



(12) 发明专利

(10) 授权公告号 CN 108778287 B

(45) 授权公告日 2021.08.31

(21) 申请号 201780018749.0

(22) 申请日 2017.02.15

(65) 同一申请的已公布的文献号
申请公布号 CN 108778287 A

(43) 申请公布日 2018.11.09

(30) 优先权数据
201621005426 2016.02.16 IN

(85) PCT国际申请进入国家阶段日
2018.09.20

(86) PCT国际申请的申请数据
PCT/IB2017/050840 2017.02.15

(87) PCT国际申请的公布数据
W02017/141169 EN 2017.08.24

(73) 专利权人 塔塔咨询服务服务有限公司
地址 印度马哈拉施特拉邦孟买

(72) 发明人 沙尔米拉·谢卡尔·曼德
穆罕默德·蒙佐尔·哈克
阿尼班·杜塔
尼什尔·库马尔·平纳
米塔利·麦钱特

(74) 专利代理机构 北京品源专利代理有限公司
11332

代理人 王小衡 王天鹏

(51) Int.Cl.
A61K 31/702 (2006.01)
C12Q 1/6883 (2018.01)
G01N 33/53 (2006.01)
G01N 33/68 (2006.01)
A61B 34/10 (2016.01)

审查员 贾星航

权利要求书6页 说明书29页 附图6页

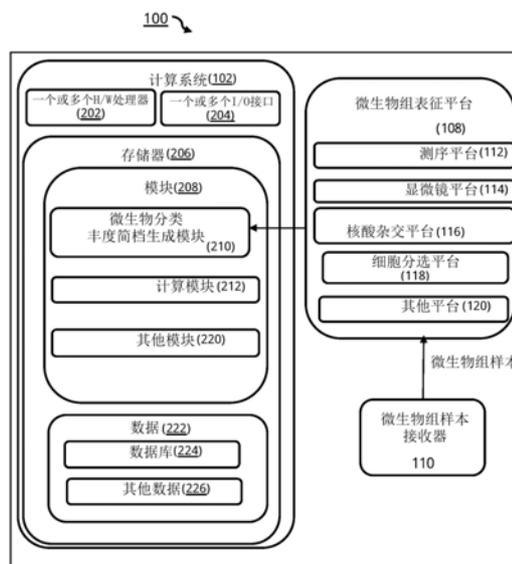
(54) 发明名称

用于早产结果的早期风险评估的方法和系统

(57) 摘要

公开了用于评估怀孕受试者的早产风险的系统和方法。现有的早产风险评估方法在妊娠的晚第二个或第三个三个月提供结果，因此可用于医疗建议的时间很少。本文公开的方法和系统在妊娠的15周内预测早产风险。从来自怀孕受试者的微生物组样本获得微生物组表征数据。根据微生物组表征数据生成的“微生物分类丰度简档”包含微生物组样本中存在的微生物的丰度值。针对“微生物分类丰度简档”计算“分类构成偏度”值和分布特征值，量化来自微生物分类丰度简档的微生物的丰度值的偏差。基于集合DSR的分布特征值或“分类构成偏度”值确定早产的风险，其中集合DSR包括量化了来自微生物组样本的微生物的丰度值中的偏差的值。

CN 108778287 B



1. 一种用于对从怀孕受试者获得的微生物组样本中的多种微生物的丰度值的偏差进行量化的方法,所述方法包括:

由硬件处理器(202)接收微生物组样本的“微生物分类丰度简档”,其中所述微生物组样本从怀孕受试者获得;并且其中所述微生物分类丰度简档包含所述微生物组样本中存在的多种微生物中的每种的丰度值;

由所述硬件处理器(202)针对所述“微生物分类丰度简档”,计算“分类构成偏度”值和分布特征值中的至少一个,其中所述“分类构成偏度”值和所述分布特征值量化了来自所述微生物分类丰度简档的所述多种微生物的丰度值的偏差,其中对所述“分类构成偏度”值和所述分布特征值的计算包括,

a) 通过计算来自所述微生物分类丰度简档的所述多种微生物的丰度值的总和来获得总种群计数“M”;

b) 创建包含来自所述微生物分类丰度简档的所述多种微生物中的每种的丰度值的经排序的列表“L”,其中所述经排序的列表“L”包括存在于所述微生物分类丰度简档中按丰度值的递增次序或递减次序排名的多种微生物中的每种的丰度值,并且其中,当列表L按丰度值的递增次序排序时,经排序的列表“L”中排名靠前的一个或多个条目对应于“稀疏分类群”并且经排序的列表“L”中排名垫底的一个或多个条目对应于“优势分类群”,而当列表L按丰度值的递减次序排序时,经排序的列表“L”中排名靠前的一个或多个条目对应于“优势分类群”并且经排序的列表“L”中排名垫底的一个或多个条目对应于“稀疏分类群”;

c) 计数其累积丰度 \geq 总种群计数“M”的*i*%的“稀疏分类群”的最小数量(ST_i),其中当列表L按递增次序排序时,丰度值按经排序的列表“L”中每种微生物的递增排名的次序逐渐累积,或者,计数其累积丰度 \geq 总种群计数“M”的*i*%的“稀疏分类群”的最小数量(ST_i),其中当列表L按递减次序排序时,丰度值按经排序的列表“L”中每种微生物的递减排名的次序逐渐累积,并且其中“*i*”是整数或分数;

d) 计数其累积丰度 \geq 总种群计数“M”的*j*%的“优势分类群”的最小数量(DT_j),其中当列表L按递增次序排序时,丰度值按经排序的列表“L”中每种微生物的递减排名的次序逐渐累积,或者计数其累积丰度 \geq 总种群计数“M”的*j*%的“优势分类群”的最小数量(DT_j),其中当列表L按递减次序排序时,丰度值按经排序的列表“L”中每种微生物的递增排名的次序逐渐累积,并且其中“*j*”是整数或分数;

e) 重复步骤“c”和“d”以获得针对预定义范围的“*i*”值和“*j*”值的 ST_i 和 DT_j 的计数,以分别生成两个值集合“ST”和“DT”,其中*i*和*j*属于预定义范围内的两个相同的算术级数,并且其中所述算术级数的连续项之间的公差是预定义的整数或预定义的分数值;

f) 计算集合“DT”和集合“ST”的笛卡尔乘积,以获得包含对应于 DT_j 值和 ST_i 值的多个有序对的集合“C”;

g) 通过针对所述集合“C”中的每个有序对而计算($DT_j \div ST_i$)来导出集合“DSR”;

h) 计算所述集合“DSR”的所述分布特征值,其中所述集合“DSR”的所述分布特征值包括所述集合DSR的集中趋势值;

i) 通过对包括所述集合“DSR”的集中趋势值的所述分布特征值应用数学变换来计算所述“分类构成偏度”值;其中所述集合DSR的所述集中趋势值、所述集合“DSR”的所述分布特征值或所述“分类构成偏度”值中的至少一个指示出所述怀孕受试者向着早产的趋势。

2. 根据权利要求1所述的方法,其中,所述微生物组样本从处于妊娠的前15周内或处于妊娠的前三个月内的所述怀孕受试者获得,其中所述怀孕受试者是孕妇。

3. 根据权利要求1所述的方法,其中,所述微生物组样本从处于妊娠期期间的任何时间处的所述怀孕受试者获得。

4. 根据权利要求1所述的方法,其中,所述怀孕受试者是孕妇。

5. 根据权利要求1所述的方法,其中,所述怀孕受试者是雌性胎生动物。

6. 根据权利要求1所述的方法,其中,所述微生物组样本从所述怀孕受试者的口腔、皮肤、肠、阴道或其他身体部位获得,并且其中所述微生物组样本选自包括以下的组:阴道拭子样本、宫颈拭子样本、尿液样本、羊水样本、血液样本、血清样本、血浆样本、胎盘拭子、脐带拭子、粪便样本、皮肤拭子、口腔拭子、唾液样本、牙周拭子、咽喉拭子、鼻拭子、囊泡液样本、鼻咽拭子、鼻孔拭子、结膜拭子、直肠拭子以及支气管拭子。

7. 根据权利要求1所述的方法,包括生成所述微生物组样本的“微生物分类丰度简档”,其中,生成所述“微生物分类丰度简档”还包括:

从所述怀孕受试者获得所述微生物组样本,

经由测序平台(112)从所述微生物组样本获得经测序的微生物DNA数据,其中获得所述经测序的微生物DNA数据包括:

使用至少一种DNA提取技术分离对应于所述微生物组样本中存在的所述多种微生物的微生物DNA;以及

使用至少一种DNA测序技术对所述微生物DNA进行测序以获得经测序的微生物DNA数据;并且

通过由所述硬件处理器(202)使用一种或多种分类学分类技术以计算方式分析所述经测序的微生物DNA数据,而生成所述微生物组样本的所述“微生物分类丰度简档”。

8. 根据权利要求1所述的方法,包括生成所述微生物组样本的所述“微生物分类丰度简档”,其中,生成所述“微生物分类丰度简档”还包括:

经由微生物组样本接收器(110),从所述怀孕受试者获得所述微生物组样本;

通过应用以下一种或多种技术经由微生物组表征平台(108)而获得所述微生物组样本的微生物组表征数据,所述一种或多种技术包括:显微镜检查、流式细胞术方法、基于体外培养的方法、一种或多种酶或荧光测定、或涉及对来自复杂微生物种群的信号的光谱鉴定和筛选的一种或多种测定,其中所述微生物组表征数据包括:显微镜成像数据、流式细胞术细胞测量数据、体外培养中生长的微生物的菌落计数和细胞表型数据、信号强度数据;并且

通过由所述硬件处理器使用一种或多种分类学分类技术以计算方式分析所述微生物组表征数据,而生成所述微生物组样本的所述“微生物分类丰度简档”。

9. 根据权利要求7所述的方法,还包括以下中的至少一种:1) 根据所述微生物DNA对细菌16S rRNA或23S rRNA标记基因进行扩增和测序,2) 根据所述微生物DNA对细菌16S rRNA或23S rRNA标记基因的全长或一个或多个特定区域进行扩增和测序,或者3) 根据所述微生物DNA对一个或多个系统发育标记基因进行扩增和测序。

10. 根据权利要求7所述的方法,其中,所述至少一种DNA测序技术选自下一代测序技术,其中所述下一代测序技术包括:全基因组鸟枪法(WGS)测序、基于片段库的测序技术、配对库或基于双端库的测序技术或其组合。

11. 根据权利要求1所述的方法,其中,“i”值和“j”值的预定义范围在0到100的范围内,并且其中i和j的预定义范围被定义为使得 $i+j$ 等于100。

12. 根据权利要求1所述的方法,其中,所述分类构成偏度在0到1的范围内取值。

13. 根据权利要求1所述的方法,其中,所述集合“DSR”的所述集中趋势值是平均值、中值、众数或将单个值标识为对所述集合DSR中的值的整体分布的表示的任何统计测度中的一个。

14. 根据权利要求1所述的方法,其中,所述集合DSR的所述分布特征值包括以下中的一个:所述集合DSR的最大值、所述集合DSR的最小值、所述集合DSR的方差、所述集合DSR的偏度、或者表征所述集合DSR中的值的分布的模式、对称性、变异性的度量。

15. 根据权利要求1所述的方法,其中,应用于所述集中趋势值的所述数学变换包括以下中的一个的数学变换方法:从单位一中减去所述集中趋势值、计算所述集中趋势值的倒数、或导致了针对所述集合DSR的较低集中趋势值生成较高数值的分类构成偏度以及反之亦然。

16. 根据权利要求1所述的方法,其中,所述怀孕受试者被分类为针对向着早产的趋势的一个或多个风险类别,其中所述分类基于所述集合DSR的所述集中趋势值、所述集合“DSR”的所述分布特征值或所述“分类构成偏度”值中的至少一个,并且其中所述一个或多个风险类别包括低早产风险类别、中度早产风险类别和高早产风险类别。

17. 根据权利要求16所述的方法,其中,针对从孕妇获得的阴道微生物组样本,对于大于0.922的所述“分类构成偏度”值,指示所述低早产风险类别,对于在0.887至0.922的范围内的所述“分类构成偏度”值,指示所述中度早产风险类别,并且对于小于0.887的所述“分类构成偏度”值,指示所述高早产风险类别,并且其中所述“分类构成偏度”值通过从单位一减去计算出的所述集合DSR的中值而获得。

18. 根据权利要求16所述的方法,其中,针对从孕妇获得的阴道微生物组样本,对于大于0.113的所述集合DSR的所述集中趋势值,指示所述高早产风险类别,对于在0.078至0.113范围内的所述集合DSR的所述集中趋势值,指示所述中度早产风险类别,对于小于0.078的所述集合DSR的所述集中趋势值,指示所述低早产风险类别,并且其中所述集合DSR的所述集中趋势值是所述集合DSR的中值。

19. 一种用于评估怀孕受试者的早产风险的系统(100),所述系统(100)包括计算系统(102),所述计算系统(102)进一步包括:

硬件处理器(202);以及

存储器(204),其被耦合到所述硬件处理器(202),其中所述硬件处理器(202)执行被存储在所述存储器(204)中的多个模块(208),并且其中所述多个模块(208)包括:

计算模块(212),用于:

接收从怀孕受试者获得的微生物组样本的“微生物分类丰度简档”,其中所述微生物分类丰度简档包含所述微生物组样本中存在的多种微生物的丰度值;并且

针对所述“微生物分类丰度简档”,计算“分类构成偏度”值和分布特征值中的至少一个,其中所述“分类构成偏度”值和所述分布特征值量化了来自所述微生物分类丰度简档的所述多种微生物的丰度值的偏差,其中对所述“分类构成偏度”值和所述分布特征值的计算包括:

(a) 通过计算来自所述微生物分类丰度简档的所述多种微生物的丰度值之和来获得总种群数“M”；

(b) 创建包含来自所述微生物分类丰度简档的所述多种微生物的丰度值的经排序的列表“L”，其中所述经排序的列表“L”包括存在于所述微生物分类丰度简档中按丰度值的递增次序或递减次序排名的多种微生物，并且其中，当列表L按丰度值的递增次序排序时，经排序的列表“L”中排名靠前的一个或多个条目对应于“稀疏分类群”，并且经排序的列表“L”中排名垫底的一个或多个条目对应于“优势分类群”，并且当列表L按丰度值的递减次序排序时，经排序的列表“L”中排名靠前的一个或多个条目对应于“优势分类群”，并且经排序的列表“L”中排名垫底的一个或多个条目对应于“稀疏分类群”；

(c) 计数其累积丰度 \geq 总种群计数“M”的*i*%的“稀疏分类群”的最小数量(ST_i)，其中当列表L按递增次序排序时，丰度值按经排序的列表“L”中每种微生物的递增排名的次序逐渐累积，或者计数其累积丰度 \geq 总种群计数“M”的*i*%的“稀疏分类群”的最小数量(ST_i)，其中当列表L按递减次序排序时，丰度值按经排序的列表“L”中每种微生物的递减排名逐渐累积，并且其中“*i*”是整数或分数；

(d) 计数其累积丰度 \geq 总种群计数“M”的*j*%的“优势分类群”的最小数量(DT_j)，其中当列表L按递增次序排序时，丰度值按经排序的列表“L”中每种微生物的递减排名的次序逐渐累积，或者计数其累积丰度 \geq 总种群计数“M”的*j*%的“优势分类群”的最小数量(DT_j)，其中当列表L按递减次序排序时，丰度值按经排序的列表“L”中每种微生物的递增排名逐渐累积，并且其中“*j*”是整数或分数；

(e) 重复步骤c和d以获得针对预定义范围的“*i*”值和“*j*”值的 ST_i 和 DT_j 的计数，以分别生成两个值集合“ST”和“DT”，其中*i*和*j*属于预定义范围内的两个相同的算术级数，并且其中所述算术级数的连续项之间的公差是预定义的整数或预定义的分数值；

(f) 计算集合“DT”和集合“ST”的笛卡尔乘积，以获得包含对应于 DT_j 值和 ST_i 值的多个有序对的集合“C”；

(g) 通过针对所述集合“C”中的每个有序对而计算($DT_j \div ST_i$)来导出集合“DSR”；

(h) 计算所述集合“DSR”的所述分布特征值，其中所述集合“DSR”的所述分布特征值包括所述集合DSR的集中趋势值；

(i) 通过对包括所述集合“DSR”的集中趋势值的所述分布特征值应用数学变换来计算所述“分类构成偏度”值；并且

基于所述集合DSR的所述集中趋势值、所述集合“DSR”的所述分布特征或所述“分类构成偏度”值中的至少一个来确定所述怀孕受试者的早产风险。

20. 根据权利要求19所述的系统，包括：

微生物组样本接收器(110)，用于从所述怀孕受试者接收所述微生物组样本；

微生物组表征平台(108)，用于：

从所述微生物组样本接收器(110)接收所述微生物组样本；以及

通过应用来自以下技术组的至少一种技术来获得所述微生物组样本的微生物组表征数据，所述技术组包括：测序技术、显微镜检查、流式细胞术方法、基于体外培养的方法、一种或多种酶或荧光测定、或涉及对来自复杂微生物种群的信号的光谱鉴定和筛选的一种或多种测定，其中所述微生物组表征数据包括显微镜成像数据、流式细胞术细胞测量数据、体

外培养中生长的微生物的菌落计数和细胞表型数据、信号强度数据;并且

所述微生物分类丰度简档生成模块(210)由所述硬件处理器使用一种或多种分类学分类技术以计算方式分析所述微生物组表征数据,以生成所述微生物组样本的所述微生物分类丰度简档。

