Supplementary Methods S1 for "Personalized mapping of drug metabolism by the human gut microbiome"

Contents

A Culturing replicability

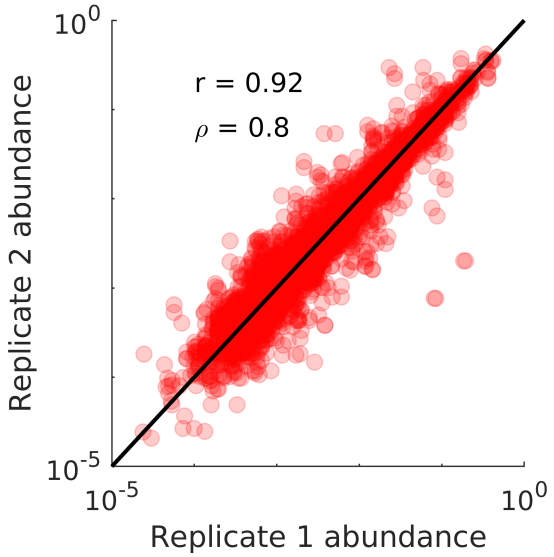

Figure 1: **Correlation between abundances of replicate BG cultures.** For each donor, three replicate BG *ex vivo* cultures were grown and sequenced. The replicates were inoculated from the same glycerol vial. Shown is the relationship between abundances of ASVs in all possible pairs of replicates (three comparisons per donor). r is the Pearson correlation and ρ is the Spearman correlation. Due to the log scale, the plot only shows ASVs present in both members of a given pair. However, the correlations shown in the figure were calculated with all ASVs present in at least one of the members of a pair.

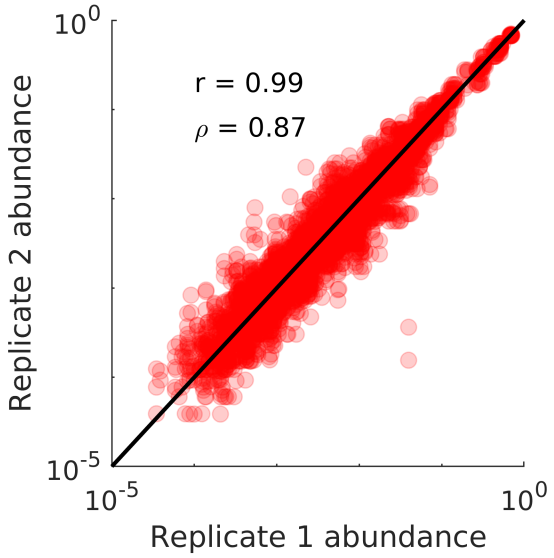

Figure 2: **Correlation between abundances of replicate mGAM cultures.** For each donor, six replicate mGAM *ex vivo* cultures were grown and sequenced. The replicates were inoculated from two different glycerol vial stocks of the same donor (three replicates from each vial). Shown is the relationship between abundances of ASVs in all possible pairs of replicates. r is the Pearson correlation and ρ is the Spearman correlation. Due to the log scale, the plot only shows ASVs present in both members of a given pair. However, the correlations shown in the figure were calculated with all ASVs present in at least one of the members of a pair.

B Design of BG media

In this section we detail the rationale behind the composition of BG as a 70/30 mixture of BB and mGAM media. Two media consistently performed well in our initial sequencing of D1-20 samples: mGAM, as previously seen with PD, and BB. Interestingly, mGAM had moderate ASV richness and high biomass, while BB yielded a much lower biomass with high richness and did not suffer from the Enterobacteriacae expansion observed in mGAM. We sought to combine these two media to create a hybrid medium with moderate biomass, high richness, and a reduced Enterobacteriacae expansion. To determine the proportions, we relied on the community compositions of BB and mGAM cultures from the first nine donors (this is all we had sequenced at the time). We assumed that a community growing in a media composed of m fraction BB would have a composition of $\vec{x}_{new} = m\vec{x}_{BB} + (1-m)\vec{x}_{mGAM}$ and sought to find the average m at which the new community's Shannon entropy is maximized (this was done before ENDS was developed). Below we show the diversity of the theoretical mixture communities as a function of BB fraction for the nine donors. The average optimal mixture fraction was 0.68, which we rounded to 0.70 for simplicity.

