



(12)发明专利

(10)授权公告号 CN 106399304 B

(45)授权公告日 2019.02.01

(21)申请号 201611027831.3

CN 101535500 A,2009.09.16,

(22)申请日 2016.11.18

CN 106048060 A,2016.10.26,

(65)同一申请的已公布的文献号

dbSNP:ss1691080190.《dbSNP》.2015,第1-3页.

申请公布号 CN 106399304 A

Anne Marie McCarthy等.Incremental

(43)申请公布日 2017.02.15

impact of breast cancer SNP panel on risk classification in a screening population of white and African American women.

(73)专利权人 深圳市第二人民医院

地址 广东省深圳市福田区笋岗西路3002号

《Breast Cancer Res Treat》.2014,第138卷(第3期),全文.

(72)发明人 何劲松 陈伟财 欧阳依雯

毛有胜 罗雪莹 潘悦 刘宝儿

李锋

吴桐等.全基因组相关研究SNP多态性阳性位点与黑龙江省女性乳腺癌发病相关性研究.

《中国卫生标准管理》.2015,第6卷(第27期),全文.

(51)Int.Cl.

C12N 15/11(2006.01)

C12Q 1/6886(2018.01)

审查员 程珂

(56)对比文件

CN 105969859 A,2016.09.28,

权利要求书1页 说明书6页

序列表2页

(54)发明名称

一种与乳腺癌相关的SNP标记

(57)摘要

本发明公开了一种与乳腺癌相关的生物标记物,所述生物标记物包括位于基因C12orf45核苷酸序列自5'端起第8187位碱基为G的SNP位点,所述位点为C12orf45:NM\_152318:exon4:c.C368G。本发明研究SNP在乳腺癌辅助诊断的应用前景,阐述SNP对于乳腺癌进展的影响,揭示其诊断价值。因此,本发明通过SNP生物标志物和诊断试剂盒的研制和应用,可使得乳腺癌的诊断更加方便易行,为临床医生快速准确掌握患者病情,为临床治疗效果评价奠定基础,并为发现具有潜在治疗价值的新型小分子药物靶标提供帮助。

1. 一种与乳腺癌相关的SNP分子标记,其特征在于,所述SNP分子标记包括位点位于SEQ ID NO:1所示的核苷酸序列的第239位的SNP,其可以是G或C。
2. 权利要求1所述SNP分子标记在制备预测乳腺癌诊断试剂中的应用。
3. 如权利要求2所述的应用,其特征在于,所述的试剂包括用于扩增所述SNP分子标记的引物对。
4. 如权利要求2所述的应用,其特征在于,所述的试剂包括用于扩增所述SNP分子标记的引物对和限制性内切酶。
5. 如权利要求3或4所述的应用,其特征在于,扩增所述SNP分子标记的引物对为SEQ ID NO:3-4所示的核苷酸序列。

## 一种与乳腺癌相关的SNP标记

### 技术领域

[0001] 本发明涉及生物医学技术领域,具体涉及一种与乳腺癌相关的SNP标记。

### 背景技术

[0002] 乳腺癌是一种全身性疾病、其发生和发展是一个涉及多因素、多环节的复杂过程,包括癌基因的激活以及抑癌基因的失活等。因此,基因突变在乳腺癌的发生、发展过程中起着非常重要的作用。

[0003] 乳腺癌是一个多因素遗传变异性疾病,只有不足10%是由于单基因缺陷引起的。随着高通量基因技术的发展,越来越多与乳腺癌相关基因被发现,这些基因上潜在的遗传变异(单核苷酸多态和拷贝数变异)可能引起乳腺癌药物治疗效果的差异。由于遗传变异的存在使抗肿瘤药物的代谢途径以及药物作用的目标基因可能受到影响,进而影响疗效以及预后。

[0004] SNP(singlenucleotidepolymorphism,SNP,即单核苷酸多态性)是1996年由美国麻省理工学院的人类基因组研究中心学者Lander提出的一类分子遗传标记,主要是指基因组水平上由单个核苷酸的变异所引起的DNA序列多态性。SNP表现出的多态性仅涉及到单个碱基的变异,表现是有转换、颠换、插入和缺失等。单核苷酸多态性为第三代遗传标志,人体许多表型差异、对药物或疾病的易感性等等都可能与SNP有关。目前对于不同分型乳腺癌预后、疗效的预测性研究主要集中在SNP水平。

