Long read de novo sequencing from blood and cells using SQK-ULK114 – expert telomere-to-telomere method

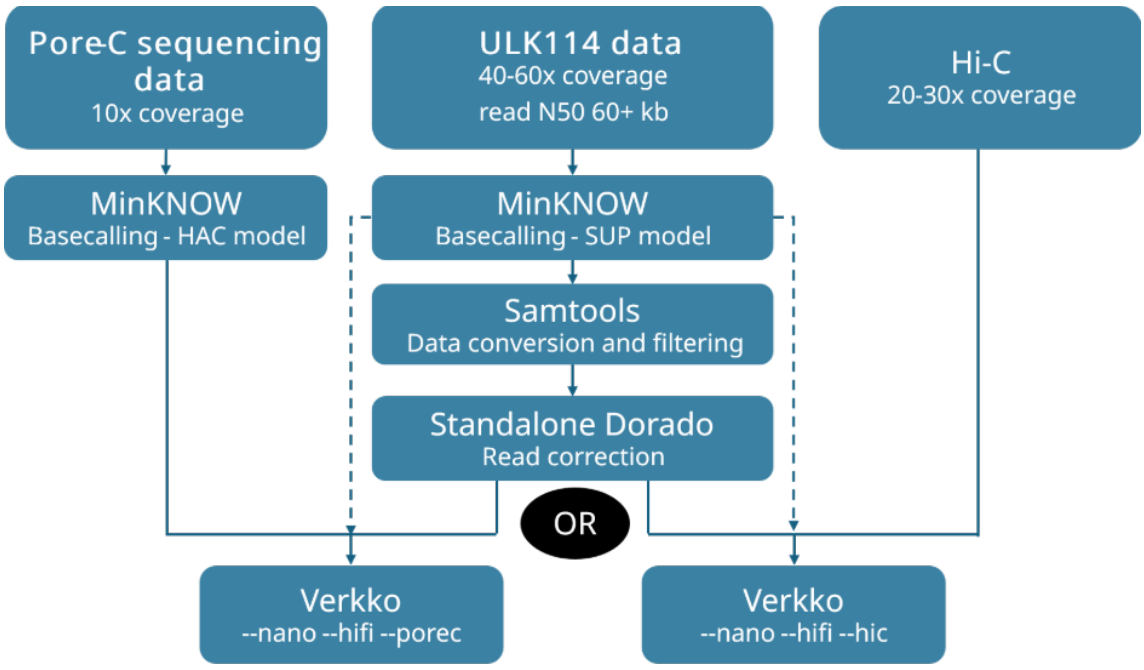
Downstream analysis recommendations

To carry out the expert long read assembly downstream analysis, we recommend previous bioinformatics experience.

The commands below combine the ultra-long DNA and Pore-C datasets. It is possible to use HiC data instead of Pore-C data, and the commands for assembling ultra-long DNA with Hi-C data are also provided.

Our recommended assembly method when using ultra-long data with Pore-C/Hi-C data is Verkko. It is possible to use parental sequencing reads in Verkko. However, we have found that this functionality is tricky to implement (requires multiple pre-processing steps) and gives finicky performance. Therefore, we are restricting our recommendations to the Pore-C/Hi-C mode. If you need to use trio data, we recommend you follow the assembly instructions in our protocol: "scalable (near) telomere-to-telomere method" instructions, which do work fine with Ultra-long + parental sequencing data (just slightly less well than Verkko with ultra-long + Pore-C/Hi-C data).

Please note: The information outlined in this document are recommendations for 3rd party analysis tools and methods to use in conjunction with the library preparation method. Oxford Nanopore Technologies do maintain or support these analysis tools. For further information and support please refer to the tools repositories.



Note: The analysis workflow described here are recommended guidelines using existing 3rd party tools.

Note: All basecalling must have been completed before starting downstream analysis.

