(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2018 年 3 月 22 日 (22.03.2018)

WIPO | PCT

(10) 国际公布号
WO 2018/049947 A1

(51) 国际专利分类号:
C12Q 1/68 (2006.01) *C12N 15/11* (2006.01)

(21) 国际申请号: PCT/CN2017/096249

(22) 国际申请日: 2017 年 8 月 7 日 (07.08.2017)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201610831390.6 2016年9月19日 (19.09.2016) CN

(71) 申请人: 深圳华大基因研究院(BGI SHENZHEN) [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。

(72) 发明人: 贾慧珏(JIA, Huijue); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。 钟焕姿(ZHONG, Huanzi); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。 宋晓蕾(SONG, Xiaolei); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。 王子榕(WANG, Zirong); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。 陈晨(CHEN, Chen); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。

(74) 代理人: 深圳鼎合诚知识产权代理有限公司(DHC IP ATTORNEYS); 中国广东省深圳福田区金田路与福华路交汇处现代商务大厦2201, Guangdong 518048 (CN)。

(81) 指定国(除另有指明，要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,

(54) **Title:** BIOMARKER COMPOSITION FOR DETECTION OF ENDOMETRIOSIS AND APPLICATION THEREOF

(54) **发明名称:** 用于子宫内膜异位症检测的生物标志物组合及应用

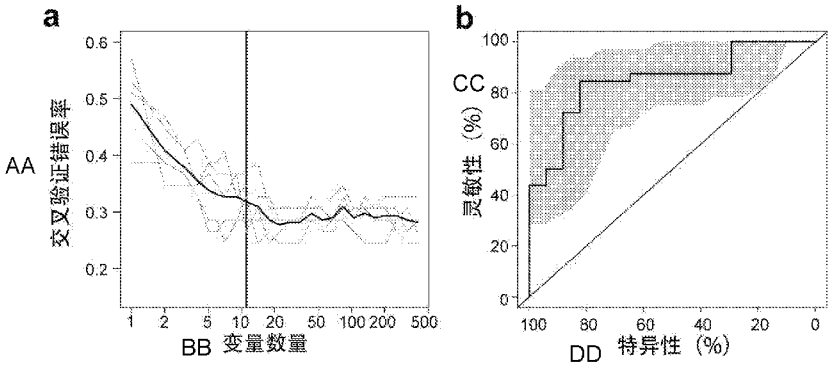


图 3

AA Cross-validation error rate
BB Number of variables
CC Sensitivity (%)
DD Specificity (%)

(57) **Abstract:** A biomarker composition for detection or risk assessment of endometriosis, and application thereof. The biomarker composition comprises at least one of 24 nucleic acids. The 24 nucleic acids have sequences represented by Seq ID No. 1 to Seq ID No. 24 respectively, or sequences having similarity of 97% or more to those represented by Seq ID No. 1 to Seq ID No. 24. The biomarker composition has excellent specificity, and can be used for early diagnosis of endometriosis, thereby providing a new approach for detection or risk assessment of endometriosis. The biomarker composition has high sensitivity and specificity, good reproducibility, and important application value. By using a genital tract specimen as a biomarker detection specimen, the material is easily available, the operation steps are simple, and continuous in-vitro detection can be implemented.

WO 2018/049947 A1

[见续页]

SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国(除另有指明，要求每一种可提供的地区保护)：ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

— 包括说明书序列表部分(细则5.2(a))。

(57) 摘要：用于子宫内膜异位症检测或患病风险评估的生物标志物组合及应用。生物标志物组合，包括二十四条核酸中的至少一条，二十四条核酸分别为Seq ID No.1至Seq ID No.24所示序列，或分别为与Seq ID No.1至Seq ID No.24所示序列有97%以上相似性的序列。生物标志物组合，具有很好的特异性，能用于子宫内膜异位症早期诊断，为子宫内膜异位症检测或风险评估提供了一条新途径。生物标志物组合，具有高灵敏性、高特异性的优点，重复性好，具有重要应用价值；以生殖道样品作为生物标志物检测样本具有取材方便、操作步骤简单和可连续体外检测等优点。

用于子宫内膜异位症检测的生物标志物组合及应用

技术领域

本申请涉及生物标志物领域，特别是涉及一种用于子宫内膜异位症检测或患病风险评估的生物标志物组合及其应用。

背景技术

子宫内膜异位症是育龄妇女的多发病，引发多种临床症状。近年发病率呈明显上升趋势，已达10%-15%，占妇科手术的30%以上。子宫内膜异位症与不孕症的关系非常密切，约40%-50%的不孕患者合并子宫内膜异位症导致的不孕。目前研究认为，微生物感染在子宫内膜异位症导致的不孕中起到非常重要的作用。此外，诸多研究显示，子宫内膜异位症腹腔液中的炎症因子，如肿瘤坏死因子（TNF-α）、白细胞介素6（IL-6）等，可通过干扰生殖过程中某些环节导致不孕。尽管子宫内膜异位症是一种良性疾病，但其病变广泛、形态多样，极具侵袭和复发性，对脏器结构和功能的破坏与恶性肿瘤相当，故有“良性癌”之称。子宫内膜异位症的临床表现与疾病程度不成正比，因此缺乏有效的临床诊断方法。

对于子宫内膜异位症，尽管目前有部分应用于临床的标志物，如癌胚抗原125（CA125）、胎盘蛋白14（PP14）、子宫内膜抗体和白细胞介素6（IL-6）等，它们对于临床诊断和跟踪复发等方面具有一定的参考价值，但仍缺乏特异性。

因此，寻找敏感、特异的子宫内膜异位症的生物标志物是目前急需解决的问题。

发明内容

本申请的目的是提供一种用于子宫内膜异位症检测的生物标志物及其制备方法和应用。

为了实现上述目的，本申请采用了以下技术方案：

本申请的一方面公开了一种用于子宫内膜异位症检测的生物标志物，生物标志物组合包括二十四条核酸中的至少一条，二十四条核酸分别为Seq ID No.1至Seq ID No.24所示序列，或者分别为与Seq ID No.1至Seq ID No.24所示序列具有97%以上相似性的序列。

需要说明的是，本申请的二十四条核酸是经过研究得出的，和子宫内膜异位症有关联的核酸序列，其中每条核酸序列都与子宫内膜异位症有关联性，因

此，在不考虑判断准确性的情况下或者对此要求较低的情况下，可以单独或者组合用于子宫内膜异位症检测或者患病风险评估。但是，本申请的一种优选方案中，不仅二十四条核酸一起使用，而且，还将二十四条核酸按照特定的规律进行分类，分成多个标志物组，各个标志物组一起用于子宫内膜异位症检测或者患病风险评估，这将在后面的优选技术方案中详细描述。

还需要说明的是，本申请的二十四条核酸是根据 97%以上相似性进行聚类分析，然后从每个分类单元（缩写 OTU）中选取最具代表性的序列作为种子序列，其中与子宫内膜异位症具有关联性的二十四个种子序列，即组成本申请的生物标志物组合；因此，本申请的生物标志物组合中，二十四条核酸不仅限于 Seq ID No.1 至 Seq ID No.24 所示序列，还可以是与 Seq ID No.1 至 Seq ID No.24 所示序列具有 97%以上相似性的序列。

需要补充说明的是，本申请的用于子宫内膜异位症检测或患病风险评估的生物标志物组合，并不是直接根据检测生物标志物组合的有或者无进行子宫内膜异位症检测或患病风险评估的，而是，在检测到生物标志物组合后，通过分析其相对丰度，并将相对丰度带入随机森林模型进行判断，根据随机森林模型输出的概率判断待测对象是否患有子宫内膜异位症或评估待测对象患子宫内膜异位症的风险，这将在后面的技术方案中详细说明。

优选的，本申请的另一面公开了一种用于子宫内膜异位症检测或患病风险评估的生物标志物组合，包括第一标志物组、第二标志物组和第三标志物组中的至少一组；第一标志物组由十四条核酸组成，十四条核酸分别为 Seq ID No.1、Seq ID No.2、Seq ID No.6、Seq ID No.7、Seq ID No.12、Seq ID No.13、Seq ID No.15、Seq ID No.17 至 Seq ID No.22、Seq ID No.24 所示序列，或者分别为与 Seq ID No.1、Seq ID No.2、Seq ID No.6、Seq ID No.7、Seq ID No.12、Seq ID No.13、Seq ID No.15、Seq ID No.17 至 Seq ID No.22、Seq ID No.24 所示序列具有 97%以上相似性的序列；第二标志物组由两条核酸组成，两条核酸分别为 Seq ID No.1、Seq ID No.7 所示序列，或者分别为与 Seq ID No.1、Seq ID No.7 所示序列具有 97%以上相似性的序列；第三标志物组由十一条核酸组成，十一条核酸分别为 Seq ID No.3 至 Seq ID No.5、Seq ID No.8 至 Seq ID No.12、Seq ID No.14、Seq ID No.16、Seq ID No.23 所示序列，或者分别为与 Seq ID No.3 至 Seq ID No.5、Seq ID No.8 至 Seq ID No.12、Seq ID No.14、Seq ID No.16、Seq ID No.23 所示序列具有 97%以上相似性的序列。

需要说明的是，本申请的优选方案中，将二十四条核酸可重复选择的分为三个标志物组，即第一标志物组、第二标志物组和第三标志物组；通过三个标

志物组的综合判断，可以大大提高本申请的生物标志物组合检测子宫内膜异位症或者评估患病风险的准确性。

优选的，第一标志物组为 CL 标志物组，用于对来自阴道下 1/3 的样品进行子宫内膜异位症检测或患病风险评估。

优选的，第二标志物组为 CU 标志物组，用于对来自阴道后穹窿的样品进行子宫内膜异位症检测或患病风险评估。

优选的，第三标志物组为 CV 标志物组，用于对来自宫颈管的样品进行子宫内膜异位症检测或患病风险评估。

需要说明的是，本申请的生物标志物组合中的二十四条核酸实际上代表的是阴道下 1/3、阴道后穹窿和宫颈管三个部位的 14 种微生物；本申请通过对阴道下 1/3、阴道后穹窿和宫颈管三个部位的 14 种微生物的二十四条核酸进行检测，并对其相对丰度与子宫内膜异位症的关系进行统计分析，建立随机森林模型，以此判断待测对象是否患有子宫内膜异位症或是否具有患子宫内膜异位症的风险。因此，三个标志物组，实际上就是分别对应三个采样部位；来自于三个部位的样品，分别对应各自的标志物组，独立进行分析判断。只是，根据三者的结果进行综合判断，能够提高本申请的生物标志物组合检测子宫内膜异位症或者评估患病风险的准确性。

还需要说明的是，在阴道下 1/3、阴道后穹窿和宫颈管这三个部位中，其微生物数量远不止 14 个种，14 种微生物的核酸也远不止本申请所记载的 24 个；但是，本申请根据随机森林模型从中筛选出 14 种微生物的二十四条核酸，以作为子宫内膜异位症检测的生物标志物，为子宫内膜异位症的检测和评估提供了一条新的途径。

需要补充说明的是，三个标志物组中，CL 标志物组即阴道下 1/3 样品的标志物组，阴道下 1/3 缩写为 CL；CU 标志物组即阴道后穹窿样品的标志物组，阴道后穹窿缩写为 CU；CV 标志物组即宫颈管样品的标志物组，宫颈管缩写为 CV。

本申请的另一面公开了一种用于子宫内膜异位症检测或患病风险评估的试剂盒，该剂盒中包含用于检测本申请的生物标志物组合的引物对，引物对的正向引物为 SEQ ID No.25 所示序列，反向引物为 SEQ ID No.26 所示序列。

