Longitudinal profiling of the microbiome at four body sites reveals core stability and individualized dynamics during health and disease

Short title: Personal longitudinal microbiome dynamics at four body sites

Xin Zhou[1,2,3,4,16], Xiaotao Shen[1,2,16], Jethro S. Johnson[4,5], Daniel J. Spakowicz[4,6], Melissa Agnello[7], Wenyu Zhou[1,2], Monica Avina[1], Alexander Honkala1[8,9], Faye Chleilat[1], Shirley Jingyi Chen[8], Kexin Cha[8], Shana Leopold[4,10], Chenchen Zhu[1], Lei Chen[11], Lin Lyu[11], Daniel Hornburg[1], Si Wu[1], Xinyue Zhang[1], Chao Jiang[12], Liuyiqi Jiang[12], Lihua Jiang[1], Ruiqi Jian[1], Andrew W. Brooks[1], Meng Wang[1], Kévin Contrepois[1], Peng Gao[1], Sophia Miryam Schüssler-Fiorenza Rose[1], Thi Dong Binh Tran[4], Hoan Nguyen[4], Alessandra Celli[1], Bo-Young Hong[4,13], Eddy J. Bautista[4,14], Yair Dorsett[4,15], Paula Kavathas[10], Yanjiao Zhou[4, 15], Erica Sodergren[4], George M. Weinstock[4], Michael P. Snyder[1,2,3,8,*]

1. Department of Genetics, Stanford University School of Medicine, Stanford, California, United States
2. Stanford Center for Genomics and Personalized Medicine, Stanford, California, United States
3. Stanford Diabetes Research Center, Stanford, California, United States
4. The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut, United States
5. Oxford Centre for Microbiome Studies, Kennedy Institute of Rheumatology, University of Oxford, Roosevelt Drive, Headington, Oxford, United Kingdom
6. Division of Medical Oncology, Ohio State University Wexner Medical Center, James Cancer Hospital and Solove Research Institute, Columbus, Ohio, United States
7. Medical Affairs, uBiome, San Francisco, California, United States
8. Stanford Healthcare Innovation Labs, Stanford University School of Medicine, Stanford, California, United States
9. Department of Biomedical Engineering, Oregon Health & Science University, Portland, Oregon, United States
10. Department of Immunobiology, Yale University School of Medicine, New Haven, Connecticut, USA
11. Shanghai Institute of Immunology, Shanghai Jiao Tong University, Shanghai, China
12. Life Sciences Institute, Zhejiang University, Hangzhou, China
13. Woody L Hunt School of Dental Medicine, Texas Tech University Health Science Center, El Paso, TX, United States
14. Corporación Colombiana de Investigación Agropecuaria (Agrosavia), Headquarters-Mosquera, Cundinamarca, Colombia
15. Department of Medicine, University of Connecticut Health Center, Farmington, CT, United States
16. These authors contributed equally: Xin Zhou, Xiaotao Shen

*** Correspondence should be sent to:**
Michael P. Snyder, Ph.D.
Stanford W. Ascherman Professor
Chair, Department of Genetics
Stanford University School of Medicine
Stanford, CA, 94305
mpsnyder@stanford.edu

Abstract (310 words):
The human microbiome is a dynamic ecosystem that exhibits remarkable individuality, yet its dynamics in health and disease across multiple body sites has not been previously explored[1-8]. We analyzed the microbiome at four body sites in a cohort of 86 well characterized insulin sensitive (IS) and insulin resistant (IR) participants for up to 6 yrs. The individuals were sampled quarterly during healthy periods and more frequently during periods of respiratory illness (Total 3058 timepoints). For each participant, we characterized the microbial composition and temporal dynamics of the stool, skin, nasal and oral microbiome to study their individual vs common genera, their stability, as well as the associations between body site microbiomes, host multi-omics datasets (proteomics, metabolomics, lipidomics), and clinical markers. We found that the stool microbiome was the most stable over time, whereas the skin and nasal microbiomes were the least stable. At each body site we identified bacterial taxa that were individual-specific as well as those that were commonly shared across the cohort; importantly, at each site, the highly individualized taxa were found to be more stable over time through health and disease than taxa commonly shared between participants. In addition, microbiome dynamics are highly correlated across sites, suggestive of global coordination through host-microbial-environment interactions. We discovered a multitude (>700) of unique microbial, molecular, and clinical relationships, (e.g., the skin microbiome and plasma lipids, oral microbiome and proinflammatory cytokines), demonstrating that microbial and host molecular and clinical components are highly interconnected. Finally, IR individuals exhibited lower microbial stability in both stool and skin, and their microbial, molecular, and clinical relationships were disrupted, suggesting their balance is altered in metabolic disease. Overall, our study provides a comprehensive assessment of multi-site microbial dynamics and its relationship with host molecules during health and disease.

Keywords: microbiome stability, stool microbiome, skin microbiome, oral microbiome, nasal microbiome, precision medicine, longitudinal profiling

Study Highlights:

1. The stability of the human microbiome varies among individuals and body sites.
2. Highly individualized microbial species are more stable over time.
3. At each of the four body sites, systematic interactions between the host and bacteria can be detected.
4. Individuals with insulin resistance have lower microbiome stability, a more diversified skin microbiome, and significantly altered host-microbiome interactions.

Introduction

Human microbiomes are composed of remarkably dynamic microbial communities that live in and on various body sites[5,9-12]. At each site, the microbial and host cell interactions exhibit territory-specific complexity[13-17]. The molecular foundations of microbial ecology and their interactions with the host are being elucidated with new technology-enabled multi-omics profiling, shedding light on their role in both normal physiological processes such as aging[18-20] as well as in diseases such as inflammatory bowel disease (IBD)[21-23], cardiovascular disease[24-26], and type 2 diabetes mellitus (T2DM)[27-30].

The etiology and pathogenesis of insulin resistance and T2DM have been closely associated with the human microbiome[31-34]. Patients with impaired glucose homeostasis and insulin resistance show microbiome composition alterations in the gut[28,30,32,33,35], skin[36,37], and other body sites[38-44], which reflect an ecological dysbiosis characterized by altered microbiome alpha diversity[32,45], decreased microbiome compositional stability[7,46], and greater inter-individual variability[32]. Damaged mucosal and skin barrier integrity associated with insulin resistance may also lead to increased microbial translocation that exacerbates systemic inflammation[47-50]. Although human microbiome studies are often, by necessity, observational, the causal

relationship between microbiome dysbiosis and impaired glucose/insulin homeostasis has been demonstrated in patients and animal models and through human microbiome manipulation[45,51,52].

While transformative, past studies of the microbiome and glucose homeostasis have limitations. Firstly, they lacked the longitudinal sampling of microbiomes required to precisely capture features of individuality and stability[28,53,54], thereby limiting fundamental insights into host-microbe connections[55-57]. Second, they have primarily focused on the microbiome from a single body location[1,5,53,58-60], and have not simultaneously sampled multiple body regions to assess important microbiome territory specificity, dynamics and interactions from numerous host microbial microenvironments[16,61-64]. Finally, they have not simultaneously measured multiple host clinical and molecular phenotypes[15,25,28,65-67] in order to investigate the molecular relationships underpinning host-microbiome interactions associated with health and disease[68,69].

Large collaborative efforts, such as the Integrative Human Microbiome Project (iHMP), offered the opportunity to overcome these restrictions through integrative Personal Omics Profiling (iPOP, https://med.stanford.edu/ipop.html) of a well-characterized human longitudinal cohort[70-72]. In order to provide a quantitative understanding of the individuality and stability of microbiomes across multiple body sites and through perturbations in host phenotype in both healthy individuals and those at risk for T2DM, we have characterized the microbiome of four body sites and its relationships with omics and clinical features in 86 adults followed for up to six years. For each individual, the stable core/shared and personalized microbiomes were identified, compared between body sites and studied for their relationship with host multi-omics and clinical features. We followed dynamics across both individuals and diseases (*e.g.*, respiratory viral infections (RVI)), as well as in IR and IS participants.

Results

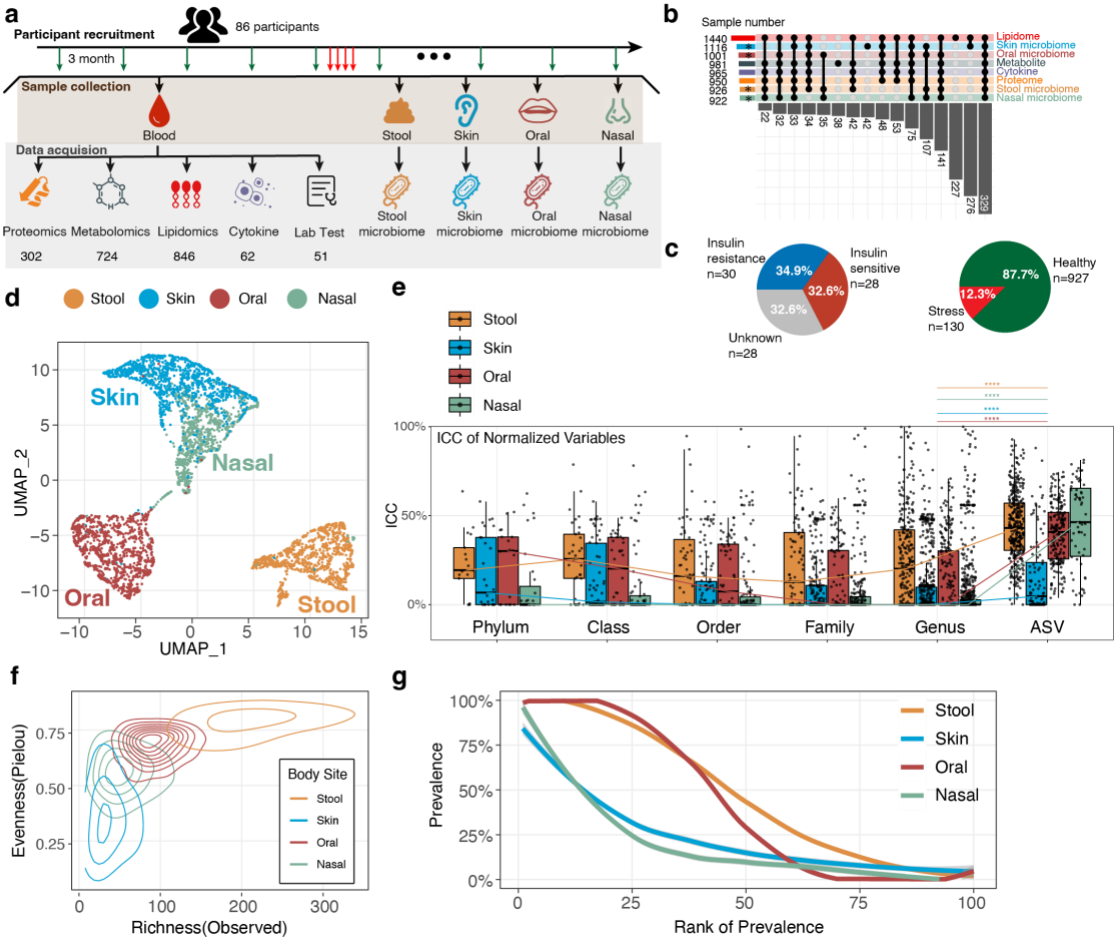


Figure 1. Longitudinal Profiles of the Microbiome at Four Body Sites

a. Graphical representation of the study design (Icons used in this figure are adapted from www.iconfont.cn). A group of 86 participants was examined every 3 months, and those with self-reported stress such as viral infected donated samples more frequently when symptomatic. At each visit, blood, oral/nasal/skin swabs, and stool samples were collected. Host multi-omics (proteomics, metabolomics, lipidomics, cytokines (also include chemokines and growth factors)) were generated using the sample stored. Blood samples were additionally sent for clinical laboratory testing (https://stanfordhealthcare.org/medical-tests/b/blood-test.html).

b. Data overlap between different types of omics. Each omics data type is indicated in the legend and *X*-axis using the same color and order. The number of samples from each omics type are marked on the *X*-axis, where microbiome samples are highlighted with an asterisk. Interaction size marks the number of overlap samples where dots are crossed by a vertical line.

c. The proportion of stress (infection, immunization, antibiotic use, etc.) and healthy samples, as well as the proportion of insulin-sensitive and insulin-resistant participant samples.

d. UMAP of the four body sites' microbiomes. Each dot represents a single microbiome sample collected from a specific body site, as indicated by its color.

e. Intraclass correlation of microbiome at each taxonomy level. The dot at each taxonomy level represents a taxonomic unit, as indicated on the *X*-axis. The Wilcoxon two-way nonparametric analysis was performed to test if the given two columns are significantly different. *P*-values are adjusted by the BH method, and the significance are summarized as follows: *: adjusted *p*-values < 0.05; **: adjusted *p*-values < 0.01, ***: adjusted *p*-values < 0.001, ****: adjusted *p*-values < 0.0001.

f. The density distribution of microbiome samples from the four body sites. The *X*-axis shows the observed species (ASV) richness (metric: ACE), and the *Y*-axis shows the evenness (metric: Pielou).
g. Rank Prevalence Curve of genera at each body site. The 100 microbiome genera with the highest longitudinal prevalence at each body site are sorted from high prevalence to low prevalence on the *X*-axis. The colored curve connects the genera from the same body site.

Description of the study design

We comprehensively profiled the host microbiome at four different human body sites (stool, nasal, skin, and tongue/oral cavity) from a cohort of 86 participants who were sampled quarterly for up to six years (median 1,126.6 ± 455.8 days). The cohort consisted of 41 males and 45 females ranging in age from 29 to 75 years old (55 ± 9.8 years old) with BMIs between 19.1- 40.8 kg/m^2 (28.31 ± 4.44 kg/m^2) (The cohort description is in **Extended Data Table 1**). The gut microbiome was assessed through the collection of stool samples, whereas oral, skin, and nasal microbiome were sampled through swabs of the retroauricular crease, oral cavity/tongue, and anterior nares, respectively (**Fig. 1a,** see **Methods**). Sampling was performed quarterly when individuals were healthy with 3-7 additional samples collected within five weeks (12% of the dataset) when encountering a period of stress such as respiratory illness, vaccination, or antibiotic usage. We performed bacterial 16S ribosomal RNA gene sequencing which ,importantly, includes a variable region (see **Methods**) that was used to identify the amplicon sequence variations (ASVs)[73]. The ASV analysis generally enables most bacterial taxa to be resolved at the genus/species level[74,75].

One important and unique feature of this cohort is that participants were also deeply phenotyped at each timepoint using multi-omics and clinical marker analyses[72] (**Fig. 1b**; see **Methods**). Multi-omics analyses used in this study included untargeted proteomics (302 proteins), untargeted metabolomics (724 annotated metabolic features), and targeted lipidomics (846 annotated lipids), as well as 62 targeted cytokine and growth factors measurements[72,76]. Fifty-one clinical markers (e.g., C-reactive protein (CRP), fasting glucose (FG), hemoglobin A1C (HbA1C), low-density lipoprotein (LDL), and high-density lipoprotein (HDL)) were also measured from blood samples collected at each timepoint (**Fig. 1a**). Advanced glucose control measurements included an annual oral glucose tolerance test on the entire cohort as well as steady-state plasma glucose (SSPG) measurement (performed on 58 individuals), the gold standard test for measuring insulin sensitivity/resistance. In this cohort, 28 of the 86 individuals were categorized as IS, and 30 were IR[77]. (**Fig 1c**) Overall we analyzed a total of 3,058 visits, 5,432 biological samples (1,467 plasma samples, 926 stool samples, 1,116 skin samples, 1,001 oral samples, and 922 nasal samples) as well as clinical tests, and generated a total of 118,124,374 measurements. The microbial and other data can be found at our data portal https://med.stanford.edu/ipop.html.

Microbiome distribution and ecology on different body sites

We first analyzed the general demographic composition of the microbiome from each of the four body sites using Uniform Manifold Approximation and Projection (UMAP) dimensional reduction. Consistent with previous reports[14,64,76,78,79], there was a clear separation between body sites, reinforcing the strong territory specificity of the microbiome at each site[13,14,80]. For the stool samples, a subset was dominated by genera from Bacteroidetes and Firmicutes with a gradient of abundance distributions; these samples were distinct from those with high *Prevotella* abundance[53,78,79]. The newly defined genus *Phocaeicola*[81] had little effect on the general Firmicutes/Bacteroidetes gradient distribution. However, among samples with high Bacteroidetes, *Phocaeicola* levels did distinguish samples with high *Bacteroides* levels from those with lower levels. The oral microbiome was dominated by *Prevotella*, *Streptococcus*, *Veillonella*, *Haemophilus*, *Neisseria*, and *Leptotrichia* as described previously[82-84]. The skin and nasal microbiome samples were the most similar and exhibited a triangular distribution driven largely by three dominant genera: *Cutibacterium*, *Corynebacterium*, and *Staphylococcus*[5,15,64,80] (**Fig. 1d**, **Extended Data Fig. 1**). These microbes and their distribution have also been reported in other cohorts[15,61,85]. However, we were able to extend these findings by comparing their inter- and intra-individual covariance longitudinally.

The total variance of microbiome increases at amplicon sequence variant level

In order to gain a more comprehensive understanding of the variance across individuals, we estimated the intraclass (intra-individual) correlation coefficient (ICC) at various taxonomic levels. We previously found that the stool microbiome from each individual was more similar to each other than to those of other participants[57,72]. In our expanded microbiome data, we found that microbial personalization was greater at the ASV level than at broader taxonomic resolution (ICC 43.7% in stool, 15.7% in skin, 39.9% in oral, and 45.8% in nasal), demonstrating stronger microbiome individualization with finer taxonomic resolution (**Fig. 1e**). Interestingly, although the nasal and skin microbiome are most similar in taxa (**Extended Data Fig. 1**), the nasal microbiome displayed a much higher personalization/ICC at the ASV level (ICC: 45.8% vs 15.7%), validating[80] that environmental factors may influence the skin microbiome more than the nasal microbiome despite their similar ecological features (i.e. air exposure/aerobic) and phylogenetically similar dominant bacterial members.

Several studies have revealed seasonal changes in the gut microbiome[86-88], however one study claimed no such variability[53]. We found that the season had a weak effect on the microbiome in each of the four body sites compared with the impact of individuals: among the four body sites, the seasonal effect was higher in the skin microbiome (4.08%) than in the other samples (stool: 0.39%, oral: 2.73%, nasal: 0.86%) (**Extended Data Fig. 2**), indicating that direct external exposures may have the greatest impact on personal skin microbiome relative to internal sites. The higher seasonal effect on oral relative to nasal is interesting and likely reflects the influence of external seasonal diets[54,89,90], as the oral microbiome is more closely related to diet compared to other body sites (**Extended Data Fig. 3a**). Additionally, two individuals had microbiome data that overlapped airborne environmental exposure data, and one with chemical exposure data[91]. In both individuals, variance decomposition revealed that the covariance of the exposome associated with the skin microbiome was much larger than that of other sites, (**Extended Data Fig. 3b**) indicating that the environment exerts a much greater influence on the skin microbiome than at other body sites, although we expect that the environment may still affect internal sites such as the oral or stool microbiome, as previously reported[92]. Overall, our observations support the concept of high niche specificity and individuality of the human microbiome at single nucleotide resolution and show that this is evident at multiple body sites over time[1,93].

Microbiome diversity demographics across body site and individuals

In addition to the personal longitudinal variation, other ecological features of the microbiome differed between the four body sites (**Fig. 1f**). The stool microbiome was the most diverse population in both richness, *i.e.*, the number of different microbes (measured by Abundance-based Coverage Estimator, ACE), and evenness, *i.e.*, how well represented species are, relative to other species (measured by Pielou's evenness index). On the other hand, although the skin and nasal microbiome showed similar richness distributions, the skin microbiome showed a more skewed population, due to its lower evenness compared to the nasal microbiome. These ecological characteristics are important health indicators and are often altered with the progression of disease. Consistent with previous reports[32,72], we found the gut dysbiosis associated with IR was characterized by a significant decrease in alpha diversity of the stool microbiome ($t = 2.8462$, p-value = 0.0067). In addition, the centroid of diversity-evenness scatter plots shifts at all four body sites in IR relative to IS individuals such that skin microbiome diversity ($t = -2.9102$, p-value = 0.0057) and evenness ($t = -2.4393$, p-value = 0.019) were significantly higher in IR individuals (**Extended Data Fig. 4**). These shifts likely suggest that the dysbiosis associated with IR is systemic and not limited to the microbiome of the intestine.

