

(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
2020 年 12 月 10 日 (10.12.2020)



(10) 国际公布号
WO 2020/244018 A1

- (51) 国际专利分类号:
C12Q 1/68 (2018.01) *A61P 25/00* (2006.01)
- (21) 国际申请号: PCT/CN2019/094924
- (22) 国际申请日: 2019 年 7 月 5 日 (05.07.2019)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201910491545.X 2019年6月6日 (06.06.2019) CN
- (71) 申请人: 西安交通大学医学院第一附属医院(XI'AN JIAO TONG UNIV, HOSP 1, MED COLL) [CN/CN]; 中国陕西省西安市雁塔区雁塔西路 277 号, Shaanxi 710061 (CN)。 深圳华大生命科学研究院(BGI-SHENZHEN) [CN/CN]; 中国广东省深圳市北山工业区, Guangdong 518083 (CN)。
- (72) 发明人: 马现仓(MA, Xianchang); 中国陕西省西安市雁塔区雁塔西路 277 号, Shaanxi 710061 (CN)。 王奇(WANG, Qi); 中国广东省深圳市北山工业区,

Guangdong 518083 (CN)。 朱峰(ZHU, Feng); 中国陕西省西安市雁塔区雁塔西路 277 号, Shaanxi 710061 (CN)。 鞠艳梅(JU, Yanmei); 中国广东省深圳市北山工业区, Guangdong 518083 (CN)。 王崴(WANG, Wei); 中国陕西省西安市雁塔区雁塔西路 277 号, Shaanxi 710061 (CN)。 郭锐进(GUO, Ruijin); 中国广东省深圳市北山工业区, Guangdong 518083 (CN)。 范雅娟(FAN, Yajuan); 中国陕西省西安市雁塔区雁塔西路 277 号, Shaanxi 710061 (CN)。 贾慧珏(JIA, Huijue); 中国广东省深圳市北山工业区, Guangdong 518083 (CN)。 马青艳(MA, Qingyan); 中国陕西省西安市雁塔区雁塔西路 277 号, Shaanxi 710061 (CN)。 郭丽阳(GUO, Liyang); 中国陕西省西安市雁塔区雁塔西路 277 号, Shaanxi 710061 (CN)。 高成阁(GAO, Chengge); 中国陕西省西安市雁塔区雁塔西路 277 号, Shaanxi 710061 (CN)。

(74) 代理人: 西安智大知识产权代理事务所(XI'AN CYDA INTELLECTUAL PROPERTY AGENCY); 中国陕西省西安市雁塔区

(54) Title: SMALL-SCALE SCHIZOPHRENIA BIOMARKER COMBINATION, APPLICATION THEREOF AND METAPHLAN2 SCREENING METHOD THEREFOR

(54) 发明名称: 一种精神分裂症的小规模生物标志物组合、其应用及metaphlan2筛选方法

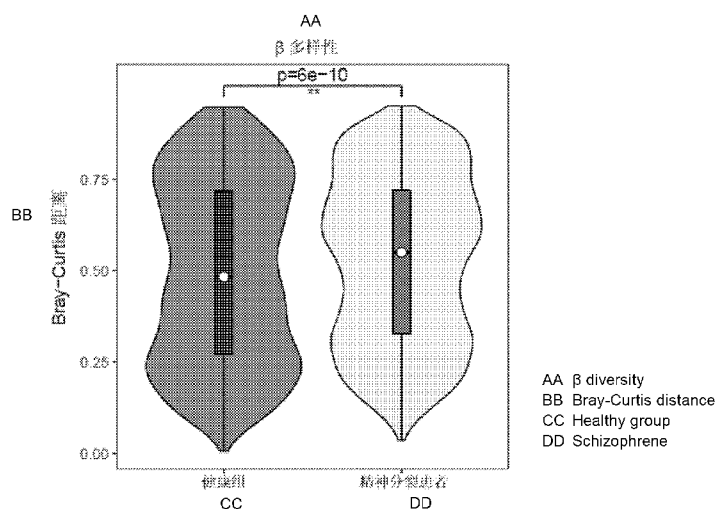


图 1

(57) Abstract: Disclosed are a small-scale schizophrenia biomarker combination, an application thereof and a metaphlan2 screening method therefor. The biomarker comprises one or more of *Veillonella parvula*, *Ruminococcus lactaris*, *Bacillus cereus thuringiensis*, *Bacillus nealsonii*, and *Bacteroides vulgatus*.

(57) 摘要: 本发明公开了一种精神分裂症的小规模生物标志物组合、其应用及 metaphlan2 筛选方法, 所述生物标志物包括 *Veillonella parvula*、*Ruminococcus lactaris*、*Bacillus cereus thuringiensis*、*Bacillus nealsonii*、*Bacteroides vulgatus*中的一种或多种。

科技路 195 号世纪颐园 A707, 段俊涛,
Shaanxi 710075 (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, 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))。

一种精神分裂症的小规模生物标志物组合、其应用及 *metaphlan2* 筛选方法**技术领域**

本发明属于生物医药技术领域，涉及一种基于肠道菌群的精神分裂症的生物标志物、诊断或预测精神分裂症风险的试剂盒及应用。

背景技术

精神分裂症（英语：Schizophrenia）是一组病因未明的重性精神病，多在青壮年缓慢或亚急性起病，临床上往往表现为症状各异的综合征，涉及感知觉、思维、情感和行为等多方面的障碍以及精神活动的不协调。患者一般意识清楚，智能基本正常，但部分患者在疾病过程中会伴有认知功能的损害。病程一般迁延，呈反复发作、加重或恶化，多数患者最终出现衰退和精神残疾，仅少数患者经过治疗后可达到痊愈或基本痊愈状态。

精神分裂症的全球患病率约为 0.3-0.7%。截止 2016 年，全球估计有超过 2100 万名精神分裂症患者，其平均预期寿命较正常人缩短 10 年至 25 年。

尽管既往研究表明精神分裂症发病是由遗传因素及环境因素共同作用所致，且患者的血清及脑组织都呈现部分异常改变，但目前对精神分裂症的诊断仍依赖于症状学评价，尚无可靠的生物学标记物标识。此外，现有的诊断标准不能早期预测精神分裂的发病，疗效及预后。

发明内容

本发明解决的问题在于提供一种精神分裂症的小规模生物标志物组合及其应用，仅使用最多 5 种标志物，即可克服现有精神分裂症诊断不能做到早期预警、不能预测精神分裂症发病以及发展趋势等缺点，能够帮助疾病病理分型以及为药物作用靶点研究、精准用药、发病机理的研究等。

本发明是通过以下技术方案来实现：

一种精神分裂症的小规模生物标志物组合，该生物标志物组合用于提供相对丰度信息，其包含选自以下的一种或多种：

生物标志物 1: *Veillonella*_parvula;

生物标志物 2: *Ruminococcus*_lactaris;

生物标志物 3: *Bacillus*_cereus_thuringiensis;

生物标志物 4: *Bacillus*_nealsonii;

生物标志物 5: *Bacteroides*_vulgatus。

所述的生物标志物组合提供的相对丰度信息用于和参考值进行比较。

所述的生物标志物 1~5 的相对丰度信息是基于能够对其进行丰度计算的基因序列来提供的。

所述的精神分裂症的小规模生物标志物组合作为检测靶点或检测目标在制备检测试剂盒中的应用。

所述的精神分裂症的小规模生物标志物组合的筛选方法，步骤如下：

1) 样本收集: 采集粪便样品后冷冻运输并迅速转移到-80℃保存, 进行 DNA 提取, 得到提取的 DNA 样本, 样本受试者包括精神分裂症病人和健康人;

2) 宏基因组测序与组装

3) 将高质量的测序片段输入到软件 Metaphlan2 计算出物种的相对丰度:

3.1) 将高质量测序片段比对到参考的标记基因上;

3.2) 根据比对结果统计插入片段的数量;

