



**HAL**  
open science

## Advances in the integration of metabolomics and metagenomics for human gut microbiome and their clinical applications

Francesc Puig-Castellví, Romina Pacheco-Tapia, Maxime Deslande, Manyi Jia, Petros Andrikopoulos, Kanta Chechi, Amélie Bonnefond, Philippe Froguel, Marc-Emmanuel Dumas

### ► To cite this version:

Francesc Puig-Castellví, Romina Pacheco-Tapia, Maxime Deslande, Manyi Jia, Petros Andrikopoulos, et al.. Advances in the integration of metabolomics and metagenomics for human gut microbiome and their clinical applications. Trends in Analytical Chemistry, 2023, 167, pp.117248. 10.1016/j.trac.2023.117248 . hal-04442375

**HAL Id: hal-04442375**

**<https://hal.science/hal-04442375v1>**

Submitted on 9 Feb 2024

**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.

## **Advances in the integration of metabolomics and metagenomics for human gut microbiome and their clinical applications**

Francesc Puig-Castellvi<sup>1,2,\*†</sup>, Romina Pacheco-Tapia<sup>1,2,†</sup>, Maxime Deslande<sup>1</sup>, Manyi Jia<sup>2</sup>, Petros Andrikopoul<sup>2,3</sup>, Kanta Chechi<sup>2,3</sup>, Amélie Bonnefond<sup>1,4</sup>, Philippe Froguel<sup>1,4</sup>, Marc-Emmanuel Dumas<sup>1,2,4,5\*</sup>

<sup>1</sup>Inserm UMR1283, CNRS UMR8199, Institut Pasteur de Lille, Lille University Hospital, EGENODIA, European Genomic Institute for Diabetes (EGID), Université de Lille, 59045 Lille, France

<sup>2</sup> Section of Biomolecular Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College London, W12 0NN, UK

<sup>3</sup> Section of Genomic and Environmental Medicine, National Heart and Lung Institute, Imperial College London, SW3 6LY, UK

<sup>4</sup> Section of Genetics and Genomics, Department of Metabolism, Digestion and Reproduction, Imperial College London, W12 0NN, UK

<sup>5</sup> McGill Genome Centre, McGill University, Montréal (Qc) Canada,

†F.P-C and R. P-T contributed equally to this work.

\*corresponding author: [m.dumas@imperial.ac.uk](mailto:m.dumas@imperial.ac.uk) / [marc-emmanuel.dumas@cnrs.fr](mailto:marc-emmanuel.dumas@cnrs.fr)

\*corresponding author: [francesc.puig-castellvi@inserm.fr](mailto:francesc.puig-castellvi@inserm.fr)

### **Abstract**

The human microbiome and its metabolic output play an important role in shaping human health and disease. The host-microbe crosstalk is often characterised using metabolomics and metagenomics, yet a large proportion of microbial metabolites and the molecular functions of gut microbes remain unknown. This review summarises the strategies for annotation and discovery of the novel metabolites from the human gut microbiome, either using metabolomics alone or in combination with metagenomics, and presents data analysis methods for combining these two types of data to obtain biological insights. Applications of the gut microbiome research in biomarker screening, precision medicine, microbiome medicine, and drug discovery are also discussed, along with the perspectives, challenges and limitations of this research field.

### **Keywords**

Gut, microbiome, metabolomics, metagenomics, integration, clinical, BGCs, multi-omics

## List with abbreviations

BGCs: biosynthetic gene clusters  
BiG-SCAPE: biosynthetic gene similarity clustering and prospecting engine  
BiG-SLiCE: biosynthetic genes super-linear clustering engine  
BLAST: basic local alignment search tool  
CA: covariate adjusted  
CAG: co-abundant gene group  
ChEBI: Chemical Entities of Biological Interest  
CORASON: core analysis of syntenic orthologs to prioritise natural product gene clusters  
DDA: data-dependent acquisition  
FMT: faecal microbiome transplantation  
GCF: gene cluster family  
GMM: Gut Metabolic Module  
GNPS: global natural product social molecular networking  
HMDB: human metabolome database  
InChI: IUPAC international chemical identifier  
KEGG: kyoto encyclopaedia of genes and genomes  
LC: liquid chromatography  
MAG: metagenome-assembled genome  
MASST: Mass Spectrometry Search Tool  
MiMeDB: human microbial metabolome database  
MMWAS: metabolome- and metagenome-wide association study  
MS: mass spectrometry  
MS/MS: tandem mass spectrometry  
MSI: metabolomics standards initiative  
NMDS: non-metric multidimensional scaling  
NRPS: nonribosomal peptide synthetase  
O-PLS: orthogonal partial least squares regression  
O-PLS-DA: orthogonal partial least squares discriminant analysis  
PCA: principal component analysis  
PCoA: principal coordinates analysis  
PKS: polyketide synthase  
PLS: partial least squares regression  
PLS-DA: partial least squares discriminant analysis  
PoDP: paired omics data platform  
QC: quality control  
QMP: quantitative microbiome profiling  
RiPPs: ribosomally and post-translationally modified peptide  
SMILES: Simplified Molecular Input Line Entry System  
SNAP-MS: structural similarity network annotation platform for mass spectrometry  
SRM: selected reaction monitoring  
TMA: trimethylamine  
TMAO: trimethylamine *N*-oxide  
UHGG: unified human gastrointestinal genome atlas  
WGCNA: weighted correlation network analysis

## 1 Introduction

The human gut microbiome is a vast collection of viruses, bacteria, archaea and eukaryotes, where the outcomes of their metabolic processes shape their host's health through its metabolic output (e.g., production of microbial compounds [1], bioaccumulation and biotransformation of drugs [2]). The vast majority of the dominant human gut microbiota belong to a few phyla [3]. Still, low-abundant microbes should not be overlooked since they can still play major functions on the microbial community [4]. At the genus level, only a minority of microbes is shared across individuals. This inter-individuality is influenced by environmental factors, geographical location, diet, health status, life style, among others [5]. Inter-individual variations are also a consequence of the functional redundancy between the taxonomically distinct microorganisms able to colonise the human gut [6].

In gut microbiome studies, the two preferred high-throughput analytical approaches are metagenomics and metabolomics. These two approaches measure the set of microorganisms and metabolites present in the biological samples, respectively. In terms of metagenomics, gut microbiome has largely been characterised using next-generation sequencing methods such as shotgun sequencing [7]. Shotgun sequencing collects information about broad genome regions allowing a high accurate microbial identification at species level therefore yielding detailed information about the microbial community [8]. For metabolomics, both mass spectrometry (MS) and nuclear magnetic resonance spectroscopy (NMR) are used, the latter being more reproducible whilst the former being widely used due to higher sensitivity [9].

Similarly to that of humans, microbial metabolomes are intrinsically entangled to their genomes. This is evidenced by the countless studies that have investigated genome differences at the metabolome level in all kinds of organisms (i.e., gene knock-out studies [10,11], quantitative trait loci studies [12,13], genome wide association studies [14,15]). Both -omes are also marked by 'dark' zones: the 'dark metabolome' and the 'biosynthetic dark matter'. The 'dark metabolome' refers to the portion of the metabolome that has not been identified using metabolomics to-date, considered to be between 90 to 98% in LC-MS based metabolomics [16,17], and additional metabolites that are below the limit of detection for the current methods. The 'biosynthetic dark matter' [18], on the other hand, refers to the genetic material that encodes for enzymes and other proteins involved in the biosynthesis of natural products not yet identified. By studying the biosynthetic dark matter, we can gain insights into the potential functions and structures of the 'dark' molecules in the metabolome, and vice-versa.

In the gut, microbially-derived metabolites form part of the host metabolome. In this context, microbial signatures have been characterised by the inspection of the host's biological fluids, being faecal [19] and plasma/serum [20] samples the most studied biological materials. The faecal metabolome and metagenome can reflect the metabolic interplay among the host, diet and gut microbiota while the circulating metabolome can inform how commensal microbiota impact other organs distant from the gut. In this line, a recent screening of 8,583 patients based on the metagenomic sequencing of their faecal matter and the corresponding mass spectrometry-based metabolite profiling of the plasma samples revealed that the microbiome explains up to 46% of the variance of circulating metabolites [21]. Another independent study showed that the circulating (or serum) levels of 182 metabolites were explained by the microbiome composition, ranging their  $R^2$  between 0.05 and 0.56 ( $P < 0.05$ , 10% FDR correction) [22]. Collectively, these studies highlight the need to further study the gut microbiome to better comprehend the human body in the context of health. The release of the Unified Human Gastrointestinal Genome atlas [23], which encompasses 289,232 microbial genomes and can be used for genotyping the human gut microbiome, is set to significantly facilitate investigations that connect the microbial genotypes and the hosts' phenotypes.

Most of the current knowledge about the chemical diversity of metabolites has been obtained by sequential fractionation and chemical characterization of extracts from microorganisms and plants found in nature. Natural product and drug discovery research have historically contributed to the development of metagenomics [24,25] and metabolomics [26], leading to the elucidation, definition and eventually manipulation of biosynthetic pathways so as to facilitate the production and simultaneous detection of target metabolites [27] investigated nowadays in gut microbiome metabolomics studies [28].

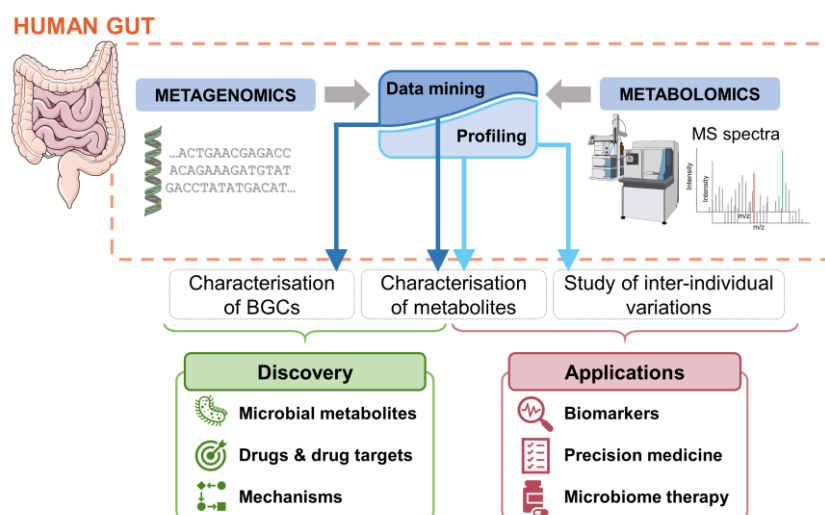
Central metabolism, also known as primary metabolism, is required for survival and contributes to the synthesis of intermediate compounds that act as precursors for specialised metabolism. Specialised metabolic pathways have evolved from primary metabolism and play a key role in growth, development, and the interaction with their environment. In microorganisms, specialised metabolites, also known as secondary metabolites [29] are commonly, but not exclusively, synthesised by polyketide synthases (PKSs), nonribosomal peptide synthetases (NRPSs) [30] and ribosomally and post-translationally modified peptides (RiPPs) [31] which are encoded by highly conserved biosynthetic gene clusters (BGCs). Genome mining, which involves using computational and predictive informatic tools, forms the basis for the search of these genes, and this search can be expanded to whole genera, strain collection and microbiomes [32].

Some of the breakthroughs contributed by the metagenomics era for microbiome research include the revelation of how biogeography and individuality shapes human skin microbiome [33], the gut microbial contribution to the response to cancer therapy [34] and the novel set of uncultured microbial species present in the human gut [35]. Until recently, the specialised metabolism of the gut microbiome has not been much investigated. In 2014, Donia and colleagues reported that typically the human gut microbiome contains on average 599 BGCs [36]. In 2017, another study led to the discovery of peptide aldehydes produced under conditions of host colonisation and capable of inhibiting human proteases [37], highlighting the clinical relevance of characterising the specialised metabolites produced by the gut microbiota. These results evidence the potential of investigating the specialised metabolites in gut microbiome metabolomics studies to enhance our understanding of the metabolic capabilities of the gut microbiome, its interactions with the host, and its implications for health and disease.

The integration of metabolomics and metagenomics data is often challenging due to the biological complexity of the gut microbiome, the heterogeneity of these data, and the lack of standard pipelines dedicated to this integration. Regarding the latter, common strategies to integrate the metagenomics and the metabolomics data range from classical bivariate statistics (e.g., calculation of correlations) to more sophisticated machine learning and pathway analysis methods. An overview of these methods will be given in section 4.2.

Recent advances in metabolomics and metagenomics data analyses for gut microbiome research greatly facilitated the development and refinement of modern medicine as well as natural products and drug development efforts (**Figure 1**). Given that each individual harbours a unique ‘microbial fingerprint’, gut microbiome holds significant potential for precision medicine. Precision or personalised medicine aims to account for the inter-individuality across patients (e.g., due to environment and lifestyle) while focusing on providing the best treatment possible, improving treatment efficacy and reducing toxicity [38]. Additionally, interventions targeting the gut microbiome, such as the use of prebiotics, probiotics and faecal microbiome transplantation (FMT) can be used for treating or preventing diseases as part of the microbiome medicine [39].

Integration of metagenomics and metabolomics is increasingly being applied for the identification and characterization of metabolites and microbiota co-variations and their implications in the host health [40–42]. The development of high-throughput methods and more powerful bioinformatic tools for integrative microbiome research can be potentially translated into clinical practice and better inform health and disease management (e.g. by designing more personalised treatments that account for the microbiome).



**Figure 1. Metagenome – metabolome crosstalk for human gut microbiome study and their clinical applications.** Both data can be mined to discover new metabolites, drugs, drug targets and biological mechanisms. A profiling strategy can also be applied to seek microbial biomarkers of certain health conditions or treatment efficacy, and to identify how gut microbes modify disease prognosis and use this knowledge to define personalised treatments based on precision medicine or microbiome therapies to counteract the effects of these microbes.

Here we present a comprehensive review of the current practices in gut microbiome research and identify methodological and knowledge gaps, by: 1) outlining the strategies to identify and annotate the gut microbial metabolites by either using metabolomics alone or in conjunction with metagenomics, 2) describing methods that enable the simultaneous analysis of metabolomics and metagenomics data to gain a better understanding of their biological implications, 3) highlighting the relevance of integrative gut microbiome research in health and disease, precision medicine, drug development and other clinical applications, and 4) identifying the perspectives, challenges, and limitations of this research field.

## 2 Study and identification of microbial metabolites in health and disease

Microbiota-derived metabolites studies have been focused on a few chemical classes (secondary bile acids, short-chain fatty acids, branched-chain amino acids, trimethylamines and aromatic amino acid derivatives due to their pathophysiological relevance in metabolic disorders [1]. Currently, the trend is to profile a larger number of metabolites (typically between 500 and 1500) by LC-MS. If then a metabolite is found to be associated with a pathology, different methodologies are applied to determine whether the metabolite is of microbial origin or not. While it is possible to establish the nature of some of these metabolites by a simple database search (since databases such as the Kyoto Encyclopedia of

Genes and Genomes (KEGG)[43] and Chemical Entities of Biological Interest (ChEBI)[44] compile this information), for some other metabolites this information is not available. In these cases, an experimental demonstration is needed.

In the following sections, the typical steps for establishing the microbial origin of metabolites in gut microbiome studies are described: the experimental analysis of the biological samples, the annotation of the measured metabolite signals, the use of molecular networking to improve the number of metabolite annotations, and the experimental validation of the metabolite-microbiome association.

## 2.1 Metabolomics experimental strategies

To achieve the aforementioned comprehensive characterizations, a combination of many empirical and data analysis strategies is applied (e.g., use different sample preparation methods, instrumental settings (i.e. liquid chromatography (LC)-MS and gas chromatography-MS), and/or chromatographic columns with different properties; combine targeted and untargeted methods [21,45–47]). Still, these methods are biased towards the major metabolite classes (e.g., organic acids, lipids) while other classes of metabolites (in particular, the specialised microbial metabolites) are poorly characterised. Thus, further development is needed to achieve much larger metabolic coverages. For example, knowing that many microbial enzymes convert diverse L-amino acids to D-configurations, more development is needed for comprehensive methods capable of distinguishing both enantiomers [48].

In metabolomics, tandem mass spectrometry (MS/MS) is widely used to generate spectral data that provide structural information about the detected compounds. In terms of mass spectra acquisition, comprehensive characterization can be reached either by using iterative data-dependent acquisition (DDA) strategies on selected samples [49], or by measuring the mass transitions employing selected reaction monitoring (SRM) methods for hundreds of metabolites (e.g., 1,036 metabolites in González-Domínguez *et al.* (2020) [50]). These two approaches are considered to be untargeted and targeted methods.

Targeted and untargeted methods can also be combined in various approaches, often referred by the umbrella terms of semi-targeted or pseudo-targeted methods. For example, when the metabolomics analysis consists of measuring mass transitions in SRM mode derived from a previous DDA analysis on selected biological samples [51]. This approach offers the advantage of having the same sensitivity as targeted methods but does not require the chemical standards for defining the mass transitions. Same terminology has also been used for untargeted methods applied in combination with a metabolite annotation strategy (based on chemical standards [52] or on an in-house spectral library [53]). Most recently, the Simultaneous Quantification and Discovery (SQUAD) approach has also been proposed. SQUAD allows, within a single LC-MS injection, to confidently annotate metabolites (using chemical standards) and to quantify them (using calibration curves, as in targeted metabolomics) and, at the same time, to identify additional metabolites from the experimental MS/MS spectra obtained from QC samples [54].

## 2.2 Metabolite annotation

In untargeted metabolomics, metabolite annotation is a critical step since flawed annotations can compromise the reliability of the results [55]. According to the Metabolomics Standards Initiative (MSI, a community-driven effort aimed at developing and promoting standards and best practices for metabolomics research [56]), different confidence levels of metabolite identification are proposed [57].

**Levels of confidence.** Unequivocal identification (Level 1) is only reached if the compound and the pure reference standard present two identical *orthogonal* characteristics (e.g., retention time in LC-MS

methods, collision cross section in ion mobility MS methods, accurate mass and MS/MS spectra), and both have been analysed in the same laboratory. To achieve this, several companies provide pure chemical compound libraries focused on the microbiome (i.e., the Fecal Metabolite Library from MetaSci, the Microbiome Metabolite Library of Standards from IROA Technologies) to facilitate the building of in-house reference libraries. Putative annotations (Level 2) are based on the data from public databases or on the physicochemical properties when reference standards are lacking. Despite their usefulness, these chemical libraries are incomplete as the number of detected metabolites is smaller than that found for the largest comprehensive studies [58].

Regarding MS/MS, a key advantage is that fragmentation of molecules using MS is highly consistent when repeated MS/MS acquisitions are performed under same or similar experimental conditions including the instrument technology, type of mass analyser, ionization method, collision method, and other parameters to be considered. This facilitates the annotation and identification of metabolites based on their fragmentation mass spectra using compound databases and pure standards, respectively.

**Harmonisation of metabolite identifiers.** In gut microbiome cohort studies, just like any other epidemiological study, biological findings are typically replicated in an independent cohort. Besides the complexity of the logistics for organising a replication study, comparing the metabolomics data from the two cohorts analysed in two different laboratories is not straightforward since metabolite annotations between studies need to be harmonised first. To facilitate the metabolite name harmonisation process, it is recommended to provide curated identifiers such as database identifiers (HMDB) or structural identifiers (e.g., the IUPAC International Chemical Identifier (InChI) [59], SMILES) in addition to the trivial names as these are often non-standardized (i.e. one molecule can have several trivial names). The InChIKey is a condensed alphabetic code for the InChI identifier string and are stereochemistry-specific (e.g., ROHFNLRFUQHCH-YFKPBYRVSA-N and ROHFNLRFUQHCH-RXMQYKEDSA-N for L- and D-leucine). However, it is also important to consider that some the spatial chemical structure can be neglected in some chemical identifiers such as the SMILES or when only the first layer of the InChIKey is used. Although the stereochemical characterization is not a common practice in metabolomics, it is necessary to use identifiers containing this information which is relevant for the biological activity of metabolites in the microbiome research [60].

**Metabolite ontologies.** Interpreting the metabolomics data can be overwhelming due to the large number of metabolites included in these studies. To address this, metabolites can be arranged in smaller sets and explored separately. Metabolites can be grouped according to the biological pathways they belong (as defined in BioCyc [61] or KEGG [43] databases) or to the chemical ontology, the latter inferred directly from the MS/MS data (Qemistree [62], CANOPUS [63]) or from their structural identifiers (ClassyFire [64]). ClassyFire chemical ontologies present a hierarchical classification (Kingdom, SuperClass, Class, and SubClass) format that resembles that of taxonomic data such as for metagenomics.

**Public databases.** General public databases contain MS/MS spectra of metabolites from different origins (e.g., MassBank [65], METLIN [66], HMDB [67]) and can be consulted for metabolite annotation. However, public databases for microbial metabolite annotation, such as the Human Microbial Metabolome Database (MiMeDB) [68] and the Natural Products Atlas [69], are scarcer. MiMeDB contains information from the taxonomy of 1,904 microbes of the human microbiome as well as the metabolites they produce. The Natural Products Atlas, as opposed to MiMeDB, is oriented to the specialised metabolites produced by fungi and bacteria. Additional databases of microbial metabolites, either public or not, can be consulted in the review of van Santen *et al.* [70].

Open spectral libraries are not only used for consulting metabolite annotation, but they are also a source of machine learning data for training and method validation of novel methods for compound



identification in metabolomics. Some computational advances worth mentioning are MS2LDA [71], MESSAR [72], Spec2Vec [73], MS2DeepScore [74] and CFM-ID [75] which represent the next generation of metabolomics analyses, however these will need to be validated in follow-up studies as the spectral databases keep growing in richness and diversity with high-quality data [17,76].

**Open-source software.** MS-DIAL [77], MZmine [78] and the MetaboAnnotation R-package [79] allow the processing of MS data including metabolite annotation. Other free tools have been reviewed by Misra, B [80]. Compound Discoverer (Thermo Scientific™) and MassHunter Workstation (Agilent® Technologies) have been developed as alternative commercial software solutions. In addition, the MASST interface can be used to search MS/MS data on public repositories (such as MassIVE [81], MetabolomicsWorkbench [82] and MetaboLights [83]), similarly to the Basic Local Alignment Search Tool (BLAST) method for investigating genes and protein sequences [84].

**Metabolomics reports.** One decade after the MSI guidelines were initially proposed, some updates have been made to better fit the needs that have been emerging across the community including the “METabolomics standaRds Initiative in Toxicology” (MERIT) that provides the most relevant practical guidelines for metabolomics specifically in regulatory toxicology [85]. To comply with the minimum standards of reporting scientific research and promote reuse of the metabolomics data, the spectral data must be deposited in open repositories. Recently, discussions around the good practices and recommendations regarding the compound identification in untargeted metabolomics have also emerged from multiple recognised researchers from different academic institutions working on this subject [17,76,86]. Some of these recommendations include the introduction of spectral entropy as a measure of the MS/MS spectra quality (low quality equals high entropy) [87] or reporting the ranking of annotations instead of keeping the top MS/MS match [76].