[0005] SNP赋予个体对环境暴露、药物治疗等的不同反应,从而产生不同的表型,因此SNP可能是导致个体疾病发生发展差异的重要遗传基础。利用疾病易感的SNP谱诊断疾病,具有快速、灵敏、准确等特点,因而应用前景广阔。近年来,利用SNP诊断疾病的发生发展已成为临床和科研工作者的研究热点。

[0006] 然而,目前还没有将SNP应用于乳腺癌诊断的报道,若能筛选出乳腺癌易感的SNP作为生物标志物,并研制相应的诊断试剂盒,必将有力地推动我国乳腺癌早期诊断的现状,并为其药物筛选、药效评价及靶向治疗开辟新的途径。

### 发明内容

[0007] 本发明的目的是针对上述技术问题,提出一种与乳腺癌辅助诊断相关的生物标志物。

[0008] 本发明的第二个目的是提供所述生物标志物在预测乳腺癌诊断试剂中的应用。

[0009] 本发明的第三个目的是提供乳腺癌辅助诊断试剂盒。

[0010] 发明人通过分离和研究乳腺癌患者及与其年龄匹配的健康女性对照外周血DNA中的单核苷酸多态性,寻找一组与乳腺癌高度相关的高特异性和敏感性的SNP,并研制出可便于临床应用的乳腺癌辅助诊断试剂盒,为乳腺癌的筛查和诊断提供数据支持。

[0011] 本发明的目的是通过下列技术方案实现的:

[0012] 一种与乳腺癌相关的SNP标记,所述生物标志物包括位于基因C12orf45核苷酸序

列自5'端起第8187位碱基发生了由C到G的SNP位点突变,所述SNP位点突变为C12orf45:NM\_152318:exon4:c.C368G:p.S123W。

[0013] 其中,基因C12orf45(chromosome 12open reading frame 45),是位于人类第12号染色体上的一个开放阅读框,其基因组序列为NC\_000012.12,共8408bp;NM\_152318是C12orf45的一个转录本,所述SNP位点突变发生在该转录本的第4个外显子上,并且第368位发生了由C到G的错义突变,该突变导致了编码蛋白由丝氨酸S到色氨酸W的转变。

[0014] 进一步地,本发明提供了所述生物标记物在预测乳腺癌诊断试剂中的应用。

[0015] 优选的,所述的试剂包括用于扩增所述SNP位点的引物对,或是包括用于扩增所述SNP位点的引物对和限制性内切酶。

[0016] 优选的,扩增所述SNP位点的引物对具有SEQIDNO:3-4所示的核苷酸序列。

[0017] 优选的,所述引物对扩增的核苷酸序列SEQIDNO:1所示。

[0018] 更进一步地,本发明提供了一种乳腺癌辅助诊断的试剂盒,其包括检测位于基因C12orf45核苷酸序列自5'端起第8187位SNP位点基因型的试剂。

[0019] 优选的,所述的试剂包括用于扩增所述SNP位点的引物对,或是包括用于扩增所述SNP位点的引物对和限制性内切酶。

[0020] 优选的,扩增所述SNP位点的引物对具有SEQIDNO:3-4所示的核苷酸序列。

[0021] 优选的,所述试剂盒还包括PCR反应常用的酶和试剂,如dNTPs、Taq酶、Mg<sup>2+</sup>、PCR反应缓冲液等;还可以含有标准品和/或对照品。

[0022] 本发明有益效果:

[0023] 本发明研究SNP在乳腺癌辅助诊断的应用前景,阐述SNP对于乳腺癌进展的影响,揭示其诊断价值。因此,本发明通过SNP生物标志物和诊断试剂盒的研制和应用,可使得乳腺癌的诊断更加方便易行,为临床医生快速准确掌握患者病情,为临床治疗效果评价奠定基础,并为发现具有潜在治疗价值的新型小分子药物靶标提供帮助。