3.3) 将插入片段的数量对标记基因的长度进行标准化得到对应的丰度;

4) 从样本集中随机地选取精神分裂症病人和健康人作为训练集, 其余样品作为验证集, 计算训练集中每个样本中物种的相对丰度, 然后将训练集的物种输入随机森林分类器, 对分类器进行 5 次 10 折交叉验证, 利用 RF 模型筛选的物种相对丰度对每一个体计算其精神分裂症患病风险, 绘制 ROC 曲线, 并计算出 AUC 作为判别模型效能评价参数, 选取标志物组合数<30, 且判别效能最佳的组合, 在模型中输出每个物种的重要性指数, 重要性指数越高, 代表该标志物用来判别精神分裂症和非精神分裂症的重要性越高。

所述样本集中, 样本受试者包括 90 个精神分裂症病人和 81 个健康人, 验证集中, 样本受试者包括 10 个精神分裂症病人和 10 个健康人。

上述标志物的使用方法即诊断对象是否患有精神分裂症或者预测对象是否患有精神分裂症的风险的方法为:

1) 从对象中采集样本;

2) 确定步骤 1) 中获得的所述样本中的生物标志物的相对丰度信息;

3) 将步骤 2) 中所述的相对丰度信息与参考数据集或参考值进行比较。所述方法不仅仅可以用于专利法意义上的疾病诊断, 同时可以用作科学研究或者其他个人遗传信息的丰富以及遗传信息库的丰富等非疾病诊断。利用检测对象中的各生物标志物的相对丰度信息与参考数据集或参考值进行比较, 来确定对象是否患有精神分裂症, 或者预测其患有精神分裂症的风险。

所述参考数据集包括来自多个精神分裂症患者和多个健康对照的样本中的生物标志物的相对丰度信息。

所述参考数据集指的是对已确诊为患病个体和健康个体的样本进行操作, 所获得的各生物标志物的相对丰度信息, 用来作为每种生物标志物的相对丰度的参考。具体的, 参考数据集可以是指训练数据集。根据本发明, 所述训练集是指和验证集具有本领域公知的含义。在本发明的一个实施方案中, 所述训练集是指包含一定样本数的精神分裂症受试者和非精神分裂症受试者待测样本中的各生物标志物的含量的数据集合。所述验证集是用来测试训练集性能的独立数据集合。

本发明中所述参考值指的是健康对照的参考值或正常值。本领域技术人员已知, 当样本容量足够大时, 可利用本领域公知的检测和计算方法获得样品中每个生物标志物的正常值(绝

对值)的范围。当采用测定方法检测生物标志物的水平时,可将样品中的生物标志物水平的绝对值直接与参考值进行比较,以评估患病风险以及诊断或早期诊断精神分裂症,任选地,可以包括统计方法。

步骤 2)中所述的相对丰度信息与参考数据集进行比较的步骤中,还包括执行多元统计模型以获得患病概率。利用多元统计模型可以实现快速高效检测。具体的,所述多元统计模型为随机森林模型。

所述患病概率大于阈值表明所述对象患有精神分裂症或相关疾病或者有患有精神分裂症或相关疾病的风险。具体的,所述阈值为 0.5。

步骤 2)中所述生物标志物的相对丰度信息是利用测序方法得到的,进一步包括:从所述对象的所述样本中分离得到核酸样本,基于所获得的所述核酸样本,构建 DNA 文库,对所述 DNA 文库进行测序,以便获得测序结果;以及基于所述测序结果,将测序结果与参考基因集进行比对,以确定所述生物标志物的相对丰度信息。

根据本发明的一种实施例,可以利用 SOAP2 和 MAQ 的至少一种将测序结果与参考基因集进行比对,由此,可以提高比对的效率,进而可以提高精神分裂症检测的效率。根据本发明的实施例,可以同时多种(至少两种)生物标志物进行检测,可以提高精神分裂症检测的效率。

所述参考基因集包括从多个精神分裂症患者和多个健康对照的样本中进行宏基因组测序,获得非冗余基因集,然后将所述非冗余基因集与肠道微生物基因集合并,得到所述参考基因集。本发明中的参考基因集可以是已有的基因集,如现有的已经公开的肠道微生物参考基因集;也可以是将多个精神分裂症患者和多个健康对照的样品进行宏基因组测序,获得非冗余基因集,然后将所述非冗余基因集与肠道微生物基因集合并,得到所述参考基因集,由此获得的参考基因集信息更全面,检测结果更可靠。

所述非冗余基因集作本领域技术人员通常的理解来解释,简单来说去除冗余基因后的剩余基因的集合。冗余基因通常指的是一条染色体上出现的一个基因的多个复份。

具体的,所述样本为粪便样本。所述测序方法是通过第二代测序方法或第三代测序方法进行的。进行测序的手段并不受特别限制,通过二代或者三代测序的方法进行测序,可以实现快速高效的测序。

所述测序方法是通过选自 Hiseq2000、SOLiD、454、和单分子测序装置的至少一种进行的。由此,能够利用这些测序装置的高通量、深度测序的特点,从而有利于对后续测序数据进行分析,尤其是进行统计学检验时的精确性和准确度。

本发明提出了所述的精神分裂症的小规模生物标志物组合作为检测靶点或检测目标在制备检测试剂盒中的应用,所述试剂盒用于诊断对象是否患有精神分裂症或相关疾病或者预测对象是否患有精神分裂症或相关疾病的风险。

即,本发明提出了一种包括用于检测生物标志物的试剂的试剂盒,利用该试剂盒,可以确定这些标志物在肠道菌群中的相对丰度,由此,可以通过所得到的相对丰度值,从而确定

对象是否患有或者易感精神分裂症，以及用于监控精神分裂症患者的治疗效果的效率。

本发明提出了精神分裂症的小规模生物标志物组合作为靶点在筛选治疗和/或者预防精神分裂症的药物中的应用。所述生物标志物为上述本发明提出的生物标志物，可以利用候选药物使用前和使用后对这些生物标志物的影响，从而确定候选药物是否可以用于治疗或预防精神分裂症。

所述的生物标志物组合相对丰度的变化为确定候选药物是否有效提供依据。

与现有技术相比，本发明具有以下有益的技术效果：

本发明基于以下事实和问题的发现和认识作出的：肠道微生物是存在于人体肠道中的微生物群落，是人体的“第二基因组”。人体肠道菌群和宿主构成一个相互关联的整体。肠道微生物能够产生在人脑中发现的大多数神经递质。越来越多的证据支持肠道微生物影响中枢神经化学和行为的观点，肠易激综合征被认为是大脑-肠道微生物轴调节紊乱的典型病例。转化研究表明，某些特定的菌群可能会对压力反应和认知功能产生影响。用益生菌或抗生素改变肠道微生物群，为提高大脑功能和治疗抑郁及自闭症等肠-脑轴疾病提供了一种新的方法。因此，本发明通过对精神分裂症患者和健康人群的肠道菌群以及基因序列进行分析，从而筛选出与精神分裂症相关性高的生物标志物，并且利用该标志物能够准确地诊断精神分裂症或者预测患病风险，并且可以用于监测治疗效果。

粪便是人体的代谢产物，其内不仅包含人体的代谢产物，还包括对我们的机体代谢和免疫以及脑功能的变化密切相关的肠道微生物。对粪便进行研究，发现在精神分裂症患者和健康人群的肠道菌群的组成上存在明显的差异，可以准确地对精神分裂症患者进行患病风险评估、早期诊断。本发明基于对精神分裂症患者和健康人群肠道菌群的比较和分析，得到多种相关的肠道微生物，结合高质量的精神分裂症人群和健康人群物种相对丰度数据作为训练集，能够准确地对精神分裂症患者进行患病风险评估、早期诊断。该方法与目前常用的诊断方法相比，具有方便、快捷的特点。

