



**HAL**  
open science

## Metaproteomics of the human gut microbiota: Challenges and contributions to other OMICS

Ngom Issa Isaac, Decloquement Philippe, Armstrong Nicholas, Didier Raoult,  
Chabrière Eric

► **To cite this version:**

Ngom Issa Isaac, Decloquement Philippe, Armstrong Nicholas, Didier Raoult, Chabrière Eric. Metaproteomics of the human gut microbiota: Challenges and contributions to other OMICS. Clinical Mass Spectrometry, 2019, 14, pp.18-30. 10.1016/j.clinms.2019.06.001 . hal-02457995

**HAL Id: hal-02457995**

**<https://amu.hal.science/hal-02457995>**

Submitted on 20 Jul 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 **Title: Metaproteomics of the human gut microbiota: challenges and contributions to**  
2 **other OMICS**

3

4 **Author names and affiliations:** Ngom Issa Isaac<sup>1</sup>, Decloquement Philippe<sup>1</sup>, Armstrong  
5 Nicholas<sup>1</sup>, Didier Raoult<sup>1</sup>, and Chabrière Eric<sup>1</sup>

6 <sup>1</sup> Aix Marseille Univ, IRD, AP-HM, MEPHI (Microbes Evolution Phylogénie et Infection),  
7 IHU-Méditerranée Infection, Marseille, France.

8

9 **Corresponding author:** Chabrière Éric, Institut Hospitalo-Universitaire Méditerranée  
10 Infection, 19-21 Boulevard Jean Moulin 13385 Marseille Cedex 05, France.

11 Phone: + 33 (0) 4 13 73 24 01. Fax: + 33 (0) 4 13 73 24 02.

12 E-mail address: [eric.chabriere@univ-amu.fr](mailto:eric.chabriere@univ-amu.fr)

13 **Abstract**

14 Our digestive tract hosts more than a billion microorganisms comprising non-pathogenic  
15 bacteria, viruses, fungi and parasites. Understanding and characterizing the human gut  
16 microbiota has become a fundamental common theme to establish a link between its dysbiosis  
17 and certain pathologies, especially autoimmune and inflammatory diseases. Meta-Omics  
18 studies have, so far, provided great progress in this field. Genomics is conventionally used to  
19 determine the composition of the microbiota and, subsequently, metatranscriptomics lists the  
20 transcribed genes. However, to better understand the relationship between microbiota and  
21 health, protein-based studies are being applied. Proteomics enables the functional study of  
22 proteins as they are expressed by microbial communities. Metaproteomics exploits the power  
23 of mass spectrometry to identify broad protein profiles in complex samples, such as gut  
24 microbiota. The latest technological advances in the field of mass spectrometry have opened  
25 the field of large-scale characterization of microbial proteins. Despite these hardware  
26 improvements, bioinformatics analysis remains a primary challenge. Herein, we describe the  
27 state-of-the-art concerning specific sample preparation and powerful shotgun analysis  
28 techniques. We also review several scientific studies of the human gut microbiota. Moreover,  
29 we discuss the advantages and limitations encountered in this research area, concerning new  
30 methods of sample preparation and innovative bioinformatic tools. Finally, prospects are  
31 addressed regarding the application of metaproteomic in the field of clinical microbiology and  
32 its integration with other meta-Omics.

33 **Keywords:** Metaproteomics; Human Gut Microbiota; OMICS; liquid chromatography; mass  
34 spectrometry; Unipept

35

36 **Abbreviations:**

37 CD: Crohn's disease

38 BLAST: Basic Local Alignment Search Tool

39 DDA/DIA: Data Dependent Acquisition / Data Independent Acquisition

40 FASP: Filter Aided Sample Preparation

41 FDR: False Discovery Rate

42 IBD: Inflammatory Bowel Disease

43 IMS/MS: Ion-Mobility Spectrometry–Mass Spectrometry

44 LCA: Lowest Common Ancestor

45 MALDI/ESI: Matrix-Assisted Laser Desorption Ionisation/Electrospray Ionization

46 NGS: Next Generation Sequencing

47 (U)HPLC: (Ultra) High Performance Liquid Chromatography

48 RP/SCX: Reverse Phase /Strong Cation Exchange

49 SDS-PAGE: Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis

50 TOF: Time-Of-Flight

## 51 **Introduction**

52 The human gut microbiota harbors complex communities of billions of microorganisms.  
53 These microorganisms, bacteria, viruses, archaea, yeasts and protozoa, are ten times more  
54 numerous than human cells [1,2]. Under normal conditions, this complex population lives in  
55 mutual coexistence with the body and plays several fundamental roles that **have** a  
56 considerable impact on human health and physiology [3,4]. Most of the microorganisms in  
57 the human microbiome are beneficial and play major metabolic and physiological roles. For  
58 example, the commensal microflora of the gut participate in the digestion of food [5], are  
59 involved in gut-brain intercommunication [6], and play an interactive role with immune  
60 system [3]. However, many factors can disturb the intestinal microbiota composition, known  
61 as dysbiosis. This microbial imbalance disrupts the microbiota composition and can lead to  
62 intestinal permeability. Alterations of the microbial ecosystem can occur due to several  
63 factors, such as environment, aging, diet and the immune system. As a result, changes in the  
64 bacterial composition of the gut microbiota have been associated with dysfunction of the  
65 digestive system, such as inflammatory bowel diseases, but also with obesity, metabolic,  
66 immune and neurological diseases and cancers [7-10]. **(Figure 1)**

67 **The different OMICs approaches have led to important advances in the study of the intestinal**  
68 **microbiome, the host and the intestinal environment. As well, next-generation sequencing**  
69 **(NGS) has allowed the use of genomic approaches to better understand the complex microbial**  
70 **environment from different biological samples.** Mainly, metagenomics provides a  
71 comprehensive overview of the **taxonomy** and functional potential of microbial ecosystems  
72 [11, 12]. However, despite these advances, metagenomics cannot address all biological  
73 questions. The different NGS platforms used in laboratories, or the choice of bioinformatics  
74 tools, remain the main limitations [13]. Moreover, the least abundant microorganisms are  
75 statistically less likely to be detected, constituting a depth bias for high-throughput

76 sequencing methods. In this respect, the metatranscriptomic (RNAseq) provides access to the  
77 metatranscriptome of the microbiome, allowing whole-genome profiling of the active  
78 microbial community and expressed biological signatures in the human microbiome [14].  
79 However, bioinformatics tools for metatranscriptome data analysis are similar to those of  
80 metagenomics. Culturomics is also a culture-based omics approach that uses multiple culture  
81 conditions, MALDI-TOF mass spectrometry and 16S rRNA sequencing for the rapid  
82 identification of bacterial species [15]. Proteomics, initially defined by microbiologists as the  
83 study of all proteins expressed by a single organism, is in full emergence thanks to its  
84 application to complex bacterial communities. As a result, the analysis of the protein content  
85 of the microbial communities, such as gut microbiota is now named “metaproteomics” [16].  
86 A metaproteomic analysis typically comprises 4 steps: 1) extraction and purification of  
87 proteins, 2) enzymatic digestion of proteins into peptides, 3) separation of peptides, usually by  
88 chromatography, followed by mass spectrometric analysis and 4) protein identification by  
89 database sequence comparison [17-19]. Metaproteomics is a rising technique but has some  
90 disadvantages related to the complexity of the sample, including both the complexity of the  
91 matrix as well as the microbial community itself. First, metaproteomes includes up to more  
92 than one thousand different species, each containing several hundred proteins, generating a  
93 myriad of peptides after digestion [20]. In addition, many peptides are common to many  
94 bacterial species or similar protein sequences, making data processing even more complex  
95 with a resultant high false-positive rate. Second, mass spectrometry generates hundreds of  
96 thousands of spectra, but the data analysis requires considerable bioinformatic effort to  
97 develop algorithms that will allow a reduction in the computational time needed. Third, one of  
98 the main elements of a successful metaproteomic study is the availability of a relevant  
99 database in order to match sequences with mass spectra. Moreover, a drawback of  
100 metaproteomics is its potential to generate numerous false positives from the use of large

101 databases. In addition, data interpretation is recognized as a major limitation for  
102 metaproteomic analysis because huge amounts of data often result in high False Discovery  
103 Rates (FDR). Solutions are required to validate protein identifications across different MS and  
104 database search algorithms. Furthermore, metagenomics, metatranscriptomics, metabolomics  
105 and culturomics data can be integrated with metaproteomic to provide insight into the  
106 functioning of bacterial communities in the gut.

107 In light of these considerations, this review presents the current status of shotgun  
108 metaproteomic (bottom-up) studies applied to the human gut microbiota and highlights  
109 experimental and bioinformatics approaches, providing several examples. Finally, we address  
110 the prospects of gut metaproteomic analysis and future directions for clinical microbiology  
111 research.

## 112 **1. Metaproteomics methodologies**

### 113 **1.1 Stool sample preparation**

114 The study of the metaproteome of gut microbiota is primarily performed using faecal samples.  
115 However, stool comprises a complicated environmental matrix that can interfere with protein  
116 characterization studies [18]. Several challenges should be expected, such as: 1) a complex  
117 microbial composition, as faeces consist of a mix of gram-positive and gram-negative cells  
118 with various envelopes structures, 2) an abundance of host proteins, 3) the presence of  
119 proteins derived from consumed and undigested foods, 4) various physico-chemical  
120 properties of proteins involved in their solubility. Metaproteomic analysis can be altered by  
121 large inter- and intra-donor variabilities. Microbial species abundance in the gut can also vary  
122 more than 10 orders of magnitude across samples. The more complex and diverse the  
123 microbial community, the fewer proteins that can be identified for each taxa [21]. The  
124 performance of the metaproteomic analysis from human gut microbiota is also dependent on  
125 sample preparation [22]. Storage conditions can affect the sample, hence, strict protocol for  
126 stool storage following defecation is required, typically at  $-80\text{ }^{\circ}\text{C}$  [23]. To date, numerous  
127 metaproteomic studies have achieved separation of microbial cells from feces by differential  
128 centrifugation, where insoluble material and large particles are separated at low speed,  
129 followed by pelleting microorganisms at higher centrifuge forces [24]. For example, *Tanca* et  
130 al. showed that stool samples previously treated by differential centrifugation revealed more  
131 proteins/peptides with a significantly higher microbial diversity than a direct conventional  
132 protein extraction step [25]. Additionally, Xiong et al developed a metaproteomic sample  
133 preparation strategy based on a double filtering (DF) differential separation step that  
134 selectively depletes human cells and proteins while enriching microbial biomass in the fecal  
135 sample [129]. The DF process constituted (1) a  $20\text{ }\mu\text{m}$  vacuum filter unit to remove larger  
136 fibrous material and intact human cells, and (2) a  $0.22\text{ }\mu\text{m}$  vacuum filter unit that permitted

137 human proteins to be washed through while microbial cells were captured on the filter. This  
138 method resulted in greater than a 2-fold increase in microbial proteins that were identified and  
139 quantitated compared to the direct method whose protein extraction was performed using high  
140 speed centrifugation. Cell lysis should be adapted for gut microbiota. For instance, Gram  
141 positive bacteria, such as *Firmicutes* and *Actinobacteria*, which are two major phyla of the  
142 intestinal tract, have a thick peptidoglycan layer that is difficult to break down. Hence, a wide  
143 range of physical, mechanical and chemical methods are generally used in metaproteomic  
144 studies to disrupt cells, such as: heating, bead beating and ultrasonication with lysis detergent  
145 such as Sodium Dodecyl Sulfate (SDS) and chaotropic agents such as urea [20, 25-30]. Many  
146 studies have reported that the use of SDS combined with mechanical disruption methods, such  
147 as bead beating or ultrasonication, provided better cell lysis yields than other buffers, in the  
148 case of gut microbial protein extraction [20,30,31].

149 Extracted proteins are then usually enzymatically digested into peptides, before or after the  
150 pre-fractionation step. The most frequently used enzyme is trypsin, because it generates many  
151 peptides, has great cleavage specificity and is easy to handle. The resulting peptides mostly  
152 have a molecular mass between 700 and 1500 Daltons, which is amenable to mass  
153 spectrometric analysis [32]. Nevertheless, other enzymes can be used alone or in combination  
154 with trypsin to enhance desired protein digestion effects [33].

155 After protein extraction, additional pre-concentration steps, such as filter-aided sample  
156 preparation (FASP) are often performed to obtain more concentrated peptides samples. This  
157 step allows a deeper coverage of metaproteomes [26]. Detergents and salt, which are  
158 commonly used during protein extraction, can interfere with mass spectrometric analysis,  
159 should be removed to the greatest extent possible during this step to increase analysis  
160 sensitivity [34].

