Title
Impact of the gut microbiome on chemotherapy outcomes

Author
Kyaw, Than Soe

Publication Date
2022

Peer reviewed|Thesis/dissertation

Impact of the gut microbiome on chemotherapy outcomes

by
Than Kyaw

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biomedical Sciences

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Goga, Andrei
2582F8BD11124B6...

Goga, Andrei
Chair

DocuSigned by:
Deanna Kroetz
DocuSigned by:AA...

Deanna Kroetz

DocuSigned by:
Shaeri Mukherjee
DocuSigned by:471...

Shaeri Mukherjee

DocuSigned by:
Peter J. Turnbaugh
8C4E22B44AE0441...

Peter J. Turnbaugh

Committee Members

Dedication

To my family and fellow villagers in Leiktho, Myanmar

Acknowledgements

I am grateful to my PhD advisor, Dr. Peter J. Turnbaugh, for being the best mentor I could have ever asked for throughout my thesis research. Thank you for offering me a home in your wonderful lab and being a shining bright light throughout my PhD journey. Your unwavering support and encouragement have made me feel like a curious child exploring the natural wonders of the microcosmos in a safe playground where I could confidently take risks and push boundaries. After all, “play is the highest form of research” according to Albert Einstein.

As I explored the microscopic unseen world, I met a host of characters who helped, supported, and cheered me on along the way. Thank you to the past and present members of the Turnbaugh lab for joining along and supporting me throughout. Thanks to Ernesto Valencia and Daryll Gempis for going above and beyond to create a welcoming environment. A special shoutout to Dr. Jordan Bisanz who patiently mentored me and motivated me to be a better scientist; I will forever be grateful. I would also like to thank the Medical Scientist Training Program staff for their support; the gnotobiotic team led by Dr. Jessie Turnbaugh for their help in mouse experiments; my clinical collaborators for giving me an opportunity to conduct translational research; and my thesis committee – Drs. Andrei Goga, Shaeri Mukherjee, and Deanna Kroetz – for helping me make sense of the discoveries.

Some might need a village to raise a PhD, but I needed a village and a city to even *get to* graduate school. I would like to thank my parents and siblings for taking a giant leap of faith to leave the familiar comforts of our village in Leiktho, Myanmar, to immigrate to Los Angeles; without their sacrifice, I would, quite literally, not be here in the MD/PhD

program at UCSF. When I arrived in the US without knowing how to read, write, or speak English, it was a lonely and isolating experience. It was Mr. Tracy Kawasaki, my 8th-grade science teacher, who invested in me and taught me English communication skills during recess (he even gave me my first computer). Mr. K, I am forever grateful for your encouragement when I needed it the most. As I fumbled my way to UCSD, I was fortunate to meet and work with Dr. Douglas Bartlett who kindled my love for research and exploration of natural wonders. Thank you, Dr. Bartlett, for instilling the spirit of exploration and discovery (and for taking me on an expedition to the Mariana Trench!). He encouraged me to apply to the Julia Brown Undergraduate Research Fellowship, where I met my first research sponsor. Julia, your fellowship not only provided financial assistance but also gave me faith that I could be a scientist and solidified my decision to apply for the MD/PhD programs. Thank you for your kindness.

Just as the microcosmos are captivating, our world is also full of wonders that I have been fortunate to explore with friends and family. I'd like to thank Billy Morrison for his constant moral support and for pushing me to try the colorful things in life. I've enjoyed our travels to New Hampshire, Hawaii, and the Caribbean; I look forward to many more thrilling trips. I'd also like to thank my friends new and old – Jenny, Debs, Greg, Kevin, and Alina for sticking with me since high school and college; Xixi for our fun game nights and British Bakeoff drafts; Vikas, Irek, Nicole, Richard, and Jen for our whitewater rafting adventures and the trip to the Greek Islands; Lindsey, Paola, Veronica, and Janice for our dinners; and Keith and Cecilia for our yummy hotpot runs. Finally, I am grateful to Lexo, Princess of the Lexo Tribe, and Oro, Lord of the House of Atreus, for their wisdom and love of adventures.

Contributions

Chapter 1 contains work previously published in:

- Kyaw TS & Turnbaugh PT. Tiny Gatekeepers: Microbial Control of Host Drug Transporters. *Clinical Pharmacology & Therapeutics* 2022; 112(3): 443-445. doi: 10.1002/cpt.2647.

Chapter 2 contains work previously published in:

- Spanogiannopoulos P*, Kyaw TS*, Guthrie B, Bradley P, Lee JV, Melamed J, Malig Y, Lam K, Gempis D, Sandy M, Kidder W, Van Blarigan E, Atreya C, Venook A, Genora R, Goga A, Pollard K, Turnbaugh P. Host and gut bacteria share metabolic pathways for anti-cancer drug metabolism. *Nature Microbiology* 2022; 7(10):1605-1620. doi: 10.1038/s41564-022-01226-5. *Equal contribution.

Chapter 3 contains work from a manuscript under review:

- Kyaw TS, Sandy M, Trepka K, Goh J, Yu K, Dimassa V, Bess E, Bisanz J, Turnbaugh PJ. Human gut Actinobacteria boost drug absorption by secreting P-glycoprotein ATPase inhibitors. *bioRxiv* 2022. doi:10.1101/2022.10.13.512142.

Chapter 4 contains work from a manuscript in preparation:

- Kyaw TS, Upadhyay V, Tolstykh I, Loon K, Laffan A, Stanfield D, Gempis D, Kenfield S, Chan J, Piawah S, Atreya C, Ng K, Venook A, Pollard K, Kidder W, Turnbaugh PJ, Blarigan E. Variety of fruit and vegetables and alcohol intake predict gut microbial species and gene abundance in colorectal cancer survivors. 2022.

Abstract

Impact of the gut microbiome on chemotherapy outcomes

Than Soe Kyaw

Understanding the determinants of inter-individual variations in drug response is critical to improve efficacy and reduce toxicity of pharmaceuticals, including anti-cancer agents. One underappreciated factor is the gut microbiome, which we demonstrate influences drug metabolism and absorption through two independent, mechanistic studies. First, we show that 5-fluorouracil (5-FU), a standard of care therapy for colorectal cancer, is inactivated by a gut bacterial enzyme PreTA, which led to decreased 5-FU bioavailability and reduced efficacy in mouse tumor model. PreTA encoded by diverse gut bacteria was also prevalent at varying levels in colorectal cancer patients, which may contribute to observed variations in treatment outcomes. Second, we test whether the gut microbiome could influence absorption of oral drugs. Mice mono-colonized with a prevalent human gut Actinobacterium *Eggerthella lenta* had higher intestinal absorption of digoxin, a cardiac drug with a narrow therapeutic window used in heart failure. Using an *in vitro* system with human colonocytes, we find that *E. lenta* secretes small polar molecules that inhibits P-glycoprotein, a key mammalian efflux drug transporter involved in cancer resistance, by inhibiting its ATPase activity. Untargeted metabolomics revealed the *E. lenta* derived inhibitor to be an isoflavonoid, a class of natural metabolites with known P-gp inhibitors. Our findings emphasize the importance of gut microbiome for drug disposition, representing an untapped rich resource to enable precision medicine.

Table of Contents

List of Figures

List of Supplementary Figures

Chapter 1: Introduction

1.1 The role of the gut microbiome in colorectal cancer pathogenesis

Colorectal cancer (CRC) is a highly prevalent and life-changing diagnosis, affecting more than a million people each year[1]. In 2020 alone, an estimated 1,880,725 people were newly diagnosed with CRC and 915,880 people died from CRC worldwide[1]. Despite a common occurrence, more than 90% of the cases occurred spontaneously and the exact etiology is unknown. Epidemiological studies have identified many factors, such as old age, family history of CRC, inactive lifestyle, and a diet rich in red meat and processed meat, that increased the risk of CRC development[2]. One historically underappreciated factor is the impact of trillions of microorganisms residing in our gastrointestinal tract (the gut microbiota) and their aggregate genomes and metabolic activities (microbiome) on CRC development.

There is now a large and rigorous literature providing support for a causal role of the microbiome in CRC pathogenesis[3,4]. Compared to healthy individuals, CRC patients have altered microbiota, with lower bacterial diversity and expansion of pro-carcinogenic bacteria, including members of the *Fusobacterium nucleatum*[5], *Escherichia coli*[6], and *Bacteroides fragilis*[6,7] species. Broadly speaking, the mechanisms responsible include microbial effects on the host immune system[8] and the production of genotoxic metabolites[3,4] that could result in tumor formation. Given the ease and the non-invasive nature of sampling paired with the rapidly declining costs of sequencing, the gut microbiota may be used as an additional diagnostic marker of colorectal cancer.

1.2 The role of the gut microbiome in cancer treatment outcomes

Cytotoxic anti-cancer agents (chemotherapeutics) are critically important in the treatment of CRC, but they have unpredictable treatment responses including treatment-related morbidity and mortality. Factors such as patient demographics, comorbidities, tumor types, genetic polymorphisms, and the gut microbiome may contribute to interindividual variations in treatment response. This variation is important in explaining why one patient may tolerate high doses of cytotoxic agents and have their cancers successfully treated while another may not tolerate treatment and experience adverse effects from chemotherapy. Accumulating evidence supports that the gut bacteria can metabolize hundreds of host-targeted drugs, including anti-cancer agents, in some cases leading to clinically relevant changes in the treatment efficacy or toxicity[9–12].

A well-studied example is the microbial metabolism of irinotecan, which is one of the three first-line treatments for metastatic colorectal cancer[13,14]. This prodrug is converted to its active form (SN-38), which in turn is subsequently inactivated into its glucuronidated form (SN-38G) and excreted into the intestine[13,14]. Gut bacteria can de-glucuronate and reactivate SN-38G into its active form in the gut, contributing to its dose-limiting gastrointestinal toxicity[13,14]. Such seminal studies have established a conceptual framework to dissect the mechanisms by which human-associated microbial communities influence cancer therapies. However, the primary focus in the field has been limited to direct microbial metabolism with a narrow focus on reactions unique to bacteria. In this manuscript, I pushed the research at the interface of microbiome and pharmacology to two new frontiers: *(i)* the impact of drug-metabolizing pathways that are shared between the host and the microbiota on treatment efficacy (**Chapter 2**) and *(ii)*

the impact of the microbiota on absorption, rather than direct metabolism, of oral drugs (**Chapter 3**).

We focused on oral fluoropyrimidine capecitabine (CAP), another first-line agent for CRC treatment, given its extensive interaction with the microbiota[15,16]. CAP is metabolized into its active form 5-fluorouracil through a series of enzymatic reactions in the liver and the tumor[12]. 5-fluorouracil is inactivated into dihydrofluorouracil through the dihydropyrimidine dehydrogenase (DPYD) enzyme in the liver. CAP meets the FDA criteria for a "highly variable" drug, with extensive variation in efficacy and toxicity between subjects[17] that cannot be explained by the known host risk factors[18,19]. In **Chapter 2**, we characterized the gut bacterial enzymes that are homologs of host DPYD and measured their effects on treatment efficacy in a novel mouse tumor model.

One potential outcome of chemotherapy is treatment resistance. A well-documented protein involved in multidrug resistance is P-glycoprotein (P-gp), a key mammalian drug efflux transporter that limits the absorption of a wide range of structurally diverse compounds, including anticancer agents doxorubicin and paclitaxel[20]. Cancer cells take advantage of this defense mechanism and upregulate ABCB1 expression, which encodes P-gp, to lower intracellular drug concentrations, rendering anticancer drugs less effective[20]. Studies comparing germ-free mice with conventionally raised mice showed that the gut microbiota modulates P-gp expression[20]. However, the functional consequences of microbiome-driven changes in Abcb1a expression and the mechanisms responsible remain largely unexplored. In **Chapter 3**, we begin to address this major gap in our knowledge by investigating the mechanism of microbe-transporter interaction and the resulting effect on drug absorption.

1.3 Potential role of the gut microbiome on cancer survival

For patients diagnosed with colorectal cancer, one critical question is "what can I do to increase my chance of survival?". Fortunately, health behaviors are a modifiable factor that is significantly associated with the risk of CRC mortality. For example, the American Cancer Society Nutrition and Physical Activity Guidelines for Cancer Survivors (ACS guidelines) recommend avoiding obesity, being physically active, and consuming a healthy diet rich in fruits, vegetables, and whole grains[21]. Adherence to these guidelines was associated with a 42% lower risk of death and longer 5-year survival among stage III colon cancer patients[22]. However, the mechanisms through which health behaviors may alter CRC survival are not well understood. We hypothesized that dietary components of the ACS guidelines act in part by changing the microbiome, motivated by the rapidly expanding literature on the role of the microbiome in cancer and the key role diet plays in shaping gut microbial community structure and function.

Diet is perhaps the most important modifiable factor that influences the gut microbiome[23,24]. Short-term dietary interventions in healthy individuals can lead to significant shifts in gut bacterial abundance and gene expression[25]. Dietary interventions in obese subjects have revealed marked changes in the gut microbiome in response to caloric restriction[26] and the consumption of a high-fat, low-carbohydrate ketogenic diet[27]. Diets rich in whole grains and fruits and vegetables were associated with a positive quality of life and lower recurrence in CRC patients and survivors[22,28,29]. However, data on the effects of dietary interventions on the gut microbiomes of CRC patients remain lacking, especially in the context of survivorship. In **Chapter 4**, we report the results of a pilot clinical study of 28 CRC survivors at the University of California, San Francisco

(UCSF) to investigate whether adherence to the ACS guidelines was associated with inter-individual variations in the gut microbiomes of CRC survivors.

1.4 References

1. Cancer.Net Editorial Board. *Cancer.Net* https://www.cancer.net/about-us/cancernet-editorial-board (2014).
2. Sánchez-Alcoholado, L. *et al.* The Role of the Gut Microbiome in Colorectal Cancer Development and Therapy Response. *Cancers* **12**, (2020).
3. Cao, Y. *et al.* Commensal microbiota from patients with inflammatory bowel disease produce genotoxic metabolites. *Science* **378**, eabm3233 (2022).
4. Wilson, M. R. *et al.* The human gut bacterial genotoxin colibactin alkylates DNA. *Science* **363**, (2019).
5. Kostic, A. D. *et al.* Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome Res.* **22**, 292–298 (2012).
6. Dejea, C. M. *et al.* Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. *Science* **359**, 592–597 (2018).
7. Chung, L. *et al.* Bacteroides fragilis Toxin Coordinates a Pro-carcinogenic Inflammatory Cascade via Targeting of Colonic Epithelial Cells. *Cell Host Microbe* **23**, 421 (2018).
8. Kostic, A. D. *et al.* Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe* **14**, 207–215 (2013).
9. Spanogiannopoulos, P., Bess, E. N., Carmody, R. N. & Turnbaugh, P. J. The microbial pharmacists within us: a metagenomic view of xenobiotic metabolism. *Nat. Rev. Microbiol.* **14**, 273–287 (2016).
10. Lam, K. N., Alexander, M. & Turnbaugh, P. J. Precision Medicine Goes

Microscopic: Engineering the Microbiome to Improve Drug Outcomes. *Cell Host Microbe* **26**, 22–34 (2019).

11. Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R. & Goodman, A. L. Mapping human microbiome drug metabolism by gut bacteria and their genes. *Nature* (2019) doi:10.1038/s41586-019-1291-3.
12. Javdan, B. *et al.* Personalized Mapping of Drug Metabolism by the Human Gut Microbiome. *Cell* **181**, 1661–1679.e22 (2020).
13. Wallace, B. D. *et al.* Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. *Science* **330**, 831–835 (2010).
14. Biernat, K. A. *et al.* Structure, function, and inhibition of drug reactivating human gut microbial β-glucuronidases. *Sci. Rep.* **9**, 825 (2019).
15. Horowitz, J., Saukkonen, J. J. & Chargaff, E. Effects of fluoropyrimidines on the synthesis of bacterial proteins and nucleic acids. *J. Biol. Chem.* **235**, 3266–3272 (1960).
16. Bloch, A. & Hutchison, D. J. A MECHANISM OF RESISTANCE TO FLUOROPYRIMIDINES. *Cancer Res.* **24**, 433–439 (1964).
17. 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).
18. Haller, D. G. *et al.* Potential regional differences for the tolerability profiles of fluoropyrimidines. *J. Clin. Oncol.* **26**, 2118–2123 (2008).
19. Jennings, B. A. *et al.* Evaluating Predictive Pharmacogenetic Signatures of Adverse Events in Colorectal Cancer Patients Treated with Fluoropyrimidines. *PLoS ONE*

vol. 8 e78053 Preprint at https://doi.org/10.1371/journal.pone.0078053 (2013).

20. Kyaw, T. S. & Turnbaugh, P. J. Tiny Gatekeepers: Microbial Control of Host Drug Transporters. *Clin. Pharmacol. Ther.* **112**, 443–445 (2022).
21. Rock, C. L. *et al.* American Cancer Society guideline for diet and physical activity for cancer prevention. *CA Cancer J. Clin.* **70**, 245–271 (2020).
22. Van Blarigan, E. L. *et al.* Association of Survival With Adherence to the American Cancer Society Nutrition and Physical Activity Guidelines for Cancer Survivors After Colon Cancer Diagnosis: The CALGB 89803/Alliance Trial. *JAMA Oncol* **4**, 783–790 (2018).
23. Rothschild, D. *et al.* Environment dominates over host genetics in shaping human gut microbiota. *Nature* **555**, 210–215 (2018).
24. Zeevi, D. *et al.* Personalized Nutrition by Prediction of Glycemic Responses. *Cell* **163**, 1079–1094 (2015).
25. David, L. A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014).
26. von Schwartzenberg, R. J. *et al.* Caloric restriction disrupts the microbiota and colonization resistance. *Nature* **595**, 272–277 (2021).
27. Ang, Q. Y. *et al.* Ketogenic Diets Alter the Gut Microbiome Resulting in Decreased Intestinal Th17 Cells. *Cell* **181**, 1263–1275.e16 (2020).
28. Balhareth, A., Aldossary, M. Y. & McNamara, D. Impact of physical activity and diet on colorectal cancer survivors' quality of life: a systematic review. *World J. Surg. Oncol.* **17**, 153 (2019).
29. Piawah, S., Walker, E. J., Van Blarigan, E. L. & Atreya, C. E. The Gut Microbiome

in Colorectal Cancer. *Hematol. Oncol. Clin. North Am.* **36**, 491–506 (2022).

Chapter 2: The gut microbiome alters anti-cancer drug efficacy

2.1 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 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 toward 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].

In this study, we focused on oral fluoropyrimidine capecitabine (CAP), which is the standard of care therapy for colorectal cancer. This prodrug is activated into 5-fluorouracil (5-FU) through a series of enzymatic reactions mainly in the liver and the tumor and is inactivated into dihydrofluorouracil (DHFU) via dihydropyrimidine dehydrogenase (DPYD) enzyme in the liver[19,20] (**Figure 2.1**). 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[21] that cannot be explained by the known host risk factors[19,20]. 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[22,23]. Host dihydropyrimidine dehydrogenase (*DPYD*) and thymidine phosphorylase genes have the strongest association with toxicities, but they only account for 33.9% of adverse events[22,23]. Given that the gut microbiome could similarly activate CAP into 5-FU and inactivate it into DHFU via bacterial *preTA* operon (**Figure 2.1**), we investigated whether bacterial *preTA* alters the efficacy of CAP in mouse tumor models.

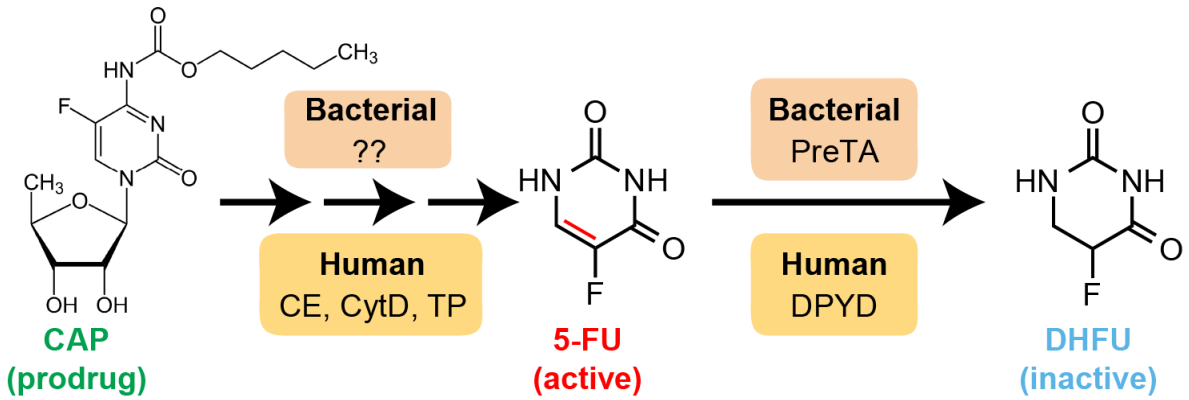


Figure 2.1. CAP metabolism in humans and bacteria. Capecitabine (CAP) is activated into 5-fluorouracil (5-FU) via carboxylesterase (CE), cytidine deaminase (CytD), and thymidine phosphorylase (TP) in humans. The enzymes responsible for this reaction in bacteria are currently unknown. 5-FU is inactivated into dihydrofluorouracil (DHFU) via dihydropyrimidine dehydrogenase (DPYD) in humans and the PreTA enzyme in bacteria.

2.2 Results

2.2.1 Bacterial PreTA interferes with cancer treatment in mice

We used a novel tumor xenograft model to test the impact of bacterial PreTA on CAP treatment efficacy in three independent experiments that varied in the enrollment criteria and streptomycin dose (**Figure 2.2a**, **Supplemental Figures S1a,g**). Mice injected with HCT-116 human CRC cells were randomly split into two colonization (Δ*preTA* or *preTA*$^{++}$ streptomycin-resistant *E. coli*) and two treatment (CAP or vehicle) groups. As expected, baseline tumor volumes were highly variable in experiments utilizing a set enrollment date (**Supplemental Figures S1b,h**), in contrast to rolling enrollment based on tumor size (**Figure 2.2b**); there were no significant differences between groups. *E. coli* colonization level was dose-dependent on streptomycin concentration and similar between groups (**Supplemental Figures S1c,i,** and **Figure 2.4a**). 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 (**Figures 2.2c-e**, rolling enrollment) corresponding to significantly prolonged survival (**Figure 2.2f**). Similar trends were observed in the high streptomycin, set enrollment experiment (**Supplemental Figures S1d-f**); however, low streptomycin did not result in detectable differences between groups (**Supplemental Figures S1j-l**). Thus, high levels of colonization by *preTA-encoding* bacteria may be necessary to interfere with drug efficacy.

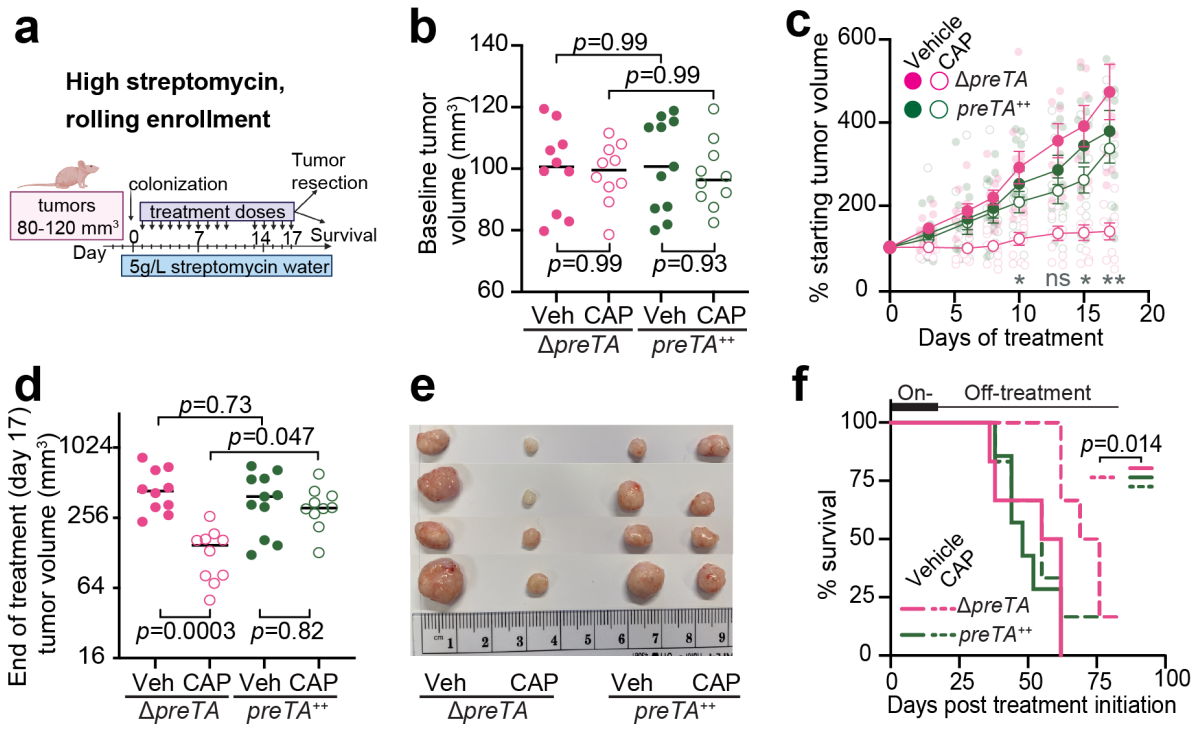


Figure 2.2. PreTA interferes with capecitabine (CAP) efficacy in mice. (a) Schematic of xenograft mice colonized with ΔpreTA or preTA^{++} E. coli and treated with capecitabine (CAP) or vehicle (Veh). **(b, d)** Tumor volumes **(b)** pre- and **(d)** post-treatment (n=11 mice preTA^{++}-Veh group; n=10 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 ΔpreTA-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 ΔpreTA-CAP to all other groups). Two mice were censored as they did not reach the endpoint when the experiment ended on day 83.

