Journal of Crohn's and Colitis, 2016, 735–746
doi:10.1093/ecco-jcc/jjw024
Advance Access publication January 22, 2016
Review Article

Review Article

Meta-omics in Inflammatory Bowel Disease Research: Applications, Challenges, and Guidelines

Mireia Valles-Colomer,[a,b,*] Youssef Darzi,[a,b,c,*] Sara Vieira-Silva,[a,b] Gwen Falony,[a,b] Jeroen Raes,[a,b,†] Marie Joossens[a,b,c,†]

[a]KU Leuven, Department of Microbiology and Immunology, Rega Institute, Leuven, Belgium [b]VIB, Center for the Biology of Disease, Leuven, Belgium [c]Microbiology Unit, Faculty of Sciences and Bioengineering Sciences, Vrije Universiteit Brussel, Brussels, Belgium

Corresponding author: Jeroen Raes, PhD, KU Leuven, Department of Microbiology and Immunology, Rega Institute. VIB, Center for the Biology of Disease. Campus Gasthuisberg: Herestraat 49, 3000 Leuven, Belgium. Tel: +32 16 37 22 22; Email: jeroen.raes@vib-kuleuven.be

*These authors contributed equally to this work.
†Raes J. and Joossens M. are co-last authors.

Abstract

Meta-omics [metagenomics, metatranscriptomics, and metaproteomics] are rapidly expanding our knowledge of the gut microbiota in health and disease. These technologies are increasingly used in inflammatory bowel disease [IBD] research. Yet, meta-omics data analysis, interpretation, and among-study comparison remain challenging. In this review we discuss the role these techniques are playing in IBD research, highlighting their strengths and limitations. We give guidelines on proper sample collection and preparation methods, and on performing the analyses and interpreting the results, reporting available user-friendly tools and pipelines.

Key Words: Metagenomics; metatranscriptomics; metaproteomics; 16S; meta-omics; microbiota

1. Introduction

Even in the initial description of Crohn's disease [CD], the importance of gut-associated bacteria in the aetiology and pathophysiology of inflammatory bowel diseases [IBD] was already suggested.[1] Intestinal dysbiosis is being studied to disentangle the various contributors in CD and ulcerative colitis [UC]. Profound analyses of the gut microbial community [microbiota] and the mechanistic insights these provide on the interaction between the human gut and the hosted microbiota are essential to develop new strategies to tackle the rise in IBD prevalence.[2]

Microbiology has been revolutionised by the development of high-throughput technologies that permit study of the microbial communities as a whole. Known as meta-omics, they aim at the direct analysis of genes, transcripts, or proteins recovered from environmental samples, by skipping cultivation and the bias it introduces altogether. Metagenomics starts by sequencing the DNA extracted from a microbial community, and assessing what microorganisms are present in a sample as well as their functional potential. By sequencing the community RNA, metatranscriptomics allows the monitoring of the microbiota's current gene expression; whereas metaproteomics, based on protein spectrum profiles, provides information about the proteins that are synthesised. 16S amplicon sequencing is based on the specific amplification of a hypervariable region of the ribosomal RNA gene that is universally present in Bacteria and Archaea. Although it does not bring the functional insights that shotgun metagenomics provides and has lower phylogenetic resolution, it is an order of magnitude more cost-effective for getting first insights into the phylogenetic composition of a sample. As the different meta-omics characterise complementary aspects of microbial communities [Figure 1], combining multiple meta-omic techniques holds great promise for understanding the role of the gut microbial community in

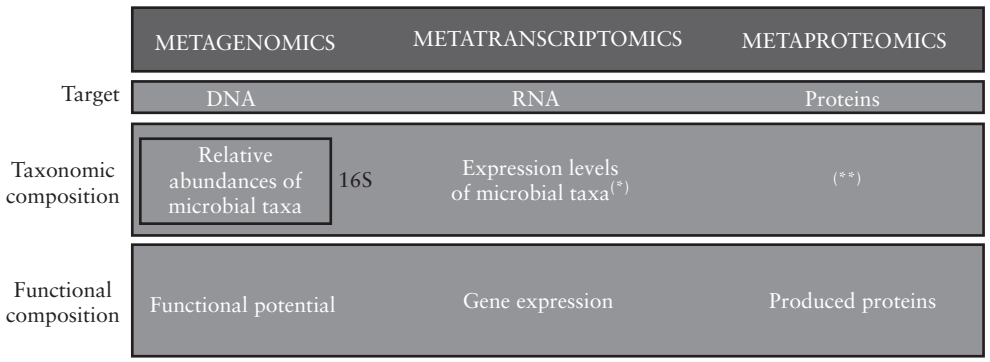

Figure 1. Information provided by the different meta-omic technologies, regarding taxonomic and functional composition, when analysing a microbiome. *Metatranscriptomics and **metaproteomics are often coupled to 16S rRNA gene sequencing or metagenomics to provide the sample's full taxonomic composition, independently of cellular activity levels.

IBD. Today, meta-omic technologies are becoming accessible to most research laboratories and some hospitals around the world, with the associated standard bioinformatics analysis provided as a service.

Clinical practice—from prognosis or diagnosis to treatment—could benefit from a closer integration with meta-omics. The purpose of this review is to describe how meta-omics can be applied to provide better insight into the microbial contributions in IBD. Starting with a description of adequate sample collection methods, we discuss the potential and specificity of different meta-omic analyses, accompanied with guidelines on how to interpret the results, and a selection of user-friendly tools that are available to perform data analysis beyond the standard analysis delivered by sequencing facilities.

2. Sample Pre-Processing

2.1. Sample collection and storage

Bacteria make up to 60% of the dry mass in stool samples, with one gram of faeces containing 10^{11}–10^{12} bacteria.[3] The microbial composition of stool samples is generally considered representative of the gut community. Hence, meta-omic analyses of faecal material have the potential to reflect composition and metabolism of the colon microbiota. However, meta-omic readouts of faeces are sensitive to inappropriate sampling logistics that allow bacterial growth and cell lysis during transport or storage, affecting both microbial composition and diversity. Therefore, faecal sample collection and storage protocols are of key importance to the success of subsequent analyses.

Several studies have investigated the effect of storage temperature and/or the use of buffers on microbiota composition.[4,5] In general, such studies conclude that inter-individual variability remains observable in faecal material despite preservation-induced biases. However, many disease-associated microbial signatures are more subtle than microbiota variation among different individuals. As such disease-specific signatures could get distorted, immediate storage of faecal samples at -20°C or -80°C for respective short- [weeks] or longer-term conservation is recommended.[4] Freeze-thaw cycles have a considerable impact on microbial cell integrity, resulting in increased proportions of degraded DNA, altered relative abundances of microbial taxa, and a decreased RNA Integrity Number.[5] In order to avoid multiple freeze-thaw cycles, aliquoting fresh stool samples before initial freezing to facilitate subsequent analyses is recommended.

In IBD research, intestinal biopsy samples are sometimes collected during colonoscopy. Such samples contain ~ 100 times less microbial cells than stool samples,[6] and thus require even more strict protocols for optimal preservation and correct subsequent characterisation of the gut microbial community. For such samples, immediate freezing in liquid nitrogen [snap-freezing] followed by storage at -80°C is recommended.[7] Alternatively, intestinal biopsies can be stored in aqueous solution at -20°C for a few weeks, or at 4°C if processed the same day.[8] In any case, processing biopsy samples as soon as possible is advised in order to minimise lysis of the reduced microbial cell fraction.

2.2. Sample preparation: the extraction protocol

Several protocols have been established for extracting microbial DNA, RNA, and proteins from complex samples. Nevertheless, it remains challenging to find a protocol to extract cell content without taxon-specific biases. For example, whereas harsh lysis methods are required to disrupt Gram-positive cells, only subtle swift extraction processes allow preservation of DNA or RNA from taxa that are more prone to cell lysis. Although any extraction method will create a bias towards some taxa, a balanced extraction protocol is essential for meta-omic analyses that ultimately provide relative abundances of features in a sample. As a case in point, the protocols used by the first two major human microbiome sequencing projects—MetaHIT [Metagenomics of the Human Intestinal Tract][9] and HMP [Human Microbiome Project][10]—had an important difference in extraction efficiency of different bacterial phyla. As a result, the relative abundance of the Bacteroidetes phylum was significantly higher in samples extracted with the HMP protocol, translated into a distinct Firmicutes:Bacteroidetes ratio distribution across MetaHIT and HMP subjects.[11] To optimise data comparability in future microbiome studies, sample collection standard operating procedures [SOPs] are being developed [http://www.microbiome-standards.org/]. Overall, an extraction protocol including mechanical disruption of cells through bead-beating, essential to extract DNA from Gram-positive cells with acceptable efficiency,[12] can be recommended for the analysis of faecal and biopsy samples.

