



肠道微生物功能宏基因组学与代谢组学关联分析方法研究进展

侯璐文¹, 吴长新^{2,3}, 秦雪梅^{1,3}, 李建国^{2,3*}

¹山西大学中医药现代研究中心, 山西 太原 030006

²山西大学生物医学研究院, 山西 太原 030006

³山西大学化学生物学与分子工程教育部重点实验室, 山西 太原 030006

摘要:近年来基于高通量基因测序的微生物组学研究极大加深了人们对微生物与健康 and 疾病关系的认识。然而基因测序方法不能直接测定微生物的功能活性, 难以鉴定微生物中的关键功能分子, 单独使用无法回答肠道微生物何种成员通过何种方式影响宿主等关键科学问题。单一组学研究弊端尽显, 多组学联用势在必行。肠道微生物代谢组学以微生物群落所有小分子代谢物为研究对象, 可发现肠道微生物随宿主病理生理变化的关键代谢物, 为微生物组-宿主互作机制研究提供线索, 成为微生物组学研究的重要补充。肠道微生物功能基因组学与代谢组学关联分析在宿主生理、疾病病理、药物药理等方面取得众多进展, 展现良好应用前景。然而目前肠道微生物功能基因组学与代谢组学关联分析存在方法滥用、相关性结论与生物学知识相悖等突出问题。为帮助正确应用肠道微生物功能宏基因组学与代谢组学关联分析, 本文综述了各种多组学数据整合分析方法的原理、优缺点与适用范围, 并给出了应用建议。

关键词: 肠道微生物, 功能宏基因组学, 代谢组学, 多组学关联分析

肠道微生物是一个复杂的生物系统, 对人体发挥重要作用, 如大分子消化、维生素合成以及免疫系统驯化等。高通量测序技术加深了对微生物组结构和功能及其与健康疾病关系的认识^[1-3]。宏基因组学研究显示, 肥胖小鼠肠道微生物的代谢通路编码基因与健康对照相比发生显著变化^[4-5], 然而宏转录组学、宏蛋白质组学和代谢组学等下游组学研究显示, 宏基因组学发现的重要编码基因或通路的富集不一定促成相应功能分子水平的

变化。因此, 基于基因测序的宏基因组学方法具有先天局限性, 不能直接测定微生物功能活性, 难以鉴定行使关键功能的微生物分子^[6-7]。

宏基因组与代谢组等多组学联用可在一定程度上克服上述单一组学研究的局限性, 在肠道微生物与健康疾病关系研究等方面取得众多进展, 显示良好应用前景。Pederson 等^[8]应用肠道宏基因组与血清代谢组关联分析 277 例非糖尿病丹麦人样品发现, 随胰岛素敏感性降低, 肠道微生物中

基金项目: 国家自然科学基金(81301441); 国家科技重大专项(2018ZX10305409); 山西省软科学研究计划(2018041031-2)

*通信作者。Tel/Fax: +86-351-7018958; E-mail: lijg@sxu.edu.cn, jglee@126.com

收稿日期: 2018-08-10; 修回日期: 2018-10-25; 网络出版日期: 2018-12-01

支链氨基酸生物合成与内向转运通路编码基因富集；并且血浆代谢组中支链氨基酸水平升高。作者进一步研究发现 *Prevotella corpi* 和 *Bacteroides vulgatus* 对支链氨基酸生物合成潜能与胰岛素抵抗的相关性最大。动物实验证明，前者可增加血液中支链氨基酸水平，加剧葡萄糖不耐受，导致胰岛素抵抗。Liu 等^[9]运用血清代谢组学和宏基因组关联研究，发现产谷氨酸盐共生菌 *Bacteroides thetaiotaomicron* 在中国肥胖患者肠道微生物中丰度降低，并与血清谷氨酸盐水平成反比。动物实验证实，给服此共生菌可升高血清谷氨酸盐水平，减轻高脂饮食导致的肥胖。上述宏基因组学与代谢组学关联分析发现的与健康疾病关系密切的关键菌株，为靶向肠道微生物治疗疾病提供了靶标。

尽管宏基因组学与代谢组学关联分析展现强大生命力，但因忽视方法应用范围等导致的多组学关联分析方法滥用，以及所得相关性结论与生物学知识相悖等问题不容忽视^[10-12]。为帮助正确应用宏基因组学与代谢组学关联分析方法研究肠道微生物，本文综述了各种多组学整合分析方法的原理、优缺点和适用范围，并给出了应用建议。

1 微生物组学研究中的多组学数据整合

多组学数据是指同一样品的多种不同生物学指标，如基因、代谢物或操作分类单元(OTU)。多组学数据整合方法按照是否基于已有知识分为统计学检验方法和知识驱动的分析方法^[13]。统计学检验方法采用单变量或多变量分析阐释不同组学层级所属生物学指标之间的相关性，所得结论不具生物学意义，需与其他技术手段联合使用^[14]；而知识驱动的分析方法将从单一组学层级获

