(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 117203348 A
(43)申请公布日 2023.12.08

(21)申请号 202280029777.3
(22)申请日 2022.03.01
(30)优先权数据
63/154,821 2021.03.01 US
(85)PCT国际申请进入国家阶段日
2023.10.20
(86)PCT国际申请的申请数据
PCT/US2022/018313 2022.03.01
(87)PCT国际申请的公布数据
WO2022/187227 EN 2022.09.09
(71)申请人 PFS基因组学公司
地址 美国威斯康星州
(72)发明人 S·L·张 L·J·皮尔斯 C·斯皮尔斯 F·冯 P·马尔姆斯特罗姆 M·弗尔诺 E·霍姆伯格 P·O·卡尔松
(74)专利代理机构 北京市柳沈律师事务所 11105
专利代理师 邹宗亮
(51)Int.Cl.
C12Q 1/68(2018.01)

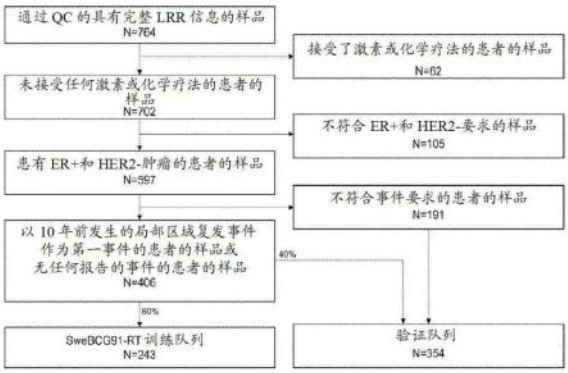
权利要求书3页 说明书42页 附图6页

(54)发明名称

用于乳腺癌预后和鉴定不可能受益于放射疗法的受试者的方法和基因组分类器

(57)摘要

本公开涉及用于提供乳腺癌复发(例如,局部区域复发)的个体化预后评估的系统和方法。所述系统和方法涉及测量来自患者样品的基因表达以产生基因表达签名,所述基因表达签名鉴定不可能受益于乳腺癌手术后放射疗法的受试者。



CN 117203348 A

1.一种方法,所述方法包括:

a)测量来自患有乳腺癌(BC)或处于患乳腺癌(BC)风险的人患者的生物样品中的一种或多种基因的表达水平,其中所述一种或多种基因选自AGR2、B4GALT1、CLDN7、EZR、GNG11、JUN、MMP11、PKIB、PRPS1、PSMD10、SH3BP5、SLC16A3、SLC7A11、SPP1、TNNT1和UBE2E1;以及

b)基于所述一种或多种基因的所述表达水平确定所述患者的BC复发的可能性。

2.如权利要求1所述的方法,

其中AGR2、CLDN7、EZR、MMP11、PKIB、PRPS1、PSMD10、SLC16A3、SLC7A11、SPP1、TNNT1和UBE2E1的表达水平增加各自与乳腺癌复发风险或可能性增加相关;

其中B4GALT1、GNG11、JUN和SH3BP5的表达水平增加各自与乳腺癌复发风险或可能性降低相关。

3.如权利要求1所述的方法,

还包括如果所述患者被表征为BC复发高风险,则用辅助放射疗法治疗所述患者;或者

还包括如果所述患者被表征为BC复发低风险,则不用辅助放射疗法治疗治疗所述患者。

4.如权利要求1所述的方法,其中所述BC复发是局部或局部区域复发或远处复发(转移)。

5.如权利要求1所述的方法,其中所述生物样品是活检或肿瘤样品。

6.如权利要求1所述的方法,其中测量所述表达水平包括进行以下中的一项或多项:原位杂交、基于PCR的方法、基于阵列的方法、免疫组织化学方法、RNA测定方法或免疫测定方法。

7.如权利要求6所述的方法,其中测量所述表达水平包括使用选自由核酸探针、一种或多种核酸引物和抗体组成的组的试剂。

8.如权利要求1所述的方法,其中测量所述表达水平包括测量RNA转录物的水平。

9.一种用于确定患有乳腺癌的患者的预后以及是否用辅助放射疗法治疗所述患者的试剂盒,所述试剂盒包含用于测量选自AGR2、B4GALT1、CLDN7、EZR、GNG11、JUN、MMP11、PKIB、PRPS1、PSMD10、SH3BP5、SLC16A3、SLC7A11、SPP1、TNNT1和UBE2E1的一种或多种基因的表达水平的剂。

10.如权利要求9所述的试剂盒,其中所述试剂盒包含用于测量全部所述一种或多种基因的所述表达水平的剂。

11.如权利要求10所述的试剂盒,其中所述剂包括用于进行原位杂交、基于PCR的方法、基于阵列的方法、免疫组织化学方法、RNA测定方法或免疫测定方法的试剂。

12.如权利要求10所述的试剂盒,其中所述剂包括微阵列、核酸探针、核酸引物或抗体中的一者或多者。

13.如权利要求10所述的试剂盒,其中所述试剂盒包含至少一组PCR引物,所述至少一组PCR引物能够扩增包含所述一种或多种基因的序列的核酸。

14.如权利要求10所述的试剂盒,其中所述试剂盒包含至少一种探针,所述至少一种探针能够与包含所述一种或多种基因的序列的核酸杂交。

15.如权利要求10所述的试剂盒,其还包含电子或纸质形式的信息,所述信息包含如何确定患有乳腺癌的受试者的预后以及是否用辅助放射疗法治疗所述受试者的说明书。

16.如权利要求10所述的试剂盒,其还包含一种或多种对照参考样品。

17.一种用于确定患有乳腺癌的受试者的预后以及是否用辅助放射疗法治疗所述受试者的探针组,所述探针组包含用于检测多种靶核酸的多种探针,其中所述多种靶核酸包含选自AGR2、B4GALT1、CLDN7、EZR、GNG11、JUN、MMP11、PKIB、PRPS1、PSMD10、SH3BP5、SLC16A3、SLC7A11、SPP1、TNNT1和UBE2E1的一个或多个基因序列。

18.如权利要求17所述的探针组,其中所述探针组包含用于检测包含所述一种或多种基因的基因序列或其互补序列的多种靶核酸的多种探针。

19.如权利要求17所述的探针组,其中至少一种探针被可检测地标记。

20.一种用于确定患有乳腺癌的受试者的预后以及是否用辅助放射疗法、化学疗法或内分泌疗法治疗所述受试者的试剂盒,所述试剂盒包含如权利要求17所述的探针组。

21.一种系统,所述系统包括:

a)如权利要求17所述的探针组;以及

b)计算机模型或算法,所述计算机模型或算法用于分析来自患有乳腺癌的所述受试者的生物样品中与所述多种探针杂交的所述多种靶核酸的表达水平或表达谱,并基于所述表达水平或表达谱确定所述受试者是否处于癌症复发低风险,并预测所述受试者是否将受益于辅助放射疗法。

22.一种用于生成风险评分以评估患者的BC复发风险的方法,所述方法包括以下步骤:

鉴定从所述BC患者获得的样品中存在的一种或多种基因的基因表达水平,其中所述一种或多种基因选自AGR2、B4GALT1、CLDN7、EZR、GNG11、JUN、MMP11、PKIB、PRPS1、PSMD10、SH3BP5、SLC16A3、SLC7A11、SPP1、TNNT1和UBE2E1;

通过将所述表达水平的值乘以对应于表8中所选择的一种或多种基因的系数值获得加权表达水平;以及

通过将所述一种或多种基因的所述加权表达水平相加生成风险评分。

23.如权利要求22所述的方法,其还包括基于所述风险评分评估所述患者的BC复发风险。

24.一种用于预测患有BC或处于患BC风险的患者的BC复发可能性的方法,所述方法包括:

(a)测量从所述患者获得的样品中一种或多种以下基因的表达水平:AGR2、B4GALT1、CLDN7、EZR、GNG11、JUN、MMP11、PKIB、PRPS1、PSMD10、SH3BP5、SLC16A3、SLC7A11、SPP1、TNNT1和UBE2E1;以及

(b)基于所述一种或多种基因的所述表达水平预测所述患者的BC复发可能性,其中AGR2、CLDN7、EZR、MMP11、PKIB、PRPS1、PSMD10、SLC16A3、SLC7A11、SPP1、TNNT1和UBE2E1的表达增加与BC复发风险增加相关,并且其中B4GALT1、GNG11、JUN和SH3BP5的表达增加与乳腺癌复发风险降低相关。

25.如权利要求24所述的方法,

还包括如果所述患者被表征为BC复发高风险,则用辅助放射疗法治疗所述患者;或者

还包括如果所述患者被表征为BC复发低风险,则不用辅助放射疗法治疗治疗所述患者。

26.如权利要求24所述的方法,其中所述BC复发是局部或局部区域复发或远处复发(转

移)。

27.如权利要求24所述的方法,其中所述样品是活检或肿瘤样品。

28.如权利要求24所述的方法,其中测量所述表达水平包括进行以下中的一项或多项:原位杂交、基于PCR的方法、基于阵列的方法、免疫组织化学方法、RNA测定方法或免疫测定方法。

29.如权利要求28所述的方法,其中测量所述表达水平包括使用选自由核酸探针、一种或多种核酸引物和抗体组成的组的试剂。

30.如权利要求24所述的方法,其中测量所述表达水平包括测量RNA转录物的水平。

用于乳腺癌预后和鉴定不可能受益于放射疗法的受试者的方法和基因组分类器

[0001] 相关申请的交叉引用

[0002] 本申请要求2021年3月1日提交的序列号为63/154,821的美国临时专利申请的优先权,该美国临时专利申请据此全文以引入方式并入本文。

技术领域

[0003] 本公开涉及用于提供乳腺癌复发(例如,局部区域复发)的个体化预后评估的系统和方法。系统和方法涉及测量来自患者样品的基因表达以产生基因表达签名,所述基因表达签名鉴定不可能受益于乳腺癌手术后放射疗法的受试者。

背景技术

[0004] 尽管在乳腺癌的治疗和诊断方面取得了进展,但仍需要改善对乳腺癌复发(例如局部区域复发)的预后评估。此外,需要可预示局部区域复发和放射敏感性的新基因签名。

[0005] 本发明满足了这些需要。

发明内容

[0006] 本公开涉及用于提供乳腺癌(BC)复发(例如,局部区域复发(LRR))的个体化预后评估的系统和方法。这些系统和方法涉及测量来自患者样品的基因表达以产生基因表达签名,该基因表达签名鉴定不太可能受益于乳腺癌手术(BCS)后放射疗法(RT)的受试者。

[0007] 本公开涉及用于受试者所患BC的诊断、预后和治疗的方法、系统和试剂盒。本公开还提供了用于鉴定处于乳腺癌复发低风险并且不太可能受益于辅助放射疗法的受试者的生物标志物和分类器。在某些情况下,本文还公开了用于检测此类生物标志物以确定受试者乳腺癌复发风险的探针组。本公开还提供了用于鉴定处于局部区域复发(LRR)风险的受试者和预测对放射疗法的响应的生物标志物和分类器。还提供了基于表达谱和/或年龄来确定乳腺癌复发风险的乳腺癌治疗方法。

[0008] 在某些实施方案中,本发明提供了方法,其包括:a)测量来自患有乳腺癌(BC)或处于患乳腺癌(BC)风险的人患者的生物样品中的一种或多种基因的表达水平,其中所述一种或多种基因选自AGR2、B4GALT1、CLDN7、EZR、GNG11、JUN、MMP11、PKIB、PRPS1、PSMD10、SH3BP5、SLC16A3、SLC7A11、SPP1、TNNT1和UBE2E1;和b)基于所选择的一种或多种基因的表达水平确定患者BC复发的可能性。在某些实施方案中,该方法可另外包括将所选择的一种或多种基因的所测量的表达水平与符合BC复发风险的预定基因表达水平进行比较,或者可包括将所选择的一种或多种基因的表达水平归一化,例如,以产生所选择的一种或多种基因的归一化表达水平。在另外的实施方案中,确定BC复发的可能性可基于所选择的一种或多种基因的比较的表达水平。在另一个实施方案中,确定BC复发的可能性可基于所选择的一种或多种基因的归一化表达水平。

[0009] 在某些实施方案中,本公开提供了用于预测患有BC或处于患BC风险的患者的BC复

发可能性的方法,其包括:(a)测量从患者获得的样品中以下基因中的一种或多种的表达水平:AGR2、B4GALT1、CLDN7、EZR、GNG11、JUN、MMP11、PKIB、PRPS1、PSMD10、SH3BP5、SLC16A3、SLC7A11、SPP1、TNNT1和UBE2E1;以及(b)基于所述一种或多种基因的表达水平预测患者的BC复发可能性,其中AGR2、CLDN7、EZR、MMP11、PKIB、PRPS1、PSMD10、SLC16A3、SLC7A11、SPP1、TNNT1和UBE2E1的表达增加与BC复发风险增大相关,并且其中B4GALT1、GNG11、JUN和SH3BP5的表达增加与BC复发风险降低相关。在某些实施方案中,该方法可另外包括使所选择的一种或多种基因的表达水平归一化以获得所选择的一种或多种基因的归一化表达水平。在另外的实施方案中,基于所选择的一种或多种基因的归一化表达水平预测BC复发可能性。

[0010] 在一些实施方案中,BC复发是局部或局部区域复发或远处复发(转移)。在一些实施方案中,生物样品是活检或肿瘤样品。

[0011] 如本文所用,"BC复发风险"或"BC复发可能性"是指在延长的时间段(例如,1个月、1年、5年、10年等)内BC复发的统计概率(例如,可能性)。在一些实施方案中,BC复发风险涉及基线,其中患者已成功进行干预(例如,手术干预)并且特征在于不具有BC和/或积极进展的癌细胞。因此,复发风险或可能性涉及癌症以某种方式复发的可能性。此类风险可被表征为极低风险、低风险、中低风险、平均风险、中高风险、高风险和极高风险。

[0012] 在一些实施方案中,AGR2、CLDN7、EZR、MMP11、PKIB、PRPS1、PSMD10、SLC16A3、SLC7A11、SPP1、TNNT1和UBE2E1的表达水平增加各自与乳腺癌复发风险或可能性增加相关。

[0013] 在一些实施方案中,AGR2、CLDN7、EZR、MMP11、PKIB、PRPS1、PSMD10、SLC16A3、SLC7A11、SPP1、TNNT1和UBE2E1的表达水平降低各自与乳腺癌复发风险或可能性降低相关。

[0014] 在一些实施方案中,B4GALT1、GNG11、JUN和SH3BP5的表达水平增加各自与乳腺癌复发风险或可能性降低相关。

[0015] 在一些实施方案中,B4GALT1、GNG11、JUN和SH3BP5表达水平降低各自与乳腺癌复发风险或可能性增加相关。

[0016] 在一些实施方案中,该方法还包括如果患者被表征为BC复发风险或可能性增加或高风险或可能性,则用辅助放射疗法治疗患者。在某些实施方案中,如果患者被表征为处于BC复发中高风险、高风险或极高风险,则可对患者进行辅助放射疗法治疗。在一些实施方案中,该方法还包括如果患者被表征为BC复发风险或可能性降低或低风险或可能性,则不用辅助放射疗法治疗治疗患者。在某些实施方案中,如果患者被表征为处于BC复发中低风险、低风险或极低风险,则患者可以不用辅助放射疗法治疗或被鉴定为不会受益于辅助放射疗法的患者。

[0017] 在一些实施方案中,测量生物样品中所有基因的表达水平。在一些实施方案中,一种或多种基因的数量选自由以下组成的组:1个基因、1-2个基因、1-3、1-4、1-5、1-6、1-7、1-8、1-9、1-10、1-11、1-12、1-13、1-14、1-15、1-16、2-3、2-4、2-5、2-6、2-7、2-8、2-9、2-10、2-11、2-12、2-13、2-14、2-15、2-16、3-4、3-5、3-6、3-7、3-8、3-9、3-10、3-11、3-12、3-13、3-14、3-15、3-16、4-5、4-6、4-7、4-8、4-9、4-10、4-11、4-12、4-13、4-14、4-15、4-16、5-6、5-7、5-8、5-9、5-10、5-11、5-12、5-13、5-14、5-15、5-16、6-7、6-8、6-9、6-10、6-11、6-12、6-13、6-14、6-15、6-16、7-8、7-9、7-10、7-11、7-12、7-13、7-14、7-15、7-16、8-9、8-10、8-11、8-12、8-13、8-14、8-15、8-16、9-10、9-11、9-12、9-13、9-14、9-15、9-16、10-11、10-12、10-13、10-14、10-15、10-16、11-12、11-13、11-14、11-15、11-16、12-13、12-14、12-15、12-16、13-14、13-15、

13-16、14-15、14-16、15-16和16个基因。

[0018] 在一些实施方案中,测量表达水平包括进行以下中的一项或多项:原位杂交、基于PCR的方法、基于阵列的方法、免疫组织化学方法、RNA测定方法或免疫测定方法。在一些实施方案中,测量表达水平包括使用选自由核酸探针、一种或多种核酸引物和抗体组成的组的试剂。在一些实施方案中,测量表达水平包括测量RNA转录物的水平。

[0019] 在某些实施方案中,本公开提供了用于预断和/或预测患有BC的受试者是否受益于辅助放射疗法的方法,该方法包括:a)获得或已获得来自受试者的样品中选自AGR2、B4GALT1、CLDN7、EZR、GNG11、JUN、MMP11、PKIB、PRPS1、PSMD10、SH3BP5、SLC16A3、SLC7A11、SPP1、TNNT1和UBE2E1的一种或多种基因的表达水平;以及b)基于所述表达水平确定受试者处于癌症复发低风险还是高风险,和/或基于所述表达水平确定受试者可能受益于还是不可能受益于辅助放射疗法,从而预断和/或预测受试者是否受益于辅助放射疗法。在一些实施方案中,该方法还包括如果受试者被鉴定为不可能受益于辅助放射疗法和/或正在施用除辅助放射疗法之外的癌症治疗,则停止辅助放射疗法治疗。在其他实施方案中,癌症复发是局部或局部区域复发或远处复发(转移)。在一些实施方案中,与对照相比,一种或多种基因的表达水平可增加或降低。在一些实施方案中,测量生物样品中一种或多种基因中的所有基因的表达水平。在一些实施方案中,该方法还包括用辅助放射疗法治疗受试者。在一些实施方案中,该方法还包括用除辅助放射疗法之外的癌症疗法治疗受试者。在其他实施方案中,该方法还包括用乳房切除术、放射增强或辅助全身性疗法治疗受试者。在一些实施方案中,在乳房保留手术(BCS)后不对受试者进行放射治疗。在一些实施方案中,该方法还包括基于受试者的年龄确定受试者处于癌症复发低风险,或基于受试者的年龄确定受试者不处于癌症复发低风险。在一些实施方案中,癌症复发是局部或局部区域复发或远处复发(转移)。在其他实施方案中,受试者处于癌症复发高风险并用放射疗法治疗。

[0020] 在一些实施方案中,本公开提供了一种方法,其包括:a)获得或已获得来自受试者的样品中选自AGR2、B4GALT1、CLDN7、EZR、GNG11、JUN、MMP11、PKIB、PRPS1、PSMD10、SH3BP5、SLC16A3、SLC7A11、SPP1、TNNT1和UBE2E1的一种或多种基因的表达水平;以及b)基于所述表达水平确定受试者处于癌症复发低风险并且不可能受益于使用辅助放射疗法的治疗,或基于所述表达水平确定受试者不处于癌症复发高风险并且可能受益于使用辅助放射疗法的治疗。在其他实施方案中,该方法还包括如果受试者被鉴定为不可能受益于辅助放射疗法,则施用除辅助放射疗法之外的癌症疗法。在其他实施方案中,癌症复发是局部或局部区域复发或远处复发(转移)。在一些实施方案中,与对照相比,一种或多种基因的表达水平可增加或降低。在一些实施方案中,测量生物样品中选自表5的所有基因的表达水平。在其他实施方案中,该方法还包括用乳房切除术、放射增强或辅助全身性疗法治疗受试者。在一些实施方案中,在乳房保留手术(BCS)后不对受试者进行放射治疗。在一些实施方案中,该方法还包括基于受试者的年龄确定受试者处于癌症复发低风险,或基于受试者的年龄确定受试者不处于癌症复发低风险。在一些实施方案中,癌症复发是局部或局部区域复发或远处复发(转移)。

[0021] 在一些实施方案中,本公开提供了一种治疗受试者乳腺癌的方法,其包括:a)获得或已获得来自受试者的样品中选自AGR2、B4GALT1、CLDN7、EZR、GNG11、JUN、MMP11、PKIB、PRPS1、PSMD10、SH3BP5、SLC16A3、SLC7A11、SPP1、TNNT1和UBE2E1的一种或多种基因的表达

水平;b)基于所述表达水平确定受试者处于癌症复发低风险并且不太可能受益于使用辅助放射疗法的治疗;以及c)如果基于所述表达水平将受试者鉴定为不可能受益于辅助放射疗法,则施用除辅助放射疗法之外的癌症治疗。在其他实施方案中,癌症复发是局部或局部区域复发或远处复发(转移)。在一些实施方案中,与对照相比,一种或多种基因的表达水平可增加或降低。在一些实施方案中,测量生物样品中所有基因的表达水平。在其他实施方案中,该方法还包括用乳房切除术、放射增强或辅助全身性疗法治疗受试者。在一些实施方案中,在乳房保留手术(BCS)后不对受试者进行放射治疗。在一些实施方案中,该方法还包括基于受试者的年龄确定受试者处于癌症复发低风险,或基于受试者的年龄确定受试者不处于癌症复发低风险。在一些实施方案中,癌症复发是局部或局部区域复发或远处复发(转移)。

[0022] 在一些实施方案中,本公开的方法和基因组分类器中使用的多种基因选自由以下组成的组:前梯度2,蛋白二硫键异构酶(AGR2);β-1,4-半乳糖基转移酶1(B4GALT1);密封蛋白7(CLDN7);埃兹蛋白(EZR);G蛋白亚基γ11(GNG11);Jun原癌基因(JUN);基质金属肽酶11(MMP11);CAMP依赖性蛋白激酶抑制剂β(PKIB);磷酸核糖焦磷酸合成酶1(PRPS1);蛋白酶体26S亚基,非ATP酶10(PSMD10);SH3结构域结合蛋白5(SH3BP5);溶质载体家族16成员3(SLC16A3);溶质载体家族7成员11(SLC7A11);分泌型磷蛋白1(SPP1);肌钙蛋白T1,慢骨骼肌型(TNNT1)和泛素结合酶E2 E1(UBE2E1)。

