665 | GCATTTTCTTCCCACAGGTGGAAAAGGAGGGAGCTGCTCTCAGGCTGCGTCCAGCAACAGTGCCCAGGCTACTACCAGTCACACCTAGACCTGGAGGATC SEQ ID NO: 124 |
| 子宫内膜子宫内膜样腺癌 | HLA-C/MUC16 | chr6 | HLA-C | NM_002117 | 31,237,270 | chr19 | MUC16 | NM_024690 | 8,959,665 | GCATTTTCTTCCCACAGGTGGAAAAGGAGGGAGCTGCTCTCAGGCTGCGTCCAGCAACAGTGCCCAGGCTACTACCAGTCACACCTAGACCTGGAGGATC SEQ ID NO: 125 |
| 子宫内膜子宫内膜样腺癌 | HLA-C/MUC16 | chr6 | HLA-C | NM_002117 | 31,237,270 | chr19 | MUC16 | NM_024690 | 8,959,665 | GCATTTTCTTCCCACAGGTGGAAAAGGAGGGAGCTGCTCTCAGGCTGCGTCCAGCAACAGTGCCCAGGCTACTACCAGTCACACCTAGACCTGGAGGATC SEQ ID NO: 126 |
| 子宫内膜子宫内膜样腺癌 | HLA-C/MUC16 | chr6 | HLA-C | NM_002117 | 31,237,270 | chr19 | MUC16 | NM_024690 | 8,959,665 | GCATTTTCTTCCCACAGGTGGAAAAGGAGGGAGCTGCTCTCAGGCTGCGTCCAGCAACAGTGCCCAGGCTACTACCAGTCACACCTAGACCTGGAGGATC SEQ ID NO: 127 |
| 子宫内膜子宫内膜样腺癌 | HLA-C/MUC16 | chr6 | HLA-C | NM_002117 | 31,237,270 | chr19 | MUC16 | NM_024690 | 8,959,665 | GCATTTTCTTCCCACAGGTGGAAAAGGAGGGAGCTGCTCTCAGGCTGCGTCCAGCAACAGTGCCCAGGCTACTACCAGTCACACCTAGACCTGGAGGATC SEQ ID NO: 128 |
| 子宫内膜子宫内膜样腺癌 | HLA-C/MUC16 | chr6 | HLA-C | NM_002117 | 31,237,270 | chr19 | MUC16 | NM_024690 | 8,959,665 | GCATTTTCTTCCCACAGGTGGAAAAGGAGGGAGCTGCTCTCAGGCTGCGTCCAGCAACAGTGCCCAGGCTACTACCAGTCACACCTAGACCTGGAGGATC SEQ ID NO: 129 |
| 子宫内膜子宫内膜样腺癌 | HLA-C/MUC16 | chr6 | HLA-C | NM_002117 | 31,237,270 | chr19 | MUC16 | NM_024690 | 8,959,665 | GCATTTTCTTCCCACAGGTGGAAAAGGAGGGAGCTGCTCTCAGGCTGCGTCCAGCAACAGTGCCCAGGCTACTACCAGTCACACCTAGACCTGGAGGATC SEQ ID NO: 130 |
| 侵袭性乳癌 | HOOK3/IKBKB | chr8 | HOOK3 | NM_032410 | 42,798,568 | chr8 | IKBKB | NM_001556 | 42,147,725 | GATGCAGCAGAGCTTGGAAGGATGCTTCAGCTCATCTTAGGCTGTGCTGTCTTCAAGCTTGGCGCCCAATGACCTGCCCCTGCTGGCCATGGAGTACTGCCAAG SEQ ID NO: 131 |
| 侵袭性乳癌 | HOOK3/IKBKB | chr8 | HOOK3 | NM_032410 | 42,798,588 | chr8 | IKBKB | NM_001556 | 42,162,705 | GGATGCTTCAGCTCATCTTAGGCTGTGCTGTGAACTGTGAACAGAAGCAACAGCCTCTGCGCTTAGATACCTTCATGAAAACAGAATCATCCATCGGGATCT SEQ ID NO: 132 |
| 卵巢浆液性囊腺癌 | IGFBP2/SPP1 | chr2 | IGFBP2 | NM_000597 | 217,528,783 | chr4 | SPP1 | NM_000582 | 88,896,866 | GGGAGCCCCCACCATCCGGGGGGGACCCCGAGTGTCATCTCTTCTACAATGAGCAGCAGGAGGAGGCAGAGCACAGCATCGTCGGGACCAGACTCGTCTCA SEQ ID NO: 133 |
| 卵巢浆液性囊腺癌 | IGFBP2/SPP1 | chr2 | IGFBP2 | NM_000597 | 217,528,783 | chr4 | SPP1 | NM_000582 | 88,896,866 | TGAGACGAGTCTGGTCCCGACGATGCTGTGCTCTGCCTCCTCCTGCTGCT\|CATTGTAGAAGAGATGACACTCGGGGTCCCCCCGGATGGTGGGGCTCCC SEQ ID NO: 134 |
| 侵袭性乳癌 | KRIT1/CDK6 | chr7 | KRIT1 | NM_004912 | 91,842,555 | chr7 | CDK6 | NM_001259 | 92,462,486 | ATATTTACAAAGGCAAGCCCCAGCAATCATAAAGTCATCCCTGTGTATGT\|AGGAGGGCATGCCGCTCTCCACCATCCGCGAGGTGGCGGTGCTGAGGCAC SEQ ID NO: 135 |
| 头颈鳞状细胞癌 | LYN/NTRK3 | chr8 | LYN | NM_002350 | 56,866,524 | chr15 | NTRK3 | NM_002530 | 88,670,398 | AGATCCCCCGGGAGTCCATCAAGTTGGTGAAAAGGCTTGGCGCTGGGCAG\|TTTGGGGTATCCATAGCAGTTGGACTTGCTGCTTTTGCCTGTGTCCTGTT SEQ ID NO: 136 |

[0126]

| 基因5癌症 | 融合体名称 | 5'基因染色体 | 5'基因符号 | 5'登录号 | 5'基因断裂点 | 3'基因染色体 | 3'基因 | 3'登录号 | 3'基因断裂点 | 断裂点序列 |
|---|---|---|---|---|---|---|---|---|---|---|
| 侵袭性乳癌 | MLLT6/ACE | chr17 | MLLT6 | NM_005937 | 36,868,267 | chr17 | ACE | NM_000789 | 61,573,755 | CCACGCAGCAGGAGAAGCACCCCACCCACCACGAGAGGGGCCAG\|AAGAAG\|GTACTTTGTCAGCTTCATCATCCAGTTCCAGTTCCACGAG\|GCACTGTGCC SEQ ID NO: 137 |
| 侵袭性乳癌 | MLLT6/ACE | chr17 | MLLT6 | NM_005937 | 36,868,267 | chr17 | ACE | NM_000789 | 61,573,755 | CCTGGCACAGTGCCTCGTGGAACTGGAATGATGAAGCTGA\|CAAAG\|TACCTTCTTCTGGCCCCTCTCGTGGTGGGTGGGTGCTTCT\|CCTGCTGCG SEQ ID NO: 138 |
| 卵巢浆液性囊腺癌 | MUC16/OR7G2 | chr19 | MUC16 | NM_024690 | 9,024,134 | chr19 | OR7G2 | NM_001005193 | 9,213,935 | AGTGGATCTCAGAACCTCAGGGACTCCATCCTCCCTCTCCAGCCCC\|ACAA\|ATTCATCATCAACAGCATGGAAGCGAGAAACCAAACAGCTA\|TTTCAAAAT SEQ ID NO: 139 |
| 卵巢浆液性囊腺癌 | MUC16/OR7G2 | chr19 | MUC16 | NM_024690 | 9,045,564 | chr19 | OR7G2 | NM_001005193 | 9,213,935 | ATTTTGAAATAGCTGTTTGGTTTCTCGCTTCCATGCTGTTGATGATGA\|AT\|TTGTTCTTGAGGTCACACTCTCAGAGGCCAAGGTGGACATCCCA\|GGTGTG SEQ ID NO: 140 |
| 侵袭性乳癌 | NARS2/TOP1 | chr11 | NARS2 | NM_024678 | 78,189,672 | chr20 | TOP1 | NM_003286 | 39,721,138 | GGAACTGTTCAAGGCTACAACAATGATGGTTCTCTCAAAATGTCCT\|GAAG\|GCATCAAGTGGAAATTCCTAGAACATAAAGGTCCAGTATTTG\|CCCCACCA SEQ ID NO: 141 |
| 侵袭性乳癌 | SRD5A1/ PAPD7 | chr5 | SRD5A1 | NM_001047 | 6,633,982 | chr5 | PAPD7 | NM_006999 | 6,738,796 | GCGCCCAACTGCATCCTCCTGGCCATGTTCCTCGTCCACTACGGGC\|ATCG\|GTACAGATATTTGGCAGCTTTAGTACAGGTCTTTATCTTCCAA\|CTAGCGA SEQ ID NO: 142 |
| 侵袭性乳癌 | PAPD7/SRD5A1 | chr5 | PAPD7 | NM_006999 | 6,746,451 | chr5 | SRD5A1 | NM_001047 | 6,662,933 | GGGAGAAATTTTAATTACTTGAAAACCGGTATTAGAATCAAAGAAG\|GAGG\|CTTATTTGAATACGTAACTGCAGCCAACTATTTTGGAGAAAT\|CATGGAGT SEQ ID NO: 143 |
| 胃腺癌 | PRKAR2A/RHOA | chr3 | PRKAR2A | NM_004157 | 48,845,082 | chr3 | RHOA | NM_001664 | 49,405,981 | GACGAGGACTTGGAAGTTCCAGTTCCTAGCAGATTTAATAGACGAG\|TATC\|AGGTAGAGTTTGGCTTTGTGGGACACAGCTGGGCAGGAAGATT\|ATGATCGC SEQ ID NO: 144 |
| 胃腺癌 | TRAPPC9/PTK2 | chr8 | TRAPPC9 | NM_031466 | 141,460,889 | chr8 | PTK2 | NM_005607 | 141,900,868 | CTCTGTGTCCCGTTTGAGAAAAAGGACTTTGTAGGACTGGACACAG\|ACAG\|CAGAATATGACAGATACCTAGCATCTAGCAAAATAATGGCAG\|CTGCTTAC SEQ ID NO: 145 |
| 胃腺癌 | PTK2/TRAPPC9 | chr8 | PTK2 | NM_005607 | 142,011,224 | chr8 | TRAPPC9 | NM_031466 | 141,034,176 | CCGCCCCGTCGTCGTCTGCCTTCGCTTCACGGCGCCGAGCCGCGGT\|CCGA\|ACCCTGGAAGCTGTCCTGAATTTCAAATACTCTGGAGGCCCG\|GGCCACAC SEQ ID NO: 146 |
| 卵巢浆液性囊腺癌 | RAB11B/MDK | chr19 | RAB11B | NM_004218 | 8,468,319 | chr11 | MDK | NM_002391 | 46,404,173 | AGGAAGCATTCAAGAACATCCTCACAGAGATCTACCGCATCGTGTC\|ACAG\|GTGATGGGGGCACAGGCACCAAAGTCCGCCAAGGCACCCT\|GAAGAAGGCG SEQ ID NO: 147 |
| 卵巢浆液性囊腺癌 | RAB11B/MDK | chr19 | RAB11B | NM_004218 | 8,468,374 | chr11 | MDK | NM_002391 | 46,404,248 | GATCGCAGACCGCGCTGCCCACGACGAGTCCCCGGGGAACAACGT\|GGTGG\|CCATCCGCGTCACCAAGCCCTGCACCCCCAAGACCAAAGC\|AAAGGCCAAA SEQ ID NO: 148 |
| 鳞状细胞肺癌 | RB1/GADD45GIP1 | chr13 | RB1 | NM_000321 | 48,955,574 | chr19 | GADD45GIP1 | NM_052850 | 13,065,313 | AAAACATTTAGAACGATGTGAACATCGAATCATGGAATCCCTTGCAT\|GGC\|CAAGATGCCACAGATGATTGTGAACTGGCAGCAGCAGCAGCG\|GGAGAACT SEQ ID NO: 149 |
| 皮肤黑素瘤 | SHANK3/MAPK1 | chr22 | SHANK3 | NM_033517 | 51,115,121 | chr22 | MAPK1 | NM_002745 | 22,153,417 | TTTATGCCCAGAACCTCATCGATGATAAGCAGTTTGCAAAGCTTCAC\|ACA\|AAGATCTGTGACTTTGGCCTGGCCCGTGTTGCAGATCCAGACC\|ATGATCA SEQ ID NO: 150 |

[0127]

| 表5 癌症 | 融合体名称 | 5'基因染色体 | 5'基因符号 | 5'登录号 | 5'基因断裂点 | 3'基因染色体 | 3'基因 | 3'登录号 | 3'基因断裂点 | 断裂点序列 |
|---|---|---|---|---|---|---|---|---|---|---|
| 甲状腺腺体癌 | SPECC1L/RET | chr22 | SPECC1L | NM_015330 | 24,734,416 | chr10 | RET | NM_020630 | 43,610,055 | TGCAGCTGCAATTCCTCGAACGCCCCTGAGCCCAAGTCCTATGAAA ACCC\|CTCCTCAGCTGAGATGACCTTCCGGAGGCCCGCCCAGGCCT TCCCGGTCA SEQ ID NO: 151 |
| 成胶质细胞瘤 | TAOK1/RARA | chr17 | TAOK1 | NM_020791 | 27,718,042 | chr17 | RARA | NM_000964 | 38,504,568 | GGGAGGGCTGGGCACTATCTCTTCAGAACTGCTGCTCTGGGTCTCA ATGG\|CCTTTCGCCGACAGGTCTGGGGCGGAGCAGGCAGGCGCAGC CCCCTGCAG SEQ ID NO: 152 |
| 胃腺癌 | THRA/CDK12 | chr17 | THRA | NM_003250 | 38,245,586 | chr17 | CDK12 | NM_015083 | 37,686,884 | CAACCACCGCAAACACAACATTCCGCACTTCTGGCCCAAGCTGCTG ATGA\|AGAGAAGAGGCCCCCTGAGCCCCCGGACCTCCACCGCCGC CACCTCCAC SEQ ID NO: 153 |
| 侵袭性乳癌 | WRN/ADAM9 | chr8 | WRN | NM_000553 | 30,982,516 | chr8 | ADAM9 | NM_003816 | 38,871,484 | TCCTTGGGAATTATGGGAACTGAAAAATGCTGTGATAATTGCAGGT CCAG\|AGACCTTTTGCCTGAAGATTTTGTGGTTTATACTTACAACAA GGAAGGGA SEQ ID NO: 154 |
| 结肠和直肠腺癌 | YWHAE/MAP2K2 | chr17 | YWHAE | NM_006761 | 1,303,359 | chr19 | MAP2K2 | NM_030662 | 4,123,868 | CGCTATGGATGATCGAGAGGGATCTGGTGTACCAGGCGAAGCTGGCC GAGC\|TGGCCCGGAGGAAGCCGGTGCTGCCGGCGCTCACCATCAAC CCTACCATC SEQ ID NO: 155 |
| 甲状腺腺体癌 | ZC3HAV1/BRAF | chr7 | ZC3HAV1 | NM_020119 | 138,758,639 | chr7 | BRAF | NM_004333 | 140448284 | ACCAAGCCAGCCAATTCTGTCTTCACCACCAAATGGATTTGGTATTG GAA\|GAATGAAAACACTTGGTAGACGGGACTCGAGTGATGATTGGG AGATTCCT SEQ ID NO:193 |
| 甲状腺腺体癌 | BRAF/SND1 | chr7 | SND1 | NM_014390 | 127,361,454 | chr7 | BRAF | NM_004333 | 140448734 | TTCACCTGTCCAGCATCCGACCACCGAGGCTGGAGGGGGAGAACA CCCAG\|GACTTGATTAGAGACCAAGGATTTCGTGGTGATGGAGGATC AACCACAGG SEQ ID NO:194 |
| 甲状腺腺体癌 | BRAF/SND1 | chr7 | BRAF | NM_004333 | 140,487,348 | chr7 | SND1 | NM_014390 | 127724776 | GTCAATATTGATGACTTGATTAGAGACCAAGGATTTCGTGGTGATGG AGG\|CACCCAGTTGGAGAAGCTGATGGAGAACATGCGCAATGACAT TGCCAGTC SEQ ID NO:195 |
| 甲状腺腺体癌 | SND1/BRAF | chr7 | SND1 | NM_014390 | 127,361,454 | chr7 | BRAF | NM_004333 | 140448734 | CACCTGTCCAGCATCCGACCACCGAGGCTGGAGGGGGAGAACACC CAGGA\|CTTGATTAGAGACCAAGGATTTCGTGGTGATGGAGGATCA ACCACAGGTT SEQ ID NO:196 |
| 甲状腺腺体癌 | MEMO1/ALK | chr2 | MEMO1 | NM_015955 | 32,168,371 | chr2 | ALK | NM_004304 | 29543748 | GGCTTTCACAAGTACAGTCTACAAAAAGACCTGCTAGAGCCATTAT TGCC\|CCGGAAACTGCCTGTGGGTTTTTACTGCAACTTTGAAGATGG CTTCTGTG SEQ ID NO:197 |
| 头颈鳞状细胞癌 | CLIP4/ALK | chr2 | CLIP4 | NM_024692 | 29,404,563 | chr2 | ALK | NM_004304 | 29462609 | GAGGGGTCTCAGGTCCTGCTCACGAGCTCCAATGAGATGGGTACTG TTAG\|GTTGAAGATGCCCCAGCACAGACACGCCGTGGGACCGCATCA TGGTGTTCT SEQ ID NO:198 |
| 鳞状细胞肺癌 | CLIP4/ALK | chr2 | CLIP4 | NM_024692 | 29,404,561 | chr2 | ALK | NM_004304 | 29462607 | ACGAGGGGTCTCAGGTCCTGCTCACGAGCTCCAATGAGATGGGTAC TGTT\|AGGTTGAAGATGCCCAGCACAGACACGCCGTGGGACCGCAT CATGGTGTT SEQ ID NO:199 |

[0128]

表 6：表 3 的断裂点序列

| 表 6 融合体名称 | 5'基因染色体 | 5'基因符号 | 5'登录号 | 5'基因断裂点 | 3'基因染色体 | 3'基因符号 | 3'登录号 | 3'基因断裂点 | 断裂点序列 | SEQ ID NO: |
|---|---|---|---|---|---|---|---|---|---|---|
| SEC16A - NOTCH1 | chr9 | SEC16A | NM_014866 | 139357445 | chr9 | NOTCH1 | NM_017617 | 1.39E+08 | ATTGATTTCACGAATGAGGCAGTGGAGC<br>AGGTGGAAGAGGAGGAGTCTGG\|CCCGC<br>GATGCTCCCAGCCCGGTGAGACCTGCCT<br>GAATGGCGGGAAGTGTG | 28 |
| ERC1 - RET | chr12 | ERC1 | NM_178039 | 1,250,953 | chr10 | RET | NM_020630 | 43612032 | GGACATGTTGGATGTGAAGGAGCGGAAG<br>GTTAATGTTCTTCAGAAGAAGG\|AGGATC<br>CAAAGTGGGAATTCCCTCGGAAGAACTT<br>GGTTCTTGGAAAAACT | 29 |
| ESR1/CCDC170 | chr6 | ESR1 | NM_000125 | 152,332,929 | chr6 | CCDC170 | NM_025059 | 151,907,024 | CATGGAGCACCCAGGGAAGCTACTGTTT<br>GCTCCTAACTTGCTCTTGGACA\|GATGGT<br>CTCCCAGCTTGAAGCCCAAATATCTGAGC<br>TTGTTGAACAGTTGG | 156 |
| RPS6KB1/VMP1 | chr17 | RPS6KB1 | NM_003161 | 57,990,165 | chr17 | VMP1 | NM_030938 | 57,915,656 | CAGAATGTTTTGAGCTACTTCGGGTACTT<br>GGTAAAGGGGGCTATGGAAAG\|TGCTGTC<br>CCCGGCATAGGTCCATCTCTGCAGAAGC<br>CATTTCAGGAGTACC | 157 |
| VMP1/RPS6KB1 | chr17 | VMP1 | NM_030938 | 57,915,758 | chr17 | RPS6KB1 | NM_003161 | 57,987,923 | GTTCATATGGTCCAACTCCCCCCATGGTCC<br>ATGCTTTCATTTAACTGACCC\|TGTGGTGT<br>GCCCATTTCGCTTTTGTGGTGAAGCTTCT<br>GCCGTTGAGCCTC | 158 |
| RPS6KB1/VMP1 | chr17 | RPS6KB1 | NM_003161 | 57,970,686 | chr17 | VMP1 | NM_030938 | 57,915,656 | AGACCTGGACCAGCCAGAGGACGCGGG<br>CTCTGAGGATGAGCTGGAGGAGG\|GGTG<br>CTGTCCCCGGCATAGGTCCATCTCTGCAG<br>AAGCCATTTCAGGAGTA | 159 |
| VMP1/RPS6KB1 | chr17 | VMP1 | NM_030938 | 57,915,758 | chr17 | RPS6KB1 | NM_003161 | 57,987,923 | AAGTTCATATGGTCCAACTCCCCCCATGGT<br>CCATGCTTTCATTTAACTGAC\|CCTGTGGT<br>GTGCCCATTTCGCTTTTGTGGTGAAGCTT<br>CTGCCGTTGAGCC | 160 |
| RPS6KB1/VMP1 | chr17 | RPS6KB1 | NM_003161 | 57,970,606 | chr17 | VMP1 | NM_030938 | 57,915,656 | GGTACTCCTGAAATGGCTTCTGCAGAGAT<br>GGACCTATGCCGGGGACAGCA\|CTTCCCT<br>GTCTCGGAAGTCCGGGGCTGGGTAAAAG<br>CCGTCCCGCCTCCTT | 161 |
| RPS6KB1/VMP1 | chr17 | RPS6KB1 | NM_003161 | 57,992,064 | chr17 | VMP1 | NM_030938 | 57,915,656 | GTAACAGGAGCAAATACTGGGAAAATAT<br>TTGCCATGAAGGTGCTTAAAAAG\|GTGCTG<br>TCCCCGGCATAGGTCCATCTCTGCAGAAG<br>CCATTTCAGGAGTAC | 162 |
| RPS6KB1/VMP1 | chr17 | RPS6KB1 | NM_003161 | 58,003,943 | chr17 | VMP1 | NM_030938 | 57,917,129 | GCCTTTCAGACTGGTGGAAAACTCTACC<br>TCATCCTTGAGTATCTCAGTGG\|GAGAAA<br>ACTGGTTGTCCTGGATGTTTGAAAAGTTG<br>GTCGTTGTCATGGTG | 163 |

[0129]

| 表6 融合体名称 | 5'基因染色体 | 5'基因符号 | 5'登录号 | 5'基因断裂点 | 3'基因染色体 | 3'基因符号 | 3'登录号 | 3'基因断裂点 | 断裂点序列 | SEQ ID NO: |
|---|---|---|---|---|---|---|---|---|---|---|
| RPS6KB1/VMP1 | chr17 | RPS6KB1 | NM_003161 | 57,970,686 | chr17 | VMP1 | NM_030938 | 57,915,656 | AGACCTGGACCCAGCCAGAGGACGCGGG CTCTGAGGATGAGCTGGAGGAGG\|GGTG CTGTCCCCGGCATAGGTCCATCTCTGCAG AAGCCATTTCAGGAGTA | 164 |
| RPS6KB1/VMP1 | chr17 | RPS6KB1 | NM_003161 | 57,990,165 | chr17 | VMP1 | NM_030938 | 57,915,656 | CAGAATGTTTTGAGCTACTTCGGGTACTT GGTAAAGGGGGCTATGGAAAG\|TGCTGTC CCCGGCATAGGTCCATCTCTGCAGAAGC CATTTCAGGAGTACC | 165 |
| RPS6KB1/VMP1 | chr17 | RPS6KB1 | NM_003161 | 58,003,943 | chr17 | VMP1 | NM_030938 | 57,917,129 | ATGCCTTTCAGACTGGTGGAAAACTCTA CCTCATCCTTGAGTATCTCAGT\|GGAGA AAACTGGTTGTCCTGGATGTTTTGAAAAG TTGGTCGTTGTCATGG | 166 |
| RPS6KB1/VMP1 | chr17 | RPS6KB1 | NM_003161 | 58,009,009 | chr17 | VMP1 | NM_030938 | 57,917,215 | ATATTTATGGAAGACACTGCCTGCTTTTA CTTGGCAGAAATCTCCATGGC\|ACAAAGT TATGCCAAACGAATCCAGCAGCGGTTGA ACTCAGAGGAGAAAA | 167 |
| RPS6KB1/VMP1 | chr17 | RPS6KB1 | NM_003161 | 58,009,061 | chr17 | VMP1 | NM_030938 | 57,895,132 | TGGGGCATTTACATCAAAAGGGGATCATC TACAGAGACCTGAAGCCGGAG\|TGGTGCT GTCCCCCGGCATAGGTCCATCTCTGCAGAA GCCATTTCAGGAGT | 168 |
| RPS6KB1/VMP1 | chr17 | RPS6KB1 | NM_003161 | 57,970,625 | chr17 | VMP1 | NM_030938 | 57,915,703 | TACCCAGCCCCGGACTTCCGAGACAGGG AAGCTGAGGACATGGCAGGAGT\|ACCTG GAGGCTCAACGGCAGAAGCTTCACCACA AAAGCGAAATGGGCACA | 169 |
| RPS6KB1/VMP1 | chr17 | RPS6KB1 | NM_003161 | 57,970,643 | chr17 | VMP1 | NM_030938 | 57,915,710 | CCTGTGGTGTGCCCATTTCGCTTTTGTGG TGAAGCTTCTGCCGTTGAGCC\|TCCAGGT CTATGTCAAACACTCCTGCCATGTCCTCA GCTTCCCTGTCTCG | 170 |
| RPS6KB1/VMP1 | chr17 | RPS6KB1 | NM_003161 | 57,992,064 | chr17 | VMP1 | NM_030938 | 57,886,157 | AACAGGAGCAAATACTGGGAAAATATTT GCCATGAAGGTGCTTAAAAAGG\|ACTTTG CCTCCCGGGCCAAACTGGCAGTTCAAAA ACTAGTACAGAAAGTT | 171 |
| RPS6KB1/VMP1 | chr17 | RPS6KB1 | NM_003161 | 58,007,535 | chr17 | VMP1 | NM_030938 | 57,915,656 | CTATTTATGCAGTTAGAAAGAGAGGGAAT ATTTATGGAAGACACTGCCTG\|TGCTGTCC CCGGCATAGGTCCATCTCTGCAGAAGCCA TTTCAGGAGTACC | 172 |
| RPS6KB1/VMP1 | chr17 | RPS6KB1 | NM_003161 | 57,970,625 | chr17 | VMP1 | NM_030938 | 57,915,703 | TGTGCCCATTTCGCTTTTGTGGTGAAGCT TCTGCCGTTGAGCCTCCAGGT\|ACTCCTG CCATGTCCTCAGCTTCCCTGTCTCGGAAG TCCGGGGCTGGGTA | 173 |
| RPS6KB1/VMP1 | chr17 | RPS6KB1 | NM_003161 | 57,990,165 | chr17 | VMP1 | NM_030938 | 57,917,129 | CCAGAATGTTTTGAGCTACTTCGGGTACT TGGTAAAGGGGGCTATGGAAA\|GGGAGA | 174 |

[0130]

表6 融合体名称 | 5'基因染色体 | 5'基因符号 | 5'登录号 | 5'基因断裂点 | 3'基因染色体 | 3'基因符号 | 3'登录号 | 3'基因断裂点 | 断裂点序列 | SEQ ID NO:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| RPS6KB1/VMP1 | chr17 | RPS6KB1 | NM_003161 | 57,992,037 | chr17 | VMP1 | NM_030938 | 57,851,147 | AAACTGGTTGTCCTGGATGTTTGAAAAG TTGGTCGTTGTCATGG | |
| | | | | | | | | | ATGGAAAGGTTTTTCAAGTACGAAAAGT AACAGGAGCAAATACTGGGAAA\|ATATTT CATGGCCAGAGCAGCTCGCCTCTCAGGT GCTGAACCAGATGATG | 175 |
| RPS6KB1/VMP1 | chr17 | RPS6KB1 | NM_003161 | 57,970,686 | chr17 | VMP1 | NM_030938 | 57,889,031 | ACCTGGACCAGCCAGAGGACGCGGGCTC TGAGGATGAGCTGGAGGAGGGG\|ATTCC AAATCCTTTATTTGATCTGGCTGGAATAA CGTGTGGACACTTTCT | 176 |
| RPS6KB1/VMP1 | chr17 | RPS6KB1 | NM_003161 | 57,970,686 | chr17 | VMP1 | NM_030938 | 57,886,157 | ACCTGGACCAGCCAGAGGACGCGGGCTC TGAGGATGAGCTGGAGGAGGGG\|GACTT TGCCTCCCGGGCCAAACTGGCAGTTCAA AAACTAGTACAGAAAGT | 177 |
| RPS6KB1/VMP1 | chr17 | RPS6KB1 | NM_003161 | 57,970,686 | chr17 | VMP1 | NM_030938 | 57,889,031 | GAAAGTGTCCACACGTTATTCCAGCCAG ATCAAATAAAGGATTTGGAATC\|CCCTCCT CCAGCTCATCCTCAGAGCCCGCGTCCTCT GGCTGGTCCAGGTC | 178 |
| RPS6KB1/VMP1 | chr17 | RPS6KB1 | NM_003161 | 57,970,686 | chr17 | VMP1 | NM_030938 | 57,886,157 | CTGGACCAGCCAGAGGACGCGGGCTCTG AGGATGAGCTGGAGGAGGGGGA\|CTTTG CCTCCCGGGCCAAACTGGCAGTTCAAAA ACTAGTACAGAAAGTTG | 179 |
| RPS6KB1/VMP1 | chr17 | RPS6KB1 | NM_003161 | 57,970,686 | chr17 | VMP1 | NM_030938 | 57,886,157 | CCTGGACCAGCCAGAGGACGCGGGCTCT GAGGATGAGCTGGAGGAGGGGG\|ACTTT GCCTCCCGGGCCAAACTGGCAGTTCAAA AACTAGTACAGAAAGTT | 180 |
| RPS6KB1/VMP1 | chr17 | RPS6KB1 | NM_003161 | 57,970,686 | chr17 | VMP1 | NM_030938 | 57,886,157 | ACCTGGACCAGCCAGAGGACGCGGGCTC TGAGGATGAGCTGGAGGAGGGGG\|GACTT TGCCTCCCGGGCCAAACTGGCAGTTCAA AAACTAGTACAGAAAGT | 181 |
| RPS6KB1/VMP1 | chr17 | RPS6KB1 | NM_003161 | 57,970,686 | chr17 | VMP1 | NM_030938 | 57,886,157 | CGCGGGCTCTGAGGATGAGCTGGAGGAG GGGGA\|CTTTGCCTCCCGGGCCAAACTGG CAGTTCAAAAACTAGTACAGAAAGTTG | 182 |
| RPS6KB1/VMP1 | chr17 | RPS6KB1 | NM_003161 | 57,992,064 | chr17 | VMP1 | NM_030938 | 57,915,656 | AGTAACAGGAGCAAATACTGGGAAAATA TTTGCCATGAAGGTGCTTAAAA\|AGTGCT GTCCCCGGCATAGGTCCATCTCTGCAGAA GCCATTTCAGGAGTA | 183 |

[0131] 本发明提供新颖基因融合体和基因融合变异体(即,搭配物基因中的一者或两者上的不同断裂点位置),其选自基因融合体的表1-表3、表19和表22中所展示的那些,如

TPM1/ALK、PRKAR1A/ALK、NCOA1/ALK、LPP/CASR、MDM2/EGFR、FGFR3/ELAVL3、B2M/GNAS、DOCK8/JAK2、HNF1B/NOTCH1、NFASC/NTRK1、SSBP2/NTRK1、SQSTM1/NTRK1、TBL1XR1/PIK3CA、AKAP13/RET、FKBP15/RET、TBL1XR1/RET、CEP85L/ROS1、CLCN6/RAF1、TRAK1/RAF1、PRKACA/AKT1、PRKACA/AKT2、MLL/FYN、ECHD1/FYN、TTC13/JAK2、SEC16A/NOTCH1、ERC1/RET、GTF2IRD1/ALK、HTATSF1/BRS3、CDH1/CCDC132、CCDC132/CDH1、ERBB2/SLC29A3、MET/TFG;TFG/MET、NOTCH2/MNDA、IRF2BP2/NTRK1、EIF2C2/PTK2、RARA/HOXB3、STAT3/ETV4、以及GFAP/VIM;VIM/GFAP、TOP1/C17orf64、以及TP53/KIAA0753。由于这些发现,本发明提供经分离基因融合体核酸和与其互补的序列、扩增子、转录物、反应混合物以及特异性地识别基因融合体、与其互补的序列、扩增子和转录物的核酸序列的探针。本发明进一步涵盖用于治疗相关疾病的反义核苷酸。

[0132] 表1-表3、表19和表22提供基因融合体的列表(基因A/基因B),表明所涉及的基因(基因A和基因B)、染色体位置、断裂点位置、融合体类型和距离。基因融合体与相关TCGA疾病一起展示(癌症基因组图谱(The Cancer Genome Atlas))。癌症用3-4个字母的缩写展示,所述缩写更详细地阐述在诊断学部分中。

[0133] 一般来说,表1-3、19和22提供一种或多种新颖基因融合体和/或基因融合体与TCGA疾病的相关性。举例来说,表19呈现新颖基因融合体,且表22呈现基因融合体与TCGA疾病的新颖相关性。

[0134] 表4-6、20和23提供表1-3、19和22中的基因融合体的断裂点序列。断裂点序列被鉴别为SEQ ID NO:1-289。

[0135] 分析和试剂盒

[0136] 在某些实施例中,提供检测的分析和方法。本文提供的检测基因融合体的方法是所属领域中已知的。作为非限制性实例,所述分析可以包括5'核酸酶PCR分析(加利福尼亚州福斯特市的应用生物系统公司(Applied Biosystems,Foster City,CA))、下一代测序分析(加利福尼亚州卡尔斯巴德的离子激流公司;加利福尼亚州圣地亚哥的伊路米那公司)或微阵列分析(斯科泰姆(Skotheim)等人,《分子癌症》(Molecular Cancer)2009,8:5)。在至少一个实施例中,所述分析或方法包括至少一种引物或探针,其与表1-6中的基因融合体和/或断裂点互补或编码所述基因融合体和/或断裂点。

[0137] 在至少一个实施例中,提供定量基因融合体的表达量的分析和方法。所述方法可能涉及一种或多种外显子的定量表达。举例来说,TaqMan™基因表达分析可经设计以用于一组已知融合体转录物以便定量分析。所述分析可经设计使得引物和探针跨越断裂点区域,不过在某些说明性实施例中,引物和探针不直接置放在断裂点上。

[0138] 在某些实施例中,本发明提供一种引物、一种探针或一种探针或引物组,其特异性地识别本文所公开的基因融合体和/或断裂点中的一者或多者。

[0139] 在一个实施例中,本发明提供一种组合物和一种试剂盒,其包含特异性地识别选自表1-3、19和22的基因融合体和/或表4-6、20和23中的断裂点的一个探针组。所述探针组可例如为一个扩增引物组。在另一实施例中,本文提供一种组合物,其包括侧接目标核酸中选自表1-3、19和22的基因融合体的一个引物组。此实施例的反应混合物可以进一步包括检测器探针,所述检测器探针结合于选自表1-3、19和22的基因融合体中的断裂点的任一侧,或结合跨越选自表1-3、19和22的基因融合体中的断裂点的结合区。包括检测器探针或不包

括检测器探针的反应混合物可以进一步包括聚合酶、dNTP和/或尿嘧啶DNA去糖基化酶(UDG)。聚合酶和UDG通常不来自人类来源。反应混合物可以进一步包括目标核酸，例如人类目标核酸。人类目标核酸可例如从来自怀疑患有癌症的个体的生物样品中分离。

[0140] 在另一实施例中，本文提供qPCR分析，如TaqMan™分析或MolecularBeacons™分析，其特异性地扩增并检测包括SEQ ID NO:1-289的目标核酸。

[0141] 本发明还提供一种经分离核酸，其包含至少一种选自SEQ ID NO:1-289的序列。经分离核酸可以在5'末端上包括第一引物。此外，核酸可能为单链或双链的。

[0142] 在其他实施例中，本发明提供一种试剂盒，其包括一个检测器探针和/或一个探针组，例如一个扩增引物组，其特异性地识别包含选自表1-3、19和22的基因融合体的断裂点的核酸。举例来说，在某些实施例中，检测器探针或扩增引物组经设计以扩增和/或检测包括SEQ ID NO:1-289中的至少一者的核酸。所述试剂盒可以在另一容器中或在同一容器中进一步包括来自扩增反应混合物的组分，如聚合酶(通常不来自人类来源)、dNTP和/或UDG。此外，所述试剂盒可以包括对照核酸。举例来说，对照核酸可以包括包含选自表1-3、19和22的基因融合体中的断裂点的序列。

[0143] 在一些实施例中，提供一种试剂盒，其涵盖至少2个引物对和2种可检测标记的探针。在这些非限制性实施例中，2个引物对和/或2种可检测标记的探针形成2种扩增检测分析。

[0144] 本发明的试剂盒还可包含用于执行本文所述的一种或多种方法的说明和/或本文所述的一种或多种组合物或试剂的描述。说明和/或描述可以呈印刷形式且可以包括于试剂盒插页中。试剂盒还可以包括提供此类说明或描述的互联网位置的书面描述。

[0145] 在一些实施例中，所述试剂盒和分析包含一种或多种特异性地识别目标(如基因融合体核酸序列)的探针。在至少一个实施例中，所述试剂盒和分析为诊断试剂盒和分析。

[0146] 提供一种试剂盒，其包含特异性地识别包含来自表4-6、20和23的断裂点的核酸的一个探针组(例如一个扩增引物组)。所述试剂盒可以在另一容器中或在同一容器中进一步包括来自扩增反应混合物的组分，如聚合酶(通常不来自人类来源)、dNTP和/或UDG。此外，所述试剂盒可以包括对照核酸。举例来说，对照核酸可以包括包含选自表4-6、20和23的断裂点的序列。

[0147] 在另一实施例中，提供一种基因融合体，其包含表4-6、20和23中的断裂点中的至少一者。

[0148] 在一些实施例中，提供一种反应混合物和一种试剂盒。在一些实施例中，所述试剂盒涵盖选择性结合基因融合体的可检测探针。在一些实施例中，基因融合体是表4、表5、表6、表20或表23中的基因融合体中的任一者。

[0149] 因此，在一些实施例中，提供一种试剂盒，其涵盖选择性结合基因融合体的反应混合物和可检测探针，所述基因融合体为表4、表5、表6、表20或表23中的基因融合体中的任一者。

[0150] 诊断学

[0151] 本文中涵盖诊断、治疗和检测基因融合体和相关疾病的方法。所述方法可以包括检测受试者样品中的基因融合体。

[0152] 受试者样品可为包括来自受试者的核酸的任何身体组织或体液。在某些实施例

中，样品将是包含循环肿瘤细胞或无细胞DNA的血液样品。在其他实施例中，样品可为组织，如癌组织。癌组织可以来自肿瘤组织，并且可以经新鲜冷冻或福尔马林固定、石蜡包埋(FFPE)。

[0153] 所述疾病可为癌症或肿瘤。癌症可以包括(但不限于)黑素瘤、宫颈癌、胰脏癌、头颈鳞状癌、肺腺癌、结肠腺癌、子宫癌、卵巢癌、成胶质细胞瘤、低级别胶质瘤、肺腺癌、甲状腺癌和胃癌。

[0154] 癌症可以包括(但不限于)膀胱癌、乳癌、子宫颈细胞癌、结肠腺癌、多形性成胶质细胞瘤、头颈鳞状细胞癌、透明细胞肾细胞癌、肾脏肾乳头状细胞癌、急性骨髓性白血病、脑低级别胶质瘤、肝脏肝细胞癌、肺腺癌、鳞状细胞肺癌、卵巢浆液性腺癌、前列腺腺癌、直肠腺癌、皮肤黑素瘤、胃腺癌、甲状腺癌和子宫体子宫内膜样癌。如本文所用，BLCA＝膀胱癌、BRCA＝乳癌、CESC＝子宫颈细胞癌、COAD＝结肠腺癌、GBM＝多形性成胶质细胞瘤、HNSC＝头颈鳞状细胞癌、KIRK＝透明细胞肾细胞癌、KIRP＝肾脏肾乳头状细胞癌、LAML＝急性骨髓性白血病、LGG＝脑低级别胶质瘤、LIHC＝肝脏肝细胞癌、LUAD＝肺腺癌、LUSC＝鳞状细胞肺癌、OV＝卵巢浆液性腺癌、PRAD＝前列腺腺癌、READ＝直肠腺癌、SKCM＝皮肤黑素瘤、STAD＝胃腺癌、THCA＝甲状腺癌和UCEC＝子宫体子宫内膜样癌。

[0155] 在一些实施例中，提供一种检测新颖基因变异体或基因融合体的方法，所述方法涵盖反应混合物，其中所述新颖基因变异体或基因融合体通过生成延伸产物来检测。

[0156] 在另一实施例中，本发明提供利用所公开的基因融合体和基因变异体的诊断和治疗目标。基因融合体、基因变异体和相关疾病病况为诊断和治疗两者提供目标。举例来说，基因融合体目标或基因变异体的存在、不存在、或表达升高或降低可以用于诊断疾病病况或可以用于预测或检测疾病病况。在至少一个实施例中，基因融合体或基因变异体可在特定癌症中具有高发生率(频率)、中等发生率或低发生率。在至少一个实施例中，基因融合体或基因变异体可在一种癌症或肿瘤中具有高频率，而在另一种癌症或肿瘤中具有低或中等发生率。在至少一个实施例中，基因融合体或基因变异体可以具有与癌症或肿瘤相关联的中等或低频率。在至少一个实施例中，低或中等频率基因融合体或基因变异体可与癌症的一种或多种不同高频率生物标记物组合使用以有助于诊断、预测或鉴别疾病的倾向性。所述方法可用于筛选患者的癌症或预测癌症的特定结果的相对展望。举例来说，BRCA1或BRCA2突变的存在可与乳癌的基因融合体JAK2/TTC13组合分析。

[0157] 提供一种检测癌症的方法，其包含扩增跨越选自表1-3、19和22的基因融合体中的断裂点的核酸，例如，所述核酸可以包括选自SEQ ID NO:1-289的序列；以及检测所述核酸的存在，其中所述核酸的存在指示癌症存在于样品中。在另一方法中，本文提供一种检测癌症的方法，其包括产生包括选自SEQ ID NO:1-289的序列的扩增子，以及检测所述核酸的存在，其中所述核酸的存在指示癌症或癌细胞存在于样品中。所述扩增子通常包括经延伸以形成所述扩增子的引物。所述癌症是选自膀胱癌、乳癌、子宫颈细胞癌、结肠腺癌、多形性成胶质细胞瘤、头颈鳞状细胞癌、透明细胞肾细胞癌、肾脏肾乳头状细胞癌、急性骨髓性白血病、脑低级别胶质瘤、肝脏肝细胞癌、肺腺癌、鳞状细胞肺癌、卵巢浆液性腺癌、前列腺腺癌、直肠腺癌、皮肤黑素瘤、胃腺癌、甲状腺癌和子宫体子宫内膜样癌。

[0158] 在另一实施例中是一种方法，其通过检测选自表1-3、19和22的基因融合体的存在来检测样品中的选自膀胱癌、乳癌、子宫颈细胞癌、结肠腺癌、多形性成胶质细胞瘤、头颈鳞

状细胞癌、透明细胞肾细胞癌、肾脏肾乳头状细胞癌、急性骨髓性白血病、脑低级别胶质瘤、肝脏肝细胞癌、肺腺癌、鳞状细胞肺癌、卵巢浆液性腺癌、前列腺腺癌、直肠腺癌、皮肤黑素瘤、胃腺癌、甲状腺癌和子宫体子宫内膜样癌的癌症。

[0159] 新基因融合体

[0160] 尽管先前已报导基因融合体中的一些，但本文提供其中断裂点不同和/或并非先前已知的基因融合体的许多变化。其中断裂点不同和/或并非先前已知的基因融合体的非限制性实例包括：TPM1/ALK、PRKAR1A/ALK、NCOA1/ALK、LPP/CASR、MDM2/EGFR、FGFR3/ELAVL3、B2M/GNAS、DOCK8/JAK2、HNF1B/NOTCH1、NFASC/NTRK1、SSBP2/NTRK1、SQSTM1/NTRK1、TBL1XR1/PIK3CA、AKAP13/RET、FKBP15/RET、TBL1XR1/RET、CEP85L/ROS1、CLCN6/RAF1、TRAK1/RAF1、PRKACA/AKT1、PRKACA/AKT2、MLL/FYN、ECHD1/FYN以及TTC13/JAK2为具有在表4和5中以SEQ ID NO:1-289形式提供的断裂点的新颖变异体。

[0161] 本文还提供与一种或多种癌症相关的许多基因融合变异体。

[0162] 癌症相关性

[0163] 本文中呈现与癌症相关的新基因融合体。基因融合体中的一些可能已经与先前的特异癌症或疾病病况相关。本文中的方法已鉴别出可以用于帮助诊断和/或治疗特定癌症的新相关性。表1-3、19和22中所展示的基因融合体提供与融合体有关的基因以及基因融合体与一种或多种特定癌症的相关性。举例来说，融合体PRKACA/AKT1展示出与结肠腺癌和子宫内膜子宫内膜样腺癌相关。

[0164] 表3中所展示的基因融合体为已展示出与新癌症相关的先前已知的基因融合体。举例来说，SEC16A/NOTCH1先前经鉴别与乳癌相关。当前方法鉴别出基因融合体SEC16A/NOTCH1与甲状腺腺体癌的相关性。另外，ERC1/RET先前经鉴别与甲状腺癌相关。当前方法鉴别出基因融合体ERC1/RET与侵袭性乳癌的相关联(参见表3和6)。

[0165] 反应混合物和扩增子

[0166] 在另一实施例中，本发明提供一种反应混合物，其包含特异性地识别选自表1-表3、表19和表22的基因融合体的一个探针或一个探针组。所述探针组可为例如一个扩增引物组或一个经标记探针。在另一实施例中，本文提供一种反应混合物，其包括侧接目标核酸中选自表1-表3、表19和表22的基因融合体的一个引物组。举例来说，所述引物组可以各自在表4-6、20和23中鉴别的融合体断裂点中的一者的相对侧的1000、750、500、250、100、90、80、75、70、65、50或25个核苷酸内结合于人类基因组中的目标序列。此实施例的反应混合物可以进一步包括检测器探针，所述检测器探针结合于选自表1-表3、表19和表22的基因融合体中的断裂点的任一侧，或结合跨越选自表1-表3、表19和表22的基因融合体中的断裂点的结合区，包括其中断裂点鉴别于表4-6、20和23中的特定实施例。在例示性实施例中，检测器探针在表4-6、20和23中鉴别的融合体断裂点中的一者的1000、750、500、250、100、90、80、75、70、60、50或25个核苷酸内结合于人类基因组中的目标序列。包括检测器探针或不包括检测器探针的反应混合物可以进一步包括聚合酶、逆转录酶、dNTP和/或尿嘧啶DNA去糖基化酶(UDG)。聚合酶、逆转录酶和UDG通常不来自人类来源。说明性实施例中的聚合酶是热稳定聚合酶，如Taq聚合酶。在某些实施例中，反应混合物中的dNTP包括dUTP，且在某些实例中，反应混合物可以没有dTTP。

[0167] 反应混合物可以进一步包括目标核酸，例如人类目标核酸。人类目标核酸可例如

从怀疑患有选自以下各者的癌症的个体的生物样品(如肿瘤样品)中分离:BLCA=膀胱癌、BRCA=乳房癌、CESC=子宫颈细胞癌、COAD=结肠腺癌、GBM=多形性成胶质细胞瘤、HNSC=头颈鳞状细胞癌、KIRK=透明细胞肾细胞癌、KIRP=肾脏肾乳头状细胞癌、LAML=急性骨髓性白血病、LGG=脑低级别胶质瘤、LIHC=肝脏肝细胞癌、LUAD=肺腺癌、LUSC=鳞状细胞肺癌、OV=卵巢浆液性腺癌、PRAD=前列腺腺癌、READ=直肠腺癌、SKCM=皮肤黑素瘤、STAD=胃腺癌、THCA=甲状腺癌和UCEC=子宫体子宫内膜样癌。在某些实施例中,目标核酸来自肿瘤,例如先前句子中列出的癌症类型中的一者的肿瘤。此外,目标核酸可从来自肿瘤的生物样品(如FFPE样品)中提取。

[0168] 本发明的反应混合物可以包括扩增子。扩增子可为例如经分离核酸。扩增子的长度可例如在25与2500个核苷酸之间、在25与2000个核苷酸之间、在25与1000个核苷酸之间、在50与1000个核苷酸之间、在50与500个核苷酸之间、在50与250个核苷酸之间、在50与200个核苷酸之间、在50与150个核苷酸之间、在50与100个核苷酸之间或在50与75个核苷酸之间。

[0169] 扩增子可在表4-6、20和23中鉴别的融合体断裂点中的一者的相对侧的1000、750、500、250、100、90、80、75、70、65、50或25个核苷酸内具有与人类基因组中的目标序列一致或互补的核苷酸序列。在某些实施例中,扩增子包括图4-6中所提供的核苷酸序列中的25到250个、25到100个、25到75个、50到250个、50到200个、50到150个、50到100个或50到75个或其互补序列。在某些实施例中,扩增子包括自然界中出现的序列变异体。举例来说,扩增子可以包括对应于单核苷酸变异体或天然存在的等位基因的可变的核苷酸序列。

[0170] 在某些说明性实施例中,本发明的扩增子具有未自然界中发现和/或未在哺乳动物(如人类)中发现的化学结构。举例来说,某些说明性扩增子包括未自然界中发现的或未在哺乳动物中发现的或被发现可能不结合于扩增子的糖-磷酸酯主链的类型的碱基。举例来说,扩增子可能是DNA扩增子,其包括结合于糖磷酸酯主链的尿嘧啶碱基,由此在模板中含有一个胸苷残基的至少一个位置且在说明性实例中所有位置处具有一个尿苷残基。

[0171] 因此,在说明性实施例中,扩增子是DNA扩增子,其包括一个或多个脱氧尿苷(“dU”)残基。dU残基可通过在用于生成扩增子的引物中包括所述残基来添加。在某些实施例中,反应混合物包括DNA扩增子,其包括针对对应人类基因组序列中的每一脱氧胸苷残基的一个或多个dU残基。当使用扩增反应(如PCR)产生扩增子时,这些扩增子可例如通过使用包括dUTP而非dTTP的dNTP混合物产生。

[0172] 在某些实施例中,扩增子包括所对应序列并未在人类基因组中发现的区段,如寡核苷酸序列,例如DNA条形码序列。非人类区段的长度可例如为5-10,000、5-5000、5-1000、5-500、5-100、5-50、5-25、5-10、10-10,000、10-5000、10-1000、10-500、10-100、10-50或10-25个核苷酸。

[0173] 在某些实施例中,扩增子包括对应于人类基因组中跨越内含子的区域的区段,但扩增子并不包括对应于内含子的区段。

[0174]

**基因变异体（表 7 和/或表 11）**

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 前列腺腺癌 | ACOT7 | 1 | 6387379 | SNP | A | G | G | Oncomine | NM_007274 | p.V202A | p.V202 | 错义突变 | 热点 |
| 鳞状细胞肺癌 | ACOT7 | 1 | 6387379 | SNP | A | G | G | Oncomine | NM_007274 | p.V202A | p.V202 | 错义突变 | 热点 |
| 透明细胞肾细胞癌 | ACOT7 | 1 | 6387379 | SNP | A | G | G | Oncomine | NM_007274 | p.V202A | p.V202 | 错义突变 | 热点 |
| 前列腺腺癌 | ANAPC1 | 2 | 112625621 | SNP | G | C | C | Oncomine | NM_022662 | p.P222A | p.P222 | 错义突变 | 热点 |
| 成神经管细胞瘤 | ANAPC1 | 2 | 112625621 | SNP | G | C | C | Oncomine | NM_022662 | p.P222A | p.P222 | 错义突变 | 热点 |
| 胃腺癌 | ANAPC1 | 2 | 112625621 | SNP | G | C | C | Oncomine | NM_022662 | p.P222A | p.P222 | 错义突变 | 热点 |
| 肺腺癌 | ANAPC1 | 2 | 112625621 | SNP | G | C | C | Oncomine | NM_022662 | p.P222A | p.P222 | 错义突变 | 热点 |
| 乳头状肾细胞癌 | ANAPC1 | 2 | 112625621 | SNP | G | C | C | Oncomine | NM_022662 | p.P222A | p.P222 | 错义突变 | 热点 |
| 结肠直肠腺癌 | C2orf69 | 2 | 200498052 | SNP | G | A | A | Oncomine | NM_153689 | p.R119H | p.R119 | 错义突变 | 热点 |
| 胃腺癌 | C2orf69 | 2 | 200789806 | SNP | C | T | T | Oncomine | NM_153689 | p.R119C | p.R119 | 错义突变 | 热点 |
| 胃腺癌 | C2orf69 | 2 | 200789807 | SNP | G | A | A | Oncomine | NM_153689 | p.R119H | p.R119 | 错义突变 | 热点 |
| 皮肤黑素瘤 | C4orf22 | 4 | 81791162 | SNP | C | T | T | Oncomine | NM_152770 | p.R117C | p.R117 | 错义突变 | 热点 |
| 皮肤黑素瘤 | C4orf22 | 4 | 81791162 | SNP | C | T | T | Oncomine | NM_152770 | p.R117C | p.R117 | 错义突变 | 热点 |
| 皮肤黑素瘤 | C4orf22 | 4 | 81504291 | SNP | C | T | T | Oncomine | NM_152770 | p.T96M | p.T96 | 错义突变 | 热点 |
| 甲状腺乳头状癌 | C4orf22 | 4 | 81504291 | SNP | C | T | T | Oncomine | NM_152770 | p.T96M | p.T96 | 错义突变 | 热点 |
| 肺腺癌 | C4orf3 | 4 | 120221 | SNP | C | T | T | Oncomine | NM_00100 | p.R18Q | p.R18 | 错义突变 | 热点 |

[0175]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 638 | | | | | | 1701 | | | | |
| 导管乳癌 | C4orf3 | 4 | 120221638 | SNP | C | C | G | Oncomine | NM_001001701 | p.R18P | p.R18 | 错义突变 | 热点 |
| 肺腺癌 | C4orf3 | 4 | 120221638 | SNP | C | T | T | Oncomine | NM_001001701 | p.R18Q | p.R18 | 错义突变 | 热点 |
| 前列腺腺癌 | CACNG3 | 16 | 24373167 | SNP | C | T | T | Oncomine | NM_006539 | p.R311C | p.R311 | 错义突变 | 热点 |
| 皮肤黑素瘤 | CACNG3 | 16 | 24372868 | SNP | C | T | T | Oncomine | NM_006539 | p.S211F | p.S211 | 错义突变 | 热点 |
| 肺腺癌 | CACNG3 | 16 | 24372930 | SNP | C | T | T | Oncomine | NM_006539 | p.R232W | p.R232 | 错义突变 | 热点 |
| 成胶质细胞瘤 | CACNG3 | 16 | 24366270 | SNP | G | A | A | Oncomine | NM_006539 | p.A138T | p.A138 | 错义突变 | 热点 |
| 星形细胞瘤 | CACNG3 | 16 | 24373167 | SNP | C | T | T | Oncomine | NM_006539 | p.R311C | p.R311 | 错义突变 | 热点 |
| 结肠直肠粘液腺癌 | CACNG3 | 16 | 24273772 | SNP | C | T | T | Oncomine | NM_006539 | p.A138V | p.A138 | 错义突变 | 热点 |
| 结肠直肠腺癌 | CACNG3 | 16 | 24273771 | SNP | G | A | A | Oncomine | NM_006539 | p.A138T | p.A138 | 错义突变 | 热点 |
| 肺腺癌 | CACNG3 | 16 | 24372930 | SNP | C | T | T | Oncomine | NM_006539 | p.R232W | p.R232 | 错义突变 | 热点 |
| 鳞状细胞肺癌 | CACNG3 | 16 | 24373168 | SNP | G | C | C | Oncomine | NM_006539 | p.R311P | p.R311 | 错义突变 | 热点 |
| 鳞状细胞肺癌 | CACNG3 | 16 | 24373168 | SNP | G | A | A | Oncomine | NM_006539 | p.R311H | p.R311 | 错义突变 | 热点 |
| 皮肤黑素瘤 | CACNG3 | 16 | 24372930 | SNP | C | T | T | Oncomine | NM_006539 | p.R232W | p.R232 | 错义突变 | 热点 |
| 皮肤黑素瘤 | CACNG3 | 16 | 24372868 | SNP | C | T | T | Oncomine | NM_006539 | p.S211F | p.S211 | 错义突变 | 热点 |
| 皮肤黑素瘤 | CCDC61 | 19 | 46498687 | SNP | G | A | A | Oncomine | NM_001080402 | p.E29K | p.E29 | 错义突变 | 热点 |
| 皮肤黑素瘤 | CCDC61 | 19 | 46498687 | SNP | C | T | T | Oncomine | NM_001080 | p.S33F | p.S33 | 错义突变 | 热点 |

[0176]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 00 | | | | | | 0402 | | | | |
| 皮肤黑素瘤 | CCDC61 | 19 | 46498687 | SNP | G | A | A | Oncomine | NM_001080402 | p.E29K | p.E29 | 错义突变 | 热点 |
| 前列腺癌 | CDC27 | 17 | 45234367 | SNP | A | A | T | Oncomine | NM_001256 | p.S252T | p.S252 | 错义突变 | 热点 |
| 皮肤黑素瘤 | CDC27 | 17 | 45234366 | SNP | G | A | A | Oncomine | NM_001256 | p.S252F | p.S252 | 错义突变 | 热点 |
| 嫌色性肾细胞癌 | CDC27 | 17 | 45234367 | SNP | A | A | T | Oncomine | NM_001256 | p.S252T | p.S252 | 错义突变 | 热点 |
| 皮肤黑素瘤 | CNTN5 | 11 | 100169975 | SNP | G | A | A | Oncomine | NM_014361 | p.E823K | p.E823 | 错义突变 | 热点 |
| 皮肤黑素瘤 | CNTN5 | 11 | 100170080 | SNP | G | A | A | Oncomine | NM_014361 | p.G858R | p.G858 | 错义突变 | 热点 |
| 皮肤黑素瘤 | CNTN5 | 11 | 99932099 | SNP | C | T | T | Oncomine | NM_014361 | p.S379F | p.S379 | 错义突变 | 热点 |
| 皮肤黑素瘤 | CNTN5 | 11 | 99715827 | SNP | G | A | A | Oncomine | NM_014361 | p.R137Q | p.R137 | 错义突变 | 热点 |
| 结肠直肠腺癌 | CNTN5 | 11 | 99221037 | SNP | G | T | T | Oncomine | NM_014361 | p.R137L | p.R137 | 错义突变 | 热点 |
| 结肠直肠腺癌 | CNTN5 | 11 | 99221037 | SNP | G | A | A | Oncomine | NM_014361 | p.R137Q | p.R137 | 错义突变 | 热点 |
| 皮肤黑素瘤 | CNTN5 | 11 | 99690287 | SNP | C | T | T | Oncomine | NM_014361 | p.S23F | p.S23 | 错义突变 | 热点 |
| 皮肤黑素瘤 | CNTN5 | 11 | 100169975 | SNP | G | A | A | Oncomine | NM_014361 | p.E823K | p.E823 | 错义突变 | 热点 |
| 皮肤黑素瘤 | CNTN5 | 11 | 99932099 | SNP | C | T | T | Oncomine | NM_014361 | p.S379F | p.S379 | 错义突变 | 热点 |
| 皮肤黑素瘤 | CNTN5 | 11 | 100170081 | SNP | G | A | A | Oncomine | NM_014361 | p.G858E | p.G858 | 错义突变 | 热点 |
| 皮肤黑素瘤 | CNTN5 | 11 | 99715827 | SNP | G | A | A | Oncomine | NM_014361 | p.R137Q | p.R137 | 错义突变 | 热点 |
| 皮肤黑素瘤 | CNTN5 | 11 | 100126527 | SNP | G | A | A | Oncomine | NM_014361 | p.E681K | p.E681 | 错义突变 | 热点 |

[0177]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 皮肤黑素瘤 | CNTN5 | 11 | 100170080 | SNP | G | A | A | Oncomine | NM_014361 | p.G858R | p.G858 | 错义突变 | 热点 |
| 星形细胞瘤 | CXCR2 | 2 | 219000407 | SNP | G | C | C | Oncomine | NM_001557 | p.A295P | p.A295 | 错义突变 | 热点 |
| 子宫内膜子宫内膜样腺癌 | CXCR2 | 2 | 218999763 | SNP | G | G | A | Oncomine | NM_001557 | p.R80H | p.R80 | 错义突变 | 热点 |
| 鳞状细胞肺癌 | CXCR2 | 2 | 218999763 | SNP | G | A | A | Oncomine | NM_001557 | p.R80H | p.R80 | 错义突变 | 热点 |
| 皮肤黑素瘤 | CXCR2 | 2 | 219000488 | SNP | C | T | T | Oncomine | NM_001557 | p.R322C | p.R322 | 错义突变 | 热点 |
| 皮肤黑素瘤 | CXCR2 | 2 | 219000408 | SNP | C | T | T | Oncomine | NM_001557 | p.A295V | p.A295 | 错义突变 | 热点 |
| 皮肤黑素瘤 | DCD | 12 | 55039462 | SNP | C | T | T | Oncomine | NM_053283 | p.E43K | p.E43 | 错义突变 | 热点 |
| 皮肤黑素瘤 | DCD | 12 | 55039462 | SNP | C | T | T | Oncomine | NM_053283 | p.E43K | p.E43 | 错义突变 | 热点 |
| 皮肤黑素瘤 | DSCR6 | 21 | 38390367 | SNP | G | A | A | Oncomine | NM_018962 | p.E145K | p.E145 | 错义突变 | 热点 |
| 肺腺癌 | DUX4L2 | 10 | 135491125 | SNP | G | A | A | Oncomine | NM_001127386 | p.A246T | p.A246 | 错义突变 | 热点 |
| 肺腺癌 | DUX4L2 | 10 | 135491123 | SNP | G | A | A | Oncomine | NM_001127386 | p.G245D | p.G245 | 错义突变 | 热点 |
| 浸润性膀胱尿道上皮癌 | DUX4L2 | 10 | 135491113 | SNP | G | T | T | Oncomine | NM_001127386 | p.A242S | p.A242 | 错义突变 | 热点 |
| 成胶质细胞瘤 | DUX4L2 | 10 | 135491113 | SNP | G | A | A | Oncomine | NM_001127386 | p.A242T | p.A242 | 错义突变 | 热点 |
| 成胶质细胞瘤 | DUX4L2 | 10 | 135491125 | SNP | G | A | A | Oncomine | NM_001127386 | p.A246T | p.A246 | 错义突变 | 热点 |
| 成胶质细胞瘤 | DUX4L2 | 10 | 135491123 | SNP | G | A | A | Oncomine | NM_001127386 | p.G245D | p.G245 | 错义突变 | 热点 |
| 星形细胞瘤 | DUX4L2 | 10 | 135491112 | SNP | C | A | A | Oncomine | NM_001127386 | p.F241L | p.F241 | 错义突变 | 热点 |

[0178]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 头颈鳞状细胞癌 | DUX4L2 | 10 | 135491125 | SNP | G | A | A | Oncomine | NM_001127386 | p.A246T | p.A246 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | DUX4L2 | 10 | 135491123 | SNP | G | A | A | Oncomine | NM_001127386 | p.G245D | p.G245 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | DUX4L2 | 10 | 135491112 | SNP | C | A | A | Oncomine | NM_001127386 | p.F241L | p.F241 | 错义突变 | 热点 |
| 皮肤黑素瘤 | DUX4L2 | 10 | 135491107 | SNP | G | A | A | Oncomine | NM_001127386 | p.A240T | p.A240 | 错义突变 | 热点 |
| 皮肤黑素瘤 | DUX4L2 | 10 | 135491123 | SNP | G | A | A | Oncomine | NM_001127386 | p.G245D | p.G245 | 错义突变 | 热点 |
| 皮肤黑素瘤 | DUX4L2 | 10 | 135491125 | SNP | G | A | A | Oncomine | NM_001127386 | p.A246T | p.A246 | 错义突变 | 热点 |
| 皮肤黑素瘤 | DUX4L2 | 10 | 135491113 | SNP | G | A | A | Oncomine | NM_001127386 | p.A242T | p.A242 | 错义突变 | 热点 |
| 皮肤黑素瘤 | DUX4L2 | 10 | 135491112 | SNP | C | A | A | Oncomine | NM_001127386 | p.F241L | p.F241 | 错义突变 | 热点 |
| 乳头状肾细胞癌 | DUX4L2 | 10 | 135491112 | SNP | C | A | A | Oncomine | NM_001127386 | p.F241L | p.F241 | 错义突变 | 热点 |
| 甲状腺乳头状癌 | DUX4L2 | 10 | 135491125 | SNP | G | A | A | Oncomine | NM_001127386 | p.A246T | p.A246 | 错义突变 | 热点 |
| 甲状腺乳头状癌 | DUX4L2 | 10 | 135491107 | SNP | G | A | A | Oncomine | NM_001127386 | p.A240T | p.A240 | 错义突变 | 热点 |
| 甲状腺乳头状癌 | DUX4L2 | 10 | 135491113 | SNP | G | A | A | Oncomine | NM_001127386 | p.A242T | p.A242 | 错义突变 | 热点 |
| 甲状腺乳头状癌 | DUX4L2 | 10 | 135491123 | SNP | G | A | A | Oncomine | NM_001127386 | p.G245D | p.G245 | 错义突变 | 热点 |
| 皮肤黑素瘤 | EDDM3A | 14 | 21216002 | SNP | G | A | A | Oncomine | NM_006683 | p.R88Q | p.R88 | 错义突变 | 热点 |
| 成胶质细胞瘤 | EDDM3A | 14 | 21216002 | SNP | G | A | A | Oncomine | NM_006683 | p.R88Q | p.R88 | 错义突变 | 热点 |
| 结肠直肠粘液腺癌 | EDDM3A | 14 | 20285842 | SNP | G | G | A | Oncomine | NM_006683 | p.R88Q | p.R88 | 错义突变 | 热点 |

[0179]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 皮肤黑素瘤 | EDDM3A | 14 | 21216002 | SNP | G | A | A | Oncomine | NM_006683 | p.R88Q | p.R88 | 错义突变 | 热点 |
| 导管乳癌 | ENDOU | 12 | 48110712 | SNP | G | G | A | Oncomine | NM_006025 | p.P130L | p.P130 | 错义突变 | 热点 |
| 子宫内膜子宫内膜样腺癌 | ENDOU | 12 | 48110712 | SNP | G | G | A | Oncomine | NM_006025 | p.P130L | p.P130 | 错义突变 | 热点 |
| 皮肤黑素瘤 | ENDOU | 12 | 48110713 | SNP | G | C | C | Oncomine | NM_006025 | p.P130A | p.P130 | 错义突变 | 热点 |
| 结肠直肠腺癌 | ERAS | X | 48572767 | SNP | C | T | T | Oncomine | NM_181532 | p.A97V | p.A97 | 错义突变 | 热点 |
| 子宫内膜子宫内膜样腺癌 | ERAS | X | 48687822 | SNP | G | G | A | Oncomine | NM_181532 | p.A97T | p.A97 | 错义突变 | 热点 |
| 肺腺癌 | ERAS | X | 48687822 | SNP | G | A | A | Oncomine | NM_181532 | p.A97T | p.A97 | 错义突变 | 热点 |
| 皮肤黑素瘤 | FABP1 | 2 | 88425751 | SNP | C | T | T | Oncomine | NM_001443 | p.E62K | p.E62 | 错义突变 | 热点 |
| 皮肤黑素瘤 | FABP1 | 2 | 88425751 | SNP | C | T | T | Oncomine | NM_001443 | p.E62K | p.E62 | 错义突变 | 热点 |
| 成神经管细胞瘤 | FAM22F | 9 | 97080945 | DEL | AGA | * | * | Oncomine | NM_017561 | p.S691_框内缺失 | p.S691_框内缺失 | 框内缺失 | 热点 |
| 宫颈鳞状细胞癌 | FAM22F | 9 | 97082793 | SNP | C | G | G | Oncomine | NM_017561 | p.K355N | p.K355 | 错义突变 | 热点 |
| 结肠直肠腺癌 | FAM22F | 9 | 96122614 | SNP | C | G | G | Oncomine | NM_017561 | p.K355N | p.K355 | 错义突变 | 热点 |
| 皮肤黑素瘤 | FAM22F | 9 | 97080945 | DEL | AGA | - | - | Oncomine | NM_017561 | p.S691_框内缺失 | p.S691_框内缺失 | 框内缺失 | 热点 |
| 前列腺腺癌 | FAM22F | 9 | 97080945 | DEL | AGA | - | - | Oncomine | NM_017561 | p.S691_框内缺失 | p.S691_框内缺失 | 框内缺失 | 热点 |

[0180]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 甲状腺腺体癌，NOS | FAM22F | 9 | 97080945 | DEL | AGA | - | - | Oncomine | NM_017561 | p.S691_框内缺失 | p.S691_框内缺失 | 框内缺失 | 热点 |
| 导管乳癌 | FBXW8 | 12 | 117465850 | SNP | G | G | A | Oncomine | NM_012174 | p.R491H | p.R491 | 错义突变 | 热点 |
| 结肠直肠腺癌 | FBXW8 | 12 | 115950233 | SNP | G | A | A | Oncomine | NM_012174 | p.R491H | p.R491 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | FBXW8 | 12 | 117465849 | SNP | C | T | T | Oncomine | NM_012174 | p.R491C | p.R491 | 错义突变 | 热点 |
| 鳞状细胞肺癌 | FBXW8 | 12 | 117465849 | SNP | C | T | T | Oncomine | NM_012174 | p.R491C | p.R491 | 错义突变 | 热点 |
| 皮肤黑素瘤 | FBXW8 | 12 | 117465849 | SNP | C | T | T | Oncomine | NM_012174 | p.R491C | p.R491 | 错义突变 | 热点 |
| 成胶质细胞瘤 | FHL3 | 1 | 38463709 | SNP | G | A | A | Oncomine | NM_004468 | p.P143S | p.P143 | 错义突变 | 热点 |
| 肺腺癌 | FHL3 | 1 | 38463709 | SNP | G | A | A | Oncomine | NM_004468 | p.P143S | p.P143 | 错义突变 | 热点 |
| 甲状腺乳头状癌 | FHL3 | 1 | 38463709 | SNP | G | C | C | Oncomine | NM_004468 | p.P143A | p.P143 | 错义突变 | 热点 |
| 结肠直肠腺癌 | GGT1 | 22 | 23340828 | SNP | G | A | A | Oncomine | NM_005265 | p.G84S | p.G84 | 错义突变 | 热点 |
| 皮肤黑素瘤 | GK2 | 4 | 80327859 | SNP | C | G | G | Oncomine | NM_033214 | p.R499P | p.R499 | 错义突变 | 热点 |
| 皮肤黑素瘤 | GK2 | 4 | 80328367 | SNP | G | A | A | Oncomine | NM_033214 | p.R330C | p.R330 | 错义突变 | 热点 |
| 皮肤黑素瘤 | GK2 | 4 | 80327860 | SNP | G | A | A | Oncomine | NM_033214 | p.R499C | p.R499 | 错义突变 | 热点 |
| 肺腺癌 | GK2 | 4 | 80328367 | SNP | G | A | A | Oncomine | NM_033214 | p.R330C | p.R330 | 错义突变 | 热点 |
| 成胶质细胞瘤 | GK2 | 4 | 80328891 | SNP | C | A | A | Oncomine | NM_033214 | p.R155L | p.R155 | 错义突变 | 热点 |
| 结肠直肠腺癌 | GK2 | 4 | 80547121 | SNP | G | A | A | Oncomine | NM_033214 | p.R420C | p.R420 | 错义突变 | 热点 |

[0181]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 子宫内膜子宫内膜样腺癌 | GK2 | 4 | 80328892 | SNP | G | G | A | Oncomine | NM_033214 | p.R155C | p.R155 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | GK2 | 4 | 80327860 | SNP | G | A | A | Oncomine | NM_033214 | p.R499C | p.R499 | 错义突变 | 热点 |
| 肺腺癌 | GK2 | 4 | 80328679 | SNP | G | A | A | Oncomine | NM_033214 | p.P226S | p.P226 | 错义突变 | 热点 |
| 鳞状细胞肺癌 | GK2 | 4 | 80328678 | SNP | G | A | A | Oncomine | NM_033214 | p.P226L | p.P226 | 错义突变 | 热点 |
| 皮肤黑素瘤 | GK2 | 4 | 80328892 | SNP | G | A | A | Oncomine | NM_033214 | p.R155C | p.R155 | 错义突变 | 热点 |
| 皮肤黑素瘤 | GK2 | 4 | 80328367 | SNP | G | A | A | Oncomine | NM_033214 | p.R330C | p.R330 | 错义突变 | 热点 |
| 皮肤黑素瘤 | GK2 | 4 | 80327860 | SNP | G | A | A | Oncomine | NM_033214 | p.R499C | p.R499 | 错义突变 | 热点 |
| 皮肤黑素瘤 | GK2 | 4 | 80328097 | SNP | G | A | A | Oncomine | NM_033214 | p.R420C | p.R420 | 错义突变 | 热点 |
| 皮肤黑素瘤 | GK2 | 4 | 80328679 | SNP | G | A | A | Oncomine | NM_033214 | p.P226S | p.P226 | 错义突变 | 热点 |
| 成胶质细胞瘤 | GOLGA6L10 | 15 | 83014132 | SNP | C | G | G | Oncomine | NM_001164465 | p.E151Q | p.E151 | 错义突变 | 热点 |
| 导管乳癌 | GOLGA6L10 | 15 | 83014132 | SNP | C | C | G | Oncomine | NM_001164465 | p.E151Q | p.E151 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | GOLGA6L10 | 15 | 83014132 | SNP | C | G | G | Oncomine | NM_001164465 | p.E151Q | p.E151 | 错义突变 | 热点 |
| 透明细胞肾细胞癌 | GOLGA6L10 | 15 | 83014132 | SNP | C | G | G | Oncomine | NM_001164465 | p.E151Q | p.E151 | 错义突变 | 热点 |
| 甲状腺乳头状癌 | GOLGA6L10 | 15 | 83014132 | SNP | C | G | G | Oncomine | NM_001164465 | p.E151Q | p.E151 | 错义突变 | 热点 |
| 皮肤黑素瘤 | GPX7 | 1 | 53072530 | SNP | C | T | T | Oncomine | NM_015696 | p.R105C | p.R105 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | GPX7 | 1 | 53072531 | SNP | G | T | T | Oncomine | NM_015696 | p.R105L | p.R105 | 错义突变 | 热点 |

[0182]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 肺腺癌 | GPX7 | 1 | 53072531 | SNP | G | A | A | Oncomine | NM_015696 | p.R105H | p.R105 | 错义突变 | 热点 |
| 皮肤黑素瘤 | GTSF1 | 12 | 54858877 | SNP | G | A | A | Oncomine | NM_144594 | p.P31S | p.P31 | 错义突变 | 热点 |
| 皮肤黑素瘤 | GTSF1 | 12 | 54858877 | SNP | G | A | A | Oncomine | NM_144594 | p.P31S | p.P31 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | H3F3A | 1 | 226252059 | SNP | C | T | T | Oncomine | NM_002107 | p.R3C | p.R3 | 错义突变 | 热点 |
| 星形细胞瘤 | H3F3A | 1 | 226252059 | SNP | C | T | T | Oncomine | NM_002107 | p.R3C | p.R3 | 错义突变 | 热点 |
| 宫颈鳞状细胞癌 | H3F3A | 1 | 226252059 | SNP | C | T | T | Oncomine | NM_002107 | p.R3C | p.R3 | 错义突变 | 热点 |
| 小细胞肺癌 | HDDC2 | 6 | 125661566 | SNP | C | G | G | Oncomine | NM_016063 | p.R101P | p.R101 | 错义突变 | 热点 |
| 小细胞肺癌 | HDDC2 | 6 | 125619867 | SNP | C | G | G | Oncomine | NM_016063 | p.R101P | p.R101 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | HDDC2 | 6 | 125619867 | SNP | C | T | T | Oncomine | NM_016063 | p.R101Q | p.R101 | 错义突变 | 热点 |
| 皮肤黑素瘤 | HEXDC | 17 | 80400154 | SNP | A | C | C | Oncomine | NM_173620 | p.T482P | p.T482 | 错义突变 | 热点 |
| 鳞状细胞肺癌 | HEXDC | 17 | 80400154 | SNP | A | C | C | Oncomine | NM_173620 | p.T482P | p.T482 | 错义突变 | 热点 |
| 透明细胞肾细胞癌 | HEXDC | 17 | 80400154 | SNP | A | C | C | Oncomine | NM_173620 | p.T482P | p.T482 | 错义突变 | 热点 |
| 小细胞肺癌 | HIST1H4C | 6 | 26212357 | SNP | G | C | C | Oncomine | NM_003542 | p.R68P | p.R68 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | HIST1H4C | 6 | 26104378 | SNP | G | C | C | Oncomine | NM_003542 | p.R68P | p.R68 | 错义突变 | 热点 |
| 皮肤黑素瘤 | HNRNPCL1 | 1 | 12907971 | SNP | C | T | T | Oncomine | NM_001013631 | p.D58N | p.D58 | 错义突变 | 热点 |
| 黑素瘤 | HNRNPCL1 | 1 | 12907847 | SNP | C | T | T | Oncomine | NM_001013631 | p.R99Q | p.R99 | 错义突变 | 热点 |

