EXTRACTING DNA FOR NANOPORE SEQUENCING IN THE REDWOOD GENOME PROJECT

Winston Timp

Johns Hopkins University

Department of Biomedical Engineering

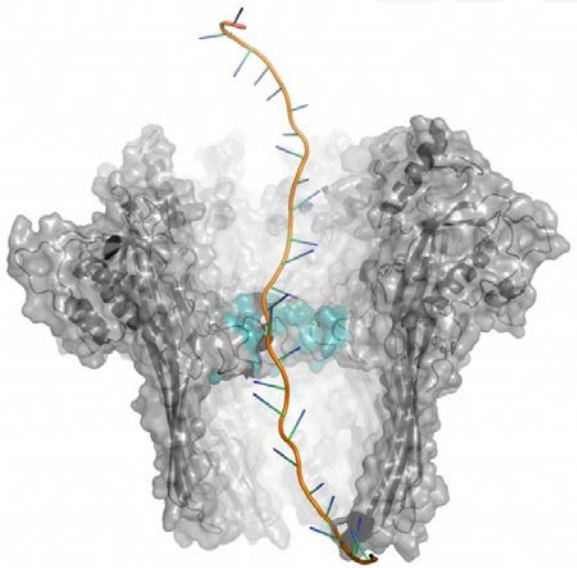
NANOPORE: SINGLE MOLECULE SEQUENCING

Oxford Nanopore Technologies, CsgG biological pore

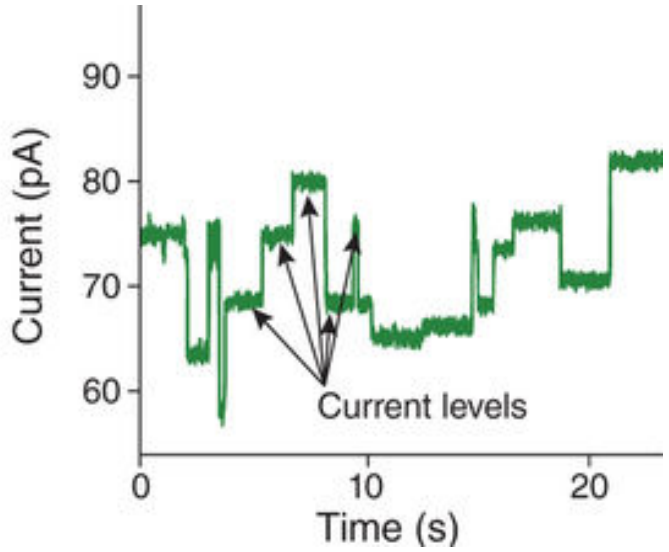
No theoretical upper limit to sequencing read length, practical limit only in delivering DNA to the pore intact