3. Metagenomics

3.1. Pre-metagenomic gut microbiota studies

Initial studies on the role of the gut microbiota in IBD depended on culturing methods only, introducing a clear bias given the fact that the vast majority of the gut bacteria cannot [easily] be grown *in vitro*. Notwithstanding their shortcomings, culture-dependent studies led to breakthroughs in IBD research such as the discovery of the

role of adherent-invasive *Escherichia coli* [AIEC] in CD.[13] However, the emergence of culture-independent approaches to microbial profiling was crucial to in-depth assessment of microbiota involvement in IBD. Such methods, mostly relying on the prokaryotic universal marker gene [16S rRNA], made the characterisation of microbial diversity in community samples more tractable. In IBD research, this led to the identification of the predominant gut microbiota alterations associated to the pathology.

Pre-metagenomic, culture-independent approaches in IBD research employed finger-printing techniques including temperature/denaturing gradient gel electrophoresis [T/DGGE][14,15] and terminal restriction fragment-length polymorphism [T-RFLP],[16] direct visualisation using fluorescent in situ hybridization [FISH],[17] phylogenetic arrays like HITChip,[18] and high-throughput cloning.[19] The role of specific bacteria in IBD pathology was subsequently investigated by a combination of culture-independent techniques and targeted microbiology. For example, *Faecalibacterium prausnitzii* was found to be associated with ileal CD recurrence and to exhibit anti-inflammatory effects.[17]

3.2. Marker gene [16S rRNA] sequencing

3.2.1. 16S rRNA gene sequencing and its application to IBD research

In the mid 1980s, Lane and colleagues successfully sequenced universal marker genes [typically the hypervariable regions of the 16S rRNA gene] from microbial community samples,[20] leading to new insights into the phylogenetic diversity of bacterial ecosystems. 16S rRNA sequencing provides information about the taxonomic composition of a sample, i.e. the microorganisms that are present and their relative abundances. As only the 16S rRNA gene is sequenced, it fails to inform about the functions that can be performed by the community [functional composition], in contrast to shotgun metagenomics. In addition, strain-level analysis is more informative using the latter [see further]. Still, 16S rRNA sequencing is a cost-effective technology for taxonomic profiling and it is widely used, especially in large cohorts, which can also serve to select a subset of samples for meta-omic analysis.

Using 16S gene rRNA sequencing, Bacteroidetes and Firmicutes were identified as the most abundant phyla in the gut microbiota of healthy subjects [based on analyses of both mucosal and faecal samples].[21] Technological advances and the associated reduction in sequencing costs have made microbiome amplicon profiling widely available, permitting fast and extensive phylogenetic analysis of clinical or environmental samples. A growing number of projects in IBD research depend on 16S rRNA sequencing, and the technique has revealed alterations in the microbiota of IBD patients compared with healthy controls. A reduced bacterial diversity in IBD patients has been reported by several studies[22,23] [see Table 1 for an overview of publications in IBD based on 16S rRNA gene sequencing and/or meta-omic technologies up to this date]. However, consistency between studies is lacking with regard to the alteration of relative abundances of microbial taxa. In gastrointestinal biopsies from IBD patients, Frank *et al.*[24] found lower relative abundances of both Bacteroidetes and Firmicutes than in control subjects, whereas Walker *et al.*[19] observed decreased Bacteroidetes but increased Firmicutes abundances. Moreover, Morgan *et al.*[25] reported lower Firmicutes in biopsies but higher in faecal samples of IBD patients compared with healthy controls. This illustrates again the importance of the nature of the samples as well as the sample preparation protocol when comparing the outcome of different studies: increased Firmicutes abundances in IBD patients were consistently reported when extraction protocols included a bead-beating step, but the opposite was observed when mechanical disruption was lacking.

3.2.2. 16S rRNA gene data processing

Nowadays, several well-established tools are available to process 16S rRNA gene sequencing data [see Table 2 for an overview of the tools for IBD meta-omic data processing, analysis, and visualisation]. The more user-friendly among them include QIIME,[26] Mothur,[27] and LotuS.[28] These tools also provide comprehensive instructions to perform the necessary sequential steps to correctly process amplicon barcode data. Basic procedures include de-multiplexing [assigning sequence reads to each sample that is pooled in sequencing reactions], quality control, and chimera filtering (removing artefactual DNA sequences containing fragments from several organisms, produced during polymerase chain reaction [PCR] amplification), clustering of 16S rRNA gene reads into operational taxonomic units [OTUs], and assigning OTU taxonomy through mapping onto 16S rRNA gene reference databases [eg Greengenes[29]]. Such data preprocessing is usually covered by the sequencing facility, in addition to a few basic statistics. To explore microbiota composition tables, a few user-friendly R packages are available. Alpha diversity [within-sample diversity, including richness—number of taxa—and evennes—sdistribution of taxa], beta diversity [dissimilarity among samples], statistical analysis of between-group differences, ordination analysis, and visualisation of results can all be performed using the R packages vegan[30] and phyloseq.[31]

For ordination of 16S rRNA gene sequencing data, the compositional dissimilarity between samples is most commonly calculated using Bray-Curtis dissimilarity or Unifrac distance. The former provides a measure of the differences in abundance of taxa between two samples, being bound between 0 and 1, where 0 means the two samples have exactly the same composition. The Unifrac distance was created to incorporate the notion of phylogenetic distance into the samples distance metric. That is to give a higher weight to compositional differences between distantly related taxa, for example if one sample's composition is dominated by Ruminococcus [Firmicutes phylum] while the other is Bacteroides-dominated [Bacteroidetes phylum]. These distances are then often used to graphically represent sample compositional dissimilarity, most often with a principal coordinates analysis [PCoA] or its non-metric alternative, non-metric dimensional scaling [NMDS]. The linear assumption underlying PCA [principal component analysis] ordination—the usual default method in general statistical programs—makes it unsuitable for most ecological data.[32] In addition, to show the effect of specific factors on community composition, methods such as RDA [redundancy analysis], CAP [canonical analysis of principal coordinates], and others can be used. Of particular interest, variation partitioning with an RDA allows estimation of the portion of the compositional variation that is explained by several factors (eg gender, age, body mass index [BMI], IBD) and allows singling out of the impact of one variable of interest from other factors influencing microbiota composition. For example, in a study including IBD patients and healthy subjects, this method could be used to calculate the percentage of variation in microbiota variation that is explained by the disease, by gender, and by age. Also, tools such as MaAsLin can be used to find associations between certain metadata parameters [disease status, diet, lifestyle] and microbiota composition, while deconfounding the effect of all other metadata parameters. Network analysis can be performed to detect co-occurrence and mutual exclusion

Table 1. Overview of the outcome of the main studies on the gut microbiota in IBD[a] based on 16S rRNA gene sequencing and/or meta-omic technologies

