Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

An optimized toolkit for high molecular weight DNA extraction and ultra-long-read nanopore sequencing using glass beads and hexamminecobalt(III) chloride

Inswasti Cahyani, John Tyson, Nadine Holmes, et al.

Genome Res. published online March 24, 2025
Access the most recent version at doi:10.1101/gr.279943.124

P<P Published online March 24, 2025 in advance of the print journal.

Accepted Manuscript Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.

Open Access Freely available online through the *Genome Research* Open Access option.

Creative Commons License This manuscript is Open Access.This article, published in *Genome Research*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International license), as described at http://creativecommons.org/licenses/by-nc/4.0/.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

CRISPR and RNAi Genetic Screening.
Your new superpower.

To subscribe to *Genome Research* go to:
https://genome.cshlp.org/subscriptions

Published by Cold Spring Harbor Laboratory Press

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

An optimised toolkit for high molecular weight DNA extraction and ultra-long-read nanopore sequencing using glass beads and hexamminecobalt(III) chloride

Inswasti Cahyani[1], John Tyson[2], Nadine Holmes[1], Josh Quick[3], Chris Moore[1], Nick Loman[3], Matthew Loose[1]

[1]School of Life Sciences, University of Nottingham NG7 2UH, United Kingdom

[2]UBC, Dept of Pathology and Laboratory Medicine, Vancouver, Canada

[3]School of Biosciences, University of Birmingham B15 2TT, United Kingdom

Corresponding author: matt.loose@nottingham.ac.uk

Abstract

Since the advent of long read sequencing, achieving longer read lengths has been a key goal for many users. Ultra-long read sets (N50 $\geq$ 100 kb) produced from Oxford Nanopore sequencers have improved genome assemblies in recent years. However, despite progress in extraction protocols and library preparation methods, ultra-long sequencing remains challenging for many sample types. Here we compare various methods and introduce the FindingNemo protocol that: (1) optimises ultra-high molecular weight (UHMW) DNA extraction and library clean-up by using glass beads and Hexamminecobalt(III) chloride (CoHex), (2) can deliver high ultra-long sequencing yield of >20 Gb of reads from a single MinION flow cell or >100 Gb from PromethION devices (R9.4 to R10.4 pore variants), and (3) is scalable to using fewer input cells or lower DNA amounts, with extraction to sequencing possible in a single working day. By comparison, we demonstrate that this protocol surpasses previous methods by enabling precise determination of input DNA quantity and quality through cell counting, sample dilution, and homogenization techniques.

Introduction

The first ultra-long (UL) sequencing protocols on nanopore sequencing instruments resulted in highly contiguous human genome assemblies (Jain et al. 2018a; Quick 2018). Later methods improved on these assemblies through increased yields (Logsdon et al. 2021; Miga et al. 2020). Oxford Nanopore Technologies (ONT) subsequently launched UL sequencing kits, starting with ULK001 on R9.4 flow cells and more recently ULK114 on R10.4.1 flow cells. The ULK001 protocol utilised extraction and clean-up steps using silicon coated discs from Circulomics Inc. (PacBio

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

30 2023), which was subsequently replaced with the ONT star-shaped matrix for library precipitation in
31 the ULK114 kit. These UL kits introduced a range of protocol improvements both in library
32 preparation and sequencing performance to maximise read length and generated reads with an N50
33 > 100kb from both the MinION and PromethION flow cells.
34 In our experience, we have found that the most critical step in obtaining high throughput, high
35 occupancy UL sequencing libraries is the DNA fragmentation step, followed by the final clean-up
36 step prior to loading the library. The UL protocol uses a transposase enzyme complex
37 (transposome) to fragment DNA. During fragmentation, the DNA should be homogeneous and
38 accessible to this fast-acting transposome. Two methods are effective in reducing the number of
39 transposase cuts to increase read length: either using a high concentration of high molecular weight
40 (HMW) DNA as shown by Jain et al in sequencing the human genome (Jain et al. 2018a) or diluting
41 the transposase relative to DNA (Jain et al. 2018b). UL libraries are difficult to clean up using
42 conventional magnetic solid phase reversible immobilisation (SPRI) approaches as the DNA can be
43 hard to elute and so sheared during this; SPRI beads are mostly used for PCR clean-up where the
44 fragment length is much shorter (Rodrigue et al. 2010; Stortchevoi et al. 2020). In addition, the
45 library is a complex of motor protein and DNA, and so alcohol-based precipitation and washes are
46 not compatible. In our search for suitable alternative approaches, we found that hexamine cobalt (III)
47 (CoHex) cations stabilise and condense DNA and reasoned they may enable alcohol-free
48 precipitation of nanopore sequencing libraries (Allers and Lichten 2000; Kankia et al. 2001).
49 Early versions of the ONT ultra-long (UL) protocol required extraction and library clean-up starting
50 from 6 million human cells, or an equivalent input amount of DNA (PacBio 2023, Methods). The
51 complete extraction and library preparation protocol takes between two to three days in the
52 laboratory. However, different sample types often require tailored extractions to provide optimal DNA
53 for sequencing, so we sought to use a wide range of sample extraction protocols to obtain UL reads
54 on ONT platforms. We also explored reducing both sample input requirements and the total time for
55 library preparation for sequencing scalability.
56 Here we describe the FindingNemo protocol (named after the characteristic orange colour of the
57 CoHex buffer) for the generation of high occupancy ultra-long reads on nanopore platforms. This
58 protocol can generate equivalent or more throughput to disc-based methods and may have

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

59 additional advantages in tissues and non-human cell material. The protocol can also be tuned to
60 enable extraction from as few as one million human cell equivalents or 5 µg of human ultra-high
61 molecular weight (UHMW) DNA as input and enables extraction to sequencing in one working day.
62 The clean-up method can be used to generate ultra-long libraries from DNA extracted with phenol
63 chloroform, SDS lysis and CTAB approaches as well as commercial kits such as the NEB Monarch
64 HMW gDNA kits and those from Circulomics (now Pacific Biosciences). Oxford Nanopore
65 Technologies released similar protocols using a spermine-based precipitation buffer (PPT) instead
66 of CoHex and no DNA binding substrate (ONT, ULK001) and then incorporating a star (ONT,
67 ULK114) after our initial protocol release on protocols.io (Cahyani et al. 2021). We compare all
68 these methods here.
69

70 ## Results

71 ## The FindingNemo protocol compiles and optimises available ultra-long sequencing
72 ## protocols

73 ### *Previously developed protocols*

74 Both the standard and ultra-long (UL) sequencing workflows consist of three main steps: DNA
75 extraction, library preparation, and library clean-up prior to flow cell loading (Figure 1a). Ultra-long
76 sequencing necessitates minimal DNA shearing throughout the library preparation process and
77 during the extraction methods to maintain ultra-high molecular weight (UHMW) DNA. Early
78 experiments pioneering UL sequencing used the transposase-based rapid kit (ONT, e.g., RAD004)
79 to obtain reads with an N50 greater than 100 kb (Jain et al. 2018a). This method eliminated the
80 need for DNA shearing before library preparation. However, the yield from rapid runs was typically
81 low, which might be a consequence of the protocol itself. The lack of purification or clean-up of rapid
82 libraries prior to flow cell loading likely contributed to reduced yields compared to ligation-based
83 sequencing approaches. To better understand the complex interplay of DNA quality, library
84 preparation, read length, and yield, we tested a range of UL protocols aimed at maximising both
85 read length and yield (Figure 1B).

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

86 The goal in obtaining UL reads with the rapid kit is to ensure that each DNA molecule is cut only
87 once by a transposase complex, maximising read length. Initially, Quick et al achieved this by
88 saturating the transposase reaction with UHMW DNA (Quick 2018) (Table 1A). An alternative
89 approach taken by Logsdon et al. was to reduce the amount of transposase for a fixed amount of
90 DNA (Logsdon et al. 2021) (Table 1B). However, the sequencing output of this method, when
91 applied to the CHM13 cell line (Logsdon et al. 2021), was lower overall compared to the publicly
92 available nanopore data for GM24385 and GM24631 ligation-based libraries (Supplemental Figure
93 S1A). We hypothesized that this was most likely due to sub-optimal pore occupancy (i.e., the
94 number of available pores sequencing at any time), resulting from the reduced number of adapted
95 DNA ends available for sequencing due to the lower amount of transposase. In testing the Quick
96 protocol, we observed two different read length distributions (see Supplemental Figure S1B, C).
97 Whilst both had N50s of approximately 90 kb, one was dominated by shorter reads. This suggests
98 that the DNA was not cut uniformly. Moreover, although these protocols generated UL reads, they
99 suffered from high variability in output N50, occupancy and yield (Supplemental Figure S1B, C).
100 Typically, as N50 increases, yield and occupancy drop (Supplemental Table S1A, B).
101 The Rocky Mountain protocol (Tyson 2019) (Table 1C) was developed to use the ligation kit (ONT,
102 e.g. LSK109) to address the yield issues experienced with the rapid kit. This protocol utilised varying
103 concentrations of salts and polyethylene glycol (PEG) to precipitate longer reads after the library
104 preparation step. The ligation kit requires light shearing of the input DNA, resulting in non-UL N50s
105 albeit with much improved yields (Tyson 2019). Therefore, this protocol was further modified to use
106 the rapid kit followed by precipitation of the tagmented DNA using a filter paper disc or star as a form
107 of SPRI matrix (KrazyStarFish; Table 1D). The KrazyStarFish (KSF) protocol worked consistently in
108 producing UL N50s, but occupancies and yields were still suboptimal, perhaps a consequence of
109 having DNA clean-up prior to sequencing adapter ligation, consequently producing many free
110 adapters (Supplemental Figure S1D, E; Supplemental Table S1C, D; Methods).
111

112 *New and developing protocols*

113 ONT released a rapid kit for UL sequencing, SQK-ULK001 (hereafter ULK001) in which the
114 transposase reaction is performed in a large volume (Table 1E), keeping the DNA concentration at

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

115 approximately 50 ng/µl. We hypothesised that larger reaction volumes facilitated even diffusion and
116 mixing of DNA and transposase complexes in a more homogeneous state in solution, a feature that
117 we included in our Nemo protocols (Table 1F, G). Efficient transposase reactions result in consistent
118 UL sequencing output as shown by the high rates of occupancy (Supplemental Figure S1F-H, S2,
119 Figure 2, and Supplemental Table S1). Occupancy as a parameter is specifically defined by the
120 percentage of sequencing pores (i.e. pores in 'adapter' and 'strand' states) compared to all pores in
121 available states (i.e. 'adapter', 'strand' and 'pore' state).
122 We next tested the performance of RAD004 and ULK001 rapid-based kit protocols, modifying
123 parameters such as the number of input cells, the RAD004 dilution buffer, and the clean-up steps
124 whilst maintaining the large volume ratio, and compared these to the control ULK001 protocol
125 (Figure 2A). Regardless of clean-up method (Nanobind disc or glass beads), ULK001-based
126 protocol outputs were comparable in terms of N50s and occupancies (Figure 2B-E). Meanwhile, the
127 output of the library prepared with the RAD004 protocol showed slightly lower performance in
128 general (Figure 2B-E and Supplemental Table S1G).
129 The total yield of an UL sequencing is not a direct measure of the library quality as yield is also flow
130 cell dependent. The control ULK001 library produced the highest total yield (Figure 2B-C) likely
131 because of the highest pore count of the flow cell (Supplemental Table S1E, Supplemental Figure
132 S3A). Normalising yields between flow cells by taking a ratio of yield to the number of pores utilised
133 during a period of data collection (i.e. yield per pore) allows comparison of run performance. When
134 normalised, the yields per pore of the two ULK001 libraries were comparable (Supplemental Table
135 S1E, F). In all three libraries, the distribution of read lengths against read quality were similar
136 (Supplemental Figure S3B), as well as sequencing speed (Supplemental Figure S3C). These
137 similarities imply that either of the rapid-based kits could be used to successfully produce UL
138 sequencing outputs. Considering all sequencing output parameters, the ULK001 was the optimal
139 protocol for UL sequencing compared to the RAD004 (Supplemental Table S2).

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

140 ## The FindingNemo protocol efficiently purifies and accurately quantifies UHMW DNA
141 and UL libraries

142 *DNA precipitation and recovery: maintaining quality and homogeneity*

143 Prior to the launch of the ULK001 kit we developed a protocol to clean the library after adapter
144 ligation using DNA precipitation chemistry compatible with the motor protein-DNA complex. The filter
145 paper from the KSF protocol was replaced with glass beads, adapting the glass bead's effectiveness
146 to act as a DNA-binding substrate as in the Monarch extraction kit protocol (NEB, T3050). We also
147 surveyed compounds with well-studied DNA precipitating (Ouameur and Tajmir-Riahi 2004; Deng
148 and Bloomfield 1999) and selected spermine and hexamminecobalt(III) chloride (CoHex) for testing
149 DNA precipitation and recovery profiles using glass beads (Supplemental Figure S4A). Optimum
150 final concentrations of the compounds were tested empirically based on previous studies (Pelta et
151 al. 1996).

152 The ULK001 protocol is carried out in a dilute condition of ~50 ng/µl UHMW DNA. We tested the
153 impact of DNA concentration (10, 20 and 50 ng/µl) on precipitation and recovery rate. CoHex is the
154 best precipitant for both recovery and homogenisation of UHMW DNA at a concentration range of 20
155 ng/µl to less than 50 ng/µl (Supplemental Figure S4B-D). The optimal DNA recovery of ~75% was
156 obtained using CoHex at DNA concentrations of 20 ng/µl. Spermine shows a significantly higher
157 recovery rate at 50 ng/µl DNA but greater heterogeneity and sample-to-sample variation
158 (Supplemental Figure S4B-D).

