Video Article

Ultra-Long Read Sequencing for Whole Genomic DNA Analysis

Liang Gong[1], Chee-Hong Wong[1], Jennifer Idol[1], Chew Yee Ngan[1], Chia-Lin Wei[1]

[1]Genome Technologies, Jackson Laboratory for Genomic Medicine

Correspondence to: Chia-Lin Wei at Chia-lin.Wei@jax.org

Abstract

Third generation single-molecule DNA sequencing technologies offer significantly longer read length that can facilitate the assembly of complex genomes and analysis of complex structural variants. Nanopore platforms perform single-molecule sequencing by directly measuring the current changes mediated by DNA passage through the pores and can generate hundreds of kilobase (kb) reads with minimal capital cost. This platform has been adopted by many researchers for a variety of applications. Achieving longer sequencing read lengths is the most critical factor to leverage the value of nanopore sequencing platforms. To generate ultra-long reads, special consideration is required to avoid DNA breakages and gain efficiency to generate productive sequencing templates. Here, we provide the detailed protocol of ultra-long DNA sequencing including high molecular weight (HMW) DNA extraction from fresh or frozen cells, library construction by mechanical shearing or transposase fragmentation, and sequencing on a nanopore device. From 20-25 μg of HMW DNA, the method can achieve N50 read length of 50-70 kb with mechanical shearing and N50 of 90-100 kb read length with transposase mediated fragmentation. The protocol can be applied to DNA extracted from mammalian cells to perform whole genome sequencing for the detection of structural variants and genome assembly. Additional improvements on the DNA extraction and enzymatic reactions will further increase the read length and expand its utility.

Introduction

Over the past decade, massively parallel and highly accurate second-generation high-throughput sequencing technologies have driven an explosion of biomedical discovery and technological innovation[1,2,3]. Despite the technical advances, the short-read data generated by the second-generation platforms are ineffective in resolving complex genomic regions and are limited in the detection of genomic structural variants (SVs), which play important roles in human evolution and diseases[4,5]. Furthermore, short-read data are unable to resolve repeat variation and are unsuitable for discerning haplotype phasing of genetic variants[6].

Recent progress in single-molecule sequencing offers significantly longer read length, which can facilitate the detection of the full spectrum of SVs[7,8,9], and offers accurate and complete assembly of complex microbial and mammalian genomes[6,10]. The nanopore platform performs single-molecule sequencing by directly measuring the current changes mediated by DNA passage through the pores[11,12,13]. Unlike any existing DNA sequencing chemistry, nanopore sequencing can generate long (tens to thousands of kilobases) reads in real-time without relying on polymerase kinetics or artificial amplification of the DNA sample. Therefore, nanopore long-read sequencing (NLR-seq) holds great promise for generating ultra-long read lengths well beyond 100 kb, which would greatly advance genomic and biomedical analyses[14], particularly in the low-complexity or repeat-rich regions of the genomes[15].

The unique feature of nanopore sequencing is its potential to generate long reads without a theoretical length limitation. Therefore, the read length is dependent on the physical length of the DNA which is directly affected by the DNA integrity and sequencing template quality. Moreover, depending on the extent of manipulation and the number of steps involved, such as pipetting forces and extraction conditions, the quality of the DNA is highly variable. Therefore, it is challenging for one to yield long reads by just applying the standard DNA extraction protocols and manufacturer's supplied library construction methods. Toward this end, we have developed a robust method to generate ultra-long read (hundreds of kilobases) sequencing data starting from harvested cell pellets. We adopted multiple improvements in the DNA extraction and library preparation procedures. We streamlined the protocol to exclude unnecessary procedures that cause DNA degradation and damages. This protocol is composed of high molecular weight (HMW) DNA extraction, ultra-long DNA library construction, and sequencing on a nanopore platform. For a well-trained molecular biologist, it typically takes 6 h from cell harvesting to the completion of HMW DNA extraction, 90 min or 8 h for library construction depending on the shearing method, and up to a further 48 h for DNA sequencing. The use of the protocol will empower the genomics community to improve our understanding of genome complexity and gain new insight into genome variation in human diseases.

Protocol

NOTE: The NLR-seq protocol consists of three consecutive steps: 1) extraction of high-molecular weight (HMW) genomic DNA; 2) ultra-long DNA library construction, which includes fragmentation of the HMW DNA into the desired sizes and ligation of sequencing adapters to the DNA ends; and 3) loading of the adapter-ligated DNA onto the arrays of nanopores (**Figure 1**).

