

DOI: 10.13376/j.cbbls/2017084

文章编号: 1004-0374(2017)07-0619-05

· 前言 ·



赵立平, 美国新泽西州立罗格斯大学生物化学与微生物学系埃弗里芬腾冠名讲席教授, 上海交通大学生命科学技术学院微生物学特聘教授。1989年南京农业大学博士毕业, 1993—1995康奈尔大学访问学者, 2004—2012担任上海交通大学生命科学技术学院副院长, 2005—2009担任上海系统生物医学研究中心常务副主任, 2009年度上海市优秀学科带头人, 2006—2012担任国际微生物生态学会 (ISME) 常务理事。ISME Journal 资深编辑、Microbiome 编辑、Scientific Reports 编委。2014年被选为美国微生物科学院 fellow。长期从事肠道微生物组与代谢健康研究, 发现首例可以引起肥胖的人体肠道细菌, 发展了以肠道菌群为靶点的肥胖症营养干预方案。在 PNAS、ISME Journal、Nature Communications、Nature Reviews Microbiology 等刊物发表论文 70 余篇。应邀在国际微生物生态大会、国际人类微生物组大会、国际糖尿病技术大会、美国微生物学会年会等一系列国际会议做大会报告、特邀报告。2012年6月, 美国《科学》周刊对他的研究工作做过专题报道。

## 微生物组研究热潮中的冷思考

赵立平\*, 张晨虹, 杨鑫

(上海交通大学生命科学技术学院, 上海 200240)

微生物组 (microbiome) 是指人体内外“定居”的所有微生物的总称, 也可以称为菌群 (microbiota)<sup>[1-3]</sup>。菌群与疾病的关系是近年来生物医学研究领域的大热门课题<sup>[4]</sup>。截至 2017 年 4 月底, 我们以“Gut microbiome”或“Gut microbiota”和“Obesity”或“Diabetes”作为检索关键词, 在 PubMed 数据库检索出 1 949 篇论文, 其中综述性论文有 1 076 篇, 占总篇数的 55.2%; 研究性论文 873 篇, 占 44.8%。在研究性论文中以人群为对象的有 292 篇, 仅占 15%; 其他研究对象的有 581 篇, 占到 30%。这些文献的年度增长曲线 (图 1) 反映出, 绝大多数文献是 2011 年以后发表的。这样的文献检索分析结果告诉我们, 人体微生物组研究在最近 5 年的确大热, 但是, 大部分文献是综述类的, 研究类的论文也是以动物模型研究为主, 真正的人体有关的研究只是一小部分, 而且是以相关性研究为主, 报道了很多病人和健康人之间的菌群的差别, 但是, 涉及到因果关系和分子机制的研究非常之少。这样的文献分析其实反映了目前菌群研究存在过度“虚热”的问题。

之所以说目前的菌群与疾病关系研究以“虚热”

为主, 是因为菌群与疾病关系研究中的最重要的科学问题, 也就是“因果关系”问题没有显著进展。而“因果关系”问题是菌群与疾病关系研究中的“终极科学问题”, 必须优先回答才行<sup>[5]</sup>。

病人的菌群与健康人有差别, 可能有两个原因。(1) 患病以后, 人体生理条件的变化影响了菌群的生长环境, 造成菌群结构的改变, 因此, 这种差别是发病的结果。如果菌群结构在病人和健康人之间的差别仅仅是发病的结果, 假如这种菌群结构特征与发病有着很高的、稳定的相关性, 就可以作为疾病诊断的辅助标准, 但也仅此而已, 在临床上的实际用途很有限。(2) 菌群结构受到各种因素的冲击而发生变化, 出现与健康人体不一样的特征, 这种新的菌群结构引起人体的病理性变化, 最终导致某种疾病的发生; 或者, 人体首先出现某种疾病, 然后菌群结构发生变化, 新的菌群结构开始诱发和加重人体疾病, 使疾病的症状加重或者更加复杂; 这

