



(12) 发明专利

(10) 授权公告号 CN 117649948 B

(45) 授权公告日 2024.05.10

(21) 申请号 202410115239.7

(22) 申请日 2024.01.29

(65) 同一申请的已公布的文献号  
申请公布号 CN 117649948 A

(43) 申请公布日 2024.03.05

(73) 专利权人 深圳市早知道科技有限公司  
地址 518000 广东省深圳市福田区沙头街  
道车公庙天安创新科技广场A1002

(72) 发明人 陈若言 李哲夫 李腾 王理中  
唐森威 陈钢

(74) 专利代理机构 深圳叁众知识产权代理事务  
所(普通合伙) 44434  
专利代理师 陈慧

(51) Int. Cl.

G16H 50/80 (2018.01)

G16H 15/00 (2018.01)

G16B 20/20 (2019.01)

G16B 20/30 (2019.01)

G16B 20/40 (2019.01)

G16B 40/00 (2019.01)

G16B 50/30 (2019.01)

(56) 对比文件

CN 115873947 A, 2023.03.31

CN 108009404 A, 2018.05.08

CN 116052903 A, 2023.05.02

CN 107256323 A, 2017.10.17

CN 107345248 A, 2017.11.14

CN 112331290 A, 2021.02.05

CN 112382337 A, 2021.02.19

CN 112553327 A, 2021.03.26

CN 116052757 A, 2023.05.02

卫园园;范晋楠;李亚蕊.ADD3基因  
rs17095355位点多态性与胆道闭锁相关性研究  
的Meta分析.临床小儿外科杂志.2020,(第06  
期),

范焯;张闻;闫书山;张鹏军;王平.中国人群  
XRCC3 Thr241Met基因多态性与肝癌易感性的荟  
萃分析.南京医科大学学报(自然科学版).2013,  
(第10期),

康振;雒艳萍;时春虎;田金徽;杨克虎;祝秉  
东.CD209-871A/G位点多态性与结核易感性的  
Meta分析.循证医学.2011,(第06期), (续)

审查员 李晓霞

权利要求书4页 说明书18页 附图3页

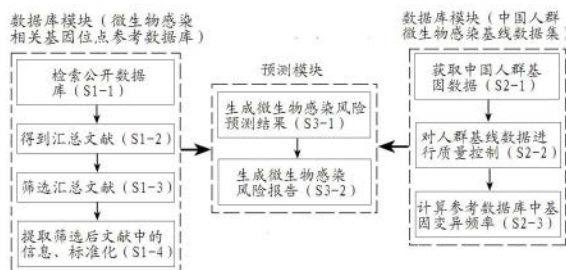
(54) 发明名称

一种基于基因检测的微生物感染风险预测  
方法及系统

(57) 摘要

本发明公开了一种基于基因检测的微生物  
感染风险预测方法及系统,该方法包括S1、通过  
公开文献获取微生物感染相关的标准化信息,包  
括微生物感染相关的变异名、病原微生物类型、  
病原微生物名、基因与病原微生物关联的风险类  
型、等位基因、与等位基因关联的表型、等位基因  
的遗传贡献OR值、参考文献编号,并构建微生物  
感染参考数据库;S2、获取中国人群的基因数据,  
并进行质量控制,建立中国人群微生物感染基线  
数据集;然后计算微生物感染相关的变异在中国  
人群中的频率;S3、获取用户的基因数据,对用户

进行病原微生物的感染风险预测,并生成微生物  
感染风险报告。本发明从多维度预测微生物感染  
风险,实现了及时、个体化的感染风险评估和预  
警。



CN 117649948 B

[接上页]

**(56) 对比文件**

张海燕;李红玉;李迅.HLA-Ⅱ类基因多态性与霍奇金淋巴瘤相关性的Meta分析.实用肿瘤杂志.2016,(第02期),

康振等.CD209-871A/G位点多态性与结核易感性的Meta分析.循证医学.2011,(06),第44-49页.

卫园园等.ADD3基因rs17095355位点多态

性与胆道闭锁相关性研究的Meta分析.临床小儿外科杂志.2020,(06),第34-40页.

范烨等.中国人群XRCC3 Thr241Met基因多态性与肝癌易感性的荟萃分析.南京医科大学学报(自然科学版).2013,(10),第152-156页.

张海燕等.HLA-Ⅱ类基因多态性与霍奇金淋巴瘤相关性的Meta分析.实用肿瘤杂志.2016,(02),第76-80页.

1. 一种基于基因检测的微生物感染风险预测方法,其特征在于,包括如下步骤:

S1、通过公开文献获取微生物感染相关的标准化信息,包括微生物感染相关的变异名、病原微生物类型、病原微生物名、基因与病原微生物关联的风险类型、等位基因、与等位基因关联的表型、等位基因的遗传贡献OR值、参考文献编号,并构建微生物感染参考数据库;

步骤S1中,所述与等位基因关联的表型包括风险较低、风险较高;

当等位基因的遗传贡献OR值小于等于1时,将与等位基因关联的表型定义为风险较低;

当等位基因的遗传贡献OR值大于1时,将与等位基因关联的表型定义为风险较高,并将对应的等位基因称作风险等位基因;

步骤S1中,所述基因与病原微生物关联的风险类型包括感染风险、感染严重程度、继发疾病遗传风险;

当公开文献记载由于个体遗传特征导致个体对病原微生物本身带来的易感性差别时,将基因与病原微生物关联的风险类型定义为感染风险;

当公开文献记载由于个体遗传特征导致个体在病原微生物感染发生后有严重程度差别时,将基因与病原微生物关联的风险类型定义为感染严重程度;

当公开文献记载由于个体遗传特征导致个体在病原微生物感染发生后有关联疾病发生倾向差别时,将基因与病原微生物关联的风险类型定义为继发疾病遗传风险;

S2、获取中国人群的基因数据,并进行质量控制,建立中国人群微生物感染基线数据集;然后根据中国人群微生物感染基线数据集,计算微生物感染相关的变异在中国人群中的频率;

S3、获取用户的基因数据,根据微生物感染参考数据库,对用户进行病原微生物感染风险预测,并生成微生物感染风险报告;

步骤S3中,所述对用户进行病原微生物感染风险预测,包括:

(一)对于一种基因与病原微生物关联的风险类型只对应一个等位基因的情况:

当用户携带与等位基因关联的表型为风险较低的等位基因,即用户没有携带风险等位基因时,则在某一种基因与病原微生物关联的风险类型下,预测结果为影响程度较低;

当用户携带与等位基因关联的表型为风险较高的等位基因,即用户携带了风险等位基因时,则在某一种基因与病原微生物关联的风险类型下,预测结果为影响程度较高;

(二)对于一种基因与病原微生物关联的风险类型对应两个及以上等位基因的情况:

首先,根据中国人群微生物感染基线数据集,计算人群平均携带风险等位基因数M,公式如下:

$$M = \sum_{i=1}^j 2p_i$$

式中,j表示与病原微生物的感染相关的j个变异; $p_i$ 表示第i个变异的风险等位基因在中国人群中的频率;

其次,将用户携带的风险等位基因数B与对应的中国人群平均携带风险等位基因数M进行比较:

当B大于M时,则在某一种基因与病原微生物关联的风险类型下,预测结果为影响程度

较高;

当B小于或等于M时,则在某一种基因与病原微生物关联的风险类型下,预测结果为影响程度较低。

2. 根据权利要求1所述的一种基于基因检测的微生物感染风险预测方法,其特征在于,步骤S1中,所述通过公开文献获取微生物感染相关的标准化信息,包括如下步骤:

S1-1、通过在公开数据库中检索微生物感染相关的主题词,得到微生物感染相关的检索文献;

S1-2、通过文献管理软件去除检索文献中的重复文献,得到去重文献;

S1-3、根据微生物感染参考数据库的纳入条件,通过人工阅读去重文献的摘要和正文内容,筛选出符合条件的纳入文献;

S1-4、提取纳入文献中的信息,并进行标准化数据格式处理。

3. 根据权利要求2所述的一种基于基因检测的微生物感染风险预测方法,其特征在于,步骤S1-3中,所述微生物感染参考数据库的纳入条件,包括:

1) 有明确标注样本的族裔信息;

2) 研究样本规模大于1000人;

3) 针对GWAS研究,基因位点的统计显著性p值小于 $5 \times 10^{-8}$ ;

针对候选基因研究,基因位点的统计显著性p值小于0.05;

4) 基因位点的连锁不平衡LD分析 $r^2$ 小于0.8。

4. 根据权利要求1所述的一种基于基因检测的微生物感染风险预测方法,其特征在于,步骤S2中,所述质量控制,包括:

a) 去除性别错误的个体;

b) 去除SNP缺失率大于0.05的个体;

c) 去除等位基因频率MAF小于0.01的SNP位点;

d) 去除哈迪温伯格平衡HWE检验的阈值小于 $1 \times 10^{-5}$ 的SNP位点。

5. 根据权利要求1所述的一种基于基因检测的微生物感染风险预测方法,其特征在于,步骤S3中,所述微生物感染风险报告包括:报告ID、用户ID、检测方式、报告日期、感染风险预测结果、感染严重程度预测结果、继发疾病遗传风险预测结果、预防建议、相关知识、检测细节和局限、参考文献;

