



Building up a clinical microbiota profiling: a quality framework proposal

Valentin Scherz, Gilbert Greub & Claire Bertelli

To cite this article: Valentin Scherz, Gilbert Greub & Claire Bertelli (2022) Building up a clinical microbiota profiling: a quality framework proposal, *Critical Reviews in Microbiology*, 48:3, 356-375, DOI: [10.1080/1040841X.2021.1975642](https://doi.org/10.1080/1040841X.2021.1975642)

To link to this article: <https://doi.org/10.1080/1040841X.2021.1975642>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



[View supplementary material](#)



Published online: 09 Nov 2021.



[Submit your article to this journal](#)



Article views: 1180



[View related articles](#)



[View Crossmark data](#)



Citing articles: 1 [View citing articles](#)

Building up a clinical microbiota profiling: a quality framework proposal

Valentin Scherz , Gilbert Greub  and Claire Bertelli 

Institute of Microbiology, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland

ABSTRACT

Extensive characterization of the human microbiota has revealed promising relationships between microbial composition and health or disease, generating interest in biomarkers derived from microbiota profiling. However, microbiota complexity and technical challenges strongly influencing the results limit the generalization of microbiota profiling and question its clinical utility. In addition, no quality management scheme has been adapted to the specificities of microbiota profiling, notably due to the heterogeneity in methods and results. In this review, we discuss possible adaptation of classical quality management tools routinely used in diagnostic laboratories to microbiota profiling and propose a specific framework. Multiple quality controls are needed to cover all steps, from sampling to data processing. Standard operating procedures, primarily developed for wet lab analyses, must be adapted to the use of bioinformatic tools. Finally, requirements for test validation and proficiency testing must take into account expected discrepancies in results due to the heterogeneity of the processes. The proposed quality management framework should support the implementation of routine microbiota profiling by clinical laboratories to support patient care. Furthermore, its use in research laboratories would improve publication reproducibility as well as transferability of methods and results to routine practice.

ARTICLE HISTORY

Received 17 January 2021
Revised 4 May 2021
Accepted 8 July 2021
Published online 3 November 2021

KEYWORDS

Metagenomics; microbiome; microbiota profiling; quality management; biomarkers



1. Introduction


Sequencing-based methods have been successfully used to describe the microbial composition of complex samples for 30 years already (Stahl et al. 1984). The advent of rapid, affordable and performant next-generation sequencing (NGS) technologies prompted a surge in microbiota profiling studies (Sekirov et al. 2010), and unveiled the complexity of microbial communities inhabiting human niches (Escobar-Zepeda et al. 2015). However, the uncoordinated and rapidly-evolving jargon used to describe the many concepts in sequencing methods and microbiota analyses is source of misunderstanding (Marchesi and Ravel 2015). In this review, *microbiota profiling* will refer to the NGS-based analysis of microbial composition, either by targeted (*amplicon-based*) or untargeted (*shotgun*) metagenomics (Table 1).

1.1. Promises of microbiota profiling

Research on model organisms, microbiota transplant experiments and large-scale metagenomics studies contributed to evidence the possible role of microbiota in

health and disease (Sekirov et al. 2010; Yamashiro 2017), generating high expectations. Microbiota is now viewed as a complex ecosystem, which can affect the overall wellbeing of its host by modulating physiological systems (Marchesi et al. 2016; Yamashiro 2017; Durack and Lynch 2019; Fan and Pedersen 2021). Furthermore, metagenomics-derived metrics have been proposed as biomarkers (i) to screen for and diagnose diseases, (ii) to prognose disease outcome (iii) to predict treatment response or (iv) to follow therapeutic response (Ziegler et al. 2012; Rogers and Wesselingh 2016; Zmora et al. 2016; Kashyap et al. 2017; Malla et al. 2019) (Figure 1(E)). Metagenomics-guided dietary intervention for weight-management (Zmora et al. 2016; Bashiardes et al. 2018), prognosis of pulmonary exacerbation in cystic fibrosis (Acosta et al. 2018) or prediction of response to immune-checkpoint inhibitors in cancer therapies (Routy et al. 2018) are examples of proposed applications for microbiota-derived biomarkers. Microbiota is also regarded as a therapeutic target, for example using faecal microbiota transplant to treat recurrent or refractory *Clostridioides difficile* infections

CONTACT Claire Bertelli  Claire.Bertelli@chuv.ch  Institute of Microbiology, University Hospital Center and University of Lausanne, Rue du Bugnon 48, Lausanne 1011, Switzerland

 Supplemental data for this article can be accessed [here](#).

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Table 1. Definition of sequencing-based microbial assays.

	<i>Input sample</i>	<i>Culture isolate or normally sterile sample</i>	<i>Normally sterile sample</i>	<i>Complex sample</i>	<i>Complex sample</i>
<i>Methodological determinants</i>	Culture isolate Random shredding NGS	Culture isolate or normally sterile sample Targeted amplification Sanger	Random shredding Random shredding NGS	Targeted amplification NGS	Random shredding NGS
<i>Details</i>	WHOLE GENOME SEQUENCING WGS, Genomics Genomic characterization: <i>Taxonomic classification, typing, phylogeny, virulence, resistance</i> Epidemiological investigation, strain characterization Limited by culture Routine	BROAD-RANGE PCR Eubacterial PCR, Panfungal PCR Conventional molecular diagnostic: <i>Detect, identify and classify pathogens. Identify virulence and resistance factors</i> Strain identification, CNS infection, endocarditis Limited by primer specificity and amplicon discriminative power Routine	SHOTGUN METAGENOMICS Unbiased pathogen detection, clinical metagenomics Direct pathogen detection: <i>Detect, identify and classify pathogens. Identify virulence and resistance factors</i> CNS infection Different libraries required for RNA viruses Series of cases	AMPLICON-BASED METAGENOMICS Metabarcoding, Post-PCR metagenomics, 16S rRNA metagenomics Microbiota profiling: <i>Describe taxonomically or functionally (metabolism, resistance, virulence) microbiota constituents. Measure clinical biomarkers.</i> Gut microbiota profiling Limited by primers specificity, amplicon discriminative power Research	SHOTGUN METAGENOMICS Whole metagenome sequencing, metagenomics Microbiota profiling: <i>Describe taxonomically or functionally (metabolism, resistance, virulence) microbiota constituents. Measure clinical biomarkers.</i> Gut microbiota profiling Different libraries required for RNA viruses, limited by content of reference databases Research
<i>Phase of Development</i>	Routine	Routine	Series of cases	Research	Research

The definitions of terms used for sequencing-based microbiota assays lack strict consensus and, therefore, require clarification. Microbiota profiling approaches discussed in this review can rely either on amplicon-based or shotgun metagenomics, as described in the last two columns of the table.

(McDonald et al. 2018). Hence, microbiota profiling could contribute to tailor transplanted faecal products to the recipient and could become a routine safety assessment of the transplanted product (Woodworth et al. 2017; Smillie et al. 2018). Despite many applications of microbiota profiling being envisioned to determine biomarkers for tailored medical treatments in personalised medicine (Ziegler et al. 2012; Rogers and Wesselingh 2016; Zmora et al. 2016; Kashyap et al. 2017; Malla et al. 2019), their precise indications and clinical utility remains to be established and significant challenges must be addressed before their adoption in routine practice (Quigley 2017).

1.2. Hurdles of microbiota profiling

The intrinsic complexity of the microbiota, added to the complexity of metagenomics data and technical pitfalls in metagenomic assays, explain the difficulties encountered by the field to meet expectations for clinical applications (Schmidt et al. 2018). Indeed, the microbiota is a highly complex and dynamic assemblage of a large diversity of eukaryotes, bacteria, archaea and viruses (Lloyd-Price et al. 2016). Human microbiota composition is the product of seeding at birth, followed by complex symbiotic interactions among its microbial constituents as well as with the host (Lloyd-Price et al. 2016). This equilibrium is constantly altered throughout life by external factors including therapies (Maier et al. 2018), lifestyle (David et al. 2014), and diet (Danneskiold-Samsøe et al. 2019), leading to a high level of inter-individual variability (Lloyd-Price et al. 2016; Wagner et al. 2018). In return, microbial constituents can modulate the host's response to external stimuli including diet (Danneskiold-Samsøe et al. 2019) or therapies (Wilkinson et al. 2018). The influence of external factors added to the crosstalk between a myriad of microbial players within the host constitute as many potential confounding factors to be considered in microbiota studies. This intricacy renders difficult to go beyond associations and to define causality between microbiota composition and the occurrence of a disease or its evolution.

Metagenomics is used to describe the abundance of thousands of taxonomical (e.g. variants, OTUs, species) or functional (e.g. metabolic, resistance, virulence genes) features (Figure 1(B)). To be used as clinical biomarkers, these counts can be summarized in a variety of metrics, including composite scores such as the "Bacteroidetes to Firmicutes ratio" (Ley et al. 2005) or alpha-diversity indexes (Figure 1(C)). Alpha-diversity indexes are a group of metrics used to describe the

