in Microbiology
frontiers

Check for updates

Computational Biology and Machine Learning Approaches to Understand Mechanistic Microbiome-Host Interactions

Padhmanand Sudhakar[1,2,3]*, Kathleen Machiels[1], Bram Verstockt[1,4], Tamas Korcsmaros[2,3] and Séverine Vermeire[1,4]

[1] Department of Chronic Diseases, Metabolism and Ageing, Translational Research Center for Gastrointestinal Disorders (TARGID), KU Leuven, Leuven, Belgium, [2] Earlham Institute, Norwich, United Kingdom, [3] Quadram Institute Bioscience, Norwich, United Kingdom, [4] Department of Gastroenterology and Hepatology, University Hospitals Leuven, KU Leuven, Leuven, Belgium

The microbiome, by virtue of its interactions with the host, is implicated in various host functions including its influence on nutrition and homeostasis. Many chronic diseases such as diabetes, cancer, inflammatory bowel diseases are characterized by a disruption of microbial communities in at least one biological niche/organ system. Various molecular mechanisms between microbial and host components such as proteins, RNAs, metabolites have recently been identified, thus filling many gaps in our understanding of how the microbiome modulates host processes. Concurrently, high-throughput technologies have enabled the profiling of heterogeneous datasets capturing community level changes in the microbiome as well as the host responses. However, due to limitations in parallel sampling and analytical procedures, big gaps still exist in terms of how the microbiome mechanistically influences host functions at a system and community level. In the past decade, computational biology and machine learning methodologies have been developed with the aim of filling the existing gaps. Due to the agnostic nature of the tools, they have been applied in diverse disease contexts to analyze and infer the interactions between the microbiome and host molecular components. Some of these approaches allow the identification and analysis of affected downstream host processes. Most of the tools statistically or mechanistically integrate different types of -omic and meta -omic datasets followed by functional/biological interpretation. In this review, we provide an overview of the landscape of computational approaches for investigating mechanistic interactions between individual microbes/microbiome and the host and the opportunities for basic and clinical research. These could include but are not limited to the development of activity- and mechanism-based biomarkers, uncovering mechanisms for therapeutic interventions and generating integrated signatures to stratify patients.

Keywords: health, disease, microbiome-host interactions, molecular mechanisms, computational approaches, machine learning, basic and clinical research

INTRODUCTION: MICROBIOME-HOST INTERACTIONS

Across different niches and ecosystems, micro-organisms including bacteria, viruses, archaea inhabit a wide range of hosts (Braga et al., 2016). This community of microbes imparts various functions such as making nutrients accessible to the host (Martin et al., 2019), modulating the host immune system (Mendes et al., 2019), warding off pathogens (Pickard et al., 2017), maintaining homeostasis (Ohland and Jobin, 2015; Penny et al., 2018) among others. These functions are in turn driven primarily by molecular interactions between microbial and host molecules such as proteins, RNA and metabolites (Hughes and Sperandio, 2008; Braga et al., 2016). Deciphering these interactions could not only reveal the microbe-host cross-talk but also provide us with insights into formulating therapeutic strategies aimed at maintaining health and/or ameliorating disease states. The past decades have witnessed a surge in research interest to study microbial communities (and their interactions) which inhabit various niches – from the gut to the soil ecosystem. This was made possible by technological advancements leading to plummeting costs of 16S and metagenomic sequencing, higher sequencing depth and resolution (Levy and Myers, 2016; Jacob et al., 2019; Valli et al., 2020), novel *in vitro* systems (Shah et al., 2016; Eain et al., 2017; May et al., 2017), and new methodologies for high-throughput profiling of multiple -omic data types such as metaproteomics, metabolomics, lipidomics (Muller et al., 2013; Roume et al., 2015). However, due to many other limitations related to scale, scope, feasibility and sample availability for parallel omic read -outs, experimentally determining the inter-species microbe-host interactions is a challenging task (Fritz et al., 2013). Computational methods can overcome some of these limitations thereby enhancing our understanding of microbe-host interactions (Dix et al., 2016). In this review, we outline some key concepts, tools, and methods involved in computationally inferring the molecular mechanisms mediating microbe-host interactions.

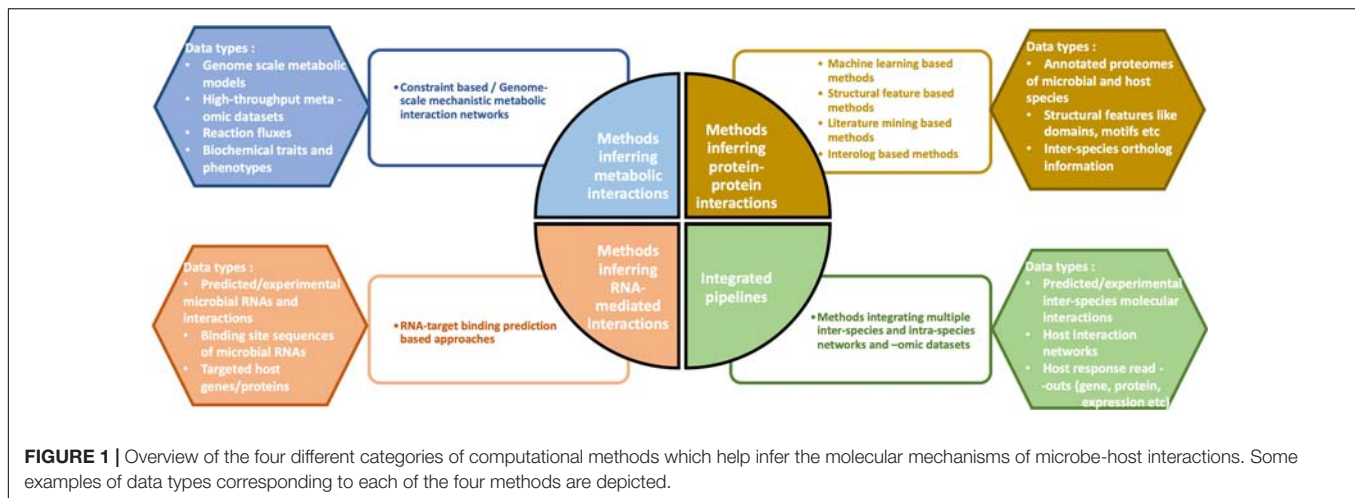# BIOLOGICAL NETWORKS: CONCEPTS AND APPLICATIONS

Biological networks represent relationships (termed edges) between any two biological entities (species, organisms, and molecules, etc.) which are usually called as nodes. At the level of molecules (genes, proteins, metabolites, RNAs, and small molecules, etc.), biological networks could either denote the physical interactions (e.g., protein–protein, protein-DNA, and RNA-protein, etc.) between molecules or any measure of association (e.g., co-expression and co-occurrence) between molecules (Gosak et al., 2018). In this paper, we will refer only to physical interactions. Physical interactions can be classified based on various criteria such as molecular types (protein–protein, protein-DNA, and RNA-protein, etc.), experimental scale (high-throughput or low-throughput), source (experimentally determined or computationally predicted), directionality (directed or undirected), relational signs (positive or negative relationships) and coverage (genome-wide or targeted). Since biological networks provide the larger context in which genes or proteins tend to exert their action, researchers can thereby fine-tune their hypotheses. Networks have largely been used in the domain of biological sciences (a) as a scaffold to integrate either singular or multiple contextual -omic datasets such as gene expression, proteomics, etc., measured in response to intrinsic or extrinsic stimuli (Charitou et al., 2016), (b) as a graph to trace potential signaling and regulatory pathways connecting any two nodes (Azeloglu and Iyengar, 2015), (c) to perform functional analysis at a local or global level (Emmert-Streib and Glazko, 2011), (d) to reconstruct the networks of non-model organisms from those of model organisms (Thompson et al., 2015), (e) to discover drug and disease targets (Huang et al., 2018), and (f) to infer globally or locally conserved signatures such as modules, motifs, etc (Wong et al., 2012). Various resources of molecular interactions and tools for integrative network analysis have been compiled and developed by the research community of network biologists. Since a very detailed description of the resources and tools is out of scope of the current review, readers are hereby referred to Pedamallu and Ozdamar (2014), Miryala et al. (2018), Romano et al. (2019).

Due to their utility in capturing contextual backgrounds and communication between molecular entities, biological networks have been used to not only study intra-species interactions but also inter-species cross-talks. Molecular ecological networks (Deng et al., 2012; Heleno et al., 2014) are a case in point by which the concept of networks are used to study the interactions between molecules (derived from different species or even kingdoms) in a larger ecological context (Yang et al., 2017; Meyer et al., 2020; Yu et al., 2020; Zheng et al., 2020). At the very core of it, a typical molecular ecological network inference workflow (Zhou et al., 2010; Deng et al., 2012; Chen et al., 2017) starts with the generation of meta -omic datasets (such as metagenomics, metatranscriptomics, and metaproteomics, etc.) followed by differential abundance testing between samples from contrasting conditions. Various measures of correlations and associations can then be applied to determine the distance between samples based on the differences and similarities in terms of the molecular features measured in the -omic datasets across the sample classes. Such correlations or associations can be used as a primary point of reference to investigate the possibility of mechanistic interactions which could in turn be driving the associative relationships. Furthermore, a network based representation of the feature-space can be used to compare samples with each other or to associate network properties such as the presence of motifs and modules to higher-level ecological traits/phenotypes. However, since molecular ecological networks do not directly infer molecular mechanisms which is the topic of this review, a detailed discussion on the topic is not undertaken.

COMPUTATIONAL METHODS IN MICROBIOME-HOST INTERACTIONS: FILLING THE GAPS

Computational methods bring in various advantages to the analysis of interactions between the host and individual microbes

FIGURE 1 | Overview of the four different categories of computational methods which help infer the molecular mechanisms of microbe-host interactions. Some examples of data types corresponding to each of the four methods are depicted.

and/or the microbial community. These include their attributes of (a) enhancing scalability, i.e., perform the computational inferences for a large number of variables and samples, (b) improving reproducibility (if complemented by inter-operability, automation, proper version control and sufficient documentation), (c) assessing performance by using a series of metrics, (d) shortlisting and prioritizing interactions, (e) and thereby (f) enabling the fine-tuning of hypothesis for experimental and/or epidemiological studies. Although most of the methods hitherto have focused on inferring the interactions between individual microbial species (mostly well studied pathogens) and the host, a few methods have been developed to predict the interactions at a community level. In principle, many of the methods which have been used to infer interactions of single species can be scaled up (with appropriate modifications) to infer community level interactions.

CLASSIFICATION OF COMPUTATIONAL METHODS IN MICROBIOME-HOST INTERACTIONS

From a mechanistic view-point, the most widely studied interaction types in interspecies cross-talks include (a) microbial metabolite-mediated networks, (b) protein–protein interactions (PPIs), and (c) RNA-mediated interactions. Accordingly, many of the computational methods developed to investigate microbe-host interactions have focused on the three above-mentioned interaction types (**Figure 1**). As a fourth method approach, integrated pipelines combine multiple microbial and host -omic data types and networks to infer the cumulative functional effects of inter-species interactions/communication on the host.

Approaches Inferring Mechanistic Metabolic Interactions

The metabolomic layer (which comprises the enzymes, metabolites, and the reactional interactions between them) has a prominent influence on both health and disease states

associated with alterations in microbiota composition (Wong et al., 2016; Martinez et al., 2017). Metabolic networks can thus represent and capture the underlying mechanisms driving various phenotypes (Pey et al., 2013; Samal et al., 2017; Zampieri et al., 2019). Computational approaches aimed at inferring the microbe-host co-metabolic networks can be classified into three prominent categories namely (a) Community-wide metabolic network modeling using metagenomic datasets: this approach is based on the assumption that the metagenomic read-outs represent the gene-distribution structure of the entire microbial community. The autonomy of species – i.e., information about which gene is derived from which species, are disregarded. Thus, the metabolic network reconstructed using this approach consists of relationships (reactions) catalyzed by enzymes (encoded by the measured genes) between molecular entities (metabolites) at a community level. (b) High throughput data driven approaches using metabolic datasets – this data-driven methodology uses targeted or untargeted profiling of metabolites from different groups of samples. Subsequently, multi-variate modeling methods and various statistical methods including simple PCAs are applied to identify biomarkers which distinguish different sample groups from each other. (c) Genome scale reconstruction applying constraint-based modeling approaches which are described below. The first two methods do not provide direct mechanistic insights and hence are not covered further in this review.

Genome-scale reconstruction models provide mechanistic information by integrating multiple inputs. These inputs include the curated genome scale metabolic models of both the host and microbial species, high-throughput meta -omic datasets including metabolites, reaction fluxes, biochemical traits and accessory phenotypic data. However, due to the strenuous nature of various steps involved in constructing the models and in scaling it up to multiple species or multiple hosts, only a handful of studies have applied this concept to infer microbe-host co-metabolic interactions (**Table 1**). The AGORA (assembly of gut organisms through reconstruction and analysis) collection is a resource of genome-scale metabolic models for 773 human gut bacterial species using a combination of metagenomics and

experimental data from literature. Furthermore, the framework employed by AGORA is amenable to scale-up given its easy adaptability to novel species of interest. AGORA also serves as a source of genome scale metabolic models reconstructed in a standardized manner. Thus, various studies have in turn used the genome scale models from the AGORA resource to construct context-specific models (Bauer et al., 2017; Bunesova et al., 2018; Tramontano et al., 2018; Pryor et al., 2019; Yilmaz et al., 2019). Recently, the authors of AGORA and their collaborators extended the framework to 7206 strains by incorporating information on the drug-metabolizing potential of the bacterial strains (Heinken et al., 2020).

The reported studies on genome-scale reconstruction models have been distributed across many different ecological contexts such as the human and rumen gut ecosystems (Islam et al., 2019), microbe-plant interactions, human alveolar macrophages, the effect of viral demands on the metabolism of human macrophages, microbe-host interactions in Parkinson's Disease to name a few. Due to the mechanistic nature of such models, they can be used as a template for further integrating other -omic datasets. This not only refines the models thereby increasing their predictive power but also assigns contextuality.

By incorporating the individual reconstructed metabolic models of tomato (*Solanum lycopersicum*) and the tomato late blight pathogen *Phytophthora infestans*, Rodenburg et al. (2019) pointed out specific pathways which mediate the dependencies of the pathogen on the metabolism of *S. lycopersicum*. The individual metabolic models for *S. lycopersicum* and *P. infestans* were derived by manually adding reactions and sub-cellular localization of metabolites and reactions (based on curation of literature) to the corresponding genome-scale models. Furthermore, by over-laying dual RNA-seq transcriptomic datasets from the host-pathogen duo into the co-metabolic network, various metabolic changes characterizing the scavenging nature of *P. infestans* were revealed. A similar study was performed in a mammalian setting wherein co-metabolic interactions and metabolic exchanges were inferred between the respiratory pathogen *Mycobacterium tuberculosis* and human alveolar macrophages (Bordbar et al., 2010). The metabolic model for the alveolar macrophages was derived from Recon1, the global human metabolic model (Thiele et al., 2013b). Briefly, a curated version of Recon1 was overlaid with gene expression data for healthy, inactivated alveolar macrophages and combined with information on flux limits for major pathways of central metabolism and a host of heterogeneous datasets such as immunohistological staining, transporter proteins, etc (Bordbar et al., 2010). The macrophage model was then combined with that of *Francisella tularensis* and corrected for compartment-specific reactions and metabolites. Unsurprisingly, given the advancement in terms of data generated and metabolic models made available, most of the genome-scale metabolic reconstruction studies (**Table 1**) were carried out for the gut ecosystem (Heinken et al., 2013; Heinken and Thiele, 2015; Ding et al., 2016; Islam et al., 2019).

