

Supplementary Information for

An approach for evaluating the effects of dietary fiber polysaccharides on the human gut microbiome and plasma proteome

Omar Delannoy-Bruno, Chandani Desai, Juan J. Castillo, Garret Couture, Ruteja A. Barve, Vincent Lombard, Bernard Henrissat, Jiye Cheng, Nathan Han, David K. Hayashi, Alexandra Meynier, Sophie Vinoy, Carlito B. Lebrilla, Stacey Marion, Andrew C. Heath, Michael J. Barratt, and Jeffrey I. Gordon*

*Corresponding author

email: jgordon@wustl.edu

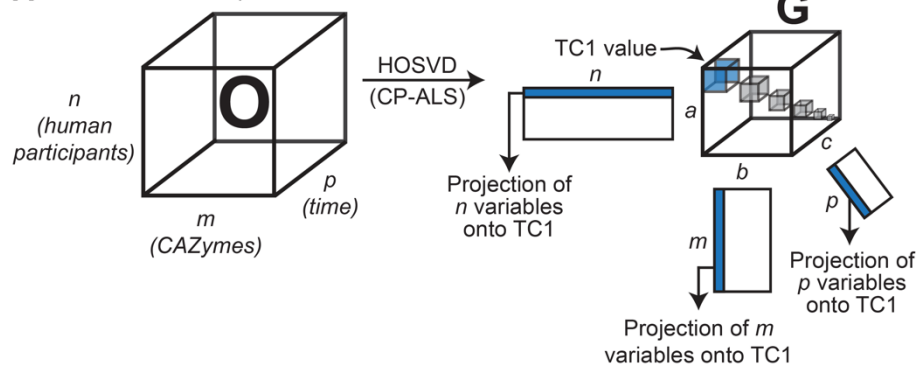
This PDF file includes:

Supplementary Figs. S1-S6

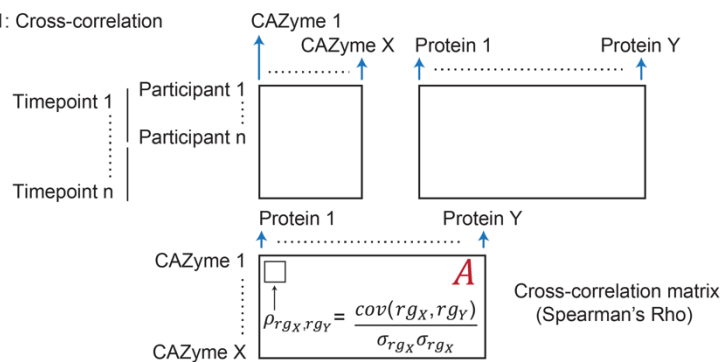
Datasets S1-S7

SUPPLEMENTARY FIGURES

A HOSVD of $n \times m \times p$ tensor



B Step 1: Cross-correlation



Step 2: Singular Value Decomposition (SVD)

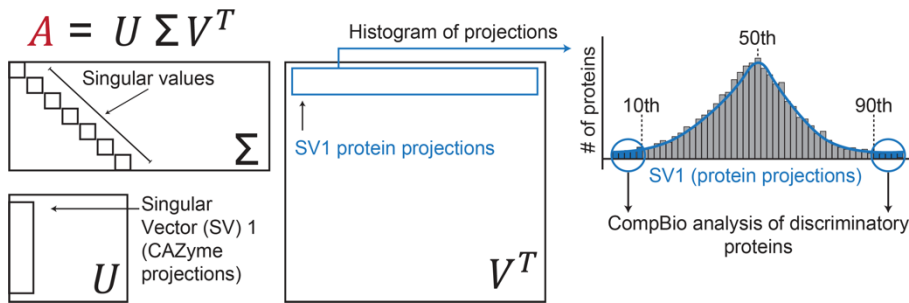


Fig. S1. Analytic methods. (A) Schematic of HOSVD analysis. (B) Schematic of CC-SVD and CompBio analyses.