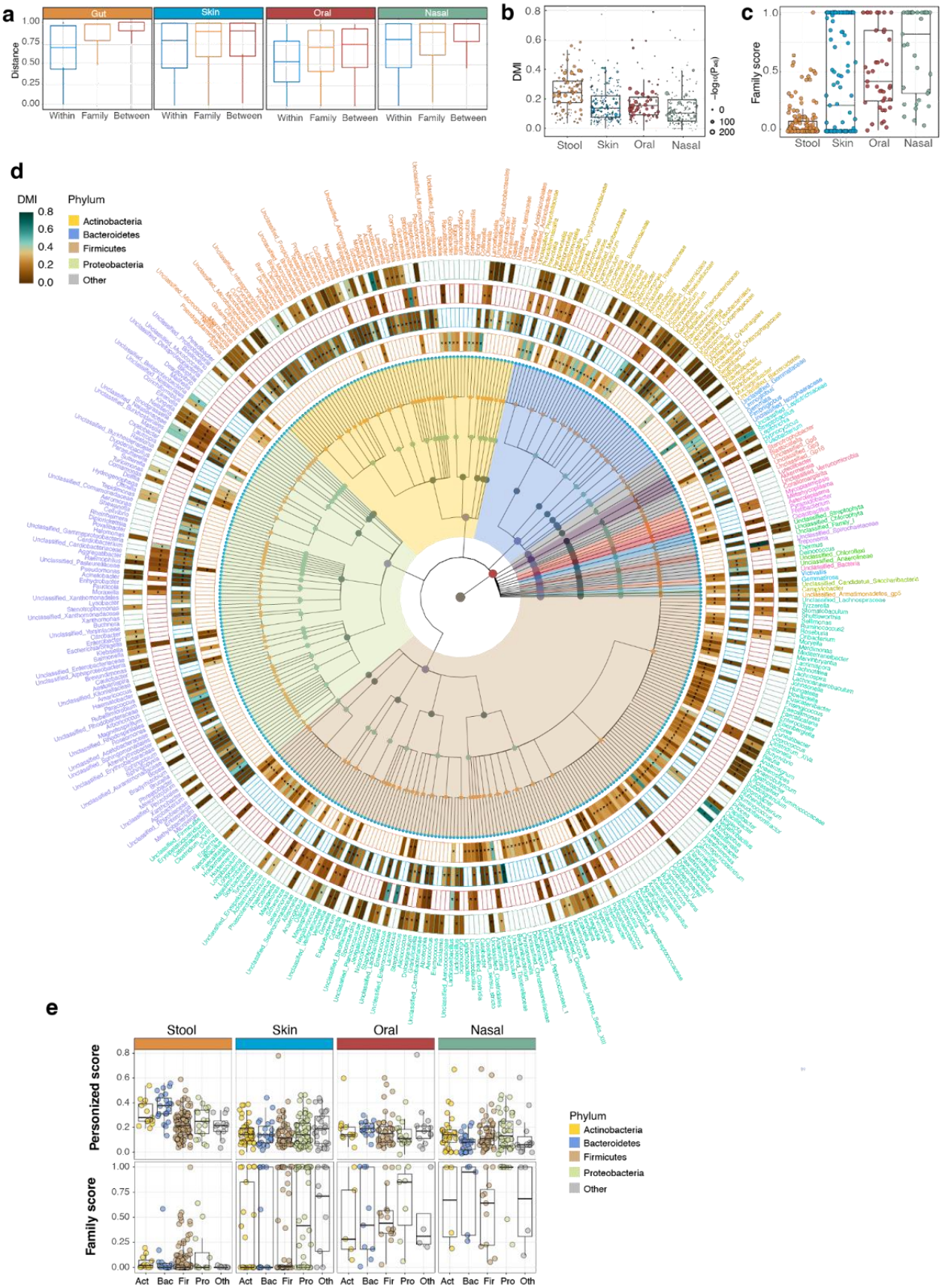
Stool and oral microbiomes maintain rich prevalent core communities

A limited number of previous studies have examined longitudinal microbiome dynamics and have mostly focused on the microbiome from single body site[1,2,94,95]. In our analysis, the microbes that show a high consistent presence over participants' longitudinal timepoints are referred to as the "core microbiome" (see **Methods**) and are believed to represent genera with indispensable functional potential for the individual[96,97].

Longitudinal analysis (3.1±1.25 years) of microbial prevalence at the genus level across the four body sites revealed that stool and oral microbiomes maintain a rich core microbiome comprised of more than 25 genera that are highly prevalent over time (Longitudinal Prevalence > 80%), compared to nasal and skin microbiomes, which only had three core genera (**Fig. 1g**). Interestingly, such highly prevalent genera are not necessarily high in relative abundance (*i.e.*, *Coprococcus* Mean Prevalence = 80.75%; Mean Relative Abundance = 0.544%), indicating that even low abundance strains may be indispensable (**Extended Data Fig. S5**). We found that the number of core genera in stool and oral microbiome is negatively associated with steady-state plasma glucose (SSPG) (Spearman Rho = -0.52, *p*-value = 0.00047) and BMI (Spearman Rho = -0.40, *p*-value = 0.005), respectively (**Extended Data Fig. S6**).

Given the link between microbiome longitudinal prevalence with SSPG and BMI, we further examined whether microbiome ecology differs significantly between IR and IS individuals. Compared with the IS population, IR subjects showed a significantly higher number of skin core genera (t = -2.5856, *p*-value = 0.014) and lower number of stool core microbiomes (t = 2.9659, *p*-value = 0.0051) (**Extended Data Fig. 7a**). Importantly, butyrate-producing bacteria such as *Coprococcus*, *Parasutterella*, and *Butyricicoccus* were more likely to be stool core microbiome members in IS individuals; whereas diabetes-related opportunistic pathogens such as *Finegoldia*[98-100] and *Acinetobacter* [101-104] were more prevalent in the skin core microbiome of IR individuals (**Extended Data Table 2**). Butyrate is an essential metabolite in the human colon and contributes to gut barrier function and has anti-inflammatory properties[105]. In addition to the genera mentioned, we found a clear divergence of the rank prevalence curves in stool and skin microbiome (**Extended Data Fig. 7b**), demonstrating an IR-associated global microbial prevalence shift at these two sites.

Additionally, the relative abundance of several taxa also differed between IR and IS individuals. The stool microbiome of IR individuals showed an increase of genus *Phocaeicola* (LEfSe_effect size: 0.03; BH-adjusted *p*-value = 0.017) and a reduction of the genus Unclassified *Ruminococcaceae* (LEfSe_effect size: 0.017; BH-adjusted *p*-value = 0.0039) whereas the skin microbiome exhibited a decrease in the genus *Cutibacterium* (LEfSe_effect size: 0.069; BH-adjusted *p*-value = 0.007) with an increase of the genus *Peptoniphilus* (LEfSe_effect size: 0.0076; BH-adjusted *p*-value = 0.0022); *Peptoniphilus* has been previously associated with skin in individuals with diabetes[99,106,107] and necrotizing infections[108] (**Extended Data Figure 8**; **Extended Data Table 3**). Interestingly, we did not find a comparable statistical difference in the prevalence patterns of oral and nasal microbiomes in IR vs IS participants. Overall, these results indicate that the skin and stool microbiomes are less stable in IR patients.


a
Gut
Skin
Oral
Nasal
Distance
Within Family Between
b
DMI
Stool Skin Oral Nasal
−log10(P_adj)
c
Family score
d
DMI
Phylum
Actinobacteria
Bacteroidetes
Firmicutes
Proteobacteria
Other
e
Stool Skin Oral Nasal
Personized score
Family score
Act Bac Fir Pro Oth

293

Figure 2. The individuality of microbiome differs significantly across genera and body sites

a. The Bray Curtis (BC) dissimilarity between sample-pairs from various groups. Within: sample pairs are from the same participants; Family: Sample pairs are from different participants living in the same household; Between: sample pairs are from different participants.

b. The Degree of Microbial Individuality (DMI) scores at different body sites. Only those genera whose intra- and inter-individual BC distances differ significantly are included. Size of the dot indicates the Log-10 transformed, BH-adjusted *p*-value when comparing the $BC_{intra\text{-}individual}$ and $BC_{inter\text{-}individual}$. The significance was generated by Wilcoxon signed-rank test, BH-adjusted *p*-value are marked as *: adjusted *p*-values < 0.05; **: adjusted *p*-values < 0.01.

c. Family Score (FS) of each genus is grouped by body sites. The significance was generated by Wilcox signed-rank test, *p*-values are marked as *: BH adjusted *p*-values < 0.05; **: BH adjusted *p*-values < 0.01.

d. Cladogram of DMI across all genera. The cladogram within the circles illustrates the phylogenetic relationships among all nodes (genus). The heatmap circles represent the color-coded DMI score. Four heatmap circles indicate the DMI of the genus for the stool, skin, oral, and nasal microbiome in order from inner circle to outer circle. The asterisk on the heatmap for the species and body site indicates a significant difference between intra- and inter-individual BC distances based on the permutation test (see **Methods** for detail).

e. The DMI and FS for different phyla and body sites. Act: Actinobacteria, Bac: Bacteroidetes, Fir: Firmicutes, Pro: Proteobacteria, Oth: other genera not belonging to the aforementioned four phyla.

Longitudinal profiling reveals microbial stability and individuality

Our understanding of the temporal stability of the microbiome at multiple body sites and their dynamic relationship with the host is limited. We assessed the stability of the microbiome at the genus level across time points in the same individuals and compared the variance across individuals. Our hypothesis is that the microbiome stability varies between different taxa and is highly site-specific. For each individual, we calculated the "degree of microbial individuality" (DMI, see **Methods**) for each genus. Higher DMI values imply that longitudinally, a particular bacterium genus is more homogenous inside the same individual than it is across individuals. We also had three participating families totaling six genetically unrelated individuals and calculated a preliminary "family score" (FS, see **Methods**) to examine the microbial dissimilarity of individuals within the same household. Regardless of the high variability in intra-individual dissimilarity across time (median BC distance 0.29-0.47), all four body sites show a significantly lower intra-individual and within-family variability compared to that observed across individuals, as expected[109] (**Fig. 2a, Extended Data Table 4**). Surprisingly, we also identified cases of intra-individual distance reaching 1.0, indicating a total replacement of the ASV composition within the genus of an individual's microbiome over time. This replacement was more likely to occur among the genera of the nasal microbiome (35.5% of all pairs) as compared to other body sites (stool is 24.2%, skin is 24.4%, and oral is 2.7%). In contrast to the replacement of a single genus, full replacement of the entire microbial community was rarely detected (1 nasal full replacement, 8 oral full replacements, and 1 skin replacement out of 45,455 possible distance pairings; **see methods**). Such a contrast of replacement rate revealed that despite the overall stability, the microbiome was very dynamic at the genus level.

The overall DMI of different genera was independent of their relative abundance and was greater in the stool microbiome compared to other body sites. (**Fig. 2b**) This observation was further supported by the fact that the stool microbiome had the lowest family score (FS) compared to other body sites. (**Fig. 2c**) This distribution of FS among genera within the stool microbiome reveals a relatively high degree of dissimilarity amongst genus members within the same household. On the other hand, all genera of nasal and oral microbiome have FS scores greater than 0, indicating that the oral and nasal microbial communities within the same household share a high-level degree of similarity, presumably influenced by the common living environment[1,110] or direct microbiome exchanges[111].

The large differences of the DMI values at the four body sites (Stool: 0.23 ± 0.13; Skin: 0.11 ± 0.12; Oral: 0.14 ± 0.14; Nasal: 0.09 ± 0.12) might be attributable to the great taxonomic complexity of microbiome ecosystems. Therefore, we evaluated the microbial populations based on their phylogenetic relationships and found that DMI of specific taxa was frequently associated with niche specificity (**Fig. 2d**, **Extended Data Table 5**). For instance, *Corynebacterium* showed the highest DMI among nasal microbiomes, and *Bacteroides* showed the highest DMI among stool microbiomes. On the other hand, environment-associated bacteria, such as *Klebsiella*[112] and *Haemophilus*[113], showed a generally low DMI in all four body sites, indicating that the environment exerts weaker effects on the individuality of the native host microbiome. Overall, members belonging to the phylum Bacteroidetes had significantly higher DMI (W = 2,140.5, BH-adjusted *p*-value = 1.871×10^{-6}) than those of Firmicutes in stool, but not in other body sites, even though other body sites also harbor several genera belonging to Bacteroidetes (**Fig. 2e**). This may be the result of Bacteroidetes' high level of adaptive evolution[114,115] or high colonization resistance[116]. Overall, the measurement of DMI and FS for each specific genus specific microbial host specificity within and across body sites. Furthermore, DMI constitutes an important physiological metric for the taxonomic composition of the community (*i.e.*, enterotype in stool microbiome[79,117]) as well as potential insights into the environmental impact on the hosts microbiome.

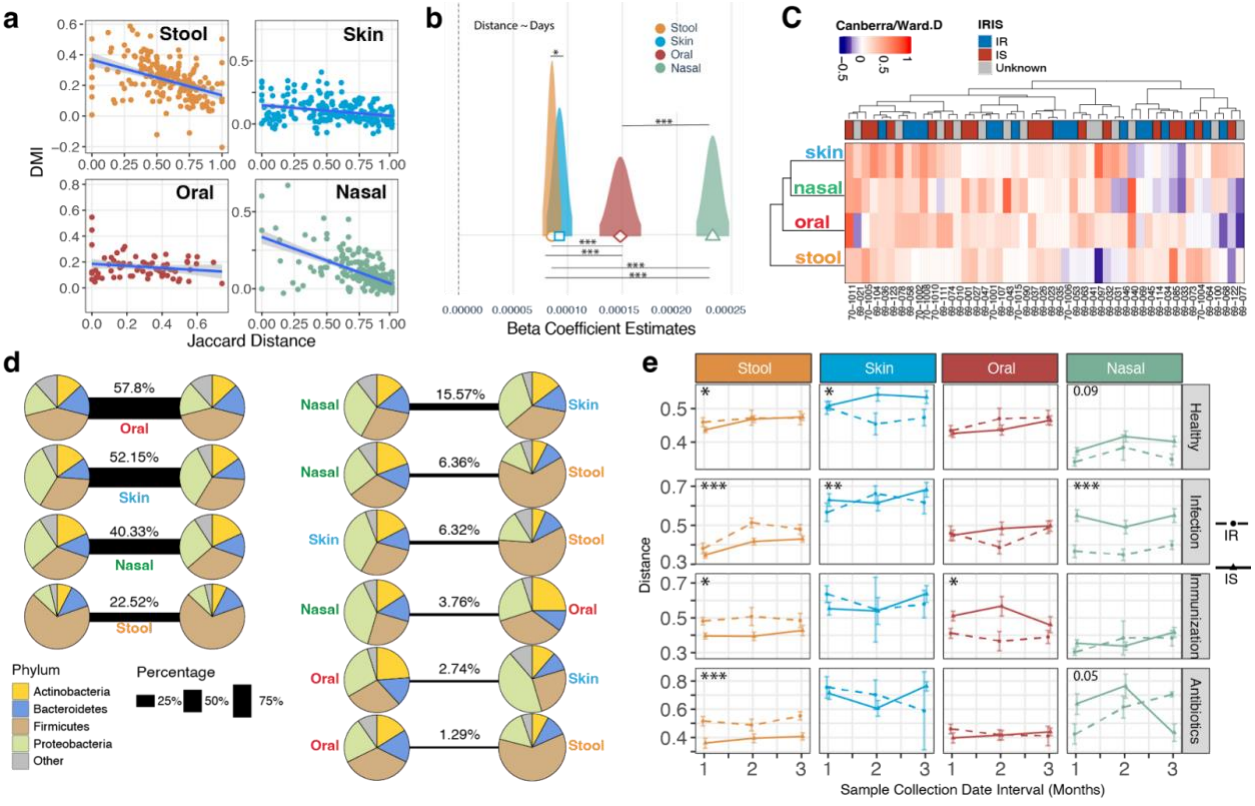


Figure 3. Temporal stability of microbiomes was associated with individuality and disrupted by stress events.

a. Correlation plot between the mean Jaccard distance of a given genus in all intra-individual sample pairs (*X*-axis) and the mean DMI of this genus (*Y*-axis). The beta coefficient and *p*-value for each correlation are Stool (n = 185): beta estimate: -0.23, *p*-value = 8.77×10^{-10}; Skin (n = 214): beta estimate = -0.085, *p*-value = 4.01×10^{-5}; Oral (n =70): beta estimate = -0.080, *p*-value = 0.155; Nasal (n = 222): beta estimate = -0.31. *p*-value < 4.12×10^{-24}.

b. Relationship between the dissimilarity and collection date interval of intra-individual sample pairs. The linear regression was estimated using a linear mixed effects model (see **Methods**). The beta coefficient

distribution was plotted and the between body site comparison was performed by setting body site variable as an interaction term. BH adjusted *p*-value was annotated as *: adjusted *p*-values < 0.05; **: adjusted *p*-values < 0.01, ***: adjusted *p*-values <0.001.

c. Inter-body site correlation of BC-distance growing trends in function of time interval. The heatmap summarizes the individual based on the correlation coefficient between sample pair's BC distances and the collection date intervals. The hierarchical clustering represents the Canberra correlation of Ward distance. Individual's insulin sensitivity status is marked at the bottom of the column clustering dendrogram. (IR [blue]: Insulin Resistant; IS [red]: Insulin Sensitive; Unknown [gray]: do not have accurate insulin sensitivity measurement).

d. The overview of Spearman correlation between intra-individual microbiome relative abundance. The left panel shows correlation within the same body site, and the right panel shows the correlation between body sites of the same individual. Pie chart shows the percentage of phylum among significantly correlated pairs. The percentage indicates the number of significant correlations relative to the total number of possible correlation pairs.

e. The microbiome shifts from healthy to healthy or stress events over three months. The BC distance (the *Y*-axis) of intra-individual sample points was examined for all subjects at 1, 2, and 3 months, grouped by their insulin sensitivity. Four comparisons were made and arranged vertically for each body site: 1) healthy to healthy time point, 2) healthy to infection time point, 3) healthy to immunization time point, 4) healthy to antibiotics time point. Two-way ANOVA was performed to determine if IR and IS show significantly different mean values of BC distance. BH adjusted *p*-values were annotated as *: adjusted *p*-values < 0.05; **: adjusted *p*-values < 0.01, ***: adjusted *p*-values < 0.001.

The highly personalized microbiome is more stable over time

A number of earlier studies demonstrated that microbiome stability is highly individualized[4,5,58]. Therefore, we investigated the relationship between microbial individuality and stability; of particular interest was microbial stability at specific body sites. We first examined the ASV consistency of a specific genus during recolonization; recolonization is defined as the re-identification of a genus after being undetectable in one or more consecutive samples. The overall recolonization rate (measured by 1 - Pairwise Jaccard Distance)[118,119] was significantly associated with DMI on all three body sites except for the oral microbiome (**Fig. 3a**), which had a high recolonization rate compared to the other three body sites. Additionally, this relationship was strongest in the nasal microbiome (**Fig. 3a**), which may explain the high ICC observed in this body site in our analysis above. We did not observe any difference in recolonization rate between IR and IS individuals (**Extended Data Figure S9**). Overall, our results indicate that highly individualized strains are always more likely to recolonize after falling below the detection limit, which may deflect a host selection on the microbiome. These observations are similar to the observation made with human fecal transplantation studies, in which the strains that colonize hosts are those strains most closely related to the host's native microbial communities[120,121].

With our longitudinal design, we were also able to evaluate the stability of the microbiome over time by tracking the rate at which the dissimilarity between pairs of samples changed over the interval between collection dates. This rate has been previously reported to be higher in IBD-related gut dysbiosis[22]. Based on our analysis, the stool microbiome changed at a slower rate across time than any other sites and was substantially slower than that of the nasal site, which changed the fastest (*p*-value < 0.001) (**Fig. 3b**). Additionally, IR individuals showed significantly decreased stability in both stool and skin microbiome compared with IS demonstrated by linear mixed models (Stool *p*-value: 1.82×10^{-06}, Skin *p*-value 2.84×10^{-12}). Consistent with the instability of the stool and skin microbiome in IR individuals, linear discriminant analysis revealed that microbial abundance differences between IR and IS participants were larger in these samples than in oral or nasal samples (**Extended Data Fig. S8**).

Systematic cooperative dynamics is detected in microbiome within and between body sites

For each individual, we also investigated whether microbiome dynamics was associated across body sites, i.e., whether temporal variations observed at one site also occurred at other sites, both at the general level as well as for individual taxa. Microbial correlations were found to occur across all body sites (**Fig. 3c**). Using hierarchical clustering, we discovered that the time-related alterations of the skin microbiome were most strongly associated with those of the nasal microbiome, and the stability of the stool microbiome is less related to that of the other body sites (Canberra distance: skin vs nasal, 24.2, skin vs oral: 27.9, skin vs stool: 32.0, oral vs nasal: 26.1. **Fig. 3c**). The correlation of time-dependent microbiome stability between body sites strongly suggests systematic ecological changes across the body, presumably in coordination with the host. Since these relationships may be altered under chronic disease conditions, we compared the inter-body site correlation of time-related stability between individuals classified as IS and IR. Surprisingly, although the strong link between the time-related stability of the skin and nasal microbiome holds true for both IR and IS participants, the correlation between skin and oral microbiome can only be detected in IS but not IR (**Extended Data Fig. S10**). This result suggests that IR status may weaken the stability of the skin or oral microbiomes, due to impaired host modulation of the microbiome stability during IR disease state.

Since microbiome members are interdependent and usually associated as functional guilds[122,123], intra-individual microbiome community fluctuations typically happen as a group. However, it is not known which body sites have the most microbial interdependence. Therefore, we systematically analyzed the longitudinal dynamics between all bacterial members per body site. Consistent with the well-documented stool microbiome interdependency, we found that 22.52% of all stool genus pairings (5,671) are significantly correlated within individuals, and these are mostly genera belonging to Firmicutes **(Fig. 3d).** Surprisingly, this co-association was highest in the oral microbiome (57.8%), followed by the skin (52.15%) and nasal (40.33%), highlighting the strong synergistic interaction of microbiome members at each body site, particularly in oral and skin.

We further investigated the microbial genera correlation between body sites within same individuals using a linear mixed model. As expected, members from the skin and nasal sites were the most correlated (15.57% of all possible pairs) **(Fig. 3d)**. However, there is still a remarkable territory specificity among the core microbiome of each body site. For example, the three most abundant genera (*Cutibacterium*, *Corynebacterium*, *Staphylococcus*) are not longitudinally correlated between skin and nasal site. Likewise, the relative abundance of *Prevotella* at the oral site does not correlate with its abundance in the stool microbiome, consistent with our observation on the high FS of *Prevotella* in oral but not stool (**Extended Data Table 6**). Although microbiome translocation between body sites has been reported[124,125], our results suggest that these translocation cases are likely not to happen among the dominant taxa of a body site, possibly due to the niche specificity associated with each body site such as temperature, moisture, nutrition availability, and local immunological microenvironment, as well as the strong microbial interdependence at each body site.

