Generating Ultra Long Reads on Oxford Nanopore MinION/GridION/PromethION

03/01/2020

Dr. Kelvin Liu
kliu@circulomics.com

circulomics

DNA Quality

For long reads, DNA quality is paramount.

Library Prep!

ACGT

DNA Quality

Size
Purity
Damage

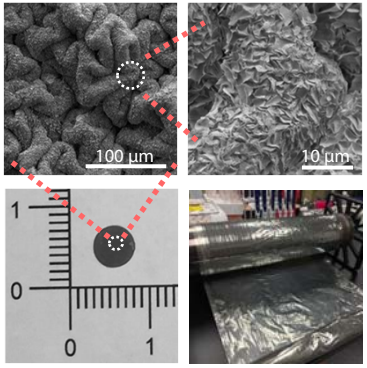
Sequence Metrics

Read Length
Throughput
QV

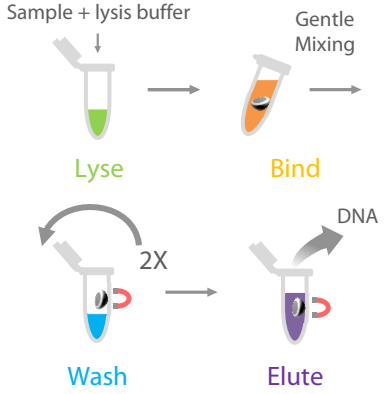
circulomics

Nanobind Magnetic Disks

Nanostructured magnetic disks for rapid HMW and UHMW DNA extraction



High Surface Area + Low Shear Force



Rapid Bind, Wash, and Elute

Oxford NANOPORE Technologies

bionano GENOMICS

10X GENOMICS

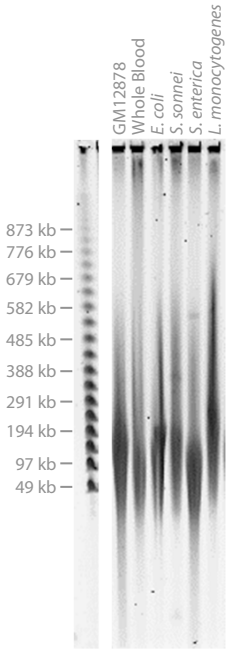
PACBIO

Specifically Optimized for Long-Range Technologies

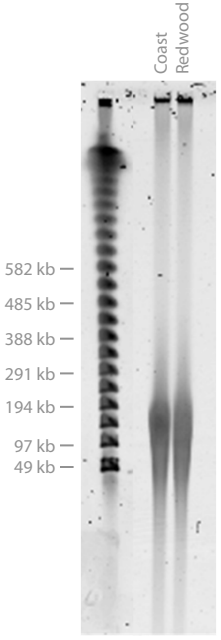
circulomics

HMW DNA Extraction (50 – 300+ kb)

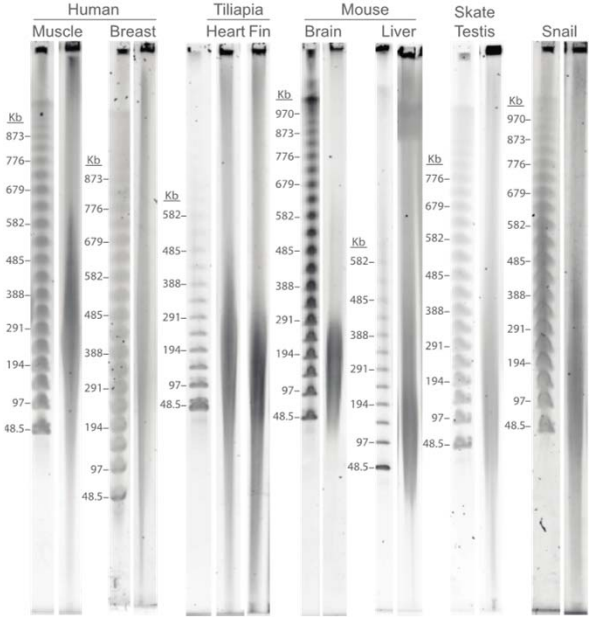
Recommended for: Oxford NANOPORE Technologies | PACBIO | 10X GENOMICS



