Performance evaluation of the G4 sequencing platform for microbiome community analysis

Timothy Looney (timl@singulargenomics.com), Singular Genomics, **Timothy Looney**, Singular Genomics, **Ann Tong**, Singular Genomics, **Chrystal Day**, Singular Genomics, **Elanor Sievers**, Singular Genomics, **Sabrina Shore**, Singular Genomics, **Bubba Brooks**, Singular Genomics, **Jiangwei Yao**, Singular Genomics, **Martin Fabani**, Singular Genomics

Background

A key challenge to next-generation sequencing (NGS)-based microbiome analysis is the presence of high and low GC content organisms, which are traditionally underrepresented in NGS data. Here we evaluate the performance of the Singular Genomics G4 Platform for microbial community analysis by sequencing of a National Institute for Standards and Technology (NIST) microbial reference material comprising low and high GC organisms, comparing results to those from the Illumina NextSeq 2000.

Methods

Whole genome sequencing libraries were prepared in triplicate for the G4 and Illumina NextSeq 2000 platforms from 200 ng Covaris-sheared gDNA comprising an equimolar pool of 19 microbial taxa spanning broad GC and phylogenetic diversity (NIST Reference Material 8376; QuantaBio sparQ DNA Library Prep kit used for library preparation). Libraries were sequenced via the G4 with F2 flow cell or the NextSeq 2000 with P1 flow cell with 2x150bp read format. Data was downsampled to 1M reads prior to analysis via Kraken v2.1.2, Bracken v2.6.2 and metaphlan4.

Results

QC analysis indicated high quality reads for the G4 Platform (90% >=Q30 basecalls; raw accuracy 99.88%). Estimated species abundance closely matched expectation, with 18 of 19 species within 2-fold of expected abundance values across both platforms. No significant correlation between GC content and species representation was observed for either platform (p-value 0.58 and 0.69 for G4 and Illumina replicates, respectively; cross-platform correlation = 0.984). The G4 data showed a higher correlation to expected abundance values than the Illumina data (R2 = 0.93 and 0.89 for G4 and NextSeq datasets respectively).

Conclusions

We observe accurate and reproducible quantification of low and high GC microorganisms using the G4 platform, with results highly correlated to those from the Illumina platform. We expect the speed and flexible throughput of the G4 platform will help reduce turnaround times in basic and translational microbiome research.