Article

Intestinal permeability of N-acetylcysteine is driven by gut microbiota-dependent cysteine palmitoylation

Yu-Hang Zhang [1,2,8] ✉, Chen-Shu Dai[1,3,8], Ya-Jie Wang[4], Wen-Yu Wang [5], Tian-Tian Qi[1,6], Man-Cheng Xia[4], Gan Zhou[7] ✉ & Yi-Min Cui [1,2] ✉

Trillions of intestinal microbiota are essential to the permeability of orally administered drugs. However, identifying microbial-drug interactions remains challenging due to the highly variable composition of intestinal flora among individuals. Using single-pass intestinal perfusion (SPIP) platform, we establish the microbiota-based permeability screening framework involving germ-free (GF) and specific-pathogen-free (SPF) rats to compare in-situ P_{eff}-values and metabolomic profiles of 32 orally administered drugs with disputable classifications of permeability, prior to the verifications of bioorthogonal chemistry and LC-MS/MS. In contrast with SPF controls, N-Acetylcysteine (NAC) exhibits significantly increased permeability in GF rats, which is inversely related to reduced cysteine-3-ketosphinganine by *Bacteroides*. To further validate these microbiome features, we integrate clinical descriptors from a prospective cohort of 319 participants to optimize a 15-feature eXtreme Gradient Boosting (XGB) model, which reveal that cysteine palmitoylation by intestinal microbiota has significantly affected NAC permeability. By comparison of net reclassification improvement (NRI) index, this machine learning (ML) model of clinical prediction model encompassing intestinal microbial features outperforms other three commercial models in predicting NAC permeability. Here we have developed an intestinal microbiota-based strategy to evaluate uncharacterized NAC permeability, thus accounting for its discordant biopharmaceutics classification.

The pharmaceutical industry has put great demands on reducing the cost, time and number of clinical drug trials in the development/approval of novel dosage forms of generic drug products around the world[1,2]. Biopharmaceutics Classification System (BCS) evaluation marks a major breakthrough in drugs classification according to their water solubility and intestinal permeability[3,4]. However, the current

evaluation system of intestinal permeability has been neglecting the metabolic effects of approximately 10^{13} symbionts from more than 250 different species of bacteria as well as fungi, viruses and archaea[5–7]. The integral mapping of the interactions between drugs and bacteria indicated that bioaccumulation and metabolism of intestinal bacteria have greatly altered drugs bioavailability, with

[1]Institute of Clinical Pharmacology, Peking University First Hospital, 100191 Beijing, China. [2]State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Sciences, Peking University, 100191 Beijing, China. [3]Department of Pharmacy, Wenzhou Medical University, Wenzhou 325035, China. [4]Department of Gastrointestinal Surgery, Peking University First Hospital, 100034 Beijing, China. [5]Beijing Anzhen Hospital, Capital Medical University, 100029 Beijing, China. [6]Department of Pharmacy, Xuzhou Medical University, Xuzhou 221004, China. [7]Department of Pharmacy, Xiangya Hospital of Central South University, Changsha 410008, China. [8]These authors contributed equally: Yu-Hang Zhang, Chen-Shu Dai. ✉e-mail: yuhang@pkufh.cn; zhougan77@163.com; cui.pharm@pkufh.com

implications for their pharmacokinetics, side effects and clinical responses in an individual manner[8,9]. Therefore, how to convincingly integrate intestinal bacteria into the assessment of drug permeability warrants systematic investigations.

To better identify intestinal microbiota-mediated drug permeability, we developed the single-pass intestinal perfusion (SPIP)-based screening platform as a framework to characterize in situ P_{eff}-values of the selected drugs between GF and SPF rats. In our screen, we substantiated three drugs with the most significantly altered intestinal permeation, including N-Acetylcysteine (NAC), a derivative of the amino acid L-cysteine approved by Food and Drug Administration (FDA) as the powerful antioxidant to treat respiratory diseases for decades[10]. In 2021, NAC was even added as dietary supplementation to prevent serious liver damage from acetaminophen/alcohol poisoning due to its high aqueous solubility (163.9 mg/ml at pH 7.4)[10,11]. Nonetheless, FDA banned NAC in the United States for this supplementary use from 2023, due to its partially low bioavailability[12]. Once NAC enters into the bloodstream, it binds to plasmatic proteins and creates disulfide bridges through sulfhydryl, leading to the side effects of facial flushing, nausea and vomiting[10,13]. Most importantly, BCS I/III class of oral NAC has been inconclusive in pharmaceutical industry. Causes of the discordant permeability should be corroborated and development of NAC delivery would be the major concern in improving its bioavailability.

In this work, we unearth insights into the underlying mechanism of intestinal microbiota in NAC permeability by analyzing their host-microbiome interactions, which is essential to understanding the full scope of NAC's effects within the gastrointestinal environment. By identifying the specific microbial strains that are prominently altered by NAC perfusion, we target these strains to elucidate the mechanisms by which NAC influences gut microbiota composition and function. Then we comprehensively combine certain microbial strains and metabolites with potential clinical descriptors to train NAC permeability prediction model by nine machine learning (ML) algorithms. Ultimately, ML predicting accuracy is cross-wisely compared with the commercially used physiologically based gastrointestinal models, including advanced compartmental absorption and transit (ACAT) model[14,15], gastrointestinal transit and absorption (GITA) model[16], and advanced dissolution, absorption and metabolism (ADAM) model[17,18], to implement more optimum clinical predictions on NAC permeability.

Results
Differential screening for intestinal microbiota-mediated drug permeability
To establish the screening platform for 32 oral drugs with individual variable permeability which were potentially affected by intestinal microbiota, we reared male Wistar rats in the absence of intestinal microbial colonization (germ free, GF) and conventionally-colonized (specific pathogen free, SPF) controls, prior to single-pass intestinal perfusion for microbiome and metabolome analysis (Fig. 1a). In our screening process, we systematically profiled in vitro interactions between 29 oral drugs and 18 representative microbial strains. The resulting 60 potential bacteria-drug pairs were subjected to two independent screenings (Fig. 1b). A false discovery rate (FDR)-corrected P value cut-off of 0.05 and a threshold of 30% depletion were defined as indicative of a potential interaction. These targeted interactions were further investigated in bacterially colonized GF rats, revealing an in vivo network that spanned all tested strains and 91% of the tested drugs (29 of 32). Notably, multiple drugs, including NAC, showed typical interactions with Bacteroides species. Among 32 tested in vivo communities, we identified 12 drugs with Z factor > 0.5 in 27 individual pairs and 3 drugs with a Z factor >0.8 in 36 individual pairs (Fig. 1c). Differential intestinal P_{eff}-values between SPF and GF rats of three oral drugs with the highest Z factor were illustrated in Fig. 1d, including NAC, Lithium carbonate and Penicillin vk. For their absorption in SPF rats, three drugs were all determined with slow increase to

the P_{eff} plateau of 0.878×10^{-4}–1.073×10^{-4} cm/s within 2 h (Fig. 1d, red lines). In stark contrast to Lithium carbonate and Penicillin vk, only NAC was featured with sharp growth to the P_{eff} plateau of 1.446×10^{-4} cm/s in GF rats, with a two-fold increase than their SPF controls (Fig. 1d, top panel).

Intestinal microbiota has impaired NAC permeability by activating sphingolipids biosynthesis
To investigate the underlying mechanism of intestinal microbiota metabolism on NAC permeability, we sought to elucidate the metabolic alterations of intestinal fluid-derived metabolites between NAC-perfused GF and SPF rats. Metabolome datasets of the heat-map plots showcased the top 21 substantially discrepant intestinal metabolites (Fig. 2a). Thereinto, a series of intestinal sphingolipid biosynthesis molecules were derived from intestinal fluids of GF rats, such as higher palmitoyl-CoA, serine, gentamicin A, N-Oleoyl Glutamine and lower 3-ketosphinganine, sphingosine, 6-hydroxysphingosine than SPF rats (Fig. 2a, b). KEGG enrichment of these multiple differences were majorly concentrated on sphingolipids (SLs) metabolism (12.40%) and amino acid metabolism (22.32%) (Fig. 2c). Then we compartmentalized the microbial taxonomy of intestinal fluid samples derived from NAC-perfused and NAC-unperfused segments of SPF rats (Fig. 2d). By ultra-deep metagenomic sequencing, we applied LEfSe algorithm to explore the top enrichment of intestinal microbial communities, demonstrating the most significantly enriched bacteria in NAC-perfused segments were dominated by Bacteroides thetaiotaomicron (B. thetaiotaomicron) and Bacteroides fragilis (B. fragilis) relative to NAC-unperfused segments (Fig. 2e, f). Real-time PCR was performed to confirm B. thetaiotaomicron and B. fragilis as the significantly differential bacteria after NAC perfusion in SPF rats (Fig. 2g). Phylogenetically, the majority of bacterial species cannot produce sphingolipids. But Bacteroides, one of the most predominant commensal genera in the microbiomes, can produce and provide a source of ceramides with both odd (C17:0) and even (C18:0) numbers of hydrocarbons. This is identical to mammalian ceramides of inducing long-lasting albeit moderate deterioration in colonized mucosa to interfere with drug permeability[19,20].

In addition, we compared their intestinal microbial α diversity to reveal that ACE index, Chao1 index, Shannon index and Simpson index of NAC-perfused group were all tantamount to those of NAC-unperfused group (Supplementary Fig. 1a). β-diversity of the intestinal microbiota evaluated by ANOSIM was also non-differential between NAC-perfused group and NAC-unperfused group based on their PCoA scores (PC1: 12.12% and PC2: 8.38%, Supplementary Fig. 1b), NMDS stress score (0.1883, Supplementary Fig. 1c) or multi samples rarefaction curves (Supplementary Fig. 1d). The microbial co-occurrence network illustrating all intestinal microbiota correlated with Bacteroides in the NAC-perfused group is presented in Supplementary Fig. 1e. This network suggests that commensal intestinal bacteria, such as B. thetaiotaomicron and B. fragilis, play significant roles in stabilizing the composition of intestinal microbiota under the influence of NAC perfusion. Pearson correlation coefficient manifested significantly positive correlation between B. thetaiotaomicron/B. fragilis with 3-ketosphinganine and sphingosine, but negative correlation with palmitoyl-CoA and serine ($P < 0.01$; Fig. 2h).