In addition to examining intra-individual correlation across body sites, we also examined inter-individual correlation across body sites. It has been shown[15,79] that the stool microbiome is comprised of three enterotypes according to their predominant taxa: *Bacteroides*, *Prevotella* and Unclassified *Ruminococcaceae*, and as discovered above, such micro-biotypes exist at other body sites. How these dominant microbes correlate with themselves and other taxa across body sites is not known. We found consistent associations between dominant taxa. For instance, individuals who host high levels of *Cutibacterium* in the skin microbiome host high levels of *Cutibacterium* in the nasal microbiome (beta = 0.56, p-value < 0.0001); individuals with high skin *Cutibacterium* maintain high *Bacteroides* in the stool microbiome (beta = 0.52, p-value < 0.001) and *Leptotrichia* in the oral microbiome (beta = 0.43, p-value < 0.0001). Likewise, individuals with high stool Unclassified *Ruminococcaceae* have low *Veillonella* in the oral microbiome (beta = -0.35, p-value < 0.0001) (**Extended Data Table 7**). These findings clearly suggest that the establishment of micro-biotypes (*i.e.*, enterotypes in the stool microbiome) are governed by host-

specific genetic and/or environmental confounding factors. Our study design also allowed us to capture events that are known to perturb the stability of the microbiome, such as infection, vaccinations, or using antibiotics[72,76,126,127]. We found that within a three-month period, temporal microbial shifts during healthy visits and perturbations were significantly higher in the stool microbiome of IR individuals than IS, but this trend is not obvious on other body sites (**Fig. 3e**).

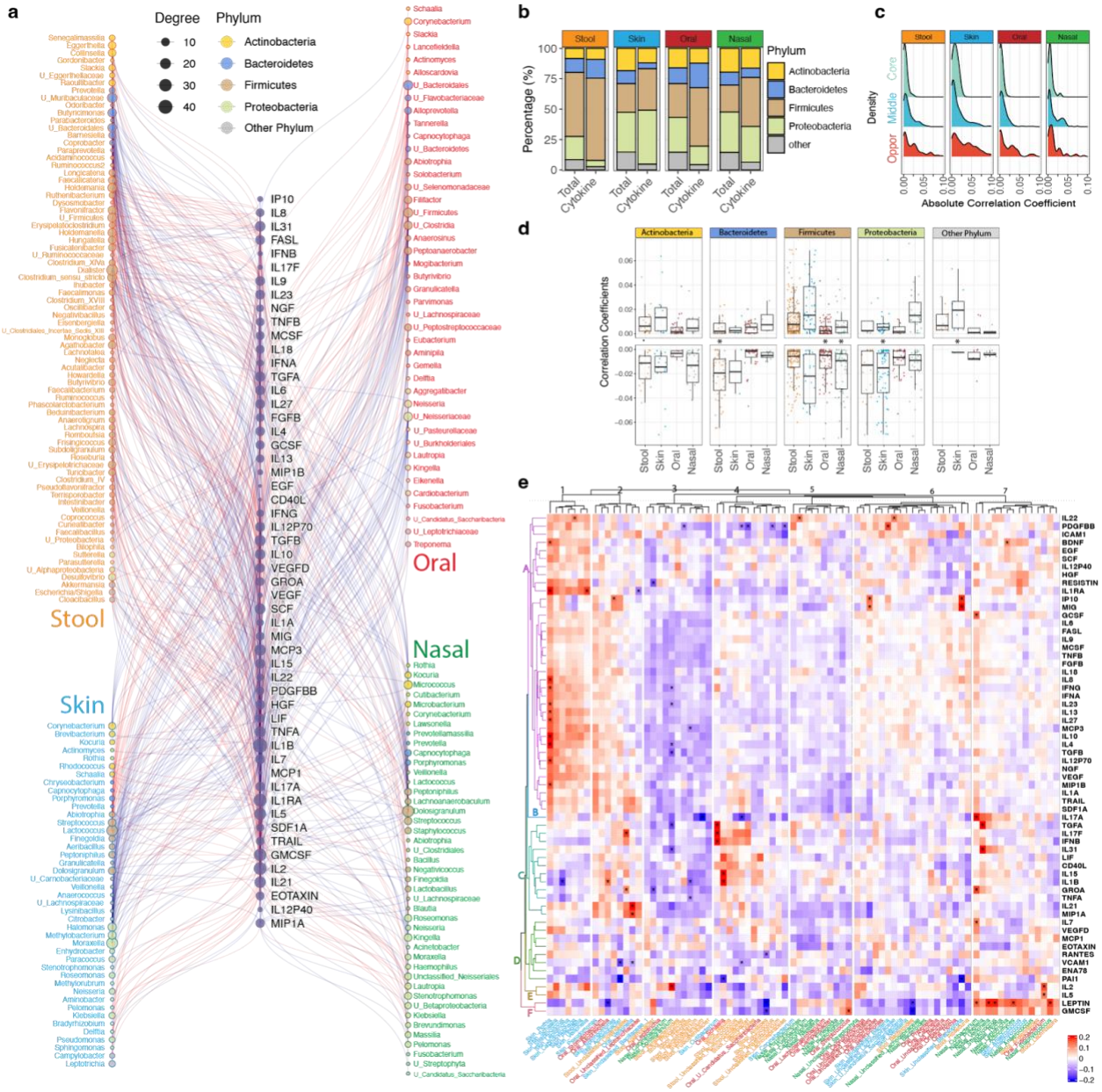


Figure 4. Systematic connections between circulating cytokines and microbiomes

a. The correlation between cytokines and microbiomes from various body sites. A customized mixed effects model was used to evaluate the correlation (see **Methods** for details). If the coefficient is positive, the association between a particular cytokine and microbiome is depicted by a red line, and if it is negative, by a blue line. The circle size corresponds to the number of significantly correlated cytokines (BH adjusted *p*-value > 0.05) with the microbiome genera for which the absolute correlation coefficient is greater than 0.005. The origin of the annotated genus' body site is indicated by the color of the genus label. (Orange: Stool; Blue: Skin; Red: Oral; Green: Nasal)

b. Percentage of cytokine-related genera summarized by their phylum. Total: this column contains all genera within a body site. Cytokine: this column contains only genera that are significantly associated with cytokines.
c. Microbiomes with a low prevalence have a stronger association with cytokines. Density plot displays the absolute value of the correlation coefficient for all pairs of cytokine and microbiome that are significantly associated. The result was separately plotted based on the mean prevalence of the genus. Core: Core microbiome with a mean longitudinal prevalence more than 80%. Oppor: Opportunistic microbiome with a mean longitudinal prevalence less than 20%. Middle: All microbiome genera that do not belong to the core or opportunistic group. Comparisons between the correlation coefficient (absolute value) of core microbiome and opportunistic microbiome are performed using the Wilcoxon signed-rank test: stool: W = 1,715, *p*-value = 1.181×10^{-12}, skin: W = 223, *p*-value = 0.004041, oral: W = 1,055, *p*-value = 7.877×10^{-6}, nasal: W = 482, *p*-value = 0.004691.
d. The comparison of correlation coefficient by body site and phylum. BH adjusted *p*-value (. adjusted *p*-value<0.1; * adjusted *p*-value<0.05) in the middle row represents the Wilcoxon signed rank sum test between correlation coefficients that are positive and absolute values of all negative correlation coefficients in the same column.
e. The Spearman correlation coefficient between cytokines and the most diverse (by observed OTU) 20 genera in each body site. The column and row are clustered using hierarchical clustering with a set k (k=7 for column, k=6 for row), and the number of clusters is annotated. Red indicates the positive correlation and blue indicates the negative correlation. Annotations on the bottom are color-coded based on the body sites to which each genus belongs. (Orange: Stool; Blue: Skin; Dark Red: Oral; Green: Nasal)

Systematic interactions between circulating cytokines and microbiomes at four body sites

The dynamics of microbial stability are likely to interact with multiple host regulatory mechanisms, particularly the immune system[128-131]. We profiled a panel of 62 circulating cytokines, chemokines, and growth factors to assess the immune status of each individual at each timepoint. We previously identified a link between the plasma level of interleukin (IL)-17/IL-22 and stool *Clostridia* which is absent in IR individuals[132]. Using the same modeling strategy, we examined the interplay between host circulating cytokines and microbiome abundance at all four body sites. Overall, we identified 477 stool, 226 skin, 318 oral, and 221 nasal specific microbiome-cytokine associations (**Fig. 4a, Extended Data Table 8**). Interestingly, the cytokines associated with epithelial/endothelial growth and vascular inflammation (EGF: 42 interactions, VCAM-1: 39 interactions, IL-22: 39 interactions), IL-1 family members (IL-1b: 44 interactions, IL-Ra:34 interactions), and leptin (39 interactions) exhibited the closest relationship (*i.e.*, most associations) with the microbiome, corroborating our previous stool findings[133]. Beta correlation coefficient comparison across all cytokines further identified a subgroup of cytokines including IL-1B, IL-1Ra, MCP3(CCL-7), and IL-23 as the strongest correlative cytokines with the microbiome (**Extended Data Fig. 11**). There was also a clear pattern of body site specific interactions that may explain the microbiome niche specificity. For example, *Moraxella* negatively correlated with 23 cytokines on the skin but only three cytokines in the nasal cavity (**Extended Data Table 9**). The fewer cytokine-*Moraxella* correlations in the nasal cavity suggest a lower immune response which may explain why *Moraxella* is specific to the nasal cavity[14,134].

We also investigated if members of certain phyla are more likely to engage with cytokines, which through interactions with the host immune system, may contribute to their stability and individuality. As expected, we found that members from the phylum Firmicutes are significantly correlated with cytokines (**Fig. 4b**) [X_{stool}^2 = 19.343, $p_{stool}^2 = 1.092 \times 10^{-5}$; X_{skin}^2 = 10.418, p_{skin} = 0.001248, X_{oral}^2 = 30.935, $p_{oral} = 2.668 \times 10^{-8}$; X_{nasal}^2= 31.396, $p_{nasal}^2 = 2.104 \times 10^{-8}$], consistent with Firmicutes' higher FS and lower DMI compared with other phyla in stool. This trend in the stool microbiome was mostly driven by the class *Clostridia*, which comprises 129 genera of the 194 total genera of Firmicutes, suggesting that *Clostridia* members may be more prone to engage with the host immune system. The increased correlation of Firmicutes with cytokines is also found in the skin, nasal and oral sites. Of note, the increase of Firmicutes on microbiome

from different body sites has previously been associated with obesity[135], oral dysbiosis in IBD[136] and skin dysbiosis in psoriasis[137], all of which involve inflammation. Importantly, we found that opportunistic microbes (longitudinal prevalence < 20%) have a stronger correlation coefficient with cytokines as compared to core microbiomes (longitudinal prevalence > 80%). This effect appears to be universal at all four body sites (**Fig. 4c**). This correlation is more likely due to association with Proteobacteria rather than driven by Firmicutes, since genera belonging to Proteobacteria consistently occupy a larger proportion of the opportunistic microbiome than in the core microbiome (**Extended Data Fig. 12**).

It was previously discovered that members from the phylum Proteobacteria are highly immunogenic[138-141], partially because the lipopolysaccharides (LPS) carried by Proteobacteria are potent TLR4 activators which trigger the downstream immune cascade[142]. Numerous prior studies[126,143,144] have also shown Proteobacteria increases in inflammation. Surprisingly, we found the correlations between cytokines and Proteobacteria abundance are mostly negative, except the Proteobacteria members in the nasal microbiome which do not differ between positive and negative (**Fig. 4d**). The negative correlation of Proteobacteria was stronger in stool (W = 25,315, *p*-value = 0.04439), skin (W = 4,995, *p*-value = 0.005214), and oral sites (W = 9,443, *p*-value = 9.649×10^{-5}), but weaker in nasal (W = 5,218, *p*-value = 0.2693). In addition, we found that all cytokine correlations from opportunistic stool Proteobacteria are negative (n = 10) whereas many of the high prevalence Proteobacteria members are positive (**Extended Data Fig. 13)**. Thus, we conclude that the overall negative correlation is due to the opportunistic *Proteobacteria* whereas the relatively more prevalent Proteobacteria contribute to the positive associations as previously found by us[126] and others[143,145].

The host response by cytokines and chemokines may impact the ASV complexity (observed richness) of bacteria, in addition to interacting with the relative abundance. Therefore, we examined the correlation between cytokines and richness for the 20 most-diverse genera per body site. Many of the associations in **Fig. 4e** (Column 3), which includes several members belonging to the stool Bacteroidetes (*Prevotella*, *Phocaeicola*, and *Parabacteroides)*, are negative, consistent with our observation that Bacteroidetes are more likely to be negatively correlated with cytokines (**Fig. 4e**). Leptin, an adipokine, and the cytokine granulocyte-macrophage colony-stimulating factor (GM-CSF) which are both strongly associated with BMI (**Extended Data Fig. S14**), show the strongest overall richness correlation (Row Cluster F, **Fig. 4e**). Intriguingly, seven skin-associated genera (column cluster 1, **Fig. 4e**) positively correlated with a cytokine cluster (row cluster A, **Fig. 4e**), indicating the richness of specific genera (*e.g., Rothia*, *Veillonella* and *Streptococcus*) is linked with the level of plasma cytokines (Row Cluster A). These cytokines are associated with inflammation and the increase in richness may represent a reduced skin microbiome selection by the host during periods of inflammation.

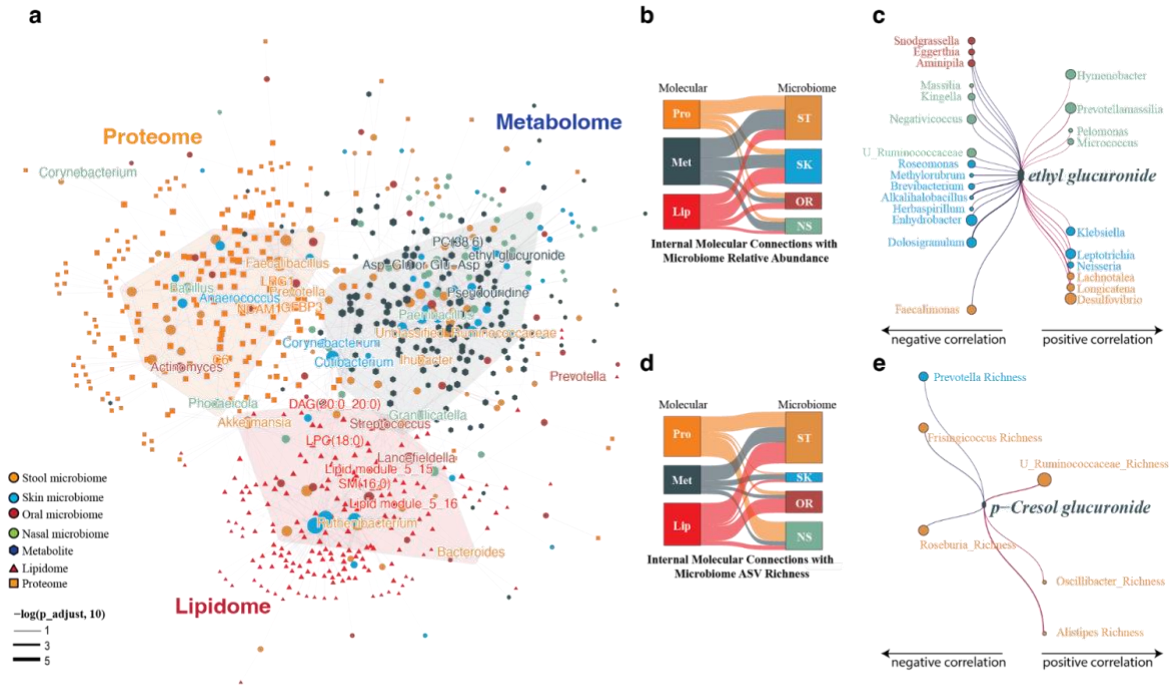


Figure 5. Interactions between plasma metabolites, lipids, proteomics, and microbiome over time

a. Network showing correlations between the relative abundance of microbiome genera at four body locations (Dark yellow filled circle: Stool; Blue filled circle Skin; Dark red filled circle: Oral filled circle; Green: Nasal) and plasma analytes (dark blue filled hexagon: Metabolome; orange filled square: Proteome; Red filled triangle: lipidome). The confidence (BH-adjusted p-value) of the correlation was represented by lines between nodes. Convex hull plot was used to annotate the Metabolome, Proteome, and lipidome zone.

b. and d) Summary of the correlations between plasma analytes and microbiome (**b:** relative abundance and **d:** observed ASV richness). Pro (light orange): Proteome; Met (Gray): metabolome; Lip (red): lipidome, ST (dark orange): stool microbiome; SK (blue): skin microbiome; OR (dark red): oral microbiome; NS (light green): nasal microbiome. The number of significantly (BH-adjusted *p*-value < 0.2) associated pairs is shown by the heights of the squares and thickness of lines.

c. Correlations (BH-adjusted *p*-value < 0.2) between genera (relative abundance) and the metabolite ethyl glucuronide. Genera with a positive correlation are placed on the right with red lines connected. Genera with a negative correlation are placed on the left with blue lines connected. Size of the dots indicates the number of significantly associated pairs that are related to each genus.

d. refer to legend section of b)

e. Correlations (BH-adjusted *p*-value < 0.2) between genera (observed richness) and the metabolite *p*-Cresol glucuronide. Genera with a positive correlation are placed on the right with red lines connected. Genera with a negative correlation are placed on the left with blue lines connected. Size of the dots indicates the number of significantly associated pairs that are related to each genus.

Longitudinal interactions between the microbiome and plasma metabolites, lipids, and proteome

To understand the inter-connection[68,143,146] between the microbiome and internal host molecules and its role in IR, we investigated the correlations between microbiome genera and plasma proteins, lipids, and metabolites in the host. To reduce the high collinearity between the numerous lipid species, we first performed dimensional reduction on the lipidomics data and the correlated lipid species were subsequently categorized into modules (**Extended Data Fig. 15, Extended Data Table 10**). To focus on longitudinal

correlations, we next reduced the between-person variation by residualizing the data in a linear mixed model with a random (individual) effect (see **Methods** for details). The interactions discovered after this data processing are therefore focused on longitudinal correlations.

Interestingly, the network was clearly partitioned according to the internal molecular types rather than microbiome body sites, suggesting that taxa-specificity is largely affected by interactions with internal molecules rather than taxa driving the host molecular composition. Strikingly, the taxa that traditionally drive enterotype: *Bacteroides, Prevotella*, and Unclassified *Ruminococcaceae*, were preferentially located in the lipidome, proteome, and metabolome regions, respectively, suggesting a specific molecular interaction for each of these taxa across all body sites (**Fig 5a, Extended Data Table 11**). While the close relationship between *Prevotella* and proteins[147,148], and *Bacteroides* with lipids have been documented previously[149,150], our results extend this observation to additional taxa as well as multiple body sites, indicating these linkages are general, systemic and robust.

Interestingly, the taxa-molecule interactions across different body sites were often consistent. For example, *Haemophilus* from skin and oral were most connected with the lipidome, whereas Unclassified *Ruminococcaceae* from stool and nasal were most connected with the metabolome. For the microbe-metabolites interactome, skin was the most connected body site (154 nodes, 235 edges) --mostly connected to the to the lipidome, followed by stool (143 nodes, 160 edges) which was mostly connected to the metabolome and proteome (**Fig. 5b**). Pathway enrichment analysis of proteins associated with microbial taxa provided additional support for these connections and their functional potentials. Analysis of the top pathways revealed that the skin microbiome-related proteins were enriched for pathways that regulate lipid metabolism. The stool, nasal, and oral microbiomes were more connected to the host immune response, including complement activation and humoral immune response. Additionally, the oral microbiome is significantly associated with functions such as the regulation of peptidase, enzyme, and proteolysis (**Extended Data Fig. S16**). Strikingly, the stool microbiome and host molecular relationships were disrupted in individuals with IR such that the complexity of their network was significantly reduced compared to IS individuals (**Extended Data Fig. S17**).

We also identified several metabolites that interact with microbiomes from all four body sites, such as ethyl glucuronide, a metabolite of ethanol[151]. The global effects of alcohol consumption towards the gut and oral microbiome were previously described[152-154]. Notably, *Neisseria* was increased in the ethanol-using group. *Klebsiella*, a genus that includes several ethanol-producing species[155,156], was positively correlated with plasma ethyl glucuronide level on skin. (**Fig. 5c**). Meanwhile, *Faecalimonas* (*a.k.a. Eubacterium*[157]), generally considered as an acetate producing bacteria[158,159], and vulnerable to alcohol consumption[160], showed a significant negative correlation with ethyl glucuronide. Finally, *Desulfovibrio*, a key bacteria genus promoting microbiome-related metabolic syndrome[161,162], was positively correlated with ethyl glucuronide. Our result strengthened the hypothesis that diseases related to alcohol metabolism might originate from the dysbiosis from different body sites.

A similar interaction was observed for the association of microbiome genera richness, another important indicator of host metabolic status[163,164], and internal plasma analytes (metabolome, proteome and lipidome). We found that interaction between ASV richness and all three molecule classes was strongest in the stool microbiome, possibly due to the high richness of the stool microbiome (**Fig. 5d, Extended Data Fig. S19, Extended Data Table S12**). Of note, we found that p-Cresol glucuronide (a metabolite only produced by the anaerobic gut microbiome with bacteriostatic properties[165], and contributing to the insulin resistance[34,166]) is positively associated with the diversity of stool Unclassified *Ruminococcaceae* and *Oscillibacter*, while negatively associated with *Frisingicoccus* and *Roseburia* (**Fig. 5e**), implying that this metabolite can modulate the tolerance of certain genera within the *Clostridia* class, as previously speculated[165]. These results not only validate several previous findings, but also generate many novel associations for validations.