161 In summary, differential centrifugation, enzymatic digestion (trypsin) and removal of  
162 detergents and salts play an important role during stool sample preparation for the  
163 metaproteomic characterization of the human gut microbiota. Table 1 summarizes the key  
164 metaproteomics studies and their sample preparation protocols. **Despite the heterogeneity of**  
165 **these different sample preparation and proteins extraction protocols, many groups use SDS**  
166 **and ultrasonication as a general procedure.**

## 167 **1.2 Pre-fractionation and Mass spectrometry**

### 168 **1.2.1 Pre-fractionation**

169 In the field of metaproteomics, it is advisable to analyse sample fractions containing  
170 fewer proteins and peptides in order to increase the sensitivity of low abundance peptides and  
171 to increase the proteomic depth of the analysis. However, analysis of complex and  
172 multispecies samples, such as those from the human gut, where the total number of microbial  
173 genes may vastly exceed the number of human genes, is complicated for a variety of reasons,  
174 including the wide dynamic range of microbial proteins present and the high levels of protein  
175 sequence homology. The separation of proteins can be obtained by gel electrophoresis, while  
176 peptides are generally separated using liquid chromatography. Gel electrophoresis can  
177 separate proteins along one or two dimensions. The gel containing the resulting protein bands  
178 or spots can then be cut and subjected to enzymatic digestion. Two-dimensional gel  
179 electrophoresis is predominantly dedicated to the study of highly expressed proteins. The first  
180 published work in the field of metaproteomics was performed with two-dimensional  
181 polyacrylamide gel electrophoresis (2D-PAGE) protein separation followed by matrix  
182 assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS)  
183 analysis [35]. Subsequently, the first metaproteomics study on human fecal microbiota was  
184 carried out using the same technique [36]. However, the disadvantages of this technique  
185 include significant sample handling and limited reproducibility. In addition, some proteins are

186 difficult to separate, such as those in low abundance, with high or low molecular weight, and  
187 particularly those of high hydrophobicity (e.g., membrane proteins) [37].

188 More recently, high performance liquid chromatography (HPLC) was used to separate tens of  
189 thousands of peptides from the enzymatic digestion of thousands of proteins. The main  
190 purpose of HPLC is to separate peptides so that fewer of them are entering the ionization  
191 source of the mass spectrometer at any one time. HPLC separates the compounds according to  
192 their affinity with a stationary phase and a mobile phase. In the field of metaproteomics,  
193 reverse phase (RP) liquid chromatography (LC) is the most commonly used technique for  
194 peptide separation due to its excellent resolving power, stability and ease of use [38,39].

195 Analytical columns are composed of a C18-grafted silica stationary phase (apolar) and of a  
196 mobile phase generally composed of two solvents (water and acetonitrile). Both solvents are  
197 combined to adjust the hydrophobicity of the mobile phase and, thus, separate the peptides  
198 according to their interaction with the column and their affinity with this mobile phase. The  
199 emergence of nano, or capillary, liquid chromatography columns and adapted systems is also  
200 interesting because of their ability to separate very small quantities of peptides without a  
201 decrease in sensitivity. Nevertheless, nano chromatography is challenging and has numerous  
202 drawbacks (e.g., less stable LC systems, low column capacity, tricky maintenance) [40].

203 Recently, micro LC has become increasingly favoured because of its ease of use compared to  
204 nano chromatography and the sensitivity gain over conventional HPLC [41]. Ultra-High  
205 Performance Liquid Chromatography (UHPLC) has become a standard hardware update  
206 necessary to achieve greater chromatographic selectivity [42]. Furthermore, liquid  
207 chromatography can also combine two or three orthogonal separation dimensions (2D-LC or  
208 2/3-phase MudPIT). The most frequently used configurations for the separation of peptides  
209 combines a strong cation exchange (SCX) column, usually in the first dimension, and a RP  
210 (C18) column in the last dimension [25, 43,44]. Another interesting 2D configuration

211 combining two columns in reverse phase with opposite pH values allows for increased  
212 identification when fractions are concatenated [45,46]. Thus, it enables a straightforward  
213 depth screening of the metaproteomes analysed. In summary, the interest of sample pre-  
214 fractionation is to be able to analyse less complex mixtures and, thus, to detect more peptides.  
215 However, **pre-fractionation** increases the number of analysis steps for a single sample, which  
216 can significantly extend the overall analysis time, **as well as greatly increase the cost** [38,47].

### 217 **1.2.2 Mass spectrometry**

218 Regarding peptide detection, tandem mass spectrometry is currently the preferred technique.  
219 It monitors the mass of the peptides and their induced fragments. **Generally, three main**  
220 **elements constitute a mass spectrometer: (i) an ion source, (ii) a mass analyzer and (iii) an ion**  
221 **current detector. The combination of two analyzers allows one to perform tandem**  
222 **spectrometry (MS/MS). This makes it possible to obtain structural information by**  
223 **fragmenting the parent ions and by measuring the fragment masses. The parent and fragment**  
224 **ions are detected by the detector and a mass spectrum is assembled. While there are several**  
225 **ionization techniques, the best for the analysis of biomolecules, such as proteins and peptides,**  
226 **are the soft ionization techniques: electrospray ionization (ESI) [134] and MALDI ionization**  
227 **(Matrix Assisted Laser Desorption Ionisation) [48]. Both techniques allow the ionization and**  
228 **transfer of intact biomolecules from the gas phase into the mass analyzer. Electrospray is an**  
229 **ionization technique that operates at atmospheric pressure, which makes it easy to couple with**  
230 **liquid chromatography. MALDI requires an overlay and co-crystallization the sample with an**  
231 **organic matrix and irradiation of the analyte-matrix crystal with a pulsed laser beam under**  
232 **vacuum.** In the field of metaproteomics concerning the human gut, the mass analyzers **most**  
233 **frequently used** are the tandem hybrid mass analysers that enable high resolution analysis:  
234 Quadrupole Time-Of-Flight (Q-TOF) [49], Linear Ion Trap/Fourier Transform Ion Cyclotron  
235 Resonance (LIT/FTICR) [50] and especially Linear Trapping Quadrupole-Orbitrap (LTQ-

236 Orbitrap) [51] or Quadrupole-Orbitrap (Q-Orbitrap). Each instrument offers a different mass  
237 accuracy, mass resolution, sensitivity or dynamic range. In any case, the purpose of mass  
238 spectrometry is to obtain sufficient selectivity and sensitivity to distinguish as many peptides  
239 as possible in complex samples [52]. In fact, sensitivity is important for the analysis of  
240 samples with limited quantities of peptide in order to increase the depth of metaproteome  
241 analysis at taxonomic and functional levels.

242 Currently, Ion mobility spectrometry (IMS) has been incorporated into a few instruments. Ion  
243 mobility is based on the separation of the molecular ions according to their mobility in a gas  
244 under the action of an electric field. The incorporation of ion mobility into MS/MS  
245 workflows allows an increase in selectivity. This enhanced selectivity can facilitate depth  
246 analysis for complex samples [53, 54]. Until now, the application of IMS for human gut  
247 metaproteomics has not yet been reported.

248 The MS/MS ion survey comprises three steps: i) selection of the peptide ion, ii) induced  
249 dissociation of the selected ion by collisions with an inert gas, iii) detection of the resulting  
250 ions. Two distinct MS/MS acquisition methods are used to collect peptide MS information:  
251 data dependent acquisition (DDA) or data independent acquisition (DIA). DDA mode is the  
252 most commonly used method in the field of shotgun proteomics due to its speed and  
253 sensitivity. In DDA, the precursors, usually the top 10–20 peptides per cycle, are sequentially  
254 selected from a full mass MS scan for fragmentation and acquisition in MS/MS mode. The  
255 selection parameters are pre-defined by the user. DDA mode generates fewer false positive  
256 rates because only the most intense peptides are fragmented. Nevertheless, it is a mode that  
257 often presents a loss of information especially in the case of weak peptide signals. Unlike the  
258 DDA mode, fragmentation of peptides is performed without pre-selection of the precursor  
259 during DIA [55]. DIA has recently been selected for a metaproteomic study on host-microbial  
260 interactions [56], but has not yet been used for the study of the human gut microbiota. Using

261 this approach, all peptide ions are fragmented in the collision cell and all the resulting  
262 fragment ions are then recorded with alternate scans. This acquisition mode allows the  
263 recording of MS/MS data of all peptide signals, which greatly reduces **information loss** [57].  
264 **However, with this acquisition mode, many fragments are non-informative. Generally, the**  
265 **high number of fragment ions generated complicates the analysis in a classical database**  
266 **search strategy. However, this problem can be solved by the use of a reference spectral**  
267 **library, previously generated by a thorough analysis of the same / similar samples by the**  
268 **DDA. Currently, DIA mode is preferentially used with Q-TOF mass spectrometers.**

269 In brief, research teams working in the field of metaproteomics, as applied to stool, have  
270 analysed their samples using LC-MS/MS systems. **Table 1** summarizes the key  
271 metaproteomic studies, their pre-concentrations and MS methods. **Most of the metaproteomic**  
272 **studies mentioned above use an Orbitrap as the mass analyzer with a DDA acquisition mode.**

### 273 **1.3. Metaproteomics data computation**

#### 274 **1.3.1 Conventional sequence database search**

275 The human gut microbiota is a complex environment and can be associated with a high  
276 number of protein sequences. On the other hand, mass spectrometry generates hundreds of  
277 thousands of peptide spectra that need to be compared with protein sequences. **The number of**  
278 **identified proteins, as well as the identified taxonomies and functional annotations result from**  
279 **protein database selection. Large databases searches (>10<sup>6</sup> sequences), such as NCBI or**  
280 **Uniprot/Trembl are a challenge for metaproteomics studies in terms of computation times and**  
281 **the large number of peptides sequences matches (PSM) [58,59].** Another important limitation  
282 with large databases concerns the evaluation of FDR, which **may lead to the rejection of true**  
283 **protein identifications [60].** The assessment of FDR is performed by “Target-Decoy” methods  
284 [61, 62,63]. FDR can then be calculated based on the matching scores. The sequence  
285 identifications are filtered according to the matching score to get an FDR lower than the

286 defined threshold (generally <1% on the peptides). However, strict filtering based on the FDR  
287 to avoid false-positive matches and the use of a restricted database would compromise the  
288 identification of microbial proteins by an increase of false negatives, thus limiting the number  
289 of peptide matches.

290 Protein databanks based on metagenomic data tend to get closer to the real protein content of  
291 samples. However, this approach may not provide a complete coverage of the protein content  
292 in the sample because there can be many different species in a single sample, for most of  
293 which a full genome is not available. Indeed, sequencing, assembly and annotation of the  
294 genome still generates incomplete or false metagenomic data [64]. Over the last few years,  
295 software has been developed in order to facilitate the automated analysis of high-throughput  
296 mass spectrometry-based proteomic data. Specialized algorithms have been implemented in  
297 research software to meet the requirements of mass spectrometry data. Generally, search  
298 engines assign a score to the peptide identifications [65]. The computation of this score differs  
299 between search algorithm. The most commonly used software programs in the field of  
300 proteomics are: Mascot [66], OMSSA [67], Sequest, X!Tandem [68], and ProteinLynx Global  
301 Server [69]. Table 1 describes the software used in the case of metaproteomics of the human  
302 gut. Despite the availability of software dedicated to metaproteomics, and the advances in  
303 DNA and RNA sequencing, metagenomes of the human gut usually contains hundreds of  
304 organisms with more than  $10^6$  proteins sequences. The determination of peptide sequence  
305 matches by searching against such databases could lead to an increased risk of false positives,  
306 but also a number of false-negative PSM. To address this challenge, metaproteomic studies  
307 based on iterative methods, where matches are derived from a primary search against a large  
308 database in order to create a smaller subset database, are increasing. The latter is called an in-  
309 house or customized database.

