HHS Public Access

Author manuscript

Nat Microbiol. Author manuscript; available in PMC 2023 April 01.

Published in final edited form as:

Nat Microbiol. 2022 October ; 7(10): 1605–1620. doi:10.1038/s41564-022-01226-5.

Host and gut bacteria share metabolic pathways for anti-cancer drug metabolism

Peter Spanogiannopoulos[1,+], **Than S. Kyaw**[1,+], **Ben G. H. Guthrie**[1], **Patrick H. Bradley**[2,3], **Joyce V. Lee**[4], **Jonathan Melamed**[5], **Ysabella Noelle Amora Malig**[5], **Kathy N. Lam**[1], **Daryll Gempis**[1], **Moriah Sandy**[6], **Wes Kidder**[6,7], **Erin L. Van Blarigan**[7,8,9], **Chloe E. Atreya**[6,7], **Alan Venook**[6,7], **Roy R. Gerona**[5], **Andrei Goga**[4,7], **Katherine S. Pollard**[2,8,10,11,12], **Peter J. Turnbaugh**[1,12,*]

[1]Department of Microbiology & Immunology, University of California San Francisco, San Francisco, CA 94143, USA

[2]Gladstone Institutes, San Francisco, CA 94158, USA

[3]Department of Microbiology, The Ohio State University, Columbus, OH 43210, USA

[4]Department of Cell & Tissue Biology, University of California San Francisco, San Francisco, CA 94143, USA

[5]Clinical Toxicology and Environmental Biomonitoring Laboratory, University of California, San Francisco, CA, 94115, USA

[6]Department of Medicine, University of California, San Francisco, San Francisco, CA 94143, USA

[7]UCSF Helen Diller Family Comprehensive Cancer Center, San Francisco, CA 94158, USA

[8]Department of Epidemiology & Biostatistics, University of California, San Francisco, San Francisco, CA, 94158, USA

[9]Department of Urology, University of California, San Francisco, San Francisco, CA 94158, USA

[10]Institute for Human Genetics, University of California, San Francisco, CA 94143, USA

[*]Correspondence to: Peter J. Turnbaugh, Professor, Department of Microbiology & Immunology, 513 Parnassus Avenue HSW 1529, San Francisco, CA 94143-0552, Peter.Turnbaugh@ucsf.edu, Office: (415) 502-3237, Fax: (415) 476-6185.

[+]Equal contributions

Author Contributions Statement

P.J.T conceived of the project and was the primary supervisor for the study. R.R.G., A.G., and K.S.P. also supervised components of this work. P.S. led the *in vitro* screens and *E. coli* strain construction and established protocols for the pharmacokinetics and xenograft experiments. T.S.K. led the final pharmacokinetics, xenograft, transcriptomics, and amplicon sequencing data generation and analysis. B.G.H.G. led the biochemical characterization of PreTA. P.H.B. led the bioinformatic analysis of *preTA* operons across genomes and microbiomes. J.M., Y.N.A.M., T.S.K., B.G.H.G., and M.S. performed mass spectrometry. K.N.L. sequenced and analyzed the drug resistant *E. coli* mutants. J.V.L. assisted with the tumor xenograft measurements. C.E.A., A.V., and W.K. (GO Study PI) oversaw the conception and design of the GO Study and contributed patient samples. E.L.V.B. contributed to developing the study protocol and supervision of data collection for the GO Study. D.G. designed the GO Study Specimen Collection Kits and managed biospecimen collection, storage, and retrieval. P.S. wrote the initial draft. T.S.K. and P.J.T. revised the manuscript with input from all authors.

Competing Interests Statement

P.J.T. is on the scientific advisory boards for Pendulum, Seed, and SNIPRbiome; there is no direct overlap between the current study and these consulting duties. K.S.P. is on the scientific advisory board for Phylagen; there is no direct overlap between the current study and these consulting duties. C.E.A serves on the scientific advisory board for Pionyr Immunotherapeutics and has received research funding (institution) from Bristol Meyer Squibb, Guardant Health, Kura Oncology, Merck, and Novartis; there is no direct overlap with the current study. All other authors have no relevant declarations.

[11]Bakar Computational Health Sciences Institute, University of California, San Francisco, CA 94143, USA

[12]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA

Abstract

Pharmaceuticals have extensive reciprocal interactions with the microbiome, but whether bacterial drug sensitivity and metabolism are driven by pathways conserved in host cells remains unclear. Here, we show that anti-cancer fluoropyrimidine drugs inhibit the growth of gut bacterial strains from 6 phyla. In both *Escherichia coli* and mammalian cells, fluoropyrimidines disrupt pyrimidine metabolism. Proteobacteria and Firmicutes metabolized 5-fluorouracil (5-FU) to its inactive metabolite dihydrofluorouracil (DHFU), mimicking the major host mechanism for drug clearance. The *preTA* operon was necessary and sufficient for 5-FU inactivation by *E. coli*, exhibited high catalytic efficiency for the reductive reaction, decreased the bioavailability and efficacy of oral fluoropyrimidine treatment in mice, and was prevalent in the gut microbiomes of colorectal cancer patients. The conservation of both the targets and enzymes for metabolism of therapeutics across domains highlights the need to distinguish the relative contributions of human and microbial cells to drug efficacy and side effect profiles.

Editor summary:

Anti-cancer fluoropyrimidine drugs have anti-bacterial effects on the gut microbiome and these drugs can be metabolized by gut bacteria via conserved pathways also found in mammalian hosts.

Introduction

Decades of pharmacogenetics and pharmacogenomics research have revealed the fundamental mechanisms through which drugs are metabolized by hepatocytes, enterocytes, and other cell types throughout the body and the importance of drug transporters in mediating the absorption, distribution, and elimination of drugs and other foreign compounds (xenobiotics)[1]. These studies have also helped illuminate drug mechanisms of action and helped to identify novel drug targets. Robust links between human genotype and drug outcomes have enabled diagnostic tests that inform drug selection and dosing. Yet, these studies have largely ignored the trillions of microorganisms found in and on the human body (microbiota) or their aggregate genomes and metabolic activities (microbiome), representing a major gap in our scientific knowledge and an untapped resource to improve efforts towards precision medicine.

Numerous studies have indicated that the gut microbiome is an underexplored contributor to inter-individual variations in the efficacy and toxicity of cancer therapy. Gut bacteria can metabolize hundreds of host-targeted drugs, in some cases leading to clinically relevant changes in first-pass metabolism or the clearance of drugs in the gut following biliary excretion[2–5]. A well-studied example of this is highly relevant to colorectal cancer (CRC): gut bacteria re-activate the downstream hepatic metabolite of irinotecan (SN-38G), contributing to its dose-limiting gastrointestinal (GI) toxicity[6,7]. Gut bacteria can influence other anti-cancer drugs (chemotherapeutics) and even biologics (immunotherapies) through

interactions with the host immune system[8–11]. These seminal cancer studies, together with the research we and others have performed in other disease areas[12–16] have established an experimental and conceptual framework to dissect the mechanisms by which human-associated microbial communities influence the treatment of disease. However, the primary focus has been on reactions unique to bacteria, far less is known about the potential contributions of pathways conserved in bacterial and mammalian cells to drug metabolism and disposition[17,18].

Despite major innovations in immunotherapy, chemotherapeutics like fluoropyrimidines remain the cornerstone of cancer therapy[19]. Fluoropyrimidines have extensive interactions with the human gut microbiome that may have downstream consequences for treatment outcomes. The oral fluoropyrimidine capecitabine (CAP) is extensively metabolized to 5-FU, 5-fluorodeoxyuridine (FUDR), and other metabolites[5] (Extended Data Fig. 1a). CAP meets the FDA criteria for a “highly variable” drug, with a coefficient of variation >30% in intra-subject pharmacokinetic parameters combined with extensive variation between subjects[20] that cannot be explained by the known host risk factors[21,22]. Adverse reactions to CAP require dose adjustments in ~35% of patients and complete discontinuation of therapy in ~10% of patients; GI side effects are common[23,24]. Fluoropyrimidines were designed to target conserved pathways essential for gut bacterial growth[25,26] and evidence in rats has shown that these drugs alter the gut microbiota[27–29]. Cross-sectional analyses of the human gut microbiome have linked the use of anti-cancer drugs (including fluoropyrimidines) to changes in gut microbial community structure and function[30–33]. The bacterial metabolism of fluoropyrimidines may also have downstream consequences for host physiology; studies in *Caenorhabditis elegans* have linked genetic differences in their bacterial food source, *E. coli,* to host drug toxicity[34–36]. Furthermore, the increasing administration of CAP and other oral fluoropyrimidines[37] may enhance the potential for interactions with the gut microbiota prior to first pass metabolism and absorption into general circulation[5].

Results

Fluoropyrimidines inhibit human gut bacterial growth.

We screened 47 human gut bacterial strains for sensitivity to the fluoropyrimidines CAP, 5-FU, and FUDR (Extended Data Fig. 1a). There was extensive variation in the minimal inhibitory concentration (MIC) of 5-FU and FUDR (Fig. 1a and Supplementary Table 1). A total of 13 (5-FU) and 15 (FUDR) strains tolerated the highest concentration assayed; 12/13 (5-FU) and 12/15 (FUDR) of these strains displayed partial growth inhibition (range: 22.8–96.0% of the growth control; Supplementary Table 1). The MICs of 5-FU and FUDR were significantly associated (r=0.475, p=0.0008), as expected based on their shared substructure and downstream metabolites (Extended Data Fig. 1a). Bacterial phylogeny did not predict 5-FU sensitivity, with no significant differences between phyla (Extended Data Fig. 1b) and multiple nearest neighbors with opposite phenotypes; *e.g., Parabacteroides distasonis* (5-FU MIC=300 ng/ml) and *P. merdae* (5-FU MIC>1 mg/ml) (Fig. 1a). In contrast, the Bacteroidetes were significantly more sensitive to FUDR relative to either Actinobacteria or Firmicutes (Extended Data Fig. 1c). Consistent with a recent report[36], short-term exposure of *E. coli* and two representative *Bacteroides* (*B. fragilis* and *B. ovatus*) to 5-FU was

sufficient to select for drug resistant mutants (MIC>1 mg/ml; Extended Data Fig. 1d and Supplementary Table 2).

The bioactivation of the prodrug CAP is thought to be uniquely catalyzed by mammalian enzymes expressed in the intestine and liver[5,38]. Surprisingly, we found that multiple gut bacteria were susceptible to CAP at concentrations below the estimated maximum in the proximal gut (~14 mg/ml following oral administration)[39]. With the exception of *Providencia rettgeri* we were able to determine a MIC for all of the tested strains (Supplementary Table 1). CAP MIC was significantly associated with FUDR MIC (r=0.478, p=0.0007), but not 5-FU MIC (r=0.014, p=0.9233), further emphasizing the compound-specific antibacterial effects of these fluoropyrimidines. Consistent with these data, *E. coli* was able to convert CAP to 5-FU during *in vitro* growth (Extended Data Fig. 1e). More work is needed to elucidate the mechanisms responsible for gut bacterial CAP activation and the metabolic fate of this prodrug during *in vitro* and *in vivo* growth, building upon the recent identification of a gut bacterial enzyme capable of CAP deglycosylation[5].

Next, we sought to gain insight into the mechanism of action of fluoropyrimidines against human gut bacteria, building on studies in *C. elegans*[34–36] and testing their generalizability to CAP and FUDR. In mammalian cells, the primary target of fluoropyrimidines is thymidylate synthase[40], an essential enzyme for DNA, RNA, and protein biosynthesis[41]. This mechanism of action is nutrient-dependent; excess uracil decreases drug sensitivity[42] (Extended Data Fig. 2a). As expected, the *E. coli* MIC for all three fluoropyrimidines was higher in rich media and increased by uracil in a dose-dependent manner (Extended Data Figs. 2b–d and Supplementary Table 3).

Transcriptional profiling (RNA-seq) enabled a more comprehensive view of the metabolic pathways impacted by fluoropyrimidines. *E. coli* was grown to mid-exponential phase and then exposed to sub-MIC levels of CAP and 5-FU under both aerobic (n=2 independent experiments) and anaerobic conditions (Supplementary Table 4). Despite marked differences in transcript levels between growth conditions and experiments (Fig. 2a), we were able to identify significant effects of both drugs relative to vehicle controls (Fig. 2b,c). Pathway enrichment analysis revealed that pyrimidine metabolism was consistently impacted by 5-FU during aerobic growth; whereas flagellar assembly was consistently impacted by CAP irrespective of growth condition and by 5-FU under anaerobic conditions (Fig. 2d). Across both experiments, we identified 1,720 differentially expressed genes (DEGs; FDR<0.1, |log2 fold-change|>1, DESeq; Fig. 2e and Supplementary Table 5). In our repeat experiment, 620/892 (69.5%) DEGs were unique to a single condition. We also identified 112 (CAP) and 12 (5-FU) genes consistently differentially expressed in both growth conditions. CAP exposure downregulated the flagellar biosynthetic pathway, including the master flagellar regulator *flhDC* (Fig. 2f). 5-FU exposure upregulated *carA*, a key gene involved in the first committed step of the *de novo* pyrimidine biosynthetic pathway.

Consistent with these results and prior work[36,43], a targeted screen of 22 non-essential DEGs involved in pyrimidine metabolism revealed that the deletion of *upp* (uracil phosphoribosyltransferase) leads to a high level of resistance to CAP, 5-FU, and FUDR (Extended Data Figs. 2e–g and Supplementary Table 3). Targeted and whole

genome sequencing of 5-FU-resistant mutants revealed multiple unique single nucleotide polymorphisms, deletions, and frameshifts within the *upp* genes of *E. coli* and *B. fragilis* (Supplementary Table 2). In strains with wild-type *upp,* we identified mutations within other pyrimidine metabolism genes, including uridine phosphorylase (*E. coli*), uridylate kinase (*E. coli*), and thymidine kinase (*B. ovatus*) (Supplementary Table 2). Together, our results support pyrimidine metabolism as a key target for fluoropyrimidines, while demonstrating broader impacts of these compounds on genes involved in flagellar assembly, consistent with another recent report[36].

The bacterial *preTA* operon inactivates 5-FU.

Drug resistant bacteria can often catalyze drug inactivation[44,45], potentially interfering with the efficacy of cancer therapy due to decreased bioavailability and/or enhanced clearance[17]. To test if the identified fluoropyrimidine-resistant gut bacteria were capable of drug inactivation, we screened the top 23 strains (5-FU MIC≥62.5 μg/ml) using a 5-FU disk diffusion assay. We identified two active Proteobacterial strains: *E. coli* MG1655 and *Salmonella enterica* LT2 (Fig. 1b). LC-QTOF/MS confirmed the depletion of 5-FU with near quantitative conversion to DHFU (Fig. 1c). In mammalian cells, dihydropyrimidine dehydrogenase (DPYD) is responsible for the biotransformation of 5-FU to the inactive metabolite DHFU[46–48]. Patients with *DPYD* sequence variants are at risk for elevated systemic exposure to 5-FU and adverse events[46–48]. In *E. coli*, DPYD is encoded by two neighboring genes: *preT* and *preA* found within the *preTA* operon[49]. Purified PreTA protein is sufficient to catalyze the reduction of the pyrimidines uracil and thymine, as well as 5-FU[49]; however, the genes necessary and sufficient for this activity in bacterial cells remained unclear.

We generated a clean deletion of the *preTA* operon in *E. coli* MG1655 (Δ*preTA*), which led to a complete loss of function in our bioassay that was validated by LC-QTOF/MS (Fig. 1d). Chromosomal complementation with a constitutively expressed *preTA* operon (*preTA*$^{++}$) resulted in an increased rate of metabolism relative to wild-type *E. coli* (Fig. 1d). An integrated empty vector control (*E. coli* Δ*preTA*/pINT1) had no impact on either assay. We confirmed the role of the *preTA* operon in 5-FU inactivation with a second strain of *E. coli* (BW25113; Extended Data Fig. 3a) and demonstrated that both genes of the operon were required for conversion to DHFU (Extended Data Fig. 3b). Comparable growth was observed for each isogenic strain in the presence or absence of 5-FU, with the exception of *preTA*$^{++}$, which showed an increase in drug tolerance in minimal media (Extended Data Fig. 3c). *E. coli* MG1655 growth was unaffected by DHFU (MIC>1 mg/ml). To test the potential relevance of this reaction to drug efficacy, we added cell-free supernatants (CFS) from each strain to the CRC cell line HCT-116. Sterile media and Δ*preTA* CFS supplemented with 5-FU inhibited cell proliferation; however, this effect was fully rescued by wild-type and *preTA*$^{++}$ *E. coli* (Fig. 1e). These results indicate that *preTA* is necessary and sufficient for 5-FU inactivation by *E. coli* (Fig. 1f) with potential impacts for host drug exposure but minimal impacts on bacterial drug sensitivity.

We were surprised that our cell-based assays demonstrated near quantitative conversion of 5-FU to DHFU given that *E. coli* PreTA has been characterized as a reversible enzyme[49],

prompting further *in vitro* studies with *E. coli* PreTA (Fig. 3a). Purifying heterologously expressed *E. coli* PreTA to homogeneity showed that PreTA purifies as a heterotetramer and contains both flavin and iron-sulfur cofactors as seen by characteristic shifts in the oxidized and reduced UV-visible absorption spectra (Fig. 3b, Extended Data Figs. 4a–c). We confirmed the activity of the heterologous PreTA by anaerobically incubating the protein with substrates and reducing equivalents, observing the concomitant decreases of substrate peaks with increases of product peaks by high pressure liquid chromatography (HPLC) (Fig. 3c, Extended Data Fig. 4d). The correct product masses of these reactions were identified by Liquid Chromatography High Resolution Mass Spectrometry (LC-HRMS), indicating that the protein was catalyzing the expected reaction (Extended Data Fig. 4e). Finally, we measured steady-state kinetics in both the reductive and oxidative reaction directions and saw that while the catalytic efficiency for uracil and 5-FU was comparable in the reductive direction, the catalytic efficiency in the oxidative direction decreased by two orders of magnitude for DHU compared to the reductive direction (Figs. 3d,e and Supplementary Table 6). While literature has stated DPYD is a reversible enzyme[50–52], quantitative studies indicate that the reaction proceeds to near completion in the reductive direction[53,54]. Thus, under physiological conditions DHFU is an extremely poor substrate for PreTA and the metabolism of 5-FU by bacteria at physiologically relevant concentrations is essentially irreversible (Fig. 3f), consistent with our cell-based assay data (Figs. 1c,d).

Bacterial PreTA interferes with cancer treatment in mice.

Measurements of tumor growth following CAP administration provided support for the potential physiological relevance of *E. coli* PreTA. We opted to use the *preTA*$^{++}$ and Δ*preTA E. coli* strains to avoid the possible confounding effects of transcriptional regulation; the wild-type *preTA* operon was expressed at a significantly higher level under aerobic conditions (Fig. 2g). We used a tumor xenograft model to test the impact of PreTA on drug efficacy in three independent experiments that varied in the enrollment criteria and streptomycin dose (Fig. 4a and Extended Data Figs. 5a,g). Mice injected with HCT-116 human CRC cells were randomly split into two colonization (Δ*preTA* or *preTA*$^{++}$ *E. coli*) and two treatment (CAP or vehicle) groups. As expected, baseline tumor volumes were highly variable in experiments utilizing a set enrollment date (Extended Data Figs. 5b,h), in contrast to rolling enrollment based on tumor size (Fig. 4b); there were no significant differences between groups. *E. coli* colonization level was dose-dependent and similar between groups (Extended Data Figs. 5c,i and 6a). Mice colonized with high levels of Δ*preTA E. coli* that received CAP had a significant decrease in tumor growth relative to vehicle controls or CAP treated animals colonized with the *preTA*$^{++}$ strain (Figs. 4c–e, rolling enrollment) corresponding to significantly prolonged survival (Fig. 4f). Similar trends were observed in the high streptomycin, set enrollment experiment (Extended Data Figs. 5d–f); however, low streptomycin did not result in detectable differences between groups (Extended Data Figs. 5j–l). Thus, high levels of colonization by *preTA* encoding bacteria may be necessary to interfere with drug efficacy.

The most parsimonious explanation for these results is that PreTA contributes to the first-pass metabolism of CAP decreasing the dose available to the tumor. Three independent pharmacokinetics experiments were performed to test the effect of PreTA on 5-FU

bioavailability (Figs. 4g,k). Conventionally-raised (CONV-R) specific pathogen free mice colonized with *preTA*$^{++}$ *E. coli* had decreased plasma concentrations of 5-FU relative to Δ*preTA* controls following a high dose of oral CAP (1100 mg/kg, Fig. 4h). Total circulating drug levels (Fig. 4i) and peak plasma concentration (Fig. 4j) were significantly lower in *preTA*$^{++}$ colonized mice relative to Δ*preTA* controls. Similar trends were observed at a lower CAP dose (500 mg/kg, Extended Data Fig. 5m). To remove the potential confounding effect of the background microbiota, we conducted a similar study using gnotobiotic mice (Fig. 4k). Gnotobiotic mice mono-colonized with Δ*preTA, wt,* or *preTA*$^{++}$ *E. coli* had comparable colonization level (Fig. 4l) and weight (Fig. 4m). Plasma 5-FU concentration was significantly decreased in *preTA*$^{++}$ colonized mice relative to Δ*preTA* controls; the *wt* strain had an intermediate phenotype that did not reach statistical significance (Fig. 4n). DHFU was undetectable in plasma from all experiments.

To identify potential impacts of PreTA on the rest of the gut microbiota, we performed 16S rRNA gene sequencing on longitudinal stool samples collected prior to and during treatment with streptomycin and CAP (Supplementary Table 7). Gut microbial diversity and community structure were significantly altered over time (Extended Data Figs. 6b–f), consistent with prior data in the streptomycin model[55]. We did not detect any significant differences in community structure, diversity, or *E. coli* abundance between colonization groups (Extended Data Figs. 6b–f). Only 3/69 amplicon sequence variants (ASVs) were differentially abundant between groups (Extended Data Figs. 6g). While these data support a direct role for *E. coli* PreTA in first-pass metabolism, more work is needed to assess the level of *preTA* abundance and expression necessary to impact drug metabolism and potential impacts of PreTA on host drug absorption and metabolism.

preTA is encoded by diverse taxa and varies in abundance.