## 具体实施方式

[0024] 以下实施例用于说明本发明,但不用来限制本发明的范围。若未特别指明,实施例中所用的技术手段为本领域技术人员所熟知的常规手段。

[0025] 本发明的技术方案具体包括:采集符合标准的血液样本,系统收集完整的人口学资料和临床资料;基因型检测:选择乳腺癌病例、与乳腺癌病例年龄匹配的健康女性对照,利用外显子测序,找出与乳腺癌相关的SNP;对筛选出的阳性关联SNP,进一步采用基因分型进行检测,验证其应用于临床诊断的可重复性;乳腺癌辅助诊断试剂盒的研制:根据乳腺癌病例和健康女性对照中基因型分布频率有显著差异的SNP开发SNP辅助诊断试剂盒。

[0026] 数据分析中各数值表示如下:

[0027] 1、1jb23\_sift:SIFT分值(version 2.3),表示该变异对蛋白序列的影响,包含三个值,一是SIFT初始分值,二是转换后的值(1-SIFT),三是T或者D。当该变异同时影响多个蛋白序列时,对每条蛋白序列有一个SIFT值,取最小值。SIFT分值越小越“有害”,表明该SNP导致蛋白结构或功能改变的可能性大;D:Deleterious(sift≤0.05);T:tolerated(sift>0.05);

[0028] 2、1jb23\_pp2hvar:利用PolyPhen2基于HumanVar数据库预测该变异对蛋白序列的

影响,用于单基因遗传病。该列包含两个值,第一个是PolyPhen2分值,数值越大越“有害”,表明该SNP导致蛋白结构或功能改变的可能性大;第二个是D或P或B(D:Probably damaging ( $\geq 0.909$ ),P:possibly damaging ( $0.447 \leq \text{pp2\_hvar} \leq 0.909$ );B:benign ( $\text{pp2\_hvar} < 0.446$ ));

[0029] 3、ljb23\_pp2hdiv:利用PolyPhen2基于HumanDiv数据库预测该变异对蛋白序列的影响,用于复杂疾病。该列包含两个值,第一个是PolyPhen 2分值,数值越大越“有害”,表明该SNP导致蛋白结构或功能改变的可能性大;第二个是D或P或B(D:Probably damaging ( $\geq 0.957$ ),P:possibly damaging ( $0.453 \leq \text{pp2\_hdiv} \leq 0.956$ );B:benign ( $\text{pp2\_hdiv} < 0.452$ ));

[0030] 4、ljb23\_mt:tionTaster分值(version 2.3),表示该变异对蛋白序列的影响,包含三个值,一是Mutation Taster初始分值,二是转换后的值,三是A、D、N或者P。第二个值越大越“有害”,表明该SNP导致蛋白结构或功能改变的可能性大,其中“A”(“disease\_causing\_automatic”);“D”(“disease\_causing”);“N”(“polymorphism”);“P”(“polymorphism\_automatic”)。

[0031] 具体来说研究的实验方法主要包括以下几个部分:

[0032] 1. 研究样本的选择

[0033] (1) 经病理学明确诊断的乳腺癌病例25例和与乳腺癌病例年龄匹配的健康女性10例作为对照,其中乳腺癌病例中有3例病人具有癌症家族史;

[0034] (2) 采血前未接受过放疗或化疗、无既往肿瘤病史;

[0035] (3) 与病例年龄匹配的健康女性对照

[0036] 2. 酚-氯仿法提取外周血基因组DNA,按常规方法操作。通常能得到20-50ng/ $\mu$ LDNA,纯度(紫外2600D:2800D)在1.6-2.0。

[0037] 3. 全外显子芯片检测

[0038] (1) 取受试者全基因组DNA样本;

[0039] (2) 在全外显子芯片(北京诺禾致源科技股份有限公司,下同)上进行扫描;

[0040] (3) 检测并比较各基因型在乳腺癌病例与健康女性对照中的分别差异。

[0041] 4. 单个SNP的基因分型

[0042] (1) 取受试者DNA样本;

[0043] (2) 设计单个SNP的特异性扩增引物;

[0044] (3) 进行PCR反应,回收产物进行测序;