### 310 1.3.2 Customized iterative database approach

311 To meet **computational** challenges, customized databases are increasingly being used for  
312 protein identification. Taxonomic assignment is essential before any sequence search.  
313 Indeed, when searching for sequences, peptide sequences were often not proteotypic [70,71].  
314 It may, therefore, be difficult to distinguish certain organisms from a single peptide. In fact,  
315 shared peptides can be identified (potentially between homologous proteins between species).  
316 **Thereby, bioinformatic pipelines were developed to assign peptides to their lowest**  
317 **unambiguous taxonomic rank, using an implementation of the Lowest Common Ancestor**  
318 **(LCA) algorithm generating taxonomic profiles at different levels. These pipelines are**  
319 **generally divided into two distinct steps: a first step where the peptides are matched to a**  
320 **reference database containing complete bacterial genome sequences assembled from NCBI or**  
321 **a reference protein database such as Uniprot or from metagenomes; a second step where**  
322 **peptides that passed the script of the first step are assigned to taxonomic ranks in an**  
323 **interactive tree in which a given tryptic peptide occurs [72]. These pipelines allow the**  
324 **generation of complete taxonomic profiles and lists of species-unique peptides (i.e.,**  
325 **discriminative peptides).** Metagenomic taxonomy-guided research strategies are increasingly  
326 being used in metaproteomics to improve the construction of protein databases. These  
327 strategies interactively explore the taxonomical content of the data using an algorithm based  
328 on the LCA peptides in order to assign each peptide to a taxon [73]. Unipept is an open source  
329 web application using the LCA algorithm to determine the taxonomic specificity of peptides  
330 [72]. For example, *Tanca et al* used the Unipept taxonomic assignment to generate a  
331 customized "host-microbiome" database containing sequences from specific microbial taxa  
332 and the host [25]. Thereby, iterative workflows can be used to build specific databases of  
333 biological samples. The study conducted by *Xiao et al* showed that a metagenomic taxonomy-  
334 guided database search strategy allows the construction of databases able to provide high  
335 sensitivity and precision in peptide identification in metaproteomic studies [74]. This strategy

336 merges both taxonomy-guided reference protein sequences from public databases and  
337 metagenome assembly. *Zhang et al.* have also developed a universal workflow (MetaPro-IQ)  
338 to expand the sensitivity of peptide identification and greatly increase proteins identified for  
339 each sample [30]. A similar pipeline was used in the metaproteomics of saliva [75]. In brief,  
340 the implementation of Unipept algorithm became essential in the field of metaproteomics to  
341 prepare custom databases and simplify data processing pipelines.

### 342 **1.3.3 De novo sequencing search**

343 Otherwise, *de novo* peptide sequencing has become an alternative and complementary option  
344 for the assignment of peptide sequences to MS/MS spectra [76]. **Peptide de novo sequencing**  
345 **in the analytical process derives a peptide's amino acid sequence from its tandem mass**  
346 **spectrum without the assistance of sequence database. A clear advantage of de novo**  
347 **sequencing is that it works for both database and novel peptides.** For example, a study  
348 combining protein databases search and peptide *de novo* sequencing, showed, respectively,  
349 the identification of 421 theoretical sequences and 333 new non-redundant proteins from  
350 faecal samples. The new peptides could not be mapped to the metagenomic sequence data  
351 [44]. As such, search engines for taxonomic and functional analysis are challenged by the vast  
352 amount of unannotated sequences [77]. *De novo sequencing is often used to provide new*  
353 *unidentified sequences into databanks. This is possible thanks to a wide range of software*  
354 *tools.* The most commonly used *de novo* peptide sequencing software **programs** are: PEAKS  
355 [78], PepNovo [79], Novor [80], NovoHMM [81], UniNovo [82] and MSNovo [83]. *De novo*  
356 peptide sequences are searched against databases using the Basic Local Alignment Search  
357 Tool Protein (BLAST p) algorithm [84]. **However, during the process of de novo sequencing,**  
358 **some factors can cause difficulties including: incorrect assignment of ions, absence of ion**  
359 **fragments, existence of noise peaks in the spectrum, and post-translational modifications can**

360 contribute to the mass ambiguity and complicate the peptide fragmentation pattern. Moreover,  
361 the short length of tryptic peptides can impede MS-BLAST identification.

362 In summary, the use of customized databases with an iterative workflow should be  
363 encouraged in order to gain computational efficiency and focus on the protein content. This  
364 approach helps to reduce the rate of false identification associated with large databases and  
365 provides appropriate information. However, the processing of metagenomic data must be  
366 carried out carefully to ensure the best quality of the resulting metaproteomic databases.  
367 Besides, since databases generally do not cover all metaproteomes, *de novo* sequencing is  
368 highly useful for the detection of unknown peptide sequences directly from MS / MS spectra.

## 369 **2. Metaproteomics of the human gut microbiota**

### 370 **2.1 Exploration of the gut metaproteome**

371 A pioneering study of the human gut microbiota was conducted on two infants to investigate  
372 the functional role of gut microbiota during early growth [36]. However, despite the relatively  
373 simple faecal protein profile, the analysis was limited in depth due to the absence of an  
374 appropriate reference database. A few years later, with the development of analytical  
375 techniques and the availability of protein data from metagenomes, a study on the fecal  
376 microbiota of a preterm infant was performed. It revealed a much more detailed profile of the  
377 intestinal metaproteome and host microbiota interactions [85]. It will be interesting and  
378 valuable to collect more proteomic data that will allow a comparative study of microbial  
379 community functions between healthy preterm infants and those who develop diseases, such  
380 as neonatal necrotizing enterocolitis.

381 The first comprehensive intestinal metaproteome from a human adult was extracted from two  
382 healthy monozygotic twins [24]. This study explained an asymmetric and distinctive, but  
383 relatively stable distribution of proteins for each individual. The study also highlighted  
384 discrepancies between predicted protein levels from the metagenome and actual results. This

385 confirms the importance of metaproteomics in the understanding of proteins expression  
386 because several unknown proteins represented previously undescribed microbial pathways.  
387 Another study showed a highly comparable clustering of the metaproteomic and phylogenetic  
388 profiles at the phylum level. The study showed differences in the relative share of  
389 *Actinobacteria* [86]. Soon after, a comparative study was performed between one lean and  
390 one obese adolescent. Their fecal samples showed subject-specific metaproteome differences  
391 that correlated with compositional differences of the microbiota [8]. In the lean subject,  
392 proteins classified as *Bacteroidetes* were in high representation (81%), while according to  
393 metagenomics, this phylum represents only about 20% of the microbial community. In the  
394 obese subject, the total microbiota was more abundant in the phylum Firmicutes (94%) and  
395 protein expression was predominantly attributed (56 %). These previous studies show that  
396 metaproteomes provide complementary information about potentially active and functional  
397 bacteria in the gut microbial community. This study should be supported by further studies  
398 dealing with large cohorts of different unrelated individuals and alternative integrated omics  
399 approaches, such as metatranscriptomic and metabolomics in order to determine the metabolic  
400 links between obesity and gut microbiota.

401 In another context, metaproteomics could play an important role in the characterization of the  
402 gut microbial community in health and disease [87-90]. The number of taxa in fecal samples  
403 is estimated at more than 21,000 with more than 63,000,000 unique proteins [91]. Studies  
404 concerning bacterial phylotypes and their identification in relation to the host, therefore,  
405 remain a vast expanse to be explored. The composition of microbial communities has been  
406 studied by different methodologies, such as culture, microscopy and especially,  
407 metagenomics. At present, metaproteomics might have a considerable contribution to explore  
408 the diversity of the gut microbiota. It also provides new information, such as the description  
409 of new functional genes. Zhang et al, recently combined an efficient sample preparation

410 technique, high-resolution mass spectrometry and bioinformatics tools for the ultra-deep  
411 metaproteomic characterization of the human gut microbiome [92]. They reported the deepest  
412 analysis of the microbiome to date with an average of 20,558 protein groups identified per  
413 analyzed sample. Using an LCA approach with the Unipept tool, the taxonomic  
414 characterization of peptides pointed to 155 different microbe species with at least 3 distinctive  
415 peptides. This work also revealed variations in the microbiome from different individuals.  
416 However, because of the relatively long MS time for deep metaproteomics, this application  
417 for clinical samples analyses is limited.

418 In comparison to metagenomics, the study of the metaproteome for the characterization of  
419 microbial communities still has a long way to go. Nevertheless, in the field of microbial  
420 ecology, metaproteomics delivers a great amount of valuable data for in-depth analysis of  
421 microbiomes in response to human and microbial changes [93]. It appears as though a  
422 complementary approach to metagenomics, and a tool for large-scale taxonomic  
423 characterization of proteins in microbial ecosystems [94], could respond to diverse biologic  
424 questions concerning the host biology in health and disease. Rapid technical advances are  
425 expected and should focus on detection methods for protein modifications, which should  
426 reduce analysis cost and time. The integration of other omics platforms, such as  
427 metatranscriptomics metabolomics and culturomics could also allow in-depth study of diverse  
428 microbial communities at different pathological states.

## 429 **2.2 Gut microbiota in health and diseases**

430 A few years later, in addition to characterizing the microbial intestinal metaproteome of  
431 healthy subjects, comparative studies have increased in number. These studies were carried  
432 out to determine the expression of microbial proteins in case of gut dysbiosis. It has been  
433 suggested that an imbalance of the microbiota plays a central role in the chronic inflammation  
434 associated with the disease commonly named Inflammatory Bowel Disease (IBD). The first

435 study compared healthy and unhealthy adults and was based on Crohn's disease (CD) [95]. In  
436 this study, *Erikson et al.* combined shotgun metagenomics and metaproteomics to identify  
437 potential functional signatures of CD. Stool samples were collected from six twins, either  
438 healthy or affected by CD in the ileum or colon. The study revealed several genes of the  
439 microbial community, as well as microbial and human proteins, that differentiated CD from  
440 healthy subjects, including depletion of many proteins in CD in the ileum. Another study  
441 focused on host–microbe relationships in Inflammatory Bowel Disease and was performed  
442 through bacterial characterisation and metaproteomics analysis. It reported that the  
443 examination of relationships between the bacteria and metaproteomes allowed identification  
444 of a high frequency of 14 bacterial phylotypes that significantly differentiate human subjects  
445 by disease type, namely Crohn's disease and ulcerative colitis [96]. Furthermore, gut  
446 microbiota dysbiosis was reported in patients with cystic fibrosis. Fecal metaproteomics  
447 allows the analysis of host and microbial proteins to elucidate the functional changes resulting  
448 from this dysbiosis. For example, *Debyser et al* demonstrated that fecal protein from patients  
449 with cystic fibrosis were dominated by host proteins involved in inflammation and mucus  
450 formation [97]. Taxonomic analysis of the microbial proteins, based on LCA, confirmed  
451 significant differences in the gut microbial diversity with a strong reduction of butyrate  
452 reducers, such as *Faecalibacterium prausnitzii* and an increase of *Enterobacteriaceae*,  
453 *Ruminococcus gnavus* and *Clostridia* species. This study also highlights a list of host and  
454 microbial proteins that could be potential biomarkers for cystic fibrosis. So, metaproteomics  
455 enhances the understanding of the microbial world and establishes a link between microbial  
456 communities to its function. The functional distribution of COGs (clusters of orthologous  
457 groups) allows identification of responsible bacterial members of health status under altered  
458 physiological conditions revealing differential protein profiles. For example, the extraction of  
459 the metaproteome allows functional classification of bacterial proteins from a classification of

460 COG [86]. The shotgun metaproteomics approach has identified several COG categories that  
461 are more highly represented in the microbial metaproteome, compared to the average  
462 metagenome, in fecal samples from a female twin pair [24]. In this study, 50% of total  
463 proteins detected in the metaproteome were involved in translation, carbohydrate metabolism,  
464 or energy production. The other categories of COG were underrepresented in the  
465 metaproteomes, relative to metagenomes, including proteins involved in inorganic ion  
466 metabolism, cell wall and membrane biogenesis, cell division, and biosynthesis of secondary  
467 metabolites. Moreover, the best understanding of the study of gut microbiota function is to  
468 associate metaproteomics with COGs classification [121].

469 The bidirectional communication between the host and its microbiota is complex [98]. It  
470 involves a third partner, which is the immune system of the host, via innate immunity  
471 receptors. The immune system protects us against the constant aggressions of our  
472 environment and the gut microbiota plays an essential role in maintaining immunity.  
473 Influencing factors, such as stress, inappropriate diet, repetitive consumption of drugs and  
474 toxic substances can cause an imbalance of the microbiota or intestinal permeability. An  
475 important aspect to keep in mind is that the human microbiome is overly exposed to  
476 antibiotics that can rapidly alter its composition with potential immediate effects on health.  
477 Gut microbiota alterations induced by antibiotics can also indirectly affect health on long-  
478 term basis [99, 100]. The effects of antibiotic-induced microbiota alterations have an impact  
479 on the immune system and, therefore, cause an increased susceptibility to infections, inducing  
480 metabolic deregulation of the host. For example, a metaproteomic analysis study showed  
481 important changes in the protein profiles of the gut microbiota responses following  $\beta$ -lactam  
482 therapy [101]. The authors demonstrated that antibiotics targeting specific pathogenic  
483 infections and diseases may alter gut microbial ecology. Metaproteome results suggest the  
484 restoration of the microbiota indicating that the initial profile was recovered at the end of the

485 treatment. To date, one of the largest clinical metaproteomic studies on the human gut  
486 microbiota was conducted on acute leukaemia patients with multidrug-resistant  
487 *Enterobacteriaceae* gut colonization [102]. This study allowed the authors to describe the  
488 taxonomic composition and functional process of patients during the *Enterobacteriaceae* gut  
489 colonization. The analysis showed that public metagenome databases are incomplete and that  
490 sample-specific metagenomes improve results. This supports the idea that large database sizes  
491 come with several issues.

