Identifying and Filling the Chemobiological Gaps of Gut Microbial Metabolites

Cristian Orgaz,[†] Andrés Sánchez-Ruiz,[†] and Gonzalo Colmenarejo[*]

Biostatistics and Bioinformatics Unit

IMDEA Food

CEI UAM+CSIC

E28049 Madrid, Spain

[*]Corresponding Author

e-mail: gonzalo.colmenarejo@imdea.org

[†]These two authors contributed equally to this work

ABSTRACT

Human gut microbial metabolites are currently undergoing much research due to their involvement in multiple biological processes important for health, including immunity, metabolism, nutrition, and the nervous system. Metabolites exert their effect through the interaction with host and bacterial proteins, suggesting the use of "metabolite-mimetic" molecules as drugs and nutraceutics. In the present work, we retrieve and analyze the full set of published interactions of these compounds with human and microbiome-relevant proteins, and find patterns in their structure, chemical class, target class, and biological origins. In addition, we use virtual screening to expand (> 4-fold) the interactions, validate them with retrospective analyses, and use bioinformatic tools to prioritize them based on biological relevance. In this way, we fill many of the chemobiological gaps observed in the published data. By providing these interactions we expect to speed up the full clarification of the chemobiological space of these compounds, by suggesting many reliable predictions for fast, focused experimental testing.

KEYWORDS:

Metabolite mimetic, gut microbiome, gut metabolome, new drug modalities, drug design, nutraceutic design, drug target

INTRODUCTION

Research on microbial metabolites is currently having a large impetus, given the growing body of knowledge about their relevance in multiple biological processes, including immunity, metabolism, nutrition, and the nervous system, and therefore in human health.[1–13] The human body hosts trillions of microbial cells,[14] mainly concentrated in the gut, and this microbial community exert their influence in host biological processes through the establishment of a myriad of molecular interactions between metabolites and protein receptors.[3,15–24] This, in addition to the fact that the gut microbiome exceeds by three orders of magnitude the human genome (> 22 million vs 21000 human genes),[25] has suggested the use of "metabolite-mimetic" molecules as drugs and nutraceutics that could modulate these interactions in pathological processes.[10,26–36] As a matter of fact, this paradigm has spurred multiple drug discovery programs in the pharma industry that have reached clinical phases, and that include diseases such as inflammatory bowel disease (IBD), Crohn's disease, autism, Parkinson disease, multiple sclerosis, thrombosis, and atherosclerosis.[37]

In order to decode this interactome, multiple experimental approaches are being used and developed, including chemoproteomics,[38,39] reverse screening of panels of proteins,[17,40] and combinations of bioinformatic analyses and synthetic biology with screening.[18] However, in this area the use of virtual screening tools could speed up the identification of interactions, as the predicted interactions could provide a set of hypothesis for fast, focused experimental tests, instead of blind and expensive all vs all screening. In this regard, we have recently provided a large set of virtual screening predictions to speed up the full characterization of the chemobiological space of food

compounds.[41] These interactions were retrospectively validated, and right now many of these interactions are being confirmed experimentally.[42]

In the present work we retrieve and analyze the full published set of interactions of gut metabolites with proteins, both human and gut microbial. A total of 2193 interactions are found, comprising 405 and 128 metabolites interacting with 451 human- and 56 bacterial targets, respectively. Patterns in this chemobiological space are identified, in terms of structures and chemical class of the metabolites, and of target class and biological origin of the proteins. After applying ligand-based virtual screening techniques, this set is expanded up to 9711 interactions (> 4-fold increase), now comprising 422 and 330 compounds, interacting with 1047 human- and 180 microbial ones, respectively. The interactions are analyzed again, and retrospectively validated experimentally, showing confirmation for predictions even for rather extrapolated chemical spaces. Finally, bioinformatic tools are used to identify subsets of "high priority" proteins and interactions of high biological relevance, both host and microbial. We expect that by providing the results of this work in an open, collaborative fashion, the clarification of the full chemobiological space of gut microbial metabolites will be significantly accelerated. This work will complement another recent one where we have analyzed and developed predictive models for the same set of molecules from the point of view of their structures, physicochemical properties, and distribution.[43]

METHODS

All data analysis was performed with Python 3.10, with RDKit 2022.03.2 as cheminformatic toolkit. Compound structures for metabolites were the same as in our previous work,[43] that is, the subset of "detected and quantified" and "detected but not quantified" molecules from the gut compartment in the Human Metabolome Database (HMDB)[44] plus some additions obtained from literature searches,[15,16] resulting in a total of 6663 molecules. Metabolites were retrieved irrespective of their origin: host, microbial, both, xenobiotics, and unknown, since in this work we are interested in the bidirectional communication between the host and the microbiome, that could also involve human metabolites, with bacterial proteins. Some comparative analyses used a subset of the DrugBank[45] (1410 molecules) that included the small molecules in approved, not-withdrawn, non-illicit status, and that were administered orally and acted systemically. A few shared compounds were manually assigned to the metabolite or drug set. Molecular structures were subject to the same pipeline as that described before[41,43,46,47] for standardization.

Biological interactions of these molecules with human or bacterial proteins were retrieved from both ChEMBL (release 33)[48] and BindingDB,[49] and were complemented with literature searches.[15,16] Only data from well-defined protein or protein complexes, and obtained through dose-response analysis (standardized in the form of pCHEMBL or pCHEMBL-like values, that is $-\log_{10}$ (molar IC50, XC50, EC50, AC50, Ki, Kd or Potency)), were retrieved. In the case of bacterial proteins, only those of a set of 332 gut genera (collected in Table S1 in Supporting Information) present in typical microbiome data were retrieved.

For the predicted interactions we used the SEA (Similarity Ensemble Approach) method, available in https://sea.bkslab.org/. Only interactions with p-value <= 1e-16 were retained, as they are considered by the program to be the most reliable ones.

Compounds were classified into 17 chemical classes derived from the ClassyFire chemical hierarchy.[50] Proteins were classified into a set of 24 target classes derived from the ChEMBL target hierarchy, after mapping each protein into a unique UniProt accession number (or set of numbers, in the case of protein complexes). Redundant entries obtained from different sources were removed, and multiple entries of each compound-protein pair were averaged to obtain a unique mean affinity value.

Target analysis was performed with PHAROS (https://pharos.nih.gov/)[51] in the case of human proteins. For bacterial proteins, essentiality was predicted using Geptop 2.0.[52]

Post-hoc analysis of contingency tables were performed through a Fisher exact approach as described elsewhere,[53] and p-values were adjusted by Benjamini-Hochberg correction; significance level was 0.05.

RESULTS

Figure 1 displays the distribution of the gut microbial metabolites used in this work across 17 chemical classes derived from the ClassyFire[50] chemical taxonomy. These very different chemical classes reflects the wide chemical space spanned by the metabolites detected in the intestinal milieu.

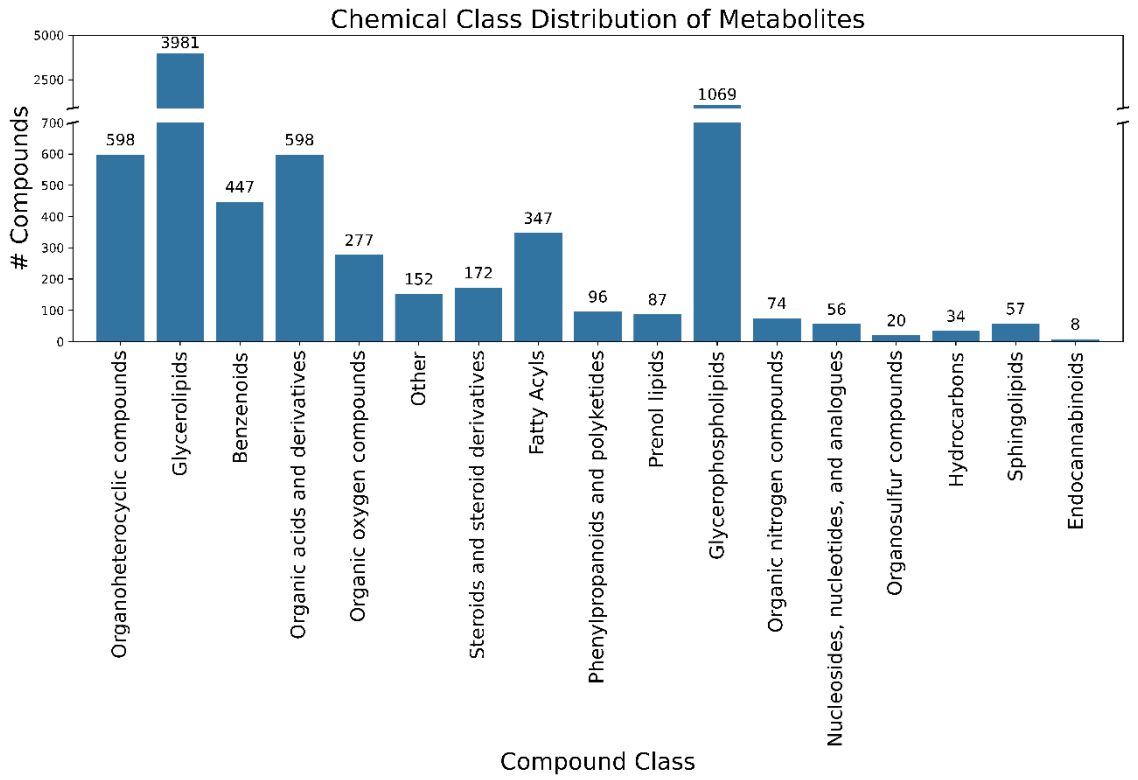

Figure 1. *Distribution by chemical classes (based on the ClassyFire taxonomy) of the 6663 gut metabolites used in this work.*

The distribution is overwhelmingly dominated by the "Glycerolipids" and "Glycerophospholipids" classes (~4000 and ~1000 compounds each, respectively) as expected, given the large number of different structures of these molecules in the diet. The following most populated classes are "Organoheterocyclic compounds" and "Organic acids and derivatives, both with ~600 compounds, and after them

"Benzenoids", "Fatty Acyls" and "Organic oxygen compounds" (> 200 compounds the three of them). The remaining classes have < 200 molecules each.

Different types of well-known microbial metabolites map into the different chemical classes. For example, among the lipids, the "Steroids and steroid derivatives" class include the bile acids, sterol lipids, and cholesterol derivatives; short-chain fatty acids (SCFA) map into both the "Organic acids and derivatives" and "Fatty Acyls" (that include fatty acids) classes, depending on the size; fatty amides and long-chain fatty acids (LCFA) are both classified as "Fatty Acyls"; and glycerolipids, glycerophospholipids, and sphingolipids are classified in the same-name classes. As regarding non-lipidic molecules, both trimethylamine-N-oxide (TMAO) and polyamines map into the "Organic nitrogen compounds", while indole derivatives are classified in the "Organoheterocyclic compounds", and branched-chain amino acids and derivatives are assigned to the "Organic acids and derivatives" class, etc.

Analysis of the published bioactivities of gut microbial metabolites

As described in Materials and Methods, the interactions of this set of metabolites with proteins of *Homo sapiens* and bacterial species belonging to a set of genuses typical in intestinal microbiome samples were retrieved from ChEMBL, BindingDB, and literature searches. A total of 2139 interactions were identified, comprising 507 targets and 426 metabolites. This results in ratios of 0.32 interactions per metabolite, 0.076 targets per metabolite, and 6.4% of metabolites with at least one interaction described (of all the 6663 metabolites). There are almost 10-fold interactions more with human proteins than with bacterial ones (1942 vs 197, respectively), corresponding to 451 human targets vs 56 bacterial ones, and 405 vs 128 metabolites interacting with human vs

bacterial proteins (note 451 + 128 > 426 as some compounds interact with both human and bacterial proteins). The distributions of compounds per target and targets per compound display typical long right tails. The target with the largest number of compounds (80) is aldehyde dehydrogenase 1A1, an oxidoreductase displaying a wide range of chemotypes as ligands, followed by the thyrotropin receptor (60), and prelamin-A/C (52 compounds). The compound with the largest number of interacting proteins (89) is luteolin, a flavonoid, followed by apigenin (86), another flavonoid. In both cases these compounds interact with a wide range of protein classes, including kinases, oxidoreductases, and proteases.

Figure 2 shows the distribution of unique protein targets of the metabolites among 25 target classes derived from the ChEMBL target hierarchy (orange bars). For comparison purposes, the same distribution is shown for a set of oral drugs with systemic action (blue bars). We can see that the set of drugs has many more interacting targets than the metabolites, reflecting that the former set has been much more tested, as well as the scarce known target space of the later. As a matter of fact, 78.2% of the drug compounds have one or more interacting target, in stark contrast with the 6.4% mentioned before for the metabolites.

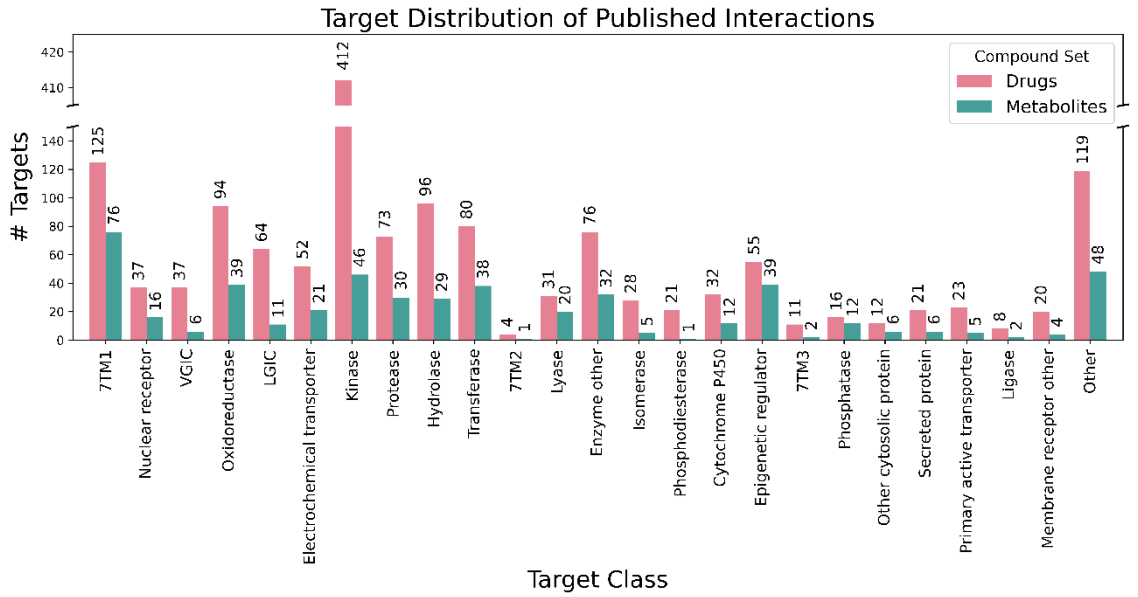

Figure 2. *Distribution of unique targets among target classes for interactions published for gut metabolites (green bars). The same distribution is shown for a set of oral systemic drugs (magenta bars).*

There is a significant correlation between these two distributions (Spearman rank correlation = 0.89), as well as some clear discrepancies: “Kinase” is, by far, the most populated class in drugs, followed by “7TM1”, but in metabolites the first class is the later, while “Kinase” is the third one; another large difference is “Hydrolases”, the 4th most populated target class in drugs but the 9th in metabolites; and “Epigenetic regulator”, the 4^{h} class in metabolites (together with “Oxidoreductase”), but the 10th in drugs. From the counts of shared vs not shared targets between the two compound sets (Table S2 in Supporting Information) it is possible to see that the target class having the largest counts of targets only interacted with metabolites is “Other” (24), followed by “Transferase” (23), “7TM1” (18) and “Epigenetic regulator” (17).

