**Relative enrichment of oral bacteria in feces indicates loss of intestinal commensals with implications for host health**

Chen Liao1,8, Thierry Rolling2,3,4,8, Ana Djukovic1, Hongbin Liu5, Lei Dai5, Bing Zhai2,3,5, Jonathan U. Peled6,7, Marcel R.M. van den Brink6,7, Tobias M. Hohl2,3,7,\*, and Joao B. Xavier1,9,\*

1Program for Computational and Systems Biology, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA

2Infectious Disease Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

3Immunology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

4Division of Infectious Diseases, First Department of Medicine, University Medical Center, Hamburg-Eppendorf, Hamburg 20251, Germany

5CAS Key Laboratory of Quantitative Engineering Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

6Adult Bone Marrow Transplantation Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

7Weill Cornell Medical College, New York, NY 10065, USA

8These authors contributed equally to this work

9Lead contact

\*Correspondence: xavierj@mskcc.org (J.B.X.), hohlt@mskcc.org (T.M.H.)

**SUMMARY**

Comparative analysis of microbiome compositions reveals relative enrichment of bacterial taxa and establishes associations with human disease states. As a notable example, a high relative abundance of oral bacteria in the feces is typically considered abnormal and has been linked to intestinal disorders. However, this reported association has two alternate explanations: ectopic oral bacteria may increase in absolute numbers in the intestine and drive disease (the *driver* hypothesis) or, instead, intestinal commensals may decrease in absolute numbers and the relative rise in oral bacteria is simply an indicator of this loss (the *marker* hypothesis). Here we use experiments in mice and clinical data from human patients to provide evidence in favor of the *marker* hypothesis. We demonstrate that the depletion of intestinal bacteria by antibiotics increases the relative abundance of oral bacteria in the host feces without increasing their absolute abundance, and the relative enrichment can indicate host health (stool consistency, pathogen overgrowth, and immunity). Therefore, host disorders linked to the relative enrichment of oral bacteria in the intestine do not result from the expansion of ectopic bacteria but instead arise from the loss of endogenous intestinal commensal bacteria. Our results shift the paradigm of interpreting microbiome sequencing data with implications for intestinal disorders and microbiome therapies.

**INTRODUCTION**

Healthy individuals harbor distinct bacterial communities at different body sites, reflecting distinct ecologic niches and conditions (Costello et al., 2009; The Human Microbiome Project Consortium, 2012). A number of health disorders and associated treatments cause the translocation of bacteria between distinct body sites, leading to a mixture of previously separated microbiome communities (Imai et al., 2021). It is well established that bacteria inhabiting the oral cavity can translocate to the lower gastrointestinal tract (Atarashi et al., 2017; Kitamoto et al., 2020b; Li et al., 2019), and the finding of oral bacteria in feces has been linked to intestinal disorders such as inflammatory bowel disease (IBD) (Read et al., 2021) and colorectal cancer (CRC) (Komiya et al., 2019). Compared to the hematogenous route (oral-blood axis) (Abed et al., 2020), this enteral route (oral-gut axis) is a natural pathway (Jin et al., 2022; Kitamoto et al., 2020a). On average, humans swallow ~1011 of oral bacteria cells per day (Segata et al., 2012), yet oral bacteria remain rare in the healthy intestine under homeostatic conditions. Gastric acids, alkaline bile salts, and antimicrobial peptides kill many­ ectopic oral bacteria, and the survivors must overcome colonization resistance that is mediated by the intestinal microbiome (Jin *et al.*, 2022). As a result, compositional sequencing data shows typically <2% of 16S rDNA from oral bacteria in the feces of healthy people (Rashidi et al., 2021; Schmidt et al., 2019).

Antibiotics, diet, aging, and intestinal inflammation can disrupt the oral-intestinal barrier and cause a relative enrichment of oral bacteria in the feces (Kitamoto *et al.*, 2020a). This enrichment has been observed in patients with a variety of diseases, including Crohn’s disease (CD) (Gevers et al., 2014), ulcerative colitis (UC) (Schirmer et al., 2018), irritable bowel syndrome (IBS) (Priya et al., 2022), CRC (Kostic et al., 2013) and liver cirrhosis (Qin et al., 2014). Recently, the enrichment signature has been found associated with unfavorable clinical response to anti-PD-1 (programmed cell death protein-1) therapy in a melanoma cohort (McCulloch et al., 2022). However, nearly all those associations were obtained from compositional sequencing data, which reports relative abundance that by itself can neither determine if there was an expansion of oral bacteria in absolute numbers nor reveal the true causes of human disorders.

Two alternate explanations exist for this finding (Fig. 1). In the *driver* hypothesis, the high relative abundance of oral bacteria reflects a high absolute abundance. The relative enrichment of oral bacteria reflects an active expansion of the oral bacterial population in the intestine. When the oral population reaches a high density, the *driver* hypothesis predicts that this high density causes an associated disorder. In the *marker* hypothesis, the relative enrichment of oral bacteria is driven by loss of intestinal commensals, and it is this bacterial loss that is associated with the disease state. Determining which hypothesis is true has important implications for our understanding of human health and the pathology of disorders linked to inflammatory and infectious pathology.

Here, we compare the *driver* and *marker* hypotheses by analyzing the dynamics of oral bacteria in feces of both mice and humans. By quantifying the absolute abundance of oral bacteria in feces after antibiotic treatment, we show that their absolute numbers do not increase when their relative numbers do, due to depleted intestinal bacterial populations. Our findings unequivocally prove that the *marker*, not the *driver*, hypothesis is true. The validated *marker* hypothesis further indicates that the link between the relative enrichment of oral bacteria in feces and intestinal disorders is mediated by the loss of intestinal commensals and their associated functions. Recognizing that the increased relative abundance of oral bacteria in the intestine does not reflect an ectopic bloom will critically inform the interpretation of microbiome compositional data and future microbiome-based interventions to treat intestinal disorders.

Diagram

Description automatically generated

**Fig. 1| The *driver* and *marker* hypotheses provide alternative explanations for the reported association of oral bacterial enrichment in the feces and intestinal disorders.** Oral bacteria flow from the oral cavity to the intestine where they encounter colonization resistance that is mediated by intestinal commensal bacteria. The relative abundance of oral bacteria in the feces can increase in patients with intestinal disorders or after treatment with certain antibiotics. This enrichment has two possible mechanisms: The relative enrichment may be due to increased absolute abundance of oral bacteria in the intestine microbiome and this increased population density drives the disorders (the *driver* hypothesis, in red). Alternatively, it may be caused by decreased absolute numbers of intestinal bacterial commensals whose depletion drives the disorders (the *marker* hypothesis, in blue).

**RESULTS**

**Antibiotics deplete intestinal bacterial commensals in mice and raises relative abundance of oral bacteria in feces without raising absolute abundances**

The two hypotheses in Fig. 1 give rise to different predictions about the relationships between relative and absolute abundance of oral bacterial populations in the intestine. To distinguish these two possibilities, we treated 3 individually housed mice for seven days with an antibiotic cocktail of ampicillin, vancomycin, and neomycin that is known to deplete murine intestinal bacteria (Reikvam et al., 2011) and measured quantitative profiles via qPCR (Fig. 2A). We first compared the bacterial compositions of the oral and fecal microbiomes on day 0 prior to antibiotics and the fecal microbiomes on day 8. As expected, the baseline composition of the oral and fecal bacterial communities was distinct, reflecting their different habitats (Fig. 2B). Surprisingly, after administration of antibiotics, the post-treatment fecal microbiomes were more similar to pre-treatment oral microbiomes than to pre-treatment fecal microbiomes (Fig. 2C). 10%-20% of ASVs (amplicon sequence variants) detected in the post-treatment feces (the majority from 6 members; Fig. S1) were exclusively found in the corresponding pre-treatment oral samples (Fig. 2D). We next quantified the relative and absolute abundances of ASVs typical of the oral cavity (mean relative abundance >0.01% across pre-treatment oral samples and ≤ 0.01% across pre-treatment fecal samples). Antibiotic treatment enriched the total relative abundance of these oral-typical ASVs in the feces from an average of 0.03% at baseline to an average of 27.2% (Fig. 2E). However, the relative enrichment of oral bacteria was not due to an increase in their absolute abundance. In fact the mean absolute abundance of oral bacteria in the feces decreased by 72% compared to pre-treatment samples (Fig. 2F). Therefore, the relative enrichment was driven by a much greater decrease (>5,000 fold) in the absolute abundances of intestinal commensals after antibiotic treatment.

The result of this mouse experiment supports the *marker* hypothesis. If this observation were universally true, we should be able to determine the antibiotic-induced intestinal bacterial loss by quantifying the relative abundance of oral bacteria in feces. However, most microbiome studies lack paired oral and intestinal samples, which prevent us from applying the same paired-sample-analysis to quantify the oral fraction. To identify oral bacterial DNAs in mouse feces, we compiled a reference set of 149 non-redundant, full-length 16S rRNA genes of culturable bacteria that reside in the oral cavity of mice but do not reside in the murine intestinal tract (Fig. S2, Table S1, see STAR Methods). This reference database allows us to identify ASVs of oral origin by exact sequence match and to estimate their total fractions in fecal samples.