本发明提出的精神分裂症相关的生物标记物对早期诊断是有价值的。第一，本发明的标记物具有较高的特异性和灵敏性。第二，粪便的分析保证准确性、安全性、可负担性和患者依从性。并且粪便的样本是可运输的。基于聚合酶链反应（PCR）的试验舒适且无创，所以人们会更容易参与给定的筛选程序。第三，本发明的标记物还可以用作于对精神分裂症患者进行治疗监测的工具以检测对治疗的响应。由于丰度度量的原因，本发明中 5 种标记物的组合特别适用于基于 *metaphlan2* 方法度量丰度的情况。

附图说明

图 1 为根据本发明一个实施例物种属水平上精神分裂症患者和健康对照 β 多样性的情况。图示表明，精神分裂症患者和健康对照存在显著差异。

图 2 为根据本发明的一个实施例随机森林分类器中 5 次 10 折交叉验证的错误率分布情况。

图 3 为根据本发明的一个实施例基于随机森林模型（5 个肠道标志物），由精神分裂症患者和健康对照组成的训练集的接收者操作特征（Receiver Operating Characteristic, ROC）曲线

和曲线下面积 (Area under Curve, AUC)。

图 4 为根据本发明的一个实施例基于随机森林模型 (5 个肠道标志物), 由精神分裂症患者和健康对照组成的验证集的 ROC 曲线和 AUC。

具体实施方式

本发明所用术语具有相关领域普通技术人员通常理解的含义。然而, 为了更好地理解本发明, 对一些定义和相关术语的解释如下:

“精神分裂症”, 是一组病因未明的重性精神病, 多在青壮年缓慢或亚急性起病, 临床上往往表现为症状各异的综合征, 涉及感知觉、思维、情感和行为等多方面的障碍以及精神活动的不协调。

“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。这样的生物标志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, 例如, 核酸标志物 (也可以称为基因标志物, 例如 DNA), 蛋白质标志物, 细胞因子标志物, 趋化因子标志物, 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。其中, 核酸标志物的含义并不局限于现有可以表达为具有生物活性的蛋白质的基因, 还包括任何核酸片段, 可以为 DNA, 也可以为 RNA, 可以是经过修饰的 DNA 或者 RNA, 也可以是未经修饰的 DNA 或者 RNA, 以及由它们组成的集合。在本文中核酸标志物有时也可以称为特征片段。在本发明中, 生物标志物也可以用“肠道标志物”来表示, 因为所发现的与精神分裂症相关的生物标志物均存在于受试者的肠道内。生物标志物经过测量和评估, 经常用以检查正常生物过程, 致病过程, 或治疗干预药理响应, 而且在许多科学领域都是有用的。

所述的生物标志物, 可以运用高通量测序, 批量分析健康人群和精神分裂症患者的粪便样本。基于高通量测序数据, 对健康人群与精神分裂症患者群进行比对, 从而确定与精神分裂症患者群相关的特异性核酸序列。简言之, 其步骤如下:

样品的收集与处理: 收集健康人群与精神分裂症患者群的粪便样本, 使用试剂盒进行 DNA 提取, 得到核酸样本;

文库构建和测序: DNA 文库构建和测序是利用高通量测序进行, 以便得到粪便样品中所包含肠道微生物的核酸序列;

通过生物信息学的分析方法, 确定与精神分裂症患者相关的特异性肠道微生物核酸序列。首先, 将测序序列 (reads) 与参照基因集 (也称为参考基因集, 可以为新构建的基因集或任何已知序列的数据库, 例如, 采用已知的人肠道微生物群落非冗余基因集) 进行比对。接下来, 基于比对结果, 分别确定来自健康人群和精神分裂症患者群粪便样品的核酸样本中各基因的相对丰度。通过将测序序列与参照基因集进行比对, 可以将测序序列与参照基因集中的基因建立对应关系, 从而针对核酸样本中的特定基因, 与其相对应的测序序列的数目可以有效地反映该基因的相对丰度。由此, 可以通过比对结果, 按照常规的统计分析, 确定在核酸样本中基因的相对丰度。最后, 在确定核酸样本中各基因的相对丰度后, 对来自健康人群和

精神分裂症患者群粪便的核酸样本中各基因的相对丰度进行统计检验，由此，可以判断在健康人群和精神分裂症患者人群中是否存在相对丰度有显著差异的基因，如果存在基因是显著差异的，则该基因被当作是异常状态的生物标志物，即核酸标志物。

另外，对于已知或新构建的参照基因集，其通常包含基因物种信息和功能注释，由此，在确定基因相对丰度的基础上，可以进一步通过将基因的物种信息和功能注释进行分类，从而确定肠道菌群中各微生物的物种相对丰度和功能相对丰度，也就可以进一步确定异常状态的物种标志物和功能标志物。简言之，确定物种标志物和功能标志物的方法进一步包括：将健康人群和精神分裂症患者群的测序序列与参照基因集进行比对；基于比对结果，分别确定健康人群和精神分裂症患者群的核酸样本中各基因的物种相对丰度和功能相对丰度；对来自健康人群和精神分裂症患者群的核酸样本中各基因的物种相对丰度和功能相对丰度进行统计学检验；以及分别确定在健康人群和精神分裂症患者群的核酸样本之间相对丰度存在显著差异的物种标志物和功能标志物。根据本发明的实施例，可以采用对来自相同物种的基因的相对丰度和具有相同功能注释的基因的相对丰度进行统计检验，例如加和、取平均值、中位数值等，来确定功能相对丰度和物种相对丰度。

若未特别指明，实施例中所采用的技术手段为本领域技术人员所熟知的常规手段，可以参照《分子克隆实验指南》第三版或者相关产品进行，所采用的试剂和产品也均为可商业获得的。未详细描述的各种过程和方法是本领域中公知的常规方法，所用试剂的来源、商品名以及有必要列出其组成成分者，均在首次出现时标明，其后所用相同试剂如无特殊说明，均与首次标明的内容相同。

本发明采用宏基因组关联分析(Metagenome-Wide Association Study, MWAS)的分析方法，经测序分析粪便样本的菌群组成，功能差异；用随机森林判别模型判别精神分裂症群体和非精神分裂症群体，获得患病概率，用于精神分裂症的患病风险评估、诊断、早期诊断或者寻找潜在药物靶点。

根据本发明，术语“个体”指动物，特别是哺乳动物，如灵长类动物，最好是人。

根据本发明，术语如“一”、“一个”和“这”不仅指单数的个体，而是包括可以用来说明特定实施方式的通常的一类。

在本发明中，所述的测序（二代测序）和 MWAS 具有本领域所公知，本领域技术人员可以根据具体情况进行调整。根据本发明的实施例，可以依据文献（Jun Wang, and Huijue Jia. Metagenome-wide association studies: fine-mining the microbiome. Nature Reviews Microbiology 14.8 (2016): 508-522.）中记载的方法进行。

在本发明中，随机森林模型和 ROC 曲线的使用方法为本领域所公知，本领域技术人员可以根据具体情况进行参数设置和调整。根据本发明的实施例，可以根据文献（Drogan D, Dunn

WB, Lin W, Buijsse B, Schulze MB, Langenberg C, Brown M, Floegel a., Dietrich S, Rolandsson O, Wedge DC, Goodacre R, Forouhi NG, Sharp SJ, Spranger J, Wareham NJ, Boeing H: Untargeted Metabolic Profiling Identifies Altered Serum Metabolites of Type 2-Diabetes Mellitus in a Prospective, Nested Case Control Study. Clin Chem 2015, 61:487-497.; Mihalik SJ, Michaliszyn SF, de las Heras J, Bacha F, Lee S, Chace DH, DeJesus VR, Vockley J, Arslanian SA: Metabolomic profiling of fatty acid and amino acid metabolism in youth with obesity and type 2 diabetes: evidence for enhanced mitochondrial oxidation. Diabetes Care 2012, 35:605-611., 通过引用全文并入此处) 中记载的方法进行。