[0045] (4) 比较乳腺癌病例与健康女性对照中不同基因型的分布差异。

[0046] 5. 诊断试剂盒制备方法

[0047] 全外显子芯片进行扫描和单个SNP检测后确定乳腺癌病例与健康女性对照中基因型分布频率有显著差异的SNP,作为乳腺癌诊断的指标。筛选出的与乳腺癌发病有关的SNP辅助诊断试剂盒,其包括检测位于基因C12orf45核苷酸序列自5'端起第8187位SNP位点基因型的试剂,诊断试剂盒还可以包括这些SNP的特异性扩增引物,以及Taq酶、dNTPs等试剂。

[0048] 6. 临床应用例

[0049] 利用本发明人制备的乳腺癌辅助诊断试剂盒检测待筛查的乳腺癌患者并与实际临床检测相比较以确定了乳腺癌辅助诊断试剂盒的有效性。具体包括测定受试者血标本

cDNA中上述SNP的特异性扩增引物和其他检测试剂,为临床医生快速准确掌握患者的疾病状态和病情严重程度,及时采取更具个性化的防治方案提供支持。

[0050] 实施例1样品的收集和样品资料的整理

[0051] 发明人于2010年1月至2015年12月在深圳市第二人民医院收集了大量的新发乳腺癌患者血标本,通过对样品资料的整理,发明人从中选择了25例符合下列标准的样本,同时选择10例年龄在25-55岁健康女性作对照进行全外显子芯片检测,样本选择标准如下:

[0052] 1、经病理学明确诊断的乳腺癌病例,其中有3例病人具有癌症家族史并分别标记为X1、X2、X3;

[0053] 2、采血前未接受过放疗或化疗、无既往肿瘤病史;

[0054] 3、与病例年龄匹配的健康女性对照

[0055] 并系统采集了这些样本的人口学资料和临床资料等情况。

[0056] 实施例2外周血DNA的提取和纯化

[0057] 在上述符合条件的25例乳腺癌患者和10例健康女性对照中,两组年龄均衡可比。

[0058] 具体步骤为:

[0059] 1、向储存于2mL冻存管中的外周血加入溶血试剂(即裂解液,40份量配置方法如下:蔗糖219.72g、氯化镁2.02g和曲拉通X-100(amresco0694) 20mL混合后,用TrisHCl溶液定容至2000mL,下同),颠倒混匀后完全转入。

[0060] 2、去除红细胞:用溶血试剂将5mL离心管补至4mL,颠倒混匀,4000rpm离心10分钟,弃上清。向沉淀中加入4mL溶血试剂,再次颠倒混匀清洗一次,4000rpm离心10分钟,弃上清。

[0061] 3、抽提DNA:向沉淀中加1mL抽提液(每300mL中含有122.5mL0.2M氯化钠,14.4mL 0.5M乙二胺四乙酸,15mL10%十二烷基硫酸钠,148.1mL双蒸水,下同)和8μL蛋白酶K,振荡器上充分振荡混匀,37℃水浴过夜。

[0062] 4、去除蛋白质:加1mL饱和酚充分混匀(手轻摇15分钟),4000rpm离心10分钟,取上清转入新的5mL离心管中。在上清液中加入等体积氯仿与异戊醇混合液(氯仿:异戊醇=24:1,v/v,下同),充分混匀后(手摇15分钟),4000rpm离心10分钟,取上清(分入两个1.5mL的离心管)。

[0063] 5、DNA沉淀:在上清液中加入3M的醋酸钠60μL,再加入与上清液等体积的冰无水乙醇,上下轻摇,可见白色絮状沉淀物,再以12000rpm离心10min。

[0064] 6、DNA洗涤:在沉淀中加入冰无水乙醇1mL,12000rpm离心10min,弃上清后真空抽干或置于清洁干燥环境中蒸干。

[0065] 7、测量浓度:通常能得到20-50ng/μLDNA,纯度(紫外2600D:2800D)在1.8-2.0。

[0066] 实施例3SNP的全外显子组检测

[0067] 将实施例2中两组人群经全外显子芯片检测获得相关结果。

[0068] 1、文库构建

[0069] 北京诺禾致源科技股份有限公司采用Agilent的液相芯片捕获系统,对人的全外显子区域DNA进行高效富集,然后在Illumina HiSeq平台上进行高通量、高深度测序。建库和捕获实验采用Agilent SureSelect Human All ExonV5试剂盒,严格使用说明书推荐的试剂和耗材,并参照最新的经过优化的实验流程进行操作。