492 **Despite** the symbiosis between the host and the gut microbiota, major changes can affect the  
493 functionality of the microbiome. From this dysbiosis, an inappropriate immune response of  
494 the host may result. It is now suggested that pathologies related to disorders of immunity or  
495 metabolism can be triggered or aggravated by the bacteria that we host. For instance,  
496 alteration of the gut microbiota has been implicated in metabolic diseases, such as obesity and  
497 diabetes [103]. A metaproteomic and genomic study of the gut microbiota showed that  
498 microbial taxa associated with host proteins involved in the function of the mucus barrier **and**  
499 microvilli adhesion were depleted in patients with new-onset type 1 diabetes [104]. Recently,  
500 a study of the gut microbiota was correlated with liver cirrhosis. The authors used  
501 metaproteomics to detect proteome changes in the case of affected patients [105]. They found  
502 that the abundances of 14 proteins were increased **in the fecal microbiota from** liver cirrhosis  
503 **patients**. Seven proteins, such as ketol-acid reductoisomerase, phosphoglycerate kinase,  
504 ribose-phosphate pyrophosphokinase, **and probable thiol peroxidase** were **more highly**  
505 expressed in **patient's intestinal microbiota compared with normal**. These specific proteins  
506 can serve as potential biomarkers and therapeutic targets for the development of treatments.  
507 Furthermore, metaproteomic analysis of the gut microbiota has been increasingly applied to  
508 the identification of specific proteins as targets for treatment. Several pathologies or  
509 functional disorders have been linked with gut microbiota dysbiosis, such as Alzheimer's

510 disease, cardiovascular diseases, Parkinson's disease, depression and anxiety [98,106,107].  
511 The gut-metaproteome is a key element in maintaining the relationship between the host and  
512 the microbiota. Consequently, advances need to focus on the identification of human gut  
513 biomarkers. This could lead to the implementation of new clinical diagnostic tests and  
514 treatments to heal microbiota-related diseases.

### 515 **3. Metaproteomics combined with other Omics**

516 The human gut microbiota has been conceptualized as a dynamic ecological community  
517 consisting of several taxa, potentially interacting with each other, the host and the  
518 environment [108]. The fundamental objectives of human microbiome research focus on the  
519 various changes in the abundance and composition of the microbiota in relation to health and  
520 disease. Four key omics technologies are used to study the functions of cells: genomics for  
521 DNA, transcriptomics for RNA, proteomics for proteins, and metabolomics for small  
522 molecules/metabolites. To explore the dynamics of the microbial community, meta-omics  
523 approaches have been used to analyze large-scale gene or protein expressions and metabolite  
524 compositions [90]. Therefore, metagenomics, metatranscriptomics, metaproteomics and  
525 metabolomics are closely linked and metaproteomics plays a central role to more effectively  
526 decipher the composition and functions of microbial communities. Indeed, recent  
527 technological progress in the field of mass spectrometry and computational informatics has  
528 allowed metaproteomics to become a significant approach for the characterization of the  
529 human gut microbiome. The application of metaproteomics, combined with metagenomic  
530 analysis, has shown that the gut microbiome contains distinctive sets of active  
531 microorganisms between individuals [95]. However, the study of the relationship between  
532 taxonomic alterations and functional repercussions linked to the disease remains difficult. To  
533 resolve the taxonomic and functional attributes of gastrointestinal microbiota, *Heintz-*  
534 *Buschart et al.* combined data from genomics, metagenomics, metatranscriptomics and

535 metaproteomics, and showed that the associated microbial functional signatures were linked  
536 to metabolic traits in distinct taxa [109]. The use of multi-omics approaches would also  
537 identify small molecules and bacterial peptides affecting the physiology of the host, such as  
538 gastrointestinal motility induced by metabolites (e.g., CH<sub>4</sub>, H<sub>2</sub>, H<sub>2</sub>S, SFCA) from the  
539 microbiota, or deregulation of the microbiota-gut-brain in neurodegenerative diseases [110].  
540 Another multi-omics study provided novel insights into metabolic changes caused by  
541 antibiotic disturbance [101]. In this study, the integrative analysis showed an oscillatory  
542 imbalance between Gram-negative and Gram-positive bacteria after initiation of the  $\beta$ -lactam  
543 therapy. During this process, metabolic disorders associated with the different stages of the  
544 therapy were noted, such as an overall attenuation of energetic metabolism of gut bacteria and  
545 their capacity to transport and metabolize bile acid, cholesterol, hormones and vitamins.

546 In fact, the study of the function of the gut microbiota is better understood by combining  
547 metaproteomics with other OMICS approaches. Among these combined analyses, it is  
548 possible to identify potential genes, proteins and metabolic pathways that can be associated  
549 with a healthy condition. For example, the study of the functionality of the gut microbiota by  
550 combining multi-omics has shown a considerable divergence between potential functions and  
551 active expression in the gut microbiota of a healthy human cohort [111]. The authors found an  
552 overlap between the metagenome and the metaproteome regarding the most abundant phyla  
553 and genera. Nevertheless, they found considerable differences that highlight a divergence of  
554 microbial functions, especially with a carbohydrate metabolism.

555 Overall, analyses exploring the interactions between the intestinal microbiota and its functions  
556 towards humans would be more relevant if metaproteomics was included with other OMICS  
557 (Figure 2). The latest of the omics approaches is culturomics, developed at our institute [15].  
558 This in-depth study tested 212 culture conditions to select 18 best conditions for the isolation  
559 of prokaryotes. This approach was effectively combined to rapid identification of bacterial

560 colonies by MALDI TOF MS. Indeed, metaproteomics provides distinct and complementary  
561 microbial functional information to metagenomics and other approaches [112]. Each "omics"  
562 dataset can be studied separately, but relevant information can certainly be extracted from a  
563 joint analysis of several of them. The power of these integrative tools on the descriptive level  
564 allows today a deep and large-scale characterization of biological systems. However, these  
565 technologies face some explanatory limitations. Among the most frequently mentioned  
566 difficulties, these must be emphasized: data management requires powerful bioinformatics  
567 tools, formulation of objective hypotheses is required to comprehend biological systems; large  
568 sets of data obtained under different experimental conditions are difficult to compare;  
569 relationship between a molecular signature and the biological interpretation of biomarkers is  
570 not always obvious.

#### 571 4. Future directions for human gut metaproteomics research

##### 572 4.1 Application challenges

573 Despite the great potential of metaproteomics for characterizing microbial ecosystems and  
574 their various roles within the human host, many challenges remain. Metaproteomics should  
575 enable measurement of proteome expression for the entire microbial community at a specific  
576 moment in the gut microbial ecosystem. Beyond the difficulties encountered during protein  
577 extraction, analytical platforms also show limitations for a sensitivity analysis of protein  
578 samples in such a complex dynamic range. Thanks to the recent emerging mass spectrometry  
579 technologies, the depth of metaproteome analysis can be improved with a data independent  
580 acquisition mode [113]. The DIA mode usually results in an increased sensitivity and enables  
581 a significant improvement concerning reproducibility and quantification of proteins in  
582 complex samples, as compared to DDA [114]. Nevertheless, it is not widely used as a routine  
583 method because of technological drawbacks, such as the high number of clinical samples.  
584 These numerous samples require efficient tools, such as software that can handle very large

585 amounts of data with a computational time that does not exceed 24 hours. In addition to this  
586 analytical aspect, identifying the proteins of the complex microbial consortium, comprising  
587 hundreds or thousands of species, has also proved a difficult task. The absence of complete  
588 genomic sequences, particularly of poorly characterized and uncultivated species, is a major  
589 challenge for researchers. The availability of a relevant database is one of the essential  
590 elements of metaproteomics for the complete analysis of the gut microbiota samples. Some  
591 researchers use protein databases from completed genomes, while others use sequences from  
592 different metagenomes, complete or not, and from diverse origins [115]. A solution could be  
593 to use standardized metaproteomic databases from non-redundant and complete metagenomes  
594 for each type of microbial community [116]. All of this is intended to simplify and speed up  
595 computational analysis. Moreover, future improvements in software and algorithms will  
596 significantly contribute to the development of advanced meta-proteomic analyses  
597 [30,67,78,79,117].

#### 598 **4.2 Clinical diagnosis**

599 Advances in scientific research have shown the important role of intestinal flora in the  
600 regulation of many functions of the body, including the immune system. Changes in the  
601 composition of the intestinal microbiota (dysbiosis) are often due to environmental or dietary  
602 factors that can lead to chronic diseases (e.g., metabolic, inflammatory, cardiac). Therefore, it  
603 is important to consider the role of the gut microbiome when selecting therapy. Gut  
604 microbiota represents a variable factor and in case of a dysbiosis, the solution to prevent any  
605 pathogenesis of chronic diseases seems to be probiotics, prebiotics and diet. Furthermore,  
606 fecal transplantation (from healthy individuals) is used as a community replacement approach  
607 to restore the composition of intestinal flora (in particular for *Clostridium difficile* infection).  
608 The gut microbiota is, therefore, used as a tool for diagnosis and personalized treatment  
609 strategies [128].

610 **In this context**, metaproteomics research has already led to some remarkable discoveries about  
611 the functional **and** taxonomic characteristics of gut microbiota. **Considered as a tool to**  
612 **observe the consequences of the modulation of the intestinal flora**, metaproteomics could help  
613 **adapt a personalized treatment in cases of dysbiosis**, since it is a powerful tool for observing  
614 **modulation of the intestinal flora**. However, this is still an emerging area where an increase in  
615 the number of studies involving complex microbial communities is expected. The number of  
616 samples in clinical microbiology labs can reach hundreds or even thousands per day.  
617 Consequently, metaproteomics might be an application tool for the routine diagnosis of fecal  
618 samples, such as MALDI-TOF-MS, which became a standard tool in clinical microbiology  
619 laboratories [118]. Indeed, metaproteomics could help identify markers for clinical diagnosis  
620 and provide an overview of antigens, functions and taxa. However, not all the conditions are  
621 set at this time to allow routine metaproteomic analyses. Such a quantity of samples would  
622 require qualified personnel, standardization of **sample** preparation with a short processing  
623 time at low cost. One of the main challenges would also be the implementation of powerful  
624 and automated software. Moreover, software and the databases should conform to high-  
625 quality standards and specific privacy regulations for the medical applications.

#### 626 **4.3 Multi-omics contribution**

627 The study of metaproteomes helps to better understand the molecular interactions of the  
628 bacterial communities with the host [119]. Apart from protein identification, metaproteomics  
629 can determine the main microbial actors contributing to the gut metabolic functions  
630 [120,121]; this is not possible with metagenomics based on the 16S RNA. For instance, the  
631 advent of "culturomics", allowed culturing of many human microbial species that were **not**  
632 **previously culturable** [15, 122,123]. **The combination of culturomics and metagenomics,**  
633 **showed that both approaches are complementary, each providing data/results confirming the**  
634 **other approach, however also providing unique information [124]. For example, Li et al**

635 showed that the combination of culturomics with metaproteomics allowed a systemic  
636 understanding of the human microbiome thanks to the evaluation of the nutritional  
637 composition of the culture medium [125]. They demonstrated that the metaproteomic profile  
638 changed with the nutritional components of the culture medium. Therefore, metaproteomics  
639 has become a complementary approach to metagenomic data and other omics approaches. In  
640 summary, the combination of OMICS approaches allows an exhaustive understanding of the  
641 intestinal microbiota thanks to the complementarities of the results.  
642 As for metagenomics, rapid technical progress is needed in the field of metaproteomics in  
643 order to facilitate integration with other OMICS.

644 **Conclusion**

645 This review on the metaproteomics of the human gut microbiome shows a recent and  
646 powerful approach **that can be used** to characterize and better understand the human intestinal  
647 environment. Given the complexity of samples, metaproteomics of the human gut still faces  
648 several challenges, such as sample preparation, limitations of analytical tools and data  
649 interpretation. To date, major improvements and developments have made it possible to  
650 rigorously validate metaproteomic analyses, thanks to (i) optimised extraction, lysis and cell  
651 purification procedures, (ii) improved separation methods by liquid chromatography, and (iii)  
652 broader analysis of metaproteomes by rapid, accurate and sensitive mass spectrometry.  
653 Moreover, the availability of tailored sequence databases from high-quality metagenomics  
654 and development of bioinformatics tools, as well as efficient workflow pipelines, have  
655 improved the number of proteins identified.