[0023] 在某些实施方案中,受试者患有雌激素受体阳性(ER+)乳腺癌、人表皮生长因子受体2阴性(HER2-)乳腺癌、I-II期乳腺癌或淋巴结阴性乳腺癌和/或是绝经后的。

[0024] 此类方法不限于特定样品或生物样品类型。例如,在一些实施方案中,样品或生物样品是组织样品、体液样品、血液样品、器官分泌物样品、CSF样品、唾液样品、血浆样品、血清样品或尿液样品。在一些实施方案中,样品可包括乳腺组织或周围组织、乳腺活检组织、肿瘤样品或含有乳腺细胞或乳腺癌细胞的组织。

[0025] 在某些实施方案中,在分析之前,从生物样品中分离和/或纯化和/或扩增包含选自表5的基因的序列或其互补序列的核酸。在一些实施方案中,核酸可包含RNA转录物。

[0026] 在其他实施方案中,生物标志物的表达水平通过原位杂交、基于PCR的方法、基于阵列的方法、免疫组织化学方法、RNA测定法或免疫测定法来确定。在其他实施方案中,使用一种或多种试剂确定基因表达水平。在某些实施方案中,一种或多种试剂是核酸探针、核酸引物和/或抗体。在其他实施方案中,确定生物标志物的表达水平包括测量核酸的水平。在一些实施方案中,核酸是RNA转录物。

[0027] 在一些实施方案中,与对照相比,至少一种基因的表达水平降低。在其他实施方案中,与对照相比,至少一种基因的表达水平增加。

[0028] 在某些实施方案中,本文所述的方法在用辅助放射疗法治疗受试者之前进行。在某些实施方案中,本文所述的方法在用乳房切除术、放射增强或辅助全身性疗法治疗受试者之前进行。

[0029] 在其他实施方案中,该方法还包括计算受试者的风险评分,其中如果基于风险评分和生物样品中选自AGR2、B4GALT1、CLDN7、EZR、GNG11、JUN、MMP11、PKIB、PRPS1、PSMD10、SH3BP5、SLC16A3、SLC7A11、SPP1、TNNT1和UBE2E1的一种或多种基因的表达水平,受试者被鉴定为处于癌症复发低风险并且不可能受益于辅助放射疗法,则不对受试者进行辅助放射

疗法,并对受试者施用除辅助放射疗法之外的癌症疗法。在一些实施方案中,癌症复发是局部或局部区域复发或远处复发(转移)。

[0030] 一种或多种生物标志物基因的表达水平的显著性可通过使用例如T检验、P值、KS(Kolmogorov Smirnov)P值、准确度、准确度P值、阳性预测值(PPV)、阴性预测值(NPV)、灵敏度、特异性、AUC、AUC P值(Auc.pvalue)、Wilcoxon检验P值、中位倍数差异(MFD)、Kaplan Meier(KM)曲线、生存AUC(survAUC)、Kaplan Meier P值(KM P值)、单变量分析比值比P值(uvaORPval)、多变量分析比值比P值(mvaORPval)、单变量分析风险比P值(uvaHRPval)和多变量分析风险比P值(mvaHRPval)来评估。一种或多种靶标的表达水平的显著性可基于选自包含AUC、AUC P值(Auc.pvalue)、Wilcoxon检验P值、中位倍数差异(MFD)、Kaplan Meier(KM)曲线、生存AUC(survAUC)、单变量分析比值比P值(uvaORPval)、多变量分析比值比P值(mvaORPval)、Kaplan Meier P值(KMP值)、单变量分析风险比P值(uvaHRPval)或多变量分析风险比P值(mvaHRPval)的组的两个或更多个度量。

[0031] 在另一方面,本公开包括用于确定患有BC的受试者的预后以及是否用放射疗法治疗该受试者的探针组,该探针组包含用于检测多种靶核酸的多种探针,其中所述多种靶核酸包含选自AGR2、B4GALT1、CLDN7、EZR、GNG11、JUN、MMP11、PKIB、PRPS1、PSMD10、SH3BP5、SLC16A3、SLC7A11、SPP1、TNNT1和UBE2E1的一种或多种基因的一个或多个基因序列或它们的互补序列。为了便于检测,可对探针进行可检测地标记。在一些实施方案中,预后包括癌症复发预后。在另外的实施方案中,癌症复发是局部或局部区域复发或远处复发(转移)。

[0032] 在另一方面,本公开包括用于确定患有BC的受试者的预后以及是否用放射疗法治疗该受试者的系统,该系统包括:a)本文所述的探针组;和b)计算机模型或算法,其用于分析来自患有BC的受试者的生物样品中与所述多种探针杂交的多种靶核酸的表达水平或表达谱,并基于表达水平或表达谱确定受试者是否处于癌症复发低风险以及是否应当用放射疗法治疗。在一些实施方案中,癌症复发是局部或局部区域复发或远处复发(转移)。

[0033] 在一些实施方案中,本公开包括用于确定患有乳腺癌的受试者的预后以及是否用辅助放射疗法治疗该受试者的试剂盒,该试剂盒包含用于测量选自AGR2、B4GALT1、CLDN7、EZR、GNG11、JUN、MMP11、PKIB、PRPS1、PSMD10、SH3BP5、SLC16A3、SLC7A11、SPP1、TNNT1和UBE2E1的一种或多种基因的表达水平的剂。在其他实施方案中,试剂盒可包括用于测量多种基因的表达水平的一种或多种剂(例如,杂交探针、PCR引物或微阵列)、用于容纳包含从人受试者分离的用于测试的乳腺癌细胞的生物样品的容器,和/或印刷的说明书,其用于使剂与生物样品或生物样品的一部分反应以确定受试者是否处于乳腺癌的癌症复发低风险并且可能受益于使用辅助放射疗法的治疗。在一些实施方案中,癌症复发是局部或局部区域复发或远处复发(转移)。在其他实施方案中,剂被包装在单独的容器中。在其他实施方案中,试剂盒还包含一种或多种对照参考样品或用于测量基因表达的其他试剂(例如,用于进行PCR、RT-PCR、微阵列分析、RNA印迹、免疫测定或免疫组织化学的试剂)。在另一个实施方案中,试剂盒包含用于测量所有以下基因的表达水平的剂:AGR2、B4GALT1、CLDN7、EZR、GNG11、JUN、MMP11、PKIB、PRPS1、PSMD10、SH3BP5、SLC16A3、SLC7A11、SPP1、TNNT1和UBE2E1。在某些实施方案中,试剂盒包含如本文所述的用于检测多种靶核酸的探针组,其中所述多种靶核酸包含选自AGR2,B4GALT1,CLDN7,EZR,GNG11,JUN,MMP11,PKIB,PRPS1,PSMD10,SH3BP5,SLC16A3,SLC7A11,SPP1,TNNT1和UBE2E1或它们的任何组合的一种或多种基因的一

个或多个基因序列或它们的互补序列。

[0034] 在其他实施方案中,试剂盒还包含系统,其中该系统包含:a)包含用于检测多种靶核酸的多种探针的探针组,其中所述多种靶核酸包含选自AGR2、B4GALT1、CLDN7、EZR、GNG11、JUN、MMP11、PKIB、PRPS1、PSMD10、SH3BP5、SLC16A3、SLC7A11、SPP1、TNNT1和UBE2E1的一种或多种基因的一个或多个基因序列或它们的互补序列;和b)计算机模型或算法,其用于分析来自患有乳腺癌的受试者的生物样品中与所述多种探针杂交的多种靶核酸的表达水平或表达谱,并基于所述表达水平或表达谱确定受试者是否处于癌症复发低风险并且不可能受益于使用辅助放射疗法的治疗。在一些实施方案中,癌症复发是局部或局部区域复发或远处复发(转移)。

[0035] 鉴于本文的公开内容,本领域技术人员将很容易想到本公开内容的这些和其他实施方案。

附图说明

[0036] 图1:用于在SweBCG91-RT中选择训练队列和验证队列的图。

[0037] 图2:用于选择Princess Margaret验证队列的图。

[0038] 图3:用于选择要包括在模型中的基因的图。

[0039] 图4:对于按POLAR分类为低风险(A)或高风险(B)的患者,在SweBCG91-RT验证队列中使用或不使用辅助放射疗法(RT)时局部区域复发的累积发生率。使用原因特异性Cox比例风险回归模型计算风险比和p值。

[0040] 图5:对于按POLAR分类为低风险(A)或高风险(B)的患者,在Princess Margaret队列中使用或不使用辅助放射疗法(RT)时局部区域复发的累积发生率。使用原因特异性Cox比例风险回归模型计算风险比和p值。

具体实施方式

[0041] 多项III期随机临床试验一致证实了保乳手术(BCS)后全乳腺放射疗法(RT)在减少局部区域复发(LRR)方面的益处,并且RT被认为是早期浸润性乳腺癌女性的护理标准(参见,Malmström,P.等人,European journal of cancer(Oxford,England:1990)39,1690-1697(2003);Fyles,A.W.等人,The New England journal of medicine 351,963-970(2004);Holli,K.等人,Br J Cancer 84,164-169(2001);Forrest,A.P.等人,Lancet 348,708-713(1996);Fisher,B.等人,Journal of clinical oncology:official journal of the American Society of Clinical Oncology 20,4141-4149(2002);Liljegren,G.等人,J Clin Oncol 17,2326-2333(1999);Veronesi,U.等人,Annals of oncology:official journal of the European Society for Medical Oncology/ESMO 12,997-100(2001);Clark,R.M.等人,J Natl Cancer Inst 88,1659-1664(1996))。Oxford早期乳腺癌试验组(EBCTCG)荟萃分析证明,接受RT的患者在10年时局部复发的相对风险降低了三分之二,从30%降至10%,这转化为15年时约5%的生存获益(参见Clarke,M.等人,Lancet 366,2087-2106(2005);Early Breast Cancer Trialists'Collaborative,Lancet 378,1707-1716(2011))。除了局部区域疾病控制的改善之外,这些数据还证明了反应的异质性和受益于辅助放射的异质性。即使在以前全身性化学疗法效果较差和内分泌疗法使用有限的时

代,高达70%的女性在没有RT的情况下不会局部复发,然而仍缺乏鉴定哪些女性可安全放弃辅助放射疗法的临床工具。

[0042] 临床风险因素诸如年龄较大、肿瘤大小较小和雌激素受体阳性(ER+)乳腺癌与LRR的低风险相关(参见Fredriksson,I.等人,英国外科杂志90,1093-1102(2003))。尝试使用这些特征来识别BCS后可能省略放射疗法的女性,结果好坏参半。CALGB 9343试验招募了70岁或70岁以上的患有小ER+乳腺癌的女性,切除时切缘为阴性,然后随机接受全乳腺RT和他莫昔芬5年或仅接受他莫昔芬(参见,Hughes,K.S.等人,Journal of clinical oncology: official journal of the American Society of Clinical Oncology 31,2382-2387(2013))。在这项600多名女性的试验中,在不影响保乳率的情况下,RT将10年的LRR风险从10%降至2%。类似地,PRIME II试验包括65岁或65岁以上的小ER+乳腺癌患者,切除时切缘为阴性,接受内分泌疗法治疗,并随机分为RT或非RT。最近报道的10年结果表明,在添加RT的情况下,IBTR从9.8%降至0.9%(参见Kunkler,I.等人,The lancet oncology 16,266-273(2015))。由于在两项试验中,未接受RT治疗的患者在10年时的局部复发率均为10%,因此可以认为,在美国中位预期寿命超过80岁的女性中,10年时添加放射疗法使相对风险降低80%至90%可能过于显著,以至于不能省略RT。迄今为止,临床实践中还没有结合亚型或临床病理学变量来预测辅助乳腺放射缺乏益处,使得做出省略放射决定在临床上具有挑战性(参见,Killander,F.等人,European journal of cancer(Oxford,England:1990)67,57-65(2016);Sjöström,M.等人,Journal of Clinical Oncology 35,3222-3229(2017);Liu,F.F.等人,Journal of clinical oncology:official journal of the American Society of Clinical Oncology 33,2035-2040(2015))。

[0043] 鉴于此类方法在确定乳腺癌女性全身化疗获益可能性方面的成功,近年来对使用基因组知情的风险分层的兴趣有所增加。在以前的临床试验中,已显示此类分子知情测试为乳腺癌女性患者结果的预后和/或化学疗法获益的预测(参见,Albain,K.S等人,The Lancet Oncology 11,55-65,(2010);Cardoso,F等人,New England Journal of Medicine 375,717-729,(2016);Dowsett,M等人,Journal of Clinical Oncology 28,1829-1834(2010);Dowsett,M等人,Journal of Clinical Oncology 31,2783-2790(2013);Paik,S等人,The New England journal of medicine 351,2817-2826,(2004);Sparano,J.A等人,New England Journal of Medicine 373,2005-2014(2015))。虽然先前已经进行了建立局部复发或放射敏感性签名的尝试,但是还没有广泛使用的签名来选择可能省略放射的早期浸润性乳腺癌患者(参见Cui,Y.,Li,B.,Pollom,E.L.,Horst,K.C.&Li,R.Clinical cancer research:an official journal of the American Association for Cancer Research 24,4754-4762(2018);Eschrich,S.A.等人,International journal of radiation oncology,biology,physics 75,489-496(2009);Sjöström,M.等人,Journal of Clinical Oncology 37,3340-3349,(2019);Speers,C.等人,Clinical cancer research:an official journal of the American Association for Cancer Research 21,3667-3677(2015);Tramm,T.等人,Clinical cancer research:an official journal of the American Association for Cancer Research 20,5272-5280(2014))。ARTIC是一种27个基因的临床基因组签名,以前被开发和验证为LRR的预后和从RT获益的预测。该签名能够识别具有较高LRR风险的患者,其从放射疗法中获益减少,因此可用于识别需要强化治疗的患

者。通过ARTIC鉴定为低风险的患者从RT中获益较大,不适合作为省略RT的候选者。设计用于估计全身治疗后远处复发风险的签名已被建议用于避免RT,并且这些签名正在进行的临床试验中进行测试,但这些试验将在若干年内不会成熟(参见Speers,C.&Pierce,L.J.Current Breast Cancer Reports 12,255-265(2020))。

[0044] 在开发本发明实施方案的过程中进行的实验(参见实施例I)进行了SweBCG91-RT和Princess Margaret队列的分析,SweBCG91-RT是BCS后淋巴结阴性I-II期浸润性乳腺癌(BC)女性患者+/-RT的随机试验,Princess Margaret队列是年龄为50岁或50岁以上的BCS后T1-T2淋巴结阴性BC女性患者+/-RT的随机试验。SweBCG91-RT试验中的患者没有用辅助全身性疗法治疗,而Princess Margaret试验中的患者用他莫昔芬治疗。使用Affymetrix人外显子1.0ST微阵列进行肿瘤的全转录组谱分析。SweBCG91-RT队列分成243名患者的训练队列和354名患者的验证队列,并使用弹性网络回归训练16基因签名(AGR2、B4GALT1、CLDN7、EZR、GNG11、JUN、MMP11、PKIB、PRPS1、PSMD10、SH3BP5、SLC16A3、SLC7A11、SPP1、TNNT1和UBE2E1;参见表5和实施例I)以预测LRR,称为省略局部辅助放射的曲线(POLAR)。结果表明,分类为POLAR低风险并且未接受RT治疗的患者的10年LRR为6%,POLAR低风险患者没有显著的RT受益,分类为POLAR高风险的患者在接受RT治疗时LRR风险显著降低。这些结果表明,基于LRR生物学的新POLAR基因组签名能够鉴定尽管未接受RT但具有低LRR风险的患者,并因此是省略RT的主要候选者。

[0045] 因此,本公开涉及用于提供BC复发(例如,LRR)的个体化预后评估的系统和方法。这些系统和方法涉及测量来自患者样品的基因表达以产生基因表达签名,该基因表达签名鉴定不可能受益于RT的受试者。

[0046] 在某些实施方案中,本发明提供了方法,其包括:a)测量来自患有乳腺癌(BC)或处于患乳腺癌(BC)风险的人患者的生物样品中的一种或多种基因的表达水平,其中所述一种或多种基因选自AGR2、B4GALT1、CLDN7、EZR、GNG11、JUN、MMP11、PKIB、PRPS1、PSMD10、SH3BP5、SLC16A3、SLC7A11、SPP1、TNNT1和UBE2E1;和b)基于所述一种或多种基因的表达水平确定患者BC复发的可能性。

[0047] 在某些实施方案中,本公开提供了预测具有患BC风险的BC患者的BC复发可能性的方法,其包括:(a)测量从患者获得的样品中以下基因中的一种或多种的表达水平:AGR2、B4GALT1、CLDN7、EZR、GNG11、JUN、MMP11、PKIB、PRPS1、PSMD10、SH3BP5、SLC16A3、SLC7A11、SPP1、TNNT1和UBE2E1;和(b)基于所述一种或多种基因的表达水平预测患者的BC复发可能性,其中AGR2、CLDN7、EZR、MMP11、PKIB、PRPS1、PSMD10、SLC16A3、SLC7A11、SPP1、TNNT1和UBE2E1的表达增加与BC复发风险增加相关,并且其中B4GALT1、GNG11、JUN和SH3BP5的表达增加与乳腺癌复发风险降低相关。

[0048] 在某些实施方案中,本公开公开了使用一种或多种基因/基因靶标的基于表达的分析来诊断、预测和/或监测受试者BC的状态或结果的系统和方法。通常,该方法包括(a)获得来自受试者的样品中一种或多种基因的表达水平;和(b)基于所述一种或多种基因的表达水平确定受试者癌症复发的风险。在一些实施方案中,该方法还可以包括如果基于所述一种或多种基因的表达水平将受试者鉴定为具有癌症复发风险则实施RT,或者如果基于所述一种或多种基因的表达水平将受试者鉴定为处于癌症复发低风险则不施用RT。

[0049] 在某些实施方案中,提供了用于确定受试者是否处于乳腺癌复发的低或高风险的

方法。一般来讲,该方法包括:(a)提供包含来自受试者的乳腺癌细胞的样品;(b)测定样品中一种或多种基因的表达水平;和(c)基于所述多种靶标的表达水平确定受试者是否处于乳腺癌复发的低风险。在某些实施方案中,该方法可另外包括确定是否用辅助放射疗法、化学疗法或内分泌疗法治疗受试者。例如,根据本公开的方法被鉴定为处于乳腺癌复发低风险的受试者可能不太可能或不可能对辅助放射疗法有反应,而被鉴定为处于乳腺癌复发的较高风险的受试者可能更可能对辅助放射疗法有反应。

[0050] 在某些实施方案中,本公开涉及用于提供BC复发的个体化预后评估的系统和方法,以及不可能受益于RT的受试者的鉴定。本发明的此类系统和方法不限于评估BC复发风险的特定方式。在一些实施方案中,BC复发评估的风险以基因表达签名的形式提供,该基因表达签名给出关于患者肿瘤的重要且易于理解的信息。

[0051] 在某些实施方案中,本发明的系统和方法涉及测量和分析来自患者样品的基因表达以产生基因表达签名。在一些实施方案中,基因表达签名以易于理解的评分或风险类别提供。例如,基因表达签名可作为数字风险评分提供,或者风险评分可用于将患者分配至BC复发风险的类别。例如,患者可被分类为BC复发的高或低风险。在其他实施方案中,风险类别可被表征为BC复发的极低风险、低风险、中低风险、平均或中风险、中高风险、高风险或极高风险。风险评分或风险类别可使得容易且有效地与患者谈论他们的测试结果,并且还可以通过减少向患者解释结果所需的时间量而有助于医生更快地做出治疗决定。

[0052] 因此,在某些实施方案中,本发明的方法涉及产生对乳腺癌的临床治疗有用的风险评分。基因表达签名可预测例如疾病复发的风险,并且因此,基因表达签名可用于选择最佳疗程。例如,风险评分可用于识别处于复发高风险的患者,从而识别RT的良好候选者。替代性地,风险评分可用于识别处于复发低风险的患者,从而识别省略RT的良好候选者。因此,风险评分可用于对患者进行分类并选择适当的治疗。

[0053] 例如,在一些实施方案中,可将所选择的基因或其组合的表达测量值制成线性或非线性模型或算法,并转换成可能性或风险评分。

[0054] 在某些实施方案中,可使用线性算法来计算可能性或风险评分,其中基于基因与BC复发的个体相关性来为所测定的每种基因分配权重或系数(参见表8)。然后可将所测定的一种或多种基因的加权表达水平加在一起以产生可能性或风险评分。在一个实施方案中,如果测定了单个基因的表达水平,则该基因的加权表达水平可被认为是风险评分,并可用于确定患者的风险类别。

[0055] 在某些实施方案中,所得评分可包括数值,并且可落在-5.0与10.0之间、或-4.0与9.0之间、或-3.0与8.0之间、或-2.0与7.0之间、或-1.0与6.0之间、或-0.5与5.0之间、或-0.5与4.0之间、或-0.5与3.0之间、或0与2.5之间、或0.5与2.0之间、或1.0与1.5之间的范围内。在另外的实施方案中,评分可包括例如,诸如-5.0、-4.5、-4.0、-3.5、-3.0、-2.5、-2.0、-1.5、-1.0、-0.95、-0.90、-0.85、-0.80、-0.75、-0.70、-0.65、-0.60、-0.55、-0.50、-0.45、-0.40、-0.35、-0.30、-0.25、-0.20、-0.15、-0.10、-0.05、0、0.05、0.10、0.15、0.20、0.25、0.30、0.35、0.40、0.45、0.50、0.55、0.60、0.65、0.70、0.75、0.80、0.85、0.90、0.95、1.0、1.5、2.0、2.5、3.0、3.5、4.0、4.5、5.0、5.5、6.0、6.5、7.0、7.5、8.0、8.5、9.0、9.5或10.0的数值。

