



ELSEVIER

Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng



Research
Microecology—Perspective

微生物组分析技术的发展趋势：从单细胞功能成像到菌群大数据

徐健^{a,e,*}, 马波^{a,e}, 苏晓泉^{a,e}, 黄适^{a,e}, 徐欣^d, 周学东^d, 黄巍^c, Rob Knight^{b,*}

^a Single-Cell Center, CAS Key Laboratory of Biofuels, Shandong Key Laboratory of Energy Genetics, Qingdao Institute of BioEnergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, Shandong 266101, China

^b Center for Microbiome Innovation, Department of Pediatrics, Department of Computer Science and Engineering, University of California, San Diego, CA 92093, USA

^c Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, UK

^d State Key Laboratory of Oral Diseases, West China Hospital of Stomatology, Sichuan University, Chengdu 610041, China

^e University of Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Article history:

Received 14 January 2016

Revised 25 January 2017

Accepted 17 February 2017

Available online 21 February 2017

关键词

微生物组

方法学创新

单细胞分析

大数据

中国微生物组计划

摘要

方法学创新一直是微生物组学研究的核心驱动力。我们认为在未来五到十年，微生物组的方法学体系在研究理念与技术平台方面将发生三大变革：①从监测菌群“结构”变化向监测菌群“功能/状态”变化的变革；②从细胞“群体”分析精度向细胞“个体”分析精度的转变；③从“数据分析”向“数据科学”的跨越。在这里我们针对实现上述三大方法学变革需要克服的关键科学或技术挑战，重点介绍了部分中国微生物组分析方法学研究团队及其国际合作伙伴的最新工作进展。我们相信中国微生物组计划应把握住当前这一重要机遇，通过在微生物组分析方法学前沿开展富有雄心、远见、创意与竞争力的交叉合作研究，为国际微生物组计划贡献一系列“中国制造”的新方法、新工具和新仪器。

© 2017 THE AUTHORS. Published by Elsevier LTD on behalf of the Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. 背景介绍

新技术与新方法一直是微生物组科学(microbiome science)发展的核心驱动力。从16S DNA作为进化标尺基因的发现，到元基因组技术的问世，再到基于单细胞基因组揭示微生物界的“暗物质”，每一个关键技术的出现都对各种生态环境及应用领域的微生物组研究和相关产业带来了广泛与深远的影响。因此，微生物组分析方法学一直是国内外微生物组计划的关键组分、支持重点和先行者[1–3]。但是，总体而言，与纯培养微生物即细胞种群(population)层面相比，微生物群落(consortium)层面的分

析方法和操作工具，如功能成像、组学分析、细胞培养，都仍然较为单一、也非常局限，既无法满足微生物组功能实时监测的需求，也难以支撑适用于自然界应用之功能菌群的理性设计与组装。同时，跨学科与跨应用领域的微生物组数据整合分析依然面临着巨大挑战，体现在微生物组数据科学(microbiome data science)虽呼之欲出，却还缺乏核心的支撑工具和平台。因此，微生物组分析方法学领域既承受着巨大的挑战和压力，也面临着前所未有的发展机遇。

我们认为，当前微生物组分析方法学体系正处于变革的前夜，体现在三大研究理念与技术平台的跨越已经迫在眉睫(图1)：①从监测菌群“结构”变化向直接监

* Corresponding author.

