# Leveraging host omic-response signatures to identify microbiome metabolizers of drugs in the vaginal and gut microbiome

Dali Nemecio† and Douglas K. Brubaker
*Department of Biomedical Engineering, School of Medicine,*
*Case Western Reserve University, 10900 Euclid Avenue,*
*Cleveland, OH 44106, United States of America*
*Email: dxn254@case.edu, dkb50@case.edu*

The human microbiome interacts with drugs and contributes to side effects and loss of efficacy. Methods to identify strains with potential to biotransform drugs are inefficient, often fail to scale, have species-level specificity, and rarely account for host factors. We present an *in silico* method to predict the microbial metabolism of drugs using a Microbe-Drug-Host (MDH) signature. The signature integrates non-treated human omics, microbiome composition, and drug-treated *in vitro* transcriptomics to capture features that predict metabolic potential of a given microbe-drug pair. We built support vector machines (SVM) using MDH signatures to predict drug metabolism by vaginal and gut microbes. We established the framework with a single-drug vaginal test case and predicted metabolism of tenofovir with models with high performance metrics (AUC, precision, accuracy, recall, specificity equal to and above 0.84). Generalizing to multidrug gut models, 27 drugs across 130 bacteria were used to generate metabolism probabilities with overall high performance metrics with AUC, precision, accuracy, specificity above or equal to 0.81 and recall at 0.75. The gut microbes most implicated to metabolize drugs include *Bacteroides cellulosilyticus, Bacteroides thetaiotaomicron, and Blautia hansenii*, consistent with existing literature. Our method using the MDH signature addresses deficiencies of existing methods by providing a disease-agnostic, tissue-specific method to identify microbe-drug metabolism interactions.
*Keywords:* machine learning, microbiome, drug metabolism, host-microbiome interactions

## 1. Background

The microbiome dynamically contributes to host heath outcomes in beneficial and deleterious manners. The gut microbiome aids in digestion of dietary fibers, synthesizes micronutrients, and assists with bile metabolism and creates a barrier of protection against pathogens.[1–14] In the vaginal microbiome, bacterial strains provide protection against pathogens by maintaining a low pH environment through lactic acid production and releasing anti-microbial compounds such as hydrogen peroxide and bacteriocins.[15–17] Microbiomes can biotransform exogenous drugs through reduction or oxidation reactions, resulting in side effects on the host such as toxicity or reduced treatment efficacy.[18,19] These interactions have been best characterized in the gut microbiome.[20–23] For instance, sulfasalazine requires an azo bond cleavage facilitated by gut microbiota-secreted azoreductases to become activated. [24,25] Digoxin has a well-established inactivating metabolic relationship with *E. lenta* strains that possess a two-gene cytochrome-encoding operon.[26–29] In the

† Work supported by Startup funding granted by Case Western Reserve University and University Hospitals Cleveland Medical Center
© 2026 The Authors. Open Access chapter published by World Scientific Publishing Company and distributed under the terms of the Creative Commons Attribution Non-Commercial (CC BY-NC) 4.0 License.

vaginal microbiome, the efficacy of the HIV pre-exposure prophylaxis drug tenofovir (TV) is compromised through metabolism by microbes such as *Gardnerella vaginalis*, *Prevotella* (*bivia* and *amnii*), *Mobiluncus mulieris* and *Escherichia coli. These microbes* are implicated in increasing HIV incidence rate to 61%, compared to 18% in *Lactobacillus*-dominant microbiomes.[30 31] In both the gut and vagina many drugs experience biotransformation by the whole microbiome, but few explicit bacterial strains have been identified. This remains a blind-spot for pharmacology towards comprehensively understanding the impact the microbiome has on drug therapies.[19,32,33]

High-throughput methods for identifying microbe-drug interactions are extremely limited. Experimental methods use *in vitro* and *ex vivo* whole microbiome and single-species culturing with drugs.[34 35] Computational methods for inference scale more effectively, but existing tools primarily predict *human* enzymatic interactions with drugs, and rarely predict *bacterial* metabolic capacity.[36–38 39,40] Models that predict drug-microbiome interactions and resulting metabolic products primarily explore whole-microbiome meta-metabolome output, or use functional annotations or metagenomic screens to identify bacteria carrying pre-characterized enzymes.[41–45 46] *In silico* models specific to microbial metabolic products primarily use physiologically based pharmacokinetic (PBPK) models and constraint-based reconstruction and analysis (COBRA).[47 48] Vaginal models that predict bacterial metabolic capacity primarily use the metabolome to explore host-microbiome interactions using genome-scale metabolic models and flux balance analysis methods.[49,50 51] Gut models are more commonly found with wider applications and input data types. Some models explore prediction of drug metabolic product but only in the context of therapeutic effect in disease states.[47,52] Models using ML methods require pre-existing molecular feature identification of the drugs (such as SMILE notations).[53,54] Models like AGORA and AGORA2 identify species-level drug biotransformations but require manual curation of tissue-specific bacterial species, rather than utilizing tissue-derived microbiome composition.[44,55] Other models instead focus on biotransformation due to host diet or microbe population cross-feeding.[56–61] Overall, most existing vaginal and gut models identify whole-microbiome biotransformations and few are species-specific. Our method does not require manual curation of bacterial species, nor pre-identified drug molecular features, and is disease-agnostic. There is no existing *in silico* methods that integrate *both* host multi-omics and *in vitro* data, making our method exceptionally unique in its approach.

Here, we develop a novel computational framework using host, microbiome, and *in vitro* drug screening data to predict species-specific metabolism of drugs across multiple microbiomes using a “microbe-drug-host (MDH)” signature. Each microbe-drug pair contains a unique MDH signature that integrates the human, microbe, and reference drug data into a set of class-separating features that best reflects the metabolic relationship. We establish the framework with a case study of vaginal microbiome metabolism of TV and generalize the pipeline to a multidrug classifier for species-specific drug metabolism in the gut microbiome. This method takes advantage of the non-disease state of human *in vivo* omics combined with high-throughput *in vitro* drug transcriptomics to create a scalable framework to support large-scale screening and prediction for clinically meaningful microbe-drug metabolic relationships.

## 2. Materials and Methods

### *2.1. Pipeline*

Fig. 1 depicts overall data preprocessing, feature selection, testing and training, and performance.

#### *2.1.2. Data pre-processing*

We generated a matrix that contains the integrated host omics, microbiome composition, and *in vitro* drug-treated transcriptomics information. Human transcriptomics or proteomics and paired microbiome composition undergo a Spearman correlation to produce a microbe-host ρ matrix. These ρ values are multiplied by each drug's Library of Network-based Cellular Signatures characteristic direction (LINCS CD) to produce a product matrix. The CD is a mean gene expression over many cell types and treatment conditions. The point-wise product combines drug-treated *in vitro* data with *in vivo* microbe-host ρ. The relationship of a microbe-drug pair is denoted in the model with a binary class, +1 indicates a metabolic and -1 a non-metabolic relationship.

Microbe-drug pairs without literature-established relationships, denoted class of NA, are considered unknown and excluded from the training and testing data. Any microbe-drug pairs with unknown metabolic class are sequestered to the auxiliary predictions dataset which will be used for microbe-drug pair predictions after the model is generated.

