



(12) 发明专利

(10) 授权公告号 CN 106661624 B

(45) 授权公告日 2021. 04. 06

(21) 申请号 201580036201.X

(22) 申请日 2015.05.18

(65) 同一申请的已公布的文献号
申请公布号 CN 106661624 A

(43) 申请公布日 2017.05.10

(30) 优先权数据
61/994,838 2014.05.17 US

(85) PCT国际申请进入国家阶段日
2016.12.30

(86) PCT国际申请的申请数据
PCT/US2015/031413 2015.05.18

(87) PCT国际申请的公布数据
W02015/179312 EN 2015.11.26

(73) 专利权人 加利福尼亚大学董事会
地址 美国加利福尼亚州

(72) 发明人 张卫国 加里·卡彭 毛建华

(74) 专利代理机构 北京英赛嘉华知识产权代理
有限责任公司 11204
代理人 王达佐 洪欣

(51) Int.Cl.
C12Q 1/6886 (2018.01)
G01N 33/53 (2006.01)
G01N 33/574 (2006.01)
C12N 15/11 (2006.01)

审查员 季璇馨

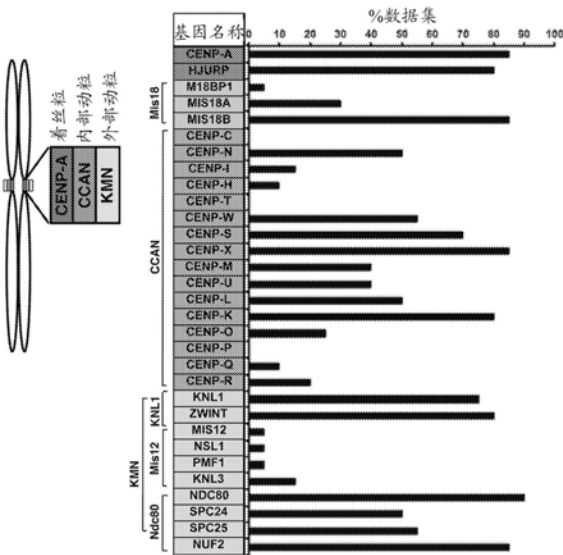
权利要求书3页 说明书62页 附图18页

(54) 发明名称

用于癌症诊断、预后和治疗选择的着丝粒/
动粒蛋白基因

(57) 摘要

本申请所述的发明涉及患有肿瘤的患者的一组基因表达标志物。因此,本发明提供用于评估标志物的基因表达水平的方法和组合物(例如试剂盒),以及使用此类基因表达水平来评估疾病进展或对化疗或放疗的应答的可能性的方法。可以将此类信息用于为癌症患者确定预后和治疗选择。



1. 为患有乳腺癌、肺癌、胃癌或卵巢癌肿瘤的患者评估疾病进展的可能性的试剂盒,其使用包括以下的方法:

检测来自所述患者的肿瘤样本中一组14种基因或所述组的至少9种基因的子集中各成员的RNA表达水平,其中所述14种基因是:CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-W、CENP-U、CENP-L、CENP-K、ZWINT、NDC80、NUF2、SPC24和SPC25;以及

将所述表达水平与所述疾病进展的可能性相关联,其包括确定所述组中各成员的所述表达水平相对于由正常细胞确定的参考水平的标准化值;并将各基因表达的所述标准化值相加以产生风险评分,其中在来自与所述患者具有相同癌症的参考对象群体的组的基因表达评分的参考标度前5%、前10%、前15%、前20%、前25%、前30%、前35%、前40%、前45%、前50%或前60%的高风险评分指示所述患者具有疾病进展的可能性。

2. 评估患有乳腺癌、肺癌或卵巢癌肿瘤的患者中响应化疗或放疗的可能性的试剂盒,其使用包括以下的方法:

检测来自所述患者的肿瘤样本中一组9种生物标志物的RNA表达水平,其中所述9种生物标志物是:CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-U、ZWINT、NDC80和SPC25;以及

将所述表达水平与响应的可能性相关联,其包括确定所述组中各成员的所述表达水平相对于由正常细胞确定的参考水平的标准化值;并将各生物标志物表达的所述标准化值相加以产生风险评分,其中在来自与所述患者具有相同癌症的参考对象群体的组的基因表达评分的参考标度前5%、前10%、前15%、前20%、前25%、前30%、前35%、前40%、前45%、前50%或前60%的高风险评分指示所述患者具有响应化疗或放疗的可能性。

3. 如权利要求1或2所述的试剂盒,其中检测所述RNA表达水平包括进行扩增分析、杂交分析、测序分析或阵列。

4. 评估患有乳腺癌、肺癌、胃癌或卵巢癌肿瘤的患者中疾病进展的可能性的试剂盒,其使用包括以下的方法:

检测来自所述患者的肿瘤样本中一组14种生物标志物或所述组的至少9种生物标志物的子集的表达水平,其中所述14种生物标志物是:CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-W、CENP-U、CENP-L、CENP-K、ZWINT、NDC80、NUF2、SPC24和SPC25;以及

将所述表达水平与所述疾病进展的可能性相关联,其包括确定所述组中各成员的所述表达水平相对于由正常细胞确定的参考水平的标准化值;并将各生物标志物表达的所述标准化值相加以产生风险评分,其中当所述风险评分在来自与所述患者具有相同癌症的参考对象群体的组的基因表达评分的参考标度的最高三分位数中时,所述患者具有疾病进展的可能性。

5. 评估患有乳腺癌、肺癌或卵巢癌肿瘤的患者中响应化疗或放疗的可能性的试剂盒,其使用包括以下的方法:

检测来自所述患者的肿瘤样本中一组9种生物标志物的表达水平,其中所述9种生物标志物是:CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-U、ZWINT、NDC80和SPC25;以及

将所述表达水平与所述响应相关联,其包括确定所述组中各成员的所述表达水平相对于由正常细胞确定的参考水平的标准化值;并将各生物标志物表达的所述标准化值相加以产生风险评分,其中当所述风险评分在来自与所述患者具有相同癌症的参考对象群体的组的基因表达评分的参考标度的最高三分位数中时,所述患者具有响应化疗或放疗的可能

性。

6. 如权利要求4或5所述的试剂盒,其中所述表达水平是蛋白表达水平。

7. 如权利要求6所述的试剂盒,其中检测所述蛋白表达水平包括进行免疫分析。

8. 如权利要求4或5所述的试剂盒,其中所述表达水平是RNA表达水平。

9. 如权利要求8所述的试剂盒,其中检测所述RNA表达水平包括进行扩增分析、杂交分析、测序分析或阵列。

10. 如权利要求4或5所述的试剂盒,其中所述参考标度是来源于患有与所述患者相同类型的肿瘤的参考患者群体的多个风险评分。

11. 为患有乳腺癌、肺癌、胃癌或卵巢癌肿瘤的患者评估疾病进展的可能性的试剂盒,其使用包括以下的方法:

检测来自所述患者的肿瘤样本中一组14种基因或所述组的至少9种基因的子集中各成员的RNA表达水平,其包括进行定量聚合酶链式反应(qPCR)以检测所述组中各基因的表达水平,并且其中所述14种基因是:CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-W、CENP-U、CENP-L、CENP-K、ZWINT、NDC80、NUF2、SPC24和SPC25;以及

将所述RNA表达水平与所述疾病进展的可能性相关联,其包括确定所述组中各成员的所述表达水平相对于由正常细胞确定的参考水平的标准化值;并将各基因表达的所述标准化值相加以产生范围从0到1的风险评分,其中当所述风险评分是0.66或以上时,所述患者具有疾病进展的可能性。

12. 为患有乳腺癌、肺癌或卵巢癌肿瘤的患者评估响应化疗或放疗的可能性的试剂盒,其使用包括以下的方法:

检测来自所述患者的肿瘤样本中一组9种基因中各成员的RNA表达水平,其包括进行定量聚合酶链式反应(qPCR)以检测所述组中各基因的所述表达水平,并且其中所述9种基因是:CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-U、ZWINT、NDC80和SPC25;以及

将所述RNA表达水平与响应的可能性相关联,其包括确定所述组中各成员的所述表达水平相对于由正常细胞确定的参考水平的标准化值;并将各基因表达的所述标准化值相加以产生范围从0到1的风险评分,其中当所述风险评分是0.66或以上时,所述患者具有响应化疗或放疗的可能性。

13. 确定患者中肿瘤的基因组不稳定性水平的试剂盒,其使用包括以下的方法:

检测取自所述患者的肿瘤的肿瘤样本中一组14种基因或所述组的至少9种基因的子集中各成员的RNA表达水平,其包括进行定量聚合酶链式反应(qPCR)以检测所述组中各基因的所述表达水平,并且其中所述14种基因是:CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-W、CENP-U、CENP-L、CENP-K、ZWINT、NDC80、NUF2、SPC24和SPC25;以及

将所述RNA表达水平与所述基因组不稳定性水平相关联,其包括确定所述组中各成员的所述表达水平相对于由正常细胞确定的参考水平的标准化值;并将各基因表达的所述标准化值相加以产生风险评分,其中在来自与所述患者具有相同癌症的参考对象群体的组的基因表达评分的参考标度前5%、前10%、前15%、前20%、前25%、前30%、前35%、前40%、前45%、前50%或前60%的高风险评分指示所述患者的肿瘤具有高水平的基因组不稳定性,

其中所述肿瘤是乳腺癌、非小细胞肺癌、胃癌、卵巢癌、胶质瘤、结直肠癌、前列腺癌、黑

色素瘤、头颈癌、膀胱癌、肾上腺皮质癌或肾癌。

14. 如权利要求1、2、4、5、11、12和13任一项所述的试剂盒,其中所述肿瘤为早期乳腺癌。

15. 如权利要求14所述的试剂盒,其中所述乳腺癌是管腔A型。

16. 如权利要求14所述的试剂盒,其中所述乳腺癌是管腔B型。

17. 如权利要求14所述的试剂盒,其中所述乳腺癌是ER阳性乳腺癌。

18. 如权利要求1、2、4、5、11、12和13中任一项所述的试剂盒,其中所述肿瘤是早期非小细胞肺癌。

19. 如权利要求1、2、4、5、11、12和13中任一项所述的试剂盒,其中所述肿瘤是早期卵巢癌。

20. 如权利要求1、2、4、5、11、12和13中任一项所述的试剂盒,其中所述肿瘤样本是肿瘤组织样本或肿瘤细胞样本。

21. 用于检测预测存活的一组基因的表达的微阵列,其中所述一组基因由CENP-A、HJURP、M1S18B、CENP-N、CENP-M、CENP-W、CENP-U、CENP-L、CENP-K、ZWINT、NDC80、NUF2、SPC24和SPC25组成,或者由所述一组基因的至少9种基因的子集组成;以及任选地包含用于检测组成型表达的参考基因的表达的探针。

22. 试剂盒,其包含用于检测14种基因或者所述14种基因的至少9种基因的子集的表达的引物和探针,其中所述14种基因是:CENP-A、HJURP、M1S18B、CENP-N、CENP-M、CENP-W、CENP-U、CENP-L、CENP-K、ZWINT、NDC80、NUF2、SPC24和SPC25。

23. 如权利要求22所述的试剂盒,其还包含用于检测组成型表达的参考基因表达的引物和探针。

24. 如权利要求23所述的试剂盒,其还包含说明书。

用于癌症诊断、预后和治疗选择的着丝粒/动粒蛋白基因

[0001] 相关申请的交叉引用

[0002] 本申请要求于2014年5月17日提交的美国临时申请第61/994,838号的优先权,为了所有目的,该公开内容通过引用整体并入。

[0003] 根据联邦资助的研究和开发作出的关于发明的权利的声明

[0004] 本发明是根据由美国能源部授予的合同第DE-AC02-05CH11231号在政府支持下进行的。政府在本发明中有一定的权利。

[0005] 发明背景

[0006] 染色体不稳定性(CIN)是大多数人类癌症的特点,并导致其他癌症性质,如肿瘤异质性和耐药性。着丝粒和动粒由着丝粒染色质,以及内部和外部动粒结构组成,其是在有丝分裂中附着于有丝分裂纺锤体和正常染色体分离所需要的。在正常细胞中,着丝粒和动粒蛋白的水平被严格控制以确保正确的染色体分离。然而,不同类型的肿瘤细胞显示显著改变的着丝粒/动粒蛋白水平。尽管着丝粒和动粒(CEN/KT)基因的调控异常对癌症的进展和结果的直接影响在很大程度上仍然未知,但在多种真核生物中的研究已经证明其与数值CIN(非整倍性,或者整个染色体或片段的增加和丢失)和结构CIN(染色体重排和局部拷贝数改变)两者有直接关系。通过突变或消耗,许多着丝粒和动粒蛋白的损失可以导致整条染色体丢失或增加,而增加的水平产生具有两个着丝粒(双着丝粒)的染色体,导致突变、染色体断裂和易位的频率增加。在癌症领域,本领域需要评估患者癌症的染色体不稳定性的,用于患者诊断、预后和对治疗响应的有效方法。

[0007] 发明概述

[0008] 一方面,本文提供为患有肿瘤的患者评估疾病进展的可能性的方法。该方法包括检测来自患者的肿瘤样本中一组14种基因或该组的至少9种基因的子集中各成员的RNA表达水平,其中所述14种基因是:CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-W、CENP-U、CENP-L、CENP-K、ZWINT、NDC80、NUF2、SPC24和SPC25;以及将表达水平与疾病进展的可能性相关联,其包括确定该组中各成员的表达水平相对于参考水平的标准化值,并将各基因表达的标准化值相加以产生风险评分,其中高风险评分指示患者具有疾病进展的可能性。在一些情况下,所述至少9种基因包括CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-U、ZWINT、NDC80和SPC25。

[0009] 本文还提供评估在患有肿瘤的患者中响应化疗或放疗的可能性的方法,该方法包括检测来自患者的肿瘤样本中一组9种基因的RNA表达水平,其中所述9种基因是:CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-U、ZWINT、NDC80和SPC25;以及将表达水平与响应的可能性相关联,其包括确定该组中各成员的表达水平相对于参考水平的标准化值,并将各生物标志物表达的标准化值相加以产生风险评分,其中高风险评分指示患者具有响应化疗或放疗的可能性。在其他实施方案中,该方法包括检测来自患者的肿瘤样本中一组14种基因或该组的至少9种基因的子集,其中这14种基因是:CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-W、CENP-U、CENP-L、CENP-K、ZWINT、NDC80、NUF2、SPC24和SPC25。所述至少9种基因可以是CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-U、ZWINT、NDC80和SPC25。

[0010] 在一些实施方案中,检测RNA表达水平的步骤包括进行扩增分析、杂交分析、测序分析或微阵列。

[0011] 在一些实施方案中,上述方法还包括向患者推荐施用化疗或放疗。

[0012] 另一方面,本文提供为患有肿瘤的患者评估疾病进展的可能性的方法。该方法包括检测来自患者的肿瘤样本中一组14种生物标志物或该组的至少9种生物标志物的子集的表达水平,其中这14种生物标志物是:CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-W、CENP-U、CENP-L、CENP-K、ZWINT、NDC80、NUF2、SPC24和SPC25;以及将表达水平与疾病进展的可能性相关联,其包括确定该组中各成员的表达水平相对于参考水平的标准化值,并将各生物标志物表达的标准化值相加以产生风险评分,其中当风险评分在参考标度的最高三分位数中时,该患者具有疾病进展的可能性。在一些情况下,所述至少9种生物标志物包括CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-U、ZWINT、NDC80和SPC25。

[0013] 本文还提供评估在患有肿瘤的患者中响应化疗或放疗的可能性的方法,该方法包括检测来自患者的肿瘤样本中一组9种生物标志物的表达水平,其中所述9种生物标志物是:CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-U、ZWINT、NDC80和SPC25;以及将表达水平与响应相关联,其包括确定该组中各成员的表达水平相对于参考水平的标准化值,并将各生物标志物表达的标准化值相加以产生风险评分,其中当风险评分在参考标度的最高三分位数中时,该患者具有响应化疗或放疗的可能性。在一些实施方案中,该方法还包括向患者推荐施用化疗或放疗。在其他实施方案中,该方法包括检测来自患者的肿瘤样本中一组14种生物标志物或该组的至少9种生物标志物的子集,其中这14种基因是:CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-W、CENP-U、CENP-L、CENP-K、ZWINT、NDC80、NUF2、SPC24和SPC25。

[0014] 在一些实施方案中,表达水平是蛋白表达水平。检测蛋白表达水平的步骤可以包括进行免疫分析。在其他实施方案中,表达水平是RNA表达水平。检测RNA表达水平的步骤包括扩增分析、杂交分析、测序分析或微阵列。

[0015] 在一些实施方案中,参考标度是来源于患有与患者相同类型的肿瘤的参考患者群体的多个风险评分。参考标度的风险评分可以是与患者的肿瘤样本相同类型的肿瘤样本的参考标度。

[0016] 另一方面,本文提供为患有肿瘤的患者评估疾病进展的可能性的方法。该方法包括检测来自患者的肿瘤样本中一组14种基因或该组的至少9种基因的子集中各成员的RNA表达水平,其包括进行定量聚合酶链式反应(qPCR)以检测该组中各基因的表达水平,并且其中这14种基因是:CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-W、CENP-U、CENP-L、CENP-K、ZWINT、NDC80、NUF2、SPC24和SPC25;以及将RNA表达水平与疾病进展的可能性相关联,其包括确定该组中各成员的表达水平相对于参考水平的标准化值,并将各基因表达的标准化值相加以产生范围从0到1的风险评分,其中当风险评分是0.66或以上时,例如0.66、0.67、0.70、0.75、0.8、0.85、0.90、0.95、0.99或1.0时,该患者具有疾病进展的可能性。在一些实施方案中,所述至少9种基因包括CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-U、ZWINT、NDC80和SPC25。

[0017] 另一方面,本文提供为患有肿瘤的患者评估响应化疗或放疗的可能性的方法。该方法包括检测来自患者的肿瘤样本中一组9种基因中各成员的RNA表达水平,其包括进行定

量聚合酶链式反应(qPCR)以检测该组中各基因的表达水平,并且其中这9种基因是:CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-U、ZWINT、NDC80和SPC25;以及将RNA表达水平与响应的可能性相关联,其包括确定该组中各成员的表达水平相对于参考水平的标准化值,并将各基因表达的标准化值相加以产生范围从0到1的风险评分,其中当风险评分是0.66或以上时,例如0.66、0.67、0.70、0.75、0.8、0.85、0.90、0.95、0.99或1.0时,该患者具有响应化疗或放疗的可能性。在一些实施方案中,该方法还包括向患者推荐施用化疗或放疗。在另外的实施方案中,一组基因包括14种基因,如CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-W、CENP-U、CENP-L、CENP-K、ZWINT、NDC80、NUF2、SPC24和SPC25。

[0018] 在一些方面,本文提供的方法是用于测定患者中肿瘤的基因组不稳定性水平。该方法包括检测取自患者的肿瘤的肿瘤样本中一组14种基因或该组的至少9种基因的子集中各成员的RNA表达水平,以检测该组中各基因的表达水平,并且其中这14种基因是:CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-W、CENP-U、CENP-L、CENP-K、ZWINT、NDC80、NUF2、SPC24和SPC25;以及将RNA表达水平与基因组不稳定性水平相关联,其包括确定该组中各成员的表达水平相对于参考水平的标准化值,并将各基因的标准化值相加以产生风险评分,其中高风险评分指示患者的肿瘤具有高水平的基因组不稳定性。在一些实施方案中,使用RT-PCR评估RNA表达。在一些实施方案中,所述至少9种,基因包括CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-U、ZWINT、NDC80和SPC25。

[0019] 在上述任一方法中,肿瘤可以是早期乳腺癌。在一些实施方案中,乳腺癌是管腔A型(luminal A)。在其他实施方案中,乳腺癌是管腔B型(luminal B)。在一些实施方案中,乳腺癌是ER阳性乳腺癌。在一些情况下,乳腺癌是已经经受辅助激素疗法的ER阳性肿瘤。在一些情况下,肿瘤是早期非小细胞肺癌。在其他情况下,肿瘤是早期卵巢癌。在其他情况下,肿瘤是在任何阶段的卵巢癌。在一些实施方案中,肿瘤是早期癌症。在其他实施方案中,肿瘤是晚期癌症。

[0020] 在一些实施方案中,肿瘤样本是肿瘤组织样本或肿瘤细胞样本。可以由正常、健康细胞确定参考水平。在一些情况下,可以由来自同一患者的正常、健康细胞确定参考水平。

[0021] 一方面,本文公开的是用于检测预测生存期的一组基因的表达的微阵列,其中所述一组基因由CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-W、CENP-U、CENP-L、CENP-K、ZWINT、NDC80、NUF2、SPC24和SPC25组成,或者该组基因的至少9种基因的子集组成;以及任选地包含用于检测参考基因的表达的探针。在一些情况下,所述一组基因的9种基因是CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-U、ZWINT、NDC80和SPC25。

[0022] 另一方面,本文公开的是试剂盒,其包含用于检测14种基因或者这14种基因的至少9种基因的子集的表达的引物和探针,其中这14种基因是:CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-W、CENP-U、CENP-L、CENP-K、ZWINT、NDC80、NUF2、SPC24和SPC25。在一些实施方案中,试剂盒还包含用于检测参考基因的表达的引物和探针。在其他实施方案中,试剂盒还包含说明书。在一些情况下,所述9种基因是CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-U、ZWINT、NDC80和SPC25。

[0023] 上述方法可以用于实体瘤癌症的诊断或预后。此外,这些方法可用于由临床医生或在临床环境中的其他个体所进行的治疗选择。

[0024] 根据以下详细描述和附图,本发明的其他目的、特征和优势对本领域技术人员将

是显而易见的。

[0025] 附图简要描述

[0026] 图1A-1C提供了CEN/KT基因列表及其在癌症中的调控异常。图1A显示了复制的有丝分裂姐妹染色单体上的着丝粒和动粒结构的示意图。放大图显示了作为着丝粒染色质和动粒形成的结构基础的CENP-A核小体,以及连接CENP-A染色质和位于外部动粒中的KMN网络(KNL1、ZWINT、MIS12、NSL1、PMF1、KNL3、NDC80、SPC24、SPC25和NUF2)的内部动粒中的CCAN网络(CENP-C、CENP-N、CENP-I、CENP-H、CENP-T、CENP-W、CENP-S、CENP-X、CENP-M、CENP-U、CENP-L、CENP-K、CENP-O、CENP-P、CENP-Q和CENP-R)。HJURP分子伴侣(直接结合CENP-A)和Mis18复合物是CENP-A装配因子。图1B显示了31个CEN/KT基因的列表。显示的区域包括Mis18(灰色)和HJURP(紫色),其瞬时定位于着丝粒用于CENP-A装配。用于CENP-P的Affymetrix探针不通过我们的限定过滤器,并且随后被从所有分析中移除。与相应的正常组织相比,三十一个CEN/KT基因中的十五个在不同癌症的>50%数据集中表达异常(倍数变化>2倍,FDR $p < 0.05$) (图1C)。

[0027] 图2显示了比较三个CEN/KT基因列表的维恩图,并鉴定了14个CES基因作为表达水平与癌症进展相关的核心子集。列表包括31个CEN/KT基因;如在图1C中,在不同癌症和相应的正常组织之间差异表达的CEN/KT基因;使用BC GenExMiner,用于乳腺癌预后的CEN/KT基因;以及使用K-M Plotter,用于肺癌预后的CEN/KT基因。

[0028] 图3A-图3F显示,使用乳腺癌和肺癌数据集,着丝粒和动粒基因表达评分(CES)是用于总存活率、无病存活率和无远端存活率的预后值。对于每个指定的GEO数据集,将患者分为高CES组、中CES组和低CES组。显示了GEO数据集和对数秩p值。卡普兰-迈耶生存曲线显示CES是用于乳腺癌和肺癌患者的总存活率(OS)的预后值(图3A)。卡普兰-迈耶生存曲线显示CES是用于乳腺癌和肺癌患者的无病存活率(DFS)的预后值(图3B)。卡普兰-迈耶生存曲线显示CES是用于ER⁺乳腺癌患者的无远端转移存活率(DMFS)的预后值(图3C)。图3D-图3F是森林图,其中正方形表示HR,以及棒的左右限度指示95%CI。森林图显示了与具有亚型或阶段信息的各种癌症中的高CES相关的风险比(图3D)。特别地,高CES预测早期(第I阶段和第II阶段组合)卵巢癌、非小细胞肺癌(NSCLC)、第I阶段肺腺癌(ADC)、胃癌和乳腺癌管腔亚型的不良OS。降低的HR与更具侵袭性的HER2+乳腺癌和基底样乳腺癌相关。高CES与不同癌症以及亚型或阶段的复发或疾病进展的增加了的风险相关(图3E)。高CES与管腔A型和管腔B型乳腺癌的DMFS的增加了的风险相关(图3F)。HER2+和基底样亚型的趋势相反。

[0029] 图4A-图4D显示,CES值预测CCLE细胞系中对Topo I抑制剂的敏感性,并且预测在辅助化疗之后肺癌患者的结果和乳腺癌患者对放疗的响应。(图4A)平均IC₅₀值的柱状图显示在最高CES四分位数中的CCLE(癌细胞系百科全书)细胞系比在最低CES四分位数中的CCLE细胞系对Topo I抑制剂拓扑替康(左)和伊立替康(右)显著地更敏感。显示了双尾学生t检验的p值。误差棒指示标准误差。(图4B-图4C)辅助化疗显著降低了与早期阶段NSCLC的高CES相关的风险。将JBR.10临床试验早期阶段NSCLC患者分为高CES(最高三分位数)组和低CES(较低的两个三分位数)组(图4B)。左边的卡普兰-迈耶曲线是高CES组,其显示与无ACT相比(OBS),在辅助化疗(ACT)之后,患者总存活率(OS)显著改善。图4B右侧的图显示,与OBS相比,ACT未显著改善低CES组的总存活率。(图4C)荟萃分析梳理了JBR.10和UT SPORE早期NSCLC患者。将患者分为高CES(最高三分位数)组和低CES(较低的两个三分位数)组。在图

4C中的卡普兰-迈耶曲线,左侧的是高CES组,其显示与无ACT相比(OBS),在辅助化疗(ACT)之后,患者的总存活率显著改善。图4C右侧的图显示,与OBS相比,ACT未改善低CES组的总存活率。(图4D)放疗降低了与乳腺癌患者的高CES相关的风险。将乳腺癌患者分为高CES组、中间CES组或低CES组。图4D是卡普兰-迈耶生存曲线,左侧是总存活率(OS),以及在图4D中,右侧是无病存活率(DFS)。标明各图的患者数量、对数秩p值和治疗信息。对于无疾病存活率和总存活率,与无RT相比,高CES值与RT之后显著更高的存活率相关(图4D,顶行)。对于具有中间CES值的患者,与RT相关的益处不是统计学显著的(图4D,中间行)。对于无疾病存活率和总存活率,在有或没有RT的低CES患者之间,未检测到显著益处(图4D,底行)。

[0030] 图5A和5B提供了在最高(75-100%)CES四分位数和最低(0-25%)CES四分位数中的CCLE细胞系的伊立替康(图5A)和拓扑替康(图5B)IC₅₀的百分比频率柱状图,证明了CES和药物IC₅₀之间负相关。

[0031] 图6A和图6B提供了生存分析,显示ACT特异性地改善了JBR.10试验的高CES NSCLC患者的OS。高CES与无辅助治疗(OBS)的患者的不良OS相关(图6A)。高CES与化疗患者(ACT)的不良OS无关(图6B)。

[0032] 图7A-图7E提供了森林图,其显示对于数种癌症类型,高CES与对辅助疗法的敏感性相关。正方形表示HR,以及棒的左右限度指示95%CI。图7A显示,通过荟萃分析,辅助化疗(黑色条)否定了与第I阶段ADC和NSCLC(具有星号的黑色条)的高CES相关的较差OS。另外,辅助化疗和放疗(RT)(黑色条)改善了与NSCLC的高CES相关的较差OS(具有星号的黑色条)。辅助化疗和放疗降低了NSCLC中与高CES相关的第一进展(FP)的风险(图7B)。与未治疗或仅用他莫昔芬进行治疗相比,使用或不用激素疗法的辅助化疗否定了与ER+乳腺癌患者的高CES相关的较差RFS(图7C)。单独的他莫昔芬不能改善与高CES ER+乳腺癌患者相关的不良RFS。包括5-FU的辅助化疗未显著改善与高CES胃癌患者相关的不良OS(图7D)。使用铂或拓扑替康的辅助治疗提高了高CES晚期卵巢癌的总存活率,并且在降低风险方面拓扑替康优于铂(图7E)。

[0033] 图8A-图8D提供了在有或无RT的情况下,具有高CES值和低CES值的乳腺癌患者存活的卡普兰-迈耶生存曲线。DFS=无病存活率,以及OS=总存活率。在无RT的情况下,高CES值分别与不良DFS和OS相关($p < 0.05$) (图8A和图8C)。在RT之后,分别对于DFS和OS,具有高CES值的患者显示与具有低CES值的患者无统计学区别的风险比($p > 0.05$) (图8B和8D)。

[0034] 图9显示了可用于根据本公开的实施方案的系统和方法的示例性计算机系统100的框图。

[0035] 发明详述

[0036] I. 引言

[0037] 本文描述了用于预测对象中疾病再发生、疾病复发、疾病进展和/或癌症转移的风险的方法。该方法也包括预测患有癌症的对象响应癌症治疗(如放疗和/或化疗)的可能性。该方法包括:确定14个CEN/KT基因的表达水平(如RNA表达水平或蛋白表达水平),将所述水平转化为着丝粒和动粒基因表达评分(CES),以及基于与CES值的参考范围相比,具有高CES值来确定对象具有不良预后的可能性。在一些情况下,高CES值表示在CES值参考范围的最高三分位数中的任何CES值。高CES值也可以预测在接受辅助癌症治疗时存活的可能性。

[0038] II. 定义

[0039] 如本文所使用的,除非另有说明,以下术语具有归于其自身的含义。

[0040] 本文使用的术语“一个/一种(a)”、“一个/一种(an)个或“所述(the)”不仅包括含有一个/一种成员的方面,也包括含有多个/多种成员的方面。例如,除非上下文另外明确规定,否则单数形式“一个/一种(a)”、“一个/一种(an),或“所述(the)”包括复数指示物。因此,例如,提及“细胞”,其包括多个此类细胞,以及提及“试剂,其包括提及本领域技术人员已知的一种或多种试剂等。

[0041] 术语“癌症”是指涉及细胞异常生长的任何疾病,并且包括影响体内任何组织、器官或细胞的疾病的所有阶段和所有形式。该术语包括所有已知的癌症和肿瘤病况,无论其特征为恶性肿瘤、良性瘤、软组织瘤或实体瘤以及包括转移前和转移后癌症的所有阶段和等级的癌症。不同类型的癌症的实例包括但不限于乳腺癌、肺癌(例如非小细胞肺癌);消化道和胃肠道癌症,例如结肠直肠癌、胃肠道间质瘤、胃肠道类癌瘤、结肠癌、直肠癌、肛门癌、胆管癌、小肠癌和胃(stomach)(胃(gastric))癌;食道癌;胆囊癌;肝癌;胰腺癌;阑尾癌;卵巢癌;前列腺癌;肾癌(例如肾细胞癌);中枢神经系统癌症;皮肤癌;淋巴瘤;绒毛膜癌;头颈癌;成骨肉瘤以及血癌。如本文所使用的,“肿瘤”包含一个或多个癌细胞。

[0042] 术语“肿瘤样本”包括包含一个或多个肿瘤细胞的生物样本或者来自生物来源的样本。生物样本包括来自体液(例如血液、血浆、血清或尿液)的样本,或者例如通过活检,来源于细胞、组织或器官,优选疑似包含癌细胞或者基本由癌细胞组成的肿瘤组织的样本。

[0043] 术语“测定(determining)”、“评估(assessing)”、“分析(assaying)”、“测量(measuring)”和“检测(detected)”可以互换使用,并且涉及定量测定和半定量测定两者。在意图定量测定或半定量测定时,可以使用短语“测定目标多核苷酸或目标多肽的水平”,或者“检测目标多核苷酸或目标多肽的水平”。

[0044] 术语“数量”或“水平”是指样本中存在的目标多核苷酸或目标多肽的量。此类量可以以绝对值,即样本中多核苷酸或多肽的总量表示,或者以相对值,即样本中多核苷酸或多肽的浓度表示。

[0045] 本文所使用的术语“核酸”或“多核苷酸”指单链或双链形式的脱氧核糖核苷酸或核糖核苷酸。该术语包括含有天然核苷酸的已知类似物的核酸,出于期望的目的,其具有与参考核酸相似或改善的结合性质。出于期望的目的,该术语也包括以与天然存在的核苷酸类似的方式或者以提高的速率代谢的核酸。该术语也包括具有合成的骨架的核酸样结构。本发明提供的DNA骨架类似物包括磷酸二酯、硫代磷酸酯、二硫代磷酸酯、甲基膦酸酯、氨基磷酸酯、烷基磷酸三酯、氨基磺酸酯、3'-硫缩醛、亚甲基(甲基亚氨基)、3'-N-氨基甲酸酯、吗啉代氨基甲酸酯以及肽核酸(PNA);参见由F.Eckstein编辑的Oligonucleotides and Analogues, a Practical Approach, IRL Press at Oxford University Press(1991); Antisense Strategies, Annals of the New York Academy of Sciences, 第600卷, Eds. Baserga and Denhardt (NYAS 1992); Milligan(1993) J. Med. Chem. 36:1923-1937; Antisense Research and Applications(1993, CRC Press)。PNA包含非离子骨架,如N-(2-氨基乙基)甘氨酸单元。硫代磷酸酯键描述于WO 97/03211、WO 96/39154; Mata(1997) Toxicol. Appl. Pharmacol. 144:189-197中。包含于该术语中的其它合成骨架包括甲基-膦酸酯键或者交替的甲基膦酸酯键和磷酸二酯键(Strauss-Soukup(1997) Biochemistry 36: 8692-8698)以及苄基膦酸酯键(Samstag(1996) Antisense Nucleic Acid Drug Dev 6:

153-156)。

[0046] 术语“蛋白质”、“肽”或“多肽”在本文可以互换使用,是指氨基酸残基的聚合物。该术语适用于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工化学模拟物的氨基酸聚合物,也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。如本文所使用的,该术语包含任何长度的氨基酸链,包括全长蛋白(即抗原),其中氨基酸残基通过共价肽键连接。

[0047] 术语“基因产物”或“基因表达产物”是指由基因编码的RNA或蛋白质。

[0048] 术语“杂交”是指在严格的条件下,核酸分子优先与特定核苷酸序列结合、双链化或杂交。术语“严格的条件”是指这样的条件,在该条件下,在混合物(例如,来自组织活检的细胞裂解物或DNA制备物)中,探针将优先与其靶子序列杂交,并且在较小程度上或完全不与其它序列杂交。在核酸杂交的背景下(例如,在阵列、微阵列、Southern或Northern杂交中),“严格杂交”和“严格杂交洗涤条件”是序列依赖的,并且在不同的环境参数下是不同的。核酸杂交的广泛指导可见于例如Tijssen(1993) *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes*,第I部分第2章,“Overview of principles of hybridization and the strategy of nucleic acid probe assays”Elsevier,N.Y.通常,选择的高度严格的杂交和洗涤条件低于限定的离子强度和pH下特定序列的热熔点(T_m)约5°C。 T_m 是50%的靶序列与完全匹配的探针杂交的温度(在限定的离子强度和pH下)。选择非常严格的条件等于具体探针的 T_m 。使用标准杂交溶液,用于阵列上或者Southern或northern印迹中的过滤器上具有超过100个互补残基的互补核酸杂交的严格杂交条件的实例是42°C(参见,例如,Sambrook和Russell(2001) *分子克隆实验指南*(第三版)第1-3卷,Cold Spring Harbor Laboratory,Cold Spring Harbor Press,NY)。高度严格的洗涤条件的实例是在72°C下,0.15M NaCl,持续约15分钟。严格的洗涤条件的实例是在65°C下,在0.2×的实例中洗涤15分钟。通常,在高度严格的洗涤之前进行低度严格洗涤,以移除背景探针信号。对于例如超过100个核苷酸的双链,中等严格洗涤的实例是在45°C下,1格洗格洗,持续15分钟。对于例如超过100个核苷酸的双链,低度严格洗涤的实例是在40°C下,4x SSC至6x SSC,持续15分钟。

[0049] 术语“互补性”指核酸通过常规的沃森-克里克或其它非常规类型与另一核酸序列形成氢键的能力。互补百分比指核酸分子中可与第二核酸序列形成氢键(例如,沃森-克里克碱基配对)的残基的百分比(例如10个中的5个、6个、7个、8个、9个、10个,是50%、60%、70%、80%、90%和100%互补)。“完全互补”意为核酸序列的所有连续残基将与第二核酸序列中相同数目的连续残基氢键键合。如本文所使用的“基本上互补”是指在8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、30个、35个、40个、45个、50个或更多个核苷酸区域内至少60%、65%、70%、75%、80%、85%、90%、95%、97%、98%、99%或100%互补的程度,或者指在严格条件下杂交的两条核酸。

