HIGH DATA THROUGHPUT AND LOW COST ULTRA LONG NANOPORE SEQUENCING

Duncan Kilburn1, Jeff Burke1, Renee Fedak2, Michelle Kim1, Hugh Olsen3, Miten Jain3, Simon Mayes3, Kelvin Liu3
1Circulonics Inc, Baltimore, MD; 2UC Santa Cruz Genomics Institute, Santa Cruz, CA; 3Oxford Nanopore Technologies, Oxford, UK

**Nanobind UHMW Extraction**

- Ultra High Molecular Weight (UHMW), megabase DNA was extracted from human cells, bacteria, and human whole blood using the Circulonics Nanobind CBB Big DNA Kit. Tissues were extracted using the Circulonics Nanobind Tissue Big DNA Kit.

**Nanobind UL Library Prep**

- Library prep purification was performed using the Circulonics Nanobind UL Library Prep Kit, NB-900-601-01, and the Oxford Nanopore Ultra Long DNA Sequencing Kit, SQK-UL01D.
- Nanobind method can purify megabase DNA without shearing it or damaging the sequencing motor protein.

**Extraction Results**

- A new UHMW extraction protocol and modified chemistry were developed to further increase size and reduce sample viscosity.
- UHMW, megabase-sized DNA can be identified on PFGE images by 1) streaking up to 1+ Mb and 2) compression zone banding.

**Nanobind-Enhanced Ultra Long Library Preparation**

- UHMW UL sequencing protocol was beta tested on 30 genomes at UCSC as part of the human pan-genome reference project. Each cell line was sequenced on 3 PromethION flowcells. N50s were between 61 kb (HG02148) and 93 kb (HG03453).
- New method uses 10 – 100X fewer flow cells to generate comparable UL coverage, dramatically decreasing cost, time, and sample consumption for both large and small projects.

**Oxford Nanopore Sequencing Metrics**

- 150
- 120
- 90
- 60
- 30
- 0
- 0
- 1000
- 2000
- 3000

**Multiple sample types**

- Multiple sample types have generated ultra long sequencing data using current protocol.
- Tissue samples and gram-positive bacteria generate N50s in line with the best LSK109 runs, and the amount of data > 200kb is up to an order of magnitude higher.

**High Coverage Ultra Long Reads**

- Previous record read of 2.27 Mb had been in place since 2018.
- New approach has led to continuous improvements in longest read.
- Since Feb 2020, record read length has been broken 3 times.
- Current record is a 4.15 Mb continuous read mapping to chr3 generated on PromethION.
- Modified version of minKNOW was used to reduce artificial splitting of reads.
- Inexpensive, high coverage ultra long reads can be used to help assemble and phase difficult regions with repetitive sequences.
- Shown on right is data from a single MiniION run that generated 9 Gb with N50 of 87 kb.
- Representative reads shown in the MHC region.
- Ongoing work will test suitability of UL reads to fully phase this region.

**Multiple sample types**

<table>
<thead>
<tr>
<th>Sample Type</th>
<th>Sample</th>
<th>Sequence</th>
<th>Total Data</th>
<th>Data &gt; 100kb</th>
<th>Data &gt; 200kb</th>
<th>Read Length</th>
<th>Max Read Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mammalian cells</td>
<td>HG02148</td>
<td>PromethION</td>
<td>95</td>
<td>45</td>
<td>17</td>
<td>93</td>
<td>3.7</td>
</tr>
<tr>
<td>Mammalian cells</td>
<td>HG02723</td>
<td>PromethION</td>
<td>95</td>
<td>45</td>
<td>17</td>
<td>93</td>
<td>3.7</td>
</tr>
<tr>
<td>Blood</td>
<td>Bovine</td>
<td>PromethION</td>
<td>95</td>
<td>45</td>
<td>17</td>
<td>93</td>
<td>3.7</td>
</tr>
<tr>
<td>Gram-negative bacteria</td>
<td>E. coli</td>
<td>PromethION</td>
<td>77</td>
<td>47</td>
<td>22</td>
<td>128</td>
<td>3.3</td>
</tr>
<tr>
<td>Gram-positive bacteria</td>
<td>L. monocytogenes</td>
<td>PromethION</td>
<td>77</td>
<td>47</td>
<td>22</td>
<td>128</td>
<td>3.3</td>
</tr>
<tr>
<td>Tissue (in development) Bovine Fetal Lung</td>
<td>PromethION</td>
<td>38</td>
<td>8.9</td>
<td>1.9</td>
<td>51</td>
<td>0.9</td>
<td></td>
</tr>
</tbody>
</table>

**High Coverage Ultra Long Reads**

- 4.15 Mb read
- 2% of Chromosome in single read
- 1 hours in pore

- 400 Gb x