Article

MicrobioLink: An integrated computational pipeline to infer functional effects of microbiome-host interactions

Tahila Andrighetti[1,2,#], Balazs Bohar[1,3], Ney Lemke[2], Padhmanand Sudhakar[1,4,5,$] , Tamas Korcsmaros[1,4,$]

[1] Earlham Institute, Norwich Research Park, Norwich, NR4 7UZ, UK
[2] São Paulo University (UNESP), Institute of Biosciences, Botucatu, Brazil.
[3] Department of Genetics, Eötvös Loránd University, Budapest 1117, Hungary
[4] Quadram Institute Bioscience, Norwich Research Park, Norwich, NR4 7UQ, UK
[5] Department of Chronic Diseases, Metabolism and Ageing, KU Leuven, BE-3000, Leuven, Belgium.
[#] Current address: Laboratory of Structural Bioinformatics and Computational Biology (SBCB), Institute of Informatics, Federal University of Rio Grande do Sul (UFRGS), Av. Bento Gonçalves 9500, 91501-970 - Porto Alegre, RS - Brazil
[$]Joint corresponding authors

Corresponding authors

Tamas Korcsmaros
Group Leader
Earlham Institute
Norwich Research Park
NR4 7UZ
Norwich, UK
Email : Tamas.Korcsmaros@earlham.ac.uk
ORCID : https://orcid.org/0000-0003-1717-996X

Padhmanand Sudhakar
Earlham Institute
Norwich Research Park
NR4 7UZ
Norwich, UK
Email : Padhmanand.Sudhakar@earlham.ac.uk
ORCID : https://orcid.org/0000-0003-1907-4491

Abstract

Microbiome-host interactions play significant roles in health and in various diseases including auto-immune disorders. Uncovering these inter-kingdom cross-talks propels our understanding of disease pathogenesis, and provides useful leads on potential therapeutic targets. Despite the biological significance of microbe-host interactions, there is a big gap in understanding the downstream effects of these interactions on host processes. Computational methods are expected to fill this gap by generating, integrating and prioritizing predictions - as experimental detection remains challenging due to feasibility issues. Here, we present MicrobioLink, a computational pipeline to integrate predicted interactions between microbial and host proteins together with host molecular networks. Using the concept of network diffusion, MicrobioLink can analyse how microbial proteins in a certain context are influencing cellular processes by modulating gene or protein expression. We demonstrated the applicability of the pipeline using a case study. We used gut metaproteomic data from Crohn's disease patients and healthy controls to uncover the mechanisms by which the microbial proteins can modulate host genes which belong to biological processes implicated in disease pathogenesis. MicrobioLink, which is agnostic of the microbial protein sources (bacterial, viral etc), is freely available on GitHub (https://github.com/korcsmarosgroup/HMIpipeline).

Keywords: Microbiota-host interactions, protein-protein interactions, systems biology, networks, network diffusion, computational pipeline

1. Introduction

Microbiota-host interactions happen in almost every known organism, shaping their metabolism and evolution [1,2]. In many ecosystems, the microbiome plays an important role as manifested by its dynamic interactions with different hosts [3]. The community of microorganisms are almost indispensable to human life since they modulate and influence immunity and nutrient acquisition. For example, the gastrointestinal microbiome plays a crucial role in nutrient assimilation and energy yield by actively participating in metabolic pathways [4]. Dysbiosis (compositional alterations) of gut microbial communities is associated with diseases such as type 2 diabetes, obesity and inflammatory bowel diseases like Crohn's disease [5,6]. In addition to infections caused by pathogenic microbes, exclusion of beneficial species from the community are also known to have negative impacts on the host [7]. Most of the inferences between microbial composition changes and disease phenotypes have been based on associations and correlations, with little explanations of the mechanisms driving the phenotypes.

Studying microbiome-host interactions and its influence on host biological mechanisms is important for monitoring health and disease and also for discovering/fine-tuning therapeutic interventions. Such interactions are mediated by the interplay between various molecular components expressed by the host and the microbiome. For example, bacterial molecules such as metabolites [8], proteins [9–12] and small RNAs [13], can interact with host molecules, and through host intracellular pathways and regulatory networks, modulate the expression of genes in various biological processes [14–16], thus maintaining

healthy states or transition to diseased states. Besides the metabolite mediated interactions, protein-protein interactions (PPIs) are one of the most relevant types of molecular interplay between microbes and host organisms [12]. Experimental techniques to test inter-species PPIs are time consuming and have limitations imposed by cost [1,17–21]. Hence, there is a dearth of validated inter-species PPIs in publicly available databases, and these are mostly limited to pathogens. However, using predicted microbe-host PPIs, it is possible to build interaction networks to better understand the cross-talk between the microbial and host proteins, and how these interactions interfere with host metabolism and physiology [22]. Recent studies have also shown how protein-protein interactions between emerging pathogens such as COVID-19 and the host have enabled the discovery of potential drug candidates for clinical testing and validation [23].

Various resources and pipelines which generate or store microbe-host interaction predictions are currently available. They include PHISTO [24], PATRIC [25], Proteopathogen2 [26], VirBase [27] and HPIDB [28]. Most of these resources store PPIs and/or genomic analysis of virulence, but are limited to pathogens [22]. Tools such as COBRA [29], RAVEN [30], NetCooperate [31] and Kbase [32] which consider commensal microbes are based on the use of metabolomics data alone [33]. Moreover, most of the existing resources and pipelines are confined to predicting the direct molecular interactions at the microbe-host interface, and do not infer the downstream effects on functional processes and host signalling pathways.

To fill these gaps in understanding the effect of the microbiome on the host, we introduce MicrobioLink - a computational pipeline to analyse microbiome-host interactions at a cellular level using network and systems biology approaches. Using MicrobioLink, it is possible to identify potential pathways by which microbes can modulate the expression of key host molecules such as genes, proteins or microRNAs. In order to demonstrate the applicability of MicrobioLink, we performed a case study investigating how the gut microbiome potentially modulates autophagy genes in Crohn's disease (CD). By providing custom-made lists of microbial proteins, host receptor proteins, molecular interaction networks and the target node sets, users can harness the functionalities of MicrobioLink to understand the mechanisms which mediate the influence of microbial proteins derived from the individual microbes or the microbiome.

2. Materials and Methods

MicrobioLink enables users to integrate host responses into the interaction networks representing the microbe-host interface, and thereby, helps users to expand the list of experimentally verifiable hypotheses. It provides the user with the option to infer indirect effects of microbial proteins on the host by prioritizing pathways and signalling chains. This prioritization is based on various criteria such as chain length, i.e, the number of steps between the host receptors modulated by the microbial protein and the host target genes/proteins, host protein localization (depending on the mode of infection) and contextual gene expression.

2.1. Compilation of input proteins and genes

The first step of the workflow involves the compilation of bacterial and host proteins that can potentially interact under the studied biological conditions (**Figure 1**). The bacterial proteins can be obtained from annotated or experimentally derived bacterial proteomes, or from metaproteomic read-outs (set of proteins secreted by an entire microbial community). As for the host proteins, the list is compiled based on their localization: the chosen proteins must be located in a cellular compartment which is physically accessible to the bacterial proteins. Localization based filtering can be applied also to bacterial proteins depending on the context (based on whether the microbe is extracellular or intracellular within the host). In the case of extracellular microbes, the provided host proteins must be present at the extracellular matrix or plasma membrane since these locations are more prone to interacting with extracellular, secreted or membrane bound bacterial proteins. In the case of intracellular microbes, the provided host proteins must be located inside the host cell, as in the cytoplasm for example. These host proteins provided by the user can be derived either from experimental data or publicly available datasets (including but not limited to ProteomicsDB [34], ComPPI [35], Human Protein Atlas (HPA) [36] and MatrixDB [37]). In addition, the localization of host proteins can be inferred by bioinformatic tools such as PSORTdb[38], SignalP [39] and Secretome [40], which use sequence-based features to predict the localization of the given proteins. Users can also provide their own pre-processed lists of host proteins after localization prediction.

Subsequently, users can compile a list of important proteins or genes for the studied conditions or related cellular processes for use as target nodes within the host. This target list can be compiled either from *a priori* knowledge (derived from phenotypic observations) or contextual data obtained from genome-wide association studies (GWAS) which correlate genetic loci to observed phenotypes, literature search or -omic expression datasets such as gene expression (transcriptomics) or proteomics measured under the studied condition(s).