[0183]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 结肠直肠腺癌 | HNRNPCL1 | 1 | 12830231 | SNP | G | A | A | Oncomine | NM_001013631 | p.R167W | p.R167 | 错义突变 | 热点 |
| 子宫内膜子宫内膜样腺癌 | HNRNPCL1 | 1 | 12907644 | SNP | G | G | A | Oncomine | NM_001013631 | p.R167W | p.R167 | 错义突变 | 热点 |
| 胃腺癌 | HNRNPCL1 | 1 | 12907847 | SNP | C | T | T | Oncomine | NM_001013631 | p.R99Q | p.R99 | 错义突变 | 热点 |
| 肺腺癌 | HNRNPCL1 | 1 | 12907643 | SNP | C | A | A | Oncomine | NM_001013631 | p.R167L | p.R167 | 错义突变 | 热点 |
| 皮肤黑素瘤 | HNRNPCL1 | 1 | 12907847 | SNP | C | T | T | Oncomine | NM_001013631 | p.R99Q | p.R99 | 错义突变 | 热点 |
| 皮肤黑素瘤 | HNRNPCL1 | 1 | 12907865 | SNP | C | T | T | Oncomine | NM_001013631 | p.G93E | p.G93 | 错义突变 | 热点 |
| 皮肤黑素瘤 | HNRNPCL1 | 1 | 12907971 | SNP | C | T | T | Oncomine | NM_001013631 | p.D58N | p.D58 | 错义突变 | 热点 |
| 前列腺腺癌 | HRCT1 | 9 | 35906559 | SNP | A | C | C | Oncomine | NM_001039792 | p.H92P | p.H92 | 错义突变 | 热点 |
| 成胶质细胞瘤 | HRCT1 | 9 | 35906348 | DEL | CTG | - | - | Oncomine | NM_001039792 | p.L22_内缺失 | p.L22_框内缺失 | 框内缺失 | 热点 |
| 导管乳癌 | HRCT1 | 9 | 35906348 | DEL | CTG | CTG | - | Oncomine | NM_001039792 | p.L22_内缺失 | p.L22_框内缺失 | 框内缺失 | 热点 |
| 宫颈鳞状细胞癌 | HRCT1 | 9 | 35906559 | SNP | A | C | C | Oncomine | NM_001039792 | p.H92P | p.H92 | 错义突变 | 热点 |
| 胃腺癌 | HRCT1 | 9 | 35906584 | DEL | CCA | - | - | Oncomine | NM_001039792 | p.L100_内缺失 | p.L100_框内缺失 | 框内缺失 | 热点 |
| 皮肤黑素瘤 | HRCT1 | 9 | 35906348 | DEL | CTG | - | - | Oncomine | NM_001039792 | p.L22_内缺失 | p.L22_框内缺失 | 框内缺失 | 热点 |
| 皮肤黑素瘤 | HRCT1 | 9 | 35906559 | SNP | A | C | C | Oncomine | NM_001039792 | p.H92P | p.H92 | 错义突变 | 热点 |

[0184]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 乳头状肾细胞癌 | HRCT1 | 9 | 35906584 | DEL | CCA | - | - | Oncomine | NM_001039792 | p.L100_框内缺失 | p.L100_框内缺失 | 框内缺失 | 热点 |
| 乳头状肾细胞癌 | HRCT1 | 9 | 35906559 | SNP | A | C | C | Oncomine | NM_001039792 | p.H92P | p.H92 | 错义突变 | 热点 |
| 甲状腺腺体癌，NOS | HRCT1 | 9 | 35906584 | DEL | CCA | - | - | Oncomine | NM_001039792 | p.L100_框内缺失 | p.L100_框内缺失 | 框内缺失 | 热点 |
| 结肠直肠腺癌 | IL3 | 5 | 131425967 | SNP | G | A | A | Oncomine | NM_000588 | p.A90T | p.A90 | 错义突变 | 热点 |
| 胃腺癌 | IL3 | 5 | 131398068 | SNP | G | A | A | Oncomine | NM_000588 | p.A90T | p.A90 | 错义突变 | 热点 |
| 胰腺导管腺癌 | JAM3 | 11 | 134014849 | SNP | G | A | G | Oncomine | NM_032801 | p.R191H | p.R191 | 错义突变 | 热点 |
| 小叶乳癌 | JAM3 | 11 | 134014849 | SNP | G | G | A | Oncomine | NM_032801 | p.R191H | p.R191 | 错义突变 | 热点 |
| 胃腺癌 | JAM3 | 11 | 134014848 | SNP | C | T | T | Oncomine | NM_032801 | p.R191C | p.R191 | 错义突变 | 热点 |
| 皮肤黑素瘤 | KCNK9 | 8 | 140631316 | SNP | C | T | T | Oncomine | NM_016601 | p.D104N | p.D104 | 错义突变 | 热点 |
| 子宫内膜子宫内膜样腺癌 | KCNK9 | 8 | 140630833 | SNP | C | C | T | Oncomine | NM_016601 | p.A265T | p.A265 | 错义突变 | 热点 |
| 肺腺癌 | KCNK9 | 8 | 140630832 | SNP | G | A | A | Oncomine | NM_016601 | p.A265V | p.A265 | 错义突变 | 热点 |
| 鳞状细胞肺癌 | KCNK9 | 8 | 140630833 | SNP | C | T | T | Oncomine | NM_016601 | p.A265T | p.A265 | 错义突变 | 热点 |
| 皮肤黑素瘤 | KCNK9 | 8 | 140631316 | SNP | C | T | T | Oncomine | NM_016601 | p.D104N | p.D104 | 错义突变 | 热点 |
| 成胶质细胞瘤 | KLK6 | 19 | 51466671 | SNP | C | T | T | Oncomine | NM_002774 | p.R111H | p.R111 | 错义突变 | 热点 |
| 结肠直肠粘液腺癌 | KLK6 | 19 | 56158484 | SNP | G | A | A | Oncomine | NM_002774 | p.R111C | p.R111 | 错义突变 | 热点 |

[0185]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 子宫内膜子宫内膜样腺癌 | KLK6 | 19 | 51466671 | SNP | C | C | T | Oncomine | NM_002774 | p.R111H | p.R111 | 错义突变 | 热点 |
| 皮肤黑素瘤 | KLK6 | 19 | 51462556 | SNP | G | A | A | Oncomine | NM_002774 | p.P200L | p.P200 | 错义突变 | 热点 |
| 前列腺腺癌 | KLK6 | 19 | 51462556 | SNP | G | A | A | Oncomine | NM_002774 | p.P200L | p.P200 | 错义突变 | 热点 |
| 透明细胞肾细胞癌 | KLK6 | 19 | 51462556 | SNP | G | A | A | Oncomine | NM_002774 | p.P200L | p.P200 | 错义突变 | 热点 |
| 结肠直肠腺癌 | KRTAP12-4 | 21 | 44898950 | SNP | T | G | G | Oncomine | NM_198698 | p.T4P | p.T4 | 错义突变 | 热点 |
| 卵巢浆液性腺癌 | KRTAP12-4 | 21 | 44898949 | SNP | G | G | A | Oncomine | NM_198698 | p.T4I | p.T4 | 错义突变 | 热点 |
| 皮肤黑素瘤 | KRTAP4-11 | 17 | 39274424 | SNP | G | C | C | Oncomine | NM_033059 | p.S48R | p.S48 | 错义突变 | 热点 |
| 皮肤黑素瘤 | KRTAP4-11 | 17 | 39274206 | SNP | C | T | T | Oncomine | NM_033059 | p.R121K | p.R121 | 错义突变 | 热点 |
| 肺腺癌 | KRTAP4-11 | 17 | 39274150 | SNP | T | A | A | Oncomine | NM_033059 | p.S140C | p.S140 | 错义突变 | 热点 |
| 肺腺癌 | KRTAP4-11 | 17 | 39274206 | SNP | C | T | T | Oncomine | NM_033059 | p.R121K | p.R121 | 错义突变 | 热点 |
| 成胶质细胞瘤 | KRTAP4-11 | 17 | 39274424 | SNP | G | C | C | Oncomine | NM_033059 | p.S48R | p.S48 | 错义突变 | 热点 |
| 少突神经胶质瘤 | KRTAP4-11 | 17 | 39274087 | SNP | G | C | C | Oncomine | NM_033059 | p.L161V | p.L161 | 错义突变 | 热点 |
| 导管乳癌 | KRTAP4-11 | 17 | 39274087 | SNP | G | G | C | Oncomine | NM_033059 | p.L161V | p.L161 | 错义突变 | 热点 |
| 宫颈鳞状细胞癌 | KRTAP4-11 | 17 | 39274424 | SNP | G | C | C | Oncomine | NM_033059 | p.S48R | p.S48 | 错义突变 | 热点 |
| 宫颈鳞状细胞癌 | KRTAP4-11 | 17 | 39274087 | SNP | G | C | C | Oncomine | NM_033059 | p.L161V | p.L161 | 错义突变 | 热点 |
| 宫颈鳞状细胞癌 | KRTAP4-11 | 17 | 39274150 | SNP | T | A | A | Oncomine | NM_033059 | p.S140C | p.S140 | 错义突变 | 热点 |

[0186]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 头颈鳞状细胞癌 | KRTAP4-11 | 17 | 39274424 | SNP | G | C | C | Oncomine | NM_033059 | p.S48R | p.S48 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | KRTAP4-11 | 17 | 39274206 | SNP | C | T | T | Oncomine | NM_033059 | p.R121K | p.R121 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | KRTAP4-11 | 17 | 39274087 | SNP | G | C | C | Oncomine | NM_033059 | p.L161V | p.L161 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | KRTAP4-11 | 17 | 39274150 | SNP | T | A | A | Oncomine | NM_033059 | p.S140C | p.S140 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | KRTAP4-11 | 17 | 39274291 | SNP | T | C | C | Oncomine | NM_033059 | p.M93V | p.M93 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | KRTAP4-11 | 17 | 39274416 | SNP | C | T | T | Oncomine | NM_033059 | p.R51K | p.R51 | 错义突变 | 热点 |
| 肺腺癌 | KRTAP4-11 | 17 | 39274424 | SNP | G | C | C | Oncomine | NM_033059 | p.S48R | p.S48 | 错义突变 | 热点 |
| 肺腺癌 | KRTAP4-11 | 17 | 39274206 | SNP | C | T | T | Oncomine | NM_033059 | p.R121K | p.R121 | 错义突变 | 热点 |
| 皮肤黑素瘤 | KRTAP4-11 | 17 | 39274150 | SNP | T | A | A | Oncomine | NM_033059 | p.S140C | p.S140 | 错义突变 | 热点 |
| 皮肤黑素瘤 | KRTAP4-11 | 17 | 39274206 | SNP | C | T | T | Oncomine | NM_033059 | p.R121K | p.R121 | 错义突变 | 热点 |
| 皮肤黑素瘤 | KRTAP4-11 | 17 | 39274424 | SNP | G | C | C | Oncomine | NM_033059 | p.S48R | p.S48 | 错义突变 | 热点 |
| 皮肤黑素瘤 | KRTAP4-11 | 17 | 39274087 | SNP | G | C | C | Oncomine | NM_033059 | p.L161V | p.L161 | 错义突变 | 热点 |
| 皮肤黑素瘤 | KRTAP4-11 | 17 | 39274416 | SNP | C | T | T | Oncomine | NM_033059 | p.R51K | p.R51 | 错义突变 | 热点 |
| 透明细胞肾细胞癌 | KRTAP4-11 | 17 | 39274291 | SNP | T | C | C | Oncomine | NM_033059 | p.M93V | p.M93 | 错义突变 | 热点 |
| 透明细胞肾细胞癌 | KRTAP4-11 | 17 | 39274206 | SNP | C | T | T | Oncomine | NM_033059 | p.R121K | p.R121 | 错义突变 | 热点 |
| 透明细胞肾细胞癌 | KRTAP4-11 | 17 | 39274150 | SNP | T | A | A | Oncomine | NM_033059 | p.S140C | p.S140 | 错义突变 | 热点 |

[0187]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 乳头状肾细胞癌 | KRTAP4-11 | 17 | 39274087 | SNP | G | C | C | Oncomine | NM_033059 | p.L161V | p.L161 | 错义突变 | 热点 |
| 甲状腺乳头状癌 | KRTAP4-11 | 17 | 39274424 | SNP | G | C | C | Oncomine | NM_033059 | p.S48R | p.S48 | 错义突变 | 热点 |
| 乳头状肾细胞癌 | KRTAP4-7 | 17 | 39240900 | SNP | T | G | G | Oncomine | NM_033061 | p.L148V | p.L148 | 错义突变 | 热点 |
| 皮肤黑素瘤 | LAD1 | 1 | 201354881 | SNP | C | T | T | Oncomine | NM_005558 | p.R360Q | p.R360 | 错义突变 | 热点 |
| 皮肤黑素瘤 | LAD1 | 1 | 201352246 | SNP | C | T | T | Oncomine | NM_005558 | p.E448K | p.E448 | 错义突变 | 热点 |
| 透明细胞肾细胞癌 | LAD1 | 1 | 201354881 | SNP | C | A | A | Oncomine | NM_005558 | p.R360L | p.R360 | 错义突变 | 热点 |
| 黑素瘤 | LELP1 | 1 | 153177244 | SNP | C | T | T | Oncomine | NM_001010857 | p.P21S | p.P21 | 错义突变 | 热点 |
| 皮肤黑素瘤 | LELP1 | 1 | 153177437 | SNP | C | T | T | Oncomine | NM_001010857 | p.S85F | p.S85 | 错义突变 | 热点 |
| 皮肤黑素瘤 | LELP1 | 1 | 153177245 | SNP | C | T | T | Oncomine | NM_001010857 | p.P21L | p.P21 | 错义突变 | 热点 |
| 皮肤黑素瘤 | LELP1 | 1 | 153177244 | SNP | C | T | T | Oncomine | NM_001010857 | p.P21S | p.P21 | 错义突变 | 热点 |
| 皮肤黑素瘤 | LOC100509575 | X | 47972582 | SNP | G | A | A | Oncomine | NM_001205103 | p.R96H | p.R96 | 错义突变 | 热点 |
| 小叶乳癌 | LOC100509575 | X | 47972582 | SNP | G | G | A | Oncomine | NM_001205103 | p.R96H | p.R96 | 错义突变 | 热点 |
| 子宫内膜子宫内膜样腺癌 | LOC100509575 | X | 47972581 | SNP | C | C | T | Oncomine | NM_001205103 | p.R96C | p.R96 | 错义突变 | 热点 |
| 成胶质细胞瘤 | MUC4 | 3 | 195516064 | SNP | C | T | T | Oncomine | NM_018406 | p.R796Q | p.R796 | 错义突变 | 热点 |
| 导管乳癌 | MUC4 | 3 | 195516064 | SNP | C | C | T | Oncomine | NM_018406 | p.R796Q | p.R796 | 错义突变 | 热点 |
| 肺腺癌 | MUC4 | 3 | 195516064 | SNP | C | T | T | Oncomine | NM_018406 | p.R796Q | p.R796 | 错义突变 | 热点 |

[0188]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 成胶质细胞瘤 | NAB2 | 12 | 57485446 | SNP | T | C | C | Oncomine | NM_005967 | p.F208L | p.F208 | 错义突变 | 热点 |
| 少突神经胶质瘤 | NAB2 | 12 | 57485446 | SNP | T | C | C | Oncomine | NM_005967 | p.F208L | p.F208 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | NAB2 | 12 | 57485446 | SNP | T | C | C | Oncomine | NM_005967 | p.F208L | p.F208 | 错义突变 | 热点 |
| 肺腺癌 | NAB2 | 12 | 57485446 | SNP | T | C | C | Oncomine | NM_005967 | p.F208L | p.F208 | 错义突变 | 热点 |
| 皮肤黑素瘤 | NAB2 | 12 | 57485446 | SNP | T | C | C | Oncomine | NM_005967 | p.F208L | p.F208 | 错义突变 | 热点 |
| 成胶质细胞瘤 | NBPF10 | 1 | 145324371 | SNP | T | C | C | Oncomine | NM_001039703 | p.V1189A | p.V1189 | 错义突变 | 热点 |
| 星形细胞瘤 | NBPF10 | 1 | 145360584 | SNP | G | A | A | Oncomine | NM_001039703 | p.G3070E | p.G3070 | 错义突变 | 热点 |
| 皮肤黑素瘤 | NBPF10 | 1 | 145360584 | SNP | G | A | A | Oncomine | NM_001039703 | p.G3070E | p.G3070 | 错义突变 | 热点 |
| 皮肤黑素瘤 | NSFL1C | 20 | 142636 0 | SNP | G | A | A | Oncomine | NM_016143 | p.R301W | p.R301 | 错义突变 | 热点 |
| 结肠直肠腺癌 | NSFL1C | 20 | 137436 0 | SNP | G | A | A | Oncomine | NM_016143 | p.R301W | p.R301 | 错义突变 | 热点 |
| 子宫内膜子宫内膜样腺癌 | NSFL1C | 20 | 142636 0 | SNP | G | G | A | Oncomine | NM_016143 | p.R301W | p.R301 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | NSFL1C | 20 | 142636 0 | SNP | G | A | A | Oncomine | NM_016143 | p.R301W | p.R301 | 错义突变 | 热点 |
| 成神经管细胞瘤 | OBP2B | 9 | 136081795 | SNP | A | G | G | Oncomine | NM_014581 | p.S133P | p.S133 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | OBP2B | 9 | 136081795 | SNP | A | G | G | Oncomine | NM_014581 | p.S133P | p.S133 | 错义突变 | 热点 |
| 肺腺癌 | OBP2B | 9 | 136081795 | SNP | A | G | G | Oncomine | NM_014581 | p.S133P | p.S133 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OR2L13 | 1 | 248262729 | SNP | C | A | A | Oncomine | NM_175911 | p.P18T | p.P18 | 错义突变 | 热点 |

[0189]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 皮肤黑素瘤 | OR2L13 | 1 | 248263173 | SNP | C | T | T | Oncomine | NM_175911 | p.P166S | p.P166 | 错义突变 | 热点 |
| 肺腺癌 | OR2L13 | 1 | 248263401 | SNP | A | G | G | Oncomine | NM_175911 | p.T242A | p.T242 | 错义突变 | 热点 |
| 小细胞肺癌 | OR2L13 | 1 | 248262832 | SNP | C | A | A | Oncomine | NM_175911 | p.P52H | p.P52 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | OR2L13 | 1 | 248262831 | SNP | C | T | T | Oncomine | NM_175911 | p.P52S | p.P52 | 错义突变 | 热点 |
| 肺腺癌 | OR2L13 | 1 | 248263401 | SNP | A | T | T | Oncomine | NM_175911 | p.T242S | p.T242 | 错义突变 | 热点 |
| 肺腺癌 | OR2L13 | 1 | 248263401 | SNP | A | G | G | Oncomine | NM_175911 | p.T242A | p.T242 | 错义突变 | 热点 |
| 鳞状细胞肺癌 | OR2L13 | 1 | 248262831 | SNP | C | T | T | Oncomine | NM_175911 | p.P52S | p.P52 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OR2L13 | 1 | 248263371 | SNP | G | A | A | Oncomine | NM_175911 | p.G232R | p.G232 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OR2L13 | 1 | 248263174 | SNP | C | T | T | Oncomine | NM_175911 | p.P166L | p.P166 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OR2L13 | 1 | 248263173 | SNP | C | T | T | Oncomine | NM_175911 | p.P166S | p.P166 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OR2L13 | 1 | 248262730 | SNP | C | T | T | Oncomine | NM_175911 | p.P18L | p.P18 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OR2L13 | 1 | 248262729 | SNP | C | A | A | Oncomine | NM_175911 | p.P18T | p.P18 | 错义突变 | 热点 |
| 卵巢浆液性腺癌 | OR2L13 | 1 | 246329995 | SNP | G | G | A | Oncomine | NM_175911 | p.G232E | p.G232 | 错义突变 | 热点 |
| 结肠直肠腺癌 | OR2T27 | 1 | 246880778 | SNP | C | T | T | Oncomine | NM_001001824 | p.D11N | p.D11 | 错义突变 | 热点 |
| 子宫内膜浆液性腺癌 | OR2T27 | 1 | 248813822 | SNP | G | G | A | Oncomine | NM_001001824 | p.R122C | p.R122 | 错义突变 | 热点 |
| 胃腺癌 | OR2T27 | 1 | 248813821 | SNP | C | T | T | Oncomine | NM_001001824 | p.R122H | p.R122 | 错义突变 | 热点 |

[0190]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 胃腺癌 | OR2T27 | 1 | 248813773 | SNP | C | T | T | Oncomine | NM_001001824 | p.R138H | p.R138 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | OR2T27 | 1 | 248813773 | SNP | C | G | G | Oncomine | NM_001001824 | p.R138P | p.R138 | 错义突变 | 热点 |
| 肺腺癌 | OR2T27 | 1 | 248814155 | SNP | C | A | A | Oncomine | NM_001001824 | p.D11Y | p.D11 | 错义突变 | 热点 |
| 肺腺癌 | OR2T27 | 1 | 248813773 | SNP | C | A | A | Oncomine | NM_001001824 | p.R138L | p.R138 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OR2Z1 | 19 | 8841458 | SNP | C | T | T | Oncomine | NM_001004699 | p.S23L | p.S23 | 错义突变 | 热点 |
| 成胶质细胞瘤 | OR2Z1 | 19 | 8841802 | SNP | C | T | T | Oncomine | NM_001004699 | p.R138C | p.R138 | 错义突变 | 热点 |
| 胃腺癌 | OR2Z1 | 19 | 8841802 | SNP | C | T | T | Oncomine | NM_001004699 | p.R138C | p.R138 | 错义突变 | 热点 |
| 肺腺癌 | OR2Z1 | 19 | 8841802 | SNP | C | T | T | Oncomine | NM_001004699 | p.R138C | p.R138 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OR2Z1 | 19 | 8841458 | SNP | C | T | T | Oncomine | NM_001004699 | p.S23L | p.S23 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OR2Z1 | 19 | 8841889 | SNP | C | T | T | Oncomine | NM_001004699 | p.P167S | p.P167 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OR4E2 | 14 | 22133748 | SNP | G | A | A | Oncomine | NM_001001912 | p.G151E | p.G151 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OR4E2 | 14 | 22133973 | SNP | G | A | A | Oncomine | NM_001001912 | p.R226Q | p.R226 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OR4E2 | 14 | 22133747 | SNP | G | A | A | Oncomine | NM_001001912 | p.G151R | p.G151 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OR4E2 | 14 | 22133748 | SNP | G | A | A | Oncomine | NM_001001912 | p.G151E | p.G151 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OR51B2 | 11 | 5345263 | SNP | C | T | T | Oncomine | NM_033180 | p.E89K | p.E89 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OR51B2 | 11 | 534504 0 | SNP | G | A | A | Oncomine | NM_033180 | p.S163L | p.S163 | 错义突变 | 热点 |
| 肺腺癌 | OR51B2 | 11 | 534477 | SNP | G | T | T | Oncomine | NM_03318 | p.T252K | p.T252 | 错义突变 | 热点 |

[0191]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 3 | | | | | | 0 | | | | |
| 成胶质细胞瘤 | OR51B2 | 11 | 5344773 | SNP | G | A | A | Oncomine | NM_033180 | p.T252I | p.T252 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | OR51B2 | 11 | 5344774 | SNP | T | C | C | Oncomine | NM_033180 | p.T252A | p.T252 | 错义突变 | 热点 |
| 鳞状细胞肺癌 | OR51B2 | 11 | 5345101 | SNP | C | T | T | Oncomine | NM_033180 | p.G143R | p.G143 | 错义突变 | 热点 |
| 鳞状细胞肺癌 | OR51B2 | 11 | 5345100 | SNP | C | A | A | Oncomine | NM_033180 | p.G143V | p.G143 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OR51B2 | 11 | 5345263 | SNP | C | T | T | Oncomine | NM_033180 | p.E89K | p.E89 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OR51B2 | 11 | 5345040 | SNP | G | A | A | Oncomine | NM_033180 | p.S163L | p.S163 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OR51B2 | 11 | 5345100 | SNP | C | T | T | Oncomine | NM_033180 | p.G143E | p.G143 | 错义突变 | 热点 |
| 成胶质细胞瘤 | OR52A1 | 11 | 5172692 | SNP | C | T | T | Oncomine | NM_012375 | p.R303H | p.R303 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OR52A1 | 11 | 5172912 | SNP | G | A | A | Oncomine | NM_012375 | p.R230C | p.R230 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OR52A1 | 11 | 5172693 | SNP | G | A | A | Oncomine | NM_012375 | p.R303C | p.R303 | 错义突变 | 热点 |
| 前列腺腺癌 | OR52A1 | 11 | 5172911 | SNP | C | T | T | Oncomine | NM_012375 | p.R230H | p.R230 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OR5AN1 | 11 | 59132584 | SNP | C | T | T | Oncomine | NM_001004729 | p.S218F | p.S218 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OR6T1 | 11 | 123814182 | SNP | G | A | A | Oncomine | NM_001005187 | p.R122C | p.R122 | 错义突变 | 热点 |
| 导管乳癌 | OR6T1 | 11 | 123813896 | SNP | G | G | T | Oncomine | NM_001005187 | p.S217Y | p.S217 | 错义突变 | 热点 |
| 结肠直肠腺癌 | OR6T1 | 11 | 123318974 | SNP | C | C | T | Oncomine | NM_001005187 | p.R261H | p.R261 | 错义突变 | 热点 |
| 结肠直肠粘液腺 | OR6T1 | 11 | 123319221 | SNP | G | A | A | Oncomine | NM_001005187 | p.R179C | p.R179 | 错义突变 | 热点 |

[0192]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 癌 | | | | | | | | | | | | | |
| 结肠直肠腺癌 | OR6T1 | 11 | 123319106 | SNP | G | T | T | Oncomine | NM_001005187 | p.S217Y | p.S217 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OR6T1 | 11 | 123813765 | SNP | G | A | A | Oncomine | NM_001005187 | p.R261C | p.R261 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OR6T1 | 11 | 123814011 | SNP | G | A | A | Oncomine | NM_001005187 | p.R179C | p.R179 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OR6T1 | 11 | 123813896 | SNP | G | A | A | Oncomine | NM_001005187 | p.S217F | p.S217 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OR6T1 | 11 | 123814182 | SNP | G | A | A | Oncomine | NM_001005187 | p.R122C | p.R122 | 错义突变 | 热点 |
| 透明细胞肾细胞癌 | OR6T1 | 11 | 123814182 | SNP | G | T | T | Oncomine | NM_001005187 | p.R122S | p.R122 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OTUD5 | X | 48792073 | SNP | C | T | T | Oncomine | NM_017602 | p.R274Q | p.R274 | 错义突变 | 热点 |
| 结肠直肠腺癌 | OTUD5 | X | 48668111 | SNP | G | A | A | Oncomine | NM_017602 | p.R412W | p.R412 | 错义突变 | 热点 |
| 结肠直肠腺癌 | OTUD5 | X | 48677018 | SNP | G | A | A | Oncomine | NM_017602 | p.R274W | p.R274 | 错义突变 | 热点 |
| 子宫内膜子宫内膜样腺癌 | OTUD5 | X | 48792074 | SNP | G | G | A | Oncomine | NM_017602 | p.R274W | p.R274 | 错义突变 | 热点 |
| 子宫内膜子宫内膜样腺癌 | OTUD5 | X | 48783167 | SNP | G | G | A | Oncomine | NM_017602 | p.R412W | p.R412 | 错义突变 | 热点 |
| 肺腺癌 | OTUD5 | X | 48783166 | SNP | C | A | A | Oncomine | NM_017602 | p.R412L | p.R412 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OXA1L | 14 | 23235902 | SNP | C | T | T | Oncomine | NM_005015 | p.P58S | p.P58 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OXA1L | 14 | 23235899 | SNP | C | T | T | Oncomine | NM_005015 | p.L57F | p.L57 | 错义突变 | 热点 |
| 皮肤黑素瘤 | OXA1L | 14 | 23235902 | SNP | C | T | T | Oncomine | NM_005015 | p.P58S | p.P58 | 错义突变 | 热点 |
| 宫颈鳞状细胞癌 | PBX2 | 6 | 321555 | SNP | T | A | A | Oncomine | NM_002585 | p.Y262F | p.Y262 | 错义突变 | 热点 |

[0193]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 09 | | | | | | 6 | | | | |
| 胃腺癌 | PBX2 | 6 | 32155509 | SNP | T | A | A | Oncomine | NM_002586 | p.Y262F | p.Y262 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | PBX2 | 6 | 32155509 | SNP | T | A | A | Oncomine | NM_002586 | p.Y262F | p.Y262 | 错义突变 | 热点 |
| 鳞状细胞肺癌 | PBX2 | 6 | 32155509 | SNP | T | A | A | Oncomine | NM_002586 | p.Y262F | p.Y262 | 错义突变 | 热点 |
| 透明细胞肾细胞癌 | PBX2 | 6 | 32155509 | SNP | T | A | A | Oncomine | NM_002586 | p.Y262F | p.Y262 | 错义突变 | 热点 |
| 前列腺腺癌 | PDHA2 | 4 | 96761513 | SNP | G | A | A | Oncomine | NM_005390 | p.R71H | p.R71 | 错义突变 | 热点 |
| 黑素瘤 | PDHA2 | 4 | 96761738 | SNP | G | A | A | Oncomine | NM_005390 | p.G146E | p.G146 | 错义突变 | 热点 |
| 皮肤黑素瘤 | PDHA2 | 4 | 96761737 | SNP | G | A | A | Oncomine | NM_005390 | p.G146R | p.G146 | 错义突变 | 热点 |
| 成胶质细胞瘤 | PDHA2 | 4 | 96761557 | SNP | C | T | T | Oncomine | NM_005390 | p.R86C | p.R86 | 错义突变 | 热点 |
| 结肠直肠腺癌 | PDHA2 | 4 | 96980580 | SNP | C | T | T | Oncomine | NM_005390 | p.R86C | p.R86 | 错义突变 | 热点 |
| 子宫内膜浆液性腺癌 | PDHA2 | 4 | 96761738 | SNP | G | G | A | Oncomine | NM_005390 | p.G146E | p.G146 | 错义突变 | 热点 |
| 肺腺癌 | PDHA2 | 4 | 96761513 | SNP | G | A | A | Oncomine | NM_005390 | p.R71H | p.R71 | 错义突变 | 热点 |
| 皮肤黑素瘤 | PDHA2 | 4 | 96761854 | SNP | G | A | A | Oncomine | NM_005390 | p.D185N | p.D185 | 错义突变 | 热点 |
| 皮肤黑素瘤 | PDHA2 | 4 | 96761738 | SNP | G | A | A | Oncomine | NM_005390 | p.G146E | p.G146 | 错义突变 | 热点 |
| 甲状腺腺体癌，NOS | PDHA2 | 4 | 96761513 | SNP | G | A | A | Oncomine | NM_005390 | p.R71H | p.R71 | 错义突变 | 热点 |
| 甲状腺乳头状癌 | PDHA2 | 4 | 96761557 | SNP | C | T | T | Oncomine | NM_005390 | p.R86C | p.R86 | 错义突变 | 热点 |
| 肺腺癌 | POTEC | 18 | 145430 | SNP | T | C | C | Oncomine | NM_001113 | p.M43V | p.M43 | 错义突变 | 热点 |

[0194]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 19 | | | | | | 7671 | | | | |
| 成胶质细胞瘤 | POTEC | 18 | 14543019 | SNP | T | C | C | Oncomine | NM_00113 7671 | p.M43V | p.M43 | 错义突变 | 热点 |
| 星形细胞瘤 | POTEC | 18 | 14513734 | SNP | C | T | T | Oncomine | NM_00113 7671 | p.G487E | p.G487 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | POTEC | 18 | 14513734 | SNP | C | T | T | Oncomine | NM_00113 7671 | p.G487E | p.G487 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | POTEC | 18 | 14543019 | SNP | T | C | C | Oncomine | NM_00113 7671 | p.M43V | p.M43 | 错义突变 | 热点 |
| 肺腺癌 | POTEC | 18 | 14513734 | SNP | C | T | T | Oncomine | NM_00113 7671 | p.G487E | p.G487 | 错义突变 | 热点 |
| 皮肤黑素瘤 | POTEC | 18 | 14543019 | SNP | T | C | C | Oncomine | NM_00113 7671 | p.M43V | p.M43 | 错义突变 | 热点 |
| 皮肤黑素瘤 | POTEC | 18 | 14542791 | SNP | C | T | T | Oncomine | NM_00113 7671 | p.A119T | p.A119 | 错义突变 | 热点 |
| 透明细胞肾细胞癌 | POTEC | 18 | 14542791 | SNP | C | T | T | Oncomine | NM_00113 7671 | p.A119T | p.A119 | 错义突变 | 热点 |
| 成胶质细胞瘤 | POTEM | 14 | 20010235 | SNP | A | G | G | Oncomine | NM_00114 5442 | p.V308A | p.V308 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | POTEM | 14 | 20010235 | SNP | A | G | G | Oncomine | NM_00114 5442 | p.V308A | p.V308 | 错义突变 | 热点 |
| 皮肤黑素瘤 | POTEM | 14 | 20019948 | SNP | C | T | T | Oncomine | NM_00114 5442 | p.M91I | p.M91 | 错义突变 | 热点 |
| 胃腺癌 | PPIL1 | 6 | 36842542 | SNP | C | T | T | Oncomine | NM_01605 9 | p.A3T | p.A3 | 错义突变 | 热点 |
| 卵巢浆液性腺癌 | PPIL1 | 6 | 36950519 | SNP | G | G | A | Oncomine | NM_01605 9 | p.A3V | p.A3 | 错义突变 | 热点 |
| 乳头状肾细胞癌 | PPIL1 | 6 | 36842542 | SNP | C | T | T | Oncomine | NM_01605 9 | p.A3T | p.A3 | 错义突变 | 热点 |
| 皮肤黑素瘤 | PRAMEF 20 | 1 | 13743091 | SNP | C | T | T | Oncomine | NM_00109 9852 | p.R94C | p.R94 | 错义突变 | 热点 |
| 成胶质细胞瘤 | PRAMEF | 1 | 137430 | SNP | G | A | A | Oncomine | NM_00109 | p.R94H | p.R94 | 错义突变 | 热点 |

[0195]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 20 | | 92 | | | | | | 9852 | | | | |
| 黑素瘤 | PRB3 | 12 | 11420548 | SNP | C | T | T | Oncomine | NM_006249 | p.G212E | p.G212 | 错义突变 | 热点 |
| 皮肤黑素瘤 | PRB3 | 12 | 11420963 | SNP | G | A | A | Oncomine | NM_006249 | p.R74C | p.R74 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | PRB3 | 12 | 11420963 | SNP | G | A | A | Oncomine | NM_006249 | p.R74C | p.R74 | 错义突变 | 热点 |
| 皮肤黑素瘤 | PRB3 | 12 | 11420548 | SNP | C | T | T | Oncomine | NM_006249 | p.G212E | p.G212 | 错义突变 | 热点 |
| 皮肤黑素瘤 | PRB4 | 12 | 11461597 | SNP | C | T | T | Oncomine | NM_002723 | p.G107E | p.G107 | 错义突变 | 热点 |
| 皮肤黑素瘤 | PRB4 | 12 | 11461475 | SNP | C | T | T | Oncomine | NM_002723 | p.G148R | p.G148 | 错义突变 | 热点 |
| 皮肤黑素瘤 | PRB4 | 12 | 11461474 | SNP | C | T | T | Oncomine | NM_002723 | p.G148E | p.G148 | 错义突变 | 热点 |
| 皮肤黑素瘤 | PROL1 | 4 | 71275418 | SNP | C | T | T | Oncomine | NM_021225 | p.P125S | p.P125 | 错义突变 | 热点 |
| 皮肤黑素瘤 | PROL1 | 4 | 71275418 | SNP | C | T | T | Oncomine | NM_021225 | p.P125S | p.P125 | 错义突变 | 热点 |
| 皮肤黑素瘤 | PROL1 | 4 | 71275428 | SNP | C | A | A | Oncomine | NM_021225 | p.P128H | p.P128 | 错义突变 | 热点 |
| 皮肤黑素瘤 | PROL1 | 4 | 71275427 | SNP | C | T | T | Oncomine | NM_021225 | p.P128S | p.P128 | 错义突变 | 热点 |
| 皮肤黑素瘤 | PRSS37 | 7 | 141536973 | SNP | C | T | T | Oncomine | NM_001008270 | p.G169E | p.G169 | 错义突变 | 热点 |
| 皮肤黑素瘤 | PRSS37 | 7 | 141540847 | SNP | C | T | T | Oncomine | NM_001008270 | p.M1I | p.M1 | 错义突变 | 热点 |
| 皮肤黑素瘤 | PRSS37 | 7 | 141536973 | SNP | C | T | T | Oncomine | NM_001008270 | p.G169E | p.G169 | 错义突变 | 热点 |
| 皮肤黑素瘤 | PRSS37 | 7 | 141540847 | SNP | C | T | T | Oncomine | NM_001008270 | p.M1I | p.M1 | 错义突变 | 热点 |
| 皮肤黑素瘤 | RAB39A | 11 | 107832799 | SNP | C | G | G | Oncomine | NM_017516 | p.R119G | p.R119 | 错义突变 | 热点 |

[0196]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 结肠直肠腺癌 | RAB39A | 11 | 107338009 | SNP | C | T | T | Oncomine | NM_017516 | p.R119W | p.R119 | 错义突变 | 热点 |
| 皮肤黑素瘤 | RAB39A | 11 | 107832799 | SNP | C | T | T | Oncomine | NM_017516 | p.R119W | p.R119 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | RALB | 2 | 121036297 | SNP | G | A | A | Oncomine | NM_002881 | p.M19I | p.M19 | 错义突变 | 热点 |
| 鳞状细胞肺癌 | RALB | 2 | 121036296 | SNP | T | C | C | Oncomine | NM_002881 | p.M19T | p.M19 | 错义突变 | 热点 |
| 皮肤黑素瘤 | RALB | 2 | 121036296 | SNP | T | A | A | Oncomine | NM_002881 | p.M19K | p.M19 | 错义突变 | 热点 |
| 成神经管细胞瘤 | RANGAP1 | 22 | 41652800 | SNP | A | C | C | Oncomine | NM_002883 | p.V268G | p.V268 | 错义突变 | 热点 |
| 宫颈鳞状细胞癌 | RANGAP1 | 22 | 41652800 | SNP | A | C | C | Oncomine | NM_002883 | p.V268G | p.V268 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | RANGAP1 | 22 | 41652800 | SNP | A | C | C | Oncomine | NM_002883 | p.V268G | p.V268 | 错义突变 | 热点 |
| 肺腺癌 | RANGAP1 | 22 | 41652800 | SNP | A | C | C | Oncomine | NM_002883 | p.V268G | p.V268 | 错义突变 | 热点 |
| 透明细胞肾细胞癌 | RANGAP1 | 22 | 41652800 | SNP | A | C | C | Oncomine | NM_002883 | p.V268G | p.V268 | 错义突变 | 热点 |
| 胃腺癌 | RAP1B | 12 | 69042539 | SNP | G | A | A | Oncomine | NM_015646 | p.G12E | p.G12 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | RAP1B | 12 | 69042539 | SNP | G | A | A | Oncomine | NM_015646 | p.G12E | p.G12 | 错义突变 | 热点 |
| 急性骨髓性白血病 | RAP1B | 12 | 67328806 | SNP | G | G | A | Oncomine | NM_015646 | p.G12E | p.G12 | 错义突变 | 热点 |
| 皮肤黑素瘤 | RBMY1D | Y | 23702641 | SNP | C | T | T | Oncomine | NM_001006120 | p.P124L | p.P124 | 错义突变 | 热点 |
| 鳞状细胞肺癌 | RBMY1D | Y | 23702641 | SNP | C | A | A | Oncomine | NM_001006120 | p.P124H | p.P124 | 错义突变 | 热点 |
| 皮肤黑素瘤 | RBMY1D | Y | 23702640 | SNP | C | T | T | Oncomine | NM_001006120 | p.P124S | p.P124 | 错义突变 | 热点 |

[0197]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 前列腺腺癌 | RQCD1 | 2 | 219447749 | SNP | C | G | G | Oncomine | NM_005444 | p.S87C | p.S87 | 错义突变 | 热点 |
| 黑素瘤 | RQCD1 | 2 | 219447749 | SNP | C | G | G | Oncomine | NM_005444 | p.S87C | p.S87 | 错义突变 | 热点 |
| 皮肤黑素瘤 | RQCD1 | 2 | 219449406 | SNP | C | T | T | Oncomine | NM_005444 | p.P131L | p.P131 | 错义突变 | 热点 |
| 皮肤黑素瘤 | RQCD1 | 2 | 219447748 | SNP | T | C | C | Oncomine | NM_005444 | p.S87P | p.S87 | 错义突变 | 热点 |
| 皮肤黑素瘤 | S100A7L2 | 1 | 153409566 | SNP | C | T | T | Oncomine | NM_001045479 | p.G103R | p.G103 | 错义突变 | 热点 |
| 皮肤黑素瘤 | S100A7L2 | 1 | 153409566 | SNP | C | T | T | Oncomine | NM_001045479 | p.G103R | p.G103 | 错义突变 | 热点 |
| 皮肤黑素瘤 | S100A7L2 | 1 | 153409565 | SNP | C | T | T | Oncomine | NM_001045479 | p.G103E | p.G103 | 错义突变 | 热点 |
| 非小细胞肺癌，NOS | S100A8 | 1 | 153362715 | SNP | T | C | C | Oncomine | NM_002964 | p.K49R | p.K49 | 错义突变 | 热点 |
| 成胶质细胞瘤 | S100A8 | 1 | 153362715 | SNP | T | C | C | Oncomine | NM_002964 | p.K49R | p.K49 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | S100A8 | 1 | 153362715 | SNP | T | C | C | Oncomine | NM_002964 | p.K49R | p.K49 | 错义突变 | 热点 |
| 甲状腺乳头状癌 | S100A8 | 1 | 153362715 | SNP | T | C | C | Oncomine | NM_002964 | p.K49R | p.K49 | 错义突变 | 热点 |
| 少突神经胶质瘤 | SAA2 | 11 | 18269491 | SNP | G | A | A | Oncomine | NM_030754 | p.S23L | p.S23 | 错义突变 | 热点 |
| 肺腺癌 | SDR16C5 | 8 | 57228627 | SNP | C | A | A | Oncomine | NM_138969 | p.A94S | p.A94 | 错义突变 | 热点 |
| 胃腺癌 | SDR16C5 | 8 | 57228626 | SNP | G | T | T | Oncomine | NM_138969 | p.A94D | p.A94 | 错义突变 | 热点 |
| 皮肤黑素瘤 | SDR16C5 | 8 | 57228829 | SNP | C | T | T | Oncomine | NM_138969 | p.M26I | p.M26 | 错义突变 | 热点 |
| 皮肤黑素瘤 | SDR16C5 | 8 | 57228854 | SNP | G | A | A | Oncomine | NM_138969 | p.S18L | p.S18 | 错义突变 | 热点 |
| 透明细胞肾细胞 | SDR16C5 | 8 | 572286 | SNP | C | G | G | Oncomine | NM_13896 | p.A94P | p.A94 | 错义突变 | 热点 |

[0198]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 癌 | | | 27 | | | | | | 9 | | | | |
| 皮肤黑素瘤 | SHH | 7 | 155596253 | SNP | G | A | A | Oncomine | NM_000193 | p.R244C | p.R244 | 错义突变 | 热点 |
| 肺腺癌 | SHH | 7 | 155596253 | SNP | G | A | A | Oncomine | NM_000193 | p.R244C | p.R244 | 错义突变 | 热点 |
| 皮肤黑素瘤 | SHH | 7 | 155596253 | SNP | G | A | A | Oncomine | NM_000193 | p.R244C | p.R244 | 错义突变 | 热点 |
| 前列腺腺癌 | SLC35G3 | 17 | 33520323 | SNP | C | T | T | Oncomine | NM_152462 | p.R335K | p.R335 | 错义突变 | 热点 |
| 浸润性膀胱尿道上皮癌 | SLC35G3 | 17 | 33520323 | SNP | C | T | T | Oncomine | NM_152462 | p.R335K | p.R335 | 错义突变 | 热点 |
| 成胶质细胞瘤 | SLC35G3 | 17 | 33520323 | SNP | C | T | T | Oncomine | NM_152462 | p.R335K | p.R335 | 错义突变 | 热点 |
| 胃腺癌 | SLC35G3 | 17 | 33520323 | SNP | C | T | T | Oncomine | NM_152462 | p.R335K | p.R335 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | SLC35G3 | 17 | 33520392 | SNP | G | C | C | Oncomine | NM_152462 | p.A312G | p.A312 | 错义突变 | 热点 |
| 肺腺癌 | SLC35G3 | 17 | 33520323 | SNP | C | T | T | Oncomine | NM_152462 | p.R335K | p.R335 | 错义突变 | 热点 |
| 皮肤黑素瘤 | SLC35G3 | 17 | 33520392 | SNP | G | C | C | Oncomine | NM_152462 | p.A312G | p.A312 | 错义突变 | 热点 |
| 皮肤黑素瘤 | SLC35G3 | 17 | 33520323 | SNP | C | T | T | Oncomine | NM_152462 | p.R335K | p.R335 | 错义突变 | 热点 |
| 皮肤黑素瘤 | SPATA8 | 15 | 97326937 | SNP | G | A | A | Oncomine | NM_173499 | p.E18K | p.E18 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | SPATA8 | 15 | 97326937 | SNP | G | A | A | Oncomine | NM_173499 | p.E18K | p.E18 | 错义突变 | 热点 |
| 皮肤黑素瘤 | SPATA8 | 15 | 97326937 | SNP | G | A | A | Oncomine | NM_173499 | p.E18K | p.E18 | 错义突变 | 热点 |
| 宫颈鳞状细胞癌 | SPINK13 | 5 | 147665577 | SNP | G | A | A | Oncomine | NM_001040129 | p.R84H | p.R84 | 错义突变 | 热点 |
| 皮肤黑素瘤 | SPINK13 | 5 | 147665 | SNP | C | T | T | Oncomine | NM_00104 | p.R84C | p.R84 | 错义突变 | 热点 |

[0199]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 576 | | | | | | 0129 | | | | |
| 皮肤黑素瘤 | ST6GAL2 | 2 | 107459497 | SNP | C | T | T | Oncomine | NM_032528 | p.E313K | p.E313 | 错义突变 | 热点 |
| 结肠直肠腺癌 | ST6GAL2 | 2 | 106816941 | SNP | G | A | A | Oncomine | NM_032528 | p.S346L | p.S346 | 错义突变 | 热点 |
| 子宫内膜子宫内膜样腺癌 | ST6GAL2 | 2 | 107460402 | SNP | C | C | T | Oncomine | NM_032528 | p.R11Q | p.R11 | 错义突变 | 热点 |
| 肺腺癌 | ST6GAL2 | 2 | 107459730 | SNP | C | A | A | Oncomine | NM_032528 | p.G235V | p.G235 | 错义突变 | 热点 |
| 肺腺癌 | ST6GAL2 | 2 | 107460276 | SNP | G | A | A | Oncomine | NM_032528 | p.P53L | p.P53 | 错义突变 | 热点 |
| 鳞状细胞肺癌 | ST6GAL2 | 2 | 107460166 | SNP | G | A | A | Oncomine | NM_032528 | p.H90Y | p.H90 | 错义突变 | 热点 |
| 鳞状细胞肺癌 | ST6GAL2 | 2 | 107459731 | SNP | C | A | A | Oncomine | NM_032528 | p.G235W | p.G235 | 错义突变 | 热点 |
| 鳞状细胞肺癌 | ST6GAL2 | 2 | 107423361 | SNP | C | T | T | Oncomine | NM_032528 | p.E455K | p.E455 | 错义突变 | 热点 |
| 皮肤黑素瘤 | ST6GAL2 | 2 | 107459497 | SNP | C | T | T | Oncomine | NM_032528 | p.E313K | p.E313 | 错义突变 | 热点 |
| 皮肤黑素瘤 | ST6GAL2 | 2 | 107460402 | SNP | C | T | T | Oncomine | NM_032528 | p.R11Q | p.R11 | 错义突变 | 热点 |
| 皮肤黑素瘤 | ST6GAL2 | 2 | 107450509 | SNP | G | A | A | Oncomine | NM_032528 | p.S346L | p.S346 | 错义突变 | 热点 |
| 皮肤黑素瘤 | ST6GAL2 | 2 | 107423361 | SNP | C | T | T | Oncomine | NM_032528 | p.E455K | p.E455 | 错义突变 | 热点 |
| 皮肤黑素瘤 | ST6GAL2 | 2 | 107459496 | SNP | T | A | A | Oncomine | NM_032528 | p.E313V | p.E313 | 错义突变 | 热点 |
| 皮肤黑素瘤 | ST6GAL2 | 2 | 107459731 | SNP | C | T | T | Oncomine | NM_032528 | p.G235R | p.G235 | 错义突变 | 热点 |
| 皮肤黑素瘤 | ST6GAL2 | 2 | 107460166 | SNP | G | A | A | Oncomine | NM_032528 | p.H90Y | p.H90 | 错义突变 | 热点 |
| 皮肤黑素瘤 | SYPL1 | 7 | 105739 | SNP | G | A | A | Oncomine | NM_00675 | p.P81S | p.P81 | 错义突变 | 热点 |

[0200]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 611 | | | | | | 4 | | | | |
| 皮肤黑素瘤 | SYPL1 | 7 | 105739611 | SNP | G | A | A | Oncomine | NM_006754 | p.P81S | p.P81 | 错义突变 | 热点 |
| 黑素瘤 | SYT1 | 12 | 79689912 | SNP | C | T | T | Oncomine | NM_005639 | p.P180S | p.P180 | 错义突变 | 热点 |
| 黑素瘤 | SYT1 | 12 | 79679683 | SNP | G | A | A | Oncomine | NM_005639 | p.E95K | p.E95 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | SYT1 | 12 | 79611355 | SNP | C | T | T | Oncomine | NM_005639 | p.A19V | p.A19 | 错义突变 | 热点 |
| 急性骨髓性白血病 | SYT1 | 12 | 78135485 | SNP | G | G | A | Oncomine | NM_005639 | p.A19T | p.A19 | 错义突变 | 热点 |
| 皮肤黑素瘤 | SYT1 | 12 | 79689912 | SNP | C | T | T | Oncomine | NM_005639 | p.P180S | p.P180 | 错义突变 | 热点 |
| 皮肤黑素瘤 | SYT1 | 12 | 79679683 | SNP | G | A | A | Oncomine | NM_005639 | p.E95K | p.E95 | 错义突变 | 热点 |
| 前列腺腺癌 | SYT1 | 12 | 79611355 | SNP | C | T | T | Oncomine | NM_005639 | p.A19V | p.A19 | 错义突变 | 热点 |
| 肺腺癌 | TCEAL8 | X | 102508844 | SNP | G | T | T | Oncomine | NM_153333 | p.R22S | p.R22 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | TCEAL8 | X | 102508843 | SNP | C | T | T | Oncomine | NM_153333 | p.R22H | p.R22 | 错义突变 | 热点 |
| 透明细胞肾细胞癌 | TCEAL8 | X | 102508844 | SNP | G | A | A | Oncomine | NM_153333 | p.R22C | p.R22 | 错义突变 | 热点 |
| 前列腺腺癌 | TMEM147 | 19 | 36037641 | SNP | C | T | T | Oncomine | NM_032635 | p.A92V | p.A92 | 错义突变 | 热点 |
| 成胶质细胞瘤 | TMEM147 | 19 | 36037641 | SNP | C | T | T | Oncomine | NM_032635 | p.A92V | p.A92 | 错义突变 | 热点 |
| 皮肤黑素瘤 | WFDC5 | 20 | 43739300 | SNP | G | A | A | Oncomine | NM_145652 | p.R68C | p.R68 | 错义突变 | 热点 |
| 导管乳癌 | WFDC5 | 20 | 43739300 | SNP | G | G | A | Oncomine | NM_145652 | p.R68C | p.R68 | 错义突变 | 热点 |
| 嫌色性肾细胞癌 | WFDC5 | 20 | 437392 | SNP | C | C | T | Oncomine | NM_145652 | p.R68H | p.R68 | 错义突变 | 热点 |

[0201]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 99 | | | | | | 2 | | | | |
| 透明细胞肾细胞癌 | ZFAND2B | 2 | 220072989 | SNP | T | C | C | Oncomine | NM_138802 | p.I149T | p.I149 | 错义突变 | 热点 |
| 乳头状肾细胞癌 | ZFAND2B | 2 | 220072989 | SNP | T | G | G | Oncomine | NM_138802 | p.I149S | p.I149 | 错义突变 | 热点 |
| 非小细胞肺癌，NOS | ZNF780A | 19 | 40581109 | SNP | T | C | C | Oncomine | NM_001010880 | p.I414V | p.I414 | 错义突变 | 热点 |
| 肺腺癌 | ZNF780A | 19 | 40581529 | SNP | C | T | T | Oncomine | NM_001010880 | p.V274I | p.V274 | 错义突变 | 热点 |
| 肺腺癌 | ZNF780A | 19 | 40581535 | SNP | A | C | C | Oncomine | NM_001010880 | p.S272A | p.S272 | 错义突变 | 热点 |
| 少枝星形细胞瘤 | ZNF780A | 19 | 40580552 | SNP | T | G | G | Oncomine | NM_001010880 | p.Q599H | p.Q599 | 错义突变 | 热点 |
| 宫颈鳞状细胞癌 | ZNF780A | 19 | 40580552 | SNP | T | G | G | Oncomine | NM_001010880 | p.Q599H | p.Q599 | 错义突变 | 热点 |
| 胃腺癌 | ZNF780A | 19 | 40581529 | SNP | C | T | T | Oncomine | NM_001010880 | p.V274I | p.V274 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | ZNF780A | 19 | 40581109 | SNP | T | C | C | Oncomine | NM_001010880 | p.I414V | p.I414 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | ZNF780A | 19 | 40580552 | SNP | T | G | G | Oncomine | NM_001010880 | p.Q599H | p.Q599 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | ZNF780A | 19 | 40581529 | SNP | C | T | T | Oncomine | NM_001010880 | p.V274I | p.V274 | 错义突变 | 热点 |
| 肺腺癌 | ZNF780A | 19 | 40581109 | SNP | T | C | C | Oncomine | NM_001010880 | p.I414V | p.I414 | 错义突变 | 热点 |
| 鳞状细胞肺癌 | ZNF780A | 19 | 40581535 | SNP | A | C | C | Oncomine | NM_001010880 | p.S272A | p.S272 | 错义突变 | 热点 |
| 皮肤黑素瘤 | ZNF780A | 19 | 40581535 | SNP | A | C | C | Oncomine | NM_001010880 | p.S272A | p.S272 | 错义突变 | 热点 |
| 皮肤黑素瘤 | ZNF780A | 19 | 40581109 | SNP | T | C | C | Oncomine | NM_001010880 | p.I414V | p.I414 | 错义突变 | 热点 |
| 甲状腺滤泡癌 | ZNF780A | 19 | 40581535 | SNP | A | C | C | Oncomine | NM_001010 | p.S272A | p.S272 | 错义突变 | 热点 |

[0202]

表 11：功能突变获得

| 癌症类型 | 基因符号 | 染色体 | 起始位置 | 变异体类型 | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | CBI 注释来源 | 转录物 | 变异体变化 | 变异体位置 | 变异体分类 | 泛疾病 CBI 变异体类别 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 35 | | | | | | 0880 | | | | |
| 甲状腺乳头状癌 | ZNF780A | 19 | 40580552 | SNP | T | G | G | Oncomine | NM_001010880 | p.Q599H | p.Q599 | 错义突变 | 热点 |
| 肺腺癌 | ZNF844 | 19 | 12187394 | SNP | T | C | C | Oncomine | NM_001136501 | p.F487L | p.F487 | 错义突变 | 热点 |
| 成胶质细胞瘤 | ZNF844 | 19 | 12187394 | SNP | T | C | C | Oncomine | NM_001136501 | p.F487L | p.F487 | 错义突变 | 热点 |
| 成胶质细胞瘤 | ZNF844 | 19 | 12187275 | SNP | G | C | C | Oncomine | NM_001136501 | p.R447P | p.R447 | 错义突变 | 热点 |
| 宫颈鳞状细胞癌 | ZNF844 | 19 | 12187394 | SNP | T | C | C | Oncomine | NM_001136501 | p.F487L | p.F487 | 错义突变 | 热点 |
| 头颈鳞状细胞癌 | ZNF844 | 19 | 12187275 | SNP | G | C | C | Oncomine | NM_001136501 | p.R447P | p.R447 | 错义突变 | 热点 |
| 肺腺癌 | ZNF844 | 19 | 12187275 | SNP | G | C | C | Oncomine | NM_001136501 | p.R447P | p.R447 | 错义突变 | 热点 |
| 皮肤黑素瘤 | ZNF844 | 19 | 12187275 | SNP | G | C | C | Oncomine | NM_001136501 | p.R447P | p.R447 | 错义突变 | 热点 |
| 皮肤黑素瘤 | ZNF844 | 19 | 12187394 | SNP | T | C | C | Oncomine | NM_001136501 | p.F487L | p.F487 | 错义突变 | 热点 |
| 少突神经胶质瘤 | ZNF845 | 19 | 53855196 | SNP | T | C | C | Oncomine | NM_138374 | p.M423T | p.M423 | 错义突变 | 热点 |
| 甲状腺乳头状癌 | ZNF845 | 19 | 53855196 | SNP | T | C | C | Oncomine | NM_138374 | p.M423T | p.M423 | 错义突变 | 热点 |
| 甲状腺乳头状癌 | ZNF845 | 19 | 53855197 | SNP | G | A | A | Oncomine | NM_138374 | p.M423I | p.M423 | 错义突变 | 热点 |

[0203] 本发明提供新颖基因变异体和基因变异体-疾病病况相关性。基因变异体可以具有一个或多个产生变异蛋白的突变。本文提供的基因变异体与某些癌症相关。基因变异体

产生蛋白变异体。本发明进一步提供探针(如扩增引物组和检测探针)以及检测、诊断和治疗的方法以及包括或检测本文所公开的基因变异体的试剂盒。

[0204] 变异体在表7和11中用所提供的亲本或野生型基因的登录号或Entrez核苷酸和/或蛋白质序列显示为氨基酸变异体。与各种癌症的相关性展示在表7和11中。表7和11提供使用实例2中概述的方法鉴别的超过99种基因的列表。在对应的正常组织中未发现所述变异或突变。这是重要的,因为在典型患者中,肿瘤样品可以具有10's-100's的肿瘤特定变异。然而,在多个患者中在同一位置处(且不在正常组织中)出现的变异更显著。分析4445个样品(来自4445个患者),且制备热点列表。许多再现的突变在15-20种不同癌症类型中在同一位置处发现。

[0205] 诊断学和试剂盒

[0206] 本文中涵盖诊断、治疗和检测基因变异体和相关疾病的方法。所述方法可以包括检测受试者样品中的基因融合体和/或基因变异体。可在本文所公开的反应混合物、组合物和试剂盒中的任一者中检测任何数量和组合的基因融合体和/或基因变异体。

[0207] 在一个实施例中,本发明提供一种组合物和一种试剂盒,其包含特异性地识别编码选自表7和/或表11的基因变异体的核苷酸序列的一个探针组。所述探针组可例如为一个扩增引物组。在另一实施例中,本文提供一种组合物,其包括侧接编码表7和/或表11中的一种或多种变异体的基因变异体的一个引物组。此实施例的反应混合物可以进一步包括检测器探针,其结合于包括选自表7和/或表11的基因变异体的核苷酸序列。包括检测器探针或不包括检测器探针的反应混合物可以进一步包括聚合酶、dNTP和/或尿嘧啶DNA去糖基化酶(UDG)。聚合酶和UDG通常不来自人类来源。反应混合物可以进一步包括目标核酸,例如人类目标核酸。人类目标核酸可例如从来自怀疑患有癌症的个体的生物样品中分离。所述癌症可选自:BLCA=膀胱癌、BRCA=乳癌、CESC=子宫颈细胞癌、COAD=结肠腺癌、GBM=多形性成胶质细胞瘤、HNSC=头颈鳞状细胞癌、KIRK=透明细胞肾细胞癌、KIRP=肾脏肾乳头状细胞癌、LAML=急性骨髓性白血病、LGG=脑低级别胶质瘤、LIHC=肝脏肝细胞癌、LUAD=肺腺癌、LUSC=鳞状细胞肺癌、OV=卵巢浆液性腺癌、PRAD=前列腺腺癌、READ=直肠腺癌、SKCM=皮肤黑素瘤、STAD=胃腺癌、THCA=甲状腺癌和UCEC=子宫体子宫内膜样癌。

[0208] 在一些实施例中,提供一种试剂盒,其中所述试剂盒涵盖一个或多个探针。在一些实施例中,所述试剂盒涵盖1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、50、75、100、150、200、250、500种或更多种融合体基因的探针。在一些实施例中,探针经可检测标记。在一些实施例中,所述探针与基因融合体中所存在的断裂点杂交。

[0209] 在一些实施例中,表7和11中所公开的基因变异体中的任一者的检测可与那些表中所公开的基因变异体中的另一者或本文所公开的基因融合体中的任一个的检测组合。也就是说,可在同一反应中检测2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、50、75、100、150、200、250、500种或更多种基因变异体。在一些实施例中,所检测的基因变异体为表4-6、7和11、20和23中所公开的那些,且可与那些表中所公开的基因融合体中的另一者的检测组合。也就是说,可在同一反应中检测2、3、使得4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、50、75、100、150、200、250、500种或更多种基因融合体。

[0210] 编码表7和/或表11中的一种或多种基因变异体的核苷酸序列可为涵盖所述变异的任何大小。举例来说,核苷酸序列可为可容易地使用引物拷贝和/或使用探针检测的任何

大小。

[0211] 在另一实施例中，提供一种探针组，其特异性地识别编码选自表7和/或表11(基因变异体)的基因变异体的核酸。在另一实施例中，本文提供一种引物组，其特异性地扩增编码选自表7和/或表11的基因变异体的目标核酸。在另一实施例中，本文提供qPCR分析，如TaqMan™分析或MolecularBeacons™分析，其特异性地扩增且检测编码选自表7和/或表11的基因变异体的目标核酸。

[0212] 本发明还提供一种经分离核酸，其包含至少一种包括在选自表7和/或表11的一种或多种基因变异体中发现的变异的序列。经分离核酸可以在5'末端上包括第一引物。此外，核酸可能为单链或双链的。

[0213] 在其他实施例中，本发明提供一种试剂盒，其包括一个检测器探针和/或一个探针组，例如一个扩增引物组，其特异性地识别编码选自表7和/或表11的基因变异体的核酸。举例来说，在某些实施例中，检测器探针或扩增引物组经设计以扩增和/或检测包括编码表7和/或表11中的基因变异体的核酸中的至少一者的核酸。所述试剂盒可以在另一容器中或在同一容器中进一步包括来自扩增反应混合物的组分，如聚合酶(通常不来自人类来源)、dNTP和/或UDG。此外，所述试剂盒可以包括对照核酸。举例来说，对照核酸可以包括包含选自表7和/或表11的基因变异体的序列。

[0214] 提供一种检测癌症的方法，其包含扩增编码选自表7和/或表11的基因变异体的核酸，例如所述核酸可以包括来自表7和/或表11中的登录号中的一者的序列，除了所述序列含有编码表7和/或表11中的基因变异体的变异体以外；以及检测所述核酸的存在，其中所述核酸的存在指示癌症存在于样品中。在另一方法中，本文提供一种检测癌症的方法，其包括产生包括选自编码表7和/或表11中的基因变异体的序列的序列的扩增子，以及检测所述核酸的存在，其中所述核酸的存在指示膀胱癌、头颈癌或肺鳞状细胞癌存在于样品中。所述扩增子通常包括经延伸以形成所述扩增子的引物。所述癌症是选自膀胱癌、乳癌、子宫颈细胞癌、结肠腺癌、多形性成胶质细胞瘤、头颈鳞状细胞癌、透明细胞肾细胞癌、肾脏肾乳头状细胞癌、急性骨髓性白血病、脑低级别胶质瘤、肝脏肝细胞癌、肺腺癌、鳞状细胞肺癌、卵巢浆液性腺癌、前列腺腺癌、直肠腺癌、皮肤黑素瘤、胃腺癌、甲状腺癌和子宫体子宫内膜样癌。

[0215] 提供一种试剂盒，其包含特异性地识别包含来自表7和/或表11的基因变异体的核酸的一个探针组(例如一个扩增引物组)。所述试剂盒可以在另一容器中或在同一容器中进一步包括来自扩增反应混合物的组分，如聚合酶(通常不来自人类来源)、dNTP和/或UDG。此外，所述试剂盒可以包括对照核酸。举例来说，对照核酸可以包括包含来自表7和/或表11的基因变异体的序列。在某些实施例中，提供一种探针组，其特异性地识别包含来自表7和/或表11的基因变异体的核酸。

[0216] 在另一实施例中，提供一种基因变异体，其包含表7和/或表11中的基因变异体中的至少一者。

[0217] 在另一实施例中是一种方法，其通过检测选自表7和/或表11的基因变异体的存在来检测样品中的选自膀胱癌、乳癌、子宫颈细胞癌、结肠腺癌、多形性成胶质细胞瘤、头颈鳞状细胞癌、透明细胞肾细胞癌、肾脏肾乳头状细胞癌、急性骨髓性白血病、脑低级别胶质瘤、肝脏肝细胞癌、肺腺癌、鳞状细胞肺癌、卵巢浆液性腺癌、前列腺腺癌、直肠腺癌、皮肤黑素

瘤、胃腺癌、甲状腺癌和子宫体子宫内膜样癌的癌症。基因变异体例如可以包括(但不限于)ZNF479变异体R11Q、R295K、R295T、R295I、R345I、R345T、K438T和T466K(参见表8)。

[0218] 表18

[0219]

| 癌症类型 | 基因符号 | 可药化性 | KM 证据 |
|---|---|---|---|
| 星形细胞瘤 | CXCR2 | Y | |
| 子宫内膜子宫内膜样腺癌 | CXCR2 | Y | |
| 鳞状细胞肺癌 | CXCR2 | Y | |
| 皮肤黑素瘤 | CXCR2 | Y | |
| 皮肤黑素瘤 | CXCR2 | Y | |
| 结肠直肠腺癌 | IL3 | Y | |
| 胃腺癌 | IL3 | Y | |
| 皮肤黑素瘤 | KCNK9 | Y | 有利的结果 |
| 子宫内膜子宫内膜样腺癌 | KCNK9 | Y | |
| 肺腺癌 | KCNK9 | Y | |
| 鳞状细胞肺癌 | KCNK9 | Y | 较差的结果 |
| 非小细胞肺癌，NOS | S100A8 | Y | |
| 成胶质细胞瘤 | S100A8 | Y | |
| 头颈鳞状细胞癌 | S100A8 | Y | |
| 甲状腺乳头状癌 | S100A8 | Y | |
| 皮肤黑素瘤 | SHH | Y | |
| 肺腺癌 | SHH | Y | |
| 皮肤黑素瘤 | CCDC61 | | 较差的结果 |
| 皮肤黑素瘤 | CCDC61 | | 较差的结果 |
| 皮肤黑素瘤 | CNTN5 | | 较差的结果 |
| 皮肤黑素瘤 | CNTN5 | | 较差的结果 |
| 皮肤黑素瘤 | CNTN5 | | 较差的结果 |
| 皮肤黑素瘤 | CNTN5 | | 较差的结果 |
| 结肠直肠腺癌 | CNTN5 | | 较差的结果 |
| 结肠直肠腺癌 | CNTN5 | | 较差的结果 |
| 皮肤黑素瘤 | CNTN5 | | 较差的结果 |
| 皮肤黑素瘤 | CNTN5 | | 较差的结果 |
| 皮肤黑素瘤 | CNTN5 | | 较差的结果 |
| 皮肤黑素瘤 | EDDM3A | | 较差的结果 |
| 皮肤黑素瘤 | FABP1 | | 较差的结果 |
| 肺腺癌 | OR2L13 | | 较差的结果 |
| 皮肤黑素瘤 | OR4E2 | | 较差的结果 |

[0220]

| 癌症类型 | 基因符号 | 可药化性 | KM 证据 |
|---|---|---|---|
| 皮肤黑素瘤 | OR4E2 | | 较差的结果 |
| 皮肤黑素瘤 | OR4E2 | | 较差的结果 |
| 皮肤黑素瘤 | PRSS37 | | 较差的结果 |
| 皮肤黑素瘤 | PRSS37 | | 较差的结果 |
| 皮肤黑素瘤 | SPINK13 | | 较差的结果 |
| 子宫内膜子宫内膜样腺癌 | ST6GAL2 | | 较差的结果 |

[0221] 表18提供从表11过滤的可药化性或预后相关性。表18提供癌症类型、在表11中鉴别为可药化的基因的基因符号、可药化性(Y＝有)和KM证据。KM证据栏提供卡普兰-迈耶证据(Kaplan-Meier evidence)。KM证据指示所述事件类型是否支持特定癌症类型中的良好或较差的预后。

[0222] 靶向治疗

[0223] 在至少一个实施例中，基因融合体和/或基因变异体可以用于鉴别靶向疗法。靶向疗法可以包括鉴别与基因融合体和/或基因变异体特异性地相互作用的试剂。靶向疗法可

以包括(但不限于)抗体疗法、反义疗法和小分子疗法。反义疗法更详细地在标题“反义”下论述。

[0224] 用于使核酸分子失活的组合物和方法在某种程度上涉及使用具有具备与作为期望失活的对象的核酸分子(即,目标核酸分子)互补的序列的核酸区域的分子。本发明的方法可用于使与特定癌症相关的基因融合体和/或基因变异体失活。因此,可鉴别与本文中鉴别的基因融合体或基因变异体中的任一个互补的反义分子。

[0225] 小分子为可用作酶底物或生物过程的调节剂的低分子量(<800道尔顿(Dalton))有机化合物,且大小约为$10^{-9}$m。在药理学中,所述术语通常用于结合于蛋白质或核酸且充当效应子(改变蛋白质或核酸的活性或功能)的分子。可通过在细胞分析中表达基因融合体或变异体且鉴别抑制基因融合体或变异体的表达或活性的小分子来测试小分子的效应子功能。

[0226] 可药化性是用于药物发现的术语,用以描述以高亲和力已知会结合或预测会结合于药物的生物目标,如蛋白质。此外,药物结合于可药化目标会以对患者的治疗效益改变目标的功能。本文中的术语“药物”包括小分子(低分子量有机物质),而且已经扩展到包括生物医学产品,如治疗性单克隆抗体。在至少一个实施例中,基因融合体或基因变异体可以用于鉴别可药化目标。表8提供已从表1-3和7中鉴别的可药化目标的列表。举例来说,TPM1/ALK基因融合体是可药化目标,因为如表8中所示,涉及ALK的疾病可用克卓替尼(crizotinib)治疗。因此,如果基因融合体包括ALK,那么癌症可用克卓替尼治疗。另外,如果基因变异体包括ALK中的突变,那么癌症可用克卓替尼治疗。

[0227] 类似地,表21提供已从表19中鉴别的可药化目标的列表,且表24提供已从表22中鉴别的可药化目标的列表。

[0228]

表 8：来自表 1 的可药化基因

| 可药化基因 | 已批准的 | 注册前（批准前） | III 期 | II 期 | I 期 | 临床前 |
|---|---|---|---|---|---|---|
| ALK | 克卓替尼 | N | N | AP-26113；RG-7853；LDK-378；TSR-011 | X-396；ASP-3026 | NMS-E628；极光激酶 + ALK 抑制剂（萨勒姆公司（Sareum）、阿斯利康公司（AstraZeneca））；ALK 抑制剂（阿斯利康公司（AstraZeneca）、赛法隆公司（Cephalon）、奥里基因公司（Aurigene））；ARN-5032；DLX-521 |
| CASR | 盐酸西那卡塞（cincacalcet hydrochloride） | N | N | N | N | N |
| EGFR | 埃罗替尼（erlotinib）；帕尼单抗（panitumumab）；西妥昔单抗（cetuximab）；奈皮德明（nepidermin）；吉非替尼（gefitinib）；尼妥珠单抗（nimotuzumab）；凡德他尼（vandetanib）；二甲苯磺酸拉帕替尼（lapatinib ditosylate）；盐酸埃克替尼（icotinib hydrochloride）； | 阿法替尼（afatinib） | 扎鲁突木单抗（zalutumumab）；来那替尼（neratinib）；多韦替尼乳酸盐（dovitinib lactate）；XL-647；林多派皮姆（rindopepimut）；莱西突木单抗（necitumumab）；达可替尼（dacomitinib） | BMS-690514；瓦尼替尼（varlitinib）；AC-480；AZD-8931；Sym-004；伊姆加土珠单抗（imgatuzumab）；AVL-301；AVL-301；泊西替尼（poziotinib）；MEHD-7945A；PR-610； | 马瑞佐米（marizomib）；CUDC-101；MM-151；AL-6802；S-222611；ABT-806；安卓奎诺尔（antroquinonol）；GT-MAB 5.2-GEX；依吡替尼（epitinib）；西栗替尼（theliatinib）；西帕替尼（cipatinib）；AMG-595 | STP-503；SN-29966；MT-062；STP-801 |
| FGFR3 | 帕纳替尼（ponatinib） | 马赛替尼（masitinib） | 多韦替尼乳酸盐 | ENMD-2076；AZD-4547 | JNJ-42756493；BGJ-398；LY-2874455；S-49076 | N |

[0229]

| 可药化基因 | 已批准的 | 注册前（批准前） | III 期 | II 期 | I 期 | 临床前 |
|---|---|---|---|---|---|---|
| GNAS | N | N | N | N | N | N |
| JAK2 | 芦可替尼（ruxolitinib）（用于特发性骨髓纤维化） | N | SAR-302503；帕瑞替尼（pacritinib） | AT-9283；莫罗替尼（momelotinib）；甘多替尼（gandotinib）；BMS-911543；NS-018 | AC-430；SB-1317 | ON-044580；INCB-16562；NVP-BSK805；TP-0413；MRLB-11055；CPL-407-22 |
| NOTCH1 | N | N | N | N | OMP-52M51 | Debio-0826；TR-4；Notch 抗体（AVEO）；Notch1 抑制剂（因特普罗廷公司（Interprotein）） |
| NTRK1 | N | N | N | 马来酸米西西尼（milciclib maleate） | N | 酪氨酸激酶抑制剂（布里斯托尔-迈尔斯·斯奎布公司（Bristol-Myers Squibb））；PLX-7486 |
| PIK3CA | N | N | 哌立福新（perifosine）；布帕昔布（buparlisib）； | ZSTK-474；PX-866；皮克立西（pictilisib）；XL-765；XL-147；BEZ-235；PKI-587；PF-04691502；PF-04691502；BAY-80-6946；BYL-719； | INK-1117；GSK-2126458；CUDC-907；GDC-0032；PWT-33597；DS-7423；GDC-0084；BAY-1082439；PI3 激酶/mTOR 抑制剂（礼来公司（Lilly）） | LOR-220；AEZS-129；SB-2343；WX-037；PI3/Mnk 激酶抑制剂（普洛金尼克斯公司（Progenics））；AEZS-132；CLR-1401；PI3/mTOR 激酶抑制剂（安进公司（Amgen））；AEZS-136；HM-032；AMG-511；抗癌剂疗法（斯法若药厂（Sphaera Pharma））；HMPL-518；GNE-317；mTOR 抑制剂/PI3 激酶抑制剂（礼来公司）；CUDC908；PF-06465603；AEZS-134； |

[0230]

| 可药化基因 | 已批准的 | 注册前（批准前） | III 期 | II 期 | I 期 | 临床前 |
|---|---|---|---|---|---|---|
| RET | 索拉非尼（sorafenib）；凡德他尼；苹果酸舒尼替尼（sunitinib malate）；卡博替尼（cabozantinib）；瑞格非尼（regorafenib） | N | 二磷酸莫替沙尼（motesanib diphosphate）；SAR-302503；阿帕替尼（apatinib） | | N | MG-516；RET 激酶抑制剂；NMS-173 |
| ROS1 | 克卓替尼 | N | N | N | N | N |
| ALK | 克卓替尼 | N | N | AP-26113；RG-7853；LDK-378；TSR-011；NMS-E628 | X-396；ASP-3026 | NMS-E628；极光激酶 + ALK 抑制剂（萨勒姆公司、阿斯利康公司）；ALK 抑制剂（阿斯利康公司、赛法隆公司、奥里基因公司）；ARN-5032；DLX-521 |
| NTRK1 | N | N | N | 马来酸米西西尼 | N | 酪氨酸激酶抑制剂（布里斯托尔-迈尔斯·斯奎布公司）；PLX-7486 |
| VIM | N | N | N | 普托木单抗（pritumumab） | N | N |
| PTK2 | | | | PF-04554878 | GSK-2256098；BI-853520；VS-4718 | CFAK-C4；FAK 抑制剂（瓦拉斯代姆公司（Varastem）、武田公司（Takeda））；CTX-0294945；CTX-0294945 |
| BRS3 | N | N | N | N | N | N |
| TP53 | 今又生（Gendicine） | N | N | 奎纳克林（quinacrine）；APR-246；ISA-102 | RG-7388；SGT-53；CBLC-137；SAR-405838 | PXN-527；ORCA-010；TR-2；ALT-802；OBP-702 |
| STAT3 | N | N | N | 溴夫定（brivudine）；OPB-31121；柠檬酸安 | OPB-51602 | CLT-005；GLG-101；GLG-202；GLG-302； |