Palm sized sequencer

Predicted sequencing output 3-6Gb



Oxford Nanopore Google Hangout March 2016

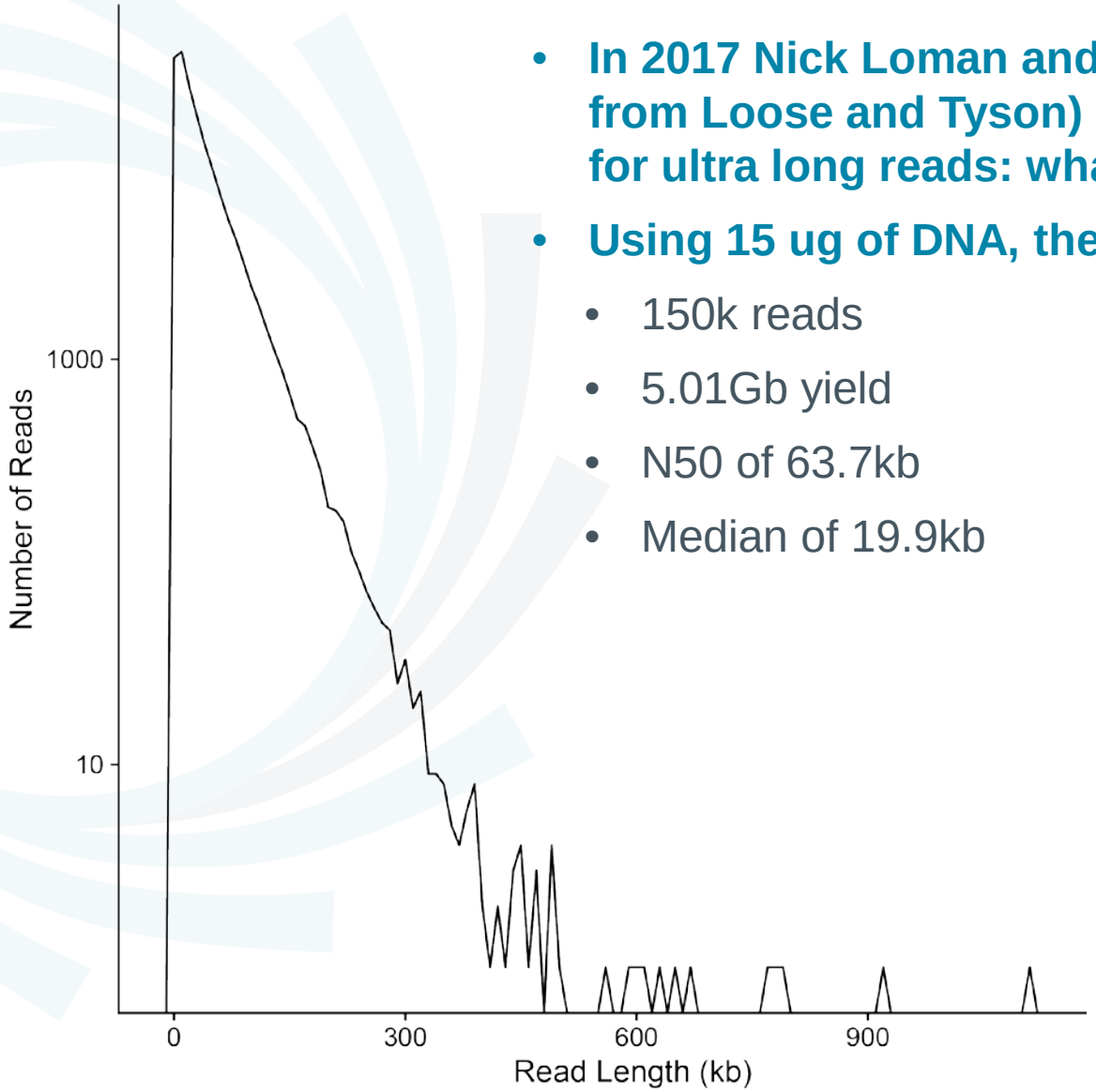


Deamer et al 2016, Nature Biotech

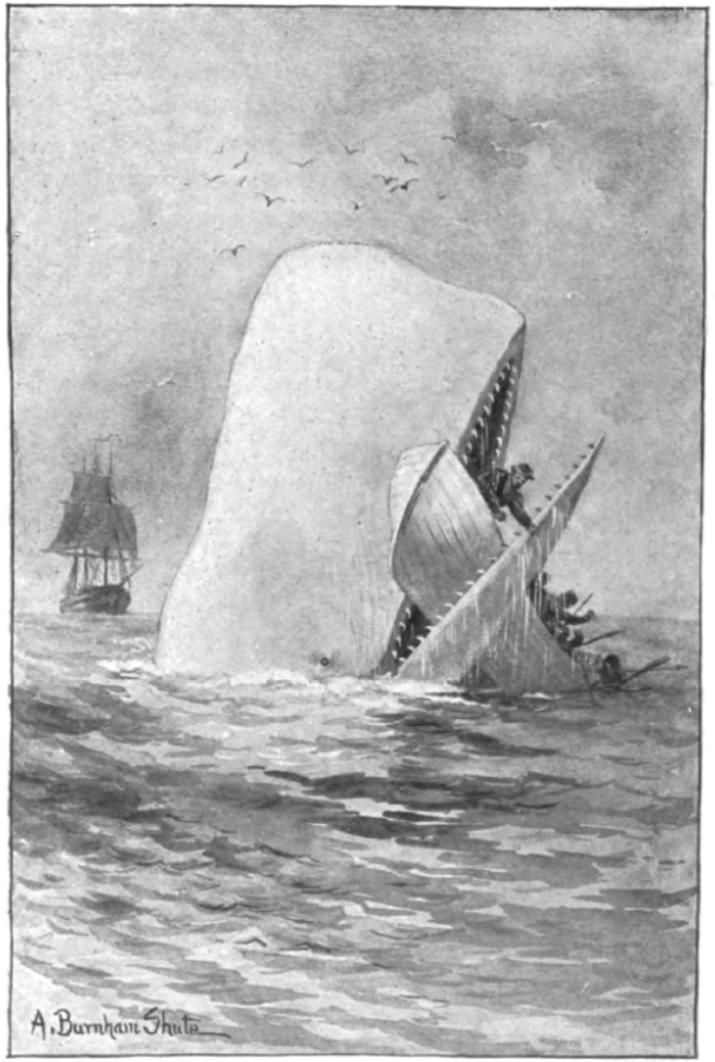
ATCGATCGATAGTAT
TAGATACGACTAGC
GATCAG

Disclosure: Timp has two patents (US 2011/0226623 A1; US2012/0040343 A1) licensed to ONT

LOMAN LAB WHALE SEARCH



- **In 2017 Nick Loman and Josh Quick (with help from Loose and Tyson) described their search for ultra long reads: whales**
- **Using 15 ug of DNA, they got:**
  - 150k reads
  - 5.01Gb yield
  - N50 of 63.7kb
  - Median of 19.9kb



"Both jaws, like enormous shears, bit the craft completely in twain."

—*Page 510.*

PROBLEM

Want: High molecular weight (HMW) DNA

- **100kb + average**

Need: High yield

- **At least 10ug purified gDNA from 1 g of leaf tissue**
- **Quick/Loman protocol assumed infinite cell mass (?)**

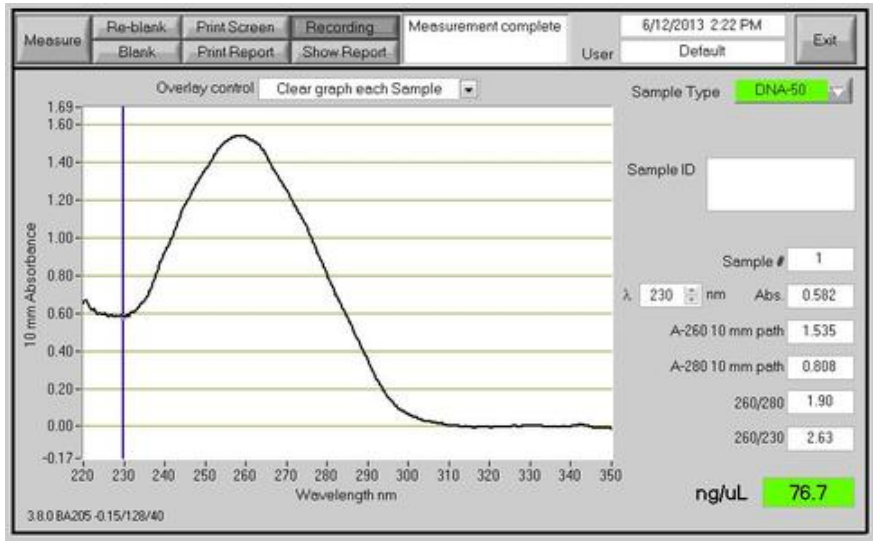
Need: High quality

- **Based on nanodrop spectra and gel migration**
- 260/280 ~1.8
- 260/230 ~2.0+
- No aberrant migration

Reproducible/Robust protocol

- **Want no wizards**

High and consistent sequencing yield (- at least 3-5Gb per run)



CHALLENGES: DNA EXTRACTION

Most protocols optimized for:

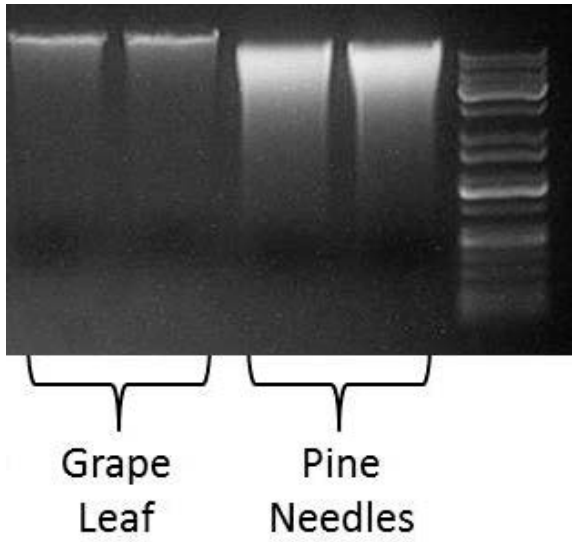
- **10-15kb average fragment size**
- **Seedlings grown in dark**
- **Anything but conifers, etc**

Trade-off between fragment length and yield

Each batch of sample behaves slightly differently

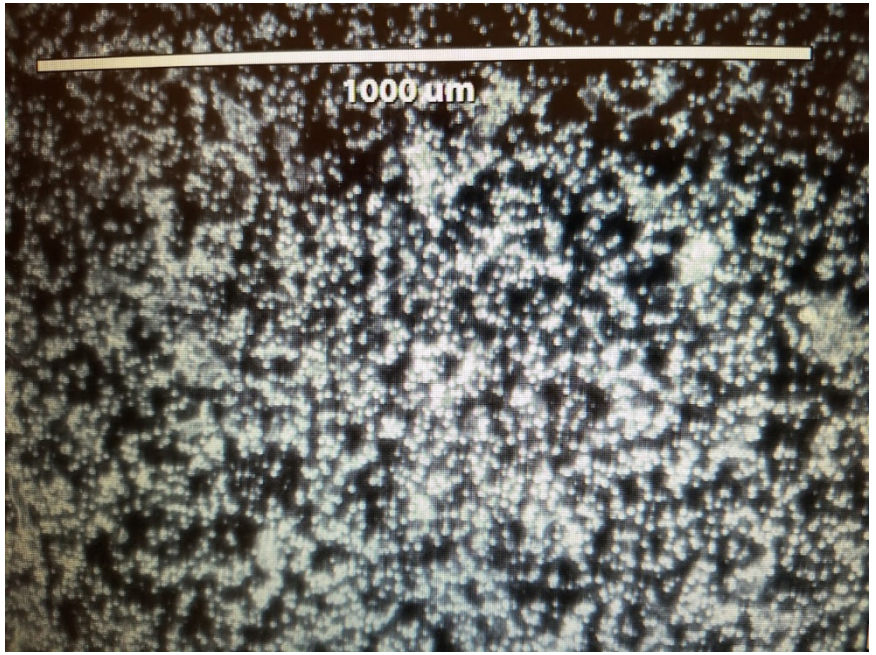
Polyphenolics and polysaccharides co-precipitate with DNA in many chemistries