[0070] 实验基本流程:将基因组DNA经Covaris破碎仪随机打断成长度为180-280bp的片

段,经末端修复和加A尾后在片段两端分别连接上接头制备DNA文库。带有特异index的文库pooling后与多达543,872个生物素标记的探针进行液相杂交,再使用带链霉素的磁珠将20,965个基因的334,378个外显子捕获下来,经PCR线性扩增后进行文库质检,合格即可进行上机测序。

[0071] 2、库检

[0072] 文库构建完成后,先使用Qubit2.0进行初步定量,稀释文库至1ng/ $\mu$ L,随后使用Agilent 2100对文库的insert size进行检测,insert size符合预期后,使用Q-PCR方法对文库的有效浓度进行准确定量(文库有效浓度 $>2$ nM),以保证文库质量。

[0073] 3、上机测序

[0074] 库检合格,根据文库的有效浓度及数据产出需求进行Illumina HiSeq平台测序。

[0075] 4、数据分析与处理

[0076] 经过数据筛选、深度加工和生物信息学序列比对,最终确定“乳腺癌病例”组和“健康女性对照”组中发现的基因型分布频率有显著差异的53个SNP位点为优选敏感级位点。其中,位于基因C12orf45核苷酸序列自5'端起第8187位碱基为G的SNP位点,所述位点为C12orf45:NM\_152318:exon4:c.C368G,该位点变异对蛋白影响值如下:

[0077] ljb23\_sift:0,1.00,D

[0078] ljb23\_pp2hvar:0.951,D

[0079] ljb23\_pp2hdiv:0.999,D

[0080] ljb23\_mt:0.996,0.004,N。

[0081] 该位点经过生物信息学分析,可以确认为乳腺癌候选标志物。

[0082] 实施例4利用危险度评分方法进一步分析SNP与乳腺癌的发病风险

[0083] 本发明人通过对2组样品(“乳腺癌病例组”和“健康女性对照组”)基因型分布频率的比较,选择阳性关联的SNP,以全外显子扫描样本中单个SNP回归系数为权重,进一步求得危险分值,绘制ROC来评价诊断的灵敏性和特异性,进而诊断这些SNP对乳腺癌发病的判断能力。对所有SNP标志物的联合分析发现,位于基因C12orf45:NM\_152318:exon4第368位碱基为G/C的突变,其灵敏度和特异度都达到60%以上。

[0084] 因此,本发明人证明了该位点标志物能够很好地将健康女性对照和乳腺癌患者区分。

[0085] 实施例5单个SNP的基因分型

[0086] 1、取5例乳腺癌患者和5例健康女性对照DNA样本同实施例2;

[0087] 2、PCR扩增

[0088] 利用Primer Premier 5软件对C12orf45:NM\_152318:exon4:c.C368G设计单个SNP的特异性扩增引物如表1所示。

[0089] 表1引物序列

[0090]

Gene	编号	序列	扩增长度
C12orf45	SEQ ID NO.3	TTGCTCATCAACTCCCAACCTAA	392bp
	SEQ ID NO.4	CTTCCACCTTTTCCGTTCTAACT	

[0091] PCR反应体系如表2所示。PCR扩增程序为：95℃预变性10min，94℃变性15s，61℃退火15s，72℃延伸30s，进行30个循环，最后72℃延伸30min，于4℃保存，过夜需放置-20℃冷冻。

[0092] 表2反应体系

[0093]

组分	加入量
2×mix	25μL
上游引物(10uM)	3.0μL
下游引物(10uM)	3.0μL
模板	5μL
加入灭菌蒸馏水	至50μL

[0094] 3、测序

[0095] PCR扩增结束后，取5μL扩增产物，1%琼脂糖凝胶电泳，电泳30min，染色20min，然后将凝胶块置于凝胶成像仪中观察，根据比对Marker的片段大小情况，初步判断扩增片段是否正确。进而对符合要求的扩增产物进行纯化：采用Mag-BindOligonucleotidePurificationKit试剂盒，并按试剂盒要求进行操作。上样测序：采用ABI公司BigDye3.1SequencingKit试剂盒，并按试剂盒要求进行操作；用ABI公司3730型测序仪进行测序。