需要说明的是，本申请的生物标志物组合，可以作为一个标准参考存在于试剂盒中，而引物对则是直接用于 PCR 扩增待测样品中的生物标志物组合的。

本申请的另一面公开了本申请的生物标志物组合在子宫内膜异位症药物筛选或者在制备子宫内膜异位症检测或患病风险评估的试剂盒或检测工具中的应

用。

可以理解，本申请的生物标志物组合本身就是针对子宫内膜异位症而研究的，当然可以用于子宫内膜异位症的检测或风险评估；而本申请的生物标志物组合也可以整合到一些专门用于子宫内膜异位症检测的试剂盒或工具中，以方便子宫内膜异位症的检测和评估，只要采用了本申请的生物标志物组合，都在本申请的保护范围内。与此同时，由于本申请的生物标志物组合可以检测子宫内膜异位症或者对子宫内膜异位症进行患病风险评估；当然，可以对比检测用药前和用药后的子宫内膜异位症患病情况或者患病风险变化，从而判断所用药物是否有效，以达到药物筛选的目的。

本申请的再一面公开了一种子宫内膜异位症的检测方法，包括以下步骤，

(1)对待测对象进行样品采集，检测所采集的样品中本申请的生物标志物组合，并分析生物标志物组合中各核酸的水平；

(2)将步骤(1)测得的各核酸的水平与参考数据集或参考值进行比较，获得检测结果；

优选的，各核酸的水平为各核酸的相对丰度；参考数据集或参考值为来源于子宫内膜异位症人群和非子宫内膜异位症人群对照的生物标志物组合中各核酸的水平。

更优选的，步骤(2)中的参考数据集或参考值为表 5、表 6 或表 7 中的至少一组；将各核酸的水平与参考数据集或参考值进行比较获得检测结果，具体包括，利用多元统计模型计算得出患病概率，优选的，多元统计模型为随机森林模型。

更优选的，步骤(1)中对待测对象进行样品采集，包括采集待测对象阴道下 1/3 样品、阴道后穹窿样品和宫颈管样品。

需要说明的是，本申请的生物标志物组合中的各核酸片段都是与子宫内膜异位症相关联的核酸片段，因此，通过分析待测对象中生物标志物组合的水平，即相对丰度，就可以检测待测对象是否患病或判断其患病风险。

本申请的再一面公开了一种通过检测生物标志物判断子宫内膜异位症的方法在制备子宫内膜异位症检测或患病风险评估试剂盒或工具中的应用；其中，生物标志物为本申请的生物标志物组合；

通过检测生物标志物判断子宫内膜异位症的方法包括以下步骤，

(1)对待测对象进行样品采集，检测所采集的样品中本申请的生物标志物组合，并分析生物标志物组合中各核酸的水平；

(2)将步骤(1)测得的各核酸的水平与参考数据集或参考值进行比较，获得检测结果；

优选的，各核酸的水平为各核酸的相对丰度；参考数据集或参考值为来源于子宫内膜异位症患者和非子宫内膜异位症对照的生物标志物组合中各核酸的水平。

更优选的，步骤(2)中的参考数据集或参考值为表 5、表 6 或表 7 中的至少一组；将各核酸的水平与参考数据集或参考值进行比较获得检测结果，具体包括，利用多元统计模型计算得出患病概率，优选地，多元统计模型为随机森林模型。

本申请的再一面公开了一种筛选治疗子宫内膜异位症的候选药物的方法，包括以下步骤，

1)分别测定用药前和用药后的样品中本申请的生物标志物组合，并分析生物标志物组合中各核酸的水平；

2)根据比较用药前和用药后的样品中各核酸的水平，判断候选药物；

步骤 2)中，比较用药前和用药后的样品中各核酸的水平，具体包括，利用多元统计模型计算得出患病概率，优选地，多元统计模型为随机森林模型。

本申请的再一面公开了一种女性生殖道内微生物群的检测方法，包括以下步骤，

(1)采集待测对象生殖道内微生物样品，检测所采集的样品中本申请的生物标志物组合，并分析生物标志物组合中各核酸的水平；

(2)将步骤(1)测得的各核酸的水平与参考数据集或参考值进行比较，获得检测结果；

优选的，各核酸的水平为各核酸的相对丰度；参考数据集或参考值为来源于子宫内膜异位症人群和非子宫内膜异位症人群对照的生物标志物组合中各核酸的水平。

优选的，步骤(2)中的参考数据集或参考值为表 5、表 6 或表 7 中的至少一组；将各核酸的水平与参考数据集或参考值进行比较获得检测结果，具体包括，利用多元统计模型计算得出患病概率，更优选地，多元统计模型为随机森林模型。

优选的，步骤(1)中采集待测对象生殖道内微生物样品，具体包括采集待测对象阴道下 1/3 样品、阴道后穹窿样品和宫颈管样品。其中，生殖道内微生物样品的采集可以采用常规的尼龙绒屑拭子，在此不做具体限定。

需要说明的是，本申请的生物标志物组合，实际上就是根据女性生殖道内微生物群 DNA 与子宫内膜异位症之间的关系而得出的，即本申请的生物标志物实际上就是女性生殖道内能够体现子宫内膜异位症状态的微生物 OTU；因此，本申请提出了一种女性生殖道内微生物群检测方法，通过微生物群的检测为子宫内膜异位症或其患病风险提供判断和评估依据。

本申请的再一面公开了一种制备子宫内膜异位症生物标志物组合的方法，包括以下步骤，

(1)分别对子宫内膜异位症病患和非病患进行生殖道内微生物样品采集，对所有采集的样品分别进行16S测序；

(2)将16S测序结果进行聚类分析，获得OTU单元以及每个OTU的种子序列，并计算每个OTU单元的相对丰度；

(3)利用随机森林模型对每个 OTU 单元的相对丰度与子宫内膜异位症状态进行拟合，并进行5次十折交叉验证，得到最优的OTU组合，最优OTU组合中各OTU的种子序列，即组成子宫内膜异位症的生物标志物组合。

优选的，步骤(1)中，生殖道内微生物样品采集，具体包括采集待测对象阴道下1/3样品、阴道后穹窿样品和宫颈管样品。

需要说明的是，本申请子宫内膜异位症生物标志物组合的制备方法，其关键在于利用随机森林模型对生殖道内微生物群DNA与子宫内膜异位症的关联进行拟合、验证，最终得到能够对子宫内膜异位症患病或风险进行评估的生物标志物组合。可以理解，本申请的制备方法或其基本思路，不只限于制备子宫内膜异位症的生物标志物组合；还可以用于制备类似的与生殖道内微生物群DNA存在关联的病症的生物标志物组合，例如子宫腺肌症的生物标志物组合。

由于采用以上技术方案，本申请的有益效果在于：

本申请的用于子宫内膜异位症检测的生物标志物，具有很好的特异性，能够很好的检测出子宫内膜异位症，为子宫内膜异位症的检测或风险评估提供了一条新的途径，并且，能够用于子宫内膜异位症的早期诊断。

本申请的其它主要优点包括：

(a)本申请的生物标志物组合用于子宫内膜异位症的检测或患病风险评估，具有高灵敏性、高特异性的优点，具有重要的应用价值。

(b)生殖道样品作为生物标志物组合检测样本具有取材方便、操作步骤简单和可连续体外检测等优点。

(c)本申请的生物标志物组合用于子宫内膜异位症的检测或患病风险评估具有重复性好的特点。

附图说明

图1是本申请实施例中基于阴道下1/3处CL的标志物组鉴别子宫内膜异位症引发不孕的结果图，图中，a为随着OTU数量的增加，对随机森林鉴别子宫内膜异位症（不孕）进行5次10折交叉验证的错误率分布情况，b为经过交叉

验证过的组合的接收者操作曲线（缩写 ROC 曲线），曲线下面积（缩写 AUC）为 0.8272，阴影面积代表 95%置信区间，对角线代表 AUC 为 0.5 的曲线；

图 2 是本申请实施例中基于阴道后穹窿CU的标志物组鉴别子宫内膜异位症（不孕）的结果图，图中，a 为随着 OTU 数量的增加，对随机森林鉴别子宫内膜异位症（不孕）进行 5 次 10 折交叉验证的错误率分布情况，b 为经过交叉验证过的组合的接收者操作曲线（缩写 ROC 曲线），曲线下面积（缩写 AUC）为 0.5919，阴影面积代表 95%置信区间，对角线代表 AUC 为 0.5 的曲线；

图 3 是本申请实施例中基于宫颈管CV标志物组鉴别子宫内膜异位症(不孕)的结果图，图中，a 为随着 OTU 数量的增加，对随机森林鉴别子宫内膜异位症（不孕）进行 5 次 10 折交叉验证的错误率分布情况，b 为经过交叉验证过的组合的接收者操作曲线，曲线下面积为 0.8493，阴影面积代表 95%置信区间，对角线代表 AUC 为 0.5 的曲线；

图 4 是本申请实施例中阴道下 1/3 处 CL 标志物组在第二群体中对子宫内膜异位症（不孕）进行鉴别的 ROC 曲线；

图 5 是本申请实施例中阴道后穹窿CU标志物组在第二群体中对子宫内膜异位症（不孕）进行鉴别的 ROC 曲线；

图 6 是本申请实施例中宫颈管CV标志物组在第二群体中对子宫内膜异位症（不孕）进行鉴别的 ROC 曲线；

图中，变量数量是指 OTU 数量，其中，灵敏性=真阳性/(真阳性+假阴性)；特异度=真阴性/(真阴性+假阳性)。

具体实施方式

本申请的生物标志物，是根据采集对象三个部位的微生物群 DNA 与子宫内膜异位症之间的关系而得出的，本申请的生物标志物实际上就是这三个部位的能够体现子宫内膜异位症状态的微生物 OTU。具体的，本申请的一种制备方法中，这种对应关系或者生物标志物的获得，是以 OTU 种子序列的相对丰度为一个对象，子宫内膜异位症状态（患病或非患病）为第二个对象，通过随机森林模型对两者进行拟合，最终通过 5 次十折交叉验证而得出的。本申请通过严格的计算和试验研究，最终获得了三个部位的 14 种微生物的二十四条核酸作为本申请的生物标志物。

本申请的一种实现方式中，三个部位的标志物组可以独立的对子宫内膜异位症患病或风险进行评估，但是，结合三个部位的概率，判断待测对象是否患有子宫内膜异位症或是否具有患子宫内膜异位症的风险，这样准确性会更高。

本申请所用术语是本领域普通技术人员通常理解的含义。为了更好地理解本申请，对一些定义和相关术语的解释如下：

本申请的“子宫内膜异位症”，是一种常见的妇科疾病，定义为子宫内膜组织生长在子宫腔以外引起的病症。最常见的子宫内膜异位症是出现于卵巢及输卵管，也有可能出现于子宫肌层、盆腔腹膜，甚至是膀胱及大肠。其中因卵巢的子宫内膜异位，可形成内有棕色液体的子宫内膜瘤，所以又有“巧克力囊肿”或“朱古力瘤”之称，会影响怀孕。本申请的“子宫内膜异位症”和“子宫内膜异位症（不孕）”特指本领域常见的子宫内膜异位症引发不孕。本申请的“非子宫内膜异位症”特指本领域常见的没有子宫内膜异位症，且可怀孕。

本申请的生物标志物质的水平通过相对丰度指示。

在本申请的一个实施方式中，参考值是指健康对照的参考值或正常值。本领域的技术人员清楚，在样品数量足够多情况下，每个生物标志物的正常值，即绝对值，的范围可以通过检验和计算方法得到。