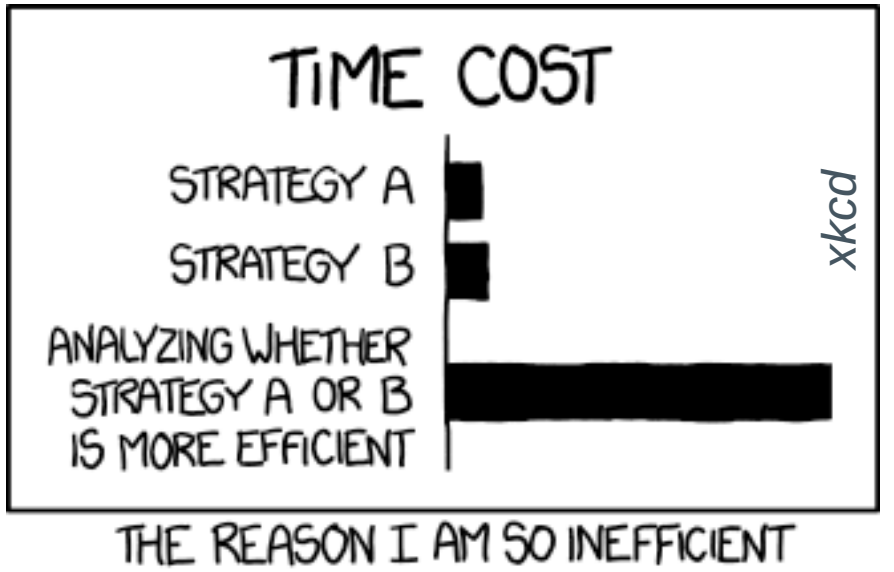
DNeasy PowerPlant kit



TRIALS: DNA EXTRACTION

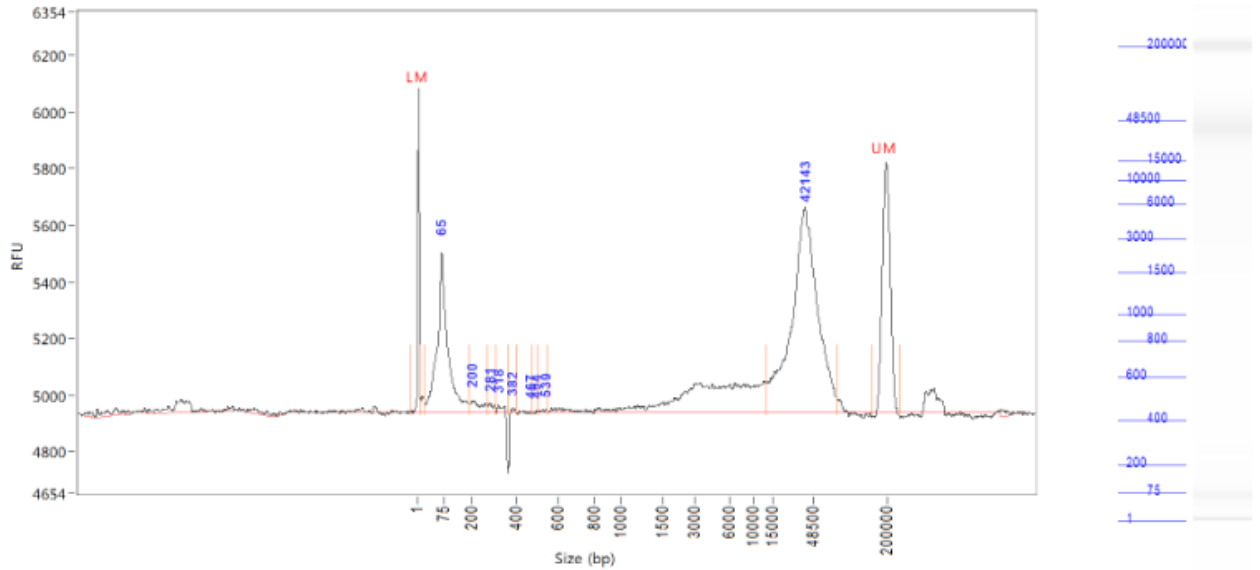
- Detergent (SDS, SLS, CTAB, Triton)
- Extract nuclei first (yes or no)
- Phenol chloroform (yes or no)
- Salt used for alcohol precipitation (NaCl, sodium acetate, ammonium acetate, none)
- Modifications for improving fragment lengths (agarose embedding, nanobind)
- Synchronous coefficient of drag alteration (SCODA) for purification





TOP 3 EXTRACTION PROTOCOLS: ZHANG

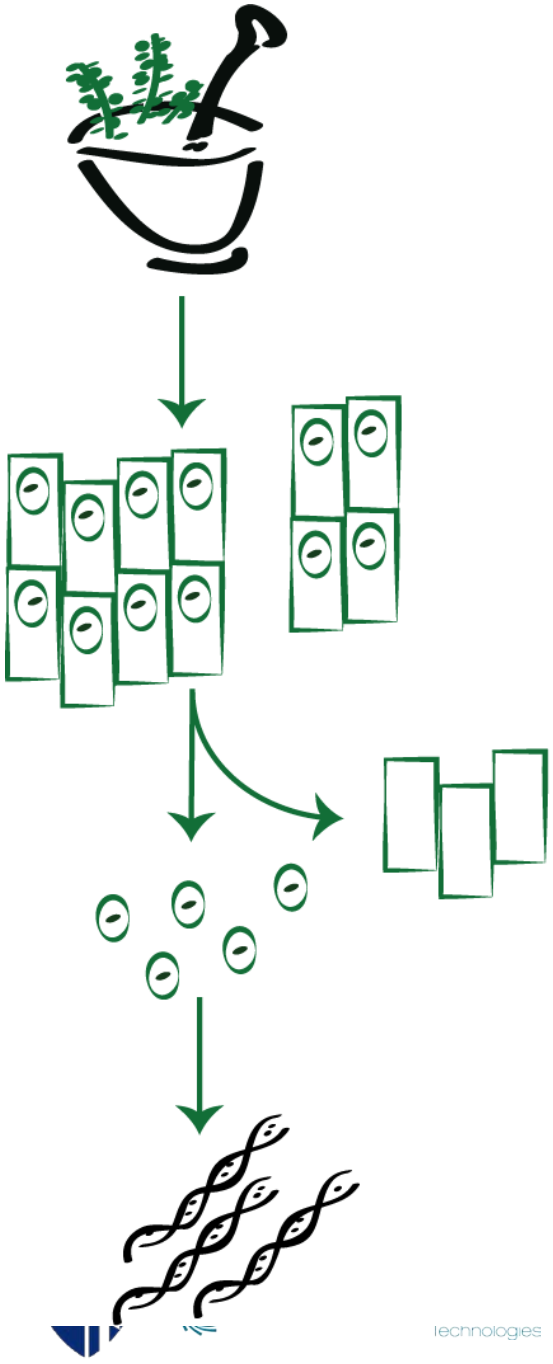
Sample: seq_zhang
Well Location: E5
Created: Tuesday, August 22, 2017 12:17:00 PM