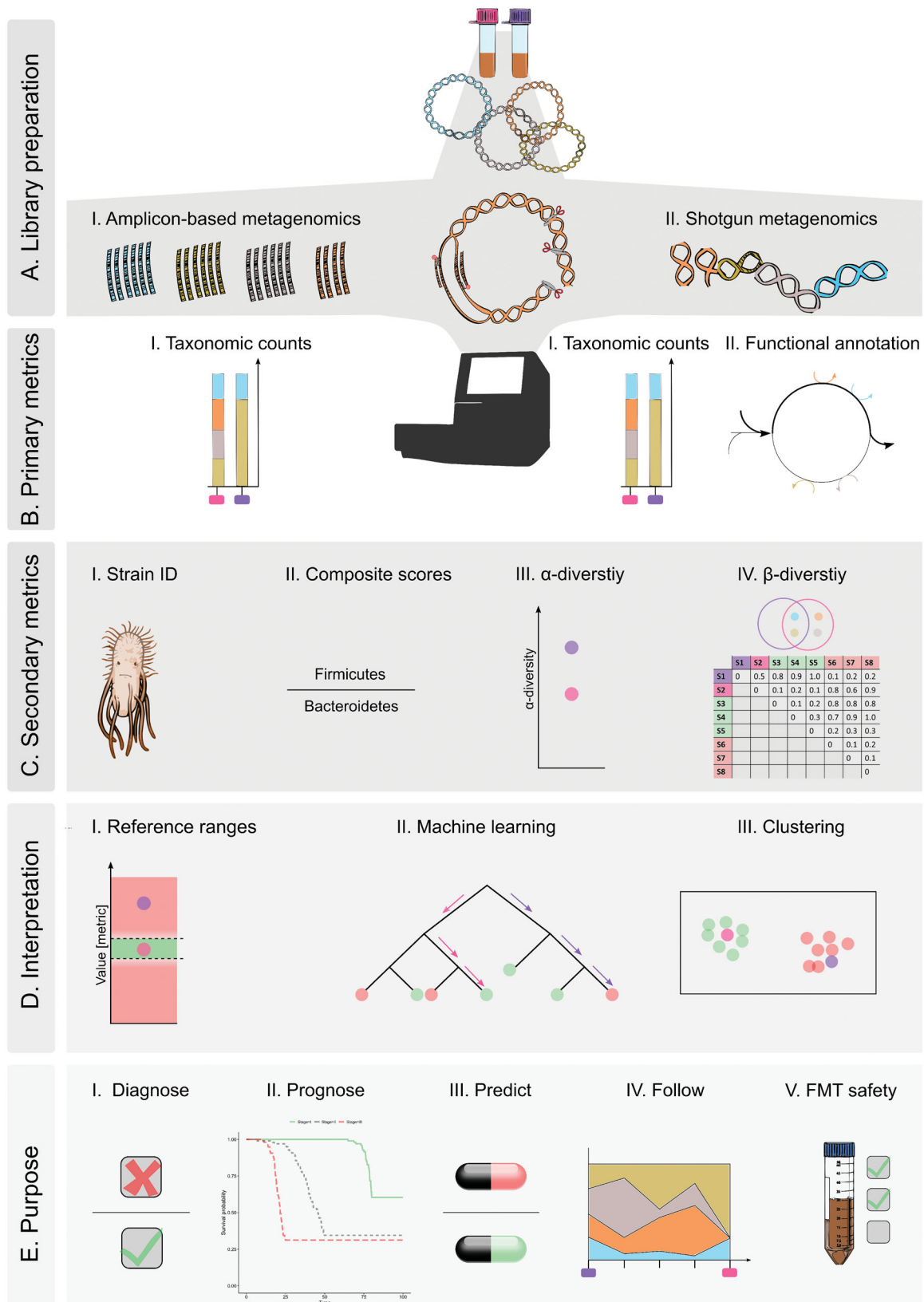


Figure 1. Microbiota profiling metrics as biomarkers. Measurement of microbiota profiling-derived biomarkers applied to two samples (pink and violet caps). A. Two main metagenomics approaches can be used to perform clinical microbiota profiling. A.I. Amplicon-based metagenomics rely on the amplification and sequencing of a target gene used as a “barcode” to identify microbial taxa. A.II. In shotgun metagenomics, DNA is randomly shredded and sequenced allowing taxonomic but also functional (e.g. metabolic) characterization of the samples. B.I. NGS reads are processed into data describing the taxonomic composition (amplicon-based metagenomics) or B.II. taxonomic and functional profiles (shotgun metagenomics). C. Various secondary metrics

Figure 1. (Continued)

computed from metagenomics data can serve as biomarkers: C.I. the abundance of specific taxonomic ranks or functional features; C.II. composite scores based on taxonomic or functional entities; C.III. α -diversity measures describing the microbial diversity for a given sample; C.IV. β -diversity approaches to compare patients' samples (e.g. healthy and diseased populations) by computing a matrix of pairwise distances between samples based on indexes such as Jaccard which equals the union (features in common) divided by the intersect (total number of unique features in both samples). D. Metrics require interpretation methods adapted to their nature: D.I. Abundance of specific features, α -diversity metrics or composite scores can be compared to pre-established reference ranges; D.II. Machine learning-based methods can be used to classify samples from pre-processed metrics or directly from whole metagenomics data; D.III. Clustering or multidimensional scaling methods could be used in β -diversity approaches. E. Microbiota profiling-based biomarkers could serve distinct clinical purposes: D.I. screen or diagnose diseases; D.II. prognose patient outcomes; D.III. predict treatment response; D.IV. follow up patients; D.V. tailor and insure safety of fecal microbiota transplant. Tubes and plasmids in this figure are modified from "Servier Medical Art by Servier", which is licensed under CC BY SA 3.0 license. The "Salmonella" drawing is reproduced from www.Krobs.ch; copyrights are owned by "The Institute of Microbiology (G.Greub)", who allowed its use for this figure only.

richness (the number of distinct features), the evenness (homogeneity of the distribution among features) or the diversity (richness weighted by the evenness) contained in a sample (Goodrich et al. 2014). These alpha-diversity indexes represent distinct, potentially relevant, dimensions of the microbiota which could provide actionable clinical information, for instance to prognose clinical outcome and predict therapeutic response of cancers (Routy et al. 2018; Riquelme et al. 2019; Peled et al. 2020) or to prognose long-term outcomes in cystic fibrosis (Acosta et al. 2018). Nevertheless, alpha-diversity indexes are reductive measurements that could hamper the identification of true microbial determinants of diseases (Hooks and O'Malley 2017; Shade 2017). To identify putative biomarkers from microbiota profiling data counts, it is hence tempting to directly correlate the abundance of all taxonomical or functional features to clinical phenotypes (Figure 1(C)). However, such correlations can lead to spurious results due to the high dimensionality, compositionality and sparsity of metagenomics data, requiring the use of adapted statistical methods (Gloor et al. 2017; Luz Calle 2019). Alternatively, various beta-diversity indexes (e.g. Jaccard, Bray-Curtis) can be used to assess the compositional overlap of samples, for instance to compare samples collected longitudinally from one patient (Goodrich et al. 2014). Stability or changes in microbiota composition reflected by beta-diversity indexes could support clinical follow-up and prognose exacerbations in chronic respiratory (Carmody et al. 2015) or digestive diseases (Kiely et al. 2018). Machine-learning models and network analyses are rapidly developing as new tools to identify microbial features as putative biomarkers (Ren et al. 2019; Marcos-Zambrano et al. 2021) or to prognose clinical outcomes directly from metagenomics data (Zhou and Gallins 2019; Lo and Marculescu 2019). While promising, machine-learning adds a new

layer of complexity, for instance regarding interpretation, validation and performance reporting (Chen et al. 2019; Wiens et al. 2019; Cammarota et al. 2020; Topçuoğlu et al. 2020). Altogether, these different approaches could be complementary to capture microbiota complexity. However, not all statistical methods are adapted to metagenomics data and the multiplicity of metrics and statistical methods limits comparisons between studies (Hooks and O'Malley 2017; Wagner et al. 2018).

Finally, the overlooked importance of methodological and technical determinants in metagenomics protocols undermines the translation of microbiota profiling into robust biomarkers. Recent observations raised concerns regarding the validity, reproducibility and comparability of metagenomics analyses (Pollock et al. 2018; Poussin et al. 2018; Schloss 2018; Hornung et al. 2019). For instance, associations between specific microbiota composition and diseases potentially resulted from contaminations in several studies testing low microbial biomass samples (Salter et al. 2014; Eisenhofer et al. 2019). In addition, simple protocol modifications can significantly alter the observed metrics and taxa in microbiota profiles (Brooks et al. 2015; Sinha et al. 2017; Pollock et al. 2018; Hornung et al. 2019) (Supplementary Table 1). Yet, despite the proposal of unified protocols for example by the Human Microbiome Project (Peterson et al. 2009), heterogeneous methods are still applied in research, which may hamper the identification and validation of clinically relevant biomarkers. Several reviews and initiatives have recently called for a better characterization of the effect of technical parameters on microbiota-derived metrics, systematic inclusion of control samples and standardization in the field (Kashyap et al. 2017; Pollock et al. 2018; Leigh Greathouse et al. 2019). Standardization will be critical to facilitate the

identification of microbiota-based biomarkers but also for their application in clinical laboratories which otherwise would have to faithfully implement multiple original protocols to offer biomarker assays.

1.3. Anticipating the need for quality management in clinical microbiota profiling

The establishment of clinically-validated microbiota-based biomarkers requires the resolution of many challenges that entails extensive efforts in translational research (Quigley 2017). Yet, we postulate that clinical microbiology laboratories will eventually be requested to provide microbiota profiling analyses for routine clinical applications. In this setting, the characterization of the microbiota must be based on appropriate analyses of data generated by accredited metagenomics protocols. Thus, laboratories will need quality management (QM) schemes adapted to the specific challenges of microbiota profiling by metagenomics. No recommendation exists yet to guide the implementation of QM for clinical microbiota profiling. The only existing quality recommendations for microbiota profiling by metagenomics have been tailored to the translational research setting (Nayfach and Pollard 2016; Hugerth and Andersson 2017; Sinha et al. 2017; Pollock et al. 2018; Poussin et al. 2018; Bharti and Grimm 2019; Eisenhofer et al. 2019), or for pathogen detection by metagenomics, which present distinct aims and requirements (Schlaberg et al. 2017; Chiu and Miller 2019; Miller et al. 2019) (Table 1).

Motivated by the promises of microbiota profiling, this review aims at proposing the basis of a QM adapted to clinical microbiota profiling by metagenomics. Building upon recommendations for microbiota profiling in translational research (Nayfach and Pollard 2016; Hugerth and Andersson 2017; Sinha et al. 2017; Pollock et al. 2018; Poussin et al. 2018; Bharti and Grimm 2019; Eisenhofer et al. 2019) as well as for clinical NGS-based assays, including metagenomics for pathogens identification (Schlaberg et al. 2017; Chiu and Miller 2019; Miller et al. 2019), our proposal covers pre-analytical, analytical and post-analytical phases. Its structure is based on classical QM, with validation and quality assurance (QA) composed of standard operating procedures (SOPs), quality control (QC), internal quality assessment (IQA) and external quality assessment (EQA). The principal factors identified as pitfalls for the translation of microbiota profiling into a diagnostic tool with identified solutions were summarized into [Supplementary Table 1](#).

2. Quality management for clinical microbiota profiling

2.1. Validation

As for any routine diagnostic method, the implementation of novel microbiota profiling-based assays will require analytical and clinical validations, completed by an assessment of clinical utility (gray frames in [Figure 2](#)) (Lundberg 1998; Jennings et al. 2009; Burd 2010; Glossary – BEST 2020). While the cumbersome assessment of microbiota-derived biomarkers will likely be conducted by large laboratories, or even consortia, individual clinical laboratories will have to verify that their implementation of externally-defined protocols reaches the expected level of analytical and clinical performance (Jennings et al. 2009; Burd 2010; Glossary – BEST 2020). In this context, the availability of reference samples and datasets for validation are key to assess adequately the implementation of any new assay (Gargis et al. 2016).

Formal guidance for the validation of microbiota profiling-based biomarkers should be provided by experts from the field and regulatory bodies. Existing recommendations for NGS-based clinical assays in general (Rehm et al. 2013; Aziz et al. 2015; Gargis et al. 2015; Endrullat et al. 2016), or in the specific setting of oncology (Jennings et al. 2009, 2017), heritable diseases (Centers for Disease Control and Prevention (CDC) 2009), clinical microbiology at large (Gargis et al. 2016) or specifically orientated towards metagenomics-based pathogen detection (Schlaberg et al. 2017; Chiu and Miller 2019; Miller et al. 2019) and forensics (Budowle et al. 2014) could serve as a basis to establish guidelines. Validation processes will have to be adapted to the specificities of the planned methodological approach and clinical purpose. Many putative clinical applications of microbiota profiling have been envisioned (Carmody et al. 2015; Kashyap et al. 2017; Woodworth et al. 2017; Kiely et al. 2018; Routy et al. 2018; Smillie et al. 2018; Riquelme et al. 2019; Peled et al. 2020), but to our knowledge none have been submitted to formal clinical validation yet. Hence, we propose at the end of this review three case scenarios exemplifying the potential pitfalls and existing solutions for the clinical validation of diverse microbiota-derived biomarkers.