收稿日期: 2017-05-08

\*通信作者: E-mail: lpzhao@sjtu.edu.cn

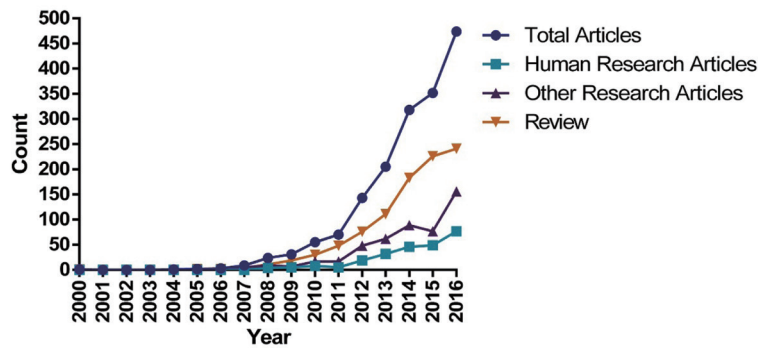


图1 菌群与代谢性疾病关系研究文献分析

样，病人所具有的与健康人不同的菌群结构实际上对疾病的发生、发展具有推动作用，从而成为一个致病因素。也就是说，菌群结构的变化成为人体发病的原因。如果是这样的因果关系，病人所具有的菌群结构特征可以有很多重要的临床应用价值。病人的菌群结构特征可以用于诊断疾病，甚至用于预测疾病的进展。更为重要的是，病人的菌群结构特征可以作为疾病治疗和预防的靶点。把病人特有的菌群结构改变为与健康人类似的结构，疾病就会得到改善，甚至治愈。因此，确认菌群与疾病的关系仅仅是“相关关系”还是“因果关系”，就成为菌群研究领域最为重要的科学问题<sup>[5]</sup>。

证明因果关系的逻辑框架早就已经建立好了，这就是一百多年以前，第一个证明传染病是微生物引起的德国医生科赫建立的“科赫法则”<sup>[6]</sup>。科赫在研究炭疽病发病原因的时候，经受了许多的批评和怀疑而不气馁，不断地建立新的方法，提供新的证据，直到让反对者无话可说。

他首先用显微镜观察罹患炭疽病死亡的小鼠的血液样品，发现里面都有一种杆菌，而健康小鼠的血液里观察不到这种杆菌。因此，他发表论文称，这种杆菌可能是小鼠得炭疽病的病因。批评者则质疑说，患病小鼠血液中的这种杆菌也有可能是得病以后才在血液里长起来的，因此，不是发病的原因，而是发病的结果。科赫做了进一步的实验，他把带有这种杆菌的患病小鼠的血液注射给健康的小鼠，结果健康小鼠都患炭疽病死亡，病死的小鼠血液里也能观察到一模一样的杆菌。这个结果发表以后，批评者认为，血液里除了这种杆菌，还有很多其他成分，有可能是显微镜没有看到的成分导致了疾病，不一定是这种杆菌。为了回答这个质疑，科赫就需要从血液里把这种杆菌分离纯化出来。如果能够在实验室里培养这种细菌，完全摆脱了血液里的其他

成分，再用这种纯培养的细菌去接种健康的小鼠，又能引发病，就能证明这个杆菌是炭疽病的病因。可是，当时的科学家还没有很好地建立分离纯化细菌的方法。科赫只好自己想办法分离这种细菌。一开始，他试着用马铃薯切片做固体培养基，在上面划线分离这种细菌，效果一直不好。有一天，苦思冥想不能解脱的他，在厨房里看到他的太太在用加热的方法化开琼脂做果冻，冷却以后形成了光滑平整的表面，而且加热过程也杀死了培养基里面的混杂的细菌。他高兴地把太太抱起来转了几个圈，然后丢下莫名其妙的太太，一头钻进实验室，去用琼脂做固体培养基平板，再通过划线培养，很快得到了炭疽病小鼠血液里的杆菌的纯培养物。他把这种纯化培养的细菌注射到健康小鼠的血液里，引发了炭疽病，从这种人工诱导发病的小鼠的血液里可以看到同样的杆菌，分离纯化以后，与一开始接种进去的杆菌没有差别。这个结果发表以后，批评者就都销声匿迹了。

