

# Beyond Homopolymer Errors: a Systematic Investigation of Nanopore-based DNA Sequencing Characteristics Using HLA-DQA2

Péter Sárközy<sup>1\*</sup>, Viktor Molnár<sup>2</sup>, Dóra Fogl<sup>1</sup>, Csaba Szalai<sup>2</sup>, Péter Antal<sup>1</sup>

Received 13 September, accepted after revision 01 May 2017

## Abstract

Electronic, nanopore based single molecule real-time DNA sequencing technology offers very long, albeit lower accuracy reads in sharp contrast to existing next-generation sequencing methods, which offer short, high-accuracy reads in abundance. We provide a systematic review of the error characteristics of this new sequencing platform, and demonstrate the most challenging aspects in the field of whole gene sequencing through the human HLA-DQA2 gene using long-range PCR products on multiplexed samples. We consider the limitations of these errors for the applications of this technology, and also indicate prospective improvements and expected thresholds with respect to these errors.

## Keywords

nanopore, DNA sequencing, next-generation sequencing, error characteristics

## 1 Introduction

After decades of DNA sequencing approaches based on sequencing-by-synthesis [1], Oxford Nanopore Technologies (ONT) introduced a commercially available electronic DNA sequencing technology in 2014: the MinION portable single-molecule real-time DNA sequencing platform. Current next-generation DNA sequencing methods –except for Pacific Biosciences' Helicos technology– use clonal clusters of identical DNA strands to perform sequencing by synthesis. In the case of Life Technology's IonTorrent platform, template strands are amplified on immobilized beads by emulsion polymerase chain reaction (PCR). The sequencing is performed by flowing alternating deoxynucleotides over a plate with a single bead in each well, and recording the pH change induced by the incorporation of nucleotides. In the current next-generation sequencing market leader Illumina's platform, the target DNA fragments of interest are amplified in situ using bridge PCR on glass plate with immobilized probes, while the sequencing is performed by fluorescent labelled reversible terminator sequencing.

**Table 1** The main features of the ONT platform, compared to a widespread sequencing-by-synthesis NGS platform

Feature	Illumina MiSeq	ONT MinION
Read length	2 x 300bp	>100kB
Mean PHRED	75% > Q30	~Q18 with R9
Throughput/run	Up to 15 Gb	Up to 1Gb
Initial capital cost	~\$100,000	~\$1,000
Approximate cost/run	~\$1,000	~\$1,000

Since its release, the MinION has seen sequential improvements, and post-processing pipelines have been created by the scientific community to support the analysis of the data coming from this DNA sequencer (see Table 1 for comparison of its characteristics against a sequencing-by-synthesis NGS platform).

We describe the results obtained using a MinION Mk I sequencer, using R7.3 flow cells and the SQK-006 Genomic

<sup>1</sup> Department of Measurement and Information Systems, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics H-1521 Budapest, P.O.B. 91, Hungary

<sup>2</sup> Department of Genetics, Cell- and Immunobiology, Semmelweis University, H-1085 Budapest, Üllői út 26., Hungary

\* Corresponding author, e-mail: [psarkozy@mit.bme.hu](mailto:psarkozy@mit.bme.hu)

DNA Sequencing Kit, all provided in the framework of the MinION Access Programme [2].

The ONT sequencing platform works by immobilizing protein pores in a membrane applied over a semiconductor substrate [3, 4]. An adenosine triphosphate (ATP)-driven ratcheting motor protein and a tether adapter is ligated onto the end of the DNA molecules of interest. An optional hairpin adapter can be ligated onto the free end of a double stranded DNA molecule, which allows both the template and the complement strands to be sequenced inside the same pore. The tether bound to a strand has affinity for the membrane, thus the concentration of target molecules is the largest around the sequencing pores.

A charge of  $\sim 100\text{mV}$  is applied across the membrane, which facilitates the initial entry of the tether protein into the pore. The potential difference drives the DNA strand through the pore while providing a bias voltage for sensing changes in the current flow. Without a ratcheting motor protein, the raw traversal rate of DNA through the nanopore is approximately  $1\text{MB/sec}$ . The motor protein consumes 1 ATP for each base that is ratcheted through, thus the concentration of ATP in the fuel mix can be adjusted for optimal traversal speed.