To validate this approach we analyzed published data from mice treated with an antibiotic cocktail of ampicillin, vancomycin, neomycin, and metronidazole for 5 days (Reese et al., 2018) (Fig. 2G). Consistent with our mouse experiment, antibiotics enriched three oral ASVs of *Lactobacillus*, *Proteus* and *Escherichia-Shigella* in the feces (Fig. 2H). In agreement with the *marker* hypothesis, their absolute abundances did not increase but in fact decreased by 24.7-fold on average (Fig. 2I, left). The absolute abundance of intestinal bacteria decreased even more, by 189.2-fold (Fig. 2I, right), which caused an observed relative enrichment of oral bacteria. Interestingly, the fraction of oral bacteria in feces showed a negative linear correlation (Pearson’s r = -0.52, P = 3.5e-17) with the total bacterial loads on the log10-log10 scale (Fig. 2J, Fig. S3), and the anti-correlation remained significant by using the control group data alone (Fig. 2J, data in green, Pearson’s r = -0.25, P = 5.9e-3). Taken together, the fraction of oral bacteria in feces can be used to quantitatively determine the level of intestinal commensal bacterial loss in antibiotic-treated subjects and to assess natural fluctuations in their loads even in untreated subjects.

Graphical user interface

Description automatically generated

**Fig. 2| Mice treated with antibiotics support the *marker* hypothesis.** (A) Three mice (labeled as M1, M2, M3) were treated with an antibiotic cocktail of ampicillin, vancomycin, and neomycin (AVN) for a week. The pre-treatment feces (fecal\_d0) and oral swabs (oral\_d0) were collected on day 0 and the post-treatment feces (fecal\_d8) were collected on day 8. (**B**) Microbiome compositions of all feces and oral swabs show different oral and fecal microbiomes prior to antibiotics, but an enrichment of oral bacteria was observed in the feces post antibiotics**.** (**C**) Principle component analysis of all microbiome samples. (**D**) Tissue of origin for observed ASVs in the post-treatment fecal samples. (**E,F**) Relative (E) and absolute (F) abundances of oral bacterial populations in the feces. Horizontal bars: mean. (**G-J**) Analysis of a previously published dataset (Reese *et al.*, 2018) leads to the same conclusion. (**G**) Experimental design. Mice were treated with either water (the control group; n=11) or antibiotic cocktails of ampicillin, vancomycin, neomycin, and metronidazole (AVNM) (the antibiotic group; n=10) for 5 days. (**H**) Compositional dynamics of oral bacterial ASVs in feces. (**I**) Absolute abundances of oral (left) and intestinal (right) bacteria in feces averaged across mice in the control or treatment group. Lines and dots: mean; shading: 95% confidence interval (CI). (**J**) The relationship between total bacterial loads and oral bacterial fractions indicates that the fractions can be used to estimate the absolute amounts of bacteria in the fecal microbiome of both antibiotic-treated (orange) and untreated (green) subjects. Line: best linear fit; shading: 95% CI. Unit of bacterial load: 16S copies per gram of feces.

**A quantitative approach to estimate the fraction of oral bacteria in human feces**

To quantify the oral bacterial fraction in human feces, we followed a similar strategy as used in mice to address the limitation of missing paired oral samples in human intestinal microbiome studies, especially those with quantitative data. To build a reference set of bacterial 16S rRNA gene sequences that typically colonize the human oral cavity (see STAR Methods), we leveraged the Human Microbiome Project (HMP) that sequenced paired oral (multiple subsites) and intestinal (stool) microbiomes from 237 healthy volunteers (The Human Microbiome Project Consortium, 2012). Distinct bacterial communities inhabit the two body sites (Fig. 3A), which provides the basis to define oral-typical ASVs based on their abundance and prevalence in healthy people. We defined ASVs of oral origin if their mean relative abundance per oral sample was greater than 0.01% and if their mean prevalence was higher than 5%. We excluded ASVs if their relative abundance or their prevalence met either cutoff in fecal samples (Fig. 3B). This filtering step gave rise to 178 oral ASVs (Table S2). The leading genera were *Prevotella* and *Streptococcus* (Fig. 3C). 219 out of 280 HMP fecal samples contained no oral ASVs. The mean relative abundance of oral ASVs was 0.05%±0.14% (Fig. 3D), which agrees with previous reports that oral bacteria are rare in the feces of healthy people (Rashidi *et al.*, 2021; Schmidt *et al.*, 2019). By contrast, identifying oral taxa via sequence overlaps in the absence of filtering leads to an overestimation of 15.63%±22.07% oral ASVs in HMP fecal samples. Thus, the filtering step makes our estimate of oral bacteria in the feces a conservative approach for both healthy and diseased individuals (Supplementary Texts, Fig. S4). In the following, we use the reference set to infer the oral ASVs by exact sequence match and to estimate their total fraction in feces by summing the abundances of individual taxa.

Graphical user interface

Description automatically generated

**Figure 3. Data from healthy individuals reveals bacterial DNAs typically found in the human oral cavity but not the intestine.** (**A**) An overview of the 4,354 microbiome samples from the Human Microbiome Project (HMP). T-distributed stochastic neighbor embedding (tSNE) (Van der Maaten and Hinton, 2008) collapses all samples (dots) into distinct clusters based on their compositions. (**B**) The mean relative abundance and prevalence of ASVs (dots) among all HMP oral cavity samples (x axis) and the fecal samples (y axis). 178 ASVs were identified as oral-typical and highlighted in orange. (**C**) Genus-level distribution of the 178 oral-typical ASVs. (**D**) Distribution of the total fractions of oral-typical ASVs in the feces of HMP subjects (n=280). Each circle represents a fecal sample. We detected no oral ASVs in 219 samples.

**The intestinal domination of oral bacteria in bone marrow transplant recipients reveals intestinal bacterial depletion**

Using the quantification method developed above, we determined the relative abundance of oral bacteria in a large-scale microbiome dataset from adult cancer patients receiving allogeneic hematopoietic cell transplantation (allo-HCT) (Liao et al., 2021; Yan et al., 2022). Allo-HCT is a potentially curative therapy for a variety of hematologic disorders including leukemias, lymphomas, and myeloma (Shono and van den Brink, 2018). Allo-HCT disrupts the intestinal microbiome composition of the recipients through the administration of conditioning chemotherapy and the use of antibiotics to prevent and treat bacterial infections. During allo-HCT, *Enterococcus* and *Streptococcus* are the two most abundant genera that frequently dominate (>30% relative abundance of a single ASV; (Taur et al., 2012)) the intestinal bacterial community (Fig. 4A, top) (Peled et al., 2020). Interestingly, among 10,433 fecal samples from 1,276 allo-HCT patients, 901 (8.6%) were dominated by at least one oral ASV. *Streptococcus*, *Actinomyces*, and *Abiotrophia* were the leading genera of these dominant oral ASVs(Fig. 4A, middle). The total bacterial loads (in 3,126 analyzed samples) showed that the fecal samples highly enriched with oral ASVs had a lower bacterial biomass than fecal samples without enrichment of oral ASVs (Fig. 4A, bottom). The total fractions of oral ASVs detected in the feces of allo-HCT recipients was not correlated with sequencing depth (Fig. S5).

To test further whether the identified genera were of oral origin, we used co-occurrence analysis (see STAR Methods) to identify oral ASVs that were simultaneously present in the feces more frequently than expected by chance. The co-occurring oral ASVs, if such existed, would indicate collective bacterial transmission from the oral cavity to the intestine. We found 71 such oral-ASV groups that included 2 to 5 ASVs each. *Streptococcus* accounted for 55 of these 71 ASV groups (Table S3). To identify the species in the *Streptococcus* ASV groups, we extracted shotgun metagenomes from 19 samples which contained at least 10% (measured by 16S amplicon sequencing) ASV\_8 (the most abundant *Streptococcus* ASV inferred as of oral origin) from the same allo-HCT cohort. We were able to recover 22 high-quality metagenome-assembled genomes (MAGs) of *Streptococcus spp.*, among which 4 contains ASV\_8. All four MAGs were annotated as *S. thermophilus*—a probiotic species of lactic acid bacteria (Siranosian et al., 2022). To assess the viability of these MAGs, we computed the ratio of metagenomic reads near the replication origin to those near the replication terminus (i.e., peak-to-trough (PTR) ratio). This ratio estimates the simultaneous replication events averaged across the entire population ([Gao and Li, 2018](https://elifesciences.org/articles/45931#bib3); [Korem et al., 2015](https://elifesciences.org/articles/45931" \l "bib4)). We found that the averaged PTR ratio was 1.35 0.014 (Table S4), consistent with a mixture of 35% 2-fork and 65% single-fork bacterial cells. The PTR-based growth estimate is consistent with the idea that the *Streptococcus* species that encodes ASV\_8 was growing slowly in the intestine.

**Piperacillin/tazobactam depletes intestinal bacterial commensals and drives relative enrichment of oral bacteria in feces**

Allo-HCT recipients receive prophylactic and directed antibiotics to minimize the risk of developing infections before immune system reconstitution or to treat invasive bacterial infections (Shono and van den Brink, 2018). Allo-HCT patients commenced antibiotic prophylaxis seven days prior to transplantation per institutional guidelines. The timing of antibacterial antibiotic administration (Fig. 4B, top) corresponded well with the decline in relative (Fig. 4B, middle) and absolute (Fig. 4B, bottom) intestinal bacterial abundances. Since the mean oral bacterial loads fluctuated around a stable average, the increased relative abundance of oral ASVs was driven by the decline in intestinal bacterial load. This finding supports the *marker* hypothesis and generates a negative correlation (Pearson’s r = -0.27, P = 2.2e-54) between the oral bacterial fraction and the total bacterial loads in feces (Fig. 4C). The inverse correlation of the oral bacterial fraction and the total bacterial load stands in contrast to pathogens such as *Enterococcus* (Stein-Thoeringer et al., 2019) and *Candida spp.* (Rolling et al., 2021; Zhai et al., 2020), in which an increase in the relative abundance reflects an increase in the absolute quantity of these organisms.

