HUMAN GENOMICS

APPLICATION NOTE

Optimising read length for human genome sequencing and analysis

INTRODUCTION

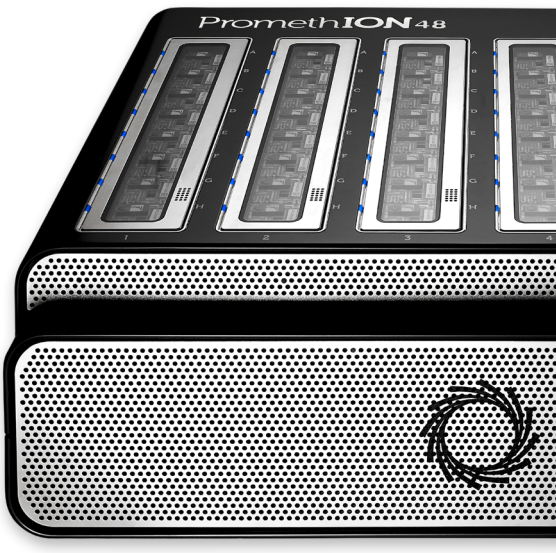
The ability to obtain sequencing reads of any length with nanopore technology, whilst maintaining base modification information, provides a unique opportunity to access the human genome in a way that has not been previously possible. Such comprehensive characterisation of the human genome, from detecting single nucleotide variation (SNV) and nucleotide-level methylation differences to calling large-scale structural variants (SVs) and chromosomal rearrangements, requires careful consideration of the end-to-end experimental workflow — from sample preparation to sequencing and data analysis.

The high-throughput, modular Oxford Nanopore PromethION™ benchtop sequencer is ideal when it comes to human genome analysis, providing the capacity to run up to 24 (P24) or 48 (P48) flow cells at any time, allowing unprecedented flexibility for users' experimental requirements.

Objectives

The aim of this Application note is to familiarise users with different sample and library preparation options available for human whole-genome sequencing, based on the combination of Circulomics and Oxford Nanopore Technologies kits and workflows, and the typical sequencing metrics that are produced from each.

With the inclusion of application-based recommendations, this Application note also intends to highlight key experimental considerations at each stage of a genome sequencing workflow.



Key experimental considerations

When planning a sequencing run, there are a number of important considerations, including:

1. Analysis requirements

Anticipated bioinformatics needs will drive sequencing choices; these will be translated into sequencing data requirements:

a) Sequencing depth of coverage

b) Read length distribution desired

2. Sample limitations

Any constraints imposed by the sample to be sequenced will also affect experimental planning:

a) Type of sample: the availability of protocols for high molecular weight (HMW) and ultra-high molecular weight (UHMW) DNA extraction from the sample

b) Amount of sample: some samples are scarce, limiting the total DNA available

3. Application-based considerations

Enriching for long and ultra-long fragments is important to maximise the overlap of sequencing reads in the analysis, enabling the resolution of repetitive sequences and large SVs, plus enhancing the phasing of variants and modification calls.

For phasing, the relationship between increasing read length and phase-block length is not linear. As can be seen from **Table 1**, small increases in read length can lead to large gains in phase-block length.

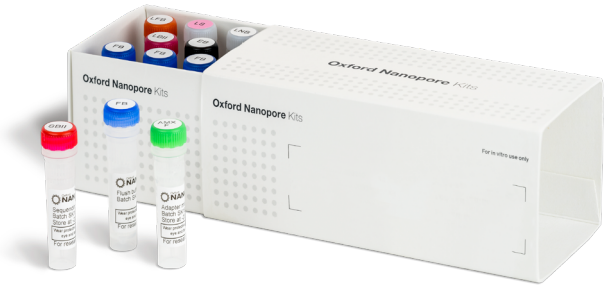
Table 1: Read length greatly influences phasing performance