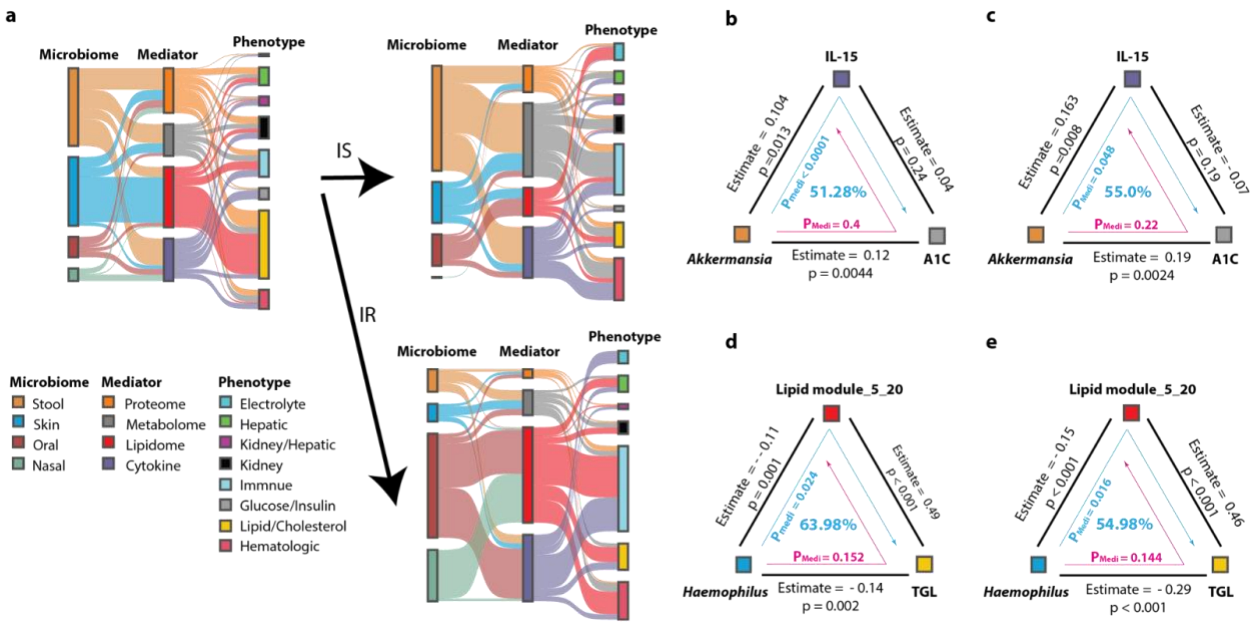


Figure 6. Causal inference decodes microbiome-driven phenotypic modifications mediated by internal molecules and cytokines

a. Microbiome and phenotype linkage analysis mediated by internal molecules and cytokines. Each column's color represents the body site of the microbiome as modulator (left), the type of analytes as mediator (middle), and the class of phenotypes as consequence (right). The heights of each column represent the number of detected associations. The left upper panel displays mediation linkage for the entire dataset, whereas the right panel displays the same type of mediation linkage for only insulin-sensitive (IS) and insulin-resistant (IR) participants, respectively.

Relative abundance of *Akkermansia muciniphila* from stool microbiome causally contributes to the blood A1C level via plasma IL-15 among **b)** all individuals (P_{medi} < 0.0001, 51.28% mediation effect) and **c)** Insulin Sensitive individuals (P_{medi} =0.048, 55.0% mediation effect).

Relative abundance of Haemophilus parainfluenzae from skin microbiome causally contribute to the plasma level of triglycerides (TGL) among **d)** all individuals (P_{medi} < 0.024, 63.98% mediation effect) and **e)** Insulin Sensitive individuals (P_{medi} =0.016, 54.98% mediation effect).

The causal effect of the microbiome on clinical markers mediated by internal multiomics

To further investigate the possible causal linkages between the correlation networks described above, we conducted mediation analyses[167,168], based on correlations between the microbiome and other omics. This analysis measures the amount to which a variable (set as internal omics and cytokines/chemokines) contributes to the transmission of change from a cause (set as the microbiome) to its effect (set as the clinical phenotypes, **Extended Data Table S13**) and provides a confidence level based on the comparison to the analysis result when the mediator and effect are switched[1,167,169]. Three hundred thirty significant mediation effects were identified in our dataset and involved many microbial taxa from all four body sites (**Fig. 6a**). 207 and 164 mediation effects were found in IS and IR participants, respectively. Interestingly, the internal omics showed that microbiome-mediated glucose regulatory functions are completely absent in IR individuals. In addition, the lipidome-mediated effect between skin microbiome and host plasma lipid/cholesterol was not evident in IR individuals, suggesting a general microbiome-metabolic dysregulation. On the other hand, the oral microbiome mediated a large proportion of immune profiles (i.e.,

neutrophil counts, white blood cell counts) via the modulation of lipidome and cytokines in IR (23.17%) relative to IS (3.86%), largely through a negative relationship among major oral core microbiome such as *Veillonella* (**Fig. 6a, Extended Data Table S14**). Such causal relationships are consistent with several previous observations indicating that diabetes-related oral dysbiosis may result from a combination of the loss of commensal oral taxa[38,41,170] and an increase in the pathogenicity of resident oral bacteria during impaired glucose metabolism, as demonstrated by a human observational study[171,172] and animal research[39,40]. These observations also strengthen the hypothesis[173,174] that oral-originated inflammation might be a strong mediator in complications of systemic insulin resistance.

In addition, this analysis revealed a causal effect of *Akkermansia* on mildly elevating host A1C through cytokines such as IL-15 (**Fig. 6b**, **Extended Data Table S14**). Despite the widely accepted beneficial correlation between high *Akkermansia* and host glucose homeostasis[175-178], we believe the mild effect observed here implies the spontaneous increase of *Akkermansia* (i.e. without supplementation) might be a sign of the over-consumption of mucins that has been noted in several autoimmune-related studies[109,179-182]. Further, the proportion of *Akkermansia* mediated effects was also significant in IS individuals but not in IR individuals (**Fig. 6c**), indicating such an interaction might still be a signature that the glucose metabolism is under control.

Interestingly, skin *Haemophilus* was closely connected to host triglycerides (TAG) through a group of correlated lipids (*Module 5_20*, **Extended Data Table 10**) (**Fig. 6d**), and was only significant in IS individuals. (**Fig. 6e**) This mediation linkage might go beyond the common observation of commensal *Haemophilus* outcompeting pathogenic strains, as one of the major characteristics for *Haemophilus* is their ability to synthesize heme[183], which is found in the skin[184,185] and is beneficial to host TAG metabolism[186,187] and insulin homeostasis[188-191]. Therefore, this mediation linkage might provide additional evidence for the observed link between *H. parainfluenzae* and lower cardio-metabolic risk, as observed in large consortium datasets such as the META genomics of the Human Intestinal Tract (MetaHIT)[192] and Personalised REsponses to DIetary Composition Trial. (PREDICT)[54]. Our data provides a logical model for the benefits of a microbe-lipid interaction in healthy hosts. Overall, these protein, metabolic, lipid, cytokine-microbial relationships provide valuable information concerning microbes and their potential benefits.

Discussion

This study provided the first systematic overview of multi-site microbiome ecology among the iHMP prediabetic cohort. With date-matched microbiome and host omics data, we not only expand our knowledge of the ecology, stability, and individuality of microbiome from various body sites, but also provide mechanism-generating hypotheses on host-microbiome interactions in the context of T2D risks. We demonstrate a few novel observations: 1) There is a "core" microbiome that is highly stable (in terms of presence) over time and opportunistic microbiome that is highly variable and more likely interacting with the immune system; these are unique to each body sites with some consistent correlations of microbes between sites. 2) Correlations between microbiomes across body sites and extensive interactions with host factors indicate systemic coordination and interactions throughout the human body. 3) Highly individualized microbiomes are differentially associated with distinct environmental factors (i.e., season, diet, chemical and biological exposome), and presumably these factors shape the microbiome at various body sites differently. However, these effects do not override the variance contributed by individuals, suggesting that the host is still the largest confounding factor for the variation observed in the microbiome. 4) Finally, individuals with IR have a less stable microbiome with more diverse microbiome members, as well as significantly altered host-microbiome interactions.

The stool microbiome is the most distinctly individualized microbiome, followed by the oral microbiome; the skin and nasal microbiomes are the least individualized. We propose that individualized diet and host factors drive this observation for the microbiome from the digestive system. Skin and nasal exposure, although likely also personal[91,110] may be less individualized, suggesting that diet and host factors may have

a lower effect on microbiomes at these body sites. Interestingly, the family score was greater for skin, oral and nasal than stool microbiomes, suggesting these are heavily influenced by environment.

Dense longitudinal sampling also provided an improved quantification of the microbiome stability and degree of individuality at different body sites. We found that high stability is correlated with the individuality of the microbiome, strongly suggesting the host participated in the adoption of commensal bacteria. We also found that the individuality of the microbiome is taxonomic specific, meaning the individualized stability might be a consequence of the dominant taxa one carries. For example, Bacteroidetes exhibited a higher DMI in stool compared with members from other phyla. It has previously been shown that genera *Bacteroides (*Bacteroidetes*),* *Prevotella (*Bacteroidetes*),* and Unclassified *Ruminococcaceae (*Firmicutes*)* in the stool microbiome have dominant patterns within individuals and mutually exclusive patterns between individuals[53,78,79]. Our results imply that individuals dominantly carrying *Bacteroides* or *Prevotella* may maintain larger inter-individual variabilities, possibly including microbiome-related metabolic status. More importantly, by examining the correlational relationship between body sites within the same individual, we further demonstrated that such micro-biotypes are strongly related across distant human body regions (**Extended Data Table 7**). For example, individuals carrying high *Bacteroides* in the stool microbiome also carry more *Cutibacterium* on the skin and *Prevotella* in the oral microbiome. Combined with our result on the correlations between individuality and stability, we suggest that the colonization and adaptation of bacteria with multiple body sites is not completely random, and that factors from the host actively participate in such processes even in adults with an established microbiome.

One important system for driving microbial individuality and stability is the host immune system. The immune system is well known to interact with microbes at multiple body sites[128,193,194], and this interaction modulates both the microbes that are present, as well as the functional benefits from the individualized microbiome (*e.g.*, beneficial microbial signals such as those that maintains barrier functions[47,195]). The interaction of inflammatory cytokines with the microbiome demonstrated that low prevalent microbes, such as members from the stool *Proteobacteria*, are likely reduced during host inflammatory events. We also revealed a systematic relationship between cytokines and the diversity within bacteria genera at each body site, and that a subset of the skin microbiome is positively correlated, and the stool microbiome negatively correlated with the same group of cytokines. This result strongly suggests the stool and skin microbiome are highly influenced by cytokines of the immune system.

We also identified a surprisingly large number (2180) of interactions between host plasma analytes and microbiomes from different body sites **(Extended Data Table 12)**. Some of the relationships, such as alcohol metabolites and the gut microbiome, have been previously documented[196,197]. Stool *Bacteroides*, *Prevotella* and Unclassified *Ruminococcaceae* are separately located in lipidomics, proteomics and metabolomics zones of the correlation network, indicating that these host factors strongly interact with the different types of microbes. We also find many correlations across body sites with host factors: for example, the correlated *Bacteroides* in the stool and *Cutibacterium* in the skin are both closely related to lipids metabolism[198-202]. Our multi-omics analyses raise the possibility that the systematic host regulatory factors may be causal for such relationships, strengthening the concept of microbial cross-body site communication and regulation, as has been suggested for the gut-brain axis[203,204] and gut-lung axis[205,206]. Since many of the metabolites, lipids and proteins are signaling molecules (*e.g.*, cytokine/chemokines, hormones, peptides), these molecules may play important roles in organismal communication across the entire host-microbiome ecosystem. Importantly, this interaction presumably occurs at the individual specific level.

Intriguingly, we found the relative abundance of *Klebsiella* on skin was positively associated with metabolites of alcohol. This indicates that alcohol intake may change the host into a *Klebsiella*-tolerant environment, resulting in the adaptation and expansion of pathogenic *Klebsiella* as previously described[207] . This correlative relationship supports previous findings in alcohol-associated pneumonia, where alcohol

consumption and increased susceptibility to *Klebsiella* in the lungs may be a result of either intestinal *Klebsiella*-specific T-cell sequestration[207,208] or alcohol-related impairment of tryptophan catabolite production/processing in the gut microbiome, which restricts pulmonary immune cell trafficking[209,210].

Our mediation analysis also found out that the dysbiosis of the oral microbiome may be a contributor to the low-grade inflammation. Consistent with a previous study[211], we observed a higher interdependency between oral microbiome members. The oral microbiome has structural complexity[84]; perhaps this leads to high constraints and interdependency. Oral dysbiosis has been implicated in several non-oral cavity diseases that involve inflammation[212,213], such as atherosclerosis[214], Alzheimer's Disease[215,216], and COVID-19 infections[217]. In addition, oral dysbiosis associated with diabetes has been previously reported. Subjects with T2D had increased oral bacteria such as *T. denticola* and *P. nigrescens* [218], which were also identified as periodontitis related species[219]. Dysbiosis in the oral microbiome can induce local and systemic inflammation, and this can provoke hyperglycemia. Indeed, periodontitis treatment has been shown to impact glycemic conditions[220].

The microbiome-host relationship is severely impacted in individuals with IR, with decreased microbiome stability. The stool and skin microbiomes of IR individuals exhibit not only different members, overall diversity, and core microbiomes relative to those who are IS, but also a systematic shift in the rank prevalence curve. This taxonomy-independent shift clearly shows that the prevalence of the whole community has changed, as opposed to just one or two microbiome members, strongly suggesting a dysbiotic microbiome. This dysbiosis may result in an altered microbiome-host interaction, as evidenced not only by the reduced number of host-microbe correlations detected and the absence of relationships between oral and skin microbiome stability in IR individuals, but also by the increased proinflammatory microbiome in the upper respiratory tract (oral and nasal microbiome) in IR individuals, as determined by our causal inference results.

Our study has several limitations. We used the 16S rRNA sequencing, which is useful for three body sites that have the most human DNA (oral, nasal, skin) in bulk preparations. However, certain variable regions have low coverage of specific genera and may be missing in our analyses. Such genera-specific biases are well-documented for the stool[74] and skin[221] microbiomes. We also note that participants in our study may be limited in their geographical location and lifestyle[222,223], which in turn limits our ability to generalize our conclusions to a more diverse and complex population. Thus, the field will benefit from additional datasets that focus on similar topics as recently proposed[224], including validation of the novel findings here. Finally, many of our analyses define associations or statistical mediation effects, and do not directly demonstrate causation which can be pursued in follow-up studies.

In spite of these limitations our studies provide a number of novel observations concerning the individuality and stability of the microbiome across multiple body sites during health and disease and in individuals with different IR/IS status. These observations have important implications in modulation human molecular health using personalized prebiotics and probiotics. Our data also provide a valuable and unique resource for the general scientific community.

Methods

Participant recruitment

Participants were recruited as part of the National Institutes of Health (NIH) integrated Human Microbiome Project. Participants in this study were recruited under the protocol IRB-23602 at Stanford University under the integrative personal omics profiling (iPOP) study. Individuals who sign up for the study are either at risk for developing type 2 diabetes or are voluntarily interested in diabetes-related research. Participants were excluded if they had hypertriglyceridemia $> 4.0\ \mathrm{mg\ ml^{-1}}$, uncontrolled hypertension, uncontrolled psychiatric disease or had previously undergone bariatric surgery. Other exclusion criteria included pregnant or lactating women, individuals with an eating disorder (*i.e.*, binge eating disorder, anorexia

nervosa, or bulimia nervosa), or alcohol use disorder. Participants who do not achieve five consecutive samples of at least one body site (*i.e.*, stool, skin, nasal, and oral) were not included in this manuscript. Therefore, data from 86 participants were used for complete analysis in this manuscript.

Microbiome sample collection and sequencing

Stool samples were self-collected by participants and others were by study coordinators following iPOP study SOPs, as adapted from iHMP standard operating procedure corresponding sections (HMP_MOP_Version12_0_072910)[72]. Briefly, retroauricular areas were rubbed with pre-moistened swabs under pressure for skin sampling, anterior nares for nasal sampling, and rear of the oropharynx for oral sampling. Samples are stored at -80 C immediately after arrival. Stool and nasal samples were further processed and sequenced in-house at the Jackson Laboratory for Genomic Medicine (JAX-GM, Farmington, CT, USA) and detailed methods are described previously[72], while oral and skin samples were sent to uBiome (uBiome, San Francisco, CA, USA) for further processing.

After 30 minutes of beads-beating lysis, skin and oral samples were processed using a silica-guanidinium thiocyanate-based nucleic acid isolation protocol[225-227] on a liquid-handling robot. The 16S rRNA variable region V4 was amplified by PCR using the primer 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACHVGGGTWTC TAAT-3')[228]. The DNA from each sample was barcoded and combined to create a sequencing library. The sequencing library was then purified using columns and microfluidic DNA fractionation[229] to reduce unwanted DNA fragments. Bio-Rad MyiQ was used to quantify the DNA concentration of the library using the Kapa iCycler qPCR kit (Bio-Rad Laboratories, Hercules, CA, USA). Sequencing was performed on the Illumina NextSeq 500 Platform (Illumina, San Diego, CA, USA) via 2 * 150 bp paired-end sequencing protocol[230].

Raw sequencing data from the stool samples and nasal samples are acquired from our previous publication[72]. Briefly, 16S rRNA gene from V1~V3 hyper-variable region was amplified with primer pair of 27F (5′-AGAGTTTGATCCTGGCTCAG-3′) and 534R (5′- ATTACCGCGGCTGCTGG-3′) and being barcoded and sequenced on the Illumina MiSeq sequencing platform through a V3 2 × 300 sequencing protocol. The same cutoff used in skin and oral sequencing data was applied to stool and nasal sequencing data in demultiplexing. After demultiplexing, reads with Q-scores less than 35 and ambiguous bases (Ns) are trimmed for additional analysis.

Microbiome data processing

Sequenced samples were demultiplexed and saved as FASTQ files using BCL2FASTQ software (Version 2.20, Illumina, CA, USA). Primers were also removed following filtering. Mismatched sequence reads in barcode or more than one mismatch in primer were removed for downstream analysis. Reads with Q-score < 30 were excluded from the analysis. The forward reads were selected for further processing due to the low overlap between the forward and reverse reads according to FLASH (V 1.2.11).

Sequencing data of the microbiome from four body sites are combined for downstream data processing. To process sequence data, raw sequence reads were processed using the R package, DADA2 (version 1.16)[231]. Sequences were further filtered by removing reads with ambiguous bases (maxN=0), and more than two expected errors (EE) (maxEE=2). Filtered data was then submitted to inter-sample composition based on the learned error rate. An amplicon sequence variant (ASV) table was constructed, and chimeras were removed using the *consensus method* from the DADA2 workflow. Reads passing all the above filters were aligned using 100% identity over 100% of the length against a trained database of target 16S rRNA gene sequences and taxonomic annotations derived from Version 18 of The Ribosomal Database Project (RDP) Taxonomy release (Aug 14, 2020)[72]. The relative abundance of each ASV was determined by dividing the count linked to those taxa by the total number of filtered reads. The samples with depths below 1000 reads were removed due to lack of sufficient sequencing depths[232] in consistent with the previous iHMP reports[22,72]. Local Outlier Factor (LOF) of each point was calculated on a depth-richness (observed ASV)

plot. Samples with a LOF greater than 3 were removed (n=7) due to an abnormal richness-sequencing depth relationship. Average sample sequencing depth after quality control was 23554 for stool microbiome, 74515 for skin microbiome, 132912 for oral microbiome, and 24899 for nasal microbiome.

The collection of Lipidomics data

Lipid extraction and data generation were conducted as previously described[233,234]. Briefly, complex lipids were extracted from 40uL of EDTA-plasma using a mixture of methyl tertiary-butyl ether, methanol, and water following biphasic separation. Lipids were then analyzed with the Lipidyzer platform consisting of a DMS device (SelexION Technology, Framingham, MA, USA) and a QTRAP 5500 (Sciex). Lipids were quantified using a mixture of 58 labeled internal standards provided with the platform (cat# 5040156, Sciex, Redwood City, CA, USA), and lipid abundances were reported in nmol/g.

To reduce the significantly high collinearity of the lipidomic data, we designed a customized clustering method. Specifically, the lipidomics data was binned into 6 clusters using the Fuzzy c-means clustering (R package "Mfuzz" (v3.15)). Then for the lipids in each cluster, the correlation was computed, and lipids with high correlative relationships at Spearman correlation > 0.8 and BH-adjusted *p*-values < 0.05 were clustered into the same module. Then the community analysis (fastgreedy.community function from R package "igraph"(v1.3.5)) was utilized to detect the modules. For the lipids that are not assigned to any of the modules, we used their original lipid species annotation for the downstream analysis. (**Extended Data Table 10**)

The collection of Metabolomics data

Untargeted metabolic profiling was performed using a broad-spectrum LC-MS platform using a combination of reverse-phase liquid chromatography (RPLC) and hydrophilic interaction liquid chromatography (HILIC) separations and high-resolution MS[72,235]. Briefly, plasma metabolites were extracted following solvent precipitation using a mixture of ice-cold acetone, acetonitrile, and methanol (1:1:1, v/v). Hydrophilic metabolites were separated on a ZIC-HILIC (2.1 × 100 mm, 3.5 µm, 200 Å; Merck Millipore) while hydrophobic metabolites were separated on a Zorbax SBaq columns (2.1 × 50 mm, 1.7 µm, 100 Å; Agilent Technologies). Data were acquired on a Thermo Q Exactive plus mass spectrometer for HILIC and a Thermo Q Exactive mass spectrometer for RPLC. Raw data were processed using Progenesis QI (v2.3, Nonlinear Dynamics, Waters) and metabolites were formally identified by matching fragmentation spectra and retention time to analytical-grade standards or matching experimental MS/MS to fragmentation spectra in publicly available databases. A total of 726 annotated metabolites were retained for downstream analysis.