其中,所述感染风险预测结果、感染严重程度预测结果、继发疾病遗传风险预测结果均包括影响程度和查看详情;所述影响程度为较低或较高;所述查看详情用于展示与某病原微生物预测结果相关的基因变异的详细信息,包括检测基因、检测位点、检测结果、变异类型、影响描述。

6. 一种基于基因检测预测微生物感染风险的分析系统,其特征在于,包括数据库构建模块、微生物感染风险预测模块、报告生成模块;

所述数据库构建模块,用于通过公开文献获取微生物感染相关的标准化信息,包括微生物感染相关的变异名、病原微生物类型、病原微生物名、基因与病原微生物关联的风险类型、等位基因、与等位基因关联的表型、等位基因的遗传贡献OR值、参考文献编号,并构建微生物感染参考数据库;

所述与等位基因关联的表型包括风险较低、风险较高;

当等位基因的遗传贡献OR值小于等于1时,将与等位基因关联的表型定义为风险较低;  
 当等位基因的遗传贡献OR值大于1时,将与等位基因关联的表型定义为风险较高,并将对应的等位基因称作风险等位基因;

所述基因与病原微生物关联的风险类型包括感染风险、感染严重程度、继发疾病遗传风险;

当公开文献记载由于个体遗传特征导致个体对病原微生物本身带来的易感性差别时,将基因与病原微生物关联的风险类型定义为感染风险;

当公开文献记载由于个体遗传特征导致个体在病原微生物感染发生后有严重程度差别时,将基因与病原微生物关联的风险类型定义为感染严重程度;

当公开文献记载由于个体遗传特征导致个体在病原微生物感染发生后有关联疾病发生倾向差别时,将基因与病原微生物关联的风险类型定义为继发疾病遗传风险;

所述微生物感染风险预测模块,用于获取中国人群的基因数据,并进行质量控制,建立中国人群微生物感染基线数据集;然后根据中国人群微生物感染基线数据集,计算微生物感染相关的变异在中国人群中的频率;

所述报告生成模块,用于获取用户的基因数据,根据微生物感染参考数据库,对用户进行病原微生物感染风险预测,并生成微生物感染风险报告;

所述对用户进行病原微生物感染风险预测,包括:

(一)对于一种基因与病原微生物关联的风险类型只对应一个等位基因的情况:

当用户携带与等位基因关联的表型为风险较低的等位基因,即用户没有携带风险等位基因时,则在某一种基因与病原微生物关联的风险类型下,预测结果为影响程度较低;

当用户携带与等位基因关联的表型为风险较高的等位基因,即用户携带了风险等位基因时,则在某一种基因与病原微生物关联的风险类型下,预测结果为影响程度较高;

(二)对于一种基因与病原微生物关联的风险类型对应两个及以上等位基因的情况:

首先,根据中国人群微生物感染基线数据集,计算人群平均携带风险等位基因数M,公式如下:

$$M = \sum_{i=1}^j 2p_i$$

式中,j表示与病原微生物的感染相关的j个变异; $p_i$ 表示第i个变异的风险等位基因在中国人群中的频率;

其次,将用户携带的风险等位基因数B与对应的中国人群平均携带风险等位基因数M进行比较:

当B大于M时,则在某一种基因与病原微生物关联的风险类型下,预测结果为影响程度较高;

当B小于或等于M时,则在某一种基因与病原微生物关联的风险类型下,预测结果为影响程度较低。

7.一种基于基因检测的微生物感染风险预测装置,其特征在于,包括客户端、服务器和计算设备,其中:

所述客户端用于连接服务器获得并查看微生物感染风险报告;

所述服务器分别连接客户端和计算设备,用于接收或发送信息;

所述计算设备用于接收并存储客户端发送的信息,或者向客户端发送信息,包括:处理器、通信器及存储器,其中:

所述处理器是计算设备的控制中心;

所述通信器是计算设备与服务器进行通信连接的通信中心;

所述存储器用于存储权利要求1-5任一项所述的一种基于基因检测的微生物感染风险预测方法的计算机软件程序,并由处理器进行加载和执行。

## 一种基于基因检测的微生物感染风险预测方法及系统

### 技术领域

[0001] 本发明涉及一种微生物感染风险预测方法及系统,尤其涉及一种基于基因检测的微生物感染风险预测方法、系统及装置,属于基因检测技术与微生物感染病学相结合的医学应用领域。

### 背景技术

[0002] 微生物感染是指身体受到病原微生物(比如细菌、病毒、真菌、寄生虫等)侵入并繁殖引起的疾病,已成为全球范围内的重要健康问题,对人类健康和社会经济造成了严重影响。具体来说,病原微生物可以通过各种途径进入人体,比如空气飞沫、接触感染、食物或水源、昆虫叮咬等方式,一旦病原微生物侵入人体,就会与宿主体的细胞或组织发生相互作用,从而导致病理过程的发生。个体对微生物感染的易感性和抵抗力存在明显的差异,差异的来源包括遗传因素、环境因素和行为习惯。

[0003] 传统的微生物感染风险评估一般是评估特定微生物对人群的危害鉴定、特征描述、暴露评估和风险特征等,无法从个体角度区分易感人群,更无法从感染风险、感染严重程度、继发疾病遗传风险等不同维度评估微生物感染的个体风险,这些局限制约了传染病的预防和个体化治疗。并且,预防和治疗微生物感染的方法主要基于广谱措施,缺乏个体化和精准的预测手段,这导致了许多挑战,包括过度和不必要的使用抗生素、治疗失败、医疗资源浪费以及感染的传播和暴发。因此,亟待需要寻找进行个体化和精准化预测的研究方向。

[0004] 研究表明,个体对微生物感染易感性和抵抗力(即“感染风险”)存在明显的个体遗传差异,比如:诺如病毒一种高致病、传染性极强的肠道病毒,可引起人急性胃肠炎。人群中存在一部分人可以免疫诺如病毒,他们被称为“非分泌者”。非分泌者携带 FUT2基因G428A(rs601338)、C571T(rs1800028)等纯合子变异,FUT2基因变异影响其产生 ABH 抗原,而ABH 抗原是诺如病毒的受体,因此非分泌者表现出对诺如病毒的抵抗能力,不易被感染。

[0005] 此外,个体的遗传因素还可以影响炎症反应和抗菌物质产生等方面的生理过程,从而影响微生物感染后的严重程度(比如:不携带 HLA-B\*27:05 单倍型感染艾滋病毒后有病毒贮积的风险,更容易积累病毒,影响其感染后的严重程度),以及可能导致的继发疾病风险(比如:携带rs2294008 CC基因型的人,感染幽门螺杆菌后发展为十二指肠溃疡(继发疾病)的遗传风险较高)。

[0006] 随着高通量基因检测技术的发展,该技术已被广泛应用于各种遗传特征的分析,使得个体层面的疾病风险预测成为可能。因此,利用基因检测技术快速获得个体的遗传特征信息,以实现微生物感染风险的评估,将成为进行个体化和精准化预测的研究方向。

### 发明内容

[0007] 针对上述现存的技术问题,本发明提供一种基于基因检测的微生物感染风险预测方法、系统及装置,通过从多维度预测微生物感染的风险,以期克服传统方法的局限,达到

更加及时、个体化的感染风险评估和预警的技术目的。

[0008] 为实现上述目的,首先,本发明提供一种基于基因检测的微生物感染风险预测方法,包括如下步骤:

[0009] S1、通过公开文献获取微生物感染相关的标准化信息,包括微生物感染相关的变异名、病原微生物类型、病原微生物名、基因与病原微生物关联的风险类型、等位基因、与等位基因关联的表型、等位基因的遗传贡献OR值、参考文献编号,并构建微生物感染参考数据库;

[0010] S2、获取中国人群的基因数据,并进行质量控制,建立中国人群微生物感染基线数据集;然后根据中国人群微生物感染基线数据集,计算微生物感染相关的变异在中国人群中的频率;

[0011] S3、获取用户的基因数据,根据微生物感染参考数据库,对用户进行病原微生物感染风险预测,并生成微生物感染风险报告。

[0012] 本发明方法进一步的,步骤S1中,所述通过公开文献获取微生物感染相关的标准化信息,包括如下步骤:

[0013] S1-1、通过在公开数据库中检索微生物感染相关的主题词,得到微生物感染相关的检索文献;

[0014] S1-2、通过文献管理软件去除检索文献中的重复文献,得到去重文献;