[0056] 在其他实施方案中,可将风险评分与预定阈值进行比较,以确定患者是处于BC复

发高风险(即,评分高于预定阈值)还是低风险(即,评分低于预定阈值),并且因此可能或不可能受益于RT。在某些实施方案中,预定阈值可以是0,使得具有正风险评分的患者被视为高风险,而具有负风险评分的患者被视为低风险。在另一个实施方案中,预定阈值可落在-1.0与5.0之间的范围内,例如-0.5与3.0之间、0与2.0之间、0与1.0之间、0.25与0.75之间、0.30与0.70之间、0.40与0.60之间、0.50与0.60之间、0.55与0.65之间、0.55与0.60之间以及0.60与0.65之间。在另一实施方案中,预定阈值可包括-1.0、-0.90、-0.80、-0.70、-0.60、-0.55、-0.50、-0.45、-0.40、-0.35、-0.30、-0.25、-0.20、-0.15、-0.10、-0.05、0、0.05、0.10、0.15、0.20、0.25、0.30、0.35、0.40、0.41、0.42、0.43、0.44、0.45、0.46、0.47、0.48、0.49、0.50、0.51、0.52、0.53、0.54、0.55、0.56、0.57、0.58、0.59、0.60、0.61、0.62、0.63、0.64、0.65、0.66、0.67、0.68、0.69、0.70、0.75、0.80、0.85、0.90、0.95、1.0、1.5、2.0、2.5或3.0。

[0057] 在另外的实施方案中,复发风险或可能性可被表征为极低风险、低风险、中低风险、平均或中风险、中高风险、高风险或极高风险。在此类实施方案中,可将风险评分与多于一个预定阈值进行比较,以确定患者是否处于BC复发的极低风险、低风险、中低风险、平均或中风险、中高风险、高风险或极高风险,或者这些风险水平的任何子集或组合。例如,可将风险评分与多于一个预定阈值进行比较,以确定患者是否处于BC复发的低风险、平均或中风险或高风险。在另一个实施方案中,可将风险评分与多于一个预定阈值进行比较,以确定患者是否处于BC复发的极低风险、低风险、平均或中风险、高风险或极高风险。在将风险评分与多于一个预定阈值进行比较的此类情况下,预定阈值可落在-1.0与5.0之间的范围内,例如-0.5与3.0之间、0与2.0之间、0与1.0之间、0.25与0.75之间、0.30与0.70之间、0.40与0.60之间、0.50与0.60之间、0.55与0.65之间、0.55与0.60之间以及0.60与0.65之间。在另一实施方案中,预定阈值可包括-1.0、-0.90、-0.80、-0.70、-0.60、-0.55、-0.50、-0.45、-0.40、-0.35、-0.30、-0.25、-0.20、-0.15、-0.10、-0.05、0、0.05、0.10、0.15、0.20、0.25、0.30、0.35、0.40、0.41、0.42、0.43、0.44、0.45、0.46、0.47、0.48、0.49、0.50、0.51、0.52、0.53、0.54、0.55、0.56、0.57、0.58、0.59、0.60、0.61、0.62、0.63、0.64、0.65、0.66、0.67、0.68、0.69、0.70、0.75、0.80、0.85、0.90、0.95、1.0、1.5、2.0、2.5或3.0或它们的任何组合。在某些实施方案中,多于一个预定阈值或阈值范围中的每一个可以是不同的并且可与不同的风险水平相关联。

[0058] 在一些实施方案中,风险或可能性评分可被缩放或调整为易于解释的标度。例如,风险评分可被缩放或调整到1与100之间的值。此类易于解释的标度可使治疗决策和与患者谈论他们的测试结果和治疗选择变得容易和有效。

[0059] 在某些实施方案中,可类似地缩放或调整用于确定风险水平的预定阈值,例如使得它们匹配或对应于为获得最终标度而进行的调整。例如,如果风险评分被缩放或调整为具有1与100之间的值的标度,则可以相应地缩放或调整预定阈值或多于一个预定阈值,使得这些阈值的值落入1-100标度内。

[0060] 本发明的方法还可以包括将表达签名或风险评分与其他临床因素组合以给出单一风险评价。

[0061] 如本文所用,“BC复发风险”或“BC复发可能性”是指在延长的时间段(例如,1个月、1年、5年、10年等)内BC复发的统计概率(例如,可能性)。在一些实施方案中,BC复发风险涉

及基线,其中患者已成功进行干预(例如,手术干预)并且特征在于不具有BC和/或积极进展的癌细胞。因此,复发风险或可能性涉及癌症以某种方式复发的可能性。

[0062] 在一些实施方案中,AGR2、CLDN7、EZR、MMP11、PKIB、PRPS1、PSMD10、SLC16A3、SLC7A11、SPP1、TNNT1和UBE2E1的表达水平增加各自与乳腺癌复发风险或可能性增加相关。

[0063] 在一些实施方案中,AGR2、CLDN7、EZR、MMP11、PKIB、PRPS1、PSMD10、SLC16A3、SLC7A11、SPP1、TNNT1和UBE2E1的表达水平降低各自与乳腺癌复发风险或可能性降低相关。

[0064] 在一些实施方案中,B4GALT1、GNG11、JUN和SH3BP5的表达水平增加各自与乳腺癌复发风险或可能性降低相关。

[0065] 在一些实施方案中,B4GALT1、GNG11、JUN和SH3BP5表达水平降低各自与乳腺癌复发风险或可能性增加相关。

[0066] 此类系统和方法不限于使用来自受试者的特定类型或种类的样品。例如,在一些实施方案中,样品或生物样品是组织样品、体液样品、血液样品、器官分泌物样品、CSF样品、唾液样品、血浆样品、血清样品或尿液样品。在一些实施方案中,样品可包括乳腺组织或周围组织、乳腺活检组织、肿瘤样品或含有乳腺细胞或乳腺癌细胞的组织。

[0067] 在一些实施方案中,受试者是人受试者。在一些实施方案中,受试者是处于患BC风险的人受试者。在一些实施方案中,受试者是处于患BC风险的女性人受试者。在一些实施方案中,受试者是被诊断为BC的人受试者。在一些实施方案中,受试者是被诊断为BC的女性人受试者。受试者可能由于各种症状(包括检测到肿物或肿块)而被怀疑患有癌症。在一些实施方案中,癌症是早期乳腺癌,即完全包含在乳腺内的癌症。

[0068] 在一些实施方案中,基因表达分析或比较可以用无监督的层次聚类算法进行,诸如K均值聚类算法。聚类算法是一种将一组对象进行聚类或分组的算法,使得同一组(称为聚类)中的对象彼此比其他组(聚类)中的对象更相似。聚类算法可将来自患者样品的RNA表达水平与癌症的一个或多个阶段中预期的RNA表达水平进行聚类。具有BC复发低风险或BC复发高风险的患者预期的RNA表达水平可来自与已知结果相关联的一个或多个肿瘤样品。RNA表达水平可基于它们的表达相似性进行聚类。在一些实施方案中,聚类算法将RNA表达水平聚类成与已知结果相关联的不同组。这些组可反映指示预后的连续结果。

[0069] 此类系统和方法不限于用于测量和/或确定基因表达的特定方式或技术。如本文所用,“基因表达”是指生物样品中基因的相对表达水平和/或表达模式。

[0070] 测量基因表达的技术是本领域已知的。实际上,用于测量基因表达的任何已知技术都被考虑并且并入本文。基因表达可通过任何合适的技术来确定,包括但不限于包括基于PCR的技术(例如实时PCR)、凝胶电泳技术、色谱技术、基于抗体的技术、离心技术或它们的组合的技术。用于测量基因表达的方法可包括测量由组织分离的RNA制备的cDNA的量。在一些实施方案中,基因表达测量技术涉及使用或不使用基因芯片的基因表达测定(参见,Onken等人,J Molec Diag 12(4):461-468(2010);和Kirby等人,Adv Clin Chem 44:247-292(2007)。在一些实施方案中,基因表达测量技术涉及affymetrix基因芯片和RNA芯片以及基因表达测定试剂盒(例如,Applied Biosystems™TaqMan®基因表达测定)。

[0071] 确定生物样品中基因表达水平的附加技术涉及核酸扩增过程,例如,通过RT-PCR(美国专利4,683,202)、连接酶链反应(Barany,Proc.Natl.Acad.Sci.USA 88:189-93,1991)、自主序列复制(Guatelli等人,Proc.Natl.Acad.Sci.USA 87:1874-78,1990)、转录

扩增系统(Kwoh等人,Proc.Natl.Acad.Sci.USA 86:1173-77,1989)、Q-β复制酶(Lizardi等人,Bio/Technology 6:1197,1988)、滚环复制(美国专利5,854,033),或任何其他核酸扩增方法进行核酸扩增,随后使用本领域技术人员熟知的技术检测扩增的分子。

[0072] 在一些实施方案中,用定量等位基因特异性实时靶标和信号扩增(QuARTS)测定法确定基因表达。在每个QuARTS测定中依次发生三个反应,包括初级反应中的扩增(反应1)和靶探针切割(反应2);以及次级反应中的FRET切割和荧光信号产生(反应3)。当用特异性引物扩增靶核酸时,具有瓣序列的特异性检测探针与扩增子松散地结合。在靶结合位点处存在特异性侵入性寡核苷酸导致5'核酸酶,例如FEN-1内切核酸酶通过在检测探针与瓣序列之间切割而释放瓣序列。瓣序列与相应的FRET盒的非发夹部分互补。因此,瓣序列在FRET盒上起着侵入性寡核苷酸的作用,并实现FRET盒荧光团与猝灭剂之间的切割,这产生荧光信号。切割反应可切割每种靶标上的多个探针,因此每个瓣释放多个荧光团,从而提供指数级信号放大。通过使用具有不同染料的FRET盒,QuARTS可检测单个反应孔中的多种靶标。参见例如Zou等人,(2010)"Sensitive quantification of methylated markers with a novel methylation specific technology"Clin Chem 56:A199)和美国专利8,361,720;8,715,937;8,916,344;和9,212,392,其各自出于所有目的以引用方式并入本文。

[0073] 在某些实施方案中,可将基因表达水平归一化以使样品之间的误差或变化最小化。因此,归一化可用于对样品之间的基因表达进行准确比较,因为基于每个样品的基因表达可能受到在样品收集和处理期间出现的非生物变量的影响,这可能会给真实信号添加噪声。测定可通过并入某些归一化基因的表达来提供归一化,这些归一化基因在相关条件下表达水平没有显著差异。本文所公开的示例性归一化基因包括管家基因。(参见例如E.Eisenberg等人,Trends in Genetics19(7):362-365(2003).)归一化可基于所有测定的基因或其大子集的平均或中值信号(Ct或Cp)(全局归一化方法)。通常,归一化基因(也称为参考基因)应该是已知与非癌性乳腺组织相比在BC中不表现出显著不同表达的基因,并且其不受各种样品和工艺条件的显著影响,从而提供排除外来影响的归一化条件。例如,可用于本文所公开的方法的参考基因可包括本领域常用于使基因表达模式归一化的基因,包括但不限于甘油醛-3-磷酸脱氢酶(GAPDH)和β-肌动蛋白,或本领域已知的任何其他参考基因。本领域技术人员将认识到,可以多种方式来实现归一化,并且上述技术仅旨在是示例性的,而非穷举性的。

[0074] 在一些实施方案中,测定样品中一种或多种基因靶标的表达水平可包括将样品应用于微阵列。

[0075] 在某些实施方案中,通过微阵列测定的基因表达可使用单通道阵列归一化(SCAN)方法进行归一化。SCAN方法利用单样品技术,而不是将微阵列样品以组的形式处理,从而避免了可能由组处理引入的偏差。该方法通过模拟探针-核苷酸组合物对荧光强度的影响并除去探针特异性和阵列特异性背景噪声来单独地归一化每个样品。(Piccolo SR,Sun Y,Campbell JD,Lenburg ME,Bild AH,Johnson WE(2012)."A single-sample microarray normalization method to facilitate personalized-medicine workflows."Genomics,100(6),337-344)。

[0076] 在一些情况下,测定一种或多种基因的表达水平包括使用算法。该算法可用于产生基因组分类器。替代性地,分类器可包括探针选择区域。在一些情况下,测定多种靶标的

表达水平包括检测和/或定量一种或多种基因。在一些实施方案中,测定一种或多种基因的表达水平包括对多种靶标进行测序。在一些实施方案中,测定一种或多种基因靶标的表达水平包括扩增所述多种靶标。在一些实施方案中,测定一种或多种基因靶标的表达水平包括定量靶标。在一些实施方案中,测定一种或多种基因靶标的表达水平包括对多种靶标进行多重反应。

[0077] 在一些情况下,测定多种基因的表达水平包括检测和/或定量多种靶分析物。在一些实施方案中,测定多种基因的表达水平包括对多种靶核酸进行测序。在一些实施方案中,测定多种生物标志物基因的表达水平包括扩增多种靶核酸。在一些实施方案中,测定多种生物标志物基因的表达水平包括对多种靶分析物进行多重反应。

[0078] 本文所公开的方法通常包括测定多种靶标的表达水平。所述多种靶标可包括蛋白编码基因或非蛋白编码基因的编码靶标和/或非编码靶标。蛋白编码基因结构可包含外显子和内含子。外显子还可以包含编码序列(CDS)和非翻译区(UTR)。蛋白编码基因可被转录以产生前体mRNA,并且前体mRNA可被加工以产生成熟mRNA。成熟mRNA可被翻译以产生蛋白质。

[0079] 非蛋白编码基因结构可包含外显子和内含子。通常,非蛋白编码基因的外显子区域主要含有UTR。非蛋白编码基因可被转录以产生前体mRNA,并且前体mRNA可被加工以产生非编码RNA(ncRNA)。

[0080] 编码靶标可包含外显子的编码序列。非编码靶标可包含外显子的UTR序列、内含子序列、基因间序列、启动子序列、非编码转录物、CDS反义、内含子反义、UTR反义或非编码转录物反义。非编码转录物可包含非编码RNA(ncRNA)。

[0081] 此类系统和方法不限于特定基因的分析。在一些实施方案中,所述一种或多种基因选自AGR2、B4GALT1、CLDN7、EZR、GNG11、JUN、MMP11、PKIB、PRPS1、PSMD10、SH3BP5、SLC16A3、SLC7A11、SPP1、TNNT1和UBE2E1。此类系统和实施方案不限于使用一种或多种基因的特定数量或组合(例如,仅1种基因、1-2种基因、1-3、1-4、1-5、1-6、1-7、1-8、1-9、1-10、1-11、1-12、1-13、1-14、1-15、1-16、2-3、2-4、2-5、2-6、2-7、2-8、2-9、2-10、2-11、2-12、2-13、2-14、2-15、2-16、3-4、3-5、3-6、3-7、3-8、3-9、3-10、3-11、3-12、3-13、3-14、3-15、3-16、4-5、4-6、4-7、4-8、4-9、4-10、4-11、4-12、4-13、4-14、4-15、4-16、5-6、5-7、5-8、5-9、5-10、5-11、5-12、5-13、5-14、5-15、5-16、6-7、6-8、6-9、6-10、6-11、6-12、6-13、6-14、6-15、6-16、7-8、7-9、7-10、7-11、7-12、7-13、7-14、7-15、7-16、8-9、8-10、8-11、8-12、8-13、8-14、8-15、8-16、9-10、9-11、9-12、9-13、9-14、9-15、9-16、10-11、10-12、10-13、10-14、10-15、10-16、11-12、11-13、11-14、11-15、11-16、12-13、12-14、12-15、12-16、13-14、13-15、13-16、14-15、14-16、15-16或16种基因)。在一些实施方案中,该方法包括16种或更少、15种或更少、14种或更少、13种或更少、12种或更少、11种或更少、10种或更少、9种或更少、8种或更少、7种或更少、6种或更少、5种或更少、4种或更少、3种或更少或2种或更少的基因的组合。在其他实施方案中,该方法可包含2种或更多种、3种或更多种、4种或更多种、5种或更多种、6种或更多种、7种或更多种、8种或更多种、9种或更多种、10种或更多种、11种或更多种、12种或更多种、13种或更多种、14种或更多种或15种或更多种基因。在另一个实施方案中,所公开的方法的一种或多种基因由16种基因组成。

[0082] 在一些情况下,一种或多种基因靶标包括编码靶标、非编码靶标或它们的任何组

合。在一些情况下,编码靶标包含外显子序列。在其他情况下,非编码靶标包含非外显子或外显子序列。替代性地,非编码靶标包含UTR序列、内含子序列、反义或非编码RNA转录物。在一些情况下,非编码靶标包含与UTR序列或内含子序列部分重叠的序列。非编码靶标还包括非外显子和/或外显子转录物。外显子序列可包含蛋白编码基因上的区,诸如外显子、UTR,或它们的一部分。非外显子序列可包含蛋白编码基因、非蛋白编码基因上的区,或它们的一部分。例如,非外显子序列可包含内含子区、启动子区、基因间区、非编码转录物、外显子反义区、内含子反义区、UTR反义区、非编码转录物反义区,或它们的一部分。在其他情况下,多种靶标包含非编码RNA转录物。

[0083] 基因靶标可包括选自本文所公开的分类器的一种或多种靶标。分类器可由一个或多个模型或算法生成。一个或多个模型或算法可以是朴素贝叶斯(NB)、递归分区(Rpart)、随机森林(RF)、支持向量机(SVM)、k最近邻(KNN)、高维判别分析(HDDA)、线性模型或它们的组合。分类器可具有等于或大于0.60的AUC。分类器可具有等于或大于0.61的AUC。分类器可具有等于或大于0.62的AUC。分类器可具有等于或大于0.63的AUC。分类器可具有等于或大于0.64的AUC。分类器可具有等于或大于0.65的AUC。分类器可具有等于或大于0.66的AUC。分类器可具有等于或大于0.67的AUC。分类器可具有等于或大于0.68的AUC。分类器可具有等于或大于0.69的AUC。分类器可具有等于或大于0.70的AUC。分类器可具有等于或大于0.75的AUC。分类器可具有等于或大于0.77的AUC。分类器可具有等于或大于0.78的AUC。分类器可具有等于或大于0.79的AUC。分类器可具有等于或大于0.80的AUC。基于其95%置信区间(CI),AUC可能具有临床意义。分类器的准确度可以是至少约70%。分类器的准确度可以是至少约73%。分类器的准确度可以是至少约75%。分类器的准确度可以是至少约77%。分类器的准确度可以是至少约80%。分类器的准确度可以是至少约83%。分类器的准确度可以是至少约84%。分类器的准确度可以是至少约86%。分类器的准确度可以是至少约88%。分类器的准确度可以是至少约90%。分类器的p值可小于或等于0.05。分类器的p值可小于或等于0.04。分类器的p值可小于或等于0.03。分类器的p值可小于或等于0.02。分类器的p值可小于或等于0.01。分类器的p值可小于或等于0.008。分类器的p值可小于或等于0.006。分类器的p值可小于或等于0.004。分类器的p值可小于或等于0.002。分类器的p值可小于或等于0.001。

[0084] 一种或多种基因靶标可包括选自线性模型分类器的一种或多种靶标。多种靶标可包括选自线性模型分类器的两种或更多种靶标。多种靶标可包括选自线性模型分类器的三种或更多种靶标。多种靶标可包括选自线性模型分类器的2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、27种或更多种靶标。线性模型分类器可以是LM2和LM3或LM4分类器。线性模型分类器可以是LM16分类器(例如,具有16种靶标的线性模型分类器)。例如,本公开的线性模型分类器可包括选自表5的两种或更多种靶标。

[0085] 一种或多种基因靶标可包括选自SVM分类器的一种或多种靶标。多种靶标可包括选自SVM分类器的2、3、4、5、6、7、8、9、10种或更多种靶标。多种靶标可包括选自SVM分类器的12、13、14、15、17、20、22、25种或更多种靶标。SVM分类器可以是SVM2分类器。本公开的SVM分类器可包括选自表5的两种或更多种靶标。

[0086] 一种或多种基因靶标可包括选自KNN分类器的一种或多种靶标。多种靶标可包括选自KNN分类器的2、3、4、5、6、7、8、9、10种或更多种靶标。多种靶标可包括选自KNN分类器的

12、13、14、15、17、20、22、25种或更多种靶标。例如,本公开的KNN分类器可包括选自表5的两种或更多种靶标。

[0087] 一种或多种基因靶标可包括选自朴素贝叶斯(NB)分类器的一种或多种靶标。多种靶标可包括选自NB分类器的2、3、4、5、6、7、8、9、10种或更多种靶标。多种靶标可包括选自NB分类器的12、13、14、15、17、20、22、25种或更多种靶标。例如,本公开的NB分类器可包括选自表5的两种或更多种靶标。

[0088] 一种或多种基因靶标可包括选自递归分区(Rpart)分类器的一种或多种靶标。多种靶标可包括选自Rpart分类器的2、3、4、5、6、7、8、9、10种或更多种靶标。多种靶标可包括选自Rpart分类器的12、13、14、15、17、20、22、25种或更多种靶标。例如,本公开的Rpart分类器可包括选自表5的两种或更多种靶标。

[0089] 一种或多种基因靶标可包括选自高维判别分析(HDDA)分类器的一种或多种靶标。多种靶标可包括选自高维判别分析(HDDA)分类器的两种或更多种靶标。多种靶标可包括选自高维判别分析(HDDA)分类器的三种或更多种靶标。多种靶标可包括选自高维判别分析(HDDA)分类器的5、6、7、8、9、10、11、12、13、14、15、17、20、22、25种或更多种靶标。例如,本公开的Rpart分类器可包括选自表5的两种或更多种靶标。

[0090] 本公开提供了用于诊断、监测和/或预测受试者乳腺癌的状态或结果的探针组,其包含多个探针,其中(i)所述组中的探针能够检测至少一种靶标的表达水平;并且(ii)该表达水平以至少约40%的特异性确定受试者的癌症状态(例如,复发风险)。

[0091] 探针组可包含一个或多个多核苷酸探针。单个多核苷酸探针包含衍生自靶序列的核苷酸序列或其互补序列的核苷酸序列。多核苷酸探针的核苷酸序列被设计成使得其与靶序列对应或互补。多核苷酸探针可在严格的或严格性降低的杂交条件下与靶序列的一个区、与其互补序列或与从其衍生的核酸序列(诸如cDNA)特异性杂交。