Cells, Bacteria, Blood



Plants

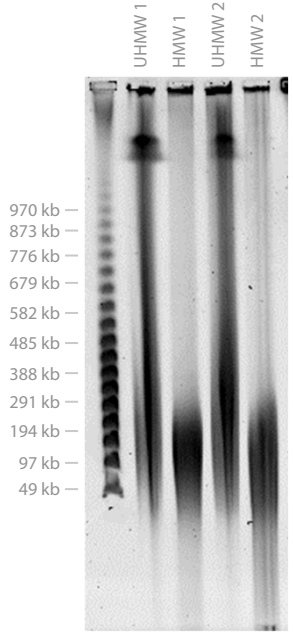


Tissues

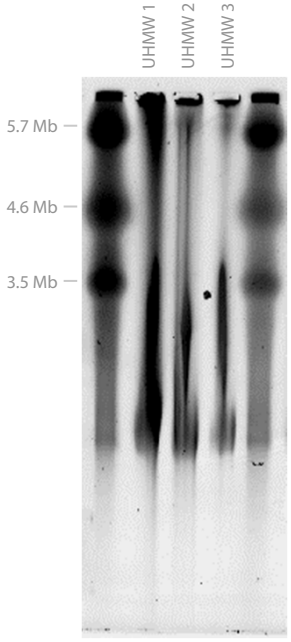
circulomics

UHMW DNA Extraction (50 kb - 1+ Mb)

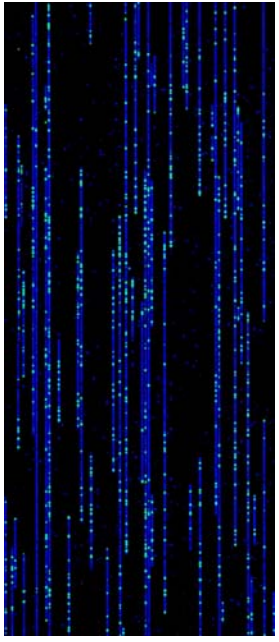
Recommended for: Oxford NANOPORE Technologies bionano GENOMICS



22 Hr CHEF Gel



72 Hr CHEF Gel



Bionano Optical Map

circulomics

Whale Watching On ONT – Josh Quick Protocol

Generate 100 kb – 1+ Mb reads (whales). Longest read to date = 2.27 Mb
ONT Rapid Sequencing Kit with high DNA concentration with low transposase = ultra long reads

10 µg UHMW DNA (400 ng) → 1 µg/µL DNA

+ 1.5 µL FRA (2.5 µL) → DNA Tagmentation

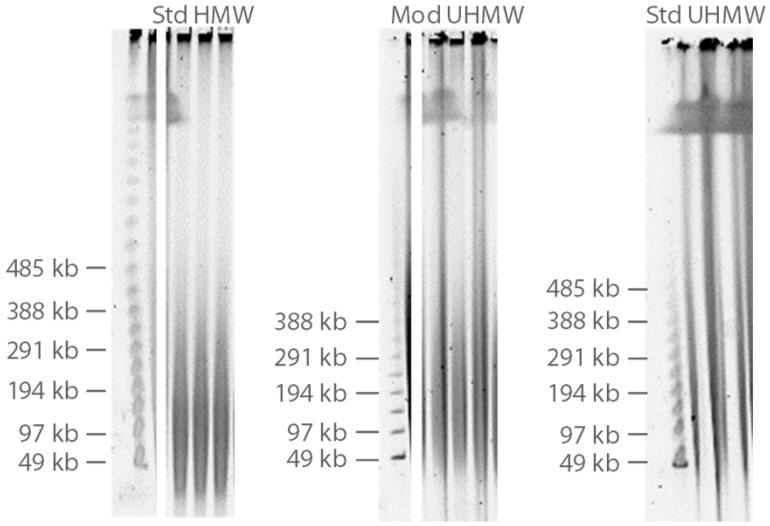
+ 1 µL RAP → Adapter Attachment

circulomics

Bigger DNA = More Whales

Used modified Quick Protocol to test effect of DNA size.