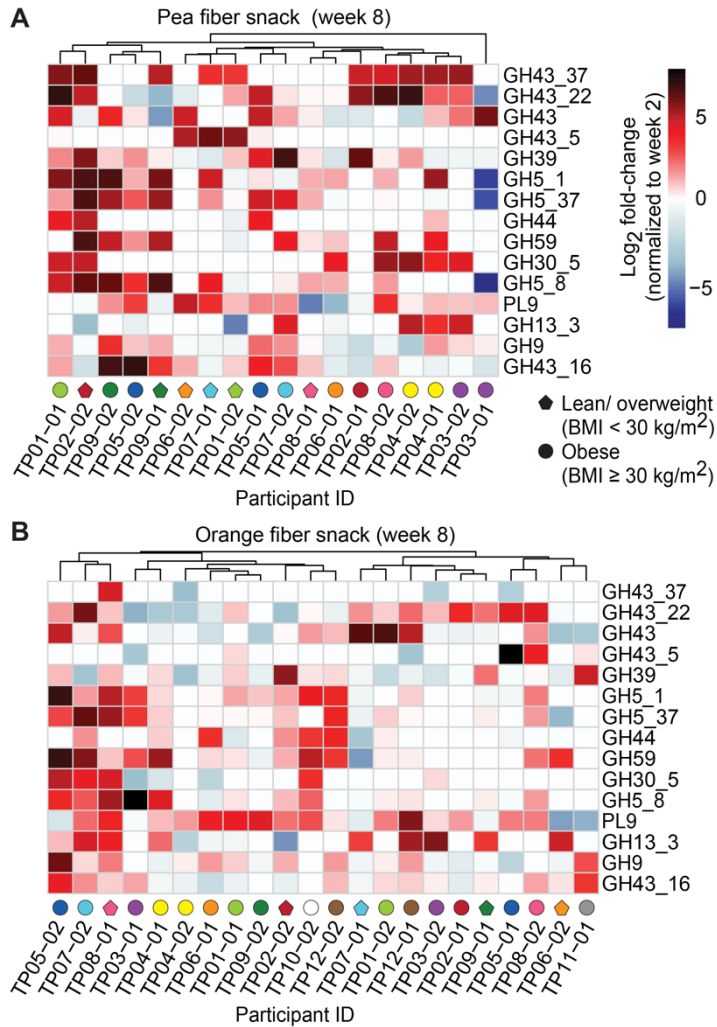


Fig. S2. Interpersonal variation in CAZyme gene responses to the two fiber snack prototypes. (A,B) Heatmap plotting the log₂-fold differences [week 8 versus week 2 (baseline)] in the fecal abundances of CAZyme genes after consumption of the pea fiber snack prototype (panel A) and the orange fiber snack prototype (panel B). Participants are grouped based on hierarchical clustering (Euclidean distances) of their CAZyme gene profiles. Symbols with matching colors adjacent to the participant IDs denote members of the same twin pair; circles represent obese, while pentagons indicate non-obese participants, respectively.

participant during, and at the end of supplementation with pea fiber (panel b) and orange fiber (panel c). Participants are grouped based on hierarchical clustering (Euclidean distances) of their ASV profiles. n=18 and 24 participants for study 1 (pea fiber) and study 2 (orange fiber), respectively; n=126 fecal samples analyzed.