取的生物学指标投射到已有知识库中，以解释各生物学指标之间的相互联系，所得结论可靠性较高，但通常为关联网络，单独使用难以聚焦生物学指标之间的关键相互联系^[15]。

2 统计学检验方法

2.1 单变量相关分析方法

单变量相关分析是最简单的组学数据整合方法，包括皮尔逊相关和斯皮尔曼相关。皮尔逊相关以两变量与各自平均值的离差反映两连续变量的线性相关程度，可定量描述此相关强度与方向，但适用范围较窄，要求两变量相互独立、均为连续变量且各自总体呈正态分布。当样品数小于 30 时所得结论不可靠。斯皮尔曼相关分析反映称名数据或顺序数据中各变量排列顺序的相关程度，不需变量正态分布假设，对样品数量无要求，适用范围较广，但结果精确度低于皮尔逊相关。

单变量相关分析应用于肠道微生物功能宏基因组与代谢组关联分析可确定微生物组组成成分或基因与代谢组单个代谢物之间是否存在显著的线性关系(皮尔逊相关)或单调关系(斯皮尔曼相关)。Theriot 等^[16]运用斯皮尔曼相关分析研究小鼠肠道微生物组 OTU 与代谢组代谢物之间的成对关系。对这些相关性的进一步无监督聚类分析显示这些代谢物-OTU 对倾向于按照不同的环境进行组合，显示肠道微生物组成与代谢应对环境影响的一致性。Mao 等^[17]在羊瘤胃微生物组的多组学研究中，应用单变量相关方法建立微生物属与代谢物之间的皮尔逊相关矩阵，发现随着碳水化合物摄入量的增加，瘤胃微生物群落结构的改变与代谢物分布变化之间存在明显的相关性。

单变量相关性分析相对简单，但假阳性率高，

需要多重校正检验以控制 I 型错误率。同时, 基于单变量相关性分析得到的代谢物与微生物组之间的相关性, 无法阐述其生物学合理性和机制内涵, 与其他知识驱动的关联分析方法联用可在一定程度克服上述局限。

2.2 多变量相关分析方法

由于生物系统的复杂性, 不受其他因素影响的单指标相互作用几乎不存在, 因而需同时考虑多个指标之间的相关性, 多变量相关分析方法应运而生。尽管多变量相关分析方法复杂程度远高于单变量方法, 但可同时考察数据矩阵内和矩阵间的多个变量之间的相互关系, 因而具有一定先进性。由于组学数据的高维特点, 数据降维已成为其主要分析方法。降维技术将大量变量以信息丢失最少的方式转变为少量新变量, 包括主成分分析(PCA)和偏最小二乘分析(PLS)。PCA 是一种经典数据降维方法, 识别数据矩阵中包含最大化方差变量的线性组合。PLS 是一种监督方法, 通过提取两组数据之间的最大协方差实现降维^[18]。双向正交 PLS (O2-PLS)是 PLS 的一种扩展方法, 将每个数据集的变量分为共有、特异和其他三个区块, 分析识别可以预测各组数据共有变异的组分^[19-21]。多变量相关分析方法是上述常用降维技术的扩展, 主要包括典范对应分析、协惯量分析和普鲁克分析等。

典范对应分析 (canonical correspondence analysis, CCA)^[22]是一种基于多元限制性排序的特征提取方法。CCA 将对应分析与多元回归分析相结合, 在对应分析的迭代过程中, 每次计算均进行两组数据的多元线性回归, 基于最大相关性提取对两组数据最佳线性关系贡献值最大的典型变量。CCA 要求数据集内变量线性独立, 并且样本

数不少于变量数。由于两个条件在多组学数据分析中很难同时满足, 人们开发了 CCA 的稀疏变异算法以克服此局限, 包括稀疏 CCA (sparse CCA)^[23]、内核 CCA (kernel CCA)^[24]和正则 CCA (regularized version of CCA, RCCA)^[25]。Kostic 等^[26]使用 Sparse CCA 对易患 1 型糖尿病(T1D)婴儿的肠道微生物组和代谢组数据进行关联分析, 鉴定瘤胃球菌属 (*Ruminococcus* 和韦荣球菌属 (*Veillonella*)为典型变量, 它们与鞘磷脂水平升高和石胆酸水平降低显著相关。