21. 根据权利要求20所述的系统,包括

测序平台(112),用于:

从所述微生物组样本接收器(110)接收所述微生物组样本;

使用至少一种DNA提取技术分离对应于所述微生物组样本中存在的所述多种微生物的微生物DNA;并且

使用至少一种DNA测序技术对所述微生物DNA进行测序以获得经测序的微生物DNA数据;并且

所述微生物分类丰度简档生成模块(210)使用一种或多种分类学分类技术以计算方式分析所述经测序的微生物DNA数据,以生成所述微生物组样本的所述微生物分类丰度简档。

22. 根据权利要求19所述的系统,其中,所述怀孕受试者是女性或雌性胎生动物。

23. 根据权利要求20所述的系统,其中,所述微生物组表征平台(108)包括测序平台(112)、显微镜平台(114)、核酸杂交平台(116)或细胞分选平台(118)中的至少一个。

24. 根据权利要求19所述的系统,其中,“i”值和“j”值的预定义范围在0到100的范围内,并且其中i和j的预定义范围被定义为使得 $i+j$ 等于100。

25. 根据权利要求19所述的系统,其中,所述分类构成偏度在0到1的范围内取值。

26. 根据权利要求19所述的系统,其中,通过将所述怀孕受试者分类为针对早产的一个或多个风险类别来确定所述怀孕受试者的早产风险,其中所述怀孕受试者的分类基于所述集合DSR的所述集中趋势值、所述集合“DSR”的所述分布特征值或所述“分类构成偏度”值中的至少一个,并且其中所述一个或多个风险类别包括低早产风险类别、中度早产风险类别和高早产风险类别。

27. 根据权利要求26所述的系统,其中,针对从孕妇获得的阴道微生物组样本,对于大于0.922的所述“分类构成偏度”值,指示所述低早产风险类别,对于在0.887至0.922的范围内的所述“分类构成偏度”值,指示所述中度早产风险类别,并且对于小于0.887的所述“分类构成偏度”值,指示所述高早产风险类别,并且其中所述“分类构成偏度”值通过从单位一减去计算出的所述集合DSR的中值而获得。

28. 根据权利要求26所述的系统,其中,针对从孕妇获得的阴道微生物组样本,对于大于0.113的所述集合DSR的所述集中趋势值,指示所述高早产风险类别,对于在0.078至0.113范围内的所述集合DSR的所述集中趋势值,指示所述中度早产风险类别,对于小于0.078的所述集合DSR的所述集中趋势值,指示所述低早产风险类别,并且其中所述集合DSR的所述集中趋势值是所述集合DSR的中值。

29. 一种非暂时性计算机可读介质,其体现可由硬件处理器在计算设备中执行以用于评估怀孕受试者的早产风险的程序,所述程序包括:

用于接收从怀孕受试者获得的微生物组样本的“微生物分类丰度简档”的程序代码,其中所述微生物分类丰度简档包含所述微生物组样本中存在的多种微生物的丰度值;以及

用于针对所述“微生物分类丰度简档”计算“分类构成偏度”值和分布特征值中的至少

一个的程序代码,其中所述“分类构成偏度”值和所述分布特征值量化了来自所述微生物分类丰度简档的所述多种微生物的丰度值的偏差,其中对所述“分类构成偏度”值和所述分布特征值的计算包括:

a) 通过计算来自所述微生物分类丰度简档的所述多种微生物的丰度值的总和来获得总种群计数“M”;

b) 创建包含来自所述微生物分类丰度简档的所述多种微生物的丰度值的经排序的列表“L”,其中所述经排序的列表“L”包括存在于所述微生物分类丰度简档中按丰度值的递增次序或递减次序排名的多种微生物,并且其中,当列表L按丰度值的递增次序排序时,经排序的列表“L”中排名靠前的一个或多个条目对应于“稀疏分类群”并且经排序的列表“L”中排名垫底的一个或多个条目对应于“优势分类群”,并且当所述列表L按丰度值的递减次序排序时,经排序的列表“L”中排名靠前的一个或多个条目对应于“优势分类群”并且经排序的列表“L”中排名垫底的一个或多个条目对应于“稀疏分类群”;

c) 计数其累积丰度 \geq 总种群计数“M”的*i*%的“稀疏分类群”的最小数量(ST_i),其中当所述列表L按递增次序排序时,丰度值按经排序的列表“L”中每种微生物的递增排名的次序逐渐累积,或者,计数其累积丰度 \geq 总种群计数“M”的*i*%的“稀疏分类群”的最小数量(ST_i),其中当所述列表L按递减次序排序时,丰度值按经排序的列表“L”中每种微生物的递减排名逐渐累积,并且其中“*i*”是整数或分数;

d) 计数其累积丰度 \geq 总种群计数“M”的*j*%的“优势分类群”的最小数量(DT_j),其中当所述列表L按递增次序排序时,丰度值按经排序的列表“L”中每种微生物的递减排名的次序逐渐累积,或者计数其累积丰度 \geq 总种群计数“M”的*j*%的“优势分类群”的最小数量(DT_j),其中当所述列表L按递减次序排序时,丰度值按经排序的列表“L”中每种微生物的递增排名逐渐累积,并且其中“*j*”是整数或分数;

e) 重复步骤c和d以获得针对预定义范围的“*i*”值和“*j*”值的 ST_i 和 DT_j 的计数,以分别生成两个值集合“ST”和“DT”,其中*i*和*j*属于预定义范围内的两个相同的算术级数,并且其中所述算术级数的连续项之间的公差是预定义的整数或预定义的分数值;

f) 计算集合“DT”和集合“ST”的笛卡尔乘积,以获得包含对应于 DT_j 值和 ST_i 值的多个有序对的集合“C”;

g) 通过针对所述集合“C”中的每个有序对而计算($DT_j \div ST_i$)来导出集合“DSR”;

h) 计算集合“DSR”的所述分布特征值,其中所述集合“DSR”的所述分布特征值包括所述集合DSR的集中趋势值;

i) 通过对包括所述集合“DSR”的所述集中趋势值的所述分布特征值应用数学变换来计算所述“分类构成偏度”值;

用于基于所述集合DSR的所述集中趋势值、所述集合“DSR”的所述分布特征或所述“分类构成偏度”值中的至少一个来确定所述怀孕受试者的早产风险的程序代码。

用于早产结果的早期风险评估的方法和系统

[0001] 相关申请的交叉引用

[0002] 本申请要求2016年2月16日提交的印度临时专利申请号201621005426的优先权，其整体通过引用借此被并入。

技术领域

[0003] 本文所描述的实施例总体涉及用于怀孕受试者的早产结果的风险评估的方法和系统，并且更具体地涉及用于怀孕受试者的早产结果的早期风险评估的基于微生物组的方法。

背景技术

[0004] 根据世界卫生组织 (WHO) 最近的估计，早产 (即在妊娠的37个完整周之前出生) 是导致全球新生儿死亡的主要原因。与痢疾、疟疾或艾滋病相比，由于早产相关并发症导致更多儿童死亡。在妊娠总数的5%-18%中观察到早产 (PTD) 病例 (www.who.int/mediacentre/factsheets/fs363/en/)。与PTD结果相关联的产前和/或产后成本显著较高。PTD通常与母亲以及所分娩的儿童两者的风险相关联。显著部分的早产儿会出现几种危及生命的并发症，包括中度或重度呼吸窘迫、免疫系统受损、黄疸、贫血、脑瘫以及视力和听力问题。鉴于此，每年有超过一百万儿童由于早产引起的并发症而死亡，集中于早产的诊断、预测和预防的研究具有很高重要性。用于预测早产风险的物理和/或生物化学诊断标记物很少，然而这些物理和/或生物化学诊断标记物大多适合应用于中期或晚期妊娠阶段。

[0005] 目前，尚未完全了解导致早产 (PTD) 的病理-生理事件的级联 (cascade)。然而，一些可以潜在地促进健康的足月妊娠结果的干预方法 (例如孕酮补剂、宫颈环扎术等) 是可用的。鉴于此，能够在妊娠早期准确地指示早产结果的风险/“进展可能性”的诊断标记物具有很大意义。目前，可以在妊娠的各个阶段采用以预测早产结果的风险的多个诊断方法是可用的。然而，其中只有少数适合在妊娠早期 (即前三个月) 应用。此外，它们的总体准确性 (就敏感性和特异性而言) 较低使得它们不适合在常规临床实践中使用。此外，可以注意到，在现有的PTB诊断方法中，那些具有相对较好的诊断/预测能力的方法，大多仅适用于妊娠的中期或晚期 (即妊娠的第二和第三个三个月)。虽然这种诊断能力 (在妊娠的中期或晚期) 有助于将孕妇重新安置到适合的医疗环境 (其中孕妇可以被更好/持续地监视)，但诊断和分娩之间相对较短的持续时间使得难以采取促进足月的健康妊娠的有意义的干预策略。这突出了目前对诊断方法的需求，该方法可以在妊娠的非常“早”期 (理想地是前三个月) 预测早产结果的可能性。这种“早期”诊断/风险评估程序将为执业医师提供足够的时间来采用可以潜在地/可能促进足月的健康妊娠的可用的干预/监视策略。

[0006] 此外，评估早产结果风险的许多现有诊断技术依赖于怀孕受试者 (a) 自身呈现出阴道感染，(b) 在放射学程序中检测到异常，或 (c) 在更早的分娩中具有早产史。因此，依赖于上述讨论的指标使得这种诊断技术不适用于第一次怀孕或不是由已知风险因素 (即阴道感染和/或“胎儿/子宫”异常) 驱动的早产结果的情况。此外，一些现有的诊断方法需要本质

上是侵入性的采样技术,从而可能对经受测试的受试者造成/施加身体和/或情感上的不便。

[0007] 现有的早产风险评估方法的主要缺点在于这些方法仅在妊娠的“后”期(即在妊娠的晚第二个三个月或第三个(即最后的)三个月)提供具有实际效用的诊断结果。因此,受试者很少有时间采取预防性或矫正性医疗建议来减少或避免早产风险。因此,在妊娠早期(特别是妊娠的15周内(即在妊娠中尽可能早))准确预测早产结果的风险极为重要。对早产结果的早期预测为孕妇/受试者(检测到有PTD风险)提供了足够的时间来采取可能会降低或消除早产结果的风险的必要的预防性或矫正性医疗程序。

发明内容

[0008] 以下描述呈现出本公开的一些实施例的简化概述以提供对这些实施例的基本理解。该概述不是对这些实施例的广泛概要。不旨在识别这些实施例的关键/重要元件或描绘这些实施例的范围。其唯一目的在于以简化的形式呈现一些实施例,作为以下呈现的更详细的描述的前奏。鉴于上文,本文的实施方式提供了用于怀孕受试者的早产结果的早期风险评估的系统和方法。

[0009] 在本公开的一个方面,公开了一种用于评估怀孕受试者的早产风险的方法。为了评估早产风险,从处于妊娠的前15周内或处于妊娠的前三个月内或处于妊娠的第二个三个月内的怀孕受试者获得微生物组样本,并且其中所述怀孕受试者是孕妇。该方法包括由硬件处理器接收微生物组样本的“微生物分类丰度简档(microbial taxonomic abundance profile)”。从怀孕受试者获得微生物组样本。微生物分类丰度简档包含微生物组样本中存在的多种微生物中的每种的丰度值。该方法还包括由硬件处理器针对“微生物分类丰度简档”计算“分类构成偏度”值和分布特征值中的至少一个。“分类构成偏度”值和分布特征值量化来自微生物分类丰度简档的多种微生物的丰度值中的偏差。“分类构成偏度”值和分布特征值的计算包括,在步骤(a)中,通过计算来自微生物分类丰度简档的多种微生物的丰度值之和来获得总种群数“M”。此外,在步骤(b)中,该方法包括创建包含来自微生物分类丰度简档的多种微生物中的每种的丰度值的经排序的列表“L”。经排序列表“L”包括存在于“微生物分类丰度简档”中按丰度值的递增次序或递减次序排名的多种微生物中的每种的丰度值。当列表L按丰度值的递增次序排序时,经排序的列表“L”中排名靠前的一个或多个条目对应于“稀疏分类群”,并且经排序的列表“L”中排名垫底的一个或多个条目对应于“优势分类群”,并且当列表L按丰度值的递减次序排序时,经排序的列表“L”中排名靠前的一个或多个条目对应于“优势分类群”,并且经排序的列表“L”中排名垫底的一个或多个条目对应于“稀疏分类群”。此外,在步骤(c)中,该方法包括:计数以下“稀疏分类群”的最小数量(ST_i),其累积丰度 \geq 总种群计数“M”的*i*%,其中当列表L按递增次序排序时,丰度值按经排序的列表“L”中每种微生物的递增排名的次序逐渐累积,或者,计数以下“稀疏分类群”的最小数量(ST_i),其累积丰度 \geq 总种群计数“M”的*i*%,其中当列表L按递减次序排序时,丰度值按经排序的列表“L”中每种微生物的递减排名的次序逐渐累积,并且其中“*i*”是整数或分数。此外,在步骤(d)中,在一个选项中,该方法包括:计数以下“优势分类群”的最小数量(DT_j),其累积丰度 \geq 总种群计数“M”的*j*%,其中当所述列表L按递增次序排序时,丰度值按经排序的列表“L”中每种微生物的递减排名的次序逐渐累积,或者在另一选项中,该方法包括:计数以

下“优势分类群”的最小数量(DT_j),其累积丰度 \geq 总种群计数“M”的j%,其中当所述列表L按递减次序排序时,丰度值按经排序的列表“L”中每种微生物的递增排名的次序逐渐累积,并且其中“j”是整数或分数。此外,在步骤(e)中,重复步骤“c”和“d”以获得针对预定义范围的“i”值和“j”值的ST_i和DT_j的计数,以分别生成两个值集合“ST”和“DT”,其中i和j属于预定义范围内的两个相同的算术级数,并且其中所述算术级数的连续项之间的公差是预定义的整数或预定义的分数值。此外,在步骤(f)中,该方法包括计算集合“DT”和集合“ST”的笛卡尔乘积,以获得包含对应于DT_j值和ST_i值的多个有序对的集合“C”,并且在步骤(g)中,该方法包括通过对集合“C”中的每个有序对而计算(DT_j÷ST_i)来导出集合“DSR”。在获得集合DSR之后,该方法还包括,在步骤(h)中,计算集合“DSR”的分布特征值,其中集合“DSR”的分布特征值包括集合DSR的集中趋势值。此外,在步骤(i)中,该方法包括通过对包括集合“DSR”的集中趋势值的分布特征值应用数学变换来计算“分类构成偏度”值。该方法还包括由硬件处理器基于集合DSR的集中趋势值、集合“DSR”的分布特征值或“分类构成偏度”值中的至少一个来确定怀孕受试者的早产风险。

[0010] 该方法包括经由微生物组表征平台108,通过应用以下一种或多种技术获得微生物组样本的微生物组表征数据,所述一种或多种技术包括:测序技术、显微镜检查、流式细胞术方法、基于体外培养的方法、一种或多种酶或荧光测定、或涉及对来自复杂微生物种群的信号的光谱鉴定和筛选的一种或多种测定,其中所述微生物组表征数据包括经测序的微生物DNA数据、显微镜成像数据、流式细胞术细胞测量数据、体外培养中生长的微生物的菌落计数和细胞表型数据、信号强度数据。

[0011] 该方法还包括通过由硬件处理器使用一种或多种分类学分类技术以计算方式分析微生物组表征数据,以生成微生物组样本的“微生物分类丰度简档”。

[0012] 在另一方面,该方法还包括获得微生物组样本的经测序的微生物DNA数据。获得经测序的微生物DNA数据包括使用至少一种DNA提取技术分离对应于微生物组样本中存在的多种微生物的微生物DNA;并且使用至少一种DNA测序技术对微生物DNA进行测序以获得经测序的微生物DNA数据。该方法还包括通过由硬件处理器使用一种或多种分类学分类技术以计算方式分析经测序的微生物DNA数据,以生成微生物组样本的“微生物分类丰度简档”。

[0013] 在本公开的另一方面,公开了一种用于评估怀孕受试者的早产风险的系统100。系统100包括计算系统102,所述计算系统102还包括硬件处理器202;以及被耦合到硬件处理器202的存储器206。硬件处理器202执行存储在存储器206中的多个模块208。多个模块208包括微生物分类丰度简档生成模块210和计算模块212。计算模块212接收从怀孕受试者获得的微生物组样本的“微生物分类丰度简档”。微生物分类丰度简档包含微生物组样本中存在的多种微生物的丰度值。计算模块212还针对“微生物分类丰度简档”计算“分类构成偏度”值和分布特征值中的至少一个。“分类构成偏度”值和分布特征值量化来自微生物分类丰度简档的多种微生物的丰度值中的偏差。为了计算“分类构成偏度”值和分布特征值,计算模块212在步骤(a)中通过计算来自微生物分类丰度简档的多种微生物的丰度值的总和来获得总种群数“M”。此外,在步骤(b)中,计算模块212创建包含来自微生物分类丰度简档的多种微生物中的每种的丰度值的经排序的列表“L”。经排序的列表“L”包括存在于“微生物分类丰度简档”中按丰度值的递增次序或递减次序排名的多种微生物(即,分类群)中的每种的丰度值。当列表L按丰度值的递增次序排序时,经排序的列表“L”中排名靠前的一个