1. HMW DNA extraction

1. Reagent setup. Make 1x phosphate buffered saline (PBS) buffer (1,000 mL) by adding 100 mL of PBS (10x) to 900 mL of water and mix well. Make lysis buffer (500 mL) by adding 43.5 mL of water to a 50 mL tube. Add 500 µL of Tris (1 M, pH 8.0), 1 mL of sodium chloride (NaCl) (5 M), 2.5 mL of ethylenediaminetetraacetic acid (EDTA) (0.5 M, pH 8.0) and 2.5 mL of sodium dodecyl sulfate (SDS) (10%, wt/vol) to the tube and mix well.
   NOTE: This PBS buffer can be stored at 4 °C for up to 6 months. The premade lysis buffer can be stored at RT for up to 2 months.
2. Check the cell mortality and count the cells. Ensure that the live ratio is > 85% and the total cell number is 30 x 10^6.
   NOTE: The cells used in this protocol are from the HG00733 cell line, a human lymphoblastoid cell line of Puerto Rican origin widely used in the 1000 Genome consortium for structural variation analysis (see table of materials for ordering information), which belongs to International Genome Sample Resource.
3. Collect the cells by centrifuging at 200 x g for 5 min at RT. Discard the medium and resuspend the cell pellet (30 x 10^6 cells) with 5 mL of 1x PBS buffer. Centrifuge again at 200 x g for 5 min at RT and discard the supernatant.
   NOTE: 25-35 x 10^6 of cells are acceptable for this approach. Further variation in the amount of cells used will need further optimization. The cell pellet can be stored at −80 °C for up to 6 months.
4. Resuspend the cell pellet in 200 µL of 1x PBS buffer. If using a frozen cell pellet, wash with 5 mL of 1x PBS buffer. Centrifuge the solution at 200 x g for 5 min at RT, discard the supernatant and resuspend the cells in 200 µL of 1x PBS buffer.
5. Prepare 10 mL of lysis buffer in a 50 mL tube. Add the 200 µL cell suspension to the lysis buffer and vortex at the highest speed for 3 s. Incubate the solution at 37 °C for 1 h.
6. Add 2 µL of RNase A (100 mg/mL) to the lysate. Gently rotate the 50 mL tube to mix the sample. Incubate the solution at 37 °C for 1 h.
7. Add 50 µL proteinase K (20 mg/mL) to the lysate. Gently rotate the 50 mL tube to mix the sample. Incubate the solution at 50 °C for 2 h. During incubation, gently mix the sample every 30 min.
8. Remove the 50 mL tube from 50 °C and let stand at RT for 5 min.
9. Add 10 mL of the phenol layer of phenol:chloroform:isoamyl alcohol (25:24:1, vol/vol/vol) to the lysate and rotate the tube on a rotator mixer (see **Table of Materials**) at RT in a fume hood at 20 rpm for 10 min. Wrap the tube cap with parafilm to prevent leakage during rotation.
10. Prepare two 50 mL gel tubes (see **Table of Materials**) by centrifuging at 1,500 x g for 2 min at RT.
    NOTE: The gel forms a stable barrier between the nucleic acid-containing aqueous phase and the organic solvent.
11. Pour the sample/phenol solution into one of the prepared 50 mL gel tubes from Step 1.10. Centrifuge the solution at 3,000 x g for 10 min at RT.
12. Pour the supernatant into a new 50 mL tube. Add 10 mL of the phenol layer of phenol:chloroform:isoamyl alcohol (25:24:1, vol/vol/vol) and rotate the tube on a rotator mixer at RT in a fume hood at 20 rpm for 10 min.
13. Repeat step 1.11 once with the second prepared gel tube.
14. Pour the supernatant into a new 50 mL tube. Add 25 mL of ice-cold 100% ethanol and gently rotate the tube by hand until the DNA precipitates (**Figure 2**).
    NOTE: The precipitation approach helps to stabilize the HMW DNA.
15. Bend a 20 µL tip to make a hook. Carefully take out the HMW DNA with the hook and let the liquid drop off.
16. Place the HMW DNA into a 50 mL tube containing 40 mL of 70% ethanol. Wash the DNA by gently inverting the tube 3 times.
17. Repeat step 1.15 once to collect DNA from the 70% ethanol tube.
18. Place the HMW DNA into a 2 mL tube containing 1.8 mL of 70% ethanol.
19. Centrifuge the washed HMW DNA at 10,000 x g for 3 s at RT. Remove as much of the residual ethanol as possible by pipetting.
    NOTE: Do not disturb the DNA pellet when pipetting the residual ethanol.
20. Incubate the 2 mL tube at 37 °C for 10 min with the lid open to dry the sample.
    1. If continuing with step 2.1 (with mechanical shearing and 1D Ligation Sequencing Kit), add 1 mL of TE (10 mM Tris and 1 mM EDTA, pH 8.0) to the 2 mL tube.
    2. If continuing with Step 2.2 (with transposase-based fragmentation and Rapid Sequencing Kit), add 200 µL of 10 mM Tris (pH 8.0) with 0.02% Triton X-100.
       NOTE: Do not disturb the DNA pellet. Letting the tube stand at 4 °C in dark for 48 h will help the sample fully resuspend. The HMW DNA can be stored at 4 °C for up to 2 weeks. Longer storage time or other storage conditions may introduce more short fragments.

2. Ultra-long DNA library construction

NOTE: There are two ways to construct the ultra-long DNA libraries based on two different shearing methods coupled with nanopore sequencing kits. A mechanical shearing-based library produces data with an N50 of 50-70 kb, taking about 8 h for the library construction. A transposase fragmentation-based library produces an N50 of 90-100 kb data, taking only 90 min for the library construction. The mechanical shearing protocol gives higher yield from the same DNA input using identical versions of sequencing adapter and quality of nanopore flow cells.

1. Mechanical shearing-based library construction
   1. Thaw and mix the reagents from the ligation kit (see **Table of Materials**). Thaw FFPE DNA repair buffer and end repair/dA-tailing buffer on ice, then vortex and spin down to mix. Thaw adapter mix (AMX) and adapter bead binding buffer (ABB) on ice, then pipette and spin down to mix. Thaw running buffer with fuel mix (RBF) and elution buffer (ELB) at RT, then vortex and spin down to mix. Thaw library loading beads (LLB) at RT and pipette to mix before use.
      1. Once thawed, keep all kit components on ice. Take out the enzymes only when needed. Bring the magnetic beads to RT for use.
         NOTE: For recommendations on the magnetic beads to use see the table of materials.
   2. Check the quality and quantity of the HMW DNA from step 1.21.1. Pipette out 20 µL of DNA into new 1.5 mL tubes from three different locations in the HMW DNA tube using P200 wide bore tips. Take 1 µL from the three aliquots to detect the concentration using a fluorometer and the quality using a UV reading. Check multiple times to confirm the results.