2.2. Bacterial-host interaction prediction

The next step in the pipeline involves the interaction prediction between the microbial and host proteins. We use two qualitative approaches, namely domain-domain and domain-motif methods [12,41,42], that use secondary structure based features to predict the interactions between microbial and host proteins. The domain-domain method is based on the assumption that microbe-host protein pairs bearing interacting domain pairs, also interact. Similarly, the domain-motif method enables the identification of potential interactions between microbial and host proteins if the microbial protein contains domains previously known to interact with eukaryotic motifs on the host proteins. Gold standard information on interacting domain pairs and interacting domain-motif pairs are retrieved from the DOMINE [43] and ELM [44] databases.

The above mentioned methods help determine which bacterial proteins, *via* their respective domains, interact with host receptor proteins. For quality control purposes, the interactions are then filtered

to sterically possible sequence features [45,46]. This is performed by excluding interactions involving motifs outside disordered regions (using IUPRED [47]) or within globular domains based on information from PFAM [48] and InterPro [49]. Previous studies have successfully used this structure-based approach to predict PPIs between microbial and host proteins, and validated the PPIs experimentally [9,10,12].

2.3. Network compilation and path tracing using diffusion

In this step, we use multi-layered network resources and network diffusion tools to trace the effect of the interactions between microbial and host proteins on other host processes further downstream. For this purpose, the TieDIE [50] tool, built into the pipeline, is used to infer the signaling paths that connect the host receptors with the host target genes. TieDIE uses the network diffusion approach to infer the signaling paths from directed multi-layered networks containing PPIs and transcriptional regulatory interactions (TRIs) from publicly available molecular interaction databases such as OmniPath [51] and DoRothEA [52] respectively. Users can also select from a vast number of other molecular interaction resources which are available in the public domain. The use of high-quality curated PPI resources such as OmniPath enables the analysis of post-translational modifications as many original sources in OmniPath contain such information. Furthermore, integration of proteomics and/or phosphoproteomics data to be analyzed in the MicrobioLink pipeline can improve the computational predictions by imparting activity information into signalling proteins and prioritizing the predictions (i.e activated signalling paths) for experimental validation.

Importantly, the network diffusion approach helps in avoiding prioritization biases stemming from network topological properties such as degree (i.e the number of interacting first neighbours of a particular gene or protein for example). This circumvents the positively biased inclusion of hub proteins into signalling paths irrespective of their actual relevance in the studied condition [50]. TieDIE also incorporates various parameters along with the input data. For example, the magnitude and direction of differential expression of the target genes/proteins or the number of microbial proteins predicted to bind to the host receptors can be specified into the input dataset. This information is then used to prioritize the signaling pathways connecting the receptors to the target genes/proteins.

After network compilation, depending on the user's discretion, a chain selection step can be performed to select the most relevant chains for the target modulation. This can be done by filtering the network to keep only signalling chains stimulated by bacterial proteins detected in any single condition, especially in cases in which the users are comparing different conditions. Chain selection can also be specified by the user depending on the dataset provided. For example, in our use case, the host expression dataset is determined from transcriptomic datasets. Hence, we confined the interactions immediately upstream to the target genes to transcriptional regulatory connections, in order to capture the regulatory biology underlying the modulation of the target genes' expression.

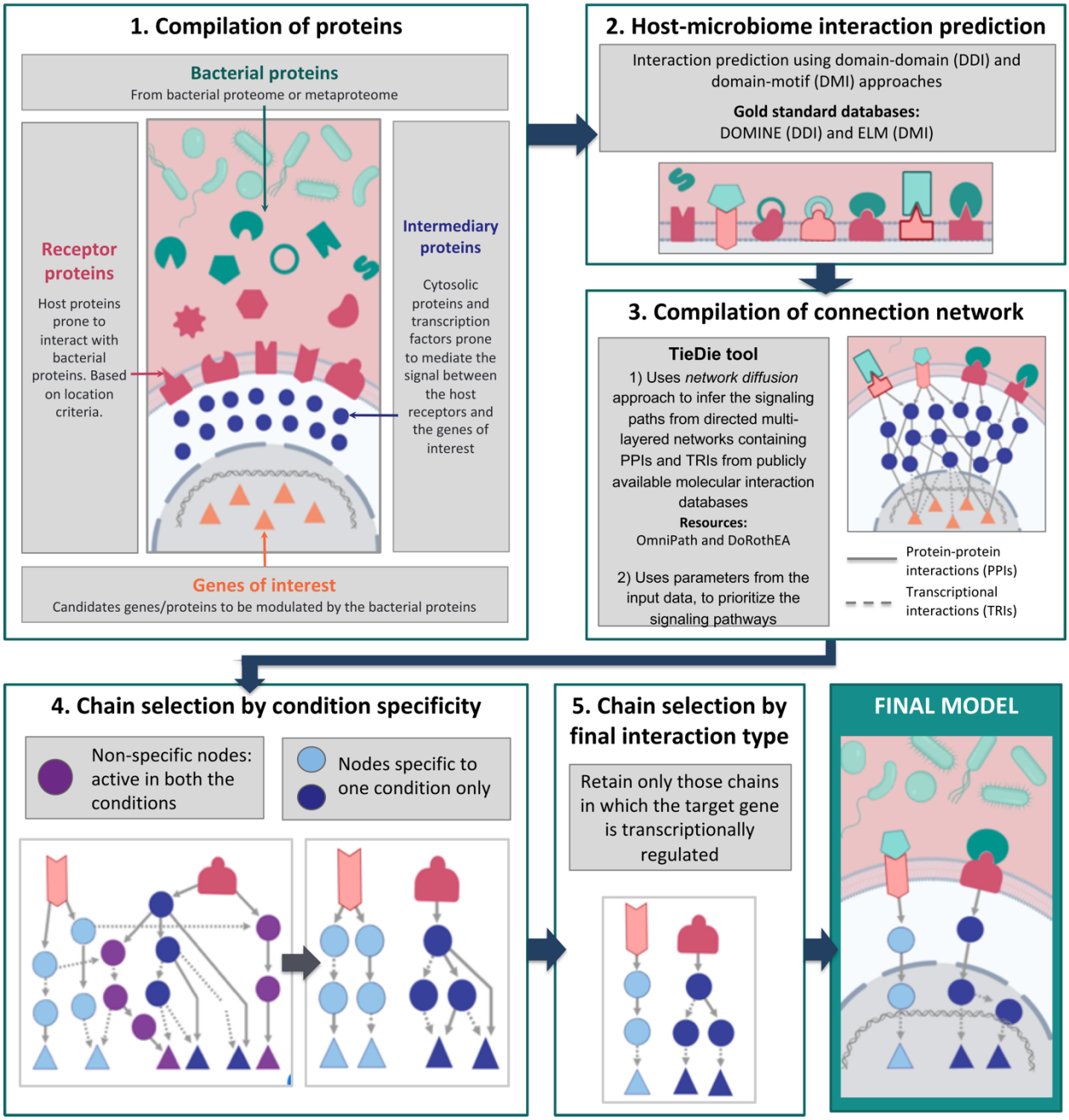

Figure 1: Graphical representation of the MicrobioLink workflow.

Finally, the obtained model can be visualized as a multi-layered network, starting with the bacterial protein-host receptor interactions, and ultimately reaching the selected host target genes through PPIs and TRIs. A summary of the MicrobioLink pipeline is presented in **Figure 1**.

3. Results

3.1. Case study

To demonstrate the applicability of the MicrobioLink pipeline, we performed a case study showing how different bacterial proteins modulate molecular processes important for the development or progression of Crohn's disease (CD). CD is a chronic intestinal inflammatory disease (IBD) associated with a multitude of factors including microbial dysbiosis and dysregulation of autophagy due to mutational effects [53–55] [56,57]. Using MicrobioLink, we identified pathways by which autophagy could be indirectly modulated by microbial proteins, thereby implying that host autophagy can be potentially modulated by microbial influence independent of mutations. This also highlighted the potential role of the microbial proteins in disease development and/or progression.

To perform this study, we first obtained the microbial proteins present only in healthy individuals or in CD patients from a Swedish twin study [58]. In total, 336 microbial proteins from healthy individuals and 226 proteins from CD patients were obtained (**Supplementary table 1**) . Then, we compiled human proteins located in the extracellular matrix and cellular membrane using matrixDB, resulting in a total of 8008 proteins (**Supplementary table 2**).