[0015] S1-3、根据微生物感染参考数据库的纳入条件,通过人工阅读去重文献的摘要和正文内容,筛选出符合条件的纳入文献;

[0016] S1-4、提取纳入文献中的信息,并进行标准化数据格式处理。

[0017] 本发明方法更进一步的,步骤S1-3中,所述微生物感染参考数据库的纳入条件,包括:

[0018] 1) 有明确标注样本的族裔信息;

[0019] 2) 研究样本规模大于1000人;

[0020] 3) 针对GWAS研究,基因位点的统计显著性p值小于 $5 \times 10^{-8}$ ;

[0021] 针对候选基因研究,基因位点的统计显著性p值小于0.05;

[0022] 4) 基因位点的连锁不平衡LD分析 $r^2$ 小于0.8。

[0023] 本发明方法进一步的,步骤S1中,所述与等位基因关联的表型包括风险较低、风险较高;

[0024] 当等位基因的遗传贡献OR值小于等于1时,将与等位基因关联的表型定义为风险较低;

[0025] 当等位基因的遗传贡献OR值大于1时,将与等位基因关联的表型定义为风险较高,并将对应的等位基因称作风险等位基因。

[0026] 本发明方法更进一步的,步骤S1中,所述基因与病原微生物关联的风险类型包括感染风险、感染严重程度、继发疾病遗传风险;

[0027] 当公开文献记载由于个体遗传特征导致个体对病原微生物本身带来的易感性差别时,将基因与病原微生物关联的风险类型定义为感染风险;

[0028] 当公开文献记载由于个体遗传特征导致个体在病原微生物感染发生后有严重程度差别时,将基因与病原微生物关联的风险类型定义为感染严重程度;



[0029] 当公开文献记载由于个体遗传特征导致个体在病原微生物感染发生后有关联疾病发生倾向差别时,将基因与病原微生物关联的风险类型定义为继发疾病遗传风险。

[0030] 本发明方法进一步的,步骤S2中,所述质量控制,包括:

[0031] a) 去除性别错误的个体;

[0032] b) 去除SNP缺失率大于0.05的个体;

[0033] c) 去除等位基因频率MAF小于0.01的SNP位点;

[0034] d) 去除哈迪温伯格平衡HWE检验的阈值小于 $1 \times 10^{-5}$ 的SNP位点。

[0035] 本发明方法更进一步的,步骤S3中,所述对用户进行病原微生物感染风险预测,包括:

[0036] (一)对于一种基因与病原微生物关联的风险类型只对应一个等位基因的情况:

[0037] 当用户携带与等位基因关联的表型为风险较低的等位基因,即用户没有携带风险等位基因时,则在某一种基因与病原微生物关联的风险类型下,预测结果为影响程度较低;

[0038] 当用户携带与等位基因关联的表型为风险较高的等位基因,即用户携带了风险等位基因时,则在某一种基因与病原微生物关联的风险类型下,预测结果为影响程度较高;

[0039] (二)对于一种基因与病原微生物关联的风险类型对应两个及以上等位基因的情况:

[0040] 首先,根据中国人群微生物感染基线数据集,计算人群平均携带风险等位基因数M,公式如下:

$$[0041] \quad M = \sum_i^j 2p_i$$

[0042] 式中,j表示与病原微生物的感染相关的j个变异; $p_i$ 表示第i个变异的风险等位基因在中国人群中的频率;

[0043] 其次,将用户携带的风险等位基因数B与对应的中国人群平均携带风险等位基因数M进行比较:

[0044] 当B大于M时,则在某一种基因与病原微生物关联的风险类型下,预测结果为影响程度较高;

[0045] 当B小于或等于M时,则在某一种基因与病原微生物关联的风险类型下,预测结果为影响程度较低。

[0046] 本发明方法更进一步的,步骤S3中,所述微生物感染风险报告,包括:报告ID、用户ID、检测方式、报告日期、感染风险预测结果、感染严重程度预测结果、继发疾病遗传风险预测结果、预防建议、相关知识、检测细节和局限、参考文献;

[0047] 其中,所述感染风险预测结果、感染严重程度预测结果、继发疾病遗传风险预测结果均包括影响程度和查看详情;所述影响程度为较低或较高;所述查看详情用于展示与某病原微生物预测结果相关的基因变异的详细信息,包括检测基因、检测位点、检测结果、变异类型、影响描述。

[0048] 其次,为实现上述目的,本发明提供一种基于基因检测预测微生物感染风险的分析系统,包括数据库构建模块、微生物感染风险预测模块、报告生成模块;

[0049] 所述数据库构建模块,用于通过公开文献获取微生物感染相关的标准化信息,包

括微生物感染相关的变异名、病原微生物类型、病原微生物名、基因与病原微生物关联的风险类型、等位基因、与等位基因关联的表型、等位基因的遗传贡献OR值、参考文献编号,并构建微生物感染参考数据库;

[0050] 所述微生物感染风险预测模块,用于获取中国人群的基因数据,并进行质量控制,建立中国人群微生物感染基线数据集;然后根据中国人群微生物感染基线数据集,计算微生物感染相关的变异在中国人群中的频率;

[0051] 所述报告生成模块,用于获取用户的基因数据,根据微生物感染参考数据库,对用户进行病原微生物感染风险预测,并生成微生物感染风险报告。

[0052] 再者,为实现上述目的,本发明提供一种基于基因检测的微生物感染风险预测装置,包括客户端、服务器和计算设备,其中:

[0053] 所述客户端用于连接服务器获得并查看微生物感染风险报告;

[0054] 所述服务器分别连接客户端和计算设备,用于接收或发送信息;

[0055] 所述计算设备用于接收并存储客户端发送的信息,或者向客户端发送信息,包括:处理器、通信器及存储器,其中:

[0056] 所述处理器是计算设备的控制中心;

[0057] 所述通信器是计算设备与服务器进行通信连接的通信中心;

[0058] 所述存储器用于存储所述的一种基于基因检测的微生物感染风险预测方法的计算机软件程序,并由处理器进行加载和执行。

[0059] 综上,本发明通过从多维度(包括感染风险、感染严重程度、继发疾病遗传风险)预测微生物感染的风险,实现了更及时、个体化的感染风险评估和预警,并具有如下有益效果和技术优势:

[0060] 1、本发明从遗传风险角度,对普通人群的多种微生物感染风险进行综合预测和分析,对易感人群起到预防警示作用,对个体是否采取进阶微生物感染筛查起到指导作用,以及可作为个体健康遗传咨询解读的重要组成部分。

[0061] 2、本发明利用现在的人口遗传资源及其遗传多样性,去寻找各种感染发病的分子机制,革新人类对微生物感染的认识,从而为个体抗感染预防、治疗或者是预防感染流行病学事件提供更多参考信息。

[0062] 3、本发明方法构建得到的分析系统简单易用,能够为临床防治感染提供参考和借鉴,具有较高的参考价值,填补了现有技术的空缺。

## 附图说明

[0063] 附图用来提供对本发明的进一步理解,并且构成说明书的一部分,与本发明的实施例一起用于解释本发明,并不构成对本发明的限制。

[0064] 图1为本发明基于基因检测的微生物感染风险预测方法的流程图;

[0065] 图2为本发明实施时“肝炎病毒”的微生物感染风险报告模版示意图;

[0066] 图3为本发明基于基因检测的微生物感染风险预测系统的原理框图;

[0067] 图4为本发明基于基因检测的微生物感染风险预测装置的原理框图。

## 具体实施方式

[0068] 以下结合附图对本发明的优选实施例进行说明,应当理解,此处所描述的优选实施例仅用于说明和解释本发明,并不用于限定本发明。

[0069] 除非在说明书中另有说明,否则说明书中使用的所有科学和技术术语具有与本发明所属技术领域的技术人员通常理解的含义相同的含义。

[0070] 本文所述术语“脱氧核糖核酸(DNA)”是传递遗传信息的重要载体,是一种由脱氧核苷酸组成的大分子聚合物。脱氧核苷酸由碱基、脱氧核糖和磷酸构成。其中碱基有四种形式,分别为腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),这四种形式组合形成了遗传信息的多样性。

[0071] 本文所述术语“基因型(Genotype)”是指个体在其基因(具有功能性的DNA序列)上的具体遗传组合方式。人类是二倍体生物,即个体从父母各继承一套染色体(一套染色体为23条)。因此,在同一基因位点上,两条DNA分子可能存在不同的碱基排列,这些不同的基因形式被称为等位基因。比如,对于某个SNP上,某个体遗传自母亲的序列为“C”(一种等位基因),而遗传自父亲的序列为“T”(另一种等位基因),则该个体的基因型结果是CT(一种基因型)。

[0072] 本文所述术语“单核苷酸多态性(SNP)”是指DNA序列中单个核苷酸发生的变异所引起的DNA序列多态性,即单碱基的变异。SNP是一种常见的遗传变异形式,约90%的人类基因变异是SNP类型。