NOTE: The expected results are shown in **Figure 3A**. The $OD_{260/280}$ value is approximately 1.9 and the $OD_{260/230}$ value is approximately 2.3.

3. Transfer the remaining 940 µL of HMW DNA into a 50 mL tube cap with a P1000 wide bore tip.
4. Aspirate all DNA into a 1 mL syringe without the needle.
5. Put the 27 G needle onto the syringe and eject all DNA into the cap gently and slowly (~10 s). Take off the 27 G needle from the syringe.
6. Repeat steps 2.1.4 and 2.1.5 for 29 times for a total of 30 passes through the needle.
   NOTE: The sheared HMW DNA can be stored at 4 °C in the dark for up to 24 h. Quality control (QC) is highly recommended by pulsed-field gel electrophoresis, but it is costly and time consuming. If performing QC on an automated pulse field gel electrophoresis machine use a 5-150 kb protocol for a 20 h run. The expected results are shown in **Figure 4**.
7. Prepare the DNA repair reaction in a 0.2 mL tube by adding 100 µL of sheared HMW DNA (20 µg), 15 µL of FFPE DNA repair buffer, 12 µL of FFPE DNA repair mix, and 16 µL of nuclease-free water. Mix the reaction by flicking gently 6 times and spin down to remove bubbles.
8. Incubate the reaction at 20 °C for 60 min. Transfer the sample into a new 1.5 mL tube with a P200 wide bore tip.
9. Resuspend the magnetic beads by pipetting or vortexing. Add 143 µL beads (1x) to the DNA repair reaction and mix gently by flicking the tube 6 times. Rotate the tube on a rotator mixer at RT at 20 rpm for 30 min.
10. Spin down the sample at 1,000 x *g* for 2 s at RT. Place the tube on a magnetic rack for 10 min. Keep the tube on the magnetic rack and discard the supernatant.
11. Keeping the tube on the magnetic rack, add 400 µL of freshly prepared 70% ethanol without disturbing the pellet. Remove the 70% ethanol after 30 s.
12. Repeat step 2.1.11 once.
13. Spin down the sample at 1,000 x *g* for 2 s at RT. Place the tube back on the magnetic rack. Remove any residual ethanol and air dry for 30 s. Do not over dry the pellet.
14. Remove the tube from the magnetic rack and add 103 µL of TE (10 mM Tris and 1 mM EDTA, pH 8.0). Gently flick the tube to ensure that beads are covered in the buffer, and incubate on a rotator mixer at RT for 30 min. Gently flick the tube every 5 min to aid resuspension of the pellet.
15. Pellet the beads on the magnetic rack for at least 10 min. Transfer 100 µL of eluate with a P200 wide bore tip into a 0.2 mL tube.
16. Prepare the end repair and dA-tailing reaction in a 0.2 mL tube by adding 100 µL of repaired HMW DNA, 14 µL of end repair/dA-tailing buffer and 7 µL of end repair/dA-tailing mix. Mix the reaction by flicking gently 6 times, and spin down to remove bubbles.
17. Incubate the reaction at 20 °C for 60 min followed by 65 °C for 20 min, and then hold at 22°C. Transfer the sample into a new 1.5 mL tube using a P200 wide bore tip.
18. Resuspend the magnetic beads by pipetting or vortexing. Add 48 µL of beads (0.4x) to the end repair/dA-tailing reaction and mix gently by flicking the tube 6 times. Rotate the tube on a rotator mixer at RT at 20 rpm for 30 min.
19. Repeat steps 2.1.10-2.1.13 once.
20. Remove the tube from the magnetic rack and add 33 µL of TE (10 mM Tris and 1 mM EDTA, pH 8.0). Gently flick the tube to ensure that beads are covered in the buffer, and incubate on a rotator mixer at RT for 30 min. Gently flick the tube every 5 min to aid resuspension of the pellet.
21. Pellet the beads on the magnetic rack for at least 10 min. Transfer 30 µL of eluate with a P200 wide bore tip into a new 1.5 mL tube. Take the extra 1-2 µL to detect the concentration using a fluorometer.
   NOTE: Recovery of 5-6 µg at this step is expected.
22. Prepare the ligation reaction in the 1.5 mL sample tube by adding 30 µL of end-repaired HMW DNA, 20 µL of adapter mix (AMX 1D), and 50 µL of blunt/TA ligation master mix. Mix the reaction by flicking gently 6 times between each sequential addition and spin down to remove bubbles.
23. Incubate the reaction at RT for 60 min.
24. Resuspend the magnetic beads by pipetting or vortexing. Add 40 µL beads (0.4x) to the ligation reaction and mix gently by flicking the tube 6 times. Rotate the tube on a rotator mixer at RT at 20 rpm for 30 min.
25. Repeat step 2.1.10 once.
26. Add 400 µL of adapter bead binding (ABB) buffer into the tube. Flick the tube gently 6 times to resuspend the beads. Place the tube back on the magnetic rack to separate the beads from the buffer and discard supernatant.
27. Repeat step 2.1.26 once.
28. Spin down the sample at 1,000 x *g* for 2 s at RT. Place the tube back on the magnetic rack. Remove any residual buffer and air dry for 30 s. Do not over dry the pellet.
29. Remove the tube from the magnetic rack and resuspend the pellet in 43 µL of elution buffer. Gently flick the tube to ensure that beads are covered in the buffer and incubate on a rotator mixer at RT for 30 min. Gently flick the tube every 5 min to aid resuspension of the pellet.
30. Pellet the beads on the magnetic rack for at least 10 min. Transfer 40 µL of eluate with a P200 wide bore tip into a new 1.5 mL tube. Take the extra 1-2 µL to detect the concentration using a fluorometer.
   NOTE: Recovery of 1-2 µg at this step is expected. The mechanical shearing-based library is ready for loading. The library can be stored on ice for up to 2 h until loading for sequencing if needed.