Technique	Material	Finding	Reference
16S sequencing, FISH	Stool samples	Reduced complexity of Firmicutes in CD patients	Manichanh *et al.* [2006][94]
16S sequencing, qPCR	GI biopsies	Decreased Bacteroidetes and Firmicutes abundances in CD and UC patients	Frank *et al.* [2007][24]
16S sequencing, histopathology, flow cytometry	Colon biopsies	Altered microbial functions in UC patients [increased lipid and amino acid metabolism], but not in CD patients	Davenport *et al.* [2009][95]
16S sequencing	Colon biopsies	Lower transcription levels in UC and CD, Firmicutes, and Actinobacteria inactive in CD patients	Rehman et al. [2010][75]
16S sequencing, qPCR	Colon biopsies	Higher Firmicutes but lower Bacteroidetes abundance in IBD patients	Walker *et al.* [2011][19]
16S sequencing, microarrays for host RNA	Colon biopsies	Lower diversity and increased Actinobacteria and Proteobacteria in UC than their healthy twins. Interaction between the transcription profile of the mucosa and the microbiota, which is lost in UC	Lepage *et al.* [2011][96]
16S sequencing, metagenomics	Stool samples, GI biopsies	Higher Firmicutes in stool samples but lower in biopsies of IBD patients; higher perturbations in microbial functions than in microbiota composition	Morgan *et al.* [2012][25]
16S sequencing	Ileum and colon biopsies	Lower Bacteroidetes and Firmicutes and higher Proteobacteria abundances in CD patients	Ricanek *et al.* [2012][97]
16S sequencing, qPCR	Ileum, colon and cecum biopsies	Lower microbial diversity in IBD patients, decreasing with increased inflammation	Zitomersky *et al.* [2013][22]
16S sequencing, immunoassays	Saliva	Higher Bacteroidetes and lower Proteobacteria abundance in IBD patients	Said *et al.* [2014][98]
16S sequencing	Stool samples	Increased Escherichia/Shigella and decreased Fecalibacterium in newly diagnosed CD patients	Thorkildsen *et al.* [2013][99]
16S sequencing	Stool samples	No changes in microbial composition or diversity between IBD patients in quiescent disease or remission state	Wills *et al.* [2014][100]
16S sequencing, metagenomics	Stool samples, GI biopsies	Increased Fusobacteriaceae and inflammation-related pathways, decreased Bacteroidales, Clostridiales abundances in CD	Gevers *et al.* [2014][46]
16S sequencing, microarray	GI biopsies	Lower microbial diversity in CD patients, increasing after surgery. Higher saccharolytic bacteria in CD patients in remission; higher proteolytic and lactic acid-producing bacteria in recurrent disease patients	De Cruz *et al.* [2014][23]
16S sequencing, RNA-Seq	Colon biopsies	Bacterial community composition differing in UC and CD patients vs controls according to geographical origin	Rehman *et al.* [2015][76]
Metagenomics	Stool samples	Gene and network-level topological differences in IBD patients, identification of biomarkers for IBD	Greenblum *et al.* [2012][44]
Metagenomics, metaproteomics	Stool samples	Identification of metabolic pathways that differentiate CD patients from controls and of possible targets	Erickson *et al.* [2012][45]
Metaproteomics, qPCR	GI biopsies	Identification of potential bacterial and protein biomarkers for CD and UC patients	Presley *et al.* [2012][82]
Metaproteomics, 16S sequencing	Stool samples	Over-representation of proteins derived from Bacteroides species, under-represented proteins from Firmicutes and Prevotella in CD patients	Juste *et al.* [2014][81]

IBD, inflammatory bowel disease; CD, Crohn's disease; UC, ulcerative colitis; GI, gastro-intestinal.

[a]Based on publications available on PubMed–NCBI [http://www.ncbi.nlm.nih.gov/pubmed] on 31 July, 2015, referring to inflammatory bowel disease and meta-omics technologies.

relationships between microbial taxa[33] and specific tools have been developed to this end [eg CoNet[34]]. All analyses mentioned above are most often performed at genus or OTU level. Tools have recently emerged to extrapolate lower-level [towards strain-level] variation from 16S rRNA sequencing data to reveal potential relevant taxonomic diversity within OTUs [Oligotyping[35]]. Additionally, tools have been developed to predict functional profiles from 16S rRNA data [eg PICRUSt[36]]; however, given the substantial functional variation across microbial species and strains, such metagenome predictions should be interpreted with great caution.

3.2.3. Pitfalls and limitations of 16S rRNA gene sequencing

Primer selection is crucial to successful microbiota characterisation through 16S rRNA gene sequencing. Commonly used primers

Table 2. Tools for IBD meta-omic data processing, analysis, and visualisation.

Tool	Function	Link	Availability
16S rRNA gene sequencing			
QIIME	Data processing and analysis	http://qiime.org/	Command-line tool
Mothur	Data processing and analysis	http://www.mothur.org/	Command-line tool
LotuS	Data processing and analysis	http://www.raeslab.org/software/lotus.html	Command-line tool
phyloseq	Data analysis and visualisation	http://www.bioconductor.org/packages/release/bioc/html/phyloseq.html	R package
vegan	Data analysis and visualisation	http://cran.r-project.org/web/packages/vegan/index.html	R package
Oligotyping	Taxonomic level profiling at higher resolution	http://oligotyping.org	Command-line tool
CoNet	Detection of microbial co-occurrence patterns	http://www.raeslab.org/software/conet.html	Command-line tool/ Cytoscape plugin
Metagenomics			
Trimmomatic	Trimming	http://www.usadellab.org/cms/?page = trimmomatic	Command-line tool
DeconSeq	Decontamination	http://deconseq.sourceforge.net/	User interface
MetAMOS	Assembly and data analysis	http://cbcb.umd.edu/software/metAMOS	Command-line tool
MOCAT	Assembly, taxonomic and functional profiling	http://vm-lux.embl.de/~kultima/MOCAT/	Command-line tool
Ray Meta	Assembly, taxonomic and functional profiling	http://denovoassembler.sourceforge.net/	Command-line tool
PBcR	Error correction and assembly	http://www.cbcb.umd.edu/software/PBcR/	Command-line tool
SmashCommunity	Annotation and analysis	http://www.bork.embl.de/software/smash/	Command-line tool
MetaPhlAn2	Taxonomic profiling	http://segatalab.cibio.unitn.it/tools/metaphlan2/	Galaxy module
MEDUSA	Taxonomic and functional profiling	http://www.metabolicatlas.com/medusa	Command-line tool
MLTreeMap	Phylogenetic analysis	http://mltreemap.org/	Web interface
PhyloSift	Phylogenetic analysis	https://phylosift.wordpress.com/	Command-line tool
MetaGeneMark	Gene prediction	http://exon.gatech.edu/meta_gmhmmp.cgi	Web interface
MetaProdigal	Gene prediction	http://prodigal.ornl.gov/	Command-line tool
Glimmer-MG	Gene prediction	http://www.cbcb.umd.edu/software/glimmer-mg/	Command-line tool
HUMAnN	Functional and metabolic analysis	http://huttenhower.sph.harvard.edu/humann	Command-line tool
FishTaco	Linking taxonomic and functional shifts	http://elbo.gs.washington.edu/software_fishtaco.html	Command-line tool
MetaNetSam	Network analysis	http://omics.informatics.indiana.edu/mg/MetaNetSam/	Command-line tool
MaAsLin	Multivariate associations	https://huttenhower.sph.harvard.edu/maaslin	Galaxy module
Metagenomics/metatranscriptomics			
MetaVelvet	Assembly	http://metavelvet.dna.bio.keio.ac.jp/	Command-line tool
EBI metagenomics	Phylogenetic and functional profiling	http://www.ebi.ac.uk/metagenomics/	Web interface
MG-RAST	Phylogenetic and functional profiling	http://metagenomics.anl.gov/	Web interface
LEfSe	Data analysis and visualisation	http://huttenhower.sph.harvard.edu/galaxy/	Web interface
MinPath	Metabolic pathway inference	http://omics.informatics.indiana.edu/MinPath/	Web interface
Metatranscriptomics			
IDBA-MT	Assembly	http://i.cs.hku.hk/~alse/hkubrg/projects/idba_mt/index.html	Command-line tool
Trinity	Assembly	http://trinityrnaseq.github.io/	Web interface
SortMeRNA	Data processing	http://bioinfo.lifl.fr/RNA/sortmerna/	Command-line tool
Metaproteomics			
SEQUEST	Peptide/protein identification	http://fields.scripps.edu/researchtools.php	Command-line tool
DTASelect	Protein assembly	http://www.scripps.edu/cravatt/protomap/dtaselect_instructions.html	Command-line tool
MASCOT	Protein annotation	http://www.matrixscience.com/	Web interface
MetaSPS	*De novo* protein sequencing [identification]	http://proteomics.ucsd.edu/software-tools/metasps/	Web interface
PepNovo+	*De novo* peptide sequencing [identification]	http://proteomics.ucsd.edu/software-tools/531–2/	Command-line tool
OpenMS	Quantification	http://open-ms.sourceforge.net/	Web interface
MetaProteomeAnalyzer	Peptide/protein identification and analysis	https://code.google.com/p/meta-proteome-analyzer/	User interface
Meta-omics integration			
ADE4	Data analysis and integration	http://pbil.univ-lyon1.fr/ade4/	Web interface
GOmixer	Gut-specific data analysis, integration and visualisation	http://www.raeslab.org/gomixer/	Web interface
HUMAnN2	Functional profiling, integration	http://huttenhower.sph.harvard.edu/humann	Command-line tool
MixOmics	Data analysis, integration, and visualisation	http://mixomics.qfab.org/	Web interface

target hypervariable regions of the 16S rRNA gene for a compromise between conserved flanking regions [to amplify all microorganisms present] and central variability [to distinguish closely related taxa]. Given the lack of truly universal primers, PCR amplification—to obtain DNA amplicon concentrations suited for library preparation—might introduce a bias towards specific taxonomic groups. Furthermore, rRNA operon copy number is known to vary according to bacterial growth strategies,[37] with fast growers like *Escherichia coli* encoding as many as seven rRNA operon copies and *Faecalibacterium prausnitzii* sp SL3/3 carrying a single one.[38] Genomes with higher 16S rRNA gene copy numbers are inevitably favoured in the process of exponential PCR amplification, introducing biases when estimating final bacterial abundances from amplicon sequencing data. A third confounder can be introduced in 16S sequencing data during OTU clustering and taxonomic assignment: as 16S rRNA gene sequences of abundant taxa are probably best represented in reference databases, it can be expected that they will be assigned more accurately.