Sensing of the individual bases is performed by sampling the change in the current flow across the pore. The large size of the pore relative to the individual nucleotides results in a compound signal from 6 bases (6-mer) which occupy the centre of the pore. Sampling of the current across each pore is performed at 3 kHz, although the platform allows for higher sampling rates up to 10 kHz. Each flow cell contains 2048 pores, which are multiplexed in groups of 4, so 512 pores can sequence DNA simultaneously. Pores are selected by multiplexers to provide the highest number of active pores at the start of a run. Switching to the next group of pores is performed automatically every 12 hours.

The MinKNOW control software compresses the raw sampled currents into 'event space'. An event is a length of time where the current flowing across a pore is constant, and is characterized by the length, the mean and the standard deviation of the current signal. The capturing of raw current levels is possible through custom software configuration, but results in prohibitively large file sizes.

Transformation of the event space to individual bases is performed by using the characteristic current levels corresponding to each 6-mer (events) to decode a Hidden Markov Model (HMM), where the hidden states are the sequences of nucleotides forming the 6-mers. The model used by the Metrichor cloud-based basecalling algorithm assumes that at most two translocations occurred per event. The model allows for the possibility of multiple events being produced by a single step in the DNA strand.

Current next-generation sequencing (NGS) platforms provide read lengths on the order of 400 base pairs in length,

primarily stemming from an inherent technological limitation: the synthesis of individual clonal templates inside a single clonal cluster is increasingly asynchronous due to the stochastic nature of nucleotide incorporation into individual strands during synthesis, and thus the quality of the sequencing drops at the end of the reads. In contrast, the MinION sequencer allows reads in excess of  $100\text{kB}$ , and is limited only by library preparation protocols (especially bead enrichment) which tends to shear DNA fragments above  $10\text{kb}$  length [5].

In this paper we provide a systematic investigation of nanopore-based DNA sequencing errors using the MinION with R7.3 flow cells and using the HLA-DQA2 gene as a target. First, we describe the measurement protocol and biomed background. The Results section presents the main outcomes of our experiment. The Discussion contains the evaluation of our experimental results in light of the performance characteristics of ONT in literature and its concise summary. In Conclusion, we overview the prospective applications of ONT, specifically highlighting the limitations of the explored error characteristics.

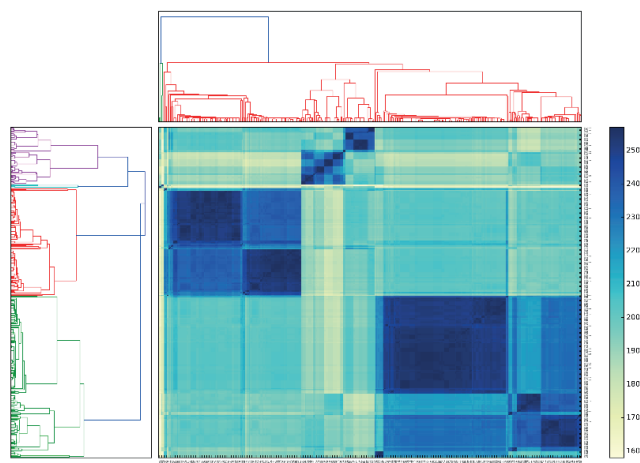
## 2 Materials and methods

The target for this sequencing study was the human Major Histocompatibility Class II (MHC) gene HLA-DQA2. This genomic region was chosen because it contains a high number and diverse classes of mutations, ranging from single nucleotide polymorphisms (SNP's) to multi-base insertions and deletions (indels) inside homopolymer tracts. Additionally, the phasing of MHC genes is a clinically relevant task for bone marrow transplant donor compatibility.

Human DNA samples were used from the DesensIT project. This study investigated allergen-specific immunotherapy using the Affymetrix Genome-Wide Human SNP Assay Kit 6.0. The 24 samples were chosen to cover the full range of haplotypes encountered in the genome-wide (GW) measurement of the DesensIT study. The a priori availability of GW genetic data was also utilized to ensure in silico that no SNP's were present underneath the primer sequences.