2. **Transposase fragmentation-based library construction**
   1. Thaw the reagents from the transposase kit (see **Table of Materials**). Thaw fragmentation mix (FRA) and rapid adapter (RAP) on ice and pipette to mix. Thaw sequencing buffer (SQB), loading beads (LB), flush buffer (FLB) and flush tether (FLT) at RT and pipette to mix. Thaw loading beads (LB) at RT and pipette to mix before use. Once thawed, keep all kit components on ice. Take out the enzymes only when needed.
   2. Check the quality and quantity of the HMW DNA from step 1.21.2. Pipette out 20 µL of DNA into new 1.5 mL tubes from three different locations in the HMW DNA tube using P200 wide bore tips. Take 1 µL from the three aliquots to detect the concentration using a fluorometer and the quality using a UV reading. Check multiple times to confirm the results.
   NOTE: The expected results are shown in **Figure 3B**. The $OD_{260/280}$ value is approximately 1.9 and the $OD_{260/230}$ value is approximately 2.3.

3. Prepare the DNA tagmentation reaction in a 0.2 mL tube by adding 22 µL of HMW DNA, 1 µL of 10 mM Tris (pH 8.0) with 0.02% Triton X-100 and 1 µL of fragmentation mix (FRA). Mix by pipetting with a P200 wide bore tip as slowly as possible 6 times, taking care not to introduce bubbles.
4. Incubate the reaction at 30 °C for 1 min followed by 80 °C for 1 min, and then hold at 4 °C. Transfer the mix into a new 1.5 mL tube with a P200 wide bore tip and go to next step immediately.
5. Add 1 µL of rapid adapter (RAP) to the 1.5 mL sample tube. Mix by pipetting with a P200 wide bore tip as slowly as possible 6 times, taking care not to introduce bubbles.
6. Incubate the reaction at RT for 60 min.
   NOTE: The transposase fragmentation-based library is ready for loading. The library can be stored on ice for up to 2 h until loading for sequencing if needed.

3. Sequencing on the nanopore device

1. Check the nanopore sequencing device (see **Table of Materials**). Make sure both the software and hardware are working and there is enough storage space.
2. Check the flow cell. Open a new flow cell and insert the flow cell into the nanopore device. Check the box of the location the flow cell was inserted into (X1-X5). Select the correct flow cell type. Click on the **Check Flow Cells** workflow. Click on the **Start Test** button to start the flow cell QC analysis.
   NOTE: If the reported total active pore number is less than 800, use a different new flow cell for sequencing.
3. Prepare the priming buffer. For a mechanical shearing-based library, add 576 µL of running buffer with fuel mix (RBF) and 624 µL of nuclease-free water into a 1.5 mL tube. Vortex and spin down to mix the priming buffer. For a transposase fragmentation-based library, add 30 µL of flush tether (FLT) to the tube of flush buffer (FLB). Vortex and spin down to mix the priming buffer.
4. On the flow cell, move the priming port cover clockwise to expose the priming port.
5. Set a P1000 pipette to 100 µL and insert the tip into the priming port. Draw back a small volume of buffer (less than 30 µL) to remove any bubbles from the flow cell. Stop pipetting once a small amount of yellow fluid enters the tip.
6. Use a P1000 pipette to load 800 µL of the priming mix into the flow cell via the priming port. To avoid introducing bubbles, add 30 µL of the priming mix to cover the top of the priming port first, then insert the tip into the priming port and slowly add the rest of the priming mix. Take out the tip when there is about 50 µL left. Add the rest of priming mix on the top of the priming port. The fluid will go inside by itself.
7. Leave the setup to incubate for 5 min. In the meantime, prepare the library mix in the 1.5 mL tube containing the library.
   NOTE: For a mechanical shearing-based library add 35 µL of running buffer with fuel mix (RBF) to 40 µL of the DNA library. For a transposase fragmentation-based library add 34 µL of sequencing buffer (SQB) and 16 µL of nuclease-free water to 25 µL of the DNA library.
8. Open the flow cell sample port cover gently to expose the sample port. Use a P1000 pipette to add 200 µL of the priming mix through the priming port into the flow cell as described in step 3.5. Make sure that the priming mix is not loaded into the flow cell through the sample port.
9. Set a P200 pipette to 80 µL. Mix the library gently with a wide bore tip by pipetting up and down 6 times just prior to loading.
10. Load the library mix dropwise through the sample port into the flow cell. Add each drop only after the previous drop is completely loaded into the port.
11. Put back the sample port cover gently and make sure the sample port is fully covered. Move the priming port cover anticlockwise to cover the priming port. Close the device lid.
12. Click on the **New Experiment** workflow. Type the library name, select the correct kit according to procedures used, and check that the settings are correct (48 h run, real-time base-calling ON).
13. Click **Start Run**. After 10 min, record the flow cell ID and the active nanopore numbers (total number and each four groups' numbers) from the run information.
14. Data analysis. Copy the data to a local computer or a cluster at any time of the sequencing or when the run is complete. Use Minimap2[16] (https://github.com/lh3/minimap2) to align the sequence data to the reference genome. Summarize the sequencing performance from the raw sequence data and the alignments by NanoPlot[17] (https://github.com/wdecoster/NanoPlot).