[0231]

| 可药化基因 | 已批准的 | 注册前（批准前） | III 期 | II 期 | I 期 | 临床前 |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | 西他滨（anatabine citrate）；ISIS-STAT3Rx | | GLG-401；PNT-500 |
| NOTCH2 | N | N | N | OMP-59R5 | N | N |
| MET | 卡博替尼；克卓替尼 | N | 提瓦替尼（tivantinib）；里乐木单抗（rilotumumab）；奥那组单抗（onartuzumab）； | MGCD-265；弗雷替尼（foretinib）；费拉妥珠单抗（ficlatuzumab）；BMS-777607；格瓦替尼（golvatinib）；INCB-028060；LY-2875358 | AMG-208；TAS-115；沃利替尼（volitinib）；SAR-125844；S-49076 | X-379；麦他替尼（metatinib）；PRS-110；ASP-08001；ARGX-111；DCC-2701；DCC-2721；MG-516；AL-2846；CG-206481；T-1840383；cMet-EGFR 双重抑制剂（水晶基因组公司（CrystalGenomics））；双特异性抗体（霍夫曼-拉罗奇公司（Hoffmann-La Roche）） |
| CDH1 | N | N | N | N | N | N |
| TOP1 | 盐酸贝洛替康（belotecan hydrochloride）；盐酸伊立替康（irinotecan hydrochloride）；拓扑替康（topotecan） | N | 考西替康（cositecan）；伊立替康，HyACT；伊立替康，智擎生技制药公司（PharmaEngine）；培舍埃特伊立替康（etirinotecan pegol） | 吉马替康（gimatecan）；喜树碱（camptothecin），卡兰多公司（Calando）；伊立替康盐酸盐 + 氟尿苷（floxuridine），赛拉特公司（Celator）；培舍弗特康（firtecan pegol）；TLC-388 盐酸盐；hRS7-SN-38；伊立替康珠粒，生物相容公司（Biocompatibles） | 伊立替康脂质体，养乐多公司（Yakult）；HM-30181A；纳迷特康（namitecan）；喜树碱前药，默萨那公司（Mersana）；拉贝珠单抗（labetuzumab）-SN-38；Genz-644282；盐酸辛米特康前药（simmitecan hydrochloride | 喜树碱（艾菲沃斯公司（Aphios））；伊立替康（生物联盟公司（BioAlliance））；顺铂（cisplatin）+ 伊立替康（赛拉特公司）；APH-0804；伊立替康（冠军公司（Champions））；SER-203；SN-38；拓扑替康 + （长春新碱（vincristine））（利泊丘尔公司（LipoCure））；拓扑替 |

[0232]

| 可药化基因 | 已批准的 | 注册前（批准前） | III 期 | II 期 | I 期 | 临床前 |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | prodrug） | 康（恩杜克斯医药公司（EnduRx Pharmaceuticals）） |
| RARA | 他米巴罗汀（tamibarotene） | N | N | IRX-5183 | N | N |
| ERBB2 | 曲妥珠单抗（trastuzumab）；曲妥珠单抗恩他新（trastuzumab emtansine）；帕妥珠单抗（pertuzumab）；二甲苯磺酸拉帕替尼；卡托莫西单抗（catumaxomab）；阿法替尼 | 曲妥珠单抗，因汉兹公司（Enhanze） | 来那替尼；XL-647；达可替尼；奈立派姆-S（nelipepimut-S）；曲妥珠单抗（赛尔群公司（Celltrion）、拜沃卡德公司（Biocad）、必奥康公司（Biocon）、辛赛恩公司（Synthon）、哈维斯姆公司（Harvest Moon）、阿约金公司（Aryogen）） | 拉普亮赛-T（lapuleucel-T）；AVX-901；AE-37；BMS-690514；MVA-BN-HER2；瓦尼替尼；MM-111；AC-480；卵巢癌疫苗（尖锐公司（Generex））；（玛格士西单抗（margetuximab））；泊西替尼；PR-610 | Her-VAXX；VM-206；ARRY-380；JNJ-26483327；S-222611；多柔比星（doxorubicin）（莫瑞麦克公司（Merrimack））；西帕替尼；TrasGEX；曲妥珠单抗（韩华化学公司（Hanwha Chemical））；曲妥珠单抗（辉瑞公司（Pfizer））；IDN-6439 | 洛瓦新 B（Lovaxin B）；TH-1（阿勒塔公司（Algeta））；曲妥珠单抗体结合物（辛赛恩公司）；CUDC-101；Her-2/neu 斯塔多体（Stradobody）（格利克尼克公司（Gliknik））；ARX-788；Etbx-021；SN-34003；IBI-302；NT-004；ICT-140；ONS-1050；Sym-013；抗 HER2 X 抗 CD3（艾默金生物方案公司（Emergent Biosolutions））；Z-650；乳癌疫苗（赛尔科学公司（Cel-Sci））；JNJ-28871063；曲妥珠单抗（普兰特弗姆公司（PlantForm）、拜沃艾克斯普莱斯公司（BioXpress）、双绿洲技术公司（biOasis Technologies）、斯塔达公司（Stada）、纳特克西公司（Natco）、酷瑞克西公司（Curaxys）、肿瘤生物公司 |

[0233]

| 可药化基因 | 已批准的 | 注册前（批准前） | III 期 | II 期 | I 期 | 临床前 |
|---|---|---|---|---|---|---|
| | | | | | | （Oncobiologics）、阿替沃金公司（Alteogen）、玛比恩公司（Mabion）） |
| ALK | 克卓替尼 | N | N | AP-26113；RG-7853；LDK-378；TSR-011；NMS-E628 | X-396；ASP-3026 | NMS-E628；极光激酶 + ALK 抑制剂（萨勒姆公司、阿斯利康公司）；ALK 抑制剂（阿斯利康公司、赛法隆公司、奥里基因公司）；ARN-5032；DLX-521 |
| NTRK1 | N | N | N | 马来酸米西西尼 | N | 酪氨酸激酶抑制剂（布里斯托尔-迈尔斯·斯奎布公司）；PLX-7486 |
| LTK | 克卓替尼 | N | N | N | N | N |
| BRAF | 帕唑帕尼（pazopanib）；维罗非尼（vemurafenib）；达拉菲尼（dabrafenib） | N | N | RAF-265；XL-281；LGX-818 | ARQ-761；ARQ-736 | AB-024；b-raf 抑制剂（萨勒姆公司）；BRAF 激酶抑制剂（赛勒克萨金治疗公司（Selexagen Therapeutics））；BeiGene-283；DP-4978；TL-241 |

[0234] 表8提供在表1-3中的基因融合体或表7和11中的基因变异体中鉴别的11个可药化目标的列表。表16和17提供表1-3或表7和11中的基因变异体内的其他可药化目标的分析。

表8、16和17提供关于可药化目标的信息，包括基因名称、药物是否已经被美国食品和药物管理局(U.S.Food and Drug Administration,FDA)批准(N＝否)，如果药物尚未被批准，那么临床试验处于哪个阶段(注册前、III期、II期、I期和临床前)。举例来说，与NOTCH1基因相关的药物尚未被批准，但到此次编写为止处于1期临床试验(参见OMP-52M51)。

[0235] 已批准的药物包括(但不限于)用于具有ALK基因融合体的疾病的克卓替尼以及用于具有CASR基因融合体的疾病的盐酸西那卡塞。许多被批准的药物已经鉴别用于具有EGFR的基因融合体，包括(但不限于)埃罗替尼；帕尼单抗；西妥昔单抗；奈皮德明；吉非替尼；尼妥珠单抗；凡德他尼；二甲苯磺酸拉帕替尼；以及盐酸埃克替尼。已批准的药物帕纳替尼已经鉴别用于具有FGFR3的疾病，芦可替尼已经鉴别用于具有JAK2基因融合体的疾病。许多已批准的药物已经鉴别用于具有RET的基因融合体，包括(但不限于)索拉非尼；凡德他尼；苹果酸舒尼替尼；卡博替尼；以及瑞格非尼。已批准的药物克卓替尼已经鉴别用于具有ROS1的疾病。可证明适用的其他药物包括(但不限于)克卓替尼、阿法替尼、马赛替尼、扎鲁突木单抗、来那替尼、多韦替尼乳酸盐、XL647、林多派皮姆、莱西突木单抗、达可替尼、SAR-302503、帕瑞替尼、哌立福新、布帕昔布、二磷酸莫替沙尼和阿帕替尼。

[0236] 本文中提供的方法可以包括向受试者或患者递送药物。所述药物可为根据政府药物监管部门(如FDA)已获批准的药物，或者所述药物可处于批准阶段之前的任何一个阶段。在说明性方面中，所述药物是FDA批准的药物。在其他方面中，所述药物可处于临床前、I期、II期、III期或批准前阶段。在某些方面中，本文中提供的方法包括向受试者递送表8、16和17中列出的一种或一种以上药物。当在受试者中鉴别出涉及表8、16和17中列出的一种以上基因的基因事件时，本文中提供的方法可以包括递送一种以上药物，尤其递送与通过所鉴别的基因事件影响的不同基因相关的药物。

[0237] 反义

[0238] 反义技术已应用于抑制各种致癌基因的表达。举例来说，相对于经对照正义转染物转染的细胞，在腺病毒2晚期启动子的控制下引入到人类鳞状癌中的呈反义定向的Craf-1cDNA片段使得致瘤潜能极大地降低。类似地，Cmyc反义构筑体在弗兰德鼠类红白血病细胞(Friend Murine Erythroleukemia cell)中加速分化且抑制$G_1$进展。反义方法采用核酸倾向于与“互补”序列配对的优点。互补序列为根据标准沃森-克里克互补规则(Watson-Crick complementarity rules)能够碱基配对的那些聚核苷酸。用聚核苷酸靶向双链(ds)DNA引起三螺旋形成；靶向RNA将引起双螺旋形成。反义聚核苷酸当引入到目标细胞中时特异性地结合于其目标聚核苷酸且干扰转录、RNA加工、转运、翻译和/或稳定性。反义RNA构筑体或编码所述反义RNA的DNA可用于在体外或体内(如在宿主动物包括人类受试者内)抑制宿主细胞内的基因转录或翻译或两者。

[0239] 反义可在启动子的转录控制下。“启动子”是指通过细胞的合成机制或引入的合成机制所识别的、为启动基因的特定转录所需要的DNA序列。短语“在转录控制下”意指启动子处于相对于核酸的正确位置和定向以控制RNA聚合酶启动以及基因的表达。

[0240] 在某些情况下，反义表达构筑体将包含病毒或来源于病毒基因组的经工程化构筑体。当采用cDNA插入时，可包括实现基因转录物的适当聚腺苷酸化的聚腺苷酸化信号。聚腺苷酸化信号的性质并未被认为是至关重要的且可采用任何所述序列。终止子可以用于增强信使水平且使从盒到其他序列中的同步读取减到最少。

[0241] 反义构筑体可经设计以结合于本文所公开的基因融合体或基因变异体的启动子和其他控制区、外显子、内含子或甚至外显子-内含子界限。最有效的反义构筑体包括与内含子/外显子剪接点互补的区域。一个实施例包括与在内含子-外显子剪接点的50-200个碱基内的区域互补的反义构筑体。已观察到一些外显子序列可在不严重地影响其目标选择性的情况下包括于构筑体中。所包括的外显子材料的量将变化，取决于所使用的特定外显子和内含子序列。可容易地通过在体外测试构筑体来测试是否有过多的外显子DNA简单地包括在内，以测定正常细胞功能是否受影响或者具有互补序列的相关基因的表达是否受影响。

[0242] 词语“互补”就反义来说意指在其全部长度上实质上互补且具有极少碱基错配的聚核苷酸序列。举例来说，长度为十五个碱基的序列当其在十三或十四个位置处具有互补核苷酸时可称为互补的。天然地，完全互补的序列将为在其整个长度中完全互补且不具有碱基错配的序列。还涵盖具有较低同源程度的其他序列。举例来说，可设计具备有限高同源区域而且含有非同源区域(例如核糖核酸酶)的反义构筑体。这些分子尽管具有小于50％的同源性，仍将在适当条件下结合于目标序列。

[0243] 反义的体内、离体或体外递送可涉及载体的使用。一种用于反义递送的有效载体是腺病毒表达载体。“腺病毒表达载体”意指包括含有足以支持构筑体包装以及表达已在其中克隆的反义聚核苷酸的腺病毒序列的那些构筑体。表达载体可以包括腺病毒的经基因工程化的形式。腺病毒因为其中等大小的基因组、易于操纵、高效价、宽目标细胞范围和高感染性而尤其适用作基因转移载体。病毒基因组的两个末端含有100-200个碱基对反转重复序列(ITR)，其为病毒DNA复制和包装所必需的顺式元件。

[0244] 腺病毒易于生长和操控，且在体外和体内展现较宽的宿主范围。此病毒群组可以高效价(例如$10^9$-$10^{11}$个噬斑形成单元/ml)获得，且其是高度感染性的。腺病毒的生命周期并不需要整合到宿主细胞基因组中。通过腺病毒载体递送的外来基因为游离型的，且因此对宿主细胞具有低基因毒性。在野生型腺病毒的情况下在疫苗接种研究中已报导无副作用，展示其作为体内基因转移载体的安全性和治疗潜力。

[0245] 逆转录病毒为一组单链RNA病毒，其具有能够通过逆转录方法在经感染细胞中将其RNA转化成双链DNA的特征。所得DNA接着以原病毒形式稳定整合到细胞染色体中且引导病毒蛋白质的合成。所述整合引起病毒基因序列保留在接受体细胞和其后代中。逆转录病毒基因组含有三种基因gag、pol和env，其分别编码衣壳蛋白、聚合酶和包膜组分。从gag基因上游发现的序列含有用于将基因组包装到病毒粒子中的信号。两个长末端重复(LTR)序列存在于病毒基因组的5’和3’末端。这些含有强启动子和增强子序列，且其也为在宿主细胞基因组中的整合所需。

[0246] 为了构筑逆转录病毒载体，将编码相关基因的核酸插入到病毒基因组中的某些病毒序列的位置中以产生有复制缺陷的病毒。为了产生病毒粒子，构筑含有gag、pol和env基因但没有LTR和包装组分的包装细胞系。当含有人类cDNA以及逆转录病毒LTR和包装序列的重组质粒被引入到细胞系中(例如通过磷酸钙沉淀)时，所述包装序列允许重组质粒的RNA转录物被包装到病毒粒子中，其接着分泌到培养基中。重组逆转录病毒接着被收集，任选地浓缩，且用于基因转移。逆转录病毒载体能够感染广泛的各种细胞类型。然而，整合和稳定表达需要宿主细胞的分裂。

[0247] 基于通过将乳糖残基化学添加到病毒包膜中来化学修饰逆转录病毒，近来开发了被设计成用于允许逆转录病毒载体的特定靶向的新颖方法。此修饰可允许经由唾液酸糖蛋白受体特异性感染肝细胞。

[0248] 其他病毒载体可用作表达载体。可采用来源于病毒(如牛痘病毒、腺相关病毒(AAV)和疱疹病毒)的载体。

[0249] 为了实现正义或反义基因构筑体的表达，表达载体可递送到细胞中。此递送可在体外(如在用于转化细胞系的实验室程序中)或在体内或离体(如在某些疾病病况的治疗中)实现。如上文所述，一种递送机制是经由病毒感染进行，其中表达构筑体在感染性病毒粒子中衣壳化。

[0250] 还涵盖用于将表达构筑体转移到经培养哺乳动物细胞中的数种非病毒方法。这些方法包括磷酸钙沉淀DEAE-葡聚糖、电穿孔、直接微注射、装载DNA的脂质体和脂染胺-DNA复合物、细胞超声处理、使用高速微粒的基因轰击以及受体介导的转染。这些技术中的一些可成功地经调适用于体内或离体用途。

[0251] 医药组合物-当涵盖临床应用时，可制造医药组合物，基因递送载体或经工程化细胞，呈适合于既定应用的形式。一般来说，这需要制备基本上不含热原质以及其他可能对人类或动物有害的杂质的组合物。

[0252] 适当盐和缓冲液用于使得递送载体稳定且允许被目标细胞摄取。当将重组细胞引入到患者中时，也将采用缓冲液。本发明的水性组合物包含有效量的到细胞的载体，其溶解或分散于医药学上可接受的载剂或水性介质中。短语“医药学上或药理学上可接受的”是指当投与动物或人时不生产不利、过敏或其他不良反应的分子实体和组合物。如本文所使用，“医药学上可接受的载剂”包括任何和所有溶剂、分散介质、涂料、抗细菌剂和抗真菌剂、等张剂以及吸收延迟剂等。此类介质和药剂用于医药学上活性物质的使用是本领域中熟知的。除非任何常规介质或试剂与本发明的载体或细胞不相容，预期其在治疗组合物中的用途。

[0253] 表达载体和递送媒剂可经由任何常见途径投与，只要目标组织经由所述途径可用即可。这包括了经口、经鼻、颊内、经直肠、经阴道或局部。可替代地，投与可通过原位、皮内、皮下、肌内、腹膜内或静脉内注射进行。所述组合物通常将以医药学上可接受的组合物形式投与。

[0254] 基于既定目的，确定治疗剂的有效量。术语“单位剂量”是指适用于受试者的物理上分散的单位，每一单位含有经计算以产生与其投与有关的所需反应的预定量的治疗性组合物，即，适当途径和治疗方案。待投与的量(均根据疗法数量和单位剂量)取决于待治疗的受试者、受试者的状态和所需保护。治疗性组合物的确切量还取决于医师的判断并且对于每一个体都是特有的。

[0255] 治疗试剂盒-抑制肿瘤细胞增殖所需的所有必需的材料和试剂可共同组装在试剂盒中。这通常将包含所选择的表达载体、病毒或细胞。还包括的可为供表达载体复制和用于所述复制的宿主细胞用的各种培养基。所述试剂盒将包含用于每一个别试剂的不同容器。所述试剂盒还可以包括说明书页，其定义(i)投与反义表达载体构筑体；(ii)反义表达病毒；以及(iii)反义表达细胞。

[0256] 在一些实施例中，提供干扰(iRNA或siRNA)。在一些实施例中，iRNA与融合体基因

的断裂点互补。

[0257] 与临床结果发现相关的方法

[0258] 本文提供的表15和39含有各种癌症类型的超过100种基因事件，包括获得功能的突变、失去功能的突变、峰内基因扩增/缺失和融合体事件，其以高统计显著性(q<0.1)与临床结果相关。因此，本文提供了用于向受试者递送治疗的方法、用于判定受试者是否接受治疗的方法、用于判定是否递送治疗的方法以及用于递送报导的方法。在某些说明性实施例中，所述治疗是药物。作为非限制性实例，药物可为表8、16和17中列出的药物，特别是当方法涉及会影响针对表8、16和17中的药物列出的基因的基因事件时。在其他实例中，如本文中所论述，药物可为由监管机构批准的任何药物，或处于批准之前开发阶段的任何药物。

[0259] 因此，在另一实施例中，提供一种向受试者递送治疗的方法，其中所述方法包括检测表15中鉴别的基因事件，以及治疗所述受试者，其中所述治疗被认为积极地影响具有所述基因事件的癌症患者的临床结果和/或被认为影响与所述基因事件相关的生物学路径。此实施例可视为一种用于判定受试者是否接受治疗的方法或一种用于判定是否对受试者递送治疗或执行治疗的方法。因此，本文提供一种判定受试者是否接受药物的方法，所述方法包括检测表15和/或39中鉴别的基因事件，且接着如果所检测的基因事件在表15和/或39中列出，那么向所述受试者递送药物，其中所述药物被认为积极地影响具有所述基因事件的患者的临床结果。在这些实施例的说明性方面中，基因事件与表8、16和17中发现的基因相关，且药物在表8、16和17中列出，作为所述基因的伴侣。受试者通常是患有表15和/或39中列出的类型的癌症的受试者。在此实施例的说明性方面中，基因事件与受试者的较差的预后相关，所述受试者罹患癌症，通常是表15和/或39中列出的癌症，所述受试者的较差的预后与所述基因事件相关。

[0260] 在另一实施例中，本文提供一种递送报导的方法，其中所述方法包括检测表15和/或39中鉴别的基因事件，且向医学专业人士递送会提供与受试者的癌症的所述基因事件相关的预测临床结果的报导。作为非限制性实例，医学专业人士可为医师、基因顾问或其他医学专业人士。通常，医师、基因顾问或其他医学专业人士与受试者具有专业关系，如患者/医生关系。所述报导可为纸质报导或可为经计算机网络递送给医学专业人士的电子报导。所述方法和报导可以包括表15和/或39中鉴别的基因事件和相关的预后中的一者或多者。

[0261] 在另一实施例中，本文提供一种确定向受试者投与哪种治疗的方法，所述方法包括检测表15中列出的基因事件，且取决于所检测的基因事件，投与治疗。在说明性实施例中，如果所检测的基因事件与较差的预后相关，那么治疗是侵袭性治疗，如患者将由于治疗而陷入更多疼痛和痛苦的治疗。在相关实施例中，如果所检测的基因事件与较差的预后相关，那么所述治疗是侵袭性较高的治疗，且如果所检测的基因事件是另一基因事件，特别是如果所检测的基因事件在表15和/或39中鉴别为表明良好的预后，那么所述治疗是侵袭性较低的治疗。举例来说，如果在肺癌腺癌患者中检测到AADAC基因缺失、CHD1L基因、FMO5基因或PRKAB2基因的扩增或其组合，那么患者可用侵袭性化学治疗药物方案治疗。如果在患者中未检测到这些基因事件，那么患者可被监测，但可能不投与化学治疗药物。

[0262] 在另一实施例中，本文提供一种用于判定是否治疗癌症患者的方法，所述方法包括检测表15和/或39中列出的基因事件，以及如果检测到在表15中与较差的预后相关的基因事件，那么治疗所述受试者。在另一实施例中，本文提供一种用于判定是否治疗癌症患者

的方法，所述方法包括检测表15和/或39中列出的基因事件，以及如果检测到在表15和/或23中与良好的预后相关的基因事件，那么不治疗所述受试者。在另一实施例中，本文提供一种用于判定是否治疗或监测癌症患者的方法，所述方法包括检测表15和/或39中列出的基因事件，以及如果检测到在表15和/或39中与良好的预后相关的基因事件，那么监测但不治疗所述受试者。如果监测检测到癌症的复发或进展，那么可稍后投与治疗。

[0263] 在基于与表15和/或39中的基因事件相关的临床结果涉及本文中提供的方法(例如向受试者递送治疗的方法、或判定是否向受试者递送治疗的方法、或确定投与或递送哪种治疗的方法或用于递送报导的方法)的本发明的这些实施例的某些方面中，受试者可被鉴别为具有表15和/或39中列出的任何一种类型的基因事件和任何一种特异基因事件。举例来说，基因事件可为获得功能的突变、失去功能的突变、基因扩增或缺失(通常是峰内基因扩增/缺失)、或融合体事件。在某些说明性实施例中，在表15和/或39中鉴别出q值为$1\times10^{-3}$或更小、$1\times10^{-4}$或更小或$1\times10^{-5}$或更小的基因事件。在某些方面中，基因事件在表15和/或39被列为涉及可药化基因。举例来说，基因事件可为表15和/或39中列出的与作为临床前药物目标的基因相关的基因事件。作为一个非限制性实例，本文提供一种确定向患有卵巢癌(例如卵巢浆液性囊腺癌)的患者投与哪种治疗或治疗疗程的方法，其中所述方法包括检测或以其他方式测定ID1或BCL2L1基因的扩增以及投与治疗。在说明性实施例中，治疗是用于BCL2L1的经批准的治疗，如目前FDA批准的BCL2L1治疗，其中检测到BCL2L1扩增。

[0264] 熟练的业内人士已知用于检测表15和/或39中列出的基因事件的类型的方法。那些方法可以包括核酸测序方法或扩增方法，如PCR或等温扩增方法或其组合。那些方法可以包括提供引物，所述引物经设计以结合于表15和/或39中鉴别的基因或结合表15和/或39中鉴别的基因的上游。因此，本文提供了反应混合物和试剂盒，其包括受试者的核酸样品以及结合于表15和/或39中鉴别的基因或其上游的一种或多种引物。通常，所述基因与表15和/或39中的基因事件相关，且所述受试者患有在表15和/或39中鉴别为具有与所述基因事件相关的预后的癌症。所述试剂盒还可包括通过如本文中针对本发明的各种实施例所公开的引物结合的对照核酸。所述反应混合物还可包括如本文中针对本发明的各种实施例所公开的聚合酶。

[0265] 在基于与表15和/或39中的基因事件相关的临床结果涉及本文中提供的方法(例如向受试者递送治疗的方法、或判定是否向受试者递送治疗的方法、确定递送哪种治疗的方法或用于向医学专业人士递送报导的方法)的本发明的这些实施例的某些方面中，基因事件可以包括表15和/或39中鉴别的一种以上基因事件。在某些方面中，根据这一实施例的方法检测表15中鉴别的2、3、4、5、6、7、8、9、10种或更多种基因事件，特别是被鉴别为对于既定癌症类型具有相同预后的那些基因事件。举例来说，所述方法可以包括检测乳癌患者中的基因事件，以及当所检测的基因事件包括BRF2、ERLIN2、GPR124、PROSC和TAB11FI基因中的两者或更多者的基因扩增时，向所述患者投与治疗。在另一实例中，所述方法包括检测罹患低级别胶质瘤的受试者中的两种或更多种基因事件，以及向所述受试者投与治疗，其中所述基因事件是以下各项中的至少两者：EGFR或SEC61G基因的扩增；CDK4、CYP27B 1、MARCH9、TSPAN31或AGAP2基因的扩增；EGFR基因中的获得功能的突变；或CDKN2A、CDKN2B或MTAP基因的缺失。在另一方面中，所述方法包括检测与较差的预后相关的基因事件，且所述基因事件在表8、16、17、表15和/或39中被鉴别为当前临床前试验中的药物或已批准的药物

(如FDA批准的药物)的目标。

[0266] 在基于与表15和/或39中的基因事件相关的临床结果涉及本文中提供的方法(例如向受试者递送治疗的方法、或判定是否向受试者递送治疗的方法、或确定投与或递送哪种治疗的方法或用于递送报导的方法)的本发明的这些实施例的某些方面中,基因事件可为本文中其他表中的一者中鉴别的特定基因事件。熟练的业内人士可以鉴别出表15和/或39中的哪种一般类型基因事件将落在其他表中的一者中的特定基因事件中。

[0267] 计算机实施的系统

[0268] 计算机系统可用于本发明的某些实施例中。在各种实施例中,计算机系统可以包括用于传达信息的总线或其他通信机制,以及与总线耦合以用于处理信息的处理器。在各种实施例中,计算机系统100还可包括耦合到总线以便测定碱基调用的存储器(其可为随机存取存储器(RAM)或其他动态存储装置)以及待通过处理器执行的指令。存储器也可以用于在执行打算由处理器执行的指令期间存储临时变量或其他中间信息。在各种实施例中,计算机系统可以进一步包括耦合到总线以便存储用于处理器的静态信息和指令的只读存储器(ROM)或其他静态存储装置。可提供存储装置(如磁盘或光盘),且其耦合到总线以便存储信息和指令。

[0269] 在各种实施例中,计算机系统可以经由总线耦合到显示器,如阴极射线管(CRT)或液晶显示器(LCD)以将信息显示给计算机用户。输入装置(包括字母数字和其他按键)可为与总线耦合以便将信息和命令选择传达给处理器。另一类型的用户输入装置是用于将方向信息和命令选择传达到处理器且用于控制显示器上的光标移动的光标控制件,如鼠标、轨迹球或光标方向键。这一输入装置通常具有在两个轴线(第一轴线(例如,x)和第二轴线(例如,y))上的两个自由度,其允许所述装置指定一个平面中的位置。

[0270] 计算机系统可以执行本发明教示。与本发明教示的某些实施方式一致,结果可通过计算机系统100回应于执行含于存储器中的一个或多个指令的一个或多个序列的处理器来提供。此类指令可以从另一计算机可读媒体(如存储装置)读取到存储器中。存储器中含有的指令序列的执行可以使得处理器执行本文所描述的方法。或者,可以使用硬连线电路代替或结合软件指令以实现本发明教示。因此,本发明教示的实施方式不限于硬件电路和软件的任何特定组合。

[0271] 在各种实施例中,如本文所用的术语“计算机可读媒体”是指参与为处理器提供指令以便执行的任何媒体。此类媒体可以呈许多形式,包括(但不限于)非易失性媒体、易失性媒体和传输媒体。非易失性媒体的实例可以包括(但不限于)光盘或磁盘,如存储装置。易失性媒体的实例可以包括(但不限于)动态存储器,如存储器。传输媒体的实例可以包括(但不限于)同轴电缆、铜线和光纤,包括包含总线的导线。

[0272] 非暂时性计算机可读媒体的常见形式包括(例如)软盘、软磁盘、硬盘、磁带、或任何其他磁性媒体、CD-ROM、任何其他光学媒体、穿孔卡片、纸带、具有孔洞图案的任何其他物理媒体、RAM、PROM和EPROM、闪存EEPROM、任何其他存储器芯片或盒带或计算机可以读取的任何其他有形媒体。

[0273] 根据各种实施例,被配置成被处理器执行以进行方法的指令存储在计算机可读媒体上。计算机可读媒体可以是存储数字信息的装置。举例来说,计算机可读媒体包括用于存储软件的如所属领域中已知的只读光盘(CD-ROM)。计算机可读媒体被适合于执行被配置成

被执行的指令的处理器访问。

[0274] 根据体现在本申请案中的教示和原理，提供可以在多个地点和实体中有效收集、分析、存储、转移、检索和/或分发信息(包括基因组学和/或患者信息)的方法、系统和计算机可读媒体。

[0275] 在一个实施例中，提供一种用于测定一种或多种基因融合体和/或变异体是否存在于样品中的系统。所述系统可以进一步确定鉴别与一种或多种基因融合体和/或基因变异体相关的疾病病况(如癌症)以及根据所述突变状态确定适当的治疗。在某些实施例中，所述系统包含与接收测序数据的测序仪器通信的处理器。

[0276] 在一些实施例中，所述处理器可以执行一次或多次变异体调用。在一些实施例中，处理器可以提供过滤和/或注释预测。

[0277] 实例

[0278] 在以下实例中，方法用于鉴别4,225个癌症患者样品中与一组19种癌症相关的基因融合体和基因变异体。所述基因融合体和基因变异体接着用于产生用以确认癌症倾向性、诊断癌症、癌症分期、提供预后以及鉴别可药化癌症的诊断方法。提供方法以基于基因融合体的鉴别来提供癌症的靶向疗法。

[0279] 实例1:表征癌症中的基因融合体景观的双端下一代测序数据的高通量系统分析

[0280] 跨越19种疾病的4,225个癌症患者样品用deFuse(麦克弗森(McPherson)等人“deFuse:用于肿瘤RNASeq数据中的基因融合体发现的算法(deFuse:an algorithm for gene fusion discovery in tumor RNASeq data)”《公共科学图书馆·比较生物学》(PLoS Comp.Bio.)2011)和TopHat(金(Kim)等人“TopHat-Fusion:用于发现新颖融合体转录物的算法(TopHat-Fusion:an algorithm for discovery of novel fusion transcripts)”《基因组生物学》(Genome Biology)2011)基因融合体调用软件使用基于云的计算基础设施处理。确认基因融合体事件的过滤准则，其富集高置信度的化学上经验证的基因融合体事件。

[0281] 基因融合体编码血液学和实体肿瘤中的致癌驱动基因，且通常与显著临床反应和适当靶向试剂相关。大规模平行双端测序可以鉴别肿瘤基因组和转录组中的结构重排。然而，鉴别基因融合体的计算方法会变化(仍然在演进中)且基本上对细胞系数据进行训练。已开发了表征已知致癌基因融合体以及发现癌症中的新颖基因融合体的系统方法。从癌症基因组图谱(the Cancer Genome Atlas,TCGA)的癌症基因组学中心(Cancer Genomics Hub,CGHub)获得来自16种癌症类型的约3,400个临床病例的RNASeq数据。调查数种基因融合体调用程序的性能，且选择两种(deFuse和TopHat)以用于目标是支持单端和双端数据两者的进一步方法开发。开发分析流水线，且在高性能计算簇上平行执行。对聚集的数据进行过滤和注释，作为允许探索性分析各种过滤的后处理步骤。对包括已知标准物(例如前列腺腺癌中的TMPRSS2.ERG、急性骨髓性白血病中的PML.RARA等)的数据集优化过滤方法，从而富集具有正确5'-3'定向的这些和其他基因融合体，同时排除具有模糊断裂点以及来自相邻基因的跨越读段、比对错误和通读转录物的情况。基于与多个搭配物一起参与融合体的独特基因和独特基因对各自在特定疾病内的出现，概述所预测的融合体。在所预测的融合体和所添加的重要验证证据阳性的情况下，在3'基因的所预测的断裂点之后观察到表达升高。泛疾病融合体和多搭配物融合体事件加宽基因融合体事件的临床群体范围。

[0282] 所有单端数据均使用TopHat处理，且所有双端数据均使用deFuse处理。TopHat已展示出在较长的75bp单端数据的情况下有效。deFuse算法与单端数据并不相容且已被设计成用于充分利用读段对。预处理数据且检测融合体：对高性能计算簇平行执行所有样品的deFuse TopHat步骤。对聚集的数据进行过滤和注释，作为启用过滤准则以使假阳性融合体减到最少的后处理步骤。用RNASeq外显子表达数据验证优先融合体的列表。

[0283] TCGA数据来源：从癌症基因组学中心(CGHub，当前的TCGA基因组数据的存储库-安全超文本传输协议://cghub.ucsc.edu/)获得基因融合体分析的所有RNASeq数据。表9列出针对M2和M3下载和处理的TCGA样品计数：

[0284] 表9：经处理的TCGA样品

[0285]

| 癌症类型 | 癌症类型缩写 | 样品 | 中心 | 仪器 |
|---|---|---|---|---|
| 膀胱尿道上皮癌 | BLCA | 122 | UNC-LCCC | Illumina HiSeq 2000 |
| 侵袭性乳癌 | BRCA | 841 | UNC-LCCC | Illumina HiSeq 2000 |
| 宫颈鳞状细胞癌和宫颈内腺癌 | CESC | 88 | UNC-LCCC | Illumina HiSeq 2000 |
| 结肠腺癌 | COAD* | 196 | UNC-LCCC | Illumina GA IIx |
| 多形性成胶质细胞瘤 | GBM | 167 | UNC-LCCC | Illumina HiSeq 2000 |
| 头颈鳞状细胞癌 | HNSC | 302 | UNC-LCCC | Illumina HiSeq 2000 |
| 嫌色性肾癌 | KICH | 66 | UNC-LCCC | Illumina HiSeq 2000 |
| 肾脏肾透明细胞癌 | KIRC | 480 | UNC-LCCC | Illumina HiSeq 2000 |
| 肾脏肾乳头状细胞癌 | KIRP | 76 | UNC-LCCC | Illumina HiSeq 2000 |
| 急性骨髓性白血病 | LAML | 179 | BCCAGSC | Illumina GA IIx |
| 脑低级别胶质瘤 | LGG | 184 | UNC-LCCC | Illumina HiSeq 2000 |
| 肝脏肝细胞癌 | LIHC | 34 | UNC-LCCC | Illumina HiSeq 2000 |
| 肺腺癌 | LUAD | 345 | UNC-LCCC | Illumina HiSeq 2000 |
| 肺鳞状细胞癌 | LUSC | 221 | UNC-LCCC | Illumina HiSeq 2000 |
| 卵巢浆液性囊腺癌 | OV | 417 | BCCAGSC | Illumina HiSeq 2000 |
| 胰腺癌 | PAAD | 31 | UNC-LCCC | Illumina HiSeq 2000 |
| 前列腺腺癌 | PRAD | 140 | UNC-LCCC | Illumina HiSeq 2000 |
| 直肠腺癌 | READ* | 71 | UNC-LCCC | Illumina GA IIx |
| 皮肤黑素瘤 | SKCM | 267 | UNC-LCCC | Illumina HiSeq 2000 |
| 胃腺癌 | STAD | 41 | BCCAGSC | Illumina HiSeq 2000 |
| 甲状腺癌 | THCA | 373 | UNC-LCCC | Illumina HiSeq 2000 |
| 子宫体子宫内膜样癌 | UCEC* | 317 | UNC-LCCC | Illumina GA IIx |

[0286] *单端TCGA疾病

[0287] BAM文件从CGHub使用其基因激流软件(Gene Torrent Software)下载

[0288] 在支持单端和双端数据两者的目标下，4,374个双端样品用deFuse处理，且584个单端样品用TopHat处理。

[0289] 广泛地，分析流水线由5个主要步骤组成：1.预处理原始数据以获得FASTQ文件；2.运行融合体调用程序；3.过滤断裂点到相关基因区域；4.用Oncomine转录物组注释断裂点；以及5.对可能令人感兴趣的新颖融合体进行概述和优先级排序。

[0290] 进到融合体调用程序中的输入由呈FASTQ格式的RNASeq读段组成，其需要将由TCGA提供的BAM文件转换成(分别)单端或双端数据的一或两个FASTQ文件。

[0291] 开发了定制SamToFastq转换器以从TCGA BAM文件生成FASTQ文件。除了允许系统地转换所有双端RNASeq TCGA BAM之外，SamToFASTQ转换器具有优于其他转换工具的其他

优点。首先,其以C编写且经编译以运行较快且减少处理时间。其次,其结合了数个验证步骤以确保适当的配偶配对和输出FASTQ文件中一致的配偶对排序,其两者均为融合体调用程序的输入需求。

[0292] 有3种癌症类型(COAD、READ、UCEC)仅以单端RNASeq数据形式可用。对于单端BAM文件转换,使用程序BamTools(安全超文本传输协议://github.com/pezmaster31/bamtools)来生成FASTQ文件。

[0293] 整合-图1图式是通过两种调用程序进行的结果过滤的相对层面。作为分析的一部分,整合“层面I”数据,来自TopHat-Fusion Post的输出potential_fusion.txt文件和来自deFuse的输出results.classify.tsv文件。整合步骤涉及将所报导的断裂点转化成基于基因组学坐标系的断裂点且合并成常见文件格式。

[0294] 断裂点过滤-过滤来自调用程序的“层面I”输出的约550万预测,仅保留其中每一断裂点是在RefSeq转录物的5'UTR或CDS区中的那些调用(refGene大约于2012年7月18日,从UCSC获得)。如此进行以针对含有功能基因区域的融合体富集所预测的融合体。还排除经预测会出现在内含子序列中的断裂点,产生423,587个所预测嵌合体的集合。

[0295] 断裂点注释-对于每对断裂点,每Entrez ID仅保留一种转录物。这确保了在相同位置处的注释断裂点中的一致性。然而,相同基因搭配物的不同位置处的所预测的断裂点仍可产生多个代表一对基因的转录物,这是替代转录物的可能的证据。

[0296] 丢弃来自调用程序的基本注释,因为其是基于每一对应调用程序的默认注释源。然而,保留TopHat和deFuse两者的某些输出字段以帮助对所预测的融合体进行优先级排序。另外,从其他调用程序特性推断未由调用程序明确报导的某些注释特性。

[0297] 推断的特性-从每一调用程序获得支持和跨越读段计数,且概述在读段跨度(Reads Span)和读段跨度支持(Reads Span Support)中。后者是跨越融合体的读段和支持融合体的读段的总和。将由调用程序报导的断裂点序列修整以在融合体的每一侧上包括50个碱基,且合并为断裂点序列。融合体断裂点通过“|”描绘。由于调用程序两者中无一提供决定性的‘5’’或‘3’’标记,故通过将调用程序参数与基因链注释组合来推断融合体搭配物的相对5'-3'定向。如果存在既定基因融合体调用的推断的5'和3'搭配物,那么将有效定向字段标记为“Y”。

[0298] RepeatMasker注释-每一预测的断裂点位置还用在断裂点附近的RepeatMasker特征注释。如此进行以鉴别高度重复的基因组区域中的断裂点,其中比对错误很可能会影响嵌合转录物的预测。在每一融合体预测中,如果侧接序列的断裂点中的任一者通过12个或更多个碱基与RepeatMasker元件重叠,那么将RepeatMasker重叠字段设定为1。重叠融合体调用的频率用于下文所述的Oncomine优先级排序,使得将具有较低重叠频率的基因融合体预测视为较高质量。

[0299] 融合体外显子表达不平衡-再现的Oncomine优先级融合体使用利用GDAC Firehose工具下载的RNASeq外显子表达数据观测,以通过寻找在断裂点调用之前和之后的外显子表达不平衡来提供真阳性融合体事件的次要证据。具体来说,如果3'搭配物的表达受到5'搭配物的启动子区影响,那么外显子表达应在所预测的断裂点之后增强。当查看融合对比非融合患者样品时,这一效应尤其可见。

[0300] 将每一患者的RPKM RNASeq值以对应于来自数个不同基因定义(包括Refseq)的

UCSC外显子的复合物的基因注释格式(GAF)特征形式列出。肯潘迪亚公司(Compendia)处理的融合体断裂点映射到GAF特征。396,298个Refseq外显子中的80.8%完美地映射到下文展示的曲线中的GAF特征。选择且报导会产生最大重叠的Refseq外显子和GAF特征对。

[0301] 称为rg_pct的值基于下式提供给定Refseq外显子与GAF特征的映射质量的量度：rg_pct＝重叠/长度$_{refseq}$*重叠/长度$_{GAF特征}$

[0302] rg_pct值为1的映射完美地重叠,而值小于1指示refseq外显子或GAF特征不映射到严格相同的基因组区域,且RPKM值可为可疑的。归因于条形图中展示的疾病覆盖率,除了OV、STAD和LAML,为所有疾病选择RNASeq V2数据。

[0303] 针对满足以下准则的Oncomine优先级融合体子集的表达不平衡人工地审查融合体外显子表达:1.再现的Oncomine优先级融合体;2.在Mitelman数据库中列出的Oncomine优先级融合体;3.一种融合体搭配物是Oncomine获得功能的致癌基因且涉及至少3种Oncomine优先级融合体;以及4.一种融合体搭配物在桑格癌症基因普查(Sanger Cancer Gene Census,超文本传输协议://www.sanger.ac.uk/genetics/CGP/Census/)中列出且涉及至少3种Oncomine优先级融合体。

[0304] 总共994种基因融合体满足这些准则,且通过将“支持的”、“反驳的”、“中立的”或“不经测试的”评级分配给基因融合体调用来人工地审查其外显子表达不平衡。

[0305] 专家使用以下准则来分配评级:支持的:融合的样品在断裂点之后具有高度表达的3’融合体搭配物,使得融合的样品是患者群体的离群值。在断裂点之前,3’搭配物的表达与在断裂点之后相比应较低。反驳的:5’搭配物的极低平均表达(<5RPKM)或一种搭配物的平均表达与其他相比远为较低(约1/10)。中立的:既不满足支持也不满足反驳准则。将未经人工审查的融合体分配不经测试的评级。

[0306] 融合体概述-基于独特基因对的出现以及基于个别基因(有可能与多个搭配物一起)的出现,概述疾病内的融合体。

[0307] 融合体层面概述-对于独特融合体对(根据Entrez ID对唯一),根据任一调用程序具有至少一个融合体预测的疾病内的样品数量是融合的样品计数。由于在一个样品中以及跨越多个样品可报导相同基因对的多个断裂点,故由424K+融合体调用代表的每种疾病内的独特融合体对的数量为49,588个。表10展示针对跨越个别预测的给定融合体搭配物对概述的特性：

[0308] 表10：

[0309]

| 特性 | 概述方法 |
|---|---|
| DEFUSE翻转 | 总融合体调用%＝‘Y’ |
| DEFUSE有效定向 | 总融合体调用%＝‘Y’ |
| DEFUSE多重映射数量 | 总融合体调用%>0 |
| TOPHAT有效定向 | 总融合体调用%＝‘Y’ |
| 3P/5P_REPEATMASKER重叠 | 总融合体调用%＝1 |

[0310] 如果基因组上的基因相隔<1Mb,那么为融合体设定邻近标记,且在这些融合体搭配物的≤75%的个别融合体预测中设定defuse翻转标记。

[0311] Mitelman交叉参考-个别独特融合体交叉参考基因组学畸变的Mitelman数据库(超文本传输协议://cgap.nci.nih.gov/Chromosomes/Mitelman,2013年2月25日下载)。匹

配基于基因名称而非疾病类型进行。因此，Mitelman中报导的某一疾病中的基因融合体可能已出现在TCGA数据集中的不同疾病类型中。在基因层面处概述的基因融合体基于基因名称交叉参考Mitelman数据库。因此，如在Mitelman中报导为与所预测的独特融合体对相比具有不同组织学或完全不同畸变类型(例如较大染色体层面缺失而非融合体)的基因有更多潜力。

[0312] 正常样品融合体黑名单-为了减少假阳性融合体的数量，跨越10种疾病的344个双端正常样品使用上文所述相同的deFuse流水线处理。观察由6,024个独特融合体组成的总共56,579个总融合体调用。在49,588个独特基因融合体事件中，在正常样品中观察到这些调用中的11,801个。这些正常样品融合体调用用于生成黑名单，且由此从Oncomine优先级基因融合体中移除这些假阳性。

[0313] 平行同源融合体搭配物黑名单-在平行同源基因家族成员之间的融合体的黑名单使用两种策略汇编：1)人工地检测高频率融合体搭配物基因名称以及2)比较所有优先级融合体搭配物基因名称的前3个字符。在后一种策略中，融合体搭配物使用HomoloGene、Ensembl和SIMAP验证为“平行同源”，随后包括在最终黑名单中。黑名单由375个独特平行同源基因融合体组成且用于从Oncomine优先级基因融合体中移除假阳性。

[0314] 实例2：用于鉴别与癌症相关的基因变异体的NGS突变方法

[0315] 基因变异体的数据整合的目的是为了形成目前可从TCGA获得的一组最完整的突变数据。

[0316] 数据来源-对此版本，整合以下各者：来自博德GDAC Mutation_Packager2013_02_22stddata build的TCGA突变数据、到2013年3月1日为止可从TCGA DCC获得的2级(公共的、实验上未经验证的)数据，以及针对前列腺腺癌的由来自TCGA原始数据的肯潘迪亚公司生成的突变数据。

[0317] 肯潘迪亚公司(CBI)突变调用-存在可从TCGA获得的前列腺腺癌突变调用具有低质量且产生假阳性‘获得功能’预测的问题。因此，对此疾病的所有调用均来源于肯潘迪亚公司的自身突变调用流水线，其与由TCGA癌症类型工作组使用的用于生成公开可获得的突变调用的方法紧密相似。

[0318] 表12：数据来源选择

[0319]

| 癌症类型 | TCGA 疾病 | Mutation Packager (2013_02_22) | DCC (20130301) | 肯潘迪亚公司 | 总计 |
|---|---|---|---|---|---|
| 膀胱尿道上皮癌 | BLCA | 28 | | | |
| 侵袭性乳癌 | BRCA | 772 | | | |
| 宫颈鳞状细胞癌和宫颈内腺癌 | CESC | 39 | | | |
| 结肠腺癌 | COAD | 153 | | | |
| 多形性成胶质细胞瘤 | GBM | 290 | | | |
| 头颈鳞状细胞 | HNSC | | 306 | | |

[0320]

| 癌症类型 | TCGA 疾病 | Mutation Packager (2013_02_22) | DCC (20130301) | 肯潘迪亚公司 | 总计 |
|---|---|---|---|---|---|
| 癌 | | | | | |
| 肾脏肾透明细胞癌 | KIRC | 293 | | | |
| 肾脏肾乳头状细胞癌 | KIRP | 100 | | | |
| 急性骨髓性白血病 | LAML | 196 | | | |
| 脑低级别胶质瘤 | LGG | 169 | | | |
| 肺腺癌 | LUAD | | 379 | | |
| 肺鳞状细胞癌 | LUSC | 178 | | | |
| 卵巢浆液性囊腺癌 | OV | 316 | | | |
| 胰腺癌 | PAAD | 34 | | | |
| 前列腺腺癌 | PRAD | | | 170 | |
| 直肠腺癌 | READ | 68 | | | |
| 皮肤黑素瘤 | SKCM | 252 | | | |
| 胃腺癌 | STAD | | 136 | | |
| 甲状腺癌 | THCA | 323 | | | |
| 子宫体子宫内膜样癌 | UCEC | 235 | | | |
| | | | | | 4,437 |

[0321] 数据清洗-执行一些简单净化操作以移除源数据中存在的一式两份突变记录。移除来自同一患者样品的各种肿瘤/正常等分试样对的一式两份突变。还从下游分析流水线中排除总共25个“超突变基因”样品(每个样品突变计数>5,000)。在某些疾病(如子宫体子宫内膜样癌)中,数个高度突变的样品可能在整体突变计数中占主导地位且稀释肯潘迪亚公司突变和基因分类流程所必需的突变再现分析的结果。

[0322] 突变注释:A.肯潘迪亚公司注释-因此,定义突变的肯潘迪亚公司方法依赖于准确的变异体注释;突变使用标准注释流水线再注释,所述标准注释流水线确保跨越疾病类型的突变被一致地评估且在潜在致癌基因或肿瘤抑制基因的提名期间经受共同的解释。

[0323] 从TCGA获得的突变通过肯潘迪亚公司根据以下通用步骤处理:1.每一突变首先使用肯潘迪亚公司转录物组再注释。成功注释的突变接受源于肯潘迪亚公司的注释,而其余部分保留从TCGA获得的注释。注释包括:变异体分类、变异体位置、变异体变化。2.移除多种转录物中的突变的冗余注释。3.移除位于相关基因区域外的突变。4.移除无有效基因Entrez ID的突变。

[0324] “突变”在本文中定义为基因组位置处的特定变化,即:染色体、启动、终止、参考碱基、替代碱基、变异体类型(SNP、INS、DEL)等。

[0325] “注释”在本文中定义为一组转录物特有的特性,其描述突变的效应,即:基因、转录物、变异体分类、变异体变化、变异体密码子位置等。

[0326] 在突变注释步骤中,针对通过肯潘迪亚公司编译的一组标准转录物再注释从TCGA获得的突变。这个转录物组包括从UCSC获得的来自hg18和hg19基因组构建的RefGene转录物。

[0327] 每一突变个别地针对规定的基因组构建内的CBI转录物组中的重叠群映射。SNP突变直接映射到其起始位置,同时针对小插入(INS)和缺失(DEL)突变选择相关位置以便映

射。

[0328] 对于成功地映射到转录物的突变，相对于所述转录物推断CBI突变注释。对于无法映射的突变，保留更有限的TCGA注释，且基于基因组学坐标构筑用于热点计算的变异体位置。

[0329] 下文是用于注释映射到CBI转录物组的突变的准则的描述：

[0330] 变异体分类：对于成功映射到转录物的每一突变，使用突变的位置和序列变异体类型推断变异体分类。此方法鉴别出以下主要突变变异体分类：

[0331] 表13：主要突变变异体分类：

[0332]

| 变异体分类 | 转录物区域 |
|---|---|
| 剪接位点 | 外显子或内含子 |
| 3'UTR、5'UTR | UTR外显子 |
| 内含子 | 内含子 |
| 错义、无义、无终止、沉默 | 编码外显子 |
| 框移插入/缺失 | 编码外显子 |
| 框内插入/缺失 | 编码外显子 |
| 非编码外显子 | 非编码基因的外显子 |

[0333] 变异体位置：突变的变异体位置是用于鉴别具有热点突变的基因的位置，热点突变为在多个肿瘤样品中在同一位置处观察到的某一分类的突变。为了有效地鉴别再现且定义每一突变的热点，构筑涵盖突变位置、所影响的氨基酸或碱基身份以及变异体分类的突变点识别符。使在同一位置处出现的与其产生的特定碱基变化无关的突变聚集。因此，仅参考碱基或氨基酸用于定义变异体位置。这确保了将针对可能的热点对会影响同一密码子或基因组位置的突变进行计数，即使其产生的替代等位基因不同。举例来说，对于给定基因，错义突变V600E、V600F和V600G将均具有V600的变异体位置，且将由此在鉴别热点突变时聚集在一起。在氨基酸层面位置不可获得时，使用RNA层面或基因组学层面位置。

[0334] 对于并不映射到CBI转录物组且因此并不具有基于转录物的位置的突变，将基因组位置(起始位置)和参考核苷酸(参考等位基因)用作与编码区或剪接位点接近性无关的变异体位置。接着，将TCGA注释的变异体分类作为后缀添加。这些突变的变异体变化(参见下文)未经定义。

[0335] 变异体变化：变异体变化提供关于突变的替代等位基因变化(例如V600E)的HGVS样信息。对于编码区中的SNP突变，变异体变化为完整HGVS蛋白质层面序列变异体描述，指示替代氨基酸。对于在编码区之外的SNP，提供替代等位基因核苷酸碱基。对于并不映射到CBI转录物组的突变，保留来自TCGA的变异体分类。

[0336] 转录物过滤：为了避免检索多种转录物，且因此避免检索一种基因内单一突变的多个注释，保持每一基因(唯一Entrez ID)每一突变仅一种转录物。如果突变映射到一种基因的数种转录物，那么仅选择一种转录物。然而，如果突变映射到数种基因，那么每一基因仅选择一种转录物。因此，一种突变可接受两种不同注释，但仅当其来自具有不同Entrez ID的转录物时才如此。实际上，在同一基因组位置处的相同变异体分类的任何突变总是分配给相同转录物，且因此在计算热点鉴别的再现时将处于同一参考框架中。

[0337] 基因区域过滤：所有突变均通过变异体类型和类别进一步过滤以避免将较不相关

的突变包括到基因功能分析中。滤除并不解析到基因区域的突变，或者因为其显著远落在转录物之外，或者因为其处于与RefSeq基因不相关的位置中。这些突变通过其缺乏基因识别符或以下变异体类别中的成员资格而为明显的：内含子、5'侧接序列、IGR和miRNA。还滤除具有DNP、TNP、ONP、复杂取代(Complex_substitution)和Indel的变异体类型的突变，因为其注释不被流水线所支持。

[0338] 将突变归类为热点、有害或其他-分析流水线中的下一个步骤基于突变的变异体位置鉴别出多个样品中重现的突变，且将其归类到热点、有害或其他变异体类别中。对于此步骤和后续频率计算，独立地处理每种疾病类型的突变。仅将相同变异体分类的突变共同运算，因此例如分别计数在相同位置处的错义突变和沉默突变。

[0339] 为了鉴别驱动基因事件，将给定Entrez Gene Id的每一突变归类为“有害”或“热点”。如果在3个或更多个肿瘤样品的相同变异体位置中观察到一种突变，那么认为所述突变是‘再现的’。如果一种突变是再现的且经以下变异体分类中的一者注释：框内插入/缺失、无终止、错义、非编码外显子，那么所述突变属于“热点”变异体类别。如果一种突变经以下变异体分类中的一者注释：框移插入/缺失、无义，那么所述突变属于“有害”类别。如果一种突变不符合上述准则，那么所述突变被视为在“其他”变异体类别中。

[0340] 命名“获得功能”和“失去功能”基因-将个别基因归类到所预测的功能类别中，即“获得功能”、“再现的其他”和“失去功能”，以反映其在潜在活化或有害突变中的相对富集。

[0341] 突变的频率：相对于一种疾病类型内的所有样品中的给定变异体分类和变异体类别计算每一基因的突变频率。一种疾病内的一种基因的整体突变频率通过组合所有突变来计算。

[0342] 突变显著性：通过选择再现最多的突变m且使用取样来计算一种疾病内的每一基因的热点p值，以测定在所述位置处观察r个或更多个突变的概率p。更确切地说，$p=\frac{100{,}000-\sum_{m=1}^{r-1}c_m}{100{,}000}$，其中$c_m$是具有最大多重性m的重复的计数。最大多重性为一的转录物的P值被定义为1.0。具有从未观察到的最大多重性的转录物的P值被定义为1e-5。

[0343] 通过对突变至少一次的转录物的数量(N)进行计数且计算每一p值的等级来计算每种疾病内的热点Q值。给定p值的q值则是Q＝p*N/等级。

[0344] 为了评估一种基因与其他基因相比是否针对有害突变显著富集，鉴于背景突变率，执行费舍尔精确检验(Fisher's exact test)，比较所讨论的基因的有害突变频率与其他基因的有害突变频率。将无义突变、框移插入和框移缺失归类为有害突变，而任何其他类型(错义等，但非基因间)的突变作为其他计数。

[0345] 通过计数具备有害突变的基因的数量(N)且计算每一相关的等级来计算每种疾病内的有害Q值。给定p值的q值则为Q＝p*N/等级。

[0346] 基因分类：一旦将突变分类，就将个别基因命名为三种类别中的一种：“获得功能”、“失去功能”和“再现的其他”。分类是基于在基因中观察到的突变的相对频率和显著性的组合。每一基因突变的显著性通过p值评估。图2中的分类流程指明用于获得功能和失去功能基因的准则。

[0347] “获得功能”基因将具有相对较高的热点错义突变频率和较低的有害突变频率，而“失去功能”基因含有较大部分的有害突变。“再现的其他”倾向于含有再现的插入/缺失突

变,其中的一些(例如1个碱基的再现的框移插入缺失)展现出可能由局部比对错误引起的潜在假阳性调用的迹象。

[0348] 泛癌症分析-为了概述跨越疾病的突变,对于疾病内分析执行相同的计算,但不根据疾病将突变记录分层。对于泛疾病基因分类,概述所有癌症类型中的基因(根据Entrez ID唯一)。

[0349] 实例3:用于鉴别癌症中的基因融合体和/或基因变异体的诊断分析

[0350] 文库制备

[0351] PCR扩增基因组DNA目标

[0352] 所公开的变异体和融合体聚核苷酸可通过核酸的测序来检测。这可通过下一代测序(以下是其描述)来实现。用于下一代测序的核酸来源可以包括新鲜冷冻的石蜡包埋的(FFPE)样品。

[0353] 执行多重聚合酶链反应以扩增一个基因组DNA样品内的384个个别扩增子。开发大于32,000个引物的池,涵盖超过100种基因变异体或融合体聚核苷酸。引物池中的每个引物经设计以在每个引物末端附近含有至少一个尿苷核苷酸。每个引物还经设计以通过形成引物对来与核酸样品的特定基因、基因变异体或融合多肽选择性地杂交,且促进扩增。

[0354] 向96孔PCR板的单一孔中添加5微升引物池(含有浓度为15μM于TE中的384个引物对)、10-50ng基因组DNA和10微升扩增反应混合物(2×AmpliSeq HiFi Master Mix,其可以包括甘油、dNTP和Platinum®Taq高保真度DNA聚合酶,英杰公司(Invitrogen),目录号11304)与无DNA酶/RNA酶的水(加利福尼亚州的生命技术公司,部件号600004)达到20微升的最终体积。

[0355] 将PCR板密封且装载到热循环仪(GeneAmp®PCR系统9700双重96孔热循环仪(加利福尼亚州的生命技术公司,部件号N8050200和4314445))中,且使用以下温度特性曲线运行以产生预扩增的扩增子文库。

[0356] 初始保持阶段在98℃下执行2分钟,接着16个变性循环在98℃下执行15秒,且粘接和延伸阶段在60℃下执行4分钟。在循环之后,将预扩增的扩增子文库保持在4℃下,直到前进到下文概述的纯化步骤。

[0357] 纯化来自输入DNA和引物的扩增子

[0358] 发现两轮的Agencourt® AMPure®XP试剂(加利福尼亚州的贝克曼库尔特公司(Beckman Coulter,CA)结合,洗涤以及以0.6×和1.2×体积比洗脱会移除基因组DNA和未结合或过量的引物。本文中概述的扩增和纯化步骤产生长度为约100bp到约600bp的扩增子。

[0359] 在1.5ml LoBind试管(艾本德公司(Eppendorf),部件号022431021)中,预扩增的扩增子文库(20微升)与12微升(0.6×体积)的Agencourt® AMPure®XP试剂(加利福尼亚州的贝克曼库尔特公司)组合。上下吹吸珠粒悬浮液以将珠粒悬浮液与预扩增的扩增子文库彻底混合。接着,将样品脉冲自旋且在室温下孵育5分钟。

[0360] 将含有样品的试管置于磁性机架(如DynaMag™-2自旋磁体(加利福尼亚州的生命技术公司,部件号123-21D)上2分钟以捕获珠粒。一旦溶液澄清,就将上清液转移到新试管中,其中向上清液中添加24微升(1.2×体积)的AgenCourt® AMPure®XP珠粒(加利福尼亚州的贝克曼库尔特公司)。吹吸混合物以确保珠粒悬浮液与预扩增的扩增子文库混合。接

着，将样品脉冲自旋且在室温下孵育5分钟。将含有样品的试管置于磁性机架上2分钟以捕获珠粒。一旦溶液澄清，就将上清液小心地丢弃，且不扰乱珠粒集结粒。所需预扩增的扩增子文库接着结合于珠粒。在不从磁性机架移出试管的情况下，将200微升新鲜制备的70％乙醇引入到样品中。将样品孵育30秒，同时将试管在磁性机架上温和地旋转。在溶液澄清之后，丢弃上清液，且不扰乱集结粒。执行第二次乙醇洗涤，且丢弃上清液。通过脉冲自旋试管来移除任何剩余乙醇，且小心地移出残余乙醇而不扰乱集结粒。将集结粒在室温下风干约5分钟。

[0361] 一旦试管干燥，就将试管从磁性机架移出，且添加20微升无DNA酶/RNA酶的水(加利福尼亚州的生命技术公司，部件号600004)。对试管进行涡旋和吹吸以确保样品彻底混合。将样品脉冲自旋且置于磁性机架上两分钟。在溶液澄清之后，将含有洗脱的DNA的上清液转移到新试管中。

[0362] 磷酸化扩增子

[0363] 向洗脱的DNA(约20微升)中添加3微升DNA连接酶缓冲液(英杰公司，目录号15224041)、2微升dNTP混合物和2微升FuP试剂。将反应混合物彻底混合以确保均匀性且在37℃下孵育10分钟。

[0364] 将衔接子连接于扩增子且纯化经连接的扩增子

[0365] 在孵育之后，反应混合物直接进行到连接步骤。这里，将现含有磷酸化的扩增子文库的反应混合物与1微升A/P1衔接子(各自20μm)(作为Ion片段文库试剂盒的组分销售，生命技术公司，部件号4466464)和1微升DNA连接酶(作为Ion片段文库试剂盒的组分销售，生命技术公司，部件号4466464)组合，且在室温下孵育30分钟。

[0366] 在孵育步骤之后，将52微升(1.8×样品体积)AgenCourt® AMPure®试剂(加利福尼亚州的贝克曼库尔特公司)添加到经连接的DNA中。将混合物彻底吹吸以将珠粒悬浮液与经连接的DNA混合。将混合物脉冲自旋且在室温下孵育5分钟。样品再经历脉冲自旋且置于磁性机架(如DynaMag™-2自旋磁体(加利福尼亚州的生命技术公司，部件号123-21D))上两分钟。在溶液澄清之后，丢弃上清液。在不从磁性机架移出试管的情况下，将200微升新鲜制备的70％乙醇引入到样品中。将样品孵育30秒，同时将试管在磁性机架上温和地旋转。在溶液澄清之后，丢弃上清液，且不扰乱集结粒。执行第二次乙醇洗涤，且丢弃上清液。通过脉冲自旋试管来移除任何剩余乙醇，且小心地移出残余乙醇而不扰乱集结粒。将集结粒在室温下风干约5分钟。

[0367] 将集结粒再悬浮在20微升无DNA酶/RNA酶的水(加利福尼亚州的生命技术公司，部件号600004)中且涡旋，以确保样品彻底混合。将样品脉冲自旋且置于磁性机架上两分钟。在溶液澄清之后，将含有经连接的DNA的上清液转移到新Lobind试管(艾本德公司，部件号022431021)中。

[0368] 缺口翻移且扩增扩增子文库并纯化文库

[0369] 将经连接的DNA(约20微升)与76微升Platinum®PCR高保真度超混合液(SuperMix High Fidelity)(加利福尼亚州的生命技术公司，部件号12532-016，作为Ion片段文库试剂盒的组分销售，生命技术公司，部件号4466464)和4微升文库扩增引物混合物(各自5μM)(加利福尼亚州的生命技术公司，部件号602-1068-01，作为Ion片段文库试剂盒的组分销售，生命技术公司，部件号4466464)组合，将混合物彻底吹吸以确保均匀的溶液。将溶液施加到96

孔PCR板的单一孔中且密封。将板装载到热循环仪(GeneAmp®PCR系统9700双重96孔热循环仪(加利福尼亚州的生命技术公司,部件号N8050200和4314445))中,且在以下温度特性曲线上运行以产生最终扩增子文库。

[0370] 缺口翻译在72℃下执行1分钟,接着酶活化阶段在98℃下执行2分钟,接着5-10个循环的变性在98℃下执行15秒,且粘接和延伸阶段在60℃下执行1分钟。在循环之后,将最终的扩增子文库保持在4℃下,直到行进到下文概述的最终纯化步骤。

[0371] 在1.5ml LoBind试管(艾本德公司,部件号022431021)中,最终扩增子文库(约100微升)与180微升(1.8×样品体积)的Agencourt® AMPure®XP试剂(加利福尼亚州的贝克曼库尔特公司)组合。上下吹吸珠粒悬浮液以将珠粒悬浮液与最终扩增子文库彻底混合。接着,将样品脉冲自旋且在室温下孵育5分钟。

[0372] 将含有最终扩增子文库的试管置于磁性机架(如DynaMag™-2自旋磁体(加利福尼亚州的生命技术公司,部件号123-21D))上2分钟以捕获珠粒。一旦溶液澄清,就将上清液小心地丢弃,且不扰乱珠粒集结粒。在不从磁性机架移出试管的情况下,将400微升新鲜制备的70%乙醇引入到样品中。将样品孵育30秒,同时将试管在磁性机架上温和地旋转。在溶液澄清之后,丢弃上清液,且不扰乱集结粒。执行第二次乙醇洗涤,且丢弃上清液。通过脉冲自旋试管来移除任何剩余乙醇,且小心地移出残余乙醇而不扰乱集结粒。将集结粒在室温下风干约5分钟。

[0373] 一旦试管干燥,就将试管从磁性机架移出,且添加20微升Low TE(加利福尼亚州的生命技术公司,部件号602-1066-01)。对试管进行吹吸和涡旋以确保样品彻底混合。将样品脉冲自旋且置于磁性机架上两分钟。在溶液澄清之后,将含有最终扩增子文库的上清液转移到新Lobind试管(艾本德公司,部件号022431021)中。

[0374] 评估文库大小分布且测定模板稀释系数

[0375] 对最终扩增子文库进行定量以测定文库稀释度(模板稀释系数),其产生用于模板制备的优化目标范围内的浓度(例如将文库分子PCR介导的添加到Ion Sphere™粒子上)。最终扩增子文库通常针对下游模板制备程序使用Ion文库定量试剂盒(qPCR)(生命技术公司,部件号4468802)和/或Bioanalyzer™(安捷伦技术公司(Agilent Technologies),Agilent 2100生物分析仪)定量,以测定扩增子文库的摩尔浓度,由其计算模板稀释系数。举例来说,可在Ion文库定量试剂盒使用者指导(生命技术公司,部件号4468986)中发现通过定量实时PCR(qPCR)测定模板稀释系数的说明书,其以全文引用的方式并入本文中。

[0376] 在此实例中,在2100Bioanalyzer™上用Agilent高敏感DNA试剂盒(安捷伦技术公司,部件号5067-4626)分析1微升最终扩增子文库制备物,以产生大小在135-205bp范围内且浓度为每微升约$5\times10^9$个拷贝的峰值。

[0377] 继续进行到模板制备

[0378] 最终文库的等分试样用于制备DNA模板,所述DNA模板在Ion Sphere™粒子上使用乳液PCR(emPCR)以克隆的方式扩增。本发明实例中的模板的制备根据制造商的说明书使用Ion Xpress模板试剂盒(生命技术公司,部件号4466457)制备,所述说明书以全文引用的方式并入本文中。一旦模板阳性Ion Sphere粒子被富集,就如Ion测序使用者指导(部件号4467391,其全文并在本文中)中所述将Ion Spheres的等分试样装载到Ion 314™芯片(生命技术公司,部件号4462923)上,且如Ion Torrent PGM测序仪使用者指导(生命技术公司,部

件号4462917,其全文并在本文中)中所述经历分析和测序。

[0379] 实例4:鉴别与临床结果相关的基因事件的Oncomine NGS整合分析方法

[0380] Oncomine NGS整合分析经设计以使得最大可能的核心NGS数据集在一起,从而启用询问数据类型和疾病间的关系的科学工作流,在多个抽象生物学层面(如基因和路径)下概述所述分析。

[0381] 数据来源(Oncomine可从生命技术公司/肯潘迪亚生物科学公司(Compendia Biosciences)-密歇根州安娜堡(Ann Arbor,Michigan)和超文本传输协议://www.oncomine.org获得)

[0382] 整合分析的数据从以下来源获取:

[0383] 融合体:Oncomine驱动基因融合体

[0384] 突变:Oncomine泛癌症驱动基因突变

[0385] CNV:由Oncomine处理的拷贝数数据获得的峰值扩增和缺失数据

[0386] DNA:Oncomine处理的DNA序列(DNA-seq)连续数据

[0387] RNA:归一化的基因层面RNA序列(RNAseq)连续数据

[0388] 临床:Oncomine管理的临床和结果元数据

[0389] 路径:Oncomine路径定义

[0390] 融合体数据和过滤

[0391] 用于整合分析的融合体数据从Oncomine NGS融合体数据获得。Oncomine优先级排序的融合体是在肯潘迪亚公司开发的优先级流程,其用于捕获已知真阳性融合体事件的属性且将所观察到的基因融合体的子集表征为高置信度优先级融合体。用于定义优先级融合体的准则包括:有效的5'到3'定向、非邻近融合体搭配物、独特映射跨越读段、非平行同源融合体搭配物、在正常组织中观察不到、以及不与基因组中的冗余区域重叠。

[0392] 如果融合体通过deFuse或Tophat调用,具有“支持的”或“中立的”的外显子表达证据且满足以下4个准则中的一者,那么将所述融合体包括在内且视为驱动基因融合体:

[0393] Oncomine优先级排序的融合体+再现的

[0394] Oncomine优先级排序的融合体+Mitelman注释的

[0395] Oncomine优先级排序的融合体+一种搭配物是涉及3种或更多种泛疾病优先级融合体的Oncomine获得功能的基因

[0396] Oncomine优先级排序的融合体+一种搭配物是涉及3种或更多种泛疾病优先级融合体的桑格致癌基因(http://goo.gl/JQBw9)

[0397] 突变数据和过滤

[0398] 用于整合分析的突变数据从Oncomine NGS突变数据获得。将个别基因归类到所预测的功能类别中,即“获得功能”和“失去功能”,以反映其在潜在活化或有害突变中的相对富集。此分类是基于在通过p值评估的基因中观察到的突变的相对频率和显著性的组合。“获得功能”基因将具有相对较高的热点错义突变频率和较低的有害突变频率,而“失去功能”基因含有较大部分的有害突变。

[0399] 拷贝数区段化和定量

[0400] 每一TCGA样品的DNA拷贝数数据从Oncomine获得。将单一基因的多个报告基因的测量值平均化。

[0401] 最小共有区域(MCR)峰值生成

[0402] 在复现性扩增(4个或更多个拷贝)或缺失(1个或更少的拷贝)的基因中,在25种癌症类型中通过对Oncomine临床样品应用MCR分析来独立地鉴别峰值。为了定义峰值,首先鉴别多个基因显著异常的连续基因组区域(共有区域)。在每一共有区域中,峰值被定义为异常样品计数满足或超出峰值阈值的一种或多种基因。在每种癌症中,共有区域被定义为异常样品计数满足或超出共有区域阈值的区域。计算每种癌症中的每一染色体的每一臂的所有基因中观察到的异常样品的基线平均数。

[0403] mRNA基因表达

[0404] 表达数据从博德GDAC的TCGA标准数据获得。

[0405] 临床数据管理

[0406] 患者临床数据从TCGA获得且由肯潘迪亚公司管理。管理的数据类型包括人口统计资料、主要临床和组织学疾病亚型以及临床结果数据。将所有特性标准化以在各疾病中一致。

[0407] 临床上相关的子集的构筑

[0408] 使用从TCGA获得的经管理的临床数据和Oncomine NGS数据,且应用表14中的规则以定义临床子集:

[0409] 表14-定义临床子集的规则

[0410]

| 疾病 | 临床亚型 | 来源 | 规则 |
|---|---|---|---|
| 侵袭性乳癌 | 三阴性 | Phenomic 数据 | ERBB2 状态 = ERBB2 阴性<br>雌激素受体状态 = 雌激素受体阴性<br>孕酮受体状态 = 孕酮受体阴性 |
| | ER 阳性 | Phenomic 数据 | 雌激素受体状态 = 雌激素受体阳性 |
| | ER 阳性和 HER2 阴性 | Phenomic 数据 | 雌激素受体状态 = 雌激素受体阳性<br>ERBB2 状态 = ERBB2 阴性 |
| 胃腺癌 | 高突变基因 | Oncomine NGS 数据 | 患者突变计数 >=400 |
| 肺腺癌 | KRAS 突变(无 ALK 融合且无 EGFR 突变) | Oncomine NGS 数据 | Oncomine 突变分类 = 热点 |
| | 三阴性 | Oncomine NGS 数据 | 无 EGFR 突变(且)无 KRAS 突变(且)无 ALK 融合 |
| 直肠腺癌 | KRAS 突变 | Oncomine NGS 数据 | Oncomine 突变分类 = 热点 |
| | KRAS 突变,3 期或 4 期 | Oncomine NGS 数据/Phenomic 数据 | Oncomine 突变分类 = 热点(且)<br>分期 = III 期(或)IV 期 |
| | KRAS 野生型 | Oncomine NGS 数据 | 无 KRAS 突变 |

[0411] 路径

[0412] 人工管理的肯潘迪亚公司路径定义用于概述整合分析中的基因层面畸变。所述路径代表临床上相关的路径模块,且数个模块可覆盖主要生物学路径,且单一基因可存在于一个或多个路径模块定义中。

[0413] 数据整合

[0414] 图3中的图式概述将各种数据类型整合到基因事件数据库(Genetic Event Database,GEDB)中的数据流。所有进一步分析均使用GEDB进行。所述方法具有4个主要步骤。

[0415] 将数据映射到内部IA基因和患者维度

[0416] 定义每一数据类型中的事件和驱动基因事件

[0417] 将个别事件上卷到基因和路径层面

[0418] 将事件合并到基因事件数据库中。

[0419] 基因和患者维度

[0420] 构筑涵盖在所有疾病和数据类型中测量的所有患者和基因的单一基因和患者维度。给予基因和患者内部识别符，且IA中的所有数据相对于这些识别符参考以实现基因名称和患者条形码一致性。一种基因的唯一识别符是所述基因的Entrez ID。一个患者的唯一识别符是TCGA患者条形码(TCGA条形码的前12个数位)。

[0421] 驱动基因事件定义

[0422] 突变、融合体和拷贝数事件基于基因组学事件的以下准则定义：

[0423] 融合体:Oncomine再现的优先级融合体

[0424] 突变:来自泛癌症驱动基因的Oncomine驱动基因突变

[0425] CNV:CBI鉴别的峰值和峰值内的基因扩增/缺失

[0426] 基因事件定义和上卷

[0427] 基因事件是基因组学畸变，代表个别突变、融合体或拷贝数事件、或在基因或路径层面下的事件的组合。所述事件根据图4中展示的流程图‘上卷’。当组合多个事件以构筑上卷事件时，所述上卷事件的所测量的患者组变为针对所有3种数据类型测量的患者的相交点。患者阳性仅在被完全测量时才包括在内。

[0428] 分析

[0429] 一旦构筑所有驱动基因事件，就对每一基因事件执行一组分析，计算疾病(且适当时为泛癌症)内的频率、相关性和关系。以下为每一分析的简短描述：

[0430] 频率

[0431] 频率是驱动基因事件在被测量所述驱动基因事件的患者当中的出现率。计算疾病和泛癌症内的频率。

[0432] 临床相关性分析

[0433] 测试每一驱动基因事件相对于一组可获得的临床亚型的相关性。通过比较一种临床亚型对比另一种临床亚型的患者中的基因事件的出现率，使用费舍尔精确检验对每一相关性进行测试。举例来说，可在吸烟者对比非吸烟者中、或在I期对比II期肺癌中测试功能缺失突变的过度表达。针对每一事件测试总共136个亚型对，下文列出定义所述亚型的特性(一些特性可为疾病特有的)。需要总共至少4个患者(且每种类别中至少1个患者)来执行测试。

[0434] 临床结果分析

[0435] 使用对数秩检验对每一驱动基因事件与临床结果的相关性进行测试。仅具有可用的临床数据的患者组才用于计算，故包括于测试中的患者数量可能少于针对驱动基因事件测量的患者数量。需要至少4个对驱动基因事件阳性的患者来执行所述测试。存活时间以年呈现，且个别存活/死亡事件清楚地标记在卡普兰-迈耶曲线(Kaplan-Meier curve)上。校正多个测试的p值(q值)。考虑q值小于0.1的事件。

[0436] 分析的结果展示在表15和39中。在表15和39中，各栏提供以下信息：

[0437] “子集”栏提供临床上相关的癌症类型。

[0438] p值栏是p值。

[0439] q值栏是经校正的p值。q<0.1的事件包括于表中。

[0440] 阳性数量栏是对一种事件类型阳性的患者数量。

[0441] 患者总数量栏是所评估的患者总数。

[0442] Cytoband栏是基因的染色体位置。

[0443] 基因(Entrez ID)栏是基因和对应Entrez id的列表。

[0444] 可药化基因栏指示任何基因是否为处于活性试验、已批准或以其他方式市售的药物的目标。

[0445] KM证据栏提供卡普兰-迈耶证据。KM证据指示所述事件类型是否支持特定癌症类型中的良好或较差的预后。

[0446] 表15和39含有q<0.1的各种癌症类型的超过100个获得功能的突变、失去功能的突变、峰内基因扩增/缺失和融合体事件。每一事件和癌症类型内的基因与其染色体位置、可药化性信息和临床结果相关性一起包括在内,如上述栏信息中所指示。

[0447] 实例5-整合的数据分析

[0448] Oncomine NGS整合的分析.Oncomine NGS整合分析经设计以使得最大可能的核心整合的基因组学和phenomic数据集在一起,从而启用询问数据类型和疾病间的关系的科学工作流,在多个抽象生物学层面(如基因和路径)下概述所述分析。

[0449] 术语:

[0450] 畸变-DNA的基因组学结构变异或变化;实例包括:mRNA过度表达/表达不足、拷贝数扩增/缺失、突变和基因融合体。

[0451] 驱动基因-通过本文档中描述的Oncomine方法鉴别为潜在癌症驱动基因的畸变;实例包括获得功能的突变、峰值扩增区域中的基因扩增、或基因融合体。

[0452] 上卷-基因或路径的所有突变、融合体或拷贝数畸变的的概述;仅针对所有三种畸变类型测量的患者包括于所述上卷中。

[0453] 热点突变-再现的(n≥3)且归类为框内插入/缺失、无终止或错义的突变。

[0454] 患者空组-针对一种基因畸变测量的患者组

[0455] 患者阳性组-含有基因畸变的患者组

[0456] 基因空组-通过用于评估基因畸变的实验平台测量的基因组

[0457] Mitelman-从文献中人工管理的癌症中的染色体畸变和基因融合体的数据库(超文本传输协议://goo.gl/PnXMT)