在本发明中, 构建了精神分裂症受试者和非精神分裂症受试者的生物标志物的训练集, 并以此为基准, 对待测样本的生物标志物含量值进行评估。

本领域技术人员知晓, 当进一步扩大样本量时, 利用本领域公知的样本检测和计算方法, 可以得出每种生物标志物在样本中的正常含量值区间(绝对数值)。可以将检测得到的生物标志物含量的绝对值与正常含量值进行比较, 任选地, 还可以结合统计学方法, 以得出精神分裂症患病风险评价、诊断以及用于监控精神分裂症患者的治疗效果的效率等。

不希望受任何理论的限制, 发明人指出这些生物标志物是存在于人体中的肠道菌群。通过本发明所述的方法对受试者肠道菌群进行关联分析, 得到精神分裂症群体的所述生物标志物在菌群检测中表现出一定的含量范围值。

下面结合具体的实施例对本发明做进一步的详细说明, 所述是对本发明的解释而不是限定。

实施例 1

1.1 样本收集

参照文献 A metagenome-wide association study of gut microbiota in type 2 diabetes (Qin J et al. Nature. 2012, 490, 55-60) 记载的方法, 采集粪便样品后冷冻运输并迅速转移到-80℃保存, 进行 DNA 提取, 得到提取的 DNA 样本。本发明的精神分裂症和非精神分裂症受试者的粪便样本来自中国。共计 171 例, 其中健康样本 81 例和精神分裂症样本 90 例。

1.2 宏基因组测序与组装

利用所提取的 DNA 样本构建测序文库, 在 Illumina HiSeq2000 测序平台上进行双向 (Paired-end) 宏基因组测序 (插入片段 350bp, 读长 100bp)。对测序产生的数据进行过滤 (quality-controlled), 去除 adapter 污染序列、低质量序列和宿主基因组污染序列, 得到高质量的测序片段 (reads)。

1.3 基因组比对与丰度计算

将上述“1.2 宏基因组测序与筛选”的高质量测序片段 (reads) 输入到软件 Metaphlan2

(<http://segatalab.cibio.unitn.it/tools/metaphlan2/>) 即可计算出物种的相对丰度。参照文献 MetaPhlAn2 for enhanced metagenomic taxonomic profiling, Nature Methods 12, 902-903 (2015) 记载的方法。其丰度计算步骤如下：1) 将高质量测序片段比对到参考的标记基因上；2) 根据比对结果统计插入片段的数量；3) 将插入片段的数量对标记基因的长度进行标准化（按平均基因长度进行标准化，并按照向下取整的方式得到对应物种的丰度）得到对应的丰度。

1.4 利用随机森林（ROC/AUC）筛选精神分裂症发生发展的潜在生物标志物

为进一步筛选潜在疾病肠道生物标志物，本实施例构建了精神分裂症受试者和非精神分裂症受试者的生物标志物的训练集，并以此为基准，对待测样本的生物标志物含量值进行评估。其中，在本发明中，所述训练集和所述验证集具有本领域所公知的含义。在本发明的实施方案中，训练集是指包含一定样本数的精神分裂症受试者和非精神分裂症受试者待测样本中的各生物标志物的含量的数据集合。验证集是用来测试训练集性能的独立数据集合。其中，非精神分裂症受试者为精神状态良好的受试者，受试者可以为人或者模型动物，在本实施例中是以人为受试者进行实验的。

具体包括如下步骤：

本发明从 171 个样品（90 个精神分裂症病人和 81 个健康人）中，随机地选取 80 个精神分裂症病人和 71 个健康人作为训练集（表 1），其余样品作为验证集（10 个精神分裂症病人和 10 个健康人）。

1.4.1 利用训练集数据筛选得到的生物标志物

首先，按照 1.3 描述的方法计算训练集中每个样本中物种的相对丰度。然后将训练集的物种输入随机森林（randomForest 4.6-12 in R 3.2.5, RF）分类器。对分类器进行 5 次 10 折交叉验证，利用 RF 模型筛选的物种相对丰度对每一个体计算其精神分裂症患病风险，绘制 ROC 曲线，并计算出 AUC 作为判别模型效能评价参数。选取标志物组合数 < 30 ，且判别效能最佳的组合为本发明组合。在模型中输出每个物种的重要性指数，重要性指数越高，代表该标志物用来判别精神分裂症和非精神分裂症的重要性越高。

本发明所得 RF 分类器包含了 5 个代谢物（即 5 个生物标记物），这 5 个生物标记物的相对丰度分别如表 1 所示，其详细信息如表 2 所示。表 3 示出了 5 种生物标记物结合来预测训练集的患病概率，其中患病概率 ≥ 0.5 确认个体具有患精神分裂的风险或者患有精神分裂症。

图 1 示出了物种属水平上精神分裂症患者和健康对照 β 多样性的情况。图示表明，精神分裂症患者和健康对照存在显著差异。

图 2 示出了随机森林分类器中 5 次 10 折交叉验证的错误率分布情况。该模型用训练集样品在经 MWAS 流程处理得到的满足目标的物种相对丰度进行训练，粗黑色实曲线代表 5 次试验（细黑色曲线代表 5 次试验）的平均值，黑色竖线代表所选最佳组合中物种数目。

图 3 示出了基于随机森林模型（5 个生物标志物）由精神分裂症患者和健康对照组成的训练集的 ROC 曲线和 AUC，其中特异性表征的是对于不患病判对的概率，敏感性指的是对于患病判对的概率，对训练集样本的判别效能为：AUC=82.66%，95%置信区间 CI=76.06-89.25%。结果表明该模型所得代谢物组合可作为区分精神分裂症与非精神分裂症的潜在生物标志物。

1.4.2 利用验证集数据验证筛选得到的生物标志物

本发明随即使用独立人群对该模型进行验证，患病概率 ≥ 0.5 预测个体具有精神分裂症患病风险或者患有精神分裂症。首先，按照 1.3 描述的方法计算训练集中每个样本中各生物标志物的相对丰度。然后按照 1.4.1 的方法利用随机森林模型对验证集数据进行验证。

基于该模型：

5 个生物标志物在验证集中的相对丰度如表 4 所示。表 5 示出了基于 5 个生物标志物、*Veillonella_parvula* 和 *Ruminococcus_lactaris* 预测验证集的患病概率。

图 4 示出了基于随机森林模型（5 个生物标志物）由精神分裂症患者和健康对照组成的独立验证集的 ROC 曲线和 AUC，模型的判别 AUC=78% (95% CI= 56.81%-99.19%)。

在 3.2.5 版本 R 中使用“randomForest 4.6-12 package”进行随机森林模型分类和回归。输入包括训练集数据（即训练样本中选定的物种标记物的相对丰度，见表 1），样本疾病状态（训练样本的样本疾病状态为矢量，‘1’代表精神分裂症，‘0’代表健康对照），以及一个验证集（验证集中所选物种标记物的相对丰度，见表 4）。然后，发明人利用 R 软件中随机森林包的随机森林函数建立分类和预测函数对验证集数据进行预测，输出即为预测结果（患病概率）；阈值为 0.5，如果疾病的概率 ≥ 0.5 ，则认为有患精神分裂症的风险或者患有精神分裂症。

表 1 随机森林模型训练集肠道标志物相对丰度数据

样本编号*	<i>Bacteroides_vulgatus</i>	<i>Bacillus_cereus_thuringiensis</i>	<i>Bacillus_nealsonii</i>	<i>Ruminococcus_lactaris</i>	<i>Veillonella_parvula</i>
H_1	0.52067	0.04657	0.00485	0.02693	0.00517
H_10	1.53882	0.00098	0	0.00313	0.00921
H_11	0.78626	0.00235	0	0.38521	0.00127
H_12	6.8102	0.00352	0	0.00571	0.00294
H_13	6.92272	0	0	0	0.01318
H_14	0.7211	0.01244	0	0	0.00109
H_15	2.32101	0	0	0	0.00798
H_16	7.97813	0	0	0	0
H_17	0.76136	0	0	0	0.00203
H_18	0.26233	0	0	0	0
H_19	0.62962	0.0072	0	1.95868	0.17759
H_2	2.37379	0.01611	0	0	0
H_20	10.21559	0.00466	0	0	0.00415
H_21	16.93799	0.00804	0	0	0