We next conducted a time-varying survival analysis to quantify the effects of the different antibiotics that are associated with enriched oral ASVs in feces. Piperacillin/tazobactam (TZP, P=5.7e-7) strongly increased the risk of intestinal domination by oral-derived bacteria (Fig. 4D). We confirmed this association in an independent cohort of pediatric allo-HCT recipients (Bekker et al., 2019). Children who have received oral TZP had much higher relative abundance of oral bacteria in feces than those who received oral polymyxin/neomycin (Fig. S7). Other than TZP, oral vancomycin (P=0.049) but not intravenous (P=0.58) vancomycin was associated with intestinal domination of oral bacteria, likely because intravenous vancomycin does not accumulate in the intestine. We also saw a negative association between quinolones and the fraction of oral bacteria in feces (P=6.1e-4, Fig. 4D).

Graphical user interface

Description automatically generated

**Figure 4. Allo-HCT injures the intestinal microbiome and increases the relative, but not the absolute abundances of oral bacteria in feces.** (**A**)Fecal microbiome compositions of all 10,433 fecal samples. The taxonomic composition (top), the estimated fractions of oral bacteria in the feces (middle), and the total bacterial loads (bottom) were aligned for each sample. (**B**) Antibiotic administration (top) and population dynamics of the relative (middle) and absolute (bottom) abundances (16S copies per gram feces) of oral (light blue line) and intestinal (orange line) bacteria in the feces. Line and dots: mean; shading: 95% confidence interval (CI). (**C**) Negative association between proportions of oral bacteria in feces and total bacterial loads across 3,055 samples (dots) with loads no less than 1,000 16S copies per gram feces. Red line: best linear fit; shading: 95% CI. (**D**) Association between antibiotic exposure and intestinal domination of oral ASVs. Vertical line: hazard ratio; bar width: 95% CI. \*\*\*\*: P<0.0001; \*\*\*: P<0.001, \*: P<0.05.

**The relative abundance of oral bacteria in the feces is associated with host health**

We next sought to determine whether the relative abundance of oral bacteria is correlated with the health status of allo-HCT recipients. The stool consistency was looser in patient samples with higher relative abundance of oral bacteria than in those with a low abundance of oral bacteria (Fig. S7), consistent with the presence of diarrhea. This association may be partially attributed to the loss of intestinal bacteria: Previous work (Vandeputte et al., 2017) showed that the water content, which strongly correlated with stool consistency, is negatively associated with microbiome density. We further found that the intestinal domination by any oral ASV over 30% of relative abundance reduced the risk of intestinal domination (relative abundance >30%) by *Enterococcus* but increased the risk of *Candida* domination (relative abundance >90%) (Table S5), both of which are predisposing factors of bloodstream infections (Taur *et al.*, 2012; Zhai *et al.*, 2020). It is likely that the administration of antibiotics successfully eliminated bacterial pathogens including *Enterococcus spp.* when intestinal bacterial commensals were depleted but opened the niche for fungal pathobionts to expand.

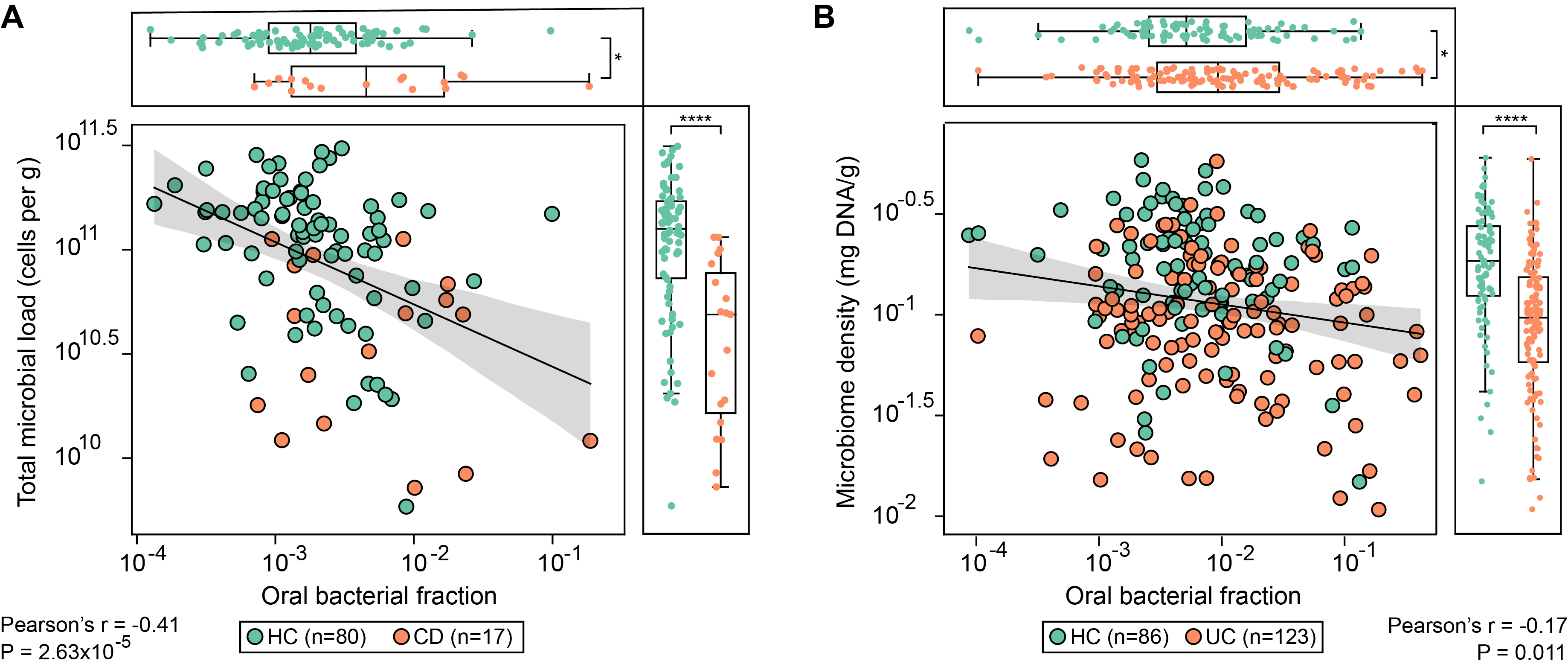
Our previous work has studied the interactions between specific intestinal bacteria and immune cell dynamics (Schluter et al., 2020). Using oral bacteria in feces as a new metric to quantify intestinal microbiome damage (i.e., loss of commensal bacteria), we found that the oral bacterial fraction was negatively correlated with circulating white blood cell counts, including neutrophils, lymphocytes, and the total white blood cells (Fig. S8). We next investigated whether the metric is associated with white blood cell dynamics (i.e., rate of white blood cell count changes). When we accounted for immunomodulatory drugs as potential covariates, the oral bacterial fraction remained positively associated with the day-to-day change in the total white blood cell, neutrophil, and monocyte counts (Table S6).

Finally, we reanalyzed a public dataset from bone-marrow-transplanted mice that received antibiotic prophylaxis with a single drug (Staffas et al., 2018) (Fig. S9A). In that study, antibiotics reduced intestinal bacterial loads (Fig. S9B) without altering oral bacterial loads (Fig. S9C), causing a strong linear relationship (Pearson’s r = -0.81, P = 2.9e-11) between oral bacterial fractions and total bacterial loads in the log10-log10 space (Fig. S9D). Most importantly, the number of bone marrow cells at day 14 showed a negative linear relationship (Pearson’s r = -0.90, P = 5.9e-5) with log10-transformed oral bacterial fractions in feces (Fig. S9E). These data from human and mice indicated that the relative abundance of oral bacteria in the feces, which indicates the total bacterial load, can inform host immune system status.

**The *marker* hypothesis unifies competing microbiome biomarkers of IBD**

Intestinal microbiome dysbiosis is a hallmark of intestinal disorders, but there are competing ways to determine dysbiosis. For example, clinical studies of IBD have revealed its associations with relative enrichment of oral-associated bacterial species (Read *et al.*, 2021) or low absolute values of microbial loads (Contijoch et al., 2019; Vandeputte *et al.*, 2017). Since these associations were reported by separate studies, they raise a critical question: are the two IBD biomarkers independent or related?