协惯量分析(CIA)^[27]是一种多变量方法, 用于识别包含相同样本或相同时间点的多个数据集中的趋势或共同关系。CIA 的基本原理是基于两组变量的协方差矩阵, 提出两变量的协同结构, 并投影到同一空间, 通过组分之间最大协方差来描述两个数据集之间的共有结构, 以反映数据集的共有趋势或共同关系。Hill 等^[28]使用 CIA 分析早产儿和足月儿的尿液代谢组与肠道微生物宏基因组数据之间的关系, 发现代谢组和宏基因组数据之间存在显著协方差, 可以明显区分早产儿和足月儿。Liu 等^[9]使用 CIA 评估肥胖和健康个体微生物组基因和代谢物之间的协方差, 协方差结果表明这两组数据之间有显著差异, 差异富集的微生物与酪氨酸、苯丙氨酸、谷氨酸和支链氨基酸等代谢物相关。CIA 未对多组学数据关联分析结果进行聚类, 不适用于大规模数据分析; 且通常只能获得少数显著差异的特征值, 难以解释数据整体差异^[29]。

普鲁克分析(Procrustes analysis, PA)是一种形状分布分析方法, 基本思想是通过连续迭代, 发现标准形状(canonical shape), 利用最小二乘法确定样本形状到标准形状的仿射变换形式。PA 利用

PCA 和 CCA 等数据降维结果对多组学数据进行可视化整合^[30], 方法是将两个数据集的主成分叠加在低维空间, 以便快速检查多组学数据集的一致性。McHardy 等运用 PA 比较不同肠段微生物宏基因组与代谢组数据, 发现两组学数据的相关性在盲肠高于乙状结肠^[31]。Quinn 等^[32]应用 PA 分析 16S rRNA 数据和代谢组学数据, 研究发酵食物对人体的作用, 发现了微生物组和代谢组随食物干预的协同变化。PA 本身不足以得出强力结论, 需与 CCA 等其他多变量分析方法结合分析。

上述基于统计检验的多组学数据整合方法, 常得出生物学上不可能出现的结论。例如, 微生物组不同层面的变化通常不会同时发生, 单、多变量分析方法无法对此作出准确评估, 而整合已有知识来解释这种时间尺度上的差异具有可行性。比如 Garali 等^[33]提出广义正则 CCA (regularized generalized CCA, RGCCA)及稀疏广义正则 CCA (sparse regularized generalized CCA, SRGCCA)两个多模块方法, 通过整合已有知识, 并与 PCA、PLS、CCA 等方法联用, 实现对多组数据的探索性分析。Li 等^[34]运用时序代谢组研究大鼠抑郁样行为发生发展过程中代谢组的动态变化, 发现了与疾病进程相关的关键代谢物。

3 基于已有知识的多组学数据整合方法

由于前述基于统计检验的关联方法常得不到不具有生物学意义的结论, 人们开发了基于已有知识的多组学数据整合方法。本方法充分利用对微生物、代谢物和/或基因间相关关系的已有认识, 通过文献挖掘或计算机预测从公共数据库中收集相关信息, 用于多组学数据整合。微生物之间及

其与宿主间关系通过相关网络的形式直观展示。微生物、代谢物、基因作为网络的节点, 它们之间已知或预测的相互关系作为网络的连线。这种网络化表现形式有利于通过直观展示和拓扑分析研究微生物群落中的关键成员或重要的相互作用关系。节点度(node degree)是相关网络的常用度量指标^[35], 它表征一个节点与周围节点之间的连接数(number of connections)。高度连接的节点(又称中心点)对整个网络的影响最大。模块化(modularity)是另一种常用度量指标^[35], 模块内部各节点之间的连接数显著多于节点与模块外节点之间的连接数, 并具有更强的功能相关性。可视化拓扑分析常被用于直观展示不同疾病状态间的关键差异, 并提出潜在生物学机制。例如, Greenblum 等^[36]揭示, 肥胖患者肠道微生物的模块化程度显著低于健康对照, 提示肥胖可能导致肠道微生物群落内部的互相交流减少。

3.1 基于已有知识的微生物组相关网络分析

相关网络(correlation network)是最简单的基于已有知识的多组学数据整合方法, 通过多组学数据直接计算或借助已有知识, 建立所研究生物体间的配对关系。例如, McHardy 等^[31]使用配对斯皮尔曼相关构建了盲肠和乙状结肠中微生物组与代谢组的交互网络, 鉴定了网络中各节点相互关系的显著性和方向性(正/负相关)。与此相比, 基于共有生物学相互关系连接节点构建相关网络的方法更具生物学意义。共有生物学相互关系包括微生物通过互养、共生、竞争等对其他微生物生长的正向/负向调控。Sung 等^[37]通过对肠道微生物可吸收或释放的代谢物和大分子的大量文献调研, 构建了肠道微生物代谢物转运网络, 并基于此信息获得肠道微生物成员间的正负相互关系,

构建了肠道菌群相关关系网络。上述两种方法都成功地鉴定了多组学数据之间的新关联。然而只关注节点间的配对关系, 忽略微生物组内部的复杂相互关系是两种方法的共同缺点^[38]。

