A picture containing text

Description automatically generated

**Figure S1. The baseline gut microbiota composition of the four vendors (Beijing, Guangdong, Hunan, Shanghai) at the family level.** Bars represent individual mice. Adonis test indicates significant difference in the baseline gut microbiota composition across the four vendors (P<0.001).

A screenshot of a computer

Description automatically generated with medium confidence

**Figure S2. Effects of inulin on (A) body weight, (B) daily food intake, (C) daily energy intake, and (D) 48-hr fecal sample weight of mice receiving diet supplementation.** Each symbol represents the mean body weight in panel A or a single data point in panels B-D. All food intakes were converted to energy intakes by multiplying food weight and its energy density (3.8 and 3.9 kcal/g for the cellulose- and inulin-based diets, respectively). The body weight data were analyzed by ordinary one-way ANOVA (Analysis of variance) with Turkey post hoc test between inulin and cellulose group. \* *P* < 0.05.

A picture containing text

Description automatically generated

**Figure S3. Dynamics of (A) evenness and (B) number of observed ASVs following inulin intervention.** Lines represent mean values across mice within the same vendor and shading areas represent standard error of the mean.

A screenshot of a video game

Description automatically generated with low confidence

**Figure S4.** Relative abundance of gene families in the gut following inulin intervention. **A**. Gene family composition represented by robust PCoA (principal coordinate analysis) plot. R2 and P-value were obtained from Adonis analysis, which tests for the difference in gene abundances among three representative timepoints during intervention (day 0: baseline, day 5: short-term response, day 31: long-term response). **B**. Increased expression of inulinase genes following inulin treatment. Each dotted line represents an individual mouse. \*: P < 0.05; \*\*: P < 0.01; \*\*\*: P < 0.001.

A screen shot of a video game

Description automatically generated with low confidence

**5**D from

**A picture containing text

Description automatically generated**

**Figure S6. Reconstructed time series of total bacterial load and three major short-chain fatty acids by sequential Non-negative matrix factorization.** Dots represent observations. Both lines and dots are color-coded on a per-mouse basis.

Graphical user interface, application

Description automatically generated

**Figure S7.** **Bacterial taxa with significant difference in relative abundance between the inulin group and the cellulose group.** Relative abundance changes were calculated between day 0 and day 1 (A), day 0 and day 5 (B), day 0 and day 31 (C). *P*-values were obtained from Wilcoxon rank-sum test after multiple test correction via false discovery rate (FDR) estimation. \*, FDR < 0.05; \*\*, FDR < 0.01; \*\*\*, FDR < 0.001.

A picture containing indoor, colorful, dark

Description automatically generated

**Figure S8. Dynamical responses of the five inulin degraders.** Lines/dots: absolute abundance averaged across mice from the same vendor. Shading area: standard error of the mean values.

**Background pattern

Description automatically generated with low confidence**

**Figure S10. Poor performance of Random Forest (RF) regression model in predicting short-chain fatty acids (SCFAs) concentration (see Fig. 5B of the main text for the results) cannot be rescued by using (A) alternative predictors, (B) alternative regression models, and (C,D) weighting of training samples. A**. Prediction accuracy of a RF model trained on different taxonomic- (ASV, Species, Genus, Family) or functional- (Gene, Pathway, Phenotype) predictors. For each taxonomic level, unclassified or uncultured taxa at this level were grouped by the lowest classified rank above this level. The abundances of genes, pathways and phenotypes were predicted using PICRUSt2. **B**. Prediction accuracy of the MelonnPan algorithm [45] trained on the same predictors as used in panel A. **C**. Weights assigned to the training data. The gut microbiota composition of all samples was shown in a reduced two-dimensional UMAP (Uniform Manifold Approximation and Projection) space [49]. The bigger the weights, the larger circle sizes. See Methods in the main text for details of weight calculation. **D**. Prediction accuracy of an RF model built from weighted training data. The absolute abundance of bacterial taxa (grouped by the lowest classified taxonomic level) was used as predictors.

**A picture containing chart

Description automatically generated**

**Figure S11. Prediction of short-chain fatty acid (SCFA) concentration from gut microbiota using data from resistant starch-treated mice**. The same figure legend applies as in the main text Fig. 5B-D (the same order).

**A screenshot of a video game

Description automatically generated**

**Figure S12. Inference of SCFA producers in inulin-treated mice by different approaches** (Random forest regression vs. Repeated correlation analysis)**. A.** Random forest (RF) regression. For each SCFA, we showed top 10 bacterial taxa with highest Gini importance score in model training using all data. The absolute abundances of bacterial taxa were standardized and filtered (threshold 10-5) by LASSO (least absolute shrinkage and selection operator) regression before passing to RF model. Several key hyperparameters in LASSO and RF were optimized using grid search cross-validation with R2 as the score metric. The vendor-level prevalence scores were obtained from Fig. 5C in the main text. **B**. Repeated correlation analysis [50]. Longitudinal data and correlation trend lines are color-coded on a per-mouse basis. Repeated measures correlation coefficients (*r*rm) and FDR-corrected P-values are indicated in the plot.

A screenshot of a video game

Description automatically generated with medium confidence

**Figure S14. Prediction of time-averaged growth response of gut microbiota by the relative abundance of dietary fiber responders in the baseline community**. **A**. Inulin intervention. **B**. Resistant starch intervention. The time averaged growth response is obtained by the area under the curve of total bacterial density divided by the observation time**.** The combination of responders highlighted in red has the highest Pearson correlation coefficient. In both scatter plots, gray lines represent the best fitting line.

A computer screen capture

Description automatically generated with low confidence

**Figure S15**. **The relationship between microbiome and metabolome is time-dependent.** **A**. Dynamics of gut microbiota composition (x-axis) and total SCFA concentration (y-axis) plotted on the same graph. We used the first principal coordinate score from PCoA (principal coordinate analysis) ordination to represent changes in gut microbiota composition (relative abundance) along the direction of maximum variance. Note that SCFAs were substantially produced between day 0 and 1 while gut microbiota composition only changes slightly. Points represent the mean PCoA coordinate score across mice within each vendor and error bars represent the standard error of the mean. **B**. Correlation of baseline unclassified (Un.) Parabacteroides absolute abundance with initial propionate production rates on day 0 (upper panel) and rates in later days (lower panel). Gray line: linear regression.

**Text

Description automatically generated**

**Figure S16. Rarefaction analysis of 16S rRNA amplicon sequencing data.** Rarefaction curves were generated using the iNEXT package [81]. Solid lines represent the observed alpha diversity with the number of reads sampled, and dashed lines represent the extrapolation of the solid lines until 25% more reads. To avoid sample-to-sample bias due to variable sequencing depth (different number of reads per sample), all samples were rarefied to 38,980 sequences (black dashed line) per sample before downstream analysis.

A picture containing text, outdoor object

Description automatically generated