Representative Results

The ultra-long DNA sequencing protocol applies HMW DNA for library construction. Therefore, it is critical to choose well-cultured cells with the live ratio >85% at the cell harvesting step. The amount of cells used for DNA extraction will affect the quality and the quantity of the HMW DNA. The cell lysis does not work well if starting with too many cells. Using too few cells does not generate enough DNA for library construction because the HMW DNA precipitation is performed using gentle rotation by hand instead of high-speed centrifugation. An example of the HMW DNA after adding ice-cold 100% ethanol and rotating is shown as the white cotton-like precipitate in **Figure 2**.

It is important to check the quality of the input DNA before beginning the library construction. Degradation, incorrect quantification, contamination (e.g., proteins, RNAs, detergents, surfactant, and residual phenol or ethanol) and low molecular weight DNA can have a significant effect on the subsequent procedures and on the final read length. We recommend performing the QC analysis using the DNA from three different locations in the tube containing HMW DNA. From UV reading results for the HMW DNA, the OD_{260}/OD_{280} value is approximately 1.9 and the OD_{260}/OD_{230} value is approximately 2.3 (**Figure 3A,B**). These ratio values are consistent among the three tests for a good HMW DNA sample. Different shearing methods requires different volumes of input DNA. The concentration of HMW DNA needs to be >200 ng/µL for mechanical shearing while it needs to be >1 µg/µL for transposase fragmentation. The concentration detected by a fluorometer is a little lower than UV reading. However, the coefficient of variation of the concentration of the same HMW DNA sample is required to be less than 15% with both the fluorometer and the UV reading assays. Mechanical shearing applies a syringe with a needle to break the HMW DNA so that the number of passes through the needle will impact the size of the sheared DNA and the final read length. It is recommended to perform size QC after needle shearing to ensure the majority of the HMW DNA is larger than 50 kb as illustrated in **Figure 4**. In the mechanical shearing method, 30 passes generated the best sequencing results considering both the length and output.

The N50 of a mechanical shearing-based library is 50-70 kb while a transposase fragmentation-based library is 90-100 kb. The results of four runs using the HG00733 cell line are shown in **Table 1**. All four runs have over 2,300 reads with length longer than 100 kb. The maximum length is longer in the transposase fragmentation-based libraries (455 kb and 489 kb) compared with the mechanical shearing-based libraries (348 kb and 387 kb) while the latter produced more total reads, indicating a higher yield. The transposase fragmentation-based library construction has fewer steps and shorter preparation time so that it will introduce fewer short fragments. The two runs using transposase have a longer mean length (>30 kb) and median length (>10 kb). In addition, the data shows consistent high quality in all runs (mean quality score is approximately 10.0, ~90% base accuracy). More than 97% of the total bases were aligned to the human reference genome (hg19) using Minimap2[16] with the default settings. The expected size distributions of the raw reads are shown in **Figure 5**. All runs have a large proportion of data above 50 kb while transposase fragmentation-based libraries have a higher ratio of ultra-long reads (e.g. > 100 kb). This protocol has been successfully applied in multiple human cell lines (Supplementary Table 1).
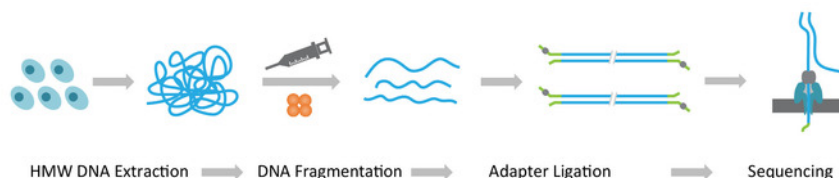