We sought to determine if *E. coli* is the primary source of *preTA* in cancer patients. Given the conserved sequence and function of mammalian DPYD and bacterial PreTA we were surprised that only two moderately drug resistant strains (MIC=62.5 μg/ml) were identified in our screen for 5-FU inactivating bacteria. A tBLASTn search querying *E. coli* PreTA against the draft genomes of the other 22 tested strains only revealed a single pair of genomic loci from *S. enterica* (84% and 94% full length amino acid identity to PreT and PreA, respectively). No significant hits were found in the 24 drug sensitive strains. To identify other putative 5-FU inactivating strains, we built profile Hidden Markov Models (HMMs) of *preT* and *preA* and used them to search for orthologs across 9,082 bacterial isolate genomes. To minimize false positives due to the fact that both PreT and PreA are part of large protein families with a range of functions and substrate specificities[56], we required that the *preT* and *preA* orthologs were adjacent in the genome and on the same strand. This analysis revealed 1,704 putative *preTA* operons from 406 species, mainly in Proteobacteria (1,507 operons from 290 species) and Firmicutes (172 operons from 92 species) (Fig. 5 and Supplementary Table 8). This phylum-level distribution was similar for species with an estimated gut prevalence over 5% (n=23) and for species with a gut prevalence of 1% or lower (n=109)[57]. As expected, *preTA* was conserved in close relatives of *E. coli*, including other *Escherichia*, *Salmonella*, and *Citrobacter*. However, we also found operons in more distantly related Proteobacteria (*e.g., Oxalobacter formigenes*) and diverse

Firmicutes: *Anaerostipes hadrus*, *Eubacterium hallii*, and *Lactobacillus reuteri*. These results were consistent in an independent analysis of 60,664 metagenome-assembled genomes (MAGs)[58], resulting in the identification of 74 additional *preTA* operons, all from human gut MAGs, 89% of which were Firmicutes or Proteobacteria (Extended Data Fig. 7 and Supplementary Table 9).

We validated 5 putative *preTA* orthologs found in bacterial isolate genomes through a combination of heterologous expression and cell-based assays. Heterologous expression in the *E. coli* Δ*preTA* strain enabled us to validate *preTA* orthologs detected in two additional Proteobacteria (*S. enterica* LT2, *O. formigenes* ATCC35274) and the *L. reuteri* DSM20016 (phylum, Firmicutes) using both our disk diffusion and LC-QTOF/MS assays (Extended Data Fig. 8a and Supplementary Tables 10,11). Incubation of *S. enterica* LT2 and two additional *preTA*-positive strains that we did not evaluate using heterologous expression supported the ability of the Firmicutes *Anaerostipes caccae* DSM14662 and *Clostridium sporogenes* DSM795 to metabolize 5-FU (Extended Data Fig. 8b). *A. caccae* depleted 5-FU without producing any detectable DHFU, potentially due to further downstream metabolism. Together, these results demonstrate that bacterial PreTA has maintained its activity against 5-FU despite considerable phylogenetic and primary sequence divergence (66–93% similarity).

Next, we quantified *preTA* abundance in 575 metagenomic samples from 179 CRC patients and healthy controls[59] using a conservative pipeline to improve specificity for the *preTA* operon. While *preTA* was detected in all individuals, its abundance varied over two orders of magnitude (Figs. 6a,b). On average, *preTA* was less abundant in CRC patients relative to controls (Fig. 6b). At the class level, Clostridia were the dominant source of *preTA* (Fig. 6c). At the species level, most of these reads came from an *Anaerostipes* operational taxonomic unit (OTU). These results, together with the significant depletion of *Anaerostipes* in CRC cases[60], likely explain the overall differences between cases and controls in *preTA* abundance (Fig. 6b). *E. coli* and *L. reuteri,* two of the bacterial species we had functionally validated (Extended Data Figs. 8a,b) explained *preTA* abundance in a minority of samples.

Finally, we examined the dynamics of *preTA* abundance over time in CRC patients treated with fluoropyrimidines. Stool samples were collected as part of the Gut Microbiome and Oral Fluoropyrimidine Study in Patients with Colorectal Cancer (GO; ClinicalTrials.gov NCT04054908). At the time of our interim analysis, longitudinal samples collected before, during, and after the first treatment cycle were available from 11 patients treated with CAP-based regimens. In total, 54 samples were subjected to deep metagenomic sequencing (Supplementary Table 12). Consistent with the larger study of CRC patients and healthy controls (Fig. 6b), we detected marked variation in *preTA* abundance between individuals and *preTA* reads in most individuals mapped best to *Anaerostipes* (Fig. 6d and Extended Data Fig. 9). In these initial 11 patients, we did not observe any consistent and statistically significant shifts in *preTA* abundance over time (p=0.42, linear model with fixed time effects and random patient effects, Satterthwaite approximation; Fig. 6d). In 3 individuals, *preTA* abundance varied over an order of magnitude across time points. Incubations of pooled bacterial isolates from the baseline samples of an expanded set of 22 CRC patients revealed 5-FU inactivation in 17/22 (77%) individuals (Extended Data Fig. 8c). LC-QTOF/MS

analysis of incubations of pooled (Fig. 6e) and clonal (Fig. 6f) bacterial isolates from a representative patient confirmed the metabolism of 5-FU to DHFU. More work is needed to assess if these variations in baseline *preTA* abundance and/or temporal fluctuations contribute to differences in treatment outcomes.

Discussion

Our results demonstrate how pathways for sensitivity to and metabolism of fluoropyrimidine drugs are conserved across two domains of life. These findings have broad implications across multiple disciplines. Elucidating drug mechanisms of action in bacteria may provide translational insights for host tissues, raising questions as to the broader view of the off-target effects of therapeutics and the degree to which drug-induced shifts in microbial community structure and function have downstream consequences for drug efficacy and/or toxicity. While our *in vitro* and *in vivo* results both support the hypothesis that bacterial drug inactivation in the gastrointestinal tract or even within diseased tissue[61] could interfere with drug efficacy, more work is needed to assess the relative impact of this biotransformation on efficacy versus gastrointestinal or other side effects, as previously demonstrated for the anti-cancer drug irinotecan[6]. Continued biochemical and structural characterization of the PreTA holoenzyme could enable the development of bacteria-specific enzyme inhibitors, while insights into the ecological role of PreTA may unlock strategies for strain replacement prior to therapy. These results are also consistent with another recent study[18], which highlighted the unexpected overlap between host and bacterial drug metabolites. Traditional approaches for assessing drug disposition do not distinguish these two alternatives[17], which may explain the difficulties in predicting fluoropyrimidine toxicity using only human genotypic information. Furthermore, our discovery of diverse *preTA* positive bacterial strains could open the door towards preventing the severe, and at times lethal, toxicity observed from patients undergoing fluoropyrimidine chemotherapy with loss-of-function mutations in mammalian DPYD[46].

This study has multiple important limitations. Bacterial drug sensitivity was primarily evaluated during *in vitro* culture, as in prior studies[62]. More work is needed to assess the *in vivo* sensitivity of the gut microbiota to fluoropyrimidines and other drugs that exhibit *in vitro* effects on bacterial growth. This could be readily studied in gnotobiotic mice colonized with the same strains tested *in vitro* or through longitudinal analyses of human cohorts. While we focused on the *preTA* operon as one mechanism through which gut bacteria inactivate 5-FU, this is unlikely to be the only mechanism through which the gut microbiome impacts the activity of fluoropyrimidines. Consistent with this, a recent manuscript described the deglycosylation of CAP[5], suggesting that these results are likely the tip of the iceberg in terms of the full set of microbial metabolic pathways involved. Another key limitation is our focus on the model organism *E. coli,* which we showed is rarely the most abundant source of the *preTA* operon in humans. Future experiments focusing on the *Anaerostipes* genus and other members of the Firmicutes phyla that encode *preTA* will help to understand and predict the metabolism of 5-FU in the setting of a complex gut microbiota. These efforts will require the development of novel genetic tools, as demonstrated for gut Clostridia[63]. Our drug efficacy experiments in mice were limited to a single xenograft model in the context of streptomycin depletion of the gut microbiota. Future

work in gnotobiotic mice using genetically engineered models of cancer and/or patient derived xenografts would be a valuable extension of this work. Finally, while our data in humans was essential to establish that *preTA* is prevalent, highly variable in abundance, dynamic over time, and primarily derived from Firmicutes in cancer patients, it did not directly test if *preTA* explains inter-individual variations in drug outcomes. Adequately powered studies in diverse clinical cohorts will be essential to test the predictive power of *preTA* as a biomarker and whether or not interventions that alter the abundance of *preTA*-encoding bacterial species have a clinically relevant effect.

Despite these limitations, our work provides new mechanistic insights into genes and enzymes involved in bacterial metabolism of anti-cancer drugs. Importantly, we demonstrated that an enzyme conserved in bacterial and mammalian cells significantly alters drug bioavailability and efficacy in mice. This is a critical step forward, demonstrating the feasibility of moving beyond simply cataloging bacterial drug metabolism reactions *in vitro* to testing the physiological and clinical relevance of gut bacterial drug biotransformations. Continued mechanistic insights paired with studies in translationally-relevant model animals and human cohorts, will be essential to follow the well-trod path established by human pharmacogenetic research, achieving the ambitious goal of a more comprehensive, microbiome-informed, approach to precision medicine[64].

Methods

Chemicals.

All materials were purchased from the specified supplier and used without further purification unless otherwise stated. 5-Fluorouracil (5-FU; ≥99% purity), 5-Fluorodeoxyuridine (FUDR; ≥99% purity), imidazol, and dithiothreitol were purchased from Millipore Sigma (St. Louis, MO). Capecitabine (CAP; ≥98% purity) used in this study was from Santa Cruz Biotechnology (Dallas, TX). Dihydrofluorouracil (DHFU; ≥99% purity) was purchased from Toronto Research Chemicals (Toronto, Canada). Internal standards 5-FU C^{13}, N_2^{15} (5-FU-IS; ≥99% purity) and 5-fluorodihydropyrimidine-2,4-dione-C^{13}, N_2^{15} (DHFU-IS; ≥99% purity) were purchased from Santa Cruz Biotechnology Inc. (Santa Cruz, CA). Solvents used for sample preparation were methanol (MeOH), acetonitrile (ACN), and water and were purchased from Honeywell-Burdick & Jackson (Muskegon, MI). Benzamidine and 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride (AEBSF) were purchased from RPI (Mt Prospect, IL). 2-mercaptoethanol was purchased from BioRad (Hercules, CA). Isopropyl β-D-1-thiogalactopyranoside (IPTG) and 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES) were purchased from Thermo Fisher Scientific (Waltham, MA).

Bacterial strain collection.

Gut bacterial strains were obtained from the DSMZ (Braunschweig, Germany) and routinely cultured in BHI^+ media (Supplementary Table 1). Strains were mapped to the MIDAS v1.0 database[65]. If only one MIDAS species matched the strain species name, that MIDAS species was selected. If multiple MIDAS species matched, genomes in the MIDAS v1.0 database were filtered by whether the “culture collection” or “strain” metadata contained

the DSMZ identifier, and then the most common remaining MIDAS species ID was kept. If no genomes matched the DSMZ identifier, the most common MIDAS species ID for that species name was selected. Abundance data came from a meta-analysis of five healthy cohorts described previously[57]. Abundances were summed across runs corresponding to the same individual, and then normalized per-individual to yield relative abundances for each species. The weighted mean of relative abundances was then taken per-species, with weights corresponding to the square root of the sample size. Four species were not represented in the MIDAS v1.0 database.

Minimum inhibitory concentration (MIC) determinations.

All procedures and incubations were performed in an anaerobic chamber (COY Laboratory Products Inc.) with the following atmosphere: 10% H_2, 5% CO_2, and 85% N_2. Reagents were equilibrated in the anaerobic chamber for at least 24 hours before using. MIC determinations were performed in BHI^+ media: Brain Heart Infusion broth supplemented with L-cysteine hydrochloride (0.05% w/v), hemin (5 μg/ml), and vitamin K (1 μg/ml). 5-FU was dissolved in dimethyl sulfoxide (DMSO), supplemented at 1% (v/v) in BHI^+ media, and assayed at concentrations ranging from 0.1–1000 μg/ml. CAP was directly resuspended in BHI^+ media and assayed at concentrations ranging from 0.02–10 mg/ml. Each strain was inoculated in 5 ml of BHI^+ media in a Hungate tube and incubated at 37°C for 24–48 hours, depending on the time required to reach stationary phase. After reaching stationary phase, each strain was diluted to OD_{600nm} of 0.1, further diluted 1:100, and 50 μl was used to inoculate a 96-well plate containing drug for a final volume of 100 μl. Plates were incubated at 37°C for 24–48 hours depending on the strain. The MIC was determined as the concentration of drug that inhibited the growth of a bacterial strain >90% relative to the growth control supplemented with vehicle. MIC assays were performed in duplicate. *Escherichia coli* BW25113 was used to perform the media effects and uracil rescue experiments. These MICs were conducted in triplicate in M9 minimal salts media (M9MM) provided by BD Difco (Franklin Lakes, NJ) supplemented with 0.4% glucose. Uracil from Acros Organics (Geel, Belgium) was supplemented at the indicated concentrations. Assays were performed in a 96-well plate with 100 μl total volume and incubated aerobically at 37°C for 24 hours. *E. coli* mutants deficient in various pyrimidine metabolism genes were obtained from the National BioResource Project (NBRP; Shizuoka, Japan) were assayed in duplicate in M9MM supplemented with 0.4% glucose and 5-FU or CAP.

Transcriptional profiling (RNA-seq).

In triplicate, 100 ml of LB media was inoculated with an overnight culture of *E. coli* MG1655 at a ratio of 1:100. These cultures were incubated aerobically at 37°C with shaking at 250 rpm and anaerobically at 37°C without shaking. Two independent experiments were conducted for treatments under aerobic conditions. Cultures were grown to a mid-exponential phase (OD_{600nm} of 0.5–0.6 and 0.26–0.28 under aerobic and anaerobic conditions respectively) and then 25 ml of culture was added to 25 ml of pre-incubated media with the following conditions: (1) fresh LB (control); (2) 5-FU (final concentration of 16 μg/ml); and (3) CAP (final concentration of 2.5 mg/ml). Drug concentrations were selected to represent 0.5X MIC under the selected media conditions. Cultures

were incubated for an additional 30 min and then 4 ml of cells were harvested and immediately frozen. Bacterial pellets were resuspended with TRI reagent and then subject to mechanical lysis using MP Biomedicals Lysing Matrix E tubes (Solon, OH). Following extraction with chloroform and precipitation with ethanol, RNA was purified using the Life Technologies PureLink RNA Mini Kit (Carlsbad, CA). DNA was removed by using the Life Technologies PureLink DNase Set (Carlsbad. CA). Depletion of rRNA was accomplished using the Illumina Bacterial Ribo-Zero rRNA Removal Kit (San Diego, CA; experiment 1) and Invitrogen Ribominus Bacterial Transcriptome Isolation Kit (experiment 2). RNA fragmentation, cDNA synthesis, and library preparation were performed using the NEBNext Ultra RNA library Prep Kit for Illumina and NEBNext Multiplex Oligos for Illumina (Dual Index Primers) (Ipswich, MA). Samples were dual-end sequenced (2 × 75 bp) using the MiSeq V3 platform (experiment 1) and NextSeq Mid Output platform (experiment 2; Supplementary Table 2). Reads were mapped to the *E. coli* MG1655 genome sequence (NCBI Reference Sequence: NC_000913.3) using Bowtie2[66] and HTSeq[67] was used to count the number of reads to *E. coli* genes. Differential gene expression was analyzed using DESeq[68]. Differentially expressed genes following drug treatment were defined as transcripts exhibiting an absolute log2 fold change ≥1 and a FDR<0.1 relative to the no drug control.

Generation and analysis of 5-FU-resistant mutants.

For the generation of spontaneous 5-FU-resistant mutants on agar media, cultures of *E. coli* MG1655, *Bacteroides fragilis* DSM2151 and *Bacteroides ovatus* DSM1896 were grown anaerobically in BHI$^+$ liquid media overnight. The following day, 100 μl of cultures (~1×10^9 CFUs/ml) were plated on BHI$^+$ agar media supplemented with 1 mg/ml of 5-FU and incubated anaerobically for 48 hours. Representative 5-FU-resistant colonies were picked and restreaked on BHI$^+$ agar supplemented with 5-FU. For the generation of 5-FU-resistant mutants through passage in liquid media, an overnight culture was subcultured 1:100 into fresh BHI$^+$ liquid media with 50 μg/ml of 5-FU. Following overnight incubation, bacterial cultures were subcultured 1:100 into fresh BHI$^+$ liquid media with 250 μg/ml of 5-FU. Finally, a third 1:100 subculturing was performed in fresh BHI$^+$ liquid media supplemented with 1 mg/ml of 5-FU. Following overnight incubation, a bacterial suspension was streaked for single colonies on BHI$^+$ agar media supplemented with 1 mg/ml of 5-FU. 5-FU MICs were determined as described above. Genomic DNA was prepared using the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's directions for Gram-negative bacteria. The *upp* gene, encoding uracil phosphoribosyltransferase, was amplified by PCR using primers upp_Ecoli_F and upp_Ecoli_R for *E. coli* and upp_Bfrag_F and Bfrag_R for *B. fragilis* (see Supplementary Table 10 for primer list). Sanger sequencing of the PCR products was performed by GENEWIZ (South Plainsfield, NJ). For whole genome sequencing, genomic DNA was fragmented to 350 bp on a Covaris S2, and libraries prepared using a TruSeq DNA PCR-Free Library Prep Kit (Illumina) following the manufacturer's instructions. Samples were quantified using a Qubit dsDNA HS Assay, pooled, quantified by qPCR using a KAPA Library Quantification Kit (KAPA Biosystems), and submitted to the UCSF Institute for Human Genetics Genomics Core Facility for 100-base sequencing on an Illumina HiSeq platform. Reads were adapter trimmed using Trimmomatic[69] and mutations identified using Snippy[70] against reference

genome sequences for *Bacteroides ovatus* ATCC 8483 (NZ_CP012938.1) and *Escherichia coli* K12 MG1655 (NC_000913.3).

5-FU inactivation assay.

Starter cultures of gut bacterial strains were grown anaerobically in 5 ml of BHI^+ media in Hungate tubes at 37°C. The following day, bacterial cultures were diluted 1:100 into Hungate tubes with 5 ml of fresh BHI^+ media supplemented with 5-FU at 20 μg/ml. Over time, conditioned media samples were harvested, centrifuged to remove cells, and assayed for residual 5-FU levels. *E. coli* BW25113 Δ*yjjG*, which is a 5-FU-hypersensitive strain[71], was obtained from National BioResource Project (NBRP, Shizuoka, Japan) and used as the indicator organism for the disk diffusion assay. An overnight culture of *E. coli* BW25113 Δ*yjjG* was diluted to an OD_{600nm} of 0.1 in saline. Using sterile swabs, the bacterial suspension was used to inoculate M9MM plus 0.4% glucose and 0.2% casamino acids agar plates and a paper disk was overlaid. Residual 5-FU concentration was assayed by diluting conditioned media samples 1 in 5 in water and 1 μl was applied on the paper disk. Agar plates were incubated aerobically overnight at 37°C. Under these conditions, our assay allowed the detection of between 1–20 μg/ml of 5-FU. Strain designations for bacteria carrying the *preTA* operon that were tested for inactivation but not included in our initial strain collection are: *Anaerostipes caccae* DSM14662 and *Clostridium sporogenes* ATCC15579.

Liquid chromatography-mass spectrometry analysis of 5-FU and metabolites.

