Supplementary Information for *Relative enrichment of oral bacteria in feces denounces loss of gut commensals with implications to host health*

LIAO ET AL.

The supplementary material includes **Supplementary Texts**, **Supplementary Figures**, **Supplementary Tables**, and **Supplementary References**.

SUPPLEMENTARY TEXTS

In Fig. 3 of the main text, we applied a filter based on relative abundance and prevalence to identify ASVs typical of the oral cavity. The rationale of filtering is that the more typical ASV of the oral cavity, the less likely it colonizes the gut due to niche specificity, and the more likely its presence in feces indicates oral-gut translocation. To show that filtering is key to 16S rRNA-based inference of oral bacteria, we took an alternative approach by counting every single ASV in feces as of oral origin if it is also found in any of paired HMP oral cavity subsite samples. This naïve approach led to an average of 247 oral ASVs per HMP fecal sample and their total fraction averaged across all HMP subjects rises >300-fold from 0.05% to 15.6%. The severe overestimation suggests that the majority of shared ASVs between human gut and oral cavity are not indicators of oral-gut transmission but coincidences of the same or closely related bacteria occupying both niches.

To validate our approach in patients, we used a public study [1] that has sequenced paired oral (saliva)-gut (stool) samples. The subjects in the study included patients with Crohn's disease (CD) and ulcerative colitis (UC) as well as their own healthy controls (HC). We first showed that the estimated oral bacterial fractions in feces were robust against variations of the filtering cutoffs used to generate the reference set of oral bacterial sequences from the HMP data (Fig. S4A). Using the default reference set (used throughout the study), we found that the oral bacterial fractions in the feces of patients with inflammatory bowel disease (IBD) were averagely 4 times (CD: 4.2%, UC: 4.3%) as high as the fractions in the feces of HC (1.1%) (Fig. S4B), confirming the notion that IBD patients are enriched with oral bacteria in their gut [2]. Among 99 fecal samples (HC:41, CD:16, UC:42) that contained at least one oral ASV, the proportion of oral ASVs that were also found in paired saliva samples is >70% in 87 samples (HC:36, CD:15, UC:36; Fig. S4C). Similarly, >70% of the total fractions of oral ASVs in 90 out of 99 fecal samples was contributed by those found in the paired saliva samples (HC:38, CD:15, UC:37; Fig. S4D). Both computational validations indicate that our approach to detect oral-gut transmission is conservative and less prone to false-positive predictions (i.e., inferred oral ASVs absent from the oral cavity of the same person).

SUPPLEMENTARY FIGURES

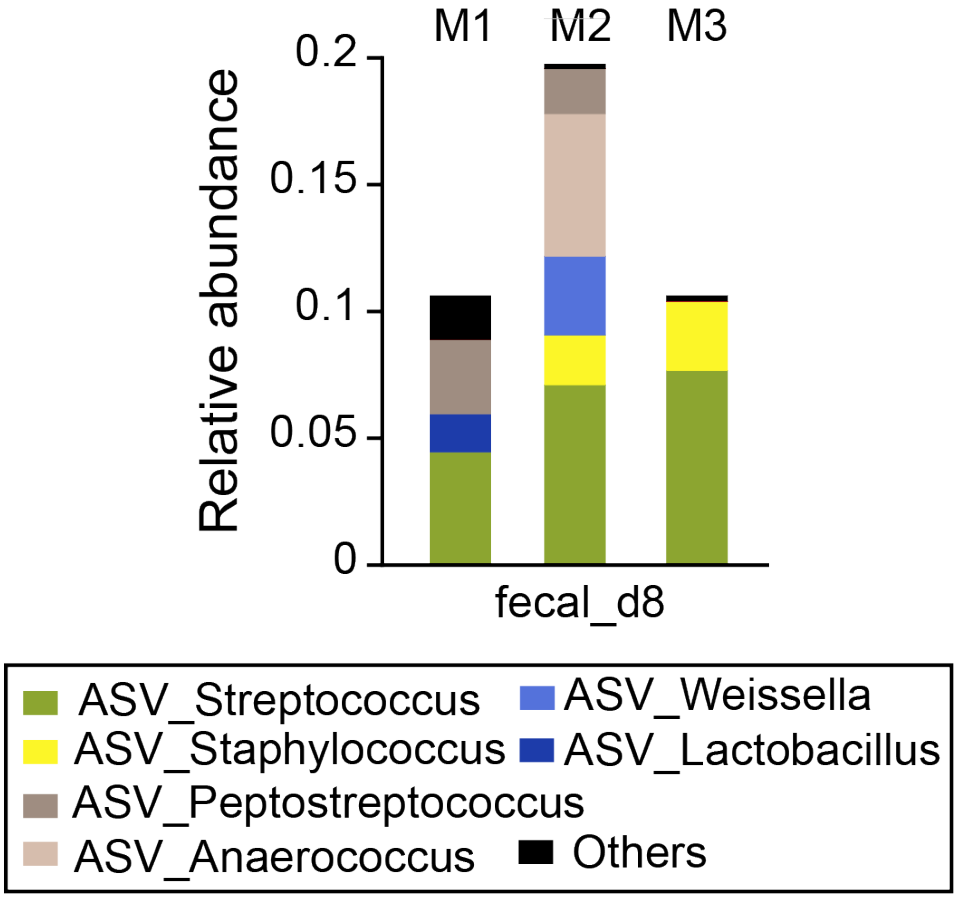

Fig. S1. (Related to Fig. 2) Taxonomy of ASVs in the post-treatment fecal samples that were found in the pre-treatment oral samples, but not the pre-treatment fecal samples, of the same mice. M1, M2, and M3 represent three mice. fecal_d8: post-treatment fecal sample on day 8.