Longer DNA = Longer Reads



Read Length N50 (bp)	33.7 kb	48.5 kb	60.7 kb
Throughput	2.9 Gb	3.1 Gb	1.9 Gb
Data >100 kb	13%	25%	34%
Reads >1 Mb	0	2	4
Longest Read (Mb)	0.78	1.1	1.3

Whale Watching Challenges

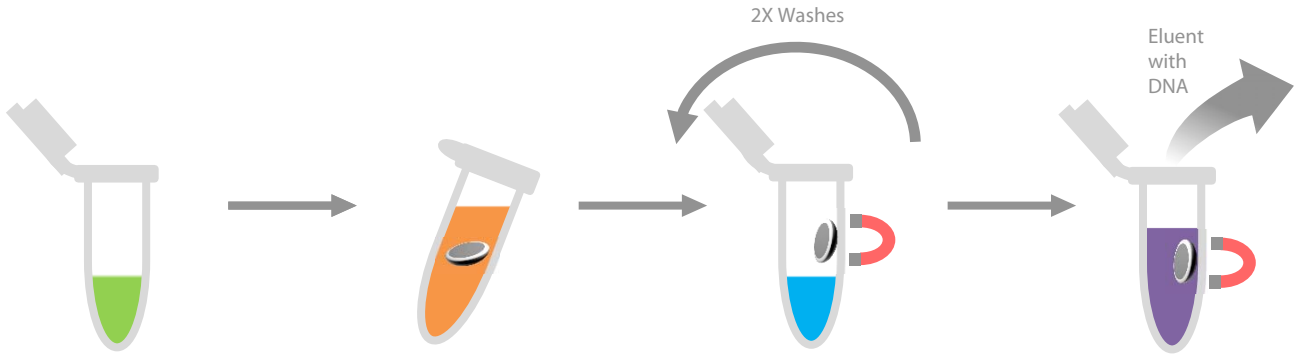
2 major challenges make this protocol difficult for widespread adoption

1. High concentrations of megabase DNA = Very viscous
   - Extraction changes
   - Library prep changes

2. Low throughput and few whales
   - Library prep changes
   - Sequencing changes

Extraction Changes to Reduce Sample Viscosity

Made tweaks to the Lysis and Binding chemistry to reduce viscosity of megabase DNA without affecting size



Step 1
Lyse

Step 2
Bind

Step 3
Wash

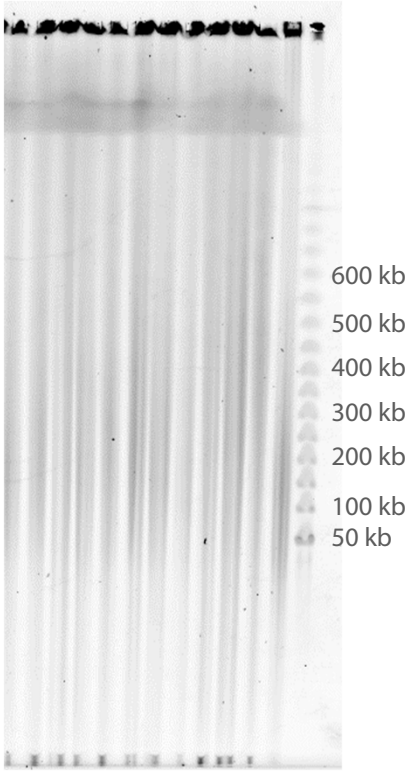
Step 4
Elute

circulomics

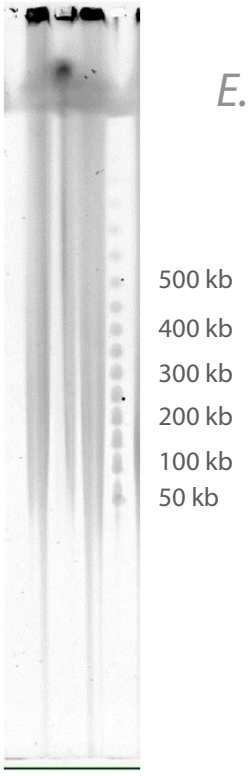
Megabase Sized DNA

New chemistry reduces sample viscosity but maintains megabase sized DNA.