After compiling the protein lists, we performed interaction predictions between microbial and human proteins, resulting in 8478 predicted interactions involving 140 microbial proteins and 2998 host receptor proteins (**Supplementary tables 3-4**). The predicted interactions were refined by passing them through a disordered region based quality control step to eliminate sterically improbable interactions. In parallel, as potential target genes affected by the microbial proteins, we focused on autophagy genes given that autophagy is known to be a dysregulated process in CD [53–55,59,60]. In total we selected 38 autophagy genes, encoding the core components of the autophagy machinery [61]. To detect the most relevant target autophagy genes for CD, we used three different transcriptomic datasets (GSE9686 [62], GSE36807 [63] and GSE75214 [64] from GEO [65]), that contained the gene expression profiles of CD patients and healthy individuals, to select the most differentially expressed core autophagy genes. From the above mentioned profiles, we selected autophagy genes which were modulated with coherent trends in at least two datasets. Based on this criteria, we obtained five target autophagy genes namely WIPI1, MAP1LC3A, MAP1LC3B, ATG7 and ATG4D (**Supplementary table 5**).

Subsequently, contextual signaling networks that potentially mediate the signal transduction from the host receptors to the target autophagy genes were compiled using the network diffusion model inferred by TieDie (**Figure 2A**). Based on these results, we were able to infer that two of the five target autophagy genes could be potentially modulated by the microbial proteins. We also observed a group of host proteins that are connected to microbial proteins occurring in both CD patients and healthy subjects. To retain specificity, we excluded the chains initiated from microbial proteins enriched in both conditions. As a result, we obtained a network with clearly separated signalling paths exclusive to the disease and healthy contexts as shown in **Figure 2B**. In a final filtering step, we retained only those chains where the last

interaction is a TRI to capture the transcriptional regulatory effect of the expression of the target genes. Post this final filtering step, we were able to obtain a final model, as depicted in **Figure 2C**.

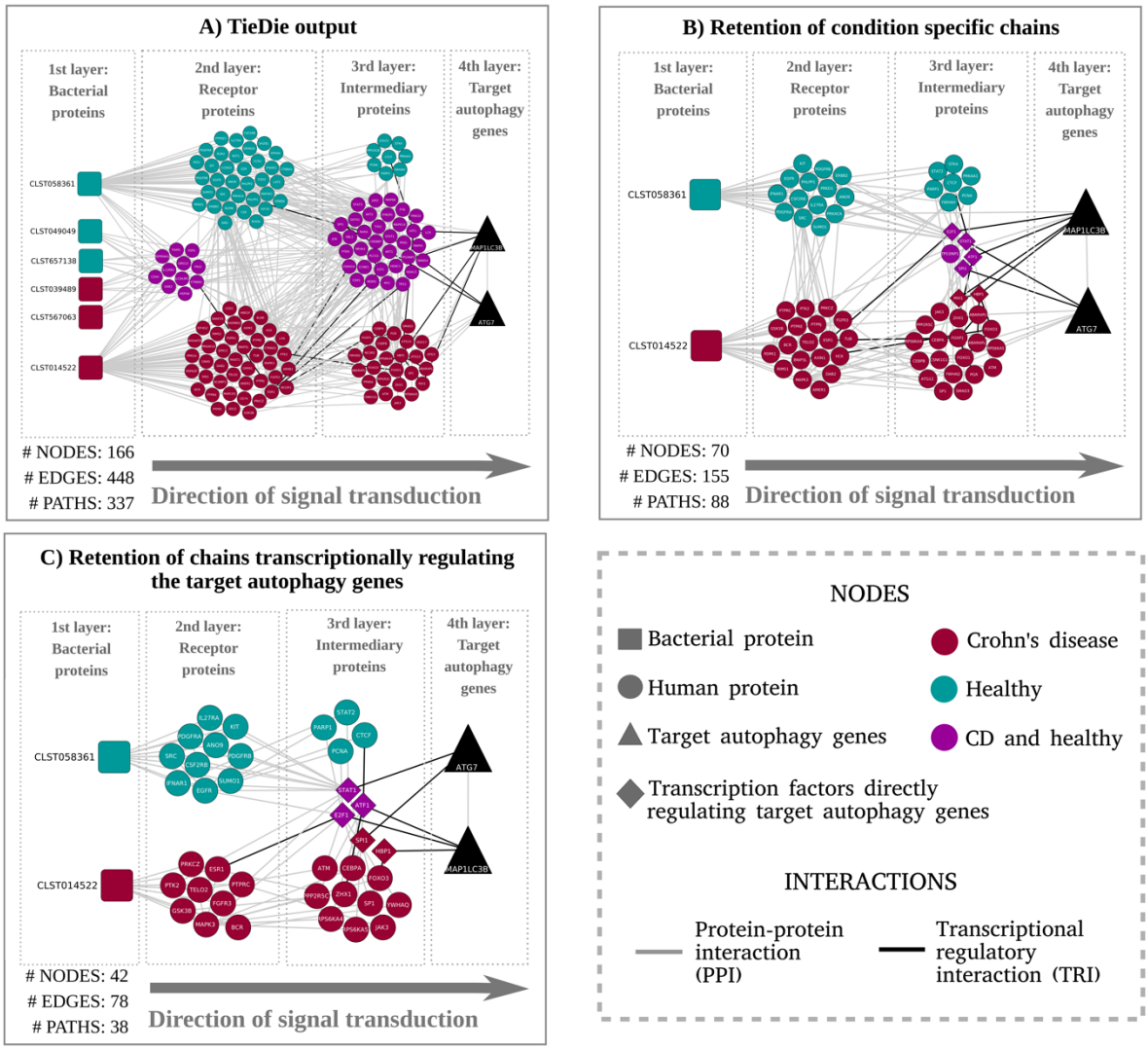

Figure 2: **(A).** Graphical representation of the signaling paths between curated receptor proteins (host proteins predicted to be modulated by microbial proteins) and target autophagy genes. The network compilation was performed by tracing the signaling chains from the human receptors (predicted to interact with the bacterial proteins) to the autophagy genes using the TieDIE tool which adopts a network diffusion [51]. For brevity, only the results corresponding to the domain-motif interaction analysis are discussed in the case study. **(B).** Network representing the signalling chains after exclusion of proteins connected with bacterial proteins detected in both CD and healthy conditions. Proteins present in both conditions that were retained were those directly regulating the target autophagy genes. **(C).** Network obtained by retaining only chains with transcriptional regulatory interactions between the intermediary protein (3rd layer) and the target autophagy genes (4th layer). The immediate upstream proteins from the autophagy target genes were confined to transcription factors modulating the target autophagy genes via a transcriptional regulatory interaction.

To analyse the functional significance of the final model, we did a Gene Ontology (GO) enrichment analysis of the human proteins to identify feature sets such as biological processes *via* which the target autophagy genes are potentially modulated by the microbial proteins. Over-represented feature sets (**Figure 3, Supplementary table 7**) included apoptosis, which is known to be upregulated in intestinal epithelial cells (IECs) in IBD patients leading to increased epithelial barrier permeability [66]. Apoptosis has a peculiar interplay with autophagy, in that it is usually up-regulated when autophagy is deactivated and *vice versa* [67]. In our final model, we also discovered several signaling chains among host networks specific to the healthy condition. These signalling chains included proteins such as SUMO1, PARP1 and E2F1. SUMO1, known to induce autophagy levels [68], also activates PARP1 and subsequently E2F1, proteins which are both known to stimulate autophagy and inhibit apoptosis [69,70]. This allowed us to propose the hypothesis that the metaproteome-derived bacterial protein CLST058361 which was expressed in healthy subjects and not in CD patients, could be a key protein involved in stimulating autophagy and decreasing apoptosis. Interestingly, CLST058361 belongs to a family of trypsin-like serine proteases whose expression level changes are associated with IBD [71] [72] [73],