Other microbe-host co-metabolic studies have been performed using publicly available tools based on constraint-based modeling approaches. The Constraint-based

TABLE 1 | Studies using genome-scale metabolic models and constraint based approaches to infer mechanistic co-metabolic interactions between microbial and host species.

Study	Context
Rodenburg et al. (2019)	Integrated metabolic model of *P. infestans* infecting tomato (*S. lycopersicum*)
Islam et al. (2019)	Genome-scale metabolic model between key members in the rumen microbiome and the viral phages
Hertel et al. (2019)	Integrated constraint-based model revealing microbe-host interactions in Parkinson's Disease
Aller et al. (2018)	Genome-scale model integrating biochemical demands arising from virus production and human macrophage cell metabolism
Ding et al. (2016)	Simulation of co-metabolic model of different enteropathogens in response to various host environments
Heinken and Thiele (2015)	*In silico* microbe-host gut co-metabolic model to predict effects of different host dietary schemes
Heinken et al. (2013)	Experimentally validated gut co-metabolic model between commensal bacterium *B. thetaiotaomicron* and mouse
Bordbar et al. (2010)	*Francisella tularensis* infecting human alveolar macrophage supported by high-throughput data from infected conditions

reconstruction and analysis (COBRA) toolbox (Heirendt et al., 2019) is one such compendium of methods containing various user-guided steps to reconstruct genome-scale metabolic models. It is characterized by properties such as interoperability, customized reconstruction, modeling, visualization, modeling, simulation, and integration of -omic datasets in various contexts (compartments, cell-types, etc.). By harnessing these properties, researchers have used the COBRA toolbox to model and investigate microbe-host metabolic interactions (Heinken et al., 2013; Thiele et al., 2013a) in the context of mammalian health with implications on human health. A representative study of the gut ecosystem using the COBRA toolbox integrated two previously published constraint-based models of mouse and a gut commensal *Bacteroides thetaiotaomicron* (Heinken et al., 2013). The *B. thetaiotaomicron* model was generated by the manual curation of a seed model produced by Model Seed (Henry et al., 2010) from the genome sequence annotated using RAST (Aziz et al., 2008) (which is a prokaryotic genome annotation tool). The mouse metabolic model was compiled by integrating a previously annotated and reconstructed model with gene essentiality data from experiments followed by corrections for duplicate reactions. The two models were then brought together by setting rules based on the subcellular localization of metabolites and reactions. The integrated metabolic model could capture many of the phenotypes exhibited *in vivo* namely the dependence of *B. thetaiotaomicron* on glycans derived from the metabolism of the host as well as the host diet itself (Heinken et al., 2013). It is noteworthy to mention that the authors also introduced novel methodologies such as Pareto analysis to complement the power of the COBRA toolbox. Pareto analysis is a bi-objective linear programming-based methodology which enables the analysis and identification of growth dependencies and trade-offs between the microbe and the host as captured by their metabolic networks.

A similar study (Hertel et al., 2019) was performed using the COBRA toolbox in conjunction with other supplementary tools such as the Microbiome Modeling Toolbox (Baldini et al., 2019) which can integrate the individual reconstructed models together into one reconstructed model in addition to other useful properties (such as inferring interactions by taxa, reconstruction of pairwise/community co-metabolic networks, compartment-based modeling, pareto analysis, and various downstream operations) to extend the constraint-based modeling framework. The study integrated the microbiome and longitudinal metabolomic datasets from patients with Parkinson's disease (Hertel et al., 2019). This microbiome-host -omic integration study provided clues as to how alterations in particular co-metabolized pathways (by both the host and microbiome) such as sulfur metabolism could contribute to the varying severity of the disease. In particular, the authors were able to identify that changes in the co-metabolized pathways could be driven by particular members of the gut microbiota. This opens up possibilities to design gut microbiome-based therapies to treat or even prevent Parkinson's disease.

Approaches Inferring Protein–Protein Interactions (PPIs)

Protein–protein interactions are one of the most well-studied interaction types mediating inter-species communication (Schweppe et al., 2015). Accordingly, a large number of computational microbe-host interaction studies have focused on PPIs. Congruently, PPI-based approaches have also been propelled by the adoption of concepts from other domains of computational biology and computational sciences in general. Hence, PPI-based approaches can be sub-classified into four predominant methods (**Table 2**) depending on the concepts used (1) Machine learning based PPI methods, (2) Structural feature based PPI methods, (3) Data/Literature mining based PPI methods, and (4) Interolog based PPI methods. In this section, we provide a brief overview of the concepts involved in each of these methods (**Table 2**) and provide a few representative examples.

Structural Feature Based PPI Methods

Interactions between proteins are usually a by-product of physical interactions between structural features of the proteins and/or could be characterized indirectly by co-occurring functional features of the proteins (Ding and Kihara, 2018). Structural features of the proteins include their domain and motif architectures/compositions, amino acid composition and frequencies, post-translational modification signatures, amino acid k-mers, mimicry motifs and 3D structural properties (Ding and Kihara, 2018). Structural feature-based PPI prediction, applied initially for intra-species PPIs, was subsequently extended to inter-species studies. Essentially, the fundamental principle on which structural feature-based PPI prediction methods work involves the use of mechanistic evidence between structural features to identify potentially interacting proteins. These could include for example interactions between domains, between domains and motifs, post-translational modifications and pairwise structural similarity (Ding and Kihara, 2018). Such structural studies have been confined to considerably well studied

TABLE 2 | Computational approaches and methods inferring protein–protein interactions mediating inter-kingdom cross-talk between microbial and host organisms.

Method and corresponding studies	Reported use-case (host-microbe)
Machine learning based methods	
Leite et al. (2018)	Bacteria–phage
Tastan et al. (2009); Qi et al. (2010), Dyer et al. (2011); Nouretdinov et al. (2012), Shoombuatong et al. (2012); Mei (2013), Hongjaisee et al. (2019)	Human–HIV
Kshirsagar et al. (2013)	Human–*F. tularensis,* Human–*Y. pestis,* Human–*B. anthracis,* Human–*S. typhi*
Wuchty (2011)	Human–*Plasmodium falciparum*
Kösesoy et al. (2019)	Human–*Y. pestis,* Human–*B. anthracis*
Cui et al. (2012); Emamjomeh et al. (2014), Kim et al. (2017)	Human–Hepatitis C virus
HOPITOR (Basit et al., 2018)	Generic (Human–virus PPIs)
Liao et al. (2011)	Human–*Schistosoma japonicum*
Mei et al. (2018); Sun et al. (2018)	Human–*Francisella tularensis*
Kargarfard et al. (2016)	3 hosts and 674 influenza strains
Cui et al. (2012); Dong et al. (2015), Kim et al. (2017)	Human–Human papillomavirus
Lai et al. (2012)	Human–Influenza A virus
Mei and Zhu (2014a)	Human–HTLV retroviruses
Mei and Zhu (2014b)	Human–*Salmonella*
Lian et al. (2019)	Human–*Y. pestis*
Structural feature based methods (features used)	
Dyer at al. (2007) (DDI)	Human–*Plasmodium falciparum*
Nourani et al. (2016) (DDI)	Human–multiple viruses
Sudhakar et al. (2019) (DDI and DMI)	Human–multiple bacterial pathogens
Doolittle and Gomez (2011) (PSS)	Human–Dengue virus, *Aedes aegypti*–Dengue virus
Cui et al. (2016) (PSS)	Human–HIV, Human–*Francisella tularensis*
P-HIPSTer (Lasso et al., 2019) (PSS)	Human–multiple viruses
Chen at al. (2019) (PSS)	Human–Dengue virus 2, Human–West Nile virus
Guven-Maiorov et al. (2017) (Mimicry)	Human–*Helicobacter pylori*
Mahajan and Mande (2017) (DDI)	Human–*Francisella tularensis*
Zhang et al. (2017a) (DMI)	Grass carp–Grass carp reovirus
Mehrotra et al. (2017) (PSS, DDI, and localization)	Human–*Leptospira interrogans,* Human–*Leptospira biflexa*
Halehalli and Nagarajaram (2015) (DDI, DMI)	Human–multiple viruses
SugarBindDB (Mariethoz et al., 2016) (glycan mediated PPIs)	Generic
Rajasekharan et al. (2013) (PSS)	Human–Chandipura virus
Carducci et al. (2010) (DDI)	Human–papillomavirus type 16
Franzosa and Xia (2011) (PSS and sequence identity)	Human–multiple viruses
Sahu et al. (2014) (DDI)	Arabidopsis-*Pseudomonas syringae*
Zhou et al. (2018) (DDI)	Human–Dengue virus, *Aedes aegypti*–Dengue virus
Kim et al. (2017) (DDI)	Human–multiple viruses
Kerr et al. (2015) (Computational docking)	Human–Dengue virus 2, Human–West Nile virus
Evans et al. (2009) (DMI)	Human–HIV
Doxey and McConkey (2013) (Mimicry)	Human–*Francisella tularensis*

(Continued)

TABLE 2 | Continued

Method and corresponding studies	Reported use-case (host-microbe)
Mei and Zhang (2020) (Mimicry)	Human-*S. typhimurium* and Human-Human respiratory syncytial virus
Data/Literature mining based methods	
Thieu et al. (2012)	Generic
Viruses.STRING (Cook et al., 2018)	319 hosts and 239 viruses
Li et al. (2018)	Human–Epstein-Barr virus
Saik et al. (2016)	Human–Hepatitis C virus
García-Pérez et al. (2018)	Human–Influenza A virus
"Interolog" based methods	
Krishnadev and Srinivasan (2008); Lee et al. (2008)	Human–*Plasmodium falciparum*
Krishnadev and Srinivasan (2011)	Human–*E. coli,* Human–*S. typhimurium,* Human–*Y. pestis*
Tyagi et al. (2009)	Human–*Helicobacter pylori*
Cui et al. (2016)	Human–HIV, Human–*Francisella tularensis*
Schleker et al. (2012)	Human–*Salmonella, Salmonella–A. thaliana*
Li et al. (2012)	*A. thaliana–Ralstonia solanacearum*
Wallqvist et al. (2017)	Human–*Coxiella burnetii*
Cuesta-Astroz et al. (2019)	Human and 15 eukaryotic parasites
Zhou et al. (2014); Cui et al. (2016)	Human–*Francisella tularensis*
Barh et al. (2013)	Human–*Corynebacterium pseudotuberculosis,* Human–*Corynebacterium diphtheriae,* Human–*Francisella tularensis,* Human–*Corynebacterium ulcerans,* Human–*Y. pestis,* and Human–*E. coli*

DDI, domain–domain interaction; DMI, domain-motif interaction; PSS, pairwise structural similarity. **Supplementary Table 1** *provides further details into the novelty of the methods and results.*
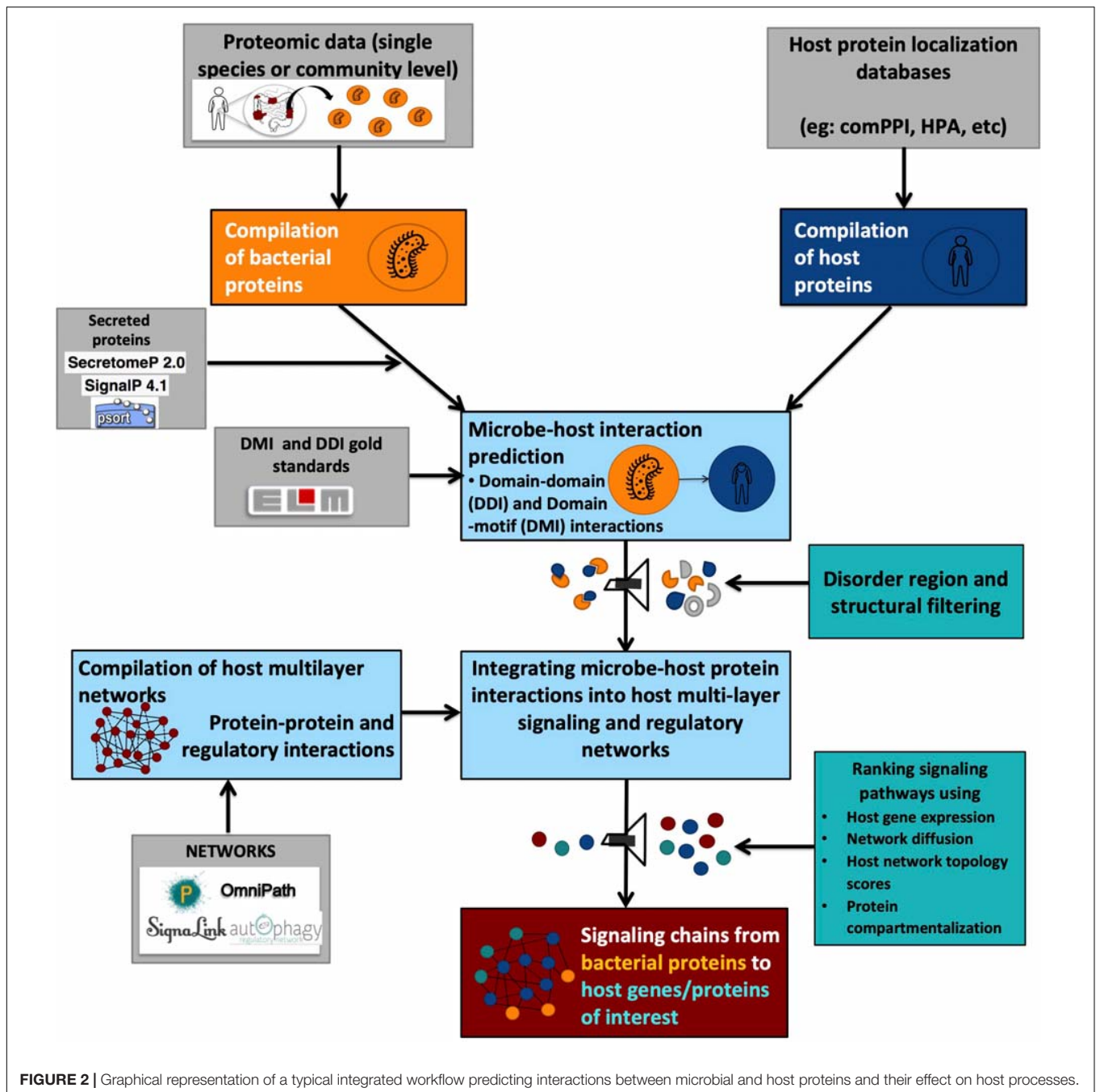
species pairs involving H. sapiens and prominent viral and bacterial pathogens (**Table 2**). Along with pairwise structural similarity-based methods using 3D protein complexes, domain–domain interaction (DDI) and domain-motif interaction (DMI) based methods are one of the most commonly used methods within the structural feature based methodological framework for predicting inter-species PPIs. Due to the ease of annotating domains and motifs, DDI- and DMI-based methods have been harnessed widely (**Table 2**). While DDI based methods have been applied to infer PPIs for a large number of species-pairs including Human–*Plasmodium falciparum* (Dyer et al., 2007), Human–*Francisella tularensis* (Zhou et al., 2013; Mahajan and Mande, 2017), Human–*Leptospira interrogans* (Mehrotra et al., 2017), Human–*Leptospira biflexa* (Mehrotra et al., 2017), Human–papillomavirus type 16 (Carducci et al., 2010), Arabidopsis–*Pseudomonas syringae* (Sahu et al., 2014), Rice–*Xanthomonas oryzae* (Kim et al., 2008), they have the inherent disadvantage of not being able to explicitly discern directionality.