Microbial serine palmitoyltransferase (SPT) catalyzed the covalent binding between NAC-derived cysteine and palmitoyl-CoA
As previously described, the first step of the sphingolipid biosynthesis is conserved from prokaryotes to eukaryotes, namely the condensation of serine and palmitoyl CoA[21]. This rate-limiting step is catalyzed by the pyridoxal-5′-phosphate (PLP) dependent enzyme−SPT (Supplementary Fig. 2a). However, the molecular basis for the microbial palmitoylation of NAC was still unrecognized, we thus analyzed the induced fit docking of palmitoyl-CoA with serine and cysteine in SPT, respectively. Structural comparison with the serine/cysteine-bound
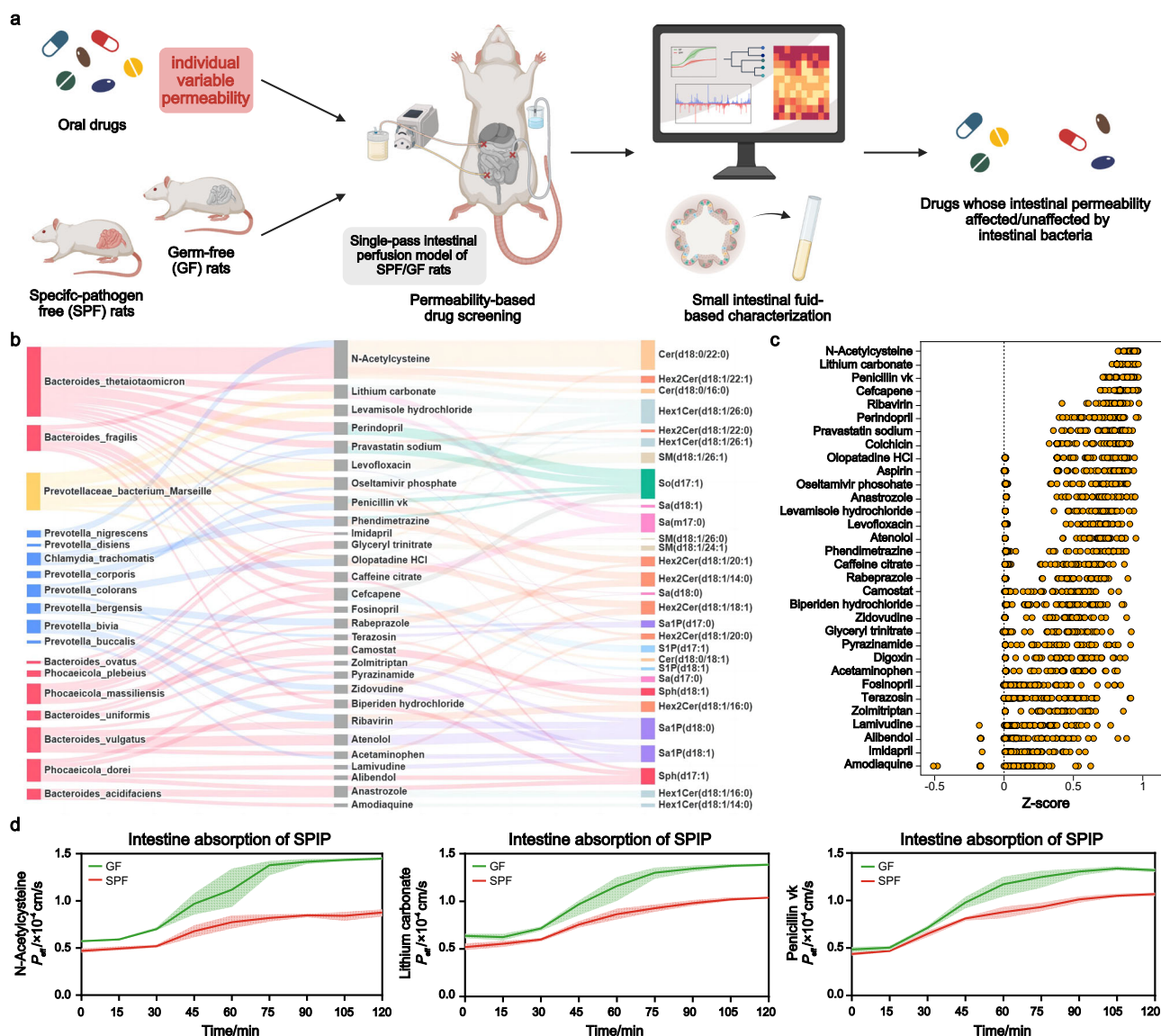
Fig. 1 | Intestinal microbiota-mediated drug permeability screening revealed the potential dependency on microbiota for N-acetylcysteine (NAC).
a Schematic workflow of the microbiota-mediated oral drug permeability screening platform. Created in BioRender. Dai, C. (2025) https://BioRender.com/r8tvfj3. For each drug with individual variable permeability, single-pass intestinal perfusion (SPIP) model was conducted in age-matched specific pathogen-free (SPF) and germ-free (GF) rats. $n = 6$ per group (the same applies hereinafter). After 2 h perfusion at 15-min intervals, the perfused and unperfused intestinal segments of rats were collected for microbiological detection, and the permeability of drugs was determined by SPIP. **b** Microbiota-drug-metabolite interaction network identified

in this study. Left network: Effects of intestinal bacteria on drug permeability. Significant interactions in two independent screenings ($n = 3$ per screen) were validated in a follow-up assay ($n = 3$; FDR-corrected $P < 0.05$) are shown (Spearman's rank tests). Right network: Differential intestinal metabolites between SPF and GF rats of drug exposure detected by two independent screenings (Spearman's rank tests). **c** Z factor of intestinal microbiota effect on the given drugs in each SPF-GF rats pair, $n = 6_{SPF} * 6_{GF} = 36$. The value of Z factor > 0.5 indicates the positive effect. **d** 2 h of intestinal P_{eff}-values for NAC (top panel), Lithium carbonate (middle panel) and Penicillin vk (bottom panel) were monitored at 15-min intervals. Data are the means ± SD. Source data are provided as a Source Data file.

SPT revealed similar recognition features as follows: sulfhydryl moiety of cysteine adopted nearly identical interactions with R123, Y401, L484 and A486, which correspond to serine-bound R123, Y401, E487 and A486 in SPT (Supplementary Fig. 2b). We thus proposed a "cysteinate sphingolipid metabolism path", which was different from the conventional de novo sphingolipid synthesis path derived from serine (Fig. 2i). By employing ultra-performance liquid chromatography-coupled time-of-flight mass spectrometry (UPLC-ESI-QTOF/MS) profiling, we further quantified the significantly increased palmitoyl-CoA but decreased cysteine-3-ketosphinganine in GF rats compared with SPF rats (Fig. 2j). These indicated oral NAC-derived cysteine could be involved into microbial SLs biosynthesis by covalently binding with palmitoyl-CoA, thus impairing the intestinal permeation of NAC.

Compared with wild-type intestinal *B. thetaiotaomicron* or *B. fragilis*, co-culture of SPT-deficient mutants heightened NAC permeation

Having shown in vivo permeation of NAC in close correlation with SLs biosynthesis of *B. thetaiotaomicron* and *B. fragilis*, we biosynthesized mutant bacterial strains with inactivated SPT (BTΔSPT and BFΔSPT) to impair their SL productions (Fig. 3a and Supplementary Fig. 3a). Excluding the intervention of SPT knockout on other metabolomes in *Bacteroides* (Supplementary Fig. 3b), we co-cultured differentiated Caco-2 cells with wild-type *B. thetaiotaomicron* (BTWT), *B. fragilis* (BFWT), BTΔSPT or BFΔSPT cells within the three-well transfer system. Then we established bioorthogonal chemistry by adding NAC and palmitoyl-CoA azide (PAA; a proxy for palmitoyl-CoA) through the
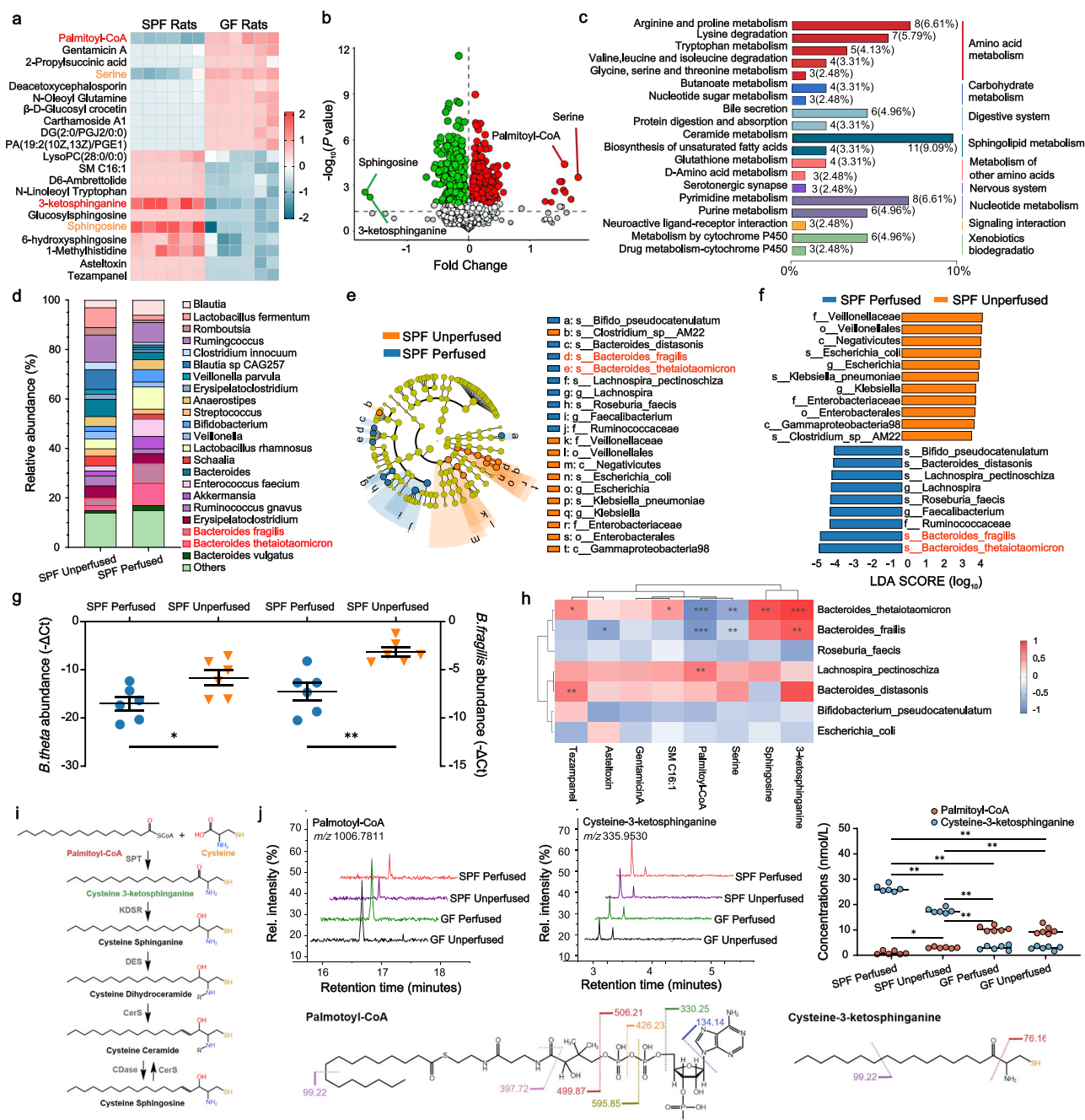
Fig. 2 | B. thetaiotaomicron or B. fragilis leveraged on sphingolipids biosynthesis to hinder NAC permeability. a Heat-map analysis of top 21 metabolites in intestinal fluid-derived metabolites between NAC-perfused GF and SPF rats. Each column represents one independent sample, and each row represents one metabolite. The color indicates the relative abundance of metabolites in each group. **b** Volcano plots to illustrate metabolites difference between NAC-perfused GF and SPF rats. Dots corresponding to significant lipids (*P* < 0.05, Student's *t* tests) were colored, in which lipids with increased fold change were colored as red, and sphingolipids with decreased fold change pertained to green. **c** Histogram presentation of the KEGG pathway. A total of 99 differentiated functional pathways were successfully annotated and grouped into 10 functional categories. *P* values were determined using two-sided Fisher's exact tests with Benjamini-Hochberg correction for multiple testing. **d** Summary of genus and species taxonomic changes in the gut microbiome of SPF rats following perfusion with 30 mg/L NAC. Taxonomic cladogram (**e**) and histogram (**f**) were generated by LEfSe of metagenomic analysis data in intestinal fluid samples derived from