We used two separate high-resolution mass spectrometry instruments based upon staff and instrument time constraints. Protocol A: 5-FU and DHFU from conditioned media samples were analyzed by liquid chromatography-quadrupole time-of-flight mass spectrometry (LC-QTOF/MS) using an Agilent LC 1260-QTOF/MS 6550 instrument (Santa Clara, CA). Conditioned media samples (25 μl) were spiked with 2.5 μl of a MeOH mix containing internal standards (10 μg/ml 5-FU $^{13}C^{15}N_2$, and 100 μg/ml DHFU $^{13}C^{15}N_2$). Proteins from conditioned media samples were precipitated with an equal volume of MeOH. The resulting extracts were dried down and thereafter reconstituted with 25 μl of a 10% ACN in water solution and further diluted 20X prior to LC-QTOF/MS analysis. Chromatographic separation was achieved using a Phenomenex Luna NH-$_2$ (50 × 2 mm, 3 mm) connected to a Phenomenex Securityguard™ guard cartridge (4×2 mm) (Torrance, CA) at 35°C. The mobile phase consisted of 20 mM ammonium hydroxide, 20 mM ammonium acetate, 5% ACN in water as solvent A, and 100% ACN as solvent B. A flow rate of 0.8 ml/min was used at the following gradient elution profile: 90% B at 0–0.5 min, gradient to 30% B from 0.5–2 min, 90% B at 2–5 min. 2.5 μl of clarified conditioned media was injected for LC-QTOF/MS analysis. The autosampler was maintained at an internal temperature of 4°C. Eluate from the column was ionized in the QTOF/MS using an electrospray ionization source (ESI) in negative polarity. Qualitative confirmation of 5-FU and DHFU in each sample was done using the Agilent MassHunter Qualitative Analysis (Santa Clara, CA) software. The criteria used for confirmation were as follows: mass error ≤ 10 ppm; retention time within 0.15 min; and target score ≥ 75 (overall indication of match based on mass error, retention time match and isotope abundance match). Quantitative analysis for both 5-FU and DHFU was performed using an isotope dilution method with an 11-point

calibration curve run in triplicate. Data analysis was performed using Agilent MassHunter Quantitative Analysis software. The linear regression of the peak area ratio was weighted 1/x for both 5-FU and DHFU with respective linear regression coefficients of R^2=0.996 and R^2=0.992. The lower limits of quantitation (LLOQ) for 5-FU and DHFU were 100 ng/ml and 4 μg/ml, while the observed upper limits of quantitation (ULOQ) were 1.25 μg/ml and 30 μg/ml, respectively. Protocol B: CAP, 5-FU, and DHFU from conditioned media samples were analyzed by liquid chromatography-triple quadrupole mass spectrometry (LC-MS/MS) using a SCIEX Triple Quad 7500 instrument with a linear ion QTRAP (Redwood City, CA). Conditioned media samples (10 μl) were spiked with 1 μl of a MeOH mix containing internal standards (10 μg/ml 5-FU $^{13}C^{15}N_2$, 100 μg/ml DHFU $^{13}C^{15}N_2$, and 10 μg capecitabine-$^2H_{11}$). Proteins from conditioned media samples were precipitated with 60 μL of MeOH. The resulting extracts were dried down and thereafter reconstituted with 500 μl of a 10% ACN in water solution. Chromatographic separation was achieved using a Phenomenex Synergi column 4 μM Fusion RP-80 (50 × 2 mm) at 35°C. The mobile phase consisted of methanol + 0.1% formic acid for solvent A and HPLC-grade water + 0.1% formic acid as solvent B. A flow rate of 0.4 ml/min was used at the following gradient elution profile: 0% B at 0–2 min, gradient to 100% B from 2–5.9 min, gradient to 0% B at 5.9–6 min. The autosampler was maintained at an internal temperature of 4°C. Eluate from the column was ionized in the LC-MS/MS using an electrospray ionization source (ESI) in positive polarity. Quantitative analysis for both 5-FU and DHFU is performed using an isotope dilution method with a 9-point calibration curve run in duplicate. Data analysis was performed using the built-in SCIEX OS Software. The linear regression of the peak area ratio was weighted 1/x for CAP with a linear regression coefficient of R^2 = 0.995. The lower limits of quantification (LLOQ) for CAP was 1 nM and the upper limit of quantification (ULOQ) for CAP was 2.5 μM. The quartic regression of the peak area ratio was weighted 1/x for both 5-FU and DHFU with quartic regression coefficients of R^2=0.998 and R^2=0.987, respectively. The lower limits of quantitation (LLOQ) for 5-FU and DHFU were 1 μM and 31 μM, while the observed upper limits of quantitation (ULOQ) were 250 μM for both.

Liquid chromatography-mass spectrometry analysis of 5-FU from plasma samples.

Protocol A: The validated LC-QTOF/MS method for the conditioned media described above was adopted to measure 5-FU in the plasma samples. Mouse plasma (9–18 μl) was extracted using protein precipitation with methanol at 3:1 methanol: plasma volume ratio Similarly, the resulting extracts were dried down and thereafter reconstituted in 10% acetonitrile (ACN) at the same volume as the plasma sample. The final extract obtained was then run using the same qualitative and quantitative analyses applied to the conditioned media. The LLOQ and ULOQ observed for 5-FU in mouse plasma were 20 ng/ml and 1500 ng/ml, respectively. Protocol B: The validated LC-MS/MS method for conditioned media was adapted to measure 5-FU in plasma. Mouse plasma (10 μl) was extracted by protein precipitation with methanol (60 μl). The extract was dried under a gentle stream of nitrogen gas and reconstituted in 30 μl of 10% acetonitrile. The final extract was then run using the same quantitative analyses applied to the conditioned media. The LLOQ and ULOQ observed for 5-FU in mouse plasma was 5 ng/ml and 1500 ng/ml, respectively.

Bacterial genetics.

Markerless mutant strains of *E. coli* BW25113 and *E. coli* MG1655 devoid of the *preTA* operon were constructed using the recombineering protocol[72] with the pSIJ8 vector as previously described[73]. The *preTA* operon was replaced using allelic exchange by electroporation of a PCR product containing the kanamycin cassette[74,75], which was generated using the primers preTA-P1-KEIO_F and preTA-P1-KEIO_R (Supplementary Table 10) and using the *E. coli* BW25113 Δ*yjjG* genomic DNA as a template. Allelic exchange of the *preTA* operon with the kanamycin cassette was confirmed by PCR. *preTA* operons identified in diverse gut bacterial genomes were amplified and cloned into the pINT1 vector[76] using primers listed in Supplementary Table 10. A list of *preTA* operons including sequences used for heterologous expression experiments can be found in Supplementary Table 10. The *preTA* operon from *O. formigenes*, *preTA-Of*, was codon optimized for expression in *E. coli* and synthesized by GENEWIZ (South Plainfield, NJ). Integration of the *preTA*-pINT1 constructs into *E. coli* MG1655 Δ*preTA* was performed as previously described[73,76]. Engineered strains were made streptomycin resistant by using the recombineering protocol[72] using pSIJ8[73], as described above. A spontaneous streptomycin-resistant mutant of *E. coli* MG1655 was created by plating an overnight culture on LB agar supplemented with 100 μg/ml of streptomycin and incubated at 37°C overnight. Streptomycin-resistant colonies were picked and restreaked on LB agar supplemented with 100 μg/ml of streptomycin to confirm their resistant phenotype. The *rpsL* gene from streptomycin-resistant isolates were amplified by PCR using the primers rpsL_F (5'-CGTGGCATGGAAATACTCCG-3') and rpsL_R (5'-GCATCGCCCTAAAATTCGGC-3'). PCR products were sequenced by GENEWIZ (South Plainsfield, NJ) using Sanger sequencing with primers rpsL_F and rpsL_R. We selected the RpsL K42R mutation for future experiments[77]. Mutant *rpsL* PCR product was transferred to *preTA* operon engineered strains described above using the recombineering protocol[72] with the pSIJ8 vector as previously described[73] and selected by plating on LB agar supplemented with 100 μg/ml streptomycin. Streptomycin-resistant colonies were picked and restreaked to confirm resistance phenotype. The *rpsL* gene from these colonies were also amplified by PCR and sequenced to confirm point mutation.

Cancer cell culture assays.

HCT-116 cells (ATCC CCL-247) were acquired from the ATCC (Manassas, VA) and grown in McCoy's 5A medium supplemented with 10% fetal bovine serum (FBS) and Gibco antibiotic-antimycotic solution. Cells were grown at 37°C with 5% CO_2. Approximately 2.4×10^4 HCT-116 cells were seeded into a Corning Costar TC-treated 48-well plate in a total volume of 250 ml culture media. Following 24 hours, 28 μl of conditioned media was added to HCT-116 cells to assay for residual 5-FU activity. Conditioned media samples from *E. coli* MG155 cultures and related *preTA* constructs were grown anaerobically for 3 days in BHI^+ media with 5-FU at a concentration of 20 μg/ml with a final DMSO concentration of 0.05%. Conditioned media samples were sterilized by passage through a 0.2 μm filter prior to assaying against HCT-116 cells. Cell proliferation was measured using the Promega CellTiter Non-Radioactive Cell Proliferation Assay (Madison, WI). Cell proliferation was normalized to HCT-116 cells that were supplemented with water instead of conditioned or sterile culture media.

Cloning, expression, and purification of *E. coli* PreTA.

Polycistronic pCOLADuet-1 constructs containing His_6_*Ec*PreT and untagged *Ec*PreA and were kindly provided by Professor Catherine Drennan's Lab at MIT. pDB1282 was kindly provided by Professor Dennis Dean at Virginia Tech. Polymerase Chain Reaction followed by Gibson Assembly was used to remove the N-terminal *Ec*PreT His_6 tag and to include a C-terminal tobacco etch virus (TEV) protease site followed by a His_6-tag to *Ec*PreA such that the final expression vector was pCOLADuet-1_*Ec*PreT_*Ec*PreA-TEV-His_6 (see Supplementary Table 10 for primer list). The final expression construct was confirmed by Sanger Sequencing. pCOLADuet-1_*Ec*PreT_*Ec*PreA-TEV-His_6 was co-transformed with pDB1282 into *E. coli* Rosetta2 (DE3) cells and selected for with kanamycin, chloramphenicol, and ampicillin. pDB1282 (Amp^R) contains the iron sulfur cluster (*isc*) biosynthetic operon from *Azotobacter vinelandii* under the control of the *E. coli* arabinose promoter P_{ara} which has been shown to increase the incorporation of Fe-S clusters during heterologous expression[78]. A single colony of *E. coli* Rosetta2 (DE3) harboring the expression construct and pDB1282 was inoculated in TB media (10 g/L NaCl) supplemented with 50 μg/mL kanamycin, 25 μg/mL chloramphenicol, and 100 μg/mL ampicillin and grown overnight at 37°C with shaking at 225 rpm. The overnight culture was diluted 1:100 into 1 L of TB media supplemented with the same antibiotics as before and grown at 37°C with shaking at 225 rpm until OD_{600nm} = 0.6. Arabinose was added to a final concentration of 0.2% (w/v) and the cultures were allowed to continue growing until OD_{600nm} > 2.0. At this point, the cultures were cooled in an ice bath for 30 min and the following was added to the cultures: 10 μM uracil, 20 μM riboflavin, 10 μM cysteine, 10 μM ferric chloride, and 500 μM IPTG. The temperature of the incubator shaker was lowered to 18°C and the protein was allowed to express with shaking at 225 rpm overnight. Cell pellets were harvested by spinning down 1 L cultures in a pelleting centrifuge, collected in tubes, snap frozen in liquid nitrogen, and stored at –80°C. All following purification steps were carried out at 4°C. Cell pellets were thawed in an ice water bath and then resuspended by vortexing in 150 mL of Lysis Buffer that consisted of Buffer A (50 mM K_2HPO_4 pH 8.0, 200 mM NaCl, 5% glycerol, 0.22 μm filtered) supplemented with 25 mM benzamidine, 1 mM AEBSF, 0.01 mg/mL bovine DNase, 0.1 mg/mL lysozyme, and 5 mM 2-mercaptoethanol. Cells were lysed by sonication using a QSonica Q500 Sonicator at 30% amplitude for 5 minutes of total process time (2 seconds on, 4 seconds off). Whole cell lysate was clarified by centrifuging at 30,000 *g* for 30 minutes. Clontech TALON Superflow resin (3 mL bed volume) was equilibrated with 10 column volumes (CV) of lysis buffer and then added to the supernatant of the clarified lysate and rocked for 30 minutes. The supernatant and resin were applied to a gravity column which was allowed to flow until the resin was collected. Next, the column was washed with 10 CV of Buffer A supplemented with 5 mM 2-mercaptoethanol and steps of increasing imidazole concentration until the protein was eluted off with Buffer B (50 mM K_2HPO_4 pH 8.0, 200 mM NaCl, 150 mM imidazole, 5% glycerol, 0.22 μm filtered). The elution fractions were dialyzed in a ratio of sample to dialysate of 1:500 for 2 hours against Buffer C (50 mM Tris pH 7.5, 150 mM NaCl, 1 mM 2-mercaptoethanol) prior to the addition of TEV protease, and the dialysis was allowed to continue overnight. The cleaved protein was collected in the flowthrough of a TALON Superflow subtractive column and then spin concentrated to 0.5 mL. Finally, the protein was injected onto a GE Superdex 200 Increase 10/300 GL size exclusion column that was pre-equilibrated with Buffer D

(50 mM Hepes pH 7.4, 150 mM NaCl, 5 mM DTT, 5% glycerol, 0.22 μm filtered). Purity was assessed by SDS-PAGE and cofactor incorporation was evaluated with UV-Visible absorption spectroscopy ($A_{378}/A_{280} > 0.35$)[79]. The main band was pooled and concentrated to a concentration of 3–5 mg/mL as assessed by Nanodrop A_{280} ($\varepsilon = 114.6\ M^{-1}cm^{-1}$, MW = 178.74 kDa), snap frozen in liquid nitrogen, and stored at −80°C.

Biochemical characterization of *E. coli* PreTA.

To collect UV-Visible absorption spectra in different oxidation states, *Ec* PreTA was brought into an anaerobic chamber and $Na_2S_2O_4$ was added in an approximately equimolar amount. Samples were placed in a septum-sealed 1 cm path-length quartz cuvette inside the glovebag, and UV-Visible absorption spectra were recorded on a Cary 3500 spectrophotometer (Agilent Technologies). Analytical Size Exclusion Chromatography was performed by injecting three subsequent samples onto a GE Superdex 200 Increase 10/300 column pre equilibrated with Buffer D: 0.25 mg/mL blue dextran for void volume determination, 0.25 mg/mL Protein Standard Mix 15–600 kDa (Millipore Sigma), and a 2 mg/mL *Ec* PreTA sample. Samples were run at 0.5 mL/min. The standard curve was generated by plotting $\log_{10}$ of the molecular weight of the protein standard in Da versus the fractional elution volume ($V_{elution}/V_{void}$). Enzyme assays to confirm product formation were performed by bringing the necessary reagents into a COY anaerobic chamber and diluting into anaerobic Buffer E (25 mM Hepes pH 7.4, 25 mM NaCl, 0.22 μm filtered). Substrates were prepared at only twice the final concentration to ensure complete solubility. The final reaction contained Buffer E supplemented with 5 mM DTT, 800 μM NADH, 400 μM substrate, and was initiated by addition of 100 nM *Ec* PreTA at 25°C. Aliquots were quenched into 2% formic acid. Reactions were analyzed by reverse-phase HPLC to observe product formation by running an isocratic method of 100% Buffer F (0.1% formic acid (FA) in dH_2O, 0.22 μM filtered, degased) followed by a cleaning cycle of 100% Buffer G (0.1% formic in acetonitrile, degased) and reequilibration in Buffer F over a Phenomenex Kinetex 150 × 4.6 mm, 2.6 μM C18 column. Reaction product masses were confirmed by liquid chromatography high resolution mass spectrometry (LC-HRMS) on a SCIEX TripleTOF 6600+. Reactions were applied to a Phenomenex Kinetex F5 (2.6 uM, 100A, 150 × 2.1 mm) column and products were separated with a gradient of 100% water (+ 0.1% FA) to 95% acetonitrile (+0.1% FA) over 12 minutes. Independent data acquisition (IDA) of TOF MS (DP 30, CE 10, ISG1 60, ISG2 80, CUR 35, TEM 400, ISVF 5500) and product ion (DP 50, CE 35, CES 15, IRD 66, IRW 24) spectra were obtained in positive ionization mode. To measure the steady-state kinetics of *Ec* PreTA, depletion of NADH was measured by following absorbance decreases at 340 nm throughout the course of the reaction. The final reaction contained Buffer G (100 mM K_2HPO_4 pH 7.4) supplemented with 5 mM DTT, 400 μM NADH, variable concentrations of substrate, and were initiated by addition of 10 nM *Ec* PreTA for the forward reactions at 25°C and Buffer G (100 mM K_2HPO_4 pH 7.4) supplemented with 5 mM DTT, 1600 μM NAD^+, variable concentrations of substrate, and were initiated by addition of 100 nM *Ec* PreTA for the reverse reactions at 25°C. Initial rates were normalized to a control reaction where no substrate was added and calculated from the linear phase of the time course during which 5–10% of the NADH substrate was consumed.

CAP pharmacokinetic studies.

All animal experiments were approved by the University of California, San Francisco (UCSF) IACUC. Sample sizes were chosen based on pilot experiments to have 95% power at a significance level of 0.01 using an unpaired t-test (G*Power3.1). Mice were randomized based on age, cage of origin, and weight to experimental groups. Pharmacokinetic studies were not blinded. Mice were housed at 67–74°F, 30–70% humidity, and 12/12-hr light/dark cycle. Gnotobiotic and CONV-R mice were fed LabDiet 5021 and LabDiet 5058, respectively. In the gnotobiotic pharmacokinetics experiment, 8–10 week-old female germ-free BALB/c mice were mono-colonized with isogenic *E. coli* strains with various *preTA* operon statuses (Δ*preTA, wt*, *preTA*$^{++}$*).* 7 days post-colonization, the mice were fasted overnight prior to drug administration. CAP used in animals was from LC Laboratories (Woburn, MA) and was freshly prepared in 40 mM citrate buffer (pH 6.0) containing 5% (w/v) gum arabic. CAP was delivered by oral gavage (0.2 ml) at a dose of 500 mg/kg. Blood samples (~25 μl per time point) were obtained from the mouse tail vein at 0, 0.5, 1, 2, 3, 4, 6, 8 hours post CAP administration. Blood samples were collected in Fisherbrand heparinized glass microhematocrit capillary tubes (Waltham, MA) and centrifuged at 3500×g for 15 min to recover blood plasma, which was stored at −80°C until analysis. In the conventionally-raised (CONV-R) specific-pathogen-free pharmacokinetic experiments, 8–10 week-old male BALB/c mice (Taconic Biosciences, Model#: BALB-M) were colonized with either streptomycin-resistant *ΔpreTA* or *preTA*$^{++}$ *E. coli* MG1655 isogenic strains using the streptomycin mouse model (as described in the CAP xenograft tumor model below). They were colonized for 1 week prior to pharmacokinetic experiments. After the overnight fast, CAP was delivered by oral gavage (0.2ml) at the dose of 500 or 1100 mg/kg. Blood was collected at 0, 0.5, 1, 1.5, 2, 3, and 5 hours post CAP administration and processed in the same procedure as described above for gnotobiotic pharmacokinetic experiment. Data distribution was assumed to be normal but this was not formally tested. No data points were excluded.

CAP tumor xenograft model.

All animal experiments were approved by the University of California, San Francisco (UCSF) IACUC. Sample sizes were chosen based on pilot experiments to have 95% power at a significance level of 0.01 using an unpaired t-test (G*Power3.1). Mice were randomized based on weight and baseline tumor size to treatment groups. Tumors were measured blinded without knowledge of colonization or treatment groups. Three independent experiments (see Extended Data Figs. 5a, 5g, and Fig. 4a for designs of experiment 1, 2, and 3, respectively) were performed with female athymic nude mice at 6 weeks of age (Taconic Biosciences, model #: NCRNU-F). HCT-116 cells were grown in McCoy's 5A medium supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin in 3×100 mm tissue-culture treated dishes at 37°C with 5% CO_2 to 85% confluence. Cells were detached with 2 ml of 0.05% trypsin and quenched with 10 ml of media. Cells were pelleted by centrifugation (200 g for 5 min), media removed, resuspended in 3.5 ml of PBS, and kept on ice for the remainder of the procedure. Cold cells were mixed with BD Matrigel™ Basement Membrane Matrix (Franklin lakes, NJ) at a 1:1 ratio. Each animal was put under isoflurane anesthesia, given artificial tears ointment, and received 100 μl of cell-Matrigel™ mix (1×10^6 cells) with a 27G1/2 needle in the subcutaneous space on

the right flank. Tumors were grown for 19 days and 13 days before bacterial colonization and CAP treatment initiation in experiment 1 and 2, respectively. In experiment 3, bacterial colonization and CAP treatment were initiated when the tumors reached between 80–120 mm^3 on a rolling enrollment basis. Mice were colonized with the engineered streptomycin-resistant *E. coli* strains described above (*E. coli* MG1655 Δ*preTA* and *E. coli* MG1655 preTA++) using the streptomycin mouse model[80]. The day before colonization, mice were put on filtered streptomycin (5 g/L for experiment 1 and 3, 5 mg/L for experiment 2) water. Overnight cultures of engineered *E. coli* strains were pelleted by centrifugation, washed with an equal volume of sterile 0.85% saline, pelleted by centrifugation again, and resuspended in 1:10 sterile saline. Each mouse was colonized with 200 μl of this bacterial suspension by gavage. Mice were housed (n=4–5 mice/cage) according to colonization status and drug or vehicle treatments and were on streptomycin tap water for the duration of the tumor xenograft experiment. Streptomycin tap water was freshly prepared and replenished weekly. *E. coli* colonization levels were determined by culturing from fecal pellets. Freshly collected fecal pellets (1–2 pellets per animal) were weighed in a microcentrifuge tube and 1 ml of 1% (w/v) Bacto tryptone solution was added to each tube. Fecal pellets were resuspended by rigorously vortexing for 5 min and large particles were removed by allowing the suspension to settle by gravity for 5 min. 10-fold serial dilutions of fecal pellet suspensions were performed in 1% Bacto tryptone solution and 100 μl of these were spread on MacConkey agar supplemented with 100 μg/ml of streptomycin. Culture plates were incubated aerobically 37°C overnight. Colony-forming units (CFUs) per gram fecal sample were calculated by dividing CFUs by 0.1 ml (volume plated), dilution, and by sample weight. Colonization levels were monitored before and after CAP treatment. CAP used in animals was from LC Laboratories (Woburn, MA) and was freshly prepared daily in 40 mM citrate buffer (pH 6.0) containing 5% (w/v) gum arabic. CAP was delivered by oral gavage (0.2 ml) at a dose of 100 mg/kg for a total of 15 doses over 17 days in experiments 1 and 3 and 18 doses over 22 days in experiment 2. Tumor dimensions were measured with a digital caliper 2–3 times a week. Tumor volume was calculated with the following formula: volume (mm^3) = (length × $width^2$)/2[81]. The remainder of the mice were allowed to reach the humane endpoint defined as tumor length ≥20 mm, tumor ulceration, and body condition score of ≤2. Maximal tumor burden was not exceeded in any of the animals. Data distribution was assumed to be normal but this was not formally tested. No data points were excluded.

16S rRNA gene sequencing of fecal pellets from CAP tumor xenograft model.

