HIGH DATA THROUGHPUT AND LOW COST ULTRA LONG NANOPORE SEQUENCING

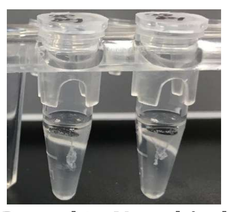
Duncan Kilburn[1], Jeff Burke[1], Renee Fedak[1], Michelle Kim[1], Hugh Olsen[2], Miten Jain[2], Simon Mayes[3], Kelvin Liu[1]

[1]Circulomics Inc, Baltimore, MD; [2]UC Santa Cruz Genomics Institute, Santa Cruz, CA; [3]Oxford Nanopore Technologies, Oxford, UK

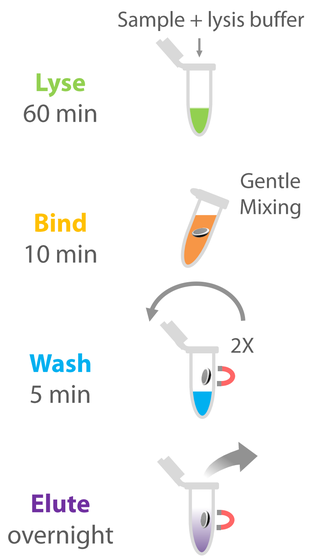
circulomics 701 E Pratt St, Baltimore, MD 21202 USA, +1-410-996-4762, www.circulomics.com, info@circulomics.com

Nanobind UHMW Extraction

Circulomics Nanobind CBB Big DNA Kit



DNA Bound to Nanobind Disk



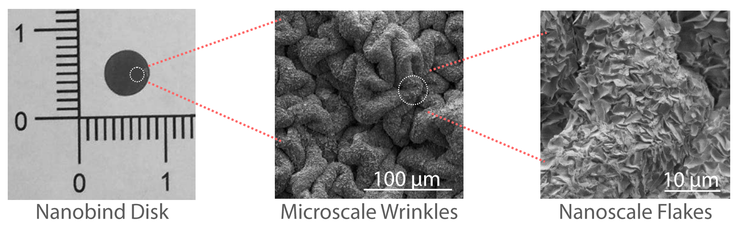
Ultra High Molecular Weight (UHMW), megabase DNA was extracted from human cells, bacteria, and human whole blood using the Circulomics Nanobind CBB Big DNA Kit. Tissues were extracted using the Circulomics Nanobind Tissue Big DNA Kit.

Nanobind UL Library Prep

Circulomics Nanobind UL Library Prep Kit

Nanobind Magnetic Disks

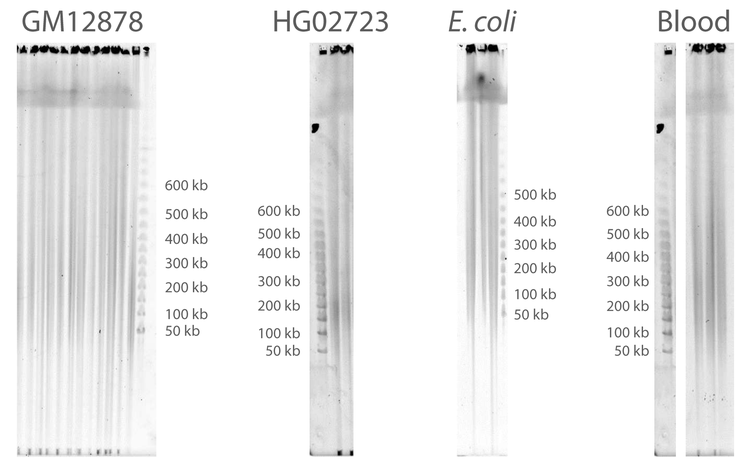
- Does not shear megabase DNA
- Micro- and nanowrinkles protect DNA from shearing
- Rapid bind, wash, and elute process
- 1 disk per tube
- Manual or automated processing