NAC-perfused and NAC-unperfused segments of SPF rats. **g** RT-qPCR to determine the 16S rRNA of *B. thetaiotaomicron* (left coordinate) and *B. fragilis* (right coordinate) absolute abundance in each group. *n* = 6 per group. Data are the means ± *SD*, **P* < 0.05, ***P* < 0.01 (Student's *t* tests). **h** Correlation heat-map of microbial species abundance and metabolites after NAC perfusion in the intestine. Statistical significance was assessed using two-sided Pearson correlation analyses, with **P* < 0.05, ***P* < 0.01, ****P* < 0.001 indicating significant differences. **i** Cysteine-SL de novo synthesis pathway. SPT (serine palmitoyl transferase), KDSR (3-keto-dihydrosphignosine reductase), DES (desaturase), CerS (ceramide synthase), CDase (ceramidases). **j** Liquid chromatography-tandem mass spectrometry (LC-MS/MS) was utilized to identify and quantify palmitoyl-CoA (left panel) and cysteine-3-ketosphinganine (middle panel). The curve of left panel is representative image of each group. Data of right panel are presented as mean ± *SD*. Statistical significance was assessed using one-way ANOVA, with **P* < 0.05, ***P* < 0.01 indicating significant differences. Source data are provided as a Source Data file.
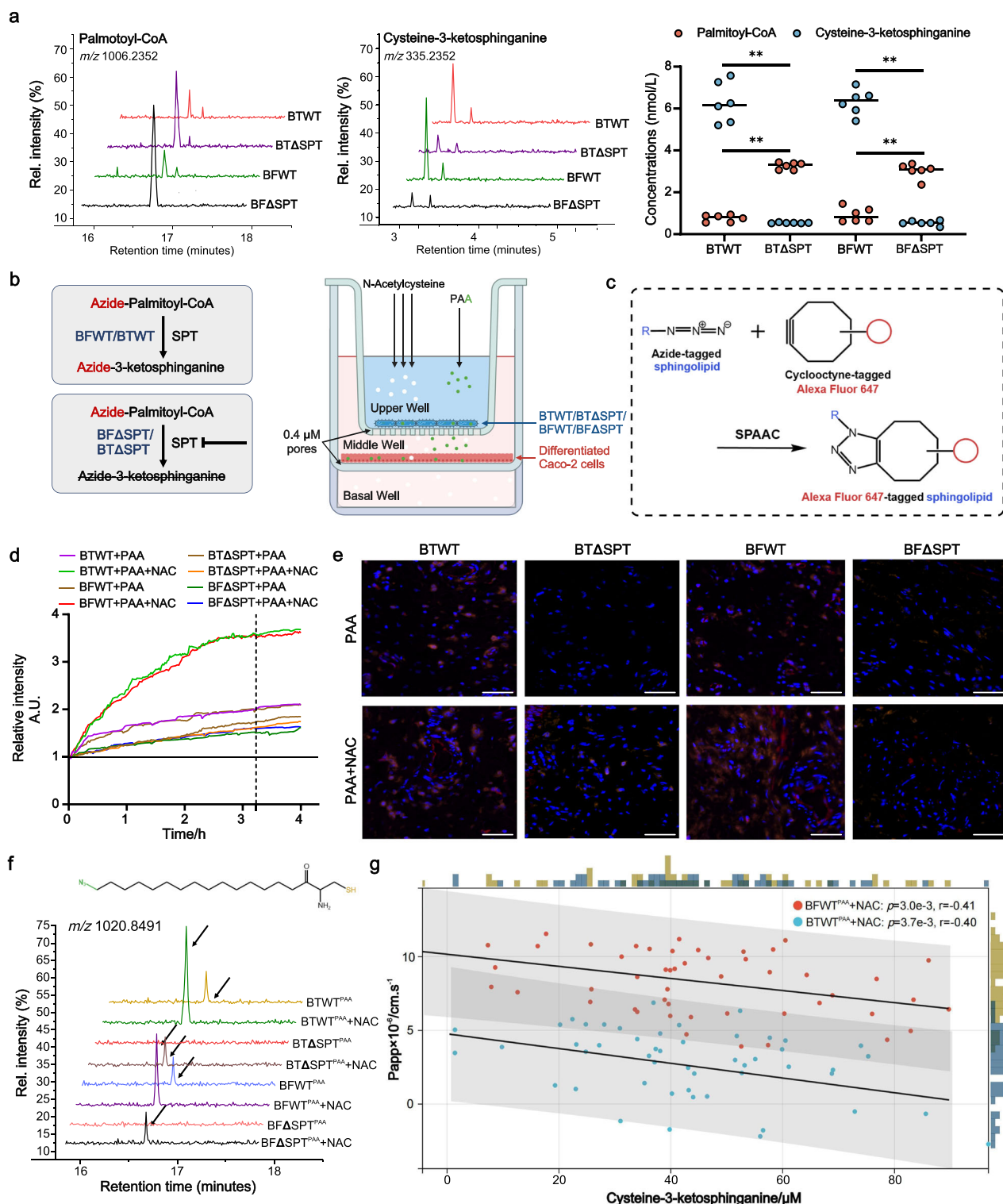
cycloaddition of cyclooctyne-conjugated detection reagent and Alexa Flour 647 cyclooccyne (AF647-Cyc), by which immunofluorescent intensity of each group was continually quantified (Fig. 3b, c). As shown in Fig. 3d, azide-labeled SLs of Caco-2 cells co-cultured with BTWT/BFWT cells and NAC underscored the rapid increase to the plateau at 3.227 h (light green and red lines). In contrast, azide-labeled SLs levels in BTΔSPT/BFΔSPT co-cultured group were determined with slow rise within 4 h (orange and blue lines). The fluorescent intensity suggested that PAA-derived SLs in BTWT/BFWT co-cultured groups were transferred into Caco-2 cells in the presence of NAC, but not

BTΔSPT/BFΔSPT co-cultured groups (Fig. 3e). To further elucidate the effects of bacterial SPT on NAC permeability, we added isotopically-labeled ^{15}N-NAC in the medium to monitor newly synthesized SLs (Supplementary Fig. 3c). At the end of the 120-min incubation, ^{15}N-cysteine-3-ketosphinganine, ^{15}N-cysteine--sphinganine, ^{15}N-cysteine-sphingosine and ^{15}N-cysteine-sphingosine-1-phosphate in SL pools were significantly down-regulated by SPT deficiency in co-culturing with both *B. thetaiotaomicron* and *B. fragilis* (Supplementary Fig. 3d–g). High-resolution MS2 spectra also revealed microbial SPT ablation has decreased the concentrations of azide-bearing cysteine-3-

Fig. 3 | Compared with wild-type intestinal *B. thetaiotaomicron* or *B. fragilis*, SPT deficiency enhanced NAC permeation. a LC-MS/MS to quantify palmitoyl-CoA and cysteine-3-ketosphinganine in the separated BTWT/BFWT or BTΔSPT/BFΔSPT treated by 25 μM palmitoyl-CoA and 10 μM NAC. Data are the means ± *SD*. One-way ANOVA: *$P < 0.05$, **$P < 0.01$. **b** Schematic representation of the three-well co-culture system. Created in BioRender. Dai, C. (2025) https://BioRender.com/r8tvfj3. NAC is transferred through BTWT/BFWT or BTΔSPT/BFΔSPT cells in the upper well to differentiated Caco-2 cells in the middle well, followed by infiltration into the basal well for concentration measurement. **c** The reaction of azide-tagged SLs with cyclooccyne-tagged Alexa Flour 647 by strain-promoted azide-alkyne cycloaddition (SPAAC) principle, which facilitates the visualization and quantification of sphingolipid uptake and transport across the intestinal epithelium.

d, **e** Exposed to 10 μM NAC, each group of azide-tagged lipids were continuously recorded with their relative immunofluorescent intensity changes (**d**). Representative images with the most significant inter-group differences of relative immunofluorescent intensity, selecting the dotted line of (**d**) at the time point of 3.227 h. Azide-tagged lipids were detected by Alexa Fluor 647 (red) and DNA was stained using DAPI (blue). Images are representative of three independent experiments, with the scale bar of 50 μm (**e**). $n = 6$ per group. **f** LC-MS/MS to quantify cysteine-3-ketosphinganine azide of the co-cocultured differentiated Caco-2 cells in each group. The curve is representative image of each group. **g** Correlation scatter plot of cysteine-3-ketosphinganine azide with *Papp* values in BTWT^PAA + NAC group and BFWT^PAA + NAC group (Pearson correlation analyses, $n = 50$). Source data are provided as a Source Data file.

ketosphinganine obtained from Caco-2 cells (Fig. 3f). Through Pearson correlation analysis, we found the negative linear dependence between azide-bearing cysteine-3-ketosphinganine and P_{eff}-values of NAC in both BTWT ($p = 3.0 \times e^{-3}$, $r = -0.41$) and BFWT ($p = 3.7 \times e^{-3}$, $r = -0.40$) co-cultured groups (Fig. 3g).

SL-production capacity of colonized *B. thetaiotaomicron* and *B. fragilis* affected intestinal absorption of NAC

To tightly control the SL-production of microbiome, we mono-colonized 6-week-old GF rats with BTWT/BFWT or BTΔSPT/BFΔSPT, previous to fat-free diet with only PAA (Fig. 4a). Assayed by SPIP, in situ P_{eff}-values of intestinal NAC showed a sharp increase to reach the P_{eff} plateau of 1.216×10^{-4} cm/s and 1.417×10^{-4} cm/s in GF^BTΔSPT and GF^BFΔSPT rats, respectively, both approximately with the two-fold increase than GF^BTWT and GF^BFWT littermates (Fig. 4b). We then determined the Alexa Fluor-647 azide (AF647-Az) intensity resulting from the increased alkyne-labeled SLs through copper-catalyzed cyclo-addition in the intestines of rats colonized with WT strains (Fig. 4c). In contrast, alkyne-labeled SLs levels in GF^BTΔSPT and GF^BFΔSPT rats were hardly detectable within 2 h, indicating that the knockdown of *SPT* in *Bacteroides* effectively blocked SLs synthesis (Fig. 4c). To further specifically identify these alkyne-bearing SLs, differential features were subjected to tandem mass spectrometry. High-resolution MS2 spectra quantified the dramatic decline in alkyne-bearing cysteine-3-ketosphinganine for NAC-perfused intestines of both GF^BTΔSPT and GF^BFΔSPT rats, as compared with their wild-type controls (Fig. 4d). Levels of two long-chain-base SLs synthesized by *B. thetaiotaomicron* and *B. fragilis*, cysteine-sphinganine (cysteine-Sa) (d18:0) and cysteine-sphingosine (cysteine-So) (d18:1) were significantly higher in the NAC-perfused intestines of GF^BTWT and GF^BFWT rats than GF^BTΔSPT and GF^BFΔSPT rats, respectively (Fig. 4e, Supplementary Figs. 4 and 5a, b). Cysteine-dihydroceramides (cysteine-DHCer), cysteine-ceramides (cysteine-Cer) and cysteine-sphingomyelin (cysteine-SM) with acyl chains of C16:0 and C24:0 were elevated in GF^BTWT and GF^BFWT rats (Fig. 4f–h, Supplementary Figs. 4 and 5c–h). In contrast, cysteine-SLs with acyl chains of C18:0 and C24:1 were all similar across the groups. These findings thus demonstrated that SL-production capacity of intestinal microbes affected levels of NAC absorption in the intestine.