Mouse fecal pellets were collected throughout the tumor xenograft experiment and stored at −80°C. DNA was extracted using a ZymoBIOMICS 96 MagBead DNA Kit (Zymo D4302) and 16S rRNA amplicon library was constructed[82] using dual error-correcting barcodes. Briefly, primary PCR was performed as a quantitative PCR using KAPA HiFi Hot Start kit (KAPA KK2502) and V4 515F/806R Nextera primers. The amplified products were diluted 1:100 in UltraPure DNase/RNase-free water and were indexed using sample-specific dual indexing primers. The reactions were quantified using Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen P11496) and pooled at equimolar concentrations. The pooled library was quantified by qPCR using KAPA Library Quantification Kit for Illumina Platforms (KAPA KK4824) and sequenced on Illumina MiSeq platform. The demultiplexed sequences were processed using a 16S rRNA gene analysis pipeline (https://

github.com/turnbaughlab/AmpliconSeq) and analyzed using qiime2R (https://github.com/jbisanz/qiime2R), phyloseq[83], and DESeq[68] in R. Briefly, the ASV tables, phylogenetic tree, taxonomy files from qiime2R outputs along with the sample metadata were used to build a phyloseq object using the phyloseq package. To control for uneven sequencing depth, the samples were sub-sampled to 10,000 sequencing reads. Alpha diversity, beta diversity, and community composition were analyzed using the phyloseq built-in functionalities. DESeq package was used to determine the differentially abundant taxa between the treatment groups.

Bioinformatic analysis of the *preTA* operon in genomes and microbiomes.

Amino acid sequences of PreT and PreA from *Escherichia coli* (Genbank protein ID: AVI56642.1 and AVI56643.1), *Salmonella enterica* (AUX97235.1 and AUX97234.1), *Anaerostipes caccae* (EDR96331.1 and EDR96330.1), *Clostridium sporogenes* (EDU36024.1 and EDU36025.1), and *Oxalobacter formigenes* (ARQ46500.1 and ARQ46501.1), as well as the human DPYD protein (NP_000101.2), were downloaded from the NCBI Protein database[84]. T-COFFEE[85,86] was used to construct multiple alignments of PreT and PreA, each containing the five bacterial PreTA sequences plus human DPYD. Non-aligning columns were removed with trimAl[87], using “automated1” parameter selection. Using HMMER3[88], amino acid profile hidden Markov models (HMMs) were built from the trimmed alignments, and these HMMs were used to search against protein sequences from 9,411 complete genomes in RefSeq[84] (9,032 of which were bacterial). Significant (*e*-value < $1e^{-10}$) hits (5,622 to the PreA HMM and 12,849 to the PreT HMM) were filtered using RefSeq GFF3-formatted genome annotations to find PreT and PreA coding sequences that were from adjacent genes on the same strand in a bacterial genome (yielding 1,704 putative operons). Previously, gut prevalences[57] were calculated for species in the MIDAS v1.0 database[65]. To link our results with these, we first used the tool BURST to look across the RefSeq genomes for close matches to the 15 PhyEco marker genes[89] that MIDAS uses to perform taxonomic identification. We then matched genomes to MIDAS species by requiring at least 10 hits to the same MIDAS species above a 95% nucleotide identity threshold. Finally, to find putative *preTA* operons from uncultured bacterial strains, we analyzed a collection of draft metagenome-assembled genomes (MAGs; N=60,664 in total), each assembled from a single gut microbiome and representing 2,962 distinct gut OTUs in total[58], were searched in the same way as the RefSeq genomes. The phylogenetic species trees used for visualization were both constructed using IQ-TREE[90] on a concatenated protein alignment of single-copy marker genes. Specifically, for the tree of RefSeq genomes, we used HMMER3 and the PhyEco HMMs to identify the 15 single-copy universal marker genes selected in MIDAS, built alignments using Clustal Omega[91], trimmed columns with <60% occupancy, and finally built an approximate-maximum-likelihood tree from the concatenated alignment using IQ-TREE. Two *Methanobrevibacter smithii* genomes were incorporated into the alignment and used as an outgroup to root the tree. Certain Proteobacterial genera of mainly obligate endosymbionts (*Buchnera* and the *Candidatus* genera *Tremblaya, Nasuia, Carsonella, Portiera, Sulcia, Riesia,* and *Babela*) had very long branch lengths and were dropped from the tree to aid visualization. Also, so that species with hundreds of genomes did not dominate the visualization, when there were multiple genomes that had exactly the same species-level taxonomic annotation (both using

the RefSeq NCBI annotations and using the MIDAS mapping) and the same prevalence estimates, a single genome was retained arbitrarily. The tree of MAGs was previously described[58,65]. To find orthologs of the *preTA* operon from our collection of gut bacteria used for screening, we used the *E. coli preTA* operon as the query in a tBLASTn[92] search of our specific organism list (Supplementary Table 1), which is publicly available.

To quantify *preTA* abundance in metagenomes without being misled by other *preT* or *preA* homologs, we applied a conservative five-step pipeline as follows. First, we built a database of genomic regions around the putative *preTA* operons that we identified from MAGs and Refseq genomes above. These regions included any intergenic sequence between the operon and the next annotated genes (or, for MAGs, the end of the contig, whichever was shorter). Second, we then extended this database by adding "decoy" regions, which included the coding and intergenic sequences around *preT* or *preA* homologs that matched our profile HMMs above but were not clustered into an operon. Third, we used vsearch[93] to align shotgun metagenomic reads to the database, retaining only reads that aligned to a putative operon and did not align better to a decoy. To reduce compute time while still allowing us to determine the probable origin of preTA reads, we ran the search in two stages: a) we initially reduced the database size using vsearch, clustering regions at 95% nucleotide identity, and then aligned metagenomes to this reduced database at a 95% ID threshold; then, b) we realigned any reads matching the reduced database against the full database, using a more stringent 97% ID threshold. Fourth, we added an additional coverage filter at the level of regions, dropping any reads mapping to regions that had highly uneven or partial coverage. Regions were kept if, across all samples in the dataset, a) both the preA and preT ORFs had non-zero coverage, and b) the median per-nucleotide coverage was >50% of the mean coverage. Fifth, the resulting read counts were normalized for average genome size using MicrobeCensus[94]. The pipeline was applied to two datasets: a previously published CRC case-control study[59] and our own longitudinal study. The CRC dataset included 575 samples from 179 individuals, 141 of which were from a French population (88 controls and 53 with CRC) and the remaining 38 of which were from a German cohort (all with CRC). We followed the original authors' classification of samples, dropping large adenomas and classifying small adenomas as controls. The longitudinal study had 54 samples from 11 CRC patients treated either with CAP alone (n=2) or CAP plus immunotherapy (n=9). For the taxonomic classification of preTA reads, we used the Genome Taxonomy Database (GTDB) taxonomy[95] annotation of the best hit in our region database.

Gut Microbiome and Oral Fluoropyrimidine Study in Patients with Colorectal Cancer (GO) sample collection and metagenomic sequencing.

This ongoing observational study is registered at ClinicalTrials.gov under the identifier NCT04054908, which includes detailed descriptions of the study population and enrollment criteria. All subjects provided informed consent to participate in the study, which was approved by the University of California, San Francisco (UCSF) Institutional Review Board. Participants were recruited and sampled, without any compensation, at UCSF from the study start date (04/13/2018) through March 2019. Patients with CRC who met the following criteria were recruited and screened for eligibility: (i) male or female aged 18 or older, (ii) histologically confirmed colorectal adenocarcinoma, (iii) expected to receive

oral fluoropyrimidine therapy, (iv) able to read and speak English, and (v) willing and able to provide informed consent. Additionally, patients who met any of the following criteria were excluded: (i) known HIV positive diagnosis, (ii) prior chemotherapy, biologic or immunotherapy in the previous 2 weeks, (iii) concurrent rectal radiation therapy, (iv) exposure to antibiotics longer than 2 weeks in the last 6 months, or (v) exposure to antibiotics in the past 4 weeks prior to starting oral chemotherapy. Patients were enrolled to one of the following cohorts: Cohort A received oral CAP treatment as part of standard-of-care therapy; cohort B received TAS-102 (trifluridine-tipiracil) with or without Y-90 radioembolization; and cohort C received CAP plus immunotherapy (pembrolizumab) and bevacizumab as part of a clinical trial. The enrollment criterion was amended to enroll patients in cohort A who received concurrent rectal radiation therapy with CAP (see Supplementary Table 12 which includes one patient who received concurrent rectal radiation therapy). CAP and TAS-102 tablets were taken per the FDA labels for oral dosing. CAP tablets were taken twice daily on days 1–14 of a 21 day cycle; TAS-102 tablets were taken twice daily on days 1–5 and 8–12 of a 28 day cycle. Stool collection occurred at baseline before chemotherapy initiation and on Day 1 of Cycles 1, 2 and 3. During cycle 1, stools were collected at two additional time points: (i) 2 days after therapy initiation on Day 3, and (ii) at the midpoint (Day 7 for cohort A and C, Day 10 for cohort B). Stool samples were collected on fecal occult blood test (FOBT) cards for all timepoints. Additional bulk stool scoop samples were collected at baseline for culturing work. Stool samples were stored at −80°C upon receipt at UCSF. At the time of sequencing, there were 11 GO study participants who completed at least one cycle of treatment: 9 from Cohort A and 2 from Cohort C. 54 stool samples collected on FOBT cards from these 11 participants were used to extract DNA using Protocol Q from International Human Microbiome Standards (http://microbiome-standards.org/). 300ng of normalized DNA was used in the Nextera DNA Flex library prep kit along with DNA UD Indexes Set A barcodes (Illumina) to assemble the metagenomic library. A blank FOBT card was included as a negative control. ZymoBIOMICS microbial community standard containing 8 bacteria and 2 yeasts (Zymo Research) was included as a positive control to assess bias and errors in the metagenomic library preparation. This mock community was significantly correlated with the theoretical composition (Pearson's correlation = 0.805; *p*-value 0.00496). The libraries were individually assessed with PicoGreen (ThermoFisher) and TapeStation 4200 (Agilent) for quantity and quality checks. The pooled library was sequenced using S1 flow cell on NovaSeq 6000 system (Illumina) at the Chan Zuckerberg Biohub.

Selective culturing from colorectal cancer patient samples.

Baseline bulk stool scoop samples from 22 GO study participants were thawed on ice. Three participants did not submit proper baseline samples and were excluded. All processing and isolations were done under aerobic conditions. A single sample from each patient stool was taken and resuspended in BHI^+ media at a ratio of 15 ml/g stool. Stool samples were rigorously vortexed for 5 min and then allowed to settle for 10 min. Stool slurries were serially diluted 10-fold and 1 ml samples of these various dilutions were spread on MacConkey agar on 150 × 15 mm petri plates. Agar plates were incubated aerobically at 37°C for 20 hours. The following day CFUs were counted, and CFUs/gram of sample were calculated by dividing CFUs by dilution and sample weight. Isolated colonies were

pooled by adding 1 ml of BHI^+ media supplemented with 15% (w/v) glycerol on top of the agar plate, resuspending and pooling colonies with a cell spreader, and recovering the cell suspension. A starter culture was initiated by using 50 μl of these cell suspensions to inoculate 5 ml of anaerobically equilibrated BHI^+ media in Hungate tubes and incubated anaerobically at 37°C for 24 hours. The following day 50 μl of starter culture was used to initiate the 5-FU inactivation assay as described above.

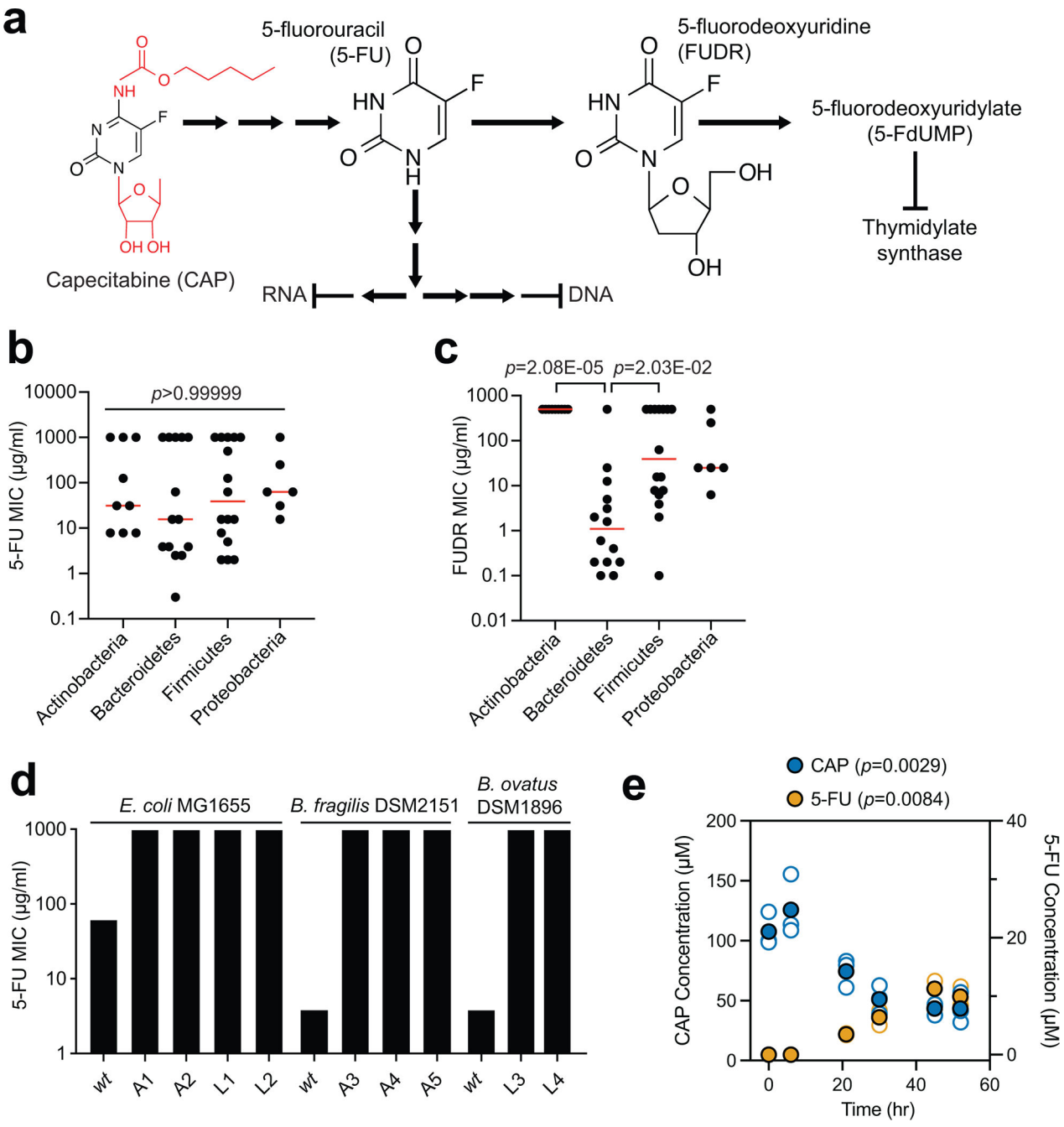
Data Availability

The raw data for our mouse experiments as well as uncropped gel images and tumor photographs can be found in the source data section. The Genome Taxonomy Database[95], Kyoto Encyclopedia of Genes and Genomes (KEGG) database[98], and MIDAS v1.0 database[65] are publicly available. Sequencing data have been deposited under the NCBI BioProjects PRJNA576932 (RNA-seq, 16S rRNA gene sequences, and isolate genomes) and PRJNA720145 (GO Study metagenomic data).

Code Availability

Source code for our analyses of bacterial *preTA* can be found at https://bitbucket.org/pbradz/preta/.

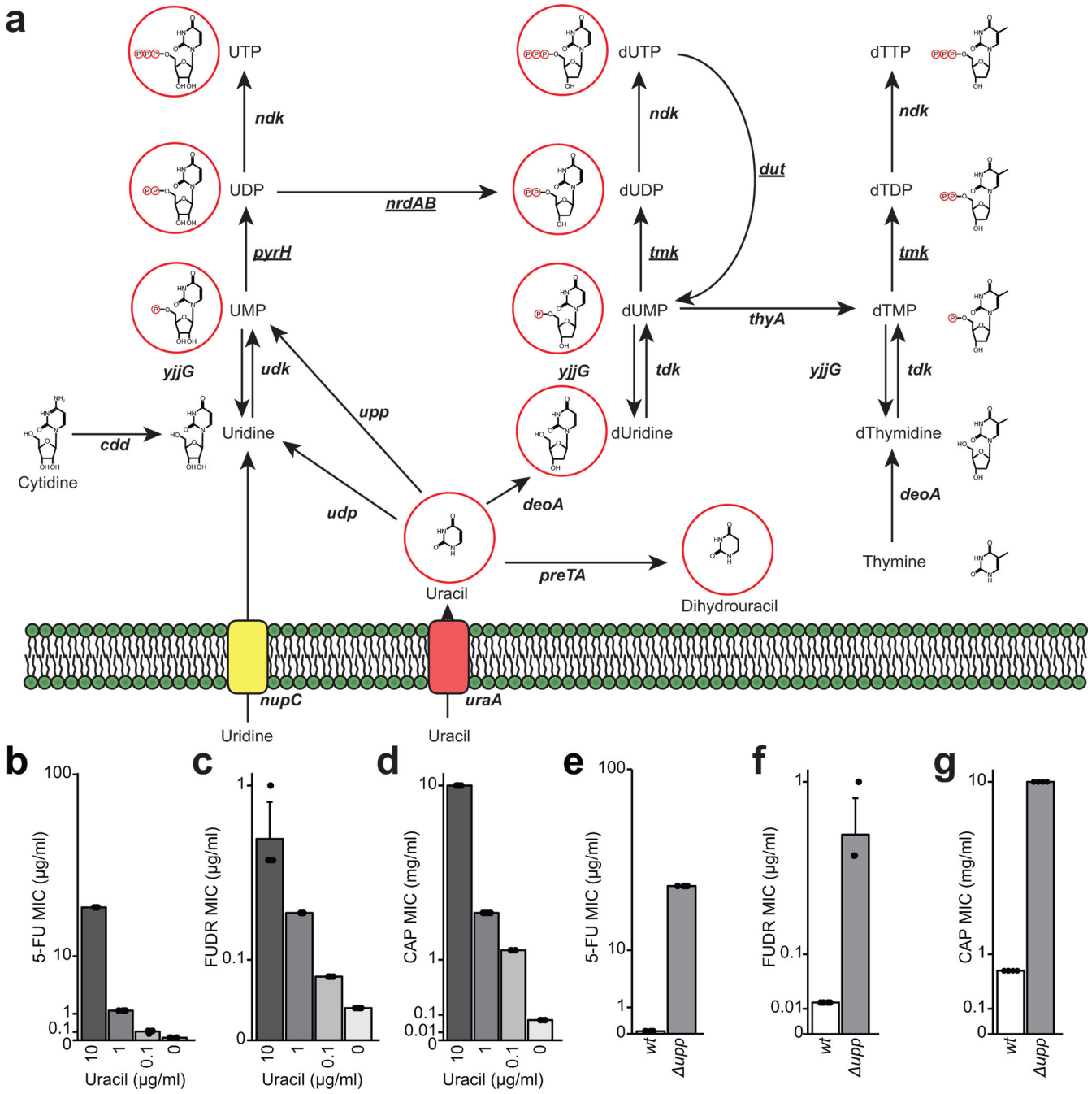
Extended Data



Extended Data Fig. 1. Bacterial taxa vary in their sensitivity to fluoropyrimidines despite the rapid emergence of resistance during *in vitro* growth.

(a) Simplified metabolic pathway for the bioactivation of the oral prodrug capecitabine (CAP) to 5-fluorouracil (5-FU), 5-fluorodeoxyuridine (FUDR), and downstream metabolites[82]. Red indicates chemical groups hydrolyzed during the conversion of CAP to 5-FU. Sequential reactions are indicated by multiple arrows. **(b,c)** Variation in **(b)** 5-FU and **(c)** FUDR minimal inhibitory concentration (MIC) at the phylum level. Bacterial strains

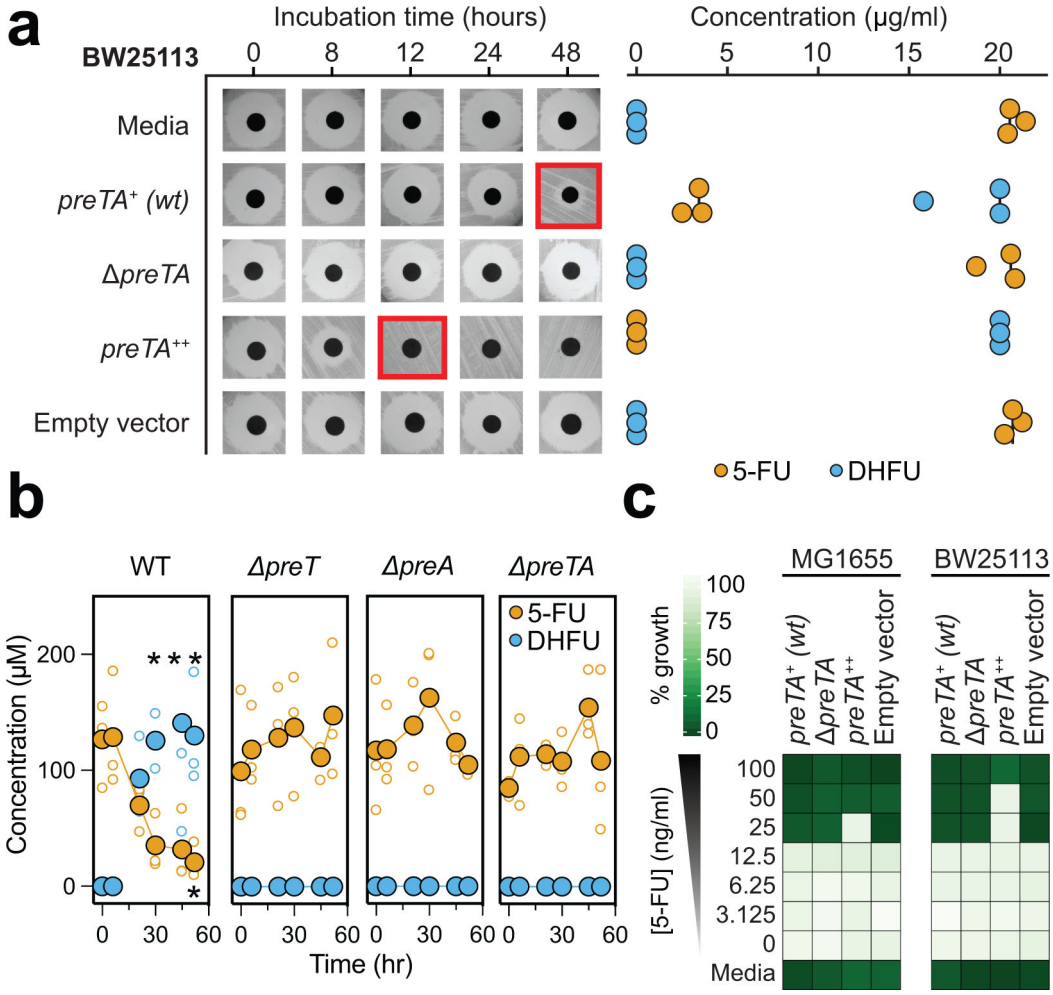
where no MIC could be determined were set to the maximum tested concentration. Each dot represents a bacterial isolate. Red lines indicate the median. *p*-values, Kruskal-Wallis test with Dunn's correction for multiple comparisons. **(d)** 5-FU MICs of parent and 5-FU-resistant strains of *E. coli* MG1655, *B. fragilis* DSM2151, and *B. ovatus* DSM1896 (see Supplementary Table 2). The type of selection is indicated by the strain identifier: A=agar; L=liquid. **(e)** Wild-type *E. coli* BW25113 was assayed for conversion of CAP to 5-FU by LC-MS/MS (n=3 biological replicates per group). Open circles represent individual values, filled circles represent mean values. *p*-values, two-tailed paired *t* test comparing final vs. baseline values for each analyte.