On the other hand, Figure 3 displays the distribution of metabolite targets by their target classes and biological group: human (blue bars) vs bacterial (orange bars).

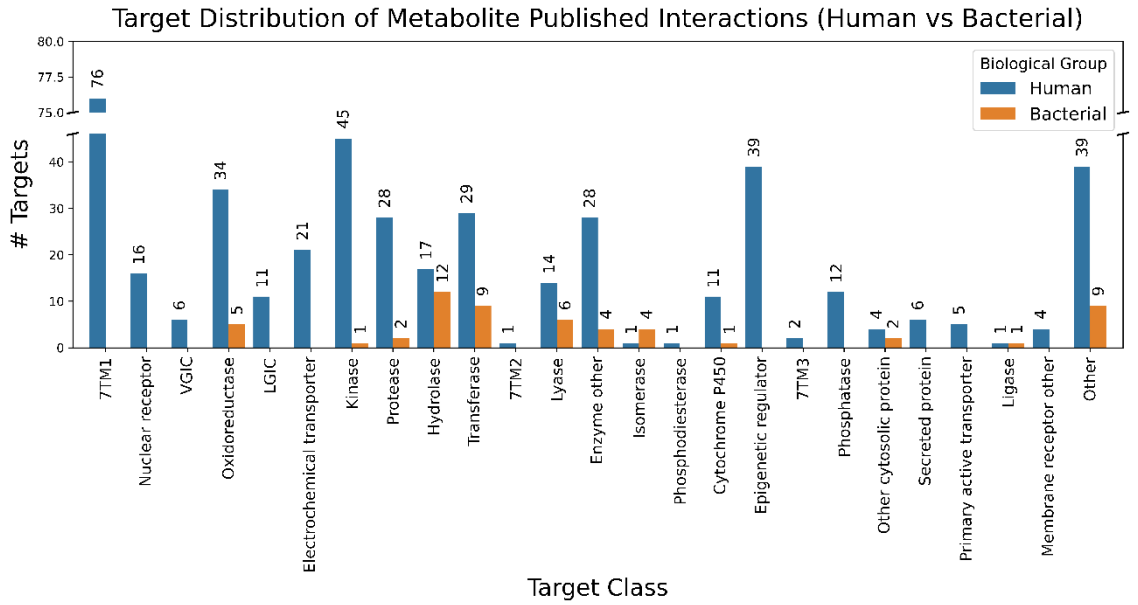

Figure 3. *Distribution of unique targets among target classes for interactions published for gut metabolites. Both human (blue bars) and bacterial targets (orange bars) are shown.*

It is observed that the majority of the targets interacting with metabolites in the literature are human, as above described. The two distributions are very different. As expected, bacterial targets are missing from typical eukaryotic target classes: “7TM1-3”, “Nuclear receptor”, “Epigenetic regulator”, “VGIC”, “LGIC”. Other target classes are missing in this dataset as far as bacterial proteins are concerned: “Electrochemical transporter”, “Phosphodiesterase”, “Phosphatase”, “Secreted protein”, “Primary active transporter” and “Membrane receptor other”. There is one bacterial kinase, streptokinase A, and one cytochrome P450-like protein, the steroid C26-monooxygenase. While the human proteins have “7TM1” as the most frequent target class (76 proteins), followed by “Kinase” (45) > “Epigenetic regulator” = “Other” (39) > “Oxidoreductase” (34), bacterial proteins have “Hydrolase” as the largest target class

(12 proteins) followed by “Transferase” = “Other” (9) > “Lyase” (6) and “Oxidoreductase” (5).

Figure 4 shows the distribution of unique compounds targeting the different human or bacterial proteins (blue vs orange bars, respectively), across the 25 target classes.

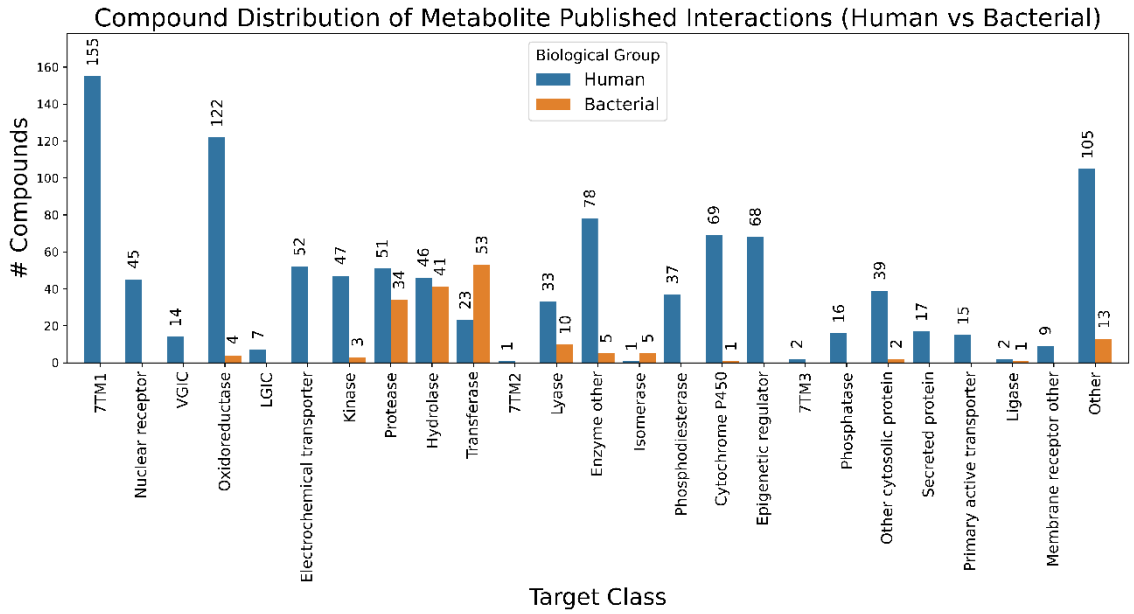

Figure 4. *Distribution of unique compounds among target classes for interactions published for gut metabolites. Both human (blue bars) and bacterial targets (orange bars) are shown.*

Here we find that the target class share of metabolites interacting with human proteins does not follow strictly the number of unique targets in each class, and goes in the following decreasing order: “7TM1” (155 compounds) > “Oxidoreductase” (122) > “Other” (105); between 10 and 100 compounds we have “Enzyme other” (78) > “Cytochrome P450” (69) > “Epigenetic regulator” (68) > “Electrochemical transporter” (52) > “Protease” (51) > “Kinase” (47) > “Nuclear receptor” (45) > “Hydrolase” (46) > “Other cytosolic protein” (39) > “Phosphodiesterase” (37) > “Lyase” (33) > “Secreted

Protein” (17) >“Phosphatase” (16) > “Primary active transporter” (15) > “VGIC” (14). The rest of the target classes have less than 10 compounds.

As regarding the metabolites interacting with bacterial targets, these are restricted to “Transferase” (53 molecules) > “Hydrolase” (41) > “Protease” (34) > “Other” (13) > “Lyase” (10); the rest of the classes have less than 10 compounds.

It is also possible to analyze the interactions in terms of both chemical and target classes. Figure 5 depicts the count distribution of the interactions for human proteins and published data, while Figure 6 displays the same distribution but with bacterial proteins.

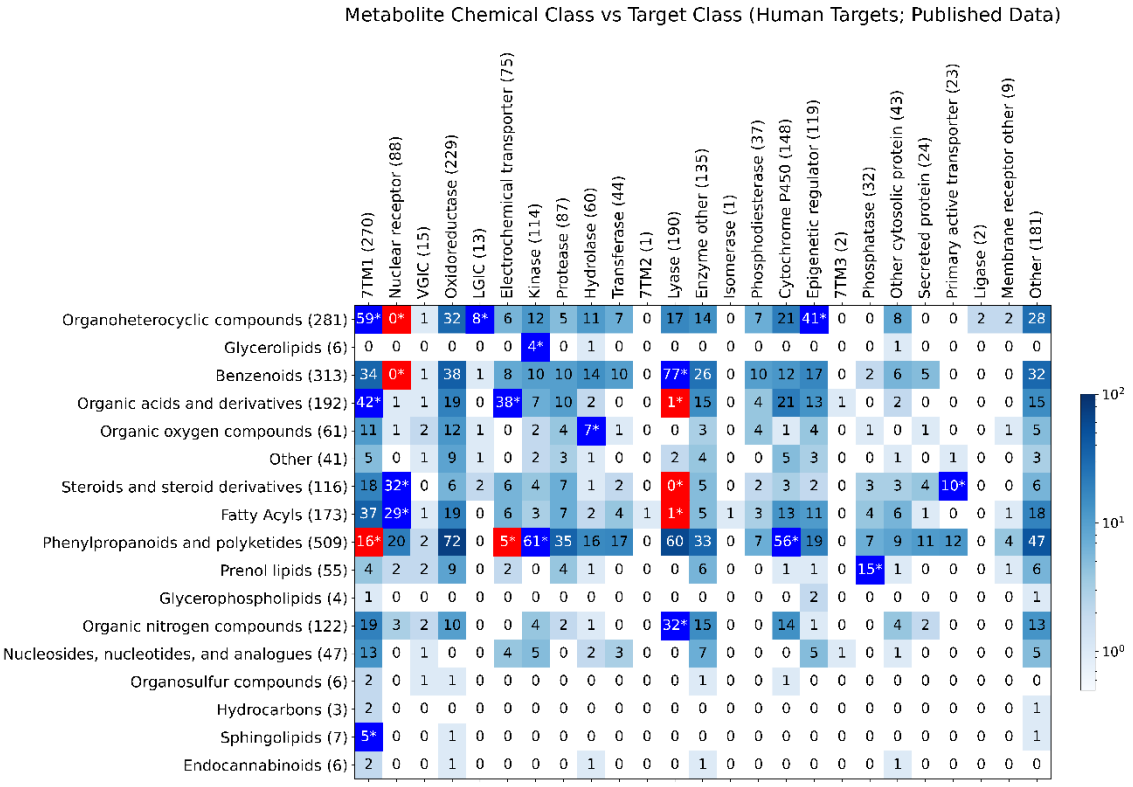

Figure 5. *Distribution of unique interactions among chemical and target classes for gut metabolites and human proteins present in published data. Counts correspond to unique InChi + UniProt accession combinations. Cells in bright blue or bright red*

with starred counts correspond to statistically significant combinations (blue overrepresented and blue underrepresented, respectively) in a post-hoc test after Benjamini-Hochberg p-value correction. Numbers within parenthesis after each label correspond to marginal counts.

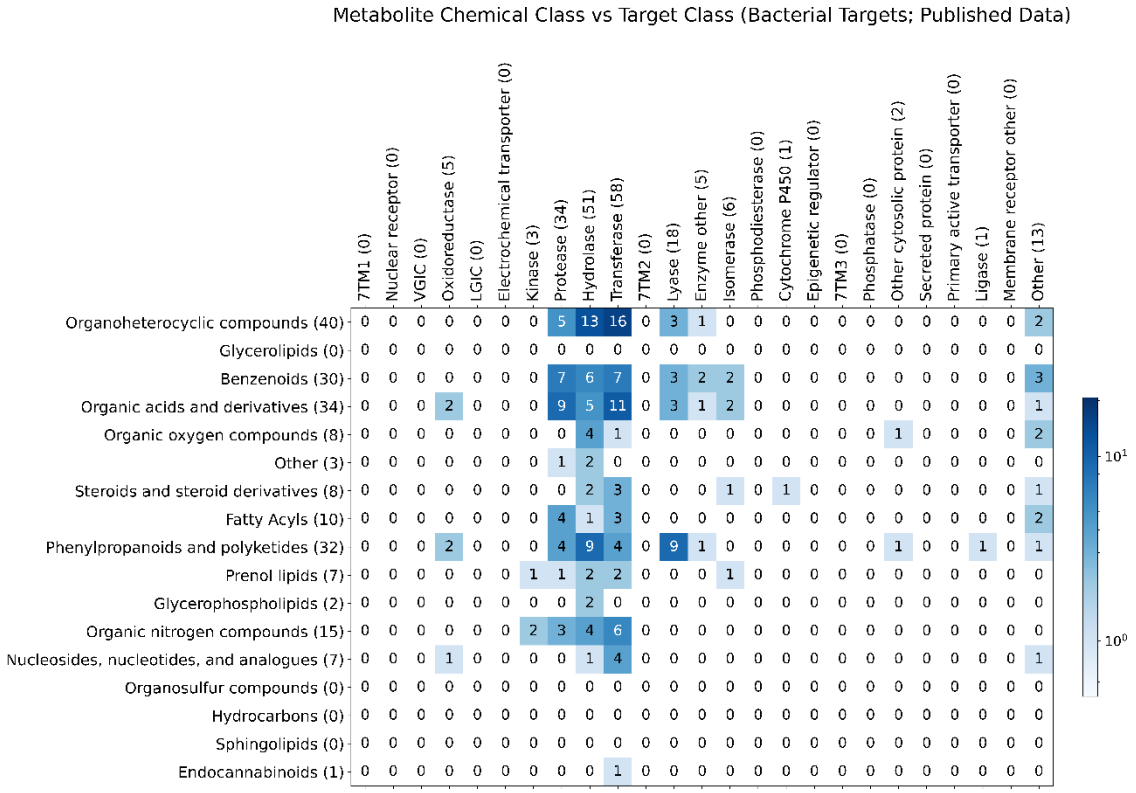

Figure 6*. Distribution of unique interactions among chemical and target classes for gut metabolites and bacterial proteins present in published data. Counts correspond to unique InChi + UniProt accession combinations. No significant cells after Benjamini-Hochberg p-value correction are obtained. Numbers within parenthesis after each label correspond to marginal counts.*

In the case of human proteins, we see that, of the 17 chemical classes, “Phenylpropanoids and polyketides” is the one with the largest number of interactions

(509), followed by “Benzenoids” (313), “Organoheterocyclic compounds” (281), “Organic acids and derivatives” (192), and “Fatty Acyls” (173). These five classes together comprise ~75% of the published interactions with human proteins. In terms of target classes, the marginal distribution is wider, where “7TM1” has the largest share of interactions (270), followed by “Oxidoreductase” (229), “Lyase” (190), “Other” (181), and “Cytochrome P450” (148). Together they comprise ~52% of the interactions. Statistically overrepresented combinations are: “Organoheterocyclic compounds” vs “7TM1”, “LGIC”, and “Epigenetic regulator”; “Glycerolipids” vs “Kinase”; “Benzenoids” vs “Lyase”; “Organic acids and derivatives” vs “7TM1” vs “Electrochemical transporter”; “Organic oxygen compounds” vs “Hydrolase”; “Steroid and steroid derivatives” vs “Nuclear receptor” and “Primary active transporter”; “Phenylpropanoids and polyketides” vs “Kinase” and “Cytochrome P450”; “Prenol lipids” vs “Phosphatase”; “Organic nitrogen compounds” vs “Lyase”; and “Sphingolipids” vs “7TM1”. Statistically underrepresented combinations are: “Organoheterocyclic compounds” vs “Nuclear receptor”; “Benzenoids” vs “Nuclear receptor”; “Organic acids and derivatives” vs “Lyase”; “Steroids and steroid derivatives” vs “Lyase”; “Fatty Acyls” vs “Lyase”; and “Phenylpropanoids and polyketides” vs “7TM1” and “Electrochemical transporter”.