Clinical utility

A biomarker offered by clinical laboratories should not only be clinically valid but also useful. In other words, a result should be clinically accurate (e.g. correctly diagnose a disease), but also have a positive impact for the patient (e.g. improve care) or the healthcare system

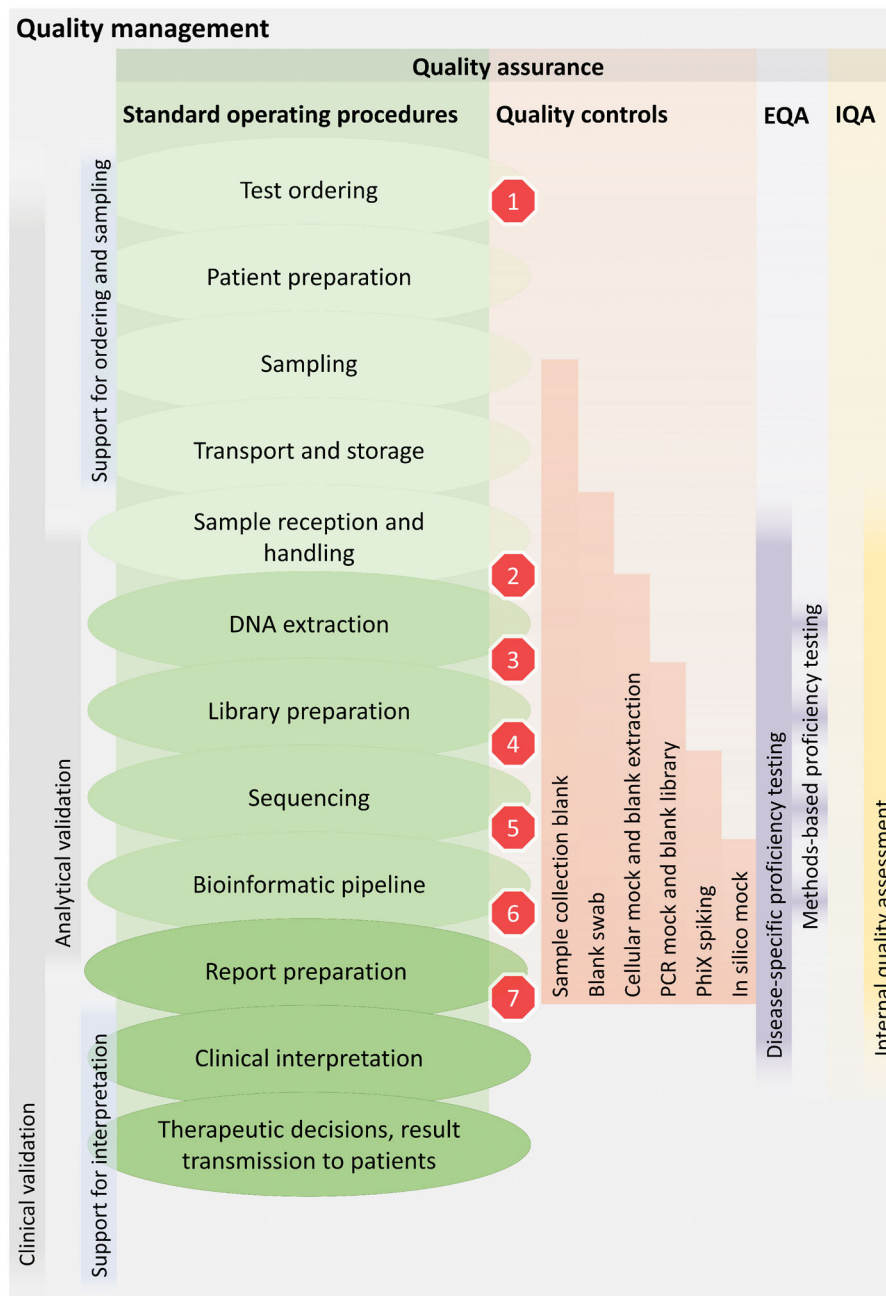


Figure 2. Quality management for microbiota profiling. Comprehensive quality management covers all pre-analytical (light green), analytical (green), and post-analytical (dark green) steps, from test prescription to clinical interpretation. These steps, including bioinformatics, must be standardized and recorded in standard operating procedures. Quality controls (QC) are constituted of positive and negative controls to cover most of the workflow (red). QC also includes “checkpoints” based on acceptance criteria (red hexagons representing stop signs): (1) Clinical indication; (2) Sample collection, transport and storage; (3) Extracted DNA concentration; (4) DNA library length profile and concentration; (5) Sequencing quality and yield; (6) Reads passing bioinformatic processing; (7) Reporting and clinical interpretation. External quality assessment (EQA) is either organized as “disease-specific” or as “method-based” proficiency testing. Internal quality assessment (IQA) completes EQA, especially in cases where an EQA program is not available. Well conducted analytical and clinical validations are prerequisites for any proposed assay. Laboratories must advise clinicians on the pre- and post-analytical steps to support adequate test ordering, sampling and results interpretation.

(e.g. reduce costs) (Lundberg 1998). This last validation stage is typically evaluated by clinical trials conducted once analytical and clinical validity have been

established. These, ideally prospective, studies are designed to evaluate the benefit provided by the use of a test (e.g. see Dobbin et al. (2016) for an example of

study design guidance adapted to predictive biomarkers in oncology).

2.2. Standard operating procedures

The definition of strict standard operating procedures (SOPs) for the entire microbiota profiling workflow (green frame in Figure 2) is particularly important given the known sensitivity of the results to the slightest changes in both wet lab protocols and bioinformatics pipelines (Kim et al. 2017; Sinha et al. 2017; Pollock et al. 2018; Bharti and Grimm 2019; Hornung et al. 2019). Indeed, simple inconsistencies, for instance in storage conditions or DNA extraction could compromise intra-laboratory reproducibility (Kim et al. 2017). Hence, loose protocol implementation in routine laboratories could render invalid the comparison of measured biomarker values with externally defined reference ranges.

Besides the traditional requirements for SOPs (written protocols, traceability), the stability of bioinformatic tools and related reference databases is a specific requirement for metagenomics. To facilitate the many applications of microbiota profiling, software for metagenomics analyses should be configurable and flexible, while remaining easy-to-use, rapid, stable and traceable (Grüning et al. 2018). Commercial tools such as CLC microbial genomics (Qiagen, Hilden, Germany) are an option to reach these requirements (Gargis et al. 2016). However, home-made or consortium-based pipelines are interesting as they offer more flexibility through a continuous and tailored development. In such cases, bioinformatic pipeline integration in open-source workflow managers such as Snakemake (Köster and Rahmann 2012) or Nextflow (Di Tommaso et al. 2017) ensure performance by automating task parallelization and usability with a single command to launch complex analyses. The traceability of changes in embedded bioinformatic tools and scripts must be ensured by versioning systems such as Git (Perez-Riverol et al. 2016). Furthermore, containers such as Docker (Docker, Inc, San Francisco, USA) or Singularity (Sylabs, San Francisco, USA) should be favoured to ensure software stability and reproducibility (Piccolo and Frampton 2016; Grüning et al. 2018).

2.3. Quality controls

Quality controls (QCs) are the cornerstone of QM. The analysis of internal quality controls represented by negative and positive samples allows to validate the successful completion of the entire workflow, from

sampling to bioinformatics analyses (red bars in Figure 2) (Pollock et al. 2018; Hornung et al. 2019). In addition, QCs also include “checkpoints”, where the process can only continue if acceptance criteria are met (checkpoint in Figure 2) (CDC 2009; Jennings et al. 2009; Rehm et al. 2013; Aziz et al. 2015). These controls monitored with a strict application of acceptance criteria ensure the intra-laboratory reproducibility as well as external validity.

Internal quality controls

The most commonly used positive controls are standardized mock communities, composed of a known proportion of complete microbes (cellular mocks) (Pollock et al. 2018; Yeh et al. 2018; Hornung et al. 2019), such as those provided by the American Type Culture Collection (ATCC, Manassas, USA). Alternatively, pre-extracted controls composed of standard microbial DNA (DNA mocks) or pre-generated sequencing reads (*in silico* mocks) can be used as positive controls (Motro and Moran-Gilad 2018; Pollock et al. 2018). However, only cellular mock communities assess comprehensively all steps of the process from DNA extraction to bioinformatic analysis. The need for additional controls can be technology-specific, like the PhiX spiked in libraries for Illumina sequencing (San Diego, CA, United States) (Illumina 2020).

Negative controls are central to metagenomics, due to the constant occurrence of contaminants. These contaminants originate from the lab environment, reagents (“kitome”) or neighbouring samples (cross-contamination) and add up during sampling, DNA extraction, library preparation and sequencing (“index-hopping”) (Salter et al. 2014; Minich et al. 2019). Adequate use of negative controls has been well described in a review by Eisenhofer et al. (2019). Briefly, extraction and library preparation blanks, made from reagents without template, are the minimal negative controls to detect contaminants (Pollock et al. 2018; Eisenhofer et al. 2019; Hornung et al. 2019). Additional sampling negative controls could be included (Hornung et al. 2019), for instance aspirated water used for bronchoalveolar lavage (Charlson et al. 2011). It is recommended to sequence all negative controls, even if no DNA is detected, to assess the occurrence of stochastic contaminations (Salter et al. 2014; Kim et al. 2017). A surveillance of biases and identification of unexpected contamination events should be implemented by recording laboratory contaminants over time into a database (Chiu and Miller 2019).

The use and interpretation of results from negative controls remain an open research field (Karstens et al. 2019), but the following strategies have been proposed; Putative contaminants can be removed by filtering all taxa found in negative controls, with the risk of generating false negatives, or taxa below a cut-off, either arbitrarily hard coded or calculated based on the content of negative controls (Karstens et al. 2019). Recently published tools such as *decontam* (Davis et al. 2017), *microDecon* (McKnight et al. 2019) and *Recentrifuge* (Martí 2019) proposed elaborate methods for the identification of contaminants. These tools differ in their compatibility with amplicon-based or shotgun metagenomics data, the method to identify contaminants (based on taxa prevalence in negative controls or on the inverse correlation of contaminant abundance with library DNA concentration) and their filtering (blunt removal of taxa classified as contaminants or subtraction of the proportion of counts explained by contamination). However, only *Recentrifuge* is designed to identify cross-contaminations (Martí 2019). Furthermore, the need for careful evaluation of negative controls depends on the specimen analysed and the intended use. Low biomass samples such as skin or lower respiratory tract samples are more likely to be significantly impacted by contaminants compared to faecal samples (Eisenhofer et al. 2019; Karstens et al. 2019). Biomarkers considering the number of distinct taxa, even at low abundance (richness), are also more likely to be affected by contaminants than broader biomarkers, for instance based on Simpson alpha-diversity metrics (Karstens et al. 2019).

Checkpoints

QC “checkpoints” determine at each step of the workflow if the process can be pursued based on acceptance criteria (checkpoints in Figure 2). There is currently no consensus on metagenomics-derived metrics to observe for quality assessment of wet-lab, sequencing and bioinformatics steps and on acceptance criteria to apply. Definition of objective, validated criteria is difficult to achieve since they depend on the applied approach (amplicon-based versus shotgun metagenomics), the sequencing technology and the bioinformatic tools. Furthermore, their interpretation may depend on the intended approach (e.g. a given sequencing yield may be sufficient to generate a broad taxonomic profile but not to assemble genomes from metagenomics data (Hillmann et al. 2018)). In this context, laboratories have to start with arbitrary criteria and then adapt these criteria based on their experience, cut-off values adopted in reference publications or expert opinions and

conventions. To help other centres in establishing their own workflow, we have compiled acceptance criteria currently applied in our laboratory in Table 2.

Illustrated by the first checkpoint in Figure 2, reviewing the appropriateness of test ordering is critical to ensure the analysis can provide an answer to the clinical questions, given the complexity of microbiota profiling interpretation and the costs of the analysis (Dickerson et al. 2014). Appropriate test prescription should be promoted by clear indications for testing, ordering restrictions, mandatory counselling and critical reviews of past orders (Riley et al. 2015). Second, sample type and quality must be assessed in view of the clinical question (checkpoint 2 in Figure 2). A precise sampling site could be needed due to the significant differences in microbial profiles between anatomical niches (Jones et al. 2018). Patient preparation (e.g. delay after a therapeutic procedure, time of the day) could also be relevant, due to potential confounding effects of therapeutic procedures, meals or circadian rhythms (Harrell et al. 2012; Liang and FitzGerald 2017; Takayasu et al. 2017). Finally, adequacy of storage time and conditions must be reviewed, given their effect on the observed microbiota (Jenkins et al. 2018).

