

Mapping human microbiome drug metabolism by gut bacteria and their genes

Michael Zimmermann^{1,3}, Maria Zimmermann-Kogadeeva^{1,3}, Rebekka Wegmann^{1,2} & Andrew L. Goodman^{1*}

Individuals vary widely in their responses to medicinal drugs, which can be dangerous and expensive owing to treatment delays and adverse effects. Although increasing evidence implicates the gut microbiome in this variability, the molecular mechanisms involved remain largely unknown. Here we show, by measuring the ability of 76 human gut bacteria from diverse clades to metabolize 271 orally administered drugs, that many drugs are chemically modified by microorganisms. We combined high-throughput genetic analyses with mass spectrometry to systematically identify microbial gene products that metabolize drugs. These microbiome-encoded enzymes can directly and substantially affect intestinal and systemic drug metabolism in mice, and can explain the drug-metabolizing activities of human gut bacteria and communities on the basis of their genomic contents. These causal links between the gene content and metabolic activities of the microbiota connect interpersonal variability in microbiomes to interpersonal differences in drug metabolism, which has implications for medical therapy and drug development across multiple disease indications.

Following administration, drug molecules typically undergo chemical modification(s); the resulting metabolites can have functional and toxicological properties that are distinct from those of their parent drug¹. Most drugs are delivered orally and can encounter commensal microorganisms in the small and large intestine. These microorganisms collectively encode 150-fold-more genes than the human genome; this genetic diversity encompasses a rich enzyme repository with drug-metabolizing potential. Anecdotal examples of interactions between the gut microbiome and drugs or drug metabolites, with intestinal and systemic pharmacological effects, have previously been reported. Such compound modifications by gut microorganisms can lead either to their activation (for example, sulfasalazine²), inactivation (for example, digoxin³) or toxification (for example, sorivudine and brivudine^{4,5}, and irinotecan⁶). For a few drugs, microbial biotransformation has been assigned to specific bacterial strains and gene products^{3,5,7}. However, these examples are the exception, as there is little systematic understanding of the scope, specificity or microbial and/or chemical determinants of microbiome–drug interactions⁸.

We set out to systematically assay interactions between drugs and microorganisms by measuring the ability of representative human gut bacteria to metabolize structurally diverse drugs, and by identifying drug-metabolizing microbial gene products. We establish that these drug-metabolizing microbial proteins can contribute to the *in vivo* drug metabolism of gnotobiotic mice, and provide evidence that metagenomics and genomics sequence data can explain the capacity of both isolated gut bacteria and complete communities to convert specific drugs. This could provide a means to mechanistically connect information about the microbiome to interpersonal variation in drug metabolism and toxicity.

Drug-metabolizing bacteria from the gut microbiome

We first assessed the capacity of 76 bacterial species and/or strains—which represent the major phyla of the human gut microbiome—to chemically modify medical drugs *in vitro* (Fig. 1a, Supplementary Table 1). We used a previously established combinatorial pooling strategy⁹ to assign 271 drugs across 21 pools, such that each drug is represented in quadruplicate but shares a pool with any other drug twice

at most (Extended Data Fig. 1a). The 271 drugs were selected to span chemical drug space, which resulted in a selection of diverse clinical indications (excluding antibiotics), physicochemical properties and predicted intestinal concentrations (Fig. 1b, Extended Data Fig. 1b–d, Supplementary Table 2). We incubated each gut species or strain with each drug pool and three vehicle controls under anaerobic conditions, and measured drug concentrations before and after a 12-h incubation by liquid-chromatography-coupled mass spectrometry (LC–MS). The 3,840 samples we analysed comprise a total of 20,596 bacteria–drug interactions, measured in quadruplicate.

We discovered that, for two thirds (176/271) of the assayed drugs, the level of the drug after incubation was significantly reduced (>20%, FDR-corrected *P* value ≤ 0.05) by at least one bacterial strain, and that each strain metabolizes 11–95 drugs (Extended Data Fig. 1e–g, Supplementary Table 3). Drug levels were largely unchanged in no-bacteria controls that were buffered to pH 4–7, controlling for acidification of the culture medium. By contrast, levels of positive-control drugs that were expected to be metabolized by gut bacteria¹⁰—such as sulfasalazine, lovastatin, omeprazole and risperidone—were significantly decreased over time (>20%, FDR-corrected *P* value ≤ 0.05) (Extended Data Fig. 2a). Clustering the bacterial isolates according to their drug-metabolizing activities recapitulates their phylogenetic relationships to the strain level, and reveals phylum-specific metabolic activities (Fig. 1c, Extended Data Fig. 3a). Clustering the drugs on the basis of these data revealed groups of compounds that share structural features, as shown by functional group and maximum common substructures analysis (Fig. 1d, e, Extended Data Fig. 3b, Supplementary Table 4). This suggests possible chemical targets for metabolic modifications by bacteria. For example, drugs that are specifically metabolized by Bacteroidetes (cluster I in Fig. 1c) contain ester or amide groups that can be hydrolysed, whereas the compounds that are metabolized by most bacteria (except Proteobacteria) (cluster II in Fig. 1c) all contain a nitro or azo group, which is prone to reduction in anaerobic metabolism. Functional group analysis suggests that particular chemical substructures (such as lactones, and nitro, azo and urea groups) predispose compounds for microbial metabolism (Extended Data Fig. 2b). Chemical groups that

¹Department of Microbial Pathogenesis and Microbial Sciences Institute, Yale University School of Medicine, New Haven, CT, USA. ²Present address: Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland. ³These authors contributed equally: Michael Zimmermann, Maria Zimmermann-Kogadeeva. *e-mail: andrew.goodman@yale.edu

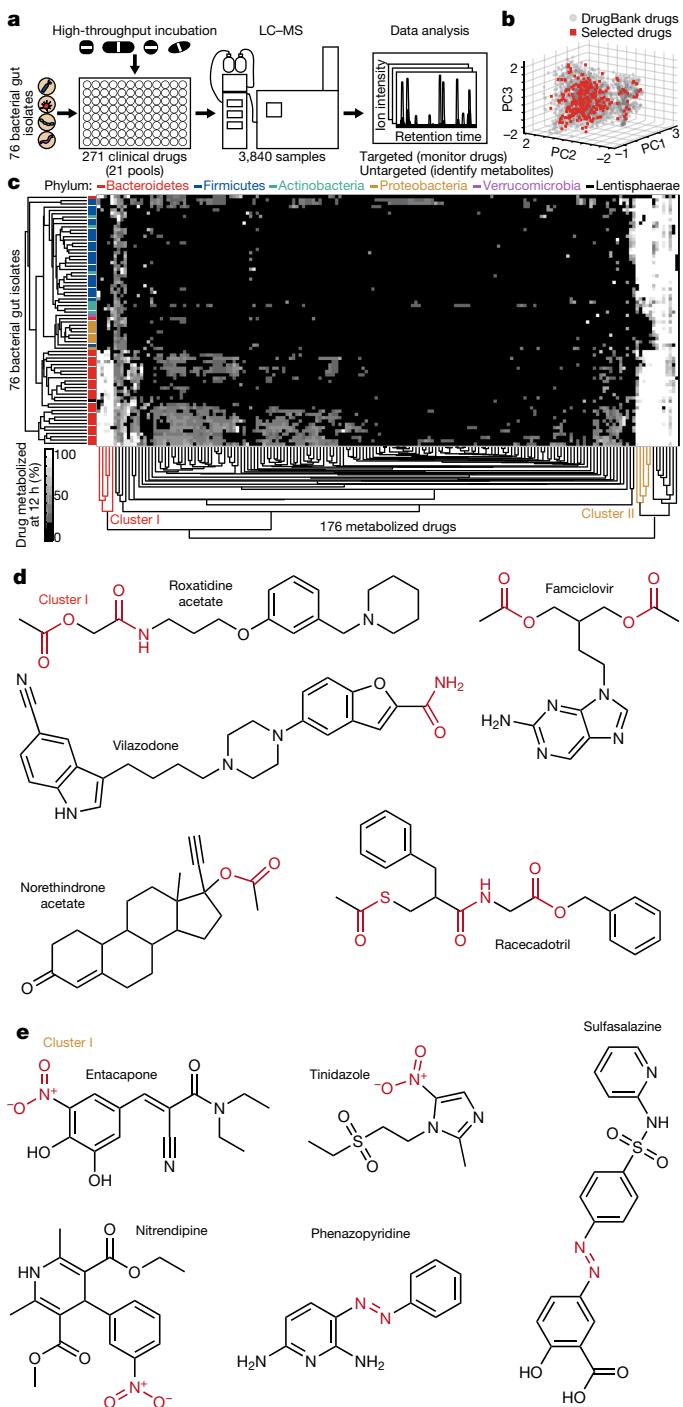


Fig. 1 | Drug-metabolizing activities of human gut bacteria.

a, Schematic of the assay. **b**, Chemical diversity of tested compounds compared to 2,099 clinical drugs (from DrugBank²¹). PC, principal component. **c**, Heat map of the 176 drugs metabolized by at least 1 of the 76 human gut bacterial strains. Strains and drugs are arranged by hierarchical clustering according to metabolic activities. **d**, **e**, Examples of drugs that cluster together according to their metabolism. See Methods and Supplementary Table 3 for statistics and reproducibility.

have previously been reported to be targeted by microbial metabolism¹⁰ (for example, esters and amides) are also found among drugs that are not metabolized by any of the bacteria that we tested, which suggests additional structural specificity to microbial drug metabolism.

Identifying bacteria-produced drug metabolites

To identify products of bacterial drug metabolism and gain insights into the chemistry of observed biotransformations, we conducted

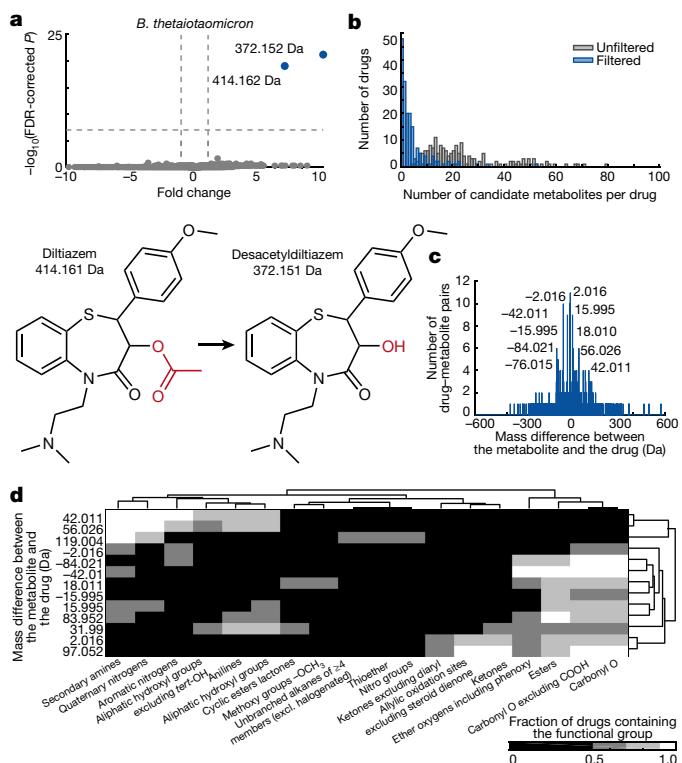


Fig. 2 | Bacteria-derived drug metabolites. **a**, Representative volcano plot showing compounds that were detected by untargeted metabolomics associated with a specific drug metabolized by a given bacterial strain (for example, diltiazem metabolism by *B. thetaiotaomicron*). The four pools containing a given drug are compared to all other pools (x axis, fold change (\log_2 diltiazem pools versus other pools, at 12 h)); y axis, $-\log_{10}(\text{FDR-corrected } P \text{ value})$). Experimentally determined masses of diltiazem-associated compounds are indicated. Compounds significantly associated with diltiazem pools are shown in blue; others are shown in grey. **b**, Identification of candidate drug metabolites. Number of drug-specific compounds detected before (grey) and after (blue) the elimination of measurement artefacts. **c**, Mass shift distribution of candidate metabolites. Mass shifts were calculated between the candidate metabolite and the corresponding drug. **d**, Chemical group enrichment analysis of drugs undergoing the same mass shift upon microbial conversion. Mass differences and chemical groups are arranged by hierarchical clustering according to the fraction of metabolized drugs that contains a specific chemical group. See Methods and Supplementary Tables 5, 6 for statistics and reproducibility.

untargeted metabolomics analysis of all samples. For each bacterial isolate, we identified compounds that solely occur in the presence of a specific drug (see Fig. 2a for an example). This resulted in 6,572 different drug–compound pairs for all 20,596 tested bacteria–drug interactions (fold change (expressed in \log_2) ≥ 1 , FDR-corrected P value $\leq 10^{-6}$) (grey bars in Fig. 2b). To eliminate measurement artefacts, we applied data filtering on the basis of chromatographic retention, mass defects of drugs and their putative metabolites, and ion fragmentation. In addition, we included only compounds that significantly accumulated (fold change (expressed in \log_2) ≥ 1 , FDR-corrected P value ≤ 0.05), whereas their associated parent drug decreased upon bacterial incubation. This analysis correctly identifies reported microorganism-derived drug metabolites¹⁰ (for example, metabolites of sulfasalazine, paliperidone and pantoprazole; Supplementary Table 5). We found 871 candidate drug metabolites that are specific to the presence of a given drug; these represent direct products of bacterial drug modifications or bacterial responses that are unique to a specific drug (blue bars in Fig. 2b, Supplementary Tables 5, 6).

To gain insights into the chemistry of microbial drug metabolism, we calculated the mass difference between each drug-specific metabolite identified and its associated drug. The resulting differences were

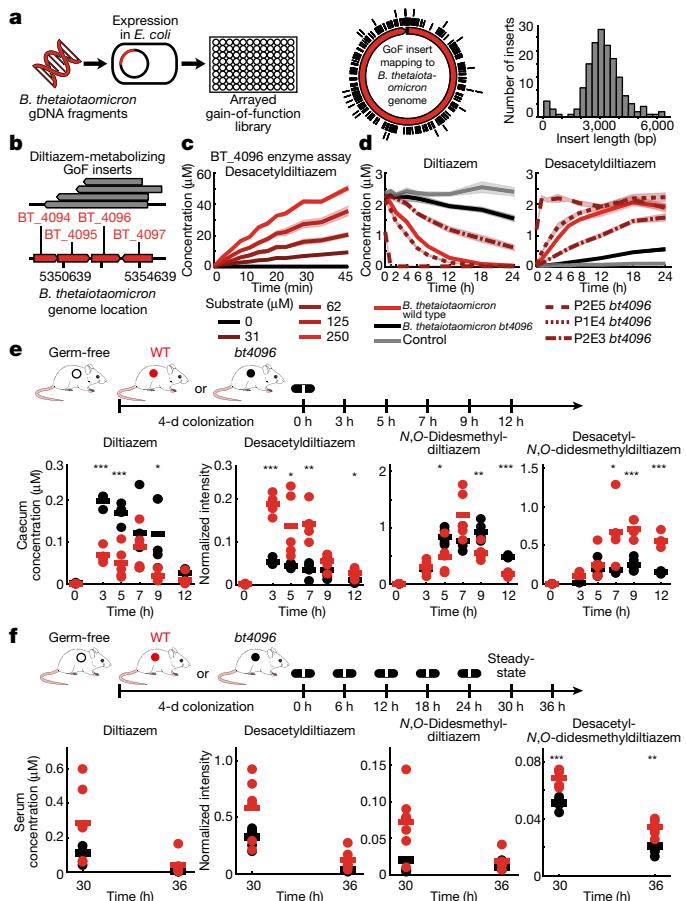


Fig. 3 | Identification and in vivo characterization of microbial drug-metabolizing gene products, with *B. thetaiotaomicron* diltiazem metabolism as an example. **a**, Scheme for generation of an arrayed gain-of-function (GoF) library, source genome coverage and insert size distribution of the *B. thetaiotaomicron* library. **b**, Mapping of active insert sequences to the *B. thetaiotaomicron* genome. **c**, Enzymatic validation using purified BT_4096. **d**, Diltiazem-metabolizing activity of *B. thetaiotaomicron* wild-type, *bt4096*-mutant and *bt4096*-complemented strains at different expression levels (P2E5 > P1E4 > P2E3). **e**, Intestinal kinetics of diltiazem and deacetylated diltiazem metabolites after single oral diltiazem administration to mice mono-colonized with either *B. thetaiotaomicron* wild-type (WT) or *bt4096*-mutant strains. **f**, Serum levels of diltiazem and deacetylated diltiazem metabolites after serial oral diltiazem administration to mice mono-colonized with either *B. thetaiotaomicron* wild-type or *bt4096*-mutant strains. In **c**, **d**, shaded areas depict the mean and s.d. of independent assays or cultures ($n = 4$). In **e**, **f**, times reflect hours after oral diltiazem administration and horizontal lines depict the mean per time point. * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$ (FDR-corrected P value). See Methods and Supplementary Tables 10, 11 for statistics and reproducibility.

not randomly distributed: discrete mass differences appeared multiple times, which suggests conserved metabolic transformations of different drugs (Fig. 2c). The high mass accuracy of the measurements, together with the drug structures, suggest chemical modifications that result in decreased or increased masses. The negative mass differences -2.016 , -15.995 and -42.011 suggest oxidation ($-H_2$), reduction ($-O$) and deacetylation ($-C_2H_2O$), respectively; the positive mass differences $+2.016$, $+15.995$, $+18.010$, $+42.011$ and $+56.026$ suggest hydrogenation ($+H_2$), oxidation ($+O$), hydroxylation ($+H_2O$), acetylation ($+C_2H_2O$) and propionylation ($+C_3H_4O$), respectively (Fig. 2c, Supplementary Tables 5, 6). To systematically identify the structural targets of microbial metabolism, we performed functional group analysis on sets of drugs that shared a specific mass difference with respect to their metabolite (Fig. 2d, Supplementary Table 4). As expected, all drugs that undergo either deacetylation or hydrogenation contain an

acetyl ester or an alkene group (Extended Data Fig. 4a). Intriguingly, most of the drugs that are predicted to be acylated contain an aliphatic amine or hydroxyl, which suggests *N*- and *O*-acylation (see below).