E-mail address: xujian@qibebt.ac.cn; robknight@ucsd.edu

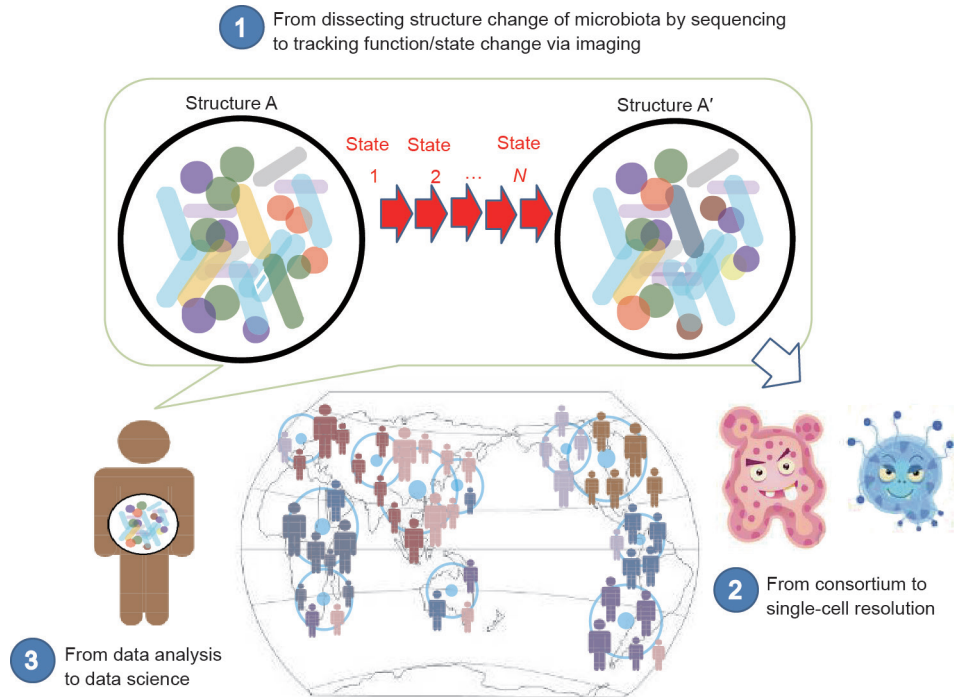


图1. 当前微生物组分析方法学的发展趋势体现于研究理念与技术平台的三大变革。

测菌群“功能/状态”变化的转变：时间分辨率的突破，即在菌群结构发生变化之前就检测到菌群功能的动态变化，是研究菌群进化原理与微进化机制的关键；②从细胞“群体”精度向细胞“个体”精度的转变：空间分辨率的突破，亦即从在群落或纯培养群体精度的功能分析推进到单个细胞精度的功能分析，对于理解菌群的功能异质性（如耐药性的产生机制等）至关重要；③从“数据分析”向“数据科学”的跨越：这一数据分析理念的转变，亦即从单一项目内数据比较所得的局部认知推进到与人类也已积累的所有菌群比较所得的全局认知，是实时理解每一批新数据之贡献内涵与程度的前提，也是“鸟瞰式”地理解地球上各种生态系统中微生物组之共性和特性的基础。

2. 从菌群“结构”监测向菌群“功能/状态”及其内部代谢互作之实时监测的转变：非侵入性快速微生物组功能成像技术的出现

对于许多慢性疾病来说，病程的发生与发展往往伴随着菌群结构的变化，然而群落结构每一变化均是由一系列菌群状态/功能的变化所导致的。因此，对菌群状态变化的检测是深入研究菌群运作机制和微进化的前提。但是，元基因组手段通常只能检测到菌群的物种结构(通过16S/18S基因扩增子测序)或功能基因结构(通

过鸟枪法测序)，因此无法直接检测菌群之状态或功能。而元转录组、元蛋白组、元代谢组等手段因为对细胞的破坏性、操作的繁琐性以及费用高昂等问题，也通常难以实现菌群功能/状态的实时检测。此外，菌群中细胞种类繁多、功能复杂、功能标识物也通常未知，因此依赖于外加标记的细胞功能成像技术体系对于微生物组通常也无能为力。例如，荧光显微镜虽被广泛应用于细胞生物学研究中，但在菌群功能分析中却受到诸多局限。这主要是因为自然界中大多数微生物既难以通过共性的策略对细胞进行标记，也没有特异性的功能探针。因此，迫切需要发展非标记式、快速、适用于菌群中所有细胞类型的功能成像策略与方法。

与此同时，菌群内不同种类的微生物之间存在着复杂、精妙的相互作用与影响，这一跨物种的细胞间代谢互作网络是群落功能和进化的基础，但是，当前微生物组功能分析的主要工具，如元基因组、元转录组、元代谢组、稳定性同位素核酸探针(DNA/RNA-SIP)技术以及高通量培养技术(或称培养组)等，通常难以揭示原位状态下群落中各个成员在代谢功能上的互作关系。