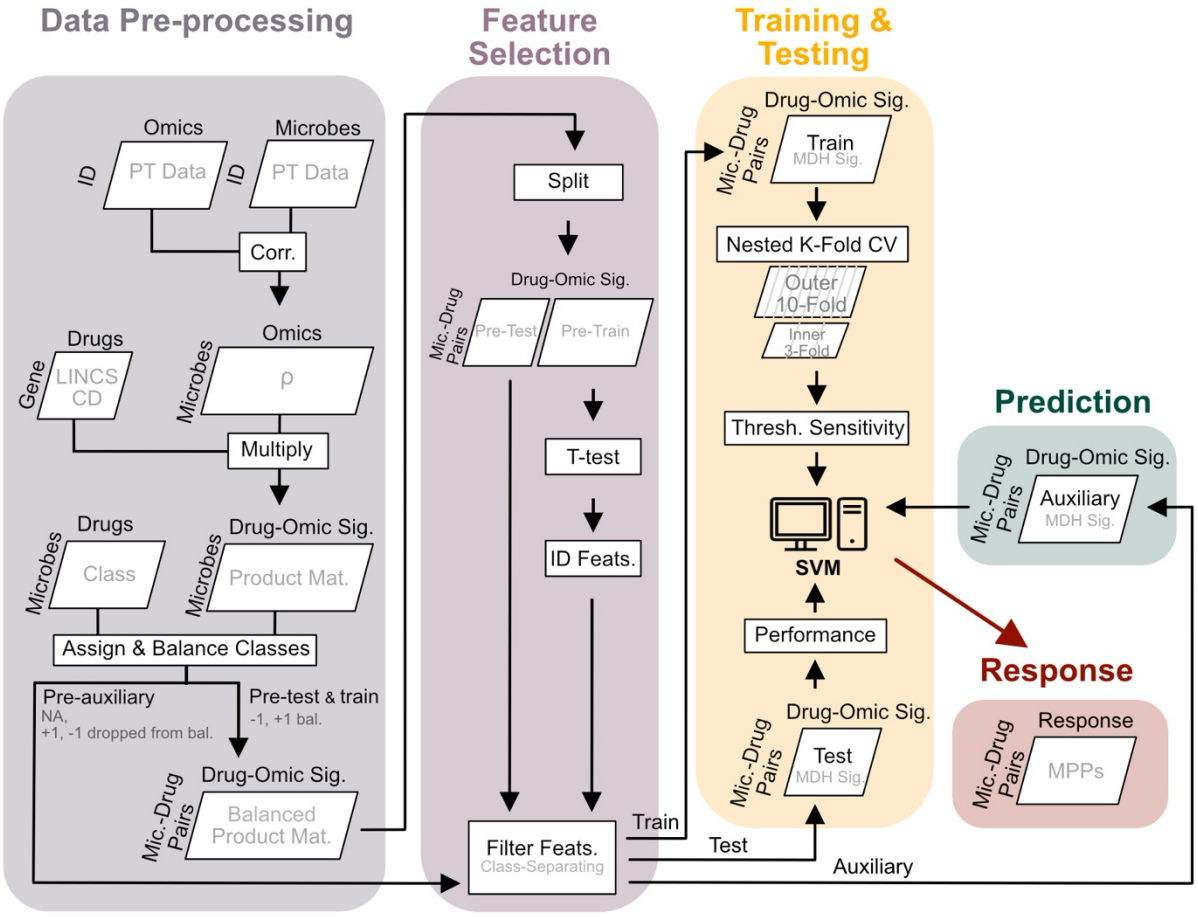



Fig.1: Generic pipeline to generate the MDH signature and SVM model to predict the metabolic relationship of the microbe-drug pairs. The Tv-VagGene, TV-VagProt and MultiDrugV1 and V2 models are adapted from this pipeline.

*2.1.3. Feature selection of the MDH signatures*

We split the product matrix into testing and training data (split 20/80% respectively). The training data undergoes feature selection against the classes associated with the microbe-drug pairs via T-test with Benjamini-Hochberg false discovery rate (FDR) correction. Features with the lowest adjusted p-values define and compose the MDH signatures for model training. SVMs lack built-in feature selection, and by reducing data dimensionality we mitigate potential model overfitting.

*2.1.4. Training and testing*

A SVM model was selected due to its ability to handle high dimensional data well and its inherent property of identifying the largest margin of separation between classes of interest. We used 10-fold cross validation (CV) with nested 3-fold hold out for hyperparameter optimization and determined performance metrics which include precision, accuracy, recall, and specificity for both test and training data. Overall performance calculated from the aggregated CV data uses a prediction threshold of probability 0.5, where >=0.5 is metabolic and < 0.5 probability is non-metabolic. We performed sensitivity analysis to determine the optimal threshold for each metabolic class at 0.05 increments. Optimal prediction thresholds were selected based on best F1 score for metabolic prediction responses and negative predictive value for non-metabolic prediction responses, with recall and specificity used to resolve ties respectively. If ties are not resolved, the lowest threshold for the metabolic class is selected and the highest threshold for the non-metabolic class is selected.

*2.1.5. Predictions and response*

Novel predictions are made with the optimized model using microbe-drug pairs from the auxiliary data that were excluded from the testing and training data. We assessed the prediction responses of the model independently on the testing and training data in addition to the auxiliary predictions. Predictions contain metabolic prediction probabilities (MPPs) for each microbe-drug pair based on the respective MDH signature. Prediction responses are finalized based on the optimal prediction thresholds determined from the sensitivity analysis. MPPs and the corresponding microbe-drug pairs that fall between the metabolic and non-metabolic classification thresholds are considered indeterminate.

### *2.2 Vaginal Data Participant Cohort*

The vaginal models used HIV negative vaginal swab samples from the Partner's PrEP study.[30,62,63] The TV-VagGene cohort contains 20 woman (25-47 years old) with 27 microbes represented and 9451 genes in vaginal tissue transcriptomics. (Fig. S1A) The TV-VagProt cohort contains 192 woman (18-49 years old) with 150 microbes represented and 434 identified proteins from host mucosa. (Fig. S1C)

### *2.3 Gut Data Participant Cohort*

The gut models use 10 control, non-IBD adults from the Inflammatory Bowel Disease Multiomics Database (IBDMDB) as part of the Integrative Human Microbiome Project.[64] In total, 19 samples of ileum and rectum tissue transcriptomics with corresponding microbiome abundance was used. In the microbiome composition, 27 of the microbes were also found in the *in vitro* data, which will assign the microbe-drug "true" relationship for the testing and training data. (Fig. S1B) This is further described below in the 2.4.2 *Preparing Zimmerman Data* section.

### *2.4 Adapted Pipelines*

The pipelines to generate the four models follow the same steps of data preprocessing, feature selection, testing and training, prediction and response. The vaginal models add bootstrapping and a modified feature selection, with testing and training data partitioned earlier in the pipeline. The gut model uses Zimmerman et al 2019 *in vitro* data from a microbe-drug screen to define which microbe-drug pairs to include in the testing and training data.[65] Adapted pipeline details are visualized in Fig. S2.

#### *2.4.1 Vaginal Models Pipeline*

*Pre-processing with Bootstrapping*

The vaginal models, TV-VagGene and TV-VagProt, undergo modified data preprocessing steps that include early testing and training partitioning, and bootstrapping as depicted in Fig. S2A. For Tv-VagGene we generated class-balanced training data from 15 participants and test data from 5 participants via independent bootstrapping. For Tv-VagProt we generated a class-balanced training data from 144 participants and test data from 48 participants via independent bootstrapping. The participant omics data and microbe composition ρ were bootstrapped for 100 iterations with 3 or 10 participants selected for each iteration to produce a single microbe-host ρ matrix for the training and test data respectively. The microbe-host correlations were multiplied by the LINCS CD to generate the product matrix. These steps are repeated to generate a set of 10 data frames. Metabolic relationships of TV were based on *in vitro* validated metabolic capabilities defined by Klatt et al 2017.[30] Microbes with metabolic capacity for TV include *G. vaginalis, M. mulieris, P. amnii, P. Bivia,* and *E. coli*. Microbes with non-metabolic capacity for TV include *L. crispatus* and *L. iners.* These classes are balanced in the training data and then filtered to exclude microbes with unknown metabolic capacity.

*Feature Selection*

Each of the 10 data frames in the training data underwent feature selection, where features with lowest 500 adjusted p-values were selected. The same features found across all 10 data frames were chosen as the MDH signature. This allows for identification of *consistent* features that distinguish the metabolic relationships *independent* of the noise introduced with bootstrapping. The first data frame is arbitrarily chosen to represent the training partition to build the model. TV-VagProt training and testing data both represented 7 microbes with the respective class composition, (200 +1, 200 -1) and (200 +1, 497 -1). TV-VagGene training data represented 4/3 microbes with the respective class composition (169 +1, 170 -1) and (195 +1, 100 -1).