[0092] 多核苷酸探针序列的选择和它们的独特性的确定可使用本领域已知的技术在计算机上进行,例如,基于针对基因序列数据库(诸如人类基因组序列、UniGene、dbEST或NCBI的非冗余数据库)对所讨论的多核苷酸序列的BLASTN搜索。在本公开的一个实施方案中,多核苷酸探针与衍生自探针组中的靶序列的靶mRNA的一个区互补。计算机程序也可用于选择不能交叉杂交或不能非特异性杂交的探针序列。

[0093] 在一些情况下,从乳腺癌组织样品中提取并扩增的RNA的微阵列杂交可产生数据集,然后通过fRMA技术对其进行总结和归一化。在去除(或过滤)交叉杂交的PSRs和含有少于4个探针的PSR后,剩余的PSR可用于进一步分析。在fRMA和过滤之后,可将数据分解成其主成分,并使用方差模型的分析来确定批次效应在前10个主成分中保持存在的程度。

[0094] 然后可通过CR(临床复发)与非CR样品之间的T检验对这些剩余的PSR进行过滤。使用0.01的p值截止值,可进一步细化剩余特征(例如,PSR)。可通过使用弹性网络惩罚的正则化逻辑回归来执行特征选择。正则化回归可使用所有训练数据自举1000次以上;对于自举的每次迭代,可将在3重交叉验证之后具有非零系数的特征制成表。在一些情况下,在总运行的至少25%中选择的特征用于建模。

[0095] 本公开的多核苷酸探针的长度范围可以为约15个核苷酸至编码靶标或非编码靶标的全长。在本公开的一个实施方案中,多核苷酸探针的长度为至少约15个核苷酸。在另一个实施方案中,多核苷酸探针的长度为至少约20个核苷酸。在另一个实施方案中,多核苷酸

探针的长度为至少约25个核苷酸。在另一个实施方案中,多核苷酸探针的长度在约15个核苷酸与约500个核苷酸之间。在其他实施方案中,多核苷酸探针的长度为约15个核苷酸至约450个核苷酸,约15个核苷酸至约400个核苷酸,约15个核苷酸至约350个核苷酸,约15个核苷酸至约300个核苷酸,约15个核苷酸至约250个核苷酸,约15个核苷酸至约200个核苷酸。在一些实施方案中,探针的长度为至少15个核苷酸。在一些实施方案中,探针的长度为至少15个核苷酸。在一些实施方案中,探针的长度为至少20个核苷酸、至少25个核苷酸、至少50个核苷酸、至少75个核苷酸、至少100个核苷酸、至少125个核苷酸、至少150个核苷酸、至少200个核苷酸、至少225个核苷酸、至少250个核苷酸、至少275个核苷酸、至少300个核苷酸、至少325个核苷酸、至少350个核苷酸、至少375个核苷酸。

[0096] 探针组的多核苷酸探针可包含RNA、DNA、RNA或DNA模拟物,或它们的组合,并且可以是单链或双链的。因此,多核苷酸探针可由天然存在的核碱基、糖和共价核苷间(主链)键以及具有类似功能的非天然存在部分的多核苷酸探针组成。此类修饰或取代的多核苷酸探针可提供所需的特性,诸如,例如对靶基因的亲和力增强和稳定性增加。探针组可包含编码靶标和/或非编码靶标。

[0097] 在一些实施方案中,探针组包含与至少约5个编码靶标和/或非编码靶标杂交的多个靶序列。替代性地,探针组包含与至少约10个编码靶标和/或非编码靶标杂交的多个靶序列。在一些实施方案中,探针组包含与至少约15个编码靶标和/或非编码靶标杂交的多个靶序列。在一些实施方案中,探针组包含与至少约20个编码靶标和/或非编码靶标杂交的多个靶序列。在一些实施方案中,探针组包含与至少约30个编码靶标和/或非编码靶标杂交的多个靶序列。

[0098] 本公开的系统还提供了能够扩增由探针组定义的靶序列或其片段或子序列或互补序列的引物和引物对。探针组的核苷酸序列可在计算机可读介质中提供,用于计算机模拟应用,并作为设计用于扩增探针组的一种或多种靶序列的合适引物的基础。

[0099] 可以设计基于靶序列的核苷酸序列的引物用于靶序列的扩增。为了用于扩增反应诸如PCR,可使用一对引物。引物序列的确切组成对于本公开不是关键的,但是对于大多数应用,引物可在如本领域已知的严格条件下,特别是在高严格条件下与探针组的特定序列杂交。通常选择引物对,以便产生至少约50个核苷酸,更通常至少约100个核苷酸的扩增产物。用于选择引物序列的算法通常是已知的,并且可在商业软件包中获得。这些引物可用于标准的定量或定性的基于PCR的测定,以评估由探针组定义的RNA的转录物表达水平。替代性地,这些引物可与探针组合使用,诸如使用实时PCR的扩增中的分子信标。

[0100] 在一个实施方案中,当在扩增反应中使用时,引物或引物对特异性地扩增靶标(或如本文所述的其子组)、其RNA形式或其任一种的互补物的核酸序列的至少一部分。

[0101] 标记可任选地附接或掺入探针或引物多核苷酸,以允许检测和/或定量代表感兴趣的靶序列的靶多核苷酸。靶多核苷酸可以是表达的靶序列RNA本身、其cDNA拷贝或由其衍生的扩增产物,并且可以是正链或负链,只要它可以在所使用的测定中特异性地检测。类似地,可标记抗体。

[0102] 在某些多重格式中,用于检测不同靶标的标记可以是可区分的。标记可(例如通过共价键)直接附接或例如通过桥接分子或一系列分子(例如可结合到测定组分的分子或复合物,或通过可掺入测定组分中的结合对的成员,例如生物素-抗生物素蛋白或链霉抗生物

素蛋白)间接附接。许多标记可以活化形式商购获得,其可以容易地用于此类缀合(例如通过胺酰化),或者标记可通过已知的或可确定的缀合方案附接,其中许多方案是本领域已知的。

[0103] 可用于本文所述的公开内容中的标记包括当结合至或掺入感兴趣的生物分子时可被检测的任何物质。可使用任何有效的检测方法,包括光学、光谱、电学、压电、磁学、拉曼散射、表面等离子体共振、比色、量热等。标记通常选自发色团、发光团、荧光团、淬灭系统的一个成员、色原、半抗原、抗原、磁性颗粒、显示非线性光学的材料、半导体纳米晶体、金属纳米颗粒、酶、抗体或其结合部分或等同物、适体和结合对的一个成员,以及它们的组合。可使用淬灭方案,其中可在探针上使用作为淬灭对成员的淬灭剂和荧光团,使得在与靶标结合时发生光学参数的变化,引入或淬灭来自荧光团的信号。此类系统的一个实例是分子信标。合适的淬灭剂/荧光团是本领域已知的。标记可通过多种中间键结合。例如,多核苷酸可包含生物素结合物种,并且光学可检测标记可与生物素缀合,然后与标记的多核苷酸结合。类似地,多核苷酸传感器可包含免疫物种,诸如抗体或片段,并且可添加含有光学可检测标记的第二抗体。

[0104] 可用于本文所述方法的发色团包括可吸收能量并发光的任何物质。对于多重测定,可使用具有可检测的不同发射光谱的多个不同信号发色团。发色团可以是发光团或荧光团。典型的荧光团包括荧光染料、半导体纳米晶体、镧系元素螯合物、多核苷酸特异性染料和绿色荧光蛋白。

[0105] 在一些实施方案中,本公开的多核苷酸包含靶标或其互补物的核酸序列的至少20个连续碱基。多核苷酸可包含靶标的核酸序列的至少21、22、23、24、25、27、30、32、35、40、45、50个或更多个连续碱基。

[0106] 多核苷酸可以多种形式提供,包括作为固体、在溶液中或在阵列中。多核苷酸可任选地包含一种或多种标记,其可以化学地和/或酶促地掺入多核苷酸中。

[0107] 在一些实施方案中,本文提供的一种或多种多核苷酸可提供在基底上。基底可包含宽范围的材料,可以是生物材料、非生物材料、有机材料、无机材料或这些材料的任意组合。例如,基底可以是聚合的Langmuir Blodgett膜、官能化玻璃、Si、Ge、GaAs、GaP、SiO_2、SiN_4、改性硅,或多种凝胶或聚合物中的任一种,诸如(聚)四氟乙烯、(聚)偏二氟乙烯、聚苯乙烯、交联聚苯乙烯、聚丙烯酸、聚乳酸、聚乙醇酸、聚(丙交酯-乙交酯)、聚酐、聚(甲基丙烯酸甲酯)、聚(乙烯-共-乙酸乙烯酯)、聚硅氧烷、聚合二氧化硅、胶乳、葡聚糖聚合物、环氧树脂、聚碳酸酯,或它们的组合。可使用导电聚合物和光电导材料。

[0108] 基底可采取阵列、光电二极管、光电传感器诸如光电半导体芯片或光电薄膜半导体,或生物芯片的形式。探针在基底上的位置可以是可寻址的;这可以高密度格式实现,并且位置可以是可微寻址的或可纳米寻址的。

[0109] 在某些实施方案中,从需要癌症治疗的受试者收集含有乳腺癌细胞的样品(例如,生物样品),以基于表达水平或表达谱和受益于辅助放射疗法的可能性来评价受试者是否处于癌症复发的低风险或高风险。本公开的系统和方法中使用的诊断样品包含适于提供RNA表达信息的核酸。原则上,从中获得表达的RNA并分析靶基因表达的生物样品可以是怀疑包含癌性乳腺组织或细胞的任何材料。诊断样品可以是在本公开的方法中直接使用的生物样品。替代性地,诊断样品可以是由生物样品制备的样品。

[0110] 在一个实施方案中,包含或疑似包含癌性组织或细胞的样品或样品部分可以是任何来源的生物材料,包括细胞、组织或流体,包括体液。样品来源的非限制性实例包括抽吸物、针吸活检、细胞学沉淀、例如通过手术或尸检获得的大块组织制备物或其切片、淋巴液、血液、血浆、血清、肿瘤和器官。在一些实施方案中,样品来自乳腺肿瘤活检。

[0111] 样品可以是具有已知的和记录的医疗结果的存档样品,或者可以是来自最终医疗结果还未知的当前受试者的样品。

[0112] 在一些实施方案中,可在分子分析之前解剖样品。样品可通过大块肿瘤样本或其部分的大体解剖来制备,或者可通过显微解剖来处理,例如通过激光捕获显微解剖(LCM)。

[0113] 样品最初可以各种状态提供,如新鲜组织、新鲜冷冻组织、细针抽吸物,并且可以是固定的或未固定的。通常,医学实验室常规地制备固定状态的医学样品,这有利于组织存储。多种固定剂可用于固定组织以稳定细胞的形态,并且可单独使用或与其他剂组合使用。示例性的固定剂包括交联剂、醇、丙酮、Bouin′s溶液、Zenker溶液、Hely溶液、锇酸溶液和Carnoy溶液。

[0114] 交联固定剂可包括适于形成两个或更多个共价键的任何剂,例如醛。通常用于固定的醛的来源包括甲醛、多聚甲醛、戊二醛或福尔马林。在一些实施方案中,交联剂包含甲醛,其可以其天然形式或以多聚甲醛或福尔马林的形式包含。本领域技术人员将理解,对于其中已使用交联固定剂的样品,可能需要特殊的制备步骤,包括例如加热步骤和蛋白酶-k消化;参见方法。

[0115] 一种或多种醇可单独或与其他固定剂组合用于固定组织。用于固定的示例性醇包括甲醇、乙醇和异丙醇。

[0116] 福尔马林固定常用于医学实验室。福尔马林包含醇(通常为甲醇)和甲醛,两者均可用于固定生物样品。

[0117] 无论固定的还是未固定的,生物样品可任选地包埋在包埋介质中。用于组织学的示例性包埋介质包括石蜡、Tissue-Tek®V.I.P、Paramat、Paramat Extra、Paraplast、Paraplast X-tra、Paraplast Plus、Peel Away Paraffin Embedding Wax、聚酯蜡、Carbowax聚乙二醇、Polyfin、组织冷冻介质TFMFM、Cryo-Gef以及OCT化合物(Electron Microscopy Sciences,Hatfield,PA)。在分子分析之前,可通过本领域已知的任何合适的技术除去包埋材料。例如,当样品包埋在蜡中时,可通过用有机溶剂例如二甲苯萃取来除去包埋材料。用于从组织中去除包埋介质的试剂盒可商购获得。样品或其切片可根据需要进行进一步的处理步骤,例如连续的水合或脱水步骤。

[0118] 在一些实施方案中,样品是固定的蜡包埋的生物样品。通常,来自医学实验室的样品作为固定的、蜡包埋的样品提供,最通常作为福尔马林固定的、石蜡包埋的(FFPE)组织提供。

[0119] 无论生物样品的来源如何,最终测定的靶多核苷酸都可以合成制备(在对照序列的情况下),但通常从生物来源中纯化并进行一个或多个制备步骤。可纯化RNA以从生物样品中除去或减少一种或多种不需要的组分或将其浓缩。相反,当RNA对于特定测定而言过于浓缩时,可将其稀释。

[0120] 可使用任何合适的技术从生物样品中提取和纯化RNA。许多技术是本领域已知的,并且有几种是可商购获得的(例如,FormaPure核酸提取试剂盒,Agencourt Biosciences,

Beverly MA;高纯FFPE RNA微量试剂盒,Roche Applied Science,Indianapolis,IN)。可使用TRIzol(Invitrogen,Carlsbad,CA)从冷冻组织切片中提取RNA,并使用RNeasy Protect试剂盒(Qiagen,Valencia,CA)纯化。可使用DNAse I处理(Ambion,Austin,TX)进一步纯化RNA,以消除任何污染DNA。可使用Nanodrop ND-1000分光光度计(Nanodrop Technologies,Rockland,DE)进行RNA浓缩。可通过冷乙酸钠沉淀进一步纯化RNA,以消除干扰cDNA合成的污染物。可通过运行电泳图谱来评估RNA完整性,并且可使用Bioanalyzer 2100(Agilent Technologies,Santa Clara,CA)的RNA 6000PicoAssay确定RNA完整性数值(RIN,指示mRNA完整性的相关量度)。

[0121] 还提供了用于进行所需方法的试剂盒,其包括用于容纳试剂盒组分的容器或外壳、包含一种或多种核酸的一个或多个容器以及任选的包含一种或多种试剂的一个或多个容器。试剂包括以上物质组成部分中所述的那些,和可用于进行所述方法的那些试剂,包括扩增试剂,并且可包括一种或多种探针、引物或引物对、酶(包括聚合酶和连接酶)、嵌入染料、标记探针以及可掺入扩增产物中的标记。

[0122] 在一些实施方案中,试剂盒包含专门针对本文所述的靶序列的那些子集和组合的引物或引物对。引物或引物对适于选择性扩增靶序列。试剂盒可包含适于选择性扩增一种或多种靶标的至少两个、三个、四个或五个引物或引物对。试剂盒可包含适于选择性扩增一种或多种靶标的至少5、10、15、20、30、40、50、60、70、80、90、100、110个或更多个引物或引物对。

[0123] 在一些实施方案中,当用于扩增反应时,试剂盒的引物或引物对特异性地扩增本文所述的非编码靶标、编码靶标、外显子或非外显子靶标、对应于选自表5的靶标的核酸序列、其RNA形式,或其任一者的互补物。试剂盒可包括多个此类引物或引物对,其可特异性地扩增相应的本文所述的多种不同扩增非编码靶标、编码靶标、外显子或非外显子转录物、对应于选自表5的靶标的核酸序列、其RNA形式或其互补物。适于选择性扩增一种或多种靶标的至少两个、三个、四个或五个引物或引物对可以试剂盒形式提供。在一些实施方案中,试剂盒包含适于扩增一种或多种靶标的五至五十个引物或引物对。

[0124] 试剂可独立地为液体或固体形式。试剂可以混合物形式提供。试剂盒中可任选地提供对照样品和/或核酸。对照样品可包括得自或代表未显示疾病证据的受试者的肿瘤样品的组织和/或核酸,以及得自或代表发生全身性癌症的受试者的肿瘤样品的组织和/或核酸。

[0125] 核酸可以阵列形式提供,因此试剂盒中可包括阵列或微阵列。用于预测癌症受试者的疾病结果和/或用于指定治疗方式的试剂盒任选地可由政府机构认证。

[0126] 使用试剂盒执行本公开的一种或多种方法的说明书可随容器一起提供,并且可在任何固定介质中提供。说明书可位于容器或外壳的内部或外部,和/或可印刷在其任何表面的内部或外部。试剂盒可以是多重形式,用于同时检测和/或定量代表所表达的靶基因的一种或多种不同的靶多核苷酸。

[0127] 在样品收集和核酸提取之后,可使得用于或可用于制备感兴趣的靶多核苷酸的包含RNA的样品的核酸部分经受一个或多个制备性反应。这些制备性反应可包括体外转录(IVT)、标记、片段化、扩增和其他反应。可在检测、定量和/或扩增之前首先用逆转录酶和引物处理mRNA以产生cDNA;这可以用纯化的mRNA在体外或原位进行,例如在附着于载玻片的

细胞或组织中。

[0128] “扩增”是指产生核酸诸如表达的RNA的至少一个拷贝的任何过程,并且在许多情况下产生多个拷贝。扩增产物可以是RNA或DNA,并且可包括与表达的靶序列互补的链。DNA扩增产物可以最初通过反向翻译产生,然后任选地从进一步的扩增反应产生。扩增产物可包括靶序列的全部或一部分,并且可任选地被标记。各种扩增方法都适用,包括基于聚合酶的方法和基于连接的方法。示例性的扩增技术包括聚合酶链式反应方法(PCR)、脂肪酶链式反应(LCR)、基于核酶的方法、自主序列复制(3SR)、基于核酸序列的扩增(NASBA)、使用Qβ复制酶、逆转录、切口平移等。

[0129] 不对称扩增反应可用于优先扩增代表作为靶多核苷酸用于检测的靶序列的一条链。在一些情况下,扩增产物本身的存在和/或量可用于确定给定靶序列的表达水平。在其他情况下,扩增产物可用于与包含用于检测和/或定量靶序列表达的传感器多核苷酸的阵列或其他基底杂交。

[0130] 基于聚合酶的方法中的第一扩增循环通常形成与模板链互补的引物延伸产物。如果模板是单链RNA,则在第一次扩增中使用具有逆转录酶活性的聚合酶以将RNA逆转录成DNA,并且可进行附加的扩增循环以复制引物延伸产物。当然,用于PCR的引物必须被设计成与它们相应的模板中能够产生可扩增片段的区杂交;因此,每个引物必须杂交,使得其3′核苷酸与其互补模板链中的核苷酸配对,该互补模板链位于PCR中用于复制该互补模板链的引物的3’核苷酸的3’上。

[0131] 可通过使靶多核苷酸的一条或多条链与引物和具有合适活性的聚合酶接触来扩增靶多核苷酸,以延伸引物并复制靶多核苷酸,产生全长互补多核苷酸或其较小部分。可使用具有能够复制靶多核苷酸的聚合酶活性的任何酶,包括DNA聚合酶、RNA聚合酶、逆转录酶、具有多于一种类型的聚合酶或酶活性的酶。酶可以是热不稳定的或热稳定的。也可以使用酶的混合物。示例性酶包括:DNA聚合酶,诸如DNA聚合酶I(“Pol I”)、Pol I的Klenow片段、T4、T7、Sequenase®T7、Sequenase®Version 2.0T7、Tub、Taq、Tth、Pfic、Pfu、Tsp、Tfl、Tli和火球菌属物种(Pyrococcus sp)GB-D DNA聚合酶;RNA聚合酶,诸如大肠杆菌(E.coil)、SP6、T3和T7 RNA聚合酶;以及逆转录酶,诸如AMV、M-MuLV、MMLV、RNAse H MMLV(SuperScript®)、SuperScript®II、ThermoScript®、HIV-1和RAV2逆转录酶。所有这些酶都是可商购获得的。具有多种特异性的示例性聚合酶包括RAV2和Tli(外切)聚合酶。示例性的热稳定聚合酶包括Tub、Taq、Tth、Pfic、Pfu、Tsp、Tfl、Tli和火球菌属物种GB-D DNA聚合酶。

[0132] 选择合适的反应条件以允许靶多核苷酸的扩增,包括用于包含聚合酶链式反应的扩增方案的pH、缓冲液、离子强度、一种或多种盐的存在和浓度、反应物和辅因子诸如核苷酸和镁和/或其他金属离子(例如锰)的存在和浓度、任选的共溶剂、温度、热循环曲线,并且可部分取决于所使用的聚合酶以及样品的性质。助溶剂包括甲酰胺(通常为约2%至约10%)、甘油(通常为约5%至约10%)和DMSO(通常为约0.9%至约10%)。可在扩增方案中使用一些技术,以便最小化在扩增期间产生的假阳性或伪像的产生。这些包括“降落”PCR、热启动技术、使用巢式引物,或设计PCR引物,使得它们在引物-二聚体形成的情况下形成茎-环结构,因此不被扩增。可使用加速PCR的技术,例如离心PCR,其允许样品内更大的对流,并且包括用于快速加热和冷却样品的红外加热步骤。可进行一个或多个扩增循环。过量的一

种引物可用于在PCR期间产生过量的一种引物延伸产物。在一些实施方案中,过量产生的引物延伸产物是待检测的扩增产物。多种不同的引物可用于扩增样品内不同的靶多核苷酸或特定靶多核苷酸的不同区。

[0133] 扩增反应可在允许任选标记的传感器多核苷酸在扩增循环的至少一部分期间与扩增产物杂交的条件下进行。当以这种方式进行测定时,如本领域已知的,可通过监测扩增期间的光发射或荧光来实时检测这种杂交事件。