本申请的“生物标志物”，也称为“生物学标志物”，是指个体的生物状态的可测量指标。这样的生物标记物可以是在个体中的任何物质，只要它们与被检个体的特定生物状态，例如疾病，有关系即可。这样的生物标记物可以是，例如，核酸标志物（例如 DNA）、蛋白质标志物、细胞因子标记物、趋化因子标记物、碳水化合物标志物、抗原标志物、抗体标志物、物种标志物（种/属的标记）和功能标志物（KO/OG 标记）等。本申请的生物标志物具体的为 DNA 核酸标志物。

本申请的“OTU”是指操作分类单元(operational taxonomic units 缩写 OTU)，是在系统发生学研究或群体遗传学研究中，为了便于进行分析，人为给某一个分类单元，如品系、种、属、分组等，设置的同一标志。本申请中按照 97%的相似性阈值将序列划分为一个 OTU，由此使得三个部位的样品分别可以获得多个 OTU，每一个 OTU 被视为一个微生物物种。样品中的微生物多样性和不同微生物的丰度都是基于对 OTU 的分析。

本申请中提到的“个体”指动物，特别是哺乳动物，如灵长类动物，本申请的实施例中所指为人。

下面通过具体实施例和附图对本申请作进一步详细说明。以下实施例仅对本申请进行进一步说明，不应理解为对本申请的限制。

实施例

1.材料与方法

1.1 样品收集

本例的样品采集由深圳北大医院妇产科医生协助进行。排除炎症病例，研究对象均为非经期、非妊娠期、非哺乳期女性，无内分泌和自身免疫性疾病，肝肾功能正常。取样前一段时间没有使用激素及抗生素，没有进行阴道用药、阴道灌洗及宫颈治疗，取样前 48 小时内没有进行性生活。根据以上标准，本例筛选出 49 例育龄女性，作为第一群体。所有符合以上标准的个体都进行详细的表型信息登记，以了解其病史、家族史、用药史及生活习惯等，并且均签署了知情同意书。

下生殖道采样是在个体入院后，不经过消毒处理，排空小便后，在妇科检查床采集阴道下 1/3（缩写 CL）、阴道后穹窿（缩写 CU）、宫颈管（缩写 CV）三个部位的分泌物样品。具体的，49 个采集对象的样品编号及采样信息为，编号 C026、C028、C033、C035、C038、C041、C045、C048、C050、C055、C056、C058、C059、C062、C063、C064、C065 的十七个采集对象为非子宫内膜异位症患者，十七个采集对象都采集了 CL、CU 和 CV 三个部位的样品；编号 T022、T024、T027、T028、T032、T033、T036、T039、T041、T042、T045、T053、T056、T058、T059、T061、T062、T063、T067、T069、T070、T076、T078、T084、T085、T086、T087、T088、T090、T092、T094、T095 的三十二个采集对象为子宫内膜异位症患者，三十二个采集对象都采集了 CL、CU 和 CV 三个部位的样品。

样品采集是利用尼龙绒屑拭子进行样品收集，尼龙绒屑拭子购自晨阳全球集团 CY-93050 和 CY-98000 两种型号。取样后将拭子头用液氮进行速冻，并保存于-80℃，用干冰运送至深圳华大基因研究院进行后续的试验。

1.2 DNA 提取与 16S 测序

本例利用 QIAamp DNA Mini Kit 试剂盒（购自 QIAGEN）进行 DNA 提取。具体提取步骤参照生产厂商提供的说明书进行。采用 16S rRNA 基因 V4-V5 高变区特异引物进行扩增，两条引物分别为 V4-515F 和 V5-907R，V4-515F 为 Seq ID No.25 所示序列，V5-907R 为 Seq ID No.26 所示序列。

Seq ID No.25：5’-GTGCCAGCMGCCGCGGTAA-3’

Seq ID No.26：5’-CCGTCAATTCMTTTRAGT-3’

PCR 程序如下，94℃变性 3min；然后进入 25 个循环：94℃变性 45s，50℃退火 60s，72℃延伸 90s；循环结束后，72℃延伸 10min。得到的 PCR 产物利用 AMPure Beads (Axygen)进行纯化，测序采用芯片泳道测序的方法，将多个样品混合后测序。所以文库构建需要在各样品的引物序列外端连接 10bp 的 barcode 序列后，添加接头序列。通过对每个样品添加不同的 barcode 序列，即样本识别

序列，区分不同样本。文库构建完成后，通过 Ion torrent PGM 测序平台，进行 V5-V4 反向测序，以上文库构建和测序等由深圳华大基因进行。

1.3 16S 测序数据处理

利用 Mothur 软件（V1.33.3）从 PGM 系统中提取原始数据并进行预处理，高质量序列的标准包括：1）长度大于 200bp；2）与简并 PCR 错配碱基少于 2 个；3）平均质量分数大于 25。基于 16S rRNA 基因序列，利用 QIIME 的 uclust 方法对 OTU 进行聚类，相似阈值设置为 97%。选取每个 OTU 的种子序列（Seed sequence），利用 Greengene 数据库中的参照基因信息 gg_13_8_otus 进行注释。计算每个样本中每一个 OTU 的相对丰度，其中某一 OTU 的相对丰度为某个样本中该 OTU 的丰度与该样本中所有 OTU 丰度之和的比值。

1.4 不同位点样品间微生物群一致性分析

基于 OTU 的存在或缺失，本例利用 Sorenson 指数（Sørensen - Dice 指数）来测量同一个体不同位点样品微生物群的相似性，计算方法如下：

$$\sum QS = \frac{2C}{A+B} = \frac{2|A \cap B|}{|A| + |B|}$$

其中 A 和 B 分别代表样品 A 和 B 中 OTU 的数目，C 代表两个样品中共有的 OTU 数目。QS 是相似性指数，取值范围为 0~1。本例中分别计算了 CL 和 CU 的相似性指数，CL 和 CV 的相似性指数，以及 CU 和 CV 的相似性指数。相似性指数约接近 1，表示两个采样部位的微生物群的相似性越高。

1.5 随机森林分类器

为了建立一个能够鉴别异常状态样品的模型，对于每个采样部位，利用 R 软件（3.1.2 RC）中 randomForest 工具包对每个样品的 OTU 的相对丰度与子宫内膜异位症状态进行拟合，采用默认参数；其中，每个样品的 OTU 是至少存在于 10%的样品中的 OTU，也就是说，剔除在各个部位的所有待测样品中只在有不到 10%的样品中才能检出的 OTU。之后进行 5 次 10 折交叉验证，将 5 次 10 折交叉验证的误差曲线进行平均，将平均后曲线的最低误差加上该点的标准误差作为可接受误差的域值。在分类误差小于域值的各组 OTU 中，其中，OTU 数目最少的为最优 OTU 组合，作为鉴别子宫内膜异位症的生物标志物组合。

1.6 生物标志物验证

为了验证本例得到的生物标志物，本例另外采用了独立的受试群体，即第二群体进行验证。第二群体中，对于 CL 和 CU，各有 11 位子宫内膜异位症患者和 11 位非子宫内膜异位症个体；对于 CV，有 12 位子宫内膜异位症患者和 10

位非子宫内膜异位症个体。

2．实验结果

2.1 同一个体内上下生殖道微生物结构特征及变化趋势

为了探索生殖道不同区域微生物群之间的关系，本例计算了同一个体的样品之间的距离。相对于下阴道 1/3（CL）样品而言，从阴道后穹窿（CU）、宫颈管（CV）粘液到子宫和腹腔液的加权 UniFrac 距离依次增加。这也再一次指明随着解剖学结构由下至上，女性生殖道的群落结构呈现连续变化性。

同一个体中不同部位的样品呈现高度相关性，不同部位样品之间的 Sorenson 指数与它们的解剖学结构相一致。宫颈（CV）粘液与腹腔液样本具有显著的相关性，平均 Sorenson 指数为 0.255，表明在普通人群中可以通过分析易取得的宫颈粘液样本来评价宫腔和腹腔的健康状况。

此外，本例还分别通过阴道、宫腔底部对宫颈粘液取样，发现两种途径取样取得样品的细菌分布显示出高度的相似性，进一步表明可以通过分析易获得的宫颈管样本来评价宫腔微生物的情况。

2.2 与疾病相关的微生物

为了得到用来鉴别子宫内膜异位症的 OTU 生物标志物，本例建立随机森林模型，具体步骤为：(1) 以 OTU 相对丰度作为输入特征，设计基于第一群体的随机森林模型；(2)对于随机森林模型，设计 10 折交叉验证算法，把第一群体分为子宫内膜异位症个体与非子宫内膜异位症个体两类，并分别得到随机森林模型的 ROC 曲线，以各 ROC 曲线下面积 AUC 值作为评价指标。

本例利用随机森林模型，并结合 10 折交叉验证，得到了各个部位最优的生物标志物，如表 1 所示，用于鉴别子宫内膜异位症。表 2 至表 4 分别为三个部位的标志物组在样品中的富集信息，表 5 至表 7 分别为三个部位的标志物组在第一群体样品的相对丰度信息。本例中，三个部位的生物标志物，鉴别子宫内膜异位症的结果，如图 1 至图 3 所示，图 1 为阴道下 1/3 处（CL）的标志物组鉴别子宫内膜异位症，图 2 为阴道后穹窿（CU）的标志物组鉴别子宫内膜异位症，图 3 为宫颈管（CV）的标志物组鉴别子宫内膜异位症。

表 1　生物标志物及其所属的各个部位

Seq ID No.	OTU 编号	OTU 分类	CL	CU	CV
1	1	Lactobacillus sp.	√	√	--
2	5	Leptotrichiaceae	√	--	--
3	11	Vagococcus sp.	--	--	√
4	12	Delftia sp.	--	--	√

5	27	Dysgonomonas sp.	--	--	√
6	33	Aerococcus sp.	√	--	--
7	35	Prevotella sp.	√	√	--
8	38	Lactobacillus sp.	--	--	√
9	42	Tissierellaceae	--	--	√
10	48	Comamonadaceae	--	--	√
11	54	Erysipelotrichaceae	--	--	√
12	61	Lactobacillus sp.	√	--	√
13	64	Dialister sp.	√	--	--
14	70	Erysipelothrix sp.	--	--	√
15	86	Anaerococcus sp.	√	--	--
16	108	Dysgonomonas sp.	--	--	√
17	221	Lactobacillus sp.	√	--	--
18	233	Lactobacillus sp.	√	--	--
19	344	Lactobacillus sp.	√	--	--
20	424	Lactobacillus iners	√	--	--
21	464	Prevotella sp.	√	--	--
22	520	Lactobacillus iners	√	--	--
23	628	Lactobacillus iners	--	--	√
24	663	Lactobacillus sp.	√	--	--

表 1 中，CL、CU、CV 三个部位的标记物可以单独分别做判断，“√”是表示针对该部位进行判断时所需用到的生物标志物，“--”表示不需要用到的。

在进行样品检测时，要计算各部位的“√”的 OTU 的相对丰度，将相对丰度输入随机森林模型，得到结果，判断是否为子宫内膜异位症。

表 2 CL 中标志物组各 OTU 丰度信息

OTU 编号	P 值	子宫内膜异位症组（相对丰度轶和数均值）	对照组（相对丰度轶和数均值）	富集方向（0：对照组富集；1：子宫内膜异位症富集）
1	0.3283	26.4688	22.2353	1
5	0.0198	22.8750	29.0000	0
33	0.0072	21.9688	30.7059	0
35	0.0181	22.6094	29.5000	0
61	0.1909	26.9531	21.3235	1
64	0.0022	21.2656	32.0294	0
86	0.0709	22.6250	29.4706	0
221	0.4159	26.0781	22.9706	1
233	0.2980	23.6250	27.5882	0
344	0.5252	24.2500	26.4118	0
424	0.5129	25.9375	23.2353	1
464	0.0004	22.0000	30.6471	0
520	0.0110	28.4688	18.4706	1
663	0.1293	23.2813	28.2353	0

