**Optional fragmentation of gDNA**

Flow cell output is governed by various factors, including DNA/RNA library input, loading amounts and pore blocking. Although DNA fragmentation is not a requirement for nanopore library preparation, it can be useful when starting with low input amounts of DNA, or for handling viscous samples where the DNA is very high molecular weight, or to make samples more uniform in fragment size, e.g. for barcoding. This document is a review of how controlling the input material size distribution by fragmentation impacts flow cell output, and what Oxford Nanopore Technologies’ recommendations are to generate the best data, based on experimental aims. Below, we have included protocols we have developed using different fragmentation methods depending on the needs of the user.

**Increasing read N50**

It has been observed that some shearing of gDNA samples can lead to an increase in observed read length: this seems counterintuitive – how can breaking up the DNA fragments give longer reads? It has been suggested that certain fragments may be so long that they become “lost” during the library preparation and therefore are not observed, leaving only the short fragments (for example the very longest fragments may not efficiently bind to or elute from the SPRI beads used after end-prep or ligation). Light shearing, for example using the Megaruptor, can break up the very longest molecules into chunks that the library preparation can more readily process, leading to increased read N50s.

This approach is suggested for users where samples appear to be very high molecular weight in gel or FEMTO Pulse analysis but the observed read length N50 is <15 kb. Other users within the Nanopore Community have also attempted other shearing methods to increase read lengths. However, the more aggressive the fragmentation, the higher the risk of over-fragmenting, leading to a reduction in observed read lengths.



**Figure 1. The effect of shearing very high molecular weight gDNA with Megaruptor 3.**
Human gDNA was extracted from cell culture, with the aim to recover the longest possible fragments. The resulting gDNA was sheared with Megaruptor 3 using a selection of shearing speeds. The sheared DNA was analysed by FEMTO Pulse (panel A) and sequencing libraries were prepared using the Ligation Sequencing Kit and run on the MinION; read N50 values were recorded (panel B). The read length distribution of the input (no shear) shows that most of the DNA is above 100 kb, with a spike at 165 kb (area where fragments become compressed). However, this does not correspond to a high read N50 in sequencing. The lowest shearing speed had little-to-no effect on the fragment length distribution or the observed read N50, suggesting unsuccessful fragmentation. However, increasing the shearing speed to 20–30 did show successful fragmentation and led to an increase in observed read N50. Increasing the shearing speed still further leads to over-fragmentation and a drop in observed read N50.

**Input amount and pore occupancy**

Loading too much or too little library can compromise flow cell performance: 5–50 fmol is optimal (for R9 flow cells). For gDNA preparation using the Ligation Sequencing Kit, we recommend starting with 500-1000 ng. If insufficient starting material is available users can start with as little as 100 ng, however we have found that data output can drop at lower inputs as there are insufficient molecules available to maximise pore occupancy. Fragmenting the sample (for example using a Covaris g-TUBE or Megaruptor® ) can be used to increase the number of molecules/ends to thread into the nanopores in order to increase pore occupancy and recover the output: an input of 100 ng unsheared Lambda DNA results in a flow cell load of ~1 fmol, which can be increased to ~6 fmol by shearing with a g-TUBE. We would generally consider shearing for gDNA samples where input is 100–500 ng. However, fragmenting the DNA to boost the output means that it will not be possible to achieve ultra-long reads. If you have <100 ng of DNA, we advise performing PCR to increase the amount of DNA available for sequencing.



**Figure 2. The relationship between input and output for sheared and unsheared libraries.**
Panel A: as the input of unsheared gDNA into the library preparation drops below ~500 ng, the pore occupancy decreases, leading to a decrease in flow cell output (Gbases). Shearing the sample using a Covaris g-TUBE increases the molar concentration of the sample leading to more efficient use of the pores and an increase in flow cell output. Panel B: shearing samples with a g-TUBE can have an impact on the read length distribution.