2.2.2 PreTA reduces anti-cancer drug bioavailability in mice

The most parsimonious explanation for our tumor 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 (**Figure 2.3a,e**). 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, **Figure 2.3b**). Total circulating drug levels **(Figure 2.3c**) and peak plasma concentration (**Figure 2.3d**) were significantly lower in *preTA*$^{++}$ colonized mice relative to Δ*preTA* controls. Similar trends were observed at a lower CAP dose (500 mg/kg, **Supplemental Figure S1m**). To remove the potential confounding effect of the background microbiota, we conducted a similar study using gnotobiotic mice **(Figure 2.3e)**. Gnotobiotic mice mono-colonized with Δ*preTA, wt,* or *preTA*$^{++}$ *E. coli* had comparable colonization levels (**Figure 2.3f**) and weights (**Figure 2.3g**). 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 (**Figure 2.3h**). DHFU was undetectable in plasma from all experiments.

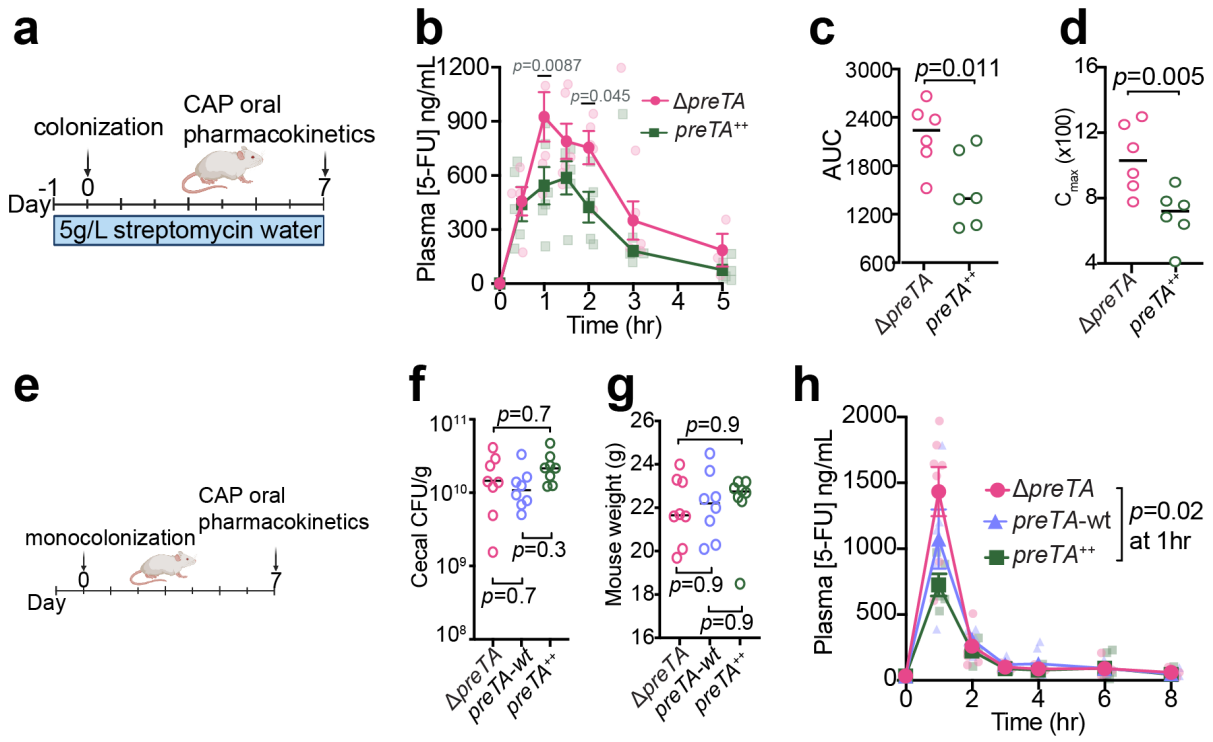


Figure 2.3. PreTA reduces 5-FU bioavailability in mice. (a) Pharmacokinetics schematic in CONV-R mice colonized with Δ*preTA* or *preTA*$^{++}$ *E. coli* and administered with 1100 mg/kg CAP. **(b)** LC-MS/MS quantification of plasma 5-FU (n=6 mice/group; opaque dots and lines represent mean±SEM; mixed-effects analysis with Sidak's correction). **(c, d)** Pharmacokinetics parameters such as **(c)** area under the curve (AUC) and **(d)** peak plasma concentrations (C_{max}) for plasma 5-FU in ng/mL (n=6 mice/group; lines are medians; Welch's one-tailed *t-test*). **(e)** 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. **(f, g)** Potential confounders like **(f)** colonization levels assessed by colony-forming units (CFU) per gram stool and **(g)** mouse weight were measured (n=8 mice/group; lines are medians; one-way ANOVA with Tukey's correction). **(h)** 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).

2.2.3 PreTA has minimal impact on the rest of the microbiota

To identify the 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. Gut microbial diversity and community structure were significantly altered over time (**Figures 2.4b-f**), consistent with prior data in the streptomycin model[24]. We did not detect any significant differences in community structure, diversity, or *E. coli* abundance between colonization groups (**Figures 2.4b-f**). Only 3/69 amplicon sequence variants (ASVs) were differentially abundant between groups (**Figure 2.4g**). 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 the potential impacts of PreTA on host drug absorption and metabolism.

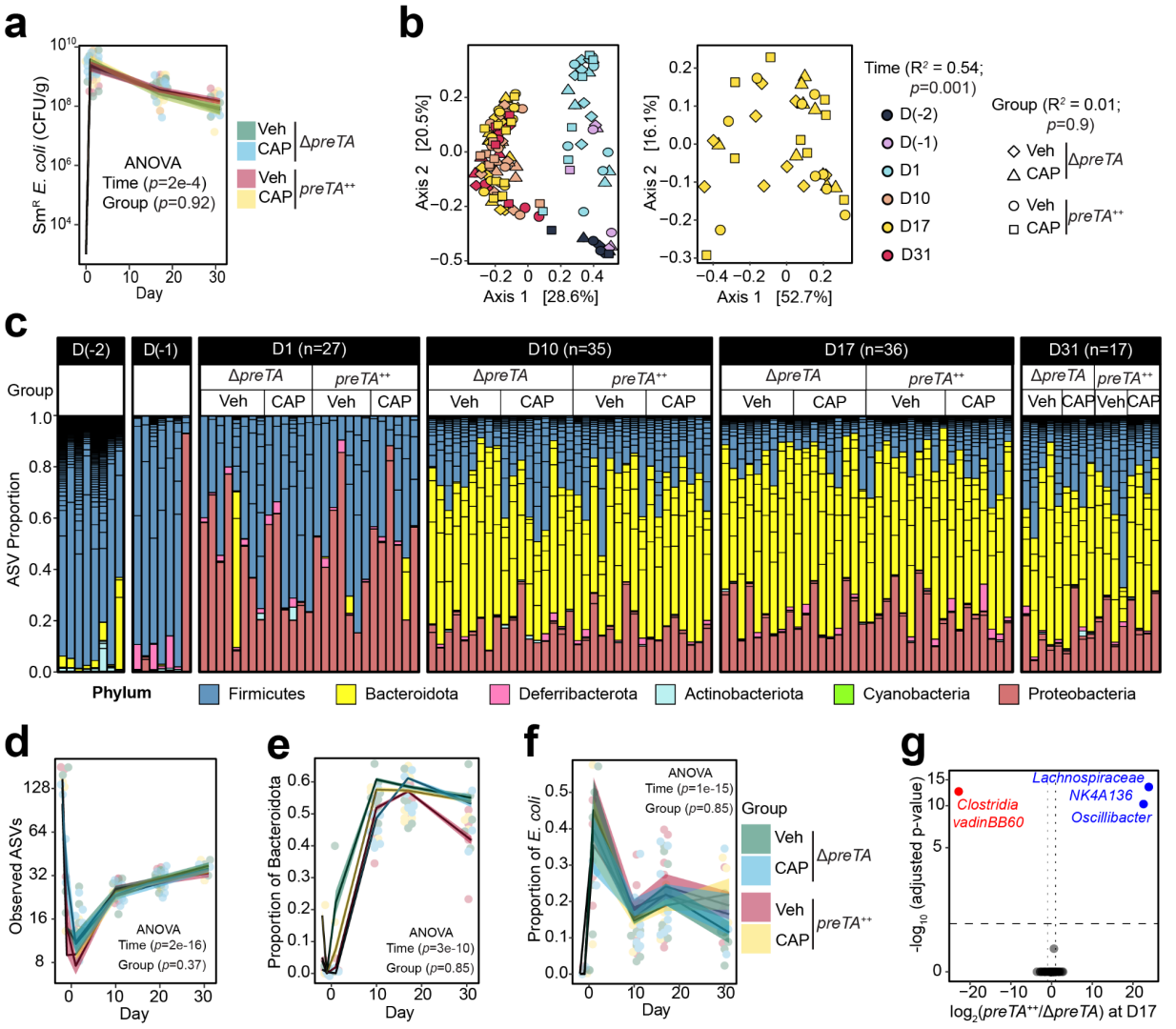


Figure 2.4. Consistent shifts in the gut microbiota across groups. (a) Quantification of streptomycin-resistant *E. coli* in feces across time (n=10-11 mice/group, 2-way ANOVA test). During log transformation, we replaced the zeros with 1000 CFU/g, which is our limit of detection. **(b)** Principal coordinate 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 using ADONIS statistical package). **(c)** Microbial community composition at the phylum level. Each bar represents a stool from each mouse. Short horizontal lines within bars represent different sequence variants. **(d)** The number of sequence variants observed over time for each treatment group (n=5-9 mice/group/time). **(e,f)** The proportion of **(e)** Bacteroidota and **(f)** *E. coli* over time for each treatment group (n=5-9 mice/group/time). *p*-values in panels e-g are from 2-way ANOVA tests. **(g)** Volcano plots of differentially abundant sequence variants at day 17 (FDR<0.1, |log2 fold-change|>1).

2.2.4 CAP impacts bacterial transcription in *Escherichia coli*

Given that *E. coli* metabolizes CAP, resulting in altered treatment efficacy and bioavailability in mice, we used transcriptional profiling (RNA-seq) to test whether exposure to CAP or its active metabolite 5-FU impacts the metabolic pathways of *E. coli*. *E. coli* was grown to the mid-exponential phase and then exposed to sub-minimal inhibitory concentration levels of CAP and 5-FU under both aerobic (n=2 independent experiments) and anaerobic conditions. Despite marked differences in transcript levels between growth conditions and experiments (**Figure 2.5a**), we were able to identify significant effects of both drugs relative to vehicle controls (**Figures 2.5b,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 (**Figure 2.5d)**. Across both experiments, we identified 1,720 differentially expressed genes (DEGs; FDR<0.1, |log2 fold-change|>1, DESeq; **Figure 2.5e**). 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* (**Figure 2.5f**). 5-FU exposure upregulated *carA*, a key gene involved in the first committed step of the *de novo* pyrimidine biosynthetic pathway.

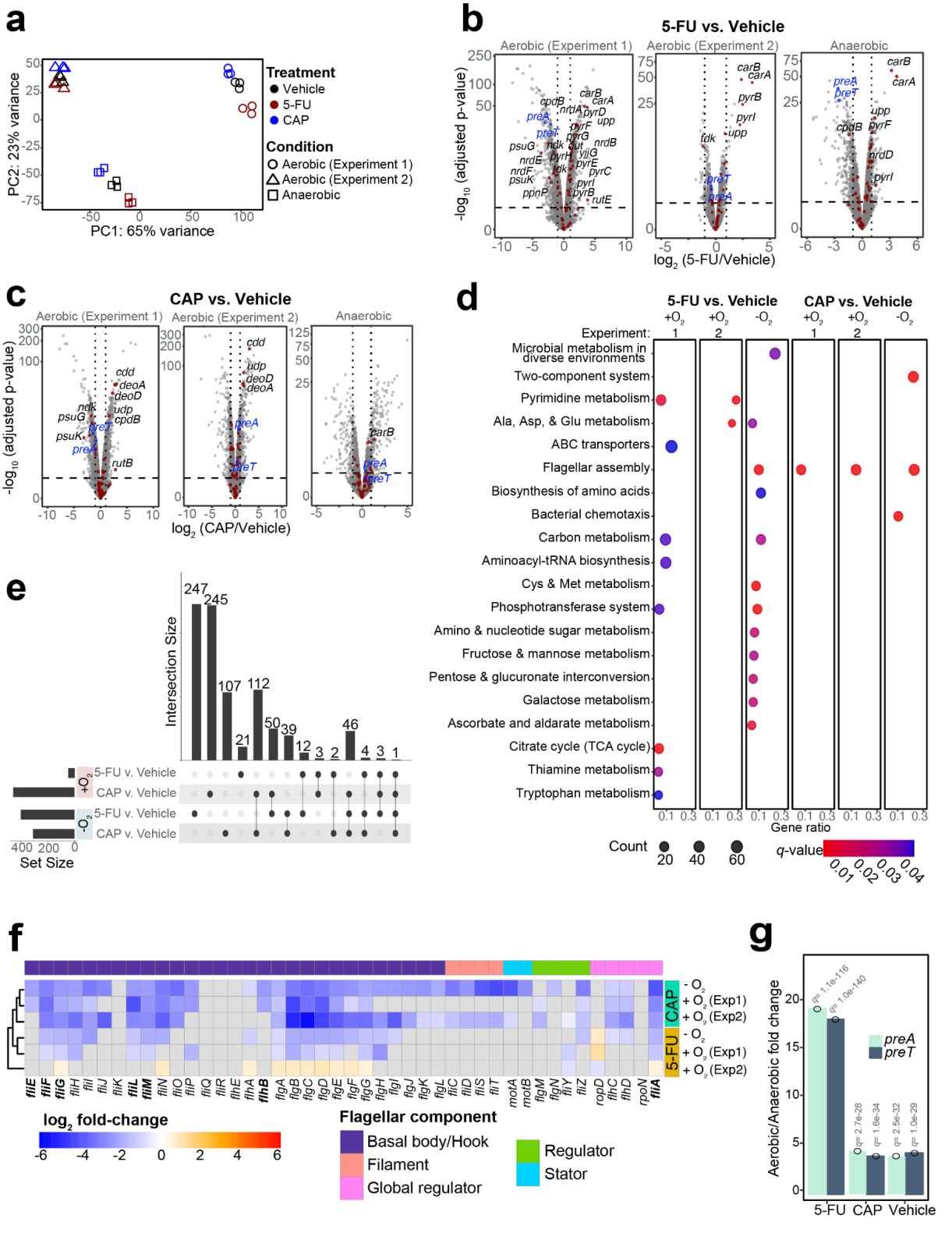

a
PC2: 23% variance
PC1: 65% variance
Treatment
Vehicle
5-FU
CAP
Condition
Aerobic (Experiment 1)
Aerobic (Experiment 2)
Anaerobic
b
5-FU vs. Vehicle
Aerobic (Experiment 1)
Aerobic (Experiment 2)
Anaerobic
-log10 (adjusted p-value)
log2 (5-FU/Vehicle)
c
CAP vs. Vehicle
log2 (CAP/Vehicle)
d
5-FU vs. Vehicle
CAP vs. Vehicle
Experiment:
Microbial metabolism in diverse environments
Two-component system
Pyrimidine metabolism
Ala, Asp, & Glu metabolism
ABC transporters
Flagellar assembly
Biosynthesis of amino acids
Bacterial chemotaxis
Carbon metabolism
Aminoacyl-tRNA biosynthesis
Cys & Met metabolism
Phosphotransferase system
Amino & nucleotide sugar metabolism
Fructose & mannose metabolism
Pentose & glucuronate interconversion
Galactose metabolism
Ascorbate and aldarate metabolism
Citrate cycle (TCA cycle)
Thiamine metabolism
Tryptophan metabolism
Gene ratio
Count
q-value
e
Intersection Size
Set Size
5-FU v. Vehicle
CAP v. Vehicle
f
log2 fold-change
Flagellar component
Basal body/Hook
Filament
Global regulator
Regulator
Stator
g
Aerobic/Anaerobic fold change
5-FU
CAP
Vehicle
preA
preT

Figure 2.5. Shared and unique transcriptional response to the related fluoropyrimidines 5-FU and 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[25]). **(e)** Upset plot comparing the fluoropyrimidine-responsive DEGs under different growth conditions. 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 the 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[26]. **(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).

2.3 Discussion

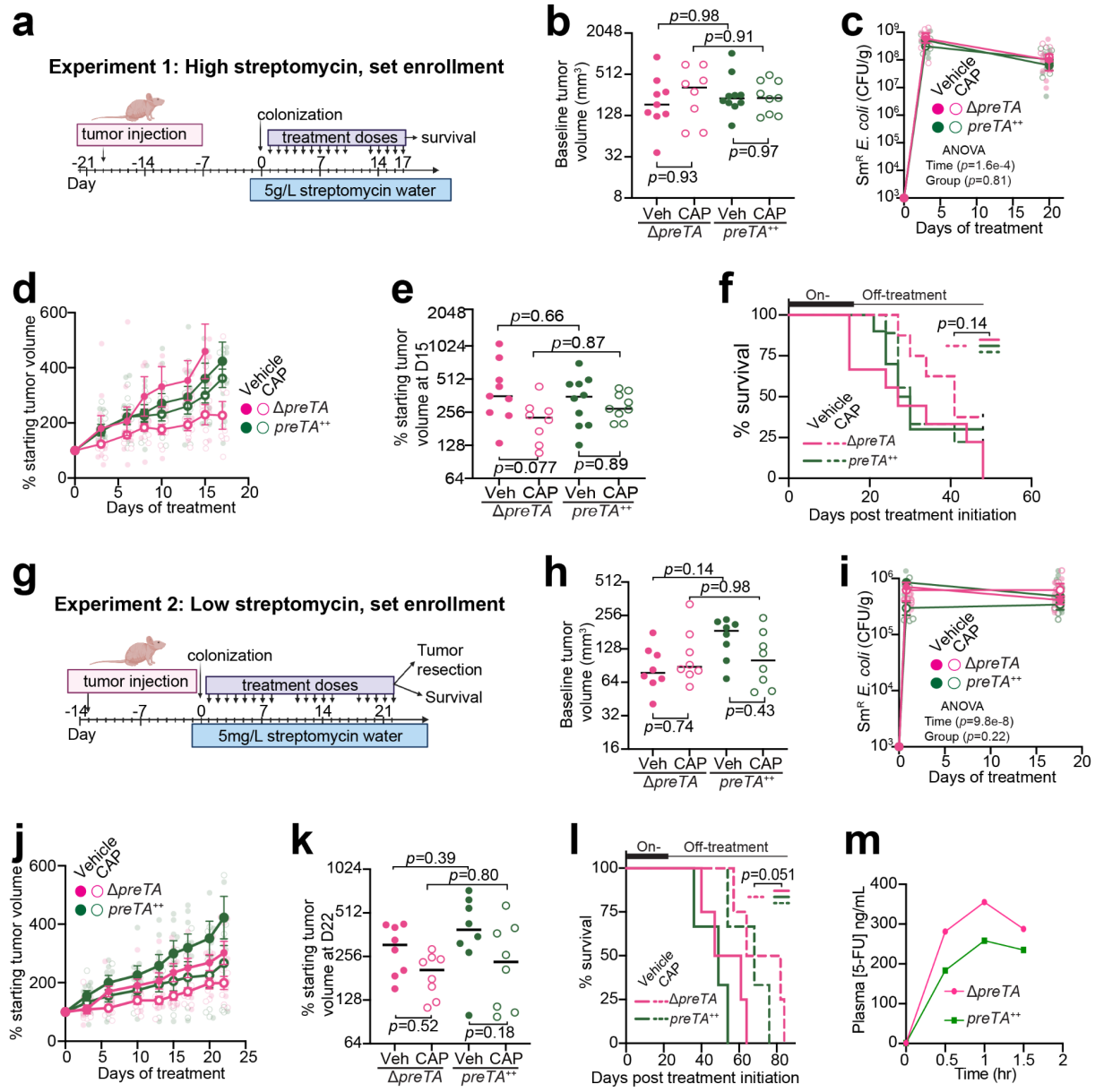
Our results demonstrate how pathways for the metabolism of fluoropyrimidine drugs are conserved across two domains of life. These findings have broad implications across multiple disciplines. Elucidating drug mechanisms of action in bacteria may provide translational insights for host tissues, raising 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 results support the hypothesis that bacterial drug inactivation in the gastrointestinal tract or even within diseased tissue[27] could interfere with drug efficacy, more work is needed to assess the relative impact of this biotransformation on efficacy versus gastrointestinal or other adverse 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.

This study has multiple important limitations. 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 complex gut microbiota. These efforts will require the development of novel genetic tools, as demonstrated for gut Clostridia[28]. 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.

Despite these limitations, 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 tests of 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[29].

2.4 Supplementary Figures



Supplemental Figure S1. PreTA decreases the efficacy and oral bioavailability of CAP. (a, g) Experimental designs for our xenograft experiments 1 and 2, respectively. **(b, h)** Baseline tumor volumes prior to treatment (n=8-10 mice/group for experiment 1, n=8 mice/group for experiment 2; lines represent the median). **(c, i)** Quantification of streptomycin-resistant *E. coli* in feces across time (n=6 mice/group for experiment 1, n=8 mice/group for experiment 2, 2-way ANOVA with Tukey's multiple comparison correction). No colonies were detected at time zero. During the log transformation, we replaced the

zeros with 1000 CFU/g, which is our limit of detection. **(d, j)** Percentage of starting tumor volumes over time (n=8-10 mice/group for experiment 1, n=8 mice/group for experiment 2; error bars represent the standard error). **(e, k)** Percentage of starting tumor volumes on **(e)** day 15 (n=8-10 mice/group) and **(k)** day 22 (n=8 mice/group); timepoint selected to capture all mice prior to euthanasia; lines represent the median. **(f, l)** Percentage of mice reaching the humane endpoint defined as tumor length >20 mm (n=8-10 mice/group for experiment 1, n=3-4 mice/group for experiment 2, Log-rank Mantel-Cox test comparing Δ*preTA*-CAP to all other groups). In experiment 1, 8 mice that did not reach the endpoint when the experiment ended on day 48 were censored. Censored mice were indicated by black boxes in the graph. *p*-values in panels (**b, e, h, k**) are from two-way ANOVA tests with Tukey's multiple comparison corrections. **(m)** 8-10-week-old male BALB/c mice were given streptomycin water one day prior to colonization with Δ*preTA* or *preTA*$^{++}$ *E. coli*. 7 days post-colonization, 500 mg/kg CAP was orally administered and longitudinal plasma samples were collected from each mouse (see the experimental design in **Figure 2.3**). LC-QTOF/MS was used to quantify 5-FU from pooled plasma samples following 500 mg/kg CAP administration (n=5 mice/group).

2.5 Materials and Methods

Chemicals. All materials were purchased from the specified supplier and used without further purification unless otherwise stated. 5-Fluorouracil (5-FU; ≥99% purity) was 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 were purchased from Honeywell-Burdick & Jackson (Muskegon, MI).

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 minimal inhibitory concentration (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 subjected 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 x 75 bp) using the MiSeq V3 platform (experiment 1) and the NextSeq Mid Output platform (experiment 2). Reads were mapped to the *E. coli* MG1655 genome sequence (NCBI Reference Sequence: NC_000913.3) using Bowtie2[30] and HTSeq[31] was used to count the number of reads to *E. coli* genes. Differential gene expression was analyzed using DESeq[32]. Differentially expressed genes following drug treatment were defined as transcripts exhibiting an absolute log2 fold change ≥1 and an FDR<0.1 relative to the no drug control.

Liquid chromatography-mass spectrometry analysis of 5-FU and metabolites. We used two separate high-resolution mass spectrometry instruments based on 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 $^{13}C^{15}N_2$). Protein from conditioned media samples was 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 x 2 mm, 3 mm) connected to a Phenomenex Securityguard™ guard cartridge (4x2 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. The 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 a 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-$^{2}H_{11}$). Protein from conditioned media samples was 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 x 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. The 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 limit 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 a 3:1 methanol: plasma volume ratio. 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 were 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[33] with the pSIJ8 vector as previously described[34]. The *preTA* operon was replaced using allelic exchange by electroporation of a PCR product containing the kanamycin cassette[35,36], which was generated using the primers preTA-P1-KEIO_F (5'-TTTACTAAACGCTCGCCTTAATTACCTATAGCATTAAGGAAGATCACATG-3') and preTA-P1-KEIO_R (5'-GGAACCAAATGCCGGATGCGTCTCTGGTTTTACAACGTTACCGGGTGTTC-3') 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. Engineered strains were made streptomycin resistant by using the recombineering protocol[33] using pSIJ8[34], as described above.

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, the 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*$^{++}$*).* Seven 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, and 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 **Supplemental Figure S1a, S1g, and Figure 2.2a** for designs of experiments 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 3x100 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 the cell-Matrigel™ mix ($1x10^6$ 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[37]. 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 vigorous 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 at 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 $\times$ $width^2$)/2[38]. 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. The 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[39] using dual error-correcting barcodes. Briefly, primary PCR was performed as a quantitative PCR using the 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 the 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[40], and DESeq[32] 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.

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[41], Kyoto Encyclopedia of Genes and Genomes (KEGG) database[25], and MIDAS v1.0 database[42] are publicly available. Sequencing data have been deposited under the NCBI BioProjects PRJNA576932 (RNA-seq and 16S rRNA gene sequences).

2.6 References

1. Thorn, C. F., Klein, T. E. & Altman, R. B. PharmGKB: the Pharmacogenomics Knowledge Base. *Methods Mol. Biol.* **1015**, 311–320 (2013).
2. Spanogiannopoulos, P., Bess, E. N., Carmody, R. N. & Turnbaugh, P. J. The microbial pharmacists within us: a metagenomic view of xenobiotic metabolism. *Nat. Rev. Microbiol.* **14**, 273–287 (2016).
3. Lam, K. N., Alexander, M. & Turnbaugh, P. J. Precision Medicine Goes Microscopic: Engineering the Microbiome to Improve Drug Outcomes. *Cell Host Microbe* **26**, 22–34 (2019).
4. Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R. & Goodman, A. L. 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).
6. Wallace, B. D. *et al.* Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. *Science* **330**, 831–835 (2010).
7. Biernat, K. A. *et al.* Structure, function, and inhibition of drug reactivating human gut microbial β-glucuronidases. *Sci. Rep.* **9**, 825 (2019).
8. Iida, N. *et al.* Commensal bacteria control cancer response to therapy by modulating the tumor microenvironment. *Science* **342**, 967–970 (2013).
9. Sivan, A. *et al.* Commensal Bifidobacterium promotes antitumor immunity and facilitates anti-PD-L1 efficacy. *Science* **350**, 1084–1089 (2015).
10. Vétizou, M. *et al.* Anticancer immunotherapy by CTLA-4 blockade relies on the gut

microbiota. *Science* **350**, 1079–1084 (2015).

