

**Supplemental Figure and Table legends**

**Figure S1. Sequencing and variant call quality metrics.** 

a Percentage of contamination and chimeric reads. 

b Various metrics based on variant calling.

**Figure S2. Distribution of genic and intergenic variants.** 

Proportion of intronic, exonic and intergenic variants in HMP300, GoNL, and 1000 Genome cohorts.

**Figure S3. Combined PCA analysis between 1000 Genomes and HMP300.** 

Both panels correspond to the same underlying coordinate system defined by a PCA of a merger of common variants from 1000 Genomes and HMP300 datasets.

**Figure S4. Correlation between high-level genetic features and microbial species in non-gut body sites.** 

Same analysis as in Figure 2b, performed using other body sites.

**Figure S5. Correlation between high-level genetic features and microbial metabolic pathways in non-gut body sites.** 

Same analysis as in Figure 2c, performed using other body sites.

**Figure S6. Quantile-quantile plots for association analysis between microbial species and GWAS catalog SNVs.**

**Figure S7. Quantile-quantile plots for association analysis between microbial metabolic pathways and GWAS catalog SNVs.**

**Figure S8. Putative SNV–microbial species associations.** 

All associations between single nucleotide variants and microbial species with $p < 5 \times 10^{-8}$ are shown, with color and size of the edges displaying sampling site and association strength respectively.

**Figure S9. Putative SNV–microbial metabolic pathway associations.** 

All associations between single nucleotide variants and microbial metabolic pathways with $p < 5 \times 10^{-8}$ are shown, with color and size of the edges displaying sampling site and association strength respectively.
Table S1. Raw statistics for genetic principal component analysis. Sheet 1 contains statistics for average $R^2$ calculations shown in Figure 2b. The additional 12 sheets provide statistics for individual species- and pathway-level analyses for all six body sites, corresponding to Figure 2c and Figures S4-S5.

Table S2. Top SNV and microbiome association results. Top associations ($p < 10^{-6}$) are shown for both species and pathways in all six body sites.
Figure S2
Figure S3

![Figure S3](image-url)
Figure S4

Empirical p value for variance explained by genomic PC1−PC5

- **Oral - Buccal mucosa**
  - Porphyromonas catoniae
  - Propionibacterium propionicum
  - Gemella unclassified
  - Ethnicity $R^2$
    - 0.10
    - 0.05
  - Count
  - $p < 0.0008$

- **Oral - Supragingival plaque**
  - Ethnicity $R^2$
    - 0.15
    - 0.10
    - 0.05
  - Count
  - $p < 0.022$

- **Oral - Tongue dorsum**
  - Ethnicity $R^2$
    - 0.10
    - 0.05
  - Count
  - $p < 0.0034$

- **Skin - Anterior nares**
  - Ethnicity $R^2$
    - 0.06
    - 0.04
    - 0.02
  - Count
  - $p = 0.16$

- **Vaginal - Posterior fornix**
  - Ethnicity $R^2$
    - 0.15
    - 0.10
    - 0.05
  - Count
  - $p = 0.39$
Figure S5

Empirical p-value for variance explained by genomic PC1-PC5

- Oral - Buccal mucosa
- Oral - Supragingival plaque
- Oral - Tongue dorsum
- Skin - Anterior nares
- Vaginal - Posterior fornix