656 Furthermore, the contribution of metaproteomic data to other meta-omics datasets provides an  
657 exhaustive and complementary view of the functional state of the intestinal microbiome.  
658 Altogether, metaproteomics is the cornerstone in the study of microbial ecosystems. It has  
659 great potential to become a valuable tool for routine diagnosis in clinical microbiology  
660 laboratories. **However, the multifaceted, diverse and complex metaproteomics approaches**  
661 **should be standardized to enable a more conclusive understanding of the function of the**  
662 **microbial communities in the human gut.**

663 **References**

664 1. Grice EA, Segre JA. The human microbiome: our second genome. *Annu Rev Genomics,*  
665 *Hum Genet.* 2012;13:151-70. doi: 10.1146/annurev-genom-090711-163814. Epub 2012, Jun  
666 6. Review.

667 2. Human Microbiome Project Consortium. Structure, function and diversity of the healthy  
668 human microbiome. *Nature.* 2012 Jun 13;486(7402):207-14. doi:10.1038/nature11234.

669 3. Macpherson AJ, Harris NL. Interactions between commensal intestinal bacteria and the  
670 immune system. *Nature Reviews Immunology.* 2004;4:478–85. doi:10.1038/nri1373.

671 4. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett C, Knight R, Gordon JI. The human  
672 microbiome project: exploring the microbial part of ourselves in a changing world. *Nature.*  
673 2007;449:804–10. doi: 10.1038/nature06244.

674 5. Snelson M, Coughlan MT. Dietary Advanced Glycation End Products: Digestion,  
675 Metabolism and Modulation of Gut Microbial Ecology. *Nutrients.* 2019 Jan 22;11(2). pii:  
676 E215. doi: 10.3390/nu11020215. Review.

677 6. Candeias EM, Sebastião IC, Cardoso SM, Correia SC, Carvalho CI, Plácido AI, Santos  
678 MS, Oliveira CR, Moreira PI, Duarte AI. Gut-brain connection: The neuroprotective effects  
679 of the anti-diabetic drug liraglutide. *World J Diabetes.*2015 Jun 25;6(6):807-27. doi:  
680 10.4239/wjd.v6.i6.807.

681 7. de Vos WM, de Vos EAJ. Role of the intestinal microbiome in health and disease: from  
682 correlation to causation. *Nutr Rev.* 2012 Aug;70 Suppl 1:S45-56. doi:10.1111/j.1753-  
683 4887.2012.00505.x. Review.

684 8. Ferrer M, Ruiz A, Lanza F, Haange S-B, Oberbach A, Till H, et al. Microbiota from the  
685 distal guts of lean and obese adolescents exhibit partial functional redundancy besides clear  
686 differences in community structure. *Environ Microbiol.* 2013 Jan;15(1):211-26. doi:  
687 10.1111/j.1462-2920.2012.02845.

688 9. Dicks LMT, Geldenhuys J, Mikkelsen LS, Brandsborg E, Marcotte H. Our gut microbiota:  
689 a long walk to homeostasis. *Benef Microbes.* 2018 Jan 29;9(1):3-20. doi:  
690 10.3920/BM2017.0066.

691 10. Hansen TH, Gøbel RJ, Hansen T, Pedersen O. The gut microbiome in cardio-metabolic  
692 health. *Genome Med.* 2015 Mar 31;7(1):33. doi: 10.1186/s13073-015-0157-z. eCollection  
693 2015. Review.

694 11. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, et al. Diversity  
695 of the human intestinal microbial flora. *Science.* 2005 Jun 10;308(5728):1635-8 doi:  
696 10.1126/science.1110591.

697 12. Mandal RS, Saha S, Das S. Metagenomic surveys of gut microbiota. *Genomics*  
698 *Proteomics Bioinformatics.* 2015 Jun;13(3):148-58. doi: 10.1016/j.gpb.2015.02.005. Epub  
699 2015 Jul 13. Review.

- 700 13. Jovel J, Patterson J, Wang W, Hotte N, O’Keefe S, Mitchel T, et al. Characterization of  
701 the Gut Microbiome Using 16S or Shotgun Metagenomics. *Front Microbiol.* 2016 Apr  
702 20;7:459. doi:10.3389/fmicb.2016.00459.
- 703 14. Gosalbes MJ, Durbán A, Pignatelli M, Abellan JJ, Jiménez-Hernández N, Pérez-Cobas  
704 AE, Latorre A, Moya A. Metatranscriptomic approach to analyze the functional human gut  
705 microbiota. *PLoS One.* 2011 Mar 8;6(3):e17447. doi: 10.1371/journal.pone.0017447.
- 706 15. Lagier J-C, Khelaifia S, Alou MT, Ndongo S, Dione N, Hugon P, et al. Culture of  
707 previously uncultured members of the human gut microbiota by culturomics. *Nat Microbiol.*  
708 2016 Nov 7;1:16203. doi: 10.1038/nmicrobiol.2016.203.
- 709 16. Schneider T, Riedel K. Environmental proteomics: analysis of structure and function of  
710 microbial communities. *Proteomics.* 2010 Feb;10(4):785-98. doi: 10.1002/pmic.200900450.  
711 Review.
- 712 17. Petriz BA, Franco OL. Metaproteomics as a Complementary Approach to Gut Microbiota  
713 in Health and Disease. *Front Chem.* 2017 Jan 26;5:4. doi: 10.3389/fchem.2017.00004.
- 714 18. Kolmeder CA, de Vos WM. Metaproteomics of our microbiome - developing insight in  
715 function and activity in man and model systems. *J Proteomics.* 2014 Jan 31;97:3-16. doi:  
716 10.1016/j.jprot.2013.05.018.
- 717 19. Lee PY, Chin S-F, Neoh H-M, Jamal R. Metaproteomic analysis of human gutmicrobiota:  
718 where are we heading? *J Biomed Sci.* 2017 Jun 12;24(1):36. doi:10.1186/s12929-017-0342-z.  
719 Review.
- 720 20. Zhang X, Li L, Mayne J, Ning Z, Stintzi A, Figeys D. Assessing the impact of protein  
721 extraction methods for human gut metaproteomics. *J Proteomics.* 2018 May 30;180:120-127.  
722 doi: 10.1016/j.jprot.2017.07.001.
- 723 21. Haange SB, Jehmlich N. Proteomic interrogation of the gut microbiota: potential clinical  
724 impact. *Expert Rev Proteomics.* 2016 Jun;13(6):535-7. doi: 10.1080/14789450.2016.1190652
- 725 22. Cañas B, Piñeiro C, Calvo E, López-Ferrer D, Gallardo JM. Trends in sample preparation  
726 for classical and second generation proteomics. *Journal of Chromatography A.*  
727 2007;1153:235–58. <https://doi.org/10.1016/j.chroma.2007.01.045>.
- 728 23. Morris LS, Marchesi JR. Assessing the impact of long term frozen storage of faecal  
729 samples on protein concentration and protease activity. *J Microbiol Methods.* 2016  
730 Apr;123:31-8. doi: 10.1016/j.mimet.2016.02.001.
- 731 24. Verberkmoes NC, Russell AL, Shah M, Godzik A, Rosenquist M, Halfvarson J, et al.  
732 Shotgun metaproteomics of the human distal gut microbiota. *ISME J.* 2009 Feb;3(2):179-89.  
733 doi:10.1038/ismej.2008.108.
- 734 25. Tanca A, Palomba A, Pisanu S, Addis MF, Uzzau S. Enrichment or depletion? The impact  
735 of stool pretreatment on metaproteomic characterization of the human gutmicrobiota.  
736 *Proteomics.* 2015 Oct;15(20):3474-85. doi: 10.1002/pmic.201400573.
- 737 26. Wiśniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method  
738 for proteome analysis. *Nat Methods.* 2009 May;6(5):359-62. doi:10.1038/nmeth.1322.

- 740 27. Deusch S, Seifert J. Catching the tip of the iceberg - evaluation of samplepreparation  
741 protocols for metaproteomic studies of the rumen microbiota. *Proteomics*. 2015  
742 Oct;15(20):3590-5. doi: 10.1002/pmic.201400556.
- 743 28. Chourey K, Jansson J, VerBerkmoes N, Shah M, Chavarria KL, Tom LM, et al. Direct  
744 cellular lysis/protein extraction protocol for soil metaproteomics. *J Proteome Res*. 2010 Dec  
745 3;9(12):6615-22. doi: 10.1021/pr100787q.
- 746 29. Tanca A, Palomba A, Pisanu S, Deligios M, Fraumene C, Manghina V, et al. A  
747 straightforward and efficient analytical pipeline formetaproteome characterization.  
748 *Microbiome*. 2014 Dec 10;2(1):49. doi:10.1186/s40168-014-0049-2.
- 749 30. Zhang X, Ning Z, Mayne J, Moore JI, Li J, Butcher J, et al. MetaPro-IQ: a universal  
750 metaproteomic approach to studying human and mouse gut microbiota. *Microbiome*. 2016 Jun  
751 24;4(1):31. doi: 10.1186/s40168-016-0176-z.
- 752 31. Chaiyarit S, Thongboonkerd V. Comparative analyses of cell disruption methods for  
753 mitochondrial isolation in high-throughput proteomics study. *Anal Biochem*. 2009 Nov  
754 15;394(2):249-58. doi: 10.1016/j.ab.2009.07.026.
- 755 32. Burkhart JM, Schumbrutzki C, Wortelkamp S, Sickmann A, Zahedi RP. Systematic and  
756 quantitative comparison of digest efficiency and specificity reveals the impact of trypsin  
757 quality on MS-based proteomics. *J Proteomics*. 2012 Feb 2;75(4):1454-62. doi:  
758 10.1016/j.jprot.2011.11.016.
- 759 33. Giansanti P, Tsiatsiani L, Low TY, Heck AJR. Six alternative proteases for mass  
760 spectrometry-based proteomics beyond trypsin. *Nat Protoc*. 2016 May;11(5):993-1006. doi:  
761 10.1038/nprot.2016.057.
- 762 34. Antharavally BS. Removal of detergents from proteins and peptides in a spin-column  
763 format. *Curr Protoc Protein Sci*. 2012 Aug;Chapter 6:Unit 6.12. doi:  
764 10.1002/0471140864.ps0612s69.
- 765 35. Wilmes P, Bond PL. The application of two-dimensional polyacrylamide gel  
766 electrophoresis and downstream analyses to a mixed community of prokaryotic  
767 microorganisms. *Environ Microbiol*. 2004;6:911-20. [https://doi.org/10.1111/j.1462-  
768 2920.2004.00687.x](https://doi.org/10.1111/j.1462-2920.2004.00687.x).
- 769 36. Klaassens ES, Vos WM de, Vaughan EE. Metaproteomics Approach To Study the  
770 Functionality of the Microbiota in the Human Infant Gastrointestinal Tract. *Appl Environ  
771 Microbiol*. 2007;73:1388-92. doi: 10.1128/AEM.01921-06.
- 772 37. Gygi SP, Corthals GL, Zhang Y, Rochon Y, Aebersold R. Evaluation of two-dimensional  
773 gel electrophoresis-based proteome analysis technology. *Proc Natl Acad Sci U S A*.  
774 2000;97:9390-5. <https://doi.org/10.1073/pnas.160270797>.
- 775 38. Bandhakavi S, Markowski TW, Xie H, Griffin TJ. Three-dimensional peptide  
776 fractionation for highly sensitive nanoscale LC-based shotgun proteomic analysis of complex  
777 protein mixtures. *Methods Mol Biol*. 2011;790:47-56. doi: 10.1007/978-1-61779-319-6\_4.