To assess whether drug metabolism of axenic bacterial cultures can translate to animal models and to complete human gut microbial communities, we focused on dexamethasone, a corticosteroid that was uniquely metabolized by *Clostridium scindens* (ATCC (American Type Culture Collection) 35704) in our screen. This desmolytic (side-chain cleaving) activity produces the androgen form of the drug^{11,12} (Extended Data Fig. 4b, c). We administered an oral dose of dexamethasone or vehicle to germ-free mice or gnotobiotic mice mono-colonized with *C. scindens* ($GN^C. scindens$), and quantified dexamethasone and its androgen metabolite in different body compartments (seven hours after administration, corresponding to two serum drug half-lives) (Extended Data Fig. 5a, Supplementary Table 7). Dexamethasone was detected in the caecum of mice from both groups, with a significant reduction in gnotobiotic mice mono-colonized with *C. scindens* relative to germ-free controls. The androgen metabolite accumulated to greater levels in the caecum and serum of gnotobiotic mice mono-colonized with *C. scindens*, as compared to the germ-free controls. This demonstrates that the drug reaches the lower intestine, which carries high bacterial density ($1.24 \pm 0.3 \times 10^9$ colony-forming units per gram of luminal contents; $n = 4$ mice), and that dexamethasone is metabolized in vivo by an intestinal microorganism, which affects serum metabolite levels (Extended Data Fig. 5b, Supplementary Table 8). This probably extends to other corticosteroids, as we also found that prednisone, prednisolone, cortisone and cortisol were desmolytically metabolized by *C. scindens* (Extended Data Fig. 5c). Notably, anaerobic incubation of dexamethasone with faecal cultures from 28 healthy human donors illustrates substantial interpersonal variation in drug-metabolizing activity, but this capacity correlates with neither bacterial culture density nor the abundance of *C. scindens* in a community (Extended Data Fig. 5d-f, Supplementary Table 9). This is consistent with previous reports that *C. scindens* metabolizes endogenous steroid hormones in a strain-specific manner¹³, and suggests that other bacterial taxa may also metabolize dexamethasone. These results, together with previous reports³, emphasize that species identity is often insufficient to explain bacterial drug metabolism, and that the identification of gene markers that are directly associated with enzymatic drug conversion may instead be necessary.

Identification of drug-metabolizing gene products

Many of the drug modifications (such as hydrolyses and reductions) found in the initial screen were generic, which makes it challenging to predict the gene products responsible from genomic sequences alone. Therefore, we developed a gain-of-function approach to identify DNA fragments from any source species that confer drug metabolic capacity to a heterologous host. To establish this protocol, we selected *Bacteroides thetaiotaomicron*—which metabolized 46 drugs (including diltiazem)—as an exemplary source species (Extended Data Fig. 6a). First, we isolated and sheared *B. thetaiotaomicron* genomic DNA to 2–8-kb fragments, cloned them into an *Escherichia coli* expression vector and arrayed 51,000 transformed *E. coli* clones in a 384-well format. Sequencing 160 randomly selected clones revealed a mean insert length of 3.1 kb, which suggested a homogenous, approximately 25-fold genome coverage for the entire library (Fig. 3a). Second, we assembled 133 pools of 384 clones, incubated them with a mixture of the drugs metabolized by *B. thetaiotaomicron*, and measured drug and drug metabolite levels over time to identify active library plates that included clones that gained specific drug-metabolizing capacities (see Extended Data Fig. 6b for an example). Third, we pooled rows and columns of active library plates, and repeated the drug metabolism assay to identify the plate position (bacterial clone) that gained drug-metabolizing function (see Extended Data Fig. 6c for an example). Fourth, we sequenced the genomic DNA inserts carried by these active clones (see Fig. 3b for an example). To validate the identified genes (for example, *bt4096* for diltiazem metabolism; in gene locus tag abbreviations,

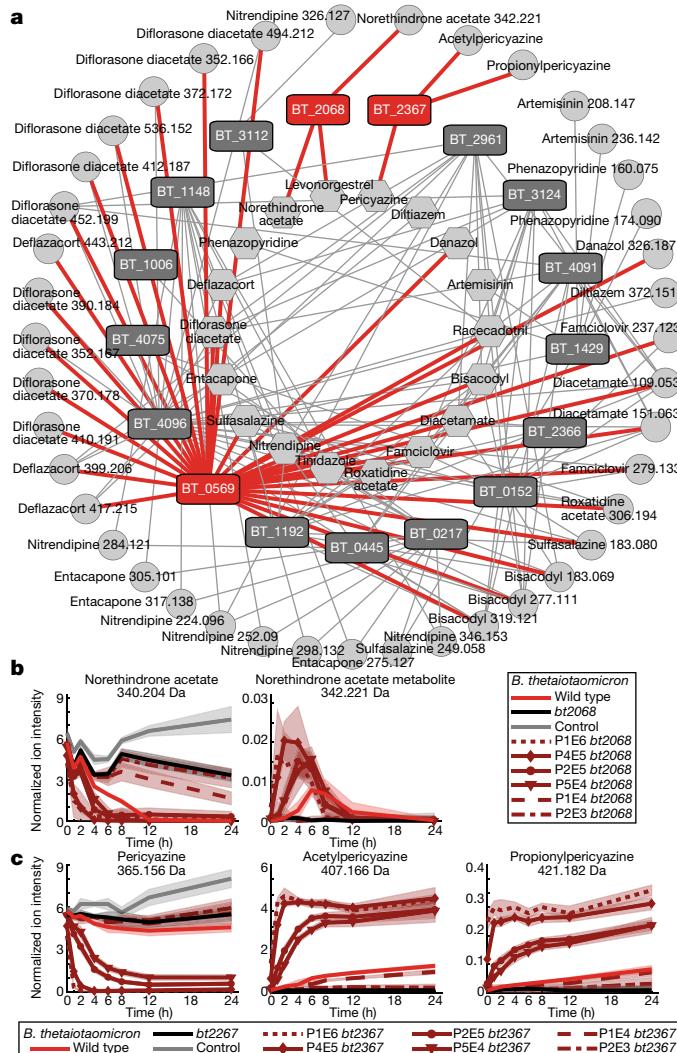


Fig. 4 | Genome-wide identification of drug-metabolizing gene products in *B. thetaiotaomicron*. **a**, Network of enzyme–substrate–product drug metabolic interactions for *B. thetaiotaomicron*. Each node represents an enzyme (rectangles), a drug substrate (hexagons) or a metabolite product (circles), and each edge represents a validated metabolic interaction (targeted cloning of the gene into *E. coli* results in metabolism of a given drug or production of a specific drug metabolite). **b**, Norethindrone-acetate-metabolizing activity of *B. thetaiotaomicron* wild-type, *bt2068*-mutant and *bt2068*-complemented strains. **c**, Pericyazine-metabolizing activity of *B. thetaiotaomicron* wild-type, *bt2367*-mutant and *bt2367*-complemented strains. In **b**, **c**, promoter strengths are P1E6 > P4E5 > P2E5 > P5E4 > P1E4 > P2E3. Shaded areas depict the mean and s.d. of independent cultures ($n = 4$). See Methods for statistics and reproducibility.

bt denotes *B. thetaiotaomicron*), we repeated drug-metabolizing assays with expression constructs that carried PCR-amplified gene sequences (Extended Data Fig. 6d, e); demonstrated that the purified enzyme catalyses the drug transformation (Fig. 3c); and established that in-frame deletion of *bt4096* in *B. thetaiotaomicron* leads to a loss of diltiazem-metabolizing activity in vitro, which is restored by gene complementation in a heterologous genomic location (Fig. 3d).

Diltiazem is an oral calcium channel blocker that is used in the treatment of hypertension, arrhythmia and angina pectoris. The drug is metabolized in vivo into multiple metabolites that maintain a variable inhibition of calcium channels and that are targets of distinct hepatic cytochromes, which gives rise to numerous potential drug–drug interactions^{14,15}. Therefore, we assessed the effect of bacterial activity on intestinal and systemic levels of diltiazem and its metabolites. We colonized germ-free mice with either *B. thetaiotaomicron* wild type or the

bt4096-deletion strain, orally administered diltiazem and quantified the kinetics of the drug and nine drug metabolites in nine body compartments (Extended Data Figs. 6f, 7). Intestinal drug and metabolite levels demonstrate that the deacetylation of both diltiazem and diltiazem metabolites in the gut is dependent on *bt4096* (Fig. 3e, Extended Data Figs. 6g, 8a, Supplementary Tables 10, 11). Bacterial contribution to diltiazem metabolism is further accentuated when repeated oral doses (simulating typical treatment schemes) are administered; this also demonstrates the contribution of a single gut bacterial gene to systemic drug metabolism (Fig. 3f, Extended Data Fig. 8b, Supplementary Tables 10, 11).

Using this gain-of-function approach, we identified 16 additional *B. thetaiotaomicron* gene products that—together with *BT_4096*—metabolize 18 drugs to 41 distinct metabolites; each validated by targeted cloning and expression in *E. coli* (Supplementary Tables 12, 13). The resulting network of bacterial gene products, metabolized drugs and drug metabolites produced reveals the specificity and cross-activity of these enzymes (Fig. 4a). For example, *BT_0569* shows promiscuous hydrolase activity towards many structurally diverse drugs; *BT_2068* targets 2, and *BT_2367* only 1, of the 18 drugs metabolized by the gain-of-function library. Although the gain-of-function approach can identify redundant enzymes, none were found in the *B. thetaiotaomicron* genome for *BT_2068* or *BT_2367* activities (Extended Data Fig. 9a, b). In-frame gene deletion and complementation studies in *B. thetaiotaomicron* confirm that *bt2068* and *bt2367* expression is required and rate-limiting for metabolizing norethindrone acetate and pericyazine, respectively (Fig. 4b, c). Wild-type and *bt2068*-complemented strains accumulate a norethindrone acetate metabolite that is 2.016 Da heavier than the parent drug, which suggests that the parent drug has been reduced. *BT_2068* also metabolizes the structurally related compounds levonorgestrel and progesterone (Extended Data Fig. 9c, d). *bt2367* encodes a putative acyltransferase; notably, *B. thetaiotaomicron* converts pericyazine to metabolites that have masses that are consistent with acetyl- and propionylpericyazine (Fig. 4c). Incubation of purified *BT_2367* with pericyazine and structurally related substrates together with reaction product liquid chromatography–tandem mass spectrometry (LC–MS/MS) analysis demonstrates that this enzyme uses acetyl-CoA and propionyl-CoA as cofactors to O-acylate pericyazine (Extended Data Fig. 9e–g). The approach we developed systematically identifies drug-metabolizing gene products encoded by the microbiome, resulting in gene–drug–metabolite networks that provide mechanistic insights into microbiome drug metabolism.

Genes explain species and community activity

To further expand our understanding of microbiome drug metabolism, we used the gain-of-function approach to identify 13 drug-metabolizing gene products from *Bacteroides dorei* and *Collinsella aerofaciens* that collectively metabolize 16 drugs (Fig. 5a, Extended Data Fig. 10a, b, Supplementary Table 13). We next assessed whether the genomic presence of homologues of these identified drug-metabolizing gene products can explain drug metabolism activities across the 76 bacterial strains (Fig. 1c). For example, the presence of a homologue of the diltiazem-metabolising enzyme *BT_4096* indicates whether a gut bacterium metabolizes diltiazem (Fig. 5b, Supplementary Table 14). To systematically investigate whether the identified drug-metabolizing gene products explain the drug metabolism activities of the tested species, we performed gene-set enrichment analysis for the identified gene products among bacterial strains that metabolize specific drugs¹⁶. Many of the identified drug-metabolizing gene products show significant enrichment, and therefore probably contribute to the observed microbial drug metabolism (Fig. 5c, Supplementary Table 15). For example, the set of bacterial strains that metabolize norethindrone acetate is significantly enriched for a homologue of *BACDOR_03091* (*BACDOR* refers to the gene locus tag abbreviation for *B. dorei*) and *BT_2068* (FDR-corrected P value $< 10^{-7}$), which are also homologous to one another (Fig. 5d, Extended Data Fig. 10c, d, Supplementary Tables 14, 15). All strains that carry a homologue of *BACDOR_03091*