针对上述瓶颈问题，最近我们提出了一种非标记、单细胞水平、基于功能成像的表型组学：拉曼组(ramanome)与元拉曼组(meta-ramanome)[4]。它们在概念上类似于给细胞的群体(ramanome)或群落(meta-ramanome)拍集体照，是特定条件和时间点下这些细胞之单细胞拉曼光谱的集合

[4]。每个单细胞拉曼光谱由分别对应于一类化学键的超过1500个拉曼谱峰组成，反映的是特定细胞内化学物质的成分及含量的多维信息。因此单细胞拉曼光谱表征的是该细胞在该条件和该时间点下所有代谢物的组成及相对含量。因为代谢物组成能够对细胞生理状态、细胞所处的环境变化，以及遗传背景的改变做出快速响应，所以每个单细胞拉曼光谱可以被视为一张超过1500个“像素”的高分辨率“个人照片”，用于表征状态或表型。同时，其测量无需破坏细胞、无需标记细胞、无需预先知道目标表型的生物标识物，而且通常仅需毫秒乃至数秒钟。所以对于一个微生物群体或群落，一个拉曼组相当于单细胞精度的、可快速、低成本测定与监控的“代谢物组”或“元代谢物组”，其变化可直接且“全景式”地反映和表征细胞群体与群落的“状态”与“功能”。例如，我们近期工作表明，拉曼组可以区分微生物类别[5]、测量细胞的代谢活性以及细胞分解特定底物的活性[6]、监测合成代谢产物如甘油三酯[7]及淀粉的含量[8]、区分细胞环境应激机制[4]等。我们最近发明的反向标记拉曼成像技术，证明元拉曼组可在单细胞水平上解析与示踪微生物群落中细胞之间的代谢互动与共生关系[9]。尤其重要的是，由于拉曼光谱中的每个拉曼谱峰或若干拉曼谱峰的集合均潜在代表一种表型，因此，正如同一个1500像素的肖像照能够同时展示一个人脸的诸多特征一样，拉曼组与元拉曼组能够在单个细胞精度“全景式”地表征细胞群体或群落的“状态”与“功能”。这些特色充分体现了拉曼组与元拉曼组手段与元基因组、元代谢组等现有组学技术的互补性，使其有望成为一种通用手段和崭新的表型组学数据类型，来定义、表征乃至监测微生物组的功能及其异质性。

3. 从细胞“群体”精度向细胞“个体”精度分析的转变：在最深层次分析微生物组

依仗针对菌群中所有细胞之集合的总DNA测序，元基因组手段能较为全面地揭示自然状况下微生物群落的物种结构和潜在功能[10,11]。元基因组数据本身通常具有体量大、遗传异质性强、物种丰度极其不均匀等特点，因此如何针对特定生境来优化样品预处理方法、如何利用PacBio和Nanopore等新一代测序技术、如何改进测序和拼装的策略、如何进行深度数据挖掘等均是目前业界的热点。为了解决以上问题，中国一些计算工具开发团队在元基因组分析工具开发方面取得了令人瞩目的进展。例如，发展了基于超短序列的物种谱分类及注释方

法[12]，建立了16S rRNA-侧翼序列环化测序技术及计算方法(RiboFR-Seq)[13]、基于机器学习和路径拓扑方法的复杂元基因组基因重建算法(inGAP-CDG)[14]、以及基于降低物种复杂度的实验和计算策略的metaSort方法[15]。同时，开发了包括质控(QC-Chain)[16]、海量样本比较(Meta-Storms算法，实现了微生物群落样本基于进化二叉树的量化比较)[17,18]、大规模微生物组的系统性分析(MDV数据计算模型)[19]与可视化(MetaSee)[20]等在内的一系列分析算法和软件包。