Based on the high linkage disequilibrium present in HLA-DQA2, as identified by Haploview [6], we used Impute2 [7] to infer all mutations in the target region using the CEU reference panel. This resulted in a total of 257 mutations, most of which had low ( $<1\%$ ) minor allele frequencies. The putative haplotypes of each sample were reconstructed in silico (Fig. 1) with the help of PHASE [8].

Primers were designed using Primer3 [9] to ensure identical annealing temperatures for both forward and reverse primers, with the incorporation of the ONT supplied barcode adapter sequences (Fig. 2).



**Fig. 1** Reconstructed haplotypes were grouped for phased similarity, and samples were selected from major groups to provide even sample coverage of all putative haplotypes. Darker colors indicate high similarity.

ONT Adapter – Forward primer:

5' TTTCTGTTGGTGCTGATATTGC-AACTGGCAAACAATATGTCACCA 3'

ONT Adapter – Reverse primer:

5' ACTTGCCTGTCGCTCTATCTTC-GATACACTGTACTTGGTTGCTCC 3'

**Fig. 2** The forward and reverse primers used for the whole-gene amplification of HLA-DQA2 with the ONT specific adapters

DNA was extracted from frozen whole blood with the Purelink Genomic DNA Mini Kit, and the recommended 100ng of input DNA per initial PCR reaction was diluted. PCR was performed on the samples for 35 cycles, with the annealing temperature of 58°C, extension temperature of 68°C with 6 minute extension times, based on manufacturer protocols for PrimeSTAR GXL Polymerase by Takara. The resulting samples were cleaned from unincorporated primers and PCR contaminants by gel electrophoresis. A control experiment was performed with only the HLA-DQA2 specific primers without any barcode adapter. PCR efficiency was significantly higher for primers without the barcode adapter sequences.

Barcoding allows the multiplexing of up to 96 samples on a single flow cell via PCR ligation of custom barcode oligonucleotide sequences. A second barcoding PCR was performed to allow the multiplexing of 12 samples per flow cell. 15 PCR cycles with annealing at 62°C and 6 minute extension times were performed, and the results cleaned up with gel electrophoresis. Amplicons were pooled in for each flow cell after clean-up, and the Lambda phage control DNA strands were added to the pools.

Library preparation began with 2 µg of input DNA, starting with end repair (NEBNext End Repair Module), dA-tailing (NEBNext dA-Tailing module), as per manufacturer protocols, with 1.6x Ampure XP bead clean-up after each step. The ligation of hairpin adapters and tethers to either ends of the fragments

was performed according to ONT protocols with NEB Blunt/TA Ligase, using the supplied reagents from the SQK-006 Genomic DNA Sequencing Kit. After a final clean-up with MyOne C1 beads, approximately 180ng of DNA remained, and the pre-sequencing mix was prepared using the supplied Fuel Mix and Running Buffers. MinION version R7.3 flow cells were used for the experiment. The traversal speed of the DNA strands through the pore was approximately 70 bases/sec.

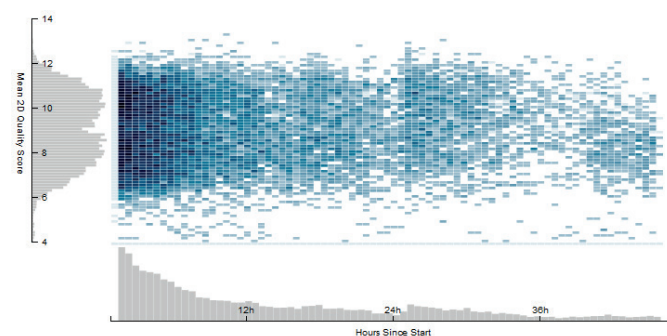
### 3 Results

Platform QC of the flow cells indicated 362 and 306 single pores detected, respectively. The first group of pores contained 201 and 210 active single pores.