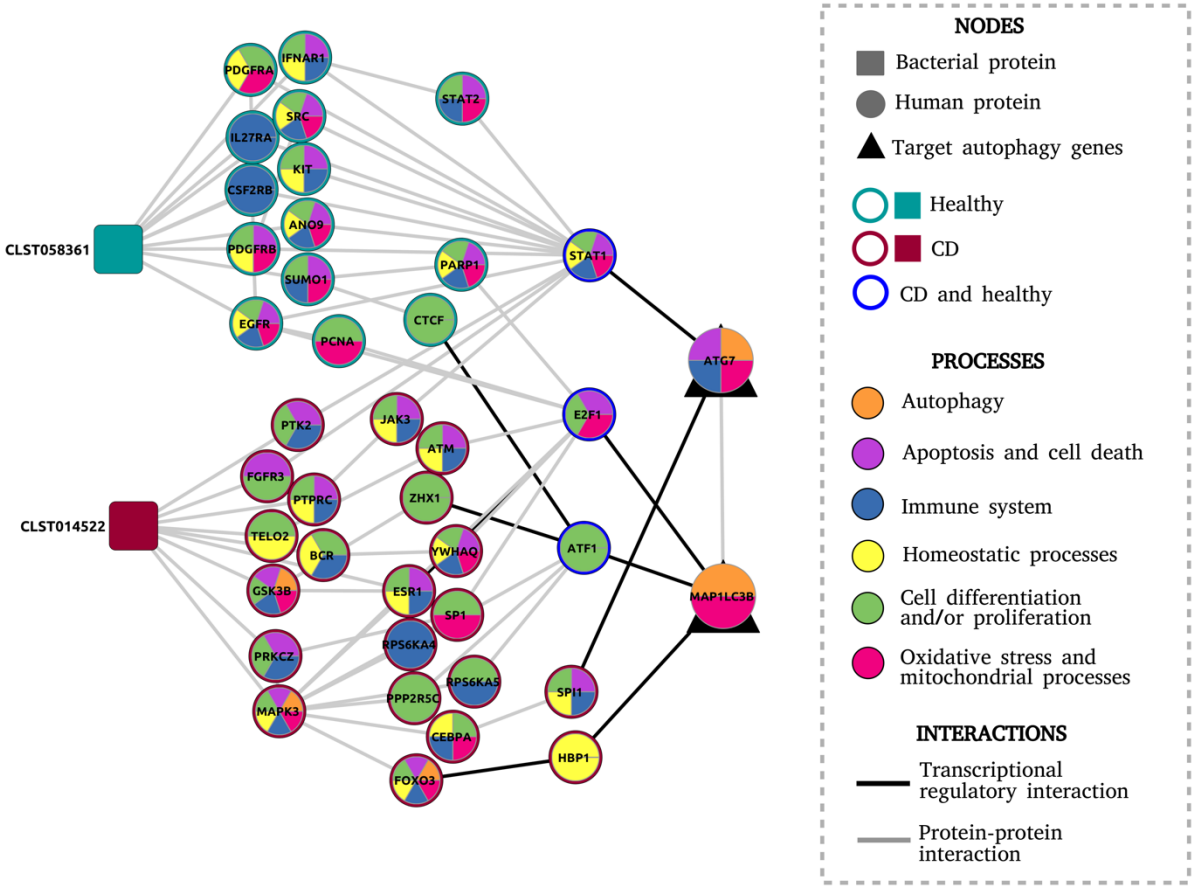

Figure 3. Final network model consisting of the proteins and the biological processes in which they are involved, inferred from the GO enrichment test.

Meanwhile, a signalling chain specific to the CD condition included MAPK3 and FOXO3, which are key proteins for apoptosis activation and apoptosis-autophagy interplay [67,74]. This pathway was modulated by the bacterial protein labeled as CLST14522 (and containing a FOG:_FHA_domain) and uniquely found in CD patients. Other biological processes over-represented in CD networks included those related to cellular differentiation and proliferation, response and generation of oxidative stress, mitochondria related processes, homeostasis and immune system regulation. More comprehensive studies are warranted to ascertain the role of these processes in how the bacterial proteins modulate autophagy in CD (**Figure 3**).

In conclusion, using MicrobioLink, we obtained a mechanistic model showing how bacterial proteins can modulate the expression of autophagy genes in the context of CD. Previous research has shown that host autophagy, while being implicated in the pathogenesis of CD, was primarily dysregulated by defects caused by genetic mutations [53–55] [56,57]. MicrobioLink reveals the possibility that autophagy can be modulated by the microbial community through signal transduction propagated *via* host molecular networks. Although experimental validation is required to ascertain this putative mechanism, we demonstrate the applicability of MicrobioLink to integrate heterogeneous datasets and generate testable hypotheses about microbe-host interactions.

4. Discussion

In this paper, we presented MicrobioLink - a pipeline for the analysis of the functional effects of the microbiome on host cellular processes. MicrobioLink integrates microbiome-host protein-protein interaction predictions with network diffusion to infer signaling networks which capture the systemic effects of microbial proteins on host processes.

Microbiome-host interactions have considerable impact on host signaling and understanding these interactions are crucial to our advances in studying disease pathogenesis and discovering drug targets. Various computational tools and approaches have enabled and furthered our understanding of microbiome-host interactions. For example, NetCooperate [31] uses metabolic network modelling to infer the nutritional interdependencies and co-feeding potential between any two microbial species or between microbial and host species. Zanzoni et al [75] harnessed the power of microbe-host protein-protein interactions to infer the perturbations induced on the host by the pathogen *Fusobacterium nucleatum*. However, researchers aiming to study the functional effects of microbiome-host interactions face technological and methodological boundaries. First, the high demand of time and cost of experimental studies makes it infeasible to test such a high number of potential microbiome-host interactions [19]. MicrobioLink is an effective alternative to study the molecular mechanisms by which microorganisms interact with the host cells and how these interactions affect the host cellular mechanisms. It can be applied either as a tool for prioritizing the most relevant non-canonical *de novo* pathways or for establishing hypotheses about how microorganisms can modulate particular cellular processes. Other limitations that MicrobioLink helps overcome include the possibility for the user to customize the pipeline depending on his/her organism or community of interest as long as the focus of the microbial-host molecular interface are

protein-protein interactions. Due to the mechanism driven approach of MicrobioLink, it provides the users with experimentally testable hypotheses.

Several existing computational methods and tools to study microbe-host interactions are currently available although they are limited to specific organisms (with a focus on pathogens) and specific types of interactions. Therefore, to perform an analysis where the microorganism-host interaction is unknown and to study its influence on host processes, the researcher would have to combine different and incompatible tools. In **Table 1**, we benchmarked some of the key resources, and compared them with the functions of MicrobioLink. With MicrobioLink, it is possible to perform functional microbiome-host analysis without previous knowledge about its interactions for any microorganism and host species as long as the corresponding annotated proteomes/genomes are available. The user can provide a protein dataset as input from a single species (proteome) or from an entire microbial community (metaproteome). These microorganisms can be bacteria, archaea, viruses or fungi, which makes MicrobioLink applicable for any type of microorganism. Subsequently, the user can select a dataset that represents the resulting host phenotype. It is possible to choose any dataset adequate for the study, such as differentially expressed genes from transcriptomic measurements or protein expression from proteomics. The host data can also correspond to any host, and hence MicrobioLink can be applied for studying the influence of the microbiome on plants, mammals or even fungi.

Table 1: Summary of selected tools and resources used in microbe-host interaction research.

Resource/ tool	Standalone version ?	Description	Can user - provided datasets be handled ?	Non-pathogenic species included / handling ?	Protein -protein interactions ?	Inferring downstream effects ?	Microorganisms supported	Host organisms supported
PHISTO [24]	online	Web-tool for mining and retrieving host-pathogen interactions	no	no	yes	no	Viral, bacterial, fungal and protozoan pathogens	Human
PATRIC [25]	online	Genome focussed infectious disease research database	yes	no	yes	no	Bacterial pathogens	Actinopterygii, Arachnida, Chromadorea, Insecta, Mammalia

Proteopathogen2 [26]	online	Database and web application to store and display fungal pathogen proteomics data.	no	no	no	no	Fungal pathogens	Mammalian species
VirBase [27]	online	Database of virus-host ncRNA-associated interactions and interaction networks during viral infections.	no	yes	no	no	Virus	Vertebrates, plants and arthropods
NetCooperate [31]	python module	Web-based tool and software package for determining host-microbe and microbe-microbe cooperative potential from metabolic networks.	yes	yes	no	yes	Any microorganism	Any host species
Kbase [32]	Online, python and java	Software and data platform that enables data sharing, integration, and analysis of microbes, plants, and their communities by creating workflows consisting of a series of analysis	yes	yes	no	yes	Any microorganism	Any host

		tool runs and code blocks						
M²IA [80]	web-based server	Statistical analysis methods for microbiome and metabolome data integration, including correlation analysis and functional network analysis	yes	yes	no	yes	Any microorganism	Any host species
COMETS [81]	Matlab and a python toolbox	Modelling framework that integrates dynamic flux balance analysis with diffusion to communities	yes	yes	no	yes	Any microorganism	Any host species
MicrobioLink (this paper)	Python, Docker	Integrated evaluation of microbe-host interaction networks	yes	yes	yes	yes	Any microorganism	Any host species

For reproducibility and interoperability [76], we have implemented the pipeline within the framework of the Docker system. Docker functions as an integrated platform with all the needed dependencies and software packages pre-installed to execute the scripts within the pipeline. Users can thereby quickly test different components of the pipeline within the docker container without having to search and install the required dependencies. The datasets corresponding to the use-case have also been provided within the container so that the users can execute the pipeline readily and cross-check the results.