159 The early ULK001 kit version used a spermine-based precipitation buffer (PPT buffer). We tested
160 this alongside our FindingNemo and spermine clean-up protocols, using a DNA input of 40 ng/µl,
161 and split the adapted library into three clean-up preparations. These runs yielded similar UL
162 sequencing metrics (Figure 3A). Albeit with a slightly longer N50, ONT's PPT protocol showed the
163 lowest yield per pore, presumably from the lack of any washing step after the library precipitation
164 that might cause the library to have a higher rate of blocking than the others (Figure 3B).

165 Homogeneity is measured by calculating the coefficient of variation percentage (%CV) of the DNA
166 concentration. We take at least three concentration measurements from the top, middle and bottom
167 part of the solution using a NanoDrop™ spectrophotometer device. The smaller the %CV, the more

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

168 homogeneous, and vice versa. Recovered DNA became more homogeneous after longer in
169 solution, except when precipitated using CoHex at 10 ng/µl DNA concentration (Supplemental
170 Figure S4E). Previous studies showed that a charge neutralisation of 88-90% is required for DNA
171 condensation (Deng and Bloomfield 1999). We hypothesise that at 10 ng/µl DNA, excess CoHex
172 cations affected the water hydration at the DNA-cations interface (Deng and Bloomfield 1999)
173 resulting in a more condensed, difficult-to-dissolve DNA as observed during pipetting. Moreover,
174 NanoDrop is more sensitive when measuring absorbance of the differentially condensed DNA in the
175 solution which explains the larger variation in DNA concentration compared to Qubit™
176 measurements (Supplemental Figure S4C-D). In contrast, spermine-precipitated DNA shows the
177 least homogeneity in solution (Supplemental Figure S4E). Spermine is also less stable than CoHex;
178 as a solution it is readily oxidised at room temperature (sigmaaldrich.com). CoHex was chosen to
179 precipitate DNA at a concentration range of 20-40 ng/µl UHMW DNA.
180 The importance of DNA homogeneity and optimum concentration during UL library preparation was
181 shown in the use of a concentrated and non-homogeneous UHMW DNA in sequencing, i.e. more
182 than 60 ng/µl DNA in reaction (Supplemental Figure S5). The sequencing output of this library
183 typifies the output of an undercut library where N50 reaches above 100 kb, but accumulated yield
184 and occupancy are low, as well as a rapid decrease of sequencing pore number (Supplemental
185 Figure S5D). In essence, ensuring the homogeneity of extracted UHMW DNA and setting the
186 reaction concentration below 50 ng/µl provide the optimum conditions for UL sequencing. Longer
187 elution of the library (e.g. overnight) at room temperature with gentle rotation may also help untangle
188 long DNA molecules, facilitating further homogenisation and eventually improving UL sequencing
189 output.
190
191 *Quantification of UHMW DNA*
192 One of the complications in UL sequencing is in accurately quantitating the UHMW DNA. There is
193 often large variation between the two most used quantification methods: the spectrophotometric and
194 fluorometric based methods, represented by NanoDrop™ and Qubit™ 3 respectively (Supplemental
195 Figure S4F). The viscous nature of UHMW DNA also makes it difficult to withdraw a homogeneous
196 small volume to be measured, even when using a positive displacement pipette. Therefore, we

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

197 followed a novel Qubit concentration measurement of UHMW DNA to complement the standard
198 NanoDrop method. This method uses Jurkat genomic DNA as the baseline standard for UHMW
199 DNA and a glass bead to homogenise samples before each measurement (Koetsier and Cantor
200 2021). We modified it by combining small volumes from 3-4 different parts of the DNA solution to
201 average out concentration differences within the solution (Cahyani et al. 2021). An accurate
202 concentration measurement requires an ideal ratio of NanoDrop to Qubit DNA values between 1-1.5
203 (Simbolo et al. 2013).
204 As expected, sheared DNA homogenised faster and better than UHMW DNA as shown by the
205 smaller %CV of the sample a few hours after elution (Supplemental Figure S6A). This fits with the
206 fact that shorter DNA molecules will diffuse faster than longer ones. Longer DNA molecules may
207 also have more complex structure and behaviour in solution (Robertson and Smith 2007). The %CV
208 of both types of samples were all below 50% indicating that the FindingNemo method could
209 homogenise DNA solution relatively quickly (Supplemental Figure S6A). It also filtered out
210 contaminants in the input DNA as shown by the increase of the NanoDrop absorbance values after
211 purification (Supplemental Figure S6B) (Koetsier and Cantor 2019). Lastly, the FindingNemo
212 protocol did not interfere with the DNA quality as shown by their size distribution on the pulsed-field
213 electrophoresis gel before and after DNA clean-up (Supplemental Figure S6C).

214 *The* FindingNemo *protocol vs undercut UL libraries*

215 Combining the FindingNemo approach with either the Quick's or KSF protocols improved
216 sequencing yield but did not significantly improve occupancy or N50 (Supplemental Table S3). We
217 hypothesise that the transposase reaction upstream of the library clean-up is a more critical step in
218 producing optimal ultra-long sequencing outputs. Suboptimal transposase fragmentation may result
219 in long DNA molecules without adapters in the library, due to an insufficient number of transposome
220 complexes compared to DNA molecules, a situation we refer to as an 'undercut'.

221 An undercut library may also occur if the transposome complex-to-DNA ratio is balanced, but the
222 DNA solution is not homogeneous. Viscous, non-homogeneous UHMW DNA is difficult to dilute and
223 can have varying DNA densities at different points in the solution. Consequently, the cutting
224 frequency of the transposome complex will be higher where DNA is more accessible and lower

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

225 where long DNA molecules remain in a compacted, less accessible mass. This scenario
226 counterintuitively results in a library with a short N50 despite the high molecular weight nature of the
227 input DNA (Supplemental Figure S8 - 0 rpm sample).

228 It is also important to consider the interaction between UHMW DNA molecules and the sequencing
229 pores, especially just before a mux scan occurs. During this period, only pores with actively
230 threading (sequencing) DNA molecules will continue sequencing; pores without DNA are
231 inactivated. This process terminates at a fixed time, which is 10 minutes by default. After this period,
232 all pores are inactivated. Molecules still traversing a pore at that point will be ejected, but UHMW
233 DNA increases the risk of stalling or blocking nanopores. At a sequencing speed of 400 bases per
234 second, any molecule longer than 240 kb is at risk of blocking a pore and negatively impacting yield.
235 Therefore, in the FindingNemo protocol, we modified this script parameter (“wind_down”) from 600
236 to 1800 seconds (Cahyani et al. 2024).

237

238 The FindingNemo protocol in UHMW DNA extraction

239 *Extractions without kits*

240 We next compared the effect of DNA extraction methods on UL sequencing. The original Quick et al.
241 and Logsdon et al. approaches utilised phenol-chloroform isolation and alcohol-based precipitation
242 with some differences (Table 2A, B). We succeeded in scaling-down these protocols in combination
243 with glass beads to use less cell input and shorten hands-on time to at least half (Table 2C). To
244 reduce the toxicity from phenol use, we also developed phenol-free glass beads-based extraction
245 protocols utilising either SDS, CTAB, or CTAB with CoHex as the lysis buffer (Table 2D, E). These
246 extraction protocols produced (U)HMW DNA, characterised by the distinctly narrow vertical
247 lines/smears and DNA in the wells of the pulsed-field gel electrophoresis results (Supplemental
248 Figure S7).
249 We also successfully produced UL runs using these DNA samples with varying degrees of
250 occupancies and yields (Figure 4A-J). Only the extraction protocol using a combination of CTAB and
251 CoHex in the lysis buffer could not produce UL N50 (Figure 4F). The inclusion of CoHex during lysis
252 was to test if stable DNA toroid aggregates could be induced (Deng and Bloomfield 1999; Ouameur

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

253 and Tajmir-Riahi 2004), so that DNA length could be kept intact during extraction and eventually
254 improve sequencing output. The CoHex precipitated DNA was then washed with ethanol, and it
255 rendered DNA elution from the beads more difficult. Ethanol is required in the wash buffer as it
256 rinses salt, contaminants and other impurities from the cell extract while simultaneously affecting
257 DNA condensation (Arscott et al. 1995; Oda et al. 2016). More optimisation is needed to obtain a
258 consistent rate of elution and recovery in the CTAB-CoHex lysis protocol. Nevertheless, the option
259 of using one of two oppositely charged surfactants as lysis agents, i.e. cationic CTAB (Arseneau et
260 al. 2017) and anionic SDS (Xia et al. 2019), provides flexibility in extracting UHMW DNA from
261 diverse sample types and context.

262 *Extraction using kits*

263 We tested a range of (U)HMW DNA extraction kits during our method development and found that
264 Nanobind CBB Big DNA (Circulomics) and Monarch (NEB) were best suited for our purposes (Table
265 2F-G). The Nanobind CBB Big DNA Circulomics kit could be used to extract HMW or UHMW DNA,
266 depending on the chosen protocol. DNA extracted with the UHMW protocol showed higher size
267 distribution on a pulsed-field electrophoresis gel compared to the HMW protocol (Supplemental
268 Figure S7A vs S7B). Libraries prepared from UHMW DNA samples extracted using the Circulomics
269 kit produced UL reads and good sequencing metrics (Figure 4I-J).

270 The Monarch kit protocol initially involved a two-step DNA extraction: first extracting nuclei from
271 cells, then isolating DNA from the nuclei ("Nuclei Prep"). It was later refined to a "Direct Lysis"
272 method, which bypassed the "Nuclei Prep" step by lysing cells and nuclei simultaneously to release
273 DNA. This method yielded high-quality UHMW DNA from GM12878 cells, as shown in pulsed-field
274 gel samples (Supplemental Figure S7A, B). Libraries prepared from these DNA samples indicated
275 that the Direct Lysis approach performed slightly better than the Nuclei Prep method (Supplemental
276 Figure S8).

277 We also found that the shaking speed during lysis in the Monarch protocol influenced DNA
278 homogeneity, which affected sequencing N50 and occupancy (Supplemental Figure S8; Figure 4G-
279 H; Supplemental Table S4). Lower lysis speeds yielded longer DNA fragments by keeping DNA
280 more intact. However, no shaking during lysis could result in a non-homogeneous UHMW DNA
281 solution, negatively affecting the transposase reaction during library preparation (Supplemental

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

282 Figure S8 - 0 rpm). This resulted in a library with non-UL read N50, the lowest occupancy, and yield
283 compared to other libraries (Supplemental Figure S8). The best UL output was achieved by
284 preparing libraries from DNA lysed at a medium speed of 600 to 900 rpm (Supplemental Figure S8;
285 Figure 4G-H; Supplemental Table S4).

286 Sequencing parameters and performance of the FindingNemo protocol

287 *Maximum yield is obtained at a read N50 between 90-110 kb*

288 On a MinION, yields in excess of 3 Mb per pore after 6 hours of sequencing could be obtained at
289 93% occupancy, reinforcing that high occupancy is essential for high yield (Figure 4J). However,
290 UL libraries with many short fragments had reduced occupancy as short reads pass quickly
291 through the pores, lowering total yields (Figure 4B, D, E, J). This underscores the importance of
292 homogeneously cut UHMW DNA in UL library preparation. Additionally, targeting an N50 between
293 90-110 kb is ideal for optimal ultra-long sequencing output, as N50 is inversely proportional to yield
294 (Figure 4J; Supplemental Table S4)

295 *Scalable library loading amount with the* FindingNemo *protocol*

296 We also tested the impact of the amount of input DNA and library loaded per flow cell. When the
297 input amount of UHMW DNA was less than 5 μg, sequencing output parameters were markedly
298 reduced (Supplemental Table S4, Monarch-01 vs Monarch-05). To produce the optimum N50,
299 occupancy and yield, DNA equivalent to at least one million GM12878 cells (or at least 5 μg)
300 should be used for library preparation, and a minimum of 2 μg of this library loaded onto a MinION
301 flow cell. At least two million cells and 4 μg of library should be used on a PromethION flow cell
302 (Supplemental Table S4, Monarch-01 vs Monarch-02).

303 FindingNemo *in one day: from cells to UL sequencing in one working day*

304 The relatively fast homogenization rate of the FindingNemo clean-up method, combined with the
305 fast DNA extraction when using the Monarch kit can support the preparation of UL libraries from
306 fresh or frozen cells to sequencing in just a working day (Supplemental Figure S9A) (Cahyani
307 2021). We can obtain UL N50s using this protocol, however yields are markedly reduced
308 (Supplemental Figure S9B-D), likely as a result of the heterogeneity of DNA samples. Yield
309 reductions were exacerbated by the low occupancy level and the abundance of short reads and
310 would be interpreted as undercut libraries (Supplemental Figure S9B-D; Run 2 and 3). Moreover,

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

311 the FindingNemo clean-up protocol might not significantly remove these highly abundant short
312 reads due to their high concentration in the libraries.
313

314 Using the FindingNemo protocol in diverse sample types, chemistry and platforms

315 *The FindingNemo protocol improved N50, occupancy and yield with ULK114 kit*

316 In 2022, ONT released a new pore and chemistry running on the R10.4.1 flow cells as well as a
317 new ultra-long kit, ULK114. This included new sequencing scripts and a sample rate shift from 4
318 kHz to 5 kHz. We tested the ULK114 kit according to the manufacturer's instructions and with the
319 FindingNemo protocol on three human cell lines (Figure 5). Libraries prepared with the
320 FindingNemo protocol showed longer N50s, optimal yield, and occupancy (Figure 5A-B). This
321 protocol consistently produced higher yields, regardless of kit chemistry and sequencing programs.
322 Additionally, the 5 kHz script increased the overall mapping rate of reads compared to the 4 kHz
323 program (Figure 5B).

