Abstract

Background
The sequencing and haplotype phasing of entire gene sequences improves the understanding of the genetic basis of disease and drug response. One example is cystic fibrosis (CF). CFTR (cystic fibrosis transmembrane conductance regulator) modulator therapy has revolutionized CF treatment, but only in a minority of CF subjects. Observed heterogeneity in CFTR modulator efficacy is related to the range of CFTR mutations; revertant mutations can modify the response to CFTR modulators, and other intronic variations in the ~200 kb CFTR gene have been linked to disease severity. Heterogeneity in the CFTR gene may also be linked to differential responses to CFTR modulators.

Experimental Design and Methods
The TLA process produced amplicons consisting of 5-10 proximity ligated DNA fragments. TLA was performed on cell line and genomic DNA from Coriell GM12878, which has few heterozygous SNVs in CFTR, and the IB3 cell line, with known haplotypes but heterozygous for the delta508 mutation. All sample types were prepared with high and low density TLA primer sets, targeting coverage of >100 kb of the (~200 kb) CFTR gene.

Conclusion
We demonstrate the power and utility of TLA with long-read SMRT Sequencing as a valuable research tool in sequencing and phasing across very long regions of the human genome. This process can be done in an efficient manner, multiplexing multiple genes and samples per SMRT Cell in a process amenable to high-throughput sequencing.

TLA Technology

Workflow

TLA Amplicons of varying length and consisting of different DNA sequences originating from the same allele are the input material for SMRTbell library preparation.

Bioanalyzer QC of TLA samples indicates a broad range of fragment sizes, some >10 kb.

One SMRTbell library was prepared from 10 samples using barcoded adapters, following standard protocols.

Long read sequencing on Sequel System generated >300,000 reads

Results from one Sequel 1M cell:

Results from 1 SMRT Cell 1M

Conclusions
We obtained >200,000 high-accuracy CCS reads from TLA amplicons on one Sequel SMRT Cell 1M, with an average read quality of Q44.

We have demonstrated the use of Cergentis’ TLA technology combined with PacBio high accuracy long read sequencing to enable phasing across large physical distances and the targeted efficient sequencing and phasing of entire genes.