[0073] 本文所述术语“单倍型(Haplotype)”是指位于同一条染色体上的多个SNP特定组合形成的DNA序列多样性,即多碱基的变异,这个组合通常作为一个整体传递给后代,用\*单倍型形式(如HLA-A\*02:01)作为整体来描述。Haplotype由SNP信息结合特定规则计算得到,也是常规的基因变异形式之一。Haplotype的基因型一般描述为“携带”、“未携带”两种类型。

[0074] 本文所述术语“感染风险”指个体暴露在特定病原体环境中发生感染的风险。本结论仅从遗传层面解析感染风险的个体差异。比如,研究表明,CCR5基因的rs333基因位点,如果个体携带了DD基因型(纯合子缺失型变异),会阻止细胞表达CCR5受体蛋白,进而阻断人类免疫缺陷病毒1型(HIV-1)进入CD4<sup>+</sup>T细胞的感染途径,这些个体对HIV-1的感染表现出显著的抵御能力,即携带rs333 DD基因型的个体感染风险相对低。

[0075] 本文所述术语“感染严重程度”指感染发生后,病原体导致病症发展的严重程度。人与人之间遗传信息的个体差异,可能造成不同的人感染同一种病原体后,有不同的病症进展,表现出不同的严重程度。比如,在急性感染乙肝病毒后自愈的人群中发现,他们身体内针对乙肝病毒的适应性T细胞免疫应答特别强,而慢性感染者体内针对乙肝病毒的适应性T细胞免疫应答很弱。这就说明,携带了编码免疫强化的适应性T细胞基因的人群,可能能够更好的抵御乙肝病毒感染。

[0076] 本文所述术语“继发疾病遗传风险”指感染发生后,病原体导致其他关联疾病发生的可能性。人与人之间遗传信息的个体差异,可能使个体对特定疾病更易感。比如,机体发生炎症时,促进炎症的肿瘤坏死因子 $\alpha$ (TNF- $\alpha$ )会激活NOX01蛋白,NOX01蛋白可参与产生活性氧(ROS),而ROS是一种组织损伤性分子,所引发的氧化性应激反应会使胃部细胞的DNA发生损伤、变异,导致肿瘤形成。幽门螺杆菌感染所引发的炎症也会产生ROS,从而增加胃部的

氧化性应激反应,提升癌变风险。

[0077] 本文所述术语“比值比(OR)”指通过比较病例组和对照组的暴露因素(本发明中为基因变异数据)来建立起暴露因素与疾病之间的联系,以及量化暴露因素对疾病的贡献。若OR值等于1,则说明暴露因素与疾病无显著关联,对疾病的发生不起作用;若OR值大于1,则说明暴露因素与疾病正相关,是疾病的危险因素,且OR值越大,关联程度越大;若OR值小于1,则说明暴露因素与疾病负相关,是疾病的保护因素。

[0078] 本文所述术语“参考数据库”指的是通过文献检索、整理、筛选、标准化之后,得到的关于微生物感染相关的基因变异及其对应表型解析相关的数据库。

[0079] 实施例1:本发明基于基因检测的微生物感染风险预测方法。

[0080] 如图1所示,本实施例提供一种基于基因检测的微生物感染风险预测方法,包括以下步骤:

[0081] S1、通过公开文献获取微生物感染相关的标准化信息,包括微生物感染相关的变异名、病原微生物类型、病原微生物名、基因与病原微生物关联的风险类型、等位基因、与等位基因关联的表型、等位基因的遗传贡献OR值、参考文献编号,并构建微生物感染参考数据库。

[0082] 具体实施时,本步骤S1中通过对公开文献进行检索、筛选、提取得到标准化信息,包括如下步骤:

[0083] S1-1、通过在公开数据库中检索微生物感染相关的主题词,得到微生物感染相关的检索文献。

[0084] 最终纳入微生物感染参考数据库的相关文献,主要是通过公开数据库中对微生物感染相关的主题词进行中英文检索,再根据特定条件筛选整理后得到的。

[0085] 举例来说,所述公开数据库包括PubMed、Google scholar、中国知网数据库、万方数据库、维普数据库等。

[0086] 检索方式包括:对PubMed数据库、中国知网数据库、万方数据库、维普数据库分别使用MeSH医学主题词检索、布尔逻辑运算符连接词、自由词结合的方式进行检索。

[0087] 检索主题词包括:“感染+风险”、“传染性疾病+风险”、“感染+遗传+易感性”、“遗传+感染+风险”、“遗传+感染+程度”、“遗传+感染+引发+疾病”、“遗传+感染+并发症”、“polymorphism\*+infectious disease\*”、“genetic variation\*+infectious disease\*”、“GWAS+infectious disease\*”、“predisposition to infection\*”、“risk+infection\*”、“severity+infection\*”、“risk+infection\*+secondary disease\*”、“infection+comorbidity”、“传染性疾病”、“infectious disease”、“传染性疾病的治疗”、“response to treatment”、“传染性疾病的预防”、“response to vaccine”。

[0088] S1-2、通过文献管理软件去除检索文献中的重复文献,得到去重文献。

[0089] 具体实施时,将步骤S1-1得到的检索文献导入文献管理软件EndNote X9中,根据研究者姓名、机构、杂志期刊名、发表时间等信息去除重复文献,并对剩余文献进行编号汇总,得到去重文件。

[0090] S1-3、根据微生物感染参考数据库的纳入条件,通过人工阅读去重文献的摘要和正文内容,筛选出符合条件的纳入文献。

[0091] 所述微生物感染参考数据库的纳入条件包括:1)有明确标注样本的族裔信息;2)

研究样本规模大于1000人;3) 针对GWAS研究,基因位点的统计显著性p值小于 $5 \times 10^{-8}$ ;针对候选基因研究,p值小于0.05;4) 基因位点的连锁不平衡LD分析 $r^2$ 小于0.8。

[0092] 具体实施时,由2名经过专业培训的研究者各自独立对汇总文献的题目和摘要进行阅读,从而初步筛选出符合参考数据库纳入条件的文献;再各自独立对初步筛选的文献全文进行阅读,从而二次筛选出符合参考数据库纳入条件的文献,得到符合条件的纳入文献。并且,所述符合参考数据库纳入条件的文献需要被2名经过专业培训的研究者同时认同才会被采用,对于有分歧的文献需要通过2名经过专业培训的研究者讨论决定是否纳入,或由第3名经过专业培训的研究者决定是否纳入。

[0093] S1-4、提取纳文献中的信息,并进行标准化数据格式处理,从而获取微生物感染相关的标准化信息。

[0094] 具体实施时,所述标准化信息,包括:微生物感染相关的变异名、病原微生物类型、病原微生物名、基因与病原微生物关联的风险类型、等位基因、与等位基因关联的表型、等位基因的遗传贡献OR值、参考文献编号等。

[0095] S1-5、经过上述步骤S1-1至步骤S1-4后,利用获取微生物感染相关的标准化信息,即可构建微生物感染参考数据库,示例见表3。

[0096] 所述微生物感染参考数据库,包括:(1) 微生物感染相关的基因变异信息;(2) 基因变异的遗传贡献;(3) 基因变异影响的微生物感染风险表型信息;(4) 对应文献的来源。具体介绍如下。

[0097] (1) 微生物感染相关的基因变异信息,一般包括基因名、变异名、染色体信息、参考基因组物理位置、参考基因组序列、变异序列、参考基因组版本,但是由于“染色体信息、参考基因组物理位置、参考基因组序列、变异序列”会随不同文献使用的不同参考基因组版本变化,因此本实施例中,对微生物感染相关的基因变异信息使用“变异名”作为统一识别符,方便跨基因组版本使用。

[0098] 需要说明的是,基因变异包括两类:SNP变异和Haplotype变异,SNP变异一般使用dbSNP数据库中的rsid来表示,Haplotype变异一般使用\*单倍型的形式来表示。本实施例中,微生物感染参考数据库包括119个微生物感染相关的变异,分别为86个SNP变异,如表1所示,以及33个Haplotype变异,如表2所示。

[0099] 表1

[0100]

rs2858318	rs1012068	rs9302752	rs4720118	rs4951074	rs4282438
rs9274748	rs3219487	rs3762318	rs55894533	rs8176719	rs9277952
rs74597329	rs557011	rs2275606	rs6807915	rs372091	rs3117027
rs12614	rs9272785	rs2058660	rs142179458	rs601338	rs9271898
rs422951	rs877356	rs6871626	rs3764147	rs1047781	rs2516448
rs378352	rs4331426	rs2735591	rs3095309	rs1188447	rs3130196
rs1883832	rs12437118	rs2221593	rs2844633	rs11896375	rs115625939
rs2856718	rs6114027	rs73058713	rs2394945	rs333	rs6107516
rs7453920	rs2057178	rs10817758	rs58994923	rs5968255	rs6951643
rs3077	rs4733781	rs58600253	rs4845320	rs9357152	rs4921542
rs12979860	rs4921437	rs663743	rs10900585	rs59661306	rs7565981
rs7770370	rs2269497	rs77061563	rs2365860	rs7457728	rs5743604
rs743862	rs41553512	rs35829419	rs10900589	rs13117307	rs10004195
rs3129293	rs4240897	rs10100465	rs2365858	rs8067378	rs368433
rs2294008	rs2671655	-	-	-	-