Hairpin adapters allow each double stranded DNA molecule to be sequenced on the template and complement strands. Hairpin adapter detection allows the alignment of the template and complement strands. Single stranded molecules, or double stranded molecules without hairpin adapters will only be sequenced along one strand. The Metrichor basecalling application reports multiple read statuses while performing basecalling:

1. Calibration strands in the form of Lambda phage control DNA. This is used for HMM parameter estimation for basecalling.
2. Single stranded (1D) reads without complementary strand data. These reads have lower accuracy.
3. 1D Basecalling failure is often reported for fragments shorter than 100bp in length.
4. 2D Basecalling failure is reported if the hairpin is found but the template and complement sequences differ too much in length or sequence.
5. If sample multiplexing is used, then barcode classification errors are reported.
6. The workflow is considered successful if 2D basecalling and barcode classification are successful.

Generally, only 2D reads are considered as passing in the initial quality filtering, as they achieve lower per-base error rates (Fig. 3).



**Fig. 3** The 2D mean quality score vs. time since the start of sequencing. The bottom histogram shows the 2D yield per hour of sequencing; note that the library top-up at 24 hours yielded additional high-quality reads



Some customized applications such as direct RNA sequencing and rapid library preparation do not perform hairpin adapter ligation, thus will only result in 1D reads.

The first flow cell resulted in 40MB of 2D reads. During the run there was a long tail of additional sequencing throughput, though the ratio of 2D reads dropped off after 30 hours. Library top-ups were performed at 15, 30 and 40 hours.

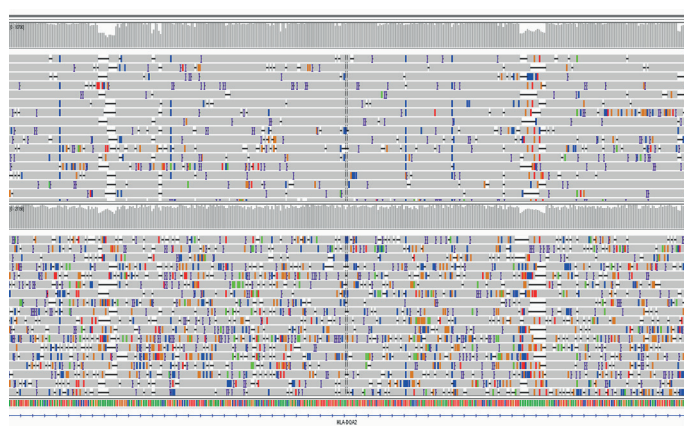
The second cell yielded only 12MB of 2D reads. The number of 2D reads dropped off sharply after 10 hours. Library top-ups were added at 11, 19 and 30 hours, but these produced no additional high-quality 2D sequences. Flow cell shelf life is limited to 8 weeks, and the measurement was performed near the end of this period. This partly explains the relatively low number of pores and low yields compared to the expected results.

The barcoding of 24 samples performed as expected. We observed excellent separation between barcodes as evidenced by the singular sequences which had unused barcodes assigned to them.

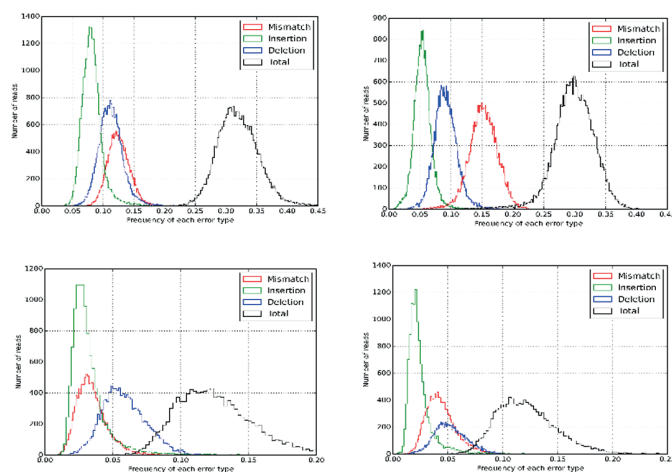
Of the 12% of 2D reads where barcodes classification failed, the majority did have the barcode sequence present on the 2D basecalled read, but there were too many errors in the barcodes or the barcodes were truncated. Thus incomplete barcode ligation in PCR can be ruled out as a major source of error.