或多个条目对应于“稀疏分类群”并且经排序的列表“L”中排名垫底的一个或多个条目对应于“优势分类群”，而当列表L按丰度值的递减次序排序时，经排序的列表“L”中排名靠前的一个或多个条目对应于“优势分类群”并且经排序的列表“L”中排名垫底的一个或多个条目对应于“稀疏分类群”。此外，在步骤(c)中，计算模块212计数其累积丰度 \geq 总种群计数“M”的*i*%的“稀疏分类群”的最小数量(ST_i)，其中当列表L按递增次序排序时，丰度值按经排序的列表“L”中每种微生物的递增排名的次序逐渐累积，或者，计数其累积丰度 \geq 总种群计数“M”的*i*%的“稀疏分类群”的最小数量(ST_i)，其中当列表L按递减次序排序时，丰度值按经排序的列表“L”中每种微生物的递减排名逐渐累积，并且其中“*i*”是整数或分数。此外，在步骤(d)中，计算模块212计数其累积丰度 \geq 总种群计数“M”的*j*%的“优势分类群”的最小数量(DT_j)，其中当列表L按递增次序排序时，丰度值按经排序的列表“L”中每种微生物的递减排名的次序逐渐累积，或者，计数其累积丰度 \geq 总种群计数“M”的*j*%的“优势分类群”的最小数量(DT_j)，其中当列表L按递减次序排序时，丰度值按经排序的列表“L”中每种微生物的递增排名逐渐累积，并且其中“*j*”是整数或分数。此外，在步骤(e)中，计算模块212重复执行步骤(c)和(d)，以获得针对预定义范围的“*i*”值和“*j*”值的 ST_i 和 DT_j 的计数，以分别生成两个值集合“ST”和“DT”，其中*i*和*j*属于预定义范围内的两个相同的算术级数，并且其中所述算术级数的连续项之间的公差是预定义的整数或预定义的分数值。此外，在步骤(f)中，计算模块212，计算集合“DT”和集合“ST”的笛卡尔乘积，以获得包含对应于 DT_j 值和 ST_i 值的多个有序对的集合“C”。此外，在步骤(g)中，计算模块212通过针对集合“C”中的每个有序对而计算($DT_j \div ST_i$)来导出集合“DSR”。此外，在步骤(h)中，计算模块212计算集合“DSR”的分布特征值，其中集合“DSR”的分布特征值包括集合DSR的集中趋势值。此外，在步骤(i)中，计算模块212通过对包括集合“DSR”的集中趋势值的分布特征值应用数学变换来计算“分类构成偏度”值。计算模块212基于集合DSR的集中趋势值、集合“DSR”的分布特征值或“分类构成偏度”值中的至少一个来确定怀孕受试者的早产风险。

[0014] 系统100包括微生物组样本接收器110和微生物组表征平台108。微生物组样本接收器110从怀孕受试者接收微生物组样本。微生物组表征平台108包括测序平台112、显微镜平台114、核酸杂交平台116或细胞分选平台118中的至少一个，以从微生物组样本获得/生成微生物组表征数据。

[0015] 微生物组表征平台108从微生物组样本接收器110接收微生物组样本。微生物组表征平台(108)通过对微生物组样本应用以下至少一个技术来从微生物组样本中获得微生物组表征数据，其中所述至少一种技术选自包括以下的技术组：测序技术、显微镜检查、流式细胞术方法、基于体外培养的方法、一种或多种酶或荧光测定、或涉及对来自复杂微生物种群的信号的光谱鉴定和筛选的一种或多种测定。微生物组表征数据包括经测序的微生物核酸数据、显微镜成像数据、流式细胞术细胞测量数据、体外培养中生长的微生物的菌落计数和细胞表型数据、以及信号强度数据。

[0016] 在一个方面，测序平台112从微生物组样本接收器110接收微生物组样本。测序平台112使用至少一种DNA提取技术来分离对应于微生物组样本中存在的多种微生物的微生物DNA。测序平台112还使用至少一种DNA测序技术对微生物DNA进行测序，以获得经测序的微生物DNA数据。微生物分类丰度简档生成模块210使用一种或多种分类学分类技术以计算方式分析经测序的微生物DNA数据，以生成微生物组样本的微生物分类丰度简档。

[0017] 在本公开的另一方面,公开了一种非暂时性计算机可读介质,其体现可由硬件处理器在计算设备中执行以用于评估怀孕受试者的早产风险的程序。该程序包括用于接收从怀孕受试者获得的微生物组样本的“微生物分类丰度简档”的程序代码,其中微生物分类丰度简档包含微生物组样本中存在的多种微生物中的每个的丰度值。该程序还包括用于针对“微生物分类丰度简档”计算“分类构成偏度”值和分布特征值中的至少一个的程序代码,其中“分类构成偏度”值和分布特征值量化来自微生物分类丰度简档的多种微生物的丰度值的偏差。“分类构成偏度”值和分布特征值的计算包括以下步骤(a)至(i)。在步骤(a)中,通过计算来自微生物分类丰度简档的多种微生物的丰度值之和来获得总种群数“M”。此外,在步骤(b)中,创建包含来自微生物分类丰度简档的多种微生物中的每种的丰度值的经排序的列表“L”,其中所述经排序的列表“L”包括存在于所述微生物分类丰度简档中按丰度值的递增次序或递减次序排名的多种微生物中的每种的丰度值,并且其中,当列表L按丰度值的递增次序排序时,经排序的列表“L”中排名靠前的一个或多个条目对应于“稀疏分类群”并且经排序的列表“L”中排名垫底的一个或多个条目对应于“优势分类群”,而当列表L按丰度值的递减次序排序时,经排序的列表“L”中排名靠前的一个或多个条目对应于“优势分类群”并且经排序的列表“L”中排名垫底的一个或多个条目对应于“稀疏分类群”。此外,在步骤(c)中,计数其累积丰度 \geq 总种群计数“M”的*i*%的“稀疏分类群”的最小数量(ST_i),其中当列表L按递增次序排序时,丰度值按经排序的列表“L”中每种微生物的递增排名的次序逐渐累积,或者计数其累积丰度 \geq 总种群计数“M”的*i*%的“稀疏分类群”的最小数量(ST_i),其中当列表L按递减次序排序时,丰度值按经排序的列表“L”中每种微生物的递减排名逐渐累积,并且其中“*i*”是整数或分数。此外,在步骤(d)中,计数其累积丰度 \geq 总种群计数“M”的*j*%的“优势分类群”的最小数量(DT_j),其中当列表L按递增次序排序时,丰度值按经排序的列表“L”中每种微生物的递减排名的次序逐渐累积,或者计数其累积丰度 \geq 总种群计数“M”的*j*%的“优势分类群”的最小数量(DT_j),其中当列表L按递减次序排序时,丰度值按经排序的列表“L”中每种微生物的递增排名逐渐累积,并且其中“*j*”是整数或分数。此外,在步骤(e)中,重复步骤(c)和(d),以获得针对预定义范围的“*i*”值和“*j*”值的 ST_i 和 DT_j 的计数,以分别生成两个值集合“ST”和“DT”,其中*i*和*j*属于预定义范围内的两个相同的算术级数,并且其中所述算术级数的连续项之间的公差是预定义的整数或预定义的分数值。此外,在步骤(f)中,计算集合“DT”和集合“ST”的笛卡尔乘积,以获得包含对应于 DT_j 值和 ST_i 值的多个有序对的集合“C”。此外,在步骤(g)中,通过对集合“C”中的每个有序对而计算 $(DT_j \div ST_i)$ 来导出集合“DSR”。此外,在步骤(h)中,计算集合“DSR”的分布特征值,其中集合“DSR”的分布特征值包括集合DSR的集中趋势值。此外,在步骤(i)中,通过对包括集合“DSR”的集中趋势值的分布特征值应用数学变换来计算“分类构成偏度”值。该程序还包括用于基于集合DSR的集中趋势值、集合“DSR”的分布特征值或“分类构成偏度”值中的至少一个来确定怀孕受试者的早产风险的程序代码。

[0018] 该程序还包括以下程序代码,其用于通过由硬件处理器使用一种或多种分类学分类技术以计算方式分析经测序的微生物DNA数据,以生成微生物组样本的“微生物分类丰度简档”。使用测序平台112从微生物组样本获得经测序的微生物DNA数据。该程序还包括以下程序代码,其用于通过由硬件处理器使用一种或多种分类学分类技术以计算方式分析微生物组表征数据,以生成微生物组样本的“微生物分类丰度简档”,其中所述微生物组表征数

据从微生物组样本获得。通过对微生物组样本应用至少一种技术来从微生物组样本获得微生物组表征数据,其中所述至少一种技术选自包括以下的技术组:测序技术、显微镜检查、流式细胞术方法、基于体外培养的方法、一种或多种酶或荧光测定、或涉及对来自复杂微生物种群的信号的光谱鉴定和筛选的一种或多种测定。微生物组表征数据包括经测序的微生物核酸数据、显微镜成像数据、流式细胞术细胞测量数据、体外培养中生长的微生物的菌落计数和细胞表型数据、以及信号强度数据。

附图说明

[0019] 本领域技术人员应理解,本文的任何框图表示体现本主题的的原理的说明性系统和方法的概念视图。类似地,应理解,任何流程图、流程图表、状态转换图、伪代码等表示可以基本上在计算机可读介质中表示并且因此由计算设备或处理器执行的各种过程,而不管这种计算设备或处理器是否被明确地示出。

[0020] 参照附图从下列详细描述中将更好地理解本文所提到的实施例,其中:

[0021] 图1示出了根据本公开的实施例的用于评估怀孕受试者的早产风险的系统的网络实施方式。

[0022] 图2示出了根据本公开实施例的用于评估怀孕受试者的早产风险的系统。

[0023] 图3示出了根据本公开的实施例的用于针对给定多样性度量的给定值集合计算分离程度(ES)的公式。

[0024] 图4示出了根据本公开一个实施例的用于评估怀孕受试者早产结果的风险的方法。

[0025] 图5-a和图5-b示出了根据本公开的实施例的用于从怀孕受试者的微生物组样本中存在的多种微生物中的每种的丰度值计算“分类构成偏度”(taxonomic composition skew, TCS)值的方法,以进一步确定怀孕受试者的早产风险。

具体实施方式

[0026] 参照在附图中示出并且在下列描述中详细说明的非限制性实施例来更充分地解释本文的实施例及其各种特征和有利细节。本文使用的示例仅旨在促进对以之可以实践本文的实施例的方式的理解,并且进一步使本领域技术人员能够实践本文的实施例。因此,这些示例不应被解释为限制本文的实施例的范围。

[0027] 根据一个实施例,公开了一种用于评估怀孕受试者的早产(PTD)风险的系统和方法。在一个方面,公开了用于定性和定量评估孕妇早产(PTD)风险的系统和方法。更具体地,公开了通过分析从处于妊娠的前15周内或妊娠的前三个月内的怀孕受试者获得的微生物组样本来评估PTD风险的系统和方法。该系统生成微生物组样本的“微生物分类丰度简档”。微生物组样本从怀孕受试者获得,并且“微生物分类丰度简档”包含被识别为存在于微生物组样本中的多种微生物中的每种的丰度值。该系统进一步计算分布特征值(诸如集合DSR的集中趋势值和集合DSR的“分类构成偏度”值),其中集合DSR包括量化对应于微生物组样本中存在的多种微生物的丰度值中的偏差的值(以微生物分类丰度简档形式表示)。此外,系统基于如此计算的分布特征值或“分类构成偏度”值中的至少一个来确定怀孕受试者的早产风险。该系统和方法提供早期风险评估程序,以基于对从处于妊娠的前15周内或妊娠的

前三个月内的怀孕受试者收集到的微生物组样本的分析,准确且可靠地预测怀孕受试者的早产风险。因此,从本文公开的方法和系统实现的风险评估使得怀孕受试者能够很好地及时地采取预防性或矫正性医疗建议或预防性程序,以降低或消除早产风险。

[0028] 根据本公开,示出了本公开的一些目的。主要目的在于提供用于针对怀孕受试者的早产结果的早期风险评估的方法和系统。在一个实施例中,怀孕受试者是孕妇。更具体地,在妊娠的尽可能早的阶段(例如,当怀孕受试者是孕妇时的妊娠的前三个月)针对怀孕受试者的早产结果提供准确的风险评估。

[0029] 本公开的另一目的在于提供用于通过量化从怀孕受试者获得的微生物组样本中的微生物的多样性来对怀孕受试者的早产进行早期风险评估的方法和系统。更具体地,本公开的目的在于通过从对应于从怀孕受试者获得的微生物组样本的微生物分类丰度简档中量化多种微生物的丰度值的偏差来对怀孕受试者的早产进行早期风险评估。

[0030] 本公开的另一目的在于提供一种早期诊断程序,其可以基于从处于妊娠的前15周内或处于妊娠的前三个月内的怀孕受试者收集到的微生物组样本而准确且可靠地预测怀孕受试者的早产风险,其中怀孕受试者是孕妇。

[0031] 本公开的另一目的在于通过根据丰度值量化“分类构成偏度”(TCS)值来评估怀孕受试者的早产(PTD)风险,所述丰度值以微生物分类丰度简档的形式表示,对应于微生物组样本中存在的多种微生物。该方法和系统采用TCS值来执行对早产(PTD)结果风险的定性和定量评估。

[0032] 本公开的另一目的在于通过计算分布特征值(诸如集合DSR的集中趋势值)来评估怀孕受试者的PTD风险,其中集合DSR包括量化对应于微生物组样本中存在的多种微生物的丰度值中的偏差的值(以微生物分类丰度简档形式表示)。该方法和系统采用分布特征值(诸如集合DSR的集中趋势值)来提供对早产(PTD)结果风险的定性和定量评估。

[0033] 本公开的另一目的在于提供一种方法和系统,其在妊娠的前15周内或在妊娠的前三个月内预测怀孕受试者对早产结果的预处置,从而使怀孕受试者能够采取可以可能降低或避免早产结果的风险的预防性或矫正性医疗建议或程序。

[0034] 当结合附图阅读以下描述时,本发明的其他目的和优点将更加显而易见,所述附图并不旨在限制本公开的范围。

[0035] 根据本公开的实施例,参照图1和图2,描述了用于评估怀孕受试者的早产(PTD)风险的系统100。示出了优选实施例,并且在以下示例性系统和/或方法的上下文中描述了这些实施例。

[0036] 然而,应理解,保护的范围被扩展至其中具有消息的这种程序以及计算机可读的装置,这种计算机可读的存储装置包含程序代码装置,其用于当程序在服务器或与移动设备或任何适合的可编程的设备上运行时实施本方法的一个或多个步骤。硬件设备可以是可以被编程的任何种类的设备,包括例如任何种类的像服务器或个人计算机等的计算机,或其任何组合。该设备还可以包括可以是例如像例如专用集成电路(ASIC)、现场可编程门阵列(FPGA)的硬件装置或硬件装置和软件装置的组合,例如,ASIC和FPGA,或者具有硬件模块位于其中的至少一个微处理器和至少一个存储器。因此,该装置可以包括硬件装置和软件装置二者。可以在硬件和软件中实施本文所描述的方法实施例。该设备还可以包括软件装置。可替代地,可以在不同硬件设备上实施这些实施例,例如使用多个CPU。

[0037] 本文的实施例可以包括硬件元件和软件元件。在硬件中实施的实施例包括但不限于固件、常驻软件、微码等。可以在其他模块或其他模块的组合中实施由本文所描述的各种模块执行的功能。为了本描述的目的，计算机可用的或计算机可读的介质可以是包括、存储、通信、传播或传输程序的用于由指令执行系统、装置或设备使用或与其连接的任何装置。

[0038] 介质可以是电、磁、光、电磁、红外或半导体系统(或装置或设备)或传播介质。计算机可读的介质的示例包括半导体或固态存储器、磁带、可移动计算机软磁盘、随机存取存储器(RAM)、只读存储器(ROM)、硬磁盘和光盘。光盘当前的示例包括光盘只读存储器(CD-ROM)、可读写光盘(CD-R/W)和DVD。

[0039] 适于存储和/或执行程序代码的数据处理系统包括通过系统总线而被直接或间接耦合至存储元件的至少一个处理器。存储元件可以包括在程序代码的实际执行期间采用的本地存储器、大容量存储器和提供对至少一些程序代码的临时存储以减少在执行期间必须从大容量存储器检索代码的次数的高速缓冲存储器。

[0040] 虽然所描述的系统和方法的各方面可以被实施在任何数量的不同计算系统、环境和/或配置中，但是在以下示例性系统的上下文中描述了实施例。

[0041] 参照图1，示出了根据本主题的实施例的用于评估怀孕受试者的早产(PTD)风险的系统100的网络实施方式。在一个实施例中，系统100是计算系统102、微生物组表征平台108和微生物组样本接收器110的组合。

[0042] 系统100包括计算系统102，但是在在一个实施例中，考虑到系统100通过计算系统102实施来解释本主题，可以理解的是，计算系统102也可以被实施为多种计算系统，诸如膝上型计算机、台式计算机、笔记本、工作站、大型计算机、服务器、网络服务器、平板电脑、移动电话等。在一个实施方式中，计算系统102可以被实施在基于云的环境中。应当理解，计算系统102可以由多个用户通过一个或多个用户设备104-1, 104-2...104-N(下文中统称为用户104)或驻留在用户设备104上的应用来访问。用户设备104的示例可以包括但不限于便携式计算机、个人数字助理、手持设备和工作站。用户设备104通过网络106被通信地耦合到计算系统102。