324 *Ratio of transposase to DNA*

325 The ratio of transposase to DNA is crucial for optimising UL sequencing. The ULK001 protocol
326 specifies 1 µl of transposase fragmentation mix (FRA) per 1 million human cells, corresponding to
327 5-6 µg of UHMW DNA for a 3.2 Gb genome (Table 3). We used this ratio for all GM12878 (Figures
328 2-5) and HEK293 (Figure 6) runs. Additionally, we tested different FRA to DNA ratios in Mongolian
329 gerbil, tick *Amblyomma variegatum*, and yeast *Saccharomyces cerevisiae* (Table 3). These ratios
330 produced UL sequencing outputs for the gerbil and tick samples (Figure 6A, C). Although the yeast
331 run did not yield UL read N50s (Figure 6B, C), it showed a significant increase in read length
332 compared to existing nanopore-sequenced data (Giordano et al. 2017; Salazar et al. 2017).

333 *Genome ploidy*

334 Genome ploidy anomalies must be considered when preparing UL libraries. For instance, about
335 4.2% of HEK293 cells are hypotriploid (Synthego HEK293). This means the DNA mass extracted
336 from the same number of cells is higher than from the GM12878 cell line, altering the FRA to DNA
337 ratio (Table 3) and potentially affecting sequencing output by increasing read length and
338 decreasing yield. Sequencing HEK293 DNA with the Monarch kit confirmed this, producing N50s
339 over 190 kb but with lower yields (Figure 6B, C). To maximise yield and N50, it is better to use the

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

340 FRA volume to DNA mass ratio rather than the absolute cell number when genome ploidy
341 information is available.

342 *Sample types: to count or to weigh?*

343 The UL protocol requires the same number of human cells per microliter of FRA, whether from a
344 cell line or nucleated blood. However, counting cells in blood samples can be impractical, as seen
345 in the gerbil UL run (Table 3). In such cases, the FRA to DNA mass ratio can be used, provided the
346 DNA is extracted to maintain native chromosome length and properly quantified (Methods). Using
347 this ratio, UL libraries were produced from gerbil blood following the original ULK001 protocol
348 (Figure 6B, C).

349 *The FindingNemo protocol works on different nanopore platforms*

350 We ran pairwise sequencing of identical libraries on both the MinION and PromethION platforms,
351 showing that PromethION produced more yield per pore (Supplemental Table S5). With nearly six
352 times as many pores as a MinION, PromethION is expected to generate more data. The best
353 PromethION results had an N50 of 126 kb and a yield of 3.85 Mb per pore at 88% occupancy
354 (Figure 6C - Tick cells ULK001-Nemo). On the MinION platform, the best results ranged from 106-
355 147 kb N50, with yields of 3.19-3.89 Mb per pore and at least 95% occupancy (Figure 4J). These
356 metrics were all obtained after 6 hours of sequencing.

357 *Relative performance of extraction kits in tick samples*

358 We compared two extraction kits, Circulomics and Monarch, for UL sequencing of a tick cell line
359 with a ~6 Gb genome (Figure 6A, C). Although flow cell occupancies were similar, the Monarch-
360 extracted DNA had longer N50s, higher yields, and a greater proportion of UL reads (Figure 6A
361 right panel, C). The Circulomics library had more short reads, likely due to undercutting (Figure 6C
362 middle panel). The Circulomics-extracted DNA concentration was ~50 ng/µl, while Monarch-
363 extracted DNA was ~26 ng/µl, reflecting differences in sample homogeneity. Moreover, Monarch
364 extractions yielded DNA with a longer size distribution, as shown by pulsed-field electrophoresis
365 gel (Supplemental Figure S10). The Monarch protocol is also more flexible, allowing modifications
366 to target different read-length distributions by adjusting lysis speed (Methods).

367

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

368 ## Discussion

369 High-quality, uniformly long DNA is essential for ultra-long reads on Oxford Nanopore sequencers.
370 Our results support the hypothesis that to obtain this length uniformity, transposase reactions
371 should occur in a dilute and homogeneous DNA solution. This can be achieved in a large reaction
372 volume and by thorough mixing to obtain a 'properly cut' library in contrast to 'undercut' as
373 explained before, or 'overcut' which is a higher ratio of FRA to DNA molecules (Figure 7A-B).
374 Additionally, non-adapted DNA molecules may 'crowd' the space around the pores thereby
375 decreasing the diffusion rate of the adapted DNA molecules to be tethered to the pores. The
376 adapted long DNA molecules themselves may inactivate or block the pores during sequencing
377 because of their lengths.
378 With these challenges, control of DNA quality and quantity is of primary importance. Proper
379 quantification of the number of input cells or tissue weight has direct consequence to controlling
380 DNA yield and to some extent, length distribution. Extended elution and homogenisation are also
381 beneficial; the longer that an UHMW DNA sample is allowed to equilibrate, the more homogeneous
382 it becomes. This reduces the benefit of our FindingNemo in One Day protocol where both maximal
383 read lengths and yields are important. However, as we show, a rapid UL library preparation is
384 possible, albeit with lower total yield. Further optimisation is required, especially at the
385 homogenization of DNA post-extraction, to make this protocol more robust.
386 Chromosome length distribution of a genome may limit the maximum read N50 that can be
387 obtained, as was the case with the yeast *S. cerevisiae* ultra-long library. Prior knowledge of the
388 genome size and chromosome lengths combined with empirical tests may be required to find the
389 optimum ratio of transposase to DNA amount and produce maximum sequencing N50 and yield.
390 This is a challenging part of ultra-long sequencing as it needs fine tuning and adjustments
391 depending on the sample types used.
392 We anticipate that the FindingNemo protocol will be generally applicable to tissue samples, could
393 be readily multiplexed, and is likely compatible with automation. Finally, open-source development
394 and sharing of protocols is crucial to enable the widest access to cost-effective high throughput
395 ultra long sequencing and troubleshooting. We established the LongRead Club on protocols.io to
396 house these protocols and promote reproducibility (https://www.protocols.io/workspaces/long-read-

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

397 club/about). The FindingNemo protocol described here is a valuable addition to the repertoire of
398 ultra-long nanopore sequencing methods available.
399

400 ## Methods

401 ## Genomic DNA extraction from cells

402 ### *Cell sources*

403 GM12878 cells were grown by Darren Crowley (University of Nottingham) and additionally
404 purchased from Coriell Institute for Medical Research, USA. Yeast *S. cerevisiae* cells were given
405 by Stephen Gray (University of Nottingham). Tick *A. variegatum* (AVL/CTVM17) cells were
406 obtained from Alistair Darby's group (University of Liverpool). The HG00233 and HG04054 cells
407 were obtained from Danny Miller's lab (University of Washington, USA). Lastly, HEK293 cells were
408 obtained from New England Biolab (NEB) in Ipswich, MA, USA.

409 ### *Kit-based extraction*

410 Kits used in the extraction of UHMW DNA were Monarch® HMW DNA Extraction Kit for Cells &
411 Blood (NEB T3050) and Nanobind CBB Big DNA Kit (Circulomics SKU NB-900-001-01) plus the
412 auxiliary Nanobind UL Library Prep Kit (Circulomics SKU NB-900-601-01) according to the
413 manufacturer's protocols with modifications as previously described for the Monarch protocol
414 (Cahyani et al. 2021). The Monarch nuclei prep approach followed the original protocol (NEB
415 T3050) while the direct lysis approach combined the prep and lysis buffers into one step (NEB
416 Direct Lysis). Lysis speed at 600 up to 900 rpm resulted in an optimum UL sequencing output and
417 a minimum of overnight incubation of DNA in solution helped homogenization.

418 ### *Phenol-chloroform extraction with glass beads*

419 This is a scaled-down version of Quick's original protocol (Quick 2018) and the modifications are
420 as previously described (Cahyani et al. 2021). In summary, pellet of 5 million cells was washed
421 with PBS [Fisher Scientific #15453819] and lysed with an SDS buffer [0.5% SDS, 100mM NaCl,
422 25mM EDTA pH 8.0, 10mM Tris-HCl pH 8.0] in the presence of RNase A [Qiagen #19101, 20
423 µg/ml] at 37 ^{o}C for 5 min. Proteinase K [Qiagen #19131, 200 µg/ml] was added and incubated at
424 56 ^{o}C for 15 min. Cell lysate was split into two 5PRIME phase-lock gels (Quanta-Bio 2302820).
425 BioUltra TE-saturated phenol (Merck 77607) was added at 1:1 volume ratio. Sample was

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

426 homogenised by vertical rotation at 20-30 rpm for 10 min, then centrifuged at 4000 × g for 10 min.
427 The aqueous phase was transferred into new phase-lock gel tubes. DNA purification was repeated
428 with a second round of phenol:chloroform:isoamyl alcohol = 25:24:1 volume ratio. In the third
429 round, only chloroform:isoamyl alcohol=24:1 was used. After the last centrifugation, the aqueous
430 phase from the two tubes was combined in a 5 ml or 15 ml tube and added with 5M ammonium
431 acetate [Sigma-Aldrich #A-7330] (0.4 volume), 3 glass beads (3-mm diameter), and absolute
432 ethanol (2.5 volume). The tube was inverted by hand 20-30 times to precipitate DNA onto the glass
433 beads (or placed in a tube rotator at 10 rpm for 3 min). The supernatant was removed, and bound
434 DNA was washed twice with 60-70% ethanol. Beads with DNA were poured into a bead retainer
435 and quickly spun to remove excess ethanol (or absorbed with a filter paper or tissue). DNA was
436 quickly eluted from beads by pouring them into a 2 ml LoBind [Eppendorf #0030108051] tube
437 containing 10mM Tris-HCl pH 9.0 and incubating at 37 °C for 30 min with regular wide-bore pipette
438 mixing. Incubation was continued overnight at room temperature. Afterwards, using a bead
439 retainer, DNA was spun down at maximum speed for 1 min.

440 *Phenol-free, home brew extractions*

441 This protocol is adapted from the Monarch® HMW DNA Extraction Kit for Cells & Blood as
442 described previously (Cahyani et al. 2021). In summary, a pellet of 1-3 million cells was washed
443 with PBS and lysed in either SDS or CTAB buffer with 200 µg Proteinase K and incubated at 56 °C
444 for 15 min (preferably with shaking at 600-700 rpm). The SDS buffer composition was the same as
445 used in the phenol-based extraction described in the previous section. The CTAB buffer
446 compositions were 2% CTAB, 1.4M NaCl, 25mM EDTA pH 8.0, and 10mM Tris-HCl pH 8.0. After
447 lysis, 100 µg RNase A was added and incubated at 37 °C for 5 min. To precipitate DNA, 5M
448 ammonium acetate was added (0.4 volume), 2-3 glass beads (depending on cell number used),
449 and isopropanol (0.9 volume). Sample was mixed on a rotator at 9 rpm for 5 min (or inverted by
450 hand 20-30 times). Liquid was removed by pipetting and bound DNA was washed twice with 60-
451 70% ethanol. Beads with DNA were poured into a bead retainer and quickly spun to remove
452 excess ethanol (or absorbed with a filter paper or tissue). DNA was quickly eluted with 10mM Tris-
453 HCl pH 8.0 in a 2 ml tube and incubated at 37 °C for 30 min with regular wide-bore pipette mixing.

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

454 Incubation was continued overnight at room temperature. Afterwards, using a bead retainer, DNA
455 was spun down at maximum speed for 1 min.

456 *CTAB-CoHex extraction*

457 This protocol is similar to the home-brew CTAB method described above with the following
458 modifications. The CTAB-CoHex buffer compositions used were 2% CTAB, 20mM CoHex, 1.4M
459 NaCl, 25mM EDTA pH 8.0, and 10mM Tris-HCl pH 8.0. To precipitate DNA, sample was diluted
460 with 10mM Tris-HCl pH 8.0, so CoHex concentration became ~5mM and NaCl was ~400mM, *i.e.*
461 ratio of CTAB-CoHex lysis buffer to Tris-HCl was 1:2.5. Then, 2-3 glass beads (depending on cell
462 number used) were added. Sample was mixed on a rotator at 9 rpm for 5 min (or inverted by hand
463 20-30 times). Liquid was removed by pipetting and bound DNA was washed twice with an ethanol
464 buffer (50% ethanol, 50% of 1mM CoHex). Beads with DNA were poured into a bead retainer and
465 quickly spun to remove excess ethanol (or absorbed with a filter paper or tissue). DNA was quickly
466 eluted with 10mM Tris-HCl pH 8.0 in a 2 ml tube and incubated at 37 ^{o}C for 30 min with regular
467 wide-bore pipette mixing. Incubation was continued overnight at room temperature. Afterwards,
468 using a bead retainer, DNA was spun down at maximum speed for 1 min.