H_22	3.05481	0	0	0.05198	0
H_23	11.7051	0.00343	0	0	0.00235
H_24	0.20928	0.00252	0	0	0.01474
H_25	8.08408	0	0	0.69785	0.03665
H_26	0.86157	0	0	0.06325	0.009
H_27	0.40934	0.00171	0	0.02532	0.00301
H_28	54.76343	0.00242	0	0.06669	0.01347
H_29	0.38915	0.05755	0	16.22277	0.02788
H_3	5.73307	0.05538	0.00032	0.00178	0.19369
H_30	16.21343	0.00363	0	0.04618	0.02047
H_31	0.2384	0.0007	0	0.03353	0.02508
H_32	1.06	0	0	0.02646	0.05809
H_33	1.90714	0.0158	0	1.46603	0.06622
H_34	7.00649	0.03857	0	0.04566	0.11871
H_35	0.42508	0	0	0.00032	0.04264
H_36	0.43994	0	0	0.0005	0.05177
H_37	6.41437	0.00235	0	0	0.06829
H_38	0.28736	0	0	0.04608	0.13219
H_39	1.28611	0	0	0.00037	0.07744
H_4	7.30747	0.0734	0.00665	0.00093	0
H_40	35.0747	0.00189	0	0.00015	0.07776
H_41	3.7058	0	0	0	0.06797
H_42	6.49454	0	0	0.5446	0.01416
H_43	1.34793	0	0	0	0.00779
H_44	2.11402	0	0	0.00234	0.00617
H_45	1.2393	0	0	0	0.05326
H_46	0.85138	0	0	1.21108	0.00294
H_47	0.23987	0	0	0	0.04441
H_48	0.40044	0	0	0	0.00805
H_49	1.59713	0	0	0	0
H_5	2.05641	0.01161	0.00039	0.12137	0
H_50	2.21994	0	0	0	0
H_51	3.29144	0	0	0.01695	0.00364
H_52	0.09335	0	0	0.11646	0.04621
H_53	2.83512	0	0	0.24813	0.01141
H_54	0.07566	0	0	0	0
H_55	9.9835	0	0	0	0.01563
H_56	0.58004	0	0	0.00063	0.01822
H_57	0.99011	0	0	0.00018	0.04454
H_58	0.14343	0	0	0.00097	0.01211
H_59	0.11856	0	0	0.44203	0.00506
H_6	3.23622	0.04669	0	0.15344	0.00195
H_60	4.92199	0	0	0.00246	0.04867
H_61	1.0653	0	0	1.8839	0.00621
H_62	14.8859	0	0	0	0

H_63	3.93515	0	0	0	0
H_64	5.23453	0	0	0	0.05159
H_65	12.75465	0	0	1.40712	0.03997
H_66	0.75725	0	0	0.06037	0.00606
H_67	1.04207	0	0	0	0
H_68	2.11769	0	0	0.52495	0.00425
H_69	13.56747	0	0	0.02986	0.03965
H_7	2.05109	0	0	0.64396	0.00854
H_70	2.66965	0	0	0.0001	0.00024
H_71	5.0423	0	0	0	0
H_8	1.05414	0	0	0.00411	0.00213
H_9	2.89738	0.00744	0	0.00279	0
SZ_1	5.0708	0.01859	0.00417	0	0.02761
SZ_10	0.04143	0.03453	0.00534	0.57207	0.07441
SZ_11	1.87432	0.04172	0.00647	0.00136	0.01854
SZ_12	8.34254	0.06872	0.00861	0	0.04027
SZ_13	0.08937	0.05475	0.00158	0	0
SZ_14	0.8392	0.04359	0.00227	0.00043	0.00862
SZ_15	2.9692	0.07879	0.00979	0.55047	0.032
SZ_16	1.63137	0.03582	0.00159	0	0.07968
SZ_17	3.51122	0.07434	0.0084	0	0
SZ_18	0.06033	0.0426	0.00468	0.00141	0.0307
SZ_19	0.44684	0.05639	0.00452	0	0.01414
SZ_2	0.03022	0.01852	0	0	0.00517
SZ_20	0.09381	0.04455	0.0072	0	0.16308
SZ_21	0.07263	0.17161	0.02565	0.17444	0.30065
SZ_22	0.13153	0.02763	0.00032	0.01182	0.02437
SZ_23	0.3133	0.09161	0.00959	0.00314	0.00266
SZ_24	1.66699	0.07262	0.00524	0.00144	0
SZ_25	3.04615	0.0127	0	0	20.26271
SZ_26	2.36475	0	0	0.00419	0.1422
SZ_27	1.21034	0	0	0	0.01299
SZ_28	6.42166	0	0	0	0.17789
SZ_29	0.3994	0	0	0	0.26852
SZ_3	6.02213	0.03969	0.00081	0	0.05291
SZ_30	3.75298	0	0	0	0.39291
SZ_31	1.03049	0	0	0	0
SZ_32	0.78421	0	0	0.13206	0.24724
SZ_33	0.12797	0.02633	0	0	0.00386
SZ_34	0.53218	0	0	0	0.05209
SZ_35	4.62687	0.00838	0	0	0
SZ_36	2.24148	0.01227	0	0	0.03446
SZ_37	1.02069	0.00214	0	0	0.00271
SZ_38	0.08347	0.00465	0	0	0.01682
SZ_39	2.6547	0	0	0	0.04698

SZ_4	0.0152	0.05757	0.00478	0	0.00053
SZ_40	0.16448	0	0	0	0.13119
SZ_41	0.03185	0.01013	0	0	0.14378
SZ_42	0.74964	0	0	0	0
SZ_43	0.57503	0	0	0.01239	0.01213
SZ_44	2.72403	0	0	0	0.02986
SZ_45	4.64031	0	0	0	0
SZ_46	0.30439	0	0	0	0.04079
SZ_47	1.37871	0	0	0	0.04326
SZ_48	24.16883	0	0	0	0.28327
SZ_49	0.03877	0	0	0	0.04528
SZ_5	0.38088	0.00677	0.0017	0.06231	0
SZ_50	0.02375	0	0	0	0
SZ_51	0.17467	0	0	0.00019	0.07729
SZ_52	0.48988	0	0	0.45416	1.23334
SZ_53	3.62131	0	0	0	37.35146
SZ_54	0.15025	0	0	0	0.14307
SZ_55	0.10671	0	0	0	0.05699
SZ_56	1.45159	0	0	0	0.09488
SZ_57	4.52419	0	0	0	0.37746
SZ_58	0.19853	0	0	0	0.11344
SZ_59	0.71781	0	0	0	0.04792
SZ_6	0.18389	0.04895	0.00648	0	0.00765
SZ_60	22.55269	0	0	0	0.42554
SZ_61	2.13716	0	0	0	0.05143
SZ_62	0.00044	0	0	0.06211	0.28277
SZ_63	1.41915	0	0	0.19492	0.00111
SZ_64	0.03898	0	0	0.28672	0
SZ_65	2.42733	0	0	0	0.08683
SZ_66	0.86857	0	0	0	0.00292
SZ_67	0.05125	0	0	0	0.00885
SZ_68	4.11919	0	0	0	1.60514
SZ_69	5.75246	0	0	0	0.92507
SZ_7	0.07105	0.0544	0.00732	0.00013	0.04138
SZ_70	0.20781	0	0	0	0.2865
SZ_71	0.07558	0	0	0	0.20045
SZ_72	0.94241	0.0042	0	0	0.00617
SZ_73	3.23848	0.02332	0	0.14759	0.01386
SZ_74	8.17603	0.01255	0	0.00083	0.13691
SZ_75	0.27822	0.00723	0	0	0.53299
SZ_76	0.03545	0.0078	0	0.01093	0.08021
SZ_77	2.80253	0	0	0	0.03818
SZ_78	6.40691	0	0	0	0.00214
SZ_79	1.16785	0.0392	0	0	0
SZ_8	0.04302	0.03026	0.00367	0.72333	0.02513