11. Viaud, S. *et al.* The intestinal microbiota modulates the anticancer immune effects of cyclophosphamide. *Science* **342**, 971–976 (2013).
12. Haiser, H. J. *et al.* Predicting and manipulating cardiac drug inactivation by the human gut bacterium Eggerthella lenta. *Science* **341**, 295–298 (2013).
13. Maurice, C. F., Haiser, H. J. & Turnbaugh, P. J. Xenobiotics shape the physiology and gene expression of the active human gut microbiome. *Cell* **152**, 39–50 (2013).
14. Maini Rekdal, V., Bess, E. N., Bisanz, J. E., Turnbaugh, P. J. & Balskus, E. P. Discovery and inhibition of an interspecies gut bacterial pathway for Levodopa metabolism. *Science* **364**, (2019).
15. Nayak, R. R. *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).
17. Bisanz, J. E., Spanogiannopoulos, P., Pieper, L. M., Bustion, A. E. & Turnbaugh, P. J. How to Determine the Role of the Microbiome in Drug Disposition. *Drug Metab. Dispos.* **46**, 1588–1595 (2018).
18. Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R. & Goodman, A. L. Separating host and microbiome contributions to drug pharmacokinetics and toxicity. *Science* **363**, (2019).
19. Haller, D. G. *et al.* Potential regional differences for the tolerability profiles of

fluoropyrimidines. *J. Clin. Oncol.* **26**, 2118–2123 (2008).

20. Jennings, B. A. *et al.* Evaluating Predictive Pharmacogenetic Signatures of Adverse Events in Colorectal Cancer Patients Treated with Fluoropyrimidines. *PLoS ONE* vol. 8 e78053 Preprint at https://doi.org/10.1371/journal.pone.0078053 (2013).
21. 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).
22. Saif, M. W., Syrigos, K., Mehra, R., Mattison, L. K. & Diasio, R. B. DIHYDROPYRIMIDINE DEHYDROGENASE DEFICIENCY (DPD) IN GI MALIGNANCIES: EXPERIENCE OF 4-YEARS. *Pak. J. Med. Sci. Q.* **23**, 832–839 (2007).
23. Leonard, R., Hennessy, B. T., Blum, J. L. & 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).
24. Lam, K. N. *et al.* Phage-delivered CRISPR-Cas9 for strain-specific depletion and genomic deletions in the gut microbiome. *Cell Rep.* **37**, 109930 (2021).
25. 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).
26. Rosener, B. *et al.* Evolved bacterial resistance against fluoropyrimidines can lower chemotherapy impact in the Caenorhabditis elegans host. *Elife* **9**, (2020).
27. Geller, L. T. *et al.* Potential role of intratumor bacteria in mediating tumor resistance to the chemotherapeutic drug gemcitabine. *Science* **357**, 1156–1160 (2017).

28. Guo, C.-J. *et al.* Depletion of microbiome-derived molecules in the host using Clostridium genetics. *Science* **366**, (2019).
29. Nayak, R. R. & Turnbaugh, P. J. Mirror, mirror on the wall: which microbiomes will help heal them all? *BMC Med.* **14**, 72 (2016).
30. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
31. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
32. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
33. Sharan, S. K., Thomason, L. C., Kuznetsov, S. G. & Court, D. L. Recombineering: a homologous recombination-based method of genetic engineering. *Nat. Protoc.* **4**, 206–223 (2009).
34. Jensen, S. I., Lennen, R. M., Herrgård, M. J. & Nielsen, A. T. Seven gene deletions in seven days: Fast generation of Escherichia coli strains tolerant to acetate and osmotic stress. *Sci. Rep.* **5**, 17874 (2015).
35. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, (2006).
36. Datsenko, K. A. & Wanner, B. L. 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).
37. Myhal, M. L., Laux, D. C. & Cohen, P. S. 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).

38. 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).
39. Gohl, D. M. *et al.* Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. *Nat. Biotechnol.* **34**, 942–949 (2016).
40. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**, e61217 (2013).
41. Parks, D. H. *et al.* A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat. Biotechnol.* **38**, 1079–1086 (2020).
42. Nayfach, S., Rodriguez-Mueller, B., Garud, N. & Pollard, K. S. An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. *Genome Res.* **26**, 1612–1625 (2016).

Chapter 3: The gut microbiome alters drug absorption

3.1 Introduction

Drug metabolism and disposition depend upon both genetic and environmental factors that shape intestinal metabolism, transport across the intestinal epithelium, and first-pass metabolism in the liver[1,2]. For the past decade, research at the interface of the microbiome and pharmacology has focused on the metabolism of drugs within the gastrointestinal tract[3–5]. While this focus is understandable given the immense and poorly characterized enzymatic diversity within the microbiome, emerging data has begun to suggest that the gut microbiome has far broader impacts on drug disposition[6,7]. Gut bacterial metabolism of common food and drug additives (*i.e.,* excipients) increases drug bioavailability by lifting the inhibition of an intestinal drug influx transporter[8]. However, it remains unclear if the diverse chemicals produced by the gut microbiome[9] can also control drug influx and/or efflux transporters, and if so, whether or not these host-microbiome interactions have physiologically relevant consequences for drug bioavailability[7].

Here, we begin to address this major gap in our knowledge. We opted to focus on P-gp (encoded by the *ABCB1* gene in humans and *Abcb1a* in mice) due in part to its broad relevance for >300 known endogenous and exogenous substrates, including the cardiac drug digoxin and the anti-cancer agents doxorubicin and paclitaxel[10]. Abcb1a can be differentially expressed between germ-free (GF) and conventionally-raised (CONV-R) mice[11,12]; however, a recent study failed to detect a significant difference in Abcb1a transcript levels[13]. Variation in the specific gut microbiota found between facilities could provide one potential reason for these discrepant results, consistent with experiments administering single pathogenic, probiotic, and commensal bacteria that can tune Abcb1a

expression in either direction[11,14–16] and a recent paper implicating butyrate-producing bacteria in colonic P-gp protein levels[17]. More importantly, the functional consequences of microbiome-driven changes in Abcb1a expression and the mechanisms responsible remain largely unexplored.

Our prior work on the cardiac drug digoxin also motivated us to consider P-gp and its relationship to the gut microbiome. Digoxin is both a substrate for gut bacterial metabolism[18] and a model substrate for P-gp efflux[19]. We previously identified a two-gene operon (the cardiac glycoside reductase operon, *cgr*) which predicts strain-level variation in the metabolism of digoxin by the prevalent gut Actinobacterium *Eggerthella lenta*[20]. The *cgr2* gene was sufficient to catalyze digoxin reduction in a heterologous expression system and biochemical characterization of the Cgr2 enzyme suggested that its substrate scope was restricted to cardenolides[21]. Surprisingly, we also discovered that Cgr2 is necessary and sufficient for the induction of colonic T-helper 17 cells[22,23], suggesting that dietary and/or host substrates can be reduced by this enzyme. Yet despite these mechanistic insights into the metabolic activity of *E. lenta,* the relative impacts of bacterial metabolism versus other types of host-microbiome interactions on drug disposition remain poorly understood, even for a drug as well-characterized as digoxin.

3.2 Results

3.2.1 *Eggerthella lenta* increases drug absorption in mice

Given that *E. lenta* metabolizes the cardiac drug digoxin[20], we hypothesized that mono-colonization of germ-free (GF) mice with *E. lenta* DSM2243 would significantly decrease digoxin bioavailability by enhancing first-pass metabolism. Nine-week-old mixed sex GF Swiss Webster mice were mono-colonized with *E. lenta* for 4 weeks followed by oral administration of 200 µg/kg digoxin. Colonization was confirmed by qPCR (**Figure 3.1a**). Surprisingly, *E. lenta* increased serum digoxin levels at the first time-point (1-hour post-drug administration; **Figure 3.1b**), resulting in a higher maximum concentration (C_{max}; **Figure 3.1c**) and lower time to reach the maximum concentration (t_{max}; **Figure 3.1d**) relative to GF controls. The digoxin area under the curve (AUC) was unaffected (**Figure 3.1e**) due to significantly decreased digoxin concentrations at 4- and 5-hours post-administration (**Figure 3.1b**). Taken together, these results suggest that *E. lenta* increases the rate of absorption of digoxin in gnotobiotic mice, balancing out the effects of bacterial drug metabolism[20].

Next, we sought to test if the observed change in digoxin absorption in response to *E. lenta* was unique to GF mice, which have broad changes to gut physiology, including changes in the expression of numerous drug transporters relative to conventionally-raised (CONV-R) specific pathogen-free animals[12,13]. To ensure consistent exposure to high levels of *E. lenta* we orally gavaged 10^9 *E. lenta* colony-forming units (CFU) or sterile media control for 7 consecutive days to 8-10-week-old female CONV-R Swiss-Webster-mice prior to the oral administration digoxin. High levels of *E. lenta* were still detected in the cecum collected at the end of the experiment (**Figure 3.1f**). In this repeat experiment,

we opted to use a lower dose of digoxin (30 μg/kg) to avoid saturating transporter effects for digoxin. The results were broadly consistent with data from gnotobiotic mice (**Figures 3.1a-e**). *E. lenta* significantly increased serum digoxin at the 1-hour time-point following drug administration (**Figure 3.1g**) and significantly increased C_{max} (**Figure 3.1h**). There was a trend toward decreased t_{max} (*p*-value=0.10, Wilcoxon test; **Figure 3.1i**). Overall AUC was unchanged, due to higher levels in the control mice at the later time points (**Figure 3.1j**). These results demonstrate that *E. lenta* increases the rate of digoxin absorption in the presence or absence of the mouse gut microbiota.

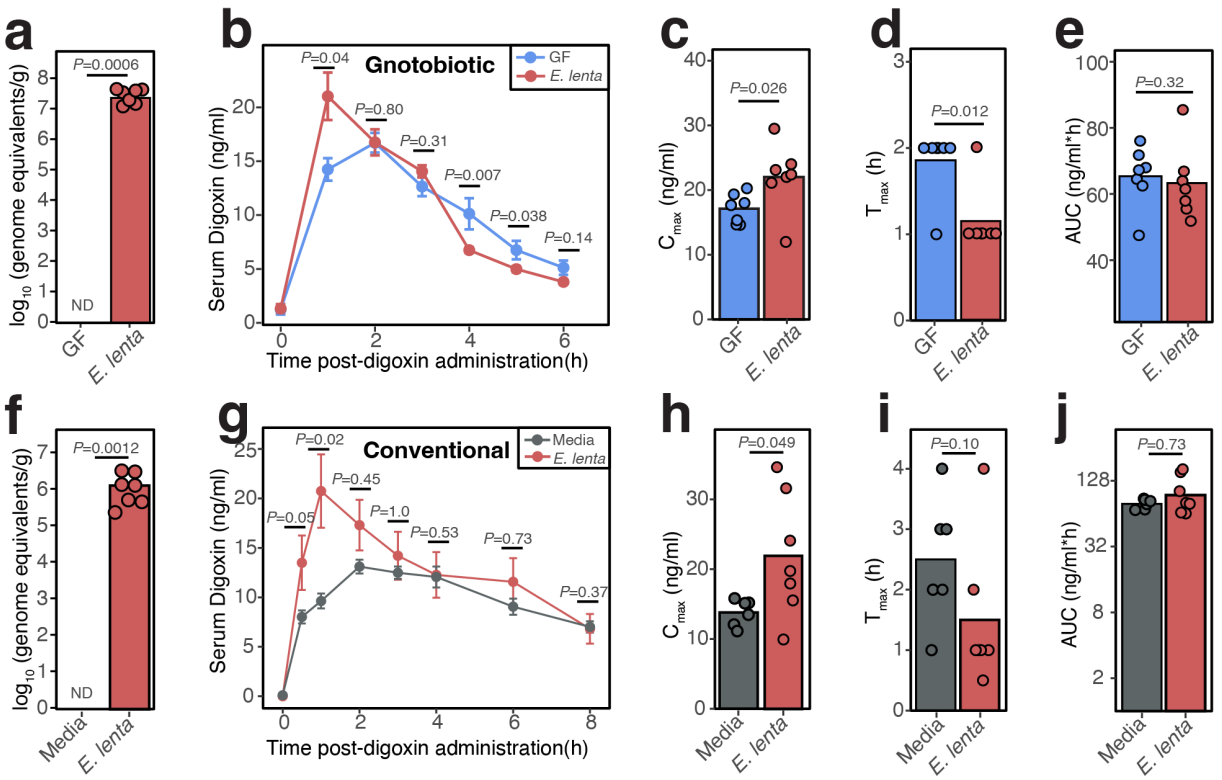


Figure 3.1. *E. lenta* increases the rate of drug absorption. (a-e) 9-week old GF mixed-sex Swiss Webster mice were mono-colonized with *E. lenta* DSM2243 for 4 weeks followed by a single oral administration of 200 µg/kg digoxin. **(a)** Quantification of *E. lenta* DSM2243 in grams of cecal content via qPCR. No amplification was detected (ND) in the entire GF group (n=6-7 mice/group; Mann-Whitney test). **(b)** Serum digoxin was quantified at different time points in GF or *E. lenta* mono-colonized mice post-digoxin administration (n=6-7 mice/group; Wilcoxon test). From the pharmacokinetic (PK) curves, key PK parameters were quantified, such as **(c)** maximal serum concentration (C_{max}), **(d)** the time at which concentrations are highest (t_{max}), and **(e)** area under the curve (AUC). **(f-j)** In a separate mouse model, 8-10 week-old female CONV-R Swiss Webster mice were gavaged 10^9 live *E. lenta* DSM2243 cells or sterile BHI^+ media control for 7 consecutive days prior to oral administration of digoxin at 30 µg/kg. **(f)** *E. lenta* engraftment was quantified via qPCR using cecal content. BHI-gavaged group had no detectable (ND) *E. lenta* (n=6-7 mice/group; Mann-Whitney test). **(g)** Serum digoxin was quantified by LC-MS/MS at different timepoints post-digoxin administration (n=6-7 mice/group; Wilcoxon test). From the PK curves, **(h)** C_{max}, **(i)** t_{max}, and **(j)** AUC were quantified (Wilcoxon test).

3.2.2 Human gut Actinobacterial metabolites inhibit P-gp efflux

Digoxin is a model substrate of the P-gp transporter[19], which is the only known mammalian mechanism that controls the absorption of digoxin. Drug-drug interactions resulting in altered P-gp function are known to disrupt digoxin pharmacokinetics, which can be dangerous due to its narrow therapeutic range[19]. We hypothesized that *E. lenta* inhibits P-gp efflux, which would explain the elevated rate of digoxin absorption in mice (**Figure 3.1**). To avoid the many confounding factors in mice, we leveraged an established cell culture assay[15]. Briefly, human enterocytes (T84 cells) were treated with Rhodamine 123 (Rh123), a fluorescent P-gp substrate, to track efflux activity under different media conditions. As expected, the P-gp inhibitors cyclosporin A[15], vinblastine[15], and verapamil[24] significantly increased intracellular Rh123 levels compared to the untreated or vehicle controls (**Figure 3.2a**). *E. lenta* cell-free supernatant (CFS) but not cell lysates resulted in significantly increased intracellular Rh123 accumulations (**Figure 3.2a**)**.** The impact of *E. lenta* CFS was dose-dependent (**Figure 3.2b**) and comparable in effect size to the purified pharmacological inhibitor controls (**Figure 3.2a**). These results suggest that one or more secreted metabolites produced by *E. lenta* inhibit P-gp.

We used three independent approaches to validate the inhibition of P-gp by *E. lenta* CFS: (*i*) gene deletion; (*ii*) alternative cell lines; and (*iii*) alternative substrates. *First*, an *ABCB1* knockout monoclonal cell line from parental T84 cells was generated via lentiviral delivery of *cas9* and a guide RNA against *ABCB1*. Successful gene deletion was confirmed via sequencing of *ABCB1* gene locus, Western blot, and activity testing (**Supplemental Figure S2**). Similar to the verapamil positive control, incubation of *E. lenta* CFS no longer increased intracellular Rh123 levels in the *ABCB1* knockout cell line,

suggesting that the mechanism is P-gp dependent and consistent with the increased basal accumulation of Rh123 in *ABCB1* knockout cells (**Figure 3.2c**). *Second*, incubation with *E. lenta* CFS led to high intracellular Rh123 concentrations in 3 other human enterocyte models (Caco-2, HCT-15, and HCT-116), indicating that inhibition is reproducible across cell lines (**Figure 3.2d**). *Third*, we validated the impact of *E. lenta* CFS on P-gp by measuring the accumulation of two additional P-gp substrates: calcein-AM[25] (**Figure 3.2e**) and the anti-cancer drug doxorubicin[25] (**Figure 3.2f**). Inhibition of P-gp efflux by *E. lenta* CFS led to a significant decrease in cancer cell viability in response to doxorubicin, consistent with the known P-gp inhibitor verapamil[24] (**Figure 3.2g**). Taken together, these results indicate that one or more metabolites secreted by *E. lenta* robustly inhibit P-gp efflux across multiple cell lines.

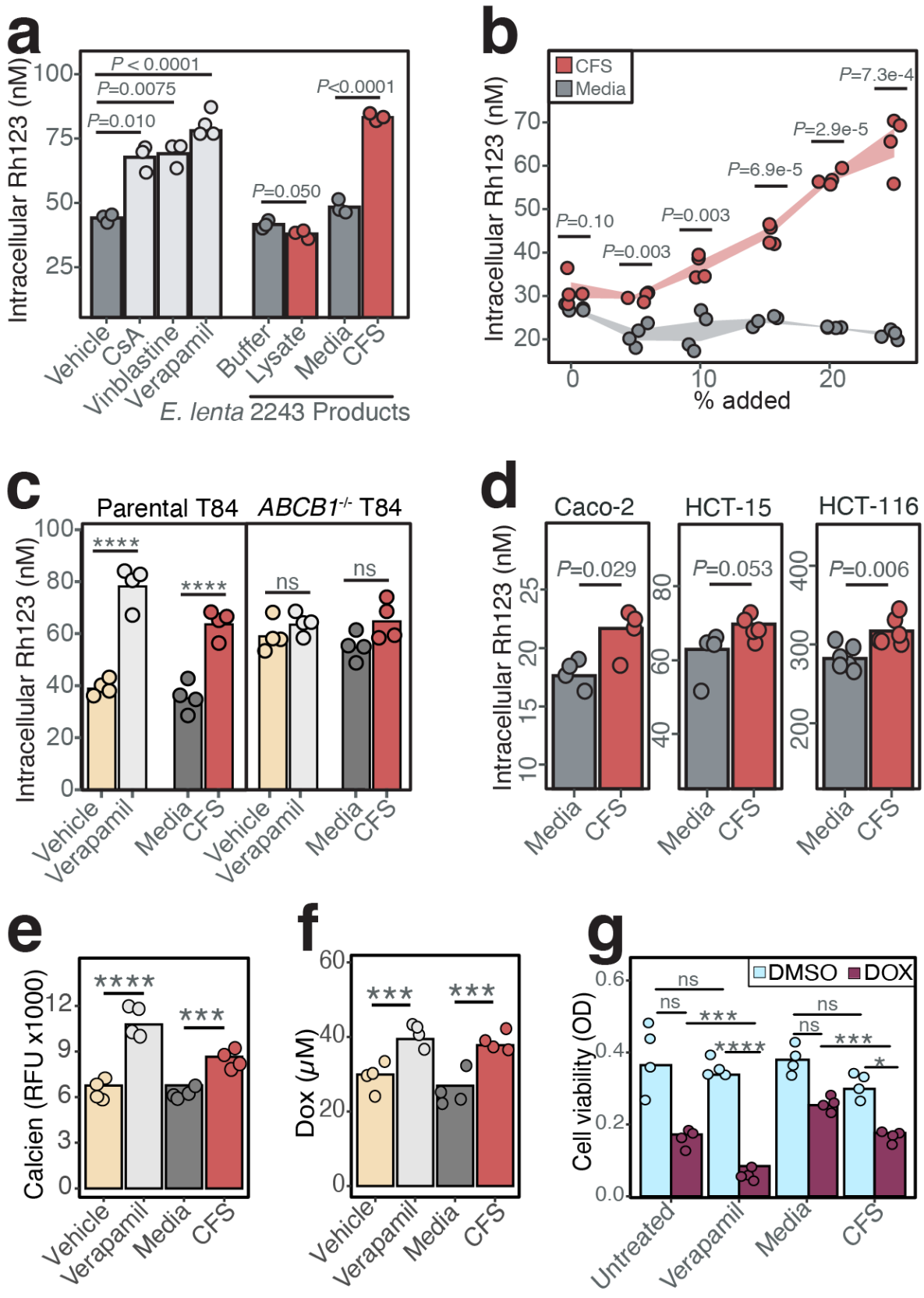

a
Intracellular Rh123 (nM)
P < 0.0001
P=0.0075
P=0.010
P=0.050
P<0.0001
Vehicle
CsA
Vinblastine
Verapamil
Buffer
Lysate
Media
CFS
E. lenta 2243 Products
b
Intracellular Rh123 (nM)
CFS
Media
P=0.10
P=0.003
P=0.003
P=6.9e-5
P=2.9e-5
P=7.3e-4
% added
c
Parental T84
ABCB1-/- T84
Intracellular Rh123 (nM)
****
****
ns
ns
Vehicle
Verapamil
Media
CFS
d
Caco-2
HCT-15
HCT-116
Intracellular Rh123 (nM)
P=0.029
P=0.053
P=0.006
Media
CFS
e
Calcien (RFU x1000)
****
***
Vehicle
Verapamil
Media
CFS
f
Dox (μM)
***
***
Vehicle
Verapamil
Media
CFS
g
DMSO
DOX
Cell viability (OD)
ns
ns
***
****
ns
ns
***
*
Untreated
Verapamil
Media
CFS

Figure 3.2. *Eggerthella lenta* inhibits P-gp efflux. (a) P-gp inhibition test of *E. lenta* DSM2243 cell pellet lysate and cell-free filter-sterilized supernatant (CFS) using rhodamine (Rh123) accumulation assay in T84 human enterocyte model (N=3-4/condition, ANOVA with Tukey's post hoc test). Vehicle (DMSO), cyclosporin A (CsA), vinblastine, and verapamil were included as controls. **(b)** Dose-response test of *E. lenta* CFS on Rh123 accumulation (P_{dose}=8.6e-13 ANOVA, N=4 wells/condition). **(c)** CFS Rh123 accumulation test on T84 *ABCB1*$^{-/-}$ knockout cells compared to parental T84 cells. **(d)** CFS activity test in 3 other human enterocyte cell lines. Activity tests in T84 cells were also completed using the accumulation of other P-gp substrates, such as **(e)** calcein-AM and **(f)** doxorubicin (DOX). **(g)** Synergistic effects of *E. lenta* CFS and DOX were tested using MTT viability cell assay. DMSO was used as vehicle control. Unless otherwise noted, *E. lenta* was cultured in BHI^{+}. N=4/condition and 2-way ANOVA with Sidak's multiple testing correction. *<0.05, **<0.01, ***<0.001, ****<0.0001.

3.2.3 *E. lenta* inhibits P-gp ATPase activity

Small molecules and bacterial toxins can inhibit P-gp function through multiple mechanisms (**Figure 3.3a**). Gene expression and protein levels did not explain the observed P-gp inhibition by *E. lenta.* ABCB1 transcript levels were significantly increased in response to *E. lenta* CFS in 3/4 cell lines tested, with a trend towards increased expression which may indicate a compensatory mechanism (**Figure 3.3b**). We did not observe any changes in P-gp protein levels in response to *E. lenta* CFS (**Figure 3.3c**), suggesting that the major mechanism of P-gp inhibition is post-translational in nature.

To test the downstream impact of *E. lenta* on P-gp activity, we turned to a cell-free vesicle system wherein the ATPase activity of P-gp can be directly quantified[26] (**Figure 3.3d**). The positive control[26] sodium orthovanadate significantly blocked P-gp ATPase activity whereas verapamil significantly stimulated ATPase activity (**Figure 3.3d**). A <3kDa fraction of small molecules from *E. lenta* CFS significantly increased free ATP in a dose-dependent manner (**Figure 3.3d**). These results suggest that *E. lenta* inhibits the ATPase activity of P-gp.