**Blocking**

Pore blocking is another factor that can affect flow cell output. During a sequencing run, pores can become “blocked”, preventing the pore from accepting a new strand for sequencing or continuing to sequence the occupying strand. Such blocks are detected by the MinKNOW software, which changes the channel state from “single pore” to “unavailable”. For the duration of a blockade, the pore acquires no sequencing data. MinKNOW attempts to drive out whatever has blocked the pore by reversing the voltage. The unblocking scheme is progressive, increasing the duration of the voltage reversal until the blockage is cleared. Most of the time (~98%), attempts to unblock a pore are successful and it reverts to the single pore state, where it is available to accept new strands and continue sequencing. However, in a minority of cases, the progressive unblocking scheme will not be able to recover a blocked pore. In the case when a pore becomes terminally blocked and cannot be recovered, a new pore is swapped in from a different well in the channel, if available. Typically, a blockade occurs every 250–500 kb (Figure 3) and is successfully removed ~98% of the time (in other words, around 1 in 50 attempts will be unsuccessful). This gives an average output of ~10–20 Mbases per pore, which for a flow cell containing ~1500 pores could lead to a total output of ~15–30 Gbases (note, other factors may limit the actual throughput obtained). If there is an increase in the rate of blocking, then pores spend less time sequencing and more unblocks are triggered. If there are more unblocks, or if the success of unblocking decreases, then the rate at which pores are lost increases and total flow cell output is reduced.

To determine if fragment length played a role in the rate of blocking, we took DNA extracted from human cells grown in culture (GM12878) and sheared it with Megaruptor 2. We were not able to establish a relationship between read length and blocking rate (Figure 3), although we observed a decrease in the success rate of the unblock for the longer libraries, indicating that our unblocking scheme is less capable of removing blocks from longer fragments (Figure 3). Given this observation, if users are obtaining a low output, then some shearing of the sample could be performed to see if unblocking success can be improved to help boost output.



**Figure 3. The effect of read length on blocking, unblocking and flow cell output.** Extracted gDNA was sheared with Megaruptor 2 using different shearing settings, or with a Covaris g-TUBE. Libraries were then prepared using the Ligation Sequencing Kit and sequenced on the MinION. Panel A: the read N50 values for the differently sheared libraries. Note, some Megaruptor shearing produces a slightly elevated read N50. Panel B: the block frequency (kb/block) for the differently sheared libraries suggests that there is little relationship between the frequency at which blockades occur and the length of the fragments that are being sequenced, at least for this sample. Panel C: as the fragment length of the library increased, a decrease in rate of the success of the unblock (number of blocks before a terminal block was encountered) was observed. Panel D: as a result of the decrease in unblock success with increased read lengths, it follows that the output from any pore (and hence whole cell) may also decrease with increased read length.

**Shearing genomic DNA using MegaruptorⓇ 3**

**Introduction**

Ligation Sequencing Kit protocols contain no designated fragmentation step. However, fragmentation of genomic DNA has been shown in some cases to provide benefits in:

* Reduction in library input requirements
* Increasing read length N50
* Reduction in pore blocking

This protocol contains information regarding the use of Megaruptor® 3 (Diagenode) for fragmenting gDNA for subsequent library preparation using the Ligation Sequencing Kit. Size selection of gDNA (for example, using the protocol ‘Size selection of HMW DNA by semi-selective DNA precipitation’) can be performed before Megaruptor® 3 shearing to help remove the shortest fragments.

**Materials**

* 2000 ng gDNA in 100 μl nuclease-free water
* Diagenode Megaruptor® 3
* Megaruptor 3 Shearing Kit : Cat. No. E07010003
* Nuclease-free water
* DNA QC equipment such as the Agilent FEMTO Pulse
* Microfuge

**Method**

1. *Optional step:* Perform a size selection on your DNA sample using the [Size selection of HMW DNA by semi-selective DNA precipitation protocol](https://community.nanoporetech.com/extraction_methods/size-selection2).
2. Transfer 2000 ng of genomic DNA into a 0.5 ml sample tube supplied in the Megaruptor 3 shearing kit, and adjust the volume to 100 μl with nuclease-free water. Mix the DNA thoroughly by pipetting the sample. Spin down briefly in a microfuge.
3. Transfer the sample to the Megaruptor 3, enter the shearing parameters into the instrument, and start the run.
4. When the shearing step is complete, remove the fragmented DNA from the instrument.
5. Analyse 1 μl of the fragmented DNA for fragment size (for example using the Agilent FEMTO Pulse).
6. Use 1 µg of fragmented DNA in Ligation Sequencing Kit library preparation.

**Results**



**Figure 1.** Fragment length profile after Megaruptor 3 shearing. Input DNA was fragmented using a variety of shearing speeds and the resulting DNA was analysed by FEMTO Pulse. A shearing speed of 3 appears to have had little impact on the fragment length profile, suggesting unsuccessful fragmentation. As the shearing speed is increased, the sample becomes progressively more fragmented.

**Change log**

| **Version** | **Change** |
| --- | --- |
| v1, September 2019 | Initial protocol publication |