- 778 39. Josic D, Kovac S. Reversed-phase High Performance Liquid Chromatography of proteins.  
779 *Curr Protoc Protein Sci.* 2010;. <https://doi.org/10.1002/0471140864.ps0807s61>.
- 780 40. Gaspari M, Cuda G. Nano LC-MS/MS: a robust setup for proteomic analysis. *Methods*  
781 *Mol Biol.* 2011;790:115-26. doi: 10.1007/978-1-61779-319-6\_9.
- 782 41. Arnold DW, Needham SR. Micro-LC-MS/MS: the future of bioanalysis. *Bioanalysis.*  
783 2013 Jun;5(11):1329-31. doi: 10.4155/bio.13.31.
- 784 42. Walter TH, Andrews RW. Recent innovations in UHPLC columns and instrumentation.  
785 *TrAC Trends in Analytical Chemistry.* <https://doi.org/10.1016/j.trac.2014.07.016>.
- 786 43. Nägele E, Vollmer M, Hörth P, Vad C. 2D-LC/MS techniques for the identification of  
787 proteins in highly complex mixtures. *Expert Review of Proteomics.* 2004;1:37-46.  
788 <https://doi.org/10.1586/14789450.1.1.37>.
- 789 44. Cantarel BL, Erickson AR, VerBerkmoes NC, Erickson BK, Carey PA, Pan C, et al.  
790 Strategies for metagenomic-guided whole-community proteomics of complex microbial  
791 environments. *PLoS One.* 2011;6(11):e27173. doi: 10.1371/journal.pone.0027173.
- 792 45. Yang F, Shen Y, Camp DG, Smith RD. Reversed-phase chromatography with multiple  
793 fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics.*  
794 2011 May;11(10):2019-26. doi: 10.1002/pmic.201000722.
- 795 46. Gilar M, Olivova P, Daly AE, Gebler JC. Orthogonality of Separation in Two-  
796 Dimensional Liquid Chromatography. *Analytical Chemistry.* 2005;77:6426-34. doi:  
797 10.1021/ac050923i.
- 798 47. Righetti PG, Castagna A, Antonioli P, Boschetti E. Prefractionation techniques in  
799 proteome analysis: The mining tools of the third millennium. *ELECTROPHORESIS.*  
800 2005;26:297-319. <https://doi.org/10.1002/elps.200406189>.
- 801 48. Hillenkamp F, Karas M. Mass spectrometry of peptides and proteins by matrix-assisted  
802 ultraviolet laser desorption/ionization. *Methods Enzymol.*1990;193:280-95.  
803 [https://doi.org/10.1016/0076-6879\(90\)93420-P](https://doi.org/10.1016/0076-6879(90)93420-P)
- 804 49. Rietschel B, Baeumlisberger D, Arrey TN, Bornemann S, Rohmer M, Schuerken M, et al.  
805 The benefit of combining nLC-MALDI-Orbitrap MS data with nLC-MALDI-TOF/TOF data  
806 for proteomic analyses employing elastase. *J Proteome Res.* 2009 Nov;8(11):5317-24. doi:  
807 10.1021/pr900557k.
- 808 50. Makarov A, Denisov E, Kholomeev A, Balschun W, Lange O, Strupat K, et al.  
809 Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal Chem.*  
810 2006;78:2113-20. doi: 10.1021/ac0518811.
- 811 51. Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Graham Cooks R. The Orbitrap: a new  
812 mass spectrometer. *J Mass Spectrom.* 2005;40:430-43. <https://doi.org/10.1002/jms.856>.
- 813 52. Graham C, McMullan G, Graham RLJ. Proteomics in the microbial sciences.  
814 *Bioengineered Bugs.* 2011;2:17-30. <https://doi.org/10.4161/bbug.2.1.14413>.

- 815 53. Sans M, Feider CL, Eberlin LS. Advances in mass spectrometry imaging coupled to ion  
816 mobility spectrometry for enhanced imaging of biological tissues. *Curr Opin Chem Biol.* 2018  
817 Feb;42:138-146. doi: 10.1016/j.cbpa.2017.12.005.
- 818
- 819 54. Meier F, Brunner A-D, Koch S, Koch H, Lubeck M, Krause M, et al. Online Parallel  
820 Accumulation-Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass  
821 Spectrometer. *Mol Cell Proteomics.* 2018 Dec;17(12):2534-2545. doi:  
822 10.1074/mcp.TIR118.000900.
- 823 55. Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, et al. Targeted data extraction  
824 of the MS/MS spectra generated by data-independent acquisition: a new concept for  
825 consistent and accurate proteome analysis. *Mol Cell Proteomics.* 2012  
826 Jun;11(6):O111.016717. doi:10.1074/mcp.O111.016717.
- 827 56. Starr AE, Deeke SA, Li L, Zhang X, Daoud R, Ryan J, et al. Proteomic and  
828 Metaproteomic Approaches to Understand Host-Microbe Interactions. *Anal Chem.* 2018  
829 Jan2;90(1):86-109. doi: 10.1021/acs.analchem.7b04340.
- 830 57. Geromanos SJ, Vissers JPC, Silva JC, Dorschel CA, Li G-Z, Gorenstein MV, et al. The  
831 detection, correlation, and comparison of peptide precursor and product ions from data  
832 independent LC-MS with data dependant LC-MS/MS. *Proteomics.* 2009 Mar;9(6):1683-95.  
833 doi: 10.1002/pmic.200800562.
- 834 58. Jagtap P, Goslinga J, Kooren JA, McGowan T, Wroblewski MS, Seymour SL, Griffin TJ.  
835 A two-step database search method improves sensitivity in peptide sequence matches for  
836 metaproteomics and proteogenomics studies. *Proteomics.* 2013 Apr;13(8):1352-7. doi:  
837 10.1002/pmic.201200352.
- 838 59. Chatterjee S, Stupp GS, Park SK, Ducom JC, Yates JR 3rd, Su AI, Wolan DW. A  
839 comprehensive and scalable database search system for metaproteomics. *BMC Genomics.*  
840 2016 Aug 16;17(1):642. doi: 10.1186/s12864-016-2855-3.
- 841 60. Cargile BJ, Bundy JL, Stephenson JL. Potential for False Positive Identifications from  
842 Large Databases through Tandem Mass Spectrometry. *J Proteome Res.* 2004;3:1082-5. doi:  
843 10.1021/pr049946o.
- 844 61. Wang G, Wu WW, Zhang Z, Masilamani S, Shen R-F. Decoy methods for assessing false  
845 positives and false discovery rates in shotgun proteomics. *Anal Chem.* 2009 Jan 1;81(1):146-  
846 59. doi: 10.1021/ac801664q.
- 847 62. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale  
848 protein identifications by mass spectrometry. *Nature Methods.* 2007;4:207-14. doi:  
849 10.1038/nmeth1019.
- 850 63. Elias JE, Gygi SP. Target-decoy search strategy for mass spectrometry-based proteomics.  
851 *Methods Mol Biol.* 2010;604:55-71. doi: 10.1007/978-1-60761-444-9\_5.
- 852 64. Wooley JC, Godzik A, Friedberg I. A Primer on Metagenomics. *PLOS Computational*  
853 *Biology.* 2010; 6:e1000667. <https://doi.org/10.1371/journal.pcbi.1000667>.

- 854 65. Chepanoske CL, Richardson BE, von Rechenberg M, Peltier JM. Average peptide score: a  
855 useful parameter for identification of proteins derived from database searches of liquid  
856 chromatography/tandem mass spectrometry data. *Rapid Communications in Mass*  
857 *Spectrometry*. 2005;19:9–14. <https://doi.org/10.1002/rcm.1741>.
- 858 66. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification  
859 by searching sequence databases using mass spectrometry data. *Electrophoresis*.1999;20:3551–67.[https://doi.org/10.1002/\(SICI\)1522-](https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2)  
860 [2683\(19991201\)20:18<3551::AID-ELPS3551>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2)
- 862 67. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, et al. Open mass  
863 spectrometry search algorithm. *J Proteome Res*. 2004;3:958–64. doi: 10.1021/pr0499491
- 864 68. Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra.  
865 *Bioinformatics*. 2004;20:1466–7. <https://doi.org/10.1093/bioinformatics/bth092>
- 866 69. Prieto G, Aloria K, Osinalde N, Fullaondo A, Arizmendi JM, Matthiesen R. PAnalyzer: a  
867 software tool for protein inference in shotgun proteomics. *BMC Bioinformatics*. 2012 Nov  
868 5;13:288. doi: 10.1186/1471-2105-13-288.
- 869 70. Kuster B, Schirle M, Mallick P, Aebersold R. Scoring proteomes with proteotypic peptide  
870 probes. *Nature Reviews Molecular Cell Biology*. 2005;6:577–83. doi: 10.1038/nrm1683.
- 871 71. Boulund F, Karlsson R, Gonzales-Siles L, Johnning A, Karami N, Al-Bayati O, Åhrén C,  
872 Moore ERB, Kristiansson E. Typing and Characterization of Bacteria Using Bottom-up  
873 Tandem Mass Spectrometry Proteomics. *Mol Cell Proteomics*. 2017 Jun;16(6):1052-1063.  
874 doi: 10.1074/mcp.M116.061721.
- 875 72. Mesuere B, Willems T, Van der Jeugt F, Devreese B, Vandamme P, Dawyndt P. Unipept  
876 web services for metaproteomics analysis. *Bioinformatics*. 2016 Jun 1;32(11):1746-8. doi:  
877 10.1093/bioinformatics/btw039.
- 878 73. Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. *Genome*  
879 *Res*. 2007;17:377–86.
- 880 74. Xiao J, Tanca A, Jia B, Yang R, Wang B, Zhang Y, et al. Metagenomic Taxonomy-  
881 Guided Database-Searching Strategy for Improving Metaproteomic Analysis. *J Proteome Res*.  
882 2018 Apr 6;17(4):1596-1605. doi: 10.1021/acs.jproteome.
- 883 75. Jagtap P, McGowan T, Bandhakavi S, Tu ZJ, Seymour S, Griffin T, et al. Deep  
884 metaproteomic analysis of human salivary supernatant. *Proteomics*. 2012 Apr;12(7):992-  
885 1001. doi: 10.1002/pmic.201100503.
- 886 76. Ma B, Johnson R. De novo sequencing and homology searching. *Mol Cell Proteomics*.  
887 2012 Feb;11(2):O111.014902. doi: 10.1074/mcp.O111.014902.
- 888 77. Muth T, Renard BY, Martens L. Metaproteomic data analysis at a glance: advances in  
889 computational microbial community proteomics. *Expert Rev Proteomics*. 2016  
890 Aug;13(8):757-69. doi: 10.1080/14789450.2016.1209418.

891

- 892 78. Ma B, Zhang K, Hendrie C, Liang C, Li M, Doherty-Kirby A, et al. PEAKS: powerful  
893 software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun Mass*  
894 *Spectrom.* 2003;17:2337–42. <https://doi.org/10.1002/rcm.1196>.
- 895 79. Frank A, Pevzner P. PepNovo: de novo peptide sequencing via probabilistic network  
896 modeling. *Anal Chem.* 2005;77:964–73. doi: 10.1021/ac048788h.
- 897 80. Ma B. Novor: real-time peptide de novo sequencing software. *J Am Soc Mass Spectrom.*  
898 2015 Nov;26(11):1885-94. doi: 10.1007/s13361-015-1204-0.
- 899 81. Fischer B, Roth V, Roos F, Grossmann J, Baginsky S, Widmayer P, et al. NovoHMM: A  
900 Hidden Markov Model for de Novo Peptide Sequencing. *Anal Chem.* 2005;77:7265–73. doi:  
901 10.1021/ac0508853.
- 902 82. Jeong K, Kim S, Pevzner PA. UniNovo: a universal tool for de novo peptide sequencing.  
903 *Bioinformatics.* 2013 Aug 15;29(16):1953-62. doi: 10.1093/bioinformatics/btt338.
- 904 83. Mo L, Dutta D, Wan Y, Chen T. MSNovo: A Dynamic Programming Algorithm for de  
905 Novo Peptide Sequencing via Tandem Mass Spectrometry. *Anal Chem.* 2007;79:4870–8. doi:  
906 10.1021/ac070039n.
- 907 84. Shevchenko A, Sunyaev S, Loboda A, Shevchenko A, Bork P, Ens W, et al. Charting the  
908 Proteomes of Organisms with Unsequenced Genomes by MALDI-Quadrupole Time-of-Flight  
909 Mass Spectrometry and BLAST Homology Searching. *Analytical Chemistry.* 2001;73:1917–  
910 26. doi: 10.1021/ac0013709.
- 911 85. Young JC, Pan C, Adams R, Brooks B, Banfield JF, Morowitz MJ, et al. Metaproteomics  
912 reveals functional shifts in microbial and human proteins during a preterm infant gut  
913 colonization case. *Proteomics.* 2015 Oct;15(20):3463-73. doi:10.1002/pmic.201400563.
- 914 86. Kolmeder CA, de Been M, Nikkilä J, Ritamo I, Mättö J, Valmu L, et al. Comparative  
915 metaproteomics and diversity analysis of human intestinal microbiota testifies for its  
916 temporal stability and expression of core functions. *PLoS One.* 2012;7(1):e29913. doi:  
917 10.1371/journal.pone.0029913.
- 918 87. Mulle JG, Sharp WG, Cubells JF. The gut microbiome: a new frontier in autism research.  
919 *Curr Psychiatry Rep.* 2013 Feb;15(2):337. doi:10.1007/s11920-012-0337-0.
- 920 88. Mao L, Franke J. Symbiosis, dysbiosis, and rebiosis-the value of metaproteomics in  
921 human microbiome monitoring. *Proteomics.* 2015 Mar;15(5-6):1142-51. doi:  
922 10.1002/pmic.201400329.
- 923 89. Baothman OA, Zamzami MA, Taher I, Abubaker J, Abu-Farha M. The role of Gut  
924 Microbiota in the development of obesity and Diabetes. *Lipids Health Dis.* 2016 Jun  
925 18;15:108. doi: 10.1186/s12944-016-0278-4.
- 926 90. Xiao M, Yang J, Feng Y, Zhu Y, Chai X, Wang Y. Metaproteomic strategies and  
927 applications for gut microbial research. *Appl Microbiol Biotechnol.* 2017 Apr;101(8):3077-  
928 3088. doi: 10.1007/s00253-017-8215-7.
- 929 91. Wilmes P, Heintz-Buschart A, Bond PL. Microbiota from the distal guts of lean and obese  
930 adolescents exhibit partial functional redundancy besides clear differences in community

931 structure. *Environ Microbiol.* 2013 Jan;15(1):211-26. doi: 10.1111/j.1462-  
932 2920.2012.02845.x.