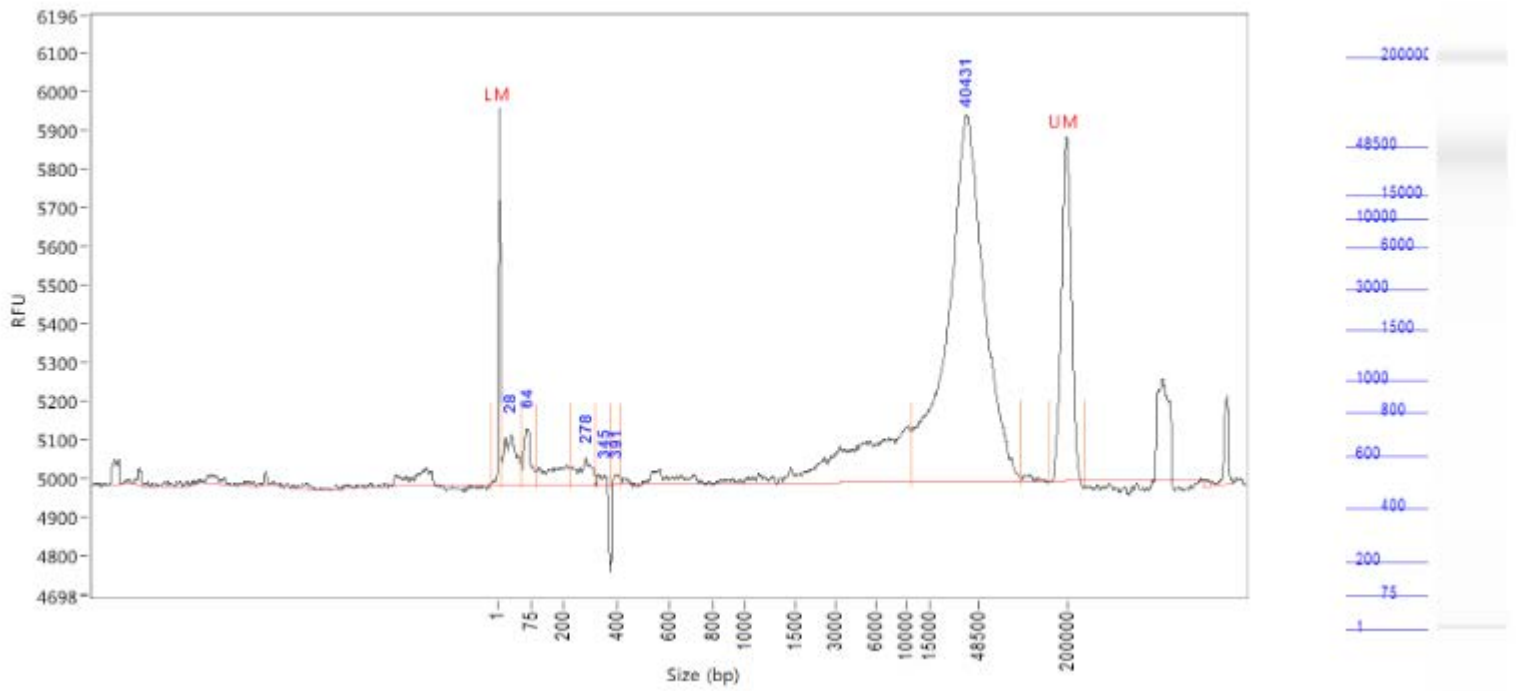
Tissue ground in LN2

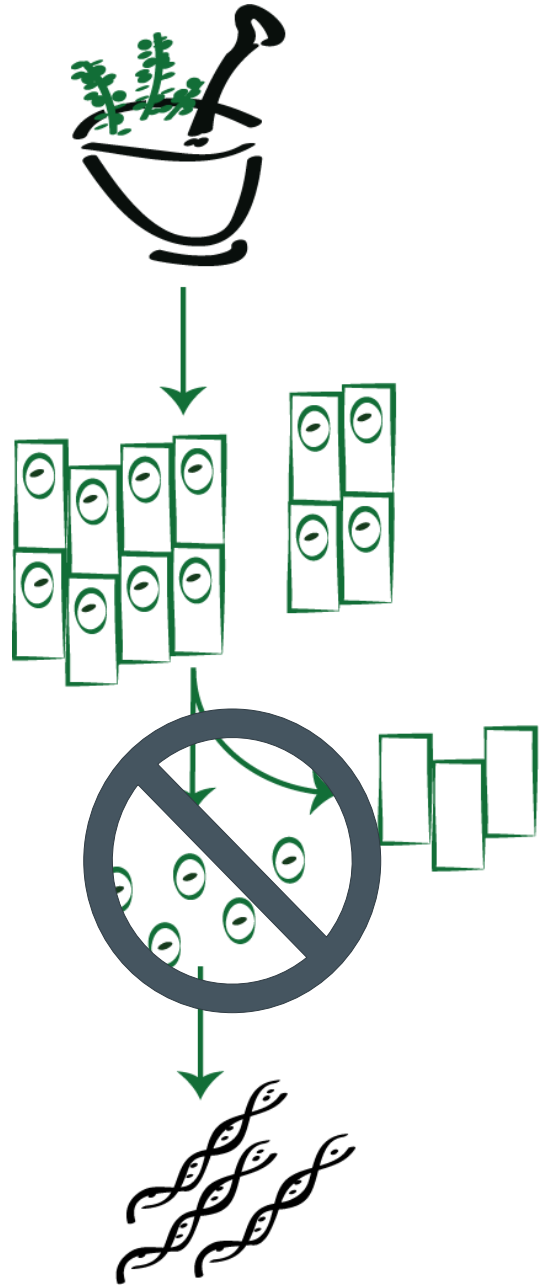
Nuclei: Cell wall lysis -> filtration -> differential centrifugation

DNA: Overnight SLS lysis, phenol chloroform, alcohol precipitation + sodium acetate



TOP 3 EXTRACTION PROTOCOLS: HEALEY





Preferred plant protocol: Skips the differential nuclei extraction

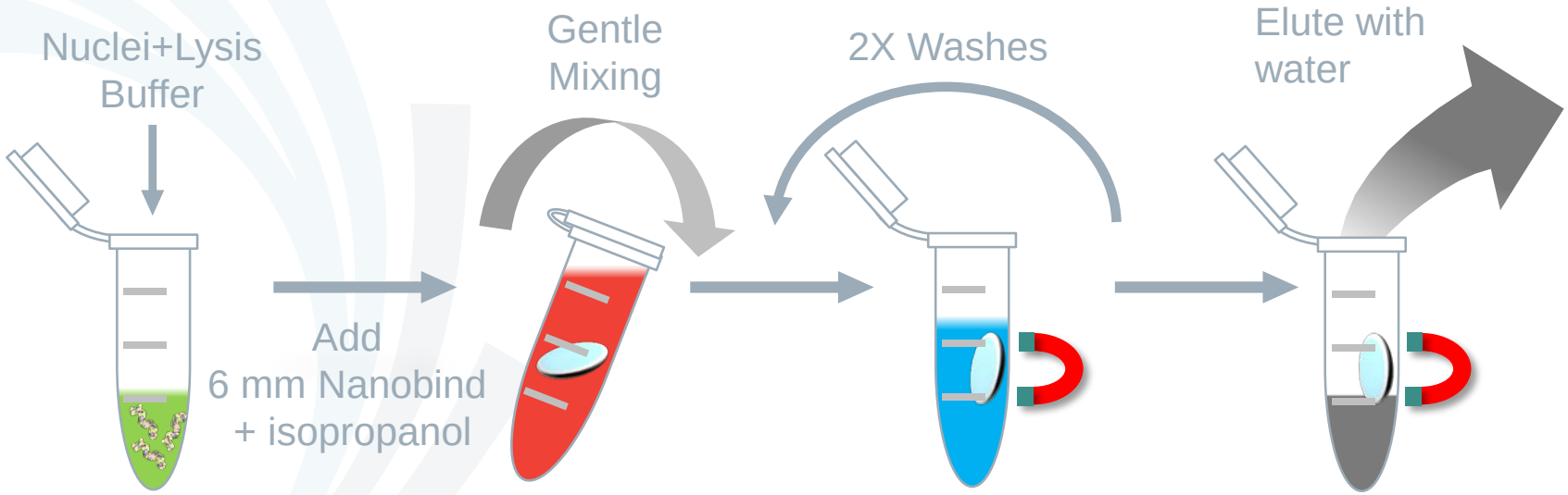
CTAB (cetyl trimethylammonium bromide) – cationic detergent