[0134] 当扩增产物用于与阵列或微阵列杂交时,可获得许多合适的可商购获得的扩增产物。这些包括可从NuGEN,Inc.(San Carlos,CA)获得的扩增试剂盒,包括WT-OvationTm系统、WT-OvationTm系统v2、WT-OvationTm Pico系统、WT-OvationTm FFPE外显子模块、WT-OvationTm FFPE外显子模块RiboAmp和RiboAmp ^{Plus} RNA扩增试剂盒(MDS Analytical Technologies(以前为Arcturus)(Mountain View,CA),Genisphere,Inc.(Hatfield,PA),包括单独或组合的RampUp Plus和SenseAmp RNA扩增试剂盒。扩增的核酸可在扩增和标记后进行一个或多个纯化反应,例如使用磁珠(例如,RNAClean磁珠,Agencourt Biosciences)。

[0135] 可使用实时定量多重RT-PCR平台和其他多重技术分析多种RNA生物标志物,诸如GenomeLab GeXP遗传分析系统(Beckman Coulter,Foster City,CA)、SmartCycler®9600或GeneXpert®系统(Cepheid,Sunnyvale,CA)、ABI 7900HT快速实时PCR系统(Applied Biosystems,Foster City,CA)、LightCycler®480系统(Roche Molecular Systems,Pleasanton,CA)、xMAP 100系统(Luminex,Austin,TX)Solexa基因组分析系统(Illumina,Hayward,CA)、OpenArray实时qPCR(BioTrove,Woburn,MA)和BeadXpress系统(Illumina,Hayward,CA)。

[0136] 检测和/或定量所编码的靶基因的表达的任何方法原则上可用于本公开。可直接检测和/或定量表达的靶基因,或者可复制和/或扩增以允许检测表达的靶基因的扩增拷贝。

[0137] 用于检测和/或定量靶基因的方法可包括Northern印迹、测序、阵列或微阵列杂交、通过酶切特定结构(例如Clariom S测定,ThermoFisher Scientific,Invader®测定,Third Wave Technologies,例如,如美国专利5,846,717、6,090,543、6,001,567、5,985,557和5,994,069中所述)和扩增方法,例如RT-PCR,包括在TaqMan®测定(PE Biosystems,Foster City,Calif.,例如,如美国专利5,962,233和5,538,848中所述)中,并且可以是定量的或半定量的,并且可根据可用的生物样品的来源、量和条件而变化。也可以使用这些方法的组合。例如,可以扩增、标记核酸并进行微阵列分析。用于检测和/或定量靶基因的方法可包括使用微阵列杂交的注释基因的基因水平表达分析(例如,GeneChip人外显子1.0ST测定或Clariom S测定,ThermoFisher Scientific)。

[0138] 在一些情况下,可通过测序检测靶基因。测序方法可包括全基因组测序或外显子组测序。也可以使用诸如Maxim-Gilbert、链终止或高通量系统的测序方法。另外,合适的测序技术包括使用标记的终止子或引物并在平板或毛细管中进行凝胶分离的经典双脱氧测序反应(Sanger方法)、通过使用可逆终止的标记的核苷酸的合成进行测序、焦磷酸测序、454测序、与标记的寡核苷酸探针文库的等位基因特异性杂交、通过使用与标记的克隆文库

的等位基因特异性杂交进行合成,随后连接,实时监测聚合步骤期间标记的核苷酸的掺入来测序,以及SOLiD测序。

[0139] 用于检测和/或定量靶基因的附加方法包括单分子测序(例如Helicos、PacBio)、合成测序(例如Illumina、Ion Torrent)、连接测序(例如ABI SOLID)、杂交测序(例如Complete Genomics)、原位杂交、珠阵列技术(例如Luminex xMAP、Illumina BeadChips)、分支DNA技术(例如Panomics、Genisphere)。测序方法可使用检测核苷酸的荧光(例如,Illumina)或电子(例如,Ion Torrent、Oxford Nanopore)方法。

[0140] 逆转录可通过本领域已知的任何方法进行。例如,RT-PCR可使用Omniscript试剂盒(Qiagen,Valencia,CA)、Superscript III试剂盒(Invitrogen,Carlsbad,CA)进行逆转录。可进行靶特异性引发,以增加靶基因的检测灵敏度并生成靶特异性cDNA。

[0141] TaqMan®RT-PCR可使用Applied Biosystems Prism(ABI)7900HT仪器以5 1.11体积进行,其中靶基因特异性cDNA相当于1ng总RNA。

[0142] 添加用于TaqMan分析的引物和探针浓度,以使用PCR循环条件扩增荧光扩增子,所述PCR循环条件为诸如95℃持续10分钟进行一个循环,95℃持续20秒,和60℃持续45秒进行40个循环。可测定参考样品,以确保试剂和过程稳定性。应测定阴性对照(例如,无模板)以监测任何外源核酸污染。

[0143] 本公开设想可以阵列形式提供探针组或由其衍生的探针。在本公开的上下文中,“阵列”是空间上或逻辑上组织的多核苷酸探针的集合。可使用包含专用于编码靶标、非编码靶标或它们的组合的探针的阵列。替代性地,可使用包含专用于靶标的两种或多种转录物或其衍生产物的探针的阵列。理想地,阵列可专用于靶基因的5、10、15、20、25、30种或更多种转录物。这些基因的表达可单独检测或与其他转录物组合检测。在一些实施方案中,使用的阵列包含用于乳腺特异性表达产物的宽范围的传感器探针以及合适的对照序列。在一些情况下,阵列可包含人外显子1.0ST阵列(HuEx 1.0ST,Thermo Fisher Scientific,Santa Clara,CA.)。

[0144] 通常,将多核苷酸探针附接到固体基底,并对其进行排序,使得各自(在基底上的)的位置和身份是已知的。可将多核苷酸探针附接到能够耐受使用阵列所需的试剂和条件的多种固体基底中的一种。实例包括但不限于聚合物,诸如(聚)四氟乙烯、(聚)偏二氟乙烯、聚苯乙烯、聚碳酸酯、聚丙烯和聚苯乙烯;陶瓷;硅;二氧化硅;改性硅;(熔融的)二氧化硅、石英或玻璃;功能化玻璃;纸,诸如滤纸;重氮化的纤维素;硝化纤维滤器;尼龙膜;以及聚丙烯酰胺凝胶垫。对光透明的基底可用于可在涉及光学检测的测定中使用的阵列。

[0145] 阵列形式的实例包括膜或滤器阵列(例如,硝化纤维、尼龙阵列)、板阵列(例如,多孔,诸如24、96、256、384、864或1536孔,微量滴定板阵列)、针阵列和珠阵列(例如,在液体“浆液”中)。基底诸如玻璃或陶瓷载玻片上的阵列通常称为芯片阵列或“芯片”。此类阵列是本领域公知的。在本公开的一个实施方案中,癌症预后阵列是芯片。

[0146] 在一些实施方案中,一种或多种模式识别方法可用于分析靶基因的表达水平。模式识别方法可包括表达水平的线性组合,或表达水平的非线性组合。在一些实施方案中,将RNA转录物或RNA转录物水平的组合的表达测量值构建成线性或非线性模型或算法(例如,“表达签名”)并转换成似然评分。该可能性评分指示生物样品来自可能未表现出疾病证据、可能表现出全身性癌症或可能表现出生物化学复发或局部区域复发的受试者的可能性。似

然评分可用于区分这些疾病状态。模型和/或算法可以机器可读格式提供,并且可用于将表达水平或表达谱与疾病状态相关联,和/或指定用于受试者或受试者类别的治疗方式。

[0147] 测定多种靶基因的表达水平可包括使用算法或分类器。可使用本领域已知的技术来管理、分类及分析阵列数据。测定多个基因靶标的表达水平可包括探针组建模和数据预处理。探针组建模和数据预处理可使用Robust Multi-Array(RMA)算法或变体GC-RMA、fRMA、Probe Logarithmic Intensity Error(PLIER)算法或变体iterPLIER来导出。方差或强度过滤器可应用于使用RMA算法预处理数据,例如通过分别去除归一化数据范围的标准偏差<10或平均强度<100强度单位的目标基因。

[0148] 替代性地,测定多个基因靶标的表达水平可包括使用机器学习算法。机器学习算法可包括监督学习算法。监督学习算法的实例可包括平均单依赖估计(AODE)、人工神经网络(例如,反向传播)、贝叶斯统计(例如,朴素贝叶斯分类器、贝叶斯网络、贝叶斯知识库)、基于病例的推理、决策树、归纳逻辑编程、高斯过程回归、数据处理的分组方法(GMDH)、学习自动机、学习向量量化、最小消息长度(决策树、决策图等)、懒惰学习、基于病例的学习最近邻算法、模拟建模、概率近似正确学习(PAC)学习、链波下降规则、知识获取方法、符号机器学习算法、子符号机器学习算法、支持向量机、随机森林、分类器集合、自举汇聚(装袋)和Boosting。监督学习可包括有序分类,诸如回归分析和信息模糊网络(IFN)。替代性地,监督学习方法可包括统计分类,诸如AODE、线性分类器(例如,Fisher线性判别、逻辑回归、朴素贝叶斯分类器、感知器,以及支持向量机、Lasso和弹性网络正则化一般线性模型)、二次分类器、k最近邻、Boosting、决策树(例如,C4.5,随机森林)、贝叶斯网络,以及隐马尔可夫模型。

[0149] 机器学习算法还可以包括无监督学习算法。无监督学习算法的实例可包括人工神经网络、数据聚类、期望最大化算法、自组织映射、径向基函数网络、向量量化、生成地形图、信息瓶颈方法和IBSEAD。无监督学习还可以包括关联规则学习算法,诸如Apriori算法、Eclat算法和FP-growth算法。也可以使用层次聚类,诸如单链聚类和概念聚类。替代性地,无监督学习可包括分区聚类,诸如K-means算法和模糊聚类。

[0150] 在一些情况下,机器学习算法包括强化学习算法。强化学习算法的实例包括但不限于时间差分学习、Q学习和学习自动机。替代性地,机器学习算法可包括数据预处理。

[0151] 在一些实施方案中,机器学习算法可包括但不限于平均单依赖估计(AODE)、Fisher线性判别、逻辑回归、感知器、多层感知器、人工神经网络、支持向量机、二次分类器、Boosting、决策树、C4.5、贝叶斯网络、隐马尔可夫模型、高维判别分析和高斯混合模型。机器学习算法可包括支持向量机、朴素贝叶斯分类器、k最近邻、高维判别分析或高斯混合模型。在一些情况下,机器学习算法包括随机森林。

[0152] 分子分型是将乳腺癌分类为多种遗传上不同的类别或亚型之一的方法。每种亚型对不同类型的治疗有不同的反应,并且特定亚型的存在预示着例如个体的放疗抵抗或化疗抵抗、更高或更低的复发风险或者个体的良好或不良预后。表5中列出的多种基因靶标的差异表达分析允许鉴定具有癌症复发低风险的受试者,其例如可能从辅助放射疗法中获益最少。在一些情况下,本公开的分子分型方法与其他生物标志物(如肿瘤等级和激素水平)组合用于分析乳腺癌。例如,患有雌激素受体阳性(ER+)、人表皮生长因子受体2阴性(HER2-)乳腺癌、淋巴结阴性乳腺癌的绝经后受试者可能更可能具有较低的复发风险(例如局部区

域复发)。

[0153] 癌症复发是指在体内没有检测到癌细胞一段时间后癌症的复发。在可手术乳腺癌的手术之后,疾病可局部地、区域地和/或在远处转移部位复发。局部复发是癌症在同侧胸壁上重新出现。相比之下,区域复发表示肿瘤累及区域淋巴结,通常是同侧腋窝或锁骨上淋巴结,较不常见的是锁骨下和/或内乳淋巴结。一些乳腺癌患者在首次诊断为乳腺癌十年内,在保乳手术和放射疗法后将具有局部或局部区域复发。如果在初次治疗过程中切除乳房,这些女性将在十年内在腋窝或胸壁中具有局部复发。在一些实施方案中,用本公开的方法治疗的受试者患有淋巴结阴性乳腺癌。

[0154] 诊断、预测或监测癌症的状态或结果可包括治疗癌症或预防癌症进展。此外,诊断、预测或监测癌症的状态或结果可包括鉴定或预测受试者处于复发(例如局部区域复发)的低或高风险。在一些情况下,诊断、预测或监测可包括确定治疗方案。确定治疗方案可包括施用抗癌疗法。替代性地,确定治疗方案可包括修改、推荐、继续或中止抗癌方案。例如,如本文所述,基于表达谱确定为处于乳腺癌复发低风险的受试者可以不用辅助放射疗法。在一些情况下,如本文所述,基于表达谱确定为处于乳腺癌复发高风险的受试者可以用乳房切除术、放射增强或辅助全身性疗法治疗。在一些情况下,如果样品表达模式与已知疾病或疾病结果的表达模式一致,则表达模式可用于指定一种或多种治疗方式(例如治疗方案、抗癌方案)。抗癌方案可包括一种或多种抗癌疗法。抗癌疗法的实例包括激素/内分泌疗法、手术、化学疗法、放射疗法、免疫疗法/生物疗法和光动力疗法。

[0155] 激素疗法或内分泌疗法可涉及施用激素,诸如类固醇激素或激素拮抗剂以调节某些激素的水平,从而阻止激素响应性癌细胞的生长或诱导其凋亡。例如,乳腺癌可用选择性雌激素受体调节剂(SERM)治疗,诸如但不限于他莫昔芬、雷洛昔芬和托瑞米芬。或者或另外,激素合成抑制剂诸如芳香酶抑制剂可用于治疗乳腺癌,包括但不限于来曲唑、阿那曲唑、依西美坦和氨鲁米特。在一些情况下,用孕激素(诸如但不限于醋酸甲地孕酮和醋酸甲羟孕酮)补充激素可用于治疗激素响应性晚期乳腺癌。特别地,ER+乳腺癌可用雌激素受体拮抗剂(例如他莫昔芬)或阻断雌激素产生的药物诸如芳香酶抑制剂(例如阿那曲唑或来曲唑)来治疗。激素疗法还可包括手术切除内分泌器官(例如睾丸切除术或卵巢切除术)。

[0156] 外科肿瘤学使用外科方法来诊断、分期和治疗癌症,并缓解某些癌症相关症状。手术可用于去除肿瘤(例如,根治术、切除术、减瘤手术)、重建身体的一部分(例如,恢复性手术),和/或缓解症状诸如疼痛(例如,姑息性手术)。手术也可包括冷冻手术。冷冻手术(也称为冷冻疗法)可使用液氮(或氩气)产生的极冷来破坏异常组织。冷冻手术可用于治疗外部肿瘤,诸如皮肤上的肿瘤。对于外部肿瘤,可以用棉签或喷雾装置将液氮直接施用于癌细胞。冷冻手术也可用于治疗体内肿瘤(内部肿瘤和骨内肿瘤)。对于内部肿瘤,液氮或氩气可通过称为冷冻探针的中空仪器循环,该仪器被放置成与肿瘤接触。可使用超声或MRI引导冷冻探针并监测细胞的冷冻,从而限制对附近健康组织的损伤。在探针周围会形成冰晶球,冻结附近的细胞。有时使用多于一个探针将液氮输送到肿瘤的各个部分。探针可在手术期间或通过皮肤(经皮)放入肿瘤中。在冷冻手术之后,冷冻的组织解冻,并且可被身体自然吸收(对于内部肿瘤),或者可溶解并形成结痂(对于外部肿瘤)。

[0157] 化学治疗剂也可用于治疗癌症。化学治疗剂的实例包括烷化剂、抗代谢物、植物生物碱和类萜、长春花生物碱、鬼臼毒素、紫杉烷、拓扑异构酶抑制剂和细胞毒性抗生素。顺

铂、卡铂和奥沙利铂是烷化剂的实例。其他烷化剂包括氮芥、环磷酰胺、苯丁酸氮芥、异环磷酰胺。烷化剂可通过与生物学上重要的分子中的氨基、羧基、巯基和磷酸根基团形成共价键而损害细胞功能。替代性地,烷化剂可以化学修饰细胞的DNA。

[0158] 抗代谢物是化学治疗剂的另一个实例。抗代谢物可伪装成嘌呤或嘧啶,并且可防止嘌呤和嘧啶在(细胞周期的)“S”期期间并入DNA中,从而停止正常发育和分裂。抗代谢物也可能影响RNA合成。代谢物的实例包括硫唑嘌呤和巯嘌呤。

[0159] 生物碱可来源于植物,并且阻断细胞分裂也可用于治疗癌症。烷酰基可阻止微管功能。生物碱的实例是长春花生物碱和紫杉烷。长春花生物碱可与微管蛋白上的特定位点结合并抑制微管蛋白组装成微管(细胞周期的M期)。长春花生物碱可来源于马达加斯加长春花(Madagascar periwinkle)、长春花(Catharanthus roseus)(以前称为玫瑰红长春花(Vinca rosea))。长春花生物碱的实例包括但不限于长春新碱、长春碱、长春瑞滨或长春地辛。紫杉烷是由紫杉属植物(红豆杉)产生的二萜类。紫杉烷可来源于天然来源或人工合成。紫杉烷包括紫杉醇(Taxol)和多西他赛(Taxotere)。紫杉烷可能破坏微管功能。微管是细胞分裂所必需的,并且紫杉烷可稳定微管中GDP结合的微管蛋白,从而抑制细胞分裂过程。因此,本质上,紫杉烷可以是有丝分裂抑制剂。紫杉烷也可以是放射增敏的并且通常含有许多手性中心。

[0160] 可供选择的化学治疗剂包括鬼臼毒素。鬼臼毒素是植物来源的化合物,其可以帮助消化并且可用于产生细胞抑制药物,诸如依托泊苷和替尼泊苷。它们可阻止细胞进入G1期(DNA复制的开始)和DNA的复制(S期)。

[0161] 拓扑异构酶是维持DNA拓扑的必需酶。I型或II型拓扑异构酶的抑制可通过扰乱适当的DNA超螺旋而干扰DNA的转录和复制。一些化学治疗剂可抑制拓扑异构酶。例如,一些I型拓扑异构酶抑制剂包括喜树碱:伊立替康和拓扑替康。II型抑制剂的实例包括安吖啶、依托泊苷、磷酸依托泊苷和替尼泊苷。激酶抑制剂可用于治疗乳腺癌。

[0162] 化学治疗剂的另一个实例是细胞毒性抗生素。细胞毒性抗生素是用于治疗癌症的一组抗生素,因为它们会干扰DNA复制和/或蛋白质合成。细胞毒性抗生素包括但不限于放线菌素、蒽环类药物、多柔比星、柔红霉素、戊柔比星、伊达比星、表柔比星、博来霉素、普卡霉素和丝裂霉素。

[0163] 在一些情况下,抗癌治疗可包括放射疗法。放射可来自身体外部的机器(外部束放射疗法)或来自置于体内癌细胞附近的放射性材料(内放射疗法,更通常称为近距离放射疗法)。全身性放射疗法使用放射性物质,通过口腔或静脉给药,其在血液中行进至全身组织。

[0164] 外部束放射疗法可以光子束(x射线或γ射线)的形式递送。光子是光和其他形式电磁辐射的基本单位。外部束放射疗法的实例称为三维适形放射疗法(3D-CRT)。3D-CRT可使用计算机软件和先进的治疗机将辐射递送到非常精确成形的目标区域。目前正在测试许多其他的外部束放射治疗方法并将其用于癌症治疗。这些方法包括但不限于调强放射疗法(IMRT)、图像引导放射疗法(IGRT)、立体定向放射外科(SRS)、立体定向体部放射疗法(SBRT)和质子疗法。

[0165] 调强放射疗法(IMRT)是外部束放射的实例,并且可使用数百个微小的放射束成形装置(称为准直器)来递送单剂量的放射。准直器可以是固定的或者可在治疗期间移动,从而允许放射束的强度在疗程改变。这种剂量调制允许肿瘤或附近组织的不同区域接收不同

剂量的放射。IMRT以逆向进行计划(称为逆向治疗计划)。在逆向治疗计划中,预先计划到肿瘤和周围组织的不同区域的放射剂量,然后高功率计算机程序计算放射治疗所需的射束数量和角度。相比之下,在传统的(前向)治疗计划期间,预先选择放射束的数量和角度,并且计算机计算可从每个计划的射束递送多少剂量。IMRT的目的是增加对需要放射的区域的放射剂量并减少对周围正常组织的特定敏感区域的放射暴露。

[0166] 外部束放射的另一个实例是图像引导放射疗法(IGRT)。在IGRT中,可在治疗期间进行重复成像扫描(CT、MRI或PET)。这些成像扫描可由计算机处理,以识别由于治疗而引起的肿瘤大小和位置的变化,并允许在治疗期间根据需要调节受试者的位置或计划的放射剂量。重复成像可提高放射治疗的准确性,并且可允许减少计划的待治疗组织体积,从而减少对正常组织的总放射剂量。

[0167] 断层治疗是一种图像引导的IMRT。断层治疗机是CT成像扫描仪与外部束放射治疗机之间的混合体。递送用于成像和治疗的放射的断层治疗机的部分可以与普通CT扫描仪相同的方式完全围绕受试者旋转。断层治疗机可在疗程之前立即捕获受试者肿瘤的CT图像,以允许非常精确的肿瘤靶向和正常组织的保留。

[0168] 立体定向放射外科(SRS)可向小肿瘤递送一个或多个高剂量的放射。SRS使用非常准确的图像引导肿瘤靶向和受试者定位。因此,可给予高剂量的放射而不会过度损伤正常组织。SRS可用于治疗边缘清晰的小肿瘤。它最常用于治疗脑或脊柱肿瘤和其他癌症类型的脑转移。对于一些脑转移的治疗,除了SRS之外,受试者还可以接受对整个脑的放射疗法(称为全脑放射疗法)。SRS需要在治疗期间使用头部框架或其他装置来固定受试者,以确保准确递送高剂量放射。

[0169] 在大多数情况下,立体定向体部放射疗法(SBRT)使用比3D-CRT更小的辐射场和更高的剂量在更少的疗程中进行放射治疗。SBRT可治疗位于脑和脊髓之外的肿瘤。因为这些肿瘤更可能随着身体的正常运动而移动,并且因此不能与脑或脊柱内的肿瘤一样精确地靶向,所以SBRT通常以多于一个剂量给予。SBRT可用于治疗小的、分离的肿瘤,包括肺癌和肝癌。SBRT系统可通过其商标名称而为人所知,诸如CyberKnife®。