The pipeline can be applied to different diseases modulated by the microbiome. In the case study, we generated a model explaining how the gut microbiome from CD patients is potentially modulating autophagy, a cellular process known to be altered and dysregulated in CD patients. Besides, the tool can be extended to diverse diseases that are shown to be influenced by the microbiome such as diabetes, depression and cardiovascular disorders [77]. Moreover, beneficial effects can also be studied with MicrobioLink. In humans, it can be applied to demonstrate how probiotic species maintain health and homeostasis. In livestock and poultry, it can be used to study how the microbiome can influence productivity [78]. In plants, MicrobioLink can be used to analyse how the commensal microbiome in the soil influences plant health and disease [79]. Eventhough experimental validation of the inferences from the pipeline is out of the scope of this work, we previously demonstrated [12] the validity of the computational inference of microbe-host protein protein interactions and their downstream effects by *in vitro* verification. As a next step towards improving the pipeline, we are planning to establish use cases in different domains of biological research (including aiding generation of hypotheses for experiments) wherein the microbiome plays a crucial role in the phenotypes of different hosts such as plants as well as mammalian host species. In particular, MicrobioLink will be very useful in studying the role of the microbiome in autoimmune disorders, such as rheumatoid arthritis, vasculitis, multiple sclerosis and systemic lupus erythematosus. With MicrobioLink, we could infer the microbiome-mediated mechanisms in such disorders and thereby point out key microbial inferences, cellular pathways transmitting normal microbial signals affected in these disorders, and potential targets for further therapeutic interventions.

5. Conclusion

For many conditions and organisms, MicrobioLink presents novel additional functionalities which makes it possible not only to predict microbiome-host interactions but also to infer mechanisms driving functional effects further downstream. MicrobioLink provides an integrated approach by embedding the host proteins modulated by the user-provided microbial proteins within molecular interaction networks and relevant -omic datasets customized for the studied conditions and organisms. Using MicrobioLink, it is possible to evaluate how an entire microbial community or even a single microorganism, either a commensal or pathogen, can interfere with host processes *via* protein mediated signal transduction.

Data and Software Availability

Code for the MicrobioLink pipeline is available at https://github.com/korcsmarosgroup/HMIpipeline. The code is implemented in Python and dockerized in

order to ensure easy testing, circumvent dependency issues and achieve reproducibility. The implementation is provided as two versions - firstly as a sequential step-by-step workflow and secondly as a one-click version. Both versions have been supplemented with corresponding documentation including the dependencies and instructions for execution.

Author contributions

TK and PS designed the methodology; PS and TA developed the codes; TA and PS performed the case study; BB performed the dockerization. TA, TK, NL and PS wrote the manuscript; all authors read and contributed to the manuscript.

Competing interests

None.

Grant information

This work was supported with scholarships to TA by CAPES – the Brazilian Federal Agency for Support and Evaluation of Graduate Education within the Ministry of Education of Brazil and by a PhD scholarship supported by CNPq - Brazilian National Council for Scientific and Technological Development. This work was also supported by a fellowship to TK in computational biology at the Earlham Institute (Norwich, UK) in partnership with the Quadram Institute (Norwich, UK), and strategically supported by the UKRI Biotechnological and Biosciences Research Council (BBSRC), UK grants (BB/J004529/1, BB/P016774/1 and BB/CSP17270/1). TK and PS were also funded by a BBSRC ISP grant for Gut Microbes and Health BB/R012490/1 and its constituent project(s), BBS/E/F/000PR10353 and BBS/E/F/000PR10355. PS was also supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 694679).

Supplementary tables

Supplementary table 1. List of microbial proteins derived from the Swedish twin cohort study [58].

Supplementary table 2. Localization filtered human receptor proteins retrieved from MatrixDB [37].

Supplementary table 3. Domain-domain based protein-protein interaction predictions between microbial and human receptor proteins.

Supplementary table 4. Domain-motif based protein-protein interaction predictions between microbial and human receptor proteins.

Supplementary table 5. Gene expression measurements, as represented by the logFCs, of the core autophagy genes from three different GEO datasets.

Supplementary table 6. Signaling network obtained by TieDIE [50], connecting the host receptor proteins (predicted to bind to the microbial proteins) and the target autophagy genes.

Supplementary table 7. GO enrichment results of the human proteins from the final network model. The 1st sheet contains all enriched gene ontology terms (biological process) in the final network. The 2nd sheet

contains selected processes related to CD pathogenesis. The 3rd sheet contains the assignment of proteins to the selected biological process terms enriched in the network.

References

1. Weimer, B. C.; Chen, P.; Desai, P. T.; Chen, D.; Shah, J. Whole Cell Cross-Linking to Discover Host-Microbe Protein Cognate Receptor/Ligand Pairs. *Front. Microbiol.* **2018**, *9*, 1585, doi:10.3389/fmicb.2018.01585.
2. Simon, J.-C.; Marchesi, J. R.; Mougel, C.; Selosse, M.-A. Host-microbiota interactions: from holobiont theory to analysis. *Microbiome* **2019**, *7*, 5, doi:10.1186/s40168-019-0619-4.
3. Wooley, J. C.; Godzik, A.; Friedberg, I. A primer on metagenomics. *PLoS Comput. Biol.* **2010**, *6*, e1000667, doi:10.1371/journal.pcbi.1000667.
4. Krajmalnik-Brown, R.; Ilhan, Z.-E.; Kang, D.-W.; DiBaise, J. K. Effects of gut microbes on nutrient absorption and energy regulation. *Nutr. Clin. Pract.* **2012**, *27*, 201–214, doi:10.1177/0884533611436116.
5. Devaraj, S.; Hemarajata, P.; Versalovic, J. The human gut microbiome and body metabolism: implications for obesity and diabetes. *Clin. Chem.* **2013**, *59*, 617–628, doi:10.1373/clinchem.2012.187617.
6. Loh, G.; Blaut, M. Role of commensal gut bacteria in inflammatory bowel diseases. *Gut Microbes* **2012**, *3*, 544–555, doi:10.4161/gmic.22156.
7. Pickard, J. M.; Zeng, M. Y.; Caruso, R.; Núñez, G. Gut microbiota: Role in pathogen colonization, immune responses, and inflammatory disease. *Immunol. Rev.* **2017**, *279*, 70–89, doi:10.1111/imr.12567.
8. Heinken, A.; Sahoo, S.; Fleming, R. M. T.; Thiele, I. Systems-level characterization of a host-microbe metabolic symbiosis in the mammalian gut. *Gut Microbes* **2013**, *4*, 28–40, doi:10.4161/gmic.22370.
9. Wojcik, J.; Schächter, V. Protein-protein interaction map inference using interacting domain profile pairs. *Bioinformatics* **2001**, *17 Suppl 1*, S296-305, doi:10.1093/bioinformatics/17.suppl_1.s296.
10. Evans, P.; Dampier, W.; Ungar, L.; Tozeren, A. Prediction of HIV-1 virus-host protein interactions using virus and host sequence motifs. *BMC Med. Genomics* **2009**, *2*, 27, doi:10.1186/1755-8794-2-27.
11. Jones, E. J.; Matthews, Z. J.; Gul, L.; Sudhakar, P.; Treveil, A.; Divekar, D.; Buck, J.; Wrzesinski, T.; Jefferson, M.; Armstrong, S. D.; Hall, L. J.; Watson, A. J. M.; Carding, S. R.; Haerty, W.; Di Palma, F.; Mayer, U.; Powell, P. P.; Hautefort, I.; Wileman, T.; Korcsmaros, T. Integrative analysis of Paneth cell proteomic and transcriptomic data from intestinal organoids reveals functional processes dependent on autophagy. *Dis. Model. Mech.* **2019**, *12*, doi:10.1242/dmm.037069.
12. Sudhakar, P.; Jacomin, A.-C.; Hautefort, I.; Samavedam, S.; Fatemian, K.; Ari, E.; Gul, L.; Demeter, A.; Jones, E.; Korcsmaros, T.; Nezis, I. P. Targeted interplay between bacterial pathogens and host autophagy. *Autophagy* **2019**, *15*, 1620–1633, doi:10.1080/15548627.2019.1590519.
13. Katiyar-Agarwal, S.; Jin, H. Role of small RNAs in host-microbe interactions. *Annu. Rev. Phytopathol.* **2010**, *48*, 225–246, doi:10.1146/annurev-phyto-073009-114457.
14. Ruch, T. R.; Engel, J. N. Targeting the mucosal barrier: how pathogens modulate the cellular polarity network. *Cold Spring Harb. Perspect. Biol.* **2017**, *9*, doi:10.1101/cshperspect.a027953.
15. Killick, K. E.; Magee, D. A.; Park, S. D. E.; Taraktsoglou, M.; Browne, J. A.; Conlon, K. M.; Nalpas, N. C.; Gormley, E.; Gordon, S. V.; MacHugh, D. E.; Hokamp, K. Key Hub and Bottleneck Genes

Differentiate the Macrophage Response to Virulent and Attenuated Mycobacterium bovis. *Front. Immunol.* **2014**, *5*, 422, doi:10.3389/fimmu.2014.00422.

