Assembling the human genome using long nanopore sequencing reads

To gain a comprehensive insight into human genetic variation, and its potential impact on disease risk, it is important to obtain fully characterised, complete genomes. However, the presence of large structural variants (SVs) and repeat sequences have posed a significant challenge to assembling the human genome to completion.

Unlike short-read sequencing, nanopore sequencing produces long and ultra-long reads, which enhance the resolution of SVs and repeats. With the high-output PromethION™ device, sequencing and assembling highly contiguous human genomes is now possible, with unprecedented efficiency.

Here we present a simple workflow for human genome assembly from a blood sample, using the PromethION sequencing platform.

**EXTRACTION:** obtaining high molecular weight DNA

Selecting a suitable extraction method for obtaining high molecular weight DNA greatly depends on sample type. For DNA extraction from whole blood, we recommend the QIAGEN Gentra Puregene Blood Kit, which we have found maximises the production of long sequencing reads.

**LIBRARY PREPARATION:** selecting a kit

Enriching for long and ultra-long (≥50 kb) gDNA fragments is important for performing genome assembly, to maximise the overlap of sequencing reads in analysis. We recommend size selection and light shearing of the extracted gDNA, which we have found to improve read length N50 — we suggest aiming for a read N50 of 25–35 kb. Internal testing has yielded good results using the Circulomics Short Read Eliminator Kit for size selection, to select for all fragments >10 kb, and the Diagenode Megaruptor 3 for shearing.

To prepare gDNA for sequencing, we recommend the Ligation Sequencing Kit (SQK-LSK110), providing the greatest throughput and control over read lengths.
For assembly, we recommend sequencing a human genome to a minimum depth of 30x of 25–35 kb reads; this can be achieved by sequencing a single genomic sample on one PromethION Flow Cell for 72 hours, after following the extraction and library preparation protocols above. However, sequencing to a depth of 60x is advisable to obtain the best assembly metrics. We also recommend basecalling in high accuracy mode. Throughput can be maximised by performing a nuclease flush and loading fresh library every 24 hours.

The on-demand PromethION sequencing platform has the capacity to run up to 24 (P24) or 48 (P48) flow cells at any time, providing ultimate flexibility and adaptability to your sequencing requirements.

*Circulomics Short Read Eliminator Kit + Diagenode Megaruptor 3

Oxford Nanopore Technologies and the Wheel icon, and PromethION are registered trademarks of Oxford Nanopore Technologies Limited in various countries. All other brands and names contained are the property of their respective owners. © 2021 Oxford Nanopore Technologies Limited. All rights reserved. Oxford Nanopore Technologies products are not intended for use for health assessment or to diagnose, treat, mitigate, cure, or prevent any disease or condition.

WF_1066(EN)_V2_03Jun2021

References:

Find out more about nanopore sequencing service providers: nanoporetech.com/services/providers

Find out more at: nanoporetech.com/assembly