#### *2.4.2 Gut model's in vitro-defined relationships*

The literature-established metabolic relationships needed for building the gut models is defined by *in vitro* data published by Zimmerman et al 2019.[65] This group assessed drug-metabolic activity of 76 human gut bacteria against 271 oral drugs, and reported percentage of drug consumed for each microbe-drug pair. The preprocessing is outlined in Fig. S2B. For our application, this *in vitro* data must be filtered to contain the same drugs and microbes found in the LINCS CD and IBDMDB cohorts respectively. The resulting filtered *in vitro* Zimmerman matrix contains 27 microbes shared with the IBDMDB cohort, and 130 drugs shared with LINCS CD.

*Microbe Proportion and Percent Consumed Thresholds*

We apply to the prepared *in vitro* Zimmerman matrix the *microbe proportion threshold* and *percent consumed threshold.* These thresholds impose cut-offs on the percent consumed values to in order to assign metabolic classes to the microbe-drug pairs. The first threshold is the *microbe proportion threshold,* which specifies for each drug the percent of microbes with a nonzero value of drug consumed. For example, a threshold of 0.25 means 25% of the 27 microbes tested with a drug contained *some* drug consumed. Any drugs and their respective microbes that do not meet this threshold are assigned a metabolic class of NA, and are dropped from the testing and training data.

The *percent consumed threshold* is applied to the remaining drugs and microbes, which specifies the percent of drug consumed to be considered metabolized by the respective microbe. For example, a threshold of 0.50 means 50% or more of the drug must be consumed to be considered a metabolic relationship and assigned a "true" class of +1, otherwise it is denoted as non-metabolic and assigned a class of -1. Each microbe-drug pair in the filtered *in vitro* Zimmerman matrix will have a metabolic class assigned based on the *percent consumed threshold*. The metabolic relationships are either metabolic (+1 class), non-metabolic (-1 class), or unknown (NA class). Microbe-drug pairs with 1) unknown metabolic relationships and 2) pairs dropped due to class balancing, compose the auxiliary data and are used for prediction at the end of the pipeline. Multi-GutV1 model contains *microbe proportion threshold* of 0.50 and *percent consumed threshold of 0.50*. After sensitivity analysis was conducted on these thresholds, Multi-GutV2 contains *microbe proportion threshold* of 0.25 and *percent consumed threshold of 0.50*. For each gut model the testing, training, and auxiliary data together account for 3510 unique combinations of microbe-drug pairs derived from 27 taxa and 130 drugs. Multi-GutV1 training and test data respectively contain representation of 27/25 microbes, 18/17 drugs, with a class composition of (120 +1, 111 -1) and (31 +1, 28 -1). Multi-GutV2 training and test data respectively contain representation of 27/27 microbes, 49/48 drugs, with a class composition of (310 +1, 326 -1) and (78 +1, 82 -1).

## *3.* Results

### *3.1 Vaginal Models' Performance*

Feature selection identified 455 and 91 features for the TV-VagGene and TV-VagProt model respectively. The hyperparameter optimization determined the TV-VagGene model to contain the following parameters: kernel = polynomial, degree = 1, cost=12.1, and gamma=1.41. The hyperparameter optimization determined the TV-VagProt model to contain the following parameters: kernel = polynomial, degree = 3, cost = 8.55, and gamma = 0.507. The performance metrics determined from the testing and aggregated CV training folds for both models are comparable and reflect useful classification from both methods. Notably, precision, accuracy, recall and specificity are all above 0.84 from the testing data and above 0.99 for the aggregated CV training data (Fig. 2C, Fig. S3B) Both models reflect an area under the curve (AUC) above 0.99 in their respective testing and aggregated CV fold performances. The recall performance indicates that both models can identify metabolic microbes within the dataset (0.84 TV-VagGene, 0.99 TV-VagProt). Similarly, the high specificity indicates reliable identification of non-metabolic prediction responses (1.00 TV-VagGene, 0.87 TV-VagProt). The patient pool bootstrapped to produce the training data cannot introduce biological variation beyond what is represented in the original cohort. The small patient pool results in limited heterogeneity and potentially reduces the generalizability of the model to unseen data. However, the learning curve for both models reflects

negligible overfitting, as the generalization gap between the accuracy and mean misclassification error are less than 0.01 for both models. (Fig. S5)

### *3.2 Vaginal Model Sensitivity Analysis*

Sensitivity analysis on the classification threshold in both TV-VagGene and TV-VagProt reflect minimal impact on the overall performance. (Fig. 2A/B) The TV-VagGene model optimal prediction threshold for response is 0.50, with an associated F1 value of 0.91 and recall of 0.84. The non-metabolic prediction threshold is also 0.50 with a determined negative predictive value of 0.92 and specificity of 1.00. The TV-VagProt model optimal metabolic prediction threshold is 0.80 with an associated F1 value of 0.98 and recall of 0.97. The nonmetabolic prediction threshold is 0.25 with an associated negative predictive value of 0.99 and specificity of 0.79.

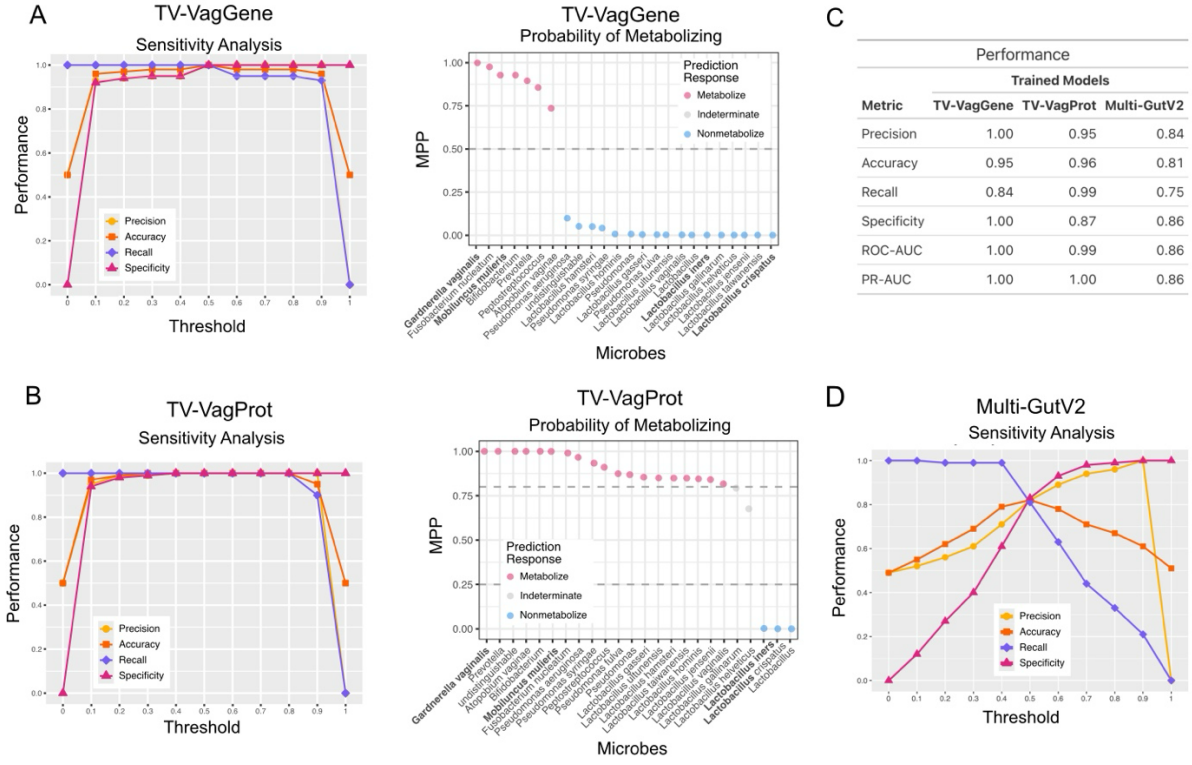



Fig. 2: Vaginal model results and performance metrics from TV-VagGene, TV- VagProt, and MultiDrug-V2. A/B) The sensitivity analysis on the classification threshold are determined from the aggregated CV folds. The MPPs (A, B right) for all direct participant MDH signatures are depicted for each model. Microbes found in both transcriptomics and proteomics data are displayed in the plots. Bolded microbes were used in testing and training. Complete predictions are found in Supplementary Tables 1 and 2. Optimized prediction thresholds determined from the sensitivity analysis are denoted as horizontal dashed lines. C) Performance metrics from three models are depicted from test data on threshold of 0.5. (D) The sensitivity analysis on Multi-GutV2 classification threshold for aggregated CV folds reflects optimal performance at 0.45.