16. Durmuş Tekir, S.; Cakir, T.; Ulgen, K. Ö. Infection Strategies of Bacterial and Viral Pathogens through Pathogen-Human Protein-Protein Interactions. *Front. Microbiol.* **2012**, *3*, 46, doi:10.3389/fmicb.2012.00046.
17. Foley, S. L.; Johnson, T. J.; Ricke, S. C.; Nayak, R.; Danzeisen, J. Salmonella pathogenicity and host adaptation in chicken-associated serovars. *Microbiol. Mol. Biol. Rev.* **2013**, *77*, 582–607, doi:10.1128/MMBR.00015-13.
18. Nourani, E.; Khunjush, F.; Durmuş, S. Computational approaches for prediction of pathogen-host protein-protein interactions. *Front. Microbiol.* **2015**, *6*, 94, doi:10.3389/fmicb.2015.00094.
19. Gould, A. L.; Zhang, V.; Lamberti, L.; Jones, E. W.; Obadia, B.; Korasidis, N.; Gavryushkin, A.; Carlson, J. M.; Beerenwinkel, N.; Ludington, W. B. Microbiome interactions shape host fitness. *Proc Natl Acad Sci USA* **2018**, *115*, E11951–E11960, doi:10.1073/pnas.1809349115.
20. Dyer, M. D.; Neff, C.; Dufford, M.; Rivera, C. G.; Shattuck, D.; Bassaganya-Riera, J.; Murali, T. M.; Sobral, B. W. The human-bacterial pathogen protein interaction networks of Bacillus anthracis, Francisella tularensis, and Yersinia pestis. *PLoS ONE* **2010**, *5*, e12089, doi:10.1371/journal.pone.0012089.
21. Dyer, M. D.; Murali, T. M.; Sobral, B. W. The landscape of human proteins interacting with viruses and other pathogens. *PLoS Pathog.* **2008**, *4*, e32, doi:10.1371/journal.ppat.0040032.
22. Guven-Maiorov, E.; Tsai, C.-J.; Nussinov, R. Structural host-microbiota interaction networks. *PLoS Comput. Biol.* **2017**, *13*, e1005579, doi:10.1371/journal.pcbi.1005579.
23. Zhou, Y.; Hou, Y.; Shen, J.; Huang, Y.; Martin, W.; Cheng, F. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov.* **2020**, *6*, 14, doi:10.1038/s41421-020-0153-3.
24. Durmuş Tekir, S.; Çakır, T.; Ardiç, E.; Sayılırbaş, A. S.; Konuk, G.; Konuk, M.; Sarıyer, H.; Uğurlu, A.; Karadeniz, İ.; Özgür, A.; Sevilgen, F. E.; Ülgen, K. Ö. PHISTO: pathogen-host interaction search tool. *Bioinformatics* **2013**, *29*, 1357–1358, doi:10.1093/bioinformatics/btt137.
25. Wattam, A. R.; Abraham, D.; Dalay, O.; Disz, T. L.; Driscoll, T.; Gabbard, J. L.; Gillespie, J. J.; Gough, R.; Hix, D.; Kenyon, R.; Machi, D.; Mao, C.; Nordberg, E. K.; Olson, R.; Overbeek, R.; Pusch, G. D.; Shukla, M.; Schulman, J.; Stevens, R. L.; Sullivan, D. E.; Sobral, B. W. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res.* **2014**, *42*, D581-91, doi:10.1093/nar/gkt1099.
26. Vialas, V.; Gil, C. Proteopathogen2, a database and web tool to store and display proteomics identification results in the mzIdentML standard. *EuPA Open Proteom.* **2015**, *8*, 22–27, doi:10.1016/j.euprot.2015.04.002.
27. Li, Y.; Wang, C.; Miao, Z.; Bi, X.; Wu, D.; Jin, N.; Wang, L.; Wu, H.; Qian, K.; Li, C.; Zhang, T.; Zhang, C.; Yi, Y.; Lai, H.; Hu, Y.; Cheng, L.; Leung, K.-S.; Li, X.; Zhang, F.; Li, K.; Wang, D. ViRBase: a resource for virus-host ncRNA-associated interactions. *Nucleic Acids Res.* **2015**, *43*, D578-82, doi:10.1093/nar/gku903.
28. Kumar, R.; Nanduri, B. HPIDB--a unified resource for host-pathogen interactions. *BMC Bioinformatics* **2010**, *11 Suppl 6*, S16, doi:10.1186/1471-2105-11-S6-S16.
29. Heirendt, L.; Arreckx, S.; Pfau, T.; Mendoza, S. N.; Richelle, A.; Heinken, A.; Haraldsdóttir, H. S.;

Wachowiak, J.; Keating, S. M.; Vlasov, V.; Magnusdóttir, S.; Ng, C. Y.; Preciat, G.; Žagare, A.; Chan, S. H. J.; Aurich, M. K.; Clancy, C. M.; Modamio, J.; Sauls, J. T.; Noronha, A.; Fleming, R. M. T. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat. Protoc.* **2019**, *14*, 639–702, doi:10.1038/s41596-018-0098-2.

30. Agren, R.; Liu, L.; Shoaie, S.; Vongsangnak, W.; Nookaew, I.; Nielsen, J. The RAVEN toolbox and its use for generating a genome-scale metabolic model for Penicillium chrysogenum. *PLoS Comput. Biol.* **2013**, *9*, e1002980, doi:10.1371/journal.pcbi.1002980.
31. Levy, R.; Carr, R.; Kreimer, A.; Freilich, S.; Borenstein, E. NetCooperate: a network-based tool for inferring host-microbe and microbe-microbe cooperation. *BMC Bioinformatics* **2015**, *16*, 164, doi:10.1186/s12859-015-0588-y.
32. Arkin, A. P.; Cottingham, R. W.; Henry, C. S.; Harris, N. L.; Stevens, R. L.; Maslov, S.; Dehal, P.; Ware, D.; Perez, F.; Canon, S.; Sneddon, M. W.; Henderson, M. L.; Riehl, W. J.; Murphy-Olson, D.; Chan, S. Y.; Kamimura, R. T.; Kumari, S.; Drake, M. M.; Brettin, T. S.; Glass, E. M.; Yu, D. Kbase: the united states department of energy systems biology knowledgebase. *Nat. Biotechnol.* **2018**, *36*, 566–569, doi:10.1038/nbt.4163.
33. Sen, P.; Orešič, M. Metabolic modeling of human gut microbiota on a genome scale: an overview. *Metabolites* **2019**, *9*, doi:10.3390/metabo9020022.
34. Schmidt, T.; Samaras, P.; Frejno, M.; Gessulat, S.; Barnert, M.; Kienegger, H.; Krcmar, H.; Schlegl, J.; Ehrlich, H.-C.; Aiche, S.; Kuster, B.; Wilhelm, M. ProteomicsDB. *Nucleic Acids Res.* **2018**, *46*, D1271–D1281, doi:10.1093/nar/gkx1029.
35. Veres, D. V.; Gyurkó, D. M.; Thaler, B.; Szalay, K. Z.; Fazekas, D.; Korcsmáros, T.; Csermely, P. ComPPI: a cellular compartment-specific database for protein-protein interaction network analysis. *Nucleic Acids Res.* **2015**, *43*, D485-93, doi:10.1093/nar/gku1007.
36. Uhlén, M.; Fagerberg, L.; Hallström, B. M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, Å.; Kampf, C.; Sjöstedt, E.; Asplund, A.; Olsson, I.; Edlund, K.; Lundberg, E.; Navani, S.; Szigyarto, C. A.-K.; Odeberg, J.; Djureinovic, D.; Takanen, J. O.; Hober, S.; Alm, T.; Pontén, F. Proteomics. Tissue-based map of the human proteome. *Science* **2015**, *347*, 1260419, doi:10.1126/science.1260419.
37. Launay, G.; Salza, R.; Multedo, D.; Thierry-Mieg, N.; Ricard-Blum, S. MatrixDB, the extracellular matrix interaction database: updated content, a new navigator and expanded functionalities. *Nucleic Acids Res.* **2015**, *43*, D321-7, doi:10.1093/nar/gku1091.
38. Peabody, M. A.; Laird, M. R.; Vlasschaert, C.; Lo, R.; Brinkman, F. S. L. PSORTdb: expanding the bacteria and archaea protein subcellular localization database to better reflect diversity in cell envelope structures. *Nucleic Acids Res.* **2016**, *44*, D663-8, doi:10.1093/nar/gkv1271.
39. Armenteros, J. J. A.; Tsirigos, K. D.; Sønderby, C. K.; Petersen, T. N.; Winther, O.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **2019**, *37*, 420–423, doi:10.1038/s41587-019-0036-z.
40. Bendtsen, J. D.; Kiemer, L.; Fausbøll, A.; Brunak, S. Non-classical protein secretion in bacteria. *BMC Microbiol.* **2005**, *5*, 58, doi:10.1186/1471-2180-5-58.
41. Bader, S.; Kühner, S.; Gavin, A.-C. Interaction networks for systems biology. *FEBS Lett.* **2008**, *582*, 1220–1224, doi:10.1016/j.febslet.2008.02.015.
42. Korcsmaros, T.; Dunai, Z. A.; Vellai, T.; Csermely, P. Teaching the bioinformatics of signaling