3.2 基于已有知识的微生物组代谢网络分析

上述基于统计相关性构建的网络虽然涉及微生物之间的相互作用, 但无法提供这些相互作用更深入的机制信息。基于有机体代谢综合重建的代谢模型(metabolic models), 因其提供微生物组功能和活性的关键机制细节信息, 可作为微生物组功能基因组学与代谢组学关联分析的备选方法。人们在此基础上进一步构建了全基因组代谢模型(genome-scale metabolic models, GEMs), 包含了机体全部代谢反应的完整代谢图谱, 可用于代谢组学和宏基因组学数据的整合, 并使这种整合更具生物学意义。现已构建人、小鼠和人体肠道微生物的全基因组代谢模型^[39-41], 联合应用这些全基因组代谢模型可构建群体代谢网络, 并进一步与宿主组织特异性全基因组代谢模型联用以助力微生物组与宿主间关系研究^[42-43]。

3.2.1 代谢模型拓扑分析: 种子集框架(seed set framework)^[44]是全基因组代谢模型研究的常用拓扑方法。此框架应用图形理论算法鉴定生成代谢网络中所有化合物所需外源化合物的最小子集。

Greenblum 等^[36]使用种子集框架研究肠道微生物群落代谢网络, 发现种子集中的酶在肥胖与 IBD 患者体内均高表达。种子集框架还可用于计算不同微生物共生与竞争得分, 预测微生物间相互作用关系^[45]。竞争得分来源于不同种子集合化合物的重叠程度, 共生得分来源于一种微生物消耗的化合物中, 可由另一种微生物提供的比例。此方法还可进一步扩展用于预测菌种/菌株水平上的微

生物相互作用。在计算机模拟微生物群落代谢网络构建时, 可使用拓扑分析敲除特定基因或反应, 以图形算法鉴定敲除前后多组学数据的变化, 评价目标基因或反应的功能^[46]。这种拓扑分析新方法可识别潜在的生物标记物或关键基因/代谢物, 以作为未来研究的靶点。

3.2.2 基于相对代谢更新率预测的代谢模型构建: 相对代谢更新率预测(predicted relative metabolic turnover, PRMT)利用宏基因组数据获取酶的相对丰度和平均酶功能计数(average enzyme function counts), 预测微生物菌属的代谢活性, 进而通过计算 PRMT 得分构建群落代谢组模型^[47]。PRMT 最早用于研究西英吉利海峡环境微生物功能, 使用 PRMT 准确预测了微生物群落随季节变化的代谢更新, 并鉴定与特定代谢物合成及分解密切相关的微生物^[47]。MIMOSA 对 PRMT 的功能进行了扩展, 整合代谢组学数据研究微生物组成和代谢活性^[48]。使用 PRMT 分析宏基因组数据中的基因丰度信息, 基于微生物在 GEMs 或 KEGG 中已知代谢相关信息, 表征微生物群落成员合成与代谢特定化合物的能力, 生成代谢物的微生物群落代谢潜能(community-wide metabolic potential, CMP)得分。通过比较 CMP 得分差异与仪器检测的代谢水平差异, 发现 CMP 得分足以解释微生物群落代谢潜能的代谢物, 进而锁定产生此代谢物的关键微生物菌属或菌株^[48]。此方法还可整合宿主信息获得更准确 CMP 得分, 从而提高结果准确度。

4 总结和展望

本文结合应用实例对基于统计相关性和基于已有生物学知识网络的微生物功能宏基因组学与代谢组学数据整合方法的研究进展进行了综述。

鉴于多组学数据整合方法各有优点与不足, 为避免方法滥用, 获得更具生物学意义的相关性结果以促进微生物组与健康疾病关系研究, 我们提出功能宏基因组学与代谢组学数据整合分析方法的应用建议。首先运用基于已有生物学知识的多组学数据整合分析方法鉴定研究指标所属具有生物学相关性的相关网络, 再结合基于统计检验的多组学数据整合方法明确研究指标之间相关性的统计学显著性, 发现影响此相关性的关键指标或指标间相互关系, 为深入研究微生物组功能提供线索。不可否认的是, 直接运用基于统计检验的多组学数据整合分析结合实验研究也可获得研究指标间的重要相关性结果, 但易掩盖关键相互作用, 得到以偏概全的结论, 不利于对肠道微生物组的全面认识。