### *3.3 Vaginal Model Predictions*

TV-VagGene and TV-VagProt model prediction responses on various microbes reflect a range of metabolic and non-metabolic responses. (Fig. 2 A/B) Both MPP plots reflect predictions in shared microbes from the TV-VagGene and TV-VagProt cohorts, all microbe MPPs are in Table 1 and 2 respectively. As expected, microbes used for training are predicted correctly with MPPs that support the response. Microbes predicted to metabolize TV includes *G. vaginalis* (1.00 MPP, 1.00 MPP), and *M. Mulieris* (0.93 MPP, 1.00 MPP) for TV-VagGene and TV-VagProt respectively. TV-VagProt also predicted metabolic response for *P. amnii, P. Bivia,* (both 1.00 MPP) and *E. coli* (0.86 MPP). Prediction response of nonmetabolic class is as expected for *L. Crispatus* (5.6e-4 MPP,

3.5e-4 MPP) and *L. iners* (1.2e-3 MPP, 2.4e-3 MPP) for TV-VagGene and TV-VagProt respectively. Comparatively, TV-VagGene returns more conservative predictions across all microbes, despite this, there are consistent predictions between both models.

Novel metabolic predictions shared by both models include *A. vaginae* (0.74 MPP, 1.00 MPP*)*, Bifidobacterium species (0.93 MPP, 1.00 MPP), Prevotella species (0.90 MPP, 1.00 MPP) and Peptostreptococcus species (0.86 MPP, 0.91 MPP) from TV-VagGene and TV-VagProt respectively. Notably, Bifidobacterium and Atopobium genus are highly predicted to be metabolizers of TV. Due to the lack of availability in the cohort data, the TV-VagGene model was not trained on *P. amanii* and *P. bivia*; However, the model correctly identifies the metabolic potential of Prevotella genus. This supports the integrity of the method as the identification of this genus is correct, despite its absence from the training data.

The two models agreed on non-metabolic prediction of Lactobacillus genus. However, there are 11 microbes with conflicting predictions, primarily where TV-VagProt response predicts metabolic relationship (MPP range 0.82-1.00) and TV-VagGene does not (MPP range 1.0e-3 to 0.10). These include Pseudomonas species*, P. aeruginosa, P. syringae, P. fulva, L. gasseri, L. ultunensis, L. hamsteri, L. taiwanensis, L. hominis, L. jensenii,* and *L. vaginalis.* A summary of the predictions between the TV-VagGene and TV-VagProt models is depicted in Fig. S3C.

### *3.4 Multi-GutV1*

The results from Multi-GutV1 model with the *microbe proportion* threshold of 0.5 and *percent consumed* threshold of 0.5 depicts decent performance. (Fig. S6) The model's parameters are: kernel = polynomial, degree = 2, cost = 19.2, gamma = 0.491 with 500 features. The testing fold performance depicts a range of 0.75 - 0.77 for precision, accuracy, recall and specificity. (Fig. S6A) The AUC of the sensitivity and specificity curve and precision-recall curve are 0.83 and 0.87 respectively. These performance values and AUC support moderate performance by the model, and its ability to distinguish the metabolic class. (Fig. S6C) The sensitivity analysis on the prediction threshold reflects a 0.5 threshold is optimal for moderate performance across metrics. (Fig. S6B) The model's predictions grouped by drug reflect higher mean MPPs compared to grouping by microbe. (Fig. S6D/E) This reflects that MPPS are more influenced by the drug type rather than microbe species.

### *3.5 Multi-GutV2 Performance*

During Multi-Gut model development we interrogated *microbe proportion threshold* and *percent consumed threshold* to identify a model with improved performance. Denoted as Multi-GutV2, this model contains a *microbe proportion threshold* of 0.25 and a *percent consumed threshold* of 0.50. The model's parameters are: kernel = polynomial, degree = 2, cost = 15.3, gamma = 0.83 and contains 500 features. This improved Multi-GutV2 model contains performance metrics of precision, accuracy, recall, and specificity between 0.75-0.86. (Fig. 2C, Fig. S7A) The learning curves reflect some minor potential overfitting with a generalization gap of 0.21, likely due to the high model complexity. (Fig. S5) The model built on a balanced dataset can differentiate between the classes well (accuracy 0.81, specificity 0.86) and most positive predictions are correct (precision 0.84). This provides reassurance the prediction responses may provide meaningful MPPs worth interpreting. Sensitivity analysis on the prediction threshold reflects optimal performance at prediction threshold of 0.45 (F1 0.79, recall 0.83). (Fig. 2D) The optimal performance for nonmetabolic class is prediction threshold of 0.25 (negative predictive value 0.9,

specificity 0.31). Performance metrics and confusion matrices for the aggregated CV folds are depicted in Fig. S3A/B.

### *3.6 Multi-GutV2 Trends in Responses*

The Multi-GutV2 responses from the testing and training data show metabolic trends when MPPs are grouped by drugs rather than microbes. When grouped by microbes, the two leading microbes with the highest mean MPPs are *B. cellulosilyticus* and *B thetaiotaomicron* (mean MPP 0.58, mean MPP 0.55). (Fig. S7B, Table 3) Interestingly, when the same MPPs are grouped by drug identity, certain drugs are favored for metabolic relationships by multiple species of microbes. (Fig. 3A, Table 3) The leading drugs with highest mean MPPs are artemisinin and tenatoprazole (mean MPP 0.88, mean MPP 0.84). 37% (18/49) of drugs contain mean MPPs that are above the metabolic prediction threshold of 0.45. There are no structural or pharmaceutical classes overrepresented in these 18 drugs. However, some of these drugs share metabolic mechanisms, specifically involving hydrolysis and reduction reactions. These 18 include prodrugs that require activation and contain the respective mean MPPs which include: bisacodyl (0.76), sulfasalazine (0.61), famciclovir (0.59), racecadotril (0.56), deflazacort (0.51). These also include known drug substrates for reduction and cleavage with the following mean MPPs: sulfasalazine (0.61), phenazopyridine (0.62), and tinidazole (0.75).

The predictions on microbe-drug pairs from the testing data reflects consistent classification of the model's response and the corresponding true class. For microbe-drug pairs assigned a class "truth" of metabolic, the median MPP is 0.60 and median percent consumed is 0.96. For the "true" non-metabolic microbe-drug pairs, the median MPP is 0.13 and median percent consumed is 0.0. A Spearman correlation between the MPPs for testing microbe-drug pairs and the corresponding percent consumed *in vitro* value returns a $\rho$ of 0.53 with a p-value of 7.3e-12. The positive direction, meaningful magnitude and notable p-value reflect the *in vitro* percent consumed values associated with each microbe-drug pair are highly correlated to the MPP produced by the model. This correlation reflects the MDH signature can capture wide metabolic range, and corresponding MPP's are reflective of measured activity in the *in vitro* values used to assign the classes. Together, these values support the useful classification abilities of the model.

Compared to the testing and training MPPs, the predictions on the auxiliary microbe-drug pairings assigned a truth of NA reflect a smaller range of MPPs. This is evident when grouping the MPPs by microbe and by drug. The two leading microbes identified to metabolize the drugs include *B. hansenii* (mean MPP 0.50) and *B. thetaiotaomicron* (mean MPP 0.50). (Fig. S7C, Table 3) The two leading drugs predicted to be metabolized by the microbes are metaxalone (mean MPP 0.64) and fenofibrate (mean MPP 0.54). (Fig. 3B) 44% (36/81) of the drugs contain mean MPPs above the metabolic prediction threshold of 0.45. The auxiliary microbe-drug pairs were selected based on failure to meet the *microbe proportion threshold* and *percent consumed threshold*, and therefore it is reasonable that we see more MPPs trending towards non-metabolic responses. Similar to the testing and training predictions, the auxiliary microbe-drug pair responses are more distinguishable based on the drug rather than the microbes. Experimentally these auxiliary microbe-drug pairs had very little metabolism representation across the microbes. The MPPs are less certain compared to those produced by the testing and training data, but are still above our classification threshold for metabolic predictions.