[0043] 在一个实施方式中，网络106可以是无线网络、有线网络或其组合。网络106可以被实施为不同类型的网络(诸如内联网、局域网(LAN)、广域网(WAN)、英特网等)中的一个。网络106可以是专用网络或共享网络。共享网络表示使用各种协议(例如，超文本传输协议(HTTP)、传输控制协议/网际协议(TCP/IP)、无线应用协议(WAP)等)来相互通信的不同类型网络的关联。此外，网络106可以包括各种网络设备，包括路由器、网桥、服务器、算设备、存储设备等。

[0044] 现在参照图2，示出了根据本主题的实施例的计算系统102。在一个实施例中，计算系统102可以包括至少一个硬件处理器202、输入/输出(I/O)接口204和存储器206。至少一个硬件处理器202可以被实施为一个或多个微处理器、微计算机、微控制器、数字信号处理器、中央处理单元、状态机、逻辑电路和/或基于操作指令而操纵信号的任何设备。除其他能力外，至少一个硬件处理器202被配置为获取并执行被存储在存储器206中的计算机可读指令或模块208。

[0045] I/O接口204可以包括各种软件和硬件接口，例如，web接口、图形用户接口等。I/O

接口204可以允许系统102直接与用户进行交互或通过客户端设备104与用户进行交互。此外,I/O接口204可以使得计算系统102能够与诸如web服务器和外部数据服务器(未示出)的其他计算设备进行通信。I/O接口204可以促进各种网络和协议类型中的多种通信,包括有线网络(例如LAN,电缆等),以及无线网络(例如WLAN、蜂窝或卫星等)。I/O接口204可以包括一个或多个端口,以用于将多个设备彼此连接或连接到另一服务器。

[0046] 存储器206可以包括本领域中已知的任何计算机可读介质或计算机程序产品,包括例如易失性存储器(诸如静态随机存取存储器(SRAM)和动态随机存取存储器(DRAM))和/或非易失性存储器(诸如只读存储器(ROM)、可擦除可编程ROM、闪存、硬盘、光碟、光盘(CD)、数字多功能盘或数字视频光盘(DVD)和磁带)。存储器206可以包括模块208和数据222。

[0047] 模块208包括执行特定任务或实施特定抽象数据类型的例程、程序、编程指令、对象、组件、数据结构等。在一个实施方式中,模块208可以包括微生物分类丰度简档生成模块210、计算模块212和其他模块220。其他模块220可以包括补充计算系统102的应用和功能的程序或编码指令。

[0048] 除其他外,数据222用作用于存储由模块208和微生物组表征平台108中的一个或多个处理、接收和生成的数据的储存库。数据222还可以包括数据库224和其他数据226。

[0049] 在一个实施例中,系统100包括微生物组表征平台108和微生物组样本接收器110。在一个实施例中,微生物分类丰度简档生成模块210被动态连接到微生物组表征平台108。微生物组表征平台108可以包括测序平台(sequencing platform)112、显微镜平台(114)、核酸杂交平台(116)、细胞分选平台(118)和其他平台(120)中的至少一种。在本公开的一个实施例中,微生物组表征平台108经由微生物组样本接收器110接收从怀孕受试者获得的微生物组样本。

[0050] 在一个实施例中,为了评估怀孕受试者的早产(PTD)的风险,微生物组样本接收器110接收从怀孕受试者获得的微生物组样本。在一个方面,怀孕受试者是孕妇。在另一方面,怀孕受试者可以是任何雌性胎生怀孕动物。在一个方面,其中怀孕受试者是孕妇,可以在妊娠的前15周内或妊娠的前三个月内的任何时间点处获得微生物组样本。在另一方面,其中怀孕受试者是孕妇,可以在妊娠的前28周内从孕妇获得微生物组样本。在又一方面,其中怀孕受试者是孕妇,可以在妊娠的前37周内从孕妇获得微生物组样本。在另一方面,可以在整个妊娠期期间的任何时间点处从怀孕受试者收集微生物组样本,以用于评估PTD结果的风险。在一个方面,微生物组样本是怀孕受试者的阴道拭子样本。从怀孕受试者的口腔、皮肤、肠、阴道或任何其他身体部位获得生物组样本。从怀孕受试者获得的微生物组样本可以是以下中的至少一个:阴道拭子样本、宫颈粘液样本、宫颈拭子样本、阴道拭子(包括阴道穹窿的拭子样本)、尿液样本、羊水样本、血液样本(全血样本)、血清样本、血浆样本、胎盘拭子、脐带拭子、粪便样本、皮肤拭子、口腔拭子、唾液样本、牙周拭子、咽喉拭子、鼻拭子、囊泡液样本、鼻咽拭子、鼻孔拭子、结膜拭子、生殖器拭子、直肠拭子、气管抽吸物以及支气管拭子。

[0051] 此外,在另一实施例中,可以在整个妊娠期期间的不同时间点处从怀孕受试者收集多种微生物组样本。早产具有如现有技术中所描述的含义。一般而言,根据定义,早产是在“胎儿活力”日期和妊娠的第37周末尾之间发生的分娩,其中“胎儿活力”被定义为胎儿在自然出生或诱发生后在子宫外存活的可能性。胎儿活力通常在妊娠的约7个月(28周完成)时进行,但在一些情况下,胎儿活力可能更早发生,甚至在妊娠的24周完成。早产给予婴

儿在子宫中发育较少的时间。

[0052] 在本公开的一个实施例中,微生物组表征平台(108)从微生物组样本接收器110接收微生物组样本,并且从微生物组样本获得微生物组表征数据。微生物组表征平台(108)通过对微生物组样本应用至少一个技术来从微生物组样本中获得微生物组表征数据,其中所述至少一种技术包括测序技术、显微镜检查、流式细胞术方法、基于体外培养的方法、一种或多种酶或荧光测定、或涉及对来自复杂微生物种群的信号的光谱鉴定和筛选的一种或多种测定。微生物组表征数据包括经测序的微生物DNA数据、显微镜成像数据、流式细胞术细胞测量数据、体外培养中生长的微生物的菌落计数和细胞表型数据、以及信号强度数据。本领域技术人员认识到,存在许多技术可用于从微生物组样本中获得微生物组表征数据,并且因此这里不再详细描述。

[0053] 微生物组表征平台(108)包括测序平台(112)、显微镜平台(114)、核酸杂交平台(116)或细胞分选平台(118)中的至少一个,以从微生物组样本获得/生成微生物组表征数据。各种微生物组表征平台在相关的研究领域中广泛实践并且已经被很好地建立,因此没有详细描述。

[0054] 在一个实施例中,微生物组表征平台(108)包括测序平台112,并且测序平台112从微生物组样本接收器110接收微生物组样本,其中微生物组样本从怀孕受试者获得。测序平台110使用至少一种DNA提取技术来分离对应于微生物组样本中存在的多种微生物的微生物DNA。在分离微生物DNA后,测序平台112使用至少一种DNA测序技术对微生物DNA进行测序以获得经测序的微生物DNA数据。测序技术可以选自下一代测序技术。下一代测序技术可以包括全基因组鸟枪法(WGS)测序技术、基于片段库的测序技术、配对库、基于双端库的测序技术或其组合。本领域技术人员认识到存在多个用于DNA测序的方法,并且因此这里不再详细描述。

[0055] 在一个方面,测序平台112根据微生物DNA对细菌16S rRNA标记基因进行扩增和测序,其中微生物DNA提取自微生物组样本。在另一方面,测序平台112可以根据微生物DNA对细菌23S rRNA标记基因进行扩增和测序。测序平台可以使用下一代测序技术来根据微生物DNA对细菌16S rRNA或23S rRNA标记基因的全长或特定可变区域进行测序。此外,在另一方面,可以根据微生物DNA对不同于细菌16S rRNA和23S rRNA标记基因或除细菌16S rRNA和23S rRNA标记基因之外还有的一个或多个系统发育(phylogenetic)标记基因进行扩增和测序。测序平台112将经测序的微生物DNA数据提供给微生物分类丰度简档生成模块210。

[0056] 在另一实施例中,显微镜平台114可以从微生物组样本接收器110接收从怀孕受试者获得的微生物组样本。显微镜平台114被用于执行自动或半自动显微镜检查或成像,其中可以从显微镜成像数据计算接收到的微生物组样本中存在的微生物细胞的数量和类型。

[0057] 在又一实施例中,核酸杂交平台116可以接收从怀孕受试者获得的微生物组样本。核酸杂交平台116从微生物组样本接收器110接收微生物组样本。核酸杂交平台116执行来源于微生物组样本中存在的总微生物核酸含量的短单链核酸序列与预先合成的核酸序列探针阵列(其用作特定微生物分类群的标记)的杂交,其中使用诸如X射线、放射自显影或显色方法的成像技术来捕获和可视化杂交事件,并且其中可以从成像数据计算微生物组样本中存在的微生物细胞的数量和类型。

[0058] 在又一实施例中,细胞分选平台118可以接收从怀孕受试者获得的微生物组样本。

细胞分选平台118从微生物组样本接收器110接收微生物组样本。细胞分选平台118执行基于体外培养的方法、一种或多种酶或荧光测定、或涉及对来自复杂微生物种群的信号的光谱鉴定和筛选的一种或多种测定,以获得微生物组表征数据(诸如在体外培养物中生长的微生物的菌落计数和细胞表型数据以及信号强度数据)。

[0059] 在又一实施例中,其他平台120可以接收从怀孕受试者获得的微生物组样本。其他平台120可以对微生物组样本执行本领域技术人员已知的一种或多种微生物组表征征技术,以获得微生物组表征数据。在又一实施例中,其他平台120可以接收从怀孕受试者获得的微生物组样本。其他平台执行一种或多种微生物组表征技术以获得从其可以计算出接收到的微生物组样本中存在的微生物细胞的数量和类型的数据。

[0060] 在本公开的另一实施例中,微生物分类丰度简档生成模块210通过分析从微生物组表征平台(108)接收到的微生物组表征数据,生成微生物组样本的“微生物分类丰度简档”。微生物分类丰度简档生成模块210由硬件处理器使用一种或多种分类学分类技术以计算方式分析微生物组表征数据,以生成微生物组样本的“微生物分类丰度简档”。计算模块212从微生物分类丰度简档生成模块210接收“微生物分类丰度简档”。

[0061] 根据本公开的另一实施例,微生物分类丰度简档生成模块210由硬件处理器使用一种或多种分类学分类技术以计算方式分析经测序的微生物DNA数据,以生成“微生物分类丰度简档”。“微生物分类丰度简档”包括指示对应于微生物组样本的经测序的微生物DNA数据中存在的多种微生物的各种单独“分类群(taxonomic group)”的丰度的值。微生物分类丰度简档包含微生物组样本中存在的多种微生物中的每种的丰度值。在一个实施例中,使用基于分配的分类学分类(分箱)方法来生成“微生物分类丰度简档”。基于分配的分类学分类(分箱)方法可以涉及将来自经测序的微生物DNA数据的序列与现有的参考序列数据库进行比较和/或经测序的微生物DNA数据与现有的参考序列数据库的构成水平相似性。在一个示例中,可以通过将DNA序列分离成操作分类单位(OTU),以计算方式分析对应于16S rRNA标记基因、23S rRNA标记基因或任何其他系统发育标记基因中的至少一个的经测序的微生物DNA数据,来生成“微生物分类丰度简档”,并且分离成“操作分类单位”(OTU)基于对序列的聚类。对序列的聚类进一步基于操作分类单位(OTU)的序列水平相似性。

[0062] 在另一实施例中,可以通过基于DNA序列的构成相似性对DNA序列进行分离或聚类或者通过将经测序的微生物DNA数据与现有的参考序列数据库进行比较,而以计算方式分析对应于系统发育标记基因或对应于微生物组样本中存在的多种微生物的微生物DNA的全基因组鸟枪法(WGS)序列数据的经测序的微生物DNA数据,来生成“微生物分类丰度简档”。计算模块212从微生物分类丰度简档生成模块210接收“微生物分类丰度简档”。

[0063] 在接收到“微生物分类丰度简档”之后,计算模块212针对“微生物分类丰度简档”计算分布特征值和“分类构成偏度”值。“分类构成偏度”值和分布特征值量化来自微生物分类丰度简档的多种微生物的丰度值中的偏差。下面解释“分类构成偏度”值和分布特征值的计算。

[0064] 为了针对对应于微生物组样本中存在的多种微生物的“微生物分类丰度简档”计算“分类构成偏度”值和分布特征值,在步骤(a)中,计算模块212通过计算来自微生物分类丰度简档的多种微生物的丰度值的总和来获得总种群计数“M”。

[0065] 此外,在步骤(b)中,计算模块212根据“微生物分类丰度简档”创建各种单独分类

群(存在于微生物组样本中)的列表“L”。列表“L”被创建为包含来自微生物分类丰度简档的多种微生物中的每种的丰度值。以下将单独分类群称为与来自微生物组样本的每种微生物相关联的“分类单元/分类群(taxon/taxa)”。此外,在步骤b中,计算模块212将包含来自微生物分类丰度简档的多种微生物的丰度值的列表“L”进行排序以创建经排序的列表L。经排序的列表“L”包括存在于“微生物分类丰度简档”中按丰度值的递增次序或递减次序排名的多种微生物中的每种的丰度值,因此列表L分别按丰度值的递增次序或按丰度值的递减次序进行排序。当列表L按丰度值的递增次序排序时,经排序的列表“L”中的一个或多个排名靠前的条目对应于“稀疏分类群”,并且经排序的列表“L”中的一个或多个排名垫底的条目对应于“优势分类群”。当列表L按丰度值的递减次序排序时,经排序的列表“L”中的一个或多个排名靠前的条目对应于“优势分类群”,并且经排序的列表“L”中的一个或多个排名垫底的条目对应于“稀疏分类群”。

[0066] 在步骤(c)中,计算模块212计数其累积丰度 \geq 总种群计数“M”的*i*%的“稀疏分类群”的最小数量(ST_i),其中当列表L按递增次序排序时,丰度值按经排序的列表“L”中每种微生物的递增排名的次序逐渐累积。否则,计算模块212计数其累积丰度 \geq 总种群计数“M”的*i*%的“稀疏分类群”的最小数量(ST_i),其中当列表L按递减次序排序时,丰度值按经排序的列表“L”中每种微生物的递减排名的次序逐渐累积。这里,“*i*”是整数或分数。

[0067] 此外,在步骤(d)中,计算模块212计数其累积丰度 \geq 总种群计数“M”的*j*%的“优势分类群”的最小数量(DT_j),其中当列表L按递增次序排序时,丰度值按经排序的列表“L”中每种微生物的递减排名的次序逐渐累积。否则,计算模块216计算其累积丰度 \geq 总种群计数“M”的*j*%的“优势分类群”的最小数量(DT_j),其中当列表L按递减次序排序时,丰度值按经排序的列表“L”中每种微生物的递增排名的次序逐渐累积。这里,“*j*”是整数或分数。

[0068] 此外,在步骤(e)中,计算模块212重复执行如上所述的步骤c和d,以获得针对预定义范围的“*i*”值和“*j*”值的 ST_i 和 DT_j 的计数,以分别生成两个值集合“ST”和“DT”。这里,*i*和*j*属于预定义范围内的两个相同的算术级数,并且其中算术级数的连续项之间的公差是预定义的整数或预定义的分数值。在一个示例中,集合ST和集合DT分别被示出在下面的等式(1)和(2)中。

[0069] $ST = \{ST_1, ST_2, ST_3, \dots, ST_{50}\} \dots \dots \dots$ 等式(1),其中,根据定义, ST_i 是非零整数,并且 $ST_{(i+1)} \geq ST_i$,并且“*i*”在1到50的范围内变化(使得 $0 < i \leq 50$)。 $DT = \{DT_1, DT_2, DT_3, \dots, DT_{50}\} \dots \dots \dots$ 等式(2),其中,根据定义, DT_j 是非零整数,并且 $DT_{(j+1)} \geq DT_j$,并且“*j*”在1到50的范围内变化(使得 $0 < j \leq 50$)。

[0070] 在一个方面,“*i*”值和“*j*”值的预定义范围在0到50之间(使得 $0 < i \leq 50$ 且 $0 < j \leq 50$)。在另一方面,“*i*”值和“*j*”值的预定义范围在0到100的范围内(使得 $0 < i \leq 100$ 且 $0 < j \leq 100$)。*i*和*j*的预定义范围被定义为使得*i*+*j*等于100。

[0071] 此外,在步骤(f)中,计算模块212计算集合“DT”和集合“ST”的笛卡尔乘积,以获得包含对应于 DT_j 值和 ST_i 值的多个有序对的集合“C”。

[0072] 在一个示例中,参考来自等式(1)和(2)的集合ST和集合DT,计算集合“ST”和“DT”的笛卡尔乘积以获得集合“C”。因此,参考等式1和2以及步骤(f),如下面的等式(3)所示,集合“C”包含对应于 DT_j 值和 ST_i 值的2500个对(即有序对)。在另一实施例中,集合ST和集合DT的笛卡尔乘积中的有序对的数量可以取决于集合ST和集合DT的大小。

[0073] $C = DT \times ST = \{ (DT_j, ST_i) \mid \text{其中 } DT_j \in DT \text{ 且 } ST_i \in ST \text{ 且 “i” 和 “j” 是正整数 } \leq 50 \}$ ……………等式 (3)