[0050] 如本文所使用的术语“标准对照”是指存在于已建立的正常组织样本中的预先确定量或浓度的多核苷酸序列或多肽序列,所述正常组织样本例如健康的非癌组织样本或者二倍体、非转化的、非癌的、基因组稳定的健康人细胞系。标准对照值适合于本发明的方法,以用作比较测试样本中存在的特定mRNA或蛋白质的量的基础。用作标准对照的已建立的样本提供在正常组织样本中典型的特定mRNA或蛋白质的平均量。标准对照值可以根据样本的

性质以及其他因素(例如,建立这样的对照值所基于的对象性别、年龄、种族)而变化。

[0051] 术语“治疗(treatment)”、“治疗(treat)”或“治疗(treating)”是指减少癌症(例如乳腺癌、肺癌、卵巢癌等)或癌症症状的影响的方法。因此,在所公开的方法中,治疗可以指癌症或癌症症状的严重性10%、20%、30%、40%、50%、60%、70%、80%、90%或100%的降低。例如,如果与对照相比,对象中疾病的一种或多种症状减少10%,则治疗疾病的方法被认为是治疗。因此,减少可以是与天然水平或对照水平相比,10%、20%、30%、40%、50%、60%、70%、80%、90%、100%或者10%至100%的任何百分比的减少。应当理解,治疗不一定是指疾病、病况或者疾病或病况的症状的治愈或完全消除。

[0052] 在疾病治疗的背景下,术语“推荐”或“建议”是指建议或推荐特别适用于患者的治疗干预(例如,药物治疗、辅助疗法等)和/或疾病管理。

[0053] 在患者响应癌症疗法的背景下,术语“响应的”、“临床应答”、“阳性临床应答”等可互换使用,并且是指与不利应答(即不良事件)相反的针对治疗的有利患者应答。在患者中,可以以许多临床参数来表示有益应答,包括可检测的肿瘤的丧失(完全应答,CR),肿瘤大小和/或癌细胞数目的减少(部分应答,PR),肿瘤生长停滞(稳定的疾病,SD),抗肿瘤免疫应答增强,可能导致肿瘤的消退或排斥;与肿瘤相关的一种或多种症状在一定程度上的缓解;治疗后的生存期的长度增加;和/或在治疗之后给定的时间点,死亡率降低。肿瘤大小和/或癌细胞数目和/或肿瘤转移的持续增加指示缺乏对治疗的有益应答。在群体中,可以基于一个或多个端点来评估药物的临床益处,即其功效。例如,总缓解率(overall response rate)(ORR)的分析将那些在用药物治疗后经历CR或PR的患者归类为应答者。疾病控制(DC)的分析将那些在用药物治疗后经历CR、PR或SD的患者归类为应答者。

[0054] 可以使用任何端点来评估指示对患者的益处的阳性临床应答,包括但不限于:(1)在一定程度上抑制肿瘤生长,包括减缓和完全生长停滞;(2)减少肿瘤细胞的数目;(3)减小肿瘤大小;(4)抑制(即,减少、减缓或完全停止)肿瘤细胞至相邻外周器官和/或组织的浸润;(5)抑制转移;(6)增强抗肿瘤免疫应答,可能导致肿瘤消退或排斥;(7)在一定程度上缓解与肿瘤相关的一种或多种症状;(8)增加治疗后生存期的长度;和/或(9)降低在治疗后给定时间点的死亡率。也可以以临床结果的各种度量来表示阳性临床应答。阳性临床结果在个体结果的背景下也可以被认为相对于患者群体的结果具有可比较的临床诊断,并且可以使用各种端点进行评估,如无复发间隔持续时间(RFI)的增加,与群体中的总体生存期(OS)相比生存时间的增加,无病生存期(DFS)时间的增加,远端无复发间隔(DRFI)持续时间的增加等。其他端点包括无任何事件(AE)生存期的可能性、无转移性复发(MR)生存期(MRFS)的可能性、无病生存期(DFS)的可能性、无复发生存期(RFS)的可能性、第一进展(FP)的可能性以及无远端转移生存期(DMFS)的可能性。阳性临床应答的可能性的增加与癌症再发生或复发的可能性的降低相对应。

[0055] 术语“良好预后”是指预测临床应答、疾病特异性生存期、总体生存期或无病生存期的可能性,包括癌细胞增殖和/或转移的部分缓解、完全缓解和阻抑。对于患有实体瘤癌症的患者,良好预后包括在疾病缓解或肿瘤缩小方面的阳性应答率,或者评估减少的肿瘤负荷或生长的任何其他形式。可以将良好预后测量为生存期的长度(时间)。

[0056] 术语“不良预后”是指预测可归因于癌症的死亡或进展(包括再发生、复发和/或转移性扩散)的可能性。

[0057] 术语“总体生存期”是指从诊断或治疗开始的时间开始,患者仍然活着的时间间隔。

[0058] 术语“无进展生存期”是指从患者治疗开始直到癌症进展或患者死亡(以先发生者为准)的时间间隔。

[0059] 术语“应答者”或“应答”是指患有癌症并且在用癌症疗法进行治疗之后显示出有益的临床应答的患者。

[0060] 术语“无应答者”或“无应答”是指患有癌症并且在用癌症疗法进行治疗之后未显示出有益的临床应答的患者。

[0061] 术语“对象”旨在包括动物。对象的实例包括哺乳动物,例如人、狗、牛、马、猪、绵羊、山羊、猫、小鼠、兔、大鼠和转基因非人动物。在某些实施方案中,对象是人,例如罹患癌症、处于罹患癌症的风险或者可能会罹患癌症的人。

[0062] 术语“风险评分”是指可以为医生和护理者提供有价值的诊断和预后洞察力的统计学导出的数值。在一些情况下,该评分提供疾病再发生/复发或疾病进展的预计风险,疾病进展的预计速速率和/或对具体治疗的预计应答。可以将个体评分与参考评分或参考评分标度进行比较,以确定疾病再发生/复发的风险或者以协助选择治疗干预或疾病管理方法。

[0063] 术语“高CES评分”、“高CES值”、“高风险评分”是指这样的着丝粒和动粒表达评分(CES),其具有在具体对象群体的CES值范围的最高百分位数范围(例如最高三分位数(例如,最高33%))内的数值。在一些实施方案中,具有高CES评分的癌症患者也处于不良预后或阴性疾病结果的高风险中。术语“中间CES评分”或“中等CES评分”是指在范围的中间三分位数中的CES值,以及术语“低CES评分”是指在范围的较低三分位数中的CES值。

[0064] III. 实施方案详述

[0065] A. CEN/KT基因和基因产物

[0066] 本文所述的方法部分基于下述令人惊讶的发现:14个着丝粒和动粒结构蛋白基因(CEN/KT基因)在不同癌症中过表达并且预示特定癌症阶段的疾病进展。本文所述的着丝粒和动粒基因表达评分(CES)可以将14种基因的表达水平或者14种基因中至少9种基因的表达水平转化为预测评分。CES评分也预示治疗应答和治疗后的生存期。

[0067] 本文所述的方法可用于评估患有癌症的患者。癌症的非限制性实例包括:乳腺癌,肺癌(例如非小细胞肺癌);前列腺癌,卵巢癌,消化道和胃肠道癌症,例如结肠直肠癌、胃肠道间质瘤、胃肠道类癌瘤、结肠癌、直肠癌、肛门癌、胆管癌、小肠癌和胃(stomach)(胃(gastric))癌;食道癌;胆囊癌;肝癌;胰腺癌;阑尾癌;膀胱癌;前列腺癌;宫颈癌;子宫癌;肾癌(例如肾细胞癌);中枢神经系统癌症,例如较低级脑癌;皮肤癌;淋巴瘤;头颈癌;肾上腺皮质癌以及甲状腺癌。癌症可以是非常早期的癌症、早期癌症、晚期癌症或转移癌。在一些实施方案中,癌症是非常早期的癌症或早期癌症。在一些实施方案中,乳腺癌是诸如基底样乳腺癌、HER2阳性(HER2⁺)乳腺癌、管腔B型乳腺癌、管腔A型乳腺癌和正常样乳腺癌的乳腺癌亚型。乳腺癌可以包括雌激素受体阳性(ER⁺)肿瘤或雌激素受体阴性(ER⁻)肿瘤。在一些实施方案中,乳腺癌是乳腺导管内原位癌。在一些实施方案中,癌症是早期卵巢癌(I期或II期)或早期(例如I期)肺癌。

[0068] 本公开涉及测量患有肿瘤的患者中生物标志物组的基因产物,例如包括CENP-A、

HJURP、MIS18B、CENP-N、CENP-M、CENP-U、ZWINT、NDC80和SPC25的至少9种基因的一组人CEN/KT基因,或包括CENP-A、CENP-K、CENP-L、CENP-M、CENP-N、CENP-U、CENP-W、MIS18B、HJURP、ZWINT、NDC80、SPC24、SPC25和NUF2 14个人CEN/KT基因的一组基因。在一些情况下,在患者经历诸如手术、化疗、放疗、药物治疗、免疫疗法等的治疗干预之前进行测量。

[0069] 人组蛋白H3样着丝粒蛋白A (CENP-A) 多肽序列示于例如Genbank登录号第AAH02703.1号和NCBI参考序列号NP_001035891和NP_001800。人CENP-A mRNA (编码) 序列示于例如NCBI参考序列号NM_001042426和NM_001809。

[0070] 人着丝粒蛋白K (CENP-K) 多肽序列示于例如NCBI参考序列号NP_001253967和NP_071428。人CENP-K mRNA (编码) 序列示于例如NCBI参考序列号NM_00126703和NM_022145。

[0071] 人着丝粒蛋白L (CENP-L) 多肽序列示于例如NCBI参考序列号NP_001120653、NP_001164653和NP_201576。人CENP-L mRNA (编码) 序列示于例如NCBI参考序列号NM_001127181、NM_001171182和NM_033319。

[0072] 人着丝粒蛋白M (CENP-M) 多肽序列示于例如NCBI参考序列号NP_001002876、NP_001103685和NP_076958。人CENP-M mRNA (编码) 序列示于例如NCBI参考序列号NM_001002876、NM_001110215和NM_024053。

[0073] 人着丝粒蛋白U (CENP-U) 多肽序列示于例如NCBI参考序列号NP_078905。人CENP-U mRNA (编码) 序列示于例如NCBI参考序列号NM_024629。

[0074] 人着丝粒蛋白W (CENP-W) 多肽序列示于例如NCBI参考序列号NP_0001012525、NP_001273453和NP_001273454。人CENP-W mRNA (编码) 序列示于例如NCBI参考序列号NM_001012507、NM_001286524和NM_001286525。

[0075] 人Mis18- β (MIS18B) 多肽序列示于例如NCBI参考序列号NP_009211。人MIS18B mRNA (编码) 序列示于例如NCBI参考序列号NM_007280。MIS18B多肽也被称为癌/睾丸抗原86、CT86、Opa相互作用蛋白5和OIP-5。

[0076] 人霍利迪连接体识别 (HJURP) 多肽序列示于例如NCBI参考序列号NP_001269891、NP_001269892和NP_060880。人HJURP mRNA (编码) 序列示于例如NCBI参考序列号NM_001282962、NM_001282963和NM_018410。

[0077] 人霍利迪连接体识别蛋白 (HJURP) 多肽序列示于例如NCBI参考序列号NP_001269891、NP_001269892和NP_060880。人HJURP mRNA (编码) 序列示于例如NCBI参考序列号NM_001282962、NM_001282963和NM_018410。

[0078] 人ZW10相互作用蛋白 (interactor) (ZWINT) 多肽序列示于例如NCBI参考序列号NP_001005413、NP_008988和NP_127490。人ZWINT mRNA (编码) 序列示于例如NCBI参考序列号NM_001005413、NM_007057和NM_032997。

[0079] 人动粒蛋白NDC80同源物 (NDC80) 多肽序列示于例如NCBI参考序列号NP_006092。人NDC80 mRNA (编码) 序列示于例如NCBI参考序列号NM_006101。NDC80多肽也被称为动粒蛋白Hec1、HsHec1、动粒相关蛋白2、KNTC2和视网膜母细胞瘤相关蛋白HEC。

[0080] 人动粒蛋白Spc24 (SPC24) 多肽序列示于例如NCBI参考序列号NP_872319和XP_005259810。人SPC24 (编码) 序列示于例如NCBI参考序列号NM_182513和XM_005259753。

[0081] 人动粒蛋白Spc25 (SPC25) 多肽序列示于例如NCBI参考序列号NP_065726。人SPC25 (编码) 序列示于例如NCBI参考序列号NM_020675。

[0082] 人动粒蛋白Nuf2同源物 (NUF2) 多肽序列示于例如NCBI参考序列号NP_113611和NP_663735。人NUF2 (编码) 序列示于例如NCBI参考序列号NM_031423和NM_145697。

[0083] 至少9个CEN/KT生物标志物或14个CEN/KT生物标志物在本公开的用于诊断或预测患者中的癌症的方法以及用于通过选择疗法而使治疗管理个性化的方法中特别有用。因此, 在一些实施方案中, 实施14个CEN/KT基因产物的表达水平 (例如RNA表达水平或蛋白表达水平) 的检测或测定。

[0084] B. 定量RNA水平的方法

[0085] 本公开的方法包括检测获自疑似患有癌症或处于患有癌症的风险的患者的肿瘤样本中14个CEN/KT基因或者14个CEN/KT基因的子集 (例如, 14种基因的至少9种基因的子集) 的RNA表达水平。在一些实施方案中, 患者已被诊断为患有癌症。

[0086] 肿瘤样本可以是包含癌细胞的生物样本。在一些实施方案中, 肿瘤样本是例如通过肿瘤活检或细针抽吸获自肿瘤的新鲜或保存的样本。样本也可以是含有癌细胞的任何生物流体。肿瘤样本分离自或获自任何数目的原发性肿瘤, 包括但不限于乳腺、肺、前列腺、脑、肝、肾、肠、结肠、脾、胰腺、胸腺、睾丸、卵巢、子宫等的肿瘤。在一些实施方案中, 肿瘤样本来自肿瘤细胞系。依照医院或诊所通常遵循的标准协议进行来自对象的肿瘤样本的采集, 例如在活检期间)。

[0087] 可以使用本领域普通技术人员已知的任何方法来检测RNA表达水平。在一些实施方案中, 从肿瘤样本中分离RNA。可以使用多种方法从肿瘤样本中分离RNA。用于从组织或细胞中提取RNA的标准方法描述于例如Ausubel等人, Current Protocols of Molecular Biology, John Wiley&Sons, 1997以及Sambrook和Russell, 分子克隆实验指南, 第三版, 2001中。也可以使用商购的试剂盒来分离RNA, 例如, RNeasy[®] 微型柱 (Qiagen)、PureLink[®] RNA微型试剂盒 (Thermo Fisher Scientific) 等。

[0088] 可以通过多种方法检测或测量上述14个CEN/KT基因的RNA (例如mRNA) 表达水平, 所述方法包括但不限于扩增分析、杂交试验、测序分析或阵列。此类方法的非限制性实例包括逆转录聚合酶链反应 (RT-PCR); 定量实时PCR (qRT-PCR); 定量PCR, 如 TaqMan[®]; Northern印迹; 原位杂交分析; 微阵列分析, 例如, 来自NanoString Technologies的微阵列; 基于多重杂交的分析, 例如来自Panomics的QuantiGene 2.0 Multiplex Assay); 基因表达系列分析 (SAGE); cDNA介导的退火、选择、延伸和连接; 直接测序或焦磷酸测序; 大规模平行测序; 下一代测序; 高效液相色谱 (HPLC) 片段分析; 毛细管电泳等。

[0089] 可以使用涉及扩增反应和/或将探针与固体支持物连接并用于定量RNA的响应的各种方法。可选地, 可以将RNA连接到固体支持物上, 并使用针对目的序列的探针进行定量。

[0090] 在一些实施方案中, 首先将靶RNA逆转录, 并定量所得的cDNA。在一些实施方案中, 使用RT-PCR或其他定量扩增技术定量靶RNA。使用PCR扩增cDNA众所周知 (参见美国专利第4,683,195号和第4,683,202号; PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS (Innis等人编辑, 1990))。定量扩增的方法公开于, 例如美国专利第6,180,349号、第6,033,854号和第5,972,602号, 以及公开于例如Gibson等人, Genome Research 6:995-1001 (1996); DeGraves等人 Biotechniques 34 (1):106-10, 112-5 (2003); Deiman B等人, Mol Biotechnol. 20 (2):163-79 (2002)。用于测定样本中目标mRNA水平的可选择的方法可以包

括其他核酸扩增方法,如连接酶链式反应(Barany (1991) Proc. Natl. Acad. Sci. USA 88: 189-193)、自主序列复制(Guatelli等人(1990) Proc. Natl. Acad. Sci. USA 87:1874-1878)、转录扩增系统(Kwoh等人(1989) Proc. Natl. Acad. Sci. USA 86:1173-1177)、Q- β 复制酶(Lizardi等人(1988) Bio/Technology 6:1197)、滚环复制(美国专利第5,854,033号)或者任何其它核酸扩增方法,随后使用本领域技术人员熟知的技术检测扩增的分子。

[0091] 通常,定量扩增基于扩增(例如PCR)反应的循环中监测代表模板拷贝的信号(例如,探针的荧光)。检测扩增产物的一种方法是5' -3' 外切核酸酶“水解”PCR分析(也称为 TaqMan[®]分析)(美国专利第5,210,015号和第5,487,972号;Holland等人,PNAS USA 88: 7276-7280 (1991);Lee等人,Nucleic Acids Res.21:3761-3766 (1993))。这个分析通过在扩增反应期间双标记荧光探针(“TaqMan[®]”探针)的杂交和切割来检测特定PCR产物的积累。荧光探针由用荧光报告染料和猝灭染料两者标记的寡核苷酸组成。在PCR期间,当且仅当探针与被扩增的区段杂交时,该探针被DNA聚合酶的5' -外切核酸酶活性切割。探针的切割产生报告染料荧光强度的增加。

[0092] 依赖于使用能量转移的检测扩增产物的另一种方法是“信标探针”方法,其描述于Tyagi和Kramer,Nature Biotech 14:303-309 (1996),也是美国专利第5,119,801号和第5,312,728号的主体。该方法使用可形成发夹结构的寡核苷酸杂交探针。在杂交探针的一端(5' 端或3' 端),存在供体荧光团,而在另一端存在受体部分。在Tyagi和Kramer方法的情况下,该受体部分是猝灭剂,即,受体吸收由供体释放的能量,然而其本身不发荧光。因此,当信标处于开放构象时,供体荧光团的荧光是可检测的,而当信标处于发夹(闭合)构象时,供体荧光团的荧光淬灭。当用于PCR时,与PCR产物的一条链杂交的分子信标探针处于“开放构象”,并且检测到荧光,而仍未杂交的那些分子信标探针不发荧光(Tyagi和Kramer,Nature Biotechnol.14:303-306 (1996))。因此,荧光的量将随着PCR产物的量的增加而增加,并且因此可以用作PCR进程的量度。本领域技术人员将认识到,其他定量扩增方法也是可用的。

[0093] 用于进行核酸定量扩增的各种其它技术也是已知的。例如,一些方法使用一种或多种结构化的探针寡核苷酸,以在当所述寡核苷酸与靶核酸杂交时产生荧光的变化。例如,一种此类方法涉及利用荧光共振能量转移(FRET)的双荧光团方法,例如LightCycler[™]杂交探针,其中两个寡探针退火成扩增子。寡核苷酸被设计成以头至尾的方向与与有效能量转移相容的距离分开的荧光团杂交。结构化以当与核酸结合或掺入延伸产物中时发射信号的标记的寡核苷酸的其他实例包括:Scorpions[™]探针(例如,Whitcombe等人,Nature Biotechnology 17:804-807,1999和美国专利第6,326,145号)、Sunrise[™](或Amplifluor[™])探针(例如,Nazarenko等人,Nuc. Acids Res.25:2516-2521,1997和美国专利第6,117,635号)以及形成二级结构的探针,所述二级结构探针导致在没有淬灭剂的情况下信号降低,并且当与靶标杂交时发射增加的信号(例如,Lux探针[™])。

[0094] 在其他实施方案中,可以使用当插入双链DNA中时产生信号的嵌入剂。示例性试剂包括SYBR GREEN[™]和SYBR GOLD[™]。由于这些试剂不是模板特异性的,因此假设基于模板特异性扩增产生信号。这可以通过监测随温度变化的信号来确认,因为模板序列的熔点通常显著高于例如引物二聚体等。

[0095] 在其他实施方案中,将mRNA固定在固体表面,并与例如斑点印迹或Northern格式

中的探针接触。在可选实施方案中,将探针固定在固体表面,并使mRNA与例如在基因芯片阵列中的探针接触。本领域技术人员可以容易地改变已知的mRNA检测方法,以用于检测编码生物标志物或其他目的蛋白的mRNA水平。

[0096] 在一些实施方案中,例如使用微阵列。DNA微阵列提供了一种同时测量大量基因的表达水平的方法。每个阵列由附着于固体支持物的可复制模式的捕获探针组成。标记的RNA或DNA与阵列上的互补探针杂交,然后通过激光扫描对其进行检测。测定阵列上每个探针的杂交强度,并转化为表示相对基因表达水平的数值。参见美国专利第6,040,138号、第5,800,992号和第6,020,135号、6,033,860号以及第6,344,316号。高密度寡核苷酸阵列对测定样本中大量RNA的基因表达谱特别有用。

[0097] 使用机械合成方法合成这些阵列的技术描述于例如美国专利第5,384,261号中。尽管经常使用平面阵列表面,但是可将阵列制造在实际上任何形状的表面或者甚至多个表面上。阵列可以是珠子、凝胶、聚合物表面、纤维(如光纤)、玻璃或任何其它合适的基底上的肽或核酸,参见美国专利第5,770,358号、第5,789,162号、第5,708,153号、第6,040,193号和第5,800,992号。可以以允许诊断或者全包式设备的其它操纵的方式来包装阵列。

[0098] 在一些实施方案中,在杂交分析中使用基因特异性探针和/或引物以检测RNA表达。可以用任何可检测的部分或化合物(如放射性同位素、荧光团、化学发光剂和酶)标记探针和/或引物。

[0099] 可以使用众所周知的技术合成和标记实施本发明所必需的探针和引物。如Needham-Van Devanter等人,Nucleic Acids Res.12:6159-6168,1984中所述,可以根据由Beaucage和Caruthers,Tetrahedron Letts.,22:1859-1862,1981首次描述的固相亚磷酸三酯法,使用自动合成仪化学合成用作探针和引物的寡核苷酸。

[0100] 在一些实施方案中,方法还包括检测可用作对照以确定表达水平的一种或多种参考基因的表达水平。这样的基因通常以高水平组成型表达,并且可以作为参照以确定准确的基因表达水平估计。对照基因的非限制性实例包括ARPC2、ATF4、ATP5B、B2M、CDH4、CELF1、CLTA、CLTC、COPB1、CTBP1、CYC1、CYFIP1、DAZAP2、DHX15、DITM1、EEF1A1、FLOT2、GAPDH、GUSB、HADHA、HDLBP、HMBS、HNRNPC、HPRT1、HSP90AB1、MTCH1、MYL12B、NACA、NDUFB8、PGK1、PPIA、PPIB、PTBP1、RPL13A、RPLP0、RPS13、RPS23、RPS3、S100A6、SDHA、SEC31A、SET、SF3B1、SF3B3、SNRNP200、STARD7、SUMO1、TBP、TFRC、TMBIM6、TPT1、TRA2B、TUBA1C、UBB、UBC、UBE2D2、UBE2D3、VAMP3、XP01、YTHDC1、YWHAZ和18S rRNA基因。因此,确定目的基因的RNA表达水平,例如一组14个CEN/KT基因的基因表达水平,还可包括测定上文公开的一种或多种参考基因的表达水平。

[0101] 可以将14个CEN/KT基因中的每一个的mRNA表达水平标准化为对照基因的参考水平。对照值可以是预先确定的、同时确定的或者在从对象获得样本之后确定的。标准可以在同一分析中运行,或者可以是来自先前分析的已知标准。可以将CEN/KT基因mRNA表达的标准化水平转化为评分,例如CES评分。

[0102] C. 定量蛋白水平的方法

[0103] 在一些实施方案中,本文公开的方法包括测定由14个CEN/KT基因编码的14条多肽的水平或者由14个CEN/KT基因编码的14条多肽中至少9条多肽的子集的水平,所述14个CEN/KT基因如CENP-A、CENP-K、CENP-L、CENP-M、CENP-N、CENP-U、CENP-W、MIS18B、HJURP、

ZWINT、NDC80、SPC24、SPC25和NUF2。在一些实施方案中,检测或测量CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-U、ZWINT、NDC80和SPC25多肽的水平。可以在各种肿瘤样本中检测这些多肽。

[0104] 可以使用本领域普通技术人员已知的任何方法检测蛋白表达水平。可应用技术的一般概述可见于Harlow&Lane, *Antibodies: A Laboratory Manual* (1988) 和Harlow&Lane, *Using Antibodies* (1999) 中找到。产生与等位基因变体特异性反应的多克隆和单克隆抗体的方法是本领域技术人员已知的(参见例如Coligan, *Current Protocols in Immunology* (1991); Harlow&Lane, 同上; Goding, *Monoclonal Antibodies: Principles and Practice* (第2版, 1986); 以及Kohler&Milstein, *Nature* 256:495-497 (1975))。此类技术包括通过从噬菌体或类似载体中的重组抗体文库选择抗体进行的抗体制备, 以及通过免疫兔或小鼠进行的多克隆和单克隆抗体制备(参见例如Huse等人, *Science* 246:1275-1281 (1989); Ward等人, *Nature* 341:544-546 (1989))。

[0105] 可以通过多种方法检测此类多肽的水平, 其包括但不限于蛋白质印迹, 免疫分析例如酶联免疫吸附分析(ELISA)、酶免疫分析(EIA)、放射免疫分析(RIA)、夹心分析、竞争分析、免疫组织化学, 质谱, 2-D凝胶电泳, 蛋白质阵列, 抗体阵列等。

[0106] 对于免疫学和免疫分析程序的综述, 参见Basic and Clinical Immunology (Stites&Terr eds., 第7版, 1991)。此外, 可以以几种配置中的任一种进行免疫分析, 其在Enzyme Immunoassay (Maggio, ed., 1980) 中以及Harlow&Lane, 同上中进行了广泛综述。对于一般免疫分析的综述, 也参见Methods in Cell Biology: Antibodies in Cell Biology, 第37卷(Asai, ed. 1993); Basic and Clinical Immunology (Stites&Terr, eds., 第7版, 1991)。

[0107] 与将CEN/KT基因的mRNA水平标准化相似, 也可以将蛋白表达水平与标准对照组进行比较, 并将其标准化为标准对照值。

[0108] D. 建立CES评分

[0109] 在确定一组CEN/KT基因的表达水平之后, 本文提供的方法包括基于14个CEN/KT基因表达产物的水平计算CES评分, 例如风险评分。在CES评分中, 14个CEN/KT基因的表达水平可以相等地加权。在一些情况下, 用预定义的系数对每个基因的表达水平进行加权。对于基因, 预定义的系数相同或不同。

[0110] 在一些实施方案中, 使用以下公式生成CES评分:

[0111] $\text{CES评分} = 14 \text{个CEN/KT基因 (例如CENP-A、CENP-K、CENP-L、CENP-M、CENP-N、CENP-U、CENP-W、MIS18B、HJURP、ZWINT、NDC80、SPC24、SPC25和NUF2) 的} \log_2(\text{mRNA水平}) \text{ 的总和。}$

[0112] 在一些实施方案中, 患者的CES评分被归类为相对于参考标度(例如来自与患者具有相同癌症的参考对象群体的CES评分范围)“高”、“中等”或“低”。在一些情况下, 高评分对应于参考标度的最高三分位数(例如, 最高的1/3)中的数值; 中间评分对应于参考标度的中间三分位数(例如, 中间的1/3); 以及低评分对应于参考标度的低三分位数(例如, 最低的1/3)。在其他实施方案中, 高评分表示基于0到1标度的归一化、标准化参考标度的0.66或更高的风险评分, 例如0.66、0.67、0.70、0.75、0.80、0.85、0.90、0.95、0.99或1.0。在其他实施方案中, 将患者的CES评分与一个或多个阈值进行比较, 以提供癌症进展和/或响应化疗或放疗的可能性。在一些情况下, 高风险评分对应于这样的数值, 例如在参考标度前5%、前

10%、前15%、前20%、前25%、前30%、前35%、前40%、前45%、前50%或前60%的风险评分。

[0113] 为了建立用于实践本发明的方法的参考CES标度或阈值,可以使用对象的参考群体。在一些实施方案中,参考群体患有与测试患者相同类型的癌症或肿瘤。参考群体可以具有与测试患者相同的癌症或肿瘤亚型和/或阶段。出于使用本文提供的方法筛选和/或监测癌症的目的,如果适用,则参考群体中的对象可以在适当的参数内。任选地,参考对象具有相同的性别,相似的年龄或相似的种族背景。参考对象可以具有与测试对象相同的性别,相似的年龄或相似的种族背景。

[0114] 可以通过充分建立的常规使用方法来确认参考对象的状态,其包括但不限于个体的一般体格检查及其病史的一般回顾。此外,参考对象组必须具有合理的大小,从而使从所述组中获得的样本中14个CEN/KT多核苷酸(mRNA)或相应多肽的平均水平可合理地被视为代表该对象群体中的正常或平均水平。

[0115] 为了建立阈值,基于所选择的参考组每个对象中发现的个体值来确定平均值。例如,超过阈值的风险评分可以指示超出平均可能性的癌症进展,而低于阈值的风险评分可以指示平均可能性或低于平均可能性的癌症进展。在一些实施方案中,也在同一过程中确定标准差。在一些情况下,可以为具有不同特征(例如年龄、性别或种族背景)的单独定义的组建立单独的阈值。

[0116] 根据本文所述的方法,将患者的风险评分与一个或多个阈值进行比较。在一些实施方案中,如果风险评分高于阈值至少1个、2个、3个、4个、5个、10个、15个、20个或更多个标准差,则认为风险评分“高”。在其他实施方案中,如果风险评分低于阈值至少1个、2个、3个、4个、5个、10个、15个、20个或更多个标准差,则风险评分“低”或低于阈值。

[0117] 在一些实施方案中,使用基于计算机的分析程序将由本文所述的检测方法产生的原始数据(例如,给定标志物的存在、不存在或量)转化为对临床医生具有预测价值的风险评分。

[0118] E. 将CES评分与预后结果或治疗选择相关联

[0119] 根据上述方法确定的CES评分可以预测患者具有高于平均值或统计学显著的不良预后的可能性,例如肿瘤进展、总体生存期、再发生、复发和/或转移。如果用化疗或放疗进行治疗,此患者具有统计学显著的响应治疗的可能性,并且在一些情况下具有良好的预后,例如长期总体生存期、无任何事件生存期、无转移性复发生存期、无远端转移生存期和/或无病生存期。

[0120] 在一些实施方案中,如果对象具有高CES评分,则存在对象将具有不良预后的可能性,例如将经历死亡、癌症复发、癌症再发生、癌症进展和/或局部或远端转移。在一些情况下,高CES评分预测,对象将具有不良无转移性复发生存期(MRFS),如短期MRFS、无MRFS或低于平均概率的MRFS。在其他情况下,高CES评分预测,对象将具有不良无任何事件(AE)生存期,例如短期无AE生存期、没有无AE生存期或低于平均概率的无AE生存期。对象可能患有早期癌症,如早期乳腺癌、肺癌、胃癌或卵巢癌。

[0121] 在一些实施方案中,与具有低CES值的乳腺癌对象相比,预测患有乳腺癌并具有高CES值的对象将具有不良无远端转移生存期(DMFS),例如较短期的DMFS、无DMFS或低于平均概率的DMFS。具有高CES的乳腺癌患者也可能患有ER阴性肿瘤和/或晚期肿瘤等级。此类患

者可能患有侵袭性肿瘤。预测患有ER阳性乳腺癌并具有高CES的患者具有不良无AE生存期、不良DMFS或不良无复发生存期。在一些情况下,患有ER阳性乳腺癌的患者具有正常乳腺样肿瘤、管腔A型肿瘤或管腔B型肿瘤。在一些情况下,已经用辅助性的他莫昔芬疗法对患有ER阳性乳腺癌的患者进行了治疗。预测患有管腔A型肿瘤并具有高CES的患者具有不良总体生存期。类似地,预期患有管腔B型肿瘤并具有高CES的患者具有不良总体生存期。

[0122] 在一些实施方案中,在接受癌症治疗后,患有特异性乳腺癌亚型(例如,基底样亚型)并具有高CES值的乳腺癌患者将具有总体生存期和/或无病生存期(DFS)。也预测这些患者在接受治疗之后具有良好的DMFS。另外,预测这些患者将响应癌症疗法。在一些情况下,癌症疗法包括单独的放疗、化疗,放疗和化疗,或上述任何一种与药物疗法、激素疗法和/或手术的组合。

[0123] 在一些实施方案中,与未接受辅助疗法相比,预测患有肺癌(如非小细胞肺癌(NSCLC))并具有高CES值的患者响应辅助性癌症疗法,并且经历改善的总体生存期和/或改善的DFS。如果确定患有肺癌的患者具有高CES值,则推荐或建议患者接受辅助性放疗、辅助性化疗或其任何组合。

[0124] F. 计算机实施的方法、系统和设备

[0125] 可以用包括一个或多个处理器的计算机系统全部或部分实施本文所述的任何方法,可以配置该计算机系统来实施这些步骤。因此,实施方案涉及配置为实施本文所述的任何方法的步骤的计算机系统,潜在地具有执行各自步骤或各步骤组的不同组件。尽管呈现为编号的步骤,但是可以同时或者以不同的顺序实施本文的方法的步骤。此外,这些步骤的部分可以与来自其他方法的其他步骤的部分一起使用。另外,全部或部分步骤可以是任选的。可以用模块、电路或实施这些步骤的其它方式来实现任何方法的任何步骤。

[0126] 本文提到的任何计算机系统可以利用任何合适数量的子系统。此类子系统的实例在图9的计算机装置100中示出。在一些实施方案中,计算机系统包括单个计算机装置,其中子系统可以是计算机装置的组件。在其他的实施方案中,计算机系统可以包括多个计算机装置,每个计算机装置都是具有内部组件的子系统。

[0127] 图9中所示的子系统经系统总线175互连。示出了另外的子系统,如打印机174、键盘178、存储设备179、耦合到显示适配器182的监视器176以及其他子系统。可以通过任何数量的本领域已知的方式(如串行端口177)将耦合到I/O控制器171的外围设备和输入/输出(I/O)设备连接到计算机系统。例如,可以使用串行端口177或外部接口181(例如,以太网、Wi-Fi等)将计算机系统100连接到诸如因特网、鼠标输入设备或扫描仪的广域网。经由系统总线175的互连允许中央处理器173与每个子系统通信,并控制来自系统存储器172或存储设备179(例如,固定磁盘,如硬盘驱动器或光盘)的指令的执行以及子系统之间的信息交换。系统存储器172和/或存储设备179可以包含计算机可读介质。本文提到的任何数据均可以从一个组件输出到另一个组件,并且可以输出给用户。

[0128] 计算机系统可以包括例如通过外部接口181或通过内部接口连接在一起的多个相同的组件或子系统。在一些实施方案中,计算机系统、子系统或装置可以通过网络进行通信。在这种情况下,可将一个计算机可认为是客户端,另一个计算机是服务器,其中每个计算机可以是同一计算机系统的一部分。客户端和服务器可以各自包括多个系统、子系统或组件。

[0129] 应当理解,可以使用硬件(例如,专用集成电路或现场可编程门阵列)和/或模块化或集成方式的通用可编程处理器利用计算机软件以控制逻辑的形式来实施本公开的任何实施方案。作为本文中的用户,处理器包括在同一集成芯片上的多核处理器,或在单一电路板上或联网的多个处理单元。基于本文提供的公开和教导,本领域普通技术人员将了解并理解使用硬件以及硬件和软件的组合来实施本公开的实施方案的其他方式和/或方法。

[0130] 可以将本申请中描述的任何软件组件或功能作为由使用任何合适的计算机语言(如例如Java、C++或Perl),使用例如常规或面向对象的技术,通过处理器执行的软件代码来实施。可以将软件代码作为一系列指令或命令存储在计算机可读介质上,以存储和/或传输,合适的介质包括随机存取存储器(RAM)、只读存储器(ROM)、磁介质(例如硬盘驱动器或软盘)或光学介质(如光盘(CD)或DVD(数字通用光盘))、闪速存储器等。计算机可读介质可以是此类存储或传输设备的任何组合。

[0131] 也可以使用适于经由符合各种协议的有线网络、光学网络和/或无线网络(包括因特网)进行传输的载波信号编码和传输的此类程序。因此,可以使用利用此类程序编码的数据信号来创建根据本发明的实施方案的计算机可读介质。编码有程序代码的计算机可读介质可以与兼容设备一起打包或者与其他设备分开提供(例如,经由因特网下载)。任何此类计算机可读介质可以驻留在单个计算机产品(例如,硬盘驱动器、CD或整个计算机系统)上或其内部,并且可以存在于系统或网络内的不同计算机产品上或其内部。计算机系统可以包括监视器、打印机或用于向用户提供本文提到的任何结果的其他合适的显示设备。

[0132] G. 选择癌症疗法

[0133] 在一些实施方案中,本公开的方法也包括为患有癌症且具有高CES评分或高死亡、疾病复发、再发生、进展和/或转移癌症可能性的对象选择治疗干预。对于具有高CES值的癌症对象,可以选择放疗、化疗、药物疗法(例如激素疗法)、免疫疗法、手术或其任何组合。在一些情况下,向具有高CES评分的对象建议手术和辅助疗法,例如放疗、化疗、药物疗法(例如激素疗法)、免疫疗法或其任何组合。