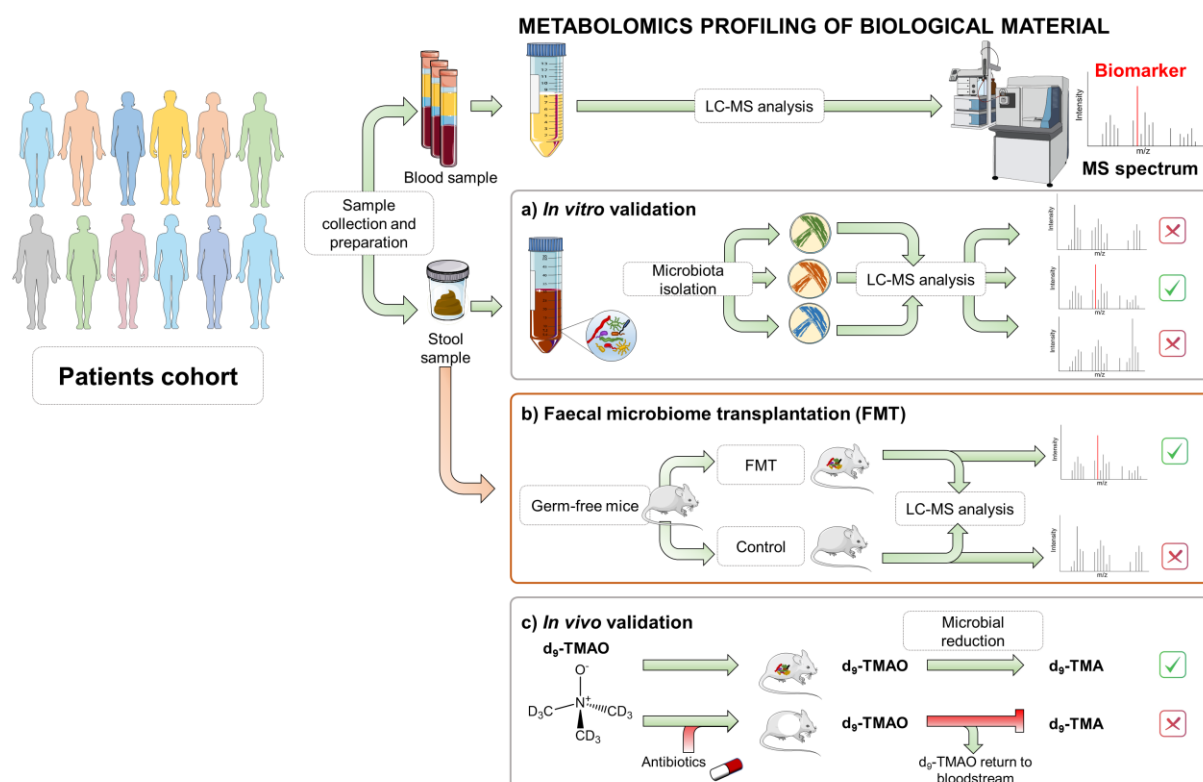
In addition to these specific recommendations for reporting metabolomics data, when working with data from human cohorts, it must be verified that the data is anonymized before submitting the data so that no one can identify individual patients.

### 2.3 Molecular networking

Expanding on these approaches, a different strategy called “molecular networking” was introduced based on the concept of spectral matching and similarity of fragmentation patterns. This approach consists of using the spectral data to build molecular networks where each feature ( $m/z$  - retention time pairs) is represented by nodes connected by spectral similarity (MS/MS information) through edges. Structurally related molecules can exhibit similar fragmentation; therefore, nodes of annotated metabolites can extend the annotation to those unknown features, guiding the chemical annotations of the adjacent nodes at both the structure and substructure level. Although molecular networks can be manually constructed, the Global Natural Product Social Molecular Networking (GNPS) [88] provides a dedicated online environment for this purpose. GNPS molecular networks can be interrogated with the Structural similarity Network Annotation Platform for Mass Spectrometry (SNAP-MS) tool to identify candidate metabolite classes of microbial origin from the Natural Products Atlas [89]. Similarly, the MS2LDA tool from the GNPS environment can be used to identify molecular substructures from MS/MS spectra [90]. Alternatively, if the expected reference spectra of the compounds are unavailable, these can be generated with software tools for *in silico* fragmentation such as SIRIUS [91] and MetFrag [92], among others. Another important development within the GNPS environment is the ion identity molecular networking (IIMN) which addresses the bottleneck of unconnected ion adducts given that one compound can generate multiple ion species and each precursor can exhibit a different fragmentation behaviour [93]. This information can be further integrated with other bioinformatic tools to validate the network connectivity shown in the molecular networking.

## 2.4 Experimental evidence of the microbial origin of a metabolite

After the annotation of relevant features and potential biomarkers, several strategies can be implemented in order to identify their microbial origin and bioactivity. The simplest approach consists of *in vitro* culturing of microorganisms isolated from the faeces and detecting the sought metabolites in the corresponding microbial extracts [94] (**Figure 2A**). Otherwise, faecal microbiome transplantation (FMT) can be carried out on mice to investigate whether the metabolomic signature is actually determined by the microbial species in the transplanted faeces [94,95] (**Figure 2B**). Last, *in vivo* experiments with germ-free or antibiotic-treated can be used to assess microbial transformations. In Hoyles *et al.* (2018) [96], deuterated TMAO was administered to antibiotic-treated mice to unambiguously prove that deuterated TMAO is retro-converted into deuterated trimethylamine (TMA) by the gut microbiota (**Figure 2C**).



**Figure 2. Experimental validation of gut microbial metabolites bioactivity involved in disease.** Metabolomics profiling of biological material such as blood samples by liquid chromatography coupled to mass spectrometry (LC-MS) leads to the detection and quantification of microbiome metabolites linked to disease (biomarkers) with a specific MS spectrum. **a)** Isolation of microorganisms from stool samples for *in vitro* validation of their production using LC-MS analysis, **b)** Faecal microbiome transplantation (FMT) to germ-free mice in order to validate bioconversions by the human gut microbiota, **c)** Deuterated trimethylamine N-oxide (TMAO) reduced to trimethylamine (TMA) by mice gut microbiota described in Hoyles *et al.*, 2018 as *in vivo* validation of TMAO retroconversion [96].

Regarding the *in vitro* culturing of isolates, current large-scale metagenomic studies from geographically and phenotypically diverse humans provides evidence about the low representation of microbial diversity by cultured isolates [97,98]. This is in part due to the complexity of cultivating some microbial strains outside their biome. General or specific culture media used for the growth of these microbial isolates may not contain the necessary nutrients or molecules needed for their development

or the production of metabolites that could be exclusively induced under naturally occurring complex community interactions [99,100] such as in the human gut environment. This represents a limitation of the current cultured-based techniques.

### **3 Metagenomes and genome mining**

Metagenomics has facilitated the exploration of several environments, enabling researchers to comprehend the microbial associations and interactions taking place within the microbial community, shifting our view from single organisms to microbial communities [101]. Progress in sequencing technologies has facilitated the generation of vast amounts of genome data. The parallel development of genome mining tools has enabled extending the study of genomes to the investigation of the human microbiota in health and disease [102]. As a result, substantial variations in human gut microbiome have been revealed according to different anatomical sites [103], geographical locations and lifestyles [104], and health status [28].

Metagenomic studies require multiple fundamental steps including sample collection, isolation of DNA, library preparation for sequencing followed by taxonomic and functional data analysis [105]. Then, gene data can be combined into larger biological entities, such as co-abundant genes (CAGs) and metagenome assembled genomes (MAGs).

CAGs are sets of genes that occur at a similar abundance across multiple samples, likely to represent a group of genes found on the same chromosome or DNA section (operon, plasmid, core genome). CAGs are meaningful reflections of the composition of the microbiome and have been used for studying gut microbial changes in association with colorectal cancer and inflammatory bowel disease [106], chronic obstructive pulmonary disease [107] and depression [108], among others.

MAGs represent microbial genomes. For the construction of MAGs, sequencing reads are assembled into scaffolds which are later grouped into candidate MAGs to be used for further analyses [109]. The common computational tools used in metagenomics can be classified into *upstream* analyses, using tools for sequence quality control (QC), metagenome assembly, QC assembly and metagenome binning and focusing on the construction of MAGs. They can also be classified into *downstream* analyses designed to annotate MAGs including gene prediction, gene annotation, functional and taxonomic analysis, MAG taxonomic classification and MAG abundance profiling [109].

In the following paragraphs, we will discuss the discovery and annotation of BGCs found in microbial genomes that encode the biosynthesis of specialised compounds and the tools that help the automation of this research from a functional point of view. This is also referred to as genome mining (section 3.1). BGCs annotation is important to increase the biological knowledge that can be extracted from gut microbiome studies employing metagenomics data. In these studies, several metagenomes from multiple samples are profiled to identify associations between the differences in the abundances of the microbial species or the molecular functions and the changes in the host phenotypes across samples. The strategies used to identify these associations are presented in section 3.2.

#### **3.1 Genome mining of microbial specialised metabolites**

The concept of genome mining emerged with the understanding of the biosynthetic capabilities and genetic basis of microbial metabolite production and from the interest in the microbial genetic potential

to produce metabolites predicted by genetics but not detected by analytical methods [32]. Genome mining typically relies on assembled and annotated genome sequences as its primary raw material. Consequently, it is crucial to understand that the accuracy and thoroughness of the assembly and annotation process can significantly impact the results of any genome-based study [110].

**Software for automated genome mining.** As previously mentioned, BGCs are defined as two or more genes from a particular genome that are physically clustered and encode a biosynthetic pathway for the production of specialised metabolites [111]. Computational identification of thousands of BGCs in microbial genome sequences is possible thanks to the development of automated bioinformatic software. One of the first tools developed in the early 2000s for automated cluster mining was the Decipher® pipeline and database [112]. Their strategy consisted of first, scanning microbial genomes coding for known chemically-diverse metabolites to build a database of gene clusters and then, to purify and confirm the metabolite structure through a genomics-guided discovery platform. Since then, different open source software and pipelines have emerged:

- BAGEL [113]
- CLUster SEquence ANalyzer (CLUSEAN) [114]
- NP.searcher [115]
- Secondary Metabolite Unknown Regions Finder (SMURF) [116]
- PRediction Informatics for Secondary Metabolomes (PRISM) [117]
- antibiotics & secondary metabolite analysis shell (antiSMASH) [118]

AntiSMASH was introduced and refined over the last 10 years, being currently the most widely used and the gold standard tool for mining bacterial, fungal (fungiSMASH) and plant (plantiSMASH) genomes for specialised metabolite BGCs [101], including automated identification of primary metabolic gene clusters from the gut microbiota (gutSMASH) [119]. Briefly, different types of sequence files (including FASTA files) or GenBank accession numbers can be provided to the web server (<https://antismash.secondarymetabolites.org>) or to the desktop downloadable version. From this information, genes encoding proteins are searched and a library of models specific for each type of BGCs is constructed with high accuracy (71 BGCs types for antiSMASH version 6) [101] along with the prediction of the chemical structures of metabolites produced by the BGCs. The outputs can be visualised as interactive and user-friendly maps, images and hyperlinks.

Despite the development of similar tools, antiSMASH provides a comprehensive analysis by integrating multiple algorithms and databases for studying microbial genomes and secondary metabolite biosynthesis covering a wide range of metabolites. This is supported by the accurate prediction by combining methods such as Hidden Markov Models (HMMs) considered as “the Lego of computational sequence analysis” [120] with other bioinformatic techniques. Besides this, the extensive database of antiSMASH is regularly updated, facilitating the annotation and prediction of specific compounds and therefore providing insights about their potential biological activities. As developed for the GNPS, antiSMASH has also provided community support mainly in the field of natural product discovery but also applied in microbiome research and microbial-related disciplines. It is also worth mentioning that antiSMASH is constantly updated and enriched by the multiple complementary emerging tools used to improve the genome mining techniques that researchers may simultaneously employ. These constant improvements position this tool as one of the most widely used.

**Expansion of genome mining.** The advances achieved regarding the creation of the ‘Minimum information about a Biosynthetic Gene cluster’ (MIBiG) allowed the standardised data annotation and

storage of experimentally validated BGCs of known functions [121] but also the cross-links that are provided between the Natural Products Atlas [69], GNPS [88] and PubChem [122]. Latest updates include the addition of structural identifiers in SMILES format directly from cross-linked databases, avoiding conflicts caused by SMILES formatting. Moreover, more compounds were linked with their bioactivities for bioactivity predictions from structure and finally, the molecular targets of more compounds were annotated [123].

As the analysis of genomes is expanding to large-scale pan-genomic mining such as those of the microbiomes, the ‘Biosynthetic Gene Similarity Clustering And Prospecting Engine’ (BiG-SCAPE) has been developed to facilitate the exploration of BGCs similarity networks at multiple hierarchical levels [124]. The ‘core analysis of syntenic orthologs to prioritise natural product gene clusters’ (CORASON) software tool, where syntenic refers to the genes present on the same chromosome of a given species, helps the elucidation of the evolutionary relationships between BGC phylogenies and gene cluster families (GCFs, groups of BGCs coding for chemically similar metabolites) [124]. The streamlined computational workflow using BiG-SCAPE and CORASON allows the integration of gene cluster identification and empirical biosynthetic data comparison including antiSMASH and MIBiG [124]. Despite recent advances in the previously mentioned software tools, pairwise BGCs comparisons used to build similarity networks and perform clustering analysis leads to a quadratic time complexity and therefore, to an impractical time of analysis. Thus, the recently developed Biosynthetic Genes Super-Linear Clustering Engine (BiG-SLiCE) enables the use of clustering algorithm to run in a near-linear time complexity, facilitating the analysis of large BGCs datasets in a shorter time and in a truly global GCFs on all available microbial genomes [125].

### **3.2 Functional mapping of metagenomic gene abundances**

Metagenomic datasets from gut microbiomes studies contain millions of genes. While the major sequencing consortia efforts between 2010 and 2016 reported that human gut microbiome is comprised of 10 million non-redundant genes [126–128], the latest efforts have put this number to be close to 22 million [129]. In addition to their massive sizes, metagenomics datasets also present an intrinsically large data dimensionality (e.g., genes are annotated at the same time at the taxonomic and functional levels, genes from different species can encode orthologous proteins). In order to cope with this high dimensionality, the taxonomic (or ecological) and functional variations of the microbiome are often evaluated separately.

Gene abundances in shotgun metagenomics data are affected by multiple external sources of sample variability (e.g., differences in sequence depth, systematic variability introduced during the DNA extraction, sample-specific differences) [130]. Hence, metagenomics datasets are routinely transformed in order to account for this variability and be able to pinpoint the intrinsic differences between study samples. Examples of such transformation methods include total count, cumulative sum scaling and rarefying, among many others [130].

Regarding these transformations, two considerations must be made. First, it is advised to use samples with a similar total cell count as transforming metagenomics data from samples with very different cell counts can lead to misleading interpretations. For example, after transformation, low abundance species may appear to be absent in samples with low cell count. Second, metagenomics variables describe relative proportions rather than absolute abundances due to the aforementioned external sources of variability. In the aim of recovering the absolute gene abundances, the strategy named Quantitative Microbiome Profiling (QMP) was introduced in 2017 [131]. In this strategy, the relative proportions

are converted to absolute counts by multiplying each sample's proportions with their corresponding microbial cell number, usually determined *via* flow cytometry.

Ecological changes in the microbiome can be determined by studying the differences across samples of either the CAGs, the MAGs or the gene markers for taxonomic classification (e.g., 16S gene, ITS). Analogously, genes can be grouped by their encoding molecular function to subsequently compare the abundances of these groups to determine the functional role that drives the changes in the studied microbiome. Many criteria exist for gene grouping and the most common are based on the system classifications defined by the KEGG [43] and the MetaCyc [132] pathways. KEGG provides two systems, KEGG PATHWAYS and KEGG MODULES, that group KOs (identifiers relative to functional orthologs defined from experimentally characterised genes and proteins) by the function units of the enzymes they encode or by the chemical reactions these enzymes catalyse, respectively. Aiming at the gut microbiome, a new KEGG-based classification has been recently made available that groups KOs into Gut Metabolic Modules (GMMs) [133]. Each GMM, manually curated, represents one cellular enzymatic process present in the gut, such as the anaerobic catabolism of carbohydrates, amino acids and lipids, cross-feeding interactions, the production of fermentation end products, archaeal methanogenesis and enzymatic processes related in mucus degradation [133].

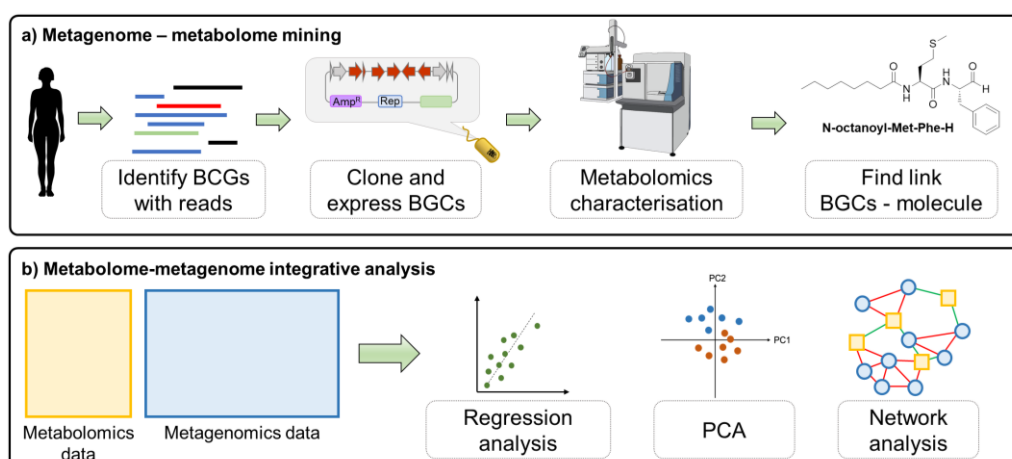
Due to the high-dimensionality of the metagenomics data, a common practice in gut microbiome studies is also to use these data for determining descriptors that can be more easily visualised and interpreted than the raw metagenomics data. These descriptors can be numeric (such as the classical ecological descriptors) or categorical (such as by defining enterotypes). Both approaches are presented in the following paragraphs. Alternatively, the metagenomics data can also be examined with multivariate statistical tools capable of reducing the large number of variables into just a few components that are a linear combination of the former, increasing the dataset interpretability and enabling their visualisation with relatively simple plots. The most frequent multivariate statistical methods used in metagenomics studies will be presented in section 4.2.

**Ecological descriptors.** Alpha- and beta-diversity indices are calculated to study intra- and inter-individual differences in the microbiome structure, respectively. Examples of alpha-diversity metrics are the number of gene-based units, Chao1, Simpson's and Shannon's indices. Shannon index measures uncertainty regarding the species identity while Simpson measures a probability of finding different species [134]. Chao1 index is the most popular asymptotic richness estimator but it is strongly influenced by the rarest and least-known species. Shannon and Simpson indices are more robust and less influenced by underrepresented species. Conversely, these two are affected by the most dominant species and, as consequence, do not reflect proportional changes when a non-dominant community is lost. Examples of beta-diversity metrics are the Bray–Curtis dissimilarity, Jaccard and (unweighted and weighted) UniFrac [135]. The Jaccard index is a measure of similarity between two communities based on the presence or absence of species, coded in a binary scale. In contrast, the Bray-Curtis dissimilarity index uses species abundance information, making it suitable for comparing similar communities with different abundances. Considering that the shared gut microbiome across individuals is small [5], classic binary distances are less recommended than the species abundance for investigating the gut microbiome inter-individual variability. UniFrac is a dissimilarity metric based on the evolutionary relatedness of species. Regarding all these beta-diversity metrics, it must be noted that gut microbiome varies along the gastrointestinal tract or even in stool samples from the same patient due to different factors such as acidity or transit time [136]. Therefore, all these sources of variations need to be considered prior to sampling in order to calculate reliable beta-diversity values.

**Enterotypes of the gut microbiome.** The inter-individual variation of the gut microbiome can be sometimes explained by discrete classifications of distinct microbiome compositions or enterotypes [137], which can be determined using a Dirichlet multinomial mixtures modelling framework [138]. This discretized information can be applied to stratify the cohort dataset into subgroups [139] or to add it as an independent variable in multivariable logistic regression models [140]. Besides being representative of a particular microbiome composition, enterotypes can also be linked to specific health conditions. For instance, Bact2 enterotype (marked by a high level of *Bacteroides*) is associated with severe obesity and inflammation [131,139].

## 4 Metabolome and metagenome integration

Metabolomics data can be combined with (meta)genomics data for mainly two reasons: (1) To increase the existing knowledge of microbial molecular functions by improving gene annotations and characterising BGC - metabolite pairs; (2) To understand the role of the microbes from a metabolomics perspective (**Figure 3**).



**Figure 3. Metabolome and metagenome integration strategies.**

### 4.1 (meta)genome-metabolome mining

This approach consists of computing the associations between all of the predicted BGCs and known microbial metabolites, which are computed and ranked. Later, the associations are prioritised for the hypothetical BGCs-MS2 pairs with the largest associations for subsequent validation [124,141]. These associations can be feature-based, correlation-based, or a combination of both [142]. Feature-based methods rely on homologies of already characterised pathways and therefore biased towards the most annotated BGCs classes (mainly for NRPs, PKs, and RiPPs); other more unusual BGC classes have also been mined [143]. The feature-based methods predict the hypothetical metabolite products from the BGCs and their theoretical mass spectra. Then, experimental mass spectra obtained from the same microbe are matched against the theoretical spectra, and a molecular network of the significant identifications is built [110]. Examples of feature-based tools are SANDPUMA [144] and NRPminer for NRPs [145], GNP for PKs [146], and MetaMiner for RiPPs [147]. In order to identify BGCs of novel classes, machine learning-based approaches and deep learning strategies have been developed [148]. With the correlation-based methods, BGCs from different strains are clustered according to their similarity into GCFs, assuming that similar BGCs will produce identical or structurally similar



metabolites [149]. Previous examples of correlation-based methods are BiG-SCAPE [124] and BiG-SLICE [125].

Already characterised links between BGCs and metabolites pairs can be consulted in the MIBiG repository [123]. Up to present, MIBiG contains around 2,700 pairs [123] although it is believed that most BGCs in nature are not yet related to any MIBiG reference gene cluster [110]. Knowing that less than 10% of the MS2 spectra are annotated in untargeted MS analyses [16,17], further advances in the development of methods for annotation of MS/MS data will concomitantly ease the path for annotating the BGCs from which the resulting metabolite product is not yet known.

It is worth mentioning here that the levels of the metabolites from the human circulating metabolome depend to some extent on the microbiota composition [22], reflecting that gut microbiome metabolomics studies should not be only focused on the microbially produced metabolites, but include as many metabolites as possible in order to capture as well the host metabolites modulated by microbial action [150]. Among the most known metabolite classes produced and transformed by the gut microbiota with reported effects on human physiology include organic acids, short chain fatty acids, branched-chain fatty acids, lipids, branched-chain and aromatic amino acids, vitamins, bile acids and neurotransmitters. However, at least 90% of the microbiome-derived metabolome is not characterised and many of these unknowns are highly likely to represent the missing link between the human-microbiome interactions [151].

In this regard, the Paired Omics Data Platform (PoDP) is a community-curated platform that links metabolomic and genomic or metagenomic datasets, together with the experimental details on culture, extraction, and instrumentation settings [152]. The main objective of this platform is to facilitate the accessibility of meticulously curated paired data sets, deposited in various public repositories, in the aim of aiding identification of natural product biosynthetic origins and metabolite structures [152]. As of March 2023, the PoPD dataset contained 76 projects comprising 4,932 unique metabolome samples and 1,306 metagenomes (among them, 481 metagenomes from human gut microbiomes), and provided empirical evidence for 115 BGCs and MS/MS spectra [152]. With a similar purpose as PoPD, microbeMASST [153], a domain-specific Mass Spectrometry Search Tool (MASST)[84], allows users to query MS/MS spectra against a reference database of MS/MS data acquired from bacterial and fungal monocultures and aims to identify the microorganisms responsible of the mass spectral signature.

#### **4.2 Metabolome-metagenome integrative analysis**

The analysis of metagenomics data for integrative purposes is challenging since data from all genes and metabolites are used, as opposed to the (meta)genome-metabolome mining approach where analyses are focused on BGCs and on a few metabolites. The approaches used for integrating these two types of data can be classified in either data-driven or knowledge-based approaches [154]. Data-driven approaches use statistical and machine learning algorithms to identify patterns and relationships between the data, while knowledge-based approaches are founded on external information that provides the link between the two sets of -omics data (e.g., biological databases, scientific literature) [154].

Data-driven integration approaches reveal data relationships by statistical and machine learning means. Most of these approaches can be grouped into statistical analysis, exploratory analysis, clustering analysis, regression analysis, discriminant analysis, multi-block analysis. The most used knowledge-based methods consist of pathway visualization, over-representation, and enrichment analysis and metabolic modelling.



In order to integrate metabolomics and metagenomics data in the context of human microbiota studies, epidemiological data analysis recommendations relative to the use of human cohort samples must be put into practice. In cross-sectional cohort studies, differences between samples can be driven by the patient interindividual differences (e.g., demographics, medication) also known as covariates. In this aim, three strategies are normally implemented to cope with these covariates either via: 1) stratification, or splitting the data into groups according to the covariates and to study each group separately [40]; 2) matching, or pairing samples across groups according to the covariates and excluding the unpaired samples from further analyses [41]; or 3) adjusting the omics variables for covariates using generalised linear model methods [155].