当前对多组学数据整合的理解深度不足和相关性分析方法不足是影响肠道微生物组功能基因组学与代谢组学关联分析的重要障碍, 开发更加可靠的生物信息学工具迫在眉睫。从代谢网络拓扑结构和连接模式分析微生物组数据是多组学整合生物信息学工具开发的优先方向, 其中利用已有生物学知识构建相关网络最为直观可行, 可为多组学数据整合提供丰富内涵。有两类方法在微生物组功能基因组学与代谢组学数据整合方面显示巨大潜能。第一类方法是概率图模型, 如基于已有知识的贝叶斯网络(Bayesian network, BN)模型, 其中 BN 是表征变量之间因果关系概率的图形模型, 用于鉴定对观测数据预测效果最好的概率网络。贝叶斯网络成功用于模拟婴儿肠道中共生细菌的定殖^[49], Zhu 等^[50]整合代谢组学和转录组数据, 建立了酵母细胞调控的因果概率网络模型。第二类方法是基于约束的建模(constraint-based

modeling), 流平衡分析(flux balance analysis, FBA)是基于约束建模的常用方法^[51], 通过整合酶结合动力学、反应流(reaction flux)和化学计量学等大量参数, 模拟生物系统内部代谢流, 可用于模拟微生物生长或预测代谢产物。然而, 基于约束的多物种模型仍不完善, 仅可用于有限物种间的相互作用研究^[52-55]。当前应用基于约束的建模方法研究肠道菌群之间的相互作用, 存在计算量大、已有生物学知识储备不足等突出问题。

本文综述了基于统计检验和基于已有生物学知识的多组学数据整合方法的原理、优缺点和适用范围, 并结合肠道微生物组功能宏基因组学与代谢组学关联分析给出了应用建议, 以帮助合理应用多组学数据整合分析方法, 避免得出不具生物学意义的纯统计相关结论, 以促进微生物组学研究。

参 考 文 献

- [1] Yatsunencko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J, Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JI. Human gut microbiome viewed across age and geography. *Nature*, 2012, 486(7402): 222-227.
- [2] Gill SR, Pop M, DeBoy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI, Relman DA, Fraser-Liggett CM, Nelson KE. Metagenomic analysis of the human distal gut microbiome. *Science*, 2006, 312(5778): 1355-1359.
- [3] The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*, 2012, 486(7402): 207-214.
- [4] Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, 2006, 444(7122): 1027-1131.
- [5] Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R,

- Gordon JI. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Science Translational Medicine*, 2009, 1(6): 6ra14.
- [6] Franzosa EA, Morgan XC, Segata N, Waldron L, Reyes J, Earl AM, Giannoukos G, Boylan MR, Ciulla D, Gevers D, Izard J, Garrett WS, Chan AT, Huttenhower C. Relating the metatranscriptome and metagenome of the human gut. *Proceedings of the National Academy of Sciences of the United States of America*, 2014, 111(22): E2329–E2338.
- [7] Verberkmoes NC, Russell AL, Shah M, Godzik A, Rosenquist M, Halfvarson J, Lefsrud MG, Apajalahti J, Tysk C, Hettich RL, Jansson JK. Shotgun metaproteomics of the human distal gut microbiota. *The ISME Journal*, 2009, 3(2): 179–189.
- [8] Patti GJ, Yanes O, Siuzdak G. Innovation: metabolomics: the apogee of the omics trilogy. *Nature Reviews Molecular Cell Biology*, 2012, 13(4): 263–269.
- [9] Liu RX, Hong J, Xu XQ, Feng Q, Zhang DY, Gu YY, Shi J, Zhao SQ, Liu W, Wang XK, Xia HH, Liu ZP, Cui B, Liang PW, Xi LQ, Jin JB, Ying XY, Wang XL, Zhao XJ, Li WY, Jia HJ, Lan Z, Li FY, Wang R, Sun YK, Yang ML, Shen YX, Jie ZY, Li JH, Chen XM, Zhong HZ, Xie HL, Zhang YF, Gu WQ, Deng XX, Shen BY, Xu X, Yang HM, Xu GW, Bi YF, Lai SH, Wang JQ, Qi L, Madsen L, Wang JQ, Ning G, Kristiansen K, Wang WQ. Gut microbiome and serum metabolome alterations in obesity and after weight-loss intervention. *Nature Medicine*, 2017, 23(7): 859–868.
- [10] Sun YV, Hu YJ. Integrative analysis of multi-omics data for discovery and functional studies of complex human diseases. *Advances in Genetics*, 2016, 93: 147–190.
- [11] Bersanelli M, Mosca E, Remondini D, Giampieri E, Sala C, Castellani G, Milanese L. Methods for the integration of multi-omics data: mathematical aspects. *BMC Bioinformatics*, 2016, 17(Suppl 2): S15.
- [12] Huang SJ, Chaudhary K, Garmire LX. More is better: recent progress in multi-omics data integration methods. *Frontiers in Genetics*, 2017, 8: 84.
- [13] Chong J, Xia JG. Computational approaches for integrative analysis of the metabolome and microbiome. *Metabolites*, 2017, 7(4): 62.
- [14] Cambiaghi A, Ferrario M, Masseroli M. Analysis of metabolomic data: tools, current strategies and future challenges for omics data integration. *Briefings in Bioinformatics*, 2017, 18(3): 498–510.
- [15] Geerts H, Hofmann-Apitius M, Anastasio TJ, Brain Health Modeling Initiative. Knowledge-driven computational modeling in Alzheimer's disease research: Current state and future trends. *Alzheimer's & Dementia*, 2017, 13(11): 1292–1302.
- [16] Theriot CM, Koenigsnecht MJ, Carlson Jr PJ, Hatton GE, Nelson AM, Li B, Huffnagle GB, Li JZ, Young VB. Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to *Clostridium difficile* infection. *Nature Communications*, 2014, 5: 3114.
- [17] Mao SY, Huo WJ, Zhu WY. Microbiome-metabolome analysis reveals unhealthy alterations in the composition and metabolism of ruminal microbiota with increasing dietary grain in a goat model. *Environmental Microbiology*, 2016, 18(2): 525–541.
- [18] Trygg J, Wold S. O2-PLS, a two-block (X-Y) latent variable regression (LVR) method with an integral OSC filter. *Journal of Chemometrics*, 2003, 17(1): 53–64.
- [19] Bylesjö M, Eriksson D, Kusano M, Moritz T, Trygg J. Data integration in plant biology: the O2PLS method for combined modeling of transcript and metabolite data. *The Plant Journal*, 2007, 52(6): 1181–1191.
- [20] El Bouhaddani S, Houwing-Duistermaat J, Salo P, Perola M, Jongbloed G, Uh HW. Evaluation of O2PLS in omics data integration. *BMC Bioinformatics*, 2016, 17 Suppl 2: 11.
- [21] El Aidy S, Derrien M, Merrifield CA, Levenez F, Dore J, Boekschoten MV, Dekker J, Holmes E, Zoetendal EG, van Baarlen P, Claus SP, Kleerebezem M. Gut bacteria-host metabolic interplay during conventionalisation of the mouse germfree colon. *The ISME Journal*, 2013, 7(4): 743–755.
- [22] Hotelling H. Relations between two sets of variates. *Biometrika*, 1936, 28(3/4): 321–377.
- [23] Lin DD, Zhang JG, Li JY, Calhoun VD, Deng HW, Wang YP. Group sparse canonical correlation analysis for genomic data integration. *BMC Bioinformatics*, 2013, 14: 245.
- [24] Yamanishi Y, Vert JP, Kanehisa M. Protein network inference from multiple genomic data: a supervised approach. *Bioinformatics*, 2004, 20 Suppl 1: i363–i370.
- [25] Golugula A, Lee G, Master SR, Feldman MD, Tomaszewski JE, Speicher DW, Madabhushi A. Supervised regularized canonical correlation analysis: integrating histologic and proteomic measurements for predicting biochemical recurrence following prostate surgery. *BMC Bioinformatics*,