As regarding the heatmap of interactions with bacterial proteins, the top-5 chemical classes are now “Organoheterocyclic compounds” (40), “Organic acids and derivatives” (34), “Phenylpropanoids and polyketides” (32), “Benzenoids” (30), and “Organic nitrogen compounds” (15). The sum of these interactions corresponds ~77% of all ones. In terms of target classes, now the distribution is even more concentrated in a few of them: “Transferase” (58), “Hydrolase” (51), and “Protease” (34) alone comprise

~72% of the interactions. Although in this dataset no significance is obtained from any of the chemical class vs target class combinations in the *post hoc* analysis of the contingency table represented in Figure 6, the four most populated combinations are the following ones: "Organoheterocyclic compounds" vs "Transferase" (16 interactions) and "Hydrolase" (13), "Organic acids and derivatives" vs "Transferase" (11); and 9 interactions are observed for "Phenylpropanoids and polyketides" vs both "Hydrolase" and "Lyase", as well as "Organic acids and derivatives" vs "Protease".

The full set of published interactions can be found in Table S3 in the Supporting Information.

Structural analysis of the metabolites interacting with the different target classes

The structures of the metabolites interacting with the different target classes show specific patterns worth mentioning. In Figure 7 the number of unique Bemis-Murcko scaffolds[54,55] for the metabolites interacting with each target class are shown as bars, together with the percentage of molecules devoid of scaffolds (molecules with no rings), as dot-connected lines.

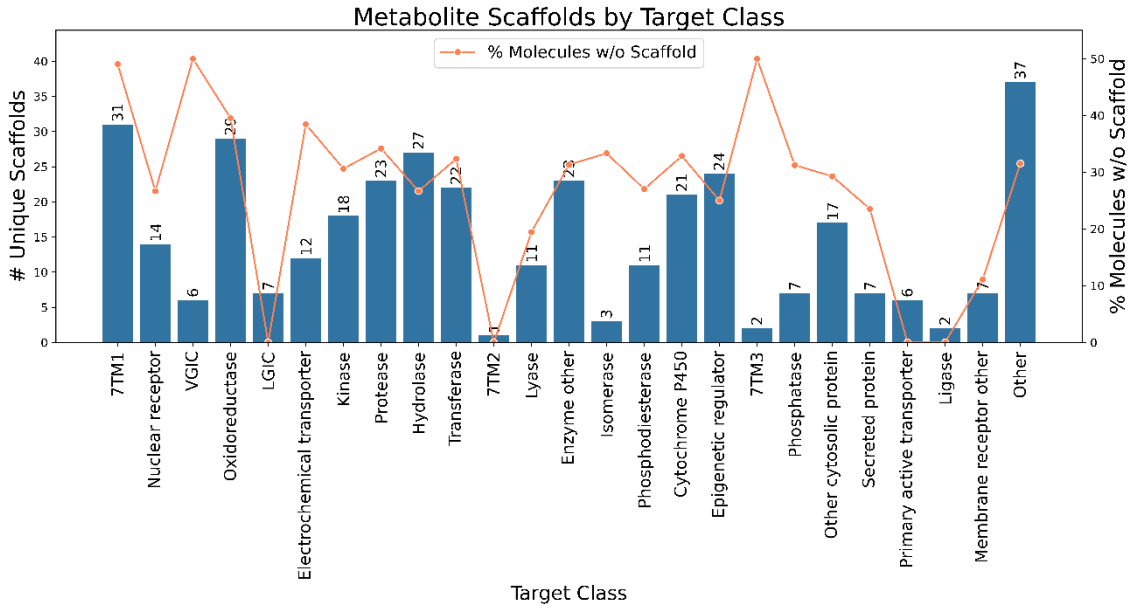

***Figure 7**. Distribution of unique scaffolds by target class for gut metabolites (blue bars) and percentage of these molecules without scaffold (orange dot-connected line)*

The largest scaffold diversities are observed in the “Other”, “7TM1”, “Oxidoreductase” and “Hydrolase” target classes. These are target classes that bind multiple chemotypes. In the case of “7TM1”, they are known to interact with organoheterocyclic metabolites like indole derivatives, plus short- and long-chain fatty acids, fatty amides, hydroxycarboxylic acids, bile acids, benzenoids, etc.[15–17,24] On the other hand, the “Other” class is by definition quite variable as it contains multiple protein families, so as expected it has the highest scaffold diversity. In turn, “Oxidoreductase” contains proteins quite different like cyclooxygenases, monoamine oxidases, lipoxygenases, etc., and these are targeted by diverse chemical families such as prenol lipids, flavonoids, etc.; while the “Hydrolase” target class include a wide range of inhibitors and substrates as well (e.g. flavonoids, xanthine derivatives, etc.)

that interact with very different redox enzymes: cholinesterases, phospholipases, endonucleases, etc.

This is in contrast with other target classes, where the number of unique scaffolds is much more reduced: e.g. in “Nuclear receptor” proteins are targeted mainly by molecules in the “Steroids and steroid derivatives”; or that have just few compounds in the gut metabolite set: “VGIC”, “LGIC”, “7TM2”, “7TM3”, “Ligase”, etc.

On the other hand, and focusing on the target classes with large numbers of compounds, the ones with the largest percentages of molecules without scaffolds are “7TM1” (49%), “Oxidoreductase” (40%), and “Electrochemical transporter” (38%). We have seen before the large number of fatty acids and derivatives (“Fatty Acyls” and “Organic acids and derivatives”) targeting the “7TM1” class, and this applies also the two other classes.

To complete the scaffold-based structural analysis, in Figure 8 the fraction of scaffolds shared between target classes is depicted, which is also a measure of the structural similarity shared between the metabolites of the different target classes.

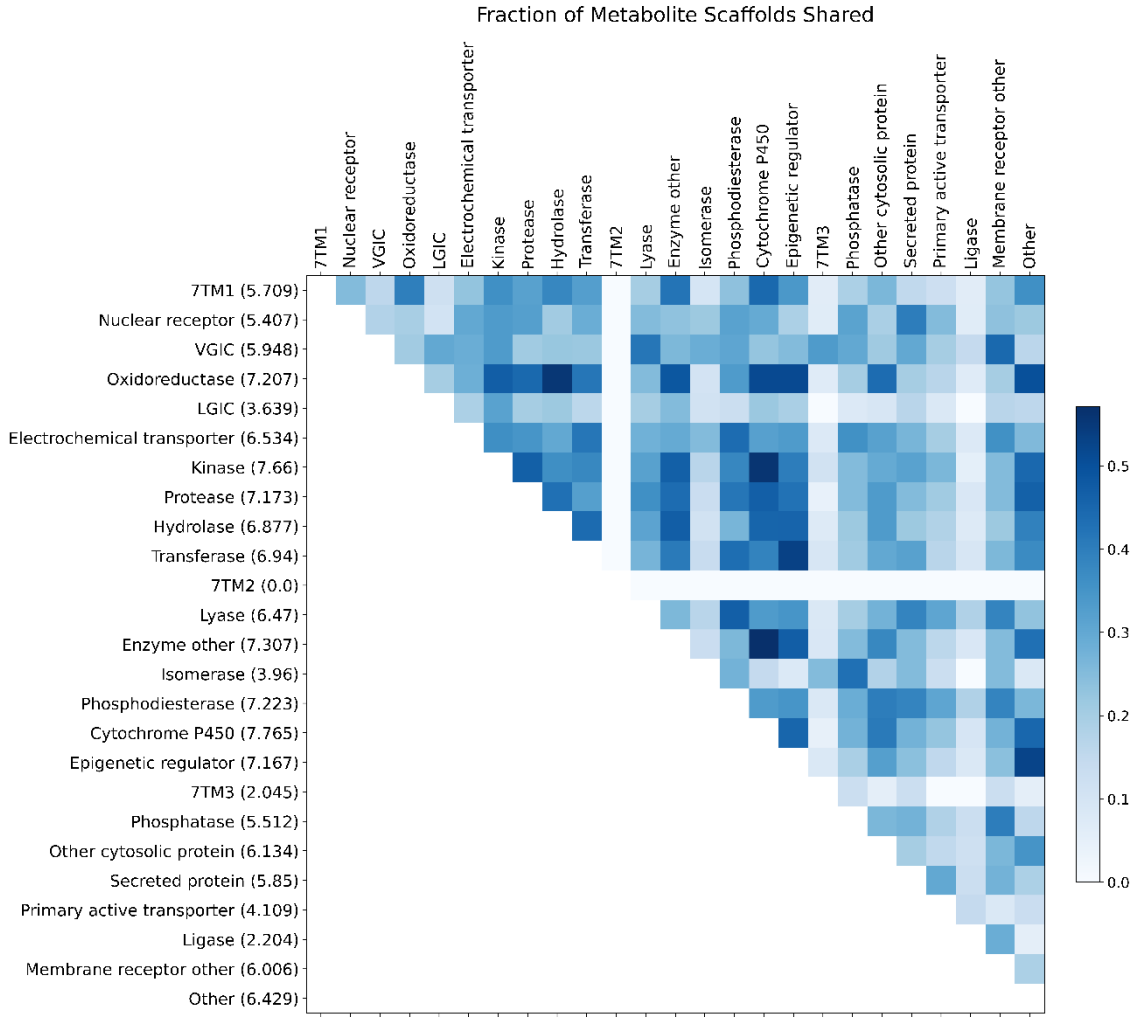

Figure 8*. Distribution of the fraction of unique scaffolds shared between pairs of target classes. For each pair of targets, the cell value is obtained as the number of shared scaffolds (intersection) divided by the total number of unique scaffolds (union) in both target classes. The absence of scaffold is counted as another scaffold. The vertical axes labels include the marginal sum of fractions.*

Scaffolds are shared the most between “Cytochrome P450” and “Enzyme other” (0.57); “Cytochrome P450” and “Kinase” (0.56); and “Hydrolase” and “Oxidoreductase” (0.55). Also between “Epigenetic regulator” and “Transferase” (0.533), and the former and “Other” (0.52). The target classes with the highest marginal sum of fractions of shared scaffolds are “Cytochrome P450”, followed by “Kinase”, and “Enzyme other”.

On the other side are “7TM2”, “7TM3”, “Ligase”, “LGIC”, and “Isomerase”, all of them with few compounds, but also there are well populated target classes like “Nuclear receptor” and “Primary active transporter” with small values, a signature of being targeted by more specific chemotypes associated to these classes.

Analysis of predicted bioactivities of gut microbial metabolites

It is possible to expand the set of published interactions by using well-established target prediction methods. In this regard, we have used the SEA method (Similarity Ensemble Approach),[56,57] which is based on testing the difference in the distribution of similarities of the query compound to the set of ligands of a target, when compared to the distribution of similarities to all the targets. This method has been successfully used in the past to, among others, identify the mechanism-of-action targets, predict bioactivities in large databases like ZINC, and identify bioactivities in inactive ingredients of drugs.[56,58,59] In this regard, after applying this method to our dataset we were able to obtain a total of 8185 interactions, comprising 376 compounds and 1029 targets. Making the union of these predicted interactions with the published ones, we obtain a total of 9711 interactions (a bigger than four-fold increase); 7572 of these are strictly new interactions, which in turn split into 6343 interactions with human proteins and 2139 with bacterial targets. The merged set of published + predicted interactions now correspond to 426 compounds and 1227 proteins. Thus, although the percentage of compounds with at least one interaction remains in 6.4%, the number of targets has more than doubled and now we have 0.184 targets per metabolite, and 1.46 interactions per metabolite. In terms of biological origin, in the merged set the number of human vs bacterial proteins is 1047 vs 180, respectively (to be compared to

the 451 vs 56 ones in the published data). Although the *total* number of compounds with at least one target annotation remains the same, the counts per biological group increase significantly as well, especially in the case of metabolites with bacterial targets: 422 vs 330, respectively for human and bacterial targets (they were 405 vs 128 metabolites, see above). Now the protein with the largest number of ligands (83) is peptidyl-glycine alpha-amidating monooxygenase, followed by aldehyde dehydrogenase 1A1 (the same 80 interactions as before), while the compound with the largest number of targets (137) is vanillic acid, a benzenoid, followed by luteolin (now with 107 interactions).

Figure 9A displays the distribution of metabolite targets by biological group of the target after including the predicted interactions (human proteins as blue bars vs bacterial proteins as orange bars), while Figure 9B displays the distribution of unique compounds targeting the different human or bacterial proteins (blue vs orange bars, respectively).

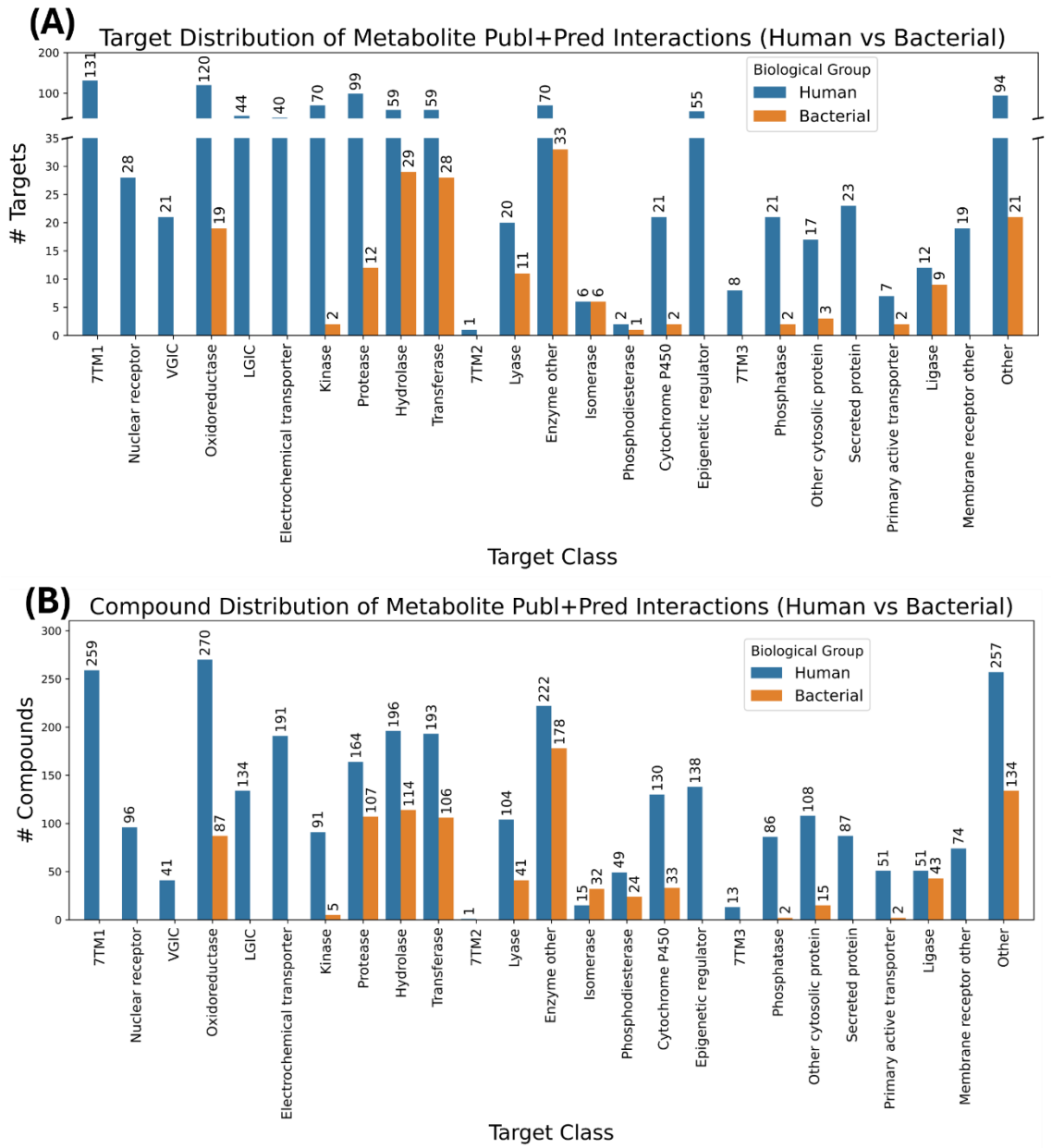

Figure 9. *(A) Distribution of unique targets among target classes for interactions published + predicted for gut metabolites. Both human (blue bars) and bacterial targets (orange bars) are shown. (B) Distribution of compounds among target classes for interactions published + predicted for gut metabolites. Both human (blue bars) and bacterial targets (orange bars) are shown.*

Looking at Figure 9A, after the incorporation of the predicted interactions, it can be seen a large boost in the number of unique human oxidoreductases (from 34 to 120),

which now are the most populated target class, followed by 7TM1, that increases from 76 to 131 targets. Another class that has experimented a notable growth is human "Protease", which was before the 9th most populated target class and now is the third one (from 51 to 99 proteins). In the case of bacterial proteins, the largest boost is in "Enzyme other" (from 4 to 33 proteins and now the most populated target class). Regarding target classes missing bacterial proteins in the published data, now we see the appearance of one bacterial phosphodiesterase, (phospholipase C), two phosphodiesterase's (tyrosine-protein phosphatase YopH and low molecular weight protein-tyrosine phosphatase A), and two primary active transporters (multidrug resistance protein MdtK and protein translocase subunit SecA 1). Typical eukaryotic target classes ("7TM1-3", "Nuclear receptor", "Epigenetic regulator", "VGIC", "LGIC") plus "Electrochemical transporter", "Secreted protein", and "Membrane receptor other" remain unpopulated in bacteria, while the rest of the classes have an increase of the number of proteins. The later also happens with all the target classes of human proteins, with the only exception of 7TM2.

