**Supplementary Text**

**Methods:**

**DNA Extraction, PCR, and Multitag Sequencing**

DNA was extracted from samples using FastDNA spin kit for Soil (MP Biomedicals) according to manufacturer’s protocol with slight modifications. The soil kit was selected for both fecal and saliva samples extractions because this bead beating methodology extracts DNA from all microbial cells in the sample. The DNA was diluted with ultra pure water (1:5-1:10 ratio) to use in PCRs.

Approximately 10 ng of DNA was used in a 20 μl reaction for PCR amplification. AmpliTaq Gold™ DNA Polymerase (Thermo Fisher Scientific, Waltham, Massachusetts, USA) was used for the amplification. Master mix for the 16S rRNA gene amplification was subjected to ultraviolet (UV) light (Stratagene Stratalinker®) prior to the addition of primers and dNTPs to eliminate possible bacterial DNA contamination. UV exposure was titrated to ensure no impact on enzyme activity. Universal 16S rRNA bacterial primers 27F and 357R (Lane, 1991) targeting variable regions 1 and 2, were used for bacterial identification. A standard reaction was performed using a GeneAmp 9700 thermocycler (Applied Biosystems Inc.) with 11 minutes of initial denaturation at 95°C, followed by 32 cycles of 30sec at 95°C, 30sec at 48°C, and 2 minutes at 72°C with added 5sec/cycle, and one cycle of 30 minutes at 72°C and it was hold at 4°C. The longer initial denaturation time is recommended by manufacturer to use with Taq Gold polymerase. This enzyme is attached to an antibody and needs to be activated at 95°C with a longer incubation time. This helps with longer reaction preparations and to avoid producing non-specific products. Most polymerases as well as Taq Gold add an A base to the end of PCR product and the longer (30 minutes) extension time ensures adding an A to all amplified fragments to avoid the size difference when examining the product length in fingerprints which might be only one base different.

*Escherichia coli (E.coli)* DNA was used as a positive PCR control and no DNA was used as a negative control for all PCRs. Duplicate LH-PCRs (Length Heterogeneity PCR) were completed and fingerprinted prior to sequencing in order to select the most consistent products to represent the samples for sequencing. This quality control step used fusion primers for the bacterial 16S rRNA gene (27F and 357R). Fusion primers contain an adapter joined to an 8 base “barcode”, as well as the appropriate primers. The reverse primer (357R) was FAM labelled on the 5-prime end. The fingerprint was run on an ABI 3130*xl* Fluorescent Sequencer (Applied BioSystems). PCR products were selected based on the fingerprints and pooled. The pool was purified twice (to ensure elimination of primer dimers and short products) with Agencourt AMPure solution (Beckman Coulter) in preparation for sequencing. The purified product was quantified using a DTX880 Multimode Fluorescent detector (Beckman Coulter) and the correct concentration was calculated to use in emulsion PCR prior to sequencing. We used Ion Torrent technology (Thermo Fisher Scientific) with the Personal Genome Machine (PGM) for high-throughput sequencing. All emulsion PCR and sequencing steps were executed using the kits and manufacturer’s protocols for the PGM. A customized PERL script was used to “demultiplex” raw sequence data from each pooled sample and to separate the sequences into individual samples based on the barcodes used for each sample at initial PCR.

**Table S1**: MAAsLin2 with National Percentile as the dependent variable shows none of the factors pass FDR correction (q-value)

|  |  |  |  |
| --- | --- | --- | --- |
| **National Percentile** | **Neighborhood disadvantage (High or low)** | **p-value** | **q-value** |
| **Male sex** | High | 0.00387 | 0.474963 |
| *Lachnospiraceae\_family* | High | 0.004106 | 0.474963 |
| *Coprococcus* | High | 0.00196 | 0.474963 |
| *Leuconostoc* | High | 0.004185 | 0.474963 |
| *Fusobacterium* |  Low | 0.005779 | 0.524714 |
| *Massiliprevotella* | High | 0.007296 | 0.552096 |
| *Negativibacillus* | High | 0.009419 | 0.568691 |
| *Allisonella* | High | 0.010021 | 0.568691 |
| **Education** | Low | 0.012556 | 0.586841 |
| *Ruminococcus* | High | 0.01477 | 0.586841 |
| *Kineothrix* | High | 0.015511 | 0.586841 |
| *Butyrivibrio* | High | 0.012954 | 0.586841 |
| *Colidextribacter* | High | 0.022226 | 0.711645 |
| *Papillibacter* | Low  | 0.02049 | 0.711645 |
| *Lactonifactor* | High | 0.028729 | 0.815198 |
| *Enterococcus* | High | 0.030847 | 0.823789 |
| *Butyricicoccus* | High | 0.036044 | 0.883047 |
| *Oscillospiraceae\_family* | High | 0.037865 | 0.883047 |
| *Desulfovibrio* | High | 0.038901 | 0.883047 |
| *Prevotella* | High | 0.040923 | 0.884721 |

|  |  |  |  |
| --- | --- | --- | --- |
| **State Decile** | **Neighborhood disadvantage (High or low)** | **p-value** | **q-value** |
| *Lachnospiraceae\_family* | High | 4.49E-04 | 0.203923 |
| **Male sex** | High | 0.012995 | 0.737472 |
| *Rothia* | High | 0.009837 | 0.737472 |
| *Oscillospiraceae\_family* | High | 0.010007 | 0.737472 |
| *Massiliprevotella* | High | 0.011724 | 0.737472 |
| *Catenibacterium* | High | 0.004529 | 0.737472 |
| *Coprococcus* | High | 0.006877 | 0.737472 |
| *Allisonella* | High | 0.005679 | 0.737472 |
| **Education** | Low | 0.024943 | 0.928815 |
| **Age** | High | 0.019805 | 0.928815 |
| *Dorea* | High | 0.026596 | 0.928815 |
| *Negativibacillus* | High | 0.023289 | 0.928815 |
| *Longicatena* | High | 0.026481 | 0.928815 |
| *Eubacterium* | High | 0.031068 | 0.949255 |
| *Coprobacter* | Low  | 0.031363 | 0.949255 |

**Table S2:** MAAsLin2 with State Decile as the dependent variable shows none of the factors pass FDR correction (q-value)

**Supplementary figure 1**: No significant changes between ADI rankings at state and

P=0.48

**b**

**a**

P=0.35

national levels and MELD score based on deciles