### 3.1 Alignment

Alignment of the individual 2D reads to the reference sequence (Fig. 4) was performed with Graphmap [10] and the Burrows-Wheeler Aligner with the option `-x ont2d` (BWA-MEM) [11]. The former was found to align reads faster and with fewer mismatches (Fig. 5). The scoring functions used when calculating alignments use different values for matches, mismatches, gap opening and gap extending for each aligner. The default values were used for both sequence alignment algorithms and their difference explains the varying ratio of mismatches to indels when comparing the resulting alignments. Alignments were examined with Integrative Genomics Viewer (IGV) [12].

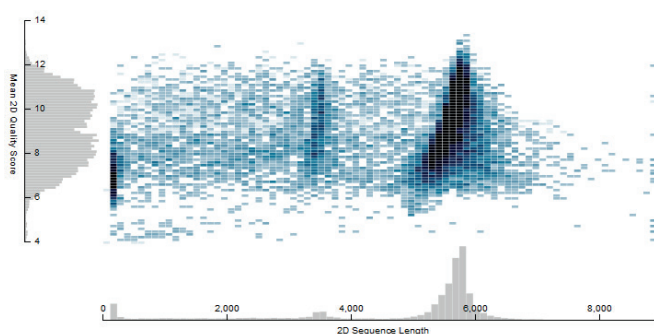


**Fig. 4** Visual representation of the alignment of 2D (top) and 1D (bottom) reads to the reference sequence of an area with a 17A homopolymer stretch (green on bottom track). Each horizontal track represents a single read, showing deviations from the reference.



**Fig. 5** The distribution of mismatch, insertion and deletion errors in 1D reads when mapped with Graphmap (top left) and BWA (top right), and 2D reads mapped with Graphmap (bottom left) and BWA (bottom right).

The coverage of the targeted gene was complete and uniform, indicating that the long-range PCR performed as expected. The read length histogram shows that most reads are of the expected length of the PCR product. Marginally shorter reads appear to have lower average quality scores (Fig. 6), but this relationship is confounded by compression of raw current data into event space: e.g. if more events are missed during the event detection phase, then these reads will appear shorter during the basecalling as they will contain more deletions from the reference. The standard coverage requirement in NGS for clinical diagnostics is greater than 100x. The achievable coverage uniformity is dependent on the read lengths produced by any sequencing platform.



**Fig. 6** Mean 2D quality score vs. sequence length.

Long homopolymer stretches present a challenge for all NGS platforms, and the MinION is no exception. The initial event detection step can miss steps inside homopolymer regions, as they produce no marked change in the ionic current. The dwell times of each k-mer inside the pores –despite the large spread of the dwell time of each k-mer– provides a feasible approximation of homopolymer lengths at high coverage depths. Homopolymers longer than 3 bases already show marked false deletions in alignments and could easily be interpreted as deletions. Homopolymer stretches longer than 6 bases

are fully collapsed, as the basecalling HMM model does not step between identical homopolymer states.

### 3.2 Variant calling

The relatively high error rate does not present an excessive burden on calling individual single nucleotide polymorphisms, as the even distribution of sequencing errors can be leveraged with higher coverage rates to average out base substitution errors. However, low frequency variant calling, especially the identification of somatic mutations presents a challenge to this platform. The identification of insertions and deletions inside homopolymer stretches is further confounded by alignment bias produced by the local alignment algorithm.

### 3.3 Phasing

A grand promise of ONT technology is the direct measurement and resolution of the haplotypes of variants across the entire gene, due to the full length coverage achieved by long-range PCR. This, however was made extremely difficult due to PCR chimera formation, most likely during the second bar-coding PCR, reported already in recent papers [13, 14]. Partly related to the unique ability of using long strands, PCR chimeras are formed when the complementary strand synthesis is incomplete or when the PCR enzyme complex switches template strands during the amplification. This strand-switching effect was observed even with heterozygous SNPs closer than 100bp, and the effect extended to complete linkage disequilibrium at SNPs further than 1000bp apart. Nevertheless, the remaining linkage disequilibrium was sufficient to assign high probability haplotypes to each sample, and these haplotypes matched the in silico haplotypes predicted from the initial GWAS.