933 92. Zhang X, Chen W, Ning Z, Mayne J, Mack D, Stintzi A, et al. Deep Metaproteomics  
934 Approach for the Study of Human Microbiomes. *Anal Chem.* 2017 Sep 5;89(17):9407-9415.  
935 doi: 10.1021/acs.analchem.7b02224.

936 93. Armengaud J. Microbiology and proteomics, getting the best of both worlds! *Environ*  
937 *Microbiol.* 2013 Jan;15(1):12-23. doi: 10.1111/j.1462-2920.2012.02811.x.Epub 2012 Jun 19.

938 94. Guirro M, Costa A, Gual-Grau A, Mayneris-Perxachs J, Torrell H, Herrero P, et al. Multi-  
939 omics approach to elucidate the gut microbiota activity: Metaproteomics and metagenomics  
940 connection. *Electrophoresis.* 2018 Jul;39(13):1692-1701. doi: 10.1002/elps.201700476.

941 95. Erickson AR, Cantarel BL, Lamendella R, Darzi Y, Mongodin EF, Pan C, et al.  
942 Integrated metagenomics/metaproteomics reveals human host-microbiota signatures of  
943 Crohn's disease. *PLoS One.* 2012;7(11):e49138. doi:10.1371/journal.pone.0049138.

944 96. Presley LL, Ye J, Li X, Leblanc J, Zhang Z, Ruegger PM, et al. Host-microbe  
945 relationships in inflammatory bowel disease detected by bacterial and metaproteomic analysis  
946 of the mucosal-luminal interface. *Inflamm Bowel Dis.* 2012 Mar;18(3):409-17. doi:  
947 10.1002/ibd.21793.

948 97. Debyser G, Mesuere B, Clement L, Van de Weygaert J, Van Hecke P, Duytschaever G, et  
949 al. Faecal proteomics: A tool to investigate dysbiosis and inflammation in patients with cystic  
950 fibrosis. *J Cyst Fibros.* 2016 Mar;15(2):242-50. doi: 10.1016/j.jcf.2015.08.003.

951 98. Valles-Colomer M, Falony G, Darzi Y, Tigchelaar EF, Wang J, Tito RY, et al. The  
952 neuroactive potential of the human gut microbiota in quality of life and depression. *Nat*  
953 *Microbiol.* 2019 Feb 4. doi:10.1038/s41564-018-0337-x.

954 99. Francino MP. Antibiotics and the Human Gut Microbiome: Dysbioses and Accumulation  
955 of Resistances. *Front Microbiol.* 2016 Jan 12;6:1543. doi:10.3389/fmicb.2015.01543.

956 100. Jernberg C, Löfmark S, Edlund C, Jansson JK. Long-term ecological impacts of  
957 antibiotic administration on the human intestinal microbiota. *The ISME Journal.* 2007;1:56-  
958 66. doi: 10.1038/ismej.2007.3.

959 101. Pérez-Cobas AE, Gosalbes MJ, Friedrichs A, Knecht H, Artacho A, Eismann K, et al.  
960 Gut microbiota disturbance during antibiotic therapy: a multi-omic approach. *Gut.* 2013  
961 Nov;62(11):1591-601. doi: 10.1136/gutjnl-2012-303184.

962 102. Rechenberger J, Samaras P, Jarzab A, Behr J, Frejno M, Djukovic A, Sanz J, González-  
963 Barberá EM, Salavert M, López-Hontangas JL, Xavier KB, Debrauwer L, Rolain JM, Sanz  
964 M, Garcia-Garcera M, Wilhelm M, Ubeda C, Kuster B. Challenges in Clinical  
965 Metaproteomics Highlighted by the Analysis of Acute Leukemia Patients with Gut  
966 Colonization by Multidrug-Resistant Enterobacteriaceae. *Proteomes.* 2019 Jan 8;7(1). pii: E2.  
967 doi: 10.3390/proteomes7010002

968 103. Meijnikman AS, Gerdes VE, Nieuwdorp M, Herrema H. Evaluating Causality of Gut  
969 Microbiota in Obesity and Diabetes in Humans. *Endocr Rev.* 2018 Apr 1;39(2):133-153. doi:  
970 10.1210/er.2017-00192.

- 971 104. Gavin PG, Mullaney JA, Loo D, Cao K-AL, Gottlieb PA, Hill MM, et al. Intestinal  
972 Metaproteomics Reveals Host-Microbiota Interactions in Subjects at Risk for Type 1  
973 Diabetes. *Diabetes Care*. 2018 Oct;41(10):2178-2186. doi: 10.2337/dc18-0777.
- 974 105. Wei X, Jiang S, Chen Y, Zhao X, Li H, Lin W, et al. Cirrhosis related functionality  
975 characteristic of the fecal microbiota as revealed by a metaproteomic approach. *BMC*  
976 *Gastroenterol*; <https://doi.org/10.1186/s12876-016-0534-0>.
- 977 106. Mancuso C, Santangelo R. The long way between preclinical studies and clinical  
978 evidence. *Pharmacol Res*. 2018 Mar;129:329-336. doi: 10.1016/j.phrs.2017.12.009.
- 979 107. Yoshida N, Yamashita T, Hirata K. Gut Microbiome and Cardiovascular Diseases.  
980 *Diseases*. 2018 Jun 29;6(3). pii: E56. doi: 10.3390/diseases6030056.
- 981 108. Foxman B, Goldberg D, Murdock C, Xi C, Gilsdorf JR. Conceptualizing human  
982 microbiota: from multicelled organ to ecological community. *Interdiscip Perspect Infect Dis*.  
983 2008;2008:613979. doi: 10.1155/2008/613979.
- 984 109. Heintz-Buschart A, May P, Laczny CC, Lebrun LA, Bellora C, Krishna A, et al.  
985 Integrated multi-omics of the human gut microbiome in a case study of familial type 1  
986 diabetes. *Nat Microbiol*. 2016 Oct 10;2:16180. doi: 10.1038/nmicrobiol.2016.180
- 987 110. Reigstad CS, Kashyap PC. Beyond phylotyping: understanding the impact of  
988 gutmicrobiota on host biology. *Neurogastroenterol Motil*. 2013 May;25(5):358-72.  
989 doi:10.1111/nmo.12134.
- 990 111. Tanca A, Abbondio M, Palomba A, Fraumene C, Manghina V, Cucca F, et al. Potential  
991 and active functions in the gut microbiota of a healthy human cohort. *Microbiome*. 2017 Jul  
992 14;5(1):79. doi: 10.1186/s40168-017-0293-3.
- 993 112. Del Chierico F, Gnani D, Vernocchi P, Petrucca A, Alisi A, Dallapiccola B, et al. Meta-  
994 omic platforms to assist in the understanding of NAFLD gut microbiota alterations: tools and  
995 applications. *Int J Mol Sci*. 2014 Jan 7;15(1):684-711. doi: 10.3390/ijms15010684.
- 996 113. Heyer R, Schallert K, Zoun R, Becher B, Saake G, Benndorf D. Challenges and  
997 perspectives of metaproteomic data analysis. *J Biotechnol*. 2017 Nov 10;261:24-36. doi:  
998 10.1016/j.jbiotec.2017.06.1201.
- 999 114. Sajic T, Liu Y, Aebersold R. Using data-independent, high-resolution mass spectrometry  
1000 in protein biomarker research: perspectives and clinical applications. *Proteomics Clin Appl*.  
1001 2015 Apr;9(3-4):307-21. doi:10.1002/prca.201400117.
- 1002 115. Pible O, Armengaud J. Improving the quality of genome, protein sequence, and  
1003 taxonomy databases: a prerequisite for microbiome meta-omics 2.0. *Proteomics*.  
1004 2015;15:3418–23. <https://doi-org.lama.univ-amu.fr/10.1002/pmic.201500104>.
- 1005 116. Tanca A, Palomba A, Fraumene C, Pagnozzi D, Manghina V, Deligios M, et al. The  
1006 impact of sequence database choice on metaproteomic results in gut microbiota studies.  
1007 *Microbiome*. 2016;4:51. <https://doi.org/10.1186/s40168-016-0196-8>.
- 1008 117. Mesuere B, Van der Jeugt F, Devreese B, Vandamme P, Dawyndt P. The unique  
1009 peptidome: Taxon-specific tryptic peptides as biomarkers for targeted metaproteomics.  
1010 *Proteomics*. 2016 Sep;16(17):2313-8. doi: 10.1002/pmic.201600023.

- 1011 118. Seng P, Rolain J-M, Fournier PE, La Scola B, Drancourt M, Raoult D. MALDI-TOF-  
1012 mass spectrometry applications in clinical microbiology. *Future Microbiol.* 2010  
1013 Nov;5(11):1733-54. doi: 10.2217/fmb.10.127.
- 1014
- 1015
- 1016 119. Haange S-B, Jehmlich N. Proteomic interrogation of the gut microbiota: potential  
1017 clinical impact. *Expert Rev Proteomics.* 2016 Jun;13(6):535-7.  
1018 doi:10.1080/14789450.2016.1190652.
- 1019 120. Mayers MD, Moon C, Stupp GS, Su AI, Wolan DW. Quantitative Metaproteomics and  
1020 Activity-Based Probe Enrichment Reveals Significant Alterations in Protein Expression from  
1021 a Mouse Model of Inflammatory Bowel Disease. *J Proteome Res.* 2017Feb 3;16(2):1014-  
1022 1026. doi: 10.1021/acs.jproteome.6b00938.
- 1023 121. Hagen LH, Frank JA, Zamanzadeh M, Eijsink VGH, Pope PB, Horn SJ, et al.  
1024 Quantitative Metaproteomics Highlight the Metabolic Contributions of Uncultured  
1025 Phylotypes in a Thermophilic Anaerobic Digester. *Appl Environ Microbiol.* 2016 Dec  
1026 30;83(2). pii: e01955-16. doi: 10.1128/AEM.01955-16.
- 1027 122. Lagier J-C, Armougom F, Million M, Hugon P, Pagnier I, Robert C, et al. Microbial  
1028 culturomics: paradigm shift in the human gut microbiome study. *Clin Microbiol Infect.* 2012  
1029 Dec;18(12):1185-93. doi: 10.1111/1469-0691.12023.
- 1030 123. Lagier J-C, Hugon P, Khelaifia S, Fournier P-E, La Scola B, Raoult D. The rebirth of  
1031 culture in microbiology through the example of culturomics to study human gut microbiota.  
1032 *Clin Microbiol Rev.* 2015 Jan;28(1):237-64. doi: 10.1128/CMR.00014-14.
- 1033 124. Lagier J-C, Dubourg G, Million M, Cadoret F, Bilen M, Fenollar F, et al Culturing the  
1034 human microbiota and culturomics. *Nat Rev Microbiol.* 2018 May 1:540-550. doi:  
1035 10.1038/s41579-018-0041-0.
- 1036 125. Li L, Zhang X, Ning Z, Mayne J, Moore JI, Butcher J, et al. Evaluating in Vitro Culture  
1037 Medium of Gut Microbiome with Orthogonal Experimental Design and a Metaproteomics  
1038 Approach. *J Proteome Res.* 2018 Jan 5;17(1):154-163. doi: 10.1021/acs.jproteome.7b00461.
- 1039
- 1040
- 1041
- 1042
- 1043 126. Rooijers K, Kolmeder C, Juste C, Doré J, de Been M, Boeren S, et al. An iterative  
1044 workflow for mining the human intestinal metaproteome. *BMC Genomics.* 2011 Jan 5;12:6.  
1045 doi:10.1186/1471-2164-12-6.
- 1046 127. Juste C, Kreil DP, Beauvallet C, Guillot A, Vaca S, Carapito C, et al. Bacterial protein  
1047 signals are associated with Crohn's disease. *Gut.* 2014 Oct;63(10):1566-77. doi:  
1048 10.1136/gutjnl-2012-303786.