Limitations notwithstanding, several primer sets with high coverage rates in microbial datasets are available [including 338f/r, 515f, 519r, 816r, 907r, 1062r], as well as tools for 16S rRNA gene copy number correction [CopyRighter[39]]. Regarding the third confounder, the human gut-associated microbiota has been extensively characterised, both by culturing and culture-independent methods, and coverage in 16S rRNA gene databases is estimated to encompass average colon microbiota composition. Only analyses of 16S rRNA gene profiles obtained from the gut content of model organisms such as the mouse remain comparatively harder.[40] 16S rRNA gene sequencing is nowadays a relatively straightforward and affordable approach to obtain information on the taxonomic composition of the gut microbiota.

3.3. Metagenomic sequencing

3.3.1. Metagenomic sequencing and its application to IBD research

Using metagenomics, randomly sequenced DNA obtained from the complete content of an environmental or clinical sample is analysed. Metagenomics allows characterisation not only of the taxonomic composition, but also of the full functional metabolic potential as represented by the combined gene pool of a microbiota sample, including for example antibiotic resistance profiles or virulence factors. Although sequencing costs and run times of metagenomic analyses are substantially higher and longer when compared with 16S sequencing, the benefits of information unlocked far outweigh the additional investment in resources.

Several large-scale metagenomic studies have provided insights into the complexity and diversity of the gut microbiota[9,10] and paved the way for targeted analyses in a clinical context. The MetaHIT project characterised the gut microbiota of almost 300 individuals [including 25 IBD patients].[9] It revealed discrete or continuous stratification patterns within individual microbiota configurations [enterotypes], independent of age, sex, or nationality.[41] The derived gene catalogue allowed not only definition of a core metagenome[9], but also identification of correlations between low microbial functional richness and markers of metabolic syndrome and inflammation.[42] The HMP[10] studied 18 microbial habitats associated to the human body in 242 individuals, performing metagenomic sequencing on over 1200 samples. Importantly, around 800 human-associated isolates were fully sequenced in the framework of this project. These microbial reference genomes aid annotation of metagenomic reads from faecal samples by sequence homology. Recently, an updated catalogue of the genes found in the human gut microbiome has been defined and released using data obtained in all major large-scale metagenomic projects, containing ~ 10M genes.[43] These pioneering efforts have generated a solid framework for future metagenomic studies.

In the context of IBD, in-depth analysis of the MetaHIT data set extended with 18 additional samples led to identification of genes and processes that are depleted or enriched in the disease,[44] most of which were related to interaction between the microbiota and the gut environment. Of the enriched processes, ~ 21% were PTS [phosphotransferase system] transporters and FrvX, a fructose-specific PTS protein which had already been identified as biomarker for IBD. Another 18% of the pathways enriched in IBD belonged to the nitrate reductase pathway, in line with high nitric oxide levels [the product of nitrate reductase] that have been associated with IBD. As the enzymes that were associated with IBD are only found in a relatively small subset of microbial genomes, the authors suggested that a few species might be responsible for the disease. Another study analysed the microbiota of 12 CD patients and found *Faecalibacterium* to be significantly depleted in ileum samples in their metagenomic analyses.[45] In addition, in CD subjects a lower proportion of the reads could not be assigned at phylum, family, or genus level compared with in controls, possibly reflecting the reduced bacterial diversity in CD.[45] Gevers *et al.*[46] performed shotgun sequencing on 33 CD patients and 10 controls, and found that the taxa enriched in CD—e.g. Fusobacteriaceae—were associated to pathways related to inflammation [lipopolysaccharide and glycerophospholipid metabolisms], whereas the depleted taxa—including Bacteroidales and Clostridiales—contributed amino acid and bile acid synthesis pathways. Again, reduced species richness of the mucosal microbiota was observed in CD patients.[46]

3.3.2. Metagenomic data sequencing

Random shotgun analysis of the whole community DNA is performed using next-generation platforms, mostly Illumina or 454 pyrosequencing [Roche]. Currently, Illumina platforms [MiSeq, HiSeq] can provide 25 million paired reads/run of up to 300 bp, whereas Roche GS FLX sequencers provide considerably longer reads [up to 1000 bp] but at a lower throughput of 1M reads/run. Although the protocols used are considerably different, they yield comparable assemblies and gene abundances.[47] Still, Illumina platforms have lower error rates [especially for low-complexity genome regions] and generate more accurate and longer contigs—sets of overlapping DNA segments—despite the shorter read length.[47] Alternatively, Ion Torrent PGM [Life Technologies], using a semiconductor-based technology, provides fast sequencing [2-6-h runs], and can yield 5M reads/run of up to 400bp. However, similar to 454 pyrosequencing, it has low accuracy for homopolymers and it generates more frameshift errors than Illumina platforms.[48] The so-called next-next-generation sequencing [single molecule sequencing] platforms perform sequencing reactions skipping any amplification step, avoiding PCR-induced biases. PacBio [Pacific Biosciences] produces ultra-long reads [15 kbp] and yields 70.000 reads in 30 min at the moment, at a ~ 95% accuracy. MinIon [Oxford Nanopore Technologies] is a pocket-size device—still in test phase—that also produces reads of tens of kbp and could potentially become faster and cheaper than the current sequencing platforms, but has an error rate of ~ 38%.[49] Illumina remain the most popular sequencing platforms nowadays, owning about 66% of the sequencing instruments that are currently in use.[50]

3.3.3. Metagenomic data processing

Several tools have been developed over the years to facilitate the processing of shotgun metagenomic data. The short reads generated

by Illumina and 454 platforms are usually trimmed from adapter sequences and low quality ends with Trimmomatic,[51] decontaminated from host sequences [human DNA] by, for example, DeconSeq[52]and assembled into larger contigs using tools such as Ray Meta.[53] The longer reads produced by PacBio and MinIon could be error-corrected by PBcR[54] and then assembled by Ray Meta or the same PBcR. Shotgun metagenomics reaches species-level resolution, but estimating species abundances from metagenomic datasets is not as straightforward as it might appear. As sequencing probability correlates to genome size, species abundances cannot be correctly calculated as the total number of reads matching microbial reference genomes. To circumvent this problem, tools have been developed to estimate community composition from single copy marker genes, such as MLTreeMap[55] and MetaPhlAn.[56] Of note, the extension of the latter, MetaPhlAn2,[57] reaches strain-level resolution. Adding another level of refinement, metagenomic operational taxonomic units [mOTUs][58] are reconstructed as groups of co-varying single-copy marker genes, allowing classification and quantification of known and unknown microorganisms at species level.[58,59] Additionally, species growth rates can be extrapolated from differential sequencing coverage by mapping reads on reference genomes, revealing the rates at which bacteria are duplicating in each sample, which could provide insights into pathology.[60]

For functional analyses, contigs are used for gene prediction [eg Glimmer-MG[61]], annotation, and quantification. Predicted genes are functionally annotated by sequence homology using functional databases such as COG,[62] KEGG,[63] or eggNOG.[64] Alternatively, assembly and gene calling can be skipped and reads are then directly used to generate functional profiles using EBI metagenomics[65] or by mapping them to the most recent version of the catalogue of genes in the human gut microbiome.[43] Typically, functional profiles are compared at the gene level, but they can also be used for metabolic pathway reconstruction [eg with MinPath[66]]. The latter allows comparison of the metabolic potential of the microbiota in different samples, using tools such as HUMAnN[67] or GOmixer [www.raeslab.org/gomixer/]—see Darzi *et al.*[68] for an example application. Data analysis and visualisation can be performed with MaAsLin, MG-RAST,[69] or GOmixer [Table 2]. FishTaco[70] can be used to link observed differences in taxonomic and functional composition, so as to identify the taxonomic groups that drive the functional shifts observed between patients and controls. The most informative—but computationally intensive—analytical strategy for metagenomic datasets involves clustering of all genes identified by abundance co-variation across all public gut metagenomic samples into co-abundance gene groups [CAGs]. This allows performance of functional profiling within boundaries, potentially corresponding to either genomes [taxonomic boundary] or co-varying genomes [eg genomes and plasmids, syntrophy partners]. In addition, when genome coverage tends to be optimal, this strategy permits assembling genomes *de novo*. After posterior assembly—using standard metagenome assembly tools—181 full genomes of previously unsequenced species were obtained.[71] Figure 2 provides an overview of a typical metagenomic workflow.