Before using any data analysis method, datasets need to be preprocessed to improve the interpretability of the models obtained with these methods. Due to the distinct nature of both data, the preferred preprocessing strategies for each data type are also different. Metabolomics data are typically first transformed (log-, glog-, or power-transformed) and then scaled (e.g., auto-scaling, pareto scaling) [156]. Additional steps of metabolomics data cleaning can be added on the metabolomics preprocessing (e.g. imputation of missing data, removal of variables present in the blanks and those with a large variability in the quality controls) [157]. Metagenomics data are preprocessed with more adapted methods to address their skewness and sparse nature (e.g., trimmed mean of M-values and relative log expression) [130]. Nevertheless, in human cohort studies, both datasets are often transformed with the rank-based inverse normal transform method [42,158] which converts the distribution of each continuous variable to appear more normally distributed.

**Statistical analysis.** The simplest data-driven integration strategy consists of performing bivariate statistics, i.e., correlating one -omic variable of each set (metabolomics and metagenomics) at a time. When associations between all possible pairs of metabolomics and metagenomics variables are evaluated systematically at once in a brute-force manner, this is referred to as Metabolome- and Metagenome-Wide Association Study (MMWAS) [159]. In all these analyses,  $p$ -values need to be corrected by multiple comparison testing.

**Exploratory analysis.** At a multivariate exploratory level, metabolomics data is typically inspected using Principal Component Analysis (PCA) whereas metagenomics data are mainly investigated with Principal Coordinates Analysis (PCoA) and Non-metric MultiDimensional Scaling (NMDS) [160]. While PCA sample-sample distances are based on Euclidean distances, both PCoA and NMDS use a similarity matrix as additional input for ordering the samples in low-dimensional space. For metagenomics analyses, the similarity matrix usually contains the ecological information of the Bray-Curtis dissimilarity distances, which represent the compositional differences between two different samples [160]. With both analyses performed separately, these can be combined with a Procrustes analysis to find a low-dimensional consensus between the metagenomics and metabolomics data and assess the sample differences across them [161].

**Clustering.** Clustering approaches (such as k-means and partitioning around medoids (PAM)) can be used to investigate sample similarity. In Battaglioli *et al.* (2018) [162], a PAM analysis revealed that clinical samples could be clustered in two groups, one related to samples from healthy patients and the other related to patients with dysbiosis. Similarly, clustering can be applied to identify the omics features (e.g., metabolites and gene abundances) that follow similar trends across the studied sample set. Nevertheless, inspecting this clustering can be tedious due to the large number of variables involved. In this line, weighted correlation network analysis (WGCNA) is a powerful approach to find clusters of co-regulated omics variables and to extract a common sample variation profile from each cluster,

considerably reducing the number of omics variables to be inspected for phenotype associations [163]. In Zheng *et al.* (2021), 230,868 microbial genes and 559 metabolites were grouped into 18 and 8 clusters, respectively [108]. Among them, 3 metagenomic and 4 metabolomic clusters were associated with depressive behaviours. Moreover, the functional analysis of the 3 relevant metagenomic clusters showed that genes were related to energy, neurotransmitter, nucleotide, and lipid metabolism pathways; whereas 4 metabolomic clusters were enriched in compounds from the peripheral and central fatty acyl, glycerophospholipid and sphingolipid metabolism [108].

A powerful strategy to represent how key microbial and metabolomics variables relate to each other is through the implementation of networks, where these variables are represented by nodes and the edges depict the degree of association between nodes in the network [159].

**Regression analysis.** At multivariate regression and classification levels, the preferred method is Partial Least Squares (PLS) regression, and the variants of it. In PLS, a descriptor matrix ( $\mathbf{X}$ , e.g., the metabolomics data) is regressed against a property variable ( $\mathbf{Y}$ , e.g., the abundances of one microbe) [164]. Unlike PCA, the obtained components in PLS (called ‘latent variables’) retain maximal covariance between  $\mathbf{X}$  and  $\mathbf{Y}$ . In this manner, by analysing the latent variables it is possible to determine the original variables in  $\mathbf{X}$  associated with  $\mathbf{Y}$  [164]. Orthogonal PLS (O-PLS) is a variant of the PLS method that removes the variation from  $\mathbf{X}$  that is not related to  $\mathbf{Y}$ , resulting in more interpretable latent variables, *i.e.*, predictive and orthogonal scores [165]. Within the context of human samples, the covariate adjusted PLS (CA-PLS) method is used to correct for known confounders using PLS [166].

**Discriminant analysis.** Analogous to the PLS methods described above, discriminant analysis (DA) can be performed using PLS-DA, O-PLS-DA and CA-PLS-DA. These methods use the same underlying algorithm as the corresponding PLS regression methods, but the  $\mathbf{Y}$  matrix contains discretized information (e.g., sample groups) instead. PLS-DA methods are used for sample class prediction and to determine the variables that maximise the separation among sample groups. For example, CA-PLS-DA was used to identify the metabolomic differences caused by the administration of bacterial  $\beta$ -glucuronidases in mice [167]. In another example, O-PLS-DA was used to build a model with pre-selected metabolomics and microbial features for predicting acute coronary syndrome [40].

Other methods can be used for prediction, such as random forest that classifies samples on the basis of an ensemble classifier consisting of multiple decision trees using a randomly selected subset of training samples and variables [168]. In Segata *et al.* (2019), random forest was used to predict colorectal cancer samples from metagenomics data [169].

**Multi-block analysis.** The multivariate methods presented above can be regarded as single-block chemometric methods, where models are built from one descriptor matrix only (e.g., a single omics data (metabolomics or metagenomics), or a data matrix containing all of the different sets of omics data. As a consequence, single-block methods do not account for the intrinsic properties of each omics data (e.g., data distribution, number of variables) nor the differences in these properties among the omics data. To address this situation, several multi-blocks methods have been proposed. The multi-block method ComDim enables an exploratory analysis of the variance common to all of the omics data [170]. In Puig-Castellví *et al.* (2021), ComDim was used to assess the batch effect across four different types of omics data [171]. Then, the consensus OPLS-DA method was developed for investigating multi-block data in a similar manner as the traditional single-block PLS methods [172]. Overall, the emergence of multi-block methods for analysing multi-omics data is a growing subject, partially driven by the decreasing

costs associated with generating such data. It is anticipated that additional methods will be developed in the near future to further advance this field.

Other data-driven methods have gained attention in the last few years. For example, the tools *mmvec* [173] and *MelonnPan* [174], which uses neural networks and elastic net regularised regressions, respectively, have been proposed to predict metabolic profiles associated with a given (meta)genomic data after training a model with paired metabolomics and (meta)genomics data. In another example, gradient-boosted decision trees were used to identify the determinants of human serum metabolome [22]. This study evaluated the microbial relative abundances as potential determinants of the metabolome, along with other variables such as diet, demography, genetics and lifestyle [22]. Gradient-boosted decision trees have been also used to find the major determinant of fasting circulating TMAO in adults, being mainly driven by kidney function with microbiota composition and diet playing minor, albeit significant roles [175].

**Pathway visualization.** With pathway visualization tools, it is possible to visualize how genes and metabolites are related within a biological pathway. Examples of these tools can be found in KEGG [43], WikiPathways [176] and MetaCyc [132], among others.

**Over-representation and enrichment analysis.** Over-representation analysis is based on the identification of the pathways where the number of significant biological variables is statistically larger (as compared to the other pathways). Set enrichment analysis is similar to over-representation analysis, but the statistical significance of each biological variable is considered in the analysis [177].

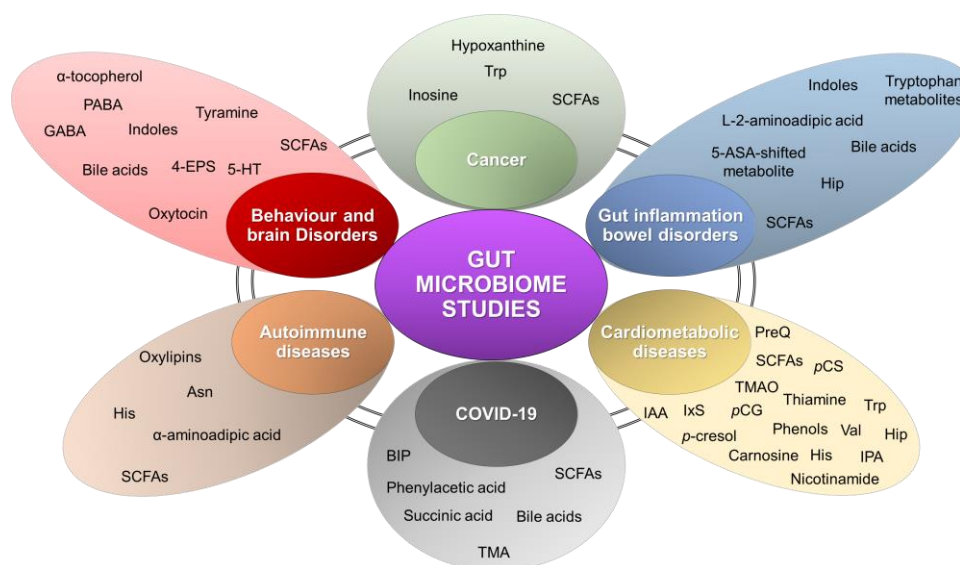
It must be said that, while pathway visualization, over-representation and enrichment analyses based on molecular functions encoded in genetic and/or gene expression data in well-established, their utilization for metabolomics data is currently constrained. This limitation stems from the relatively low mapping of metabolites within these pathways. To illustrate the limited coverage of the metabolomics in these pathway analysis tools, as of July 2023, KEGG lists 19,119 compounds while HMDB 5.0 includes a significantly larger number of compounds, totalling 217,920. We anticipate that ongoing and forthcoming research in metabolomics will contribute to the enhancement of biological pathway maps, consequently leading to important improvements in these tools.

**Metabolic modelling.** Bioinformatic tools that use the biological knowledge of the enzymes encoded in the gut microbiome as input have been developed to predict whether the variations in the metabolite profiles can be explained by the metabolic potential of the microbial communities as inferred from the KEGG KOs they encode. Examples of these tools are MIMOSA [178] and AMON [179]. Other bioinformatic tools predict the metabolite contributions of the microbes present in the gut by using metagenome-scale metabolic models. One of the most popular tools is COBRA, a system biology framework that employs curated (meta)genome-scale reconstructions of organisms to predict their metabolic fluxes after imposition of condition-specific constraints (for instance, calculated from experimental metabolomics data and/or from nutritional data) [180]. The COBRA framework can be used to explore the metabolic activity of human gut microbiome through the analysis of AGORA2 (assembly of gut organisms through reconstruction and analysis, version 2), which contains the genome-scale metabolic reconstructions of 7,302 human gut microorganisms. Specifically oriented for studying the gut microbiome, the COBRA-based framework MICOM (that uses AGORA reconstructions) was developed. MICOM has been used to predict the production of the short chain fatty acids by the gut in individuals with diabetes [181]. Another tool that uses genome-scale metabolic modelling is CASINO (Community And Systems-level INteractive Optimization) toolbox, which allowed for the quantitative

description of the altered amino acid levels in faeces and serum in response to diet intervention on 45 overweight and obese individuals [182].

## 5 Applications in Health and Disease

The symbiotic relationship between humans and their microbiomes have coevolved to perform numerous essential functions. Human microbial communities, especially the gut microbiome, have shown to be a key piece in host health, whose balance is influenced by different factors intrinsic and extrinsic to the host [28]. The emergence of new high-throughput technologies and the employment of meta-omics approaches are the foundation of modern microbiome research, making it possible to study human microbiomes on an epidemiological scale. During the last 20 years, omics-based research, including metagenomics, metatranscriptomics, metaproteomics and metabolomics, has revealed many disease-microbiome relationships [7]. **Figure 4** summarises some of the most recent metabolites – disease associations mediated by the gut microbiome described in the literature (**Table 1**).



**Figure 4.** Gut microbe-modulated metabolites involved in cardiometabolic diseases, gut inflammation and bowel disorders, autoimmune diseases, COVID-19, behaviour and brain disorders and cancer. Abbreviations: 4-EPS, 4-Ethylphenyl sulphate; 5-HT, 5-hydroxytryptamine; Asn, asparagine; BIP, 2,5-bis(3-indolyl methyl)pyrazine; GABA, gamma-aminobutyric acid; Hip, hippuric acid; His, histidine; IAA, indole-3-acetic acid; IxS, indoxyl sulphate; PABA, p-aminobenzoic acid; SCFAs, short chain fatty acids; IPA, indole-3-propionic acid; pCG, p-cresyl glucuronide; pCS, p-cresol sulphate; PreQ, 7-cyanodeazaguanine; TMA, trimethylamine; TMAO, trimethylamine-N-oxide; Trp, tryptophan; Val, valine.

**Table 1.** Literature covering metabolites – disease associations.

Literature covering metabolites – disease associations	
Autoimmune diseases	<ul style="list-style-type: none"> <li>• Rheumatoid arthritis [183]</li> <li>• Type 1 diabetes [184]</li> <li>• Atopic eczema [185]</li> </ul>

	<ul style="list-style-type: none"> <li>• Atopic asthma [186]</li> </ul>
<b>Gut inflammation / bowel disorders</b>	<ul style="list-style-type: none"> <li>• Irritable bowel syndrome [187]</li> <li>• Inflammatory bowel disease [188]</li> </ul>
<b>Cardiometabolic diseases</b> [40]	<ul style="list-style-type: none"> <li>• Cardiovascular disease [189]</li> <li>• Type 2 diabetes (T2D) [190]</li> <li>• Non-alcoholic fatty liver disease [191]</li> <li>• Chronic kidney disease [192]</li> <li>• Hypertension [193]</li> </ul>
<b>Behaviour and brain Disorders</b> [194]	<ul style="list-style-type: none"> <li>• Schizophrenia [195]</li> <li>• Attention-deficit hyperactivity disorder [196]</li> <li>• Depression [197]</li> <li>• Autism spectrum disorder [198]</li> <li>• Social anxiety disorder [199]</li> <li>• Parkinson's disease [200]</li> <li>• Alzheimer's disease [201]</li> <li>• Bipolar disorder [202]</li> <li>• Multiple sclerosis [203]</li> </ul>
Other	<ul style="list-style-type: none"> <li>• Cancer [204]</li> <li>• COVID-19 [205]</li> </ul>

Some metabolites are associated with multiple diseases. For example, hippuric acid is a glycine conjugate of benzoic acid produced in the mitochondria of the liver and kidneys linked to both gut inflammation bowel disorders [206] and cardiometabolic diseases [207]. This metabolite has been shown to have a positive association with microbial diversity but a negative association with blood pressure, non-alcoholic fatty liver disease and Crohn's disease. High hippuric acid levels are a marker of metabolic health, mainly in people having a diet elevated in saturated fats [208]. Another example is indole-3-propionic acid (IPA), derived from tryptophan metabolism by the gut microbiota, which has shown protection towards inflammation, lipid peroxidation and production of free radicals [209] but also promotes nerve regeneration and repair [210] having potential therapeutic implication in brain disorders. Moreover, IPA production significantly decreases in diabetes [211] while its circulating levels are negatively associated with advanced atherosclerosis [212].

## 5.1 Precision medicine

Precision medicine seeks to identify the best treatment according to the patient's genetic and clinical features using an individualised approach to improve treatment efficacy and toxicity [35]. Still in its infancy, -developments in precision medicine are driven by the accessibility of high-throughput -omics technologies (e.g., genomics, proteomics, metabolomics) as well as by the development of computational tools capable of integrating the heterogeneous data, resulting in insightful interpretations that can be used for clinical diagnosis and for anticipating patients' response to the available treatments.

Besides the patient's genotype and phenotype, the investigation of the patient's gut microbiome in precision medicine should not be neglected as gut metagenomes contain around 1,000 times more genes than their hosts and, in fact, up to 46% of the variance in the host's circulating metabolites can be explained by the gut microbiome [21]. Each individual possesses a highly unique microbiome and many microbiome functions remain well conserved due to functional redundancy among microbial species [194]. Although this, the existence of individual variations found in some patients rely on the strain-specific functions [195].

The research made on the microbiome has provided a link with diseases and illnesses. Some examples from metabolomics studies include atherosclerosis and TMAO [213]; cancer as a metabolic disorder and the discovery of metabolites associated with cancer [204]; some amino acids (branched-chain, aromatic and aminoadipic) as biomarkers for T2D [214]; and heart failure with both imidazole propionic acid [215] and phenylacetylglutamine [216]. The applications of omics technologies have lead, for example, to the discovery of the gene cluster responsible for the synthesis of TMA [217], whose abundance was correlated with the plasma levels of TMAO and the cardiovascular risk in humans [218].

## 5.2 Microbiome therapies

The practice of preventing or treating diseases through the manipulation of the microbiome comprises the emerging therapeutic activities of the microbiome medicine. They can be broadly divided into those that use living viable organisms, known as probiotics; those that use non-living components to selectively stimulate the growth and/or activity of the indigenous gut bacteria, known as prebiotics; and those that use metabolic products of probiotic microorganisms with biologic activity in the host, known as postbiotics [219]. Several examples of gut microbiome therapies are discussed below.

**Living components.** Nowadays, probiotics such as microorganisms from the *Lactobacillus* and *Bifidobacterium* genera are used as supplements in foods such as yoghurt, cheese, snacks and nutrition bars, breakfast cereals and infant formulas, among others [220]. The use of probiotics has been associated with several health benefits upon administration to humans, such as prevention of treatment of acute gastroenteritis, *Clostridium difficile*-associated diarrhoea, irritable bowel syndrome and other digestive complains, reduction of risk for neonatal late-onset sepsis, reduction in incidence and severity of respiratory infections, alleviation of depression, and reduction of cardiovascular risk factors associated with the cardiometabolic syndrome [220]. Supplementation with *Akkermansia muciniphila* reduced the levels of blood markers for liver dysfunction and inflammation in a pilot study comprising overweight/obese insulin-resistant volunteers [221]. Despite the wide range of applications and potential benefits for probiotics, several studies have pointed out that their effectiveness may be dependent on factors such as the genetic background and the gut microbiome composition and, as a consequence, probiotic administration should be carefully adapted according to the individual's condition [220]. Recent advances in gene editing technology have fostered the development of engineered probiotics, and some have already entered the clinical trial stage [222]. Engineered probiotics can exert specific activities including binding to targets, responding to pathogens or producing molecules with a therapeutic purpose, exhibiting promising results for the treatment of tumours, inflammation and metabolic diseases [222]. Another strategy to develop next-generation probiotics is to adapt strictly anaerobic bacteria associated with healthy microbiomes to tolerate oxygen exposure [223].

FMT is a medical procedure to treat microbiome dysbiosis through the transference of human faecal matter from a healthy donor to the patient. However, it is not formally considered within the probiotics

category as it includes undefined populations of microbiota. Currently, FMT is clinically applied with some restrictions, as a treatment for recurrent *Clostridium difficile* infection. This results in the suppression of indigenous intestinal microbiota and loss of colonisation of *C. difficile* and in the restoration of bacterial diversity in the gut microbiota [224]. Besides, FMT has been researched for IBD, IBS, melanoma, Tourette syndrome and metabolic syndrome [225]. Still, FMT treatment efficacy is dependent on the composition of the donor's gut microbiome and the route of delivery. It might also entail the transmission of communicable diseases and other attributes mediated by the microbiome [39].

**Non-living components.** Some drugs are positively influenced by the gut microbiome as is the case of metformin, used as an antihyperglycemic for T2D. In this example, metformin was found to alter gut bacterial composition by the promotion of beneficial bacteria growth in the gut, influencing inflammatory pathways and improving glucose control. Moreover, studies in multiple model organisms have shown that metformin can combat disorders related to age, delaying stem cell ageing, modulating mitochondrial function, lowering telomere senescence, among other mechanisms of ageing [226]. Statins have also been shown to have a protective effect against *Bacteroides 2* enterotype, validated in two independent cohorts [139].

Phage therapy represents an alternative to antibiotic treatment for their specificity in targeting bacteria and fewer side effects due to their lack of reaction to eukaryotic cells. Patients with inflammatory bowel disease exhibit abnormal gut viromes and an increase of bacteriophages. The administration of lytic bacteriophages to these patients could eliminate in a targeted way the bacteria that contribute to the disease pathogenesis [227]. In the case of patients with alcoholic liver disease, the increase of *Enterococcus faecalis* numbers is correlated with severe disease and mortality. This bacterium secretes an exotoxin named cytolysin, thus, phage therapy can target cytolytic *E. faecalis* strains. In the study of Duan *et al.*, this treatment resulted in a precise method for editing gut microbiota in patients with alcoholic liver disease [228]. Phage therapies, however, faces major challenges, mainly when the observed *in vitro* activities are not replicated *in vivo* experiments due to dosage effects, mutagenesis of phages, host antibodies, or the emergence of bacterial strains resistant to phages.

Last, postbiotic therapies consist in the administration of one or more microbial metabolites (e.g. exopolysaccharides, enzymes, short chain fatty acids, phenolic-derived metabolites, aromatic amino acids, vitamins, bacterial lysates, and cell-free supernatants containing secreted microbial metabolites). For instance, *Lactobacillus casei* supernatants showed anti-inflammatory and antioxidant effects on intestinal epithelial cells, macrophages, and neutrophils by reducing the secretion of TNF- $\alpha$  and increasing the secretion of IL-10 [229]. As another example, butyrate, a short chain fatty acid, increases the expression of immunosuppressive cytokines and downregulates several cytokines and proinflammatory receptors. In this line, rectal administration of butyrate reduced the inflammatory response in the large intestine of patients with ulcerative colitis [230].

### 5.3 Microbiome-Inspired Drug discovery

The pairing of genomic and metabolomic workflows have also enhanced the microbial natural product discovery under different approaches including: the reduction of re-discovery rates through the dereplication with genomic data, the focus on hotspots (such as novel or poorly known BGCs) for the search and discovery of new compounds, the identification of the putative BGC responsible for the production of a compound of interest for drug development, and the matching mass spectral data with predicted structures by the BGCs and MS spectra pairing tools [231].

## 6 Perspectives, challenges and limitations

In gut microbiome research, metagenomics and metabolomics can be regarded as two interdependent analytical strategies, where advancements in one can contribute to the progress of the other. On the one hand, better insight of the microbial composition can be informative of the chemical diversity in the metabolomics samples, and therefore be used as an assessment to design experimental methods that optimally capture this chemical information. In this line, tools capable of predicting the chemical diversity from metagenomics data can potentially be used to guide the development of the future LC-MS metabolomics methods.

On the other hand, the combination of metabolomics with metagenomics can also improve microbial genome assembly, gene prediction and the identification of genes that are induced or repressed under specific conditions thanks to the detection of metabolites using metabolomics [7]. Thus, more powerful tools for gene prediction are required, as particular specialised metabolite classes such as the RiPPs and fungal or plant genomes presenting introns within their sequence remain a major challenge [110]. In addition to this, BGCs having no core enzymes or non-canonical enzymes with unique protein sequences and unknown metabolite category, represent the dark matter for microbial genome mining [232]. To sum up, further development in (meta)genome-metabolome mining is needed to increase the number of annotated microbial genes and metabolites in integrative studies and, therefore greater insights of the role of the gut microbes and their interplay with the human host are obtained.

The gut microbiota is considered a crucial factor in the link between chemical exposure and disease development, such as obesity, diabetes, cancer, and their impact on immune and reproductive health. The interplay between the gut microbiota and these chemicals occur at many levels. First, chemicals can alter the composition and diversity of the gut microbiota, affecting metabolic pathways and potentially leading to an individually unique microbiome [233]. Second, the gut microbiota can transform the chemicals and bioaccumulate them, modifying the bioactivity of the compound within the human body [2]. We therefore envision that understanding how the gut microbiome mediate the activity of these chemicals will become an important piece in toxicology research.