[0101] 表2

[0102]

HLA-A*02:01	HLA-B*27:02	HLA-B*57:03	HLA-C*07:02
HLA-DQB1*03:03	HLA-DRB1*04:01	HLA-A*03:01	HLA-B*27:05
HLA-B*58:01	HLA-DQB1*02:01	HLA-DQB1*04:01	HLA-DRB1*04:03
HLA-A*11:01	HLA-B*35:01	HLA-B*58:02	HLA-DQB1*02:01
HLA-DQB1*06:02	HLA-DRB1*09:01	HLA-A*30:01	HLA-B*57:01
HLA-C*01:02	HLA-DQB1*03:01	HLA-DRB1*01:01	HLA-DRB1*11:01
HLA-A*33:01	HLA-B*57:01	HLA-C*06:02	HLA-DQB1*03:01
HLA-DRB1*03:01	HLA-DRB1*11:02	HLA-DRB1*12:01	HLA-DRB1*13:01
HLA-DRB1*15:01	-	-	-

[0103] (2) 基因变异的遗传贡献,包括具体的效应基因型或效应等位基因,及其对应的比值比OR。实施时,可以用等位基因、等位基因的遗传贡献OR值、与等位基因关联的表型来表示。

[0104] 需要说明的是,与等位基因关联的表型是一种影响程度的描述,取决于等位基因的遗传贡献OR值。当等位基因的遗传贡献OR值为小于等于1时,定义与等位基因关联的表型为风险较低;当等位基因的遗传贡献OR值大于1时,定义与等位基因关联的表型为风险较高,并将对应的等位基因称作风险等位基因。

[0105] 比如表3中rs1883832这个SNP位点上,等位基因C的遗传贡献OR值为1。当SNP位点携带了等位基因C时,表明该SNP位点影响的表型(乙型肝炎病毒感染风险)相关风险(影响)较低。

[0106] 尤其是,把可以提升某个风险类型的等位基因称作风险等位基因。比如表3中rs1883832这个SNP位点上,T就是一种风险等位基因。该风险等位基因T的遗传贡献OR值为1.19,当SNP位点携带了风险等位基因T时,表明该SNP位点影响的表型(乙型肝炎病毒感染风险)相关风险较高。

[0107] 表3

[0108]

变异名	病原微生物类型	病原微生物名	风险类型	等位基因	表型	OR	参考文献编号
rs1883832	病毒	乙型肝炎病毒	感染风险	C	风险较低	1	20
rs1883832	病毒	乙型肝炎病毒	感染风险	T	风险较高	1.19	20
HLA-DRB1*03:01	病毒	丙型肝炎病毒	感染严重程度	携带	严重程度较高	1.78	21
HLA-DRB*03:01	病毒	丙型肝炎病毒	感染严重程度	不携带	严重程度正常	1	21
rs2294008	细菌	幽门螺杆菌	继发疾病遗传风险	T	感染后发展为十二指肠溃疡的遗传风险较低	1	22
rs2294008	细菌	幽门螺杆菌	继发疾病遗传风险	C	感染后发展为十二指肠溃疡的遗传风险较高	1.41	22

[0109] (3) 基因变异影响的微生物感染风险信息,包括病原微生物类型、病原微生物名、基因与病原微生物关联的风险类型。

[0110] 本实施例中,微生物感染参考数据库包括14种病原微生物,分别为丙型肝炎病毒、乙型肝炎病毒、结核分枝杆菌、麻风分枝杆菌、溶组织痢疾阿米巴虫、念珠菌、恶性疟原虫、诺如病毒、人类免疫缺陷病毒1型、人类乳头瘤病毒、男性人类乳头瘤病毒、朊病毒、隐球菌、幽门螺杆菌,以及这14种病原微生物所属的病原微生物类型,包括病毒、细菌、真菌、寄生虫、其他。

[0111] 需要说明的是,基因与病原微生物关联的风险类型的划分不是必须的,或者本领域技术人员可以自行根据文献研究内容归纳定义风险类型。

[0112] 在本实施例中,将基因与病原微生物关联的风险类型分为感染风险、感染严重程度、继发疾病遗传风险这三个维度,这三个维度是由文献的研究内容总结,并结合实际进行风险预测后的干预方向归纳而来。

[0113] 具体来说,若某些文献研究的是由于个体遗传特征导致个体对病原微生物A本身带来的易感风险,则将基因与病原微生物关联的风险类型定义为感染风险。

[0114] 若某些文献研究中由于个体遗传特征导致个体在病原微生物A感染发生后会造成



严重后果(如发展成为慢性感染、病原微生物更容易积累等),则将基因与病原微生物关联的风险类型定义为感染严重程度。

[0115] 若某些文献研究侧重于感染病原微生物A后会引发B疾病(如感染幽门螺杆菌后发展为十二指肠溃疡),则将基因与病原微生物关联的风险类型定义为继发疾病遗传风险。

[0116] 如此一来,即可根据文献的研究内容归纳分类出感染风险、感染严重程度、继发疾病遗传风险这三个维度。但是,受限于目前公开文献的研究进展,并非所有病原微生物都有“感染风险”、“感染严重程度”、“继发疾病遗传风险”三个维度的研究证据,因此预测结果可能出现:一个维度的预测、二个维度的预测、三个维度的预测。也就是说,三个维度是基因与病原微生物关联的风险类型的细分内容,实际某种病原微生物可能只有一条证据,对应其中的某一维度的解读结果;或者有多条证据,对应其中一个或者多个维度的解读。

[0117] (4)对应文献的来源,实施时可以用参考文献编号来表示。

[0118] S2、获取中国人群的基因数据,并进行质量控制,建立中国人群微生物感染基线数据集。

[0119] 需要说明的是,由于微生物感染参考数据库是通过步骤S1检索筛选公开文献得到,造成涉及的人群样本多样(包括非洲人群、欧洲人群、亚洲人群等),而不同人群具有不同的遗传特征,因此在中国人群中应用参考数据库时,还需要基于中国人群的数据进行数据建模,从而建立一套适合中国人群特征的微生物感染风险预测和解读模型,具体步骤如下。

[0120] S2-1、获取中国人群的基因数据作为人群基线数据。

[0121] 本领域技术人员可以根据实际情况自行招募中国人群样本。本实施例中,中国人群样本招募自深圳市早知道科技有限公司的用户数据库,并从中抽取10万例用户的基因数据作为人群基线数据。需要说明的是,深圳市早知道科技有限公司通过基因检测为用户提供个人基因组解读服务。在日常服务过程中,用户均已经通过Illumina基因芯片完成自身基因数据的检测,因此用户数据库中的基因分型和单倍型数据可以用于人群基线数据的构建。

[0122] S2-2、对人群基线数据进行质量控制,得到中国人群微生物感染基线数据集。

[0123] 具体实施时,所述质量控制,包括如下内容:

[0124] a)去除性别错误的个体:根据X染色体杂合率进行遗传性别计算,与用户自我汇报的性别进行比较,去除遗传性别与自我汇报性别不一致的样本。由于性别错误可能意味着存在数据污染或者特殊的性别情况,因此本步骤可以防止性别错误对分析结果产生影响。

[0125] b)去除SNP缺失率过高的个体:设置SNP缺失率的阈值为0.05,表示对于缺失率超过0.05的个体进行删除。由于较高的SNP缺失率可能提示DNA质量和基因分型质量不佳,因此本步骤可以防止SNP缺失率过高对分析结果产生影响。

[0126] c)去除等位基因频率(MAF)较小的SNP位点:设置MAF的阈值为0.01,即删除MAF低于0.01的SNP位点。由于较小的MAF可能限制了位点的统计分析能力,因此本步骤可以防止MAF较小对分析结果产生影响。

[0127] d)去除偏离哈迪温伯格平衡(HWE)的SNP位点:设置HWE检验的阈值为 $1 \times 10^{-5}$ ,即删除HWE检验中小于 $1 \times 10^{-5}$ 的SNP位点。HWE是群体遗传学概念,描述了在一个稳态遗传系统中,基因型分布是否符合预期的遗传平衡状态。由于偏离HWE的SNP位点可能反映出基因测

量偏差或者与特定表型相关,因此本步骤可以防止偏离HWE对分析结果产生影响。

[0128] S2-3、根据中国人群微生物感染基线数据集,计算微生物感染参考数据库中,微生物感染相关的变异在中国人群中的频率,即SNP变异的频率,以及Haplotype变异的频率。