[0134] 有用的化疗剂的非限制性实例包括:烷化剂,例如环磷酰胺、氮芥、苯丁酸氮芥、异环磷酰胺、美法仑、链脲菌素、卡莫司汀、洛莫司汀、白消安、达卡巴嗪、替莫唑胺、塞替派和六甲蜜胺;抗代谢物,例如5-氟尿嘧啶(5-FU)、6-巯基嘌呤(6-MP)、卡培他滨、阿糖胞苷、氟尿苷、吉西他滨、羟基脲、甲氨蝶呤和培美曲塞;蒽环类,例如柔红霉素、阿霉素、表柔比星和伊达比星;其它抗肿瘤抗生素,例如放线菌素-D、博来霉素、丝裂霉素-c和米托蒽醌;拓扑异构酶抑制剂,例如拓扑替康、伊立替康、依托泊苷、替尼泊苷和metoxantrone;有丝分裂抑制剂,例如紫杉醇、多西他赛、伊沙匹隆、长春花碱、长春新碱、长春瑞滨(vinorelbine)和雌莫司汀;皮质类固醇,例如泼尼松、甲基泼尼松和地塞米松,L-天冬酰胺酶(L-asparaginase)以及蛋白酶体抑制剂硼替佐米。不受限制,放疗的实例包括外照射放疗、内照射放疗或全身放疗。

[0135] 治疗干预可包括一种或多种DNA损伤化合物或DNA损伤剂,如顺铂、卡铂、奥沙利铂、吡铂、其他基于铂的化合物、阿霉素、柔红霉素、其它蒽环类化合物、它们的变体及其衍生物。

[0136] 在一些情况下,推荐施用一种或多种蒽环类化合物;环磷酰胺;一种或多种紫杉烷;甲氨蝶呤;5-氟尿嘧啶;一种或多种蒽环类化合物和环磷酰胺;一种或多种蒽环类化合

物、环磷酸胺和一种或多种紫杉烷；环磷酸胺、甲氨蝶呤和5-氟尿嘧啶；或者它们的任何组合。

[0137] 在一些实施方案中，推荐一种或多种治疗干预。可以同时或顺序施用各疗法。

[0138] H. 试剂盒

[0139] 为了用于上述诊断应用、预后应用和疗法选择应用，本文还公开了试剂盒。本发明的试剂盒可以包含实施本文所述的方法任何或所有试剂。在这样的应用中，试剂盒可以包含以下中的任何或全部：分析试剂、缓冲剂、与本文所述的基因中的至少一种结合的核酸、杂交探针和/或引物、与由本文所述的基因编码的多肽中的至少一种特异性结合的抗体或其它部分等。另外，试剂盒可以包含诸如与参考基因或参考多肽特异性结合的核酸、杂交探针、引物、抗体等的试剂。试剂盒可以包含针对本文鉴定的一种或多种参考基因的探针，所述参考基因如ARPC2、ATF4、ATP5B、B2M、CDH4、CELF1、CLTA、CLTC、COPB1、CTBP1、CYC1、CYFIP1、DAZAP2、DHX15、DITM1、EEF1A1、FLOT2、CAPDH、GUSB、HADHA、HDLBP、HMBS、HNRNPC、HPRT1、HSP90AB1、MTCH1、MYL12B、NACA、NDUFB8、PGK1、PPIA、PPIB、PTBP1、RPL13A、RPLP0、RPS13、RPS23、RPS3、S100A6、SDHA、SEC31A、SET、SF3B1、SFRS3、SNRNP200、STARD7、SUMO1、TBP、TFRC、TMBIM6、TPT1、TRA2B、TUBA1C、UBB、UBC、UBE2D2、UBE2D3、VAMP3、XP01、YTHDC1、YWHAZ和18S rRNA。

[0140] 在检测试剂的背景下，本文所使用的术语“试剂盒”旨在指如下事物：多种基因表达产物检测试剂的组合，或者一种或多种基因表达产物检测试剂与一种或多种其它类型的要素或组件（例如，其它类型的生化试剂、容器、包装（例如旨在用于商业销售的包装），附着有基因表达检测产物试剂的基底、电子硬件组件等）的组合。

[0141] 在一些实施方案中，本公开提供了附着于固体支持物（例如阵列载玻片或芯片）的寡核苷酸探针，例如如DNA Microarrays: A Molecular Cloning Manual, 2003, Eds. Bowtell和Sambrook, Cold Spring Harbor Laboratory Press中所述。此类设备的构造在本领域中众所周知，例如如美国专利和专利出版物，美国专利第5,837,832号；PCT申请W095/11995；美国专利第5,807,522号；美国专利第7,157,229号、第7,083,975号、第6,444,175号、第6,375,903号、第6,315,958号、第6,295,153号和第5,143,854号、第2007/0037274号、第2007/0140906号、第2004/0126757号、第2004/0110212号、第2004/0110211号、第2003/0143550号、第2003/0003032号和第2002/0041420号中所述。下述参考文献中也综述了核酸阵列：Biotechnol Annu Rev 8:85-101 (2002)；Sosnowski等人，Psrychiatr Genet 12(4):181-92 (Dec. 2002)；Heller, Annu Rev Biomed Eng 4:129-53 (2002)；Kolchinsky等人，Hum. Mutat 19(4):343-60 (Apr. 2002)；和McGail等人，Adv Biochem Eng Biotechnol 77:21-42 (2002)。

[0142] 微阵列可以由大量固定于固体支持物上的独特的单链多核苷酸组成，通常是合成的反义多核苷酸或cDNA片段。典型的多核苷酸优选长度为约6-60个核苷酸长度，更优选长度为约15-30个核苷酸，以及最优选长度为约18-25个核苷酸。对于某些类型的阵列或其他检测试剂盒/系统，可以优选使用长度仅为约7-20个核苷酸的寡核苷酸。在其它类型的阵列中，例如与化学发光检测技术结合使用的阵列，优选的探针长度可以是例如长度为约15-80个核苷酸，优选长度为约50-70个核苷酸，更优选长度为约55-65个核苷酸，以及最优选长度为约60个核苷酸。

[0143] 另外,试剂盒可以包括含有用于实践本文提供的方法的指导(即方案)的说明材料。虽然说明材料通常包括书面或印刷材料,但它们不限于此。本发明考虑能够存储此类指令并将它们传达给终端用户的任何介质。这样的介质包括但不限于电子存储介质(例如,磁盘、磁带、盒(cartridge)、芯片)、光学介质(例如,CD ROM)等。这样的介质可以包括提供此类说明材料的互联网网站的地址。

[0144] 尽管出于清楚理解的目的通过说明和示例在一定程度上详细描述了上述发明,但是本领域技术人员将理解,可以在所附权利要求的范围内实施某些改变和修改。另外,将本文提供的各参考文献通过引用整体并入,其程度如同各参考文献通过引用单独并入。

IV. 实施例

[0145] 提供以下实施例以说明而非限制要求保护的发明。

[0146] 实施例1.特定着丝粒和动粒蛋白的错误表达预测癌症患者的生存期以及对放疗和化疗的应答

[0147] 染色体不稳定性(CIN)是大多数人类癌症的标志,并且有助于肿瘤异质性和耐药性。在有丝分裂和染色体重排过程中,异常的着丝粒和动粒功能通过染色体错误分离引起CIN。然而,我们以前缺乏对着丝粒和动粒结构蛋白基因(CEN/KT基因)在患者癌症进展中的影响的系统理解。利用来自许多癌症类型的大量患者数据,基于癌症组织和正常组织之间的差异基因表达以及基因错误表达的预后值,我们鉴定了14个CEN/KT基因的核心子集,并设计与各种类型的癌症中的拷贝数变化(CNA)和突变频率密切相关CES系统(用于着丝粒和动粒基因表达评分)。CES评分是大范围癌症类型(包括乳腺癌、肺癌、胃癌和早期卵巢癌)的总体患者生存期和疾病进展的有效预测因子。最重要的是,高CES值与癌症细胞系中对DNA损伤化合物的敏感性相关,并且有效预测化疗和放疗后癌症患者的治疗结果。这些结果提示,CES系统提供了用于癌症预后和选择治疗选择的新的且有效的生物标志物。

[0148] 引言

[0149] 基因组不稳定性是大多数人类癌症的特性,并且被认为能够获得其他癌症标志,如不受抑制的细胞生长和增殖、异质性和耐药性(Hanahan和Weinberg,2011)。基因组不稳定的主要类型是染色体不稳定性(CIN),其特征是染色体异常率增加,包括全染色体或大片段(非整倍性)的获得/丧失、结构重排和病灶畸变(例如,扩增和缺失)(Beroukhi等人,2010;Geigl等人,2008)。已在癌前期病变和恶性生长中观察到CIN(Pihan等人,2003)。已经假设非整倍性促进肿瘤发生(Boveri,1902,1914;Siegel和Amon,2012;Weaver等人,2007)。最近的癌症基因组测序研究提示,CIN有助于肿瘤间和肿瘤内的异质性(de Bruin等人,2014;Hiley和Swanton,2014;Nik-Zainal等人,2012;Yachida等人,2010;Zhang等人,2014)。异常染色体行为可以引起异常染色体结构,增加突变频率或表观遗传修饰基因活性。

[0150] 在癌症发展过程中,有缺陷的染色体分离和完整性通路(integrity pathway)可以与改变的信号转导通路协同作用,或者它们可以独立地起作用。例如,许多致癌基因编码由染色体易位或倒位导致的融合或调控异常的信号转导分子(Pierotti,2003)。在另一个实施例中,胶质母细胞瘤细胞中EGFR基因座的动态扩增有助于抵抗EGFR抑制剂(Nathanson等,2014)。另一方面,纺锤体检查点组分(如Bub1、BubR1和Mad1)中的稀有突变促进小鼠和

人类的非整倍体并且和肿瘤发生相联系 (Kops等人, 2005)。因此, 基因组景观的连续修改可以允许促进癌症进展和生长的有害变化快速积累 (Schvartzman等人, 2010), 并在治疗期间赋予药物抗性 (Lee等人, 2011; Nathanson等人, 2014; Swanton等人, 2007; Weaver和Cleveland, 2006)。因此, 基因组景观的连续修改可以允许促进癌症进展和生长的有害变化的快速积累 (Schvartzman等人, 2010), 并在治疗期间赋予药物抗性 (Lee等人, 2011; Nathanson等人, 2014; Swanton等人, 2007; Weaver和Cleveland, 2006)。一个矛盾的现象是, 极端的CIN可以引起生长不利, 大概是由于来自遗传毒性效应和蛋白质组不平衡的过度应激 (Hiley和Swanton, 2014; Siegel和Amon, 2012)。这些结果和选择性杀死显示CIN的癌细胞的可能性证明CIN既是癌症治疗的挑战也是潜在的机会 (Carter等人, 2006; Roschke和Kirsch, 2005)。

[0151] 虽然在大多数散发性癌症 (sporadic cancer) 中CIN的确切原因仍不清楚, 但提出的机制包括致癌基因诱导的复制应激、由端粒功能障碍或易位引起的断裂-融合-桥循环以及异常有丝分裂 (Artandi和DePinho, 2010; Kops等人, 2005; Negrini等人, 2010)。染色体错误分离也可以伴随着通过染色体断裂的DNA损伤 (Janssen等人, 2011)。在真核细胞的有丝分裂期间, 着丝粒和相关的动粒是正常染色体集合、有丝分裂检查点功能和姐妹染色单体分离所需的 (Allshire, 1997; Cleveland等人, 2003; Earnshaw等人, 1991)。因此, CIN的另一潜在原因是着丝粒和动粒功能的调控异常, 其导致细胞系和模式生物体中的染色体异常。

[0152] 着丝粒和动粒由着丝粒染色质以及内部和外部动粒结构组成 (图1A)。在活性着丝粒富集的组蛋白H3变体CENP-A, 是决定着丝粒身份及其忠实增殖的关键表观遗传标记 (Allshire和Karpen, 2008; Black等人, 2010)。所有其他着丝粒和动粒蛋白最终需要CENP-A用于它们的定位 (Allshire和Karpen, 2008)。CCAN (组成型着丝粒相关网络) 将染色质与外部动粒连接, 并且包含几个亚复合物。CENP-T/-W/-S/-X复合物位于散布在CENP-A核小体之间的H3结构域中 (Blower等人, 2002; Foltz等人, 2006; Nishino等人, 2012; Sullivan和Karpen, 2004)。CENP-N/-L/-M决定CENP-H/-I/-K的定位, 其转而是CENP-O/-P/-Q/-R/-U募集所需要的。CCAN确保KMN网络 (KNL1复合物、MIS 12复合物和NDC80复合物) 至外部动粒的募集 (Cheeseman和Desai, 2008; Hori等人, 2008; Perpelescu和Fukagawa, 2011), 并且NDC80复合物与纺锤体微管相互作用。

[0153] 使用HJURP作为其分子伴侣和装配因子, 将CENP-A并入到着丝粒染色质 (Dunleavy等人, 2009; Foltz等人, 2009; Mellone等人, 2009)。由包含MIS18A、MIS18B和M18BP1亚基的MIS18复合物将HJURP募集到着丝粒 (Barnhart等人, 2011; Fujita等人, 2007; Moree等人, 2011)。此外, CENP-A的装配还需要几个CCAN组件, 如CENP-C以及CENP-H/-I/-K和CENP-N/-L/-M的成员 (Black和Cleveland, 2011; Cheeseman等人, 2008; Moree等人, 2011; Okada等人, 2006)。

[0154] 着丝粒和动粒蛋白的水平似乎被严格调控以确保忠实的染色体分离。通过突变或缺失引起的许多着丝粒和动粒蛋白的减少导致染色体分离缺陷和细胞死亡, 并且一些关键蛋白的过表达也可以损害功能 (Allshire和Karpen, 2008)。例如, 果蝇CENP-A的过表达或异位束缚 (ectopic tethering) 导致错误并入通常非着丝粒染色体区域, 引起新着丝粒形成, 双着丝粒行为以及染色体断裂和异常 (Heun等人, 2006; Mendiburo等人, 2011)。细胞系中高水平的人CENP-A或HJURP也错误定位到基因组的其他区域并引起染色质桥和微核

(Mishra等人,2011)。重要的是,几种着丝粒蛋白(包括CENP-A、HJURP和CENP-H)的过表达与几种类型的癌症的不良预后相关,提示在癌症病因中的潜在作用(Hu等人,2010;Mcgovern等人,2012;Tomonaga等人,2005;Tomonaga等人,2003)。

[0155] 我们假设,着丝粒和动粒蛋白基因的调控异常引起助于人癌症进展和/或肿瘤维持的染色体异常。我们系统地研究了癌组织和正常组织之间着丝粒和动粒结构蛋白(CEN/KT)基因的差异表达,并研究了其基因表达水平的预后影响。我们发现,在许多不同类型的癌症中一致观察到14个CEN/KT基因的过表达。我们获得了CES系统(用于着丝粒和动粒基因表达评分),其总结了跨几种癌症类型的着丝粒和动粒基因错误表达的程度。我们将CES值与大范围的癌症类型中的基因组不稳定性的部分相关联。我们还显示肿瘤CES值预测患者生存期和疾病进展。放疗(RT)和DNA损伤化合物通过损伤DNA无法修复来杀死癌细胞,随后是细胞死亡,但与对患者的潜在严重附带损伤相关。我们假设具有高CES值的肿瘤已经遭受高水平的遗传毒性应激,并且可能对进一步的损伤更敏感。高CES癌细胞系对引起DNA损伤并抑制复制和转录的Topo I抑制剂的敏感性支持这一想法(Mathijssen等人,2002)。使用乳腺癌和肺癌临床数据集,我们发现CES系统有效地预测了在RT或辅助化疗之后的患者结果,并建议这个系统可能对优化临床肿瘤学中的治疗方案有价值。这里鉴定的CEN/KT基因是决定动态着丝粒和动粒结构的组件,并在忠实的染色体分离中发挥关键作用。许多CES基因参与CENP-A核小体装配过程,提示该过程在癌症进展中的潜在重要性。这些染色体功能不同于参与信号转导和致癌或肿瘤阻抑通路的调控的许多现有药物靶标。因此,本文鉴定的CEN/KT蛋白可能是用于癌症治疗药物开发的新靶标,当与靶向正交通路(例如信号转导和细胞生长)的组分的药物组合时,其可能特别有效。

[0156] 结果

[0157] A. 人类癌症中着丝粒和动粒蛋白基因子集的调控异常

[0158] 我们手动编辑了31个着丝粒和动粒(CEN/KT)蛋白基因的列表(图1B)以研究它们在癌症预后中的潜在作用。该列表限于被证明定位于着丝粒或动粒并决定着丝粒或动粒结构的蛋白(Black等人,2010)。该列表包括CENP-A和下游的CCAN和KMN组分,以及CENP-A核小体装配和着丝粒增殖所需的因子(例如HJURP和MIS18),其瞬时定位于着丝粒(Dunleavy等人,2009;Foltz等人,2009;Fujita等人,2007;Mellone等人,2009)。

[0159] 使用GEO数据集,对大范围人癌症类型,特别是乳腺癌、肺癌、肝癌、鼻咽癌、胃癌、头颈癌、宫颈癌、前列腺癌、胶质母细胞瘤、结肠直肠癌、胰腺癌和卵巢癌,进行CEN/KT基因表达分析(表1)。与相应的正常组织相比,在大多数类型的癌症中我们一致地观察到CEN/KT基因子集的调控异常。另外,我们检测了癌症进展期间不同疾病阶段的肿瘤之间的差异。例如,在乳腺组织中,乳腺导管内原位癌(DCIS)中的CEN/KT的调控异常在正常组织和浸润性导管癌(IDC)中间。在肝脏中,CEN/KT基因在肝硬化或发育不良的组织中没有显示显著的调控异常,但是在从非常早期阶段到非常晚期的癌症中显著地调控异常。在前列腺中,仅转移癌显示出CEN/KT基因的显著调控异常,而恶性局部损伤显示有限的调控异常。这些结果提示CEN/KT蛋白的调控异常在癌症进展的开始或早期阶段的潜在作用。另外,肝发育不良中CEN/KT基因调控异常的缺乏提示有缺陷的CEN/KT调控不仅仅是细胞过度增殖的结果。

[0160] 表1. 用于研究CEN/KT基因的GEO数据集

[0161]

癌症类型	GEO系列登录号
------	----------

乳腺癌	GSE21422和GSE3744
肺癌	GSE19188
卵巢癌	GSE14407
肝癌	GSE6764
胰腺癌	GSE16515
结肠直肠癌	GSE8671
鼻咽癌	GSE12452
胃癌	GSE13911
头颈癌	GSE6791
宫颈癌	GSE6791
前列腺癌	GSE3325
胶质母细胞瘤	GSE4290

[0162] 为了处理CEN/KT基因在癌症进展中的潜在作用,我们分析了包含代表疾病进展的不同阶段的组织样本的数据集。具体来说,我们对来自9种癌症类型的13个Affymetrix基因表达微阵列数据集的分析显示,在超过50%的将癌症与相应的正常组织以及早期与晚期肿瘤样本相比较的癌症数据集中,15个CEN/KT基因的表达显著地调控异常(FDR $p < 0.05$,至少2倍的差异)(图1C和表2)。这些结果指示,有缺陷的CEN/KT基因调控在大量癌症中是保守的,并且可能在疾病进展中起重要作用。

[0163]

表 2. 用于研究差异表达的 CEN/KT 基因的 GEO 数据集

基因	探针 ID	乳腺				宫颈	头&颈		结肠	胃	脑脊 CNS				肝脏			肺				胰腺																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
		GSE17080	GSE21422	GSE3744	GSE6791	GSE6791	GSE6791	GSE12452	GSE8671	GSE13911	GSE4290			GSE6764			GSE19188				GSE31210		GSE16515																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
CENP-A	204962_s_at	30.98	2.83	5.27	8.75	BC 对比正常																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													</

[0164]

CENP-S	213454_at	12.06	0.28	0.95	2.80	5.74	2.06	0.29	16.29	0.47	0.17	0.67	0.28	0.83	0.30	2.77	11.49	3.84	3.64	6.57	0.32
CENP-X	209478_at	14.56	2.44	4.71	0.87	0.03	0.08	0.89	4.53	1.51	0.05	1.68	1.92	0.20	1.80	3.87	10.68	11.88	13.76	8.85	3.24
CENP-C	204739_at	6.71	1.25	1.54	0.91	2.86	0.03	0.56	1.46	1.89	0.61	1.07	0.19	0.15	0.57	1.48	2.53	4.33	2.45	0.88	2.75
CENP-N	219555_s_at	21.13	1.53	3.69	5.96	3.62	3.23	3.81	18.46	6.20	0.82	7.90	0.57	0.26	1.35	8.56	15.33	15.21	31.04	4.06	3.18
	222118_at	5.59	0.55	2.54	2.49	3.96	5.11	3.88	10.95	3.97	0.11	4.82	0.30	0.03	1.08	5.35	4.85	3.00	12.01	0.74	0.46
	228559_at	21.26	2.71	2.59	5.66	4.53	2.12	2.53	13.61	4.35	0.83	1.27	1.52	0.23	1.32	1.61	9.43	7.21	23.60	1.94	1.46
CENP-M	234811_at	2.59	0.02	0.60	0.51	0.28	0.15	0.57	1.45	0.50	0.72	2.54	1.41	0.47	0.51	0.73	1.68	0.21	1.06	0.16	1.78
	218741_at	24.37	2.19	3.40	4.38	3.36	2.26	0.94	8.71	7.41	3.42	6.77	3.38	0.38	2.46	5.08	10.61	13.00	22.74	8.90	4.17
CENP-U	218883_s_at	37.45	3.12	3.19	7.30	5.54	0.32	7.22	11.01	4.63	8.43	15.47	7.14	0.12	4.37	10.03	13.16	15.13	23.22	7.44	6.89
	229304_s_at	7.37	1.98	1.65	1.71	1.70	0.51	4.02	7.01	0.35	3.24	7.81	3.88	1.47	0.79	3.30	6.97	5.79	12.70	2.37	2.37
CENP-H	229305_at	24.09	1.43	1.31	3.17	3.39	0.94	6.02	8.36	1.61	2.52	8.77	3.55	1.27	0.67	2.54	12.41	8.95	19.81	4.74	2.92
	231772_x_at	17.85	1.74	2.07	2.02	3.38	3.51	5.46	11.62	6.93	5.25	8.20	3.13	0.07	1.69	5.11	15.83	10.18	19.21	3.95	1.23
CENP-I	1555046_at	1.01	0.46	0.20	1.07	0.81	0.44	0.91	1.23	0.74	0.47	0.12	0.22	0.00	0.70	1.82	2.48	1.23	0.75	0.96	3.30
	1563223_a_at	0.07	0.25	0.09	0.94	1.43	1.54	0.85	1.33	2.20	0.25	1.23	0.70	0.33	1.68	1.43	1.45	0.89	0.86	0.53	2.77
	207590_s_at	12.03	0.98	1.52	2.32	1.75	1.53	3.70	8.26	1.47	0.53	2.88	0.52	0.22	0.09	3.50	13.56	13.93	21.23	5.26	0.43
CENP-L	214804_at	17.29	0.65	1.22	2.32	5.19	2.70	4.71	6.92	4.35	1.64	6.96	2.04	0.30	0.34	6.83	14.58	14.02	30.52	3.89	2.20
	1554271_a_at	20.53	3.57	2.99	4.54	0.91	2.43	1.92	10.63	3.93	4.22	11.16	3.79	0.47	1.31	6.83	15.96	18.96	22.31	6.30	3.26
CENP-K	232065_x_at	9.57	4.56	2.84	3.83	4.45	3.39	4.20	8.42	9.45	4.21	12.98	2.86	0.11	3.29	8.56	14.14	13.80	22.30	6.57	3.07
	222848_at	29.45	2.84	2.61	3.76	11.72	1.57	6.76	16.12	4.87	5.49	17.13	3.66	0.35	2.55	8.39	14.39	16.39	19.78	6.94	7.03
CENP-O	219472_at	1.89	0.24	0.35	0.12	1.49	0.20	0.01	0.50	1.38	1.45	2.69	2.05	0.30	1.05	0.04	6.35	4.93	9.48	0.68	1.78
	226118_at	15.16	1.81	1.98	2.75	4.76	2.79	3.05	4.28	3.02	3.50	4.92	1.88	0.05	1.44	4.61	16.95	10.16	18.59	1.14	2.83
CENP-Q	219294_at	5.36	0.10	0.27	1.77	8.13	3.39	4.03	9.66	4.18	5.63	5.71	4.69	0.05	0.74	5.61	10.00	4.87	5.66	0.66	1.12
CENP-R	205176_s_at	0.30	0.60	1.59	0.06	6.54	0.59	5.11	12.20	0.97	7.38	14.69	0.97	1.22	2.97	6.25	5.86	4.27	4.78	3.83	0.46
KNL1	1552680_a_at	32.52	0.58	0.12	3.16	1.62	0.47	1.31	3.15	4.37	1.78	7.99	1.84	1.41	1.82	2.77	14.73	14.03	24.03	5.33	1.66
	1552682_a_at	2.97	0.84	0.64	0.98	0.32	0.04	0.82	4.67	1.57	0.03	0.92	1.94	0.09	1.37	2.35	8.81	6.42	9.34	2.26	0.69
	220247_at	0.14	0.45	0.37	0.73	1.52	3.81	0.25	0.17	0.56	0.25	1.97	0.79	0.01	0.85	1.98	0.24	0.71	0.11	0.09	4.67
ZWINT	228323_at	38.73	3.05	4.18	9.34	8.85	1.10	5.12	13.74	5.34	5.67	14.71	6.61	0.08	2.50	2.96	16.43	17.40	29.41	7.99	5.42
	204026_s_at	44.43	3.62	3.42	9.20	8.74	3.10	9.20	14.70	4.08	3.44	9.55	4.53	0.02	3.48	9.28	14.75	17.70	21.83	8.72	6.53
MIS12	221559_s_at	0.01	0.43	0.81	0.25	8.92	1.59	1.52	1.41	0.05	3.83	8.02	3.20	1.37	0.10	0.56	0.13	0.68	0.10	0.84	0.76

[0165]

NSL1	209483_s_at	5.65	0.40	0.15	1.81	0.98	0.12	0.09	0.42	0.70	0.51	0.40	0.04	0.14	0.40	0.45	4.67	1.29	1.62	6.30	0.49
	209484_s_at	4.87	0.53	0.51	1.37	4.00	1.56	1.82	0.45	1.90	3.76	3.82	3.58	0.46	6.25	7.26	10.22	6.56	5.23	6.01	1.25
	230592_at	0.29	0.10	0.87	0.81	3.52	0.63	0.90	0.54	0.31	1.31	2.81	2.93	0.30	0.54	1.44	3.16	0.59	3.73	0.05	0.64
	235799_at	0.43	0.09	0.45	0.07	2.28	2.93	0.34	0.83	2.87	0.48	0.62	0.47	0.10	0.10	0.31	1.78	0.28	1.13	1.07	1.64
PMF1	202337_at	5.86	2.30	3.72	0.33	0.51	2.40	0.00	6.97	0.38	5.13	7.42	7.37	2.28	3.23	11.39	4.97	5.03	0.09	6.20	3.51
KNL3	219512_at	14.65	1.43	1.28	3.67	5.02	4.92	3.76	7.22	3.94	1.89	9.53	1.47	0.95	1.62	5.64	8.30	5.97	12.44	1.22	4.61
NDC80	204162_at	28.46	3.16	3.56	5.33	6.95	2.68	6.97	15.36	8.77	10.01	27.43	10.70	0.05	6.56	11.78	16.17	15.58	27.49	4.39	5.69
SPC24	235572_at	20.13	1.79	2.81	4.01	3.21	0.39	2.75	3.03	11.63	3.85	14.29	6.71	0.37	5.78	7.97	15.80	15.17	29.58	8.76	1.65
SPC25	209891_at	34.65	1.93	2.94	4.58	4.22	1.90	2.87	15.78	7.47	2.78	7.66	5.66	0.15	2.00	7.60	22.18	13.90	28.24	5.43	2.16
NUF2	223381_at	31.63	2.91	6.65	10.62	8.28	1.88	6.00	17.42	11.67	6.77	13.94	6.85	0.41	3.79	13.32	23.72	21.31	38.00	8.13	7.02

数字表示-log(p 值)。灰色突出指示>2 倍变化

[0166] 然后我们使用跨不同类型癌症的TCGA RNA-seq数据研究了CEN/KT基因的调控异常。最近的研究证实了FoxM1转录因子和动粒基因表达之间的强相关性,并提出在癌症中

CEN/KT基因被FoxM1同时上调(Thiru等人,2014)。与这一观察一致,使用共表达相关性网络分析,我们还在多种癌症类型中检测到了许多CEN/KT基因之间的强相关性。然而,这个网络中的基因数目和相关系数在不同癌症中变化很大,提示在癌症类型内部和癌症类型之间共表达的显著不同的强度。例如,在诸如膀胱癌、宫颈癌和子宫癌的几种癌症中,这一子网络比诸如AML、肺腺癌或低级脑癌的癌症包含少得多的节点和边缘。这一结果提示癌症类型中和/或在相同类型的个体中CEN/KT基因表达的显著调控差异。

[0167] B. CEN/KT基因的子集对于多种人类癌症的生存期和转移具有显著的预后价值

[0168] 为了确定CEN/KT基因调控异常对癌症患者是否具有预后价值,我们使用来自多个癌症GEO数据集的CEN/KT表达数据进行了荟萃分析。首先,我们使用乳腺癌基因表达Miner v3.0(BC-GenExMiner 3.0)(Jezequel等人,2013)对超过3,000份人乳腺癌临床样本中31个CEN/KT基因表达水平的预后影响进行了荟萃分析。22个个体的CEN/KT基因的过表达和CENP-C降低的表达与不良无任何事件(AE)生存期($p < 0.05$)和不良的无转移性复发(MR)生存期(MRFS) ($p < 0.05$)显著相关(表3)。

[0169] 表3. 乳腺癌中CEN/KT基因的无任何事件生存期和无转移性复发生存期

[0170]

基因	乳腺癌 AEFS		乳腺癌 MRFS		
	HR (95%CI)	P 值	HR (95%CI)	P 值	
CENP-A	1.47 (1.31-1.66)	<0.0001	1.56 (1.33-1.83)	<0.0001	过表达
HJURP	1.55 (1.37-1.76)	<0.0001	1.69 (1.43-1.98)	<0.0001	过表达
M18BP1	0.96 (0.85-1.09)	0.5668	0.93 (0.80-1.09)	0.3998	
MIS18A	1.34 (1.18-1.51)	<0.0001	1.48 (1.26-1.74)	<0.0001	过表达
MIS18B	1.42 (1.26-1.61)	<0.0001	1.51 (1.28-1.77)	<0.0001	过表达
CENP-C	0.82 (0.75-0.91)*	0.0001	0.78 (0.69-0.88)*	0.0001	降低的表达
CENP-N	1.61 (1.46-1.77)	<0.0001	1.82 (1.60-2.07)	<0.0001	过表达

[0171]

CENP-I	1.44 (1.28-1.63)	<0.0001	1.46 (1.25-1.72)	<0.0001	过表达
CENP-H	1.27 (1.08-1.50)	0.005	1.32 (1.01-1.72)	0.0413	过表达
CENP-T	1.02 (0.90-1.17)	0.7206	1.08 (0.91-1.28)	0.3921	
CENP-W	1.95 (1.40-2.72)	0.0001	2.13 (1.41-3.19)	0.0003	过表达
CENP-S	1.14 (0.98-1.32)	0.0899	1.10 (0.91-1.32)	0.3313	
CENP-X	1.22 (1.12-1.34)	0.0001	1.20 (1.06-1.36)	0.0038	过表达
CENP-M	1.39 (1.23-1.57)	<0.0001	1.49 (1.27-1.75)	<0.0001	过表达
CENP-U	1.35 (1.19-1.52)	<0.0001	1.51 (1.29-1.77)	<0.0001	过表达
CENP-L	1.44 (1.22-1.70)	<0.0001	1.43 (1.15-1.79)	0.0014	过表达
CENP-K	1.28 (1.10-1.48)	0.0012	1.34 (1.10-1.65)	0.0046	过表达
CENP-O	1.31 (1.16-1.48)	<0.0001	1.28 (1.09-1.50)	0.0024	过表达
CENP-P	0.96 (0.79-1.16)	0.6576	0.88 (0.69-1.13)	0.3268	
CENP-Q	0.99 (0.88-1.12)	0.8696	0.98 (0.84-1.15)	0.8346	
CENP-R	0.96 (0.85-1.08)	0.4688	0.90 (0.77-1.05)	0.1652	
KNL1	1.19 (1.08-1.32)	0.0006	1.21 (1.07-1.38)	0.0034	过表达
ZWINT	1.51 (1.37-1.67)	<0.0001	1, 66 (1, 46-1, 89)	<0.0001	过表达
MIS12	1.14 (1.00-1.30)	0.045	1.17 (0.98-1.39)	0.0769	
PMF1	0.96 (0.87-1.05)	0.3813	0.99 (0.87-1.12)	0.8274	
NSL1	1.02 (0.90-1.14)	0.7956	1.03 (0.88-1.21)	0.6793	
KNL3	1.26 (1.11-1.42)	0.0003	1.44 (1.23-1.70)	<0.0001	过表达
NDC80	1.32 (1.17-1.48)	<0.0001	1.38 (1.18-1.62)	<0.0001	过表达
SPC24	1.54 (1.25-1.88)	<0.0001	1.66 (1.20-2.30)	0.0024	过表达
SPC25	1.32 (1.17-1.48)	<0.0001	1.36 (1.16-1.59)	0.0001	过表达
NUF2	1.32 (1.14-1.54)	0.0003	1.41 (1.13-1.76)	0.0025	过表达

[0172] 在使用TCGA乳腺癌数据进行的共表达相关性网络分析中,也分离了所鉴定的对于无AE生存期和无MR生存期具有最高风险比的22个基因的大多数。在着丝粒处的初生CENP-A的稳定装配需要11个鉴定的基因(CENP-A、HJURP、MIS18A、MIS18B、CENP-C、CENP-N、CENP-H、CENP-I、CENP-M、CENP-K、CENP-L),暗示着丝粒染色质装配在乳腺癌进展中的重要作用。值得注意的是,其他9个CEN/KT基因(CENP-T、-S、-P、-Q、-R、M18BP1、PMF1、MIS12和NSL1)的错误表达没有显示出明显的预后价值,尽管它们也是着丝粒或动粒的重要或必需组分。例如,即使第三组分M18BP1对着丝粒装配也是必需的,但是只有Mis18复合物的MIS18A亚基和MIS18B亚基显示出显著的预后价值。使用基于不同算法的K-M Plotter软件进行的分析鉴定了大部分相同的基因加上几种基因(表4)(Gyorffy等人,2010)。我们推断,即使功能密切相关,不同的CEN/KT基因在癌症病因中可以显示不同的调控和功能。

[0173] 表4.使用K-M plotter,CEN/KT基因对于乳腺癌的预后价值

[0174]