469 *Yeast extraction using CTAB and glass beads (Cahyani 2024)*

470 Yeast pellet (~50 µl volume or equivalent to 200 million cells) was washed twice in cold PBS and
471 resuspended in 480 µl SpheroBuffer (1M Sorbitol, 100mM Na_2HPO_4, 100mM EDTA). Lyticase
472 enzyme (5U/µl) was added (20 ul). Sample was incubated at 30°C for 30 minutes and spheroplast
473 formation was checked from time to time. Sample was lysed by adding 480 µl of 2× CTAB lysis
474 buffer (4% CTAB, 2M NaCl, 50 mM EDTA pH 8.0, and 20 mM Tris-HCl pH 9.0), mixed well with
475 wide-bore pipet tip as the buffer is viscous, added with 400 µg Proteinase K (20 mg/ml) and
476 incubated at 56°C for 15 minutes whilst shaking at 700 rpm. To remove RNA, 400 µg of RNase A
477 (100 mg/ml) was added followed by incubation at 56°C for 10 minutes without shaking. Sample
478 was cooled off at room temperature and added with 400 µl of 5M ammonium acetate and 100 µl of
479 5M NaCl [Thermofisher #AM9759] and mixed by a vertical rotator for 1 minute at 9 rpm. Sample
480 was then centrifuged for 7 minutes at 16,000×g and the lysate supernatant was removed to a new
481 2 ml tube. Three glass beads and 0.5 ml isopropanol were added, and the sample was mixed on a

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

482 vertical rotating mixer at 9 rpm for 5 minutes. Between 0.5-0.7 ml liquid was removed and replaced
483 with the same amount of isopropanol and was rotated for an additional 3 minutes. When DNA had
484 bound tightly around the beads, the liquid was discarded, and bound DNA was washed twice with
485 1 ml of 70% ethanol. Beads with DNA were poured into a bead retainer and quickly spun to
486 remove excess ethanol (or absorbed with a filter paper or tissue). DNA was quickly eluted with EB
487 (ONT) or 10mM Tris-HCl pH 9.0 in a 2 ml tube and incubated at 37 ^{o}C for 30 min with regular wide-
488 bore pipette mixing. Incubation was continued overnight at room temperature. Afterwards, using a
489 bead retainer, DNA was separated from the beads by centrifugation at maximum speed for 1
490 minute.
491

492 Quantification of UHMW DNA

493 Two nucleic acid quantification methods: fluorometric-based Qubit [Thermo Fisher Scientific, MA,
494 USA] and spectrophotometric-based NanoDrop [Thermo Fisher Scientific], were used in parallel to
495 assess both the quantity and purity of the extracted DNA. The quantification follows the published
496 protocol (Koetsier and Cantor 2021) with some modification as previously described (Cahyani et al.
497 2021). In brief, DNA is sampled from four different points in the tube: top, upper-middle, lower-
498 middle, and bottom part of the solution. Then, a glass bead was added, and sample was vortexed
499 at full speed for 1 min for concentration measurements. For Qubit measurement, Jurkat genomic
500 DNA (Thermofisher #SD1111) was used as a concentration standard instead of the lambda DNA
501 provided by the manufacturer (Koetsier and Cantor 2021). To measure DNA homogeneity in
502 solution, %CV of concentration was calculated by directly measuring concentrations of 3-4 different
503 points of the solution using NanoDrop.

504 Quality check of UHMW DNA using PFGE

505 To assess DNA quality and read-length distribution, 200-300 ng extracted DNA was run on 0.75%
506 gold agarose gel (SeaKem #50150). Pulsed-field gel electrophoresis was carried out using Pippin
507 Pulse device (Sage Science) at a pre-set mode of “5-430 kb” for 16 hours. Gel was stained with
508 RedSafe DNA stain (ChemBio #21141) and visualised with a Bio-Rad GelDoc imaging system.

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

509 ## Rapid kit-based UL library preparation

510 *Ultra-long kit-Nemo*

511 The ONT ultra-long kits, i.e. ULK001 and ULK114, are based on the cleaving and tagging of the
512 DNA by the transposome complexes to insert transposase adapters, which then are ligated to the
513 sequencing adapters (ONT ULK001). Our modifications of the ULK001 method were described in
514 previously published protocol (Cahyani et al. 2021). To summarise, DNA extracted from 1-6 million
515 cells was diluted to a final concentration of 20-40 ng/µl (in the final reaction volume). All mixing was
516 done with wide-bore pipette tips. Meanwhile on ice, the dilution buffer (FDB) was well mixed with
517 FRA (1 µl per 1 million human cells DNA), then combined to the DNA solution, and quickly
518 incubated at 23 ^{o}C for 5-10 min (depending on the reaction volume). Enzyme was inactivated at 70
519 ^{o}C for 5 min and cooled down to room temperature. Sequencing adapter (RAP-F) was added (0.83
520 µl per 1 µl FRA used) and the sample was incubated at room temperature for at least 30 min.
521 The ULK114 modifications are similar to ULK001 and are described in the newly published and
522 updated protocol also on protocols.io (Cahyani et al. 2024). In summary, library preparation to run
523 on a PromethION flow cell needs extracted DNA from 6 million cells that was diluted to a final
524 concentration of 20-40 ng/µl (in the final reaction volume). On ice, the dilution buffer (FDB) was
525 well mixed with FRA (6 µl per 6 million human cells DNA), then combined to the DNA solution, and
526 quickly incubated at 23 ^{o}C for 10 min with 9 rpm rotation. Enzyme was inactivated at 75 ^{o}C for 10
527 min and cooled down to room temperature. Sequencing adapter (RA) was added (5 µl per 6 µl
528 FRA) and the sample was rotated at 9 rpm for 30 minutes.

529 *RAD004-Nemo*

530 The RAD004-Nemo protocol followed the same route as the ULK001-Nemo protocol described
531 previously (Cahyani et al. 2021). However, the transposase and adapter used were the ones from
532 the RAD004 kit (*i.e.* FRA and RAP, respectively). The dilution buffer used to replace FDB was the
533 4× MuA buffer [100 mM Tris-HCl pH 8.0, 40 mM $MgCl_2$, 440 mM NaCl, 0.2% Triton X-100, 40%
534 Glycerol].

535 *Krazy StarFish (KSF)*

536 The KSF protocol is an alternative UL sequencing that uses a filter paper instead of glass beads. It
537 is a cheap, quick-and-dirty method to produce UL sequencing based on the RAD004 kit. This

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

538 method was described previously (Cahyani et al. 2021) and summarised as follows. An amount of
539 ~7.5 µg DNA in 75 µl volume was mixed well with 25 µl 4× MuA buffer. Meanwhile in another tube
540 on ice, 25 µl 4× MuA buffer was diluted with 74 µl water and added with 1 µl FRA (or up to 1.2 µl
541 FRA if DNA was quite viscous). These two solutions were mixed well on ice, aliquoted into two
542 PCR tubes, and treated at 30° C for 1 min, 80° C for 1 min then cooled to room temperature.
543 Reactions were pooled into a single tube, 12 µl 5M NaCl added, and mixed gently. A Krazy
544 StarFish filter (made from Whatman filter paper no. 3 with a star hole puncher) was submerged in
545 the solution, 142 µl isopropanol added, and mixed by inversion 20-30 times allowing UHMW DNA
546 to collect and condense onto the filter. Liquid was removed and filter paper was washed twice with
547 60-70% ethanol. Excess ethanol was removed by pipetting and paper was air-dried for 2 min. After
548 the paper was moved to a clean tube, the library was eluted by adding 120 µl of elution buffer
549 [10mM Tris-HCl pH 8.0]. Before sequencing, 37.5 µl library was added with 0.5 µl RAP and
550 incubated for at least 30 min at room temperature.

551 Nemo Clean Up: UHMW DNA Library Purification with Glass Beads and CoHex

552 The Nemo protocol is at the heart of this study and was previously described as an alcohol-free
553 library clean-up method (Cahyani et al. 2021, 2024). Typically, for DNA equivalent to 1 million cells
554 or more, 3 glass beads were used. In brief, glass beads were added to the library to be cleaned-up
555 in a 1:1 volume ratio of 10 mM CoHex (2×), i.e. final CoHex concentration is 5 mM. Sample was
556 rotated at 9 rpm for 3 min, or 20-25 times if using hand inversion. Liquid was removed and beads
557 were washed twice with PEGW buffer [10% PEG-8000, 0.5M NaCl] for 3 minutes each. Excess
558 buffer was removed with a fine pipette tip after a quick spin of the tube. Alternatively, the beads
559 with DNA could be poured into a bead retainer (home-made or using the NEB-T3004) and the
560 buffer was absorbed by tapping on a filter paper/tissue. Beads should not be let dry, were quickly
561 eluted with 10mM Tris-HCl pH 9.0 or EB (ONT) and incubated at 37 ^{o}C for 30 min with regular
562 wide-bore pipette mixing. Incubation was continued overnight (preferable) at room temperature.
563 Afterwards, using a bead retainer on a new tube, DNA was spun down at maximum speed for 1
564 min.

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

565 ## Flowcell Priming and Library Loading

566 All flow cells were primed according to the ULK001 and ULK114 protocols. Library was loaded
567 onto a primed FLO-MIN106 or FLO-PRO002 (R9.4.1), or FLO-MIN114 or FLO-PRO114 (R10.4.1)
568 flow cells for sequencing on a GridION or PromethION, respectively, using the accompanying
569 MinKNOW version software. Loaded UL library was let to tether on the flow cell for at least 30 min
570 before run was started using the ONT ultra-long script with mux scan of every 6 hours for R9.4.1
571 and the default 1.5 hours for R10.4.1.

572 ## FindingNemo in one day

573 To run UL sequencing within the same day as the DNA extraction, FindingNemo *in one day*
574 protocol was developed as previously described (Cahyani 2021). In brief, cells were extracted
575 using the Monarch kit and lysed at 700-800 rpm. Eluted DNA was incubated at 37 ^{o}C and regularly
576 mixed for at least two hours. After quantification showed that DNA was sufficiently homogeneous
577 (%CV < 100%), the library was prepared following the ULK001-Nemo protocol. Eluted library was
578 further incubated at 37 ^{o}C for at least 30 min with regular pipette mixing, quantified and loaded on
579 a primed flow cell.

580 ## NanoPlot Data Analysis and Plotting

581 Violin plots of sequencing speed and distribution graphs comparing read-lengths and read quality
582 were created using the summary file data processed by the NanoPlot tool (De Coster et al. 2018).

583 ## R Data Analyses and Plotting

584 Read-length histograms, occupancy, yield and other sequencing metrics were analysed from the
585 summary, mux, and duty time (or in newer filename version, pore activity) output files using in-
586 house R (R Core Team 2024) scripts available on GitHub
587 (https://github.com/cinswasti/FindingNemo) or in the Supplemental Codes.

588 ## Sequence sources and alignment

589 For human work, the reference genome used is the hg38 (NCBI RefSeq assembly
590 GCF_000001405.40). Sequence data were aligned to this reference genome using minimap2 (Li
591 2021). Nanopore data of GM24385 and GM24631 are publicly available on AWS registry and can

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

592 be accessed using the command line `aws s3 ls --no-sign-request s3://ont-open-
593 data`.
594

595 ## Data Access

596 The FASTQ data generated in this study had been submitted to the European Nucleotide Archive
597 (ENA; https://www.ebi.ac.uk/ena/browser/home) under accession number PRJEB76809. The
598 FindingNemo protocol can be found in Supplemental_Protocol_FindingNemo or on the protocols.io
599 website (https://dx.doi.org/10.17504/protocols.io.5jyl8p38rg2w/v1).
600

601 ## Competing Interest Statement

602 M.L., J.T, J.Q. and N.L were members of the MinION early access program and have received free
603 flow cells and sequencing reagents in the past. All have received reimbursement for travel and
604 accommodation to speak at events organised by ONT.
605

606 ## Acknowledgements

607 We would like to thank Giron Koetsier (NEB), Simon Mayes (ONT), and Kelvin Liu (Circulomics) for
608 lending their expertise and/or advanced product trials. We thank Darren Crowley and Luke
609 Simpson for supplying the GM12878 cells, Stephen Gray for the yeast cells, Danny Miller and
610 Miranda Galey (University of Washington) for the Human Genome (HG) cells, Lesley Bell-Sakyi
611 and Alistair Darby (University of Liverpool) for the tick *A. variegatum* cells, Giron Koetsier (NEB) for
612 the HEK293 cells, and John Mulley (Bangor University) for the use of the gerbil's sequencing
613 metrics. John Tyson is grateful to Terrance Snutch. This work was supported by the Wellcome
614 Trust (grant number 204843/Z/16/Z, I.C., N.H., N.L and M.L.).
615

616 ## Author contributions

617 N.L, M.L, J.T and J.Q conceived the project. I.C. led the work with contributions from J.Q, J.T, N.H
618 and C.M. I.C drafted the manuscript with input from all authors. All authors reviewed, revised and
619 approved the final version of the manuscript.
620

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

621 ## References

622 Allers T, Lichten M. 2000. A method for preparing genomic DNA that restrains branch
623 migration of Holliday junctions. *Nucleic Acids Res* **28**: 6.
624 Arscott PG, Ma C, Wenner JR, Bloomfield VA. 1995. DNA condensation by cobalt
625 hexaammine(III) in alcohol–water mixtures: Dielectric constant and other solvent effects.
626 *Biopolymers* **36**: 345–364. https://onlinelibrary.wiley.com/doi/10.1002/bip.360360309
627 (Accessed February 25, 2022).
628 Arseneau J-R, Steeves R, Laflamme M. 2017. Modified low-salt CTAB extraction of high-
629 quality DNA from contaminant-rich tissues. *Mol Ecol Resour* **17**: 686–693.
630 https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.12616.
631 Cahyani I. 2021. FindingNemo in OneDay: Ultra-Long ONT Library Preparation from Cell to
632 Flowcell in One Day. https://www.protocols.io/view/findingnemo-in-oneday-ultra-long-ont-
633 library-prepa-14egnzqzyg5d/v1 (Accessed November 4, 2024).
634 Cahyani I, Tyson J, Holmes N, Quick J, Loman N, Loose M. 2021. FindingNemo: A Toolkit of
635 CoHex- and Glass Bead-based Protocols for Ultra-Long Sequencing on ONT Platforms.
636 *protocols.io*. https://www.protocols.io/view/findingnemo-a-toolkit-of-cohex-and-glass-
637 bead-base-dm6gpwkn1lzp/v1 (Accessed March 11, 2024).
638 Cahyani I, Tyson J, Holmes N, Quick J, Loman N, Loose M. 2024. FindingNemo (v.kit14): A
639 Toolkit for DNA Extraction, Library Preparation and Purification for Ultra Long Nanopore
640 Sequencing. *protocols.io*. https://www.protocols.io/view/findingnemo-v-kit14-a-toolkit-for-
641 dna-extraction-l-5jyl8p38rg2w/v1/metadata (Accessed April 17, 2024).
642 De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack:
643 visualizing and processing long-read sequencing data. Bioinformatics 34: 2666–2669.
644 https://dx.doi.org/10.1093/bioinformatics/bty149 (Accessed November 4, 2024).
645 Deng H, Bloomfield VA. 1999. Structural Effects of Cobalt-Amine Compounds on DNA
646 Condensation. *Biophys J* **77**: 1556–1561.
647 https://linkinghub.elsevier.com/retrieve/pii/S0006349599770037 (Accessed March 11,
648 2024).