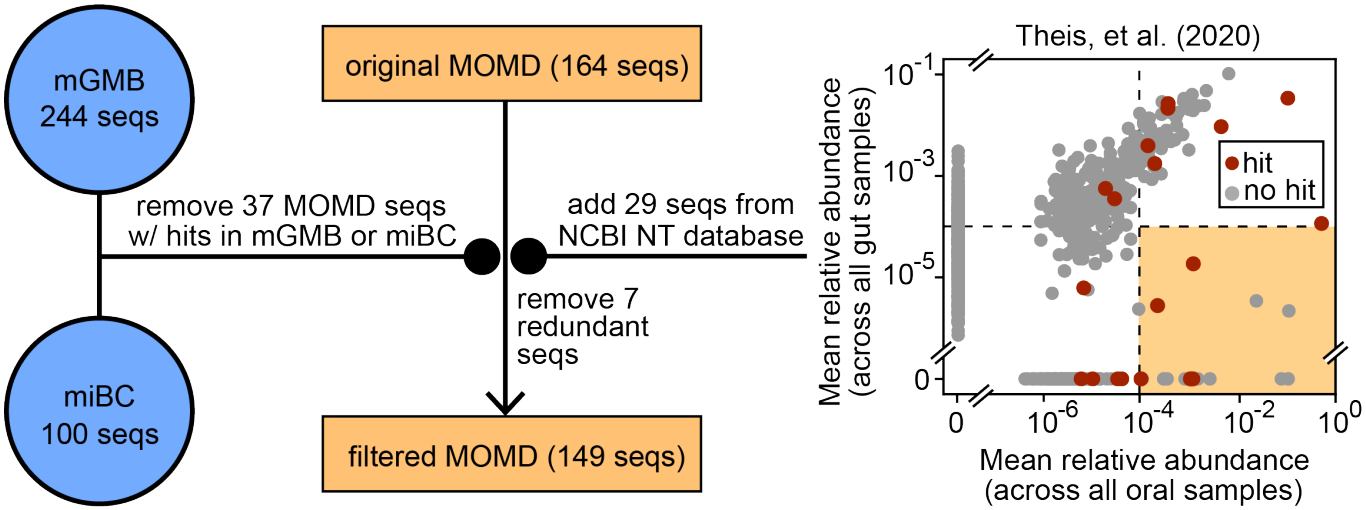

Fig. S2. (Related to Fig. 2) The schematic of our approach for constructing a reference set of oral bacterial 16S rRNA sequences in mouse. Briefly, we started with 164 full length 16S rRNA sequences from cultivable oral bacteria in MOMD (Mouse Oral Microbiome Database [3]). We removed sequences from cultivable gut bacteria in mouse (those collected in mGMB and miBC) and added a few new sequences missed by MOMD but representative of the oral cavity of 11 mice in a public study [4]. In the scatter plot, hits refer to ASVs in the 11 mice matching at least one sequence in the original MOMD. The dashed lines (equal to 10^{-4}) represent the relative abundance cutoffs to identify oral-typical ASVs (those in the shading area). Here we did not use prevalence-based filter due to the small sample size. mGMB: the mouse Gut Microbial Biobank [5]; miBC: the mouse intestinal Bacterial Collection [6]; Seqs: sequences. See STAR Methods of the main text for details.