“个体”精度的分析，即单细胞功能分选、测序和培养技术，能从根本上突破上述元基因组手段通常难以分辨与验证群落中细胞个体功能这一核心局限性。目前，国际上较为通用的细胞分选手段是基于荧光信号的荧光流式细胞分选(fluorescence-activated cell sorting, FACS)[21]。然而问题在于，由于需要对目标蛋白、代谢产物或核酸进行基于荧光探针的细胞标记，FACS严重依赖于对目标功能之生物标识物的预先认知，而对于绝大部分微生物组，生物标识物经常未知，也没有普适性的细胞标记方法，因此FACS通常难以用于研究复杂菌群的功能。针对这些技术屏障，一系列基于拉曼光谱的单细胞分选技术和核心器件，包括拉曼激活弹射分选(RACE)[5]、拉曼激活微流分选(RAMS)[22]等已成功研制；并在此基础上，我们研制了首台“单细胞拉曼分选仪”(Raman-activated cell sorter, RACS)实验室原型机，能够根据单细胞拉曼光谱对微生物细胞进行功能分选[5,22]。如前所述，由单细胞拉曼光谱组成的拉曼组和元拉曼组手段不需要生物标识物，无需标记细胞，适用于所有的细胞，而且潜在能够表征的细胞功能几近无限[23]，因此RACS为从菌群中非标记式地分选出各种特定功能的单个细胞或细胞亚群提供了崭新的共性技术平台。

与上述基于FACS或RACS的细胞功能分选直接耦合的单细胞测序策略能够从细胞“个体”角度诠释微生物组功能机制，从而克服元基因组手段的核心局限性。单细胞基因组扩增方法，如“多重置换扩增”(MDA)等，在多种复杂环境中识别出全新的细菌和古菌门类[24]，从深海热泉中发现了原油降解途径[25]，在菌株(strain)水平上鉴别不同病菌[26]，还揭示了海洋浮游微藻和共生病毒之间[27]、人体共生菌与寄生病毒之间[28]、固氮蓝细菌和光合微藻之间的互动与共进化机制[29]。最近发展的MALBAC(multiple annealing and looping based amplification cycles)方法对于哺乳动物细胞的扩增偏好性低于MDA方法[30,31]，但在微生物细胞中并无明显

改善,且来自环境DNA污染的序列比例较高[32]。因此,与上游功能分选相结合,而且低偏好性、高覆盖度和高通量的微生物单细胞基因组扩增与测序流程目前尚存诸多问题。针对这些难点,基于集成化动态微液滴阵列的高通量微生物单细胞功能基因组分析体系将具相当前景。该体系集成基于自供能方式的微液滴产生、融合、内含物回收等模块,能够实现单细胞分离、细胞裂解、核酸扩增试剂的添加、在线预扩增以及扩增产物回收等多种功能[33],因此在不需昂贵和笨重仪器或专用设施的前提下能降低受到环境中细胞与DNA污染的概率,从而支撑在研究条件受限的极端环境下便携式、原位和实时的单细胞分析。同时,不依赖于微流控芯片的低成本动态界面打印(cross-interface emulsification, XiE)微滴制备技术,可将单细胞基因组扩增反应控制在纳升级的微流控液滴反应系统,并与上游荧光流式细胞分选相耦合[34]。这些新技术与新方法将有助于移动式乃至手持式器件与仪器的研制,服务于各种应用场景下菌群中特定细胞或细胞亚群16S/18S rRNA基因、功能基因或基因组的扩增与测序。

微生物培养,无论是在细胞分选前还是细胞分选后,既是微生物组功能检测与验证的手段之一,也是纯化与制备其功能组分的重要步骤。最近的研究表明,通过培养条件的多重优化或者原位培养,诸多原本认为难以培养的土壤或肠道微生物能实现实验室培养[35,36]。但是目前针对菌群中所有组分的培养,亦即培养组技术,大多依赖于基于传统平板涂布的群体培养方法,其问题在于优势菌和非优势菌的竞争可能导致物种的多样性丧失,因此经常难以获得生长缓慢或数量稀少的物种。针对这些挑战,一些解决办法如单细胞水平的并行式微液滴培养已提出[37,38]。例如,在一种简易的微流控液滴“划板”培养方法(microfluidic streak plate, MSP)中,细胞被包裹在液滴或凝胶微球中进行独立并行培养,在减少菌群内部营养竞争时允许物种间代谢产物的交换,从而比传统的平板涂布培养方法获得更高的多样性[39]。微液滴技术还可以服务于快速培养。首先,基于液滴单细胞包裹培养与成像的数字化涂布计数(digital spreading plate counting, dSPC)技术,可将培养检测周期大幅缩短(从数天缩到数小时)。其次,基于纳米微球团聚的单细胞直接成像技术可无需培养而直接识别与计数单细胞包裹液滴,从而在数分钟完成微生物绝对计数。我们实验室开发的上述两种快速培养方法都无需