SZ_80	1.34019	0	0	0.00205	0.60506
SZ_9	0.29642	0.0933	0.01439	0.00212	0

*SZ: 精神分裂患者; H: 健康对照

表 2 5 种生物标志物详细信息

标志物编号	模型预测中的重要性	显著富集方向	验证集 AUC(%)#	95% CI(%)	基因序列信息 (NCBI)
Veillonella_parvula	27.85785472	SZ	72	48.61-95.39	txid29466
Ruminococcus_lactaris	15.3045984	H	63.5	38.66-88.34	txid46228
Bacillus_cereus_thuringiensis	9.243517102	H	55	31-79	txid1428
Bacillus_nealsonii	15.04336932	SZ	54	33.88-74.12	txid115979
Bacteroides_vulgatus	10.51326795	SZ	51.5	24.48-78.52	txid821

#验证集 AUC, 表示在训练集数据得到模型下, 对验证集数据的判别程度。

&txid, 表示该生物标记物在 NCBI 数据库中的编号。

表 3 基于 5 种标志物组合训练集患病概率

样本编号	患病概率		样本编号	患病概率
H_1	0.88372093		SZ_1	0.769736842
H_10	0.066326531		SZ_10	0.853932584
H_11	0.316939891		SZ_11	0.861904762
H_12	0.047058824		SZ_12	0.552238806
H_13	0.351851852		SZ_13	0.869791667
H_14	0.640718563		SZ_14	0.683615819
H_15	0.357541899		SZ_15	0.659793814
H_16	0.201149425		SZ_16	0.899371069
H_17	0.72826087		SZ_17	0.758064516
H_18	0.423076923		SZ_18	0.945945946
H_19	0.630208333		SZ_19	0.85483871
H_2	0.702702703		SZ_2	0.838888889
H_20	0.298913043		SZ_20	0.936046512
H_21	0.333333333		SZ_21	0.893617021
H_22	0.165714286		SZ_22	0.645502646
H_23	0.171717172		SZ_23	0.85106383
H_24	0.743315508		SZ_24	0.662790698
H_25	0.032085561		SZ_25	0.938888889
H_26	0.062146893		SZ_26	0.45398773
H_27	0.172413793		SZ_27	0.350515464
H_28	0.059139785		SZ_28	0.728915663
H_29	0.518918919		SZ_29	0.905027933
H_3	0.83		SZ_3	0.679347826
H_30	0		SZ_30	0.95959596
H_31	0.135		SZ_31	0.288135593
H_32	0.078125		SZ_32	0.642487047
H_33	0.365		SZ_33	0.772486772
H_34	0.554285714		SZ_34	0.627027027

H_35	0.192708333		SZ_35	0.276243094
H_36	0.232044199		SZ_36	0.641176471
H_37	0.357894737		SZ_37	0.52173913
H_38	0.730964467		SZ_38	0.635359116
H_39	0.447852761		SZ_39	0.757575758
H_4	0.863157895		SZ_4	0.902298851
H_40	0.395480226		SZ_40	0.965517241
H_41	0.674285714		SZ_41	0.983425414
H_42	0.072625698		SZ_42	0.259803922
H_43	0.502793296		SZ_43	0.037837838
H_44	0.04519774		SZ_44	0.755319149
H_45	0.865979381		SZ_45	0.17989418
H_46	0.096590909		SZ_46	0.525139665
H_47	0.816216216		SZ_47	0.632124352
H_48	0.604519774		SZ_48	0.838323353
H_49	0.326732673		SZ_49	0.795454545
H_5	0.355191257		SZ_5	0.344632768
H_50	0.277777778		SZ_50	0.586592179
H_51	0.164835165		SZ_51	0.322580645
H_52	0.5		SZ_52	0.7
H_53	0.209039548		SZ_53	0.974489796
H_54	0.833333333		SZ_54	0.976047904
H_55	0.190954774		SZ_55	0.694444444
H_56	0.398773006		SZ_56	0.767045455
H_57	0.449438202		SZ_57	0.960199005
H_58	0.493902439		SZ_58	0.914438503
H_59	0.306451613		SZ_59	0.746987952
H_6	0.584158416		SZ_6	0.912568306
H_60	0.149171271		SZ_60	0.878787879
H_61	0.040462428		SZ_61	0.689473684
H_62	0.126903553		SZ_62	0.821989529
H_63	0.43814433		SZ_63	0.015544041
H_64	0.642458101		SZ_64	0.561497326
H_65	0.016666667		SZ_65	0.851282051
H_66	0.039325843		SZ_66	0.396449704
H_67	0.614035088		SZ_67	0.6375
H_68	0.053763441		SZ_68	0.964102564
H_69	0.016949153		SZ_69	0.920454545
H_7	0.029761905		SZ_7	0.947643979
H_70	0.157360406		SZ_70	0.921875
H_71	0.505434783		SZ_71	0.979274611
H_8	0.173184358		SZ_72	0.463687151
H_9	0.278947368		SZ_73	0.212435233
			SZ_74	0.417085427
			SZ_75	0.877906977

			SZ_76	0.715189873
			SZ_77	0.755102041
			SZ_78	0.052325581
			SZ_79	0.538860104
			SZ_8	0.857142857
			SZ_80	0.781609195
			SZ_9	0.886486486

表 4 随机森林模型验证集肠道标志物相对丰度数据

样本编号	<i>Bacteroides_vulgatus</i>	<i>Bacillus_cereus_thuringiensis</i>	<i>Bacillus_nealsonii</i>	<i>Ruminococcus_lactaris</i>	<i>Veillonella_parvula</i>
SZ_81	0.72437	0.06655	0.01188	2.58931	0
SZ_82	7.7121	0.07201	0.01357	0.04804	0.0063
H_72	0.93279	0.10383	0.01281	0.00094	0.0088
H_73	8.64522	0.33867	0.0446	0.00015	0.00523
SZ_83	1.89049	0.01231	0	0.13471	0.01031
SZ_84	4.45675	0.0831	0.00573	0.00479	0.00035
H_74	0.25003	0.00087	0	0.85931	0.27916
H_75	0.11235	0.01071	0	0.00065	5.39226
SZ_85	1.92825	0	0	0	0
SZ_86	0.94776	0	0	0	0.03645
SZ_87	0.11391	0	0	0	0.00085
SZ_88	3.39953	0	0	0.28245	0.31003
SZ_89	0.03365	0	0	0.08543	0.08512
H_76	38.91024	0	0	0	0.16138
SZ_90	0.99436	0	0	0.00106	0.13754
H_77	0.0782	0	0	0.0021	0.25356
H_78	0.04442	0	0	0	0
H_79	0.21019	0	0	0	0
H_80	5.8593	0	0	0	0.16698
H_81	63.83726	0	0	0	0.46716

表 5 随机森林模型验证集数据患病概率

样本编号	患病概率(基于 5 种标志物组合)	患病概率(基于标志物 <i>Veillonella_parvula</i>)	患病概率(基于标志物 <i>Ruminococcus_lactaris</i>)
H_72	0.902	0.896	0.096
H_73	0.636	0.446	0.266
H_74	0.716	1	0.654
H_75	0.91	1	0.114
H_76	0.826	0.752	1
H_77	0.888	1	0.892
H_78	0.758	0.296	1
H_79	0.308	0.296	1
H_80	0.888	0.75	1
H_81	0.922	1	1
SZ_81	0.692	0.296	0.004