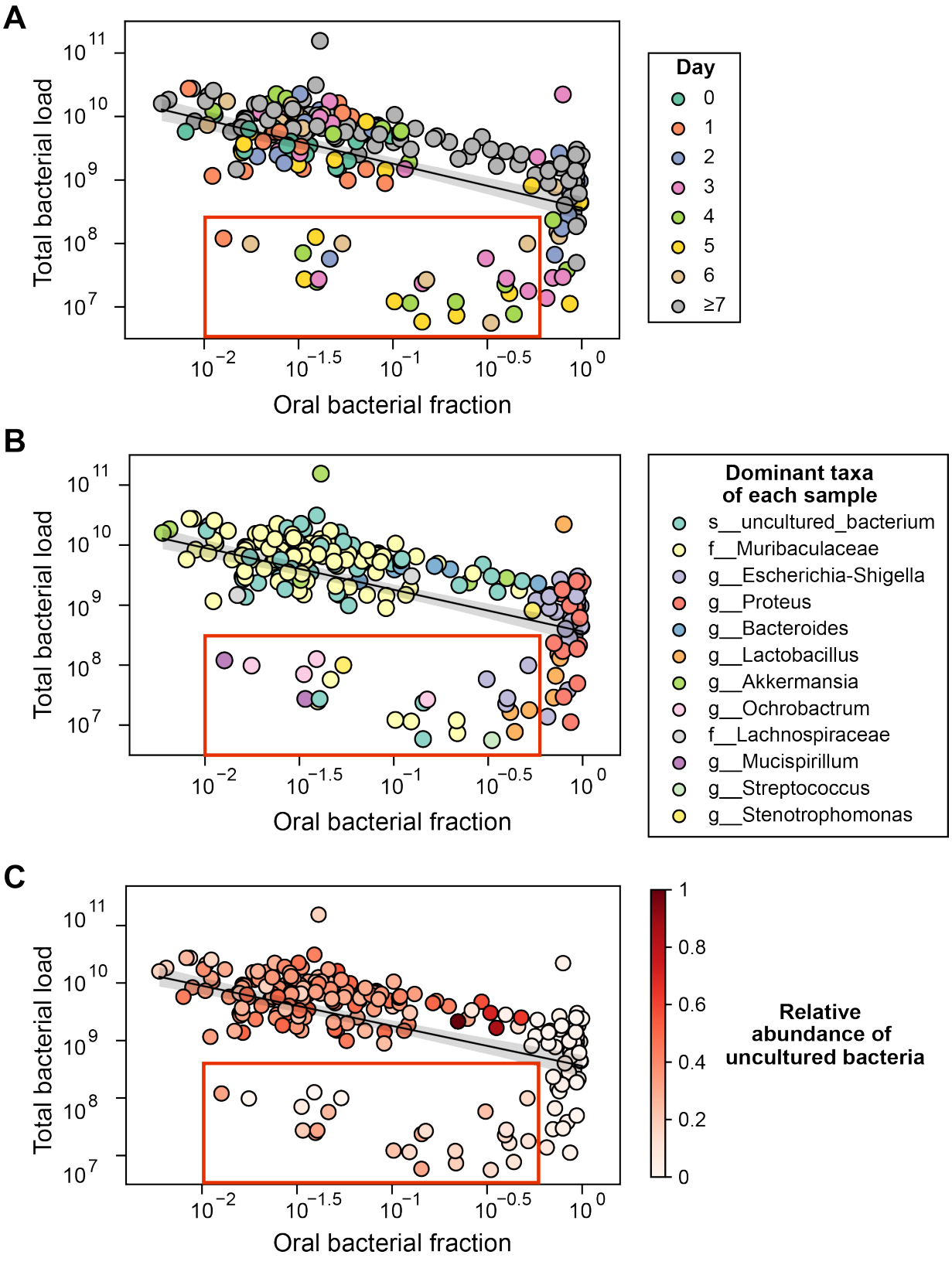

Fig. S3. (Related to Fig. 2) In Fig. 2J of the main text, we noted that 11% samples (outlined in red box) with low fractions of oral bacteria in feces ($< 10^{-0.25}$) and low total bacterial loads ($< 10^{8.5}$) deviated from the trend line of the association (black line: best linear fit; gray shading: 95% confidence interval). According to the *marker* hypothesis, they should be detected with very high proportions of oral bacteria. These samples were collected between day 1-6 (**A**). We speculated that the most abundant ASVs of these outliers (**B**) were orally derived but undetected by our inference approach. The omit of uncultured bacteria in the reference set is unlikely the major cause of the potential inference failure, because the proportions of uncultured bacteria based on taxonomic annotation are low to intermediate in these samples (**C**). Unit of total bacterial load: 16S copies per gram feces.

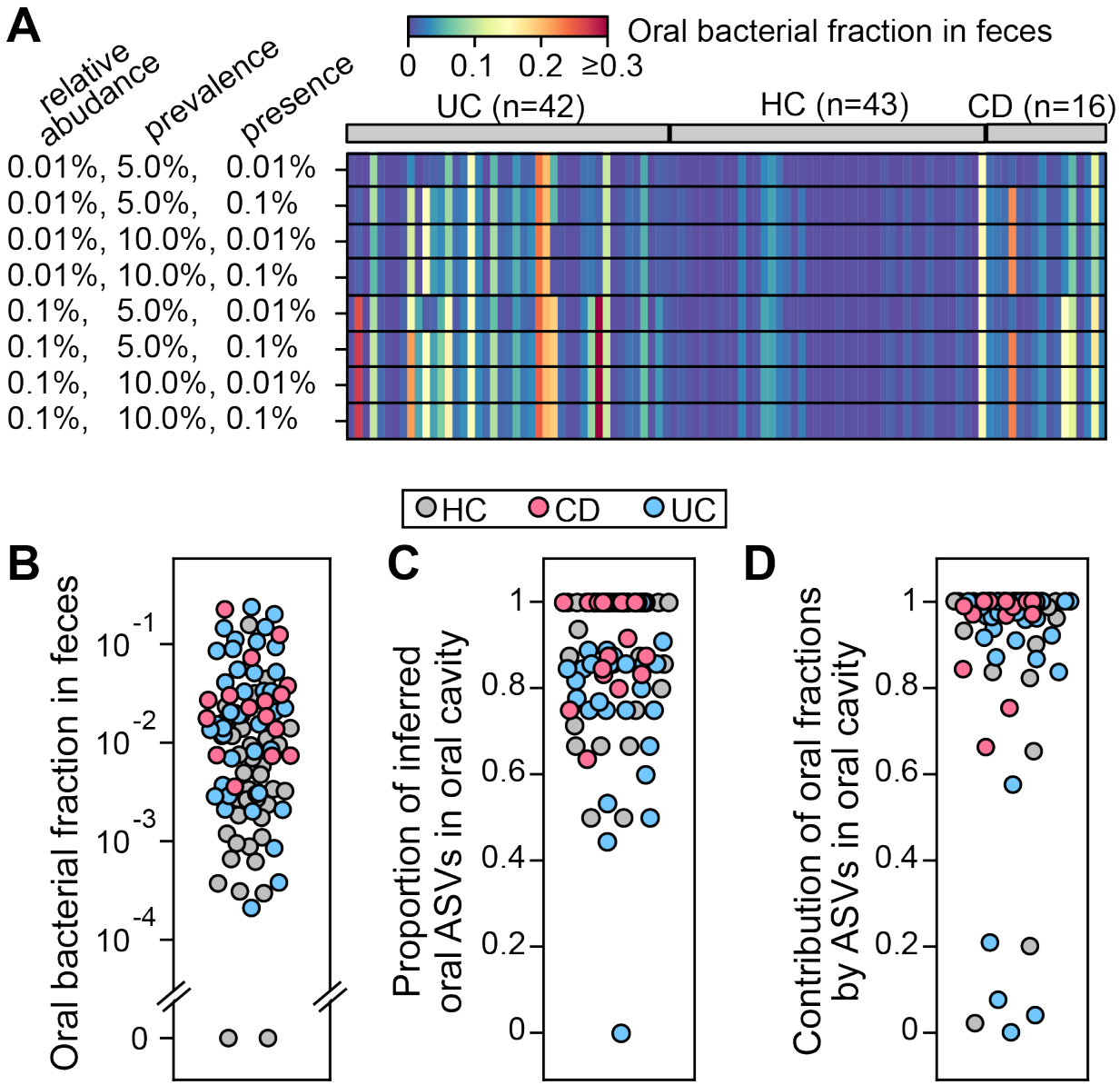

Fig. S4. (Related to Fig. 3) Validations of our approach to identify oral bacteria in human feces using a cohort of patients with inflammatory bowel disease and their healthy controls [1]. (**A**) Effects of filtering cutoffs to construct the reference 16S rRNA sequences of oral-typical bacteria. We varied the thresholds of mean relative abundance (0.01%, 0.1%), prevalence (5%, 10%), and the definition of ASV presence (0.01%, 0.1%) in the computation of prevalence. The reference oral-typical ASVs generated by each combination of the thresholds was used to compute the total fractions of oral bacteria in the feces of all subjects. We saw that the estimated oral proportions formed vertical colored bands in the heatmap, suggesting that they are largely insensitive to the variations in filtering thresholds. CD: Crohn's disease; UC: ulcerative colitis; HC: healthy controls. (**B**) A strip plot showing the distributions of oral bacterial fraction in the feces. (**C**) Proportions of fecal ASVs inferred as of oral origin that can also be found in paired saliva samples. (**D**) Total fractions of inferred oral ASVs contributed by those found in the paired saliva samples (total fraction of inferred oral ASVs that were also present in the paired oral samples divided by total fraction of all inferred oral ASVs).