Extended Data Fig. 2. Multiple fluoropyrimidines disrupt the pyrimidine metabolism pathway in *E. coli*.

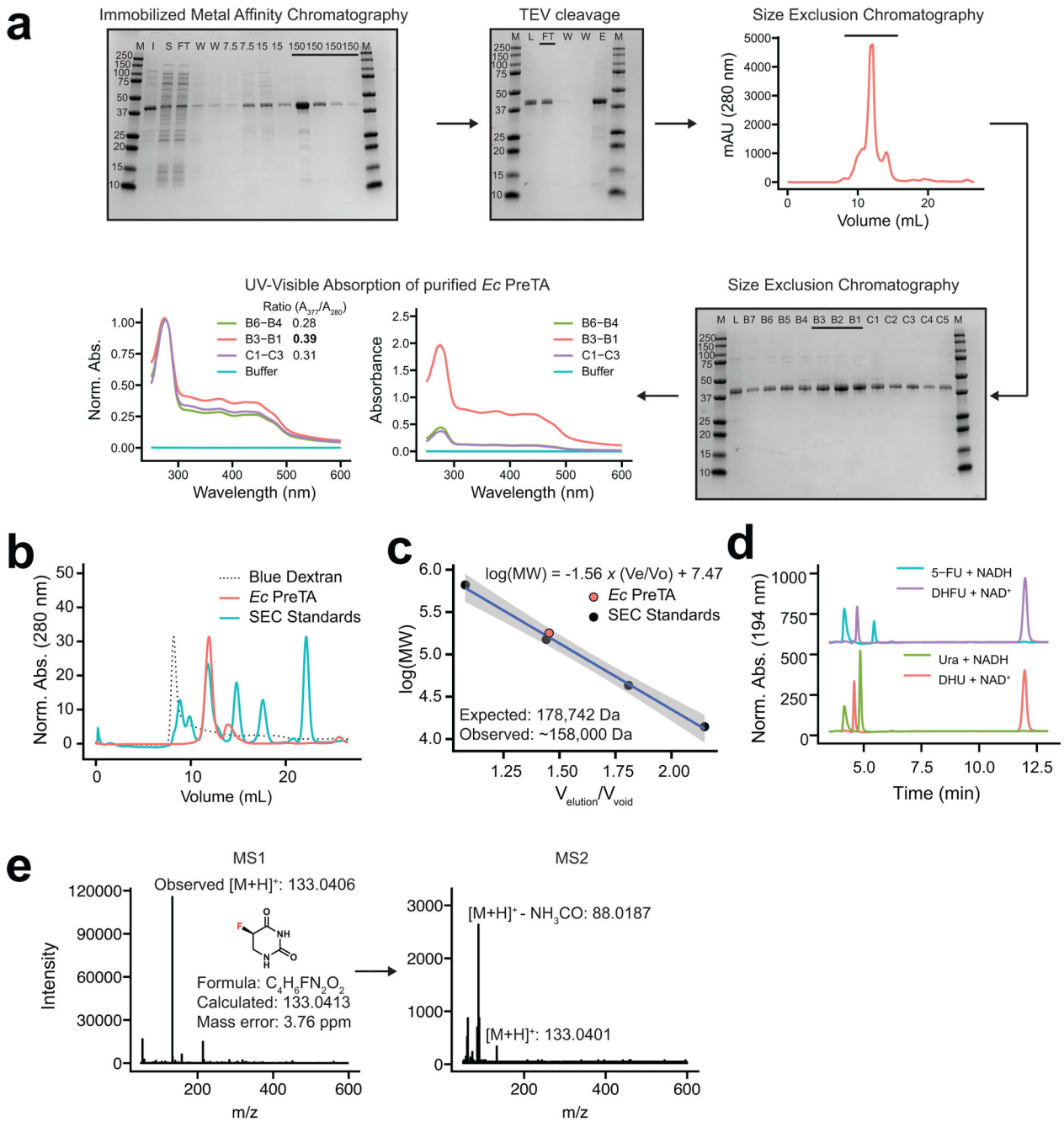
(a) A working model for uracil and uridine import and metabolism in bacterial and mammalian cells. Underlined genes are essential for *E. coli* growth and metabolites with red circles indicate putative 5-FU metabolites. **(b-d)** Uracil rescues the growth of *E. coli* BW25113 in the presence of **(b)** 5-FU, **(c)** FUDR, and **(d)** CAP in a dose-dependent manner. M9MM plus glucose was used as the base media with added uracil from 0.1–10 μg/ml. **(e-g)** A loss-of-function mutation in uracil phosphoribosyltransferase gene (Δ*upp*) rescues the growth of *E. coli* BW25113 in the presence of **(e)** 5-FU, **(f)** FUDR, and **(g)** CAP.

MIC assays performed in M9MM plus glucose. Values in panels b-g are mean±stdev (n=4 biological replicates per group).



Extended Data Fig. 3. The *preTA* operon is necessary and sufficient for the inactivation of 5-fluorouracil in an independent *E. coli* background and has a modest impact on growth.

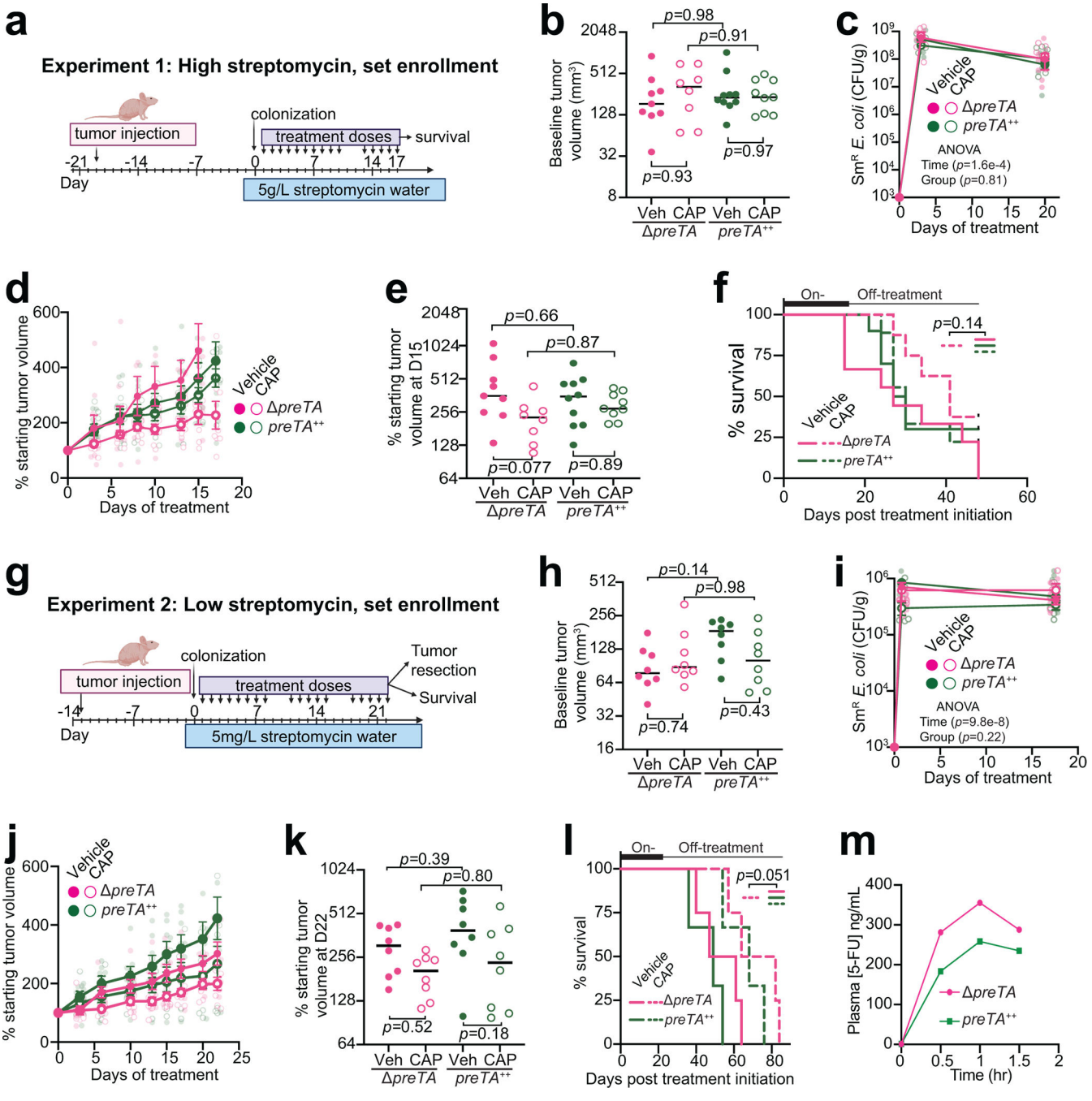
(a) Wild-type (*preTA*$^{+}$), deletion (*ΔpreTA*), complemented (*preTA*$^{++}$), and empty vector *E. coli* BW25113 strains were assayed for residual 5-FU using disk diffusion (0–48 hours incubation) and LC-QTOF/MS (48 hours). Complementation and empty vector were on the *ΔpreTA* background. **(b)** Wild-type, single gene deletion (*ΔpreT*, *ΔpreA*) and operon deletion (*ΔpreTA*) strains of *E. coli* BW25113 were assayed for conversion of 5-FU to DHFU by LC-MS/MS. Open circles represent individual values, filled circles represent means [n=3; *p-value<0.05; 5-FU: 52 hr, p=0.041; DHFU: 30 hr, p=0.011; 45 hr, p=0.004; 52 hr, p=0.009, 2-way ANOVA with Tukey's correction relative to baseline of the same analyte]. **(c)** 5-FU MIC determination of the *E. coli* strains shown in panel a and Fig. 1d in minimal (M9MM) media. Values are normalized to the growth control (no 5-FU) with darker colors indicating growth inhibition. Sterile media is shown in the final row.



Extended Data Fig. 4. Purification and biochemical characterization of *E. coli* PreTA.

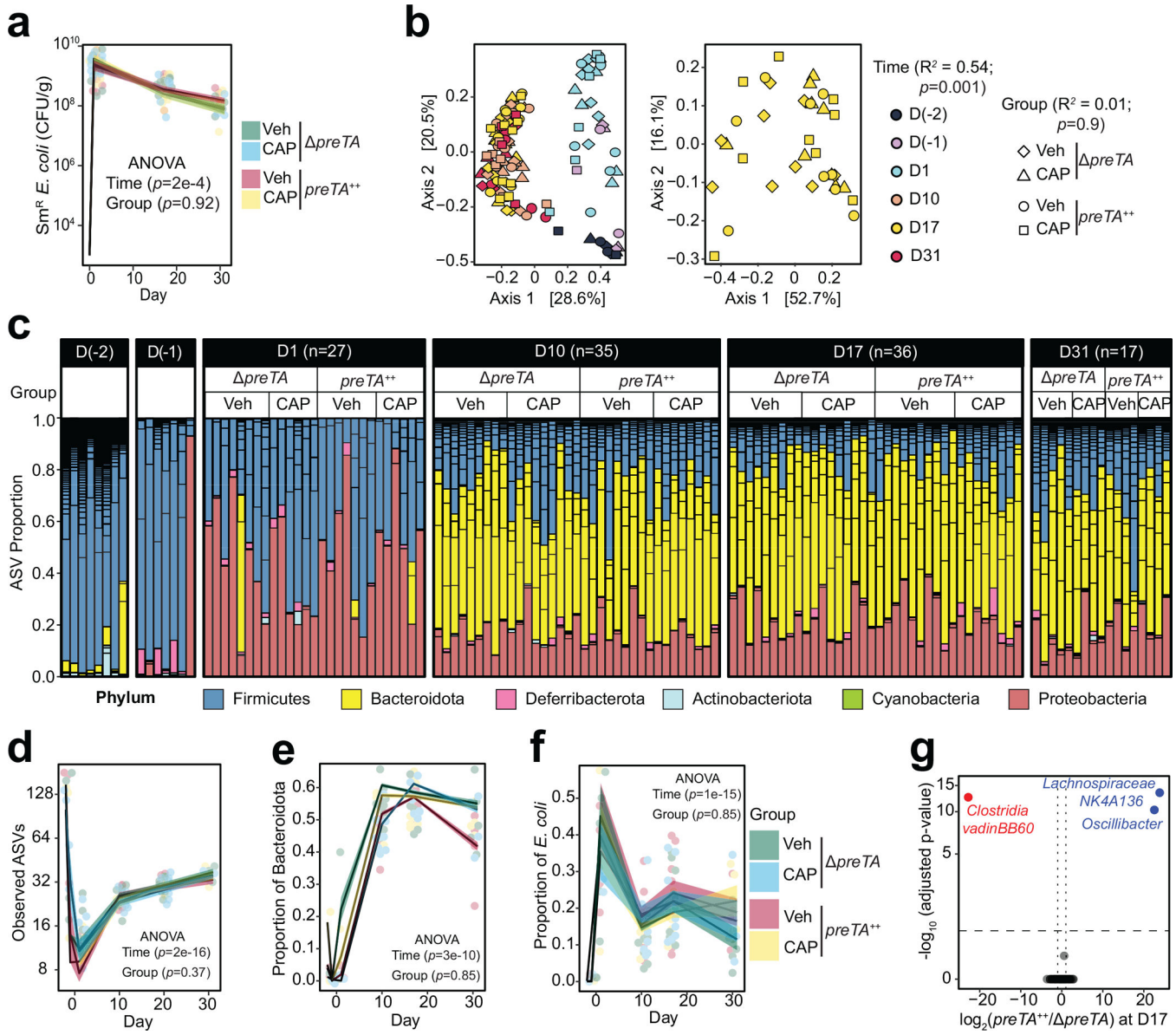
(a) Purification workflow for *E. coli* PreTA including immobilized metal affinity chromatography (M: marker, I: insoluble fraction, S: soluble fraction, FT: flowthrough, W: wash, numbers across the top indicate imidazole concentration), TEV cleavage (M: marker, L: loaded, FT: flowthrough, W: wash, E: elute), size exclusion chromatography (M: marker, L: loaded, lanes are labeled by fraction from a 96 well plate collected via FPLC), and by UV-visible absorption spectroscopy where normalizing protein levels showed a ratio of greater than 0.35 for A_{280}/A_{377} indicated holoenzyme. In all gels, numbers down the side indicate protein molecular weight in kDa and solid lines indicate lanes that

were carried forward in the preparation. **(b-c)** Analytical size exclusion chromatography (SEC) traces **(b)** and analysis **(c),** characterizing the main peak as a heterotetramer. Grey shaded region indicates 95% confidence intervals of the linear model. **(d)** High pressure liquid chromatography (HPLC) of NADH, uracil (Ura), dihydrouracil (DHU), NAD^+, 5-fluorouracil (5-FU), and dihydrofluorouracil (DHFU). **(e)** Liquid Chromatography Mass Spectrometry (LC-MS/MS) of enzymatic reaction with 5-FU confirms the presence of the exact mass of DHFU.



Extended Data Fig. 5. PreTA decreases the efficacy and oral bioavailability of capecitabine (CAP).

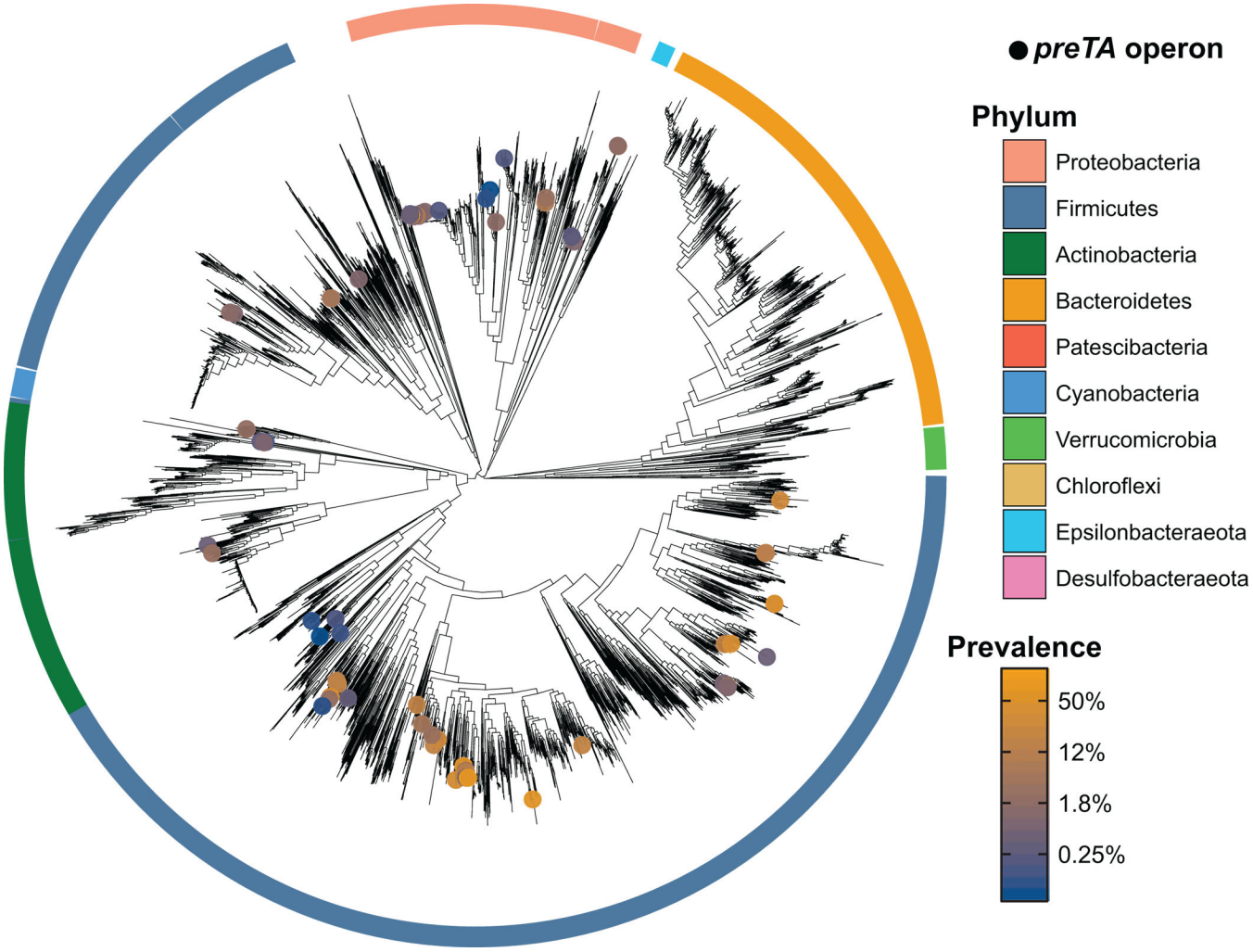
(a, g) Xenograft experiment (expt) 1 and 2 design, respectively. **(b, h)** Pre-treatment tumor volumes: **(b)** expt1, **(h)** expt2 (n=8 mice Δ*preTA*-CAP; n=10 mice *preTA*$^{++}$-Veh; n=9 mice/group remainder expt1; n=8 mice/group expt2; lines represent medians; 2-way ANOVA with Tukey's correction). **(c, i)** Colonization levels of streptomycin-resistant (Sm^R) *E. coli* in colony-forming units (CFU)/gram stool in **(c)** expt1 and **(i)** expt2 (n=6 mice/group expt1; n=8 mice/group expt2; opaque dots and lines represent mean±SEM; 2-way ANOVA with Tukey's correction). Zero values at baseline replaced with our limit of detection (1000 CFU/g). **(d, j)** Percentage of starting tumor volumes over time in **(d)** expt1 and **(j)** expt2 (n=8 mice Δ*preTA*-CAP group; n=10 mice *preTA*$^{++}$-Veh; n=9 mice/group remainder expt1; n=8 mice/group expt2; opaque dots and lines represent mean±SEM; 2-way ANOVA with Tukey's correction did not reach significance). **(e, k)** Percentage of starting tumor volumes on **(e)** day 15 expt1 and **(k)** day 22 expt2 (n=8 mice Δ*preTA*-CAP group; n=10 mice *preTA*$^{++}$-Veh; n=9 mice/group remainder expt1; n=8 mice/group expt2; timepoints selected to capture all mice prior to euthanasia; lines represent medians; 2-way ANOVA with Tukey's correction). **(f, l)** Percentage of mice reaching the humane endpoint in **(f)** expt1 and **(l)** expt2 (expt1: n=8 mice Δ*preTA*-CAP group; n=10 mice *preTA*$^{++}$-Veh; n=9 mice/group remainder; expt2: n=3 *preTA*$^{++}$ groups, n=4 Δ*preTA* groups; log-rank Mantel-Cox test comparing Δ*preTA*-CAP to all other groups. 8 mice were censored (black boxes) as they did not reach the endpoint in expt1. **(m)** LC-QTOF/MS quantification of 5-FU from pooled plasma samples following 500 mg/kg CAP administration in mice (same design as Fig. 4g; n=5 mice/group pooled into n=1 sample/group).