表 3 CU 中标志物组各 OTU 丰度信息

OTU 编号	P 值	子宫内膜异位症组（相对丰度轶和数均值）	对照组（相对丰度轶和数均值）	富集方向（0：对照组富集；1：子宫内膜异位症富集）
1	0.2112	26.8750	21.4706	1
35	0.0001	21.5000	31.5882	0

表 4 CV 中标志物组各 OTU 丰度信息

OTU 编号	P 值	子宫内膜异位症组（相对丰度轶和数均值）	对照组（相对丰度轶和数均值）	富集方向（0：对照组富集；1：子宫内膜异位症富集）
11	0.0021	29.5781	16.3824	1
12	0.0895	27.5313	20.2353	1
27	0.0004	30.2500	15.1176	1
38	0.0814	27.5938	20.1176	1
42	0.0053	28.9688	17.5294	1
48	0.0014	29.4063	16.7059	1
54	0.0011	29.7344	16.0882	1
61	0.0349	28.1250	19.1176	1
70	0.0007	29.8594	15.8529	1
108	0.0060	28.9375	17.5882	1
628	0.0181	22.6094	29.5000	0

表 2 至表 4 中，子宫内膜异位症组是指第一群体的 49 个采集对象中患有子宫内膜异位症的样品，对照组是指第一群体的 49 个采集对象中没有患子宫内膜异位症的样品。

表 5 CL 中标志物组各 OTU 在第一群体中的丰度信息

OTU 编号 / 样品编号	1	5	33	35	61	64	86	221	233	344	424	464	520	663
C026CL	0.025755	0	0	0	0.003848	0	0	0	0	0	0.002664	0	0.000888	0.000296
C028CL	3.37E-05	0.352313	0.004986	0.081298	0	0.000708	0.000236	0	0	0	0	0.000101	0	0
C033CL	0	0	0	0.000326	0	0.00013	0.000391	0	0	0	0	0.000326	0	0
C035CL	2.18E-05	0	0	0.000174	0	0.000153	0.005754	0	0	0	0	0	0	0
C038CL	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C041CL	0.379448	0	0.120236	0	0.003264	0.000421	0.002211	0	0.000316	0	0	0	0.000211	0
C045CL	0.782185	0	0.003031	0	0.004744	0.000132	0	0.000132	0.000395	0	0.000264	0	0	0.000132
C048CL	0	0.230799	0.007552	0.097958	6.40E-05	0.00464	0.003648	0	0	0	0	0.000352	0	0
C050CL	0.984574	0	0	0	0.001248	0	0	0.000113	0.000295	2.27E-05	0	0	0	2.27E-05
C055CL	0.928318	0	0	0	0.003861	1.22E-05	0	8.55E-05	0.000403	0.000318	0	0	0	9.78E-05
C056CL	0.993735	0	1.46E-05	0	0.000583	0	0	7.28E-05	0.000291	0.000321	0	0	0	0.000102

C058CL	0	0.27721	0.000354	0.002922	0	0.008905	0.002814	0	0	0	3.08E-05	0.001292	0	0
C059CL	0.014053	0.000562	0	0	0.021922	0	0	0	0	0	0.005621	0	0.001124	0
C062CL	0.000134	0	0	0	4.45E-05	0.000712	0.001335	0	0	0	4.45E-05	8.90E-05	0	0
C063CL	0.003079	0	0	0	0.008621	0	0	0	0	0	0.002771	0	0	0
C064CL	0.006181	0	0	0	0.013736	0	0	0	0	0	0.006181	0	0	0
C065CL	0.983899	0	0.000459	0	0.001613	0	4.74E-05	0.000142	0.000332	0.000269	0	0	0	6.33E-05
T022CL	0.003799	0	0	0	0.003419	0.00038	0.149696	0	0	0	0.00038	0	0.00152	0
T024CL	0.382185	0	0	0	0.005329	0	0.000381	0	0	0	0.000381	0	0.001523	0
T027CL	0.011324	0	0	0	0.014808	0	0	0	0	0	0	0	0.000871	0
T028CL	0.578748	0	0.000759	0	0.021632	0	0	0	0.00038	0	0.001518	0	0.001518	0
T032CL	0.987933	0	0	0	0.003736	0	0	0.000215	0.000376	0	0	0	0	0
T033CL	0	0	0	0	0	0.000395	0	0	0	0	0	0	0	0
T036CL	0.013047	0	0	0	0.005372	0	0	0	0	0	0.002302	0	0.000767	0
T039CL	0.951261	2.29E-05	0	0	0.007837	2.29E-05	0.00016	9.17E-05	0.000183	4.58E-05	0	0	0	4.58E-05
T041CL	0.990375	0	0	0	0.003774	0	0	0.000252	0.000231	4.19E-05	0	0	0	2.10E-05
T042CL	0.978821	0	0	0	0.004269	0	0	0.000356	0.000261	4.74E-05	2.37E-05	0	0	4.74E-05
T045CL	0.539787	0	0	0	0.00256	0	0.000577	0.000144	7.21E-05	0	0	0	0	0
T053CL	0.00275	0	0	0	0.005 5	0	0	0	0	0	0.00275	0	0.005 5	0
T056CL	0.994051	0	0	0	0.000739	0	0	0.000233	3.89E-05	7.78E-05	7.78E-05	0	0	0
T058CL	0.838965	0	0	0	0.002992	0	0.000176	0.000176	0.000176	0.000176	0.000528	0	0.000176	0
T059CL	0.666895	0	0	0	0.014065	0	0	0.000343	0	0	0.000343	0	0.000343	0
T061CL	0.993 94	0	0	0	0.001771	0	0	0.000233	0.00028	6.99E-05	0	0	0	0
T062CL	0.004123	0	0	0	0.003207	0	0	0	0	0	0.001374	0	0.001374	0
T063CL	0.008763	0	0	0	0.009737	0	0	0	0	0	0.001947	0	0.003895	0
T067CL	0.012739	0	0	0	0.003892	0	0	0	0	0	0.001415	0	0.001415	0
T069CL	0.00017	0	0	0	0	0	0	0	0	0	0	0	0	0
T070CL	0.008056	0	0.038268	0	0.008056	0	0.005035	0	0	0	0.001007	0	0.002014	0
T076CL	0.157246	0	0	0	0.003083	0	0	0	0	0	0.000685	0	0.003768	0
T078CL	0.011249	0	0	0	0.00225	0	0	0	0	0	0.001687	0	0.001687	0
T084CL	0.003517	0	0	0.000703	0.006683	0	0.000352	0	0	0	0.002462	0	0.000703	0
T085CL	0.970126	0	0	1.17E-05	0.003122	0	0.000282	0.000176	0.000141	2.35E-05	0	0	1.17E-05	5.87E-05
T086CL	0.000107	0	0	0	0	0	0	0	0	0	0	0	0	0
T087CL	0.586498	0	0.000162	0	0.028238	0	0	0.000974	0	0	0	0	0.000487	0
T088CL	0.969059	0	0	0	0.002888	0	0.000301	0.0001	0.000221	0.00014	2.01E-05	0	4.01E-05	2.01E-05
T090CL	0.968334	4.35E-05	0.001391	0	0.00113	0	2.17E-05	0.000239	0.000283	0	2.17E-05	0	0	0.000152
T092CL	5.05E-05	0	0	0.000303	0	0.003586	0	0	0	0	0	0	0	0
T094CL	0	0	0	0	0	0	0	0	0	0	0	0	0	0
T095CL	0.002795	0	0	0	0.036688	0.003145	0	0	0	0	0.001747	0	0.000349	0

表 6 CU 中标志物组各 OTU 在第一群体中的丰度信息

样品编号 \ OTU编号	1	35	样品编号 \ OTU编号	1	35
C026CU	0.023428	0	T041CU	0.992301	0
C028CU	0	0.113628	T042CU	0.980139	0
C033CU	0	0	T045CU	0.506525	0
C035CU	0	0.03033	T053CU	0.005803	0
C038CU	0	0	T056CU	0.994821	0
C041CU	0.290916	0	T058CU	0.89253	0
C045CU	0.764798	0	T059CU	0.652016	0
C048CU	3.51E-05	0.141157	T061CU	0.994157	0
C050CU	0.992995	3.18E-05	T062CU	0.009134	0
C055CU	0.965383	0	T063CU	0.007898	0
C056CU	0.995162	0	T067CU	0.007309	0
C058CU	2.91E-05	0.004011	T069CU	5.69E-05	0
C059CU	0.012563	0	T070CU	0.003514	0
C062CU	0.000111	0.000111	T076CU	0.0735	0
C063CU	0.00404	0	T078CU	0.005145	0
C064CU	0.007359	0	T084CU	0.005479	0
C065CU	0.986511	2.34E-05	T085CU	0.968011	0
T022CU	0.002283	0	T086CU	0.00049	0
T024CU	0.611603	0	T087CU	0.576768	0
T027CU	0.004378	0	T088CU	0.978815	0
T028CU	0.614412	0	T090CU	0.986075	0
T032CU	0.988763	0	T092CU	0	0
T033CU	0.000216	0	T094CU	0.000173	0
T036CU	0.010967	0	T095CU	0.003645	0
T039CU	0.954426	0			

表 7 CV 中标志物组各 OTU 在第一群体中的丰度信息

样品编号 \ OTU编号	11	12	27	38	42	48	54	61	70	108	628
C026CV	0.007045	0	0.002202	0.018054	0.004844	0	0.003523	0.001761	0.000881	0	0.00044
C028CV	0.000154	3.84E-05	0.000192	0	0	0	3.84E-05	0	0	0.000115	0
C033CV	0.003516	0	0.002041	0	0.000907	0.000454	0.00068	0	0.001134	0.000113	0
C035CV	0.000182	3.64E-05	0.000291	0	7.27E-05	0.000182	0	0	7.27E-05	3.64E-05	0
C038CV	0.00698	0.001454	0.002617	0	0	0	0.0016	0	0.000291	0.000145	0
C041CV	0.001128	0	0	0.007219	0.001128	0	0.000226	0.001353	0	0.000451	0.000451
C045CV	0	0	0	0.010986	0	0	0	0.002104	0	0	0
C048CV	0	0	0	0.000192	0	0	0	0	0	0	3.85E-05
C050CV	0.000118	2.94E-05	5.88E-05	8.82E-05	0	0	0	0.000852	0	2.94E-05	0
C055CV	0.000597	0	0.000163	5.42E-05	8.14E-05	0.000136	5.42E-05	0.002631	0	2.71E-05	0
C056CV	0	0.000429	0	2.86E-05	0	0	0	0.000515	0	0	0
C058CV	0	1.71E-05	0	0	0	0	0	0	0	0	0
C059CV	0	0.020377	0	0.016811	0	0	0	0.001528	0	0	0.000509
C062CVU	0.013808	0.053198	0	0	0	0	0	0	0	0	0