Machine learning models for clinically predicting NAC permeability

In an attempt to comprehensively verify the above microbial determinants of NAC permeability in rats, we subsequently enrolled the prospective clinical cohort of 240 healthy volunteers (aged between 22 and 69 years, with body mass index of 22.1–24.8 kg/m^2) to exclude participants with clinically detectable gastrointestinal conditions or diseases. More participants' information was elaborated in Supplementary Data 1. To develop the machine learning model of incorporating intestinal microbial features, we initially included experimentally confirmed microbial strains and SL metabolites into the multi-parameter models (Supplementary Fig. 6). All participants of inputting 19 features to predict fraction of dose entering the systemic circulation (*Fsys*) of NAC,

including available age, prescription, NAC dose, intestinal blood flow, intestinal content viscosity, intestinal pH, intestinal transit rate, two microbial strains and ten SL metabolites, were randomly divided into training set for feature selection and hyperparameter optimization ($n = 192$ participants, 80%; Fig. 5a), and test set for evaluation ($n = 48$ participants, 20%; Fig. 5a). As the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) guideline on *M9: Biopharmaceutics Classification System-Based Biowaivers* statement that drugs with *Fsys* higher than 85% are considered to be highly permeable[3,4], feature values of the participants in test set were ranked in order according to their true *Fsys* values from high to low. We divided the test set into high (left panel) and low (right panel) permeability groups using *Fsys* = 85% as the boundary (Fig. 5b).

We then trained and evaluated nine different ML regression algorithms, including Multiple Linear Regression (MLR), Neural Network (NN), k-Nearest Neighbors (k-NN), Random Forest (RF), Decision Tree (DT), Partial Least Squares (PLS), Support Vector Regressor (SVR), eXtreme Gradient Boosting (XGB) and Light Gradient Boosting Machine (LGBM) using cross-validation strategy. The feature selection (SelectKBest) and hyperparameter optimization procedure (Randomized Search cross-validation and Grid Search cross-validation) were performed by fivefold cross-validation with training set. The selected features for all models, changes in model accuracy after each feature deletion and range of all considered hyperparameters, along with their final selected values, are presented in Supplementary Tables 4, 5 and Supplementary Data 3. Whereafter, the models with these fixed features and hyperparameter structures were assessed by mean absolute error (MAE) of test set (Fig. 5c). Thereinto, 15-feature XGB model manifested the best predictive performance with a MAE value of 2.453 (±2.011) and the least median in comparison with 13-feature LGBM, 7-feature k-NN and 7-feature SVR models of similar MAE values (Fig. 5c). To further verify the prediction accuracy of these four algorithms, we compared predicted values with true values surpassing *Fsys* = 85% as the threshold for high permeability. Receiver operating characteristic (ROC) curve analysis demonstrated 15-feature XGB model as the most superior predictive performance with the highest AUC value of 0.867 (Fig. 5d).

The contribution of each input feature in predicting *Fsys* values was then visualized by Shapley additive explanation (SHAP) analysis of the established 15-feature XGB model. The swarm plot revealed the overall contribution of input features in terms of SHAP values (Fig. 5e). Cysteine-3-ketosphinganine is listed as the top determinant, indicating its greatest impact on *Fsys* predictions, and higher Cysteine-3-ketosphinganine was accompanied with lower predicted *Fsys* values. *B. thetaiotaomicron* was observed with similar contributions to *Fsys*. On the contrary, higher levels of Palmitoyl-CoA, intestinal blood flow and transit rate predicted higher *Fsys*. Then decision path function was applied to better visualize the collective effect of the 13 features on model output, which also identified intestinal microbial features as the top three pivotal features (Fig. 5f). In agreement with the swarm plot of ranked importance, hierarchical clustering demonstrated that those with combinations of lower levels of Cysteine-3-ketosphinganine, *B. thetaiotaomicron*, and higher levels of Palmitoyl-CoA, intestinal blood
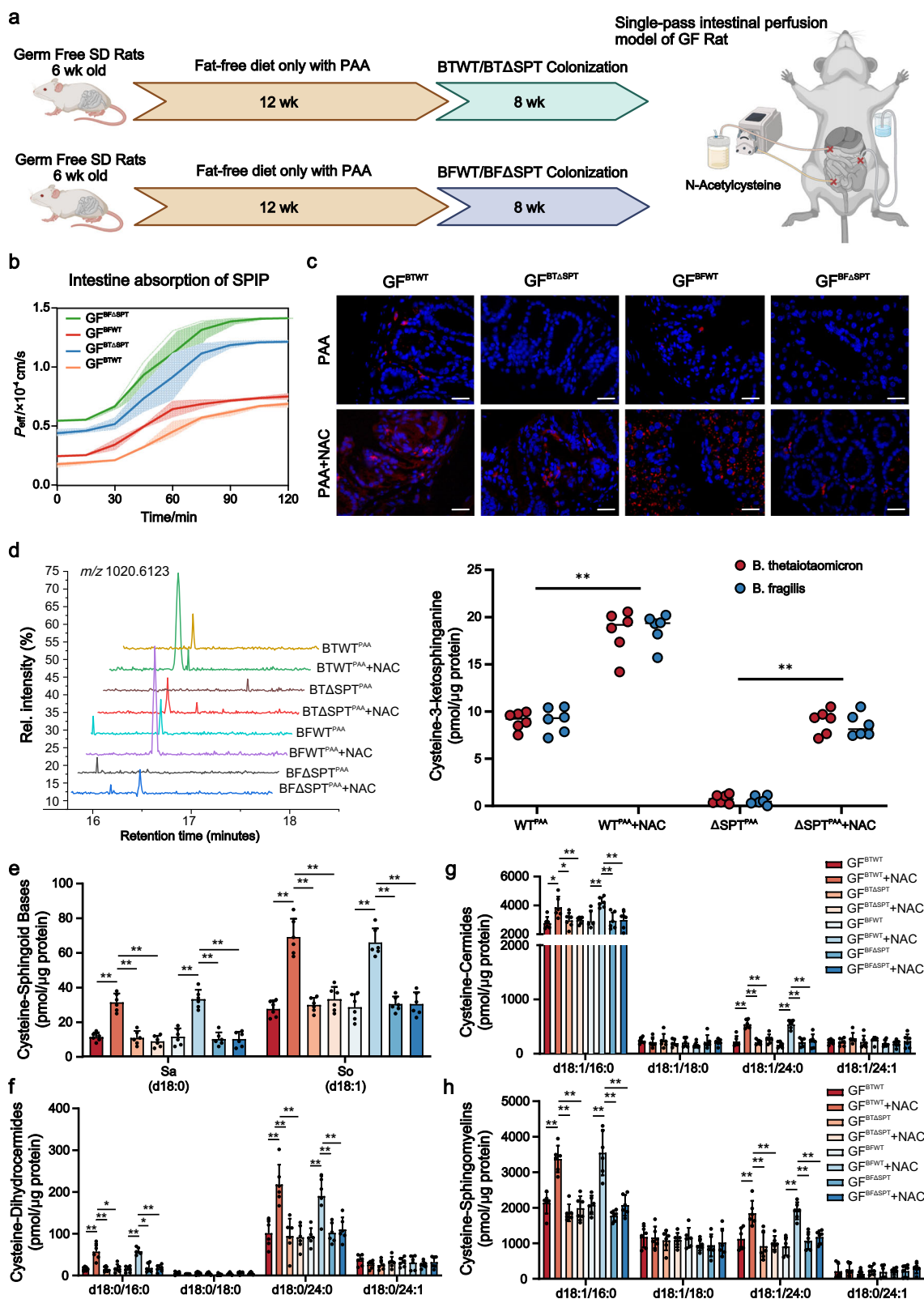
flow, intestinal transit rate tended to comprise the population subsets with increasingly predicted chance of high *Fsys* (Fig. 5g, red labels).

Cross comparison of multi-parameter XGB model to predict NAC permeability classification

ACAT model, ADAM model and GITA model are acknowledged as transit compartment models for commercially predicting oral drug absorption. These mechanism-based models have taken account of physicochemical properties (e.g., drug solubility, pKa), dosage form (e.g., dose, type of dosage form) and physiological parameters (e.g., intestinal transit time, intestinal pH)[22], but none of them were capable of quantitatively incorporating intestinal microbial characteristics into the model construction. Therefore, the above ML approach (the selected 15-feature XGB model) was introduced to compensate for this

Fig. 4 | SL-production capacity of colonized *B. thetaiotaomicron* and *B. fragilis* affected intestinal absorption of NAC. a Experimental setting: age-matched GF rats were simultaneously fed with 12 weeks of fat-free diet with PAA (palmitoyl-CoA alkyne) before 8 weeks of BTWT/BFWT or BTΔSPT/BFΔSPT colonization. Then SPIP model of 30 mg/L NAC was conducted in each group. Created in BioRender. Dai, C. (2025) https://BioRender.com/r8tvfj3. **b** Monitoring of intestinal P_{eff}-values over a 2-h period for each group, with measurements taken at 15-min intervals. Data are the means ± SD. **c** Intestinal tissue of GF rats inoculated with BTWT/BFWT or BTΔSPT/BFΔSPT grown in PAA. PAA-based metabolites were detected with Alexa Fluor 647 azide (red) using click chemistry, and nuclei of the intestinal epithelial cells were stained using DAPI (blue). Scale bar is 20 μm. Each experiment was repeated 6 times independently. The curve of left panel is representative image of each group. **d** LC-MS/MS to quantify cysteine-3-ketosphinganine alkyne across the groups. Data are the means ± SD. One-way ANOVA: *$P < 0.05$, **$P < 0.01$. Cysteine-sphinganine (cysteine-Sa) (d18:0) and cysteine-sphingosine (cysteine-So) (d18:1) (**e**), cysteine-dihydroceramides (cysteine-DHCer) (**f**), cysteine-ceramides (cysteine-Cer) (**g**) and cysteine-sphingomyelin (cysteine-SM) (**h**) in the NAC-perfused intestines of each group. Bar charts represent SL abundance ± SD for 6 rats per condition (two-way ANOVA, Tukey's multiple comparison test, *$P < 0.05$, **$P < 0.01$). Source data are provided as Supplementary Fig. 5 and a Source Data file.

deficiency. We enrolled the other prospective cohort of 79 healthy participants (Supplementary Data 2) to calculate their predicting F_{sys} values by ACAT model, ADAM model, GITA model (Supplementary Fig. 7) and XGB model. All input parameters to depict NAC permeability simulations and simulated pharmacokinetic parameters were detailed in Supplementary Tables 6, 7. Error scatter density plots have illustrated more adjacent distribution of XGB model scatters to the regression line of "Predictive value = True value" than other three models (Fig. 6a). Further comparison of absolute error (AE) values confirmed the most accurate predictive performance of XGB model ($P < 0.01$, Fig. 6b). Subsequently, we divided the 79 participants into high ($F_{sys} \geq 85\%$) and low ($F_{sys} < 85\%$) permeability cohorts based on true/predicted values, as shown in the confusion matrices of four models (Fig. 6c). The top block means low-permeability samples that were correctly classified (left block) or misclassified (right block), and the bottom block indicated other high-permeability cases that were correctly classified (right block) or misclassified (left block). The mean accuracy, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of XGB model were 77.2%, 77.8%, 76.7%, 73.7% and 80.5%, respectively. Its average performance was much superior to those of ACAT (57.0%, 36.1%, 74.4%, 54.2% and 58.2%), GITA (77.2%, 100.0%, 58.1%, 66.7% and 100.0%) and ADAM (73.4%, 100.0%, 51.2%, 63.2% and 100.0%) models (Fig. 6c). To quantify whether XGB model could provide the clinically relevant enhancement in predictability, we further introduced the net reclassification improvement (NRI) index (Fig. 6d), which revealed that XGB model has prominently refined the performance of F_{sys} prediction as compared with ACAT model (NRI = 0.551, $P < 0.001$), but its improvements over GITA (NRI = 0.075, $P = 0.464$) and ADAM models (NRI = 0.168, $P = 0.135$) were negligible. Taken together, these findings demonstrated the predictive performance of XGB model that included intestinal microbial features outperformed other three models in NAC permeability prediction.