	乳腺癌 OS (n=1115)		乳腺癌 RFS (n=3455)		乳腺癌 DMFS (n=1609)	
	HR (95% CI)	P 值	HR (95% CI)	P 值	HR (95% CI)	P 值
CENP-A	<i>1.89 (1.49-2.42)</i>	<i>1.60E-07</i>	<i>1.97 (1.75-2.22)</i>	<i>0.00E+00</i>	<i>1.87 (1.52-2.3)</i>	<i>1.90E-09</i>
HJURP	<i>1.73 (1.36-2.2)</i>	<i>7.10E-06</i>	<i>1.72 (1.53-1.93)</i>	<i>0.00E+00</i>	<i>1.87 (1.52-2.29)</i>	<i>1.90E-09</i>
M18BP1	1.07 (0.85-1.36)	5.50E-01	<i>1.3 (1.16-1.46)</i>	<i>6.20E-06</i>	<i>1.3 (1.06-1.59)</i>	<i>1.10E-02</i>
MIS18A	<i>1.75 (1.38-2.23)</i>	<i>4.10E-06</i>	<i>1.16 (1.03-1.3)</i>	<i>1.30E-02</i>	<i>1.82 (1.48-2.24)</i>	<i>7.50E-09</i>
MIS18B	<i>1.89 (1.49-2.41)</i>	<i>1.50E-07</i>	<i>1.43 (1.27-1.61)</i>	<i>1.30E-09</i>	<i>1.62 (1.33-1.99)</i>	<i>2.50E-06</i>
CENP-C	0.88 (0.69-1.12)	3.00E-01	0.71 (0.63-0.8)	8.10E-09	0.91 (0.74-1.11)	3.40E-01
CENP-N	<i>1.85 (1.45-2.36)</i>	<i>4.90E-07</i>	<i>1.95 (1.73-2.19)</i>	<i>0.00E+00</i>	<i>1.8 (1.46-2.21)</i>	<i>1.70E-08</i>
CENP-I	<i>1.53 (1.2-1.94)</i>	<i>5.10E-04</i>	<i>1.45 (1.29-1.62)</i>	<i>3.30E-10</i>	<i>1.76 (1.44-2.17)</i>	<i>4.30E-08</i>
CENP-H	NA	NA	NA	NA	NA	NA
CENP-T	0.92 (0.72-1.17)	4.90E-01	0.73 (0.65-0.83)	1.70E-07	1.09 (0.89-1.33)	4.20E-01
CENP-W	NA	NA	NA	NA	NA	NA
CENP-S	HR=1.09 (0.86-1.39)	4.70E-01	1.1 (0.98-1.24)	9.20E-02	0.92 (0.75-1.12)	4.00E-01
CENP-X	1.07 (0.84-1.36)	5.70E-01	<i>1.39 (1.24-1.56)</i>	<i>2.30E-08</i>	<i>1.28 (1.05-1.57)</i>	<i>1.70E-02</i>
CENP-M	<i>1.55 (1.22-1.97)</i>	<i>3.30E-04</i>	<i>1.49 (1.32-1.67)</i>	<i>1.70E-11</i>	<i>1.4 (1.14-1.72)</i>	<i>1.00E-03</i>
CENP-U	<i>1.75 (1.38-2.23)</i>	<i>3.90E-06</i>	<i>1.98 (1.76-2.23)</i>	<i>0.00E+00</i>	<i>1.79 (1.46-2.2)</i>	<i>2.00E-08</i>
CENP-L	NA	NA	NA	NA	NA	NA
CENP-K	NA	NA	NA	NA	NA	NA
CENP-O	0.92 (0.72-1.16)	4.80E-01	0.8 (0.71-0.9)	1.20E-04	0.93 (0.76-1.14)	5.10E-01
CENP-P	NA	NA	NA	NA	NA	NA
CENP-Q	1.05 (0.83-1.33)	7.00E-01	<i>1.14 (1.02-1.28)</i>	<i>2.40E-02</i>	1.14 (0.93-1.39)	2.20E-01
CENP-R	1.06 (0.84-1.35)	6.20E-01	<i>1.17 (1.05-1.32)</i>	<i>6.50E-03</i>	0.94 (0.77-1.15)	5.80E-01
KNL1	0.98 (0.77-1.24)	8.60E-01	0.86 (0.76-0.96)	9.70E-03	0.97 (0.79-1.18)	7.40E-01
ZWINT	<i>1.55 (1.22-1.97)</i>	<i>2.80E-04</i>	<i>1.7 (1.51-1.91)</i>	<i>0.00E+00</i>	<i>1.56 (1.27-1.91)</i>	<i>1.70E-05</i>
MIS12	0.94 (0.74-1.2)	6.40E-01	1.05 (0.93-1.17)	4.40E-01	0.99 (0.81-1.21)	9.10E-01
NSL1	0.94 (0.74-1.19)	5.80E-01	<i>1.2 (1.07-1.35)</i>	<i>2.20E-03</i>	1.09 (0.89-1.34)	3.90E-01
PMF1	0.78 (0.61-0.99)	4.00E-02	0.92 (0.82-1.03)	1.50E-01	0.9 (0.73-1.1)	3.00E-01
KNL3	<i>1.36 (1.07-1.72)</i>	<i>1.20E-02</i>	<i>1.36 (1.21-1.52)</i>	<i>2.30E-07</i>	<i>1.31 (1.07-1.61)</i>	<i>8.00E-03</i>
NDC80	<i>1.51 (1.19-1.91)</i>	<i>6.80E-04</i>	<i>1.87 (1.66-2.11)</i>	<i>0.00E+00</i>	<i>1.82 (1.48-2.24)</i>	<i>6.40E-09</i>
SPC24	NA	NA	NA	NA	NA	NA
SPC25	<i>1.49 (1.18-1.9)</i>	<i>9.40E-04</i>	<i>1.44 (1.28-1.61)</i>	<i>8.30E-10</i>	<i>1.55 (1.26-1.9)</i>	<i>2.10E-05</i>
NUF2	NA	NA	NA	NA	NA	NA

[0175] OS=总体生存期,RFS=无复发生存期,DMFS=无远端转移生存期。对于显著相关性,当HR>1时,HR格是斜体,当HR<1且p<0.05时,HR格是粗体。NA=不适用。NA表示在U133A平台上不包括的基因。对于具有多个探针的基因,显示最敏感的探针和相关数值。

[0176] 此外,使用K-M Plotter,我们研究了来自>1,600名肺癌患者,>350名胃癌患者以及较小数目(n<150)的I期和II期卵巢癌患者的数据中,CEN/KT基因表达对于总体生存期和疾病进展的预后价值(Gyorffy等人,2010)。我们鉴定了20个CEN/KT基因,其错误表达指示肺癌预后(p<0.05)(表5)的,23个CEN/KT基因用于胃癌预后(p<0.05)(表6)以及20个CEN/KT基因用于早期卵巢癌预后(p<0.05)(表7)。用于不同类型癌症的预后的大多数CEN/KT基因重叠(表8)。这些结果提示,许多CENK/T基因表达水平是乳腺癌、肺癌、胃癌和早期卵巢癌预后的有效预测因子。

[0177] 表5.使用K-M plotter,CEN/KT基因对于肺癌的预后价值

[0178]

基因	肺癌 OS (n=1405)		肺癌 FP (n=982)	
	HR (95% CI)	P 值	HR (95% CI)	P 值
CENP-A	<i>1.57 (1.35-1.83)</i>	<i>5.60E-09</i>	<i>1.87 (1.51-2.31)</i>	<i>6.00E-09</i>
HJURP	<i>1.73 (1.49-2.02)</i>	<i>1.20E-12</i>	<i>1.96 (1.61-2.38)</i>	<i>6.70E-12</i>
M18BP1	0.74 (0.63-0.86)	8.40E-05	0.61 (0.49-0.75)	3.80E-06
MIS18A	<i>1.38 (1.18-1.61)</i>	<i>3.20E-05</i>	<i>1.37 (1.11-1.7)</i>	<i>0.0031</i>
MIS18B	<i>1.67 (1.43-1.94)</i>	<i>3.80E-11</i>	<i>1.81 (1.46-2.24)</i>	<i>3.20E-08</i>
CENP-C	0.62 (0.53-0.72)	5.50E-10	0.5 (0.4-0.62)	1.40E-10
CENP-N	<i>1.47 (1.27-1.72)</i>	<i>5.60E-07</i>	<i>1.62 (1.31-2.01)</i>	<i>6.90E-06</i>
CENP-I	<i>1.41 (1.21-1.64)</i>	<i>8.30E-06</i>	<i>1.73 (1.4-2.14)</i>	<i>3.50E-07</i>
CENP-H	NA	NA	NA	NA
CENP-T	<i>1.34 (1.15-1.56)</i>	<i>1.50E-04</i>	<i>1.56 (1.26-1.93)</i>	<i>3.90E-05</i>
CENP-W	NA	NA	NA	NA
CENP-S	0.71 (0.61-0.83)	1.10E-05	0.64 (0.52-0.79)	3.60E-05
CENP-X	<i>1.55 (1.33-1.8)</i>	<i>1.80E-08</i>	<i>1.74 (1.4-2.15)</i>	<i>3.00E-07</i>
CENP-M	<i>1.55 (1.33-1.8)</i>	<i>1.30E-08</i>	<i>1.53 (1.24-1.89)</i>	<i>7.50E-05</i>
CENP-U	<i>1.7 (1.46-1.98)</i>	<i>8.60E-12</i>	<i>1.57 (1.27-1.94)</i>	<i>2.60E-05</i>
CENP-L	NA	NA	NA	NA
CENP-K	NA	NA	NA	NA
CENP-O	<i>1.19 (1.02-1.38)</i>	<i>2.50E-02</i>	<i>1.52 (1.23-1.88)</i>	<i>8.60E-05</i>
CENP-P	NA	NA	NA	NA
CENP-Q	0.82 (0.71-0.96)	1.20E-02	0.86 (0.69-1.08)	2.00E-01
CENP-R	1 (0.86-1.17)	3.60E-01	1.1 (0.9-1.36)	3.50E-01
KNL1	0.93 (0.8-1.08)	0.12	0.77 (0.6-1)	0.05

[0179]

ZWINT	<i>1.5 (1.32-1.71)</i>	<i>3.20E-10</i>	<i>1.52 (1.16-1.99)</i>	<i>2.40E-03</i>
MIS12	0.73 (0.63-0.85)	4.90E-05	0.62 (0.5-0.77)	8.70E-06
NSL1	0.68 (0.59-0.8)	9.40E-07	0.64 (0.52-0.79)	2.80E-05
PMF1	0.86 (0.76-0.97)	1.70E-02	1.01 (0.83-1.22)	9.40E-01
KNL3	1.08 (0.93-1.25)	3.30E-01	1.09 (0.88-1.34)	4.40E-01
NDC80	<i>1.29 (1.11-1.51)</i>	<i>8.90E-04</i>	<i>1.21 (0.98-1.5)</i>	<i>7.10E-02</i>
SPC24	NA	NA	NA	NA
SPC25	<i>1.69 (1.45-1.97)</i>	<i>1.30E-11</i>	<i>1.99 (1.61-2.47)</i>	<i>1.40E-10</i>
NUF2	NA	NA	NA	NA

[0180] OS=总体生存期,RFS=无复发生存期,DMFS=无远端转移生存期。对于显著相关性,当HR>1且p<0.05时,HR格是斜体,当HR<1且p<0.05时,HR格是粗体。NA=不适用。NA指示在U133A平台上不包括的基因。对于具有多个探针的基因,显示最敏感的探针和相关数值。

[0181] 表6. 使用K-M plotter,CEN/KT基因对于胃癌的预后价值

[0182]

基因	胃癌 OS (n=593)		胃癌 FP (n=359)	
	HR (95% CI)	P 值	HR (95% CI)	P 值
CENP-A	1.5 (1.23-1.82)	5.90E-05	1.86 (1.45-2.37)	5.00E-07
HJURP	1.7 (1.36-2.12)	2.60E-06	2.23 (1.74-2.86)	7.20E-11
M18BP1	1.18 (0.95-1.47)	1.40E-01	0.73 (0.56-0.96)	2.50E-02
MIS18A	1.29 (1.04-1.6)	1.90E-02	1.33 (1.04-1.7)	0.024
MIS18B	1.28 (1.05-1.56)	1.50E-02	1.55 (1.18-2.03)	1.50E-03
CENP-C	0.78 (0.64-0.96)	1.80E-02	0.8 (0.63-1.03)	8.40E-02
CENP-N	1.54 (1.26-1.89)	2.90E-05	2.06 (1.53-2.76)	8.80E-07
CENP-I	1.52 (1.25-1.84)	2.20E-05	1.75 (1.36-2.24)	7.70E-06
CENP-H	NA	NA	NA	NA
CENP-T	1.15 (0.95-1.39)	1.60E-01	1.61 (1.25-2.08)	2.00E-04
CENP-W	NA	NA	NA	NA
CENP-S	1.44 (1.16-1.78)	1.00E-03	1.37 (1.07-1.75)	1.20E-02
CENP-X	0.67 (0.54-0.84)	3.00E-04	0.56 (0.43-0.74)	2.70E-05
CENP-M	1.38 (1.1-1.73)	5.60E-03	1.51 (1.18-1.93)	9.50E-04
CENP-U	1.48 (1.19-1.85)	4.40E-04	1.59 (1.22-2.06)	4.60E-04
CENP-L	NA	NA	NA	NA
CENP-K	NA	NA	NA	NA
CENP-O	1.16 (0.94-1.43)	1.60E-01	1.53 (1.2-1.96)	5.90E-04
CENP-P	NA	NA	NA	NA
CENP-Q	1.59 (1.29-1.96)	1.30E-05	1.52 (1.14-2.02)	4.20E-03
CENP-R	1.43 (1.16-1.77)	7.50E-04	1.66 (1.29-2.13)	6.80E-05
KNL1	0.82 (0.66-1.01)	6.60E-02	0.8 (0.61-1.05)	1.10E-01
ZWINT	1.39 (1.13-1.72)	1.80E-03	1.56 (1.22-2)	3.00E-04

[0183]

MIS12	1.28 (1.04-1.57)	1.80E-02	1.33 (1.02-1.74)	3.70E-02
NSL1	0.88 (0.71-1.08)	2.20E-01	0.74 (0.57-0.95)	1.70E-02
PMF1	0.82 (0.67-1)	5.20E-02	0.71 (0.55-0.91)	6.70E-03
KNL3	1.24 (1.02-1.5)	3.00E-02	1.35 (1.05-1.75)	2.10E-02
NDC80	1.54 (1.26-1.89)	2.10E-05	1.88 (1.44-2.45)	2.00E-06
SPC24	NA	NA	NA	NA
SPC25	1.45 (1.19-1.77)	2.10E-04	1.94 (1.45-2.58)	4.60E-06
NUF2	NA	NA	NA	NA

[0184] OS=总体生存期,FP=首次进展。NA指示,对应于该基因的探针没有通过限定过滤器。对于显著相关性,当 $HR > 1$ 且 $p < 0.05$ 时,HR格是斜体,当 $HR < 1$ 且 $p < 0.05$ 时,HR格是粗体。

[0185] 表7.使用K-M plotter,CEN/KT基因对于第一阶段和第二阶段合并的卵巢癌的预后价值

[0186]

基因	卵巢癌 OS (n=133)		卵巢癌 PFS (n=126)	
	HR (95% CI)	P 值	HR (95% CI)	P 值
CENP-A	9.2 (1.24-68.13)	8.20E-03	17.24 (2.37-125.29)	1.10E-04
HJURP	5.17 (1.54-17.35)	3.00E-03	3.12 (1.7-5.73)	1.10E-04
M18BP1	3.65 (1.37-9.75)	5.60E-03	3.03 (1.28-7.16)	8.10E-03
MIS18A	7.59 (1.03-56.14)	1.90E-02	3.24 (1.71-6.13)	1.30E-04
MIS18B	4.78 (1.13-20.28)	1.90E-02	4.63 (1.65-12.95)	1.30E-03
CENP-C	0.43 (0.19-0.96)	3.40E-02	3.34 (1.84-6.06)	2.70E-05
CENP-N	5.17 (1.54-17.29)	3.00E-03	2.36 (1-5.6)	4.40E-02
CENP-I	3.85 (1.15-12.88)	1.80E-02	4.13 (1.63-10.51)	1.20E-03
CENP-H	NA	NA	NA	NA
CENP-T	1.49 (0.66-3.38)	3.40E-01	1.35 (0.71-2.55)	3.60E-01
CENP-W	NA	NA	NA	NA
CENP-S	1.49 (0.66-3.33)	3.30E-01	1.81 (0.99-3.3)	4.90E-02
CENP-X	1.84 (0.83-4.11)	1.30E-01	0.49 (0.25-0.95)	3.10E-02
CENP-M	4.36 (1.3-14.59)	9.20E-03	2.96 (1.37-6.38)	3.80E-03
CENP-U	4.94 (1.16-20.97)	1.60E-02	1.93 (1.06-3.5)	2.90E-02
CENP-L	NA	NA	NA	NA
CENP-K	NA	NA	NA	NA
CENP-O	1.97 (0.89-4.35)	8.80E-02	0.73 (0.39-1.37)	3.30E-01
CENP-P	NA	NA	NA	NA
CENP-Q	1.92 (0.86-4.28)	1.10E-01	1.88 (1.02-3.46)	3.80E-02
CENP-R	0.38 (0.13-1.11)	6.60E-02	1.33 (0.74-2.41)	3.40E-01
KNL1	1.66 (0.69-3.97)	2.50E-01	0.47 (0.23-0.93)	2.60E-02
ZWINT	2.6 (1.14-5.9)	1.80E-02	4.03 (2.02-8)	1.80E-05
MIS12	3.79 (1.13-12.71)	2.00E-02	1.83 (0.98-3.42)	5.40E-02

[0187]

NSL1	3.21 (1.27-8.07)	9.10E-03	3.83 (2.07-7.08)	4.60E-06
PMF1	3.36 (1.44-7.84)	2.90E-03	2.86 (1.57-5.19)	3.20E-04
KNL3	4.16 (1.53-11.29)	2.60E-03	3.92 (2-7.68)	2.00E-05
NDC80	5.15 (2.03-13.09)	1.40E-04	3.94 (1.99-7.78)	2.40E-05
SPC24	NA	NA	NA	NA
SPC25	2.98 (1.24-7.16)	1.00E-02	4.17 (1.63-10.63)	1.20E-03
NUF2	NA	NA	NA	NA

[0188] OS=总体生存期,PFS=无进展生存期.NA指示,对应于该基因的探针没有通过限定过滤器。对于显著相关性,当HR>1且p<0.05时,HR格是斜体,当HR<1且p<0.05时,HR格是粗体。

[0189] 表8.来自上述表格的不同基因列表的比较

[0190]

基因	乳腺癌 (GenExMiner)	乳腺癌 (K-M plotter)	肺癌 (K-M plotter)	胃癌 (K-M plotter)	差别	CES 基因
CENP-A	x	x	x	x	x	是
HJURP	x	x	x	x	x	是
M18BP1		x	#	#		
MIS18A	x	x	x	x		
MIS18B	x	x	x	x	x	是
CENP-C	#	#	#	#		
CENP-N	x	x	x	x	x	是
CENP-I	x	x	x	x		
CENP-H	x	NA	NA	NA		
CENP-T		x	x	x		
CENP-W	x	NA	NA	NA	x	是
CENP-S			#	x		
CENP-X	x	x	x	#		
CENP-M	x	x	x	x	x	是
CENP-U	x	x	x	x	x	是
CENP-L	x	NA	NA	NA	x	是
CENP-K	x	NA	NA	NA	x	是
CENP-O	x	#	x	x		
CENP-P		NA	NA	NA		
CENP-Q		x	#	x		
CENP-R		#		x		
KNL1	x	x			x	
ZWINT	x	x	x	x	x	是
MIS12			#	x		

[0191]

NSL1		x	#	#		
PMF1		#	#	#		
KNL3	x	x		x		
NDC80	x	x	x	x	x	是
SPC24	x	NA	NA	NA	x	是
SPC25	x	x	x	x	x	是
NUF2	x	NA	NA	NA	x	是

[0192] NA=不适用。“X”指示过表达。“#”指示降低的表达。

[0193] C. 着丝粒和动粒基因表达评分 (CES) 指示许多癌症的基因组不稳定性水平

[0194] 我们比较了使用不同方法获得的基因列表,并鉴定了共有的14个CEN/KT基因(图2和表8)。共有的基因列表包含参与CENP-A装配的7种基因和NDC80复合物的所有4个组分,但不含MIS12复合物成员。基于重叠,我们认为这14个CEN/KT基因的调控异常是为多种癌症类型的癌症进展和患者结果提供有效预后特征的核心子集(表9和图2)。由于列表中的个体基因在癌症样本中显示出过表达,我们认为通路的总体调控异常和程度可以由个体基因表达水平的总和而非任何单个基因来表示。因此,我们将给定样本中的着丝粒和动粒基因表达评分 (CES) 定义为14个CEN/KT基因的 $\log_2$ (mRNA水平) 的总和。

[0195] 表9.CES基因表

[0196]

CES 基因	
CENP-A	着丝粒
HJURP	
MIS18B	Mis18
CENP-N	内部动粒
CENP-M	
CENP-W	
CENP-U	
CENP-L	
CENP-K	外部动粒
ZWINT	
NDC80	
NUF2	
SPC24	
SPC25	

[0197] 适当的着丝粒和动粒功能确保忠实的染色体分离和基因组稳定性。我们研究了在癌症样本中CEN/KT基因的调控异常是否与基因组不稳定性的程度相关。将TCGA数据集用于比较CES值与具有CNA的癌症基因组的分数和突变频率(表10)。对于许多癌症类型(包括乳腺癌、低级脑部胶质瘤和胃腺癌),我们检测到了CES值与CNA分数和突变频率两者之间的显著正相关性(斯皮尔曼 ρ , $p < 0.05$, 学生t检验)。对于一些癌症,如肺腺癌、肾上腺皮质癌、结肠直肠癌、肾脏肾透明细胞(RCC)癌、肾脏肾乳头状细胞(RPC)癌和卵巢浆液性囊腺癌,我们仅检测到CES值与CAN分数或突变频率之间的显著相关性。我们观察到,对于许多癌症,没有相关性,如宫颈鳞状细胞癌、胶质母细胞瘤和甲状腺癌。总体而言,我们推断,TCGA数据集中CES值与多种癌症类型的基因组不稳定性水平之间存在显著相关性。

[0198] 表10.TCGA数据集中CES值与突变频率和拷贝数变化之间的斯皮尔曼相关系数

[0199]

癌症类型	突变频率			拷贝数变化		
	斯皮尔曼 ρ	P 值	N	斯皮尔曼 ρ	P 值	N
肾上腺皮质癌	0.468	3.34E-06	72	0.046	0.703	72
膀胱尿路上皮癌	0.240	6.23E-03	129	0.244	3.36E-08	353
低级胶质瘤	0.323	4.75E-08	273	0.483	1.51E-30	499
乳腺腺癌	0.443	3.69E-48	975	0.539	3.40E-82	1076
宫颈鳞状细胞癌和子宫颈内腺癌 (Endocervical Adenocarcinoma)	0.095	0.191	191	0.091	0.155	244
结肠直肠癌(colectoral carcinoma)	0.145	0.05	182	-0.031	0.674	182
胶质母细胞瘤	0.090	0.28	147	0.128	0.12	148
头颈 SCC	0.124	0.031	304	0.208	3.15E-06	494
肾脏 RCC	0.036	0.466	410	0.179	4.70E-05	513
肾脏 RPC	0.049	0.542	121	0.182	2.35E-03	276
肺 ADC	0.338	6.33E-06	171	0.290	7.51E-11	485
肺 SCC	0.236	1.53E-03	178	0.509	3.56E-34	498
卵巢浆液性囊腺癌	0.071	0.368	161	0.133	0.031	262
前列腺腺癌	0.349	1.60E-08	248	0.487	1.38E-29	474
皮肤皮肤黑色素瘤	0.196	2.73E-04	339	0.066	0.152	468
胃腺癌	0.386	2.04E-08	198	0.377	2.76E-10	262
甲状腺癌	0.056	0.273	391	0.036	0.427	489
子宫肉瘤	0.432	8.97E-04	56	0.263	0.05	56

[0200] 用灰色突出显示斯皮尔曼系数相关性的显著双尾p值 ($p < 0.05$)。

[0201] D.CES是用于癌症患者总体生存期、复发和远端转移的预后值

[0202] 使用乳腺癌和肺癌微阵列数据集,我们评估了CES系统对于癌症患者生存期的预后。通过将数据集的完整CES范围分为三分位数,我们将每个测试GEO数据集的患者分为高CES组、中CES组和低CES组。卡普兰-迈耶存活曲线表明,对于乳腺癌和肺癌数据集,CES系统有效地预测了总体生存期(OS)和无病生存期(DFS)。通常,具有较低CES值的患者对OS(图3A)和DFS(图3B)两者具有显著更好的预后。值得注意的是,对于肺癌数据集GSE19188,高CES组和低CES组的患者均比具有中间CES的那些患者显示出显著更好的DFS,这可能是由于高CES患者对辅助治疗的更好的应答(Botling等人,2013)(见下文)。我们还使用数据集GSE6532(Loi,2007)确定了CES对于乳腺癌无远端转移生存期(DMFS)的预后价值。卡普兰-迈耶曲线显示,高CES的患者比具有低CES的那些患者具有显著更差的DMFS($p = 1.21E-05$,图3C)。

[0203] 为了更好地理解CEN/KT基因调控异常对癌症进展的影响,我们分析了含有临床病理学信息乳腺癌数据。ER-肿瘤倾向于更具侵袭性,并且阳性淋巴结指示癌症侵袭。我们发现,高CES值与晚期肿瘤等级和阴性ER状态显著相关($p < 0.05$),但与患者年龄、肿瘤大小、淋巴结状态或肿瘤阶段没有显著相关性(数据未显示)。此外,不管淋巴结状态,对于ER+乳腺癌患者,我们发现了许多CEN/KT基因的调控异常与不良无AE生存期和无MR生存期之间的显著相关性(表11, $p < 0.05$)。这些分析提示,CEN/KT基因的错误表达可以预测病理分类的乳腺癌的患者生存期和疾病进展。

[0204] 表11.单个CEN/KT基因对于具有临床病理学信息的乳腺癌的预后价值

基因名称	p 值无 AE 生存期				P 值无 MR 生存期			
	N+, ER+	N-, ER+	N+, ER-	N-, ER-	N+, ER+	N-, ER+	N+, ER-	N-, ER-
[0205] CENP-A	0.0407	<0.0001	0.247	0.245	0.1092	<0.0001	0.0423	0.196
HJURP	0.0001	<0.0001	0.7176	0.534	0.0022	<0.0001	0.3109	0.676
M18BP1	0.8638	0.8429	0.0822	0.523	0.9256	0.5233	0.1423	0.695
[0206]	MIS18A	0.0246	0.2008	0.0823	0.103	0.0899	0.0343	0.2567
	MIS18B	0.0173	0.0001	0.07	0.334	0.0486	0.0003	0.0151
	CENP-C	0.6629	0.0966	0.2412	0.346	0.4351	0.0036	0.4788
	CENP-N	0.0601	<0.0001	0.5045	0.441	0.4172	<0.0001	0.2694
	CENP-I	0.0029	0.0002	0.4031	0.881	0.0465	0.0034	0.6094
	CENP-H	0.0417	0.7757	0.87	0.751	0.1094	0.1989	0.6952
	CENP-T	0.5857	0.1209	0.0525	0.748	0.2311	0.2491	0.2937
	CENP-W	0.2808	0.0054	0.1119	0.929	0.0779	0.0045	0.0668
	CENP-S	0.7168	0.1122	0.6003	0.499	0.4346	0.3601	0.6644
	CENP-X	0.0502	0.0036	0.9954	0.527	0.0155	0.0421	0.2276
	CENP-M	0.0077	0.0003	0.6798	0.551	0.0071	0.0005	0.322
	CENP-U	0.1553	0.0007	0.9789	0.941	0.0134	0.0001	0.7316
	CENP-L	0.0179	0.403	0.8631	0.11	0.1092	0.5438	0.7616
	CENP-K	0.6084	0.1594	0.5005	0.645	0.6674	0.5904	0.3571
	CENP-O	0.0059	0.1174	0.5991	0.452	0.0426	0.5381	0.2694
	CENP-P	0.8499	0.4767	0.2822	0.246	0.8073	0.1364	0.6465
	CENP-Q	0.4007	0.8576	0.1232	0.782	0.4123	0.2321	0.3785
	CENP-R	0.6737	0.6343	0.5745	0.99	0.2943	0.4	0.1213
	KNL1	0.0332	0.2529	0.5922	0.336	0.0035	0.3585	0.4865
	MIS12	0.2789	0.0833	0.386	0.475	0.0949	0.3634	0.9375
	NSL1	0.3705	0.2677	0.445	0.046	0.7039	0.4334	0.581
	KNL3	0.0457	0.0044	0.1049	0.973	0.0031	0.0036	0.9868
	NDC80	0.0967	0.0005	0.4815	0.817	0.0873	0.0072	0.8999
	SPC24	0.1865	0.0016	0.2107	0.726	0.0753	0.8263	0.9996
	SPC25	0.0193	<0.0001	0.2205	0.76	0.0551	0.0003	0.3055
	NUF2	0.0409	0.1901	0.3714	0.957	0.0767	0.5886	0.5009

[0207] 基于基因表达特征,还可以将乳腺癌分为亚型(Parker等人,2009)。基底样亚型、HER2+亚型(通常均为ER-)和管腔B(ER+)亚型肿瘤比管腔A型(ER+)和正常样的肿瘤更具侵袭性。我们观察到,更具侵袭性的分子亚型与较高的平均CES值相关($p < 0.05$)。使用BC GenExMiner软件,我们发现与正常样和管腔A型肿瘤相比,17个CEN/KT基因在基底样乳腺癌、HER2+乳腺癌和管腔B型乳腺癌中表现出更高的表达,其中数个在正常乳腺样亚型肿瘤、管腔A型亚型肿瘤和管腔B型亚型肿瘤中显示出与不良无AE生存期的显著相关性($p < 0.05$) (表12)。这些结果指示,在不同的乳腺癌亚型中,CEN/KT的调控异常是不同的。

[0208]

表 12. 单个 CEN/KT 基因对于具有亚型信息的乳腺癌的预后价值

BC 亚型	基底样		HER2+		管腔 A 型		管腔 B 型		正常乳腺样	
	无 AE 生存期 HR (95%CI)	P 值	无 AE 生存期 HR (95%CI)	P 值	无 AE 生存期 HR (95%CI)	P 值	无 AE 生存期 HR (95%CI)	P 值	无 AE 生存期 HR (95%CI)	P 值
CENP-A	0.98 (0.73-1.32)	0.9081	.89 (0.66-1.19)	0.4214	1.24 (0.93-1.65)	0.1461	1.40 (1.09-1.80)	0.0075	1.59 (1.06-2.39)	0.026
HJURP	1.01 (0.76-1.33)	0.9705	0.84 (0.61-1.14)	0.2553	1.37 (1.03-1.83)	0.0323	1.41 (1.10-1.81)	0.0073	1.96 (1.32-2.90)	0.0008
M18BP1	0.93 (0.71-1.23)	0.6235	0.86 (0.63-1.17)	0.3444	1.27 (0.96-1.67)	0.0978	0.66 (0.51-0.85)	0.0015	0.96 (0.66-1.39)	0.8147
MIS18A	1.14 (0.86-1.52)	0.3598	1.09 (0.80-1.50)	0.5719	1.12 (0.84-1.48)	0.4436	1.19 (0.93-1.53)	0.161	1.15 (0.79-1.69)	0.4585
MIS18B	1.11 (0.84-1.48)	0.4577	0.94 (0.70-1.28)	0.7108	1.38 (1.03-1.85)	0.0289	1.35 (1.06-1.73)	0.0159	1.13 (0.76-1.68)	0.5476
CENP-T	1.25 (0.92-1.70)	0.1539	0.98 (0.70-1.37)	0.9037	0.79 (0.59-1.06)	0.1166	0.86 (0.65-1.13)	0.2667	0.91 (0.60-1.36)	0.635
CENP-W	0.68 (0.33-1.42)	0.3088	1.40 (0.67-2.90)	0.3678	0.98 (0.40-2.38)	0.9667	1.15 (0.63-2.09)	0.6425	1.09 (0.40-2.97)	0.8626
CENP-S	0.80 (0.57-1.13)	0.2034	1.16 (0.80-1.66)	0.4354	1.13 (0.82-1.55)	0.4668	0.95 (0.71-1.27)	0.7359	0.76 (0.46-1.26)	0.2829
CENP-X	1.03 (0.77-1.36)	0.8553	0.91 (0.68-1.23)	0.5469	1.10 (0.83-1.46)	0.512	1.25 (0.98-1.60)	0.0762	1.12 (0.77-1.64)	0.5596
CENP-C	1.24 (0.95-1.64)	0.1179	0.88 (0.65-1.18)	0.3905	0.89 (0.67-1.18)	0.4287	0.79 (0.62-1.01)	0.0636	1.11 (0.77-1.59)	0.5851
CENP-N	0.90 (0.67-1.20)	0.4716	1.04 (0.76-1.43)	0.8006	1.26 (0.94-1.69)	0.1267	1.48 (1.15-1.92)	0.0027	1.00 (0.67-1.47)	0.9804
CENP-M	1.00 (0.75-1.33)	0.991	1.18 (0.86-1.61)	0.3101	1.11 (0.83-1.49)	0.4707	1.37 (1.06-1.76)	0.015	1.26 (0.85-1.88)	0.2473
CENP-U	1.03 (0.77-1.36)	0.8564	0.99 (0.72-1.36)	0.941	1.15 (0.87-1.53)	0.3226	1.22 (0.95-1.58)	0.1213	1.28 (0.87-1.90)	0.2132
CENP-H	1.13 (0.73-1.76)	0.5861	0.98 (0.65-1.47)	0.9058	1.38 (0.93-2.04)	0.1082	1.03 (0.73-1.46)	0.8652	1.27 (0.79-2.03)	0.3308
CENP-I	1.07 (0.81-1.41)	0.6476	1.24 (0.92-1.65)	0.1538	1.54 (1.16-2.03)	0.0028	1.16 (0.91-1.48)	0.2249	1.30 (0.89-1.89)	0.171
CENP-L	1.05 (0.71-1.56)	0.7951	1.01 (0.68-1.50)	0.9466	1.52 (1.04-2.22)	0.0299	1.23 (0.89-1.71)	0.2055	1.39 (0.83-2.32)	0.2123
CENP-K	1.30 (0.86-1.95)	0.2097	0.96 (0.65-1.41)	0.8263	1.53 (1.05-2.24)	0.0276	0.94 (0.69-1.28)	0.6992	1.21 (0.77-1.88)	0.4082
CENP-O	0.94 (0.71-1.25)	0.6862	1.08 (0.78-1.48)	0.655	1.01 (0.76-1.34)	0.9583	1.35 (1.05-1.74)	0.0175	1.41 (0.96-2.06)	0.0777
CENP-P	0.83 (0.51-1.36)	0.4675	0.93 (0.60-1.45)	0.7648	1.09 (0.68-1.73)	0.7246	0.93 (0.65-1.33)	0.6902	1.58 (0.86-2.88)	0.1387
CENP-Q	0.80 (0.60-1.07)	0.1273	0.81 (0.59-1.10)	0.1803	1.01 (0.76-1.34)	0.9339	0.95 (0.74-1.22)	0.6922	0.83 (0.56-1.21)	0.3252
CENP-R	0.95 (0.73-1.24)	0.6977	0.87 (0.65-1.16)	0.3527	1.11 (0.85-1.46)	0.4483	0.88 (0.69-1.12)	0.2892	0.74 (0.50-1.07)	0.1097
KNL1	0.77 (0.58-1.04)	0.0896	0.72 (0.52-1.00)	0.0508	1.25 (0.93-1.68)	0.1357	0.95 (0.74-1.23)	0.7076	1.13 (0.76-1.68)	0.5546
MIS12	0.96 (0.72-1.29)	0.8007	1.25 (0.89-1.76)	0.1938	1.22 (0.91-1.64)	0.1814	1.02 (0.78-1.33)	0.904	1.21 (0.81-1.80)	0.3432
NSL1	1.12 (0.86-1.47)	0.4023	0.94 (0.70-1.26)	0.6779	1.02 (0.78-1.34)	0.8801	0.98 (0.77-1.24)	0.8407	0.91 (0.63-1.31)	0.6015

[0209]

KNL3	0.96 (0.73-1.27)	0.7981	0.82 (0.60-1.12)	0.2116	1.21 (0.91-1.61)	0.1895	1.16 (0.90-1.49)	0.2504	0.94 (0.63-1.39)	0.7496
NDC80	0.93 (0.70-1.22)	0.591	1.15 (0.85-1.56)	0.355	1.04 (0.79-1.37)	0.7725	0.98 (0.77-1.25)	0.8681	1.13 (0.77-1.65)	0.5474
SPC24	0.81 (0.45-1.48)	0.4995	1.20 (0.74-1.93)	0.4646	1.63 (1.00-2.63)	0.0478	1.25 (0.85-1.84)	0.2578	1.86 (1.01-3.42)	0.0466
SPC25	1.07 (0.81-1.40)	0.6358	0.92 (0.68-1.24)	0.5797	0.97 (0.73-1.29)	0.8357	1.22 (0.96-1.56)	0.1115	1.49 (1.00-2.22)	0.051
NUF2	0.84 (0.57-1.23)	0.3732	0.95 (0.66-1.39)	0.8041	1.45 (1.00-2.11)	0.0507	1.24 (0.91-1.69)	0.1728	1.18 (0.73-1.91)	0.4893

该表显示了各 CEN/KT 基因对于不同乳腺癌亚型的无 AE 生存期的风险比(HR)和相关 p 值。对于乳腺癌亚型具有显著预后价值(p<0.05)的 HR 和 p 值以粗体显示。

[0210] 最后,为了确认和扩展观察到的CES系统的预后价值,我们使用K-M Plotter对乳腺癌、肺癌、胃癌和卵巢癌患者进行了荟萃分析(图3D-3F)。对于乳腺癌,我们关注ER+固有

亚型。高CES值有效地预测了管腔A型和管腔B型肿瘤的不良患者生存期。在基底样和HER2+亚型中,我们观察到CES值与OS和DMFS之间的相反相关性,这可能是由于高CES肿瘤对治疗的敏感性和/或非常高水平的CIN对癌细胞适合度的有害影响(参见下面)。另外,荟萃分析显示CES系统有效预测肺癌、胃癌和早期卵巢癌患者的患者生存期(图3D和3E)。值得注意的是,我们的结果提示,CES系统预测第1阶段肺腺癌的患者生存期。总之,分析提示,CES可用于预测多种癌症类型的患者生存期、复发和转移。

[0211] E. CES预测对DNA损伤剂的敏感性以及患者对辅助化疗疗法的应答

[0212] 高CES值和基因组不稳定性之间的强关联预测高CES癌细胞处于遗传毒性应激下,并且因此将对额外的DNA损伤更敏感。我们挖掘了癌细胞系百科全书(CCLE)数据,其包含关于药物治疗和基因表达两者的信息(Barretina等人,2012)。根据CCLE细胞系的CES值,将其划分成四分位数,我们比较了Topo I抑制剂和DNA损伤剂伊立替康和拓扑替康的IC₅₀的最高和最低CES四分位数。值得注意的是,在最高CES四分位数中的癌细胞系比在最低四分位数中的癌细胞系显示出显著更低的IC₅₀(学生t检验,分别地,对于伊立替康 $p=2.19E-8$,对于拓扑替康 $p=1.71E-06$)(图4A、5A和5B)。一致地,Topo I抑制剂的细胞系CES值和IC₅₀与强显著性负相关(斯皮尔曼 ρ ,分别地,对于伊立替康 $r=-0.3836$, $p<0.000001$,以及对于拓扑替康, $r=-0.3394$, $p<0.000001$)(表12)。此外,在根据细胞系的组织起源将其破碎之后,对于几种癌症类型,我们检测到细胞系CES与伊立替康和拓扑替康的IC₅₀之间的显著相关性。我们推断,与低CES癌细胞系相比,高CES癌细胞系对Topo I抑制剂更敏感。