[0074] 在另一实施例中,累积丰度之间的差异(在生成连续的 ST_i 和/或 DT_j 值时计算)可以具有除1之外的任何适当的分数值或整数值。例如,在替代实施方式中,其中i和j的最大值是50,集合ST可以包含100个 ST_i 值(其中 $i = 0.5, 1, 1.5, 2, \dots, 50$)并且集合DT可以包含100个 DT_j 值(其中 $j = 0.5, 1, 1.5, 2, \dots, 50$)。在另一示例中,集合ST和集合DT的笛卡尔乘积中的有序对的数量可以取决于集合ST和集合DT的大小。

[0075] 此外,在步骤(g)中,计算模块212基于来自集合C的值 DT_j 和 ST_i 导出集合“DSR”。在一个实施例中,在步骤(g)中,计算模块212通过针对集合“C”中的每个有序对而计算 $(DT_j \div ST_i)$ 来导出集合“DSR”。如下面的等式(4)所示,导出包括通过针对集合“C”中的每个有序对而计算 $(DT_j \div ST_i)$ 来获得的值的集合“DSR”。因此,如下面的等式(4)所示,集合“DSR”包括通过针对集合“C”中的每个有序对将 DT_j 值除以 ST_i 值而获得的值。

[0076] $DSR = \{ (DT_j \div ST_i) \mid \text{其中 } (DT_j, ST_i) \in C, \text{ 并且 “i” 和 “j” 是正整数 } \leq 50 \}$ ……………等式 (4)

[0077] 此外,在步骤(h)中,计算模块212计算集合“DSR”的分布特征值。集合“DSR”的分布特征值包括集合DSR的集中趋势值。集合DSR的分布特征值可以包括以下中的一个:集合DSR的最大值、集合DSR的最小值、集合DSR的方差、集合DSR的偏度、或者表征集合DSR中的值的分布的模式、对称性、变异性的度量。

[0078] 计算模块212计算统计测度(诸如集合“DSR”的集中趋势值)。集合“DSR”的集中趋势值是平均值、中值、众数或将单个值标识为对集合DSR中的值的整体分布的表示的任何统计测度中的一个。

[0079] 计算模块212通过对分布特征值应用数学变换来计算“分类构成偏度”值。计算模块212通过对包括集合“DSR”的集中趋势值的分布特征值应用数学变换来计算“分类构成偏度”值。在计算集中趋势值之后,在步骤(i)中,计算模块212通过对集合DSR的集中趋势值应用数学变换来计算“分类构成偏度”(TCS)值。应用于集中趋势值的数学变换包括:从1中减去集中趋势值或计算集中趋势值的倒数、或导致针对较低集中趋势值生成较高数值的分类构成偏度(反之亦然)的数学变换方法。“分类构成偏度”(TCS)值使用如下面的等式(5)或等式(6)中所示的以下公式来计算。

[0080] $TCS = 1 - \text{集中趋势值}$ ……………等式 (5) 或

[0081] $TCS = 1 / \text{集中趋势值}$ ……………等式 (6)

[0082] 在另一示例性实施例中,在步骤(i)中,计算模块212通过使用集合DSR的分布特征值来计算TCS值。计算模块212通过对集合DSR的分布特征值应用数学变换来计算TCS值。集合DSR的分布特征值可以包括:集合DSR的最大值、集合DSR的最小值、集合DSR的方差、集合DSR的偏度、或者表征集合DSR中的值的分布的模式、对称性、变异性的度量。

[0083] 在一个示例性实施例中,在微生物组样本中的分类群完全均匀分布的情况下,对应于 $(DT_j \div ST_i)$ 的值可具有最大值1。因此,任何微生物组样本的分类构成偏度(TCS)值可在以下范围内变化: $0 \leq \text{分类构成偏度 (TCS) 值} < 1$ 。任何微生物组样本的集合DSR的集中趋势值可在以下范围内变化: $0 \leq \text{集中趋势值} < 1$ 。

[0084] 此外,计算模块212基于集合DSR的集中趋势值、集合“DSR”的分布特征值或“分类

构成偏度”值中的至少一个来确定怀孕受试者的早产风险。计算模块212还基于集合DSR的集中趋势值、集合“DSR”的分布特征值或“分类构成偏度”值中的至少一个将怀孕受试者分类为针对早产的一个或多个风险类别。一个或多个风险类别包括低早产风险类别、中度早产风险类别和高早产风险类别。

[0085] 在一个示例中,当从孕妇获得阴道微生物组样本时,并且当使用集合DSR的“中值”作为集合DSR的集中趋势值(如步骤“h”中计算的)时,并且当通过使用等式(5)生成TCS值时,并且其中早产风险类别基于基于上述集合DSR的集中趋势值获得的TCS值,对于大于0.922的分类构成偏度值,指示低早产风险类别,对于在0.887至0.922(含两个值)的范围内的分类构成偏度值,指示中度早产风险类别,并且对于小于或等于0.887的分类构成偏度值,指示高早产风险类别。

[0086] 在另一实施例中,可以基于集合DSR的集中趋势值来指示怀孕受试者的早产风险类别。对于大于0.113的集合DSR的集中趋势值,指示高早产风险类别,对于在0.078至0.113(含两个值)范围内的集合DSR的集中趋势值,指示中度早产风险类别,对于小于0.078的集合DSR的集中趋势值,指示低早产风险类别,并且其中集合DSR的集中趋势值是集合DSR的中值,并且其中从孕妇获得的微生物组样本是阴道微生物组样本。

[0087] 下面通过示例描述用于确定上述各种早产风险类别的阈值所采用的程序。仅通过说明提供而非通过限制提供以下实施例。

[0088] 根据一个示例性实施例,基于具有几乎相似的测序覆盖度的三个假设微生物组样本(序列数:每个样本中约49350)计算“分类构成偏度”值所涉及的各种步骤如下所示。下面示出的计算出的示例描绘了通过分析从怀孕受试者获得的微生物组样本中各种微生物的丰度值来计算“分类构成偏度”值/度量所采用的各种步骤。多种微生物的丰度值已经以表格的形式表示(下文称为“微生物分类丰度简档”)。表1(a-c)(下面)描绘了三种微生物组样本的微生物分类丰度简档。

表 1-a

表 1-b

表 1-c

[0089]

样本 1 (S1)		样本 2 (S2)		样本 3 (S3)	
分类群 id	丰度	分类群 id	丰度	分类群 id	丰度
1	1	1	1	1	1
2	1	2	1	2	1
3	1	3	1	3	1
4	1	4	1	4	1
5	76	5	6	5	6
6	241	6	7	6	7
7	745	7	23	7	23
8	1027	8	56	8	56
9	2786	9	110	9	78
10	3598	10	2028	10	226
11	4863	11	3075	11	897
12	6721	12	7354	12	3954
13	7458	13	10326	13	9876
14	9852	14	12456	14	14725
15	11974	15	13879	15	19521
M_(S1) = 49345		M_(S2) = 49324		M_(S3) = 49373	

[0090] 表1 (a-c) 描绘了微生物分类丰度简档,其包含存在于三个微生物组样本 (S1-S3) 中的各种微生物的丰度值 (按每种微生物的丰度的升序排序)。在每个丰度简档下面,还指示“M” (即总种群计数) 的值。通过计算来自微生物分类丰度简档的多种微生物的丰度值的总和来获得M的值。必须指出的是,针对每个分类群ID (微生物) 提供的“丰度”的数值表示: 通过所采用的分类学分类技术分配给该分类群ID的经测序微生物组样本中的DNA序列的绝对计数。

[0091] 从给定的“微生物分类丰度简档”开始 (如表1a-c所示), 以下步骤涉及对“分类构成偏度”值的计算。

[0092] 在步骤1中,通过计算来自微生物分类丰度简档的多种微生物的丰度值的总和来获得总种群计数“M”。表1a-c中示出了对“M”的计算。此外,在步骤2中,创建了包含来自微生物分类丰度简档的多种微生物中的每种的丰度值的经排序的列表“L”。经排序的列表“L”包括存在于微生物分类丰度简档中按丰度值的递增次序或递减次序排名的多种微生物中的每种的丰度值,并且其中,当列表L按丰度值的递增次序排序时,经排序的列表“L”中排名靠前的一个或多个条目对应于“稀疏分类群”,并且经排序的列表“L”中排名垫底的一个或多个条目对应于“优势分类群”,并且当列表L按丰度值的递减次序排序时,经排序的列表“L”中排名靠前的一个或多个条目对应于“优势分类群”,并且经排序的列表“L”中排名垫底的一个或多个条目对应于“稀疏分类群”。在本示例中,表1a-c中指示的3个样本 (S1,S2和S3) 的经排序的列表 (L) 如下所示:

[0093] $L(S1) = \{1, 1, 1, 1, 76, \dots, 9852, 11974\}$

[0094] $L(S2) = \{1, 1, 1, 1, 6, \dots, 12456, 13879\}$

[0095] $L(S3) = \{1, 1, 1, 1, 6, \dots, 14725, 19521\}$

[0096] 在上述集合的每个中,对应的列表L按相应的微生物组样本中各个微生物的丰度值的递增次序排序。因此,一个或多个排名靠前的条目(即,从所描绘的列表的左手侧开始)对应于“稀疏分类群”,并且一个或多个排名垫底的条目(即,从所描绘的列表的右手侧开始)对应于“优势分类群”。这里,L(S1)的“稀疏分类群”对应于 $\{1, 1, 1, 1, 76 \dots\}$,并且L(S1)的“优势分类群”对应于 $\{\dots, 9852, 11974\}$ 。

[0097] 在步骤3中,当丰度值按经排序的列表“L”中的每种微生物的丰度值的递增次序的次序(当列表按递增次序排序时)逐渐累积时,计数其累积丰度 \geq 总种群计数“M”的i%的“稀疏分类群”的最小数量(ST_i)。

[0098] 在步骤3(a)中,考虑i值=1,下面描绘了三个微生物组样本中的每个中包括 \geq 总种群计数“M”的1%的“稀疏分类群”的最小数量(ST_1) (其中经排序的微生物丰度按每种微生物的丰度值的递增次序累积)。

[0099] $S1 = \{1, 1, 1, 1, 76, 241, 745, 1027, 2786, 3598, 4863, 6721, 7458, 9852, 11974\}$ 。

[0100] 在上面的样本中,前7个稀疏分类群(以粗体字体突出显示)的累积丰度(1066)即包括 \geq 总种群计数(即49345)的1%。因此,S1的 $ST_1 = 7$ 。

[0101] 类似地, $S2 = \{1, 1, 1, 1, 6, 7, 23, 56, 110, 2028, 3075, 7354, 10326, 12456, 13879\}$ 。对于S2,前10个稀疏分类群(以粗体字体突出显示)的累积丰度(2234)包括 \geq 总种群计数(即49324)的1%。因此,S2的 $ST_1 = 10$ 。

[0102] 类似地, $S3 = \{1, 1, 1, 1, 6, 7, 23, 56, 78, 226, 897, 3954, 9876, 14725, 19521\}$ 。对于S3,前11个稀疏分类群(以粗体字体突出显示)的累积丰度(1297)包括 \geq 总种群计数(即49373)的1%。因此,S3的 $ST_1 = 11$ 。

[0103] 在步骤3(b)中,类似地,考虑i值=2,下面描绘了三个微生物组样本中的每个中包括总种群计数“M”的 $\geq 2\%$ 的“稀疏分类群”的最小数量(ST_2) (其中经排序的微生物丰度按每种微生物的丰度值的递增次序累积)。

[0104] $S1 = \{1, 1, 1, 1, 76, 241, 745, 1027, 2786, 3598, 4863, 6721, 7458, 9852, 11974\}$ 。

[0105] 在上面的样本中,前7个稀疏分类群(以粗体字体突出显示)的累积丰度(1066)包括 \geq 总种群计数(即49345)的2%。因此,S1的 $ST_2 = 7$ 。

[0106] 类似地, $S2 = \{1, 1, 1, 1, 6, 7, 23, 56, 110, 2028, 3075, 7354, 10326, 12456, 13879\}$ 。对于S2,前10个稀疏分类群(以粗体字体突出显示)的累积丰度(2234)包括 \geq 总种群计数(即49324)的2%。因此,S2的 $ST_2 = 10$ 。

[0107] 类似地, $S3 = \{1, 1, 1, 1, 6, 7, 23, 56, 78, 226, 897, 3954, 9876, 14725, 19521\}$ 。对于S3,前11个稀疏分类群(以粗体字体突出显示)的累积丰度(1297)包括 \geq 总种群计数(即49373)的2%。因此,S3的 $ST_2 = 11$ 。

[0108] 在步骤3(c)中,类似地,考虑i值=3,下面描绘了三个微生物组样本中的每个中包括 ≥ 3 总种群计数“M”的%的“稀疏分类群”的最小数量(ST_3) (其中经排序的微生物丰度按每种微生物的丰度值的递增次序累积)。

[0109] $S1 = \{1, 1, 1, 1, 76, 241, 745, 1027, 2786, 3598, 4863, 6721, 7458, 9852, 11974\}$ 。在

对于每个样本,通过针对相应的集合 C_{S_1} 、集合 C_{S_2} 和集合 C_{S_3} 中的每个有序对而计算 $(DT_j \div ST_i)$ 来导出集合“DSR”。

[0138] 例如,对于样本1,通过针对集合 C_{S_1} 中的每个有序对而计算 $(DT_j \div ST_i)$ 来导出集合 DSR_{S_1} (在表3中表示如下)。

	DT_1	DT_2	DT_3	DT_4	DT_5	DT_6	DT_7	DT_8	DT_{50}
[0139] ST_1	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.4286
ST_2	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.4286
ST_3	0.1250	0.1250	0.1250	0.1250	0.1250	0.1250	0.1250	0.1250	0.3750
ST_4	0.1250	0.1250	0.1250	0.1250	0.1250	0.1250	0.1250	0.1250	0.3750
ST_5	0.1111	0.1111	0.1111	0.1111	0.1111	0.1111	0.1111	0.1111	0.3333
[0140] ST_6	0.1111	0.1111	0.1111	0.1111	0.1111	0.1111	0.1111	0.1111	0.3333
ST_7	0.1111	0.1111	0.1111	0.1111	0.1111	0.1111	0.1111	0.1111	0.3333
ST_8	0.1111	0.1111	0.1111	0.1111	0.1111	0.1111	0.1111	0.1111	0.3333
.....
.....
ST_{50}	0.0769	0.0769	0.0769	0.0769	0.0769	0.0769	0.0769	0.0769	0.2308

[0141] 表3:集合 DSR_{S_1}

[0142] 以类似的方式,通过针对集合 C_{S_2} 和集合 C_{S_3} (分别对应于样本2和样本3)中的每个有序对而计算 $(DT_j \div ST_i)$ 来导出集合 DSR_{S_2} 和集合 DSR_{S_3} 。

[0143] 此外,在步骤7中,针对每个单独的样本计算集合“DSR”的集中趋势值(或分布特征值)。在该示例中,集合DSR的“中值”被认为是集合DSR的集中趋势值。

[0144] 集合 DSR_{S_1} 的集中趋势(中值)值=0.1538

[0145] 集合 DSR_{S_2} 的集中趋势(中值)值=0.0909

[0146] 集合 DSR_{S_3} 的集中趋势(中值)值=0.0769

[0147] 此外,在步骤8中,通过对集中趋势值应用数学变换来计算“分类构成偏度”(TCS)值。在该示例中,数学变换步骤包括从1(即,单位一(unity))减去集中趋势值。

[0148] $TCS = 1 - (\text{集合DSR的集中趋势值})$

[0149] 样本S1的TCS值/度量=1-中值(集合 DSR_{S_1})=1-0.1538=0.8461

[0150] 样本S2的TCS值/度量=1-中值(DSR_{S_2})=1-0.0909=0.909

[0151] 样本S3的TCS值/度量=1-中值(DSR_{S_3})=1-0.0769=0.932

[0152] 表4(下面)描绘了使用三个样本S1、S2和S3的微生物分类丰度简档计算出的集合DSR的“集中趋势值”(集合DSR的中值)、“分类构成偏度”(TCS)值的值,如表4所示。

样本名称	度量名称	
	集中趋势值 (即中值 (DSR))	分类构成偏度值 (TCS) 度量
[0153] S1	0.1538	0.8461
S2	0.0909	0.909
S3	0.0769	0.932

[0154] 表4: 样本S1、S2和S3的集中趋势值 (中值 (DSR)) 和分类构成偏度。

[0155] 此外, 在步骤9中, 应指出的是, 预定的早产风险类别阈值如下-

[0156] “高PTD风险”: 集中趋势值 (中值 (DSR)) > 0.113;

[0157] “中度PTD风险”: $0.078 \leq \text{集中趋势值 (中值 (DSR))} \leq 0.113$; 并且

[0158] “低PTD风险”: 集中趋势值 (中值 (DSR)) < 0.078。

[0159] “高PTD风险”: 分类构成偏度 (TCS) 值 < 0.887;