- 1049 128. Kashyap PC, Chia N, Nelson H, Segal E, Elinav E. Microbiome at the Frontier of  
1050 Personalized Medicine. *Mayo Clin Proc.* 2017 Dec;92(12):1855-1864. doi:  
1051 10.1016/j.mayocp.2017.10.004
- 1052 129. Xiong W, Giannone RJ, Morowitz MJ, Banfield JF, Hettich RL. Development of an  
1053 enhanced metaproteomic approach for deepening the microbiome characterization of the  
1054 human infant gut. *J Proteome Res.* 2015 Jan 2;14(1):133-41. doi:10.1021/pr500936p.
- 1055 130. Kolmeder CA, Ritari J, Verdam FJ, Muth T, Keskitalo S, Varjosalo M, et al. Colonic  
1056 metaproteomic signatures of active bacteria and the host in obesity. *Proteomics.* 2015  
1057 Oct;15(20):3544-52. doi: 10.1002/pmic.201500049.
- 1058 131. Zhang X, Ning Z, Mayne J, Deeke SA, Li J, Starr AE, et al. In Vitro Metabolic Labeling  
1059 of Intestinal Microbiota for Quantitative Metaproteomics. *Anal Chem.* 2016 Jun  
1060 21;88(12):6120-5. doi: 10.1021/acs.analchem.6b01412.
- 1061 132. Kolmeder CA, Salojärvi J, Ritari J, de Been M, Raes J, Falony G, et al. Faecal  
1062 Metaproteomic Analysis Reveals a Personalized and Stable Functional Microbiome and  
1063 Limited Effects of a Probiotic Intervention in Adults. *PLoS One.* 2016 Apr  
1064 12;11(4):e0153294. doi: 10.1371/journal.pone.0153294.
- 1065 133. Li X, LeBlanc J, Truong A, Vuthoori R, Chen SS, Lustgarten JL, et al. A metaproteomic  
1066 approach to study human-microbial ecosystems at the mucosal luminal interface. *PLoS One.*  
1067 2011;6(11):e26542. doi: 10.1371/journal.pone.0026542.  
1068
- 1069 134. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for  
1070 mass spectrometry of large biomolecules. *Science.* 1989 Oct 6;246(4926):64-71. Review.  
1071 DOI: 10.1126/science.2675315

1072 **Table 1:** Metaproteomics studies of the human gut (classified in a chronological order)

<b>Samples (Feces)</b>	<b>Sample preparation</b>	<b>Pre- fractionation</b>	<b>Mass spectrometry</b>	<b>Database</b>	<b>Search engines</b>	<b>Results of analysis  Number of validated proteins (&gt; n peptides)</b>	<b>References</b>
Childs N=2	Chemical and mechanical lysis  Tryptic digestion	2D PAGE,	MALDI-TOF-  MS/MS	NCBI  Swissprot  Uniprot KB	PDQuest	> 200 spots 1 protein  identified	[36]
Healthy monozygotic twin N=2	Differential centrifugation  Chemical and mechanical lysis  Trypsin digestion	Nano2D-LC  MudPIT RPC18-  SCX-RPC18	LTQ Orbitrap; DDA	In-house database :  db1 and metadb	SEQUEST	600-900 proteins per sample and replicate (db1)  970-1340 proteins per sample and replicate (metadb)	[24]
Healthy adult N=2	Chemical lysis (urea, thiourea)  Tryptic digestion	SDS-PAGE,  nanoLC-RPC18	LTQ Orbitrap; DDA	In-house database  from genomics	OMSSA	2331 and 1870 peptides  1120 and 922 peptides	[132]
Healthy human N=2	Differential centrifugation,	SDS-PAGE,  Nano2D-LC	LTQ Orbitrap; DDA	In-house database	SEQUEST  PepNovo+	5233 proteins (2 peptides)  6186 proteins ( $\geq$ 1 peptide)	[44]

	Direct extraction Chemical and mechanical lysis Trypsin digestion	MudPIT RPC18- SCX-RPC18			PEAKS	3706 proteins ( $\geq 2$ peptides)	
Healthy adult N=3	Mechanical lysis: PBS+zirconium-silica beads Tryptic digestion	SDS-PAGE Nano2D-LC RPC18	LTQ Orbitrap; DDA	In-house databases	OMSSA	1790 microbial proteins ( $>2$ peptides)	[86]
Adolescents N=2 1 lean (female) and 1 obese (male)	Differential Centrifugation Mechanical lysis (sonication) Tryptic digestion	SDS-PAGE nanoUPLC-C18	LTQ Orbitrap; DDA	Matched metagenomes unmatched metagenomes	Maxquant	613 proteins ( $>2$ peptides)	[8]
Patients with CD N =6 (4 women and 2 men) Healthy Controls N=6	Ultracentrifugation Chemical and mechanical lysis Tryptic digestion	2D-DIGE; nanoLC-C18	LTQ Orbitrap; DDA LTQ-Linear Ion Trap; DDA	MetaHit database, Human SwissProt, In-house contaminant database	X!Tandem;	141 proteins spots 89 bacterial proteins spots	[127]
Child N=1	Chemical and mechanical lysis	nano2D-LC : SCX ; RPC18	LTQ Orbitrap; DDA	in-house database from genomics	SEQUEST	and 4,031 proteins ( $> 1$ peptides)	[85]

	Tryptic digestion						
Healthy volunteer N=1	Direct extraction(DE) Differential Centrifugation (DC) Chemical and mechanical lysis Tryptic digestion	nanoLC: RPC18	LTQ Orbitrap; DDA	UniProtKB, SwissProt, customized host- microbiome Unipept	SEQUEST	- DE:, 3911 proteins - DC: 4587 proteins	[25]
Premature infants N=2	Direct extraction Differential filtering Chemical and mechanical lysis Tryptic digestion	nano2D-LC SCX; RPC18	LTQ Orbitrap; DDA	Customized database	Myrimatch	807 proteins groups (DE) 1264 proteins groups (DF) (1 unique peptide per proteins)	[129]
Healthy individuals N=29 (9 normal, 4 overweigh)	Mechanical lysis (bead beating)	SDS-PAGE; nanoHPLC: C18	Q-Orbitrap; DDA	In-house human intestinal	OMSSA X!Tandem	91.86% human and microbial proteins	[130]

and 16 obese)	Tryptic digestion			metaproteome database (HIMPdb) Unipept		73.90% assigned to <i>Bacteroidetes</i> phylum (obese)	
Children with cystic fibrosis their unaffected siblings N=30	Differential centrifugation Chemical lysis Tryptic digestion	SDS-PAGE nanoLC-C18	Linear Ion trap- FTICR; DDA	NCBI Unipept	Mascot X!Tandem	1,676 proteins 495 unique to patients 793 unique to siblings (≥4 peptides)	[93]
Mucosal lavage from distal colon of different adolescent N=5	Differential centrifugation; Chemical lysis Tryptic digestion	SDS-PAGE ; LC- RPC18	LTQ Orbitrap; DDA	In-house Database (HIMPD);Target- decoy database Unipept	X!Tandem	4 014 protein groups (≥2 unique peptides)	[131]
Healthy adults N=16 (8 probiotic and 8 placebo)	Mechanical lysis (bead beating)  Tryptic digestion	SDS-PAGE Nano2D-LC:C18	LTQ Orbitrap; DDA	In-house metaproteome database	Mascot	66 % identified peptides with LCA: 80,9% bacteria 1% Archaea 13.8% Eukaryotic 5.3% could not be assigned	[132]
Children with IBD N=4	Differential centrifugation; Chemical and mechanical lysis	nanoLC-C18	Q-Orbitrap; DDA	Human gut gene catalog; human proteome In-house database	Maxquant,	: 20 558 protein groups (>2 peptides)	[88]

	Tryptic digestion			Unipept			
Mucosal lavage from healthy subjects N=38 (205 lavage samples)	Differential centrifugation (DC), Mechanical lysis (magnetic beads) Tryptic digestion	nano2D-LC: RPC18	LTQ Orbitrap; DDA	SwissProt (human and bacteria)	SEQUEST	117 unique proteins: 63% human proteins 30% bacterial proteins 7% others	[133]
56 patients with Gut Colonization by Multidrug-Resistant <i>Enterobacteriaceae</i> (N=212 stool samples)	Centrifugation Chemical and mechanical lysis Tryptic digestion	SDS-PAGE 2D- UHPLC RPC18	Hybrid quadrupole Orbitrap ; DDA	-Genome Reference Catalog - SWISS-PROT bacteria and human - Metagenomes	-Maxquant -Unipept	-60% of the identified peptides to a taxonomy level -80% of the peptides mapped to at least one Gene Ontology term	[102]

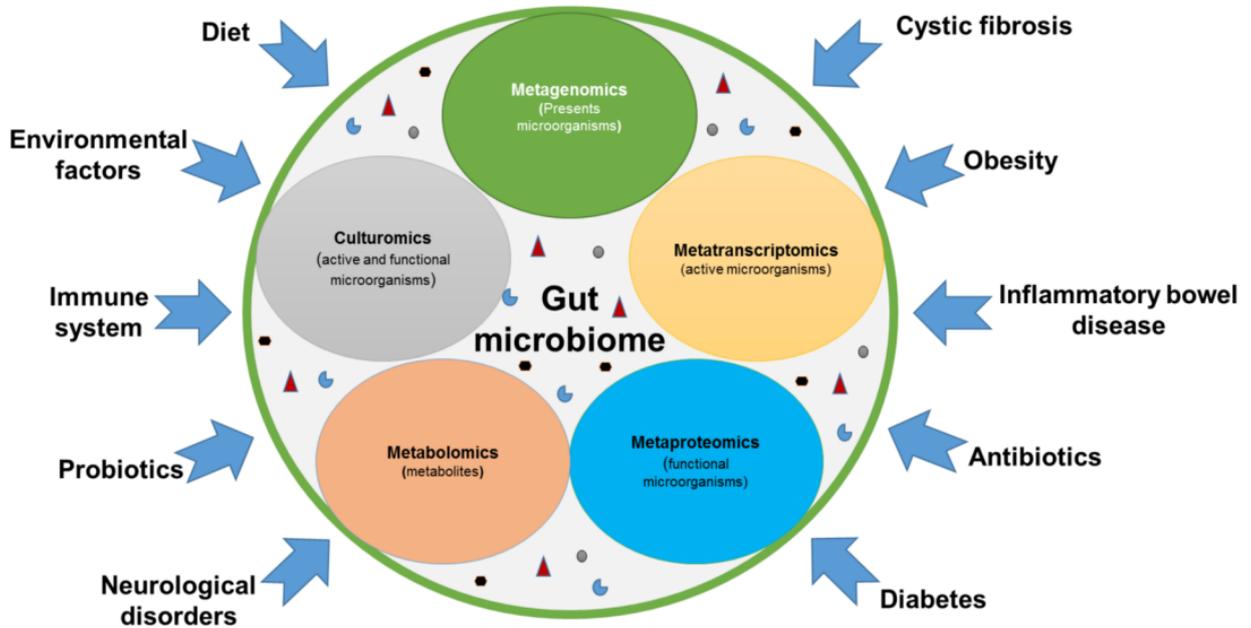
1073

1074 **List of figures**

1075 **Figure 1:** Understanding human gut microbiome: different omics that are involved and the  
1076 factors influencing microbiota.

1077 **Figure 2:** Representation of a typical workflow in a metaproteomic analysis of fecal sample

1078



➔ Influencing factors of gut microbiota modulation


 Microorganisms

## Stool sample preparation



Fecal sample

- Differential centrifugation
- Cell lysis
- Protein extraction
- Enzymatic digestion
- Sample clean-up

## Liquid Chromatography Mass Spectrometry

- Pre-fractionation LC  
SCX, RPC18
- Analytical LC  
RPC18
- Mass Spectrometry  
Orbitrap (DDA); Q-TOF (DDA or DIA)

## Integrative multi-Omics

- Metagenomics
- Metatranscriptomics
- Metabolomics
- Culturomics

## Data computation