GM12878



E. coli / L. monocytogenes



circulomics

Even more whales?

1) Dilute reaction

2) Concentrate and purify library

10 µg UHMW DNA
1 µg/µL

+ 1.5 µL FRA

+ 1 µL RAP

10 µg UHMW DNA
0.1 µg/µL

+ 1.5 µL FRA

Concentrate and purify

+ 1 µL RAP

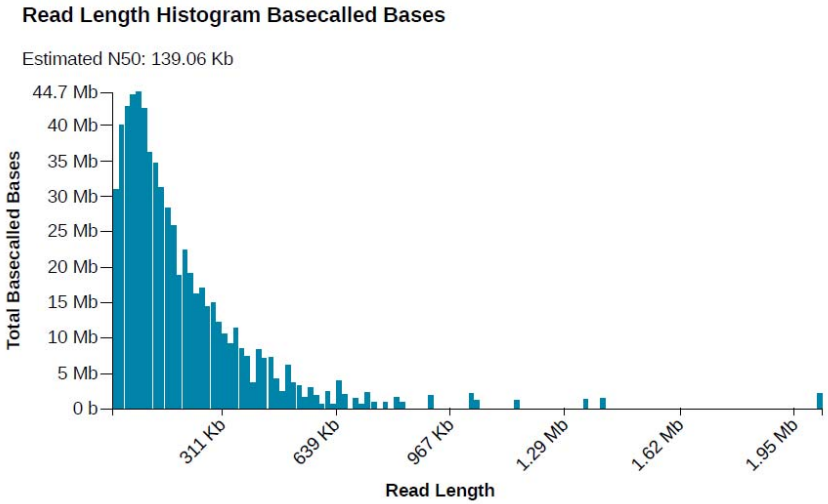
Remove free adapters
and buffer exchange

circulomics

A Small Pod of GM12878 Whales

Improved mixing and extraction leads to large increase in ultra long reads

Generated 2.1 Mb unfused read

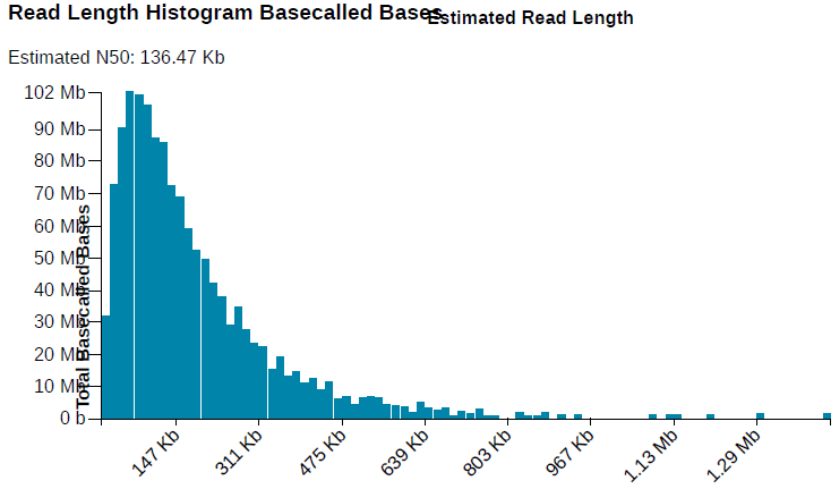


	Read Length N50	Number of Mapped Whales	Longest Mapped Read
Before Whale Watch	139 kb	7	2.1 Mb
After Whale Watch	179 kb	13	2.1 Mb