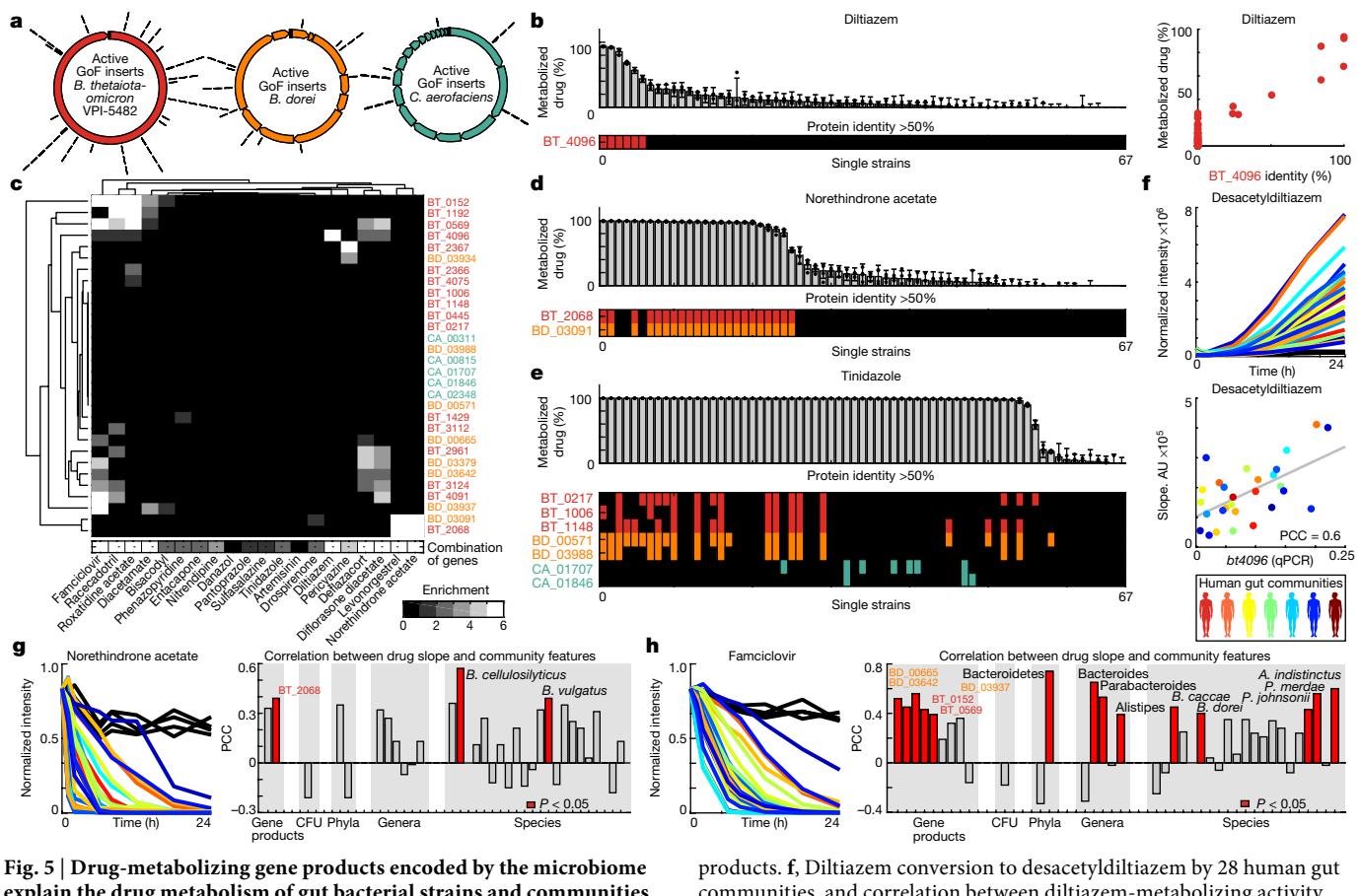


Fig. 5 | Drug-metabolizing gene products encoded by the microbiome explain the drug metabolism of gut bacterial strains and communities. **a**, Genes identified from libraries of *B. thetaiotomicron*, *B. dorei* and *C. aerofaciens* genomic DNA. **b**, Diltiazem metabolism rates for the 67 bacterial strains from Fig. 1 with available genome sequences, presence of BT_4096 homologues in their genomes, and comparison between maximal BT_4096 identity and diltiazem metabolism of each strain. **c**, Gene-set enrichment analysis for specific drug-metabolizing gene products and their combinations among the bacterial strains that metabolize each drug. Enrichment is shown for single (top) and combined (bottom) drug-metabolizing gene products among sets of bacterial strains that metabolize a given drug. BD, BACDOR; CA, COLAER locus tags. **d**, **e**, Norethindrone acetate and tinidazole metabolism of the 67 genome-sequenced gut bacteria and presence of homologues of identified drug-metabolizing gene

and BT_2068 also metabolize norethindrone acetate, and only three of the norethindrone-acetate-metabolizing strains do not encode such an enzyme homologue. We repeated the enrichment analysis using combinations of identified (including non-homologous) enzymes that metabolize the same drug. The combination of enzyme sequences that originate from different bacterial species that target the same drug(s) further increases enrichment significance, as well as the ability to explain microbial drug metabolism from genomes (Fig. 5c, Extended Data Fig. 10d, e, Supplementary Table 15). For example, we identified tinidazole-metabolizing gene products from three bacterial species (Fig. 4a, Extended Data Fig. 10b). None of these individual products is significantly enriched among tinidazole-metabolizing bacteria; however, their combination significantly (FDR-corrected P value = 0.0045) explains tinidazole metabolism across species (Fig. 5c, e, Extended Data Fig. 10c, d).

To test whether the abundance of encoded drug-metabolizing gene products can also explain drug metabolism of a complex microbial gut community, we used the diltiazem-metabolizing enzyme BT_4096 as an example. We measured the ex vivo diltiazem deacetylation kinetics of faecal samples collected from 28 unrelated human donors. These gut communities exhibited substantial differences in their diltiazem-metabolizing capacity, which correlates with the abundance of bt4096

products. **f**, Diltiazem conversion to desacetyldiltiazem by 28 human gut communities, and correlation between diltiazem-metabolizing activity of the microbiota and abundance of bt4096 homologues (measured by quantitative PCR). AU, arbitrary unit. **g**, **h**, Metabolism of norethindrone acetate and famciclovir by 28 human gut communities. Right, correlation coefficients between community drug metabolizing activity and the metagenomic abundance of drug metabolizing bacterial gene products, species, genera and phyla identified in this study. CFU, colony-forming unit; PCC, Pearson correlation coefficient. The species *Alistipes indistinctus*, *Bacteroides caccae*, *Bacteroides cellulosilyticus*, *B. dorei*, *Bacteroides vulgatus*, *Parabacteroides merdae* and *Parabacteroides johnsonii* are highlighted. In **f–h**, each colour represents a human donor, and lines depict the mean of $n = 4$ independent replicates. See Methods and Supplementary Tables 14, 16, 18, 19 for statistics and reproducibility.

homologues (as measured by quantitative PCR) but not with bacterial density or the abundance of *B. thetaiotomicron* (Fig. 5f, Extended Data Fig. 11a, b, Supplementary Table 16). Although diltiazem deacetylation is a chemically simple reaction, it requires a specific enzyme that is heterogeneously represented across strains and communities. As a result, gene abundance can partially explain the drug-metabolizing capacity of individual gut isolates and of human gut microbial communities. To generalize this approach, we performed metagenomic sequencing of the 28 bacterial communities (Extended Data Fig. 11c, Supplementary Tables 17, 18). Consistent with the quantitative PCR analysis, the abundance of BT_4096 sequence homologues in these metagenomic data was highly correlated with the diltiazem deacetylation activity of the bacterial communities (Extended Data Fig. 11d, Supplementary Table 19). Unlike quantitative PCR, metagenomic analysis does not rely on specific primers; this enabled us to examine whether less-conserved drug-metabolizing genes also correlate with the drug-metabolizing capacity of an individual's microbiome. To this aim, we measured the ex vivo activity of the 28 communities to metabolize 2 additional drugs (norethindrone acetate and famciclovir) and quantified the metagenomic abundance of bacterial phyla, genera, species and gene products that we identified in the strain- and enzyme-targeted screens to metabolize these 2 drugs. We found that the gene abundance of identified

drug-metabolizing proteins significantly ($P < 0.05$) correlated with the capacity of the bacterial community to metabolize the respective drug (Fig. 5g, h, Extended Data Fig. 11e, f, Supplementary Table 19). In the case of famciclovir, we also found significant correlations between drug-metabolizing activity of the microbiota and the abundance of particular bacterial species and broader phylogenetic groups (consistent with results of the initial screen, shown in Fig. 1c). These results provide a prospective approach for better understanding—and potentially predicting—microbiome drug metabolism, in cases in which a drug-modifying reaction is catalysed by a single gene product and also in cases that are mediated through the metabolic activity of defined members of a microbial community.

Discussion

We provide an outline of the drug-metabolizing activity of human gut bacteria, and find that about two thirds of the assayed drugs are metabolized by at least one strain in our survey. Notably, food and endogenous compounds probably serve as physiological substrates for many of these chemical transformations, as illustrated for cortisol and progesterone metabolism. We further developed an approach for microbiome-wide identification of drug-metabolizing gene products, and we validated 30 microbiome-encoded enzymes that collectively convert 20 drugs to 59 candidate metabolites. Depending on the drug and its formulation, we anticipate that such microbiome drug metabolism could have a role in determining intestinal and systemic drug and drug-metabolite exposure. Together, these results complement previous studies that highlight the effect of drugs on bacterial fitness and microbiome composition^{17,18}, and that identify microbiome-induced changes in hepatic drug metabolism^{19,20}. In addition, this study provides a mechanistic understanding of microbiome drug metabolism that may enable rational strategies aimed at manipulating individuals' microbiota to beneficially alter metabolic microbiome–host interactions.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1291-3>.

Received: 31 December 2018; Accepted: 22 May 2019;

Published online 3 June 2019.

- Obach, R. S. Pharmacologically active drug metabolites: impact on drug discovery and pharmacotherapy. *Pharmacol. Rev.* **65**, 578–640 (2013).
- Sousa, T. et al. On the colonic bacterial metabolism of azo-bonded prodrugs of 5-aminosalicylic acid. *J. Pharm. Sci.* **103**, 3171–3175 (2014).

- Haiser, H. J. et al. Predicting and manipulating cardiac drug inactivation by the human gut bacterium *Eggerthella lenta*. *Science* **341**, 295–298 (2013).
- Okuda, H., Ogura, K., Kato, A., Takubo, H. & Watabe, T. A possible mechanism of eighteen patient deaths caused by interactions of sorivudine, a new antiviral drug, with oral 5-fluorouracil prodrugs. *J. Pharmacol. Exp. Ther.* **287**, 791–799 (1998).
- Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R. & Goodman, A. L. Separating host and microbiome contributions to drug pharmacokinetics and toxicity. *Science* **363**, eaat9931 (2019).
- Wallace, B. D. et al. Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. *Science* **330**, 831–835 (2010).
- Klatt, N. R. et al. Vaginal bacteria modify HIV tenofovir microbicide efficacy in African women. *Science* **356**, 938–945 (2017).
- Aziz, R. K., Hegazy, S. M., Yasser, R., Rizkallah, M. R. & ElRakaiby, M. T. Drug pharmacomicrobiomics and toxicomicrobiomics: from scattered reports to systematic studies of drug–microbiome interactions. *Expert Opin. Drug Metab. Toxicol.* **14**, 1043–1055 (2018).
- Goodman, A. L. et al. Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* **6**, 279–289 (2009).
- Wilson, I. D. & Nicholson, J. K. Gut microbiome interactions with drug metabolism, efficacy, and toxicity. *Transl. Res.* **179**, 204–222 (2017).
- Winter, J. et al. Mode of action of steroid desmolase and reductases synthesized by *Clostridium "scindens"* (formerly *Clostridium* strain 19). *J. Lipid Res.* **25**, 1124–1131 (1984).
- Morris, G. N., Winter, J., Cato, E. P., Ritchie, A. E. & Bokkenheuser, V. D. *Clostridium scindens* sp. nov., a human intestinal bacterium with desmolytic activity on corticoids. *Int. J. Syst. Evol. Microbiol.* **35**, 478–481 (1985).
- Ridlon, J. M. et al. *Clostridium scindens*: a human gut microbe with a high potential to convert glucocorticoids into androgens. *J. Lipid Res.* **54**, 2437–2449 (2013).
- Boyd, R. A. et al. The pharmacokinetics and pharmacodynamics of diltiazem and its metabolites in healthy adults after a single oral dose. *Clin. Pharmacol. Ther.* **46**, 408–419 (1989).
- Molden, E., Åberg, A. & Christensen, H. Desacetyl-diltiazem displays severalfold higher affinity to CYP2D6 compared with CYP3A4. *Drug Metab. Dispos.* **30**, 1–3 (2002).
- Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
- Maier, L. et al. Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature* **555**, 623–628 (2018).
- Wu, H. et al. Metformin alters the gut microbiome of individuals with treatment-naïve type 2 diabetes, contributing to the therapeutic effects of the drug. *Nat. Med.* **23**, 850–858 (2017).
- Clayton, T. A., Baker, D., Lindon, J. C., Everett, J. R. & Nicholson, J. K. Pharmacometabolicomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism. *Proc. Natl. Acad. Sci. USA* **106**, 14728–14733 (2009).
- Björkholm, B. et al. Intestinal microbiota regulate xenobiotic metabolism in the liver. *PLoS ONE* **4**, e6958 (2009).
- Wishart, D. S. et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).

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

© The Author(s), under exclusive licence to Springer Nature Limited 2019

METHODS

Chemicals. Screened drugs were picked from the Pharmakon 1600 library (MicroSource Discovery Systems) and pooling was performed by the Yale Center for Molecular Discovery. Compounds for follow-up studies were purchased individually (see Supplementary Table 20 for details). LC–MS-grade solvents were from Fisher Scientific and other chemicals were from Sigma Aldrich, if not otherwise indicated.

Bacterial culture conditions. Bacterial strains used in this study are listed in Supplementary Table 1.

Anaerobic culture conditions. A flexible anaerobic chamber (Coy Laboratory Products) containing 20% CO₂, 10% H₂ and 70% N₂ was used for all anaerobic microbiology steps. All anaerobic culturing was performed on brain–heart infusion (BHI; Becton Dickinson) agar supplemented with 10% horse blood (Quad Five). Liquid cultures of bacterial gut isolates and whole communities for drug degradation assays were grown in gut microbiota medium (GMM)²². To make unmarked deletion and complementation strains, *B. thetaiotaomicron* VPI-5482 (ATCC 29148)-derived strains were grown anaerobically at 37 °C in liquid TYG medium and on TYG agar supplemented with haemin and vitamin K²³. For selection, gentamicin 200 µg/ml, erythromycin 25 µg/ml and/or 5-fluoro-2-deoxyuridine (FUDR) 200 µg/ml were added as indicated. CFU counting to determine mouse gut colonization levels and in vitro culture densities was performed anaerobically on BHI blood agar.

Aerobic culture conditions. *E. coli* strains for molecular cloning were grown aerobically at 37 °C in LB medium (200 r.p.m. shaking) and on LB agar supplemented with carbenicillin (100 µg/ml) or kanamycin (50 µg/ml).

Preparing gain-of-function libraries, strain pooling and hit validation. Bacterial strains, plasmids, and primers are listed in Supplementary Tables 1 and 12.

Preparing gain-of-function libraries. Heterologous expression libraries were prepared as previously described²⁴. In brief, genomic DNA was extracted from overnight cultures of the source bacterial strain²⁵. DNA was sheared to 2–8 kb by focused ultrasonication (Covaris E220 with miniTUBE red) and fragments were cloned into PCR-linearized expression vector pZE21 (primer no. 1 and 2) by blunt-ended ligation (Epicentre FastLinkTM kit). Before transformation, the ligation products were separated on a 0.5% agarose gel, the region between 5 and 10 kb was excised, and DNA was extracted using a gel extraction kit (Qiagen). Ligation products were transformed into *E. cloni* 10G Elite competent cells (Lucigen) by electroporation. Overnight grown colonies were picked and arrayed in 384-well format in liquid LB medium supplemented with kanamycin using a colony-picking robot (Molecular Devices QPix 420). After incubation overnight at 37 °C, plates were replicated in duplicate onto LB agar plates supplemented with kanamycin. The first plate served for the initial drug assay (described below in ‘Bacterial drug metabolism assays’) to identify plates that contained drug-metabolizing gain-of-function hits. The second plate was stored at 4 °C for use in the secondary assay to localize drug-metabolizing clones within active plates as described in ‘Bacterial drug metabolism assays’. Primer numbers 3 and 4 were used for Sanger sequencing of representative library clones and identified drug-metabolizing clones.

Hit validation by targeted gene expression in *E. coli*. To validate identified drug-metabolizing gene products from the gain-of-function screen, gene sequences were PCR-amplified (primer no. 5 to 54 and 95 to 150), cloned into the pZE21 expression plasmid by either blunt-end ligation (T4 polynucleotide kinase and ligase, NEB) or Gibson cloning (NEBuilder HiFi DNA Assembly Kit, NEB), and electro-transformed *E. cloni* 10G Elite cells. Resulting strains were tested for specific drug-metabolizing activities.

Construction of *B. thetaiotaomicron* targeted mutants and complementation strains. Bacterial strains, plasmids, and primers are listed in Supplementary Tables 1 and 12.

Gene deletions and complementations. A counter-selectable (FUDR, Sigma Aldrich) allelic exchange procedure²⁶ was used to generate in-frame, unmarked deletions in a *B. thetaiotaomicron* VPI-5482 tdk background (wild type) as previously described⁵. In brief, primer numbers 55 to 60, 63 to 68, 71 to 76, 79 and 80 were used to generate pExchange-tdk suicide plasmids²⁶ for deletion of *bt2068*, *bt2367* and *bt4096* by splicing by overlap extension PCR²⁷, cloning via restriction sites BamHI and XbaI, plasmid sequencing and PCR screening of resolved clones after second DNA recombination event. Gene complementations at various expression levels were performed as previously described^{5,28} using primer numbers 61, 62, 69, 70, 77, 78, 81 and 82 and pNBU2-derived plasmids that integrate into the genome in single copy. *bt4096* complementation using the highest-expressing promoters was not successful, possibly owing to enzyme toxicity at high levels of expression.