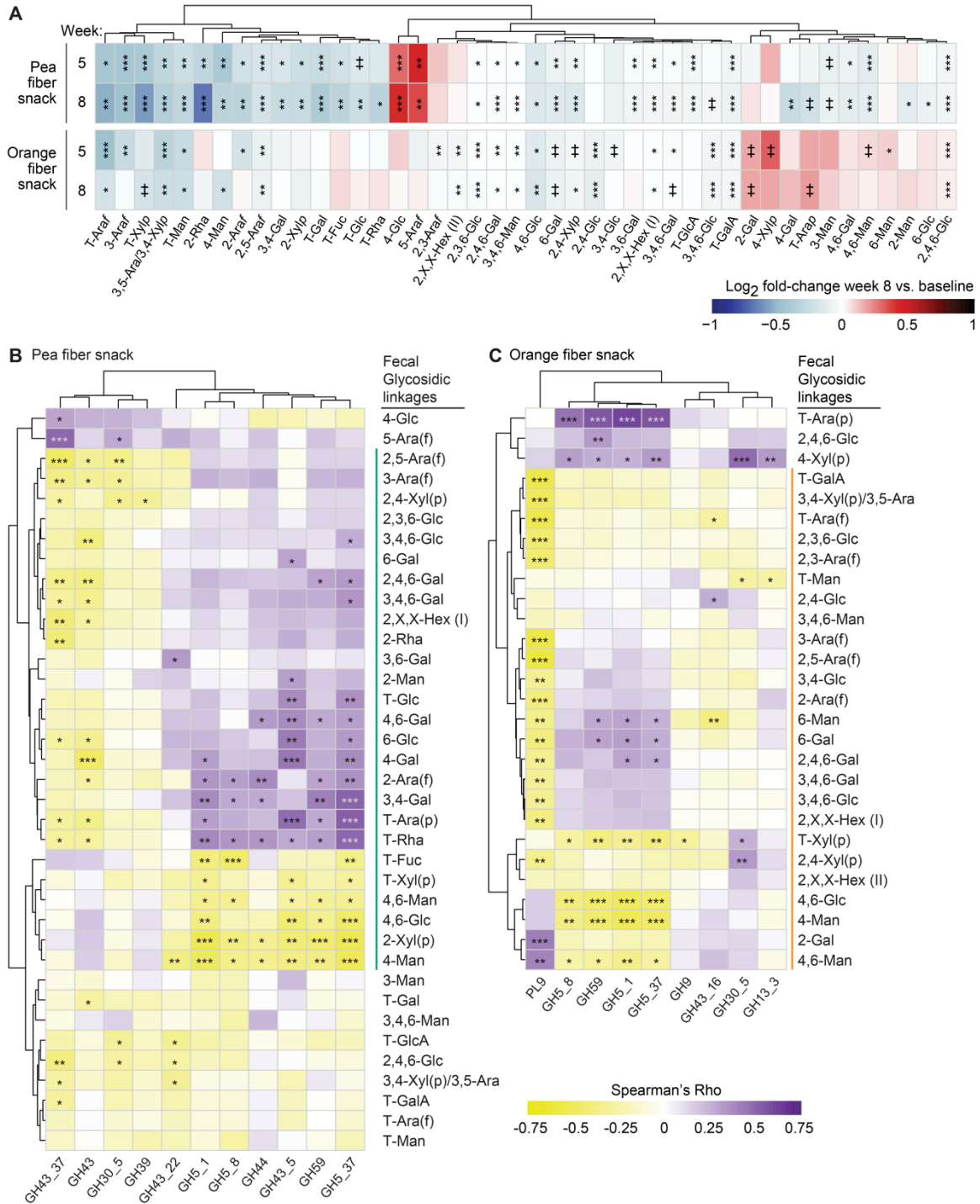


Fig S4. Correlating changes in the abundances of fecal glycosidic linkages with changes in the abundances of discriminatory CAZyme genes in participants consuming the pea or orange fiber snack prototypes. (A) Heatmap plotting the mean log₂ fold-change [week 8 versus pre-intervention, week 2] in fecal levels of glycosidic linkages in participants consuming the pea or orange fiber snack prototypes. Hierarchical clustering (Euclidean distances) was used to group linkages whose representation was similar after consumption of each fiber type. † $q < 0.10$ * $q < 0.05$, ** $q < 0.05$, *** $q < 0.001$ (linear mixed effects model; FDR-corrected). (B,C) Spearman

correlations between the abundances of fiber-responsive CAZyme genes and levels of glycosidic linkages, matched by time and participant (weeks 1, 5 and 8 in both studies). The green line in panel b indicates significant correlations between CAZyme genes that are significantly increased or decreased by one or the other treatment and pea fiber-associated glycosidic linkages. The orange line in panel c indicates significant correlations between these CAZyme genes and orange fiber-associated glycosidic linkages. Spearman correlation test: *, $P < 0.05$; **, $P < 0.01$, ***, $P < 0.001$. Abbreviations: terminal (T), pyranose (p), furanose (f), undefined linkage (X), glucose (Glc), arabinose (Ara), xylose (Xyl), galactose (Gal), mannose (Man), rhamnose (Rha), glucuronic acid (GlcA), galacturonic acid (GalA), hexose (Hex) and fucose (Fuc).