3.3.4. Pitfalls and limitations of metagenomics

Metagenomic data processing and data analysis remain computationally intensive and require advanced bioinformatics skills. Data pre-processing requires access to computer clusters able to handle big data. Both cost and expertise thus limit the generalised application of metagenomics in clinical research, and 16S rRNA gene sequencing is often used as an exploratory step before investing in metagenomic sequencing. Future advances in length of sequencing reads are expected to reduce assembly and gene prediction complexity. However, sequencing developments also imply readapting all tools to the changing configurations of datasets generated. It is therefore recommended to computational non-experts to use sequencing technologies for which the bioinformatics pipelines are already fully developed.

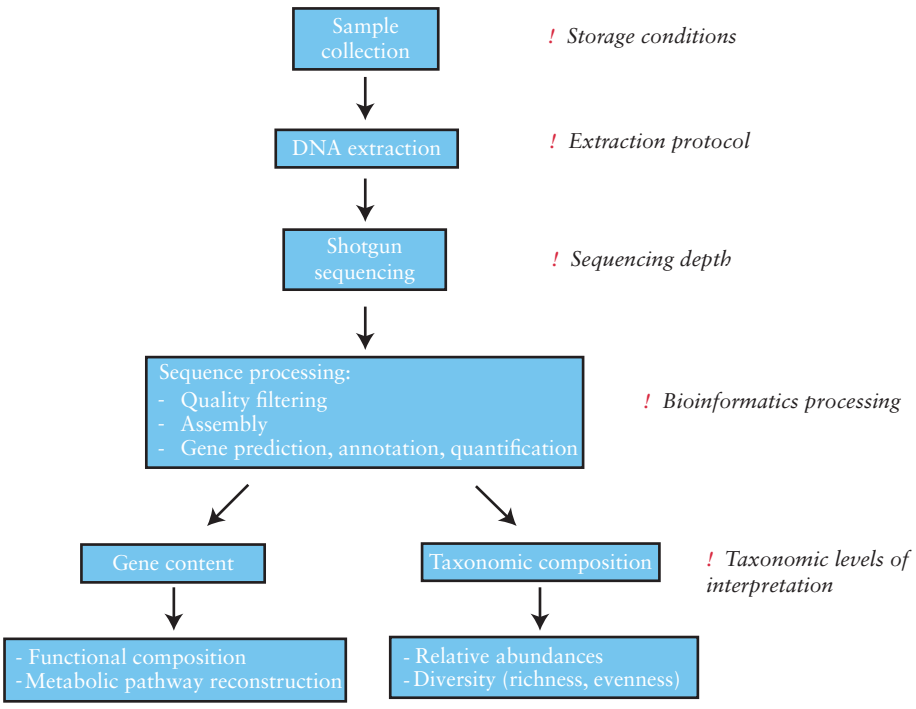

Figure 2. A typical metagenomic workflow and the main factors to be taken into account when comparing results of multiple gut microbiota studies.

Since they are not affected by gene copy number variation bias, metagenomic species abundance estimations based on single-copy marker genes are more accurate than their 16S rRNA gene-based counterparts. However, other biases, including differential DNA extraction efficiency between Gram-positive and Gram-negative bacteria or potential GC-content bias in sequencing efficiency, still remain. Furthermore, DNA extraction is performed from a certain weight and volume of faecal material with unknown [and variable across individuals / time points] microbial load. In addition, sequencing library preparation requires dilution of the DNA extracted to optimal process concentrations and equimolar pooling of DNA extracted from multiple samples. A similar number of sequencing reads is obtained from samples with high and low microbial density, leading to variation in sequencing depths. Consequently, metagenomics [but also 16S rRNA gene sequencing] inherently provides relative abundance estimations. Caution is thus needed when interpreting sample differences: the lower relative abundance of members of the Firmicutes phylum in CD samples reported by Erickson *et al.*,[45] based on metagenomic analyses, does not necessarily mean that these taxa are all depleted in CD patients, as differences could also result from the blooming of other taxa. Taking this into account, metagenomics can help disentangle the taxa and metabolic processes that are associated with IBD pathology.

4. Metatranscriptomics

4.1. Metatranscriptomic sequencing and its application to IBD research

Metatranscriptomics analyses the RNA transcript pool expressed by a community. As the mere presence of a gene in a metagenome does not guarantee its expression, and hence does not provide information about its expression pattern, metatranscriptomics provides valuable information that metagenomics overlooks. It allows monitoring regulation of and changes in microbial gene expression over time, which is particularly interesting when studying changes in the microbiota in response to perturbations. Whereas metagenomics reflects functional potential, metatranscriptomics provides information concerning the microbial processes that are active at a given time point. Formerly, community messenger RNA [mRNA] was studied using microarrays[72] or cDNA-AFLP [amplified fragment length polymorphism],[73] but recently, high-throughput sequencing technologies have been successfully applied to cDNA sequencing [RNA-seq].[74] However, its application to study the human microbiota in health and disease is still rather limited.

Gosalbes *et al.*[74] analysed the metatranscriptomes of faecal material from healthy individuals and compared them with their metagenomes. Most transcripts belonged to the metabolic processes of carbohydrate metabolism, energy production, and synthesis of cellular components, and amino acid and lipid metabolisms were under-represented in the metatranscriptome compared with the metagenome. They further found that the Bacteroidetes and the Firmicutes were the most active phyla. However, whereas the families Bacteroidaceae, Porphyromonadaceae, Clostridiaceae, and Bifidobacteriaceae had a higher relative abundance at the RNA level than at the DNA level, Lachnospiraceae were substantially lower at the RNA level—i.e. less active. They also reported the presence of small RNAs that play an important regulatory role in prokaryotic physiology and pathogenicity.[74] In IBD, following a similar strategy to 16S rRNA gene sequencing, Rehman *et al.* extracted microbial RNA from colon biopsies from IBD patients and, after reverse transcription to cDNA, amplified the 16S rRNA gene.[75,76] The Bacteroidetes were identified as the most active phylum both in CD patients and in healthy controls, whereas Actinobacteria and Firmicutes were comparatively less active in patients.[75] Using the same technique, colon biopsies of IBD patients were found to harbour a lower bacterial diversity at the RNA level as compared with DNA abundances, but more associations between microbiota and disease could be identified.[76] This highlights the importance of studying the active component of the microbiota in an IBD context.

4.2. Metatranscriptomic sample and data processing

Several techniques for RNA isolation are available nowadays. mRNA constitutes only a small fraction of the total sample RNA pool; therefore, to avoid sequencing mostly rRNA, effective protocols and commercial kits have been developed for mRNA enrichment. These protocols are based either on rRNA removal or on mRNA amplification, often using oligo-dT primers after introducing a polyA tail. After enrichment, the RNA is reverse-transcribed to cDNA and RNA-seq is performed using high-throughput sequencing platforms. Quality assessment and host sequences decontamination is performed using the same tools as metagenomics. Notwithstanding mRNA enrichment, it remains important to filter rRNA from mRNA to obtain meaningful results. This can be done using SortMeRNA.[77] For assembly, IDBA-MT[78]—which also performs chimera removal—is commonly used. The metatranscriptomic assembler Trinity[79] could increase the rate of assembled contigs at the risk of increased chimera. However, for the generation of gene expression profiles, mapping of the transcripts to the gene catalogue[43] is a faster and more accurate approach. At this point, the analysis and visualisation tools used in metagenomics could be applied for the downstream analysis after adapting the thresholds, as metatranscriptomic sequencing coverage is inherently lower due to the shorter half-life of mRNA compared with DNA.

4.3. Pitfalls and limitations of metatranscriptomics

Isolation of prokaryotic mRNA is hindered by the lack of the polyA tail that eases separation of mRNA from rRNA in eukaryotes. The low stability of mRNA makes RNA preparation tricky and, as gene expression profiles can change rapidly, the mRNA pool that is recovered could reflect the expression patterns of the microbiota in response to sampling-induced stress conditions rather than the metatranscriptome in the sampled individual. This also points to the main pitfall of metatranscriptomics: given the rapid change in mRNA pool, it is uncertain how well a faecal metatranscriptome represents the processes that were active in the ileum or colon hours to days before. Furthermore, as transcript abundance is a function of both the expression level of the gene as well as the abundance of the host organism, the interpretation of such data is challenging. Bioinformatic skills remain necessary for metatranscriptomic data processing and analysis at the moment, but several tools start including user-friendly web interfaces [Table 2].