[0129] 由于步骤 S1 得到的微生物感染参考数据库只包含位点信息,不包含人群频率信息,这里使用步骤S2-1、S2-2步骤得到的中国人群微生物感染基线数据集,计算微生物感染参考数据库包含的位点的相关频率。

[0130] 本实施例中,经过计算得到86个SNP变异的频率,见表4所示,以及33个Haplotype变异的频率,见表5所示。其中,“A、T、C、G、D、I”列分别对应不同等位基因的频率,“-”表示频率为0。

[0131] 表4

[0132]

变异名	A	T	C	G	D	I
rs10004195	0.53	0.47	-	-	-	-
rs10100465	0.263	-	-	0.737	-	-
rs1012068	-	0.75	-	0.25	-	-
rs1047781	0.553	0.447	-	-	-	-
rs10817758	-	0.399	0.601	-	-	-
rs10900585	-	0.982	-	0.018	-	-
rs10900589	0.002	0.998	-	-	-	-
rs115625939	0.854	-	-	0.146	-	-
rs11884476	-	-	0.992	0.008	-	-
rs11896375	0	-	-	1	-	-
rs12437118	0.303	-	-	0.697	-	-
rs12614	-	0.068	0.932	-	-	-
rs12979860	-	0.059	0.941	-	-	-
rs13117307	-	0.12	0.88	-	-	-
rs142179458	0.004	-	-	0.996	-	-
rs1883832	-	0.378	0.622	-	-	-
rs2057178	0.049	-	-	0.951	-	-
rs2058660	0.509	-	-	0.491	-	-
rs2221593	-	0.199	0.801	-	-	-
rs2269497	0.937	-	-	0.063	-	-
rs2275606	0.225	-	-	0.775	-	-
rs2294008	-	0.267	0.733	-	-	-
rs2365858	-	-	0.001	0.999	-	-
rs2365860	-	0.999	-	0.001	-	-
rs2394945	-	-	0.771	0.229	-	-
rs2516448	-	0.243	0.757	-	-	-
rs2671655	-	0.657	0.343	-	-	-
rs2735591	0.287	-	-	0.713	-	-

rs2844633	-	0.314	0.686	-	-	-
rs2856718	-	0.542	0.458	-	-	-
rs2858318	-	0.214	-	0.786	-	-
rs3077	0.35	-	-	0.65	-	-
rs3095309	-	0.277	0.723	-	-	-
rs3117027	0.209	-	0.791	-	-	-
rs3129293	0.523	-	-	0.477	-	-
rs3130196	-	0.941	0.059	-	-	-
rs3219487	-	0.152	0.848	-	-	-
rs333	-	-	-	-	0.001	0.999
rs35829419	0.001	-	0.999	-	-	-
rs368433	-	0.999	0.001	-	-	-
rs372091	0	-	-	1	-	-
rs3762318	0.901	-	-	0.099	-	-
rs3764147	0.687	-	-	0.313	-	-
rs378352	0.382	-	-	0.618	-	-
rs41553512	-	0.013	0.987	-	-	-
rs422951	-	0.785	0.215	-	-	-
rs4240897	0.497	-	-	0.503	-	-
rs4282438	-	0.617	-	0.383	-	-
rs4331426	0.97	-	-	0.03	-	-
rs4720118	-	0.316	0.684	-	-	-
rs4733781	0.383	-	0.617	-	-	-
rs4845320	0.843	-	0.157	-	-	-
rs4921437	-	0.003	0.997	-	-	-
rs4921542	-	0.285	-	0.715	-	-
rs4951074	0.001	-	-	0.999	-	-
rs557011	-	0.493	0.507	-	-	-
rs55894533	0.573	-	0.427	-	-	-
rs5743604	0.437	-	-	0.563	-	-
rs58600253	-	0.149	0.851	-	-	-
rs58994923	-	-	-	-	0.081	0.919
rs59661306	0.866	-	-	0.134	-	-
rs5968255	-	0.494	0.506	-	-	-
rs601338	0.008	-	-	0.992	-	-
rs6107516	0.019	-	-	0.981	-	-
rs6114027	-	0.812	0.188	-	-	-
rs663743	0.159	-	-	0.841	-	-
rs6807915	-	0.512	0.488	-	-	-

rs6871626	0.315	-	0.685	-	-	-
rs6951643	0.926	-	-	0.074	-	-
rs73058713	0.143	-	0.857	-	-	-
rs743862	-	0.9	0.1	-	-	-
rs7453920	0.128	-	-	0.872	-	-
rs7457728	-	-	0.36	0.64	-	-
rs74597329	-	0.942	-	0.058	-	-
rs7565981	0.515	-	-	0.485	-	-
rs77061563	-	0.366	0.634	-	-	-
rs7770370	0.487	-	-	0.513	-	-
rs8067378	0.726	-	-	0.274	-	-
rs8176719	-	-	-	-	0.566	0.434
rs877356	-	0.427	0.573	-	-	-
rs9271898	0.368	-	-	0.632	-	-
rs9272785	0.283	-	-	0.717	-	-
rs9274748	0.684	-	-	0.316	-	-
rs9277952	0.42	-	-	0.58	-	-
rs9302752	-	0.711	0.289	-	-	-
rs9357152	0.638	-	-	0.362	-	-

[0133] 表5

[0134]

变异名	频率
HLA-A*02:01	0.1201
HLA-A*11:01	0.2093
HLA-A*03:01	0.0296
HLA-A*30:01	0.0591
HLA-A*33:01	0.0019
HLA-B*57:01	0.0117
HLA-B*57:03	0.0001
HLA-B*58:01	0.0613
HLA-B*58:02	0
HLA-B*35:01	0.0244
HLA-B*27:05	0.0072
HLA-B*27:02	0.0002
HLA-C*06:02	0.0885
HLA-C*07:02	0.1519
HLA-C*01:02	0.1589
HLA-DQB1*02:01	0.0493
HLA-DQB1*03:01	0.2109
HLA-DQB1*03:03	0.1588

HLA-DQB1*04:01	0.045
HLA-DQB1*06:02	0.0762
HLA-DRB1*03:01	0.0511
HLA-DRB1*11:02	0
HLA-DRB1*13:01	0.0146
HLA-DRB1*01:01	0.0206
HLA-DRB1*04:01	0.01
HLA-DRB1*11:01	0.0564
HLA-DRB1*12:01	0.0243
HLA-DRB1*09:01	0.1474
HLA-DRB1*15:01	0.116
HLA-DRB1*04:03	0.0156

[0135] S3、获取用户的基因数据,根据微生物感染参考数据库,对用户进行病原微生物感染风险预测,并生成微生物感染风险报告。

[0136] S3-1、获取用户的基因数据,根据微生物感染参考数据库,对用户进行某个病原微生物感染风险预测,得到预测结果。

[0137] 具体实施时,所述对用户进行病原微生物感染风险预测分为两种情况,包括:

[0138] (一)对于一种基因与病原微生物关联的风险类型只对应一个等位基因的情况。

[0139] 需要说明的是,具体的情况是变化的,并非每一种病原微生物都有“感染风险”、“感染严重程度”、“继发疾病遗传风险”三个风险类型维度的解读。

[0140] 对于某种病原微生物的某一个风险类型维度下,与之相关联的变异的数量为一个时,该变异的等位基因携带情况决定预测结果。换言之,对于某种病原微生物的某表型只有一个变异与之相关的时候,预测结果取决于用户是否携带该变异的风险等位基因。

[0141] 具体实施时,当用户携带与等位基因关联的表型为风险较低的等位基因,即用户没有携带风险等位基因时,则在某一种基因与病原微生物关联的风险类型(“感染风险”、“感染严重程度”、“继发疾病遗传风险”)下,预测结果为影响程度较低。

[0142] 当用户携带与等位基因关联的表型为风险较高的等位基因,即用户携带了风险等位基因时,则在某种病原微生物的某一个风险类型维度(“感染风险”、“感染严重程度”、“继发疾病遗传风险”)下,预测结果为影响程度较高。

[0143] 以“念珠菌”的“感染风险”为例,在微生物感染参考数据库中,“念珠菌”只有1个微生物感染相关的变异,即SNP变异rs4845320,见表6所示。

[0144] 表6

变异名	病原微生物类型	病原微生物名	风险类型	等位基因	表型	OR	参考文献编号
rs4845320	真菌	念珠菌	感染风险	A	风险较低	1	16
rs4845320	真菌	念珠菌	感染风险	C	风险较高	4.25	16

[0146] 由表6可知,和等位基因A相比,等位基因C提升了风险类型,可以视作一种风险等

位基因。当用户检测结果为AA时,表示其没有携带“较高”风险的风险等位基因,对表型的影响为风险较低,因此,预测念珠菌的感染风险较低;当用户检测结果为AC或CC时,表示其携带了“较高”风险的风险等位基因,对表型的影响为风险较高,因此,该用户念珠菌的感染风险较高。