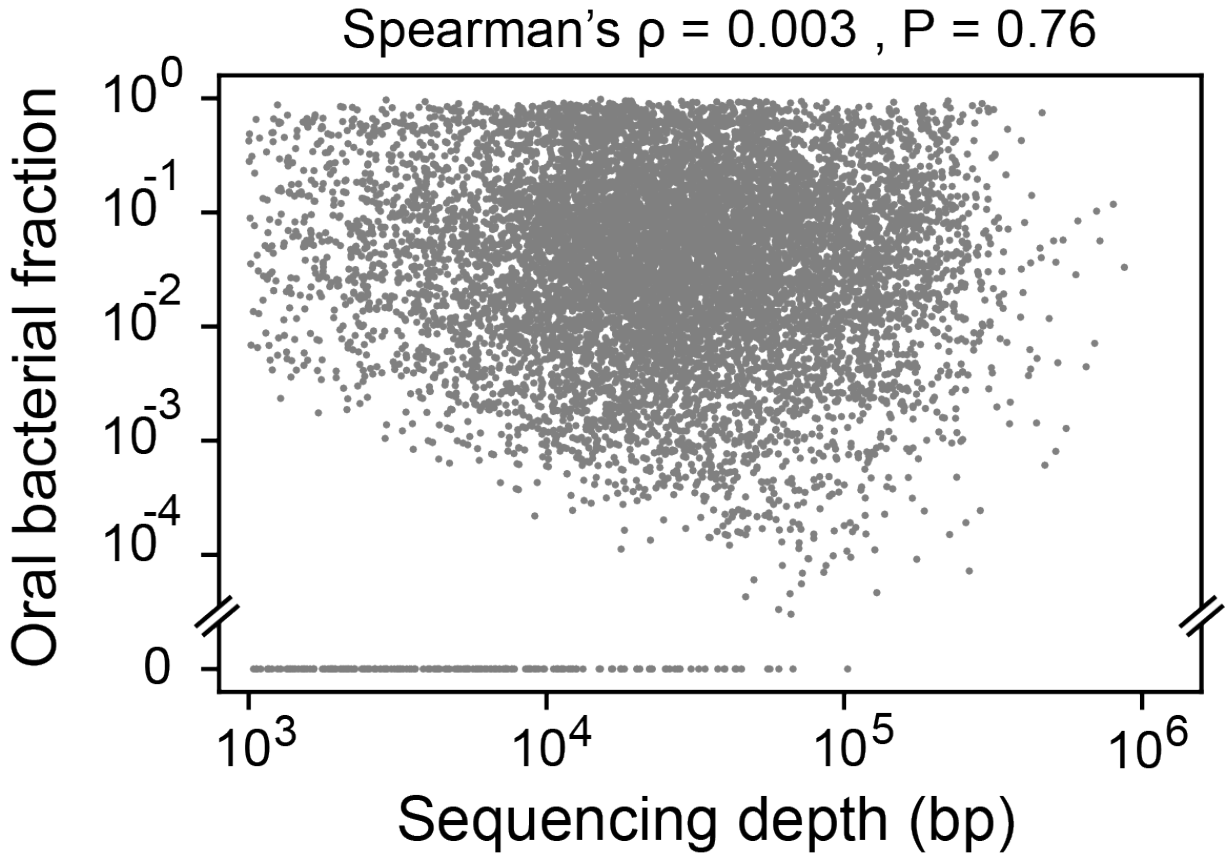

Fig. S5. (Related to Fig. 4) Total fractions of oral bacteria in feces are not associated with sequencing depths. Each dot represents a fecal sample from adult allo-HCT (allogeneic hematopoietic cell transplantation) recipients hospitalized at Memorial Sloan Kettering Cancer Center.

