ADVANCES IN HIGH DATA THROUGHPUT AND LOW COST ULTRA LONG NANOPORE SEQUENCING

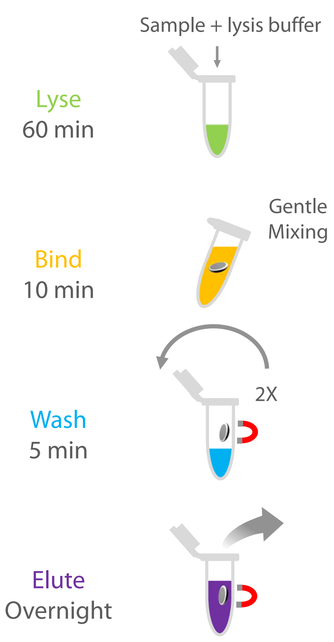
Duncan Kilburn[1], Jeff Burke[1], Renee Fedak[1], Michelle Kim[1], Justin Sibert[1], Hugh Olsen[2], Miten Jain[2], Simon Mayes[3], Kelvin Liu[1]

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

circulomics

www.circulomics.com
info@circulomics.com

Nanobind UHMW Extraction



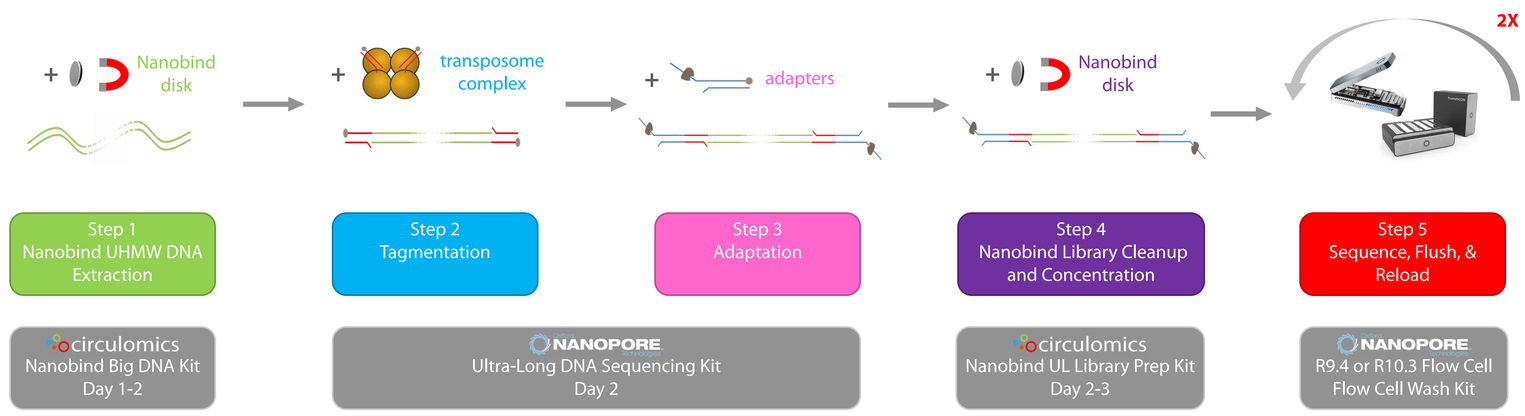
Nanobind Magnetic Disks

- Micro- and nanowrinkles protect DNA from shearing.
- 1 – 5 mm in diameter.
- 1 disk per tube.
- High binding capacity.
- High purity for single molecule sequencing.

Simple Magnetic Purification

- Bind, wash, and elute process.
- 1 disk per tube.
- UHMW protocols generate 50 kb – 1+ Mb sized DNA.
- Manual processing using magnetic rack.
- Optimized protocols for cells, bacteria, and blood.
- Developmental protocols for tissue and plant samples.

Nanobind Ultra Long Sequencing



- Nanobind Ultra Long Sequencing enables generation of large amounts of ultra long (100+ kb) reads and max reads lengths of 4+ Mb.
- First, UHMW DNA is extracted using the appropriate Circulomics Nanobind Big DNA Kit (CBB, Plant, or Tissue).
- Then, library preparation is performed using the Oxford Nanopore Technologies Ultra-Long DNA Sequencing Kit from (SQK-ULK001) and Circulomics Nanobind UL Library Prep Kit (NB-900-601-01).
- New advances presented here: 1) reduce total process time, 2) reduce DNA input requirements and 3) allows multiplexed ultra long sequencing.