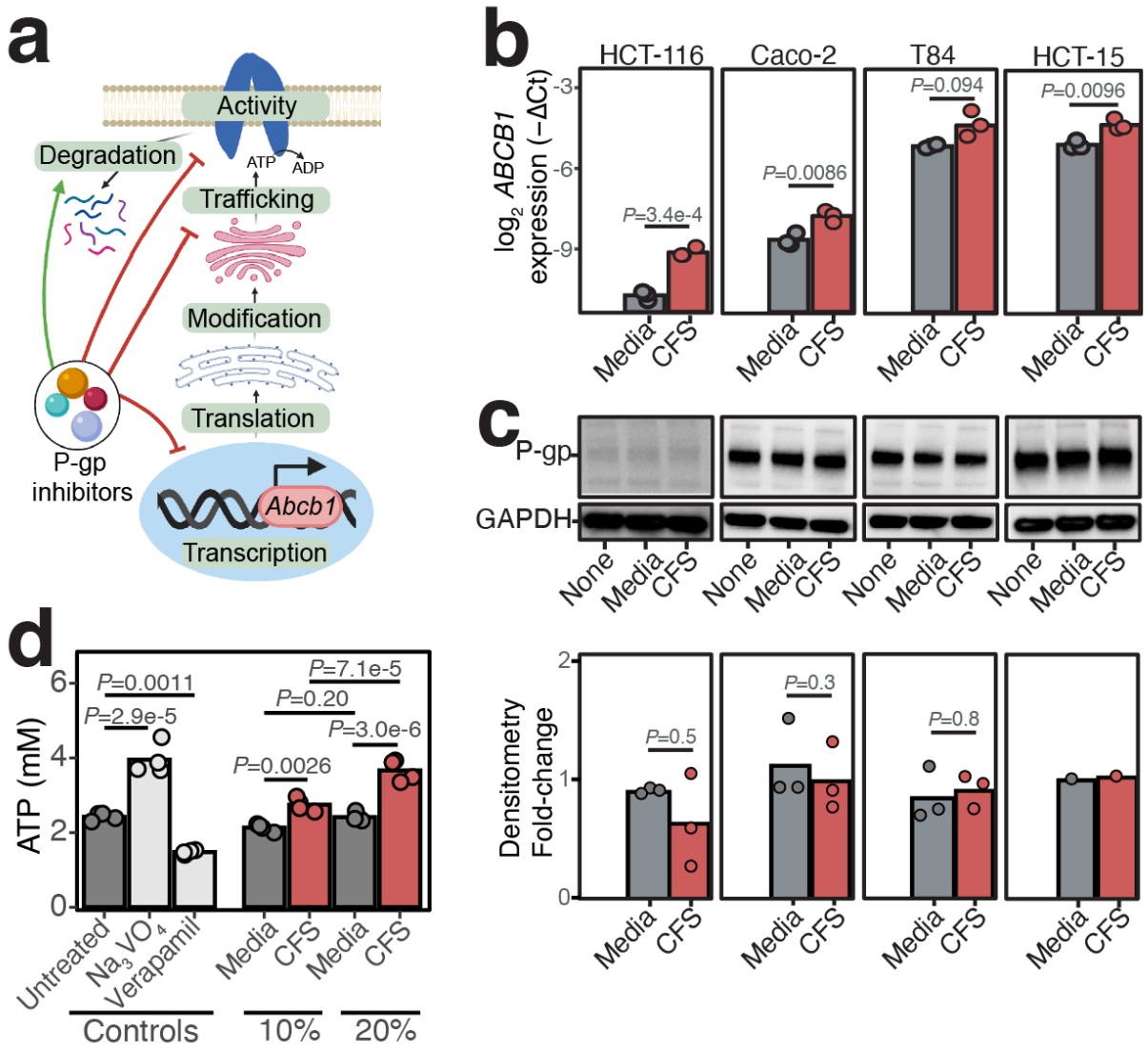


Figure 3.3. *E. lenta* inhibits P-gp ATPase activity. (a) Possible mechanisms of P-gp inhibition reported in the literature. **(b)** RT-qPCR measurement of *ABCB1* transcription in multiple human enterocytes with varying basal expression when treated with *E. lenta* CFS. *GAPDH* and beta-actin primers were used as loading controls. Amplification cycles (Ct) were normalized to untreated conditions for each cell line (N=3 biological replicates/condition; 2-way ANOVA with Sidak's correction). **(c)** Representative Western blots of untreated, BHI sterile media, or *E. lenta* CFS treated human enterocytes using C219 P-gp antibody and GAPDH antibody and densitometry quantifications normalized to untreated control for each cell line (N=1-3 biological replicates/condition; Wilcoxon test). **(d)** Cell-free P-gp vesicles were incubated with 3kDa filtrate of *E. lenta* CFS prepared in EDM media in the presence of ATP. At the end of the incubation period, the amount of remaining ATP was quantified. Sodium orthovanadate (Na_3VO_4) and verapamil P-gp inhibitors were included as controls for an ATPase inhibitor and an activator, respectively. N=4/condition; ANOVA with Tukey's HSD post hoc test.

3.2.4 Comparative genomics reveals genes linked to P-gp inhibition

Next, we sought to determine the extent to which the P-gp inhibitory trait is conserved in gut bacteria. We leveraged our previously published strain collection[27], testing 24 *E. lenta* strains and 10 *Coriobacteriia* relatives for their ability to inhibit P-gp, using our T84 cell-based assay. All of the tested *Eggerthellaceae* strains had a comparable degree of inhibition, on par with our positive control verapamil (**Figure 3.4a**). However, no inhibitory activity was observed in *Olsenella uli* (a *Coriobacteriia*) and 15 other non-*Coriobacteriia* strains, highlighting that the activity is conserved in the *Eggerthellaceae* clade but not found in more distantly-related members of the human gut microbiota (**Figure 3.4a**).

Next, we sought to use comparative genomics to determine if gene presence/absence was associated with P-gp inhibition. Using ElenMatchR, a comparative genomic tool we previously developed to identify functional genes in *E. lenta*[27], we identified 36 genes shared among all 33 P-gp inhibitory *Coriobacteriia* but absent in *Olsenella uli* (**Figure 3.4a**). From these 36 hits, 7 genes were absent from the genomes of all 15 non-inhibitory non-*Coriobacteriia* organisms (**Figure 3.4a**). All 7 of these hits were transcriptionally active in *E. lenta* DSM2243. Notably, two genes predictive of P-gp inhibition were found within the same genomic locus and annotated as impacting molybdate transport: *modA* (molybdate-binding protein) and *modB* (molybdenum transport system permease protein; **Figures 3.4b,c**). This suggests molybdate, a redox-active metal that complexes with diverse bacterial enzymes[28,29], is important for the biosynthetic pathways that *Coriobacteriia* use to produce the P-gp inhibitors.

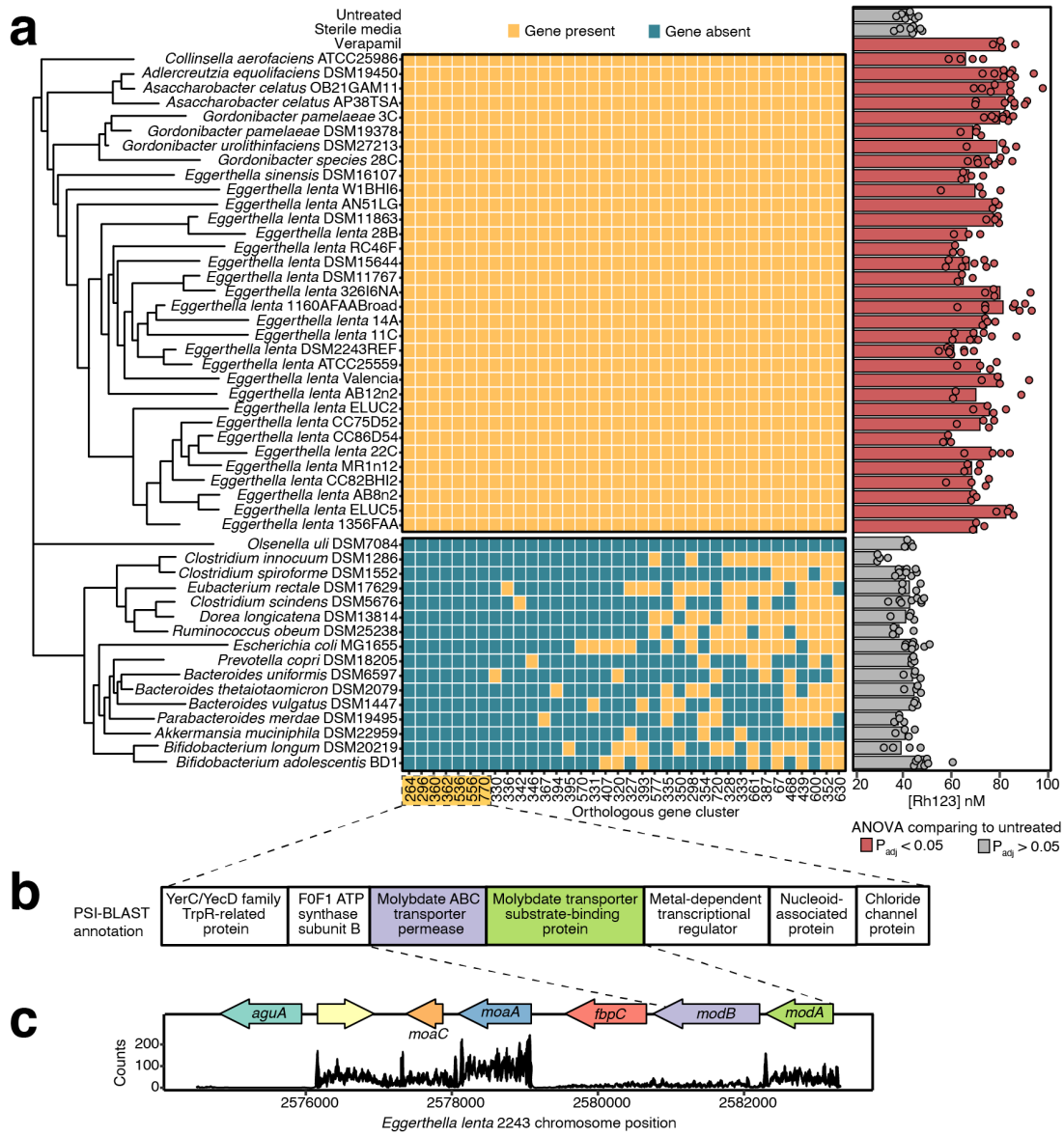


Figure 3.4. Comparative genomics reveals genes linked to P-gp inhibitory activity. **(a)** Identification of genes shared among strains with P-gp inhibitory activity assessed via Rh123 accumulation assay. Strains cultured in BHI^{++} were considered P-gp inhibitory if Rh123 accumulation is significantly higher than the untreated control. Sterile BHI^{++} media and 10μM verapamil were included as negative and positive controls, respectively (n=3-8/strain, ANOVA with Tukey's correction). ElenMatchR and NCBI BlastX were used to identify genes unique to PGP-inhibitory bacterial strains. A cladogram shows the phylogenetic relationships between all profiled strains. **(b)** Functional annotation of 7 putative genes that were uniquely present in P-gp inhibitory strains. **(c)** Locus diagram of *modA* and *modB* conserved genes in the *E. lenta* DSM2243 genome. Counts of transcripts from *E. lenta* DSM2243 mapped to the genome are shown.

3.2.5 Polar metabolites from *E. lenta* CFS inhibit P-gp

Given that the P-gp inhibitory activity was detected in the CFS (**Figure 3.2a**), we used activity-guided fractionation to enrich for and better understand the chemical nature of the *E. lenta* P-gp inhibitor (**Figure 3.5a**). To narrow down the extracellular products, *E. lenta* CFS was filtered through a 3kDa membrane. Both size fractions were tested in T84 and Caco-2 cells, which showed the activity to be in a fraction containing small molecules <3kDa in size **(Figure 3.5b)**. When the 3kDa filtrate was separated into polar and nonpolar fractions via liquid-liquid extraction, the activity remained in the polar fraction (**Figure 3.5c**). The active polar fraction was further separated via size exclusion chromatography and the activity was detected in fractions 3 and 4 (**Figure 3.5d**). These results indicate that *E. lenta* secretes one or more polar small molecules that inhibit P-gp.

The arginine-supplemented BHI media used to cultivate *E. lenta* is rich and undefined, potentially complicating further efforts to identify the active compounds via mass spectrometry[27,30]. To address this issue, we switched to a recently developed defined minimal media that supports robust growth of *E. lenta* (EDM)[30]. Consistent with our results in BHI, we noted significant P-gp inhibitory activity in *E. lenta* CFS following growth in EDM (**Supplemental Figure S3**). We then repeated our activity-guided fractionation steps (**Figure 3.5a**) and validated the activity in our T84 assay (**Supplemental Figure S3**). The biochemical fractionations not only reveal the chemical nature of the P-gp inhibitor but also enrich the metabolites that could be detected via mass spectrometry.

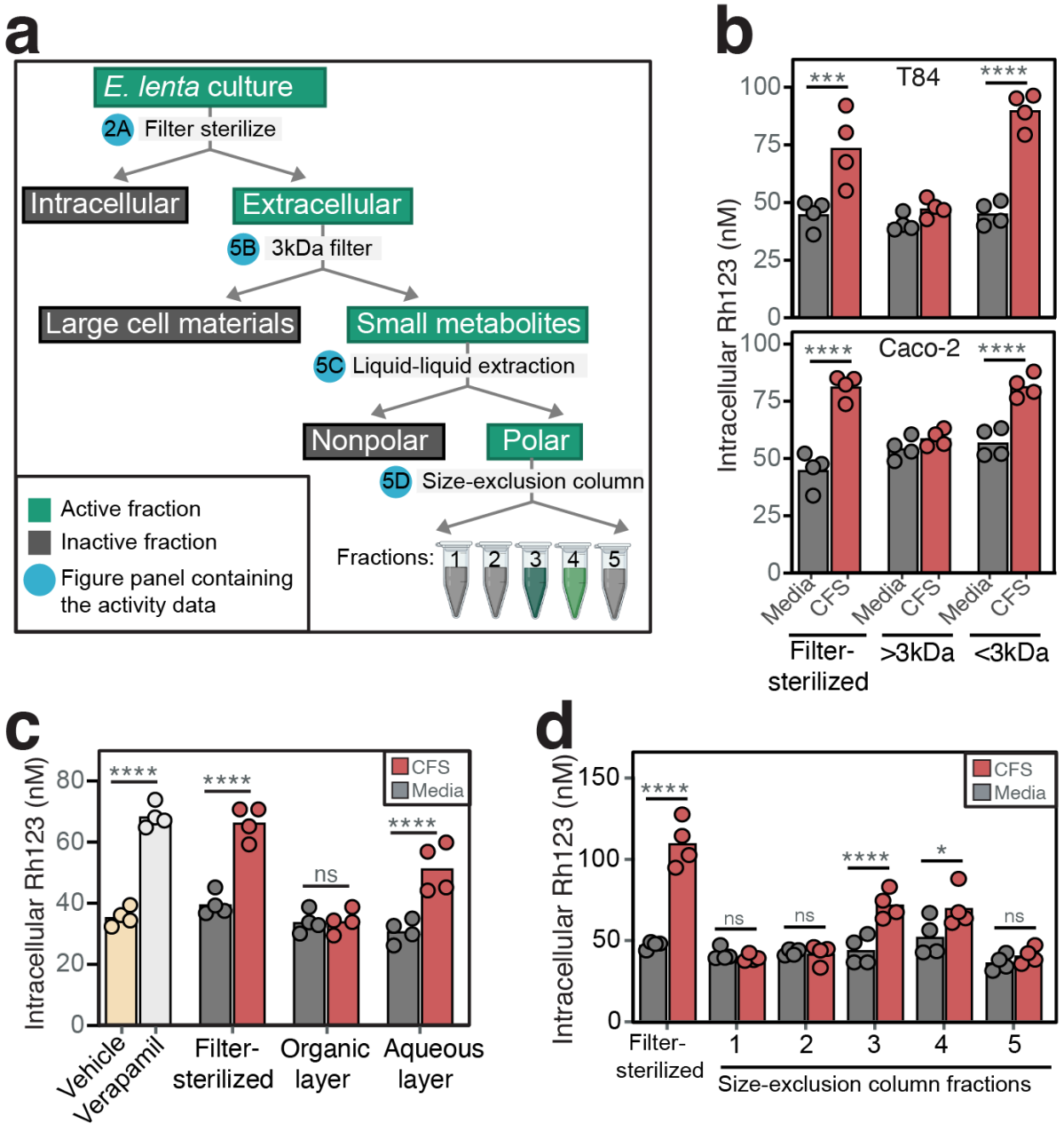


Figure 3.5. *E. lenta* secretes small polar metabolites that inhibit P-gp. (a) Activity-guided biochemical fractionation pipeline; the fractions are considered active when it leads to high Rh123 accumulation compared to the sterile media control. *E. lenta* CFS cultured in BHI^{+} was tested for activity in T84 cells after it was **(b)** filtered through 3kDa filter cartridge, **(c)** separated into polar and non-polar fractions via liquid-liquid extractions with MTBE:MeOH: H_2O, and **(d)** separated into 5 fractions via size exclusion chromatography. Verapamil and filter sterilized CFS were included as positive controls. Similar fractionations and activity testing using *E. lenta* CFS cultured in EDM are shown in **Supplemental Figure S3**. N=4/condition and 2-way ANOVA with Sidak's multiple testing correction. *<0.05, **<0.01, ***<0.001, ****<0.0001.

3.2.6 Untargeted mass spectrometry reveals isoflavonoids in P-gp inhibitory fractions

Finally, we used untargeted metabolomics to identify putative P-gp inhibitors in our enriched CFS fraction in two independent experiments. In the first experiment comparing *E. lenta* DSM2243 to the EDM sterile media control, there were 25 out of 175 total features (14.3%) detected only in the active fractions (**Figure 3.6a**). Molecular networking and annotations revealed that 56% of the features (14/25) unique to the active fractions were isoflavonoids (**Figure 3.6a**). The remaining annotated features in the active fractions were carboxylic acids and derivatives (3/25; 12%) and benzene and substituted derivatives (3/25; 12%) (**Figure 3.6a**). The final 20% (5/25 features) were unannotated.

To test whether similar metabolites are produced in other strains with P-gp inhibitory activity, we expanded our metabolomic analysis to 3 strains in an independent experiment, including a repeat of the type strain (DSM2243). We selected *E. lenta* 1356 and *E. lenta* W1BHI6 given that they inhibit P-gp to a similar degree as *E. lenta* DSM2243 (**Figures 3.4a; Supplemental Figure S4a**), grow robustly in the EDM media (maximum OD_{600nm}: 0.8-1.0), and are furthest from DSM2243 in the phylogeny of *E. lenta* (**Figure 3.4a**). The EDM media is selective for *E. lenta*, prohibiting us from including non-inhibitory control strains. In this experiment, 39 out of 221 total features were detected in all 3 strains but not in the media control, including 16 isoflavonoids and 9 carboxylic acids (**Supplemental Figures S4b-c**). A combined analysis of both experiments revealed 13 features present in all active fractions but absent in all media controls: 7 isoflavonoids, 2 carboxylic acids, and 4 unannotated features (**Figure 3.6b**). Benzene and benzene derivatives from the first experiment were not detected in the second experiment (**Figure**

3.6b). Thus, these results point towards metabolites in the carboxylic acid and/or isoflavonoid groups as putative P-gp inhibitors.

To further narrow down which chemical classes could explain the observed P-gp inhibition, we tested pure compounds from each group in our Rh123 assay. The carboxylic acids cluster, which contains both unique and shared features (**Figure 3.6a**), most likely reflects the biotransformation of arginine to other carboxylic acids. We selected the carboxylic acids citrulline, ornithine, and sarcosine for activity since these compounds are known by-products of the arginine dihydrolase pathway that *E. lenta* uses for growth[30,31]. Although benzene and derivatives were not uniquely present in CFS in the second experiment, we included benzoyl-coA, resorcinol, and phloroglucinol, which are benzene derivatives that bacteria can synthesize via the benzene degradation pathway[32]. None of the tested carboxylic acids or benzene derivatives exhibited any detectable P-gp inhibitory activity (**Figure 3.6c**).

Finally, we turned our attention to the isoflavonoid cluster, which represented the most common type of feature enriched in our active fraction relative to media controls. Isoflavonoids are a diverse family of natural bioactive products that have been previously shown to inhibit P-gp[33,34]. Incubating cells with acacetin and apigenin flavonoids led to high Rh123 accumulation in our assay (**Figure 3.6c**), suggesting that the related isoflavonoids produced by *E. lenta* are the most likely source of P-gp inhibition.

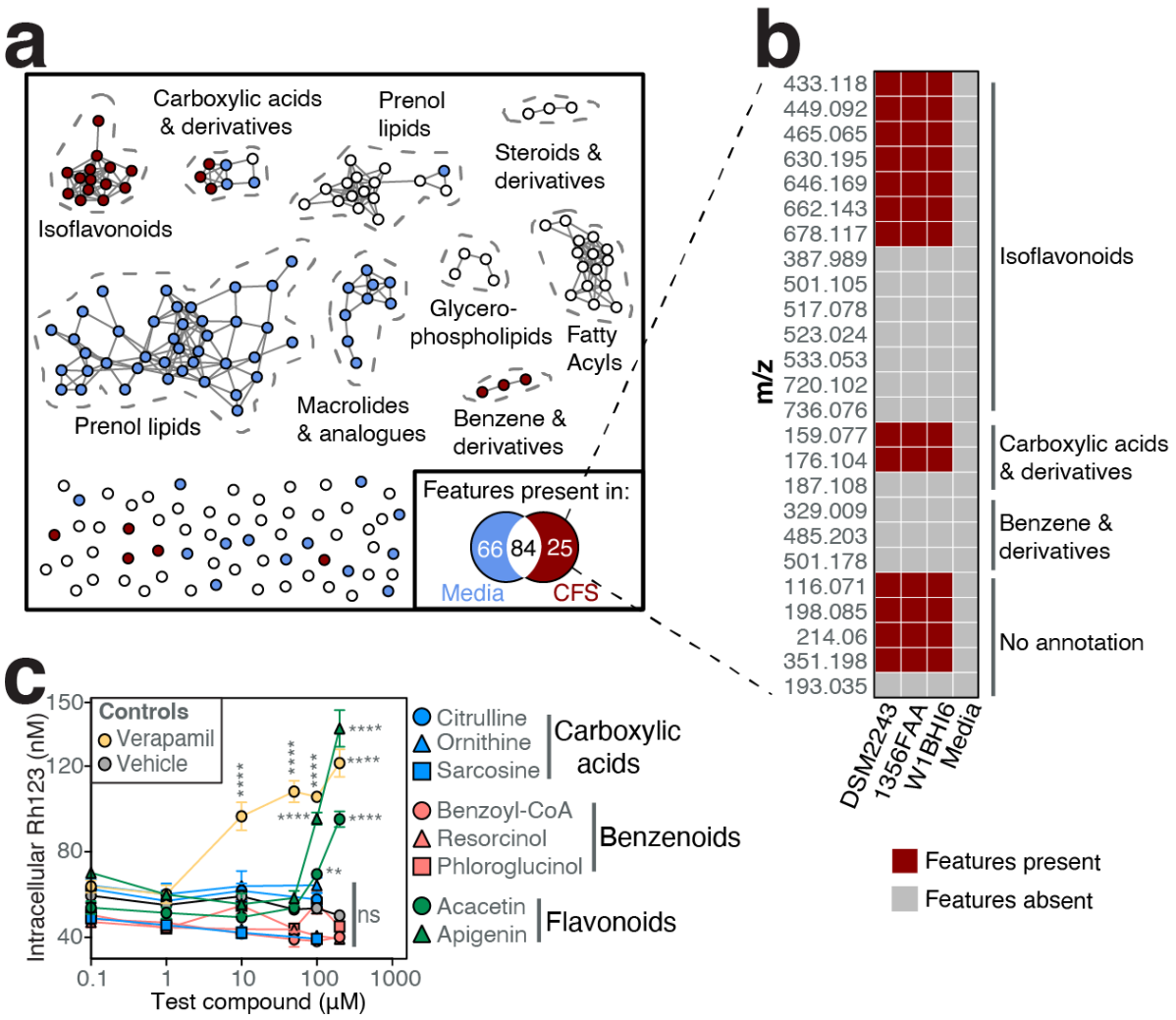


Figure 3.6. Discovery of a cluster of isoflavonoids produced by *E. lenta* linked to P-gp inhibition. Two independent untargeted metabolomics experiments were conducted (N=4/condition/experiment). In experiment 1, *E. lenta* DSM2243 CFS cultured in EDM along with sterile media control were fractionated using activity-guided fractionations (**Figure 3.5a**). Active fractions (**Supplemental Figure S3**) were analyzed via mass spectrometry. **(a)** Feature network of features detected only in DSM2243 CFS (red points) or sterile media control (blue points), or in both (white points). Inset shows the number of unique and shared features. In experiment 2, a similar untargeted metabolomics screen was expanded to 3 *E. lenta* strains (DSM2243, 1356FAA, and W1BHI6) (**Supplemental Figure S4**). 25 candidate features identified in experiment 1 were searched against strains in experiment 2. **(b)** Presence/absence heatmap of 25 candidate features in additional strains of *E. lenta*. **(c)** Dose-response curves of select carboxylic acids and derivatives, benzene and substituted derivatives, and flavonoids in T84 cells using the Rh123 accumulation assay (N=4/condition; ANOVA with Tukey's correction). 0μM concentrations were replaced with the pseudo-value of 0.1μM during log transformation.

3.3 Discussion

Our work uncovers how a prevalent gut Actinobacterium boosts absorption of digoxin, a cardiac glycoside with a narrow therapeutic window[19], by blocking a host drug efflux transporter. Mechanistic investigation revealed that *E. lenta* inhibits P-gp ATPase activity without affecting protein levels. The inhibitory activity was limited to the *Coriobacteriia* class and conserved across the *Eggerthellaceae* clade. Three distinct but complementary approaches, namely comparative genomics, activity-guided biochemical fractionations, and untargeted mass spectrometry, linked P-gp inhibition to genes involved in molybdate transport and metabolites in the isoflavonoid class.

Our results emphasize the importance of considering the vast biosynthetic potential within the human microbiome [9] for explaining inter-individual variations in drug absorption and metabolism. Seminal studies have shown that microbiome-derived bioactive molecules can directly interact with human cells or receptors (host-microbe interactions) or indirectly affect the host through impacting other members of the microbiome (microbe-microbe interactions)[9]. One example is the production of a metabolite structurally similar to the anti-cancer drug doxorubicin, a well-known P-gp substrate[9], that could compete for the absorption of other P-gp substrates. As the microbiota is highly diverse and dynamic, inter-individual genetic differences in the microbiome could account for variations in drug disposition[6]. By focusing on a prevalent member of the gut microbiota, this study demonstrates host-microbe interactions that resulted in altered drug absorption.

While bacterial metabolism of drugs is well-documented, with examples ranging from the bioactivation of the antibiotic prontosil to the re-activation of the anti-cancer drug

irinotecan[3], little is known about whether and how gut microbes could alter drug absorption without biotransformation. We found that the same gut bacterial species, *E. lenta*, can have opposing effects on drug bioavailability, contributing to the intestinal metabolism of digoxin [20,21] while also boosting drug absorption due to the inhibition of P-gp efflux. *E. lenta*, which colonizes the entire intestinal tract[27], could be inhibiting P-gp in the proximal small intestine, which is the major site of oral digoxin absorption[19]. The remaining digoxin in the gut lumen would then be subjected to bacterial metabolism in more distal regions; however, more work is needed to map out the specific site of the conversion of digoxin to dihydrodigoxin.

