Methods

1.1 Sample Quality Control

Please refer to Novogene's QC report for methods of sample quality control.

1.2 Library Construction, Quality Control and Sequencing

Sampling 200 ng of genomic DNA, the sample is randomly fragmented into segments of approximately 350 bp using a Covaris ultrasonic disruptor to construct the library. The entire library preparation is completed through steps including end repair, addition of A-tails, ligation of sequencing adapters, purification, and PCR amplification. After library construction, the integrity of the library fragments and the size of the inserted fragments are assessed using AATI analysis. If the insert size meets expectations, the accurate concentration of the effective library is quantified using Q-PCR (effective library concentration > 3 nM) to ensure the library quality. After the library passes the quality check, different libraries are pooled according to their effective concentrations and target data output requirements, and then subjected to PE150 sequencing.

2 Bioinformatics Analysis Pipeline

Metagenomic samples were quality-controlled and trimmed for adaptors using fastp (https://github.com/OpenGene/fastp). Considering the possibility of host contamination in samples, clean data needs to be blasted to the host database to filter out reads that may come from the host origin. Bowtie2 software (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml) is used by default, with the following parameter settings: --end-to-end, --sensitive, -I 200, and - X 400 (Karlsson FH et al.,2012; Karlsson FH et al., 2013; Scher JU et al., 2013). Using Kraken 2 (Lu J et al., 2022), microbial sequences (including bacteria, archaea, fungi, and viruses) were compared against a database built with kraken2-build and the species annotation results were corrected using Braken (Lu J et al., 2022). Diamond was used to align sequences against the UniProt database (https://www.uniprot.org/), and the annotation results for various functional databases, including KEGG, GO, eggNOG, and Pfam, were obtained based on the correspondence with UniRef90 IDs and other functional databases.

Based on the species or functional data at each taxonomy level, alpha diversity analysis is performed within samples, including the calculation of alpha index (R vegan package) , significant box plots of alpha indexes (R ggplot package) , and rarefaction curves (R vegan package). Beta diversity analysis is performed between samples, including relative abundance overview and abundance clustering heatmap analysis, combined with PCA (R ade4 package) (Rao CR et al., 1964), PCoA (R ade4 package), and NMDS (R vegan package) dimension reduction analysis (Legendre P, 1998). Anosim analysis (R vegan package) is used to test the differences between groups. MetaGenomeSeq and LEfSe analysis are used to search for species or functional differences between groups. MetaGenomeSeq analysis is used to perform permutation tests between groups on each taxonomy level and to obtain a p-value. LEfSe software is used for LEfSe analysis (Segata N et al., 2011).

3 Reference

Blanco-Míguez A, Beghini F, Cumbo F, et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. Nat Biotechnol. 2023;41(11):1633-1644. doi:10.1038/s41587-023-01688-w

Franzosa EA, McIver LJ, Rahnavard G, et al. Species-level functional profiling of metagenomes and metatranscriptomes. Nat Methods. 2018;15(11):962-968. doi:10.1038/s41592-018-0176-y

Karlsson FH, Fåk F, Nookaew I, et al. Symptomatic atherosclerosis is associated with an altered gut metagenome. Nat Commun. 2012;3:1245. doi:10.1038/ncomms2266

Karlsson FH, Tremaroli V, Nookaew I, et al. Gut metagenome in European women with normal, impaired and diabetic glucose control. Nature. 2013;498(7452):99-103. doi:10.1038/nature12198

Legendre P, Legendre L. Numerical Ecology. 2nd ed. Elsevier; 1998.

Paulson JN, Stine OC, Bravo HC, Pop M. Differential abundance analysis for microbial marker-gene surveys. Nat Methods. 2013;10(12):1200-1202. doi:10.1038/nmeth.2658

Rao CR. The Use and Interpretation of Principal Component Analysis in Applied Research. Sankhy ā : The Indian Journal of Statistics, Series A. 1964;26(4):329-358.

Scher JU, Sczesnak A, Longman RS, et al. Expansion of intestinal Prevotella copri correlates with enhanced susceptibility to arthritis. Elife. 2013;2:e01202. doi:10.7554/eLife.01202

Segata N, Izard J, Waldron L, et al. Metagenomic biomarker discovery and explanation. Genome Biol. 2011;12(6):R60. doi:10.1186/gb-2011-12-6-r60