SZ_82	0.488	0.158	0.002
SZ_83	0.228	0.052	0.874
SZ_84	0.788	0.298	0.668
SZ_85	0.182	0.296	1
SZ_86	0.848	0.348	1
SZ_87	0.416	0.314	1
SZ_88	0.834	1	0.738
SZ_89	0.71	0.954	0.094
SZ_90	0.696	0.916	0.134

以上结果表明，本发明公开的生物标志物具有较高的准确度和特异性，具有良好的开发为诊断方法的前景，从而为精神分裂症的患病风险评估、诊断、早期诊断，寻找潜在药物靶点提供依据。

因此，本发明提出以下应用：

所述的精神分裂症的小规模生物标志物组合作为检测靶点或检测目标在制备检测试剂盒中的应用。

所述的精神分裂症的小规模生物标志物组合作为靶点在筛选治疗和/或者预防精神分裂症的药物中的应用。

所述的生物标志物组合相对丰度的变化为确定候选药物是否有效提供依据。

尽管上面已经示出和描述了本发明的实施例，可以理解的是，上述实施例是示例性的，不能理解为对本发明的限制，本领域的普通技术人员在本发明的范围内可以对上述实施例进行变化、修改、替换和变型。

1、一种精神分裂症的小规模生物标志物组合，其特征在于，该生物标志物组合用于提供相对丰度信息，其包含选自以下的一种或多种：

生物标志物 1: *Veillonella*_parvula;

生物标志物 2: *Ruminococcus*_lactaris;

生物标志物 3: *Bacillus*_cereus_thuringiensis;

生物标志物 4: *Bacillus*_nealsonii;

生物标志物 5: *Bacteroides*_vulgatus。

2、如权利要求 1 所述的精神分裂症的小规模生物标志物组合，其特征在于，所述的生物标志物组合提供的相对丰度信息用于和参考值进行比较。

3、如权利要求 1 所述的精神分裂症的小规模生物标志物组合，其特征在于，所述的生物标志物 1~5 的相对丰度信息是基于能够对其进行丰度计算的基因序列来提供的。

4、权利要求 1 所述的精神分裂症的小规模生物标志物组合作为检测靶点或检测目标在制备检测试剂盒中的应用。

5、权利要求 1 所述的精神分裂症的小规模生物标志物组合作为靶点在筛选治疗和/或者预防精神分裂症的药物中的应用。

6、如权利要求 5 所述的应用，其特征在于，所述的生物标志物组合相对丰度的变化为确定候选药物是否有效提供依据。

7、权利要求 1 所述的精神分裂症的小规模生物标志物组合的筛选方法，其特征在于，步骤如下：

1) 样本收集：采集粪便样品后冷冻运输并迅速转移到-80℃保存，进行 DNA 提取，得到提取的 DNA 样本，样本受试者包括精神分裂症病人和健康人；

2) 宏基因组测序与组装

3) 将高质量的测序片段输入到软件 *Metaphlan2* 计算出物种的相对丰度：

3.1) 将高质量测序片段比对到参考的标记基因上；

3.2) 根据比对结果统计插入片段的数量；

3.3) 将插入片段的数量对标记基因的长度进行标准化得到对应的丰度；

4) 从样本集中随机地选取精神分裂症病人和健康人作为训练集，其余样品作为验证集，计算训练集中每个样本中物种的相对丰度，然后将训练集的物种输入随机森林分类器，对分类器进行 5 次 10 折交叉验证，利用 RF 模型筛选的物种相对丰度对每一个体计算其精神分裂症患病风险，绘制 ROC 曲线，并计算出 AUC 作为判别模型效能评价参数，选取标志物组合数<30，且判别效能最佳的组合，在模型中输出每个物种的重要性指数，重要性指数越高，代表该标志物用来判别精神分裂症和非精神分裂症的重要性越高。