The study of human gut microbiota interactions is complex and expensive as it requires the use of clinical cohorts that include hundreds to thousands of individuals. Recently, the use of 'gut-on-a-chip' technology has been proposed to simulate the human gut microenvironment at the *in vitro* level [234]. For example, a device with two channels separated by a porous membrane was devised to model the host intestinal cavity in one channel and the blood vessel in the other [235]. In the future, the combination of such technologies with advanced metabolomics strategies (e.g., single-cell metabolomics [236], MS imaging [237]) and metagenomics could provide a unique opportunity for investigating the human-gut microbiome crosstalk *ex vivo*. This will allow the investigation of conditions that would be challenging to study in the corresponding *in vivo* scenario (e.g., time-course monitoring of the microbial response to a drug). Moreover, this technology could be used, for instance, to study how the gut microbiome regulates the susceptibility of the human body to environmental pollutants, which could be crucial in explaining the distribution and frequency of certain diseases in society.

A challenge of the integrative analyses lies in the interpretation of the biological and clinical data, derived from the little information available about the role of each microbe in the gut. This knowledge is biased, towards pathogens and the most abundant microbes. However, these abundant microorganisms are not proxies for studying the functional complexity of the whole gut microbiome



since low-abundant microbes can also be the primary contributors of abundant proteins [137] with a specific function, to the extent of causing a disproportionate effect in the host [238]. A strategy to overcome this difficulty is the isolation, *in vitro* culture, and characterization of the gut microbes. However, this strategy is not feasible for all microbes as a large number are not culturable and, for those that can be cultured, their metabolism may differ from that presented *in vivo*.

Another aspect that needs to be addressed in the future is the development of diagnostic tools based on the two -omics discussed above to complement existing clinical risk scores. The levels of certain gut microbial metabolites in the host can be potentially measured to identify the host's health status, to monitor the patient's response to treatments (e.g., chemotherapy, immunotherapy), or even to anticipate their response to the treatments. Although the impact of the gut microbiome in health and disease has been well supported by the literature provided in this review, the wider demographic applicability of these diagnostic tools must be assessed as the screened cohorts were not representative of the whole human population [104,239].

To foster advancement in this research field and ultimately, to promote the application of the strategies discussed herein for clinical use, the development of the experimental protocols to study the human gut microbiome must be made in a synchronised manner. All the research groups should be involved, analogously to the 'European Biomonitoring for Europe initiative for the human monitoring of emerging contaminants' to improve the comparability of the results across laboratories [240].

This review represents a snapshot focusing on the intersection of metabolomics and metagenomics, narrowing the current knowledge gaps in the context of gut microbiome research. As this is a rapidly evolving field, these advances pave the way for translation into accurate patient diagnosis and prognosis as well as personalised treatments.

### **Declaration of competing interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### **Acknowledgments**

M.-E.D. is supported by the NIHR Imperial Biomedical Research Centre, GutsUK, Diabetes UK and by Isite ULNE (R-002-20-TALENT-DUMAS), jointly funded by ANR (ANR16-IDEX-0004-ULNE), the Hauts-de-France Regional Council (20002845) and by the European Metropolis of Lille (MEL). The European Genomic Institute for Diabetes is funded by a grant from the French National Research Agency (ANR-10-LABX-46, European Genomics Institute for Diabetes). The National Center for Diabetes Precision Medicine – PreciDIAB, is jointly supported by the French National Agency for Research (ANR-18-IBHU-0001), by the European Union (FEDER), by the Hauts-de-France Regional Council (Agreement 20001891/NP0025517) and by the European Metropolis of Lille (MEL, Agreement 2019\_ESR\_11) This research was conducted within the context of the CNRS–Imperial International Research Project METABO-LIC.

## References

- [1] A. Agus, K. Clément, H. Sokol, Gut microbiota-derived metabolites as central regulators in metabolic disorders, *Gut*. 70 (2021) 1174–1182. <https://doi.org/10.1136/gutjnl-2020-323071>.
- [2] I.D. Wilson, J.K. Nicholson, Gut microbiome interactions with drug metabolism, efficacy, and toxicity, *Translational Research*. 179 (2017) 204–222. <https://doi.org/10.1016/j.trsl.2016.08.002>.
- [3] C.A. Lozupone, J.I. Stombaugh, J.I. Gordon, J.K. Jansson, R. Knight, Diversity, stability and resilience of the human gut microbiota, *Nature*. 489 (2012) 220–230. <https://doi.org/10.1038/nature11550>.
- [4] G. Falony, M. Joossens, S. Vieira-Silva, J. Wang, Y. Darzi, K. Faust, A. Kurilshikov, M.J. Bonder, M. Valles-Colomer, D. Vandeputte, R.Y. Tito, S. Chaffron, L. Rymenans, C. Verspecht, L. De Sutter, G. Lima-Mendez, K. D’hoë, K. Jonckheere, D. Homola, R. Garcia, E.F. Tigchelaar, L. Eeckhaudt, J. Fu, L. Henckaerts, A. Zhernakova, C. Wijmenga, J. Raes, Population-level analysis of gut microbiome variation, *Science*. 352 (2016) 560–564. <https://doi.org/10.1126/science.aad3503>.
- [5] A. Zhernakova, A. Kurilshikov, M.J. Bonder, E.F. Tigchelaar, M. Schirmer, T. Vatanen, Z. Mujagic, A.V. Vila, G. Falony, S. Vieira-Silva, J. Wang, F. Imhann, E. Brandsma, S.A. Jankipersadsing, M. Joossens, M.C. Cenit, P. Deelen, M.A. Swertz, LifeLines cohort study, R.K. Weersma, E.J.M. Feskens, M.G. Netea, D. Gevers, D. Jonkers, L. Franke, Y.S. Aulchenko, C. Huttenhower, J. Raes, M.H. Hofker, R.J. Xavier, C. Wijmenga, J. Fu, Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity, *Science*. 352 (2016) 565–569. <https://doi.org/10.1126/science.aad3369>.
- [6] S. Louca, M.F. Polz, F. Mazel, M.B.N. Albright, J.A. Huber, M.I. O’Connor, M. Ackermann, A.S. Hahn, D.S. Srivastava, S.A. Crowe, M. Doebeli, L.W. Parfrey, Function and functional redundancy in microbial systems, *Nat Ecol Evol*. 2 (2018) 936–943. <https://doi.org/10.1038/s41559-018-0519-1>.
- [7] X. Zhang, L. Li, J. Butcher, A. Stintzi, D. Figeys, Advancing functional and translational microbiome research using meta-omics approaches, *Microbiome*. 7 (2019) 154. <https://doi.org/10.1186/s40168-019-0767-6>.
- [8] G.M. Weinstock, Genomic approaches to studying the human microbiota, *Nature*. 489 (2012) 250–256. <https://doi.org/10.1038/nature11553>.
- [9] A. Bauermeister, H. Mannocho-Russo, L.V. Costa-Lotufo, A.K. Jarmusch, P.C. Dorrestein, Mass spectrometry-based metabolomics in microbiome investigations, *Nat Rev Microbiol*. 20 (2022) 143–160. <https://doi.org/10.1038/s41579-021-00621-9>.
- [10] A. Valle, Z. Soto, H. Muhamadali, K.A. Hollywood, Y. Xu, J.R. Lloyd, R. Goodacre, D. Cantero, G. Cabrera, J. Bolivar, Metabolomics for the design of new metabolic engineering strategies for improving aerobic succinic acid production in *Escherichia coli*, *Metabolomics*. 18 (2022) 56. <https://doi.org/10.1007/s11306-022-01912-9>.
- [11] R. Porosk, A. Terasmaa, R. Mahlapuu, U. Soomets, K. Kilk, Metabolomics of the Wolfram Syndrome 1 Gene (Wfs1) Deficient Mice, *OMICS: A Journal of Integrative Biology*. 21 (2017) 721–732. <https://doi.org/10.1089/omi.2017.0143>.
- [12] A. Koulman, M. Cao, M. Faville, G. Lane, W. Mace, S. Rasmussen, Semi-quantitative and structural metabolic phenotyping by direct infusion ion trap mass spectrometry and its application in genetical metabolomics, *Rapid Communications in Mass Spectrometry*. 23 (2009) 2253–2263. <https://doi.org/10.1002/rcm.4142>.
- [13] L. Bartle, E. Peltier, J.F. Sundstrom, K. Sumbly, J.G. Mitchell, V. Jiranek, P. Marullo, QTL mapping: an innovative method for investigating the genetic determinism of yeast-bacteria interactions in wine, *Appl Microbiol Biotechnol*. 105 (2021) 5053–5066. <https://doi.org/10.1007/s00253-021-11376-x>.
- [14] C. Sauvage, V. Segura, G. Bauchet, R. Stevens, P.T. Do, Z. Nikoloski, A.R. Fernie, M. Causse, Genome-Wide Association in Tomato Reveals 44 Candidate Loci for Fruit Metabolic Traits, *Plant Physiology*. 165 (2014) 1120–1132. <https://doi.org/10.1104/pp.114.241521>.
- [15] S.-Y. Shin, E.B. Fauman, A.-K. Petersen, J. Krumsiek, R. Santos, J. Huang, M. Arnold, I. Erte, V. Forgetta, T.-P. Yang, K. Walter, C. Menni, L. Chen, L. Vasquez, A.M. Valdes, C.L. Hyde, V. Wang, D. Ziemek, P. Roberts, L. Xi, E. Grundberg, M. Waldenberger, J.B. Richards, R.P.

- Mohney, M.V. Milburn, S.L. John, J. Trimmer, F.J. Theis, J.P. Overington, K. Suhre, M.J. Brosnan, C. Gieger, G. Kastenmüller, T.D. Spector, N. Soranzo, An atlas of genetic influences on human blood metabolites, *Nat Genet.* 46 (2014) 543–550. <https://doi.org/10.1038/ng.2982>.
- [16] R.R. da Silva, P.C. Dorrestein, R.A. Quinn, Illuminating the dark matter in metabolomics, *Proceedings of the National Academy of Sciences.* 112 (2015) 12549–12550. <https://doi.org/10.1073/pnas.1516878112>.
- [17] N.F. de Jonge, K. Mildau, D. Meijer, J.J.R. Louwen, C. Bueschl, F. Huber, J.J.J. van der Hooft, Good practices and recommendations for using and benchmarking computational metabolomics metabolite annotation tools, *Metabolomics.* 18 (2022) 103. <https://doi.org/10.1007/s11306-022-01963-y>.
- [18] F. Hemmerling, J. Piel, Strategies to access biosynthetic novelty in bacterial genomes for drug discovery, *Nat Rev Drug Discov.* 21 (2022) 359–378. <https://doi.org/10.1038/s41573-022-00414-6>.
- [19] J. Zierer, M.A. Jackson, G. Kastenmüller, M. Mangino, T. Long, A. Telenti, R.P. Mohney, K.S. Small, J.T. Bell, C.J. Steves, A.M. Valdes, T.D. Spector, C. Menni, The fecal metabolome as a functional readout of the gut microbiome, *Nat Genet.* 50 (2018) 790–795. <https://doi.org/10.1038/s41588-018-0135-7>.
- [20] N. Psychogios, D.D. Hau, J. Peng, A.C. Guo, R. Mandal, S. Bouatra, I. Sinelnikov, R. Krishnamurthy, R. Eisner, B. Gautam, N. Young, J. Xia, C. Knox, E. Dong, P. Huang, Z. Hollander, T.L. Pedersen, S.R. Smith, F. Bamforth, R. Greiner, B. McManus, J.W. Newman, T. Goodfriend, D.S. Wishart, The human serum metabolome, *PLoS One.* 6 (2011) e16957. <https://doi.org/10.1371/journal.pone.0016957>.
- [21] K.F. Dekkers, S. Sayols-Baixeras, G. Baldanzi, C. Nowak, U. Hammar, D. Nguyen, G. Varotsis, L. Brunkwall, N. Nielsen, A.C. Eklund, J. Bak Holm, H.B. Nielsen, F. Ottosson, Y.-T. Lin, S. Ahmad, L. Lind, J. Sundström, G. Engström, J.G. Smith, J. Ärnlöv, M. Orho-Melander, T. Fall, An online atlas of human plasma metabolite signatures of gut microbiome composition, *Nat Commun.* 13 (2022) 5370. <https://doi.org/10.1038/s41467-022-33050-0>.
- [22] The IMI DIRECT consortium, N. Bar, T. Korem, O. Weissbrod, D. Zeevi, D. Rothschild, S. Leviatan, N. Kosower, M. Lotan-Pompan, A. Weinberger, C.I. Le Roy, C. Menni, A. Visconti, M. Falchi, T.D. Spector, J. Adamski, P.W. Franks, O. Pedersen, E. Segal, A reference map of potential determinants for the human serum metabolome, *Nature.* 588 (2020) 135–140. <https://doi.org/10.1038/s41586-020-2896-2>.
- [23] A. Almeida, S. Nayfach, M. Boland, F. Strozzi, M. Beracochea, Z.J. Shi, K.S. Pollard, E. Sakharova, D.H. Parks, P. Hugenholtz, N. Segata, N.C. Kyrpides, R.D. Finn, A unified catalog of 204,938 reference genomes from the human gut microbiome, *Nat Biotechnol.* 39 (2021) 105–114. <https://doi.org/10.1038/s41587-020-0603-3>.
- [24] A.L. Harvey, R. Edrada-Ebel, R.J. Quinn, The re-emergence of natural products for drug discovery in the genomics era, *Nat Rev Drug Discov.* 14 (2015) 111–129. <https://doi.org/10.1038/nrd4510>.
- [25] L.K. Caesar, R. Montaser, N.P. Keller, N.L. Kelleher, Metabolomics and genomics in natural products research: complementary tools for targeting new chemical entities, *Nat. Prod. Rep.* 38 (2021) 2041–2065. <https://doi.org/10.1039/D1NP00036E>.
- [26] A. Aharoni, R. Goodacre, A.R. Fernie, Plant and microbial sciences as key drivers in the development of metabolomics research, *Proceedings of the National Academy of Sciences.* 120 (2023) e2217383120. <https://doi.org/10.1073/pnas.2217383120>.
- [27] A.G. Atanasov, S.B. Zotchev, V.M. Dirsch, C.T. Supuran, Natural products in drug discovery: advances and opportunities, *Nat Rev Drug Discov.* 20 (2021) 200–216. <https://doi.org/10.1038/s41573-020-00114-z>.
- [28] H. VanEvery, E.A. Franzosa, L.H. Nguyen, C. Huttenhower, Microbiome epidemiology and association studies in human health, *Nat Rev Genet.* 24 (2023) 109–124. <https://doi.org/10.1038/s41576-022-00529-x>.
- [29] J. Davies, Specialized microbial metabolites: functions and origins, *J Antibiot.* 66 (2013) 361–364. <https://doi.org/10.1038/ja.2013.61>.
- [30] K.J. Weissman, The structural biology of biosynthetic megaenzymes, *Nat Chem Biol.* 11 (2015) 660–670. <https://doi.org/10.1038/nchembio.1883>.

- [31] C. Ongpipattanakul, E.K. Desormeaux, A. DiCaprio, W.A. van der Donk, D.A. Mitchell, S.K. Nair, Mechanism of Action of Ribosomally Synthesized and Post-Translationally Modified Peptides, *Chem. Rev.* 122 (2022) 14722–14814. <https://doi.org/10.1021/acs.chemrev.2c00210>.
- [32] N. Ziemert, M. Alanjary, T. Weber, The evolution of genome mining in microbes – a review, *Nat. Prod. Rep.* 33 (2016) 988–1005. <https://doi.org/10.1039/C6NP00025H>.
- [33] J. Oh, A.L. Byrd, C. Deming, S. Conlan, H.H. Kong, J.A. Segre, Biogeography and individuality shape function in the human skin metagenome, *Nature.* 514 (2014) 59–64. <https://doi.org/10.1038/nature13786>.
- [34] B. Routy, E. Le Chatelier, L. Derosa, C.P.M. Duong, M.T. Alou, R. Daillère, A. Fluckiger, M. Messaoudene, C. Rauber, M.P. Roberti, M. Fidelle, C. Flament, V. Poirier-Colame, P. Opolon, C. Klein, K. Iribarren, L. Mondragón, N. Jacquelot, B. Qu, G. Ferrere, C. Clémenson, L. Mezquita, J.R. Masip, C. Naltet, S. Brosseau, C. Kaderbhai, C. Richard, H. Rizvi, F. Levenez, N. Galleron, B. Quinquis, N. Pons, B. Ryffel, V. Minard-Colin, P. Gonin, J.-C. Soria, E. Deutsch, Y. Lioriot, F. Ghiringhelli, G. Zalcman, F. Goldwasser, B. Escudier, M.D. Hellmann, A. Eggermont, D. Raoult, L. Albiges, G. Kroemer, L. Zitvogel, Gut microbiome influences efficacy of PD-1–based immunotherapy against epithelial tumors, *Science.* 359 (2018) 91–97. <https://doi.org/10.1126/science.aan3706>.
- [35] A. Almeida, A.L. Mitchell, M. Boland, S.C. Forster, G.B. Gloor, A. Tarkowska, T.D. Lawley, R.D. Finn, A new genomic blueprint of the human gut microbiota, *Nature.* 568 (2019) 499–504. <https://doi.org/10.1038/s41586-019-0965-1>.
- [36] M.S. Donia, P. Cimermancic, C.J. Schulze, L.C. Wieland Brown, J. Martin, M. Mitreva, J. Clardy, R.G. Linington, M.A. Fischbach, A Systematic Analysis of Biosynthetic Gene Clusters in the Human Microbiome Reveals a Common Family of Antibiotics, *Cell.* 158 (2014) 1402–1414. <https://doi.org/10.1016/j.cell.2014.08.032>.
- [37] C.-J. Guo, F.-Y. Chang, T.P. Wyche, K.M. Backus, T.M. Acker, M. Funabashi, M. Taketani, M.S. Donia, S. Nayfach, K.S. Pollard, C.S. Craik, B.F. Cravatt, J. Clardy, C.A. Voigt, M.A. Fischbach, Discovery of Reactive Microbiota-Derived Metabolites that Inhibit Host Proteases, *Cell.* 168 (2017) 517–526.e18. <https://doi.org/10.1016/j.cell.2016.12.021>.
- [38] D.A. Schupack, R.A.T. Mars, D.H. Voelker, J.P. Abeykoon, P.C. Kashyap, The promise of the gut microbiome as part of individualized treatment strategies, *Nat Rev Gastroenterol Hepatol.* 19 (2022) 7–25. <https://doi.org/10.1038/s41575-021-00499-1>.
- [39] N. Zmora, E. Soffer, E. Elinav, Transforming medicine with the microbiome, *Science Translational Medicine.* 11 (2019) eaaw1815. <https://doi.org/10.1126/scitranslmed.aaw1815>.
- [40] S. Fromentin, S.K. Forslund, K. Chechi, J. Aron-Wisnewsky, R. Chakaroun, T. Nielsen, V. Tremaroli, B. Ji, E. Prifti, A. Myridakis, J. Chilloux, P. Andrikopoulos, Y. Fan, M.T. Olanipekun, R. Alves, S. Adiouch, N. Bar, Y. Talmor-Barkan, E. Belda, R. Caesar, L.P. Coelho, G. Falony, S. Fellahi, P. Galan, N. Galleron, G. Helft, L. Hoyles, R. Isnard, E. Le Chatelier, H. Julienne, L. Olsson, H.K. Pedersen, N. Pons, B. Quinquis, C. Rouault, H. Roume, J.-E. Salem, T.S.B. Schmidt, S. Vieira-Silva, P. Li, M. Zimmermann-Kogadeeva, C. Lewinter, N.B. Søndertoft, T.H. Hansen, D. Gauguier, J.P. Gøtze, L. Køber, R. Kornowski, H. Vestergaard, T. Hansen, J.-D. Zucker, S. Hercberg, I. Letunic, F. Bäckhed, J.-M. Oppert, J. Nielsen, J. Raes, P. Bork, M. Stumvoll, E. Segal, K. Clément, M.-E. Dumas, S.D. Ehrlich, O. Pedersen, Microbiome and metabolome features of the cardiometabolic disease spectrum, *Nat Med.* 28 (2022) 303–314. <https://doi.org/10.1038/s41591-022-01688-4>.
- [41] Y. Talmor-Barkan, N. Bar, A.A. Shaul, N. Shahaf, A. Godneva, Y. Bussi, M. Lotan-Pompan, A. Weinberger, A. Shechter, C. Chezar-Azerrad, Z. Arow, Y. Hammer, K. Chechi, S.K. Forslund, S. Fromentin, M.-E. Dumas, S.D. Ehrlich, O. Pedersen, R. Kornowski, E. Segal, Metabolomic and microbiome profiling reveals personalized risk factors for coronary artery disease, *Nat Med.* 28 (2022) 295–302. <https://doi.org/10.1038/s41591-022-01686-6>.
- [42] Q. Qi, J. Li, B. Yu, J.-Y. Moon, J.C. Chai, J. Merino, J. Hu, M. Ruiz-Canela, C. Rebholz, Z. Wang, M. Usyk, G.-C. Chen, B.C. Porneala, W. Wang, N.Q. Nguyen, E.V. Feofanova, M.L. Grove, T.J. Wang, R.E. Gerszten, J. Dupuis, J. Salas-Salvadó, W. Bao, D.L. Perkins, M.L. Daviglius, B. Thyagarajan, J. Cai, T. Wang, J.E. Manson, M.A. Martínez-González, E. Selvin, K.M. Rexrode, C.B. Clish, F.B. Hu, J.B. Meigs, R. Knight, R.D. Burk, E. Boerwinkle, R.C. Kaplan, Host and gut microbial tryptophan metabolism and type 2 diabetes: an integrative analysis of host genetics,

- diet, gut microbiome and circulating metabolites in cohort studies, *Gut*. 71 (2022) 1095–1105. <https://doi.org/10.1136/gutjnl-2021-324053>.
- [43] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, M. Tanabe, KEGG as a reference resource for gene and protein annotation, *Nucleic Acids Res.* 44 (2016) D457–462. <https://doi.org/10.1093/nar/gkv1070>.
- [44] K. Degtyarenko, P. de Matos, M. Ennis, J. Hastings, M. Zbinden, A. McNaught, R. Alcántara, M. Darsow, M. Guedj, M. Ashburner, ChEBI: a database and ontology for chemical entities of biological interest, *Nucleic Acids Res.* 36 (2008) D344–D350. <https://doi.org/10.1093/nar/gkm791>.
- [45] W.R. Wikoff, A.T. Anfora, J. Liu, P.G. Schultz, S.A. Lesley, E.C. Peters, G. Siuzdak, Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites, *Proceedings of the National Academy of Sciences*. 106 (2009) 3698–3703. <https://doi.org/10.1073/pnas.0812874106>.
- [46] F. Chen, X. Dai, C.-C. Zhou, K. Li, Y. Zhang, X.-Y. Lou, Y.-M. Zhu, Y.-L. Sun, B.-X. Peng, W. Cui, Integrated analysis of the faecal metagenome and serum metabolome reveals the role of gut microbiome-associated metabolites in the detection of colorectal cancer and adenoma, *Gut*. 71 (2022) 1315–1325. <https://doi.org/10.1136/gutjnl-2020-323476>.
- [47] J. Medina, R. Borreggine, T. Teav, L. Gao, S. Ji, J. Carrard, C. Jones, N. Blomberg, M. Jech, A. Atkins, C. Martins, A. Schmidt-Trucksass, M. Giera, A. Cazenave-Gassiot, H. Gallart-Ayala, J. Ivanisevic, Omic-Scale High-Throughput Quantitative LC–MS/MS Approach for Circulatory Lipid Phenotyping in Clinical Research, *Anal. Chem.* 95 (2023) 3168–3179. <https://doi.org/10.1021/acs.analchem.2c02598>.
- [48] Y. Gonda, A. Matsuda, K. Adachi, C. Ishii, M. Suzuki, A. Osaki, M. Mita, N. Nishizaki, Y. Ohtomo, T. Shimizu, M. Yasui, K. Hamase, J. Sasabe, Mammals sustain amino acid homochirality against chiral conversion by symbiotic microbes, *Proceedings of the National Academy of Sciences*. 120 (2023) e2300817120. <https://doi.org/10.1073/pnas.2300817120>.
- [49] I. Ten-Doménech, T. Martínez-Sena, M. Moreno-Torres, J.D. Sanjuan-Herráez, J.V. Castell, A. Parra-Llorca, M. Vento, G. Quintás, J. Kuligowski, Comparing Targeted vs. Untargeted MS2 Data-Dependent Acquisition for Peak Annotation in LC–MS Metabolomics, *Metabolites*. 10 (2020) 126. <https://doi.org/10.3390/metabo10040126>.
- [50] R. González-Domínguez, O. Jáuregui, M.I. Queipo-Ortuño, C. Andrés-Lacueva, Characterization of the Human Exposome by a Comprehensive and Quantitative Large-Scale Multianalyte Metabolomics Platform, *Anal. Chem.* 92 (2020) 13767–13775. <https://doi.org/10.1021/acs.analchem.0c02008>.
- [51] F. Zheng, X. Zhao, Z. Zeng, L. Wang, W. Lv, Q. Wang, G. Xu, Development of a plasma pseudotargeted metabolomics method based on ultra-high-performance liquid chromatography–mass spectrometry, *Nat Protoc.* 15 (2020) 2519–2537. <https://doi.org/10.1038/s41596-020-0341-5>.
- [52] W.B. Dunn, A. Erban, R.J.M. Weber, D.J. Creek, M. Brown, R. Breitling, T. Hankemeier, R. Goodacre, S. Neumann, J. Kopka, M.R. Viant, Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics, *Metabolomics*. 9 (2013) 44–66. <https://doi.org/10.1007/s11306-012-0434-4>.
- [53] N. Che, Y. Ma, H. Ruan, L. Xu, X. Wang, X. Yang, X. Liu, Integrated semi-targeted metabolomics analysis reveals distinct metabolic dysregulation in pleural effusion caused by tuberculosis and malignancy, *Clinica Chimica Acta*. 477 (2018) 81–88. <https://doi.org/10.1016/j.cca.2017.12.003>.
- [54] B. Amer, R.R. Deshpande, S.S. Bird, Simultaneous Quantitation and Discovery (SQUAD) Analysis: Combining the Best of Targeted and Untargeted Mass Spectrometry-Based Metabolomics, *Metabolites*. 13 (2023) 648. <https://doi.org/10.3390/metabo13050648>.
- [55] G. Theodoridis, H. Gika, D. Raftery, R. Goodacre, R.S. Plumb, I.D. Wilson, Ensuring Fact-Based Metabolite Identification in Liquid Chromatography–Mass Spectrometry-Based Metabolomics, *Anal. Chem.* 95 (2023) 3909–3916. <https://doi.org/10.1021/acs.analchem.2c05192>.
- [56] O. Fiehn, D. Robertson, J. Griffin, M. van der Werf, B. Nikolau, N. Morrison, L.W. Sumner, R. Goodacre, N.W. Hardy, C. Taylor, J. Fostel, B. Kristal, R. Kaddurah-Daouk, P. Mendes, B. van