According to the *marker* hypothesis, the two quantities should be dependent and negatively correlated with each other. Indeed, the fraction of oral bacteria and total bacterial loads in the feces of patients with CD (Pearson’s r = -0.41, P = 2.63e-5, Fig. 6A) and UC (Pearson’s r = -0.17, P = 0.011; Fig. 6B) were negatively correlated, and each feature was also significantly different between the patients and their healthy controls. For the UC cohort, the association became insignificant (P=0.32) after excluding patients with prior antibiotics or an unclear antibiotic history, due to the loss of significant differences of oral bacterial fractions between the patients and controls. Since oral bacteria can be significantly enriched in the feces of IBD patients who have not taken antibiotics weeks prior to study (Fig. S10), antibiotics may amplify the intestinal microbiome dysbiosis in IBD patients (Gevers *et al.*, 2014) and strengthen the negative relationships between the two biomarkers. In conclusion, the increased oral bacterial fraction and the reduced total number of microbes in the intestine indicate a loss of intestinal bacteria. This loss of intestinal commensal bacteria, rather than the reported association of oral bacterial enrichment in fecal samples, likely represents a fundamental risk factor of IBD.



**Figure 5. The marker hypothesis unifies previous work showing associations between oral bacterial enrichment or low microbial loads in the intestine of patients with inflammatory bowel disease.** (**A**)Crohn’s disease (CD)(Vandeputte *et al.*, 2017)and (**B**)Ulcerative colitis (UC) (Contijoch *et al.*, 2019).Each circle represents a fecal sample and samples that contained no oral bacteria were omitted in the plots. Lines: best linear fits; shading: 95% confidence intervals; HC: healthy control. \*\*\*\*P<0.0001; \*P<0.05; Welch’s t-test.

**DISCUSSION**

Our study combines mouse experiments, human cohort studies, and quantitative microbiome analyses to provide strong evidence that the high relative abundance of oral bacteria in the intestine does not indicate a proportionally high density of oral bacteria in the intestine. Instead, the high relative abundance of oral bacteria represents a marker of the loss of intestinal commensal bacteria. These findings challenge the view that the oral cavity is a reservoir of opportunistic pathogens, whose translocation to and ectopic colonization of the intestine drives intestinal inflammation and pathologic immune responses (Atarashi *et al.*, 2017; Kitamoto *et al.*, 2020b; Rojas-Tapias et al., 2022). Consistent with the *marker* hypothesis, oral bacterial populations in the intestine did not expand after antibiotic treatment, although their relative abundance increased substantially. As in a previous study (Siranosian *et al.*, 2022), we recovered *S. thermophilus* from fecal metagenomic samples of adult allo-HCT recipients. *S. thermophilus* is generally not considered as a pathogen and is a commercially available probiotic that has been reported to attenuate inflammation (Han et al., 2020). Our findings do not formally exclude the possibility that translocated oral bacteria may exert pathologic effects in the intestine, even when present at low absolute levels. More experimental evidence is needed to broadly understand the pathogenicity of translocated oral bacteria at the strain level.

Amplicon-based sequencing produces compositional data, which requires careful interpretation. Quantitative microbiome profiling bypasses compositionality effects and can thus calibrate the associations learned from relative microbiome profiling. One notable example is the negative association between *Bacteroides* and *Prevotella*, which is an outcome of relative microbiome analysis and disappears by taking absolute cell counts into accounts (Vandeputte *et al.*, 2017). Resonating with this finding, our study revealed another discrepancy between relative and absolute abundances: the relative enrichment of oral-derived bacteria in the intestine does not reflect their active population expansion, but simply indicates a microbiome state of depleted intestinal bacteria. Therefore, various oral bacterial species (e.g., *Streptococcus spp.*, *Veillonella parvula*, *Fusobacterium nucleatum*) found enriched in the intestine of IBD patients (Read *et al.*, 2021) may represent biomarkers of a reduced intestinal bacterial load.

To our best knowledge, there is a limited number of quantitative microbiome studies in the literature (Contijoch *et al.*, 2019; Jian et al., 2020; Rao et al., 2021; Schluter *et al.*, 2020; Vandeputte *et al.*, 2017; Vieira-Silva et al., 2019). Among these studies, cell-counting (e.g., flow cytometry) and molecular-based methods (e.g., qPCR, DNA mass) have been applied to measure absolute microbial abundances (Galazzo et al., 2020). All methods have their own limitations, but qPCR is arguably superior for quantifying oral bacteria in feces thanks to the compatibility with the 16S rRNA sequencing (Jian et al., 2021). Technical biases introduced in the extraction, purification, and amplification steps of 16S sequencing are shared with qPCR, but not other approaches. In addition, 16S qPCR measures bacterial loads, not total microbial loads, which is again compatible with 16S data. Therefore, we exclusively used 16S qPCR-based quantification unless the data type is unavailable (e.g., Fig. 5).

Our study discovered that TZP and quinolones promote and reduce the risk of intestinal dominations of oral bacteria respectively (Fig. 3D). According to the *marker* hypothesis, the positive association of TZP reflects the strong anaerobe-killing capacity of this b-lactam/b-lactam inhibitor drug combination (Morjaria et al., 2019). Consistently, TZP used to treat bacterial infections is known as a double-edged sword that significantly increases the GVHD (graft-versus-host disease)-related mortality at 5 years after allo-HCT (Shono et al., 2016). Quinolones are widely used as prophylaxis to reduce the incidences of gram-negative bacterial infections (e.g., *Pseudomonas* and *Enterobacteriaceae*) in patients with neutropenia (Cometta et al., 1994). One possible explanation for this negative association is that most quinolones (Appelbaum, 1999) are more effective against aerobic and facultative anaerobic bacteria (e.g., oral flora) than against anaerobic bacteria.

Research over the past two decades has accumulated unequivocal evidence that the intestinal microbiome contributes to human health (Clemente et al., 2012; Fan and Pedersen, 2021). Compared to biodiversity, the absolute load of intestinal bacteria has received much less attention as a different indicator of microbiome dysfunction (Contijoch *et al.*, 2019). Losing intestinal commensals can fundamentally alter host-microbiome interactions and host health (Contijoch *et al.*, 2019; Tang et al., 2019; Zarrinpar et al., 2018). Consistent with our findings, the total microbial load in the intestine was shown to be positively correlated with the abundance of host immune cells, such as the proportion of mucosal RORγt+ Treg cells in mice (Britton et al., 2020). For host disorders such as IBD, the loss of intestinal bacteria may contribute to the pathogenesis via its regulations of host immunity. Since the total fraction of oral bacteria in feces reflects the total bacterial load in the intestine, it provides a convenient way to assess the size of intestinal bacterial population and even host status. The robust relationship between the two quantities will lead to new opportunities for diagnosis, prognosis, and treatment of microbiome-related human diseases. For example, the finding of a high relative abundance of oral-typical bacteria such as *Streptococcus salivarius* and *Dialister invisus* in the recipient microbiome, as an indicator of low bacterial load, could facilitate engraftment in studies of therapeutic fecal microbiota transplantation (Schmidt et al., 2021).

**ACKNOWLEDGMENTS**

C.L. is supported by National Institutes of Health (NIH) grant nos. U01 AI124275 (J.B.X.), R01 AI137269 (J.B.X.) and U54 CA209975 (J.B.X.). T.R. is funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) grant no. RO-5328/1-2 (T.R.), NIH grant nos. R01 AI093808 (T.M.H.) and R21 AI156157 (T.M.H.). We thank Dr. Zhenjiang Xu for fruitful discussions.

**AUTHOR CONTRIBUTIONS**

Conceptualization, C.L., T.R., T.M.H., and J.B.X.; Mouse experiment: A.D.; Microbiome data processing: C.L. and H.L.; Microbiome data analysis, C.L.; Methodology, C.L.; Investigation, C.L. and T.R.; Writing – Original Draft, C.L., T.R., A.D., H.L.; Writing – Review and Editing, T.M.H., J.B.X., J.U.P., B.Z., L.D., and M.R.M.v.d.B.; Supervision, J.B.X. and T.M.H.

**DECLARATION OF INTERESTS**

J.U.P. reports research funding, intellectual property fees and travel reimbursement from Seres Therapeutics and consulting fees from DaVolterra, CSL Behring and from Maat Pharma. He has filed intellectual property applications related to the microbiome (reference nos. 62/843,849, 62/977,908 and 15/756,845). M.R.M.v.d.B. has received research support from Seres Therapeutics; he has consulted, received honorarium from or participated in advisory boards for Seres Therapeutics, WindMIL Therapeutics, Rheos, Frazier Healthcare Partners, Nektar Therapeutics, Notch Therapeutics, Forty Seven Inc., Priothera, Ceramedix, Lygenesis, Pluto Immunotherapeutics, Magenta Therapeutics, Merck & Co., Inc. and DKMS Medical Council (Board); and he has IP Licensing with Seres Therapeutics, Juno Therapeutics and stock options from Seres and Notch Therapeutics. T.M.H. has participated in a scientific advisory board for Boehringer-Ingelheim Inc. T.R. is currently an employee of BioNTech SE.