[0160] “中度PTD风险”: $0.887 \leq \text{分类构成偏度 (TCS) 值} \leq 0.922$; 并且

[0161] “低PTD风险”: 分类构成偏度 (TCS) > 0.922。

[0162] 因此, 在应用这些风险类别后, 如上面的示例中分析的3个样本可分类如下-

[0163] 样本S1 → 高TD风险

[0164] 样本S2 → 中度TD风险

[0165] 样本S3 → 低TD风险

[0166] 如下所示来评估作为在应用中确定早产结果风险上的度量的TCS值的效率。此外, 在示例性实施例中, 可以采用交叉验证策略来评估分类构成偏度 (TCS) 值在使用孕妇的阴道微生物组样本 (在妊娠的15周内取样) 预测早产 (病例) 和足月产 (对照) 结果上的效率。该示例性实施例还可以被用于将TCS值的不同早产风险类别阈值确定为度量。

[0167] 针对对应于具有已知/记录的分娩结果的怀孕受试者/孕妇的微生物组样本计算若干多样性/不等性度量值。将基于给定多样性/不等性度量预测的分娩结果与已知/记录的分娩结果进行比较, 以评估预测效率。对于给定的多样性/不等性度量, 马修斯相关系数 (MCC) 是捕获用于预测分娩结果的度量的特异性和敏感性的测度。为+1的完美MCC值指示对应于从其妊娠进行到记录的早产结果的怀孕受试者获得的微生物组样本以及从其妊娠进行到记录的足月产结果 (即正常的足月妊娠) 的怀孕受试者获得的微生物组样本的多样性/不等性度量值的完全分离。此外, 参照图3, 根据本公开的示例性实施例, 分离程度 (ES) 值指示两个值集合 D_{TD} 和 D_{PTD} 之间的分离程度, 其中 D_{TD} 表示针对来自对应于具有已知/记录的足月产结果的怀孕受试者的样本的给定多样性度量计算的值的范围, 并且其中 D_{PTD} 表示针对来自对应于具有已知/记录的一个或多个早产结果的怀孕受试者的样本的给定多样性度量计算的值的范围。较高的ES值指示值集合之间的更好分离。观察分类构成偏度 (TCS) 值以获得为+1的完美MCC值 (例如在妊娠的第15周之前的每周)。因此, 可以假设在第15孕周之时或之前的妊娠中的任何时间点处从怀孕受试者 (孕妇) 获得的单个阴道微生物组样本可以被用于准确预测早产 (PTD) 结果的风险 (使用TCS值作为度量)。

[0168] 对取自数据语料库 (data corpus) 的阴道微生物组样本 (n=211) 的子集执行验证

研究。如本文所述导出数据语料库。来自三个先前公布的研究的公开可用的阴道微生物组样本数据被用于本评估。在每个研究中，对应于足月产和早产结果的样本数量如下表5所示。

研究	样本总数	足月产样本数	早产样本数	参考文献(DOI)
研究 1	208	165	43	Romero <i>et al.</i> , Microbiome, Feb 2014(DOI:10.1186/2049-2618-2-4)
研究 2	139	139	-	Romero <i>et al.</i> , Microbiome, May 2014(DOI:10.1186/2049-2618-2-18)
研究 3	698	517	181	DiGiulio <i>et al.</i> , PNAS, Sep 2015(DOI: 10.1073/pnas.1502875112)

[0170] 表5: 对应于足月产和早产结果的样本数量

[0171] 获得针对上述阴道微生物组样本数据的分类简档的操作分类单位 (OTU) 水平 (对应于Greengenes OTU版本13.5)。在整个评估过程期间分析对应于总共1045个样本的数据，每个样本具有至少500个分类学分配的序列。计算1045个微生物组样本中的每个的“分类构成偏度” (TCS) 值。除了TCS值之外，还针对可用的微生物组样本计算了以下已建立和广泛使用的生态多样性指数，即分别表示物种多样性、均匀度和丰富度的Shannon、Simpson和Chao。还计算了用于确定经济不等性的统计测度，即基尼系数、Ricci-Schutz、Atkinson、Theil和Decile比率 (90:10)，以检查这些指数是否也能捕获到分类构成分布的偏度。

[0172] 将可用的微生物组样本分成33个周组。组“第N周”包括在妊娠的第N周内的任何时间点处取样的阴道微生物组 (N范围在8-40之间)。每组中的微生物组样本被标记为分别表示报告的“足月”产结果或“早产”结果的“足月产”或“早产”。对于每个组，依据“马修斯相关系数 (MCC)”来估计各个指数区分足月产与早产的诊断值/能力。如下面的等式 (6) 所示，马修斯相关系数 (MCC) 是使用选定的阈值来捕获预测/分类的特异性和敏感性两者的测度。如等式6中所示的术语TP、TN、FP、FN通常被用于计算敏感性/特异性或类似测度，其用于量化分类技术的预测准确性。

[0173]
$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \dots\dots\dots \text{等式 (6)}$$

[0174] 其中，在等式 (6) 中，TP、TN、FP和FN分别表示真阳性预测、真阴性预测、假阳性预测和假阴性预测的数量。

[0175] 为+1的完美MCC值指示对应于早产和足月产的微生物组样本之间的完全分离。对于每个组，通过首先针对给定的指数针对不同的选定的阈值计算“马修斯相关系数 (MCC)”并且随后选择在其处获得最大MCC值的阈值，来估计各个指数用以区分足月产与早产的诊断值/能力。在给定指数 (多样性或不等性测度) 的选定阈值可以完全区分/分离对应于足月产和早产的微生物组样本 (即MCC=+1) 的情况下，分离程度 (ES) 被进一步评估。较高的ES值指示两组微生物组样本 (即足月产样本和早产样本) 之间的更好分离。

[0176] 参照图3，分离程度 (ES) 是示意性地描绘的附加特征，其可以量化生态多样性/经济不等性度量的区分能力。通过使用如下所示的等式7和等式8来计算分离程度 (ES)。

[0177] 分离程度 (ES) =
$$\frac{\min [\delta (\max D_{TD}, \min D_{PTD}), \delta (\max D_{PTD}, \min D_{TD})]}{\max [\delta (\max D_{TD}, \min D_{PTD}), \delta (\max D_{PTD}, \min D_{TD})]} \times 100$$

.....等式 (7)

[0178] 分离程度 (ES) =
$$\frac{B}{A} \times 100$$
.....等式 (8)

[0179] 其中,在等式7中, D_{TD} 表示针对对应于足月产结果的所有样本的给定多样性/不等性度量而计算的集合。 D_{PTD} 表示针对对应于早产结果的所有样本的给定多样性/不等性度量而计算的集合。 $\delta (\max D_{TD}, \min D_{PTD})$ 表示集合 D_{TD} 的最大值与集合 D_{PTD} 的最小值之间的绝对差。 $\delta (\max D_{PTD}, \min D_{TD})$ 表示集合 D_{PTD} 的最大值与集合 D_{TD} 的最小值之间的绝对差。此外,B和A的值可以参照图3。

[0180] 鉴于本公开的主要目的在于提供用于孕妇早产结果的“早期”风险评估的系统和方法,使用仅对应于妊娠的前15周的样本获得的评估结果被示出在表6中。表6提供了各种多样性和不等性指数在根据处于妊娠的前15周内的孕妇的阴道微生物组预测早产结果上的效用的比较评估。表6中所示的数据子集仅包括那些在妊娠15周内获得的阴道微生物组样本。样本被适当地标记为分别表示(报告)“足月”产结果或“早产”结果的“足月产”或“早产”。

[0181] 在表6中,为+1的完美MCC值指示对应于病例(早产)和对照(足月产)的微生物组样本之间的完全分离。在每个妊娠周处指示的样本的数量包括在相应的妊娠周之时或之前的任何时间点处采样的所有阴道微生物组样本。在阈值多样性或不等性测度可以完全区分/分离对应于足月产和早产结果的样本(即MCC=+1)的情况下,分离程度(ES)被进一步评估并且被指示在与MCC值相邻的括号中。较高的ES值指示两组样本之间的分离更好。在每个孕周处,获得的最佳MCC和ES值以粗体字体指示。

[0182] 如表6中所示获得的结果指示出从具有足月产或早产结果的孕妇获得的阴道微生物组样本之间的生态多样性/经济不等性测度的显著差异。观察一个或多个指数以获得为+1的完美MCC值(即,使用选定阈值将与足月产相关联的样本和与早产相关联的样本完全分离),其中在妊娠的15周内(即第8周至第15周)收集微生物组样本。表6示出了各种多样性和不等性指数在根据处于妊娠的前15周内的孕妇的阴道微生物组预测早产结果上的效用的比较性评估。在每个妊娠周处,获得的最佳MCC值以粗体字体指示,并且ES值被指示在与MCC值相邻的括号中。在每个妊娠周处指示的样本的数量包括在相应的妊娠周之时或之前的任何时间点处采样的阴道微生物组样本。

孕周 (第N周)	评估的阴道微生物组样本的数量			使用各种多样性/不等性测度获得的马修斯相关系数 (MCC)								
	早产	足月	总计	TCS	Shannon	Simpson	Chao1	基尼系数	Ricci-Schutz	Atkinson	Theil	Decile 比率 (90:10)
8	11	17	28	1(24.28)	0.929	0.856	1(15.40)	0.856	0.699	0.551	0.804	0.718
9	16	29	45	1(21.96)	0.953	0.812	1(9.24)	0.905	0.658	0.612	0.858	0.626
10	22	41	63	1(8.39)	0.966	0.797	1(4.36)	0.897	0.720	0.660	0.863	0.540
11	29	64	93	1(7.64)	0.926	0.721	1(2.89)	0.925	0.773	0.749	0.901	0.532
12	39	87	126	1(1.68)	0.871	0.719	0.964	0.890	0.776	0.720	0.871	0.518
13	50	105	155	1(1.68)	0.868	0.690	0.971	0.883	0.779	0.732	0.869	0.542
14	61	133	194	1(1.68)	0.822	0.635	0.976	0.882	0.771	0.720	0.870	0.548
15	62	149	211	1(1.68)	0.817	0.635	0.977	0.887	0.780	0.731	0.875	0.563

[0184] 表6:各种多样性和不等性指数在根据处于妊娠的前15周内的孕妇的阴道微生物

组预测早产结果上的效用的比较评估。

[0185] 如表6中所示,在所比较的指数中,观察到如本公开中公开的作为度量的“分类构成偏度”(TCS)值优于所有其他指数。此外,观察TCS值以获得为+1的完美MCC值(每周,直到妊娠的第15周为止),从而表明了在第15孕周之时或之前的妊娠的任何时间点处从孕妇获得的单个微生物组样本(例如阴道微生物组样本)可以被用于以有效和可靠的方式准确地诊断或预测早产结果的风险。更重要的是,针对各种度量/指数获得的ES(分离程度)值的比较(在获得为+1的MCC值的情况下)指示出以TCS值获得的ES比以其他比较的度量/指数获得的ES高5-9%。较高的ES值指示出关于度量/指数的区分能力的较高置信度。虽然原则上,从个体阴道微生物组样本(在第15孕周之前获得)的分类简档中获得的TCS值可以帮助确定早产风险,但在现实环境中,只要有可能,可以理想的是,在发起任何预防性、治疗性或咨询性措施之前,从相同受试者获得多个样本(例如每周)以确定早产风险。

[0186] 观察TCS值以获得为+1的完美MCC值(例如每周,直到妊娠的第15周为止)。因此,可以假设在第15孕周之时或之前的妊娠中的任何时间点处从受试者(孕妇)获得的单个阴道微生物组样本可以被用于准确地预测早产(PTD)结果的风险(使用TCS值作为度量)。

[0187] 执行重复的随机子采样交叉验证实验,以进一步评估TCS的效率,以及预测妊娠结果的各种多样性和不等性指标。这些实验还被用于确定TCS度量的早产风险类别阈值。对于一个或多个交叉验证实验,随机选择三分之二的样本,并且随机选择的样本用作“训练集”,以用于确定提供了对应于早产(病例)结果的微生物组样本与足月产(对照)结果之间的最大分离的TCS值的阈值(依据MCC值进行量化)。随后使用包括剩余的三分之一的微生物组样本的“测试集”评估如此确定的阈值的效率。迭代上述程序1000次。在每次迭代中,在六个参数(即准确性、敏感性、特异性,阳性预测值(PPV)、阴性预测值(NPV)和马修斯相关系数(MCC))方面对TCS值的阈值的效率进行评估。在下面的表7中提供了1000次迭代中的六个评估参数的平均值。表7还包括以其他“生态多样性/经济不等性”度量(包括Shannon、Simpson、Chao1、基尼系数、Ricci-Schutz、Atkinson、Theil、Decile比率)执行的一个或多个相同交叉验证实验的结果。

[0188] 表7示出了交叉验证实验的结果(阴道微生物组样本取自妊娠15周内的孕妇),提供了各种评估度量在准确性、敏感性、特异性、阳性预测值(PPV)、阴性预测值(NPV)和马修斯相关系数(MCC)方面的效率的比较。 $PPV = TP / (TP + FP)$; $NPV = TN / (TN + FN)$; $准确性 = (TP + TN) / (TP + TN + FP + FN)$; $敏感性 = TP / (TP + FN)$; $特异性 = TN / (TN + FP)$ 。在表7中提供了以每个度量的1000次交叉验证迭代中的六个评估参数的平均值,并且以粗体字体指示最佳的获得值。标准偏差值被指示在相应的平均值下方的括号中。表7中的值已四舍五入到小数点后第二位。

度量	准确性	敏感性	特异性	阳性预测值 (PPV)	阴性预测值 (NPV)	马修斯相 关系数
TCS	1.00	0.99	1.00	0.99	1.00	0.99
	(0.01)	(0.03)	(0.01)	(0.02)	(0.01)	(0.02)
Shannon	0.91	0.85	0.94	0.86	0.94	0.79
	(0.03)	(0.07)	(0.03)	(0.07)	(0.03)	(0.06)
Simpson	0.84	0.52	0.98	0.91	0.83	0.60
	(0.04)	(0.09)	(0.02)	(0.08)	(0.04)	(0.08)
Chao1	0.98	0.96	0.98	0.96	0.98	0.95
	(0.02)	(0.04)	(0.02)	(0.05)	(0.02)	(0.04)
基尼系数	0.94	0.84	0.98	0.95	0.94	0.85
	(0.02)	(0.07)	(0.03)	(0.07)	(0.03)	(0.06)
Ricci-Schutz	0.90	0.72	0.97	0.91	0.89	0.74
	(0.03)	(0.09)	(0.03)	(0.09)	(0.04)	(0.07)
Atkinson	0.87	0.69	0.94	0.84	0.88	0.68
	(0.03)	(0.10)	(0.05)	(0.11)	(0.04)	(0.08)
Theil	0.93	0.81	0.98	0.95	0.93	0.83
	(0.03)	(0.08)	(0.04)	(0.08)	(0.03)	(0.07)
Decile 比率	0.82	0.40	0.99	0.95	0.80	0.53
	(0.04)	(0.09)	(0.01)	(0.08)	(0.04)	(0.08)

[0189] 表7:交叉验证实验的结果(其中阴道微生物组样本取自妊娠的15周内的孕妇)

[0191] 从表7中显示的数据可以看出,用作度量的TCS值获得了显著更高的准确性、敏感性、特异性、PPV和NPV值,这清楚地指示TCS值在根据在妊娠的15周内取出的单个阴道微生物组样本而准确诊断/预测早产风险中的效用。

[0192] 根据示例性实施例,采用以下程序来定义针对早产“风险分类”的TCS值的阈值。根据一个示例,在如上所述的交叉验证实验中,从1000个随机选择的训练集获得的TCS值(度量)(其在1000次迭代中的每次中提供最高的MCC值)的最小阈值和最大阈值被识别为0.887和0.922。这些最小TCS阈值和最大TCS阈值被考虑用于定义三种早产“风险类别”。这些类别旨在为孕妇提供临床环境中早产(PTD)结果风险的定性评估。类别详述如下:TCS<0.887→“高PTD风险”, $0.887 \leq TCS \leq 0.922$ →“中度PTD风险”和TCS>0.922→“低PTD风险”。

[0193] 在本公开的另一实施例中,早产风险类别模块可以利用适合于以下的不同组的预定义风险类别阈值:1)在妊娠期间的不同时间点处收集的一个或多种微生物组样本,2)从除了阴道以外的身体部位收集的一个或多种微生物组样本,以及3)从属于不同地理区域、种族的受试者或者以各种早产结果的产妇风险因素进行评估的受试者收集的一个或多种微生物组样本。可以通过分析从相关群组(cohort)获得的微生物组数据来计算这种预定义的风险分类阈值。

[0194] 根据实施例,本公开的一个或多个优点如本文所公开。然而,本文所描述的本公开的优点不限于下面提到的优点。

[0195] 在一个实施例中,该系统和方法评估妊娠的15周内或早于15周的早产结果的风险,并且以显著高的阳性预测值 (PPV) 准确预测早产风险或倾向。

[0196] 在另一实施例中,该系统和方法在妊娠期间的任何时间评估早产结果的风险,并且以显著高的阳性预测值 (PPV) 准确预测早产风险或倾向。

[0197] 该系统和方法使能够基于从微生物组样本的微生物分类丰度简档获得的集合“DSR”的分布特征值或基于微生物组样本的微生物分类丰度简档计算出的分类构成偏度值中的至少一个来确定怀孕受试者的早产风险。集合“DSR”的分布特征值包括集合DSR的集中趋势值。

[0198] 该系统和方法使能够基于分布特征值 (诸如从微生物组样本的微生物分类丰度简档获得的集合“DSR”的集中趋势值) 或者分类构成偏度值中的至少一个的对怀孕受试者的早产风险的确定在预测怀孕受试者的足月产或早产上示出足够的分离程度,因此集合“DSR”的集中趋势值、分布特征值和分类构成偏度具有预测准确性以及在怀孕受试者的足月产和早产上更好的区分能力。