[0147] (二)对于一种基因与病原微生物关联的风险类型对应两个及以上等位基因的情况:

[0148] 在本实施例中,对于某种病原微生物的某一个风险类型维度(“感染风险”、“感染严重程度”、“继发疾病遗传风险”)下,与之关联的变异数量为多个(两个及以上)的情况,需要综合多个位点的遗传贡献决定预测结果。

[0149] 首先,根据步骤S2建立的中国人群微生物感染基线数据集,计算人群平均携带风险等位基因数M,公式如下所示:

$$[0150] \quad M = \sum_i^j 2p_i$$

[0151] 式中,j表示与某个病原微生物的感染风险有关的j个变异; $p_i$ 表示第i个变异的风险等位基因在中国人群中的频率;2表示一对等位基因。

[0152] 其次,将用户携带的风险等位基因数B与对应的中国人群平均携带风险等位基因数M进行比较:

[0153] 当B大于M时,表明用户携带的风险等位基因数大于中国人群的平均水平,则预测结果为:某一个风险类型维度(“感染风险”、“感染严重程度”、“继发疾病遗传风险”)下,影响程度“较高”。

[0154] 当B小于或等于M时,表明用户携带的风险等位基因数小于或等于中国人群的平均水平,则预测结果为:某一个风险类型维度(“感染风险”、“感染严重程度”、“继发疾病遗传风险”)下,影响程度“较低”。

[0155] 换而言之,对于某个与等位基因关联的表型有2个及以上的变异时,需要通过比较用户携带的风险等位基因数量与中国人群平均携带的风险等位基因数量,才能生成预测结果。

[0156] 以“人类乳头瘤病毒”的“感染风险”为例,在步骤S1所得的微生物感染参考数据库中,“人类乳头瘤病毒”有5个变异,分别为3个SNP变异(rs9357152、rs59661306、rs7457728),以及2个Haplotype变异(HLA-DRB1\*04:03、HLA-DQB1\*06:02),如表7所示。

[0157] 表7

[0158]

变异名	病原微生物类型	病原微生物名	风险类型	等位基因	表型	OR	参考文献编号
rs9357152	病毒	人类乳头瘤病毒	感染风险	A	风险较低	1	14
rs9357152	病毒	人类乳头瘤病毒	感染风险	G	风险较高	1.37	14
rs59661306	病毒	人类乳头瘤病毒	感染风险	G	风险较高	1.2	13
rs59661306	病毒	人类乳头瘤病毒	感染风险	A	风险较低	1	13
rs7457728	病毒	人类乳头瘤病毒	感染风险	G	风险较低	1	12
rs7457728	病毒	人类乳头瘤病毒	感染风险	C	风险较高	1.21	12
HLA-DRB1*04:03	病毒	人类乳头瘤病毒	感染风险	携带	风险较高	1.89	11
HLA-DRB1*04:03	病毒	人类乳头瘤病毒	感染风险	不携带	风险较低	1	11
HLA-DQB1*06:02	病毒	人类乳头瘤病毒	感染风险	携带	风险较高	2.09	11

HLA-DQB1*06:02	病毒	人类乳头瘤病毒	感染风险	不携带	风险较低	1	11
----------------	----	---------	------	-----	------	---	----

[0159] 由表7可知,首先,计算人群平均携带风险等位基因数 $M=2 \times 0.362+2 \times 0.134+2 \times 0.360+2 \times 0.0156+2 \times 0.0762=1.8956$ ,即人群中平均携带1.8956个风险等位基因。其次,若某用户检测结果为rs9357152 AG、rs59661306 AG、rs7457728 CG、HLA-DRB1\*04:03 不携带、HLA-DQB1\*06:02不携带,得到风险等位基因携带数为3,则风险等位基因携带数(3)大于人群平均携带风险等位基因数(1.8956)。可知,该用户的人类乳头瘤病毒的“感染风险”的预测结果为风险较高。

[0160] S3-2、根据某个病原微生物的预测结果,生成用户的微生物感染风险报告。

[0161] 具体实施时,汇总特定病原微生物在步骤S3-1中的预测结果,并根据每个病原微生物相应的报告模版,生成该病原微生物感染风险预测的电子版报告。并且,预测结果可能包含一个或多个风险类型维度,即“感染风险”、“感染严重程度”和/或“继发疾病遗传风险”,以及每个风险类型维度对影响程度的定义,即较低或较高。

[0162] 如图2所示,以“肝炎病毒”的微生物感染风险报告模版为例,报告模版包括以下板块:报告ID、用户ID、检测方式、报告日期、感染风险预测结果、感染严重程度预测结果、继发疾病遗传风险预测结果、预防建议、相关知识、检测细节和局限、参考文献。

[0163] 其中,所述感染风险预测结果、感染严重程度预测结果、继发疾病遗传风险预测结果均含有影响程度和查看详情;所述影响程度为较低或较高;所述查看详情用于展示与肝炎病毒特定预测结果相关的详细信息,包括:影响程度、报告详情描述、检测基因、检测位点、检测结果、变异类型、影响描述。所述报告详情描述用于展示基因变异的表型(即影响程度的描述),表型的展示优先级从高到低为:1)“风险较高”的基因变异表型;2)“风险较低”的基因变异表型。

[0164] 由此,既可以通过微生物感染风险报告了解检测的整体概况,还可以根据需要,通过查看详情获取具体的检测结果和详细依据。

[0165] 实施例2:本发明基于基因检测预测微生物感染风险的分析系统。

[0166] 如图3所示,本实施例提供一种基于基因检测预测微生物感染风险的分析系统,包括数据库构建模块、微生物感染风险预测模块、报告生成模块,具体介绍如下。

[0167] 所述数据库构建模块,用于通过公开文献获取微生物感染相关的标准化信息,包括微生物感染相关的变异名、病原微生物类型、病原微生物名、基因与病原微生物关联的风险类型、等位基因、与等位基因关联的表型、等位基因的遗传贡献OR值、参考文献编号,并构建微生物感染参考数据库。

[0168] 所述微生物感染风险预测模块,用于获取中国人群的基因数据,并进行质量控制,建立中国人群微生物感染基线数据集;然后根据中国人群微生物感染基线数据集,计算微生物感染相关的变异在中国人群中的频率。

[0169] 所述报告生成模块,用于获取用户的基因数据,根据微生物感染参考数据库,对用户进行病原微生物感染风险预测,并生成微生物感染风险报告。

[0170] 实施例3:本发明基于基因检测预测微生物感染风险的分析装置。

[0171] 如图4所示,本实施例提供一种基于基因检测预测微生物感染风险的分析装置,包括客户端、服务器和计算设备,其中:

[0172] 所述客户端是一种软件应用,用于连接服务器获得并查看微生物感染风险报告。

具体实施时,客户端可以是智能手机、平板电脑、笔记本电脑、台式计算机、智能手表,但并不局限于此。客户端的操作系统可以为iOS系统、Android系统、Windows系统,但并不局限于此。

[0173] 所述服务器是客户端和计算设备连接的中转站,用于接收或发送请求信息。

[0174] 所述计算设备用于(经服务器接力)接受客户端的信息或向客户端发出信息(包括预测结果(报告)),并存储客户端发送及获取到的信息。具体实施时,计算设备为具有较强计算能力的服务器,可以是独立的物理服务器,也可以是多个物理服务器构成的服务器集群或者分布式系统,还可以是提供基础云计算服务的云服务器。

[0175] 具体实施时,所述计算设备包括:处理器、通信器及存储器,其中:

[0176] 所述存储器是计算设备的存储中心,用于存储所述基于基因检测的微生物感染风险预测方法的计算机软件程序,该计算机软件程序存储于计算机可读存储介质中,并由处理器进行加载和执行。

[0177] 所述处理器是计算设备的控制中心,用于利用各种接口和线路连接整个计算设备的各个部分,通过运行存储在存储器内的基于基因检测的微生物感染风险预测方法的计算机软件程序,或执行来自通信器的命令或请求,执行计算设备的各种功能和数据处理,从而对计算设备进行整体监控。可选的,处理器可包括一个或多个处理核心。

[0178] 所述通信器是计算设备与外部设备进行通信连接的通信中心,用于获取设备功能的功能名称;根据配置请求以及功能名称,生成对应的命令类型信息;根据配置传输命令信息。

[0179] 虽然本发明已经提供了具体的实施例进行描述,但这些实施例仅用于解释本发明,并不对发明进行限制。在任何一个或多个实施例或示例中,可以以合适的方式结合描述的具体特征、结构、材料或特点。本领域的技术人员在阅读完本说明书后,可以在不违背本发明原理和宗旨的前提下,根据需要对实施例进行修改、替换和变型等,这些改变虽然没有创造性贡献,但只要在本发明的权利要求范围内,仍然受到专利法的保护。