- 2011, 12: 483.
- [26] Kostic AD, Gevers D, Siljander H, Vatanen T, Hyötyläinen T, Hämäläinen AM, Peet A, Tillmann V, Pöhö P, Mattila I, Lähdesmäki H, Franzosa EA, Vaarala O, de Goffau M, Harmsen H, Ilonen J, Virtanen SM, Clish CB, Orešič M, Huttenhower C, Knip M, Xavier RJ. The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host & Microbe*, 2015, 17(2): 260–273.
- [27] Dolédec S, Chessel D. Co-inertia analysis: an alternative method for studying species-environment relationships. *Freshwater Biology*, 1994, 31(3): 277–294.
- [28] Hill CJ, Lynch DB, Murphy K, Ulaszewska M, Jeffery IB, O’Shea CA, Watkins C, Dempsey E, Mattivi F, Tuohy K, Ross RP, Ryan CA, O’Toole PW, Stanton C. Evolution of gut microbiota composition from birth to 24 weeks in the INFANTMENT cohort. *Microbiome*, 2017, 5(1): 4.
- [29] Kutalik Z, Beckmann JS, Bergmann S. A modular approach for integrative analysis of large-scale gene-expression and drug-response data. *Nature Biotechnology*, 2008, 26(5): 531–539.
- [30] Gower JC. Generalized procrustes analysis. *Psychometrika*, 1975, 40(1): 33–51.
- [31] McHardy IH, Goudarzi M, Tong M, Ruegger PM, Schwager E, Weger JR, Graeber TG, Sonnenburg JL, Horvath S, Huttenhower C, McGovern DP, Fornace AJ Jr, Borneman J, Braun J. Integrative analysis of the microbiome and metabolome of the human intestinal mucosal surface reveals exquisite inter-relationships. *Microbiome*, 2013, 1: 17.
- [32] Quinn RA, Navas-Molina JA, Hyde ER, Song SJ, Vázquez-Baeza Y, Humphrey G, Gaffney J, Minich JJ, Melnik AV, Herschend J, DeReus J, Durant A, Dutton RJ, Khosroheidari M, Green C, da Silva R, Dorrestein PC, Knight R. From sample to multi-omics conclusions in under 48 hours. *mSystems*, 2016, 1(2): e00038–16.
- [33] Garali I, Adanyeguh IM, Ichou F, Perlberg V, Seyer A, Colsch B, Moszer I, Guillemot V, Durr A, Mochel F, Tenenhaus A. A strategy for multimodal data integration: application to biomarkers identification in spinocerebellar ataxia. *Briefings in Bioinformatics*, 2018, 19(6): 1356–1369.
- [34] Li JG, Hou LW, Wang C, Jia XY, Qin XM, Wu CX. Short term intrarectal administration of sodium propionate induces antidepressant-like effects in rats exposed to chronic unpredictable mild stress. *Frontiers in Psychiatry*, 2018, 9: 454.
- [35] Agler MT, Ruhe J, Kroll S, Morhenn C, Kim ST, Weigel D, Kemen EM. Microbial hub taxa link host and abiotic factors to plant microbiome variation. *PLoS Biology*, 2016, 14(1): e1002352.
- [36] Greenblum S, Turnbaugh PJ, Borenstein E. Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. *Proceedings of the National Academy of Sciences of the United States of America*, 2012, 109(2): 594–599.
- [37] Sung J, Kim S, Cabatbat JJT, Jang S, Jin YS, Jung GY, Chia N, Kim PJ. Global metabolic interaction network of the human gut microbiota for context-specific community-scale analysis. *Nature Communications*, 2017, 8: 15393.
- [38] Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, Huttenhower C. Microbial co-occurrence relationships in the human microbiome. *PLoS Computational Biology*, 2012, 8(7): e1002606.
- [39] Thiele I, Swainston N, Fleming RMT, Hoppe A, Sahoo S, Aurich MK, Haraldsdottir H, Mo ML, Rolfsson O, Stobbe MD, Thorleifsson SG, Agren R, Bölling C, Bordel S, Chavali AK, Dobson P, Dunn WB, Endler L, Hala D, Hucka M, Hull D, Jameson D, Jamshidi N, Jonsson JJ, Juty N, Keating S, Nookaew I, Le Novère N, Malys N, Mazein A, Papin JA, Price ND, Selkov E, Sigurdsson MI, Simeonidis E, Sonnenschein N, Smallbone K, Sorokin A, Van Beek JHGM, Weichart D, Goryanin I, Nielsen J, Westerhoff HV, Kell DB, Mendes P, Palsson BØ. A community-driven global reconstruction of human metabolism. *Nature Biotechnology*, 2013, 31(5): 419–425.
- [40] Sigurdsson MI, Jamshidi N, Steingrimsson E, Thiele I, Palsson BØ. A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1. *BMC Systems Biology*, 2010, 4: 140.
- [41] Magnúsdóttir S, Heinken A, Kutt L, Ravcheev DA, Bauer E, Noronha A, Greenhalgh K, Jäger C, Baginska J, Wilmes P, Fleming RMT, Thiele I. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nature Biotechnology*, 2017, 35(1): 81–89.
- [42] Agren R, Bordel S, Mardinoglu A, Pornputtapong N, Nookaew I, Nielsen J. Reconstruction of genome-scale active

- metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Computational Biology*, 2012, 8(5): e1002518.
- [43] Wang YL, Eddy JA, Price ND. Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Systems Biology*, 2012, 6: 153.
- [44] Borenstein E, Kupiec M, Feldman MW, Ruppin E. Large-scale reconstruction and phylogenetic analysis of metabolic environments. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, 105(38): 14482–14487.
- [45] Steinway SN, Biggs MB, Loughran TP, Papin JA, Albert R. Inference of network dynamics and metabolic interactions in the gut microbiome. *PLoS Computational Biology*, 2015, 11(6): e1004338.
- [46] Zhang C, Hua Q. Applications of genome-scale metabolic models in biotechnology and systems medicine. *Frontiers in Physiology*, 2016, 6: 413.
- [47] Larsen PE, Collart FR, Field D, Meyer F, Keegan KP, Henry CS, McGrath J, Quinn J, Gilbert JA. Predicted relative metabolomic turnover (PRMT): determining metabolic turnover from a coastal marine metagenomic dataset. *Microbial Informatics and Experimentation*, 2011, 1(1): 4.
- [48] Noecker C, Eng A, Srinivasan S, Theriot CM, Young VB, Jansson JK, Fredricks DN, Borenstein E. Metabolic model-based integration of microbiome taxonomic and metabolomic profiles elucidates mechanistic links between ecological and metabolic variation. *mSystems*, 2016, 1(1): e00013–15.
- [49] McGeachie MJ, Sordillo JE, Gibson T, Weinstock GM, Liu YY, Gold DR, Weiss ST, Litonjua A. Longitudinal prediction of the infant gut microbiome with dynamic Bayesian networks. *Scientific Reports*, 2016, 6: 20359.
- [50] Zhu J, Sova P, Xu QW, Dombek KM, Xu EY, Vu H, Tu ZD, Brem RB, Bumgarner RE, Schadt EE. Stitching together multiple data dimensions reveals interacting metabolomic and transcriptomic networks that modulate cell regulation. *PLoS Biology*, 2012, 10(4): e1001301.
- [51] Orth JD, Thiele I, Palsson B. What is flux balance analysis? *Nature Biotechnology*, 2010, 28(3): 245–248.
- [52] El-Semman IE, Karlsson FH, Shoaie S, Nookaew I, Soliman TH, Nielsen J. Genome-scale metabolic reconstructions of *Bifidobacterium adolescentis* L2-32 and *Faecalibacterium Prausnitzii* A2-165 and their interaction. *BMC Systems Biology*, 2014, 8: 41.
- [53] Harcombe WR, Riehl WJ, Dukovski I, Granger BR, Betts A, Lang AH, Bonilla G, Kar A, Leiby N, Mehta P, Marx CJ, Segrè D. Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. *Cell Reports*, 2014, 7(4): 1104–1115.
- [54] Shoaie S, Ghaffari P, Kovatcheva-Datchary P, Mardinoglu A, Sen P, Pujos-Guillot E, de Wouters T, Juste C, Rizkalla S, Chilloux J, Hoyles L, Nicholson JK, MICRO-Obes Consortium, Dore J, Dumas ME, Clement K, Bäckhed F, Nielsen J. Quantifying diet-induced metabolic changes of the human gut microbiome. *Cell Metabolism*, 2015, 22(2): 320–331.
- [55] Shoaie S, Karlsson F, Mardinoglu A, Nookaew I, Bordel S, Nielsen J. Understanding the interactions between bacteria in the human gut through metabolic modeling. *Scientific Reports*, 2013, 3: 2532.