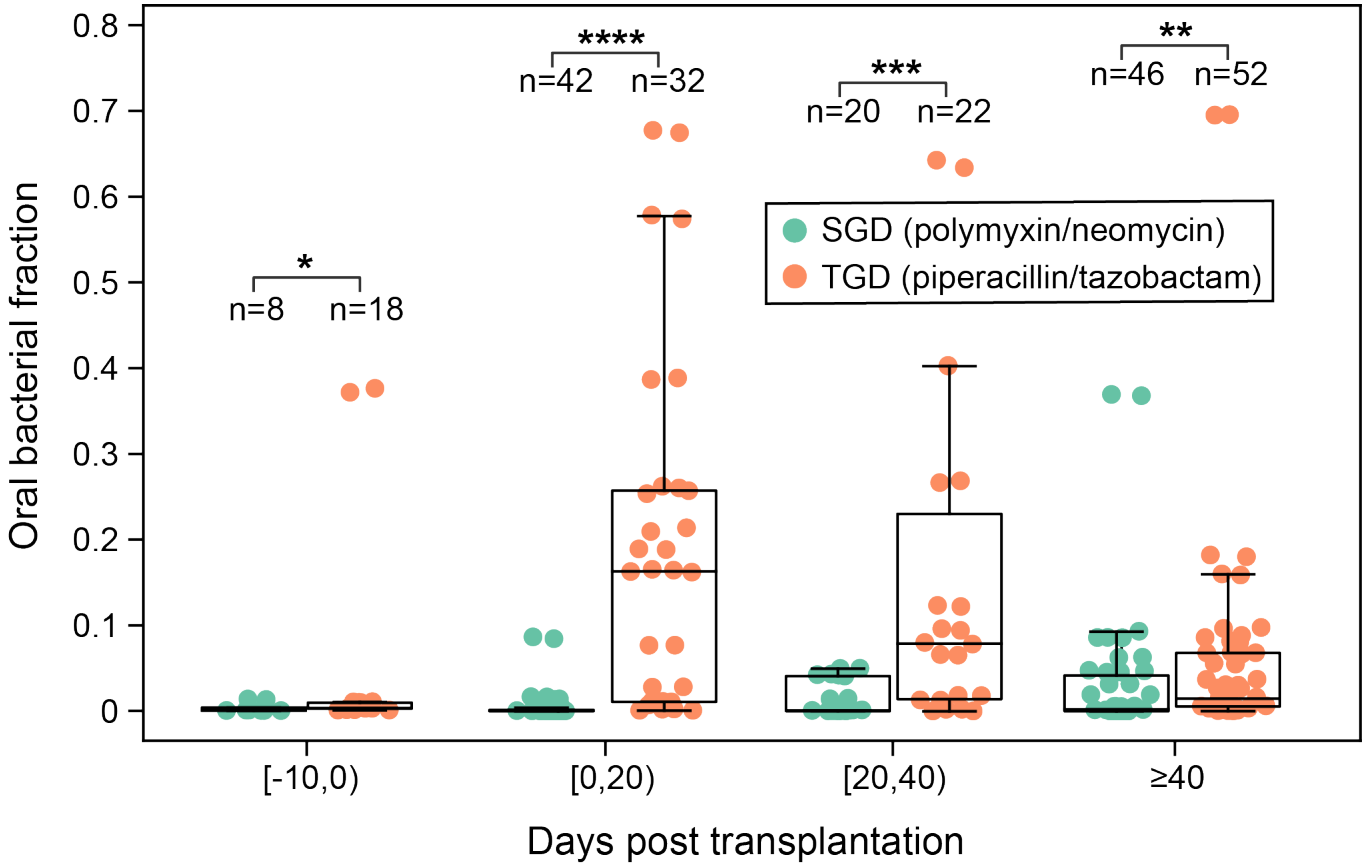

Fig. S6. (Related to Fig. 4) Piperacillin/tazobactam (TZP) caused significant enrichment of oral bacteria in the feces of 19 pediatric allo-HCT (allogeneic hematopoietic cell transplantation) recipients (1-17 year old, 10.1 year old on average) [7]. Samples were grouped into four different stages of transplantation. TZP was used in total gut decontamination (TGD) as the only antibacterial antibiotic, while polymycin/neomycin was used in selective gut decontamination (SGD). The administration of SGD and TGD started 10 days before transplantation until engraftment or 21 days after transplantation, whichever occurred later. The comparison is visualized using box plot overlaid on top of a strip plot (each dot represents a fecal sample), with central line representing the median, box limits representing the first and third quartiles, and whiskers extending to the smallest and largest values or at most to 1.5× the interquartile range, whichever is smaller. ****, P<0.0001; ***, P<0.001, **, P<0.01; *, P<0.05. Kruskal-Wallis test.

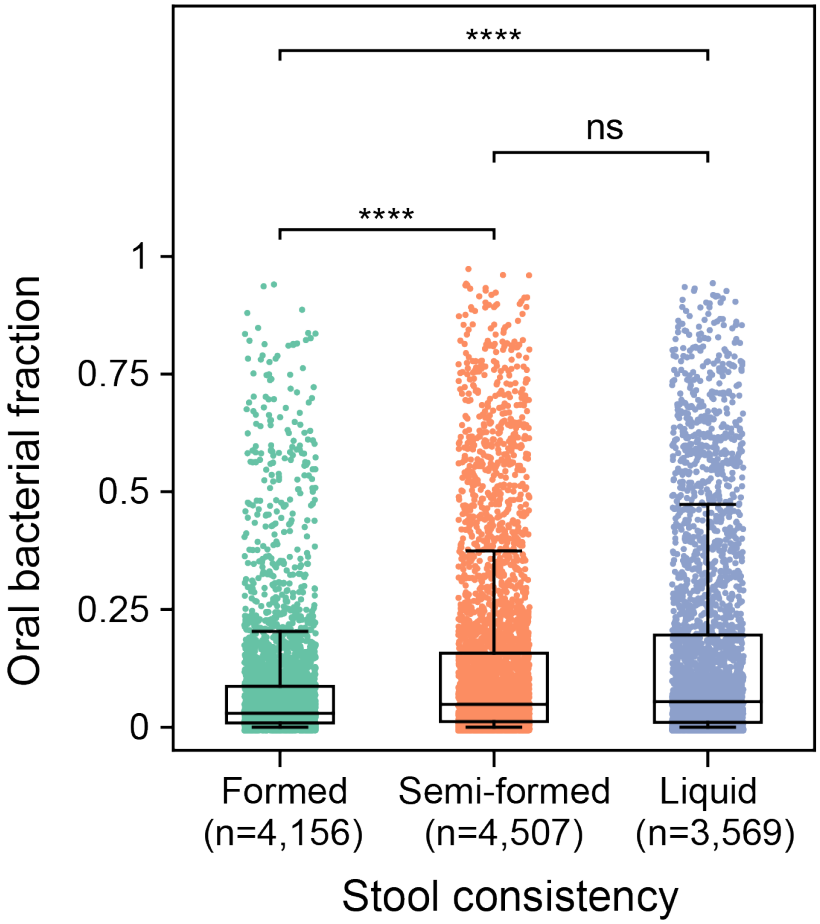

Fig. S7. (Related to Fig. 4) Distribution of oral bacterial fractions in feces stratified by stool consistency. Each dot represents a a fecal sample from adult allo-HCT (allogeneic hematopoietic cell transplantation) recipients hospitalized at Memorial Sloan Kettering Cancer Center. The comparison is visualized using box plot overlaid on top of a strip plot (each dot represents a fecal sample), with central line representing the median, box limits representing the first and third quartiles, and whiskers extending to the smallest and largest values or at most to 1.5× the interquartile range, whichever is smaller. ****, P<0.0001; ns, not significant. Kruskal-Wallis test.