### 3.4 Quality Scores

We investigated the reported and empirical PHRED quality scores from the 1D and 2D datasets by comparing the reads to the reference sequence in positions that were not covered by known SNP's, to show the reliability of the reported quality scores. The quality scores for C and G bases in 1D reads are accurate, while A and T scores are reported lower than the actual error rate (Fig. 7). The 2D reads show similar characteristics with higher mean quality scores (Fig. 8).

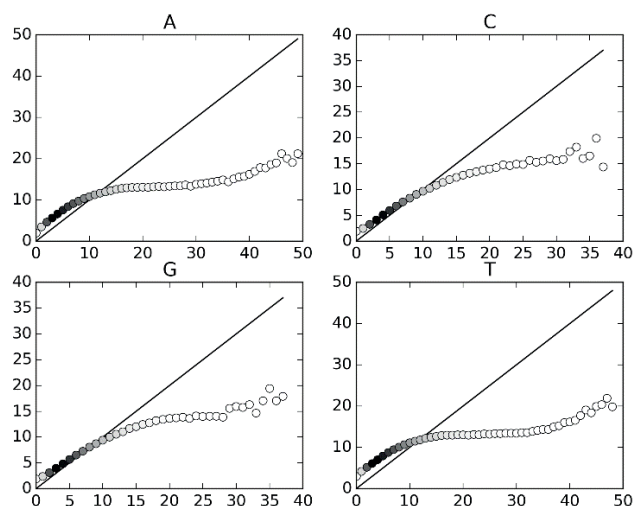
The raw data obtained from this sequencing run and any intermediate data is available from the authors upon request.

## 4 Discussion

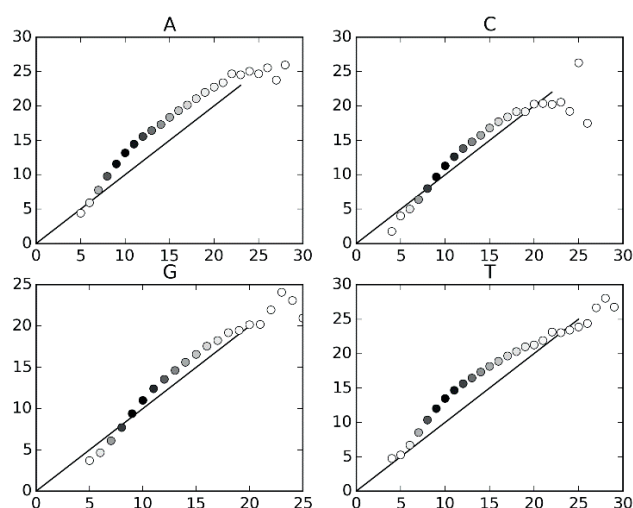
The MinION platform performed according to its specifications in the experiment, and produced usable, long reads spanning the entire target gene. The in silico predicted variants along with their predicted phase could be resolved from the data for every sample, albeit not without major difficulties stemming from the formation of highly chimeric reads during PCR. Homopolymer stretches showed characteristic high

deletion rates, but the aggregated dwell times in pores provided evidence to their exact lengths. We intend to fully characterize and utilize these features in a follow-up paper.

Read identity on the ONT platform has evolved considerably. The initial 70% identity rate experienced with the R6 chemistry [15], increases to the 98% identity rate with the newly released R9.4 chemistry. The identity rates may pose limitations on the identification of somatic mutations, and can confound the calling of rare variants. Irrespective of homopolymer errors and tandem repeats, the errors (especially mismatches) are evenly distributed.



**Fig. 7** The reported (x-axis) vs. the empirical (y-axis) PHRED quality scores for 1D reads, broken down for each nucleotide. The shading of the circles represents the frequency of each reported PHRED score, with darker circles indicating higher occurrence rates.



**Fig. 8** The reported (x-axis) vs. the empirical (y-axis) PHRED quality scores for 2D reads, broken down for each nucleotide. The shading of the circles represents the frequency of each reported PHRED score, with darker circles indicating higher occurrence rates.

## 5 Conclusion

The initial capital cost of NGS starts above 50k USD, while the MinION device promises capital costs under 1k USD. Reagent and consumables costs are similar for all sequencing technologies, but PCR-free library preparation and simplified library preparation protocols allow for laboratories with simplified infrastructure – up to the point of using mobile field laboratories.