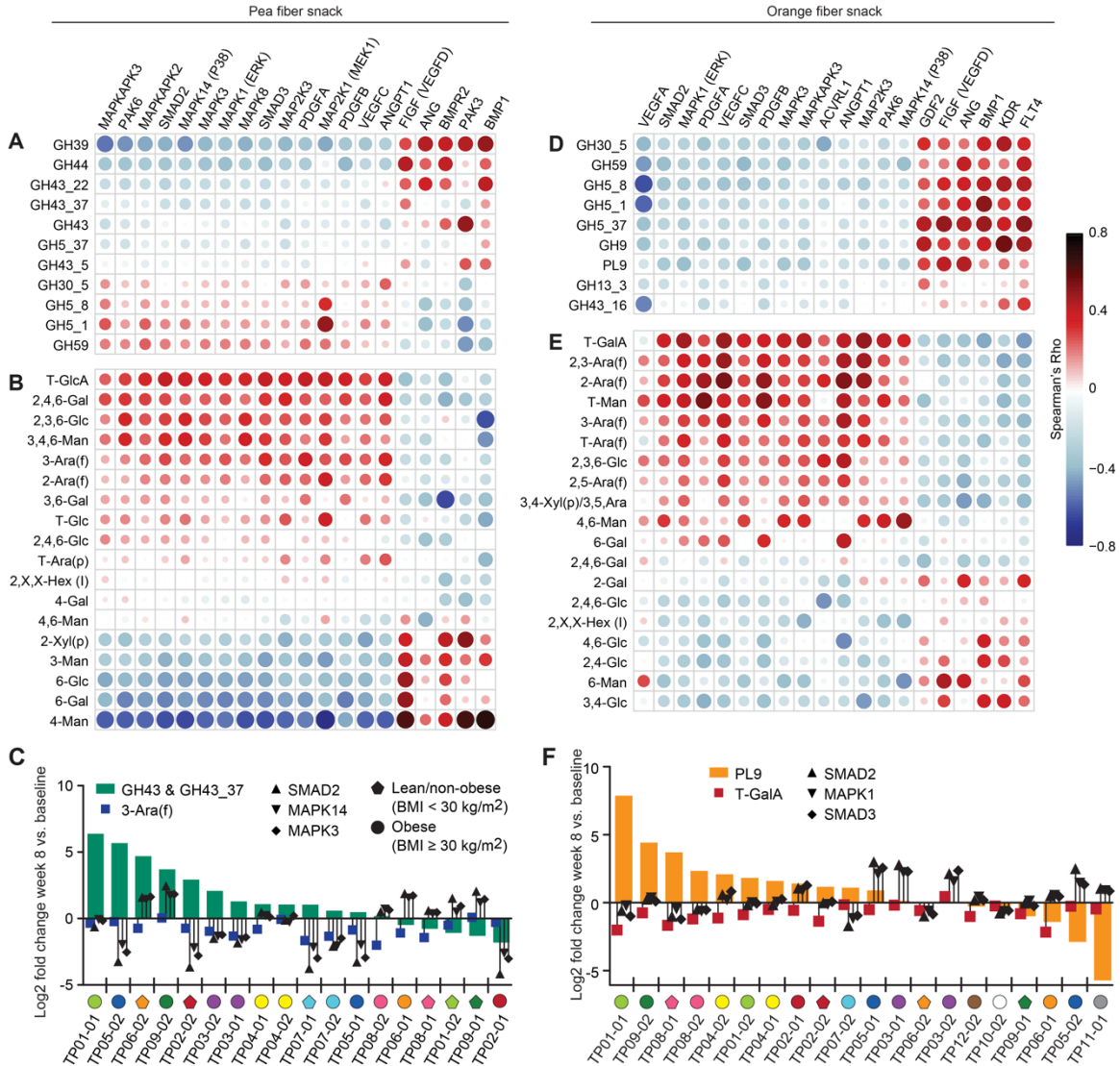


Fig S5. Correlating changes in host plasma proteins related to TGFβ/BMP1 signaling, p38/MAPK signaling and VEGF-mediated angiogenesis, CAZyme genes, and fecal glycosidic linkages. (A,D) Plasma proteins and CAZymes, **(B,E)** plasma proteins and glycosidic linkages in participants consuming the pea fiber **(A-B)** and the orange fiber **(D-E)** snack prototypes. The order of presentation of CAZymes in panel a and panel d follows their projection along SV1 in the CC-SVD analysis. Color denotes the direction of the correlation. The size of each circle and its color intensity represent the strength of the correlation. **(C)** Plot showing \log_2 fold-changes [week 8 versus pre-intervention] in the abundances of fecal microbiome GH43 plus GH43_37 genes, fecal glycans with 3-Araf linkages, and plasma levels of SMAD2, MAPK14 and MAPK3 in individual participants in the pea fiber study. **(F)** Analogous changes in PL9, T-GalA and SMAD2, SMAD3 and MAPK1 in the orange fiber study. Symbols with matching colors adjacent to the participant IDs denote members of the same twin pair; circles represent obese while pentagons indicate non-obese participants, respectively; n=18 and 20 participants for study 1 (pea fiber) and 2 (orange fiber), respectively; n=76 fecal samples and n=76 plasma samples analyzed.