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

649 Giordano F, Aigrain L, Quail MA, Coupland P, Bonfield JK, Davies RM, Tischler G, Jackson
650 DK, Keane TM, Li J, et al. 2017. De novo yeast genome assemblies from MinION,
651 PacBio and MiSeq platforms. *Sci Rep* **7**: 1–10.
652 Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, Tyson JR, Beggs AD, Dilthey AT,
653 Fiddes IT, et al. 2018a. Nanopore sequencing and assembly of a human genome with
654 ultra-long reads. *Nat Biotechnol* **36**: 338–345. http://www.nature.com/articles/nbt.4060.
655 Jain M, Olsen HE, Turner DJ, Stoddart D, Bulazel K V, Paten B, Haussler D, Willard HF,
656 Akeson M, Miga KH. 2018b. Linear assembly of a human centromere on the Y
657 chromosome. *Nat Biotechnol* **36**: 321–323. https://www.nature.com/articles/nbt.4109
658 (Accessed March 11, 2024).
659 Kankia BI, Buckin V, Bloomfield VA. 2001. Hexamminecobalt(III)-induced condensation of
660 calf thymus DNA: circular dichroism and hydration measurements. *Nucleic Acids Res* **29**:
661 2795. /pmc/articles/PMC55774/ (Accessed July 19, 2022).
662 Koetsier G, Cantor E. 2019. A Practical Guide to Analyzing Nucleic Acid Concentration and
663 Purity with Microvolume Spectrophotometers. *New England Biolabs* **1**: 1–8.
664 https://www.neb.com/-/media/catalog/application-
665 notes/mvs_analysis_of_na_concentration_and_purity.pdf?rev=be7c8e19f4d34e5585274
666 96ea51623dc.
667 Koetsier PA ‘Giron’, Cantor EJ. 2021. A simple approach for effective shearing and reliable
668 concentration measurement of ultra-high-molecular-weight DNA. *Biotechniques* **71**: 439–
669 444. https://www.future-science.com/doi/10.2144/btn-2021-0051 (Accessed August 4,
670 2021).
671 Logsdon G. 2020. HMW gDNA purification and ONT ultra-long-read data generation v1.
672 https://www.protocols.io/view/hmw-gdna-purification-and-ont-ultra-long-read-data-
673 bchhit36 (Accessed December 13, 2024).
674 Logsdon GA, Vollger MR, Hsieh PH, Mao Y, Liskovykh MA, Koren S, Nurk S, Mercuri L,
675 Dishuck PC, Rhie A, et al. 2021. The structure, function and evolution of a complete
676 human chromosome 8. *Nature*. http://dx.doi.org/10.1038/s41586-021-03420-7.

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

677 Oda Y, Sadakane K, Yoshikawa Y, Imanaka T, Takiguchi K, Hayashi M, Kenmotsu T,
678 Yoshikawa K. 2016. Highly Concentrated Ethanol Solutions: Good Solvents for DNA as
679 Revealed by Single-Molecule Observation. *Chemphyschem* **17**: 471.
680 /pmc/articles/PMC4770436/ (Accessed February 25, 2022).
681 Ouameur AA, Tajmir-Riahi H-A. 2004. Structural Analysis of DNA Interactions with Biogenic
682 Polyamines and Cobalt(III)hexamine Studied by Fourier Transform Infrared and Capillary
683 Electrophoresis. *Journal of Biological Chemistry* **279**: 42041–42054.
684 https://linkinghub.elsevier.com/retrieve/pii/S002192582072675X (Accessed February 25,
685 2022).
686 Pelta J, Livolant F, Sikorav J-L. 1996. DNA Aggregation Induced by Polyamines and
687 Cobalthexamine. *Journal of Biological Chemistry* **271**: 5656–5662.
688 https://linkinghub.elsevier.com/retrieve/pii/S0021925818979068 (Accessed March 11,
689 2024).
690 Quick J. 2018. Ultra-long read sequencing protocol for RAD004. *Protocols Io* 1–16.
691 https://www.protocols.io/view/ultra-long-read-sequencing-protocol-for-rad004-mrvc566
692 (Accessed August 4, 2021).
693 R Core Team (2024). R: A language and environment for statistical computing. R Foundation
694 for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.
695 Robertson RM, Smith DE. 2007. Strong effects of molecular topology on diffusion of
696 entangled DNA molecules. *Proc Natl Acad Sci U S A* **104**: 4824–4827.
697 www.pnas.orgcgidoi10.1073pnas.0700137104 (Accessed March 12, 2024).
698 Rodrigue S, Materna AC, Timberlake SC, Blackburn MC, Malmstrom RR, Alm EJ, Chisholm
699 SW. 2010. Unlocking Short Read Sequencing for Metagenomics. *PLoS One* **5**: e11840.
700 https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0011840 (Accessed
701 February 7, 2022).
702 Salazar AN, Gorter de Vries AR, van den Broek M, Wijsman M, de la Torre Cortés P,
703 Brickwedde A, Brouwers N, Daran J-MG, Abeel T. 2017. Nanopore sequencing enables
704 near-complete de novo assembly of Saccharomyces cerevisiae reference strain

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

705 CEN.PK113-7D. *FEMS Yeast Res* **17**.
706 https://academic.oup.com/femsyr/article/17/7/fox074/4157789 (Accessed July 22, 2021).
707 Simbolo M, Gottardi M, Corbo V, Fassan M, Mafficini A. 2013. DNA Qualification Workflow for
708 Next Generation Sequencing of Histopathological Samples. *PLoS One* **8**: 62692.
709 www.plosone.org (Accessed March 12, 2024).
710 Stortchevoi A, Kamelamela N, Levine SS. 2020. SPRI beads-based size selection in the
711 range of 2-10kb. *Journal of Biomolecular Techniques* **31**: 7–10.
712 /pmc/articles/PMC6944320/ (Accessed February 7, 2022).
713 Synthego HEK293. HEK293 Cells: Background, Applications, Protocols, and More.
714 *synthego.com*. https://www.synthego.com/hek293 (Accessed March 12, 2024).
715 Tyson J. 2019. Rocky Mountain adventures in Genomic DNA sample preparation, ligation
716 protocol optimisation / simplification and Ultra long read generation v1.
717 https://www.protocols.io/view/rocky-mountain-adventures-in-genomic-dna-sample-pr-
718 7euhjew (Accessed November 4, 2024).
719 Xia Y, Chen F, Du Y, Liu C, Bu G, Xin Y, Liu B. 2019. A modified SDS-based DNA extraction
720 method from raw soybean. *Biosci Rep* **39**: 1–10.
721 https://portlandpress.com/bioscirep/article/39/2/BSR20182271/110910/A-modified-SDS-
722 based-DNA-extraction-method-from.
723
724

725 ## Figure legends

726 **Figure 1**. Summary of the FindingNemo toolkit: (A) Similarities and differences between the
727 standard and ultra-long (UL) protocol that centre on the way UHMW DNA is handled. Created with
728 BioRender (https://biorender.com). (B) The route of FindingNemo toolkit consists of five main
729 steps, with options at the extraction, library preparation and clean-up steps. Options with asterisks
730 (*) are our current lab workflow for the ultra-long protocol.
731
732 **Figure 2**. Validation of the FindingNemo protocol - sequencing outputs between rapid-based
733 library preparation protocols (ULK001-Nemo and RAD004-Nemo protocol) showing comparable

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

734 performance when compared to the control ULK001 protocol. (A) Workflow of the library
735 preparation protocols, (B) Total yields sub categorised by read lengths, (C) Time lapse of yields,
736 (D) Read length distributions; dashed black vertical lines denote N50s, (E) Time lapse of relative
737 pore occupancies. Each library was loaded three times (i.e. total DNA equivalent to 3 million
738 GM12878 cells) on a MinION R.9.4.1 flow cell following a nuclease flush protocol. All data are after
739 65 hours of sequencing on the GridION platform.

740

741 **Figure 3**. Choosing the best precipitating agent that ensures fast homogenisation: (A) Read length
742 distribution of libraries precipitated with different buffers; dashed black vertical lines denote N50s,
743 (B) Sequencing metrics after 6 hours of run (excluding the first 10 minutes). All libraries were
744 extracted from GM12878 cells using Monarch direct lysis protocol with an input DNA concentration
745 of ~40 ng/µl split into three clean-up preparations, loaded on MinION R.9.4.1 flow cells and
746 sequenced on GridION platform.

747

748 **Figure 4**. Combinatorial effects of extraction and library clean-up methods, either with or without
749 the use of kits, on the UL sequencing outputs: (A-I) Read length distributions of the libraries: graph
750 titles denote the used extraction kit/method and the clean-up protocol, dashed black vertical lines
751 denote N50s. (J) Sequencing metrics after 6 hours of run (excluding the first 10 minutes). DNA was
752 extracted from GM12878 cells. Each library was loaded on a MinION R9.4.1 flow cell and run on
753 the GridION platform.

754

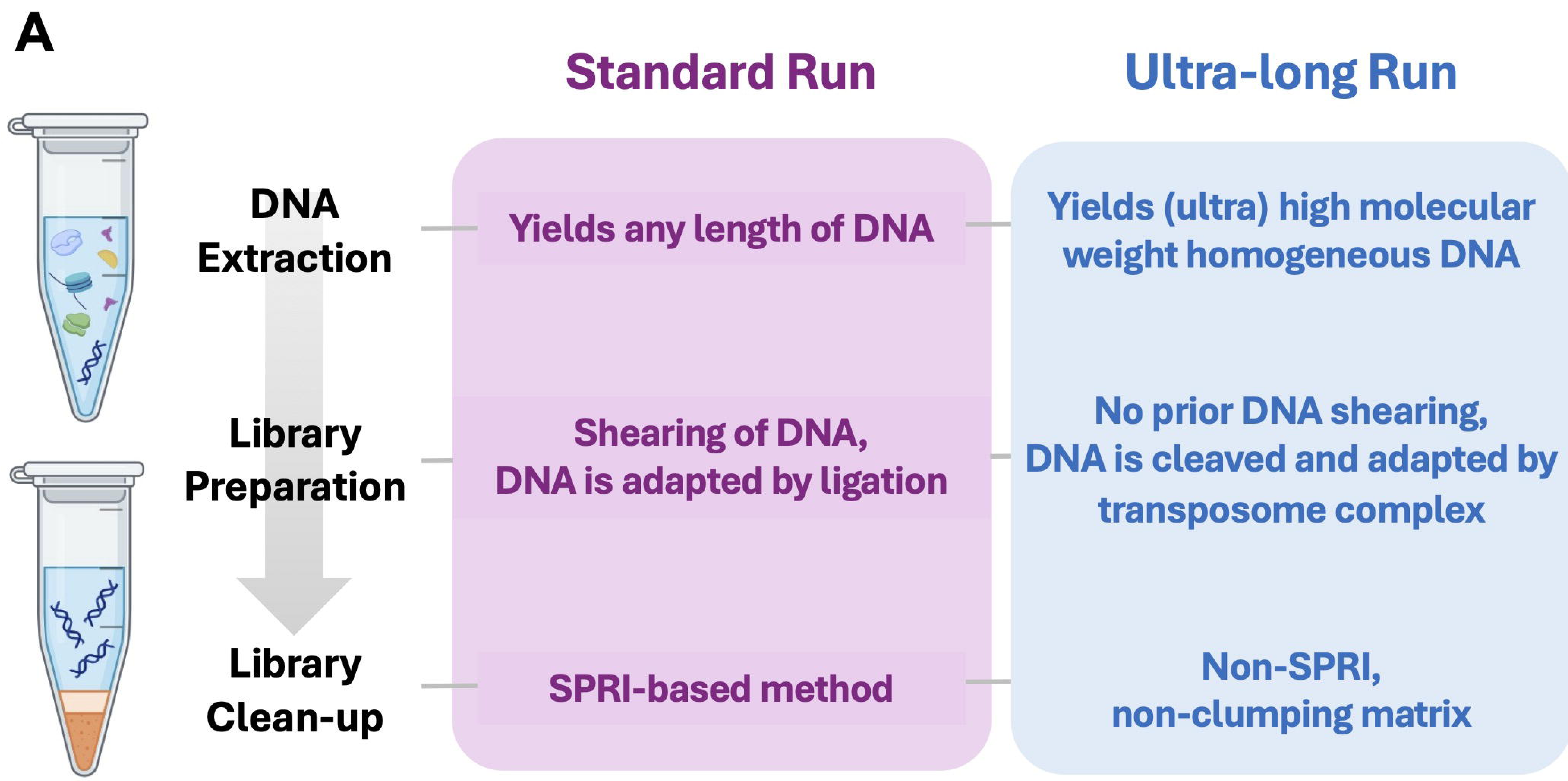
755 **Figure 5.** Ultra-long N50s were obtained using the new ULK114 kit from three cell lines: (A) Read
756 length distribution of GM12878 library run with the earlier ULK114 script version (left panel), while
757 the other two libraries (middle and right panel) were run with the most current 5 kHz version of the
758 script. Vertical dashed black lines indicate N50s. (B) Run metrics of the three libraries showing the
759 Nemo protocol performed better than ONT's.