复杂的检测仪器,基于手机无线显微成像技术便能满足应用需求,非常适合便携式、实时的微生物组现场检测。

4. 从微生物组数据分析向微生物组数据科学的跨越:温故而知新

大数据技术是当前微生物组科学与应用的关键瓶颈之一。微生物组大数据中蕴含着应对当前精准医学、环境修复、清洁能源等方面重大挑战的诸多机会。比如,基于数据科学,能够解读群落“相似性”的深刻内涵,从而理解地球上所有微生物组的全局特征;同时,运用机器学习,能够揭示微生物组结构相似性与功能相似性的关联,从而预测菌群的功能与进化方向;此外,通过人工智能,能够基于微生物组建立人类疾病与生态灾害的新一代诊断与预警模型。然而,元基因组极端高通量、多源性、多类型、多尺度、多形式存在性等特点,使得大数据整合、搜索与匹配等全局分析面临着巨大的技术屏障,阻碍着上述重大科学问题的剖析。具体来说主要是以下瓶颈。首先,元基因组数据整合与索引技术匮乏。MG-RAST[40]、CAMERA[41]等是国际上主要的元基因组数据库之一,但其中的数据来自于迥异的课题设计和采样环境,而且由于测序样品制备以及测序方法的不同,对这些数据进行整合分析以及全局比对一直异常困难。其次,元基因组高速比对与海量数据搜索技术缺失。目前已发表的Phyloshop[42]、MEGAN[43]等菌群结构分析算法和MOTHUR[44]、UniFrac[45,46]、QIIME[47]等菌群结构比对算法,主要基于“单个研究项目内元基因组数据的分析”这一设计理念,这从根本上导致现有方法难以支撑在整个人类已知元基因组数据空间的比较和搜索。再次,爆发式增长的元基因组数据来源与体量对大数据系统的设计和可持续性都提出了空前的成本与计算能力的需求。与此同时,多组学的整合挖掘,即通过元代谢组、元转录组、单细胞基因组,以及新的表型组学数据形式(如上文提到的拉曼组、元拉曼组等)的整合与多维度分析,也是目前微生物组大数据研究中涉及尚浅但亟待解决的问题。

针对这些瓶颈,我们设计了“微生物组大数据搜索引擎”(microbiome search engine, MSE)[†],服务于以整个微生物组为分析单元的智能搜索和大数据挖掘。基于数据库索引[17]、16S扩增文库比对算法开发[48-50]、相

[†] <http://mse.single-cell.cn>

似度的统计学意义评价方法[17]、基于GPU的搜索加速软件[18]等系列算法, MSE实现了以元基因组为单位的“BLAST式”搜索, 能够根据人类已经积累的海量微生物组数据, 快速辨别并返回与目标菌群“结构”或“功能”最为相似的参照菌群样本, 从而对目标菌群进行注释。同时, MSE还基于机器学习, 自动为元基因组数据建立微生物组诊断指数模型, 以判别目标样本针对特定疾病或生态灾害的状态或风险, 如“牙龈健康指数”[51]、“儿童龋病预警指数”[52]、“痛风预警指数”[53]等。MSE中的微生物组参照数据库包含了大范围内收集的、由同一生物信息分析流程产生的高质量、注释完整的数据, 从而实现数据获取、质量控制、数据比对及可视化等环节的差异最小化, 保证了跨项目、跨应用领域元基因组数据比较的科学性。因此, MSE是支撑目标菌群的深入解读和挖掘, 同时对整个已知微生物组数据空间进行全局性思考的“数据科学”支撑平台, 将成为服务于探索微生物组新大陆(包括真菌、病毒等在内[54])的大数据挖掘工具。