Multiple Sample Types Supported

Sample	Sample input	Flow cell	Longest Read (Mb)	Read Length N50 (kb)	Gb >100 kb	Gb > 200 kb	Reads >1 Mb	Yield (Gb)
HG02723 Human Cell Line	6M cells	PRO002	0.99	93.1	60.5	20.3	0	129.6
HG005 Human Cell Line[1]	6M cells	PRO002	1.9	106.0	47.5	20.7	96	90.6
HG02723 Human Cell Line	6M cells	PRO111	1.1	83.3	37.2	11.0	1	89.7
HG02723 Human Cell Line[2]	18M cells	3x PRO111	1.2	103	105.4	35.4	4	204.7
Human Blood	1 mL	PRO002	1.3	84.4	28.4	8.1	6	67.6
E. coli	1 mL 1 OD	PRO002	3.3	128	47.1	22.5	133	77.4
L. monocytogenes	1 mL 1 OD	MIN106D	1.2	50	4.8	1.3	1	22
Fish Blood	10 µL	MIN106D	1.0	98.6	8.3	3.4	1	16.8
Bovine Lung[3]	34 mg	PRO002	0.9	51	8.9	1.9	0	38.5
Sugar Beet[4]	2 g	MIN106D	0.4	47	2.3	0.3	0	11.6

[1]No-vortex Library prep.
[2]Single, 9X scale up extraction and library prep. Library split and run on three PromethION flow cells.
[3]Data generated in collaboration with USDA.
[4]Data generated in collaboration with KeyGene.

- Nanobind UMHW extractions are critical for generating UL sequencing data.
- Further development is ongoing to enable UL tissue and plant sequencing.
- 468 Gb of UL data from the HG02723 cell line data generated in collaboration with UCSC Genomics Institute are available for download, see www.circulomics.com/datasets for details.

New 1-Day Protocol

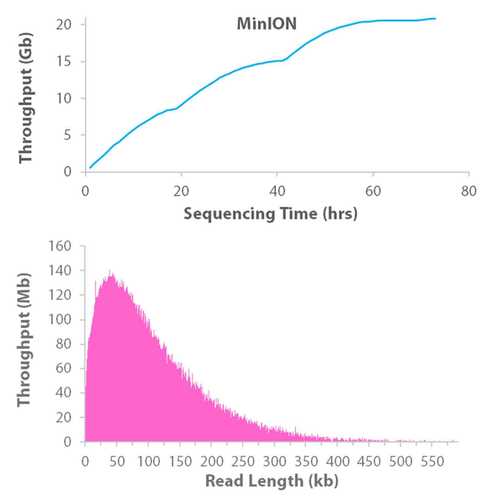
9.00 am — Extraction

11 am – 1.30 pm, Extracted DNA resting

1.30 pm — UL Library prep

3.00 pm – 4.30 pm, Library resting

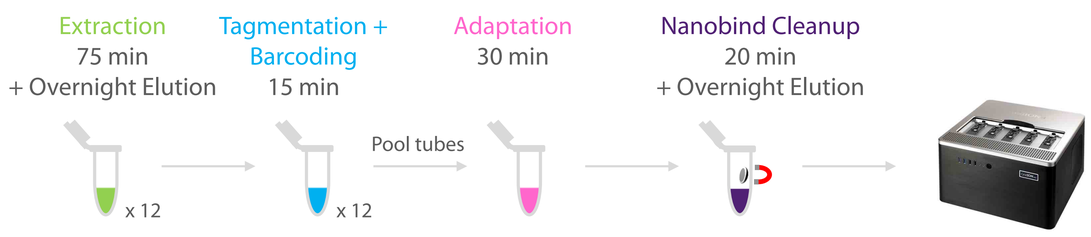
4.30 pm — Sequence



Sample	Sample input	Flow cell	Longest Read (Mb)	Read Length N50 (kb)	Gb >100 kb	Gb > 200 kb	Reads >1 Mb	Yield (Gb)
HG005 Human Cell Line	4M cells	MIN106D	1.2	87.3	9.0	2.8	1	20.6

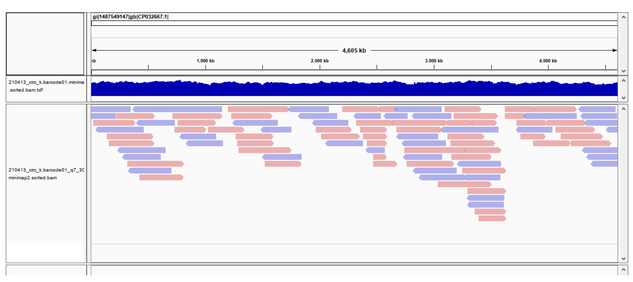
- New protocol reduces total extraction and library prep time from 3 days to 1 day.
- Overnight elution steps are eliminated. Reduced input testing is underway.
- Sequencing data are commensurate with standard protocol.