The collection of Proteomics data

Plasma proteins were profiled by LC-MS using SWATH acquisition on a TripleTOF 6600 system (Sciex) as previously described[72]. In each injection, 8-µg of tryptic peptides generated from depleted plasma were loaded on a ChromXP C18 column (0.3 × 150 mm, 3 µm, 120 Å, Sciex). Peptides were separated with a 43-min gradient from 4-32% B. Variable Q1 window SWATH Acquisition methods (100 windows) were built in high sensitivity MS/MS mode with Analyst TF Software (v1.7). PyProphet (v2.0.1)[236] and TRIC[237] were used to score and align peak groups, respectively (1% FDR at peptide level; 10% FDR at protein level). Protein abundances were computed as the sum of the three most abundant peptides.

Luminex Multiplex Assays for targeted cytokine, chemokine, and growth factors

Cytokine assays were performed using well-characterized protocols from the Stanford Human Immune Monitoring Center (HIMC). Circulating cytokines, chemokines, and growth factors were assessed in EDTA-plasma using a mixture of Human 62-plex Luminex conjugated antibodies. (Affymetrix, Santa Clara, California) Raw data from the assay were normalized to the median fluorescence intensity (MFI) value, and data were transformed by variance stabilizing transformation (VST) to eliminate the batch effect. As

we previously described[132], measurements with background noise (CHEX) that are outside 5 standard deviations from the mean (mean ± 5 × SD) are removed.

The collection of Exposome and associated environmental data

Exposome and relevant environmental data were collected as previously described[91,238]. Briefly, the chemical exposome was collected using RTI MicroPEM V3.2 personal exposure monitor (RTI international, Research Triangle Park, NC, USA) from two participants. The MicroPEM is an active air sampling device that pumps air at a flow rate of 0.5 L/min. The device was modified to include a customized cartridge filled with 200 mg zeolite adsorbent beads (Sigma 2-0304, Sigma-Aldrich Corp., St. Louis, MO USA) at the end of the air flow to collect hydrophobic and hydrophilic compounds. Each sampling period lasts around 5 days. After each session, the cartridge was removed and stored at -80 ℃ degrees till further processing. To perform chemical extraction, zeolite beads were resuspended in a clean Eppendorf LoBind tube with 1 mL methanol (Mass Spec grade) and incubated at room temperature for 20 mins. The samples were then centrifuged at 22,000 g for 20 mins at room temperature and then analyzed by a Waters UPLC-coupled Exactive Orbitrap Mass Spectrometer (Thermo, Waltham, MA, USA). Finally, 158 exposome chemicals were collected. Environmental data were collected from the following sources: 1. Temperature, humidity, and sampling flow rate recorded by MicroPEM; 2. GPS coordinates of the wearer; 3. Meteorological and demographic data from public data deposits, including Climate Data Online (CDO), US Census Bureau, and local weather stations. Finally, 10 environmental feature data were collected.

The collection of dietary data

A total of 25 food items were included in a questionnaire that was completed voluntarily by participants during their routine visit in the study, using a diet questionnaire hosted on https://www.projectredcap.org/ from our previous report[132]. The frequency of consuming each food item was scored in downstream analysis. For breads, biscuits, cakes, pies and pastries, we score the frequency from 0~4, with the associated frequency per day: 0) less than 1, 1)1 per day, 2)2~3 per day, 3)4~5 per day, 4)6 or more. For other foods we ask the participants to state the frequency from 6+ times per day to less than once per month.

Insulin sensitivity measurement

A subset of eligible, consented subjects (N=58) participated in a one-time assessment of steady-state plasma glucose (SSPG) test. Insulin-mediated glucose uptake was measured as an indicator of insulin sensitivity as previously described[72]. Briefly, fasted individuals (12-h overnight) were infused with 0.27 ug/m^2 min of octreotide, 25m U/m^2 min of insulin and 240 mg/m^2 min of glucose for 3 hours during their visit to Stanford CTRU. Blood was collected during the last 30 minutes of the infusion every 10 minutes, totaling 4 blood draws, to measure plasma glucose and insulin. The mean of the 4 SSPG and insulin concentrations were used. Individuals with SSPG < 150 mg/dl or SSPG ≥ 150 mg/dl were assigned to the insulin sensitive (IS) (n=28) or insulin resistant (IR) group (n =30) respectively. Individuals with no measurements due to personal or medical reasons were assigned to the unknown group (n= 28).

Clinical Lab test

Clinical lab tests were performed at the Stanford Clinical Lab following its guideline of blood and urine collection and submission (https://stanfordlab.com/test-directory.html). The test includes a metabolic panel, complete blood count panel, glucose, HbA1C, insulin measurements, hsCRP, IgM, lipid panel, kidney panel, liver panel. Detailed measurements are provided as a supplementary table. (**Extended Data Table 13**)

UMAP for Microbiome distribution

Uniform Manifold Approximation and Projection was calculated using the R package "Seurat (V 4.0)[239]". We first normalized the count data to relative abundance and scaled the data into 1 million reads per sample as recommended. The 2000 highly variable features were selected and stored in a Seurat object. We then generated a distance matrix using the R package "Vegan (version 2.6-2)[240]" under Bray Curtis distances.

The first 10 dimensions (1094 generated) in the distance matrix were applied to calculate neighboring relationships and projected by calling Python UMAP via reticulate with default settings. Result was displayed on the first two dimensions of the UMAP.

Intraclass correlation

Intraclass Correlation Coefficient (ICC) was calculated from Linear Mixed Models, in which we modeled random intercepts but a fixed slope, allowing different personal levels between subjects[72]. We first linearly transformed each analyte (when applicable) and standardized the total variation to 1 before applying "lmer" function from R package "lme4 (V1.1-30)[241]", with the formula as:

$$Exp \sim 1 + Days + A1C + SSPG + FPG + (1|SubjectID)$$

Where *Exp* was the linearly transformed and standardized values of each analyte, Days was the length of time individuals participated in the study, *A1C* was the HbA1C value (%), *SSPG* was the steady state plasma glucose value (mg/dl), *FPG* was the fasting plasma glucose value (mg/dl), *SubjectID* was the subject ID associated with each participant.

We then used ICC as the proportion of total variation explained by subject structure in the cohort by $V_{subjectID}/V_{total}$, in which V was the variance from the corresponding component extracted by "VarCorr" and V_{total} was 1.

Permutation test

The Bray Curtis (BC) distance was used to quantify the degree of similarity between two microbiome samples. Sample ASV was utilized as the unit for calculating dissimilarity for the complete microbiome sample or for specific taxa. Similarity metrics were calculated pair wisely from the same subject (intra-individual) and between different subjects (inter-individual). A permutation test was used to estimate the null distribution while adjusting for the varying sample sizes of each participant. Briefly, for each microbial genus, all sample labels were randomly permuted, and the BC distance was then computed pairwise. This process was repeated for 1,000 times, and the null distribution was achieved to evaluate the p-values for each inter-individual BC distance.

Degree of microbial individuality (DMI)

Degree of Microbial Individuality was measured as the mathematical difference between a given genus regarding their populational median of the inter-individual Bray-Curtis distance (BC) and median of the intra-individual BC distance. We first summarized the between-sample distance of genera whose longitudinal prevalence was > 10% at a given body site. Distance of sample pairs were then allocated into inter-individual or intra-individual groups. Total genera left for analysis per region included: stool: 172; skin: 218; oral: 101and nasal: 70. For a given genus, *i*, at one body site, the DMI was calculated using the following formula:

$$DMI_i = BC_{inter\text{-}individual} - BC_{intra\text{-}individual}$$

Family score (FS)

To estimate the influence of microbiome variability from a shared living environment, we calculated the "within-family" inter-individual BC distance for each genus, from the cohabitating pair, within the same family. Genera with a longitudinal prevalence <10% were excluded. The number of genera filtered through per region were as follows: stool: 141; skin:119; oral: 41, and nasal: 33. The median inter-individual BC distance between two family members (BC intra-family) was used to calculate the family score of one given genus, *i*, using the following formula:

$$Family\ Score_i = (BC_{inter\text{-}individual} - BC_{intra\text{-}family}) / (BC_{inter\text{-}individual} - BC_{intra\text{-}individual})$$

Any result with Family Score ≥ 1 will be assigned as 1, and values ≤0 will be assigned as 0.

Classification of the microbiome genera by their longitudinal prevalence

The microbiome genera were categorized as the core microbiome, opportunistic microbiome, and middle group based on their longitudinal prevalence as described before[242-244]. Calculation of prevalence was based on the presence or absence of reads from each sample. For each sample, the relative abundance of each genus was first transformed to 1 if it was greater than 0; then, the proportion of 1 for each genus in each participant was determined as the longitudinal prevalence. Then the genera were assigned to a group based on their longitudinal prevalence: core microbiome: longitudinal prevalence > 80%; middle group: 20% ≤ longitudinal prevalence ≤ 80%; opportunistic microbiome: longitudinal prevalence < 20%.

Body site-specific longitudinal model for bray-curtis distance

To estimate the effect of body site on the change in BC distance over time, we used a linear mixed effects regression with the BC distance as the response variable implemented in the "lme4"[241] package in R. The BC distance of all pairwise samples were transformed by: *dist_trans = log10(-log10(1-dist))* such that the transformed values were more normally distributed according to the Anderson-Darling test[245]. Fixed effects included an interaction between the body site and time, and with a random intercept for each individual. Time was normalized to the days from the first sample for each individual. The model was optimized using the "nlopwrap" method[246] and nested models were compared by likelihood ratio test using the "lmtest" package[247] in R. Differences in the slopes of the body sites over time were assessed by F-test with Satterthwaite's degrees of freedom[248]. For the comparisons between body sites, the model was constructed as:

$$dist \sim -1 + diffdays*dataset + (1 \mid subject\_id)$$

For the comparison between IR and IS participants, the model was constructed as:

$$dist \sim -1 + IRIS * diffdays + (1 \mid subject\_id) + (1 + IRIS \mid subject\_id)$$

Where *dist* was the pairwise BC distance, *diffdays* was the date interval between two samples, *subject_id* was the subject ID associated with each participant, and *IRIS* was the insulin sensitivity status of each participant.

Bayesian mixed-effects model for microbial taxa and cytokine interactions

Participants with longitudinal, date-matching measurements of both the microbiome and blood-based cytokines (n=62) were included in a Bayesian negative-binomial longitudinal mixed-effects model. Each microbe was modeled as a sparse-matrix response variable, with a plasma cytokine level MFI quantity and time as fixed effects and a random intercept for each individual, following the formula:

$$Mi = Xi+Zibi+i$$

Where M_i is a vector of the genus-level microbe relative abundances for each participant i, X_i is the design matrix for the fixed effects, β. Each row of matrix X_i contains the terms (1) time (days post–study start), D_i, and (2) cytokine measurements, Y_i, from 1 to n. Z_i is the random effects design vector of 1's denoting a random intercept, b_i is a scalar for each participant, and ε_i is a zero-centered error term.

Posterior sampling was performed using four chains, 5,000 iterations per sample, and a 1,000-iteration burn-in of a No-U-Turn Sampler implemented in the "brms (Version 2.18.0)" package in R[249-252]. The iteration plots and posterior predictive distributions were visually inspected for chain convergence. Microbe genera and cytokines above the limit of detection in less than 10% of the samples were excluded from the

analysis. A microbe-cytokine association was considered significant if the 95% credible interval on the fit coefficient of the cytokine term did not include zero.

Correlation network analysis

The correlation network between the microbiome from stool, skin, oral, and nasal samples, and internal multi-omics (proteome, metabolome, and lipidome) data from plasma was constructed using the published methods[22,72] with in-house modifications. For each pair of the microbiome and internal omics data, we removed the time points that did not match for collection date (detailed in **Fig. 1b**). The subjects with less than 5 samples for a specific microbiome type were removed from this analysis for that microbiome type related correlations. For the microbiome data, the relative abundance or observed richness at the genus level were used for calculation, respectively. Only the genus that can be detected in at least 10% of all the samples was kept for subsequent analysis. To address compositionality in the microbiome data, the centered log ratio (CLR, "clr" function from R package "compositions" (Version 2.0-4)) was utilized to normalize the data. For proteome, metabolome, and lipidome, $\log_2$ transformation was utilized for transformation processing. Next, to account for the effect of repeated samplings from the same subject, the linear mixed effect model was utilized, which takes the subject ID as the random effect ("lmer" function from R package "lme4" (Version 1.1-30)). Then the Spearman correlations were calculated, and the *p*-values were adjusted using The Benjamini-Hochberg Procedure. Here, the BH adjusted *p*-values < 0.2 was used for correlation network construction. The network visualization was achieved using the R package "ggraph" (Version 2.0.5), "igraph" (Version 1.3.2), and "tidygraph" (Version 1.2.1) under the "kk" layout.

Pathway enrichment for proteins correlated with the microbiome.

Using the R package "clusterProfiler (Version 3.15)", the proteins that corresponded with the microbiome of four different body sites were utilized to enrich pathways. Briefly, the proteins were used for pathway enrichment using the Gene Ontology (GO), where significantly associated pathways are determined by Fisher's exact test. The GO terms with BH-adjusted *p*-values < 0.05 are selected for subsequent analysis. For the significantly enriched GO terms, the "simplifyEnrichment" package in R was used to calculate the similarities between each pair of GO terms, and the edges with similarities > 0.70 were retained for network construction. The community analysis (R package "igraph (Version 1.3.4)") was used to identify the correlation network's modules. Only the GO term with the minimum BH-adjusted p-value was retained to represent each module.

Mediation Analysis

Mediation analysis was performed to explore how microbiomes from stool, skin, oral, and nasal affect phenotypes via internal multi-omics data (proteome, metabolome, lipidome, and cytokine)[1,253,254]. The phenotype data were measured by clinical laboratory tests of plasma samples. The microbiome's associations with phenotype and internal omics data were first achieved, respectively, as described in the "Correlation network analysis" section. Only the significant associations (BH adjusted *p*-values < 0.05) were included in the subsequent mediation analysis. The R package "mediation" was utilized to do the mediation analysis. Finally, only the pairs with significant Average Causal Mediation Effect (ACME, *p*-values < 0.05) were documented, which illustrates the microbiome's influence on the measurement of phenotypes via internal multi-omics. To control the false discovery rate (FDR), the reverse mediation analysis was also carried out by exchange mediator with effects (microbiome affects internal omics data through phenotype), and the pairs with significant ACME in reverse mediation (*p*-values < 0.05) were also removed from the results.

Principal variance component analysis (PVCA)

To assess the variation in microbiome data based on individual and season, the principal variance component analysis (PVCA) was performed (R package "pvca (Version 3.15)") [255]. The PVCA is a combination of the principal component analysis and variance components analysis, which were previously being employed to assess batch effects in high dimensional data[255,256]. For microbiome samples in each

body site, the season was determined by subtracting the date of collection from the first day of the year (from 1-365 days). Each sample's participant ID and season were then entered into the PVCA as variables. Then, the "ggtern (Version 3.3.5)" R package was used to visualize the data.

Deconvolute the environmental effect on the microbiome

To explore how the exposome and diet data affect the microbiome from different body sites, the exposome data (chemical and environmental data) were collected and processed as previously described[238]. The diet data were collected and described separately in the above methods section. Here, we use the exposome chemical data as an example to explain how the analysis was performed. Briefly, samples that had both exposome chemical and microbiome data within a 3-day period were included for subsequent analysis ($N_{Chemical}$ = 8, $N_{Environmental}$ = 32 ($N_{Participant1}$ = 13; $N_{Participant2}$ = 19)). Microbiome data was normalized using the centered log ratio (CLR, "clr" function from R package "compositions"). Exposome data was also log_2-transformed and auto-scaled follow our previous report[257]. Both microbiome and exposome data were used to produce a principal component analysis (PCA) analysis. Principal components (PCs) from the microbiome and exposome were further analyzed. PCs with cumulative explained variation larger than 80% were included. A linear regression model was constructed using PCs of microbiome data as Y and the corresponding exposome PCs as X. R^2 was extracted and used to represent the contributions of the exposome to microbiome data. For the diet data, the same method described above was used to evaluate how the diet affects the microbiome from four body sites.

Author Contribution:

X.Zhou, X.S., M.P.S. and G.M.W. designed the study and analysis plan. X.Zhou, W.Z, M.Agnello, S.R.L, M.Avina managed the overall sample collection and performed experimental bench work. X.Zhou, W.Z, S.R.L, Y.Z, L.C., L.L., and E.J.B. conducted microbiome data acquisition and processing. D.H, S.W, conducted lipidome data acquisition and processing. C.J, Liuyiqi.J., X.Zhang, X.S, P.G. conducted exposome data acquisition and processing. X.Zhou, W.Z, P.K conducted cytokine Luminex assay data acquisition and analysis. Lihua.J., R.J. conducted proteome data acquisition and analysis. K.C. and X.S. conducted metabolomics data acquisition and analysis. J.S.J and D.J.S, X.Zhou, and X.S. designed and executed the longitudinal data linear mixed effect modeling. W.Z., M.Avina, S.R.L., A.C., managed participants' clinical visits and sample collection/inventory. X.Zhou, X.S, J.J, C.Z performed integrative omics analysis. M.P.S., and G.M.W. provided funding support and overall study guidance. X.Zhou, A.H., S.Chen designed the graphical abstract. X.Zhou., X.S., M.P.S., J.S.J., and D.J.S. wrote the manuscript under the help of all listed authors.

Acknowledgment:

We sincerely appreciate the dedication and involvement of all research participants. We also thank the administrative assistance from Mrs. Ada Yee Ki Chen and Mrs. Lisa Stainton. We deeply appreciate the insightful comments about the experiments and analysis from all other members at Snyder Lab (Stanford University) and Weinstock Lab (The Jackson Laboratory of Genomic Medicine). We thank the Stanford Immune Monitoring Core and Ms. Yael Rosenberg-Hasson for the contribution to cytokine Luminex Multiplex Assay.

This work was supported by National Institutes of Health (NIH) Common Fund Human Microbiome Project U54_DE023789-01, NIH U54_DK102556-03, NIH R01_DK110186-05, NIH S10_OD020141-01, NIH R01 AT010232-04, NIH UL1 TR001085, NIH P30_DK116074, S10_OD023452-01. We also sincerely thank the generous support from Leona M. and Harry B. Helmsley Charitable Trust under Grant No. G-2004-03820 and Innovative Medicines Accelerator (IMA) grant at Stanford University (Gates Foundation Grant).

X.Zhou. received fellowship support from the Stanford Aging and Ethnogeriatrics (SAGE) Research Center under NIH/NIA grant P30AG059307. The SAGE Center is part of the Resource Centers for Minority Aging Research (RCMAR) Program led by the National Institute on Aging (NIA) at the National Institutes of Health (NIH). Its contents are solely the responsibility of the authors and does not necessarily represent the official views of the NIA or the NIH. S.M.S.-F.R received fellowship support from the NIH K08 ES028825. A.W.B. received fellowship support from 1F32DK126287 - 01A1. D.J.S received fellowship support from NIA-K01AG070310. J.S.J is funded by the Kennedy Trust for Rheumatology Research.

Conflict of Interest:

M.P.S. is a cofounder and scientific advisor of Personalis, SensOmics, Qbio, January AI, Fodsel, Filtricine, Protos, RTHM, Iollo, Marble Therapeutics and Mirvie. He is a scientific advisor of Genapsys, Jupiter, Neuvivo, Swaza, Mitrix. A.H. is a founder and shareholder of Arxeon. Y.Z. and G.M.W. are co-founders of General Biomics. No other potential conflicts of interest relevant to this article were reported.

Data Availability

Stool and nasal microbiome data are available at https://hmpdacc.org/. Other omics data can be accessed through Stanford iPOP website at https://med.stanford.edu/ipop.html. Other data (e.g., cytokines, lipids) will be made available on this site upon acceptance of the manuscript.

Code Availability

All Software and Algorithms are mentioned in the material section and are made available on GitHub (https://github.com/xzhou7). Custom analysis scripts are hosted on the Stanford iPOP site (http://med.stanford.edu/ipop.html). Questions with code and analysis can be addressed by contacting the corresponding author.

Extended Data Figures

Extended Data Figure 1. Relative abundance of representative genera on UMAP

Representative genera from the four body sites are displayed via Uniform Manifold Approximation and Projection (UMAP), using the same system of coordinates as **Figure 1d**. Each genus' normalized and scaled relative abundance (see **Methods** for details) was color-coded and annotated.

Extended Data Fig. 2: Microbiome variance explained by the individuality, season, and residuals

The microbiome from four body sites is analyzed separately using Principal Variance Component Analysis (PVCA) for variance decomposed by individuality (by subject ID), season (the number of days between the date of collection and January 1 of each year), and residues. The explaining power (R2) from four body sites is displayed as percentage in one ternary plot for the three variables: orange: season; black: individuality; blue residues.

Extended Data Fig. 3: Microbiome variance explained by diet and exposome

a. The microbiome from four body sites is analyzed using Principal Variance Component Analysis (PVCA) for their variance associated with diet. Differences between body sites are tested overall by the Kruskal-Wallis test and pairwisely by the Wilcoxon Rank test, and BH adjusted *p*-values are displayed.

b. The microbiome from four body sites is analyzed using Principal Variance Component Analysis (PVCA) for their variance associated with exposome chemical and exposome environment. Two participants with the date-matched microbiome and exposome data are curated for the analysis. Differences between body sites are tested overall by the Kruskal-Wallis test and pair wisely by Wilcoxon Rank-test, and BH adjusted *p*-values are displayed.