Figure 1: Schematic overview of the nanopore long-read sequencing (NLR-seq) workflow. Orange, the transposase complex. Yellow-green, the nanopore adapter. Please click here to view a larger version of this figure.
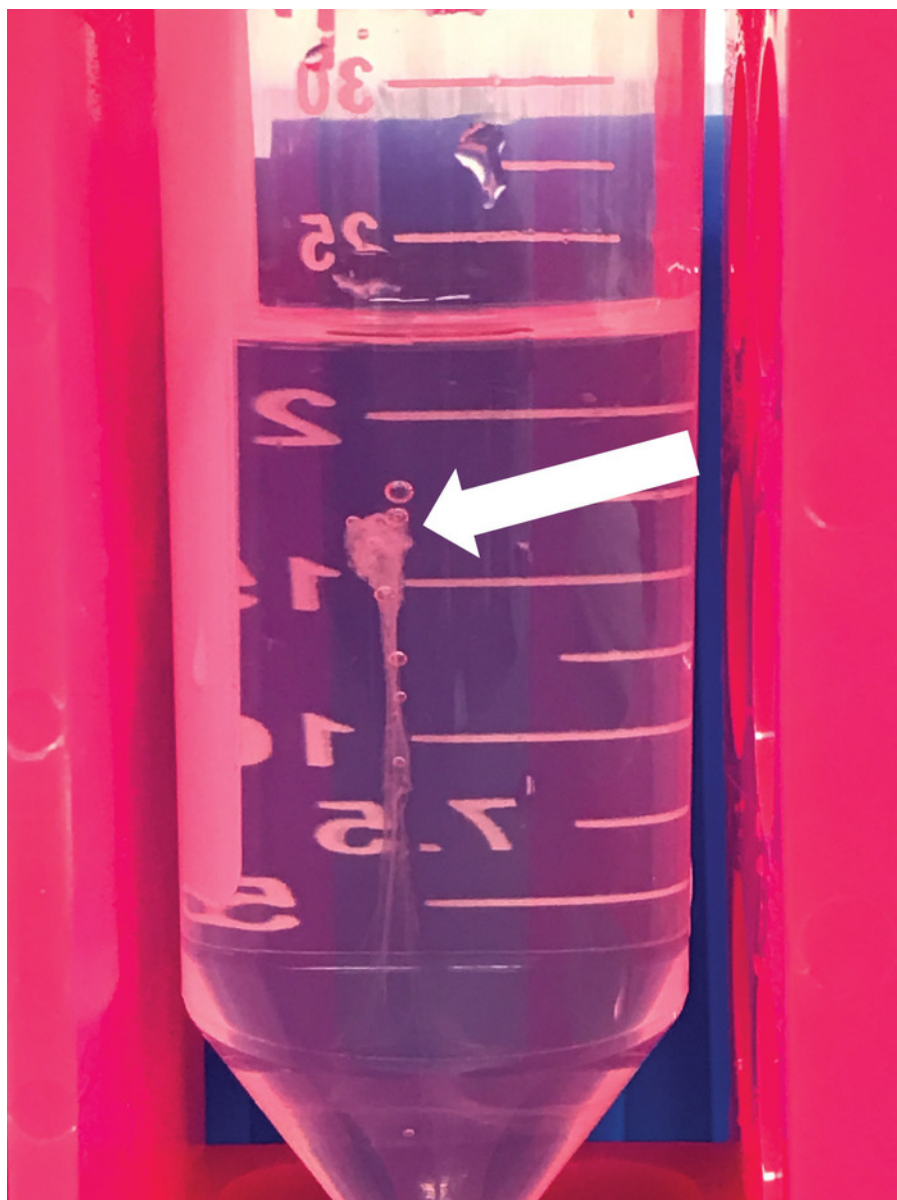
Figure 2: Representative DNA precipitation from phenol-chloroform extraction method. The white arrow indicates the HMW DNA. Please click here to view a larger version of this figure.
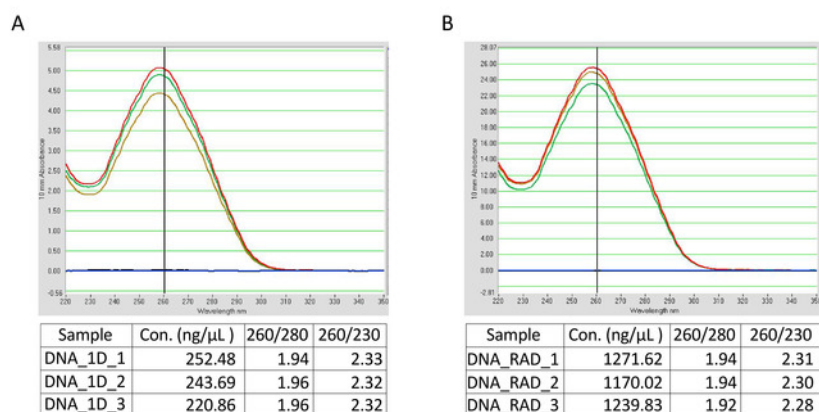
Figure 3: Example QC results of the HMW DNA from UV reading. (**A**) HMW DNA from step 1.21.1 ready for mechanical shearing-based library construction. (**B**) HMW DNA from step 1.21.2 for transposase fragmentation-based library construction. Please click here to view a larger version of this figure.
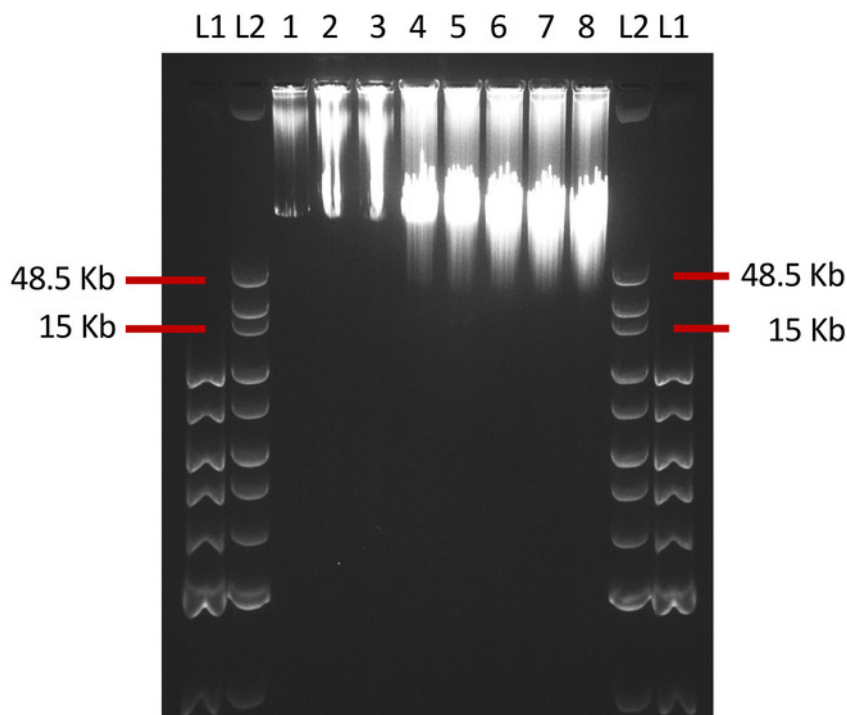