760

761 **Figure 6.** Ultra-long sequencing output and metrics of non-GM12878 cells and different
762 sequencing platforms: gerbil blood, tick *Amblyomma variegatum*, yeast *Saccharomyces cerevisiae*

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

763 and human HEK293 cells using either the ULK001 or the ULK001-Nemo protocol (as labelled). (A)
764 Read length distributions of libraries run on PromethION R9.4.1 flow cells; dashed black vertical
765 lines denote N50s, (B) Read length distribution of the HEK293 and yeast *S. cerevisiae* libraries run
766 on MinION R9.4.1 flow cells, (C) Sequencing metrics of the libraries. Each metric was shown after
767 6 hours of run (excluding the first 10 minutes).
768
769 **Figure 7**. Optimum conditions for ultra-long sequencing - to obtain an optimum UL sequencing
770 output, there needs to be a 1:1 ratio of FRA to DNA molecules in a homogeneous solution
771 ascertained by proper cell counting and/or DNA mass measurement. Image was created with
772 BioRender (https://biorender.com).
773



B

I. UHMW DNA Extractions

*Choose **one** protocol*:
- Kit NEB Monarch*
- Non-kit Phenol-based
- Non-kit Phenol-free
- Kit Circulomics

II. Quantification & QC

III. Library Prep

*Choose **one** protocol*:
- Rapid kit ULK114*
- Rapid kit ULK001
- Rapid kit RAD004
- KrazyStarFish (KSF)

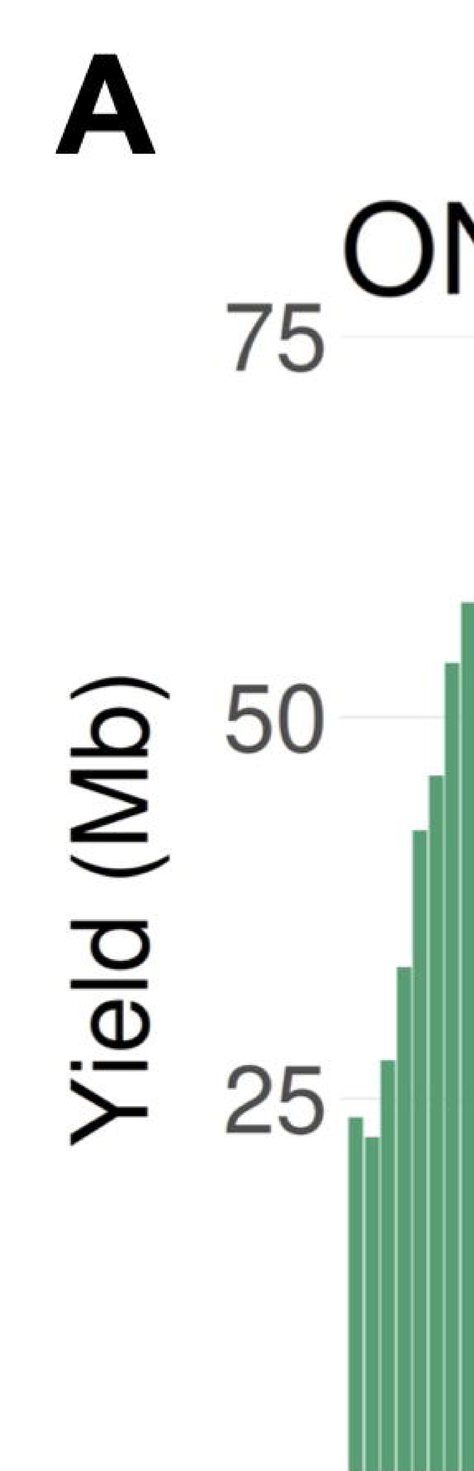
IV. Library Clean-Up

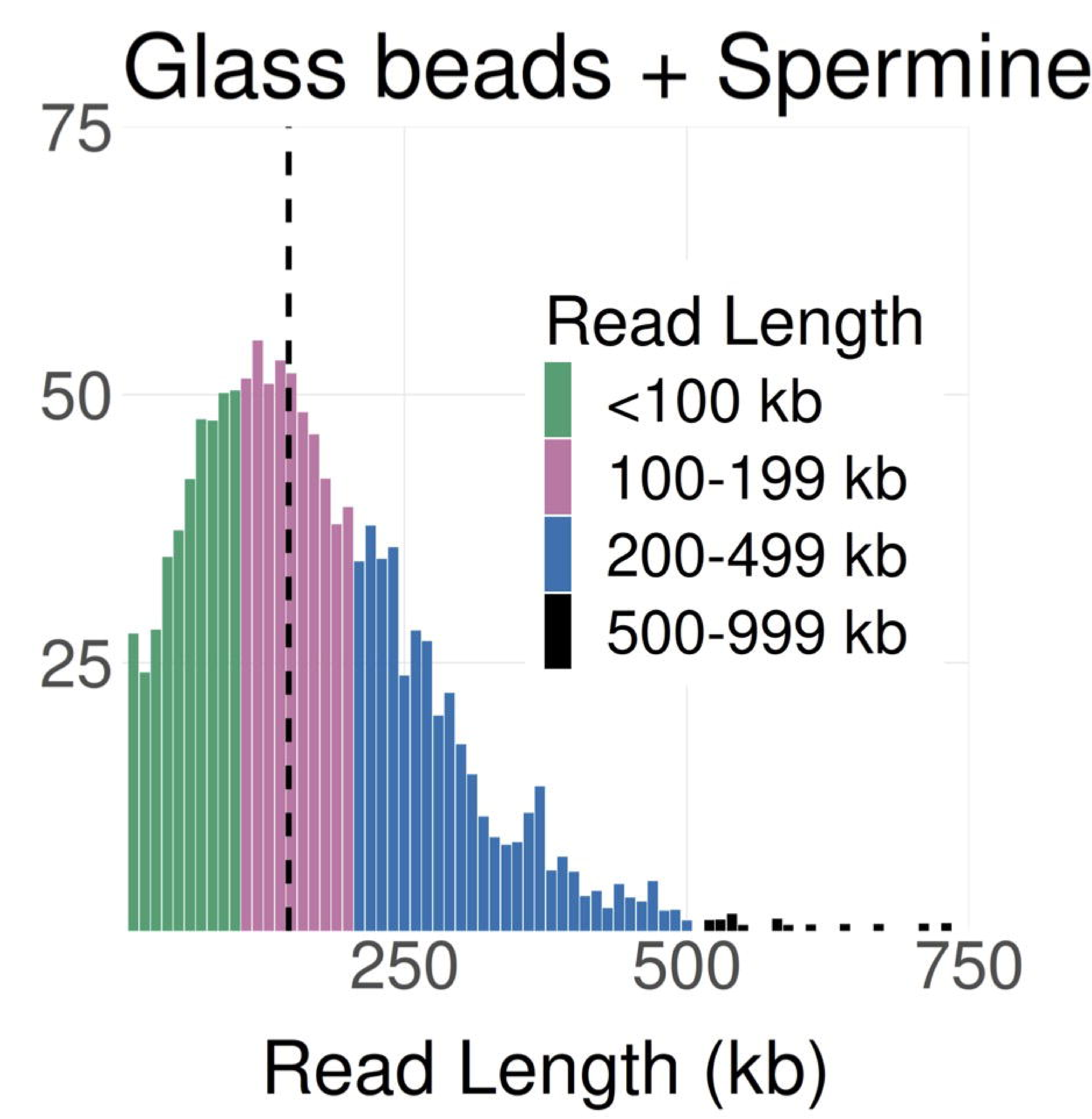
*Choose **one** protocol*:
- FindingNemo*
- ONT Ninja-star
- Circulomics Nanobind

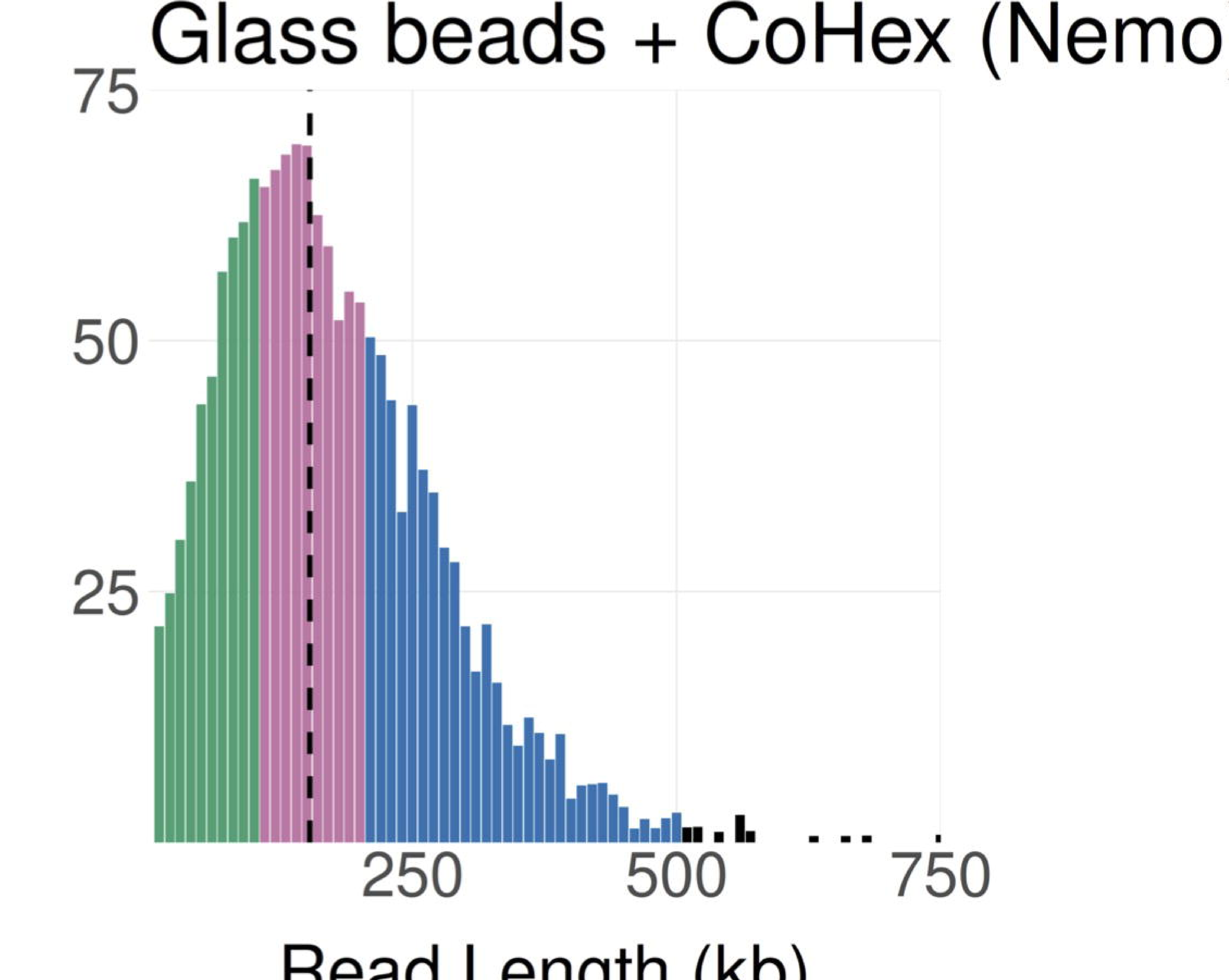
V. Priming & Loading

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press







B

Sequencing Metrics at 6 Hours of Run

Sample	N50 (bp)	Pores Used	Total Yield (% Bases ≥100 kb)	Yield per Pore (Mb)	Average Occupancy (%)	% CV of Library Concentration (Nanodrop)
ONT PPT buffer	153,295	1010	1.59 Gb (72.55%)	1.57	85.65	14.8
Glass Beads + Spermine	147,386	635	1.28 Gb (69.51%)	2.01	85.32	11.1
Glass Beads + CoHex (Nemo)	152,653	794	1.61 Gb (72.20%)	2.03	89.29	1.3