[0170] 在质子疗法中,外部束放射疗法可通过质子递送。质子是一种带电粒子。质子束与光子束的区别主要在于它们在活组织中沉积能量的方式。光子沿着其通过组织的路径以小包的形式沉积能量,而质子在其路径的末端沉积其大部分能量(称为布拉格峰)并且在沿途沉积较少的能量。质子的使用可减少正常组织对放射的暴露,从而可能允许向肿瘤递送更高剂量的放射。

[0171] 其他带电粒子束诸如电子束可用于照射表面肿瘤,诸如皮肤癌或身体表面附近的肿瘤,但它们不能穿过组织进行非常远的距离。

[0172] 内放射疗法(近距离放射疗法)是从置于体内或身体上的放射源(放射性材料)递送的放射。在癌症治疗中使用若干种近距离放射治疗技术。间质近距离放射疗法可使用放置在肿瘤组织内(诸如乳房肿瘤内)的放射源。腔内近距离放射疗法可使用放置在靠近肿瘤的手术腔或体腔诸如胸腔内的源。可用于治疗眼睛内的黑素瘤的巩膜外近距离放射疗法可使用附接到眼睛的源。在近距离放射疗法中,放射性同位素可被密封在微小的小球或“种子”中。可使用递送装置(诸如针、导管或一些其他类型的载体)将这些种子置于受试者体内。当同位素自然衰变时,它们发出可能损害附近癌细胞的放射。近距离放射疗法可能能够

向一些癌症递送比外部束放射疗法更高剂量的放射,同时对正常组织造成更少的损伤。

[0173] 近距离放射疗法可作为低剂量率或高剂量率治疗给予。在低剂量率治疗中,癌细胞在若干天的时间内接收来自放射源的连续低剂量放射。在高剂量率治疗中,附接到置于体内的递送管的机器人机器可将一个或多个放射源引导到肿瘤内或肿瘤附近,然后在每个疗程结束时移除放射源。可在一个或多个疗程中给予高剂量率治疗。高剂量率治疗的实例是MammoSite®系统。近距离放射治疗可用于治疗已进行保乳手术的乳腺癌受试者。

[0174] 近距离放射治疗源的放置可以是临时性的或永久性的。对于永久性近距离放射疗法,可通过外科手术将源密封在体内并留在那里,即使在所有的放射都已发出之后。在一些情况下,剩余材料(其中放射性同位素被密封)不会对受试者造成任何不适或伤害。永久性近距离放射疗法是一种低剂量率近距离放射疗法。对于临时性近距离放射疗法,使用管(导管)或其他载体来递送放射源,并且在治疗之后移除载体和放射源。临时性近距离放射疗法可以是低剂量率或高剂量率治疗。近距离放射疗法可单独使用或与外部束放射治疗一起使用,以提供对肿瘤的“增强”放射,同时不影响周围的正常组织。

[0175] 在全身性放射疗法中,受试者可吞咽放射性物质或接受放射性物质的注射,诸如放射性碘或结合到单克隆抗体的放射性物质。放射性碘(131I)是一种全身性放射疗法,通常用于帮助治疗癌症,诸如甲状腺癌。甲状腺细胞自然吸收放射性碘。对于一些其他类型的癌症的全身性放射疗法,单克隆抗体可帮助将放射性物质靶向到正确的位置。与放射性物质结合的抗体穿过血液行进,定位并杀死肿瘤细胞。例如,药物替伊莫单抗(ibritumomab tiuxetan)(Zevalin®)可用于治疗某些类型的B细胞非霍奇金淋巴瘤(NHL)。该药物的抗体部分可识别并结合B淋巴细胞表面上发现的蛋白质。托西莫单抗(tositumomab)和碘I 131托西莫单抗的联合用药方案(Bexxar®)可用于治疗特定类型的癌症,诸如NHL。在该方案中,可以首先给予受试者非放射性托西莫单抗抗体,然后用附接有131I的托西莫单抗抗体治疗。与替伊莫单抗相同,托西莫单抗可识别并结合B淋巴细胞上的蛋白质。抗体的非放射性形式可帮助保护正常B淋巴细胞免受来自131I放射的损害。

[0176] 一些全身性放射疗法药物可缓解已经扩散到骨的癌症(骨转移)的疼痛。这是一种姑息性放射疗法。放射性药物钐-153-来昔决南(Quadramet®)和氯化锶-89(Metastron®)是可用于治疗骨转移疼痛的放射性药物的实例。

[0177] 生物疗法(有时称为免疫疗法、生物疗法(biotherapy)、生物疗法(biologic therapy)或生物反应调节剂(BRM)疗法)直接或间接地使用身体的免疫系统来对抗癌症或减轻可能由一些癌症治疗引起的副作用。生物疗法包括干扰素、白细胞介素、集落刺激因子、单克隆抗体、疫苗、基因治疗和非特异性免疫调节剂。

[0178] 干扰素(IFN)是体内天然存在的细胞因子类型。干扰素α、干扰素β和干扰素γ是可用于癌症治疗的干扰素的实例。

[0179] 像干扰素一样,白细胞介素(IL)是体内天然存在的细胞因子,并且可在实验室中制备。许多白细胞介素已被鉴定用于治疗癌症。例如,白细胞介素-2(IL-2或阿地白细胞介素)、白细胞介素7和白细胞介素12可用于抗癌治疗。IL-2可刺激许多免疫细胞诸如淋巴细胞的生长和活性,这可摧毁癌细胞。白细胞介素可用于治疗多种癌症,包括白血病、淋巴瘤以及脑癌、结肠直肠癌、卵巢癌、乳腺癌、肾癌和前列腺癌。

[0180] 集落刺激因子(CSF)(有时称为造血生长因子)也可用于治疗癌症。CSF的一些实例包括但不限于G-CSF(非格司亭)和GM-CSF(沙格司亭)。CSF可促进骨髓干细胞的分裂并促使它们发育成白血细胞、血小板和红血细胞。骨髓是机体免疫系统的关键,因为它是所有血细胞的来源。因为抗癌药物可损害身体产生白血细胞、红血细胞和血小板的能力,CSF对免疫系统的刺激可有益于经历其他抗癌治疗的受试者,因此CSF可与其他抗癌疗法诸如化学疗法组合。CSF可用于治疗多种癌症,包括淋巴瘤、白血病、多发性骨髓瘤、黑素瘤以及脑癌、肺癌、食道癌、乳腺癌、子宫癌、卵巢癌、前列腺癌、肾癌、结肠癌和直肠癌。

[0181] 另一种类型的生物疗法包括单克隆抗体(MOAB或MoAB)。这些抗体可由单一类型的细胞产生,并且可对特定抗原具有特异性。为了产生MOAB,可将人癌细胞注射到小鼠体内。作为响应,小鼠免疫系统可产生针对这些癌细胞的抗体。可分离产生抗体的小鼠浆细胞并与实验室生长的细胞融合以产生称为杂交瘤的“杂交”细胞。杂交瘤可以无限地产生大量的这些纯抗体或MOAB。MOAB可以多种方式用于癌症治疗。例如,与特定类型的癌症反应的MOAB可增强受试者对癌症的免疫应答。MOAB可被编程以抵抗细胞生长因子,从而干扰癌细胞的生长。

[0182] MOAB可与其他抗癌疗法诸如化学疗法、放射性同位素(放射性物质)、其他生物疗法或其他毒素相关联。当抗体附着在癌细胞上时,它们将这些抗癌疗法直接递送至肿瘤,帮助摧毁肿瘤。携带放射性同位素的MOAB也可证明可用于诊断某些癌症,诸如结肠直肠癌、卵巢癌、前列腺癌和乳腺癌。

[0183] Rituxan®(利妥昔单抗)和Herceptin®(曲妥单抗)是可用作生物疗法的MOAB的实例。Rituxan可用于治疗非霍奇金淋巴瘤。Herceptin可用于治疗患有产生过量的称为HER2的蛋白质的肿瘤的受试者的转移性乳腺癌。替代性地,MOAB可用于治疗淋巴瘤、白血病、黑素瘤以及脑、乳腺、肺、肾、结肠、直肠、卵巢、前列腺和其他区域的癌症。

[0184] 癌症疫苗是另一种形式的生物疗法。癌症疫苗可被设计成促进受试者的免疫系统识别癌细胞。癌症疫苗可被设计成治疗现有的癌症(治疗性疫苗)或预防癌症的发展(预防性疫苗)。治疗性疫苗可在诊断出癌症后注射到人体内。这些疫苗可阻止现有肿瘤的生长,防止癌症复发,或消除未被先前治疗杀死的癌细胞。在肿瘤较小时给予癌症疫苗可能能够根除癌症。另一方面,在癌症发展之前给予健康个体预防性疫苗。这些疫苗被设计成刺激免疫系统攻击可引起癌症的病毒。通过靶向这些致癌病毒,可预防某些癌症的发展。例如,cervarix和gardasil是治疗人乳头瘤病毒并可预防宫颈癌的疫苗。治疗性疫苗可用于治疗黑素瘤、淋巴瘤、白血病以及脑、乳腺、肺、肾、卵巢、前列腺、胰腺、结肠和直肠的癌症。癌症疫苗可与其他抗癌疗法组合使用。

[0185] 基因疗法是生物疗法的另一个实例。基因疗法可涉及将遗传物质引入人体细胞以对抗疾病。基因治疗方法可改善受试者对癌症的免疫应答。例如,可将基因插入免疫细胞以增强其识别和攻击癌细胞的能力。在另一种方法中,可以给癌细胞注射使癌细胞产生细胞因子并刺激免疫系统的基因。

[0186] 在一些情况下,生物疗法包括非特异性免疫调节剂。非特异性免疫调节剂是刺激或间接增强免疫系统的物质。通常,这些剂靶向关键的免疫系统细胞,并可引起二次反应,诸如细胞因子和免疫球蛋白的产生增加。用于癌症治疗的两种非特异性免疫调节剂是卡介苗(Calmette-Guerin)(BCG)和左旋咪唑。BCG可用于手术后浅表性膀胱癌的治疗。BCG可通

过刺激炎症反应和可能的免疫反应而起作用。可将BCG溶液滴注到膀胱中。左旋咪唑有时与氟尿嘧啶(5-FU)化学疗法一起用于治疗手术后的III期(Dukes'C)结肠癌。左旋咪唑可恢复受抑制的免疫功能。

[0187] 光动力疗法(PDT)是可使用称为光敏剂或光增敏剂的药物和特定类型的光的抗癌治疗。当光敏剂暴露于特定波长的光时,它们可产生杀死附近细胞的氧形式。光敏剂可被特定波长的光激活。该波长确定光可行进到体内的距离。因此,光敏剂和光波长可用于用PDT治疗身体的不同区域。

[0188] 在用于癌症治疗的PDT的第一步中,可将光增敏剂注射到血流中。该剂可被全身的细胞吸收,但在癌细胞中的停留时间可比在正常细胞中更长。在注射后约24至72小时,当大部分剂离开正常细胞但仍留在癌细胞中时,可将肿瘤暴露于光。肿瘤中的光敏剂可吸收光并产生活性形式的氧,从而摧毁附近的癌细胞。除了直接杀死癌细胞之外,PDT还可以两种其他方式缩小或破坏肿瘤。光敏剂可损伤肿瘤中的血管,从而防止癌症接受必要的营养。PDT也可以激活免疫系统以攻击肿瘤细胞。

[0189] 用于PDT的光可来自激光器或其他源。可通过光纤电缆(传输光的细纤维)引导激光,以将光递送到体内的区域。例如,光纤电缆可通过内窥镜(用于观察体内组织的发光管)插入肺或食道中,以治疗这些器官中的癌症。其他光源包括发光二极管(LED),其可用于表面肿瘤,诸如皮肤癌。PDT通常作为门诊手术进行。PDT也可以重复进行,并且可与其他疗法一起使用,诸如手术、放射或化学疗法。

[0190] 体外光分离置换法(ECP)是一种PDT,其中可使用机器收集受试者的血细胞。受试者的血细胞可在体外用光增敏剂处理,暴露于光,然后返回受试者。ECP可用于帮助减轻对其他疗法没有反应的皮肤T细胞淋巴瘤的皮肤症状的严重程度。ECP可用于治疗其他血癌,也可帮助减少移植后的排斥反应。

[0191] 此外,光增敏剂诸如卟吩姆钠或Photofrin®可用于PDT,以治疗或缓解食道癌和非小细胞肺癌的症状。当食道癌阻塞食管或当单独用激光疗法治疗食道癌不能令人满意时,卟吩姆钠可缓解食道癌的症状。卟吩姆钠可用于治疗常规治疗不适合的受试者的非小细胞肺癌,并减轻患有阻塞气道的非小细胞肺癌的受试者的症状。卟吩姆钠也可用于治疗患有巴雷特食道的受试者的癌前病变,巴雷特食道是可导致食道癌的病症。

[0192] 激光疗法可使用高强度光来治疗癌症和其他疾病。激光可用于缩小或摧毁肿瘤或癌前生长。激光最常用于治疗浅表性癌症(身体表面或内部器官内层的癌症),诸如基底细胞皮肤癌和一些癌症的极早期阶段,诸如子宫颈癌、阴茎癌、阴道癌、外阴癌和非小细胞肺癌。

[0193] 激光也可用于缓解癌症的某些症状,诸如出血或阻塞。例如,激光可用于缩小或摧毁阻塞受试者气管(气管)或食道的肿瘤。激光也可用于去除阻塞结肠或胃的结肠息肉或肿瘤。

[0194] 激光疗法通常通过柔性内窥镜(用于观察体内组织的薄的发光管)进行。内窥镜装配有光纤(传输光的细纤维)。它通过身体的开口插入,诸如口、鼻、肛门或阴道。然后激光被精确地瞄准以切割或摧毁肿瘤。

[0195] 激光诱导间质热疗法(LITT)或间质激光光凝术也使用激光来治疗一些癌症。LITT类似于称为热疗的癌症治疗,其使用热通过破坏或杀死癌细胞来缩小肿瘤。在LITT期间,将

光纤插入肿瘤中。光纤顶端的激光升高了肿瘤细胞的温度并损伤或摧毁它们。LITT有时用于缩小肝脏中的肿瘤。

[0196] 激光疗法可单独使用,但最常见的是与其他治疗诸如手术、化学疗法或放射疗法组合使用。此外,激光可封闭神经末梢以减轻术后疼痛并且封闭淋巴管以减轻肿胀并限制肿瘤细胞的扩散。

[0197] 用于治疗癌症的激光可包括二氧化碳(CO_2)激光、氩激光和钕:钇-铝-石榴石(Nd:YAG)激光。这些中的每一个都可以缩小或摧毁肿瘤并且可与内窥镜一起使用。CO_2和氩激光可切割皮肤表面而不会进入更深层。因此,它们可用于除去浅表癌症,诸如皮肤癌。相比之下,Nd:YAG激光更常用于通过内窥镜治疗内部器官,诸如子宫、食道和结肠。Nd:YAG激光也可在LITT期间穿过光纤行进到身体的特定区域。氩激光通常用于激活PDT中使用的药物。

[0198] 对于患有全身性疾病的受试者,可以指定附加的治疗形式,诸如辅助化学疗法(例如多西他赛、米托蒽醌和泼尼松)和全身性放射疗法(例如钐或锶)。此类受试者可能立即用放射疗法治疗,以消除推测的微转移性疾病,该推测的微转移性疾病临床上无法检测到但可通过靶基因表达签名揭示。还可以更密切地监测此类受试者的疾病进展迹象。

[0199] 对于无全身性疾病的受试者,可施用局部辅助放射疗法(例如,定位于乳腺组织)或内分泌疗法或化学疗法。对于没有疾病证据(NED)的受试者,如本文所述的表达谱和/或风险评分的计算可用于确定乳腺癌复发的风险。对于被鉴定为具有乳腺癌复发低风险并被鉴定为具有使用本文所述方法的放射疗法的高益处的受试者,将推荐辅助放射疗法。

[0200] 可将靶基因分组,使得获得的关于该组中靶基因集合的信息可用于做出或帮助做出临床相关判断,诸如诊断、预后或治疗选择。

[0201] 还提供了受试者报告,其包含来自受试者的生物样品中多种靶基因的所测量的表达水平的表示,其中所述表示包含对应于任一种、两种、三种、四种、五种、六种、八种、十种、二十种、三十种或更多种靶基因、本文所述的子集或它们的组合的表达水平。在一些实施方案中,所测量的表达水平的表示可采取感兴趣的靶基因的表达水平的线性或非线性组合的形式。受试者报告可以机器(例如,计算机)可读格式和/或以硬(纸)拷贝提供。该报告还可包括来自具有已知疾病状态和/或结果的一组或多组受试者的所述多种靶基因的表达水平的标准测量结果。该报告可用于告知受试者和/或治疗医师所表达的靶基因的表达水平、可能的医学诊断和/或含义,并且任选地可为受试者推荐治疗方式。

[0202] 还提供了可用于治疗、诊断、预断和以其他方式评估疾病的基因表达谱的表示。在一些实施方案中,这些谱的表示被简化为可由机器自动读取的介质,诸如计算机可读介质(磁性、光学等)。制品还可包括用于评估此类介质中的基因表达谱的说明书。例如,制品可包括具有用于比较上述基因组合的基因表达谱的计算机指令的可读存储形式。制品还可以具有数字化记录在其中的基因表达谱,使得可将它们与来自受试者样品的基因表达数据进行比较。替代性地,可以不同的表示格式来记录谱。图形记录是一种此类格式。聚类算法可有助于此类数据的可视化。

[0203] 实验

[0204] 提供以下实施例以展现并且进一步阐明本发明的某些优选实施方案和方面,并且不应理解为限制其范围。代词诸如“我们”、“我们的”和“我”的使用是指本发明实体。

[0205] 实施例1:用于乳腺癌局部复发的预后和对放射疗法反应的预测的基因组分类器

的开发和验证。

[0206] 方法

[0207] SweBCG91-RT队列

[0208] SweBCG91-RT是在瑞典进行的随机试验,并且以前已有描述(参见,Malmström,P等人,European journal of cancer(Oxford,England:1990)39,1690-1697,doi:10.1016/s0959-8049(03)00324-1(2003);Killander,F.等人,European journal of cancer(Oxford,England:1990)67,57-65,(2016);Sjöström,M.等人,Journal of Clinical Oncology 37,3340-3349(2019))。简言之,该试验将1,178例淋巴结阴性的I-IIA期乳腺癌患者随机分配至BCS后的辅助全乳腺RT或无RT。所有患者的手术切缘均为阴性。内分泌疗法和化学疗法是根据当时的当地指南提供的,在原始试验中仅对8%的患者施用了治疗。如前所述通过ER、PgR、HER2和KI67的免疫组织化学染色进行了亚型分型(参见,Sjöström,M.等人,Journal of Clinical Oncology 35,3222-3229(2017))。来自患有ER+、HER2-肿瘤的患者的597个样品具有可用的基因表达数据和完整的LRR信息。将该患者子组进一步分成训练队列(N=243)和验证队列(N=354)(附加信息见下文和图1)。

[0209] Princess Margaret队列

[0210] Princess Margaret队列是在加拿大进行的随机试验,并且先前已描述过(参见,Fyles,A.W.等人,The New England journal of medicine 351,963-970(2004);Liu,F.F.等人,Journal of clinical oncology:official journal of the American Society of Clinical Oncology 33,2035-2040(2015))。简言之,该试验将769名年龄在50岁或以上的T1或T2淋巴结阴性乳腺癌女性患者随机分配至BCS后的辅助全乳腺RT或非RT。Princess Margaret试验中的所有患者都用他莫昔芬治疗。如前所述通过ER、PgR、HER2和Ki67的免疫组织化学染色进行亚型分型(参见,Liu,F.F.等人,Journal of clinical oncology:official journal of the American Society of Clinical Oncology 33,2035-2040(2015))。基因表达数据可从来自患有ER+、HER2-肿瘤的患者的132个样品和完整的LRR信息获得(图2)。

[0211] 基因表达数据

[0212] 基因表达数据(Gene Expression Omnibus,GSE119295)是从SweBCG91-RT试验中包括的764名患者的原发性肿瘤获得的(参见,Sjöström,M.等人,Journal of Clinical Oncology 37,3340-3349(2019))。与SweBCG91-RT队列相似,Princess Margaret队列的基因表达数据是在CLIA/CAP认证的实验室(Decipher Biosciences,San Diego,CA)中由GeneChip人外显子1.0ST阵列(Thermo Fisher Scientific,South San Francisco)生成的。

[0213] 训练组和验证组的开发

[0214] 在该分析中,仅包括患有ER+、HER2-肿瘤的患者。这是为了确保所得的签名可区分临床上低风险的患者的风险,这些患者将是省略RT的候选者。此外,由于全身性疗法会影响LRR并且仅有一小部分SweBCG91-RT患者接受了全身性治疗,因此从SweBCG91-RT队列分析中去除了接受化学疗法和/或内分泌疗法的患者,以改善结果解释。在这些考虑之后,SweBCG91-RT试验的597名患者被纳入本研究。

[0215] 为了提高信噪比以用于识别由SweBCG91-RT数据组成的训练队列中的LRR事件,仅包括记录的局部区域事件为10年前的第一事件的患者或随访时间大于10年的没有任何记录的事件的患者。这些患者中的60％被保留用于训练队列,其余所有患者用于验证(图1)。在SweBCG91-RT训练和验证队列的无RT组与RT组之间,临床病理学变量分布没有显著差异(表1和表2),除了训练队列中的年龄,其中相对于无RT组,RT组中40至59岁女性的比例增加。在训练和验证队列中,RT组的同侧乳腺肿瘤复发率均低于无RT(表3)。保留具有来自ER+、HER2-肿瘤患者的可用基因组和LRR数据的整个Princess Margaret队列用于验证,没有一个用于训练目的。