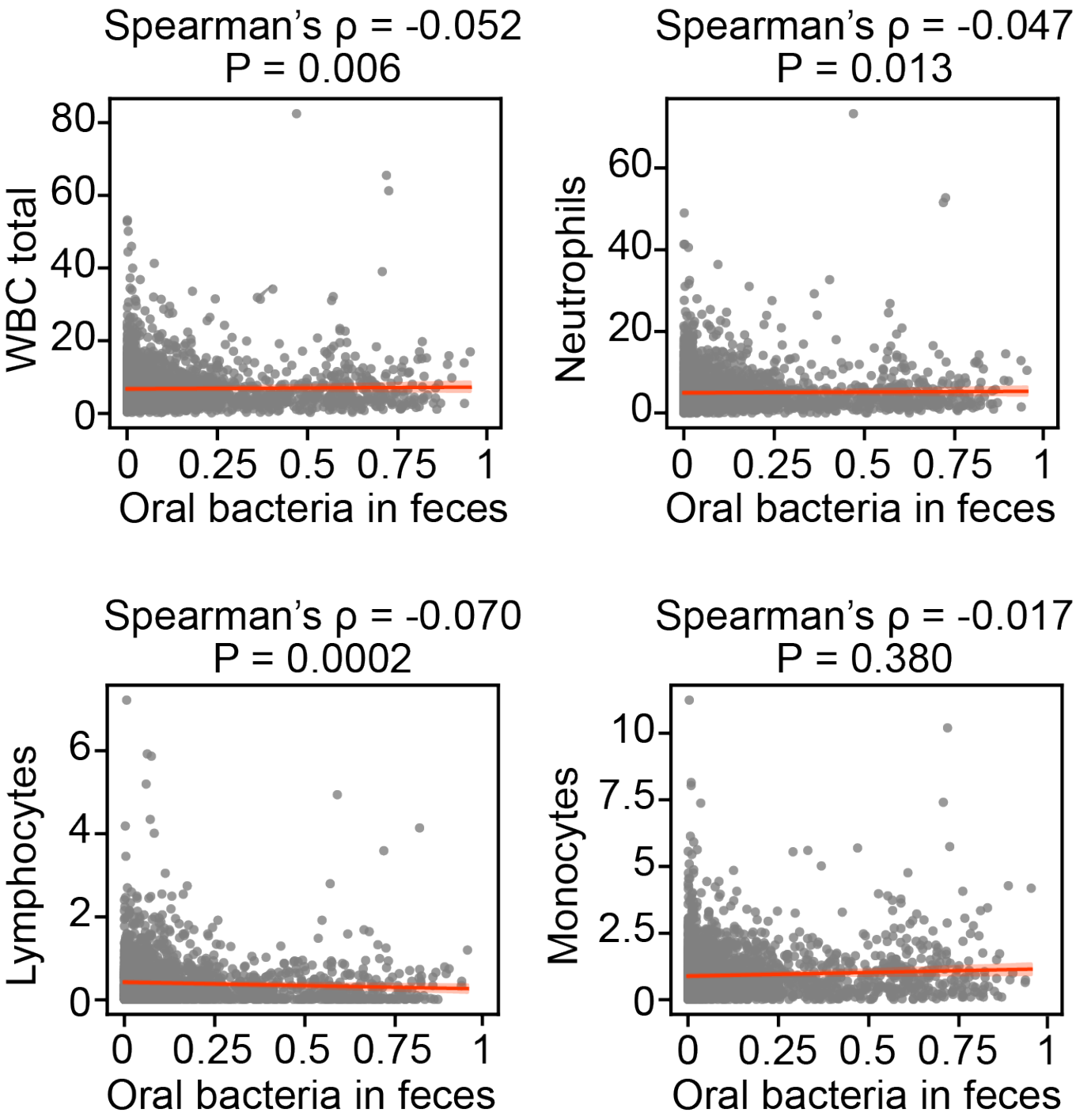

Fig. S8. (Related to Fig. 4) Correlation between proportions of oral bacteria in feces and circulating white blood cell counts measured on the same days from adult allo-HCT (allogeneic hematopoietic cell transplantation) recipients hospitalized at Memorial Sloan Kettering Cancer Center. Each dot represents a sample collected between neutrophil engraftment and 100 days post engraftment. Red lines: best linear fit; shadings: 95% confidence interval. WBC total: total white blood cell counts. Unit of white blood cells: K/μL.