networks: an integrated approach to facilitate multi-disciplinary learning. *Brief. Bioinformatics* **2013**, *14*, 618–632, doi:10.1093/bib/bbt024.

43. Raghavachari, B.; Tasneem, A.; Przytycka, T. M.; Jothi, R. DOMINE: a database of protein domain interactions. *Nucleic Acids Res.* **2008**, *36*, D656-61, doi:10.1093/nar/gkm761.
44. Gouw, M.; Michael, S.; Sámano-Sánchez, H.; Kumar, M.; Zeke, A.; Lang, B.; Bely, B.; Chemes, L. B.; Davey, N. E.; Deng, Z.; Diella, F.; Gürth, C.-M.; Huber, A.-K.; Kleinsorg, S.; Schlegel, L. S.; Palopoli, N.; Roey, K. V.; Altenberg, B.; Reményi, A.; Dinkel, H.; Gibson, T. J. The eukaryotic linear motif resource - 2018 update. *Nucleic Acids Res.* **2018**, *46*, D428–D434, doi:10.1093/nar/gkx1077.
45. Dinkel, H.; Van Roey, K.; Michael, S.; Davey, N. E.; Weatheritt, R. J.; Born, D.; Speck, T.; Krüger, D.; Grebnev, G.; Kuban, M.; Strumillo, M.; Uyar, B.; Budd, A.; Altenberg, B.; Seiler, M.; Chemes, L. B.; Glavina, J.; Sánchez, I. E.; Diella, F.; Gibson, T. J. The eukaryotic linear motif resource ELM: 10 years and counting. *Nucleic Acids Res.* **2014**, *42*, D259-66, doi:10.1093/nar/gkt1047.
46. Weatheritt, R. J.; Jehl, P.; Dinkel, H.; Gibson, T. J. iELM--a web server to explore short linear motif-mediated interactions. *Nucleic Acids Res.* **2012**, *40*, W364-9, doi:10.1093/nar/gks444.
47. Mészáros, B.; Erdos, G.; Dosztányi, Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **2018**, *46*, W329–W337, doi:10.1093/nar/gky384.
48. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S. R.; Luciani, A.; Potter, S. C.; Qureshi, M.; Richardson, L. J.; Salazar, G. A.; Smart, A.; Sonnhammer, E. L. L.; Hirsh, L.; Paladin, L.; Piovesan, D.; Tosatto, S. C. E.; Finn, R. D. The Pfam protein families database in 2019. *Nucleic Acids Res.* **2019**, *47*, D427–D432, doi:10.1093/nar/gky995.
49. Mitchell, A. L.; Attwood, T. K.; Babbitt, P. C.; Blum, M.; Bork, P.; Bridge, A.; Brown, S. D.; Chang, H.-Y.; El-Gebali, S.; Fraser, M. I.; Gough, J.; Haft, D. R.; Huang, H.; Letunic, I.; Lopez, R.; Luciani, A.; Madeira, F.; Marchler-Bauer, A.; Mi, H.; Natale, D. A.; Finn, R. D. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* **2019**, *47*, D351–D360, doi:10.1093/nar/gky1100.
50. Paull, E. O.; Carlin, D. E.; Niepel, M.; Sorger, P. K.; Haussler, D.; Stuart, J. M. Discovering causal pathways linking genomic events to transcriptional states using Tied Diffusion Through Interacting Events (TieDIE). *Bioinformatics* **2013**, *29*, 2757–2764, doi:10.1093/bioinformatics/btt471.
51. Türei, D.; Korcsmáros, T.; Saez-Rodriguez, J. OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nat. Methods* **2016**, *13*, 966–967, doi:10.1038/nmeth.4077.
52. Garcia-Alonso, L.; Holland, C. H.; Ibrahim, M. M.; Turei, D.; Saez-Rodriguez, J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res.* **2019**, *29*, 1363–1375, doi:10.1101/gr.240663.118.
53. Jones, E. J.; Matthews, Z. J.; Gul, L.; Sudhakar, P.; Treveil, A.; Divekar, D.; Buck, J.; Wrzesinski, T.; Jefferson, M.; Armstrong, S. D.; Hall, L. J.; Watson, A. J. M.; Carding, S. R.; Haerty, W.; Di Palma, F.; Mayer, U.; Powell, P. P.; Hautefort, I.; Wileman, T.; Korcsmaros, T. Integrative analysis of Paneth cell proteomic and transcriptomic data from intestinal organoids reveals functional processes dependent on autophagy. *Dis. Model. Mech.* **2019**, *12*, doi:10.1242/dmm.037069.
54. Tschurtschenthaler, M.; Adolph, T. E. The selective autophagy receptor optineurin in crohn's disease. *Front. Immunol.* **2018**, *9*, 766, doi:10.3389/fimmu.2018.00766.