Measuring DNA yields after extraction (checkpoint 3 in Figure 2) is of interest but not always necessary since indirectly assessed by the following quantification of libraries. Concentrations of extracted DNA can be measured by qPCR, fluorescence spectroscopy (e.g. Qubit, ThermoFisher Scientific Inc., Waltham, MA, USA), or spectrophotometry (e.g. NanoDrop ThermoFisher Scientific Inc., Waltham, MA, USA), with different advantages and drawbacks (Simbolo et al. 2013; Hussing et al. 2018). The need for systematic DNA quantification to detect extraction failure could depend on the nature of the sample. In high microbial biomass samples (e.g. faeces), unexpectedly low library or sequencing reads should be interpreted as a failure of DNA extraction or library preparation. Retroactive quantification of extracted DNA can be limited to these cases to disentangle the cause of failure. Conversely, in samples with low to moderate microbial biomass (e.g. respiratory or skin), low library yields can occur frequently, which could justify the prospective quantification of all extracted DNA to assess the success of library preparation and sequencing. Furthermore, low biomass samples can generate significant quantities of sequencing reads, but should be considered at higher risk of contaminations (Eisenhofer et al. 2019; Karstens et al. 2019). Targeted qPCR can quantify specific organisms (bacteria with broad-range 16S rRNA qPCR, fungi with broad-range ITS qPCR), when quantification by

Table 2. Acceptance criteria for microbiota-profiling assays.

PHASE	ITEM	READ-OUT	VALUES	DETAILS	PROPOSED ACCEPTANCE CRITERIA
Pre-analytical	<i>Clinical indication</i>	Biomedical validation	Appropriate/ not appropriate	The clinical indication for testing should rely on evidence supporting the clinical utility of the assay for screening, diagnosis, prognosis, prediction, or follow-up purpose. The demanded assay should be able to capture the organism involved in the physiopathological process of interest (e.g. include reverse complement library for RNA viruses).	<ul style="list-style-type: none"> Clinical utility of microbiota-profiling has not been formally demonstrated for any indication yet. Thus, indication should be discussed between clinicians and clinical microbiologists experimented in microbiota profiling prior to testing.
		Biomedical validation	Appropriate/ not appropriate	The clinical specimen was collected in conditions similar to the conditions in which the test was clinically validated (e.g. time of the day, distance to medical or surgical interventions).	<ul style="list-style-type: none"> NA
		Visual inspection	Appropriate/ not appropriate	Clinical specimen is appropriate in terms of quality (faecal staining of rectal swab, sputum thickness, ...) and quantity Sample collection blank can be analysed to validate that the sampling device, or tube, is appropriated to the planned analysis during workflow validation stages.	<ul style="list-style-type: none"> Faecal staining for rectal swabs, thick mucus and not salivary liquid for sputum-based analysis of lower respiratory tract microbiota. Sample collection blank yields a minimal number of reads (<1000), close to the number of reads obtained from negative library preparation controls. Taxa systematically found in these controls are recorded and considered during data interpretation. They will be considered only in low-biomass samples and/or if represented by a significant number of reads
Analytical	<i>Sample logistics</i>	Sample records	Appropriate/ not appropriate	Storage time and temperature are within acceptance criteria.	<ul style="list-style-type: none"> 20 °C freezing within 2h, DNA extraction within 7 days -80 °C within 2h for long term storage Use of preservation buffer stabilizing the sample Extracted DNA quantification is not applied in our routine. Library yield is used as a surrogate of extracted DNA loads.
		Nucleic acids quantity and purity	Undetectable or traces/ low/ high nucleic acid concentration	Quantification of specific organisms (e.g. bacteria by 16S rRNA qPCR) or whole DNA/RNA, (fluorescence spectroscopy, spectrophotometry)	<ul style="list-style-type: none"> Negative extraction controls from collection device yields a minimal number of reads (<1000), close to the number of reads obtained from negative library.
		DNA fragment length and quantity	Undetectable or traces/ low/ high DNA concentration	Undetectable or low extracted DNA amount can be expected in low biomass samples (e.g. skin swabs) or problematic in high biomass samples (e.g. faeces, rectal swab). Unexpected results can motivate DNA re-extraction or sample refusal, acceptance with expression of concerns regarding contaminants or simple acceptance.	<ul style="list-style-type: none"> Taxa found in these controls are recorded and considered during data interpretation. A positive mock community is systematically extracted with samples of interest. The taxonomic profile obtained for these samples should be stable.
<i>Library preparation</i>	DNA fragment length and quantity	Fragmentation and selection method	Fragmentation and selection method	In amplicon-based metagenomics, fragment length profiles should correspond to the length of the target sequence, added to the length of the adapters included for sequencing.	<ul style="list-style-type: none"> Undetectable concentrations based on Qubit and Fragment Analyser → DNA traces; proceed further if expected to be a low biomass sample and with caution regarding contaminants.
		Fragmentation and selection method	Fragmentation and selection method	In shotgun metagenomics, fragment length profiles will depend on the applied fragmentation and selection method.	<ul style="list-style-type: none"> 5–20 nM → low starting nucleic acid concentration; contaminants can still represent a significant (>1%) proportion of reads.

(continued)

Table 2. Continued.

PHASE	ITEM	READ-OUT	VALUES	DETAILS	PROPOSED ACCEPTANCE CRITERIA
				<p>The composition of the library depends on the method applied. In amplicon-based metagenomics, the library quantified mostly represents microbial amplicons, even if traces of host DNA can lead to over-estimations of the effective amplicon concentration. Conversely, depending on the sample type, shotgun libraries will contain a significant proportion of nucleic acids from the host. Targeted quantification by qPCR allows to measure the source organisms (e.g. bacterial, fungal or host)</p> <p>Molar concentration [nM] based on average fragment length [bp] and DNA concentration [ng/μl] is required for library normalization in Illumina sequencing.</p> <p>With Illumina, amplicon-based metagenomics will generally display poorer quality than shotgun-metagenomics or whole-genome sequencing (WGS) due to low sequence diversity. Quality systematically declines towards the end of the read and is lower for reverse reads than for forward reads.</p> <p>All samples should display comparable sequencing coverage thanks to library normalization. However, some variability is expected.</p> <p>Sequencing quality is influenced by many factors in Illumina sequencing, including the concentration of library loaded which impacts cluster density, sequence diversity, and sequencing length.</p>	<ul style="list-style-type: none"> • >20nM → high starting nucleic acid concentration: expected low impact from contaminants. • Negative library preparation controls yield a minimal number of reads (<1000). • Taxa systematically found in these controls are recorded and considered during data interpretation.
	Sequencing	Read quality and quantity	Q30, Phred score, visual inspection of quality curves, number of reads per sample (FastQC, Basespace)	<p>With Illumina, amplicon-based metagenomics will generally display poorer quality than shotgun-metagenomics or whole-genome sequencing (WGS) due to low sequence diversity. Quality systematically declines towards the end of the read and is lower for reverse reads than for forward reads.</p> <p>All samples should display comparable sequencing coverage thanks to library normalization. However, some variability is expected.</p> <p>Sequencing quality is influenced by many factors in Illumina sequencing, including the concentration of library loaded which impacts cluster density, sequence diversity, and sequencing length.</p>	<ul style="list-style-type: none"> • Sequencing run passed Illumina embedded automatic QC with successful basecalling and read demultiplexing. • Proportion of clusters passing filters (%PF) are > 85% and the proportion of bases with Phred score > 30 (Q30) is > 75%. Lower yield or quality can be acceptable if bioinformatic processing includes read filtering based on quality but should prompt investigation to optimize sequencing yield. • For faecal samples, 50 000 sequences per sample after bioinformatic read processing (see <i>Read processing</i>) is considered sufficient. Less (20 000) could be acceptable for samples with lower diversity (e.g. lower respiratory tract). • Rarefaction curves of richness should reach a plateau. Importantly, a plateau could never be reached when observing raw reads because of random sequencing errors or in highly diverse samples. Conversely, richness can be underestimated, and a plateau reached too early, when relying on ASVs generated by DADA2 due to singleton filtering and interpretation of rare variants as sequencing errors. • Other quality metrics available in Basespace and FastQC, or generated along read processing (e.g. with the DADA2 pipeline) are reviewed if an insufficient number of reads are obtained after read processing. • Average size below expected length (~297 for 300bp paired-end sequencing of the V3V4 16S rRNA fragment) is suspicious of insufficient library concentration. Typically found in negative controls.
		Sequence length	Average length of reads for each sample (FastQC)	<p>Depends on the sequencing length and the library length profile. Should be maximal (as long as the sequencing length) if the DNA fragments are significantly longer than the sequencing length.</p>	

(continued)

Table 2. Continued.

PHASE	ITEM	READ-OUT	VALUES	DETAILS	PROPOSED ACCEPTANCE CRITERIA
	Adapter Content	Adapter Content	% of library adapters in the middle of reads	Adapter sequences should not be in the middle of the reads. This typically occurs when the template DNA is insufficient leading to aberrant DNA constructs during library preparation.	<ul style="list-style-type: none"> These sequences should be filtered out during read processing. Indicate insufficient library concentration.
	<i>Bioinformatic pipeline</i>	Read processing	Sequence loss along the bioinformatic pipeline	<p>Bioinformatic processing of reads into exploitable sequence counts and classification requires multiple steps. The processing will depend on the metagenomic approach (amplicon-based vs shotgun), the sequencing technology (long versus short reads) and the expected outcome (e.g. taxonomic profiles, carried metabolic functions, presence of resistance or virulence genes).</p> <p>In the example of DADA2, a relatively recent R package, which is one of the main methods for Amplicon Sequence Variant (ASV) analysis, this typically includes:</p> <ul style="list-style-type: none"> PCR primer removal Sequence trimming (e.g. low quality ends of reads) and filtering based on the predicted rate of errors per read (calculated as estimated for the Phred score). Correction of read sequencing errors Merging of paired-end reads Chimaera filtering <p>Evaluation of the number of reads passing each processing step can point towards erroneous parameters in the pipeline, but also technical problems in wet-lab or sequencing steps.</p> <p>Alternatively, shotgun metagenomics processing based on marker-gene mapping (e.g. MetaPhiAn), read classification (e.g. Kraken) or assembly (e.g. metaSpades) have specific constraints which will require adapted QC and acceptance criteria.</p>	<ul style="list-style-type: none"> Most reads fed to the pipeline should pass all steps (e.g. > 80%). If not, this could point towards inappropriate parameters in the pipeline or technical problems during library preparation or sequencing. Read loss after PCR-primer removal can indicate a mismatch between the primers provided to the pipeline those found in the reads. Read loss after quality filtering can indicate lower than expected quality of the sequencing run. Read loss after read merging can indicate that reads do not overlap or are aberrant construct typically generated by PCRs performed on insufficient DNA concentration. Read loss after chimaera filtering can indicate high proportion of aberrant sequences, typically generated from PCR performed on insufficient DNA concentration.
	Taxonomic assignment	Taxonomic assignment	Classification scores, unexpected taxa	<p>Consensus sequences generated by the read processing pipeline are assigned to a taxonomic classification (or a functional classification) based on a reference database. Tools for classification generally generate scores indicating the level of confidence in the assignment, typically based on the similarity of the queried sequences with one, or multiple, sequences in the reference database. Unassigned sequences or sequences assigned with a low level of confidence should be the subject of concerns.</p>	<ul style="list-style-type: none"> Unassigned or loosely assigned sequences (e.g. only assigned to a Phylum) are suspicious, and potentially aberrant sequences, either chimeric, from an unspecific priming of host DNA or contaminants (e.g. mitochondria or chloroplast). These should be filtered out. A phylogenetic tree of the sequences should be reviewed as sequences unrelated to the microorganism of interest will branch away from the sequences of interest (e.g. a mitochondrial sequence will appear as distantly related to 16S rRNA sequences).