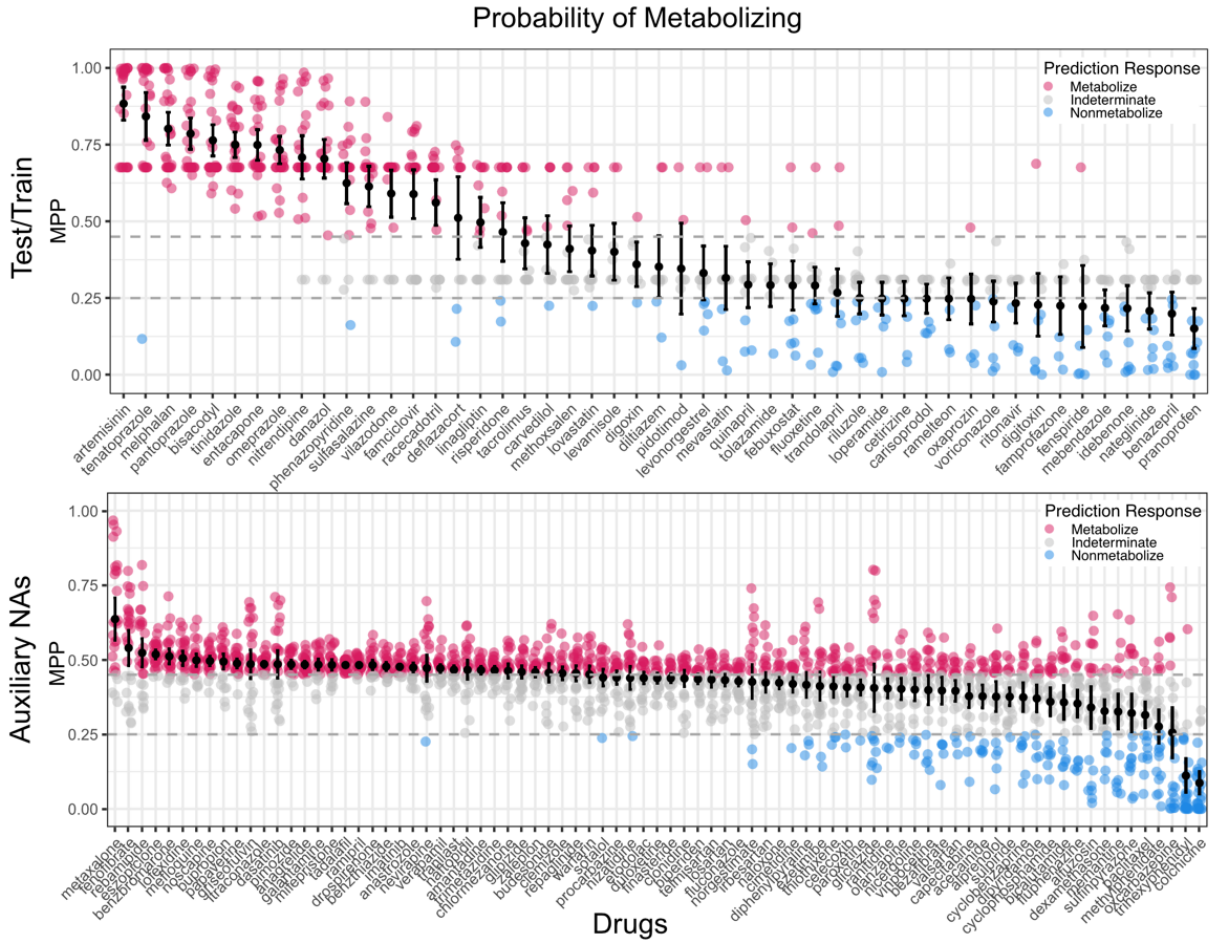



Fig. 3: Multi-GutV2 prediction responses grouped by drugs from the training and testing data (top) and auxiliary NA data (bottom). Optimized prediction thresholds determined from the sensitivity analysis are denoted as horizontal dashed lines.

### *3.7 Literature Exploration of Multi-GutV2 Responses*

We asses Multi-GutV2 responses with literature-established gut microbiome metabolic relationships to explore our results. These include microbe-drug pairs from across the testing, training and auxiliary data. (Table 3) We explore known relationships between sulfasalazine, omeprazole, digoxin and levamisole metabolism evidence. The sulfasalazine literature-established relationship with the gut microbiome is reflected in the MPPs. Microbe-sulfasalazine pairs predicted as metabolic returned a relatively high median MPP of 0.68. For sulfasalazine, 78% (21/27) of microbes are predicted as metabolic and 22% (6/27) are indeterminate. Interestingly, for these six the true assigned class is denoted as non-metabolic, but since the MPPs are not low enough meet the non-metabolic prediction threshold of 0.25, they fall within the indeterminate class. This suggests the associated MDH signatures for these six sulfasalazine-microbe pairs still contain enough "metabolic" signature so that the model is unable to confidently predict them to be non-metabolic. Our results reflect *P. distasonis* and *B. vulgatus* contain high MPPs (0.89, 0.68 respectively). Interestingly, *ex vivo* human fecal cultures treated with sulfasalazine displayed a significantly increased species abundance of these same two bacterial species.[66] Additionally, of the Bacteroides genera, *B. caccae* and *B. Fragillis* (MPP 0.77, MPP 0.75) are most likely metabolic towards sulfasalazine. Within literature, *B. caccae* has been identified as a biomarker for children with ulcerative colitis.[67] At the same time, *B. caccae* is also reported as a predictor for successful

treatment outcomes with anti-cytokine therapies.[20] There are no findings associating successful treatment outcomes with sulfasalazine in children with ulcerative colitis and *B. caccae* abundance. However, the Bacteroides genera has been identified to contain azoreductases implicated in activating sulfasalazine through reduction.[68] Indeed our predictions reflect this trend, where 83% (10/12) of the Bacteroides species are predicted to be metabolic of sulfasalazine with a median MPP of 0.68. There are two Bacteroides species that are denoted as undetermined, *B. stercoris* and *B. finegoldii*, however the *in vitro* assigned truth is non-metabolic. The lack of consistency between the truth and response may be due to metabolic capacity being captured in the MDH signature. Reductive metabolism of omeprazole has been reported in *in vitro* conditions by gut bacteria to produce sulfide metabolites however there are few species explicitly identified.[26,69] In our model, 96% (26/27) of microbe-omeprazole pairs were assigned a truth of metabolic and 100% (27/27) had a predicted response of metabolic. Digoxin has an interesting literature established relationship with *E. lenta*, where only strains that contain a two-gene cytochrome-encoding operon metabolize the drug. Our results (MPP 0.31, indeterminate) reflect this range of *E. Lenta* response, since it is not a metabolic relationship held by the entire species and thusly the MPP is low enough to be placed in the indeterminate response. Due to the MDH signature capturing some of this metabolic capacity the MPP is not below the non-metabolic threshold. Microbe-levamisole pairs had responses that deviated from the "truth" assigned. There are two microbes *A. colihominis* and *E. coli* out of the 27 levamisole-microbe pairs assigned a truth of metabolic. Interestingly, the model prediction responses identified seven microbes to be metabolic (median MPP 0.53). Evidence suggests Levamisole may be metabolized by Bacteroides species, and indeed three of these levamisole-microbe pairs included *B. uniformis* (MPP 0.72). *B. cellulosilyticus* (MPP 0.54), and *B. thetaiotaomicron* (MPP 0.50).[70] Our results with sulfasalazine, omeprazole, digoxin and levamisole are explored due to their existing literature evidence to have metabolic properties with the gut microbiome. These results highlight the model's capability to return predictions that align with literature-based findings. Drug-metabolizing capacity can vary among strains therefore taxonomic trends do not necessarily indicate conserved metabolic capacity across all members. Discrepancies between the response and truth may indicate the MDH signature's ability to discern between the *in vivo* metabolic capacity *independent* of the assigned "true" class based on the *in vitro* data.