8、权利要求 7 所述筛选方法，其特征在于，所述样本集中，样本受试者包括 90 个精神

分裂症病人和 81 个健康人,验证集中,样本受试者包括 10 个精神分裂症病人和 10 个健康人。

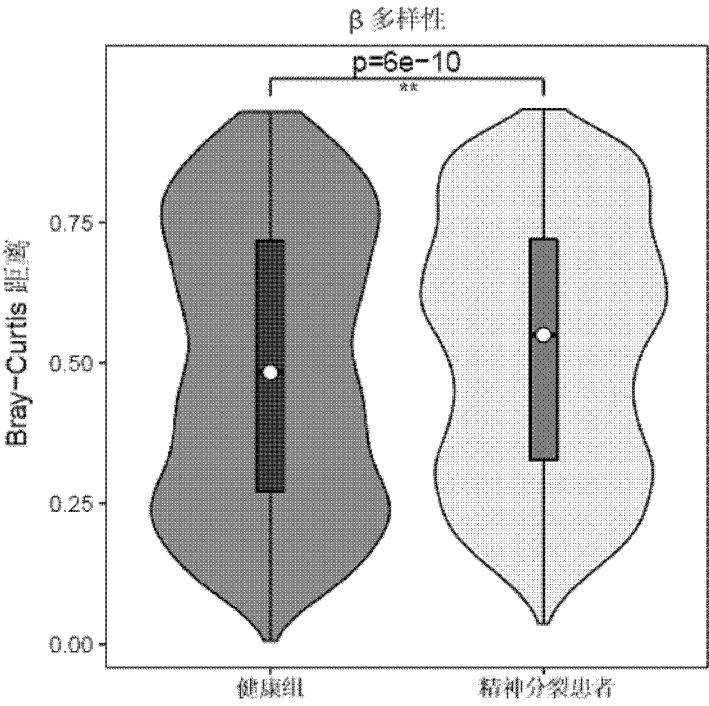


图 1

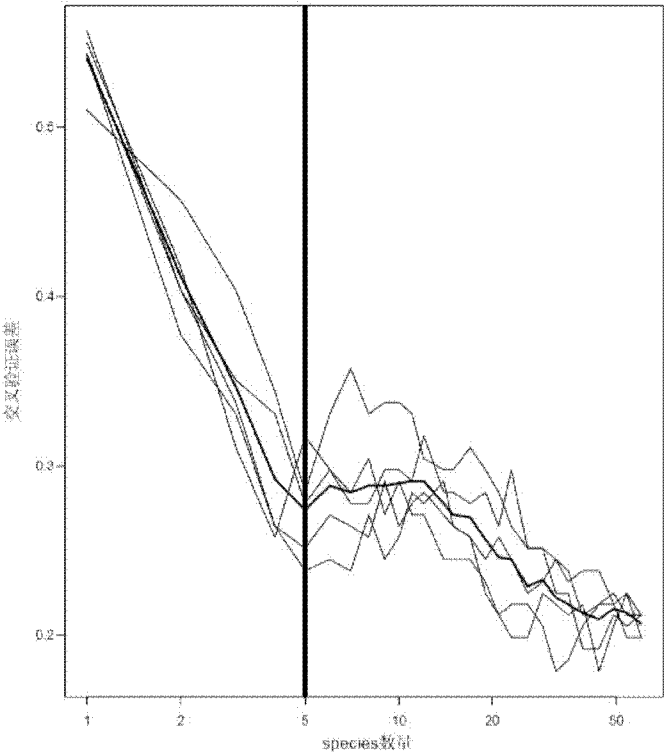


图 2

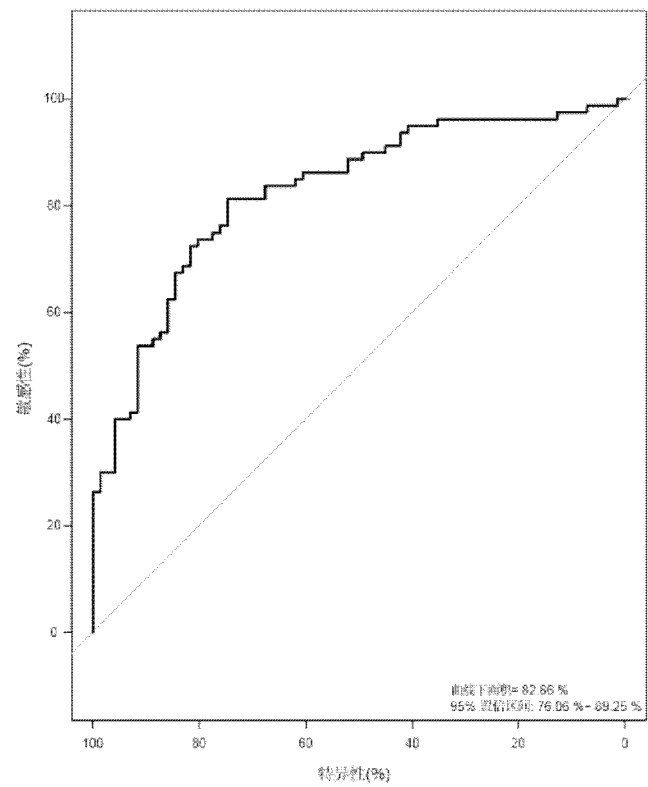


图 3

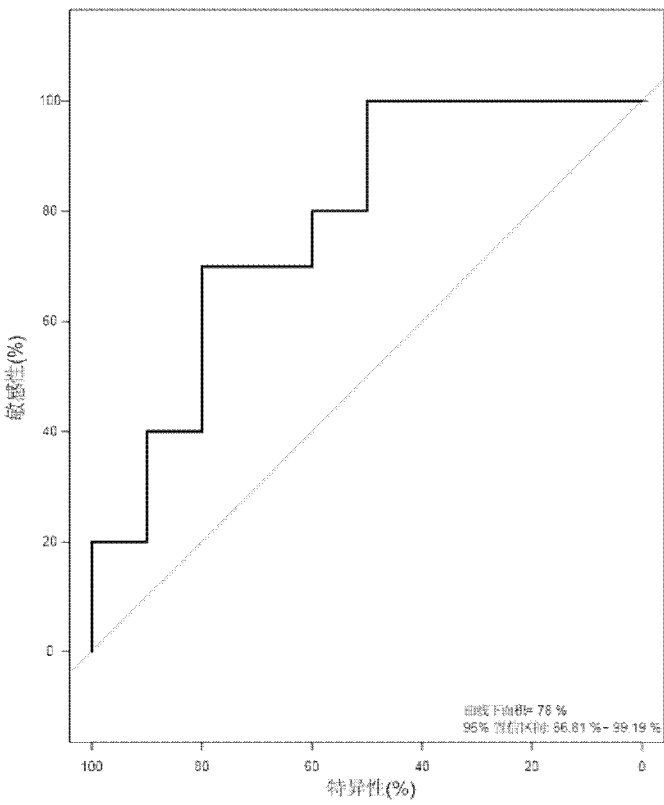


图 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/094924

A. CLASSIFICATION OF SUBJECT MATTER

C12Q 1/68(2018.01)i; A61P 25/00(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; A61P

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)

DWPI, SIPOABS, CNABS, CNKI, ISI Web of knowledge, NCBI, Google Scholar: 精神分裂, 精神病, 精神疾病, 标志物, 标记物, 菌, 微生物, 粪便, 芽孢, 拟杆菌, 胃癌, Schizophrenia, mental, marker, bacteria, microorganisms, feces, metaphlan, Veillonella_parvula, Ruminococcus_lactaris, Bacillus_cereus_thuringiensis, Bacillus_nealsonii, Bacteroides_vulgatus

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	赵星梅 等 (ZHAO, Xingmei et al.). "基于高通量测序的精神分裂症患者肠道菌群多样性 (Non-official translation: Gut Microbial Profile Analysis by High-Throughput Sequencing in Schizophrenia Patients)" 中国微生态学杂志 (Chinese Journal of Microecology), Vol. 31, No. 1, 31 January 2019 (2019-01-31), sections 1.2 and 2.2	1-8
A	CN 107746874 A (ZHANG, Meng; LIANG, Ying) 02 March 2018 (2018-03-02) entire document	1-8
A	CN 108474018 A (NATIONAL CENTER OF NEUROLOGY AND PSYCHIATRY; KK YAKULT HONSHA) 31 August 2018 (2018-08-31) entire document	1-8

☐ 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

23 August 2019

Date of mailing of the international search report

16 January 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Authorized officer

Facsimile No. (86-10)62019451

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/094924

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	107746874	A	02 March 2018	None			
CN	108474018	A	31 August 2018	US	2018100177	A1	12 April 2018
				EP	3284830	A1	21 February 2018
				EP	3284830	A4	19 December 2018
				WO	2016167365	A1	20 October 2016
				JP	WO2016167365	A1	15 February 2018

国际检索报告

国际申请号

PCT/CN2019/094924

A. 主题的分类 C12Q 1/68(2018.01)i; A61P 25/00(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; A61P 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)) DWPI、SIP0ABS、CNABS、CNKI、ISI Web of knowledge、NCBI、Google Scholar: 精神分裂、精神病、精神疾病、标志物、标记物、菌、微生物、粪便、芽孢、拟杆菌、胃瘤、Schizophrenia、mental、marker、bacteria、microorganisms、feces、metaphlan、Veillonella parvula、Ruminococcus lactaris、Bacillus cereus thuringiensis、Bacillus nealsonii、Bacteroides vulgatus														
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>赵星梅 等. “基于高通量测序的精神分裂症患者肠道菌群多样性” 中国微生态学杂志, 第31卷, 第1期, 2019年 1月 31日 (2019 - 01 - 31), 1.2、2.2节</td> <td>1-8</td> </tr> <tr> <td>A</td> <td>CN 107746874 A (张猛 梁英) 2018年 3月 2日 (2018 - 03 - 02) 全文</td> <td>1-8</td> </tr> <tr> <td>A</td> <td>CN 108474018 A (国立研究开发法人国立精神 神经医疗研究中心 株式会社益力多本社) 2018年 8月 31日 (2018 - 08 - 31) 全文</td> <td>1-8</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	赵星梅 等. “基于高通量测序的精神分裂症患者肠道菌群多样性” 中国微生态学杂志, 第31卷, 第1期, 2019年 1月 31日 (2019 - 01 - 31), 1.2、2.2节	1-8	A	CN 107746874 A (张猛 梁英) 2018年 3月 2日 (2018 - 03 - 02) 全文	1-8	A	CN 108474018 A (国立研究开发法人国立精神 神经医疗研究中心 株式会社益力多本社) 2018年 8月 31日 (2018 - 08 - 31) 全文	1-8
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求												
X	赵星梅 等. “基于高通量测序的精神分裂症患者肠道菌群多样性” 中国微生态学杂志, 第31卷, 第1期, 2019年 1月 31日 (2019 - 01 - 31), 1.2、2.2节	1-8												
A	CN 107746874 A (张猛 梁英) 2018年 3月 2日 (2018 - 03 - 02) 全文	1-8												
A	CN 108474018 A (国立研究开发法人国立精神 神经医疗研究中心 株式会社益力多本社) 2018年 8月 31日 (2018 - 08 - 31) 全文	1-8												
☐ 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 2019年 8月 23日		国际检索报告邮寄日期 2020年 1月 16日												
ISA/CN的名称和邮寄地址 中国国家知识产权局(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		授权官员 张起 电话号码 62411036												

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/094924

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	107746874	A	2018年 3月 2日	无			
CN	108474018	A	2018年 8月 31日	US	2018100177	A1	2018年 4月 12日
				EP	3284830	A1	2018年 2月 21日
				EP	3284830	A4	2018年 12月 19日
				WO	2016167365	A1	2016年 10月 20日
				JP	W02016167365	A1	2018年 2月 15日