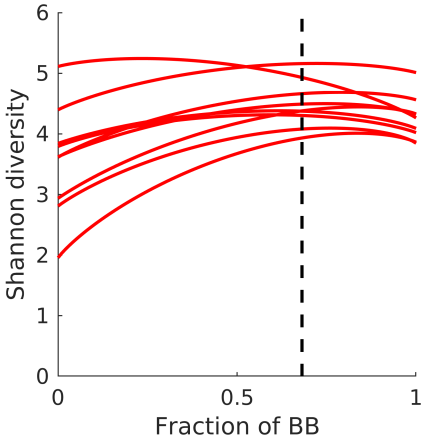

Figure 3: **Diversity of the theoretical mixture communities as a function of BB fraction for the nine initial donors.** Each red line is the mixture diversity for a single donor and the dashed black line is the average position of the point of maximum diversity.

C Theoretical relationships between taxa abundance and observed metabolite production

In this section we explore the types of relationships that could theoretically exist between a single metabolite producer and its metabolite. We start with the simplest model: production of the metabolite by only a single bacterium. We assume linear metabolite production where the parent drug is saturating. The resulting dynamics are $\frac{d[M]}{dt} = rx_1$ and if we assume that x_1 is constant we solve this to yield $[M](t) = rx_1t$. In this very simple case, there is a linear relationship between the metabolite producer and its metabolite.

What happens to the relationship if we consider additional bacteria? Consider a model with a second bacterial strain that produces the metabolite at the same rate. The dynamics defining the model are $\frac{d[M]}{dt} = r(x_1 + x_2)$ and we again assume steady-state bacterial abundances to yield $[M](t) = r(x_1 + x_2)t$. Assuming that the abundances of the two strains are independently distributed, the inclusion of the second producer will result in a reduced correlation between the first producer and its metabolite. If we introduce a metabolite consumer instead of a producer, the results are similar. If the consumer reduces the metabolite in proportion to its current level, our dynamics are $\frac{d[M]}{dt} = rx_1 - rx_3[M]$ and the resulting metabolite level is $[M](t) = \frac{x_1}{x_3}(1 - e^{-rx_3t})$. If we have both the consumer and additional producer present, the relationship between the original producer and the metabolite is further obscured. With two producers and a consumer, the dynamics are $\frac{d[M]}{dt} = r(x_1 + x_2) - rx_3[M]$ and the resulting metabolite level is $[M](t) = \frac{x_1+x_2}{x_3}(1 - e^{-rx_3t})$. We show one realization of these models in Figure S4.

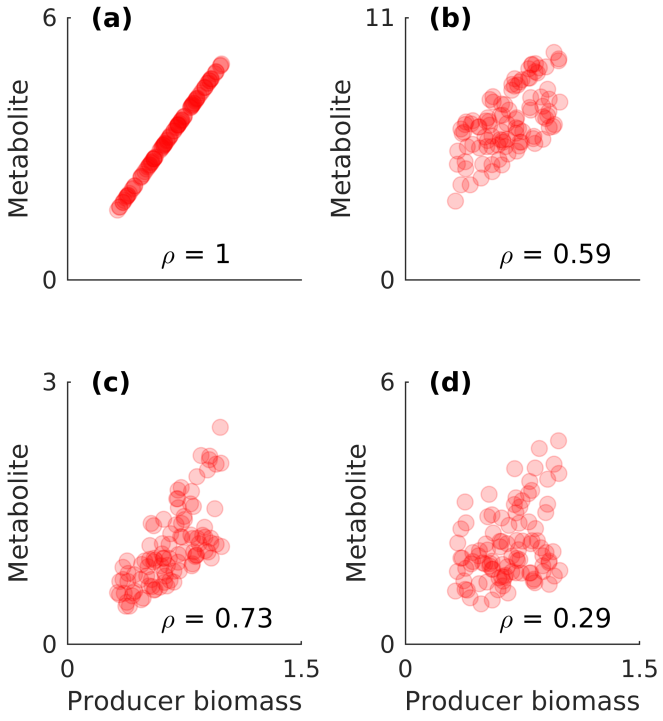

Figure 4: **Examples of theoretical relationships between producer abundance and measured metabolite level.** The abundance of all bacteria is sampled from a uniform distribution on the interval [0.2,1] and we set $rt = 1$. ρ is the Spearman correlation. **(a)** Relationship between producer abundance and metabolite production when only a single producer is present. **(b)** Relationship between producer abundance and metabolite production when there is an additional producer present. **(c)** Relationship between producer abundance and metabolite production when there is a consumer present. **(d)** Relationship between producer abundance and metabolite production when there is an additional producer and a consumer present.