**STAR METHODS**

**KEY RESOURCES TABLE**

|  |  |  |
| --- | --- | --- |
| **REAGENT or RESOURCE** | **SOURCE** | **IDENTIFIER** |
| **Biological samples** | | |
| Fecal and oral samples from mouse | This study | NA |
| **Chemicals** | | |
| Ampicillin | Nova Plus | Cat#0781-9408-80 |
| Vancomycin hydrochloride | Thermo Scientific | Cat#J62790.06 |
| Neomycin sulfate hydrate | Thermo Scientific | Cat#J61499.14 |
| MiSeq Reagent Kit v3 | Illumina | Cat#15043895 |
| KAPA LTP Library preparation kit | Roche | Cat#07961880001 |
| DNeasy Blood & Tissue Kit | Qiagen | Cat#69504 |
| PowerUp™ SYBR™ Green Master Mix | Thermo Scientific | Cat#A25742 |
| Ready-Lyse Lysozyme Solution, 10,000,000 U | Lucigen | Cat#NC9745284 |
| Investigator Lyse&Spin Basket kit | Qiagen | Cat#19598 |
| Quant-iT dsDNA Assay Kit, High Sensitivity | Invitrogen | Cat#Q33120 |
| **Deposited Data** | | |
| This study | SRA | PRJNA873058 |
| Reese *et al.* (2018) | SRA | PRJEB26446 |
| Staffas *et al.* (2018) | SRA | PRJEB24887 |
| Theis et al. (2020) | SRA | PRJNA594727 |
| The Human Microbiome Project Consortium (2012) | Qiita | 1928 |
| Imai *et al.* (2021) | SRA | PRJNA684508,  PRJNA684584 |
| Pascal et al. (2017) | SRA | PRJNA422193 |
| Liao *et al.* (2021); Yan *et al.* (2022) | SRA | PRJNA394877, PRJNA607574, PRJNA606262, PRJNA548153, PRJNA545312 |
| Bekker *et al.* (2019) | SRA | PRJEB28845 |
| Vandeputte *et al.* (2017) | SRA | PRJEB21504 |
| Contijoch *et al.* (2019) | SRA | PRJNA413199 |
| **Experimental Models: Organisms/Strains** | | |
| Female C57BL6/J Mice | The Jackson Laboratory | JAX Cat# 000664; RRID: IMSR\_JAX:000664; Room #RB13 |
| **Oligonucleotides** | | |
| Primer: 16S\_Forward\_563F 5’-AYTGGGYDTAAAGNG-3’ | Taur et al. (2018) | NA |
| Primer: 16S\_Reverse\_926R 5’-CCGTCAATTYHTTTRAGT-3’ | Taur *et al.* (2018) | NA |
| qPCR: 16S\_Forward\_27F 5’-AGAGTTTGATCMTGGCTCAG-3’ | IDT | NA |
| qPCR: 16S\_Reverse\_338R 5’TGCTGCCTCCCGTAGGAGT-3’ | IDT | NA |
| **Databases** | | |
| SILVA v138 | Quast et al. (2012) | www.arb-silva.de |
| The Mouse Intestinal Microbial Bank (mGMB) | Liu et al. (2020) | NA |
| The Mouse Intestinal Bacterial Collection (miBC) | Lagkouvardos et al. (2016) | NA |
| Mouse Oral Microbiome Database (MOMD) | Joseph et al. (2021) | momd.org |
| **Softwares** | | |
| Python v3.7.9 | Python | python.org |
| Scikit learn v0.24.0 | Pedregosa et al. (2011) | scikit-learn.org |
| Lifelines v0.25.6 | Davidson-Pilon (2019) | github.com/CamDavidsonPilon/lifelines |
| iRep v1.10 | Brown et al. (2016) | github.com/christophertbrown/iRep |
| DADA2 v1.20 | Callahan et al. (2016) | benjjneb.github.io/data2 |
| Cutadapt v3.4 | Martin (2011) | github.com/marcelm/cutadapt |
| Bhatt lab workflow | Siranosian *et al.* (2022) | github.com/bhattlab/bhattlab\_workflows |
| MEGAHIT v1.2.9 | Li et al. (2015) | github.com/voutcn/megahit |
| MetaBAT 2 v2.12.1 | Kang et al. (2019) | bitbucket.org/berkeleylab/metabat/src/master/ |
| CONCOCT v1.1.0 | Alneberg et al. (2014) | github.com/BinPro/CONCOCT |
| DAS Tool v1.1.2 | Sieber et al. (2018) | github.com/cmks/DAS\_Tool |
| Kraken2 v2.1.1 | Wood et al. (2019) | github.com/DerrickWood/kraken2 |
| QIIME2 v2021.8 | Bolyen et al. (2019) | qiime2.org |
| Matlab 2021b | Matlab | mathworks.com |

**RESOURCE AVAILABILITY**

**Lead contact**

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Joao B. Xavier (xavierj@mskcc.org)

**Materials Availability**

This study did not generate new unique reagents.

**Data and code availability**

* All raw sequences of the microbiome datasets analyzed in this study are public available with accession numbers listed in the key resource table.
* All processed data supporting the findings of this study are available within the article and its supplementary materials.
* Customized Python and Matlab scripts for all analyses included in this study are available on Github (<https://github.com/liaochen1988/marker_hypothesis>).

METHOD DETAILS

**Mouse experiment setup**

We treated C57BL/6J female mice with cocktail of ampicillin (0.5g/l), vancomycin (0.5g/l) and neomycin (1g/l) for one week in drinking water. Antibiotics were changed once during the course of the treatment. Animals were single-housed in autoclaved cages. Autoclaved water supplemented with antibiotics and 5053 irradiated food was provided *ad libitum*. Fecal pellets were collected immediately before and one week after the initiation of antibiotic treatment. Oral swabs were collected as per (Abusleme et al., 2017) before the treatment was started. Briefly, mice were hand-held while sterile swab was introduced into mouth and swiped for at least 30 seconds. After, the swab was put into 150ul of TE, the tip was cut off so that the eppendorf can be closed. Samples were put immediately to dry ice. One negative control swab was taken by pulling out the swab from the pouch and swirling through air for at least 30 sec, after which it was put in TE and on dry ice. Fecal samples and oral swabs were kept at -80ºC until further processing.

**DNA extraction and sequencing**

Fecal DNA was extracted and 16S rRNA gene was amplified using the previously described protocol (Taur *et al.*, 2018). Illumina TruSeq Sample Preparation protocol was used for generating libraries, that were later quantified, normalized, and sequenced using MiSeq Reagent Kit V3. Oral DNA was extracted by using modified DNeasy Blood and Tissue Kit protocol as described in (Abusleme *et al.*, 2017). After extraction 16S rRNA gene was amplified and KAPA LTP Library Preparation Kit was used to generate sequencing libraries that were later quantified, normalized and sequenced using MiSeq Reagent Kit V3.

**Quantitative PCR (qPCR) for determining bacterial load**

For assessing the bacterial load in fecal and oral samples, qPCR against standard curve was used to determine 16S rRNA copy number. For this purpose, the PowerUP qPCR Kit was used. Briefly, for each sample, 20μl PCR triplicates were prepared with each containing 2μl of the DNA used as template, 10μl of mix provided by the manufacturer, and 1μl of forward and reverse primers at the final concentration of 0.5μM (F- AGAGTTTGATCMTGGCTCAG; R- TGCTGCCTCCCGTAGGAGT). In order to complete the volume of the reaction, 6μl of water was added. A PCR product of the 16S rRNA gene from *Enterococcus faecium* ATCC 700221 strain was used for obtaining a standard curve by amplifying its 16S rRNA gene and purifying the product. The copy number of the PCR product was determined based on its concentration and 16S rRNA sequence. A standard curve was obtained by using 10-fold dilutions.

Cycling conditions of the qPCR were 50ºC for 2 minutes, 95ºC for 2 minutes, and 40 cycles of 95ºC for 15 seconds, 56ºC for 15 seconds and 72ºC for 60 seconds. By extrapolating results by looking the ones obtained from standard curve samples, the number of 16S rRNA genes was determined for each sample. The final number of 16S rRNA genes per 1g of fecal sample was calculated by multiplying the number of 16S rRNA molecules obtained by qPCR with DNA elution volume after DNA extraction and dividing this number with the weight of the fecal pellet from which DNA extraction was performed.

**Identification of oral bacteria in feces**

We constructed two reference sets of 16S rRNA sequences, separately for humans (V4-V5 region) and mice (full length), to identify bacteria typically colonizing the oral cavity. A 16S rRNA sequence detected in human or mouse intestinal microbiome can be classified to be oral-typical or oral-atypical by exactly matching the sequence to the corresponding reference set. The 16S rRNA regions typical of the human oral cavity were selected by filtering the microbiome profiles of the Human Microbiome Project at the ASV (V4-V5 region of 16S rRNA gene) level. An ASV is oral-typical if (1) its relative abundance averaged across all oral cavity samples (including samples from all oral cavity subsites, the same below) is greater than 1e-4; and (2) its relative abundance averaged across all fecal samples is no greater than 1e-4; and (3) its prevalence across all oral cavity samples is greater than 0.05; and (4) its prevalence across all fecal samples is no greater than 0.05. Prevalence of an ASV was computed as the proportion of samples that contain the ASV at a relative abundance above 1e-3 (Machado et al., 2021). Similar thresholds of relative abundance (1e-3 at the species level) and prevalence (0.05) were used in a previous study to identify oral-typical species (not ASVs) from metagenomic profiles (Thomas et al., 2019).

Our approach of constructing the mouse reference set is schematically shown in Fig. S2. Due to a lack of large-scale microbiome datasets with paired oral and intestinal samples, we used the Mouse Oral Microbiome Database (MOMD) (Joseph *et al.*, 2021) as the starting point for gathering oral-typical sequences. MOMD has 164 full length rRNA sequences from cultured bacterial isolates. To detect if any MOMD sequence is also typically found in the intestine, we matched them to a combined collection of 344 cultured intestinal 16S rRNA sequences from the Mouse Intestinal Microbial Biobank (mGMB) (Liu *et al.*, 2020) and the Mouse Intestinal Bacterial Collection (miBC) (Lagkouvardos *et al.*, 2016) using reciprocal blast search (-perc\_identity 99 -qcov\_hsp\_perc 100 -ungapped). We excluded 37 MOMD sequences that either have exact matches or belong to any taxa that contain at least 3 matched sequences.