- Ommen, J.C. Lindon, S.-A. Sansone, The metabolomics standards initiative (MSI), *Metabolomics*. 3 (2007) 175–178. <https://doi.org/10.1007/s11306-007-0070-6>.
- [57] L.W. Sumner, A. Amberg, D. Barrett, M.H. Beale, R. Beger, C.A. Daykin, T.W.-M. Fan, O. Fiehn, R. Goodacre, J.L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A.N. Lane, J.C. Lindon, P. Marriott, A.W. Nicholls, M.D. Reily, J.J. Thaden, M.R. Viant, Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI), *Metabolomics*. 3 (2007) 211–221. <https://doi.org/10.1007/s11306-007-0082-2>.
- [58] N. Ishii, K. Nakahigashi, T. Baba, M. Robert, T. Soga, A. Kanai, T. Hirasawa, M. Naba, K. Hirai, A. Hoque, P.Y. Ho, Y. Kakazu, K. Sugawara, S. Igarashi, S. Harada, T. Masuda, N. Sugiyama, T. Togashi, M. Hasegawa, Y. Takai, K. Yugi, K. Arakawa, N. Iwata, Y. Toya, Y. Nakayama, T. Nishioka, K. Shimizu, H. Mori, M. Tomita, Multiple High-Throughput Analyses Monitor the Response of *E. coli* to Perturbations, *Science*. 316 (2007) 593–597. <https://doi.org/10.1126/science.1132067>.
- [59] V. Koistinen, O. Kärkkäinen, P. Keski-Rahkonen, H. Tsugawa, A. Scalbert, M. Arita, D. Wishart, K. Hanhineva, Towards a Rosetta stone for metabolomics: recommendations to overcome inconsistent metabolite nomenclature, *Nat Metab.* (2023) 1–4. <https://doi.org/10.1038/s42255-023-00757-3>.
- [60] M. Matsumoto, A. Kunisawa, T. Hattori, S. Kawana, Y. Kitada, H. Tamada, S. Kawano, Y. Hayakawa, J. Iida, E. Fukusaki, Free D-amino acids produced by commensal bacteria in the colonic lumen, *Sci Rep*. 8 (2018) 17915. <https://doi.org/10.1038/s41598-018-36244-z>.
- [61] P.D. Karp, C.A. Ouzounis, C. Moore-Kochlacs, L. Goldovsky, P. Kaipa, D. Ahrén, S. Tsoka, N. Darzentas, V. Kunin, N. López-Bigas, Expansion of the BioCyc collection of pathway/genome databases to 160 genomes, *Nucleic Acids Res.* 33 (2005) 6083–6089. <https://doi.org/10.1093/nar/gki892>.
- [62] A. Tripathi, Y. Vázquez-Baeza, J.M. Gauglitz, M. Wang, K. Dührkop, M. Nothias-Esposito, D.D. Acharya, M. Ernst, J.J.J. van der Hoof, Q. Zhu, D. McDonald, A.D. Brejnrod, A. Gonzalez, J. Handelsman, M. Fleischauer, M. Ludwig, S. Böcker, L.-F. Nothias, R. Knight, P.C. Dorrestein, Chemically informed analyses of metabolomics mass spectrometry data with Qemistree, *Nat Chem Biol*. 17 (2021) 146–151. <https://doi.org/10.1038/s41589-020-00677-3>.
- [63] K. Dührkop, L.-F. Nothias, M. Fleischauer, R. Reher, M. Ludwig, M.A. Hoffmann, D. Petras, W.H. Gerwick, J. Rousu, P.C. Dorrestein, S. Böcker, Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra, *Nat Biotechnol*. 39 (2021) 462–471. <https://doi.org/10.1038/s41587-020-0740-8>.
- [64] Y. Djoumbou Feunang, R. Eisner, C. Knox, L. Chepelev, J. Hastings, G. Owen, E. Fahy, C. Steinbeck, S. Subramanian, E. Bolton, R. Greiner, D.S. Wishart, ClassyFire: automated chemical classification with a comprehensive, computable taxonomy, *J Cheminform*. 8 (2016) 61. <https://doi.org/10.1186/s13321-016-0174-y>.
- [65] MassBank: a public repository for sharing mass spectral data for life sciences - Horai - 2010 - Journal of Mass Spectrometry - Wiley Online Library, (n.d.). <https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/jms.1777> (accessed March 24, 2023).
- [66] C. Guijas, J.R. Montenegro-Burke, X. Domingo-Almenara, A. Palermo, B. Warth, G. Hermann, G. Koellensperger, T. Huan, W. Uritboonthai, A.E. Aisporna, D.W. Wolan, M.E. Spilker, H.P. Benton, G. Siuzdak, METLIN: A Technology Platform for Identifying Knowns and Unknowns, *Anal Chem*. 90 (2018) 3156–3164. <https://doi.org/10.1021/acs.analchem.7b04424>.
- [67] D.S. Wishart, A. Guo, E. Oler, F. Wang, A. Anjum, H. Peters, R. Dizon, Z. Sayeeda, S. Tian, B.L. Lee, M. Berjanskii, R. Mah, M. Yamamoto, J. Jovel, C. Torres-Calzada, M. Hiebert-Giesbrecht, V.W. Lui, D. Varshavi, D. Varshavi, D. Allen, D. Arndt, N. Khetarpal, A. Sivakumaran, K. Harford, S. Sanford, K. Yee, X. Cao, Z. Budinski, J. Liigand, L. Zhang, J. Zheng, R. Mandal, N. Karu, M. Dambrova, H.B. Schiöth, R. Greiner, V. Gautam, HMDB 5.0: the Human Metabolome Database for 2022, *Nucleic Acids Research*. 50 (2022) D622–D631. <https://doi.org/10.1093/nar/gkab1062>.

- [68] D.S. Wishart, E. Oler, H. Peters, A. Guo, S. Girod, S. Han, S. Saha, V.W. Lui, M. LeVatte, V. Gautam, R. Kaddurah-Daouk, N. Karu, MiMeDB: the Human Microbial Metabolome Database, *Nucleic Acids Res.* 51 (2023) D611–D620. <https://doi.org/10.1093/nar/gkac868>.
- [69] J.A. van Santen, G. Jacob, A.L. Singh, V. Aniebok, M.J. Balunas, D. Bunsko, F.C. Neto, L. Castaño-Espriu, C. Chang, T.N. Clark, J.L. Cleary Little, D.A. Delgadillo, P.C. Dorrestein, K.R. Duncan, J.M. Egan, M.M. Galey, F.P.J. Haeckl, A. Hua, A.H. Hughes, D. Iskakova, A. Khadilkar, J.-H. Lee, S. Lee, N. LeGrow, D.Y. Liu, J.M. Macho, C.S. McCaughey, M.H. Medema, R.P. Neupane, T.J. O'Donnell, J.S. Paula, L.M. Sanchez, A.F. Shaikh, S. Soldatou, B.R. Terlouw, T.A. Tran, M. Valentine, J.J.J. van der Hooft, D.A. Vo, M. Wang, D. Wilson, K.E. Zink, R.G. Linington, The Natural Products Atlas: An Open Access Knowledge Base for Microbial Natural Products Discovery, *ACS Cent. Sci.* 5 (2019) 1824–1833. <https://doi.org/10.1021/acscentsci.9b00806>.
- [70] J.A. van Santen, S.A. Kautsar, M.H. Medema, R.G. Linington, Microbial natural product databases: moving forward in the multi-omics era, *Nat. Prod. Rep.* 38 (2021) 264–278. <https://doi.org/10.1039/D0NP00053A>.
- [71] J.J.J. van der Hooft, J. Wandy, M.P. Barrett, K.E.V. Burgess, S. Rogers, Topic modeling for untargeted substructure exploration in metabolomics, *Proceedings of the National Academy of Sciences.* 113 (2016) 13738–13743. <https://doi.org/10.1073/pnas.1608041113>.
- [72] Y. Liu, A. Mrzic, P. Meysman, T. De Vijlder, E.P. Romijn, D. Valkenburg, W. Bittremieux, K. Laukens, MESSAR: Automated recommendation of metabolite substructures from tandem mass spectra, *PLoS One.* 15 (2020) e0226770. <https://doi.org/10.1371/journal.pone.0226770>.
- [73] F. Huber, L. Ridder, S. Verhoeven, J.H. Spaaks, F. Diblen, S. Rogers, J.J.J. van der Hooft, Spec2Vec: Improved mass spectral similarity scoring through learning of structural relationships, *PLoS Comput Biol.* 17 (2021) e1008724. <https://doi.org/10.1371/journal.pcbi.1008724>.
- [74] F. Huber, S. van der Burg, J.J.J. van der Hooft, L. Ridder, MS2DeepScore: a novel deep learning similarity measure to compare tandem mass spectra, *J. Cheminform.* 13 (2021) 84. <https://doi.org/10.1186/s13321-021-00558-4>.
- [75] F. Wang, J. Liigand, S. Tian, D. Arndt, R. Greiner, D.S. Wishart, CFM-ID 4.0: More Accurate ESI-MS/MS Spectral Prediction and Compound Identification, *Anal. Chem.* 93 (2021) 11692–11700. <https://doi.org/10.1021/acs.analchem.1c01465>.
- [76] W. Bittremieux, M. Wang, P.C. Dorrestein, The critical role that spectral libraries play in capturing the metabolomics community knowledge, *Metabolomics.* 18 (2022) 94. <https://doi.org/10.1007/s11306-022-01947-y>.
- [77] H. Tsugawa, T. Cajka, T. Kind, Y. Ma, B. Higgins, K. Ikeda, M. Kanazawa, J. VanderGheynst, O. Fiehn, M. Arita, MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis, *Nat Methods.* 12 (2015) 523–526. <https://doi.org/10.1038/nmeth.3393>.
- [78] R. Schmid, S. Heuckeroth, A. Korf, A. Smirnov, O. Myers, T.S. Dyrland, R. Bushuiev, K.J. Murray, N. Hoffmann, M. Lu, A. Sarvepalli, Z. Zhang, M. Fleischauer, K. Dührkop, M. Wesner, S.J. Hoogstra, E. Rudt, O. Mokshyna, C. Brungs, K. Ponomarov, L. Mutabdžija, T. Damiani, C.J. Pudney, M. Earll, P.O. Helmer, T.R. Fallon, T. Schulze, A. Rivas-Ubach, A. Bilbao, H. Richter, L.-F. Nothias, M. Wang, M. Orešič, J.-K. Weng, S. Böcker, A. Jeibmann, H. Hayen, U. Karst, P.C. Dorrestein, D. Petras, X. Du, T. Pluskal, Integrative analysis of multimodal mass spectrometry data in MZmine 3, *Nat Biotechnol.* (2023) 1–3. <https://doi.org/10.1038/s41587-023-01690-2>.
- [79] J. Rainer, A. Vicini, L. Salzer, J. Stanstrup, J.M. Badia, S. Neumann, M.A. Stravs, V. Verri Hernandez, L. Gatto, S. Gibb, M. Witting, A Modular and Expandable Ecosystem for Metabolomics Data Annotation in R, *Metabolites.* 12 (2022) 173. <https://doi.org/10.3390/metabo12020173>.
- [80] B.B. Misra, New software tools, databases, and resources in metabolomics: updates from 2020, *Metabolomics.* 17 (2021) 49. <https://doi.org/10.1007/s11306-021-01796-1>.
- [81] Re3data.Org, Mass Spectrometry Interactive Virtual Environment, (2018) 10.546 Public Datasets. <https://doi.org/10.17616/R3980B>.
- [82] M. Sud, E. Fahy, D. Cotter, K. Azam, I. Vadivelu, C. Burant, A. Edison, O. Fiehn, R. Higashi, K.S. Nair, S. Sumner, S. Subramaniam, Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and

- analysis tools, *Nucleic Acids Research*. 44 (2016) D463–D470. <https://doi.org/10.1093/nar/gkv1042>.
- [83] K. Haug, K. Cochrane, V.C. Nainala, M. Williams, J. Chang, K.V. Jayaseelan, C. O’Donovan, *MetaboLights: a resource evolving in response to the needs of its scientific community*, *Nucleic Acids Res.* 48 (2020) D440–D444. <https://doi.org/10.1093/nar/gkz1019>.
- [84] M. Wang, A.K. Jarmusch, F. Vargas, A.A. Aksenov, J.M. Gauglitz, K. Weldon, D. Petras, R. da Silva, R. Quinn, A.V. Melnik, J.J.J. van der Hooft, A.M. Caraballo-Rodríguez, L.F. Nothias, C.M. Aceves, M. Panitchpakdi, E. Brown, F. Di Ottavio, N. Sikora, E.O. Elijah, L. Labarta-Bajo, E.C. Gentry, S. Shalapour, K.E. Kyle, S.P. Puckett, J.D. Watrous, C.S. Carpenter, A. Bouslimani, M. Ernst, A.D. Swafford, E.I. Zúñiga, M.J. Balunas, J.L. Klassen, R. Loomba, R. Knight, N. Bandeira, P.C. Dorrestein, *Mass spectrometry searches using MASST*, *Nat Biotechnol.* 38 (2020) 23–26. <https://doi.org/10.1038/s41587-019-0375-9>.
- [85] M.R. Viant, T.M.D. Ebbels, R.D. Beger, D.R. Ekman, D.J.T. Epps, H. Kamp, P.E.G. Leonards, G.D. Loizou, J.I. MacRae, B. van Ravenzwaay, P. Rocca-Serra, R.M. Salek, T. Walk, R.J.M. Weber, *Use cases, best practice and reporting standards for metabolomics in regulatory toxicology*, *Nat Commun.* 10 (2019) 3041. <https://doi.org/10.1038/s41467-019-10900-y>.
- [86] J.C. May, J.A. McLean, *Integrating ion mobility into comprehensive multidimensional metabolomics workflows: critical considerations*, *Metabolomics.* 18 (2022) 104. <https://doi.org/10.1007/s11306-022-01961-0>.
- [87] Y. Li, T. Kind, J. Folz, A. Vaniya, S.S. Mehta, O. Fiehn, *Spectral entropy outperforms MS/MS dot product similarity for small-molecule compound identification*, *Nat Methods.* 18 (2021) 1524–1531. <https://doi.org/10.1038/s41592-021-01331-z>.
- [88] M. Wang, J.J. Carver, V.V. Phelan, L.M. Sanchez, N. Garg, Y. Peng, D.D. Nguyen, J. Watrous, C.A. Kaponov, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A.V. Melnik, M.J. Meehan, W.-T. Liu, M. Crüsemann, P.D. Boudreau, E. Esquenazi, M. Sandoval-Calderón, R.D. Kersten, L.A. Pace, R.A. Quinn, K.R. Duncan, C.-C. Hsu, D.J. Floros, R.G. Gavilan, K. Kleigrewe, T. Northen, R.J. Dutton, D. Parrot, E.E. Carlson, B. Aigle, C.F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B.T. Murphy, L. Gerwick, C.-C. Liaw, Y.-L. Yang, H.-U. Humpf, M. Maansson, R.A. Keyzers, A.C. Sims, A.R. Johnson, A.M. Sidebottom, B.E. Sedio, A. Klitgaard, C.B. Larson, C.A. Boya P, D. Torres-Mendoza, D.J. Gonzalez, D.B. Silva, L.M. Marques, D.P. Demarque, E. Pociute, E.C. O’Neill, E. Briand, E.J.N. Helfrich, E.A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J.J. Kharbush, Y. Zeng, J.A. Vorholt, K.L. Kurita, P. Charusanti, K.L. McPhail, K.F. Nielsen, L. Vuong, M. Elfeki, M.F. Traxler, N. Engene, N. Koyama, O.B. Vining, R. Baric, R.R. Silva, S.J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P.G. Williams, J. Dai, R. Neupane, J. Gurr, A.M.C. Rodríguez, A. Lamsa, C. Zhang, K. Dorrestein, B.M. Duggan, J. Almaliti, P.-M. Allard, P. Phapale, L.-F. Nothias, T. Alexandrov, M. Litaudon, J.-L. Wolfender, J.E. Kyle, T.O. Metz, T. Peryea, D.-T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Müller, K.M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P.R. Jensen, B.Ø. Palsson, K. Pogliano, R.G. Linington, M. Gutiérrez, N.P. Lopes, W.H. Gerwick, B.S. Moore, P.C. Dorrestein, N. Bandeira, *Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking*, *Nat Biotechnol.* 34 (2016) 828–837. <https://doi.org/10.1038/nbt.3597>.
- [89] N.J. Morehouse, T.N. Clark, E.J. McMann, J.A. van Santen, F.P.J. Haeckl, C.A. Gray, R.G. Linington, *Annotation of natural product compound families using molecular networking topology and structural similarity fingerprinting*, *Nat Commun.* 14 (2023) 308. <https://doi.org/10.1038/s41467-022-35734-z>.
- [90] J.J.J. van der Hooft, J. Wandy, F. Young, S. Padmanabhan, K. Gerasimidis, K.E.V. Burgess, M.P. Barrett, S. Rogers, *Unsupervised Discovery and Comparison of Structural Families Across Multiple Samples in Untargeted Metabolomics*, *Anal. Chem.* 89 (2017) 7569–7577. <https://doi.org/10.1021/acs.analchem.7b01391>.
- [91] L.P. Mallmann, A. O. Rios, E. Rodrigues, *MS-FINDER and SIRIUS for phenolic compound identification from high-resolution mass spectrometry data*, *Food Research International.* 163 (2023) 112315. <https://doi.org/10.1016/j.foodres.2022.112315>.



- [92] C. Ruttkies, E.L. Schymanski, S. Wolf, J. Hollender, S. Neumann, MetFrag relaunched: incorporating strategies beyond in silico fragmentation, *Journal of Cheminformatics*. 8 (2016) 3. <https://doi.org/10.1186/s13321-016-0115-9>.
- [93] R. Schmid, D. Petras, L.-F. Nothias, M. Wang, A.T. Aron, A. Jagels, H. Tsugawa, J. Rainer, M. Garcia-Aloy, K. Dührkop, A. Korf, T. Pluskal, Z. Kameník, A.K. Jarmusch, A.M. Caraballo-Rodríguez, K.C. Weldon, M. Nothias-Esposito, A.A. Aksenov, A. Bauermeister, A. Albarracin Orio, C.O. Grundmann, F. Vargas, I. Koester, J.M. Gauglitz, E.C. Gentry, Y. Hövelmann, S.A. Kalinina, M.A. Pendergraft, M. Panitchpakdi, R. Tehan, A. Le Gouellec, G. Aleti, H. Mannocho Russo, B. Arndt, F. Hübner, H. Hayen, H. Zhi, M. Raffatellu, K.A. Prather, L.I. Aluwihare, S. Böcker, K.L. McPhail, H.-U. Humpf, U. Karst, P.C. Dorrestein, Ion identity molecular networking for mass spectrometry-based metabolomics in the GNPS environment, *Nat Commun*. 12 (2021) 3832. <https://doi.org/10.1038/s41467-021-23953-9>.
- [94] S. Han, W. Van Treuren, C.R. Fischer, B.D. Merrill, B.C. DeFelice, J.M. Sanchez, S.K. Higginbottom, L. Guthrie, L.A. Fall, D. Dodd, M.A. Fischbach, J.L. Sonnenburg, A metabolomics pipeline for the mechanistic interrogation of the gut microbiome, *Nature*. 595 (2021) 415–420. <https://doi.org/10.1038/s41586-021-03707-9>.
- [95] L. Zhao, F. Zhang, X. Ding, G. Wu, Y.Y. Lam, X. Wang, H. Fu, X. Xue, C. Lu, J. Ma, L. Yu, C. Xu, Z. Ren, Y. Xu, S. Xu, H. Shen, X. Zhu, Y. Shi, Q. Shen, W. Dong, R. Liu, Y. Ling, Y. Zeng, X. Wang, Q. Zhang, J. Wang, L. Wang, Y. Wu, B. Zeng, H. Wei, M. Zhang, Y. Peng, C. Zhang, Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes, *Science*. 359 (2018) 1151–1156. <https://doi.org/10.1126/science.aa5774>.
- [96] L. Hoyles, M.L. Jiménez-Pranteda, J. Chilloux, F. Brial, A. Myridakis, T. Aranas, C. Magnan, G.R. Gibson, J.D. Sanderson, J.K. Nicholson, D. Gauguier, A.L. McCartney, M.-E. Dumas, Metabolic retroconversion of trimethylamine N-oxide and the gut microbiota, *Microbiome*. 6 (2018) 73. <https://doi.org/10.1186/s40168-018-0461-0>.
- [97] A.D. Steen, A. Crits-Christoph, P. Carini, K.M. DeAngelis, N. Fierer, K.G. Lloyd, J. Cameron Thrash, High proportions of bacteria and archaea across most biomes remain uncultured, *ISME J*. 13 (2019) 3126–3130. <https://doi.org/10.1038/s41396-019-0484-y>.
- [98] S. Nayfach, Z.J. Shi, R. Seshadri, K.S. Pollard, N.C. Kyrpides, New insights from uncultivated genomes of the global human gut microbiome, *Nature*. 568 (2019) 505–510. <https://doi.org/10.1038/s41586-019-1058-x>.
- [99] M.G. Chevrette, C.S. Thomas, A. Hurley, N. Rosario-Meléndez, K. Sankaran, Y. Tu, A. Hall, S. Magesh, J. Handelsman, Microbiome composition modulates secondary metabolism in a multispecies bacterial community, *Proceedings of the National Academy of Sciences*. 119 (2022) e2212930119. <https://doi.org/10.1073/pnas.2212930119>.
- [100] D. Arora, P. Gupta, S. Jaglan, C. Roullier, O. Grovel, S. Bertrand, Expanding the chemical diversity through microorganisms co-culture: Current status and outlook, *Biotechnology Advances*. 40 (2020) 107521. <https://doi.org/10.1016/j.biotechadv.2020.107521>.
- [101] K. Blin, S. Shaw, A.M. Kloosterman, Z. Charlop-Powers, G.P. van Wezel, M.H. Medema, T. Weber, antiSMASH 6.0: improving cluster detection and comparison capabilities, *Nucleic Acids Research*. 49 (2021) W29–W35. <https://doi.org/10.1093/nar/gkab335>.
- [102] I. Cho, M.J. Blaser, The human microbiome: at the interface of health and disease, *Nat Rev Genet*. 13 (2012) 260–270. <https://doi.org/10.1038/nrg3182>.
- [103] C. Martino, A.H. Dillmore, Z.M. Burcham, J.L. Metcalf, D. Jeste, R. Knight, Microbiota succession throughout life from the cradle to the grave, *Nat Rev Microbiol*. 20 (2022) 707–720. <https://doi.org/10.1038/s41579-022-00768-z>.
- [104] C.Y. Kim, M. Lee, S. Yang, K. Kim, D. Yong, H.R. Kim, I. Lee, Human reference gut microbiome catalog including newly assembled genomes from under-represented Asian metagenomes, *Genome Medicine*. 13 (2021) 134. <https://doi.org/10.1186/s13073-021-00950-7>.
- [105] G.S. Navgire, N. Goel, G. Sawhney, M. Sharma, P. Kaushik, Y.K. Mohanta, T.K. Mohanta, A. Al-Harrasi, Analysis and Interpretation of metagenomics data: an approach, *Biol Proced Online*. 24 (2022) 18. <https://doi.org/10.1186/s12575-022-00179-7>.
- [106] S.S. Minot, A.D. Willis, Clustering co-abundant genes identifies components of the gut microbiome that are reproducibly associated with colorectal cancer and inflammatory bowel disease, *Microbiome*. 7 (2019) 110. <https://doi.org/10.1186/s40168-019-0722-6>.