科赫不仅第一个发现和证明微生物是引发传染病的原因，建立了“传染病的病原学说”，而且，为寻找和鉴定各种传染病的病因建立了一套实验规则，被后人称为“科赫法则”，成为确认一种微生物是引发某种传染病的病原物的黄金法则。历史证明，遵循这个法则，就能找到某种传染病的病因，从而可以从根本上控制这种传染病。不遵循这个法则，就可能在寻找传染病病因的过程中误入歧途。例如，2003年SARS爆发期间，一开始，有人声称SARS是衣原体引起的，因为在多数SARS患者的样品里能观察到衣原体。但是，这只是相关性的结果，必须把这种衣原体分离纯化出来，再接种动物模型，得到类似的疾病，才能确认衣原体是引起SARS的病因。可惜的是，后面的实验一直做不出来。后来，加拿大科学家在多数SARS患者的样品里发

现了冠状病毒, 分离纯化以后, 接种到猴子体内, 引发了类似人类 SARS 的症状, 从而满足了科赫法则, 确认这种冠状病毒是引起 SARS 的原因<sup>[7]</sup>。如果 SARS 真是衣原体引起的, 抗生素就能有效治疗疾病。如果 SARS 是病毒引起的, 抗生素就没有治疗效果, 控制疾病的流行就需要用与细菌性传染病不同的手段。因此, 病因学研究对控制新发传染病极为重要。

同样地, 要有效控制慢性病, 也必须在病因学研究中取得突破。

在人体微生物组, 特别是肠道菌群与慢性病的关系研究领域, 科赫法则也是研究因果关系必须遵循的逻辑规则。不过, 由于肠道菌群与慢性病的关系不像病原体与传染病那样一般是“一个病菌对应一种传染病”, 而是多种细菌都可能参与了慢性病的发生和发展, 而且, 有的细菌像经典的病原体一样是促进疾病的, 还有的细菌是可以减轻或者预防疾病的。因此, 在现代条件下研究肠道菌群为代表的微生物组与慢性病的关系, 虽然也要遵循科赫法则, 但是, 需要做出适当的修改, 对证据的要求应该更高<sup>[5]</sup>。

我们认为, 遵循科赫法则的逻辑研究肠道菌群与慢性病的因果关系, 需要满足以下三个要求: (1) 通过“全微生物组关联分析”, 把肠道菌群中所有与疾病的发生发展呈正相关或者负相关的细菌种类全部鉴定出来; (2) 把这些与疾病相关的细菌分离培养成纯的菌种或者成员确定的多个菌种的组合物, 然后接种到无菌动物肠道里, 构建疾病动物模型, 在合适的条件下复制疾病的全部或者部分症状; (3) 最后, 深入研究这些细菌影响疾病发生发展的分子机制, 搞清楚这些细菌定植到体内后, 在什么条件下, 产生什么样的生物活性分子, 与哪些类型的动物细胞表面的什么分子受体结合, 启动何种分子信号转导通路, 调控哪些宿主基因的表达, 最终造成疾病症状的加重或者减轻。我们从一位重度肥胖患者肠道里分离到了一株过度生长的阴沟肠杆菌 B29 菌株, 接种到无菌动物肠道里, 可以诱导出肥胖、胰岛素抵抗和脂肪肝等症状, 虽然满足了经典的科赫法则, 但是, 分子机制还在研究之中<sup>[8]</sup>。只有全部满足了这三方面要求的细菌才能被认为是真正参与了某种慢性疾病发生发展的细菌。这样的细菌通过临床验证, 最终有可能成为生物标志物, 用于疾病的预测和诊断, 也可以作为治疗的靶点用于开发新的预防和治疗药物和技术方法。