Nanobind Disk | Microscale Wrinkles | Nanoscale Flakes

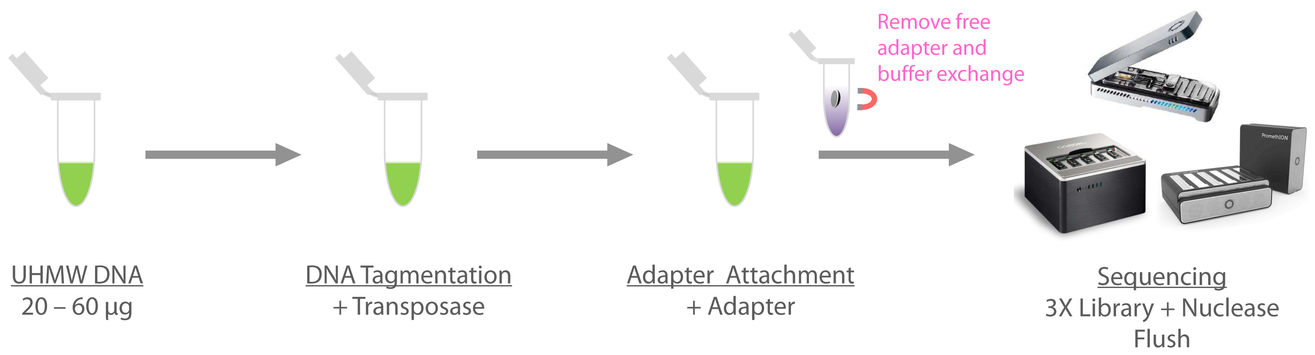
Library prep purification was performed using the Circulomics Nanobind UL Library Prep Kit, NB-900-601-01, and the Oxford Nanopore Ultra-Long DNA Sequencing Kit, SQK-ULK001. Nanobind method can purify megabase DNA without shearing it or damaging the sequencing motor protein.

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



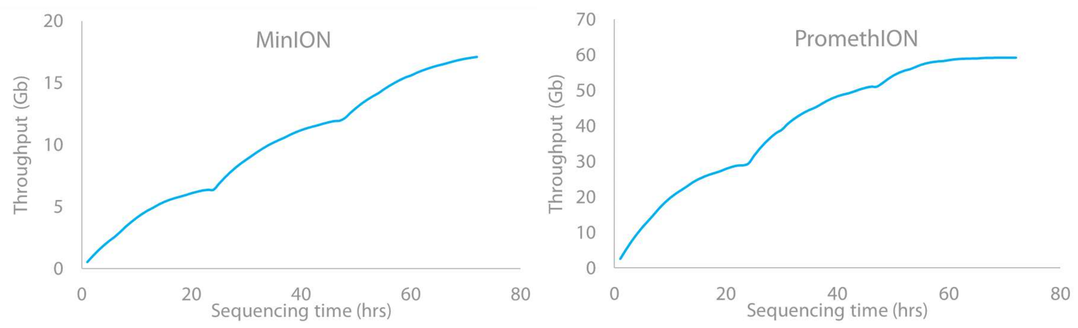
UHMW DNA
20 – 60 µg

DNA Tagmentation
+ Transposase

Adapter Attachment
+ Adapter

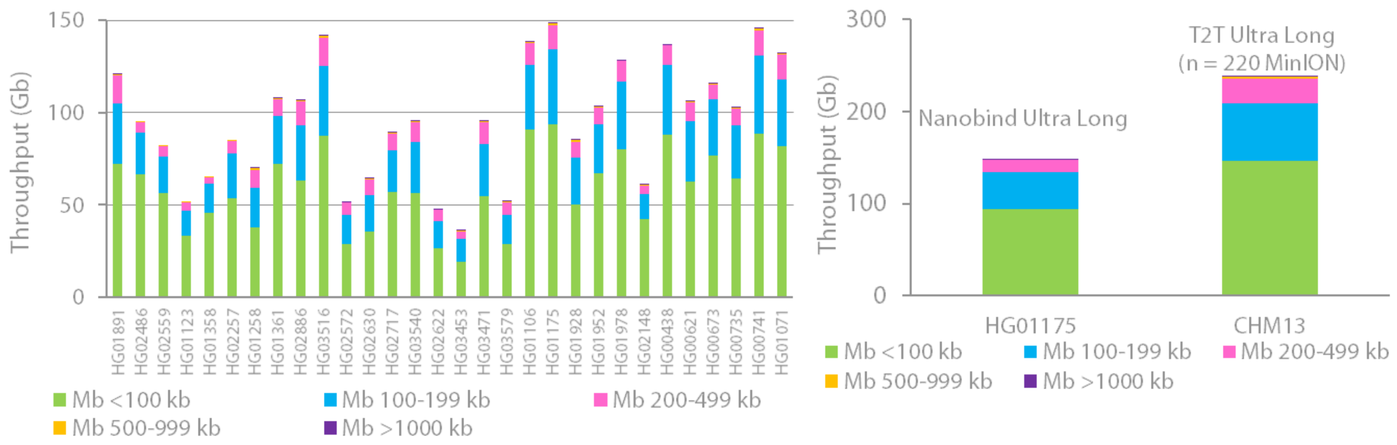
Sequencing
3X Library + Nuclease Flush

- Transposase : DNA ratio is decreased to reduce fragmentation.
- Increased reaction volume facilitates handling and mixing of viscous UHMW DNA.
- Alcohol-free Nanobind reaction cleanup concentrates, removes free adapters and performs buffer exchange without shearing megabase-sized DNA.