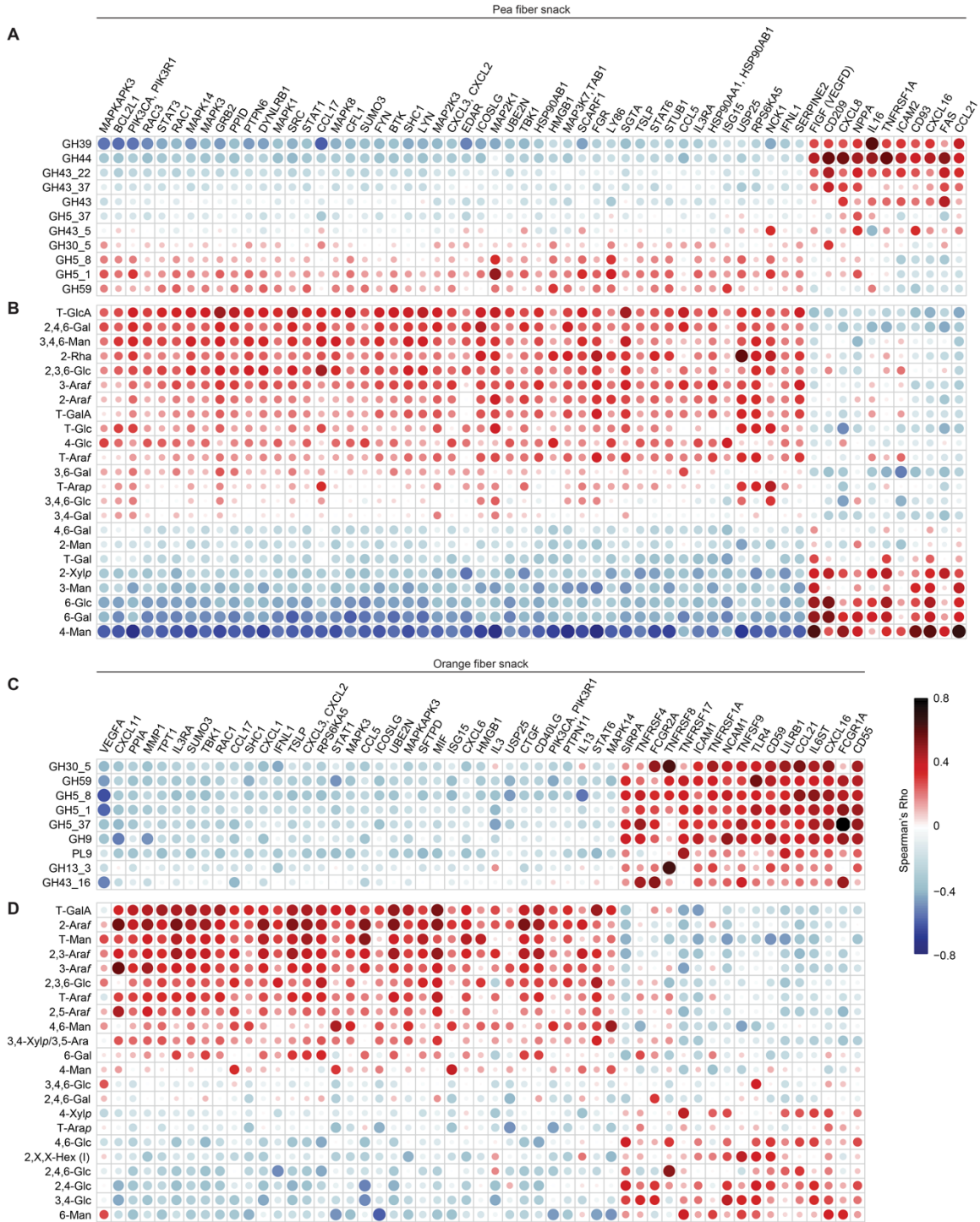


Fig S6. Correlating responses of plasma proteins related to immune biomarkers associated with interferon-stimulated genes, MyD88-dependent toll-like receptor signaling and lymphocyte/ dendritic cell markers, CAZyme genes, and fecal glycosidic linkages. Heatmaps plotting Spearman's Rho values of the correlation between changes in levels of plasma proteins and CAZyme gene abundances. Color denotes the direction of the correlation. The size of each circle and its color intensity represent the strength of the correlation. (A,C) and plasma proteins

and fecal glycosidic linkages (**B,D**) in participants after consuming the pea fiber (**A-B**) and orange fiber (**C-D**) snack prototypes, respectively. n=18 and 20 participants for study 1 (pea fiber) and 2 (orange fiber), respectively; n=76 fecal samples and n=76 plasma samples analyzed.

DATASETS

Dataset S1: Fiber snack food formulations and clinical meta-data from participants enrolled in each study.

- A. Nutritional composition of the fiber snack prototypes and carbohydrate composition of fibers.
- B. Participant characteristics.
- C. Effects of consumption of the fiber snack prototypes: clinical meta-data.
- D. Dietary histories of participants during study 1 (pea fiber snack).