C063CV	0.001947	0.000324	0	0.06392	0	0	0	0.005191	0	0	0.014925
C064CV	0	0.005098	0	0.060323	0	0	0	0.009346	0	0	0.00085
C065CVU	0	0.321645	0	6.11E-05	0	0	0	6.11E-05	0	0	0
T022CV	0.021034	0.000495	0.008414	0.009404	0.005444	0.005939	0.005444	0.003712	0.004207	0.001237	0
T024CV	0.000704	0.001056	0.003167	0.015482	0.002111	0.002111	0.000704	0.003519	0.000704	0.001407	0
T027CV	0.020788	0	0.012035	0.025164	0.001094	0.004376	0.003282	0.003282	0.002188	0	0
T028CV	0.113727	0.020583	0.079076	0.012587	0.028876	0.007848	0.023693	0.002517	0.02236	0.004442	0
T032CV	0.000238	0	0.000132	0.000739	0.000106	2.64E-05	0.000106	0.003406	2.64E-05	0	0
T033CV	0.004934	0.00185	0.00185	0	0.001028	0.002467	0.002878	0	0.000822	0.000411	0
T036CV	0.006143	0	0.006143	0.039312	0.004914	0	0.002457	0.013514	0	0.001229	0
T039CV	0.001154	0.000269	0.000269	7.69E-05	0.000192	0.000269	0.000308	0.006655	7.69E-05	3.85E-05	0
T041CV	0.000336	0	0.000196	8.41E-05	0	0	0	0.003196	8.41E-05	0	0
T042CV	0.000369	0.000164	4.10E-05	0.000451	0	0	0.000164	0.001393	0.000123	0	0
T045CV	0.00127	0.000203	0.000254	0.002845	0.00061	0.00066	0.000305	0.00188	0.000254	0.000152	0
T053CV	0.128435	0.037605	0.017067	0.009257	0.03153	0	0.016199	0.001446	0.00781	0.002025	0
T056CV	7.02E-05	0.00014	0	0.000246	0.00014	0	0	0.001616	0	0	0
T058CV	0.001052	0.00015	0.000301	0.005561	0	0.00015	0.00015	0.004208	0.00015	0	0
T059CV	0.003435	0.004723	0.001288	0.062688	0	0.000859	0.006011	0.027909	0	0.001288	0
T061CV	0.000804	0.000151	0.000301	0.0001	0	0	0.000703	0.000804	0.000151	0	0
T062CV	0.006985	0.016764	0.002235	0.021514	0.003073	0	0.002235	0.001118	0.002794	0.000838	0.000279
T063CV	0.086662	0.001507	0.013564	0.006782	0.021854	0.008289	0.023361	0	0.006029	0.000754	0
T067CV	0	0.011572	0	0.040501	0	0	0	0.007715	0	0	0
T069CV	0.071864	0.032182	0.029683	0	0.01531	0.019684	0.02609	0	0.001562	0.006093	0
T070CV	0.003908	0.000558	0.002792	0.018984	0.001117	0.000558	0	0.001117	0	0.000558	0
T076CV	0.017681	0.013521	0.00624	0.016641	0.00052	0.00156	0.00104	0.00364	0.0026	0.00104	0
T078CV	0.044974	0.027337	0.014991	0.010582	0.020282	0	0.012346	0.001764	0.007055	0.003527	0
T084CV	0.002278	0.001139	0.001519	0.027715	0	0	0	0.001898	0.000759	0.00038	0.00038
T085CV	0.000288	0.000144	8.65E-05	0.000778	2.88E-05	0.000115	0.000202	0.001326	8.65E-05	0	0
T086CV	0.038995	0.007415	0.019086	0	0.007415	0.004531	0.011259	0	0.007003	0.00151	0
T087CV	0.004893	0.002553	0.001276	0.075516	0	0.000638	0.001276	0.011487	0	0.000213	0
T088CV	0.002544	0.001479	0.000355	0.000533	0.000592	0.000178	0.000237	0.001953	0.000178	5.92E-05	0
T090CV	3.45E-05	3.45E-05	0	0.002936	0	0	0	0.000691	3.45E-05	0	0
T092CV	0.002317	0.000172	0.000944	0	0.000772	0.000601	0.000686	0	0.000172	0.000172	0
T094CV	0.00857	0.00048	0.006651	0	0.003154	0.002948	0.002057	0	0.000754	0.000411	0
T095CV	0.001759015	0.000879507	0	0.035180299	0	0	0	0.004397537	0	0	0.000439754

图 1 为阴道下 1/3 处（CL）的标志物组鉴别子宫内膜异位症，图中，a 图为随着 OTU 数量的增加，对随机森林鉴别子宫内膜异位症进行 5 次 10 折交叉验证的错误率分布情况，该模型用样品中 OTU 的相对丰度进行训练，总计采用了 17 位非子宫内膜异位症个体和 32 位子宫内膜异位症个体的 CL 样品，黑色线代表 5 次试验的平均值，灰色线则分别代表 5 次试验，黑色竖线代表最佳组合中 OTU 数目；b 图为经过交叉验证过的组合的接收者操作曲线，曲线下面积 AUC 为 0.8272，阴影面积代表 95%置信区间，对角线代表 AUC 为 0.5 的曲线。

图 2 为阴道后穹窿（CU）的标志物组鉴别子宫内膜异位症，图中， a 图为随着 OTU 数量的增加，对随机森林鉴别子宫内膜异位症进行 5 次 10 折交叉验

证的错误率分布情况，该模型用样品中 OTU 的相对丰度进行训练，总计采用了 17 位非子宫内膜异位症个体和 32 位子宫内膜异位症个体的 CU 样品，黑色线代表 5 次试验的平均值，灰色线分别为 5 次试验，黑色竖线代表最佳组合中 OTU 数目； b 图为经过交叉验证过的组合的接收者操作曲线，曲线下面积 AUC 为 0.5919，阴影面积代表 95%置信区间，对角线代表 AUC 为 0.5 的曲线。

图 3 为宫颈管（CV）的标志物组鉴别子宫内膜异位症，图中， a 图为随着 OTU 数量的增加，对随机森林鉴别子宫内膜异位症进行 5 次 10 折交叉验证的错误率分布情况，该模型用样品中 OTU 的相对丰度进行训练，总计采用了 17 位非子宫内膜异位症个体和 32 位子宫内膜异位症个体的 CV 样品，黑色线代表 5 次试验的平均值，灰色线分别为 5 次试验，黑色竖线代表最佳组合中 OTU 数目；b 图为经过交叉验证过的组合的接收者操作曲线，曲线下面积 AUC 为 0.8493，阴影面积代表 95%置信区间，对角线代表 AUC 为 0.5 的曲线。

由图 1 至图 3 的结果可以看出，三个不同位点的 OTU 生物标志物组能够鉴别子宫内膜异位症个体和非子宫内膜异位症个体；ROC 的曲线下面积 AUC 值分别为 0.8272（CL），0.5919（CU）和 0.8493（CV）。其中，AUC 是曲线下面积，该值越大，即越接近 1，表示判断能力越强，即判断越准确。

2.3 生物标志物验证

将随机森林得到的 OTU 生物标志物在第二群体样品中进行验证，结果如表 8、表 9 和表 10 所示。表 8 至表 10 中，样品编号例如 C003CL、C003CU、C003CV，分别表示采集自同样一个 C003 采样对象的 CL、CU、CV 三个部位的样品。表 8 至表 10 为三个标志物组预测个体患有子宫内膜异位症的概率，由此得到的 ROC 曲线依序为图 4 至图 6。表 8 至表 10 中，概率>0.5 认为通过该部位的标志物组判断个体具有患子宫内膜异位症的风险或者患有子宫内膜异位症。

表 8 CL 部位的 CL 标志物组预测第二群体样品患有子宫内膜异位症的概率

样品编号	是否子宫内膜异位（不孕）（N：否；Y 是）	概率
C001CL	N	0.427
C003CL	N	0.428
C004CL	N	0.261
C005CL	N	0.295
C007CL	N	0.427
C008CL	N	0.541
C009CL	N	0.577
C012CL	N	0.337
C014CL	N	0.447
C019CL	N	0.563
C021CL	N	0.649
T000CL	Y	0.575
T001CL	Y	0.691
T003CL	Y	0.851

T005CL	Y	0.526
T006CL	Y	0.731
T008CL	Y	0.366
T015CL	Y	0.808
T017CL	Y	0.746
T081CL	Y	0.330
T082CL	Y	0.784
T083CL	Y	0.761

表 9 CU 部位的 CU 标志物组预测第二群体样品患有子宫内膜异位症的概率

样品编号	是否子宫内膜异位（不孕）（N：否；Y 是）	概率
C001CU	N	0.285
C003CU	N	0.106
C004CU	N	0.369
C005CU	N	0.443
C007CU	N	0.298
C008CU	N	0.588
C009CU	N	0.675
C012CU	N	0.988
C014CU	N	0.121
C019CU	N	0.188
C021CU	N	0.584
T001CU	Y	0.943
T003CU	Y	0.514
T005CU	Y	0.969
T006CU	Y	0.885
T000CU	Y	0.475
T008CU	Y	0.135
T015CU	Y	0.838
T017CU	Y	0.943
T081CU	Y	0.722
T082CU	Y	0.943
T083CU	Y	0.992

表 10 CV 部位的 CV 标志物组预测第二群体样品患有子宫内膜异位症的概率

样品编号	是否子宫内膜异位（不孕）（N：否；Y 是）	概率
C003CV	N	0.415
C004CV	N	0.415
C005CV	N	0.351
C007CV	N	0.388
C008CV	N	0.347
C009CV	N	0.515
C012CV	N	0.418
C014CV	N	0.417
C019CV	N	0.152
C021CV	N	0.152
T000CV	Y	0.546
T001CV	Y	0.286
T003CV	Y	0.632
T004CV	Y	0.395
T005CV	Y	0.633
T006CV	Y	0.723
T008CV	Y	0.642
T015CV	Y	0.554

T017CV	Y	0.717
T081CV	Y	0.448
T082CV	Y	0.573
T083CV	Y	0.408

图 4 的结果显示 CL 部位基于 CL 标志物组判断子宫内膜异位症概率，其 AUC 值为 0.8750；图 5 的结果显示 CU 部位基于 CU 标志物组判断子宫内膜异位症概率，其 AUC 值为 0.840；图 6 的结果显示 CV 部位基于 CV 标志物组判断子宫内膜异位症概率，其 AUC 值为 0.9189；可见，这三个标志物组具有较高的鉴别能力，能够用于子宫内膜异位症的检测，该结果与表 8 至表 10 的结果相符。表 8 至表 10 的结果中，三个标志物组预测的概率，其中至少一个大于 0.5，则判断个体具有患子宫内膜异位症的风险或者患有子宫内膜异位症，由此获得的判断结果，与实际情况相符。

以上内容是结合具体的实施方式对本申请所作的进一步详细说明，不能认定本申请的具体实施只局限于这些说明。对于本申请所属技术领域的普通技术人员来说，在不脱离本申请构思的前提下，还可以做出若干简单推演或替换。

权 利 要 求 书

1.一种用于子宫内膜异位症检测或患病风险评估的生物标志物组合，其特征在于：所述生物标志物组合包括二十四条核酸中的至少一条，所述二十四条核酸分别为 Seq ID No.1 至 Seq ID No.24 所示序列，或者分别为与 Seq ID No.1 至 Seq ID No.24 所示序列具有 97%以上相似性的序列。

2.一种用于子宫内膜异位症检测或患病风险评估的生物标志物组合，其特征在于：所述生物标志物组合包括第一标志物组、第二标志物组和第三标志物组中的至少一组；

所述第一标志物组由十四条核酸组成，所述十四条核酸分别为 Seq ID No.1、Seq ID No.2、Seq ID No.6、Seq ID No.7、Seq ID No.12、Seq ID No.13、Seq ID No.15、Seq ID No.17 至 Seq ID No.22、Seq ID No.24 所示序列，或者分别为与 Seq ID No.1、Seq ID No.2、Seq ID No.6、Seq ID No.7、Seq ID No.12、Seq ID No.13、Seq ID No.15、Seq ID No.17 至 Seq ID No.22、Seq ID No.24 所示序列具有 97%以上相似性的序列；