Fused + Unfused Whales	
Mapped Read Length	**Chromosome**
2.1	3
1.7*	-
1.4 x 2	7, 12
1.3	8
1.2	3
1.1 x 6	2, 4, 5, 7, 16, 20
1.0	2

A Small Pod of *E. coli* Whales

Comparable results seen with bacteria



	Read Length N50	Number of Mapped Whales	Longest Mapped Read
Before Whale Watch	134 kb	6	1.4 Mb
After Whale Watch	175 kb	15	1.6 Mb

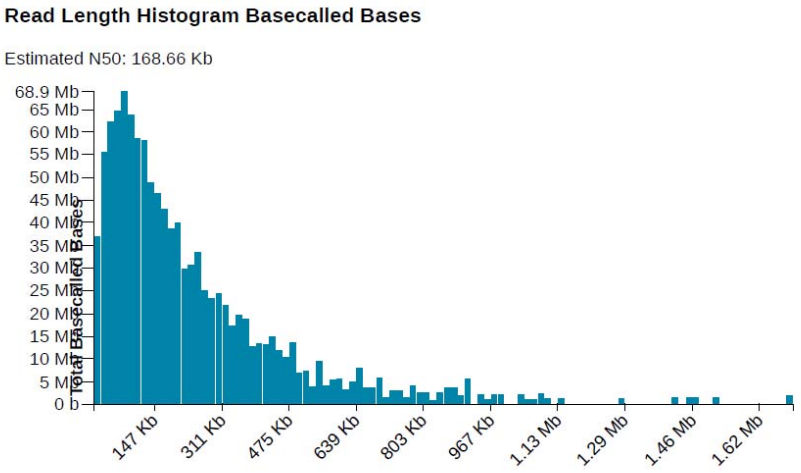
Fused + Unfused Whales
Mapped Read Length
1.6
1.5
1.3 x 3
1.2 x 6
1.1 x 4

circulomics

Longest Nanopore Read to Date?

Tested the protocol on human whole blood

Generated 2.44 Mb fused read (1.1 + 0.85 + 0.5)



	Read Length N50	Number of Mapped Whales	Longest Mapped Read
Before Whale Watch	168 kb	16	1.8 Mb
After Whale Watch	193 kb	28	2.44 Mb

Fused + Unfused Whales	
Mapped Read Length	Chromosome
2.44	13
2.2	1
2.0	2
1.7 x 3	2, 2, 3, 4
1.6	3
1.5 x 4	3, 6, 7, 18
1.4 x 2	2, 16
1.3 x 3	2, 7, 8
1.2 x 3	2, 4, 4
1.1 x 4	6, 11, 15, 18
1.0 x 4	8, 11, 14, 19