Within the auxiliary prediction data, we explore the prediction responses for microbe-drug pairs with an unknown true class. (Fig. 3, Fig. S7C, Table 3) Some of the 18 drugs identified to be metabolized include capecitabine, sulfinpyrazone, ranitidine and nizatidine. Our results reflect varying degrees of model responses that align with literature-established evidence. Capecitabine is reduced by dihydropyrimidine dehydrogenase, an enzyme previously identified in 85 species of bacteria through comparative genomics.[44] In our results, capecitabine returned 22% (6/27) microbes as having a metabolic relationship (median MPP 0.49). Sulfinpyrazone can experience s-oxide reduction by microbes found in human stool.[71,72] For sulfinpyrazone, 30% (8/27) microbes were predicted to have metabolic relationship (median MPP 0.52). Ranitidine and nizatidine, when incubated with human stool experience a nitro reduction via cleavage of an N-oxide bond.[73,74] Ranitidine-microbe pairs had 33% (9/27) predicted to be metabolic (median MPP 0.47). Nizatidine-microbe pairs had 52% (14/27) microbes predicted to metabolic (median MPP 0.51). Despite these microbe-drug pairs being sequestered to the auxiliary data during preprocessing class assignment due to their inability to meet the *percent consumed* and *microbe proportion* threshold, we still see predictions for microbes with metabolic capacity. The median MPPs are lower compared to those from the testing/training. This is somewhat reasonable, since the auxiliary data

was not used in training, and therefore the MDH signature is not tuned to detect any metabolic capacity within these drugs. Despite this, the metabolic predictions demonstrate the generalizable, MDH signature *can* distinguish metabolic capacities on drugs that did not contribute to the training of the model.

Metabolic relationships identified from *in vitro* screening with drugs can contribute to our understanding of drug-microbe interactions. However, it does not capture dynamic microbiome composition which often impacts bacterial metabolic need and activity. While we can state what the assigned "truth" is based off the *in vitro* data, the endogenous behavior of the bacteria species in the host microbiome context may be different. Due to this misalignment, it is possible that behavior observed *in vitro* is not behavior observed in the endogenous host microbiome. These results reflect the MDH signature derived from *in vivo* patient data can capture metabolic relationships, which are difficult to replicable *in vitro*. The MDH signature improves on the existing *in vitro* and *in silico* methods as it integrates the *context* of the complete endogenous microbiome interaction with the host. The Multi-GutV2 model has identified species-specific microbes with MDH signatures consistent with metabolic activity.

## *4. Discussion*

The MDH signature can predict species-specific metabolic relationships. Our model is distinct from existing methods since it is derived from human omics data and microbiome composition from disease-free subjects and integrates the drug-treatment variable through *in vitro* derived data. Our findings highlight the novelty and utility of using a MDH signature to identify drug-microbial metabolic relationships in the vaginal and gut microbiome. Using this MDH signature we build vaginal microbiome models, TV-VagGene and TV-VagProt, that predict metabolism of TV. These models contain high performance metrics on test data (precision, accuracy, recall, specificity equal to and above 0.84) and an AUC equal to and above 0.99. The two models agree on the novel predictions of TV metabolism by the vaginal bacteria: *A. vaginae*, *F. nucleatum*, Peptostreptococcus genus, and Bifidobacterium genus.

We further demonstrate the greater capacity of the MDH signature to capture multiple microbe-drug metabolic relationships in the gut microbiome with the Multi-GutV1 and Multi-GutV2 models. The Multi-GutV2 model was identified through sensitivity training of the V1 model and contains useable performance metrics from the testing data (precision, accuracy, specificity above or equal to 0.81 and recall at 0.75), with an AUC of 0.87. For the Multi-GutV2 model the leading microbes most implicated in metabolism of the assessed drugs are *B. cellulosilyticus*, *B thetaiotaomicron*, and *B. hansenii*. There are 54 total drugs with mean MPPs across the respective microbe-drug pairs assed that reflect metabolic potential with multiple microbe species. The responses also reflect trends in drug literature-established metabolic relationships. The Multi-GutV2 model produced MPPs in the testing data which are highly correlated ($\rho$ of 0.55, p-value 1.3e-53) to the percent consumed values from the *in vitro* data. This reflects the MDH signature's ability to capture and discern the range of metabolic capacity of each unique microbe-drug pair.

The method to produce the MDH signatures is highly customizable as it requires 1) non-disease paired omics and 2) a few microbe-drugs with defined metabolic relationships. It is advantageous the human omics and microbiome composition is *non-disease* as this expands the range of applicable available datasets. The human paired omics and microbiome composition are tissue-specific and disease-agnostic. The training data requires defined metabolic relationships for microbe-drug pairs to establish the model. These can be taken from literature (as shown in TV-

VagGene and TV-VagProt models) or custom *in vitro* experiments with a threshold to delineate defined metabolic relationships (as done in the Multi-GutV1 and V2 models).

Future applications of this model could adapt the existing pipeline and shift the specificity of the model to suit the research interests. The utility of this model is constrained to the microbiome diversity of patient samples. This can impact the performance of the model though limited samples for testing and training or limit the amount of novel microbe-drug pairs that are predicted on. Future applications of this method could generate models unique to specific drug classes or hypothesized enzymatic mechanisms. Drugs in similar classes or with similar functional groups have been shown to share metabolic properties consistent across different microbes with the same enzymes.[19,32,33,65,75] Future applications can incorporate *in vitro* or *ex vivo* validation of the microbe-drug prediction, which is a limitation in our interpretation. Another model constraint is the feature selection T-test, which may omit interaction effects leveraged by the SVM. Future implementations could consider multivariate approaches with built-in feature selection such as Elastic Net, LASSO, or PCA.

Another improvement could be applying tissue-specific *in vitro* drug-treated cell line data rather than the LINCS CD values. The LINCS CD readings include a vast diversity of cell types across various tissue locations. The models may have better performance if the *in vitro* values were more tissue-specific and therefore would have more contextually accurate drug-treatment values to contribute to the MDH signatures. Future applications of this model could also expand using a larger applied drug pool. Here, we limited our microbe-drug pairs to drugs that were found in both our *in vitro* drug-treated microbes from the Zimmerman et al 2019 publication to simplify interpretation of our results and to ensure we had *in vitro* values to compare our generated MPP per microbe-drug pair. Future applications could use the whole LINCS catalogue to generate microbe-drug pair predictions for all LINCS drugs to generate a reference database. This generated library of predictions could be cited to preemptively identify any potential impact of the gut microbiome on the drug of interest. We could also make individual SVMs per drug for each microbiome, similar to the individual models we made for TV metabolism. These single-drug models may contain higher precision since the class-separating features in the MDH signatures capture each specific drug's metabolic properties. The weakness of a multidrug model such as the Multi-GutV1 and V2 is the class-distinguishing features are *not* specific to the drug class nor the structures, which can contribute to loss of performance. This is reasonable, since introducing additional drugs may obscure the features that best distinguish the classification abilities. However, despite this, in a multidrug model the features are more generalizable *because* the features are not tied to one drug. We have shown reliable performance is still viable in a multidrug model.

The utility of these models is most evident within the clinical setting. We can preemptively determine how a patient's microbiome may impact the therapeutic effect of the drug. With this knowledge we can either pivot drug treatment to select a therapeutic that will not interact with the microbiome or provide probiotics to shift the microbiome populations to be more receptive to the drug of choice. This is a low-cost, high throughput method to assess the microbe metabolism of drugs in an accessible manner. Clinicians can reference this tool to incorporate microbiome state into patient treatment plans going forward, thereby improving personalized medicine and treatment success.

