**Data description**

how the study was designed? what purpose? how? why? methodology? collection procedure? measurements.

**Study background:** Bacterial infection has long been implicated with rheumatoid arthritis (RA), but a systematic analysis of the probiotic and pathogenic microbiome in RA has been lacking.

**Why conduct this research:** It is not clear whether and how the gut or oral microbial community is compositionally and functionally altered in RA, and whether and how the microbiota at different body sites overlap.

**How:** Here we perform whole genome shotgun sequencing for fecal, dental, and salivary samples from a large cohort of RA patients, analyze metagenomic linkage groups to construct an RA classifier.

**Methodology:** CCA, spearman correlation, Random forest model, Wilcoxon rank-sum test etc.

**Data collection:** Patients were diagnosed at Peking Union Medical College Hospital on the basis of the 2010 American College of Rheumatology and EULAR classification criteria. Healthy controls were those who meet the healthy criteria.

Oral samples included dental plaques and saliva.

**Metagenomic sequencing:** paired-end metagenomic sequencing was done on the Illumina platform**. (2015)**

**Metagenomic sequencing:** paired-end metagenomic sequencing was done on BGISeq-500. (2017)

**How to get mOTU profile:** After regular quality control (trim reads of low quality and remove host reads), then processing the clean reads with mOTU2 (a software) to get mOTU profile.

**Notes**:

1, we re-sequenced 293 samples which were used in the study using BGISeq-500 instead of Illumina.