5. Metaproteomics

5.1. Metaproteomics and their application to IBD research

Metaproteomics encompasses the large-scale study of the whole protein complement of environmental samples. It is a promising tool to gain knowledge on the functional diversity of the gut microbiota. Metaproteomics reveals what metabolic processes are ongoing and

how they are affected by changes in the environment such as disease conditions. Compared with metatranscriptomics, metaproteomics provides more direct information on metabolic processes that are carried out by the microbiota, as study of RNA expression does not account for post-transcriptional regulation. In addition, the metaproteome is more stable than the metatranscriptome, which is an advantage when analysing stool samples as it is more likely to reflect the actual microbial ecosystem and not sampling-induced alterations. A metaproteomics study of a healthy monozygotic twin pair revealed that the gut microbiota produces considerably more proteins for translation, energy production, and carbohydrate metabolism than was expected from metagenomic data[80]—similar to what was found using metatranscriptomics.[74] The same study identified host antimicrobial peptides,[80] providing information about the host response to the microbiota. Recently, Juste *et al.*[81] identified another 12 bacterial signals for CD using metaproteomics based on stool samples, which potentially correspond to functions of opportunistic pathogens, including mucosal layer colonisation, host barrier crossing, and mucosal invasion. They also found proteins from *Bacteroides* to be over-represented, and those from Firmicutes and *Prevotella* to be under-represented in CD patients. Using gastrointestinal biopsies, Presley *et al.*[82] identified biomarkers for IBD patients in the mucosal-luminal interface, mostly corresponding to functions related to microbe-host interactions [Table 1].

5.2. Metaproteomic sample and data processing

In metaproteomics, experimental protein spectrum profiles are normally obtained by liquid chromatography and tandem mass spectrometry [LC-MS/MS], or nuclear magnetic resonance [NMR]. Quality assessment and subsequent quality filtering are performed to remove the low quality spectra, so that only the potentially informative spectra are used as input for the time-consuming peptide identification tools. This is particularly necessary in metaproteomics due to the amount of information that is produced. To determine spectrum quality, most quality assessment algorithms use the number of peaks in the spectrum, total peak intensity, and the number of ions observed in the high-intensity peaks.[83] Then, experimental protein spectra profiles are matched to *in silico*-generated profiles for peptide identification and quantification. *In silico*-generated profiles are produced by translating a gene database [using the six coding frames] to protein sequences and simulating their cleavage by trypsin or other proteases. The identified peptides are then assembled in proteins using dedicated databases and are quantified during the process[84]; commonly used tools include SEQUEST[85] and MASCOT.[86] These databases need to contain the metagenome of the sample that is being analysed. Rooijers *et al.*[70] developed and implemented an iterative workflow that relies on the catalogue of reference genes in the human gut microbiome[43] to increase matching of protein spectra when the corresponding metagenome is not available. MetaSPS[87] permits identification of proteins with unknown sequence [*de novo* sequencing]. Tools such as QVALITY[88] are then used to assess the confidence of the identifications and to detect incorrect spectrum identifications. The inferred proteins are assigned taxonomically; however, this is challenging and generates ambiguities as proteins can be shared by multiple bacterial taxa. Finally, functional and metabolic pathway analyses are performed, in a similar manner as in metagenomics.

Recently, the first pipeline for metaproteomic analysis, MetaProteomeAnalyzer,[89] became available. The pipeline performs peptide and protein identification and annotation of protein taxonomy and function, decreases data redundancy by grouping protein hits in 'meta-proteins', and then further annotates them.

5.3. Pitfalls and limitations of metaproteomics

The high complexity and heterogeneity of metaproteomic samples hinder metaproteomic data analysis and interpretation. Peptide identification and quantification, protein inference from peptides, and taxonomic assignment of proteins that potentially belong to hundreds of species are challenging. However, first results are promising, and the increasing availability of metagenomic data and improvements in mass spectrometry are all expected to ease metaprotemic research. Moreover, tools like MetaProteomeAnalyzer further simplify the analysis of metaproteomic data by non-bioinformatics experts.

6. Meta-Omics Integration

6.1. Meta-omics integration and its application to IBD research

The integration of results of multiple meta-omic technologies can help understand the functioning of microbial communities beyond the possibilities of single meta-omic approaches. Using metaproteomics, the microbiota was found to produce more proteins than predicted from metagenomic studies[80] for some processes including energy production, carbohydrate metabolism, and translation. Furthermore, the integration of metagenomic and metatranscriptomic data showed that the relative abundance of taxonomic groups often differs between DNA and cDNA libraries.[90] In IBD, Erickson *et al.*[45] integrated metagenomic and metaproteomic data from healthy and CD twin pairs and found that only a subset of the genes differentially detected in the microbiome of CD patients with metagenomics were expressed and corresponded to identified proteins in the metaproteome. Also, pathway expression differed more between healthy and CD individuals at the metaproteomic level than at the metagenomic level, and species-specific analysis showed that a lowering of transcripts encoding for butyrate production could be attributed to both a reduction in the number of carrying organisms [*F. prausnitzii*] and the specific down-regulation of the enzymes by this organism.[45]

Likewise, integrating other 'microbial' omes [from intestinal eukaryotes and viruses] also holds great promise to provide new insights in IBD. Bacteria account for a very large fraction of the microbiome, but the viral and fungal components should not be overlooked. Analyses of the virome and mycobiome are in their infancy, and specific experimental protocols and bioinformatic pipelines and databases are still being developed. Nevertheless, a pioneer study found that the enteric virome is altered in CD and UC patients compared with healthy individuals, with increased diversity and richness of bacteriophages—opposite to what has been consistently observed for the bacterial microbiome, which suggests that the virome could be an important player in IBD.[91] The mycobiome has also been reported to be altered in CD, and similarly, increased diversity was observed at the inflamed mucosa compared with the non-inflamed mucosa.[92]

Integrating multiple layers of information—also including metadata such as clinical history—is definitely a promising approach to obtain a more complete picture of the gastrointestinal ecosystem, which is especially interesting for a multifactorial disease like IBD.

6.2. Meta-omic data integration

Several tools that were initially developed for single organism meta-omics could be applied to explore, integrate, and visualise meta-omic datasets. For instance, mixOmics [http://mixomics.org/] and ADE4[93] detect associations between sets of variables [eg proteins and species]

on matched samples [ie different measurements on the same sample], and provide graphical representation. Both are available as R packages; mixOmics is also available via a web interface. GOmixer [http://www.raeslab.org/gomixer/] is a user-friendly web application for functional analysis [with a gut-specific metabolic module framework], integration, and visualisation of gut meta-omic data. The next generation of HUMAnN [HUMAnN2] also supports integration of metatranscriptomic and metagenomic data.

6.3. Pitfalls and limitations of meta-omics integration

Integration of meta-omics remains challenging, and only a few studies have done this until now. Sample size, sequencing depth, and annotation rate are all potential limiting factors for integration of meta-omic datasets, and thus need to be taken into consideration in the design of such studies. Also, the interpretation of data at multiple levels can remain a challenge given the fact that all datasets will be inherently under-sampled and not perfectly overlapping due to technical biases. Cost is an important limitation for multiple meta-omics studies, but the constant drop in sequencing cost enables sequencing a greater number of samples and at a higher sequencing depth. In addition, the release of the 10 million gut microbial gene catalogue[43] has improved gene annotation in metagenomics/metatranscriptomics and spectral matching in metaproteomics, decreasing the proportion of unassigned reads and spectra.

7. Conclusions

Research in IBD benefits considerably from advances in meta-omic technologies, as the steep increase in IBD publications that use meta-omics and their findings demonstrate. Sample preparation protocols are critical as they can affect microbiota composition, already starting from storage of samples, and hamper comparison of different studies. 16S rRNA gene sequencing is now widely available and user-friendly bioinformatics tools and pipelines have been developed to analyse these data. Metagenomic sequencing can complement 16S by providing information on the functional capacity of the microbiota and allows species and strain-level analysis. Metatranscriptomics and metaproteomics reveal the functional dynamics of the microbiota. These technologies are still at an early stage, and the fact that some of the results obtained in different studies contradict each other highlights the need for further standardisation and benchmarking. Furthermore, although multi-omic data integration tools such as GOmixer, HUMAnN2, and FishTaco are becoming available, the bioinformatics of multi-readout experiments remains challenging. Combinatory, multi-omic approaches have the potential to help understand a complex, multifactorial disease like IBD.

Glossary

Contigs

DNA sequence recreated by assembling partially overlapping contiguous DNA fragments.

Functional diversity

Number of genes or functions encoded by any member of the community.

Gene annotation

Assignment of functional information to genes by sequence similarity [functional domains, function].

Metagenomics

Study of the collective genomic content of a microbial community.

Metaproteomics

Study of protein production by a microbial community.

Metatranscriptomics

Study of genes expressed by a microbial community.

OTUs: operational taxonomic units, clusters of sequences with reciprocal similarity above a certain threshold [usually 97%].