Extended Data Fig. 6. Consistent shifts in the gut microbiota across groups.

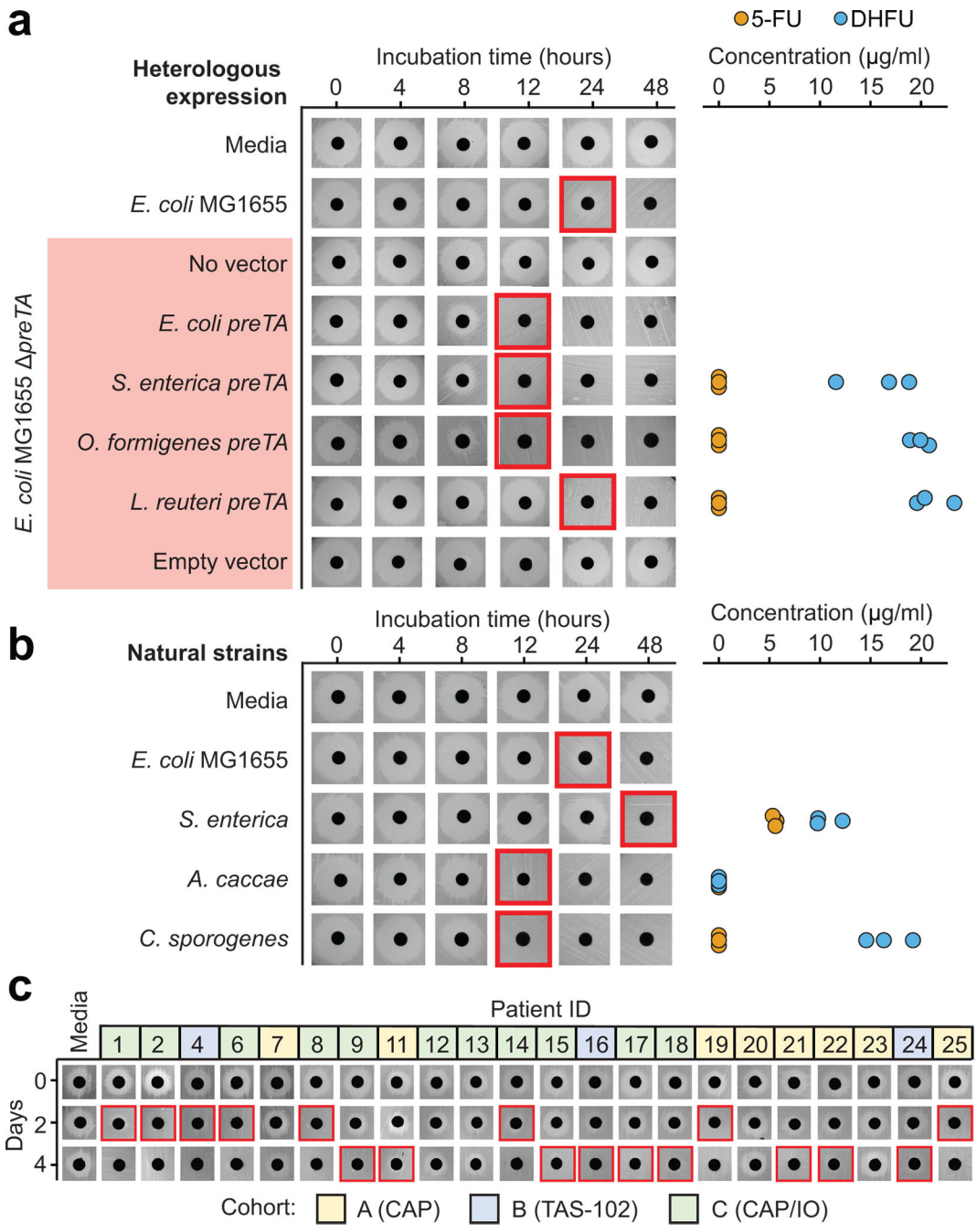
(a) Quantification of streptomycin-resistant (Sm^R) *E. coli* in feces across time (n=11 mice for for *preTA*$^{++}$-Veh group and n=10 mice/group for the rest; 2-way ANOVA test with Tukey's correction; lines and ribbons represent mean±SEM). Zero values at baseline replaced with our limit of detection (1000 CFU/g). **(b)** Principal coordinate analysis of fecal microbiota from mice treated with CAP or vehicle (Veh) and colonized with Δ*preTA* or *preTA*$^{++}$ *E. coli* across time (Bray-Curtis distance matrix, permutational multivariate analysis of variance test with Benjamini-Hochberg correction using ADONIS statistical package). **(c)** Microbial community composition at the phylum level. Each bar represents stool from each mouse. Short horizontal lines within bars represent different amplicon sequence variants (ASV). **(d)** Number of ASVs over time for each treatment group (exact n values indicated in panel c; 2-way ANOVA with Tukey's correction; lines and ribbons

represent mean±SEM). **(e, f)** Proportion of **(e)** Bacteroidota and **(f)** *E. coli* over time for each treatment group (exact n values indicated in panel c; 2-way ANOVA with Tukey's correction; lines and ribbons represent mean±SEM). **(g)** Volcano plots of differentially abundant sequence variants at day 17 (FDR<0.1, |log2 fold-change|>1, Wald test with Benjamini-Hochberg correction, significance limits are marked with dash lines). Blue and red texts represent enriched and depleted taxa, respectively.



Extended Data Fig. 7. Distribution of *preTA* operons across gut metagenome-assembled genomes (MAGs).

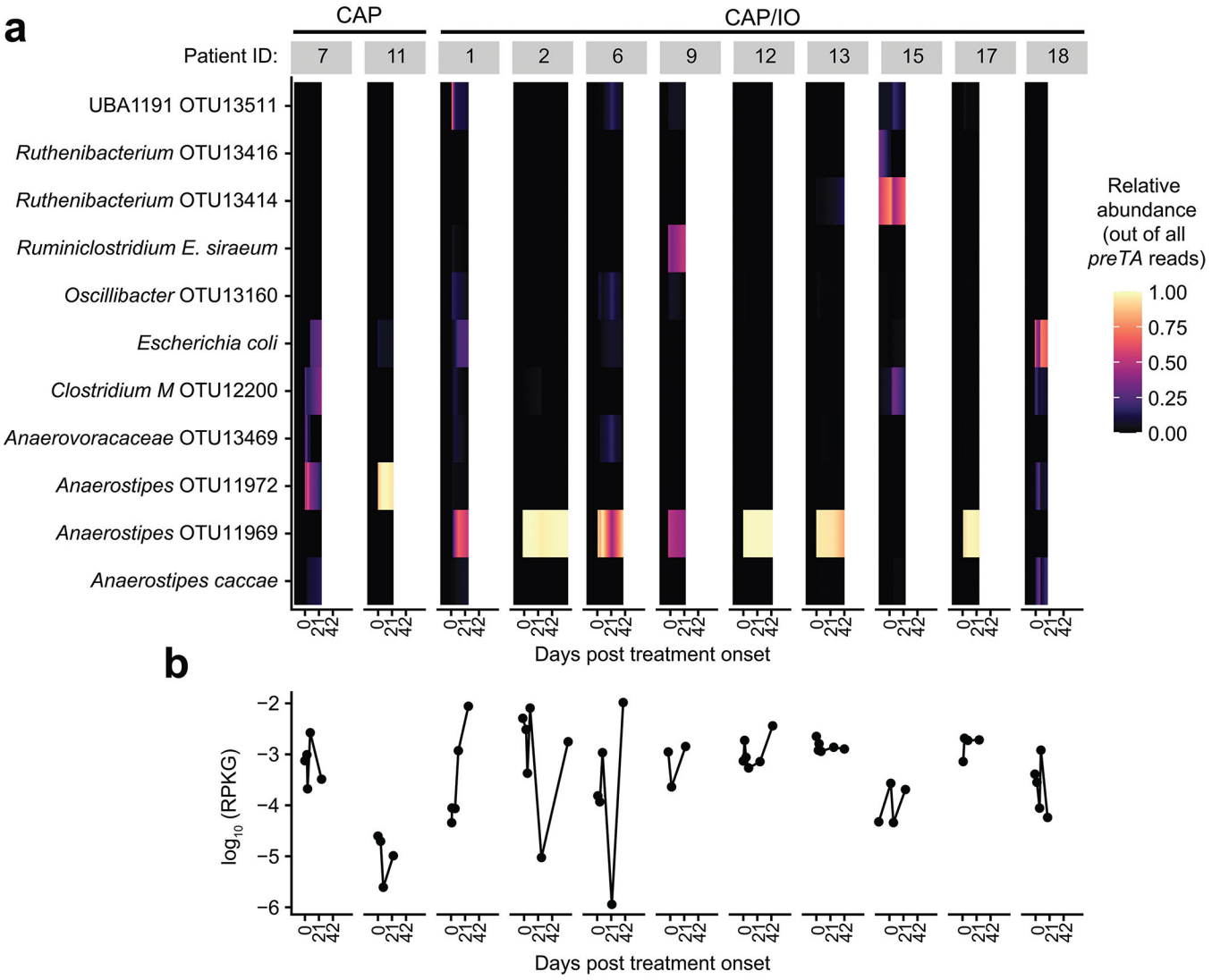
A phylogenetic tree of species from IGGdb, made using a concatenated alignment of single-copy marker genes, is shown in black. Species in which a MAG contains a putative *preTA* operon are identified with colored circles, where the color of the circle corresponds to prevalence in the gut (blue: low prevalence; orange: high prevalence). All species displayed were detected at least once from 3,810 gut metagenome samples. Phylum-level annotations are shown as colored ring segments surrounding the tree for the ten phyla with the most gut MAGs. Data can be found in Supplementary Table 9.



Extended Data Fig. 8. Bacterial *preTA* operons from diverse strains, *preTA* encoding natural strains, and pooled isolates from patient samples are capable of 5-FU inactivation.

(a) *preTA* operons from *Salmonella enterica* LT2 (DSM17058), *Oxalobacter formigenes* ATCC35274, and *Lactobacillus reuteri* DSM20016 were amplified and integrated into the chromosome of *E. coli* MG1655 Δ*preTA*. Strains were incubated with 20 µg/ml 5-FU and assayed for residual drug using disk diffusion (0–48 hours incubation) and LC-QTOF/MS (48 hours). *E. coli* MG1655 *preTA* complementation and empty vector controls are also included on the *ΔpreTA* background. **(b)** We also tested strains predicted to encode the *preTA* operon: *Salmonella enterica* LT2, *Anaerostipes caccae* DSM14662, and *Clostridium*

sporogenes DSM795. Residual 5-FU was assayed as in panel a. Of note, we did not detect either compound from *A. caccae*, suggesting that this strain may further metabolize DHFU. **(c)** Bacterial isolates from 22 CRC patient stool samples in the GO Study (Supplementary Table 12) were isolated on McConkey agar, pooled, and incubated in BHI^+ with 5-FU for 4 days. Residual 5-FU levels were assessed using a disk diffusion assay. Colors indicate the treatment cohort: (A) CAP (yellow); (B) TAS-102 (blue); and (C) CAP plus immunotherapy (green). Red boxes in panels a-c indicate the first timepoint without a clear zone of inhibition.



Extended Data Fig. 9. *preTA* sources and abundance in CRC patients during fluoropyrimidine treatment.

(a) Fraction of total *preTA* reads mapping to different species in metagenomic samples from GO Study patients undergoing CAP treatment with or without pembrolizumab immunotherapy (IO). Heatmap values were linearly interpolated between samples (filled circles in panel **b**). **(b)** Total abundance, as $\log_{10}$ RPKG, of the *preTA* operon in the gut

microbiome prior to and during treatment with the oral fluoropyrimidine CAP with or without IO, as shown in Fig. 6d. Lines connect measurements (filled circles) for the same patient. One zero RPKG value was replaced with half the minimum non-zero value prior to taking the logarithm. The first day of treatment is defined as day 1. Days are the same between panels.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We are indebted to the other members of the Gerona, Goga, Pollard, and Turnbaugh labs for their helpful suggestions during the preparation of this manuscript. We thank Brian Yu, Michelle Tan, and Rene Sit from the Chan-Zuckerberg Biohub for assistance with DNA sequencing; Gerry Wright and Linda Ejim (McMaster University) for providing pINT1; Jordan Bisanz for advice with computational methods; Kyle Spitler in the UCSF Quantitative Metabolite Analysis Center for help with analytical methods; and the GO Study clinical research coordinators (Dalila Stanfield, Paige Steiding, and Julia Whitman). This work was supported by the National Institutes of Health (R01HL122593; R21CA227232; R01CA223817) and the Searle Scholars Program (SSP-2016–1352) and CDMRP W81XWH-18–1-0713. P.J.T. is a Chan Zuckerberg Biohub Investigator and was a Nadia's Gift Foundation Innovator supported, in part, by the Damon Runyon Cancer Research Foundation (DRR-42–16). A.G. was in part supported by a MARK Foundation Endeavor Award. Fellowship support was provided by the Canadian Institutes of Health Research (P.S. and K.N.L.) and the NIH (J.V.L. - F32CA243548, T32CA108462; T.S.K. - F30CA257378). B.G.H.G. is a Connie and Bob Lurie Fellow of the Damon Runyon Cancer Research Foundation (DRG-2450–21).

References

1. Thorn CF, Klein TE & Altman RB PharmGKB: the Pharmacogenomics Knowledge Base. Methods Mol. Biol. 1015, 311–320 (2013). [PubMed: 23824865]
2. Spanogiannopoulos P, Bess EN, Carmody RN & Turnbaugh PJ The microbial pharmacists within us: a metagenomic view of xenobiotic metabolism. Nat. Rev. Microbiol. 14, 273–287 (2016). [PubMed: 26972811]
3. Lam KN, Alexander M & Turnbaugh PJ Precision Medicine Goes Microscopic: Engineering the Microbiome to Improve Drug Outcomes. Cell Host Microbe 26, 22–34 (2019). [PubMed: 31295421]
4. Zimmermann M, Zimmermann-Kogadeeva M, Wegmann R & Goodman AL Mapping human microbiome drug metabolism by gut bacteria and their genes. Nature (2019) doi:10.1038/s41586-019-1291-3.
5. Javdan B et al. Personalized Mapping of Drug Metabolism by the Human Gut Microbiome. Cell 181, 1661–1679.e22 (2020). [PubMed: 32526207]
6. Wallace BD et al. Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. Science 330, 831–835 (2010). [PubMed: 21051639]
7. Biernat KA et al. Structure, function, and inhibition of drug reactivating human gut microbial β-glucuronidases. Sci. Rep. 9, 825 (2019). [PubMed: 30696850]
8. Iida N et al. Commensal bacteria control cancer response to therapy by modulating the tumor microenvironment. Science 342, 967–970 (2013). [PubMed: 24264989]
9. Sivan A et al. Commensal Bifidobacterium promotes antitumor immunity and facilitates anti-PD-L1 efficacy. Science 350, 1084–1089 (2015). [PubMed: 26541606]
10. Vétizou M et al. Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. Science 350, 1079–1084 (2015). [PubMed: 26541610]
11. Viaud S et al. The intestinal microbiota modulates the anticancer immune effects of cyclophosphamide. Science 342, 971–976 (2013). [PubMed: 24264990]

12. Haiser HJ et al. Predicting and manipulating cardiac drug inactivation by the human gut bacterium Eggerthella lenta. Science 341, 295–298 (2013). [PubMed: 23869020]
13. Maurice CF, Haiser HJ & Turnbaugh PJ Xenobiotics shape the physiology and gene expression of the active human gut microbiome. Cell 152, 39–50 (2013). [PubMed: 23332745]
14. Maini Rekdal V, Bess EN, Bisanz JE, Turnbaugh PJ & Balskus EP Discovery and inhibition of an interspecies gut bacterial pathway for Levodopa metabolism. Science 364, (2019).
15. Nayak RR et al. Methotrexate impacts conserved pathways in diverse human gut bacteria leading to decreased host immune activation. Cell Host Microbe (2021) doi:10.1016/j.chom.2020.12.008.
16. Artacho A et al. The Pretreatment Gut Microbiome Is Associated With Lack of Response to Methotrexate in New-Onset Rheumatoid Arthritis. Arthritis Rheumatol 73, 931–942 (2021). [PubMed: 33314800]
17. Bisanz JE, Spanogiannopoulos P, Pieper LM, Bustion AE & Turnbaugh PJ How to Determine the Role of the Microbiome in Drug Disposition. Drug Metab. Dispos. 46, 1588–1595 (2018). [PubMed: 30111623]
18. Zimmermann M, Zimmermann-Kogadeeva M, Wegmann R & Goodman AL Separating host and microbiome contributions to drug pharmacokinetics and toxicity. Science 363, (2019).
19. Arruebo M et al. Assessment of the evolution of cancer treatment therapies. Cancers 3, 3279–3330 (2011). [PubMed: 24212956]
20. Gadiko C et al. Comparative bioavailability study of capecitabine tablets of 500 mg in metastatic breast cancer and colorectal cancer patients under fed condition. Clin. Res. Regul. Aff. 29, 72–76 (2012).
21. Haller DG et al. Potential regional differences for the tolerability profiles of fluoropyrimidines. J. Clin. Oncol. 26, 2118–2123 (2008). [PubMed: 18445840]
22. Jennings BA et al. Evaluating Predictive Pharmacogenetic Signatures of Adverse Events in Colorectal Cancer Patients Treated with Fluoropyrimidines. PLoS ONE vol. 8 e78053 (2013). [PubMed: 24167597]
23. Saif MW, Syrigos K, Mehra R, Mattison LK & Diasio RB DIHYDROPYRIMIDINE DEHYDROGENASE DEFICIENCY (DPD) IN GI MALIGNANCIES: EXPERIENCE OF 4-YEARS. Pak. J. Med. Sci. Q. 23, 832–839 (2007).
24. Leonard R, Hennessy BT, Blum JL & O'Shaughnessy J Dose-adjusting capecitabine minimizes adverse effects while maintaining efficacy: a retrospective review of capecitabine for metastatic breast cancer. Clin. Breast Cancer 11, 349–356 (2011). [PubMed: 21856245]
25. Horowitz J, Saukkonen JJ & Chargaff E Effects of fluoropyrimidines on the synthesis of bacterial proteins and nucleic acids. J. Biol. Chem. 235, 3266–3272 (1960). [PubMed: 13715937]
26. Bloch A & Hutchison DJ A MECHANISM OF RESISTANCE TO FLUOROPYRIMIDINES. Cancer Res. 24, 433–439 (1964). [PubMed: 14147816]
27. Stringer AM et al. Gastrointestinal microflora and mucins may play a critical role in the development of 5-Fluorouracil-induced gastrointestinal mucositis. Exp. Biol. Med. 234, 430–441 (2009).
28. Von Bültzingslöwen I, Adlerberth I, Wold AE, Dahlén G & Jontell M Oral and intestinal microflora in 5-fluorouracil treated rats, translocation to cervical and mesenteric lymph nodes and effects of probiotic bacteria. Oral Microbiol. Immunol. 18, 278–284 (2003). [PubMed: 12930518]
29. Stringer AM et al. Chemotherapy-induced diarrhea is associated with changes in the luminal environment in the DA rat. Exp. Biol. Med. 232, 96–106 (2007).
30. Zwielehner J et al. Changes in human fecal microbiota due to chemotherapy analyzed by TaqMan-PCR, 454 sequencing and PCR-DGGE fingerprinting. PLoS One 6, e28654 (2011). [PubMed: 22194876]
31. van Vliet MJ et al. Chemotherapy treatment in pediatric patients with acute myeloid leukemia receiving antimicrobial prophylaxis leads to a relative increase of colonization with potentially pathogenic bacteria in the gut. Clin. Infect. Dis. 49, 262–270 (2009). [PubMed: 19514856]
32. Stringer AM et al. Biomarkers of chemotherapy-induced diarrhoea: a clinical study of intestinal microbiome alterations, inflammation and circulating matrix metalloproteinases. Support. Care Cancer 21, 1843–1852 (2013). [PubMed: 23397098]