D Estimating rates of MDM reactions for ENDS

The ENDS metric has one free parameter: the rate of the MDM reaction that determines how much metabolite a given amount of microbial biomass produces in 24 hours. The optimal medium can vary depending on this constant (as shown in Fig 3G) and we must therefore determine what rates are physically meaningful. In our case, there are only two optimal media in the ranges we test. Below a rate constant of approximately 0.1 $\frac{\text{normalized AUC}}{\text{g/L·day}}$, mGAM is the optimal medium, while above this BG is the optimal medium. To determine which region is more relevant, we estimated the reaction rate necessary for bacteria to influence the bioavailability of a drug before it leaves the small intestine and estimated lower bounds of the reaction rates for two known MDM reactions.

What reaction rate is needed for a change in bioavailability of a drug before it leaves the small intestine? Suppose that a 40 μmol ($O(10^1)$ mg) dose of a drug is taken orally. The small intestine contains approximately 10^{11} bacteria [6]. Assuming these bacteria have a similar dry weight to *E. coli* of 2.3×10^{-13} g [3], this corresponds to a dry weight of 0.023 g. With a dry weight percentage of 30 % [1], this is a wet weight of 0.077 g. We require that the bacteria must convert 10% of the dose to a metabolite within a 3 hour small intestine transit time [2], giving us a required rate of $(4\mu\text{mol})/(0.077\text{g} \cdot 3\text{h}) \approx 17\frac{\mu\text{mol}}{\text{g·h}}$. With our incubation time gives us a reaction rate for our screen of 408 $\frac{\mu M}{\text{g/L·day}}$. In our screen, 20 μM typically corresponds to 1 normalized AUC (though this varies by compound), meaning that the reaction rate is on the order of 10^1 $\frac{\text{normalized AUC}}{\text{g/L·day}}$, well within the region where BG is superior.

We can estimate a lower bound of the sulfasalazine metabolite reaction rate from our data. Our biomass data indicates that BG cultures have an average biomass of approximately 10 g/L wet mass and in 24 hours these cultures produce approximately 0.75 normalized AUC of sulfapyridine. The lowest possible rate of sulfapyridine production, assuming all microbes are performing the reaction and reaching the final normalized AUC requires the entire 24 hours, is $(1\ \text{day}^{-1})(0.75\ \text{normalized AUC})/(10\ \text{g/L}) \approx 0.075\ \frac{\text{normalized AUC}}{\text{g/L·day}}$.

We can estimate a lower bound of the dihydrodigoxin production rate from Koppel et al. 2018 [5]. Supplementary Fig. 2 of Fig. 4 of the paper shows that *E. lenta* DSM 2243 grown in BHI media reaches a final OD_{600} of approximately 0.1 in the presence of digoxin. Assuming a similar OD_{600} to wet biomass conversion as *E. coli* (1.7 g/L wet weight per unit) [4], this corresponds to a wet biomass of 0.17 g/L. Figure 1A of the paper shows that the same strain converts 94% of 10 μM digoxin into dihydrodigoxin within 48 hours, corresponding to a dihydrodigoxin signal of 0.2 normalized AUC on our machine. The lowest possible rate of dihydrodigoxn production, assuming the culture quickly reaches its final OD and reaching the final conversion requires the entire 48 hours, is $(0.5\ \text{day}^{-1}) \times (0.2\ \text{normalized AUC})/(0.17\ \text{g/L}) \approx 0.6\ \frac{\text{normalized AUC}}{\text{g/L·day}}$.

All three of these rate estimates are either within the BG region, or have lower bounds near the border of the region, meaning that the true reaction rates almost certainly fall within the region where BG is superior. For this reason we selected BG as our optimal media.