Read N50	Shearing method	Phase block N50†
11 kb	g-TUBE	80–200 kb
17 kb	SRE + MR-3*	150–300 kb
25 kb	Unfragmented	1–2.5 Mb
26 kb	SRE + MR-3*	500–650 kb
32 kb	SRE + MR-3*	1–2.5 Mb
36 kb	SRE + MR-3*	1.5–3 Mb
97 kb	Unfragmented	13 Mb+

*Circulomics Short Read Eliminator Kit + Diagenode Megaruptor 3
†Established using data for human chr21 (total length 46.7 Mb)

Human genome assembly with/without genome-wide structural variation calling

The combination of the Ligation Sequencing Kit (SQK-LSK110) with size selection and light shearing provides users with control over read lengths and sequencing throughput; more specifically, aiming for a read N50 of 25–35 kb is recommended for human genome assembly and genome-wide SV calling.



Genome-wide methylation or SNV calling

A greater throughput, and so higher depth of coverage, can be achieved from a sequencing run by implementing a higher degree of shearing, suitable for when the experimental aim is genome-wide methylation or SNV calling, without phasing.

Scaffolding and resolving challenging regions

The new Ultra-Long DNA Sequencing Kit (SQK-ULK001) (ULK) is ideal for resolving long repeats and large-scale structural variants, making it well-suited to accessing highly challenging areas of the genome, such as repeat-rich centromeres and telomeres. The kit is the optimal solution for closing gaps in human genome assemblies and obtaining longer phase-blocks for long-range haplotype resolution. The ULK is also a good option for researchers aiming to scaffold genomes and produce initial draft genome assemblies.

Sample and library preparation

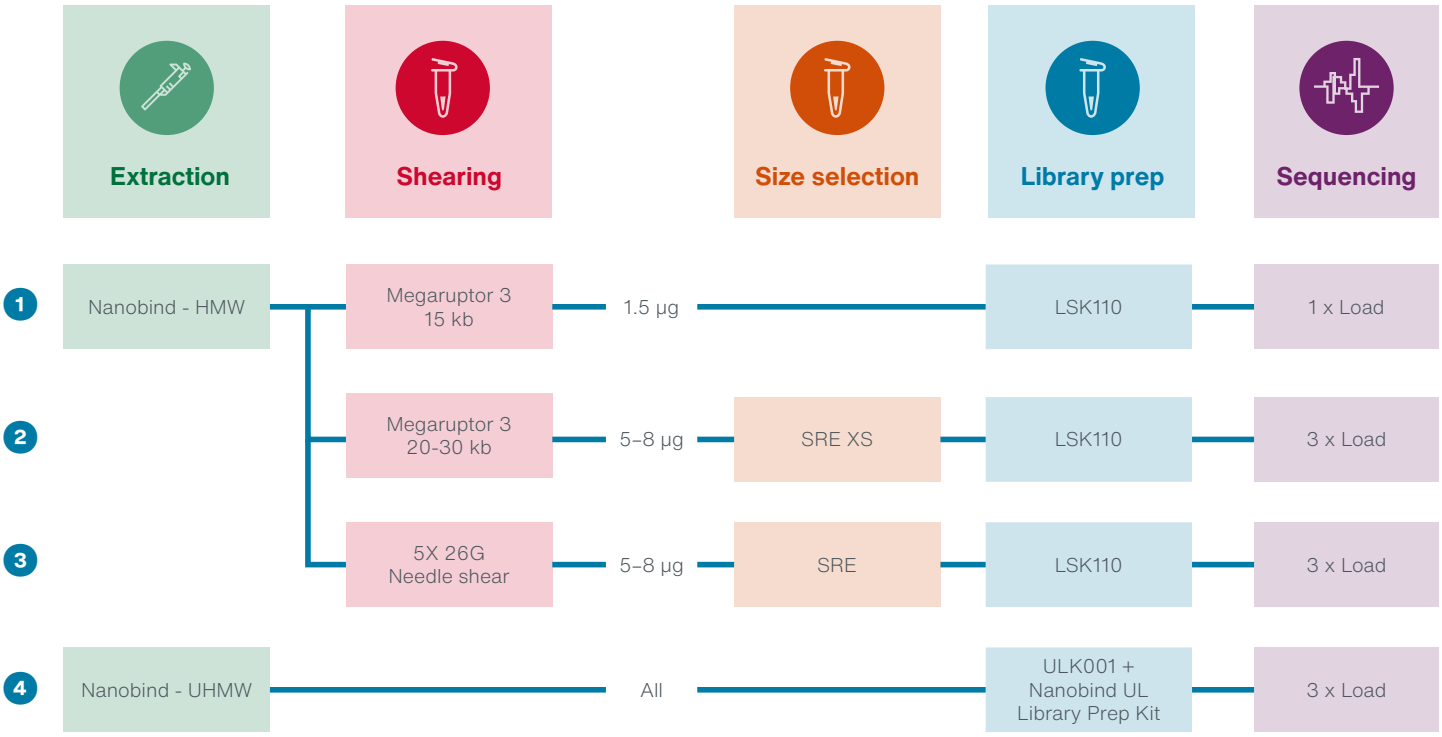
A single sample of fresh human blood was stored in K2/EDTA anticoagulant at 4°C for 24 hours prior to DNA extraction.