- [107] K.L. Bowerman, S.F. Rehman, A. Vaughan, N. Lachner, K.F. Budden, R.Y. Kim, D.L.A. Wood, S.L. Gellatly, S.D. Shukla, L.G. Wood, I.A. Yang, P.A. Wark, P. Hugenholtz, P.M. Hansbro, Disease-associated gut microbiome and metabolome changes in patients with chronic obstructive pulmonary disease, *Nat Commun.* 11 (2020) 5886. <https://doi.org/10.1038/s41467-020-19701-0>.
- [108] P. Zheng, J. Wu, H. Zhang, S.W. Perry, B. Yin, X. Tan, T. Chai, W. Liang, Y. Huang, Y. Li, J. Duan, M.-L. Wong, J. Licinio, P. Xie, The gut microbiome modulates gut–brain axis glycerophospholipid metabolism in a region-specific manner in a nonhuman primate model of depression, *Mol Psychiatry.* 26 (2021) 2380–2392. <https://doi.org/10.1038/s41380-020-0744-2>.
- [109] C. Yang, D. Chowdhury, Z. Zhang, W.K. Cheung, A. Lu, Z. Bian, L. Zhang, A review of computational tools for generating metagenome-assembled genomes from metagenomic sequencing data, *Computational and Structural Biotechnology Journal.* 19 (2021) 6301–6314. <https://doi.org/10.1016/j.csbj.2021.11.028>.
- [110] J.J.J. van der Hoof, H. Mohimani, A. Bauermeister, P.C. Dorrestein, K.R. Duncan, M.H. Medema, Linking genomics and metabolomics to chart specialized metabolic diversity, *Chem. Soc. Rev.* 49 (2020) 3297–3314. <https://doi.org/10.1039/D0CS00162G>.
- [111] M.H. Medema, R. Kottmann, P. Yilmaz, M. Cummings, J.B. Biggins, K. Blin, I. de Bruijn, Y.H. Chooi, J. Claesen, R.C. Coates, P. Cruz-Morales, S. Duddela, S. Düsterhus, D.J. Edwards, D.P. Fewer, N. Garg, C. Geiger, J.P. Gomez-Escribano, A. Greule, M. Hadjithomas, A.S. Haines, E.J.N. Helfrich, M.L. Hillwig, K. Ishida, A.C. Jones, C.S. Jones, K. Jungmann, C. Kegler, H.U. Kim, P. Kötter, D. Krug, J. Masschelein, A.V. Melnik, S.M. Mantovani, E.A. Monroe, M. Moore, N. Moss, H.-W. Nützmann, G. Pan, A. Pati, D. Petras, F.J. Reen, F. Rosconi, Z. Rui, Z. Tian, N.J. Tobias, Y. Tsunematsu, P. Wiemann, E. Wyckoff, X. Yan, G. Yim, F. Yu, Y. Xie, B. Aigle, A.K. Apel, C.J. Balibar, E.P. Balskus, F. Barona-Gómez, A. Bechthold, H.B. Bode, R. Borriss, S.F. Brady, A.A. Brakhage, P. Caffrey, Y.-Q. Cheng, J. Clardy, R.J. Cox, R. De Mot, S. Donadio, M.S. Donia, W.A. van der Donk, P.C. Dorrestein, S. Doyle, A.J.M. Driessen, M. Ehling-Schulz, K.-D. Entian, M.A. Fischbach, L. Gerwick, W.H. Gerwick, H. Gross, B. Gust, C. Hertweck, M. Höfte, S.E. Jensen, J. Ju, L. Katz, L. Kaysser, J.L. Klassen, N.P. Keller, J. Kormanec, O.P. Kuipers, T. Kuzuyama, N.C. Kyrpides, H.-J. Kwon, S. Lautru, R. Lavigne, C.Y. Lee, B. Linquan, X. Liu, W. Liu, A. Luzhetskyy, T. Mahmud, Y. Mast, C. Méndez, M. Metsä-Ketelä, J. Micklefield, D.A. Mitchell, B.S. Moore, L.M. Moreira, R. Müller, B.A. Neilan, M. Nett, J. Nielsen, F. O’Gara, H. Oikawa, A. Osbourn, M.S. Osburne, B. Ostash, S.M. Payne, J.-L. Pernodet, M. Petricek, J. Piel, O. Ploux, J.M. Raaijmakers, J.A. Salas, E.K. Schmitt, B. Scott, R.F. Seipke, B. Shen, D.H. Sherman, K. Sivonen, M.J. Smanski, M. Sosio, E. Stegmann, R.D. Süßmuth, K. Tahlan, C.M. Thomas, Y. Tang, A.W. Truman, M. Viaud, J.D. Walton, C.T. Walsh, T. Weber, G.P. van Wezel, B. Wilkinson, J.M. Willey, W. Wohlleben, G.D. Wright, N. Ziemert, C. Zhang, S.B. Zotchev, R. Breitling, E. Takano, F.O. Glöckner, Minimum Information about a Biosynthetic Gene cluster, *Nat Chem Biol.* 11 (2015) 625–631. <https://doi.org/10.1038/nchembio.1890>.
- [112] E. Zazopoulos, K. Huang, A. Staffa, W. Liu, B.O. Bachmann, K. Nonaka, J. Ahlert, J.S. Thorson, B. Shen, C.M. Farnet, A genomics-guided approach for discovering and expressing cryptic metabolic pathways, *Nat Biotechnol.* 21 (2003) 187–190. <https://doi.org/10.1038/nbt784>.
- [113] A. de Jong, S.A.F.T. van Hijum, J.J.E. Bijlsma, J. Kok, O.P. Kuipers, BAGEL: a web-based bacteriocin genome mining tool, *Nucleic Acids Res.* 34 (2006) W273–W279. <https://doi.org/10.1093/nar/gkl237>.
- [114] T. Weber, C. Rausch, P. Lopez, I. Hoof, V. Gaykova, D.H. Huson, W. Wohlleben, CLUSEAN: a computer-based framework for the automated analysis of bacterial secondary metabolite biosynthetic gene clusters, *J Biotechnol.* 140 (2009) 13–17. <https://doi.org/10.1016/j.jbiotec.2009.01.007>.
- [115] M.H. Li, P.M. Ung, J. Zajkowski, S. Garneau-Tsodikova, D.H. Sherman, Automated genome mining for natural products, *BMC Bioinformatics.* 10 (2009) 185. <https://doi.org/10.1186/1471-2105-10-185>.
- [116] N. Khaldi, F.T. Seifuddin, G. Turner, D. Haft, W.C. Nierman, K.H. Wolfe, N.D. Fedorova, SMURF: genomic mapping of fungal secondary metabolite clusters, *Fungal Genet Biol.* 47 (2010) 736–741. <https://doi.org/10.1016/j.fgb.2010.06.003>.

- [117] M.A. Skinnider, N.J. Merwin, C.W. Johnston, N.A. Magarvey, PRISM 3: expanded prediction of natural product chemical structures from microbial genomes, *Nucleic Acids Res.* 45 (2017) W49–W54. <https://doi.org/10.1093/nar/gkx320>.
- [118] M.H. Medema, K. Blin, P. Cimermancic, V. de Jager, P. Zakrzewski, M.A. Fischbach, T. Weber, E. Takano, R. Breitling, antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences, *Nucleic Acids Research.* 39 (2011) W339–W346. <https://doi.org/10.1093/nar/gkr466>.
- [119] V. Pascal Andreu, H.E. Augustijn, L. Chen, A. Zhernakova, J. Fu, M.A. Fischbach, D. Dodd, M.H. Medema, gutSMASH predicts specialized primary metabolic pathways from the human gut microbiota, *Nat Biotechnol.* (2023) 1–8. <https://doi.org/10.1038/s41587-023-01675-1>.
- [120] S.R. Eddy, What is a hidden Markov model?, *Nat Biotechnol.* 22 (2004) 1315–1316. <https://doi.org/10.1038/nbt1004-1315>.
- [121] S.A. Kautsar, K. Blin, S. Shaw, J.C. Navarro-Muñoz, B.R. Terlouw, J.J.J. van der Hooft, J.A. van Santen, V. Tracanna, H.G. Suarez Duran, V. Pascal Andreu, N. Selem-Mojica, M. Alanjary, S.L. Robinson, G. Lund, S.C. Epstein, A.C. Sisto, L.K. Charkoudian, J. Collemare, R.G. Lington, T. Weber, M.H. Medema, MIBiG 2.0: a repository for biosynthetic gene clusters of known function, *Nucleic Acids Research.* 48 (2020) D454–D458. <https://doi.org/10.1093/nar/gkz882>.
- [122] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B.A. Shoemaker, P.A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, E.E. Bolton, PubChem in 2021: new data content and improved web interfaces, *Nucleic Acids Res.* 49 (2021) D1388–D1395. <https://doi.org/10.1093/nar/gkaa971>.
- [123] B.R. Terlouw, K. Blin, J.C. Navarro-Muñoz, N.E. Avalon, M.G. Chevrette, S. Egbert, S. Lee, D. Meijer, M.J.J. Recchia, Z.L. Reitz, J.A. van Santen, N. Selem-Mojica, T. Tørring, L. Zaroubi, M. Alanjary, G. Aleti, C. Aguilar, S.A.A. Al-Salihi, H.E. Augustijn, J.A. Avelar-Rivas, L.A. Avitia-Domínguez, F. Barona-Gómez, J. Bernaldo-Agüero, V.A. Bielinski, F. Biermann, T.J. Booth, V.J. Carrion Bravo, R. Castelo-Branco, F.O. Chagas, P. Cruz-Morales, C. Du, K.R. Duncan, A. Gavriilidou, D. Gayraud, K. Gutiérrez-García, K. Haslinger, E.J.N. Helfrich, J.J.J. van der Hooft, A.P. Jati, E. Kalkreuter, N. Kalyvas, K.B. Kang, S. Kautsar, W. Kim, A.M. Kunjapur, Y.-X. Li, G.-M. Lin, C. Loureiro, J.J.R. Louwen, N.L.L. Louwen, G. Lund, J. Parra, B. Philmus, B. Pourmohsenin, L.J.U. Pronk, A. Rego, D.A.B. Rex, S. Robinson, L.R. Rosas-Becerra, E.T. Roxborough, M.A. Schorn, D.J. Scobie, K.S. Singh, N. Sokolova, X. Tang, D. Udvary, A. Vigneshwari, K. Vind, S.P.J.M. Vromans, V. Waschulin, S.E. Williams, J.M. Winter, T.E. Witte, H. Xie, D. Yang, J. Yu, M. Zdouc, Z. Zhong, J. Collemare, R.G. Lington, T. Weber, M.H. Medema, MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters, *Nucleic Acids Res.* 51 (2023) D603–D610. <https://doi.org/10.1093/nar/gkac1049>.
- [124] J.C. Navarro-Muñoz, N. Selem-Mojica, M.W. Mullooney, S.A. Kautsar, J.H. Tryon, E.I. Parkinson, E.L.C. De Los Santos, M. Yeong, P. Cruz-Morales, S. Abubucker, A. Roeters, W. Lokhorst, A. Fernandez-Guerra, L.T.D. Cappelini, A.W. Goering, R.J. Thomson, W.W. Metcalf, N.L. Kelleher, F. Barona-Gomez, M.H. Medema, A computational framework to explore large-scale biosynthetic diversity, *Nat Chem Biol.* 16 (2020) 60–68. <https://doi.org/10.1038/s41589-019-0400-9>.
- [125] S.A. Kautsar, J.J.J. van der Hooft, D. de Ridder, M.H. Medema, BiG-SLiCE: A highly scalable tool maps the diversity of 1.2 million biosynthetic gene clusters, *Gigascience.* 10 (2021) giaa154. <https://doi.org/10.1093/gigascience/giaa154>.
- [126] J. Qin, R. Li, J. Raes, M. Arumugam, K.S. Burgdorf, C. Manichanh, T. Nielsen, N. Pons, F. Levenez, T. Yamada, D.R. Mende, J. Li, J. Xu, S. Li, D. Li, J. Cao, B. Wang, H. Liang, H. Zheng, Y. Xie, J. Tap, P. Lepage, M. Bertalan, J.-M. Batto, T. Hansen, D. Le Paslier, A. Linneberg, H.B. Nielsen, E. Pelletier, P. Renault, T. Sicheritz-Ponten, K. Turner, H. Zhu, C. Yu, S. Li, M. Jian, Y. Zhou, Y. Li, X. Zhang, S. Li, N. Qin, H. Yang, J. Wang, S. Brunak, J. Doré, F. Guarner, K. Kristiansen, O. Pedersen, J. Parkhill, J. Weissenbach, P. Bork, S.D. Ehrlich, J. Wang, A human gut microbial gene catalogue established by metagenomic sequencing, *Nature.* 464 (2010) 59–65. <https://doi.org/10.1038/nature08821>.
- [127] S.C. Forster, H.P. Browne, N. Kumar, M. Hunt, H. Denise, A. Mitchell, R.D. Finn, T.D. Lawley, HPMCD: the database of human microbial communities from metagenomic datasets and

- microbial reference genomes, *Nucleic Acids Research*. 44 (2016) D604–D609. <https://doi.org/10.1093/nar/gkv1216>.
- [128] S.D. Ehrlich, MetaHIT: The European Union Project on Metagenomics of the Human Intestinal Tract, in: K.E. Nelson (Ed.), *Metagenomics of the Human Body*, Springer, New York, NY, 2011: pp. 307–316. [https://doi.org/10.1007/978-1-4419-7089-3\\_15](https://doi.org/10.1007/978-1-4419-7089-3_15).
- [129] B.T. Tierney, Z. Yang, J.M. Lubber, M. Beaudin, M.C. Wibowo, C. Baek, E. Mehlenbacher, C.J. Patel, A.D. Kostic, The Landscape of Genetic Content in the Gut and Oral Human Microbiome, *Cell Host & Microbe*. 26 (2019) 283–295.e8. <https://doi.org/10.1016/j.chom.2019.07.008>.
- [130] M.B. Pereira, M. Wallroth, V. Jonsson, E. Kristiansson, Comparison of normalization methods for the analysis of metagenomic gene abundance data, *BMC Genomics*. 19 (2018) 274. <https://doi.org/10.1186/s12864-018-4637-6>.
- [131] D. Vandeputte, G. Kathagen, K. D’hoë, S. Vieira-Silva, M. Valles-Colomer, J. Sabino, J. Wang, R.Y. Tito, L. De Commer, Y. Darzi, S. Vermeire, G. Falony, J. Raes, Quantitative microbiome profiling links gut community variation to microbial load, *Nature*. 551 (2017) 507–511. <https://doi.org/10.1038/nature24460>.
- [132] R. Caspi, R. Billington, I.M. Keseler, A. Kothari, M. Krummenacker, P.E. Midford, W.K. Ong, S. Paley, P. Subhraveti, P.D. Karp, The MetaCyc database of metabolic pathways and enzymes - a 2019 update, *Nucleic Acids Res*. 48 (2020) D445–D453. <https://doi.org/10.1093/nar/gkz862>.
- [133] S. Vieira-Silva, G. Falony, Y. Darzi, G. Lima-Mendez, R. Garcia Yunta, S. Okuda, D. Vandeputte, M. Valles-Colomer, F. Hildebrand, S. Chaffron, J. Raes, Species–function relationships shape ecological properties of the human gut microbiome, *Nat Microbiol*. 1 (2016) 1–8. <https://doi.org/10.1038/nmicrobiol.2016.88>.
- [134] E.K. Morris, T. Caruso, F. Buscot, M. Fischer, C. Hancock, T.S. Maier, T. Meiners, C. Müller, E. Obermaier, D. Prati, S.A. Socher, I. Sonnemann, N. Wäschke, T. Wubet, S. Wurst, M.C. Rillig, Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories, *Ecol Evol*. 4 (2014) 3514–3524. <https://doi.org/10.1002/ece3.1155>.
- [135] J.G. Kers, E. Saccenti, The Power of Microbiome Studies: Some Considerations on Which Alpha and Beta Metrics to Use and How to Report Results, *Frontiers in Microbiology*. 12 (2022). <https://www.frontiersin.org/articles/10.3389/fmicb.2021.796025> (accessed March 30, 2023).
- [136] J.L. Anders, M.A.M. Moustafa, W.M.A. Mohamed, T. Hayakawa, R. Nakao, I. Koizumi, Comparing the gut microbiome along the gastrointestinal tract of three sympatric species of wild rodents, *Sci Rep*. 11 (2021) 19929. <https://doi.org/10.1038/s41598-021-99379-6>.
- [137] M. Arumugam, J. Raes, E. Pelletier, D. Le Paslier, T. Yamada, D.R. Mende, G.R. Fernandes, J. Tap, T. Bruls, J.-M. Batto, M. Bertalan, N. Borruel, F. Casellas, L. Fernandez, L. Gautier, T. Hansen, M. Hattori, T. Hayashi, M. Kleerebezem, K. Kurokawa, M. Leclerc, F. Levenez, C. Manichanh, H.B. Nielsen, T. Nielsen, N. Pons, J. Poulain, J. Qin, T. Sicheritz-Ponten, S. Tims, D. Torrents, E. Ugarte, E.G. Zoetendal, J. Wang, F. Guarner, O. Pedersen, W.M. de Vos, S. Brunak, J. Doré, J. Weissenbach, S.D. Ehrlich, P. Bork, Enterotypes of the human gut microbiome, *Nature*. 473 (2011) 174–180. <https://doi.org/10.1038/nature09944>.
- [138] P.I. Costea, F. Hildebrand, M. Arumugam, F. Bäckhed, M.J. Blaser, F.D. Bushman, W.M. de Vos, S.D. Ehrlich, C.M. Fraser, M. Hattori, C. Huttenhower, I.B. Jeffery, D. Knights, J.D. Lewis, R.E. Ley, H. Ochman, P.W. O’Toole, C. Quince, D.A. Relman, F. Shanahan, S. Sunagawa, J. Wang, G.M. Weinstock, G.D. Wu, G. Zeller, L. Zhao, J. Raes, R. Knight, P. Bork, Enterotypes in the landscape of gut microbial community composition, *Nat Microbiol*. 3 (2018) 8–16. <https://doi.org/10.1038/s41564-017-0072-8>.
- [139] MetaCardis Consortium, S. Vieira-Silva, G. Falony, E. Belda, T. Nielsen, J. Aron-Wisnewsky, R. Chakaroun, S.K. Forslund, K. Assmann, M. Valles-Colomer, T.T.D. Nguyen, S. Proost, E. Prifti, V. Tremaroli, N. Pons, E. Le Chatelier, F. Andreelli, J.-P. Bastard, L.P. Coelho, N. Galleron, T.H. Hansen, J.-S. Hulot, C. Lewinter, H.K. Pedersen, B. Quinquis, C. Rouault, H. Roume, J.-E. Salem, N.B. Søndertoft, S. Touch, M.-E. Dumas, S.D. Ehrlich, P. Galan, J.P. Götze, T. Hansen, J.J. Holst, L. Køber, I. Letunic, J. Nielsen, J.-M. Oppert, M. Stumvoll, H. Vestergaard, J.-D. Zucker, P. Bork, O. Pedersen, F. Bäckhed, K. Clément, J. Raes, Statin therapy is associated with lower prevalence of gut microbiota dysbiosis, *Nature*. 581 (2020) 310–315. <https://doi.org/10.1038/s41586-020-2269-x>.