[0216] 基因集富集分析

[0217] 为了鉴定与LRR相关的基因集,在未用RT治疗的训练队列的131名患者中进行基因集富集分析(GSEA)(参见,Subramanian,A.等人,Proc Natl Acad Sci U S A 102,15545-15550,(2005);Mootha,V.K.等人,Nat Genet 34,267-273(2003))。首先,对于每个注释的基因转录物,计算LRR的Cox模型。然后将预先排序的基因列表输入GSEA,其中对于每个基因,用以下等式产生统计量:-log(p值)*(如果HR>1,则为1,否则-1),其中p值和危害比(HR)为从Cox模型到LRR(参见表8,每个基因的确定的统计量/系数值)。因此,对于较高LRR率的患者,高预后的个别基因将具有大的正值,而对于较低LRR率的患者,高预后的个别基因将具有大的负值。将排序的列表输入GSEA(版本4.0.3),并指定Hallmarks(H)、C2和C5集合(Molecular Signatures database,版本7.2,MSigDB)(参见,Liberzon,A.等人,Cell systems 1,417-425(2015))。由于基因集集合之间的固有差异,分别分析H、C2和C5列表以寻找集合间的趋势。对于所分析的基因,基于针对每个基因确定的系数值累积总富集评分(参见表8)。具有正富集评分的基因集被解释为与较差的预后(LRR增加)相关联,而具有负富集评分的基因集被解释为与较好的预后(LRR减少)相关联。

[0218] 模型开发

[0219] 在GSEA基因列表中,错误发现率q值<0.05并且在H、C2和C5集合的正和负基因列表中排在前10位的基因被鉴定为模型的潜在候选者(表4)。这些指定基因以及编码ER、PgR、HER2和Ki67的基因与满足以下要求的基因交叉:当在Cox模型中对训练组中未用RT治疗的患者的LRR终点进行评估时,在标准偏差值的顶部四分位数中具有标准偏差的基因和p值<0.05的基因(图3)。将得到的82个基因输入弹性网络回归模型(Rglmnet,α=0.5),用于在未用RT治疗的患者中的训练队列中进行训练,将LRR用作终点。弹性网络回归建模得到该模型的最终16个基因的集。通过使用SweBCG91-RT训练数据集中评分的第25百分位值,应用预定截止值,将评分二分为高与低。然后将该阈值应用于SweBCG91-RT和Princess Margaret验证组。

[0220] 统计方法

[0221] 使用R版本4.0.2(https://cran.r-project.org/bin/windows/base/old/4.0.2/)进行统计学分析。主要终点是LRR的累积发生率,使用至LRR的时间作为第一事件。使用竞争风险方法(R cmprsk软件包)和HRs计算累积发生率,并使用特定原因Cox比例风险回归计算二分风险组之间显著差异的检验。如前所述,当用图表和Schoenfeld检验检查时,对于SweBCG91RT队列中的若干变量,风险与时间不成比例,并且HRs应被解释为整个时间段内的平均值(参见,Liberzon,A.等人,Cell systems 1,417-425(2015))。远处转移和无复

发的死亡被认为是竞争事件。同时有远处转移和LRR(定义为与转移同时或在3个月内登记的LRR)的患者被认为具有LRR。风险比和点估计值用括号中的95％置信区间(CI)报告。p值小于0.05被认为具有统计学显著性。对于人口统计表,使用卡方检验测试治疗组和未治疗组之间的组分布差异,对2x2表进行Yates连续性校正。单变量和多变量Cox分析以及相互作用测试使用连续签名评分进行,因此不依赖于任何阈值。评分与RT之间的相互作用的测试包括SweBCG91-RT和Princess Margaret验证队列中的所有患者,并按队列分层。

[0222] 结果

[0223] 与局部区域复发相关联的生物学概念

[0224] 在SweBCG91-RT训练队列中未接受RT治疗的患者中,我们使用GSEA来识别与乳腺癌中LRR相关联的生物学概念。在分子签名数据库Hallmark、C2和C5集合中,靠前的正富集基因列表包括涉及细胞周期和增殖的列表(表4),而靠前的负富集基因列表包括涉及免疫系统的列表。

[0225] 在未给予RT的训练组患者中,靠前的基因列表与LRR的信息量最大和个体化预后基因相交。使用LRR终点训练具有弹性网络正则化的广义线性模型,并且所得的16基因签名包括结合从GSEA列表(表5)广泛示出的概念的基因。我们将此签名命名为省略局部辅助放射的谱(POLAR)。表8示出了用于生成用于评估患者中LRR预后的富集评分的每个基因的确定的统计量/系数值。

[0226] 使用线性算法生成风险评分,并使用POLAR将患者归类为低风险或高风险。具体地,确定签名(表5)中16个基因中的每一个的表达水平,并乘以它们各自的系数值(表8),以产生每个所测定基因的加权表达值。然后将加权表达值加在一起以生成最终的风险评分。然后将0.5622的预定阈值应用于所生成的风险评分,以将患者分配至高或低风险类别。例如,风险评分高于0.5622的患者被归类为高风险,风险评分低于0.5622的患者被归类为低风险。

[0227] 在SweBCG91-RT验证队列中,POLAR是未接受RT治疗的患者的LRR预后

[0228] 在SweBCG91-RT验证队列中,在以LRR为终点的单变量Cox模型中,POLAR是未接受放射疗法治疗的患者的LRR预后(HR＝1.7[1.3,2.2],p<0.001)。在包括年龄、等级、肿瘤大小和管腔A对管腔B的多变量模型中,该签名仍然是预后(HR＝1.7[1.2,2.3],p<0.001)(表6)。

[0229] 在SweBCG91-RT和Princess Margaret验证队列中,RT对低和高POLAR风险患者的差异性影响

[0230] 按照POLAR在验证队列中归类为低风险的SweBCG91-RT患者的LRR累积发生率在10年时小于10％,无论是否接受放射疗法,并且无法显示RT获益(10年LRR累积发生率,无RT:6％[2％-16％],RT:5％[1％-13％],HR＝1.1[0.39-3.4],p＝0.81,图4A)。然而,鉴于患者人数较少(N＝108),必须谨慎解释这些结果。对于按照签名归类为高风险的SweBCG91-RT患者,与未接受RT的患者相比,RT导致LRR风险显著降低(10年局部区域累积发生率,无RT:19％[13％-27％],RT:8％[4％-14％],HR＝0.43[0.24-0.78],p＝0.0055,图4B)。

[0231] 然后我们评估了Princess Margaret试验的132名患者的POLAR(图5)。在表7中提供了该队列的人口统计学和描述性详情。对于BCS后单独使用他莫昔芬治疗的女性,按照POLAR归类为低风险的患者的10年LRR累积发生率为7％[0％-27％](无RT)和13％[2％-

34％](有RT)。虽然患者数量少(N＝34)排除了明确结论(图5A),但在这些低风险女性中无法显示RT获益(HR＝1.5[0.14-16],p＝0.74)。按照POLAR归类为高风险的女性在10年时的LRR累积发生率为22％[10％-36％](无RT)和8％[2％-20％](有RT),并且RT显著获益(HR＝0.25[0.07-0.92],p＝0.038)。我们通过对SweBCG91-RT和Princess Margaret可评估样品的总计N＝486名患者测试RT与连续POLAR评分之间的相互作用进行了探索性分析,并且相互作用p值为0.066。

[0232] 这些结果表明,本公开的方法和分类器可用于预示局部区域复发,并且鉴定不会受益于辅助放射疗法(放射治疗)的乳腺癌患者。结果还表明,本公开的方法和分类器可鉴定局部区域复发低风险的乳腺癌患者。这些结果进一步表明,本发明的方法和分类器可用于治疗乳腺癌。

[0233] 本公开内容的基因组分类器鉴定了保乳手术后在没有辅助放射疗法的情况下处于低LRR风险的患者,因此将它们鉴定为保乳手术后省略放射疗法的候选者。此外,本公开的基因组分类器显示,LRR风险最低的乳腺癌患者没有显著受益于RT。

[0234] 表1.SweBCG91-RT训练队列人口统计学

[0235]

N (%)	无 RT N=131	RT N=112	总计 N=243	p 值
手术时年龄(岁)				
≤39	5 (3.8)	1 (0.9)	6 (2.5)	0.03
40-49	26 (19.8)	26 (23.2)	52 (21.4)	
50-59	34 (26)	47 (42)	81 (33.3)	
60-69	57 (43.5)	34 (30.4)	91 (37.4)	

[0236]

≥70	9 (6.9)	4 (3.6)	13 (5.3)	
绝经状态				
绝经前	33 (25.2)	24 (21.4)	57 (23.5)	0.63
绝经后	95 (72.5)	84 (75)	179 (73.7)	
不适用	3 (2.3)	4 (3.6)	7 (2.9)	
组织学分级				
1 级	14 (10.7)	24 (21.4)	38 (15.6)	0.07
2 级	99 (75.6)	73 (65.2)	172 (70.8)	
3 级	18 (13.7)	14 (12.5)	32 (13.2)	
不适用	0 (0)	1 (0.9)	1 (0.4)	
肿瘤大小(mm)				
≤10	55 (42)	43 (38.4)	98 (40.3)	0.75
11-20	71 (54.2)	64 (57.1)	135 (55.6)	
21-30	4 (3.1)	3 (2.7)	7 (2.9)	
不适用	1 (0.8)	2 (1.8)	3 (1.2)	
亚型 IHC				
管腔 A	93 (71)	73 (65.2)	166 (68.3)	0.41
管腔 B (HER2-)	38 (29)	39 (34.8)	77 (31.7)	
辅助内分泌疗法				
否	131 (100)	112 (100)	243 (100)	
辅助化学疗法				
否	131 (100)	112 (100)	243 (100)	

[0237] 表2.SweBCG91-RT验证队列人口统计学

[0238]

N (%)	**无 RT N=178**	**RT N=176**	**总计 N=354**	**p 值**
手术时年龄(岁)				
≤39	4 (2.2)	3 (1.7)	7 (2)	0.69
40-49	34 (19.1)	26 (14.8)	60 (16.9)	
50-59	54 (30.3)	49 (27.8)	103 (29.1)	
60-69	56 (31.5)	64 (36.4)	120 (33.9)	
≥70	30 (16.9)	34 (19.3)	64 (18.1)	
绝经状态				
绝经前	36 (20.2)	29 (16.5)	65 (18.4)	0.37
绝经后	135 (75.8)	145 (82.4)	280 (79.1)	

[0239]

不适用	7 (3.9)	2 (1.1)	9 (2.5)	
组织学分级				
1 级	32 (18)	28 (15.9)	60 (16.9)	0.36
2 级	108 (60.7)	118 (67)	226 (63.8)	
3 级	34 (19.1)	25 (14.2)	59 (16.7)	
不适用	4 (2.2)	5 (2.8)	9 (2.5)	
肿瘤大小 (mm)				
≤10	70 (39.3)	69 (39.2)	139 (39.3)	0.78
11-20	97 (54.5)	99 (56.2)	196 (55.4)	
21-30	11 (6.2)	8 (4.5)	19 (5.4)	
亚型 IHC				
管腔 A	115 (64.6)	113 (64.2)	228 (64.4)	1.00
管腔 B (HER2-)	63 (35.4)	63 (35.8)	126 (35.6)	
辅助内分泌疗法				
否	178 (100)	176 (100)	354 (100)	
辅助化学疗法				
否	178 (100)	176 (100)	354 (100)	

[0240] 表3.SweBCG91-RT训练和验证队列中的事件分布

[0241]

训练队列(N=243)				
N (%)	无 RT(N=131)	RT(N=112)	总计	p 值
同侧乳腺肿瘤复发				
否	91 (69.5)	97 (86.6)	188 (77.4)	0.0025
是	40 (30.5)	15 (13.4)	55 (22.6)	
区域复发				
否	125 (95.4)	109 (97.3)	234 (96.3)	0.66
是	6 (4.6)	3 (2.7)	9 (3.7)	
远处转				

[0242]

移				
否	120 (91.6)	107 (95.5)	227 (93.4)	0.33
是	11 (8.4)	5 (4.5)	16 (6.6)	
乳腺癌特异性死亡				
否	116 (88.5)	102 (91.1)	218 (89.7)	0.66
是	15 (11.5)	10 (8.9)	25 (10.3)	
死于任何原因				
否	89 (67.9)	84 (75)	173 (71.2)	0.28
是	42 (32.1)	28 (25)	70 (28.8)	
验证队列(N=354)				
	无 RT(N=178)	**RT(N=176)**	**总计(N=354)**	**p 值**
同侧乳腺肿瘤复发				
否	135 (75.8)	153 (86.9)	288 (81.4)	0.01
是	43 (24.2)	23 (13.1)	66 (18.6)	
区域复发				
否	170 (95.5)	172 (97.7)	342 (96.6)	0.39
是	8 (4.5)	4 (2.3)	12 (3.4)	
远处转移				
否	151 (84.8)	146 (83)	297 (83.9)	0.74
是	27 (15.2)	30 (17)	57 (16.1)	
乳腺癌特异性死亡				
否	141 (79.2)	136 (77.3)	277 (78.2)	0.75
是	37 (20.8)	40 (22.7)	77 (21.8)	
死于任何原因				
否	79 (44.4)	77 (43.8)	156 (44.1)	0.99
是	99 (55.6)	99 (56.2)	198 (55.9)	

[0243] 表4.通过GSEA鉴定为高度正富集和负富集的基因列表。

[0244]

	正 (与预后较差相关)	负 (与预后较好相关)
H	HALLMARK_E2F_TARGETS HALLMARK_MYC_TARGETS_V2 HALLMARK_G2M_CHECKPOINT HALLMARK_MYC_TARGETS_V1 HALLMARK_ESTROGEN_RESPONSE_LATE HALLMARK_MYOGENESIS HALLMARK_OXIDATIVE_PHOSPHORYLATION HALLMARK_ESTROGEN_RESPONSE_EARLY HALLMARK_MTORC1_SIGNALING HALLMARK_HEDGEHOG_SIGNALING	HALLMARK_TNFA_SIGNALING_VIA_NFKB HALLMARK_KRAS_SIGNALING_UP HALLMARK_ALLOGRAFT_REJECTION
C2	SOTIRIOU_BREAST_CANCER_GRADE_1_VS_3_UP ROSTY_CERVICAL_CANCER_PROLIFERATION_CLUSTER NIKOLSKY_BREAST_CANCER_16P13_AMPLICON ZHOU_CELL_CYCLE_GENES_IN_IR_RESPONSE_24HR BLANCO_MELO_BRONCHIAL_EPITHELIAL_CELLS_INFLUENZA_A_DEL_NS1_INFECTION_DN DUTERTRE_ESTRADIOL_RESPONSE_24HR_UP WHITEFORD_PEDIATRIC_CANCER_MARKERS ZHAN_MULTIPLE_MYELOMA_PR_UP LEE_EARLY_T_LYMPHOCYTE_UP	SMID_BREAST_CANCER_NORMAL_LIKE_UP WINTER_HYPOXIA_DN PHONG_TNF_TARGETS_UP ZHENG_FOXP3_TARGETS_IN_T_LYMPHOCYTE_DN KEGG_GRAFT_VERSUS_HOST_DISEASE BOQUEST_STEM_CELL_CULTURED_VS_FRESH_DN BERTUCCI_INVASIVE_CARCINOMA_DUCTAL_VS_LOBULAR

[0245]

	CHIANG_LIVER_CANCER_SUBCLASS_PROLIFERATION_UP	_DN NAKAYAMA_SOFT_TISSUE_TUMORS_PCA2_DN REACTOME_CHEMOKINE_RECEPTORS_BIND_CHEMOKINES KEGG_AUTOIMMUNE_THYROID_DISEASE
C5	GO_ANAPHASE_PROMOTING_COMPLEX_DEPENDENT_CATABOLIC_PROCESS GO_DNA_REPLICATION_INDEPENDENT_NUCLEOSOME_ORGANIZATION GO_CENTROMERE_COMPLEX_ASSEMBLY GO_CHROMATIN_REMODELING_AT_CENTROMERE	GO_B_CELL_RECEPTOR_SIGNALING_PATHWAY GO_POSITIVE_REGULATION_OF_INTERLEUKIN_2_PRODUCTION GO_INTERLEUKIN_1_BETA_PRODUCTION GO_REGULATION_OF_ANTIGEN_RECEPTOR_MEDIATED_SIGNALING_PATHWAY GO_NEGATIVE_REGULATION_OF_SMOOTH_MUSCLE_CELL_PROLIFERATION GO_INTERLEUKIN_1_PRODUCTION GO_REGULATION_OF_B_CELL_RECEPTOR_SIGNALING_PATHWAY GO_RESPONSE_TO_CHEMOKINE GO_LYMPHOCYTE_CHEMOTAXIS GO_DENDRITIC_CELL_CHEMOTAXIS

[0246] 表5.模型中的16个基因

[0247]

AGR2	前梯度2,蛋白二硫键异构酶

B4GALT1	β-1,4-半乳糖基转移酶1
CLDN7	密封蛋白7
EZR	埃兹蛋白
GNG11	G蛋白亚基γ11
JUN	Jun原癌基因
MMP11	基质金属肽酶11
PKIB	CAMP依赖性蛋白激酶抑制剂β
PRPS1	磷酸核糖焦磷酸合成酶1
PSMD10	蛋白酶体26S亚基,非ATP酶10
SH3BP5	SH3结构域结合蛋白5
SLC16A3	溶质载体家族16成员3
SLC7A11	溶质载体家族7成员11
SPP1	分泌型磷蛋白1
TNNT1	肌钙蛋白T1,慢骨骼肌型
UBE2E1	泛素结合酶E2 E1

[0248] 表6.

[0249]

单变量分析		**HR [95%CI]**	**p 值**
POLAR		1.7 [1.3,2.2]	<0.001
多变量分析		**HR [95%CI]**	**p 值**
POLAR		1.7 [1.2,2.3]	<0.001
年龄		0.98 [0.95, 1.0]	0.1
组织学分级	1	参考文献	
	2	1.1 [0.43, 2.9]	0.82
	3	0.74 [0.23, 2.4]	0.61
肿瘤大小	≤20mm	参考文献	
	>20mm ，≤50mm	0.41 [0.055, 3.1]	0.39

[0250]

亚型	管腔 A	参考文献	
	管腔 B	1.5 [0.75, 2.8]	0.27

[0251] 表7.

[0252]

N (%)	无 RT N=62	RT N=70	总计 N=132	p 值
手术时年龄(岁)				
50-59	16 (25.8)	14 (20)	30 (22.7)	0.76
60-69	23 (37.1)	24 (34.3)	47 (35.6)	
70-79	19 (30.6)	27 (38.6)	46 (34.8)	
≥80	4 (6.5)	5 (7.1)	9 (6.8)	
组织学分级				
I	9 (14.5)	9 (12.9)	18 (13.6)	0.43
II	34 (54.8)	45 (64.3)	79 (59.8)	
III	17 (27.4)	13 (18.6)	30 (22.7)	
不适用	2 (3.2)	3 (4.3)	5 (3.8)	
肿瘤大小(mm)				
≤10	15 (24.2)	9 (12.9)	24 (18.2)	0.41
11-20	36 (58.1)	46 (65.7)	82 (62.1)	
21-30	10 (16.1)	14 (20)	24 (18.2)	
≥31	1 (1.6)	1 (1.4)	2 (1.5)	
ER 状态				
正	62 (100)	70 (100)	132 (100)	
PR 状态				
负	7 (11.3)	10 (14.3)	17 (12.9)	0.77
正	49 (79.0)	52 (74.3)	101 (76.5)	
不适用	6 (9.7)	8 (11.4)	14 (10.6)	
HER2 状态				
负	62 (100)	70 (100)	132 (100)	

[0253] 表8.