[0199] 通过本文公开的方法和系统获得的高PPV显著地使与不具有这种早产风险的怀孕受试者相比具有潜在早产风险的怀孕受试者错误分类的一个或多个机会最小化。

[0200] 此外,本文公开的系统和方法为孕妇/受试者 (检测到具有PTD结果的风险) 提供足够的时间来采取减少或消除PTD风险的所需“预防性/矫正性”医疗建议或程序。

[0201] 作为基于微生物组的方法,本公开甚至可应用于第一次怀孕或者不是由阴道感染或胎儿或子宫异常或先前已知的风险因素 (例如先前的早产史) 驱动的早产结果。

[0202] 本公开中公开的采样方法和系统包括非侵入性样本收集技术,因此本文公开的用于评估早产风险的方法和系统对于怀孕受试者更舒适。此外,单个样本足以在本文公开的方法中进行准确诊断。

[0203] 此外,该系统和方法需要在妊娠的前15周内的任何时间获得样本,并且与现有技术采用的采样技术 (其在某些情况下是侵入性的,并且通常适用于妊娠的后期) 相比,由采样程序引起的怀孕受试者所经历的任何不适都显著降低。

[0204] 一些现有技术提出细菌性阴道病 (BV) 作为PTD的诱发因素。因此,一些现有技术方法涉及监视阴道拭子中“特定”细菌病原体的丰度或存在,以预测PTD的倾向。本公开指示涉及研究从怀孕受试者获得的微生物组样本的预测方法,但在几个方面不同于这些现有技术。本公开内容不仅仅基于特定细菌病原体的丰度变化来预测或评估PTD风险。本文公开的用于评估PTD结果风险的系统和方法基于的是微生物组样本中存在的种群水平的微生物群落的特征 (不管受试者是否具有BV)。

[0205] 使用来自本文给出的以下出版物的公开可获得的阴道微生物组数据来测试和验证用于评估怀孕受试者的早产结果风险的系统和方法所述出版物包括Romero et al., Microbiome, May 2014, Romero et al., Microbiome, February 2014, 以及D B DiGiulio et al., PNAS, September 2015。

[0206] 现在参照图4,示出了根据本主题的实施例的用于怀孕受试者的早产结果的早期风险评估的方法400。可以在计算机可执行指令的广义语境中描述方法400。通常,计算机可执行指令可以包括执行特定功能或实施特定抽象数据类型的例程、程序、对象、组件、数据结构、过程、模块、功能等。方法400还可以在分布式计算环境中实践,其中功能由通过通信

网络链接的远程处理设备执行。在分布式计算环境中,计算机可执行指令可以位于本地和远程计算机存储介质(包括存储器存储设备)中。

[0207] 描述方法400的次序不旨在被解释为限制,并且可以以任何次序组合任何数量的所描述的方法框以实施方法400或替代方法。此外,单独块可以从方法400中删除而不脱离本文所描述的主题的精神和范围。此外,该方法可以被实施在任何适合的硬件、软件、固件或其组合中。然而,为了便于解释,在下面描述的实施例中,可以认为方法400被实施在上述系统100中。

[0208] 在步骤402处,可以获得来自怀孕受试者的微生物组样本。来自怀孕受试者的微生物组样本可以在妊娠的前15周内或在妊娠的前三个月内或在妊娠的第二个三个月内获得,其中怀孕受试者是孕妇。在另一方面,可以在妊娠的前28周内或在妊娠的前37周内接收来自怀孕受试者的微生物组样本,其中怀孕受试者是孕妇。可以在妊娠期间的任何时间点处接收来自怀孕受试者的微生物组样本。在一个实施例中,怀孕受试者是孕妇。在另一实施例中,怀孕受试者是雌性胎生动物。在另一实施例中,可以在整个妊娠期的任何时间点处从怀孕受试者收集微生物组样本,以用于评估怀孕受试者的PTD结果的风险。

[0209] 在一个实施例中,微生物组样本接收器110可以从怀孕受试者收集或接收微生物组样本。微生物组样本可以从怀孕受试者的口腔、皮肤、肠、阴道或其他身体部位获得。微生物组样本选自包括以下的组:阴道拭子样本、宫颈粘液样本、宫颈拭子样本、阴道拭子(包括穹窿的拭子样本)、尿液样本、羊水样本、血液样本、血清样本、血浆样本、胎盘拭子、脐带拭子、粪便样本、皮肤拭子、口腔拭子、唾液样本、牙周拭子、咽喉拭子、鼻拭子、囊泡液样本、鼻咽拭子、鼻孔拭子、结膜拭子、生殖器拭子、直肠拭子、气管抽吸物以及支气管拭子。

[0210] 此外,在一个实施例中,步骤404处的方法包括经由微生物组表征平台108获得微生物组样本的微生物组表征数据。通过应用以下一种或多种技术获得微生物组表征数据,所述一种或多种技术包括:测序技术、显微镜检查、流式细胞术方法、基于体外培养的方法、一种或多种酶或荧光测定、或涉及对来自复杂微生物种群的信号的光谱鉴定和筛选的一种或多种测定,并且其中所述微生物组表征数据包括经测序的微生物DNA数据、显微镜成像数据、流式细胞术细胞测量数据、体外培养中生长的微生物的菌落计数和细胞表型数据、以及信号强度数据。

[0211] 在一个实施例中,在接收微生物组样本后,在步骤404处,经由针对微生物组样本的测序平台112获得经测序的微生物DNA数据。获得经测序的微生物DNA数据包括:使用至少一种DNA提取技术分离对应于微生物组样本中存在的多种微生物的微生物DNA。在一个实施例中,测序平台106使用至少一种DNA提取技术分离对应于微生物组样本中存在的多种微生物的微生物DNA。

[0212] 此外,在步骤404中,使用至少一种DNA测序技术对微生物DNA进行测序,以获得经测序的微生物DNA数据。在一个实施例中,测序平台112使用至少一种DNA测序技术对微生物DNA进行测序,以获得经测序的微生物DNA数据。在一个方面,根据微生物DNA对细菌16S rRNA或23S rRNA标记基因进行扩增和测序。在另一方面,根据微生物DNA对细菌16S rRNA或23S rRNA标记基因的全长或一个或多个特定区域进行扩增和测序。在一个实施例中,测序平台212根据从微生物组样本中提取的微生物DNA对细菌16S rRNA或23S rRNA标记基因进行扩增和测序。

[0213] 在一个实施例中,可以根据微生物DNA对一种或多种系统发育标记基因进行扩增和测序。DNA测序技术可选自下一代测序技术,其中所述下一代测序技术包括全基因组鸟枪法(WGS)测序、基于片段库的测序技术、配对库或基于双端库的测序技术、或其组合。

[0214] 此外,在一个实施例中,在获得微生物组表征数据之后,在步骤406中,可以通过由硬件处理器使用一种或多种分类学分类技术以计算方式分析微生物组表征数据,从而生成微生物组样本的“微生物分类丰度简档”。

[0215] 在又一实施例中,在步骤406中,通过由硬件处理器使用一种或多种分类学分类技术以计算方式分析经测序的微生物DNA数据,从而生成微生物组样本的“微生物分类丰度简档”。“微生物分类丰度简档”包括微生物组样本中存在的各种“分类群”的丰度值。在一个实施方式中,微生物分类丰度简档生成模块210由硬件处理器使用一种或多种分类学分类技术以计算方式分析经测序的微生物DNA数据,以生成微生物组样本的“微生物分类丰度简档”。在本公开的另一方面,可以通过使用一种或多种技术分析微生物组样本来生成微生物组样本的微生物分类丰度简档,所述一种或多种技术包括显微镜检查、流式细胞术方法、基于体外培养的方法、一种或多种酶或荧光测定、或涉及对来自复杂微生物种群的信号的光谱鉴定和筛选的一种或多种测定。

[0216] 在生成“微生物分类丰度简档”之后,在步骤408中,计算模块212接收“微生物分类丰度简档”。微生物分类丰度简档包含微生物组样本中存在的多种微生物中的每种的丰度值。此外,在步骤408中,计算针对“微生物分类丰度简档”的分布特征值和“分类构成偏度”值。“分类构成偏度”值和分布特征值量化了来自微生物分类丰度简档的多种微生物的丰度值中的偏差。“分类构成偏度”值和分布特征值的计算包括以下步骤。在步骤408中计算“分类构成偏度”值和分布特征值还包括如下所述的多个步骤。步骤408还包括步骤502-518,如图5-a和图5-b所示。

[0217] 参照图5-a和图5-b,这里解释步骤502-518。在步骤502中,通过计算来自微生物分类丰度简档的多种微生物的丰度值的总和来获得总种群计数“M”。此外,在步骤504中,创建包含来自微生物分类丰度简档的多种微生物的丰度值的经排序的列表“L”。经排序的列表“L”包括存在于“微生物分类丰度简档”中按丰度值的递增次序或递减次序排名的多种微生物中的每种的丰度值。当列表L按丰度值的递增次序排序时,经排序的列表“L”中排名靠前的一个或多个条目对应于“稀疏分类群”并且经排序的列表“L”中排名垫底的一个或多个条目对应于“优势分类群”,而当列表L按丰度值的递减次序排序时,经排序的列表“L”中排名靠前的一个或多个条目对应于“优势分类群”并且经排序的列表“L”中排名垫底的一个或多个条目对应于“稀疏分类群”。

[0218] 此外,在步骤506中,计数以下“稀疏分类群”的最小数量(ST_i),其累积丰度 \geq 总种群计数“M”的*i*%。当列表L按递增次序排序时,丰度值按经排序的列表“L”中每种微生物的递增排名的次序逐渐累积,或者计数以下“稀疏分类群”的最小数量(ST_i),其累积丰度 \geq 总种群计数“M”的*i*%,其中当列表L按递减次序排序时,丰度值按经排序的列表“L”中每种微生物的递减排名逐渐累积。“i”是整数或分数。在一个方面,“i”值和“j”值的预定义范围在0到50之间(使得 $0 < i \leq 50$ 且 $0 < j \leq 50$)。在另一方面,“i”值和“j”值的预定义范围在0与100之间(使得 $0 < i \leq 100$ 且 $0 < j \leq 100$)。

[0219] 此外,在步骤508中,计数以下“优势分类群”的最小数量(DT_j),其累积丰度 \geq 总种

群计数“M”的j%，其中当列表L按递增次序排序时，丰度值按经排序的列表“L”中每种微生物的递减排名的次序逐渐累积，或者计数以下“优势分类群”的最小数量(DT_j)，其累积丰度≥总种群计数“M”的j%，其中当列表L按递减次序排序时，丰度值按经排序的列表“L”中每种微生物的递增排名逐渐累积。“j”是整数或分数。

[0220] 此外，在步骤510中，重复步骤506和508以获得针对预定义范围的“i”值和“j”值的ST_i和DT_j的计数，以分别生成两个值集合“ST”和“DT”，其中i和j属于预定义范围内的两个相同的算术级数。算术级数的连续项之间的公差是预定义的整数或预定义的分数值。

[0221] 此外，在步骤512中，计算集合“DT”和集合“ST”的笛卡尔乘积以获得包含对应于DT_j值和ST_i值的多个有序对的集合“C”。

[0222] 此外，在步骤514中，通过针对集合“C”中的每个有序对而计算(DT_j÷ST_i)来导出集合“DSR”。通过针对一系列i值和j值，将每个DT_j项除以每个ST_i项来导出集合“DSR”。i值和j值在预定义的范围内变化。i和j的预定义范围被定义为使得i+j等于100。

[0223] 此外，在步骤516中，计算集合“DSR”的分布特征值。在一个示例中，集合“DSR”的分布特征值包括集合DSR的集中趋势值。集合DSR的分布特征值还包括：集合DSR的最大值、集合DSR的最小值、集合DSR的方差、集合DSR的偏度、或者表征集合DSR中的值的分布的模式、对称性、变异性的度量。在一个方面，集合“DSR”的集中趋势值是以下中的一个：平均值、中值、众数或将单个值标识为对集合DSR中的值的整体分布的表示的任何统计测度。在本公开的一个方面，基于“微生物分类丰度简档”计算集合“DSR”的分布特征值，并且基于集合“DSR”的分布特征值确定怀孕受试者的早产风险。

[0224] 在计算出集合DSR的分布特征值之后，在步骤518中，通过对分布特征值应用数学变换来计算“分类构成偏度”值。通过对包括集合“DSR”的集中趋势值的分布特征值应用数学变换来计算“分类构成偏度”值。通过对集合DSR的分布特征值应用数学变换来计算“分类构成偏度”值，其中集合DSR的分布特征值包括以下中的一个：集合DSR的最大值、集合DSR的最小值、集合DSR的方差、集合DSR的偏度、或者表征集合DSR中的值的分布的模式、对称性、变异性的度量。应用于集中趋势值的数学变换包括：从单位一(unity)中减去集中趋势值、计算集中趋势值的倒数、或导致针对集合DSR的较低集中趋势值生成较高数值的分类构成偏度(反之亦然)的数学变换方法。

[0225] 此外，参照图4，在步骤410中，由硬件处理器基于集合DSR的集中趋势值、集合DSR的分布特征值或“分类构成偏度”值中的至少一个来确定怀孕受试者的早产风险。

[0226] 此外，在步骤412中，确定怀孕受试者的早产风险还包括将怀孕受试者分类为针对早产的一个或多个风险类别，其中分类基于集合DSR的集中趋势值、集合DSR的分布特征值或“分类构成偏度”值中的至少一个。一个或多个风险类别包括低早产风险类别、中度早产风险类别和高早产风险类别。

[0227] 当微生物组样本是从孕妇获得的阴道微生物组样本时，对于大于0.922的“分类构成偏度”值，指示低早产风险类别，对于在0.887至0.922的范围内的“分类构成偏度”值，指示中度早产风险类别，并且对于小于0.887的“分类构成偏度”值，指示高早产风险类别，并且其中“分类构成偏度”值通过从单位一减去计算出的集合DSR的中值而获得。

[0228] 当微生物组样本是从孕妇获得的阴道微生物组样本时，对于大于0.113的集合DSR的集中趋势值，指示高早产风险类别，对于在0.078至0.113范围内的集合DSR的集中趋势

值,指示中度早产风险类别,对于小于0.078的集合DSR的集中趋势值,指示低早产风险类别,并且其中集合DSR的集中趋势值是集合DSR的中值。

[0229] 在一个实施方式中,在步骤408中,对应于计算集合DSR的“分类构成偏度”值和分布特征值的步骤502-518由计算模块212执行。

[0230] 已经参照各种实施例呈现了前面的描述。本申请所属领域的普通技术人员将理解,在不有意脱离本发明的原理、精神和范围的情况下,可以对所描述的结构和操作方法进行变更和改变。