HMW and UHMW DNA extractions were performed using Circulomics Nanobind Kits and associated protocols (see Appendix). DNA was then sheared with the Diagenode Megaruptor 3 (MR-3) (50 ng/µl sample volume) or a needle; size selection was then performed for experimental conditions 2 and 3 using a Short Read Eliminator (SRE) kit. Note that size selection post shearing slightly improves sequencing performance; selection performed prior to shearing allows for greater sample recovery.

Library preparation was carried out with either the Oxford Nanopore Ligation Sequencing Kit (LSK110) or Ultra-Long DNA Sequencing Kit (ULK001), prior to PromethION sequencing on a single flow cell per condition for a total of 72 hours. A nuclease flush and library reload every 24 hours was performed for conditions 2–4, as is recommended for libraries containing longer DNA (≥~20 kb), to optimise sequencing throughput.

The flow chart below in **Figure 1** demonstrates the experimental workflows used, from sample preparation, including shearing and size selection, through to library preparation and sequencing.

Figure 1: Experimental workflows for each of the four conditions tested
The DNA inputs for size selection (conditions 2 and 3) or library preparation (conditions 1 and 4) are also given.



Sequencing results

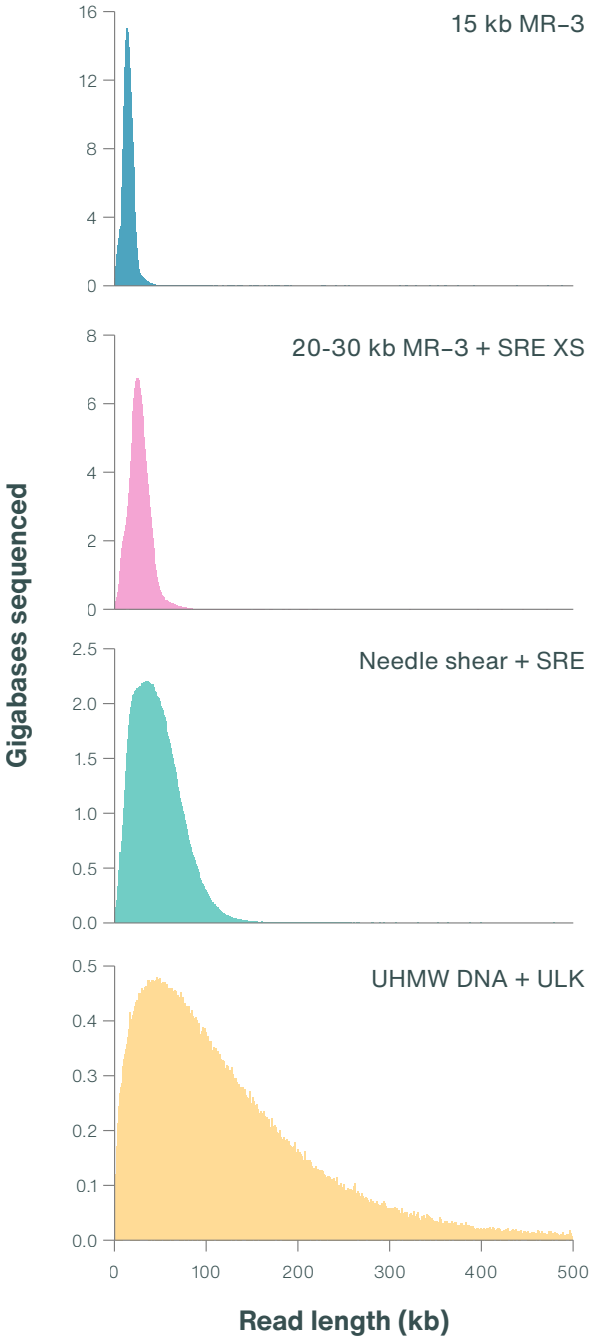
The sequencing yields obtained for each of the four experimental conditions are provided in **Table 2** and read length distributions are displayed in **Figure 2**. As expected, the longest reads were obtained for condition 4, incorporating UHMW DNA extraction and library preparation with the Ultra Long DNA Sequencing Kit. Sequencing yield was greater for libraries containing shorter DNA fragments; due to the nature of UHMW DNA, the yield was lower for condition 4.