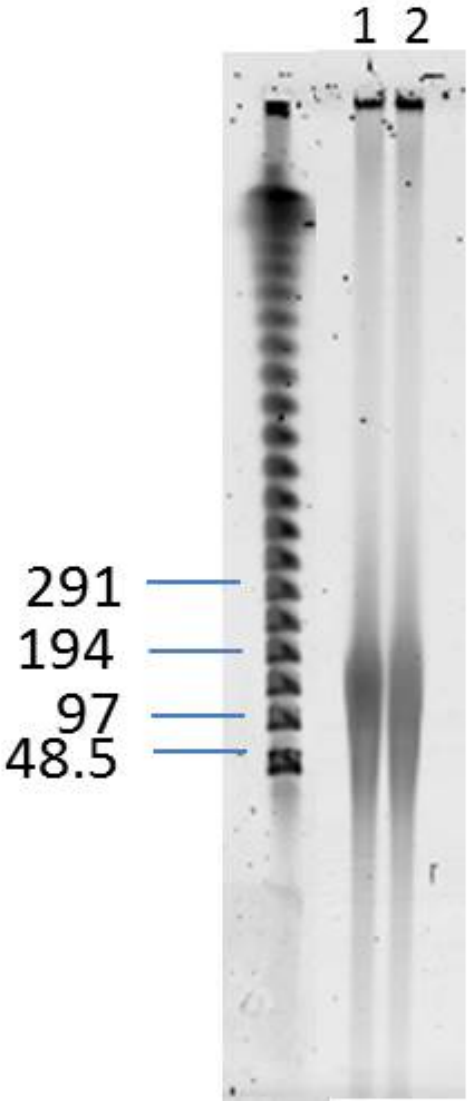
CTAB extraction, chloroform, alcohol precipitation + NaCl, elute

TOP 3 EXTRACTION PROTOCOLS: NANOBIND

circulomics

Nuclei →





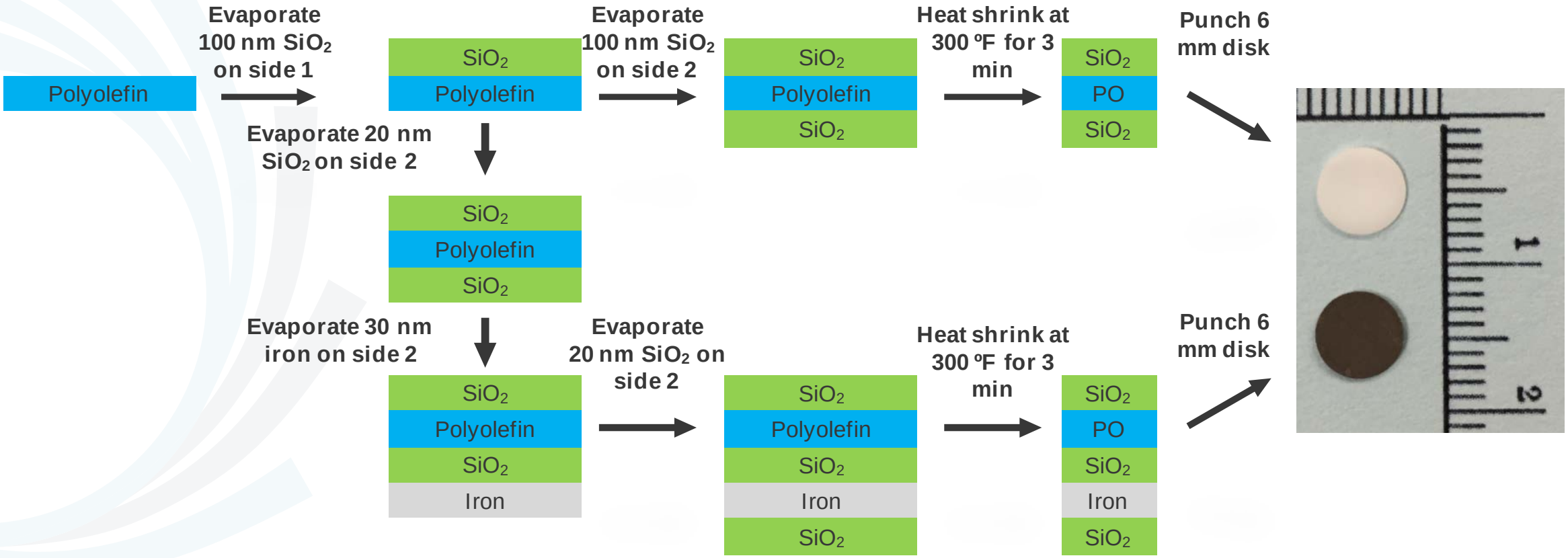
Nanobind: Nuclei isolation as previously described

Nanobind extraction with modified CTAB, EtOH washes, elute

45 minutes total time (after nuclei extract)

Oxford NANOPORE Technologies

NANOBIND: HOW DOES IT WORK



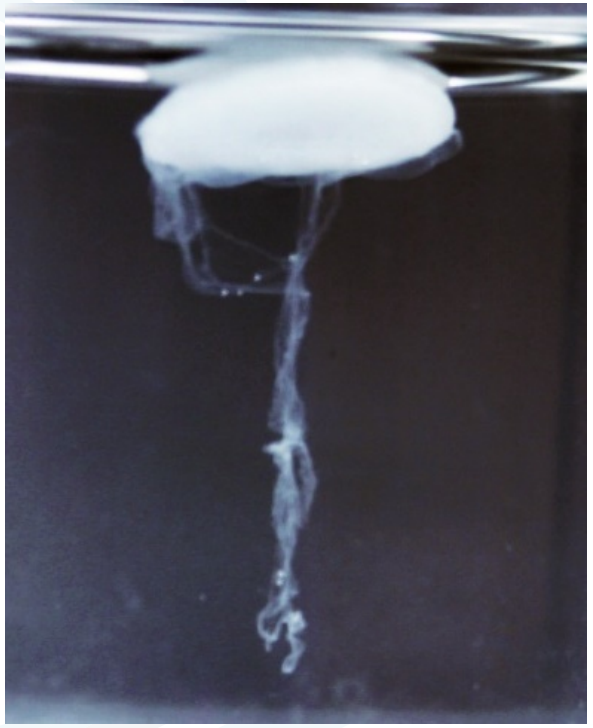
Just nanostructured silica (SiO_2), essentially.

With or without Iron layer for magnetism

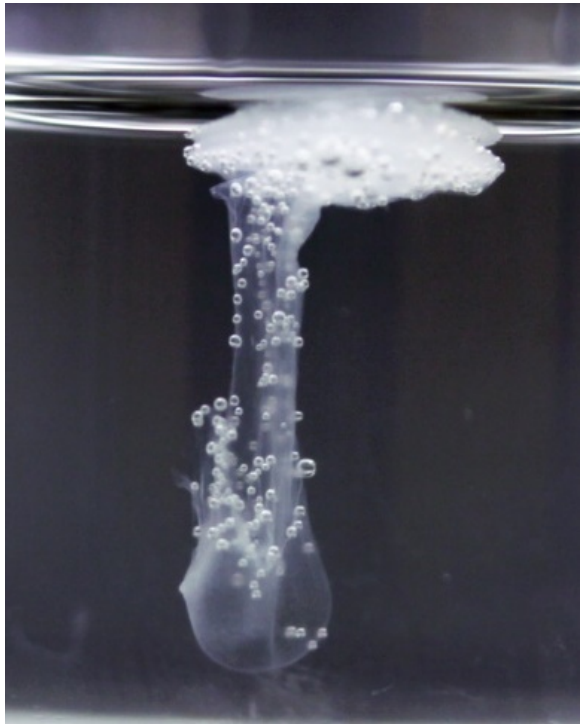
Zhang et al. Adv Mat. (2016)