Oxford Nanopore Technologies, the Wheel icon, AmPORE-TB, ElysION, EPI2ME, Flongle, GridION, MinION, MinKNOW, PromethION, and TraxION are registered trademarks or the subject of trademark applications of Oxford Nanopore Technologies plc in various countries. Information contained herein may be protected by patents or patents pending of Oxford Nanopore Technologies plc. All other brands and names contained are the property of their respective owners. © 2025 Oxford Nanopore Technologies plc. All rights reserved.

Oxford Nanopore Technologies products are not intended for use for health assessment or to diagnose, treat, mitigate, cure, or prevent any disease or condition.

Data requirements

40-60X coverage ultra-long read data, read N50 60+ kb (optimal performance seen with 50x-60x, 70-100 kb N50 reads).

- SUP basecalling >= SUP@v5.0.0 recommended. HAC basecalling with >=HAC@v5.2.0 also reasonably performant, but some decrease in contiguity observed.

10x Pore-C sequencing data OR 20-30x Hi-C sequencing data.

- HAC basecalling (>=HAC@v5.2.0) sufficient for Pore-C.

Setting up the environment

The Nanopore-only long read *de novo* expert T2T assembly workflow is summarised in the figure above and requires the following tools and their dependencies to be installed:

- Samtools (≥ v1.2) for format conversion
- Dorado (≥ v0.95) for read correction
- Verkko (≥ v2.3.2) for assembly

Fastq conversion and read filtering)

Dorado correct requires long read data in fastq format, and for convenience we'll convert everything to fastq files for further analysis. Please note that fasta format is not acceptable for Dorado correct input as quality scores are used to distinguish informative polymorphisms from sequencing errors during read self-correction. This is also an opportunity to do some quality filtering. These filtering steps assume you have bam files and have not lost the "qs" bam tag output by the Dorado basecaller. If these assumptions are not true, "seqkit seq -q 10 -m 1000" and "seqkit seq -q 9 -m 100" can be used to filter ultra-long and Pore-C fastq files, respectively – this is significantly slower, however.

The command to filter ULK reads by quality and read length, and to convert to fastq is:

```
samtools view -u -e '[qs]>=10&&length(seq)>=1000' <input_ulk.bam> | samtools fastq > ultralongreads.fastq
```

Note: If you have Pore-C data and have not already done so, the command to filter Pore-C reads by quality is:

```
samtools view -u -e '[qs]>=10&&length(seq)>=1000' <input_ulk.bam> | samtools fastq > ultralongreads.fastq
```


Oxford Nanopore Technologies, the Wheel icon, AmPORE-TB, ElysION, EPI2ME, Flongle, GridION, MinION, MinKNOW, PromethION, and TraxION are registered trademarks or the subject of trademark applications of Oxford Nanopore Technologies plc in various countries. Information contained herein may be protected by patents or patents pending of Oxford Nanopore Technologies plc. All other brands and names contained are the property of their respective owners. © 2025 Oxford Nanopore Technologies plc. All rights reserved.

Oxford Nanopore Technologies products are not intended for use for health assessment or to diagnose, treat, mitigate, cure, or prevent any disease or condition.

Ultra long read correction

The built-in pre-assembly read correction for Verkko is not appropriate for Oxford Nanopore reads, thus you need to run the Dorado correct tool (a reimplementation of the HERRO pre-assembly correction pipeline) first. You can do this in one step if you have a machine with many cores (>24 for efficient compute) and a high-performance GPU (e.g. on a PromethION analysis tower), or you can run the CPU-intensive step separately (useful if you are running on the cloud and don't want to pay for the GPU while your cheaper CPU-intensive steps are running). By default, Dorado will use all threads available for the CPU intensive steps – you can restrict this using the "-t" flag (e.g. adding "-t 96" to use 96 threads).