As a result, the distribution of unique compounds per target class vary accordingly (Figure 9B). Here, in terms of human proteins, "Oxidoreductase" surpasses now "7TM1" as the most populated human target class (270 vs 155 compounds before); the largest boost in unique compounds is experimented by "Transferase" (193 molecules now vs 23 molecules before, from the 15th to the 6th position), followed by "Other" (257 compounds now vs 105 molecules before, from the 10th to the 5th position). Regarding compounds targeting bacterial proteins, the target classes having the largest boosts are "Enzyme other" (from 5 to 178 compounds) and "Other" (from 13 to 134 compounds), and now they are the most- and second-most populated target classes.

By analyzing these interactions in terms of both chemical class and target class, we obtain the plots in Figures 10 and 11, respectively for human and bacterial targets.

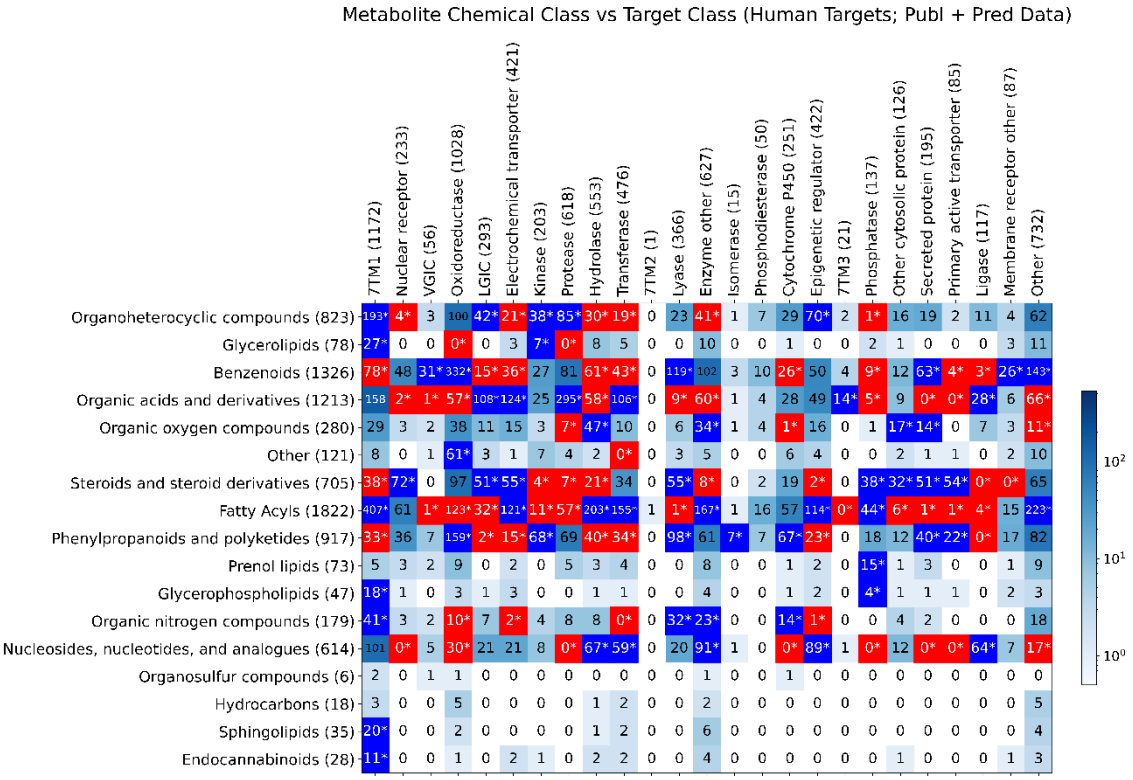

Figure 10. *Distribution of unique interactions among chemical and target classes for gut metabolites and human proteins present in published or predicted data. Counts correspond to unique InChi + UniProt accession combinations. Cells in bright blue or bright red with starred counts correspond to statistically significant combinations (blue overrepresented and blue underrepresented, respectively) in a post-hoc test after Benjamini-Hochberg p-value correction. Numbers within parenthesis after each label correspond to marginal counts.*

In both cases, we observe a large increase in the number of interactions in many cells. In particular, in the case of human proteins (Figure 10), 75 previously empty cells have now one or more interactions, and the interaction counts have increased in 230 cells. The filling of previously empty cells have been most effective in the target classes

"Membrane receptor other" and "Transferases", as well in the chemical classes "Glycerophospholipids" and "Glycerolipids", while the numbers of cells with increased counts have been largest in "7TM1", "Hydrolase", "Enzyme other", and "Other" target classes, and in the "Organoheterocyclic compounds" and "Benzenoids" chemical classes. The target classes with the largest increases in *total* counts of interactions correspond to "7TM1" again and "Oxidoreductase", while "Fatty Acyls", "Organic acids and derivatives" and "Benzenoids" are the chemical classes with largest increases in total counts.

By considering the bacterial targets (Figure 11), 50 previously empty cells have now been filled with one or more interactions, and 96 cells have increased their counts. The target classes with the most effective filling of empty cells are "Enzyme other", "Phosphodiesterase", and "Other", while the chemical classes are "Glycerophospholipids", "Benzenoids", and "Organic oxygen compounds". The target classes with largest numbers of cells with increased counts are "Enzyme other", "Other", and "Protease", while the chemical classes are "Benzenoids" and "Phenylpropanoids and polyketides". Finally, the largest increments in total counts have occurred in the "Enzyme other" and "Other" target classes, and in the "Organic acids and derivatives", "Fatty Acyls", and "Benzenoids" chemical classes.

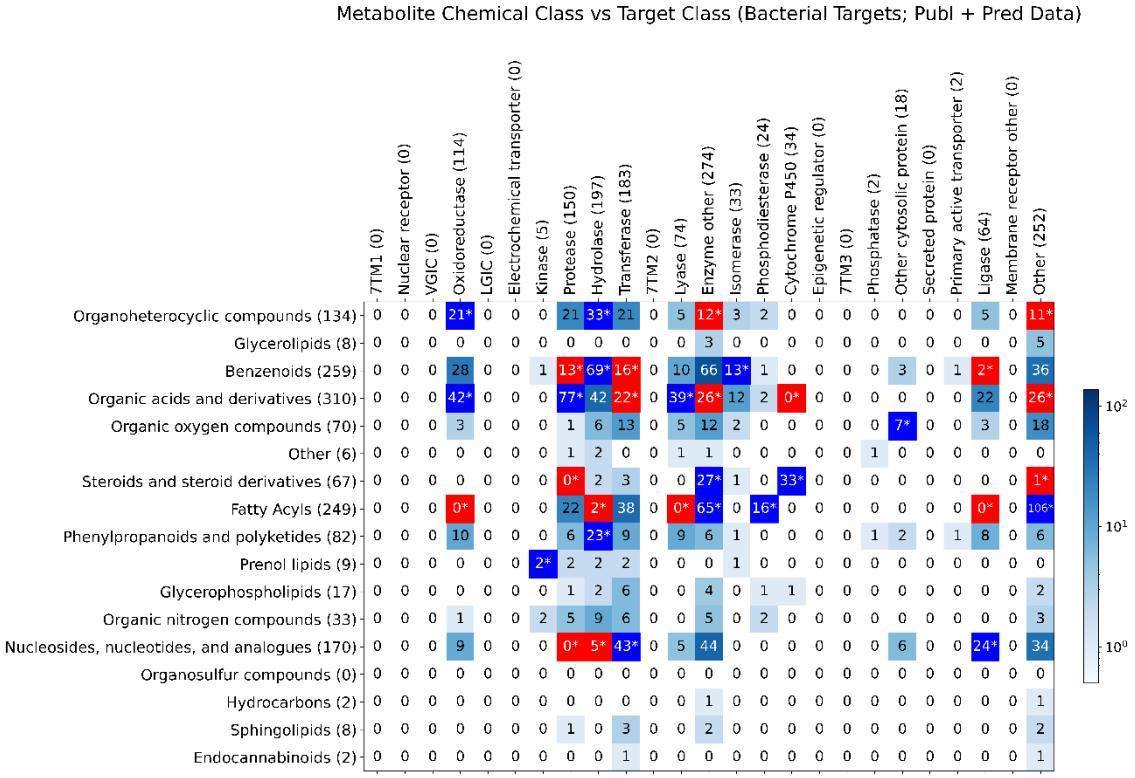

Figure 11. *Distribution of interactions among compound and target classes for gut metabolites and bacterial proteins present in published or predicted data. Counts correspond to unique InChi + UniProt accession combinations. Cells in bright blue or bright red with starred counts correspond to statistically significant combinations (blue overrepresented and blue underrepresented, respectively) in a post-hoc test after Benjamini-Hochberg p-value correction. Numbers within parenthesis after each label correspond to marginal counts.*

The full set of published interactions is available in Table S3 in the Supporting Information.

Experimental validation of predicted interactions

The provided predictions comprise a large number (7572) of hypotheses for fast focused testing. By sharing them openly in this work we expect to accelerate the full identification of the target space of gut microbial metabolites. In order to have an idea

of the experimental validity of these predictions, we took the advantage of the fact that SEA predictions are based on a ChEMBL release, namely 27, previous to the release we have used to identify published interactions (ChEMBL 33); in addition, for the published interactions we have also included BindingDB data, not present in ChEMBL and therefore not used by SEA to base its predictions. The same happens with the literature data. Therefore, we sought for SEA predicted interactions that were also present in our dataset of published interactions. As a matter of fact, a total of 613 interactions were shared between the published and predicted datasets. However, it is possible that some predicted interactions are using data already present in the SEA "training" dataset (and therefore, also present in ChEMBL27). These correspond to the entries with maximum Tanimoto coefficients of 1, meaning that the predicted interaction is already present in the SEA "training" set (a prediction of an interaction is being made of an based on a set of published interactions, of which one is with exactly the predicted compound). Therefore, *bona fide* predictions would correspond to those with maximum Tanimoto coefficient < 1, meaning that the prediction was based on non-identical compounds interacting with the target. In this way, we found a total of 127 true predictions (maximum Tanimoto coefficient < 1) that were later confirmed in our ChEMBL33/BindingDB/Literature set of published interactions. Of them, 115 were interactions with human targets, and 12 with bacterial proteins.

We can analyze the Tanimoto coefficients of these 127 compounds to see how "trivial" were the predictions, that is, if the predictions were based on very close analogs of the compound or, on the other hand, they were based on compounds rather dissimilar. Figure 12 shows the distribution of Tanimoto coefficients across these 127 confirmed

predictions, for both human (blue histogram) and bacterial (orange histogram) proteins.

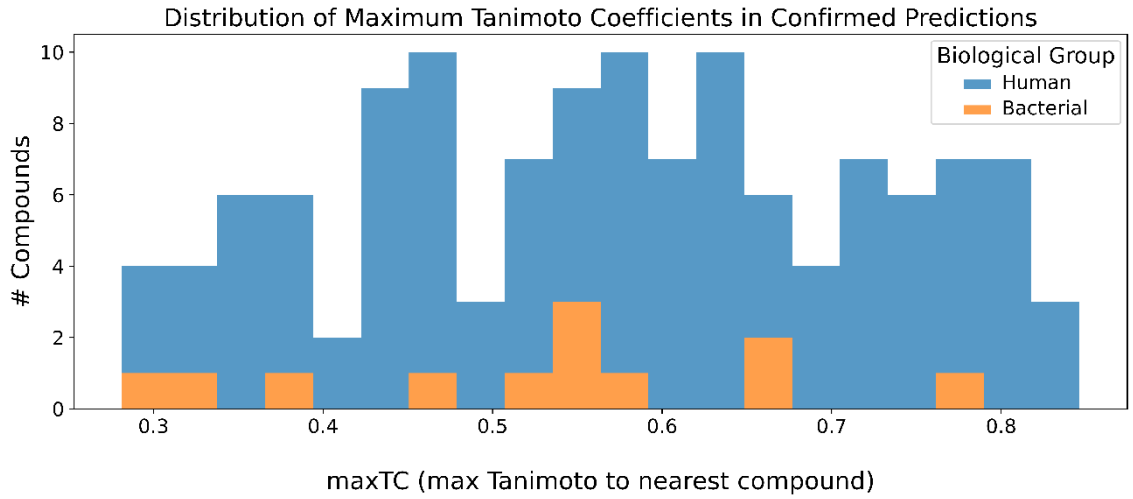

Figure 12. *Distribution of maximum Tanimoto coefficients in SEA predictions confirmed in later published interactions for both human (blue) and bacterial (orange) targets.*

It is possible to see that the distributions of Tanimoto coefficients for both human and bacterial proteins are rather uniform, and going as low as 0.28. This indicates that the experimentally confirmed predictions are not necessarily based on close analogs, but on the contrary they are based in many times on rather dissimilar parts of the chemical space. The fact that SEA uses ensembles of compounds instead of the nearest analog in the training set, makes it more capable of successful extrapolation to non-trivial predictions of bioactivities, and provides further interest to the set of hypotheses shared in this work.

Target analysis and prioritization

In order to gain more information about the targets identified as interacting with the gut microbial metabolites, we used the PHAROS tool[51] for the case of human proteins. PHAROS is a user interface to the so-called Knowledge Management Center of the Illuminating the Druggable Genome (IDG) initiative, which aims to explore the properties and biological functions of the whole potentially druggable human proteome. In this way, a total of 1019 proteins (out of the 1047 human proteins above described) were found with annotations in this system. PHAROS classifies proteins in four groups as regarding their drug development level: “Tclin”, which are proteins with at least one approved drug, and therefore well validated as therapeutic targets; “Tchem”, which are proteins with at least one known modulating small molecule; “Tbio”, proteins characterized to some extent biologically but with not known modulating small molecule; and “Tdark”, which are proteins about which little knowledge is available, in terms of biological function, structure, and involvement in disease. In our list of 1019 human proteins, 60% are in the “Tchem” group, 31% in the “Tclin”, and 9% in the “Tbio” groups. Figure 13 displays the distribution of these three target development groups across the different target classes.