Increasing Throughput – Nuclease Flush

Load multiple libraries interspersed with nuclease flush

Load Library

72 hours

Load Library

Nuclease flush Load Library

24 hours

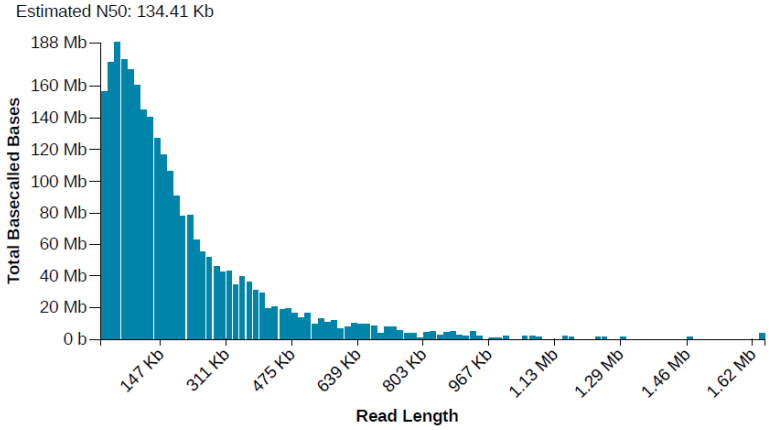
48 hours

A Medium Pod of Whales

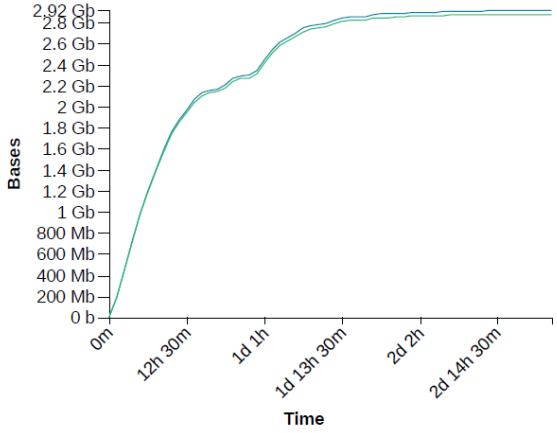
Nuclease flush increased throughput by 30%.

36 whales. 65% data >100 kb. 40% data >200 kb

Read Length Histogram Basecalled Bases



Cumulative Output Bases



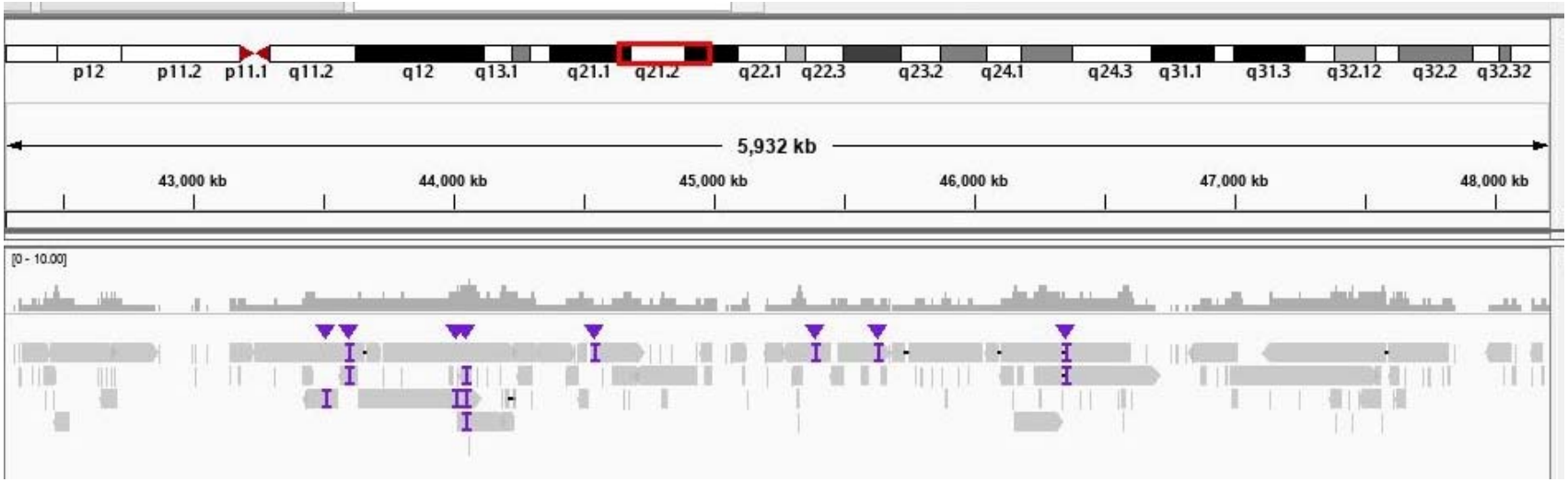
	Read Length N50	Number of Mapped Whales	Longest Mapped Read
Before Whale Watch	134 kb	16	1.7 Mb
After Whale Watch	150 kb	36	1.9 Mb

Fused + Unfused Whales	
Mapped Read Length	Chromosome
1.9	2
1.7	7
1.6	15
1.5 x 3	14, 14, 8
1.4 x 3	4, 6, 17
1.3 x 6	5, 5, 10, 11, 13, 13
1.2 x 6	6, 7, 7, 13, 14, 17
1.1 x 10	3, 3, 4, 4, 6, 10, 12, 17, 20, 21
1.0 x 5	1, 1, 2, 16, X

circulomics

Mapping Ultra Long Reads

Massive differences in scale between 1.5 Mb vs. 500 kb vs. 50 kb reads



Increasing Throughput - PromethION

RAD004 chemistry is not officially supported for PromethION yet.

