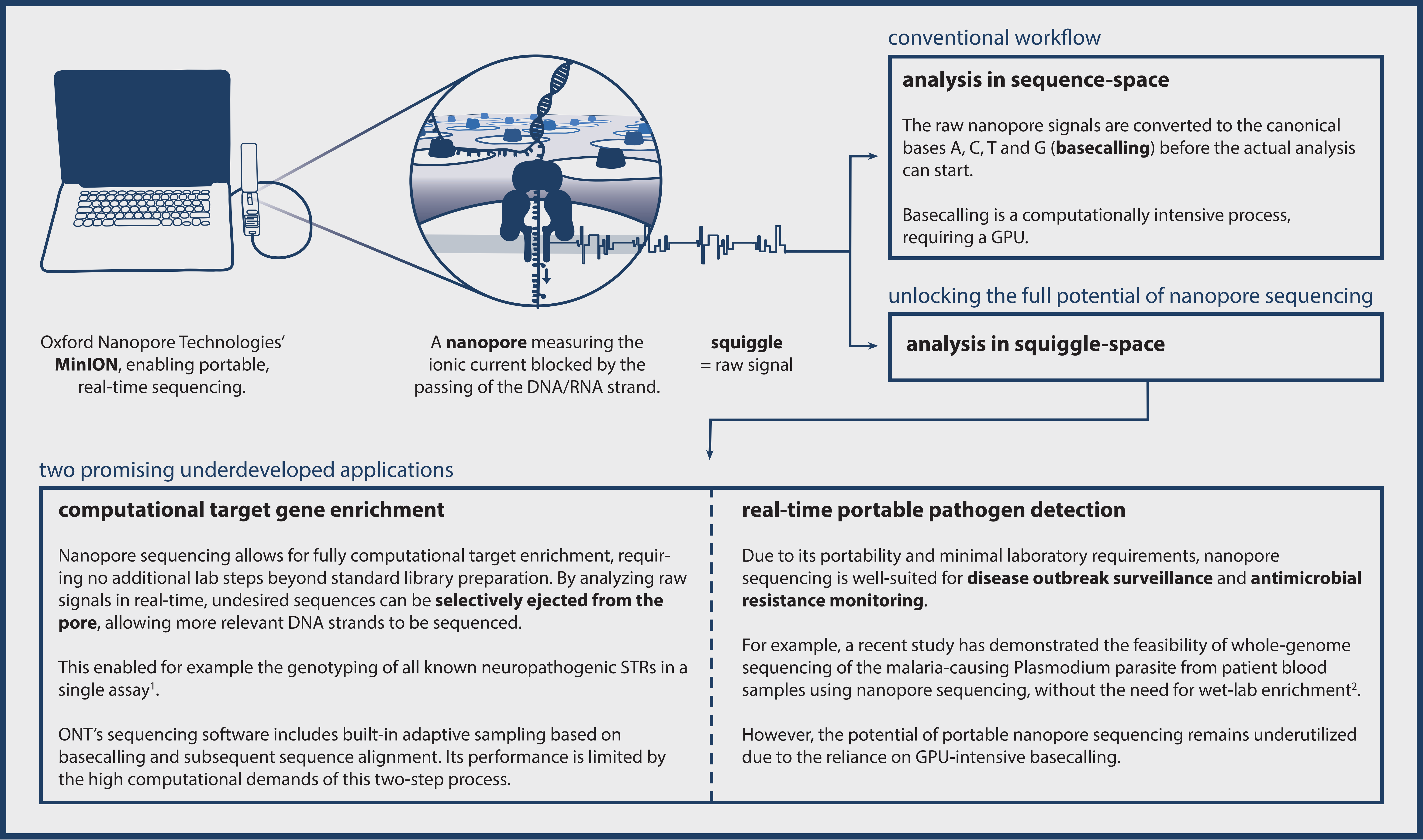


# ADVANCING NANOPORE SQUIGGLE INTERPRETATION THROUGH SQUIDBASE DATA CENTRALISATION AND ALGORITHMIC BENCHMARKING

Laura Raes, Wim L. Cuypers, Halil Ceylan, Eline Turcksin, Julia M. Gauglitz, Wout Bittremieux, Kris Laukens



## potential of nanopore squiggle analysis



## our contributions

**issue**

**lack of squiggle data**

Traditional repositories (e.g. Sequence Read Archive) are **not well-suited** for raw nanopore data:

- raw signal data is practically unfindable
- no nanopore-specific filtering options (e.g. pore model, chemistry)
- upload and download of squiggle data is cumbersome and inefficient

The absence of a centralized storage system for raw nanopore data **hinders**:

- researchers accessing the epigenetic data contained in raw signal data
- the development of new squiggle-space algorithms, particularly those based on machine learning
- transparency and reproducibility in data-driven scientific research

**contribution**

**SquidBase**

We developed SquidBase, a **curated database for data of microbial and viral origin**.

- We included publicly available raw nanopore data generated with the legacy R9 nanopore, to support the development and benchmarking of squiggle-space algorithms.
- We generated and uploaded R10.4.1 data for 60 viral strains and the malaria-causing parasite Plasmodium. This is especially important because of the lack of publicly available data generated with the new R10 nanopore.
- Since SquidBase does not support human genomic data, a preprocessing pipeline for filtering out human reads is available.

SquidBase is available at <https://squidbase.org>.

**issue**

**lack of squiggle-space algorithms**

Only a limited number of squiggle-space algorithms are available. The algorithms that do exist, remain largely **untested outside of their original studies**.

**contribution**

**benchmarking of squiggle-space classification algorithms**

We initiated a benchmarking study to assess squiggle-space classification algorithms. Further development of these algorithms would benefit computational target gene enrichment as well as real-time portable pathogen detection. Our study is still ongoing, but preliminary results reveal **both areas for improvement and significant potential**.

The figure below compares the performance of three squiggle-space classification algorithms (Sigmoni<sup>3</sup>, DeepSelectNet<sup>4</sup>, and RawAlign<sup>5</sup>) on CPU for the binary classification of SARS-CoV-2 vs. Zymo R9 reads. Their performance is evaluated against ONT's current approach: basecalling and mapping with Dorado on GPU.



QR code links to a web page containing the poster, references, abstract and additional information.

Interested in contributing to SquidBase or fellow nanopore enthusiast? We would love to get in touch!