On the other hand, DMIs provide directionality for PPIs, thus indicating the flow of signal transduction (Akiva et al., 2012; Gibson et al., 2015). For example, if a microbial protein A contains a domain known to be interacting with a motif on the host protein B, it is graphically represented as A > B, translating into "microbial protein A modulates host protein B." Due to their specificity, DMI-based methods are preferred over DDI based methods for research questions seeking to answer the role of post-translational modifications elicited on host proteins by microbial proteins or vice versa. However, due to the short sequence length of protein sequence motifs, even the most stringent search strategies have the tendency to result in thousands of false-positive hits while performing motif searches on a proteome-wide basis (Perkins et al., 2010; Idrees et al., 2018). Therefore, proper quality controls need to be applied to filter out false-positives based on structural properties such as the occurrence of truly interacting motifs within disordered regions and outside globular domains (Perkins et al., 2010; Idrees et al., 2018; **Figure 2**).

Several studies (**Table 2**) have been conducted to apply the principles of DMIs to predict PPIs for multiple microbe-host species-combinations including grass carp-grass carp reovirus (Zhang et al., 2017a), human-multiple bacterial pathogens (Sudhakar et al., 2019) and human-multiple viruses (Evans et al., 2009; Halehalli and Nagarajaram, 2015). By integrating DMI predictions between grass carp and grass carp reovirus (GCRV) proteins with differential gene expression and tissue-specific gene expression followed by functional enrichment, Zhang et al. (2017a) were able to pinpoint several signaling pathways modulated by GCRV. The authors also highlight an enrichment of host genes expressed in the intestinal niche suggesting that GCRV might have a higher influence on the gut. Recently, we conducted a study (Sudhakar et al., 2019) using DDI and DMI based methods to identify cross-talks between several bacterial pathogens including *Salmonella* and autophagy – a prominent biological process involved in host cellular homeostasis. Firstly, to identify microbial proteins targeted by selective autophagy, we scanned the bacterial proteins for the presence of the recognition motifs corresponding to the selective autophagy receptors p62 and NDP52 and the autophagy adapter protein LC3. Conversely, to infer the modulation of host autophagy by the bacterial pathogens, DMI and DDI based methods were used to identify the bacterial proteins which are able to bind to/modulate the 37 core autophagy host proteins. By overlapping the two above-mentioned sets of predictions, bacterial proteins involved in interplays were identified. Such bacterial proteins are also targeted by the host autophagy machinery for clearance and degradation. This was followed by experimentally verifying the effect on autophagy of a *Salmonella* protease involved in human-*Salmonella* interplay.

A variation of the motif-based methodologies is the use of motifs to characterize pathogen mimicry. This essentially involves the identification of eukaryotic linear motifs on microbial proteins which in turn can hijack host proteins and thereby promote antagonistic binding (Hurford and Day, 2013; Via et al., 2015). Motif-mediated molecular mimicry therefore rewires the host signaling and regulatory networks by titrating essential host proteins and enabling the microbe to create favorable micro-environments in the host cell by altering immune responses for example (Cusick et al., 2012). In addition to motifs, molecular mimicry can also be mediated at the level of protein, structural and interface levels. At the

FIGURE 2 | Graphical representation of a typical integrated workflow predicting interactions between microbial and host proteins and their effect on host processes.

protein level, specific studies investigating the role of molecular mimicry in the pathogenesis of prominent bacterial pathogens (Doxey and McConkey, 2013) including *Salmonella typhimurium* and *Human respiratory syncytial virus* (Mei and Zhang, 2020) have been carried out (**Table 2**). At the interface level, Guven-Maiorov et al. (2017) devised a computational method to infer mimicry induced by a prominent gastric cancer causing pathogen *Helicobacter pylori*. Besides DDI and DMI based methods, researchers have also used other structure-based methodologies such as pairwise structural similarity (PSS) to predict inter-species PPIs. PSS methods at their very core are based on

the premise that proteins possessing similar structures have a greater probability of interacting with the same set of protein partners (Ding and Kihara, 2018). This has been applied to infer the interactions with the host of various pathogens such as Dengue virus (Doolittle and Gomez, 2011), HIV (Cui et al., 2016), *Francisella tularensis* (Cui et al., 2016), West Nile virus (Chen et al., 2019), Chandipura virus (Rajasekharan et al., 2013), and other viral pathogens (Franzosa and Xia, 2011; Lasso et al., 2019).

As a means of ensuring proper quantitative evaluation of *de novo* PPI predictions, emerging computational methods such as

machine learning have been used in conjunction with structural-feature based PPI prediction methods. In order to avoid repetitions, methods using ML for evaluating the performance of structural feature dependent PPI predictions are discussed in the next subsection.

Machine Learning Based PPI Methods

Due to their ability to discern complex patterns among a large number of features in big datasets, machine learning (ML) methods have found favor in various applications of computational biology and bioinformatics (Shastry and Sanjay, 2020) including the prediction of microbe-host molecular interactions. A variety of supervised and unsupervised methods have been used to predict the interactions between microbial and host proteins (**Table 2**). In general, supervised machine learning methods utilize features from "gold-standard" interaction datasets to identify potential protein–protein interaction pairs from the user provided list of microbial and host proteins (Zhang et al., 2017b). In supervised methods, the "gold-standard" datasets are either compiled from high-throughput experimental methodologies or from curated lists of interactions from the literature (Zhang et al., 2017b). In the case of ML being used in combination with "interolog" based methods (explained in section 5.2.4), "gold-standard" PPI datasets can also be retrieved from other related or unrelated microbe-host species pairs depending on the scope of the study. Some of the features used to infer *de novo* PPI predictions include protein properties such as post-translational modifications, chemical composition, tissue distribution, molecular weight, domain/motif compositions, ontologies, gene expression, amino-acid frequencies, homology to human binding partners, and relevance of proteins in host network. By using these features, supervised methods are able to discern truly interacting protein pairs from all possible pairs of microbial and host proteins (Zhang et al., 2017b).

Supervised methods can also be differentiated by the kind of ML methodology/model used for the task of rightly classifying truly interacting protein pairs. Several supervised studies employing individual ML models [such as I2-regularized logistic regression (Mei et al., 2018), random forests (RF) (Kösesoy et al., 2019), etc], support vector machine (SVM) (Cui et al., 2012; Shoombuatong et al., 2012; Kim et al., 2017) have been applied to infer PPIs between microbial and host species. SVMs use a framework of searching and finding the best hyperplane (aka decision boundary represented by a mathematical equation) to separate sample with different labels corresponding to a class. Several variations of the SVM exist to handle data with underlying linear or non-linear relationships (Byvatov and Schneider, 2003).

Using four different ML models namely RF, SVM, Artificial Neural Networks (ANN) and K-Nearest Neighbors (K-NN), and multiple lines of -omic evidence including experimental PPIs as predictive features, Leite et al. (2018) devised a model based on a supervised protocol to accurately predict bacterium-phage interactions. The model, a type of ensemble learning, due to its generic nature, can also be used to predict interactions between any two given species, given the availability of informative feature sets. Ensemble learning (Che et al., 2011), combines multiple

individual classifiers to achieve a final classification and has been used to predict PPI based HIV-human and hepatitis C virus-human networks (Mei, 2013; Emamjomeh et al., 2014). Ensemble classification methods outperform individual classifiers based on several use-cases (Krawczyk, 2015; Haque et al., 2016; Yijing et al., 2016; Lin et al., 2019) and can be generalized into three distinct categories namely bagging, boosting and stacked generalization. The last of the three approaches, stacked generalization, was used by Emamjomeh et al. (2014) to predict PPIs between human and the hepatitis C virus. While bagging assigns training sets to individual classifiers based on a random selection of the initial training dataset with replacement for subsequent sampling runs, boosting involves the creation and evaluation of classifiers in a sequential manner, with the succeeding classifier assigning more weights to the misclassification errors committed by the preceding classifier. The "boosted" weights are then normalized for all the instances in the entire dataset which is then used as the training dataset for the next classifier after which the final classification step is carried out based on the weighted individual classifiers. The stacked generalization methodology is designed to overcome some of the errors committed by the individual classifiers even if they are used in the ensemble framework. The stacked approach achieves this by using a "stacks" of base learners so that its output is the input for a meta-learner which knows how best to combine the base learners' outputs. The training data may or may not overlap between the two stacks and can be specified accordingly.

Various auxiliary algorithms have been used in conjunction with machine learning methods to predict inter-species PPIs. An example of such a study includes the use of a novel protein sequence based feature extraction method called Location Based Encoding (LBE) with different classifier models including RFs. Such integrated methodologies have been used to predict protein interactions with the human host of two important pathogens – *Bacillus anthracis* and *Yersinia pestis* (Kösesoy et al., 2019). LBE is a methodology which complements the ML approaches for PPIs by differentiating proteins only based on the locations of the amino acids in the sequence (Li et al., 2009).

Supervised methods are sometimes constrained due to the small size of "gold-standard" datasets that restricts the inference and prediction of proteome-wide PPIs between the full list of proteins of any two given species. Mei and Zhu (2014a) harness the power of multi-instance AdaBoost, a type of boosting-based ensemble learning protocol, which is a multi-instance learning based ML method, to reconstruct proteome-wide Human T-cell leukemia virus-human PPI networks using homology knowledge derived protein features. AdaBoost improves classification performance by combining multiple weak classifiers into one strong classifier. It works in part by assigning more weight to instances which can only be classified with greater difficulty than to instances which can be easily classified (Kim et al., 2012). The dearth of true interacting protein-pairs has also prompted researchers to use unsupervised or semi-supervised approaches to infer microbe-host PPIs. Qi et al. (2010) complement the list of true interactions with a list of protein-pairs wherein association evidence exists with no interaction evidence between the proteins of a pair. Supervised learning is performed thereafter with a

multilayer perceptron network and by using the true interaction list. Subsequently, the semi-supervised approach uses the same network layers of the supervised classifier but instead trains on the protein-pairs with association evidence only. By using this hybrid approach, the authors report improved performance for predicting interactions between HIV and human proteins (Qi et al., 2010).

Data/Literature Mining Based PPI Methods

Even though many databases have been compiled to collect, curate and store microbe-host PPIs (Kumar and Nanduri, 2010; Durmus Tekir et al., 2013; Cook et al., 2018; Gao et al., 2018; Singh et al., 2019), these are mostly confined to well-studied pathogens and are predominantly comprised of interactions from high-throughput experiments. Contrastingly, in the literature, there exist inter-species PPIs from low-throughput experiments with some of them from non-model organisms, and commensal microbes, but mostly distributed over several individual studies. Very often, the inter-species PPI databases and repositories do not capture these sparse interactions. Hence, researchers have adapted and modified data- and text-mining tools to search for and extract microbe-host PPIs from existing literature. Retrieving such PPIs not only helps in increasing the number of true positive and true negative interactions (which helps aid the predictive performance of algorithms) but also extends our knowledge of existing microbe-host interactions. Motivated by the above explained need to mine-out microbe-host PPIs, Thieu et al. (2012) combine and compare the performance of a language based method based on a link grammar parser to a supervised ML methodology (SVM) and report that the combined approach results in a higher classification accuracy when compared to existing literature mining methods. As part of a bigger analytical framework aimed at uncovering the cellular mechanisms involved in human B lymphocytes during *Epstein-Barr virus* infection, Li et al. (2018) use a big-data mining methodology to identify a diverse range of inter-species molecular interactions including PPIs. Similar text/data mining approaches were also executed to extract PPI-mediated interactions of the human host with multiple viruses such as Hepatitis C virus (Saik et al., 2016) and Influenza A virus (García-Pérez et al., 2018; **Table 2**).

Interolog Based PPI Methods

For most species-pairs of interest, especially those belonging to the category of non-model organisms, there is a scarcity of experimentally verified PPIs. This has necessitated the development of novel bioinformatic methods, one of which is the inference of interactions from existing experimentally determined inter-species PPIs (Kshirsagar et al., 2015). These types of methodologies are usually based on the principle of homology (hence the term "interolog": meaning interacting orthologs) – either at the level of proteins or protein structural features or both. Protein features used for homology based extrapolation include but are not limited to domains, motifs, amino-acid k-mers, and 3D structural properties (Kshirsagar et al., 2015). Interolog based approaches have been applied to harness the large volume of experimentally verified PPIs

for model organisms including prominent bacterial/viral pathogens. Despite the potentially large coverage that can be achieved by such approaches, there exist several disadvantages of using interolog approaches as a silver bullet for inferring inter-species PPIs especially for novel species-pairs. These disadvantages are attributed to different pathogenic mechanisms between the microbes in the context of infecting different host species, different cellular localizations, and varying activity levels (expression, post-translational modifications, etc.) of the orthologous microbial proteins. Such differences lead to accessibility bottlenecks i.e., the ability of the proteins to physically access host proteins and thereby interact. Hence, interolog based approaches need to be complemented with additional filtering and quality control steps such as selecting proteins from infection-relevant cellular compartments, expression/activity measurements, etc.

Interolog based methods have been used to infer inter-species PPIs for many prominent pathogens and parasites (**Table 2**). Different versions of the interolog approach have been used to extrapolate PPIs corresponding to interactions between the human host and various pathogens such as *Plasmodium falciparum* (Krishnadev and Srinivasan, 2008; Lee et al., 2008), *Escherichia coli* (Krishnadev and Srinivasan, 2011), *S. typhimurium* (Krishnadev and Srinivasan, 2011; Schleker et al., 2012), *Y. pestis* (Krishnadev and Srinivasan, 2011), *Helicobacter pylori* (Tyagi et al., 2009), HIV (Cui et al., 2016), *Francisella tularensis* (Zhou et al., 2014; Cui et al., 2016), *Coxiella burnetii* (Wallqvist et al., 2017), *Corynebacterium pseudotuberculosis* (Barh et al., 2013), *Corynebacterium diphtheriae* (Barh et al., 2013), and *Corynebacterium ulcerans* (Barh et al., 2013). Using PPIs from the STRING database as the starting interaction set, Cuesta-Astroz et al. (2019) used the interolog methodology to predict PPIs between 15 different eukaryotic pathogens and the human host. To assign species-specific and lifecycle- specific contextuality, the authors confined the analysis to proteins from particular cellular compartments which are relevant to the infection process. From the analysis of the ensuing PPI networks, various invasion and evasion mechanisms adopted commonly and specifically by particular parasites were inferred (Cuesta-Astroz et al., 2019). Schleker et al. (2012) present another version of the interolog approach to predict human-*Salmonella* and *A. thaliana-Salmonella* PPI networks. As a source of template PPIs, publicly available interaction databases are used along with databases containing 3D structures between Pfam domains. As an add-on to the sequence based orthology of proteins, domain based orthology is also performed in order to reduce the false positive rates. Several additional filtering strategies such as restriction to predicted transmembrane proteins, relevance in host network and functional attributes such as gene ontology are used to make the PPIs more specific.