NGS platforms currently have the highest throughput (bases of sequenced DNA/run) of any sequencing technology, Oxford Nanopore aims to rise to this level of throughput with the recently released Promethion sequencing grid. Individual MinION flow cells offer up to 1 Gb of throughput under ideal conditions.

The ability to perform basecalling in real time on the first couple of hundred bases traversing a pore opens up possibilities for in situ enrichment. The current across the membrane can be reversed based on the sequence observed at the start of a read, ejecting the molecule from the pore [16]. A rejected molecule does not have a tether or a motor protein, so it cannot enter the pore again.

Direct RNA sequencing is possible by ligating a poly-T sequencing adapter and annealing the tether molecule to RNA strands. This offers both qualitative and quantitative gene expression analysis and transcriptomic profiling.

Real-time species identification, along with PCR free quantitative and qualitative microbiomics are possible. This opens up an array of opportunities in the field of metagenomics and pathogen identification, particularly with the feasibility of portable sequencing [17, 18]. PCR-free library preparation techniques offer quantitative and qualitative cancer profiling, immunome profiling and mtDNA profiling.

Rapid response to antibiotic resistant infections, with a turn-around time under 6 hours is possible with simplified library preparation routines, as well as accurate pathogen identification down to the exact genes responsible for antibiotic resistance [19].

Long reads allow for the phased genotyping on polyploid samples, eliminating the need for performing paired-end sequencing with large insert sizes to resolve the gametic phase of distant genetic polymorphisms [13, 14].

The extremely long reads produced by the MinION sequencer make it a suitable tool for generating scaffolds for de-novo genome sequencing. In a hybrid approach, NGS reads supplemented by long ONT reads greatly reduce the number and increase the length of contigs a large genome can be assembled into. Shorter genomes or plasmids can be assembled exclusively from long reads [20]. Large scale chromosomal rearrangements and copy number variations can also be resolved, and are only limited by the attainable read lengths.

DNA methylation is involved in multiple facets of biology, such as gene regulation, cell differentiation and development, and disease. These epigenetic modifications – e.g. the change of cytosine to 5methylcytosine and 5hydroxymethylcytosine – can be detected with the ONT sequencer [21].

## Acknowledgements

The authors acknowledge that they received the MinION sequencer, the Genomic DNA Sequencing kit and the PCR Barcoding Kit free of charge from Oxford Nanopore Technologies as participants in the MinION Access Program. The authors have no other conflicts of interest. This research was supported by the OTKA-K-112915 Grant, the János Bolyai Research Scholarship of the Hungarian Academy of Sciences (P. Antal) and the NKFIH-K112872 Grant.