综上所述, 就像近十几年来元基因组等手段给微生物组科学带来的变革一样, 上述三大微生物组分析理念和方法学的转变与革新, 将在此后五到十年贯穿并改观微生物组分析方法学领域, 同时催生一系列崭新、共性、系统的实验与计算工具。正如孔子两千多年前在《论语》中所说:“工欲善其事, 必先利其器”。如果没有在方法、技术、软件、仪器方面的创新, 我们就无法借助自然界这些“微小的巨人”之力量来迎接这一代人在健康、能源、环境、资源、农业, 以及深空、深海、深地探索中所面临的挑战。同时, 通过资助形式与管理机制的探索, 促进工具开发者与工具使用者之间的紧密合作, 也是中国微生物组计划应当考虑的。来自中国的微生物组分析方法学研究人员应当把握住当前机遇, 基于国内外紧密合作, 在中国微生物组计划中策划富有雄心、远见、创意与竞争力的方法学开拓, 从而为国际微生物组计划和广阔的微生物组应用领域贡献一系列“中国制造”的新方法、新工具和新仪器。

致谢

感谢国家自然科学基金委员会(31425002、91231205、81430011、61303161、31470220和31327001)、中国科学

院前沿科学重点项目、中国科学院土壤微生物先导专项和中国科学院生物高通量检测分析技术服务网络(STS)的支持。

Compliance with ethics guidelines

Jian Xu, Bo Ma, Xiaoquan Su, Shi Huang, Xin Xu, Xuedong Zhou, Wei Huang, and Rob Knight declare that they have no conflict of interest or financial conflicts to disclose.

References

- [1] Dubilier N, McFall-Ngai M, Zhao L. Microbiology: create a global microbiome effort. *Nature* 2015;526(7575):631-4.
- [2] Bouchie A. White House unveils National Microbiome Initiative. *Nat Biotechnol* 2016;34(6):580.
- [3] Blaser MJ, Cardon ZG, Cho MK, Dangl JL, Donohue TJ, Green JL, et al. Toward a predictive understanding of earth's microbiomes to address 21st century challenges. *MBio* 2016;7(3):e00714-16.
- [4] Teng L, Wang X, Wang X, Gou H, Ren L, Wang T, et al. Label-free, rapid and quantitative phenotyping of stress response in *E. coli* via Ramanome. *Sci Rep* 2016;6:34359.
- [5] Wang Y, Ji Y, Wharfe ES, Meadows RS, March P, Goodacre R, et al. Raman-activated cell ejection for isolation of single cells. *Anal Chem* 2013;85(22):10697-701.
- [6] Berry D, Mader E, Lee TK, Woebken D, Wang Y, Zhu D, et al. Tracking heavy water (D₂O) incorporation for identifying and sorting active microbial cells. *Proc Natl Acad Sci USA* 2015;112(2):E194-203.
- [7] Wang T, Ji Y, Wang Y, Jia J, Li J, Huang S, et al. Quantitative dynamics of triacylglycerol accumulation in microalgae populations at single-cell resolution revealed by Raman microspectroscopy. *Biotechnol Biofuels* 2014;7(1):58.
- [8] Ji Y, He Y, Cui Y, Wang T, Wang Y, Li Y, et al. Raman spectroscopy provides a rapid, non-invasive method for quantitation of starch in live, unicellular microalgae. *Biotechnol J* 2014;9(12):1512-8.
- [9] Wang Y, Song Y, Tao Y, Muhamadali H, Goodacre R, Zhou NY, et al. Reverse and multiple stable isotope probing to study bacterial metabolism and interactions at the single cell level. *Anal Chem* 2016;88(19):9443-50.
- [10] Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, et al. Enterotypes of the human gut microbiome. *Nature* 2011;473(7346):174-80.
- [11] Nayfach S, Pollard KS. Toward accurate and quantitative comparative metagenomics. *Cell* 2016;166(5):1103-16.
- [12] Liu J, Wang H, Yang H, Zhang Y, Wang J, Zhao F, et al. Composition-based classification of short metagenomic sequences elucidates the landscapes of taxonomic and functional enrichment of microorganisms. *Nucleic Acids Res* 2013;41(1):e3.
- [13] Zhang Y, Ji P, Wang J, Zhao F. RiboFR-Seq: a novel approach to linking 16S rRNA amplicon profiles to metagenomes. *Nucleic Acids Res* 2016;44(10):e99.
- [14] Peng G, Ji P, Zhao F. A novel codon-based de Bruijn graph algorithm for gene construction from unassembled transcriptomes. *Genome Biol* 2016;17(1):232.
- [15] Ji P, Zhang Y, Wang J, Zhao F. MetaSort untangles metagenome assembly by reducing microbial community complexity. *Nat Commun* 2017;8:14306.
- [16] Zhou Q, Su X, Wang A, Xu J, Ning K. QC-Chain: fast and holistic quality control method for next-generation sequencing data. *PLoS One* 2013;8(4):e60234.
- [17] Su X, Xu J, Ning K. Meta-Storms: efficient search for similar microbial communities based on a novel indexing scheme and similarity score for metagenomic data. *Bioinformatics* 2012;28(19):2493-501.
- [18] Su X, Wang X, Jing G, Ning K. GPU-Meta-Storms: computing the structure similarities among massive amount of microbial community samples using GPU. *Bioinformatics* 2014;30(7):1031-3.
- [19] Su X, Hu J, Huang S, Ning K. Rapid comparison and correlation analysis among massive number of microbial community samples based on MDV data model. *Sci Rep* 2014;4:6393.
- [20] Song B, Su X, Xu J, Ning K. MetaSee: an interactive and extendable visualization toolbox for metagenomic sample analysis and comparison. *PLoS One* 2012;7(11):e48998.
- [21] Rinke C, Lee J, Nath N, Goudeau D, Thompson B, Poulton N, et al. Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nat Protoc* 2014;9(5):1038-48.
- [22] Zhang P, Ren L, Zhang X, Shan Y, Wang Y, Ji Y, et al. Raman-activated cell