Approaches Inferring RNA Mediated Interactions

The role of RNAs, especially non-coding RNAs such as long non-coding RNAs (lncRNAs) and microRNAs (miRNAs) in mediating molecular microbe-host interactions have been

reported in the literature (Li et al., 2015b; Agliano et al., 2019). RNA molecules are either secreted by the microbial cell into the host cell or are packaged into vesicles along with other molecules which are then taken up by the host cell by endocytosis (Weiberg et al., 2014; Huang et al., 2019; Ahmadi Badi et al., 2020). Such microbial RNAs then modulate host cell activity by either binding to DNA, messenger RNAs or proteins. Thus, by salvaging and titrating host components, microbial RNAs modulate regulatory and signaling networks and subsequently host cell activity (Duval et al., 2017; Agliano et al., 2019; Shirahama et al., 2020). However, in contrast to PPI based methods, even though RNA-mediated microbe-host interactions are well studied from an experimental point of view, very few methods or studies exist that have systemically and systematically applied computational analysis (Table 3). As such, the resources which exist in the domain of RNA-mediated microbe-host interactions comprise of databases such as ViRBase (Li et al., 2015b) which is predominantly a source of experimentally verified virus–host non-coding RNA-associated interactions. In addition, it also contains predicted binding sites of virus non-coding RNAs on host proteins and RNAs. A prominent study which comprehensively examines and evaluates the role of RNAs in microbe-host interactions is that of Saçar Demirci and Adan (2020) who investigated the roles in infection of miRNA-like sequences encoded within the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) genome. They used a modified version of izMiR (Allmer et al., 2016), a SVM based ML method to predict pre-miRNAs which are homologous to the human precursor miRNAs from miRbase. The SVM based ML method identified several viral hairpin sequences which were smaller in length compared to the human miRNA precursors while many of the human and viral miRNA precursors were similar in length and shared identical minimum free energy, a feature used by the izMiR workflow (Allmer et al., 2016). Based on this observation, a revised classifier trained using only the known human miRNAs was used on the entire SARS-CoV-2 hairpin dataset which resulted in the identification of potential hairpins from which mature miRNA candidates were extracted. As a next step, the psRNATarget tool (Dai et al., 2018) was used to predict de novo the human genes targeted by the inferred viral miRNAs. Functional analysis of the human genes targeted revealed that the SARS-CoV-2miRNAs can affect various host processes including transcription, defense systems, Wnt and EGFR signaling pathways.

Approaches Utilizing Pipelines Integrating Multiple-Omic Datasets

Besides the computational methods based on particular types of molecular interactions, some integrated pipelines (Table 4) have been compiled to infer mechanistic microbe-host interactions. In general, such pipelines (Figure 2) incorporate the prediction of at least one molecular interaction type between microbial and host molecular components followed by various other functionalities such as integration of host responses. Table 5 provides a non-exhaustive overview of the different tools, databases and resources which are available in the public domain to compile integrated workflows based on PPIs for example.

KBase (Arkin et al., 2018) is an integrated bioinformatics platform enabling users to share datasets with the research community as well as facilitating the integration, and analysis of -omic datasets from microbes and plants by creating computational workflows. Recently, we developed MicrobioLink (Andrighetti et al., 2020), an integrated pipeline which carries out de novo DDI and DMI based microbe-host PPI prediction followed by quality control using information from disordered region predictions from built-in tools such as IUPred (Mészáros et al., 2018). The pipeline then utilizes network diffusion principles and tools (Paull et al., 2013) to infer the molecular mechanisms and signaling pathways which mediate the effect of microbial proteins on host responses as measured by transcriptomic or proteomic read-outs. Flexibility is provided for users to feed in the desired datasets at any given step of the pipeline. Given the advent of new computational tools in inter-species interactions and pipeline management platforms, it is expected that an increasing number of dedicated bioinformatic workflows for microbe-host interactions will be developed in the near future.

DISCUSSION: OPPORTUNITIES AND CHALLENGES

Opportunities
Clinical and Translational Research
Since the aforementioned computational tools help researchers narrow down on both microbial and host components involved in mechanistic cross-talks, the tools may discover molecules which can delineate different clinical phenotypes. In addition,

---

TABLE 3 | Examples of studies utilizing computational approaches to infer RNA-mediated interactions between microbes and hosts.

Study	Context
Saçar Demirci and Adan (2020)	Analysis revealing the potential interactions between mature micro-RNA like viral RNA sequences and host genes
ViRBase (Li et al., 2015b)	Source of experimentally verified virus–host non-coding RNA-associated interactions; also contains predicted binding sites of virus non-coding RNAs on host proteins and RNAs

---

TABLE 4 | Integrated pipelines used to infer microbe-host interactions by combining heterogeneous -omic datasets.

Methodology	Functionalities
MicrobioLink (Andrighetti et al., 2020)	Integrating microbe-host protein interaction networks with host responses and host regulatory/signaling networks using network diffusion principles
KBase (Arkin et al., 2018)	Integrated platform enabling data sharing, integration, and analysis of -omic datasets from microbes, plants, and their communities by creating computational workflows
Li et al. (2015a)	Identifying critical effectors involved in host-pathogen interactions by integrating multiple lines of -omic evidence

TABLE 5 | A non-exhaustive catalog of resources, tools and databases to compile protein–protein interaction based workflows for inferring microbe (microbiome)-host interactions.

Step in workflow	Resource/Tool/Database
Source of proteomes (sequence information)	UniProt (The UniProt Consortium, 2018), HumanPSD (Hodges et al., 2002), YPD (Payne and Garrels, 1997), PombePD (Costanzo et al., 2001), WormPD (Costanzo et al., 2001), and SWISS-PROT (Bairoch and Apweiler, 1996)
Source of proteomic datasets (expression information)	ProteomicsDB (Schmidt et al., 2018), Human Protein Atlas (HPA) (Thul and Lindskog, 2018), PRIDE (Perez-Riverol et al., 2019), PeptideAtlas (Desiere et al., 2006), MassIVE.quant (Choi et al., 2020), jPOSTrepo (Okuda et al., 2017), iProX (Ma et al., 2019), and Panorama Public (Sharma et al., 2018)
Proteomic annotations (structural features)	InterPro (Mitchell et al., 2019), Pfam (El-Gebali et al., 2019), ELM (Gouw et al., 2018), and PDB (Burley et al., 2017)
Protein sub-cellular localization (databases and prediction tools)	ComPPI (Veres et al., 2015), HPA (Thul and Lindskog, 2018), LocDB (Rastogi and Rost, 2011), LocSigDB (Negi et al., 2015), COMPARTMENTS (Binder et al., 2014), eSLDB (Pierleoni et al., 2007), SCLpred-EMS (Kaleel et al., 2020), DeepLoc (Almagro Armenteros et al., 2017), PSORTdb (Peabody et al., 2016), SecretomeP (Bendtsen et al., 2004), and Signal P (Armenteros et al., 2019)
Base information for prediction of PPIs	Domain-domain predictions – DOMINE (Raghavachari et al., 2008) and Domain-motif predictions – ELM (Gouw et al., 2018)
Quality control of inferred PPIs (using disordered region prediction)	IUPred (Mészáros et al., 2018), PrDOS (Ishida and Kinoshita, 2007), D2P2 (Oates et al., 2013), PONDR-FIT (Xue et al., 2010), DISOPRED (Ward et al., 2004), MFDp2 (Mizianty et al., 2013), and Meta-Disorder (Kozlowski and Bujnicki, 2012)
Network resources	OmniPath (Türei et al., 2016), IntAct (Orchard et al., 2014), Reactome (Fabregat et al., 2018), STRING (Szklarczyk et al., 2017), HTRI (Bovolenta et al., 2012), and DoRothEA (Garcia-Alonso et al., 2018)
Network diffusion approaches	NBS (Hofree et al., 2013), HotNet (Vandin et al., 2011), TieDie (Basha et al., 2013; Paull et al., 2013), RegMod (Qiu et al., 2010), and stSVM21 (Cun and Fröhlich, 2013)
Databases for host gene expression	GEO (Clough and Barrett, 2016) and ArrayExpress (Parkinson et al., 2007)

they can also be possible targets for therapeutic interventions. In other words, mechanistic predictions combined with clinical meta-data have a dual-purpose – they provide information on molecular components which could both represent and drive clinical phenotypes (Younesi, 2015) and thereby could potentially minimize our reliance on association-based biomarkers alone which need not explain causality (Levenson and Mori, 2014). The discovery of such mechanistic knowledge warrants the combinatorial use of different methodologies including machine learning and molecular interaction analysis. While many community level studies have been conducted on meta -omic datasets for the clinical classification of patients and the discovery of associative biomarkers (Wen et al., 2017; Yu et al., 2020; Clos-Garcia et al., 2019; Conteville et al., 2019), they have not incorporated mechanistic inferences. On the other hand, most mechanistic studies (**Tables 2**, **3**) have been carried out on

particular pathogens/microbial species without including clinical meta-data and/or clinical classifications.

Multi-omic approaches integrating heterogeneous -omic datasets from patients have been implemented for several diseases including IBD (Lloyd-Price et al., 2019) which are associated with microbial dysbiosis. However, these studies do not provide the required mechanistic insights for formulating therapeutic interventions. Beltran and Brito (2019) devised an integrated methodology to unravel the molecular mechanisms underlying the microbe-host interactions associated with various diseases such as colorectal cancer, IBD, obesity and type-2 diabetes. The aforementioned study represents one of the first and few initiatives to use community-wide microbe-host interaction predictions using meta -omic datasets from patients to discover mechanistic interactions driving the clinical phenotypes. By combining orthology based approaches to extrapolate interactions from experimental PPIs, machine learning and patient derived -omic datasets, the authors identified a subset of inter-species PPIs which are associated with disease phenotypes (Beltran and Brito, 2019). Thiele et al. (2020) published a novel study by integrating different levels of information (dietary information, physiological parameters, organ weights, and organ connectivities, etc.) and datasets such as molecular -omics (proteomics, metabolomics, metabolites produced by the gut microbiota) in an organ specific manner to arrive at a whole-body-model of human metabolism. Although not fully mechanistic, with this model, the authors were able to predict biomarkers of inherited metabolic diseases and host-microbiome co-metabolism. Such integrated studies and workflows combining statistical and mechanistic inference of multi -omic datasets awaits further adoption and application in the research on various diseases associated with microbial dysbiosis.

Research on Comparative Ecological Networks

The tools and resources listed in this review can be used to infer and predict molecular interactions between species in several contexts [microbe/microbiota in host, microbe/microbiota in several hosts, microbe (vs) microbe, and microbiota (vs) microbe, etc]. In almost all of the above-mentioned cases, molecular interactions between the autonomous entities (be it species or communities) could be driving the emergent phenotypes. Since the tools discussed in this manuscript also concern themselves with extrapolating interactions based on homology between species-pairs, it could be a right fit to predict *de novo* interaction relationships for species with very little experimental interaction information.

For example, Crohn's disease, a sub-type of IBD, is characterized by the dysbiosis of the gut microbiome (Joossens et al., 2011; Schaubeck et al., 2016; Shaw et al., 2016). This results in persistent inflammation of the gut mucosal barrier as a result of the unbalanced host responses (co-influenced by host genetic factors as well) to the dysbiosed microbiome and its various components such as proteins, metabolites, etc (Li et al., 2014; Lavelle and Sokol, 2020). Some of the CD patients also display lesions of the skin during or after therapeutic regimens (Huang et al., 2012; Gravina et al., 2016). It is known that the

skin also houses a complex microbial community which plays a role in maintaining homeostasis (Schommer and Gallo, 2013; Chen et al., 2018). Understanding the mechanisms by which CD medications impact the microbe-host interactions in the gut as well as the skin could help in avoiding the unintended side-effects of therapy in CD.

Yet another relevant context to apply the tools discussed herein is the inference of underlying molecular mechanisms which mediate the evasion of immune responses by bacterial pathogens in various hosts and their importance in transmission between hosts. We recently showed that bacterial pathogens and autophagy, a primary intracellular line of defense in the host, are engaged in an evolutionary tug of war, as evidenced by the presence of various interplays and cross-talks (Sudhakar et al., 2019). Given the exposure of host animals such as poultry and cattle to xenobiotic compounds such as antibiotics, many zoonotic pathogens are under constant selection pressure to evolve survival strategies to modulate/evade/survive within the host animal (Harada and Asai, 2010). This opens the door for impending risks of transmission (from animal hosts to human hosts or between various animal hosts) via the food chain of zoonotic species which have been selected for survival over many generations of persistence in the host (Farrell and Davies, 2019; Mollentze and Streicker, 2020). Microbe-host interaction mechanisms are at the evolutionary cross-roads of such transmission events between hosts. In this context, studying such interactions is expected to provide deeper insights into designing strategies to prevent and/or minimize spill-over transmission events.

Challenges

Over the past decade, various advances in the domain of computational analysis of microbe-host interactions have been made. However, despite this progress, there remain many challenges as described below. These challenges also present opportunities and the need to come up with innovative approaches and solutions.

Catching Up With Complex Infection Processes

Infection biology has taken new strides over the past years with new molecule classes (Katiyar-Agarwal and Jin, 2010; Rana et al., 2015; Duval et al., 2017; Long et al., 2017; Peters et al., 2019; Acuña et al., 2020) and cell-types (Chattopadhyay et al., 2018) being discovered as having a role in the infection process. With that, novel interaction types between various molecular classes are also unearthed (Silmon de Monerri and Kim, 2014). In some cases, computational methods have not caught up with molecular mechanisms. For example, hepadnaviruses utilize host DNA ligases to generate covalently closed circular DNAs which play a major role in mediating viral infection and persistence (Long et al., 2017). Similarly long non-coding RNAs are known to be involved in host-pathogen interactions (Duval et al., 2017; Agliano et al., 2019). However, till date, computational methods do not exist to predict or infer the mechanisms by which the viruses recruit the host DNA ligases or directly modulate the

biogenesis, conformation and activity of long non-coding RNAs. Hence, computational method developments are always a step behind the complexity associated with infection biology. This gap is all the more prevalent for commensal organisms in contrast to pathogens due to the constant and historically prevalent study bias.

Lack of Experimental Datasets

Non-model organisms and non-pathogenic organisms such as probiotics and commensals also suffer from a considerable knowledge gap in terms of known/experimentally verified molecular interactions. This affects the performance of computational methods considerably due to the need for large sets of true positives for the satisfactory performance and assessment of predictive algorithms (Jiao and Du, 2016). In addition, this also influences the coverage and accuracy of interolog approaches since they harness already existing true positive datasets for extrapolating to the species-pairs of interest based on orthology.

False-Positives

As with any computational algorithm, microbe-host interaction prediction methods also face the curse of false positives. This issue could be exacerbated by the availability of relatively small true positive (truly interacting) and true negative (non-interacting sets) datasets (Jiao and Du, 2016). Furthermore, the evolutionary distance and difference in infection process between the template species-pairs and the species-pair of interest as well as the absence of orthologous molecular components involved in the interactions could also contribute to the inflated false positive rates, reduced performance and coverage.

Community-Wide Interaction Prediction

Most of the microbe-host interaction computational tools have been directed at uncovering interactions corresponding to individual microbe-host pairs. This is a major drawback of existing methodologies, especially given the fact that phenotypes related to health and disease are associated with changes in community wide alterations (Clemente et al., 2012; Koboziev et al., 2014; Wang et al., 2017; Bailey and Holscher, 2018; Dominguez-Bello et al., 2019).

Modeling Dynamics of Microbe-Host Interactions

Last but not the least, current methods involved in microbe-host interaction analysis are not equipped to handle the dynamic nature of natural ecosystems and ecological niches in which the interactions are embedded. Although it is a generic drawback of many bioinformatic approaches, this challenge will need coordinated efforts between modelers, experimental biologists and bioinformaticians.

CONCLUSION

Since the advent and expansion of high-throughput sequencing technologies, various observational studies of microbial communities inhabiting various ecological niches (inside host

organisms for example) have been carried out. This has mostly resulted in associations with health- or disease-associated phenotypes. However, there is a huge gap in terms of the mechanisms mediated by these microbial communities and how these mechanisms contribute to the observed phenotypes. Despite the availability of experimental datasets which capture some of these mechanisms such as PPIs, these are either confined to model organisms or well-studied pathogens. Computational approaches provide researchers with the tools to upscale microbe-host interaction research by enabling them to make *de novo* inter-species molecular interactions and to extrapolate existing microbe-host interaction datasets to the species-pairs of interest. Computational methods may aid the study of microbe-host interaction by reducing the variable space, prioritizing interactions, and eventually building hypothesis for further experimental verification.