UNIQUE TENTACLE BINDING MECHANISM

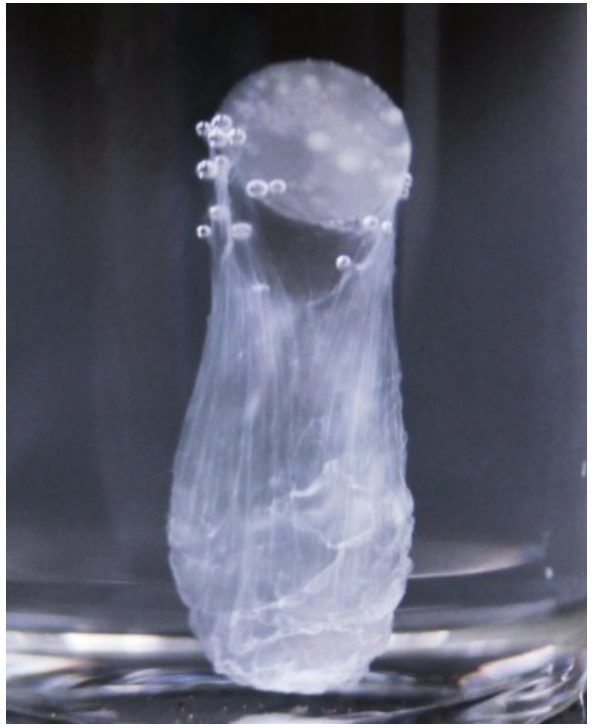
Enhances binding capacity and protects DNA from shear forces



Low Input (10 µg)



Medium Input (50 µg)



High Input (200 µg)

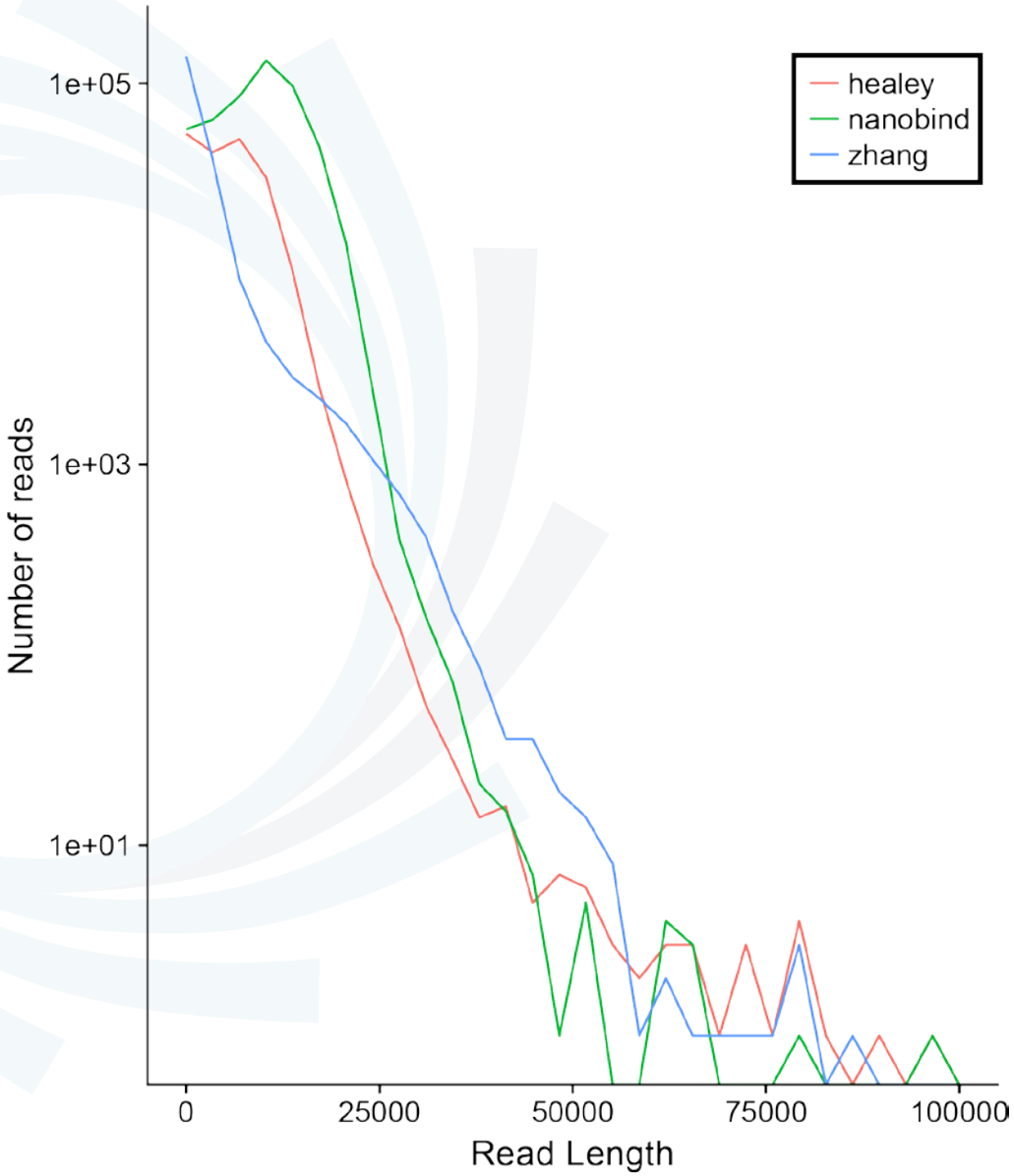
- Three material properties needed: low shear, non-porous, high surface area
- DNA tentacles form and extend from substrate to get high binding capacity
- Low shear unlike beads and columns

TOP 3 EXTRACTION PROTOCOLS: SUMMARY

From 1g input	Nanobind	Healey	Zhang
Yield (ug)	20	15	2
260/280	1.76	1.68	1.65
260/230	1.51	0.76	0.34

- Modified CTAB protocols (Nanobind and Healey) produce highest yield and Nanobind extraction produces the best quality extract with long fragment length.
- Nanodrop spectra not full picture of quality – plant samples sometimes retain visible color