Discussion

The human intestinal microbiota is a complex ecosystem due to its important role in host digestion, especially for oral drugs ingested through microbial-secreted enzymes[23,24]. Different from the microbiome in colorectum, existing studies have proved much lower diversity and abundance of microflora in the small intestine, which mainly covered microbial enzymes such as β-glucosidase, β-glycosylated enzymes, nitrogen reductase, sulfate ester enzyme, nitro reduction enzyme, nitrate reductase and palmitoyltransferases[25–27]. These variable enzymes for biotransformation have directly or indirectly influenced the oral bioavailability of drugs. To take full advantage of these effects, certain intestinal flora metabolites could be identified as determinants to improve the evaluation model of drug permeation. Specific strategies depend on the fact whether we have identified certain metabolites that are biologically active in promoting drug permeability, then we can combine these metabolites with pharmaceutical properties to construct intestinal microbiota-based permeability model for further clinical application.

In this study, we have screened 32 oral drugs with individual variable permeability and pinpointed that SL-production capacity of

intestinal *B. thetaiotaomicron* and *B. fragilis* affected intestinal absorption of NAC, which attributed to SPT-catalyzed cysteine palmitoylation. Unlike eukaryotes consisting of the heterodimer of two highly related subunits SPT1 and SPT2 (encoded by the LCB1 and LCB2 genes), *Raman* et al. have thoroughly analyzed the homodimeric bacterial SPT from *Sphingomonas paucimobilis* and reported the first high-resolution X-ray crystal structure of *S. paucimobilis* holo-SPT[28]. This structure suggested that the active site containing the PLP cofactor is located at the dimeric interface[29]. NAC contains both carboxyl terminal and N-terminal cysteine residue, however, protein palmitoylation is mainly catalyzed by S-acyltransferase DHHC (Asp-His-His-Cys) family which only expressed in eukaryotes[30–32]. We thus narrowed the focus of SPT deficiency for NAC permeability, which confirmed the negative linear dependence of P_{eff}-values of NAC on cysteine-3-ketosphinganine by covalently binding with palmitoyl-CoA. While this study provides valuable insights into the role of gut microbiota in NAC permeability using rat models, it is important to recognize the compositional differences between rat and human gut microbiomes. Notably, *Prevotella copri*, a prominent human gut commensal bacterium known to produce sphingolipids, is absent in rat microbiomes[33]. The inclusion of such species in future studies could offer additional insights into the mechanisms of NAC permeability in humans. Our screening model, therefore, serves as a foundation that can be further refined through systematic research involving fecal microbiota transplantation (FMT) from diverse human donors, which would help to more accurately model the human gut environment and refine our predictions for clinical applications. Besides, none of SPT inhibitors have been identified to efficiently inhibit microbial SPT[34,35]. Specific compounds screening and structural modification of microbial SPT inhibitors warrant further investigation to benefit more microbial sphingolipid-related treatments.

As widely known, NAC has the variable oral bio-availability but high aqueous solubility of 163.9 mg/ml[36]. After an oral dose of NAC 400 mg, the C_{max} of the reduced form was 3.47 mg/L with a time to t_{max} of 30 min[37]. Bio-availability was 4.0% for the reduced form and 9.1% for total drug. This lower bio-availability of reduced NAC compared with total drug indicated that NAC was hindered to be permeated before it reached the general circulation. Besides, large within-study coefficients of variability (CVs) of 35-45% have been reported in different populations[38,39]. For instance, the C_{max} and the AUC of the reference product (NAC tablets 200 mg) were respectively 32.1% and 26.9% lower in NAC low-bioavailability group, as compared with high-bioavailability group. Herein, we recruited the prospective cohort study of 240 participants by encompassing intestinal microbial features into the input feature profile and obtained the most interpretative XGB model for predicting NAC F_{sys} from ML training. As a data-driven approach, ML models trained with insufficient sample sizes often suffer from the problem of "overfitting", in which the models rely excessively on features derived from the under-represented training data and thus lose the ability to perform effectively in practice[40,41]. Nevertheless, ensemble methods could constitute an effective approach to mitigate overfitting in machine learning frameworks, such as XGB and LGBM algorithms. They aggregate predictions from multiple base models, enabling the ensemble to capture diverse predictive patterns and diminish individual model biases. Besides, XGB
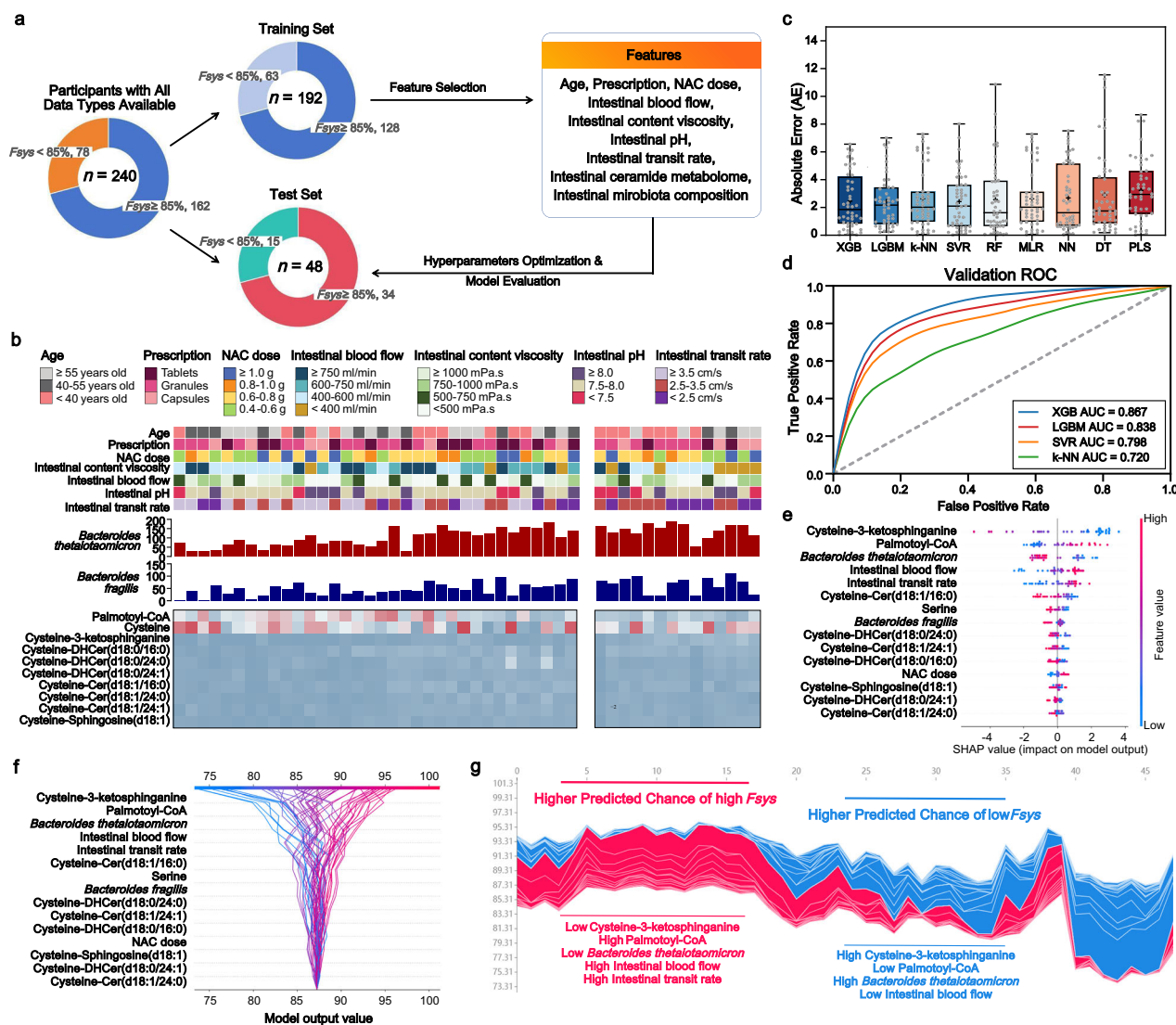
Fig. 5 | Machine learning (ML) model generated for clinically predicting NAC permeability. a A total of 240 healthy participants were enrolled for the training and test of ML models, with 192 (80%) participants as training set for feature selection and hyperparameter optimization, and 48 (20%) participants as test set for evaluation. **b** Participant characteristics of test set. Each column represents an individual participant with the potential determinants correlated with NAC absorption. Participants were ranked by true values of NAC *Fsys* from high to low (left to right) and divided by *Fsys* of 85% for biopharmaceutics classification. **c** Box-whisker plot summaries the overall predictive performance of various ML models. The data represent the absolute error (AE) of fraction of dose entering the systemic circulation (*Fsys*) predictions obtained in test set (*n* = 48). The mean absolute error (MAE) and median AE of each model are displayed in the boxes as black "+" and black dashed lines, respectively. The first and third quartiles are shown by the upper and lower edges of the respective boxes, with the minima and maxima by the upper and lower solid lines. XGB eXtreme Gradient Boosting, LGBM Light Gradient Boosting Machine, k-NN k-Nearest

Neighbors, SVR Support Vector Regressor, RF Random Forest, MLR Multiple Linear Regression, NN Neural Network, DT Decision Tree, PLS Partial Least Squares. **d** Receiver operating curve (ROC) for predicting NAC *Fsys* by XGB, LGBM, SVR and k-NN, with the area under the curve (AUC) values of 0.867, 0.838, 0.798 and 0.720, respectively. **e** Swarm plot illustrates the impact of each feature on *Fsys* according to their SHapley Additive exPlanation (SHAP) values. The color of the dots denotes the relative value of the feature within dataset (high-to-low depicted as pink-to-blue). The horizontal position of each dot represents whether the effect of each feature value contributes positively or negatively to the prediction instance. **f** Decision path taken for each *Fsys* prediction, illustrating how the XGB model combines the relative contribution of each feature to predict *Fsys*. **g** Rotate the force plot of all participants in the cohort 90° counterclockwise and obtain a global picture of the NAC *Fsys* prediction, clustered by similar risk factor combinations. Common characteristics of subpopulations of participants are referred to high (*red*) or low (*blue*) prediction probabilities. Source data are provided as a Source Data file.

algorithm selected in this study was less likely to overfit the data due to the maximum depth limit setting based on leaf-wise strategy. Within the cohort of 79 participants, we confirmed the 15-feature XGB model that included intestinal microbial features with superior performance than ACAT, GITA, ADAM models in predicting NAC permeability. Given the challenges in precisely identifying the origin of the collected intestinal fluid from either the duodenum or jejunum due to the oral catheterization, we characterized the fluid based on limited intestinal features, which is a recognized limitation of our study. Furthermore, the application of our model is contingent upon several conditions, including the

absence of disease, which can significantly alter gut microbiota and drug absorption. We also assumed normal transit times for the oral drugs and a standard diet to control for dietary variations. However, these assumptions limit the model's applicability. Future models should aim to incorporate a wider range of factors that affect drug absorption, such as drug combinations, physiological processes, and internal environment conditions, to enhance their predictive accuracy in diverse clinical scenarios.