AUTHOR CONTRIBUTIONS

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.618856/full#supplementary-material

Supplementary Table 1 | Studies using genome-scale metabolic models and constraint based approaches to infer mechanistic co-metabolic interactions between microbial and host species.

REFERENCES

Acuña, S. M., Floeter-Winter, L. M., and Muxel, S. M. (2020). MicroRNAs: biological regulators in pathogen-host interactions. *Cells* 9:113. doi: 10.3390/cells9010113

Agliano, F., Rathinam, V. A., Medvedev, A. E., Vanaja, S. K., and Vella, A. T. (2019). Long noncoding RNAs in host-pathogen interactions. *Trends Immunol.* 40, 492–510. doi: 10.1016/j.it.2019.04.001

Ahmadi Badi, S., Bruno, S. P., Moshiri, A., Tarashi, S., Siadat, S. D., and Masotti, A. (2020). Small RNAs in outer membrane vesicles and their function in host-microbe interactions. *Front. Microbiol.* 11, 1209. doi: 10.3389/fmicb.2020.01209

Akiva, E., Friedlander, G., Itzhaki, Z., and Margalit, H. (2012). A dynamic view of domain-motif interactions. *PLoS Comput. Biol.* 8, e1002341. doi: 10.1371/journal.pcbi.1002341

Aller, S., Scott, A., Sarkar-Tyson, M., and Soyer, O. S. (2018). Integrated human-virus metabolic stoichiometric modelling predicts host-based antiviral targets against Chikungunya. Dengue and Zika viruses. *J. R. Soc. Interface* 15:20180125. doi: 10.1098/rsif.2018.0125

Allmer, J., Allmer, J., and Saçar Demirci, M. D. (2016). izMiR: computational ab initio microRNA detection. *Protoc. Exch. [Preprint].* doi: 10.1038/protex.2016.047

Almagro Armenteros, J. J., Sønderby, C. K., Sønderby, S. K., Nielsen, H., and Winther, O. (2017). DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics* 33, 3387–3395. doi: 10.1093/bioinformatics/btx431

Andrighetti, T., Bohar, B., Lemke, N., Sudhakar, P., and Korcsmaros, T. (2020). MicrobioLink: an integrated computational pipeline to infer functional effects of microbiome-host interactions. *Cells* 9:1278. doi: 10.3390/cells9051278

Arkin, A. P., Cottingham, R. W., Henry, C. S., Harris, N. L., Stevens, R. L., Maslov, S., et al. (2018). Kbase: the united states department of energy systems biology knowledgebase. *Nat. Biotechnol.* 36, 566–569. doi: 10.1038/nbt.4163

Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., et al. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37, 420–423. doi: 10.1038/s41587-019-0036-z

Azeloglu, E. U., and Iyengar, R. (2015). Signaling networks: information flow, computation, and decision making. *Cold Spring Harb. Perspect. Biol.* 7:a005934. doi: 10.1101/cshperspect.a005934

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. doi: 10.1186/1471-2164-9-75

Bailey, M. A., and Holscher, H. D. (2018). Microbiome-mediated effects of the mediterranean diet on inflammation. *Adv. Nutr.* 9, 193–206. doi: 10.1093/advances/nmy013

Bairoch, A., and Apweiler, R. (1996). The SWISS-PROT protein sequence data bank and its new supplement TREMBL. *Nucleic Acids Res.* 24, 21–25. doi: 10.1093/nar/24.1.21

Baldini, F., Heinken, A., Heirendt, L., Magnusdottir, S., Fleming, R. M. T., and Thiele, I. (2019). The microbiome modeling toolbox: from microbial interactions to personalized microbial communities. *Bioinformatics* 35, 2332–2334. doi: 10.1093/bioinformatics/bty941

Barh, D., Gupta, K., Jain, N., Khatri, G., León-Sicairos, N., Canizalez-Roman, A., et al. (2013). Conserved host-pathogen PPIs. globally conserved inter-species bacterial PPIs based conserved host-pathogen interactome derived novel target in *Corynebacterium pseudotuberculosis, Corynebacterium diphtheriae, Francisella tularensis, Corynebacterium ulcerans, Y. pestis,* and *E. coli* targeted by Piper betel compounds. *Integr. Biol. (Camb)* 5, 495–509. doi: 10.1039/c2ib20206a

Basha, O., Tirman, S., Eluk, A., and Yeger-Lotem, E. (2013). ResponseNet2.0: revealing signaling and regulatory pathways connecting your proteins and genes–now with human data. *Nucleic Acids Res.* 41, W198–W203. doi: 10.1093/nar/gkt532

Basit, A. H., Abbasi, W. A., Asif, A., Gull, S., and Minhas, F. U. A. A. (2018). Training host-pathogen protein-protein interaction predictors. *J. Bioinform. Comput. Biol.* 16:1850014. doi: 10.1142/S0219720018500142

Bauer, E., Zimmermann, J., Baldini, F., Thiele, I., and Kaleta, C. (2017). BacArena: individual-based metabolic modeling of heterogeneous microbes in complex communities. *PLoS Comput. Biol.* 13:e1005544. doi: 10.1371/journal.pcbi.1005544

Bendtsen, J. D., Jensen, L. J., Blom, N., Von Heijne, G., and Brunak, S. (2004). Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng. Des. Sel.* 17, 349–356. doi: 10.1093/protein/gzh037

Beltran, J. F., and Brito, I. (2019). Host-microbiome protein-protein interactions capture mechanisms in human disease. *BioRxiv*. doi: 10.1101/821926

Binder, J. X., Pletscher-Frankild, S., Tsafou, K., Stolte, C., O'Donoghue, S. I., Schneider, R., et al. (2014). COMPARTMENTS: unification and visualization of protein subcellular localization evidence. *Database (Oxford)* 2014:bau012. doi: 10.1093/database/bau012

Bordbar, A., Lewis, N. E., Schellenberger, J., Palsson, B. Ø, and Jamshidi, N. (2010). Insight into human alveolar macrophage and *Francisella tularensis* interactions via metabolic reconstructions. *Mol. Syst. Biol.* 6:422. doi: 10.1038/msb.2010.68

Bovolenta, L. A., Acencio, M. L., and Lemke, N. (2012). HTRIdb: an open-access database for experimentally verified human transcriptional regulation interactions. *BMC Genomics* 13:405. doi: 10.1186/1471-2164-13-405

Braga, R. M., Dourado, M. N., and Araújo, W. L. (2016). Microbial interactions: ecology in a molecular perspective. *Braz. J. Microbiol.* 47(Suppl. 1), 86–98. doi: 10.1016/j.bjm.2016.10.005

Bunesova, V., Lacroix, C., and Schwab, C. (2018). Mucin cross-feeding of infant bifidobacteria and *Eubacterium hallii. Microb. Ecol.* 75, 228–238. doi: 10.1007/s00248-017-1037-1034

Burley, S. K., Berman, H. M., Kleywegt, G. J., Markley, J. L., Nakamura, H., and Velankar, S. (2017). Protein data bank (PDB): the single global macromolecular structure archive. *Methods Mol. Biol.* 1607, 627–641. doi: 10.1007/978-1-4939-7000-1_26

Byvatov, E., and Schneider, G. (2003). Support vector machine applications in bioinformatics. *Appl. Bioinform.* 2, 67–77.

Carducci, M., Licata, L., Peluso, D., Castagnoli, L., and Cesareni, G. (2010). Enriching the viral-host interactomes with interactions mediated by SH3 domains. *Amino Acids* 38, 1541–1547. doi: 10.1007/s00726-009-0375-z

Charitou, T., Bryan, K., and Lynn, D. J. (2016). Using biological networks to integrate, visualize and analyze genomics data. *Genet. Sel. Evol.* 48:27. doi: 10.1186/s12711-016-0205-201

Chattopadhyay, P. K., Roederer, M., and Bolton, D. L. (2018). A deadly dance: the choreography of host-pathogen interactions, as revealed by single-cell technologies. *Nat. Commun.* 9, 4638. doi: 10.1038/s41467-018-06214-0

Che, D., Liu, Q., Rasheed, K., and Tao, X. (2011). Decision tree and ensemble learning algorithms with their applications in bioinformatics. *Adv. Exp. Med. Biol.* 696, 191–199. doi: 10.1007/978-1-4419-7046-6_19

Chen, J., Sun, J., Liu, X., Liu, F., Liu, R., and Wang, J. (2019). Structure-based prediction of West Nile virus-human protein-protein interactions. *J. Biomol. Struct. Dyn.* 37, 2310–2321. doi: 10.1080/07391102.2018.1479659

Chen, Y. E., Fischbach, M. A., and Belkaid, Y. (2018). Skin microbiota-host interactions. *Nature* 553, 427–436. doi: 10.1038/nature25177

Chen, Z., Zheng, Y., Ding, C., Ren, X., Yuan, J., Sun, F., et al. (2017). Integrated metagenomics and molecular ecological network analysis of bacterial community composition during the phytoremediation of cadmium-contaminated soils by bioenergy crops. *Ecotoxicol. Environ. Saf.* 145, 111–118. doi: 10.1016/j.ecoenv.2017.07.019

Choi, M., Carver, J., Chiva, C., Tzouros, M., Huang, T., Tsai, T.-H., et al. (2020). MassIVE.quant: a community resource of quantitative mass spectrometry-based proteomics datasets. *Nat. Methods* 17, 981–984. doi: 10.1038/s41592-020-0955-950

Clemente, J. C., Ursell, L. K., Parfrey, L. W., and Knight, R. (2012). The impact of the gut microbiota on human health: an integrative view. *Cell* 148, 1258–1270. doi: 10.1016/j.cell.2012.01.035

Clos-Garcia, M., Andrés-Marin, N., Fernández-Eulate, G., Abecia, L., Lavín, J. L., van Liempd, S., et al. (2019). Gut microbiome and serum metabolome analyses identify molecular biomarkers and altered glutamate metabolism in fibromyalgia. *EBioMedicine* 46, 499–511. doi: 10.1016/j.ebiom.2019.07.031

Clough, E., and Barrett, T. (2016). The gene expression omnibus database. *Methods Mol. Biol.* 1418, 93–110. doi: 10.1007/978-1-4939-3578-9_5

Conteville, L. C., Oliveira-Ferreira, J., and Vicente, A. C. P. (2019). Gut microbiome biomarkers and functional diversity within an amazonian semi-nomadic hunter-gatherer group. *Front. Microbiol.* 10, 1743. doi: 10.3389/fmicb.2019.01743

Cook, H. V., Doncheva, N. T., Szklarczyk, D., von Mering, C., and Jensen, L. J. (2018). Viruses.STRING: a virus-host protein-protein interaction database. *Viruses* 10:519. doi: 10.3390/v10100519

Costanzo, M. C., Crawford, M. E., Hirschman, J. E., Kranz, J. E., Robertson, L. S., et al. (2001). YPD, PombePD and WormPD: model organism volumes of the BioKnowledge library, an integrated resource for protein information. *Nucleic Acids Res.* 29, 75–79. doi: 10.1093/nar/29.1.75

Cuesta-Astroz, Y., Santos, A., Oliveira, G., and Jensen, L. J. (2019). Analysis of predicted host-parasite interactomes reveals commonalities and specificities related to parasitic lifestyle and tissues tropism. *Front. Immunol.* 10:212. doi: 10.3389/fimmu.2019.00212

Cui, G., Fang, C., and Han, K. (2012). Prediction of protein-protein interactions between viruses and human by an SVM model. *BMC Bioinformatics* 13(Suppl. 7):S5. doi: 10.1186/1471-2105-13-S7-S5

Cui, T., Li, W., Liu, L., Huang, Q., and He, Z.-G. (2016). Uncovering new pathogen-host protein-protein interactions by pairwise structure similarity. *PLoS One* 11:e0147612. doi: 10.1371/journal.pone.0147612

Cun, Y., and Fröhlich, H. (2013). Network and data integration for biomarker signature discovery via network smoothed T-statistics. *PLoS One* 8:e73074. doi: 10.1371/journal.pone.0073074

Cusick, M. F., Libbey, J. E., and Fujinami, R. S. (2012). Molecular mimicry as a mechanism of autoimmune disease. *Clin. Rev. Allergy Immunol.* 42, 102–111. doi: 10.1007/s12016-011-8294-7

Dai, X., Zhuang, Z., and Zhao, P. X. (2018). psRNATarget: a plant small RNA target analysis server (2017 release). *Nucleic Acids Res.* 46, W49–W54. doi: 10.1093/nar/gky316

Deng, Y., Jiang, Y.-H., Yang, Y., He, Z., Luo, F., and Zhou, J. (2012). Molecular ecological network analyses. *BMC Bioinformatics* 13:113. doi: 10.1186/1471-2105-13-113

Desiere, F., Deutsch, E. W., King, N. L., Nesvizhskii, A. I., Mallick, P., Eng, J., et al. (2006). The PeptideAtlas project. *Nucleic Acids Res.* 34, D655–D658. doi: 10.1093/nar/gkj040

Ding, T., Case, K. A., Omolo, M. A., Reiland, H. A., Metz, Z. P., Diao, X., et al. (2016). Predicting essential metabolic genome content of niche-specific enterobacterial human pathogens during simulation of host environments. *PLoS One* 11:e0149423. doi: 10.1371/journal.pone.0149423

Ding, Z., and Kihara, D. (2018). Computational methods for predicting protein-protein interactions using various protein features. *Curr. Protoc. Protein Sci.* 93, e62. doi: 10.1002/cpps.62

Dix, A., Vlaic, S., Guthke, R., and Linde, J. (2016). Use of systems biology to decipher host-pathogen interaction networks and predict biomarkers. *Clin. Microbiol. Infect.* 22, 600–606. doi: 10.1016/j.cmi.2016.04.014

Dominguez-Bello, M. G., Godoy-Vitorino, F., Knight, R., and Blaser, M. J. (2019). Role of the microbiome in human development. *Gut* 68, 1108–1114. doi: 10.1136/gutjnl-2018-317503

Dong, Y., Kuang, Q., Dai, X., Li, R., Wu, Y., Leng, W., et al. (2015). Improving the understanding of pathogenesis of human papillomavirus 16 via mapping protein-protein interaction network. *Biomed. Res. Int.* 2015:890381. doi: 10.1155/2015/890381

Doolittle, J. M., and Gomez, S. M. (2011). Mapping protein interactions between dengue virus and its human and insect hosts. *PLoS Negl. Trop. Dis.* 5:e954. doi: 10.1371/journal.pntd.0000954

Doxey, A. C., and McConkey, B. J. (2013). Prediction of molecular mimicry candidates in human pathogenic bacteria. *Virulence* 4, 453–466. doi: 10.4161/viru.25180

Durmus Tekir, S., Çakir, T., Ardiç, E., Sayilirbas, A. S., Konuk, G., Konuk, M., et al. (2013). PHISTO: pathogen-host interaction search tool. *Bioinformatics* 29, 1357–1358. doi: 10.1093/bioinformatics/btt137

Duval, M., Cossart, P., and Lebreton, A. (2017). Mammalian microRNAs and long noncoding RNAs in the host-bacterial pathogen crosstalk. *Semin. Cell Dev. Biol.* 65, 11–19. doi: 10.1016/j.semcdb.2016.06.016

Dyer, M. D., Murali, T. M., and Sobral, B. W. (2007). Computational prediction of host-pathogen protein-protein interactions. *Bioinformatics* 23, i159–i166. doi: 10.1093/bioinformatics/btm208