E Sensitivity of ENDS and PD analyses to sequencing depth

Diversity metrics computed from amplicon sequencing data are potentially sensitive to sampling depth due to the discovery of new variants with additional reads. To account for this, we designed our study to minimize variability in sampling depth by processing all samples for a given experiment within a short time frame using the same protocol. We will first characterize the effect of sequencing depth on the D1-20 analyses, since these depend primarily on an ASV-based metric. We show that (1) ENDS has low correlation with sequencing depth in our dataset, (2) there are no large differences between the mean sequencing depth of different media conditions, and (3) the results of our analysis do not appreciably change in a multilinear model accounting for the possible confounding effects of sequencing depth.

We begin by characterizing the relationship between sequencing depth and ENDS computed at a moderate reaction rate ($0.5\ \frac{\text{normalized AUC}}{\text{g/L·day}}$) in Figure S5 and find a low Pearson correlation of $r = 0.29$. To test a more extreme case, we look at the relationship between the ASV richness (equivalent to ENDS with an infinite reaction rate) and depth in Figure S6. This is the limit in which ENDS is maximally dependent on sequencing depth since the metric now takes into account all low abundance variants. In this case, we find a higher but still weak correlation at $r = 0.43$.

To examine differences in sequencing depth between media, we show the sequencing depths of all samples grouped by media type in Figure S7. As can be seen, there are only minor differences in mean media sequencing depth. To more rigorously examine the impact of sequencing depth on ENDS, we constructed multilinear models of ENDS (at the same reaction rate characterized in Figure S5) and ASV richness that take into account donor, media type, and sequencing depth. The inclusion of donor is needed since not all media-donor combinations met the 10,000 read filtering threshold. We excluded the replicates of BG and GAM to avoid pseudoreplication. In both of these models, the effect of media type remains highly significant ($p < 0.001$, ANOVA F-test) while the impact of sequence depth is not significant ($p > 0.01$, ANOVA F-test). The effect size coefficient of the BG media type is either the highest or within the standard error of the highest effect size coefficient.

We also examine the effect of sequencing depth on the Shannon diversity metrics computed in the PD analyses. In this analysis, we exclude samples below 10,000 reads and the original PD fecal sample. There are no statistically significant differences in sequence depth between media types ($p > 0.05$, one-way ANOVA F-test) and the correlation between Shannon diversity and sequencing depth is low ($r = 0.37$). In a multilinear model of Shannon diversity accounting for media and sequencing depth, media type is significant ($p < 0.001$, ANOVA F-test) and sequencing depth is not ($p > 0.05$, ANOVA F-test).

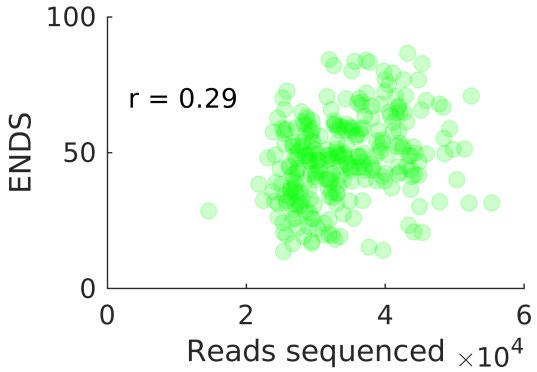

Figure 5: **Correlation between ENDS and sequencing depth**. ENDS is computed at a reaction rate of $0.5\ \frac{\text{normalized AUC}}{\text{g/L·day}}$. r is the Pearson correlation.

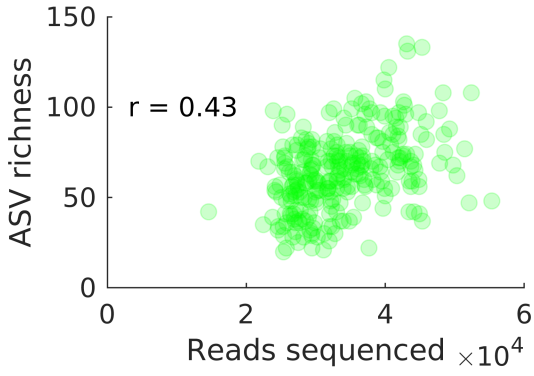

Figure 6: **Correlation between ASV richness and sequencing depth**. r is the Pearson correlation.