Figure 4: QC results of the needle sheared HMW DNA by pulsed-field gel electrophoresis. L1: Quick-load 1 kb DNA ladder; L2: Quick-load 1 kb extend DNA ladder. 1-8: DNA with different passing times through the needle shearing. 1-3, no shearing; 4, 10 times; 5, 20 times; 6, 30 times; 7, 40 times; 8, 50 times. This QC step is optional. Please click here to view a larger version of this figure.
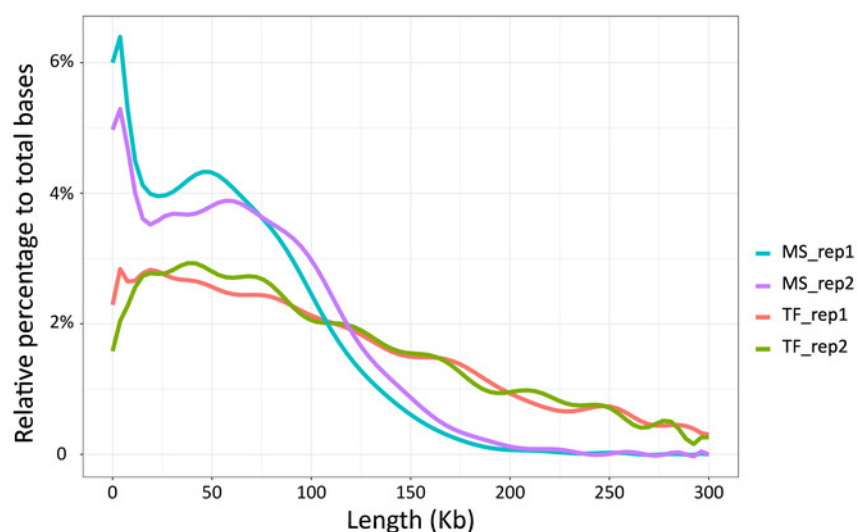
Figure 5: Expected size distributions of the nanopore ultra-long DNA libraries. MS, mechanical shearing-based libraries. TF, transposase fragmentation-based libraries. Please click here to view a larger version of this figure.

	Mechanical shearing_rep1	Mechanical shearing_rep2	Transposase fragmentation_rep1	Transposase fragmentation_rep2
Cell line	HG00733	HG00733	HG00733	HG00733
N50 of the reads	55,180	63,007	98,237	95,629
Number of reads longer than 100 Kb	2,500	3,082	2,386	2,355
Number of total reads	97,859	80,465	24,166	21,032
Maximum length (bp)	348,482	387,113	454,660	489,426
Mean length (bp)	17,861	20,395	33,528	38,175
Median length (bp)	5,335	5,894	10,249	15,656
Mean quality of the reads	10.0	10.1	9.9	10.0
Total bases of raw reads	1,747,849,822	1,641,058,932	810,229,733	802,886,304
Total bases of aligned reads	1,693,300,832	1,607,975,925	791,422,077	778,417,627
Mapped ratio of total bases (hg19, Minimap2)	96.9%	98.0%	97.7%	97.0%
Number of active pores	1225: 480, 402, 254, 89	1058: 480, 356, 176, 46	958: 452, 328, 148, 30	1092: 487, 367, 195, 43

Table 1: Performance metrics summary from runs with different shearing protocols.

	Library 1	Library 2
Cell line	K562	GM19240
Cell Ordering Information	ATCC, cat. No. CCL-243	Coriell Institute, cat. No. GM19240
Protocol	mechanical shearing	mechanical shearing
N50 of the reads	60,063	55,295
Number of total reads	193,783	120,807
Median length (bp)	1,843	4,688
Mean length (bp)	9,825	17,408
Maximum length (bp)	548,780	212,338
Total bases of raw reads	1,903,989,686	2,103,015,331
Total bases of aligned reads	1,837,350,047	1,997,419,761
Mapped ratio of total bases (hg19, Minimap2)	96.6%	95.0%
Number of active pores	1111: 482, 371, 203, 55	1032: 447, 333, 196, 56

Supplementary Table 1: Summary of two NLR-seq runs using other cell lines with the mechanical shearing protocol.

Discussion

In principle, nanopore sequencing is able to generate 100 kb to megabase reads in length[11,12,13]. Four major factors will affect the performance of the sequencing run and data quality: 1) active pore numbers and the activity of the pores; 2) motor protein, which controls the speed of DNA passing through the nanopore; 3) DNA template (length, purity, quality, mass); 4) sequencing adapter ligation efficiency, which determines the useable DNA from the input sample. The first two factors depend on the version of the flow cell and the sequencing kit provided by the manufacturer. The second two factors are critical steps in this protocol (HMW DNA extraction, shearing and ligation).