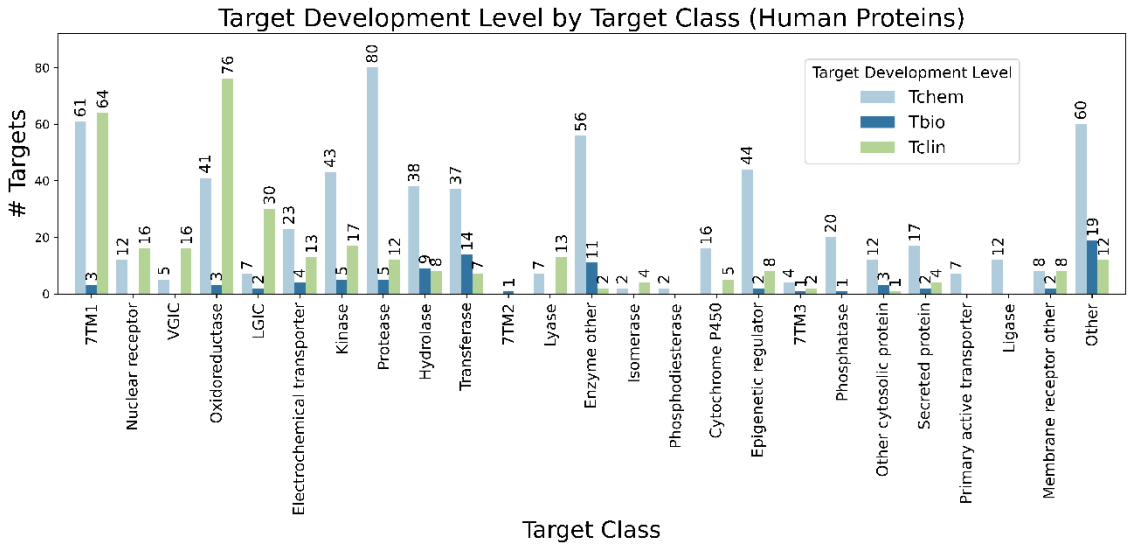

Figure 13. *Distribution of target development levels in PHAROS for the human proteins interacting with microbial metabolites across the different target classes.*

It is possible to see that, within the "Tclin" group, "7TM1" and "Oxidoreductase" are the most abundant ones. As regarding the "Tchem" group, the classes with the largest numbers of proteins are "Protease", ""7TM1", "Other", and "Enzyme other". Finally, the "Tbio" group has "Other", "Transferase", and "Enzyme other" as the most populated target classes.

In order to obtain a "high priority" subset of targets in the gut where to focus experimentation first, we searched for the anatomically defined group of ailments called "intestinal diseases", considering that, while gut metabolites can cross the gut wall and are known to interact with proteins located in other organs, the vast majority of interactions of these metabolites are expected to occur with intestinal targets. In turn, the subset of proteins known to be associated with intestinal diseases is expected to be involved in signaling pathways and interactions in the gut and therefore to be of high biological relevance. In this way, a total of 454 of the metabolite-interacting targets (~43% of the total human proteins) are associated to at least one of these diseases. Interestingly, gut metabolites display a highly significant trend to interact with targets associated with intestinal diseases (p-value = 2e-33 in Fisher exact test, by considering 20412 total proteins and 5587 intestinal disease associated proteins). Also, 57 of the 191 intestinal diseases are associated to a gut-metabolite interacting protein. The associated diseases include ulcerative colitis, Crohn disease, multiple benign and malign tumors, Lynch and Hirschsprung disease, and several additional inflammatory and metabolic diseases. As a matter of fact, the associated disease most significantly

overrepresented in this subset of 454 targets in a Fisher exact test, among all set of all associated diseases in PHAROS, is “ulcerative colitis” (Benjamini-Hochberg adjusted p-value = 2e-78), followed by “malignant colon neoplasm” (adjusted p-value = 6e-54) , “cancer” (adjusted p-value = 9e-47), and “colorectal cancer” (adjusted p-value = 2e-35). By including some additional targets recently identified to be of biological relevance in the microbial-host communication, although yet to be associated with intestinal diseases,[1,15,16,18,19,26,27,60,61] this results in a set of 467 “high-priority” proteins where to focus experimental efforts. These interact with a total of 392 metabolites.

Figure 14 displays the distribution of the 392 metabolites across chemical classes, as well as the 467 targets among target classes.

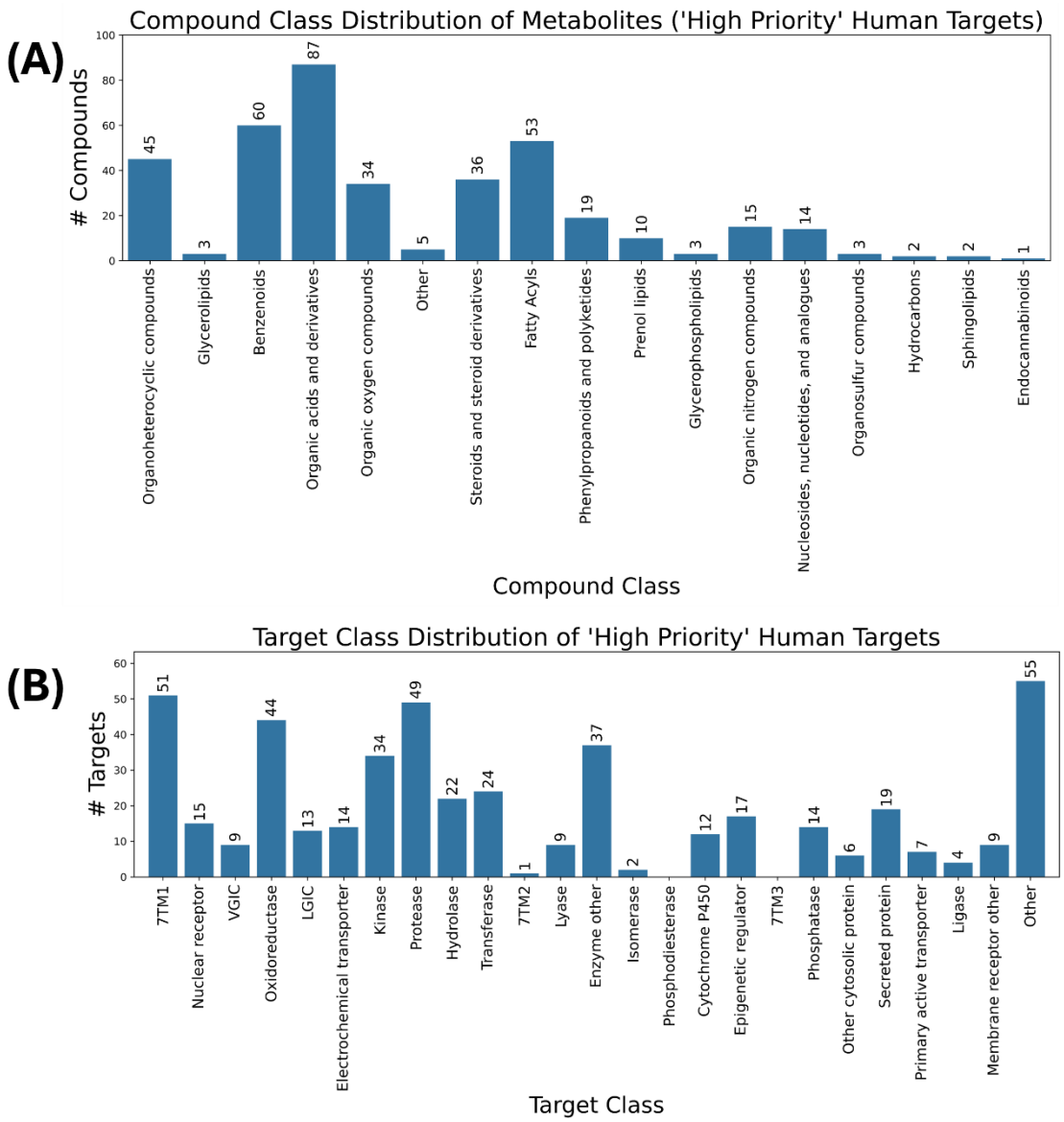

Figure 14. *(A) Distribution of chemical classes (based on the ClassyFire taxonomy) of the 392 gut metabolites interacting with "high priority" human targets. (B) Distribution of the 467 "high priority" targets among target classes, from both published and predicted interactions.*

The largest share of metabolites in this "high priority" set correspond to "Organic acids and derivatives", followed by "Benzenoids", "Fatty Acyls", and "Organoheterocyclic compounds". In turn, the target classes most abundant in this set are "Other", "7TM1", "Protease", "Oxidoreductase", and "Enzyme other".

Figure 15 displays the distribution of interactions across both chemical and target classes for the “high priority” human targets.

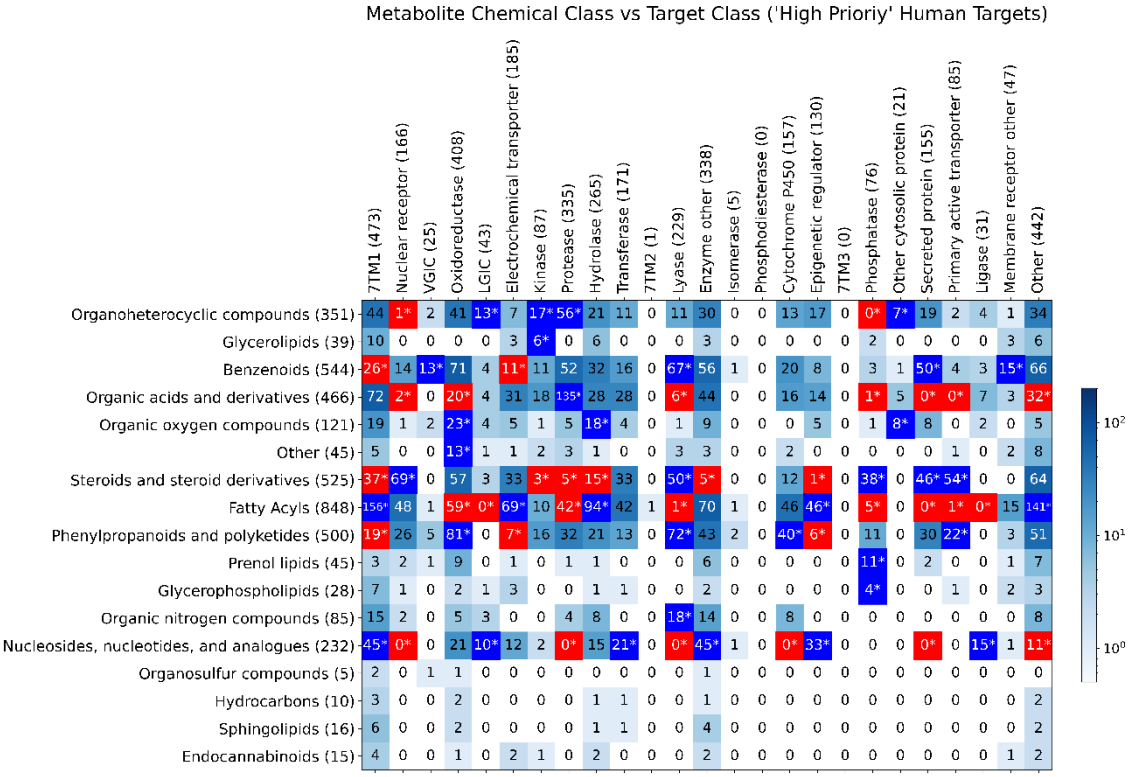

Figure 15. *Distribution of unique interactions among compound and target classes for gut metabolites and “high priority” human targets, present in published or predicted data. Counts correspond to unique InChi + UniProt accession combinations. Cells in bright blue or bright red with starred counts correspond to statistically significant combinations (blue overrepresented and blue underrepresented, respectively) in a post-hoc test after Benjamini-Hochberg p-value correction. Numbers within parenthesis after each label correspond to marginal counts.*

A total of 2804 predicted interactions are present in this “high priority” set, thus providing a large number of hypothesis to test through fast, focused in vitro assays.

The predicted interactions span 22 target classes, with “Other”, “7TM1”, and “Protease” as the most populated ones. Also they involve 16 chemical classes, where “Fatty Acyls”, “Steroids and steroid derivatives”, and “Benzenoids” are the most frequent among the interactions. All these interactions are provided as Supplementary Information.

As regarding the bacterial proteins, since they are not contemplated in PHAROS, the Geptop 2.0[52] method was used to predict their essentiality as a way of prioritization. Geptop 2.0 is a method that identifies essential genes in bacteria, by comparing orthology and phylogeny of the query protein with essential gene datasets determined experimentally (from both DEG[62] and OGEE[63] databases of essential genes). In this way, a total of 63 proteins were assigned the essential class and therefore where deemed of “high priority” within the set of 180 bacterial targets. These “high priority” proteins interact with a total of 212 metabolites. Figure 16 displays the chemical class distribution of the 212 metabolites, plus the target class distribution of the 63 “high priority” bacterial proteins. The largest chemical class are “Organic acids and derivatives”, including mainly organic acids, amino acids, and dipeptides, while the largest class is “Transferase”, that collects transferases of a wide range of chemical groups, including phosphate, methyl, nucleotidyl, isoprenyl, etc.