Extended Data Fig.4 Diversity and evenness shift between IS and IR individuals

a. Sample wise centroid shift between insulin sensitive (IS) and insulin resistant (IR) individuals. Four pairs of centroids are annotated for their sample origin: stool, oral, skin, or nasal. The blue dots represent the centroid of samples in the IR group, and red dots represent the centroid of samples in the IS group. Evenness was measured by Pielou's evenness index and Richness was measured by abundance-based coverage estimator (ACE).
b. Boxplot of ACE richness between individuals in IS and IR. The longitudinal mean of species richness (ACE) for each participant was calculated and grouped based on this participants' insulin sensitivity status. Difference between the two groups was tested by a two-sample Student-t test. Significant values are marked as: * p-value<0.05; ** p-value < 0.01.
c. Boxplot of Pielou's evenness between individuals in IS and IR. The longitudinal mean of Pielou's evenness for each participant was calculated and grouped based on this participants' insulin sensitivity status. Difference between the two groups was tested by a two-sample Student-t test. Significant values are marked as: * p-value<0.05; ** p-value < 0.01.

Extended Data Fig.5 Prevalence by relative abundance plot
Each dot represents a genus from a single participant at a given body site. Prevalence of each genus from the microbiome of a given body site was calculated and displayed on the *X* axis. Mathematical mean value of the relative abundance was calculated for each genus on a given body site too and displayed on *Y*-axis. Horizontal dash line marks the mean relative abundance of 0.01%.

Extended Data Fig.6 Relationship between number of core microbiome, steady-state plasma glucose and body mass index
The longitudinal core microbiome of an individual was designated as any genus whose prevalence was greater than 80 percent. The number of genera that can be assigned as core microbiome from each individual is compared with this individual's **a)** steady-state plasma glucose (SSPG) measurement. **b)** body mass index (BMI). Spearman correlation coefficient was computed for each correlation and significant value are marked as * p-value < 0.05; ** p-value < 0.01.

Extended Data Fig.7 Core microbiome in IS and IR individuals
a. The genera of each participant were categorized, based on their intra-individual prevalence, as core microbiome (longitudinal prevalence > 80%), opportunistic (longitudinal prevalence < 20%) or middle (80% ⩾ longitudinal prevalence ⩾ 20%). The number of genera in each participant's core microbiome was then compared based on the participant's insulin sensitivity status.
b. Rank prevalence curve of the microbiome on each body site. The rank prevalence was calculated in the same way that was described in figure 1g. Participants were first grouped by their insulin sensitivity status, and then the mean prevalence was computed for each group.

Extended Data Fig. 8: Effect size of taxa that are significantly differ in relative abundance between IR and IS individuals
The mean value of the relative abundance for all genera from each participant was calculated, and then a linear discriminant analysis (LDA) effect size (LEfSe)[258] was performed to test if certain taxa were significantly different between IR and IS group. The effect sizes of the significantly different taxa were displayed according to the body sites (*X*-axis) and the effect size (*Y*-axis).

Extended Data Fig.9 Strain replacement rate in IS and IR individuals
a. The sample before and after a given genus' relative abundance are below the limit of detection (relative abundance = 0) in each participant was paired and the Jaccard Similarity (1- Jaccard Distance) was calculated for these pairs. The mean value of all pairs from IS or IR were compared, and Mann Whitney Wilcoxon Test was performed to determine if two groups are significantly different.

Extended Data Fig.10 Time related stability correlation between body sites in IS and IR group

For a given participant, the time related stability was defined as the correlation coefficient between the Bray Curtis (BC) distance and the difference between collection dates of all paired samples collected from the same body site. Each dot in this plot represents a participant, and the correlation coefficient associated with the microbiome from different body sites are compared and the Pearson correlation coefficient was computed for individuals in **a)** IS group **b)** IR group.

Extended Data Fig. 11: Beta coefficient comparison between cytokines and genera from different body sites

Correlation coefficients between the plasma cytokines levels and bacteria genera relative abundance were computed by an in-house mixed effects model (see **Methods** for detail). The absolute value of the correlation coefficient was displayed by the heatmap in order of a hierarchical clustering.

Extended Data Fig.12 Phyla composition of core, middle and opportunistic genera of microbiome

The prevalence of genera for each participant was summarized, and the population mean of the genera prevalence was computed accordingly. The genera with mean population prevalence > 80% was assigned as the core microbiome (Core). The genera with mean prevalence < 20% are assigned as the opportunistic microbiome (Oppr). Genera do not belong to core microbiome nor opportunistic microbiome (80% ⩾ longitudinal prevalence ⩾ 20%) was assigned as or middle group (Middle). The phyla composition of core, middle and opportunistic microbiome was then computed, and color coded as follow: 1) yellow, Actinobacteria; 2) blue, Bacteroidetes, 3) dark yellow, Firmicutes; 4) light green, Proteobacteria; 5) gray, genera that do not belong to the previously mentioned phyla.

Extended Data Fig.13 Correlation between genera of stool Proteobacteria and plasma cytokines grouped by prevalence

The Proteobacteria-belonging genera were subset, and their correlation coefficients with cytokines were compared based on their longitudinal prevalence. (Oppor: opportunistic microbiome, Middle: middle group microbiome, see Methods for details)

Difference of the mean value between two groups was tested with two-sample Student-t tests, and the test statistics are: t = -1.81, p-value = 0.09.

Extended Data Fig.14 The correlation between body mass index and plasma leptin and granulocyte-macrophage colony-stimulating factor

Mean value of plasma leptin and granulocyte-macrophage colony-stimulating factor (GM-CSF) was calculated for each individual. The distribution of the body mass index (BMI) and cytokine was summarized as histogram on X-axis or Y-axis, respectively. The Spearman correlation between the two cytokines and the BMI of participants was computed. Statistics for the correlation between

1. BMI - GM CSF: S = 28774.64, p-value = 9.42×10^{-15}, rho = 0.72.
2. BMI - Leptin: S = 31,102, p-value=1.41×10^{-13}, rho = 0.70.

Extended Data Fig. 15: High collinearity of omics is specific to the lipidome data

The Spearman correlation between elements belonging to metabolome, lipidome, and proteome was calculated. The histogram of the correlation coefficient was displayed.

Extended Data Fig.16 Pathway enrichment analysis of proteomics and microbiome interactions

The correlation between the proteome and microbiome was calculated, and the names of the proteins that are significantly correlated with microbiome are curated to the GO database for their functional pathway annotation (see **Methods**). The pathway enrichment result was then summarized and visualized based on their GO term similarity (see **Methods**). A pie chart indicates, for each pathway, the number of proteins that were correlated by the microbiome of each body site. (Yellow: stool microbiome; blue: skin microbiome, red: oral microbiome, green: nasal microbiome).

Extended Data Fig.17 Different interactome of the stool microbiome and internal plasma analytes
The correlation of stool microbiome (dark yellow) and proteome (orange), metabolome (dark blue) and lipidome (red) was separately calculated (see **Methods** for detail) for the insulin sensitive (IS) and insulin resistant (IR) population. the number of nodes and edges in IS: proteome: 223 nodes and 319 edges, metabolome: 174 nodes and 386 edges, lipidome: 171 nodes and 335 edges; IR: proteome: 80 nodes and 73 edges, metabolome: 128 nodes and 187 edges, lipidome: 94 nodes and 109 edges.

Extended Data Fig.18 Correlation network of microbiome richness and multiple internal analytes on four body sites
Network showing correlations between the observed richness (number of ASV observed) of microbiome genera at four body locations (Dark yellow filled circle: Stool (71 nodes and 1017 edges); Blue filled circle: Skin (38 nodes and 186 edges); Dark red filled circle: Oral (26 nodes and 424 edges); Green filled circle: Nasal (38 nodes and 553 edges)) and plasma analytes (Dark blue filled hexagon: Metabolome (157 nodes and 557 edges); orange filled square: Proteome (230 nodes and 790 edges); Red filled triangle: lipidome (200 nodes and 833 edges)). The confidence (BH-adjusted *p*-value) of the correlation is represented by lines between nodes. (See **Methods**)

Reference

1 Chen, L. *et al.* The long-term genetic stability and individual specificity of the human gut microbiome. *Cell* **184**, 2302-2315 e2312 (2021). https://doi.org:10.1016/j.cell.2021.03.024

2 Mehta, R. S. *et al.* Stability of the human faecal microbiome in a cohort of adult men. *Nat Microbiol* **3**, 347-355 (2018). https://doi.org:10.1038/s41564-017-0096-0

3 Faith, J. J. *et al.* The long-term stability of the human gut microbiota. *Science* **341**, 1237439 (2013). https://doi.org:10.1126/science.1237439

4 Byrd, A. L. *et al.* Gut microbiome stability and dynamics in healthy donors and patients with non-gastrointestinal cancers. *J Exp Med* **218** (2021). https://doi.org:10.1084/jem.20200606

5 Oh, J. *et al.* Temporal Stability of the Human Skin Microbiome. *Cell* **165**, 854-866 (2016). https://doi.org:10.1016/j.cell.2016.04.008

6 Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K. & Knight, R. Diversity, stability and resilience of the human gut microbiota. *Nature* **489**, 220-230 (2012). https://doi.org:10.1038/nature11550

7 Frost, F. *et al.* Long-term instability of the intestinal microbiome is associated with metabolic liver disease, low microbiota diversity, diabetes mellitus and impaired exocrine pancreatic function. *Gut* **70**, 522-530 (2021). https://doi.org:10.1136/gutjnl-2020-322753

8 Sommer, F., Anderson, J. M., Bharti, R., Raes, J. & Rosenstiel, P. The resilience of the intestinal microbiota influences health and disease. *Nat Rev Microbiol* **15**, 630-638 (2017). https://doi.org:10.1038/nrmicro.2017.58

9 Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59-65 (2010). https://doi.org:10.1038/nature08821

10 Gilbert, J. A. *et al.* Current understanding of the human microbiome. *Nat Med* **24**, 392-400 (2018). https://doi.org:10.1038/nm.4517

11 Ursell, L. K., Metcalf, J. L., Parfrey, L. W. & Knight, R. Defining the human microbiome. *Nutr Rev* **70 Suppl 1**, S38-44 (2012). https://doi.org:10.1111/j.1753-4887.2012.00493.x

12 Roswall, J. *et al.* Developmental trajectory of the healthy human gut microbiota during the first 5 years of life. *Cell Host Microbe* **29**, 765-776 e763 (2021). https://doi.org:10.1016/j.chom.2021.02.021
13 Zhou, Y. *et al.* Biogeography of the ecosystems of the healthy human body. *Genome Biol* **14**, R1 (2013). https://doi.org:10.1186/gb-2013-14-1-r1
14 Human Microbiome Project, C. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207-214 (2012). https://doi.org:10.1038/nature11234
15 Zhou, Y. *et al.* Exploration of bacterial community classes in major human habitats. *Genome Biol* **15**, R66 (2014). https://doi.org:10.1186/gb-2014-15-5-r66
16 DiGiulio, D. B. *et al.* Temporal and spatial variation of the human microbiota during pregnancy. *Proc Natl Acad Sci U S A* **112**, 11060-11065 (2015). https://doi.org:10.1073/pnas.1502875112
17 Zhernakova, A. *et al.* Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* **352**, 565-569 (2016). https://doi.org:10.1126/science.aad3369
18 Wilmanski, T. *et al.* Gut microbiome pattern reflects healthy ageing and predicts survival in humans. *Nat Metab* **3**, 274-286 (2021). https://doi.org:10.1038/s42255-021-00348-0
19 Bana, B. & Cabreiro, F. The Microbiome and Aging. *Annu Rev Genet* **53**, 239-261 (2019). https://doi.org:10.1146/annurev-genet-112618-043650
20 Zhou, X. *et al.* Exploratory studies of oral and fecal microbiome in healthy human aging. *Frontiers in Aging*, 110
21 Zhang, Y. *et al.* Discovery of bioactive microbial gene products in inflammatory bowel disease. *Nature* **606**, 754-760 (2022). https://doi.org:10.1038/s41586-022-04648-7
22 Lloyd-Price, J. *et al.* Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **569**, 655-662 (2019). https://doi.org:10.1038/s41586-019-1237-9
23 Fornelos, N. *et al.* Growth effects of N-acylethanolamines on gut bacteria reflect altered bacterial abundances in inflammatory bowel disease. *Nat Microbiol* **5**, 486-497 (2020). https://doi.org:10.1038/s41564-019-0655-7
24 Nemet, I. *et al.* A Cardiovascular Disease-Linked Gut Microbial Metabolite Acts via Adrenergic Receptors. *Cell* **180**, 862-877 e822 (2020). https://doi.org:10.1016/j.cell.2020.02.016
25 Walker, R. L. *et al.* Population study of the gut microbiome: associations with diet, lifestyle, and cardiometabolic disease. *Genome Med* **13**, 188 (2021). https://doi.org:10.1186/s13073-021-01007-5
26 Fromentin, S. *et al.* Microbiome and metabolome features of the cardiometabolic disease spectrum. *Nat Med* **28**, 303-314 (2022). https://doi.org:10.1038/s41591-022-01688-4
27 Qi, Q. *et al.* Host and gut microbial tryptophan metabolism and type 2 diabetes: an integrative analysis of host genetics, diet, gut microbiome and circulating metabolites in cohort studies. *Gut* **71**, 1095-1105 (2022). https://doi.org:10.1136/gutjnl-2021-324053
28 Wu, H. *et al.* The Gut Microbiota in Prediabetes and Diabetes: A Population-Based Cross-Sectional Study. *Cell Metab* **32**, 379-390 e373 (2020). https://doi.org:10.1016/j.cmet.2020.06.011
29 Forslund, K. *et al.* Corrigendum: Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature* **545**, 116 (2017). https://doi.org:10.1038/nature22318

30 Pedersen, H. K. *et al.* Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature* **535**, 376-381 (2016). https://doi.org:10.1038/nature18646
31 Ruuskanen, M. O. *et al.* Gut Microbiome Composition Is Predictive of Incident Type 2 Diabetes in a Population Cohort of 5,572 Finnish Adults. *Diabetes Care* **45**, 811-818 (2022). https://doi.org:10.2337/dc21-2358
32 Chen, Z. *et al.* Association of Insulin Resistance and Type 2 Diabetes With Gut Microbial Diversity: A Microbiome-Wide Analysis From Population Studies. *JAMA Netw Open* **4**, e2118811 (2021). https://doi.org:10.1001/jamanetworkopen.2021.18811
33 Deng, K. *et al.* Temporal relationship among adiposity, gut microbiota, and insulin resistance in a longitudinal human cohort. *BMC Med* **20**, 171 (2022). https://doi.org:10.1186/s12916-022-02376-3
34 Khan, M. T., Nieuwdorp, M. & Backhed, F. Microbial modulation of insulin sensitivity. *Cell Metab* **20**, 753-760 (2014). https://doi.org:10.1016/j.cmet.2014.07.006
35 Forslund, K. *et al.* Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature* **528**, 262-266 (2015). https://doi.org:10.1038/nature15766
36 Gardiner, M. *et al.* A longitudinal study of the diabetic skin and wound microbiome. *PeerJ* **5**, e3543 (2017). https://doi.org:10.7717/peerj.3543
37 Redel, H. *et al.* Quantitation and composition of cutaneous microbiota in diabetic and nondiabetic men. *J Infect Dis* **207**, 1105-1114 (2013). https://doi.org:10.1093/infdis/jit005
38 Besnard, P. *et al.* Identification of an oral microbiota signature associated with an impaired orosensory perception of lipids in insulin-resistant patients. *Acta Diabetol* **57**, 1445-1451 (2020). https://doi.org:10.1007/s00592-020-01567-9
39 Watanabe, K. *et al.* Porphyromonas gingivalis impairs glucose uptake in skeletal muscle associated with altering gut microbiota. *FASEB J* **35**, e21171 (2021). https://doi.org:10.1096/fj.202001158R
40 Xiao, E. *et al.* Diabetes Enhances IL-17 Expression and Alters the Oral Microbiome to Increase Its Pathogenicity. *Cell Host Microbe* **22**, 120-128 e124 (2017). https://doi.org:10.1016/j.chom.2017.06.014
41 Almeida-Santos, A., Martins-Mendes, D., Gaya-Vidal, M., Perez-Pardal, L. & Beja-Pereira, A. Characterization of the Oral Microbiome of Medicated Type-2 Diabetes Patients. *Front Microbiol* **12**, 610370 (2021). https://doi.org:10.3389/fmicb.2021.610370
42 Negrini, T. C., Carlos, I. Z., Duque, C., Caiaffa, K. S. & Arthur, R. A. Interplay Among the Oral Microbiome, Oral Cavity Conditions, the Host Immune Response, Diabetes Mellitus, and Its Associated-Risk Factors-An Overview. *Front Oral Health* **2**, 697428 (2021). https://doi.org:10.3389/froh.2021.697428
43 Demmer, R. T. *et al.* Periodontal Bacteria and Prediabetes Prevalence in ORIGINS: The Oral Infections, Glucose Intolerance, and Insulin Resistance Study. *J Dent Res* **94**, 201S-211S (2015). https://doi.org:10.1177/0022034515590369
44 Demmer, R. T. *et al.* Subgingival Microbiota and Longitudinal Glucose Change: The Oral Infections, Glucose Intolerance and Insulin Resistance Study (ORIGINS). *J Dent Res* **98**, 1488-1496 (2019). https://doi.org:10.1177/0022034519881978
45 Vrieze, A. *et al.* Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome. *Gastroenterology* **143**, 913-916 e917 (2012). https://doi.org:10.1053/j.gastro.2012.06.031

46 Giongo, A. *et al.* Toward defining the autoimmune microbiome for type 1 diabetes. *ISME J* **5**, 82-91 (2011). https://doi.org:10.1038/ismej.2010.92
47 Kawano, Y. *et al.* Microbiota imbalance induced by dietary sugar disrupts immune-mediated protection from metabolic syndrome. *Cell* **185**, 3501-3519 e3520 (2022). https://doi.org:10.1016/j.cell.2022.08.005
48 Massier, L. *et al.* Adipose tissue derived bacteria are associated with inflammation in obesity and type 2 diabetes. *Gut* **69**, 1796-1806 (2020). https://doi.org:10.1136/gutjnl-2019-320118
49 Amar, J. *et al.* Intestinal mucosal adherence and translocation of commensal bacteria at the early onset of type 2 diabetes: molecular mechanisms and probiotic treatment. *EMBO Mol Med* **3**, 559-572 (2011). https://doi.org:10.1002/emmm.201100159
50 Linh, H. T. *et al.* Intestinal Bacterial Translocation Contributes to Diabetic Kidney Disease. *J Am Soc Nephrol* **33**, 1105-1119 (2022). https://doi.org:10.1681/ASN.2021060843
51 de Groot, P. *et al.* Donor metabolic characteristics drive effects of faecal microbiota transplantation on recipient insulin sensitivity, energy expenditure and intestinal transit time. *Gut* **69**, 502-512 (2020). https://doi.org:10.1136/gutjnl-2019-318320
52 Balaich, J. *et al.* The human microbiome encodes resistance to the antidiabetic drug acarbose. *Nature* **600**, 110-115 (2021). https://doi.org:10.1038/s41586-021-04091-0
53 Gacesa, R. *et al.* Environmental factors shaping the gut microbiome in a Dutch population. *Nature* **604**, 732-739 (2022). https://doi.org:10.1038/s41586-022-04567-7
54 Asnicar, F. *et al.* Microbiome connections with host metabolism and habitual diet from 1,098 deeply phenotyped individuals. *Nat Med* **27**, 321-332 (2021). https://doi.org:10.1038/s41591-020-01183-8
55 Fassarella, M. *et al.* Gut microbiome stability and resilience: elucidating the response to perturbations in order to modulate gut health. *Gut* **70**, 595-605 (2021). https://doi.org:10.1136/gutjnl-2020-321747
56 Price, N. D. *et al.* A wellness study of 108 individuals using personal, dense, dynamic data clouds. *Nat Biotechnol* **35**, 747-756 (2017). https://doi.org:10.1038/nbt.3870
57 Tebani, A. *et al.* Integration of molecular profiles in a longitudinal wellness profiling cohort. *Nat Commun* **11**, 4487 (2020). https://doi.org:10.1038/s41467-020-18148-7
58 Olsson, L. M. *et al.* Dynamics of the normal gut microbiota: A longitudinal one-year population study in Sweden. *Cell Host Microbe* **30**, 726-739 e723 (2022). https://doi.org:10.1016/j.chom.2022.03.002
59 Li, J. *et al.* Interplay between diet and gut microbiome, and circulating concentrations of trimethylamine N-oxide: findings from a longitudinal cohort of US men. *Gut* **71**, 724-733 (2022). https://doi.org:10.1136/gutjnl-2020-322473
60 Zhou, W. *et al.* Host-Specific Evolutionary and Transmission Dynamics Shape the Functional Diversification of Staphylococcus epidermidis in Human Skin. *Cell* **180**, 454-470 e418 (2020). https://doi.org:10.1016/j.cell.2020.01.006
61 Costello, E. K. *et al.* Bacterial community variation in human body habitats across space and time. *Science* **326**, 1694-1697 (2009). https://doi.org:10.1126/science.1177486
62 Larson, P. J., Chong, D., Fleming, E. & Oh, J. Challenges in Developing a Human Model System for Skin Microbiome Research. *J Invest Dermatol* **141**, 228-231 e224 (2021). https://doi.org:10.1016/j.jid.2020.05.096