关联分析显然是锁定“嫌疑犯”的第一步, 这一步有点像“大海捞针”, 如果做得不好, 找到的相关细菌可能是虚假的, 就会令我们在后面的疾病建模和分子验证阶段误入歧途, 不知所终。目前的菌群研究大多数是基于 16S rRNA 基因测序, 分析疾病组与健康组之间的差别。最常见的做法是先把高变区的短标签序列按照 97% 的相似性划分成操作分类单元 (OTUs)。每个 OTU 可以视作一个相当于细菌“种”的分类单元。为了寻找菌群结构特征与疾病的相关性, 往往会计算从“属”到“门”不同分类单元的丰度, 统计分析各个分类单元在疾病组和健康组之间是否有显著差异。遗憾的是, 这样的相关性研究是最有可能让研究者误入歧途的, 根本原因在于细菌功能往往是“菌株专一性的”, 在做全微生物组关联分析的时候, 需要努力做到在“菌株”水平识别和鉴定细菌, 才有可能把真正参与和影响慢性病发生发展的细菌找出来<sup>[9]</sup>。

我们在研究中认识到的所谓“菌株”, 对应的是细菌在自然界的基本生存单位种群 (population)<sup>[10]</sup>。一个种群是从同一个母细胞分裂繁殖而来的基因完全一样的一群细胞。如果我们把这个种群的细胞分离出来, 能够在人工培养基上长期传代培养加以深入研究的话, 我们就得到了一个菌株。如果没有得到一个种群的分离物, 但是, 通过测序检测到了这个种群的细胞的存在情况的话, 也可以视为得到了一个菌株。因此, 菌株是被人类检测和 (或) 分离到的自然界里存在的细菌种群。

为了研究的方便, 我们需要对菌株进行分类。一般是先把类似的菌株划归到一个“种 (species)”里; 然后把类似的种划归成一个“属 (genus)”, 以此类推, 直到把所有的菌株都划归到最高分类单元“门 (phylum)”。同一个种里的菌株遗传上的相似性有多大呢? 按照细菌分类学的研究结果, 目前大家约定, 同一个种里的两个菌株的遗传相似性不能低于 70%, 也就是说, 要有超过 70% 的基因组序列是同源的。换句话说, 同一个种里的不同菌株允许有不超过 30% 的基因序列不一样。我们知道, 人和小鼠的基因组序列差异才只有 10%。这就是说, 同一个种的细菌里面的不同菌株在遗传上的差别可以高达人和小鼠遗传差别的 3 倍。正因为这样, 同一个种里的细菌可能会具有非常不同的功能, 有的菌株会促进某种疾病的发展, 有的菌株可能会减轻疾病, 其余的也许与疾病没有任何关系。因此, 需要在菌株水平识别细菌的有益、有害或者中性的“身份”。