Option 1: Combined CPU + GPU

This assumes that you are running on a device with a high-performance GPU in the "0" slot for CUDA accessible devices – otherwise please update the "-x cuda:0" portion of the command as appropriate.

```
dorado correct -x cuda:0 ultralongreads.fastq > ultralongreads.doradocorrect.fasta
```

Option 2: CPU then GPU

CPU step:

```
dorado correct --to-paf ultralongreads.fastq > ultralongreads.dcOvl.paf
```

GPU step: This assumes that you are running on a device with a high-performance GPU in the "0" slot for CUDA accessible devices – otherwise please update the "-x cuda:0" portion of the command as appropriate.

```
dorado correct -x cuda:0 --from-paf ultralongreads.dcOvl.paf ultralongreads.fastq > ultralongreads.doradocorrect.fasta
```

Verkko assembly

Verkko handles everything from here. Note that the uncorrected reads are supplied as "--nano", while the corrected reads are supplied as "--hifi" – this is for historical reasons as Verkko was originally designed to use HiFi reads for building the graph and Oxford Nanopore reads for untangling (now we know it works best to use corrected Nanopore reads and uncorrected Nanopore reads respectively). We will provide the "--no-correction" flag to tell Verkko that the reads supplied as "--hifi" have already been corrected and Verkko should not correct them itself.

Verkko is a bash script which runs a snakemake pipeline, and if it is run on a cluster with job management (e.g. SGE, LSF, Slurm), you can configure it to run across multiple cluster nodes for more efficient execution. Cluster configuration is beyond the scope of this guide, however, so we are only providing instructions to run locally on a single machine – please see Verkko documentation for more advanced run instructions.

Oxford Nanopore Technologies, the Wheel icon, AmPORE-TB, ElysION, EPI2ME, Flongle, GridION, MinION, MinKNOW, PromethION, and TraxION are registered trademarks or the subject of trademark applications of Oxford Nanopore Technologies plc in various countries. Information contained herein may be protected by patents or patents pending of Oxford Nanopore Technologies plc. All other brands and names contained are the property of their respective owners. © 2025 Oxford Nanopore Technologies plc. All rights reserved.

Oxford Nanopore Technologies products are not intended for use for health assessment or to diagnose, treat, mitigate, cure, or prevent any disease or condition.

When running locally, Verkko will automatically use all threads available, but will automatically throttle RAM usage unless overridden with the “--local-memory” option – we recommend setting this to the max you have available.

Option 1: Pore-C

```
 verkko --nano ultralongreads.fastq --hifi ultralongreads.doradocorrect.fasta --porec porec.fastq --no-correction --local-memory <max_mem_in_GB> -d verkko_output
```

Option 2: Hi-C

Assuming here that Hi-C reads are in FASTQ format with left and right reads in separate files hic_r1.fastq and hic_r2.fastq (standard format to receive reads from sequencing providers).

```
 verkko --nano ultralongreads.fastq --hifi ultralongreads.doradocorrect.fasta --hic1 hic_r1.fastq --hic2 hic_r2.fastq --no-correction --local-memory <max_mem_in_GB> -d verkko_output
```

Outputs

Verkko outputs assembly files to the “verkko_output” directory (or whatever you supplied with “-d”).

Outputs include a pair of phased haplotype assemblies in FASTA format, assembly.haplotype1.fasta and assembly.haplotype2.fasta, as well as a FASTA file with both concatenated together (assembly.fasta).

Please note, it does not provide a collapsed assembly – if you need such an assembly, we recommend identifying the most contiguous of the two haplotypes and using that. If your sample is heterogametic (e.g. a male mammal i.e. XY), you may need to identify the sex chromosome in the non-selected assembly and add it in for a complete reference.

Change log

Date	Version	Changes
July 2026	V1	Document release

Oxford Nanopore Technologies, the Wheel icon, AmPORE-TB, ElysION, EPI2ME, Flongle, GridION, MinION, MinKNOW, PromethION, and TraxION are registered trademarks or the subject of trademark applications of Oxford Nanopore Technologies plc in various countries. Information contained herein may be protected by patents or patents pending of Oxford Nanopore Technologies plc. All other brands and names contained are the property of their respective owners. © 2025 Oxford Nanopore Technologies plc. All rights reserved.

Oxford Nanopore Technologies products are not intended for use for health assessment or to diagnose, treat, mitigate, cure, or prevent any disease or condition.