E. lenta increased digoxin absorption in both GF and CONV-R mice, which suggests that the ability to inhibit P-gp is not found in common members of the mouse gut microbiota. Interestingly, *E. lenta* is unique to humans, has a distinctive metabolic niche, and performs several unusual chemical transformations[30], indicating that the mechanism described herein may represent a human-specific host-microbe interaction relevant to pharmacology. This host species-specificity is also true for digoxin metabolism, as the bacterial genes that metabolize digoxin to dihydrodigoxin have only been detected in select *E. lenta* strains found in human gut microbiota[21].

Our comparative genomics results shed light on the genes necessary for P-gp inhibition by *E. lenta*. Using the clustering cutoff of protein identity <50%, we identified 7 transcriptionally active genes that were uniquely present in the P-gp inhibitory strains. Among these 7 genes were *modA* and *modB*, two genes involved in molybdenum transport[35]. Molybdenum, redox-active under physiological conditions, forms an active

site for many bacterial enzymes involved in diverse redox chemistry[29,36–38], potentially including isoflavonoid biosynthesis.

Although isoflavonoid biosynthetic pathways have been well-described in plants, little is known about *de novo* synthetic pathways in bacteria. Plants use phenylalanine ammonia-lyase and chalcone synthase enzymes to synthesize flavonoids from phenylalanine via the phenylpropanoid pathway[39]. Once thought to be restricted to plants, homologous enzymes were found in the sediment-derived Actinobacterium *Streptomyces maritimus*[39]. Biochemical studies revealed that *S. maritimus* produces flavonoids in a plant-like manner from phenylalanine[39]. Although we did not detect homologs of these specific genes in *Eggerthella*, our data demonstrate that human gut Actinobacteria can also produce isoflavonoids, similar to their soil-derived counterparts. More work is needed to elucidate the biosynthetic pathway(s) responsible.

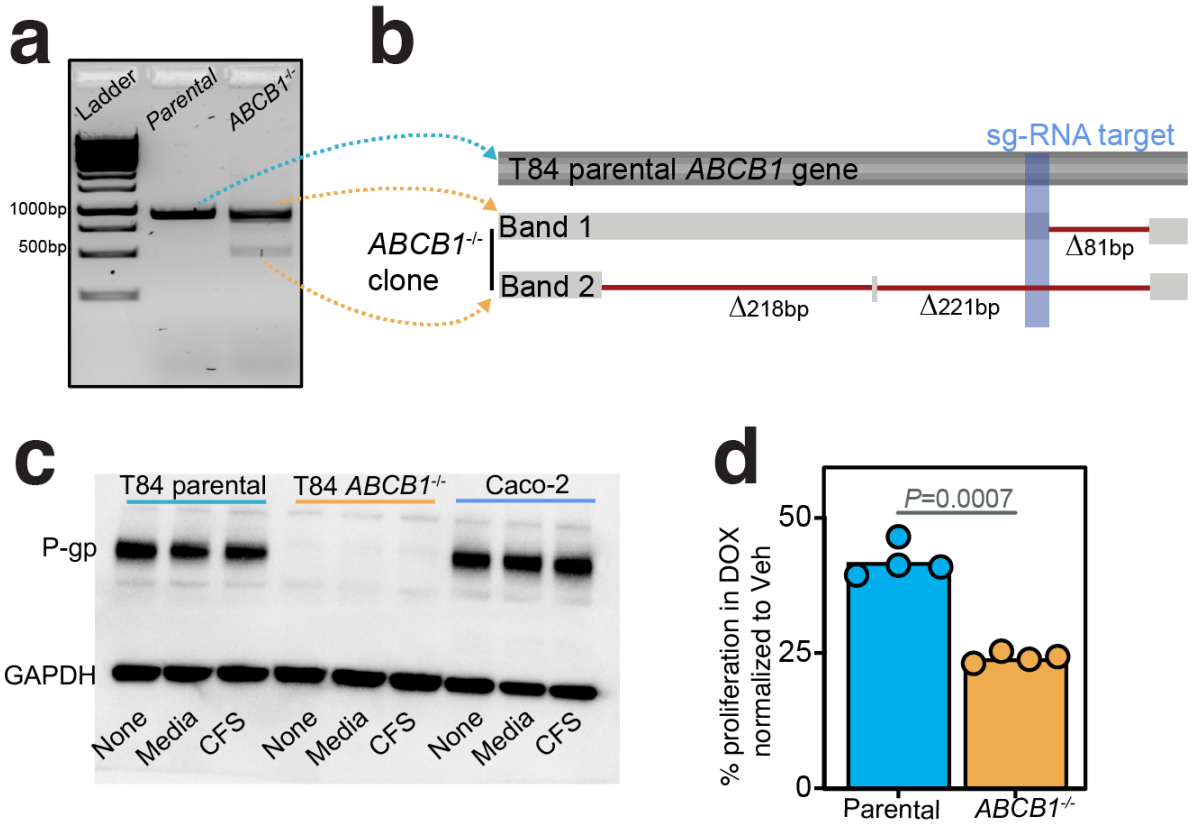
Flavonoids, typically found in fruits and vegetables, are a diverse class of natural products that exhibit numerous pharmacological properties, including anti-oxidative, anti-viral, anti-cancer, and anti-inflammatory activities[40]. Functional assays and structure-activity relationship analyses show that several flavonoids inhibit P-gp by binding to the ATP-binding site due to their structural similarity to the adenine moiety of ATP[40,41]. This mechanism is congruent with our ATPase activity assay in the vesicles.

There are multiple limitations of our study. *E. lenta*-mediated P-gp inhibition was tested in GF and CONV-R mice using the P-gp substrate digoxin; synthetic communities[42] or humanized mice[43] could be used to test how *E. lenta* affects P-gp function in the context of other members of the human gut microbiota. Although digoxin is a sensitive and physiologically-relevant P-gp probe, additional P-gp substrates could be used to assess

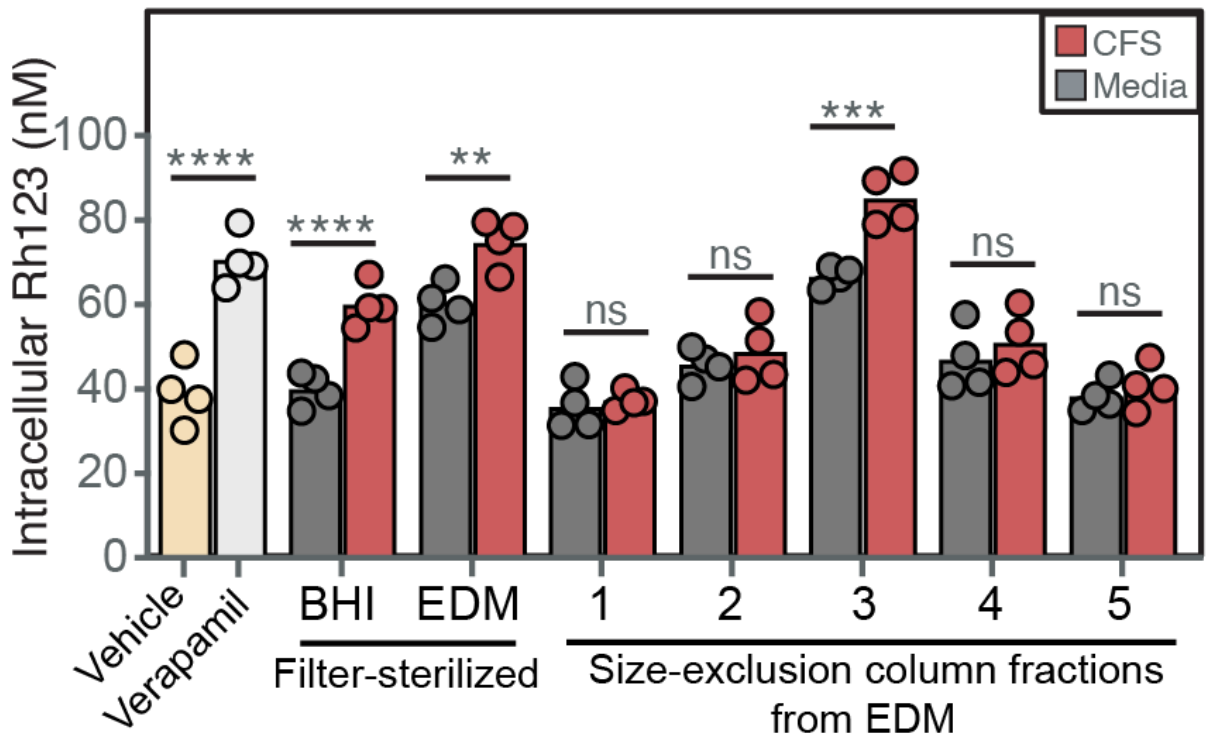
generalizability. Our comparative genomics analysis identified 7 candidate genes that have yet to be validated. Two of the genes identified, *modA* and *modB*, share <50% sequence identity to the genes with the same annotation in *E. coli*. Since these genes have never been studied in Actinobacteria, their functions may differ from canonical *modA* and *modB* functions. We did not make genetic knockouts in *E. lenta* since the genetic tools were not available until recently[23]. Through biochemical fractionations and metabolomics, we narrowed our candidate P-gp inhibitor to polar isoflavonoid compounds. Further purification efforts are required to isolate the metabolites and solve their structures, which would allow for in-depth biochemical characterization and dose-response experiments in mice.

Nonetheless, our study emphasizes the broad impact of the microbiome for multiple aspects of drug disposition, including metabolism[44] and absorption[7,8]. Continued mechanistic dissection of these and other host-microbe interactions relevant to the treatment of disease will help to explain the often-unpredictable inter-individual variations in drug efficacy and toxicity and enable a new generation of microbiome-aware medicines.

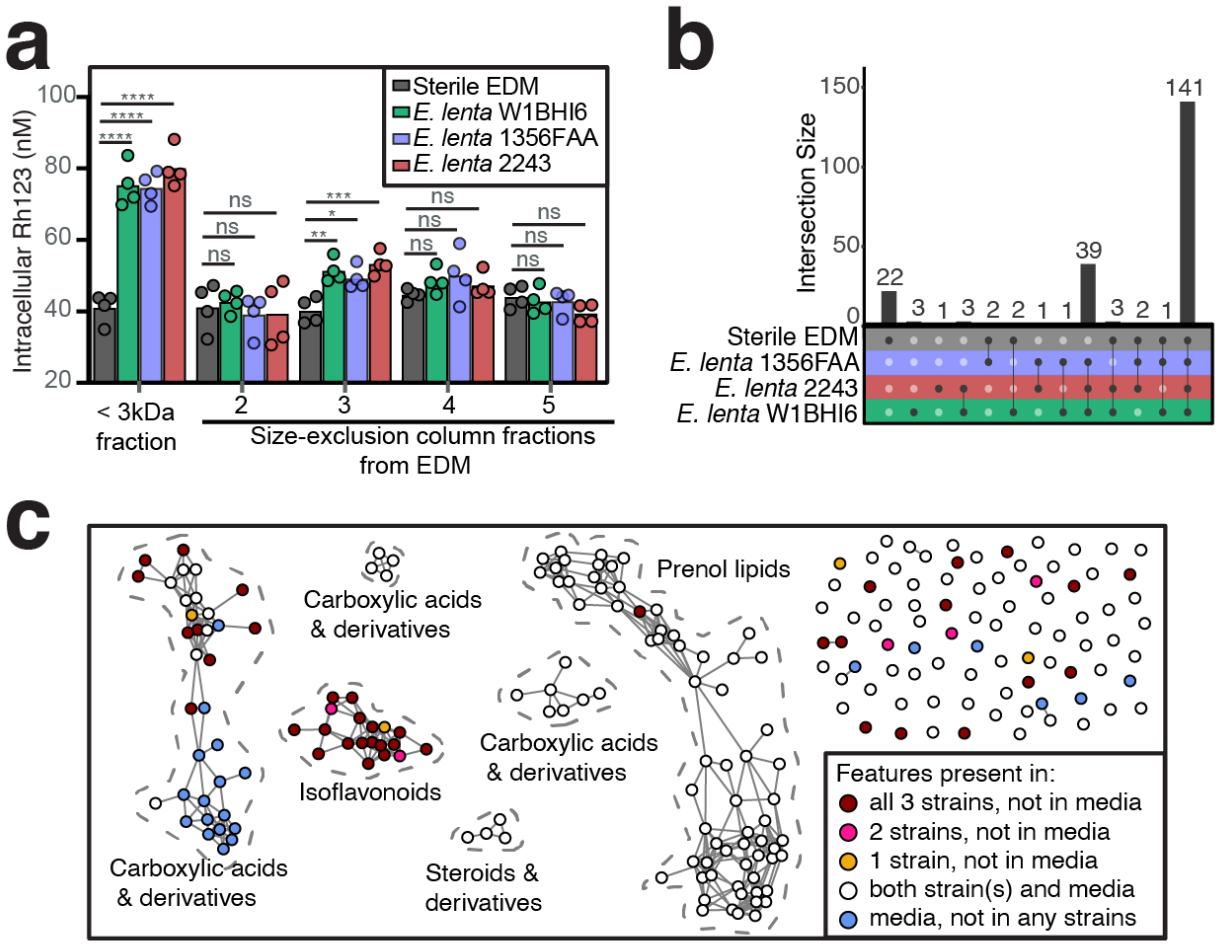
3.4 Supplemental Figures



Supplemental Figure S2, related to Figure 3.2. P-gp knockout monoclonal cell line generated from T84 parental cells is confirmed via sequencing, Western blot, and activity assay. (a) T84 cells were transduced with lentiviral particles carrying cas9 gene and small guide RNA (sg-RNA) against *ABCB1* gene. The transfected cells were sorted into a single cell per well and expanded to generate a monoclonal cell line. DNA was extracted and PCR amplified using primers flanking the sg-RNA target site. **(b)** The two visible bands in the gel were extracted, sequenced, and aligned to the parental T84 PCR product using NCBI multiple alignments tool. **(c)** Western blot of untreated (none), BHI^{+} media, or *E. lenta* CFS treated *ABCB1*$^{-/-}$ cell lysates. T84 parental cells and Caco-2 cells were included as controls. **(d)** P-gp effluxes doxorubicin (DOX) rendering cells with functional P-gp more resistant. Cells without functional P-gp are therefore more sensitive to DOX. Using these principles, P-gp function in *ABCB1*$^{-/-}$ cells was tested in a DOX sensitivity assay, and growth was normalized to DMSO vehicle (Veh) control (N=4/condition, Wilcoxon test).



Supplemental Figure S3, related to Figure 3.5. P-gp inhibitory activity is detected in *E. lenta* CFS cultured in EDM. *E. lenta* DSM2243 CFS from EDM cultures was processed through the activity-guided biochemical fractionation pipeline shown in **Figure 3.5a** (n=4/group; two-way ANOVA with Tukey's correction). Vehicle (0.02% DMSO) and 10 μM verapamil were included as controls. Filter-sterilized CFS cultured in EDM and BHI^{+} were included as controls for fractionations.



Supplemental Figure S4, related to Figure 3.6. Expanded metabolomics screen links isoflavonoids with P-gp inhibition. CFS of 3 *E. lenta* strains (DSM2243, 1354FAA, W1BHI6) cultivated in EDM was processed through the activity-guided biochemical fractionation pipeline shown in **Figure 3.5a** along with sterile EDM control. **(a)** Rh123 P-gp inhibition assay of each fraction (n=4/group; two-way ANOVA with Tukey's correction). **(b)** Shared and unique features in the active fractions detected via untargeted metabolomics. **(c)** Functional annotations and network of features.

3.5 Materials and Methods

Bacterial culturing. All the bacterial strains were cultured at 37℃ in an anaerobic chamber (Coy Laboratory Products) containing 2-5% H_2, 20% CO_2, and balance N_2 gasses for 48hr. There were 3 different media used: (1) brain heart infusion (BHI; VWR 90003-040) media supplemented with 1% arginine (denoted BHI^+), (2) BHI supplemented with 1% arginine, 0.05% L-cysteine-HCl, 5 µg/ml hemin, and 1 µg/ml vitamin K, and 0.0001% resazurin (denoted BHI^{++}), and (3) *Eggerthella* selective media (denoted EDM).

Mammalian cell culture. Human colorectal cancer cell lines (T84, HCT-15, HCT-116, and Caco-2) were obtained from American Type Culture Collection and HEK-293T cell line was a gift from the Mukherjee lab. These cell lines except for Caco-2 were routinely cultured in Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12 GlutaMAX (DMEM/F12) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin antibiotic solution. Caco-2 cells were cultured in DMEM, high glucose, pyruvate media with 10% FBS and 1% penicillin-streptomycin. Cells were grown in humidified incubators at 37℃ under 5% CO_2. Media was replaced 2-3 times a week and cells were split (1:5 for T84 and Caco2 and 1:10 for the other cell lines) using trypsin when they reached 80-90% confluence.

Rhodamine (Rh123) accumulation assay. Human colorectal cancer cell lines were cultured in 24-well tissue culture-treated plates. Once the cells reached 100% confluence in monolayers, they were incubated with 20% (unless otherwise noted) of CFS or sterile bacterial media control in DMEM/F12 with FBS for 18-20 hours. The media was then removed and the cells were incubated with fresh pre-warmed DMEM/F12 media containing fluorescent P-gp substrates (5µM Rhodamine 123) at 37℃ under 5% CO_2 for

30 minutes; 10μM verapamil hydrochloride, 20μM cyclosporin A, or 20μM vinblastine in 0.2% dimethyl sulfoxide (DMSO) were used as positive controls. After the media was removed, the cells were gently washed with 2mL of ice-cold phosphate buffer saline 3 times on ice. The final wash was carefully aspirated. The cells were then lysed with 300μL of RIPA buffer for 30 minutes and fluorescence was quantified using 100μL of supernatant lysate on the Biotek microplate reader (Tecan). To test P-gp inhibition with other substrates, the same procedures were performed with 50μM doxorubicin or 5μM Calcein-AM dye instead of 5μM Rhodamine 123.

MTT proliferation assay. T84 cells were seeded at 20,000 cells/well in 50μL in 96-well tissue culture-treated plates in DMEM/F12 supplemented with 10% FBS. To test whether *E. lenta* CFS sensitizes the cells against doxorubicin, the cells were incubated for 72 hours with 10% *E. lenta* DSM2243 CFS or sterile BHI^+ media control and treated with either 2μM doxorubicin or vehicle DMSO control; 10 μM verapamil was included as a positive control. After 3 days of incubation, the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-2H-tetrazolium bromide (MTT) viability assay was performed according to manufacturer's instructions (Promega). Briefly, 15 μL MTT dye solution was added to the wells and incubated at 37℃ under 5% CO_2 for 4 hours. An aliquot (100 μL) of solubilization solution was added and incubated for 1 hour at room temperature. Cell growth was assessed by measuring absorbance at 570 nm.

Differential expression via qPCR. Human colorectal cancer cell lines were cultured in 24-well tissue culture-treated plates. Once the cells reached confluence, 20% cell-free supernatant (CFS) or sterile bacterial media controls were added for 18-20 hours. The media was removed and RNA was extracted using the PureLink RNA Mini extraction kit

(Invitrogen), according to the manufacturer's instructions. Extracted RNA (5 µL of 150-400 ng/µL) was used to synthesize cDNA in 20µL reactions using iScript Reverse Transcription Supermix for RT-qPCR kit (Biorad). Diluted cDNA (1:20 in DNase-free water) was used to perform qPCR with SYBR Select Master Mix (Thermo Fisher Scientific) for 40 cycles. Human *ABCB1* exon-spanning primers (30 nM; 5'-CCATGCTCAGACAGGATGTGA-3' and 5'-ATCATTGGCGAGCCTGGTAG-3') were used to quantify P-gp expression along with human *GAPDH* (5'-GCTCTCTGCTCCTCCTGTTC-3' and 5'-GACCAAATCCGTTGACTCCG-3') and beta-actin primers (5'-ATGATGATATCGCCGCGCTC-3' and 5'-CCACCATCACGCCCTGG-3') as loading controls.

Western Blot analysis. Human colorectal cancer cell lines were cultured in 24-well tissue culture-treated plates. Once the cells reached 100% confluence in monolayers, they were incubated with 20% CFS or sterile bacterial media control for 18-20 hours. The media was removed and the cells were washed twice with pre-warmed PBS. The cells were lysed on ice with 100µl of RIPA lysis buffer containing a cOmplete Mini Protease Inhibitor Cocktail tablet (Millipore Sigma). Proteins in the lysate were quantified using Pierce BCA Protein Assay kit (Thermo Fisher Scientific) according to the manufacturer's instructions. After normalizing protein concentrations in RIPA buffer, 30µg of proteins in Laemmli sample buffer were loaded in a 10-well 4-15% Mini-PROTEAN gradient SDS-PAGE gel (Biorad) and ran for 15 minutes at 70V followed by 150V for 45 minutes in Tris-Glycine-SDS buffer at room temperature. Precision Plus Protein Dual Color standards were included in each gel. The proteins in the gel were then transferred to polyvinylidene difluoride membranes overnight at 23V at 4℃ in the N-cyclohexyl-3-

aminopropanesulfonic acid buffer. After blocking membranes in 5% milk, the membranes were incubated with C219 anti-ABCB1 primary antibody (Genetex) at 1:200 or anti-GAPDH primary antibody (Invitrogen) at 1:5000 in 5% milk in PBS-T blocking buffer overnight at 4℃. The membranes were washed 3 times and incubated with goat anti-mouse HRP-conjugated secondary antibody (Abcam) at 1:10,000 in a blocking buffer for 1 hour at room temperature. After washing the membranes, they were treated with Clarity Max Western ECL substrates (Biorad) for 5 minutes and visualized in the ChemiDoc (Biorad). The densitometry analysis was performed using the Biorad ImageLab software.

P-gp vesicle assay. To test whether *E. lenta* modulates P-gp ATPase activity, vesicles containing P-gp (5 mg/ml) purchased from Thermo Fisher were tested for ATP consumption in Pgp-Glo assays (Promega) according to the manufacturer's instructions. Briefly, the vesicles were thawed rapidly at 37℃ and immediately kept on ice. The vesicles were diluted to 1.25mg/ml with Pgp-Glo assay buffer before each assay. For each assay 20µl of 1.25mg/ml vesicles was mixed with 10-20µl of test agents (CFS, sterile media, 0.5mM verapamil, 0.25mM Na_3VO_4, or Pgp-Glo assay buffer control) in the presence of 10µl of 25mM MgATP in white flat-bottom 96-well plates. Pgp-Glo assay buffer was added if necessary to achieve the total volume of 50µl. The reaction was incubated for 1 hour at 37℃ under 5% CO_2.

After incubation, the reaction was diluted 1:4 (12.5µl into 37.5 µl of Pgp-Glo assay buffer) in a new white flat-bottom 96-well plate and 50µl of ATP detection reagents were added. Standard curves (3mM - 0.375mM MgATP at 1:2 dilution) in different matrices (CFS, sterile media, Pgp-Glo assay buffer) were included. The mixture was incubated at room temperature for 20 minutes and luminescence was quantified on a plate reader (Tecan).

Luminescence measurements were converted to ATP concentrations using matrix-specific standard curves.

Generation of P-gp knockout monoclonal cell line. *Escherichia coli* strains containing pCMV-dR8.91 and pMD2.G lentiviral packaging plasmids were a kind gift from the Mukherjee lab. An *E. coli* strain carrying the transfer plasmid (pCas9) with cas9 gene, eGFP, neomycin, and ampicillin marker genes and a separate *E. coli* strain carrying the transfer plasmid (pABCB1) with sgRNA against human *ABCB1* gene, mCherry, puromycin, and ampicillin maker genes were purchased from Genecopoeia. To harvest the plasmids, these strains were cultured separately in LB media supplemented with 50μg/ml carbenicillin aerobically overnight with shaking at 37℃. The resulting turbid cultures were centrifuged to collect the cell pellets, which were then processed with Midi-Prep plasmid extraction kit (Qiagen) based on the manufacturer's protocol. Plasmids were confirmed via enzymatic restriction digest.

Plasmids (4 μg of pCMV-dR8.91, 1 μg of pMD2.G, and 5 μg of either pCas9 or pABCB1) were added to a tube containing 30 μL of LT1 transfection reagent (Mirus) and 1 mL of Opti-MEM I Reduced Serum medium. The contents were vortexed and incubated at room temperature for 30 minutes. The mixture was added drop-by-drop into a 10cm cell culture dish containing HEK293T cells at 90% confluency. The cells were incubated for 48 hours and the supernatant containing recombinant lentiviral particles was filtered using 0.45μm filters. The lentiviral filtrate was flash-frozen.

To transduce the cells, 0.5 ml of lentiviral supernatant and 0.5 ml of 20 μg/ml polybrene transfection agent in cell culture media were added to T84 cells cultured in 12-well plates at 70% confluency. After 72 hr of incubation, 10 μg/ml of puromycin and 3 mg/ml of

geneticin antibiotics were added to select for successfully transduced cells. After 10 days of antibiotic selection, GFP^+/RFP^+ double-positive cells from the resulting polyclonal pool were sorted individually into 96-well plates using a cell sorter at the Flow Cytometry Core. The sorted cells were expanded into monoclonal cell lines, which were tested for *ABCB1* deletion via PCR, Western blot, and functional rhodamine accumulation assay.

Comparative genomics and associated transcriptomics. Comparative genomics among *Coriobacteriia* was performed using ElenMatchR[27]. Genes were clustered at a minimum of 50% amino acid identity and 80% query coverage to allow for identification of orthologs across genus boundaries. Hits from ElenMatchR were compared to the genomes of non-Coriobacteriia non-inhibitory strains using NCBI BlastX with identical cutoffs for amino acid identity and query coverage and a liberal E value threshold of 10 to ensure that all identified gene clusters were unique to P-gp inhibitory bacterial strains. Gene annotations were called using *E. lenta* DSM2243 genes corresponding to each orthologous gene cluster hit. Raw *E. lenta* DSM2243 RNA sequencing data in BHI was obtained and filtered and aligned as described[36]. Genes with RPKM >10 in all 3 replicates were called transcriptionally active.