Table 2: Sequencing metrics for all reads obtained from each of the four sequencing runs

As detailed in the previous section, one PromethION Flow Cell was run per library.

	1 15 kb MR–3	2 20-30 kb MR–3 + SRE XS	3 Needle shear + SRE	4 UHMW DNA + ULK
Throughput (Gb)	188	160	143	83
Read N50 (kb)	15	26	43	101
Gb > 100 kb	0	0	5	42

Figure 2: Read length distributions for sequencing data produced from each of the four conditions



Summary

A number of considerations are required at each stage of a human genome sequencing experimental workflow: from extraction, through to library preparation, sequencing, and data analysis. Different combinations of Circulomics and Oxford Nanopore Technologies sample and library preparation workflows are available, taking advantage of the unique benefits that nanopore sequencing technology affords, namely the ability to obtain reads of any length, and the simultaneous detection of base modifications.

Furthermore, PromethION provides unprecedented flexibility and scalability in sample throughput.

As demonstrated, read length and yield can be adjusted and adapted according to the experimental aims, enabling resolution of large-scale structural variants, closure of assembly gaps, and phasing of variant and modification calls — providing a truly comprehensive analysis of the human genome.

View our best practice workflows for human genome assembly and analysis, including variant calling, phasing, and methylation detection at: nanoporetech.com/workflows

Appendix: Accompanying protocols

Oxford Nanopore Technologies protocols:

- **High-output, long-read sequencing from human blood**

  Available at: community.nanoporetech.com/extraction_methods

- **Ultra-Long DNA Sequencing Kit (ULK001)**

  Available at: community.nanoporetech.com/protocols/ultra-long-sequencing-kit-ULK001

Oxford Nanopore provides detailed library preparation protocols and related documentation containing in-depth guidance on flow cell priming and loading.

For more information, please visit community.nanoporetech.com/docs/

Circulomics protocols:

1. **HMW DNA extraction and sample preparation (conditions 1-3):**

   DNA was extracted from 200 µl whole human blood using the Nanobind CBB Big DNA Kit (NB-900-001-01), following the HMW DNA Extraction for 200 µl Mammalian Whole Blood protocol (EXT-BLH-001).

2. **UMHW DNA extraction and sample preparation (condition 4):**

   DNA was extracted from 1.5 ml whole human blood using the Nanobind CBB Big DNA Kit (NB-900-001-01) and following the UHMW DNA Mammalian Whole Blood protocol (EXT-BLU-001).

Please refer to the Circulomics Support Page circulomics.com/support-nanobind for the latest protocol versions, and to the appropriate Nanobind Kit Handbook for additional data and guidance.

For more information, please contact info@circulomics.com

AN_1130(EN)_V1_01Jul2021 | APP-ONL-001