[0458] RPKM-“每百万每千碱基读段”;用于将测序读段的总读段长度和数量归一化的RNASeq数据定量的方法(莫塔扎维(Mortazavi)等人2008)

[0459] RSEM-“通过期望最大化的RNA-Seq”用于估计预期的转录物之间的读段的最佳可能的分布的RNASeq数据定量的方法提供相对转录物丰度作为总读段池的一部分(李(Li)和杜威(Dewey)2011)

[0460] 数据来源.努力收集可用于每一样品的数据的最大重叠组。NGS整合分析浏览器的此版本中的数据从癌症基因组图谱(TCGA)、癌细胞系百科全书(Cancer Cell Line Encyclopedia,CCLE)、COSMIC细胞系项目和许多研究出版物直接或在经历Oncomine处理和分析方法之后获得。由于来源数据集中所有数据类型的不均匀覆盖率,一些癌症类型具有多种数据类型中覆盖的更多的患者数量。

[0461] 用于整合分析的Oncomine NGS突变版本含有从同行审查的出版物中的NGS突变研究获得的许多人工管理的数据集。关于为整合分析贡献突变数据的出版物的完整列表，请参见Oncomine NGS突变方法资料。以下数据集含有多维度NGS数据，提供突变和拷贝数数据两者。这些数据集的拷贝数数据以与从TCGA获得的拷贝数数据相同的方式处理。

[0462] 细胞系数据包括突变、融合体和拷贝数数据集。细胞系数据以与临床肿瘤数据相同的方式处理，且突变和融合体细胞系数据分别从Oncomine™ NGS突变和Oncomine™ NGS融合体动力工具获得。细胞系的拷贝数数据使用标准Oncomine拷贝数流水线处理。尽管使用两种迥然不同的细胞系数据集CCLE和COSMIC，但我们对细胞系疾病类型和名称的标准化已使得我们能够交叉参考两个数据集且合并CCLE拷贝数数据、COSMIC突变数据和Oncomine融合体调用(基于CCLE RNASeq数据)。因此，此版本中的许多细胞系的外显子组已针对所有三种类型的畸变进行系统地表征。细胞系数据使用将与肿瘤数据直接可比的Oncomine癌症类型定义概述，不过分别针对肿瘤和细胞系执行概述。

[0463] Phenomic数据

[0464] 临床患者元数据管理.患者临床数据从原始来源获得且由肯潘迪亚公司管理。管理的数据类型包括人口统计资料、主要临床和组织学疾病亚型以及临床结果数据。将所有癌症类型无关的特性(如年龄或存活率)标准化以实现癌症类型中的一致性。合并某些疾病分期以获得一个分期内的较高患者计数。举例来说，Ia期和Ib期可组合为经修改的I期。

[0465] 以下是通过管理方法捕获的人口最稠密的特性和对应值的列表。不是所有特性都是可用于所有患者的。

[0466]

| 特性名称 | 特性值 |
|---|---|
| 年龄 | 10 - 14 岁 |
| | 15 - 19 岁 |
| | 20 - 29 岁 |
| | 30 - 39 岁 |
| | 30 -39 岁 |
| | 40 - 49 岁 |
| | 50 - 59 岁 |
| | 60 - 69 岁 |
| | 70 - 79 岁 |
| | 80 - 89 岁 |
| | 90+岁 |
| ERBB2 状态 | ERBB2 阴性 |
| | ERBB2 阳性 |
| 雌激素受体状态 | 雌激素受体阴性 |
| | 雌激素受体阳性 |
| FAB 亚型 | FAB 亚型 M0 |
| | FAB 亚型 M1 |
| | FAB 亚型 M2 |
| | FAB 亚型 M3 |
| | FAB 亚型 M4 |
| | FAB 亚型 M5 |
| | FAB 亚型 M6 |
| | FAB 亚型 M7 |
| 格里森分数（Gleason Score） | 格里森分数 10 |
| | 格里森分数 6 |
| | 格里森分数 7 |
| | 格里森分数 8 |
| | 格里森分数 9 |

[0467]

| 特性名称 | 特性值 |
|---|---|
| 级别 | 1 级 |
| | 2 级 |
| | 3 级 |
| | 3-4 级 |
| | 4 级 |
| 肝炎病毒感染状态 | 乙型肝炎病毒阳性 |
| | 丙型肝炎病毒阳性 |
| 人乳头瘤病毒感染状态 | HPV 阴性 |
| | HPV 阳性 |
| | HPV 16 和 52 型阳性 |
| | HPV 16 型阳性 |
| | HPV 45 型阳性 |
| | HPV 58 型阳性 |
| 转移性事件状态 | 转移性事件 |
| 微卫星状态 | 微卫星不稳定 |
| | 微卫星稳定 |
| 整体存活状态 | 存活 |
| | 死亡 |
| 整体存活状态（详述） | 存活 |
| | 带病存活 |
| | 不带病存活 |
| | 死亡 |
| | 带病死亡 |
| | 不带病死亡 |
| 患者治疗响应 | 未知疗法完全响应 |
| | 未知疗法部分响应 |
| | 未知疗法进展性疾病 |
| | 未知疗法稳定疾病 |
| 孕酮受体状态 | 孕酮受体阴性 |
| | 孕酮受体阳性 |
| 人种/种族 | 美洲印第安人或阿拉斯加原住民 |
| | 亚洲人 |
| | 黑人或非裔美国人 |
| | 西班牙裔或拉丁裔 |
| | 夏威夷原住民或其他太平洋岛民 |
| | 白人 |
| 复发状态 | 生物化学复发 |
| | 无生物化学复发 |
| | 复发 |
| 复发状态（详述） | 局部复发 |
| | 转移性复发 |
| | 复发 |
| 经修改的 M 期 | M0 |
| | M1 |
| 经修改的 N 期 | N0 |
| | N1 |
| | N2 |
| | N3 |

[0468]

| 特性名称 | 特性值 |
|---|---|
| 经修改的吸烟状态 | 从不吸烟者 |
| | 吸烟者 |
| 经修改的分期 | FIGO I 期 |
| | FIGO II 期 |
| | FIGO III 期 |
| | FIGO IV 期 |
| | I 期 |
| | II 期 |
| | III 期 |
| | IV 期 |
| 经修改的 T 期 | T |
| | T0 |
| | T1 |
| | T11 |
| | T12 |
| | T2 |
| | T21 |
| | T22 |
| | T3 |
| | T4 |
| 性别 | 女性 |
| | 男性 |
| *TCGA PAM50 亚型 | 基底样 |
| | HER2 富集 |
| | 管状 A 型 |
| | 管状 B 型 |
| | 正常样 |
| *TCGA RPPA 亚型 | 基底 |
| | Her2 |
| | 管状 A 型 |
| | 管状 A/B 型 |
| | 反应性 I 组 |
| | 反应性 II 组 |
| *TCGA 亚型 | 基底 |
| | CIN |
| | 经典 |
| | 侵袭性 |
| | 间充质型 |
| | MSI/CIMP |
| | 神经型 |
| | 初期 |
| | 原神经型 |
| | 分泌型 |
| 转移性事件随访时间 | (天) |
| 整体存活率随访时间 | (天) |
| 复发随访时间 | (天) |

[0469] 前缀为“TCGA”的特性是由定义侵袭性乳癌、成胶质细胞瘤、鳞状细胞肺癌和结肠直肠癌的分子亚型的TCGA出版物获得和管理的。

[0470] 基因组学事件数据:融合体数据过滤.整合分析浏览器的融合体数据从2013年11月发行的Oncomine NGS融合体数据获得。仅将鉴别为Oncomine优先级融合体的融合体包括

于整合分析浏览器中。

[0471] Oncomine优先级排序的融合体是在肯潘迪亚公司开发的优先级流程,其用于捕获已知真阳性融合体事件的属性且将所观察到的基因融合体的子集表征为高置信度优先级融合体。用于定义优先级融合体的准则包括:有效的5'到3'定向、非邻近融合体搭配物、独特映射跨越读段、非平行同源融合体搭配物、在正常组织中观察不到、以及不与基因组中的冗余区域重叠。

[0472] 融合体数据的患者空组是在融合体分析中处理的完整的患者肿瘤样品组;每个患者仅保留一个肿瘤样品(优选地,原发性、非复发性肿瘤)的数据。基因空组是到2012年5月为止的RefGene中的基因组。如果融合体是Oncomine优先级融合体,具有"支持的"或"中立的"的外显子表达证据且满足以下准则中的一者,那么将所述融合体包括在整合分析浏览器中:

[0473] 再现(在2个或更多个患者中出现)

[0474] 在已知结构变异的Mitelman数据库中注释

[0475] 含有作为涉及3种或更多种泛疾病优先级融合体的Oncomine获得功能基因的基因搭配物

[0476] 含有作为涉及3种或更多种泛疾病优先级融合体的桑格致癌基因(超文本传输协议://goo.gl/JQBw9)的基因搭配物。

[0477] 突变数据过滤.整合分析的突变数据从2013年11月发行的Oncomine NGS突变数据获得。仅将编码基因区域中的非沉默突变包括于整合分析浏览器中。

[0478] 患者空组是突变分析中处理的完整患者组;每个患者仅保留一个肿瘤样品(优选地,原发性、非复发性肿瘤)的数据。基因空组是到2012年3月为止的RefGene中的基因组。

[0479] 具有以下变异体分类的突变不包括于整合分析浏览器中:沉默、5'UTR、3'UTR、RNA、非编码外显子。

[0480] 调用扩增/缺失.每一样品的DNA拷贝数数据从2013Q4Oncomine标准数据构建获得,其中将可从TCGA和到2013年10月为止的人工管理的出版物获得的所有拷贝数数据标准化。

[0481] 此分析的患者空组为到2013年10月为止针对拷贝数数据测量的患者组以及在人工管理的出版物中测量的患者组。每个患者仅保留一个肿瘤样品(优选地,原发性、非复发性肿瘤)的数据。此数据的基因空组为Oncomine DNA拷贝数平台,基于由UCSC RefGene构建2009年7月提供的RefSeq座标(hg18),且测量18,796个基因。将单一基因的多个报告基因的测量值平均化。

[0482] 估计的拷贝值的$\log_2$用于进行扩增/缺失(amp/del)调用,且截断值分别为>1.0和<-1.0。没有amp/del调用作出≥-1.0或≤+1.0的$\log_2$(估计的拷贝)。

[0483] 基因组学连续数据:拷贝数区段化和定量.每一样品的DNA拷贝数数据从2013 Q4 Oncomine标准数据构建获得,其中将到2013年9月为止可从TCGA获得的所有拷贝数数据以及来自人工管理的出版物的所有拷贝数数据标准化。

[0484] 此分析的患者空组为到2013年10月为止针对拷贝数数据测量的TCGA患者组以及在人工管理的出版物中测量的患者组。每个患者仅保留一个肿瘤样品(优选地,原发性、非复发性肿瘤)的数据。此数据的基因空组为Oncomine DNA拷贝数平台,基于由UCSC RefGene

构建2009年7月提供的RefSeq座标(hg18),且测量18,796个基因。将单一基因的多个报告基因的测量值平均化。

[0485] 拷贝数数据使用标准Oncomine处理流水线区段化且定量。区段化是一种用于鉴别扩增或缺失的连续区域的方法。这些区域或“区段”可以包括多个基因或单个基因。基于含于区段中的报告基因的平均值,计算每个区段的拷贝数值。基因被映射到区段且分配一个值。接着报导此基因层面数据。更多信息请参见Oncomine DNA处理流水线白皮书(Oncomine DNA Processing Pipeline White Paper)。

[0486] mRNA表达数据.表达数据从来自2013年9月的博德GDAC的TCGA标准数据构建获得。

[0487] 此数据的患者空组是在博德GDAC 2013_08_09stddata build中具有可用的RNASeq数据的患者组;每患者仅保留一个肿瘤样品(优选地,原发性、非复发性肿瘤)的数据。此数据的基因空组根据疾病不同且对应于用于RNASeq定量的TCGA基因注释文件(Gene Annotation File,GAF)。

[0488] TCGA目前采用两种RNASeq定量方法V1(RPKM)和V2(RSEM),其在数值上不是直接可比的。为了避免可能不精确的数值转换,我们以每种疾病为基础使用来自单一定量方法的数据,基于最大覆盖率选择格式。符合TCGA使用RSEM(V2)处理(和再处理)所有可用的RNASeq数据的努力,RSEM(V2)数据可用于大部分样品。例外是胃腺癌,其中使用RPKM(V1)数据。获得RSEM和RPKM两者的归一化的基因层面定量值,且将其转换成$\log_2$值(将最小非零RPKM或RSEM值设定在-12)。如果基因的$\log_2$值>-12,那么所述基因被认为是表达的。

[0489] Oncomine驱动基因参考数据:最小共有区域(MCR)峰值生成和基因选择.为了鉴别经历扩增和缺失的癌症驱动基因,执行峰值簇聚方法以选择在多个癌症类型中经常异常的基因。定义许多癌症类型内的最大可用的拷贝数数据集(即,超出其的数据包括于整合分析中)中的第一拷贝数峰值。接着,将由峰值定义的基因列表聚集,以便鉴别在多个样品和多种疾病中的拷贝数峰值中出现的基因。下文更详细地描述部分方法。

[0490] 畸变可归类为“驱动基因”畸变,或根据数据类型特有的Oncomine分类方法中的一者考虑为可能令人感兴趣的畸变。驱动基因畸变将作为与其他畸变(非驱动基因畸变被称为“任何”)无关的事件捕获。举例来说,具有“驱动基因”突变的患者将是两种畸变(“驱动基因”突变和“任何”突变)阳性的。所测量的数据类型中的每一者具有一组确定驱动基因事件的规则。

[0491] 经历扩增或缺失的一组连续基因组区域使用Oncomine MCR分析通过将其应用于分组到25种癌症中的Oncomine的10,249个临床样品来鉴别。

[0492] 峰值定义的患者空组为来自Oncomine的10,249个临床样品(参见下表)。此数据的基因空组为Oncomine DNA拷贝数平台,基于由UCSC refGene构建2009年7月提供的RefSeq座标(hg18),且测量18,796个基因。

[0493] 最小共有区域(MCR)分析的数据来源于跨越不同癌症类型的含有>20,000个临床样本、异种移植物和细胞系的Oncomine DNA拷贝数浏览器。MCR分析通过分析以下三个层面下的数据来鉴别再现的拷贝数扩增或缺失的区域:泛癌症(在所有癌症类型中)、普通癌症类型(在多个癌症类型中)和中期癌症类型或特定癌症亚型。简单来说,所述方法首先计算被定义为在2个或更多个样品中扩增或缺失的连续基因组区域的共有区域(CR)。将扩增和缺失的最小阈值分别设定为log2≥0.9(3.7个拷贝或更多个)和log2≤-0.9(1个拷贝或更

少)。接着,将这些共有区域内的峰值区域定义为(i)在最高数量样品(n)中异常的一种或多种基因以及在比最高数量小一的样品(n-1)中异常的一种或多种基因以及(ii)在90%的最高异常样品计数中异常的基因。

[0494] 鉴别共有峰值区域的聚类分析.执行MCR分析。使用下表中列出的准则在三种分析类型(即,泛癌症、普通癌症类型和特定癌症类型分析)中进一步过滤通过MCR分析鉴别的峰值区域。应注意,仅包括所选数量(约40)的中期或特定癌症类型(下文也进一步列出)。

[0495] 鉴别来自MCR分析的高度扩增/缺失区域的过滤准则:

[0496]

| | 泛癌症 | 普通癌症类型 | 中期或特定癌症类型 |
|---|---|---|---|
| **异常样品计数** | ≥4 | ≥4 | ≥4 |
| **最大 log2 拷贝数-扩增** | ≥2(8 个或更多个拷贝) | ≥2(8 个或更多个拷贝) | ≥2(8 个或更多个拷贝) |
| **最大 log2 拷贝数-缺失** | N/A | ≤-1(1 个或更少的拷贝) | ≤-1(1 个或更少的拷贝) |
| 频率中位数 | ≥0.5% | ≥0.5% | ≥1.0% |
| 中期或特定癌症类型 | 包括所有 | 包括所有 | 所选 ICT(参见表 6) |

[0497] 包括于上文描述的过滤准则中的所选择的中期或特定癌症类型:

[0498]

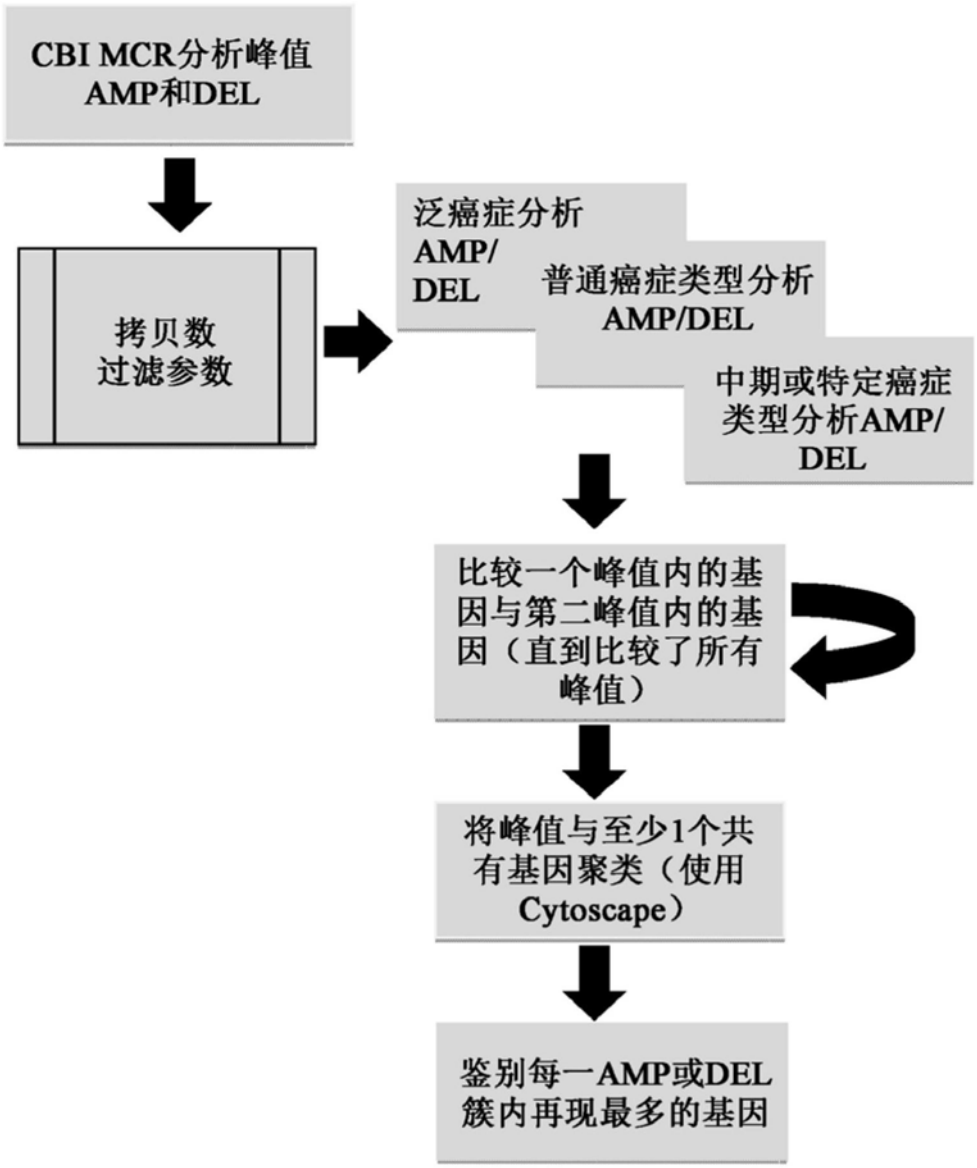
| 常规癌症类型 | 中期或特定癌症类型 |
|---|---|
| 膀胱癌 | 膀胱尿道上皮癌 |
| 脑和 CNS 癌 | 成胶质细胞瘤;成神经管细胞瘤;成神经细胞瘤 |
| 乳癌 | N/A |
| 宫颈癌 | 宫颈腺癌;宫颈鳞状细胞癌 |
| 结肠直肠癌 | 结肠直肠腺癌 |
| 食道癌 | 食道腺癌;食道鳞状细胞癌 |
| 胃癌 | 胃腺癌 |
| 头颈癌 | 头颈鳞状细胞癌;甲状腺腺体滤泡癌;甲状腺乳头状癌 |
| 肾癌 | 透明细胞肾细胞癌;乳头状肾细胞癌 |
| 白血病 | 急性淋巴母细胞性白血病;急性骨髓性白血病;慢性淋巴细胞性白血病;慢性骨髓性白血病;骨髓发育不良综合症 |
| 肝癌 | 肝细胞癌 |

[0499]

| | |
|---|---|
| 肺癌 | 肺腺癌;小细胞肺癌;鳞状细胞肺癌 |
| 淋巴瘤 | 伯基特氏淋巴瘤(Burkitt's Lymphoma);DLBCL;滤泡性淋巴瘤;霍奇金氏淋巴瘤(Hodgkin's Lymphoma);套细胞淋巴瘤; |
| 黑素瘤 | 皮肤黑素瘤;多发性骨髓瘤 |
| 其他 | 子宫内膜子宫内膜样腺癌 |
| 卵巢癌 | 卵巢透明细胞腺癌;卵巢浆液性腺癌 |
| 胰脏癌 | 胰腺导管腺癌;前列腺腺癌 |
| 肉瘤 | GIST |

[0500] 接着,为了鉴别跨越多种癌症类型的再现最多的峰值区域和基因,我们使用Cytoscape 2.8.3[马克尔(Markiel)等人2003;斯姆特(Smoot)等人2001]以构建网络簇。简单来说,所述分析比较给定峰值区域中的每一基因与其他峰值区域中的基因,且使峰值与至少一个共有基因聚类。接着,将每一簇内再现最多的扩增或缺失的基因视为潜在的候选驱动基因。所述方法概述在下文示意图中:

[0501]




[0502] 临床上相关的癌症亚型的鉴别和形成.为了提供患者子集以用于更集中的分析，使用临床表型和类别基因组数据的组合鉴别和管理数种临床上相关的癌症亚型。表型数据来源于TCGA网络门户(TCGA Web Portal)或人工管理的出版物的补充方法。

[0503] 应用以下规则以定义临床子集：

[0504]

| 癌症类型 | 临床亚型 | 数据来源 | 用于包括在亚型中的数据解释规则 |
|---|---|---|---|
| 侵袭性乳癌 | 三阴性 | TCGA 网络门户 | ERBB2 状态 = “ERBB2 阴性”<br>雌激素受体状态 = “雌激素受体阴性”<br>孕酮受体状态 = “孕酮受体阴性” |
| | ER 阳性 | Phenomic | 雌激素受体状态 = “雌激素受体阳性” |
| | HER2 阳性 | Phenomic | ERBB2 状态 = “ERBB2 阳性” |
| | ER 阳性和 HER2 阳性 | Phenomic | 雌激素受体状态 = “雌激素受体阳性”<br>ERBB2 状态 = “ERBB2 阳性” |
| | ER 阳性和 HER2 阴性 | Phenomic | 雌激素受体状态 = “雌激素受体阳性”<br>ERBB2 状态 = “ERBB2 阴性” |
| 胃腺癌 | 弥漫型 | Phenomic | 癌症类型 = “弥漫型胃腺癌” |
| | 肠型 | Phenomic | 癌症类型 = “胃肠型腺癌” |
| | 高突变基因 | Oncomine NGS | 患者突变计数 ≥400 |
| 头颈鳞状细胞癌 | HPV 阳性 | Phenomic | 人乳头瘤病毒感染状态 = “HPV 阳性” |
| | HPV 阴性 | Phenomic | 人乳头瘤病毒感染状态 = “HPV 阴性” |
| 宫颈鳞状细胞癌 | HPV 阳性 | Phenomic | 人乳头瘤病毒感染状态 = “HPV 阳性” |
| 肺腺癌 | EGFR 突变 | Oncomine NGS 突变 | Oncomine 突变分类 = “热点” |
| | KRAS 突变（无 ALK 融合且无 EGFR 突变） | Oncomine NGS 突变和融合体 | Oncomine 突变分类 = “热点” |
| | ALK 融合 | Oncomine NGS 融合体 | 具有 Oncomine 驱动基因 ALK 融合体 |
| | 三阴性 | Oncomine NGS 突变和融合体 | 无 EGFR 突变<br>且<br>无 KRAS 突变<br>且<br>无 ALK 融合 |
| 结肠和直肠腺癌 | KRAS 突变 | Oncomine NGS 突变 | Oncomine 突变分类 = “热点” |
| | KRAS 突变，3 期或 4 期 | Oncomine NGS 突变和 Phenomic | Oncomine 突变分类 = “热点”<br>且<br>分期 = “III 期”或“IV 期” |
| | KRAS 野生型 | Oncomine NGS 突变和融合体 | Oncomine 突变分类 = “热点”<br>且<br>无 KRAS 融合 |
| | 微卫星稳定 | Phenomic | 微卫星状态 = “微卫星稳定” |
| | 微卫星不稳定 | Phenomic | 微卫星状态 = “微卫星不稳定” |
| 子宫内膜子宫内膜样癌 | 微卫星稳定 | Phenomic | 微卫星状态 = “微卫星稳定” |
| | 微卫星不稳定 | Phenomic | 微卫星状态 = “微卫星不稳定” |

[0505] 路径.人工管理的肯潘迪亚公司路径定义用于概述整合分析浏览器中的基因层面畸变。所述路径代表临床上相关的路径模块，且数种模块可覆盖主要生物学路径。单一基因可存在于一个或多个路径定义中，但需小心地消除基本上冗余的路径，所述路径中一个模块是另一个模块的完全子集。存在总共67个路径，大小在42个基因(例如MAPK路径)到2个基因(例如IGF1/IGF1R和数个其他路径)范围内。

[0506] 数据整合.此部分概述了数据流，所述数据流将原始数据整合到共有的患者和基因维度上且构筑包含将经受整合分析的所有畸变的基因事件数据库(GEDB)。所述方法具有4个主要步骤：(1)使用通用基因和患者维度整合原始数据；(2)调用每一数据类型的畸变事件且定义驱动基因畸变；(3)将个别事件上卷到基因和路径层面且整合事件；以及(4)通过

定义每一事件的患者状态来构筑基因事件数据库。

[0507] 构筑且映射到基因和患者维度.可能已在不同实验平台上且对不完全重叠的患者组测量整合分析中所包括的变化的数据类型。因此,小心地包括所测量的所有患者和基因,同时避免一式两份或冲突的条目。

[0508] 对于每一数据类型,构筑基因和患者“维度”,枚举数据中测量的基因和患者。每一数据类型的维度可能不同,如通过以下患者维度重叠图式所指示(数量仅用于说明),在这种情况下,对于侵袭性乳癌而言。

[0509] 从每一癌症和数据类型收集基因和患者维度,且在整合分析中构筑所有患者和基因识别符的非冗余超集。下文图中的条柱代表针对某些畸变类型测量的患者的患者识别符(如果是通过所述识别符分选的的)的数据块。

[0510] 患者的冗余基于独特患者识别符(目前为TCGA肿瘤样品条形码的前12个字符(例如TCGA-AB-1234))确定。

[0511] 基因的冗余基于独特基因识别符(目前为Entrez Gene ID)确定。基因还相对于Oncomine基因组比较,且当发现基因符号冲突(一个Entrez ID分配两个或更多个基因符号)时,使用来自Oncomine的基因符号。数种(12种)所测量基因含有已停止的Entrez ID,且因此可能不代表最准确的基因模型。这些基因的基因符号标记有词语“停止”。

[0512] 一旦构筑,就将非冗余基因和患者维度编索引以在数据集中为每一基因和患者提供一致的内部识别符。接着,将原始数据中的所有独特患者和基因识别符映射到维度患者和基因识别符。因此,基因和患者元数据(如基因符号和患者临床数据)总是经由对应的维度映射,提供命名和注释方面的一致性。整合分析中的独特基因和患者的总数如下:

[0513]

| | |
|---|---|
| **基因** | 23,340 |
| **患者** | 11,476 |

[0514] 患者维度与患者的数据集特有的映射一起有助于正确鉴别完全野生型患者(测量了所有畸变类型但并不含有任何畸变的患者)。

[0515] 患者可因此测量任何数量的畸变,且可仅对于所测量的那些事件是异常的。为了确定分析类型测量患者畸变,所述患者将包括于以下各者中:

[0516]

| | **针对以下测量的患者“X”:** | **从以下排除的患者“X”:** | **针对以下测量的患者“Y”:** | **从以下排除的患者“Y”:** |
|---|---|---|---|---|
| **临床*** | | 临床相关性,临床结果相关性 | · | DNA 对比 RNA 相关性,差异表达,表达的频率 |
| **突变** | · | | · | |
| **融合体** | · | | · | |
| **拷贝数** | · | | · | |
| **表达** | · | | | |

[0517] *“针对临床测量”指示为患者呈现临床元数据。

[0518] 事件模型.将来自突变、融合体和拷贝数数据集的每一基因组学畸变鉴别为畸变事件(用于定义将经历各种预定义整合分析的相关事件的术语)。每一畸变是三个宽广层面事件(数据类型特有的事件、基因特有的但数据类型无关的事件以及路径特有的但基因或数据类型无关的事件)的一部分。后两个层面被视为“上卷”事件。

[0519] 下文描述畸变事件定义的特有规则以及“层面”和“驱动基因”流程。

[0520] 突变事件调用程序.Oncomine泛癌症突变分类:如果突变是:再现的(在3个或更多个样品中在相同的氨基酸位置处出现)且经以下变异体分类中的一者注释:框内插入/缺

失、无终止错义，那么将所述归类为“热点”。如果突变是：不再现的且经以下变异体分类中的一者注释：框移插入/缺失、无义，那么所述突变被归类为“有害”。再现性在所分析的所有癌症类型作为Oncomine NGS突变浏览器的一部分测量。

[0521] Oncomine泛癌症基因分类.作为Oncomine NGS突变浏览器流水线的一部分，将个别基因归类到所预测的功能类别中，即“获得功能”和“失去功能”，以反映其在潜在活化或有害突变中的相对富集。此分类是基于在通过p值评估的基因中观察到的突变的相对频率和显著性的组合。“获得功能”基因将具有相对较高频率的热点(在3个或更多个样品中再现的)错义突变和较低频率的有害突变，而“失去功能”基因含有较大部分的有害突变。泛癌症基因分类是基于在所有癌症类型中观察到的突变。

[0522] 突变畸变事件.对于每一患者基因突变(如由突变数据过滤部分所定义)，将形成一个或两个突变事件，取决于所述突变是否归类为驱动基因畸变。驱动基因突变畸变被定义为存在于“获得功能”基因中的“热点”突变、或存在于“失去功能”基因中的“热点”或“有害”突变。对于驱动基因突变，形成驱动基因事件和任何事件两者。对于非驱动基因突变，仅形成一个任何事件。将泛癌症突变和基因分类用于所有分析子集；因此，基因分类可能有时在整合分析与Oncomine NGS突变浏览器之间不同。

[0523] 下表给出可针对每一基因形成的突变事件的描述和实例：

[0524]

| 事件类型 | 事件的描述 | 示例畸变名称 | 驱动基因状态 |
|---|---|---|---|
| **获得功能的突变** | “热点”突变和“获得功能”基因分类 | EGFR 获得功能的突变 | 驱动基因 |
| **失去功能的突变** | “失去功能”基因中的“热点”或“有害”突变 | APC 失去功能的突变 | 驱动基因 |
| **任何基因突变** | 基因中的任何突变 | TTN <任何基因突变> | 任何 |

[0525] 融合体事件调用程序.仅Oncomine优先级融合体包括于整合分析中。在优先级融合体中，驱动基因融合体被定义为通过Mitelman数据库标记已知致癌基因的融合体、或不具有充分外显子表达数据且再现的融合体、或具有外显子表达数据和两种基因搭配物的外显子表达不平衡的显著p值的融合体(关于外显子不平衡分类的细节参见Oncomine NGS融合体方法资料)。对于每一基因，将针对基因搭配物的每一独特观察到的5’-3’组合形成一个事件。举例来说，对于PML-RARA平衡易位，观察到两种同工型，且因此将分别针对PML-RARA和RARA-PML调用两个融合体事件。

[0526]

| 事件类型 | 事件的描述 | 示例畸变名称 | 驱动基因状态 |
|---|---|---|---|
| **融合体** | 涉及基因的驱动基因融合体 | PML-RARA 融合体 | 驱动基因 |
| **任何融合体** | 涉及基因的任何融合体 | FRS2-LYZ 融合体 | 任何 |

[0527] 拷贝数事件调用程序.所调用的每一扩增/缺失(参见调用扩增/缺失)被定义为异常基因的任何事件。如果扩增/缺失出现在某一癌症类型中的作为峰值定义(参见MCR峰值生成)的一部分的基因中，那么还形成所述基因的驱动基因扩增缺失事件。因此，拷贝数事件的驱动基因定义是癌症类型特有的。

[0528] 以下为可针对基因扩增或缺失调用的拷贝数畸变事件：

[0529]

| 事件类型 | 事件的描述 | 示例畸变名称 | 驱动基因状态 |
|---|---|---|---|
| **峰内基因扩增** | 在相同癌症类型内的扩增峰值中观察到的基因中的扩增 | EGFR 峰内基因扩增 | 驱动基因 |
| **峰内基因缺失** | 在相同癌症类型内的缺失峰值中观察到的基因中的缺失 | CDKN1A 峰内基因缺失 | 驱动基因 |
| **任何基因扩增** | 基因中的扩增 | ERBB2 <任何基因扩增> | 任何 |
| **任何基因缺失** | 基因中的缺失 | FGFR <任何基因缺失> | 任何 |

[0530] 基因事件上卷.驱动基因和任何事件均“上卷”到基因层面和路径层面事件以捕获数据类型无关的畸变统计和相关性。举例来说,可能令人感兴趣的是看到肿瘤抑制基因中的任何畸变与临床结果的相关性,而不仅仅是有害突变的相关性。

[0531] 针对具有任何数据类型的至少一个畸变的每一基因形成基因层面畸变事件。针对其中组分基因中的至少一者具有任何数据类型的畸变的每一路径形成路径层面畸变事件。驱动基因和任何畸变独立地上卷到基因层面或路径层面驱动基因或任何事件中。以下图式展示各种畸变事件类型之间的层级关系。

[0532] 患者事件状态.患者可测量任何数量的畸变,但可仅对于所测量的那些事件是异常的。因此,将每一事件层面畸变的患者状态记录为异常、野生型或不经测量的。

[0533] 患者维度与患者的数据集特有的映射一起有助于正确鉴别完全野生型患者(测量了所有畸变类型但并不含有任何畸变的患者)。

[0534] 当基因层面和路径层面事件经定义时,仅将测量了所有3种基因数据类型(突变、融合体和拷贝数)的患者针对所述事件标记为“异常”或“野生型”。此假设具有与数据类型特有的事件层面畸变相比可能减少针对基因或路径层面畸变概述的患者数量的作用。如果患者的事件层面畸变类型(融合体、突变、扩增或缺失)中的至少一者异常,那么将所述患者视为基因层面事件异常。如果患者在作为路径定义的一部分的至少一种基因中具有畸变,那么将所述患者视为路径层面事件异常。在每一种情况下,患者必须已测量所有事件类型。

[0535] 在获得功能和融合体的情况下,畸变频率是约50%。对于驱动基因畸变事件,畸变频率也是约50%,但仅一半的患者包括于频率的分子和分母中。

[0536] 分析.一旦构筑了所有驱动基因基因事件,就对每一基因事件执行一组分析,计算癌症类型、临床上相关亚型之内和癌症类型之间(泛癌症)的频率、相关性和关系。以下为每一分析的简短描述,包括使用何种数据以及对报导的结果放置何种限制(如果有的话):频率、表达的频率、共同出现率和相互排他性、临床相关性分析等。

[0537] 频率.频率是畸变在被测量所述畸变的患者之间的出现率。计算在癌症类型、临床上相关癌症亚型和泛癌症内的频率。报导具有至少一个异常患者的所有事件。

[0538] 表达的频率.表达的频率是基因在事件阳性患者当中表达的频率。对于每一事件,基因的表达水平用于确认表达的频率。计算在癌症类型和临床上相关癌症亚型但非泛癌症内的频率。

[0539] 共同出现率和相互排他性.使用费舍尔精确检验计算每对事件的共同出现率和相互排他性。需要至少2个对每一事件阳性的患者和5个测量全部事件的患者用于计算。两个个别拷贝数缺失或扩增事件的共同出现率或排他性未经计算。此外,并未计算具有“任何”驱动基因状态的事件对之间的共同出现率和相互排他性(即,仅比较驱动基因对比驱动基因和驱动基因对比任何)。计算在癌症类型和临床上相关癌症亚型但非泛癌症内的相关性。

[0540] 临床相关性分析.测试每一驱动基因事件相对于一组可获得的临床亚型的相关性。通过比较一种临床亚型对比另一种临床亚型的患者中的基因事件的出现率,使用费舍尔精确检验对每一相关性进行测试。举例来说,可在吸烟者对比非吸烟者中、或在I期对比II期肺癌中测试功能缺失突变的过度表达。针对每一事件测试总共136个亚型对,且下文列出定义所述亚型的特性(一些特性可为疾病特有的)。需要总共至少4个患者(且每种类别中至少1个患者)来执行测试。计算在癌症类型、临床上相关癌症亚型和泛癌症内的相关性。

[0541] 临床亚型特性名称:

[0542] 人种/种族

[0543] 经修改的吸烟状态

[0544] ERBB2状态

[0545] 雌激素受体状态

[0546] 孕酮受体状态

[0547] TCGA PAM50亚型

[0548] BRAF突变状态

[0549] 经修改的T期

[0550] 经修改的N期

[0551] 经修改的M期

[0552] 经修改的分期

[0553] KRAS突变状态

[0554] EGFR扩增状态

[0555] TCGA亚型

[0556] 微卫星状态

[0557] 人乳头瘤病毒感染状态

[0558] 临床结果分析.使用对数秩检验对每一事件与临床结果的相关性进行测试。仅具有可用的临床数据的患者组才用于计算,故包括于测试中的患者数量可能少于针对驱动基因事件测量的患者数量。需要至少4个对事件异常的患者来执行所述测试。存活时间以年呈现,且个别存活/死亡事件清楚地标记在卡普兰-迈耶曲线上。计算在癌症类型和临床上相关癌症亚型但非泛癌症内的相关性。

[0559] DNA-RNA相关性分析.对于每一基因,使用皮尔森氏相关性(Pearson's correlation)针对测量疾病的这些数据类型的所有患者当中的相关性测试RNA表达和DNA拷贝数值。计算在癌症类型和临床上相关癌症亚型但非泛癌症内的相关性。

[0560] 差异表达分析.对于每一事件,使用斯图登氏T检验(Student's T-Test)针对事件阳性患者对比事件阴性患者中的差异表达测试与事件相关的每一基因。对于涉及数种基因的事件,如融合体,测试每一基因。计算在癌症类型和临床上相关癌症亚型但非泛癌症内的差异表达。

[0561]

表 15：与癌症预后相关的事件

| 表 15 亚组 | 事件类型 | p 值 | q 值 | 阳性数量 | 患者总数量 | 染色体区带 | 基因（Entrez ID） | 可药化基因 | KM 证据 |
|---|---|---|---|---|---|---|---|---|---|
| 肝细胞癌 | 峰内基因扩增 | 3.31E-02 | 9.93E-02 | 4 | 65 | 1q21.2 | ADAMTSL4（54507）、MCL1（4170） | MCL1 | 较差的预后 |
| 肝细胞癌 | 峰内基因缺失 | 2.47E-02 | 9.89E-02 | 4 | 65 | 13q14.2 | LPAR6（10161） | 无 | 较差的预后 |
| 鳞状细胞肺癌 | 失去功能的突变 | 1.60E-02 | 9.59E-02 | 7 | 175 | 4q31.3 | FBXW7（55294） | 无 | 较差的预后 |
| 鳞状细胞肺癌 | 失去功能的突变 | 3.14E-02 | 9.42E-02 | 7 | 175 | 9q34.3 | NOTCH1（4851） | NOTCH1 | 较差的预后 |
| 鳞状细胞肺癌 | 失去功能的突变 | 7.73E-03 | 9.28E-02 | 5 | 175 | 1p35.3 | ARID1A（8289） | 无 | 较差的预后 |
| 透明细胞肾细胞癌 | 峰内基因缺失 | 7.12E-03 | 9.25E-02 | 8 | 493 | 9p21 | CDKN2B（1030） | 无 | 较差的预后 |
| 侵袭性乳癌:ER 阳性 | 峰内基因扩增 | 2.17E-03 | 9.13E-02 | 15 | 635 | 17q11.2 | TIAF1（9220）、MYO18A（399687）、CRYBA1（1411） | 无 | 较差的预后 |
| 卵巢浆液性囊腺癌 | 峰内基因扩增 | 1.00E-03 | 8.99E-02 | 10 | 557 | 19q13.1-q13.2 | PSG2（5670）、PSG5（5673）、CEACAM1（634）、CEACAM8（1088）、CXCL17（284340）、RABAC1（10567）、ATP1A3（478） | CEACAM1（临床前） | 较差的预后 |
| 透明细胞肾细胞癌 | 失去功能的突变 | 2.44E-02 | 8.55E-02 | 14 | 293 | 3p21 | BAP1（8314） | 无 | 较差的预后 |
| 卵巢浆液性囊腺癌 | 峰内基因扩增 | 5.45E-04 | 8.39E-02 | 89 | 557 | 19q12 | C19orf2（8725） | 无 | 较差的预后 |
| 肺腺癌 | 峰内基因扩增 | 6.80E-03 | 8.16E-02 | 4 | 320 | 1q12 | CHD1L（1105） | 无 | 较差的预后 |
| 肺腺癌 | 峰内基因扩增 | 6.80E-03 | 8.16E-02 | 4 | 320 | 1q21.1 | FMO5（2330）、PRKAB2（5565） | 无 | 较差的预后 |
| 肺腺癌 | 峰内基因扩增 | 8.57E-03 | 7.71E-02 | 9 | 320 | 12p12.1 | KRAS（3845）、CASC1（55259）、LYRM5（144363）、LRMP（4033） | KRAS（临床前） | 较差的预后 |
| 侵袭性乳癌:三阴性 | 峰内基因扩增 | 8.10E-03 | 7.29E-02 | 5 | 88 | 8p12 | BRF2（55290）、ERLIN2（11160）、GPR124（25960）、PROSC（11212）、RAB11FIP1（80223）、ZNF703 | 无 | 较差的预后 |

[0562]

| 表 15<br>亚组 | 事件类型 | p 值 | q 值 | 阳性数量 | 患者总数量 | 染色体区带 | 基因（Entrez ID） | 可药化基因 | KM 证据 |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | （80139） | | |
| 头颈鳞状细胞癌 | 峰内基因扩增 | 1.02E-02 | 6.93E-02 | 8 | 316 | 5q35 | THOC3（84321） | 无 | 较差的预后 |
| 直肠腺癌 | 峰内基因缺失 | 2.08E-03 | 6.86E-02 | 4 | 145 | 16p13.3 | A2BP1（54715） | 无 | 较差的预后 |
| 肺腺癌 | 峰内基因缺失 | 9.37E-03 | 6.09E-02 | 5 | 320 | 3q25.1 | AADAC（13） | 无 | 较差的预后 |
| 肝细胞癌 | 峰内基因缺失 | 3.03E-02 | 6.06E-02 | 4 | 65 | 8p21.2 | GNRH1（2796） | GNRH1 | 较差的预后 |
| 卵巢浆液性囊腺癌 | 峰内基因扩增 | 5.58E-04 | 6.02E-02 | 22 | 557 | 20q11 | ID1（3397）、BCL2L1（598）、COX4I2（84701） | ID1（临床前）、BCL2L1 | 较差的预后 |
| 直肠腺癌:KRAS 野生型 | 失去功能的突变 | 2.69E-02 | 5.38E-02 | 11 | 28 | 5q21-q22 | APC（324） | 无 | 较差的预后 |
| 乳头状肾细胞癌 | 峰内基因扩增 | 2.68E-02 | 5.35E-02 | 6 | 100 | 17q21.1 | CCL3L3（414062）、CCL3L1（6349） | 有 | 较差的预后 |
| 急性骨髓性白血病 | PML+RARA 融合 | 1.26E-02 | 5.03E-02 | 15 | 169 | 17q 和 15q | RARA（5914）、PML（5371） | 无 | 良好的预后 |
| 直肠腺癌:KRAS 野生型 | 峰内基因扩增 | 4.55E-02 | 4.96E-02 | 10 | 27 | 20q | ACOT8（10005）、ADA（100）、C20orf111（51526）、C20orf123（128506）、C20orf165（128497）、CD40（958）、CDH22（64405）、CTSA（5476）、DBNDD2（55861）、DNTTIP1（116092）、ELMO2（63916）、FITM2（128486）、GDAP1L1（78997）、GTSF1L（149699）、HNF4A（3172）、IFT52（51098）、JPH2（57158）、KCNK15（60598）、KCNS1（3787）、L3MBTL（26013）、MATN4 | ADA，CD40（958）、MMP9, PI3 | 较差的预后 |

[0563]

| 表 15 亚组 | 事件类型 | p 值 | q 值 | 阳性数量 | 患者总数量 | 染色体区带 | 基因（Entrez ID） | 可药化基因 | KM 证据 |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | （ 8785 ） 、 MMP9（ 4318 ） 、 MYBL2（ 4605 ） 、 NCOA5（ 57727 ） 、 NEURL2（ 140825 ） 、 PABPC1L（ 80336 ） 、 PCIF1（63935）、PI3（5266）、PIGT（51604）、PKIG（ 11142 ） 、 PLAGL2（5326）、PLTP(5360)、POFUT1（ 23509 ） 、 R3HDML（140902）、RBPJL(11317)、RIMS4（ 140730 ） 、 SDC4（6385）、SEMG1(6406)SEMG2（ 6407 ） 、 SERINC3（ 10955 ）、SFRS6（6431）、SGK2（ 10110 ） 、 SLC12A5（ 57468 ） 、 SLC13A3（ 64849 ） 、 SLC35C2（51006）、SLPI(6590)、SNAI1（6615）、SNX21（ 90203 ） 、 SPINLW1（ 57119 ） 、 SPINT3（ 10816 ） 、 SPINT4（ 391253 ） 、 STK4（ 6789 ） 、 SYS1（ 90196 ） 、 TM9SF4（ 9777 ） 、 TNNC2（ 7125 ） 、 TOMM34（ 10953 ） 、 TOX2 | | |

[0564]

| 表 15 亚组 | 事件类型 | p 值 | q 值 | 阳性数量 | 患者总数量 | 染色体区带 | 基因（Entrez ID） | 可药化基因 | KM 证据 |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | （ 84969 ）、 TP53RK （ 112858 ）、 TP53TG5 （ 27296 ）、 TTPAL （ 79183 ）、 UBE2C （ 11065 ）、 WFDC10A （ 140832 ）、 WFDC10B （ 280664 ）、 WFDC11 （ 259239 ）、 WFDC12 （ 128488 ）、 WFDC13 （ 164237 ）、 WFDC2 （ 10406 ）、 WFDC3 （ 140686 ）、 WFDC5 （ 149708 ）、 WFDC6 （ 140870 ）、 WFDC8 （ 90199 ）、 WFDC9 （ 259240 ）、 WISP2 （ 8839 ）、 YWHAB （ 7529 ）、 ZNF334 （ 55713 ）、 ZNF335 （ 63925 ）、 ZSWIM1 （ 90204 ）、 ZSWIM3 （140831） | | |
| 胃腺癌 | 失去功能的突变 | 4.09E-03 | 4.50E-02 | 4 | 131 | 6p21 | HLA-B（3106） | 有 | 较差的预后 |
| 子宫内膜子宫内膜样腺癌 | 峰内基因缺失 | 1.36E-02 | 4.43E-02 | 6 | 446 | 16Q24 | SLC7A5（8140）、CTU2（ 66965 ）、 FAM38A（ 9780 ）、 CDT1（81620）、APRT（353）、GALNS（2588） | SLC7A5（临床前） | 较差的预后 |
| 肺腺癌 | 峰内基因缺失 | 3.15E-03 | 4.09E-02 | 8 | 320 | 19q13.4 | KIR2DS4（3809） | 无 | 较差的预后 |
| 头颈鳞状细胞癌 | 峰内基因扩增 | 1.45E-03 | 3.82E-02 | 6 | 316 | 20p12 | C20orf94（128710）、JAG1（182）、MKKS（8195）、SNAP25（6616） | JAG1（临床前） | 较差的预后 |

[0565]

| 表 15<br>亚组 | 事件类型 | p 值 | q 值 | 阳性数量 | 患者总数量 | 染色体区带 | 基因（Entrez ID） | 可药化基因 | KM 证据 |
|---|---|---|---|---|---|---|---|---|---|
| 肺腺癌;三阴性 | 峰内基因扩增 | 8.80E-04 | 3.78E-02 | 6 | 174 | 7q31 | MET（4233）、CAPZA2（830） | MET | 较差的预后 |
| 子宫内膜子宫内膜样腺癌 | 峰内基因缺失 | 1.09E-02 | 3.77E-02 | 4 | 446 | 3Q26 | APOD（347） | 无 | 较差的预后 |
| 皮肤黑素瘤 | 失去功能的突变 | 3.74E-03 | 3.74E-02 | 16 | 148 | 17q11 | NF1（4763） | 无 | 较差的预后 |
| 急性骨髓性白血病 | CBFB+MYH11 融合 | 1.83E-02 | 3.67E-02 | 11 | 169 | 16Q22 和 16P13.11 | CBFB（865）、MYH11（4629） | 无 | 良好的预后 |
| 头颈鳞状细胞癌 | 峰内基因扩增 | 6.01E-04 | 2.86E-02 | 5 | 316 | 7p12 | ABCA13（154664）、C7orf57（136288）、C7orf65（401335）、C7orf69（80099）、C7orf72（100130988）、DDC（1644）、FIGNL1（63979）、GRB10（2887）、HUS1(3364)、IKZF1（10320）、PKD1L1（168507）、SUN3（256979）、TNS3（64759）、UPP1（7378）、VWC2（375567）、ZPBP(11055) | 无 | 较差的预后 |
| 肺腺癌 | 峰内基因扩增 | 1.28E-03 | 2.76E-02 | 7 | 320 | 7q31 | MET（4233）、CAPZA2（830） | MET | 较差的预后 |
| 头颈鳞状细胞癌 | 峰内基因扩增 | 3.00E-03 | 2.31E-02 | 80 | 316 | 11q13 | FADD（8772）、PPFIA1（8500）、ANO1（55107）、CTTN（2017） | 无 | 较差的预后 |
| 胃腺癌 | 峰内基因扩增 | 1.89E-04 | 2.14E-02 | 4 | 172 | 18q11 | GATA6（2627） | 无 | 较差的预后 |
| 侵袭性乳癌 | 峰内基因缺失 | 2.27E-03 | 1.82E-02 | 11 | 863 | 10q23.31,q23.2 | ATAD1（84896）、KILLIN（100144748） | 无 | 较差的预后 |
| 头颈鳞状细胞癌 | 峰内基因扩增 | 1.89E-03 | 1.55E-02 | 6 | 316 | 2q32 | GLS（2744）、MYO1B（4430）、NAB1(4664)、STAT1（6772）、STAT4（6775）、TMEM194B | 无 | 较差的预后 |

[0566]

| 表 15<br>亚组 | 事件类型 | p 值 | q 值 | 阳性数量 | 患者总数量 | 染色体区带 | 基因（Entrez ID） | 可药化基因 | KM 证据 |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | （100131211） | | |
| 结肠腺癌 | 峰内基因缺失 | 2.27E-04 | 1.48E-02 | 4 | 412 | 3Q26 | APOD（347） | 无 | 较差的预后 |
| 胃腺癌:高突变基因 | 失去功能的突变 | 5.32E-04 | 1.22E-02 | 4 | 32 | 2q31 | HOXD8（3234） | 无 | 较差的预后 |
| 成胶质细胞瘤 | 失去功能的突变 | 1.23E-03 | 1.11E-02 | 6 | 276 | Xq25 | STAG2（10735） | 无 | 较差的预后 |
| 头颈鳞状细胞癌 | 获得功能的突变 | 2.61E-03 | 1.04E-02 | 13 | 304 | 2q31 | NFE2L2（4780） | 无 | 较差的预后 |
| 子宫内膜子宫内膜样腺癌 | 峰内基因扩增 | 9.20E-04 | 1.03E-02 | 7 | 446 | 1q21 | SSR2（6746）、ARHGEF2（9181）、UBQLN4（56893） | 无 | 较差的预后 |
| 子宫内膜子宫内膜样腺癌 | 峰内基因缺失 | 2.47E-03 | 9.17E-03 | 7 | 446 | 16p13 | LOC339047（339047） | 无 | 较差的预后 |
| 肝细胞癌 | 峰内基因扩增 | 2.57E-03 | 8.89E-03 | 4 | 65 | 1q21.3 | DCST1（149095）、ADAM15（8751）、EFNA4（1945）、EFNA3（1944）、EFNA1（1942）、RAG1AP1（55974）、DPM3（54344）、KRTCAP2（200185）、TRIM46（80128）、MUC1（4582）、THBS3（7059）、MTX1（4580）、GBA（2629） | ADAM15, MUC1 | 较差的预后 |
| 透明细胞肾细胞癌 | 峰内基因缺失 | 3.16E-04 | 8.23E-03 | 8 | 493 | 9p21 | CDKN2A（1029） | 有 | 较差的预后 |
| 成胶质细胞瘤 | 获得功能的突变 | 2.72E-03 | 8.15E-03 | 14 | 276 | 2q33 | IDH1（3417） | 临床前 | 良好的预后 |
| 肺腺癌:KRAS 突变(无 ALK 融合且无 EGFR 突变) | 峰内基因扩增 | 2.56E-03 | 5.98E-03 | 4 | 78 | 12p12.1 | LYRM5（144363）、KRAS（3845）、CASC1（55259） | KRAS（临床前） | 较差的预后 |
| 子宫内膜子宫内膜样腺癌 | 峰内基因扩增 | 4.40E-04 | 5.55E-03 | 6 | 446 | 1q22 | ROBLD3（28956）、RAB25（57111）、MEX3A（92312） | 无 | 较差的预后 |
| 结肠腺癌：KRAS 突变 | 获得功能的突变 | 4.97E-03 | 4.97E-03 | 17 | 53 | 3q26 | PIK3CA（5290） | 有 | 较差的预后 |
| 头颈鳞状细胞癌 | 失去功能的突变 | 1.79E-04 | 3.95E-03 | 161 | 304 | 17p13 | TP53（7157） | TP53 | 较差的预后 |

[0567]

| 表 15<br>亚组 | 事件类型 | p 值 | q 值 | 阳性数量 | 患者总数量 | 染色体区带 | 基因（Entrez ID） | 可药化基因 | KM 证据 |
|---|---|---|---|---|---|---|---|---|---|
| 头颈鳞状细胞癌 | 峰内基因扩增 | 6.41E-05 | 3.81E-03 | 4 | 316 | 22q11 | CRKL（1399）、PI4KA（5297）、SERPIND1（3053）、SNAP29（9342） | 无 | 较差的预后 |
| 胃腺癌 | 失去功能的突变 | 2.14E-04 | 3.53E-03 | 5 | 131 | 17q22 | RNF43（54894） | 无 | 较差的预后 |
| 低级别胶质瘤 | 失去功能的突变 | 3.00E-04 | 2.70E-03 | 5 | 166 | 17q11.2 | NF1（4763） | 无 | 较差的预后 |
| 肺腺癌:三阴性 | 获得功能的突变 | 5.06E-04 | 2.53E-03 | 11 | 175 | 3q26.3 | PIK3CA（5290） | 有 | 较差的预后 |
| 肺腺癌 | 失去功能的突变 | 5.24E-05 | 9.96E-04 | 4 | 283 | 5q21-q22 | APC（324） | 无 | 较差的预后 |
| 膀胱尿道上皮癌 | 峰内基因扩增 | 8.34E-05 | 9.31E-04 | 5 | 125 | 5p15.33 | PLEKHG4B（153478）、LRRC14B（389257）、CCDC12（151903）、SDHA（6389）、PDCD6（10016）、AHRR（57491）、C5orf55（116349）、EXOC3（11336）、SLC9A3（6550）、CEP72（55722）、TPPP（11076）、BRD9（65980）、TRIP13（9319）、NKD2（85409）、SLC12A7（10723）、SLC6A19（340024）、SLC6A18（348932）、TERT（7015）、CLPTM1L（81037）、SLC6A3（6531）、LPCAT1（79888）、MRPL36（64979）、NDUFS6（4726） | AHRR, TERT | 较差的预后 |
| 子宫内膜子宫内膜样腺癌 | 峰内基因缺失 | 1.12E-04 | 8.32E-04 | 9 | 446 | 10q23 | PTEN（5728）、ANKRD22（118932）、 | 有 | 较差的预后 |

[0568]

| 表 15 亚组 | 事件类型 | p 值 | q 值 | 阳性数量 | 患者总数量 | 染色体区带 | 基因（Entrez ID） | 可药化基因 | KM 证据 |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | STAMBPL1（57559）、ACTA2（59）、FAS（355）、ATAD1（84896）、KILLIN（100144748）、RNLS（55328） | | |
| 低级别胶质瘤 | 峰内基因扩增 | 5.69E-04 | 6.74E-04 | 5 | 206 | 1q32.1 | C1orf157（284573）、ETNK2（55224）、GOLT1A（127845）、KISS1（3814）、LAX1（54900）、LRRN2（10446）、MDM4（4194）、PIK3C2B（5287）、PLEKHA6（22874）、PPP1R15B（84919）、REN（5972）、SNRPE（6635）、SOX13（9580）、ZC3H11A（9877） | MDM4（临床前） | 较差的预后 |
| 卵巢浆液性囊腺癌 | 峰内基因缺失 | 1.05E-06 | 6.28E-04 | 7 | 557 | 9q22 | FAM75C1（441452） | 无 | 较差的预后 |
| 子宫内膜子宫内膜样腺癌 | 峰内基因扩增 | 8.93E-06 | 3.01E-04 | 25 | 446 | 8q24 | MYC（4609）、TAF2（6873）、DSCC1（79075）、DEPDC6（64798） | 无 | 较差的预后 |
| 急性骨髓性白血病 | 失去功能的突变 | 2.35E-05 | 9.42E-05 | 12 | 184 | 17P | TP53（7157） | 有 | 较差的预后 |
| 结肠腺癌 | 峰内基因扩增 | 3.93E-06 | 6.24E-05 | 7 | 412 | 12p13 | CCND2（894）、TULP3（7289）、TEAD4（7004）、TSPAN9（10867）、PRMT（563418）、EFCAB4B（84766）、PARP11（57097）、C12orf5 | 无 | 较差的预后 |

[0569]

| 表 15<br>亚组 | 事件类型 | p 值 | q 值 | 阳性数量 | 患者总数量 | 染色体区带 | 基因（Entrez ID） | 可药化基因 | KM 证据 |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | （ 57103 ） 、 FGF23（8074）、FGF6（2251）、FKBP4（2288）、ITFG2（ 55846 ） 、 NRIP2（ 83714 ） 、 FOXM1（2305） | | |
| 胃腺癌 | 失去功能的突变 | 8.74E-07 | 2.88E-05 | 4 | 131 | 2q31 | HOXD8（3234） | 无 | 较差的预后 |
| 低级别胶质瘤 | 获得功能的突变 | 9.38E-08 | 2.81E-07 | 130 | 166 | 2q33.3 | IDH1（3417） | IDH1（临床前） | 良好的预后 |
| 低级别胶质瘤 | 峰内基因扩增 | 1.31E-08 | 3.48E-08 | 14 | 206 | 7p11.2 | EGFR（1956）、SEC61G（23480） | EGFR | 较差的预后 |
| 低级别胶质瘤 | 峰内基因扩增 | 1.48E-10 | 1.18E-09 | 5 | 206 | 12q14.1 | CDK4 （ 1019 ） 、CYP27B1 （ 1594 ） 、MARCH9 （ 92979 ） 、TSPAN31 （ 6302 ） 、AGAP2（116986）、AVIL（ 10677 ） 、 CTDSP2（ 10106 ） 、 FAM119B（ 25895 ） 、 METTL14234）、OS9（10956）、TSFM（10102） | CDK4 | 较差的预后 |
| 低级别胶质瘤 | 获得功能的突变 | 1.09E-10 | 6.56E-10 | 6 | 166 | 7p12 | EGFR（1956） | EGFR | 较差的预后 |
| 肺腺癌 | 峰内基因扩增 | 1.30E-12 | 4.66E-11 | 4 | 320 | 12p11 | LOC100133893（100133893）、MRPS3（ 604885 ） 、 REP15（387849） | 无 | 较差的预后 |
| 低级别胶质瘤 | 峰内基因缺失 | 4.57E-12 | 6.85E-12 | 21 | 206 | 9p21 | CDKN2A （ 1029 ） 、CDKN2B(1030)、MTAP（4507） | CDKN2A（1029） | 较差的预后 |
| 子宫内膜子宫内膜样腺癌 | 峰内基因扩增 | 2.00E-15 | 1.01E-13 | 4 | 446 | 17q21 | CCL3L3 （ 414062 ） 、CCL3L1（6349） | 无 | 较差的预后 |
| 星形细胞瘤 | 失去功能的突变 | | 3.88E-03 | 34 | 59 | 17p13.1 | TP53（7157） | TP53 | 有利的结果 |

[0570]

| 表 15<br>亚组 | 事件类型 | p 值 | q 值 | 阳性数量 | 患者总数量 | 染色体区带 | 基因（Entrez ID） | 可药化基因 | KM 证据 |
|---|---|---|---|---|---|---|---|---|---|
| 星形细胞瘤 | 失去功能的突变 | | 8.15E-03 | 22 | 59 | Xq21.1 | ATRX（546） | 无 | 有利的结果 |
| 乳癌 | 峰内基因缺失 | | 8.14E-03 | 4 | 36 | 8p23.2 | CSMD1（64478） | 无 | 较差的结果 |
| 结肠直肠腺癌 | 峰内基因扩增 | | 5.71E-02 | 12 | 407 | 8q24.3 | PARP10（84875）、MAPK15（225689）、PTK2（5747）、KHDRBS3（10656） | PTK2 | 较差的结果 |
| 结肠直肠腺癌 | 峰内基因扩增 | | 9.18E-02 | 17 | 407 | 13q34 | FAM70B（348013） | 无 | 较差的结果 |
| 结肠直肠粘液腺癌 | 获得功能的突变 | | 8.10E-03 | 8 | 32 | 3q26.3 | PIK3CA（5290） | PIK3CA | 较差的结果 |
| 皮肤黑素瘤 | 峰内基因扩增 | | 2.60E-06 | 6 | 231 | 8q22.3 | ODF1（4956） | 无 | 较差的结果 |
| 皮肤黑素瘤 | 峰内基因扩增 | | 1.54E-04 | 7 | 231 | 8q24.3 | PARP10（84875）、MAPK15（225689）、PTK2（5747）、KHDRBS3（10656） | PTK2 | 较差的结果 |
| 皮肤黑素瘤 | 峰内基因扩增 | | 7.21E-03 | 8 | 231 | 8q21 | HEY1（23462） | 无 | 较差的结果 |
| 皮肤黑素瘤 | 峰内基因扩增 | | 7.59E-03 | 6 | 231 | 11q13.3 | FADD（8772）、CCND1（595）、ORAOV1（220064）、FGF19（9965） | CCND1 | 较差的结果 |
| 皮肤黑素瘤 | 峰内基因扩增 | | 1.82E-02 | 4 | 231 | 1q44 | OR2T27（403239） | 无 | 较差的结果 |
| 皮肤黑素瘤 | 峰内基因扩增 | | 9.36E-02 | 6 | 231 | 1q21.3 | LCE1E（353135） | 无 | 较差的结果 |
| 导管乳癌 | 峰内基因扩增 | | 2.77E-03 | 4 | 665 | 3q29 | OSTalpha（200931） | 无 | 较差的结果 |
| 导管乳癌 | 峰内基因扩增 | | 2.28E-02 | 7 | 665 | 6q23.3 | AHI1（54806） | 无 | 较差的结果 |
| 导管乳癌 | 峰内基因扩增 | | 2.64E-02 | 8 | 665 | 3q26.3 | PIK3CA（5290）、SOX2（6657）、ATP11B（23200） | PIK3CA | 较差的结果 |
| 导管乳癌:ER 阳性和 HER2 阴性 | 峰内基因扩增 | | 7.92E-06 | 6 | 263 | 1q21.3 | ADAMTSL4（54507）、MCL1（4170）、ENSA（2029） | MCL1 | 较差的结果 |
| 导管乳癌:ER 阳性和 HER2 阴性 | 峰内基因扩增 | | 4.02E-02 | 7 | 263 | 1q32 | MDM4（4194） | MDM4（临床前） | 较差的结果 |
| 导管乳癌:ER 阳性和 HER2 | 峰内基因缺失 | | 4.35E-02 | 4 | 263 | 8p11.2 | FKSG2（59347） | 无 | 较差的结果 |

[0571]

| 表 15<br>亚组 | 事件类型 | p 值 | q 值 | 阳性数量 | 患者总数量 | 染色体区带 | 基因（Entrez ID） | 可药化基因 | KM 证据 |
|---|---|---|---|---|---|---|---|---|---|
| 阴性 | | | | | | | | | |
| 导管乳癌:ER 阳性和 HER2 阳性 | 峰内基因缺失 | | 7.48E-02 | 4 | 84 | 9q22 | FAM75C1（441452） | 无 | 较差的结果 |
| 导管乳癌:HER2 阳性 | 峰内基因缺失 | | 4.47E-02 | 4 | 116 | 15q13.1 | CHRFAM7A（89832） | 无 | 较差的结果 |
| 导管乳癌:HER2 阳性 | 峰内基因缺失 | | 5.17E-02 | 4 | 116 | 9p21 | CDKN2B（1030） | CDKN2B（临床前） | 较差的结果 |
| 导管乳癌:三阴性 | 峰内基因扩增 | | 2.58E-02 | 5 | 75 | 1q23.3 | APOA2（336）、SDHC（6391）、FCGR2B（2213） | 无 | 较差的结果 |
| 导管乳癌:三阴性 | 峰内基因扩增 | | 7.21E-02 | 8 | 75 | 1q21 | ACP6（51205）、ECM1（1893）、ADAMTSL4（54507）、MCL1（4170）、ENSA（2029） | MCL1 | 较差的结果 |
| 子宫内膜子宫内膜样腺癌;微卫星稳定 | 失去功能的突变 | | 5.55E-02 | 19 | 113 | 5q13.1 | PIK3R1（5295） | 无 | 较差的结果 |
| 子宫内膜浆液性腺癌 | 峰内基因扩增 | | 6.37E-04 | 4 | 52 | 19p13.2 | DNMT1（1786） | DNMT1 | 较差的结果 |
| 胃腺癌:高突变基因 | 峰内基因缺失 | | 9.05E-02 | 8 | 106 | 9p21 | CDKN2A（1029）、CDKN2B（1030） | CDKN2A,<br>CDKN2B（临床前） | 较差的结果 |
| 成胶质细胞瘤 | 峰内基因缺失 | | 2.58E-02 | 300 | 565 | 9p21 | CDKN2A（1029）、CDKN2B（1030） | CDKN2A,<br>CDKN2B（临床前） | 较差的结果 |
| 成胶质细胞瘤 | 峰内基因扩增 | | 8.80E-02 | 189 | 565 | 7p11.2 | SEC61G（23480） | 无 | 较差的结果 |
| 肺腺癌 | 融合 | | 5.79E-02 | 7 | 343 | 17q23.1 | RPS6KB1（6198）、VMP1（81671） | RPS6KB1 | 较差的结果 |
| 肺腺癌:三阴性 | 失去功能的突变 | | 1.31E-03 | 4 | 99 | 7q36.1 | MLL3（58508） | 无 | 较差的结果 |
| 少枝星形细胞瘤 | 失去功能的突变 | | 1.97E-02 | 38 | 53 | 17p13.1 | TP53（7157） | TP53 | 有利的结果 |
| 少突神经胶质瘤 | 失去功能的突变 | | 5.90E-02 | 6 | 89 | 9q34.3 | NOTCH1（4851） | NOTCH1 | 较差的结果 |
| 少突神经胶质瘤 | 失去功能的突变 | | 6.62E-02 | 15 | 89 | 1p31.1 | FUBP1（8880） | 无 | 较差的结果 |
| 卵巢浆液性腺癌 | 峰内基因扩增 | | 1.15E-02 | 17 | 562 | 19q13.1 | FCGBP（8857）、PAK4（10298） | PAK4（临床前） | 较差的结果 |

[0572]

| 表15<br>亚组 | 事件类型 | p值 | q值 | 阳性数量 | 患者总数量 | 染色体区带 | 基因（Entrez ID） | 可药化基因 | KM证据 |
|---|---|---|---|---|---|---|---|---|---|
| 卵巢浆液性腺癌 | 峰内基因扩增 | | 6.59E-02 | 17 | 562 | 20q11.2-13.2 | ZNF217（7764）、MYLK2（85366）、KIF3B（9371） | 无 | 较差的结果 |
| 卵巢浆液性腺癌 | 峰内基因缺失 | | 7.86E-02 | 7 | 562 | 17p13.1 | ATP1B2（482） | 无 | 较差的结果 |
| 卵巢浆液性腺癌 | 峰内基因扩增 | | 8.43E-02 | 53 | 562 | 19q12 | CCNE1（898） | CCNE1 | 较差的结果 |
| 鳞状细胞肺癌 | 峰内基因扩增 | | 7.93E-02 | 63 | 320 | 3q26.2 | MECOM（2122） | 无 | 有利的结果 |

[0573] 实例5:其他融合体方法

[0574] 临床数据来源.用于基因融合体分析的所有RNASeq数据均从癌症基因组学中心

(CGHub,当前的TCGA基因组数据的存储库-https://cghub.ucsc.edu/)中获得。

[0575] 细胞系数据来源.用于基因融合体分析的所有CCLE RNASeq数据均从癌症基因组学中心(CGHub,当前的CCLE NGS数据的存储库-https://cghub.ucsc.edu/)中获得。

[0576] BAM到FASTQ转换.到融合体调用程序中的输入由呈FASTQ格式的RNASeq读段组成,其需要将由TCGA提供的BAM文件转换成一个或两个FASTQ文件以(分别)用于单端或双端数据。

[0577] BAM文件在出处和处理方面变化,且许多需要特殊处理。举例来说,由UNC提供的较早的BAM文件使用BWA(巴罗斯-惠勒对准器(Burrows-Wheeler Aligner))对准,而含有读段的较新BAM通过MapSplice对齐。TCGA近来更新了RNASeq流水线以支持替代基因表达报导。(之前的流水线依赖于基因表达的RPKM测量值,而后者使用RSEM。)这些不同的RNASeq分析流水线由UNC分别提及为V1和V2(https://wiki.nci.nih.gov/display/TCGA/RNASeq+Version+2)。当两种格式均可用于同一TCGA样品时,我们使用以下BAM优先级排序流水线来选择单一"原始BAM":1)相对于V1BAM选择V2BAM;和2)当呈现相同情况的多个文件时,选择上传日期较新的BAM。

[0578] 上文描述的定制SamToFastq转换器用于从TCGA BAM文件生成FASTQ文件。

[0579] 有2种癌症类型(COADREAD和UCEC)仅以单端RNASeq数据形式可用。对于单端BAM文件转换,使用程序BamTools(安全超文本传输协议://github.com/pezmaster31/bamtools)来生成FASTQ文件。

[0580] 在支持单端和双端数据两者的目标下,我们使用TopHat处理所有单端数据,且使用deFuse处理所有双端数据。

[0581] 广泛地,我们的分析流水线由5个主要步骤组成:

[0582] 预处理原始数据以获得FASTQ文件

[0583] 运行融合体调用程序

[0584] 过滤断裂点到相关基因区域

[0585] 用Oncomine转录物组注释断裂点

[0586] 对可能令人感兴趣的新颖融合体进行概述并区分优先级

[0587] 基于高性能云端计算簇对所有样品平行执行步骤1和2。对聚集的数据进行过滤和注释作为后处理步骤,以启用各种过滤和注释流程的作用的探索性分析。在完成过滤准则以使假阳性融合体减到最少(步骤5)之后,Oncomine优先级排序融合体的列表用RNASeq外显子表达数据验证。

[0588] TopHat.TopHat-Fusion从作者超文本传输协议://tophat.cbcb.umd.edu中获得。软件和参考数据相依性如由TopHat资料所规定配置:

[0589] 软件:

[0590] TopHat:2.0.4,包括TopHat-Fusion Post(04/09/2012发行)

[0591] bowtie:0.12.8(05/06/2012发行)

[0592] samtools:0.1.18(09/02/2011发行)

[0593] blast(2.2.26)(03/03/2012发行)

[0594] blast+(2.2.26)(10/21/11发行)

[0595] 参考和注释:

[0596] 参考基因组:UCSC hg19(2012年5月下载)

[0597] 基因模型:refGene、ensGene(2012年5月下载)

[0598] BLAST DB:nt、人类、其他(2012年5月下载)

[0599] 参数:

[0600] 我们用基本上默认的参数如TopHat资料中所规定的对单端和双端TCGA Illumina数据运行TopHat。以下是所用参数的列表。

[0601] 表25

[0602]

| TopHat参数 | 所用的值 |
|---|---|
| --fusion-search | 标记 |
| --keep-fasta-order | 标记 |
| --no-coverage-search | 标记 |
| --mate-inner-dist | 0 |
| --mate-std-dev | 80 |
| --min-anchor-length | 8 |
| --splice-mismatches | 0 |
| --min-intron-length | 70 |
| --max-intron-length | 500,000 |
| --max-insertion-length | 3 |
| --max-deletion-length | 3 |
| --num-threads | 4 |
| --max-multihits | 20 |
| --transcriptome-mismatches | 2 |
| --genome-read-mismatches | 2 |
| --read-mismatches | 2 |
| --segment-mismatches | 2 |
| --segment-length | 25 |
| --fusion-min-dist | 100,000 |
| --fusion-anchor-length | 13 |
| --fusion-read-mismatches | 2 |
| --fusion-multireads | 2 |
| --fusion-multipairs | 2 |
| --fusion-ignore-chromosomes | chrM |

[0603] --mate-inner-dist和--mate-std-dev参数不具有默认值。第一参数规定RNASeq双端读段的预期插入大小,而第二参数规定所述值的预期标准偏差。TopHat作者为大部分数据集推荐0和80的值。

[0604] 表26

[0605]

| TopHat-Fusion Post 参数 | 所用的值 | 值的阐述 |
|---|---|---|
| **--num-fusion-reads** | 3 | *推荐的值* |
| **--num-fusion-pairs** | 0 | *设定为0 以不惩罚低证据，但可能的重要融合体* |
| **--num-fusion-both** | 0 | |

[0606] 对一个样品同时执行TopHat-融合体,紧接着执行TopHat-Fusion Post。我们保留未过滤的TopHat-Fusion输出和过滤的TopHat-Fusion Post输出两者以启用更深的分析。

[0607] deFuse.deFuse从作者:超文本传输协议://defuse.sf.net获得。软件和参考数据相依性如由deFuse资料所规定配置:

[0608] 软件:

[0609] deFuse:0.5.0(04/07/2012发行)

[0610] bowtie:0.12.8(05/06/2012发行)

[0611] R 2.15.0(03/30/2012发行)

[0612] blat、faToTwoBit(05/01/2012获得)

[0613] 参考和注释:

[0614] 参考基因组:Ensembl GRCh37.62fa(2012年5月下载)

[0615] 基因模型:Ensembl gtf(2012年5月下载)

[0616] 基因组数据:

[0617] UCSC EST fasta、EST比对和重复序列(2012年5月下载)

[0618] NCBI UniGene(2012年5月下载)

[0619] 参数:

[0620] 我们用默认参数如deFuse程序资料中所规定运行deFuse。

[0621] 表27

[0622]

| deFuse参数 | 所用的值 |
|---|---|
| -bowtie_quals | phred33-quals |
| -max_insert_size | 500 |
| -discord_read_trim | 50 |
| -clustering_precision | 0.95 |
| -span_count_threshold | 5 |
| -split_count_threshold | 3 |
| -percent_identity_threshold | 0.90 |
| -max_dist_pos | 600 |
| -num_dist_genes | 500 |
| -split_min_anchor | 4 |
| -max_concordant_ratio | 0.1 |
| -splice_bias | 10 |
| -denovo_assembly | No |
| -probability_threshold | 0.5 |
| -covariance_sampling_density | 0.01 |
| -reads_per_job | 1,000,000 |
| -regions_per_job | 20 |
| -p | 4 |

[0623] 同时对一个样品执行deFuse。我们保持deFuse输出的过滤和未过滤的结果两者以启用更深的分析。

[0624] 整合.我们整合“层面I”数据,来自TopHat-Fusion Post的potential_fusion.txt文件的输出和来自deFuse的results.classify.tsv文件的输出。deFuse报导了与TopHat相比在此层面下许多更潜在的调用,且因此还可报导更多假阳性预测。选择层面I数据以在利用调用程序的内置过滤和允许遍及足够的结果之间力求平衡以鉴别具有稍微较弱证据的可能的真正融合体。

[0625] 由于每一调用程序为融合体调用提供不同层面的注释和支持证据,提取来自两种调用程序的所预测的融合体的断裂点,且将其整合成常见格式以便过滤和注释。整合步骤由以下组成:将所报导的断裂点转化成基于基因组学坐标系的断裂点且合并成常见文件格式。