所述第二标志物组由两条核酸组成，所述两条核酸分别为 Seq ID No.1、Seq ID No.7 所示序列，或者分别为与 Seq ID No.1、Seq ID No.7 所示序列具有 97%以上相似性的序列；

所述第三标志物组由十一条核酸组成，所述十一条核酸分别为 Seq ID No.3 至 Seq ID No.5、Seq ID No.8 至 Seq ID No.12、Seq ID No.14、Seq ID No.16、Seq ID No.23 所示序列，或者分别为与 Seq ID No.3 至 Seq ID No.5、Seq ID No.8 至 Seq ID No.12、Seq ID No.14、Seq ID No.16、Seq ID No.23 所示序列具有 97%以上相似性的序列。

3.根据权利要求 2 所述的生物标志物组合，其特征在于：所述第一标志物组为 CL 标志物组，用于对来自阴道下 1/3 的样品进行子宫内膜异位症检测或患病风险评估。

4.根据权利要求 2 所述的生物标志物组合，其特征在于：所述第二标志物组为 CU 标志物组，用于对来自阴道后穹窿的样品进行子宫内膜异位症检测或患病风险评估。

5.根据权利要求 2 所述的生物标志物组合，其特征在于：所述第三标志物组为 CV 标志物组，用于对来自宫颈管的样品进行子宫内膜异位症检测或患病风险评估。

6.一种子宫内膜异位症检测或患病风险评估的试剂盒，其特征在于：所述试剂盒包含用于检测权利要求 1-5 任一项所述的生物标志物组合的引物对，所述引

物对的正向引物为SEQ ID No.25所示序列,反向引物为SEQ ID No.26所示序列。

7.根据权利要求1-5任一项所述的生物标志物组合在子宫内膜异位症药物筛选或者在制备子宫内膜异位症检测或患病风险评估的试剂盒或检测工具中的应用。

8.一种子宫内膜异位症的检测方法，其特征在于：包括以下步骤，

(1)对待测对象进行样品采集，检测所采集的样品中权利要求 1-5 任一项所述的生物标志物组合，并分析生物标志物组合中各核酸的水平；

(2)将步骤(1)测得的各核酸的水平与参考数据集或参考值进行比较，获得检测结果；

优选的，所述各核酸的水平为各核酸的相对丰度；所述参考数据集或参考值为来源于子宫内膜异位症患者和非子宫内膜异位症对照的生物标志物组合中各核酸的水平。

9.根据权利要求 8 所述的检测方法，其特征在于：所述步骤(2)中的参考数据集或参考值为表 5、表 6 或表 7 中的至少一组；将各核酸的水平与参考数据集或参考值进行比较获得检测结果，具体包括，利用多元统计模型计算得出患病概率，优选地，所述多元统计模型为随机森林模型。

10.根据权利要求 8 或 9 所述的检测方法，其特征在于：所述步骤(1)中对待测对象进行样品采集，包括采集待测对象阴道下 1/3 样品、阴道后穹窿样品和宫颈管样品。

11.一种通过检测生物标志物判断子宫内膜异位症的方法在制备子宫内膜异位症检测或患病风险评估试剂盒或工具中的应用；所述生物标志物为权利要求 1-5 任一项所述的生物标志物组合；

所述通过检测生物标志物判断子宫内膜异位症的方法包括以下步骤，

(1)对待测对象进行样品采集，检测所采集的样品中权利要求 1-5 任一项所述的生物标志物组合，并分析生物标志物组合中各核酸的水平；

(2)将步骤(1)测得的各核酸的水平与参考数据集或参考值进行比较，获得检测结果；

优选的，所述各核酸的水平为各核酸的相对丰度；所述参考数据集或参考值为来源于子宫内膜异位症患者和非子宫内膜异位症对照的生物标志物组合中各核酸的水平。

12.根据权利要求 11 所述的应用，其特征在于：所述步骤(2)中的参考数据集或参考值为表 5、表 6 或表 7 中的至少一组；将各核酸的水平与参考数据集或参考值进行比较获得检测结果，具体包括，利用多元统计模型计算得出患病概率，

优选地，所述多元统计模型为随机森林模型。

13.一种筛选治疗子宫内膜异位症的候选药物的方法，其特征在于：包括以下步骤，

1)分别测定用药前和用药后的样品中权利要求1-5任一项所述的生物标志物组合，并分析生物标志物组合中各核酸的水平；

2)根据比较用药前和用药后的样品中各核酸的水平，判断候选药物；

所述步骤 2)中，比较用药前和用药后的样品中各核酸的水平，具体包括，利用多元统计模型计算得出患病概率，优选地，所述多元统计模型为随机森林模型。

14.一种女性生殖道内微生物群的检测方法，其特征在于：包括以下步骤，

(1)采集待测对象生殖道内微生物样品，检测所采集的样品中权利要求 1-5任一项所述的生物标志物组合，并分析生物标志物组合中各核酸的水平；

(2)将步骤(1)测得的各核酸的水平与参考数据集或参考值进行比较，获得检测结果；

优选的，所述各核酸的水平为各核酸的相对丰度；所述参考数据集或参考值为来源于子宫内膜异位症人群和非子宫内膜异位症人群对照的生物标志物组合中各核酸的水平。

15.根据权利要求 14 所述的检测方法，其特征在于：所述步骤(2)中的参考数据集或参考值为表 5、表 6 或表 7 中的至少一组；将各核酸的水平与参考数据集或参考值进行比较获得检测结果，具体包括，利用多元统计模型计算得出患病概率，优选地，所述多元统计模型为随机森林模型。

16.根据权利要求 14 或 15 所述的检测方法，其特征在于：所述步骤(1)中采集待测对象生殖道内微生物样品，具体包括采集待测对象阴道下 1/3 样品、阴道后穹窿样品和宫颈管样品。

17.一种制备子宫内膜异位症生物标志物组合的方法，其特征在于：包括以下步骤，

(1)分别对子宫内膜异位症病患和非病患进行生殖道内微生物样品采集，对所有采集的样品分别进行 16S 测序；

(2)将 16S 测序结果进行聚类分析，获得 OTU 单元以及每个 OTU 的种子序列，并计算每个 OTU 单元的相对丰度；

(3)利用随机森林模型对每个 OTU 单元的相对丰度与子宫内膜异位症状态进行拟合，并进行 5 次十折交叉验证，得到最优的 OTU 组合，最优 OTU 组合中各 OTU 的种子序列，即组成子宫内膜异位症的生物标志物组合。

18.根据权利要求 17 所述的方法，其特征在于：所述步骤(1)中，生殖道内微生物样品采集，具体包括采集待测对象阴道下 1/3 样品、阴道后穹窿样品和宫颈管样品。

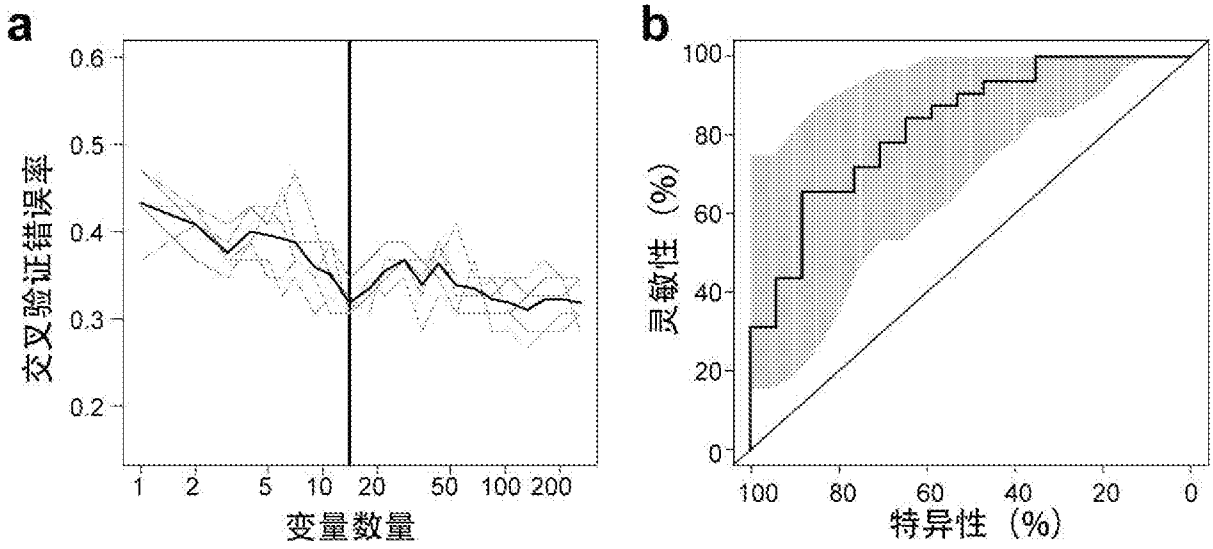

a
交叉验证错误率
0.6
0.5
0.4
0.3
0.2
1 2 5 10 20 50 100 200
变量数量
b
灵敏性（%）
100
80
60
40
20
0
100 80 60 40 20 0
特异性（%）

图 1

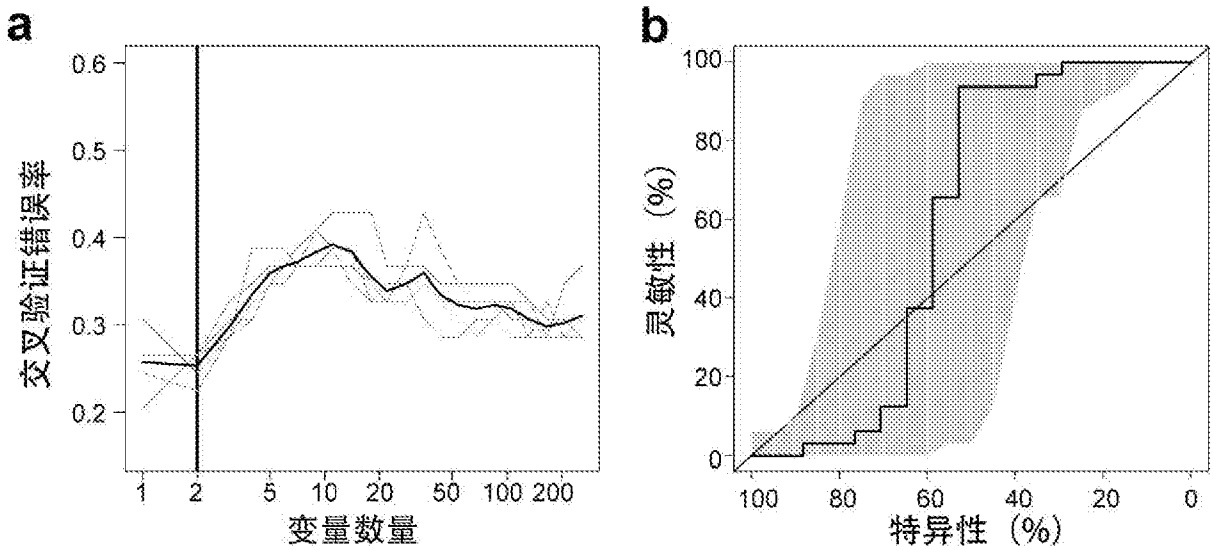

a
交叉验证错误率
0.6
0.5
0.4
0.3
0.2
1 2 5 10 20 50 100 200
变量数量
b
灵敏性（%）
100
80
60
40
20
0
100 80 60 40 20 0
特异性（%）