- E. Dietary histories of participants during study 2 (orange fiber snack).
- F. Self-reported GI symptoms of participants in study 1 (pea fiber snack).
- G. Self-reported GI symptoms of participants in study 2 (orange fiber snack).

Dataset S2: Analysis of the representation of GH and PL genes in the fecal microbiomes of participants in each study.

- A. Shotgun sequencing datasets generated from the fecal microbiomes of participants in each study.
 - i. CAZyme gene representation in the fecal microbiomes of participants in study 1.
 - ii. CAZyme gene representation in the fecal microbiomes of participants in study 2.
- B. Reported or predicted function(s) of detected CAZyme genes.
- C. HOSVD analyses of datasets of CAZyme gene abundances generated from participants in each study.
 - i. CAZyme- timepoint projections for study 1.
 - ii. CAZyme- feature projections for study 1.
 - iii. CAZyme- ranked feature projections for study 1.
 - iv. CAZyme- participant projections for study 1.
 - v. Study 1 statistics for fiber discriminatory CAZymes.
 - vi. CAZyme- timepoint projections for study 2.
 - vii. CAZyme- feature projections for study 2.
 - viii. CAZyme- ranked feature projections for study 2.
 - ix. CAZyme- participant projections for study 2.
 - x. Study 2 statistics for fiber discriminatory CAZymes.