- sorting based on dielectrophoretic single-cell trap and release. *Anal Chem* 2015;87(4):2282–9.
- [23] Zhang Q, Zhang P, Gou H, Mou C, Huang WE, Yang M, et al. Towards high-throughput microfluidic Raman-activated cell sorting. *Analyst* 2015;140(18):6163–74. Erratum in: *Analyst* 2015;140(19):6758.
- [24] Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF, et al. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 2013;499(7459):431–7.
- [25] Mason OU, Hazen TC, Borglin S, Chain PSG, Dubinsky EA, Fortney JL, et al. Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. *ISME J* 2012;6(9):1715–27.
- [26] Lasken RS, McLean JS. Recent advances in genomic DNA sequencing of microbial species from single cells. *Nat Rev Genet* 2014;15(9):577–84.
- [27] Roux S, Hallam SJ, Woyke T, Sullivan MB. Viral dark matter and virus–host interactions resolved from publicly available microbial genomes. *eLife* 2015;4:e08490.
- [28] Tang HW, Yang XB, Kirkham J, Smith DA. Probing intrinsic and extrinsic components in single osteosarcoma cells by near-infrared surface-enhanced Raman scattering. *Anal Chem* 2007;79(10):3646–53.
- [29] Li M, Canniffe DP, Jackson PJ, Davison PA, FitzGerald S, Dickman MJ, et al. Rapid resonance Raman microspectroscopy to probe carbon dioxide fixation by single cells in microbial communities. *ISME J* 2012;6(4):875–85.
- [30] McEwen GD, Wu Y, Tang M, Qi X, Xiao Z, Baker SM, et al. Subcellular spectroscopic markers, topography and nanomechanics of human lung cancer and breast cancer cells examined by combined confocal Raman microspectroscopy and atomic force microscopy. *Analyst* 2013;138(3):787–97.
- [31] Beljebbar A, Dukic S, Amharref N, Manfait M. *Ex vivo* and *in vivo* diagnosis of C6 glioblastoma development by Raman spectroscopy coupled to a microprobe. *Anal Bioanal Chem* 2010;398(1):477–87.
- [32] de Bourcy CF, De Vlaminck I, Kanbar JN, Wang J, Gawad C, Quake SR. A quantitative comparison of single-cell whole genome amplification methods. *PLoS One* 2014;9(8):e105585.
- [33] Zhang Q, Wang T, Zhou Q, Zhang P, Gong Y, Gou H, et al. Development of a facile droplet-based single-cell isolation platform for cultivation and genomic analysis in microorganisms. *Sci Rep* 2017;7:41192.
- [34] Xu P, Zheng X, Tao Y, Du W. Cross-interface emulsification for generating size-tunable droplets. *Anal Chem* 2016;88(6):3171–7.
- [35] Browne HP, Forster SC, Anonye BO, Kumar N, Neville BA, Stares MD, et al. Culturing of ‘unculturable’ human microbiota reveals novel taxa and extensive sporulation. *Nature* 2016;533(7604):543–6.
- [36] Lau JT, Whelan FJ, Herath I, Lee CH, Collins SM, Bercik P, et al. Capturing the diversity of the human gut microbiota through culture-enriched molecular profiling. *Genome Med* 2016;8(1):72.