Dyer, M. D., Murali, T. M., and Sobral, B. W. (2011). Supervised learning and prediction of physical interactions between human and HIV proteins. *Infect. Genet. Evol.* 11, 917–923. doi: 10.1016/j.meegid.2011.02.022

Eain, M. M. G., Baginska, J., Greenhalgh, K., Fritz, J. V., Zenhausern, F., and Wilmes, P. (2017). Engineering solutions for representative models of the gastrointestinal human-microbe interface. *Engineering* 3, 60–65. doi: 10.1016/J.ENG.2017.01.011

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432. doi: 10.1093/nar/gky995

Emamjomeh, A., Goliaei, B., Zahiri, J., and Ebrahimpour, R. (2014). Predicting protein-protein interactions between human and hepatitis C virus via an ensemble learning method. *Mol. Biosyst.* 10, 3147–3154. doi: 10.1039/c4mb00410h

Emmert-Streib, F., and Glazko, G. V. (2011). Network biology: a direct approach to study biological function. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 3, 379–391. doi: 10.1002/wsbm.134

Evans, P., Dampier, W., Ungar, L., and Tozeren, A. (2009). Prediction of HIV-1 virus-host protein interactions using virus and host sequence motifs. *BMC Med. Genomics* 2:27. doi: 10.1186/1755-8794-2-27

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., et al. (2018). The reactome pathway knowledgebase. *Nucleic Acids Res.* 46, D649–D655. doi: 10.1093/nar/gkx1132

Farrell, M. J., and Davies, T. J. (2019). Disease mortality in domesticated animals is predicted by host evolutionary relationships. *Proc. Natl. Acad. Sci. U S A.* 116, 7911–7915. doi: 10.1073/pnas.1817323116

Franzosa, E. A., and Xia, Y. (2011). Structural principles within the human-virus protein-protein interaction network. *Proc. Natl. Acad. Sci. U S A.* 108, 10538–10543. doi: 10.1073/pnas.1101440108

Fritz, J. V., Desai, M. S., Shah, P., Schneider, J. G., and Wilmes, P. (2013). From meta-omics to causality: experimental models for human microbiome research. *Microbiome* 1:14. doi: 10.1186/2049-2618-1-14

Gao, N. L., Zhang, C., Zhang, Z., Hu, S., Lercher, M. J., Zhao, X.-M., et al. (2018). MVP: a microbe-phage interaction database. *Nucleic Acids Res.* 46, D700–D707. doi: 10.1093/nar/gkx1124

Garcia-Alonso, L., Iorio, F., Matchan, A., Fonseca, N., Jaaks, P., Peat, G., et al. (2018). Transcription factor activities enhance markers of drug sensitivity in cancer. *Cancer Res.* 78, 769–780. doi: 10.1158/0008-5472.CAN-17-1679

García-Pérez, C. A., Guo, X., Navarro, J. G., Aguilar, D. A. G., and Lara-Ramírez, E. E. (2018). Proteome-wide analysis of human motif-domain interactions mapped on influenza a virus. *BMC Bioinformatics* 19:238. doi: 10.1186/s12859-018-2237-2238

Gibson, T. J., Dinkel, H., Van Roey, K., and Diella, F. (2015). Experimental detection of short regulatory motifs in eukaryotic proteins: tips for good practice as well as for bad. *Cell Commun. Signal.* 13:42. doi: 10.1186/s12964-015-0121-y

Gosak, M., Markoviè, R., Dolenšek, J., Slak Rupnik, M., Marhl, M., Stožer, A., et al. (2018). Network science of biological systems at different scales: a review. *Phys. Life Rev.* 24, 118–135. doi: 10.1016/j.plrev.2017.11.003

Gouw, M., Michael, S., Sámano-Sánchez, H., Kumar, M., Zeke, A., Lang, B., et al. (2018). The eukaryotic linear motif resource - 2018 update. *Nucleic Acids Res.* 46, D428–D434. doi: 10.1093/nar/gkx1077

Gravina, A. G., Federico, A., Ruocco, E., Lo Schiavo, A., Romano, F., Miranda, A., et al. (2016). Crohn's disease and skin. *United Eur. Gastroenterol. J.* 4, 165–171. doi: 10.1177/2050640615597835

Guven-Maiorov, E., Tsai, C.-J., Ma, B., and Nussinov, R. (2017). Prediction of host pathogen interactions for *Helicobacter pylori* by interface mimicry and implications to gastric Cancer. *J. Mol. Biol.* 429, 3925–3941. doi: 10.1016/j.jmb.2017.10.023

Halehalli, R. R., and Nagarajaram, H. A. (2015). Molecular principles of human virus protein-protein interactions. *Bioinformatics* 31, 1025–1033. doi: 10.1093/bioinformatics/btu763

Haque, M. N., Noman, N., Berretta, R., and Moscato, P. (2016). Heterogeneous ensemble combination search using genetic algorithm for class imbalanced data classification. *PLoS One* 11:e0146116. doi: 10.1371/journal.pone.0146116

Harada, K., and Asai, T. (2010). Role of antimicrobial selective pressure and secondary factors on antimicrobial resistance prevalence in *Escherichia coli* from food-producing animals in Japan. *J. Biomed. Biotechnol.* 2010:180682. doi: 10.1155/2010/180682

Heinken, A., Acharya, G., Ravcheev, D. A., Hertel, J., Nyga, M., Okpala, O. E., et al. (2020). AGORA2: large scale reconstruction of the microbiome highlights wide-spread drug-metabolising capacities. *BioRxiv [preprint]* doi: 10.1101/2020.11.09.375451

Heinken, A., Sahoo, S., Fleming, R. M. T., and Thiele, I. (2013). Systems-level characterization of a host-microbe metabolic symbiosis in the mammalian gut. *Gut Microbes* 4, 28–40. doi: 10.4161/gmic.22370

Heinken, A., and Thiele, I. (2015). Systematic prediction of health-relevant human-microbial co-metabolism through a computational framework. *Gut Microbes* 6, 120–130. doi: 10.1080/19490976.2015.1023494

Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S. N., Richelle, A., Heinken, A., et al. (2019). Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat. Protoc.* 14, 639–702. doi: 10.1038/s41596-018-0098-92

Heleno, R., Garcia, C., Jordano, P., Traveset, A., Gómez, J. M., Blüthgen, N., et al. (2014). Ecological networks: delving into the architecture of biodiversity. *Biol. Lett.* 10:20131000. doi: 10.1098/rsbl.2013.1000

Henry, C. S., DeJongh, M., Best, A. A., Frybarger, P. M., Linsay, B., and Stevens, R. L. (2010). High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat. Biotechnol.* 28, 977–982. doi: 10.1038/nbt.1672

Hertel, J., Harms, A. C., Heinken, A., Baldini, F., Thinnes, C. C., Glaab, E., et al. (2019). Integrated analyses of microbiome and longitudinal metabolome data reveal microbial-host interactions on sulfur metabolism in Parkinson's disease. *Cell Rep.* 29, 1767–1777.e8. doi: 10.1016/j.celrep.2019.10.035

Hodges, P. E., Carrico, P. M., Hogan, J. D., O'Neill, K. E., Owen, J. J., Mangan, M., et al. (2002). Annotating the human proteome: the human proteome survey database (HumanPSD) and an in-depth target database for G protein-coupled receptors (GPCR-PD) from incyte genomics. *Nucleic Acids Res.* 30, 137–141. doi: 10.1093/nar/30.1.137

Hofree, M., Shen, J. P., Carter, H., Gross, A., and Ideker, T. (2013). Network-based stratification of tumor mutations. *Nat. Methods* 10, 1108–1115. doi: 10.1038/nmeth.2651

Hongjaisee, S., Nantasenamat, C., Carraway, T. S., and Shoombuatong, W. (2019). HIVCoR: a sequence-based tool for predicting HIV-1 CRF01_AE coreceptor usage. *Comput. Biol. Chem.* 80, 419–432. doi: 10.1016/j.compbiolchem.2019.05.006

Huang, B. L., Chandra, S., and Shih, D. Q. (2012). Skin manifestations of inflammatory bowel disease. *Front. Physiol.* 3:13. doi: 10.3389/fphys.2012.00013

Huang, C.-Y., Wang, H., Hu, P., Hamby, R., and Jin, H. (2019). Small RNAs – Big players in plant-microbe interactions. *Cell Host Microbe* 26, 173–182. doi: 10.1016/j.chom.2019.07.021

Huang, J. K., Carlin, D. E., Yu, M. K., Zhang, W., Kreisberg, J. F., Tamayo, P., et al. (2018). Systematic evaluation of molecular networks for discovery of disease genes. *Cell Syst.* 6, 484–495.e5. doi: 10.1016/j.cels.2018.03.001

Hughes, D. T., and Sperandio, V. (2008). Inter-kingdom signalling: communication between bacteria and their hosts. *Nat. Rev. Microbiol.* 6, 111–120. doi: 10.1038/nrmicro1836

Hurford, A., and Day, T. (2013). Immune evasion and the evolution of molecular mimicry in parasites. *Evolution* 67, 2889–2904. doi: 10.1111/evo.12171

Idrees, S., Pérez-Bercoff, Å., and Edwards, R. J. (2018). SLiM-Enrich: computational assessment of protein-protein interaction data as a source of domain-motif interactions. *PeerJ* 6, e5858. doi: 10.7717/peerj.5858

Ishida, T., and Kinoshita, K. (2007). PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* 35, W460–W464. doi: 10.1093/nar/gkm363

Islam, M. M., Fernando, S. C., and Saha, R. (2019). Metabolic modeling elucidates the transactions in the rumen microbiome and the shifts upon virome interactions. *Front. Microbiol.* 10:2412. doi: 10.3389/fmicb.2019.02412

Jacob, J. J., Veeraraghavan, B., and Vasudevan, K. (2019). Metagenomic next-generation sequencing in clinical microbiology. *Ind. J. Med. Microbiol.* 37, 133–140. doi: 10.4103/ijmm.IJMM_19_401

Jiao, Y., and Du, P. (2016). Performance measures in evaluating machine learning based bioinformatics predictors for classifications. *Quant. Biol.* 4, 320–330. doi: 10.1007/s40484-016-0081-2

Joossens, M., Huys, G., Cnockaert, M., De Preter, V., Verbeke, K., Rutgeerts, P., et al. (2011). Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut* 60, 631–637. doi: 10.1136/gut.2010.223263

Kaleel, M., Zheng, Y., Chen, J., Feng, X., Simpson, J. C., Pollastri, G., et al. (2020). SCLpred-EMS: subcellular localization prediction of endomembrane system

and secretory pathway proteins by Deep N-to-1 convolutional neural networks. *Bioinformatics* 36, 3343–3349. doi: 10.1093/bioinformatics/btaa156

Kargarfard, F., Sami, A., Mohammadi-Dehcheshmeh, M., and Ebrahimie, E. (2016). Novel approach for identification of influenza virus host range and zoonotic transmissible sequences by determination of host-related associative positions in viral genome segments. *BMC Genomics* 17:925. doi: 10.1186/s12864-016-3250-3259

Katiyar-Agarwal, S., and Jin, H. (2010). Role of small RNAs in host-microbe interactions. *Annu. Rev. Phytopathol.* 48, 225–246. doi: 10.1146/annurev-phyto-073009-114457

Kerr, S. A., Jackson, E. L., Lungu, O. I., Meyer, A. G., Demogines, A., Ellington, A. D., et al. (2015). Computational and functional analysis of the virus-receptor interface reveals host range trade-offs in new world arenaviruses. *J. Virol.* 89, 11643–11653. doi: 10.1128/JVI.01408-1415

Kim, B., Alguwaizani, S., Zhou, X., Huang, D.-S., Park, B., and Han, K. (2017). An improved method for predicting interactions between virus and human proteins. *J. Bioinform. Comput. Biol.* 15:1650024. doi: 10.1142/S0219720016500244

Kim, J.-G., Park, D., Kim, B.-C., Cho, S.-W., Kim, Y. T., Park, Y.-J., et al. (2008). Predicting the interactome of Xanthomonas oryzae pathovar oryzae for target selection and DB service. *BMC Bioinformatics* 9:41. doi: 10.1186/1471-2105-9-41

Kim, T.-H., Park, D.-C., Woo, D.-M., Jeong, T., and Min, S.-Y. (2012). "Multi-class classifier-based adaboost algorithm," in *Intelligent Science and Intelligent Data Engineering Lecture Notes in Computer Science*, eds Y. Zhang, Z.-H. Zhou, C. Zhang, and Y. Li (Berlin: Springer), 122–127. doi: 10.1007/978-3-642-31919-8_16

Koboziev, I., Reinoso Webb, C., Furr, K. L., and Grisham, M. B. (2014). Role of the enteric microbiota in intestinal homeostasis and inflammation. *Free Radic. Biol. Med.* 68, 122–133. doi: 10.1016/j.freeradbiomed.2013.11.008

Kösesoy, Ý, Gök, M., and Öz, C. (2019). A new sequence based encoding for prediction of host-pathogen protein interactions. *Comput. Biol. Chem.* 78, 170–177. doi: 10.1016/j.compbiolchem.2018.12.001

Kozlowski, L. P., and Bujnicki, J. M. (2012). MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinformatics* 13:111. doi: 10.1186/1471-2105-13-111

Krawczyk, B. (2015). Forming ensembles of soft one-class classifiers with weighted bagging. *New Gener. Comput.* 33, 449–466. doi: 10.1007/s00354-015-0406-400

Krishnadev, O., and Srinivasan, N. (2008). A data integration approach to predict host-pathogen protein-protein interactions: application to recognize protein interactions between human and a malarial parasite. *In Silico Biol. (Gedrukt)* 8, 235–250.