- [140] N. Saji, K. Murotani, T. Hisada, T. Tsuduki, T. Sugimoto, A. Kimura, S. Niida, K. Toba, T. Sakurai, The relationship between the gut microbiome and mild cognitive impairment in patients without dementia: a cross-sectional study conducted in Japan, *Sci Rep.* 9 (2019) 19227. <https://doi.org/10.1038/s41598-019-55851-y>.
- [141] A.W. Goering, R.A. McClure, J.R. Doroghazi, J.C. Albright, N.A. Haverland, Y. Zhang, K.-S. Ju, R.J. Thomson, W.W. Metcalf, N.L. Kelleher, Metabologenomics: Correlation of Microbial Gene Clusters with Metabolites Drives Discovery of a Nonribosomal Peptide with an Unusual Amino Acid Monomer, *ACS Cent. Sci.* 2 (2016) 99–108. <https://doi.org/10.1021/acscentsci.5b00331>.
- [142] G.H. Eldjárn, A. Ramsay, J.J.J. van der Hooft, K.R. Duncan, S. Soldatou, J. Rousu, R. Daly, J. Wandy, S. Rogers, Ranking microbial metabolomic and genomic links in the NPLinker framework using complementary scoring functions, *PLOS Computational Biology.* 17 (2021) e1008920. <https://doi.org/10.1371/journal.pcbi.1008920>.
- [143] X. Zhu, M. Su, K. Manickam, W. Zhang, Bacterial Genome Mining of Enzymatic Tools for Alkyne Biosynthesis, *ACS Chem. Biol.* 10 (2015) 2785–2793. <https://doi.org/10.1021/acscembio.5b00641>.
- [144] M.G. Chevrette, F. Aicheler, O. Kohlbacher, C.R. Currie, M.H. Medema, SANDPUMA: ensemble predictions of nonribosomal peptide chemistry reveal biosynthetic diversity across Actinobacteria, *Bioinformatics.* 33 (2017) 3202–3210. <https://doi.org/10.1093/bioinformatics/btx400>.
- [145] B. Behsaz, E. Bode, A. Gurevich, Y.-N. Shi, F. Grundmann, D. Acharya, A.M. Caraballo-Rodríguez, A. Bouslimani, M. Panitchpakdi, A. Linck, C. Guan, J. Oh, P.C. Dorrestein, H.B. Bode, P.A. Pevzner, H. Mohimani, Integrating genomics and metabolomics for scalable non-ribosomal peptide discovery, *Nat Commun.* 12 (2021) 3225. <https://doi.org/10.1038/s41467-021-23502-4>.
- [146] C.W. Johnston, M.A. Skinnider, M.A. Wyatt, X. Li, M.R.M. Ranieri, L. Yang, D.L. Zechel, B. Ma, N.A. Magarvey, An automated Genomes-to-Natural Products platform (GNP) for the discovery of modular natural products, *Nat Commun.* 6 (2015) 8421. <https://doi.org/10.1038/ncomms9421>.
- [147] L. Cao, A. Gurevich, K.L. Alexander, C.B. Naman, T. Leão, E. Glukhov, T. Luzzatto-Knaan, F. Vargas, R. Quinn, A. Bouslimani, L.F. Nothias, N.K. Singh, J.G. Sanders, R.A.S. Benitez, L.R. Thompson, M.-N. Hamid, J.T. Morton, A. Mikheenko, A. Shlemov, A. Korobeynikov, I. Friedberg, R. Knight, K. Venkateswaran, W.H. Gerwick, L. Gerwick, P.C. Dorrestein, P.A. Pevzner, H. Mohimani, MetaMiner: A Scalable Peptidogenomics Approach for Discovery of Ribosomal Peptide Natural Products with Blind Modifications from Microbial Communities, *Cels.* 9 (2019) 600-608.e4. <https://doi.org/10.1016/j.cels.2019.09.004>.
- [148] A.M. Kloosterman, P. Cimermancic, S.S. Elsayed, C. Du, M. Hadjithomas, M.S. Donia, M.A. Fischbach, G.P. van Wezel, M.H. Medema, Expansion of RiPP biosynthetic space through integration of pan-genomics and machine learning uncovers a novel class of lanthipeptides, *PLoS Biol.* 18 (2020) e3001026. <https://doi.org/10.1371/journal.pbio.3001026>.
- [149] J.R. Doroghazi, J.C. Albright, A.W. Goering, K.-S. Ju, R.R. Haines, K.A. Tchalukov, D.P. Labeda, N.L. Kelleher, W.W. Metcalf, A roadmap for natural product discovery based on large-scale genomics and metabolomics, *Nat Chem Biol.* 10 (2014) 963–968. <https://doi.org/10.1038/nchembio.1659>.
- [150] P.C. Dorrestein, S.K. Mazmanian, R. Knight, Finding the Missing Links among Metabolites, Microbes, and the Host, *Immunity.* 40 (2014) 824–832. <https://doi.org/10.1016/j.immuni.2014.05.015>.
- [151] B.Y.L. Peisl, E.L. Schymanski, P. Wilmes, Dark matter in host-microbiome metabolomics: Tackling the unknowns—A review, *Analytica Chimica Acta.* 1037 (2018) 13–27. <https://doi.org/10.1016/j.aca.2017.12.034>.
- [152] M.A. Schorn, S. Verhoeven, L. Ridder, F. Huber, D.D. Acharya, A.A. Aksenov, G. Aleti, J.A. Moghaddam, A.T. Aron, S. Aziz, A. Bauermeister, K.D. Bauman, M. Baunach, C. Beemelmans, J.M. Beman, M.V. Berlanga-Clavero, A.A. Blacutt, H.B. Bode, A. Boullie, A. Brejnrod, T.S. Bugni, A. Calteau, L. Cao, V.J. Carrión, R. Castelo-Branco, S. Chanana, A.B. Chase, M.G. Chevrette, L.V. Costa-Lotufu, J.M. Crawford, C.R. Currie, B. Cuyper, T. Dang, T. de Rond,

- A.M. Demko, E. Dittmann, C. Du, C. Drozd, J.-C. Dujardin, R.J. Dutton, A. Edlund, D.P. Fewer, N. Garg, J.M. Gauglitz, E.C. Gentry, L. Gerwick, E. Glukhov, H. Gross, M. Gugger, D.G. Guillén Matus, E.J.N. Helfrich, B.-F. Hempel, J.-S. Hur, M. Iorio, P.R. Jensen, K.B. Kang, L. Kaysser, N.L. Kelleher, C.S. Kim, K.H. Kim, I. Koester, G.M. König, T. Leao, S.R. Lee, Y.-Y. Lee, X. Li, J.C. Little, K.N. Maloney, D. Männle, C. Martin H., A.C. McAvoy, W.W. Metcalf, H. Mohimani, C. Molina-Santiago, B.S. Moore, M.W. Mullowney, M. Muskat, L.-F. Nothias, E.C. O’Neill, E.I. Parkinson, D. Petras, J. Piel, E.C. Pierce, K. Pires, R. Reher, D. Romero, M.C. Roper, M. Rust, H. Saad, C. Saenz, L.M. Sanchez, S.J. Sørensen, M. Sosio, R.D. Süßmuth, D. Sweeney, K. Tahlan, R.J. Thomson, N.J. Tobias, A.E. Trindade-Silva, G.P. van Wezel, M. Wang, K.C. Weldon, F. Zhang, N. Ziemert, K.R. Duncan, M. Crüseemann, S. Rogers, P.C. Dorrestein, M.H. Medema, J.J.J. van der Hooft, A community resource for paired genomic and metabolomic data mining, *Nat Chem Biol.* 17 (2021) 363–368. <https://doi.org/10.1038/s41589-020-00724-z>.
- [153] Simone Zuffa, Robin Schmid, Anelize Bauermeister, Paulo Wender P. Gomes, Andres M. Caraballo-Rodriguez, Yasin El Abiead, Allegra T. Aron, Emily C. Gentry, Jasmine Zemlin, Michael J. Meehan, Nicole E. Avalon, Robert H. Cichewicz, Ekaterina Buzun, Marvic Carrillo Terrazas, Chia-Yun Hsu, Renee Oles, Adriana Vasquez Ayala, Jiaqi Zhao, Hiutung Chu, Mirte C. M. Kuijpers, Sara L. Jackrel, Fidele Tugizimana, Lerato Pertunia Nephali, Ian A. Dubery, Ntakadzeni Edwin Madala, Eduarda Antunes Moreira, Leticia Veras Costa-Lotufu, Norberto Peoporine Lopes, Paula Rezende-Teixeira, Paula C. Jimenez, Bipin Rimal, Andrew D. Patterson, Matthew F. Traxler, Rita de Cassia Pessotti, Daniel Alvarado-Villalobos, Giselle Tamayo-Castillo, Priscila Chaverri, Efrain Escudero-Leyva, Luis-Manuel Quiros-Guerrero, Alexandre Jean Bory, Juliette Joubert, Adriano Rutz, Jean-Luc Wolfender, Pierre-Marie Allard, Andreas Sichert, Sammy Pontrelli, Benjamin S Pullman, Nuno Bandeira, William H. Gerwick, Katia Gindro, Josep Massana-Codina, Berenike C. Wagner, Karl Forchhammer, Daniel Petras, Nicole Aiosa, Neha Garg, Manuel Liebeke, Patric Bourceau, Kyo Bin Kang, Henna Gadhavi, Luiz Pedro Sorio de Carvalho, Mariana Silva dos Santos, Alicia Isabel Pérez-Lorente, Carlos Molina-Santiago, Diego Romero, Raimo Franke, Mark Brønstrup, Arturo Vera Ponce de León, Phillip Byron Pope, Sabina Leanti La Rosa, Giorgia La Barbera, Henrik M. Roager, Martin Frederik Laursen, Fabian Hammerle, Bianka Siewert, Ursula Peintner, Cuauhtemoc Licon-Cassani, Lorena Rodriguez-Orduña, Evelyn Rampler, Felina Hildebrand, Gunda Koellensperger, Harald Schoeny, Katharina Hohenwallner, Lisa Panzenboeck, Rachel Gregor, Ellis Charles O’Neill, Eve Tallulah Roxborough, Jane Odoi, Nicole J. Bale, Su Ding, Jaap S. Sinninghe Damsté, Xueli Li Guan, Jerry J. Cui, Kou-San Ju, Denise Brentan Silva, Fernanda Motta Ribeiro Silva, Gilvan Ferreira da Silva, Hector H. F. Koolen, Carlismari Grundmann, Jason A. Clement, Hosein Mohimani, Kirk Broders, Kerry L. McPhail, Sidnee E. Ober-Singleton, Christopher M. Rath, Daniel McDonald, Rob Knight, Mingxun Wang, Pieter C. Dorrestein, A Taxonomically-informed Mass Spectrometry Search Tool for Microbial Metabolomics Data, *BioRxiv.* (2023) 2023.07.20.549584. <https://doi.org/10.1101/2023.07.20.549584>.
- [154] M.A. Wörheide, J. Krumsiek, G. Kastenmüller, M. Arnold, Multi-omics integration in biomedical research – A metabolomics-centric review, *Anal Chim Acta.* 1141 (2021) 144–162. <https://doi.org/10.1016/j.aca.2020.10.038>.
- [155] A. Vich Vila, V. Collij, S. Sanna, T. Sinha, F. Imhann, A.R. Bourgonje, Z. Mujagic, D.M.A.E. Jonkers, A.A.M. Masclee, J. Fu, A. Kurilshikov, C. Wijmenga, A. Zernakova, R.K. Weersma, Impact of commonly used drugs on the composition and metabolic function of the gut microbiota, *Nat Commun.* 11 (2020) 362. <https://doi.org/10.1038/s41467-019-14177-z>.
- [156] R.A. van den Berg, H.C.J. Hoefsloot, J.A. Westerhuis, A.K. Smilde, M.J. van der Werf, Centering, scaling, and transformations: improving the biological information content of metabolomics data, *BMC Genomics.* 7 (2006) 142. <https://doi.org/10.1186/1471-2164-7-142>.
- [157] A. Klåvus, M. Kokla, S. Noerman, V.M. Koistinen, M. Tuomainen, I. Zarei, T. Meuronen, M.R. Häkkinen, S. Rummukainen, A. Farizah Babu, T. Sallinen, O. Kärkkäinen, J. Paananen, D. Broadhurst, C. Brunius, K. Hanhineva, “Notame”: Workflow for Non-Targeted LC–MS Metabolic Profiling, *Metabolites.* 10 (2020) 135. <https://doi.org/10.3390/metabo10040135>.
- [158] S.K. Forslund, R. Chakaroun, M. Zimmermann-Kogadeeva, L. Markó, J. Aron-Wisnewsky, T. Nielsen, L. Moitinho-Silva, T.S.B. Schmidt, G. Falony, S. Vieira-Silva, S. Adriouch, R.J. Alves, K. Assmann, J.-P. Bastard, T. Birkner, R. Caesar, J. Chilloux, L.P. Coelho, L. Fezeu, N. Galleron,

- G. Helft, R. Isnard, B. Ji, M. Kuhn, E. Le Chatelier, A. Myridakis, L. Olsson, N. Pons, E. Prifti, B. Quinquis, H. Roume, J.-E. Salem, N. Sokolovska, V. Tremaroli, M. Valles-Colomer, C. Lewinter, N.B. Søndertoft, H.K. Pedersen, T.H. Hansen, J.P. Götze, L. Køber, H. Vestergaard, T. Hansen, J.-D. Zucker, S. Herberg, J.-M. Oppert, I. Letunic, J. Nielsen, F. Bäckhed, S.D. Ehrlich, M.-E. Dumas, J. Raes, O. Pedersen, K. Clément, M. Stumvoll, P. Bork, Combinatorial, additive and dose-dependent drug–microbiome associations, *Nature*. 600 (2021) 500–505. <https://doi.org/10.1038/s41586-021-04177-9>.
- [159] L. Cao, E. Shcherbin, H. Mohimani, A Metabolome- and Metagenome-Wide Association Network Reveals Microbial Natural Products and Microbial Biotransformation Products from the Human Microbiota, *MSystems*. 4 (2019) e00387-19. <https://doi.org/10.1128/mSystems.00387-19>.
- [160] M.L. Calle, Statistical Analysis of Metagenomics Data, *Genomics Inform*. 17 (2019) e6. <https://doi.org/10.5808/GI.2019.17.1.e6>.
- [161] A.V. Melnik, R.R. da Silva, E.R. Hyde, A.A. Aksenov, F. Vargas, A. Bouslimani, I. Protsyuk, A.K. Jarmusch, A. Tripathi, T. Alexandrov, R. Knight, P.C. Dorrestein, Coupling Targeted and Untargeted Mass Spectrometry for Metabolome-Microbiome-Wide Association Studies of Human Fecal Samples, *Anal. Chem*. 89 (2017) 7549–7559. <https://doi.org/10.1021/acs.analchem.7b01381>.
- [162] E.J. Battaglioli, V.L. Hale, J. Chen, P. Jeraldo, C. Ruiz-Mojica, B.A. Schmidt, V.M. Rekdal, L.M. Till, L. Huq, S.A. Smits, W.J. Moor, Y. Jones-Hall, T. Smyrk, S. Khanna, D.S. Pardi, M. Grover, R. Patel, N. Chia, H. Nelson, J.L. Sonnenburg, G. Farrugia, P.C. Kashyap, *Clostridioides difficile* uses amino acids associated with gut microbial dysbiosis in a subset of patients with diarrhea, *Sci Transl Med*. 10 (2018) eaam7019. <https://doi.org/10.1126/scitranslmed.aam7019>.
- [163] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, *BMC Bioinformatics*. 9 (2008) 559. <https://doi.org/10.1186/1471-2105-9-559>.
- [164] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, *Chemometrics and Intelligent Laboratory Systems*. 58 (2001) 109–130. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1).
- [165] J. Trygg, S. Wold, Orthogonal projections to latent structures (O-PLS), *Journal of Chemometrics*. 16 (2002) 119–128. <https://doi.org/10.1002/cem.695>.
- [166] J.M. Posma, I. Garcia-Perez, T.M.D. Ebbels, J.C. Lindon, J. Stamler, P. Elliott, E. Holmes, J.K. Nicholson, Optimized Phenotypic Biomarker Discovery and Confounder Elimination via Covariate-Adjusted Projection to Latent Structures from Metabolic Spectroscopy Data, *J. Proteome Res*. 17 (2018) 1586–1595. <https://doi.org/10.1021/acs.jproteome.7b00879>.
- [167] M.P.M. Letertre, A.P. Bhatt, M. Harvey, J.K. Nicholson, I.D. Wilson, M.R. Redinbo, J.R. Swann, Characterizing the metabolic effects of the selective inhibition of gut microbial  $\beta$ -glucuronidases in mice, *Sci Rep*. 12 (2022) 17435. <https://doi.org/10.1038/s41598-022-21518-4>.
- [168] M. Belgiu, L. Drăguț, Random forest in remote sensing: A review of applications and future directions, *ISPRS Journal of Photogrammetry and Remote Sensing*. 114 (2016) 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>.
- [169] A.M. Thomas, P. Manghi, F. Asnicar, E. Pasolli, F. Armanini, M. Zolfo, F. Beghini, S. Manara, N. Karcher, C. Pozzi, S. Gandini, D. Serrano, S. Tarallo, A. Francavilla, G. Gallo, M. Trompetto, G. Ferrero, S. Mizutani, H. Shiroma, S. Shiba, T. Shibata, S. Yachida, T. Yamada, J. Wirbel, P. Schrotz-King, C.M. Ulrich, H. Brenner, M. Arumugam, P. Bork, G. Zeller, F. Cordero, E. Dias-Neto, J.C. Setubal, A. Tett, B. Pardini, M. Rescigno, L. Waldron, A. Naccarati, N. Segata, Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation, *Nat Med*. 25 (2019) 667–678. <https://doi.org/10.1038/s41591-019-0405-7>.
- [170] F. Puig-Castellví, C. Midoux, A. Guenne, D. Conteau, O. Franchi, C. Bureau, C. Madigou, D. Jouan-Rimbaud Bouveresse, P. Kroff, L. Mazéas, D.N. Rutledge, G. Gaval, O. Chapleur, Metataxonomics, metagenomics and metabolomics analysis of the influence of temperature modification in full-scale anaerobic digesters, *Bioresource Technology*. 346 (2022) 126612. <https://doi.org/10.1016/j.biortech.2021.126612>.
- [171] F. Puig-Castellví, D. Jouan-Rimbaud Bouveresse, L. Mazéas, O. Chapleur, D.N. Rutledge, Rearrangement of incomplete multi-omics datasets combined with ComDim for evaluating

- replicate cross-platform variability and batch influence, *Chemometrics and Intelligent Laboratory Systems*. 218 (2021) 104422. <https://doi.org/10.1016/j.chemolab.2021.104422>.
- [172] J. Boccard, D.N. Rutledge, A consensus orthogonal partial least squares discriminant analysis (OPLS-DA) strategy for multiblock Omics data fusion, *Analytica Chimica Acta*. 769 (2013) 30–39. <https://doi.org/10.1016/j.aca.2013.01.022>.
- [173] J.T. Morton, A.A. Aksenov, L.F. Nothias, J.R. Foulds, R.A. Quinn, M.H. Badri, T.L. Swenson, M.W. Van Goethem, T.R. Northen, Y. Vazquez-Baeza, M. Wang, N.A. Bokulich, A. Watters, S.J. Song, R. Bonneau, P.C. Dorrestein, R. Knight, Learning representations of microbe–metabolite interactions, *Nat Methods*. 16 (2019) 1306–1314. <https://doi.org/10.1038/s41592-019-0616-3>.
- [174] H. Mallick, E.A. Franzosa, L.J. McIver, S. Banerjee, A. Sirota-Madi, A.D. Kostic, C.B. Clish, H. Vlamakis, R.J. Xavier, C. Huttenhower, Predictive metabolomic profiling of microbial communities using amplicon or metagenomic sequences, *Nat Commun*. 10 (2019) 3136. <https://doi.org/10.1038/s41467-019-10927-1>.
- [175] P. Andrikopoulos, J. Aron-Wisnewsky, R. Chakaroun, A. Myridakis, S.K. Forslund, T. Nielsen, S. Adriouch, B. Holmes, J. Chilloux, S. Vieira-Silva, G. Falony, J.-E. Salem, F. Andreelli, E. Belda, J. Kieswich, K. Chechi, F. Puig-Castellvi, M. Chevalier, E. Le Chatelier, M.T. Olanipekun, L. Hoyles, R. Alves, G. Helft, R. Isnard, L. Køber, L.P. Coelho, C. Rouault, D. Gauguier, J.P. Gøtze, E. Prifti, MetaCardis Consortium, J.-D. Zucker, F. Bäckhed, H. Vestergaard, T. Hansen, J.-M. Oppert, M. Blüher, J. Nielsen, J. Raes, P. Bork, M.M. Yaqoob, M. Stumvoll, O. Pedersen, S.D. Ehrlich, K. Clément, M.-E. Dumas, Evidence of a causal and modifiable relationship between kidney function and circulating trimethylamine N-oxide with implications for heart and kidney disorders, *MedRxiv*. (n.d.). <https://doi.org/10.1101/2022.11.08.22282073>.
- [176] D.N. Slenter, M. Kutmon, K. Hanspers, A. Riutta, J. Windsor, N. Nunes, J. Mélius, E. Cirillo, S.L. Coort, D. Digles, F. Ehrhart, P. Giesbertz, M. Kalafati, M. Martens, R. Miller, K. Nishida, L. Rieswijk, A. Waagmeester, L.M.T. Eijssen, C.T. Evelo, A.R. Pico, E.L. Willighagen, WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research, *Nucleic Acids Res*. 46 (2018) D661–D667. <https://doi.org/10.1093/nar/gkx1064>.
- [177] T.-M. Nguyen, A. Shafi, T. Nguyen, S. Draghici, Identifying significantly impacted pathways: a comprehensive review and assessment, *Genome Biol*. 20 (2019) 203. <https://doi.org/10.1186/s13059-019-1790-4>.
- [178] C. Noecker, A. Eng, S. Srinivasan, C.M. Theriot, V.B. Young, J.K. Jansson, D.N. Fredricks, E. Borenstein, Metabolic Model-Based Integration of Microbiome Taxonomic and Metabolomic Profiles Elucidates Mechanistic Links between Ecological and Metabolic Variation, *MSystems*. 1 (2016) e00013-15. <https://doi.org/10.1128/mSystems.00013-15>.
- [179] M. Shaffer, K. Thurimella, K. Quinn, K. Doenges, X. Zhang, S. Bokatzian, N. Reisdorph, C.A. Lozupone, AMON: annotation of metabolite origins via networks to integrate microbiome and metabolome data, *BMC Bioinformatics*. 20 (2019) 1–11. <https://doi.org/10.1186/s12859-019-3176-8>.
- [180] L. Heirendt, S. Arreckx, T. Pfau, S.N. Mendoza, A. Richelle, A. Heinken, H.S. Haraldsdóttir, J. Wachowiak, S.M. Keating, V. Vlasov, S. Magnusdóttir, C.Y. Ng, G. Preciat, A. Žagare, S.H.J. Chan, M.K. Aurich, C.M. Clancy, J. Modamio, J.T. Sauls, A. Noronha, A. Bordbar, B. Cousins, D.C. El Assal, L.V. Valcarcel, I. Apaolaza, S. Ghaderi, M. Ahookhosh, M. Ben Guebila, A. Kostromins, N. Sompairac, H.M. Le, D. Ma, Y. Sun, L. Wang, J.T. Yurkovich, M.A.P. Oliveira, P.T. Vuong, L.P. El Assal, I. Kuperstein, A. Zinovyev, H.S. Hinton, W.A. Bryant, F.J. Aragón Artacho, F.J. Planes, E. Stalidzans, A. Maass, S. Vempala, M. Hucka, M.A. Saunders, C.D. Maranas, N.E. Lewis, T. Sauter, B.Ø. Palsson, I. Thiele, R.M.T. Fleming, Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0, *Nat Protoc*. 14 (2019) 639–702. <https://doi.org/10.1038/s41596-018-0098-2>.
- [181] C. Diener, S.M. Gibbons, O. Resendis-Antonio, MICOM: Metagenome-Scale Modeling To Infer Metabolic Interactions in the Gut Microbiota, *MSystems*. 5 (2020) e00606-19. <https://doi.org/10.1128/mSystems.00606-19>.
- [182] S. Shoaie, P. Ghaffari, P. Kovatcheva-Datchary, A. Mardinoglu, P. Sen, E. Pujos-Guillot, T. de Wouters, C. Juste, S. Rizkalla, J. Chilloux, L. Hoyles, J.K. Nicholson, MICRO-Obes Consortium, J. Dore, M.E. Dumas, K. Clement, F. Bäckhed, J. Nielsen, Quantifying Diet-Induced Metabolic