(continued)

Table 2. Continued.

PHASE	ITEM	READ-OUT	VALUES	DETAILS	PROPOSED ACCEPTANCE CRITERIA
Post-analytical	Reporting	Summary compositional plots	Summary plots faithfully and adequately represent composition	Microbial composition is often summarized into compositional plots. In most cases, a grouping of sequence counts at a higher taxonomic rank (e.g. family or genus) and a filtration or grouping of rare taxa is used to maintain figure readability. However, this transformation should not hide the taxa of potential interest. Library normalization breaks the correlation between starting DNA concentrations and generated sequencing read counts. Thus, plots should generally be represented as proportions. Taxonomic classification at lower levels (species, clades) can be erroneous since different taxa can carry identical sequences, especially when considering a short or conserved sequence. Alpha-diversity indices, and especially those describing richness, are significantly influenced by starting DNA concentrations and sequencing coverage. Thus, sample coverage should be similar to the coverage of samples used to determine reference-ranges. The plateau in rarefaction curves of the index is an indicator of the coverage allowing to capture adequately sample diversity.	<ul style="list-style-type: none"> • Assignment of taxa or sequences of specific interest are double-checked by independent methods. If multiple taxa could be assigned to a sequence, all taxa are listed. • Compositional plots are browsed in detail, helped by interactive KRONA plots to explore lowly abundant but potentially relevant taxa (based on the clinical question). • Heatmaps are preferred to barplots for results transmission since they allow better visualization of lowly abundant taxa. • Reads counts are normalized, to provide the proportion (%) of taxa. • The compositional nature of results is underlined in the reporting to clinicians.
		Alpha-diversity indices	Alpha-diversity score are accurate		<ul style="list-style-type: none"> • Analysed samples displayed >20nM of library DNA and 50 000 exploitable reads (for faecal samples). • The rarefaction curve of the alpha-diversity index of interest reaches a plateau, before the minimal sample coverage. • Sample coverage is comparable to the coverage used to determine reference values, or for all samples included in the comparison. If not, rarefaction to a value at which most of samples reached a plateau is used. • Some indexes (ACE, Chao1) define a standard error that should be reported. • Analysed samples displayed >20nM of library DNA and 50 000 exploitable reads (for faecal samples). • Coverage should be comparable for all samples included in comparison. If not, it is a subject of debate, but we use rarefaction to a value at which most of samples reached a plateau in richness. • Aberrant sequences signed by the fact that they are far from any bacterial sequences in our reference database are filtered out. • NMDS are provided with a stress value and a Shepard diagram to display the goodness of fit. • PCoA are provided with the explained values to report the goodness of fit.
		Comparison by beta-diversity approaches	Beta-diversity representations are accurate	Beta-diversity based comparisons are significantly influenced by starting DNA concentrations and coverage. Thus, sample coverage should be similar for all samples included in the comparisons. A plateau in rarefaction curves in alpha-diversity richness index is an indicator of sufficient coverage. Presence of contaminating sequences, either from cross-contaminations or from host or environmental DNA (mitochondria, chloroplast) can artificially distance samples, especially in phylogenetically weighted distances (unifrac). Thus, those aberrant sequences should be controlled for and filtered if necessary.	

(continued)

Table 2. Continued.

PHASE	ITEM	READ-OUT	VALUES	DETAILS	PROPOSED ACCEPTANCE CRITERIA
		Composite scores	Composite scores computed are accurate.	Representation of distance matrices into ordinations can be unable to faithfully represent the actual distances between samples. Representation adequacy should be controlled	
		Machine-learning classification/prediction	Machine-learning based classification or prediction is accurate.	Will depend on the actual score in use.	• NA
				Machine-learning classifier used to classify samples (e.g. into diseased or healthy) should be internally and externally valid. Classifiers are trained on a training dataset and their performance is assessed on an independent confirmation dataset. The trained classifier should have a good performance on data generated from the same laboratory it was trained in (internal validity), but also in other laboratories which could be willing to use it (external validity).	• NA

Examples of acceptance criteria used in a laboratory for each step from the pre-analytical to the post-analytical stages.

fluorescence spectroscopy or spectrophotometry would be impacted by host DNA concentrations. Another benefit of bacterial or fungal quantification by targeted qPCR is to transform sequencing counts (Jian et al. 2020). Indeed, some authors proposed to multiply the bacterial or fungal loads by the relative distribution of taxa to transform relative quantification into absolute values (Jian et al. 2020) or to subtract contaminants (Lazarevic et al. 2016).

DNA library quantification is needed to normalize libraries prior to sequencing and, as mentioned above, may also substitute quantification of extracted DNA yields to check sample quality (checkpoint 4 in Figure 2) and to roughly evaluate the original microbial loads. Libraries obtained from each sample are normalized to equimolar concentrations, usually after quantification by fluorometric methods (e.g. Qubit) and DNA fragment length profiling by capillary gel electrophoresis (e.g. Fragment Analyser, Agilent, Santa Clara, USA) (Illumina Inc 2016).

NGS sequencers measure many quality metrics to check during sequencing QC (checkpoint 5 in Figure 2). In the case of Illumina, cluster density, sequencing yield, error rate, proportion of reads over Q30 and passed filter clusters can be directly evaluated from their proprietary BaseSpace platform. Tools like FastQC help completing this evaluation, with information on average read length, Phred score profiles, uncalled base content and adapter or index contaminations (Andrews 2010). In our experience, review of sequencing yield (reads per sample), quality (Q30 score and per base sequence quality plots) and reads length (average read length) are the most informative metrics to identify problematic samples. Other metrics provided by these tools can help identifying the origin of sequencing failure for isolated samples or the whole sequencing run.

Raw read processing into exploitable classified sequences and counts require a combination of computational steps to perform quality controls, read filtering (based on quality and/or length), trimming (read cutting based on quality and/or length), clustering (sequence grouping based on similarity), assembly (reads merging into larger contigs), and mapping (sequence alignment against references) (Bharti and Grimm 2019). We recommend two reviews for in-depth discussion of the numerous bioinformatic tools available to process raw amplicon-based or shotgun metagenomics reads into exploitable taxonomic, metabolic, resistance or virulence data (Bharti and Grimm 2019; Breitwieser et al. 2019). All of these algorithms generate scores and logs that must be checked to assess the successful completion of bioinformatic processing (checkpoint 6 in Figure 2).

A final QC should be conducted before reporting results to clinicians by a clinical microbiologist trained in metagenomics (checkpoint 7 in Figure 2). As for any routine assay, this final biomedical validation should ensure the adequate completion of the workflow. Furthermore, the adequacy of the reported metrics and figures should be evaluated before transmission of the report to the clinicians. For instance, rare taxa are usually hidden in graphical representation of microbiota composition. However, some of these rare taxa could be of interest in certain cases. Furthermore, microbiologists specialized in metagenomics will have to proactively support clinicians in their interpretation of metagenomics results. Indeed, in the initial times of clinical microbiota profiling, physicians will have limited knowledge and insufficient expertise to understand complex microbiota-derived biomarkers. On the long term, adapted teaching and education will be crucial to democratize the understanding of microbiota dynamics and favour critical interpretation of metagenomics results (Gargis et al. 2016).

2.4. External and internal quality assessment

External Quality Assessment (EQA), also known as Proficiency Testing (PT) programs, submit reference material to participating laboratories and compare their results (violet frames in Figure 2). EQA programs will have to be adapted to the specificities of microbiota profiling. Multiple microbiota-derived metrics can be greatly influenced by minor modifications in protocols. In response to these constraints, we propose to consider two complementary forms of EQA for microbiota profiling, adapted from those applied in NGS of host-genetics: *disease-specific* and *method-based* (Kalman et al. 2013; Schrijver et al. 2014).

Disease-specific EQA

The *disease-specific* EQA assesses the congruence of clinical conclusions between laboratories (Kalman et al. 2013). It is expected that different laboratories, using different protocols, would provide discordant intermediate results (e.g. Shannon index of alpha-diversity). These laboratories could even support their conclusions based on different metrics. Yet, these laboratories could agree on clinical conclusions (e.g. *high risk of failure for faecal microbiota transplant*) when comparing the results obtained to their internal reference ranges or cut-off values. Nevertheless, the *disease-specific* approach is currently limited by the absence of clinically validated indications for microbiota profiling.

Method-based EQA

Conversely, a *method-based* EQA compares results disregarding any clinical question but focusing on the raw performance of one or more analytical steps (Kalman et al. 2013; Schrijver et al. 2014). This EQA would assess if identical protocols generate identical results when starting from the same material. For instance, taxonomic identification, quantification and extrapolated metrics could be surrogates evaluated by *method-based* EQA. Such an EQA program would allow laboratories to verify that they have faithfully implemented shared protocols. As compared to *disease-specific* EQA, *method-based* EQA could be more adapted to the current phase of development of clinical microbiota profiling since it can be conducted without any precise clinical indication.

Internal quality assessment

Internal quality assurance (IQA) encompasses different types of quality assessments, including reprocessing of reference material or processing of split samples (Scherz et al. 2017) (yellow frame in Figure 2). IQA and challenges between local laboratories should be organized to complement or to compensate the current lack of EQA programs for microbiota profiling (Kalman et al. 2013; Sinha et al. 2017).

3. Case scenario for validation of microbiota profiling-based biomarkers

To our knowledge, none of the biomarkers recently proposed by research studies has been considered sufficiently promising to be implemented in routine practice. Hence, validation of microbiota profiling-based biomarkers remains an ongoing and cumbersome task. Meanwhile, the three case scenarios presented here, and an additional case available as [Supplementary material](#), anticipate some of the standing challenges and practicable solutions in the validation of different types of putative biomarkers.

Case 1: *Akkermansia muciniphila* as a predictive biomarker of check-point inhibitors therapeutic response

Several bacterial taxa found in gut microbiota were associated to immune check-point inhibitors response in different types of cancer (Chaput et al. 2017; Matson et al. 2018; Routy et al. 2018). The presence of *Akkermansia muciniphila* was for instance associated to a higher probability of response to anti-PDL-1 therapy in non-small cell lung cancer patients (Routy et al. 2018). The following case scenario considers the validation of *Akkermansia muciniphila* detection by

metagenomics as a predictive biomarker for anti-PDL-1 therapy in non-small cell lung cancer.

Analytical validation of microbiota profiling in this application would be relatively straight-forward and would only require to assess the performance of the workflow to recover, identify and quantify this sole taxon. Existing guidelines for NGS-based pathogens detection (Schlaberg et al. 2017; Chiu and Miller 2019; Miller et al. 2019) and detection by molecular assays in general would hence apply (Burd 2010). The analytical performance could be assessed on spike-in samples or by orthogonal testing with qPCR (Schlaberg et al. 2017; Chiu and Miller 2019; Miller et al. 2019).

Clinical validation of the detection of *Akkermansia muciniphila* by metagenomics as a biomarker to predict treatment response should be provided by adapted clinical trials designed following existing recommendation for clinical validation of predictive biomarkers (Mandrekar and Sargent 2009; Dobbin et al. 2016). For instance, the performance (e.g. sensitivity/specificity) of *Akkermansia muciniphila* detection could be evaluated in prospective studies conducting gut microbiota profiling at treatment initiation and correlating results to a 6-month treatment response (measuring features specific to the clinical response to immune checkpoint inhibitors (Nishino et al. 2017)). Importantly, the structure of the investigated population should be carefully considered for the clinical validation of microbiota-derived biomarkers to account for potential confounding factors such as ethnicity, geography or diet that can all have a significant effect on microbial composition (Gupta et al. 2017; Gaulke and Sharpton 2018).