Krishnadev, O., and Srinivasan, N. (2011). Prediction of protein-protein interactions between human host and a pathogen and its application to three pathogenic bacteria. *Int. J. Biol. Macromol.* 48, 613–619. doi: 10.1016/j.ijbiomac.2011.01.030

Kshirsagar, M., Carbonell, J., and Klein-Seetharaman, J. (2013). Multitask learning for host-pathogen protein interactions. *Bioinformatics* 29, i217–i226. doi: 10.1093/bioinformatics/btt245

Kshirsagar, M., Schleker, S., Carbonell, J., and Klein-Seetharaman, J. (2015). Techniques for transferring host-pathogen protein interactions knowledge to new tasks. *Front. Microbiol.* 6, 36. doi: 10.3389/fmicb.2015.00036

Kumar, R., and Nanduri, B. (2010). HPIDB–a unified resource for host-pathogen interactions. *BMC Bioinformatics* 11, S16. doi: 10.1186/1471-2105-11-S6-S16

Lai, Y.-H., Li, Z.-C., Chen, L.-L., Dai, Z., and Zou, X.-Y. (2012). Identification of potential host proteins for influenza a virus based on topological and biological characteristics by proteome-wide network approach. *J. Proteomics* 75, 2500–2513. doi: 10.1016/j.jprot.2012.02.034

Lasso, G., Mayer, S. V., Winkelmann, E. R., Chu, T., Elliot, O., Patino-Galindo, J. A., et al. (2019). A structure-informed atlas of human-virus interactions. *Cell* 178, 1526–1541.e16. doi: 10.1016/j.cell.2019.08.005

Lavelle, A., and Sokol, H. (2020). Gut microbiota-derived metabolites as key actors in inflammatory bowel disease. *Nat. Rev. Gastroenterol. Hepatol.* 17, 223–237. doi: 10.1038/s41575-019-0258-z

Levenson, V., and Mori, Y. (2014). The era of personalized medicine: mechanistic or correlative biomarkers? *Per. Med.* 11, 361–364. doi: 10.2217/pme.14.10

Lee, S.-A., Chan, C., Tsai, C.-H., Lai, J.-M., Wang, F.-S., Kao, C.-Y., et al. (2008). Ortholog-based protein-protein interaction prediction and its application to

inter-species interactions. *BMC Bioinformatics* 9(Suppl. 12):S11. doi: 10.1186/1471-2105-9-S12-S11

Leite, D. M. C., Brochet, X., Resch, G., Que, Y.-A., Neves, A., and Peña-Reyes, C. (2018). Computational prediction of inter-species relationships through omics data analysis and machine learning. *BMC Bioinformatics* 19:420. doi: 10.1186/s12859-018-2388-2387

Levy, S. E., and Myers, R. M. (2016). Advancements in next-generation sequencing. *Annu. Rev. Genomics. Hum. Genet.* 17, 95–115. doi: 10.1146/annurev-genom-083115-22413

Li, C.-W., Jheng, B.-R., and Chen, B.-S. (2018). Investigating genetic-and-epigenetic networks, and the cellular mechanisms occurring in Epstein-Barr virus-infected human B lymphocytes via big data mining and genome-wide two-sided NGS data identification. *PLoS One* 13:e0202537. doi: 10.1371/journal.pone.0202537

Li, Q., Wang, C., Tang, C., He, Q., Li, N., and Li, J. (2014). Dysbiosis of gut fungal microbiota is associated with mucosal inflammation in Crohn's disease. *J. Clin. Gastroenterol.* 48, 513–523. doi: 10.1097/MCG.0000000000000035

Li, W., Fan, X., Long, Q., Xie, L., and Xie, J. (2015a). Mycobacterium tuberculosis effectors involved in host-pathogen interaction revealed by a multiple scales integrative pipeline. *Infect. Genet. Evol.* 32, 1–11. doi: 10.1016/j.meegid.2015.02.014

Li, Y., Wang, C., Miao, Z., Bi, X., Wu, D., Jin, N., et al. (2015b). ViRBase: a resource for virus-host ncRNA-associated interactions. *Nucleic Acids Res.* 43, D578–D582. doi: 10.1093/nar/gku903

Li, X., Liao, B., Shu, Y., Zeng, Q., and Luo, J. (2009). Protein functional class prediction using global encoding of amino acid sequence. *J. Theor. Biol.* 261, 290–293. doi: 10.1016/j.jtbi.2009.07.017

Li, Z.-G., He, F., Zhang, Z., and Peng, Y.-L. (2012). Prediction of protein-protein interactions between Ralstonia solanacearum and *Arabidopsis thaliana*. *Amino Acids* 42, 2363–2371. doi: 10.1007/s00726-011-0978-z

Lian, X., Yang, S., Li, H., Fu, C., and Zhang, Z. (2019). Machine-Learning-Based predictor of human-bacteria protein-protein interactions by incorporating comprehensive host- network properties. *J. Proteome Res.* 18, 2195–2205. doi: 10.1021/acs.jproteome.9b00074

Liao, Q., Yuan, X., Xiao, H., Liu, C., Lv, Z., Zhao, Y., et al. (2011). Identifying *Schistosoma japonicum* excretory/secretory proteins and their interactions with host immune system. *PLoS One* 6:e23786. doi: 10.1371/journal.pone.0023786

Lin, W.-C., Lu, Y.-H., and Tsai, C.-F. (2019). Feature selection in single and ensemble learning-based bankruptcy prediction models. *Expert Systems* 36:e12335. doi: 10.1111/exsy.12335

Lloyd-Price, J., Arze, C., Ananthakrishnan, A. N., Schirmer, M., Avila-Pacheco, J., Poon, T. W., et al. (2019). Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* 569, 655–662. doi: 10.1038/s41586-019-1237-9

Long, Q., Yan, R., Hu, J., Cai, D., Mitra, B., Kim, E. S., et al. (2017). The role of host DNA ligases in hepadnavirus covalently closed circular DNA formation. *PLoS Pathog.* 13:e1006784. doi: 10.1371/journal.ppat.1006784

Ma, J., Chen, T., Wu, S., Yang, C., Bai, M., Shu, K., et al. (2019). iProX: an integrated proteome resource. *Nucleic Acids Res.* 47, D1211–D1217. doi: 10.1093/nar/gky869

Mahajan, G., and Mande, S. C. (2017). Using structural knowledge in the protein data bank to inform the search for potential host-microbe protein interactions in sequence space: application to *Mycobacterium tuberculosis*. *BMC Bioinformatics* 18:201. doi: 10.1186/s12859-017-1550-y

Mariethoz, J., Khatib, K., Alocci, D., Campbell, M. P., Karlsson, N. G., Packer, N. H., et al. (2016). SugarBindDB, a resource of glycan-mediated host-pathogen interactions. *Nucleic Acids Res.* 44, D1243–D1250. doi: 10.1093/nar/gkv1247

Martin, A. M., Yabut, J. M., Choo, J. M., Page, A. J., Sun, E. W., Jessup, C. F., et al. (2019). The gut microbiome regulates host glucose homeostasis via peripheral serotonin. *Proc. Natl. Acad. Sci. U S A.* 116, 19802–19804. doi: 10.1073/pnas.1909311116

Martinez, K. B., Leone, V., and Chang, E. B. (2017). Microbial metabolites in health and disease: navigating the unknown in search of function. *J. Biol. Chem.* 292, 8553–8559. doi: 10.1074/jbc.R116.752899

May, S., Evans, S., and Parry, L. (2017). Organoids, organs-on-chips and other systems, and microbiota. *Emerg. Top. Life Sci.* 1, 385–400. doi: 10.1042/ETLS20170047

Mehrotra, P., Ramakrishnan, G., Dhandapani, G., Srinivasan, N., and Madanan, M. G. (2017). Comparison of *Leptospira interrogans* and *Leptospira biflexa* genomes: analysis of potential leptospiral-host interactions. *Mol. Biosyst.* 13, 883–891. doi: 10.1039/c6mb00856a

Mei, S. (2013). Probability weighted ensemble transfer learning for predicting interactions between HIV-1 and human proteins. *PLoS One* 8:e79606. doi: 10.1371/journal.pone.0079606

Mei, S., Flemington, E. K., and Zhang, K. (2018). Transferring knowledge of bacterial protein interaction networks to predict pathogen targeted human genes and immune signaling pathways: a case study on *Francisella tularensis*. *BMC Genomics* 19:505. doi: 10.1186/s12864-018-4873-4879

Mei, S., and Zhang, K. (2020). In silico unravelling pathogen-host signaling cross-talks via pathogen mimicry and human protein-protein interaction networks. *Comput. Struct. Biotechnol. J.* 18, 100–113. doi: 10.1016/j.csbj.2019.12.008

Mei, S., and Zhu, H. (2014a). AdaBoost based multi-instance transfer learning for predicting proteome-wide interactions between *Salmonella* and human proteins. *PLoS One* 9:e110488. doi: 10.1371/journal.pone.0110488

Mei, S., and Zhu, H. (2014b). Computational reconstruction of proteome-wide protein interaction networks between HTLV retroviruses and Homo sapiens. *BMC Bioinformatics* 15:245. doi: 10.1186/1471-2105-15-245

Mendes, V., Galvão, I., and Vieira, A. T. (2019). Mechanisms by which the gut microbiota influences cytokine production and modulates host inflammatory responses. *J. Interferon Cytokine Res.* 39, 393–409. doi: 10.1089/jir.2019.0011

Mészáros, B., Erdos, G., and Dosztányi, Z. (2018). IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* 46, W329–W337. doi: 10.1093/nar/gky384

Meyer, J. M., Leempoel, K., Losapio, G., and Hadly, E. A. (2020). Molecular ecological network analyses: an effective conservation tool for the assessment of biodiversity, trophic interactions, and community structure. *Front. Ecol. Evol.* 8:588430. doi: 10.3389/fevo.2020.588430

Miryala, S. K., Anbarasu, A., and Ramaiah, S. (2018). Discerning molecular interactions: a comprehensive review on biomolecular interaction databases and network analysis tools. *Gene* 642, 84–94. doi: 10.1016/j.gene.2017.11.028

Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., et al. (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* 47, D351–D360. doi: 10.1093/nar/gky1100

Mizianty, M. J., Peng, Z., and Kurgan, L. (2013). MFDp2: accurate predictor of disorder in proteins by fusion of disorder probabilities, content and profiles. *Intrinsically Disord Proteins* 1:e24428. doi: 10.4161/idp.24428

Mollentze, N., and Streicker, D. G. (2020). Viral zoonotic risk is homogenous among taxonomic orders of mammalian and avian reservoir hosts. *Proc. Natl. Acad. Sci. U S A.* 117, 9423–9430. doi: 10.1073/pnas.1919176117

Muller, E. E. L., Glaab, E., May, P., Vlassis, N., and Wilmes, P. (2013). Condensing the omics fog of microbial communities. *Trends Microbiol.* 21, 325–333. doi: 10.1016/j.tim.2013.04.009

Negi, S., Pandey, S., Srinivasan, S. M., Mohammed, A., and Guda, C. (2015). LocSigDB: a database of protein localization signals. *Database (Oxford)* 2015:bav003. doi: 10.1093/database/bav003

Nourani, E., Khunjush, F., and Durmuñ, S. (2016). Computational prediction of virus-human protein-protein interactions using embedding kernelized heterogeneous data. *Mol. Biosyst.* 12, 1976–1986. doi: 10.1039/c6mb00065g

Nouretdinov, I., Gammerman, A., Qi, Y., and Klein-Seetharaman, J. (2012). Determining confidence of predicted interactions between HIV-1 and human proteins using conformal method. *Pac. Symp. Biocomput.* 2012, 311–322.

Oates, M. E., Romero, P., Ishida, T., Ghalwash, M., Mizianty, M. J., Xue, B., et al. (2013). D2P2: database of disordered protein predictions. *Nucleic Acids Res.* 41, D508–D516. doi: 10.1093/nar/gks1226

Ohland, C. L., and Jobin, C. (2015). Microbial activities and intestinal homeostasis: a delicate balance between health and disease. *Cell. Mol. Gastroenterol. Hepatol.* 1, 28–40. doi: 10.1016/j.jcmgh.2014.11.004

Okuda, S., Watanabe, Y., Moriya, Y., Kawano, S., Yamamoto, T., Matsumoto, M., et al. (2017). jPOSTrepo: an international standard data repository for proteomes. *Nucleic Acids Res.* 45, D1107–D1111. doi: 10.1093/nar/gkw1080

Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., et al. (2014). The MIntAct project –IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* 42, D358–D363. doi: 10.1093/nar/gkt1115

Parkinson, H., Kapushesky, M., Shojatalab, M., Abeygunawardena, N., Coulson, R., Farne, A., et al. (2007). ArrayExpress–a public database of microarray experiments and gene expression profiles. *Nucleic Acids Res.* 35, D747–D750. doi: 10.1093/nar/gkl995

Paull, E. O., Carlin, D. E., Niepel, M., Sorger, P. K., Haussler, D., and Stuart, J. M. (2013). Discovering causal pathways linking genomic events to transcriptional states using Tied Diffusion Through Interacting Events (TieDIE). *Bioinformatics* 29, 2757–2764. doi: 10.1093/bioinformatics/btt471

Payne, W. E., and Garrels, J. I. (1997). Yeast Protein Database (YPD): a database for the complete proteome of *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 25, 57–62. doi: 10.1093/nar/25.1.57

Peabody, M. A., Laird, M. R., Vlasschaert, C., Lo, R., and Brinkman, F. S. L. (2016). PSORTdb: expanding the bacteria and archaea protein subcellular localization database to better reflect diversity in cell envelope structures. *Nucleic Acids Res.* 44, D663–D668. doi: 10.1093/nar/gkv1271

Pedamallu, C. S., and Ozdamar, L. (2014). "A review on protein-protein interaction network databases," in *Modeling, Dynamics, Optimization and Bioeconomics I Springer Proceedings in Mathematics & Statistics*, eds A. A. Pinto and D. Zilberman (Cham: Springer International Publishing), 511–519. doi: 10.1007/978-3-319-04849-9_30

Penny, H. A., Hodge, S. H., and Hepworth, M. R. (2018). Orchestration of intestinal homeostasis and tolerance by group 3 innate lymphoid cells. *Semin. Immunopathol.* 40, 357–370. doi: 10.1007/s00281-018-0687-688

Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., et al. (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* 47, D442–D450. doi: 10.1093/nar/gky1106

Perkins, J. R., Diboun, I., Dessailly, B. H., Lees, J. G., and Orengo, C. (2010). Transient protein-protein interactions: structural, functional, and network properties. *Structure* 18, 1233–1243. doi: 10.1016/j.str.2010.08.007

Peters, J. M., Solomon, S. L., Itoh, C. Y., and Bryson, B. D. (2019). Uncovering complex molecular networks in hostŰpathogen interactions using systems biology. *Emerg. Top. Life Sci.* 3, 371–378. doi: 10.1042/ETLS20180174

Pey, J., Tobalina, L., de Cisneros, J. P. J., and Planes, F. J. (2013). A network-based approach for predicting key enzymes explaining metabolite abundance alterations in a disease phenotype. *BMC Syst. Biol.* 7:2. doi: 10.1186/1752-0509-7-62

Pickard, J. M., Zeng, M. Y., Caruso, R., and Núñez, G. (2017). Gut microbiota: role in pathogen colonization, immune responses, and inflammatory disease. *Immunol. Rev.* 279, 70–89. doi: 10.1111/imr.12567

Pierleoni, A., Martelli, P. L., Fariselli, P., and Casadio, R. (2007). eSLDB: eukaryotic subcellular localization database. *Nucleic Acids Res.* 35, D208–D212. doi: 10.1093/nar/gkl775

Pryor, R., Norvaisas, P., Marinos, G., Best, L., Thingholm, L. B., Quintaneiro, L. M., et al. (2019). Host-Microbe-Drug-Nutrient screen identifies bacterial effectors of metformin therapy. *Cell* 178, 1299–1312.e29. doi: 10.1016/j.cell.2019.08.003

Qi, Y., Tastan, O., Carbonell, J. G., Klein-Seetharaman, J., and Weston, J. (2010). Semi-supervised multi-task learning for predicting interactions between HIV-1 and human proteins. *Bioinformatics* 26, i645–i652. doi: 10.1093/bioinformatics/btq394

Qiu, Y.-Q., Zhang, S., Zhang, X.-S., and Chen, L. (2010). Detecting disease associated modules and prioritizing active genes based on high throughput data. *BMC Bioinformatics* 11:26. doi: 10.1186/1471-2105-11-26

Raghavachari, B., Tasneem, A., Przytycka, T. M., and Jothi, R. (2008). DOMINE: a database of protein domain interactions. *Nucleic Acids Res.* 36, D656–D661. doi: 10.1093/nar/gkm761

Rajasekharan, S., Rana, J., Gulati, S., Sharma, S. K., Gupta, V., and Gupta, S. (2013). Predicting the host protein interactors of *Chandipura* virus using a structural similarity-based approach. *Pathog Dis.* 69, 29–35. doi: 10.1111/2049-632X.12064

Rana, A., Ahmed, M., Rub, A., and Akhter, Y. (2015). A tug-of-war between the host and the pathogen generates strategic hotspots for the development of novel therapeutic interventions against infectious diseases. *Virulence* 6, 566–580. doi: 10.1080/21505594.2015.1062211

Rastogi, S., and Rost, B. (2011). LocDB: experimental annotations of localization for homo sapiens and *Arabidopsis thaliana*. *Nucleic Acids Res.* 39, D230–D234. doi: 10.1093/nar/gkq927