从基因组的角度看,同一个种里的所有菌株所共有的基因,叫做这个种的核心基因组(core genome),某个菌株或者部分菌株所独有的基因全部加起来叫做这个种的泛基因组(pangenome)<sup>[11]</sup>。如果参与某种疾病发生发展的功能是在一种细菌的核心基因组里编码的,这个种里的所有菌株就都有这个功能,这个时候,把同一个种里的所有菌株的丰度都加起来得到的种的丰度会依然与疾病有相关性。但是,如果一种细菌参与某种疾病发生发展的功能是在泛基因组里编码的话,就只有一部分菌株真正具有这个功能。如果不分青红皂白,把所有菌株都加起来得到这个种的丰度,然后去分析这个种与疾病的相关性,就会因为增加了噪声,减低了信号强度,可能得不到真实的相关性的结果,从而令研究者在接下来的深入研究过程中失去目标,或者追踪错误的目标,误入歧途,无功而返。如果说因为同一个种里的菌株与疾病的关系会不一样,因此,在种的水平分析与疾病的相关性有可能误入歧途的话,把同一个属里的种加在一起计算不同的属与疾病的相关性就更有可能出现大的偏差。很显然,随着分类单元的等级升高,同一个分类单元里面的细菌功能的差别就越大,与疾病的关系就越复杂,通过简单地把所有细菌加在一起计算高级分类单元与疾病的相关性,得到结果偏差就会越来越大。美国科学家2006年在*Nature*周刊报道了厚壁菌门与类杆菌门的比值与肥胖的相关性,但是,10年过去了,没有深入的机制研究可以解释为什么这两个门的比值会影响肥胖的发生<sup>[12-13]</sup>。实际上,不同的研究对这两个门的比值与肥胖的关系得出的结论是不一致,甚至是相互矛盾的<sup>[14-16]</sup>。因此,忽略菌株水平的差异,在高等级分类单元研究与疾病的相关性,往往会存在不能追踪功能细菌和深入阐明机制的问题。如果得不到机制研究的确认,菌群结构特征就难以成为稳定可靠的生物标志物,不能用于疾病的诊断和预测,也不能成为预防和治疗的靶点,这样的菌群研究就失去了科学意义和实用价值。可以预计,随着时间的推移,大量的停留在高分类单元的相关性研究,由于不能深入追踪具体的功能细菌,不能阐明分子作用机制,就会逐渐趋于沉寂,最终被遗忘,而目前的菌群研究热潮也就会冷静不少。

如何才能做到在菌株水平分析菌群组成与疾病的相关性呢?目前的技术进步为解决这个问题带来了多种可供选择的方案。其中我们实验室应用比较多的是一种叫做基于Canopy的算法<sup>[17]</sup>。首先,用

基于样品总DNA鸟枪法测序的元基因组技术(也叫宏基因组技术)测定每个样品里面的微生物基因种类和丰度。再通过元基因组测序识别出所有样品里面不冗余的微生物基因,而且计算了每个基因在每个样品里的丰度以后,开始计算基因之间的丰度共变化的关系。如果两个基因的丰度高度相关,说明它们有可能原来是由同一个基因组的DNA分子编码的。通过计算任意两个基因之间的丰度共变化的关系,可以把高度相关的基因归为一个丰度共变化的基因群(co-abundance group, CAG)。每一个这样的基因群可能对应了一个基因组。如果一个基因群带有超过700个基因,而且有典型的细菌基因组的核心基因的话,这些基因群就可以认为是来自于同一个细菌的基因组。将每一个这样的基因群对应的所有高质量的测序片段放在一起进行拼接,可能得到高质量的基因组草图。每个基因组草图对应一个菌株或者非常相似的一组菌株。通过对这样的基因组进行与疾病表型的相关性分析,就可能找到真正参与疾病发生发展的功能菌株。

随着测序成本的大幅度降低,越来越多的研究者开始用元基因组技术分析菌群的功能特征与疾病的相关性。但是,大多数元基因组研究不关心功能细菌的识别,重点分析功能基因、代谢通路在疾病组和健康组之间的差别。这样的分析策略也是有问题的。例如,三甲胺氧化物(TMAO)可能是动脉硬化的高危因素<sup>[18]</sup>。我们发现,很多细菌基因组的丰度与尿液里的TMAO的浓度具有统计学显著的相关性<sup>[19]</sup>。但是,其中只有一部分基因组里真正编码了把胆碱转化成TMAO所需要的功能基因。另一方面,基因组里编码了转化胆碱成为TMAO的功能基因的细菌,它们的丰度与尿液里的TMAO的浓度又不一定相关。这可能意味着,具有某种功能基因的细菌不一定能够表达有关基因,产生有关的代谢物。不带有所需功能基因的细菌又可能“恰好”与有关的代谢产物呈正相关。因此,在寻找疾病相关细菌的过程中,既要关注每个细菌的丰度变化是不是与疾病表型相关,又要注意这些细菌的基因组里是不是编码了参与疾病发生发展所需要的功能基因。只有那些丰度变化与疾病表型有显著的相关性,同时基因组里又编码了参与疾病过程所需要的功能基因的细菌,才值得分离培养出来并建立动物模型和深入研究分子机制。