Activity-guided biochemical fractionations. *E. lenta* DSM2243 cultures were grown anaerobically at 37℃ in BHI^+ or EDM media with sterile media control for 48 hours. To test whether *E. lenta-derived* P-gp inhibitor was intracellular or extracellular, cell pellets and supernatants were separated via centrifugation to test for P-gp inhibitory activity using our Rh123 assay. The pellet was reconstituted in 15 mM HEPES buffer (pH 7.3), lysed using bead beating with Lysing Matrix E tubes (MP Bio) in a Mini Beadbeater 16 Disrupter 16 (BioSpec Products), and tested for activity. The supernatant was pH-

adjusted to 7.6 with NaOH or HCl, filter-sterilized using 0.22 μm filters to create CFS, and tested for activity. To test whether the inhibitor was large cellular materials or small metabolites, CFS was further separated using 3kDa filter cartridges. Large materials on the filter membrane were reconstituted in sterile media in the same volume. The flowthrough and reconstituted large materials were then tested for activity. To further separate the small metabolites based on polarity, 4 ml of 3kDa flowthrough was added to 12 ml of M1 solvent (75% MTBE and 25% MeOH) and 8 ml of M2 solvent (25% MeOH in water), vortexed for 1 minute, and centrifuged at maximum speed for 5 minutes at 4℃. The polar aqueous layer was lyophilized down to 2 ml while the non-polar organic layer was dried under a gentle stream of nitrogen gas in TurboVap. Their activity was tested after reconstituting in 0.5x volume of water to account for incomplete extractions and compound loss through prior steps. To fractionate based on size, 1 ml of active polar fraction was loaded to G10 size exclusion midi-columns. Once the solution was absorbed into the column, water was added on top gently without disturbing the beads and allowed to flow through via gravity. One ml fractions were collected and tested for activity.

Untargeted metabolomics. Untargeted liquid chromatography high-resolution mass spectrometry (LC-HRMS) analysis was performed on a Sciex Exion UPLC equipped with a BDS C8 hypersil column (Thermo, 150 x 4.6 mm, particle size 5μm) and coupled to a Sciex TripleTOF 6600+ operated in positive mode electrospray ionization with IDA acquisition of MS2 (accumulation time 0.25 s, collision energy 35V (15V spread), acquisition window 100-2000 Da). A gradient of 0-100% acetonitrile over 14 min was used for compound separation.

Feature-based molecular networking analysis (FBMN)[45] was performed on shared ms/ms fragments across all metabolite features to determine the relationship between metabolite features detected in our HPLC fraction mixtures. Raw mass spectrometry data were preprocessed using MzMine2[46,47] for feature detection, chromatogram building, peak deconvolution, isotope grouping, gap filling, and alignment. The resulting intensity-based metabolite feature matrix and MS2 mgf files were used for further statistical and network analyses in Metaboanalyst[48] and GNPS[49]. Following FBMN, Network Annotation Propagation (NAP)[50], MS2LDA[51], Dereplicator[52], and MolNetEnhancer[53] were used to assign chemical classifications to the metabolite features. To summarize, a feature-based molecular network was created using the online workflow (https://ccms-ucsd.github.io/GNPSDocumentation/) on the GNPS website (http://gnps.ucsd.edu). MS/MS spectra were window filtered by choosing only the top 6 fragment ions in the +/- 50Da window throughout the spectrum. The precursor ion mass tolerance was set to 2.0 Da and an MS/MS fragment ion tolerance of 0.5 Da. A network was then created where edges were filtered to have a cosine score above 0.7 and more than 6 matched peaks. Further, edges between two nodes were kept in the network only if each of the nodes appeared in each other's respective top 10 most similar nodes. Finally, the maximum size of a molecular family was set to 100, and the lowest scoring edges were removed from molecular families until the molecular family size was below this threshold. The spectra in the network were then searched against GNPS' spectral libraries. The library spectra were filtered in the same manner as the input data. All matches kept between network spectra and library spectra were required to have a score above 0.7 and at least 6 matched peaks. The molecular networks were visualized in R.

Quantification of serum digoxin. To quantify serum digoxin from the pharmacokinetics experiments, 5 µl serum was combined with 100 µl of 10 ng/ml oleandrin internal standard and 5 µl of MeOH. For controls, 5 µl of 0-200 ng/ml (1:2 dilution series) digoxin/dihydrodigoxin mix in MeOH was added to 5 µl of blank serum along with 100 µl of 10 ng/ml oleandrin. To all tubes, 1 ml of 1:1 MTBE:ethyl acetate was added and vortexed for 2 minutes. After centrifuging for 10 minutes at 850 *xg*, the samples were dipped into the dry ice/isopropanol bath to freeze the aqueous layer. The liquid organic layer was transferred into new 2 ml tubes and dried to completion under a gentle stream of nitrogen gas at 10 psi at room temperature in a turbovap. The dried samples were reconstituted in 100 µl of 65% MeOH in water and stored at -20℃. The extracted samples were analyzed in two methods.

In method A, electrospray ionization (ESI) triple-quadrupole liquid chromatography-mass spectrometry in negative-ion mode with multiple reaction monitoring was used to detect digoxin, dihydrodigoxin, and oleandrin. Aliquots (10 µl) of the samples were injected onto a column (ODS Hypersil 4.6 x 100 mm, 5 µm particle size) using Shimadzu SIL-30AC injector set at 4℃. The column temperature was maintained at 50℃. Using 65% MeOH in water as a mobile phase in isocratic flow, each injection was run for 5 minutes. Digoxin, dihydrodigoxin, and oleandrin were expected to elute at 2.0 minute, 2.1 minute, and 3.6 minute, respectively. Expected Q1 - Q3 m/z masses were as follows: 779.4 - 475.2 for digoxin, 781.4 - 521.1 for dihydrodigoxin, and 575.2 - 531.1 for oleandrin. Peak areas of the standard curves were converted to concentrations at ng/ml using weighted linear regression ($1/x^2$).

In method B, the extracted samples were shipped on dry ice to the University of Washington Mass Spectrometry Center. Waters Xevo XS tandem mass spectrometry was used to detect digoxin, dihydrodigoxin, and oleandrin. An aliquot (10 µl) of the samples were injected into a column (Waters BEH C18, 2.1x50mm, 1.9 µm particle size) and ran for 7 minutes. Gradient mobile phase was used (A: 0.1% formic acid in H_2O and B: 0.1% formic acid in MeOH) at 50:50 for 0-4 minutes, 20:80 for 4-5 minutes, and 50:50 for 5.1-7 minute at 0.3 ml/min. A post-column infusion of 5 mM ammonium acetate was used at 10 µl/min. The expected retention times for digoxin, dihydrodigoxin and oleandrin are 2.75 min, 2.6 min, and 3.41 min respectively. NH_4-adducts for digoxin (parent m/z: 798.6; daughter m/z: 651.6) and dihydrodigoxin (parent m/z: 800.6; daughter m/z: 408.3) were monitored. Oleandrin was monitored at 577.4 m/z for the parent compound and 373.2 m/z for the daughter compound. Peak areas of the standard curves were converted to concentrations at ng/ml using linear regression.

Quantification of *E. lenta* in mice. Cecal content (0.1-0.6g) collected from mice at the end of the pharmacokinetics experiments was used to extract DNA using the ZymoBIOMICS DNA extraction kit according to the manufacturer's instructions. Extracted DNA (2 µL), 2 µL of water, 5 µL of iTaq Supermix reagents, and 1 µL of 2 µM *E. lenta*-specific primers with a FAM fluorophore probe (*elnmrk1*; 5'-GTACAACATGCTCCTTGCGG-3', 5'-CGAACAGAGGATCGGGATGG-3', and [6FAM]TTCTGGCTGCACCGTTCGCGGTCCA[BHQ1]) were amplified in 4 technical replicates using the following parameters for 40 cycles: initial denaturation at 95℃ for 5 minutes, denaturation at 95°C for 5 seconds, and anneal/extension at 60℃ for 15 seconds. A 10-fold dilution series of DNA from *E. lenta* DSM2243 culture was included to

construct a standard curve to convert amplification cycle (Cq) to genome equivalents using the formula that 1 ng of DSM2243 DNA = 2.7e5 genome equivalents.

Quantification and statistical analysis. All statistical analyses and data visualizations were performed in R v4.1.1 or PRISM 9.4.1. Statistical tests, sample size, standard error, and multiple testing correction methods are reported in the figures and figure legends.

Data and code availability. All data is available in the main text and the supplemental information and tables. Comparative genomics analyses mainly utilized the published ElenMatchR tool[27] and NCBI BlastX. Metabolomics datasets were analyzed on R; the script is available as a markdown HTML document on our lab GitHub (https://github.com/turnbaughlab). Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

3.6 References

1. Altman, R. B. PharmGKB: a logical home for knowledge relating genotype to drug response phenotype. *Nat. Genet.* **39**, 426 (2007).
2. Brunton, L., Chabner, B. A. & Knollmann, B. C. Goodman and Gilman's the pharmacological basis of therapeutics. Twelfth. Preprint at http://www.clivar.org/sites/default/files/pdf-goodman-and-gilmans-the-pharmacological-basis-of-therapeutics-t-laurence-brunton-bruce-chabner-bjorn-knollman-pdf-download-free-book-7e42e5f.pdf (2011).
3. Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R. & Goodman, A. L. Mapping human microbiome drug metabolism by gut bacteria and their genes. *Nature* (2019) doi:10.1038/s41586-019-1291-3.
4. Lam, K. N., Alexander, M. & Turnbaugh, P. J. Precision Medicine Goes Microscopic: Engineering the Microbiome to Improve Drug Outcomes. *Cell Host Microbe* **26**, 22–34 (2019).
5. Javdan, B. *et al.* Personalized Mapping of Drug Metabolism by the Human Gut Microbiome. *Cell* **181**, 1661–1679.e22 (2020).
6. Bisanz, J. E., Spanogiannopoulos, P., Pieper, L. M., Bustion, A. E. & Turnbaugh, P. J. How to Determine the Role of the Microbiome in Drug Disposition. *Drug Metab. Dispos.* **46**, 1588–1595 (2018).
7. Kyaw, T. S. & Turnbaugh, P. J. Tiny Gatekeepers: Microbial Control of Host Drug Transporters. *Clin. Pharmacol. Ther.* (2022) doi:10.1002/cpt.2647.
8. Zou, L. *et al.* Bacterial metabolism rescues the inhibition of intestinal drug absorption by food and drug additives. *Proceedings of the National Academy of Sciences* **117**,

16009–16018 (2020).

9. Sugimoto, Y. *et al.* A metagenomic strategy for harnessing the chemical repertoire of the human microbiome. *Science* **366**, (2019).
10. Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).
11. Hooper, L. V. *et al.* Molecular analysis of commensal host-microbial relationships in the intestine. *Science* **291**, 881–884 (2001).
12. Larsson, E. *et al.* Analysis of gut microbial regulation of host gene expression along the length of the gut and regulation of gut microbial ecology through MyD88. *Gut* vol. 61 1124–1131 Preprint at https://doi.org/10.1136/gutjnl-2011-301104 (2012).
13. Fu, Z. D., Selwyn, F. P., Cui, J. Y. & Klaassen, C. D. RNA-Seq Profiling of Intestinal Expression of Xenobiotic Processing Genes in Germ-Free Mice. *Drug Metab. Dispos.* **45**, 1225–1238 (2017).
14. Saksena, S. *et al.* Upregulation of P-glycoprotein by probiotics in intestinal epithelial cells and in the dextran sulfate sodium model of colitis in mice. *Am. J. Physiol. Gastrointest. Liver Physiol.* **300**, G1115–23 (2011).
15. Siccardi, D., Mumy, K. L., Wall, D. M., Bien, J. D. & McCormick, B. A. Salmonella enterica serovar Typhimurium modulates P-glycoprotein in the intestinal epithelium. *Am. J. Physiol. Gastrointest. Liver Physiol.* **294**, G1392–400 (2008).
16. Wang, Y. *et al.* Lactobacillus rhamnosus GG culture supernatant ameliorates acute alcohol-induced intestinal permeability and liver injury. *Am. J. Physiol. Gastrointest. Liver Physiol.* **303**, G32–41 (2012).
17. Foley, S. E. *et al.* Gut microbiota regulation of P-glycoprotein in the intestinal

epithelium in maintenance of homeostasis. *Microbiome* **9**, 183 (2021).

18. Saha, Saha, Butler, V., Neu, H. & Lindenbaum, J. Digoxin-inactivating bacteria: identification in human gut flora. *Science* **220**, 325–327 (1983).
19. Elmeliegy, M., Vourvahis, M., Guo, C. & Wang, D. D. Effect of P-glycoprotein (P-gp) Inducers on Exposure of P-gp Substrates: Review of Clinical Drug-Drug Interaction Studies. *Clin. Pharmacokinet.* **59**, 699–714 (2020).
20. Haiser, H. J. *et al.* Predicting and manipulating cardiac drug inactivation by the human gut bacterium Eggerthella lenta. *Science* **341**, 295–298 (2013).
21. Koppel, N., Bisanz, J. E., Pandelia, M.-E., Turnbaugh, P. J. & Balskus, E. P. Discovery and characterization of a prevalent human gut bacterial enzyme sufficient for the inactivation of a family of plant toxins. *Elife* **7**, (2018).
22. Alexander, M. *et al.* Human gut bacterial metabolism drives Th17 activation and colitis. *Cell Host Microbe* **30**, 17–30.e9 (2022).
23. Dong, X., Guthrie, B., Turnbaugh, P. & Balskus, E. Genome editing in an Actinobacterium Eggerthella lenta. *In preparation* (2022).
24. Ledwitch, K. V., Gibbs, M. E., Barnes, R. W. & Roberts, A. G. Cooperativity between verapamil and ATP bound to the efflux transporter P-glycoprotein. *Biochem. Pharmacol.* **118**, 96–108 (2016).
25. Yang, X. *et al.* Prevention of multidrug resistance (MDR) in osteosarcoma by NSC23925. *Br. J. Cancer* **110**, 2896–2904 (2014).
26. Duan, Z., Choy, E. & Hornicek, F. J. NSC23925, identified in a high-throughput cell-based screen, reverses multidrug resistance. *PLoS One* **4**, e7415 (2009).
27. Bisanz, J. E. *et al.* A Genomic Toolkit for the Mechanistic Dissection of Intractable

Human Gut Bacteria. *Cell Host Microbe* **27**, 1001–1013.e9 (2020).

28. Zhong, Q., Kobe, B. & Kappler, U. Molybdenum Enzymes and How They Support Virulence in Pathogenic Bacteria. *Front. Microbiol.* **11**, 615860 (2020).
29. Maini Rekdal, V. *et al.* A widely distributed metalloenzyme class enables gut microbial metabolism of host- and diet-derived catechols. *Elife* **9**, (2020).
30. Noecker, C. *et al.* Systems biology illuminates alternative metabolic niches in the human gut microbiome. *bioRxiv* 2022.09.19.508335 (2022) doi:10.1101/2022.09.19.508335.
31. Sperry, J. F. & Wilkins, T. D. Arginine, a growth-limiting factor for Eubacterium lentum. *J. Bacteriol.* **127**, 780–784 (1976).
32. Harwood, C. S., Burchhardt, G., Herrmann, H. & Fuchs, G. Anaerobic metabolism of aromatic compounds via the benzoyl-CoA pathway. *FEMS Microbiol. Rev.* **22**, 439–458 (1998).
33. Limtrakul, P., Khantamat, O. & Pintha, K. Inhibition of P-glycoprotein function and expression by kaempferol and quercetin. *J. Chemother.* **17**, 86–95 (2005).
34. Cui, J., Liu, X. & Chow, L. M. C. Flavonoids as P-gp Inhibitors: A Systematic Review of SARs. *Curr. Med. Chem.* **26**, 4799–4831 (2019).
35. Rech, S., Wolin, C. & Gunsalus, R. P. Properties of the periplasmic ModA molybdate-binding protein of Escherichia coli. *J. Biol. Chem.* **271**, 2557–2562 (1996).
36. Maini Rekdal, V., Bess, E. N., Bisanz, J. E., Turnbaugh, P. J. & Balskus, E. P. Discovery and inhibition of an interspecies gut bacterial pathway for Levodopa metabolism. *Science* **364**, (2019).
37. Bess, E. N. *et al.* Genetic basis for the cooperative bioactivation of plant lignans by

Eggerthella lenta and other human gut bacteria. *Nat Microbiol* **5**, 56–66 (2020).

38. Hille, R. The Mononuclear Molybdenum Enzymes. *Chem. Rev.* **96**, 2757–2816 (1996).
39. Moore, B. S. *et al.* Plant-like biosynthetic pathways in bacteria: from benzoic acid to chalcone. *J. Nat. Prod.* **65**, 1956–1962 (2002).
40. Palmeira, A., Sousa, E., Vasconcelos, M. H. & Pinto, M. M. Three decades of P-gp inhibitors: skimming through several generations and scaffolds. *Curr. Med. Chem.* **19**, 1946–2025 (2012).
41. Fox, E. & Bates, S. E. Tariquidar (XR9576): a P-glycoprotein drug efflux pump inhibitor. *Expert Rev. Anticancer Ther.* **7**, 447–459 (2007).
42. Cheng, A. G. *et al.* Design, construction, and in vivo augmentation of a complex gut microbiome. *Cell* **185**, 3617–3636.e19 (2022).
43. Turnbaugh, P. J. *et al.* The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci. Transl. Med.* **1**, 6ra14 (2009).
44. Spanogiannopoulos, P., Bess, E. N., Carmody, R. N. & Turnbaugh, P. J. The microbial pharmacists within us: a metagenomic view of xenobiotic metabolism. *Nat. Rev. Microbiol.* **14**, 273–287 (2016).
45. Nothias, L.-F. *et al.* Feature-based molecular networking in the GNPS analysis environment. *Nat. Methods* **17**, 905–908 (2020).
46. Pluskal, T., Castillo, S., Villar-Briones, A. & Oresic, M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**, 395 (2010).

47. Katajamaa, M., Miettinen, J. & Oresic, M. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics* **22**, 634–636 (2006).
48. Pang, Z. *et al.* MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res.* **49**, W388–W396 (2021).
49. Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828–837 (2016).
50. da Silva, R. R. *et al.* Propagating annotations of molecular networks using in silico fragmentation. *PLoS Comput. Biol.* **14**, e1006089 (2018).
51. Wandy, J. *et al.* Ms2lda.org: web-based topic modelling for substructure discovery in mass spectrometry. *Bioinformatics* **34**, 317–318 (2018).
52. Mohimani, H. *et al.* Dereplication of microbial metabolites through database search of mass spectra. *Nat. Commun.* **9**, 4035 (2018).
53. Ernst, M. *et al.* MolNetEnhancer: Enhanced Molecular Networks by Integrating Metabolome Mining and Annotation Tools. *Metabolites* **9**, (2019).

Chapter 4: Diet predicts the gut microbiome of colorectal cancer survivors

4.1 Introduction

Colorectal cancer (CRC) is the second-leading cause of cancer death in the United States, with 52,580 estimated deaths in 2022[1]. Health behaviors are a modifiable factor that may be associated with the risk of CRC mortality. For example, the American Cancer Society Nutrition and Physical Activity Guidelines for Cancer Survivors (hereafter referred to as the ACS guidelines) recommend avoiding obesity, being physically active, and consuming a healthy diet rich in fruits, vegetables, and whole grains[2]. Our team previously reported that adherence to these guidelines was associated with a 42% lower risk of death and longer 5-year survival among stage III colon cancer patients[3]. However, the mechanisms through which health behaviors may alter CRC survival are not well understood. We hypothesized that dietary components of the ACS guidelines act in part by changing the microbiome, motivated by the rapidly expanding literature on the role of the microbiome in cancer and the key role diet plays in shaping gut microbial community structure and function.

There is now a large and rigorous literature providing support for a causal role of the microbiome in both the etiology and treatment of CRC. Multiple procarcinogenic bacteria have been identified, including members of the *Fusobacterium nucleatum*[4], *Escherichia coli*[5], and *Bacteroides fragilis*[5,6] species. Broadly speaking, the mechanisms responsible include microbial effects on the host immune system[7] and the production of genotoxic metabolites[8,9]. Similarly, the extensive role of the microbiome in programming host immune function can modify the effects of cancer immunotherapy[10,11]. Furthermore, the tremendous metabolic diversity found within the human gut and tumor-associated

microbial communities can alter the metabolism and absorption of multiple anti-cancer drugs[12–15].

Diet is perhaps the most important modifiable factor that influences the gut microbiome[16,17]. Short-term dietary interventions in healthy individuals can lead to significant shifts in gut bacterial abundance and gene expression[18]. Dietary interventions in obese subjects have revealed marked changes in the gut microbiome in response to caloric restriction[19] and the consumption of a high-fat, low-carbohydrate ketogenic diet[20]. We and others have reported that diets rich in whole grains and fruits and vegetables were associated with a positive quality of life and lower recurrence in CRC patients and survivors[3,21,22]. However, data on the effects of dietary interventions on the gut microbiomes of CRC patients remain lacking, especially in the context of survivorship.

Here, we report the results of a pilot clinical study of 28 CRC survivors at the University of California, San Francisco (UCSF) to investigate whether adherence to the ACS guidelines was associated with inter-individual variations in the gut microbiomes of CRC survivors. We first calculated the ACS scores based on responses from the validated food frequency questionnaires. Then, we determined the covariates that were most predictive of the microbiome and assessed the specific aspects of the microbiome that were associated with the covariates.

4.2 Results

4.2.1 Dietary components of the ACS score predict variability in the gut microbiota

We contacted 163 CRC survivors at the University of California, San Francisco from 2019 through 2020. Of 45 (28%) individuals that expressed interest in participating in our study, we collected stool samples from 28 patients (**Figure 4.1a**). One outlier sample was excluded, as described above (see *Methods* and **Supplementary Figure S5a-d**). These 28 participants had a mean age of 52.7 (10.9) years, 57% were male, 79% identified as White, and 14% identified as Hispanic/Latino ethnicity. Nearly half of the participants (n=13; 48%) had an ACS score=3, with the other participants ranging from 4-6. There were no significant differences in the assessed demographic, clinical, or dietary factors between the sampled population and the population that was originally invited to participate in this study.

16S rRNA sequencing was used to profile microbial community compositions of stool from 27 CRC survivors after excluding one outlier sample (**Supplementary Figure S5a-d**). The overall ACS score was not significantly associated with the gut microbial community structure (**Figure 4.1c**). However, two components of the ACS score were statistically significant. This included (*i*) fruits and vegetables and (*ii*) alcohol, accounting for 19% and 13% of the variation seen in the distal gut microbiota, respectively (**Figure 4.1c**). The number of unique fruits and vegetables consumed was also significantly associated with the microbiota when modeled as a continuous variable (**Figure 4.1c**). Surprisingly, total fiber intake was not associated with the microbiota (**Figure 4.1c**). Similar results were obtained using a different beta-diversity metric (Bray-Curtis),

suggesting that these findings were robust and independent of the distance metrics used (**Supplementary Figure S6a**). These variables were the largest predictors of variability in the microbiome in our sample - more than clinical, demographic, or other behavioral variables. Other components of the ACS score, such as physical activity, BMI, red/processed meat intake, and whole grains consumption, were not associated with the microbiota (**Figure 4.1c**).

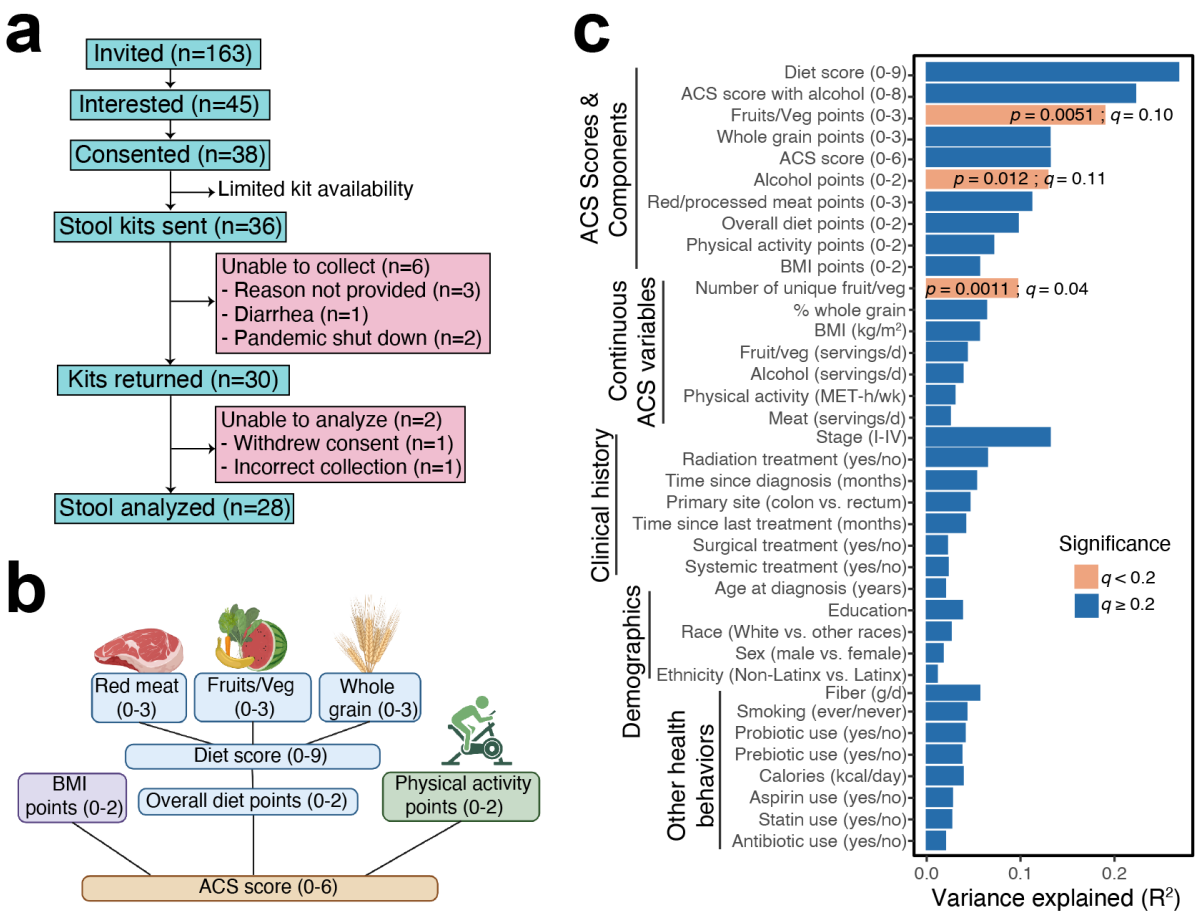


Figure 4.1. Components of ACS scores significantly account for variations in the microbiota of colorectal cancer survivors. (a) Derivation of our study population. **(b)** A schematic for calculating the American Cancer Society (ACS) score. When examining the ACS score with alcohol, the alcohol points (0-2) are added to the ACS score for a total range of 0-8. **(c)** PERMANOVA testing of ACS scores and components, clinical history, demographics, other health behavior variables, and continuous variables used to calculate ACS scores. Nominal *p*-values were reported from the ADONIS test using the weighted-UniFrac beta-diversity metric of bacterial community composition and false discovery rates *q* were calculated with Benjamini-Hochberg multiple testing corrections. A PERMANOVA test with a different beta diversity metric (Bray-Curtis) is shown in **Supplementary Figure S6a**.