33. Montassier E et al. Chemotherapy-driven dysbiosis in the intestinal microbiome. Aliment. Pharmacol. Ther. 42, 515–528 (2015). [PubMed: 26147207]
34. Scott TA et al. Host-Microbe Co-metabolism Dictates Cancer Drug Efficacy in C. elegans. Cell vol. 169 442–456.e18 (2017). [PubMed: 28431245]
35. García-González AP et al. Bacterial Metabolism Affects the C. elegans Response to Cancer Chemotherapeutics. Cell 169, 431–441.e8 (2017). [PubMed: 28431244]
36. Rosener B et al. Evolved bacterial resistance against fluoropyrimidines can lower chemotherapy impact in the Caenorhabditis elegans host. Elife 9, (2020).
37. Grothey A et al. Duration of Adjuvant Chemotherapy for Stage III Colon Cancer. N. Engl. J. Med. 378, 1177–1188 (2018). [PubMed: 29590544]
38. Walko CM & Lindley C Capecitabine: a review. Clin. Ther. 27, 23–44 (2005). [PubMed: 15763604]
39. Zhang L, Zhang YD, Strong JM, Reynolds KS & Huang S-M A regulatory viewpoint on transporter-based drug interactions. Xenobiotica 38, 709–724 (2008). [PubMed: 18668428]
40. Longley DB, Harkin DP & Johnston PG 5-fluorouracil: mechanisms of action and clinical strategies. Nat. Rev. Cancer 3, 330–338 (2003). [PubMed: 12724731]
41. Islam Z et al. Bacterial versus human thymidylate synthase: Kinetics and functionality. PLoS One 13, e0196506 (2018). [PubMed: 29715278]
42. Pinedo HM & Peters GF Fluorouracil: biochemistry and pharmacology. J. Clin. Oncol. 6, 1653–1664 (1988). [PubMed: 3049954]
43. O'Donovan GA & Neuhard J Pyrimidine metabolism in microorganisms. Bacteriol. Rev. 34, 278–343 (1970). [PubMed: 4919542]
44. Lewis K Platforms for antibiotic discovery. Nat. Rev. Drug Discov. 12, 371–387 (2013). [PubMed: 23629505]
45. Spanogiannopoulos P, Waglechner N, Koteva K & Wright GD A rifamycin inactivating phosphotransferase family shared by environmental and pathogenic bacteria. Proc. Natl. Acad. Sci. U. S. A. 111, 7102–7107 (2014). [PubMed: 24778229]
46. Lunenburg CATC et al. Prospective DPYD genotyping to reduce the risk of fluoropyrimidine-induced severe toxicity: Ready for prime time. Eur. J. Cancer 54, 40–48 (2016). [PubMed: 26716401]
47. Wei X, McLeod HL, McMurrough J, Gonzalez FJ & Fernandez-Salguero P Molecular basis of the human dihydropyrimidine dehydrogenase deficiency and 5-fluorouracil toxicity. J. Clin. Invest. 98, 610–615 (1996). [PubMed: 8698850]
48. Vreken P, Van Kuilenburg AB, Meinsma R & van Gennip AH Dihydropyrimidine dehydrogenase (DPD) deficiency: identification and expression of missense mutations C29R, R886H and R235W. Hum. Genet. 101, 333–338 (1997). [PubMed: 9439663]
49. Hidese R, Mihara H, Kurihara T & Esaki N Escherichia coli dihydropyrimidine dehydrogenase is a novel NAD-dependent heterotetramer essential for the production of 5,6-dihydrouracil. J. Bacteriol. 193, 989–993 (2011). [PubMed: 21169495]
50. Smith AE & Yamada EW Dihydrouracil dehydrogenase of rat liver. J. Biol. Chem. 246, 3610–3617 (1971). [PubMed: 4397104]
51. Podschun B, Jahnke K, Schnackerz KD & Cook PF Acid base catalytic mechanism of the dihydropyrimidine dehydrogenase from pH studies. J. Biol. Chem. 268, 3407–3413 (1993). [PubMed: 8429016]
52. Porter DJ Dehalogenating and NADPH-modifying activities of dihydropyrimidine dehydrogenase. J. Biol. Chem. 269, 24177–24182 (1994). [PubMed: 7929074]
53. Porter DJ, Harrington JA, Almond MR, Lowen GT & Spector T (R)-5-fluoro-5,6-dihydrouracil: kinetics of oxidation by dihydropyrimidine dehydrogenase and hydrolysis by dihydropyrimidine aminohydrolase. Biochem. Pharmacol. 48, 775–779 (1994). [PubMed: 8080451]
54. Beaupre BA, Forouzesh DC & Moran GR Transient-State Analysis of Porcine Dihydropyrimidine Dehydrogenase Reveals Reductive Activation by NADPH. Biochemistry 59, 2419–2431 (2020). [PubMed: 32516529]

55. Lam KN et al. Phage-delivered CRISPR-Cas9 for strain-specific depletion and genomic deletions in the gut microbiome. Cell Rep. 37, 109930 (2021). [PubMed: 34731631]
56. Barba M, Dutoit R, Legrain C & Labedan B Identifying reaction modules in metabolic pathways: bioinformatic deduction and experimental validation of a new putative route in purine catabolism. BMC Syst. Biol. 7, 99 (2013). [PubMed: 24093154]
57. Bradley PH, Nayfach S & Pollard KS Phylogeny-corrected identification of microbial gene families relevant to human gut colonization. PLoS Comput. Biol. 14, e1006242 (2018). [PubMed: 30091981]
58. Nayfach S, Shi ZJ, Seshadri R, Pollard KS & Kyrpides NC New insights from uncultivated genomes of the global human gut microbiome. Nature 568, 505–510 (2019). [PubMed: 30867587]
59. Zeller G et al. Potential of fecal microbiota for early-stage detection of colorectal cancer. Mol. Syst. Biol. 10, 766 (2014). [PubMed: 25432777]
60. Ai D et al. Identifying Gut Microbiota Associated With Colorectal Cancer Using a Zero-Inflated Lognormal Model. Front. Microbiol. 10, 826 (2019). [PubMed: 31068913]
61. Geller LT et al. Potential role of intratumor bacteria in mediating tumor resistance to the chemotherapeutic drug gemcitabine. Science 357, 1156–1160 (2017). [PubMed: 28912244]
62. Maier L et al. Extensive impact of non-antibiotic drugs on human gut bacteria. Nature 555, 623–628 (2018). [PubMed: 29555994]
63. Guo C-J et al. Depletion of microbiome-derived molecules in the host using Clostridium genetics. Science 366, (2019).
64. Nayak RR & Turnbaugh PJ Mirror, mirror on the wall: which microbiomes will help heal them all? BMC Med. 14, 72 (2016). [PubMed: 27146150]
65. Nayfach S, Rodriguez-Mueller B, Garud N & Pollard KS An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. Genome Res. 26, 1612–1625 (2016). [PubMed: 27803195]
66. Langmead B & Salzberg SL Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012). [PubMed: 22388286]
67. Anders S, Pyl PT & Huber W HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169 (2015). [PubMed: 25260700]
68. Anders S & Huber W Differential expression analysis for sequence count data. Genome Biol. 11, R106 (2010). [PubMed: 20979621]
69. Bolger AM, Lohse M & Usadel B Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120 (2014). [PubMed: 24695404]
70. Seemann T snippy. (Github).
71. Titz B, Häuser R, Engelbrecher A & Uetz P The Escherichia coli protein YjjG is a house-cleaning nucleotidase in vivo. FEMS Microbiol. Lett. 270, 49–57 (2007). [PubMed: 17286574]
72. Sharan SK, Thomason LC, Kuznetsov SG & Court DL Recombineering: a homologous recombination-based method of genetic engineering. Nat. Protoc. 4, 206–223 (2009). [PubMed: 19180090]
73. Jensen SI, Lennen RM, Herrgård MJ & Nielsen AT Seven gene deletions in seven days: Fast generation of Escherichia coli strains tolerant to acetate and osmotic stress. Sci. Rep. 5, 17874 (2015). [PubMed: 26643270]
74. Baba T et al. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. 2, (2006).
75. Datsenko KA & Wanner BL One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc. Natl. Acad. Sci. U. S. A. 97, 6640–6645 (2000). [PubMed: 10829079]
76. Cox G et al. A Common Platform for Antibiotic Dereplication and Adjuvant Discovery. Cell Chem Biol 24, 98–109 (2017). [PubMed: 28017602]
77. Funatsu G & Wittmann HG Ribosomal proteins. 33. Location of amino-acid replacements in protein S12 isolated from Escherichia coli mutants resistant to streptomycin. J. Mol. Biol. 68, 547–550 (1972). [PubMed: 4560854]

78. Cicchillo RM et al. Lipoyl Synthase Requires Two Equivalents ofS-Adenosyl-l-methionine To Synthesize One Equivalent of Lipoic Acid†. Biochemistry vol. 43 6378–6386 (2004). [PubMed: 15157071]
79. Beaupre BA, Roman JV & Moran GR An improved method for the expression and purification of porcine dihydropyrimidine dehydrogenase. Protein Expr. Purif. 171, 105610 (2020). [PubMed: 32088324]
80. Myhal ML, Laux DC & Cohen PS Relative colonizing abilities of human fecal and K 12 strains ofEscherichia coli in the large intestines of streptomycin-treated mice. Eur. J. Clin. Microbiol. (1982).
81. Measuring treatment response in Patient Derived Xenograft (PDX) models at The Jackson Laboratory. The Jackson Laboratory http://tumor.informatics.jax.org/mtbwi/live/www/html/SOCHelp.html (2017).
82. Gohl DM et al. Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. Nat. Biotechnol. 34, 942–949 (2016). [PubMed: 27454739]
83. McMurdie PJ & Holmes S phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8, e61217 (2013). [PubMed: 23630581]
84. Haft DH et al. RefSeq: an update on prokaryotic genome annotation and curation. Nucleic Acids Res. 46, D851–D860 (2018). [PubMed: 29112715]
85. Notredame C, Higgins DG & Heringa J T-Coffee: A novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. 302, 205–217 (2000). [PubMed: 10964570]
86. Madeira F et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. 47, W636–W641 (2019). [PubMed: 30976793]
87. Capella-Gutiérrez S, Silla-Martínez JM & Gabaldón T trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973 (2009). [PubMed: 19505945]
88. Finn RD, Clements J & Eddy SR HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39, W29–37 (2011). [PubMed: 21593126]
89. Wu D, Jospin G & Eisen JA Systematic Identification of Gene Families for Use as ‘Markers’ for Phylogenetic and Phylogeny-Driven Ecological Studies of Bacteria and Archaea and Their Major Subgroups. PLoS One 8, e77033 (2013). [PubMed: 24146954]
90. Nguyen L-T, Schmidt HA, von Haeseler A & Minh BQ IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274 (2015). [PubMed: 25371430]
91. Sievers F et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol 7, (2011).
92. Madden T The BLAST Sequence Analysis Tool. (National Center for Biotechnology Information (US), 2003).
93. Rognes T, Flouri T, Nichols B, Quince C & Mahé F VSEARCH: a versatile open source tool for metagenomics. PeerJ 4, e2584 (2016). [PubMed: 27781170]
94. Nayfach S, Fischbach MA & Pollard KS MetaQuery: a web server for rapid annotation and quantitative analysis of specific genes in the human gut microbiome. Bioinformatics 31, 3368–3370 (2015). [PubMed: 26104745]
95. Parks DH et al. A complete domain-to-species taxonomy for Bacteria and Archaea. Nat. Biotechnol. 38, 1079–1086 (2020). [PubMed: 32341564]
96. Guindon S, Delsuc F, Dufayard J-F & Gascuel O Estimating maximum likelihood phylogenies with PhyML. Methods Mol. Biol. 537, 113–137 (2009). [PubMed: 19378142]
97. Edgar RC MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797 (2004). [PubMed: 15034147]
98. Yu G, Wang L-G, Han Y & He Q-Y clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 16, 284–287 (2012). [PubMed: 22455463]
99. Waterhouse A et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46, W296–W303 (2018). [PubMed: 29788355]

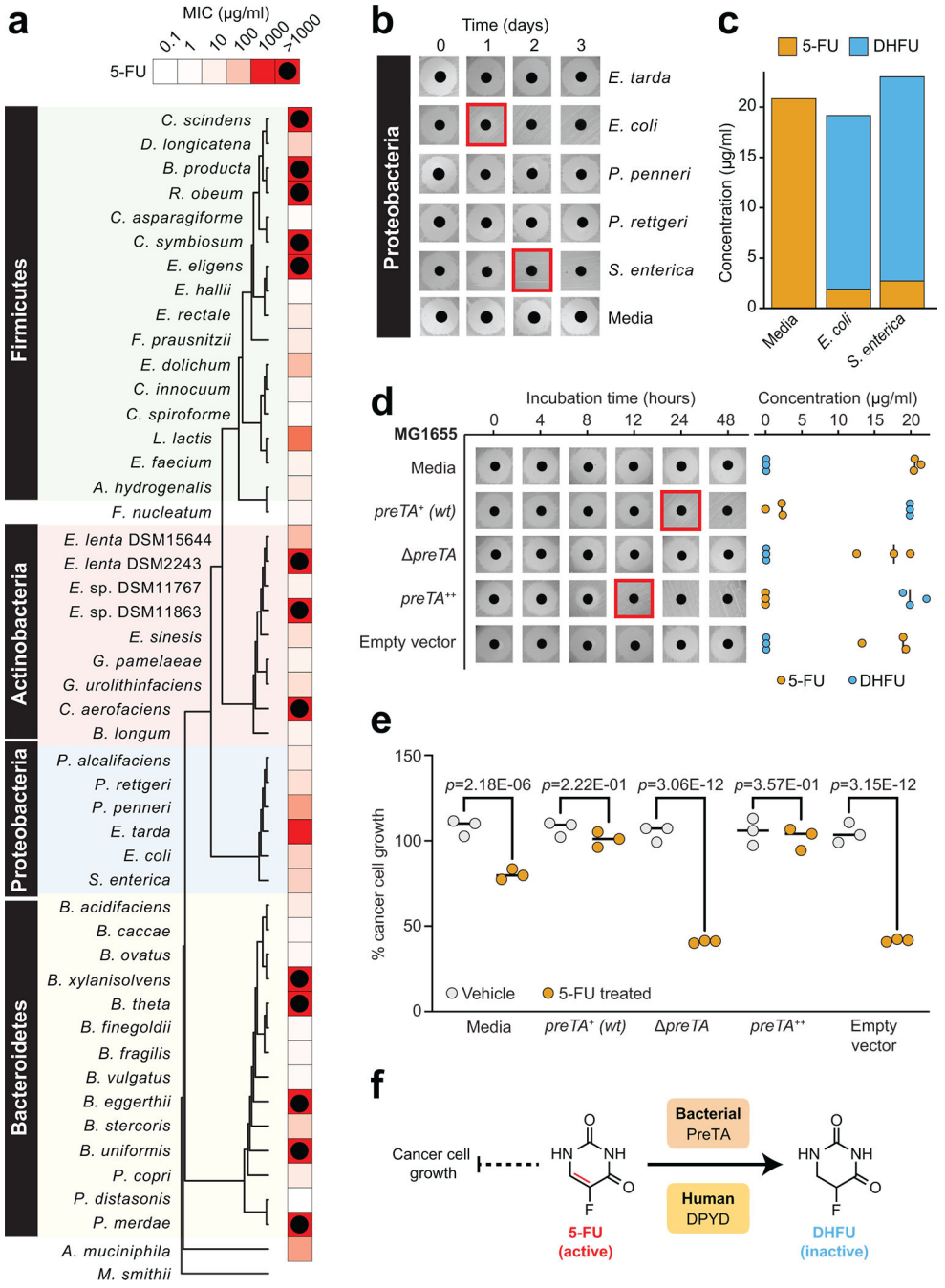


Fig. 1|. The anti-cancer drug 5-fluorouracil (5-FU) inhibits bacterial growth and is inactivated by the *preTA* operon.

(a) Heatmap representing the 5-FU minimal inhibitory concentration (MIC, <10% growth relative to vehicle controls; 47 human gut bacterial strains; 2 replicates/strain/concentration; see Supplementary Table 1). The phylogenetic tree was built using PhyML[96], based on a MUSCLE[97] alignment of full-length 16S rRNA genes. **(b)** We screened 23 strains with MIC≥62.5 μg/ml for drug inactivation using a disk diffusion assay (BHI$^+$ media; 20 μg/ml 5-FU); only Proteobacteria are shown. Red squares indicate decreased zones of inhibition. **(c)** LC-QTOF/MS detection of the predicted 5-FU metabolite DHFU after 48 hours of

anaerobic incubation [*E. coli* MG1655, *S. enterica* LT2 (DSM17058), 20 μg/ml 5-FU]. **(d)** Wild-type (*preTA*$^{+}$), deletion (*ΔpreTA*), complemented (*preTA*$^{++}$), and empty vector *E. coli* MG1655 strains were assayed for residual 5-FU using disk diffusion (0–48 hours incubation) and LC-QTOF/MS (48 hours). Complementation and empty vector were on the *ΔpreTA* background. Black lines indicate the median value (n=3 biological replicates). **(e)** The *E. coli* MG1655 strains shown in panel d were incubated for 72 hours with 5-FU (20 μg/ml) or vehicle (DMSO, 0.05%), conditioned media was added to the CRC cell line HCT-116, and cell proliferation was quantified using the MTT assay. Lines represent medians (n=3 biological replicates/strain/condition). *p*-values, one-way ANOVA with Holm-Sidak's multiple comparisons test. **(f)** We propose that bacterial PreTA contributes to the elimination of 5-FU similar to hepatic expression of dihydropyrimidine dehydrogenase (DPYD).

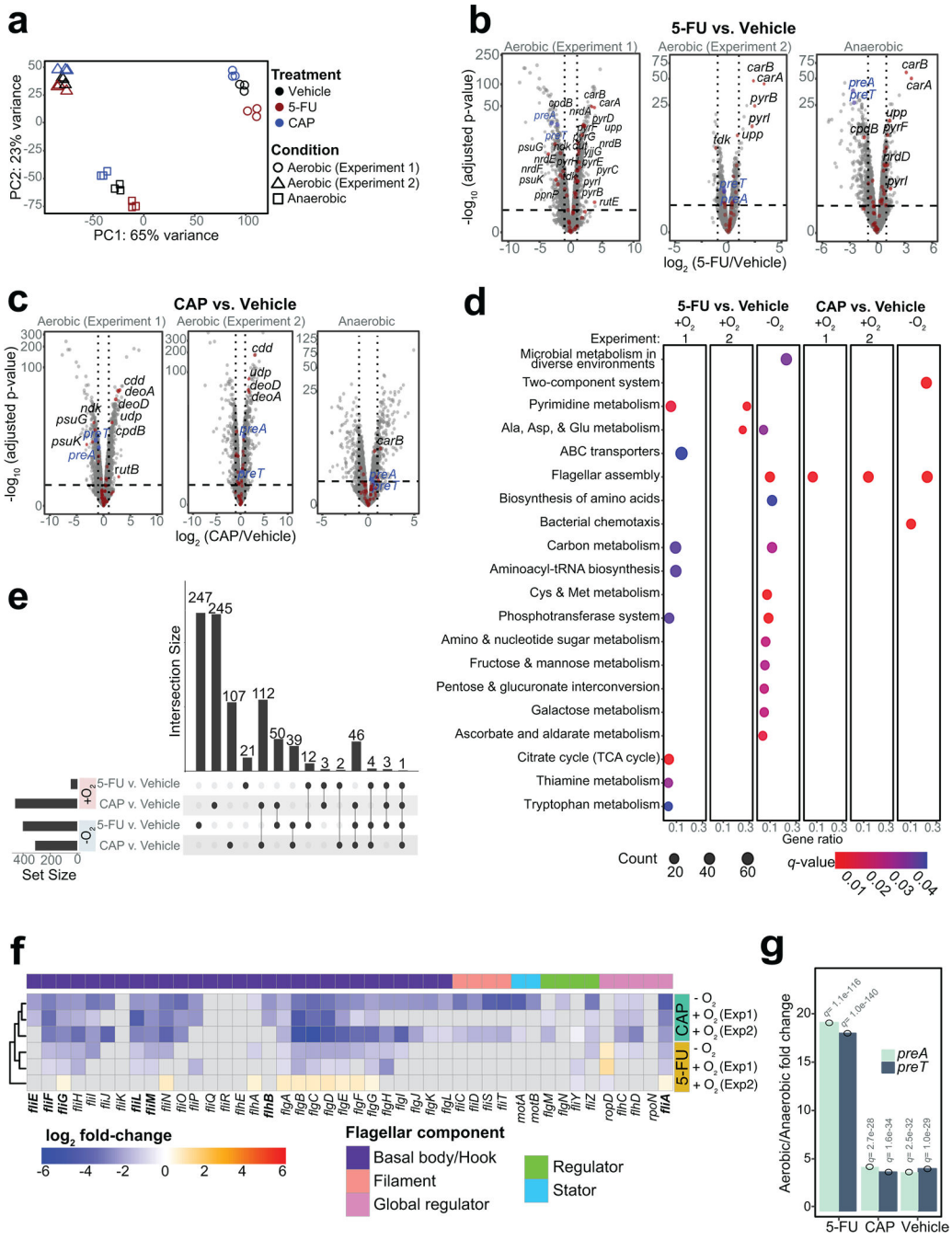


Fig. 2|. Shared and unique transcriptional response to the related fluoropyrimidines 5-fluorouracil (5-FU) and capecitabine (CAP) during aerobic and anaerobic growth.

(a) Principal component analysis of *E. coli* MG1655 transcriptomes under different treatments in aerobic and anaerobic conditions (n=3 biological replicates; 2 independent experiments were conducted for aerobic growth). **(b,c)** Volcano plots of differentially expressed genes (DEGs) [False Discovery Rate (FDR)<0.1, $|\log_2$ fold-change$|>1$, significance limits are marked with dash lines; Wald test with Benjamini-Hochberg correction]. DEGs involved in pyrimidine metabolism are labeled by gene name and marked with a red dot. *preT* and *preA* are labeled in blue. **(d)** Pathway enrichment analysis

following 5-FU and CAP exposure under aerobic and anaerobic conditions. The Kyoto Encyclopedia of Genes and Genomes (KEGG) database was used to test for pathway enrichment (*q*-value<0.1, Hypergeometric test with Benjamini-Hochberg correction from enrichKEGG function in clusterProfiler[98]). **(e)** Upset plot comparing the fluoropyrimidine responsive DEGs under different growth conditions (Supplementary Table 5). The numbers on top of the vertical bars represent the number of DEGs unique to a single comparison (single dot) or shared among multiple comparisons (connected dots). The set size refers to a total number of DEGs in a single comparison (DEGs defined as FDR<0.1, $|\log_2$ fold-change$|$ >1; Wald test with Benjamini-Hochberg correction). **(f)** Heatmap of transcriptional changes of 44 genes involved in flagellar biosynthesis in response to fluoropyrimidine treatments under anaerobic and anaerobic conditions. Gray boxes indicate non-significant changes (FDR>0.1). Bolded genes indicate DEGs that are down-regulated in 5-FU evolved strains reported in literature[36]. **(g)** *preT* and *preA* transcript levels are significantly lower during anaerobic growth relative to aerobic growth irrespective of the presence of fluoropyrimidines (*q*-value<0.1, DESeq).