Bacterial drug-metabolism assays. **Drug assays with axenic cultures and faecal communities.** Frozen glycerol stocks of bacterial strains (Supplementary Table 1) were plated on BHI blood agar and incubated at 37 °C under anaerobic conditions. Single colonies were inoculated into 6 ml pre-reduced GMM (supplemented with 1% w/v arginine for *Eggerthella lenta*) and incubated anaerobically at 37 °C for 24 h (for *Akkermansia muciniphila*, 48 h). Drug-conversion assays were

performed as previously described^{5,22}. In brief, bacterial cultures were diluted into fresh, pre-reduced GMM (1/5) containing the tested drug(s) at 2 µM, incubated anaerobically until samples were collected and samples were stored at –80 °C until further processing for analysis by LC–MS (see ‘Mass spectrometry analysis of drugs and metabolites’).

Gain-of-function screen. All 384 colonies of a single arrayed library agar plate (see ‘Preparing gain of function libraries’) were collected en masse by scraping, and resuspended in 750 µl of GMM (1/5). Two hundred and twenty-five microlitres of the cell suspension and an eightfold dilution thereof were combined with 25 µl of GMM (1/5) with a drug mixture and incubated anaerobically at 37 °C. Twenty-microlitre samples were collected after 0, 1, 2, 4, 6, 8, 12 and 24 h of incubation. Plates corresponding to pools that exhibited the capacity to metabolize one or more of the tested drugs were replicated into a 384-well plate containing 70 µl of LB supplemented with kanamycin and cultures were grown aerobically for 12 h at 37 °C. For groups of 2 × 2 active plates, pools were assembled from 20 µl of each culture corresponding to each row (2 × 16 pools) and to each column (2 × 24 pools) and tested for drug-metabolizing activity. Identified active gain-of-function clones were colony-purified, four independent colonies were retested for drug-metabolizing activity and two of the verified clonal cultures were selected for Sanger sequencing of the plasmid inserts (primer no. 3 and 4).

Cloning, purification and enzymatic activity testing of BT_4096 and BT_2367.

BT_4096. Sequence analysis of *bt4096* using the SignalP 4.0 server²⁹ suggested an N-terminal export signal with a cleavage site between residues 20 and 21. The open reading frame without this signal sequence was PCR-amplified (primer no. 85 and 86) and cloned into the pASG-IBA105 (IBA) expression plasmid (N-terminal Twin-Strep-Tag protein fusion and anhydrotetracycline-inducible expression) following the vendor’s protocol for StarGate cloning. *E. coli* (*E. cloni* 10G, Lucigen) containing a sequence-verified plasmid (primer no. 87 and 88) was aerobically cultured (220 r.p.m. shaking) in 1 l LB medium with carbenicillin, and protein expression was induced with anhydrotetracycline (200 ng/ml) during mid-exponential growth (optical density at 600 nm, OD₆₀₀ = 0.4–0.6). After 3 h, bacteria were collected by centrifugation (4,000g) and lysed on ice by sonication (8 pulses of 15 s at 45% amplitude in intervals of 60 s) in the presence of protease inhibitors (cComplete protease inhibitor cocktail, Sigma Aldrich). Affinity purification was performed using a gravity-flow Strep-TactinXT Superflow column (IBA) following the manufacturer’s recommendations. Protein purity was verified by SDS-PAGE and Coomassie staining, and quantified by Bradford protein assay using BSA as a standard (Biorad). Enzyme assays were performed in Tris-HCl buffer (10 mM, pH 7), supplemented with 2.5 mM MgSO₄ and MnCl at 37 °C; reaction volumes were 150 µl with BT_4096 at 5 µg/ml and varying substrate concentrations: 250, 125, 62.5, 31.3, 15.6, 7.8, 3.9 and 0 µM. Ten microlitres of sample was collected and quenched in ice-cold acetonitrile (10 µl) at times 30, 60, 90, 120, 150, 180, 210, 240, 300, 360, 480 and 600 s of incubation. Samples were stored at –80 °C until further processing and analysis by LC–MS (see ‘Mass spectrometry analysis of drugs and metabolites’).

BT_2367. The *bt2367* open reading frame was PCR-amplified (primer no. 83 and 84) and cloned into the pASG-IBA103 (IBA) expression plasmid (C-terminal Twin-Strep-Tag protein fusion and anhydrotetracycline-inducible expression) following the vendor’s protocol for StarGate cloning. *E. coli* (*E. cloni* 10G, Lucigen) containing a sequence-verified plasmid (primer no. 87 and 88) was used to express and purify the enzyme as described in ‘BT_4096’. Enzyme assays were performed in Tris-HCl buffer (10 mM, pH 7), supplemented with 2.5 mM MgSO₄ and MnCl at 37 °C; reaction volumes were 100 µl with BT_2367 at 10 µl/ml and 1 mM of cofactor (either acetyl-CoA or propionyl-CoA) and 100 µM of substrate (either pericyazine, cyamemazine or 1-(3-aminopropyl)piperidin-4-ol) were used to test acyltransferase activity. Five microlitres of sample was collected and quenched in ice-cold acetonitrile (15 µl) at times 0, 2.5, 5, 7.5, 10, 15, 20, 25, 30, 40, 50 and 60 min of incubation. Samples were stored at –80 °C until further processing and analysis by LC–MS and LC–MS/MS (see ‘Mass spectrometry analysis of drugs and metabolites’).

Quantitative PCR analysis. Faecal DNA was extracted from a biomass pellet of 500-µl community cultures (see ‘Bacterial drug-metabolism assays’) as previously described^{25,30}. The abundance of specific bacterial species or *bt4096* homologues was assessed by quantitative PCR (primer no. 89 to 94) as previously described^{9,31}. A CFX96 instrument (BioRad) and SYBR FAST universal master mix (KAPA Biosystems) were used.

Animal experiments. All experiments using mice were performed using protocols approved by the Yale University Institutional Animal Care and Use Committee in accordance with the highest scientific, humane and ethical principles and in compliance with federal and state regulations, including the Animal Welfare Act (AWA) and the Public Health Service (PHS).

Germ-free 9–16-week-old C57BL/6J mice were maintained in flexible plastic gnotobiotic isolators with a 12-h light/dark cycle and germ-free status monitored by PCR and culture-based methods. Conventional C57BL/6J mice (Jackson

Laboratories) were purchased at the age of 6–7 weeks and kept in the laboratory for 2–3 weeks before experiments. All mice were provided a standard, autoclaved mouse chow (5013 LabDiet, Purina) ad libitum.

Dexamethasone treatment. Serum kinetics of dexamethasone following oral administration was determined using 20 conventional C57BL/6J mice ($n = 4$ per time point) treated with 10 mg/kg of dexamethasone suspension in PBS. One blood sample was collected from each mouse into lithium heparin tubes (BD Life Sciences) by submandibular bleeding, and a second sample was collected at time of euthanization. Serum was collected by centrifugation (2,500 r.c.f., 4°C for 10 min) of heparinized blood and stored at –80°C until further processing and analysis by LC–MS (see ‘Mass spectrometry analysis of drugs and metabolites’).

To compare dexamethasone metabolism between germ-free mice and mice mono-colonized with *C. scindens*, individually caged germ-free C57BL/6J mice were either directly treated with dexamethasone (as above) or colonized with *C. scindens* by oral gavage of 200 µl of an overnight bacteria culture in GMM. After four days, bacterial loads were determined by CFU plating on BHI blood agar before dexamethasone was orally administered to the mice as described above. Mice were killed seven hours after drug administration, serum was collected as described above, and tissue samples were collected into sample tubes and snap-frozen. Faecal samples were collected before euthanization and re-suspended in PBS (1 ml) through vigorous shaking. Twenty microlitres were then plated on BHI blood agar plates and incubated aerobically and anaerobically at 37°C to check mice for contamination.

Diltiazem treatment. Germ-free C57BL/6J mice were mono-colonized with *B. thetaiotaomicron* wild-type or *bt4096*-mutant strain, and bacterial loads were determined by CFU plating four days after colonization as described in ‘Dexamethasone treatment’. For the single-dose treatment, 5 mice per time point and group were treated with 50 mg/kg of diltiazem suspension in PBS with 20% glycerol or solely vehicle for non-treated controls ($n = 5$ per group). One early blood sample (at 0.5, 1, 1.5, 2 and 2.5 h after drug administration) was collected from each mouse by submandibular bleeding, and a second sample was collected when mice were killed (at 3, 5, 7, 9 and 12 h after drug administration). In addition to serum, the following tissues were collected after euthanization: luminal content of duodenum (SI), jejunum (SII), ileum (SIII), caecum and colon, and faeces, liver and bile. Samples were collected and stored as described in ‘Dexamethasone treatment’. For the multiple dose treatment, 6 and 5 mice mono-colonized with *B. thetaiotaomicron* wild-type or *bt4096*-mutant strain, respectively, were treated 5 times with 50 mg/kg of diltiazem suspension in PBS with 20% glycerol in intervals of 6 h. Six hours after the final treatment, a blood sample was collected from each mouse by submandibular bleeding and mice were killed for sample collection after an additional 6 h.

Mass spectrometry analysis of drugs and metabolites. *Extraction of solid tissues and liquid samples.* Solid tissues and liquid samples were prepared for LC–MS and LC–MS/MS analysis by organic solvent extraction (acetonitrile:methanol, 1:1) at –20°C after the addition of internal standard mix (sulfamethoxazole, caffeine, ipriflavone and yohimbine each to a final concentration of 80 nM) as previously described^{5,30}.

LC–MS and LC–MS/MS analysis. Analyses was performed as previously described^{5,30}. In brief, chromatographic separation was performed by reversed-phase chromatography (C18 Kinetex Evo column, 100 mm × 2.1 mm, 1.7-mm particle size, and according guard columns, Phenomenex) using an Agilent 1200 Infinity UHPLC system and mobile phase A (H₂O, 0.1% formic acid) and B (methanol, 0.1% formic acid), and the column compartment was kept at 45°C. Five microlitres of sample was injected at 100% A and 0.4 ml/min flow, followed by a linear gradient to 95% B over 5.5 min and 0.4 ml/min. To ensure reproducible chromatographic separation (retention shifts between samples <2% or 0.15 min) columns were changed after 1,000 sample injections. The qTOF instrument (Agilent 6550) was operated in positive scanning mode (50–1,000 *m/z*) with the following source parameters: VCap, 3,500 V; nozzle voltage, 2,000 V; gas temperature, 225°C; drying gas 13 l/min; nebulizer, 20 psig; sheath gas temperature 225°C; sheath gas flow 12 l/min. Online mass calibration was performed using a second ionization source and a constant flow (5 µl/min) of reference solution (121.0509 and 922.0098 *m/z*). LC–MS/MS was performed using the chromatographic separation and source parameters described above, and the auto-MS/MS mode of the instrument with a preferred inclusion list for parent ions with 20 p.p.m. tolerance, iso width set to ‘narrow width’ and collision energy to either 15, 20 or 30 eV. The MassHunter Quantitative Analysis Software (Agilent, version 7.0) was used for peak integration based on retention time and accurate mass measurement of chemical standards. Quantification of *in vivo* samples was based on dilution series of chemical standards spanning 0.001 to 10 µM and measured amounts were normalized by weights of extracted tissue samples. The MassHunter Qualitative Analysis Software (Agilent, version 7.0) and Mass Profiler Professional (Agilent, version 13.0) with standard parameters were used for untargeted feature extraction and peak alignment, respectively, allowing tolerances for mass of 0.002 AMU or 20 p.p.m. and for retention time of 0.15 min or 2%. Statistical analysis and plotting were performed in Matlab 2017b (MathWorks).

Analysis of drug metabolism screen. Identification of metabolized drugs. For each bacterial strain, drug fold changes were calculated between time points 12 h and 0 h in the 4 pools that contained a specific drug (Supplementary Table 2), and between these drug-containing pools and the 3 non-drug controls at time points 0 h and 12 h. Statistical significance of the drug intensity differences was assessed with two-sided *t*-test (ttest2 function in MatLab), and *P* values were FDR-corrected for multiple hypotheses testing using the Benjamini–Hochberg procedure (mafdr function in MatLab with BHFD parameter). To account for fast drug metabolism (within seconds after exposure), fold changes to control at time point 0 were used for drug and strain combinations for which (i) $\log_2(\text{fold change to control at } t = 0) < -5$; (ii) FDR-corrected *P* value (fold change to control at $t = 0$) < 0.05; and (iii) $\log_2(\text{fold change to control at } t = 0) < \log_2(\text{fold change at } t = 12 \text{ h to } t = 0)$. To account for variability in drug measurements, for each drug an adaptive fold-change threshold was calculated as either (20%) or (mean + 2 s.d. of the fold changes, for which $\log_2(\text{fold change at } t = 12 \text{ h to } t = 0) > 0$, to account for measurement noise), whichever was greater. Hierarchical clustering was performed with the clustergram function in MatLab using Euclidean distance between the drug fold-change vectors for each bacterial strain.

Identification of candidate drug metabolites. For each drug and each bacterial strain, intensities of all compounds detected in the drug-containing pools ($n = 4$) versus all other pools ($n = 20$) were compared by calculating the fold change and statistical significance with two-sided *t*-test (ttest2 function in MatLab). All *P* values were FDR-corrected for multiple hypotheses testing with the Benjamini–Hochberg procedure (mafdr function in MatLab with BHFD parameter). The FDR-corrected *P* value threshold for candidate drug metabolites was set to 10^{-6} based on the distribution of *P* values of parent drugs in their corresponding 4 pools versus all other pools (combinatorial pooling scheme shown in Extended Data Fig. 1a, Supplementary Table 2). In brief, histograms of the FDR-corrected *P* values calculated for the candidate metabolites were compared with the histograms of the FDR-corrected *P* values of the drugs detected in the corresponding drug pools (positive controls) and the FDR-corrected *P* values of metabolites calculated for a random pooling scheme not included in the experiment (negative controls). The FDR-corrected *P* value distributions appeared to be bimodal, and the findpeaks function in MatLab2017b was used to find the FDR-corrected *P*-value threshold that corresponds to the local minimum between the distribution peaks, thus separating high-confidence FDR-corrected *P* values prevalent in the positive controls. A detailed analysis scheme and distribution plots are available on figshare (<https://doi.org/10.6084/m9.figshare.8119058>); analysis scripts are available on GitHub (https://github.com/mszimmermann/drug-bacteria-gene_mapping).

Candidate metabolites were filtered according to the following exclusion criteria: (i) metabolite intensity in the drug pools at $t = 0$ is $>10^4$ ion counts (corresponding to twofold of the minimal intensity for chromatographic feature extraction, 5×10^3 ion counts), and $\log_2(\text{fold change to other pools at } t = 0) < 1$; (ii) metabolite intensity in the drug pools at $t = 12$ h is $>10^4$ ion counts, and $\log_2(\text{fold change to other pools at } t = 12 \text{ h}) < 1$; (iii) the difference of mass defects between the drug and the metabolite is >0.2 AMU; (iv) retention time difference between the drug and the metabolite is <0.1 min; or (v) metabolite mass is similar to one of the metabolites filtered at step (iv) (mass difference <0.002 AMU) and retention time is similar to an unfiltered metabolite of the same drug (difference <0.1 min). Additionally, metabolites were filtered based on whether the parent drug was metabolized, whether the metabolite was identified for a single drug or multiple drugs, and whether the metabolite was increasing in at least 1 strain at 12 h compared to 0 h (fold change (expressing in \log_2) > 1 and FDR-corrected *P* value <0.05). Mass difference between the parent drug and candidate metabolites were calculated for each drug–metabolite pair and smoothed with 0.002-Da window.

Chemoinformatics. Chemical similarity analysis. Chemical similarity analysis between the 271 selected drugs and the 2,099 clinically approved drugs from DrugBank²¹ was performed using Morgan chemical fingerprints calculated with AllChem.GetMorganFingerprintAsBitVect function from the RDKit AllChem module (Open Source Chemoinformatics, <http://www.rdkit.org>). Chemical fingerprints were converted into a binary matrix and subjected to principal component analysis. Chemical structure similarity for the drug clusters identified with hierarchical clustering was performed using the maximum common substructure function in RDKit (rdFMCS.FindMCS).