## 5. Supplementary Information

All supplementary figures, tables, code and high-resolution main figures can be found here: https://github.com/Brubaker-Lab/microbe-host-omic_nemecio

## 6. References

1. Jiménez-Escrig, A. & Sánchez-Muniz, F. J. Dietary fibre from edible seaweeds: Chemical structure, physicochemical properties and effects on cholesterol metabolism. *Nutr. Res.* **20**, 585–598 (2000).

2. Holscher, H. D. Dietary fiber and prebiotics and the gastrointestinal microbiota. *Gut Microbes* **8**, 172–184 (2017).

3. Williams, B. A., Mikkelsen, D., Flanagan, B. M. & Gidley, M. J. 'Dietary fibre': moving beyond the 'soluble/insoluble' classification for monogastric nutrition, with an emphasis on humans and pigs. *J. Anim. Sci. Biotechnol.* **10**, 45 (2019).

4. Kumar, J., Rani, K. & Datt, C. Molecular link between dietary fibre, gut microbiota and health. *Mol. Biol. Rep.* **47**, 6229–6237 (2020).

5. Biesalski, H. K. Nutrition meets the microbiome: micronutrients and the microbiota. *Ann. N. Y. Acad. Sci.* **1372**, 53–64 (2016).

6. Barone, M., D'Amico, F., Brigidi, P. & Turroni, S. Gut microbiome–micronutrient interaction: The key to controlling the bioavailability of minerals and vitamins? *BioFactors* **48**, 307–314 (2022).

7. Rowland, I. *et al.* Gut microbiota functions: metabolism of nutrients and other food components. *Eur. J. Nutr.* **57**, 1–24 (2018).

8. Soto-Martin, E. C. *et al.* Vitamin Biosynthesis by Human Gut Butyrate-Producing Bacteria and Cross-Feeding in Synthetic Microbial Communities. *mBio* **11**, e00886-20 (2020).

9. Hofmann, A. F. The continuing importance of bile acids in liver and intestinal disease. *Arch. Intern. Med.* **159**, 2647–2658 (1999).

10. Ridlon, J. M., Kang, D. J., Hylemon, P. B. & Bajaj, J. S. Bile acids and the gut microbiome. *Curr. Opin. Gastroenterol.* **30**, 332–338 (2014).

11. Xu, Z. & Knight, R. Dietary effects on human gut microbiome diversity. *Br. J. Nutr.* **113**, S1–S5 (2015).

12. Shreiner, A. B., Kao, J. Y. & Young, V. B. The gut microbiome in health and in disease. *Curr. Opin. Gastroenterol.* **31**, 69–75 (2015).

13. Jones, M. L., Martoni, C. J., Ganopolsky, J. G., Labbé, A. & Prakash, S. The human microbiome and bile acid metabolism: dysbiosis, dysmetabolism, disease and intervention. *Expert Opin. Biol. Ther.* **14**, 467–482 (2014).

14. Karlsson, F. H., Ussery, D. W., Nielsen, J. & Nookaew, I. A closer look at bacteroides: phylogenetic relationship and genomic implications of a life in the human gut. *Microb. Ecol.* **61**, 473–485 (2011).

15. Witkin, S. & Linhares, I. Why do lactobacilli dominate the human vaginal microbiota? *BJOG Int. J. Obstet. Gynaecol.* **124**, 606–611 (2017).

16. Witkin, S. S. *et al.* Influence of vaginal bacteria and D- and L-lactic acid isomers on vaginal extracellular matrix metalloproteinase inducer: implications for protection against upper genital tract infections. *mBio* **4**, e00460-13 (2013).

17. V. Sgibnev, A. & A. Kremleva, E. Vaginal Protection by H2O2-Producing Lactobacilli. *Jundishapur J. Microbiol.* **8**, e22913 (2015).

18. Wilson, I. D. & Nicholson, J. K. Gut Microbiome Interactions with Drug Metabolism, Efficacy and Toxicity. *Transl. Res. J. Lab. Clin. Med.* **179**, 204 (2016).

19. Enright, E. F., Gahan, C. G. M., Joyce, S. A. & Griffin, B. T. The Impact of the Gut Microbiota on Drug Metabolism and Clinical Outcome. *Yale J. Biol. Med.* **89**, 375–382 (2016).

20. Lee, J. W. J. *et al.* Multi-omics reveal microbial determinants impacting responses to biologic therapies in inflammatory bowel disease. *Cell Host Microbe* **29**, 1294-1304.e4 (2021).

21. Ananthakrishnan, A. N. *et al.* Gut Microbiome Function Predicts Response to Anti-integrin Biologic Therapy in Inflammatory Bowel Diseases. *Cell Host Microbe* **21**, 603-610.e3 (2017).

22. Kostrzewska, M. *et al.* The effect of omeprazole treatment on the gut microflora and neutrophil function. *Clin. Res. Hepatol. Gastroenterol.* **41**, 575–584 (2017).

23. Koo, S. H. *et al.* Effects of proton pump inhibitor on the human gut microbiome profile in multi-ethnic groups in Singapore. *Singapore Med. J.* **60**, 512–521 (2019).

24. Sousa, T. *et al.* On the Colonic Bacterial Metabolism of Azo-Bonded Prodrugsof 5-Aminosalicylic Acid. *J. Pharm. Sci.* **103**, 3171–3175 (2014).

25. Seradj, D.-S. *et al.* Influence of Different Diets on the Degradation of Sulfasalazine by Colon Bacteria Determined Using MimiCol3. *Pharmaceuticals* **16**, 1128 (2023).

26. Guo, J. *et al.* Gut microbiota and host Cyp450s co-contribute to pharmacokinetic variability in mice with non-alcoholic steatohepatitis: Effects vary from drug to drug. *J. Adv. Res.* **39**, 319–332 (2022).

27. Saha, J. R., Butler, V. P., Neu, H. C. & Lindenbaum, J. Digoxin-inactivating bacteria: identification in human gut flora. *Science* **220**, 325–327 (1983).

28. Haiser, H. J. *et al.* Predicting and Manipulating Cardiac Drug Inactivation by the Human Gut Bacterium Eggerthella lenta. *Science* **341**, 295–298 (2013).

29. Lindenbaum, J., Rund, D. G., Butler, V. P., Tse-Eng, D. & Saha, J. R. Inactivation of Digoxin by the Gut Flora: Reversal by Antibiotic Therapy. *N. Engl. J. Med.* **305**, 789–794 (1981).

30. Klatt, N. R. *et al.* Vaginal bacteria modify HIV tenofovir microbicide efficacy in African women. *Science* **356**, 938–945 (2017).

31. Cheu, R. K. *et al.* Impact of vaginal microbiome communities on HIV antiretroviral-based pre-exposure prophylaxis (PrEP) drug metabolism. *PLOS Pathog.* **16**, e1009024 (2020).

32. Sun, C., Chen, L. & Shen, Z. Mechanisms of gastrointestinal microflora on drug metabolism in clinical practice. *Saudi Pharm. J.* **27**, 1146–1156 (2019).

33. Martinelli, F. & Thiele, I. Microbial metabolism marvels: a comprehensive review of microbial drug transformation capabilities. *Gut Microbes* **16**, 2387400.

34. Wu, Q. *et al.* Activity of GPCR-targeted drugs influenced by human gut microbiota metabolism. *Nat. Chem.* **17**, 808–821 (2025).

35. Verdegaal, A. A. *et al.* A drug–microbiome–drug interaction impacts co-prescribed medications for Parkinson's disease. *Nat. Microbiol.* **11**, 1387–1409 (2026).

36. Lettieri, M., Rodda, M. & Carlucci, V. Machine Learning in Early Prediction of Metabolism of Drugs. in *Computational Toxicology: Methods and Protocols* (ed. Nicolotti, O.) 275–291 (Springer US, New York, NY, 2025). doi:10.1007/978-1-0716-4003-6_13.

37. Kazmi, S. R., Jun, R., Yu, M.-S., Jung, C. & Na, D. *In silico* approaches and tools for the prediction of drug metabolism and fate: A review. *Comput. Biol. Med.* **106**, 54–64 (2019).