图 2

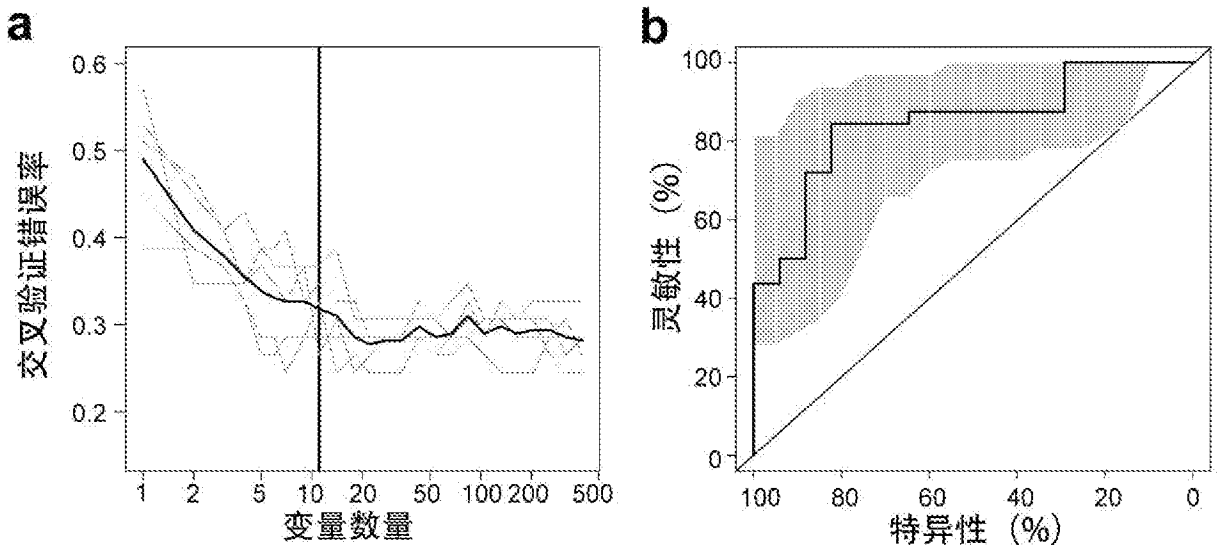

a
交叉验证错误率
0.6
0.5
0.4
0.3
0.2
1 2 5 10 20 50 100 200 500
变量数量
b
灵敏性（%）
100
80
60
40
20
0
100 80 60 40 20 0
特异性（%）

图 3

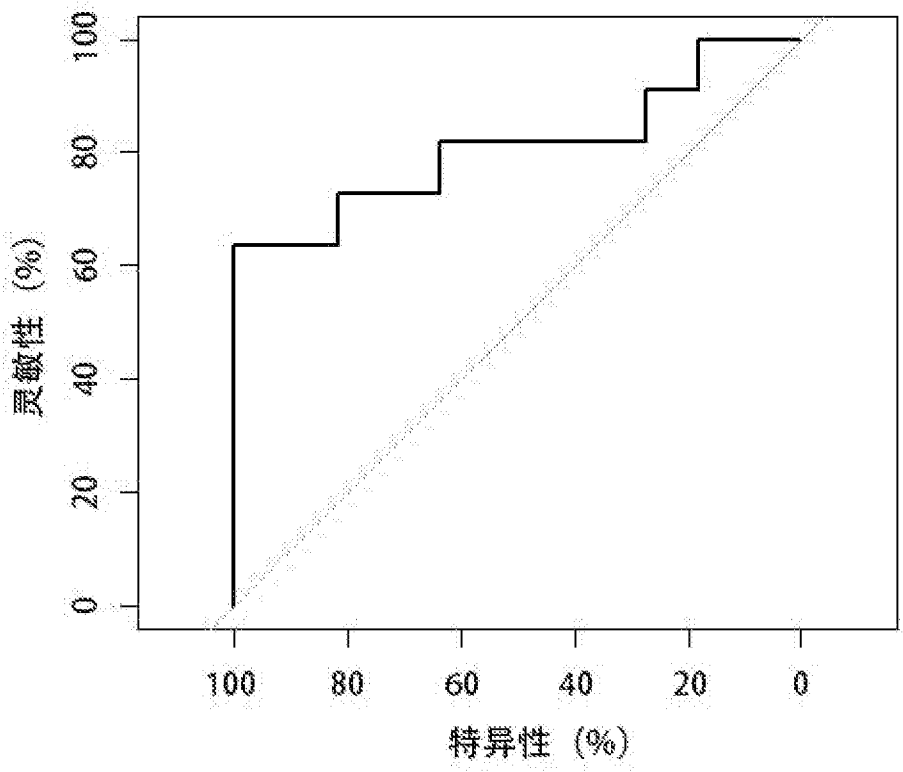

灵敏性（%）
100
80
60
40
20
0
100 80 60 40 20 0
特异性（%）

图 4

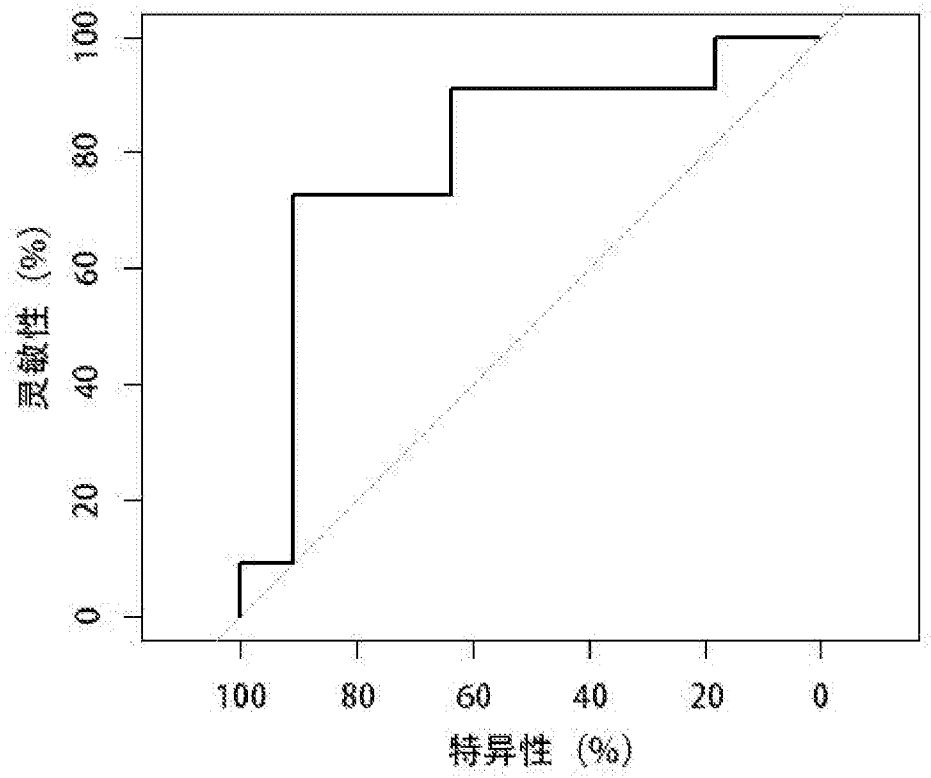

灵敏性（%）
0
20
40
60
80
100
100
80
60
40
20
0
特异性（%）

图 5

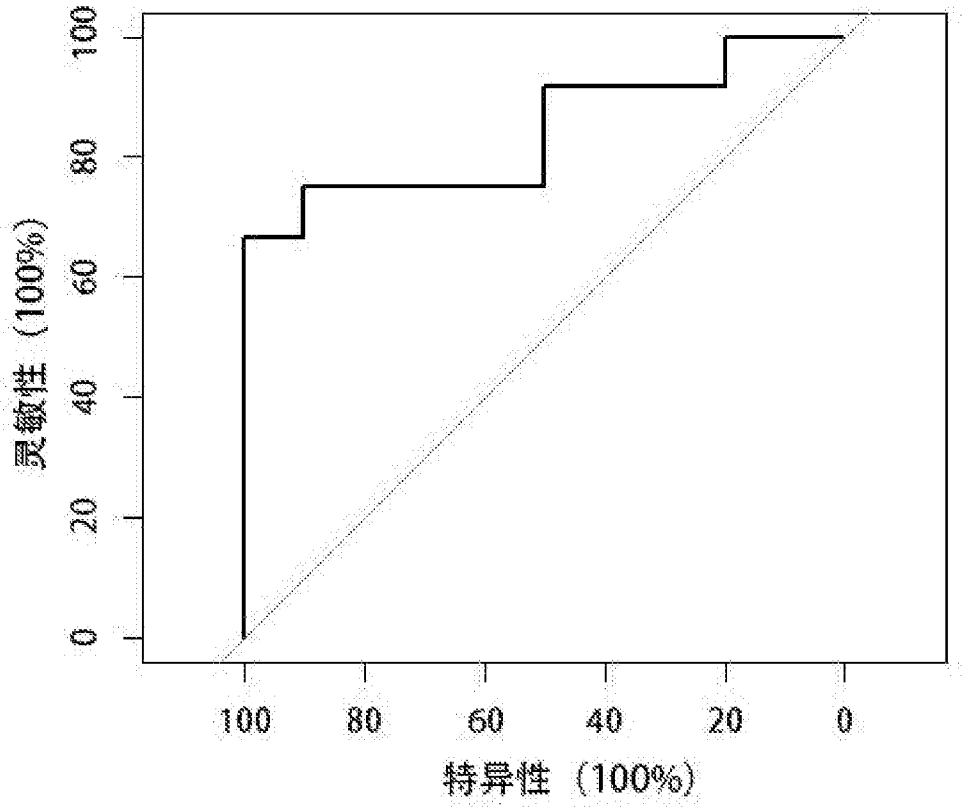

灵敏性（100%）
0
20
40
60
80
100
100
80
60
40
20
0
特异性（100%）

图 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/096249

A. CLASSIFICATION OF SUBJECT MATTER

C12Q 1/68 (2006.01) i; C12N 15/11 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12Q 1/-; C12N 15/-

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

VEN, CNABS, CNKI, PubMed, ISI web of Knowledge, Elsevier Science Direct, GenBank, EBI, Bio-Sequence Database of Chinese Patent: SEQ ID Nos: 1-26, 深圳华大基因研究院, 贾慧珏, 钟焕姿, 宋晓蕾, 王子榕, 陈晨, 子宫内膜异位, 生物标志, 分子标志, 嗜酸乳杆菌, biomarker, diagnosis, detection, endometriosis, 16S, Anaerococcus, Lactobacillus, Leptotrichiaceae, Vagococcus, Delftia, Dysgonomonas, Aerococcus, Prevotella, Tissierellaceae, Comamonadaceae, Erysipelotrichaceae, Dialister, Erysipelothrix

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 103249421 A (MEIJI CO., LTD.), 14 August 2013 (14.08.2013), claims 1-19, and description, paragraphs [0001]-[0061]	1-18
A	CN 104067124 A (SINGAPORE VOLITION PTE LIMITED), 24 September 2014 (24.09.2014), claims 1-23	1-18
A	US 2009220589 A1 (ABRAXIS BIOSCIENCE, LLC.), 03 September 2009 (03.09.2009), claims 1-81	1-18
A	US 2007287676 A1 (GUO, Sunwei et al.), 13 December 2007 (13.12.2007), claims 1-43	1-18
A	US 2008318237 A1 (REGENTS OF THE UNIVERSITY OF CALIFORNIA), 25 December 2008 (25.12.2008), claims 1-9, and tables 4-6	1-18
A	US 2015368714 A1 (WARD, K. et al.), 24 December 2015 (24.12.2015), claims 21-40	1-18

☒ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
27 September 2017	19 October 2017
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer YAO, Jinxiao Telephone No. (86-10) 62414292