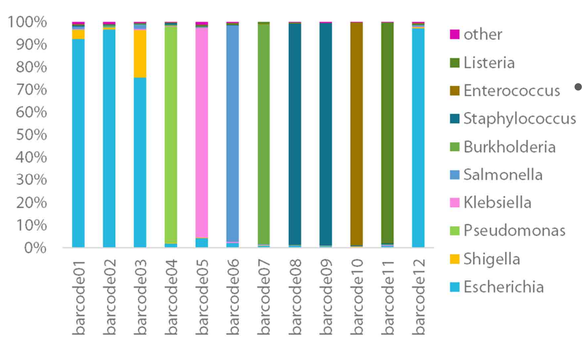
Bacterial Multiplex Sequencing



- Nanobind ultra long sequencing was performed in multiplex on 12 gram negative and gram positive bacteria samples.
- Nanobind UHMW DNA extraction was performed at 1/12th standard input.
- Tagmentation + barcoding was performed at 1/12th volume (SQK-RBK004).
- All 12 samples were pooled for adapter attachment and Nanobind clean up and then sequenced in multiplex for 72 h on GridION. Reads were de-multiplexed with Guppy.

- *E. coli* shows uniform coverage (top track) when mapped using minimap2.
- Bottom track shows alignments of 300 kb+ reads with barcode 01 to *E. coli* reference.
- The full genome can be spanned with 15-20 reads for all bacteria.





- Analysis performed using Fastq WIMP (v2021.03.05) on ~3000 randomly selected reads per barcode show:
  1. Most barcodes contain predominance of correct assigned bacteria.
  2. 75% of reads with barcode 03 *Shigella* reads are misidentified as *Escherica*.

	Sample	Read Length N50 (kb)	% Data >100 kb	Longest read (Mb)	Yield (Gb)	Assembly cutoff (kb)	Assembly depth (x)	Contig Length (Mb)	Circular
Barcode01	*Escherichia coli* K12	86	43	0.76	1.09	300	7	4.6	Y
Barcode02	*Escherichia coli* W	91	44	0.83	1.09	300	8	4.9	Y
Barcode03	*Shigella sonnei*	94	45	0.59	0.88	300	9	4.8	Y
Barcode04	*Pseudomonas aeruginosa*	105	46	0.55	0.47	300	3	6.9	Y
Barcode05	*Klebsiella pneumoniae*	97	47	0.61	1.07	300	9	5.4	Y
Barcode06	*Salmonella enterica*	95	48	1.09	1.04	300	12	4.8	Y
Barcode07	*Burkholderia cepecia*	100	49	0.75	0.98	200	17,17,23,52	3.8,3.4,1.2, 0.2	Y
Barcode08	*Staphylococcus aureus*	88	50	0.53	0.98	300	9	2.9	Y
Barcode09	*Staphylococcus aureus*	91	51	0.60	1.04	300	10	2.8	Y
Barcode10	*Enterococcus faecalis*	61	52	0.44	0.74	200	12	2.7	Y
Barcode11	*Listeria monocytogenes*	94	53	0.88	1.42	300	20	3.0	Y
Barcode12	*Escherichia coli* W	90	54	0.65	0.51	200	13	4.9	Y
Unclassified		86	55	0.57	4.16				
Total		90	56	1.09	15.48				

- Yields and read length distributions are similar for all barcodes.
- 27% of reads are unclassified.
- Subset of reads filtered by q (>7) and length (>200 or >300 kb) were assembled with Flye.
- All bacteria assembled to a single circular contig.
- *Burkholderia cepecia* has three chromosomes and one 200 kb plasmid. Reads >300 kb are too long to provide coverage of the plasmid. Reads with barcode07 and a cutoff >200 kb include 52 reads that span the entire plasmid.