However, after Nanobind purification + buffer exchange RAD004 library = LSK109 library

Load Library

24 hours

Nuclease flush
Load Library

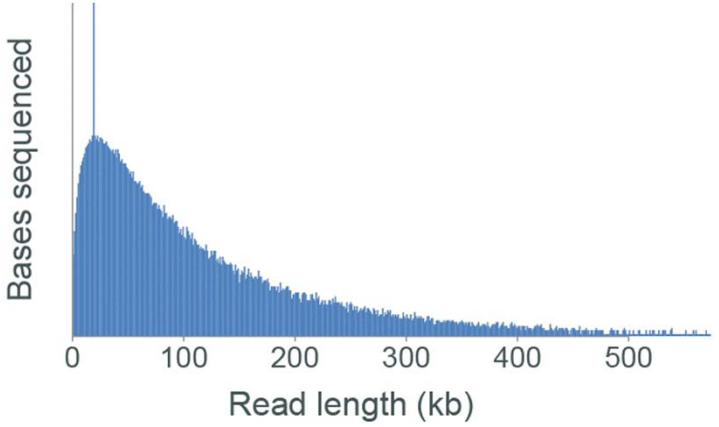
24 hours

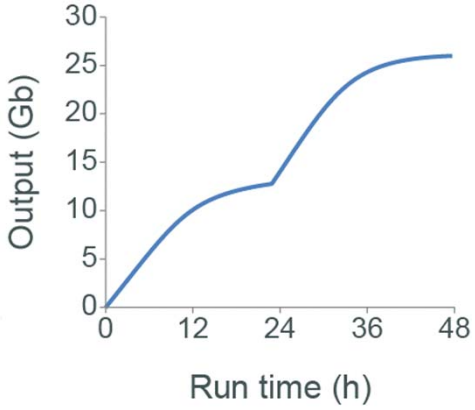
circulomics

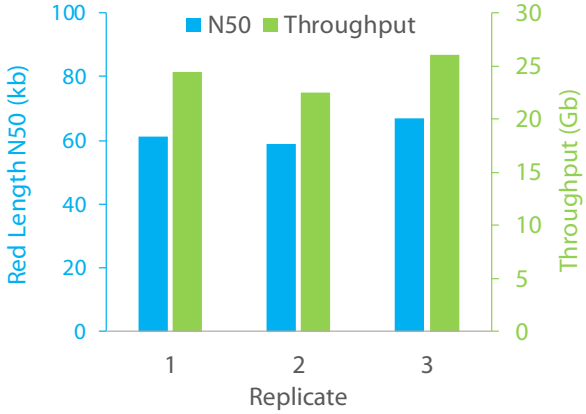
Towards A Large Pod of Whales

Protocol biased toward higher throughput (3X FRA).

14% data (3.4 Gb) in reads >200 kb. Even higher throughput w/ 3rd library?







circulomics

Conclusion

- New extraction chemistry -> Reduced viscosity and improved mixing
- Fragmentation performed at low concentration -> Improved mixing and reaction efficiency
- Nuclease flush -> Higher throughput
- Reaction purification -> Scales to PromethION

Acknowledgements

Questions?

circulomics

Jeffrey Burke
Duncan Kilburn
Renee Fedak
Kelvin Liu

Oxford NANOPORE Technologies

Simon Mayes
David Stoddart
Michelle Hiscutt
Rosemary Dokos

Johns Hopkins University

Winston Timp

University of California Santa Cruz

Karen Miga
Miten Jain
Hugh Olsen

University of Washington

Glennis Logsdon

University of Birmingham

Josh Quick

The University of Nottingham

Matt Loose

NIH National Institute of General Medical Sciences

NIH National Human Genome Research Institute