Case 2: Alpha-diversity as prognostic biomarker in allogeneic Hematopoietic-Cell transplantation

Clinical use of alpha-diversity-based biomarkers leads to specific technical pitfalls and challenges in interpretation. Alpha-diversity indexes summarises the distribution of features (e.g. species, operational taxonomic units, sequence variants or genes) in a sample in terms of richness (counts of different features), evenness (homogeneity of their distribution) or diversity (richness weighted by evenness). Various formula are used to weight or correct the observed distributions (e.g. Chao and Fischer richness or Shannon, Simpsons or inverse of Simpson diversity indexes) (Goodrich et al. 2014). These indexes are commonly used as broad descriptive measures in research but some associations with clinical evolution were observed for instance in inflammatory bowel diseases (Ananthakrishnan et al. 2017), cystic fibrosis (Cuthbertson et al. 2020) or intensive care units (Lamarche et al. 2018). In this case scenario, we will focus on a recent study by

Peled et al. who reported a significant correlation between the inverse of Simpson diversity in gut microbiota before hematopoietic-cell transplantation and overall patient survival (Peled et al. 2020).

Transformation of metagenomics data into alpha-diversity metrics such as the Simpson index or its inverse will require adapted analytical validation. Microbiota descriptive metrics typically represent values for which the ground truth is unknown (e.g. the actual number of species in a faecal sample represented by richness indexes). Reference material, for instance provided by EQA, could serve as gold standard (Burd 2010). Alternatively, the clinical diagnosis itself can serve as gold standard (Burd 2010). In this example, one should test directly the correlation between the inverse Simpson index measured by the workflow and patient outcome. Once validated, the analytical workflow should remain unchanged and applied under strict acceptance criteria for both wet-lab and bioinformatic processing since both can have a significant effect on obtained alpha-diversity results (Dybboe Bjerre et al. 2019; Prodan et al. 2020). In particular, the definition of minimal extracted DNA concentration and sequencing coverage should be strictly defined as both can lead to significant underappreciation of the diversity (Rodriguez-R and Konstantinidis 2014; Multinu et al. 2018). However, changes in the bioinformatics workflow could be validated retrospectively on existing reference datasets used for the primary validation.

When used as clinical biomarkers, alpha-diversity metrics can be compared to reference ranges (or intervals) for interpretation. In the present examples, Peled et al. defined the median of diversity values as a cut-off between two populations of high and low-diversity patients. In reality, these reference values should be determined by dedicated studies evaluating metric distribution, as recommended for molecular assays in microbiology (Burd 2010) or clinical chemistry (Henny et al. 2016).

Clinical validation of microbiota-based biomarkers is complicated by the large effect of confounding factors on microbiota, including ethnicity, geography or diet (Gupta et al. 2017; Gaulke and Sharpton 2018). In the example provided by the Peled et al. study, the results were generated over three continents, from subjects suffering from different malignant diseases and with a relatively balanced male-to-women ratio, which reinforces its transferability and generalizability. However, this approach could hide relevant population-specific patterns (Gupta et al. 2017).

Case 3: Machine-learning classification for hepatocellular carcinoma screening

Supervised and unsupervised machine-learning offers novel applications for metagenomics but also raises

challenges for their clinical validation. Supervised machine-learning algorithms could be trained on microbiota profiling datasets obtained from patients for which the clinical outcome is known. These algorithms could then theoretically be used on the microbiota profiles to diagnose a disease by categorizing patients as healthy or diseased, or to prognose the outcome of a disease (Zhou and Gallins 2019; Topçuoğlu et al. 2020; Marcos-Zambrano et al. 2021). This case scenario will focus on a recent study by Ren et al. who trained a random forest classifier to screen for hepatocellular carcinoma based on gut microbiota (Ren et al. 2019).

Clinical validation of supervised machine-learning classifiers based on microbiota profiling data is a complex task. Once trained, classifiers should be rigorously evaluated and tested (Chen et al. 2019; Wiens et al. 2019; Topçuoğlu et al. 2020). Whenever possible, factors supporting classification (e.g. microbial or clinical features) should be reviewed to identify potentially inadequate factors that could have been retained in training, such as implicit indication of the clinical outcome (i.e. "leaked labels") (Wiens et al. 2019) or taxa that will not be consistently found in microbiota profiles (e.g. contaminants). Then, the performance of the trained model should be tested on data forming a validation set distinct from the training set (Chen et al. 2019; Wiens et al. 2019; Topçuoğlu et al. 2020). In the example of the study by Ren et al. a random forest classifier was first trained (discovery phase) on a subset of patients and then tested on a different set of patients from the same healthcare institution to validate its performance (Ren et al. 2019).

Ideally, the clinical performance of the proposed machine-learning algorithm should be evaluated on large independent datasets (Topçuoğlu et al. 2020). In this evaluation, the performance of the classifier should be observed using relevant metrics, beyond the standard ROC curve (Wiens et al. 2019). In our example of microbiota-based screening for hepatocellular carcinoma, suboptimal specificity would be less problematic than imperfect sensitivity. Indeed, as for any screening test, false positives that can be refuted by confirmation tests are acceptable, while false negatives are more problematic. The population tested should be large and heterogeneous to ensure transferability of the model (Chen et al. 2019). However, models could also be adapted to some subpopulations of interest since microbial determinants could vary, for instance upon age group or geography (Wiens et al. 2019).

Machine-learning to diagnose diseases or prognose their outcome is a promising approach that will certainly find applications in clinical microbiota profiling.

However, many aspects remain to be explored and we recommend recent publications that address the use of machine learning in health-care in general (Chen et al. 2019; Wiens et al. 2019), in oncology (Cammarota et al. 2020) and in microbiota studies (Topçuoğlu et al. 2020).

4. Conclusion

This review proposes adaptations of existing QM concepts to the specificities of metagenomics-based microbiota profiling. The many aspects of QM discussed offer a framework for the implementation of microbiota profiling by clinical laboratories summarized in Figure 2. Some QM aspects are already well established (need for negative QC samples) and some other practices could be directly inspired from other clinical laboratory standards (IQA, EQA). Conversely, the definition of acceptance criteria or the validation of microbiota profiling assays raise specific challenges that require further improvements to reach maturity. To complete the translation of microbiota research into diagnostic analyses, the main actors in the field of clinical microbiota profiling need to confirm the usefulness of metagenomics, precise the potential clinical indications for testing and overcome the remaining challenges around microbiota profiling.

Besides anticipating the future need for clinical microbiota profiling, the QM scheme proposed in this review could also directly apply to research laboratories. Indeed, the inclusion of comprehensive methodological descriptions (SOPs) and controls (QC samples, acceptability criteria) in studies proposing new biomarkers will accelerate the uptake by routine clinical laboratories and help reaching similar clinical performance (Endrullat et al. 2016; Dirnagl et al. 2018), allowing to translate fundamental research findings into applicable and actionable diagnostic tools for patients.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

VS was supported by a Swiss National Science Foundation (SNSF) Grant (n° 10531 C-170280-F. Taroni, L. Falquet and G. Greub).

ORCID

Valentin Scherz  <http://orcid.org/0000-0002-9746-5423>
 Gilbert Greub  <http://orcid.org/0000-0001-9529-3317>
 Claire Bertelli  <http://orcid.org/0000-0003-0550-8981>

References

- Acosta N, Heirali A, Somayaji R, Surette MG, Workentine ML, Sibley CD, Rabin HR, Parkins MD. 2018. Sputum microbiota is predictive of long-term clinical outcomes in young adults with cystic fibrosis. *Thorax*. 73(11):1016–1025.
- Amos GCA, Logan A, Anwar S, Fritzsche M, Mate R, Bleazard T, Rijpkema S. 2020. Developing standards for the microbiome field. *Microbiome*. 8(1):13–98.
- Ananthkrishnan AN, Luo C, Yajnik V, Khalili H, Garber JJ, Stevens BW, Cleland T, Xavier RJ. 2017. Gut microbiome function predicts response to anti-integrin biologic therapy in inflammatory bowel diseases. *Cell Host Microbe*. 21(5):603–610.e3.
- Andrews S. 2010. FASTQC. A quality control tool for high throughput sequence data; [cited 2021 Aug 31]. Available from: <http://www.BioinformaticsBabrahamAcUk/Projects/Fastqc/>.
- Aziz N, Zhao Q, Bry L, Driscoll DK, Funke B, Gibson JS, Wayne WG, Hegde MR, Hoeltge GA, Leonard DGB, et al. 2015. College of American pathologists' laboratory standards for next-generation sequencing clinical tests. *Arch Pathol Lab Med*. 139(4):481–493.
- Bashiardes S, Godneva A, Elinav E, Segal E. 2018. Towards utilization of the human genome and microbiome for personalized nutrition. *Curr Opin Biotechnol*. 51:57–63.
- Bharti R, Grimm DG. 2019. Current challenges and best-practice protocols for microbiome analysis. *Brief Bioinform*. 22(1):178–193.
- Breitwieser FP, Lu J, Salzberg SL. 2019. A review of methods and databases for metagenomic classification and assembly. *Brief Bioinform*. 20(4):1125–1136.
- Brooks JP, Edwards DJ, Harwich MD, Rivera MC, Fettweis JM, Serrano MG, Reris RA, Sheth NU, Huang B, Girerd P, et al. 2015. The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiol*. 15(1):66.
- Budowle B, Connell ND, Bielecka-Oder A, Colwell RR, Corbett CR, Fletcher J, Forsman M, Kadavy DR, Markotic A, Morse SA, et al. 2014. Validation of high throughput sequencing and microbial forensics applications. *Investig Genet*. 5(1):9–18.
- Burd EM. 2010. Validation of laboratory-developed molecular assays for infectious diseases. *Clin Microbiol Rev*. 23(3):550–576.
- Cammarota G, Ianiro G, Ahern A, Carbone C, Temko A, Claesson MJ, Gasbarrini A, Tortora G. 2020. Gut microbiome, big data and machine learning to promote precision medicine for cancer. *Nat Rev Gastroenterol Hepatol*. 17(10):635–648.
- Carmody LA, Zhao J, Kalikin LM, LeBar W, Simon RH, Venkataraman A, Schmidt TM, Abdo Z, Schloss PD, LiPuma JJ. 2015. The daily dynamics of cystic fibrosis airway microbiota during clinical stability and at exacerbation. *Microbiome*. 3(1):12.
- Centers for Disease Control and Prevention (CDC) 2009. Good laboratory practices for molecular genetic testing for heritable diseases and conditions. *Mmwr*. 58(RR-6):1–37.
- Chaput N, Lepage P, Coutzac C, Soularue E, Le Roux K, Monot C, Boselli L, Routier E, Cassard L, Collins M, et al. 2017. Baseline gut microbiota predicts clinical response and colitis in metastatic melanoma patients treated with ipilimumab. *Ann Oncol*. 28(6):1368–1379.
- Charlson ES, Bittinger K, Haas AR, Fitzgerald AS, Frank I, Yadav A, Bushman FD, Collman RG. 2011. Topographical continuity of bacterial populations in the healthy human respiratory tract. *Am J Respir Crit Care Med*. 184(8):957–963.
- Chen PHC, Liu Y, Peng L. 2019. How to develop machine learning models for healthcare. *Nat Mater*. 18(5):410–414.
- Chiu CY, Miller SA. 2019. Clinical metagenomics. *Nat Rev Genet*. 20(6):341–355.
- Cuthbertson L, Walker AW, Oliver AE, Rogers GB, Rivett DW, Hampton TH, Ashare A, Elborn JS, De Soya A, Carroll MP, et al. 2020. Lung function and microbiota diversity in cystic fibrosis. *Microbiome*. 8(1):45–13.
- Danneskiold-Samsøe NB, Dias de Freitas Queiroz Barros H, Santos R, Bicas JL, Cazarin CBB, Madsen L, Kristiansen K, Pastore GM, Brix S, Maróstica Júnior MR. 2019. Interplay between food and gut microbiota in health and disease. *Food Res Int*. 115:23–31.
- David LA, Materna AC, Friedman J, Campos-Baptista MI, Blackburn MC, Perrotta A, Erdman SE, Alm EJ. 2014. Host lifestyle affects human microbiota on daily timescales. *Genome Biol*. 15(7):R89.
- Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. 2017. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *bioRxiv*. 1–14.
- Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. 2017. Nextflow enables reproducible computational workflows. *Nat Biotechnol*. 35(4):316–319.
- Dickerson JA, Cole B, Conta JH, Wellner M, Wallace SE, Jack RM, Rutledge J, Astion ML. 2014. Improving the value of costly genetic reference laboratory testing with active utilization management. *Arch Pathol Lab Med*. 138(1):110–113.
- Dirnagl U, Kurreck C, Castaños-Vélez E, Bernard R. 2018. Quality management for academic laboratories: burden or boon? *EMBO Rep*. 19(11):1–5.
- Dobbin KK, Cesano A, Alvarez J, Hawtin R, Janetzki S, Kirsch I, Masucci GV, Robbins PB, Selvan SR, Streicher HZ, et al. 2016. Validation of biomarkers to predict response to immunotherapy in cancer: Volume II – clinical validation and regulatory considerations. *J Immunother Cancer*. 4(1):77–14.
- Ducarmon QR, Hornung BVH, Geelen AR, Kuijper EJ, Zwittink RD. 2020. Toward standards in clinical microbiota studies: comparison of three DNA extraction methods and two bioinformatic pipelines. *mSystems*. 5(1):19.
- Durack J, Lynch SV. 2019. The gut microbiome: relationships with disease and opportunities for therapy. *J Exp Med*. 216(1):20–40.
- Dybboe Bjerre R, Hugerth LW, Boulund F, Seifert M, Duus Johansen J, Engstrand L. 2019. Effects of sampling strategy and DNA extraction on human skin microbiome investigations. *Sci Rep*. 9:17287.
- Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS. 2019. Contamination in low microbial biomass microbiome studies: issues and recommendations. *Trends Microbiol*. 27(2):105–117.