EXTRACTION COMPARISON

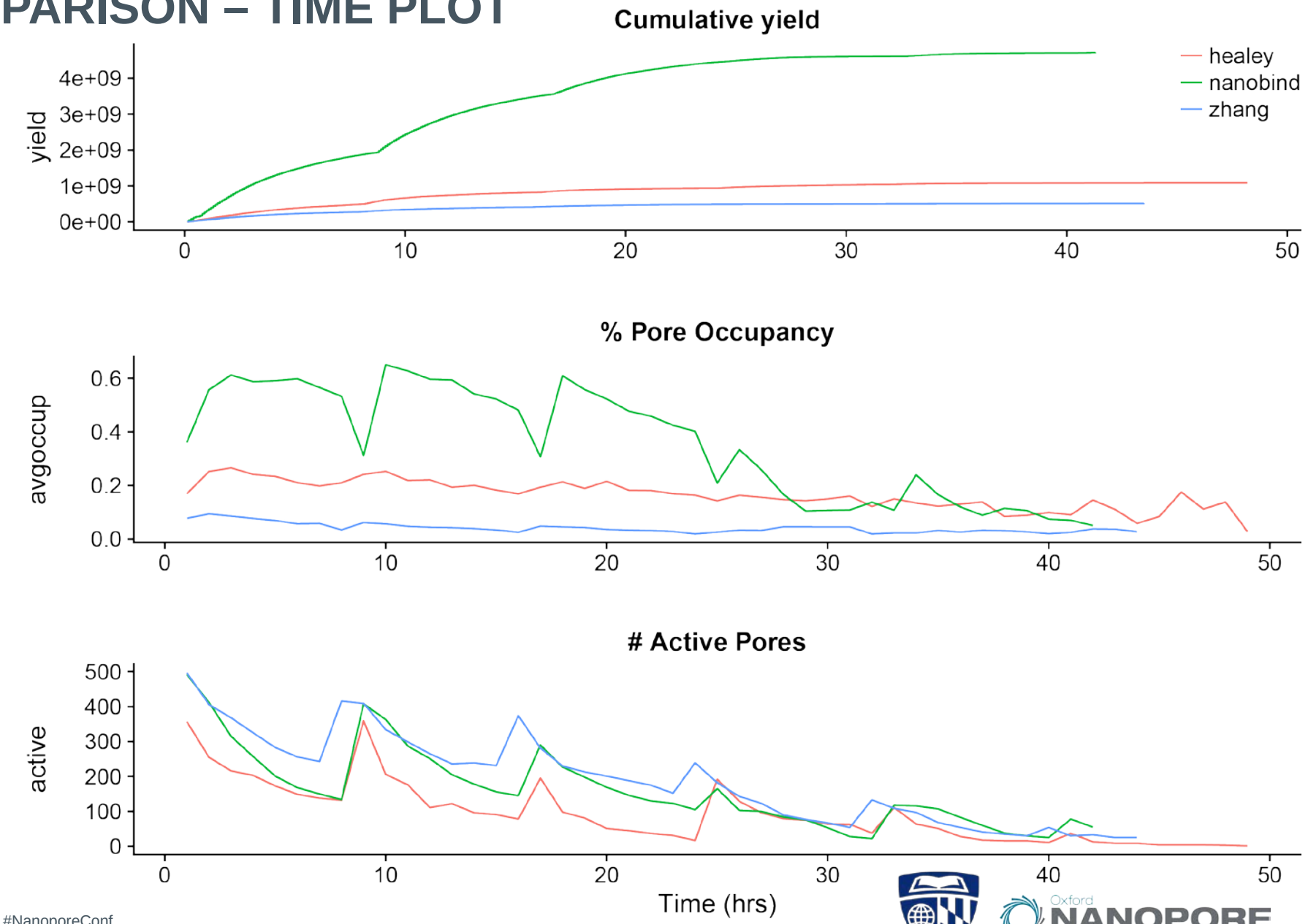


	Zhang	Nanobind	Healey
Reads	201k	500k	195k
Yield	0.51Gb	4.72Gb	1.08Gb
N50	7.1kb	12.3kb	8.6kb
Median	929	9.8kb	5.1kb

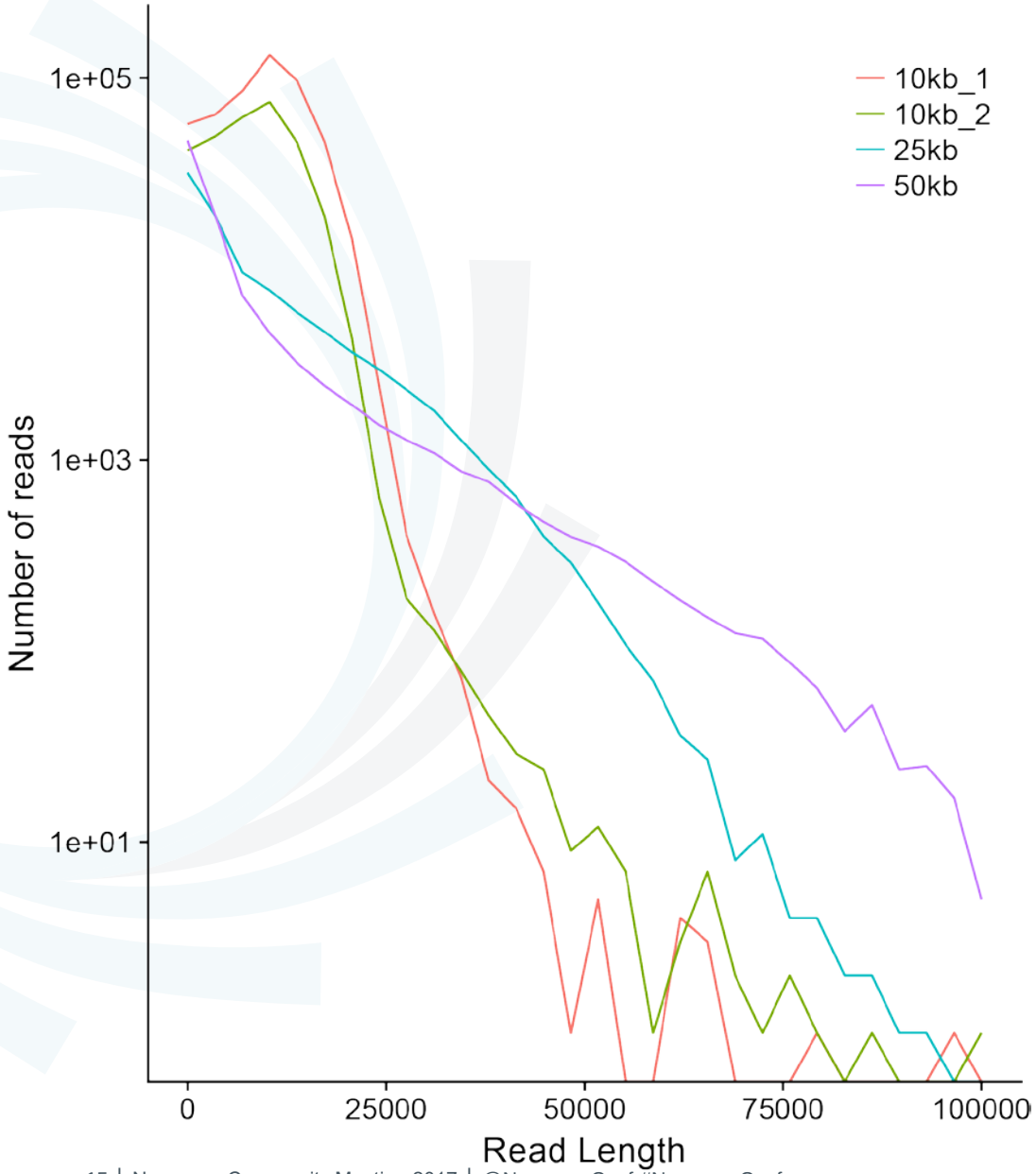
- Zhang seemed (in our hands) to fragment badly)
- Nanobind and Healey seemed to give reasonable read lengths, but doesn't match with PFGE profile size

EXTRACTION COMPARISON – TIME PLOT

Though ostensibly (from nanodrop and qubit) the same concentration, nanobind showed much higher pore occupancy, and resulting higher yield than Zhang or Healey