Functional chemical group analysis. For each drug, the existence of chemical functional groups was calculated with available functions from the rdkit.Chem.Fragments module (Open Source Chemoinformatics, <http://www.rdkit.org>). Functional group enrichment among metabolized drugs, drugs belonging to specific clusters or drugs undergoing specific mass difference transformation upon bacterial metabolism was calculated with Fisher’s exact test. All *P* values were FDR-corrected for multiple hypotheses testing using the Benjamini–Hochberg procedure (mafdr function in MatLab with BHFD parameter).

Analysis of drug-metabolizing capacity of bacterial strains and communities. Protein sequence alignments. The genomic sequences of bacterial strains and

protein sequences of the candidate genes were downloaded from NCBI. Sequence similarity between the genomes and proteins of interest was assessed by searching the translated nucleotide database using protein query (tblastn) with default parameters³². Sequence similarity between candidate proteins was assessed with protein blast (blastp) with default parameters³². Protein similarity networks and drug–gene–metabolite networks were visualized in Cytoscape v.3.4.0³³.

Gene-set enrichment analysis. Enrichment of strains encoding a given protein among strains metabolizing a given drug was assessed using the gene-set enrichment procedure¹⁶. In brief, for each drug, the strains were sorted according to the per cent of parent drug that was metabolized, and enrichment of strains encoding y per cent similar sequence to the protein of interest was calculated with Fisher's exact test for each set of strains metabolizing more than x per cent of the drug, in which x ranged between 100% and 20% (with step size = 20%) and y ranged between 100% and 50%. For enrichment analysis of gene combinations, maximal sequence similarity to the corresponding genes was used. The lowest P value was recorded for each drug–gene pair. All P values were FDR-corrected for multiple hypotheses testing using the Benjamini–Hochberg procedure (mafdr function in MatLab with BHFD parameter).

Analysis of microbial community drug metabolism. For the human gut communities, drug metabolism of each community was represented with conversion slopes of the drug and the corresponding drug metabolite. To assess the velocities of drug consumption and drug metabolite production, the concentration slope was calculated by fitting a piecewise linear function to the corresponding concentration curves with polyfit function in MatLab 2017b. Correlation between drug consumption or metabolite production slopes and specific gene abundance, bacterial 16S abundance or bacterial CFU ml⁻¹ was calculated with corr function in MatLab 2017b.

Metagenomic analysis. *Sample preparation and sequencing.* DNA from bacterial community samples was extracted as described in ‘Quantitative PCR analysis’. Library preparation and sequencing were performed at the Yale Center for Genome Analysis. The Kapa Biosystems Hyper prep kit WGS was used for the preparation of the metagenomic sequencing libraries. The 2 × 150-bp sequencing was performed on an Illumina Novaseq 6000 instrument with a S4 flow-cell to target depth of ~20 Mio reads per sample.

Data preprocessing. Metagenomic sequencing data preprocessing and analysis was performed with bioBakery tools³⁴. Paired-end reads from each sample were filtered with KneadData v.0.6.1 to trim adaptor sequences with trimmomatic-0.38 and to exclude reads mapping to the human genome³⁵. Filtered paired-end reads were merged before further processing.

Operational taxonomic unit construction and taxonomic assignment. Operational taxonomic unit (OTU) construction and taxonomic assignment was performed with MetaPhlAn2 v.2.6.0³⁶ on the filtered and merged sequencing data. OTU tables were subsequently merged into a summary OTU table with merge_metaphlan_tables function (Supplementary Table 17).

Diversity analysis. Diversity analysis was performed with Qiime 1.8³⁷. OTU tables from metagenomic analysis were converted to biom tables with biom -convert functions. Shannon's alpha-diversity metric was calculated with alpha_diversity_py function with -m shannon parameter.

Quantification of protein sequence abundance. Protein sequence abundance quantification was performed with ShortBRED v.0.9.5³⁸. Target protein sequences were downloaded from NCBI in amino acid .fasta format. ShortBRED markers were created with shortbred-identify function using Uniref90 (downloaded from <https://www.uniprot.org/downloads>) as a reference³⁹. Protein sequence abundance was quantified using the created reference markers with shortbred_quantify function using built in USEARCH v.11.0.667_i86linux32 tool⁴⁰. The resulting tables for each sample were merged into one summary table (Supplementary Table 18).

Correlation analysis. Correlation analysis between drug or metabolite conversion slopes was performed in MatLab2017b with corr function (resulting in Pearson's correlation coefficient and P value calculated using Student's t -distribution for statistics $t = r\sqrt{((n-2)/(1-r^2))}$ with $n - 2$ degrees of freedom, in which r is the correlation coefficient and n is the sample size) (Supplementary Table 19).

Metagenomics analysis workflows are available on figshare (<https://doi.org/10.6084/m9.figshare.8119058>).

Human faecal material. All samples were collected under Yale University Human Investigation Committee protocol number 1106008725 at Yale University School of Medicine and stored in the Goodman laboratory, using identifier numbers that are not associated with study volunteer names or other identifying information. Study volunteers were recruited through the campus billboard advertisement and reviewed an information sheet describing the study before deciding to participate. Eligibility criteria included age (20–60 years) and general health status (generally healthy, no long-term chronic diseases or illnesses, no cold, no flu or apparent bacterial or viral infection at the time of contact, not currently on antibiotics or no antibiotic use in the past two months, and regular bowel movement). No information was collected from the subjects except for their age and gender

(Supplementary Table 9 and 16). Faecal samples were then collected and stored as previously described²².

Statistics and reproducibility. Statistical analysis of all data was performed in MatLab R2017b. No statistical methods were used to predetermine sample size. For mouse experiments, mice were randomized before allocation to study groups and respective cages. All other experiments conducted for the study were not randomized, and the investigators were not blinded to either allocation during experiments or to outcome assessments.

All in vitro experiments were performed once with indicated replication. Mouse experiments in Fig. 3f and Extended Data Figs. 5a, 8b were performed once, in Extended Data Fig. 5b twice, and in Fig. 3e and Extended Data Fig. 8a three times with indicated replication (Supplementary Tables 7, 8, 10), which resulted in comparable observations between repeats. In Fig. 3e, f and Extended Data Figs. 5a, b, 8, horizontal lines represent mean values of independent animals. P values were calculated with a two-sided unpaired Student's t -test, and FDR-corrected for multiple hypotheses testing with the Benjamini–Hochberg procedure. The exact number of mice used per time point and calculated FDR-corrected P values per compound and time point are indicated in Supplementary Tables 7, 8, 10, 11.

In Fig. 2a, the volcano plot represents mean fold changes between $n = 4$ drug-containing pools and $n = 20$ non-drug-containing pools (vehicle controls), and P values were calculated with a two-sided unpaired Student's t -test and FDR-corrected for multiple hypotheses testing with Benjamini–Hochberg procedure. In Fig. 5b, d, e, bar plots and error bars represent mean and s.d. of $n = 4$ independent cultures. In Fig. 5f, each colour represents a human donor, and lines depict the mean of $n = 4$ independent samples.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All data generated during this study are included in the published Article and its Supplementary Tables. Data are available from figshare, at <https://doi.org/10.6084/m9.figshare.8119058>. Raw sequencing data have been deposited on the ENA server, with accession number PRJEB31790. Raw metabolomics data have been deposited in the MetaboLights repository, with accession number MTBLS896.

Code availability

Analysis pipeline schemes and input files are available from figshare (<https://doi.org/10.6084/m9.figshare.8119058>); scripts for analysing data and generating figures are available on GitHub (<https://github.com/mszimmermann/drug-bacteria-gene-mapping>) and archived at Zenodo (<https://doi.org/10.5281/zenodo.2827640>).

22. Goodman, A. L. et al. Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proc. Natl Acad. Sci. USA* **108**, 6252–6257 (2011).
23. Holdeman, L. V., Moore, W. E. C. & Cato, E. P. *Anaerobe Laboratory Manual* (Virginia Polytechnic Institute and State Univ., 1977).
24. Forsberg, K. J. et al. The shared antibiotic resistome of soil bacteria and human pathogens. *Science* **337**, 1107–1111 (2012).
25. Goodman, A. L., Wu, M. & Gordon, J. I. Identifying microbial fitness determinants by insertion sequencing using genome-wide transposon mutant libraries. *Nat. Protoc.* **6**, 1969–1980 (2011).
26. Koropatkin, N. M., Martens, E. C., Gordon, J. I. & Smith, T. J. Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure* **16**, 1105–1115 (2008).
27. Warrens, A. N., Jones, M. D. & Lechner, R. I. Splicing by overlap extension by PCR using asymmetric amplification: an improved technique for the generation of hybrid proteins of immunological interest. *Gene* **186**, 29–35 (1997).
28. Whitaker, W. R., Shepherd, E. S. & Sonnenburg, J. L. Tunable expression tools enable single-cell strain distinction in the gut microbiome. *Cell* **169**, 538–546.e12 (2017).
29. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786 (2011).
30. Lim, B., Zimmermann, M., Barry, N. A. & Goodman, A. L. Engineered regulatory systems modulate gene expression of human commensals in the gut. *Cell* **169**, 547–558.e15 (2017).
31. Cullen, T. W. et al. Antimicrobial peptide resistance mediates resilience of prominent gut commensals during inflammation. *Science* **347**, 170–175 (2015).
32. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
33. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
34. McIver, L. J. et al. bioBakery: a meta'omic analysis environment. *Bioinformatics* **34**, 1235–1237 (2018).
35. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
36. Truong, D. T. et al. MetaPhiAn2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* **12**, 902–903 (2015).

37. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
38. Kaminski, J. et al. High-specificity targeted functional profiling in microbial communities with ShortBRED. *PLOS Comput. Biol.* **11**, e1004557 (2015).
39. Suzek, B. E., Wang, Y., Huang, H., McGarvey, P. B. & Wu, C. H. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **31**, 926–932 (2015).
40. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
41. Yeung, P. K. et al. Pharmacokinetics and metabolism of diltiazem in healthy males and females following a single oral dose. *Eur. J. Drug Metab. Pharmacokinet.* **18**, 199–206 (1993).

Acknowledgements We thank the Goodman laboratory for helpful discussions and N. A. Barry, L. Valle, D. Lazo, W. B. Schofield, P. H. Degnan, T. Wu and the Yale Center for Molecular Discovery for technical assistance. This work was supported by NIH grants GM118159, GM105456, AI124275, DK114793, the Center for Microbiome Informatics and Therapeutics, the Burroughs Wellcome Fund, the Yale Cancer Center, and the HHMI Faculty and Pew Scholars Programs to A.L.G. M.Z. received Early and Advanced Postdoc Mobility Fellowships from the Swiss National Science Foundation (P2EZP3_162256 and P300PA_177915, respectively) and a Long-Term Fellowship (ALTF 670-2016)

from the European Molecular Biology Organization. M.Z.-K. received an Early Postdoc Mobility Fellowship from the Swiss National Science Foundation (P2EZP3_178482).

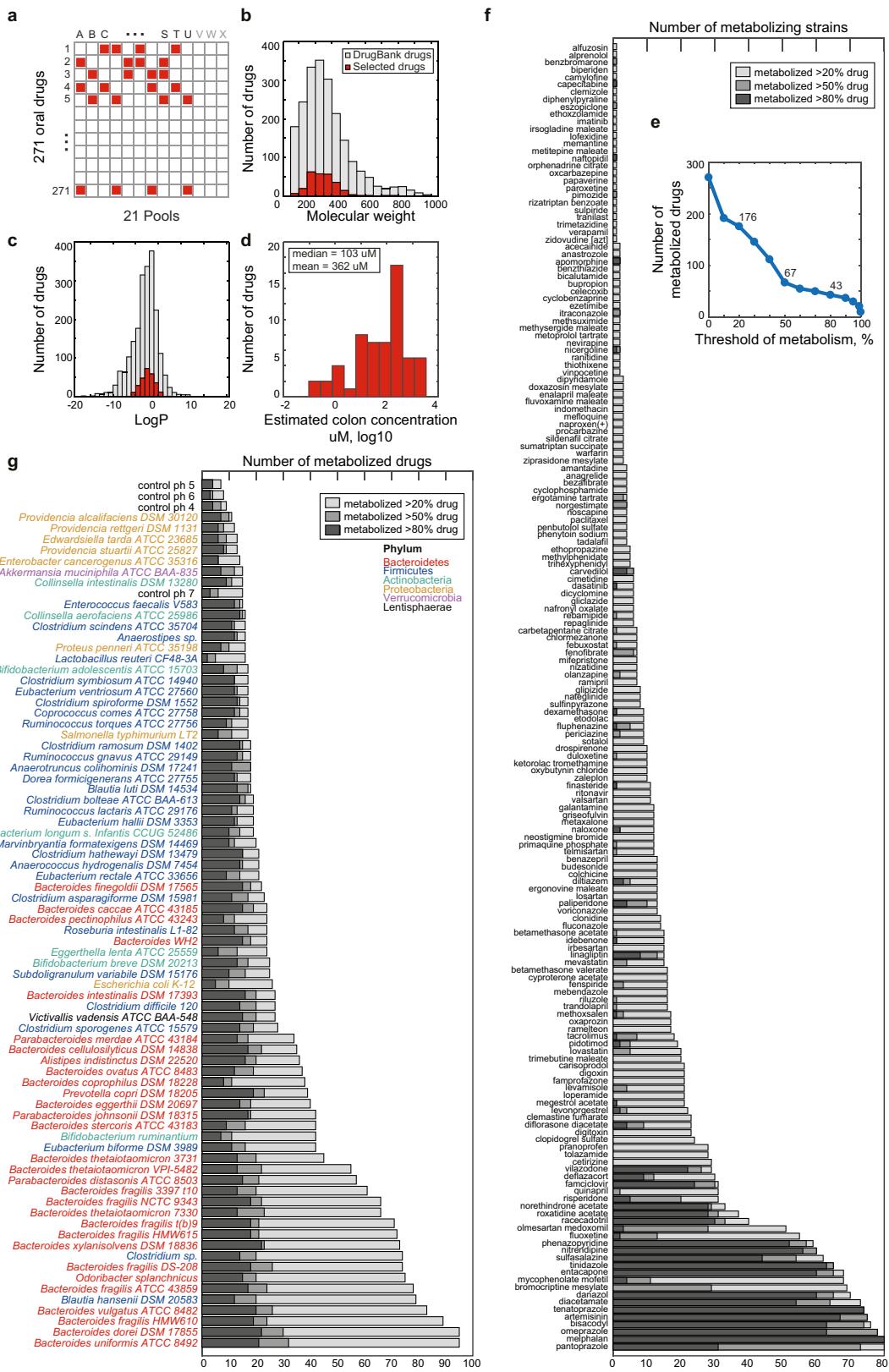
Author contributions M.Z. and A.L.G. conceived and initiated the project; M.Z. performed the experiments; M.Z. and R.W. established the gain-of-function pipeline; M.Z. and M.Z.-K. analysed the data; M.Z.-K. performed statistical analyses and prepared graphical illustrations; and M.Z., M.Z.-K. and A.L.G. wrote the manuscript.

Competing interests M.Z., M.Z.-K. and A.L.G. have filed a patent application based on these studies with the US Patent and Trademark Office (62/693,741). The specific aspects included in the patent application include methods for measuring microbial drug metabolism and for identifying microbial taxa and gene products associated with drug metabolism.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1291-3>.