The enhanced predictive accuracy of the XGB model, which includes characteristics of the gut microbiota, underscores the
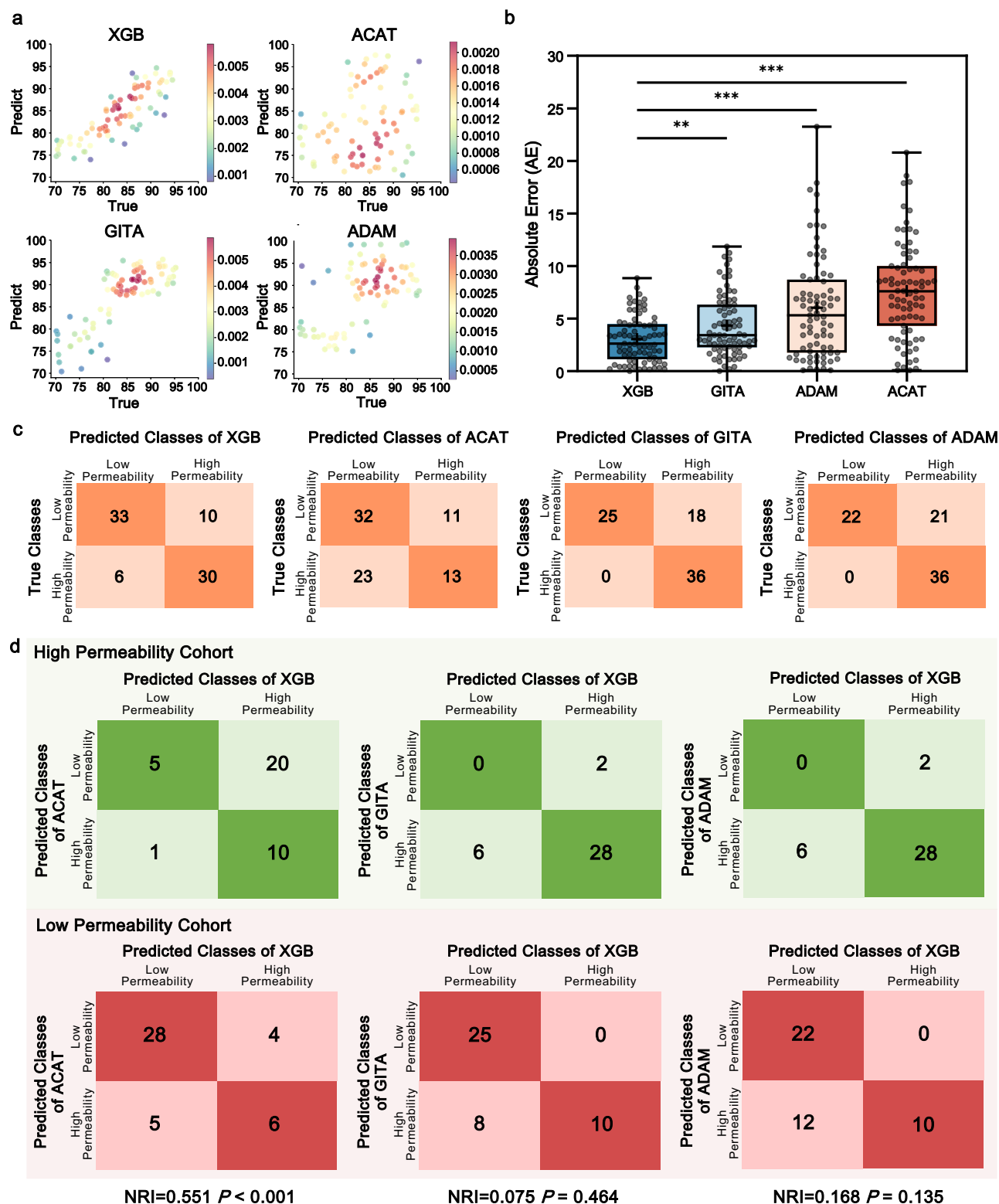
significant role of gut microbes in drug absorption and the potential for refining BCS assessment systems. Given the multitude of factors that influence drug absorption, such as diet, drug interactions, and physiological conditions including digestive juice composition, gastrointestinal blood flow, and disease states[42], there is a clear need for more sophisticated models. Machine learning offers a promising avenue for developing such comprehensive models in pharmaceutical research.

Methods

Animal experiments

All procedures involving animals were carried out according to the protocols approved by Peking University First Hospital Experimental Animal Center (Ethical No. 2023-59332). Wistar GF or SPF rats (6-week-old, 250–300 g) were housed in rigid auto-claved cages as 2 per cage by the 12-h light-dark cycle. For in situ permeability experiments in Fig. 1, the rats ($n = 6$ each group) were fasted for 12 h with free access to

Fig. 6 | The *Fsys* predictive performance of the XGB model outperformed advanced compartmental absorption and transit (ACAT), gastrointestinal transit and absorption (GITA) and advanced dissolution, absorption and metabolism (ADAM) models. a Error scatter density plots to evaluate the predicted *Fsys* with true *Fsys* of XGB, ACAT, GITA and ADAM models. The dots color is determined by the kernel density estimation values (low-to-high correlation pertains to blue-to-red color). **b** Box-whisker plot summaries the predictive performance of the XGB, ACAT, GITA and ADAM models, which represent the AE of *Fsys* predictions of 79 participants. The MAE and median AE of each model were displayed in the boxes as black "+" and black dashed lines, respectively. The first and third quartiles are shown by the upper and lower edges of the respective boxes, with the minima and maxima by the upper and lower solid lines. One-way ANOVA:

$**P < 0.01$, $***P < 0.001$. **c** Confusion matrices of the dataset by XGB, ACAT, GITA and ADAM models. The numbers in each colored box represent the number of instances between the true and predicted classes obtained from the models. **d** Confusion matrices of predicted results comparing XGB models with ACAT, GITA and ADAM models, respectively. The numbers in each colored box indicate the instances between the predicted classes of XGB model and the other model. The green confusion matrices are derived from the high permeability cohorts (denoted as a1, b1, c1 and d1 from upper left to lower right, respectively), and the red ones derived from the low permeability cohorts (denoted as a2, b2, c2 and d2 from upper left to lower right, respectively). $N1 = a1 + b1 + c1 + d1$; $N2 = a2 + b2 + c2 + d2$. $NRI = ($c1-b1$)/N1 + ($b2-c2$)/N2$. $Z = \frac{NRI}{\sqrt{\frac{b1+c1}{N1^2} + \frac{b2+c2}{N2^2}}}$, $P = (1-Z)*2$. Source data are provided as a Source Data file.

water before anesthetization by an intramuscular injection of ketamine-xylazine mixture (80 mg/kg and 20 mg/kg, details in Supplementary Table 1) and placed on the heated surface maintaining with 37 ± 1 °C. The laparotomy was implemented through a midline incision of 3-4 cm to expose the intestine and approximately 10 cm of the proximal jejunum portion was cannulated at both ends. The blank perfusion solution at 37 °C was pumped by peristaltic pump through the intestine at a flow rate of 0.5 mL/min for approximately 30 min. Then, a perfusion solution containing 30 mg/L NAC was administered at 37 °C through the intestinal lumen at a constant flow rate of 0.2 mL/min. To ascertain the steady state during the perfusion process, perfusate samples were collected from the distal portion of the jejunum at 15-min intervals (at 15, 30, 45, 60, 75, 90, 105, and 120 min). These samples were collected in pre-weighed vials to monitor when the effective permeability (P_{eff}) values at each point of the perfused segment remained constant, indicating that the system had reached a steady state. The samples were immediately frozen at -20 °C until subsequent analysis by LC-MS/MS. P_{eff} (cm/s) was calculated by the following equation:

$$P_{eff} = -\frac{Q_{in}}{A} \times \ln \frac{C_{out} \times Q_{out}}{C_{in} \times Q_{in}} \qquad (1)$$

wherein, Q_{in} = inlet perfusate flux (cm³/s), A ($2\pi rl$) = absorbable intestinal area considering radius and segment length (cm²), C_{out} = outlet drug concentration (mg/mL), Q_{out} = outlet perfusate flux (cm³/s), C_{in} = inlet drug concentration (mg/mL).

Upon sacrifice, the perfused and unperfused intestinal segments were removed to assess microbiomic changes. Each intestinal segment was perfused with 37 °C saline at a rate of 0.2 mL/min for 10 min. The intestinal fluids collected from these segments were then used for subsequent microbiological analysis (refer to 'Metabolome Analysis of Intestinal Microbiota' and 'Taxonomic Profiles of Intestinal Microbiota' sections). We uniformly sampled intestinal fluid at the endpoint of perfusion, which was 120 min, to capture the changes in microbiome composition. The network was constructed by correlating the abundance of symbiotic bacteria (or microbial metabolites) in the intestine with P_{eff} during the drug perfusion steady state. For each bacterium-drug interaction, triplicates were analyzed, along with single bacteria-free controls for each drug. Subsequently, SPIP experiments were conducted for each drug in bacterially colonized GF rats (see below for colonization method). Quantitative PCR (qPCR) was used to validate bacterial abundances (Supplementary Table 2), and LC-MS was employed to quantify metabolites (Supplementary Table 3), thereby verifying the initial screening results. Statistical significance was assessed using Wilcoxon's rank sum test with a *P* value cut-off of 0.05 after FDR correction. For in vivo experiments of *B. thetaiotaomicron* and *B. fragilis* strains colonization in Fig. 4, all rats were first fed with fat-free diet (0% kcal fat, 76% kcal carbohydrate, 24% kcal protein) but only remaining PAA for 12 weeks. BTWT/BFWT or BTΔSPT/BFΔSPT strains (10^8 CFU/μL of PBS per rat three times a week) were subsequently added to the diet that eliminated exogenous lipid interference

beyond PAA for 8 weeks prior to the in situ permeability experiments described above. Following the in vivo click-chemistry experiments, rats were euthanized, and both perfused and unperfused intestinal segments were directly embedded in O.C.T compound (Sakura Finetek, Tokyo, Japan) and placed on ice. After one hour of stabilization on ice, the samples embedded in O.C.T compound were rapidly frozen in 2-methylbutane chilled with liquid nitrogen. These frozen blocks were then wrapped in labeled foil and stored at -80 °C until further use.

Human intestinal fluids collection
The Ethics Committees in Peking University First Hospital approved the study protocols (Ethical number: 2023-166). Written informed consent were obtained from all participants in this study. We obtained intestinal fluid samples from 319 healthy participants aged 22–69 years after an overnight fast, by simultaneous duodenal and jejunal aspiration with two double-lumen catheters. The catheters were introduced orally and positioned such that the proximal site was located in the duodenum (5–10 cm from the pylorus) and the distal site 90 cm below the duodenal site. The position of the catheters was confirmed by fluoroscopy. All the healthy participants' information could be found in Supplementary Data 1, 2.