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

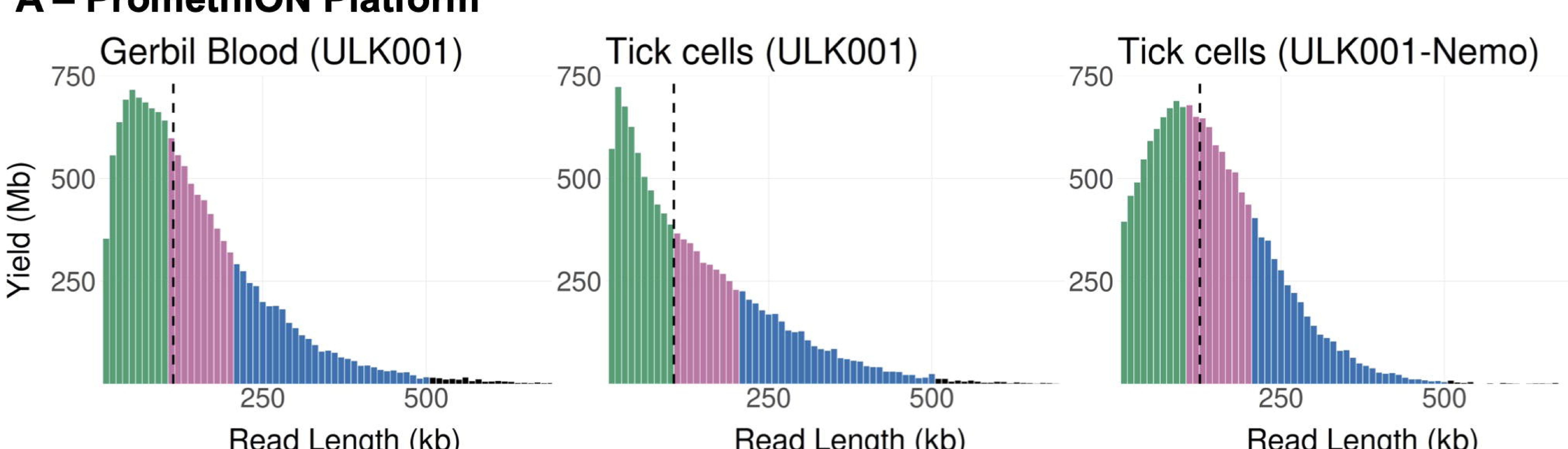
	Extraction-Clean up	N50 (bp)	Pores Used	Yield (% Bases ≥ 100 kb)	Average Occupancy (%)	Yield per Pore (Mb)
A	Phenol-Nemo	131,338	643	2.35 Gb (62.15%)	95.65	3.66
B	Phenol-Nemo 2	129,521	437	0.67 Gb (62.00%)	76.69	1.53
C	CTAB Lysis-Nemo	74,909	581	2.13 Gb (34.90%)	93.60	3.67
D	CTAB Lysis-Nemo 2	122,223	537	1.01 Gb (58.75%)	78.09	1.87
E	SDS Lysis-Nemo	106,027	461	0.92 Gb (52.55%)	76.48	1.99
F	CTAB and CoHex Lysis-Nemo	83,819	834	2.41 Gb (39.62%)	77.30	2.89
G	Monarch 300 rpm-Nemo	124,929	775	2.10 Gb (60.18%)	94.43	2.71
H	Monarch 900 rpm-Nanobind	106,493	862	3.35 Gb (52.70%)	97.36	3.89
I	Circulomics-Nanobind	147,123	1025	3.27 Gb (64.85%)	95.01	3.19

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

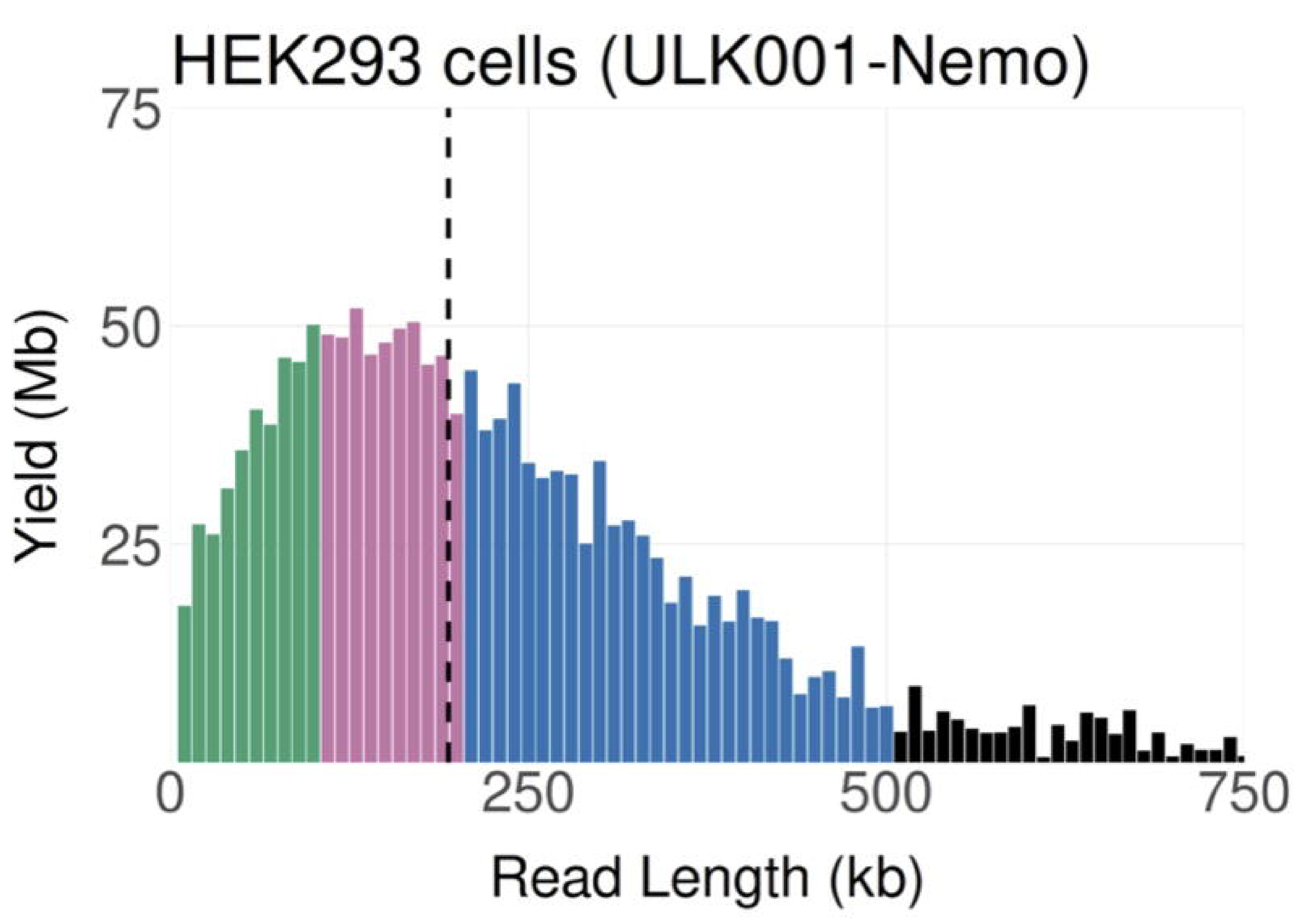
B

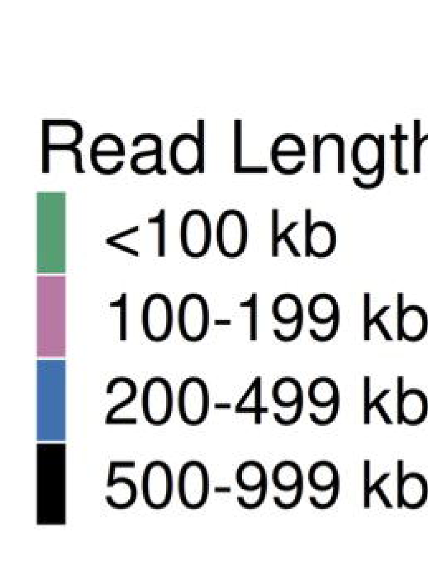
Sequencing metrics at 24 hours of run on PromethION platform

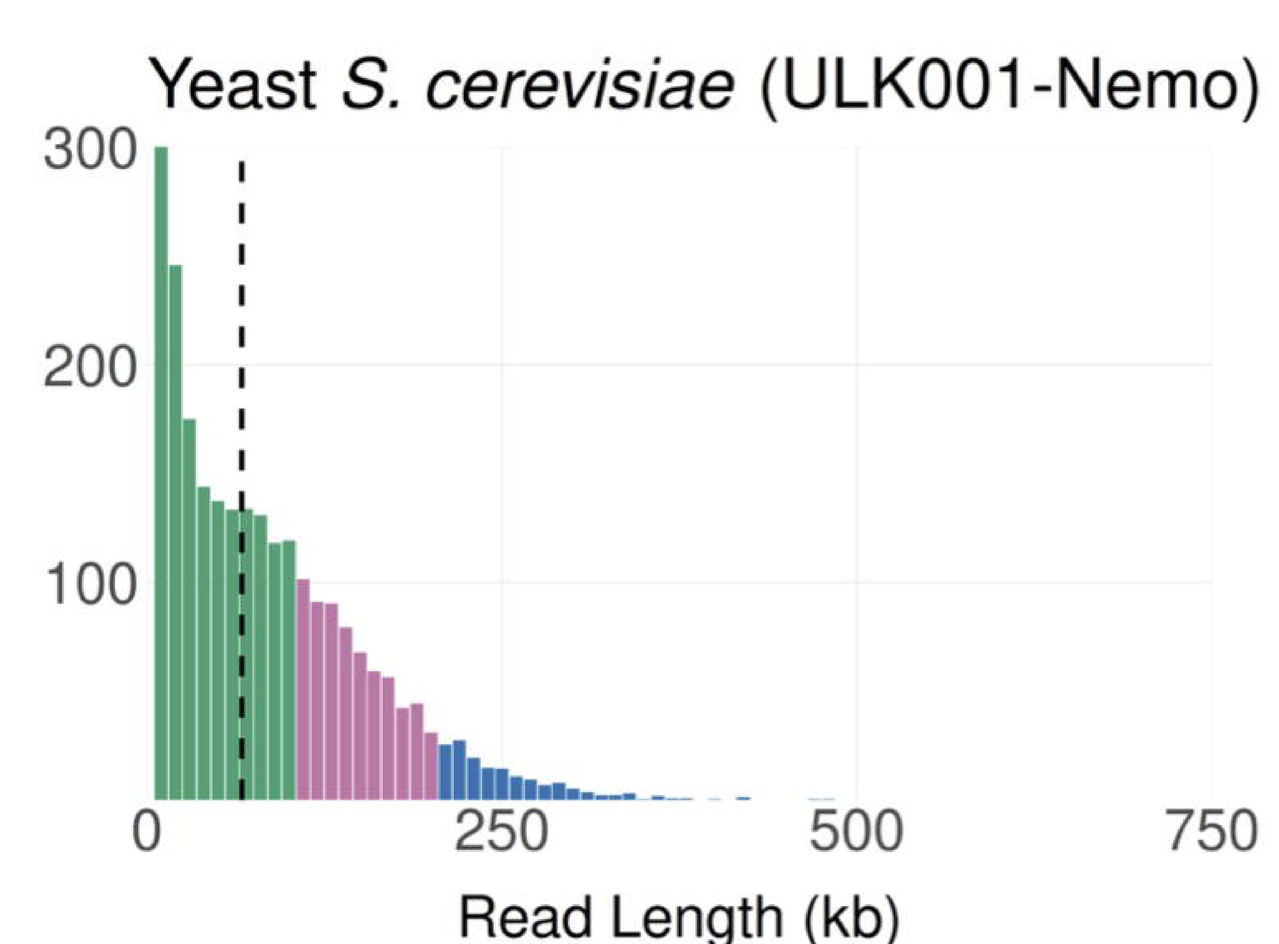
Sample	N50 (bp)	Pores Used	Yield (Reads >100 kb)	Yield per Pore (Mb)	Average Occupancy (%)	Mapping Frequency (Primary / Total Mapped)
ULK114-Nemo-4 kHz	102,692	7817	38.04 Gb (51.26%)	4.87	98.70	92.1% / 94.8%
ULK114-Nemo-5 kHz	133,881	7867	37.43 Gb (63.36%)	4.76	98.65	96.9% / 98.3%
ULK114-ONT-5 kHz	93,538	7879	32.30 Gb (47.08%)	4.10	98.74	96.9% / 98.0 %