- Endrullat C, Glöckler J, Franke P, Frohme M. 2016. Standardization and quality management in next-generation sequencing. *Appl Transl Genom.* 10:2–9.
- Escobar-Zepeda A, Vera-Ponce de León A, Sanchez-Flores A. 2015. The road to metagenomics: from microbiology to DNA sequencing technologies and bioinformatics. *Front Genet.* 6:348.
- Fan Y, Pedersen O. 2021. Gut microbiota in human metabolic health and disease. *Nat Rev Microbiol.* 19(1):55–71.
- Gargis AS, Kalman L, Bick DP, da Silva C, Dimmock DP, Funke BH, Gowrisankar S, Hegde MR, Kulkarni S, Mason CE, et al. 2015. Good laboratory practice for clinical next-generation sequencing informatics pipelines. *Nat Biotechnol.* 33(7):689–693.
- Gargis AS, Kalman L, Lubin IM. 2016. Assuring the quality of next-generation sequencing in clinical microbiology and public health laboratories. *J Clin Microbiol.* 54(12):2857–2865.
- Gaulke CA, Sharpton TJ. 2018. The influence of ethnicity and geography on human gut microbiome composition. *Nat Med.* 24(10):1495–1496.
- Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. 2017. Microbiome datasets are compositional: and this is not optional. *Front Microbiol.* 8:2224.
- Glossary – BEST (Biomarkers, EndpointS, and other Tools) Resource – NCBI Bookshelf [Internet]; [cited 2020 Aug 3]. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK338448/#IX-V>.
- Goodrich JK, Di Rienzi SC, Poole AC, Koren O, Walters WA, Caporaso JG, Knight R, Ley RE. 2014. Conducting a microbiome study. *Cell.* 158(2):250–262.
- Gopalakrishnan V, Spencer CN, Nezi L, Reuben A, Andrews MC, Karpinets TV, Prieto PA, Vicente D, Hoffman K, Wei SC, et al. 2018. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science.* 359(6371):97–103.
- Grüning B, Chilton J, Köster J, Dale R, Soranzo N, van den Beek M, Goecks J, Backofen R, Nekrutenko A, Taylor J. 2018. Practical Computational Reproducibility in the Life Sciences. *Cell Syst.* 6(6):631–635.
- Gupta VK, Paul S, Dutta C. 2017. Geography, ethnicity or subsistence-specific variations in human microbiome composition and diversity. *Front Microbiol.* 8:1162.
- Harrell L, Wang Y, Antonopoulos D, Young V, Lichtenstein L, Huang Y, Hanauer S, Chang E. 2012. Standard colonic lavage alters the natural state of mucosal-associated microbiota in the human colon. Singh SR, editor. *PLoS One.* 7(2):e32545.
- Henny J, Vassault A, Boursier G, Vukasovic I, Mesko Brguljan P, Lohmander M, Ghita I, Andreu FAB, Kroupis C, Sprongl L, et al. 2016. EFLM Recommendation Recommendation for the review of biological reference intervals in medical laboratories. *Clin Chem Lab Med.* 54(12):1893–1900.
- Hillmann B, Al-Ghalith GA, Shields-Cutler RR, Zhu Q, Gohl DM, Beckman KB, Knight R, Knights D. 2018. Evaluating the information content of shallow shotgun metagenomics. *mSystems.* 3(6):69–87. Nov 13
- Hooks KB, O'Malley MA. 2017. Dysbiosis and its discontents. *MBio.* 8(5):e01492–17.
- Hornung BVH, Zwitink RD, Kuijper EJ. 2019. Issues and current standards of controls in microbiome research. *FEMS Microbiol Ecol.* 95(5):1–7.
- Hugerth LW, Andersson AF. 2017. Analysing microbial community composition through amplicon sequencing: from sampling to hypothesis testing. *Front Microbiol.* 8(1561):1561.
- Hussing C, Kampmann ML, Mogensen HS, Børsting C, Morling N. 2018. Quantification of massively parallel sequencing libraries – a comparative study of eight methods. *Sci Rep.* 8(1):1110–1119.
- Illumina Inc. 2016. Library Quantification and Quality Control Quick Reference Guide [Internet]; [cited 2019 Oct 3]. p. 2016–7. Available from: <https://support.illumina.com/bulletins/2016/05/library-quantification-and-quality-control-quick-reference-guide.html>.
- Illumina. 2020. 16S Metagenomic sequencing library preparation preparing 16S Ribosomal RNA Gene amplicons for the illumina MiSeq System.
- Jenkins SV, Vang KB, Gies A, Griffin RJ, Jun S-R, Nookaew I, Dings RPM. 2018. Sample storage conditions induce post-collection biases in microbiome profiles. *BMC Microbiol.* 18(1):227–229.
- Jennings L, Van Deerlin VM, Gulley ML. 2009. Recommended principles and practices for validating clinical molecular pathology tests. *Arch Pathol Lab Med.* 133(5):743–755.
- Jennings LJ, Arcila ME, Corless C, Kamel-Reid S, Lubin IM, Pfeifer J, Temple-Smolkin RL, Voelkerding KV, Nikiforova MN. 2017. Guidelines for validation of next-generation sequencing-based oncology panels: a joint consensus recommendation of the association for molecular pathology and college of american pathologists. *J Mol Diagn.* 19(3):341–365.
- Jian C, Luukkonen P, Yki-Järvinen H, Salonen A, Korpela K. 2020. Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling. *PLoS One.* 15(1):e0227285–10.
- Jones RB, Zhu X, Moan E, Murff HJ, Ness RM, Seidner DL, Sun S, Yu C, Dai Q, Fodor AA, et al. 2018. Inter-niche and inter-individual variation in gut microbial community assessment using stool, rectal swab, and mucosal samples. *Sci Rep.* 8(1):4112–4139.
- Kalman LV, Lubin IM, Barker S, Du Sart D, Elles R, Grody WW, Pazzagli M, Richards S, Schrijver I, Zehnbauser B, et al. 2013. Current landscape and new paradigms of proficiency testing and external quality assessment for molecular genetics HHS public access. *Arch Pathol Lab Med.* 137(7):983–988.
- Karstens L, Asquith M, Davin S, Fair D, Gregory WT, Wolfe AJ, Braun J, McWeeney S. 2019. Controlling for Contaminants in Low-Biomass 16S rRNA Gene Sequencing Experiments. *mSystems.* 4(4):290–309.
- Kashyap PC, Chia N, Nelson H, Segal E, Elinav E. 2017. Microbiome at the frontier of personalized medicine. *Mayo Clin Proc.* 92(12):1855–1864.
- Kiely CJ, Pavli P, O'Brien CL. 2018. The role of inflammation in temporal shifts in the inflammatory bowel disease mucosal microbiome. *Gut Microbes.* 9(6):477–485.
- Kim D, Hofstaedter CE, Zhao C, Mattei L, Tanes C, Clarke E, Lauder A, Sherrill-Mix S, Chehoud C, Kelsen J, et al. 2017. Optimizing methods and dodging pitfalls in microbiome research. *Microbiome.* 5(1):52.
- Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, Gonzalez A, Kosciolek T, McCall L-I, McDonald