Meanwhile, we profiled the paired oral (swab) and intestinal (distal colon) samples from a small set (n=11) of pregnant mice in a public study (Theis *et al.*, 2020) and identified 17 oral-typical ASVs using definitions (1) and (2) for humans above. Prevalence-based filter was not implemented here due to too few samples. Our results showed that only 5 of 17 ASVs were covered by MOMD, suggesting that MOMD is incomplete. By searching the nucleotide database in NCBI (National Center for Biotechnology Information), we obtained 29 full length mouse 16S rRNA sequences that contain 6 of the 12 missed oral-typical ASVs. The 29 NCBI sequences were added to MOMD. Finally, 7 redundant sequences were removed.

**16S amplicon sequencing**

The >10,000 microbiome samples from the adult allo-HCT recipients were previously analyzed by an in-house processing pipeline and the ASV profiles were available in a recent compilation study (Liao et al., 2021). Briefly, reads were trimmed to the first 180 bp or the first point with a quality score Q<2, and removed if they contained ambiguous nucleotides (N) or if two or more errors were expected based on the quality of the trimmed reads. ASVs were identified using DADA2 (Divisive Amplicon Denoising Algorithm) (Callahan *et al.*, 2016) and classified by IDTaxa (Murali et al., 2018) and the SILVA v138 database (Quast *et al.*, 2012).

The demultiplexed and primer-trimmed HMP 16S sequences were downloaded from the Qitta repository (Gonzalez et al., 2018) and processed by QIIME (Quantitative Insights Into Microbial Ecology) 2 (Bolyen *et al.*, 2019) . DADA2 was used to denoise data and generate an ASV per sample count table, using the QIIME denoise-pyro plugin (Bolyen *et al.*, 2019). Parameter --p-trunc-len 395 was used to remove low quality tails. Taxonomy classification of the representative ASV sequences was performed using the QIIME plugin “feature-classifier” (Bokulich et al., 2018) and the SILVA v138 database (Quast *et al.*, 2012). The classification took three steps. We first extracted the V3-V5 region of the SILVA reference sequences using the extract-reads method. Then we created a classifier by using the fit-classifier-naïve-bayes method with extracted reads and the SILVA reference taxonomy. Finally, we ran the classifier on the ASV sequences using the classify-sklearn method to get their taxonomy.

All other microbiome datasets used in this study were similarly processed using QIIME 2 (Bolyen *et al.*, 2019). When needed, primers were removed from demultiplexed short reads using the QIIME cutadapt plugin (Martin, 2011) with parameters “--p-error-rate 0.1” and “--p-overlap 3”. The trimmed reads were denoised using the QIIME dada2 plugin with truncation lengths determined by per-base quality scores to generate feature tables at the ASV level. Taxonomic classification was performed using the QIIME plugin “feature-classifier classify-sklearn” (Pedregosa *et al.*, 2011) against the SILVA 138 database (Quast *et al.*, 2012) at a cutoff of 80% and inferred to the lowest possible taxonomic level.

For each sample, we only kept bacterial ASVs and removed those whose taxonomy contains keywords “Chloroplast” or “Mitochondria”. We further discarded samples whose total sequencing depths are smaller than 1,000 reads.

**Shotgun metagenomic sequencing**

We adapted a recently published pipeline (Siranosian *et al.*, 2022) to assemble the contigs from short reads using MEGAHIT (Li *et al.*, 2015) and then bin the contigs into Metagenome-assembled genomes (MAGs) using two different methods: Metabat2 (Kang *et al.*, 2019) and CONCOCT (Alneberg *et al.*, 2014). The results were then aggregated to produce an optimized, non-redundant set of MAGs by DAS Tool (Sieber *et al.*, 2018). The high-quality MAGs (75% complete, 175 fragments/Mbp sequence, and 2% contamination) classified as *Streptococcus spp.* by Kraken2 (Wood *et al.*, 2019) were analyzed by iRep (Brown *et al.*, 2016). The iRep value of a MAG represents the average number of replication events over different subpopulations of the MAGs weighted by their relative abundances.

**Co-occurrence analysis**

The algorithm for computing co-occurring bacterial communities is described in detail elsewhere (Machado et al., 2021). Briefly, it begins with ASV pairs and iteratively identifies co-occurring ASVs of larger combination sizes. The presence/absence of an ASV is determined by a relative abundance cutoff of 0.001. A combination of ASVs must satisfy the following criteria to be considered as co-occurring: (1) they must co-occur in at least 200 patients and 500 samples; (2) they must co-occur at least twice more than expected by chance, which is estimated by assuming each ASV was observed independently and its number of observation can be modeled by a binomial distribution; (3) the (false discovery rate) FDR-corrected *P*-values for the independent observation hypothesis must be less than 0.05.

**Cox’s proportional hazard model**

We used the Cox’s time-varying proportional hazard model to regress fecal domination (relative abundance > 30%) of any oral bacterial ASV as a microbial endpoint of interest against antibiotic administration as multivariable predictors. The exposure to each antibiotic takes a value of 1 on the day the antibiotic was administered and 0 otherwise. Next, we again examined domination of oral bacteria in fecal samples, this time as a univariable predictor of intestinal domination of *Enterococcus* and *Candida* ASVs using the same time-varying Cox’s hazard model. Penalty was not added for all Cox hazard calculations. The time interval of allo-HCT recipients starts from 10 days prior to transplantation and ends by 40 days post-transplantation. Patients with less than 5 samples during the time interval were excluded. Rarely administered antibiotics that have been administered less than 10 times were also excluded. We assumed no intestinal domination of oral ASVs on days when samples were not collected.

**Linear mixed-effects model**

To identify the effects of oral bacterial fractions in feces on white blood cell dynamics, we followed our previous strategy (Schluter *et al.*, 2020) and developed four linear mixed-effect models for total white blood cells, neutrophils, lymphocytes, and monocytes separately. These models were specified by the Wilkinson notation ‘logWi ~ TransplantType + GCSFi + MMi + cetirizinei + OralFractioni + Wi + (1|PatientID)’, where Wi is any of the four white blood cell counts measured on day i, TransplantType represents different graft sources (unmodified bone marrow, T-cell depleted, peripheral blood stem cell, umbilical cord) of bone marrow transplantation, GCSF (Granulocyte colony-stimulating factor), MM (Mycophenolate mofetil), and cetirizine are three immunomodulatory drugs and their values represent binary exposure on day i, and PatientID represents a unique identifier of each patient. logWi was computed as the difference between log-transformed white blood cell counts on day i and its next day i+1. The microbiome data and clinical metadata were limited to the periods between neutrophil engraftment and 100 days post engraftment. The expression ‘(1|PatientID)’ indicates that the patient heterogeneity was modeled as random effects. The model was solved by fitlme in Matlab.

**Quantification and statistical analysis**

The oral and intestinal bacterial loads in the intestine were computed by multiplying total bacterial loads (qPCR, flow cytometry, DNA mass) with relative abundance of oral-typical and oral-atypical ASVs respectively. By interpreting oral-atypical bacteria as intestinal bacterial members, we ignored bacterial transmission from other non-oral body sites to the intestine. All statistical analysis, including linear regression, correlation between variables (Pearson’s and Spearman’s correlation), and hypothesis testing (Kruskal-Wallis test and Welch’s t-test) were performed using python. The descriptive statistics was shown as mean±standard deviation.

REFERENCES

Abed, J., Maalouf, N., Manson, A.L., Earl, A.M., Parhi, L., Emgård, J.E., Klutstein, M., Tayeb, S., Almogy, G., and Atlan, K.A. (2020). Colon cancer-associated Fusobacterium nucleatum may originate from the oral cavity and reach colon tumors via the circulatory system. Frontiers in cellular and infection microbiology *10*, 400.

Abusleme, L., Hong, B.-Y., Hoare, A., Konkel, J.E., Diaz, P.I., and Moutsopoulos, N.M. (2017). Oral microbiome characterization in murine models. Bio-protocol *7*, e2655-e2655.

Alneberg, J., Bjarnason, B.S., De Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Lahti, L., Loman, N.J., Andersson, A.F., and Quince, C. (2014). Binning metagenomic contigs by coverage and composition. Nature methods *11*, 1144-1146.

Appelbaum, P.C. (1999). Quinolone activity against anaerobes. Drugs *58*, 60-64.

Atarashi, K., Suda, W., Luo, C., Kawaguchi, T., Motoo, I., Narushima, S., Kiguchi, Y., Yasuma, K., Watanabe, E., and Tanoue, T. (2017). Ectopic colonization of oral bacteria in the intestine drives TH1 cell induction and inflammation. Science *358*, 359-365.

Bekker, V., Zwittink, R.D., Knetsch, C.W., Sanders, I.M., Berghuis, D., Heidt, P.J., Vossen, J.M., de Vos, W.M., Belzer, C., and Bredius, R.G. (2019). Dynamics of the gut microbiota in children receiving selective or total gut decontamination treatment during hematopoietic stem cell transplantation. Biology of Blood and Marrow Transplantation *25*, 1164-1171.