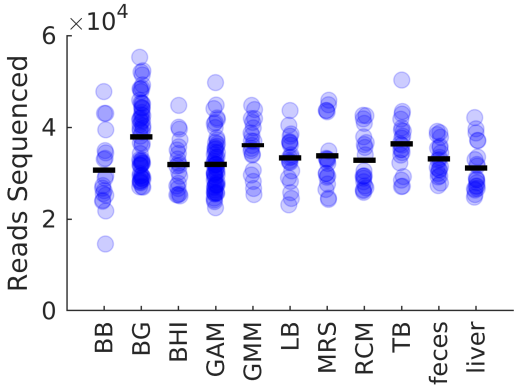

Figure 7: **Sampling depths of various media conditions**. The black bar represents the mean sequencing depth of each media.

F ENDS with empirically estimated power

At the core of ENDS is the function, $B(\mu)$, relating the mean of a signal observed on our MS to the probability that we will be able to detect it as significant compared to the control condition. In our estimation of this function we make the strong assumption that all distributions involved in the measurement process are normal. In this section, we instead estimate $B(\mu)$ empirically and show that using the empirically estimated function to compute ENDS provides similar results to those reached assuming all distributions are normal.

From the twenty donor screen, we have data from a large set of statistical tests of metabolite production. For each of these tests, we have the sample mean of the observed MS signal and the outcome of the test. We perform a logistic regression on this data to estimate $B(\mu)$, assuming $B(\mu)$ has the form of:

$$\log \frac{B(\mu)}{1 - B(\mu)} = b_1 + b_2 \log \mu \tag{1}$$

Where b_1 and b_2 are parameters to be estimated. We perform this regression using the glmfit function within Matlab using the uncorrected screen p-values to determine statistical significance (using a significance cutoff of $\alpha = 0.01$). We then compute ENDS with the resulting fit and show the results in Fig S8. These results are similar to those in Fig 3G, generally preserving the hierarchies between media.

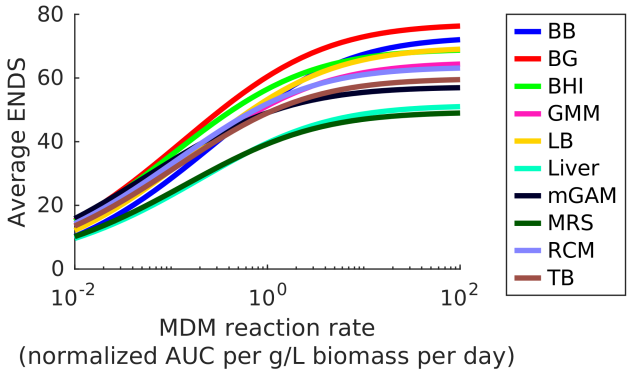

Figure 8: **ENDS computed with empirical power estimates**. Here ENDS with the same parameters as in Fig 3G in the main text except with the power function estimated empirically from the results of the twenty donor screen.

References

[1] Gunnar Bratbak and Ian Dundas. Bacterial dry matter content and biomass estimations. *Appl. Environ. Microbiol.*, 48(4):755–757, 1984.

[2] SS Davis, JG Hardy, and JW Fara. Transit of pharmaceutical dosage forms through the small intestine. *Gut*, 27:886–892, 1986.

[3] Patrick P. Dennis and Hans Bremer. Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. *EcoSal Plus*, 3(1), 2008.

[4] Julia Glazyrina, Eva-Maria Materne, Thomas Dreher, Dirk Storm, Stefan Junne, Thorsten Adams, Gerhard Greller, and Peter Neubauer. High cell density cultivation and recombinant protein production with escherichia coli in a rocking-motion-type bioreactor. *Microbial Cell Factories*, 9(1):42, 2010.

[5] Nitzan Koppel, Jordan E Bisanz, Maria-Eirini Pandelia, Peter J Turnbaugh, and Emily P Balskus. Discovery and characterization of a prevalent human gut bacterial enzyme sufficient for the inactivation of a family of plant toxins. *eLife*, 7:e33953, 2018.

[6] Ron Sender, Shai Fuchs, and Ron Milo. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biology*, 14(8):1–14, 2016.