[0213] 表13. CCLE细胞系中伊立替康和拓扑替康的IC₅₀与CES值之间的相关性

CCLE 细胞系	伊立替康			拓扑替康		
	斯皮尔曼 ρ	学生 t (p 值)	n	斯皮尔曼 ρ	学生 t (p 值)	n
乳腺	-0.6441	0.007	16	-0.2468	0.281	21
肺	-0.4186	0.005	44	-0.4252	0.0001	77
卵巢	-0.6226	0.003	20	-0.4685	0.018	25
造血细胞系和淋巴	-0.3439	0.017	48	-0.092	0.468	65
皮肤	0.2	0.327	26	-0.1634	0.364	33
CNS	-0.2034	0.436	17	-0.1255	0.568	23
胰腺	-0.4387	0.078	17	-0.1429	0.536	21
总池(overall pool)	-0.3836	<.000001	277	-0.3394	<.000001	412

[0215] 然后我们使用JBR.10临床试验数据集(GSE14814)来确定具有高CES值的早期非小细胞肺癌(NSCLC)对辅助化疗是否更敏感(图4B)。在这项前瞻性临床试验中,为手术后早期NSCLC患者随机分配辅助化疗(ACT),其包括引起DNA损伤和促进凋亡的顺铂,以及通过与微管蛋白相互作用抑制有丝分裂的长春瑞滨,或者无治疗(OBS)(Zhu等人,2010)。我们将患者分为高CES(最高三分位数)组和低CES(较低的两个三分位数)组。对于OBS组,与低CES($HR=2.728$, $p=0.017$)相比,高CES预测不良OS,验证了CES系统的预后能力(图6A)。ACT有效地消除了与高CES($HR=0.710$, $p=0.402$)相关的不良作用,提示CES系统对化疗敏感性的预测能力(图6B)。实际上,与OBS相比,ACT显著改善了高CES患者的总体生存期($HR=0.391$,对数秩 $p=0.035$)(图4B,左)。该作用对于高CES组是特异的,因为对于低CES组没有与ACT相关的显著益处($HR=1.318$,对数秩 $p=0.431$)(图4B,右)。我们还检查了UT肺SPORE NSCLC数据集(GSE41274)(Tang等人,2013)(图4C)。将这一数据集中的手术后患者用包括顺铂且主要是

紫杉烷的ACT进行治疗,或者未接受治疗(OBS)。我们观察到ACT对于高CES患者的类似趋势,但无统计显著性($HR=0.233$,对数秩 $p=0.110$),可能是由于小样本大小和相对短的ACT患者的随访时间(图6B)。因此,我们联合两个数据集进行了荟萃分析(图4C),并且我们的结果显示,使用荟萃数据集,ACT显著改善与早期NSCLC中高CES相关的不良OS($HR=0.432$, $p=0.016$)图4C,左)。再次,这一能力是高CES组特异的,因为ACT对于低CES组没有显示这样的影响($HR=1.075$, $p=0.783$) (图4C,右)。重要的是,荟萃分析显示,ACT特异地改善了高CES组的5年存活率(ACT的81.5%对比OBS的47.3%, $p=0.002$),但没有改善低CES组的5年存活率(ACT的74.4%对比OBS的68.4%, $p=0.347$)。总之,我们的结果提示,CES系统在预测患者对辅助化疗(包括针对NSCLC的顺铂)的敏感性上是有用的。

[0216] 为了处理在预测药物敏感性时个体数据集中的小样本大小,我们使用K-M Plotter对几种癌症类型的化疗敏感性进行了荟萃分析。对于NSCLC,CES在OS和第一进展的风险(FP)两者的化疗敏感性中显示出预测价值(图7A和7B)。对于乳腺癌,我们关注ER+肿瘤,因为许多患者在内分泌治疗后有时由于不清楚的原因而遭受复发。对于患有ER+肿瘤的系统性未治疗的患者,如所预期的,高CES值与显著不良无复发生存期(RFS)相关($HR=1.83$, $p=4.3E-05$) (图7C)。使用或不用激素疗法($HR=1.14$, $p=0.67$),但不仅使用他莫昔芬($HR=1.80$, $p=4E-11$)的化疗,有效地降低了与高CES相关的风险比,意味着CIN有助于激素治疗之后ER+乳腺癌患者的复发。对于胃癌,没有检测到对5-FU的敏感性的显著预测能力(图7D)。对于卵巢癌,大多数患者患有晚期肿瘤,对于其,CES不能显著预测患者的OS($HR=1.09$, $p=0.29$)。然而,铂和拓扑替康均降低了与高CES相关的危险(图7E),并且拓扑替康($HR=0.63$, $p=0.029$)显示出相对于铂的一些优越性($HR=0.85$, $p=0.075$)。我们注意到,在CCLE卵巢癌细胞系中,高CES值也与增加的拓扑替康敏感性相关(斯皮尔曼 ρ , $r=-0.469$, $p=0.018$) (表13)。这些结果提示,CES系统可用于预测对于多种癌症类型,患者对辅助化疗的敏感性。

[0217] F. CES预测乳腺癌和肺癌中患者对放疗的应答

[0218] 根据药物敏感性的结果,我们推断如果着丝粒调控异常是不同癌症类型中共有的重要特性,CES系统也可以预测损伤DNA并引起细胞凋亡的放疗(RT)。我们确定CES是否与RT之后癌症患者的结果相关。使用乳腺癌数据集,我们发现当进行RT时,与无RT相比,具有高CES值的患者显示改善的总体生存期($HR=0.279$, $p=0.008$)和无病生存期($HR=0.254$, $p=0.016$)。相比之下,对于低CES组中的患者,没有与RT相关的生存益处(对于OS, $HR=1.309$, $p=0.58$;以及对于DFS, $HR=0.950$, $p=0.98$) (图4D)。对于具有中间CES值的患者组,我们观察到中间HR,尽管相关性不显著(分别地,对于OS, $HR=0.370$, $p=0.085$;以及对于DFS, $HR=0.389$, $p=0.16$)。进一步的分析提示,RT和较好预后之间的关联对具有高CES的患者组是特异性的(图8A-8D)。然后我们使用K-M Plotter对肺癌数据进行荟萃分析,结果与乳腺癌中的结果一致(图7A和7B)。我们的分析提示,CES系统有效地预测癌症患者对RT以及化疗的敏感性。

[0219] 讨论

[0220] 在这项研究中,使用关注调控染色体分离的特定细胞结构的假说驱动的方法,我们鉴定了14个CEN/KT基因的子集,对于大范围人类癌症,其过表达是共有的特性并且是其预后值(表9)。我们将CES定义为14种基因mRNA表达水平的总和。我们显示,在数种癌症类型

中,CES值是总体生存期、癌症复发和疾病进展的预后值。它也是乳腺癌管腔亚型和早期NSCLC的预后值。此外,CES预测在癌细胞系和患者数据集中,对DNA损伤化合物的敏感性,以及患者对放疗的敏感性。重要的是,对于多种癌症类型,CES值与CNA水平和突变频率显著相关。这一结果与细胞系和模式生物体中的大量基础研究一致,表明着丝粒和动粒结构的调控异常引起染色体错误分离和重排并且促成CIN (Allshire和Karpen,2008;Black等人,2010)。例如,包括CENP-A和HJURP的CES基因的调控异常引起有缺陷的染色体分离。这种严重缺陷可能引起染色体桥和微核 (Mishra等人,2011),其转而导致结构畸变和DNA损伤 (Janssen等人,2011)。这一想法得到以下事实的进一步支持:50%的CES基因参与CENP-A核小体的装配,CENP-A核小体是有丝分裂中着丝粒功能和增殖以及动粒形成的结构基础。这些蛋白质共同包含通过每个细胞周期补充 (replenish) 着丝粒CENP-A核小体的表观遗传通路。令人惊讶的是,必需的MIS12复合物成员均未通过我们关于包含于CES基因列表的显著性阈值。结果提示,着丝粒和动粒的调控异常,特别是着丝粒补充的过程是许多癌症中CIN和基因组不稳定性的关键机制。基因组缺陷一方面促进癌症进展,但是另一方面也似乎是当通过某一阈值时的负担,如通过对进一步DNA损伤的敏感性所证明的。

[0221] 我们已经证明,CES系统代表新的生物标志物,其有效预测许多癌症的基因组不稳定性水平、患者生存期和DNA损伤疗法的影响。我们的研究结果指示,同一癌症类型中个体肿瘤之间的变化不仅预测癌症患者的生存期和疾病进展,而且还预测癌细胞系中对几种DNA损伤剂的敏感性,辅助放疗或者辅助化疗 (包括诸如顺铂的特异性DNA损伤化合物) 之后的患者结果。在RT或化疗后改善的患者生存期与高CES值的特异性关联表明CES系统可用于鉴定处于复发和疾病进展的高风险的患者,并且所述患者可能有利地响应具体治疗。

[0222] 如CES在乳腺癌、肺癌、胃癌和早期卵巢癌中的预后能力所支持的,基因组维护中对着丝粒和动粒的普遍要求意味着其应用可以超越仅仅一种类型的癌症。另外,在乳腺和其他组织的癌前期病变中观察到了显著的染色体异常和CIN (Pihan等人,2003)。重要的是,CEN/KT基因在乳腺DCIS的一部分中错误表达,早期肝癌显示较高的CES,而非癌性发育不良不显示较高的CES,并且与良性肿瘤和非转移癌相比,CEN/KT基因仅在转移性前列腺癌中过表达。这些观察结果表明,CES系统可用于早期诊断,并且有助于避免与过度诊断相关的问题。

[0223] 除了癌症管理的潜在应用之外,我们的结果具有与癌症进展和基因组不稳定性相关的重要意义。成为本研究 and CES系统基础的主要假说是CEN/KT基因的错误表达损害着丝粒和动粒功能,导致染色体异常和遗传毒性应激增加。与该假说一致,在来自多种癌症的大量临床样本中,高CES值与高水平的基因组不稳定性相关 (表2)。这一假说进一步得到了机理研究的支持。着丝粒和动粒组分的扰动引起增加基因组不稳定性的异常的功能,如有缺陷的有丝分裂检查点、纺锤体附着、染色体集合和姐妹染色单体分离。果蝇CENP-A的过表达或将人HJURP系链到非着丝粒区域导致新的着丝粒和结构重排 (Barnhart等人,2011;Heun等人,2006;Mendiburo等人,2011)。另一方面,过量着丝粒蛋白的消耗以及可能过量着丝粒蛋白的存在,可以引起内源性着丝粒功能障碍,完整染色体非整倍性或通过检查点激活引起有丝分裂阻滞 (Blower和Karpen,2001;Cheeseman等人,2008;Foltz等人,2006;Okada等人,2006)。

[0224] 极度CIN似乎对癌细胞的适合度有害 (Hiley和Swanton,2014;Siegel和Amon,

2012)。CES系统的可能应用是其在预测对用于癌症患者的放疗和化疗药物的敏感性方面的有效性。观察到的对高CES肿瘤的敏感性的简单解释是辐射、顺铂和Topo I抑制剂将DNA损伤增加到不能有效修复的水平。这意味着调节剂量或改变患者敏感性阈值可能对患有中等CES肿瘤的患者有益。

[0225] 除了“了“患阈值”模型,可以通过一些CES基因在DNA修复中的直接参与来解释高CES癌症的敏感性。例如,在癌症细胞系的DNA修复中,HJURP显示是重要的(Kato等人,2007),并且我们先前已经证明了癌症和具有较高HJURP水平的细胞系的辐射敏感性(Hu等人,2010)。CENP-A在DNA损伤和修复中的作用不清楚,并且可能依赖于遗传和细胞环境(Lacoste等人,2014;Mathew等人,2014;Zeitlin等人,2009)。另外,CENP-S/CENP-X也被称为MHF复合物,其通过DNA修复中的FANCM刺激复制叉重塑(Yan等人,2010)。可以想象,CES基因CENP-W的过表达可能有利于在着丝粒处形成CENP-S/-X/-T/-W复合物,并且消耗修复可用的CENP-S/-X库。如果一方面着丝粒调控异常增强CIN和基因组不稳定性,并且另一方面抑制有效DNA修复,则CES系统可以鉴定对进一步DNA损伤非常敏感的患者组。

[0226] 尽管CEN/KT基因对于增殖是必需的,但是它们的错误表达不太可能简单地反映癌症中较快的增殖速率。我们强调着丝粒和动粒缺陷的初级后果是染色体错误分离和基因组异常(Allshire和Karpen,2008),其诱导细胞周期阻滞和细胞死亡作为次级效应(Allshire和Karpen,2008;Heun等人,2006)。实际上,多个实验支持这个结论。首先,与相应的分化的正常组织相比,有丝分裂活跃的组织不一定过表达CEN/KT基因。例如,与相应的正常组织相比,我们没有在晚期肝发育不良的组织或许多乳腺DCIS中观察到CEN/KT的错误表达。类似地,在腹腔疾病中以非癌性、隐蔽性增生为特点的小肠炎性病变不错误表达CEN/KT基因(Bracken等人,2008)。第二,在细胞周期中,许多CEN/KT基因的mRNA水平不会波动(Thiru等人,2014)。第三,我们在同基因乳腺癌进展细胞系中测试了几种CEN/KT基因,包括CENP-A和HJURP,并且发现在癌症进展过程中它们显著上调,尽管这些细胞系具有相似的倍增时间(Zhang,W.,未发表数据;Stampfer,M.,个人通信)。第四,增殖中的纯粹作用意味着,CEN/KT基因的mRNA水平应当始终以相同的方向改变并且对于癌症显示类似的预后价值。然而,即使当在相同复合物中的它们的结合伴侣有预后价值时,几个必需的CEN/KT基因没有预后价值(图2)。此外,对于诸如CENP-C的一些基因,在数种类型的癌症中,降低的表达与更具侵袭性的肿瘤相关。最后,高CES值与高水平的CIN相关,其已知其导致增殖降低(Lee等人,2011)。例如,由于染色体错误分离,CENP-A的过表达引起细胞周期阻滞。现有证据共同认为,CEN/KT基因的功能是确保基因组稳定性而不是简单地促进增殖。

[0227] 对于癌症治疗,已经提出了特异性杀死包含染色体畸变的细胞(Roschke和Kirsch,2005)。在这里我们鉴定了一组基因,其调控异常与药物敏感性相关。未来可能开发调节着丝粒组装和着丝粒/动力结构的新型小分子,其可以帮助克服由CIN导致的药物抗性或者增加癌细胞对特定疗法的应答。

[0228] 材料和方法

[0229] A. 本研究中使用的数据集

[0230] 使用用Affymetrix GeneChip阵列(HG-U133A和HG-U133B或HG-U133 Plus 2.0)分析的几个公布的微阵列数据集来评估CEN/KT基因的调控异常及它们对预后的影响。从GEO网站下载微阵列数据集。标准化的UT SPORE NSCLC数据集直接获自UT Southwestern的

Yang Xie和Hao Tang博士。各TCGA研究中样本组的CEN/KT mRNA表达水平、拷贝数变化的分数和基因突变的频率获自cBioPortal (Gao等人,2013)。

[0231] B. 使用卡普兰-迈耶存活曲线对用于癌症预后的个体CEN/KT基因和CES进行荟萃分析

[0232] 我们按照各自的教程,使用bc-GenExMiner v3.1对17个乳腺癌数据集的乳腺癌AE-生存期、无MR生存期、乳腺癌亚型和具有临床病理学信息的乳腺癌进行了荟萃分析,以及使用KM Plotter对乳腺癌、肺癌、胃癌和卵巢癌进行了荟萃分析 (Gyorffy等人,2013; Jezequel等人,2012)。对于使用K-M Plotter进行预后的荟萃分析,我们选择使用最高CES三分位数作为高CES组,其余作为低CES组。对于使用K-M Plotter对针对辅助治疗的敏感性的荟萃分析,我们选择使用用于所有数据集中的患者分层的自动计算最佳执行CES阈值,以增加检测灵敏度,即使对于许多数据集而言,三分位数或中位数法给出了具有统计显著性的类似结果。K-M Plotter软件仅并入了在U133A+B上均存在探针的基因,以确保用于荟萃分析的数据集之间的可比性,因此其排除了7个CEN/KT基因,包括CENP-H、-W、-L、-K、-P、SPC24和NUF2,其中5个也是CES基因。因此,当使用K-M Plotter时,我们使用剩余9个CES基因,例如CENP-A、HJURP、MIS18B、CENP-N、CENP-M、CENP-U、ZWINT、NDC80和SPC25,作为CES的简化版本。

[0233] C. 基因共表达网络的构建

[0234] 使用从cBioPortal (Gao等人,2013)下载的TCGA数据集,构建每种癌症类型的CEN/KT基因共表达相关网络。使用具有表达相关性插件(ExpressionCorrelation plugin) (网址baderlab.org/Software/ExpressionCorrelation)的Cytoscape 2.8.0软件 (www.cytoscape.org) 构建CEN/KT基因的网络。超过阈值的相关系数(R相关系数)显示为由节点表示的基因之间的边界。将具有较少边界的节点安排在网络的左侧,并且将具有更多边界的那些节点安排在右侧。

[0235] D. 统计分析

[0236] 将微阵列统计分析(SAM)Excel附加包(网址www-stat.stanford.edu/~tibs/SAM/)用于鉴定正常组织和肿瘤组织之间CEN/KT基因表达水平的差异(FDR达水平瘤组以及倍数变化 $\geq$ 数)。使用Cluster 3.0和Java Treeview 1.1.6r4中的分级聚类,对CEN/KT基因表达谱进行样本聚类,以通过质心(centroid)对样本进行分组。对于CCLE药物敏感性,在Excel (Microsoft)和GraphPad中进行分析。为患者生成用于CES系统的卡普兰-迈耶存活曲线以评估在无病生存期(DFS)中的差异,将其分为高CES组(较高三分位数)、中间CES组(中等三分位数)和低CES组(较低三分位数)以评估在无病生存期(DFS)中的差异。对于用于化疗预测的NSCLC JBR.10试验和UT肺SPORE数据集,高CES患者组包含最高CES三分位数中的患者,其余两种三分位数被定义为低CES患者组。使用社会科学统计软件包11.5版(SPSS, Inc., Chicago, IL)和Prism 6.0 GraphPad进行统计分析。在Prism 6.0中生成CES预后和治疗敏感性的荟萃分析的森林图。对于对针对辅助化疗的敏感性的NSCLC JBR.10和UT SPORE荟萃分析,我们合并了分别来自荟萃组的两个数据集的最高CES三分位数和较低的两个CES三分位数,并且在Prims 6.0中制作卡普兰-迈耶存活曲线。通过卡方检验分析5年存活率的显著性。

[0237] 参考文献

- [0238] Allshire, R.C. (1997). Centromeres, checkpoints and chromatid cohesion. *Curr Opin Genet Dev* 7, 264–273.
- [0239] Allshire, R.C., and Karpen, G.H. (2008). Epigenetic regulation of centromeric chromatin: old dogs, new tricks? *Nature Reviews Genetics* 9, 923–937.
- [0240] Artandi, S.E., and DePinho, R.A. (2010). Telomeres and telomerase in cancer. *Carcinogenesis* 31, 9–18.
- [0241] Barnhart, M.C., Kuich, P.H., Stellfox, M.E., Ward, J.A., Bassett, E.A., Black, B.E., and Foltz, D.R. (2011). HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. *J Cell Biol* 194, 229–243.
- [0242] Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483, 603–607.
- [0243] Beroukhi, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., et al. (2010). The landscape of somatic copy-number alteration across human cancers. *Nature*, pp. 899–905.
- [0244] Black, B.E., and Cleveland, D.W. (2011). Epigenetic centromere propagation and the nature of CENP-a nucleosomes. *Cell* 144, 471–479.
- [0245] Black, B.E., Jansen, L.E., Foltz, D.R., and Cleveland, D.W. (2010). Centromere identity, function, and epigenetic propagation across cell divisions. *Cold Spring Harb Symp Quant Biol* 75, 403–418.
- [0246] Blower, M.D., and Karpen, G.H. (2001). The role of *Drosophila* CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. *Nat Cell Biol* 3, 730–739.
- [0247] Blower, M.D., Sullivan, B.A., and Karpen, G.H. (2002). Conserved organization of centromeric chromatin in flies and humans. *Dev Cell* 2, 319–330.
- [0248] Botling, J., Edlund, K., Lohr, M., Hellwig, B., Holmberg, L., Lambe, M., Berglund, A., Ekman, S., Bergqvist, M., Ponten, F., et al. (2013). Biomarker discovery in non-small cell lung cancer: integrating gene expression profiling, meta-analysis, and tissue microarray validation. *Clin Cancer Res* 19, 194–204.
- [0248] Boveri, T. (1902). Ueber mehrpolige Mitosen als Mittel zur Analyse des Zellkerns. *verh d phys med Ges zu Wurzburg NF* 35, 67–90.
- [0249] Boveri, T. (1914). Zur Frage der Entstehung maligner Tumoren (The Origin of Malignant Tumors) (Jena: Gustav Fischer).
- [0250] Bracken, S., Byrne, G., Kelly, J., Jackson, J., and Feighery, C. (2008). Altered gene expression in highly purified enterocytes from patients with active coeliac disease. *BMC genomics* 9, 377.
- [0251] Carter, S.L., Eklund, A.C., Kohane, I.S., Harris, L.N., and Szallasi, Z. (2006). A signature of chromosomal instability inferred from gene expression

profiles predicts clinical outcome in multiple human cancers. In Nat Genet, pp.1043-1048.

[0252] Cheeseman, I.M., and Desai, A. (2008). Molecular architecture of the kinetochore-microtubule interface. Nature reviews Molecular cell biology 9, 33-46.

[0253] Cheeseman, I.M., Hori, T., Fukagawa, T., and Desai, A. (2008). KNL1 and the CENP-H/I/K Complex Coordinately Direct Kinetochore Assembly in Vertebrates. Molecular biology of the cell 19, 587-594.

[0254] Cleveland, D.W., Mao, Y., and Sullivan, K.F. (2003). Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. Cell 112, 407-421.

[0255] de Bruin, E.C., McGranahan, N., Mitter, R., Salm, M., Wedge, D.C., Yates, L., Jamal-Hanjani, M., Shafi, S., Murugaesu, N., Rowan, A.J., et al. (2014). Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. Science 346, 251-256.

[0256] Dunleavy, E.M., Roche, D., Tagami, H., Lacoste, N., Ray-Gallet, D., Nakamura, Y., Daigo, Y., Nakatani, Y., and Almouzni-Pettinotti, G. (2009). HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. Cell 137, 485-497.

[0257] Earnshaw, W.C., Bernat, R.L., Cooke, C.A., and Rothfield, N.F. (1991). Role of the centromere/kinetochore in cell cycle control. Cold Spring Harb Symp Quant Biol 56, 675-685.

[0258] Foltz, D.R., Jansen, L.E., Bailey, A.O., Yates, J.R., 3rd, Bassett, E.A., Wood, S., Black, B.E., and Cleveland, D.W. (2009). Centromere-specific assembly of CENP-a nucleosomes is mediated by HJURP. Cell 137, 472-484.

[0259] Foltz, D.R., Jansen, L.E., Black, B.E., Bailey, A.O., Yates, J.R., 3rd, and Cleveland, D.W. (2006). The human CENP-A centromeric nucleosome-associated complex. Nat Cell Biol 8, 458-469.

[0260] Fujita, Y., Hayashi, T., Kiyomitsu, T., Toyoda, Y., Kokubu, A., Obuse, C., and Yanagida, M. (2007). Priming of centromere for CENP-A recruitment by human hMis18alpha, hMis18beta, and M18BP1. Dev Cell 12, 17-30.

[0261] Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Science signaling 6, p11.

[0262] Geigl, J.B., Obenauf, A.C., Schwarzbraun, T., and Speicher, M.R. (2008). Defining 'chromosomal instability'. Trends in genetics: TIG 24, 64-69.

[0263] Györfy, B., Lanczky, A., Eklund, A.C., Denkert, C., Budczies, J., Li, Q., and Szallasi, Z. (2010). An online survival analysis tool to rapidly assess the

effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res Treat* 123,725–731.

[0264] Gyorffy, B., Surowiak, P., Budczies, J., and Lanczky, A. (2013). Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLoS One* 8,e82241.

[0265] Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. In *Cell*, pp. 646–674.

[0266] Heun, P., Erhardt, S., Blower, M. D., Weiss, S., Skora, A. D., and Karpen, G. H. (2006). Mislocalization of the *Drosophila* centromere-specific histone CID promotes formation of functional ectopic kinetochores. *Dev Cell* 10,303–315.

[0267] Hiley, C. T., and Swanton, C. (2014). Spatial and temporal cancer evolution: causes and consequences of tumour diversity. *Clinical medicine* 14 Suppl 6,s33–37.

[0268] Hori, T., Amano, M., Suzuki, A., Backer, C. B., Welburn, J. P., Dong, Y., McEwen, B. F., Shang, W.-H., Suzuki, E., Okawa, K., et al. (2008). CCAN makes multiple contacts with centromeric DNA to provide distinct pathways to the outer kinetochore. *Cell* 135,1039–1052.

[0269] Hu, Z., Huang, G., Sadanandam, A., Gu, S., Lenburg, M. E., Pai, M., Bayani, N., Blakely, E. A., Gray, J. W., and Mao, J.-H. (2010). The expression level of HJURP has an independent prognostic impact and predicts the sensitivity to radiotherapy in breast cancer. In *Breast Cancer Res*, pp. R18.

[0270] Janssen, A., van der Burg, M., Szuhai, K., Kops, G. J., and Medema, R. H. (2011). Chromosome segregation errors as a cause of DNA damage and structural chromosome aberrations. *Science* 333,1895–1898.

[0271] Jezequel, P., Campone, M., Gouraud, W., Guerin-Charbonnel, C., Leux, C., Ricolleau, G., and Campion, L. (2012). bc-GenExMiner: an easy-to-use online platform for gene prognostic analyses in breast cancer. *Breast cancer research and treatment* 131,765–775.

[0272] Jezequel, P., Frenel, J. S., Campion, L., Guerin-Charbonnel, C., Gouraud, W., Ricolleau, G., and Campone, M. (2013). bc-GenExMiner 3.0: new mining module computes breast cancer gene expression correlation analyses. *Database (Oxford)* 2013,bas060.

[0273] Kato, T., Sato, N., Hayama, S., Yamabuki, T., Ito, T., Miyamoto, M., Kondo, S., Nakamura, Y., and Daigo, Y. (2007). Activation of Holliday junction recognizing protein involved in the chromosomal stability and immortality of cancer cells. *Cancer Res* 67,8544–8553.

[0274] Kops, G. J. P. L., Weaver, B. A. A., and Cleveland, D. W. (2005). On the road to cancer: aneuploidy and the mitotic checkpoint. In *Nature reviews Cancer*, pp. 773–785.

- [0275] Lacoste,N.,Woolfe,A.,Tachiwana,H.,Garea,A.V.,Barth,T.,Cantaloube,S.,Kurumizaka,H.,Imhof,A.,and Almouzni,G. (2014) .Mislocalization of the centromeric histone variant CenH3/CENP-A in human cells depends on the chaperone DAXX.Molecular cell 53,631-644.
- [0276] Lee,A.J.X.,Endesfelder,D.,Rowan,A.J.,Walther,A.,Birkbak,N.J.,Futreal,P.A.,Downward,J.,Szallasi,Z.,Tomlinson,I.P.M.,Howell,M.,et al. (2011) .Chromosomal instability confers intrinsic multidrug resistance.In Cancer Research,pp.1858-1870.
- [0277] Loi (2007) .Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade(vol 25,pg 1239,2007) .Journal of Clinical Oncology 25,3790-3790.
- [0278] Mathew,V.,Pauleau,A.L.,Steffen,N.,Bergner,A.,Becker,P.B.,and Erhardt,S. (2014) .The histone-fold protein CHRAC14 influences chromatin composition in response to DNA damage.Cell reports 7,321-330.
- [0279] Mathijssen,R.H.,Loos,W.J.,Verweij,J.,and Sparreboom,A. (2002) .Pharmacology of topoisomerase I inhibitors irinotecan (CPT-11) and topotecan.Current cancer drug targets 2,103-123.
- [0280] MCGovern,S.L.,Qi,Y.,Pusztai,L.,Symmans,W.F.,and Buchholz,T.A. (2012) .Centromere protein-A,an essential centromere protein,is a prognostic marker for relapse in estrogen receptor-positive breast cancer.Breast Cancer Res, pp.R72.
- [0281] Mellone,B.G.,Zhang,W.,and Karpen,G.H. (2009) .Frodo's found:Behold the CENP-a"Ring"bearers.Cell,pp.409-412.
- [0282] Mendiburo,M.J.,Padeken,J., Fülöp, S.,Schepers,A.,and Heun,P. (2011) .Drosophila CENH3 Is Sufficient for Centromere Formation.Science 334,686-690.
- [0283] Mishra,P.K.,Au,W.C.,Choy,J.S.,Kuich,P.H.,Baker,R.E.,Foltz,D.R.,and Basrai,M.A. (2011) .Misregulation of Scm3p/HJURP causes chromosome instability in Saccharomyces cerevisiae and human cells.PLoS Genet 7,e1002303.
- [0284] Moree,B.,Meyer,C.B.,Fuller,C.J.,and Straight,A.F. (2011) .CENP-Crecruits M18BP1 to centromeres to promote CENP-A chromatin assembly.The Journal of Cell Biology 194,855-871.
- [0285] Nathanson,D.A.,Gini,B.,Mottahedeh,J.,Visnyei,K.,Koga,T.,Gomez,G.,Eskin,A.,Hwang,K.,Wang,J.,Masui,K.,et al. (2014) .Targeted therapy resistance mediated by dynamic regulation of extrachromosomal mutant EGFR DNA.Science 343,72-76.
- [0286] Negrini,S.,Gorgoulis,V.G.,and Halazonetis,T.D. (2010) .Genomic instability--an evolving hallmark of cancer.Nat Rev Mol Cell Biol,pp.220-228.
- [0287] Nik-Zainal,S.,Van Loo,P.,Wedge,D.C.,Alexandrov,L.B.,Greenman,C.D.,Lau,K.W.,Raine,K.,Jones,D.,Marshall,J.,Ramakrishna,M.,et al. (2012) .The life

history of 21 breast cancers. *Cell* 149,994–1007.

[0288] Nishino, T., Takeuchi, K., Gascoigne, K.E., Suzuki, A., Hori, T., Oyama, T., Morikawa, K., Cheeseman, I.M., and Fukagawa, T. (2012). CENP-T-W-S-X forms a unique centromeric chromatin structure with a histone-like fold. *Cell* 148,487–501.

[0289] Okada, M., Cheeseman, I.M., Hori, T., Okawa, K., McLeod, I.X., Yates, J.R., 3rd, Desai, A., and Fukagawa, T. (2006). The CENP-H-I complex is required for the efficient incorporation of newly synthesized CENP-A into centromeres. *Nat Cell Biol* 8,446–457.

[0290] Parker, J.S., Mullins, M., Cheang, M.C., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., et al. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 27,1160–1167.

[0291] Perpelescu, M., and Fukagawa, T. (2011). The ABCs of CENPs. *Chromosoma* 120, 425–446.

[0292] Pierotti, M.A., Sozzi, Gabriella, Croce, Carlo M. (2003). *Holland-Frei Cancer Medicine*, 6th Edition, 6th Edition edn (Hamilton (ON): BC Decker).

[0293] Pihan, G.A., Wallace, J., Zhou, Y., and Doxsey, S.J. (2003). Centrosome abnormalities and chromosome instability occur together in pre-invasive carcinomas. *Cancer Research*, pp.1398–1404.

[0294] Roschke, A.V., and Kirsch, I.R. (2005). Targeting cancer cells by exploiting karyotypic complexity and chromosomal instability. In *Cell Cycle*, pp.679–682.

[0295] Schwartzman, J.M., Sotillo, R., and Benezra, R. (2010). Mitotic chromosomal instability and cancer: mouse modelling of the human disease. *Nature reviews Cancer* 10,102–115.

[0296] Siegel, J.J., and Amon, A. (2012). New insights into the troubles of aneuploidy. *Annu Rev Cell Dev Biol* 28,189–214.

[0297] Sullivan, B.A., and Karpen, G.H. (2004). Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nature structural & molecular biology* 11,1076–1083.

[0298] Swanton, C., Marani, M., Pardo, O., Warne, P.H., Kelly, G., Sahai, E., Elustondo, F., Chang, J., Temple, J., Ahmed, A.A., et al. (2007). Regulators of mitotic arrest and ceramide metabolism are determinants of sensitivity to paclitaxel and other chemotherapeutic drugs. *Cancer cell* 11,498–512.

[0299] Tang, H., Xiao, G., Behrens, C., Schiller, J., Allen, J., Chow, C.-W., Suraokar, M., Corvalan, A., Mao, J., White, M.A., et al. (2013). A 12-gene set predicts survival benefits from adjuvant chemotherapy in non-small cell lung cancer patients. *Clin Cancer Res*, pp.1577–1586.

[0300] Thiru, P., Kern, D.M., McKinley, K.L., Monda, J.K., Rago, F., Su, K.C., Tsinman, T., Yarar, D., Bell, G.W., and Cheeseman, I.M. (2014). Kinetochore genes are

coordinately up-regulated in human tumors as part of a FoxM1-related cell division program. *Mol Biol Cell* 25,1983-1994.

[0301] Tomonaga,T.,Matsushita,K.,Ishibashi,M.,Nezu,M.,Shimada,H.,Ochiai,T.,Yoda,K.,and Nomura,F. (2005) .Centromere protein H is up-regulated in primary human colorectal cancer and its overexpression induces aneuploidy. *Cancer Res* 65,4683-4689.

[0302] Tomonaga,T.,Matsushita,K.,Yamaguchi,S.,Oohashi,T.,Shimada,H.,Ochiai,T.,Yoda,K.,and Nomura,F. (2003) .Overexpression and mistargeting of centromere protein-A in human primary colorectal cancer. *Cancer Res* 63,3511-3516.

[0303] Weaver,B.A.A.,and Cleveland,D.W. (2006) .Does aneuploidy cause cancer? In *Curr Opin Cell Biol*,pp.658-667.

[0304] Weaver,B.A.A.,Silk,A.D.,Montagna,C.,Verdier-Pinard,P.,and Cleveland,D.W. (2007) .Aneuploidy acts both oncogenically and as a tumor suppressor. In *Cancer Cell*,pp.25-36.

[0305] Yachida,S.,Jones,S.,Bozic,I.,Antal,T.,Leary,R.,Fu,B.,Kamiyama,M.,Hruban,R.H.,Eshleman,J.R.,Nowak,M.A.,et al. (2010) .Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* 467,1114-1117.

[0306] Yan,Z.,Delannoy,M.,Ling,C.,Daee,D.,Osman,F.,Muniandy,P.A.,Shen,X.,Oostra,A.B.,Du,H.,Steltenpool,J.,et al. (2010) .A histone-fold complex and FANCM form a conserved DNA-remodeling complex to maintain genome stability. *Molecular cell* 37,865-878.

[0307] Zeitlin,S.G.,Baker,N.M.,Chapados,B.R.,Soutoglou,E.,Wang,J.Y.,Berns,M.W.,and Cleveland,D.W. (2009) .Double-strand DNA breaks recruit the centromeric histone CENP-A. *Proc Natl Acad Sci U S A* 106,15762-15767.

[0308] Zhang,J.,Fujimoto,J.,Zhang,J.,Wedge,D.C.,Song,X.,Zhang,J.,Seth,S.,Chow,C.W.,Cao,Y.,Gumbs,C.,et al. (2014) .Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* 346,256-259.

[0309] Zhu,C.Q.,Ding,K.,Strumpf,D.,Weir,B.A.,Meyerson,M.,Pennell,N.,Thomas,R.K.,Naoki,K.,Ladd-Acosta,C.,Liu,N.,et al. (2010) .Prognostic and predictive gene signature for adjuvant chemotherapy in resected non-small-cell lung cancer. *J Clin Oncol* 28,4417-4424.