4.2.2 Fruit and vegetable consumption is associated with the gut microbiome

Microbial community structure was significantly associated with the total fruit and vegetable points and variety of fruits/veg consumed, but not the total quantity (servings/day) (**Figure 4.2b**). Microbial diversity was positively correlated with all three metrics of fruits and vegetables consumed (fruits/veg points, serving size, and variety), although this relationship was not linear with respect to the total fruit and vegetable points (**Figure 4.2c**). At the phylum-level, the proportion of Firmicutes increased with higher fruit and vegetable intake, while Bacteroidota decreased with more intake **(Figure 4.2d,e)**. Differences in the abundance of three additional phyla were observed only with respect to fruits/veg points (**Figure 4.2d**). At the genus level, the abundances of four genera had a non-linear association with fruit/veg points (**Figure 4.2e**). A general linear model testing the association between the variety of fruit/veg consumed and the abundances of individual genera revealed two positively-correlated genera (*Fusicatenibacter* and *Erysipelotrichaceae*) and five negatively-correlated genera (*Eggerthella*, *Eubacterium*, *Actinomyces*, *Faecalitalea*, *Angelakisella*; **Figure 4.2g**).

To test if fruit and vegetable intake also accounted for the observed variations in the functional composition of the gut microbiome, PERMANOVA tests were used with the unstratified gene and pathway abundances. Similar to the taxonomic composition, fruit and vegetable intake accounted for 15% (p=0.018; **Figure 4.3a**) and 26% (p=0.011; **Figure 4.3b**) of the variations seen in gene and pathway abundances, respectively. Abundances of 21 gene families (q<0.2; **Figure 4.3c**) and 33/375 pathways (q<0.2; **Figure 4.3d**) were significantly associated with fruit and vegetable consumption. With the

exception of dTDP-N-acetylviosamine biosynthesis pathway, all of the pathways had a non-linear relationship with the fruit and vegetable points (**Figure 4.3d**).

Overall, fruit and vegetable consumption was associated with pathways for the synthesis of certain amino acids, nucleic acids, and short-chain fatty acids (SCFA), and the degradation of sugars (**Figure 4.3d**). Twenty one percent (7/33) of the significant pathways were involved in amino acid synthesis, such as arginine, ornithine, glutamine, putrescine, and chorismate, a precursor of aromatic amino acids and 4/33 (12%) pathways involved pyrimidine and adenosine biosynthesis. Bifidobacterium shunt, a pathway that breaks down hexose sugar to produce acetate and lactate[23], was significant. However, other pathways involved in the production of SCFA via fermentation of L-lysine, pyruvate, or acetyl-CoA had no detectable association with fruit/veg intake. Degradation of galactose and stachyose, sugars found in avocados and numerous vegetables[24], was also significantly associated with fruit and vegetable consumption.

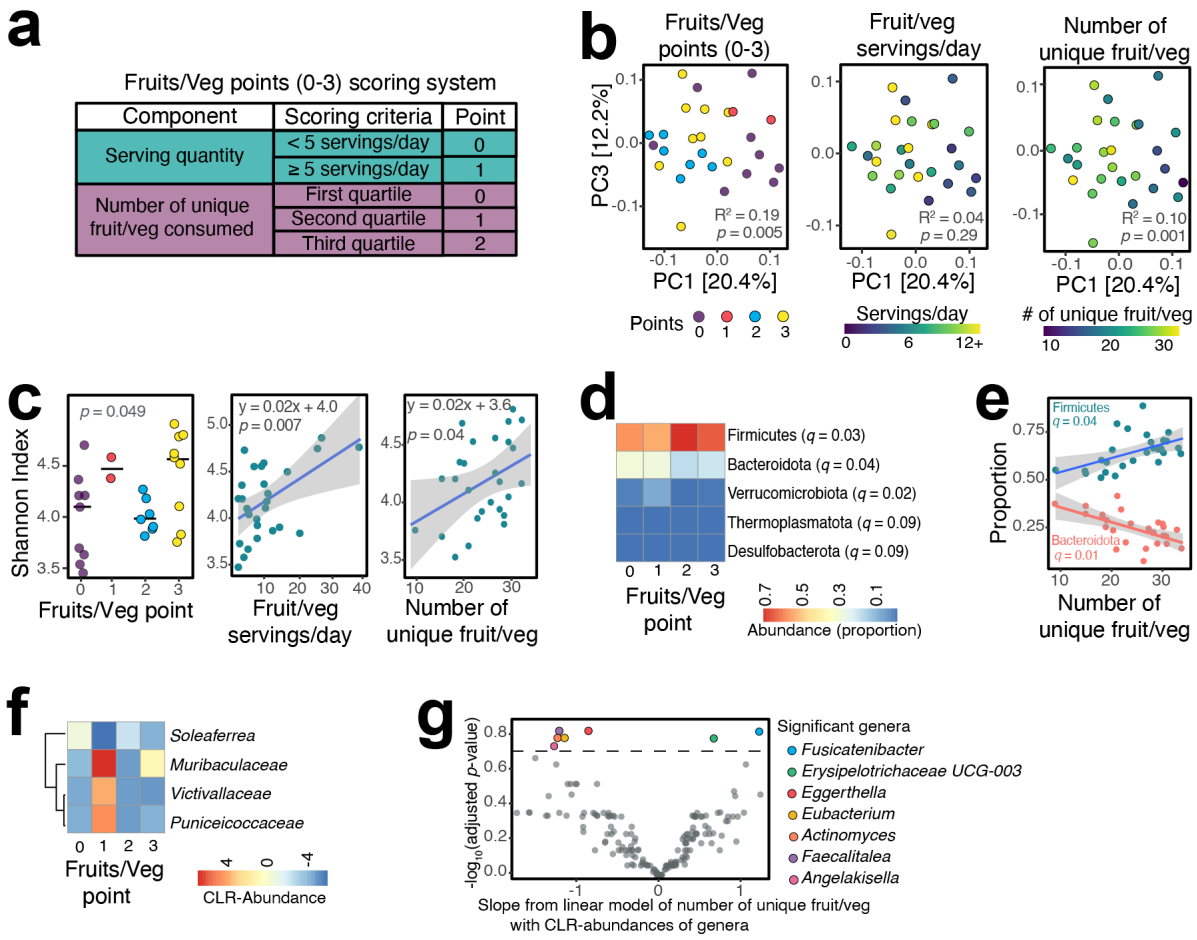


Figure 4.2. Fruit and vegetable intake is associated with gut microbial diversity and composition. (a) The Fruits and Vegetable score is calculated as 1 point for consuming 5 or more servings/day and 0-2 points for tertiles of the number of unique fruits and vegetables usually consumed in a year. **(b)** Principal coordinate analysis plots of weighted-UniFrac distance matrices of bacterial community composition colored based on fruits/veg points, servings, and the number of unique fruit/veg consumed. R^2 and *p*-values were extracted from PERMANOVA tests. **(c)** Alpha diversity analysis using Shannon diversity metric. **(d,e)** Proportions of phyla that were significantly (adjusted *p*-value $q < 0.2$) correlated with **(d)** fruit/veg points and **(e)** the number of unique fruits and vegetables consumed. **(f,g)** Centered-log-ratio (CLR) transformed abundances of genera significantly associated with **(f)** fruit/veg points and **(g)** the number of unique fruits and vegetables consumed. Unless otherwise stated, one-way ANOVA was used to test categorical variables (fruit/veg points) and continuous variables (servings, number of unique fruit/veg), respectively. Adjusted *p*-values (*q*) were calculated using Benjamini-Hochberg multiple-testing correction. The line and gray ribbons represent 95% confidence intervals.

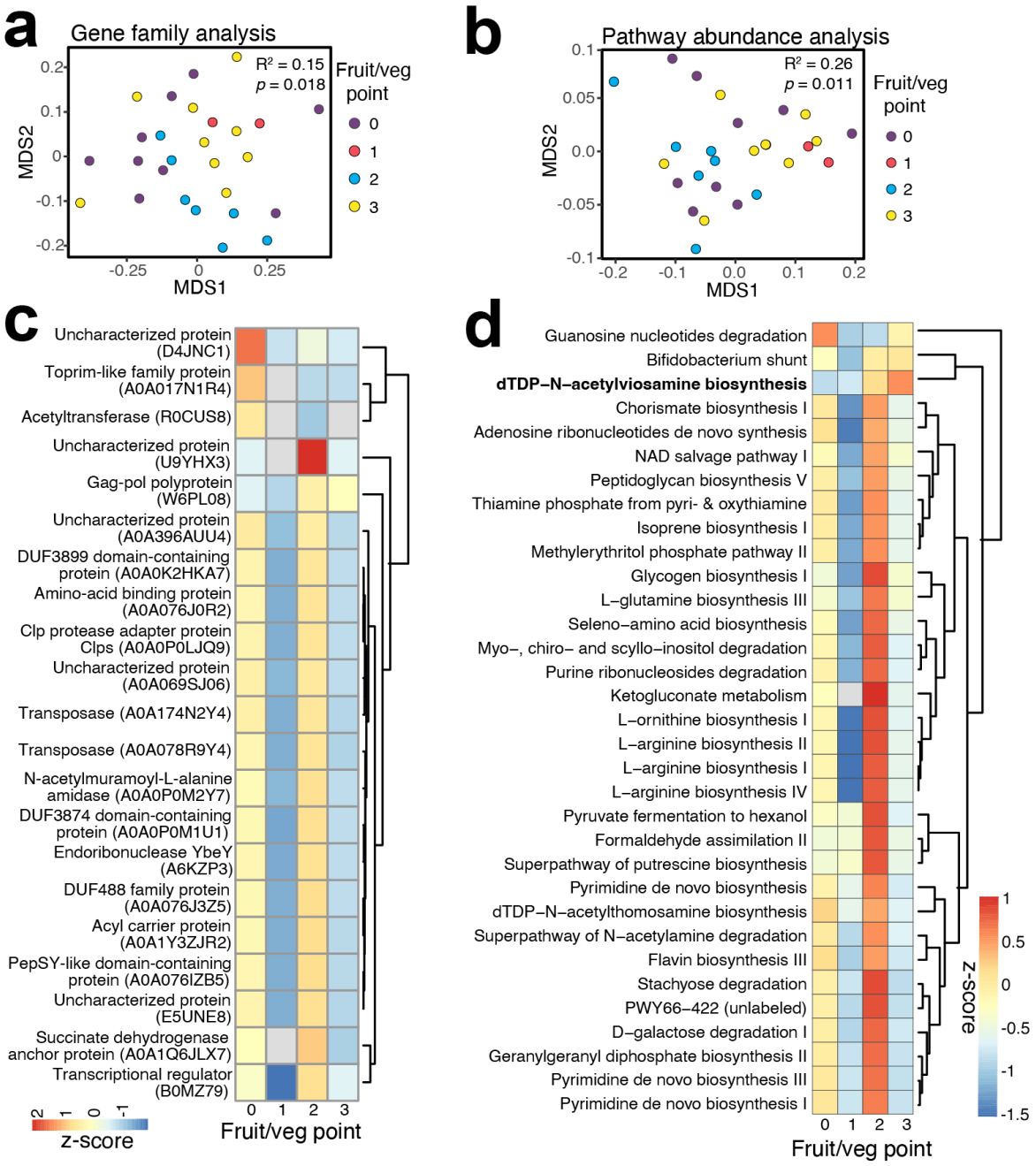


Figure 4.3. Fruit and vegetable intake is associated with microbial pathway abundances. (a,b) Non-metric multi-dimensional scaling of variations in **(a)** gene family-level abundances and **(b)** pathway abundances colored by fruit/vegetable points. R^2 and *p*-values were extracted from PERMANOVA tests. **(c,d)** Significant (false-discovery-rate $q < 0.2$; ANOVA test with Benjamini-Hochberg correction) **(c)** gene families and **(d)** pathways associated with fruit/veg points. Bolded pathways have a linear relationship. Z-score was calculated from average normalized abundances. The identifier in the parentheses refers to gene family entries in the UniRef90 database.

4.2.3 Alcohol intake predicts gut microbial species and gene abundance

Intake of alcoholic drinks was also associated with inter-individual variations in gut microbial community structure (p=0.01; **Figure 4.1c, Figure 4.4a**) and higher microbial diversity (p=0.03; **Figure 4.4b**). Alcohol intake negatively correlated with the abundance of the Bacteroidota phylum. Two other phyla (*Thermoplasmatota* and *Proteobacteria*) and one species-level taxon (*Clostridium innocuum*) were also significantly associated with alcohol use (**Figure 4.4c,d**). Notably, alcohol use and fruit and vegetable consumption were not correlated (**Supplementary Figure S6b**), suggesting they represent independent factors associated with variations in the microbiota.

To test whether alcohol consumption was linked to the observed variations in the functional composition of the gut microbiome, we first assessed microbial gene abundances. We used two statistical models: *(i)* ANOVA testing for significant differences in gene abundances among non-drinkers, light drinkers, and heavy drinkers, and *(ii)* a general linear model (GLM) that tests for a linear association between alcohol consumption and gene abundances. Of 13,533 gene families, 1,278 (9.5%) were significant by ANOVA (q_{ANOVA}<0.2), 2,812 (21%) were significant by GLM (q_{GLM}<0.2), and 1,061 (7.8%) were significant by both models (**Figure 4.4e**). Notably, more than 97% (1,031/1,061) of the significant gene families had negative correlations with alcohol consumption.

Similar analyses were conducted with the pathway abundances. Of 375 total pathways, 49 (13%) were significant by ANOVA (q_{ANOVA}<0.2), of which 6/49 (12%) had significant linear relationships (q_{ANOVA}<0.2; q_{GLM}<0.2; **Figure 4.4f**). Two pathways, lipopolysaccharide biosynthesis and 2-methylcitrate cycle involved in degrading a

common food and drink additive propionate, were positively associated with alcohol intake. In contrast, the abundances of 4 pathways involved in histidine degradation, cis-vacenate fatty acid biosynthesis, and pyrimidine salvage were negatively associated with alcohol use. Taken together, these results showed marked differences in taxonomic and functional compositions of the gut microbiome among CRC survivors linked to alcohol intake.

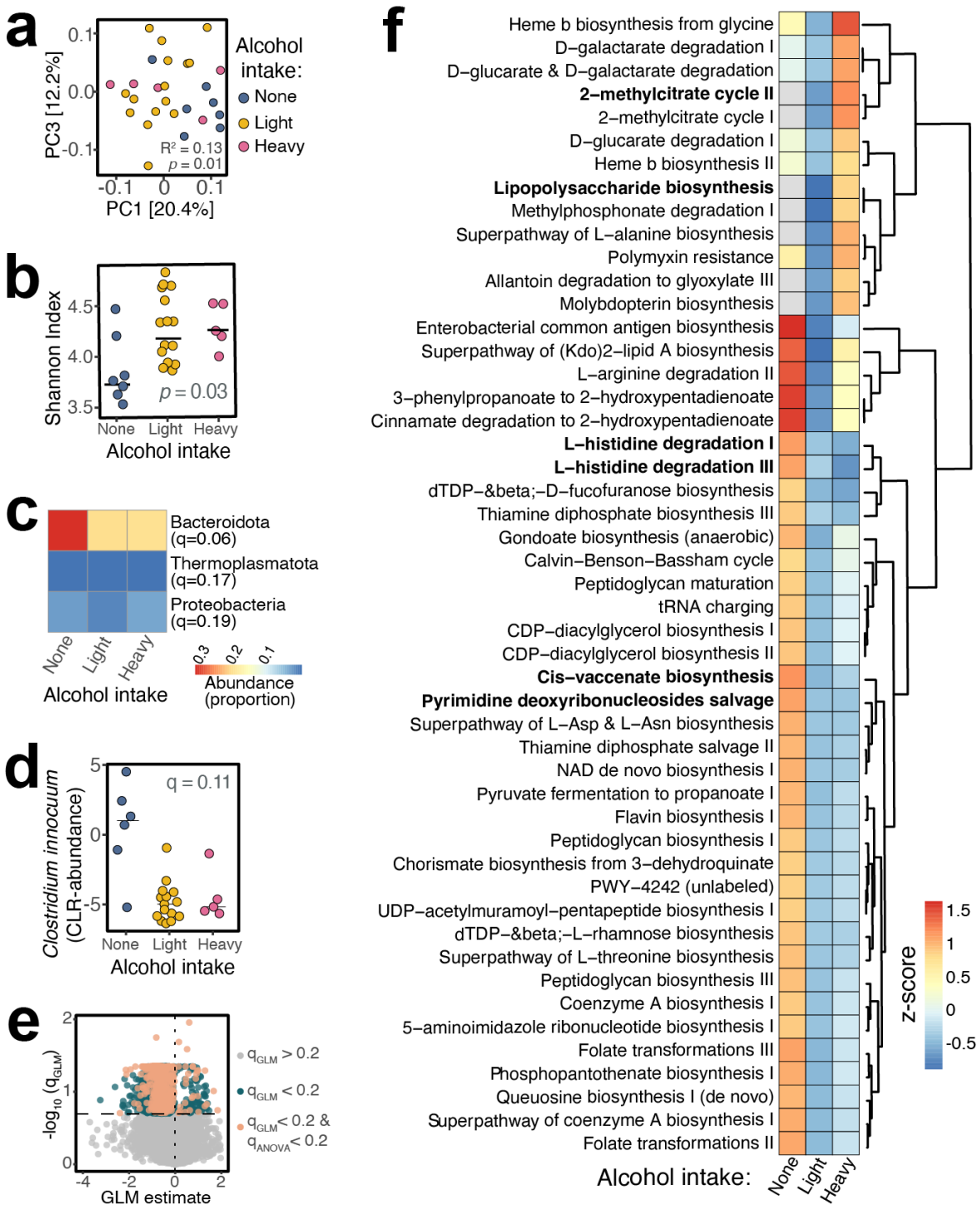


Figure 4.4. Alcohol consumption is linked to the gut microbiome of cancer survivors. (a) Principal coordinate analysis of weighted-UniFrac distance matrices of bacterial community composition from 16S sequencing colored based on alcohol intake. Light drinking is defined as ≤1 drink/day for women and ≤2 drinks/day for men; amounts above these cut-offs are labeled as heavy drinking. R^2 and p-values were extracted from

PERMANOVA tests. **(b)** Alpha diversity analysis using Shannon diversity index of community composition for alcohol intake. **(c)** The proportion of phylum-level abundances associated with alcohol intake. **(d)** Centered-log-ratio (CLR)-transformed abundance of a species significantly associated with alcohol intake. **(e)** A volcano plot of gene families tested for association with alcohol intake using two statistical models - a general linear model (GLM) and ANOVA. GLM tests for the linear relationship between normalized gene family-level abundances and alcohol intake. **(f)** Significant pathways associated with alcohol intake. Bolded pathways have a linear relationship based on GLM. Significance is defined as false-discovery-rate $q < 0.2$ using Benjamini-Hochberg multiple-testing correction.

4.3 Discussion

In this pilot study, we explored whether the ACS score or its components were associated with variations in the gut microbiome of 28 CRC survivors. While the overall ACS score was not associated with the microbiome, we found that fruit and vegetable intake and alcohol consumption were significantly associated with inter-individual variations in distal gut microbial diversity, community composition, and gene and pathway abundances. In fact, these two dietary components were the only covariates among other variables, such as clinical history, demographics, and health behaviors, that reached statistical significance in our dataset.

This study supports the feasibility of studying the gut microbiomes of CRC survivors. Despite exclusively recruiting via email a few months before the global COVID-19 pandemic began, we were able to enroll and obtain stool samples remotely from 28 CRC survivors. Importantly, these individuals did not differ from our invited population of CRC survivors in terms of demographic or clinical factors. However, incorporating other strategies such as reminder invitations, phone calls, and compensation in future studies may increase the recruitment rate.

All samples, except one, were primarily composed of Firmicutes and Bacteroidota (**Supplementary Figure S5c**), which are typically the two most abundant phyla found in the human gut microbiota[25], suggesting that the gut microbiome of CRC survivors is relatively intact. One outlier sample exhibiting a bloom of Proteobacteria and no Bacteroidota was collected from a patient that had an ileostomy, which is a surgically created opening in the abdominal wall through which the ileum is attached

(**Supplementary Figure S5c**). Given prior data that ileostomies markedly perturb the gut microbiota[26], we opted to exclude this sample from all analyses presented herein.

We observed a strong and significant association between fruit and vegetable consumption and multiple aspects of the gut microbiomes of CRC survivors. Consistent with our previous intervention studies in healthy individuals[18], we found a decreased abundance of Bacteroidota phylum in individuals that consume high levels of fruits and vegetables. Surprisingly, it was the number of unique fruits and vegetables consumed, not the quantity or fiber content of the diet, that was linked to variations in the gut microbiota. These results emphasize the need to broadly consider components of fruits and vegetables, in addition to dietary fiber[27], that may play a role in shaping the gut microbiome[28]. Eating a variety of fruits and vegetables provides diverse nutritional substrates that can support the growth of many microorganisms; for example, dietary lignans [29] or other polyphenols[30].

The fruit and vegetable score, which factors in the quantity and variety of fruits and vegetables consumed, was associated with amino acids, nucleic acids, cofactor, and SCFA biosynthesis and sugar degradations. Supported by the observation that herbivorous fecal microbiomes have enriched biosynthetic pathways for amino acids compared to their carnivorous counterparts[31], this may suggest that the gut microbes may not be able to extract sufficient amino acids and cofactors from fruits and vegetables. Enriched abundance of pathways for sugar degradations suggests the microbes are able to utilize galactose and stachyose found in fruits and vegetables. Surprisingly, microbial pathway abundances were non-linear with respect to fruit/veg points, which may suggest a certain combination of quantity and the number of unique fruits and vegetables triggered

the expansion of biosynthetic pathways. Future intervention studies would be helpful to determine the reproducibility and mechanistic basis for such non-linear effects on the gut microbiome, as we have recently demonstrated for ketogenic diets[20].

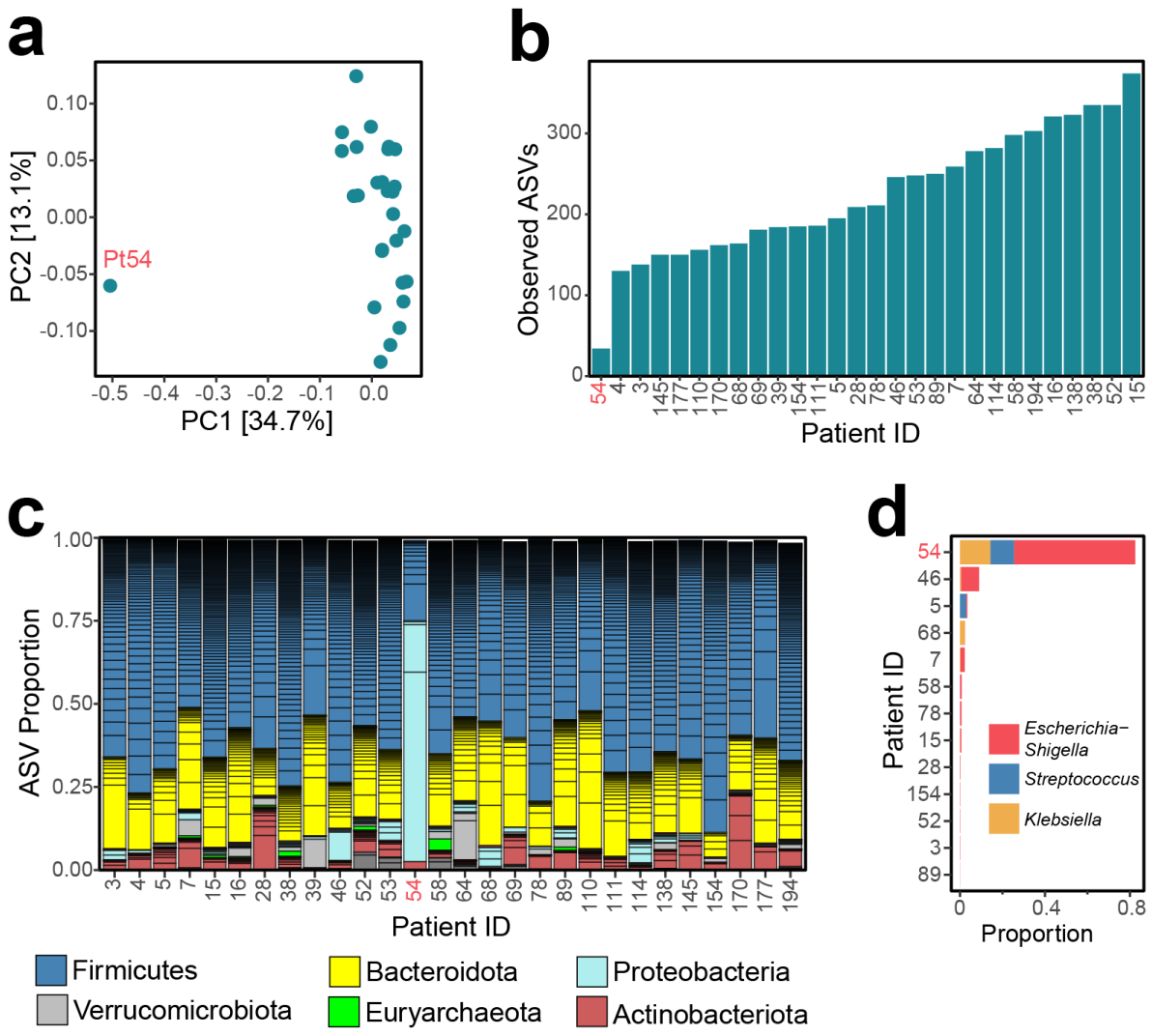
Alcohol is one of the strongest modifiable risk factors for cancer after tobacco use and excess body weight[32]. Heavy alcohol consumption is associated with a higher risk of several cancers, including CRC, but the effect of moderate drinking has conflicting reports[32]. Some studies considered moderate drinking to be optimal compared to non-drinking or heavy drinking in terms of CRC mortality while other studies considered non-drinking to be optimal[32]. In our study, we found that alcohol drinking was linearly associated with multiple aspects of the gut microbiome. The effect size of these associations is surprising given that ethanol is rapidly absorbed in the stomach and small intestine[33]. We propose that the other components of alcohol-containing drinks (*e.g.,* polyphenolic compounds in wine or hops in beer) are a more likely source of these differences. However, it is also possible that more indirect mechanisms due to changes in host physiology in response to alcohol somehow impact the microbiome. Future studies in mice aimed at deconvoluting these various factors coupled with studies of the downstream health effects of alcohol-associated differences in the gut microbiome in humans are needed.