Rodenburg, S. Y. A., Seidl, M. F., Judelson, H. S., Vu, A. L., Govers, F., and de Ridder, D. (2019). Metabolic model of the phytophthora infestans-tomato interaction reveals metabolic switches during host colonization. *mBio* 10:e00454-19. doi: 10.1128/mBio.00454-419

Romano, P., Dräger, A., Fiannaca, A., Giugno, R., La Rosa, M., et al. (2019). The 2017 Network Tools and Applications in Biology (NETTAB) workshop: aims, topics and outcomes. *BMC Bioinformatics* 20:125. doi: 10.1186/s12859-019-2681-2680

Roume, H., Heintz-Buschart, A., Muller, E. E. L., May, P., Satagopam, V. P., Laczny, C. C., et al. (2015). Comparative integrated omics: identification of key functionalities in microbial community-wide metabolic networks. *NPJ Biofilms Microb.* 1:15007. doi: 10.1038/npjbiofilms.2015.7

Saçar Demirci, M. D., and Adan, A. (2020). Computational analysis of microRNA-mediated interactions in SARS-CoV-2 infection. *PeerJ* 8:e9369. doi: 10.7717/peerj.9369

Sahu, S. S., Weirick, T., and Kaundal, R. (2014). Predicting genome-scale *Arabidopsis-Pseudomonas* syringae interactome using domain and interolog-based approaches. *BMC Bioinformatics* 15(Suppl. 11):S13. doi: 10.1186/1471-2105-15-S11-S13

Saik, O. V., Ivanisenko, T. V., Demenkov, P. S., and Ivanisenko, V. A. (2016). Interactome of the hepatitis C virus: literature mining with ANDSystem. *Virus Res.* 218, 40–48. doi: 10.1016/j.virusres.2015.12.003

Samal, S. S., Radulescu, O., Weber, A., and Fröhlich, H. (2017). Linking metabolic network features to phenotypes using sparse group lasso. Bioinformatics 33, 3445–3453. doi: 10.1093/bioinformatics/btx427

Schaubeck, M., Clavel, T., Calasan, J., Lagkouvardos, I., Haange, S. B., Jehmlich, N., et al. (2016). Dysbiotic gut microbiota causes transmissible Crohn's disease-like ileitis independent of failure in antimicrobial defence. *Gut* 65, 225–237. doi: 10.1136/gutjnl-2015-309333

Schleker, S., Garcia-Garcia, J., Klein-Seetharaman, J., and Oliva, B. (2012). Prediction and comparison of *Salmonella-human* and *Salmonella-Arabidopsis* interactomes. *Chem. Biodivers.* 9, 991–1018. doi: 10.1002/cbdv.201100392

Schmidt, T., Samaras, P., Frejno, M., Gessulat, S., Barnert, M., Kienegger, H., et al. (2018). ProteomicsDB. *Nucleic Acids Res.* 46, D1271–D1281. doi: 10.1093/nar/gkx1029

Schommer, N. N., and Gallo, R. L. (2013). Structure and function of the human skin microbiome. *Trends Microbiol.* 21, 660–668. doi: 10.1016/j.tim.2013.10.001

Schweppe, D. K., Harding, C., Chavez, J. D., Wu, X., Ramage, E., Singh, P. K., et al. (2015). Host-Microbe Protein Interactions during Bacterial Infection. *Chem. Biol.* 22, 1521–1530. doi: 10.1016/j.chembiol.2015.09.015

Shah, P., Fritz, J. V., Glaab, E., Desai, M. S., Greenhalgh, K., Frachet, A., et al. (2016). A microfluidics-based in vitro model of the gastrointestinal human-microbe interface. *Nat. Commun.* 7:11535. doi: 10.1038/ncomms11535

Sharma, V., Eckels, J., Schilling, B., Ludwig, C., Jaffe, J. D., MacCoss, M. J., et al. (2018). Panorama public: a public repository for quantitative data sets processed in skyline. *Mol. Cell Proteomics* 17, 1239–1244. doi: 10.1074/mcp.RA117.000543

Shastry, K. A., and Sanjay, H. A. (2020). "Machine learning for bioinformatics," in *Statistical Modelling and Machine Learning Principles for Bioinformatics Techniques, Tools, and Applications Algorithms for Intelligent Systems* eds. K. G. Srinivasa, G. M. Siddesh, and S. R. Manisekhar (Singapore: Springer Singapore), 25–39. doi: 10.1007/978-981-15-2445-5_3

Shaw, K. A., Bertha, M., Hofmekler, T., Chopra, P., Vatanen, T., Srivatsa, A., et al. (2016). Dysbiosis, inflammation, and response to treatment: a longitudinal study of pediatric subjects with newly diagnosed inflammatory bowel disease. *Genome Med.* 8:75. doi: 10.1186/s13073-016-0331-y

Shirahama, S., Miki, A., Kaburaki, T., and Akimitsu, N. (2020). Long non-coding RNAs involved in pathogenic infection. *Front. Genet.* 11:454. doi: 10.3389/fgene.2020.00454

Shoombuatong, W., Hongjaisee, S., Barin, F., Chaijaruwanich, J., and Samleerat, T. (2012). HIV-1 CRF01_AE coreceptor usage prediction using kernel methods based logistic model trees. *Comput. Biol. Med.* 42, 885–889. doi: 10.1016/j.compbiomed.2012.06.011

Silmon de Monerri, N. C., and Kim, K. (2014). Pathogens hijack the epigenome: a new twist on host-pathogen interactions. *Am. J. Pathol.* 184, 897–911. doi: 10.1016/j.ajpath.2013.12.022

Singh, N., Bhatia, V., Singh, S., and Bhatnagar, S. (2019). MorCVD: a unified database for host-pathogen protein-protein interactions of cardiovascular diseases related to microbes. *Sci. Rep.* 9:4039. doi: 10.1038/s41598-019-40704-5

Sudhakar, P., Jacomin, A.-C., Hautefort, I., Samavedam, S., Fatemian, K., Ari, E., et al. (2019). Targeted interplay between bacterial pathogens and host autophagy. *Autophagy* 15, 1620–1633. doi: 10.1080/15548627.2019.1590519

Sun, J., Yang, L.-L., Chen, X., Kong, D.-X., and Liu, R. (2018). Integrating multifaceted information to predict *Mycobacterium tuberculosis*-human protein-protein interactions. *J. Proteome Res.* 17, 3810–3823. doi: 10.1021/acs.jproteome.8b00497

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 45, D362–D368. doi: 10.1093/nar/gkw937

Tastan, O., Qi, Y., Carbonell, J. G., and Klein-Seetharaman, J. (2009). Prediction of interactions between HIV-1 and human proteins by information integration. *Pac. Symp. Biocomput.* 2009, 516–527. doi: 10.1142/9789812836939_0049

The UniProt Consortium (2018). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 46:2699. doi: 10.1093/nar/gky092

Thiele, I., Heinken, A., and Fleming, R. M. T. (2013a). A systems biology approach to studying the role of microbes in human health. *Curr. Opin. Biotechnol.* 24, 4–12. doi: 10.1016/j.copbio.2012.10.001

Thiele, I., Swainston, N., Fleming, R. M. T., Hoppe, A., Sahoo, S., Aurich, M. K., et al. (2013b). A community-driven global reconstruction of human metabolism. *Nat. Biotechnol.* 31, 419–425. doi: 10.1038/nbt.2488

Thiele, I., Sahoo, S., Heinken, A., Hertel, J., Heirendt, L., Aurich, M. K., et al. (2020). Personalized whole-body models integrate metabolism, physiology, and the gut microbiome. *Mol. Syst. Biol.* 16:e8982. doi: 10.15252/msb.20198982

Thieu, T., Joshi, S., Warren, S., and Korkin, D. (2012). Literature mining of host-pathogen interactions: comparing feature-based supervised learning and language-based approaches. *Bioinformatics* 28, 867–875. doi: 10.1093/bioinformatics/bts042

Thompson, D., Regev, A., and Roy, S. (2015). Comparative analysis of gene regulatory networks: from network reconstruction to evolution. *Annu. Rev. Cell Dev. Biol.* 31, 399–428. doi: 10.1146/annurev-cellbio-100913-112908

Thul, P. J., and Lindskog, C. (2018). The human protein atlas: a spatial map of the human proteome. *Protein Sci.* 27, 233–244. doi: 10.1002/pro.3307

Tramontano, M., Andrejev, S., Pruteanu, M., Klünemann, M., Kuhn, M., Galardini, M., et al. (2018). Nutritional preferences of human gut bacteria reveal their metabolic idiosyncrasies. *Nat. Microbiol.* 3, 514–522. doi: 10.1038/s41564-018-0123-129

Türei, D., Korcsmáros, T., and Saez-Rodriguez, J. (2016). OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nat. Methods* 13, 966–967. doi: 10.1038/nmeth.4077

Tyagi, N., Krishnadev, O., and Srinivasan, N. (2009). Prediction of protein-protein interactions between *Helicobacter pylori* and a human host. *Mol. Biosyst.* 5, 1630–1635. doi: 10.1039/b906543c

Valli, R. X. E., Lyng, M., and Kirkpatrick, C. L. (2020). There is no hiding if you Seq: recent breakthroughs in *Pseudomonas aeruginosa* research revealed by genomic and transcriptomic next-generation sequencing. *J. Med. Microbiol.* 69, 162–175. doi: 10.1099/jmm.0.001135

Vandin, F., Upfal, E., and Raphael, B. J. (2011). Algorithms for detecting significantly mutated pathways in cancer. *J. Comput. Biol.* 18, 507–522. doi: 10.1089/cmb.2010.0265

Veres, D. V., Gyurkó, D. M., Thaler, B., Szalay, K. Z., Fazekas, D., Korcsmáros, T., et al. (2015). ComPPI: a cellular compartment-specific database for protein-protein interaction network analysis. *Nucleic Acids Res.* 43, D485–D493. doi: 10.1093/nar/gku1007

Via, A., Uyar, B., Brun, C., and Zanzoni, A. (2015). How pathogens use linear motifs to perturb host cell networks. *Trends Biochem. Sci.* 40, 36–48. doi: 10.1016/j.tibs.2014.11.001

Wallqvist, A., Wang, H., Zavaljevski, N., Memiševiæ, V., Kwon, K., Pieper, R., et al. (2017). Mechanisms of action of *Coxiella burnetii* effectors inferred from host-pathogen protein interactions. *PLoS One* 12:e0188071. doi: 10.1371/journal.pone.0188071

Wang, B., Yao, M., Lv, L., Ling, Z., and Li, L. (2017). The human microbiota in health and disease. *Engineering* 3, 71–82. doi: 10.1016/J.ENG.2017.01.008

Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F., and Jones, D. T. (2004). The DISOPRED server for the prediction of protein disorder. *Bioinformatics* 20, 2138–2139. doi: 10.1093/bioinformatics/bth195

Weiberg, A., Wang, M., Bellinger, M., and Jin, H. (2014). Small RNAs: a new paradigm in plant-microbe interactions. *Annu. Rev. Phytopathol.* 52, 495–516. doi: 10.1146/annurev-phyto-102313-045933

Wen, C., Zheng, Z., Shao, T., Liu, L., Xie, Z., Le Chatelier, E., et al. (2017). Quantitative metagenomics reveals unique gut microbiome biomarkers in ankylosing spondylitis. *Genome Biol.* 18:42. doi: 10.1186/s13059-017-1271-6

Wong, A. C. N., Vanhove, A. S., and Watnick, P. I. (2016). The interplay between intestinal bacteria and host metabolism in health and disease: lessons from *Drosophila melanogaster*. *Dis. Model. Mech.* 9, 271–281. doi: 10.1242/dmm.023408

Wong, E., Baur, B., Quader, S., and Huang, C.-H. (2012). Biological network motif detection: principles and practice. *Brief. Bioinformatics* 13, 202–215. doi: 10.1093/bib/bbr033

Wuchty, S. (2011). Computational prediction of host-parasite protein interactions between *Plasmodium falciparum* and *H. sapiens*. *PLoS One* 6:e26960. doi: 10.1371/journal.pone.0026960

Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K., and Uversky, V. N. (2010). PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta* 1804, 996–1010. doi: 10.1016/j.bbapap.2010.01.011

Yang, G., Peng, M., Tian, X., and Dong, S. (2017). Molecular ecological network analysis reveals the effects of probiotics and florfenicol on intestinal microbiota homeostasis: an example of sea cucumber. *Sci. Rep.* 7:4778. doi: 10.1038/s41598-017-05312-5311

Yijing, L., Haixiang, G., Xiao, L., Yanan, L., and Jinling, L. (2016). Adapted ensemble classification algorithm based on multiple classifier system and feature selection for classifying multi-class imbalanced data. *Knowledge-Based Systems* 94, 88–104. doi: 10.1016/j.knosys.2015.11.013

Yilmaz, B., Juillerat, P., Øyås, O., Ramon, C., Bravo, F. D., Franc, Y., et al. (2019). Microbial network disturbances in relapsing refractory Crohn's disease. *Nat. Med.* 25, 323–336. doi: 10.1038/s41591-018-0308-z

Younesi, E. (2015). Disease systems modeling for discovery of mechanistic biomarkers. *Eur. J. Mol. Clin. Med.* 2:61. doi: 10.1016/j.nhtm.2014.11.023

Yu, H., Xue, D., Wang, Y., Zheng, W., Zhang, G., and Wang, Z.-L. (2020). Molecular ecological network analysis of the response of soil microbial communities to depth gradients in farmland soils. *Microbiologyopen* 9:e983. doi: 10.1002/mbo3.983

Zampieri, G., Vijayakumar, S., Yaneske, E., and Angione, C. (2019). Machine and deep learning meet genome-scale metabolic modeling. *PLoS Comput. Biol.* 15:e1007084. doi: 10.1371/journal.pcbi.1007084

Zhang, A., He, L., and Wang, Y. (2017a). Prediction of GCRV virus-host protein interactome based on structural motif-domain interactions. *BMC Bioinformatics* 18:145. doi: 10.1186/s12859-017-1500-1508

Zhang, M., Su, Q., Lu, Y., Zhao, M., and Niu, B. (2017b). Application of machine learning approaches for protein-protein interactions prediction. *Med. Chem.* 13, 506–514. doi: 10.2174/1573406413666170522150940

Zheng, Q., Zhang, M., Zhang, T., Li, X., Zhu, M., and Wang, X. (2020). Insights from metagenomic, metatranscriptomic, and molecular ecological network analyses into the effects of chromium nanoparticles on activated sludge system. *Front. Environ. Sci. Eng.* 14:60. doi: 10.1007/s11783-020-1239-1238

Zhou, H., Gao, S., Nguyen, N. N., Fan, M., Jin, J., Liu, B., et al. (2014). Stringent homology-based prediction of *H. sapiens-M. tuberculosis* H37Rv protein-protein interactions. *Biol. Direct* 9:5. doi: 10.1186/1745-6150-9-5

Zhou, H., Rezaei, J., Hugo, W., Gao, S., Jin, J., Fan, M., et al. (2013). Stringent DDI-based prediction of *H. sapiens-M. tuberculosis* H37Rv protein-protein interactions. *BMC Syst. Biol.* 7:S6. doi: 10.1186/1752-0509-7-S6-S6

Zhou, J., Deng, Y., Luo, F., He, Z., Tu, Q., and Zhi, X. (2010). Functional molecular ecological networks. *mBio* 1:e00169-10. doi: 10.1128/mBio.00169-110

Zhou, X., Park, B., Choi, D., and Han, K. (2018). A generalized approach to predicting protein-protein interactions between virus and host. *BMC Genomics* 19:568. doi: 10.1186/s12864-018-4924-4922