[0096] 4、结果分析

[0097] 通过Chromas序列分析软件，将测序结果与标准序列进行比对，寻找SNP位点，通过分析SNP位点处碱基的类型，就可以得到SNP位点的基因型。结果显示：5例乳腺癌患者测序得到392bp的片段的核苷酸序列如SEQ ID NO.1所示，其在第239位碱基为CG、GG；而5例健康女性对照测序得到392bp的片段的核苷酸序列如SEQ ID NO.2所示，其在第239位碱基为CC；证实了该位点为CG、GG基因型时判断为乳腺癌的易感基因型，该位点为CC基因型时判断为乳腺癌的非易感基因型，从而进一步确认所述C12orf45:NM\_152318:exon4:c.C368G的SNP位点可用于乳腺癌的检测、治疗、诊断、预后评估等辅助诊断。

[0098] 实施例6用于乳腺癌辅助诊断SNP试剂盒的制作

[0099] 基于实施例5得到的引物组，组装本发明所述的用于乳腺癌的试剂盒，所述试剂盒包括特异扩增如SEQ ID NO.1所示的核苷酸序列的引物对如SEQ ID NO:3和SEQ ID NO:4所示。所述试剂盒还可以有相应PCR技术所需的常用试剂，如：dNTPs，MgCl<sub>2</sub>，双蒸水，Taq酶等，这些常用试剂都是本领域技术人员熟知的，另外还可以有标准品和对照（如确定基因型的标准品和空白对照等）。此试剂盒的价值在于只需要外周血而不需要其他组织样品，通过最精简和特异的引物对检测SNP，再通过SNP谱辅助判断乳腺癌，不仅稳定，检测方便，且精确，大大提高疾病诊断的敏感性和特异性，因此将此试剂盒投入实践，可以帮助指导诊断和更有效的个体化治疗。

[0100] 虽然，上文中已经用一般性说明及具体实施方案对本发明作了详尽的描述，但在本发明基础上，可以对之作一些修改或改进，这对本领域技术人员而言是显而易见的。因此，在不偏离本发明精神的基础上所做的这些修改或改进，均属于本发明要求保护的范围。

## 序列表

<110>	深圳市第二人民医院	
<120>	一种与乳腺癌相关的 SNP 标记	
<130>	P1690	
<160>	4	
<170>	PatentIn version 3.5	
<210>	1	
<211>	392	
<212>	DNA	
<213>	人工序列	
<400>	1	
	ttgctcatca actccaacc taagtccaga aagacctcca ctcttcaaac agttcggata	60
	gagaggagtc ccttattgga ccaggtacag acattttctcc cacagatggc acgggcaaat	120
[0001]	gaaaagctaa gaaaagaaat ggcagctgca ccacctggtc gtttcaatat tgaaaacatt	180
	gatgggectc atagtaaagt tatacaaatg gatgtggctt tgtttgagat gaatcagtgg	240
	gattcaaaag aagtggacag ttcagaagag agttcacaag acagttcaga gaacagtcca	300
	gaatcagaag acgaagatga cagcatccca tetgaagtca ccatagataa cattaagctt	360
	cccaattctg aaggtggaaa aggcaagatt ga	392
<210>	2	
<211>	392	
<212>	DNA	
<213>	人工序列	
<400>	2	
	ttgctcatca actccaacc taagtccaga aagacctcca ctcttcaaac agttcggata	60
	gagaggagtc ccttattgga ccaggtacag acattttctcc cacagatggc acgggcaaat	120
	gaaaagctaa gaaaagaaat ggcagctgca ccacctggtc gtttcaatat tgaaaacatt	180

	gatgggcctc atagtaaagt tatacaaatg gatgtggcctt tgtttgagat gaatcagtcg	240
	gattcaaaag aagtgacag ttcagaagag agttcacaag acagttcaga gaacagttca	300
	gaatcagaag acgaagatga cagcatccca tctgaagtca ccatagataa cattaagctt	360
	cccaattctg aaggtggaaa aggcaagatt ga	392
	<210> 3	
	<211> 23	
	<212> DNA	
[0002]	<213> 人工序列	
	<400> 3	
	ttgctcatca actccaacc taa	23
	<210> 4	
	<211> 23	
	<212> DNA	
	<213> 人工序列	
	<400> 4	
	cttcacctt ttccgttcta act	23