[0626] 断裂点过滤.过滤来自调用程序的“层面I”输出的所预测的融合体,仅保留其中每一断裂点是在RefSeq转录物的5'UTR或CDS区中的那些调用(refGene大约于2012年7月18日,从UCSC获得)。如此进行以为含有功能基因区域的那些富集所预测的融合体,滤除例如其中一种基因的3'UTR被预测为与另一种基因的3'UTR融合的融合体调用。尽管在基因组DNA层面,断裂点可能出现在内含子中,但在RNASeq数据中,在最接近的外显子-内含子边界处将观察到所述断裂点。因此,还排除被预测会出现在内含子序列中的断裂点。

[0627] 断裂点注释.在排除在RefSeq转录物的5'UTR或CDS区域外的融合体之后,将来自RefSeq转录物的注释转移到具有相对于多个Entrez ID注释的一些预测的剩余断裂点。

[0628] 对于每对断裂点,每个Entrez ID仅保留一个转录物。在多个转录物的情况下,选择具有最短转录物登录号的转录物;通过以字母数字方式分选登录号且保留第一个登录号来打破进一步联系。这个流程确保了在相同位置处的注释断裂点中的一致性。然而,相同基因搭配物的不同位置处的所预测的断裂点仍可产生多个代表一对基因的转录物,这是替代转录物的可能的证据。

[0629] 丢弃来自调用程序自身的基本注释,因为其是基于每一对应的调用程序的默认注释来源。然而,保留TopHat和deFuse两者的某些输出字段以帮助对所预测的融合体进行优先级排序。另外,从其他调用程序特性推断未由调用程序明确报导的某些注释特性。

[0630] 推断的特性.从每一调用程序获得支持和跨越读段计数,且概述在读段跨度和读段跨度支持两栏中。后一栏是跨越融合体的读段和支持融合体的读段的总和(不与“其中一端跨越融合体的跨越配对物对”的TopHat的计数混淆,其有时被称为‘跨越和支持读段’)。

[0631] 将通过调用程序报导的断裂点序列修整以包括在融合体每一侧上的50个碱基,且合并到断裂点序列一栏中。融合体断裂点通过“|”描绘。应注意,此是如通过调用程序推断的断裂点序列,且并非是简单地从参考基因组获得的。因为推断的序列可能反映通过跨越读段观察到的实际序列,所以此序列可能代表参考基因组序列的互补序列。

[0632] 由于调用程序两者中无一提供决定性的‘5’’或‘3’’标记,故我们通过将调用程序参数与基因链注释组合来推断融合体搭配物的相对5'-3'定向。对于deFuse,基于基因链和deFuse输出特性‘基因组_链’的以下组合,推断每一搭配物的定向:

[0633] 表28

[0634]

| **基因链** | ***deFuse_基因组_链*** | |
|---|---|---|
| | + | - |

[0635]

| | | |
|---|---|---|
| + | 5' | 3' |
| - | 3' | 5' |

[0636] TopHat报导不同的量度,即,映射到基因搭配物的读段的相对定向,因此需要不同的规则集合来推断基因对的5'-3'顺序:

[0637] 表29

[0638]

| 基因 A/B 链 | *tophat_定向* | | | |
|---|---|---|---|---|
| | ff | fr | rr | Rf |
| +/+ | 5'-3' | | 3'-5' | |
| +/- | | 5'-3' | | 3'-5' |
| -/- | 3'-5' | | 5'-3' | |
| -/+ | | 3'-5' | | 5'-3' |

[0639] 如果存在既定基因融合体调用的推断的5'和3'搭配物,那么将有效定向字段标记为"Y"。

[0640] RepeatMasker注释.每一预测的断裂点位置还用在断裂点附近的RepeatMasker特征注释。如此进行以鉴别高度重复的基因组区域中的断裂点,其中比对错误很可能会影响嵌合转录物的预测。

[0641] 具体来说,将5'和3'搭配物断裂点的25bp序列上游或下游分别选择为'断裂点侧接序列'。这些侧接序列相对于在2012年8月24日从UCSC表浏览器下载的RepeatMasker元件组(超文本传输协议://www.repeatmasker.org/)相交。我们报导了每一断裂点中与26个碱基断裂点侧接区域重叠的元件名称、元件长度和量。目前,未针对特定元件类型(LINES、SINES、简单重复序列等)过滤RepeatMasker元件。

[0642] 对于每一融合体预测,我们设定RepeatMasker重叠字段以等于断裂点侧接序列与RepeatMasker元件重叠的碱基数量,且将12个或更多个碱基的重叠视为显著的。显著重叠融合体调用的频率用于下文所述的Oncomine优先级排序,使得将具有较低重叠频率的基因融合体视为较高质量。

[0643] 融合体外显子表达不平衡.融合体使用RNASeq外显子表达数据观测,以通过寻找在断裂点调用之前和之后的外显子表达不平衡来提供真阳性融合体事件的次要证据。具体来说,如果3'搭配物的表达受到5'搭配物的启动子区影响,那么外显子表达应在所预测的断裂点之后增强。当查看融合对比非融合患者样品时,这一效应尤其可见。

[0644] TCGA外显子表达数据.TCGA外显子表达数据从博德GDAC Firehose位置下载。将每一患者的RPKM RNASeq值以对应于来自数个不同基因定义(包括RefSeq)的UCSC外显子的复合物的基因注释格式(GAF)特征形式列出。在下载21种疾病的数据之后,我们发现4组不同的GAF特征用于注释RPKM表达。最后,患者表达数据的可用性在上文描述的V1和V2RNASeq分析流水线中按疾病变化。

[0645] 为了解决这些挑战,我们首先将UCSC RefSeq外显子映射到可用的GAF特征,且计算每一RefSeq外显子与GAF特征之间的重叠百分比。此步骤至关重要,因为所有CBI处理的融合体断裂点均映射到在7/18/2012下载的UCSC Refgene定义,且这些断裂点必须又映射到GAF特征。396,298个RefSeq外显子中的80.8%完美地映射到下文展示的曲线中的GAF特征。我们选择且报导会产生最大重叠的RefSeq外显子和GAF特征对。

[0646] 称为rg_pct的值基于下式提供给定RefSeq外显子与GAF特征的映射质量的量度:

[0647] $rg_pct = 重叠/长度_{refseq} = 重叠/长度_{GAF特征}$

[0648] rg_pct值为1的映射完美地重叠，而值小于1指示RefSeq外显子或GAF特征不映射到严格相同的基因组区域，且RPKM值可为可疑的。

[0649] 我们选择除STAD以外(归因于V2数据的不可用性)的所有疾病的RNASeq V2数据。

[0650] 细胞系外显子表达数据.细胞系样品的外显子表达数据由从CGHub获得的CCLE BAM文件产生。所采用的方法与步骤18类似，如这里可用的“UNC数据的TCGA mRNA-seq流水线”方法中所述：安全超文本传输协议://webshare.bioinf.unc.edu/public/mRNAseq_TCGA/UNC_mRNAseq_summary.pdf。

[0651] UNC方法与我们的方法之间的差异在于在我们的方法中使用RefSeq外显子BED而非由TCGA使用的复合外显子BED。

[0652] 外显子表达不平衡计算.系统地分析每一样品中的3’搭配物表达中的潜在5’启动子诱发的不平衡的证据。首先，将每一基因的表达水平转换成对数标度，且接着将z分数在每一疾病的样品组中归一化。此归一化在外显子层面执行以顾及广泛群体趋势，如3’偏差或较差RefSeq外显子/GAF特征匹配(参见下文)。

[0653] 浆液性卵巢癌患者中的PLXNB21和COL7A1的粗RPKM表达值(顶部)对比z分数归一化值(参见图8A-D)。在外显子12、17和23处的PLXNB1表达中的广泛群体突降在归一化数据中被平滑处理。被预测具有这些基因之间的融合体的样品以红色突出显示；野生型患者以蓝色展示。红色菱形指示调用程序预测的断裂点外显子。

[0654] 在归一化之前，将被视为考虑中的融合体的野生型但被预测具有涉及基因搭配物中的一者的其他融合体的样品从野生型群体中移出，以便不会污染z分数计算。

[0655] 在归一化之后，基于样品的断裂点后归一化的表达值(群体A)与断裂点前值相比具有较高平均值的假设，每一样品被分配经由单边斯图登氏t检验计算的p值($H_{*I}$ $\mu_A \leq \mu_B$)。调用程序预测的断裂点用于分离通过任一融合体调用程序鉴别的样品的表达群体。

[0656] 还计算每一野生型样品的P值以在整个群体的情形下促进融合体阳性样品的p值的分析。这允许我们将涉及展现并非融合体诱发的广泛群体外显子不平衡趋势的基因的融合体丢弃。将p值不在野生型样品p值的最前部分内评级的任何样品丢弃。将使断裂点前和断裂点后表达水平之间的差异最大化的断裂点用于野生型样品p值计算。

[0657] 融合体概述.基于独特基因对的出现以及基于个别基因(有可能与多个搭配物一起)的出现，概述疾病内的融合体。

[0658] 对于独特融合体对(根据Entrez ID对唯一)，根据任一调用程序具有至少一个融合体预测的疾病内的样品数量是融合的样品计数。由于在一个样品中以及跨越多个样品可报导相同基因对的多个断裂点，故每一疾病内的独特融合体对的数量远远小于融合体调用的总数。为了在基因对层面而非融合体调用层面对融合体进行过滤并区分优先级，概述数种融合体调用程序特性。下表展示针对跨越个别预测的给定融合体搭配物对概述的特性：

[0659] 表30

[0660]

| 特性 | 概述方法 |
|---|---|
| DEFUSE翻转 | 总融合体调用%＝‘Y’ |
| DEFUSE有效定向 | 总融合体调用%＝‘Y’ |
| DEFUSE多重映射数量 | 总融合体调用%>0 |
| TOPHAT有效定向 | 总融合体调用%＝‘Y’ |

| | |
|---|---|
| 3P/5P_REPEATMASKER重叠 | 总融合体调用%≥12 |

[0661] 如果基因组上的基因相隔<1Mb,那么为融合体设定邻近标记,且在这些融合体搭配物的≤75%的个别融合体预测中设定defuse翻转标记。

[0662] 基因层面概述.融合的样品计数还在每一疾病类型和跨越疾病(泛癌症)内的基因层面(根据Entrez gene ID唯一)概述。此概述方法与融合体内推断的定向无关。另外,仅运算Oncomine优先级融合体的融合的样品计数(如下所述)。

[0663] 个别独特融合体对交叉参考基因组畸变的Mitelman数据库(超文本传输协议://cgap.nci.nih.gov/Chromosomes/Mitelman)。匹配基于基因名称而非疾病类型进行。因此,Mitelman中报导的某一疾病中的基因融合体可能已出现在TCGA数据集中的不同疾病类型中。

[0664] 在基因层面处概述的基因融合体基于基因名称交叉参考Mitelman数据库。因此,如在Mitelman中报导为与所预测的独特融合体对相比具有不同组织学或完全不同畸变类型(例如较大染色体层面缺失而非融合体)的基因有更多潜力。

[0665] 正常样品融合体黑名单.假设TCGA正常样品中调用的所有融合体均为假阳性,我们提出以下问题:1)肿瘤样品中的融合体调用是否在正常样品中鉴别?2)肿瘤样品中鉴别的Oncomine优先级排序的融合体是否也在正常样品中鉴别?回答第一个问题提供肿瘤基因融合体调用中的技术性假阳性率的基线意义。第二个问题是对Oncomine优先级融合体过滤怎样很好地克服此问题的健全检查。下载10种疾病内的344个双端正常样品,且使用上文描述的相同的deFuse流水线处理。观察由6,024个独特融合体组成的总共56,579个总融合体调用。这些正常样品融合体调用用于生成黑名单,且从Oncomine优先级基因融合体中移除这些假阳性。

[0666] 平行同源融合体搭配物黑名单.在平行同源基因家族成员之间的融合体的黑名单使用两种策略汇编:1)人工地检测高频率融合体搭配物基因名称以及2)比较所有优先级融合体搭配物基因名称的前3个字符。在后一种策略中,融合体搭配物使用HomoloGene、Ensembl、SIMAP和GeneDecks V3验证为“平行同源”,随后包括在最终黑名单中。下表展示出前10个最常观察到的在平行同源融合体搭配物之间的基因融合体调用。整个表由超过400个独特平行同源基因融合体组成且用于从我们的Oncomine优先级基因融合体中移除这些假阳性。

[0667] 表31

[0668]

| 基因A符号 | 基因B符号 | 在正常中观察到的 | TCGA癌症类型 |
|---|---|---|---|
| **HLA-B** | HLA-C | 有 | BLCA、BRCA、CESC、COAD、GBM、HNSC、KICH、KIRC、KIRP、LAML、LGG、LIHC、LUAD、LUSC、OV、PAAD、PRAD、SKCM、STAD、THCA、UCEC |
| **HLA-A** | HLA-B | 有 | BLCA、BRCA、CESC、GBM、HNSC、KICH、KIRC、KIRP、LAML、LGG、LIHC、LUAD、LUSC、OV、PAAD、PRAD、SKCM、STAD、THCA |
| **HLA-A** | HLA-C | 有 | BLCA、BRCA、CESC、GBM、HNSC、KICH、KIRC、KIRP、LAML、LGG、LIHC、LUAD、LUSC、OV、PAAD、PRAD、SKCM、STAD、THCA |
| **TTLL12** | TTLL12 | 有 | BLCA、BRCA、CESC、GBM、HNSC、KICH、KIRC、KIRP、LAML、LGG、LIHC、LUAD、LUSC、OV、PAAD、PRAD、SKCM、STAD、THCA |
| **TRPV1** | TRPV1 | 有 | BLCA、BRCA、CESC、GBM、HNSC、KICH、KIRC、KIRP、LAML、LGG、LIHC、LUAD、LUSC、OV、PAAD、PRAD、SKCM、STAD、THCA |
| **B9D1** | B9D1 | 有 | BLCA、BRCA、CESC、GBM、HNSC、KICH、KIRC、KIRP、LGG、LIHC、LUAD、LUSC、OV、PAAD、PRAD、SKCM、THCA |
| **TGIF2-C20ORF24** | TGIF2-C20ORF24 | 有 | BLCA、BRCA、CESC、GBM、HNSC、KICH、KIRC、KIRP、LAML、LGG、LIHC、LUAD、LUSC、OV、PRAD、SKCM、STAD、THCA |
| **HLA-B** | HLA-E | 有 | BLCA、BRCA、CESC、COAD、GBM、HNSC、KICH、KIRC、KIRP、LAML、LGG、LIHC、LUAD、LUSC、OV、PAAD、PRAD、READ、SKCM、STAD、THCA、UCEC |
| **SEC16A** | SEC16A | 有 | BRCA、CESC、GBM、HNSC、KICH、KIRC、KIRP、LAML、LGG、LUAD、LUSC、OV、PRAD、SKCM、THCA |
| **LOC390940** | LOC390940 | 有 | BLCA、BRCA、CESC、GBM、HNSC、KICH、KIRC、KIRP、LGG、LUAD、LUSC、OV、SKCM、STAD、THCA |

[0669] 融合体优先级排序-Oncomine优先级流程.下文概述的Oncomine优先级流程通过以下方式设计:迭代探索层面I融合体预测中的最前结果,且系统地消除可疑的假阳性融合体,同时保留先前发现的‘真阳性’(Mitelman)融合体。此流程意欲将符合‘真阳性’融合体预期的某些特征且相反地缺乏许多‘假阳性’融合体中观察到的特征的融合体突出。

[0670] 在以下情况中时,融合体是Oncomine优先级融合体:

[0671] 表32

[0672]

| 融合体概述特性 | 值 | 阐述 |
|---|---|---|
| DEFUSE 有效定向<br>TOPHAT 有效定向 | > 0.75 | 大部分预测呈正确定向 |
| 邻近 | 'N' | |
| REPEATMASKER 频率 | < 0.25 | 所预测的断裂点中的少数或无一者在重复区域中 |
| DEFUSE 多重映射数量 | > 0 | 大部分跨越读段独特地映射到融合体断裂点 |
| 平行同源搭配物 | 不在平行同源黑名单上 | 平行同源基因之间的所预测的融合体的人工管理的黑名单 |
| 在正常中所观察到的 | 不在正常黑名单上 | 从使用 deFuse 处理 344 个正常样品获得的列表 |

[0673] 实例6:Oncomine NGS突变方法

[0674] 突变整合.数据整合的目的是形成目前可用的最完整的NGS突变数据组。我们考虑以下来源:

[0675] 原始数据来源

[0676] COSMIC细胞系项目

[0677] 来自博德GDAC Mutation_Packager(stddata build)的TCGA数据

[0678] 来自DCC 2级的TCGA数据

[0679] 基于TCGA数据的肯潘迪亚公司突变调用

[0680] 含有NGS突变数据的出版物

[0681] COSMIC细胞系项目

[0682] 癌症基因组项目已针对突变表征了超过1000种癌细胞系的外显子组。数据库提供针对质量过滤的突变数据,呈平面文件格式。细胞系数据经历用于临床突变数据的相同Oncomine管理和注释方法。细胞系名称相对于Oncomine本体审查,且将癌症类型标准化以便与临床突变数据相当。

[0683] 数据集当其在2013年11月呈现时从维康信托桑格研究所细胞系项目(Wellcome Trust Sanger Institute Cell Lines Project)网站:超文本传输协议://cancer.sanger.ac.uk/cancergenome/projects/cell_lines/获得。

[0684] 博得GDAC Mutation_Packager.博得自2011年第三季度起已致力于收集和整合来自多个来源的突变数据。

[0685] https://docs.google.com/document/d/18X1Wv-a9xLBOflNikOa9rCXOyiravMM8_-PVJx AQPPo/edit

[0686] 上述文献详述了博德整合到Mutation_Packager标准数据运行中的MAF文件的出处。博德已整合了许多MAF文件,所述MAF文件维持在中央TCGA DCC系统外,通常由分析工作组自身的成员维持。我们已执行了在对我们可用的所有MAF文件之间的深入比较。我们相信,博德具有最完整的可用的突变数据。

[0687] 对此版本,我们整合来自2013_08_09stddata build的数据。

[0688] TCGA DCC 2级.此是可从DCC获得的受控制的访问突变数据。TCGA在其维基(wiki)上具有提供关于可用的MAF文件的额外细节的页面:

[0689] https://wiki.nci.nih.gov/display/TCGA/TCGA+MAF+Files

[0690] 对此版本,我们考虑到2013年9月15日为止可用的所有MAF文件。

[0691] 肯潘迪亚公司NGS DNASeq突变调用.我们感觉到可从TCGA获得的PRAD突变调用具有低质量且产生假阳性'获得功能'预测。因此,对此疾病的所有调用均来源于肯潘迪亚公

司的自身突变调用流水线。作出肯潘迪亚公司突变调用以符合MAF文件格式以便整合。关于更多详情，请参见附录：肯潘迪亚公司NGS DNASeq突变调用(Compendia NGS DNASeq Mutation Calling)。在此版本中包括170个前列腺腺癌患者。

[0692] 所有NGS数据的人工管理.TCGA和非TCGA NGS数据集由Oncomine管理小组直接从其原始来源(主要是同行审查的癌症出版物和上述公开可访问的数据库)获得。使补充材料中通常可用的突变数据达到突变再注释和分类所需的标准作为整个NGS突变处理流水线的一部分。癌症类型使用Oncomine癌症类型本体管理，基于出版物中存在的最佳可用的临床元数据，分配适当Oncomine癌症类型。由于所有公开的实验均要求全基因组('NGS')覆盖率，故假设每一数据集的空基因组包括所有人类RefSeq基因。非TCGA数据以与用于突变分析流水线的其余部分的TCGA MAF文件数据完全相同的方式处理。

[0693] 移除一式两份突变.我们执行一些简单净化操作以移除源数据中存在的一式两份突变记录。我们还执行若干文件-栏名称再映射，因为许多来源并不遵守MAF文件标准。移除来自同一患者样品的各种肿瘤/正常等分试样对的一式两份突变。

[0694] 突变注释.从TCGA和非TCGA来源获得的数据含有由通过不同基因组测序中心或作者经若干年的过程处理和注释的数据集产生的突变。这产生使用不同基因模型和使用不同的变异体分类约定注释的突变调用。由于肯潘迪亚公司定义突变的方法依赖于准确的变异体注释，我们相对于单一转录物组和一致的变异体分类规则再注释突变。标准注释流水线确保跨越疾病类型的突变被一致地评估且在潜在致癌基因或肿瘤抑制基因的提名期间经受共同的解释。还提供不可从原始来源一致地获得的重要注释，如HGVS样式突变命名法(例如V600E)。

[0695] 从原始来源获得的突变通过肯潘迪亚公司根据以下通用步骤处理(下文提供详情)。

[0696] 我们首先使用肯潘迪亚公司的Oncomine转录物组再注释每一突变。成功注释的突变接受源于肯潘迪亚公司的注释，而其余部分保留从原始来源获得的注释。注释包括：

[0697] 变异体分类

[0698] 变异体位置

[0699] 变异体变化

[0700] 实施若干过滤步骤以移除多个转录物中的冗余注释以及位于相关基因区域之外的突变。

[0701] 排除"超突变基因"样品.在某些疾病(如子宫内膜癌)中，数个高度突变的样品可能在整个突变计数中占主导地位。我们还在肺腺癌、胃癌、黑素瘤和结肠直肠癌中观察到所述"超突变基因"样品。基于通过分析数种癌症类型中的超突变基因离群值确定的截断值，我们决定将<5,000个非沉默外显子突变作为用于将样品包括在我们的再现分析中的阈值。我们因此从我们的下游分析流水线中排除此数据集中的许多超突变基因样品。

[0702] 在突变注释步骤中，我们试图相对于通过肯潘迪亚公司编译的标准转录物组再注释从原始来源获得的突变。这组转录物包括来自2012年2月19日从UCSC获得的hg18和hg19基因组构建的RefGene转录物。

[0703] 每一突变个别地针对规定的基因组构建内的Oncomine转录物组中的重叠群映射。SNP突变直接映射到其起始位置，同时针对小插入(INS)和缺失(DEL)突变选择相关位置以

便映射。对于插入，相关位置是出现插入的碱基。取决于转录物的方向，这可以是突变的起始或终止坐标，取决于基因分别是在正链上还是在负链上。对于缺失，相关位置在转录物位于正链上时为缺失的碱基或在转录物位于负链上时为缺失的最后一个碱基。此调节确保突变位置被定义为相对于转录物翻译的方向(即5'→3')受到插入/缺失影响的第一个碱基。

[0704] 对于成功地映射到转录物的突变，相对于所述转录物推断肯潘迪亚公司突变注释。对于未能映射的突变，保留来自原始数据来源的注释，且基于基因组坐标构筑用于热点计算的变异体位置(更多详情如下)。由于仅23个染色体的标准组包括于我们的转录物组中，故位于线粒体或其他非标准重叠群上的突变不经映射。

[0705] 下文是用于注释映射到Oncomine转录物组的突变的准则的描述。

[0706] 变异体分类.对于成功映射到转录物的每一突变，使用突变和注释特性的组合推断变异体分类。我们的方法鉴别了六个主要突变变异体分类，所有均位于转录物内。目前不考虑在基因区域外部(例如基因间)的突变的变异体分类(参见下文过滤部分)。下文是用于推断变异体分类的准则：

[0707] 表33

[0708]

| 变异体分类 | 准则 | 转录物区域 |
|---|---|---|
| 剪接位点 | 突变是在剪接位点的 2 bp 内 | 外显子或内含子 |
| 3'UTR、5'UTR | 突变是在 UTR 区域中而不是在剪接位点的 2 bp 内 | UTR 外显子 |
| 内含子 | 突变是在内含子中且在距剪接位点的 3 到 10 bp 之间 | 内含子 |
| 错义、无义、无终止、沉默 | 突变是 SNP | 编码外显子 |
| 框移插入/缺失 | 突变是不可被 3 除尽的插入/缺失 | 编码外显子 |
| 框内插入/缺失 | 突变是可被 3 除尽的插入/缺失 | 编码外显子 |
| 非编码外显子 | 突变是在非编码转录物中 | 非编码外显子 |

[0709] 变异体分类的此列表是由TCGA针对MAF文件格式规定的允许的变异体分类的子集。

[0710] https://wiki.nci.nih.gov/display/TCGA/Mutation+Annotation+Format+%28MAF%29+S pecification

[0711] 此子集覆盖相关突变类别以用于潜在获得或失去功能基因的再现分析和鉴别，且因此对于映射到Oncomine转录物组的绝大部分突变是足够的。下表描述了将对比原始作者分类(假设突变映射到与用于定义分类的转录物相同的转录物)分配的可能的变异体分类以及来源数据集中的所述突变类型的相对丰度：

[0712] 表34

[0713]

| 示例 TCGA 变异体分类 | 等效的肯潘迪亚公司变异体分类 | 潜在 Oncomine 突变分类<br>热点（H）、有害（D）或其他（O） |
|---|---|---|
| **错义突变** | 错义突变 | H、O |
| **无义突变** | 无义突变 | D |
| **无终止突变** | 无终止突变 | H、O |
| **沉默** | 沉默 | O |
| **框移缺失** | 框移缺失 | D |
| **框移插入** | 框移插入 | D |
| **翻译起始位点** | 错义突变 | O |
| **框内缺失** | 框内缺失 | H、O |
| **框内插入** | 框内插入 | H、O |
| **3'UTR** | 3'UTR | O |
| **5'UTR** | 5'UTR | O |
| **非编码外显子（或“RNA”）** | 非编码外显子 | H、O |
| **剪接位点** | 剪接位点 | O |
| **内含子** | 内含子 | -- |
| **5'侧接序列** | --不被 Oncomine 转录物组支持-- | -- |
| **IGR** | --不被 Oncomine 转录物组支持-- | -- |
| **其他（存在于突变列表中但不被 TCGA 支持的分类）** | | -- |

[0714] 变异体位置.当前分析的初始目的之一是鉴别具有热点突变的基因,热点突变为在多个肿瘤样品中在同一位置处观察到的某一分类的突变。为了有效地鉴别再现且界定每一突变的热点,我们必须构筑涵盖突变位置、所影响的氨基酸或碱基身份以及变异体分类的突变点识别符。我们使在同一位置处出现的与其产生的特定碱基变化无关的突变聚集。因此,我们仅使用参考碱基或氨基酸来定义变异体位置。这确保了将针对可能的热点对会影响同一密码子或基因组位置的突变进行计数,即使其产生的替代等位基因不同。举例来说,对于给定基因,错义突变V600E、V600F和V600G将均具有V600的变异体位置,且将由此在鉴别热点突变时聚集在一起。因此,我们的变异体位置如下定义:

[0715] 变异体位置＝突变点{碱基|密码子}+参考{碱基|AA}+[变异体分类]

[0716] 如果突变是在编码区中,那么在相关碱基处的密码子数量和对应的氨基酸用于鉴别突变点,例如p.L116_框内缺失。如果突变是在非编码区(如UTR)中,那么在相关碱基处的参考核苷酸的位置和身份用于鉴别突变点,例如c.*110C。

[0717] 对于在编码区外的剪接位点突变,相对于剪接边界规定变异体位置。相对位置使用+{1|2}或-{1|2}鉴别(剪接位点突变为在剪接点的2个碱基内的那些)。如同插入和缺失,为剪接位点突变添加“_剪接位点”的后缀。对于插入和缺失突变,向变异体位置中添加指示框内(“_框内插入”或“_框内缺失”)或框移(“_框移插入”或“_框移缺失”)的后缀。

[0718] 总之,以下为不同可能的变异体位置格式的实例:

[0719] 表35

[0720]

| 变异体类型 | 接近剪接位点？ | 在编码区中？ | 变异体位置 |
|---|---|---|---|
| SNP | 是 | 是 | p.A42_剪接位点 |

[0721]

| 变异体类型 | 接近剪接位点？ | 在编码区中？ | 变异体位置 |
|---|---|---|---|
| | | 否 | c.42+1_剪接位点 |
| SNP | 否 | 是 | p.A42（错义、无终止、沉默）<br>p.Stop42（无义） |
| | | 否 | c.*42T（3'UTR)<br>c.-42C（5'UTR)<br>c.42（非编码外显子） |
| 插入 | 是 | 是 | p.A42_剪接位点 |
| | | 否 | c.42+1_剪接位点 |
| | 否 | 是 | p.A42_{框内插入\|框移插入} |
| | | 否 | c.*42G_{框内插入\|框移插入}（3'UTR）<br>c.-42G_{框内插入\|框移插入}（5'UTR）<br>c.42（非编码外显子） |
| 缺失 | 是 | 是 | p.A42_剪接位点 |
| | | 否 | c.42+1_剪接位点 |
| | 否 | 是 | p.A42_{框内缺失\|框移缺失} |
| | | 否 | c.*42T_{框内缺失\|框移缺失}（3'UTR）<br>c.-42C_{框内缺失\|框移缺失}（5'UTR）<br>c.42（非编码外显子） |

[0722] 对于并不映射到Oncomine转录物组且因此并不具有基于转录物的位置的突变，将基因组位置(起始位置)和参考核苷酸(参考等位基因)用作与编码区或剪接位点接近性无关的变异体位置。由原始数据提供的变异体分类接着作为后缀添加。举例来说，SNP错义突变将具有变异体位置，如“chr19_c.C22952756_错义突变”，且剪接位点SNP将具有变异体位置“chr1_c.A155025094_剪接位点”。这些突变的变异体变化(参见下文)未经定义。

[0723] 尽管变异体位置的后缀通常隐含地结合变异体分类，当计算热点时，变异体位置和变异体分类两者明确地用于聚集突变。因此，分别运算可能产生相同变异体位置但具有不同变异体分类(如错义和无义SNP)的突变。

[0724] 变异体变化.变异体变化提供关于突变的替代等位基因变化的HGVS样信息。对于编码区中的SNP突变，变异体变化为完整HGVS蛋白质层面序列变异体描述，指示替代氨基酸。对于在编码区之外的SNP，提供替代等位基因核苷酸碱基。

[0725] 对于插入和缺失变异体类型，将变异体位置(参见上文)用作变异体变化。在这些情况下，不推断在氨基酸层面的变化的后果。因而，插入/缺失的变异体变化并不绝对地遵循HGVS规格。

[0726] 下文为源于肯潘迪亚公司的突变注释的变异体变化的说明性实例：

[0727] 表36

[0728]

| 突变 | 变异体分类 | 变异体位置 | 变异体变化 |
|---|---|---|---|
| SNP 于 CDS 中，E>K，残基 137 | 错义突变 | p.E137 | p.E137K |
| SNP 于内含子 C>2 中距剪接位点两个 bp | 剪接位点 | c.4913-1_剪接位点 | c.4913-1 |
| 插入于 CDS 中，在残基 Gly 264 处 | 框移插入 | p.G264_框移插入 | p.G264_框移插入 |

[0729]

| 突变 | 变异体分类 | 变异体位置 | 变异体变化 |
|---|---|---|---|
| UTR 中一个碱基的缺失 | 3'UTR | c.*1007A_框移缺失 | c.*1007A_框移缺失 |

[0730] 对于并不映射到Oncomine转录物组的突变，保留来自原始数据来源的变异体分类。

[0731] 转录物过滤.为了避免检索多个转录物,且因此避免检索一个基因内单一突变的多个注释,我们保持每一基因(唯一Entrez ID)每一突变仅一种转录物。如果突变映射到一种基因的数种转录物,那么仅选择一种转录物。然而,如果突变映射到数种基因,那么每一基因仅选择一种转录物。因此,一种突变可接受两种不同注释,但仅当其来自具有不同Entrez ID的转录物时才如此。

[0732] 我们基于以下优先级流程选择突变的代表转录物:

[0733] 具有最具影响力的变异体分类的转录物:

[0734] 在编码中的较高影响:错义、无义、无终止、框移

[0735] 在编码中的较低影响:框内、沉默

[0736] 在编码区外:剪接位点、3'或5'UTR、非编码外显子

[0737] 在外显子外:内含子

[0738] 如果存在基于优先级的联系,那么选择具有最短(根据长度)RefSeq转录物登录号的转录物,接着选择进一步联系的事件中在字母数字上最小的转录物登录号。举例来说,在TTN基因的转录物NM_003319、NM_133378和NM_00125685中,我们将选择NM_003319作为代表转录物。

[0739] 这些步骤允许我们为在一个位置处的相同类型的突变反复地选择一致的转录物。选择最具影响力的转录物的一个结果是多个转录物可用于在单一基因中的多个位置处的突变。然而,此流程的益处是在相同位置处的相同变异体分类的任何突变总是分配给相同转录物,且因此当计算用于热点鉴别的再现时将处于相同的参考框架中。

[0740] 通过突变类别和类型过滤.所有突变通过变异体类型和类别进一步过滤。为了避免将较不相关的突变包括到基因功能分析中,我们滤除并不解析到基因区域的突变,或者因为其显著远落在转录物之外,或者因为其处于与RefSeq基因不相关的位置中。这些突变通过其缺乏基因识别符(Entrez ID=0或空白)或以下变异体类别中的成员资格而为明显的:内含子、5'侧接序列、IGR和miRNA。

[0741] 我们还滤除具有DNP、TNP、ONP、复杂取代和Indel的变异体类型的突变,因为其注释不被我们的流水线所支持。

[0742] 由于某些数据来源包括广泛量的内含子和基因间突变,故此过滤步骤显著减小数据集的大小,因为许多NGS数据集并不应用这些过滤预出版物。

[0743] 将突变归类为热点、有害或其他.我们的分析流水线中的下一个步骤基于突变的变异体位置鉴别出多个样品中重现的突变,且将其归类到热点、有害或其他变异体类别中。对于此步骤和后续频率计算,独立地处理每种疾病类型的突变。仅将相同变异体分类的突变共同运算,因此例如分别计数在相同位置处的错义突变和沉默突变。

[0744] 为了鉴别驱动基因事件,取决于以下准则,将给定Entrez Gene ID的每一突变归类为"有害"或"热点":

[0745] 如果在3个或更多个肿瘤样品的相同变异体位置中观察到一种突变,那么认为所述突变是'再现的'。

[0746] 如果突变是以下各者,那么所述突变属于"热点"变异体类别:

[0747] 再现的且

[0748] 经以下变异体分类中的一者注释:

[0749] 框内插入/缺失

[0750] 无终止

[0751] 错义

[0752] 非编码外显子

[0753] 如果突变是以下各者,那么所述突变属于“有害”类别:

[0754] 非再现的且

[0755] 经以下变异体分类中的一者注释:

[0756] 框移插入/缺失

[0757] 无义

[0758] 如果一种突变不符合上述准则,那么所述突变被视为在“其他”变异体类别中。

[0759] Oncomine突变分类和变异体分类可以用于在基因层面概述各种突变的相对频率。

[0760] 命名“获得功能”和“失去功能”基因.将个别基因归类到所预测的功能类别中,即“获得功能”、“再现的其他”和“失去功能”,以反映其在潜在活化或有害突变中的相对富集。下文提供用于进行分类的流程的详情。

[0761] 突变的样品频率计算.相对于一种疾病类型内的所有样品中的给定变异体分类和变异体类别计算每一基因的突变频率。通过将所有变异体分类的突变组合来计算疾病内的基因的总体突变频率。

[0762] 总体突变频率.通过将在一种基因中具有任何变异体分类的至少一个突变的样品的总数(突变的样品计数)除以给定癌症类型中的样品的总数(样品计数)来获得所述基因的总体突变频率。

[0763] 热点频率.通过将具有属于“热点”Oncomine突变分类的至少一个突变的样品的总数除以突变的样品计数(具有给定基因的至少一个突变的样品的总数)来获得基因的热点频率。举例来说,如果样品具有热点错义和热点框内缺失两者,那么其将仅计数一次。

[0764] 热点错义频率.为了获得基因的热点错义频率,将含有具有“热点”的Oncomine突变分类的至少一个错义突变的样品的数量除以突变的样品计数(在此基因中具有至少一个任何类型的突变的样品数量)。具有一个以上所述类型的突变的样品仅计数一次。

[0765] 有害频率.为了获得基因的有害频率,将含有具有“有害”的Oncomine突变分类的至少一个突变的样品的数量除以突变的样品计数(具有给定基因的至少一个突变的样品数量)。具有一个以上所述类型的突变的样品仅计数一次。

[0766] 其他频率.为了获得基因的其他频率,将具有具备“其他”的Oncomine突变分类的至少一个突变的样品的总数除以突变的样品计数(具有给定基因的至少一个突变的样品总数)。举例来说,如果样品含有剪接位点和UTR突变两者,那么其将仅计数一次。

[0767] 热点、其他、和有害频率考量.热点、其他和有害频率应并不预期总计为100%,因为样品可能已在这些类别中的一种以上中计数。

[0768] 评估热点和有害突变的显著性.通过两种独立的方法计算疾病内的每一基因的热点和有害p值。

[0769] 有害突变富集的显著性.为了评估一种基因与其他基因相比是否针对有害突变显著富集,鉴于背景突变率,我们使用以下列联表执行费舍尔精确检验:

[0770] 表37

[0771]

| | 有害 | 其他 |
|---|---|---|
| 相关基因 | A | B |
| 所有其他基因 | C | D |

[0772] 其中A、B、C和D为跨越疾病的突变计数。将无义突变、框移插入和框移缺失归类为有害突变，而任何其他类型(UTR、沉默、错义等，但非基因间)的突变作为其他计数。

[0773] 通过计数具备有害突变的基因的数量(N)且计算每一相关的等级来计算每种疾病内的Q值。给定p值的q值则是Q＝p*N/等级。

[0774] 再现的热点突变的显著性.为了计算基因特有的p值，评估所述基因上再现最多的热点的显著性。鉴于每一序列位置很可能同等地突变的假设，可测试每一基因，再现最多的是否显著大于使用多项式检验所预期的。这是已在先前版本中实施的取样算法的精确检验。此检验的优点之一是p值精确度升高到1E-16，因此不出现取底。为了获得热点突变，我们过滤突变以移除不影响编码序列的任何(即，通过移除沉默、UTR、终止密码子和剪接位点突变进行)，且接着移除我们可能不用RefSeq转录物识别符注释的基因的突变数据。我们接着计数针对在每一疾病中的每一转录物观察到的突变。我们通过将CDS长度除以三且减去1来计算氨基酸序列长度。

[0775] p值的确切计算制定如下。给出长度x的氨基酸序列、观察到的热点突变的数量n，在再现最多的点处偶然观察到r个或更多个突变的可能性是多少。对于每一基因，通过下式计算p值：

[0776]
$$\begin{aligned} p&=Pr(y_{(X)} \geq r) \\ &=1\text{-}Pr(y_1 < r, y_2 < r, \ldots, y_X < r) \\ &=1\text{-}\sum_0^{r-1} \frac{n!}{y_1!y_2!\ldots y_X!}(1/x)^n \end{aligned}$$

[0777] 其中$y_{(x)}$是在再现最多的热点处的突变计数，且$y_1,\cdots,y_x$表示在每一点1,…,x处的突变计数。

[0778] 当n和x较大时，上式可极慢，使用具有邦弗朗尼-马洛斯(Bonferroni-Mallows，BM)界限的近似值：

[0779]
$$1-\frac{n!}{n^n e^{-n}}\left\{\prod\nolimits_{i=1}^{X} P(y_i \leq r-1)\right\}P(W=n)$$

[0780] 其中$y_i$是具有平均值n/x的泊松随机变量(Poisson random variable)，且$W=\sum_{i=1}^{X} Y_i$，其中Y是截尾泊松(truncated Poisson)。P(W＝n)通过埃奇沃思展开(Edgeworth Expansion)估算。p值的邦弗朗尼-马洛斯上限和下限是：

[0781]
$$1-\text{二项式}\left(r-1,n,\frac{1}{x}\right)^x \leq p \leq x*(1-\text{二项式}\left(r-1,n,\frac{1}{x}\right))。$$

[0782] 如果近似值落在BM界限之外，那么使用下限或上限。其很少出现在我们的数据中，且其主要针对较小p值(p<1e-16)或较大p值(p约等于1)出现。

[0783] Q值使用本亚明-霍赫贝格方法(Benjamini-Hochberg method)计算，其是Q＝p*N/等级，其中N是转录物数量且等级是每一p值的等级。

[0784] 沉默热点突变.再现的沉默突变-沉默热点似乎是测序错误的指示，出现在低序列质量的区域中且对于邻近核苷酸中的假阳性错义突变峰值来说充当‘煤矿中的金丝雀’。基

于检查具有沉默热点的基因以及邻近沉默峰值的评估,我们相信这些基因经受系统测序错误,且这些基因中的热点突变应不会有助于基因分类。

[0785] Oncomine基因分类规则.一旦突变已经分类,就将个别基因命名为三种类别中的一种:“获得功能”、“失去功能”和“再现的其他”。分类是基于在基因中观察到的突变的相对频率和显著性的组合。每一基因突变的显著性通过p值评估。

[0786] 再现的沉默突变.“获得功能”基因将具有相对较高的热点错义突变频率和较低的有害突变频率,而“失去功能”基因含有较大部分的有害突变。“再现的其他”倾向于含有再现的插入/缺失突变,其中的一些(例如1个碱基的再现的框移插入缺失)展现可能由局部比对错误引起的潜在假阳性调用的迹象。一般来说,我们更确信关于归类为获得/失去功能的基因的功能重要性。

[0787] 泛癌症分析.为了概述跨越疾病的突变,我们执行了与我们对于疾病内分析所执行的相同的计算,但不根据疾病将突变记录分层。将所有突变记录聚集,且在此泛癌症情形中计算频率、变异体类别和基因类别。对于泛疾病概述,概述所有癌症类型中的基因(根据Entrez ID唯一),且每个疾病一排。然而,还提供疾病内但在泛癌症情形中的基因的概述。此意指例如总计疾病内的具有热点突变的样品,但仅针对视为在泛癌症情形中的热点的突变。<20个样品的癌症类型包括于泛癌症分析中,尽管其归因于低样品计数而不符合条件疾病内分析。

[0788] 细胞系注释.细胞系突变数据经历上文描述的相同的Oncomine管理和注释方法,除了突变和基因分类。取而代之,每当细胞系中的突变也在临床样品中被观察到,来自细胞系的突变就用Oncomine突变分类和基因分类注释。此注释仅针对具有热点或有害或其他Oncomine突变分类的突变执行。如果在肿瘤中观察不到突变,那么其将接受“在肿瘤中观察不到”的突变分类。

[0789] 如果来自细胞系和肿瘤样品的突变属于相同基因,且具有相同变异体位置和变异体分类,那么将其视为等效物。

[0790] 细胞系名称相对于内部Oncomine本体审查,且将细胞系癌症类型标准化以便与临床突变数据相当。将在整个数据库或出版物中可能不经独立地验证的身份或癌症类型的数种细胞系从我们的分析中移除。来自临床数据的突变注释在泛癌症和疾病内情形中执行。

[0791] 肯潘迪亚公司NGS DNASeq突变调用

[0792] BAM文件选择.我们查询TCGA的CGHub以鉴别具有单一肿瘤-正常BAM对的患者。我们如此做以移除归因于不同肿瘤-正常对的突变调用差异的可能性。

[0793] 参考基因组构建.我们通过解析SAM标题来鉴别用于对准BAM文件中的读段的参考基因组构建。我们对如输入到突变调用程序包中所需的所有参考基因组构建进行定位、下载且编索引。

[0794] 突变调用.我们对此分析采用以下体细胞突变调用包:

[0795] MuTect(1.0.27783),博德研究所(Broad Institute),癌症基因组分析组(Cancer Genome Analysis Group,CGA)(齐布尔斯基(Cibulskis),2013)

[0796] SomaticIndelDetector(1.6-13-g91f02df),博德研究所,基因组分析工具包(Genome Analysis Toolkit,GATK)

[0797] MuTect.MuTect执行初始预处理以移除“具有太多错配或极低质量分数的读段”

(MuTect资料)。接着,对于候选突变,计算描述存在于肿瘤样品中的突变($LOD_T$)和在正常样品中不突变($LOD_N$)的可能性的两个对数差异(LOD)分数:

[0798] $$LOD_T = log_{10}\left(\frac{P(在肿瘤中观察到的数据|位点是突变的)}{P(在肿瘤中观察到的数据|位点是参考的)}\right)$$

[0799] $$LOD_N = log_{10}\left(\frac{P(在正常中观察到的数据|位点是参考的)}{P(在正常中观察到的数据|位点是突变的)}\right)$$

[0800] MuTect预期体细胞突变以约1的速率以Mb出现且需要$LOD_T$>=6.3。MuTect要求突变不在dbSNP中且具有$LOD_N$>=2.3,因为预期非dbSNP以每Mb 100个的速率出现。选择两个截断值以保证假阳性率小于预期的体细胞突变速率的一半。最后,执行其他后处理步骤,包括测试替代等位基因在两个读段方向中观察到。针对待考虑的突变,MuTect要求至少14个肿瘤读段和8个正常读段。

[0801] SomaticIndelDetector(SID).对于给定突变位点,SID使用基于计数的阈值和indel一致表决流程考虑候选插入缺失。选择具有最大数量的支持读段或投票的indel作为推定的indel调用。如果存在以下各者,那么报导此调用:

[0802] 足够的覆盖率(默认:正常>=4个读段,肿瘤>=6个读段),

[0803] 在所述位点的大部分读段支持推定的调用(默认:>=30%)

[0804] 此部分与支持indel的任何位点的那些相比足够大(默认:>=70%)

[0805] 如果在正常样品中存在相同indel的甚至较弱的证据,那么将肿瘤样品中的Indel调用注释为“生殖系”;否则的话,将其标记为“体细胞”。忽略仅在正常样品中观察到的调用。SID采用BAM文件作为输入和输出VCF和BED格式化的推定的调用。

[0806] 突变过滤.调用程序输出所有候选突变调用,包括具有低统计置信度的生殖系突变和其他调用。我们将突变调用程序输出过滤到仅体细胞突变、由MuTect标示“保留(KEEP)”的突变以及出现在RefSeq基因的CDS内的突变。下表详述应用到MuTect和SomaticIndelDetector输出的特定过滤器:

[0807] 表38

[0808]

| MuTect 过滤 | 描述 |
|---|---|
| tumor_f > 0.1 | 肿瘤读段中的至少 10%必须是变异体 |
| t_alt_sum / t_alt_count > 28 | 变异体碱基调用的平均质量>28 |
| t_alt_count - map_Q0_reads - improper_pairs >= 3 | 保守需要至少 3 个读段，其中变异体不在 Q0 中或在不恰当成对的读段中。 |
| t_alt_count > 10 * n_alt_count | 当 MuTect 允许一个变异体正常读段时,需要至少 10 个变异体肿瘤读段。 |
| dbsnp_site NE ‘DBSNP’ | 忽略 dbSNP v132 中存在的变异体 |

[0809]

| SomaticIndelDetector 过滤 | 描述 |
|---|---|
| T_STRAND_COUNTS_C[12] / (... C1 + ... C2) > 0.1 | 肿瘤变异体读段中的至少 10%必须在每一链上 |
| T_AV_MAPQ_C > 28 | 变异体调用的平均质量>28 |

[0810] 参考文献

[0811] 齐布尔斯基K.(Cibulskis,K.)等人不纯和异质癌症样品中的体细胞点突变的敏感性检测(Sensitive detection of somatic point mutations in impure and

heterogeneous cancer samples).《自然·生物技术》(Nat Biotechnology)(2013).doi:10.1038/nbt.2514

[0812] MuTect:超文本传输协议://www.broadinstitute.org/cancer/cga/mutect

[0813] SID:超文本传输协议://gatkforums.broadinstitute.org/discussion/35/somatic-indel-detection

[0814] 表16:表2基因/融合体的可药化状态

[0815]

| 表 16 基因 | 已批准的 | 注册前（批准前） | III 期 | II 期 | I 期 | 临床前 |
|---|---|---|---|---|---|---|
| TOP1 | 盐酸贝洛替康；盐酸伊立替康；拓扑替康 | N | 考西替康；伊立替康，HyACT；伊立替康，智擎生技制药公司；培舍埃特伊立替康 | 吉马替康；喜树碱，卡兰多公司；伊立替康盐酸盐 + 氟尿苷，赛拉特公司；培舍弗特康；TLC-388 盐酸盐；hRS7-SN-38；伊立替康珠粒，生物相容公司 | 伊立替康脂质体，养乐多公司；HM-30181A；纳迷特康；喜树碱前药，默萨那公司；拉贝珠单抗-SN-38；Genz-644282；盐酸辛米特康前药 | 喜树碱（艾菲沃斯公司）；伊立替康（生物联盟公司）；顺铂 + 伊立替康（赛拉特公司）；APH-0804；伊立替康（冠军公司）；SER-203；SN-38；拓扑替康 +（长春新碱）（利泊丘尔公司）；拓扑替康（恩杜 |

[0816]

| 表 16 基因 | 已批准的 | 注册前（批准前） | III 期 | II 期 | I 期 | 临床前 |
|---|---|---|---|---|---|---|
| | | | | | | 克斯医药公司） |
| SRD5A1 | 度他雄胺（dutasteride） | N | 依卓诺西（idronoxil） | N | N | N |
| VIM | N | N | N | 普托木单抗 | N | N |
| IGFBP2 | N | N | N | N | N | OGX-225 |
| SPP1 | N | N | N | N | N | N |
| MDK | N | N | N | N | N | CAMI-103；CMS-101 |
| MUC16 | N | N | 奥戈伏单抗（oregovomab） | N | DMUC-5754A | N |
| RET | 索拉非尼；凡德他尼；苹果酸舒尼替尼；卡博替尼；瑞格非尼 | 阿帕替尼 | 二磷酸莫替沙尼；SAR-302503 | N | JNJ-26483327 | MG-516；NMS-173；RET 激酶抑制剂（拜诺米克公司（Bionomic）） |
| MAP2K2 | 曲美替尼（trametinib） | N | ARRY-438162 | 司美替尼（selumetinib）；瑞法美替尼（refametinib）；皮玛瑟替（pimasertib）；WX-554 | PD-0325901；ARRY-704；TAK-733；GDC-0623；BI-847325；AS-703988 | N |
| MAPK1 | N | N | N | N | N | AEZS-129；AEZS-136；AEZS-134；SCH-722984；SCH-772984 |
| BRAF | 帕唑帕尼；维罗非尼；达拉菲尼 | N | N | RAF-265；XL-281；LGX-818 | ARQ-761；ARQ-736 | AB-024；b-raf 抑制剂（萨勒姆公司）；BRAF 激酶抑制剂（赛勒克萨金治疗公司）；BeiGene-283；DP-4978；TL-241 |
| MUC16 | N | N | 奥戈伏单抗 | N | DMUC-5754A | N |
| MET | 卡博替尼；克卓 | N | 提瓦替尼；里乐木单 | MGCD-265；弗 | AMG-208； | X-379；麦 |

[0817]

| 表 16 基因 | 已批准的 | 注册前（批准前） | III 期 | II 期 | I 期 | 临床前 |
|---|---|---|---|---|---|---|
| | 替尼 | | 抗；奥那组单抗； | 雷替尼；费拉妥珠单抗；BMS-777607；格瓦替尼；INCB-028060；LY-2875358 | TAS-115；沃利替尼；SAR-125844；S-49076 | 他替尼；PRS-110；ASP-08001；ARGX-111；DCC-2701；DCC-2721；MG-516；AL-2846；CG-206481；T-1840383；cMet-EGFR 双重抑制剂（水晶基因组公司）；双特异性抗体（霍夫曼-拉罗奇公司） |
| PTK2 | N | N | N | PF-04554878 | GSK-2256098；BI-853520；VS-4718 | CFAK-C4；FAK 抑制剂（维拉斯泰姆公司（Verastem））；CTX-0294945；CTx-0294886；FAK 抑制剂（武田公司） |
| ACE* | 阿拉普利（alacepril）；贝那普利（benazepril）；地拉普利（delapril）+ 马尼地平（manidipine）（凯西公司（Chiesi））；卡托普利（captopril）；卡托普利 + HCTZ；卡托普利缓释剂（三共公司（Sankyo））；西拉 | N | 培哚普利 + 吲达帕胺 + 氨氯地平（赛维尔公司） | N | 氨氯地平 + 马来酸依那普利（葛兰素史克公司（GlaxoSmithKline）） | N |

[0818]

| 表 16 基因 | 已批准的 | 注册前（批准前） | III 期 | II 期 | I 期 | 临床前 |
|---|---|---|---|---|---|---|
| | 普利（cilazapril）；地拉普利；地拉普利 + 吲达帕胺（indapamide）（凯西公司）；地尔硫卓（diltiazem），阿尔扎公司（Alza）；马来酸依那普利（enalapril maleate）；马来酸依那普利 + HCTZ；依那普利（enalapril）+ 尼群地平（nitrendipine）；依那普利（克尔卡公司（KRKA））；依那普利拉（enalaprilat）；非洛地平（felodipine）+ 依那普利；福辛普利（fosinopril）；咪达普利（imidapril）；赖诺普利（lisinopril）；赖诺普利 + HCTZ；莫西普利（moexipril）；培哚普利（perindopril）；盐酸喹那普利（quinapril hydrochloride）；喹那普利拉（quinaprilat）；雷米普利（ramipril）；非洛地平 + 雷米普利；培哚普利 + 吲达帕胺，瑟伍公司（Serv）；乙酸沙拉新（saralasin acetate）；螺普利（spirapril）；替莫普利（temocapril）；群多普利（trandolapril）；佐 | | | | | |

[0819]

| 表 16 基因 | 已批准的 | 注册前（批准前） | III 期 | II 期 | I 期 | 临床前 |
|---|---|---|---|---|---|---|
| | 芬普利（zofenopril）；群多普利 + 维拉帕米（verapamil），艾文公司（Aven）；乐卡地平（lercanidipine）+ 依那普利（瑞克达提公司（Recordati））；佐芬普利 + HCTZ；吡咯他尼（piretanide）+ 雷米普利；贝那普利 + HCTZ；氨氯地平（amlodipine）+ 贝那普利；莫西普利 + HCTZ；氨氯地平 + 培哚普利，赛维尔公司（Servier）；ASA + 阿托伐他汀（atorvastatin）+ 雷米普利 + 美托洛尔 ER（metoprolol ER）（扎杜斯卡地拉公司（Zydus Cadila））；雷米普利 + 氢氯噻嗪（hydrochlorothiazide）；(S)-氨氯地平 + 雷米普利（艾姆酷尔公司（Emcure））；喹那普利（quinapril）/氢氯噻嗪 | | | | | |
| ADAM9 | N | N | N | N | N | N |
| CDK6 | N | N | 帕博西里（palbociclib） | 阿昔迪布（alvocidib）；LY-2835219 | LEE-011 | N |
| IKBKB | N | N | N | N | N | EC-70124 |
| RARA | 他米巴罗汀 | N | N | IRX-5183 | N | N |
| LYN | 达沙替尼（dasatinib） | N | 尼达尼布（nintedanib） | 巴氟替尼（bafetinib） | JNJ-26483327 | Bcr-Abl/Lyn 抑制剂（AB 科学公司（AB Science）） |

[0820]

| 表 16 基因 | 已批准的 | 注册前（批准前） | III 期 | II 期 | I 期 | 临床前 |
|---|---|---|---|---|---|---|
| NTRK3 | N | N | N | N | N | PLX-7486 |
| ERBB2 | 曲妥珠单抗；曲妥珠单抗恩他新；帕妥珠单抗；二甲苯磺酸拉帕替尼；卡托莫西单抗；阿法替尼 | 曲妥珠单抗，因汉兹公司 | 来那替尼；XL-647；达可替尼；奈立派姆-S；曲妥珠单抗（赛尔群公司、拜沃卡德公司、必奥康公司、辛赛恩公司、哈维斯姆公司、阿约金公司） | 拉普亮赛-T；AVX-901；AE-37；BMS-690514；MVA-BN-HER2；瓦尼替尼；MM-111；AC-480；卵巢癌疫苗（尖锐公司）；（玛格土西单抗）；泊西替尼；PR-610 | Her-VAXX；VM-206；ARRY-380；JNJ-26483327；S-222611；多柔比星（莫瑞麦克公司）；西帕替尼；TrasGEX；曲妥珠单抗（韩华化学公司）；曲妥珠单抗（辉瑞公司）；IDN-6439 | 洛瓦新 B；TH-1（阿勒塔公司）；曲妥珠单抗-抗体结合物（辛赛恩公司）；CUDC-101；Her-2/neu 斯塔多体（格利克尼克公司）；ARX-788；Etbx-021；SN-34003；IBI-302；NT-004；ICT-140；ONS-1050；Sym-013；抗 HER2 X 抗 CD3（艾默金生物方案公司）；Z-650；乳癌疫苗（赛尔科学公司）；JNJ-28871063；曲妥珠单抗（普兰特弗姆公司、拜沃艾克斯普莱斯公司、双绿洲技术公司、斯塔达公司、纳特克公司、酷瑞克西公司、肿瘤生物公司、阿替沃金公司、玛比恩公司） |
| RHOA | N | N | N | N | N | N |
| RB1 | N | N | N | N | SGT-RB94 | N |
| THRA | N | N | N | N | N | N |
| CBL | N | N | N | N | N | N |

[0821]

| 表 16 基因 | 已批准的 | 注册前（批准前） | III 期 | II 期 | I 期 | 临床前 |
|---|---|---|---|---|---|---|
| ALK | 克卓替尼 | N | N | AP-26113；RG-7853；LDK-378；TSR-011；NMS-E628 | X-396；ASP-3026 | NMS-E628；极光激酶 + ALK 抑制剂（萨勒姆公司、阿斯利康公司）；ALK 抑制剂（阿斯利康公司、赛法隆公司、奥里基因公司）；ARN-5032；DLX-521 |

[0822] 表17：表3基因/融合体的可药化性状态

[0823]

| 基因 | 已批准的 | 注册前（批准前） | III 期 | II 期 | I 期 | 临床前 |
|---|---|---|---|---|---|---|
| ESR1 | 雌莫司汀磷酸钠（estramustine phosphate sodium）；乙炔基雌二醇磺酸盐（ethinyl estradiol sulfonate）；氟维司群（fulvestrant）；盐酸雷诺昔酚（raloxifene hydrochloride）；他莫昔芬（tamoxifen）；柠檬酸托瑞米芬（toremifene citrate）；曲洛司坦（trilostane）； | N | 阿考比芬（acolbifene） | TAS-108；雌四醇（estetrol）；GTx-758；内昔芬（endoxifen）；阿非昔芬（afimoxifene） | 淫羊藿素（icaritin）；ARN-810 | SR-16388；VAL-201；SERM + 毒素（SEEK）；雌二醇（BHR 医药公司（BHR Pharma））；NDC-1407；抗癌单抗（珅奥基公司（Shenogen）） |
| RPS6KB1 | N | N | N | N | AZD-5363；AT-13148；LY-S6KAKT1 | p70S6 激酶抑制剂（前哨公司（Sentinel）） |

[0824] 表19：基因融合体

[0825]

| 表 19 癌症类型 | 5'基因符号 | 3'基因符号 | 可药化基因 |
|---|---|---|---|
| 前列腺腺癌 | ABCD3 | DPYD | DPYD |
| 肉瘤 | ACTG2 | ALK | ALK |
| 肺腺癌 | ADAMTS16 | TERT | TERT |
| 脑低级别胶质瘤 | ATRX | BCL2 | BCL2 |

[0826]

| 表 19<br>癌症类型 | 5'基因符号 | 3'基因符号 | 可药化基因 |
|---|---|---|---|
| 胃腺癌 | B4GALT1 | RAF1 | RAF1 |
| 胃腺癌 | BRD3 | LCN2 | BRD3 |
| 胃腺癌 | CASZ1 | MTOR | MTOR |
| 急性骨髓性白血病 | CHD1 | MTOR | MTOR |
| 子宫体子宫内膜样癌 | CPA6 | PTK2 | PTK2 |
| 侵袭性乳癌 | DAB1 | IL12RB2 | IL12RB2 |
| 肺腺癌 | DDI2 | MTOR | MTOR |
| 肉瘤 | FRS2 | MDM2 | MDM2 |
| 肉瘤 | GLIS3 | TERT | TERT |
| 肺腺癌 | HIF1A | PRKCH | HIF1A |
| 侵袭性乳癌 | HPRT1 | CTPS2 | HPRT1 |
| 侵袭性乳癌 | IL12RB2 | DAB1 | IL12RB2 |
| 侵袭性乳癌 | IL6R | C1orf112 | IL6R |
| 侵袭性乳癌 | KCMF1 | PRKDC | PRKDC |
| 肺腺癌 | KIF5B | MET | MET |
| 侵袭性乳癌 | MAPK14 | EFHA1 | MAPK14 |
| 肉瘤 | MDM2 | SPATS2 | MDM2 |
| 甲状腺癌 | MTMR12 | TERT | TERT |
| 膀胱尿道上皮癌 | NOTCH2 | EIF2B3 | NOTCH2 |
| 肉瘤 | NTRK1 | DYNC2H1 | NTRK1 |
| 肾脏肾透明细胞癌 | PDCD6 | TERT | TERT |
| 肺腺癌 | PHKB | PDE3A | PDE3A |
| 子宫癌肉瘤 | RARA | SLC9A3R1 | RARA |
| 肝脏肝细胞癌 | SLC12A7 | TERT | TERT |
| 肉瘤 | SMARCA4 | EEF2 | EEF2 |
| 侵袭性乳癌 | STARD13 | TNFRSF8 | TNFRSF8 |
| 肺腺癌 | TICAM1 | IL12RB1 | IL12RB1 |
| 肉瘤 | TRIO | TERT | TERT |
| 前列腺腺癌 | TRPM8 | UGT1A9 | TRPM8 |
| 肉瘤 | TSPAN3 | MDM2 | MDM2 |
| 侵袭性乳癌 | TTLL7 | TERT | TERT |
| 脑低级别胶质瘤 | USP46 | PDGFRA | PDGFRA |
| 胃腺癌 | WNK2 | BRD3 | BRD3 |
| 宫颈鳞状细胞癌和宫颈内腺癌 | ZNF226 | AKT2 | AKT2 |

[0827]

**表 20. 来自表 19 的基因融合体的断裂点**

| 表 20<br>融合体名称 | 癌症类型 | TCGA 肿瘤样品条形码 | 5' 基因符号 | 5'登录号 | 5'染色体 | 5'断裂点 | 3'基因符号 | 3'登录号 | 3'染色体 | 3'断裂点 | 断裂点序列 | SEQ ID NO |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PDCD6\|TERT | 透明细胞肾细胞癌 | TCGA-BP-4991-01A-01R-1334-07 | PDCD6 | 10016 | chr5 | 272852 | TERT | 7015 | chr5 | 1282548 | TTCCTGTGGAAC<br>GTTTTCCAGAGG<br>GTCGATAAAGAC<br>AGGAGTGGAGTG<br>AT\|ATCAGACAGC<br>ACTTGAAGAGGG<br>TGCAGCTGCGGG<br>AGCTGTCGGAAG<br>CAGA | 200 |
| TSPAN3\|MDM2 | 肉瘤 | TCGA-DX-A23R-01A-11R-A26T-07 | TSPAN3 | 10099 | chr15 | 77344775 | MDM2 | 4193 | chr12 | 69202269 | ACCTCTATGCTG<br>AGGGGTGTGAGG<br>CTCTAGTAGTGA<br>AGAAGCTACAAG<br>AA\|CAGGCAAAT<br>GTGCAATACCAA<br>CATGTCTGTACCT<br>ACTGATGGTGCT<br>GTAA | 201 |
| SLC12A7\|TERT | 肝细胞癌 | TCGA-BC-A3KG-01A-11R-A213-07 | SLC12A7 | 10723 | chr5 | 1111983 | TERT | 7015 | chr5 | 1282739 | CGGAGGCTCCGG<br>GCACCCCCGAGG<br>GCCCCGAGCCCG<br>AGCGCCCCAGCC<br>CG\|GGGGTTGGCT<br>GTGTTCCGGCCG<br>CAGAGCACCGTC<br>TGCGTGAGGAGA<br>TCCT | 202 |
| FRS2\|MDM2 | 肉瘤 | TCGA-DX-A3M1-01A-11R-A22K-07 | FRS2 | 10818 | chr12 | 69864310 | MDM2 | 4193 | chr12 | 69202988 | GTGGTTACAGCA<br>CCATCAGTAGGT<br>ACAGACATGTTG<br>GTATTGCACATTT<br>G\|CCGTCCGCCCA<br>GGTGCTGAGAGG<br>GAGCAGGGCGC | 203 |

[0828]

| 表 20 融合体名称 | 癌症类型 | TCGA 肿瘤样品条形码 | 5' 基因符号 | 5'登录号 | 5'染色体 | 5'断裂点 | 3'基因符号 | 3'登录号 | 3'染色体 | 3'断裂点 | 断裂点序列 | SEQ ID NO |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | GGGTCGGCGGGC GCGA | |
| CHD1\|MTOR | 急性骨髓性白血病 | TCGA-AB-2939-03A-01T-0740-13 | CHD1 | 1105 | chr5 | 98199112 | MTOR | 2475 | chr1 | 11273623 | GAATGTCTAAAA GAGTATACAAAT CCTGAACAAATT AAGCAATGGAGA AA\|GAATTCTGGG TCATGAACACCT CAATTCAGAGCA CGATCATTCTTCT CAT | 204 |
| CHD1\|MTOR | 急性骨髓性白血病 | TCGA-AB-2939-03A-01T-0740-13 | CHD1 | 1105 | chr5 | 98204199 | MTOR | 2475 | chr1 | 11273623 | TTCCCATTTTCTGA AGAATCTGAAGA GCTGGATCAGAA GACATTCAGCAT T\|GAATTCTGGGT CATGAACACCTC AATTCAGAGCAC GATCATTCTTCTC AT | 205 |
| CHD1\|MTOR | 急性骨髓性白血病 | TCGA-AB-2939-03A-01T-0740-13 | CHD1 | 1105 | chr5 | 98199112 | MTOR | 2475 | chr1 | 11273623 | AATGAGAAGAAT GATCGTGCTCTG AATTGAGGTGTT CATGACCCAGAA TT\|CTTTCTCCATT GCTTAATTTGTTC AGGATTTGTATAC TCTTTTAGACATT | 206 |
| CHD1\|MTOR | 侵袭性乳癌 | TCGA-AO-A1I29-01A-21R-A10J-07 | MAPK14 | 1432 | chr6 | 36044379 | EFHA1 | 221154 | chr13 | 22113824 | GGGATGCATAAT GGCCGAGCTGTT GACTGGAAGAAC ATTGTTTCCTGGT A\|AAACTTCAGTC AAGAAGCTGACA AAAAAGGACATC GAGGATACACTG | 207 |

[0829]

| 表 20 融合体名称 | 癌症类型 | TCGA 肿瘤样品条形码 | 5' 基因符号 | 5'登录号 | 5'染色体 | 5'断裂点 | 3'基因符号 | 3'登录号 | 3'染色体 | 3'断裂点 | 断裂点序列 | SEQ ID NO |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | TCA | |
| TICAM1\|IL12RB1 | 肺腺癌 | TCGA-05-4426-01A-01R-1206-07 | TICAM1 | 148022 | chr19 | 48316036 | IL12RB1 | 3594 | chr19 | 18180463 | GTCCTGGCCCAC<br>AGGCTGCCATTC<br>AATGCAATACGT<br>CATGCTCTGAGC<br>CC\|GGGCTGCCG<br>GCTGCGCCACTG<br>GGTCCTGGGGTC<br>CTGGGGGCTGGG<br>GCTTC | 208 |
| TICAM1\|IL12RB1 | 肺腺癌 | TCGA-05-4426-01A-01R-1206-07 | TICAM1 | 148022 | chr19 | 48316030 | IL12RB1 | 3594 | chr19 | 18182962 | CCACTGGTTCTG<br>TGTGGGTGTCGG<br>CAGGAATGTGCC<br>ACGTCTGGTTCA<br>GG\|GATCCGGGG<br>CTGCCGGCTGCG<br>CCACTGGGTCCT<br>GGGGTCCTGGGG<br>GCTGG | 209 |
| DAB1\|IL12RB2 | 侵袭性乳癌 | TCGA-AN-A0AM-01A-11R-A034-07 | DAB1 | 1600 | chr1 | 57611102 | IL12RB2 | 3595 | chr1 | 67845789 | CCCTTCACCTTTA<br>AACCTCTTTATCA<br>AAGTGGCTTCAC<br>TGCGATCCTGAC\|<br>GGGAATTTTGTC<br>TGCAAGGTGAGA<br>GGCAGTGTTAAG<br>GATGATGAGTCC<br>AC | 210 |
| IL12RB2\|DAB1 | 侵袭性乳癌 | TCGA-AN-A0AM-01A-11R-A034-07 | IL12RB2 | 3595 | chr1 | 67845806 | DAB1 | 1600 | chr1 | 57611102 | CTGCTGGTGAAA<br>GTTCCCACGGAA<br>ATGAGAGGGAAT<br>TTTGTCTGCAAG<br>GT\|CAGGATCGCA<br>GTGAAGCCACTT<br>TGATAAAGAGGT<br>TTAAAGGTGAAG | 211 |

[0830]

| 表 20 融合体名称 | 癌症类型 | TCGA 肿瘤样品条形码 | 5' 基因符号 | 5'登录号 | 5'染色体 | 5'断裂点 | 3'基因符号 | 3'登录号 | 3'染色体 | 3'断裂点 | 断裂点序列 | SEQ ID NO |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | GGGT | |
| IL12RB2\|DAB1 | 侵袭性乳癌 | TCGA-AN-A0AM-01A-11R-A034-07 | IL12RB2 | 3595 | chr1 | 67845733 | DAB1 | 1600 | chr1 | 57611052 | TCTCCCAAAATTCACATCCAATAAACAGCCTGCAGCCCCGAGTGACATAT\|GTCCGGTACAAAGCCAAATTGATCGGGATTGATGAAGTTTCCGCAGCTCG | 212 |
| GLIS3\|TERT | 肉瘤 | TCGA-DX-A3LS-01A-11R-A21T-07 | GLIS3 | 169792 | chr9 | 41177068 | TERT | 7015 | chr5 | 12827839 | CTGCTGATCCACATGAGAGTCCACTCTGGGGAGAAGCCCAACAAGTGTAC\|GGGGTTGGCTGTGTTCCGGCCGCAGAGCACCGTCTGCGTGAGGAGATCCT | 213 |
| ADAMTS16\|TERT | 肺腺癌 | TCGA-44-2662-01A-01R-0946-07 | ADAMTS16 | 170690 | chr5 | 51919 03 | TERT | 7015 | chr5 | 12827 39 | GATACAGGTCTTGGACTGGCCTTCACCATTGCCCATGAGTCTGGACAC AA\|GGGTTGGCTGTGTTCCGGCCGCAGAGCACCGTCTGCGTGAGGAGATCCTG | 214 |
| ABCD3\|DPYD | 前列腺腺癌 | TCGA-CH-5764-01A-21R-1580-07 | ABCD3 | 5825 | chr1 | 94956803 | DPYD | 1806 | chr1 | 97981497 | CTTTAGCAACGCCAAATGGAGATGTTTTGATCCGAGACCTTAATTTTGA A\|TCACAATATGGAGCTTCCGTTTCTGCCAAGCCTGAACTACCCCCTCTTT | 215 |

[0831]

| 表 20 融合体名称 | 癌症类型 | TCGA 肿瘤样品条形码 | 5' 基因符号 | 5'登录号 | 5'染色体 | 5'断裂点 | 3'基因符号 | 3'登录号 | 3'染色体 | 3'断裂点 | 断裂点序列 | SEQ ID NO |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | TA | |
| SMARCA4\|EEF2 | 肉瘤 | TCGA-K1-A3PO-01A-11R-A21T-07 | SMARCA4 | 6597 | chr19 | 11151982 | EEF2 | 1938 | chr19 | 3983208 | TCTGCCGGACCTCCTCTTCGATCTCCTCCAGCGTGCCCTCCTCGATGGCC\|CAACCTCATTGACTCCCCCGGGCATGTCGACTTCTCCTCGGAGGTGACTG | 216 |
| ZNF226\|AKT2 | 宫颈鳞状细胞癌 | TCGA-IR-A3LH-01A-21R-A213-07 | ZNF226 | 7769 | chr19 | 44669953 | AKT2 | 208 | chr19 | 4074852 9 | ATTCAGCCCTGACTTCTCAAAAAGCACTGCACAGAGGAGGAGGCAGC\|ACCCCATGGACTACAAGTGTGGCTCCCCCAGTGACTCCTCCACGACTGAG | 217 |
| ZNF226\|AKT2 | 宫颈鳞状细胞癌 | TCGA-IR-A3LH-01A-21R-A213-07 | ZNF226 | 7769 | chr19 | 44669953 | AKT2 | 208 | chr19 | 4074852 9 | AATTCTCCCTGACTTCTCAAAAAGCACTGCACAGAGGAGGAGGCAGC\|ACCCCATGGACTACAAGTGTGGCTCCCCCAGTGACTCCTCCACGACTGAG | 218 |
| ACTG2\|ALK | 肉瘤 | TCGA-IW-A3M6-01A-11R-A21T-07 | ACTG2 | 72 | chr2 | 74128558 | ALK | 238 | chr2 | 2944638 0 | GAGATGATGCCCCCCGGGCTGTCTTCCCCTCCATTGTGGGCCGCCCTCGC\|CACCAGGAGCTGCAAGCCATGCAGATGGAGCTGCAGAGCCCTGAGT | 219 |

[0832]

| 表 20 融合体名称 | 癌症类型 | TCGA 肿瘤样品条形码 | 5' 基因符号 | 5'登录号 | 5'染色体 | 5'断裂点 | 3'基因符号 | 3'登录号 | 3'染色体 | 3'断裂点 | 断裂点序列 | SEQ ID NO |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | ACAA | |
| ACTG2\|ALK | 肉瘤 | TCGA-IW-A3M6-01A-11R-A21T-07 | ACTG2 | 72 | chr2 | 74128564 | ALK | 238 | chr2 | 2944994 0 | ATGCCCCCCGGGCTGTCTTCCCCTCCATTGTGGGCCGCCCTCGCCACCAG\|TGATGGAAGGCCACGGGGAAGTGAATATTAAGCATTATCTAAACTGCAGT | 220 |
| ACTG2\|ALK | 肉瘤 | TCGA-IW-A3M5-01A-22R-A21T-07 | ACTG2 | 72 | chr2 | 74128564 | ALK | 238 | chr2 | 2944994 0 | TGATGCCCCCCGGGCTGTCTTCCCCTCCATTGTGGGCCGCCCTCGCCACCAG\|AGTGATGGAAGGCCACGGGGAAGTGAATATTAAGCATTATCTAAACTGCA | 221 |
| CASZ1\|MTOR | 胃腺癌 | TCGA-BR-8590-01A-11R-2402-13 | CASZ1 | 54897 | chr1 | 10765549 | MTOR | 2475 | chr1 | 1128897 5 | ATGAAGTGACACCCCCAGCTACATCCGAGGAGGTTCTAGGACCTGCTACG\|AGCTGACTATAGCACTAGTGAAATGCTGGTCAACATGGGAAACTTGCCTC | 222 |
| DDI2\|MTOR | 肺腺癌 | TCGA-MP-A4SW-01A-21R-A24X-07 | DDI2 | 84301 | chr1 | 15944303 | MTOR | 2475 | chr1 | 1122757 4 | ATTCTAACACTCCGGCCGCTGCCTCCGGCTGCTGTAGCTTATTATTAATG\|CTGGCTCTCGGCTGCGGGGATGCCAGACTCGAGCTCGCACAGCGCGC | 223 |

[0833]

| 表 20 融合体名称 | 癌症类型 | TCGA 肿瘤样品条形码 | 5' 基因符号 | 5'登录号 | 5'染色体 | 5'断裂点 | 3'基因符号 | 3'登录号 | 3'染色体 | 3'断裂点 | 断裂点序列 | SEQ ID NO |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | GGA | |
| B4GALT1\|RAF1 | 胃腺癌 | TCGA-HU-A4GH-01A-11R-A24K-31 | B4GALT1 | 2683 | chr9 | 33166756 | RAF1 | 5894 | chr3 | 12641914 | CTGGACAGGGCTGAAGGTGAGGCTGATTCGCTGTGACTTCGAATTGCATC\|CAAGCAGCGGGACTCCTCAGGGCAGGCGGGCAGCGACAGTGCGGTGGTG | 224 |
| HIF1A\|PRKCH | 肺腺癌 | TCGA-44-2668-01A-01R-0946-07 | HIF1A | 3091 | chr14 | 62207906 | PRKCH | 5583 | chr14 | 61995793 | AAAAATCTCATCCAAGAAGCCCTAACGTGTTATCTGTCGCTTTGAGTCA\|AGAGATCTGAAACTGGACAATGTCCTGTTGGACCACGAGGGTCACTGTAA | 225 |
| HIF1A\|PRKCH | 肺腺癌 | TCGA-44-2668-01A-01R-0946-07 | HIF1A | 3091 | chr14 | 62207766 | PRKCH | 5583 | chr14 | 61995805 | CGAAGTCTGCCAGTTTACAGTGACCCTCGTGGTCCAACAGGACATTGTCC\|AGTTTCTTTATGTATGTGGGTAGGAGATGGAGATGCAATCAATATTTTAA | 226 |
| HPRT1\|CTPS2 | 侵袭性乳癌 | TCGA-AR-A24W-01A-11R-A169-07 | HPRT1 | 3251 | chrX | 13362754 2 | CTPS2 | 56474 | chrX | 16657355 | GATGATCTCTCAACTTTAACTGGAAAGTCTAGGTTGTTGGCAGAAGATAT\|GCCCGAGCACAACCCTGGCAATTTGGGAGGAACAATGAGACTGGGA | 227 |