- Count
- Ethnicity R²

- Oral - Buccal mucosa
  - Ethnicity R²: 0.3
  - p < 0.05

- Oral - Supragingival plaque
  - Ethnicity R²: 0.125
  - p < 0.074

- Oral - Tongue dorsum
  - Ethnicity R²: 0.1
  - p < 0.036

- Skin - Anterior nares
  - Ethnicity R²: 0.15
  - p < 0.16

- Vaginal - Posterior fornix
  - Ethnicity R²: 0.3
  - p < 0.12
Figure S6

Species: Gut-Stool
QQ-plot for all 1,743,096 p-values

Species: Oral-Buccal Mucosa
QQ-plot for all 1,743,450 p-values

Species: Oral-Supragingival Plaque
QQ-plot for all 2,375,555 p-values

Species: Skin-Anterior Nares
QQ-plot for all 395,901 p-values

Species: Vaginal-Posterior Fornix
QQ-plot for all 341,895 p-values
Figure S7

Species: Gut-Stool
QQ-plot for all 7,953,762 p-values

Species: Oral-Buccal Mucosa
QQ-plot for all 6,988,575 p-values

Species: Oral-Supragingival Plaque
QQ-plot for all 7,884,415 p-values

Species: Oral-Tongue Dorsum
QQ-plot for all 7,852,959 p-values

Species: Skin- Anterior Nares
QQ-plot for all 6,120,608 p-values

Species: Vaginal-Posterior Fornix
QQ-plot for all 5,232,488 p-values
Figure S8

Site
- Gut - Stool
- Oral - Buccal mucosa
- Oral - Supragingival plaque
- Skin - Anterior nares
- Vaginal - Posterior fornix

P-value
- $10^{-11}$
- $10^{-8}$
<table>
<thead>
<tr>
<th>MAF</th>
<th>Indel</th>
<th>Multi-allelic</th>
<th>SNV</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC1</td>
<td>974,836 (7.9%)</td>
<td>0 (0%)</td>
<td>11,375,702 (92.1%)</td>
<td>12,350,538 (41.7%)</td>
</tr>
<tr>
<td>AC2</td>
<td>258,462 (8.1%)</td>
<td>54,546 (1.7%)</td>
<td>2,870,739 (90.2%)</td>
<td>3,183,747 (10.7%)</td>
</tr>
<tr>
<td>MAF &lt; 1%</td>
<td>267,166 (8.1%)</td>
<td>74,313 (2.2%)</td>
<td>2,968,917 (89.7%)</td>
<td>3,310,396 (11.2%)</td>
</tr>
<tr>
<td>1% &lt; MAF &lt; 5%</td>
<td>299,227 (8.0%)</td>
<td>125,840 (3.4%)</td>
<td>3,297,232 (88.6%)</td>
<td>3,722,299 (12.6%)</td>
</tr>
<tr>
<td>MAF &gt; 5%</td>
<td>547,164 (7.8%)</td>
<td>257,000 (3.6%)</td>
<td>6,250,382 (88.6%)</td>
<td>7,054,546 (23.6%)</td>
</tr>
<tr>
<td><strong>Summary</strong></td>
<td><strong>2,346,855 (7.9%)</strong></td>
<td><strong>511,699 (1.7%)</strong></td>
<td><strong>26,762,972 (90.3%)</strong></td>
<td><strong>29,621,526</strong></td>
</tr>
</tbody>
</table>

Table S1. Number of variants and minor allele frequency.
<table>
<thead>
<tr>
<th>Variation</th>
<th>MAF</th>
<th>LoF</th>
<th>Moderate</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indel</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC1</td>
<td>1,315</td>
<td>926</td>
<td>518</td>
<td></td>
</tr>
<tr>
<td>AC2</td>
<td>228</td>
<td>185</td>
<td>130</td>
<td></td>
</tr>
<tr>
<td>MAF &lt; 1%</td>
<td>218</td>
<td>171</td>
<td>114</td>
<td></td>
</tr>
<tr>
<td>1% &lt; MAF &lt; 5%</td>
<td>195</td>
<td>177</td>
<td>117</td>
<td></td>
</tr>
<tr>
<td>MAF &gt; 5%</td>
<td>299</td>
<td>243</td>
<td>233</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>2,255</td>
<td>1,702</td>
<td>1,112</td>
<td></td>
</tr>
<tr>
<td>SNV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC1</td>
<td>1,647</td>
<td>41,213</td>
<td>31,268</td>
<td></td>
</tr>
<tr>
<td>AC2</td>
<td>236</td>
<td>7,981</td>
<td>7,462</td>
<td></td>
</tr>
<tr>
<td>MAF &lt; 1%</td>
<td>244</td>
<td>7,671</td>
<td>7,484</td>
<td></td>
</tr>
<tr>
<td>1% &lt; MAF &lt; 5%</td>
<td>233</td>
<td>7,448</td>
<td>7,996</td>
<td></td>
</tr>
<tr>
<td>MAF &gt; 5%</td>
<td>310</td>
<td>9,863</td>
<td>13,326</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>2,670</td>
<td>74,176</td>
<td>67,536</td>
<td></td>
</tr>
</tbody>
</table>

Table S2. Coding mutation distribution according to minor allele frequency and impact on gene product. LoF, loss of function.