55. Kuballa, P.; Huett, A.; Rioux, J. D.; Daly, M. J.; Xavier, R. J. Impaired autophagy of an intracellular pathogen induced by a Crohn's disease associated ATG16L1 variant. *PLoS ONE* **2008**, *3*, e3391, doi:10.1371/journal.pone.0003391.
56. Li, Q.; Wang, C.; Tang, C.; He, Q.; Li, N.; Li, J. Dysbiosis of gut fungal microbiota is associated with mucosal inflammation in Crohn's disease. *J. Clin. Gastroenterol.* **2014**, *48*, 513–523, doi:10.1097/MCG.0000000000000035.
57. Nagao-Kitamoto, H.; Shreiner, A. B.; Gillilland, M. G.; Kitamoto, S.; Ishii, C.; Hirayama, A.; Kuffa, P.; El-Zaatari, M.; Grasberger, H.; Seekatz, A. M.; Higgins, P. D. R.; Young, V. B.; Fukuda, S.; Kao, J. Y.; Kamada, N. Functional Characterization of Inflammatory Bowel Disease-Associated Gut Dysbiosis in Gnotobiotic Mice. *Cell. Mol. Gastroenterol. Hepatol.* **2016**, *2*, 468–481, doi:10.1016/j.jcmgh.2016.02.003.
58. Erickson, A. R.; Cantarel, B. L.; Lamendella, R.; Darzi, Y.; Mongodin, E. F.; Pan, C.; Shah, M.; Halfvarson, J.; Tysk, C.; Henrissat, B.; Raes, J.; Verberkmoes, N. C.; Fraser, C. M.; Hettich, R. L.; Jansson, J. K. Integrated metagenomics/metaproteomics reveals human host-microbiota signatures of Crohn's disease. *PLoS ONE* **2012**, *7*, e49138, doi:10.1371/journal.pone.0049138.
59. Durães, C.; Machado, J. C.; Portela, F.; Rodrigues, S.; Lago, P.; Cravo, M.; Ministro, P.; Marques, M.; Cremers, I.; Freitas, J.; Cotter, J.; Tavares, L.; Matos, L.; Medeiros, I.; Sousa, R.; Ramos, J.; Deus, J.; Caldeira, P.; Chagas, C.; Duarte, M. A.; Magro, F. Phenotype-genotype profiles in Crohn's disease predicted by genetic markers in autophagy-related genes (GOIA study II). *Inflamm. Bowel Dis.* **2013**, *19*, 230–239, doi:10.1002/ibd.23007.
60. Rioux, J. D.; Xavier, R. J.; Taylor, K. D.; Silverberg, M. S.; Goyette, P.; Huett, A.; Green, T.; Kuballa, P.; Barmada, M. M.; Datta, L. W.; Shugart, Y. Y.; Griffiths, A. M.; Targan, S. R.; Ippoliti, A. F.; Bernard, E.-J.; Mei, L.; Nicolae, D. L.; Regueiro, M.; Schumm, L. P.; Steinhart, A. H.; Brant, S. R. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat. Genet.* **2007**, *39*, 596–604, doi:10.1038/ng2032.
61. Kubisch, J.; Türei, D.; Földvári-Nagy, L.; Dunai, Z. A.; Zsákai, L.; Varga, M.; Vellai, T.; Csermely, P.; Korcsmáros, T. Complex regulation of autophagy in cancer - integrated approaches to discover the networks that hold a double-edged sword. *Semin. Cancer Biol.* **2013**, *23*, 252–261, doi:10.1016/j.semcancer.2013.06.009.
62. Carey, R.; Jurickova, I.; Ballard, E.; Bonkowski, E.; Han, X.; Xu, H.; Denson, L. A. Activation of an IL-6:STAT3-dependent transcriptome in pediatric-onset inflammatory bowel disease. *Inflamm. Bowel Dis.* **2008**, *14*, 446–457, doi:10.1002/ibd.20342.
63. Montero-Meléndez, T.; Llor, X.; García-Planella, E.; Perretti, M.; Suárez, A. Identification of novel predictor classifiers for inflammatory bowel disease by gene expression profiling. *PLoS ONE* **2013**, *8*, e76235, doi:10.1371/journal.pone.0076235.
64. Vancamelbeke, M.; Vanuytsel, T.; Farré, R.; Verstockt, S.; Ferrante, M.; Van Assche, G.; Rutgeerts, P.; Schuit, F.; Vermeire, S.; Arijs, I.; Cleynen, I. Genetic and transcriptomic bases of intestinal epithelial barrier dysfunction in inflammatory bowel disease. *Inflamm. Bowel Dis.* **2017**, *23*, 1718–1729, doi:10.1097/MIB.0000000000001246.
65. Clough, E.; Barrett, T. The gene expression omnibus database. *Methods Mol. Biol.* **2016**, *1418*, 93–110, doi:10.1007/978-1-4939-3578-9_5.

66. Blander, J. M. Death in the intestinal epithelium-basic biology and implications for inflammatory bowel disease. *FEBS J.* **2016**, *283*, 2720–2730, doi:10.1111/febs.13771.
67. Fitzwalter, B. E.; Towers, C. G.; Sullivan, K. D.; Andrysik, Z.; Hoh, M.; Ludwig, M.; O'Prey, J.; Ryan, K. M.; Espinosa, J. M.; Morgan, M. J.; Thorburn, A. Autophagy inhibition mediates apoptosis sensitization in cancer therapy by relieving foxo3a turnover. *Dev. Cell* **2018**, *44*, 555-565.e3, doi:10.1016/j.devcel.2018.02.014.
68. Cho, S.-J.; Yun, S.-M.; Jo, C.; Lee, D.-H.; Choi, K. J.; Song, J. C.; Park, S. I.; Kim, Y.-J.; Koh, Y. H. SUMO1 promotes Aβ production via the modulation of autophagy. *Autophagy* **2015**, *11*, 100–112, doi:10.4161/15548627.2014.984283.
69. Polager, S.; Ofir, M.; Ginsberg, D. E2F1 regulates autophagy and the transcription of autophagy genes. *Oncogene* **2008**, *27*, 4860–4864, doi:10.1038/onc.2008.117.
70. Li, M.; Lee, A.; Kim, K. L.; Murray, J.; Shrinidhi, A.; Sung, G.; Park, K. M.; Kim, K. Autophagy Caught in the Act: A Supramolecular FRET Pair Based on an Ultrastable Synthetic Host-Guest Complex Visualizes Autophagosome-Lysosome Fusion. *Angew Chem Int Ed Engl* **2018**, *57*, 2120–2125, doi:10.1002/anie.201711629.
71. Carroll, I. M.; Maharshak, N. Enteric bacterial proteases in inflammatory bowel disease-pathophysiology and clinical implications. *World J. Gastroenterol.* **2013**, *19*, 7531–7543, doi:10.3748/wjg.v19.i43.7531.
72. Jablaoui, A.; Kriaa, A.; Mkaouar, H.; Akermi, N.; Soussou, S.; Wysocka, M.; Wołoszyn, D.; Amouri, A.; Gargouri, A.; Maguin, E.; Lesner, A.; Rhimi, M. Fecal serine protease profiling in inflammatory bowel diseases. *Front. Cell. Infect. Microbiol.* **2020**, *10*, 21, doi:10.3389/fcimb.2020.00021.
73. Solà Tapias, N.; Vergnolle, N.; Denadai-Souza, A.; Barreau, F. The interplay between genetic risk factors and proteolytic dysregulation in the pathophysiology of inflammatory bowel disease. *J Crohns Colitis* **2020**, doi:10.1093/ecco-jcc/jjaa033.
74. Achuthan, A.; Aslam, A. S. M.; Nguyen, Q.; Lam, P.-Y.; Fleetwood, A. J.; Frye, A. T.; Louis, C.; Lee, M.-C.; Smith, J. E.; Cook, A. D.; Olshansky, M.; Turner, S. J.; Hamilton, J. A. Glucocorticoids promote apoptosis of proinflammatory monocytes by inhibiting ERK activity. *Cell Death Dis.* **2018**, *9*, 267, doi:10.1038/s41419-018-0332-4.
75. Zanzoni, A.; Spinelli, L.; Braham, S.; Brun, C. Perturbed human sub-networks by Fusobacterium nucleatum candidate virulence proteins. *Microbiome* **2017**, *5*, 89, doi:10.1186/s40168-017-0307-1.
76. Di Tommaso, P.; Palumbo, E.; Chatzou, M.; Prieto, P.; Heuer, M. L.; Notredame, C. The impact of Docker containers on the performance of genomic pipelines. *PeerJ* **2015**, *3*, e1273, doi:10.7717/peerj.1273.
77. Durack, J.; Lynch, S. V. The gut microbiome: Relationships with disease and opportunities for therapy. *J. Exp. Med.* **2019**, *216*, 20–40, doi:10.1084/jem.20180448.
78. Celi, P.; Cowieson, A. J.; Fru-Nji, F.; Steinert, R. E.; Kluenter, A. M.; Verlhac, V. Gastrointestinal functionality in animal nutrition and health: New opportunities for sustainable animal production. *Anim. Feed Sci. Technol.* **2017**, *234*, 88–100, doi:10.1016/j.anifeedsci.2017.09.012.
79. Turner, T. R.; James, E. K.; Poole, P. S. The plant microbiome. *Genome Biol.* **2013**, *14*, 209, doi:10.1186/gb-2013-14-6-209.
80. Ni, Y.; Yu, G.; Chen, H.; Deng, Y.; Wells, P. M.; Steves, C. J.; Ju, F.; Fu, J. M2IA: a web server for

microbiome and metabolome integrative analysis. *Bioinformatics* **2020**, doi:10.1093/bioinformatics/btaa188.

81. Harcombe, W. R.; Riehl, W. J.; Dukovski, I.; Granger, B. R.; Betts, A.; Lang, A. H.; Bonilla, G.; Kar, A.; Leiby, N.; Mehta, P.; Marx, C. J.; Segrè, D. Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. *Cell Rep.* **2014**, *7*, 1104–1115, doi:10.1016/j.celrep.2014.03.070.