- Changes of the Human Gut Microbiome, *Cell Metab.* 22 (2015) 320–331. <https://doi.org/10.1016/j.cmet.2015.07.001>.
- [183] M.M. Zaiss, H.-J. Joywu, D. Mauro, G. Schett, F. Ciccia, The gut–joint axis in rheumatoid arthritis, *Nat Rev Rheumatol.* 17 (2021) 224–237. <https://doi.org/10.1038/s41584-021-00585-3>.
- [184] A.D. Kostic, D. Gevers, H. Siljander, T. Vatanen, T. Hyötyläinen, A.-M. Hämäläinen, A. Peet, V. Tillmann, P. Pöhö, I. Mattila, H. Lähdesmäki, E.A. Franzosa, O. Vaarala, M. de Goffau, H. Harmsen, J. Ilonen, S.M. Virtanen, C.B. Clish, M. Orešič, C. Huttenhower, M. Knip, R.J. Xavier, The Dynamics of the Human Infant Gut Microbiome in Development and in Progression toward Type 1 Diabetes, *Cell Host & Microbe.* 17 (2015) 260–273. <https://doi.org/10.1016/j.chom.2015.01.001>.
- [185] H. Song, Y. Yoo, J. Hwang, Y.-C. Na, H.S. Kim, Faecalibacterium prausnitzii subspecies–level dysbiosis in the human gut microbiome underlying atopic dermatitis, *Journal of Allergy and Clinical Immunology.* 137 (2016) 852–860. <https://doi.org/10.1016/j.jaci.2015.08.021>.
- [186] W. Barcik, R.C.T. Boutin, M. Sokolowska, B.B. Finlay, The Role of Lung and Gut Microbiota in the Pathology of Asthma, *Immunity.* 52 (2020) 241–255. <https://doi.org/10.1016/j.immuni.2020.01.007>.
- [187] R.S. Mehta, J.R. Mayers, Y. Zhang, A. Bhosle, N.R. Glasser, L.H. Nguyen, W. Ma, S. Bae, T. Branck, K. Song, L. Sebastian, J.A. Pacheco, H.-S. Seo, C. Clish, S. Dhe-Paganon, A.N. Ananthakrishnan, E.A. Franzosa, E.P. Balskus, A.T. Chan, C. Huttenhower, Gut microbial metabolism of 5-ASA diminishes its clinical efficacy in inflammatory bowel disease, *Nat Med.* 29 (2023) 700–709. <https://doi.org/10.1038/s41591-023-02217-7>.
- [188] A. Lavelle, H. Sokol, Gut microbiota-derived metabolites as key actors in inflammatory bowel disease, *Nat Rev Gastroenterol Hepatol.* 17 (2020) 223–237. <https://doi.org/10.1038/s41575-019-0258-z>.
- [189] D.V. Zhernakova, T.H. Le, A. Kurilshikov, B. Atanasovska, M.J. Bonder, S. Sanna, A. Claringbould, U. Vösa, P. Deelen, L. Franke, R.A. de Boer, F. Kuipers, M.G. Netea, M.H. Hofker, C. Wijmenga, A. Zhernakova, J. Fu, Individual variations in cardiovascular-disease-related protein levels are driven by genetics and gut microbiome, *Nat Genet.* 50 (2018) 1524–1532. <https://doi.org/10.1038/s41588-018-0224-7>.
- [190] B. Ezzamouri, D. Rosario, G. Bidkhorji, S. Lee, M. Uhlen, S. Shoaie, Metabolic modelling of the human gut microbiome in type 2 diabetes patients in response to metformin treatment, *Npj Syst Biol Appl.* 9 (2023) 2. <https://doi.org/10.1038/s41540-022-00261-6>.
- [191] G. Park, S. Jung, K.E. Wellen, C. Jang, The interaction between the gut microbiota and dietary carbohydrates in nonalcoholic fatty liver disease, *Exp Mol Med.* 53 (2021) 809–822. <https://doi.org/10.1038/s12276-021-00614-x>.
- [192] H. Krukowski, S. Valkenburg, A.-M. Madella, J. Garssen, J. van Bergenhenegouwen, S.A. Overbeek, G.R.B. Huys, J. Raes, G. Glorieux, Gut microbiome studies in CKD: opportunities, pitfalls and therapeutic potential, *Nat Rev Nephrol.* 19 (2023) 87–101. <https://doi.org/10.1038/s41581-022-00647-z>.
- [193] J.A. O’Donnell, T. Zheng, G. Meric, F.Z. Marques, The gut microbiome and hypertension, *Nat Rev Nephrol.* 19 (2023) 153–167. <https://doi.org/10.1038/s41581-022-00654-0>.
- [194] L.H. Morais, H.L. Schreiber, S.K. Mazmanian, The gut microbiota–brain axis in behaviour and brain disorders, *Nat Rev Microbiol.* 19 (2021) 241–255. <https://doi.org/10.1038/s41579-020-00460-0>.
- [195] F. Zhu, Y. Ju, W. Wang, Q. Wang, R. Guo, Q. Ma, Q. Sun, Y. Fan, Y. Xie, Z. Yang, Z. Jie, B. Zhao, L. Xiao, L. Yang, T. Zhang, J. Feng, L. Guo, X. He, Y. Chen, C. Chen, C. Gao, X. Xu, H. Yang, J. Wang, Y. Dang, L. Madsen, S. Brix, K. Kristiansen, H. Jia, X. Ma, Metagenome-wide association of gut microbiome features for schizophrenia, *Nat Commun.* 11 (2020) 1612. <https://doi.org/10.1038/s41467-020-15457-9>.
- [196] D. Gkougka, K. Mitropoulos, G. Tzanakaki, E. Panagouli, T. Psaltopoulou, L. Thomaidis, M. Tsolia, T.N. Sergentanis, A. Tsitsika, Gut microbiome and attention deficit/hyperactivity disorder: a systematic review, *Pediatr Res.* 92 (2022) 1507–1519. <https://doi.org/10.1038/s41390-022-02027-6>.
- [197] M. Valles-Colomer, G. Falony, Y. Darzi, E.F. Tigchelaar, J. Wang, R.Y. Tito, C. Schiweck, A. Kurilshikov, M. Joossens, C. Wijmenga, S. Claes, L. Van Oudenhove, A. Zhernakova, S. Vieira-

- Silva, J. Raes, The neuroactive potential of the human gut microbiota in quality of life and depression, *Nat Microbiol.* 4 (2019) 623–632. <https://doi.org/10.1038/s41564-018-0337-x>.
- [198] C.X. Yap, A.K. Henders, G.A. Alvares, D.L.A. Wood, L. Krause, G.W. Tyson, R. Restuadi, L. Wallace, T. McLaren, N.K. Hansell, D. Cleary, R. Grove, C. Hafekost, A. Harun, H. Holdsworth, R. Jellett, F. Khan, L.P. Lawson, J. Leslie, M.L. Frenk, A. Masi, N.E. Mathew, M. Muniandy, M. Nothard, J.L. Miller, L. Nunn, G. Holtmann, L.T. Strike, G.I. de Zubicaray, P.M. Thompson, K.L. McMahon, M.J. Wright, P.M. Visscher, P.A. Dawson, C. Dissanayake, V. Eapen, H.S. Heussler, A.F. McRae, A.J.O. Whitehouse, N.R. Wray, J. Gratten, Autism-related dietary preferences mediate autism-gut microbiome associations, *Cell.* 184 (2021) 5916-5931.e17. <https://doi.org/10.1016/j.cell.2021.10.015>.
- [199] M.I. Butler, T.F.S. Bastiaanssen, C. Long-Smith, S. Morkl, K. Berding, N.L. Ritz, C. Strain, D. Patangia, S. Patel, C. Stanton, S.M. O’Mahony, J.F. Cryan, G. Clarke, T.G. Dinan, The gut microbiome in social anxiety disorder: evidence of altered composition and function, *Transl Psychiatry.* 13 (2023) 95. <https://doi.org/10.1038/s41398-023-02325-5>.
- [200] Z.D. Wallen, A. Demirkan, G. Twa, G. Cohen, M.N. Dean, D.G. Standaert, T.R. Sampson, H. Payami, Metagenomics of Parkinson’s disease implicates the gut microbiome in multiple disease mechanisms, *Nat Commun.* 13 (2022) 6958. <https://doi.org/10.1038/s41467-022-34667-x>.
- [201] C. Chen, J. Liao, Y. Xia, X. Liu, R. Jones, J. Haran, B. McCormick, T.R. Sampson, A. Alam, K. Ye, Gut microbiota regulate Alzheimer’s disease pathologies and cognitive disorders via PUFA-associated neuroinflammation, *Gut.* 71 (2022) 2233–2252. <https://doi.org/10.1136/gutjnl-2021-326269>.
- [202] M.A. Ortega, M.A. Álvarez-Mon, C. García-Montero, Ó. Fraile-Martínez, J. Monserrat, L. Martínez-Rozas, R. Rodríguez-Jiménez, M. Álvarez-Mon, G. Lahera, Microbiota–gut–brain axis mechanisms in the complex network of bipolar disorders: potential clinical implications and translational opportunities, *Mol Psychiatry.* (2023) 1–29. <https://doi.org/10.1038/s41380-023-01964-w>.
- [203] J. Correale, R. Hohlfeld, S.E. Baranzini, The role of the gut microbiota in multiple sclerosis, *Nat Rev Neurol.* 18 (2022) 544–558. <https://doi.org/10.1038/s41582-022-00697-8>.
- [204] Q. Yang, B. Wang, Q. Zheng, H. Li, X. Meng, F. Zhou, L. Zhang, A Review of Gut Microbiota-Derived Metabolites in Tumor Progression and Cancer Therapy, *Advanced Science.* 10 (2023) 2207366. <https://doi.org/10.1002/advs.202207366>.
- [205] F. Zhang, R.I. Lau, Q. Liu, Q. Su, F.K.L. Chan, S.C. Ng, Gut microbiota in COVID-19: key microbial changes, potential mechanisms and clinical applications, *Nat Rev Gastroenterol Hepatol.* (2022). <https://doi.org/10.1038/s41575-022-00698-4>.
- [206] H.R.T. Williams, I.J. Cox, D.G. Walker, B.V. North, V.M. Patel, S.E. Marshall, D.P. Jewell, S. Ghosh, H.J.W. Thomas, J.P. Teare, S. Jakobovits, S. Zeki, K.I. Welsh, S.D. Taylor-Robinson, T.R. Orchard, Characterization of inflammatory bowel disease with urinary metabolic profiling, *Am J Gastroenterol.* 104 (2009) 1435–1444. <https://doi.org/10.1038/ajg.2009.175>.
- [207] T. Pallister, M.A. Jackson, T.C. Martin, J. Zierer, A. Jennings, R.P. Mohney, A. MacGregor, C.J. Steves, A. Cassidy, T.D. Spector, C. Menni, Hippurate as a metabolomic marker of gut microbiome diversity: Modulation by diet and relationship to metabolic syndrome, *Sci Rep.* 7 (2017) 13670. <https://doi.org/10.1038/s41598-017-13722-4>.
- [208] F. Brial, J. Chilloux, T. Nielsen, S. Vieira-Silva, G. Falony, P. Andrikopoulos, M. Olanipekun, L. Hoyles, F. Djouadi, A.L. Neves, A. Rodriguez-Martinez, G.I. Mouawad, N. Pons, S. Forslund, E. Le-chatelier, A.L. Lay, J. Nicholson, T. Hansen, T. Hyötyläinen, K. Clément, M. Oresic, P. Bork, S.D. Ehrlich, J. Raes, O.B. Pedersen, D. Gauguier, M.-E. Dumas, Human and preclinical studies of the host–gut microbiome co-metabolite hippurate as a marker and mediator of metabolic health, *Gut.* 70 (2021) 2105–2114. <https://doi.org/10.1136/gutjnl-2020-323314>.
- [209] P. Konopelski, I. Mogilnicka, Biological Effects of Indole-3-Propionic Acid, a Gut Microbiota-Derived Metabolite, and Its Precursor Tryptophan in Mammals’ Health and Disease, *International Journal of Molecular Sciences.* 23 (2022) 1222. <https://doi.org/10.3390/ijms23031222>.
- [210] E. Serger, L. Luengo-Gutierrez, J.S. Chadwick, G. Kong, L. Zhou, G. Crawford, M.C. Danzi, A. Myridakis, A. Brandis, A.T. Bello, F. Müller, A. Sanchez-Vassopoulos, F. De Virgiliis, P. Liddell, M.E. Dumas, J. Strid, S. Mani, D. Dodd, S. Di Giovanni, The gut metabolite indole-3 propionate

- promotes nerve regeneration and repair, *Nature*. 607 (2022) 585–592. <https://doi.org/10.1038/s41586-022-04884-x>.
- [211] V.D. de Mello, J. Paananen, J. Lindström, M.A. Lankinen, L. Shi, J. Kuusisto, J. Pihlajamäki, S. Auriola, M. Lehtonen, O. Rolandsson, I.A. Bergdahl, E. Nordin, P. Ilanne-Parikka, S. Keinänen-Kiukaanniemi, R. Landberg, J.G. Eriksson, J. Tuomilehto, K. Hanhineva, M. Uusitupa, Indolepropionic acid and novel lipid metabolites are associated with a lower risk of type 2 diabetes in the Finnish Diabetes Prevention Study, *Sci Rep*. 7 (2017) 46337. <https://doi.org/10.1038/srep46337>.
- [212] C.A. Cason, K.T. Dolan, G. Sharma, M. Tao, R. Kulkarni, I.B. Helenowski, B.M. Doane, M.J. Avram, M.M. McDermott, E.B. Chang, C.K. Ozaki, K.J. Ho, Plasma microbiome-modulated indole- and phenyl-derived metabolites associate with advanced atherosclerosis and postoperative outcomes, *J Vasc Surg*. 68 (2018) 1552-1562.e7. <https://doi.org/10.1016/j.jvs.2017.09.029>.
- [213] Y. Zhu, Q. Li, H. Jiang, Gut microbiota in atherosclerosis: focus on trimethylamine N-oxide, *APMIS*. 128 (2020) 353–366. <https://doi.org/10.1111/apm.13038>.
- [214] J. Morze, C. Wittenbecher, L. Schwingshackl, A. Danielewicz, A. Rynkiewicz, F.B. Hu, M. Guasch-Ferré, Metabolomics and Type 2 Diabetes Risk: An Updated Systematic Review and Meta-analysis of Prospective Cohort Studies, *Diabetes Care*. 45 (2022) 1013–1024. <https://doi.org/10.2337/dc21-1705>.
- [215] A. Molinaro, I. Nemet, L.P. Bel, R. Chakaroun, T. Nielsen, -Wisnewsy Judith Aron, P.-O. Bergh, L. Li, M. Henricsson, øber L. K, R. Isnard, G. Helft, M. Stumvoll, O. Pedersen, J.G. Smith, W.H.W. Tang, ément K. Cl, S.L. Hazen, äckhed F. B, R. Alves, C. Amouyal, G.E.A. Andersson, F. Andreelli, O. Barthelemy, J.-P. Bastard, J.-P. Batisse, M. Berland, R. Bittar, üher M. Bl, P. Bork, O. Bourron, M. Camus, D. Cassuto, C. Ciangura, L.P. Coelho, J.-P. Collet, M.-E. Dumas, S.D. Ehrlich, L. Engelbrechtsen, L. Fezeu, S. Forslund, S. Fromentin, P. Galan, P. Giral, øtze J.P. G, T. Hansen, T.H. Hansen, A. Hartemann, B. Hartmann, S. Hercberg, B. Holmes, J.J. Holst, M. Hornbak, L. Hoyles, J.-S. Hulot, S. Jaqueminet, M. Kerneis, J. Khemis, R. Kozlowski, H.K. Pedersen, M. Kuhn, ås-H.L. Manner, L. Marko, -Gili Robin Massey Laura Martinez, N. Maziers, -Stamminger Jonathan Medina, -Silva Lucas Moitinho, G. Montalescot, S. Moutel, A.L. Neves, M. Olanipekun, J.-M. Oppert, C. Poitou, F. Pousset, L. Pouzoulet, C. Rouault, J. Silvain, H. Vestergaard, Microbially Produced Imidazole Propionate Is Associated With Heart Failure and Mortality, *JACC: Heart Failure*. 11 (2023) 810–821. <https://doi.org/10.1016/j.jchf.2023.03.008>.
- [216] K.A. Romano, I. Nemet, P. Prasad Saha, A. Haghikia, X.S. Li, M.L. Mohan, B. Lovano, L. Castel, M. Witkowski, J.A. Buffa, Y. Sun, L. Li, C.M. Menge, I. Demuth, M. König, E. Steinhagen-Thiessen, J.A. DiDonato, A. Deb, F. Bäckhed, W.H.W. Tang, S.V. Naga Prasad, U. Landmesser, D.R. Van Wagoner, S.L. Hazen, Gut Microbiota-Generated Phenylacetylglutamine and Heart Failure, *Circulation: Heart Failure*. 16 (2023) e009972. <https://doi.org/10.1161/CIRCHEARTFAILURE.122.009972>.
- [217] L.J. Rajakovich, B. Fu, M. Bollenbach, E.P. Balskus, Elucidation of an anaerobic pathway for metabolism of l-carnitine-derived  $\gamma$ -butyrobetaine to trimethylamine in human gut bacteria, *Proceedings of the National Academy of Sciences*. 118 (2021) e2101498118. <https://doi.org/10.1073/pnas.2101498118>.
- [218] J.A. Buffa, K.A. Romano, M.F. Copeland, D.B. Cody, W. Zhu, R. Galvez, X. Fu, K. Ward, M. Ferrell, H.J. Dai, S. Skye, P. Hu, L. Li, M. Parlov, A. McMillan, X. Wei, I. Nemet, R.A. Koeth, X.S. Li, Z. Wang, N. Sangwan, A.M. Hajjar, M. Dwidar, T.L. Weeks, N. Bergeron, R.M. Krauss, W.H.W. Tang, F.E. Rey, J.A. DiDonato, V. Gogonea, G.F. Gerberick, J.C. Garcia-Garcia, S.L. Hazen, The microbial gbu gene cluster links cardiovascular disease risk associated with red meat consumption to microbiota l-carnitine catabolism, *Nat Microbiol*. 7 (2021) 73–86. <https://doi.org/10.1038/s41564-021-01010-x>.
- [219] J. Żółkiewicz, A. Marzec, M. Ruszczyński, W. Feleszko, Postbiotics—A Step Beyond Pre- and Probiotics, *Nutrients*. 12 (2020) 2189. <https://doi.org/10.3390/nu12082189>.
- [220] J. Suez, N. Zmora, E. Segal, E. Elinav, The pros, cons, and many unknowns of probiotics, *Nat Med*. 25 (2019) 716–729. <https://doi.org/10.1038/s41591-019-0439-x>.
- [221] C. Depommier, A. Everard, C. Druart, H. Plovier, M. Van Hul, S. Vieira-Silva, G. Falony, J. Raes, D. Maiter, N.M. Delzenne, M. de Barsey, A. Loumaye, M.P. Hermans, J.-P. Thissen, W.M. de Vos, P.D. Cani, Supplementation with *Akkermansia muciniphila* in overweight and obese human

- volunteers: a proof-of-concept exploratory study, *Nat Med.* 25 (2019) 1096–1103. <https://doi.org/10.1038/s41591-019-0495-2>.
- [222] J. Ma, Y. Lyu, X. Liu, X. Jia, F. Cui, X. Wu, S. Deng, C. Yue, Engineered probiotics, *Microbial Cell Factories.* 21 (2022) 72. <https://doi.org/10.1186/s12934-022-01799-0>.
- [223] M.T. Khan, C. Dwibedi, D. Sundh, M. Pradhan, J.D. Kraft, R. Caesar, V. Tremaroli, M. Lorentzon, F. Bäckhed, Synergy and oxygen adaptation for development of next-generation probiotics, *Nature.* (2023) 1–5. <https://doi.org/10.1038/s41586-023-06378-w>.
- [224] A. Khoruts, C. Staley, M.J. Sadowsky, Faecal microbiota transplantation for *Clostridioides difficile*: mechanisms and pharmacology, *Nat Rev Gastroenterol Hepatol.* 18 (2021) 67–80. <https://doi.org/10.1038/s41575-020-0350-4>.
- [225] G. Ianiro, M. Punčochář, N. Karcher, S. Porcari, F. Armanini, F. Asnicar, F. Beghini, A. Blanco-Míguez, F. Cumbo, P. Manghi, F. Pinto, L. Masucci, G. Quaranta, S. De Giorgi, G.D. Sciumè, S. Bibbò, F. Del Chierico, L. Putignani, M. Sanguinetti, A. Gasbarrini, M. Valles-Colomer, G. Cammarota, N. Segata, Variability of strain engraftment and predictability of microbiome composition after fecal microbiota transplantation across different diseases, *Nat Med.* 28 (2022) 1913–1923. <https://doi.org/10.1038/s41591-022-01964-3>.
- [226] A.S. Kulkarni, S. Gubbi, N. Barzilai, Benefits of Metformin in Attenuating the Hallmarks of Aging, *Cell Metabolism.* 32 (2020) 15–30. <https://doi.org/10.1016/j.cmet.2020.04.001>.
- [227] J.M. Norman, S.A. Handley, M.T. Baldridge, L. Droit, C.Y. Liu, B.C. Keller, A. Kambal, C.L. Monaco, G. Zhao, P. Fleshner, T.S. Stappenbeck, D.P.B. McGovern, A. Keshavarzian, E.A. Mutlu, J. Sauk, D. Gevers, R.J. Xavier, D. Wang, M. Parkes, H.W. Virgin, Disease-specific alterations in the enteric virome in inflammatory bowel disease, *Cell.* 160 (2015) 447–460. <https://doi.org/10.1016/j.cell.2015.01.002>.
- [228] Y. Duan, C. Llorente, S. Lang, K. Brandl, H. Chu, L. Jiang, R.C. White, T.H. Clarke, K. Nguyen, M. Torralba, Y. Shao, J. Liu, A. Hernandez-Morales, L. Lessor, I.R. Rahman, Y. Miyamoto, M. Ly, B. Gao, W. Sun, R. Kiesel, F. Hutmacher, S. Lee, M. Ventura-Cots, F. Bosques-Padilla, E.C. Verna, J.G. Abraldes, R.S. Brown, V. Vargas, J. Altamirano, J. Caballería, D.L. Shawcross, S.B. Ho, A. Louvet, M.R. Lucey, P. Mathurin, G. Garcia-Tsao, R. Bataller, X.M. Tu, L. Eckmann, W.A. van der Donk, R. Young, T.D. Lawley, P. Stärkel, D. Pride, D.E. Fouts, B. Schnabl, Bacteriophage targeting of gut bacterium attenuates alcoholic liver disease, *Nature.* 575 (2019) 505–511. <https://doi.org/10.1038/s41586-019-1742-x>.
- [229] S. De Marco, M. Sichert, D. Muradyan, M. Piccioni, G. Traina, R. Pagiotti, D. Pietrella, Probiotic Cell-Free Supernatants Exhibited Anti-Inflammatory and Antioxidant Activity on Human Gut Epithelial Cells and Macrophages Stimulated with LPS, *Evid Based Complement Alternat Med.* 2018 (2018) 1756308. <https://doi.org/10.1155/2018/1756308>.
- [230] C. Lee, B.G. Kim, J.H. Kim, J. Chun, J.P. Im, J.S. Kim, Sodium butyrate inhibits the NF-kappa B signaling pathway and histone deacetylation, and attenuates experimental colitis in an IL-10 independent manner, *International Immunopharmacology.* 51 (2017) 47–56. <https://doi.org/10.1016/j.intimp.2017.07.023>.
- [231] N.E. Avalon, A.E. Murray, B.J. Baker, Integrated Metabolomic–Genomic Workflows Accelerate Microbial Natural Product Discovery, *Anal. Chem.* 94 (2022) 11959–11966. <https://doi.org/10.1021/acs.analchem.2c02245>.
- [232] D.A. Yee, K. Niwa, B. Perlatti, M. Chen, Y. Li, Y. Tang, Genome mining for unknown–unknown natural products, *Nat Chem Biol.* (2023) 1–8. <https://doi.org/10.1038/s41589-022-01246-6>.
- [233] F. Giambò, C. Costa, M. Teodoro, C. Fenga, Role-Playing Between Environmental Pollutants and Human Gut Microbiota: A Complex Bidirectional Interaction, *Frontiers in Medicine.* 9 (2022). <https://www.frontiersin.org/articles/10.3389/fmed.2022.810397> (accessed March 27, 2023).
- [234] N. Ashammakhi, R. Nasiri, N. Roberto De Barros, P. Tebon, J. Thakor, M. Goudie, A. Shamloo, M.G. Martin, A. Khademhosseni, Gut-on-a-chip: Current Progress and Future Opportunities, *Biomaterials.* 255 (2020) 120196. <https://doi.org/10.1016/j.biomaterials.2020.120196>.
- [235] H.J. Kim, J. Lee, J.-H. Choi, A. Bahinski, D.E. Ingber, Co-culture of Living Microbiome with Microengineered Human Intestinal Villi in a Gut-on-a-Chip Microfluidic Device, *JoVE (Journal of Visualized Experiments).* (2016) e54344. <https://doi.org/10.3791/54344>.

- [236] A. Ali, Y. Abouleila, Y. Shimizu, E. Hiyama, S. Emara, A. Mashaghi, T. Hankemeier, Single-cell metabolomics by mass spectrometry: Advances, challenges, and future applications, *TrAC Trends in Analytical Chemistry*. 120 (2019) 115436. <https://doi.org/10.1016/j.trac.2019.02.033>.
- [237] A.R. Buchberger, K. DeLaney, J. Johnson, L. Li, Mass Spectrometry Imaging: A Review of Emerging Advancements and Future Insights, *Anal. Chem.* 90 (2018) 240–265. <https://doi.org/10.1021/acs.analchem.7b04733>.
- [238] A.S. Rolig, R. Parthasarathy, A.R. Burns, B.J. Bohannon, K. Guillemin, Individual members of the microbiota disproportionately modulate host innate immune responses, *Cell Host Microbe*. 18 (2015) 613–620. <https://doi.org/10.1016/j.chom.2015.10.009>.
- [239] F. Mobeen, V. Sharma, P. Tulika, Enterotype Variations of the Healthy Human Gut Microbiome in Different Geographical Regions, *Bioinformation*. 14 (2018) 560–573. <https://doi.org/10.6026/97320630014560>.
- [240] N. Caballero-Casero, L. Belova, P. Vervliet, J.-P. Antignac, A. Castaño, L. Debrauwer, M.E. López, C. Huber, J. Klanova, M. Krauss, A. Lommen, H.G.J. Mol, H. Oberacher, O. Pardo, E.J. Price, V. Reinstadler, C.M. Vitale, A.L.N. van Nuijs, A. Covaci, Towards harmonised criteria in quality assurance and quality control of suspect and non-target LC-HRMS analytical workflows for screening of emerging contaminants in human biomonitoring, *TrAC Trends in Analytical Chemistry*. 136 (2021) 116201. <https://doi.org/10.1016/j.trac.2021.116201>.