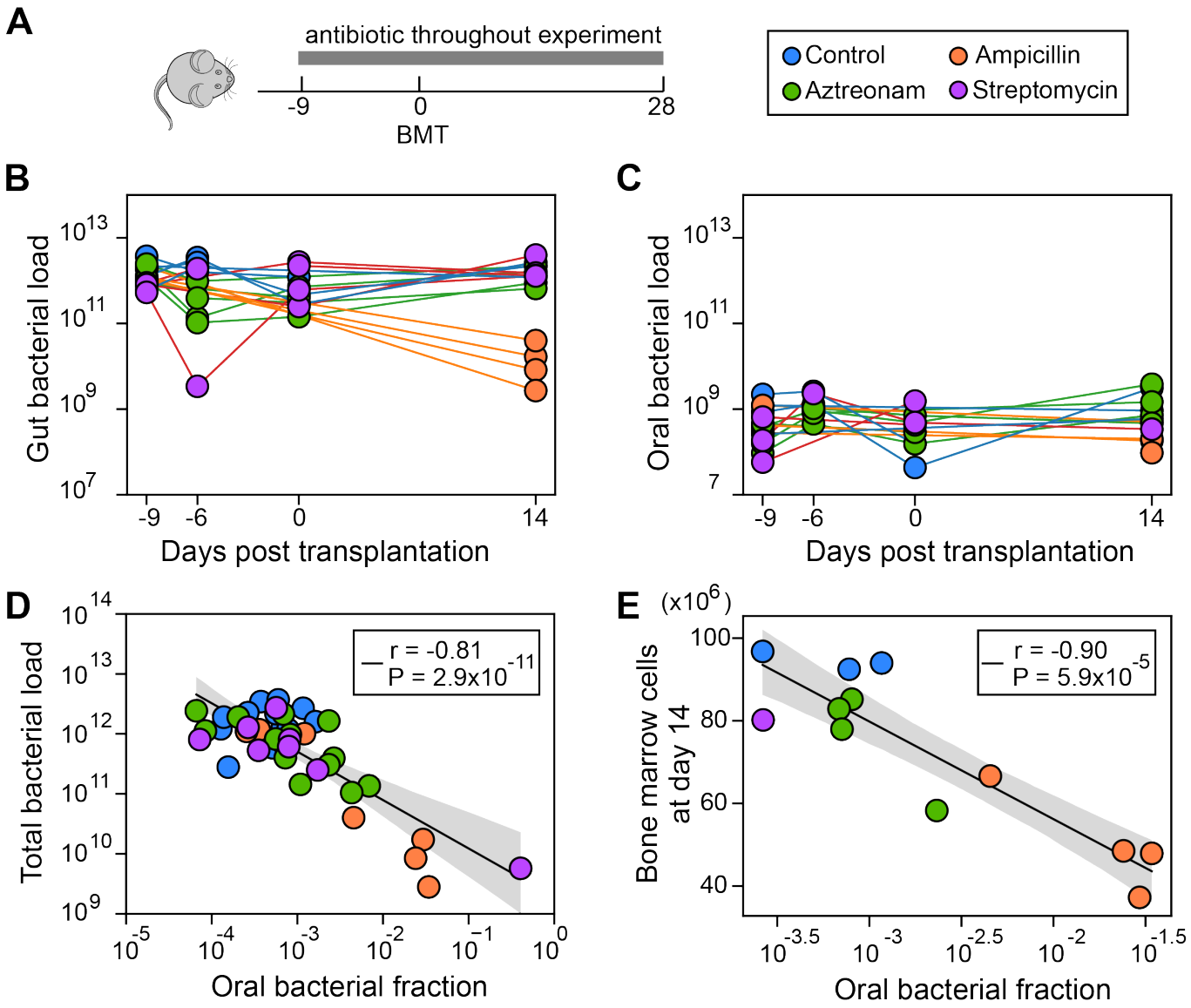

Fig. S9. (Related to Fig. 4) Correlations between oral bacteria in feces and host immunity in bone marrow transplanted mice [8]. (**A**) Experiment design. Mice were given a split 1100cGy radiation dose and administered 5×10^6 bone marrow cells via tail vein injection at day 0. Ampicillin, aztreonam, and streptomycin were started 9 days before bone marrow transplantation (BMT) and administered throughout the experiment. Except for ampicillin (4 mice), all other antibiotic treatments (including the control group) were repeated in 5 mice. (**B, C**) Dynamics of gut (B) and oral (C) bacterial loads during BMT. Each trajectory represents a mouse. (**D**) Linear relationship between oral bacterial fractions and total bacterial loads in the log-log space. Black line: best linear fit; shading: 95% confidence interval (CI). (**E**) Negative linear relationship between log-transformed oral bacterial fractions in feces and bone marrow cells at day 14. Black line: best linear fit; shading: 95% CI. Samples that do not contain oral ASVs were omitted in panels D and E. Unit of bacterial loads: 16S copies per gram feces.

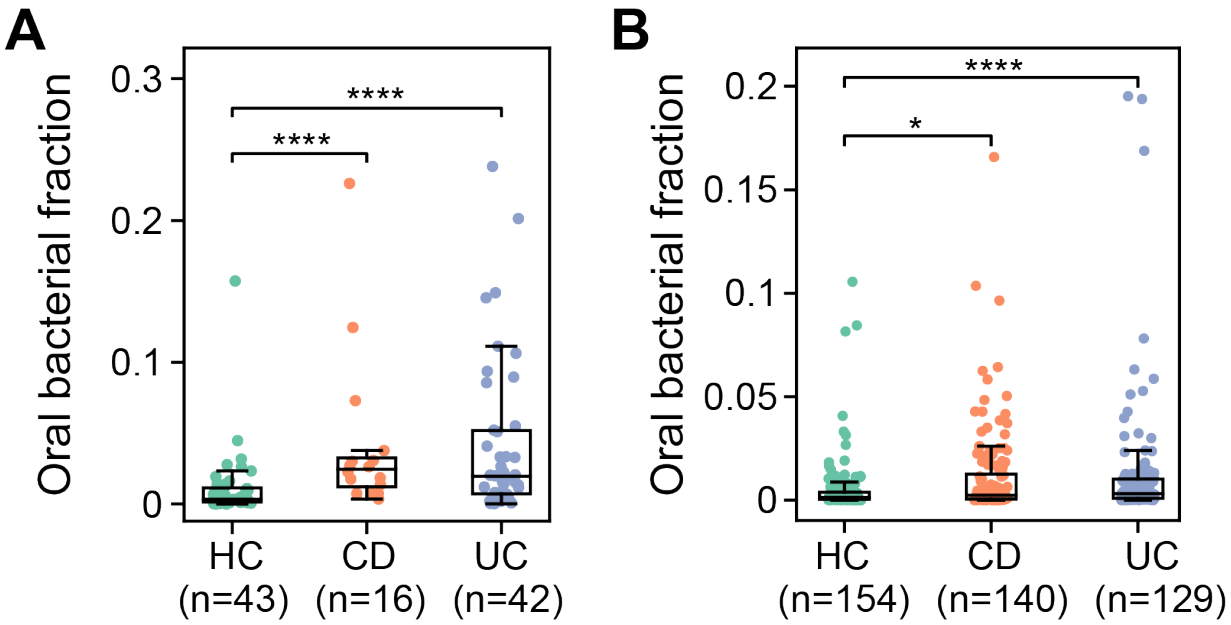

Fig. S10. (Related to Fig. 5) Oral bacteria showed significantly higher enrichment in the feces of patients with Crohn's disease (CD) and ulcerative colitis (UC) compared to their healthy controls (HC). (**A**) Imai et al. [1]. Participants who took any antibiotics within the last 3 months were excluded. (**B**) Pascal et al. [9]. Participants who were treated by antibiotics within the previous 4 weeks were excluded. The comparison is visualized using box plot overlaid on top of a strip plot (each dot represents one sample), with central line representing the median, box limits representing the first and third quartiles, and whiskers extending to the smallest and largest values or at most to 1.5× the interquartile range, whichever is smaller. ****, P<0.0001; *, P<0.05. Kruskal-Wallis test.

SUPPLEMENTARY TABLES

Predictor	Event	HR (95% CI)	*P*
Oral bacterial domination	*Enterococcus* domination	0.13 (0.03-0.52)	0.004
Oral bacterial domination	*Candida* domination	8.50 (4.43-16.31)	1.21×10^{-10}

Table S5. Intestinal domination of oral bacterial ASVs (relative abundance > 30%) reduced the risk of intestinal domination of *Enterococcus* ASVs (relative abundance > 30%), but increased the risk of intestinal domination of *Candida* ASVs (relative abundance > 90%). Text in red color: significant positive associations; text in blue color: significant negative association; CI: confidence interval; HR: hazard ratio.