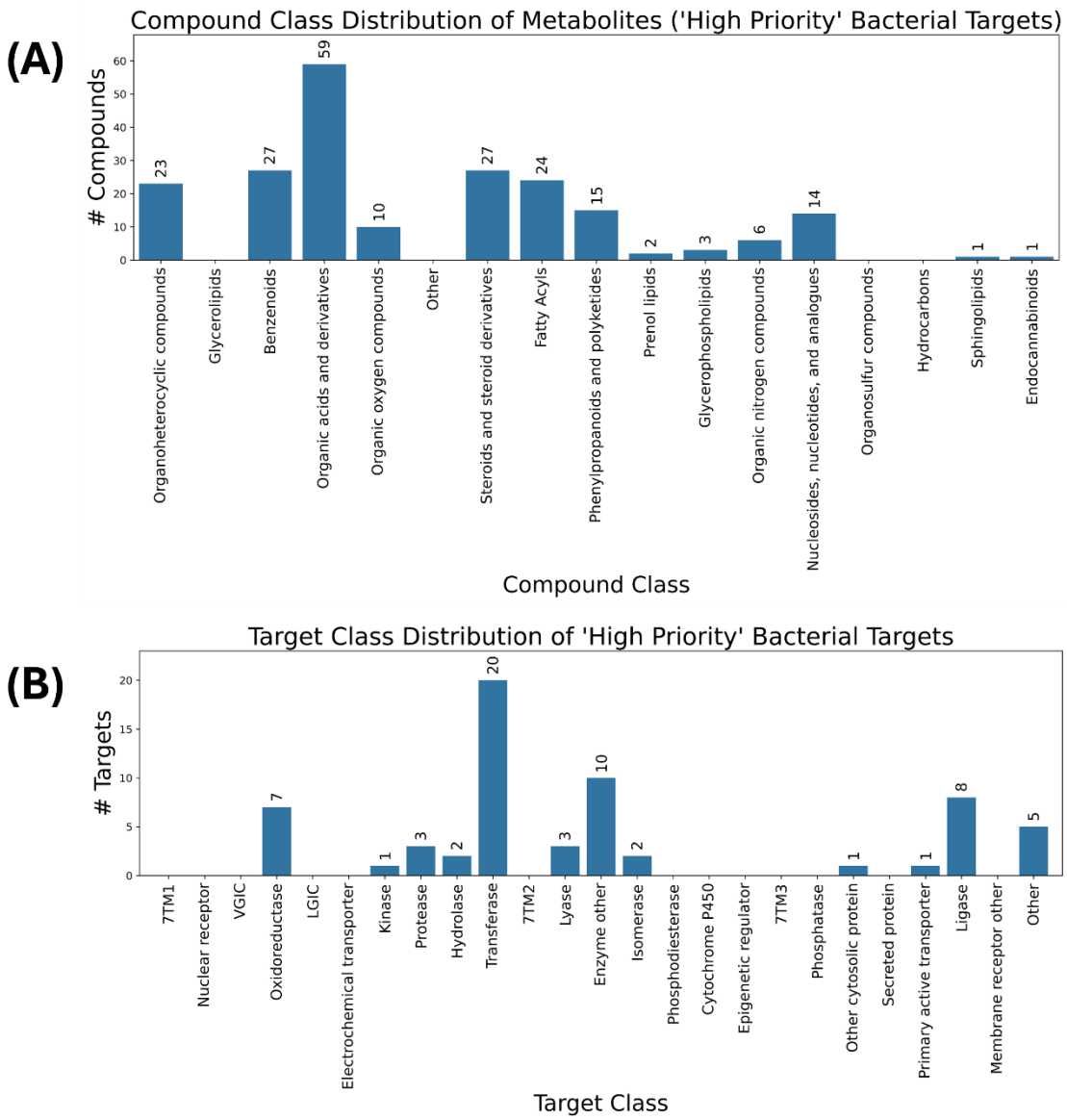

Figure 16. *(A) Distribution of chemical classes (based on the ClassyFire taxonomy) of the 212 gut metabolites interacting with "high priority" bacterial targets. (B) Distribution of the 63 "high priority" targets among target classes, from both published and predicted interactions.*

Figure 17 collects the counts of interactions by both chemical and target class. A total of 442 metabolite-high priority bacterial protein are observed, of which 374 are SEA predictions. This provides a lot of opportunities for fast focused test of putative

interactions of gut microbial metabolites with essential bacterial proteins of the metagenome.

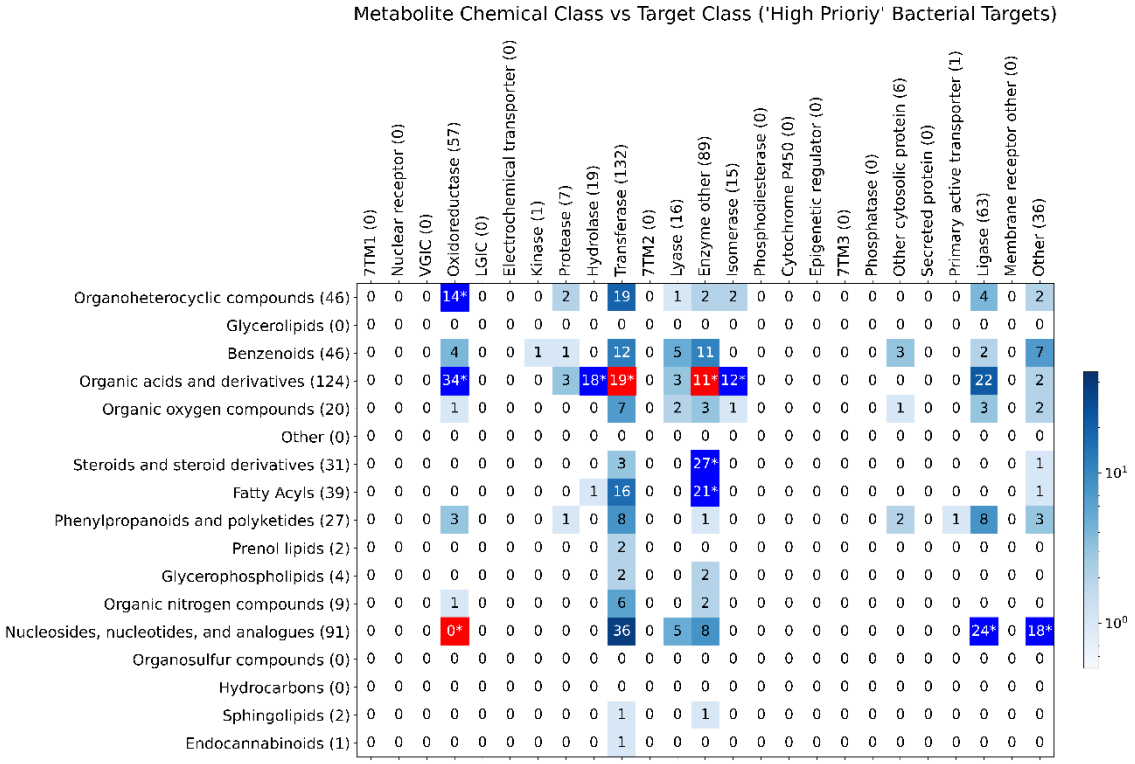

Figure 17. *Distribution of unique interactions among compound and target classes for gut metabolites and "high priority" bacteirla targets, present in published or predicted data. Counts correspond to unique InChi + UniProt accession combinations. Cells in bright blue or bright red with starred counts correspond to statistically significant combinations (blue overrepresented and blue underrepresented, respectively) in a post-hoc test after Benjamini-Hochberg p-value correction. Numbers within parenthesis after each label correspond to marginal counts.*

The set of both human and microbial “high priority” proteins are annotated in Table S3 of Supporting Information.

DISCUSSION

Human gut microbial metabolites are currently the subject of much interest. In a previous work, we have analyzed their structural, physicochemical, and distribution properties, and developed QSAR predictive models for their intestinal permanence.[43] Clarifying the full target space of these molecules could shed light to understand fundamental biological processes that result from the cross-signaling and regulatory pathways established between the microbiome and the host.[1,1,2,6,8,15,16] In addition to the improvement in our comprehension of the role of the microbiome in human health, the knowledge gained can be used in the generation of a new drug and nutraceutic modality of "metabolite-mimetic" molecules.[18,26,31,32,32,33,36] This new modality could benefit from new chemotypes, but additionally, from new targets as they could be directed to proteins encoded by the vast microbial metagenome. It is currently estimated that a total of ~900 proteins (~700 human and ~200 pathogen) are targeted by FDA-approved drugs, while ~3000 are considered druggable out of the 21000 genes in the human genome.[37,64] In turn, the human gut microbiome has been estimated to contain between 250- to 1000-fold as many genes.[65] As a matter of fact, several drug discovery programs in this area are focused on bacterial proteins as pathogenicity factors, like the *E. coli* fimbrial adhesin (FimH) for IBD, and the archaeal F420-dependent Mtd, a methane-producing enzyme for the same disease.[37] This is the reason why, in our analysis, we have not focused on just human proteins interacted with bacterial metabolites, as most of the literature does, but we have also considered all possible host-microbe bidirectional communications by including bacterial proteins of typical gut microbiome genera in our dataset. By the same token, we have also taken into account metabolites from the host, xenobiotics, and metabolites from

bacteria but modified by host enzymes, or from the host/xenobiotics but modified by bacteria. The later could be useful for instance for drug discovery programs that aim to modulate drug metabolism by the microbiota to enhance already existing drugs,[66] and again we find some examples in the development and use of inhibitors of microbial β-glucuronidases and tyrosine decarboxylase to improve the effects of irinotecan and L-dopa, respectively.[37,67]

Our analysis of published interactions has identified 2193 interactions, involving 405 and 128 metabolites, that bind to 451 human and 56 microbial proteins, respectively. In the case of human proteins, most of the interactions (~75%) involve only five chemical classes: "Phenylpropanoids and polyketides", "Benzenoids", "Organoheterocyclic compounds", "Organic acids and derivatives", and "Fatty Acyls"; while in terms of target classes the distribution is more uniform, with "7TM1", "Oxidoreductase", "Lyase", "Other", and "Cytochrome P450" comprising only ~52% of the interactions. For the case of microbial targets, ~77% of the interactions are concentrated in the same previous five chemical classes, but now "Fatty Acyls" is replaced by "Organic nitrogen compounds". However, in terms of target classes the distribution is quite nonuniform in this case, with ~72% of the interactions that involve just "Transferase", "Hydrolase", and "Protease".

The inclusion of SEA predictions expanded these interactions to 9711 ones (> 4-fold increase), now comprising 422 and 330 compounds, and for 1047 human and 180 microbial proteins, respectively. This expansion resulted in the generation of new interaction hypothesis for 125 chemical class vs target class combinations with no previously published interactions. Within these combinations, there a total of 1150

new hypothesis for fast, focused tests in uncharted chemobiological spaces. It is worth noting that most of the research in gut microbial metabolites is being focused in GPCRs and nuclear receptors, but here we provide lots of hypothesis for alternative, underexplored target classes like “Enzyme other”, “LGIC”, “Ligase”, “Transferase”, etc., in combination with well described or underexplored chemical classes (“Organic acids and derivatives”, “Nucleosides, nucleotides, and analogues”, “Benzenoids”, etc.) In addition, for other 116 combinations the interaction counts have increased, corresponding to 6422 additional new hypotheses for chemical class vs target class combinations with at least one published interaction.

The fact that these interactions have been experimentally validated by retrospective analysis (127 of the non-identical predicted interactions were identified in a later ChEMBL release, BindingDB, or the literature) gives confidence in the reliability of the predictions. Moreover, we have seen that within the confirmed predictions the maximum Tanimoto coefficient with compounds in the SEA “training” set were as low as 0.28, demonstrating the capability of this method, that works through compound ensembles, to extrapolate to rather distant parts of the chemical space and find novel chemical diversity.

The use of bioinformatic tools have also allowed to prioritize the targets to pursue in the short term, by spotting those in the host most associated with intestinal diseases on one hand, and in bacteria those that are essential on the other. This, in addition to the filled chemobiological gaps above described, provides additional criteria and guide for the prioritization of experimental tests.

To sum up, we can conclude that the current work provides a comprehensive analysis of our current knowledge about the chemobiological space of gut microbial metabolites, and in addition openly provides a large set of novel validated predictions that will allow to fill in an efficient way many uncharted regions in this space. All the results of this analysis (published interactions, predicted interactions, target prioritizations) are provided in Table S3 in the Supporting Information. It will be useful to complete our knowledge about the mode of action of these highly important molecules in human health.

AUTHOR INFORMATION

Corresponding Author:

Gonzalo Colmenarejo - Biostatistics and Bioinformatics Unit. IMDEA Food, CEI UAM+CSIC, E28049 Madrid, Spain. orcid.org/0000-0002-8249-4547. gonzalo.colmenarejo@imdea.org

Authors:

Cristián Orgaz, Andrés Sánchez-Ruiz - Biostatistics and Bioinformatics Unit. IMDEA Food, CEI UAM+CSIC, E28049 Madrid, Spain

Notes

The authors declare no competing financial interest.

DATA AND SOFTWARE AVAILABILITY STATEMENT

The results of all the analyses are collected in file SI.xlsx as Supporting Information. In that file, the following Tables can be found:
Table S1: set of microbial genera typical in human microbial metagenomics analyses and used in this work.

Table S2: distribution by target classes of target sharing between gut microbial metabolites and drugs (small molecule-, oral-, and with systemic action-type).

Table S3: set of published and predicted metabolite-target interactions. For each interaction, the following data is provided: hmdb identifier (“hmdb_id”), inchi string, chemical class (“chem_cl”), compound set (“cset”: Metabolites vs Drugs); specific

compound set (“comp_set”: Drugs vs GutFL vs GutnoFL vs Gut/Serum); uniprot accession number of the target (“uniport_id”); target name (“tar_name”); target class (“tar_cl”); target biological group (“tar_biolgr”: “b” for bacterial, “h” for human); biological species (“organism”); source of data (“src”); pchembl-like affinity data (“pbind”); maximum Tanimoto coefficient for SEA prediction (“maxTc”); name of compound (“comp_name”); aggregated source of data (“src2”); even more aggregated source of data (“src3”); high priority target (“hpr”: empty vs “hum” for high-priority human vs “bac” for high-priority bacterial).

REFERENCES

(1) Rahman, S.; O'Connor, A. L.; Becker, S. L.; Patel, R. K.; Martindale, R. G.; Tsikitis, V. L. Gut Microbial Metabolites and Its Impact on Human Health. *Ann Gastroenterol* **2023**, *36* (4), 360–368. https://doi.org/10.20524/aog.2023.0809.

(2) Garrett, W. S. Immune Recognition of Microbial Metabolites. *Nat Rev Immunol* **2020**, *20* (2), 91–92. https://doi.org/10.1038/s41577-019-0252-2.

(3) Ghosh, S.; Whitley, C. S.; Haribabu, B.; Jala, V. R. Regulation of Intestinal Barrier Function by Microbial Metabolites. *Cellular and Molecular Gastroenterology and Hepatology* **2021**, *11* (5), 1463–1482. https://doi.org/10.1016/j.jcmgh.2021.02.007.

(4) Harris, V. C.; Haak, B. W.; Boele Van Hensbroek, M.; Wiersinga, W. J. The Intestinal Microbiome in Infectious Diseases: The Clinical Relevance of a Rapidly Emerging Field. *Open Forum Infectious Diseases* **2017**, *4* (3), ofx144. https://doi.org/10.1093/ofid/ofx144.

(5) Hou, K.; Wu, Z.-X.; Chen, X.-Y.; Wang, J.-Q.; Zhang, D.; Xiao, C.; Zhu, D.; Koya, J. B.; Wei, L.; Li, J.; Chen, Z.-S. Microbiota in Health and Diseases. *Sig Transduct Target Ther* **2022**, *7* (1), 1–28. https://doi.org/10.1038/s41392-022-00974-4.

(6) Henke, M. T.; Clardy, J. Molecular Messages in Human Microbiota. *Science* **2019**, *366* (6471), 1309–1310. https://doi.org/10.1126/science.aaz4164.

(7) Agirman, G.; Hsiao, E. Y. SnapShot: The Microbiota-Gut-Brain Axis. *Cell* **2021**, *184* (9), 2524-2524.e1. https://doi.org/10.1016/j.cell.2021.03.022.

(8) Agus, A.; Clément, K.; Sokol, H. Gut Microbiota-Derived Metabolites as Central Regulators in Metabolic Disorders. *Gut* **2021**, *70* (6), 1174–1182. https://doi.org/10.1136/gutjnl-2020-323071.

(9) Gilbert, J. A.; Blaser, M. J.; Caporaso, J. G.; Jansson, J. K.; Lynch, S. V.; Knight, R. Current Understanding of the Human Microbiome. *Nat Med* **2018**, *24* (4), 392–400. https://doi.org/10.1038/nm.4517.

(10) Delzenne, N. M.; Neyrinck, A. M.; Bäckhed, F.; Cani, P. D. Targeting Gut Microbiota in Obesity: Effects of Prebiotics and Probiotics. *Nat Rev Endocrinol* **2011**, *7* (11), 639–646. https://doi.org/10.1038/nrendo.2011.126.