This protocol requires patience and practice. The quality of HMW DNA is important for ultra-long DNA libraries[6]. The protocol starts with cells collected with high viability (>85% viable cell preferred), limiting the degraded DNA from dead cells. Any harsh process which may introduce damages to the DNA (e.g., strong disturbing, shaking, vortex, multiple pipetting, repeated freezing and thawing) should be avoided. In the design of the protocol, we omit pipetting in the entire process of DNA extraction. Wide bore tips need to be used when pipetting is necessary after the mechanical shearing during library construction and sequencing. As the nanopores are sensitive to the chemistries in the chamber buffer[12], there should be as few residual contaminants (e.g., the detergents, surfactants, phenol, ethanol, proteins RNAs, etc.) as possible in the DNA. Considering the length and yield, the phenol extraction method shows the best and most reproducible results compared with multiple different extraction methods tested so far.

Despite the ability of this protocol to produce long-read sequences, several limitations still remain. First, this protocol was optimized based on the nanopore sequencing device available at time of publication; thus, it is limited to the selective nanopore-based sequencing chemistry and could be suboptimal when performed in other types of long-read sequencing devices. Second, the outcome is highly dependent on the quality of the DNA extracted from the starting material (tissues or cells). Read length will be compromised if the starting DNA is already degraded or damaged. Third, although multiple QC steps are incorporated in the protocol to check the DNA quality, the final yield and length of the reads can be affected by the flow cell and pore activity, which could be variable at this early stage of nanopore sequencing platform development.

The protocol described here uses human suspension cell line samples for DNA extraction. We have optimized the passing times in needle shearing, the ratio of HMW DNA to transposase and the ligation time to produce the described results. The protocol can be expanded in four ways. First, users can start with other cultured mammalian cells and with different amount of cells, tissues, clinical samples, or other organisms. Further optimization on lysis incubation time, reaction volume and centrifugation will be needed. Second, it is hard to predict the target size for ultra-long read sequencing. If the read lengths are shorter than expected, the users can adjust the passing times in the mechanical shearing-based method or change the ratio of the HMW DNA to transposase in the transposase fragmentation-based method. Longer binding and elution time during cleanup steps are helpful because the HMW DNA is highly viscous. Third, with different nanopore sequencing devices, one can adjust the amount and volume of the DNA to meet the criteria of the sequencer. Fourth, only those DNA ligated to sequencing adapters will be sequenced. To further improve ligation efficiency, one can attempt to titrate the adapter and ligase concentrations. Modified ligation time and molecular crowding agents such as PEG[18] can be applied in future. The ultra-long DNA sequencing protocol combined with CRISPR[19,20] may offer an effective tool for target enrichment sequencing.

Disclosures

The authors declare that they have no competing financial interests.

Acknowledgements

References

1. Mardis, E. R. Next-generation sequencing platforms. *Annual Review of Analytical Chemistry.* **6** 287-303 (2013).
2. Goodwin, S., McPherson, J. D., McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics.* **17** (6), 333-351 (2016).
3. Shendure, J. et al. DNA sequencing at 40: past, present and future. *Nature.* **550** (7676), 345-353 (2017).
4. Alkan, C., Coe, B. P., Eichler, E. E. Genome structural variation discovery and genotyping. *Nature Reviews Genetics.* **12** (5), 363-376 (2011).
5. Weischenfeldt, J., Symmons, O., Spitz, F., Korbel, J. O. Phenotypic impact of genomic structural variation: insights from and for human disease. *Nature Reviews Genetics.* **14** (2), 125-138 (2013).
6. Pollard, M. O., Gurdasani, D., Mentzer, A. J., Porter, T., Sandhu, M. S. Long reads: their purpose and place. *Human Molecular Genetics.* **27** (R2), R234-r241 (2018).
7. Cretu Stancu, M. et al. Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nature Communications.* **8** (1), 1326 (2017).
8. Gong, L. et al. Picky comprehensively detects high-resolution structural variants in nanopore long reads. *Nature Methods.* **15** (6), 455-460 (2018).
9. Sedlazeck, F. J. et al. Accurate detection of complex structural variations using single-molecule sequencing. *Nature Methods.* **15** (6), 461-468 (2018).
10. Jain, M. et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology.* **36** (4), 338-345 (2018).
11. Jain, M. et al. Improved data analysis for the MinION nanopore sequencer. *Nature Methods.* **12** (4), 351-356 (2015).
12. Deamer, D., Akeson, M., Branton, D. Three decades of nanopore sequencing. *Nature Biotechnology.* **34** (5), 518-524 (2016).
13. Jain, M., Olsen, H. E., Paten, B., Akeson, M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology.* **17** (1), 239 (2016).
14. Editorial. The long view on sequencing. *Nature Biotechnology.* **36** (4), 287 (2018).
15. Jain, M. et al. Linear assembly of a human centromere on the Y chromosome. *Nature Biotechnology.* **36** (4), 321-323 (2018).
16. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* **34**, 3094-3100 (2018).
17. De Coster, W., D'Hert, S., Schultz, D.T., Cruts, M., Van Broeckhoven, C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics.* **34**, 2666-2669 (2018).
18. Akabayov, B., Akabayov, S. R., Lee, S. J., Wagner, G., Richardson, C. C. Impact of macromolecular crowding on DNA replication. *Nature Communications.* **4**, 1615 (2013).
19. Gabrieli, T., Sharim, H., Michaeli, Y., Ebenstein, Y. Cas9-Assisted Targeting of CHromosome segments (CATCH) for targeted nanopore sequencing and optical genome mapping. *bioRxiv.* (2017).
20. Gabrieli, T. et al. Selective nanopore sequencing of human BRCA1 by Cas9-assisted targeting of chromosome segments (CATCH). *Nucleic Acids Research.* (2018).