Predictor	WBC total		Neutrophils		Lymphocytes		Monocytes	
	Estimate (95% CI)	*P*	Estimate (95% CI)	*P*	Estimate (95% CI)	*P*	Estimate (95% CI)	*P*
Oral bacterial fraction	0.112 (0.013-0.211)	0.027	0.125 (0.001-0.248)	0.048	0.122 (-0.069-0.313)	0.210	0.163 (0.033-0.292)	0.014
WBC count	0.013 (0.009-0.016)	6.15×10^{-13}	0.019 (0.014-0.024)	1.26×10^{-12}	-0.057 (-0.132-0.018)	0.134	0.006 (-0.022-0.034)	0.680
Transplant type								
PBSC unmodified	0.076 (0.011-0.142)	0.022	0.111 (0.029-0.193)	0.008	0.057 (-0.063-0.176)	0.355	-0.009 (-0.089-0.072)	0.836
TCD	0.112 (0.041-0.183)	0.002	0.158 (0.070-0.247)	0.0005	0.138 (0.008-0.268)	0.038	0.006 (-0.082-0.095)	0.886
umbilical cord	-0.098 (-0.175 – -0.020)	0.013	-0.090 (-0.187-0.007)	0.069	-0.032 (-0.174-0.110)	0.660	-0.156 (-0.251 – -0.060)	0.001
Immunomodulator								
GCSF	1.019 (0.975-1.064)	0	1.220 (1.165-1.275)	1.322×10^{-314}	0.309 (0.225-0.393)	7.004×10^{-13}	0.523 (0.466-0.581)	1.677×10^{-67}
mycophenolate mofetil	-0.089 (-0.147 – -0.031)	0.003	-0.112 (-0.184 – -0.040)	0.002	-0.070 (-0.179-0.038)	0.204	-0.032 (-0.105-0.041)	0.388
cetirizine	0.127 (-0.003 – -0.256)	0.055	0.123 (-0.038 – -0.285)	0.134	0.097 (-0.148-0.341)	0.437	0.139 (-0.026-0.305)	0.099

Table S6. Total fraction of oral bacteria in the gut microbiome is positively associated with net production rate of total white blood cells (WBC total), neutrophils and monocytes. Other clinical predictors include counts of each white blood cell type (self interaction), transplant type (unmodified bone marrow vs. others), and three immunomodulators included in our previous study [10]. Text in red color: significant positive association; text in blue color: significant negative association. PBSC: peripheral blood stem cell; TCD: T-cell depleted; GCSF: granulocyte-colony stimulating factor; CI: confidence interval.

SUPPLEMENTARY REFERENCES

1. J. Imai, H. Ichikawa, S. Kitamoto, J. L. Golob, M. Kaneko, J. Nagata, M. Takahashi, M. G. Gillilland III, R. Tanaka, H. Nagao-Kitamoto *et al.*, "A potential pathogenic association between periodontal disease and Crohn's disease," JCI Insight **6** (2021).
2. E. Read, M. A. Curtis, and J. F. Neves, "The role of oral bacteria in inflammatory bowel disease," Nat. Rev. Gastroenterol. & Hepatol. **18**, 731–742 (2021).
3. S. Joseph, J. Aduse-Opoku, A. Hashim, E. Hanski, R. Streich, S. C. Knowles, A. B. Pedersen, W. G. Wade, and M. A. Curtis, "A 16s rrna gene and draft genome database for the murine oral bacterial community," Msystems **6**, e01222–20 (2021).
4. K. R. Theis, R. Romero, J. M. Greenberg, A. D. Winters, V. Garcia-Flores, K. Motomura, M. M. Ahmad, J. Galaz, M. Arenas-Hernandez, and N. Gomez-Lopez, "No consistent evidence for microbiota in murine placental and fetal tissues," Msphere **5**, e00933–19 (2020).
5. C. Liu, N. Zhou, M.-X. Du, Y.-T. Sun, K. Wang, Y.-J. Wang, D.-H. Li, H.-Y. Yu, Y. Song, B.-B. Bai *et al.*, "The mouse gut microbial biobank expands the coverage of cultured bacteria," Nat. communications **11**, 1–12 (2020).
6. I. Lagkouvardos, R. Pukall, B. Abt, B. U. Foesel, J. P. Meier-Kolthoff, N. Kumar, A. Bresciani, I. Martínez, S. Just, C. Ziegler *et al.*, "The mouse intestinal bacterial collection (mibc) provides host-specific insight into cultured diversity and functional potential of the gut microbiota," Nat. microbiology **1**, 1–15 (2016).
7. V. Bekker, R. D. Zwittink, C. W. Knetsch, I. M. Sanders, D. Berghuis, P. J. Heidt, J. M. Vossen, W. M. de Vos, C. Belzer, R. G. Bredius *et al.*, "Dynamics of the gut microbiota in children receiving selective or total gut decontamination treatment during hematopoietic stem cell transplantation," Biol. Blood Marrow Transplantation **25**, 1164–1171 (2019).
8. A. Staffas, M. B. da Silva, A. E. Slingerland, A. Lazrak, C. J. Bare, C. D. Holman, M. D. Docampo, Y. Shono, B. Durham, A. J. Pickard *et al.*, "Nutritional support from the intestinal microbiota improves hematopoietic reconstitution after bone marrow transplantation in mice," Cell Host & Microbe **23**, 447–457 (2018).
9. V. Pascal, M. Pozuelo, N. Borruel, F. Casellas, D. Campos, A. Santiago, X. Martinez, E. Varela, G. Sarrabayrouse, K. Machiels *et al.*, "A microbial signature for Crohn's disease," Gut **66**, 813–822 (2017).
10. J. Schluter, J. U. Peled, B. P. Taylor, K. A. Markey, M. Smith, Y. Taur, R. Niehus, A. Staffas, A. Dai, E. Fontana *et al.*, "The gut microbiota is associated with immune cell dynamics in humans," Nature **588**, 303–307 (2020).