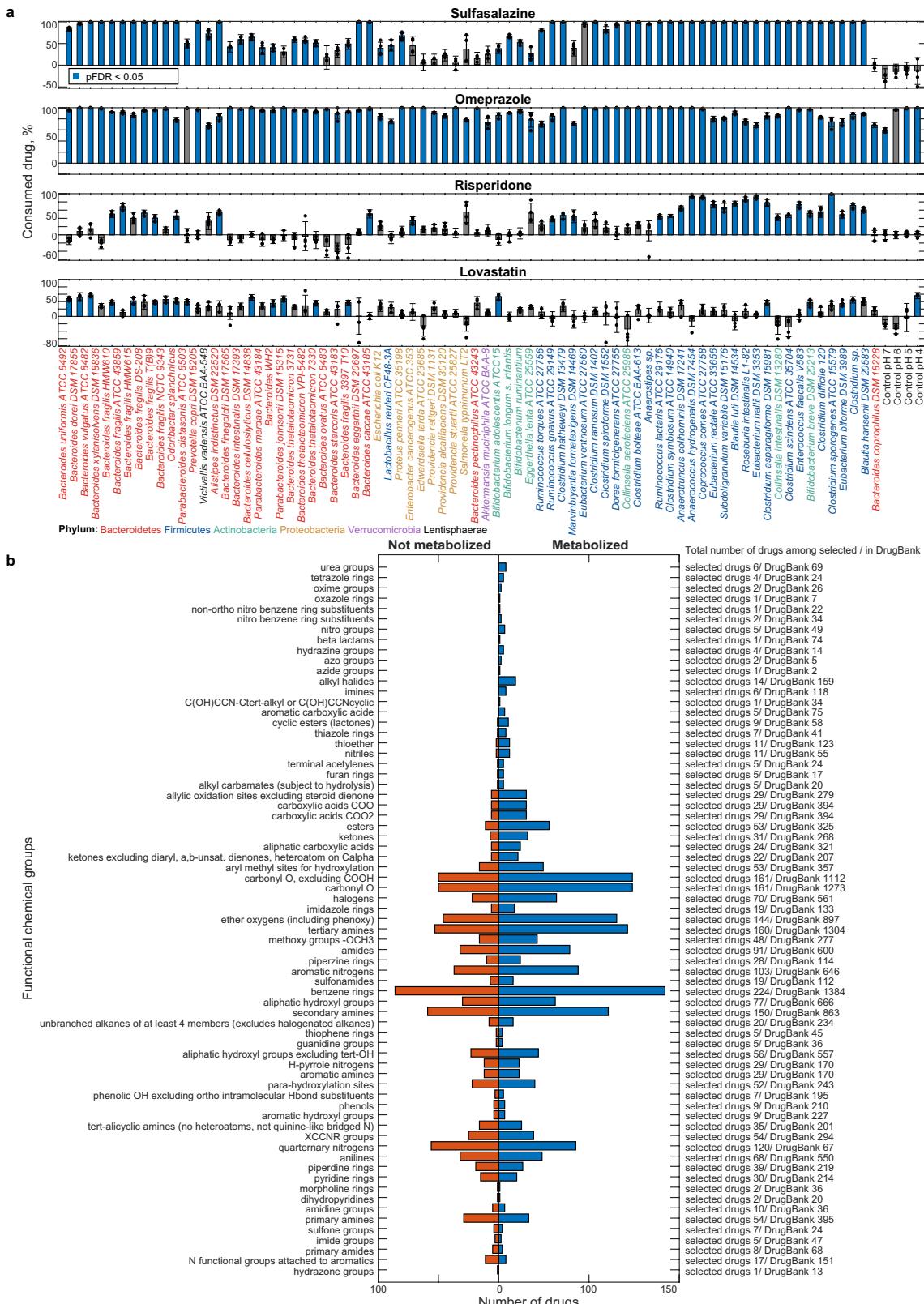
Correspondence and requests for materials should be addressed to A.L.G. **Reprints and permissions information** is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | Setup of drug assay, characterization of tested drugs and summary of metabolic bacteria–drug interactions.

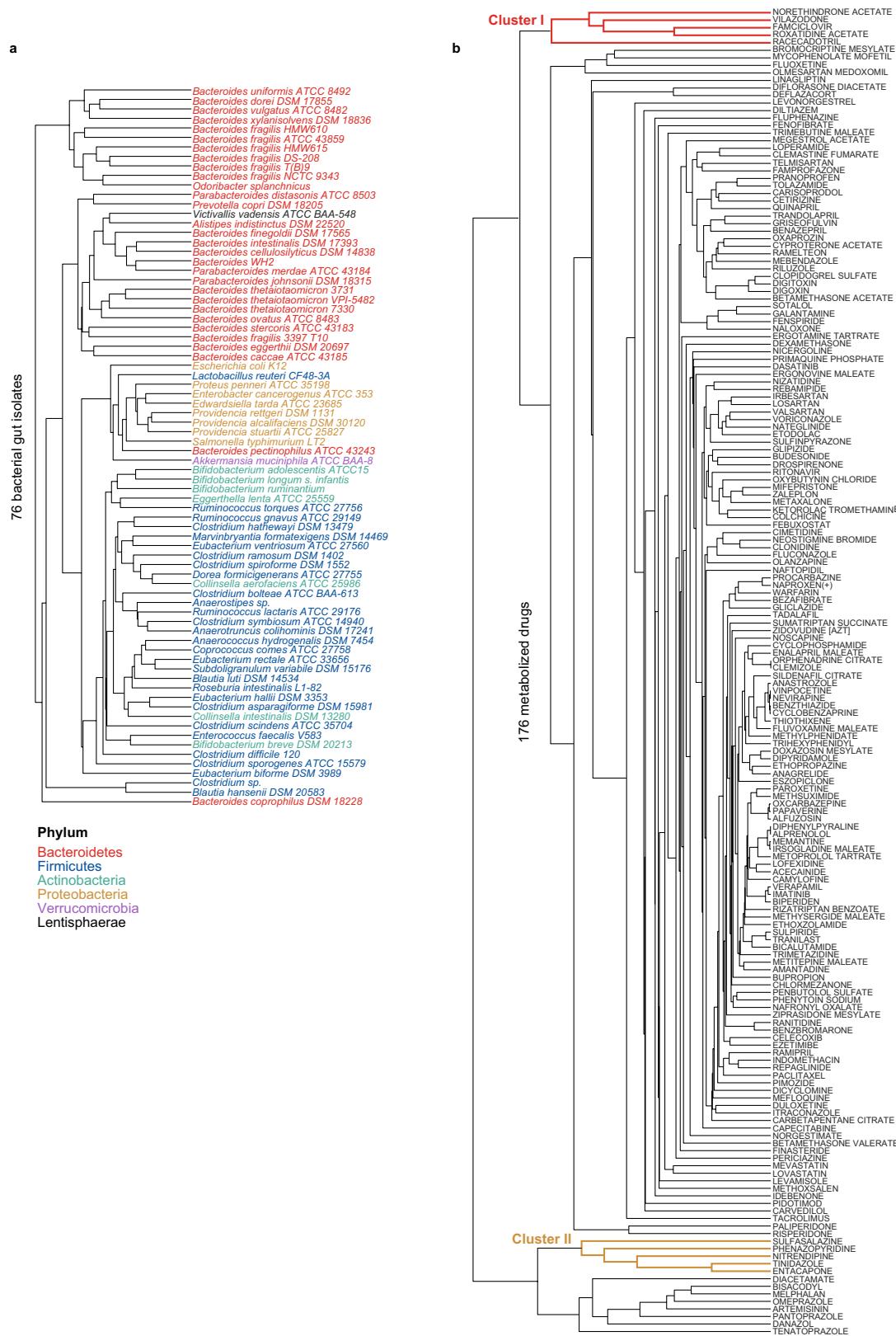
a. Schematic of combinatorial pooling scheme using 21 drug pools (A–U) and 3 non-drug controls (V–X). Each of the 271 drugs is tested in quadruplicate (present in 4 pools) and any 2 drugs are tested in the same pool twice at most (Supplementary Table 2). **b, c.** Molecular mass (**b**) and log *P* value (**c**) distribution of 271 tested drugs (red) in comparison to 2,099 clinically approved drugs (from DrugBank²¹).

d. Distribution of predicted drug concentration in the colon for 58 of the 271 drugs tested (data from a previous study¹⁷). The predicted median and mean concentration in the large intestine for these compounds is 103 μ M and 362 μ M, respectively, when each drug is administered at its standard oral dose. **e.** Number of drugs metabolized as a function of the selection threshold (metabolized fraction). **f.** Number of gut bacteria that metabolize a given drug. **g.** Number of drugs metabolized by each bacterial strain.

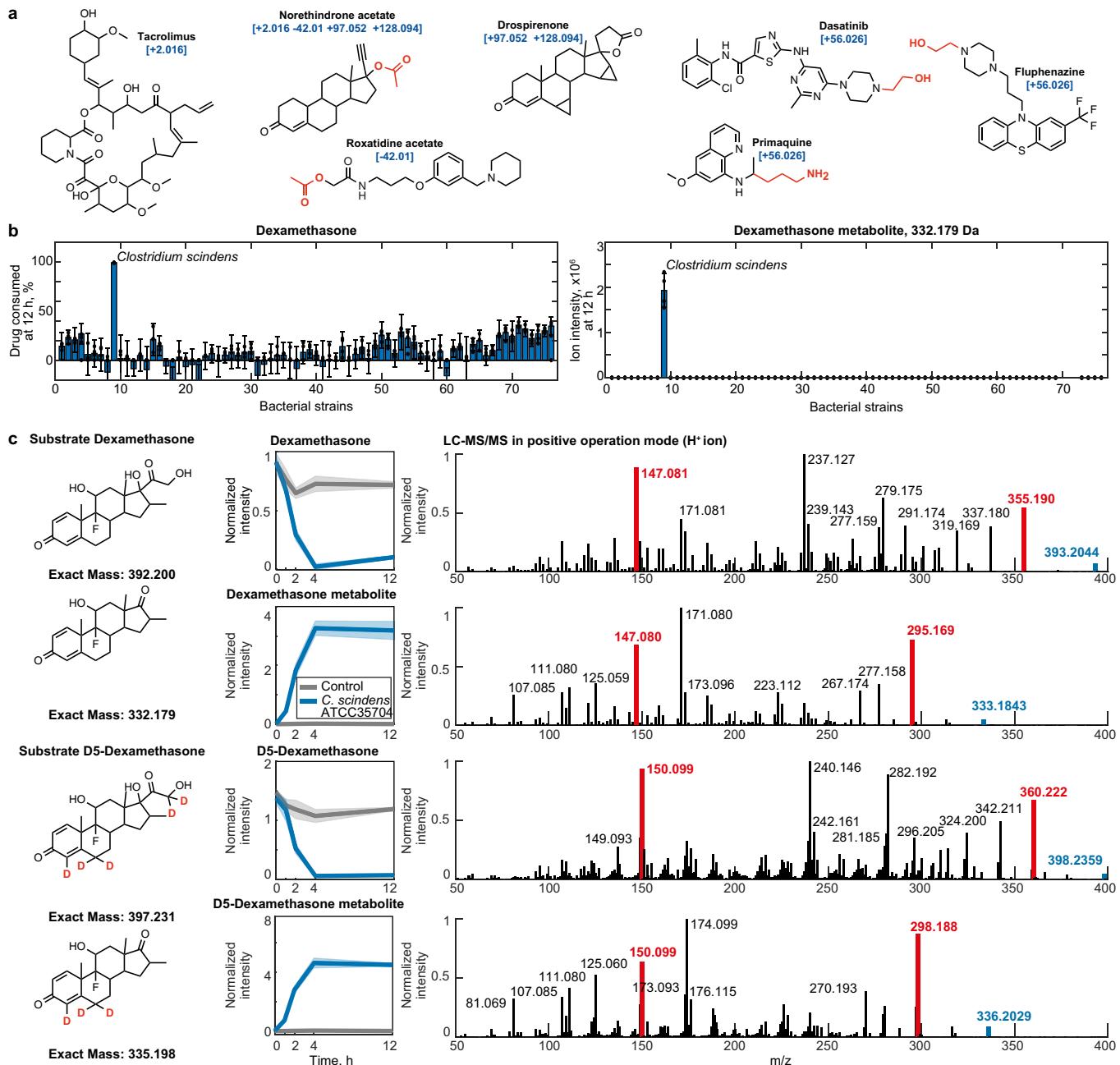


Extended Data Fig. 2 | Metabolism of drugs previously reported to be transformed by bacteria, and functional chemical group distribution. **a**, Per cent of consumption between 0 h and 12 h for each drug after incubation with each gut bacterial species or strain are shown. Bars and error bars depict the mean and s.d. of $n = 4$ assay

replicates. **b**, Distribution of functional chemical groups in drugs that are metabolized or not metabolized across the 76 bacterial strains we tested. The abundance of each chemical group among the 271 selected drugs and 2,099 clinical drugs (from DrugBank²¹) is indicated.

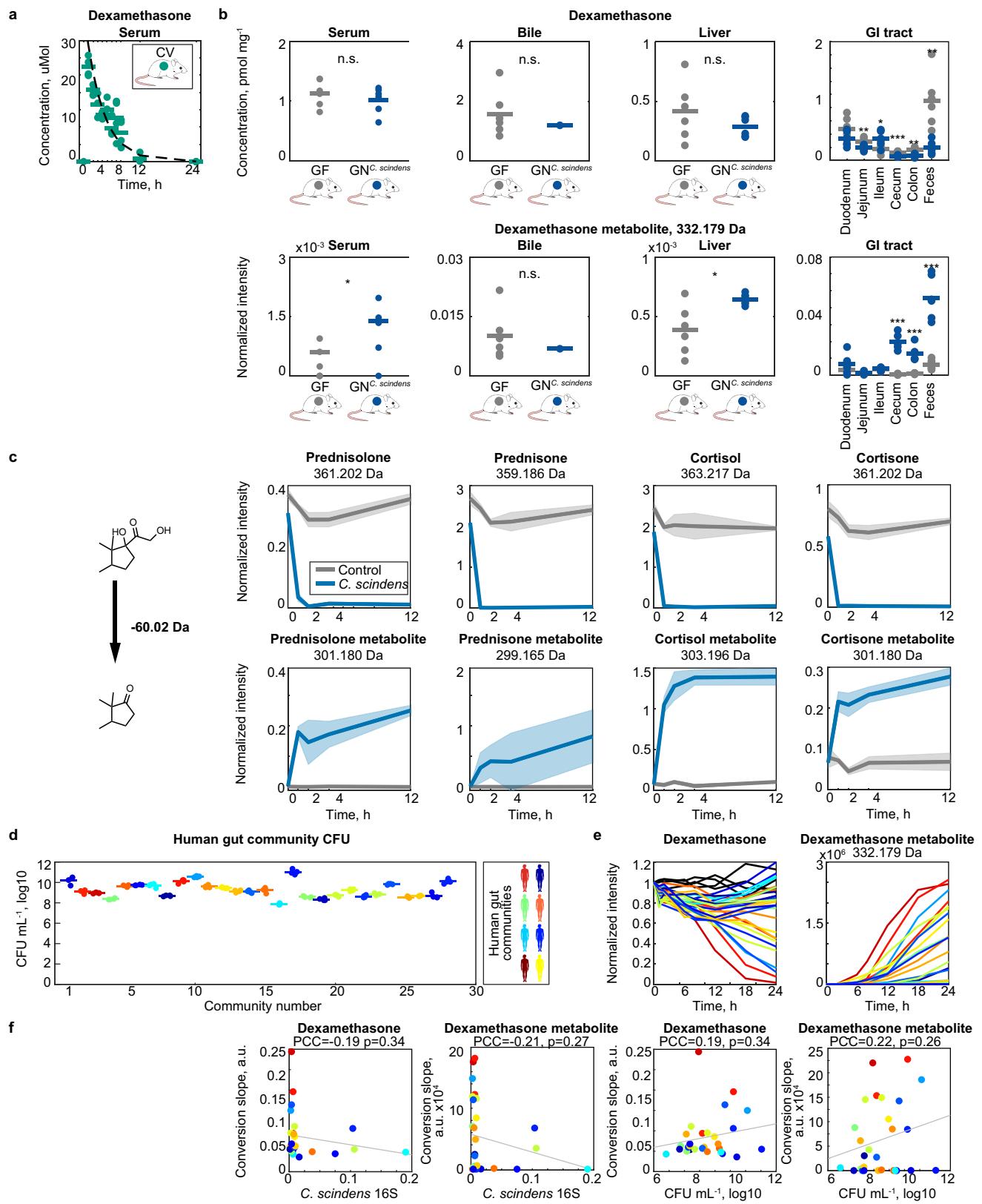


Extended Data Fig. 3 | Hierarchical clustering of bacterial strains or species and drugs according to microbial drug metabolism. **a**, Dendrogram of bacterial strains from Fig. 1c (x axis). **b**, Dendrogram of drugs from Fig. 1c (y axis).