Bokulich, N.A., Kaehler, B.D., Rideout, J.R., Dillon, M., Bolyen, E., Knight, R., Huttley, G.A., and Gregory Caporaso, J. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2’s q2-feature-classifier plugin. Microbiome *6*, 1-17.

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., and Asnicar, F. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature biotechnology *37*, 852-857.

Britton, G.J., Contijoch, E.J., Spindler, M.P., Aggarwala, V., Dogan, B., Bongers, G., San Mateo, L., Baltus, A., Das, A., and Gevers, D. (2020). Defined microbiota transplant restores Th17/RORγt+ regulatory T cell balance in mice colonized with inflammatory bowel disease microbiotas. Proceedings of the National Academy of Sciences *117*, 21536-21545.

Brown, C.T., Olm, M.R., Thomas, B.C., and Banfield, J.F. (2016). Measurement of bacterial replication rates in microbial communities. Nature biotechnology *34*, 1256-1263.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nature methods *13*, 581-583.

Clemente, J.C., Ursell, L.K., Parfrey, L.W., and Knight, R. (2012). The impact of the gut microbiota on human health: an integrative view. Cell *148*, 1258-1270.

Cometta, A., Calandra, T., Bille, J., and Glauser, M.P. (1994). Escherichia coli resistant to fluoroquinolones in patients with cancer and neutropenia. New England Journal of Medicine *330*, 1240-1241.

Contijoch, E.J., Britton, G.J., Yang, C., Mogno, I., Li, Z., Ng, R., Llewellyn, S.R., Hira, S., Johnson, C., and Rabinowitz, K.M. (2019). Gut microbiota density influences host physiology and is shaped by host and microbial factors. Elife *8*.

Costello, E.K., Lauber, C.L., Hamady, M., Fierer, N., Gordon, J.I., and Knight, R. (2009). Bacterial community variation in human body habitats across space and time. science *326*, 1694-1697.

Davidson-Pilon, C. (2019). lifelines: survival analysis in Python. Journal of Open Source Software *4*, 1317.

Fan, Y., and Pedersen, O. (2021). Gut microbiota in human metabolic health and disease. Nature Reviews Microbiology *19*, 55-71.

Galazzo, G., Van Best, N., Benedikter, B.J., Janssen, K., Bervoets, L., Driessen, C., Oomen, M., Lucchesi, M., van Eijck, P.H., and Becker, H.E. (2020). How to count our microbes? The effect of different quantitative microbiome profiling approaches. Frontiers in cellular and infection microbiology *10*, 403.

Gevers, D., Kugathasan, S., Denson, L.A., Vázquez-Baeza, Y., Van Treuren, W., Ren, B., Schwager, E., Knights, D., Song, S.J., and Yassour, M. (2014). The treatment-naive microbiome in new-onset Crohn’s disease. Cell host & microbe *15*, 382-392.

Gonzalez, A., Navas-Molina, J.A., Kosciolek, T., McDonald, D., Vázquez-Baeza, Y., Ackermann, G., DeReus, J., Janssen, S., Swafford, A.D., and Orchanian, S.B. (2018). Qiita: rapid, web-enabled microbiome meta-analysis. Nature methods *15*, 796-798.

Han, F., Wu, G., Zhang, Y., Zheng, H., Han, S., Li, X., Cai, W., Liu, J., Zhang, W., and Zhang, X. (2020). Streptococcus thermophilus attenuates inflammation in septic mice mediated by gut microbiota. Frontiers in microbiology *11*, 598010.

Imai, J., Ichikawa, H., Kitamoto, S., Golob, J.L., Kaneko, M., Nagata, J., Takahashi, M., Gillilland III, M.G., Tanaka, R., and Nagao-Kitamoto, H. (2021). A potential pathogenic association between periodontal disease and Crohn’s disease. JCI insight *6*.

Jian, C., Luukkonen, P., Yki-Järvinen, H., Salonen, A., and Korpela, K. (2020). Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling. PloS one *15*, e0227285.

Jian, C., Salonen, A., and Korpela, K. (2021). Commentary: How to count our microbes? The effect of different quantitative microbiome profiling approaches. Frontiers in Cellular and Infection Microbiology *11*, 627910.

Jin, S., Wetzel, D., and Schirmer, M. (2022). Deciphering mechanisms and implications of bacterial translocation in human health and disease. Current Opinion in Microbiology *67*, 102147.

Joseph, S., Aduse-Opoku, J., Hashim, A., Hanski, E., Streich, R., Knowles, S.C., Pedersen, A.B., Wade, W.G., and Curtis, M.A. (2021). A 16S rRNA gene and draft genome database for the murine oral bacterial community. Msystems *6*, e01222-01220.

Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., and Wang, Z. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ *7*, e7359.

Kitamoto, S., Nagao-Kitamoto, H., Hein, R., Schmidt, T., and Kamada, N. (2020a). The bacterial connection between the oral cavity and the gut diseases. Journal of dental research *99*, 1021-1029.

Kitamoto, S., Nagao-Kitamoto, H., Jiao, Y., Gillilland III, M.G., Hayashi, A., Imai, J., Sugihara, K., Miyoshi, M., Brazil, J.C., and Kuffa, P. (2020b). The intermucosal connection between the mouth and gut in commensal pathobiont-driven colitis. Cell *182*, 447-462. e414.

Komiya, Y., Shimomura, Y., Higurashi, T., Sugi, Y., Arimoto, J., Umezawa, S., Uchiyama, S., Matsumoto, M., and Nakajima, A. (2019). Patients with colorectal cancer have identical strains of Fusobacterium nucleatum in their colorectal cancer and oral cavity. Gut *68*, 1335-1337.

Kostic, A.D., Chun, E., Robertson, L., Glickman, J.N., Gallini, C.A., Michaud, M., Clancy, T.E., Chung, D.C., Lochhead, P., and Hold, G.L. (2013). Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. Cell host & microbe *14*, 207-215.

Lagkouvardos, I., Pukall, R., Abt, B., Foesel, B.U., Meier-Kolthoff, J.P., Kumar, N., Bresciani, A., Martínez, I., Just, S., and Ziegler, C. (2016). The Mouse Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota. Nature microbiology *1*, 1-15.

Li, B., Ge, Y., Cheng, L., Zeng, B., Yu, J., Peng, X., Zhao, J., Li, W., Ren, B., and Li, M. (2019). Oral bacteria colonize and compete with gut microbiota in gnotobiotic mice. International journal of oral science *11*, 1-9.

Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics *31*, 1674-1676.

Liao, C., Taylor, B.P., Ceccarani, C., Fontana, E., Amoretti, L.A., Wright, R.J., Gomes, A.L., Peled, J.U., Taur, Y., and Perales, M.-A. (2021). Compilation of longitudinal microbiota data and hospitalome from hematopoietic cell transplantation patients. Scientific data *8*, 1-12.

Liu, C., Zhou, N., Du, M.-X., Sun, Y.-T., Wang, K., Wang, Y.-J., Li, D.-H., Yu, H.-Y., Song, Y., and Bai, B.-B. (2020). The Mouse Gut Microbial Biobank expands the coverage of cultured bacteria. Nature communications *11*, 1-12.

Machado, D., Maistrenko, O.M., Andrejev, S., Kim, Y., Bork, P., Patil, K.R., and Patil, K.R. (2021). Polarization of microbial communities between competitive and cooperative metabolism. Nature ecology & evolution *5*, 195-203.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal *17*, 10-12.

McCulloch, J.A., Davar, D., Rodrigues, R.R., Badger, J.H., Fang, J.R., Cole, A.M., Balaji, A.K., Vetizou, M., Prescott, S.M., and Fernandes, M.R. (2022). Intestinal microbiota signatures of clinical response and immune-related adverse events in melanoma patients treated with anti-PD-1. Nature Medicine *28*, 545-556.

Morjaria, S., Schluter, J., Taylor, B.P., Littmann, E.R., Carter, R.A., Fontana, E., Peled, J.U., van den Brink, M.R., Xavier, J.B., and Taur, Y. (2019). Antibiotic-induced shifts in fecal microbiota density and composition during hematopoietic stem cell transplantation. Infection and immunity *87*, e00206-00219.

Murali, A., Bhargava, A., and Wright, E.S. (2018). IDTAXA: a novel approach for accurate taxonomic classification of microbiome sequences. Microbiome *6*, 1-14.

Pascal, V., Pozuelo, M., Borruel, N., Casellas, F., Campos, D., Santiago, A., Martinez, X., Varela, E., Sarrabayrouse, G., and Machiels, K. (2017). A microbial signature for Crohn's disease. Gut *66*, 813-822.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., and Dubourg, V. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research *12*, 2825-2830.

Peled, J.U., Gomes, A.L., Devlin, S.M., Littmann, E.R., Taur, Y., Sung, A.D., Weber, D., Hashimoto, D., Slingerland, A.E., and Slingerland, J.B. (2020). Microbiota as predictor of mortality in allogeneic hematopoietic-cell transplantation. New England Journal of Medicine *382*, 822-834.

Priya, S., Burns, M.B., Ward, T., Mars, R.A., Adamowicz, B., Lock, E.F., Kashyap, P.C., Knights, D., and Blekhman, R. (2022). Identification of shared and disease-specific host gene–microbiome associations across human diseases using multi-omic integration. Nature Microbiology, 1-16.