Dataset S3: Analysis of the representation of bacterial taxa in the fecal microbiota of participants in each study.

- A. Taxonomic summary of V4-16S rDNA ASV sequences.
- B. V4-16S rDNA amplicon sequencing datasets generated from the fecal microbiota of participants in each study.
 - i. Relative abundances of ASVs in the fecal microbiota of participants enrolled in study 1.
 - ii. Relative abundances of ASVs in the fecal microbiota of participants enrolled in study 2.
- C. HOSVD analyses of datasets of ASV relative abundances generated from participants in each study.

- i. ASV- timepoint projections for study 1.
 - ii. ASV- feature projections for study 1.
 - iii. ASV- ranked feature projections for study 1.
 - iv. ASV- participant projections for study 1.
 - v. Study 1 statistics for fiber discriminatory ASVs.
 - vi. ASV- timepoint projections for study 2.
 - vii. ASV- feature projections for study 2.
 - viii. ASV- ranked feature projections for study 2.
 - ix. ASV- participant projections for study 2.
 - x. Study 2 statistics for fiber discriminatory ASVs.
- D. Short chain fatty acid levels (in $\mu\text{mol/g}$) generated from fecal samples of participants in each study.
- i. Study 1: Pea fiber snack.
 - ii. Study 2: Orange fiber snack.
- E. Results of cross-correlation analysis between levels of short chain fatty acids and abundances of fiber-snack treatment-responsive ASVs in participants from each study.
- i. Study 1: Pea fiber snack.
 - ii. Study 2: Orange fiber snack.
- F. Bray-Curtis dissimilarity distances of square root transformed ASV and CAZyme datasets generated from fecal samples of participants in each study.
- i. Study 1: Pea fiber snack.
 - ii. Study 2: Orange fiber snack.

Dataset S4: Mass spectrometric analyses of fecal samples.

- A. UHPLC-QqQ-MS of glycosidic linkages in fecal samples collected from participants in each study.
- B. Statistics for the fecal glycosidic linkages for participants in each study.

Dataset S5: Results of cross-correlation analysis between levels of fecal glycosidic linkages and abundances of fiber snack treatment-responsive CAZyme genes

- A. Results from participants consuming the pea fiber snack.
- B. Results from participants consuming the orange fiber snack.

Dataset S6: Analysis of plasma proteome datasets generated from participants in each study.

- A. Log₂ of the relative fluorescence units (RFU) of SOMAmer-bound plasma proteins at baseline (Week 1) and at the end of pea fiber treatment (Week 8) in participants consuming the pea fiber snack prototype (study 1).
- B. Log₂ of the relative fluorescence units (RFU) of SOMAmer-bound plasma proteins at baseline (Week 1) and at the end of orange fiber treatment (Week 8) in participants consuming the orange fiber snack prototype (study 2).
- C. Description of target proteins.
- D. Correlating abundances of fiber responsive CAZyme genes to host responses defined by changes in the plasma proteome.
 - i. Results of CCSVD analysis of the abundances of treatment-responsive fecal microbiome CAZyme genes and levels of plasma proteins in participants consuming the pea fiber snack.

- ii. Results of CCSVD analysis of the abundances of treatment-responsive fecal microbiome CAZyme genes and levels of plasma proteins for participants consuming the orange fiber snack.
- iii. List of CAZyme-correlated biological themes identified by CompBio analysis enriched in the plasma proteomes of participants consuming the pea fiber snack.
- iv. List of CAZyme-correlated biological themes identified by CompBio analysis as enriched in the plasma proteomes of participants consuming the orange fiber snack.
- v. Comparison of biological themes enriched in each of the two studies.

Dataset S7: Correlating levels of fiber snack-associated fecal glycosidic linkages to plasma proteomic changes.

- A. Results of cross-correlation analysis between levels of glycosidic linkages and the abundances of CAZyme gene-associated plasma proteins in participants consuming the pea fiber snack.
- B. Results of cross-correlation analysis between levels of glycosidic linkages and the abundances of CAZyme gene-associated plasma proteins in participants consuming the orange fiber snack.