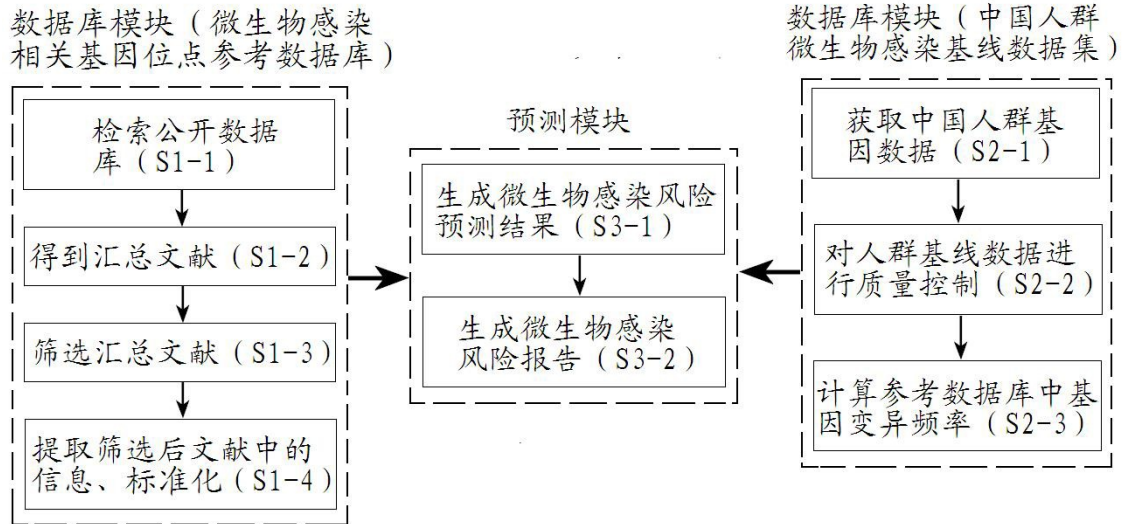


图1

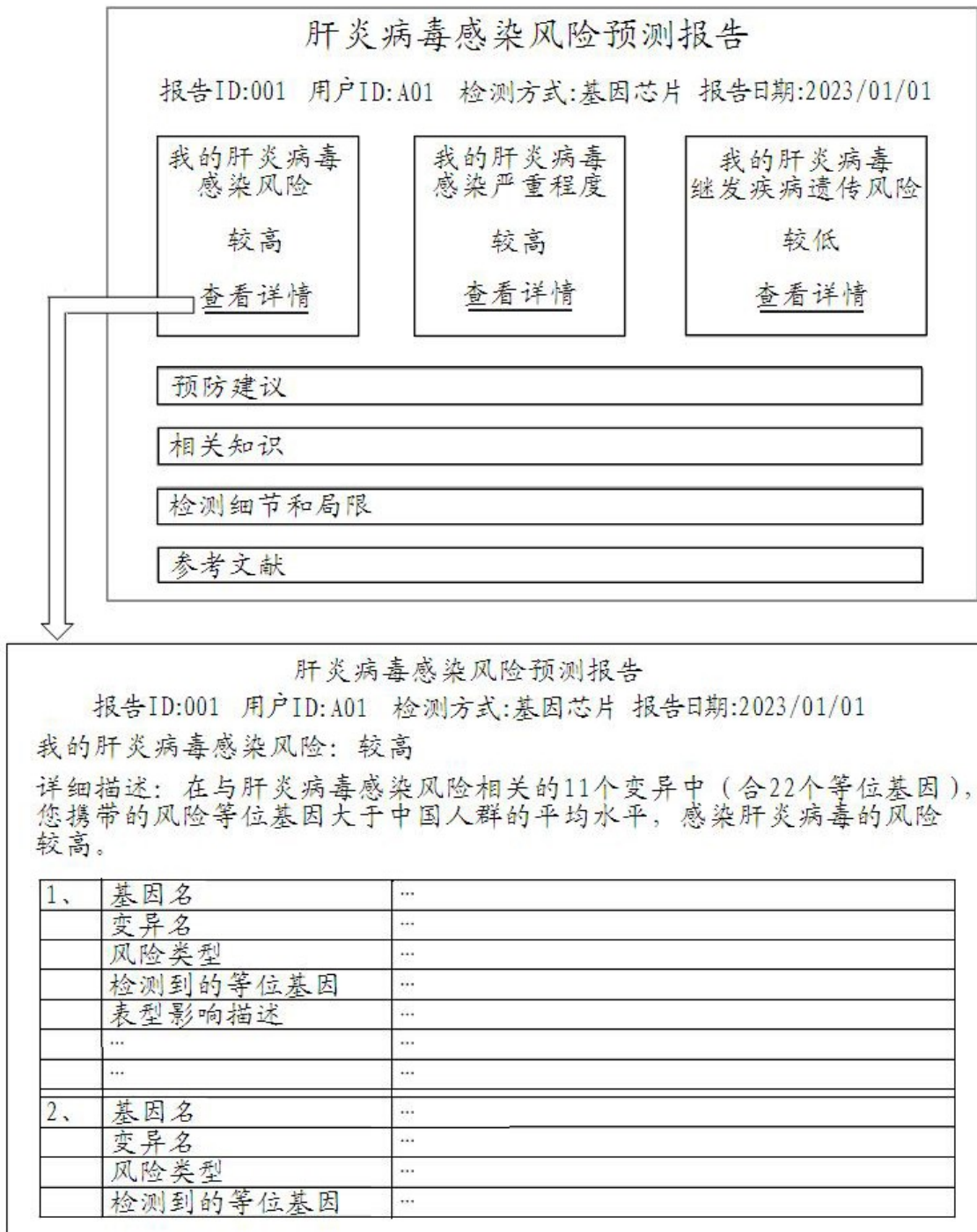


图2

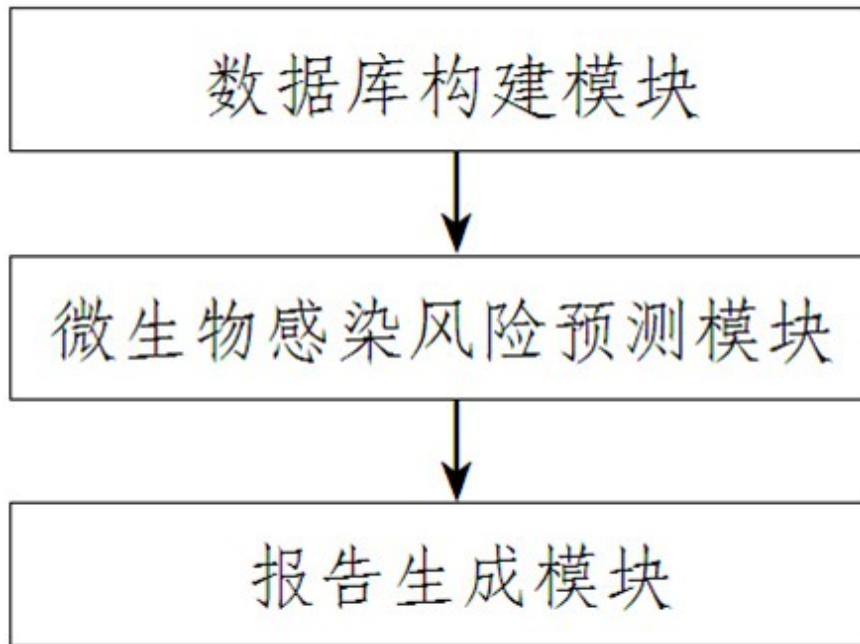


图3

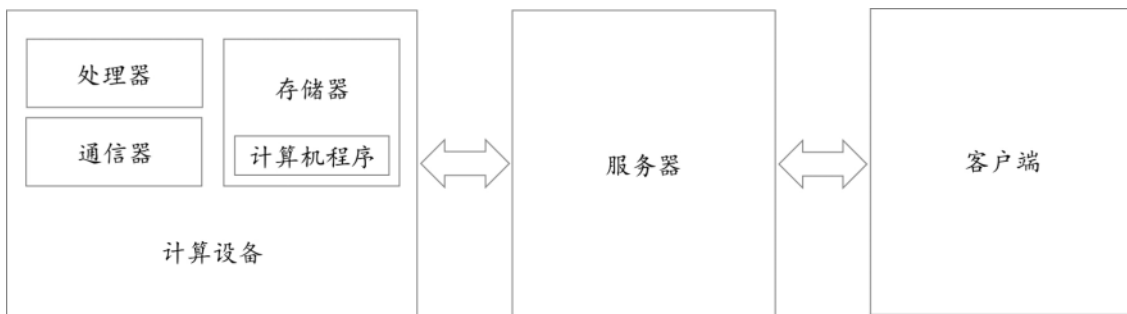


图4