Extended Data Fig. 4 | Structural drug features targeted for biotransformation and microbiome metabolism of dexamethasone.
a, Examples of drugs that are associated with a particular mass shift between the parent drug and its metabolite. Functional groups that are enriched in drugs that undergo specific mass shift (Fig. 2d) are highlighted. **b**, Dexamethasone metabolism by each of the 76 bacterial strains we tested. Bar plots and error bars represent mean and s.d. of

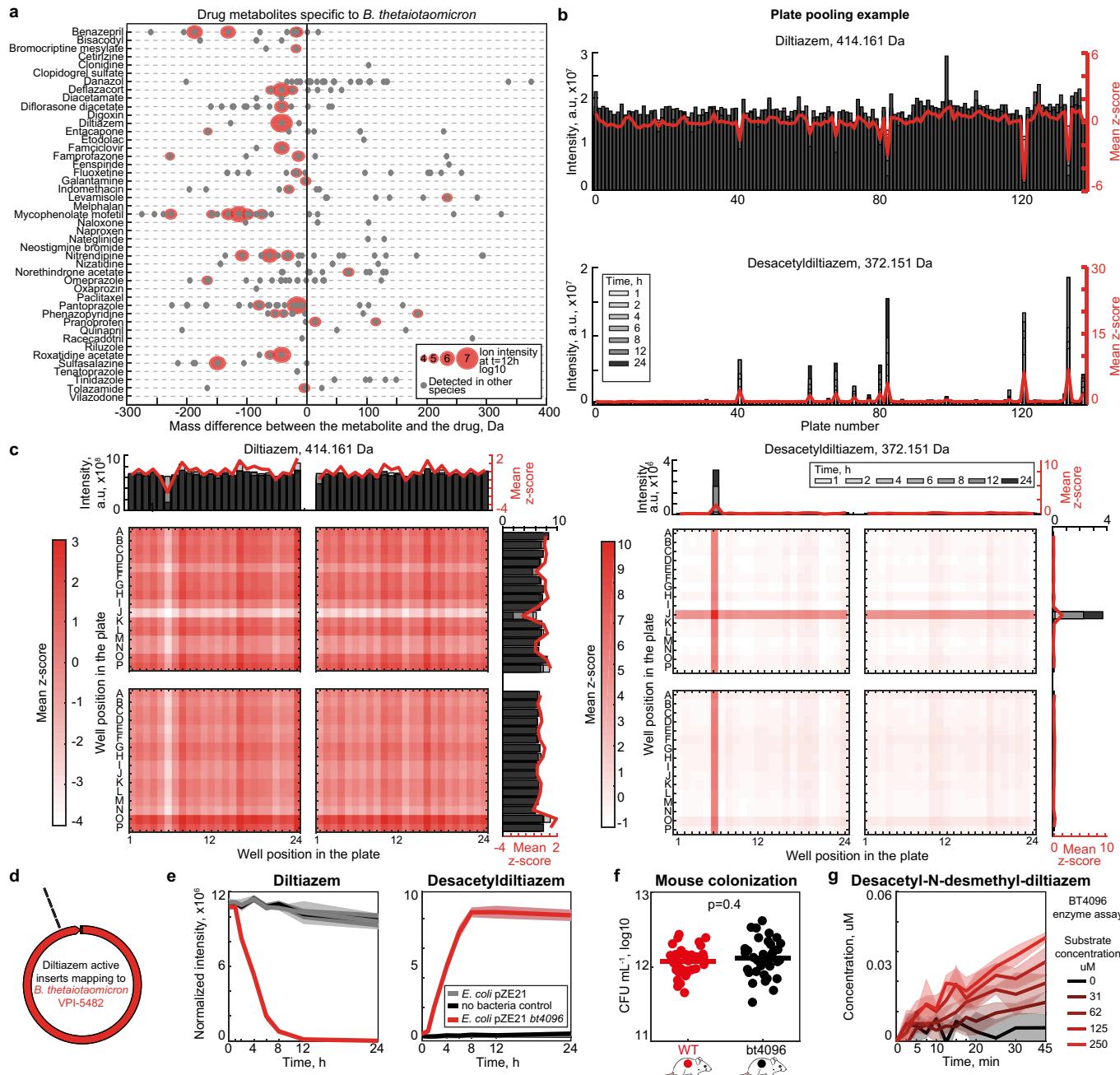
$n = 4$ assay replicates. **c**, Validation of *C. scindens* desmolase activity by mass comparison of metabolites produced from either dexamethasone or *d*₅-dexamethasone and their corresponding LC–MS/MS spectra. Shaded areas correspond to mean \pm s.d., $n = 6$ independent cultures. Representative ion fragments are highlighted in red, to illustrate the loss of the dexamethasone side chain (labelled with two deuterium atoms, compared to the steroid backbone labelled with three deuterium atoms).



Extended Data Fig. 5 | See next page for caption.

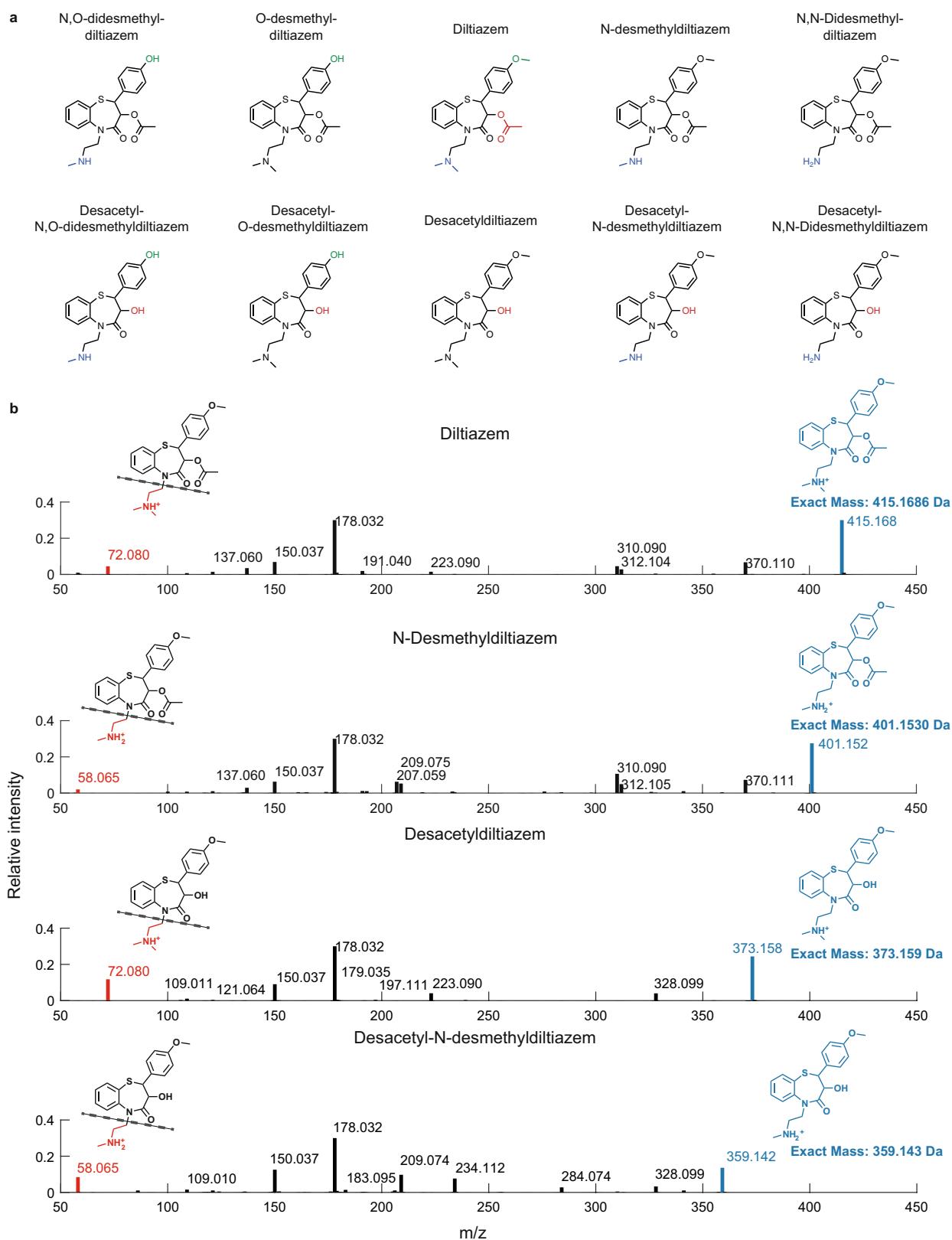
Extended Data Fig. 5 | Microbial corticosteroid metabolism in vivo and in human gut communities. **a**, Dexamethasone serum profile in conventional mice after a single oral dose of dexamethasone. Line depicts the fit of first-order drug elimination kinetics. $n = 4$ mice in total (of either sex) were used for each time point. Data are provided in Supplementary Table 7. **b**, Dexamethasone and dexamethasone metabolite levels across tissues of germ-free and *C. scindens* mono-colonized gnotobiotic mice (GN^{C. scindens}) after 7 h of drug exposure. Horizontal lines show mean values of $n = 6$ mice. * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$ (unpaired two-sided Student's *t*-test). Data and *P* values are provided in Supplementary Table 8. **c**, *C. scindens* desmolase activity for different corticosteroids. Shaded areas correspond to mean \pm s.d., $n = 6$ independent cultures. **d**, Bacterial density of human gut communities. CFU denotes colony-

forming units measured by anaerobic culturing. Horizontal bars represent mean of $n = 4$ independent cultures. **e**, Ex vivo dexamethasone metabolism of gut communities isolated from 28 humans (each colour represents a human donor, lines depict the mean of $n = 4$ replicate assays). *C. scindens* species abundance (quantified by species-specific quantitative PCR) is not sufficient to explain the dexamethasone-metabolizing activity of these human gut communities. Data are provided in Supplementary Table 9. **f**, Correlation between community CFU ml⁻¹ values and dexamethasone (left) or androgen dexamethasone metabolite (right) consumption and production slopes after 12 h of incubation with each of the 28 human gut communities. *P* values were calculated for the null hypothesis that there is zero correlation, against the two-sided alternative that there is non-zero correlation (see Methods).



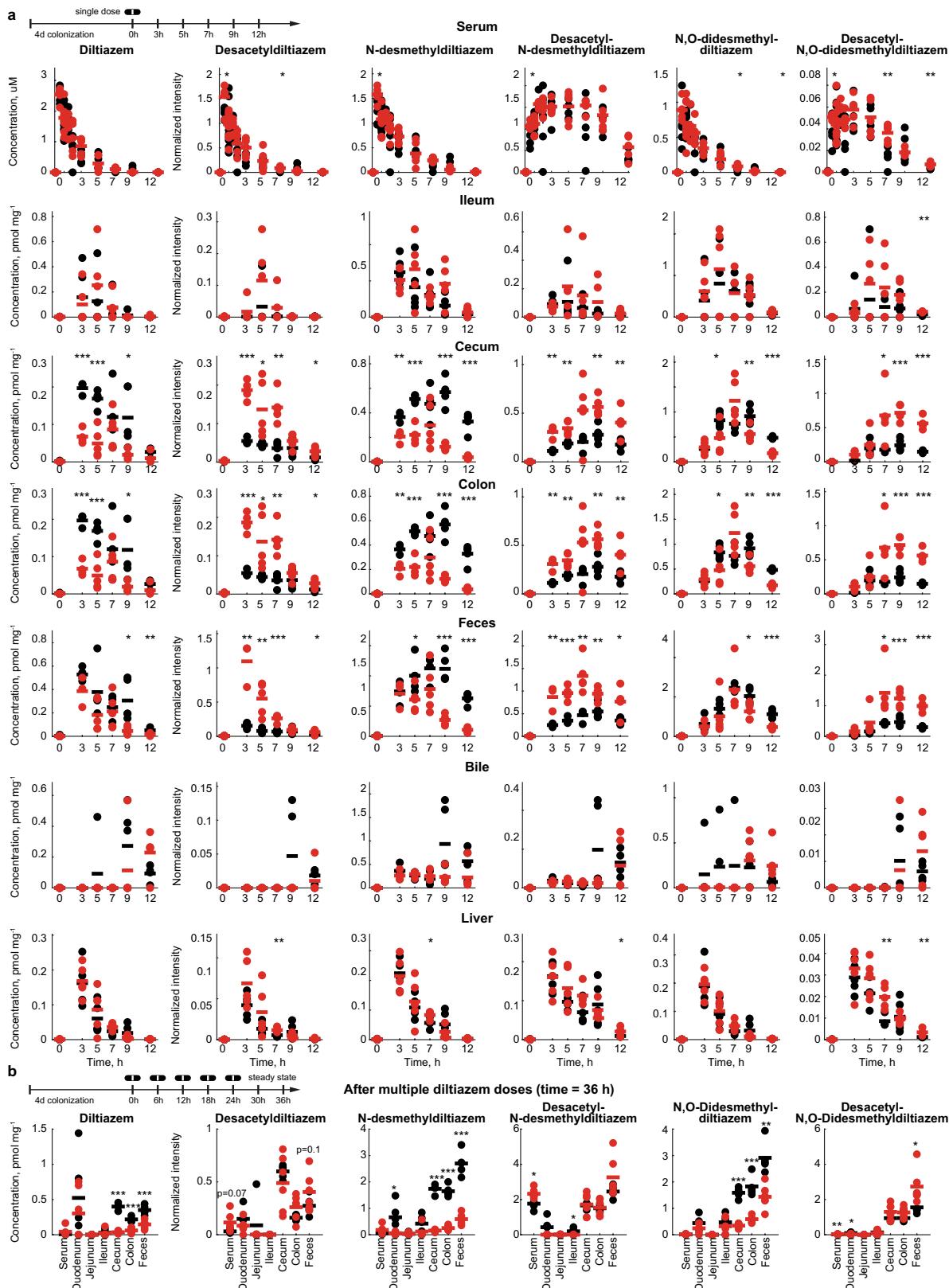
Extended Data Fig. 6 | Gain-of-function approach to identifying microbial drug-metabolizing gene products, using diltiazem metabolism by *B. thetaiotaomicron* as an example. **a**, Drugs metabolized by *B. thetaiotaomicron* and candidate drug metabolites identified by untargeted metabolomics. **b**, Identification of active 384-well library plates that include clones with diltiazem deacetylation activity. **c**, Mapping of diltiazem-converting activity within active plates to identify active clones. **d**, Four independent *E. coli* clones that demonstrate gain of diltiazem-metabolizing activity carry inserts that all map to the same region in the *B. thetaiotaomicron* genome. **e**, Validation of BT_4096 activity by targeted

expression of the open reading frame in *E. coli*. Shaded areas depict the mean and s.d. of independent cultures or assays ($n = 4$). **f**, Bacterial load of gnotobiotic mice mono-colonized with either *B. thetaiotaomicron* wild type or the *bt4096*-mutant strain. Horizontal bars represent mean of $n = 35$ independent mice per group. P value was calculated with unpaired two-sided Student's *t*-test. **g**, In vitro enzyme assay with *N*-desmethyl-diltiazem as substrate to demonstrate that BT_4096 also deacetylates *N*-desmethyl-diltiazem, which is the major metabolic product of diltiazem metabolism in the mouse. Lines and shaded areas depict the mean and s.d. of $n = 4$ assay replicates, respectively.