Advances in research on association analysis methods between gut functional metagenomics and metabolomics

Luwen Hou¹, Changxin Wu^{2,3}, Xuemei Qin^{1,3}, Jianguo Li^{2,3*}

¹ Modern Research Center for Traditional Chinese Medicine, Shanxi University, Taiyuan 030006, Shanxi Province, China

² Institutes of Biomedical Sciences, Shanxi University, Taiyuan 030006, Shanxi Province, China

³ Key Laboratory of Chemical Biology and Molecular Engineering, Ministry of Education, Shanxi University, Taiyuan 030006, Shanxi Province, China

Abstract: In recent years, microbiome study based on high-throughput gene sequencing has greatly deepened the understanding of the relationship between gut microbiota and health. However, gene sequencing method cannot directly determine the functional activity of microorganisms. It is difficult to identify the key functional molecules in microorganisms. Application of this single technology cannot answer some key scientific questions which members of gut microbiota and how they affect the host. The disadvantages of singles-omics research are obvious, and it is imperative to perform more multi-omics studies. Metabolomics of the gut microbiota takes all small molecular metabolites of microbial community as the research object. It can discover the key metabolites of intestinal microorganisms with the pathophysiological changes of host. It provides clues for the study of microbial-host interaction mechanism and becomes an important complement to gut functional metagenomics. Integration study of gut functional metagenomics and metabolomics have made a lot of progress in host physiology, disease pathology and other aspects, showing a promising future. However, there are some crucial problems in multi-omics study of gut functional metagenomics and metabolomics, such as misapplication of various methods, resulting in inconsistency between relevant conclusions and biological knowledge. In order to help the proper application of multi-omics study on gut functional metagenomics and metabolomics, this paper reviewed the principles, advantages, disadvantages and application scope of various multi-omics data integration analysis methods, and gave some application suggestions.

Keywords: gut microbiota, functional metagenomics, metabolomics, multi-omics association study

(本文责编: 张晓丽)

Supported by the National Natural Science Foundation of China (81301441), by the National Science and Technology Major Project of China (2018ZX10305409) and by the Shanxi Soft Science Research Program (2018041031-2)

*Corresponding author. Tel/Fax: +86-351-7018958; E-mail: lijg@sxu.edu.cn, jglee@126.com

Received: 10 August 2018; Revised: 25 October 2018; Published online: 1 December 2018



李建国, 山西大学生物医学研究院副教授, 2012年于北京协和医学院获得博士学位。任 *Scientific Reports*、*Brain and Behavior*、*the Chinese Journal of Chromatography* 等期刊审稿编辑。主要从事肠道微生物与疾病关系研究, 在抑郁症等心境障碍疾病的肠道微生物代谢组学、多组学关联研究等方面取得系列成果。先后主持数项国家及省市级科研项目, 以第一作者或通信作者在 *Translational Psychiatry*、*Frontiers in Psychiatry* 等知名学术期刊发表文章 10 余篇, 授权专利 7 项。