因此,关联分析、分离建模和分子机制“三位一体”的研究,才有可能令我们对菌群与疾病的关

系认识取得实质性的突破,才有可能在目前国际上菌群研究大热的局面下,保持冷静的头脑,选好自己的突破口,扎扎实实地,一步一个脚印地逐步深入,直到把参与和推动疾病发生发展的主要细菌都“验明正身”,从而为慢性病的防治带来新的诊断和预测的标志物以及预防和治疗的新靶点。唯有此,菌群研究才能真正引发生物医学领域的新的革命性变化,造福人类健康。

#### [参 考 文 献]

- [1] Lenerberg J. Infectious history. *Science*, 2000, 288: 287-93
- [2] Zhao L, Shen J. Whole-body systems approaches for gut microbiota-targeted, preventive healthcare. *J Biotechnol*, 2010, 149: 183-90
- [3] Yang X, Xie L, Li Y, et al. More than 9,000,000 unique genes in human gut bacterial community: estimating gene numbers inside a human body. *PLoS One*, 2009, 4: e6074
- [4] Nicholson JK, Holmes E, Wilson ID. Gut microorganisms, mammalian metabolism and personalized health care. *Nat Rev Microbiol*, 2005, 3: 431-8
- [5] Zhao L. The gut microbiota and obesity: from correlation to causality. *Nat Rev Microbiol*, 2013, 11: 639-47
- [6] Evans AS. Causation and disease: the Henle-Koch postulates revisited. *Yale J Biol Med*, 1976, 49: 175-95
- [7] Fouchier RA, Kuiken T, Schutten M, et al. Aetiology: Koch's postulates fulfilled for SARS virus. *Nature*, 2003, 423: 240
- [8] Fei N, Zhao L. An opportunistic pathogen isolated from the gut of an obese human causes obesity in germfree mice. *ISME J*, 2013, 7: 880-4
- [9] Zhang C, Zhao L. Strain-level dissection of the contribution of the gut microbiome to human metabolic disease. *Genome Med*, 2016, 8: 41
- [10] Madigan MT, Martinko JM, Stahl D, et al. *Brock biology of microorganisms* [M] 13th ed. New York: Pearson Education Limited, 2010
- [11] Tetz VV. The pangenome concept: a unifying view of genetic information. *Med Sci Monit*, 2005, 11: HY24-9
- [12] Ley RE, Turnbaugh PJ, Klein S, et al. Microbial ecology: human gut microbes associated with obesity. *Nature*, 2006, 444: 1022-3
- [13] Turnbaugh PJ, Ley RE, Mahowald MA, et al. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, 2006, 444: 1027-31
- [14] Ley RE, Backhed F, Turnbaugh P, et al. Obesity alters gut microbial ecology. *Proc Natl Acad Sci USA*, 2005, 102: 11070-5
- [15] Kalliomaki M, Collado MC, Salminen S, et al. Early differences in fecal microbiota composition in children may predict overweight. *Am J Clin Nutr*, 2008, 87: 534-8
- [16] Hanage WP. Microbiology: microbiome science needs a healthy dose of scepticism. *Nature*, 2014, 512: 247-8
- [17] Nielsen HB, Almeida M, Juncker AS, et al. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat Biotechnol*, 2014, 32: 822-8
- [18] Wang Z, Klipfell E, Bennett BJ, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature*, 2011, 472: 57-63
- [19] Zhang C, Yin A, Li H, et al. Dietary modulation of gut microbiota contributes to alleviation of both genetic and simple obesity in children. *EBioMedicine*, 2015, 2: 968-84