- D, et al. 2018. Best practices for analysing microbiomes. *Nat Rev Microbiol.* 16(7):410–422.
- Köster J, Rahmann S. 2012. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics.* 28(19):2520–2522.
- Lamarche D, Johnstone J, Zytaruk N, Clarke F, Hand L, Loukov D, Szamosi JC, Rossi L, Schenck LP, Verschoor CP, et al. Microbial dysbiosis and mortality during mechanical ventilation: a prospective observational study. *Respir Res.* 2018;19(1):245.
- Lazarevic V, Gaia N, Girard M, Schrenzel J. 2016. Decontamination of 16S rRNA gene amplicon sequence datasets based on bacterial load assessment by qPCR. *BMC Microbiology.* 16:Article number: 73.
- Leigh Greathouse K, Sinha R, Vogtmann E. 2019. DNA extraction for human microbiome studies: the issue of standardization. *Genome Biol.* 20(1):212.
- Ley RE, Bäckhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI. 2005. Obesity alters gut microbial ecology. *Proc Natl Acad Sci USA.* 102(31):11070–11075.
- Liang X, FitzGerald GA. 2017. Timing the microbes: the circadian rhythm of the gut microbiome. *J Biol Rhythms.* 32(6): 505–515.
- Lloyd-Price J, Abu-Ali G, Huttenhower C. 2016. The healthy human microbiome. *Genome Med.* 8(1):51.
- Lo C, Marculescu R. 2019. MetaNN: accurate classification of host phenotypes from metagenomic data using neural networks. *BMC Bioinformatics.* 20:314.
- Lundberg GD. 1998. The need for an outcomes research agenda for clinical laboratory testing. *JAMA.* 280(6): 565–566.
- Luz Calle M. 2019. Statistical analysis of metagenomics data. *Genomics Inform.* 17(1):e6.
- Maier L, Pruteanu M, Kuhn M, Zeller G, Telzerow A, Anderson EE, Brochado AR, Fernandez KC, Dose H, Mori H, et al. 2018. Extensive impact of non-antibiotic drugs on human gut bacteria Europe PMC Funders Group. *Nature.* 555(7698):623–628.
- Malla MA, Dubey A, Kumar A, Yadav S, Hashem A, Allah EFA. 2019. Exploring the human microbiome: the potential future role of next-generation sequencing in disease diagnosis and treatment. *Front Immunol.* 10:1–23.
- Mandrekar SJ, Sargent DJ. 2009. Clinical trial designs for predictive biomarker validation: theoretical considerations and practical challenges. *J Clin Oncol.* 27(24):4027–4034.
- Marchesi JR, Adams DH, Fava F, Hermes GDA, Hirschfield GM, Hold G, Quraishi MN, Kinross J, Smidt H, Tuohy KM, et al. 2016. The gut microbiota and host health: a new clinical frontier. *Gut.* 65(2):330–339.
- Marchesi JR, Ravel J. 2015. The vocabulary of microbiome research: a proposal. *Microbiome.* 3(1):31.
- Marcos-Zambrano LJ, Karadzovic-Hadziabdic K, Loncar Turukalo T, Przymus P, Trajkovic V, Aasmets O. 2021. Applications of machine learning in human microbiome studies: a review on feature selection, biomarker identification, disease prediction and treatment. *Front Microbiol.* 12: 634511.
- Martí JM. 2019. Recentrifuge: Robust comparative analysis and contamination removal for metagenomics. *PLoS Comput Biol.* 15(4):e1006967–24.
- Matson V, Fessler J, Bao R, Chongsuwat T, Zha Y, Alegre M-L, Luke JJ, Gajewski TF. 2018. The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. *Science.* 359(6371):104–108.
- McDonald LC, Gerding DN, Johnson S, Bakken JS, Carroll KC, Coffin SE, Dubberke ER, Garey KW, Gould CV, Kelly C, et al. 2018. Clinical practice guidelines for clostridium difficile infection in adults and children: 2017 Update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA). *Clin Infect Dis.* 66(7):e1–e48.
- McKnight DT, Huerlimann R, Bower DS, Schwarzkopf L, Alford RA, Zenger KR. 2019. microDecon: a highly accurate read-subtraction tool for the post-sequencing removal of contamination in metabarcoding studies. *Environ DNA.* 1(1):14–25.
- Miller S, Naccache SN, Samayoa E, Messacar K, Arevalo S, Federman S, Stryke D, Pham E, Fung B, Bolosky WJ, et al. 2019. Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. *Genome Res.* 29(5):831–842.
- Minich JJ, Sanders JG, Amir A, Humphrey G, Gilbert JA, Knight R. 2019. Quantifying and understanding well-to-well contamination in microbiome research. *mSystems.* 4(4):396.
- Motro Y, Moran-Gilad J. 2018. Microbial metagenomics mock scenario-based sample simulation (M3S3). *Clin Microbiol Infect.* 24(3):308.e1–308.e4.
- Multinu F, Harrington SC, Chen J, Jeraldo PR, Johnson S, Chia N, Walther-Antonio MR. 2018. Systematic bias introduced by genomic DNA template dilution in 16S rRNA gene-targeted microbiota profiling in human stool homogenates. *mSphere.* 3(2):17.
- Nayfach S, Pollard KS. 2016. Toward accurate and quantitative comparative metagenomics. *Cell.* 166(5):1103–1116.
- Nishino M, Ramaiya NH, Hatabu H, Hodi FS. 2017. Monitoring immune-checkpoint blockade: response evaluation and biomarker development. *Nat Rev Clin Oncol.* 14(11): 655–668.
- Peled JU, Gomes ALC, Devlin SM, Littmann ER, Taur Y, Sung AD, Weber D, Hashimoto D, Slingerland AE, Slingerland JB, et al. 2020. Cast and MRM van den B. Microbiota as Mortality Predictor in Allogeneic HCT. *N Engl J Med.* 382(9):822–834.
- Perez-Riverol Y, Gatto L, Wang R, Sachsenberg T, Uszkoreit J, Leprevost F, d V, Fufezan C, Ternent T, Eglen SJ, Katz DS, et al. 2016. Ten simple rules for taking advantage of Git and GitHub. *PLoS Comput Biol.* 12(7):e1004947–11.
- Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, Bonazzi V, McEwen JE, Wetterstrand KA, Deal C, et al. 2009. The NIH human microbiome project. *Genome Res.* 19(12):2317–2323.
- Piccolo SR, Frampton MB. 2016. Tools and techniques for computational reproducibility. *Gigascience.* 5(1):30.
- Pollock J, Glendinning L, Wisedchanwet T, Watson M. 2018. The madness of microbiome: attempting to find consensus “best practice” for 16S microbiome studies. *Appl Environ Microbiol.* 84:AEM.02627-17.
- Poussin C, Sierro N, Boué S, Battey J, Scotti E, Belcastro V, Peitsch MC, Ivanov NV, Hoeng J. 2018. Interrogating the microbiome: experimental and computational considerations in support of study reproducibility. *Drug Discov Today.* 23(9):1644–1657.

- Prodan A, Tremaroli V, Brolin H, Zwinderman AH, Nieuwdorp M, Levin E. 2020. Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. *PLoS One*. 15(1): e0227434–19.
- Quigley EMM. 2017. Gut microbiome as a clinical tool in gastrointestinal disease management: are we there yet? *Nat Rev Gastroenterol Hepatol*. 14(5):315–320.
- Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, Friez MJ, Funke BH, Hegde MR, Lyon E, 2013. ACMG clinical laboratory standards for next-generation sequencing. *Genet Med*. 15(9):733–747.
- Ren Z, Li A, Jiang J, Zhou L, Yu Z, Lu H, Xie H, Chen X, Shao L, Zhang R, et al. 2019. Gut microbiome analysis as a tool towards targeted non-invasive biomarkers for early hepatocellular carcinoma. *Gut*. 68(6):1014–1023.
- Riley JD, Procop GW, Kottke-Marchant K, Wyllie R, Lacbawan FL. 2015. Improving molecular genetic test utilization through order restriction, test review, and guidance. *J Mol Diagn*. 17(3):225–229.
- Riquelme E, Zhang Y, Zhang L, Montiel M, Zoltan M, Dong W, Quesada P, Sahin I, Chandra V, San Lucas A, et al. 2019. Tumor microbiome diversity and composition influence pancreatic cancer outcomes. *Cell*. 178(4):795–806.e12.
- Rodriguez-R LM, Konstantinidis KT. 2014. Estimating coverage in metagenomic data sets and why it matters. *Isme J*. 8(11):2349–2351.
- Rogers GB, Wesselingh S. 2016. Precision respiratory medicine and the microbiome. *Lancet Respir Med*. 4(1):73–82.
- Routy B, Le Chatelier E, Derosa L, Duong CPM, Alou MT, Daillère R, Fluckiger A, Messaoudene M, Rauber C, Roberti MP, et al. 2018. Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Science*. 359(6371):91–97.
- Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW. 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol*. 12:87.
- Scherz V, Durussel C, Greub G. 2017. Internal quality assurance in diagnostic microbiology: a simple approach for insightful data. *PLoS One*. 12(11):e0187263.
- Schlager R, Chiu CY, Miller S, Procop GW, Weinstock G. 2017. Validation of metagenomic next-generation sequencing tests for universal pathogen detection. *Arch Pathol Lab Med*. 141(6):776–786.
- Schloss PD. 2018. Identifying and overcoming threats to reproducibility, replicability, robustness, and generalizability in microbiome research. *ASM J*. 9(3):18.
- Schmidt TSB, Raes J, Bork P. 2018. The human gut microbiome: from association to modulation. *Cell*. Cell Press. 172(6):1198–1215.
- Schrijver I, Aziz N, Jennings LJ, Richards CS, Voelkerding KV, Weck KE. 2014. Methods-based proficiency testing in molecular genetic pathology. *J Mol Diagn*. 16(3):283–287.
- Sekirov I, Russell SL, Caetano M Antunes L, Finlay BB. 2010. Gut microbiota in health and disease. *Physiol Rev*. 90(3): 859–904.
- Shade A. 2017. Diversity is the question, not the answer. *Isme J*. 11(1):1–6.
- Simbolo M, Gottardi M, Corbo V, Fassan M, Mafficini A, Malpeli G, Lawlor RT, Scarpa A. 2013. DNA qualification workflow for next generation sequencing of histopathological samples. Hoheisel JD, editor. *PLoS One*. 8(6):e62692.
- Sims TT, El Alam MB, Karpinets TV, Dorta-Estremera S, Hegde VL, Nookala S, Yoshida-Court K, Wu X, Biegert GWG, Delgado Medrano AY, et al. 2021. Gut microbiome diversity is an independent predictor of survival in cervical cancer patients receiving chemoradiation. *Commun Biol*. 4(1): 210–237.
- Sinha R, Abu-Ali G, Vogtmann E, Fodor AA, Ren B, Amir A, Schwager E, Crabtree J, Ma S, Abnet CC, et al. 2017. Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. *Nat Biotechnol*. 35(11):1077–1086.
- Smillie CS, Sauk J, Gevers D, Friedman J, Sung J, Youngster I, Hohmann EL, Staley C, Khoruts A, Sadowsky MJ, et al. 2018. Strain tracking reveals the determinants of bacterial engraftment in the human gut following fecal microbiota transplantation. *Cell Host Microbe*. 23(2):229–240.e5.
- Stahl DA, Lane DJ, Olsen GJ, Pace NR. 1984. Analysis of hydrothermal vent-associated symbionts by ribosomal RNA sequences. *Science*. 224(4647):409–411.
- Takayasu L, Suda W, Takanashi K, Iioka E, Kurokawa R, Shindo C, Hattori Y, Yamashita N, Nishijima S, Oshima K, et al. 2017. Circadian oscillations of microbial and functional composition in the human salivary microbiome. *DNA Res*. 24(3):261–270.
- Topçuoğlu BD, Lesniak NA, Ruffin MT, Wiens J, Schloss PD. 2020. A framework for effective application of machine learning to microbiome-based classification problems. *MBio*. 11(3):20.
- Wagner BD, Grunwald GK, Zerbe GO, Mikulich-Gilbertson SK, Robertson CE, Zemanick ET, Harris JK. 2018. On the use of diversity measures in longitudinal sequencing studies of microbial communities. *Front Microbiol*. 9:1037.
- Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, Jung K, Heller K, Kale D, Saeed M, et al. 2019. Author correction: do no harm: a roadmap for responsible machine learning for health care. *Nat Med*. 25(10):1627.
- Wilkinson EM, Ilhan ZE, Herbst-Kralovetz MM. 2018. Microbiota-drug interactions: impact on metabolism and efficacy of therapeutics. *Maturitas*. 112:53–63.
- Woodworth MH, Neish EM, Miller NS, Dhert T, Burd EM, Carpentieri C, Sitchenko KL, Kraft CS. 2017. Laboratory testing of donors and stool samples for fecal microbiota transplantation for recurrent *Clostridium difficile* infection. *J Clin Microbiol*. 55(4):1002–1010. Apr
- Yamashiro Y. 2017. Gut Microbiota in health and disease. *Ann Nutr Metab*. 71(3–4):242–246.
- Yeh Y-C, Needham DM, Sieradzki ET, Fuhrman JA. 2018. Taxon disappearance from microbiome analysis reinforces the value of mock communities as a standard in every sequencing run. *mSystems*. 3(3):18.
- Zhou Y-HH, Gallins P. 2019. A review and tutorial of machine learning methods for microbiome host trait prediction. *Front Genet*. 10(JUN):514–579.
- Ziegler A, Koch A, Krockenberger K, Grosshennig A. 2012. Personalized medicine using DNA biomarkers: a review. *Hum Genet*. 131(10):1627–1638.
- Zmora N, Zeevi D, Korem T, Segal E, Elinav E. 2016. Taking it personally: personalized utilization of the human microbiome in health and disease. *Cell Host Microbe*. 19(1):12–20.