[0254]

基因缩写	基因名称	系数值	Entrez 基因 Id 编号

[0255]

AGR2	前梯度 2，蛋白二硫键异构酶	0.200353	10551
B4GALT1	β-1,4-半乳糖基转移酶 1	-0.33626	2683
CLDN7	密封蛋白 7	0.115625	1366
EZR	埃兹蛋白	0.003256	7430
GNG11	G 蛋白亚基 γ11	-0.41974	2791
JUN	Jun 原癌基因	-0.01851	3725
MMP11	基质金属肽酶 11	0.444806	4320
PKIB	CAMP 依赖性蛋白激酶抑制剂 β	0.170331	5570
PRPS1	磷酸核糖焦磷酸合成酶 1	0.170411	5631
PSMD10	蛋白酶体 26S 亚基，非 ATP 酶 10	0.042903	5716
SH3BP5	SH3 结构域结合蛋白 5	-0.17135	9467
SLC16A3	溶质载体家族 16 成员 3	0.214877	9123
SLC7A11	溶质载体家族 7 成员 11	0.292485	23657
SPP1	分泌型磷蛋白 1	0.183821	6696
TNNT1	肌钙蛋白 T1，慢骨骼肌型	0.396711	7138
UBE2E1	泛素结合酶 E2 E1	0.033286	7324

[0256] 等效方案

[0257] 在不脱离本发明的精神或基本特征的情况下，本发明能以其他具体形式来实施。因此，前述实施方案在所有方面都应被视为说明性的，而非是对本文中描述的本发明的限制。因此，本发明的范围是由随附权利要求而非由上述描述所指示，并且属于权利要求的等效性含义和范围内的所有变化都意图包括在其中。

[0258] 以引用方式并入

[0259] 本文提及的每个专利文献和科学文章的全部公开内容出于所有目的以引用方式并入本文。

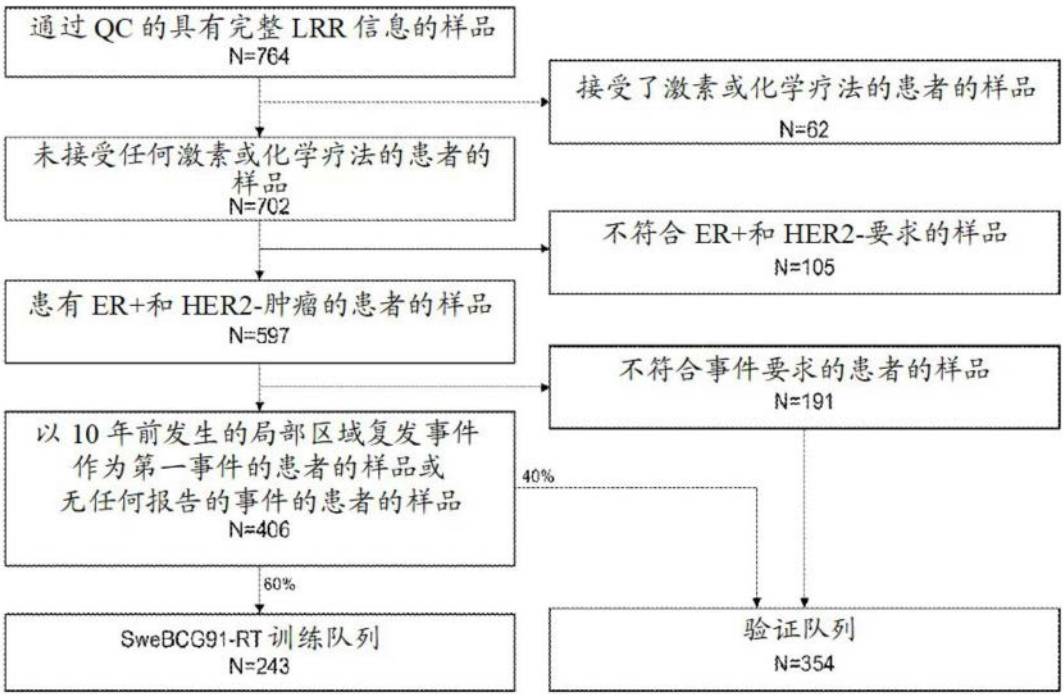

通过 QC 的具有完整 LRR 信息的样品
N=764
接受了激素或化学疗法的患者的样品
N=62
未接受任何激素或化学疗法的患者的样品
N=702
不符合 ER+和 HER2-要求的样品
N=105
患有 ER+和 HER2-肿瘤的患者的样品
N=597
不符合事件要求的患者的样品
N=191
以 10 年前发生的局部区域复发事件作为第一事件的患者的样品或无任何报告的事件的患者的样品
N=406
40%
60%
SweBCG91-RT 训练队列
N=243
验证队列
N=354

图1

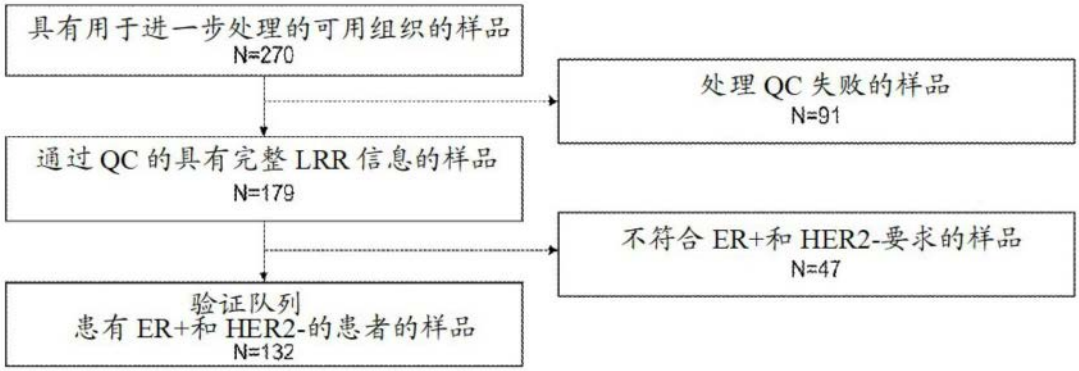

具有用于进一步处理的可用组织的样品
N=270
处理 QC 失败的样品
N=91
通过 QC 的具有完整 LRR 信息的样品
N=179
不符合 ER+和 HER2-要求的样品
N=47
验证队列
患有 ER+和 HER2-的患者的样品
N=132

图2

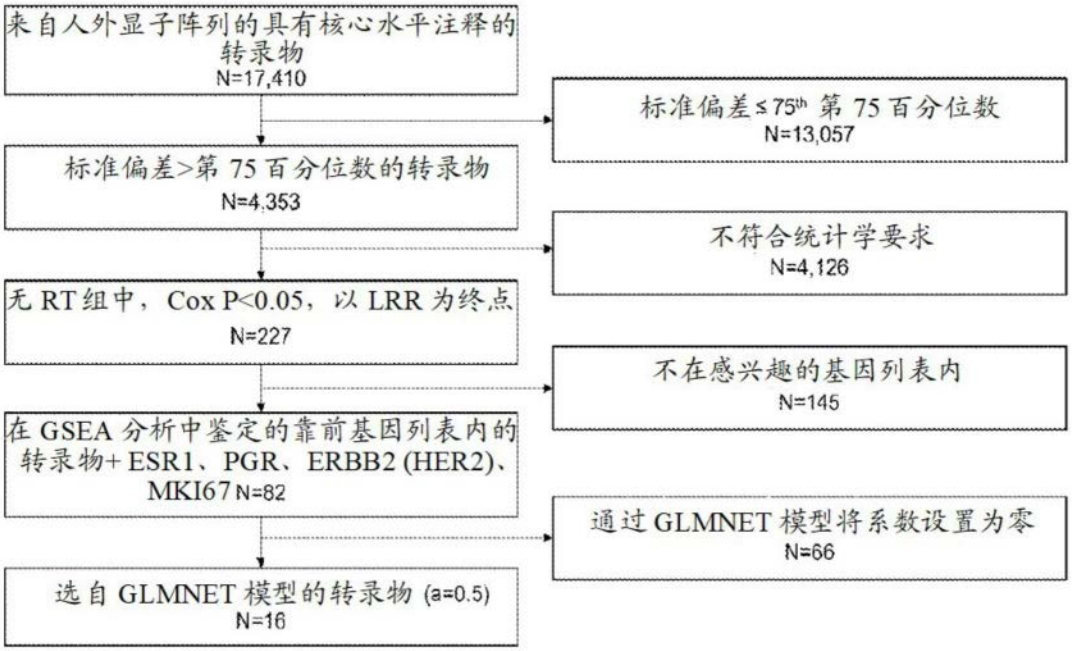

来自人外显子阵列的具有核心水平注释的转录物
N=17,410
标准偏差≤75th 第 75 百分位数
N=13,057
标准偏差>第 75 百分位数的转录物
N=4,353
不符合统计学要求
N=4,126
无 RT 组中，Cox P<0.05，以 LRR 为终点
N=227
不在感兴趣的基因列表内
N=145
在 GSEA 分析中鉴定的靠前基因列表内的转录物+ ESR1、PGR、ERBB2 (HER2)、MKI67 N=82
通过 GLMNET 模型将系数设置为零
N=66
选自 GLMNET 模型的转录物 (a=0.5)
N=16

图3

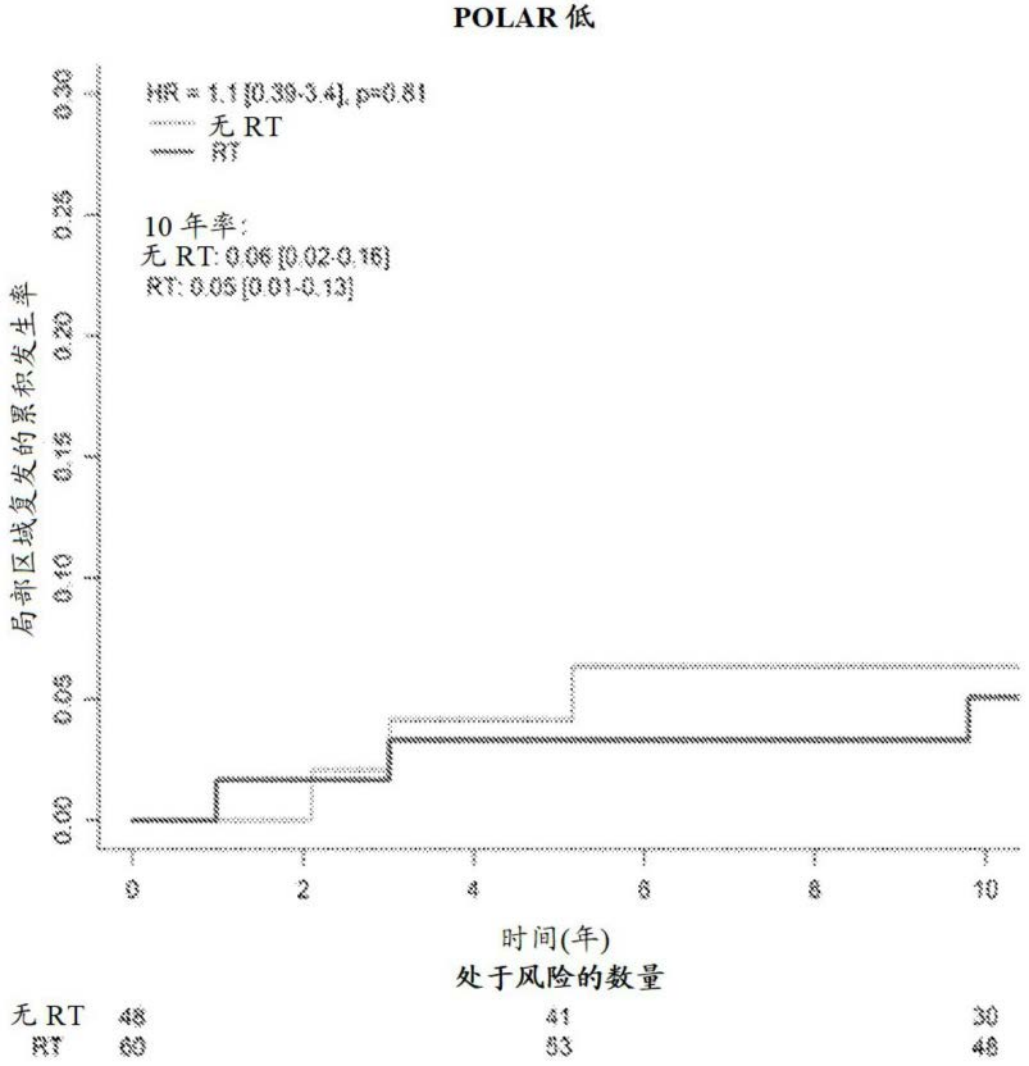

POLAR 低
HR = 1.1 [0.39-3.4], p=0.81
无 RT
RT
10 年率:
无 RT: 0.06 [0.02-0.16]
RT: 0.05 [0.01-0.13]
局部区域复发的累积发生率
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
2
4
6
8
10
时间(年)
处于风险的数量
无 RT 48 41 30
RT 83 53 46

图4A

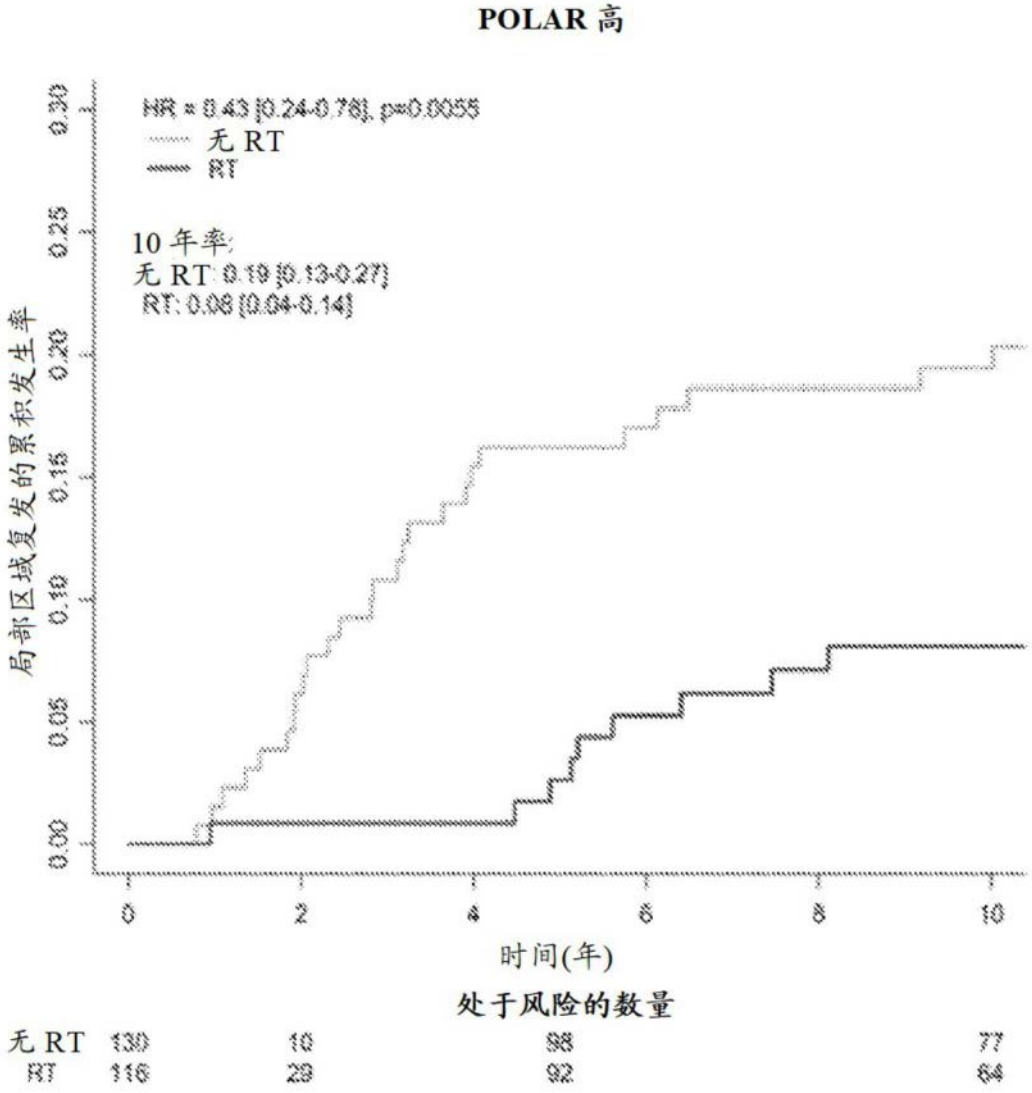

POLAR 高
HR = 0.43 [0.24-0.76], p=0.0055
无 RT
RT
10 年率:
无 RT: 0.19 [0.13-0.27]
RT: 0.08 [0.04-0.14]
局部区域复发的累积发生率
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
2
4
6
8
10
时间(年)
处于风险的数量
无 RT 130 10 98 77
37 116 29 92 64

图4B

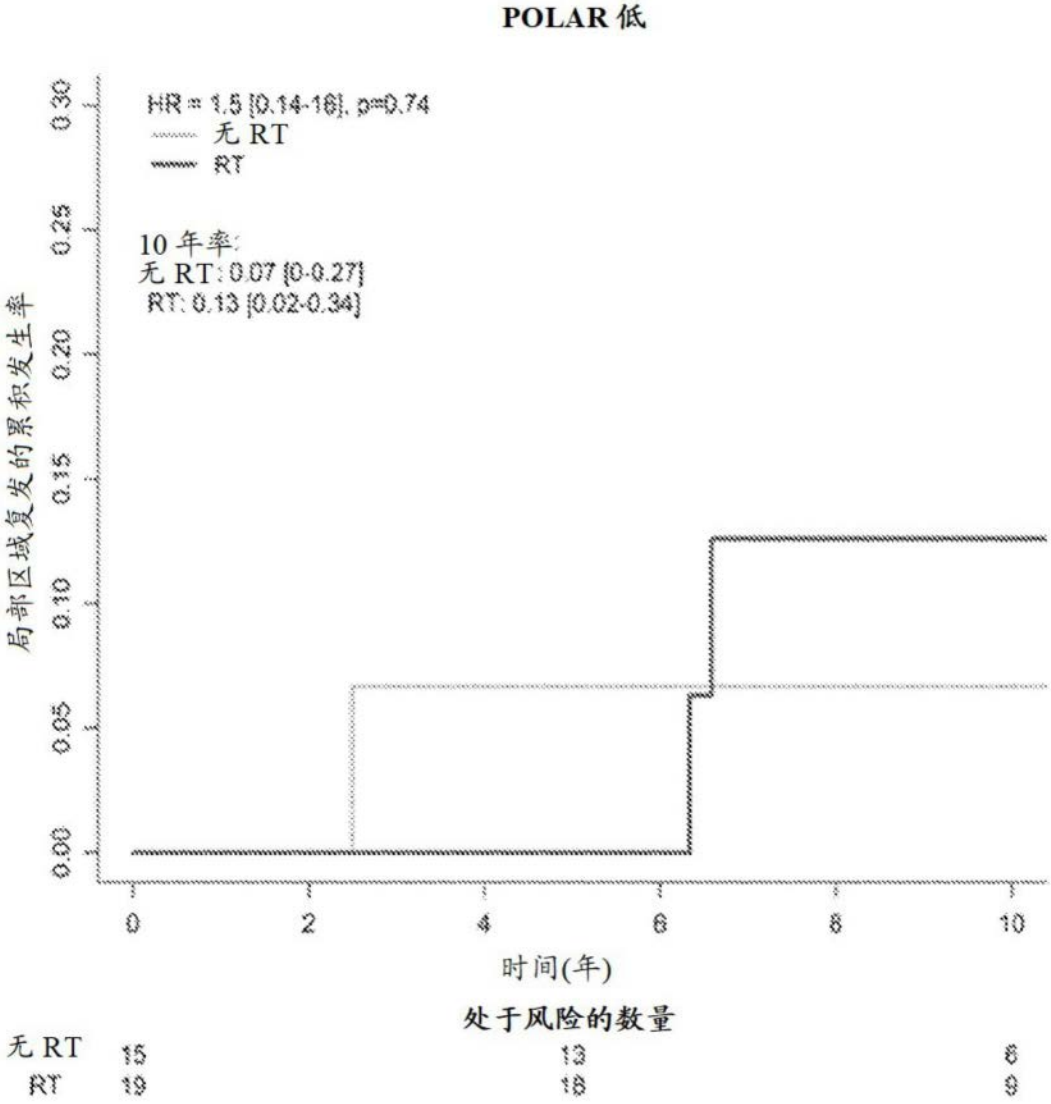

POLAR 低
HR = 1.5 [0.14-16], p=0.74
无 RT
RT
10 年率:
无 RT: 0.07 [0-0.27]
RT: 0.13 [0.02-0.34]
局部区域复发的累积发生率
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
2
4
6
8
10
时间(年)
处于风险的数量
无 RT 15 13 6
RT 19 18 9

图5A

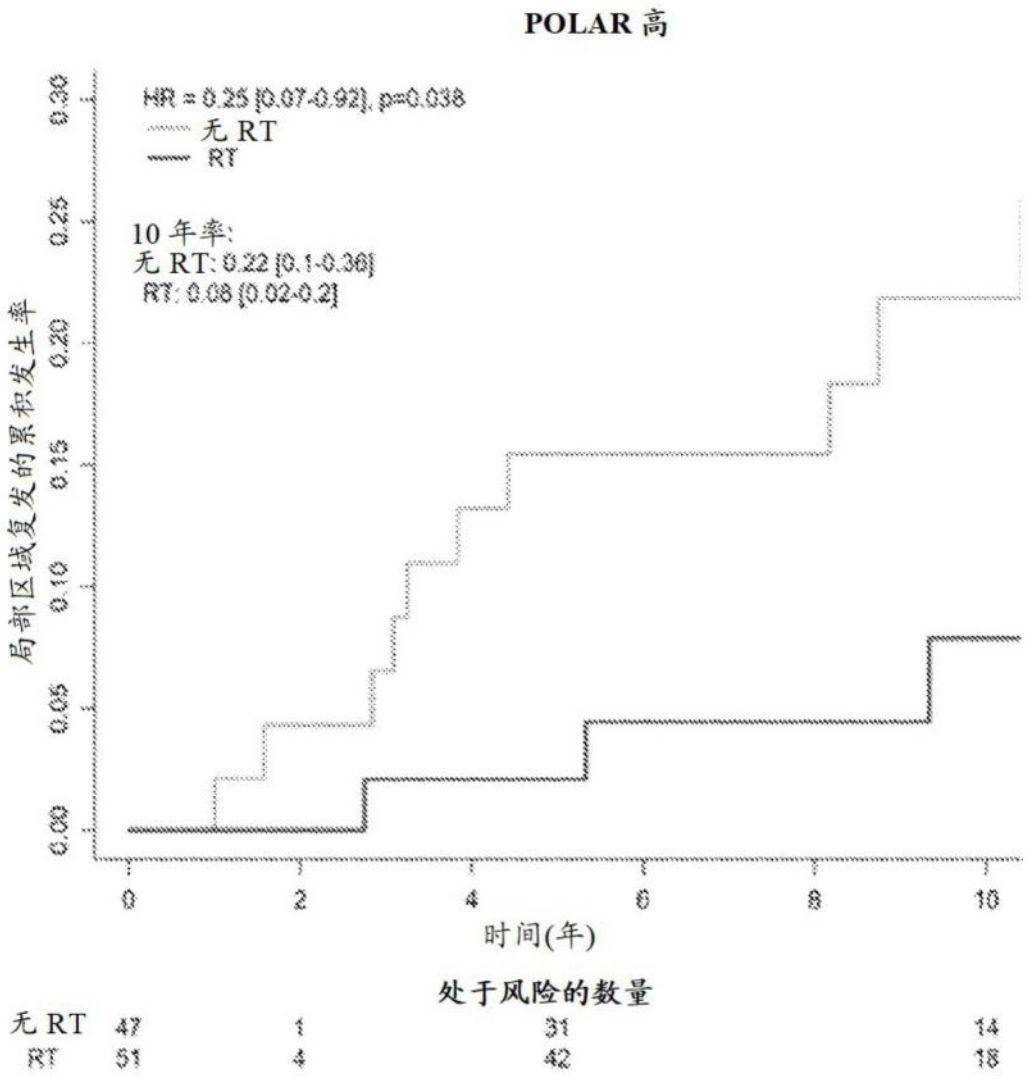

POLAR 高
HR = 0.25 [0.07-0.92], p=0.038
无 RT
RT
10 年率:
无 RT: 0.22 [0.1-0.36]
RT: 0.08 [0.02-0.2]
局部区域复发的累积发生率
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
2
4
6
8
10
时间(年)
处于风险的数量
无 RT 47 1 31 14
RT 51 4 42 18

图5B