38. Dudas, B. & Miteva, M. A. Computational and artificial intelligence-based approaches for drug metabolism and transport prediction. *Trends Pharmacol. Sci.* **45**, 39–55 (2024).

39. Litsa, E. E., Das, P. & Kavraki, L. E. Machine learning models in the prediction of drug metabolism: challenges and future perspectives. *Expert Opin. Drug Metab. Toxicol.* **17**, 1245–1247 (2021).

40. Kazmi, S. R., Jun, R., Yu, M.-S., Jung, C. & Na, D. In silico approaches and tools for the prediction of drug metabolism and fate: A review. *Comput. Biol. Med.* **106**, 54–64 (2019).

41. Guthrie, L., Wolfson, S. & Kelly, L. The human gut chemical landscape predicts microbe-mediated biotransformation of foods and drugs. *eLife* **8**, e42866 (2019).

42. Javdan, B. *et al.* Personalized Mapping of Drug Metabolism by the Human Gut Microbiome. *Cell* **181**, 1661-1679.e22 (2020).

43. Wensel, C. R., Pluznick, J. L., Salzberg, S. L. & Sears, C. L. Next-generation sequencing: insights to advance clinical investigations of the microbiome. *J. Clin. Invest.* **132**, e154944.

44. Heinken, A. *et al.* Genome-scale metabolic reconstruction of 7,302 human microorganisms for personalized medicine. *Nat. Biotechnol.* **41**, 1320–1331 (2023).

45. Malla, M. A. *et al.* Exploring the Human Microbiome: The Potential Future Role of Next-Generation Sequencing in Disease Diagnosis and Treatment. *Front. Immunol.* **9**, (2019).

46. Bustion, A. E., Nayak, R. R., Agrawal, A., Turnbaugh, P. J. & Pollard, K. S. SIMMER employs similarity algorithms to accurately identify human gut microbiome species and enzymes capable of known chemical transformations. *eLife* **12**, e82401 (2023).

47. Bhalla, P., Rengaswamy, R., Karunagaran, D., Suraishkumar, G. K. & Sahoo, S. Metabolic modeling of host–microbe interactions for therapeutics in colorectal cancer. *Npj Syst. Biol. Appl.* **8**, 1 (2022).

48. Stevanoska, M., Folz, J., Beekmann, K. & Aichinger, G. Physiologically based kinetic (PBK) modeling as a new approach methodology (NAM) for predicting systemic levels of gut microbial metabolites. *Toxicol. Lett.* **396**, 94–102 (2024).

49. Hu, T. *et al.* Genome-scale metabolic modeling reveals specific vaginal Lactobacillus strains and their metabolites as key inhibitors of Candida albicans. *Microbiol. Spectr.* **13**, e02984-24 (2025).

50. Glass, E. M., Kolling, G. L. & Papin, J. A. Genome-scale metabolic modelling identifies vaginal microbiome members as potential probiotics. *Nat. Microbiol.* **11**, 2034–2046 (2026).

51. Lee, C. Y., Dillard, L. R., Papin, J. A. & Arnold, K. B. New perspectives into the vaginal microbiome with systems biology. *Trends Microbiol.* **31**, 356–368 (2023).

52. Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R. & Goodman, A. L. Separating host and microbiome contributions to drug pharmacokinetics and toxicity. *Science* **363**, eaat9931 (2019).

53. McCoubrey, L. E. *et al.* Machine Learning Predicts Drug Metabolism and Bioaccumulation by Intestinal Microbiota. *Pharmaceutics* **13**, 2001 (2021).

54. Malwe, A. S., Srivastava, G. N. & Sharma, V. K. GutBug: A Tool for Prediction of Human Gut Bacteria Mediated Biotransformation of Biotic and Xenobiotic Molecules Using Machine Learning. *J. Mol. Biol.* **435**, 168056 (2023).

55. Magnúsdóttir, S. *et al.* Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat. Biotechnol.* **35**, 81–89 (2017).

56. Chen, X., Alonso, A. P., Allen, D. K., Reed, J. L. & Shachar-Hill, Y. Synergy between (13)C-metabolic flux analysis and flux balance analysis for understanding metabolic adaptation to anaerobiosis in E. coli. *Metab. Eng.* **13**, 38–48 (2011).

57. Vijayakumar, S., Rahman, P. K. S. M. & Angione, C. A Hybrid Flux Balance Analysis and Machine Learning Pipeline Elucidates Metabolic Adaptation in Cyanobacteria. *iScience* **23**, 101818 (2020).

58. Burgard, A. P., Pharkya, P. & Maranas, C. D. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* **84**, 647–657 (2003).

59. Oskotsky, T. T. *et al.* A transcriptomics-based computational drug repurposing pipeline identifies simvastatin and primaquine as therapeutics for endometriosis. *iScience* **29**, (2026).

60. Heinken, A. & Thiele, I. Systematic prediction of health-relevant human-microbial co-metabolism through a computational framework. *Gut Microbes* **6**, 120–130 (2015).

61. Bauer, E., Zimmermann, J., Baldini, F., Thiele, I. & Kaleta, C. BacArena: Individual-based metabolic modeling of heterogeneous microbes in complex communities. *PLOS Comput. Biol.* **13**, e1005544 (2017).

62. Heffron, R. *et al.* Efficacy of oral pre-exposure prophylaxis (PrEP) for HIV among women with abnormal vaginal microbiota: a post-hoc analysis of the randomised, placebo-controlled Partners PrEP Study. *Lancet HIV* **4**, e449–e456 (2017).

63. Baeten, J. M. *et al.* Antiretroviral prophylaxis for HIV prevention in heterosexual men and women. *N. Engl. J. Med.* **367**, 399–410 (2012).

64. Lloyd-Price, J. *et al.* Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **569**, 655–662 (2019).

65. Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R. & Goodman, A. L. Mapping human microbiome drug metabolism by gut bacteria and their genes. *Nature* **570**, 462 (2019).

66. Li, L. *et al.* RapidAIM: a culture- and metaproteomics-based Rapid Assay of Individual Microbiome responses to drugs. *Microbiome* **8**, 33 (2020).

67. Mouzan, M. E., Sarkhy, A. A. & Assiri, A. Gut microbiota predicts the diagnosis of ulcerative colitis in Saudi children. *World J. Clin. Pediatr.* **13**, (2024).

68. Braccia, D. J. *et al.* Gut Microbiome–Wide Search for Bacterial Azoreductases Reveals Potentially Uncharacterized Azoreductases Encoded in the Human Gut Microbiome. *Drug Metab. Dispos.* **51**, 142–153 (2023).

69. Watanabe, K., Yamashita, S., Furuno, K., Kawasaki, H. & Gomita, Y. Metabolism of Omeprazole by Gut Flora in Rats. *J. Pharm. Sci.* **84**, 516–517 (1995).

70. Shu, Y.-Z., Kingston, D. G. I., Van Tassell, R. L. & Wilkins, T. D. Metabolism of levamisole, an anti-colon cancer drug, by human intestinal bacteria. *Xenobiotica* **21**, 737–750 (1991).

71. Renwick, A. G., Evans, S. P., Sweatman, T. W., Cumberland, J. & George, C. F. The role of the gut flora in the reduction of sulphinpyrazone in the rat. *Biochem. Pharmacol.* **31**, 2649–2656 (1982).

72. Strong, H. A., Renwick, A. G., George, C. F., Liu, Y. F. & Hill, M. J. The reduction of sulphinpyrazone and sulindac by intestinal bacteria. *Xenobiotica* **17**, 685–696 (1987).

73. Basit, A. W., Newton, J. M. & Lacey, L. F. Susceptibility of the H2-receptor antagonists cimetidine, famotidine and nizatidine, to metabolism by the gastrointestinal microflora. *Int. J. Pharm.* **237**, 23–33 (2002).

74. Basit, A. W. & Lacey, L. F. Colonic metabolism of ranitidine: implications for its delivery and absorption. *Int. J. Pharm.* **227**, 157–165 (2001).

75. Jourova, L., Anzenbacher, P. & Anzenbacherova, E. Human gut microbiota plays a role in the metabolism of drugs. *Biomed. Pap.* **160**, 317–326 (2016).