[0310] >gi|12803733|gb|AAH02703.1|CENPA蛋白[智人]

MGPRRRSRKPEAPRRRSPSPTPTPGPSRRGPSLGASSHQHSRRRQGWLKEIRKLQKSTHLLIRKLPF SRL

[0311] AREICVKFTRGVDFNWQAQALLALQEAAEFLVHLFEDAYLLTLHAGRVTLFPKDVQLARRIRGLEEG
LG

[0312] >gi|83816964|ref|NP_060880.3|霍利迪连接体识别蛋白同种型a[智人]

- MLGTLRAMEGEDVEDDQLLQKLRRRRFQRRMQRLIEKYNQPFEDTPVVQMATLTYETPQGLRIWG
GRLIKERNEGEIQDSSMKPADRTDGSVQAAAWGPPELPSHRTVLGADSKSGEVDATSDQEEVAVALA
PAVPQSPLKNELRRKYLTQVDILLQGAIEYFECAGNRAGRDVRVTPLPSLASPAVPAPGYCSRISRKSPG
DPAKPASSPREWDPLHPSSTDMAVPRNDSLSLQETSSSSFLSSQPFEDDDICNVTISDLYAGMLHMSR
LLSTKPSSIISTKTFIMQNWNSRRRHRYKSRMNKTYCKGARRSQRSSKENFIPCSEPVKGTGALRDCKN
[0313] VLDVSCRKTGLKLEKAFLEVNRPQIHKLDPSWKERKVTSPKYSSLIYFDSSATYNLDEENRFTLKWLI
SPVKIVSRPTIRQGHGENRQREIEIRFDQLHREYCLSPRNQPRRMCLPDSWAMNMYRGGPASPGLQGL
ETRRLSLPSSKAKAKSLSEAFENLGKRSLEAGRCLPKSDSSSLPKTNPTHSAITRQQTSDLHVQGNSSG
IFRKSVPSTKLSVPDKEVPGHGRNRYDEIKEEFDKLHQKYCLKSPGQMTVPLCIGVSTDKASMEVRYQ
TEGFLGKLNPDPHFQGFQKLPSPLGCRKSLLGSTAIEAPSSTCVARAITRDGTRDHQFPAKRPRLSEPQ
GSGRQGNLGA SDGVDNTVRPGDQSSSQPNSEERGENTSYSRMEEKSDFMLEKLETKSV
[0314] >gi|545477999|ref|NP_001269891.1|霍利迪连接体识别蛋白同种型b[智人]
MLGTLRAMEGEDVEDDQLLQKLRRRRFQRRMQRLIEKYNQPFEDTPVVQMATLTYETPQGLRIWG
GRLIKERNEGEIQPAVPQSPLKNELRRKYLTQVDILLQGAIEYFECAGNRAGRDVRVTPLPSLASPAVPA
PGYCSRISRKSPGDPAKPASSPREWDPLHPSSTDMAVPRNDSLSLQETSSSSFLSSQPFEDDDICNVTIS
DLYAGMLHMSRLLSTKPSSIISTKTFIMQNWNSRRRHRYKSRMNKTYCKGARRSQRSSKENFIPCSEPV
KGTGALRDCKNVLDVSCRKTGLKLEKAFLEVNRPQIHKLDPSWKERKVTSPKYSSLIYFDSSATYNL
[0315] DEENRFTLKWLI SPVKIVSRPTIRQGHGENRQREIEIRFDQLHREYCLSPRNQPRRMCLPDSWAMNMY
RGGPASPGLQGLETRRLSLPSSKAKAKSLSEAFENLGKRSLEAGRCLPKSDSSSLPKTNPTHSAITRQ
QTSDLHVQGNSSGIFRKSVPSTKLSVPDKEVPGHGRNRYDEIKEEFDKLHQKYCLKSPGQMTVPLCIG
VSTDKASMEVRYQTEGFLGKLNPDPHFQGFQKLPSPLGCRKSLLGSTAIEAPSSTCVARAITRDGTRD
HQFPAKRPRLSEPQSGRQGNLGA SDGVDNTVRPGDQSSSQPNSEERGENTSYSRMEEKSDFMLEKL
ETKSV
[0316] >gi|545478902|ref|NP_001269892.1|霍利迪连接体识别蛋白同种型c[智人]

- MLGTLRAMEGEDVEDDQLLQKLRRRRFQRRMQRLIEKYNQPFEDTPVVQMATLTYETPQGLRIWG
GRLIKERNEGEIQCAGNRAGRDVVRVTPPLSLASPAVPAPGYCSRISRKSPGDPAPASSPREWDPLHPSS
TDMALVPRNDSLQETSSSSFLSSQPFEDDDICNVITISDLYAGMLHMSRLLSTKPSSIISTKTFIMQNW
NSRRRHRYKSRMNKTYCKGARRSQRSSKENFIPCSEPVKGTGALRDCKNVLDVSCRKTGLKLEKAFLE
[0317] VNRPQIHKLDPSWKERKVTSPKYSSLIYFDSSATYNLDEENRFRTLKWLISPVKIVSRPTIRQGHGENRQ
REIEIRFDQLHREYCLSPRNQPRRMCLPDSWAMNMYRGGPASPGLQGLETRRLSLPSSKAKAKSLSEA
FENLGKRSLEAGRCLPKSDSSSSLPTNPTHSATRPQQTSDLHVQGNSSGIFRKSVPSPKTLSPDKEVP
GHGRNRYDEIKEEFDKLHQKYCLKSPGQMTVPLCIGVSTDKASMEVRYQTEGFLGKLNPDPHFQGFQK
LPSSPLGCRKSLLGSTAIEAPSSTCVARAITRDGTRDHQFPAKRPRLSEPQSGRQGNLGLASDGDVNTV
RPGDQGS SSQPNSEERGENTS YRMEEKSDFMLEKLETKSV
- [0318] >gi|9506437|ref|NP_061817.1|蛋白Mis18- α [智人]
MAGVRSRLCRGCAGGCECGDKGKCSDSLGLKRLSESSRHQLLQKQWASMWSSMSSEDASVADMER
AQLEEEAAAAEERPLVFLCSGCRRPLGDSLSWVASQEDTNCILLRCVSCNVSDKEQKLSKREKENG
[0319] VLETLCCAGCSNLGYVYRCTPKNLDYKRDLCFLSVEAIESYVLGSSEKQIVSEDKELFNLESRVEIEKS
LTQMED VLKALQMKLWEAESKLSFATCKS
- [0320] >gi|24307929|ref|NP_009211.1|蛋白Mis18- β [智人]
MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPS
WLQPERCAVFQCAQCHAVLADSVHLAWDLRSRLGAVVFSRVTNVLEAPFLVGIEGSLKGSTYNLL
[0321] FCGSGIPVGFHLYSTHAALAALRGHFLSSDKMVCYLLKTKAIVNASEMDIQNVPLSEKIAELKEKIV
LTHNRL KSLMKILSEVTPDQSKPEN
- [0322] >gi|119625950|gb|EAX05545.1|着丝粒蛋白C1,同种型CRA_a[智人]
MAASGLDHLKNGYRRRRCRPSRARDINTEQGQNVLEILQDCFEKSLANDFSTNSTKSVPNSTRKIKDT
CIQSPSKECQKSHPKSVPVSSKKKEASLQFVVEPSEATNRSVQAHEVHQKILATDVSSKNTPDSKKISSR
[0323] NINDHHSEADEEFYLSVGSPSVLLDAKTSVSQNVIPSSAQKRETYTFENSVNMLPSSTEVSVKTKKRLNF
DDKVMLKKIEIDNKVSDEEDKTSEGQERKPSGSSQNRIRDSEYEIQRQAKKSFSTLFLETVKRKSESPIV
RHAATAPPHSCPPDDTKLIEDEFIIDESDQSFASRSWITIPRKAGSLKQRTISPAESTALFQGRKSREKH

- HNILPKTLANDKSHKPHPVETSQPSDKTVLDTSYALIGETVNNYRSTKYEMYSKNAEKPSRSKRTIKQ
KQRRKFMAKPAEEQLDVGQSKDENIHTSHITQDEFQRNSDRNMEEHEEMGNDCVSKKQMPVVGSKKS
STRKDKEESKKKRFSSSESKNKLVP E E V T S T V T K S R R I S R R P S D W W V V K S E E S P V Y S N S S V R N E L P M H H N
SSRKSTKKTNQSSKNIRKKTIP L K R Q K T A T K G N Q R V Q K F L N A E G S G G I V G H D E I S R C S L S E P L E S D E A D L
[0324] AKKKNLDCSRSTRSSKNEDNIMTAQNVPLKPQTSGYTCNIPTESNLDSGEHKTSVLEESGPSRLNNNYL
MSGKNDVDDEEVHGSSDDSKQSKVIPKNRIHKLVLPSNTPNVRRTKRTRLKPLEYWRGERIDYQGRP
SGGFVISGVLSPD T I S S K R K A K E N I G K V N K K S N K K R I C L D N D E R K T N L M V N L G I P L G D P L Q P T R V K D P E T
REIILMDLVRPQDTYQFFVKHGELKVYKTLDT P F F S T G K L I L G P Q E E K G K Q H V G Q D I L V F Y V N F G D L L C
TLHETPYIL STGDSFYVPSGNYNINLNRNEESVLLFTQIKR
- [0325] >gi|119625951|gb|EAX05546.1|着丝粒蛋白C1,同种型CRA_b[智人]
MAASGLDHLKNGYRRRFCRPSRARDINTEQGQNVLEILQDCFEKSLANDFSTNSTKSVPNSTRKIKDT
CIQSPSKECQKSHPKSVPVSSKKKEASLQFVVEPSEATNRSVQAHEVHQKILATDVSSKNTPDSKKISSR
NINDHHSEADEEFYLSVGSPSVLLDAKTSVSQNVIPSSAQKRETYTFENSVMNLPSSSTEVSVKTKKRLNF
DDKVMLKKIEIDNKVSDEEDKTSEGQERKPSGSSQNRIRDSEYEIQRQAKKSFSTLFLETVKRKSESPIV
[0326] RHAATAPPHSCPPDDTKLIEDEFIIDESDQSFASRSWITIPRKAGSLKQRTISPAESTALFQGRKSREKH
HNILPKTLANDKSHKPHPVETSQPSDKTVLDTSYALIGETVNNYRSTKYEMYSKNAEKPSRSKRTIKQ
KQRRKFMAKPAEEQLDVGQSKDENIHTSHITQDEFQRNSDRNMEEHEEMGNDCVSKKQMPVVGSKKS
STRK DKEESKKKRFSSSESKNKLVP E E V T S T V T K S R R I S R R P S D W W V V K S E E S C L K C
[0327] >gi|395132501|ref|NP_060925.2|着丝粒蛋白N同种型3[智人]
MDETVAEFIKRTILKIPMNELTILKAWDFLSENQLQTVNFRQRKESVVQHHLHLCEEKRASISDAALLD
- [0328] IIYMQFHQHQKVWEVFQMSKGPGEVDLFDMDKQFKNSFKKILQRALKNVTVSFRETEENAVWIRIAW
GTQ YTKPNQYKPTYVVYYSQTPYAFTSSSMLRRNTPLLQGELEATGKIYLRQEEHLDITEMKKACN
- [0329] >gi|395132497|ref|NP_001257403.1|着丝粒蛋白N同种型5[智人]
MDETVAEFIKRTILKIPMNELTILKAWDFLSENQLQTVNFRQRKESVVQHHLHLCEEKRASISDAALLD
- [0330] IIYMQFHQHQKVWEVFQMSKGPGEVDLFDMDKQFKNSFKKILQRALKNVTVSFRETEENAVWIRIAW
GTQYTKPNQYKPTYVVYYSQTPYAFTSSSMLRRNTPLLQGT FETHNSTTPLQERSLGLDINMDSRIHEN
IVEKERVQRITQETFGDYPQPQLEFAQYKLET K F K S G L N G S I L A E R E E P L R C L I K F S S P H L L E A L K S L A P A
[0331] GI ADAPLSPLLT CIPNKR MNYFKIRDK
- [0332] >gi|395132495|ref|NP_001257402.1|着丝粒蛋白N同种型4[智人]
MDETVAEFIKRTILKIPMNELTILKAWDFLSENQLQTVNFRQRKESVVQHHLHLCEEKRASISDAALLD
IICEDVDLFDMDKQFKNSFKKILQRALKNVTVSFRETEENAVWIRIAWGTQYTKPNQYKPTYVVYYSQT
- [0333] PYAFTSSSMLRRNTPLLQGALTIASHHQIVKMDLRSRYLDSLKAIVFKQYNQTFETHNSTTPLQERSLG
LD INMDSRIHENIVEKERVQRITQETFGDYPQPQLEFAQYKLET K F K S G L N G S I L A E R E E P L R C L I K F S S P
H L L E A L K S L A P A G I A D A P L S P L L T C I P N K R M N Y F K I R D K

- [0334] >gi|395132493|ref|NP_001094094.2|着丝粒蛋白N同种型2[智人]
MDETVAEFIKRTILKIPMNELTTLKAWDFLSENQLQTVNFRQRKESVVQHHLHLCEEKRASISDAALLD
IIYMQFHQHQKVWEVFQMSKGPGEVDLFDMDKQFKNSFKKILQRALKNVTVSFRETEENAVWIRIAW
- [0335] GTQYTKPNQYKPTYVVVYYSQTPYAFTSSSMLRRNTPLLQALTIASKHHQIVKMDLRSRYLDSLKAIVF
KQYNQTFETHNSTTPLQERSLGLDINMDSRIIHENIVEKERVQRITQETFGDYPQPQLEFAQYKLETQFK
SGLN GSILAEREEPLRCLIKFSSPHLLEALKSLAPAGIADAPLSPLLTICIPNKRMYFKIRDK
- [0336] >gi|395132491|ref|NP_001094095.2|着丝粒蛋白N同种型1[智人]
MDETVAEFIKRTILKIPMNELTTLKAWDFLSENQLQTVNFRQRKESVVQHHLHLCEEKRASISDAALLD
IIYMQFHQHQKVWEVFQMSKGPGEVDLFDMDKQFKNSFKKILQRALKNVTVSFRETEENAVWIRIAW
GTQYTKPNQYKPTYVVVYYSQTPYAFTSSSMLRRNTPLLQALTIASKHHQIVKMDLRSRYLDSLKAIVF
- [0337] KQYNQTFETHNSTTPLQERSLGLDINMDSRIIHENIVEKERVQRITQETFGDYPQPQLEFAQYKLETQFK
SGLNGSILAEREEPLRCLIKFSSPHLLEALKSLAPAALVCRIQKLLCYSGSHSQGTQDPSSWQKDLYLLF
VPLY PRC
- [0338] >gi|77416860|sp|Q92674.2|CENPI_人RecName:全称=着丝粒蛋白I;缩写=
CENP-I;AltName:全称=FSH初级应答蛋白1;AltName:全称=促卵泡激素初级应答蛋白;
AltName:全称=间期着丝粒复合蛋白19;AltName:全称=富含亮氨酸的初级应答蛋白1
MSPQKRKVNQVQAQNRSTSGSSSFQTTLSAWKVQDPSNSKNISKHGQNNPVG DYEHADDQAEEDAL
- [0339] QMAVG YFEKGPIKASQNKDKTLEKHLKTVENVAWKNGLASEEIDILLNIALSGKFGNAVNTILKCMIP
ATVISEDVVKAVSWLCVKGKSGSTKVLFYRWLVAMFDFIDRKEQINLLYGFFASLQDDALCPYVCH
LLYLLTKKENVKPFRVRKLLDLQAKMGMQPHLQALLSLYKFFAPALISVSLPVRKKIYFKNSEN LWKT
ALLAVKQRNRGPSPEPLKMLGPANVRPLKRKWNLSVIPVLNSSSYTKECGKKEMSLSDCLNRSGSFP
LEQLQSFPQLLQNIHCLELPSQMGSVLNNSLLHYNLCVRDEPVLLRFYYWLSQTLQEECIWKVNNYE
- [0340] HGKEFTNFDLTIIRAECFLQEGFYSCFAFLYKSLPLWDGLCCRSQFLQLVSWIPFSSFSEVKPLLFDHLAQ
LFFTSTIYFKCSVLQSLKELLQNWLLWLSMDIHMKPVTNSPLETTLGGSMNSVSKLIHYVGWLSTTAM
RLESNNTFLLHFILDFYEKVCDIYINYNLPLVVLFPFGIFYSALLSLDTSILNQLCFIMHRYRKNLTAACK
NELVQKTKSEFNFSKTYQEFNHYLTSMVGCLWTSKPFKGKIYIDPEILEKTGVAEYKNSLNVVHHPSF
LSYAVSFLQES PEERTVNVSIRGKKWSWYLDYLFSSQGLQGLKLFIRSSVHHSSIPRAEGINCNNQY
- [0341] >gi|74733576|sp|Q9H3R5.1|CENPH_人RecName:全称=着丝粒蛋白H;缩写=
CENP-H;AltName:全称=间期着丝粒复合蛋白35
MEEQPQM QDADEPADSGGEGRAGGPPQVAGAQAACSEDRMTLLRLRAQTKQQLLEYKSMVDASEE
- [0342] KTPEQIMQE KQIEAKIEDLENEIEEVKVAFEIKKLALDRMRLSTALKKNLEKISRQSSVLMDNMKHLLE
LNKLIMKSQQESWDLEEKLLDIRKKRLQLKQASESKLLEIQTEKNKQKIDLDSMENSERIKIIRQNLQME
IKITT VIQHVFNILGSKVNWAEDPALKEIVLQLEKNVDM
- [0343] >gi|557357675|ref|NP_001273454.1|着丝粒蛋白W同种型c[智人]

- [0344] MALSTIVSQRKQIKRKAPRGFLKRVFKRKKPQLRLEKSGDLLKSPGQTLVRVNVESLTRSMYWPQQR
- [0345] >gi|557355812|ref|NP_001273453.1|着丝粒蛋白W同种型a[智人]
MALSTIVSQRKQIKRKAPRGFLKRVFKRKKPQLRLEKSGDLLVRFHPFSGWEWGTGEVHLNCLLFVHR
- [0346] LA EESRTNACASKCRVINKEHVLAAAKVILKKS RG
- [0347] >gi|60302883|ref|NP_001012525.1|着丝粒蛋白W同种型b[智人]
MALSTIVSQRKQIKRKAPRGFLKRVFKRKKPQLRLEKSGDLLVHLNCLLFVHRLAEESRTNACASKCR
- [0348] VINKEHVLAAAKVILKKS RG
- [0349] >gi|401871068|ref|NP_001257936.1|着丝粒蛋白X同种型3[智人]
- [0350] MEGAGAGSGFRKELVSRLHLHFKDDKTKEAAVRGVRQAQAEDALRVDVDQLEKLLDF
- [0351] >gi|401871066|ref|NP_001257935.1|着丝粒蛋白X同种型1[智人]
MEGAGAGSGFRKELVSRLHLHFKDDKT KVS GDALQLMVELLKVFVVEAAVRGVRQAQAEDALRVD
- [0352] VDQL EK VLPQ LLLDF
- [0353] >gi|71559139|ref|NP_659435.2|着丝粒蛋白X同种型2[智人]
- [0354] MEGAGAGSGFRKELVSRLHLHFKDDKTKEAAVRGVRQAQAEDALRVDVDQLEK VLPQ LLLDF
- [0355] >gi|50845414|ref|NP_001002876.1|着丝粒蛋白M同种型b[智人]
MSVLRPLDKLPGLNTATILLVGTE DALLQQLAD SMLKEDCASELKVHLAKSLPLPSSVNRPRIDLIVFV
- [0356] VNLH SKYSLQ NTEESLRHVDASFFLGKVCFLATGGGRL
- [0357] >gi|13129022|ref|NP_076958.1|着丝粒蛋白M同种型a[智人]
MSVLRPLDKLPGLNTATILLVGTE DALLQQLAD SMLKEDCASELKVHLAKSLPLPSSVNRPRIDLIVFV
- [0358] VNLH SKYSLQ NTEESLRHVDASFFLGKVCFLATGAGRESHCSIHRHTVVKLAHTYQSPLLYCDLEVEGF
RA TMAQRLVRVLQICAGHVPGVSA LNLLSLLRSSEGPSLEDL
- [0359] >gi|158966691|ref|NP_001103685.1|着丝粒蛋白M同种型c[智人]
- [0360] MGRVWDLPGVLKVEGFRATMAQRLVRVLQICAGHVPGVSA LNLLSLLRSSEGPSLEDL
- [0361] >gi|74712714|sp|Q71F23.1|CENPU_人RecName: 全称=C着丝粒蛋白U; 缩写=CENP-U; AltName: 全称=50kDa的着丝粒蛋白; 缩写=CENP-50; AltName: 全称=间期着丝粒复合蛋白24; AltName: 全称=KSHV潜伏核抗原相互作用蛋白1; AltName: 全称=MLF1相互作用蛋白; AltName: 全称=Polo框相互作用蛋白1
MAPRGRRRRPRHRSEGARRSKNTLERTHSMKDKAGQKCKPIDVDFDPD NSDVSSIGRLGENEKDEETY
ET FDPPLHSTAIYADEEEFSKHCGLSLSSSTPPGKEAKRSSDTSGNEASEIESVKISAKKPGRKLRPISDDSE
SIEESDTRRKVKSAEKISTQRHEVIRTTASSELSEKPAESVTSKKTGPLSAQPSVEKENLAIESQSKTQK
- [0362] KGKISHDKRKKSRSAIGSDTSDIVHIWCPEGMKTS DIKELNIVLPEFEKTHLEHQRIESKVCKAAIAT
FYVNVKEQFIKMLKESQMLTNLKRKNAKMISDIEKKRQRMIEVQDELLRLEPQLKQLQTKYDELKERK
SSLRNAAYFLSNLKQLYQDYS DVQAQEPNVKETYDSSSLPALLFKARTLLGAESHLRNINHQLEKLLD
QG
- [0363] >gi|113958992|gb|ABI49143.1|MLF1相互作用蛋白同种型2[智人]

- MAPRGRRRRPRPHRSEGARRSKNTLERTHSMKDKAGQKCKPIDVDFDPDNDVSSIGRLGENEKDEETY
ET FDPPLHSTAIYADEEEFSKHCGLSLSTPPGKEAKRSSDTSGNEASEIESVKISAKKPGRKLRPISDDSE
- [0364] SIEESDTRRKVKSAEKISTQRHEVIRTTASSELSEKPAESVTSKKTGPLSAQPSVEKENLAIESQSKTQK
KKGISHDKRKKSRSKAIGSDTSDIVHIWCPEGMKTSDIKELNIVLPEFEKTHLEHQRIESKVCKAAIAT
FYVNVKEQFIKMLKESQMLTNLKRKNAKVRATAETTTNKI
- [0365] >gi|113958976|gb|ABI49142.1|MLF1相互作用蛋白同种型1[智人]
MAPRGRRRRPRPHRSEGARRSKNTLERTHSMKDKAGQKCKPIDVDFDPDNDVSSIGRLGENEKDEETY
ET FDPPLHSTAIYADEEEFSKHCGLSLSTPPGKEAKRSSDTSGNEASEIESVKISAKKPGRKLRPISDDSE
- [0366] SIEESDTRRKVKSAEKISTQRHEVIRTTASSELSEKPAESVTSKKTGPLSAQPSVEKENLAIESQSKTQK
KKGISHDKRKKSRSKAIGSDTSDIVHIWCPEGMKTSDIKELNIVLPEFEKTHLEHQRIESKVCKAAIAT
FYVNVKEQFIKMLKESQMLTNLKRKNAKMISDIEKKRQRMIEVQDELLRLWTGAGLW
- [0367] >gi|306482665|ref|NP_001182361.1|髓性白血病因子1同种型3[智人]
MLKEVLQREGKSYKSETLMYIKKARASENKLSESLAHRENMRQMIRSFSEPFGRDLLSISDGRGRAHN
RRGHNDGEDSLTATSCSLVPFGDFGGMHTDVSSFQTMQMVSNMRNYMQKLERNFGQLSVDPNGHS
- [0368] FCSSSVMTYSKIGDEPPKVFQASTQTRRAPGGIKETRKAMRDSGSGLEKMAIGHHHIDRAHVIKSKNK
KTGDEEVNQEFINMNESDAHAFDEEWQSEVLKYKPGRHNLGNTRMRSVGHENPGSRELKRREKPPQS
PAIEHGRRS NVLGDKLHIKGSSVKSNNK
- [0369] >gi|283945459|ref|NP_001164653.1|着丝粒蛋白L同种型2[智人]
MDSYSAPESTPSASSRPEDYFIGATPLQKRLESVRKQSSFILTPPRRKIPQCSQLQEDVDPQKVAFLHK
QWTLYSLTPLYKFSYSNLKEYSRLLNAFIVAEEKQKGLAVEVGEDFNKIVIFSTLLGMKGTQRDPEAFLV
- [0370] QIVSKSQLPENREGKVLWTGWFCVFGDSLLETVSEDFTCLPLFLANGAESNTAIIGTWFQKTFDCYFS
PLAINAFNLSWMAAMWTACKMDHYVATTEFLWSVPCSPQSLDISFAIHPEDAKALWDSVHKTPGEVT
QEEV DLFMDCLYSHFHRHFKIHLSATRLVRVSTSVASAHTDGKIKILCHKYLIGVLAYLTELAIFQIE
- [0371] >gi|187608508|ref|NP_001120653.1|着丝粒蛋白L同种型1[智人]
- [0372] MDSYSAPESTPSASSRPEDYFIGATPLQKRLESVRKQSSFILTPPRRKIPQCSQLQEDVDPQKVAFLHK
QWTLYSLTPLYKFSYSNLKEYSRLLNAFIVAEEKQKGLAVEVGEDFNKIVIFSTLLGMKGTQRDPEAFLV
QGLILSPRLEYSGLTILVDCNLCLLGSSDPSTLAFQVAGTAGACHHTRIVSKSQLPENREGKVLWTGW
- [0373] CCVFGDSLLETVSEDFTCLPLFLANGAESNTAIIGTWFQKTFDCYFSPPLAINAFNLSWMAAMWTACKM
DHYVATTEFLWSVPCSPQSLDISFAIHPEDAKALWDSVHKTPGEVTQEEVDLFMDCLYSHFHRHFKIHL
SATRL VRVSTSVASAHTDGKIKILCHKYLIGVLAYLTELAIFQIE
- [0374] >gi|74732954|sp|Q9BS16.1|CENPK_人RecName:全称=着丝粒蛋白K;缩写=CENP-K;AltName:全称=间期着丝粒复合蛋白37;AltName:全称=蛋白AF-5a;AltName:全称=p33

- [0375] MNQEDLDPDSTTDVGDVTNTEELIRECEEMWKDMEECQNKLSLIGTETLTDSNAQLSLIMQVKCLT
AELSQQWQKKTPEIPLTEDVLITLGKEEFQKLRQDLEMVLSTKESKNEKLKEDLEREQRWLDEQQQIM
ESLNVLHSELKNKVETFSERIFNELKTKMLNIKEYKEKLLSTLGEFLEDHFPLPDRSVKKKKKNIQESS
VNLI TLHEMLEILINRLFDVPHDPYVKISDSFWPPYVELLLRNGIALRHPEDPTRIRLEAFHQ
- [0376] >gi|315221159|ref|NP_001186732.1|着丝粒蛋白0同种型2[智人]
MAGILASGLGVLAHLERLETQVSRSRKQSEELQSVQAQEGALGTKIHKLRRLRDELRAVVRHRRASVK
ACIANVEPNQTVEINEQEALEEKLENVKAILQAYHFTGLSGKLTSGVCVCISTAFEGNLLDSYFVDLVI
- [0377] QKPLRIHHHSVPVFIPLEEIAAKYLQTNIQHFLFSLCEYLNAYSGRKYQADRLQSDFAALLTGPLQRNPL
CNLLSFTYKLDPGGQSFPFCARLLYKDLTATLPTDVTVTTCQGVEVLSTSWEEQRASHETLFC TKPLHQV
FAS FTRKGEKLDMSLVS
- [0378] >gi|13236565|ref|NP_077298.1|着丝粒蛋白0同种型1[智人]
MEQANPLRPDGESKGGVLAHLERLETQVSRSRKQSEELQSVQAQEGALGTKIHKLRRLRDELRAVVRH
RRASVKACIANVEPNQTVEINEQEALEEKLENVKAILQAYHFTGLSGKLTSGVCVCISTAFEGNLLDS
- [0379] YFVDLVIQKPLRIHHHSVPVFIPLEEIAAKYLQTNIQHFLFSLCEYLNAYSGRKYQADRLQSDFAALLTG
PLQRNPLCNLLSFTYKLDPGGQSFPFCARLLYKDLTATLPTDVTVTTCQGVEVLSTSWEEQRASHETLFC
TKPL HQVFASFTRKGEKLDMSLVS
- [0380] >gi|444299651|ref|NP_653091.3|蛋白CASC5同种型2[智人]
MDGVSSEANEENDNIERPVRRRHSSILKPPRSPLQDLRGGERVQESNALRNKKNSRRVSFADTIKVFQ
- [0381] TESHMKIVRKSEMEETETGENLLLIQNKKLEDNYCEITGMNTLLSAPIHTQMQQKEFSIIHTRERKHAN

DQTVIFSDENQMDLTSSHTVMITKGLLDNPISSEKSTKIDTTSFLANLKLHTEDSRMKKEVNFVSVDQNTSS
ENKIDFNDFIKRLKTGKCSAFPDPDKENFEIPIYSKEPNSSASTHQMHVSLKEDENNSNITRLFREKDD
GMNFTQCHTANIQTLIPTSSETNSRESKGNIDITYGNDFMDLTFNHTLQILPATGNFSEIENQTNAMDV
TTGYGTKASGNKTVFKSKQNTAFQDLSINSADKIHITRSHIMGAETHIVSQTNCQDARILAMTPESIYSN
PSIQGCKTVFYSSCNDAMEMTKCLSNMREEKNLLKHDSNYAKMYCNPDAMSSLTEKTIYSGEENMDI
TKSHTVAIDNQIFKQDQSNVQIAAAPTPEKEMMLQNLMTTSEDGKMNVNCNSVPHVSKERIQQSLSNP
LSISLTDRKTELLSGENMDLTESHTSNLGSQVPLAAYNLAPESTSESHSQSKSSSDECEEITKSRNEPFQR
SDIIAKNSLTDWTWNKDKDWVLKILPYLDKDSQPQADCNQEIATSHNIVYCGGVLDKQITNRNTVSWEQ
SLFSTTKPLFSSGQFSMKNHDTAISSHTVKSVLGQNSKLAELPRKSLSNPTPDYCHDKMIICSEEEQNMD
LTKSHTVIGFGPSELQELGKTNLEHTTGQLTTMNRQIAVKVEKCGKSPIEKSGVLKSNCIMDVLEDES
VQKPKFPKEKQNVKIWGRKSVGGPKIDKTIVFSEDDKNDMDITKSYTIEINHRPILLEKRDCHLVPLAGT
SETILYTCRQDDMEITRSHTTALECKTVSPDEITTRPMDKTVVFDNHVELEMTESHTVFIDYQEKERT
DRPNFELSQRKSLGTPTVICTPTEESVFFPGNGESDRLVANDSQLTPLEEWSNNRGPVEVADNMELSKS
ATCKNIKDVQSPGFLNEPLSSKSQRRKSLKLNKDKTIVFSENHKNMDMITQSCMVEIDNESALEDKEDF
[0382] HLAGASKTILYSCGQDDMEITRSHTTALECKTLLPNEIAIRPMDKTVLFTDNYSDELVTDSHTVFIDCQA
TEKILEENPKFGIGKGNLGVSPKDNQVQIEAEKQALAVGNKIVLHTEQKQQLFAATNRTTNEIIFH
SAAMDEKVIGKVVDQACTLEKAQVESCQLNNRDRRNVDTSSHATAVCGSSDNYSCLPNVISCTDNLE
GSAMLLCDKDEEKANYCPVQNDLAYANDFASEYYLESEGQPLSAPCPLEKEEVIQTSTKGQLDCVITL
HKDQDLIKDPRNLLANQTLVYSQDLGEMTKLNSKRVSFKLPKDQMKVYVDDIYVIPQPHFSTDQPPLP
KKGQSSINKEEVILSKAGNKSNIENSSAPICENKPKILNSEEWFAAACKKELKENIQTNTNYNTALDFHS
NSDVTQKVIQTHVNAGEAPDPVITSNVPCFHSIKPNLNNLNGKTGEFLAFQTVHLPPLPEQELLEGNKA
HNDMHIVQATEIHNINISSNAKDSRDEENKKSNGAETTSPPKTVFKDKVRRCSLGFPLRPLPNKRNK
SVTGIDDLEQIPADTTDINHLETQPVSSKDSGIGSVAGKLNLSQYINEENLPVYPDEINSSDSINIETEEK
ALIETYQKEISPYENKMGKTCNSQKRTWVQEEEDIHKEKKIRKNEIKFSDTTQDREIFDHHTTEEDIDKSA
NSVLIKNLSRTPSSCSSSLDSIKADGTSLDFTYRSSQMESQFLRDTICEESLREKLQDGRITIREFFILLQV
HILIQKPRQSNLPGNFTVNTPTPEDLMLSQYVYRPKIQIYREDCEARRQKIEELKLSASNQDKLLVDIN
KNLWEKMRHCSDKELKAFGIYLNKIKSCFTKMTKVFTHQGKVALYGKLVQSAQNEREKLQIKIDEMD
KILKKIDNCLTEMETETKNLEDEEKNPVVEWDSEMRAAEKELEQLKTEEEELQRNLLLEVQKEQTL
AQIDFMQKQRNRTEELLDQLSLSEWDVVEWSDDQAVFTFVYDTIQLTITFEESVVGFPFLDKRYRKIVD
VNFQSLLEDQAPPSSLLVHKLIFQYVEEKESWKKCTTQHQLPKMLEEFLSVVHHCRLLGEEIEYLKR
[0383] WGPNYNLMNIDINNELLRLFFSSSAFAKFEITLFLSAYYPSVPLPSTIQNHVGNTSQDDIATILSKVPLE
NNYLKNVVKQIYQDLFQ DCHFYH

[0384] >gi|444299649|ref|NP_733468.3|蛋白CASC5同种型1[智人]

[0385]

MDGVSSEANEENDNIERPVRRRHSSILKPPRSPLQDLRGGERVQESNALRNKKNRRVVSFADTIKVFQ
TESHMKIVRKSEMEGCSAMVPSQLQLLPFGFKRFSCSLPETETGENLLLIQNKKLEDNYCEITGMNTLL
SAPIHTQMQQKEFSIIEHTRERKHANDQTVIFSDENQMDLTSSHTVMITKGLLDNPISEKSTKIDTTSFLA
NLKLHTEDSRMKKEVNFVSDQNTSSSENKIDFNDFIKRLKTGKCSAFPDPDKENFEIPIYSKEPNSASST
HQMHSVSLKEDENNSNITRLFREKDDGMNFTQCHTANIQTLIPTSSSETNSRESKGNIDITYGNDFMDLTF
NHTLQILPATGNFSEIENQTQNAMDVTTGYGTKASGNKTVFKSKQNTAFQDLSINSADKIHITRSHIMG
AETHIVSQTCTNCDARILAMTPESIYSNPSIQGCKTVFYSSCNDAMEMTKCLSNMREEKNLLKHDSNYA
KMYCNPDAMSSLTEKTIYSGEENMDITKSHTVAIDNQIFKQDQSNVQIAAAPTEPEKEMMLQNLMTTSE
DGKMNVCNSVPHVSKERIQQSLSNPLSISLTDRKTELLSGENMDLTESHTSNLGSQVPLAAYNLAPES
TSEHSQSKSSSDECEEITKSRNEPFQSRDIIAKNSLTDWTKDKDWVLKILPYLDKDS PQSADCNQEIA
TSHNIVYCGGVLDKQITNRNTVSWEQSLFSTTKPLFSSGQFSMKNHDTAISSHTVKSVLGQNSKLAEPL
RKSLSNPTPDYCHDKMIICSEEEQNMDLTKSHTVVIGFGPSELQELGKTNLEHTTGQLTTMNRQIAVKV
EKCGKSPIEKSGVLKSNCIMDVLEDESQKPKFPKEKQNVKIWGRKSVGGPKIDKTIVFSEDDKNDMDI
TKSYTIEINHRPILLEKRDCHLVPLAGTSETILYTCRQDDMEITRSHTTALECKTVSPDEITTRPMDKT VVF
VDNHVELEMTESHTVFIDYQEKERTDRPNFELSQRKSLGTPTVICTPTEESVFFPGNGESDRLVANDSQL
TPLEEWSNNRGPVEVADNMELSKSATCKNIKDVQSPGFLNEPLSSKSQRRKSLKLNKDKTIVFSENHKN
DMDITQSCMVEIDNESALEDKEDFHLGASKTILYSCGQDDMEITRSHTTALECKTLLPNEIAIRPMDKT
VLFTDNYSDLEVTDSTVFIDCQATEKILEENPKFGIGKGKNLGVSPKDNSCVQEIAEKQALAVGNKI
VLHTEQKQQLFAATNRTTNEIIFHSAAMDEKVGKVVDQACTLEKAQVESCQLNNRDRRNVDFTSSH
ATAVCGSSDNYSCLPNVISCTDNLEGSAMLLCDKDEEKANYCPVQNDLAYANDFASEYYLESEGQPLS
APCPLLEKEEVIQTSTKGQLDCVITLHKDQDLIKDPRNLLANQTLVYSQDLGEMTKLNSKRVSFKLPKD
QMKVYVDDIYVIPQPHFSTDQPPLPKKGQSSINKEEVILSKAGNKS LNIIENSSAPICENKPKILNSEEF
AAACKKELKENIQTNTNYNTALDFHSNSDVTKQVIQTHVNAGEAPDPVITSNVPCHFHSIKPNLNNLNGKT
GEFLAFQTVHLPPLPEQLLELGNKAHNDMHIVQATEIHNINIISSNAKDSRDEENKKS HNGAETTS LPPK
TVFKDKVRRCSLGIFLRLPNKRNC SVTGIDDLEQIPADTTDINHLETQPVSSKDSGIGSVAGKLNLSPSQ
YINEENLPVYPDEINSSDSINIETEEKALIETYQKEISPYENKMGKTCNSQKRTWVQEEEDIHKEKKIRKN