[0834]

| 表 20 融合体名称 | 癌症类型 | TCGA 肿瘤样品条形码 | 5' 基因符号 | 5'登录号 | 5'染色体 | 5'断裂点 | 3'基因符号 | 3'登录号 | 3'染色体 | 3'断裂点 | 断裂点序列 | SEQ ID NO |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | ATAA | |
| HPRT1\|CTPS2 | 侵袭性乳癌 | TCGA-AR-A2 4W-01A-11R-A169-07 | HPRT1 | 3251 | chrX | 133609340 | CTPS2 | 56474 | chrX | 16685822 | ATAAATTCTTTGC TGACCTGCTGGA TTACATCAAAGC ACTGAATAGAAA T\|AGTGATAGAGT TTGCAAGAAACT GCCTTAACTTGA AAGATGCTGATT CCA | 228 |
| HPRT1\|CTPS2 | 侵袭性乳癌 | TCGA-AR-A2 4W-01A-11R-A169-07 | HPRT1 | 3251 | chrX | 133609375 | CTPS2 | 56474 | chrX | 16638444 | GCACTGAATAGA AATAGTGATAGAT CCATTCCTATGAC TGTAGATTTTAT\|G GTGATGTTCCTTT TATAGAAGAAAG ACACAGACATCG GTTCGAGGTAAA | 229 |
| HPRT1\|CTPS2 | 侵袭性乳癌 | TCGA-AR-A2 4W-01A-11R-A169-07 | HPRT1 | 3251 | chrX | 133627542 | CTPS2 | 56474 | chrX | 16657355 | GATGATCTCTCA ACTTTAACTGGA AAGAATGTCTTG ATTGTGGAAGAT AT\|GCCCGAGCAC AACCCTGGCAAT TTGGGAGGAACA ATGAGACTGGGA ATAA | 230 |
| HPRT1\|CTPS2 | 侵袭性乳癌 | TCGA-AR-A2 4W-01A-11R-A169-07 | HPRT1 | 3251 | chrX | 133609363 | CTPS2 | 56474 | chrX | 16685820 | GATTACATCAAA GCACTGAATAGA AATAGTGATAGAT CCATTCCTATGAC \|TGATAGAGTTTG CAAGAAACTGCC TTAACTTGAAAG ATGCTGATTCCA CA | 231 |

[0835]

| 表 20 融合体名称 | 癌症类型 | TCGA 肿瘤样品条形码 | 5' 基因符号 | 5'登录号 | 5'染色体 | 5'断裂点 | 3'基因符号 | 3'登录号 | 3'染色体 | 3'断裂点 | 断裂点序列 | SEQ ID NO |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| IL6R\|C1orf112 | 侵袭性乳癌 | TCGA-E9-A1RF-01A-11R-A157-07 | IL6R | 3570 | chr1 | 154420647 | C1orf112 | 55732 | chr1 | 169790820 | GGACAGAATCCAGGAGTCCTCCAGCTGAGAACGAGGTGTCCACCCCCCATG\|CAGGATAATGCTGACTACAGATTATTTCAGAAAACACTCAAATTGTGTCG | 232 |
| KIF5B\|MET | 肺腺癌 | TCGA-93-A4JN-01A-11R-A24X-07 | KIF5B | 3799 | chr10 | 32304500 | MET | 4233 | chr7 | 116411617 | CCAACTCACCCAAGTGCAATTCGTGGAGGAGGTGCATTTGTTCAGAACAG\|AGGATTGATTGCTGGTGTTGTCTCAATATCAACAGCACTGTTATTACTAC | 233 |
| KIF5B\|MET | 肺腺癌 | TCGA-93-A4JN-01A-11R-A24X-07 | KIF5B | 3799 | chr10 | 32306145 | MET | 4233 | chr7 | 116411932 | GCACTGAAAGAAGCTAAAGAAAATGCATCTCGTGATCGCAAACGCTATCA\|GCAAGAGTACACACTCCTCATTTGGATAGGCTTGTAAGTGCCCGAAGTGT | 234 |
| BRD3\|LCN2 | 胃腺癌 | TCGA-HU-A4H2-01A-11R-A251-31 | BRD3 | 8019 | chr9 | 136917428 | LCN2 | 3934 | chr9 | 130912517 | GTATGCAGGACTTCAACACCATGTTTACAAATTGTTACATTTATAACAAG\|TTCCAGGGGAAGTGGTATGTGGTAGGCCTGGCAGGGAATGCAATTCTCAG | 235 |

[0836]

| 表 20 融合体名称 | 癌症类型 | TCGA 肿瘤样品条形码 | 5'基因符号 | 5'登录号 | 5'染色体 | 5'断裂点 | 3'基因符号 | 3'登录号 | 3'染色体 | 3'断裂点 | 断裂点序列 | SEQ ID NO |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MDM2\|SPATS2 | 肉瘤 | TCGA-DX-A1KZ-01A-11R-A24X-07 | MDM2 | 4193 | chr12 | 69233549 | SPATS2 | 65244 | chr12 | 49883267 | CATTGTCCATGGCAAAACAGGACATCTTATGGCCTGCTTTACATGTGCAA\|TAGTTCCTAATAAGAGCAACAATGAAATTATCCTGGTTTTGCAGCACTTT | 236 |
| NOTCH2\|EIF2B3 | 膀胱尿道上皮癌 | TCGA-FD-A5BS-01A-21R-A26T-07 | NOTCH2 | 4853 | chr1 | 12045 8963 | EIF2B3 | 8891 | chr1 | 45392411 | CATGCCTACTAGCCTCCCTAACCTTGCCAAGGAGGCAAAAGGATGCCAAGG\|TGGAGCAGCGTGACTTCATTGGAGTGGACAGCACAGGAAAGAGGCTGCTC | 237 |
| NTRK1\|DYNC2H1 | 肉瘤 | TCGA-DX-A3LY-01B-11R-A27Q-07 | NTRK1 | 4914 | chr1 | 15685 1401 | DYNC2H1 | 79659 | chr11 | 10330 6708 | AACGCCACAGCATCAAGGATGTGCACGCCCGGCTGCAAGCCCTGGCCCAG\|AAGATCCCTTACAATACCTGAGAGGTCTTGTTGCCCGTGCCCTTGCAATA | 238 |
| PHKB\|PDE3A | 肺腺癌 | TCGA-MN-A4N5-01A-11R-A24X-07 | PHKB | 5257 | chr16 | 47723028 | PDE3A | 5139 | chr12 | 20799464 | ACTTCAGATCCGTGGCGGAGACAAGCCAGCCTTGGACTTGTATCAGCTGT\|TGGTATCTTACTACACAGCCTATTCCAGGCCTCTCAACTGTGATTAATG | 239 |

[0837]

| 表 20 融合体名称 | 癌症类型 | TCGA 肿瘤样品条形码 | 5'基因符号 | 5'登录号 | 5'染色体 | 5'断裂点 | 3'基因符号 | 3'登录号 | 3'染色体 | 3'断裂点 | 断裂点序列 | SEQ ID NO |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| USP46\|PDGFRA | 低级别胶质瘤 | TCGA-CS-6665-01A-11R-1896-07 | USP46 | 64854 | chr4 | 53522650 | PDGFRA | 5156 | chr4 | 5514357 6 | GTGGAAGCAACC ACTAATATAAAC ACCTCCCATGTAT AGGAAGGCTGG AG\|CGTTTGGGA AGGTGGTTGAAG GAACAGCCTATG GATTAAGCCGGT CCCAA | 240 |
| USP46\|PDGFRA | 低级别胶质瘤 | TCGA-CS-6665-01A-11R-1896-07 | USP46 | 64854 | chr4 | 53494288 | PDGFRA | 5156 | chr4 | 5514077 1 | GGTCAATTTTGG AAACACATGCTA CTGTAACTCCGT GCTTCAGGCATT GT\|CCTGGTTGTC ATTTGGAAACAG AAACCGAGGTAT GAAATTCGCTGG AGGG | 241 |
| MTMR12\|TERT | 甲状腺腺体癌 | TCGA-BJ-A4O9-01A-11R-A250-07 | MTMR12 | 54545 | chr5 | 32263219 | TERT | 7015 | chr5 | 1282739 | ACATGAAGTACA AAGCAGTGAGTG TCAACGAAGGCT ATAAAGTCTGTG AG\|AGGGGTTGG CTGTGTTCCGGC CGCAGAGCACCG TCTGCGTGAGGA GATCC | 242 |
| ATRX\|BCL2 | 低级别胶质瘤 | TCGA-DB-A4XF-01A-11R-A27Q-07 | ATRX | 546 | chrX | 77041468 | BCL2 | 596 | chr18 | 6079599 2 | AATCAAACAGAG GCCGCATGCTGG GGCCGTACAGTT CCACAAAGGCAT CC\|TCATGGGCTC AGCGGTCATGTT TTCGCTTGAACG CCTTGTCGGCTT CTGT | 243 |

[0838]

| 表 20 融合体名称 | 癌症类型 | TCGA 肿瘤样品条形码 | 5'基因符号 | 5'登录号 | 5'染色体 | 5'断裂点 | 3'基因符号 | 3'登录号 | 3'染色体 | 3'断裂点 | 断裂点序列 | SEQ ID NO |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TRPM8\|UGT1A9 | 前列腺腺癌 | TCGA-CH-5766-01A-11R-1580-07 | TRPM8 | 79054 | chr2 | 23489 4509 | UGT1A9 | 54600 | chr2 | 23467 5680 | CATGTTATCCACC AACATCCTGCTG GTCAACCTGCTG GTCGCCATGTTT G\|GGAATTTGAAG CCTACATTAATGC TTCTGGAGAACA TGGAATTGTGGT TT | 244 |
| KCMF1\|PRKDC | 侵袭性乳癌 | TCGA-EW-A1P4-01A-21R-A144-07 | KCMF1 | 56888 | chr2 | 85262 227 | PRKDC | 5591 | chr8 | 48772 278 | CACAGTCTTTTA CTTGTCCCTATTG TGGAAAAATGGG CTATACGGAGAC A\|GTACCCTGAGT GAGGAAATGAGT CAATTTGATTTCT CAACCGGAGTTC AG | 245 |
| CPA6\|PTK2 | 子宫内膜子宫内膜样腺癌 | TCGA-A5-A0G5-01A-11R-A040-07 | CPA6 | 57094 | chr8 | 68536 411 | PTK2 | 5747 | chr8 | 14177 4389 | AAACAGAAGAG GAAGCATATGCA CTGAAGAAAATA TCCTATCAACTTA AG\|AAACAGATG ATTATGCTGAGAT TATAGATGAAGA AGATACTTACAC CATG | 246 |
| RARA\|SLC9A3R1 | 癌肉瘤 | TCGA-N8-A4PQ-01A-11R-A28V-07 | RARA | 5914 | chr17 | 38508 759 | SLC9A3R1 | 9368 | chr17 | 72758 151 | ACCATCGCCGAC CAGATCACCCTC CTCAAGGCTGCC TGCCTGGACATC CT\|GCGCGAGCTT CGGCCTCGGCTC TGTACCATGAAG AAGGGCCCCAGT GGCT | 247 |

[0839]

| 表 20 融合体名称 | 癌症类型 | TCGA 肿瘤样品条形码 | 5'基因符号 | 5'登录号 | 5'染色体 | 5'断裂点 | 3'基因符号 | 3'登录号 | 3'染色体 | 3'断裂点 | 断裂点序列 | SEQ ID NO |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| WNK2\|BRD3 | 胃腺癌 | TCGA-HU-A4H2-01A-11R-A251-31 | WNK2 | 65268 | chr9 | 95947892 | BRD3 | 8019 | chr9 | 136910543 | ACAAGGGGCTGGACACGGAGACCTGGGTGGAGGTGGCCTGGTGTGAGCTG\|CAGAGGAAGATGGATGGCCGAGAGTACCCAGACGCACAGGGCTTTGCTGC | 248 |
| TRIO\|TERT | 肉瘤 | TCGA-DX-A1L3-01A-11R-A24X-07 | TRIO | 7204 | chr5 | 14420130 | TERT | 7015 | chr5 | 1282739 | ATCGCCCACTCCAGAAGTAGCATGGAAATGGAGGGCATCTTCAACCACAA\|AGGGGTTGGCTGTGTTCCGGCCGCAGAGCACCGTCTGCGTGAGGAGATCC | 249 |
| TTLL7\|TERT | 侵袭性乳癌 | TCGA-C8-A131-01A-11R-A115-07 | TTLL7 | 79739 | chr1 | 84464614 | TERT | 7015 | chr5 | 1282739 | CCGCTTGCAGCGGGGACGCGAGGACCCCGGGCTGGGCTTTCCTCACCCGGG\|GTTGGCTGTGTTCCGGCCGCAGAGCACCGTCTGCGTGAGGAGATCCTGG | 250 |
| STARD13\|TNFRSF8 | 侵袭性乳癌 | TCGA-BH-A0C7-01B-11R-A115-07 | STARD13 | 90627 | chr13 | 33859649 | TNFRSF8 | 943 | chr1 | 12164568 | CTCACAGACCGTGTTCTTCTGCGCCGTGCCTGGGAACTTGACAATCATCC\|GGCTCATCCTGTAAGGAGAGCGTCTTGTAGTCTGATCAAATCGCAAGTAC | 251 |

[0840]

**表 21. 表 19 基因/融合体的可药化性状态**

| 表 21 基因 | 已批准的 | 注册前（批准前） | III 期 | II 期 | I 期 | 临床前 |
|---|---|---|---|---|---|---|
| AKT2 | N | N | N | N | ARQ-092；BAY-1125976 | RX-1792；NT-113；TAS-117 |
| ALK | 克卓替尼 | N | N | AP-26113；RG-7853；LDK-378；TSR-011；NMS-E628 | X-396；ASP-3026； | NMS-E628；极光激酶+ALK 抑制剂（萨勒姆公司、阿斯利康公司）；ALK 抑制剂（阿斯利康公司、赛法隆公司、奥里基因公司）；ARN-5032；DLX-521 |
| BCL2 | N | N | N | PBI-1402；PNT-2258；R-(-)-棉籽醇；纳维克拉斯（navitoclax）；RG-7601 | N | VAL-101；BP-100-1.02；萨布托克雷斯（sabutoclax） |
| BRD3 | N | N | N | N | Y-803 | N |
| DPYD | N | N | N | 恩尿嘧啶（eniluracil） | TAS-114 | N |
| EEF2 | 地尼白介素（denileukin diftitox） | N | 莫昔土莫单抗帕苏多托（moxetumomab pasudotox） | 辛曲德克贝苏多托（cintredekin besudotox） | N | 成神经胶质细胞（Glioblast）-13 |
| FGFR3 | 帕纳替尼 | 阿西替尼（asitinib） | 乐伐替尼（lenvatinib） | 多韦替尼乳酸盐；ENMD-2076；AZD-4547 | JNJ-42756493；BGJ-398；LY-2874455；S-49076 | N |
| HIF1A | | | | 喜树碱，卡兰多公司 | 2-甲氧基雌二醇；SPC-2968 | RX-0047；ATSP-9172；ATSP-9172；P-3971 |
| HPRT1 | 甘氨硫嘌呤 | N | N | N | N | N |

[0841]

| 表 21<br>基因 | 已批准的 | 注册前（批准前） | III 期 | II 期 | I 期 | 临床前 |
|---|---|---|---|---|---|---|
| | （butocin） | | | | | |
| IL12RB1 | N | N | N | INXN-2001/1001 ；IL-12 | AS-1409；NHS-IL12 | N |
| IL12RB2 | N | N | N | IL-12 | NHS-IL-12；AS-1409 | N |
| IL6R | 托西利单抗（tocilizumab） | N | ARRY-438162 | 吉韦诺他（ givinostat ）；ALX-0061 | | L-6 抑制剂，因特普罗廷公司；IL-6 拮抗剂，普塔高尼斯特治疗公司（ Protagonist Therapeutics ）；APX-007 |
| MAPK14 | 吡非尼酮（pirfenidone） | N | N | 拉尼米替尼（ralimetinib） | ARRY-614；硫脲基丁腈 | N |
| MDM2 | N | N | N | N | SAR-405838；RG-7388；RO-5503781 ；CGM-097；DS-3032 | p53-mdm2/mdm4 双重抑制剂，阿戴姆德公司（ Adamed ）；PXN-527 ；ATSP-7041 ；MDM2 抑制剂，安进公司 |
| MET | 卡博替尼；克卓替尼 | N | 提瓦替尼；里乐木单抗；奥那组单抗； | MGCD-265 ；弗雷替尼；费拉妥珠单抗；BMS-777607；格瓦替尼 ；INCB-028060 ；LY-2875358 ；阿皮托利西（apitolisib） | AMG-208；TAS-115；沃利替尼；SAR-125844 ；S-49076 | X-379 ；麦他替尼；PRS-110 ；ASP-08001 ；ARGX-111 ；DCC-2701 ；DCC-2721 ；MG-516 ；AL-2846 ；CG-206481 ；T-1840383 ；cMet-EGFR 双重抑制剂（水晶基因组公 |

[0842]

| 表 21<br>基因 | 已批准的 | 注册前（批准前） | III 期 | II 期 | I 期 | 临床前 |
|---|---|---|---|---|---|---|
| | | | | | | 司）；双特异性抗体（霍夫曼-拉罗奇公司） |
| MTOR | 依维莫司（everolimus）；坦罗莫司（temsirolimus） | 地磷莫司（ridaforolimus） | N | 奎纳克林；XL-765；达妥昔布（dactolisib）；PKI-587；PF-04691502；CC-223 | P-7170；CBLC-137，INK-128，AZD-2014；CC-115；PWT-33957；DS-7423；GDC-0084；DS-3078；LY-3023414；PI3 激酶/mTOR 抑制剂，礼来公司 | nPT-MTOR；SB2343；STP-503；X-480；ABTL-0812；X-414；CC214；HMPL-518；PQR-309；PQR-401；mTOR 抑制剂/PI3 激酶抑制剂，Lilly-1；PIM/PI3k/mTOR 抑制剂，转调生物科学公司（Inflection Biosciences） |
| NOTCH2 | N | N | N | OMP-59R5 | N | N |
| NTRK1 | N | N | N | 马来酸米西西尼 | N | 酪氨酸激酶抑制剂（布里斯托尔-迈尔斯·斯奎布公司）；PLX-7486 |
| PDE3A | 氨利酮（amrinone）；盐酸阿那格雷（anagrelide hydrochloride）；依诺昔酮（enoximone）；西洛他唑（cilostazol）；盐酸 | N | N | 盐酸帕罗格列（parogrelil hydrochloride）；K-134；RPL-554；西洛他唑，健亚公司（Genovate） | CR-3465 | CLC-2001 |

[0843]

| 表 21<br>基因 | 已批准的 | 注册前（批准前） | III 期 | II 期 | I 期 | 临床前 |
|---|---|---|---|---|---|---|
| | 洛普力农（loprinone hydrochloride）；盐酸洛普力农；盐酸洛普力农 | | | | | |
| PDGFRA | 甲磺酸伊马替尼（imatinib mesilate）；帕唑帕尼；舒尼替尼、达沙替尼；尼罗替尼（nilotinib）；瑞格非尼 | 尼达尼布（nintedanib） | 奥兰替尼（orantinib）；莫替沙尼（motesanib）；立法尼（linifanib） | ENMD-2076；奥拉单抗（olaratumab）；X-82；克诺拉尼（crenolanib）； | N | DCC-2618；CG-206481 |
| PRKDC | N | N | 沃萨若新（vosaroxin） | N | SF-1126，迪贝特（Dbait）；CC-115 | N |
| PTK2 | N | N | N | 迪法西替尼（defactinib） | GSK-2256098；CEP-37440；BI-853520；VS-4718 | CFAK-C4；FAK 抑制剂，维拉斯泰姆公司；CTX-0294945；x-0294886 |
| RAF1 | 索拉非尼 | N | N | iCo-007；XL-281 | RO-5126766；MLN-2480 | BIB-024；STP503；DP-4978；HM-95573；TAK-632 |
| RARA | 他米巴罗汀 | N | N | IRX-5183 | N | N |
| TERT | N | N | GV-1001 | VX-001；GX-301- | TeloB-Vax | 端粒酶疫苗，杰龙公司（Geron）；hTERT DNA 疫苗，依诺维欧公司（Inovio） |

[0844]

| 表 21<br>基因 | 已批准的 | 注册前（批准前） | III 期 | II 期 | I 期 | 临床前 |
|---|---|---|---|---|---|---|
| TNFRSF8 | 贝伦妥单抗维多汀（ brentuximab vedotin） | N | N | AFM-13；XmAb-2513 | N | N |
| TRPM8 | N | N | N | N | D-3263 | N |

**表 22. 与基因融合体新相关的癌症类型**

| 癌症类型 | 基因 A | 基因 B | 定向（5'/3'） | 可药化基因 | 癌症类型先例 |
|---|---|---|---|---|---|
| 乳头状肾细胞癌 | FGFR3 | TACC3 | FGFR3/TACC3 | FGFR3 | 膀胱癌；鳞状细胞肺癌；成胶质细胞瘤；头颈鳞状细胞癌；宫颈鳞状细胞癌；低级别胶质瘤 |
| 鳞状细胞肺癌 | SEC16A | NOTCH1 | SEC16A/NOTCH1 | NOTCH1 | 乳癌；甲状腺腺体癌 |

**表 23. 来自表 22 的基因融合体的断裂点**

| 表 23<br>融合体名称 | 癌症类型 | TCGA 肿瘤样品条形码 | 5'基因符号 | 5'登录号 | 5'染色体 | 5'断裂点 | 3'基因符号 | 3'登录号 | 3'染色体 | 3'断裂点 | 断裂点序列 | SEQ ID NO |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FGFR3\|TACC3 | 乳头状肾细胞癌 | TCGA-A4-7287-01A-11R-2139-07 | FGFR3 | 2261 | chr4 | 1808661 | TACC3 | 10460 | chr4 | 1741429 | TCCTCACACC<br>TGCTCCTCAG<br>CTCCCGGTTC<br>TCCTCCTGTGT<br>CGCCTTTAC\|G<br>TCGGTGGACG<br>TCACGGTAAG<br>GACACGGTCC<br>AGGTCCTCCA<br>CCAGCTGCT | 252 |
| FGFR3\|TACC3 | 乳头状肾细胞癌 | TCGA-A4-7287-01A-11R-2139-07 | FGFR3 | 2261 | chr4 | 1808633 | TACC3 | 10460 | chr4 | 1741500 | GCCGCGCCCT<br>CCCAGAGGCC<br>CACCTTCAAG<br>CAGCTGGTGG<br>AGGACCTGGA<br>\|ACTGGGGAA | 253 |

[0845]

| 表 23<br>融合体名称 | 癌症类型 | TCGA 肿瘤样品条形码 | 5'基因符号 | 5'登录号 | 5'染色体 | 5'断裂点 | 3'基因符号 | 3'登录号 | 3'染色体 | 3'断裂点 | 断裂点序列 | SEQ ID NO |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | GATCATGGAC<br>AGGTTCGAAG<br>AGGTTGTGTA<br>CCAGGCCATG<br>G | |
| FGFR3\|TACC3 | 乳头状肾细胞癌 | TCGA-A4-7<br>287-01A-11R<br>-2139-07 | FGFR3 | 2261 | chr4 | 1808661 | TACC3 | 10460 | chr4 | 1741429 | AGCAGCTGGT<br>GGAGGACCTG<br>GACCGTGTCC<br>TTACCGTGAC<br>GTCCACCGAC\|<br>GTAAAGGCGA<br>CACAGGAGGA<br>GAACCGGGAG<br>CTGAGGAGCA<br>GGTGTGAGGA | 254 |
| FGFR3\|TACC3 | 乳头状肾细胞癌 | TCGA-A4-7<br>287-01A-11R<br>-2139-07 | FGFR3 | 2261 | chr4 | 1808637 | TACC3 | 10460 | chr4 | 1742650 | CGCCCTCCCA<br>GAGGCCCACC<br>TTCAAGCAGC<br>TGGTGGAGGA<br>CCTGGACCGT\|<br>GTCCTTCTCC<br>GACCTCTTCA<br>AGCGTTTTGA<br>GAAACAGAAA<br>GAGGTGATCG | 255 |
| FGFR3\|TACC3 | 乳头状肾细胞癌 | TCGA-A4-7<br>287-01A-11R<br>-2139-07 | FGFR3 | 2261 | chr4 | 1808561 | TACC3 | 10460 | chr4 | 1741689 | GAGGGCCACC<br>GCATGGACAA<br>GCCCGCCAAC<br>TGCACACACG<br>ACCTGTACAT\|<br>GATCATGGAC<br>AGGTTCGAAG<br>AGGTTGTGTA<br>CCAGGCCATG<br>GAGGAAGTTC | 256 |

[0846]

| 表 23<br>融合体名称 | 癌症类型 | TCGA 肿瘤样品条形码 | 5'基因符号 | 5'登录号 | 5'染色体 | 5'断裂点 | 3'基因符号 | 3'登录号 | 3'染色体 | 3'断裂点 | 断裂点序列 | SEQ ID NO |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SEC16A\|NOTCH1 | 鳞状细胞肺癌 | TCGA-NC-A5HK-01A-11R-A26W-07 | SEC16A | 9919 | chr9 | 139352036 | NOTCH1 | 4851 | chr9 | 139418396 | GTACGCCCAG TCCCTGGGTG CCGAGACCTG CCCCCTGCCT AGTTTCCAGG\| ACCCCAACCC GTGCCTCAGC ACCCCCTGCA AGAACGCCGG GACATGCCAC | 257 |

**表 24. 表 22 的基因/融合体的可药化性状态**

| 基因 | 已批准的 | 注册前（批准前） | III 期 | II 期 | I 期 | 临床前 |
|---|---|---|---|---|---|---|
| FGFR3 | 帕纳替尼 | 马赛替尼 | 乐伐替尼 | 多韦替尼乳酸盐；ENMD-2076；AZD-4547 | JNJ-42756493；BGJ-398；LY-2874455；S-49076 | N |
| NOTCH1 | N | N | N | N | OMP-52M51 | Debio-0826；TR-4；Notch 抗体（AVEO）；Notch1 抑制剂（因特普罗廷公司）；BMS871；NTR-4 |

**表 39**

| 癌症 | 事件类型 | Q | 阳性数量 | 患者的总数量 | 染色体区带 | 基因（Entrez ID） | 可药化基因 | KM 证据 |
|---|---|---|---|---|---|---|---|---|
| 子宫内膜子宫内膜样腺癌 | 融合体 | 2.18E-03 | 5 | 258 | 11p15.5、4p13 | RPLP2（6181）、ATP8A1（609542） | | 较差的结果 |
| 宫颈鳞状细胞癌 | 融合体 | 3.56E-03 | 5 | 54 | 17q21.2 | KRT15（3866）、KRT19（3880） | | 较差的结果 |

[0847]

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 结肠直肠腺癌:KRAS 突变 | 失去功能的突变 | 9.69E-03 | 4 | 105 | 11q22-q23 | ATM（472） | 较差的结果 |
| 导管乳癌:ER 阳性和 HER2 阴性 | 融合体 | 1.46E-02 | 7 | 265 | 17p11.2、17p13 | USP22（23326）、MYH10（160776） | 较差的结果 |
| 子宫内膜子宫内膜样腺癌:微卫星稳定 | 峰内基因扩增 | 3.40E-02 | 8 | 171 | 3q26.2 | MECOM（2122） | 较差的结果 |
| 子宫内膜子宫内膜样腺癌 | 失去功能的突变 | 5.04E-02 | 4 | 188 | 16p13.3 | CREBBP（1387） | 较差的结果 |
| 皮肤黑素瘤 | 获得功能的突变 | 6.69E-02 | 5 | 214 | 7q34 | PRSS37（136242） | 较差的结果 |
| 子宫内膜浆液性腺癌 | 峰内基因缺失 | 7.52E-02 | 4 | 94 | 8p11.2 | FKSG2（59347） | 较差的结果 |
| 皮肤黑素瘤 | 获得功能的突变 | 7.94E-02 | 5 | 214 | 6p21.3 | STK19（8859） | 较差的结果 |
| 子宫内膜浆液性腺癌:微卫星稳定 | 失去功能的突变 | 8.05E-02 | 30 | 38 | 17p13.1 | TP53（7157） | TP53 | 有利的结果 |
| 结肠直肠腺癌:KRAS 突变，3 或 4 期 | 峰内基因扩增 | 8.58E-02 | 4 | 45 | 13q12.3 | CDX2（1045） | 较差的结果 |
| 结肠直肠腺癌:KRAS 突变 | 失去功能的突变 | 8.77E-02 | 4 | 105 | 18q21.1 | SMAD4（4089） | 较差的结果 |
| 结肠直肠腺癌:微卫星稳定 | 获得功能的突变 | 9.10E-02 | 10 | 21 | 12p12.1 | KRAS（3845） | KRAS（临床前） | 较差的结果 |

[0848] 实例7鉴别TP53的状态

[0849] 在分子诊断学和理解癌症生物学两方面的进展以更有效的患者分层将改进结果

且加快有效癌症药物的批准的期望提高临床试验范例的门槛。

[0850] TP53的突变状态已被鉴别为治疗反应和预后的预测生物标记物。举例来说，TP53野生型(WT)患者已展示出与含有TP53突变的患者相比在包括佐剂5-氟尿嘧啶和西妥昔单抗组合疗法的疗法之后展现显著延长的无进展存活期。

[0851] TP53突变注释从ONCOMINE™ NGS突变浏览器(密歇根州的肯潘迪亚生物科学公司(Compendia Biosciences,MI)获得。在评估TP53突变状态的总共776个患者中;259个患者含有TP53中的至少一个突变且被注释为TP53突变体，而519个患者缺乏检测的TP53突变且被注释为TP53野生型。TP53野生型和TP53突变体注释接着在患者层面映射到来自TCGA乳癌数据集的对应微阵列样品。当突变注释映射到具有对应微阵列数据的患者时，327个患者被注释为TP53野生型，且188个被注释为TP53突变体。TP53野生型和TP53突变标记由TCGA乳癌数据集的差异表达分析产生。基因列表通过p值根据史都登氏两类t检验定等级。在TP53野生型患者中有差异地上调的基因贡献到TP53野生型标记，而在TP53突变体患者中上调的基因贡献到TP53突变体标记。每一标记含有定等级的基因(n＝204)中的前1％。TP53野生型和TP53突变标记中的所有基因在校正假发现之后高度显著(Q<0.0001)。Q值计算为(p值/p值评级)*所测量的基因数量。

[0852] 五种ONCOMINE™癌症类型含有足够的TP53突变状态数据以完成分析。在这些当中，与来自乳癌(p<0.001;n＝189WT,37个突变体)、肺癌(p＝0.0003;n＝23WT,18个突变)、肝癌(p＝0.0069;n＝74WT,11个突变)和卵巢癌(p＝0.05;n＝22WT,15个突变)癌症患者的TP53突变的临床样品相比，在TP53WT中发现显著升高的标记表达，且在淋巴瘤患者(p＝0.068;n＝65WT,16个突变)内发现趋势(参见图5-7和9-10)。表40含有TP53WT TCGA乳癌标记。

[0853] 临床来源的表达标记有效地区分TP53WT与突变体肿瘤样品。

[0854] 表40

[0855]

| TP53 WT 标签基因 | | | | |
|---|---|---|---|---|
| SUSD3 | BAG1 | ZNF214 | USP30 | CEP120 |
| DMXL1 | ERBB4 | SLC24A1 | MKL2 | CA12 |
| P4HTM | PCP2 | AGBL2 | SYTL4 | SLC7A2 |
| KIF12 | C1orf64 | NME5 | HEXIM2 | ANKHD1-EIF4EBP3 |
| ACBD4 | TMEM161B | RERG | BRD8 | EIF4EBP3 |
| FSIP1 | SLC16A6 | VEZF1 | LOC644189 | TMEM128 |
| CAMLG | MLPH | ZNF484 | PJA2 | HVCN1 |
| FAM47E | LRBA | FBXO38 | TCEAL5 | TCTN1 |
| C14orf25 | EXOC6 | LOC100129623 | CHIC1 | TOX4 |
| USP47 | FAM174A | WFS1 | RNF135 | SEPSECS |
| POLK | C14orf19 | TRIM4 | LOC646976 | KIAA1370 |
| SPG11 | TCEAL3 | SLC7A8 | XPC | RG9MTD2 |
| TLE3 | CCNH | ZC3H6 | MED13L | CELSR1 |
| GLIPR1L2 | ANXA9 | SFRS12 | CXXC5 | TBC1D9B |
| PCBD2 | TTC8 | LOC100131801 | C9orf68 | |
| TCEAL4 | TCEAL6 | GAMT | CACNA1D | |
| KCTD3 | MAN2B2 | ABCC8 | ANKRD42 | |

[0856]

| | | | | |
|---|---|---|---|---|
| OBFC1 | CST5 | CRY2 | LOC440459 | |
| MRFAP1L1 | SCAMP1 | LRRC48 | PCM1 | |
| GMPR2 | PTGER3 | ZNF24 | C7orf63 | |
| DDB2 | CST3 | TMEM101 | RHBDD1 | |
| TIGD6 | PTPRT | NDFIP1 | WDFY3 | |
| KIAA0232 | RAI2 | CHCHD5 | REEP5 | |
| TMEM26 | GREB1 | KCNE4 | FUT8 | |
| PCDH19 | CCDC103 | PGR | ZFYVE1 | |

[0857] 实例8

[0858] 根据实例9的方法，分析8,690份临床样品的突变状态且分析7,202份临床样品的融合体状态。

[0859] 鉴别新颖突变和融合体事件。表44展示在CDNK2A和ALK中鉴别出的新颖突变。表41-43鉴别在分析中鉴别出的新颖融合体事件。

[0860] 每一目标的可药化性和可利用的临床试验的也提供在表中。

[0861] 新颖突变和融合体事件提供了不仅进一步开发目标治疗而且可选择化学治疗剂的未来目标的机会。

[0862] 实例9

[0863] 根据先前实例1-8的方法，分析8,690份临床样品的突变状态且分析7,202份临床样品的突变状态。

[0864] 鉴别新颖突变和癌症相关事件。表45和46展示出新颖突变/癌症相关性。

[0865] 每一目标的可药化性和可能的结果也提供在表45中。

[0866] 新颖突变/癌症相关事件提供了不仅进一步开发目标治疗而且可选择化学治疗剂的未来目标的机会。

[0867] 实例10

[0868] 根据先前实例1-8的方法，分析8,690份临床样品的突变状态且分析7,202份临床样品的突变状态。

[0869] 新颖突变/癌症相关事件展示在表47中。

[0870] 新颖融合体事件和相关癌症展示在表48中。断裂点和可药化性展示在表49中。与融合体搭配物有关的其他药物的状态展示在表50中。

[0871] 在某些实施例中，本发明提供：

[0872] 一种试剂盒，其包含一个探针组，所述探针组特异性地识别选自表42或49的两种基因的至少一种融合体。

[0873] 一种组合物，其包含一个探针组，所述探针组特异性地识别选自表42或49的两种基因的至少一种融合体。

[0874] 一种探针组，其特异性地识别包含SEQ ID NO:1-289中的至少一者的核酸。

[0875] 一种经分离核酸，其包含至少一个选自SEQ ID NO:1-289的序列。

[0876] 一种检测样品中的膀胱尿道上皮癌、乳癌、子宫内膜子宫内膜样腺癌、结肠腺癌、多形性成胶质细胞瘤、透明细胞肾细胞癌、乳头状肾细胞癌、急性骨髓性白血病、脑低级别胶质瘤、肺腺癌、卵巢浆液性囊腺癌、前列腺腺癌、直肠皮肤黑素瘤和甲状腺腺体癌的方法，所述方法包含：扩增包含选自SEQ ID NO:1-289的序列的核酸；以及检测所述包含选自SEQ ID NO:1-289的序列的核酸的存在；其中检测所述包含选自SEQ ID NO:1-289的序列的核

酸，指示上述癌症中的一种或多种存在于所述样品中。

[0877] 一种试剂盒，其包含一个探针组，所述探针组与至少一种包含来自表42或49的断裂点的核酸特异性地杂交。

[0878] 一种探针组，其与至少一种包含来自表42或49的断裂点的核酸特异性地杂交。

[0879] 一种检测样品中的膀胱尿道上皮癌、乳癌、子宫内膜子宫内膜样腺癌、结肠腺癌、多形性成胶质细胞瘤、透明细胞肾细胞癌、乳头状肾细胞癌、急性骨髓性白血病、脑低级别胶质瘤、肺腺癌、卵巢浆液性囊腺癌、前列腺腺癌、直肠皮肤黑素瘤和甲状腺腺体癌的方法，所述方法包含：扩增选自表41或48的基因融合体或融合基因产物；以及检测所述样品中选自表41或48的所述基因融合体或融合基因产物的存在；其中检测选自表41的所述基因融合体的存在，指示上述癌症中的一种或多种存在于所述样品中。

[0880] 一种经分离基因融合体核酸，其长度在100与10,000个核苷酸之间，且在表42或49中的断裂点中的一者的任一侧上包含至少25个核苷酸。

[0881] 经分离的基因融合体核酸可具有基因融合体搭配物，所述基因融合体搭配物为选自以下各者的两种基因：TPM1和ALK、PRKAR1A和ALK、NCOA1和ALK、LPP和CASR、MDM2和EGFR、FGFR3和ELAVL3、B2M和GNAS、DOCK8和JAK2、HNF1B和NOTCH1、NFASC和NTRK1、SSBP2和NTRK1、SQSTM1和NTRK1、TBL1XR1和PIK3CA、AKAP13和RET、FKBP15和RET、TBL1XR1和RET、CEP85L和ROS1、CLCN6和RAF1、TRAK1和RAF1、PRKACA和AKT1、PRKACA和AKT2、MLL和FYN、以及TTC13和JAK2、SEC16A和NOTCH1、ERC1和RET、GTF2IRD1和ALK、HTATSF1和BRS3、CDH1和CCDC132、CCDC132和CDH1、ERBB2和SLC29A3、MET和TFG；TFG和MET、NOTCH2和MNDA、IRF2BP2和NTRK1、EIF2C2和PTK2、RARA和HOXB3、STAT3和ETV4、以及GFAP和VIM、VIM和GFAP、TOP1和C17orf64、以及TP53和KIAA0753。

[0882] 在上述方法中，所述样品为患者样品。

[0883] 上述方法可进一步包含当包含选自SEQ ID NO:1-289的序列的核酸存在于所述患者样品时，将所述患者诊断为患有膀胱尿道上皮癌、乳癌、子宫内膜子宫内膜样腺癌、结肠腺癌、多形性成胶质细胞瘤、透明细胞肾细胞癌、乳头状肾细胞癌、急性骨髓性白血病、脑低级别胶质瘤、肺腺癌、卵巢浆液性囊腺癌、前列腺腺癌、直肠皮肤黑素瘤或甲状腺腺体癌。

[0884] 一种试剂盒，其包含一个探针组，其中所述探针组特异性地识别选自表41、表44、表45、表46、表47或其组合的基因变异体。

[0885] 一种组合物，其包含一个探针组，所述探针组特异性地识别来自表41、表44、表45、表46、表47或其组合的基因变异体。

[0886] 一种探针组，其特异性地识别来自表41、表44、表45、表46、表47或其组合的基因变异体。

[0887] 一种检测膀胱癌、乳癌、子宫颈细胞癌、结肠腺癌、多形性成胶质细胞瘤、头颈鳞状细胞癌、透明细胞肾细胞癌、肾脏肾乳头状细胞癌、急性骨髓性白血病、脑低级别胶质瘤、肝脏肝细胞癌、肺腺癌、鳞状细胞肺癌、卵巢浆液性腺癌、前列腺腺癌、直肠腺癌、皮肤黑素瘤、胃腺癌、甲状腺癌或子宫体子宫内膜样癌的方法，所述方法包含：扩增包含选自表7和/或表11的变异体的核酸；以及检测所述选自表41、表44、表45、表46、表47或其组合的变异体的存在；其中检测所述选自表41、表44、表45、表46、表47或其组合的变异体，指示上述癌症中的一种或多种存在于所述样品中。

[0888] 一种试剂盒，其包含一个探针组，所述探针组与选自表41、表44、表45、表46、表47或其组合的变异体特异性地杂交。

[0889] 一种探针组，其与选自表41、表44、表45、表46、表47或其组合的基因变异体特异性地杂交。

[0890] 根据权利要求17所述的方法，其中所述样品为患者样品。

[0891] 根据权利要求20所述的方法，其进一步包含当选自表41、表44、表45、表46、表47或其组合的变异体存在于所述患者样品中时，将所述患者诊断为患有膀胱癌、乳癌、子宫颈细胞癌、结肠腺癌、多形性成胶质细胞瘤、头颈鳞状细胞癌、透明细胞肾细胞癌、肾脏肾乳头状细胞癌、急性骨髓性白血病、脑低级别胶质瘤、肝脏肝细胞癌、肺腺癌、鳞状细胞肺癌、卵巢浆液性腺癌、前列腺腺癌、直肠腺癌、皮肤黑素瘤、胃腺癌、甲状腺癌或子宫体子宫内膜样癌。

[0892] 一种经分离基因构筑体，其包含选自SEQ ID NO:1-289的序列的反义聚核苷酸序列。

[0893] 经分离基因构筑体可进一步包含可操作地连接于所述反义聚核苷酸的启动子。

[0894] 经分离基因构筑体可为载体。所述载体可为病毒载体。

[0895] 一种包含使包含基因融合体的细胞与药物接触的方法，其中所述基因融合体公开于表41或表49中且所述药物为表43或表49中的药物中的一者或多者。

[0896] 一种包含向具有基因融合体的患者投与药物的方法，其中所述基因融合体公开于表41或表49中。

[0897] 根据权利要求27所述的方法，其中所述患者经诊断患有癌症。

[0898] 根据权利要求28所述的方法，其中所述癌症为公开于表41或表49中的任一者中的癌症类型。

[0899] 一种包含使来自患者的核酸样品与包含两种引物的反应混合物接触的方法，其中第一引物与一种基因互补，且第二引物与第二基因互补，其中所述第一基因和所述第二基因的融合体可通过利用所述第一引物和所述第二引物生成的扩增子的存在来检测，其中所述融合体断裂点为表42或49的断裂点中的一者，且其中具有扩增子的患者被投与表43或表49中的药物中的一者或多者。

[0900] 一种包含执行生物分析以检测来自受试者的样品中的一种或多种基因融合体的方法，其中所述基因融合体中的至少一者是选自表42或表49中的那些，将所述生物分析的结果接收到计算机系统中，处理所述结果以确定输出，将所述输出呈现在可读媒体上，其中所述输出基于所述基因融合体的存在或不存在来确认推荐给所述受试者的治疗选择。

[0901] 在上述方法中，所述生物分析可包括对表42或表49的断裂点中的一者或多者具有特异性的探针。

[0902] 一种试剂盒，其包含：一个探针组，其中每一探针与包含来自表42或表49的断裂点的核酸特异性地杂交。

[0903] 在所述试剂盒中，每一非天然存在的探针可包含：经配置以与所述包含来自表42或表49的断裂点的核酸特异性地杂交的核酸序列，和共价键结于所述核酸序列的可检测部分。

[0904] 在所述试剂盒中，每一非天然存在的探针可包含：脱氧核糖核酸序列，其包含至少

一个脱氧尿苷(dU)残基代替一个脱氧胸苷残基。

[0905] 一种方法,其包含:扩增来自样品的包含来自表41或表49的至少一种基因融合体的核酸;以及通过以下各项中的至少一者检测所述至少一种基因融合体的存在:使组合物与至少一种探针接触,其中每一探针与所述核酸特异性地杂交,或观察所述核酸中的非天然或非原生化学结构的存在;其中检测所述至少一种基因融合体的存在指示至少一种来自表41或表49的癌症存在于所述样品中。

[0906] 一种系统,其包含:核酸扩增器,其经配置以扩增来自样品的包含来自表41或表49的至少一种基因融合体的核酸,得到经扩增核酸;检测器,其经配置以通过以下各项中的至少一者检测所述经扩增核酸中的所述至少一种基因融合体的存在:(i)使组合物与至少一种探针接触,其中每一探针与所述核酸特异性地杂交,或(ii)观察所述核酸中的非天然或非原生化学结构的存在,且经进一步配置以发送检测指示;以及计算机系统,其经配置以接收所述检测指示,且基于所述检测指示来确定来自表41或表49的至少一种癌症存在于所述样品中。

[0907] 一种非暂时性计算机可读程序存储单元,其经指令编码,所述指令当由计算机执行时执行一种方法,所述方法包含:接收包含至少一种癌症类型和事件类型的输入,其中所述癌症类型选自表44、45、46和/或47,且所述事件类型选自表44、45、46和/或47;在数据库中查询至少一个包含多个字段的条目,其中所述多个字段包含所述癌症类型和所述事件类型中的至少一者;发送包含来自所述至少一个条目的所述多个中的至少一个字段的输出,其中所述至少一个字段包含至少一种基因、至少一种可药化基因、至少一种靶向所述至少一种可药化基因的药物、或预后。

[0908] 一种方法,其包含:向具有选自表41中列出的基因融合体的至少一种基因融合体的患者投与选自表43中列出的药物的至少一种药物。

[0909] 一种方法,其包含:使来自患者的核酸样品与包含与第一基因互补的第一引物和与第二基因互补的第二引物的反应混合物接触,其中所述第一基因和所述第二基因的融合体可通过由所述第一引物和所述第二引物生成的扩增子的存在来检测,其中所述融合体包含选自表42或表49中列出的断裂点的断裂点。

[0910] 在上述方法中,其中回应于所述融合体的检测,将选自表43、45、46和/或47中列出的药物的至少一种药物投与所述患者。

[0911] 一种非暂时性计算机可读程序存储单元,其经指令编码,所述指令当由计算机执行时执行一种方法,所述方法包含:接收来自至少一种癌细胞系的RNA序列数据;对所述序列数据运行至少一种基因融合体调用程序,以鉴别所述经处理数据中的融合基因之间的可能的断裂点;过滤所述可能的断裂点以保留候选断裂点,其中每一候选断裂点是在功能基因区的5'非翻译区(UTR)或编码DNA序列(CDS)中,且每一候选断裂点不出现在内含子中;用适用于针对癌症诊断、癌症预后或癌症治疗中的至少一者测定基因融合体的相关性的至少一种注释对所述候选断裂点进行注释,其中所述基因融合体包含所述候选断裂点。

[0912] 在所述非暂时性计算机可读程序存储单元中,所述至少一种基因融合体调用程序可选自TopHat和deFuse。

[0913] 一种非暂时性计算机可读程序存储单元,其经指令编码,所述指令当由计算机执行时执行一种方法,所述方法包含:接收来自至少一种癌细胞系的突变数据;用变异体分

类、变异体位置或变异体变化中的至少一者对所述突变数据进行注释，得到经注释的突变数据；过滤所述经注释的突变数据，得到基因区域突变数据；将所述基因区域突变数据归类为热点、有害或其他；以及基于所述基因中的突变相对频率和所述基因中所有基因区域突变的分类，将包含所述基因区域突变的基因命名为获得功能、失去功能或再现的其他基因。

[0914] 在所述非暂时性计算机可读程序存储单元中，所述变异体分类可选自剪接位点、3'非翻译区(UTR)、5'UTR、内含子、错义、无义、无终止、沉默、框移插入、框移缺失、框内插入、框内缺失或非编码外显子。

[0915] 在所述非暂时性计算机可读程序存储单元中，过滤所述经注释的突变数据可包含将在已知转录物外的突变和与已知基因不相关的突变排除。

[0916] 在所述非暂时性计算机可读程序存储单元中，将所述突变归类为热点可包含测定其在至少三种癌细胞系中的相同变异体位置处的存在，其中所述变异体分类是错义、无终止、框内插入、框内缺失或非编码外显子。

[0917] 在所述非暂时性计算机可读程序存储单元中，其中将所述突变归类为有害可包含观察所述变异体分类是无义、框移插入或框移缺失。

[0918] 一种方法，其包含：检测来自受试者的样品中的一种或多种基因融合体，得到基因融合体检测数据，其中所述基因融合体中的至少一者是选自表41或49中列出的基因融合体，通过计算机系统接收所述基因融合体检测数据，基于所述基因融合体检测数据，通过所述计算机系统确认推荐给所述受试者的至少一种治疗选择。

[0919] 一种系统，其包含：检测器，其经配置以(i)检测来自受试者的样品中的一种或多种基因融合体，得到基因融合体检测数据，其中所述基因融合体中的至少一者选自表41或49中列出的基因融合体和(ii)发送所述基因融合体检测数据；以及计算机系统，其经配置以接收所述基因融合体检测数据且基于所述基因融合体检测数据来确认推荐给所述受试者的至少一种治疗选择。

[0920] 一种基因融合体，其通过一种方法制备，所述方法包含：分离包含所述基因融合体的RNA分子；以及合成与所述经分离RNA分子互补的互补DNA(cDNA)分子。

[0921] 除非另外指明，否则本说明书和权利要求书中所用的表示成分量、特性(如分子量)、反应条件等的所有数量应理解为在所有情况下都由术语“约”修饰。因此，除非指明为相反的，否则在说明书和所附权利要求书中所阐述的数值参数是可能取决于试图通过本发明获得的所需性质而变化的近似值。最低限度地，并且不试图限制等效物原则对权利要求书范围的应用，每一个数值参数都应至少根据所报告的有效数字的数量并且通过应用一般四舍五入技术来解释。尽管阐述本发明广泛范围的数值范围和参数是近似值，但具体实例中所阐述的数值是尽可能精确报导的。但是，任何数值固有地含有某些由其各别测试测量值中所发现的标准偏差必然造成的误差。

[0922] 除非本文另外指示或明显与上下文相矛盾，否则在描述本发明的情形下(尤其在随附权利要求书的情形下)所用的术语“一(a/an)”、“所述”和类似指示物应解释为涵盖单数和复数两者。本文中值的范围的叙述仅旨在充当个别地提及处于所述范围内的每个单独值的简写方法。除非本文另外指示，否则每一个个别值并入到本说明书中，如同其在本文中个别地列举一般。除非本文另外指出或另外明显与内容相矛盾，否则本文所描述的所有方法可以任何适合顺序执行。本文中提供的任何和所有实例或例示性语言(例如，“如”)的使

用，仅打算更好地阐明本发明，并且不对以其他方式所要求的本发明范围造成限制。本说明书中的任何语言都不应被解释为指示任何未要求的要素对于实践本发明而言是必需的。

[0923] 本文所公开的本发明的替代性要素或实施例的分组不应解释为限制。每一个群组成员都可以个别地或以与所述群组中其他成员或本文中所见其他要素的任何组合形式来提及和要求。预期，群组中的一个或多个成员可以出于便利性和/或专利性的原因而包括于群组中或由群组删除。当任何所述包括或删除发生时，本说明书被认为如所修改地含有群组，因此满足所附权利要求书中所用的所有马库什(Markush)群组的书面描述。

[0924] 本文中描述了本发明的某些实施例，包括本发明人已知的用于进行本发明的最佳模式。当然，在阅读以上描述之后，这些所描述实施例的变化对于所属领域的技术人员将变得显而易见。本发明人期望熟练的技术人员在适当时采用这类变化，并且本发明人打算以不同于本文中具体描述的其他方式来实践本发明。因此，本发明包括可适用法律所允许的随附权利要求书中所引述的主题的所有修改和等效物。此外，除非本文另外指示或另外明显与内容相矛盾，否则本发明涵盖上述要素以其所有可能的变化形式的任何组合。

[0925] 此外，已经在本说明书通篇中大量参考专利和印刷出版物。上文所引用的参考文献和印刷出版物中的每一个都个别地以全文引用的方式并入本文中。

[0926] 应理解本文所公开的本发明实施例说明本发明的原理。可以采用的其他修改处于本发明的范围内。因此，作为实例而非限制，可以根据本文中的教示来利用本发明的替代性配置。因此，本发明不限于如所精确展示和描述的内容。

[0927] 本文中所公开的具体实施例可以在权利要求书中使用由……组成或基本上由……组成语言来进一步限制。当用于权利要求书中时，不论如所提交还是根据修改添加，过渡术语“由……组成”排除权利要求书中未指定的任何要素、步骤或成分。过渡术语“基本上由……组成”将权利要求书的范围限制于规定材料或步骤和不实质上影响基本和新颖特征的材料或步骤。如此要求的本发明实施例固有地或明确地描述并实现于本文中。

[0928] 提及序列识别符(如从NM_开始的那些)是指数据库登录号和下面的序列，如其见于2013年4月18日。

[0929] 表7

[0930]

| TCGA疾病 | 基因符号 | Entrez基因ID | 起始位置 | 参考等位基因 | 肿瘤序列等位基因 | 转录物 | 变异体C |
|---|---|---|---|---|---|---|---|
| READ | ANXA1 | 301 | 74965099 | G | A | NM_000700 | p.R124H |
| SKCM | ANXA1 | 301 | 75775278 | C | T | NM_000700 | p.R124C |
| UCEC | ANXA1 | 301 | 75775279 | G | A | NM_000700 | p.R124H |
| BRCA | AR | 367 | 66765161 | A | T | NM_000044 | p.Q58L |
| HNSC | AR | 367 | 66765161 | A | T | NM_000044 | p.Q58L |
| KIRP | AR | 367 | 66765161 | A | T | NM_000044 | p.Q58L |
| LGG | AR | 367 | 66765161 | A | T | NM_000044 | p.Q58L |
| LUAD | AR | 367 | 66765161 | A | T | NM_000044 | p.Q58L |
| STAD | AR | 367 | 66765161 | A | T | NM_000044 | p.Q58L |
| UCEC | AR | 367 | 66765161 | A | T | NM_000044 | p.Q58L |
| LUAD | ARAF | 369 | 47426120 | T | A | NM_001654 | p.S214T |
| LUAD | ARAF | 369 | 47426121 | C | T | NM_001654 | p.S214F |
| SKCM | ARAF | 369 | 47426121 | C | T | NM_001654 | p.S214F |
| PAAD | ATP6V1A | 523 | 113505224 | T | C | NM_001690 | p.L237P |
| SKCM | ATP6V1A | 523 | 113505224 | T | C | NM_001690 | p.L237P |
| LUAD | CDK4 | 1019 | 58145430 | C | A | NM_000075 | p.R24L |
| SKCM | CDK4 | 1019 | 58145430 | C | A | NM_000075 | p.R24L |
| SKCM | CDK4 | 1019 | 58145431 | G | T | NM_000075 | p.R24S |
| OV | CHEK2 | 11200 | 27422947 | C | T | NM_007194 | p.R346H |
| GBM | CHEK2 | 11200 | 29083962 | G | C | NM_007194 | p.R519G |
| HNSC | CHEK2 | 11200 | 29083962 | G | C | NM_007194 | p.R519G |
| KIRC | CHEK2 | 11200 | 29083962 | G | C | NM_007194 | p.R519G |
| PAAD | CHEK2 | 11200 | 29083962 | G | C | NM_007194 | p.R519G |
| HNSC | CHEK2 | 11200 | 29091840 | T | C | NM_007194 | p.K373E |
| KIRC | CHEK2 | 11200 | 29091840 | T | C | NM_007194 | p.K373E |
| LUAD | CHEK2 | 11200 | 29091840 | T | C | NM_007194 | p.K373E |
| SKCM | CHEK2 | 11200 | 29091840 | T | C | NM_007194 | p.K373E |
| BRCA | CHEK2 | 11200 | 29092948 | G | A | NM_007194 | p.R346C |
| LUSC | CHEK2 | 11200 | 29092948 | G | C | NM_007194 | p.R346G |
| HNSC | CSNK2A1 | 1457 | 470440 | T | C | NM_001895 | p.H236R |
| LUAD | CSNK2A1 | 1457 | 470440 | T | C | NM_001895 | p.H236R |
| LUSC | CSNK2A1 | 1457 | 470440 | T | C | NM_001895 | p.H236R |
| STAD | CSNK2A1 | 1457 | 470440 | T | C | NM_001895 | p.H236R |
| THCA | CSNK2A1 | 1457 | 470440 | T | C | NM_001895 | p.H236R |
| GBM | DRD5 | 1816 | 9784478 | C | A | NM_000798 | p.S275R |
| HNSC | DRD5 | 1816 | 9784478 | C | A | NM_000798 | p.S275R |
| LUSC | DRD5 | 1816 | 9784478 | C | A | NM_000798 | p.S275R |
| STAD | DRD5 | 1816 | 9784478 | C | A | NM_000798 | p.S275R |
| COAD | ERBB3 | 2065 | 54765121 | G | A | NM_001982 | p.V104M |
| COAD | ERBB3 | 2065 | 54765121 | G | A | NM_001982 | p.V104M |
| COAD | ERBB3 | 2065 | 54765121 | G | T | NM_001982 | p.V104L |
| READ | ERBB3 | 2065 | 54765121 | G | A | NM_001982 | p.V104M |
| CESC | ERBB3 | 2065 | 56478854 | G | A | NM_001982 | p.V104M |
| STAD | ERBB3 | 2065 | 56478854 | G | T | NM_001982 | p.V104L |
| STAD | ERBB3 | 2065 | 56478854 | G | A | NM_001982 | p.V104M |
| UCEC | ERBB3 | 2065 | 56478854 | G | A | NM_001982 | p.V104M |

[0931]

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| BRCA | ERBB3 | 2065 | 56482341 | G | T | NM_001982 | p.D297Y |
| UCEC | ERBB3 | 2065 | 56482341 | G | T | NM_001982 | p.D297Y |
| UCEC | ERBB3 | 2065 | 56482341 | G | A | NM_001982 | p.D297N |
| UCEC | ERBB3 | 2065 | 56482342 | A | T | NM_001982 | p.D297V |
| HNSC | FGFR3 | 2261 | 1803568 | C | G | NM_000142 | p.S249C |
| KIRP | FGFR3 | 2261 | 1803568 | C | G | NM_000142 | p.S249C |
| LUSC | FGFR3 | 2261 | 1803568 | C | G | NM_000142 | p.S249C |
| COAD | GPRC5A | 9052 | 12952538 | G | A | NM_003979 | p.V30I |
| UCEC | GPRC5A | 9052 | 13061271 | G | A | NM_003979 | p.V30I |
| LUAD | GPX1 | 2876 | 49395482 | G | C | NM_000581 | p.P77R |
| SKCM | GPX1 | 2876 | 49395482 | G | C | NM_000581 | p.P77R |
| STAD | GPX1 | 2876 | 49395482 | G | C | NM_000581 | p.P77R |
| KIRC | HSD17B7 | 51478 | 162769603 | G | A | NM_016371 | p.S173N |
| PAAD | HSD17B7 | 51478 | 162769603 | G | A | NM_016371 | p.S173N |
| BRCA | JUN | 3725 | 59248409 | C | T | NM_002228 | p.E112K |
| LUSC | JUN | 3725 | 59248409 | C | T | NM_002228 | p.E112K |
| LUSC | JUN | 3725 | 59248409 | C | G | NM_002228 | p.E112Q |
| COAD | KDR | 3791 | 55650977 | C | T | NM_002253 | p.R1032Q |
| SKCM | KDR | 3791 | 55955863 | G | A | NM_002253 | p.S1100F |
| SKCM | KDR | 3791 | 55956220 | C | T | NM_002253 | p.R1032Q |
| LAML | KIT | 3815 | 55294077 | G | T | NM_000222 | p.D816Y |
| LAML | KIT | 3815 | 55294078 | A | T | NM_000222 | p.D816V |
| SKCM | LHCGR | 3973 | 48915500 | C | T | NM_000233 | p.R479Q |
| UCEC | LHCGR | 3973 | 48915500 | C | A | NM_000233 | p.R479L |
| SKCM | LHCGR | 3973 | 48936151 | C | T | NM_000233 | p.E206K |
| HNSC | MAP2K2 | 5605 | 4117549 | A | C | NM_030662 | p.F57L |
| SKCM | MAP2K2 | 5605 | 4117551 | A | C | NM_030662 | p.F57V |
| STAD | MAP2K2 | 5605 | 4117551 | A | C | NM_030662 | p.F57V |
| CESC | MAPK1 | 5594 | 22127164 | C | T | NM_002745 | p.E322K |
| HNSC | MAPK1 | 5594 | 22127164 | C | T | NM_002745 | p.E322K |
| COAD | MMP15 | 4324 | 56631345 | G | A | NM_002428 | p.R169H |
| SKCM | MMP15 | 4324 | 58073843 | C | T | NM_002428 | p.R169C |
| LUAD | MMP15 | 4324 | 58073844 | G | A | NM_002428 | p.R169H |
| OV | MMP3 | 4314 | 102215174 | G | A | NM_002422 | p.R316C |
| GBM | MMP3 | 4314 | 102709963 | C | T | NM_002422 | p.R316H |
| GBM | MMP3 | 4314 | 102709964 | G | A | NM_002422 | p.R316C |
| LUAD | MMP3 | 4314 | 102709964 | G | A | NM_002422 | p.R316C |
| COAD | MTOR | 2475 | 11107160 | G | T | NM_004958 | p.S2215Y |
| KIRC | MTOR | 2475 | 11184573 | G | T | NM_004958 | p.S2215Y |
| KIRP | MTOR | 2475 | 11184573 | G | T | NM_004958 | p.S2215Y |
| UCEC | MTOR | 2475 | 11184573 | G | T | NM_004958 | p.S2215Y |
| KIRC | MTOR | 2475 | 11189845 | G | C | NM_004958 | p.F1888L |
| UCEC | MTOR | 2475 | 11189845 | G | T | NM_004958 | p.F1888L |
| UCEC | MTOR | 2475 | 11189847 | A | C | NM_004958 | p.F1888V |
| OV | MTOR | 2475 | 11195525 | C | T | NM_004958 | p.A1105T |
| KIRC | MTOR | 2475 | 11217230 | C | T | NM_004958 | p.C1483Y |
| KIRC | MTOR | 2475 | 11217230 | C | A | NM_004958 | p.C1483F |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | GBM | MTOR | 2475 | 11217231 | A | G | NM_004958 | p.C1483R |
| | SKCM | MTOR | 2475 | 11272938 | C | T | NM_004958 | p.A1105T |
| | GBM | PIK3CB | 5291 | 138374244 | T | G | NM_006219 | p.D1067A |
| | HNSC | PIK3CB | 5291 | 138374244 | T | A | NM_006219 | p.D1067V |
| | THCA | PIK3CB | 5291 | 138374244 | T | A | NM_006219 | p.D1067V |
| | UCEC | PIK3CB | 5291 | 138374245 | C | A | NM_006219 | p.D1067Y |
| | LUAD | PIK3R2 | 5296 | 18273784 | G | A | NM_005027 | p.G373R |
| | UCEC | PIK3R2 | 5296 | 18273784 | G | A | NM_005027 | p.G373R |
| | COAD | POLE | 5426 | 131760362 | C | A | NM_006231 | p.V411L |
| | COAD | POLE | 5426 | 131763257 | G | T | NM_006231 | p.P286H |
| | UCEC | POLE | 5426 | 133250289 | C | A | NM_006231 | p.V411L |
| | UCEC | POLE | 5426 | 133253184 | G | C | NM_006231 | p.P286R |
| | UCEC | PPP2R1A | 5518 | 52715971 | C | G | NM_014225 | p.P179R |
| | UCEC | PPP2R1A | 5518 | 52715982 | C | T | NM_014225 | p.R183W |
| | HNSC | PPP2R1A | 5518 | 52715983 | G | A | NM_014225 | p.R183Q |
| | STAD | PPP2R1A | 5518 | 52715983 | G | A | NM_014225 | p.R183Q |
| | UCEC | PPP2R1A | 5518 | 52716323 | C | T | NM_014225 | p.S256F |
| | UCEC | PPP2R1A | 5518 | 52716323 | C | A | NM_014225 | p.S256Y |
| | UCEC | PPP2R1A | 5518 | 52716328 | C | T | NM_014225 | p.R258C |
| | LUAD | PPP2R1A | 5518 | 52716329 | G | A | NM_014225 | p.R258H |
| | COAD | PPP2R1A | 5518 | 57407794 | C | T | NM_014225 | p.R183W |
| | COAD | PPP2R1A | 5518 | 57407794 | C | T | NM_014225 | p.R183W |
| | OV | PPP2R1A | 5518 | 57407794 | C | T | NM_014225 | p.R183W |
| [0932] | COAD | PPP2R1A | 5518 | 57408141 | G | A | NM_014225 | p.R258H |
| | HNSC | PRKCA | 5578 | 64299066 | G | C | NM_002737 | p.E33Q |
| | LUAD | PRKCA | 5578 | 64299066 | G | A | NM_002737 | p.E33K |
| | LUSC | PRKCA | 5578 | 64299066 | G | A | NM_002737 | p.E33K |
| | KIRC | PRKCH | 5583 | 61789073 | C | T | NM_006255 | p.A85V |
| | PAAD | PRKCH | 5583 | 61789073 | C | T | NM_006255 | p.A85V |
| | STAD | PRKCI | 5584 | 170013719 | C | A | NM_002740 | p.R480S |
| | COAD | PRKCI | 5584 | 171496413 | C | T | NM_002740 | p.R480C |
| | COAD | PRKCI | 5584 | 171496413 | C | T | NM_002740 | p.R480C |
| | OV | PRKCI | 5584 | 171496413 | C | T | NM_002740 | p.R480C |
| | COAD | RAF1 | 5894 | 12620699 | G | A | NM_002880 | p.S257L |
| | COAD | RAF1 | 5894 | 12620699 | G | A | NM_002880 | p.S257L |
| | LUAD | RAF1 | 5894 | 12645699 | G | A | NM_002880 | p.S257L |
| | LUAD | RAF1 | 5894 | 12645699 | G | C | NM_002880 | p.S257W |
| | SKCM | RAF1 | 5894 | 12645699 | G | A | NM_002880 | p.S257L |
| | STAD | RAF1 | 5894 | 12645699 | G | A | NM_002880 | p.S257L |
| | KIRC | RHEB | 6009 | 151188050 | A | T | NM_005614 | p.Y35N |
| | UCEC | RHEB | 6009 | 151188050 | A | T | NM_005614 | p.Y35N |
| | STAD | RHOA | 387 | 49412898 | T | C | NM_001664 | p.Y42C |
| | STAD | RHOA | 387 | 49412898 | T | G | NM_001664 | p.Y42S |
| | BRCA | RHOA | 387 | 49412905 | C | G | NM_001664 | p.E40Q |
| | HNSC | RHOA | 387 | 49412905 | C | G | NM_001664 | p.E40Q |
| | COAD | SRC | 6714 | 35464354 | G | C | NM_005417 | p.D407H |
| | OV | SRC | 6714 | 35464354 | G | C | NM_005417 | p.D407H |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | SKCM | SRCIN1 | 80725 | 36704930 | C | T | NM_025248 | p.E1045K |
| | READ | SYK | 6850 | 92676932 | G | T | NM_003177 | p.K387N |
| | LGG | SYK | 6850 | 93637110 | A | G | NM_003177 | p.K387R |
| | SKCM | SYK | 6850 | 93637110 | A | G | NM_003177 | p.K387R |
| | STAD | TOP2A | 7153 | 38552660 | T | C | NM_001067 | p.K1199E |
| | THCA | TOP2A | 7153 | 38552660 | T | C | NM_001067 | p.K1199E |
| | COAD | TOP2B | 7155 | 25643731 | C | T | NM_001068 | p.R651H |
| | UCEC | TOP2B | 7155 | 25668727 | C | T | NM_001068 | p.R651H |
| | GBM | TUBA1B | 10376 | 49523423 | C | T | NM_006082 | p.G29D |
| | STAD | TUBA1B | 10376 | 49523423 | C | T | NM_006082 | p.G29D |
| [0933] | HNSC | TUBA1B | 10376 | 49523424 | C | G | NM_006082 | p.G29R |
| | BLCA | TXNRD1 | 7296 | 104725378 | G | A | NM_003330 | p.E439K |
| | CESC | TXNRD1 | 7296 | 104725378 | G | C | NM_003330 | p.E439Q |
| | UCEC | TXNRD1 | 7296 | 104725378 | G | C | NM_003330 | p.E439Q |
| | HNSC | TXNRD1 | 7296 | 104725379 | A | G | NM_003330 | p.E439G |
| | KIRC | TXNRD1 | 7296 | 104725379 | A | G | NM_003330 | p.E439G |
| | LGG | VEGFB | 7423 | 64005040 | A | C | NM_003377 | p.T187P |
| | PAAD | VEGFB | 7423 | 64005040 | A | C | NM_003377 | p.T187P |
| | HNSC | VEGFB | 7423 | 64005048 | A | C | NM_001243733 | p.T156P |
| | PAAD | VEGFB | 7423 | 64005048 | A | C | NM_001243733 | p.T156P |
| | SKCM | VEGFB | 7423 | 64005048 | A | C | NM_001243733 | p.T156P |

| | |
|---|---|
| BLCA | 膀胱尿道上皮癌 |
| BRCA | 侵袭性乳癌 |
| CESC | 宫颈鳞状细胞癌 |
| COAD | 结肠腺癌 |
| GBM | 成胶质细胞瘤 |
| HNSC | 头颈鳞状癌 |
| KIRC | 肾脏肾透明细胞癌 |
| KIRP | 肾脏肾乳头状细胞癌 |
| LAML | 急性骨髓性白血病 |
| LGG | 低级别胶质瘤 |
| LUAD | 肺腺癌 |
| LUSC | 肺鳞状细胞癌 |
| OV | 卵巢癌 |
| PAAD | 胰腺癌 |
| READ | 直肠腺癌 |
| SKCM | 皮肤黑素瘤 |
| STAD | 胃腺癌 |
| THCA | 甲状腺癌 |
| UCEC | 子宫体子宫内膜样癌 |

[0934]

[0935] 表41

[0936]

| 癌症类型 | 5’基因符号 | 3’基因符号 | 可药化基因 |
|---|---|---|---|
| 前列腺腺癌 | ACPP | AXL | AXL |
| 前列腺腺癌 | ATM | CARD18 | ATM |
| 肺腺癌 | ATXN1 | ARG1 | ARG1 |
| 低级别胶质瘤 | CLU | CST3 | CLU |
| 前列腺腺癌 | DHX15 | ETV1 | 无 |
| 低级别胶质瘤 | DOT1L | TECR | DOT1L |
| 肾上腺皮质癌 | EXOSC10 | MTOR | MTOR |

| | | | |
|---|---|---|---|
| 宫颈鳞状细胞癌 | FAT1 | NTRK3 | NTRK3 |
| 膀胱尿道上皮癌 | FHIT | RUNX1 | 无 |
| 前列腺腺癌 | GUCA2A | ERG | 无 |
| 肉瘤 | KANK2 | ALK | ALK |
| 皮肤黑素瘤 | LMNA | RAF1 | RAF1 |
| 前列腺腺癌 | MDM2 | BRI3BP | MDM2 |
| 宫颈鳞状细胞癌 | MX1 | ERG | 无 |
| 肝细胞癌 | OXR1 | MET | MET |
| 乳头状肾细胞癌 | PLS1 | PIK3CB | PIK3CB |
| 结肠和直肠腺癌 | RPS6 | FN1 | FN1 |
| 皮肤黑素瘤 | RSBN1L | HGF | HGF |
| 肺腺癌 | RUNX1 | CEP76 | 无 |
| 肉瘤 | SEC16A | RXRA | RXRA |
| 肺腺癌 | STK11 | HMHA1 | 无 |
| 卵巢浆液性囊腺癌 | TBL1XR1 | CTNNB1 | CTNNB1 |
| 前列腺腺癌 | TMPRSS2 | GUCA2A | 无 |
| 前列腺腺癌 | TMPRSS2 | TMEM109 | 无 |
| 前列腺腺癌 | TMPRSS2 | BRAF | BRAF |
| 结肠和直肠腺癌 | USP7 | PRKCB | PRKCB |
| 低级别胶质瘤 | VEGFA | STK38 | VEGFA |

[0937]

表 42

| 5'基因符号 | 3'基因符号 | 癌症类型 | TCGA 肿瘤样品条形码 | 5'登录号 | 5'染色体 | 5'断裂点 | 3'登录号 | 3'染色体 | 3'断裂点 |
|---|---|---|---|---|---|---|---|---|---|
| ACPP | AXL | 前列腺腺癌 | TCGA-J4-A6G3-01A-11R-A311-07 | 55 | chr3 | 132075687 | 558 | chr19 | 41745606 |
| ATM | CARD18 | 前列腺腺癌 | TCGA-KK-A6E8-01A-11R-A31N-07 | 472 | chr11 | 108218092 | 59082 | chr11 | 105009805 |
| ATXN1 | ARG1 | 肺腺癌 | TCGA-55-6978-01A-11R-1949-07 | 6310 | chr6 | 16658016 | 383 | chr6 | 131897847 |
| CLU | CST3 | 低级别胶质瘤 | TCGA-CS-4944-01A-01R-1470-07 | 1191 | chr8 | 27461879 | 1471 | chr20 | 23618406 |
| CLU | CST3 | 低级别胶质瘤 | TCGA-P5-A72U-01A-31R-A32Q-07 | 1191 | chr8 | 27461878 | 1471 | chr20 | 23614559 |
| DHX15 | ETV1 | 前列腺腺癌 | TCGA-G9-6339-01A-12R-A311-07 | 1665 | chr4 | 24585946 | 2115 | chr7 | 13975521 |
| DOT1L | TECR | 低级别胶质瘤 | TCGA-TM-A7CF-02A-11R-A32Q-07 | 84444 | chr19 | 2194576 | 9524 | chr19 | 14673337 |
| DOT1L | TECR | 低级别胶质瘤 | TCGA-TM-A7CF-01A-11R-A32Q-07 | 84444 | chr19 | 2194576 | 9524 | chr19 | 14673337 |
| EXOSC10 | MTOR | 肾上腺皮质癌 | TCGA-OR-A5J7-01A-11R-A29S-07 | 5394 | chr1 | 11139768 | 2475 | chr1 | 11190834 |
| FAT1 | NTRK3 | 宫颈鳞状细胞癌 | TCGA-HM-A3JK-01A-11R-A32Y-07 | 2195 | chr4 | 187627717 | 4916 | chr15 | 88680792 |
| FHIT | RUNX1 | 膀胱尿道上皮癌 | TCGA-LC-A66R-01A-41R-A30C-07 | 2272 | chr3 | 61237029 | 861 | chr21 | 36231788 |
| GUCA2A | ERG | 前列腺腺癌 | TCGA-KC-A4BR-01A-32R-A32Y-07 | 2980 | chr1 | 42629074 | 2078 | chr21 | 39817544 |
| HMHA1 | STK11 | 肺腺癌 | TCGA-62-8395-01A-11R-2326-07 | 23526 | chr19 | 1066162 | 6794 | chr19 | 1218416 |
| STK11 | HMHA1 | 肺腺癌 | TCGA-NJ-A4YF-01A-12R-A262-07 | 6794 | chr19 | 1222005 | 23526 | chr19 | 1080254 |
| KANK2 | ALK | 肉瘤 | TCGA-FX-A48G-01A-11R-A24X-07 | 25959 | chr19 | 11303507 | 238 | chr2 | 29451932 |
| LMNA | RAF1 | 皮肤黑素瘤 | TCGA-EB-A5SF-01A-11R-A311-07 | 4000 | chr1 | 156107492 | 5894 | chr3 | 12645651 |
| MDM2 | BRI3BP | 前列腺腺癌 | TCGA-KK-A7B2-01A-12R-A32O-07 | 4193 | chr12 | 69214153 | 140707 | chr12 | 125509889 |
| MX1 | ERG | 宫颈鳞状细胞癌 | TCGA-C5-A7CJ-01A-11R-A32P-07 | 4599 | chr21 | 42809050 | 2078 | chr21 | 39947620 |
| OXR1 | MET | 肝细胞癌 | TCGA-RC-A6M6-01A-11R-A32O-07 | 55074 | chr8 | 107726213 | 4233 | chr7 | 116411552 |
| PLS1 | PIK3CB | 乳头状肾细胞癌 | TCGA-G7-6793-01A-11R-1965-07 | 5357 | chr3 | 142315341 | 5291 | chr3 | 138461531 |
| RPS6 | FN1 | 结肠和直肠腺癌 | TCGA-AG-3609-01A-02R-0826-07 | 6194 | chr9 | 19378859 | 2335 | chr2 | 216259394 |
| RSBN1L | HGF | 皮肤黑素瘤 | TCGA-EB-A5SE-01A-11R-A311-07 | 222194 | chr7 | 77326339 | 3082 | chr7 | 81381543 |
| RUNX1 | CEP76 | 肺腺癌 | TCGA-O1-A52J-01A-11R-A262-07 | 861 | chr21 | 36206707 | 79959 | chr18 | 12686449 |
| SEC16A | RXRA | 肉瘤 | TCGA-DX-A3LS-01A-11R-A21T-07 | 9919 | chr9 | 139352034 | 6256 | chr9 | 137293495 |
| TBL1XR1 | CTNNB1 | 卵巢浆液性囊腺癌 | TCGA-24-2280-01A-01R-1568-13 | 79718 | chr3 | 176914909 | 1499 | chr3 | 41274832 |
| TMPRSS2 | GUCA2A | 前列腺腺癌 | TCGA-KC-A4BR-01A-32R-A32Y-07 | 7113 | chr21 | 42879877 | 2980 | chr1 | 42629281 |
| TMPRSS2 | TMEM109 | 前列腺腺癌 | TCGA-HC-8216-01A-11R-A29R-07 | 7113 | chr21 | 42860321 | 79073 | chr11 | 60687158 |
| TMPRSS2 | TMEM109 | 前列腺腺癌 | TCGA-HC-8213-01A-11R-A29R-07 | 7113 | chr21 | 42852403 | 79073 | chr11 | 60687197 |
| TMPRSS2 | BRAF | 前列腺腺癌 | TCGA-FC-A6HD-01A-11R-A31N-07 | 7113 | chr21 | 42866283 | 673 | chr7 | 140481493 |
| USP7 | PRKCB | 结肠和直肠腺癌 | TCGA-A6-3807-01A-01R-1022-07 | 7874 | chr16 | 9057064 | 5579 | chr16 | 24043457 |
| VEGFA | STK38 | 低级别胶质瘤 | TCGA-TM-A7CF-02A-11R-A32Q-07 | 7422 | chr6 | 43745395 | 11329 | chr6 | 36467720 |
| VEGFA | STK38 | 低级别胶质瘤 | TCGA-TM-A7CF-01A-11R-A32Q-07 | 7422 | chr6 | 43745395 | 11329 | chr6 | 36467720 |