Metabolome analysis of intestinal microbiota
For intestinal microflora metabolites extraction, all samples were transferred into EP tubes three times with 1000 μL extract containing internal target (methanol acetonitrile volume ratio =1:1, internal standard concentration 20 mg/L), and vortex mixed for 30 s. Add steel ball for 45 Hz grinding instrument for 10 min and ultrasonic 10 min (ice bath). Then samples were centrifuged at $200 \times g$, 4 °C for 15 min and remove 500 μL supernatant into EP tube. The extract is dried in a vacuum concentrator and 160 μL extract (acetonitrile-water ratio: 1:1) was added to the dried metabolites for resolution. Vortex 30 s, ice water bath ultrasonic 10 min and centrifuged at $200 \times g$, 4 °C for 15 min. 120 μL supernatant of each sample was mixed into QC sample for LC/MS detection.

We applied Waters Xevo G2-XS QTOF high resolution mass spectrometer to collect primary and secondary MS data in MSe mode under the control of MassLynx V4.2 (Waters). The low collision energy is 2 V, the high collision energy range is 10-40 V, and the scanning frequency is 0.2 s. The parameters of the ESI ion source are as follows: Capillary voltage: 2000V (positive ion mode) or –1500V (negative ion mode), cone voltage: 30 V, ion source temperature: 150 °C, dissolvent gas temperature 500 °C, backflush gas flow rate: 50 L/h, dissolvent gas flow rate: 800 L/h. The raw data collected using MassLynx V4.2 is processed by Progenesis QI software for peak extraction, peak alignment and other data processing operations, based on the Progenesis QI software online METLIN database (https://metlin.scripps.edu). The identified compounds are searched for classification and pathway information in KEGG, HMDB and Lipid Maps Databases (https://lipidmaps.org/databases). The screening criteria were Fold Change >1, *P* value < 0.05 and VIP > 1.

Taxonomic profiles of intestinal microbiota

For 16S rRNA extraction, all samples were centrifuged for 10 min at $1500 \times g$ using an ALC PJ180R centrifuge (Winchester, VA, USA). For the generation of paired-end clusters, the V4 kit was utilized across 8 lanes on an Illumina cBot and subsequently applied for sequencing on an Illumina HiSeq 2500 platform. The median cluster density was determined to be 908.5 for Nextera XT. The open-source software package DADA2 (version 1.10) was employed for amplicon data processing, enabling single-nucleotide resolution of amplicons. The forward and reverse reads were truncated to 200 and 150 bases in length, respectively. Read pairs were excluded if they contained ambiguous bases with expected errors higher than two or originated from PhiX spike-in controls. One million reads from each sequencing run were used to infer error profiles. Subsequent steps included replication, error correction, and merging of forward and reverse reads. For assignment of taxonomic annotation, a Bayesian classifier and the Ribosomal Database Project training set (v.16) were used. The functional genomic potential of the intestinal microbiota was predicted with PICRUSt2 using the standard workflow. All samples were normalized to 10,000 16S rRNA gene read counts for analysis. Amplicon sequence variants (ASVs) with an average relative abundance greater than 0.1% within the microbiome were selected for correlation analysis of their occurrence patterns. The SparCC algorithm was employed to estimate the correlations between intestinal microbes. Pseudo P values were calculated using 1000 bootstrap replicates. Correlations with an absolute value of the correlation coefficient (r) greater than 0.2 and a P value less than 0.01 were considered statistically significant. For each genus with a significant SparCC correlation, its degree was calculated as an indicator of its weight within the network by summing up its edges. LEfSe scores measure the consistency of differences in relative abundance between taxa, with a higher score indicating higher consistency. We considered taxa with linear discriminant analysis score >2 and $P < 0.05$ to be significant.

In silico molecular docking and microbial SPT activity measurement

SPT structure (PDB code 4A5J) for the in silico fit docking procedure was prepared using Maestro Molecular Modeling platform (version 12.8) with default parameters. Induced fit docking accounts for accurate side-chain and backbone flexibility in the ligand-receptor docking was predicted by Glide software, with refinement for residues within 5 Å at the active site by Prime. All SPT substrates were obtained from PubChem database (https://pubchem.ncbi.nlm.nih.gov/). Each ligand was subjected to LigPrep with the OPLS3e30 force field. Mn^{2+} was chosen for the center of the box, and the size of box was 20 Å. While performing induced fit docking, receptor and ligand *van der Waals* scaling was maintained to 0.5–0.7 with the maximum of 20 poses.

Microbial strains and gene-editing strategies

Isolated from the intestinal fluids of enrolled participants, strains of *B. thetaiotaomicron* and *B. fragilis* were identified by contrasting the 16S rRNA gene sequence with those in the NCBI reference database (https://www.ncbi.nlm.nih.gov/refseq). The DNA fragments encrypting full-length *SPT* gene were cloned into the pET28a vector with 6× His tag at the N-terminal end adopting standard molecular cloning procedures. *SPT* was overexpressed in *E. coli* Rosetta (DE3), which were cultured to an OD_{600} of 0.6 at 37 °C (Supplementary Fig. 2c). An internal fragment (610 bp) of the *SPT* gene from *B. thetaiotaomicron* and *B. fragilis* was cloned into the pGERM suicide vector containing selective markers of *E. coli* (*bla*). The constructed vector was subsequently transformed into the conjugative *E. coli* S17 strain. Then *E. coli* S17 as donor bacteria and *B. thetaiotaomicron/B. fragilis* as receptor bacteria were co-incubated under aerobic conditions, and the cells were transferred to brain heart infusion (BHI) medium agar plates which contained mutant selection with gentamicin (200 μg/mL) and erythromycin (25 μg/mL). Primers

targeting junction regions between pGERM and *SPT* genes were utilized to pick resistant colonies for qPCR identification. Both WT and *SPT*-depleting *B. thetaiotaomicron* and *B. fragilis* were grown at 37 °C in BHI medium supplemented with 0.3 g/mL cysteine under anaerobic conditions. The bacterial culture medium was centrifuged for 10 min at $8000 \times g$ and 4 °C, and the pellets were resuspended with oxygen-free PBS to obtain bacteria for oral administration. Colonization fitness of the four strains in rats was analyzed by 2% agarose gel electrophoresis and 3500 UV images (Supplementary Fig. 3h).

NAC permeability evaluation in intestinal microbiota and differentiated Caco-2 cells co-culture system

To establish the differentiated Caco-2 cell model, the cell line was routinely subcultured using the low-density protocol[43]. Caco-2 cells were initially seeded at 6.2×10^3 cells/cm^2 and subcultured for at least 10 passages at 50% confluence (5.4×10^4 cells/cm^2). Differentiation was achieved by seeding cells on polyethylene-terephthalate (PET) membrane inserts at a density of 3×10^5 cells/cm^2 and maintained for 21 days in complete medium, which was changed three times a week. Then intestinal microbes-supplemented differentiated Caco-2 monolayer coculture system was constructed to determine NAC intestinal permeability changes when exposed to different compositions of intestinal microbiota. The plug of coculture system inserts tightly, physically blocking the influx of external oxygen. It thus allows hypoxia maintenance in the upper well, while oxygen freely perfused the middle and basal well. 10^5 cells/cm^2 differentiated Caco-2 cells (passages 30-60) were cultured in DMEM medium (20% FBS without antibiotics). When the cell layer was close to confluence for 3 days, we equilibrated the medium in the apical well with anaerobic gas and subsequently added intestinal microbiota to the apical well which was sealed by inserting a plug made of butyl rubber (AsONE international, Santa Clara, CA). The oxygen concentration of the apical well was measured by a fiberoptic oxygen meter (PreSens. Regensburg, Germany), which was set with 0-0.2 mg/L oxygen. On the day of the permeation study, only monolayers with TEER values (using Millicell®-ERS system, Millipore, USA) higher than 600 Ω.cm^2 were used in the permeation study. For middle well to basal well transport study, 1.5 mL of the HBSS buffer was added to the basal side, and 0.5 mL test solution (0.5-20 μM NAC) was added to the apical side, vice versa. Then volume of the basal well was maintained constantly at 0.5 mL by replacing it with fresh HBSS at different time points. NAC concentrations of these samples were determined using the RP-HPLC assay. The PM2000/AQUA immunofluorescent platform was utilized to capture images of Caco-2 cells within the middle well of each core tissue microarray. The slide was automatically scanned using the PM2000 hardware (HistoRx/Genoptix), and fluorescent images were captured at ×20 magnification in two channels: DAPI (for cell nuclei) and Alexa Fluor 647 (for azide-tagged sphingolipids) at predefined locations. The ScanScopeFL fluorescence immunohistochemistry platform was employed to digitally capture images of each channel (DAPI, Alexa Fluor 647) across the entire slide at ×20 magnification using the ScanScopeFL (Aperio/Leica Biosystems). The two-color immunofluorescence images were then separated into single-channel fluorescence images to allow for the quantification of each fluorescent channel individually. The immunofluorescence intensity of Caco-2 cells was quantified using the following formula:

$$\text{Mean Gray Value} = \text{Integrated Density} / \text{Area}$$

To account for noise interference, the relative fluorescence intensity was calculated, setting the initial immunofluorescent intensity as reference value of 1.

Data splitting strategy for ML model training

For each ML model, 20% samples in the dataset were apportioned randomly as test set for evaluation using the train_test_split method

from the Scikit-learn library in Python[44]. The remaining 80% samples utilized for feature selection and hyperparameter optimization were taken as training set. In these processes, data were again split following the k-fold cross-validation strategy ($k = 5$). 80% samples of training set were employed to train model architectures by different features and hyperparameters, and the other 20% were used to test the performance of model structure with fixed features and hyperparameters.

ML model development and evaluation

A total of nine ML algorithms were trained and investigated in the study, including MLR, NN, k-NN, RF, DT, PLS, SVR, XGB and LGBM models. All of which were built and evaluated in Python. LGBM and XGB models were constructed using the LightGBM package and XGBoost package[45,46], and all the other models were constructed using the Scikit-learn library[44]. In all cases, prior to training any of the ML models, a data preprocessing step was performed to label the categorical data through Label Encoder and One Hot Encoder methods from Scikit-learn library[44]. Specifically, all data were standardized with Standard Scaler method before training MLR, NN, k-NN, PLS and SVR models. The features and hyperparameters in ML models were tuned using SelectKBest and Randomized SearchCV, Grid SearchCV method, respectively. Feature selection was conducted using SelectKBest from the Scikit-learn library in Python[44] to rank the initial input features based on their mutual information regression scores. The performance of models was assessed by 5-fold cross-validation after the removal of unselected features. Optimization of hyperparameters was performed by integrating Randomized SearchCV and Grid SearchCV methodologies: the Randomized SearchCV was initially applied to explore 100 random hyperparameter configurations in each case to increase the range of potential hyperparameters, and the Grid SearchCV was subsequently applied to identify the optimal set of hyperparameters. This enabled a more quantitative evaluation for the general prediction performance of each model by determining the values of absolute error in test set. AE and MAE were calculated as the following equations:

$$\text{Absolute error (AE)} = |predicted\ yi - true\ yi| \qquad (2)$$

$$\text{Mean absolute error (MAE)} = \frac{\sum_{i=1}^{n}|predicted\ yi - true\ yi|}{n} \qquad (3)$$

Where yi is the $Fsys$ obtained from the test datasets; n is the total number of data points.