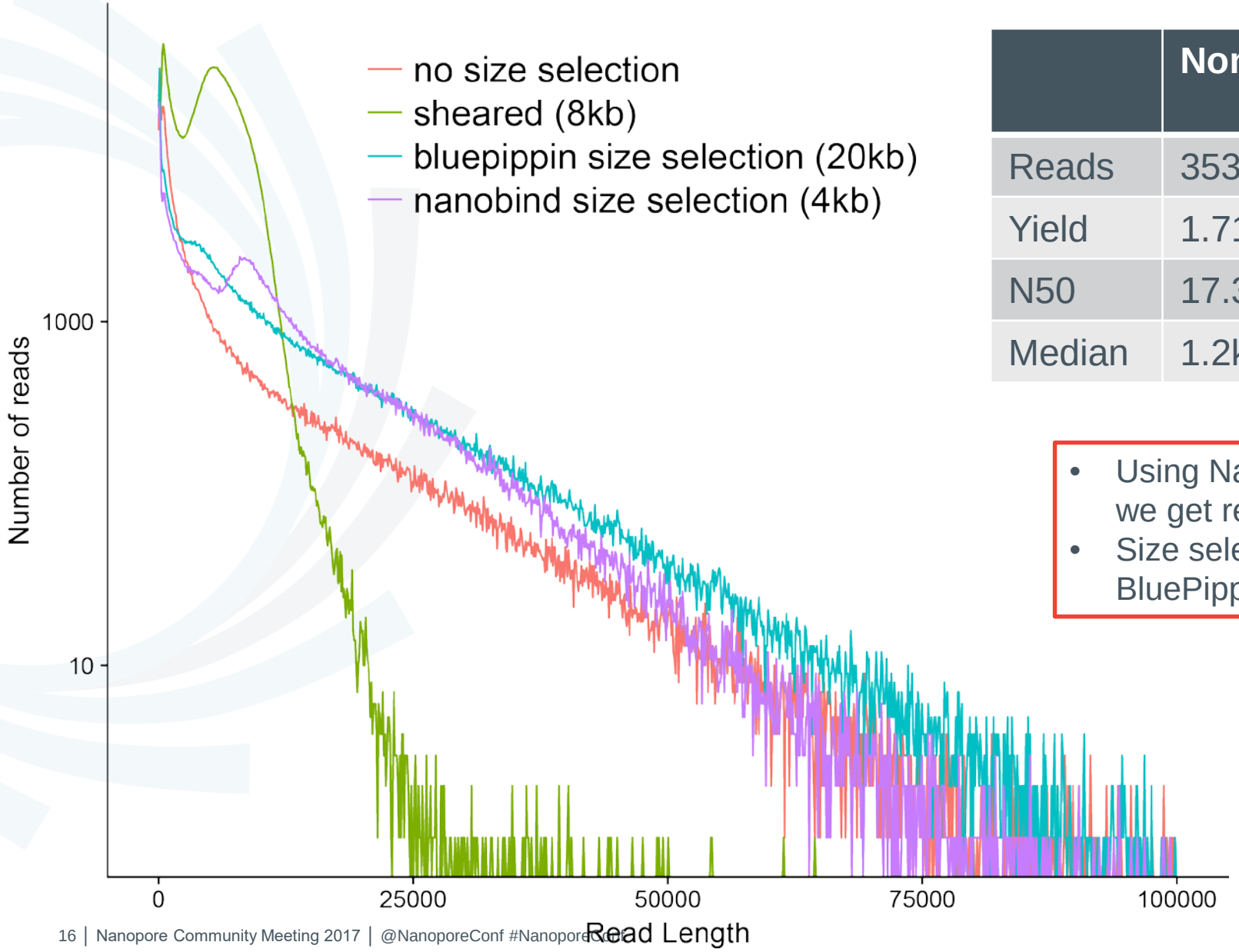
SHEAR COMPARISON



	10kb	10kb_2	25kb	50kb
Reads	500k	299k	93.7k	94.7k
Yield	4.72Gb	2.47Gb	0.82	0.66
N50	12.3kb	11.3kb	19.8	24
Median	9.8kb	8.5kb	4.1	1.7

- Currently a trade-off between long reads and high yield
- Possible to improve but would require more optimization

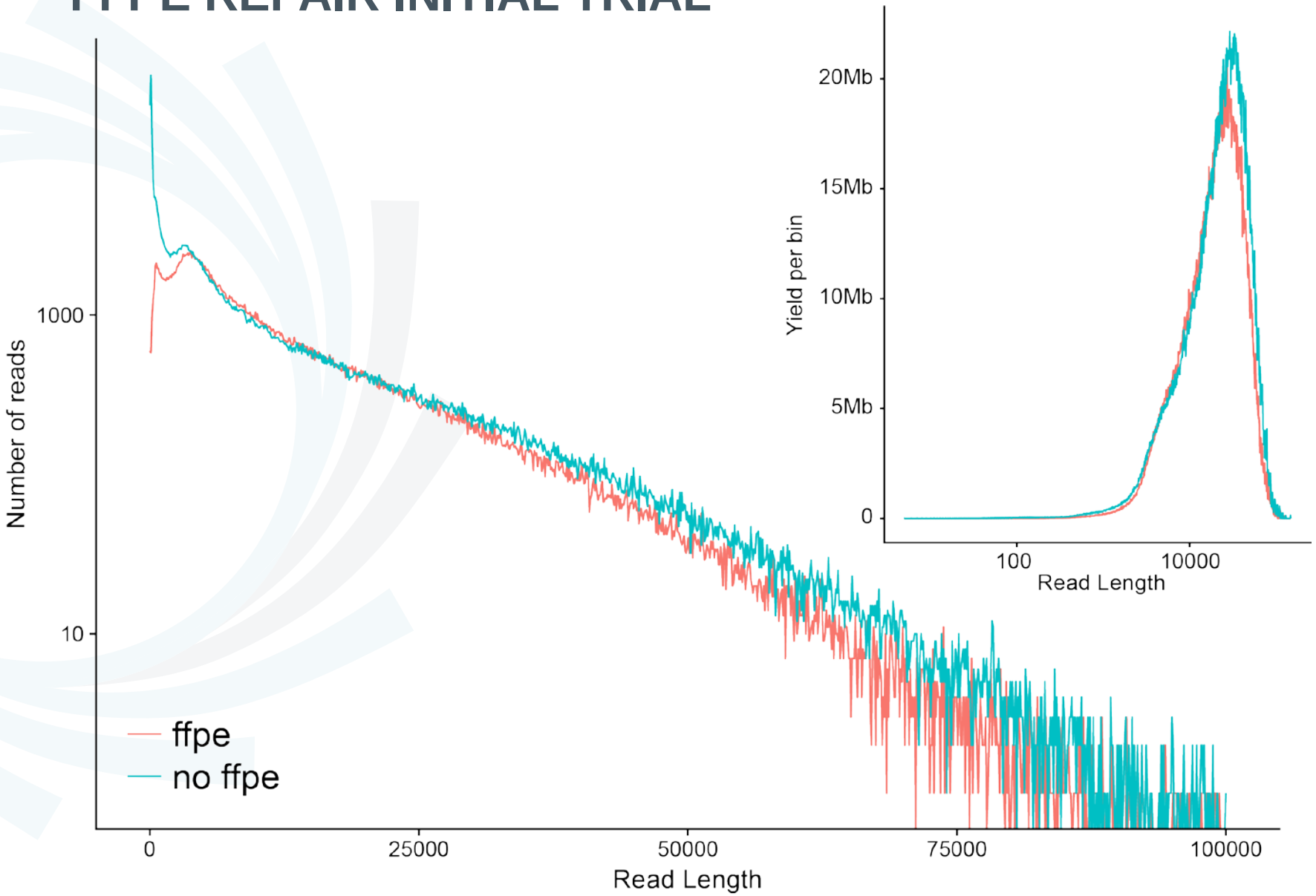
NANOBIND FOR SIZE SELECTION/AMPURE REPLACEMENT



	None	Sheared	BP (20kb)	NB (4kb)
Reads	353k	2060k	435k	400k
Yield	1.71Gb	10.1Gb	3.65Gb	3.57Gb
N50	17.3kb	6.6kb	19.0kb	15.7kb
Median	1.2kb	5.1kb	4.3kb	6.8kb

- Using Nanobind+PEG as an AMPure replacement, we get reasonable yield
- Size selection for long reads a little worse than BluePippin, but more >5kb reads

FFPE REPAIR INITIAL TRIAL



	Untreated	FFPE
Reads	385k	280k
Yield	3.69Gb	3.34Gb
N50	23.0kb	20.6kb
Median	4.5kb	7.7kb

- Adding FFPE step:
- FFPE reduced number of short reads, but did not improve HMW yield
- Resulting overall yield was lower

CONCLUSIONS

Extraction from plants is hard because of polyphenol compounds and polysaccharaides – rigorous purification is needed.

At the moment, nanopore sequencing yield is maximized with shorter fragment input. This is not merely a molarity issue.

- **Yield and median read length decrease drastically with 25kb and 50kb shearing relative to 10kb.**
- **Our best solution – target 10kb for high yield runs, then sprinkle in some long-read runs.**

We are still looking into nick repair and other likely methods to improve read length.

ACKNOWLEDGMENTS

- **Timp Lab**
- Rachael Workman
- Stephanie Hao

- **Salzberg Lab**
- Jennifer Lu

- **Neale Lab**
- Alison Scott
- Zane Moore

- **Circulomics**
- Kelvin Liu
- Duncan Kilburn
- Jefferey Burke
- Renee Fedak

Redwood Genome Project (Neale)

National Human Genome Research Institute
1R01HG009190-01A1 (Timp)
2R44GM109618-02 (Liu)