- [37] Boedicker JQ, Li L, Kline TR, Ismagilov RF. Detecting bacteria and determining their susceptibility to antibiotics by stochastic confinement in nanoliter droplets using plug-based microfluidics. *Lab Chip* 2008;8(8):1265–72.
- [38] Eun YJ, Utada AS, Copeland MF, Takeuchi S, Weibel DB. Encapsulating bacteria in agarose microparticles using microfluidics for high-throughput cell analysis and isolation. *ACS Chem Biol* 2011;6(3):260–6.
- [39] Jiang CY, Dong L, Zhao JK, Hu X, Shen C, Qiao Y, et al. High-throughput single-cell cultivation on microfluidic streak plates. *Appl Environ Microbiol* 2016;82(7):2210–8.
- [40] Glass EM, Wilkening J, Wilke A, Antonopoulos S, Meyer F. Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harb Protoc* 2010;2010(1):pdb.prot5368.
- [41] Seshadri R, Kravitz SA, Smarr L, Gilna P, Frazier M. CAMERA: a community resource for metagenomics. *PLoS Biol* 2007;5(3):e75.
- [42] Shah N, Tang H, Doak TG, Ye Y. Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics. *Pac Symp Biocomput* 2011;16:165–76.
- [43] Mitra S, Stärk M, Huson DH. Analysis of 16S rRNA environmental sequences using MEGAN. *BMC Genomics* 2011;12(Suppl 3):S17.
- [44] Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009;75(23):7537–41.
- [45] Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* 2005;71(12):8228–35.
- [46] Hamady M, Lozupone C, Knight R. Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *ISME J* 2010;4(1):17–27.
- [47] Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;7(5):335–6.
- [48] Su X, Xu J, Ning K. Parallel-META: efficient metagenomic data analysis based on high-performance computation. *BMC Syst Biol* 2012;6(Suppl 1):S16.
- [49] Su X, Pan W, Song B, Xu J, Ning K. Parallel-META 2.0: enhanced metagenomic data analysis with functional annotation, high performance computing and advanced visualization. *PLoS One* 2014;9(3):e89323.
- [50] Jing G, Sun Z, Wang H, Gong Y, Huang S, Ning K, et al. Parallel-META 3: comprehensive taxonomical and functional analysis platform for efficient comparison of microbial communities. *Sci Rep* 2017;7:40371.
- [51] Huang S, Li R, Zeng X, He T, Zhao H, Chang A, et al. Predictive modeling of gingivitis severity and susceptibility via oral microbiota. *ISME J* 2014;8(9):1768–80.
- [52] Teng F, Yang F, Huang S, Bo C, Xu ZZ, Amir A, et al. Prediction of early childhood caries via spatial-temporal variations of oral microbiota. *Cell Host Microbe* 2015;18(3):296–306.
- [53] Guo Z, Zhang J, Wang Z, Ang KY, Huang S, Hou Q, et al. Intestinal microbiota distinguish gout patients from healthy humans. *Sci Rep* 2016;6:20602.
- [54] Baker JL, Bor B, Agnello M, Shi W, He X. Ecology of the oral microbiome: beyond bacteria. *Trend Microbiol*. Epub 2017 Jan 11.