[0938]

| 5'基因符号 | 3'基因符号 | 断裂点序列 | SEQ ID NO |
|---|---|---|---|
| ACPP | AXL | GGTTGGCCCTGTGATCCCTCAAGACTGGTCCACGGAGTGTATGACCACAA\|ACAGCCAGTC | 258 |
| | | CACCAGCTGGTGAAGGAACCTTCAACTCCTGCCTTCTCGT | |
| ATM | CARD18 | GAATATCTTGATAAATGAGCAGTCAGCAGAACTTGTACATATAGATCTAG\|ACCAACTCTTG | 259 |
| | | CGTAAAAAGAGAAGAATTTTATCCATTCAGTGGGTGCA | |
| ATXN1 | ARG1 | AAGAGAGAAAGAGTGGATTTCAGCCTGCACGGATGGTCTTGAAACACAAA\|TGGTCTGCT | 260 |
| | | TGAGAAACTTAAAGAACAAGAGTGTGATGTGAAGGATTATG | |
| CLU | CST3 | ACAGAATTCATACGAGAAGGCGACGATGACCGGACTGTGTGCCGGGAGAT\|CCGCCCGCGC | 261 |
| | | CTAGTGGGAGGCCCCATGGACGCCAGCGTGGAGGAGGAGGG | |
| CLU | CST3 | CAGAATTCATACGAGAAGGCGACGATGACCGGACTGTGTGCCGGGAGATC\|CGCCTAGGG | 262 |
| | | GTCTGTACCGGGCTGGCCTGTGCCTATCACCTCTTATGCAC | |
| DHX15 | ETV1 | GGACCTAGGGGAGGATTACCCCTCTGGCAAGAAGCGTGCGGGGACCGATG\|GTGCCTATGA | 263 |
| | | TCAGAAGCCACAAGTGGGAATGAGGCCCTCCAACCCCCCC | |
| DOT1L | TECR | AGTTCAGGAAGTGGATGAAATGGTATGGAAAAAAAGCATGCAGAATACACA\|GTGGAGATT | 264 |
| | | CTGGACGCAAAGACAAGGGAGAAGCTGTGTGTTTCTTGGACAA | |
| DOT1L | TECR | AGTTCAGGAAGTGGATGAAATGGTATGGAAAAAAAGCATGCAGAATACACA\|GTGGAGATT | 265 |
| | | CTGGACGCAAAGACAAGGGAGAAGCTGTGTGTTTCTTGGACAA | |
| EXOSC10 | MTOR | ATCAACGAAATGCACCTTTTAATCCAGCAGGCCCGAGAGATGCCCCTGCT\|CAAGGCCTGG | 266 |
| | | CATGCGTGGGCAGTGATGAACTTCGAAGCTGTGTGCTACACT | |
| FAT1 | NTRK3 | TCCATTAGAGATGGCTCTGGCGTTGGTGTTTTTCAAAATAGGTGAAGAGAC\|AGGCAGTTGG | 267 |
| | | AGCAGAACTTTTTCAACTGCAGCTGTGACATCCGCTGGAT | |
| FHIT | RUNX1 | GTCCCTCAGGCGGCCACCCAGTGGGCACACTCCCAGGCGGCGCTCC\|GGCCCCGAGAACC | 268 |
| | | TCGAAGACATCGGCAGAAACTAGATGATCAGACCAAG | |
| GUCA2A | ERG | AGGCACACTCAAACAACGACTGGTCCTCACTCACAACTGATAAGGCTTCC\|CAGCCTCTGA | 269 |
| | | AGTATCTCCTGGGCATTGGGCTCCTTGCAGAGAGGCTTGA | |
| HMHA1 | STK11 | GCTGTGGCCTTCATGTCCTGTGCCCCAGAGATCTGCCCCTCCCACCCGAG\|GGAAATTCAA | 270 |
| | | CTACTGAGGAGGTTACGGCACAAAAATGTCATCCAGCTGG | |
| STK11 | HMHA1 | GAGTACGAACCGGCCAAGAGGTTCTCCATCCGGCAGATCCGGCAGCACAG\|GTCCCCCGT | 271 |
| | | CATGCGTGCCCGGAAGAGCAGCTTCAACGTGAGTGATGTGG | |
| KANK2 | ALK | CCGGCCCAGAGCAAGGAAGTGTTATCATTCCAGCCACCTCCACCACCTGC\|TGCTCCATCG | 272 |
| | | CAGCTTCGCTCTGTGATGCTAATCTTCTTCACCATATGGA | |
| LMNA | RAF1 | AAGAAGTGGCCATGCGCAAGCTGGTGCGCTCAGTGACTGTGGTTGAGGAC\|ACAGCAGGA | 273 |
| | | TGATTGAGAATAACAACCTGAGTGCTTCTCCCAGGGCGTGG | |

[0939]

| | | | |
|---|---|---|---|
| MDM2 | BRI3BP | GGAAAATATATACCATGATCTACAGGAACTTGGTAGTAGTCAATCAGCAG\|GCTGGAGCACC | 274 |
| | | TGGAGAAGCAGGTCAGACTGCTCAACATCCGTCTCAACC | |
| MX1 | ERG | TGGAGAGGCAAGGTCAGTTACCAGGACTACGAGATTGAGATTTCGGATGC\|TTCAGACTGT | 275 |
| | | CCCGGACCCAGCAGCTCATATCAAGGAAGCCTTATCAGTT | |
| OXR1 | MET | CTGAAACAATTGAGGATTCTAGTAATCAAGCAGCAGCCAGAGAATGGGAG\|TGGAAGCAA | 276 |
| | | GCAATTTCTTCAACCGTCCTTGGAAAAGTAATAGTTCAACC | |
| PLS1 | PIK3CB | TGCTCTGGTGGATATGTTTGTTTTAGCCAGTCCATCCAAGACAATCCCAC\|CGCTGTCCTGC | 277 |
| | | CTTCGAGAGGGCAGCTCCCACTGGAGATCCAAGTACCTG | |
| RPS6 | FN1 | GGTCCGAATCAGTGGTGGGAACGACAAACAAGGTTTCCCCATGAAGCAGG\|GAAATTCTT | 278 |
| | | TGGAAGAAGTGGTCCATGCTGATCAGAGCTCCTGCACTTTT | |
| RSBN1L | HGF | GAGGAGGCACGGTCTCGGTGGGGCCCGAGAGGCCGGCGGGGCCTCCCGGG\|AGGAAAAC | 279 |
| | | TACTGTCGAAATCCTCGAGGGGAAGAAGGGGGACCCTGGTGT | |
| RUNX1 | CEP76 | CCCTGAACCACTCCACTGCCTTTAACCCTCAGCCTCAGAGTCAGATGCAG\|GGATGAAAAT | 280 |
| | | GGGATAAATAGACCAGTCTGTTCCTATGTTAAACCACTTC | |
| SEC16A | RXRA | ACGCCCAGTCCCTGGGTGCCGAGACCTGCCCCCTGCCTAGTTTCCAGGTG\|AACTCCTCCC | 281 |
| | | TCACCTCCCCGACGGGGCGAGGCTCCATGGCTGCCCCCTC | |
| TBL1XR1 | CTNNB1 | CAAGACGTTGACTTGGATCTGTCAGGTGAAGTCCTAAAGCTTGCATTCCA\|CCTGGAAACG | 282 |
| | | GTGGCCTCCAACGCCGCTCCCCCCTCCCGGGAATGGAGGC | |
| TMPRSS2 | GUCA2A | AGGTCTTTGAGCTTCTTCACTGACTCCAGAGAAAAGGAGAAATTTCCATC\|CTGCCGCGCC | 283 |
| | | GCGCTCCTCACACCCGCTTTCACCTCCGGGCGGGGCAGGG | |
| TMPRSS2 | TMEM109 | GTGTGATGGCGTGTCACACTGCCCCGGCGGGGAGGACGAGAATCGGTGTG\|ACCCAGTCA | 284 |
| | | TGGCAGCCTCCAGCATCAGTTCACCATGGGGAAAGCATGTG | |
| TMPRSS2 | TMEM109 | GACTGGAACGAGAACTACGGGCGGGCCGGCCTGCAGGGACATGGGCTATAA\|GAAAGCATG | 285 |
| | | TGTTCAAAGCCATTCTGATGGTCCTAGTGGCCCTTATCCTC | |
| TMPRSS2 | BRAF | CGTCGTCTGCACGCAGCCCAAATCCCCATCCGGGACAGTGTGCACCTCAA\|AAAACACTTG | 286 |
| | | GTAGACGGGACTCGAGTGATGATTGGGAGATTCCTGATGG | |
| USP7 | PRKCB | GAAAGCGGGCGAGCAGTTGAGCGAGCCCGAGGACATGGAGATGGAAG\|GACCCCCG | 287 |
| | | CAGCAAACACAAGTTTAAGATCCACACGTACTCCAGCCCCAC | |
| VEGFA | STK38 | TGCAATGACGAGGGCCTGGAGTGTGTGCCCACTGAGGAGTCCAACATCAC\|CATGCAGAC | 288 |
| | | CGGGTACAACAAGCTCTGTGATTGGTGGTCGCTTGGGGTGA | |
| VEGFA | STK38 | TGCAATGACGAGGGCCTGGAGTGTGTGCCCACTGAGGAGTCCAACATCAC\|CATGCAGAC | 289 |
| | | CGGGTACAACAAGCTCTGTGATTGGTGGTCGCTTGGGGTGA | |

[0940] 表43

[0941]

| 基因 | 已批准的 | 注册前（批准前） | III 期 |
|---|---|---|---|
| AXL | | | |
| ARG1 | | | |
| CLU | | | 库司替森（custirsen） |
| DOT1L | | | |
| MTOR | 依维莫司（everolimus），坦罗莫司（temsirolimus） | 地磷莫司（ridaforolimus） | N |
| NTRK3 | | | |
| ALK | 克卓替尼（crizotinib） | 盐酸艾乐替尼（alectinib hydrochloride） | 色瑞替尼（ceritinib） |
| RAF1 | 索拉非尼（sorafenib） | N | N |
| MDM2 | N | N | N |
| MET | 卡博替尼（cabozantinib），克卓替尼 | N | 提瓦替尼（tivantinib），里乐木单抗（rilotumumab），奥那组单抗（onartuzumab） |
| PIK3CB | | | |
| FN1 | | | |
| HGF | | | 里乐木单抗 |
| RXRA | 贝沙罗汀（bexarotene），贝沙罗汀凝胶 | | |
| CTNNB1 | | | |
| PRKCB | | | |
| BRAF | 维罗非尼（vemurafenib），帕唑帕尼（pazopanib），达拉菲尼（dabrafenib） | | 恩可非尼（encorafenib） |
| VEGFA | 贝伐单抗（bevacizumab），哌加他尼八钠（pegaptanib octasodium），哌加他尼八钠 | | 米哚妥林（midostaurin）、XL-647、贝伐单抗，Biocad；贝伐单抗，Actavis；贝伐单抗，Reliance Life Sciences |

[0942]

| 基因 | II 期 | I 期 | 临床前 |
|---|---|---|---|
| AXL | 弗雷替尼（foretinib） | BGB-324、S-49076、宁格替尼（ningetinib） | TP-0903、SGI-7079、Q-4、BGB-109、CEP-40783 |
| ARG1 | PEG-精氨酸酶，Bio-Cancer | | |
| CLU | | | AB-16B5、CGEN-25008 |
| DOT1L | | EPZ-01 | DOT1L 抑制剂，Aurigene |
| MTOR | 奎纳克林（quinacrine）；XL-765；达妥昔布（dactolisib）；PKI-587；PF-04691502；CC-223、阿皮托利西（apitolisib）、格达托利西（gedatolisib）、INK-128 | P-7170；CBLC-137、AZD-2014；CC-115；PWT-33957；DS-7423；GDC-0084；DS-3078；LY-3023414；PI3 激酶/mTOR 抑制剂，Lilly；SF-1126、SB-2343 | nPT-MTOR；STP-503；X-480；ABTL-0812；X-414；CC214；HMPL-518；PQR-309；PQR-401；mTOR 抑制剂/PI3 激酶抑制剂，Lilly-1；PIM/PI3k/mTOR 抑制剂，Inflection Biosciences |
| NTRK3 | TSR-011 | PLX-7486 | NMS-P626 |
| ALK | AP-26113；RG-7853；TSR-011；NMS-E628，PF-06463922 | X-396；ASP-3026；CEP-37440 | NMS-E628；极光激酶 +ALK 抑制剂（Sareum、AstraZeneca）；ALK 抑制剂（AstraZeneca、Cephalon、Aurigene）；ARN-5032；DLX-521、TL-398、AZD-3463 |
| RAF1 | iCo-007；XL-281 | RO-5126766；MLN-2480、泛 Raf 抑制剂，Eli Lilly | BIB-024；STP503；DP-4978；HM-95573；TAK-632 |
| MDM2 | N | DS-3032、AMG-232 | p53-mdm2/mdm4 双重抑制剂，Adamed；PXN-527；ATSP-7041；MDM2 抑制剂，Amgen；DS-5272 |
| MET | MGCD-265；弗雷替尼；费拉妥珠单抗（ficlatuzumab）；BMS-777607；格瓦替尼（golvatinib）；INCB-028060；LY-2875358；阿皮托利西 | AMG-208；TAS-115；沃利替尼（volitinib）；SAR-125844；S-49076、麦他替尼（metatinib）、F-50064、ARGX-111、DCC-2701 | X-379；麦他替尼；PRS-110；ASP-08001；DCC-2721；MG-516；AL-2846；CG-206481；T-1840383；cMet-EGFR 双重抑制剂（CrystalGenomics）；双特异性抗体（Hoffmann-La Roche） |
| PIK3CB | 皮克立西（pictilisib）、GS-9820、GSK-2636771 | INK-1117、BAY-1082439、SAR-260301、AZD-8186 | PQR-3xx，Piqur |
| FN1 | 雷曲妥单抗（radretumab）、达路金（darleukin） | AS-1409 | |
| HGF | | HuL2G7 | NK4，Kringle |

| | | | | |
|---|---|---|---|---|
| | RXRA | | | IRX-4310 |
| | CTNNB1 | PRI-724 | | β-链蛋白抑制剂，PhaseRx；K-756，DsiRNAs，Dicerna |
| | PRKCB | 索曲妥林（sotrastaurin） | | |
| | BRAF | RAF-265、XL-281、AB-024 | ARQ-761、ARQ-736、BeiGene-283 | b-raf 抑制剂，Sareum；B-Raf 和 c-RAF 双重抑制剂，Redx Pharma；PLX-8394、DP-4978、TL-241、B-Raf 激酶抑制剂，Array BioPharma-1；EN-3352、EBI-907、PI3K/BET 布罗莫结构域抑制剂，SignalRx |
| [0943] | VEGFA | 角鲨胺（squalamine）、PTC-299、MP-0112、ESBA-1008 | ASC-06、PRS-050、赛伐单抗（sevacizumab）、RG-7221、PAN-90806、TAS-115、贝伐单抗，Pfizer | VEGF 疫苗，Neovacs；CEQ-300；贝伐单抗，BioXpress；贝伐单抗，Natco；贝伐单抗，Oncobiologics；贝伐单抗，Harvest Moon；OMP-305B83；贝伐单抗，Fujifilm Kyowa Kirin Biologics；贝伐单抗，LG Life Sciences；贝伐单抗，Biocon；贝伐单抗，Mabion；贝伐单抗，Aryogen；hVEGF-截短疫苗，Immunovo；贝伐单抗，Avesthagen；贝伐单抗，Alteogen；贝伐单抗，Nanogen |

[0944] 表44

| 癌症类型 | 基因符号 | Entrez 基因 ID | 参考等位基因 | 肿瘤序列等位基因 1 | 肿瘤序列等位基因 2 | 变异体变化 | 变异体位置 | 变异体分类 | 可药化性 |
|---|---|---|---|---|---|---|---|---|---|
| 胃腺癌 | CDKN2A | 1029 | A | C | C | p.V25G | p.V25 | 错义突变 | Y |
| 透明细胞肾细胞癌 | CDKN2A | 1030 | A | C | C | p.V25G | p.V25 | 错义突变 | Y |
| 多发性骨髓瘤 | CDKN2A | 1031 | A | C | | p.V25G | p.V25 | 错义突变 | Y |
| 前列腺腺癌 | CDKN2A | 1032 | A | A | C | p.V25G | p.V25 | 错义突变 | Y |
| 肺腺癌 | ALK | 238 | C | G | G | p.W247S | p.W247 | 错义突变 | Y |

[0945]

[0946] 表45

[0947]

| 子集 | 事件类型 | Q | 阳性数量 | 患者的总数量 |
|---|---|---|---|---|
| 急性骨髓性白血病 | 融合体 | 4.03E-02 | 11 | 168 |
| 嫌色性肾细胞癌 | 失去功能的突变 | 6.84E-02 | 18 | 65 |
| 嫌色性肾细胞癌 | 失去功能的突变 | 6.92E-02 | 4 | 65 |
| 结肠直肠腺癌:KRAS 突变，3 或 4 期 | 获得功能的突变 | 3.86E-02 | 8 | 45 |
| 结肠直肠腺癌:微卫星稳定 | 获得功能的突变 | 9.10E-02 | 5 | 21 |
| 弥漫型胃腺癌 | 峰内基因扩增 | 3.04E-02 | 4 | 52 |
| 导管乳癌:三阴性 | 失去功能的突变 | 9.90E-02 | 51 | 68 |
| 胃肠型腺癌 | 峰内基因扩增 | 3.41E-02 | 4 | 44 |
| 浸润性膀胱尿道上皮癌 | 峰内基因扩增 | 7.16E-02 | 4 | 183 |
| 平滑肌肉瘤 | 峰内基因缺失 | 1.86E-03 | 6 | 40 |
| 肺腺癌:EGFR 突变 | 失去功能的突变 | 9.57E-03 | 13 | 21 |
| 胰腺导管腺癌 | 获得功能的突变 | 8.78E-04 | 5 | 39 |
| 胰腺导管腺癌 | 失去功能的突变 | 5.97E-02 | 5 | 39 |
| 前列腺腺癌 | 失去功能的突变 | 1.25E-03 | 14 | 172 |
| 结肠直肠腺癌:KRAS 突变 | 获得功能的突变 | 3.39E-04 | 10 | 105 |

[0948]

| 染色体区带 | 基因（Entrez 基因 ID） | 可药化基因 | KM 证据 |
|---|---|---|---|
| 17q21, 15q22 | RARA (5914), PML (5371) | RARA | 有利的结果 |
| 17p13 | TP53 (7157) | TP53 | 较差的结果 |
| 10q23 | PTEN (5728) | PTEN | 较差的结果 |
| 7q32 | IRF5 (3663) | | 较差的结果 |
| 22q12 | NEFH (4744) | | 较差的结果 |
| 11p13 | APIP (51074) | | 较差的结果 |
| 17p13 | TP53 (7157) | TP53 | 有利的结果 |
| 10q26 | FGFR2 (2263) | FGFR2 | 较差的结果 |
| 20q11 | E2F1 (1869) | E2F1 | 较差的结果 |
| 9p21 | CDKN2A (1029), CDKN2B (1030) | CDKN2A | 较差的结果 |
| 17p13 | TP53 (7157) | TP53 | 较差的结果 |
| 7q22 | PTCD1 (26024), ATP5J2-PTCD1 (100526740) | | 较差的结果 |
| 4q35 | DUX2 (26583) | | 较差的结果 |
| 17p13 | TP53 (7157) | TP53 | 较差的结果 |
| 7q32 | IRF5 (3663) | | 较差的结果 |

[0949] 表46

| 子集 | 事件类型 | Q |
|---|---|---|
| 急性骨髓性白血病 | 融合体 | 4.03E-02 |
| 嫌色性肾细胞癌 | 失去功能的突变 | 6.84E-02 |
| 嫌色性肾细胞癌 | 失去功能的突变 | 6.92E-02 |
| 结肠直肠腺癌:KRAS 突变，3 或 4 期 | 获得功能的突变 | 3.86E-02 |
| 结肠直肠腺癌:微卫星稳定 | 获得功能的突变 | 9.10E-02 |
| 弥漫型胃腺癌 | 峰内基因扩增 | 3.04E-02 |
| 导管乳癌:三阴性 | 失去功能的突变 | 9.90E-02 |
| [0950] 胃肠型腺癌 | 峰内基因扩增 | 3.41E-02 |
| 浸润性膀胱尿道上皮癌 | 峰内基因扩增 | 7.16E-02 |
| 平滑肌肉瘤 | 峰内基因缺失 | 1.86E-03 |
| 肺腺癌:EGFR 突变 | 失去功能的突变 | 9.57E-03 |
| 胰腺导管腺癌 | 获得功能的突变 | 8.78E-04 |
| 胰腺导管腺癌 | 失去功能的突变 | 5.97E-02 |
| 前列腺腺癌 | 失去功能的突变 | 1.25E-03 |
| 结肠直肠腺癌:KRAS 突变 | 获得功能的突变 | 3.39E-04 |
| 结肠直肠腺癌:KRAS 突变 | 获得功能的突变 | 1.53E-02 |

| 子集 | 阳性数量 | 患者的总数量 | 染色体区带 | 基因（Entrez 基因 ID） |
|---|---|---|---|---|
| 急性骨髓性白血病 | 11 | 168 | 17q21, 15q22 | RARA (5914), PML (5371) |
| 嫌色性肾细胞癌 | 18 | 65 | 17p13 | TP53 (7157) |
| 嫌色性肾细胞癌 | 4 | 65 | 10q23 | PTEN (5728) |
| 结肠直肠腺癌:KRAS 突变，3 或 4 期 | 8 | 45 | 7q32 | IRF5 (3663) |
| 结肠直肠腺癌:微卫星稳定 | 5 | 21 | 22q12 | NEFH (4744) |
| 弥漫型胃腺癌 | 4 | 52 | 11p13 | APIP (51074) |
| 导管乳癌:三阴性 | 51 | 68 | 17p13 | TP53 (7157) |
| [0951] 胃肠型腺癌 | 4 | 44 | 10q26 | FGFR2 (2263) |
| 浸润性膀胱尿道上皮癌 | 4 | 183 | 20q11 | E2F1 (1869) |
| 平滑肌肉瘤 | 6 | 40 | 9p21 | CDKN2A (1029), CDKN2B (1030) |
| 肺腺癌:EGFR 突变 | 13 | 21 | 17p13 | TP53 (7157) |
| 胰腺导管腺癌 | 5 | 39 | 7q22 | PTCD1 (26024), ATP5J2-PTCD1 (100526740) |
| 胰腺导管腺癌 | 5 | 39 | 4q35 | DUX2 (26583) |
| 前列腺腺癌 | 14 | 172 | 17p13 | TP53 (7157) |
| 结肠直肠腺癌:KRAS 突变 | 10 | 105 | 7q32 | IRF5 (3663) |
| 结肠直肠腺癌:KRAS 突变 | 4 | 105 | 5q31 | PCDHA7 (56141) |

| 子集 | 可药化基因 | KM 证据 |
|---|---|---|
| 急性骨髓性白血病 | RARA | 有利的结果 |
| 嫌色性肾细胞癌 | TP53 | 较差的结果 |
| 嫌色性肾细胞癌 | PTEN | 较差的结果 |
| 结肠直肠腺癌:KRAS 突变，3 或 4 期 | | 较差的结果 |
| 结肠直肠腺癌:微卫星稳定 | | 较差的结果 |
| 弥漫型胃腺癌 | | 较差的结果 |
| 导管乳癌:三阴性 | TP53 | 有利的结果 |
| 胃肠型腺癌 | FGFR2 | 较差的结果 |
| 浸润性膀胱尿道上皮癌 | E2F1 | 较差的结果 |
| 平滑肌肉瘤 | CDKN2A | 较差的结果 |
| 肺腺癌:EGFR 突变 | TP53 | 较差的结果 |
| 胰腺导管腺癌 | | 较差的结果 |
| 胰腺导管腺癌 | | 较差的结果 |
| 前列腺腺癌 | TP53 | 较差的结果 |
| 结肠直肠腺癌:KRAS 突变 | | 较差的结果 |
| 结肠直肠腺癌:KRAS 突变 | | 较差的结果 |

[0952]

[0953] 表47

[0954]

| 癌症类型 | 基因符号 | Entrez基因ID |
|---|---|---|
| 乳头状肾细胞癌 | AR | 367 |
| 小叶乳癌 | AR | 367 |
| 肾上腺皮质癌 | AR | 367 |
| 结肠直肠腺癌 | AR | 367 |
| 胰脏癌 | AR | 367 |
| 结肠直肠腺癌 | AR | 367 |
| 胰腺导管腺癌 | AR | 367 |
| 前列腺癌 | JAK1 | 3716 |
| 肝细胞癌 | JAK1 | 3716 |

[0955] 表47

[0956]

| 癌症类型 | 参考等位基因 |
|---|---|
| 乳头状肾细胞癌 | - |
| 小叶乳癌 | - |
| 肾上腺皮质癌 | GCAGCAGCAGCAGCAGCAGCAGCA |
| 结肠直肠腺癌 | - |
| 胰脏癌 | GCAGCAGCA |
| 结肠直肠腺癌 | GCAGCAGCAGCAGCAGCAGCAGCAGCA |
| 胰腺导管腺癌 | GCAGCAGCA |
| 前列腺癌 | C |
| 肝细胞癌 | T |

[0957] 表47

[0958]

| 癌症类型 | 肿瘤序列等位基因1 |
|---|---|
| 乳头状肾细胞癌 | - |
| 小叶乳癌 | - |
| 肾上腺皮质癌 | GCAGCAGCAGCAGCAGCAGCA |
| 结肠直肠腺癌 | - |
| 胰脏癌 | GCAGCAGCA |
| 结肠直肠腺癌 | GCAGCAGCAGCAGCAGCAGCAGCA |
| 胰腺导管腺癌 | GCAGCAGCA |
| 前列腺癌 | T |
| 肝细胞癌 | T |

[0959] 表47

[0960]

| 癌症类型 | 肿瘤序列等位基因 2 | 变异体变化 | 变异体位置 |
|---|---|---|---|
| 乳头状肾细胞癌 | GCAGCA | p.L57_框内插入 | p.L57 |
| 小叶乳癌 | GCA | p.L57_框内插入 | p.L57 |
| 肾上腺皮质癌 | - | p.L57_框内缺失 | p.L57 |
| 结肠直肠腺癌 | GCAGCA | p.L57_框内插入 | p.L57 |
| 胰脏癌 | - | p.L57_框内缺失 | p.L57 |
| 结肠直肠腺癌 | - | p.L57_框内缺失 | p.L57 |
| 胰腺导管腺癌 | - | p.L57_框内缺失 | p.L57 |

[0961]

| | | | |
|---|---|---|---|
| 前列腺癌 | - | p.E188K | p.E188 |
| 肝细胞癌 | C;A | p.E188G；p.E188V | p.E188 |

[0962] 表47

[0963]

| 癌症类型 | 变异体分类 | 泛癌症CBI变异体类别 | 可药化性 |
|---|---|---|---|
| 乳头状肾细胞癌 | 框内插入 | 热点 | Y |
| 小叶乳癌 | 框内插入 | 热点 | Y |
| 肾上腺皮质癌 | 框内缺失 | 热点 | Y |
| 结肠直肠腺癌 | 框内插入 | 热点 | Y |
| 胰脏癌 | 框内缺失 | 热点 | Y |
| 结肠直肠腺癌 | 框内缺失 | 热点 | Y |
| 胰腺导管腺癌 | 框内缺失 | 热点 | Y |
| 前列腺癌 | 错义突变 | 热点 | Y |
| 肝细胞癌 | 错义突变 | 热点 | Y |

[0964] 表48

[0965]

| 癌症类型 | 5’基因符号 | 3’基因符号 | 可药化基因 |
|---|---|---|---|
| 食道癌 | NUP214 | ABL1 | ABL1 |
| 结肠和直肠腺癌 | SMEK2 | ALK | ALK |
| 皮肤黑素瘤 | KCNQ5 | ALK | ALK |
| 结肠和直肠腺癌 | TRIM24 | BRAF | BRAF |
| 膀胱尿道上皮癌 | CDKN2A | CAPZB | CDKN2A |

| | | | |
|---|---|---|---|
| 结肠和直肠腺癌 | PEX1 | CDK6 | CDK6 |
| 低级别胶质瘤 | EGFR | C7orf72 | EGFR |
| 肝细胞癌 | ERBB2 | PPP1R1B | ERBB2 |
| 前列腺腺癌 | ERBB2 | PPP1R1B | ERBB2 |
| 肝细胞癌 | FGFR2 | BICC1 | FGFR2 |
| 低级别胶质瘤 | FGFR3 | FBXO28 | FGFR3 |
| 前列腺腺癌 | FGFR3 | AES | FGFR3 |
| 食道癌 | FGFR3 | TACC3 | FGFR3 |
| 宫颈鳞状细胞癌 | JAK2 | RCL1 | JAK2 |
| 胃腺癌 | ZBTB17 | MTOR | MTOR |
| 子宫内膜子宫内膜样腺癌 | KAZN | MTOR | MTOR |
| 低级别胶质瘤 | EPHB2 | NTRK1 | NTRK1 |
| 食道癌 | RAF1 | C9orf153 | RAF1 |
| 胰腺癌 | RAF1 | TMEM40 | RAF1 |
| 鳞状细胞肺癌 | CDKN2A | SRGAP2 | CDKN2A |
| 肝细胞癌 | SLC7A2 | TERT | TERT |

[0966] 表49

[0967]

| 5’基因符号 | 3’基因符号 | 癌症类型 | TCGA肿瘤样品条形码 |
|---|---|---|---|
| NUP214 | ABL1 | 食道癌 | TCGA-L5-A4OQ-01A-11R-A260-31 |
| FGFR3 | AES | 前列腺腺癌 | TCGA-EJ-A7NM-01A-21R-A33R-07 |
| FGFR3 | AES | 前列腺腺癌 | TCGA-EJ-A7NM-01A-21R-A33R-07 |
| FGFR3 | AES | 前列腺腺癌 | TCGA-EJ-A7NM-01A-21R-A33R-07 |
| FGFR3 | AES | 前列腺腺癌 | TCGA-EJ-A7NM-01A-21R-A33R-07 |
| SMEK2 | ALK | 结肠和直肠腺癌 | TCGA-F5-6864-01A-11R-1928-07 |
| KCNQ5 | ALK | 皮肤黑素瘤 | TCGA-EB-A6QY-01A-12R-A32P-07 |
| FGFR2 | BICC1 | 肝细胞癌 | TCGA-ED-A7PX-01A-51R-A352-07 |
| FGFR2 | BICC1 | 肝细胞癌 | TCGA-ED-A7PX-01A-51R-A352-07 |
| EGFR | C7orf72 | 低级别胶质瘤 | TCGA-E1-A7YJ-01A-11R-A34F-07 |
| RAF1 | C9orf153 | 食道癌 | TCGA-L5-A4OS-01A-11R-A28J-31 |
| CDKN2A | CAPZB | 膀胱尿道上皮癌 | TCGA-E7-A6MD-01A-41R-A352-07 |
| CDKN2A | CAPZB | 膀胱尿道上皮癌 | TCGA-E7-A6MD-01A-41R-A352-07 |
| CDKN2A | CAPZB | 膀胱尿道上皮癌 | TCGA-E7-A6MD-01A-41R-A352-07 |
| PEX1 | CDK6 | 结肠和直肠腺癌 | TCGA-AA-3697-01A-01R-1723-07 |
| FGFR3 | FBXO28 | 低级别胶质瘤 | TCGA-TM-A84B-01A-11R-A36H-07 |
| FGFR3 | FBXO28 | 低级别胶质瘤 | TCGA-TM-A84B-01A-11R-A36H-07 |
| FGFR3 | FBXO28 | 低级别胶质瘤 | TCGA-TM-A84B-01A-11R-A36H-07 |
| FGFR3 | FBXO28 | 低级别胶质瘤 | TCGA-TM-A84B-01A-11R-A36H-07 |
| FGFR3 | FBXO28 | 低级别胶质瘤 | TCGA-TM-A84B-01A-11R-A36H-07 |
| FGFR3 | FBXO28 | 低级别胶质瘤 | TCGA-TM-A84B-01A-11R-A36H-07 |
| FGFR3 | FBXO28 | 低级别胶质瘤 | TCGA-TM-A84B-01A-11R-A36H-07 |
| FGFR3 | FBXO28 | 低级别胶质瘤 | TCGA-TM-A84B-01A-11R-A36H-07 |

| | | | |
|---|---|---|---|
| ZBTB17 | MTOR | 胃腺癌 | TCGA-BR-4357-01A-01R-1157-13 |
| KAZN | MTOR | 子宫内膜子宫内膜样腺癌 | TCGA-D1-A3JQ-01A-11R-A22K-07 |
| EPHB2 | NTRK1 | 低级别胶质瘤 | TCGA-TM-A84J-01A-11R-A36H-07 |
| ERBB2 | PPP1R1B | 肝细胞癌 | TCGA-KR-A7K2-01A-12R-A33R-07 |
| ERBB2 | PPP1R1B | 胰腺癌 | TCGA-LB-A7SX-01A-11R-A33R-07 |
| JAK2 | RCL1 | 宫颈鳞状细胞癌 | TCGA-PN-A8MA-01A-11R-A36F-07 |
| JAK2 | RCL1 | 宫颈鳞状细胞癌 | TCGA-PN-A8MA-01A-11R-A36F-07 |
| CDKN2A | SRGAP2 | 鳞状细胞肺癌 | TCGA-60-2715-01A-01R-0851-07 |
| CDKN2A | SRGAP2 | 鳞状细胞肺癌 | TCGA-60-2715-01A-01R-0851-07 |
| FGFR3 | TACC3 | 食道癌 | TCGA-LN-A5U5-01A-21R-A28J-31 |
| SLC7A2 | TERT | 肝细胞癌 | TCGA-DD-A73C-01A-12R-A33J-07 |
| RAF1 | TMEM40 | 胰腺癌 | TCGA-IB-A5SS-01A-11R-A32O-07 |
| BRAF | TRIM24 | 结肠和直肠腺癌 | TCGA-F5-6464-01A-11R-1736-07 |
| TRIM24 | BRAF | 结肠和直肠腺癌 | TCGA-F5-6464-01A-11R-1736-07 |

[0968] 表49

[0969]

| 5’基因符号 | 5’登录号 | 5’染色体 | 5’断裂点 | 3’登录号 | 3’染色体 | 3’断裂点 |
|---|---|---|---|---|---|---|
| NUP214 | 8021 | chr9 | 134027281 | 25 | chr9 | 133747516 |
| FGFR3 | 2261 | chr4 | 1808969 | 166 | chr19 | 3061232 |
| FGFR3 | 2261 | chr4 | 1808905 | 166 | chr19 | 3061233 |
| FGFR3 | 2261 | chr4 | 1808905 | 166 | chr19 | 3061233 |
| FGFR3 | 2261 | chr4 | 1808661 | 166 | chr19 | 3061255 |
| SMEK2 | 57223 | chr2 | 55804451 | 238 | chr2 | 29940563 |
| KCNQ5 | 56479 | chr6 | 73332315 | 238 | chr2 | 29498362 |
| FGFR2 | 2263 | chr10 | 123239533 | 80114 | chr10 | 60461846 |
| FGFR2 | 2263 | chr10 | 123244971 | 80114 | chr10 | 60461846 |
| EGFR | 1956 | chr7 | 55268106 | 100130988 | chr7 | 50143910 |
| RAF1 | 5894 | chr3 | 12627180 | 389766 | chr9 | 88844544 |
| CDKN2A | 1029 | chr9 | 21974677 | 832 | chr1 | 19671746 |
| CDKN2A | 1029 | chr9 | 21974673 | 832 | chr1 | 19666067 |
| CDKN2A | 1029 | chr9 | 21994138 | 832 | chr1 | 19671746 |
| PEX1 | 5189 | chr7 | 92151518 | 1021 | chr7 | 92355063 |
| FGFR3 | 2261 | chr4 | 1808661 | 23219 | chr1 | 224340844 |
| FGFR3 | 2261 | chr4 | 1808989 | 23219 | chr1 | 224340848 |
| FGFR3 | 2261 | chr4 | 1808645 | 23219 | chr1 | 224345163 |
| FGFR3 | 2261 | chr4 | 1808645 | 23219 | chr1 | 224345163 |
| FGFR3 | 2261 | chr4 | 1808586 | 23219 | chr1 | 224340892 |
| FGFR3 | 2261 | chr4 | 1808591 | 23219 | chr1 | 224318276 |
| FGFR3 | 2261 | chr4 | 1808661 | 23219 | chr1 | 224340844 |
| FGFR3 | 2261 | chr4 | 1808582 | 23219 | chr1 | 224340949 |
| ZBTB17 | 7709 | chr1 | 16269559 | 2475 | chr1 | 11227574 |
| KAZN | 23254 | chr1 | 15251068 | 2475 | chr1 | 11206848 |
| EPHB2 | 2048 | chr1 | 23111569 | 4914 | chr1 | 156843425 |

| | | | | | | |
|---|---|---|---|---|---|---|
| ERBB2 | 2064 | chr17 | 37883777 | 84152 | chr17 | 37790317 |
| ERBB2 | 2064 | chr17 | 37883798 | 84152 | chr17 | 37790337 |
| JAK2 | 3717 | chr9 | 5090911 | 10171 | chr9 | 4834141 |
| JAK2 | 3717 | chr9 | 5126446 | 10171 | chr9 | 4860125 |
| CDKN2A | 1029 | chr9 | 21994138 | 23380 | chr1 | 206634382 |
| CDKN2A | 1029 | chr9 | 21974677 | 23380 | chr1 | 206634382 |
| FGFR3 | 2261 | chr4 | 1808661 | 10460 | chr4 | 1741429 |
| SLC7A2 | 6542 | chr8 | 17354726 | 7015 | chr5 | 1282731 |
| RAF1 | 5894 | chr3 | 12632350 | 55287 | chr3 | 12791339 |
| BRAF | 673 | chr7 | 140624377 | 8805 | chr7 | 138255741 |
| TRIM24 | 8805 | chr7 | 138255748 | 673 | chr7 | 140550012 |

[0970] 表49

[0971]

| 5'基因符号 | 断裂点序列 |
|---|---|
| NUP214 | CATGACTGCAGCTTCTTTCAAGAACTCTTCCACCTCCATGGTGTCCTCCT\|GAAGCTGAGCTTCACTCTTGGGATCCAGTGGTCTTTTATAAAGCAAATGC |
| FGFR3 | GGACGACTCCGTGTTTGCCCACGACCTGCTGCCCCCGGCCCCACCCAGCA\|ACTCAAATTCACCACCTCGGACTCCTGCGACCGCATCAAAGACGAATTTC |
| FGFR3 | TGTCGGCGCCTTTCGAGCAGTACTCCCCGGGTGGCCAGGACACCCCCAGC\|AACTCAAATTCACCACCTCGGACTCCTGCGACCGCATCAAAG |
| FGFR3 | TGTCGGCGCCTTTCGAGCAGTACTCCCCGGGTGGCCAGGACACCCCCAGC\|AACTCAAATTCACCACCTCGGACTCCTGCGACCGCATCAAAGACGAATTT |
| FGFR3 | GCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTCCACCGACG\|GCTCCTCGCACCTACCCCAGCAACTCAAATTCACCACCTCGGACTCCTGC |
| SMEK2 | TGAAAAGGATAATATAGTTGGATCAAACAAAAACAACACAATTTGTCCCG\|GTCATAGCTCCTTGGAATCACCAACAAACATGCCTTCTCCTTCTCCTGAT |
| KCNQ5 | CGATGGCTCTGGATCCTTGTCCCCACCATGCGACCATCTGCAGCCAGAAC\|ACGAAAGCGTGGTAGATGAACGCCCAGCCGCGGGGTCTCTCCAGCACGTT |
| FGFR2 | AGTTGGTAGAAGACTTGGATCGAATTCTCACTCTCACAACCAATGAGGAA\|ACAAATACGCAGATTGCTTGGCCATCAAAACTGAAGATCGGAGCCAAATC |
| FGFR2 | AGATCTTCACTTTAGGGGGCTCGCCCTACCCAGGGATTCCCGTGGAGGAA\|ACAAATACGCAGATTGCTTGGCCATCAAAACTGAAGATCGGAGCCAAATC |
| EGFR | GATCATCGAATTCTCCAAAATGGCCCGAGACCCCCAGCGCTACCTTGTCA\|TTCAGGTTGGACAAGCCCTCTGAAAGTTACTCCCTTACAACCTCATCATG |
| RAF1 | GAGTGGTTCTCAGCAGGTTGAACAACCTACTGGCTCTGTCCTCTGGATGG\|AATTAGAGAATAAAATCCCAGCACGATGTTCCTCACTGGAGACACCAGTC |
| CDKN2A | TTGGCTATGTGTGGGGAGCAGTCACTCACAGTTTCATCCTTCTCCATCTG\|GATCGGCCTCCGACCGTAACTATTCGGTGCGTTGGGCAGCGCCCCCGCCT |
| CDKN2A | AGAGGTTTAGCATTGCTGCTTTCTCTTCAAAGCCTCCACCAGGTCATTCT\|CCACCTGGATCGGCCTCCGACCGTAACTATTCGGTGCGTTGGGCAGCGCC |
| CDKN2A | TTGGCTATGTGTGGGGAGCAGTCACTCACAGTTTCATCCTTCTCCATCTG\|GTCTTCTAGGAAGCGGCTGCTGCCCTAGACGCTGGCTCCTCAGTAGCATC |
| PEX1 | TGCTGCAGAATCAAGCTATAGAAGTGGTCTGGAGTCACCAGCCTGCATTC\|ACACCGAGTAGTGCATCGCGATCTAAAACCACAGAACATTCTGGTGACCA |
| FGFR3 | GCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTCCACCGACG\|TGATTGATGAGATTTATCGTGTGTTGAGATATGTCAATTCTACCAGAGCC |
| FGFR3 | ACGACCTGCTGCCCCCGGCCCCACCCAGCAGTGGGGGCTCGCGGACGTGA\|TTGATGAGATTTATCGTGTGTTGAGATATGTCAATTCTACCAGAGCCCCT |
| FGFR3 | CAGAGGCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTAC\|CGTGACTGTTCTCAGGCGTGAAATTTCTGAGCTTCGCACCAAAGTGCAAG |
| FGFR3 | CTTGCACTTTGGTGCGAAGCTCAGAAATTTCACGCCTGAGAACAGTCACG\|GTAAGGACACGGTCCAGGTCCTCCACCAGCTGCTTGAAGGTGGG |
| FGFR3 | CCAACTGCACACACGACCTGTACATGATCATGCGGGAGTGCTGGCATGCC\|GCCCCTCAACGAGCTCATGAAGTACTTCAAGAATTAAGGGATATATCCTC |
| FGFR3 | TGCACACACGACCTGTACATGATCATGCGGGAGTGCTGGCATGCCGCGCC\|CTCCCAAGGTGATTGATGAGATTTATCGTGTGTTGAGATATGTCAATTCT |
| FGFR3 | AGGGGCTCTGGTAGAATTGACATATCTCAACACACGATAAATCTCATCAA\|TCACGTCGGTGGACGTCACGGTAAGGACACGGTCCAGGTCCTCCACCAGC |
| FGFR3 | CCCGCCAACTGCACACACGACCTGTACATGATCATGCGGGAGTGCTGGCA\|ATGGAGTACTTTGATGAAAAGATTGTTCCAATTTTAAAGAGGAAATTACC |
| ZBTB17 | GGCCTTCGTGAACGTGGGGGACCTGTCCAAGCACATCATCATTCACACTG\|CATTAATAATAAGCTACAGCAGCCGGAGGCAGCGGCCGGAGTGTTAGAAT |

[0972]

| | |
|---|---|
| KAZN | CCCGCGCCGGGGTTCCCCGGGTCCGAGCGGATGGCGACTGCAGCCAGCCC\|GGTCAGTGGGACAGCATGGAAGAATACACCTGTATGATCCCTCGGGACAC |
| EPHB2 | CATGTGCAAAGCAGGCTTCGAGGCCGTTGAGAATGGCACCGTCTGCCGAG\|TCCCGGCCAGTGTGCAGCTGCACACGGCGGTGGAGATGCACCACTGGTGC |
| ERBB2 | CCTGGCCACAGGTTGTCTTTTGCCCAGCAGACTGCCTGATGACCTTCAGG\|GGGGCAACGTAGCCATCAGTCTCAGAGGGCAGGGGTACTGTGGGGTCCTC |
| ERBB2 | CCCTCTGAGACTGATGGCTACGTTGCCCCCCTGACCTGCAGCCCCCAGCC\|CTGCTGGGCAAAAGACAACCTGTGGCCAGGGTCTGGAAGGGCCCTGGGAG |
| JAK2 | GACAAAGAATACTATAAAGTAAAAGAACCTGGTGAAAGTCCCATATTCTG\|ATTGTGCGACGGGGAATGCCTCCCGGAGGAGGAGGCGAAGTGGTTTTCTC |
| JAK2 | ACTTTTGAAGAATAATGGAAGATTACCAAGACCAGATGGATGCCCAGATG\|AGGATAGAATTTTTGCGGCATTTGAAGAGCTTTTTCCAGATTATGTTTAA |
| CDKN2A | CTACTGAGGAGCCAGCGTCTAGGGCAGCAGCCGCTTCCTAGAAGACCAGG\|ATATTGAGGCAACAATGAACTCGGCCCTGAATGAGCTACGGGAACTAGAA |
| CDKN2A | AGGCGGGGGCGCTGCCCAACGCACCGAATAGTTACGGTCGGAGGCCGATC\|CAGGATATTGAGGCAACAATGAACTCGGCCCTGAATGAGCTACGGGAACT |
| FGFR3 | GCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTCCACCGACG\|TAAAGGCGACACAGGAGGAGAACCGGGAGCTGAGGAGCAGGTGTGAGGAG |
| SLC7A2 | CGCGGGCCCCCGACGCGCTGCAGCCGGCAGCCCACCGCCGCCTTCTTGGC\|TGTGTTCCGGCCGCAGAGCACCGTCTGCGTGAGGAGATCCTGGCCAAGTT |
| RAF1 | GGTGCGAGGGCAGCAGCCTCTACAAACACCTGCATGTCCAGGAGACCAAG\|GAAAAGCCATGGAGACTTCAGCATCCTCCTCCCAGCCTCAGGACAACAGT |
| BRAF | GGCCGGCGCCGGCGCCGGCGCCGCGGCCTCTTCGGCTGCGGACCCTGCCA\|TTCCGGATATTGACTGTTCAAGTACTATTATGCTGGACAATATTGTGAGG |
| TRIM24 | ATTTGAGCTCACCAGTGGGAGGGTCTTATAATCTTCCCTCTCTTCCGGAT\|GTGTGGAATATCAAACAAATGATTAAGTTGACACAGGAACATATAGAGGC |

[0973] 表50

[0974]

| 基因 | 已批准的 | 注册前（批准前） | III 期 |
|---|---|---|---|
| ABL1 | 甲磺酸伊马替尼（imatinib mesilate）<br>博舒替尼（bosutinib）<br>尼罗替尼（nilotinib）<br>帕纳替尼（ponatinib）<br>拉多替尼（radotinib） | N | N |
| ALK | 克卓替尼<br>盐酸艾乐替尼<br>色瑞替尼 | N | N |
| BRAF | 维罗非尼<br>帕唑帕尼<br>达拉菲尼 | N | 恩可非尼 |
| CDKN2A | N | N | N |
| CDK6 | N | 帕博西里（palbociclib） | LEE-011<br>LY-2835219 |
| EGFR | 帕尼单抗（panitumumab）<br>西妥昔单抗（cetuximab）<br>埃罗替尼（erlotinib）<br>奈皮德明（nepidermin）<br>吉非替尼（gefitinib）<br>尼妥珠单抗（nimotuzumab）<br>凡德他尼（vandetanib）<br>拉帕替尼（lapatinib）<br>阿法替尼（afatinib）<br>埃克替尼（icotinib） | N | 来那替尼（neratinib）<br>XL-647<br>林多派皮姆（rindopepimut）<br>莱西突木单抗（necitumumab）<br>达可替尼（dacomitinib）<br>卢瑟替尼（rociletinib）<br>AZD-9291 |

[0975] 表50

[0976]

| 基因 | II 期 | I 期 | 临床前 |
|---|---|---|---|
| ABL1 | 巴氟替尼（bafetinib） | 雷巴替尼（rebastinib）<br>ABL-001 | ON-044580 SUN-K706 |
| ALK | AP-26113<br>X-396<br>PF-06463922<br>TSR-011<br>恩确替尼（entrectinib） | CEP-37440 EBI-215 | AZD-3463 ARN-5032<br>DLX-521 |
| BRAF | RAF-265<br>XL-281<br>AB-024<br>PLX-8394 | ARQ-761<br>ARQ-736<br>BeiGene-283<br>DP-4978 | Braf 抑制剂，Sareum<br>B-Raf 和 c-RAF 双重抑制剂，Redx Pharma<br>PLX-8394<br>DP-4978<br>TL-241，B-Raf 激酶抑制剂，<br>Array BioPharma-1 |

[0977]

| | | | |
|---|---|---|---|
| | | | EN-3352<br>EBI-907<br>PI3K/BET 布罗莫结构域抑制剂 SignalRx<br>UAI-201<br>Nanolipolee-007<br>ASN-003 |
| CDKN2A | P16 37-63 | N | N |
| CDK6 | 阿昔迪布（alvocidib） | GZ38-1 | CDK4/6 抑制剂，G1 Therapeutics<br>卡皮汀 β（capridine Beta），AV Therapeutics<br>G1T38-1<br>G1T30-1 |
| EGFR | 多韦替尼（dovitinib）<br>瓦尼替尼（varlitinib）<br>马瑞佐米（marizomib）<br>福土西单抗（futuximab）<br>S-222611<br>AP-26113<br>安卓奎诺尔（antroquinonol）<br>GT-MAB 5.2-GEX<br>泊西替尼（poziotinib）<br>杜里格单抗（duligotuzumab）<br>LY-3016859<br>ABT-414<br>KD-020<br>ASP-8273<br>负载多柔比星的 EnGeneIC 传递媒剂（doxorubicin loaded EnGeneIC delivery vehicles）<br>EGF-816<br>EGFRvII CAR，Kite Pharma | JNJ-26483327<br>MM-151<br>AL-6802<br>ABT-806<br>依吡替尼（epitinib）<br>艾力替尼（allitinib）<br>西栗替尼（theliatinib）<br>ErbituxEDVsPac<br>AMG-595<br>IMGN-289<br>吡咯替尼（pyrotinib）<br>TAS-121<br>嵌合抗原受体 T 细胞疗法，EGFRVIII，Novartis<br>GC-1118A<br>LY-3164530<br>AZD-3759<br>miR-16，EDV 纳米细胞，EnGeneIC | CUDC-101<br>Sym-013<br>STI-A020X<br>SCT-200<br>哌罗替尼（pirotinib）<br>PF-06459988<br>NT-004<br>拉罗替尼（larotinib）<br>KL-ON113<br>JNJ-61186372<br>IRAD-425<br>HL-176<br>FV-225<br>表皮生长因子，USV Limited<br>EM1-mAB<br>EGFRvIII/CD3<br>CUDC-101<br>CTX-023<br>BPI-4039<br>BPI-0403<br>双特异性抗体，Zyngenia<br>AvidBiologics-2<br>AvidBiologics-1<br>抗癌基因疗法，Gradalis<br>抗 EGFR 单抗，Kadmon<br>抗 EGFR 生物仿制药，Mabion<br>AFM-21 |

[0978] 表50

[0979]

| 基因 | 已批准的 | 注册前（批准前） | III 期 |
|---|---|---|---|
| ERBB2 | 曲妥珠单抗（trastuzumab）<br>曲妥珠单抗恩他新（trastuzumab emtansine）<br>帕妥珠单抗（pertuzumab）；二甲苯磺酸拉帕替尼（lapatinib ditosylate catumaxomab）<br>阿法替尼<br>曲妥珠单抗，Enhanze | N | 来那替尼<br>XL-647<br>达可替尼<br>奈立派姆-S（nelipepimut-S）<br>曲妥珠单抗（Celltrion、Biocad、Biocon、Synthon、Harvest Moon、Aryogen）<br>多柔比星 |
| FGFR2 | 帕纳替尼（ponatinib）<br>帕利夫明（palifermin） | 乐伐替尼（lenvatinib） | N |
| FGFR3 | 帕纳替尼<br>左乙拉西坦（levetiracetam）于氯化钠中，HQ Specialty Pharma | 乐伐替尼（lenvatinib）<br>马赛替尼（masitinib） | N |
| JAK2 | 芦可替尼（ruxolitinib） | N | 莫罗替尼（momelotinib）<br>帕瑞替尼（pacritinib）<br>巴瑞替尼（baricitinib） |
| MTOR | 依维莫司<br>坦罗莫司<br>西罗莫司（sirolimus）<br>佐他莫司（zotarolimus）<br>百奥莫司（biolimus）<br>乌米莫司（umirolimus） | 地磷莫司（ridaforolimus） | TCD-10023<br>沃他立西（voxtalisib）<br>nab-雷帕霉素（nab-rapamycin）<br>阿皮托利西<br>格达托利西 |
| NTRK1 | N | N | MIM-D3 |
| RAF1 | 索拉非尼 | N | N |
| TERT | N | N | GV-1001 |

[0980] 表50

[0981]

| 基因 | II 期 | I 期 | 临床前 |
|---|---|---|---|
| ERBB2 | 拉普亮赛-T（lapuleucel-T）<br>AVX-901<br>AE-37<br>BMS-690514<br>MVA-BN-HER2<br>瓦尼替尼<br>MM-111<br>AC-480；卵巢癌疫苗（Generex）<br>玛格土西单抗（margetuximab）<br>泊西替尼（poziotinib）<br>PR-610<br>KD-020 | Her-VAXX<br>VM-206<br>ARRY-380<br>JNJ-26483327<br>S-222611<br>多柔比星（Merrimack）<br>西帕替尼（cipatinib）<br>TrasGEX<br>曲妥珠单抗（Hanwha Chemical）；<br>曲妥珠单抗（Pfizer）<br>IDN-6439<br>AVX-901 | 洛瓦新 B（Lovaxin B）<br>TH-1，Algeta<br>曲妥珠单抗抗体结合物，Synthon<br>CUDC-101<br>Her-2/neu 斯塔多体（Stradobody），Gliknik<br>ARX-788<br>Etbx-021<br>SN-34003<br>IBI-302<br>NT-004 ICT-140<br>ONS-1050<br>Sym-013<br>抗 HER2 X 抗 CD3（Emergent Biosolutions） |
| FGFR2 | ENMD-2076<br>德立替尼（lucitanib）<br>AZD-4547<br>BGJ-398<br>LY-2874455 | JNJ-42756493<br>S-49076<br>FF-284<br>BAY-1179470<br>BAY-1163877<br>ASP-5878 | FPA-144<br>角化细胞生长因子，Nanogen |
| FGFR3 | 多韦替尼乳酸盐（dovitinib lactate）<br>ENMD-2076<br>AZD-4547<br>BGJ-398<br>LY-2874455 | JNJ-42756493<br>S-49076<br>FF284<br>BAY-1163877<br>ASP-5878<br>B-701 | N |
| JAK2 | BMS-911543<br>NS-018 | AC-430<br>SB-1317<br>PF-06263276 | ON-044580<br>TP-0413<br>VR-588<br>NMS-P953 |
| MTOR | 奎纳克林<br>XL-765<br>达妥昔布<br>PKI-587<br>PF-04691502<br>INK-128<br>AZD-2014<br>CC-223 | P-7170<br>CBLC-137<br>AZD-2014<br>CC-115<br>PWT-33957<br>DS-7423<br>GDC-0084<br>DS-3078<br>LY-3023414<br>PI3 激酶/mTOR 抑制剂，Lilly<br>SF-1126<br>SB-2343<br>欧米帕立西（omipalisib）<br>ABTL-0812<br>PQR-309 | nPT-MTOR<br>STP-503<br>X-480<br>ABTL-0812<br>X-414；CC214<br>HMPL-518<br>PQR-401<br>mTOR 抑制剂/PI3 激酶抑制剂，Lilly-1<br>PIM/PI3k/mTOR 抑制剂，Inflection Biosciences<br>LOR-220<br>NV-128<br>SPR-965<br>IBL-301<br>PQR-316 |

[0982]

| | | | |
|---|---|---|---|
| | | DCBCI-0901 | PQR-311<br>TAM-01 |
| NTRK1 | 马来酸米西西尼（milciclib maleate）<br>CT-327<br>NMS-E628<br>PSR-011 | BXL-1H5<br>PLX-7486<br>LOXO-101<br>DS-6051b | NMS-P626<br>CT-340<br>FX-007<br>CRB-0089<br>泛 TRK（Merck） |
| RAF1 | iCo-007<br>XL-281 | RO-5126766<br>MLN-2480<br>DP-4978 | HM-95573<br>TAK-632<br>UAI-201<br>泛 RAF Novartis 和 Celator |
| TERT | VX-001<br>GX-301 | TeloB-Vax | 端粒酶疫苗，Geron<br>hTERT DNA 疫苗，Inovio<br>INVAC-1<br>AST-VAC2 |

序列表

<110> 生命技术公司

<120>与癌症相关的基因融合体和基因变异体

<130> LT00802.1 PCT

<140>
<141>

<150> 61/971,455
<151> 2014-03-27

<150> 61/993,732
<151> 2014-05-15

<150> 62/004,727
<151> 2014-05-29

<150> 62/092,898
<151> 2014-12-17

<160> 289

[0001] <170> PatentIn version 3.5

<210> 1
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 1
tgcggagagg tcagtaacta aattggagaa aagcattgat gacttagaag tgtaccgccg 60

gaagcaccag gagctgcaag ccatgcagat ggagctgcag 100

<210> 2
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 2
ctgagagacc catggcattc ctcagggaat actttgagag gttggagaag acctcctcca 60

tcagtgacct gaaggaggtg ccgcggaaaa acatcaccct 100

<210> 3
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 3
gtgcaacagg ttcaggtgtt tgctgacgtc cagtgtacag tgaatctggt aggcggctgt 60

ggggctgctc cagttcaatc tcagcgagct gttcagttgg 100

<210> 4
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 4
gaaactttcc tcctccacca cctcttgatg aagaggcttt caaagtacag aaggcatcac 60

aggaggcctc tgcatgatgt ggcttccaaa gactcaagga 100

[0002] <210> 5
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 5
gatggtgctg taaccacctc acagattcca gcttcggaac aagagaccct gtgagccaag 60

ggagtttgtg gagaactctg agtgcataca gtgccaccca 100

<210> 6
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 6
gccctcccag aggcccacct tcaagcagct ggtggaggac ctggaccgtg tccttggtac 60

aaatggagcc actgacgaca gcaagaccaa cctcatcgtc 100

<210> 7

<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸
<400> 7
tagctgtgct cgcgctactc tctctttctg gcctggaggc tatccagcgt gctggagaat 60

ctggtaaaag caccattgtg aagcagatga ggatcctgca 100

<210> 8
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 8
gagattttgg aatttccaac acgagaagta tatgtccctc acactgtgta cagtggcggc 60

atgattttgt gcacggatgg ataaaagtac ctgtgactca 100

<210> 9
<211> 100
[0003] <212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 9
tgccgctctg tacacctggt acgtcagaaa gcaacgagag atcctccgac gtgagaccgt 60

ggagccgccc ccgccggcgc agctgcactt catgtacgtg 100

<210> 10
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸
<400> 10
gggaagggcc ctgagccaga gtccgtcatc ggttactccg gagaagatta cactaacagc 60

acatctggag acccggtgga gaagaaggac gaaacacctt 100

<210> 11
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 11
tccaggaggt ggagggccac caggaacacc catcatgcct agtccagcag gcccggctgt 60

gctggctcca gaggatgggc tggccatgtc cctgcatttc 100

<210> 12
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 12
tttcctgaag aacgttgggg agagtgtggc agctgccctt agccctctgg acactaacag 60

cacatctgga gacccggtgg agaagaagga cgaaacacct 100

<210> 13
<211> 100
<212> DNA
<213>人工序列

[0004]

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 13
catataaaac tactttaagg aattagatgt atggttgtcc caaagcagaa acctggaaac 60

ggtggcctcc aacgccgctc ccccctcccg ggaatggagg 100

<210> 14
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸
<400> 14
cgccatctgc accttccata gccaaatcag ggtcattgga ctcagaactt ggttcttgga 60

aaaactctag gagaaggcga atttggaaaa gtggtcaagg 100

<210> 15
<211> 100
<212> DNA
<213>人工序列

<220>

<223>人工序列的描述：合成

<400> 15
aatcttacaa tggcaggacc attctgggaa ccatcatgaa tacgatcaag gaggatccaa 60

agtgggaatt ccctcggaag aacttggttc ttggaaaaac 100

<210> 16
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 16
gccctatatt tgcattaaaa tggaataaga aaggaaattt catcctaagt gctggactcc 60

atggagaacc aggtctccgt ggatgccttc aagatcctgg 100

<210> 17
<211> 100
<212> DNA
<213>人工序列

[0005] <220>
<223>人工序列的描述：合成聚核苷酸

<400> 17
ttaatatgcc agaaaaagaa agaaaaggag ttagtaacta ccgttcagag tactcttcca 60

acccaagagg agattgaaaa tcttcctgcc ttccctcggg 100

<210> 18
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 18
gagaaacaca ggaggaggag gatgagattc ttccaaggaa agactatgag gatgcaattc 60

gaagtcacag cgaatcagcc tcaccttcag ccctgtccag 100

<210> 19
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 19
tccagcatct gggggctgct aaggatgccc agcggcagct cacagccgag gatgcaattc 60

gaagtcacag cgaatcagcc tcaccttcag ccctgtccag 100

<210> 20
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 20
agggccgcac ttggaccttg tgcggcaccc ctgagtacct ggcccctgag gtgctggagg 60

acaatgacta cggccgtgca gtggactggt gggggctggg 100

<210> 21
<211> 100
<212> DNA
[0006] <213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 21
agggccgcac ttggaccttg tgcggcaccc ctgagtacct ggcccctgag gtgctggagg 60

acaatgacta cggccgtgca gtggactggt gggggctggg 100

<210> 22
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 22
agggccgcac ttggaccttg tgcggcaccc ctgagtacct ggcccctgag gtgctggagg 60

acaatgacta cggccgtgca gtggactggt gggggctggg 100

<210> 23

<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 23
agggccgcac ttggaccttg tgcggcaccc ctgagtacct ggcccctgag gtgctggagg 60

acaatgacta cggccgtgca gtggactggt gggggctggg 100

<210> 24
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 24
agggccgcac ttggaccttg tgcggcaccc ctgagtacct ggcccctgag gtgctggagg 60

acaatgacta tggccgggcc gtggactggt gggggctggg 100

[0007]

<210> 25
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 25
ccaggaagct cgatcaaatg cccgcctaaa gcagctctca tttgcaggtg gtactttgga 60

aaacttggcc gaaaagatgc tgagcgacag ctattgtcct 100

<210> 26
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 26
caaggttggg cattgggtgg aggagcagaa tttactacag catgtgattt cagggaagga 60

gattggtggg aagcccgctc cttgacaact ggagagacag 100

<210> 27
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 27
cttcatatca gaggactatg caacagccca tgaagacttt cagcagtcct ctggaaattg 60

aacttagctc attaagggaa gctttgtctt tcgtgtcatt 100

<210> 28
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 28
attgatttca cgaatgaggc agtggagcag gtggaagagg aggagtctgg cccgcgatgc 60

[0008] tcccagcccg gtgagacctg cctgaatggc gggaagtgtg 100

<210> 29
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 29
ggacatgttg gatgtgaagg agcggaaggt taatgttctt cagaagaagg aggatccaaa 60

gtgggaattc cctcggaaga acttggttct tggaaaaact 100

<210> 30
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 30

gaatgcacct atcttaacaa atacaacatt gaacgtcata agacttgttg ttctggggat 60

tcttggagga attcttgctt tgctaattct gattctgctg 100

<210> 31
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 31
aacatcaaag gcaattggct taagaatgtt catcatctgc atatattttc ttagcaaagc 60

aagaattcct ccaagaatcc ccagaatggc aggaatttgc 100

<210> 32
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

[0009]

<400> 32
gcaaattcct gccattctgg ggattcttgg aggaattctt gctttgctaa gaaaatatat 60

gcagatgatg aacattctta agccaattgc ctttgatgtt 100

<210> 33
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 33
gctgcaggat ctggtttacc cacaggctga tatatatgtt ggtttccaat cggggccggc 60

tcccgagtac atggtggcgc cgccgagggg ctccggggcc 100

<210> 34
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 34
ggccccggag cccctcggcg gcgccaccat gtactcggga gccggccccg attggaaacc 60

aacatatata tcagcctgtg ggtaaaccag atcctgcagc 100

<210> 35
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 35
ccccggagcc cctcggcggc gccaccatgt actcgggagc cggccccggt ttctggctac 60

cctggttcac atggaatcac agccatggct ggcagcatct 100

<210> 36
<211> 100
<212> DNA
<213>人工序列

<220>
[0010] <223>人工序列的描述：合成聚核苷酸

<400> 36
cgaagtacag tttttacatg ttttaattgc aaccgccaaa gctggattct ccggggccgg 60

ctcccgagta catggtggcg ccgccgaggg gctccggggc 100

<210> 37
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 37
ggccccggag cccctcggcg gcgccaccat gtactcggga gccggccccg gaagtcggct 60

tggccctgag gacattattg gccactgtgg atgagaccat 100

<210> 38
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 38
acacatgggc cgcaagaaca ggcctcatgt agtacctggc atactccagc gcccggggca 60

gggtctggac agaagaagcc ctgctggggt accagatact 100

<210> 39
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 39
gggcgctggg ggcatggtcc accacaggca ccgcagctca tctaccagga ctccctcagt 60

gcccccttcgg tggcctccag attcattgat tcccacacac 100

<210> 40
<211> 100
<212> DNA
[0011] <213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 40
gtggcggtgg ggacctgaca ctagggctgg agccctctga agaggaggcc tgttcttgcg 60

gcccatgtgt tttctggtga agaggagctt ccccaggact 100

<210> 41
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 41
ctgaggagta tctggtaccc cagcagggct tcttctgtcc agaccctgcc cagcgccctg 60

gccttcttcc tgacggccac tgtcttcctc gtgctctgca 100

<210> 42

<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 42
tgcagagcac gaggaagaca gtggccgtca ggaagaaggc cagggcgctg ggtgcagatg 60

gggggctggg gcagccgctc ccccttttcc agcaggtcag 100

<210> 43
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 43
aggagaaccg gatcaccatt cccgtgcaga ccttctccaa cctgcagatt cgaggagagc 60

aggattctct gcctcttcca aacttttcct ccctgaacct 100

[0012]

<210> 44
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 44
acgtgcggga ggcggccagt tatcaggagg cgctggcgcg gctggaggaa atggctcgtc 60

accttcgtga ataccaagac ctgctcaatg ttaagatggc 100

<210> 45
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 45
aggagaaccg gatcaccatt cccgtgcaga ccttctccaa cctgcagatt cgaggagagc 60

aggatttctc tgcctcttcc aaacttttcc tccctgaacc 100

<210> 46
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 46
aatgtcaagc tggccctgga catcgagatc gccacctaca ggaagctgct ggaaggcgag 60

gagagcagga tttctctgcc tcttccaaac ttttcctccc 100

<210> 47
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 47
atcaccattc ccgtgcagac cttctccaac ctgcagattc gagaaaccag gacactattg 60

[0013] gccgcctgca ggatgagatt cagaatatga aggaggaaat 100

<210> 48
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 48
cttctccaac ctgcagattc gaggggcaa aagcaccaaa gacggggaaa tggctcgtca 60

ccttcgtgaa taccaagacc tgctcaatgt taagatggcc 100

<210> 49
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 49

gaaccggatc accattcccg tgcagacctt ctccaacctg cagattcgag aatctggatt 60

cactccctct ggttgatacc cactcaaaaa ggacacttct 100

<210> 50
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 50
cagagatgat ggagctcaat gaccgctttg ccagctacat cgagaaggtt cgcttcctgg 60

agcagcagaa taagatcctg ctggccgagc tcgagcagct 100

<210> 51
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

[0014]

<400> 51
gaaaccagcc tggacaccaa gtctgtgtca gaaggccacc tcaagaggaa atggctcgtc 60

accttcgtga ataccaagac ctgctcaatg ttaagatggc 100

<210> 52
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 52
cacgaacgag tccctggaga ggcagatgcg cgagcaggag gagcggcacg aatgagtccc 60

tggaacgcca gatgcgtgaa atggaagaga actttgccgt 100

<210> 53
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 53
ggcagagaaa tcctgctctc ctcgccttcc agcagcttcc tgtaggtggc gtggcgatct 60

cgatgtccag ggccagcttg acattgagca ggtcctggta 100

<210> 54
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 54
ctggcttcaa ggagacccgg gccagtgagc gggcagagat gatggagctc aatgaccgct 60

tcgccaacta catcgacaag gtgcgcttcc tggagcagca 100

<210> 55
<211> 100
<212> DNA
<213>人工序列

<220>
[0015] <223>人工序列的描述：合成聚核苷酸

<400> 55
gaggagcggc acgtgcggga ggcggccagt tatcaggagg cgctggcgcg gccacctaca 60

ggaagctgct ggaaggcgag gagagcagga tttctctgcc 100

<210> 56
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 56
catcgagatc gccacctaca ggaagctgct agagggcgag gagaaccgga gacaggtgca 60

gtccctcacc tgtgaagtgg atgcccttaa aggaaccaat 100

<210> 57
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 57
gaggagcggc acgtgcggga ggcggccagt tatcaggagg cgctggcgcg tgctggaagg 60

cgaggagagc aggatttctc tgcctcttcc aaacttttcc 100

<210> 58
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 58
ggcagagaaa tcctgctctc ctcgccttcc agcagcttcc tgtaggtggc gatctcgatg 60

tccagggcca gcttgacatt gagcaggtcc tggtactcct 100

<210> 59
<211> 100
<212> DNA
[0016] <213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 59
aggagtacca ggacctgctc aatgtcaagc tggccctgga catcgagatc gccacctaca 60

ggaagctgct ggaaggcgag gagagcagga tttctctgcc 100

<210> 60
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 60
cccgggccag tgagcgggca gagatgatgg agctcaatga ccgctttgcc ctcgagcagc 60

tcaagggcca aggcaagtcg cgcctggggg acctctacga 100

<210> 61

<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 61
cgggccagtg agcgggcaga gatgatggag ctcaatgacc gctttgccag ctacatcgac 60

aaggtgcgct tcctggagca gcagaataag atcctgctgg 100

<210> 62
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 62
acgtccatgc ctccaagcgc attctcttct ccatcgtcca tgacaagtca gtgtaccgcc 60

ggaagcacca ggagctgcaa gccatgcaga tggagctgca 100

[0017]

<210> 63
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 63
ccatgagcga gttgtcatca tcaagaatat gtttcatcct atggattttg agatacaagg 60

cagttgtgaa gccacttgag cgacagccct ccaatgccat 100

<210> 64
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 64
ctcggggccc ttcgagagca agtttaagaa ggagccggcc ctgactgcag acactaacag 60

cacatctgga gacccggtgg agaagaagga cgaaacacct 100

<210> 65
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 65
aggtgtttcg tccttcttct ccaccgggtc tccagatgtg ctgttagtgt ctgcagtcag 60

ggccggctcc ttcttaaact tgctctcgaa gggccccgag 100

<210> 66
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 66
agaaatggtt tcaaatgaat ctgtagacta ccgagctact tttccagaag ggccacccag 60

[0018] tgctcctgca gaagatcgtt caggaacacc cgacagcatt 100

<210> 67
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 67
tgtaagtgcc cgaagtgtaa gcccaactac agaaatggtt tcaaatgaat ctgcagaaga 60

tcgttcagga acacccgaca gcattgcttc ctcctcctca 100

<210> 68
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 68

aatggtttca aatgaatctg tagactaccg agctactttt ccagaagatc gttcaggaac 60

acccgacagc attgcttcct cctcctcagc agctcaccca 100

<210> 69
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 69
tatatccagt ccattactgc aaaatactgt ccacattgac ctcagtgctc ctgcagaaga 60

tcgttcagga acacccgaca gcattgcttc ctcctcctca 100

<210> 70
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

[0019]

<400> 70
tattgacctt gtgaaccatt tcaagtgctc ttgcccacca ggcactcggg aatcaggaaa 60

cccaggccca acggcaggtg gatgcaagaa gaaatgttcc 100

<210> 71
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 71
gtattgacct tgtgaaccat ttcaagtgct cttgcccacc aggcactcgg gaatcaggaa 60

acccaggccc aacggcaggt ggatgcaaga agaaatgttc 100

<210> 72
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 72
ccatcgccga ccagatcacc ctcctcaagg ctgcctgcct ggacatcctg gaggggagat 60

ttgtcgcctg ccgctcgctc tggggctcga tgtgaatata 100

<210> 73
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 73
gtttggaaat aatggtgaag gtgctgaacc ctcagcagga gggcagtttg tagctttcca 60

cagccccacc accaggatca agaaggagcc ccagagtccc 100

<210> 74
<211> 100
<212> DNA
<213>人工序列

<220>
[0020] <223>人工序列的描述：合成聚核苷酸

<400> 74
agcaatacca ttgacctgcc gatgtccccc cgcactttag attcattgat gcagtttgtt 60

cctgatttcc attcagaaaa cctagctttc cacagcccca 100

<210> 75
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 75
cctgccgatg tccccccgca ctttagattc attgatgcag tttggaaata gatgtcaccg 60

ggtgcgcatc aatgtacctc cacacagagg gcttctctgg 100

<210> 76
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 76
atcaataaaa atgttatgtc agcgtttggc ttaacagatg atcaggtttc agatcagttt 60

cctaattcat ctcagaacgg ttcatgccga caagtgcagt 100

<210> 77
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 77
catccaaggt tccattaaat acatcatgct taaccctagt tcacgaatca aggtgacaaa 60

tgtgtcatgc ctggagacaa gctccagcgc cagccctgct 100

<210> 78
<211> 100
<212> DNA
[0021] <213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 78
ccaaggttcc attaaataca tcatgcttaa ccctagttca cgaatcaagg tgacaaatgt 60

gtcatgcctg gagacaagct ccagcgccag ccctgctaga 100

<210> 79
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 79
tggcatggcg catgagcgag tctctagcag ggctggcgct ggagcttgtc tccaggaggc 60

tctatcttga agttagcaat cctctctttg tggttatcca 100

<210> 80

<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 80
tcagcatatg cgattttatt atatctttga cgaacagact cctggtattt ccaatccagg 60

gaagcgtgtc accgtcgtgg aaagcacgct cccagcccga 100

<210> 81
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 81
tcccaagcaa tggatgattt gatgctgtcc ccggacgata ttgaacaatg ttccctggat 60

gaaagtgtgg gaacagagga aggatcagag aaaagagagg 100

[0022]

<210> 82
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 82
ttcgggctgg gagcgtgctt tccacgacgg tgacacgctt ccctggattg gaaataccag 60

gagtctgttc gtcaaagata taataaaatc gcatatgctg 100

<210> 83
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 83
gaactttgcc gttgaagctg ctaactacca agacactatt ggccgcctgc tcgagaaacc 60

agcctggaca ccaagtctgt gtcagaaggc cacctcaaga 100

<210> 84
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 84
aaggaggaaa tggctcgtca ccttcgtgaa taccaagacc tgctcaatgt caagctggcc 60

ctggacatcg agatcgccac ctacaggaag ctgctagagg 100

<210> 85
<211> 89
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸
oligonucleotide

<400> 85
ttgaagctgc taactaccaa gacactattg gccgcctgct cgagaaacca gcctggacac 60

[0023] caagtctgtg tcagaaggcc acctcaaga 89

<210> 86
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 86
aatgttaaga tggcccttga cattgagatt gccacctaca ggaagctgct agagggcgag 60

gagaaccgga tcaccattcc cgtgcagacc ttctccaacc 100

<210> 87
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 87

ggtgcgcttc ctggagcagc agaataagat cctgctggcc gagctcgagc gggcactcaa 60

tgctggcttc aaggagaccc gggccagtga gcgggcagag 100

<210> 88
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 88
gtgaatacca agacctgctc aatgttaaga tggcccttga cattgagatt gccacctaca 60

ggaagctgct agagggcgag gagaaccgga tcaccattcc 100

<210> 89
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

[0024]

<400> 89
gaaggcgagg agagcaggat ttctctgcct cttccaaact tttcctccct tggacatcga 60

gatcgccacc tacaggaagc tgctagaggg cgaggagaac 100

<210> 90
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 90
ggaaggcgag gagagcagga ttctctgcct cttccaaact tttcctccct tggacatcga 60

gatcgccacc tacaggaagc tgctagaggg cgaggagaac 100

<210> 91
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 91
ggggacagcg acgacgcgga ggcagagaag ggaacgcccg gcccagcccc tgtgcaccgg 60

cacagacatg aagctgcggc tccctgccag tcccgagacc 100

<210> 92
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 92
ccagaaggaa gatggcggat ctggaggagc agttgtctga tgaagagaag tggtcctttg 60

gcgtgctcct ctgggagctg atgacaagag gagccccacc 100

<210> 93
<211> 100
<212> DNA
<213>人工序列

<220>
[0025] <223>人工序列的描述：合成聚核苷酸

<400> 93
caaaatcaaa ccttcctcat ctgccaatgc catttattct ctggctgcca gggatgagga 60

gaatttctgt gccactgtgc ccaaggatgg acgttcctat 100

<210> 94
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 94
tctgcgagaa cagagaaggg agctctatag tcggagtgga gaactgcaag attctcgcct 60

ctattgagct gctggcccgc tcattgccaa aaattcaccg 100

<210> 95
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 95
cctgctggcc gagctcgagc agctcaaggg ccaaggcaag tcgcgcctgg ctcctggccg 60

ccgtctgggt cctggcaccc gcctctccct ggctcgaatg 100

<210> 96
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 96
ctgacctctc tgaggctgcc aaccggaaca atgacgccct gcgccaggca caggagtacc 60

aggacctgct caatgtcaag ctggccctgg acatcgagat 100

<210> 97
<211> 100
<212> DNA
[0026] <213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 97
caggcaaagc aggagtccac tgagtaccgg agacaggtgc agtccctcac gtaccgctcc 60

aagtttgcag acctgacaga cgctgctgcc cgcaacgcgg 100

<210> 98
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 98
tttgccgttg aagctgctaa ctaccaagac actattggcc gcctgcagga gtaccaggac 60

ctgctcaatg tcaagctggc cctggacatc gagatcgcca 100

<210> 99

<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 99
cattgagatt gccacctaca ggaagctgct ggaaggcgag gagagcagga gtaccaggac 60

ctgctcaatg tcaagctggc cctggacatc gagatcgcca 100

<210> 100
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 100
gaactttgcc gttgaagctg ctaactacca agacactatt ggccgcctgc ttcgagaaac 60

cagcctggac accaagtctg tgtcagaagg ccacctcaag 100

[0027]