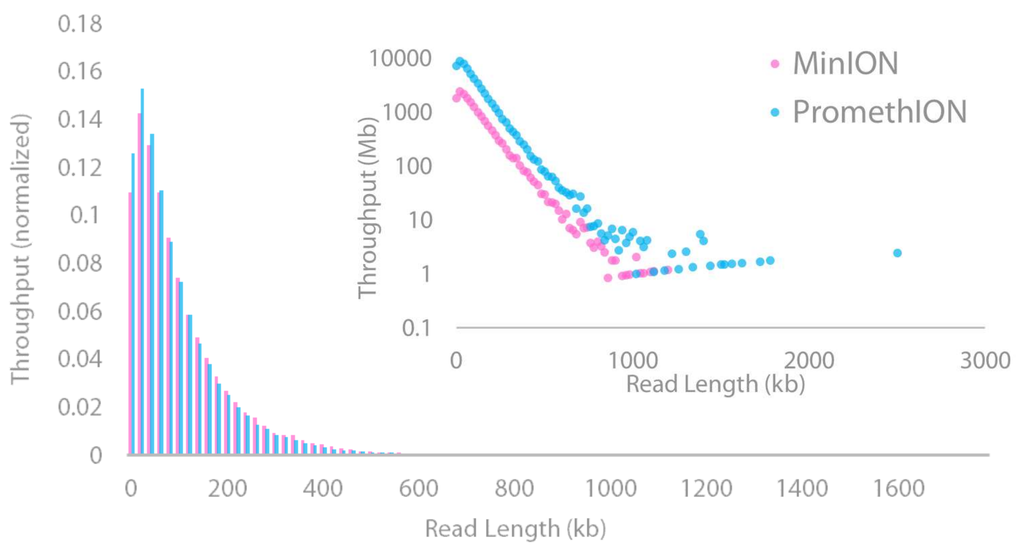


- A single 3X library is prepared, split into thirds, and loaded sequentially onto a single flow cell every 24 hours using nuclease flushes to increase throughput.
- Prepared library is compatible with MinION, GridION, and PromethION.

Oxford Nanopore Sequencing Metrics



- Nanobind 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 flow cells. N50s were between 61 kb (HG02148) and 93 kb (HG03453).
- T2T ChrX CHM13 data was generated using 220 MinION flow cells with previous ultra long approach as part of Miga, Koren *et al.* Telomere-to-telomere assembly of a complete human X chromosome. Nature, 2020.
- 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.



- HG02723 cell lines from the same extraction sequenced on MinION and PromethION instruments.
- Read length distribution is the same for both instruments. Throughput is 3.5x higher on PromethION

Multiple sample types

Sample Type	Sample	Sequencer	Total Data (GB)	Data > 100kb (GB)	Data > 200kb (GB)	Read Length N50 (kb)	Max Read Length (Mb)
Mammalian cells	GM24385	PromethION*	96	45	17	93	3.7
Mammalian cells	HG02723	PromethION	59	23	8.3	75	2.5
Blood	Bovine	PromethION	49	17	4.9	72	1.2
Gram-negative bacteria	*E. coli*	PromethION	77	47	22	128	3.3
Gram-positive bacteria	*L. monocytogenes*	PromethION	58	13	3.6	53	1.3
Tissue (in development)	Bovine Fetal Lung	PromethION	38	8.9	1.9	51	0.9
For comparison: LSK-109 with SRE Size selection from Safin *et al.* Nature Biotechnology, 2020.							
Mammalian cells	GM24385	PromethION	71	11	0.3	49	0.8

* Flow cell loaded 5 times

- 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 LSK-109 runs, and the amount of data > 200 kb 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 chr 3 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 MinION run that generated 9 Gb with an N50 of 87 kb.
- Representative reads shown in the MHC region.
- Ongoing work will test ability of UL reads to fully phase this region.