

Calling and phasing single nucleotide variants in the human genome with long nanopore reads

With an estimated 8% of the human genome still unsequenced¹, our understanding of genetic variation has been limited. Compared to short-read technologies, nanopore sequencing lacks GC bias and does not require PCR, enabling wider access to the genome for variant calling.

Assigning variants to the maternal or paternal chromosome (phasing) is important for understanding their inheritance and functional impact. Long and ultra-long nanopore sequencing reads greatly enhance phasing, especially of heterozygous variants, as an individual read is more likely to contain multiple single nucleotide variants (SNVs).

Here we present a simple workflow for calling and phasing SNVs in the human genome.

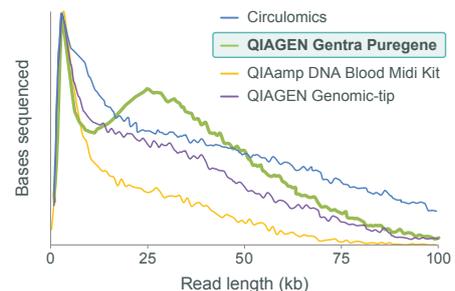


EXTRACTION: obtaining high molecular weight DNA

Long sequencing reads are required for optimal phasing. Obtaining high molecular weight (HMW) DNA from your sample is therefore crucial. Selecting the most suitable HMW DNA extraction method depends on your sample type. For DNA extraction from whole blood, we recommend the **QIAGEN Gentra Puregene Blood Kit**, which we have found maximises the production of long sequencing reads.



Find more extraction protocol recommendations for your sample type, plus guidance on DNA storage and contaminants: community.nanoporetech.com/extraction_methods



LIBRARY PREPARATION: selecting a kit

We have found that read length has a strong effect on phasing performance. To prepare your extracted gDNA, we advise light shearing and size selecting for >10 kb fragments, for a read N50 of >32 kb. Alternatively, no shearing, with optional size-selection, for a read N50 of ≥25 kb, produces similar results (see Table). Internal testing has yielded good results using the **Circulomics Short Read Eliminator Kit** for size selection (to >10 kb), and the **Diagenode Megaruptor 3** for shearing.

To prepare gDNA for sequencing, we recommend the **Ligation Sequencing Kit (SQK-LSK110)**, providing the greatest throughput.

Find out more about size selection methods for long-read sequencing: community.nanoporetech.com/extraction_methods

Read N50	Shearing method	Phase block N50
11 kb	g-TUBE	80–200 kb
17 kb	SRE + MR-3*	150–300 kb
25 kb	Unfragmented	1–2.5 Mb
26 kb	SRE + MR-3	500–650 kb
32 kb	SRE + MR-3	1–2.5 Mb
36 kb	SRE + MR-3	1.5–3 Mb

*Circulomics Short Read Eliminator Kit + Diagenode Megaruptor 3

Find out more about library prep, including rapid and multiplexing options: nanoporetech.com/products/kits

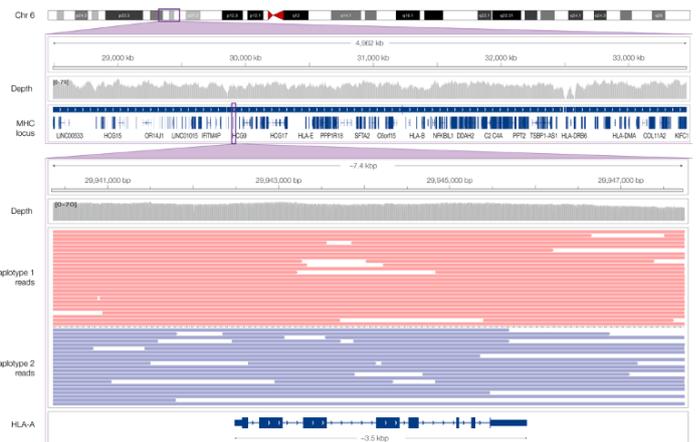
SEQUENCING: generating high yields of long reads with PromethION

For SNV calling, we recommend sequencing a human genome to 40x–60x read depth; this can be achieved by sequencing a single genomic sample on one or two PromethION™ Flow Cells for 72 hours.

Whilst longer reads significantly improve phasing, read depth is less important, with good metrics achieved at 45x and diminishing returns for higher depth.

We recommend basecalling in high accuracy mode with Guppy version 3.6.1 or higher. Throughput can be maximised by performing a nuclease flush and loading fresh library every 24 hours.

Find out more about PromethION:
nanoporetech.com/products/promethion



Find out about nanopore sequencing service providers: nanoporetech.com/services/providers

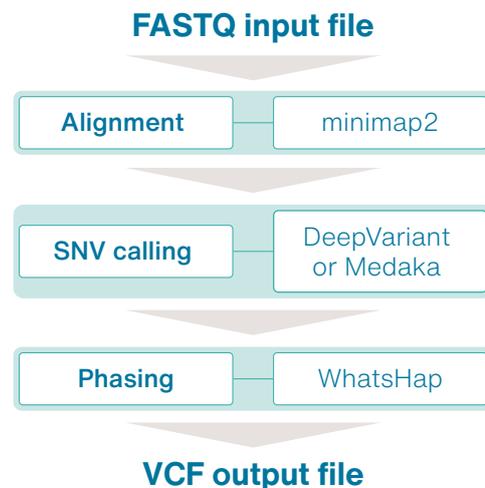
ANALYSIS: selecting tools for SNV calling & phasing

To call SNVs in the human genome we recommend **Medaka**², developed and supported by Oxford Nanopore, and with easy installation. Analysis runtime for this workflow is <48 hours using a single PromethION or GridION™ GPU and 8 CPU cores.

Alternatively, we have found that the third-party **PEPPER/DeepVariant**³ workflow provides the best SNV calling metrics for datasets with read depths <60x, basecalled with Guppy 3.6.1 and above.

We advise the third-party tool **WhatsHap**⁴ for phasing, alongside PEPPER/DeepVariant. WhatsHap is integrated into Medaka, so this additional step is unnecessary when using Medaka for SNV calling.

Find out more about data analysis solutions:
nanoporetech.com/analyse



Find out more at: nanoporetech.com/snvs-phasing

References:

1. Miga, K.H., Eisenhart, C., and Kent, W.J. Utilizing mapping targets of sequences underrepresented in the reference assembly to reduce false positive alignments. *Nucleic Acids Res.* 43(20):e133-e133 (2015).
2. © 2020 Oxford Nanopore Technologies Ltd.
3. Shafin, K. PEPPER. Software. Available at: <https://github.com/kishwarshafin/pepper/> [Accessed: 03 June 2021].
4. Martin, M. WhatsHap: fast and accurate read-based phasing. *bioRxiv*. DOI: <https://doi.org/10.1101/085050> (2020).

Oxford Nanopore Technologies and the Wheel icon, GridION, and PromethION are registered trademarks of Oxford Nanopore Technologies Limited in various countries. All other brands and names contained are the property of their respective owners. © 2021 Oxford Nanopore Technologies Limited. 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. WF_1085(EN)_V2_03Jun2021