B – MinION Platform





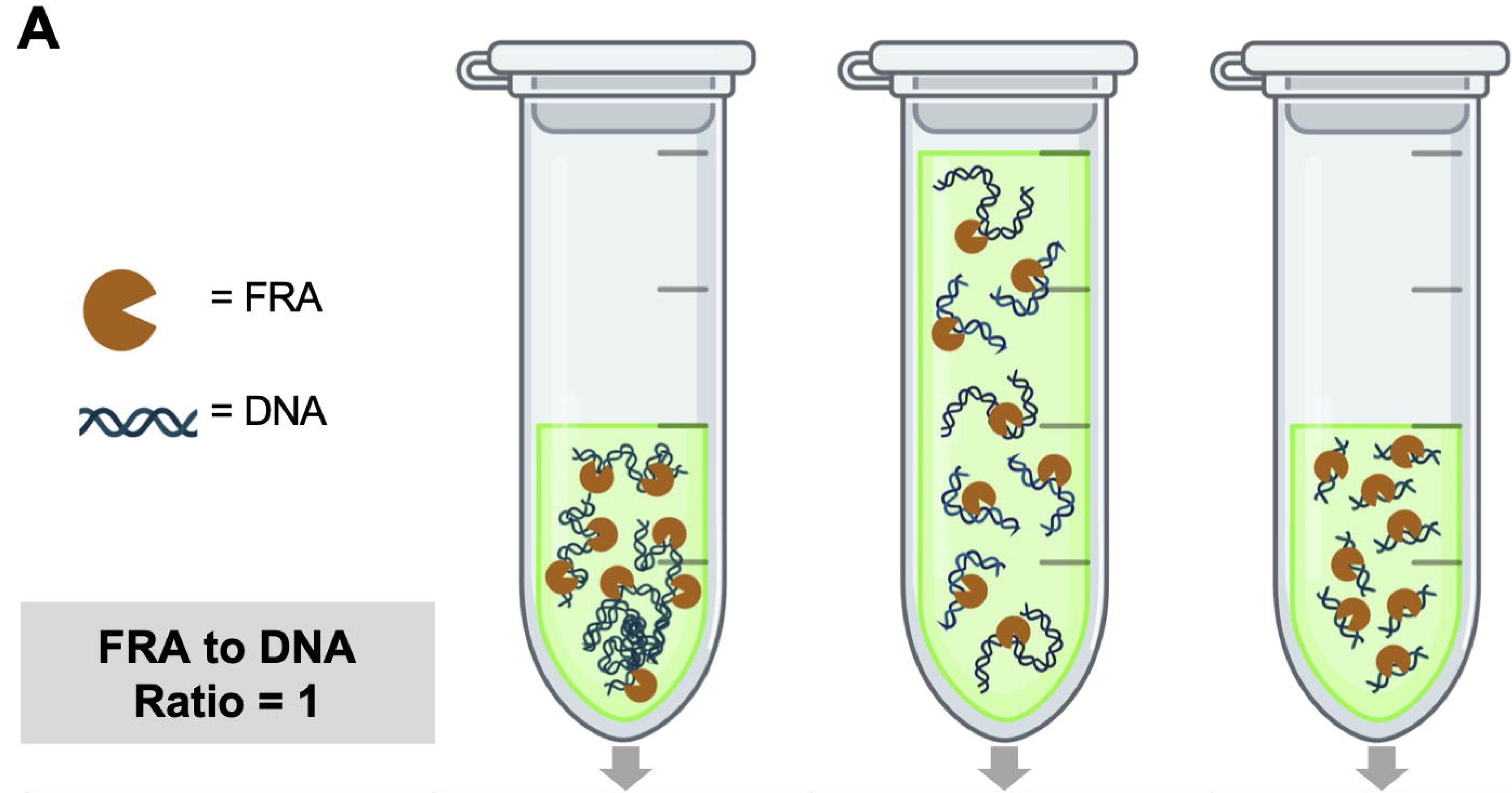


C – Sequencing Metrics at 6 Hours of Run

Sample (Library Protocol)	Platform	N50 (bp)	Pores Used	Yield (% Bases ≥ 100 kb)	Yield per Pore (Mb)	Average Occupancy (%)
Gerbil blood cells (ULK001)	PromethION	112,940	6220	14.16 Gb (55.43%)	2.28	79.89
Tick cells (ULK001)	PromethION	104,993	3492	11.11 Gb (51.64%)	3.18	87.50
Tick cells (ULK001-Nemo)	PromethION	126,008	3902	15.01 Gb (61.46%)	3.85	88.20
HEK293 cells (ULK001-Nemo)	MinION	194,200	728	1.63 Gb (77.94%)	2.24	91.53
Yeast *S. cerevisiae* (ULK001-Nemo)	MinION	66,694	762	2.47 Gb (33.36%)	3.24	92.87

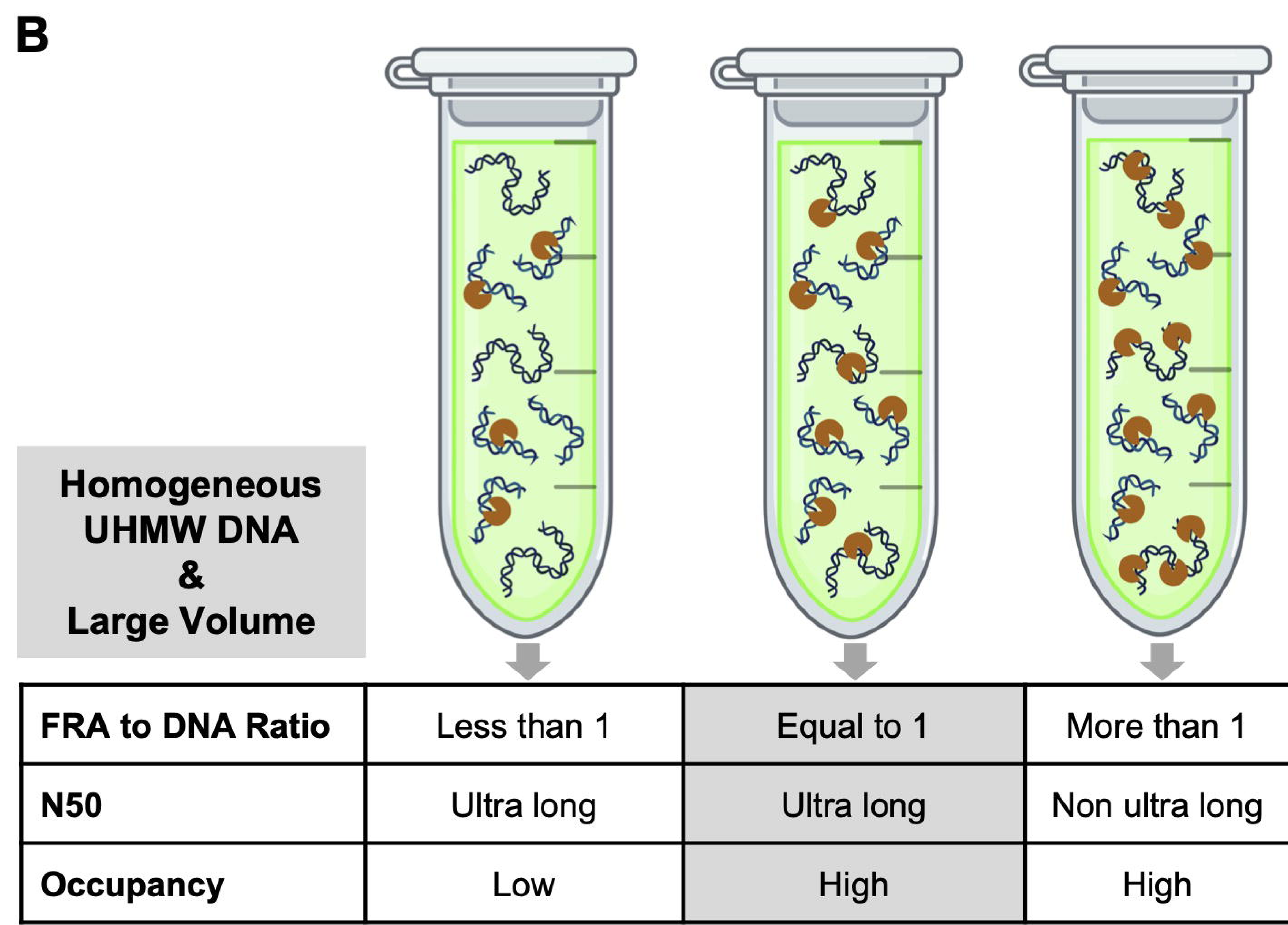
Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

A



DNA State	UHMW, non-homogeneous	UHMW, homogeneous	HMW, Homogeneous
Volume	Small	Large	Small
N50	Non ultra long	Ultra long	Non ultra long
Occupancy	Low to medium	High	High

B



FRA to DNA Ratio	Less than 1	Equal to 1	More than 1
N50	Ultra long	Ultra long	Non ultra long
Occupancy	Low	High	High

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

Table 1 Compilation of ultra-long (UL) library preparation protocols, optimized in human GM12878 cells

	Previous Protocols				New Protocols				
	Quick's UL (Quick, 2018)	Logsdon's UL (Logsdon, 2020)	Rocky Mountain (Tyson, 2019)	Krazy StarFish (KSF)	ULK001	RAD004-Nemo	ULK001-Nemo	ULK114	ULK114-Nemo
	A	B	C	D	E	F	G	H	I
Input cell number*	50 million	20-70 million	>2 million	>2 million	6 million	1-6 million	1-6 million	6 million	6 million
Input DNA amount	15 µg	2-3 µg	≥7.5 µg	≥7.5 µg	40 µg	5-40 µg	5-40 µg	40 µg	40 µg
ONT Library preparation kit	RAD004	RAD004	LSK109	RAD004	ULK001	RAD004	ULK001	ULK114	ULK114
Fragmentation Volume	Small (15 µl)	Small (18 µl)	Medium (60-100 µl)	Medium (2 × 100 µl)	Large (1 ml)	Large (0.5-1 ml)	Large (0.5-1 ml)	Large (1 ml)	Large (1-1.2 ml)
Library clean-up protocol	No	No	No	Partial using filter paper	Nanobind (Circulomics)	Nemo	Nemo	No	Nemo
Processing time (excluding incubations)	15 min	2.5 hours	18-24 hours	30 min	~2.5 hours	~2.5 hours	~2.5 hours	~2.5 hours	~2.5 hours
Input library per MinION flow cell	10-15 µg	2-3 µg	1 to 3 µg	2 to 4 µg	5 to 6 µg	1 to 5 µg	1 to 5 µg	NA	NA
No. of loads per library (MinION)	1	1	1 to 3	3	6	1 to 6	1 to 6	NA	NA
Expected yield per MinION flow cell	~3 Gb	1-2 Gb	10 to >20 Gb (depends on N50)	6-8 Gb (3 loads)	>20 Gb (3 loads)	>15 Gb (3 loads)	~20 Gb (3 loads)	NA	NA
Input library per PromethION flow cell	NA	NA	NA	NA	10-13 µg	10-13 µg	10-13 µg	10-13 µg	10-13 µg
No. of loads per library (PromethION)	NA	NA	NA	NA	3	1 to 3	1 to 3	3	3 to 4
Expected yield per PromethION flow cell (Gb)	NA	NA	NA	NA	NA	NA	24 Gb (1 load)	~85 Gb	up to >100 Gb
Ultra-long read length (≥100 kb)	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes

(*) cell number was dependent on the extraction protocol used; here it is listed as the number of cells required by the original protocol; if cell number is unknown, use the DNA amount; (NA) not available

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

Table 2 Compilation of UHMW DNA extraction methods for UL sequencing, optimised in human GM12878 cells

	Without Kits (Phenol-based)			**Without Kits (Phenol-free)**		**Kit-based**	
	Quick's Phenol Protocol (Quick, 2018)	**Logsdon's Phenol Protocol (Logsdon, 2021)**	**Phenol + Glass beads**	**CTAB or SDS Lysis + glass beads**	**CTAB-CoHex Lysis + glass beads**	**NEB Monarch Cell & Blood Kit**	**Circulomics Nanobind CBB Big DNA Kit**
	A	**B**	**C**	**D**	**E**	**F**	**G**
Standard cell input	50 million	20-70 million	1-5 million	1-5 million	1-5 million	0.5-5 million	6 million
Phenol-chloroform	Yes	Yes	Yes	None	None	None	None
Lysis buffer	SDS	SDS	SDS	CTAB or SDS	CTAB and CoHex	Proprietary	Proprietary
SPRI Substrate	None	None	3-mm glass beads	3-mm glass beads	3-mm glass beads	4-mm glass beads	Nanobind disc
Precipitation reagents	Ethanol, Ammonium Acetate	Ethanol, Ammonium Acetate	Ethanol, Ammonium Acetate	Isopropanol, Ammonium Acetate	CoHex	Isopropanol, Proprietary binding buffer	Isopropanol, Proprietary binding buffer
DNA extraction time	4 hours	4 hours + overnight	2 hours	1 hour	1 hour	~30 min	2 hours
DNA resuspension time	2 days	2 days	min. overnight	min. 2 hours	min. overnight	min. 1 hour (standard overnight)	overnight

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

Downloaded from genome.cshlp.org on June 23, 2026 . Published by Cold Spring Harbor Laboratory Press

Table 3. Ratio of FRA to DNA in different genomes

Species	**Genome Size**	**Cell Input Number**	**DNA Equivalent (µg)**	**FRA Volume (µl)**
Human GM12878	3.2 Gb	1×10^6	5-6	1
Human HEK293	3.2 Gb	1×10^6	6-8*	1
Mongolian Gerbil	~2.7 Gb	Unknown	10	1.5
Tick *A. variegatum*	~6.0 Gb	1×10^6	11-13	2.5
Yeast *S. cerevisiae* (n)	12 Mb	2×10^8	4-5	6

(*) extracted DNA yield is likely to be more than normal diploid human cell line because of its hypotriploid nature (Lin 2014)