63 Selway, C. A., Sudarpa, J. & Weyrich, L. S. Moving beyond the gut microbiome: Combining systems biology and multi-site microbiome analyses to combat non-communicable diseases. *Medicine in Microecology* **12**, 100052 (2022).
64 Oh, J. *et al.* Biogeography and individuality shape function in the human skin metagenome. *Nature* **514**, 59-64 (2014). https://doi.org:10.1038/nature13786
65 Li, Z. *et al.* Characterization of the human skin resistome and identification of two microbiota cutotypes. *Microbiome* **9**, 47 (2021). https://doi.org:10.1186/s40168-020-00995-7
66 David, L. A. *et al.* Host lifestyle affects human microbiota on daily timescales. *Genome Biol* **15**, R89 (2014). https://doi.org:10.1186/gb-2014-15-7-r89
67 Oh, J. *et al.* The altered landscape of the human skin microbiome in patients with primary immunodeficiencies. *Genome Res* **23**, 2103-2114 (2013). https://doi.org:10.1101/gr.159467.113
68 Olofsson, L. E. & Backhed, F. The Metabolic Role and Therapeutic Potential of the Microbiome. *Endocr Rev* **43**, 907-926 (2022). https://doi.org:10.1210/endrev/bnac004
69 Wozniak, J. M. *et al.* Mortality Risk Profiling of Staphylococcus aureus Bacteremia by Multi-omic Serum Analysis Reveals Early Predictive and Pathogenic Signatures. *Cell* **182**, 1311-1327 e1314 (2020). https://doi.org:10.1016/j.cell.2020.07.040
70 Chen, R. *et al.* Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell* **148**, 1293-1307 (2012). https://doi.org:10.1016/j.cell.2012.02.009
71 Integrative, H. M. P. R. N. C. The Integrative Human Microbiome Project. *Nature* **569**, 641-648 (2019). https://doi.org:10.1038/s41586-019-1238-8
72 Zhou, W. *et al.* Longitudinal multi-omics of host-microbe dynamics in prediabetes. *Nature* **569**, 663-671 (2019). https://doi.org:10.1038/s41586-019-1236-x
73 Callahan, B. J., McMurdie, P. J. & Holmes, S. P. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J* **11**, 2639-2643 (2017). https://doi.org:10.1038/ismej.2017.119
74 Johnson, J. S. *et al.* Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat Commun* **10**, 5029 (2019). https://doi.org:10.1038/s41467-019-13036-1
75 Schloss, P. D. Amplicon Sequence Variants Artificially Split Bacterial Genomes into Separate Clusters. *mSphere* **6**, e0019121 (2021). https://doi.org:10.1128/mSphere.00191-21
76 Schussler-Fiorenza Rose, S. M. *et al.* A longitudinal big data approach for precision health. *Nat Med* **25**, 792-804 (2019). https://doi.org:10.1038/s41591-019-0414-6
77 Muniyappa, R., Lee, S., Chen, H. & Quon, M. J. Current approaches for assessing insulin sensitivity and resistance in vivo: advantages, limitations, and appropriate usage. *Am J Physiol Endocrinol Metab* **294**, E15-26 (2008). https://doi.org:10.1152/ajpendo.00645.2007
78 Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174-180 (2011). https://doi.org:10.1038/nature09944
79 Costea, P. I. *et al.* Enterotypes in the landscape of gut microbial community composition. *Nat Microbiol* **3**, 8-16 (2018). https://doi.org:10.1038/s41564-017-0072-8
80 Oh, J., Conlan, S., Polley, E. C., Segre, J. A. & Kong, H. H. Shifts in human skin and nares microbiota of healthy children and adults. *Genome Med* **4**, 77 (2012). https://doi.org:10.1186/gm378

81 Garcia-Lopez, M. *et al.* Analysis of 1,000 Type-Strain Genomes Improves Taxonomic Classification of Bacteroidetes. *Front Microbiol* **10**, 2083 (2019). https://doi.org:10.3389/fmicb.2019.02083
82 Kilian, M. *et al.* The oral microbiome - an update for oral healthcare professionals. *Br Dent J* **221**, 657-666 (2016). https://doi.org:10.1038/sj.bdj.2016.865
83 Aas, J. A., Paster, B. J., Stokes, L. N., Olsen, I. & Dewhirst, F. E. Defining the normal bacterial flora of the oral cavity. *J Clin Microbiol* **43**, 5721-5732 (2005). https://doi.org:10.1128/JCM.43.11.5721-5732.2005
84 Mark Welch, J. L., Rossetti, B. J., Rieken, C. W., Dewhirst, F. E. & Borisy, G. G. Biogeography of a human oral microbiome at the micron scale. *Proc Natl Acad Sci U S A* **113**, E791-800 (2016). https://doi.org:10.1073/pnas.1522149113
85 Lahti, L., Salojarvi, J., Salonen, A., Scheffer, M. & de Vos, W. M. Tipping elements in the human intestinal ecosystem. *Nat Commun* **5**, 4344 (2014). https://doi.org:10.1038/ncomms5344
86 Sailani, M. R. *et al.* Deep longitudinal multiomics profiling reveals two biological seasonal patterns in California. *Nat Commun* **11**, 4933 (2020). https://doi.org:10.1038/s41467-020-18758-1
87 Smits, S. A. *et al.* Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. *Science* **357**, 802-806 (2017). https://doi.org:10.1126/science.aan4834
88 Koliada, A. *et al.* Seasonal variation in gut microbiota composition: cross-sectional evidence from Ukrainian population. *BMC Microbiol* **20**, 100 (2020). https://doi.org:10.1186/s12866-020-01786-8
89 Wastyk, H. C. *et al.* Gut-microbiota-targeted diets modulate human immune status. *Cell* **184**, 4137-4153 e4114 (2021). https://doi.org:10.1016/j.cell.2021.06.019
90 Lancaster, S. M. *et al.* Global, distinctive, and personal changes in molecular and microbial profiles by specific fibers in humans. *Cell Host Microbe* **30**, 848-862 e847 (2022). https://doi.org:10.1016/j.chom.2022.03.036
91 Jiang, C. *et al.* Dynamic Human Environmental Exposome Revealed by Longitudinal Personal Monitoring. *Cell* **175**, 277-291 e231 (2018). https://doi.org:10.1016/j.cell.2018.08.060
92 Fouladi, F. *et al.* Air pollution exposure is associated with the gut microbiome as revealed by shotgun metagenomic sequencing. *Environ Int* **138**, 105604 (2020). https://doi.org:10.1016/j.envint.2020.105604
93 Sunagawa, S. *et al.* Individuality and temporal stability of the human gut microbiome. *Cent Asian J Glob Health* **2**, 120 (2013). https://doi.org:10.5195/cajgh.2013.120
94 Grice, E. A. *et al.* Topographical and temporal diversity of the human skin microbiome. *Science* **324**, 1190-1192 (2009). https://doi.org:10.1126/science.1171700
95 Gajer, P. *et al.* Temporal dynamics of the human vaginal microbiota. *Sci Transl Med* **4**, 132ra152 (2012). https://doi.org:10.1126/scitranslmed.3003605
96 Lemanceau, P., Blouin, M., Muller, D. & Moenne-Loccoz, Y. Let the Core Microbiota Be Functional. *Trends Plant Sci* **22**, 583-595 (2017). https://doi.org:10.1016/j.tplants.2017.04.008
97 Priya, S. & Blekhman, R. Population dynamics of the human gut microbiome: change is the only constant. *Genome Biol* **20**, 150 (2019). https://doi.org:10.1186/s13059-019-1775-3

98 Scapaticci, M., Marchetto, S., Nardi, A., Zoppelletto, M. & Bartolini, A. A case of necrotizing fasciitis caused by Finegoldia magna in a patient with type 2 diabetes mellitus. *Infez Med* **26**, 359-363 (2018).
99 Jneid, J. *et al.* Exploring the Microbiota of Diabetic Foot Infections With Culturomics. *Front Cell Infect Microbiol* **8**, 282 (2018). https://doi.org:10.3389/fcimb.2018.00282
100 Arencibia-Perez, B., Benet-Munoz, O., Roque-Castellano, C. & Marchena-Gomez, J. Gluteal Abscess due to Finegoldia Magna in a Patient with Diabetes. *Actas Dermosifiliogr (Engl Ed)* **111**, 527-528 (2020). https://doi.org:10.1016/j.ad.2018.12.010
101 Castellanos, N. *et al.* A Study on Acinetobacter baumannii and Staphylococcus aureus Strains Recovered from the Same Infection Site of a Diabetic Patient. *Curr Microbiol* **76**, 842-847 (2019). https://doi.org:10.1007/s00284-019-01696-7
102 Perera, D. *et al.* Impaired host response and the presence of Acinetobacter baumannii in the serum microbiome of type-II diabetic patients. *iScience* **24**, 101941 (2021). https://doi.org:10.1016/j.isci.2020.101941
103 Leung, C. H. & Liu, C. P. Diabetic status and the relationship of blood glucose to mortality in adults with carbapenem-resistant Acinetobacter baumannii complex bacteremia. *J Microbiol Immunol Infect* **52**, 654-662 (2019). https://doi.org:10.1016/j.jmii.2018.06.005
104 Henig, O. *et al.* The Impact of Multidrug-Resistant Organisms on Outcomes in Patients With Diabetic Foot Infections. *Open Forum Infect Dis* **7**, ofaa161 (2020). https://doi.org:10.1093/ofid/ofaa161
105 Riviere, A., Selak, M., Lantin, D., Leroy, F. & De Vuyst, L. Bifidobacteria and Butyrate-Producing Colon Bacteria: Importance and Strategies for Their Stimulation in the Human Gut. *Front Microbiol* **7**, 979 (2016). https://doi.org:10.3389/fmicb.2016.00979
106 Min, K. R. *et al.* Association between baseline abundance of Peptoniphilus, a Gram-positive anaerobic coccus, and wound healing outcomes of DFUs. *PLoS One* **15**, e0227006 (2020). https://doi.org:10.1371/journal.pone.0227006
107 Smith, K. *et al.* One step closer to understanding the role of bacteria in diabetic foot ulcers: characterising the microbiome of ulcers. *BMC Microbiol* **16**, 54 (2016). https://doi.org:10.1186/s12866-016-0665-z
108 Brown, K., Church, D., Lynch, T. & Gregson, D. Bloodstream infections due to Peptoniphilus spp.: report of 15 cases. *Clin Microbiol Infect* **20**, O857-860 (2014). https://doi.org:10.1111/1469-0691.12657
109 i, M. C. E. a. s. b. u. e. & i, M. C. Gut microbiome of multiple sclerosis patients and paired household healthy controls reveal associations with disease risk and course. *Cell* **185**, 3467-3486 e3416 (2022). https://doi.org:10.1016/j.cell.2022.08.021
110 Rothschild, D. *et al.* Environment dominates over host genetics in shaping human gut microbiota. *Nature* **555**, 210-215 (2018). https://doi.org:10.1038/nature25973
111 Kort, R. *et al.* Shaping the oral microbiota through intimate kissing. *Microbiome* **2**, 41 (2014). https://doi.org:10.1186/2049-2618-2-41
112 Rocha, J., Henriques, I., Gomila, M. & Manaia, C. M. Common and distinctive genomic features of Klebsiella pneumoniae thriving in the natural environment or in clinical settings. *Sci Rep* **12**, 10441 (2022). https://doi.org:10.1038/s41598-022-14547-6
113 Musher, D. M. in *Medical Microbiology* (eds th & S. Baron) (1996).
114 Zhao, S. *et al.* Adaptive Evolution within Gut Microbiomes of Healthy People. *Cell Host Microbe* **25**, 656-667 e658 (2019). https://doi.org:10.1016/j.chom.2019.03.007

115 Carrow, H. C., Batachari, L. E. & Chu, H. Strain diversity in the microbiome: Lessons from Bacteroides fragilis. *PLoS Pathog* **16**, e1009056 (2020). https://doi.org:10.1371/journal.ppat.1009056
116 Lee, S. M. *et al.* Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature* **501**, 426-429 (2013). https://doi.org:10.1038/nature12447
117 Vieira-Silva, S. *et al.* Statin therapy is associated with lower prevalence of gut microbiota dysbiosis. *Nature* **581**, 310-315 (2020). https://doi.org:10.1038/s41586-020-2269-x
118 Ontiveros, V. J., Capitan, J. A., Casamayor, E. O. & Alonso, D. The characteristic time of ecological communities. *Ecology* **102**, e03247 (2021). https://doi.org:10.1002/ecy.3247
119 Real, R. & Vargas, J. M. The Probabilistic Basis of Jaccard's Index of Similarity. *Systematic Biology* **45**, 380-385 (1996). https://doi.org:10.1093/sysbio/45.3.380
120 Li, S. S. *et al.* Durable coexistence of donor and recipient strains after fecal microbiota transplantation. *Science* **352**, 586-589 (2016). https://doi.org:10.1126/science.aad8852
121 Russell, B. J. *et al.* Intestinal transgene delivery with native E. coli chassis allows persistent physiological changes. *Cell* **185**, 3263-3277 e3215 (2022). https://doi.org:10.1016/j.cell.2022.06.050
122 Zhao, L. *et al.* Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes. *Science* **359**, 1151-1156 (2018). https://doi.org:10.1126/science.aao5774
123 Wu, G., Zhao, N., Zhang, C., Lam, Y. Y. & Zhao, L. Guild-based analysis for understanding gut microbiome in human health and diseases. *Genome Med* **13**, 22 (2021). https://doi.org:10.1186/s13073-021-00840-y
124 Chen, X. *et al.* Gut dysbiosis induces the development of pre-eclampsia through bacterial translocation. *Gut* **69**, 513-522 (2020). https://doi.org:10.1136/gutjnl-2019-319101
125 Marchetti, G., Tincati, C. & Silvestri, G. Microbial translocation in the pathogenesis of HIV infection and AIDS. *Clin Microbiol Rev* **26**, 2-18 (2013). https://doi.org:10.1128/CMR.00050-12
126 Bartley, J. M., Zhou, X., Kuchel, G. A., Weinstock, G. M. & Haynes, L. Impact of Age, Caloric Restriction, and Influenza Infection on Mouse Gut Microbiome: An Exploratory Study of the Role of Age-Related Microbiome Changes on Influenza Responses. *Front Immunol* **8**, 1164 (2017). https://doi.org:10.3389/fimmu.2017.01164
127 Lynn, D. J., Benson, S. C., Lynn, M. A. & Pulendran, B. Modulation of immune responses to vaccination by the microbiota: implications and potential mechanisms. *Nat Rev Immunol* **22**, 33-46 (2022). https://doi.org:10.1038/s41577-021-00554-7
128 Schirmer, M. *et al.* Linking the Human Gut Microbiome to Inflammatory Cytokine Production Capacity. *Cell* **167**, 1125-1136 e1128 (2016). https://doi.org:10.1016/j.cell.2016.10.020
129 Geva-Zatorsky, N. *et al.* Mining the Human Gut Microbiota for Immunomodulatory Organisms. *Cell* **168**, 928-943 e911 (2017). https://doi.org:10.1016/j.cell.2017.01.022
130 Mendes, V., Galvao, I. & Vieira, A. T. Mechanisms by Which the Gut Microbiota Influences Cytokine Production and Modulates Host Inflammatory Responses. *J Interferon Cytokine Res* **39**, 393-409 (2019). https://doi.org:10.1089/jir.2019.0011
131 Bolte, L. A. *et al.* Long-term dietary patterns are associated with pro-inflammatory and anti-inflammatory features of the gut microbiome. *Gut* **70**, 1287-1298 (2021). https://doi.org:10.1136/gutjnl-2020-322670

132 Zhou, X. *et al.* Longitudinal Analysis of Serum Cytokine Levels and Gut Microbial Abundance Links IL-17/IL-22 With Clostridia and Insulin Sensitivity in Humans. *Diabetes* **69**, 1833-1842 (2020). https://doi.org:10.2337/db19-0592
133 Yang, D. *et al.* New statistical method identifies cytokines that distinguish stool microbiomes. *Sci Rep* **9**, 20082 (2019). https://doi.org:10.1038/s41598-019-56397-9
134 Lloyd-Price, J. *et al.* Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature* **550**, 61-66 (2017). https://doi.org:10.1038/nature23889
135 Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027-1031 (2006). https://doi.org:10.1038/nature05414
136 Somineni, H. K. *et al.* Site- and Taxa-Specific Disease-Associated Oral Microbial Structures Distinguish Inflammatory Bowel Diseases. *Inflamm Bowel Dis* **27**, 1889-1900 (2021). https://doi.org:10.1093/ibd/izab082
137 Gao, Z., Tseng, C. H., Strober, B. E., Pei, Z. & Blaser, M. J. Substantial alterations of the cutaneous bacterial biota in psoriatic lesions. *PLoS One* **3**, e2719 (2008). https://doi.org:10.1371/journal.pone.0002719
138 Shin, N. R., Whon, T. W. & Bae, J. W. Proteobacteria: microbial signature of dysbiosis in gut microbiota. *Trends Biotechnol* **33**, 496-503 (2015). https://doi.org:10.1016/j.tibtech.2015.06.011
139 Bradley, P. H. & Pollard, K. S. Proteobacteria explain significant functional variability in the human gut microbiome. *Microbiome* **5**, 36 (2017). https://doi.org:10.1186/s40168-017-0244-z
140 d'Hennezel, E., Abubucker, S., Murphy, L. O. & Cullen, T. W. Total Lipopolysaccharide from the Human Gut Microbiome Silences Toll-Like Receptor Signaling. *mSystems* **2** (2017). https://doi.org:10.1128/mSystems.00046-17
141 Pither, M. D. *et al.* Bacteroides thetaiotaomicron rough-type lipopolysaccharide: The chemical structure and the immunological activity. *Carbohydr Polym* **297**, 120040 (2022). https://doi.org:10.1016/j.carbpol.2022.120040
142 Vatanen, T. *et al.* Variation in Microbiome LPS Immunogenicity Contributes to Autoimmunity in Humans. *Cell* **165**, 842-853 (2016). https://doi.org:10.1016/j.cell.2016.04.007
143 Deriu, E. *et al.* Influenza Virus Affects Intestinal Microbiota and Secondary Salmonella Infection in the Gut through Type I Interferons. *PLoS Pathog* **12**, e1005572 (2016). https://doi.org:10.1371/journal.ppat.1005572
144 Winter, S. E. & Baumler, A. J. Why related bacterial species bloom simultaneously in the gut: principles underlying the 'Like will to like' concept. *Cell Microbiol* **16**, 179-184 (2014). https://doi.org:10.1111/cmi.12245
145 Scales, B. S., Dickson, R. P. & Huffnagle, G. B. A tale of two sites: how inflammation can reshape the microbiomes of the gut and lungs. *J Leukoc Biol* **100**, 943-950 (2016). https://doi.org:10.1189/jlb.3MR0316-106R
146 Park, S. Y. *et al.* Strain-level fitness in the gut microbiome is an emergent property of glycans and a single metabolite. *Cell* **185**, 513-529 e521 (2022). https://doi.org:10.1016/j.cell.2022.01.002
147 Petersen, L. M. *et al.* Community characteristics of the gut microbiomes of competitive cyclists. *Microbiome* **5**, 98 (2017). https://doi.org:10.1186/s40168-017-0320-4

148 Patra, A. K. & Yu, Z. Genomic Insights into the Distribution of Peptidases and Proteolytic Capacity among Prevotella and Paraprevotella Species. *Microbiol Spectr* **10**, e0218521 (2022). https://doi.org:10.1128/spectrum.02185-21
149 Lamichhane, S. *et al.* Linking Gut Microbiome and Lipid Metabolism: Moving beyond Associations. *Metabolites* **11** (2021). https://doi.org:10.3390/metabo11010055
150 Flint, H. J. & Duncan, S. Bacteroides and prevotella. (2014).
151 Song, B.-J., Akbar, M., Jo, I., Hardwick, J. P. & Abdelmegeed, M. A. Translational implications of the alcohol-metabolizing enzymes, including cytochrome P450-2E1, in alcoholic and nonalcoholic liver disease. *Advances in Pharmacology* **74**, 303-372 (2015).
152 Engen, P. A., Green, S. J., Voigt, R. M., Forsyth, C. B. & Keshavarzian, A. The Gastrointestinal Microbiome: Alcohol Effects on the Composition of Intestinal Microbiota. *Alcohol Res* **37**, 223-236 (2015).
153 Kosnicki, K. L. *et al.* Effects of moderate, voluntary ethanol consumption on the rat and human gut microbiome. *Addict Biol* **24**, 617-630 (2019). https://doi.org:10.1111/adb.12626
154 Fan, X. *et al.* Drinking alcohol is associated with variation in the human oral microbiome in a large study of American adults. *Microbiome* **6**, 59 (2018). https://doi.org:10.1186/s40168-018-0448-x
155 Yuan, J. *et al.* Fatty Liver Disease Caused by High-Alcohol-Producing Klebsiella pneumoniae. *Cell Metab* **30**, 675-688 e677 (2019). https://doi.org:10.1016/j.cmet.2019.08.018
156 Li, N. N. *et al.* High alcohol-producing Klebsiella pneumoniae causes fatty liver disease through 2,3-butanediol fermentation pathway in vivo. *Gut Microbes* **13**, 1979883 (2021). https://doi.org:10.1080/19490976.2021.1979883
157 Sakamoto, M., Iino, T. & Ohkuma, M. Faecalimonas umbilicata gen. nov., sp. nov., isolated from human faeces, and reclassification of Eubacterium contortum, Eubacterium fissicatena and Clostridium oroticum as Faecalicatena contorta gen. nov., comb. nov., Faecalicatena fissicatena comb. nov. and Faecalicatena orotica comb. nov. *Int J Syst Evol Microbiol* **67**, 1219-1227 (2017). https://doi.org:10.1099/ijsem.0.001790
158 Sakamoto, M., Ikeyama, N., Yuki, M. & Ohkuma, M. Draft Genome Sequence of Faecalimonas umbilicata JCM 30896(T), an Acetate-Producing Bacterium Isolated from Human Feces. *Microbiol Resour Announc* **7** (2018). https://doi.org:10.1128/MRA.01091-18
159 Spector, M. Metabolism, central (intermediary). (2009).
160 Tsuruya, A. *et al.* Ecophysiological consequences of alcoholism on human gut microbiota: implications for ethanol-related pathogenesis of colon cancer. *Sci Rep* **6**, 27923 (2016). https://doi.org:10.1038/srep27923
161 Petersen, C. *et al.* T cell-mediated regulation of the microbiota protects against obesity. *Science* **365** (2019). https://doi.org:10.1126/science.aat9351
162 Lin, Y. C., Lin, H. F., Wu, C. C., Chen, C. L. & Ni, Y. H. Pathogenic effects of Desulfovibrio in the gut on fatty liver in diet-induced obese mice and children with obesity. *J Gastroenterol* (2022). https://doi.org:10.1007/s00535-022-01909-0
163 Le Chatelier, E. *et al.* Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541-546 (2013). https://doi.org:10.1038/nature12506