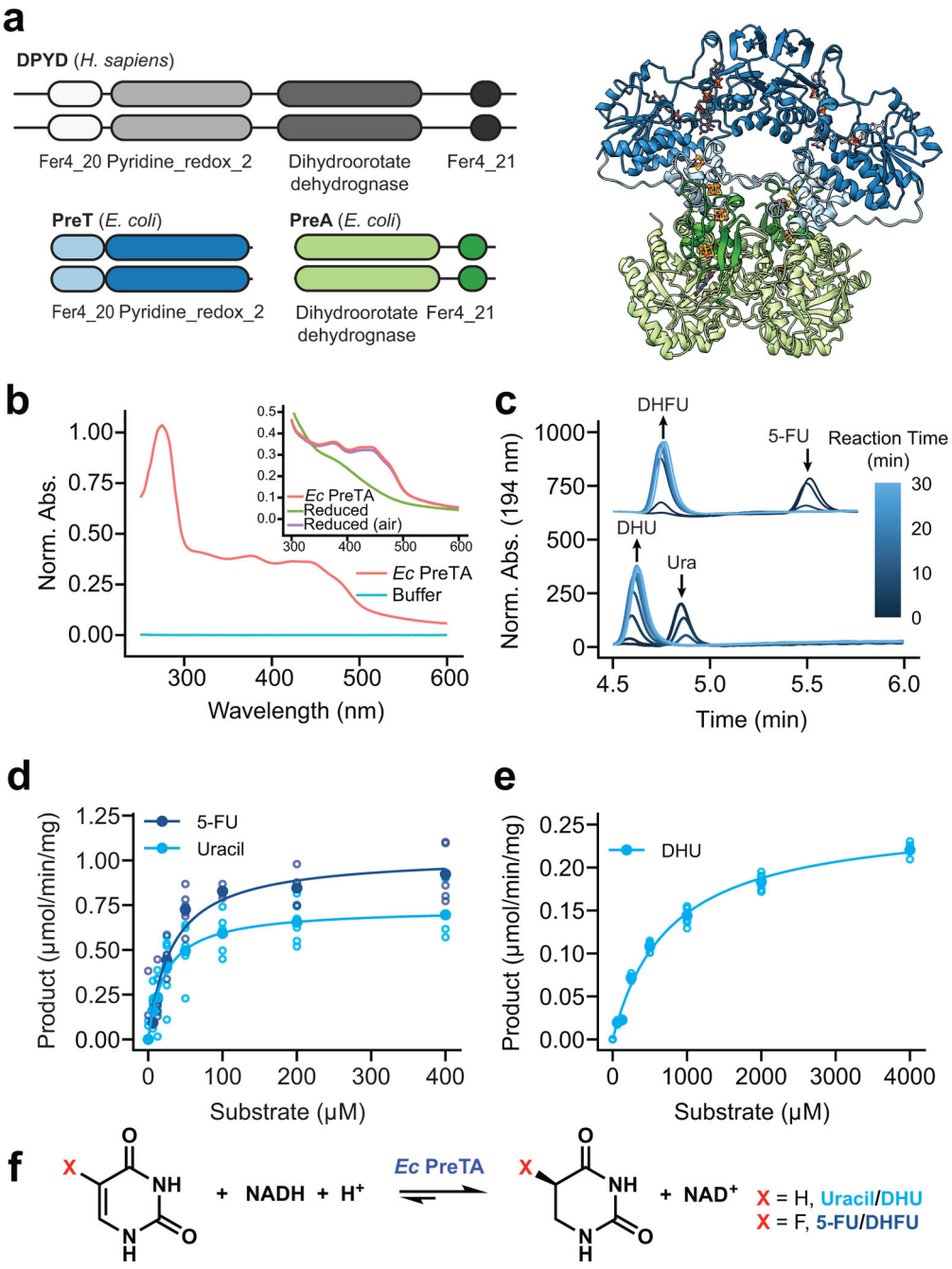


Fig. 3|. *E. coli* PreTA more rapidly catalyzes the reduction of pyrimidines.

(a) A cartoon schematic shows the domain architecture of the homodimeric *H. sapiens* DPYD and heterotetrameric *E. coli* PreTA, as well as a *E. coli* PreTA homology model (Swiss-Model[99]). Protein backbone colors are indicated by the schematic, cofactors are colored by heteroatom. **(b)** UV-visible absorption spectra indicate heterologously expressed *E. coli* PreTA has characteristic peaks due to 4Fe-4S and flavin cofactors. Inset shows reduction of flavin cofactors and reoxidation upon exposure to air. **(c)** *In vitro* enzyme assays analyzed by high pressure liquid chromatography (HPLC) show conversion of uracil and 5-FU to their reduced forms over time. **(d-e)** Michaelis-Menten analysis of NADH cofactor

consumption and accumulation *in vitro* enzyme assays in the reductive **(d)** and oxidative **(e)** directions display a marked difference in the ability for *E. coli* PreTA to catalyze the oxidative reaction. Solid points represent the mean of the hollow circles, n=6 replicates over 2 different enzyme preparations. **(f)** Schematic of the chemical reaction catalyzed by *E. coli* PreTA.

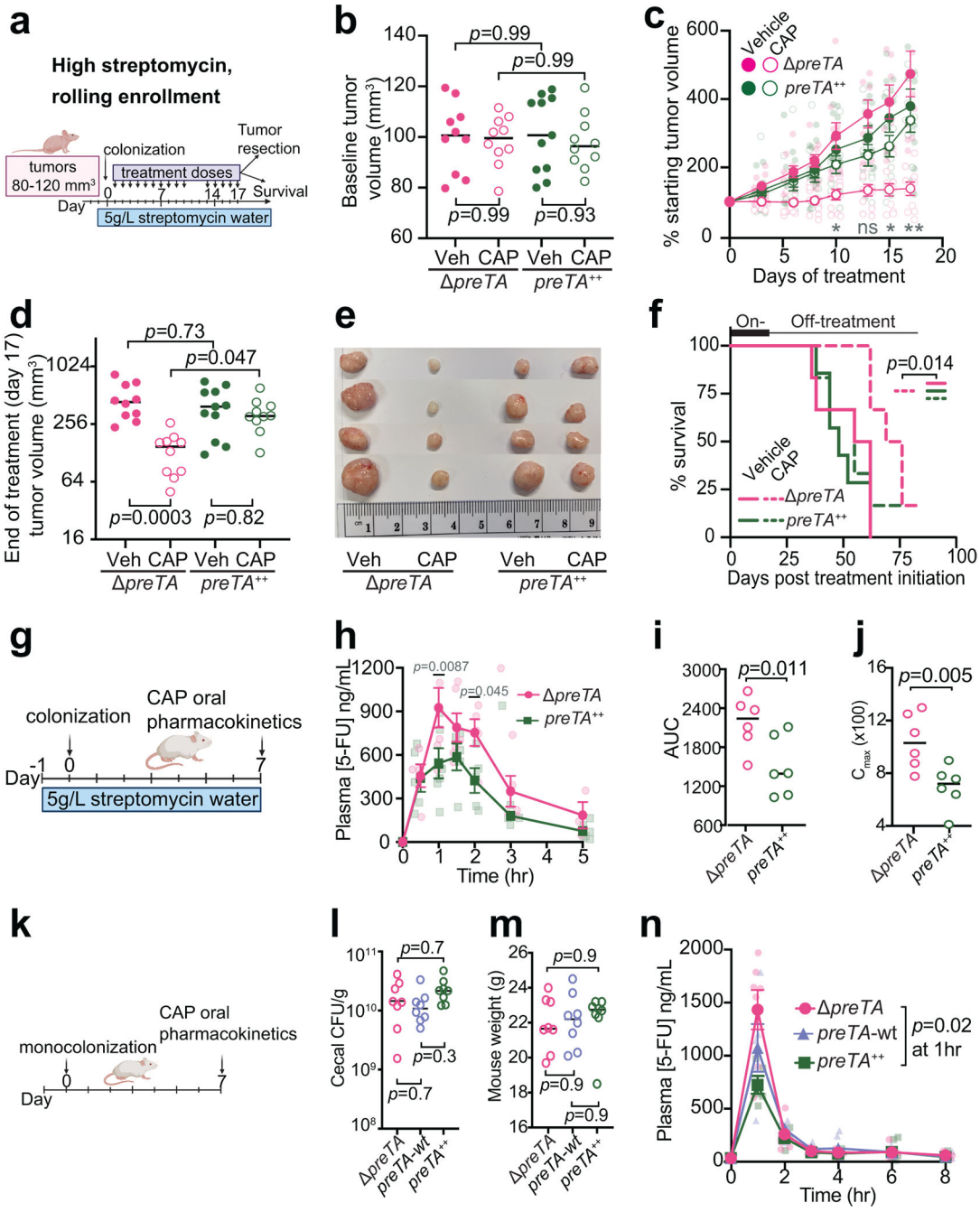


Fig. 4|. PreTA interferes with capecitabine (CAP) efficacy in mice.

(a) Schematic of xenograft mice colonized with Δ*preTA* or *preTA*$^{++}$ *E. coli* and treated with capecitabine (CAP) or vehicle (Veh). **(b, d)** Tumor volumes **(b)** pre- and **(d)** post-treatment (n=11 mice *preTA*$^{++}$-Veh group; n=10 mice/group remainder; lines represent medians; 2-way ANOVA with Tukey's correction). **(c)** Percentage of starting tumor volumes across time (n=11 mice *preTA*$^{++}$-Veh; n=10 mice/group remainder; opaque dots and lines represent mean±SEM; 2-way ANOVA with Tukey's correction; adjusted *p*-values=0.046 day 10, 0.053 day 13, 0.018 day 15, and 0.0012 day 17 for Δ*preTA*-CAP vs. *preTA*$^{++}$-CAP comparison). **(e)** Resected tumors post-treatment on day 17 (n=4 mice/group). **(f)** Percentage

of mice reaching the humane endpoint (n=7 mice *preTA*$^{++}$-Veh; n=6 mice/group remainder; log-rank Mantel-Cox test comparing Δ*preTA*-CAP to all other groups). Two mice were censored as they did not reach the endpoint when the experiment ended on day 83. **(g)** Pharmacokinetics schematic in CONV-R mice colonized with Δ*preTA* or *preTA*$^{++}$ *E. coli* and administered with 1100 mg/kg CAP. **(h)** LC-MS/MS quantification of plasma 5-FU (n=6 mice/group; opaque dots and lines represent mean±SEM; mixed-effects analysis with Sidak's correction). **(i, j)** Pharmacokinetics parameters such as **(i)** area under curve (AUC) and **(j)** peak plasma concentrations (C_{max}) for plasma 5-FU (n=6 mice/group; lines are medians; Welch's one-tailed *t* test). **(k)** Pharmacokinetics schematic in gnotobiotic mice mono-colonized with Δ*preTA*, *preTA*$^{++}$, or wild-type *E. coli* followed by oral administration of 500 mg/kg CAP. **(l, m)** Potential confounders like **(l)** colonization levels assessed by colony-forming units (CFU) per gram stool and **(m)** mouse weight were measured (n=8 mice/group; lines are medians; one-way ANOVA with Tukey's correction). **(n)** LC-MS/MS quantification of plasma 5-FU (n=8 mice/group; opaque dots and lines represent mean±SEM; mixed-effects analysis with Tukey's correction).

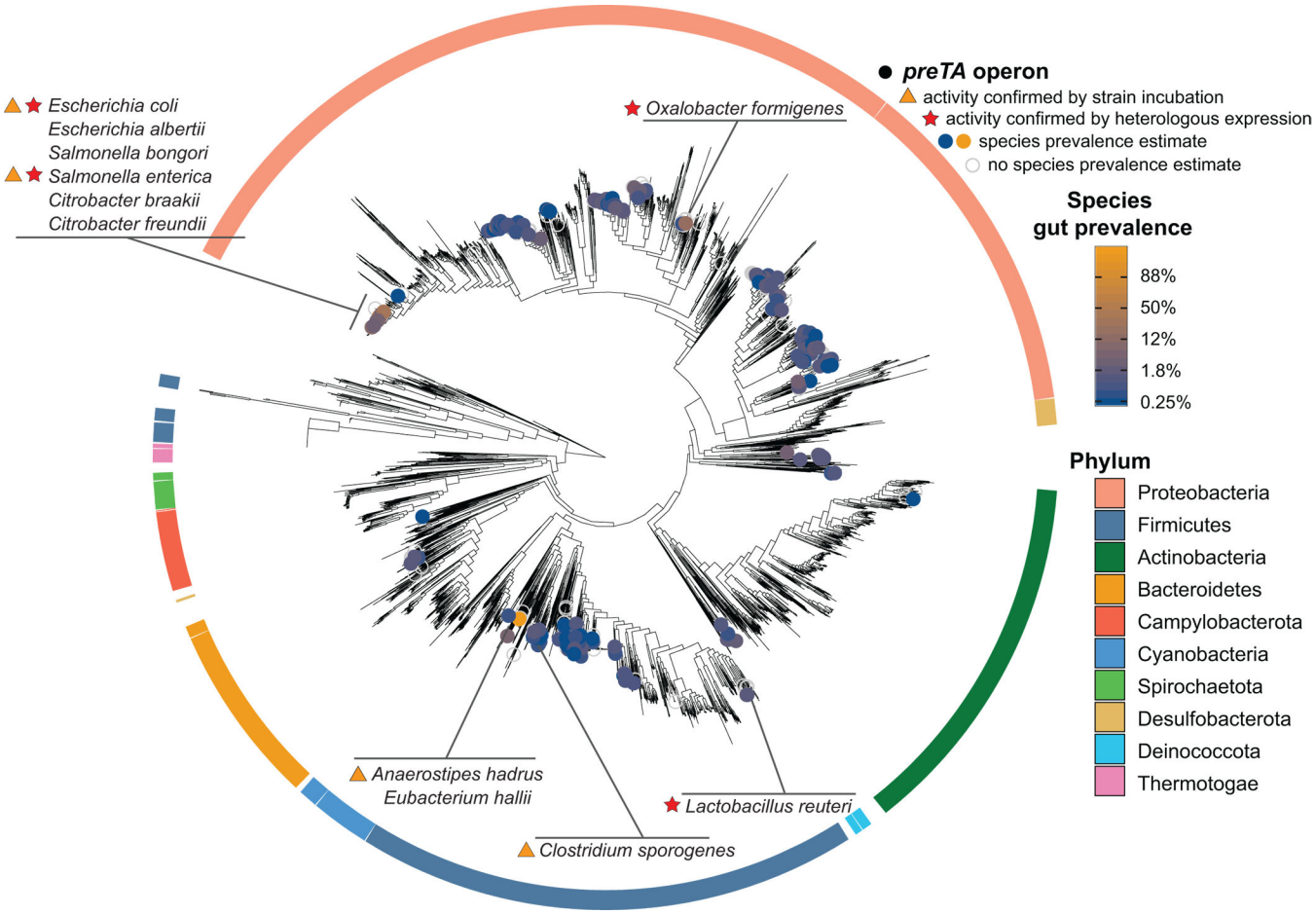


Fig. 5|. Functional orthologs of the *preTA* operon are widespread in human gut bacterial strains from the Firmicutes and Proteobacteria phyla.

Distribution of bioinformatically-identified *preTA* operons across RefSeq bacterial isolate genomes. A phylogenetic tree of these genomes made using a concatenated alignment of single-copy marker genes is shown. Bacterial species identified as carriers of putative *preTA* operons are identified with colored circles, where the color of the circle corresponds to prevalence levels from human gut microbiomes (blue: low prevalence; orange: high prevalence; unfilled grey: no prevalence estimate). Phylum-level annotations are shown as colored ring segments surrounding the tree for the ten phyla with the most species in Refseq. Specific taxa of interest are highlighted in call-out boxes. Red stars indicate *preTA* operons that have been validated to inactive 5-FU by heterologous expression from *E. coli* Δ*preTA* (see Extended Data Fig. 8a). Orange triangles indicate *preTA*-positive bacterial species (or close relatives) for which we have confirmed 5-FU inactivation *in vitro* (see Extended Data Fig. 8b). A list of *preTA*-positive bacteria can be found in Supplementary Table 8.

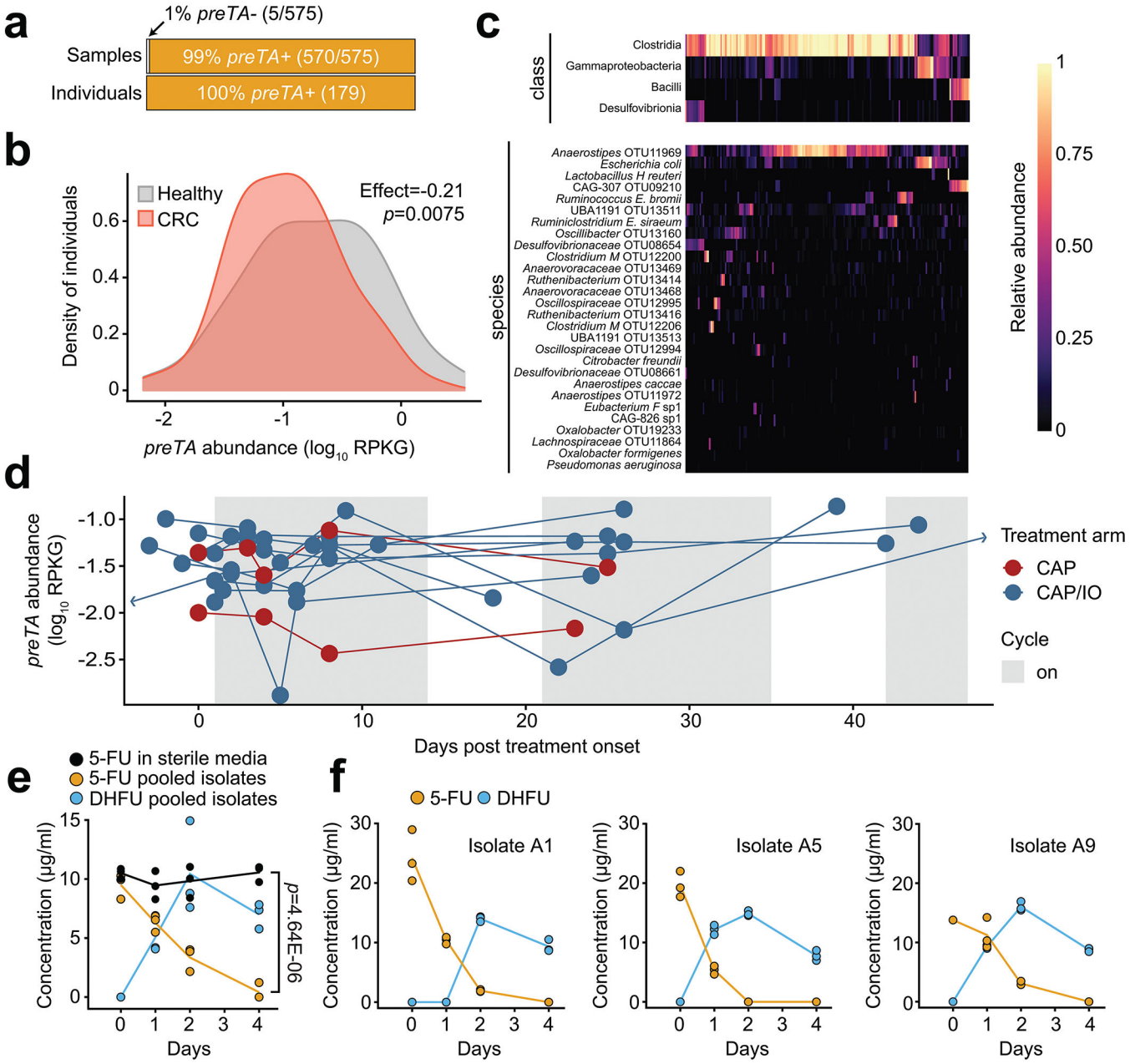


Fig. 6|. The *preTA* operon is prevalent and functional in the gut microbiomes of colorectal cancer (CRC) patients and controls, exhibiting marked inter-individual and temporal variation in abundance.

(a) Number of samples and individuals where *preTA* was detected (orange) or not detected (light gray). **(b)** Variation in *preTA* abundance, as $\log_{10}$ reads per kilobase of genome equivalents (RPKG), across CRC patients and healthy controls where *preTA* was detected (n=179 individuals[59]). **(c)** Fraction of total reads mapping to *preTA* in CRC and control samples where it was detected (x-axis) whose best hit was in a given phylogenetic class (top) or species (bottom). Taxa are ordered by decreasing mean across all samples; samples are ordered based on hierarchical clustering using Bray-Curtis dissimilarity and complete linkage. **(d)** Abundance, as $\log_{10}$ RPKG, of the *preTA* operon in the gut microbiome prior to

and during treatment with the oral fluoropyrimidine CAP (red) or a combination of CAP and immunotherapy (blue). Lines connect measurements for the same patient. One zero RPKG value was replaced with half the minimum non-zero value prior to taking the logarithm. 3/11 patients varied over an order of magnitude during treatment (Extended Data Fig. 9). The first day of treatment is defined as day 1. Two samples collected on days −16 and 66 were censored for display purposes. **(e)** Bacterial isolates from a representative CRC patient stool sample in the GO Study (patient 1, Supplementary Table 12) were isolated on McConkey agar, pooled, and incubated in BHI^{+} with 5-FU for 4 days and 5-FU/DHFU were quantified by LC-QTOF/MS. *p*-value, two-way ANOVA contrasting 5-FU levels in sterile and inoculated media. **(f)** Three bacterial isolates from the same CRC patient stool sample were incubated as described in panel e (n=3 biological replicates). Lines in e,f represent means.