<210> 101
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 101
tttgccgttg aagctgctaa ctaccaagac actattggcc gcctgcagga gtaccaggac 60

ctgctcaatg tcaagctggc cctggacatc gagatcgcca 100

<210> 102
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 102
tttgccgttg aagctgctaa ctaccaagac actattggcc gcctgcagga gtaccaggac 60

ctgctcaatg tcaagctggc cctggacatc gagatcgcca 100

<210> 103
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 103
tgaaggagga aatggctcgt caccttcgtg aataccaaga cctgctcaat ctagagggcg 60

aggagaaccg gatcaccatt cccgtgcaga ccttctccaa 100

<210> 104
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 104
aaggaggaaa tggctcgtca ccttcgtgaa taccaagacc tgctcaatgt caagctggcc 60

[0028] ctggacatcg agatcgccac ctacaggaag ctgctagagg 100

<210> 105
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 105
aatgttaaga tggcccttga cattgagatt gccacctaca ggaagctgct agagggcgag 60

gagaaccgga tcaccattcc cgtgcagacc ttctccaacc 100

<210> 106
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 106

aatgttaaga tggcccttga cattgagatt gccacctaca ggaagctgct agagggcgag 60

gagaaccgga tcaccattcc cgtgcagacc ttctccaacc 100

<210> 107
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 107
aatgttaaga tggcccttga cattgagatt gccacctaca ggaagctgct agagggcgag 60

gagaaccgga tcaccattcc cgtgcagacc ttctccaacc 100

<210> 108
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

[0029]

<400> 108
aatgttaaga tggcccttga cattgagatt gccacctaca ggaagctgct agagggcgag 60

gagaaccgga tcaccattcc cgtgcagacc ttctccaacc 100

<210> 109
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 109
taagatggcc cttgacattg agattgccac ctacaggaag ctgctggaag gcgggaggcg 60

gccagttatc aggaggcgct ggcgcggctg gaggaagagg 100

<210> 110
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 110
gaaggcgagg agagcaggat ttctctgcct cttccaaact tttcctccct tggacatcga 60

gatcgccacc tacaggaagc tgctagaggg cgaggagaac 100

<210> 111
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 111
gaaggcgagg agagcaggat ttctctgcct cttccaaact tttcctccct tggacatcga 60

gatcgccacc tacaggaagc tgctagaggg cgaggagaac 100

<210> 112
<211> 100
<212> DNA
<213>人工序列

<220>
[0030] <223>人工序列的描述：合成聚核苷酸

<400> 112
cctcaagagg aacatcgtgg tgaagaccgt ggagatgcgg gatggagagg gatacccact 60

caaaaaggac acttctgatt aagacggttg aaactagaga 100

<210> 113
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 113
ggccacctca agaggaacat cgtggtgaag accgtggaga tgcgggatgg agatgcgtga 60

aatggaagag aactttgccg ttgaagctgc taactaccaa 100

<210> 114
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 114
aaggccacct caagaggaac atcgtggtga agaccgtgga gatgcgggat ggagagcagg 60

atttctctgc ctcttccaaa cttttcctcc ctgaacctga 100

<210> 115
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 115
caagtctgtg tcagaaggcc acctcaagag gaacatcgtg gtgaagaccg ggaggaaatg 60

gctcgtcacc ttcgtgaata ccaagacctg ctcaatgtta 100

<210> 116
<211> 100
<212> DNA
[0031] <213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 116
ttataccaat acaggctcac cagattgtaa atggaacgcc gccggctcgc gaataccaag 60

acctgctcaa tgttaagatg gcccttgaca ttgagattgc 100

<210> 117
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 117
aggagaaccg gatcaccatt cccgtgcaga ccttctccaa cctgcagatt cgaggagagc 60

aggatttctc tgcctcttcc aaacttttcc tccctgaacc 100

<210> 118

<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 118
aggagaaccg gatcaccatt cccgtgcaga ccttctccaa cctgcagatt cgaggagagc 60

aggatttctc tgcctcttcc aaacttttcc tccctgaacc 100

<210> 119
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 119
aggagtacca ggacctgctc aatgtcaagc tggccctgga catcgagatc gccacctaca 60

ggaagctgct ggaaggcgag gagagcagga tttctctgcc 100

[0032]
<210> 120
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 120
gcccgccact tgcaggagta ccaggacctg ctcaatgtca agctggccct cttgacattg 60

agattgccac ctacaggaag ctgctggaag gcgaggagag 100

<210> 121
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 121
gcattttctt cccacaggtg gaaaaggagg gagctgctct caggctgcgt ccagcaacag 60

tgcccaggct actaccagtc acacctagac ctggaggatc 100

<210> 122
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 122
gcattttctt cccacaggtg gaaaaggagg gagctgctct caggctgcgt ccagcaacag 60

tgcccaggct actaccagtc acacctagac ctggaggatc 100

<210> 123
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 123
gcattttctt cccacaggtg gaaaaggagg gagctgctct caggctgcgt ccagcaacag 60

[0033] tgcccaggct actaccagtc acacctagac ctggaggatc 100

<210> 124
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 124
gcattttctt cccacaggtg gaaaaggagg gagctgctct caggctgcgt ccagcaacag 60

tgcccaggct actaccagtc acacctagac ctggaggatc 100

<210> 125
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 125

gcattttctt cccacaggtg gaaaaggagg gagctgctct caggctgcgt ccagcaacag 60

tgcccaggct actaccagtc acacctagac ctggaggatc 100

<210> 126
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 126
gcattttctt cccacaggtg gaaaaggagg gagctgctct caggctgcgt ccagcaacag 60

tgcccaggct actaccagtc acacctagac ctggaggatc 100

<210> 127
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

[0034]

<400> 127
gcattttctt cccacaggtg gaaaaggagg gagctgctct caggctgcgt ccagcaacag 60

tgcccaggct actaccagtc acacctagac ctggaggatc 100

<210> 128
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 128
gcattttctt cccacaggtg gaaaaggagg gagctgctct caggctgcgt ccagcaacag 60

tgcccaggct actaccagtc acacctagac ctggaggatc 100

<210> 129
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 129
gcattttctt cccacaggtg gaaaaggagg gagctgctct caggctgcgt ccagcaacag 60

tgcccaggct actaccagtc acacctagac ctggaggatc 100

<210> 130
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 130
gcattttctt cccacaggtg gaaaaggagg gagctgctct caggctgcgt ccagcaacag 60

tgcccaggct actaccagtc acacctagac ctggaggatc 100

<210> 131
<211> 100
<212> DNA
<213>人工序列

<220>
[0035] <223>人工序列的描述：合成聚核苷酸

<400> 131
gatgcagcag agcttggaag gatgcttcag ctcatcttag gctgtgctgt gaacttggcg 60

cccaatgacc tgcccctgct ggccatggag tactgccaag 100

<210> 132
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 132
ggatgcttca gctcatctta ggctgtgctg tgaactgtga acagaagcaa gcctctgcgc 60

ttagatacct tcatgaaaac agaatcatcc atcgggatct 100

<210> 133
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 133
gggagccccc accatccggg gggaccccga gtgtcatctc ttctacaatg agcagcagga 60

ggaggcagag cacagcatcg tcgggaccag actcgtctca 100

<210> 134
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 134
tgagacgagt ctggtcccga cgatgctgtg ctctgcctcc tcctgctgct cattgtagaa 60

gagatgacac tcggggtccc cccggatggt gggggctccc 100

<210> 135
<211> 100
<212> DNA
[0036] <213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 135
atatttacaa aggcaagccc cagcaatcat aaagtcatcc ctgtgtatgt aggagggcat 60

gccgctctcc accatccgcg aggtggcggt gctgaggcac 100

<210> 136
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 136
agatcccccg ggagtccatc aagttggtga aaaggcttgg cgctgggcag tttggggtat 60

ccatagcagt tggacttgct gcttttgcct gtgtcctgtt 100

<210> 137

<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 137
ccacgcagca ggagaagcac cccacccacc acgagagggg ccagaagaag gtactttgtc 60

agcttcatca tccagttcca gttccacgag gcactgtgcc 100

<210> 138
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 138
cctggcacag tgcctcgtgg aactggaact ggatgatgaa gctgacaaag taccttcttc 60

tggcccctct cgtggtgggt ggggtgcttc tcctgctgcg 100

[0037]

<210> 139
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 139
agtggatctc agaacctcag ggactccatc ctccctctcc agccccacaa attcatcatc 60

aacagcatgg aagcgagaaa ccaaacagct atttcaaaat 100

<210> 140
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 140
attttgaaat agctgtttgg tttctcgctt ccatgctgtt gatgatgaat ttgttcttga 60

ggtcacactc tcagaggcca aggtggacat cccaggtgtg 100

<210> 141
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 141
ggaactgttc aaggctacaa caatgatggt tctctcaaaa tgtcctgaag gcatcaagtg 60

gaaattccta gaacataaag gtccagtatt tgccccacca 100

<210> 142
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 142
gcgcccaact gcatcctcct ggccatgttc ctcgtccact acgggcatcg gtacagatat 60

[0038] ttggcagctt tagtacaggt ctttatcttc caactagcga 100

<210> 143
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 143
gggagaaatt ttaattactt gaaaaccggt attagaatca aagaaggagg cttatttgaa 60

tacgtaactg cagccaacta ttttggagaa atcatggagt 100

<210> 144
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 144

gacgaggact tggaagttcc agttcctagc agatttaata gacgagtatc aggtagagtt 60

ggctttgtgg gacacagctg ggcaggaaga ttatgatcgc 100

<210> 145
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 145
ctctgtgtcc cgtttgagaa aaaggacttt gtaggactgg acacagacag cagaatatga 60

cagataccta gcatctagca aaataatggc agctgcttac 100

<210> 146
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

[0039]

<400> 146
ccgccccgtc gtcgtctgcc ttcgcttcac ggcgccgagc cgcggtccga accctggaag 60

ctgtcctgaa tttcaaatac tctggaggcc cgggccacac 100

<210> 147
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 147
aggaagcatt caagaacatc ctcacagaga tctaccgcat cgtgtcacag gtgatggggg 60

cacaggcacc aaagtccgcc aaggcaccct gaagaaggcg 100

<210> 148
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 148
gatcgcagac cgcgctgccc acgacgagtc cccggggaac aacgtggtgg ccatccgcgt 60

caccaagccc tgcaccccca agaccaaagc aaaggccaaa 100

<210> 149
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 149
aaaacattta gaacgatgtg aacatcgaat catggaatcc cttgcatggc caagatgcca 60

cagatgattg tgaactggca gcagcagcag cgggagaact 100

<210> 150
<211> 100
<212> DNA
<213>人工序列

<220>
[0040] <223>人工序列的描述：合成聚核苷酸

<400> 150
tttatgccca gaacctcatc gatgataagc agtttgcaaa gcttcacaca aagatctgtg 60

actttggcct ggcccgtgtt gcagatccag accatgatca 100

<210> 151
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 151
tgcagctgca attcctcgaa cgcccctgag cccaagtcct atgaaaaccc ctcctcagct 60

gagatgacct tccggaggcc cgcccaggcc ttcccggtca 100

<210> 152
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 152
gggagggctg ggcactatct cttcagaact gctgctctgg gtctcaatgg cctttcgccg 60

acaggtctgg ggcggagcag gcaggcgcag ccccctgcag 100

<210> 153
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 153
caaccaccgc aaacacaaca ttccgcactt ctggcccaag ctgctgatga agagaagagg 60

ccccctgagc cccccggacc tccaccgccg ccacctccac 100

[0041]

<210> 154
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 154
tccttgggaa ttatgggaac tgaaaaatgc tgtgataatt gcaggtccag agacctttg 60

cctgaagatt ttgtggttta tacttacaac aaggaaggga 100

<210> 155
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 155
cgctatggat gatcgagagg atctggtgta ccaggcgaag ctggccgagc tggcccggag 60

gaagccggtg ctgccggcgc tcaccatcaa ccctaccatc 100

<210> 156

<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 156
catggagcac ccagggaagc tactgtttgc tcctaacttg ctcttggaca gatggtctcc 60

cagcttgaag cccaaatatc tgagcttgtt gaacagttgg 100

<210> 157
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 157
cagaatgttt tgagctactt cgggtacttg gtaaaggggg ctatggaaag tgctgtcccc 60

ggcataggtc catctctgca gaagccattt caggagtacc 100

[0042]
<210> 158
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 158
gttcatatgg tccaactccc ccatggtcca tgctttcatt taactgaccc tgtggtgtgc 60

ccatttcgct tttgtggtga agcttctgcc gttgagcctc 100

<210> 159
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 159
agacctggac cagccagagg acgcgggctc tgaggatgag ctggaggagg ggtgctgtcc 60

ccggcatagg tccatctctg cagaagccat ttcaggagta 100

<210> 160
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 160
aagttcatat ggtccaactc ccccatggtc catgctttca tttaactgac cctgtggtgt 60

gcccatttcg cttttgtggt gaagcttctg ccgttgagcc 100

<210> 161
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 161
ggtactcctg aaatggcttc tgcagagatg gacctatgcc ggggacagca cttccctgtc 60

[0043] tcggaagtcc ggggctgggt aaaagccgtc ccgcctcctt 100

<210> 162
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 162
gtaacaggag caaatactgg gaaaatattt gccatgaagg tgcttaaaaa gtgctgtccc 60

cggcataggt ccatctctgc agaagccatt tcaggagtac 100

<210> 163
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 163

gcctttcaga ctggtggaaa actctacctc atccttgagt atctcagtgg gagaaaactg 60

gttgtcctgg atgtttgaaa agttggtcgt tgtcatggtg 100

<210> 164
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 164
agacctggac cagccagagg acgcgggctc tgaggatgag ctggaggagg ggtgctgtcc 60

ccggcatagg tccatctctg cagaagccat ttcaggagta 100

<210> 165
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

[0044]

<400> 165
cagaatgttt tgagctactt cgggtacttg gtaaaggggg ctatggaaag tgctgtcccc 60

ggcataggtc catctctgca gaagccattt caggagtacc 100

<210> 166
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 166
atgcctttca gactggtgga aaactctacc tcatccttga gtatctcagt gggagaaaac 60

tggttgtcct ggatgtttga aaagttggtc gttgtcatgg 100

<210> 167
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 167
atatttatgg aagacactgc ctgcttttac ttggcagaaa tctccatggc acaaagttat 60

gccaaacgaa tccagcagcg gttgaactca gaggagaaaa 100

<210> 168
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 168
tggggcattt acatcaaaag gggatcatct acagagacct gaagccggag tggtgctgtc 60

cccggcatag gtccatctct gcagaagcca tttcaggagt 100

<210> 169
<211> 100
<212> DNA
<213>人工序列

<220>
[0045] <223>人工序列的描述：合成聚核苷酸

<400> 169
tacccagccc cggacttccg agacagggaa gctgaggaca tggcaggagt acctggaggc 60

tcaacggcag aagcttcacc acaaaagcga aatgggcaca 100

<210> 170
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 170
cctgtggtgt gcccatttcg cttttgtggt gaagcttctg ccgttgagcc tccaggtcta 60

tgtcaaacac tcctgccatg tcctcagctt ccctgtctcg 100

<210> 171
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 171
aacaggagca aatactggga aaatatttgc catgaaggtg cttaaaaagg actttgcctc 60

ccgggccaaa ctggcagttc aaaaactagt acagaaagtt 100

<210> 172
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 172
ctatttatgc agttagaaag agagggaata tttatggaag acactgcctg tgctgtcccc 60

ggcataggtc catctctgca gaagccattt caggagtacc 100

<210> 173
<211> 100
<212> DNA
[0046] <213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 173
tgtgcccatt tcgcttttgt ggtgaagctt ctgccgttga gcctccaggt actcctgcca 60

tgtcctcagc ttccctgtct cggaagtccg gggctgggta 100

<210> 174
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 174
ccagaatgtt ttgagctact tcgggtactt ggtaaagggg gctatggaaa gggagaaaac 60

tggttgtcct ggatgtttga aaagttggtc gttgtcatgg 100

<210> 175

<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 175
atggaaaggt ttttcaagta cgaaaagtaa caggagcaaa tactgggaaa atatttcatg 60

gccagagcag ctcgcctctc aggtgctgaa ccagatgatg 100

<210> 176
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 176
acctggacca gccagaggac gcgggctctg aggatgagct ggaggagggg attccaaatc 60

ctttatttga tctggctgga ataacgtgtg gacactttct 100

[0047]

<210> 177
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 177
acctggacca gccagaggac gcgggctctg aggatgagct ggaggagggg gactttgcct 60

cccgggccaa actggcagtt caaaaactag tacagaaagt 100

<210> 178
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 178
gaaagtgtcc acacgttatt ccagccagat caaataaagg atttggaatc ccctcctcca 60

gctcatcctc agagcccgcg tcctctggct ggtccaggtc 100

<210> 179
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 179
ctggaccagc cagaggacgc gggctctgag gatgagctgg aggaggggga ctttgcctcc 60

cgggccaaac tggcagttca aaaactagta cagaaagttg 100

<210> 180
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 180
cctggaccag ccagaggacg cgggctctga ggatgagctg gaggaggggg actttgcctc 60

[0048] ccgggccaaa ctggcagttc aaaaactagt acagaaagtt 100

<210> 181
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 181
acctggacca gccagaggac gcgggctctg aggatgagct ggaggaggggg gactttgcct 60

cccgggccaa actggcagtt caaaaactag tacagaaagt 100

<210> 182
<211> 83
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸
oligonucleotide

<400> 182

cgcgggctct gaggatgagc tggaggaggg ggactttgcc tcccgggcca aactggcagt 60

tcaaaaacta gtacagaaag ttg 83

<210> 183
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 183
agtaacagga gcaaatactg ggaaaatatt tgccatgaag gtgcttaaaa agtgctgtcc 60

ccggcatagg tccatctctg cagaagccat ttcaggagta 100

<210> 184
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

[0049]

<400> 184
tgattgacac tctgcagcac caagtgaaat ctctggagca acagctggcc gtggggcttg 60

gcccggccca gtcctggcct ctgccaccag gtgtcaccga 100

<210> 185
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 185
tacgggacag aattgaatca gggagatatg aagcctccaa gctatgattc tgtgtaccgc 60

cggaagcacc aggagctgca agccatgcag atggagctgc 100

<210> 186
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 186
tacgggacag aattgaatca gggagatatg aagcctccaa gctatgattc tgtgtaccgc 60

cggaagcacc aggagctgca agccatgcag atggagctgc 100

<210> 187
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 187
tagacctgga caccttatta aagaactttc taaagtaatt cgagcaatag agaaaacact 60

tggtagacgg gactcgagtg atgattggga gattcctgat 100

<210> 188
<211> 100
<212> DNA
<213>人工序列

<220>
[0050] <223>人工序列的描述：合成聚核苷酸

<400> 188
gacctggaca ccttattaaa gaactttcta aagtaattcg agcaatagag aaaacacttg 60

gtagacggga ctcgagtgat gattgggaga ttcctgatgg 100

<210> 189
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 189
ctgaaaagga aaatctgcaa agaactttcc tgcttacaac ctcaagtaaa aaaacacttg 60

gtagacggga ctcgagtgat gattgggaga ttcctgatgg 100

<210> 190
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 190
tgcaggtcct gcatccaatg gatgctgccc agagatcgca gcatatcaaa gacttgatta 60

gagaccaagg atttcgtggt gatggaggat caaccacagg 100

<210> 191
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 191
ttggacaaag ggtggatgaa attgatgctg ctattcagag atcacaacag gacttgatta 60

gagaccaagg atttcgtggt gatggaggat caaccacagg 100

[0051]

<210> 192
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 192
cagagaaggc tttggatacc ctaaacaaag ccattgtcat tgatcccaag gatttcgtgg 60

tgatggagga tcaaccacag gtttgtctgc tacccccccct 100

<210> 193
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 193
accaagccag ccaattctgt cttcaccacc aaatggattt ggtattggaa gaatgaaaac 60

acttggtaga cgggactcga gtgatgattg ggagattcct 100

<210> 194

<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 194
ttcacctgtc cagcatccga ccaccgaggc tggaggggga gaacacccag gacttgatta 60

gagaccaagg atttcgtggt gatggaggat caaccacagg 100

<210> 195
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 195
gtcaatattg atgacttgat tagagaccaa ggatttcgtg gtgatggagg cacccagttg 60

gagaagctga tggagaacat gcgcaatgac attgccagtc 100

[0052]
<210> 196
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 196
cacctgtcca gcatccgacc accgaggctg gaggggggaga acacccagga cttgattaga 60

gaccaaggat ttcgtggtga tggaggatca accacaggtt 100

<210> 197
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 197
ggctttcaca agtacagtct acaaaaagac ctgctagagc cattattgcc ccggaaactg 60

cctgtgggtt tttactgcaa ctttgaagat ggcttctgtg 100

<210> 198
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 198
gaggggtctc aggtcctgct cacgagctcc aatgagatgg gtactgttag gttgaagatg 60

cccagcacag acacgccgtg ggaccgcatc atggtgttct 100

<210> 199
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 199
acgaggggtc tcaggtcctg ctcacgagct ccaatgagat gggtactgtt aggttgaaga 60

[0053] tgcccagcac agacacgccg tgggaccgca tcatggtgtt 100

<210> 200
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 200
ttcctgtgga acgttttcca gagggtcgat aaagacagga gtggagtgat atcagacagc 60

acttgaagag ggtgcagctg cgggagctgt cggaagcaga 100

<210> 201
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 201

acctctatgc tgaggggtgt gaggctctag tagtgaagaa gctacaagaa caggcaaatg 60

tgcaatacca acatgtctgt acctactgat ggtgctgtaa 100

<210> 202
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 202
cggaggctcc gggcaccccc gagggccccg agcccgagcg ccccagcccg ggggttggct 60

gtgttccggc cgcagagcac cgtctgcgtg aggagatcct 100

<210> 203
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

[0054]

<400> 203
gtggttacag caccatcagt aggtacagac atgttggtat tgcacatttg ccgtccgccc 60

aggtgctgag agggagcagg gcgcgggtcg gcgggcgcga 100

<210> 204
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 204
gaatgtctaa aagagtatac aaatcctgaa caaattaagc aatggagaaa gaattctggg 60

tcatgaacac ctcaattcag agcacgatca ttcttctcat 100

<210> 205
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 205
ttcccatttc tgaagaatct gaagagctgg atcagaagac attcagcatt gaattctggg 60

tcatgaacac ctcaattcag agcacgatca ttcttctcat 100

<210> 206
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 206
aatgagaaga atgatcgtgc tctgaattga ggtgttcatg acccagaatt ctttctccat 60

tgcttaattt gttcaggatt tgtatactct tttagacatt 100

<210> 207
<211> 100
<212> DNA
<213>人工序列

<220>
[0055] <223>人工序列的描述：合成聚核苷酸

<400> 207
gggatgcata atggccgagc tgttgactgg aagaacattg tttcctggta aaacttcagt 60

caagaagctg acaaaaaagg acatcgagga tacactgtca 100

<210> 208
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 208
gtcctggccc acaggctgcc attcaatgca atacgtcatg ctctgagccc gggctgccgg 60

ctgcgccact gggtcctggg gtcctggggg ctggggcttc 100

<210> 209
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 209
ccactggttc tgtgtgggtg tcggcaggaa tgtgccacgt ctggttcagg gatccggggc 60

tgccggctgc gccactgggt cctggggtcc tgggggctgg 100

<210> 210
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 210
cccttcacct ttaaacctct ttatcaaagt ggcttcactg cgatcctgac gggaattttg 60

tctgcaaggt gagaggcagt gttaaggatg atgagtccac 100

<210> 211
<211> 100
<212> DNA
[0056] <213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 211
ctgctggtga aagttcccac ggaaatgaga gggaattttg tctgcaaggt caggatcgca 60

gtgaagccac tttgataaag aggtttaaag gtgaaggggt 100

<210> 212
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 212
tctcccaaaa ttcacatcca ataaacagcc tgcagccccg agtgacatat gtccggtaca 60

aagccaaatt gatcgggatt gatgaagttt ccgcagctcg 100

<210> 213

<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 213
ctgctgatcc acatgagagt ccactctggg gagaagccca acaagtgtac ggggttggct 60

gtgttccggc cgcagagcac cgtctgcgtg aggagatcct 100

<210> 214
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 214
gatacaggtc ttggactggc cttcaccatt gcccatgagt ctggacacaa gggttggctg 60

tgttccggcc gcagagcacc gtctgcgtga ggagatcctg 100

[0057]

<210> 215
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 215
ctttagcaac gccaaatgga gatgttttga tccgagacct taattttgaa tcacaatatg 60

gagcttccgt ttctgccaag cctgaactac ccctctttta 100

<210> 216
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 216
tctgccggac ctcctcttcg atctcctcca gcgtgccctc ctcgatggcc caacctcatt 60

gactcccccg ggcatgtcga cttctcctcg gaggtgactg 100

<210> 217
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 217
attcagccct gacttctcaa aaagcactgc acagaggagg aggcagcaga accccatgga 60

ctacaagtgt ggctccccca gtgactcctc cacgactgag 100

<210> 218
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 218
aattctccct gacttctcaa aaagcactgc acagaggagg aggcagcaga accccatgga 60

[0058] ctacaagtgt ggctccccca gtgactcctc cacgactgag 100

<210> 219
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 219
gagatgatgc cccccgggct gtcttccccc ccattgtggg ccgccctcgc caccaggagc 60

tgcaagccat gcagatggag ctgcagagcc ctgagtacaa 100

<210> 220
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 220

atgccccccg ggctgtcttc ccctccattg tgggccgccc tcgccaccag tgatggaagg 60

ccacggggaa gtgaatatta agcattatct aaactgcagt 100

<210> 221
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 221
tgatgccccc cgggctgtct tcccctccat tgtgggccgc cctcgccacc agtgatggaa 60

ggccacgggg aagtgaatat taagcattat ctaaactgca 100

<210> 222
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

[0059]

<400> 222
atgaagtgac acccccagct acatccgagg aggttctagg acctgctacg agctgactat 60

agcactagtg aaatgctggt caacatggga aacttgcctc 100

<210> 223
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 223
attctaacac tccggccgct gcctccggct gctgtagctt attattaatg ctggctctcg 60

gctgcgggga tgccagactc gagctcgcac agcgcgcgga 100

<210> 224
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 224
ctggacaggg ctgaaggtga ggctgattcg ctgtgacttc gaattgcatc caagcagcgg 60

ggactcctca gggcaggcgg gcagcgacag tgcggtggtg 100

<210> 225
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 225
aaaaatctca tccaagaagc cctaacgtgt tatctgtcgc tttgagtcaa agagatctga 60

aactggacaa tgtcctgttg gaccacgagg gtcactgtaa 100

<210> 226
<211> 100
<212> DNA
<213>人工序列

<220>
[0060] <223>人工序列的描述：合成聚核苷酸

<400> 226
cgaagtctgc cagtttacag tgaccctcgt ggtccaacag gacattgtcc agtttcttta 60

tgtatgtggg taggagatgg agatgcaatc aatattttaa 100

<210> 227
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 227
gatgatctct caactttaac tggaaagtct aggttgttgg cagaagatat gcccgagcac 60

aaccctggca atttgggagg aacaatgaga ctgggaataa 100

<210> 228
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 228
ataaattctt tgctgacctg ctggattaca tcaaagcact gaatagaaat agtgatagag 60

tttgcaagaa actgccttaa cttgaaagat gctgattcca 100

<210> 229
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 229
gcactgaata gaaatagtga tagatccatt cctatgactg tagattttat ggtgatgttc 60

cttttataga agaaagacac agacatcggt tcgaggtaaa 100

<210> 230
<211> 100
<212> DNA
[0061] <213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 230
gatgatctct caactttaac tggaaagaat gtcttgattg tggaagatat gcccgagcac 60

aaccctggca atttgggagg aacaatgaga ctgggaataa 100

<210> 231
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 231
gattacatca aagcactgaa tagaaatagt gatagatcca ttcctatgac tgatagagtt 60

tgcaagaaac tgccttaact tgaaagatgc tgattccaca 100

<210> 232

<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 232
ggacagaatc caggagtcct ccagctgaga acgaggtgtc cacccccatg caggataatg 60

ctgactacag attatttcag aaaacactca aattgtgtcg 100

<210> 233
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 233
ccaactcacc caagtgcaat tcgtggagga ggtgcatttg ttcagaacag aggattgatt 60

gctggtgttg tctcaatatc aacagcactg ttattactac 100

[0062]

<210> 234
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 234
gcactgaaag aagctaaaga aaatgcatct cgtgatcgca aacgctatca gcaagagtac 60

acactcctca tttggatagg cttgtaagtg cccgaagtgt 100

<210> 235
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 235
gtatgcagga cttcaacacc atgtttacaa attgttacat ttataacaag ttccagggga 60

agtggtatgt ggtaggcctg gcagggaatg caattctcag 100

<210> 236
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 236
cattgtccat ggcaaaacag gacatcttat ggcctgcttt acatgtgcaa tagttcctaa 60

taagagcaac aatgaaatta tcctggtttt gcagcacttt 100

<210> 237
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 237
catgcctact agcctcccta accttgccaa ggaggcaaag gatgccaagg tggagcagcg 60

[0063] tgacttcatt ggagtggaca gcacaggaaa gaggctgctc 100

<210> 238
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 238
aacgccacag catcaaggat gtgcacgccc ggctgcaagc cctggcccag aagatccctt 60

acaatacctg agaggtcttg ttgcccgtgc ccttgcaata 100

<210> 239
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 239

actttcagatc cgtggcggag acaagccagc cttggacttg tatcagctgt ttggtatctt 60

actacacagc ctattccagg cctctcaact gtgattaatg 100

<210> 240
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 240
gtggaagcaa ccactaatat aaacacctcc catgtatagg aaggctggag cgtttgggaa 60

ggtggttgaa ggaacagcct atggattaag ccggtcccaa 100

<210> 241
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

[0064]

<400> 241
ggtcaatttt ggaaacacat gctactgtaa ctccgtgctt caggcattgt cctggttgtc 60

atttggaaac agaaaccgag gtatgaaatt cgctggaggg 100

<210> 242
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 242
acatgaagta caaagcagtg agtgtcaacg aaggctataa agtctgtgag aggggttggc 60

tgtgttccgg ccgcagagca ccgtctgcgt gaggagatcc 100

<210> 243
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 243
aatcaaacag aggccgcatg ctggggccgt acagttccac aaaggcatcc tcatgggctc 60

agcggtcatg ttttcgcttg aacgccttgt cggcttctgt 100

<210> 244
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 244
catgttatcc accaacatcc tgctggtcaa cctgctggtc gccatgtttg ggaatttgaa 60

gcctacatta atgcttctgg agaacatgga attgtggttt 100

<210> 245
<211> 100
<212> DNA
<213>人工序列

<220>
[0065] <223>人工序列的描述：合成聚核苷酸

<400> 245
cacagtcttt tacttgtccc tattgtggaa aaatgggcta tacggagaca gtaccctgag 60

tgaggaaatg agtcaatttg atttctcaac cggagttcag 100

<210> 246
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 246
aaacagaaga ggaagcatat gcactgaaga aaatatccta tcaacttaag aaacagatga 60

ttatgctgag attatagatg aagaagatac ttacaccatg 100

<210> 247
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 247
accatcgccg accagatcac cctcctcaag gctgcctgcc tggacatcct gcgcgagctt 60

cggcctcggc tctgtaccat gaagaagggc cccagtggct 100

<210> 248
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 248
acaaggggct ggacacggag acctgggtgg aggtggcctg gtgtgagctg cagaggaaga 60

tggatggccg agagtaccca gacgcacagg gctttgctgc 100

<210> 249
<211> 100
<212> DNA
[0066] <213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 249
atcgcccact ccagaagtag catggaaatg gagggcatct tcaaccacaa aggggttggc 60

tgtgttccgg ccgcagagca ccgtctgcgt gaggagatcc 100

<210> 250
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 250
ccgcttgcag cggggacgcg aggacccggg ctgggctttc ctcacccggg ggttggctgt 60

gttccggccg cagagcaccg tctgcgtgag gagatcctgg 100

<210> 251

<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 251
ctcacagacc gtgttcttct gcgccgtgcc tgggaacttg acaatcatcc ggctcatcct 60

gtaaggagag cgtcttgtag tctgatcaaa tcgcaagtac 100

<210> 252
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 252
tcctcacacc tgctcctcag ctcccggttc tcctcctgtg tcgcctttac gtcggtggac 60

gtcacggtaa ggacacggtc caggtcctcc accagctgct 100

[0067]

<210> 253
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 253
gccgcgccct cccagaggcc caccttcaag cagctggtgg aggacctgga actggggaag 60

atcatggaca ggttcgaaga ggttgtgtac caggccatgg 100

<210> 254
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 254
agcagctggt ggaggacctg gaccgtgtcc ttaccgtgac gtccaccgac gtaaaggcga 60

cacaggagga gaaccgggag ctgaggagca ggtgtgagga 100

<210> 255
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 255
cgccctccca gaggcccacc ttcaagcagc tggtggagga cctggaccgt gtccttctcc 60

gacctcttca agcgttttga gaaacagaaa gaggtgatcg 100

<210> 256
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 256
gagggccacc gcatggacaa gcccgccaac tgcacacacg acctgtacat gatcatggac 60

[0068] aggttcgaag aggttgtgta ccaggccatg gaggaagttc 100

<210> 257
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 257
gtacgcccag tccctgggtg ccgagacctg ccccctgcct agtttccagg accccaaccc 60

gtgcctcagc acccccctgca agaacgccgg gacatgccac 100

<210> 258
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 258

ggttggccct gtgatccctc aagactggtc cacggagtgt atgaccacaa acagccagtc 60

caccagctgg tgaaggaacc ttcaactcct gccttctcgt 100

<210> 259
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 259
gaatatcttg ataaatgagc agtcagcaga acttgtacat atagatctag accaactctt 60

gcgtaaaaag agaagaattt ttatccattc agtgggtgca 100

<210> 260
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

[0069]

<400> 260
aagagagaaa gagtggattt cagcctgcac ggatggtctt gaaacacaaa tggtctgctt 60

gagaaactta aagaacaaga gtgtgatgtg aaggattatg 100

<210> 261
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 261
acagaattca tacgagaagg cgacgatgac cggactgtgt gccgggagat ccgccgcgcc 60

tagtgggagg ccccatggac gccagcgtgg aggaggaggg 100

<210> 262
<211> 100
<212> DNA
<213>人工序列

<220>

<223>人工序列的描述：合成聚核苷酸

<400> 262
cagaattcat acgagaaggc gacgatgacc ggactgtgtg ccgggagatc cgcctagggg 60

tctgtaccgg gctggcctgt gcctatcacc tcttatgcac 100

<210> 263
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 263
ggacctaggg gaggattacc cctctggcaa gaagcgtgcg gggaccgatg gtgcctatga 60

tcagaagcca caagtgggaa tgaggccctc caacccccc 100

[0070]

<210> 264
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 264
agttcaggaa gtggatgaaa tggtatggaa aaaagcatgc agaatacaca gtggagattc 60

tggacgcaaa gacaagggag aagctgtgtt tcttggacaa 100

<210> 265
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 265
agttcaggaa gtggatgaaa tggtatggaa aaaagcatgc agaatacaca gtggagattc 60

tggacgcaaa gacaagggag aagctgtgtt tcttggacaa 100

<210> 266
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 266
atcaacgaaa tgcacctttt aatccagcag gcccgagaga tgcccctgct caaggcctgg 60

catgcgtggg cagtgatgaa cttcgaagct gtgctacact 100

<210> 267
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 267
[0071] tccattagag atggctctgg cgttggtgtt ttcaaaatag gtgaagagac aggcagttgg 60

agcagaactt tttcaactgc agctgtgaca tccgctggat 100

<210> 268
<211> 96
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 268
gtccctcagg cggccaccca gtgggcacac tcccaggcgg cgctccggcc ccgagaacct 60

cgaagacatc ggcagaaact agatgatcag accaag 96

<210> 269
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 269
aggcacactc aaacaacgac tggtcctcac tcacaactga taaggcttcc cagcctctga 60

agtatctcct gggcattggg ctccttgcag agaggcttga 100

<210> 270
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 270
gctgtggcct tcatgtcctg tgccccagag atctgcccct cccacccgag ggaaattcaa 60

ctactgagga ggttacggca caaaaatgtc atccagctgg 100

<210> 271
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

[0072]

<400> 271
gagtacgaac cggccaagag gttctccatc cggcagatcc ggcagcacag gtcccccgtc 60

atgcgtgccc ggaagagcag cttcaacgtg agtgatgtgg 100

<210> 272
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 272
ccggcccaga gcaaggaagt gttatcattc cagccacctc caccacctgc tgctccatcg 60

cagcttcgct ctgtgatgct aatcttcttc accatatgga 100

<210> 273
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 273
aagaagtggc catgcgcaag ctggtgcgct cagtgactgt ggttgaggac acagcaggat 60

gattgagaat aacaacctga gtgcttctcc cagggcgtgg 100

<210> 274
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 274
ggaaaatata taccatgatc tacaggaact tggtagtagt caatcagcag gctggagcac 60

ctggagaagc aggtcagact gctcaacatc cgtctcaacc 100

[0073]
<210> 275
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 275
tggagaggca aggtcagtta ccaggactac gagattgaga tttcggatgc ttcagactgt 60

cccggaccca gcagctcata tcaaggaagc cttatcagtt 100

<210> 276
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 276
ctgaaacaat tgaggattct agtaatcaag cagcagccag agaatgggag tggaagcaag 60

caatttcttc aaccgtcctt ggaaaagtaa tagttcaacc 100

<210> 277
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 277
tgctctggtg gatatgtttg ttttagccag tccatccaag acaatcccac cgctgtcctg 60

ccttcgagag ggcagctccc actggagatc caagtacctg 100

<210> 278
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 278
[0074] ggtccgaatc agtggtggga acgacaaaca aggtttcccc atgaagcagg gaaattcttt 60

ggaagaagtg gtccatgctg atcagagctc ctgcactttt 100

<210> 279
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 279
gaggaggcac ggtctcggtg gggcccgaga ggccggcggg gcctcccggg aggaaaacta 60

ctgtcgaaat cctcgagggg aagaaggggg accctggtgt 100

<210> 280
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 280
ccctgaacca ctccactgcc tttaaccctc agcctcagag tcagatgcag ggatgaaaat 60

gggataaata gaccagtctg ttcctatgtt aaaccacttc 100

<210> 281
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 281
acgcccagtc cctgggtgcc gagacctgcc ccctgcctag tttccaggtg aactcctccc 60

tcacctcccc gacggggcga ggctccatgg ctgccccctc 100

<210> 282
<211> 100
<212> DNA
<213>人工序列

[0075]

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 282
caagacgttg acttggatct gtcaggtgaa gtcctaaagc ttgcattcca cctggaaacg 60

gtggcctcca acgccgctcc cccctcccgg gaatggaggc 100

<210> 283
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 283
aggtctttga gcttcttcac tgactccaga gaaaaggaga aatttccatc ctgccgcgcc 60

gcgctcctca cacccgcttt cacctccggg cggggcaggg 100

<210> 284
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 284
gtgtgatggc gtgtcacact gccccggcgg ggaggacgag aatcggtgtg acccagtcat 60

ggcagcctcc agcatcagtt caccatgggg aaagcatgtg 100

<210> 285
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 285
[0076] gactggaacg agaactacgg gcgggcggcc tgcagggaca tgggctataa gaaagcatgt 60

gttcaaagcc attctgatgg tcctagtggc ccttatcctc 100

<210> 286
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 286
cgtcgtctgc acgcagccca aatccccatc cgggacagtg tgcacctcaa aaaacacttg 60

gtagacggga ctcgagtgat gattgggaga ttcctgatgg 100

<210> 287
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 287
gaaagcgggc gagcagcagt tgagcgagcc cgaggacatg gagatggaag gacccccgca 60

gcaaacacaa gtttaagatc cacacgtact ccagccccac 100

<210> 288
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 288
[0077] tgcaatgacg agggcctgga gtgtgtgccc actgaggagt ccaacatcac catgcagacc 60

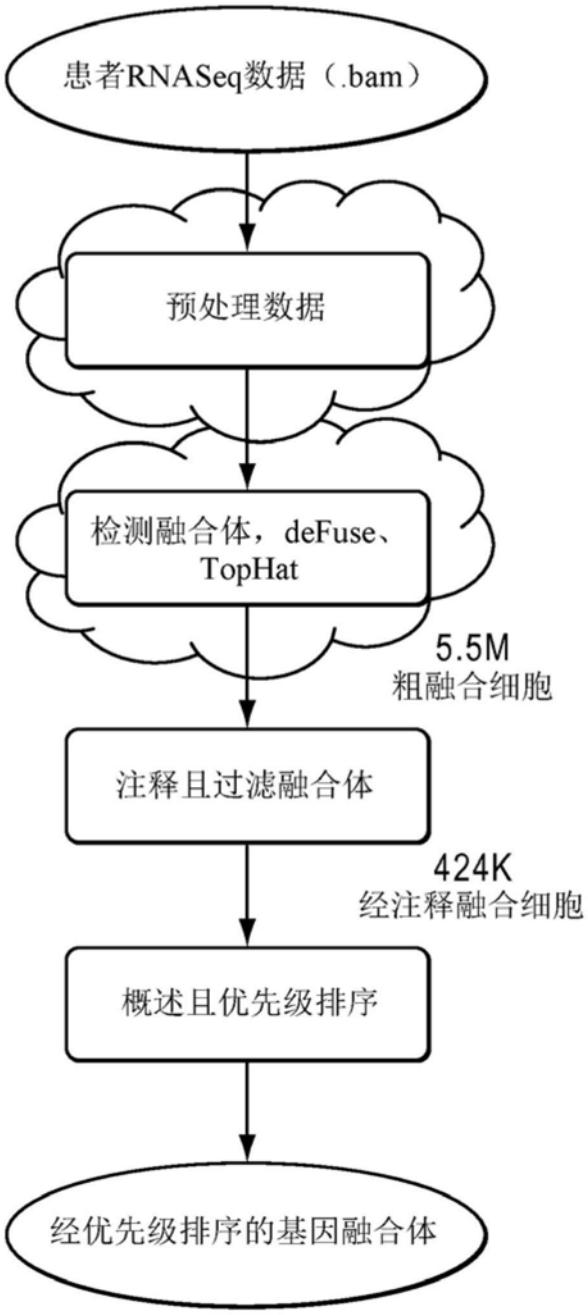
gggtacaaca agctctgtga ttggtggtcg cttggggtga 100

<210> 289
<211> 100
<212> DNA
<213>人工序列

<220>
<223>人工序列的描述：合成聚核苷酸

<400> 289
tgcaatgacg agggcctgga gtgtgtgccc actgaggagt ccaacatcac catgcagacc 60

gggtacaaca agctctgtga ttggtggtcg cttggggtga 100




基因融合体RNASeq数据处理工作流

图1

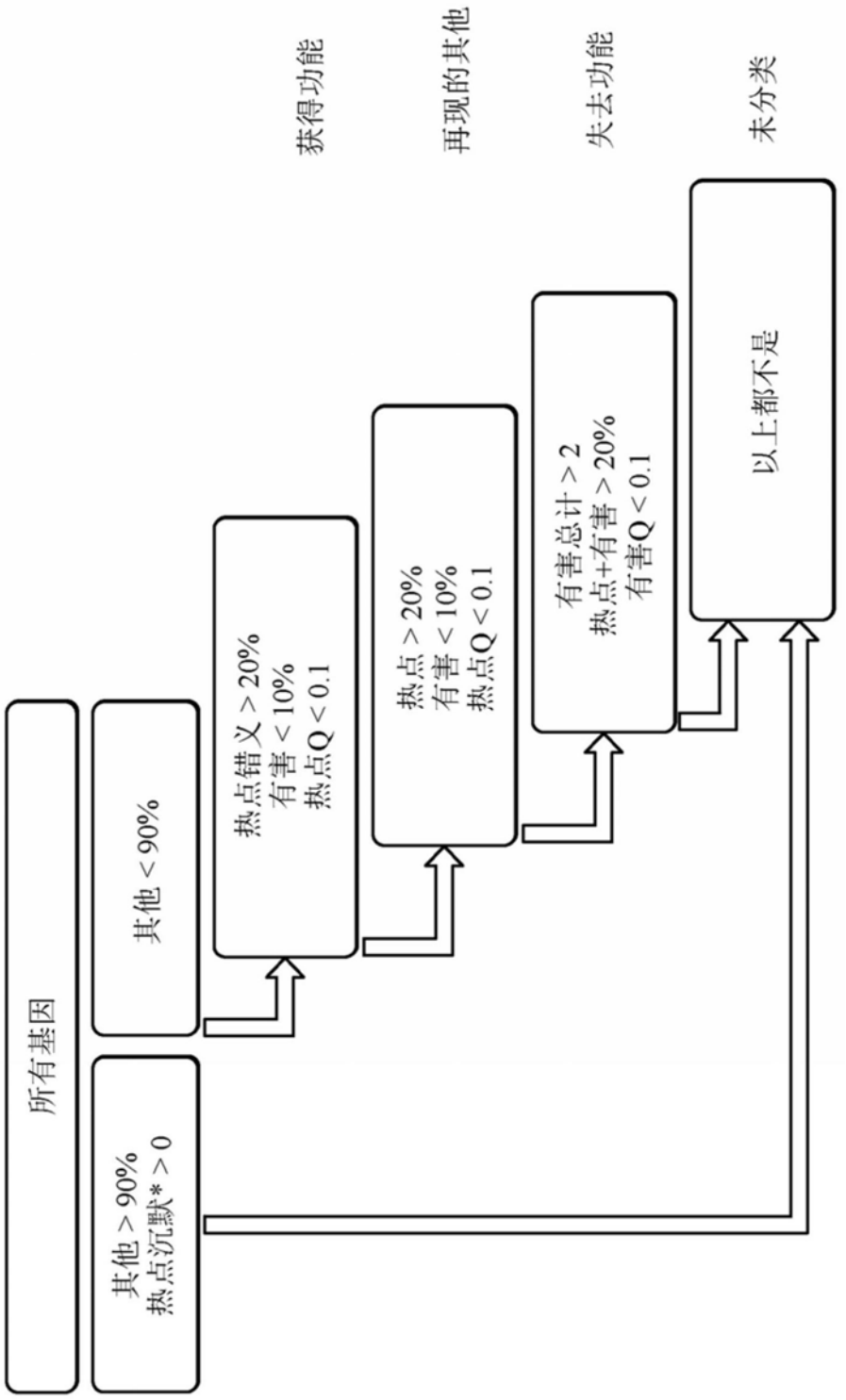

所有基因
其他 > 90%
热点沉默* > 0
其他 < 90%
热点错义 > 20%
有害 < 10%
热点Q < 0.1
热点 > 20%
有害 < 10%
热点Q < 0.1
有害总计 > 2
热点+有害 > 20%
有害Q < 0.1
以上都不是
获得功能
再现的其他
失去功能
未分类

图2

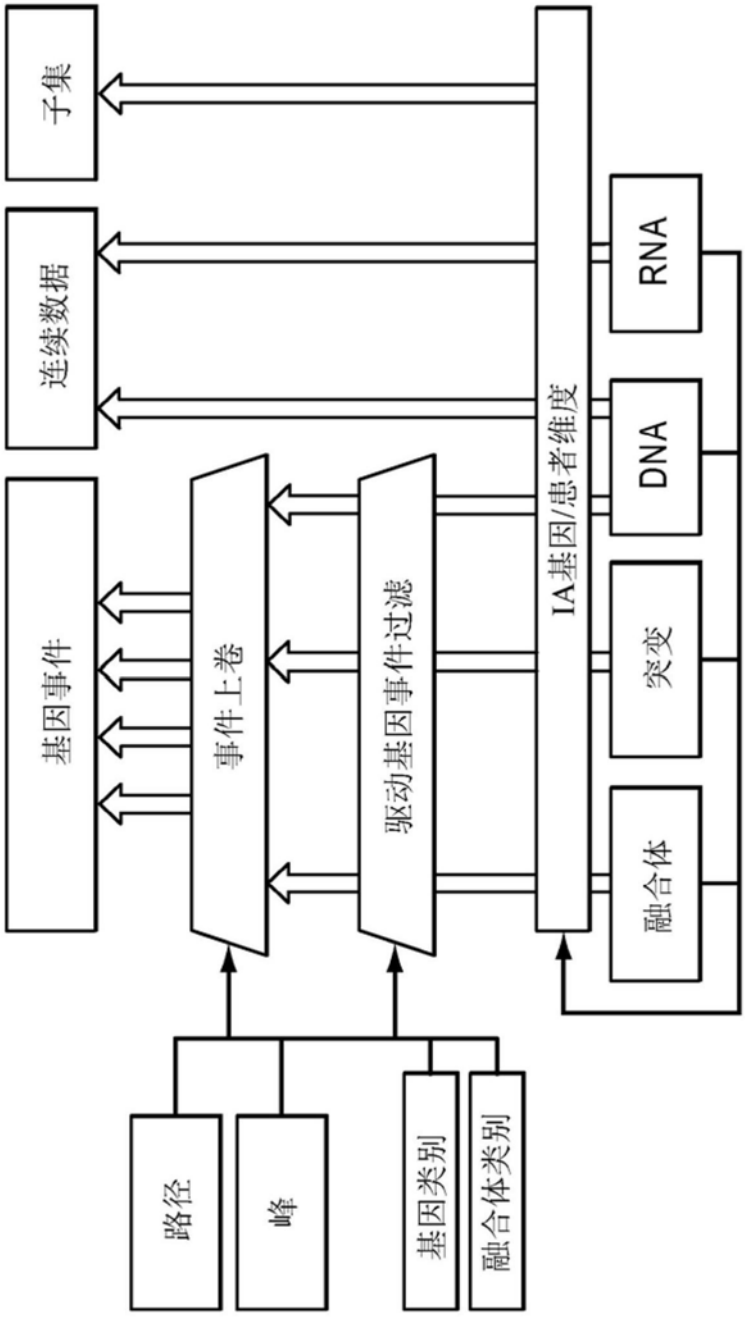

子集
连续数据
基因事件
事件上卷
驱动基因事件过滤
IA基因/患者维度
RNA
DNA
突变
融合体
路径
峰
基因类别
融合体类别

图3

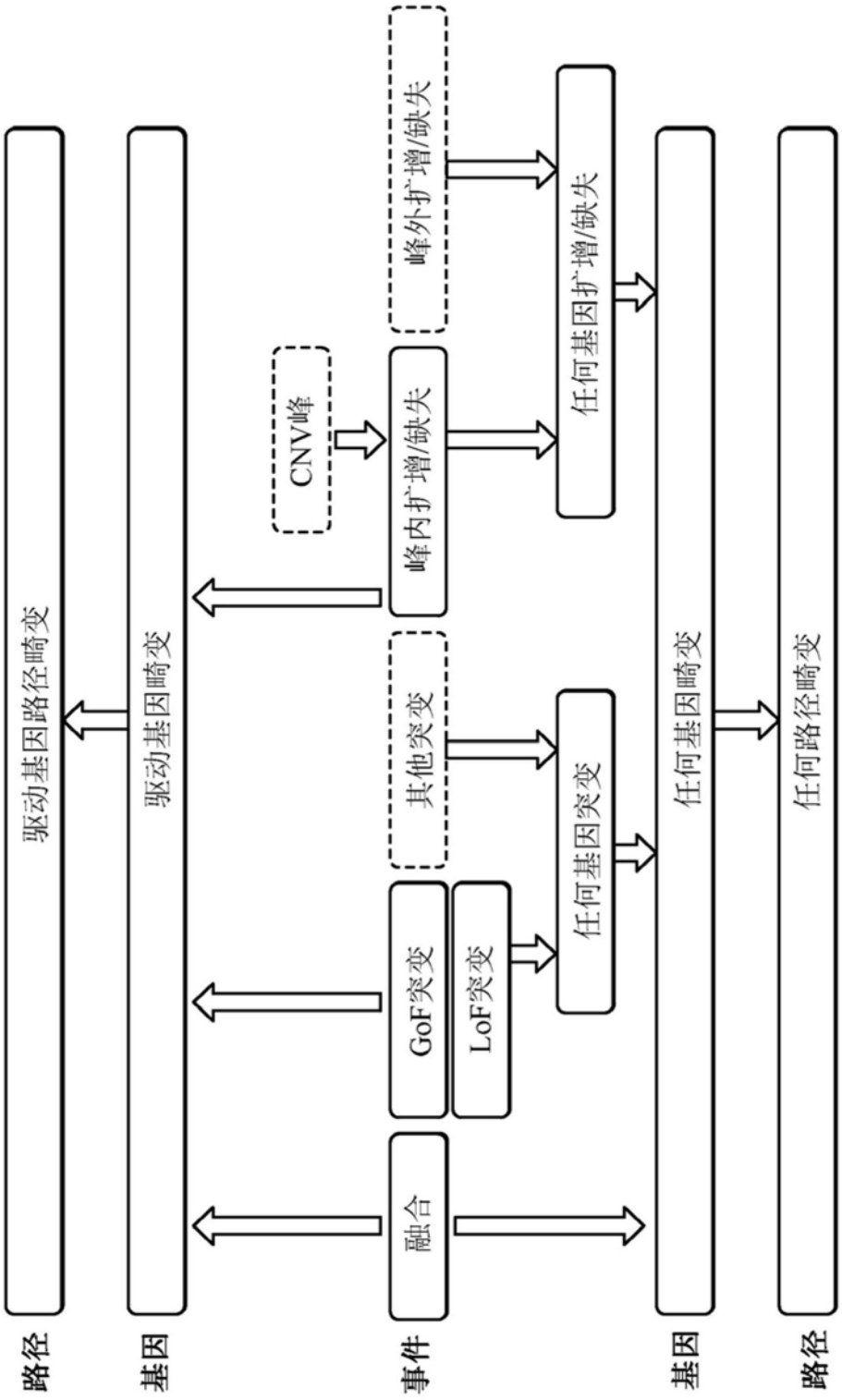

路径
基因
事件
基因
路径
驱动基因路径畸变
驱动基因畸变
融合
GoF突变
LoF突变
其他突变
CNV峰
峰内扩增/缺失
峰外扩增/缺失
任何基因突变
任何基因扩增/缺失
任何基因畸变
任何路径畸变

图4

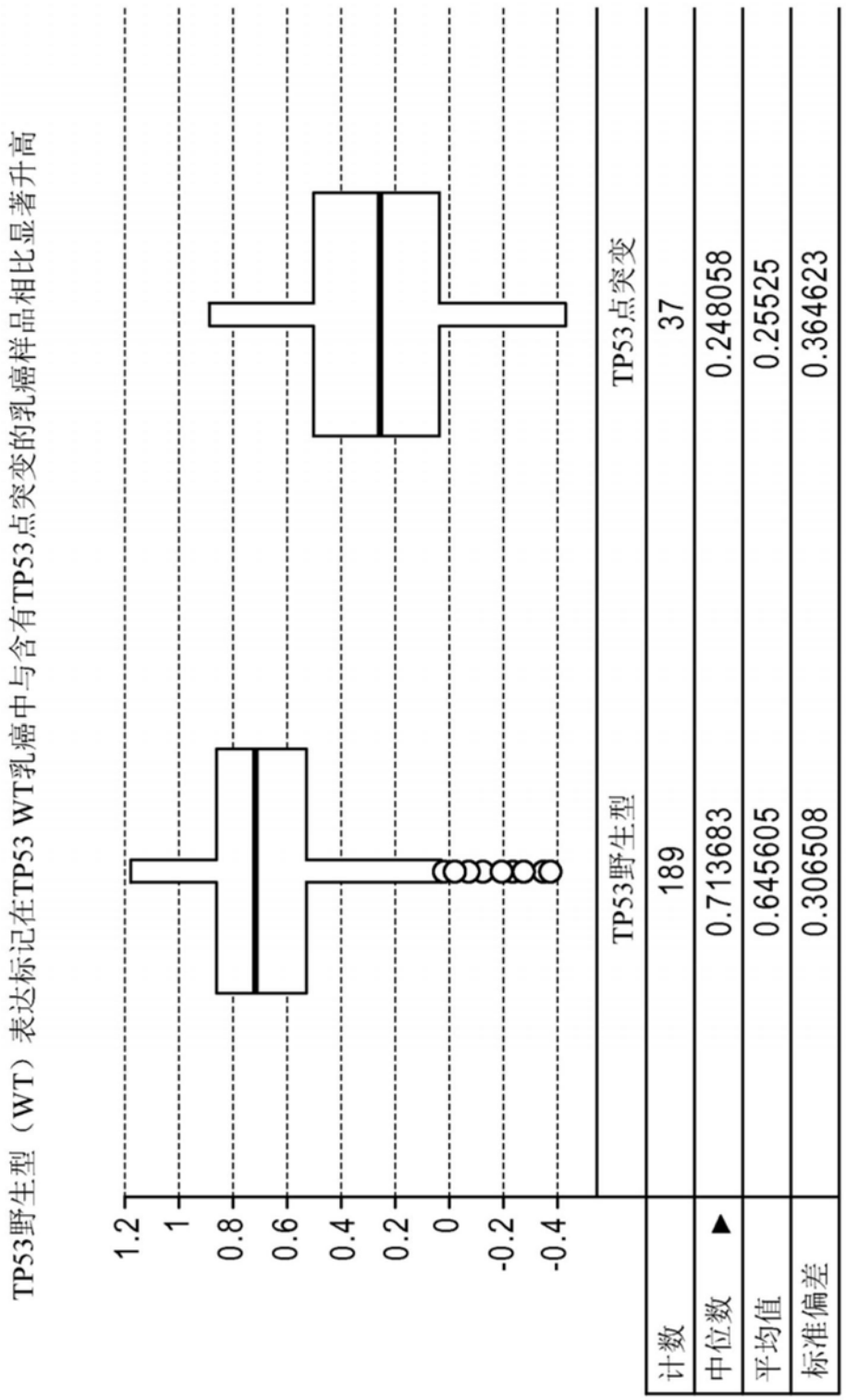

TP53野生型（WT）表达标记在TP53 WT乳癌中与含有TP53点突变的乳癌样品相比显著升高
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
TP53野生型
TP53点突变
计数 189 37
中位数 ▲ 0.713683 0.248058
平均值 0.645605 0.25525
标准偏差 0.306508 0.364623

图5

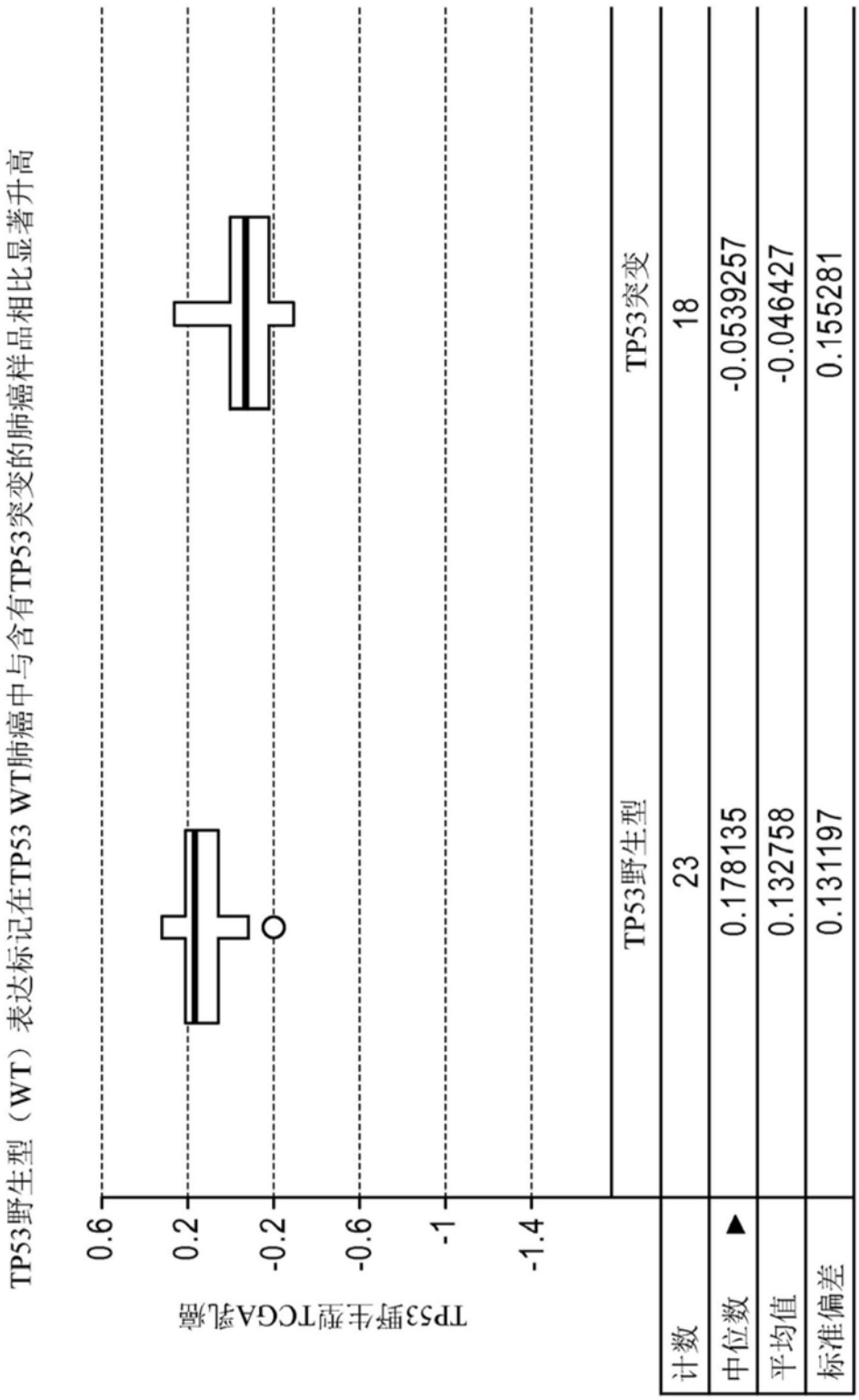

TP53野生型（WT）表达标记在TP53 WT肺癌中与含有TP53突变的肺癌样品相比显著升高
TP53野生型TCGA乳癌
0.6
0.2
-0.2
-0.6
-1
-1.4
TP53野生型
TP53突变
计数 23 18
中位数 0.178135 -0.0539257
平均值 0.132758 -0.046427
标准偏差 0.131197 0.155281

图6

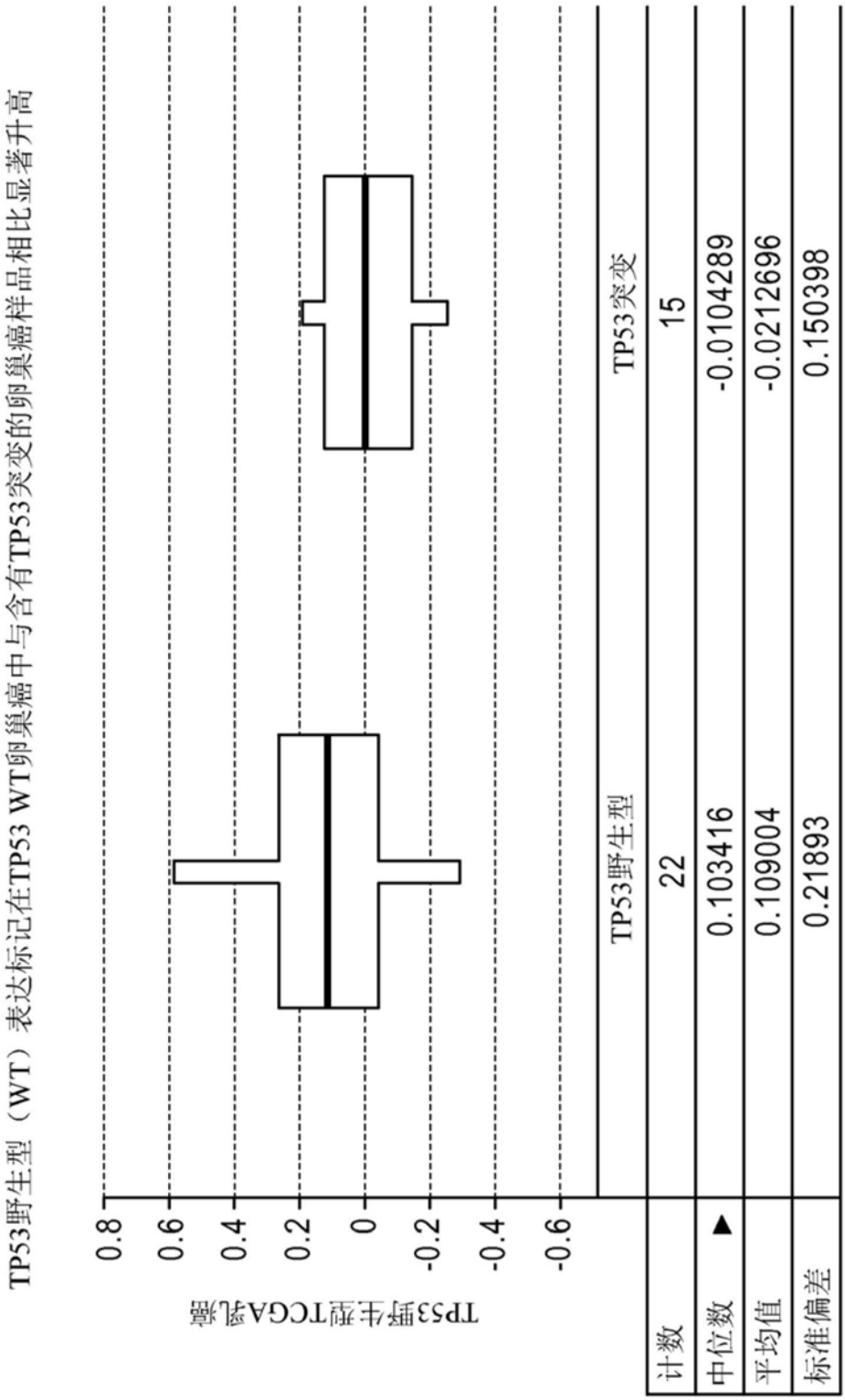

TP53野生型（WT）表达标记在TP53 WT卵巢癌中与含有TP53突变的卵巢癌样品相比显著升高
TP53野生型TCGA乳癌
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
TP53野生型
TP53突变
计数 22 15
中位数 ▲ 0.103416 -0.0104289
平均值 0.109004 -0.0212696
标准偏差 0.21893 0.150398

图7

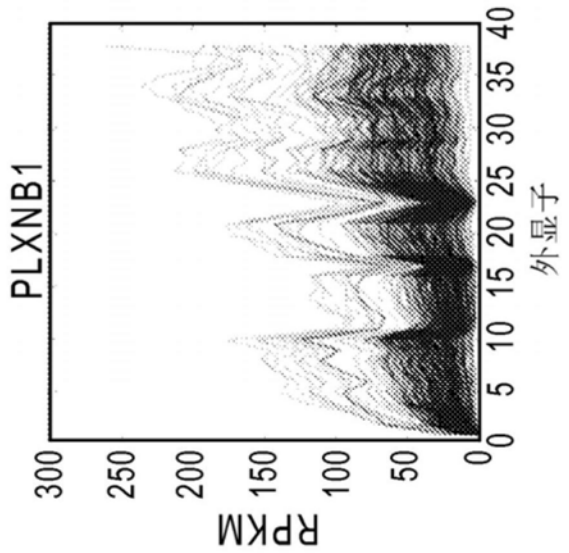

PLXNB1
RPKM
300
250
200
150
100
50
0
0
5
10
15
20
25
30
35
40
外显子

图8A

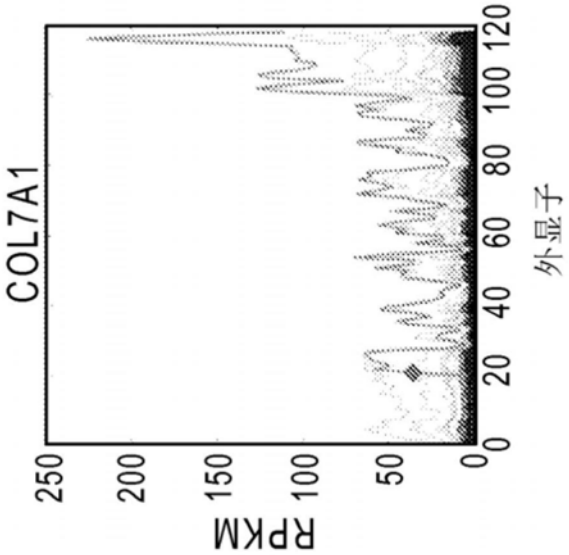

COL7A1
RPKM
250
200
150
100
50
0
0
20
40
60
80
100
120
外显子

图8B

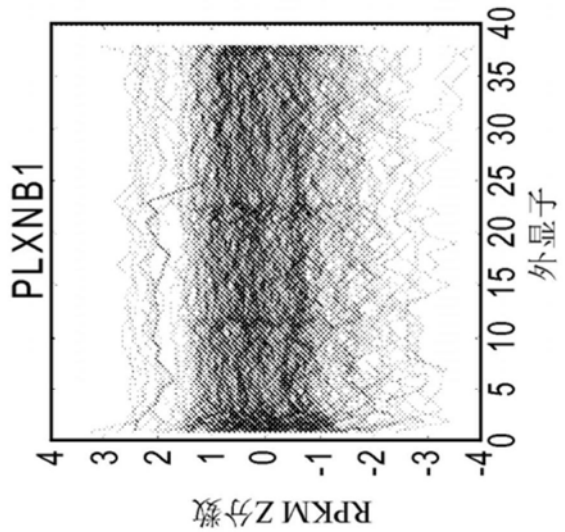

PLXNB1
RPKM Z分数
外显子
4
3
2
1
0
-1
-2
-3
-4
0
5
10
15
20
25
30
35
40

图8C

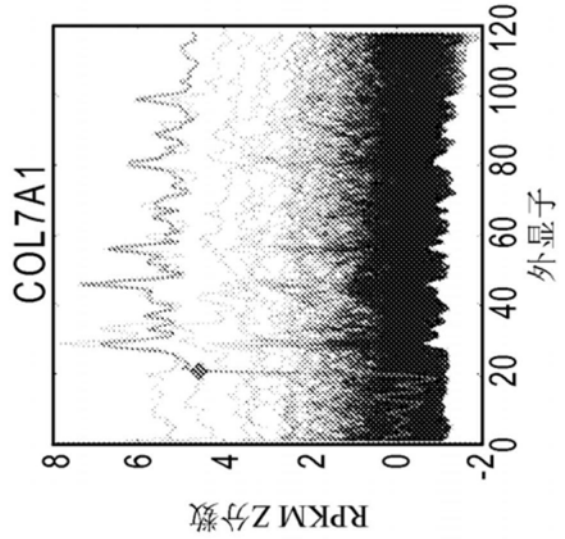

COL7A1
RPKM Z分数
外显子
8
6
4
2
0
-2
0
20
40
60
80
100
120

图8D

| 变异体位置 | 变异体等位基因的人类基因组构建和核酸起始位置 | 泛癌症患者计数 |
|---|---|---|
| p.R273 | 36_7517845, 36_7517846, 37_7577120, 37_7577121 | 134 |
| p.R248 | 36_7518263, 36_7518264, 37_7577538, 37_7577539 | 95 |
| p.R175 | 36_7519131, 36_7519132, 37_7578406, 37_7578407 | 69 |
| p.G245 | 36_7518272, 36_7518273, 37_7577547, 37_7577548 | 45 |
| p.R213 | 36_7518937, 37_7578212 | 41 |
| p.H179 | 36_7519118, 36_7519119, 37_7578393, 37_7578394, 37_7578395 | 40 |
| p.Y220 | 36_7518915, 37_7578190, 37_7578191 | 38 |
| p.H193 | 36_7518996, 36_7518997, 37_7578271, 37_7578272 | 34 |
| p.R282 | 36_7517819, 37_7577093, 37_7577094 | 34 |
| p.R196 | 36_7518988, 37_7578263 | 33 |
| p.R249 | 36_7518259, 36_7518261, 37_7577534, 37_7577535, 37_7577536 | 30 |
| p.R158 | 36_7519183, 37_7578457, 37_7578458 | 27 |
| p.R195 | 36_7518990, 36_7518991, 37_7578264, 37_7578265, 37_7578266 | 24 |
| p.C176 | 36_7519128, 37_7578403, 37_7578404 | 23 |
| p.G266 | 36_7517866, 36_7517867, 37_7577141, 37_7577142 | 22 |
| p.P278 | 36_7517830, 36_7517831, 37_7577105, 37_7577106 | 22 |
| p.T125 | 36_7520037, 36_7520038, 37_7579312, 37_7579313 | 21 |
| p.C238 | 36_7518293, 37_7577567, 37_7577568, 37_7577569 | 20 |
| p.V157 | 36_7519186, 37_7578460, 37_7578461 | 19 |
| p.V173 | 36_7519138, 37_7578412, 37_7578413 | 19 |
| p.R306 | 36_7517747, 37_7577022 | 18 |
| p.R280 | 36_7517824, 36_7517825, 37_7577098, 37_7577099,37_7577100 | 16 |
| p.Y163 | 36_7519167, 36_7519168, 37_7578442, 37_7578443 | 16 |
| p.K132 | 36_7519259, 36_7519260, 36_7519261, 37_7578534, 37_7578535, 36_7578536 | 15 |
| p.S241 | 36_7518284, 37_7577559, 37_7577560 | 15 |
| p.Y234 | 36_7518305, 36_7518306, 37_7577580 | 15 |
| p.R337 | 36_7514743, 37_7574017, 37_7574018 | 14 |
| p.V216 | 36_7518927, 36_7518928, 37_7578202, 37_7578203 | 14 |

根据氨基酸位置频繁的TP53突变

图9A

| 变异体分类 | 总体频率<br>（n = 5,309个患者） | 相对于所有观察到的TP53突变的频率<br>（n = 1,985个总TP53突变） |
|---|---|---|
| 错义突变 | 2.52% | 6.75% |
| 错义突变 | 1.79% | 4.79% |
| 错义突变 | 1.30% | 3.48% |
| 错义突变 | 0.85% | 2.27% |
| 无义突变 | 0.77% | 2.07% |
| 错义突变 | 0.75% | 2.02% |
| 错义突变 | 0.72% | 1.91% |
| 错义突变 | 0.64% | 1.71% |
| 错义突变 | 0.64% | 1.71% |
| 无义突变 | 0.62% | 1.66% |
| 错义突变 | 0.57% | 1.51% |
| 错义突变 | 0.51% | 1.36% |
| 错义突变 | 0.45% | 1.21% |
| 错义突变 | 0.43% | 1.16% |
| 错义突变 | 0.41% | 1.11% |
| 错义突变 | 0.41% | 1.11% |
| 剪接位点 | 0.40% | 1.06% |
| 错义突变 | 0.38% | 1.01% |
| 错义突变 | 0.36% | 0.96% |
| 错义突变 | 0.36% | 0.96% |
| 无义突变 | 0.34% | 0.91% |
| 错义突变 | 0.30% | 0.81% |
| 错义突变 | 0.30% | 0.81% |
| 错义突变 | 0.28% | 0.76% |
| 错义突变 | 0.28% | 0.76% |
| 错义突变 | 0.28% | 0.76% |
| 错义突变 | 0.26% | 0.71% |
| 错义突变 | 0.26% | 0.71% |

根据氨基酸位置频繁的TP53突变

图9B

| 变异体位置 | 变异体等位基因的人类基因组构建和核酸起始位置 | 患者计数 |
|---|---|---|
| p.P177 | 36_7519107, 37_7578384 | 4 |
| p.I255 | 36_7518239, 37_7577515 | 2 |
| p.L252 | 37_7577519, 37_7577525 | 2 |
| p.P128 | 37_7578540, 37_7578542 | 2 |
| p.P191 | 36_7519000 | 2 |
| p.T155 | 37_7578446, 37_7578456 | 2 |
| p.A138 | 37_7578505 | 1 |
| p.A161 | 37_7578444 | 1 |
| p.C277 | 37_7577107 | 1 |
| p.F113 | 37_7579346 | 1 |
| p.G112 | 37_7579343 | 1 |
| p.I162 | 37_7578443 | 1 |
| p.I162 | 37_7578443 | 1 |
| p.K139 | 37_7578506 | 1 |
| p.L264 | 36_7517871 | 1 |
| p.M237 | 36_7518288 | 1 |
| p.N239 | 37_7577558 | 1 |
| p.R110 | 37_7579356 | 1 |
| p.R267 | 37_7577120 | 1 |
| p.R273 | 37_7577102 | 1 |
| p.T253 | 36_7518239 | 1 |
| p.V218 | 37_7578195 | 1 |
| p.V274 | 37_7577115 | 1 |

TP53框内插入和缺失突变很罕见

图10A

| 变异体分类 | 总体频率<br>（n = 5,309个患者） | 相对于所有观察到的TP53突变的频率<br>（n = 1,985个总TP53突变） |
|---|---|---|
| 框内缺失 | 0.08% | 0.20% |
| 框内缺失 | 0.04% | 0.10% |
| 框内缺失 | 0.04% | 0.10% |
| 框内缺失 | 0.04% | 0.10% |
| 框内缺失 | 0.04% | 0.10% |
| 框内缺失 | 0.04% | 0.10% |
| 框内缺失 | 0.02% | 0.05% |
| 框内缺失 | 0.02% | 0.05% |
| 框内缺失 | 0.02% | 0.05% |
| 框内缺失 | 0.02% | 0.05% |
| 框内缺失 | 0.02% | 0.05% |
| 框内缺失 | 0.02% | 0.05% |
| 框内插入 | 0.02% | 0.05% |
| 框内缺失 | 0.02% | 0.05% |
| 框内插入 | 0.02% | 0.05% |
| 框内缺失 | 0.02% | 0.05% |
| 框内缺失 | 0.02% | 0.05% |
| 框内缺失 | 0.02% | 0.05% |
| 框内缺失 | 0.02% | 0.05% |
| 框内缺失 | 0.02% | 0.05% |
| 框内缺失 | 0.02% | 0.05% |
| 框内缺失 | 0.02% | 0.05% |
| 框内插入 | 0.02% | 0.05% |

TP53框内插入和缺失突变很罕见

图10B