Phylogenetic trees

Representation of evolutionary relationships between species, reconstructed from comparative sequence analysis.

Reference genome

Fully sequenced and assembled genome representative for a species or strain.

Taxonomic assignment

Assignment of microbial lineage information to genes by sequence similarity.

Taxonomic diversity

Number and relative abundances of taxa in a community.

Funding

The Raes laboratory is supported by the Rega Institute for Medical Research, KU Leuven, the Agency for Innovation by Science and Technology [IWT], and the Fund for Scientific Research Flanders [the FWO]. MVC, SVS, and MJ are supported by pre- and post-doctoral fellowships from the FWO, respectively.

Conflict of Interest

JR has acted as scientific advisor for GSK, Johnson & Johnson, and 23andme.

Acknowledgments

We thank all members of the Raes Laboratory for lively discussions. MVC and YD performed the literature review and drafted the manuscript. SVS, GF, MJ, and JR critically revised the manuscript for important intellectual content. All authors approved the submitted version of the manuscript.

References

1. Crohn BB, Ginzburg L, Oppenheimer GD. Regional ileitis: a pathologic and clinical entity. *Mt Sinai J Med* 1932;**67**:263–8.
2. Molodecky NA, Soon IS, Rabi DM, *et al.* Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology* 2012;**142**:46–54
3. O'Hara AM, Shanahan F. The gut flora as a forgotten organ. *EMBO Rep* 2006;**7**:688–93.
4. Carroll IM, Ringel-Kulka T, Siddle JP, Klaenhammer TR, Ringel Y. Characterization of the fecal microbiota using high-throughput sequencing reveals a stable microbial community during storage. *PLoS One* 2012;**7**:e46953.
5. Cardona S, Eck A, Cassellas M, *et al.* Storage conditions of intestinal microbiota matter in metagenomic analysis. *BMC Microbiol* 2012;**12**:158.
6. Lyra A, Forssten S, Rolny P, *et al.* Comparison of bacterial quantities in left and right colon biopsies and faeces. *World J Gastroenterol* 2012;**18**:4404–11.
7. Budding AE, Grasman ME, Eck A, Bogaards JA, Vandenbroucke-Grauls CMJE, Van Bodegraven AA, *et al.* Rectal swabs for analysis of the intestinal microbiota. *PLoS One* 2014;**9**:5–12.
8. Zoetendal EG, Heilig HGHJ, Klaassens ES, *et al.* Isolation of DNA from bacterial samples of the human gastrointestinal tract. *Nat Protoc* 2006;**1**:870–3.

9. Qin J, Li R, Raes J, *et al*. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 2010;**464**:59–65.
10. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* 2012;**486**:207–14.
11. Wesolowska-Andersen A, Bahl MI, Carvalho V, *et al.* Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis. *Microbiome* 2014;**2**:19.
12. Santiago A, Panda S, Mengels G, *et al.* Processing faecal samples: a step forward for standards in microbial community analysis. *BMC Microbiol* 2014;**14**:112.
13. Darfeuille-Michaud A, Neut C, Barnich N, *et al.* Presence of adherent Escherichia coli strains in ileal mucosa of patients with Crohn's disease. *Gastroenterology* 1998;**115**:1405–13.
14. Joossens M, Huys G, Cnockaert M, *et al.* Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut* 2011;**60**:631–7.
15. Machiels K, Joossens M, Sabino J, *et al*. A decrease of the butyrate-producing species Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis. *Gut* 2014;**63**:1275–83.
16. Dicksved J, Halfvarson J, Rosenquist M, *et al.* Molecular analysis of the gut microbiota of identical twins with Crohn's disease. *ISME J* 2008;**2**:716–27.
17. Sokol H, Pigneur B, Watterlot L, *et al*. Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U S A* 2008;**105**:16731–6.
18. Rajilić-Stojanović M, Heilig HGHJ, Molenaar D, *et al*. Development and application of the human intestinal tract chip, a phylogenetic microarray: analysis of universally conserved phylotypes in the abundant microbiota of young and elderly adults. *Environ Microbiol* 2009;**11**:1736–51.
19. Walker AW, Sanderson JD, Churcher C, *et al.* High-throughput clone library analysis of the mucosa-associated microbiota reveals dysbiosis and differences between inflamed and non- inflamed regions of the intestine in inflammatory bowel disease. *BMC Microbiol* 2011;**11**:7.
20. Lane DJ, Pace B, Olsen GJ, Stahlt DA, Sogint ML, Pace NR. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc Natl Acad Sci U S A* 1985;**82**:6955–9.
21. Eckburg PB, Bik EM, Bernstein CN, *et al*. Diversity of the human intestinal microbial flora. *Science* 2005;**308**:1635–8.
22. Zitomersky NL, Atkinson BJ, Franklin SW, *et al*. Characterization of Adherent Bacteroidales from Intestinal Biopsies of Children and Young Adults with Inflammatory Bowel Disease. *PLoS One* 2013;**8**[6]:63686.
23. De Cruz P, Kang S, Wagner J, *et al*. Specific Mucosa-Associated Microbiota in Crohn's Disease at the Time of Resection are Associated with Early Disease Recurrence: A Pilot Study. *J Gastroenterol Hepatol* 2015;**30**:26878.
24. Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci U S A* 2007;**104**:13780–5.
25. Morgan XC, Tickle TL, Sokol H, *et al.* Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol* 2012;**13**:R79.
26. Caporaso JG, Kuczynski J, Stombaugh J, *et al*. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;**7**:335–6.
27. Schloss PD, Westcott SL, Ryabin T, *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009;**75**:7537–41.
28. Hildebrand F, Tito RY, Voigt A, Bork P, Raes J. LotuS: an efficient and user-friendly OTU processing pipeline. *Microbiome* 2014;**2**:30
29. DeSantis TZ, Hugenholtz P, Larsen N, *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 2006;**72**:5069–72.
30. Oksanen J, Blanchet FG, Kindt R, *et al. vegan: Community Ecology Package*. R package version 2.2-1. Vienna: R Core Team, 2015.
31. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 2013;**8**:e61217.
32. Gauch HG. *Multivariate Analysis in Community Ecology*. Vol. **34**. Cambridge, UK: Cambridge University Press, 1982.
33. Faust K, Raes J. Microbial interactions: from networks to models. *Nat Rev Microbiol* 2012;**10**:538–50.
34. Faust K, Sathirapongsasuti JF, Izard J, *et al.* Microbial co-occurrence relationships in the human microbiome. *PLoS Comput Biol* 2012;**8**:e1002606.
35. Eren M, Maignien L, Sul WJ, *et al.* Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol Evol* 2013;**4**:1111–9.
36. Langille MGI, Zaneveld J, Caporaso JG, *et al*. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* 2013;**31**:814–21.
37. Vieira-Silva S, Rocha EPC. The systemic imprint of growth and its uses in ecological [meta]genomics. *PLoS Genet* 2010;**6**:1000808.
38. Markowitz VM, Chen I-MA, Palaniappan K, *et al*. IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res* 2014;**42**[Database issue]:D560–7.
39. Angly FE, Dennis PG, Skarshewski A, Vanwonterghem I, Hugenholtz P, Tyson GW. CopyRighter: a rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction. *Microbiome* 2014;**2**:11.
40. Nguyen TLA, Vieira-Silva S, Liston A, Raes J. How informative is the mouse for human gut microbiota research? *Dis Model Mech* 2015;**8**:1–16.
41. Arumugam M, Raes J, Pelletier E, *et al*. Enterotypes of the human gut microbiome. *Nature* 2011;**473**:174–80.
42. Le Chatelier E, Nielsen T, Qin J, *et al.* Richness of human gut microbiome correlates with metabolic markers. *Nature* 2013;**500**:541–6.
43. Li J, Jia H, Cai X, *et al*. An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol* 2014;**32**:834–41.
44. Greenblum S, Turnbaugh PJ, Borenstein E. Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. *PNAS* 2012;**109**:594–9.
45. Erickson AR, Cantarel BL, Lamendella R, *et al.* Integrated metagenomics/metaproteomics reveals human host-microbiota signatures of Crohn's disease. *PLoS One* 2012;**7**:e49138.
46. Gevers D, Kugathasan S, Denson LA, *et al.* The Treatment-Naive Microbiome in New-Onset Crohn's Disease. *Cell Host Microbe* 2014;**15**:382–92.
47. Luo C, Tsementzi D, Kyrpides N, Read T, Konstantinidis KT. Direct comparisons of Illumina vs.Roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS One* 2012;**7**:e30087.
48. Loman NJ, Misra R V, Dallman TJ, *et al*. Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* 2012;**30**:434–41.
49. Laver T, Harrison J, O'Neill P, *et al*. Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomol Detect Quantif* 2015;**3**:1–8.
50. Minoche AE, Dohm JC, Himmelbauer H. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. *Genome Biol* 2011;**12**:R112.
51. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;**30**:2114–20.
52. Schmieder R, Edwards R. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS One* 2011;**6**:e17288.
53. Boisvert S, Raymond F, Godzaridis E, Laviolette F, Corbeil J. Ray Meta: scalable de novo metagenome assembly and profiling. *Genome Biol* 2012;**13**:R122.
54. Berlin K, Koren S, Chin C-S, Drake JP, Landolin JM, Phillippy AM. Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. *Nat Biotechnol* 2015;**33**:623–30.
55. Stark M, Berger S, Stamatakis A, von Mering C. MLTreeMap-accurate maximum likelihood placement of environmental DNA sequences into taxonomic and functional reference phylogenies. *BMC Genomics* 2010;**11**:461.

56. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods* 2012;**9**:811–4.
57. Truong DT, Franzosa E, Tickle TL, *et al.* MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat Methods* 2015;**12**[:902–3.
58. Sunagawa S, Mende DR, Zeller G, *et al.* Metagenomic species profiling using universal phylogenetic marker genes. *Nat Methods* 2013;**10**:1196–9.
59. Von Mering C, Hugenholtz P, Raes J, *et al.* Quantitative phylogenetic assessment of microbial communities in diverse environments. *Science* 2007;**315**:1126–30.
60. Korem T, Zeevi D, Suez J, *et al.* Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. *Science* 2015;**349**:110-16.
61. Kelley DR, Liu B, Delcher AL, Pop M, Salzberg SL. Gene prediction with Glimmer for metagenomic sequences augmented by classification and clustering. *Nucleic Acids Res* 2012;**40**:e9
62. Tatusov RL, Galperin MY, Natale D, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* 2000;**28**:33–6.
63. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* 2014;**42**[Database issue]:D199–205.
64. Powell S, Forslund K, Szklarczyk D, *et al.* EggNOG v4.0: Nested orthology inference across 3686 organisms. *Nucleic Acids Res* 2014;**42**[Database issue]:D231–9.
65. Hunter S, Corbett M, Denise H, *et al.* EBI metagenomics-a new resource for the analysis and archiving of metagenomic data. *Nucleic Acids Res* 2014;**42**[Database issue]:D600–6.
66. Ye Y, Doak TG. A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. *PLoS Comput Biol* 2009;**5**:e1000465.
67. Abubucker S, Segata N, Goll J, *et al.* Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput Biol* 2012;**8**:e1002358.
68. Darzi Y, Falony G, Vieira-Silva S, Raes J. Towards biome-specific analysis of meta-omics data. *ISME J* 2015.
69. Meyer F, Paarmann D, D'Souza M, *et al.* The metagenomics RAST server-a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 2008;**9**:386.
70. Rooijers K, Kolmeder C, Juste C, *et al.* An iterative workflow for mining the human intestinal metaproteome. *BMC Genomics* 2011;**12**(1):6. Doi: 10.1186/1471-2164-12-6.
71. Nielsen HB, Almeida M, Juncker AS, Rasmussen S, Li J, Sunagawa S, *et al.* Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat Biotechnol* 2014;**32**;822-8.
72. Mahowald MA, Rey FE, Seedorf H, *et al.* Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. *Proc Natl Acad Sci U S A* 2009;**106**:5859–64.
73. Booijink CCGM, Boekhorst J, Zoetendal EG, Smidt H, Kleerebezem M, De Vos WM. Metatranscriptome analysis of the human fecal microbiota reveals subject-specific expression profiles, with genes encoding proteins involved in carbohydrate metabolism being dominantly expressed. *Appl Environ Microbiol* 2010;**76**:5533–40.
74. Gosalbes MJ, Durbán A, Pignatelli M, *et al.* Metatranscriptomic approach to analyze the functional human gut microbiota. *PLoS One* 2011;**6**:1–9.
75. Rehman A, Lepage P, Nolte A, Hellmig S, Schreiber S, Ott SJ. Transcriptional activity of the dominant gut mucosal microbiota in chronic inflammatory bowel disease patients. *J Med Microbiol* 2010;**59**]:1114–22.
76. Rehman A, Rausch P, Wang J, *et al.* Geographical patterns of the standing and active human gut microbiome in health and IBD. *Gut* 2015;**0**:1–11.
77. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 2012;**28**:3211–7.
78. Leung HCM, Yiu S-M, Parkinson J, Chin FYL. IDBA-MT: De Novo Assembler for Metatranscriptomic Data Generated from Next-Generation Sequencing Technology. *J Comput Biol* 2013;**20**:540–50.
79. Grabherr MG, Haas BJ, Yassour M, *et al.* Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol* 2011;**29**:644–52.
80. Verberkmoes NC, Russell AL, Shah M, *et al.* Shotgun metaproteomics of the human distal gut microbiota. *ISME J* 2009;**3**:179–89.
81. Juste C, Kreil DP, Beauvallet C, *et al.* Bacterial protein signals are associated with Crohn's disease. *Gut* 2014;**63:**1566-77.
82. Presley LL, Ye J, Li X, *et al.* Host-microbe relationships in inflammatory bowel disease detected by bacterial and metaproteomic analysis of the mucosal-luminal interface. *Inflamm Bowel Dis* 2012;**18**:409–17.
83. Bern M, Goldberg D, McDonald WH, Yates JR. Automatic quality assessment of peptide tandem mass spectra. *Bioinformatics* 2004;**20**[Suppl. 1]:49–54.
84. Tabb DL, McDonald WH, Yates JR. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res* 2002;**1**:21–6.
85. Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 1994;**5**:976–89.
86. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999;**20**:3551–67..
87. Guthals A, Clauser KR, Frank AM, Bandeira N. Sequencing-Grade De novo Analysis of MS/MS Triplets [CID/HCD/ETD] From Overlapping Peptides. *J Proteome Res* 2013;**12**:2846–57.
88. Käll L, Storey JD, Noble WS. QVALITY: non-parametric estimation of q-values and posterior error probabilities. *Bioinformatics* 2009;**25**:964–6.
89. Muth T, Behne A, Heyer R, *et al.* The MetaProteomeAnalyzer: A Powerful Open-Source Software Suite for Metaproteomics Data Analysis and Interpretation. *J Proteome Res* 2015:150223140604002.
90. Shi Y, Tyson GW, Eppley JM, DeLong EF. Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *ISME J* 2011;**5**:999–1013.
91. Norman JM, Handley SA, Baldridge MT, *et al.* Disease-Specific Alterations in the Enteric Virome in Inflammatory Bowel Disease. *Cell* 2015;**160**:447–60.
92. Li Q, Wang C, Tang C, He Q, Li N, Li J. Dysbiosis of Gut Fungal Microbiota is Associated With Mucosal Inflammation in Crohn's Disease. *J Clin Gastroenterol* 2014;**48**:513–23.
93. Dray S, Dufour B. The ade4 Package: Implementing the Duality Diagram for Ecologists. *J Stat Softw* 2007;**22**:1–20.
94. Manichanh C, Rigottier-Gois L, Bonnaud E, *et al.* Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* 2006;**55**[:205–11.
95. Davenport M, Poles J, Leung JM, *et al.* Metabolic alterations to the mucosal microbiota in inflammatory bowel disease. *Inflamm Bowel Dis* 2009;**20**:723–31.
96. Lepage P, Hösler R, Spehlmann ME, *et al.* Twin study indicates loss of interaction between microbiota and mucosa of patients with ulcerative colitis. *Gastroenterology* 2011;**141**:227–36.
97. Ricanek P, Lothe SM, Frye SA, Rydning A, Vatn MH. Gut bacterial profile in patients newly diagnosed with treatment-naïve Crohn's disease. *Clin Exp Gastroenterol* 2012;**5**:173–86.
98. Said HS, Suda W, Nakagome S, *et al.* Dysbiosis of Salivary Microbiota in Inflammatory Bowel Disease and Its Association With Oral Immunological Biomarkers. *DNA Res* 2014;**21**:15–25.
99. Thorkildsen LT, Nwosu FC, Avershina E, *et al.* Dominant fecal microbiota in newly diagnosed untreated inflammatory bowel disease patients. *Gastroenterol Res Pract* 2013;**2013**: 636785.
100. Wills ES, Jonkers DMAE, Savelkoul PH, Masclee AA, Pierik MJ, Penders J. Fecal microbial composition of ulcerative colitis and Crohn's disease patients in remission and subsequent exacerbation. *PLoS One* 2014;**9**:1–10.