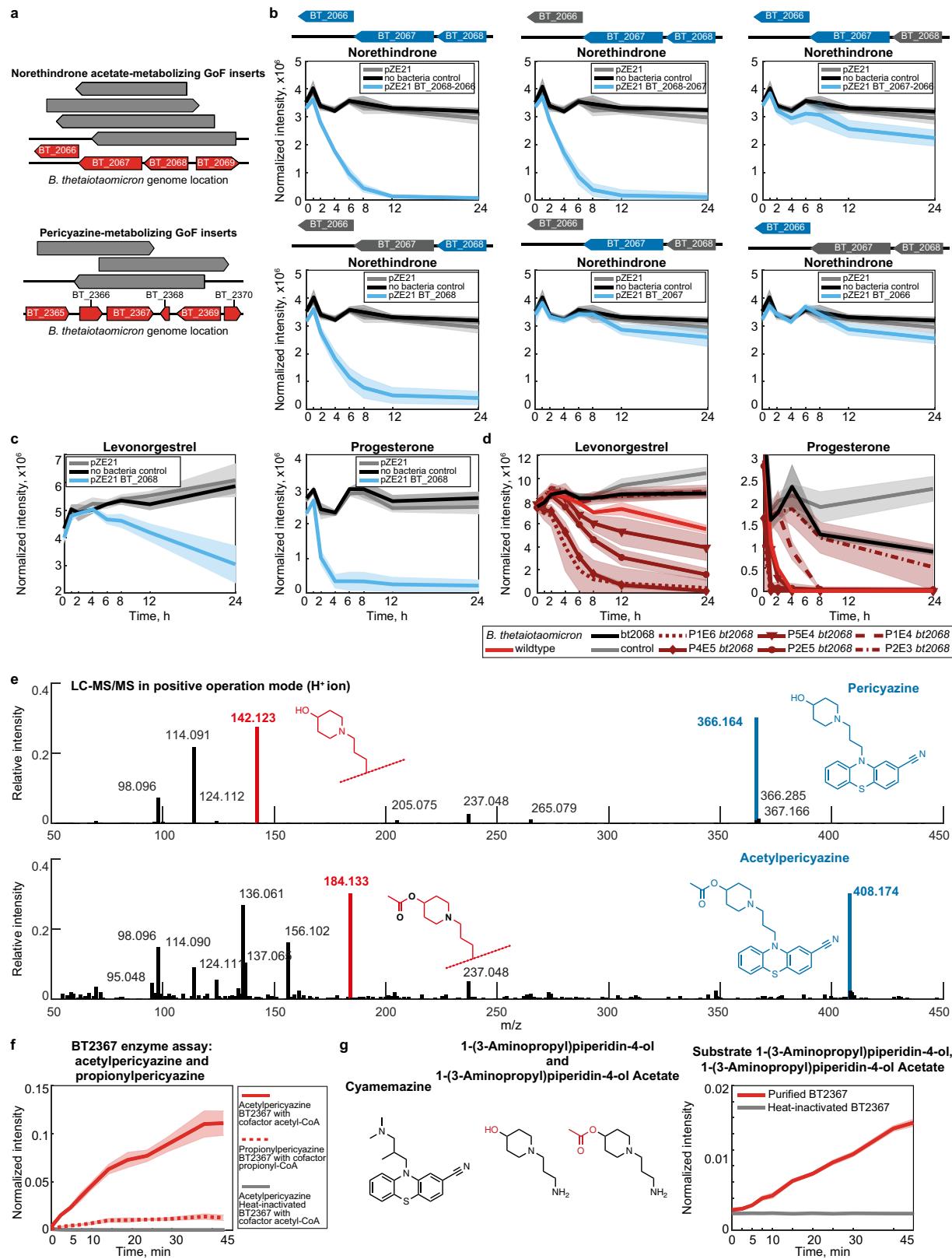
Extended Data Fig. 7 | In vivo diltiazem metabolism and MS/MS to validate metabolite identities. **a**, Structures of diltiazem in vivo metabolites⁴¹. **b**, Exemplary MS/MS analysis to validate the identities of

diltiazem metabolites. LC-MS/MS data for all diltiazem metabolites are compiled in Supplementary Table 21. The experiment was performed $n = 3$ times, with comparable results.



Extended Data Fig. 8 | bt4096-dependent in vivo diltiazem metabolism. **a**, Diltiazem and diltiazem metabolite kinetics in tissues following a single oral dose of diltiazem in gnotobiotic mice mono-colonized with *B. thetaiotaomicron* wild-type or the bt4096-mutant strain. **b**, Intestinal diltiazem and diltiazem metabolite levels following multiple oral doses of diltiazem in mice mono-colonized with *B. thetaiotaomicron* wild-type or the bt4096-mutant strain. Five oral doses were administrated to mice in

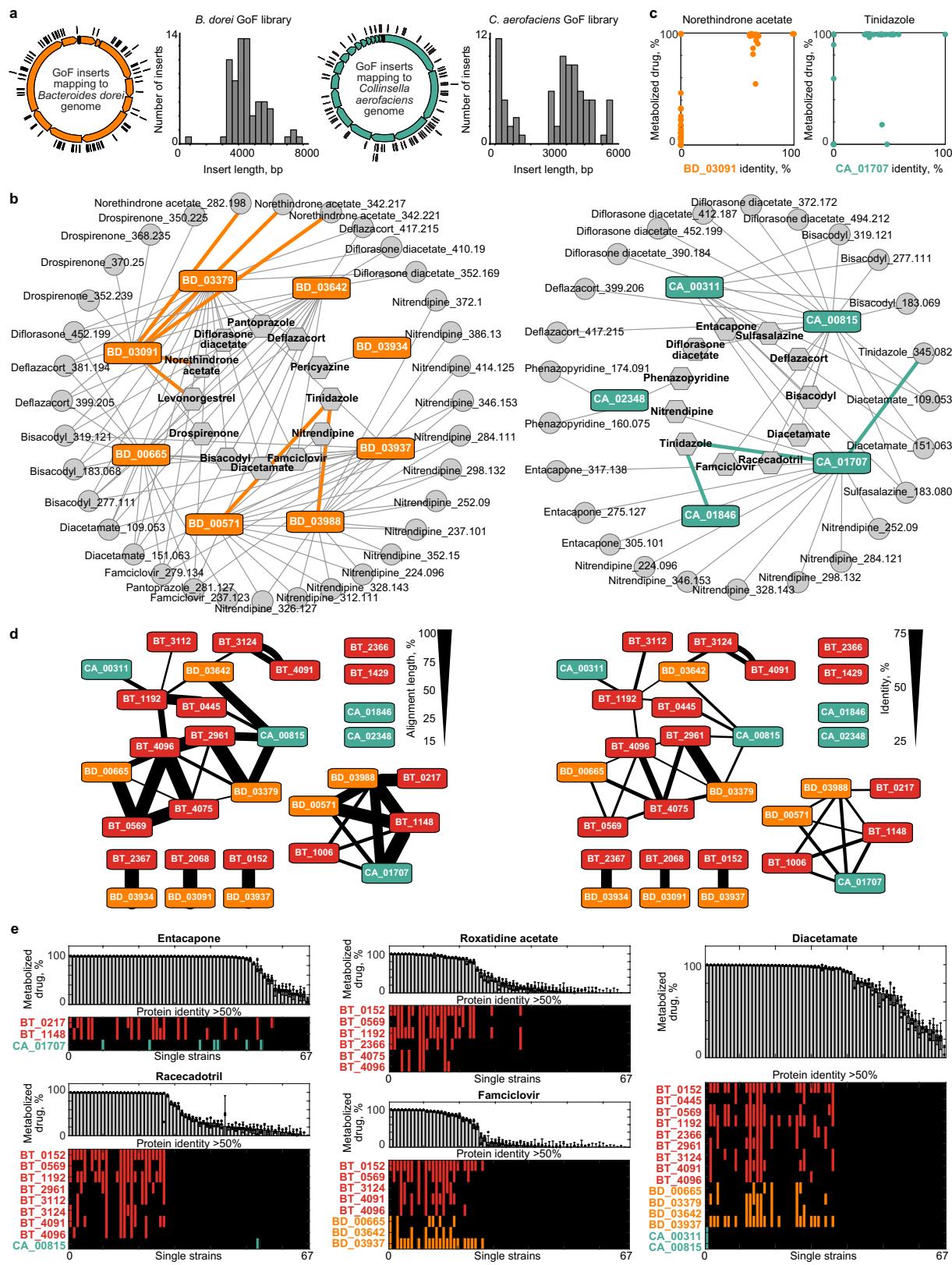
six-hour intervals. Tissues were collected 12 h after the final oral dose of diltiazem. For all mouse experiments, horizontal lines show the mean of $n = 5$ mice and times reflect hours after oral administration of diltiazem. * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$ (unpaired two-sided Student's *t*-test with FDR correction for multiple hypotheses testing). Data and *P* values are provided in Supplementary Tables 10, 11.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Validation of identified drug-metabolizing gene products. **a, B. thetaiotaomicron** genomic DNA fragments identified in norethindrone-acetate- and pericyazine-metabolizing *E. coli* clones. **b, All drug-metabolizing gain-of-function hits were validated by assays with *E. coli* expression constructs that carried PCR-amplified gene sequences, and their combinations in the case of operons (for example, BT_2068–BT_2066, metabolizing norethindrone).** **c, d,** Levonorgestrel- and progesterone-metabolizing activity of BT_2068, shown by *E. coli* that expresses *bt2068* (**c**) and *B. thetaiotaomicron* wild-type, *bt2068*-mutant and *bt2068*-complemented strains (**d**). Gene complementation at different expression levels: P1E6 > P4E5 > P2E5 > P5E4 > P1E4 > P2E3.

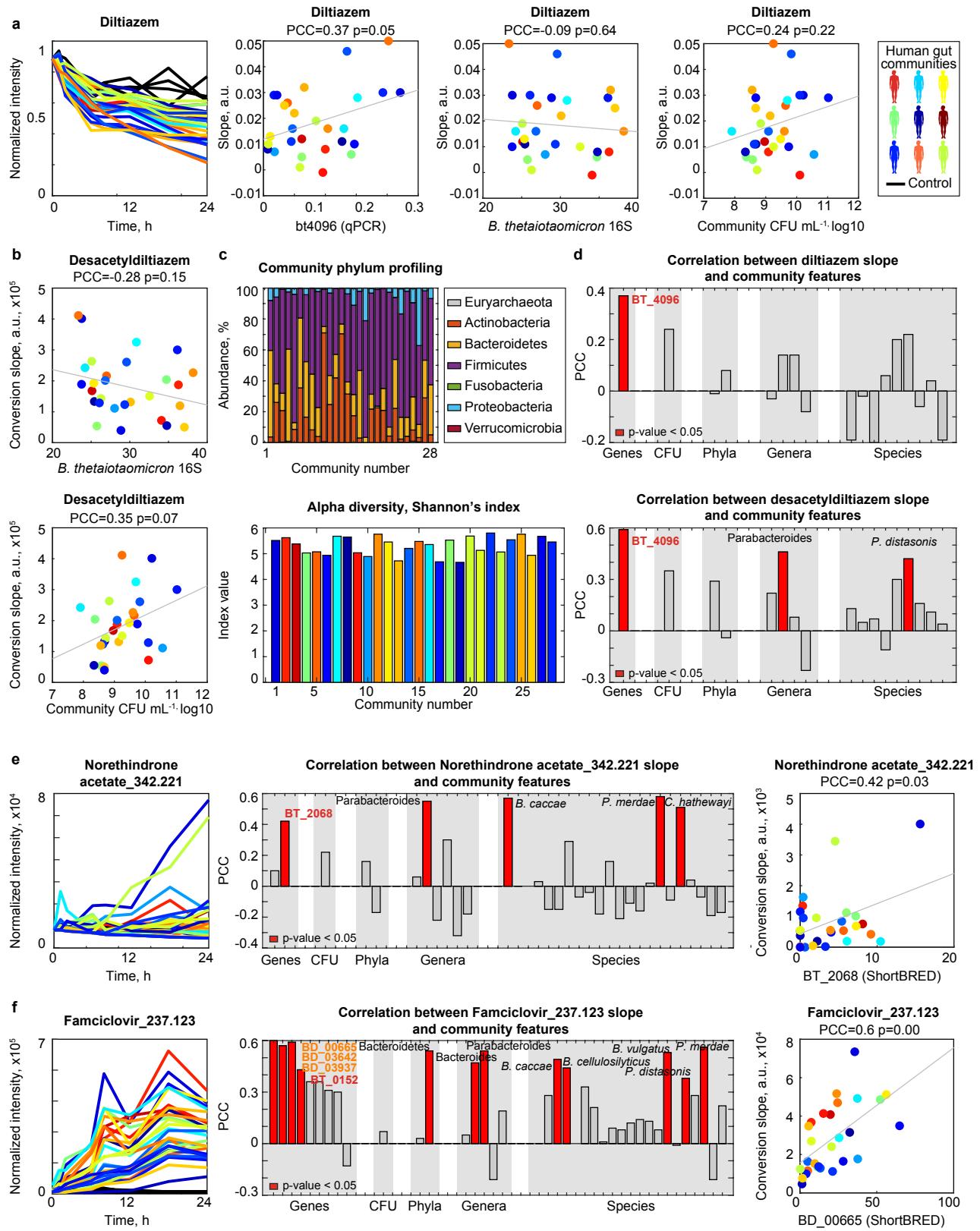
e, Example of LC–MS/MS validation of O-acetyl-periciazine. The experiment was performed $n = 3$ times, with comparable results. **f, Enzymatic validation of O-acetyl- and O-propionyl-transferase activity using purified BT_2367 and pericyazine as substrate.** **g, Enzymatic validation of O-acyl-transferase activity of purified BT_2367 using substrates structurally similar to pericyazine.** Although no acetyl-transferase activity could be measured for cyamemazine, aminopropylpiperidinol is converted to O-acetyl-aminopropylpiperidinol by BT_2367. In **a–d, f, g**, shaded areas depict the mean and s.d. of independent cultures or assays ($n = 4$).



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Identified drug-metabolizing gene products explain the observed drug metabolism of microbial strains and species. **a**, Genome coverage and fragment size distribution in *E. coli* gain-of-function libraries specific for *B. dorei* (based on 78 sequenced clones) and *C. aerofaciens* (based on 81 sequenced clones). Both libraries contained about 37,000 clones. **b**, Network of enzyme–substrate–product drug metabolic interactions for *B. dorei* and *C. aerofaciens*. Each node represents an enzyme (rectangles), a drug substrate (hexagons) or a metabolite product (circles), and each edge represents a validated metabolic interaction (targeted cloning of the gene into *E. coli* results in metabolism of a given drug or production of a specific drug metabolite). **c**, Comparison between maximal BD_03091 and CA_01707 (BD refers

to the BACDOR gene locus tag abbreviation for *B. dorei*; CA refers to the COLAER gene locus tag abbreviation for *C. aerofaciens*) identity of a given bacterial strain and its metabolism of norethindrone acetate and tinidazole, respectively. **d, e**, Reciprocal BLAST analysis of identified drug-metabolizing proteins. Line width depicts the percentage of length (**d**) and identity (**e**) of mutual protein sequence alignment. **e**, Specific drug metabolism rates of 67 genome-sequenced gut bacteria, and the presence of homologues to respective drug-metabolizing gene products. Notably, roxatidine acetate, famciclovir, diacetamate and diltiazem (Fig. 5b) all undergo the same chemical transformation (deacetylation), but distinct sets of gene products explain their microbial metabolism. Bars and error bars represent mean and s.d. of $n = 4$ assay replicates.



Extended Data Fig. 11 | See next page for caption.

Extended Data Fig. 11 | Identified drug-metabolizing gene products explain the observed drug metabolism of bacterial gut communities. **a, b**, Diltiazem conversion to desacetyldiltiazem by 28 human gut communities (each colour represents a human donor, lines depict the mean of $n = 4$ assay replicates). Diltiazem-metabolizing activity of the microbiota does not correlate with total bacterial culture densities or with microbiota abundance of *B. thetaiaomicron* (quantified by species-specific 16S RNA quantitative PCR). P values were calculated for the null hypothesis that there is zero correlation, against the two-sided alternative that there is non-zero correlation (see Methods). **c**, Composition and diversity of the 28 bacterial communities based on metagenomic

sequencing. **d**, Correlation analysis between diltiazem-metabolizing activity of the microbiota and community CFU or metagenomic abundance of BT_4096 homologues, and diltiazem-metabolizing bacterial species, genera and phyla identified in this study. **e, f**, Correlation analysis identical to that shown in **d**, but for the metabolism of norethindrone acetate and famciclovir by the 28 bacterial communities (each colour represents a human donor, lines depict the mean of $n = 4$ replicate assays). P values were calculated for the null hypothesis that there is zero correlation, against the two-sided alternative that there is non-zero correlation (see Methods). Data are available in Supplementary Tables 16–19.