Our study has limitations. We invited 163 individuals to participate and obtained stool samples from 28 (17%). A strength of our study is that the sampled individuals did not differ in terms of demographic or clinical factors from those who did not provide stool samples. However, the invited and sampled population had limited racial, ethnic, and socioeconomic diversity. Additionally, none of our participants had low ACS scores (0-2

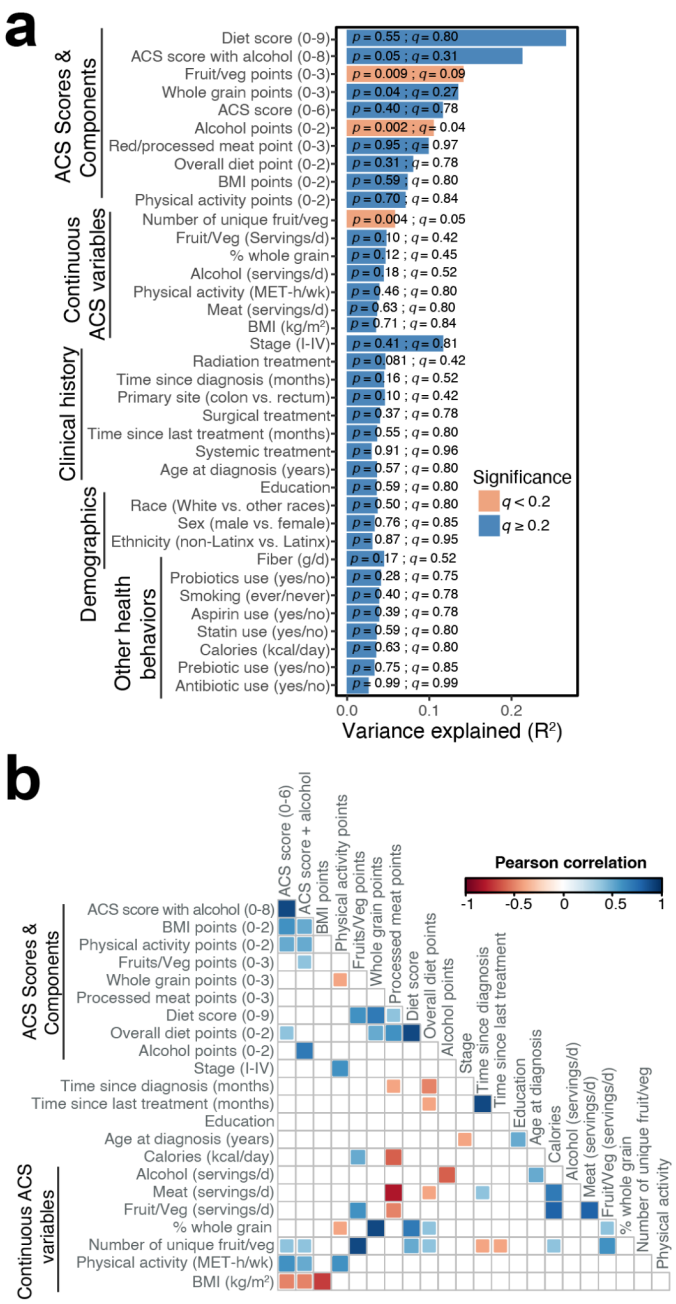
points), which limits the generalizability and may have prevented us from detecting differences in the microbiome between CRC survivors who engage in few health behaviors compared to people who are more adherent to the guidelines. Future studies will include repeated stool sample collections and measures of health behaviors in more racially/ethnically diverse populations of CRC survivors. Paired studies in mice and/or other model organisms are also critical to identify the causal relationships between the variables assessed here, their mechanistic basis, and their potential relevance to host health and disease.

Despite these limitations, we were able to detect significant associations between components of the ACS score and inter-individual variations in the human gut microbiome, which is linked to cancer pathogenesis and progression. Continued progress in this area could help to tailor dietary guidelines for each individual's microbiome or to develop microbiome-based interventions that could assist in the post-treatment recovery and survival of CRC and other cancer patients.

4.4 Supplementary Figures



Supplemental Figure S5. The microbiota from Patient 54 (Pt54) is an outlier with high levels of Proteobacteria. (a) Principal component (PC) analysis using weighted UniFrac beta-diversity metric. **(b)** Alpha diversity analysis using the observed amplicon sequence variance (ASVs). **(c)** Community composition with the color-coded phyla. The short horizontal lines in each bar denote each ASV. **(d)** Proportion of top 3 most abundant genera in Pt54 microbiome. Samples that do not contain any of these 3 genera are not displayed in the graph.



Supplemental Figure S6. A PERMANOVA test with a different beta-diversity metric reveals the same findings of fruits and vegetables and alcohol consumption significantly associated with the microbiota. (a) PERMANOVA testing of metadata variables using Bray-Curtis dissimilarity metric on the genus-level composition of amplicon sequencing dataset. Nominal p-values were reported from the ADONIS test and false discovery rates *q* were calculated with Benjamini-Hochberg multiple testing correction. **(b)** Pearson correlation of numerically-encoded metadata variables. Blank squares represent non-significant ($p>0.05$) correlations. The continuous ACS variables collected from surveys were converted to points for ACS score calculation.

4.5 Materials and Methods

Study participants: Participants for this pilot study were recruited from a single-center open cohort of gastrointestinal cancer survivors, Lifestyle and Outcomes after Gastrointestinal Cancer (LOGIC). The goal of the LOGIC study is to examine health behaviors, quality of life, and clinical outcomes among gastrointestinal cancer survivors. Adults (18 years or older) who have been diagnosed with any gastrointestinal cancer, can complete online surveys in English, and have been seen at UCSF are eligible to enroll. Participants are recruited through the Gastrointestinal Oncology Survivorship Clinic as well as through MyChart invitations sent to UCSF patients with a diagnosis of any gastrointestinal cancer.

For this feasibility study focused on the gut microbiome, we sent one email invitation in October 2019 to active participants in LOGIC who had a diagnosis of colon or rectal cancer and had consented to be contacted about future research opportunities (n=156). The study staff also invited four participants who enrolled in the LOGIC in October 2019 after the initial email invitation. In February 2020, we sent a second invitation to 64 people who had been invited in October and had not replied as well as 3 new participants, so the total number invited was 163. The stool collection sub-study was closed in March 2020 due to the onset of the COVID-19 pandemic, so no additional invitations were sent out.

Survey data: Participants in LOGIC are asked to complete surveys online using Research Electronic Data Capture (REDCap)[34] at enrollment and every six months thereafter for five years. The survey set at enrollment includes sociodemographics, medical history (including height and weight), a validated food frequency questionnaire (FFQ), a validated physical activity questionnaire, and quality-of-life surveys. Self-

reported medical history (including weight) is updated every 12 months; the FFQ is repeated once at 18 months; and the baseline physical activity questionnaire is repeated at 24 and 48 months. Participants who consented to the optional stool sub-study were asked to complete an additional survey at the time of stool collection that asked about antibiotic, prebiotic, and probiotic use. They were also asked to complete an FFQ and a physical activity survey at the time of stool collection if they had not completed one within the past 12 months as part of the parent study.

ACS Guideline score: We calculated a standardized score to estimate adherence to the ACS guidelines that have been previously described (**Figure 4.1b**)[32,35]. Briefly, participants were assigned 0-2 points for body mass index (BMI), total physical activity metabolic equivalent task-hours per week (MET-h/wk), and dietary habits. For BMI, 0 points were assigned if the participant had a BMI <18.5 kg/m^2 or ≥30 kg/m^2, 1 point for a BMI 25.0 - 29.9 kg/m^2, and 2 points for a BMI 18.5 - 24.9 kg/m^2. For physical activity, 0 points were assigned for <8.75 MET-h/wk, 1 point was assigned for 8.75 - <17.5 MET-h/wk, and 2 points were assigned to 17.5 MET-h/wk or more.

To assign the 0-2 points for diet, a diet subscore was first calculated based on intake of fruits and vegetables (1 point for 5 servings/day or more, 0 points if less than 5 servings/day), the number of unique fruits and vegetables (0-2 points assigned to tertiles; **Figure 4.2a**), percent of total grains that are whole grains (0-3 points assigned to quartiles), and intake of red and processed meat (3-0 points assigned to quartiles with the highest score given to the lowest quartile). The diet score (possible range of 0 to 9) was re-weighted to 0-2 points as follows: 0-2 = 0 points, 3-6 = 1 point, and 7-9 = 2 points.

The points for BMI, physical activity, and diet were then summed for a total score ranging from 0 to 6 points.

We also explored a secondary score that included alcohol. This score assigned 0 points to excessive drinkers (>1 drink/d for women, >2 drinks/d for men), 1 point to non-drinkers, and 2 points to low to moderate drinkers (1 drink/d or fewer for women, 2 drinks/day or fewer for men). This scoring system was previously shown to be associated with colon cancer survival[3]. Points for alcohol were added to the standard ACS score, creating a score that ranged from 0 to 8 points.

Participants in our study had completed up to 3 health behavior surveys at the time of stool collection. Therefore, we calculated the ACS score in two ways: using data collected closest in time to stool collection and taking the average of all survey data available. For both calculations, we included data collected up to 6 months after stool collection because we considered the analysis cross-sectional, and the FFQ and physical activity survey both asked participants to recall their usual behaviors over the past year.

Other dietary variables: The FFQ used in the LOGIC study is based on an FFQ that has been extensively described and validated[36]. Nutrient intakes were calculated from the FFQ using databases from the United States Department of Agriculture supplemented with other sources. In this analysis, in addition to examining the ACS score and its components, we explored intake of calories, fiber[37], glycemic index[38], glycemic load[38], long-chain omega-3 fatty acids[39], fish[39], vitamin D[40], coffee[41,42], and tree nuts[43] in relation to the gut microbiome because these factors have been associated with CRC mortality[43], but are not included in the ACS Guideline score.

Clinical data: Clinical data including date of diagnosis, cancer type, stage at diagnosis, and comprehensive treatment history were abstracted from medical records by trained staff at enrollment in LOGIC. The medical records of LOGIC participants were reviewed annually to abstract information on any treatments received in the past year as well as any diagnoses of recurrence or metastasis.

Fecal sample and data collection: Participants were provided a commercially available OMNIgene-GUT kit (DNA Genotek Inc), which contained a stool collection tube with preservatives, user instructions, a spatula, two toilet accessories, a bio-specimen bag, a 2-way mailer prepaid shipping box, and a custom label. Per user instruction, participants were asked to mix the fecal sample with preservatives immediately upon collection and ship it at room temperature to our laboratory. The samples were stored at -80°C until further analysis.

Amplicon sequencing and analysis: DNA was extracted from ethanol-preserved stool samples using the International Human Microbiome Standard operating procedure (IHMS Protocol Q)[44] and the 16S rRNA gene sequencing library was constructed using dual error-correcting barcodes. Briefly, a quantitative primary PCR was performed using KAPA HiFi Hot Start Kit (KAPA KK2502) with V4 515F/806R Nextera universal bacterial primers. The amplified products were diluted 1:100 in UltraPure DNase/RNase-free water and were indexed using unique dual indexing primers. The products were quantified using Quant-iT PicoGreen dsDNA assay kit (Invitrogen P11496) and pooled at equimolar concentrations. The pooled library was quantified via qPCR using the KAPA Library Quantification kit (KAPA KK4824) and its quantity was assessed on 1.5% agarose gel to check the predicted product size. Once quality checks were complete, the library was

sequenced on the Illumina MiSeq platform at 270 x 12 x 12 x 270 cycles. Demultiplexed sequences were processed using our 16S rRNA gene analysis pipeline[45]. High-quality reads were analyzed using qiime2R[46] and phyloseq[47] packages in R. Alpha diversity was assessed through Shannon diversity index using reads subsampled at 40,795 reads per sample. Multivariate analysis of variance (ADONIS) was used to examine the associations between health behavior, sociodemographic, and clinical variables, and microbial community structures. Differentially abundant ASVs were determined using ANOVA and linear regression models.

In our preliminary analysis, we discovered that the sample from patient 54 was different from the rest of the samples (**Supplementary Figure S5a**). Specifically, sample 54 had lower alpha diversity (**Supplementary Figure S5b**) with 80% of its community dominated by facultative anaerobes, such as *Escherichia-Shigella*, *Streptococcus*, and *Klebsiella* (**Supplementary Figure S5c,d**). This participant was the only one in the cohort who had an ileostomy (versus colostomy or no ostomy). Due to its distinct microbiome profile, we excluded this outlier from the remaining analyses described in this manuscript.

Metagenomic sequencing and analysis: 300 ng 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. ZymoBIOMICS synthetic microbial community standard containing 8 bacteria and 2 yeasts (Zymo Research) was included as a positive control to assess bias and errors in the library preparation. The actual taxonomic composition was correlated with the theoretical composition ($r = 0.83$, $p = 0.003$, Pearson correlation). Each sample was quantified with PicoGreen (ThermoFisher) and checked for quality with TapeStation 4200 (Agilent). Once every sample passed the quality checks,

the samples were pooled to the same concentration and sequenced using an S1 flow cell on NovaSeq 6000 system (Illumina) at the Chan Zuckerberg Biohub. The demultiplexed sequences were processed through our metagenomic pipeline, which trims low-quality reads with Fastp[48], removes human reads with Bowtie2[49], determines genome equivalents with Microbecensus[50], and assigns reads to taxonomic and functional profiles with Humann3[51].

Of 30 million raw reads per sample on average, 99.0% of reads passed the quality filtering step, yielding 29 million high-quality reads per sample. To account for uneven sequencing depth, reads were normalized to genome equivalents and converted to reads per kilobase per genome equivalent (RPKG). Unstratified gene families and pathways were used to analyze the functional composition of the microbiome. Genes or pathways detected in less than 3 samples were removed. The top 2.5% of the most abundant gene families were analyzed for differential abundance using ANOVA and generalized linear models.

Statistical analysis: We used descriptive statistics to describe the sociodemographic and clinical characteristics of the study sample, including the median (interquartile range, IQR) for continuous variables and N (%) for categorical variables. To examine the ACS score, its components, and other dietary factors in relation to the microbiome, we modeled the ACS score as an ordinal variable; components of the ACS score were examined categorically, using the point cut-offs defined above, as well as continuously; and the other dietary factors (described in the “Other dietary variables” above) were examined continuously.

Permutational multivariate analysis of variance was used to test the association between the health behavior, sociodemographic, and clinical variables and the beta-diversity on R using the Vegan package[52]. A general linear univariate model, one-way ANOVA, and t-tests were used to test the univariate associations between continuous variables (time since diagnosis, time since last treatment, age, fiber, calories, number of unique fruit/veg, % whole grain, BMI, fruit/veg serving, alcohol, physical activity, and meat), categorical variables (ACS score, ACS score with alcohol, diet sub-score, fruit/veg points, whole grain points, alcohol points, meat points, overall diet points, physical activity points, BMI points, stage, education, race, and ethnicity), and dichotomous variables (primary site, radiation treatment, surgical treatment, systemic treatment, sex, smoking, probiotic use, prebiotic use, aspirin use, statin use, and antibiotic use) and microbiome features, respectively. The Benjamini-Hochberg test was used to account for multiple testing.

4.6 References

1. Cancer of the Colon and Rectum - Cancer Stat Facts. *SEER* https://seer.cancer.gov/statfacts/html/colorect.html.
2. Rock, C. L. *et al.* American Cancer Society guideline for diet and physical activity for cancer prevention. *CA Cancer J. Clin.* **70**, 245–271 (2020).
3. Van Blarigan, E. L. *et al.* Association of Survival With Adherence to the American Cancer Society Nutrition and Physical Activity Guidelines for Cancer Survivors After Colon Cancer Diagnosis: The CALGB 89803/Alliance Trial. *JAMA Oncol* **4**, 783–790 (2018).
4. Kostic, A. D. *et al.* Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome Res.* **22**, 292–298 (2012).
5. Dejea, C. M. *et al.* Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. *Science* **359**, 592–597 (2018).
6. Chung, L. *et al.* Bacteroides fragilis Toxin Coordinates a Pro-carcinogenic Inflammatory Cascade via Targeting of Colonic Epithelial Cells. *Cell Host Microbe* **23**, 421 (2018).
7. Kostic, A. D. *et al.* Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe* **14**, 207–215 (2013).
8. Cao, Y. *et al.* Commensal microbiota from patients with inflammatory bowel disease produce genotoxic metabolites. *Science* **378**, eabm3233 (2022).
9. Wilson, M. R. *et al.* The human gut bacterial genotoxin colibactin alkylates DNA. *Science* **363**, (2019).

10. Sivan, A. *et al.* Commensal Bifidobacterium promotes antitumor immunity and facilitates anti-PD-L1 efficacy. *Science* **350**, 1084–1089 (2015).
11. Iida, N. *et al.* Commensal bacteria control cancer response to therapy by modulating the tumor microenvironment. *Science* **342**, 967–970 (2013).
12. Spanogiannopoulos, P. *et al.* Host and gut bacteria share metabolic pathways for anti-cancer drug metabolism. *Nature Microbiology* vol. 7 1605–1620 Preprint at https://doi.org/10.1038/s41564-022-01226-5 (2022).
13. Kyaw, T. S. *et al.* Human gut Actinobacteria boost drug absorption by secreting P-glycoprotein ATPase inhibitors. Preprint at https://doi.org/10.1101/2022.10.13.512142.
14. Geller, L. T. *et al.* Potential role of intratumor bacteria in mediating tumor resistance to the chemotherapeutic drug gemcitabine. *Science* **357**, 1156–1160 (2017).
15. Wallace, B. D. *et al.* Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. *Science* **330**, 831–835 (2010).
16. Rothschild, D. *et al.* Environment dominates over host genetics in shaping human gut microbiota. *Nature* **555**, 210–215 (2018).
17. Zeevi, D. *et al.* Personalized Nutrition by Prediction of Glycemic Responses. *Cell* **163**, 1079–1094 (2015).
18. David, L. A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014).
19. von Schwartzenberg, R. J. *et al.* Caloric restriction disrupts the microbiota and colonization resistance. *Nature* **595**, 272–277 (2021).
20. Ang, Q. Y. *et al.* Ketogenic Diets Alter the Gut Microbiome Resulting in Decreased

Intestinal Th17 Cells. *Cell* **181**, 1263–1275.e16 (2020).

21. Balhareth, A., Aldossary, M. Y. & McNamara, D. Impact of physical activity and diet on colorectal cancer survivors' quality of life: a systematic review. *World J. Surg. Oncol.* **17**, 153 (2019).
22. Piawah, S., Walker, E. J., Van Blarigan, E. L. & Atreya, C. E. The Gut Microbiome in Colorectal Cancer. *Hematol. Oncol. Clin. North Am.* **36**, 491–506 (2022).
23. de Vries, W. & Stouthamer, A. H. Pathway of glucose fermentation in relation to the taxonomy of bifidobacteria. *J. Bacteriol.* **93**, 574–576 (1967).
24. Muir, J. G. *et al.* Measurement of short-chain carbohydrates in common Australian vegetables and fruits by high-performance liquid chromatography (HPLC). *J. Agric. Food Chem.* **57**, 554–565 (2009).
25. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
26. Hartman, A. L. *et al.* Human gut microbiome adopts an alternative state following small bowel transplantation. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 17187–17192 (2009).
27. Myhrstad, M. C. W., Tunsjø, H., Charnock, C. & Telle-Hansen, V. H. Dietary Fiber, Gut Microbiota, and Metabolic Regulation-Current Status in Human Randomized Trials. *Nutrients* **12**, (2020).
28. Carmody, R. N. *et al.* Cooking shapes the structure and function of the gut microbiome. *Nat Microbiol* (2019) doi:10.1038/s41564-019-0569-4.
29. Bess, E. N. *et al.* Genetic basis for the cooperative bioactivation of plant lignans by Eggerthella lenta and other human gut bacteria. *Nat Microbiol* (2019)

doi:10.1038/s41564-019-0596-1.

30. Zhang, L. *et al.* Grape proanthocyanidin-induced intestinal bloom of Akkermansia muciniphila is dependent on its baseline abundance and precedes activation of host genes related to metabolic health. *J. Nutr. Biochem.* **56**, 142–151 (2018).
31. Muegge, B. D. *et al.* Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* **332**, 970–974 (2011).
32. McCullough, M. L. *et al.* Following cancer prevention guidelines reduces risk of cancer, cardiovascular disease, and all-cause mortality. *Cancer Epidemiol. Biomarkers Prev.* **20**, 1089–1097 (2011).
33. Pohl, K., Moodley, P. & Dhanda, A. D. Alcohol's Impact on the Gut and Liver. *Nutrients* **13**, (2021).
34. Harris, P. A. *et al.* The REDCap consortium: Building an international community of software platform partners. *J. Biomed. Inform.* **95**, 103208 (2019).
35. Thomson, C. A. *et al.* Nutrition and physical activity cancer prevention guidelines, cancer risk, and mortality in the women's health initiative. *Cancer Prev. Res.* **7**, 42–53 (2014).
36. Al-Shaar, L. *et al.* Reproducibility and Validity of a Semiquantitative Food Frequency Questionnaire in Men Assessed by Multiple Methods. *Am. J. Epidemiol.* **190**, 1122–1132 (2021).
37. Song, M. *et al.* Fiber Intake and Survival After Colorectal Cancer Diagnosis. *JAMA Oncol* **4**, 71–79 (2018).
38. Meyerhardt, J. A. *et al.* Dietary glycemic load and cancer recurrence and survival in patients with stage III colon cancer: findings from CALGB 89803. *J. Natl. Cancer*

Inst. **104**, 1702–1711 (2012).

39. Van Blarigan, E. L. *et al.* Marine ω-3 Polyunsaturated Fatty Acid and Fish Intake after Colon Cancer Diagnosis and Survival: CALGB 89803 (Alliance). *Cancer Epidemiol. Biomarkers Prev.* **27**, 438–445 (2018).
40. Ng, K. *et al.* Effect of High-Dose vs Standard-Dose Vitamin D3 Supplementation on Progression-Free Survival Among Patients With Advanced or Metastatic Colorectal Cancer: The SUNSHINE Randomized Clinical Trial. *JAMA* **321**, 1370–1379 (2019).
41. Hu, Y. *et al.* Association Between Coffee Intake After Diagnosis of Colorectal Cancer and Reduced Mortality. *Gastroenterology* **154**, 916–926.e9 (2018).
42. Guercio, B. J. *et al.* Coffee Intake, Recurrence, and Mortality in Stage III Colon Cancer: Results From CALGB 89803 (Alliance). *J. Clin. Oncol.* **33**, 3598–3607 (2015).
43. Fadelu, T. *et al.* Nut Consumption and Survival in Patients With Stage III Colon Cancer: Results From CALGB 89803 (Alliance). *J. Clin. Oncol.* **36**, 1112–1120 (2018).
44. IHMS. http://www.human-microbiome.org/index.php?id=Sop&num=006.
45. Bisanz, J. *AmpliconSeq*. (Github).
46. Bisanz, J. *qiime2R: Import qiime2 artifacts to R*. (Github).
47. phyloseq: Explore microbiome profiles using R. https://joey711.github.io/phyloseq/.
48. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
49. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

50. Nayfach, S. & Pollard, K. S. Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome. *Genome Biol.* **16**, 51 (2015).
51. Franzosa, E. A. *et al.* Species-level functional profiling of metagenomes and metatranscriptomes. *Nat. Methods* **15**, 962–968 (2018).
52. Oksanen, J. Vegan: community ecology package. R package version 1.8-5. *http://www. cran. r-project. org* (2007).

Chapter 5: Conclusion

Much of the past and current research are heavily focused on understanding the contribution of the gut microbiome on colorectal cancer pathogenesis. Our work described here expanded our view of the potential impact of the gut microbiome on not only pathogenesis but also on treatment response, resistance, and survivorship. We established two specific mechanisms through which the gut microbiome can explain variability in the efficacy, side effect profile, and bioavailability of major cytotoxic anticancer agents. This work can be utilized to better predict the treatment failure or success for millions of patients with colorectal cancer and to help develop precision medicine capable of addressing inter-individual variations in CRC treatment outcomes.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Than Kyaw
CA3CBF675B274F4...
Author Signature

12/5/2022
Date