[0386]

EIKFSDTTQDREIFDHHTEEDIDKSANSVLKLNLSRTPSSCSSSLDSIKADGTSLDFSTYRSSQMESQFLRD
TICEESLREKLQDGRITIREFFILLQVHILIQKPRQSNLPGNFTVNTPTPEDLMLSQYVYRPKIQIYREDCE
ARRQKIEELKLSASNQDKLLVDINKNLWEKMRHCSDKELKAFGIYLNKIKSCFTKMTKVFTHQGKVAL
YGKLVQSAQNEREKLQIKIDEMDKILKKIDNCLTEMETETKNLEDEEKNNPVEEWDSEMRAAEKELEQ
LKTEEEELQRNLELEVQKEQTLAQIDFMQKQRNRTEELLDQLSLSEWDVVEWSDDQAVFTFVYDTIQ
LTITFEESVVGFPFLDKRYRKIVDVNFQSLLEDQAPPSSLLVHKLIFQYVEEKESWKKTCCTTQHQLPK
MLEEFSLVVHHCRLLGEEIEYLKRWGPNYNLMNIDINNELLRLFFSSSAFAKFEITLFLSAYYPSVPLP
STIQNHVGNTSQDDIA TILSKVPLENNYLKNVVKQIYQDLFQDCHFHYH

[0387]

>gi|119612836|gb|EAW92430.1|癌症易感性候选基因5,同种型CRA_b[智人]
MDGVSSEANEENDNIERPVRRRHSSILKPPRSPQLDLRGGERVQESNALRNKKNSRRVSFADTIKVFQ
TESHMKIVRKSEMEETETGENLLLIQNKKLEDNYCEITGMNTLLSAPIHTQMQQKEFSIIEHTRERKHAN
DQTVIFSDENQMDLTSSHTVMITKGLLDNPSEKSTKIDTTSFLANLKLHTEDSRMKKEVNFSDQNTSS
ENKIDFNDFIKRLKTGKCSAFPDPDKENFEIPIYSKEPNASSTHQMHSVSLKEDENNSNITRLFREKDD
GMNFTQCHTANIQTLIPTSSETNSRESKGN DITIYGNDFMDLTFNHTLQILPATGNFSEIENQTQNAMDV
TTGYGTKASGNKTVFKSKQNTAFQDLSINSADKIHITRSHIMGAETHIVSQTNCQDARILAMTPESIYSN
PSIQGCKTVFYSSCNDAMEMTKCLSNMREEKNLLKHDSNYSKMYCNPDMAMSSLTEKTIYSGEENMDIT
KSHTVAIDNQIFKQDQSNVQIAAAPTPEKEMMLQNLMTTSEDGKMNVNCNSVPHVSKERIQQSLSNPL
SISLTDKTELLSGENMDLTESHTSNLGSQVPLAAYNLAPESTSESHSQSKSSSDECEEITKSRNEPFQRS
DIIAKNSLTDTWNKDKDWVLKILPYLDKDSPQSADCNQEIATSHNIVYCGGVLDKQITNRNTVSWEQS
LFSTTKPLFSSGQFSMKNHDTAISHTVKSVLGQNSKLAELRKSLSNPTPDYCHDKMIICSEEEQNMDL
TKSHTVVIGFGPSELQELGKTNLEHTTGQLTTMNRQIAVKVEKCGKSPIEKSGVLKSNCIMDVLEDES
QKPKFPKEKQNVKIWGRKSVGGPKIDKTIVFSEDDKNDMDITKSYTIEINHRPILLEKRDCHLVPLAGTSE
TILYTCGQDDMEITRSHTTALECKTVSPDEITTRPMDKTVVFDNVHVELEMESHTVFIDYQEKERTDR
PNFELSQRKSLGTPTVICTPTEESVFFPGNGESDRLVANDSQLTPLEWSNNRGPVEVADNMELSKSAT
CKNIKDVQSPGFLNEPLSSKSQRRKSLKLNKDKTIVFSENHKNMDITQSCMVEIDNESALEDKEDFHL
AGASKTILYSCGQDDMEITRSHTTALECKTLLPNEIAIRPMDKTVLFTDNYS DLEVTDSTHTVFIDCQATE
KILEENPKFGIGKGKNLGVSPKDNQSCVQIEAEKQALAVGNKIVLHTEQKQQLFAATNRTTNEIIFHSA
AMDEKVIGKVVDQACTLEKAQVESCQLNNRDRRNVDFTSSHATAVCGSSDNYSCLPNVISCTDNLEGS
AMLLCDKDEEKANYCPVQNDLAYANDFASEYYLESEGQPLSAPCPLEKEEVIQTSTKGQLDCVITLH

[0388]

KDQDLIKDPRNLLANQTLVYSQDLGEMTKLNSKRVSFKLPKDQMKVYVDDIYVIPQPHFSTDQPPLPK
KGQSSINKEEVILSKAGNKS LNIIENSSAPICENKPKILNSEEWF AAACKKELKENIQT TNYNTALDFHSN
SDVT KQVIQTHVNAGEAPDPVITSNVPCFHSIKPNLNNLNGKTGEFLAFQTVHLPPLPEQLLELGNAH
[0389] NDMHIVQATEIHNINIISNAKDSRDEENKKSHNGAETTS LPPKTVFKDKVRRCSLGIFLPRLPNKRNC
VTGIDDLEQIPADTTDINHLETQPVSSKDSGIGSVAGKLNLSPSQYINEENLPVYPDEINSSDSINIETEEK
ALIETYQKEISPYENKMGKTCNSQKRTWVQEEEDIHKEKKIRKNEIKFSDDTTQDREVSSVLNQRMFLNF
GFCFVFLNCGYSQILILVSGRQKIIIST

[0390] >gi | 119612835 | gb | EAW92429.1 | 癌症易感性候选基因5, 同种型CRA_a, 部分[智
人]

MDGVSSEANEENDNIERPVRRRHSSILKPPRSPLQDLRGGERVQESNALRNKKNSRRVSFADTIKVFQ
TESHMKIVRKSEMEETETGENLLLIQNKKLEDNYCEITGMNTLLSAPIHTQMQQKEFSIIEHTRERKHAN
DQTVIFSDENQMDLTSSHTVMITKGLLDNPISEKSTKIDTTSFLANLKLHTEDSRMKKEVNFSDQNTSS
ENKIDFNDFIKRLKTGKCSAFPDPDKENFEIPIYSKEPNSASSTHQMHSVSLKEDENNSNITRLFREKDD
GMNFTQCHTANIQT LIPTSSETNSRESKGN DITIYGNDFMDLTFNHTLQILPATGNFSEIENQTQNAMDV
TTGYGTKASGNKTVFKSKQNTAFQDLSINSADKIHITRSHIMGAETHIVSQT CNQDARILAMTPESIYSN
PSIQGCKTVFYSSCNDAMEMTKCLSNMREEKNLLKHDSNYSKMYCNP DAMSSLTEKTIYSGEENMDIT
KSHTVAIDNQIFKQDQSNVQIAAAPTPEKEMMLQNLMTTSEDGKMNVNCNSVPHVSKERIQQSLSNPL
SISLTD RKT ELLSGENMDLTESHTSNLGSQVPLAAYNLAPESTS ESHSQSKSSSDECEEITKSRNEPFQRS
DIIAKNSLTD TWNKDKDWVLKILPYLDKDS PQSADCNQEIATSHNIVYCGGVLDKQITNRNTVSWEQS
[0391] LFSTTKPLFSSGQFSMKNHDTA ISSHTVKS VLGQNSKLA EPLRKSLSNPTPDYCHDKMIICSEEEQNMDL
TKSHTV VIGFGPSELQELGKT NLEHTTGQLTTMNRQIAVKVEKCGKSPIEKSGVLKSNCIMDVLEDES
QKPKFPKEKQNVKIWGRKSVGGPKIDKTIVFSEDDKNDMDITKSYTIEINHRP LLEKRDCHLVPLAGTSE
TILYTCGQDDMEITRSHTTALECKTVSPDEITTRPMDKTVVFDNHVELEMTESHTVFIDYQEKERTDR
PNFELSQRKSLGTPTVICTPTEESVFFPGNGESDRLVANDSQLT PLEWSNNRGPVEVADNMELSKSAT
CKNIKDVQSPGFLNEPLSSKSQRRKSLKLNDKTIVFSENHKNDMDITQSCMVEIDNESALEDKEDFHL
AGASKTILYSCGQDDMEITRSHTTALECKTLLPNEIAIRPMDKTVLFTDNYS DLEVTDSTV FIDCQATE
KILEENPKFGIGKGKNLGVSPKDN SCVQEIAEKQALAVGNKIVLHTEQKQQLFAATNRTTNEIKFHSA
AMDEKVIGKVVDQACTLEKAQVESCQLNNRDRRNVDF TSSHATAVCGSSDNYSCLPNVISCTDNLEGS
AMLLCDKDEEKANYCPVQNDLAYANDFASEYYLESEGQPLSAPCP LLEEVIQTSTKGQLDCVITLH
KDQDLIKDPRNLLANQTLVYSQDLGEMTKLNSKRVSFKLPKDQMKVYVDDIYVIPQPHFSTDQPPLPK

- KGQSSINKEEVILSKAGNKS LNIIENSSAPICENKPKILNSEEWFAAACKKELKENIQT TNYNTALDFHSN
SDVTKQVIQTHVNAGEAPDPVITSNVPCFHSIKPNLNNLNGKTGEFLAFQTVHLPLPEQLLELGNAH
NDMHIVQATEIHNINIISNAKDSRDEENK KSHNGAETTS LPPKTVFKDKVRRCSLGIFLPRLPNKRNC S
VTGIDDLEQIPADTTDINHLETQPVSSKDSGIGSVAGKLNLSPSQYINEENLPVYPDEINSSDSINIETEEK
ALIETYQKEISPYENKMGKTCNSQKRTWVQEEEDIHKEKKIRKNEIKFSDTTQDREIFDHHT EEDIDKSA
[0392] NSVLIKNLSRTPSSCSSLSDSIKADGTS LDFSTYRSSQMESQFLRDTICEESLREKLQDGRITIREFFILLQV
HILIQKPRQSNLPGNFTVNTPTPEDLMLSQYVYRPKIQIYREDCEARRQKIEELKLSASNQDKLLVDIN
KNLWEKMRHCS DKELKAFGIYLNKIKSCFTKMTKVFTHQGKVALYGKLVQSAQNEREKLQIKIDEMD
KILKKIDNCLTEMETETKNLEDEEKNNPVEEWDSEMRAAEKELEQLKTEEEELQRNLL ELEVQKEQTL
AQIDFMQKQRNRTEELLDQLSLSEWDVVEWSDDQAVFTFVYDTIQLTITFEESVVGFPFLDKRYRKIVD
VNFQSLLEDQAPPSSL LVHKLIFQYVEEKESWKKCTCTQHQL
[0393] >gi|110349759|ref|NP_056286.3|动粒相关蛋白NSL1同源异构体1[智人]
MAGSPELVLDPPWDKELAAGTESQALVSATPREDFRVRCTSKRAVTEMLQLCGRFVQKLG DALPEEI
REPALRDAQWTFESAVQENISINGQAWQEASDNCFMDS DIKVLEDQFDEIIVDIATKRKQYPRKILECVI
[0394] KTIKAKQEILKQYHPVVHPLDLKYDPDPAPHMENLKC RGETVAKEISEAMKSLPALIEQGEGFSQVLR
MQPVIHLQRIHQEVFSSCHRKPDAPENFITQIETPTETASRKTSDMV LKRKQTKDCPQRKWYPLRPK
KINLD T
[0395] >gi|110349761|ref|NP_001036014.1|动粒相关蛋白NSL1同源异构体2[智人]
MAGSPELVLDPPWDKELAAGTESQALVSATPREDFRVRCTSKRAVTEMLQLCGRFVQKLG DALPEEI
REPALRDAQWTFESAVQENISINGQAWQEASDNCFMDS DIKVLEDQFDEIIVDIATKRKQYPRKILECVI
[0396] KTIKAKQEILKQYHPVVHPLDLKYDPDPVLNGNAFNFS PFNMMLAVDLSYMFITSSPSY GKFEMQRG
NSSK GDQ
[0397] >gi|223972618|ref|NP_079194.3|动粒相关蛋白DSN1同源异构体1[智人]
MTSVTRSEIIDEKGPVMSKTHDHQLESSLSPVEVFAKTSASLEMNQGVSEERIHLGSSPKKGGNCDLSH
QERLQSKSLHLS PQEQSASYQDRRQSWRRASMKETNRRKSLHPIHQGITELSR SISVDLAESKRLGCLLL
[0398] SSFQFSIQKLEPFLRDTKGFSLESFRAKASSLSEELKHFADGLETDGTLQKCFEDSNGKASDFSLEASVAE
MKEYITKFSLERQTWDQLLLHYQQAKEILSRGSTEAKITEVKVEPMTYLGSSQNEVLNTPDYQKILQ
NQSKVFDCMELVMDELQGSVKQLQAFMDESTQCFQKVS VQLGKRSMQQLDPSARKLLKLQLQNPP
[0399] AIHGS GSGSCQ
[0400] >gi|223890153|ref|NP_001138790.1|动粒相关蛋白DSN1同源异构体3[智人]

- MSKTHDQHLESSLSPVEVFAKTSASLEMNQGVSEERIHGSSPKKGGNCDLSHQERLQSKSLHLSPQEQ
SASYQDRRQSWRRASMKETNRRKSLHPIHQGITELSRISISVDLAESKRLGCLLLSSFQFSIQKLEPFLRDT
- [0401] KGFSLESFRAKASSLSEELKHFADGLETGTLQKCFEDSNGKASDFSLEASVAEMKEYITKFSLERQTW
DQLLLHYQQEAKEILSRGSTEAKITEVKVEPMTYLGSSQNEVLNTPDYQKILQNQSKVFDCMELVMD
ELQ GSVKQLQAFMDESTQCFQKVSQVLGKRSMQQLDPSARKLLKLQLQNPPAIHGSGSGSCQ
- [0402] >gi|223890151|ref|NP_001138789.1|动粒相关蛋白DSN1同源异构体2[智人]
MTSVTRSEIIDELSRISISVDLAESKRLGCLLLSSFQFSIQKLEPFLRDTKGFSLESFRAKASSLSEELKH
FADGLETGTLQKCFEDSNGKASDFSLEASVAEMKEYITKFSLERQTDQLLLHYQQEAKEILSRGSTE
- [0403] AKITEVKVEPMTYLGSSQNEVLNTPDYQKILQNQSKVFDCMELVMDELQGSVKQLQAFMDESTQCF
QKVS VQLGKRSMQQLDPSARKLLKLQLQNPPAIHGSGSGSCQ
- [0404] >gi|74735330|sp|014777.1|NDC80_人RecName:全称=动粒蛋白NDC80同源物;
AltName:全称=在癌症中高度表达的蛋白;AltName:全称=动粒蛋白Hec1;缩写=HsHec1;
AltName:全称=动粒相关蛋白2;AltName:全称=视网膜母细胞瘤相关蛋白HEC
MKRSSVSSGGAGRLSMQELRSQDVNKQGLYTPQTKEKPTFGKLSINKPTSERKVSFLGKRTSGHGSRN
SQLGIFSSSEKIKDPRPLNDKAFIQQCIRQLCEFLTENGYAHNVSMKSLQAPSVKDFLKIFTFLYGFLCPS
YELPDTKFEEEVPRIFKDLGYPFALSKSSMYTVGAPHTWPHIVAALVWLIDCIKIHTAMKESSPLFDDGQ
PWGEETEDGIMHNKFLDYTIKCYESFMSGADSFDEMNAELQSKLKDLFNVD AFKLESLEAKNRALNE
QIARLEQEREKEPNRLESRLKASLQGDVQKYQAYMSNLESHSAILDQKLNGLNNEEIARVELECETIK
- [0405] QENTRLQNIIDNQKYSVADIERINHERNELQQTINKLTKDLEAEQQKLWNEELKYARGKEAIETQLAEY
HKLARKLKLIPKAENSKGYDFEIKFNPEAGANCLVKYRAQVYVPLKELLNETEEEINKALNKKMGLE
DTLEQLNAMITESKRSVRTLKEEVQKLDDLYQQKIKEAEEDEKCASELESLEKHKHLESTVNQGLSE
AMNELDAVQREYQLVVQTTTEERRKVGNNLQRLLEMVATHVGSVEKHLEEQIAKVDREYEECMS
LSENIKEIRDKY EKKATLIKSSSE
- [0406] >gi|32698866|ref|NP_872319.1|动粒蛋白Spc24[智人]
MAAFRDIEEVSQGLSLLGANRAEAQQRLLGRHEQVVERLLETQDGAEKQLREILTMEKEVAQSLN
- [0407] AKEQVHQGGVELQQLEAGLQEAGEEDTRLKASLLQLTRELEELKEIEADLERQEKEVDEDTTVTIPSAV
YVA QLYHQVSKIEWDYECEPGMVKGIIHGPSVAQPIHLDSTQLSRKFISDYLSLVDTEW
- [0408] >gi|48146511|emb|CAG33478.1|Spc25[智人]
MVEDELALFDKSINEFWNKFKSTDTSCQMAGLRDITYKDSIKAFAEKLSVKLKEEERMVEMFLEYQNQI
SRQNKLIEKKDNLLKLIAEVKGGKQELEVL TANIQDLKEEYSRKKETISTANKANAERLRLKQKSADL
- [0409] YKDRGLGLEIRKIYGEKLQFIFTNIDPKNPESPFMFLHLNEARDYEVSDSAPHLEGLAEFQENVRKTNNF
SAF LANVRKAFTATVYN
- [0410] >gi|115311829|sp|Q9BZD4.2|NUF2_人RecName:全称=动粒蛋白Nuf2;缩写=
hNuf2;缩写=hNuf2R;缩写=hsNuf2;AltName:全称=细胞分裂周期相关蛋白1

METLSFPRYNVAEIVHIRNKILTGADGKNLTKNDLYPNPKPEVLHMIYMRALQIVYGIRLEHFYMPV
N SEVMYPHLMEGFLPFSNLVTHLDSFLPICRVNDFETADILCPKAKRTSRFLSGHNFHIFREACRETYME
FLWQYKSSADKMQQLNAAHQEALMKLERLDSVPVEEQEEFKQLSDGIQELQQSLNQDFHQKTIVLQE
[0411] GNSQKKSNISEKTKRLNELKLSVVSLKEIQESLTKIVDSPEKLKKNYKEKMKDVTQKLKNARQEVVEK
YEIYGDSVDCLPSCQLEVQLYQKKIQDLSDNREKLASILKESLNLEDQIESDESELKKLKTEENSFKRLM
IVKKEKLATAQFKINKKHEDVKQYKRTVIEDCNKVQEKRGAVYERVTTINQEIQKIKLGIQQLKDAAER
EKLKSQ EIFLNLKTALEKYHDGIEKAAEDSYAKIDEKTAELKRKMFKMST

[0412] 尽管出于清楚理解的目的,通过说明和实例在一定程度上详细描述了前述发明,但是本领域技术人员将理解,可以在所附权利要求的范围内实施某些改变和修改。另外,将本文提供的各参考文献通过引用整体并入,其程度如同将每个参考文献通过引用单独并入。