(11) Ferreiro, A. L.; Choi, J.; Ryou, J.; Newcomer, E. P.; Thompson, R.; Bollinger, R. M.; Hall-Moore, C.; Ndao, I. M.; Sax, L.; Benzinger, T. L. S.; Stark, S. L.; Holtzman, D. M.; Fagan, A. M.; Schindler, S. E.; Cruchaga, C.; Butt, O. H.; Morris, J. C.; Tarr, P. I.; Ances, B. M.; Dantas, G. Gut Microbiome Composition May Be an Indicator of Preclinical Alzheimer's Disease. *Science Translational Medicine* **2023**, *15* (700), eabo2984. https://doi.org/10.1126/scitranslmed.abo2984.

(12) Fobofou, S. A.; Savidge, T. Microbial Metabolites: Cause or Consequence in Gastrointestinal Disease? *American Journal of Physiology-Gastrointestinal and Liver Physiology* **2022**, *322* (6), G535–G552. https://doi.org/10.1152/ajpgi.00008.2022.

(13) Fan, Y.; Pedersen, O. Gut Microbiota in Human Metabolic Health and Disease. *Nat Rev Microbiol* **2021**, *19* (1), 55–71. https://doi.org/10.1038/s41579-020-0433-9.

(14) Sender, R.; Fuchs, S.; Milo, R. Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. *Cell* **2016**, *164* (3), 337–340. https://doi.org/10.1016/j.cell.2016.01.013.

(15) Zheng, X.; Cai, X.; Hao, H. Emerging Targetome and Signalome Landscape of Gut Microbial Metabolites. *Cell Metabolism* **2022**, *34* (1), 35–58. https://doi.org/10.1016/j.cmet.2021.12.011.

(16) Morozumi, S.; Ueda, M.; Okahashi, N.; Arita, M. Structures and Functions of the Gut Microbial Lipidome. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **2022**, *1867* (3), 159110. https://doi.org/10.1016/j.bbalip.2021.159110.

(17) Colosimo, D. A.; Kohn, J. A.; Luo, P. M.; Piscotta, F. J.; Han, S. M.; Pickard, A. J.; Rao, A.; Cross, J. R.; Cohen, L. J.; Brady, S. F. Mapping Interactions of Microbial Metabolites with

Human G-Protein-Coupled Receptors. *Cell Host & Microbe* **2019**, *26* (2), 273-282.e7. https://doi.org/10.1016/j.chom.2019.07.002.

(18) Cohen, L. J.; Esterhazy, D.; Kim, S.-H.; Lemetre, C.; Aguilar, R. R.; Gordon, E. A.; Pickard, A. J.; Cross, J. R.; Emiliano, A. B.; Han, S. M.; Chu, J.; Vila-Farres, X.; Kaplitt, J.; Rogoz, A.; Calle, P. Y.; Hunter, C.; Bitok, J. K.; Brady, S. F. Commensal Bacteria Make GPCR Ligands That Mimic Human Signalling Molecules. *Nature* **2017**, *549* (7670), 48–53. https://doi.org/10.1038/nature23874.

(19) Ay, Ü.; Leníček, M.; Classen, A.; Olde Damink, S. W. M.; Bolm, C.; Schaap, F. G. New Kids on the Block: Bile Salt Conjugates of Microbial Origin. *Metabolites* **2022**, *12* (2), 176. https://doi.org/10.3390/metabo12020176.

(20) Aleti, G.; Troyer, E. A.; Hong, S. G Protein-Coupled Receptors: A Target for Microbial Metabolites and a Mechanistic Link to Microbiome-Immune-Brain Interactions. *Brain, Behavior, & Immunity - Health* **2023**, *32*, 100671. https://doi.org/10.1016/j.bbih.2023.100671.

(21) Mager, L. F.; Burkhard, R.; Pett, N.; Cooke, N. C. A.; Brown, K.; Ramay, H.; Paik, S.; Stagg, J.; Groves, R. A.; Gallo, M.; Lewis, I. A.; Geuking, M. B.; McCoy, K. D. Microbiome-Derived Inosine Modulates Response to Checkpoint Inhibitor Immunotherapy. *Science* **2020**, *369* (6510), 1481–1489. https://doi.org/10.1126/science.abc3421.

(22) Duscha, A.; Gisevius, B.; Hirschberg, S.; Yissachar, N.; Stangl, G. I.; Dawin, E.; Bader, V.; Haase, S.; Kaisler, J.; David, C.; Schneider, R.; Troisi, R.; Zent, D.; Hegelmaier, T.; Dokalis, N.; Gerstein, S.; Del Mare-Roumani, S.; Amidror, S.; Staszewski, O.; Poschmann, G.; Stühler, K.; Hirche, F.; Balogh, A.; Kempa, S.; Träger, P.; Zaiss, M. M.; Holm, J. B.; Massa, M. G.; Nielsen, H. B.; Faissner, A.; Lukas, C.; Gatermann, S. G.; Scholz, M.; Przuntek, H.; Prinz, M.; Forslund, S. K.; Winklhofer, K. F.; Müller, D. N.; Linker, R. A.; Gold, R.; Haghikia, A. Propionic Acid Shapes the Multiple Sclerosis Disease Course by an Immunomodulatory Mechanism. *Cell* **2020**, *180* (6), 1067-1080.e16. https://doi.org/10.1016/j.cell.2020.02.035.

(23) Nemet, I.; Saha, P. P.; Gupta, N.; Zhu, W.; Romano, K. A.; Skye, S. M.; Cajka, T.; Mohan, M. L.; Li, L.; Wu, Y.; Funabashi, M.; Ramer-Tait, A. E.; Naga Prasad, S. V.; Fiehn, O.; Rey, F. E.; Tang, W. H. W.; Fischbach, M. A.; DiDonato, J. A.; Hazen, S. L. A Cardiovascular Disease-Linked Gut Microbial Metabolite Acts via Adrenergic Receptors. *Cell* **2020**, *180* (5), 862-877.e22. https://doi.org/10.1016/j.cell.2020.02.016.

(24) Husted, A. S.; Trauelsen, M.; Rudenko, O.; Hjorth, S. A.; Schwartz, T. W. GPCR-Mediated Signaling of Metabolites. *Cell Metabolism* **2017**, *25* (4), 777–796. https://doi.org/10.1016/j.cmet.2017.03.008.

(25) Tierney, B. T.; Yang, Z.; Luber, J. M.; Beaudin, M.; Wibowo, M. C.; Baek, C.; Mehlenbacher, E.; Patel, C. J.; Kostic, A. D. The Landscape of Genetic Content in the Gut and Oral Human Microbiome. *Cell Host & Microbe* **2019**, *26* (2), 283-295.e8. https://doi.org/10.1016/j.chom.2019.07.008.

(26) Dvořák, Z.; Li, H.; Mani, S. Microbial Metabolites as Ligands to Xenobiotic Receptors: Chemical Mimicry as Potential Drugs of the Future. *Drug Metab Dispos* **2023**, *51* (2), 219–227. https://doi.org/10.1124/dmd.122.000860.

(27) Guan, X.-J.; Zhang, Y.-Y.; Zheng, X.; Hao, H.-P. Drug Discovery Inspired from Nuclear Receptor Sensing of Microbial Signals. *Trends in Molecular Medicine* **2021**, *27* (7), 624–626. https://doi.org/10.1016/j.molmed.2021.03.007.

(28) Dvořák, Z.; Sokol, H.; Mani, S. Drug Mimicry: Promiscuous Receptors PXR and AhR, and Microbial Metabolite Interactions in the Intestine. *Trends in Pharmacological Sciences* **2020**, *41* (12), 900–908. https://doi.org/10.1016/j.tips.2020.09.013.

(29) Chavira, A.; Belda-Ferre, P.; Kosciolek, T.; Ali, F.; Dorrestein, P. C.; Knight, R. The Microbiome and Its Potential for Pharmacology. In *Concepts and Principles of Pharmacology: 100 Years of the Handbook of Experimental Pharmacology*; Barrett, J. E., Page, C. P., Michel, M. C., Eds.; Handbook of Experimental Pharmacology; Springer

International Publishing: Cham, 2019; pp 301–326. https://doi.org/10.1007/164_2019_317.
(30) Descamps, H. C.; Herrmann, B.; Wiredu, D.; Thaiss, C. A. The Path toward Using Microbial Metabolites as Therapies. *eBioMedicine* **2019**, *44*, 747–754. https://doi.org/10.1016/j.ebiom.2019.05.063.
(31) Dvořák, Z.; Kopp, F.; Costello, C. M.; Kemp, J. S.; Li, H.; Vrzalová, A.; Štěpánková, M.; Bartoňková, I.; Jiskrová, E.; Poulíková, K.; Vyhlídalová, B.; Nordstroem, L. U.; Karunaratne, C. V.; Ranhotra, H. S.; Mun, K. S.; Naren, A. P.; Murray, I. A.; Perdew, G. H.; Brtko, J.; Toporova, L.; Schön, A.; Wallace, B. D.; Walton, W. G.; Redinbo, M. R.; Sun, K.; Beck, A.; Kortagere, S.; Neary, M. C.; Chandran, A.; Vishveshwara, S.; Cavalluzzi, M. M.; Lentini, G.; Cui, J. Y.; Gu, H.; March, J. C.; Chatterjee, S.; Matson, A.; Wright, D.; Flannigan, K. L.; Hirota, S. A.; Sartor, R. B.; Mani, S. Targeting the Pregnane X Receptor Using Microbial Metabolite Mimicry. *EMBO Molecular Medicine* **2020**, *12* (4), e11621. https://doi.org/10.15252/emmm.201911621.
(32) Nuzzo, A.; Brown, J. R. Microbiome Metabolite Mimics Accelerate Drug Discovery. *Trends in Molecular Medicine* **2020**, *26* (5), 435–437. https://doi.org/10.1016/j.molmed.2020.03.006.
(33) Saha, S.; Rajpal, D. K.; Brown, J. R. Human Microbial Metabolites as a Source of New Drugs. *Drug Discovery Today* **2016**, *21* (4), 692–698. https://doi.org/10.1016/j.drudis.2016.02.009.
(34) Nuzzo, A.; Saha, S.; Berg, E.; Jayawickreme, C.; Tocker, J.; Brown, J. R. Expanding the Drug Discovery Space with Predicted Metabolite–Target Interactions. *Communications Biology* **2021**, *4* (1), 1–11. https://doi.org/10.1038/s42003-021-01822-x.
(35) Roberts, A. B.; Gu, X.; Buffa, J. A.; Hurd, A. G.; Wang, Z.; Zhu, W.; Gupta, N.; Skye, S. M.; Cody, D. B.; Levison, B. S.; Barrington, W. T.; Russell, M. W.; Reed, J. M.; Duzan, A.; Lang, J. M.; Fu, X.; Li, L.; Myers, A. J.; Rachakonda, S.; DiDonato, J. A.; Brown, J. M.; Gogonea, V.; Lusis, A. J.; Garcia-Garcia, J. C.; Hazen, S. L. Development of a Gut Microbe–Targeted Nonlethal Therapeutic to Inhibit Thrombosis Potential. *Nat Med* **2018**, *24* (9), 1407–1417. https://doi.org/10.1038/s41591-018-0128-1.
(36) Gulati, A.; Guo, X. LEVERAGING SMALL MOLECULES TO MODULATE THE MICROBIOME TO TREAT HUMAN DISEASES. In *2022 Medicinal Chemistry Reviews*; Medicinal Chemistry Reviews; MEDI, Inc. Published by American Chemical Society., 2022; Vol. 57, pp 389–414. https://doi.org/10.1021/mc-2022-vol57.ch16.
(37) Cully, M. Microbiome Therapeutics Go Small Molecule. *Nat Rev Drug Discov* **2019**, *18* (8), 569–572. https://doi.org/10.1038/d41573-019-00122-8.
(38) Li, Q.; Cao, L.; Tian, Y.; Zhang, P.; Ding, C.; Lu, W.; Jia, C.; Shao, C.; Liu, W.; Wang, D.; Ye, H.; Hao, H. Butyrate Suppresses the Proliferation of Colorectal Cancer Cells via Targeting Pyruvate Kinase M2 and Metabolic Reprogramming *. *Molecular & Cellular Proteomics* **2018**, *17* (8), 1531–1545. https://doi.org/10.1074/mcp.RA118.000752.
(39) Piazza, I.; Kochanowski, K.; Cappelletti, V.; Fuhrer, T.; Noor, E.; Sauer, U.; Picotti, P. A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication. *Cell* **2018**, *172* (1), 358-372.e23. https://doi.org/10.1016/j.cell.2017.12.006.
(40) Chen, H.; Nwe, P.-K.; Yang, Y.; Rosen, C. E.; Bielecka, A. A.; Kuchroo, M.; Cline, G. W.; Kruse, A. C.; Ring, A. M.; Crawford, J. M.; Palm, N. W. A Forward Chemical Genetic Screen Reveals Gut Microbiota Metabolites That Modulate Host Physiology. *Cell* **2019**, *177* (5), 1217-1231.e18. https://doi.org/10.1016/j.cell.2019.03.036.
(41) Sánchez-Ruiz, A.; Colmenarejo, G. Systematic Analysis and Prediction of the Target Space of Bioactive Food Compounds: Filling the Chemobiological Gaps. *J. Chem. Inf. Model.* **2022**, *62* (16), 3734–3751. https://doi.org/10.1021/acs.jcim.2c00888.
(42) Colmenarejo, G.; Perez-Martinez, L.; Palacios; Ramirez de Molina, A. Identification of METTL3-Mediated RNA Methylation Inhibitors for Cancer Treatment. In *SPETSES Summer School*; Spetses, Greece, 2022.

(43) Gil-Pichardo, A.; Sánchez-Ruiz, A.; Colmenarejo, G. Analysis of Metabolites in Human Gut: Illuminating the Design of Gut-Targeted Drugs. *Journal of Cheminformatics* **2023**, *15* (1), 96. https://doi.org/10.1186/s13321-023-00768-y.

(44) Wishart, D. S.; Guo, A.; Oler, E.; Wang, F.; Anjum, A.; Peters, H.; Dizon, R.; Sayeeda, Z.; Tian, S.; Lee, B. L.; Berjanskii, M.; Mah, R.; Yamamoto, M.; Jovel, J.; Torres-Calzada, C.; Hiebert-Giesbrecht, M.; Lui, V. W.; Varshavi, D.; Varshavi, D.; Allen, D.; Arndt, D.; Khetarpal, N.; Sivakumaran, A.; Harford, K.; Sanford, S.; Yee, K.; Cao, X.; Budinski, Z.; Liigand, J.; Zhang, L.; Zheng, J.; Mandal, R.; Karu, N.; Dambrova, M.; Schiöth, H. B.; Greiner, R.; Gautam, V. HMDB 5.0: The Human Metabolome Database for 2022. *Nucleic Acids Research* **2022**, *50* (D1), D622–D631. https://doi.org/10.1093/nar/gkab1062.

(45) Knox, C.; Wilson, M.; Klinger, C. M.; Franklin, M.; Oler, E.; Wilson, A.; Pon, A.; Cox, J.; Chin, N. E. (Lucy); Strawbridge, S. A.; Garcia-Patino, M.; Kruger, R.; Sivakumaran, A.; Sanford, S.; Doshi, R.; Khetarpal, N.; Fatokun, O.; Doucet, D.; Zubkowski, A.; Rayat, D. Y.; Jackson, H.; Harford, K.; Anjum, A.; Zakir, M.; Wang, F.; Tian, S.; Lee, B.; Liigand, J.; Peters, H.; Wang, R. Q. (Rachel); Nguyen, T.; So, D.; Sharp, M.; da Silva, R.; Gabriel, C.; Scantlebury, J.; Jasinski, M.; Ackerman, D.; Jewison, T.; Sajed, T.; Gautam, V.; Wishart, D. S. DrugBank 6.0: The DrugBank Knowledgebase for 2024. *Nucleic Acids Research* **2024**, *52* (D1), D1265–D1275. https://doi.org/10.1093/nar/gkad976.