164 Wilmanski, T. *et al.* Blood metabolome predicts gut microbiome alpha-diversity in humans. *Nat Biotechnol* **37**, 1217-1228 (2019). https://doi.org:10.1038/s41587-019-0233-9

165 Passmore, I. J. *et al.* Para-cresol production by Clostridium difficile affects microbial diversity and membrane integrity of Gram-negative bacteria. *PLoS Pathog* **14**, e1007191 (2018). https://doi.org:10.1371/journal.ppat.1007191

166 Koppe, L. *et al.* p-Cresyl glucuronide is a major metabolite of p-cresol in mouse: in contrast to p-cresyl sulphate, p-cresyl glucuronide fails to promote insulin resistance. *Nephrol Dial Transplant* **32**, 2000-2009 (2017). https://doi.org:10.1093/ndt/gfx089

167 VanderWeele, T. J. Mediation Analysis: A Practitioner's Guide. *Annu Rev Public Health* **37**, 17-32 (2016). https://doi.org:10.1146/annurev-publhealth-032315-021402

168 Jiang, Z. *et al.* The gut microbiota-bile acid axis links the positive association between chronic insomnia and cardiometabolic diseases. *Nat Commun* **13**, 3002 (2022). https://doi.org:10.1038/s41467-022-30712-x

169 Lee, H. *et al.* A Guideline for Reporting Mediation Analyses of Randomized Trials and Observational Studies: The AGReMA Statement. *JAMA* **326**, 1045-1056 (2021). https://doi.org:10.1001/jama.2021.14075

170 Demmer, R. T. *et al.* The subgingival microbiome, systemic inflammation and insulin resistance: The Oral Infections, Glucose Intolerance and Insulin Resistance Study. *J Clin Periodontol* **44**, 255-265 (2017). https://doi.org:10.1111/jcpe.12664

171 Demmer, R. T. *et al.* Periodontal infection, systemic inflammation, and insulin resistance: results from the continuous National Health and Nutrition Examination Survey (NHANES) 1999-2004. *Diabetes Care* **35**, 2235-2242 (2012). https://doi.org:10.2337/dc12-0072

172 Li, X., Kolltveit, K. M., Tronstad, L. & Olsen, I. Systemic diseases caused by oral infection. *Clin Microbiol Rev* **13**, 547-558 (2000). https://doi.org:10.1128/CMR.13.4.547

173 Taylor, J. J., Preshaw, P. M. & Lalla, E. A review of the evidence for pathogenic mechanisms that may link periodontitis and diabetes. *J Clin Periodontol* **40 Suppl 14**, S113-134 (2013). https://doi.org:10.1111/jcpe.12059

174 Polak, D. & Shapira, L. An update on the evidence for pathogenic mechanisms that may link periodontitis and diabetes. *J Clin Periodontol* **45**, 150-166 (2018). https://doi.org:10.1111/jcpe.12803

175 Shih, C. T., Yeh, Y. T., Lin, C. C., Yang, L. Y. & Chiang, C. P. Akkermansia muciniphila is Negatively Correlated with Hemoglobin A1c in Refractory Diabetes. *Microorganisms* **8** (2020). https://doi.org:10.3390/microorganisms8091360

176 Zhang, J. *et al.* Decreased Abundance of Akkermansia muciniphila Leads to the Impairment of Insulin Secretion and Glucose Homeostasis in Lean Type 2 Diabetes. *Adv Sci (Weinh)* **8**, e2100536 (2021). https://doi.org:10.1002/advs.202100536

177 Yoon, H. S. *et al.* Akkermansia muciniphila secretes a glucagon-like peptide-1-inducing protein that improves glucose homeostasis and ameliorates metabolic disease in mice. *Nat Microbiol* **6**, 563-573 (2021). https://doi.org:10.1038/s41564-021-00880-5

178 Cani, P. D., Depommier, C., Derrien, M., Everard, A. & de Vos, W. M. Akkermansia muciniphila: paradigm for next-generation beneficial microorganisms. *Nat Rev Gastroenterol Hepatol* **19**, 625-637 (2022). https://doi.org:10.1038/s41575-022-00631-9

179 Zheng, H. *et al.* Altered Gut Microbiota Composition Associated with Eczema in Infants. *PLoS One* **11**, e0166026 (2016). https://doi.org:10.1371/journal.pone.0166026

180 Jangi, S. *et al.* Alterations of the human gut microbiome in multiple sclerosis. *Nat Commun* **7**, 12015 (2016). https://doi.org:10.1038/ncomms12015
181 Cekanaviciute, E. *et al.* Gut bacteria from multiple sclerosis patients modulate human T cells and exacerbate symptoms in mouse models. *Proc Natl Acad Sci U S A* **114**, 10713-10718 (2017). https://doi.org:10.1073/pnas.1711235114
182 Volkova, A. & Ruggles, K. V. Predictive Metagenomic Analysis of Autoimmune Disease Identifies Robust Autoimmunity and Disease Specific Microbial Signatures. *Front Microbiol* **12**, 621310 (2021). https://doi.org:10.3389/fmicb.2021.621310
183 Norskov-Lauritsen, N. Classification, identification, and clinical significance of Haemophilus and Aggregatibacter species with host specificity for humans. *Clin Microbiol Rev* **27**, 214-240 (2014). https://doi.org:10.1128/CMR.00103-13
184 Bouslimani, A. *et al.* Molecular cartography of the human skin surface in 3D. *Proc Natl Acad Sci U S A* **112**, E2120-2129 (2015). https://doi.org:10.1073/pnas.1424409112
185 Allhorn, M., Arve, S., Bruggemann, H. & Lood, R. A novel enzyme with antioxidant capacity produced by the ubiquitous skin colonizer Propionibacterium acnes. *Sci Rep* **6**, 36412 (2016). https://doi.org:10.1038/srep36412
186 Li, X., Wang, X. & Snyder, M. P. Metformin Affects Heme Function as a Possible Mechanism of Action. *G3 (Bethesda)* **9**, 513-522 (2019). https://doi.org:10.1534/g3.118.200803
187 Tang, S. Y., Cheah, I. K. M., Ng, P. E., Hoi, A. & Jenner, A. M. Heme consumption reduces hepatic triglyceride and fatty acid accumulation in a rat model of NAFLD fed westernized diet. *International Scholarly Research Notices* **2014** (2014).
188 Ndisang, J. F. & Mishra, M. The heme oxygenase system selectively suppresses the proinflammatory macrophage m1 phenotype and potentiates insulin signaling in spontaneously hypertensive rats. *Am J Hypertens* **26**, 1123-1131 (2013). https://doi.org:10.1093/ajh/hpt082
189 Varga, C. *et al.* The Effects of Exercise Training and High Triglyceride Diet in an Estrogen Depleted Rat Model: The Role of the Heme Oxygenase System and Inflammatory Processes in Cardiovascular Risk. *J Sports Sci Med* **17**, 580-588 (2018).
190 Cimini, F. A. *et al.* Biliverdin reductase-A protein levels are reduced in type 2 diabetes and are associated with poor glycometabolic control. *Life Sci* **284**, 119913 (2021). https://doi.org:10.1016/j.lfs.2021.119913
191 Mayneris-Perxachs, J. *et al.* Iron status influences non-alcoholic fatty liver disease in obesity through the gut microbiome. *Microbiome* **9**, 104 (2021). https://doi.org:10.1186/s40168-021-01052-7
192 Qin, J. *et al.* A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55-60 (2012). https://doi.org:10.1038/nature11450
193 Lv, L. *et al.* The Salivary Microbiota, Cytokines, and Metabolome in Patients with Ankylosing Spondylitis Are Altered and More Proinflammatory than Those in Healthy Controls. *mSystems* **6**, e0117320 (2021). https://doi.org:10.1128/mSystems.01173-20
194 Elias, A. E., McBain, A. J. & O'Neill, C. A. The role of the skin microbiota in the modulation of cutaneous inflammation-Lessons from the gut. *Exp Dermatol* **30**, 1509-1516 (2021). https://doi.org:10.1111/exd.14420
195 Harris-Tryon, T. A. & Grice, E. A. Microbiota and maintenance of skin barrier function. *Science* **376**, 940-945 (2022). https://doi.org:10.1126/science.abo0693

196 Vujkovic-Cvijin, I. *et al.* Host variables confound gut microbiota studies of human disease. *Nature* **587**, 448-454 (2020). https://doi.org:10.1038/s41586-020-2881-9
197 Zhu, L. *et al.* Characterization of gut microbiomes in nonalcoholic steatohepatitis (NASH) patients: a connection between endogenous alcohol and NASH. *Hepatology* **57**, 601-609 (2013). https://doi.org:10.1002/hep.26093
198 Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105-108 (2011). https://doi.org:10.1126/science.1208344
199 Gribbon, E. M., Cunliffe, W. J. & Holland, K. T. Interaction of Propionibacterium acnes with skin lipids in vitro. *J Gen Microbiol* **139**, 1745-1751 (1993). https://doi.org:10.1099/00221287-139-8-1745
200 Mayslich, C., Grange, P. A. & Dupin, N. Cutibacterium acnes as an Opportunistic Pathogen: An Update of Its Virulence-Associated Factors. *Microorganisms* **9** (2021). https://doi.org:10.3390/microorganisms9020303
201 Brown, E. M. *et al.* Bacteroides-Derived Sphingolipids Are Critical for Maintaining Intestinal Homeostasis and Symbiosis. *Cell Host Microbe* **25**, 668-680 e667 (2019). https://doi.org:10.1016/j.chom.2019.04.002
202 Kenny, D. J. *et al.* Cholesterol Metabolism by Uncultured Human Gut Bacteria Influences Host Cholesterol Level. *Cell Host Microbe* **28**, 245-257 e246 (2020). https://doi.org:10.1016/j.chom.2020.05.013
203 Cryan, J. F. & Mazmanian, S. K. Microbiota-brain axis: Context and causality. *Science* **376**, 938-939 (2022). https://doi.org:10.1126/science.abo4442
204 Needham, B. D. *et al.* A gut-derived metabolite alters brain activity and anxiety behaviour in mice. *Nature* **602**, 647-653 (2022). https://doi.org:10.1038/s41586-022-04396-8
205 Budden, K. F. *et al.* Emerging pathogenic links between microbiota and the gut-lung axis. *Nat Rev Microbiol* **15**, 55-63 (2017). https://doi.org:10.1038/nrmicro.2016.142
206 Marsland, B. J., Trompette, A. & Gollwitzer, E. S. The Gut-Lung Axis in Respiratory Disease. *Ann Am Thorac Soc* **12 Suppl 2**, S150-156 (2015). https://doi.org:10.1513/AnnalsATS.201503-133AW
207 Atarashi, K. *et al.* Ectopic colonization of oral bacteria in the intestine drives TH1 cell induction and inflammation. *Science* **358**, 359-365 (2017). https://doi.org:10.1126/science.aan4526
208 Samuelson, D. R. *et al.* Alcohol-associated intestinal dysbiosis impairs pulmonary host defense against Klebsiella pneumoniae. *PLoS Pathog* **13**, e1006426 (2017). https://doi.org:10.1371/journal.ppat.1006426
209 Shellito, J. E. *et al.* Effect of alcohol consumption on host release of interleukin-17 during pulmonary infection with Klebsiella pneumoniae. *Alcohol Clin Exp Res* **25**, 872-881 (2001).
210 Samuelson, D. R. *et al.* Pulmonary immune cell trafficking promotes host defense against alcohol-associated Klebsiella pneumonia. *Commun Biol* **4**, 997 (2021). https://doi.org:10.1038/s42003-021-02524-0
211 Ding, T. & Schloss, P. D. Dynamics and associations of microbial community types across the human body. *Nature* **509**, 357-360 (2014). https://doi.org:10.1038/nature13178
212 Schmidt, T. S. *et al.* Extensive transmission of microbes along the gastrointestinal tract. *Elife* **8** (2019). https://doi.org:10.7554/eLife.42693

213 Kohn, J. N. *et al.* Differing salivary microbiome diversity, community and diurnal rhythmicity in association with affective state and peripheral inflammation in adults. *Brain Behav Immun* **87**, 591-602 (2020). https://doi.org:10.1016/j.bbi.2020.02.004
214 Koren, O. *et al.* Human oral, gut, and plaque microbiota in patients with atherosclerosis. *Proc Natl Acad Sci U S A* **108 Suppl 1**, 4592-4598 (2011). https://doi.org:10.1073/pnas.1011383107
215 Jungbauer, G. *et al.* Periodontal microorganisms and Alzheimer disease - A causative relationship? *Periodontol 2000* **89**, 59-82 (2022). https://doi.org:10.1111/prd.12429
216 Panza, F., Lozupone, M., Solfrizzi, V., Watling, M. & Imbimbo, B. P. Time to test antibacterial therapy in Alzheimer's disease. *Brain* **142**, 2905-2929 (2019). https://doi.org:10.1093/brain/awz244
217 Ren, Z. *et al.* Alterations in the human oral and gut microbiomes and lipidomics in COVID-19. *Gut* **70**, 1253-1265 (2021). https://doi.org:10.1136/gutjnl-2020-323826
218 Hintao, J., Teanpaisan, R., Chongsuvivatwong, V., Ratarasan, C. & Dahlen, G. The microbiological profiles of saliva, supragingival and subgingival plaque and dental caries in adults with and without type 2 diabetes mellitus. *Oral Microbiol Immunol* **22**, 175-181 (2007). https://doi.org:10.1111/j.1399-302X.2007.00341.x
219 Hong, B. Y. *et al.* Microbiome profiles in periodontitis in relation to host and disease characteristics. *PLoS One* **10**, e0127077 (2015). https://doi.org:10.1371/journal.pone.0127077
220 Al-Maskari, A. Y., Al-Maskari, M. Y. & Al-Sudairy, S. Oral Manifestations and Complications of Diabetes Mellitus: A review. *Sultan Qaboos Univ Med J* **11**, 179-186 (2011).
221 Meisel, J. S. *et al.* Skin Microbiome Surveys Are Strongly Influenced by Experimental Design. *J Invest Dermatol* **136**, 947-956 (2016). https://doi.org:10.1016/j.jid.2016.01.016
222 He, Y. *et al.* Regional variation limits applications of healthy gut microbiome reference ranges and disease models. *Nat Med* **24**, 1532-1535 (2018). https://doi.org:10.1038/s41591-018-0164-x
223 Gaulke, C. A. & Sharpton, T. J. The influence of ethnicity and geography on human gut microbiome composition. *Nat Med* **24**, 1495-1496 (2018). https://doi.org:10.1038/s41591-018-0210-8
224 Kelly, P., Alderton, G., Scanlon, S. T. & Ash, C. A multiplicity of microbiomes. *Science* **376**, 932-933 (2022). https://doi.org:10.1126/science.adc9690
225 Hummel, W. & Kula, M.-R. Simple method for small-scale disruption of bacteria and yeasts. *Journal of microbiological methods* **9**, 201-209 (1989).
226 Cady, N. C., Stelick, S. & Batt, C. A. Nucleic acid purification using microfabricated silicon structures. *Biosens Bioelectron* **19**, 59-66 (2003). https://doi.org:10.1016/s0956-5663(03)00123-4
227 Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* **6**, 1621-1624 (2012). https://doi.org:10.1038/ismej.2012.8
228 Caporaso, J. G. *et al.* Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A* **108 Suppl 1**, 4516-4522 (2011). https://doi.org:10.1073/pnas.1000080107
229 Minalla, A., Dubrow, R. S. & Bousse, L. J. in *MOEMS-MEMS.*

230 Almonacid, D. E. *et al.* 16S rRNA gene sequencing and healthy reference ranges for 28 clinically relevant microbial taxa from the human gut microbiome. *PLoS One* **12**, e0176555 (2017). https://doi.org:10.1371/journal.pone.0176555
231 Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581-583 (2016). https://doi.org:10.1038/nmeth.3869
232 Kuczynski, J. *et al.* Microbial community resemblance methods differ in their ability to detect biologically relevant patterns. *Nat Methods* **7**, 813-819 (2010). https://doi.org:10.1038/nmeth.1499
233 Contrepois, K. *et al.* Molecular Choreography of Acute Exercise. *Cell* **181**, 1112-1130 e1116 (2020). https://doi.org:10.1016/j.cell.2020.04.043
234 Contrepois, K. *et al.* Cross-Platform Comparison of Untargeted and Targeted Lipidomics Approaches on Aging Mouse Plasma. *Sci Rep* **8**, 17747 (2018). https://doi.org:10.1038/s41598-018-35807-4
235 Contrepois, K., Jiang, L. & Snyder, M. Optimized Analytical Procedures for the Untargeted Metabolomic Profiling of Human Urine and Plasma by Combining Hydrophilic Interaction (HILIC) and Reverse-Phase Liquid Chromatography (RPLC)-Mass Spectrometry. *Mol Cell Proteomics* **14**, 1684-1695 (2015). https://doi.org:10.1074/mcp.M114.046508
236 Rost, H. L., Schmitt, U., Aebersold, R. & Malmstrom, L. pyOpenMS: a Python-based interface to the OpenMS mass-spectrometry algorithm library. *Proteomics* **14**, 74-77 (2014). https://doi.org:10.1002/pmic.201300246
237 Rost, H. L. *et al.* TRIC: an automated alignment strategy for reproducible protein quantification in targeted proteomics. *Nat Methods* **13**, 777-783 (2016). https://doi.org:10.1038/nmeth.3954
238 Jiang, C., Zhang, X., Gao, P., Chen, Q. & Snyder, M. Decoding personal biotic and abiotic airborne exposome. *Nat Protoc* **16**, 1129-1151 (2021). https://doi.org:10.1038/s41596-020-00451-8
239 Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420 (2018). https://doi.org:10.1038/nbt.4096
240 Jari Oksanen, F. G. B., Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs, Helene Wagner vegan: Community Ecology Package. *R package version 2.5-1* (2018).
241 Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823* (2014).
242 Shade, A. & Handelsman, J. Beyond the Venn diagram: the hunt for a core microbiome. *Environ Microbiol* **14**, 4-12 (2012). https://doi.org:10.1111/j.1462-2920.2011.02585.x
243 Risely, A. Applying the core microbiome to understand host-microbe systems. *J Anim Ecol* **89**, 1549-1558 (2020). https://doi.org:10.1111/1365-2656.13229
244 Salonen, A., Salojarvi, J., Lahti, L. & de Vos, W. M. The adult intestinal core microbiota is determined by analysis depth and health status. *Clin Microbiol Infect* **18 Suppl 4**, 16-20 (2012). https://doi.org:10.1111/j.1469-0691.2012.03855.x
245 Nelson, L. S. The Anderson-Darling Test for Normality. *Journal of Quality Technology* **30**, 298-299 (1998). https://doi.org:10.1080/00224065.1998.11979858

246 Johnson, S. G. *The NLopt nonlinear-optimization package* (http://ab-initio.mit.edu/nlopt, 2011).
247 Zeileis, A. & Hothorn, T. Diagnostic checking in regression relationships. (2002).
248 Satterthwaite, F. E. An approximate distribution of estimates of variance components. *Biometrics* **2**, 110-114 (1946).
249 Bürkner, P.-C. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* **80**, 1-28 (2017).
250 Bürkner, P.-C. Advanced Bayesian Multilevel Modeling with the R Package brms. *arXiv preprint arXiv:1705.11123* (2017).
251 Bürkner, P.-C. Bayesian Item Response Modeling in R with brms and Stan. *Journal of Statistical Software* **100**, 1 - 54 (2021). https://doi.org:10.18637/jss.v100.i05
252 Homan, M. D. & Gelman, A. The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15**, 1593–1623 (2014).
253 Hughes, D. A. *et al.* Genome-wide associations of human gut microbiome variation and implications for causal inference analyses. *Nat Microbiol* **5**, 1079-1087 (2020). https://doi.org:10.1038/s41564-020-0743-8
254 Koh, A. & Backhed, F. From Association to Causality: the Role of the Gut Microbiota and Its Functional Products on Host Metabolism. *Mol Cell* **78**, 584-596 (2020). https://doi.org:10.1016/j.molcel.2020.03.005
255 Schaum, N. *et al.* Ageing hallmarks exhibit organ-specific temporal signatures. *Nature* **583**, 596-602 (2020). https://doi.org:10.1038/s41586-020-2499-y
256 Narunsky-Haziza, L. *et al.* Pan-cancer analyses reveal cancer-type-specific fungal ecologies and bacteriome interactions. *Cell* **185**, 3789-3806 e3717 (2022). https://doi.org:10.1016/j.cell.2022.09.005
257 Gao, P. *et al.* Precision environmental health monitoring by longitudinal exposome and multi-omics profiling. *Genome Res* **32**, 1199-1214 (2022). https://doi.org:10.1101/gr.276521.121
258 Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol* **12**, R60 (2011). https://doi.org:10.1186/gb-2011-12-6-r60