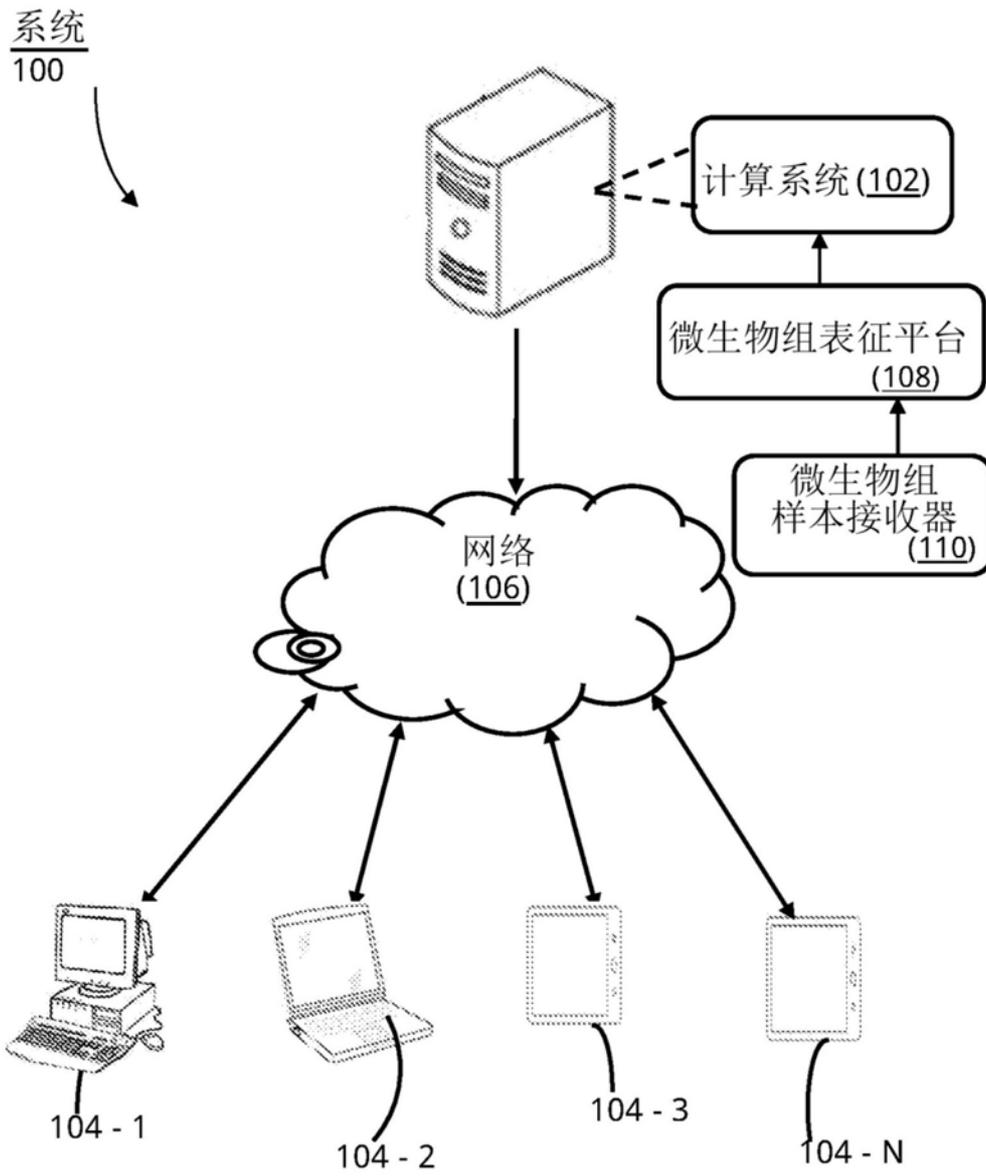


图1

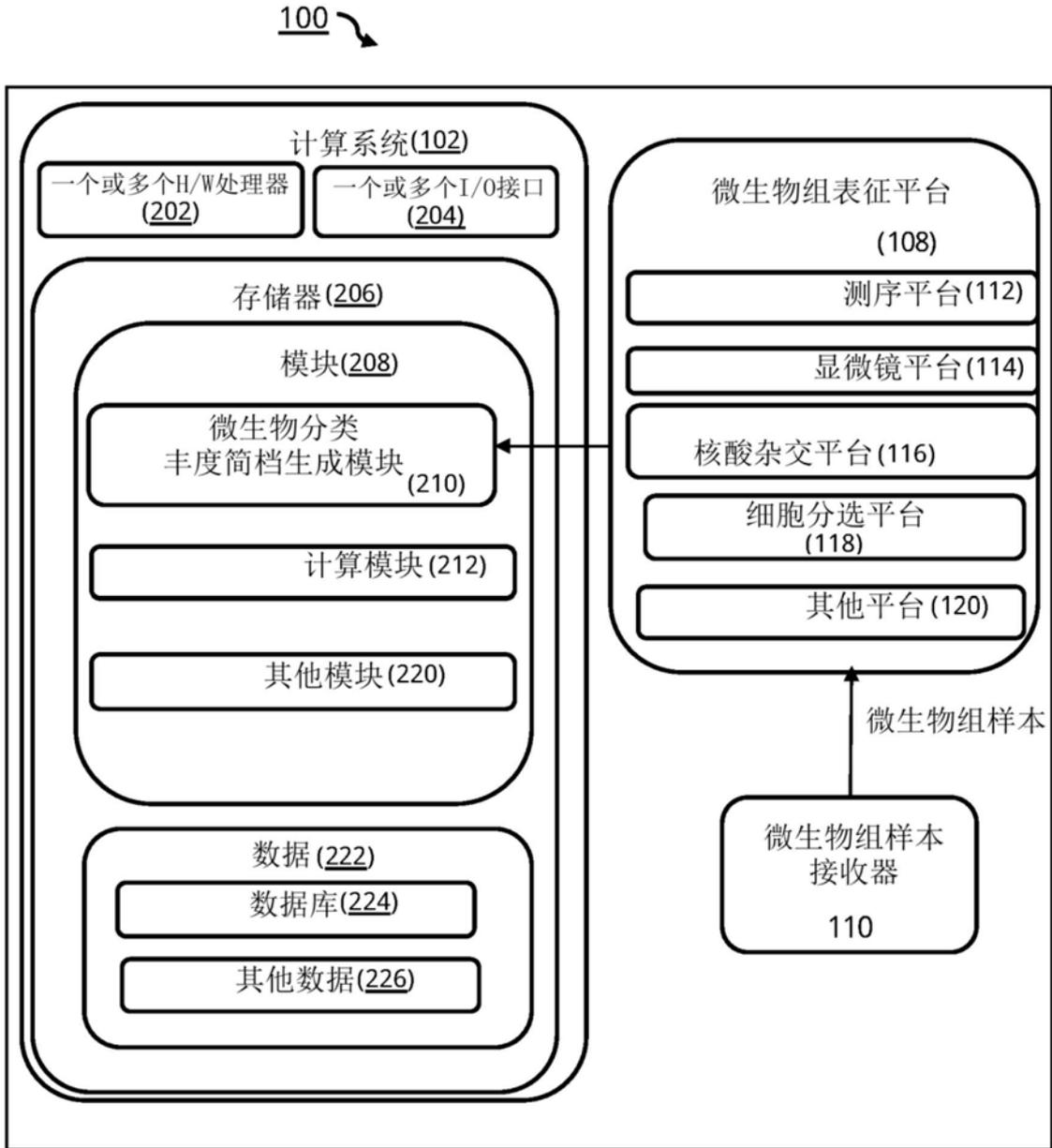


图2

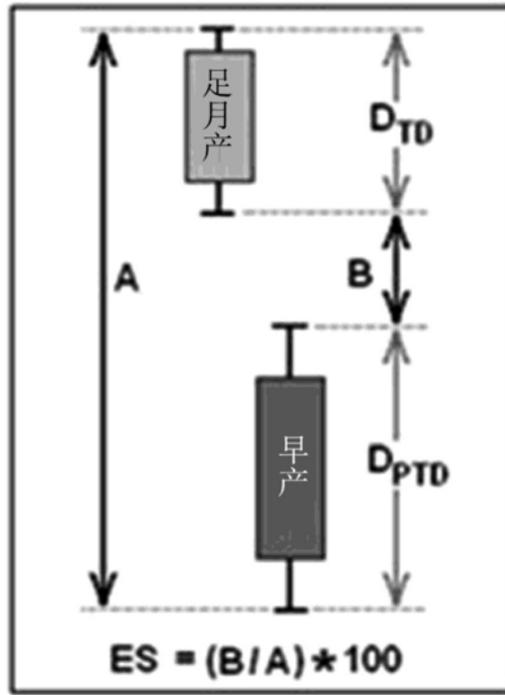


图3

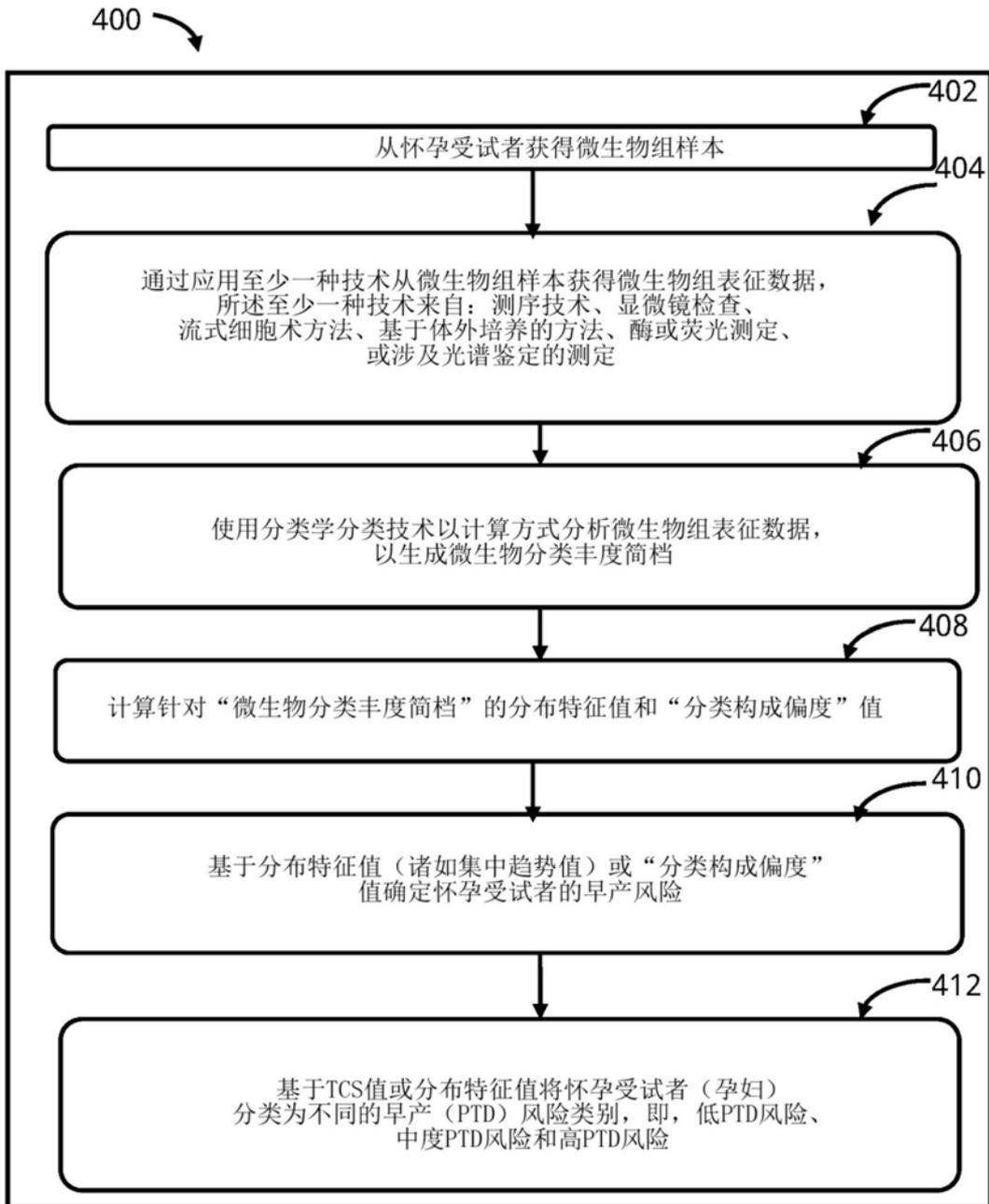


图4

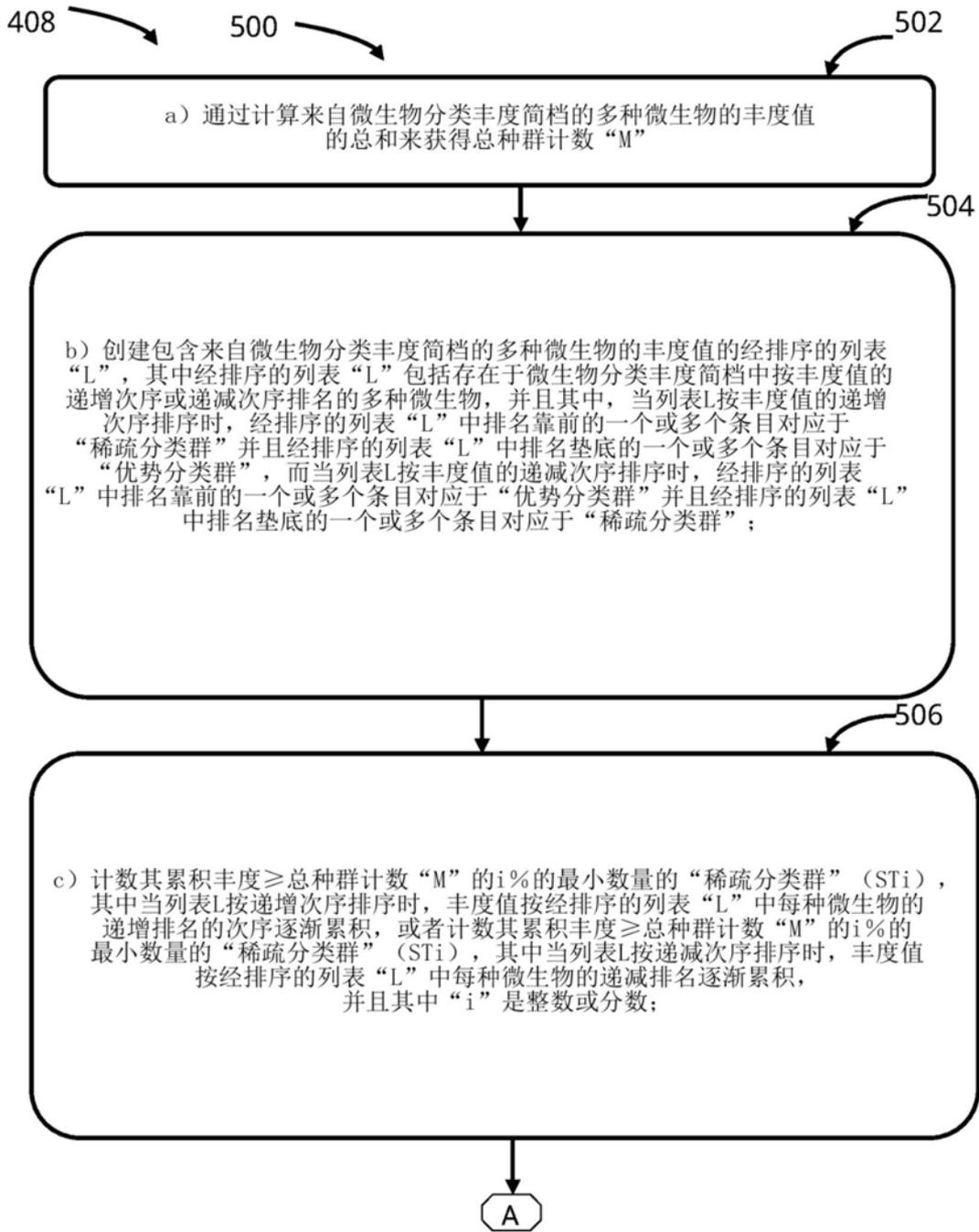


图5-a

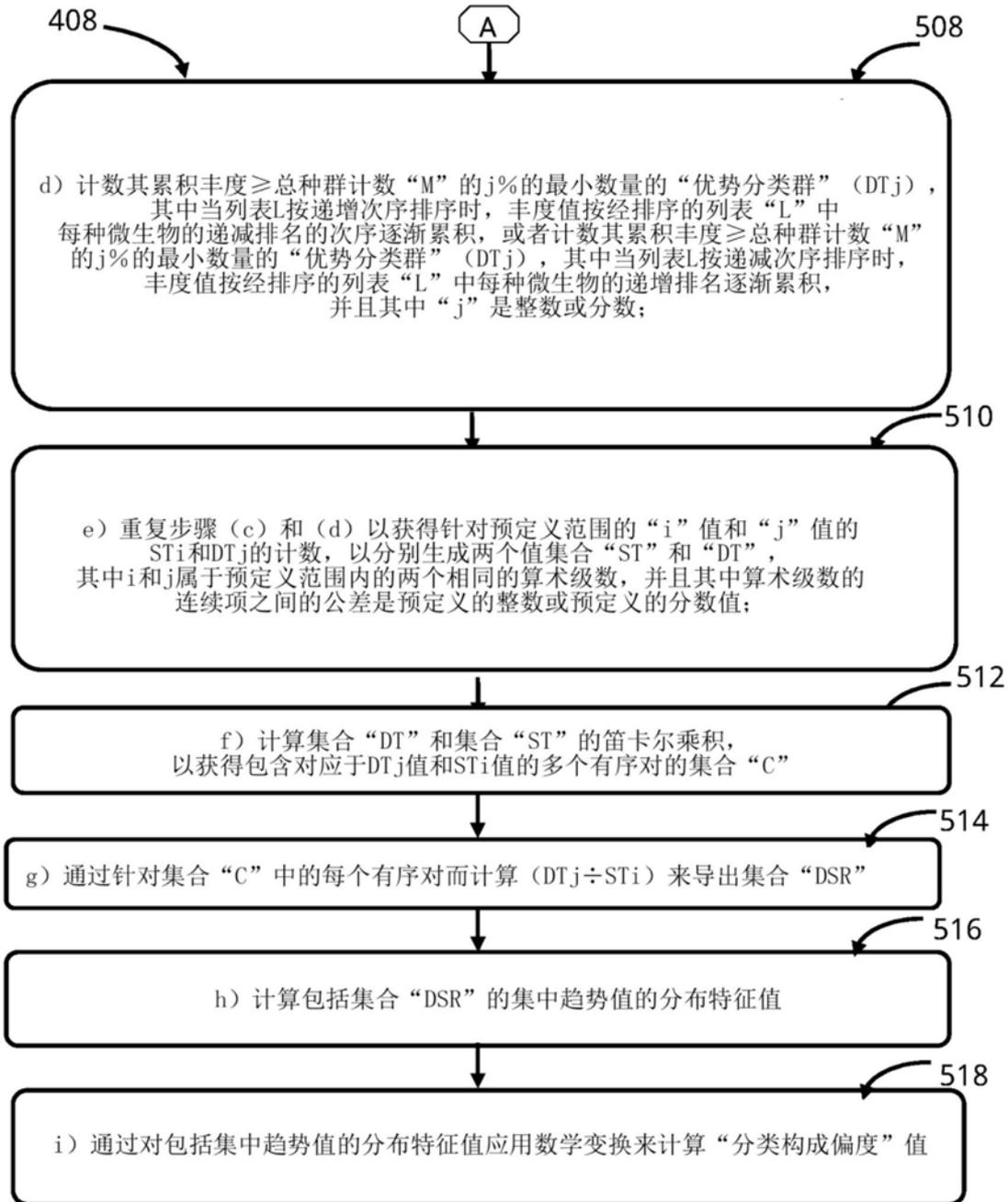


图5-b