Metagenomic Analysis of Type II Diabetes Gut Microbiota Using PacBio HiFi Reads Reveals Taxonomic and Functional Differences

Daniel Portik1, Joan Wong1, Khi Pin Chua1, Baoli Zhu2, Na Lu3, Fei Liu4, Xiaofeng Song2, Yong Yue5, Weigang Zhao5, Shixuan Liu6, Tao Yuan5, Yong Fu5, Gregory Young5, Richard Hall1

1Pacific Biosciences, Menlo Park, CA, 2Chinese Academy of Sci., Beijing, China, 3Chinese Ctr. for Disease Control and Prevention, Beijing, China, 4China Agricultural Univ., Beijing, China, 5Peking Union Med. Coll. & Chinese Academy of Med. Sci., Beijing, China

Introduction

In the past decade, the human microbiome has been increasingly shown to play a major role in health. For example, imbalances in gut microbiota appear to be associated with Type II diabetes mellitus (T2DM) and cardiovascular disease (CAD). Metagenomic analysis of the long-term gut microbiota suggests that microbial changes, such as increased short chain fatty acids (SCFAs), can lead to metabolic dysfunctions and cardiovascular disease. However, the exact microbial taxa or functions implicated in disease need further investigation.

From a large study with 523 participants (185 healthy controls, 186 T2DM patients without CAD, and 152 T2DM patients with CAD), 3 samples from each patient group were selected for long-read sequencing. Each sample was prepared and sequenced on the Sequel II System SMRT Cell. To assess whether long accurate PacBio HiFi reads could yield additional insights to those made using short reads.

Each of the 9 samples was subjected to metagenomic assembly and binning, taxonomic classification and functional profiling. Results from metagenomic assembly and binning showed that it is possible to generate a significant number of complete MAGs (Metagenomic Assembled Genomes) from each sample, with over half of the high-quality MAGs being represented by a single circular contig. We show that functional differences in T2DM versus healthy gut microbiota.

Assembly

HiFi reads for each sample were assembled using Canu (v1.8) with settings specific to PacBio HiFi data.

Data

Sampling Design and HiFi Data Yields

Three samples in each group:
- Healthy controls
- Type II diabetes + cardiovascular disease (CAD)
- Type II diabetes only

Each sample sequenced on one Sequel II System SMRT Cell 8M reads

Figure 1. The number of HiFi reads and the average HiFi read length for all 9 samples. Typical yields were 2.0–2.5 Mreads, with mean read lengths ~8–10 kb.

Figure 2. Summary of the N50 length and number of assembled contigs per sample. The N50 length and the number of assembled contigs vary considerably across samples. However, most samples had ~40–50 contigs that were greater than ~1Mbp in length.

Figure 4. Example of results for genome binning for a control sample (Figure 1A). Each point represents a unique high-quality genome bin containing one contig (blue). These represent high-quality assembled genomes.

Figure 5. The SEED viewer in MEGAN6 showing a hierarchical tree of major functional classes. The representation of the different classes is shown for each sample in the bar charts, including samples in the control group (blue), Type II diabetes + CAD (green), and Type II diabetes (purple). Functional classes can be expanded to examine representation in various subcategories.

Figure 7. A visualization of the differences in the abundance of bacterial taxa across samples is shown in A). A principle component analysis of samples based on their taxonomic profiles is shown in B). The bacterial taxa contributing the most to the variation are shown as vectors indicating the direction of the steepest increase. Table C details the 7 species with the most significant change in abundance between the healthy and diabetic patient groups.

Conclusion

The ability to generate high-quality MAGs together with taxonomic and functional data demonstrates the high value of HiFi datasets in metagenomic analysis.

References