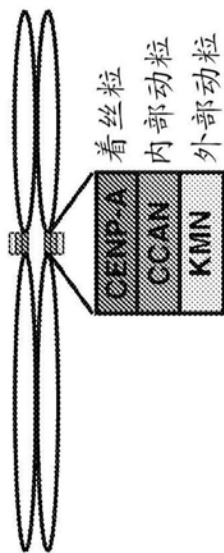


图1A

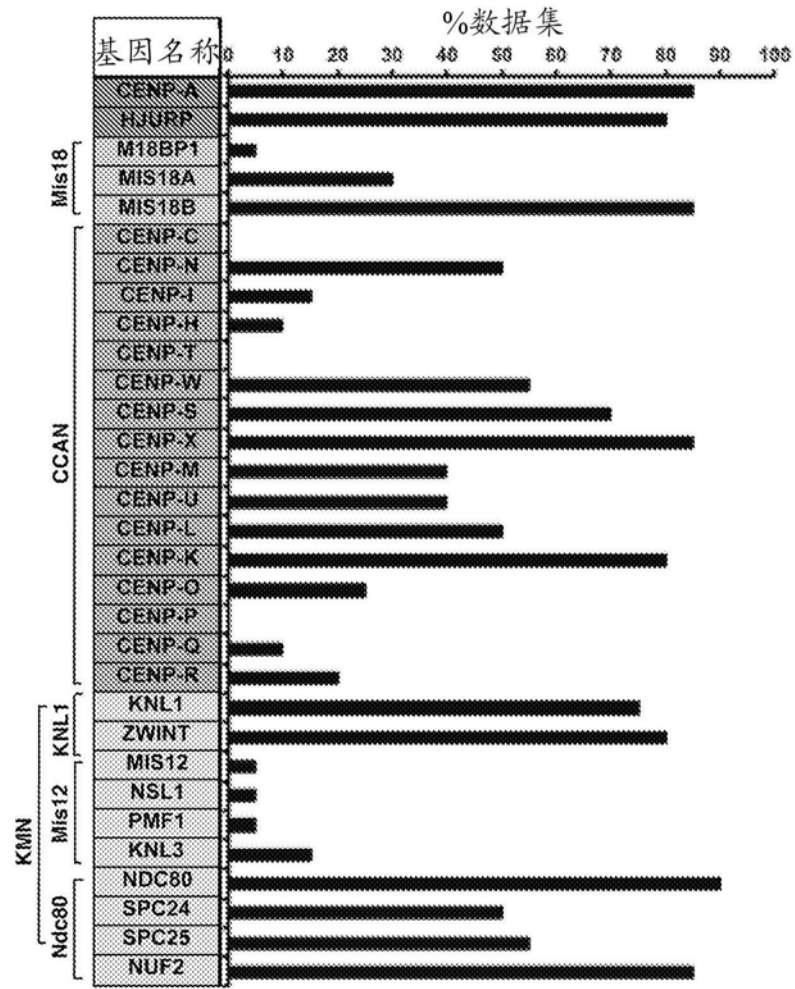


图 1B

图 1C

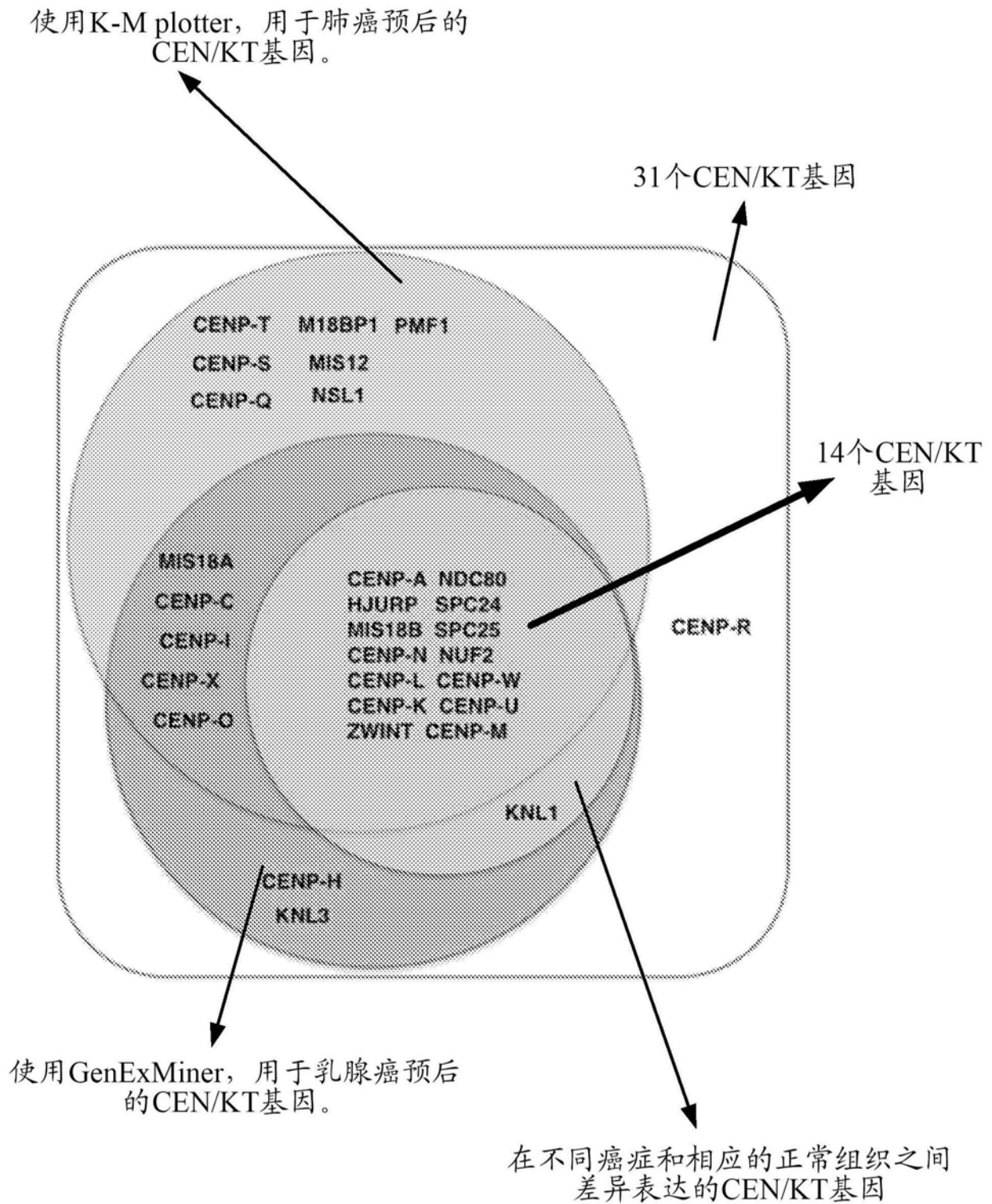


图2

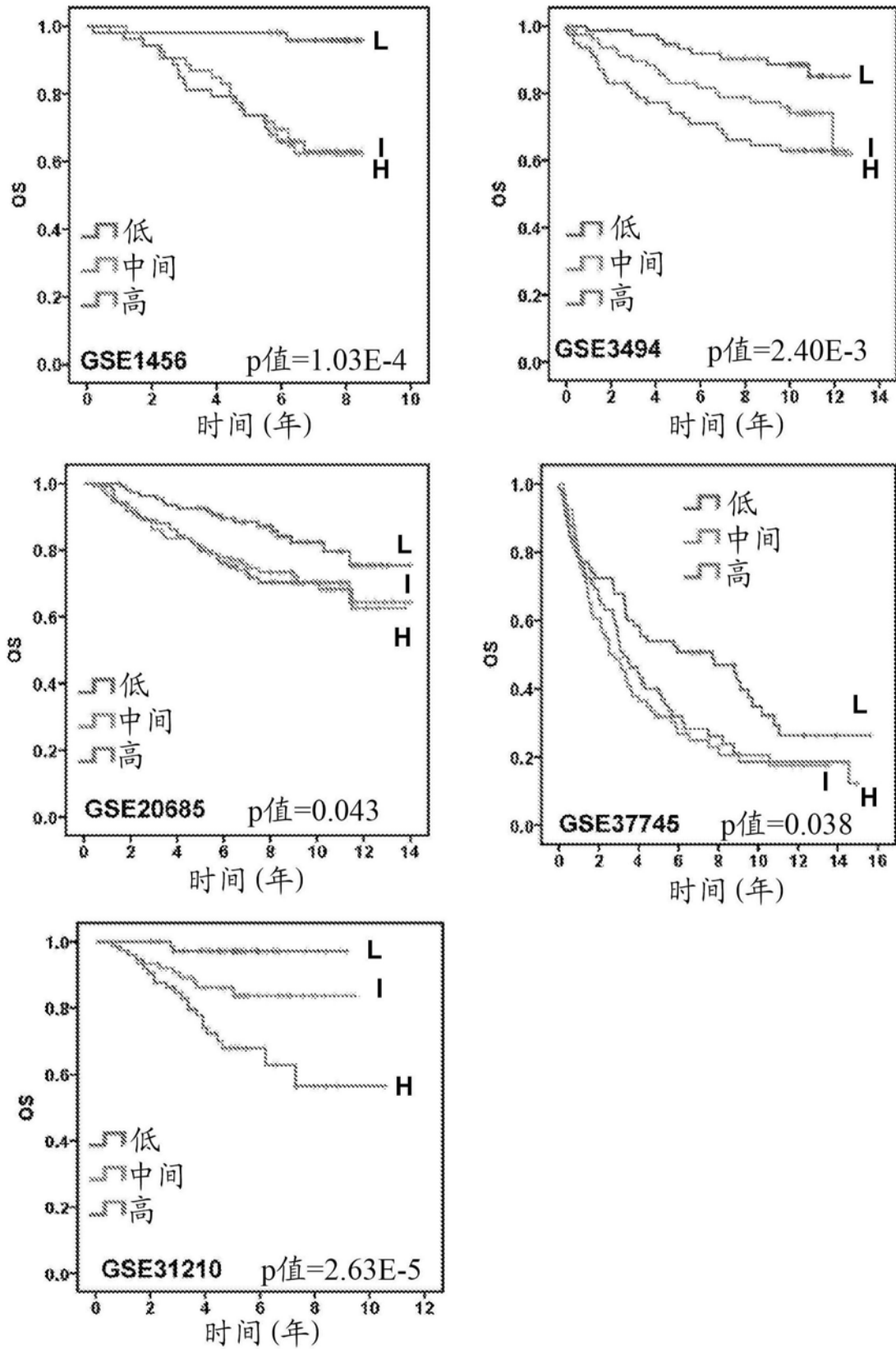


图3A

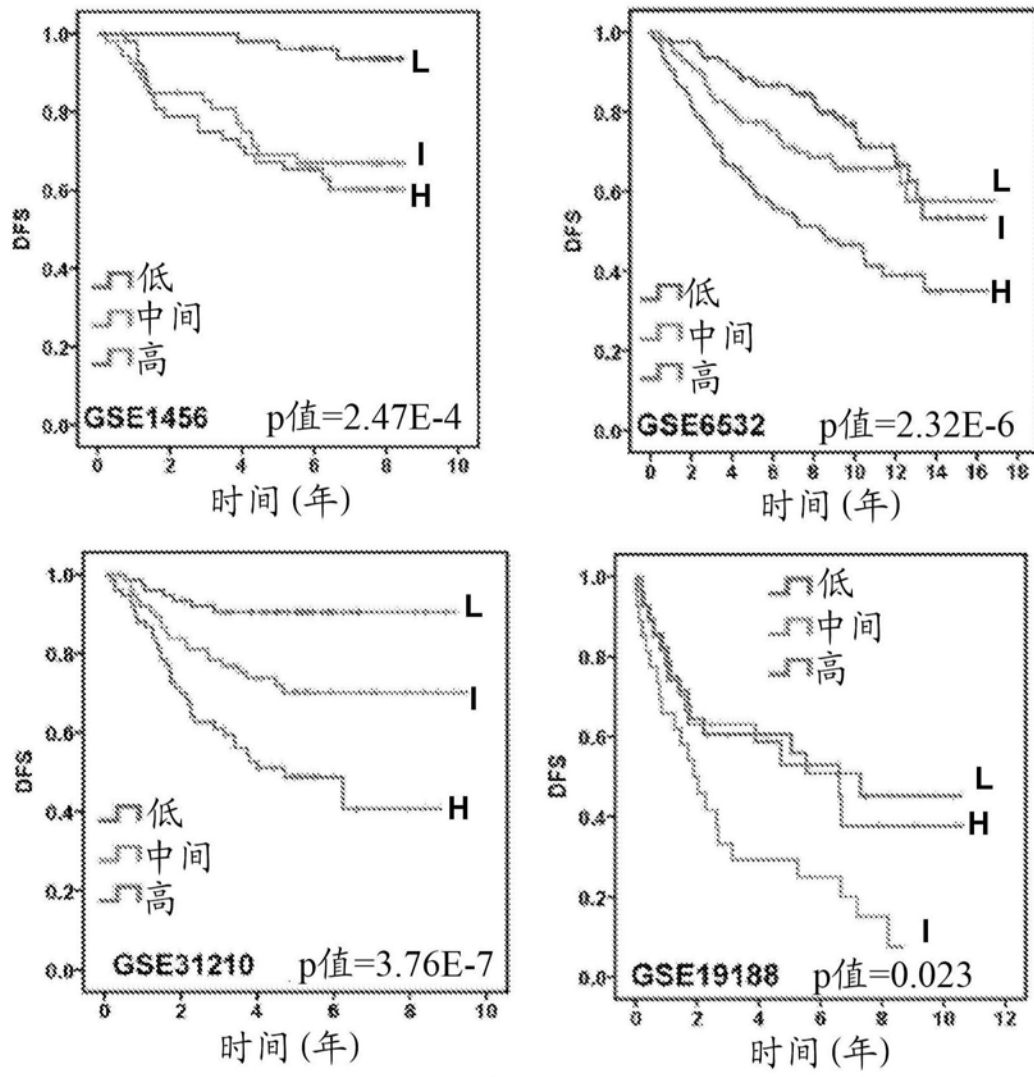


图3B

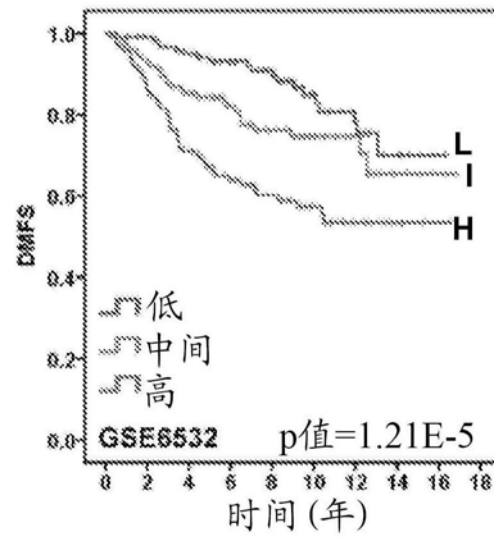


图3C

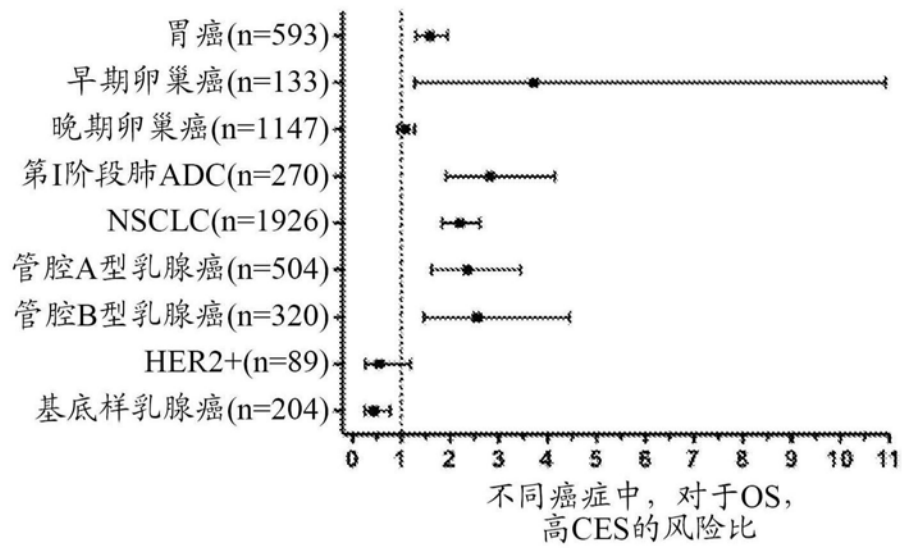


图3D

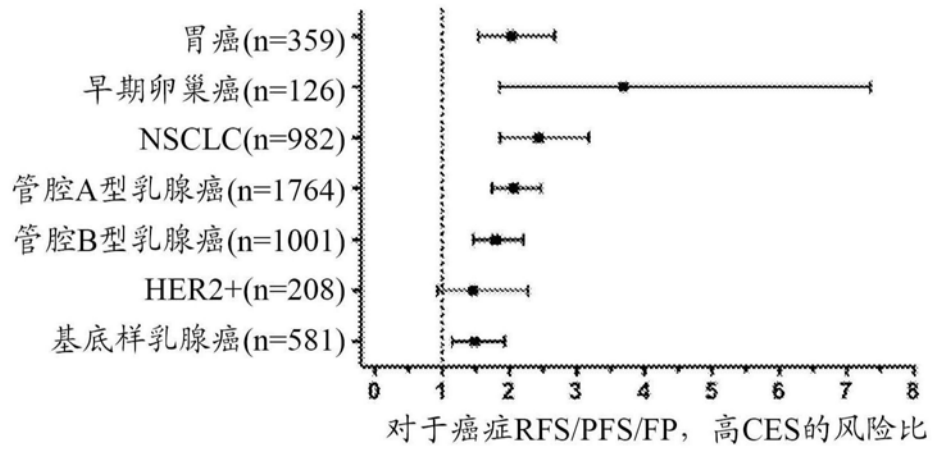


图3E

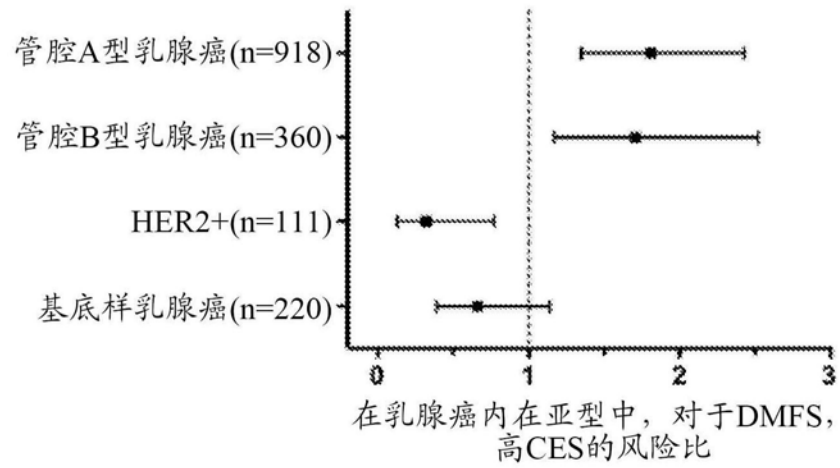


图3F

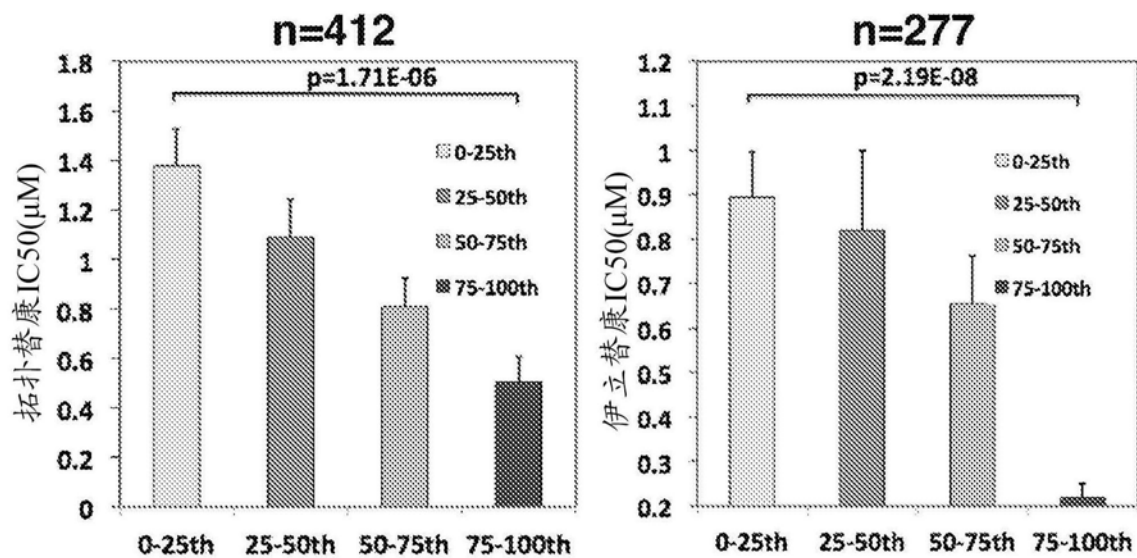


图4A

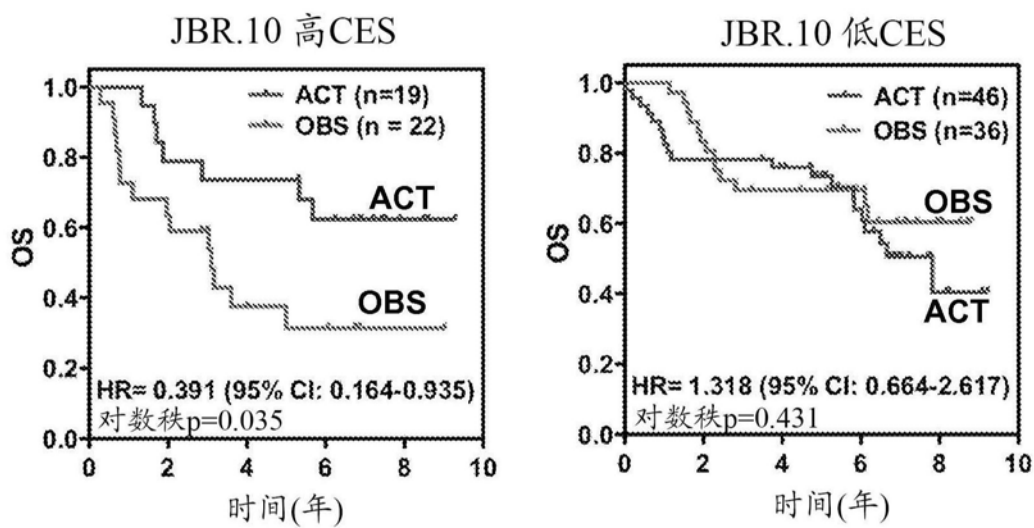


图4B

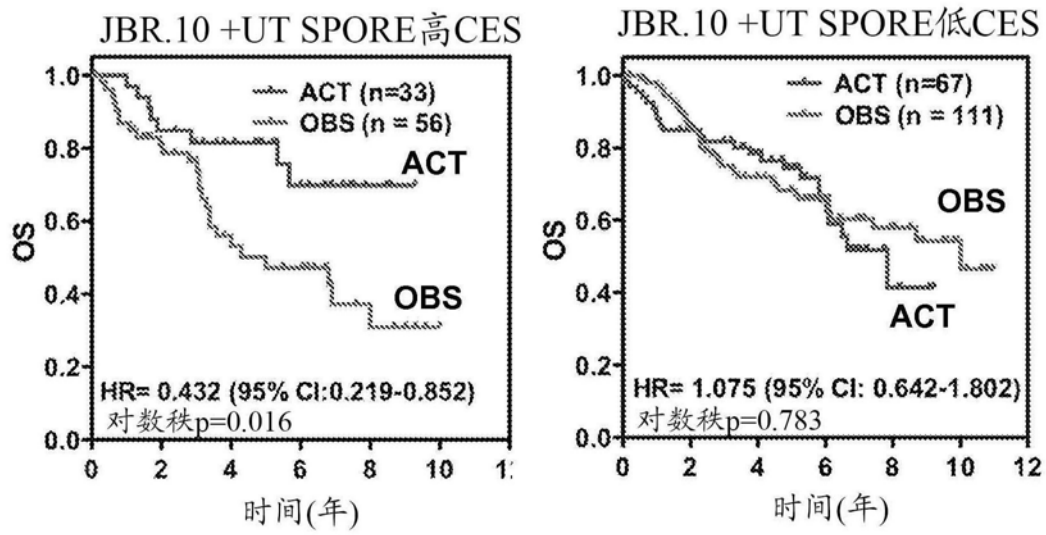


图4C

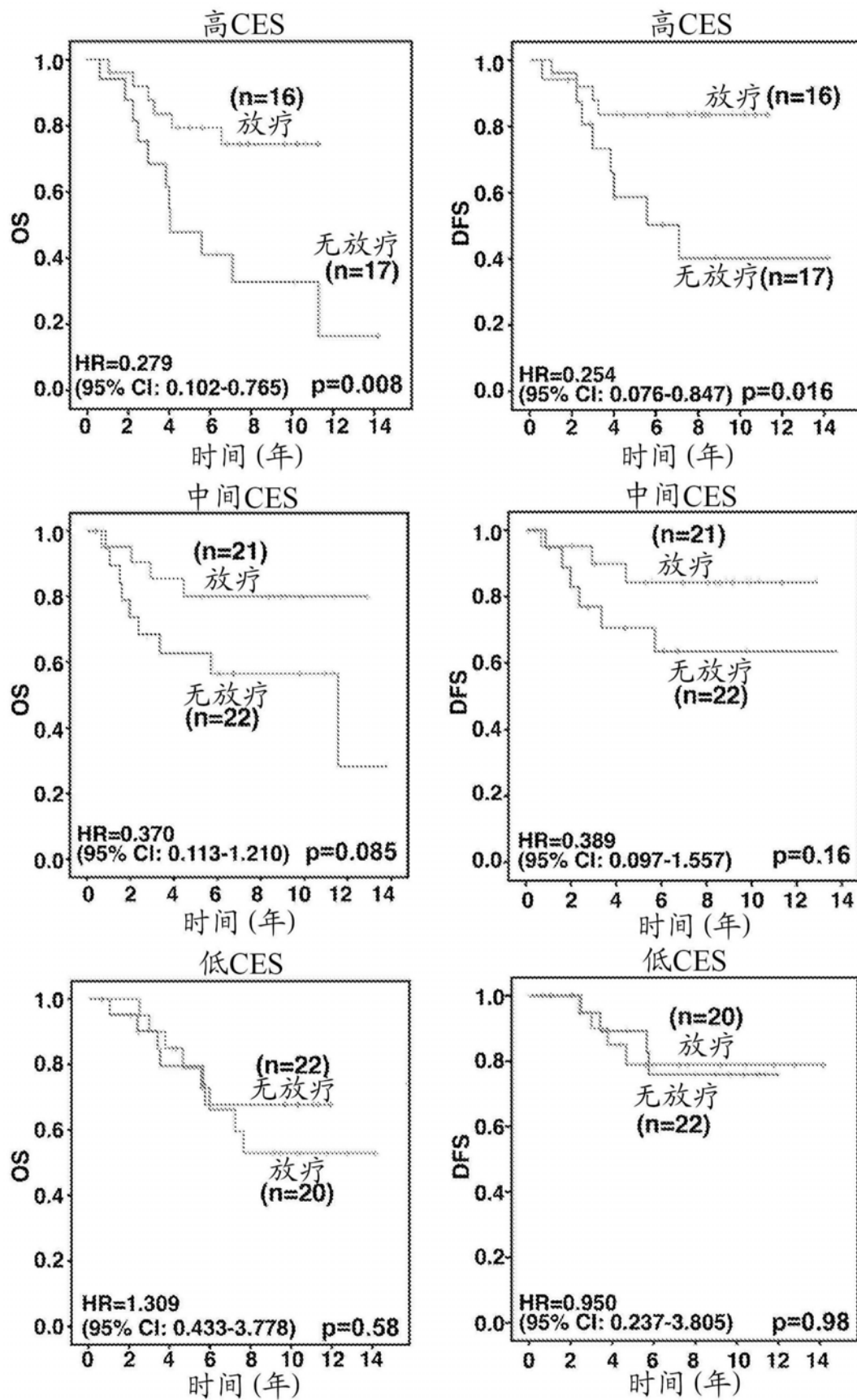


图4D

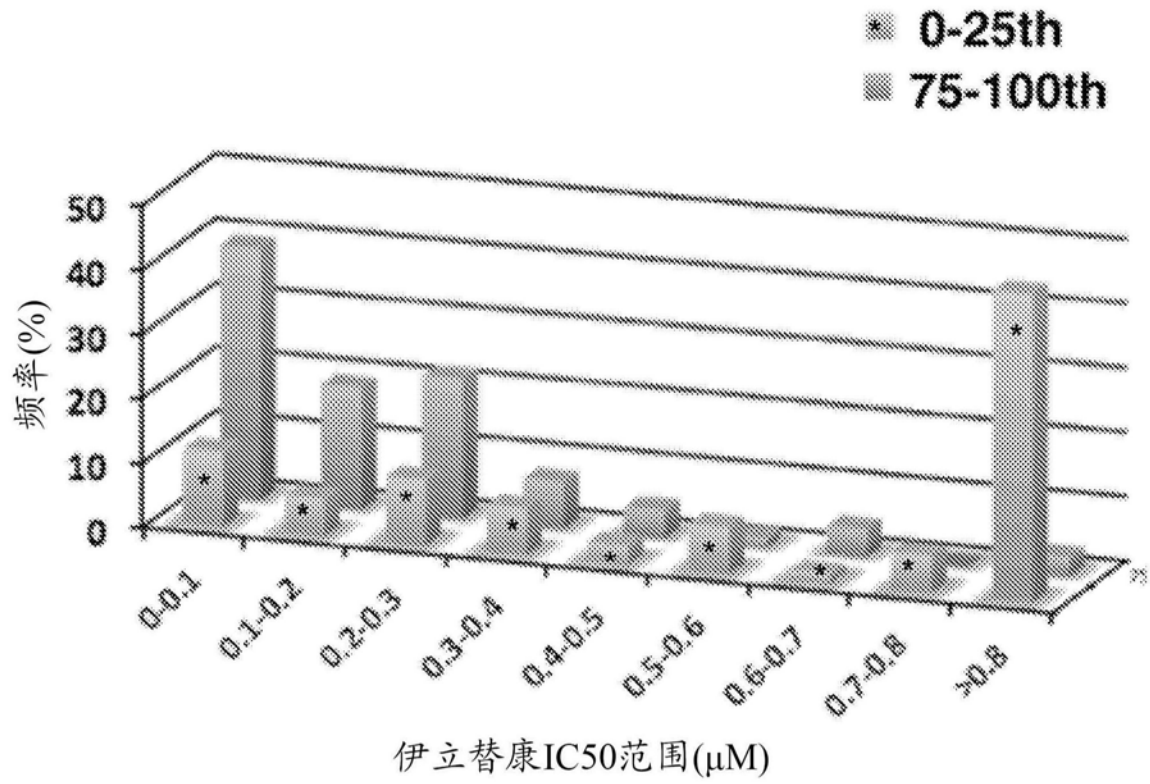


图5A

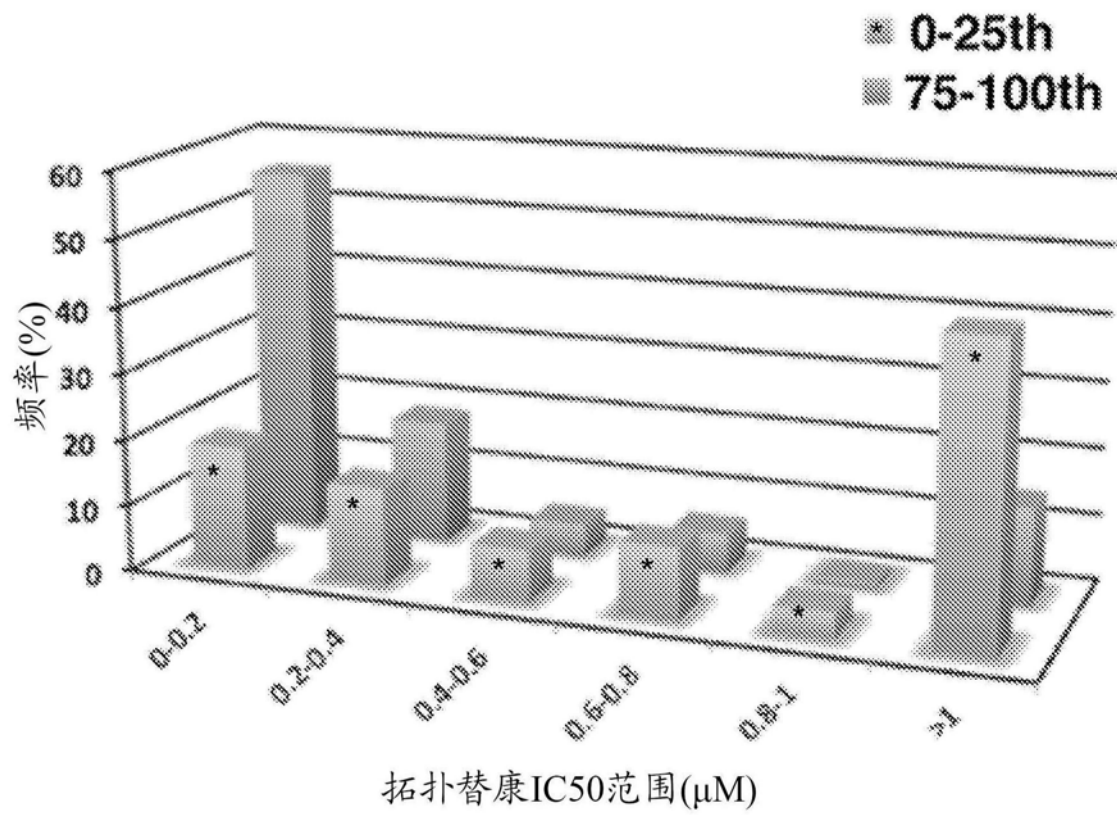


图5B

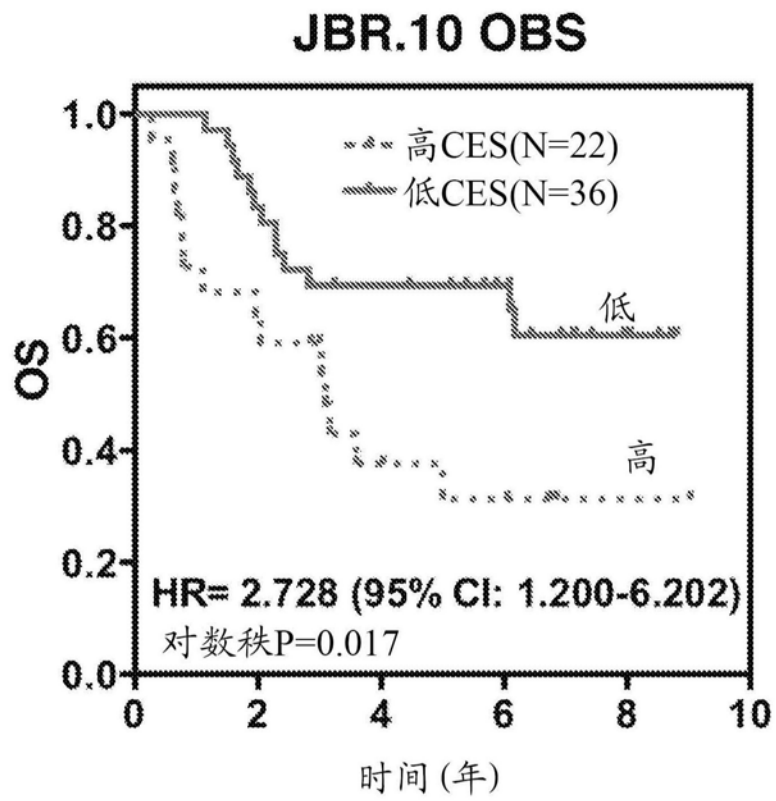


图6A

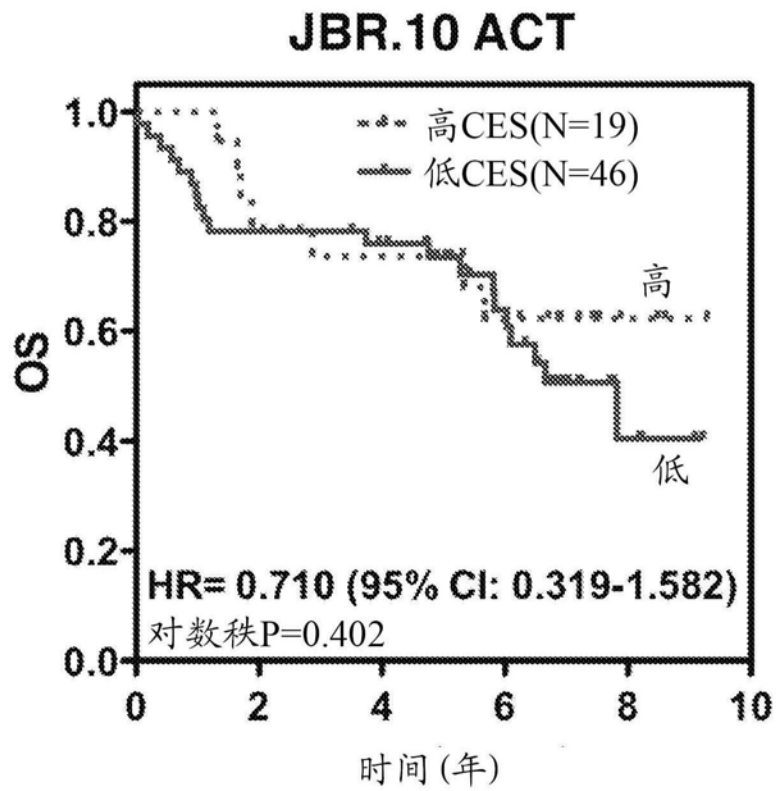


图6B

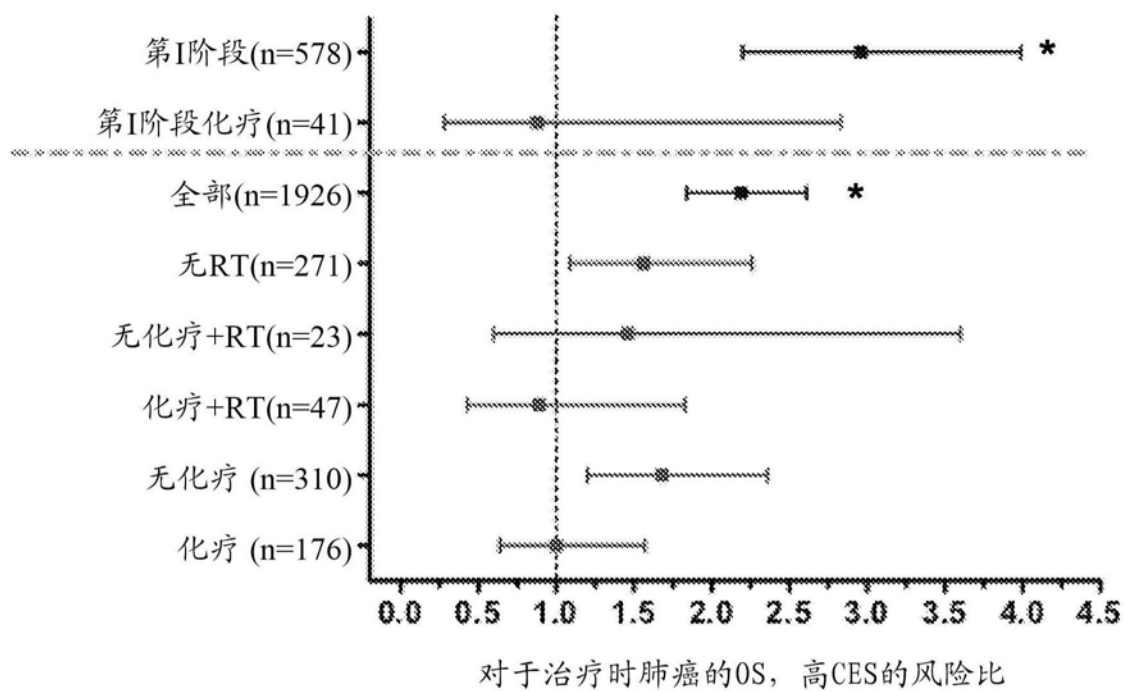


图7A

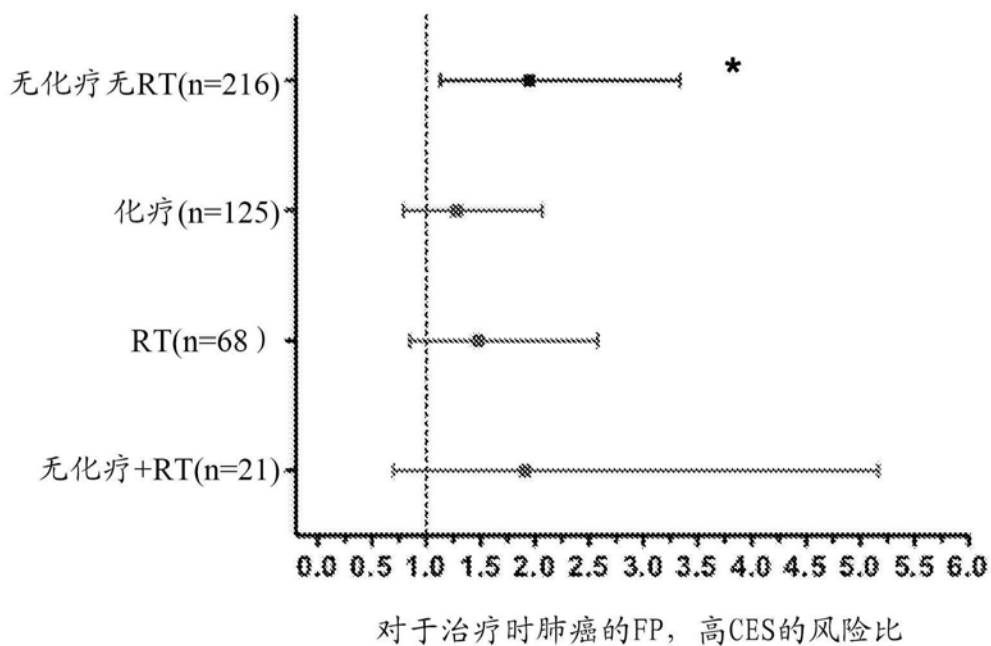


图7B

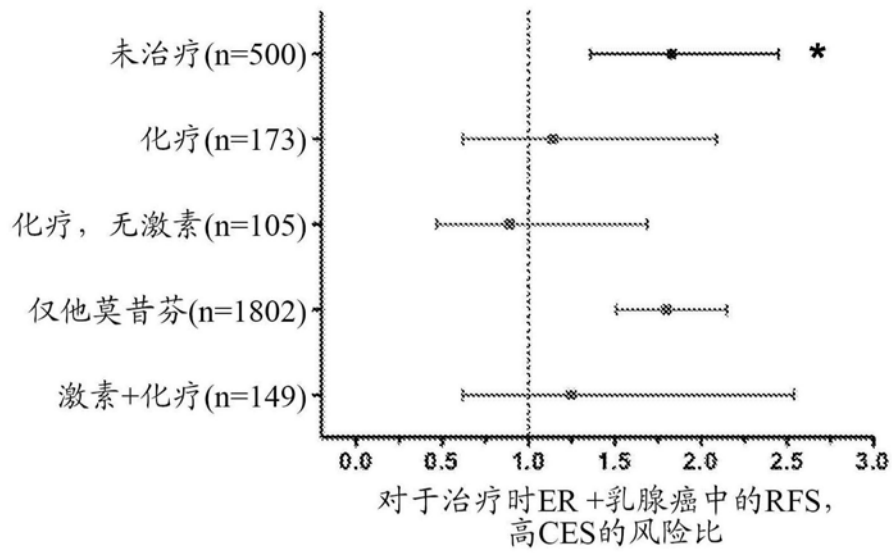


图7C

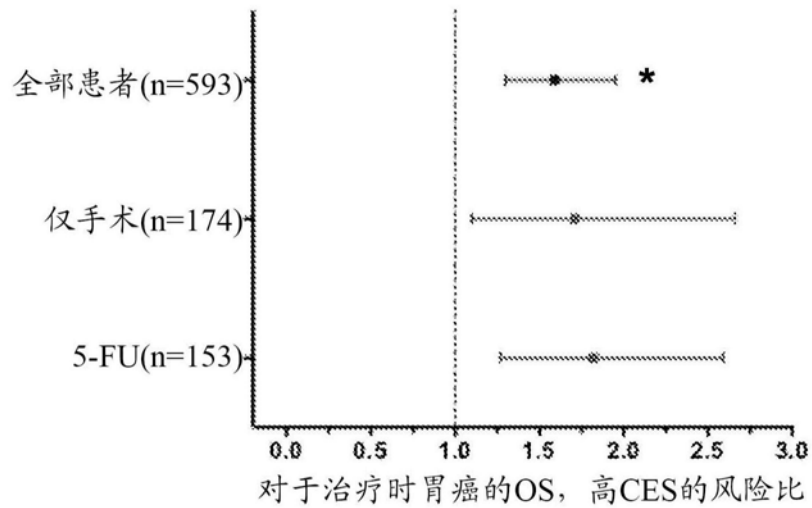


图7D

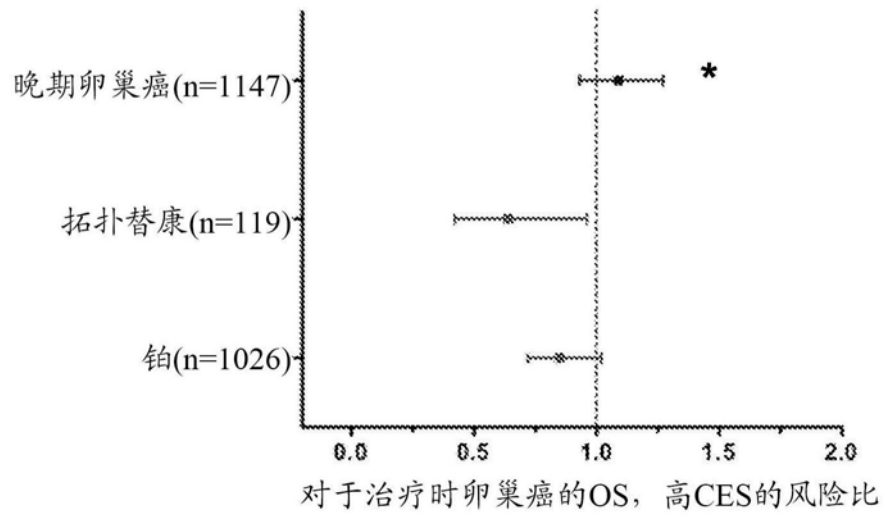


图7E

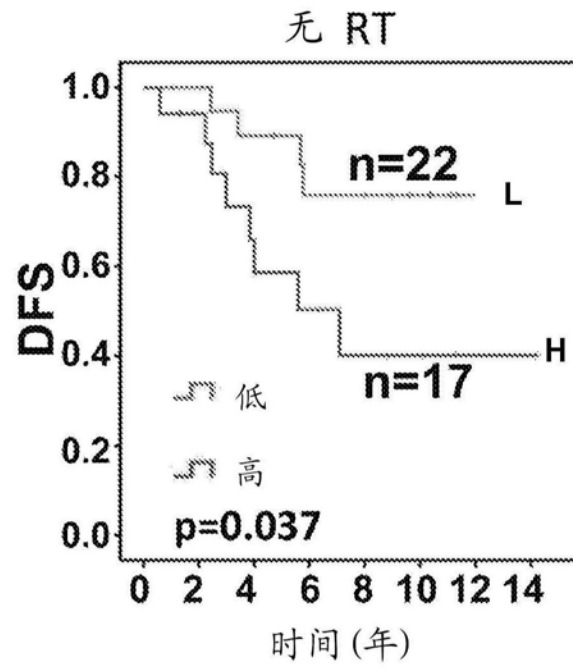


图8A

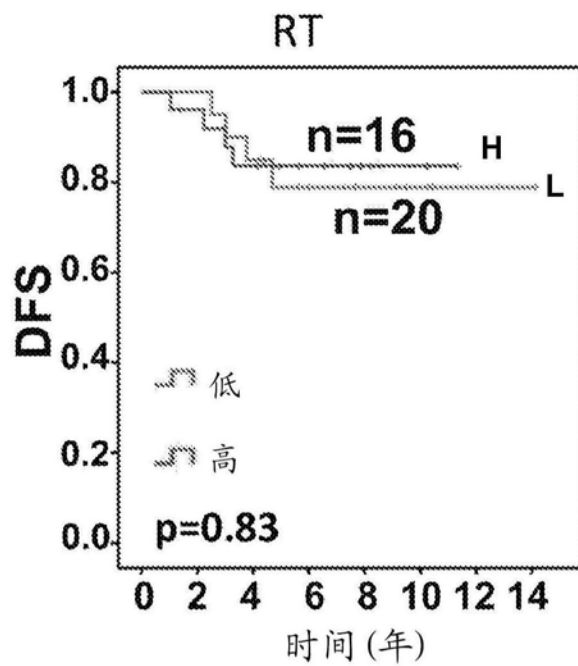


图8B

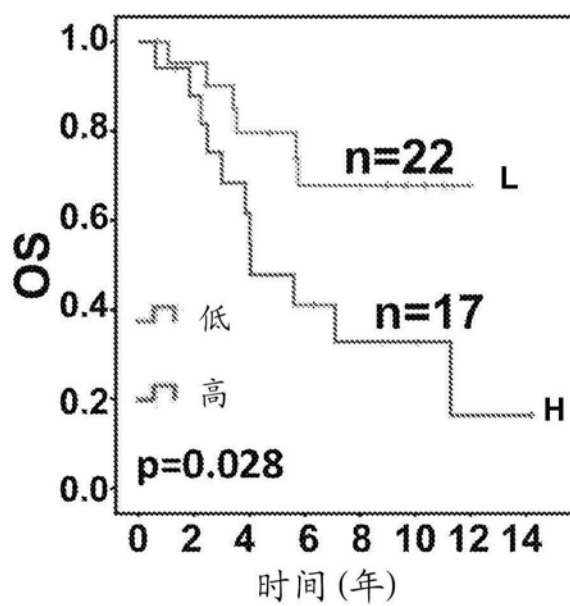


图8C

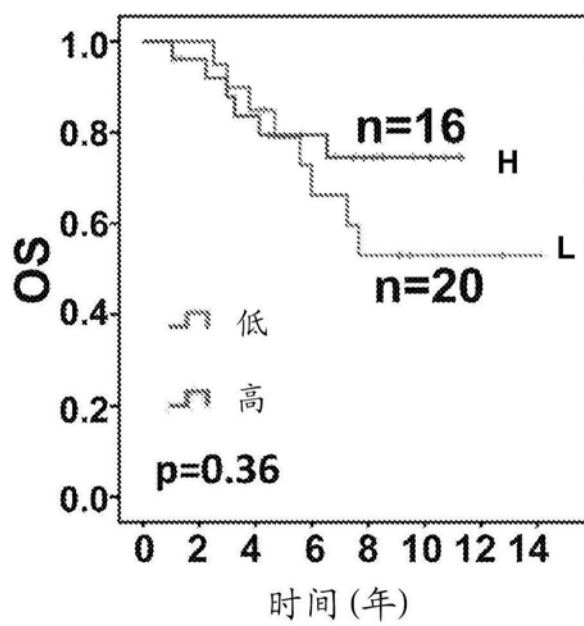


图8D

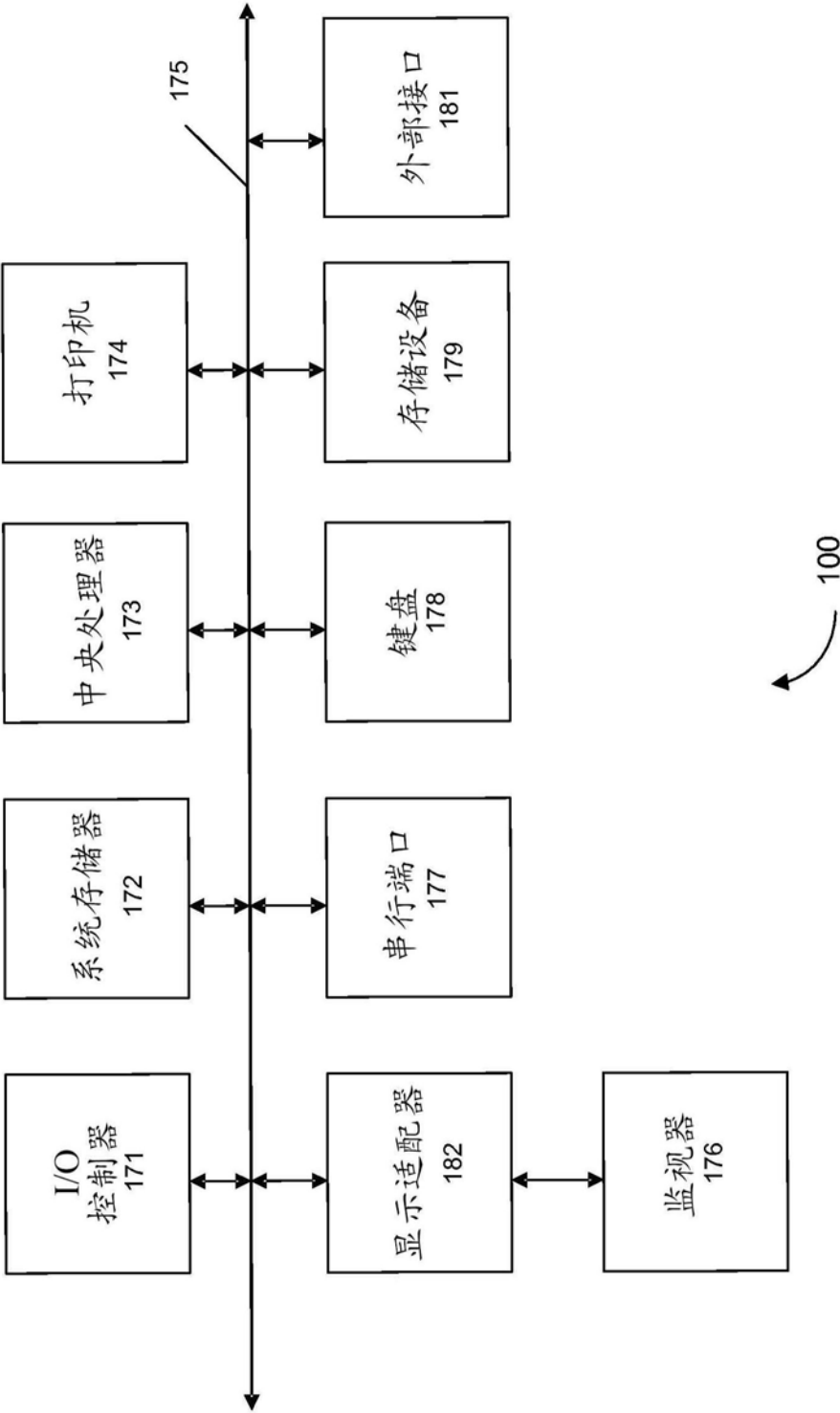


图9