(46) Kaya, I.; Colmenarejo, G. Analysis of Nuisance Substructures and Aggregators in a Comprehensive Database of Food Chemical Compounds. *J. Agric. Food Chem.* **2020**, *68* (33), 8812–8824. https://doi.org/10.1021/acs.jafc.0c02521.

(47) Sánchez-Ruiz, A.; Colmenarejo, G. Updated Prediction of Aggregators and Assay-Interfering Substructures in Food Compounds. *J. Agric. Food Chem.* **2021**, *69* (50), 15184–15194. https://doi.org/10.1021/acs.jafc.1c05918.

(48) Zdrazil, B.; Felix, E.; Hunter, F.; Manners, E. J.; Blackshaw, J.; Corbett, S.; de Veij, M.; Ioannidis, H.; Lopez, D. M.; Mosquera, J. F.; Magarinos, M. P.; Bosc, N.; Arcila, R.; Kizilören, T.; Gaulton, A.; Bento, A. P.; Adasme, M. F.; Monecke, P.; Landrum, G. A.; Leach, A. R. The ChEMBL Database in 2023: A Drug Discovery Platform Spanning Multiple Bioactivity Data Types and Time Periods. *Nucleic Acids Research* **2024**, *52* (D1), D1180–D1192. https://doi.org/10.1093/nar/gkad1004.

(49) Gilson, M. K.; Liu, T.; Baitaluk, M.; Nicola, G.; Hwang, L.; Chong, J. BindingDB in 2015: A Public Database for Medicinal Chemistry, Computational Chemistry and Systems Pharmacology. *Nucleic Acids Res* **2016**, *44* (D1), D1045–D1053. https://doi.org/10.1093/nar/gkv1072.

(50) Djoumbou Feunang, Y.; Eisner, R.; Knox, C.; Chepelev, L.; Hastings, J.; Owen, G.; Fahy, E.; Steinbeck, C.; Subramanian, S.; Bolton, E.; Greiner, R.; Wishart, D. S. ClassyFire: Automated Chemical Classification with a Comprehensive, Computable Taxonomy. *Journal of Cheminformatics* **2016**, *8* (1), 61. https://doi.org/10.1186/s13321-016-0174-y.

(51) Kelleher, K. J.; Sheils, T. K.; Mathias, S. L.; Yang, J. J.; Metzger, V. T.; Siramshetty, V. B.; Nguyen, D.-T.; Jensen, L. J.; Vidović, D.; Schürer, S. C.; Holmes, J.; Sharma, K. R.; Pillai, A.; Bologa, C. G.; Edwards, J. S.; Mathé, E. A.; Oprea, T. I. Pharos 2023: An Integrated Resource for the Understudied Human Proteome. *Nucleic Acids Research* **2023**, *51* (D1), D1405–D1416. https://doi.org/10.1093/nar/gkac1033.

(52) Wen, Q.-F.; Wei, W.; Guo, F.-B. Geptop 2.0: Accurately Select Essential Genes from the List of Protein-Coding Genes in Prokaryotic Genomes. In *Essential Genes and Genomes: Methods and Protocols*; Zhang, R., Ed.; Springer US: New York, NY, 2022; pp 423–430. https://doi.org/10.1007/978-1-0716-1720-5_23.

(53) Shan, G.; Gerstenberger, S. Fisher's Exact Approach for Post Hoc Analysis of a Chi-Squared Test. *PLoS One* **2017**, *12* (12), e0188709. https://doi.org/10.1371/journal.pone.0188709.

(54) Bemis, G. W.; Murcko, M. A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39* (15), 2887–2893. https://doi.org/10.1021/jm9602928.
(55) Bemis, G. W.; Murcko, M. A. Properties of Known Drugs. 2. Side Chains. *J. Med. Chem.* **1999**, *42* (25), 5095–5099. https://doi.org/10.1021/jm9903996.
(56) Keiser, M. J.; Roth, B. L.; Armbruster, B. N.; Ernsberger, P.; Irwin, J. J.; Shoichet, B. K. Relating Protein Pharmacology by Ligand Chemistry. *Nat Biotechnol* **2007**, *25* (2), 197–206. https://doi.org/10.1038/nbt1284.
(57) Irwin, J. J.; Gaskins, G.; Sterling, T.; Mysinger, M. M.; Keiser, M. J. Predicted Biological Activity of Purchasable Chemical Space. *J. Chem. Inf. Model.* **2018**, *58* (1), 148–164. https://doi.org/10.1021/acs.jcim.7b00316.
(58) Gregori-Puigjané, E.; Setola, V.; Hert, J.; Crews, B. A.; Irwin, J. J.; Lounkine, E.; Marnett, L.; Roth, B. L.; Shoichet, B. K. Identifying Mechanism-of-Action Targets for Drugs and Probes. *Proceedings of the National Academy of Sciences* **2012**, *109* (28), 11178–11183. https://doi.org/10.1073/pnas.1204524109.
(59) Pottel, J.; Armstrong, D.; Zou, L.; Fekete, A.; Huang, X.-P.; Torosyan, H.; Bednarczyk, D.; Whitebread, S.; Bhhatarai, B.; Liang, G.; Jin, H.; Ghaemi, S. N.; Slocum, S.; Lukacs, K. V.; Irwin, J. J.; Berg, E. L.; Giacomini, K. M.; Roth, B. L.; Shoichet, B. K.; Urban, L. The Activities of Drug Inactive Ingredients on Biological Targets. *Science* **2020**, *369* (6502), 403–413. https://doi.org/10.1126/science.aaz9906.
(60) Cohen, L. J.; Kang, H.-S.; Chu, J.; Huang, Y.-H.; Gordon, E. A.; Reddy, B. V. B.; Ternei, M. A.; Craig, J. W.; Brady, S. F. Functional Metagenomic Discovery of Bacterial Effectors in the Human Microbiome and Isolation of Commendamide, a GPCR G2A/132 Agonist. *Proceedings of the National Academy of Sciences* **2015**, *112* (35), E4825–E4834. https://doi.org/10.1073/pnas.1508737112.
(61) Quinn, R. A.; Melnik, A. V.; Vrbanac, A.; Fu, T.; Patras, K. A.; Christy, M. P.; Bodai, Z.; Belda-Ferre, P.; Tripathi, A.; Chung, L. K.; Downes, M.; Welch, R. D.; Quinn, M.; Humphrey, G.; Panitchpakdi, M.; Weldon, K. C.; Aksenov, A.; da Silva, R.; Avila-Pacheco, J.; Clish, C.; Bae, S.; Mallick, H.; Franzosa, E. A.; Lloyd-Price, J.; Bussell, R.; Thron, T.; Nelson, A. T.; Wang, M.; Leszczynski, E.; Vargas, F.; Gauglitz, J. M.; Meehan, M. J.; Gentry, E.; Arthur, T. D.; Komor, A. C.; Poulsen, O.; Boland, B. S.; Chang, J. T.; Sandborn, W. J.; Lim, M.; Garg, N.; Lumeng, J. C.; Xavier, R. J.; Kazmierczak, B. I.; Jain, R.; Egan, M.; Rhee, K. E.; Ferguson, D.; Raffatellu, M.; Vlamakis, H.; Haddad, G. G.; Siegel, D.; Huttenhower, C.; Mazmanian, S. K.; Evans, R. M.; Nizet, V.; Knight, R.; Dorrestein, P. C. Global Chemical Effects of the Microbiome Include New Bile-Acid Conjugations. *Nature* **2020**, *579* (7797), 123–129. https://doi.org/10.1038/s41586-020-2047-9.
(62) Liang, Y.-T.; Luo, H.; Lin, Y.; Gao, F. Recent Advances in the Characterization of Essential Genes and Development of a Database of Essential Genes. *iMeta* **2024**, *3* (1), e157. https://doi.org/10.1002/imt2.157.
(63) Gurumayum, S.; Jiang, P.; Hao, X.; Campos, T. L.; Young, N. D.; Korhonen, P. K.; Gasser, R. B.; Bork, P.; Zhao, X.-M.; He, L.; Chen, W.-H. OGEE v3: Online GEne Essentiality Database with Increased Coverage of Organisms and Human Cell Lines. *Nucleic Acids Research* **2021**, *49* (D1), D998–D1003. https://doi.org/10.1093/nar/gkaa884.
(64) Santos, R.; Ursu, O.; Gaulton, A.; Bento, A. P.; Donadi, R. S.; Bologa, C. G.; Karlsson, A.; Al-Lazikani, B.; Hersey, A.; Oprea, T. I.; Overington, J. P. A Comprehensive Map of Molecular Drug Targets. *Nat Rev Drug Discov* **2017**, *16* (1), 19–34. https://doi.org/10.1038/nrd.2016.230.
(65) Methé, B. A.; Nelson, K. E.; Pop, M.; Creasy, H. H.; Giglio, M. G.; Huttenhower, C.; Gevers, D.; Petrosino, J. F.; Abubucker, S.; Badger, J. H.; Chinwalla, A. T.; Earl, A. M.; FitzGerald, M. G.; Fulton, R. S.; Hallsworth-Pepin, K.; Lobos, E. A.; Madupu, R.; Magrini, V.; Martin, J. C.; Mitreva, M.; Muzny, D. M.; Sodergren, E. J.; Versalovic, J.; Wollam, A. M.; Worley, K. C.; Wortman, J. R.; Young, S. K.; Zeng, Q.; Aagaard, K. M.; Abolude, O. O.; Allen-Vercoe, E.; Alm, E. J.; Alvarado, L.; Andersen, G. L.; Anderson, S.; Appelbaum, E.;

Arachchi, H. M.; Armitage, G.; Arze, C. A.; Ayvaz, T.; Baker, C. C.; Begg, L.; Belachew, T.; Bhonagiri, V.; Bihan, M.; Blaser, M. J.; Bloom, T.; Bonazzi, V. R.; Brooks, P.; Buck, G. A.; Buhay, C. J.; Busam, D. A.; Campbell, J. L.; Canon, S. R.; Cantarel, B. L.; Chain, P. S.; Chen, I.-M. A.; Chen, L.; Chhibba, S.; Chu, K.; Ciulla, D. M.; Clemente, J. C.; Clifton, S. W.; Conlan, S.; Crabtree, J.; Cutting, M. A.; Davidovics, N. J.; Davis, C. C.; DeSantis, T. Z.; Deal, C.; Delehaunty, K. D.; Dewhirst, F. E.; Deych, E.; Ding, Y.; Dooling, D. J.; Dugan, S. P.; Michael Dunne, W.; Scott Durkin, A.; Edgar, R. C.; Erlich, R. L.; Farmer, C. N.; Farrell, R. M.; Faust, K.; Feldgarden, M.; Felix, V. M.; Fisher, S.; Fodor, A. A.; Forney, L.; Foster, L.; Di Francesco, V.; Friedman, J.; Friedrich, D. C.; Fronick, C. C.; Fulton, L. L.; Gao, H.; Garcia, N.; Giannoukos, G.; Giblin, C.; Giovanni, M. Y.; Goldberg, J. M.; Goll, J.; Gonzalez, A.; Griggs, A.; Gujja, S.; Haas, B. J.; Hamilton, H. A.; Harris, E. L.; Hepburn, T. A.; Herter, B.; Hoffmann, D. E.; Holder, M. E.; Howarth, C.; Huang, K. H.; Huse, S. M.; Izard, J.; Jansson, J. K.; Jiang, H.; Jordan, C.; Joshi, V.; Katancik, J. A.; Keitel, W. A.; Kelley, S. T.; Kells, C.; Kinder-Haake, S.; King, N. B.; Knight, R.; Knights, D.; Kong, H. H.; Koren, O.; Koren, S.; Kota, K. C.; Kovar, C. L.; Kyrpides, N. C.; La Rosa, P. S.; Lee, S. L.; Lemon, K. P.; Lennon, N.; Lewis, C. M.; Lewis, L.; Ley, R. E.; Li, K.; Liolios, K.; Liu, B.; Liu, Y.; Lo, C.-C.; Lozupone, C. A.; Dwayne Lunsford, R.; Madden, T.; Mahurkar, A. A.; Mannon, P. J.; Mardis, E. R.; Markowitz, V. M.; Mavrommatis, K.; McCorrison, J. M.; McDonald, D.; McEwen, J.; McGuire, A. L.; McInnes, P.; Mehta, T.; Mihindukulasuriya, K. A.; Miller, J. R.; Minx, P. J.; Newsham, I.; Nusbaum, C.; O'Laughlin, M.; Orvis, J.; Pagani, I.; Palaniappan, K.; Patel, S. M.; Pearson, M.; Peterson, J.; Podar, M.; Pohl, C.; Pollard, K. S.; Priest, M. E.; Proctor, L. M.; Qin, X.; Raes, J.; Ravel, J.; Reid, J. G.; Rho, M.; Rhodes, R.; Riehle, K. P.; Rivera, M. C.; Rodriguez-Mueller, B.; Rogers, Y.-H.; Ross, M. C.; Russ, C.; Sanka, R. K.; Sankar, P.; Fah Sathirapongsasuti, J.; Schloss, J. A.; Schloss, P. D.; Schmidt, T. M.; Scholz, M.; Schriml, L.; Schubert, A. M.; Segata, N.; Segre, J. A.; Shannon, W. D.; Sharp, R. R.; Sharpton, T. J.; Shenoy, N.; Sheth, N. U.; Simone, G. A.; Singh, I.; Smillie, C. S.; Sobel, J. D.; Sommer, D. D.; Spicer, P.; Sutton, G. G.; Sykes, S. M.; Tabbaa, D. G.; Thiagarajan, M.; Tomlinson, C. M.; Torralba, M.; Treangen, T. J.; Truty, R. M.; Vishnivetskaya, T. A.; Walker, J.; Wang, L.; Wang, Z.; Ward, D. V.; Warren, W.; Watson, M. A.; Wellington, C.; Wetterstrand, K. A.; White, J. R.; Wilczek-Boney, K.; Qing Wu, Y.; Wylie, K. M.; Wylie, T.; Yandava, C.; Ye, L.; Ye, Y.; Yooseph, S.; Youmans, B. P.; Zhang, L.; Zhou, Y.; Zhu, Y.; Zoloth, L.; Zucker, J. D.; Birren, B. W.; Gibbs, R. A.; Highlander, S. K.; Weinstock, G. M.; Wilson, R. K.; White, O.; The Human Microbiome Project Consortium. A Framework for Human Microbiome Research. *Nature* **2012**, *486* (7402), 215–221. https://doi.org/10.1038/nature11209.

(66) Zimmermann, M.; Zimmermann-Kogadeeva, M.; Wegmann, R.; Goodman, A. L. Mapping Human Microbiome Drug Metabolism by Gut Bacteria and Their Genes. *Nature* **2019**, *570* (7762), 462–467. https://doi.org/10.1038/s41586-019-1291-3.

(67) Maini Rekdal, V.; Bess, E. N.; Bisanz, J. E.; Turnbaugh, P. J.; Balskus, E. P. Discovery and Inhibition of an Interspecies Gut Bacterial Pathway for Levodopa Metabolism. *Science* **2019**, *364* (6445), eaau6323. https://doi.org/10.1126/science.aau6323.

FUNDING SOURCES AKNOWLEDGEMENT

Grant PID2021-127318OB-I00 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe"

AS-R acknowledges the Consejería de Ciencia, Universidades e Innovación de la Comunidad de Madrid, Spain (Ref. PEJ-2020-AI/BIO-17904), for a research assistant contract, and a predoctoral grant (PIPF-2022/SAL-GL-26278).

TABLE OF CONTENTS GRAPHIC