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/096249

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2013148151 A1 (UNIVERSITY OF PITTSBURGH-OF THE COMMONWEALTH SYSTEM OF HIGHER EDUCATION), 03 October 2013 (03.10.5013), claims 1-23	1-18
A	CN 103026227 A (BRITISH COLUMBIA CANCER AGENCY BRANCH), 03 April 2013 (03.04.2013), claims 1-44	1-18
A	US 2015291959 A1 (KAOHSIUNG MEDICAL UNIVERSITY), 15 October 2015 (15.10.2015), claims 1-3	1-18
A	WO 2015128671 A1 (QUEEN MARY UNIVERSITY OF LONDON), 03 September 2015 (03.09.2015), claims 1-20	1-18
A	US 2006105405 A1 (ONCO DETECTORS INTERNATIONAL, LLC.), 18 May 2006 (18.05.2006), abstract	1-18

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/096249

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing filed or furnished:

   a. (means)

   ☐ on paper

   ☒ in electronic form

   b. (time)

   ☒ in the international application as filed

   ☐ together with the international application in electronic form

   ☐ subsequently to this Authority for the purposes of search

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2017/096249

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 103249421 A	14 August 2013	JP WO2012073924 A1	19 May 2014
		JP 5667642 B2	12 February 2015
		JP 2015044822 A	12 March 2015
		JP 5925274 B2	25 May 2016
		WO 2012073924 A1	07 June 2012
		HK 1187825 A1	23 October 2015
		SG 190719 A1	31 July 2013
		CN 103249421 B	20 May 2015
CN 104067124 A	24 September 2014	KR 20140078637 A	25 June 2014
		MX 2014002443 A	13 October 2014
		GB 201115095 D0	19 October 2011
		SG 10201601153R A	30 March 2016
		RU 2014112350 A	10 October 2015
		EP 2751568 A1	09 July 2014
		JP 2014531573 A	27 November 2014
		SG 11201400270P A	28 March 2014
		AU 2012300641 A1	06 March 2014
		HK 1199760 A1	17 July 2015
		CA 2845993 A1	07 March 2013
		ZA 201401295 B	23 December 2014
		BR 112014005087 A2	28 March 2017
		US 2014363812 A1	11 December 2014
		WO 2013030577 A1	07 March 2013
US 2009220589 A1	03 September 2009	WO 2009111375 A2	11 September 2009
		EP 2257626 A2	08 December 2010
		AU 2009222056 A1	11 September 2009
		CA 2717169 A1	17 September 2009
		WO 2009111375 A3	29 October 2009
		JP 2011515078 A	19 May 2011
US 2007287676 A1	13 December 2007	None	
US 2008318237 A1	25 December 2008	US 2014024557 A1	23 January 2014
		US 2013022593 A1	24 January 2013
		US 8247174 B2	21 August 2012
		US 9175349 B2	03 November 2015
		US 2016032392 A1	04 February 2016
		US 7871778 B2	18 January 2011
		US 2011171631 A1	14 July 2011
US 2015368714 A1	24 December 2015	US 9434991 B2	06 September 2016
		US 2016367568 A1	22 December 2016
WO 2013148151 A1	03 October 2013	None	
CN 103026227 A	03 April 2013	BR 112012027104 A2	16 November 2016
		US 2013197056 A1	01 August 2013
		WO 2011132175 A3	15 December 2011
		WO 2011132175 A2	27 October 2011
		EP 2561351 A2	27 February 2013
		EP 2561351 A4	16 October 2013
		CN 103026227 B	30 September 2015
		EP 2561351 B1	09 December 2015
		CA 2797291 A1	27 October 2011

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2017/096249

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
		HK 1181846 A1	27 January 2017
US 2015291959 A1	15 October 2015	TW I517852 B	21 January 2016
		US 9322017 B2	26 April 2016
		TW 201538162 A	16 October 2015
WO 2015128671 A1	03 September 2015	US 2017016067 A1	19 January 2017
		JP 2017509333 A	06 April 2017
		EP 3110970 A1	04 January 2017
		AU 2015221951 A1	15 September 2016
		GB 201403489 D0	16 April 2014
US 2006105405 A1	18 May 2006	None	

国际检索报告

国际申请号
PCT/CN2017/096249

A. 主题的分类

C12Q 1/68(2006.01)i; C12N 15/11(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

C12Q 1/-; C12N 15/-

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用））

VEN,CNABS,CNKI, PubMed, ISI web of Knowledge, Elsevier Science Direct, GenBank, EBI, Bio-Sequence Database of Chinese Patent: SEQ ID Nos: 1-26, 深圳华大基因研究院,贾慧珏,钟焕姿,宋晓蕾,王子榕,陈晨,子宫内膜异位,生物标志,分子标志, 嗜酸乳杆菌,biomarker, diagnosis, detection, endometriosis, 16S, Anaerococcus, Lactobacillus, Leptotrichiaceae, Vagococcus, Delftia, Dysgonomonas, Aerococcus, Prevotella, Tissierellaceae, Comamonadaceae, Erysipelotrichaceae, Dialister, Erysipelothrix

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
A	CN 103249421 A (株式会社明治) 2013年 8月 14日 (2013-08-14) 权利要求1-19，说明书第[0001]-[0061]段	1-18
A	CN 104067124 A (新加坡意志私人有限公司) 2014年 9月 24日 (2014-09-24) 权利要求1-23	1-18
A	US 2009220589 A1 (ABRAXIS BIOSCIENCE, LLC.) 2009年 9月 3日 (2009-09-03) 权利要求1-81	1-18
A	US 2007287676 A1 (GUO, SUN-WEI等) 2007年 12月 13日 (2007-12-13) 权利要求1-43	1-18
A	US 2008318237 A1 (REGENTS OF THE UNIVERSITY OF CALIFORNIA) 2008年 12月 25日 (2008-12-25) 权利要求1-9，表4-6	1-18
A	US 2015368714 A1 (WARD, KENNETH等) 2015年 12月 24日 (2015-12-24) 权利要求21-40	1-18

☑ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

"A" 认为不特别相关的表示了现有技术一般状态的文件

"E" 在国际申请日的当天或之后公布的在先申请或专利

"L" 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

"O" 涉及口头公开、使用、展览或其他方式公开的文件

"P" 公布日先于国际申请日但迟于所要求的优先权日的文件

"T" 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

"X" 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

"Y" 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

"&" 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
2017年 9月 27日	2017年 10月 19日
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451	受权官员 姚进孝 电话号码 (86-10)62414292

表 PCT/ISA/210 (第2页) (2009年7月)

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
A	WO 2013148151 A1 (UNIVERSITY OF PITTSBURGH-OF THE COMMONWEALTH SYSTEM OF HIGHER EDUCATION) 2013年 10月 3日 (2013-10-03) 权利要求1-23	1-18
A	CN 103026227 A (不列颠哥伦比亚省癌症分社) 2013年 4月 3日 (2013-04-03) 权利要求1-44	1-18
A	US 2015291959 A1 (KAOHSIUNG MEDICAL UNIVERSITY) 2015年 10月 15日 (2015-10-15) 权利要求1-3	1-18
A	WO 2015128671 A1 (QUEEN MARY UNIVERSITY OF LONDON) 2015年 9月 3日 (2015-09-03) 权利要求1-20	1-18
A	US 2006105405 A1 (ONCO DETECTORS INTERNATIONAL, LLC.) 2006年 5月 18日 (2006-05-18) 摘要	1-18

国际检索报告

国际申请号
PCT/CN2017/096249

第I栏　核苷酸和/或氨基酸序列(续第1页第1.c项)

1. 关于国际申请中所公开的任何对要求保护的发明必要的核苷酸和/或氨基酸序列，国际检索是在下列基础上进行的：

a. (提交提供)

☐ 纸件形式

☑ 电子形式

b. (提交时间)

☑ 含在申请提交时的国际申请中

☐ 以电子形式与国际申请一起提交

☐ 为检索之用随后提交本单位

2. ☐ 另外，在提交/提供了多个版本或副本的序列表的情况下，提供了关于随后提交的或附加的副本中的信息与申请时提交的申请中的信息相同或未超出申请时提交的申请中的信息范围（如适用）的所需声明。

3. 补充意见：

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2017/096249

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	103249421	A	2013年 8月 14日	JP	WO2012073924	A1	2014年 5月 19日
				JP	5667642	B2	2015年 2月 12日
				JP	2015044822	A	2015年 3月 12日
				JP	5925274	B2	2016年 5月 25日
				WO	2012073924	A1	2012年 6月 7日
				HK	1187825	A1	2015年 10月 23日
				SG	190719	A1	2013年 7月 31日
				CN	103249421	B	2015年 5月 20日
CN	104067124	A	2014年 9月 24日	KR	20140078637	A	2014年 6月 25日
				MX	2014002443	A	2014年 10月 13日
				GB	201115095	D0	2011年 10月 19日
				SG	10201601153R	A	2016年 3月 30日
				RU	2014112350	A	2015年 10月 10日
				EP	2751568	A1	2014年 7月 9日
				JP	2014531573	A	2014年 11月 27日
				SG	11201400270P	A	2014年 3月 28日
				AU	2012300641	A1	2014年 3月 6日
				HK	1199760	A1	2015年 7月 17日
				CA	2845993	A1	2013年 3月 7日
				ZA	201401295	B	2014年 12月 23日
				BR	112014005087	A2	2017年 3月 28日
				US	2014363812	A1	2014年 12月 11日
				WO	2013030577	A1	2013年 3月 7日
US	2009220589	A1	2009年 9月 3日	WO	2009111375	A2	2009年 9月 11日
				EP	2257626	A2	2010年 12月 8日
				AU	2009222056	A1	2009年 9月 11日
				CA	2717169	A1	2009年 9月 17日
				WO	2009111375	A3	2009年 10月 29日
				JP	2011515078	A	2011年 5月 19日
US	2007287676	A1	2007年 12月 13日		无		
US	2008318237	A1	2008年 12月 25日	US	2014024557	A1	2014年 1月 23日
				US	2013022593	A1	2013年 1月 24日
				US	8247174	B2	2012年 8月 21日
				US	9175349	B2	2015年 11月 3日
				US	2016032392	A1	2016年 2月 4日
				US	7871778	B2	2011年 1月 18日
				US	2011171631	A1	2011年 7月 14日
US	2015368714	A1	2015年 12月 24日	US	9434991	B2	2016年 9月 6日
				US	2016367568	A1	2016年 12月 22日
WO	2013148151	A1	2013年 10月 3日		无		
CN	103026227	A	2013年 4月 3日	BR	112012027104	A2	2016年 11月 16日
				US	2013197056	A1	2013年 8月 1日
				WO	2011132175	A3	2011年 12月 15日
				WO	2011132175	A2	2011年 10月 27日
				EP	2561351	A2	2013年 2月 27日
				EP	2561351	A4	2013年 10月 16日
				CN	103026227	B	2015年 9月 30日
				EP	2561351	B1	2015年 12月 9日
				CA	2797291	A1	2011年 10月 27日

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2017/096249

检索报告引用的专利文件			公布日（年/月/日）	同族专利			公布日（年/月/日）
				HK	1181846	A1	2017年 1月 27日
US	2015291959	A1	2015年 10月 15日	TW	I517852	B	2016年 1月 21日
				US	9322017	B2	2016年 4月 26日
				TW	201538162	A	2015年 10月 16日
WO	2015128671	A1	2015年 9月 3日	US	2017016067	A1	2017年 1月 19日
				JP	2017509333	A	2017年 4月 6日
				EP	3110970	A1	2017年 1月 4日
				AU	2015221951	A1	2016年 9月 15日
				GB	201403489	D0	2014年 4月 16日
US	2006105405	A1	2006年 5月 18日		无		