Statistics & reproducibility

At least three independent experiments were expressed as mean ± standard deviation. Statistical Product and Service solutions (SPSS23.0) were used to determine the statistical significance. Graph-Pad Prism v.8.01 is also used for statistical analysis (GraphPad Software, La Jolla, CA, USA). Error bars in the scatter plots and the bar graphs represent *SD*. Comparisons between groups were conducted utilizing Student's t test, Mann–Whitney U-test, one-way analysis of variance (ANOVA), Bonferroni and Scheffe post-test for continuous variables. Spearman's Rank Correlation was applied to analyze the correlation between variables. Unless otherwise stated, double-sided test was applied in each analysis. No statistical method was used to predetermine sample size. No data were excluded from the analyses. Investigators were not blinded to allocation during experiments and outcome assessment. Clinical participants were randomly split by the train_test_split method in Scikit-learn library of Python software.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The 16S rRNA sequencing and metabolomics data supporting the results in this study are deposited in BIG Sub (Study ID: PRJCA039557, https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA039557) and Metabolomics Workbench (Study ID: ST003877), respectively. The minimum dataset for main figures and Supplementary Figs. that support the findings of this study are openly available in Figshare (https://doi.org/10.6084/m9.figshare.27851817). Source data are provided with this paper.

Code availability

The source code containing fixed features and hyperparameter combinations identified in this study to ensure reproducibility is openly available in the cloud-based executable platform Code Ocean (https://codeocean.com/capsule/9444975/tree/v3).

References

1. Garcia-Sanchez, S. et al. A new methodology to estimate drug cost avoidance in clinical trials: Development and application. *Front. Oncol.* **12**, 889575 (2022).
2. Ezeji, J. C. et al. Parabacteroides distasonis: intriguing aerotolerant gut anaerobe with emerging antimicrobial resistance and pathogenic and probiotic roles in human health. *Gut Microbes* **13**, 1922241 (2021).
3. Amidon, G. L., Lennernas, H., Shah, V. P. & Crison, J. R. A theoretical basis for a biopharmaceutic drug classification: The correlation of in vitro drug product dissolution and in vivo bioavailability. *Pharm. Res.* **12**, 413–420 (1995).
4. Samineni, R., Chimakurthy, J. & Konidala, S. Emerging role of biopharmaceutical classification and biopharmaceutical drug disposition system in dosage form development: a systematic review. *Turk. J. Pharm. Sci.* **19**, 706–713 (2022).
5. Bootz-Maoz, H. et al. Ex vivo intestinal permeability assay (X-IPA) for tracking barrier function dynamics. *npj Biofilms Microbi* **9**, 44 (2023).
6. Dahlgren, D. & Lennernas, H. Intestinal permeability and drug absorption: Predictive experimental, computational and in vivo approaches. *Pharmaceutics* **11**, 411 (2019).
7. Volpe, D. A. Advances in cell-based permeability assays to screen drugs for intestinal absorption. *Expert. Opin. Drug Dis.* **15**, 539–549 (2020).
8. Du Toit, A. Scooping up all the drugs. *Nat. Rev. Microbiol.* **19**, 682 (2021).
9. Klunemann, M. et al. Bioaccumulation of therapeutic drugs by human gut bacteria. *Nature* **597**, 533–538 (2021).
10. Tenorio, M., Graciliano, N. G., Moura, F. A., Oliveira, A. & Goulart, M. N-acetylcysteine (NAC): Impacts on human health. *Antioxidants.* **10**, 967 (2021).
11. Licata, A. et al. N-acetylcysteine for preventing acetaminophen-induced liver injury: a comprehensive review. *Front. Pharmacol.* **13**, 828565 (2022).
12. An, D. et al. Sphingolipids from a symbiotic microbe regulate homeostasis of host intestinal natural killer T cells. *Cell* **156**, 123–133 (2014).
13. Pedre, B., Barayeu, U., Ezerina, D. & Dick, T. P. The mechanism of action of N-acetylcysteine (NAC): The emerging role of H(2)S and sulfane sulfur species. *Pharmacol. Ther.* **228**, 107916 (2021).
14. Yu, L. X., Lipka, E., Crison, J. R. & Amidon, G. L. Transport approaches to the biopharmaceutical design of oral drug delivery systems: prediction of intestinal absorption. *Adv. Drug Deliv.* **19**, 359–376 (1996).
15. Yu, L. X. & Amidon, G. L. A compartmental absorption and transit model for estimating oral drug absorption. *Int. J. Pharm.* **186**, 119–125 (1999).
16. Kimura, T. & Higaki, K. Gastrointestinal transit and drug absorption. *Biol. Pharm. Bull.* **25**, 149–164 (2002).

17. Jamei, M. et al. The simcyp population-based ADME simulator. *Expert Opin. Drug Metab. Toxicol.* **5**, 211–223 (2009).

18. Jamei, M. et al. Population-based mechanistic prediction of oral drug absorption. *AAPS J.* **11**, 225–237 (2009).

19. Brown, E. M. et al. Bacteroides-derived sphingolipids are critical for maintaining intestinal homeostasis and symbiosis. *Cell Host Microbe* **25**, 668–680 (2019).

20. Johnson, E. L. et al. Sphingolipids produced by gut bacteria enter host metabolic pathways impacting ceramide levels. *Nat. Commun.* **11**, 2471 (2020).

21. Le, H. H., Lee, M. T., Besler, K. R. & Johnson, E. L. Host hepatic metabolism is modulated by gut microbiota-derived sphingolipids. *Cell Host Microbe* **30**, 798–808 (2022).

22. Abuhelwa, A. Y., Williams, D. B., Upton, R. N. & Foster, D. J. Food, gastrointestinal pH, and models of oral drug absorption. *Eur. J. Pharm. Biopharm.* **112**, 234–248 (2017).

23. Brody, H. The gut microbiome. *Nature* **577**, S5 (2020).

24. Heintz-Buschart, A. & Wilmes, P. Human gut microbiome: function matters. *Trends Microbiol* **26**, 563–574 (2018).

25. Kastl, A. J., Terry, N. A., Wu, G. D. & Albenberg, L. G. The structure and function of the human small intestinal microbiota: current understanding and future directions. *Cell Mol. Gastroenter.* **9**, 33–45 (2020).

26. Martinez-Guryn, K. et al. Small intestine microbiota regulate host digestive and absorptive adaptive responses to dietary lipids. *Cell Host Microbe* **23**, 458–469 (2018).

27. Zoetendal, E. G. et al. The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. *ISME J.* **6**, 1415–1426 (2012).

28. Raman, M. et al. The external aldimine form of serine palmitoyl-transferase: Structural, kinetic, and spectroscopic analysis of the wild-type enzyme and HSAN1 mutant mimics. *J. Biol. Chem.* **284**, 17328–17339 (2009).

29. Chen, C. D. et al. Crystal structures of complexes of the branched-chain aminotransferase from Deinococcus radiodurans with alpha-ketoisocaproate and L-glutamate suggest the radiation resistance of this enzyme for catalysis. *J. Bacteriol.* **194**, 6206–6216 (2012).

30. Fhu, C. W. & Ali, A. Protein lipidation by palmitoylation and myristoylation in cancer. *Front. Cell Dev. Biol.* **9**, 673647 (2021).

31. Gu, M. et al. Palmitoyltransferase DHHC9 and acyl protein thioesterase APT1 modulate renal fibrosis through regulating beta-catenin palmitoylation. *Nat. Commun.* **14**, 6682 (2023).

32. Li, Y., Scott, R., Doughty, J., Grant, M. & Qi, B. Protein S-acyltransferase 14: a specific role for palmitoylation in leaf senescence in arabidopsis. *Plant Physiol.* **170**, 415–428 (2016).

33. Xiao, X. et al. Prevotella copri variants among a single host diverge in sphingolipid production. *Mbio* **15**, e0240923 (2024).

34. Wattenberg, B. W. Kicking off sphingolipid biosynthesis: structures of the serine palmitoyltransferase complex. *Nat. Struct. Mol. Biol.* **28**, 229–231 (2021).

35. Hanada, K., Nishijima, M., Fujita, T. & Kobayashi, S. Specificity of inhibitors of serine palmitoyltransferase (SPT), a key enzyme in sphingolipid biosynthesis, in intact cells. A novel evaluation system using an SPT-defective mammalian cell mutant. *Biochem. Pharmacol.* **59**, 1211–1216 (2000).

36. Holdness, M. R. Clinical pharmacokinetics of N-acetylcysteine. *Clin. Pharmacokinet.* **20**, 123–134 (1991).

37. Mokra, D., Mokry, J., Barosova, R. & Hanusrichterova, J. Advances in the use of N-acetylcysteine in chronic respiratory diseases. *Antioxidants-Basel* **12**, 1713 (2023).

38. Papi, A., Di Stefano, A. & Radicioni, M. Pharmacokinetics and safety of single and multiple doses of oral n-acetylcysteine in healthy Chinese and Caucasian volunteers: an open-label, phaseI clinical study. *Adv. Ther.* **38**, 468–478 (2021).

39. Sahasrabudhe, S. A. et al. Population pharmacokinetic analysis of N-acetylcysteine in pediatric patients with inherited metabolic disorders undergoing hematopoietic stem cell transplant. *J. Clin. Pharmacol.* **61**, 1638–1645 (2021).

40. Rajput, D., Wang, W. J. & Chen, C. C. Evaluation of a decided sample size in machine learning applications. *BMC Bioinforma.* **24**, 48 (2023).

41. Zhang, A., Xing, L., Zou, J. & Wu, J. C. Shifting machine learning for healthcare from development to deployment and from models to data. *Nat. Biomed. Eng.* **6**, 1330–1345 (2022).

42. Chu, J. N. & Traverso, G. Foundations of gastrointestinal-based drug delivery and future developments. *Nat. Rev. Gastroenter.* **19**, 219–238 (2022).

43. Natoli, M. et al. Cell growing density affects the structural and functional properties of Caco-2 differentiated monolayer. *J. Cell. Physiol.* **226**, 1531–1543 (2011).

44. Kannt, A. et al. Activation of thyroid hormone receptor-beta improved disease activity and metabolism independent of body weight in a mouse model of non-alcoholic steatohepatitis and fibrosis. *Br. J. Pharmacol.* **178**, 2412–2423 (2021).

45. Chaurasia, B. & Summers, S. A. Ceramides in metabolism: Key lipotoxic players. *Annu. Rev. Physiol.* **83**, 303–330 (2021).

46. Collison, L. W. et al. The composition and signaling of the IL-35 receptor are unconventional. *Nat. Immunol.* **13**, 290–299 (2012).

Acknowledgements

Author contributions

C.Y.M. and G.Z. have conceived the project. Z.Y.H., W.Y.J. and W.W.Y. collected clinical samples and pathological analysis. Q.T.T. and X.M.C. have implemented experiments of pharmaceutical analysis. D.C.S. and Q.T.T. performed statistical analysis for multi-omics data and machine learning algorithms. Z.Y.H. and D.C.S. wrote the manuscript, which was edited by all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-59916-7.

Correspondence and requests for materials should be addressed to Yu-Hang Zhang, Gan Zhou or Yi-Min Cui.

Peer review information *Nature Communications* thanks Franco Scaldaferri, and the other, anonymous, reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.