Qin, N., Yang, F., Li, A., Prifti, E., Chen, Y., Shao, L., Guo, J., Le Chatelier, E., Yao, J., and Wu, L. (2014). Alterations of the human gut microbiome in liver cirrhosis. Nature *513*, 59-64.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic acids research *41*, D590-D596.

Rao, C., Coyte, K.Z., Bainter, W., Geha, R.S., Martin, C.R., and Rakoff-Nahoum, S. (2021). Multi-kingdom ecological drivers of microbiota assembly in preterm infants. Nature *591*, 633-638.

Rashidi, A., Ebadi, M., Weisdorf, D.J., Costalonga, M., and Staley, C. (2021). No evidence for colonization of oral bacteria in the distal gut in healthy adults. Proceedings of the National Academy of Sciences *118*, e2114152118.

Read, E., Curtis, M.A., and Neves, J.F. (2021). The role of oral bacteria in inflammatory bowel disease. Nature Reviews Gastroenterology & Hepatology *18*, 731-742.

Reese, A.T., Cho, E.H., Klitzman, B., Nichols, S.P., Wisniewski, N.A., Villa, M.M., Durand, H.K., Jiang, S., Midani, F.S., and Nimmagadda, S.N. (2018). Antibiotic-induced changes in the microbiota disrupt redox dynamics in the gut. Elife *7*, e35987.

Reikvam, D.H., Erofeev, A., Sandvik, A., Grcic, V., Jahnsen, F.L., Gaustad, P., McCoy, K.D., Macpherson, A.J., Meza-Zepeda, L.A., and Johansen, F.-E. (2011). Depletion of murine intestinal microbiota: effects on gut mucosa and epithelial gene expression. PloS one *6*, e17996.

Rojas-Tapias, D.F., Brown, E.M., Temple, E.R., Onyekaba, M.A., Mohamed, A.M., Duncan, K., Schirmer, M., Walker, R.L., Mayassi, T., and Pierce, K.A. (2022). Inflammation-associated nitrate facilitates ectopic colonization of oral bacterium Veillonella parvula in the intestine. Nature Microbiology, 1-13.

Rolling, T., Zhai, B., Gjonbalaj, M., Tosini, N., Yasuma-Mitobe, K., Fontana, E., Amoretti, L.A., Wright, R.J., Ponce, D.M., and Perales, M.A. (2021). Haematopoietic cell transplantation outcomes are linked to intestinal mycobiota dynamics and an expansion of Candida parapsilosis complex species. Nature microbiology *6*, 1505-1515.

Schirmer, M., Denson, L., Vlamakis, H., Franzosa, E.A., Thomas, S., Gotman, N.M., Rufo, P., Baker, S.S., Sauer, C., and Markowitz, J. (2018). Compositional and temporal changes in the gut microbiome of pediatric ulcerative colitis patients are linked to disease course. Cell host & microbe *24*, 600-610. e604.

Schluter, J., Peled, J.U., Taylor, B.P., Markey, K.A., Smith, M., Taur, Y., Niehus, R., Staffas, A., Dai, A., and Fontana, E. (2020). The gut microbiota is associated with immune cell dynamics in humans. Nature *588*, 303-307.

Schmidt, T.S., Hayward, M.R., Coelho, L.P., Li, S.S., Costea, P.I., Voigt, A.Y., Wirbel, J., Maistrenko, O.M., Alves, R.J., and Bergsten, E. (2019). Extensive transmission of microbes along the gastrointestinal tract. Elife *8*, e42693.

Schmidt, T.S., Li, S.S., Maistrenko, O.M., Akkani, W., Coelho, L.P., Dolai, S., Fullam, A., Glazek, A., Hercog, R., and Herrema, H. (2021). Drivers and Determinants of Strain Dynamics Following Faecal Microbiota Transplantation. bioRxiv.

Segata, N., Haake, S.K., Mannon, P., Lemon, K.P., Waldron, L., Gevers, D., Huttenhower, C., and Izard, J. (2012). Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. Genome biology *13*, 1-18.

Shono, Y., Docampo, M.D., Peled, J.U., Perobelli, S.M., Velardi, E., Tsai, J.J., Slingerland, A.E., Smith, O.M., Young, L.F., and Gupta, J. (2016). Increased GVHD-related mortality with broad-spectrum antibiotic use after allogeneic hematopoietic stem cell transplantation in human patients and mice. Science translational medicine *8*, 339ra371-339ra371.

Shono, Y., and van den Brink, M.R. (2018). Gut microbiota injury in allogeneic haematopoietic stem cell transplantation. Nature Reviews Cancer *18*, 283-295.

Sieber, C.M., Probst, A.J., Sharrar, A., Thomas, B.C., Hess, M., Tringe, S.G., and Banfield, J.F. (2018). Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nature microbiology *3*, 836-843.

Siranosian, B.A., Brooks, E.F., Andermann, T., Rezvani, A.R., Banaei, N., Tang, H., and Bhatt, A.S. (2022). Rare transmission of commensal and pathogenic bacteria in the gut microbiome of hospitalized adults. Nature communications *13*, 1-17.

Staffas, A., da Silva, M.B., Slingerland, A.E., Lazrak, A., Bare, C.J., Holman, C.D., Docampo, M.D., Shono, Y., Durham, B., and Pickard, A.J. (2018). Nutritional support from the intestinal microbiota improves hematopoietic reconstitution after bone marrow transplantation in mice. Cell host & microbe *23*, 447-457. e444.

Stein-Thoeringer, C., Nichols, K., Lazrak, A., Docampo, M., Slingerland, A., Slingerland, J., Clurman, A., Armijo, G., Gomes, A., and Shono, Y. (2019). Lactose drives Enterococcus expansion to promote graft-versus-host disease. Science *366*, 1143-1149.

Tang, T.W., Chen, H.-C., Chen, C.-Y., Yen, C.Y., Lin, C.-J., Prajnamitra, R.P., Chen, L.-L., Ruan, S.-C., Lin, J.-H., and Lin, P.-J. (2019). Loss of gut microbiota alters immune system composition and cripples postinfarction cardiac repair. Circulation *139*, 647-659.

Taur, Y., Coyte, K., Schluter, J., Robilotti, E., Figueroa, C., Gjonbalaj, M., Littmann, E.R., Ling, L., Miller, L., and Gyaltshen, Y. (2018). Reconstitution of the gut microbiota of antibiotic-treated patients by autologous fecal microbiota transplant. Science translational medicine *10*, eaap9489.

Taur, Y., Xavier, J.B., Lipuma, L., Ubeda, C., Goldberg, J., Gobourne, A., Lee, Y.J., Dubin, K.A., Socci, N.D., and Viale, A. (2012). Intestinal domination and the risk of bacteremia in patients undergoing allogeneic hematopoietic stem cell transplantation. Clinical infectious diseases *55*, 905-914.

The Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. nature *486*, 207-214.

Theis, K.R., Romero, R., Greenberg, J.M., Winters, A.D., Garcia-Flores, V., Motomura, K., Ahmad, M.M., Galaz, J., Arenas-Hernandez, M., and Gomez-Lopez, N. (2020). No consistent evidence for microbiota in murine placental and fetal tissues. Msphere *5*, e00933-00919.

Thomas, A.M., Manghi, P., Asnicar, F., Pasolli, E., Armanini, F., Zolfo, M., Beghini, F., Manara, S., Karcher, N., and Pozzi, C. (2019). Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. Nature medicine *25*, 667-678.

Van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. Journal of machine learning research *9*.

Vandeputte, D., Kathagen, G., D’hoe, K., Vieira-Silva, S., Valles-Colomer, M., Sabino, J., Wang, J., Tito, R.Y., De Commer, L., and Darzi, Y. (2017). Quantitative microbiome profiling links gut community variation to microbial load. Nature *551*, 507-511.

Vieira-Silva, S., Sabino, J., Valles-Colomer, M., Falony, G., Kathagen, G., Caenepeel, C., Cleynen, I., van der Merwe, S., Vermeire, S., and Raes, J. (2019). Quantitative microbiome profiling disentangles inflammation-and bile duct obstruction-associated microbiota alterations across PSC/IBD diagnoses. Nature microbiology *4*, 1826-1831.

Wood, D.E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. Genome biology *20*, 1-13.

Yan, J., Liao, C., Taylor, B.P., Fontana, E., Amoretti, L.A., Wright, R.J., Littmann, E.R., Dai, A., Waters, N., and Peled, J.U. (2022). A compilation of fecal microbiome shotgun metagenomics from hematopoietic cell transplantation patients. Scientific Data *9*, 1-11.

Zarrinpar, A., Chaix, A., Xu, Z.Z., Chang, M.W., Marotz, C.A., Saghatelian, A., Knight, R., and Panda, S. (2018). Antibiotic-induced microbiome depletion alters metabolic homeostasis by affecting gut signaling and colonic metabolism. Nature communications *9*, 1-13.

Zhai, B., Ola, M., Rolling, T., Tosini, N.L., Joshowitz, S., Littmann, E.R., Amoretti, L.A., Fontana, E., Wright, R.J., and Miranda, E. (2020). High-resolution mycobiota analysis reveals dynamic intestinal translocation preceding invasive candidiasis. Nature medicine *26*, 59-64.