## References

- [1] Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P., Gu, Y. “A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers.” *BMC Genomics*. 13(1), pp. 341-353. 2012. <https://doi.org/10.1186/1471-2164-13-341>
- [2] MinION Access Programme. Available from: <https://nanoporetech.com/getting-started-with-minion> [Accessed: 12<sup>th</sup> September 2016]
- [3] Branton, D., Deamer, D. W., Marziali, A., Bayley, H., Benner, S. A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X., Jovanovich, S. B., Krstic, P. S., Lindsay, S., Ling, X. S., Mastrangelo, C. H., Meller, A., Oliver, J. S., Pershin, Y. V., Ramsey, J. M., Riehn, R., Soni, G. V., Tabard-Cossa, V., Wanunu, M., Wiggins, M., Schloss, J. A. “The potential and challenges of nanopore sequencing.” *Nature Biotechnology*. 26, pp. 1146-1153. 2008. <https://doi.org/10.1038/nbt.1495>
- [4] Deamer, D., Akeson, M., Brenton, D. “Three decades of nanopore sequencing.” *Nature Biotechnology*. 34, pp. 518–524. 2016. <https://doi.org/10.1038/nbt.3423>
- [5] Urban, J. M., Bliss, J., Lawrence, C. E., Gerbi, S. A. “Sequencing ultra-long DNA molecules with the Oxford Nanopore MinION.” *bioRxiv* 019281, 2015. <https://doi.org/10.1101/019281>
- [6] Barrett, J. C., Fry, B., Maller, J., Daly, M. J. “Haploview: analysis and visualization of LD and haplotype maps.” *Bioinformatics*. 21(2), pp. 263-265. 2005. <https://doi.org/10.1093/bioinformatics/bth457>
- [7] Marchini, J., Howie, B. “Genotype imputation for genome-wide association studies.” *Nature Reviews Genetics*. 11, pp. 499-511. 2010. <https://doi.org/10.1038/nrg2796>
- [8] Stephens, M., Donnelly, P. “A comparison of bayesian methods for haplotype reconstruction from population genotype data.” *The American Journal of Human Genetics*. 73(5), pp. 1162–1169. 2003. <https://doi.org/10.1086/379378>
- [9] Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., Rozen, S. G. “Primer3 – new capabilities and interfaces.” *Nucleic Acids Research*. 40(15), e115. 2005. <https://doi.org/10.1093/nar/gks596>
- [10] Sovic, I., Sikic, M., Wilm, A., Fenlon, S. N., Chen, S., Nagarajan, N. “Fast and sensitive mapping of nanopore sequencing reads with Graph-Map.” *Nature Communications*. 7, Article number: 11307. 2016. <https://doi.org/10.1038/ncomms11307>
- [11] Li, H. “Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.” 2013. arXiv:1303.3997v1 [q-bio.GN]
- [12] Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., Mesirov, J. P. “Integrative Genomics Viewer.” *Nature Biotechnology*. 29, pp. 24–26. 2011. <https://doi.org/10.1038/nbt.1754>

- [13] Ammar, R., Paton, T.A., Torti, D., Shlien, A., Bader, G. D. "Long read nanopore sequencing for detection of HLA and CYP2D6 variants and haplotypes." *F1000Research*. 4, pp. 1-17. 2015.  
<https://doi.org/10.12688/f1000research.6037.1>
- [14] Laver, T. W., Caswell, R. C., Moore, K. A., Poschmann, J., Johnson, M. B., Owens, M. M., Ellard, S., Paszkiewicz, K. H., Weedon, M. N. "Pitfalls of haplotype phasing from amplicon-based long-read sequencing." *Scientific Reports*. 6, Article number: 21746. 2016.  
<https://doi.org/10.1038/srep21746>
- [15] Mikheyev, A. S., Tin, M. Y. M. "A first look at the Oxford Nanopore MinION sequencer." *Molecular Ecology Resources*. 14(6), pp. 1097-102. 2014.  
<https://doi.org/10.1111/1755-0998.12324>
- [16] Loose, M., Malla, S., Stout, M. "Real time selective sequencing using nanopore technology." *Nature Methods*. 13, pp. 751-754. 2016.  
<https://doi.org/10.1038/nmeth.3930>
- [17] Juul, S., Izquierdo, F., Hurst, A., Dai, X., Wright, A., Kulesha, E., Pettett, R., Turner, D. J. "What's in my pot? Real-time species identification on the MinION." bioRxiv 030742, 2015.  
<https://doi.org/10.1101/030742>
- [18] Nichols, D., Cahoon, N., Trakhtenberg, E. M., Pham, L., Mehta, A., Belanger, A., Kanigan, T., Lewis, K., Epstein, S. S. "Use of iChip for high-throughput in situ cultivation of "uncultivable" microbial species." *Applied and Environmental Microbiology*. 76(8), pp. 2445-2450. 2010.  
<https://doi.org/10.1128/AEM.01754-09>
- [19] Eckert, S. E., Chan, J. Z-M., Houniet, D., Breuer, J., Speight, G. "Enrichment of long DNA fragments from mixed samples for Nanopore sequencing." bioRxiv 048850, 2016.  
<https://doi.org/10.1101/048850>
- [20] Loman, N. J., Quick, J., Simpson, J. T. "A complete bacterial genome assembled de novo using only nanopore sequencing data." *Nature Methods*. 12, pp. 733-735. 2015.  
<https://doi.org/10.1038/nmeth.3444>
- [21] Rand, A. C., Jain, M., Eizenga, J., Musselman-Brown, A., Olsen, H. E., Akeson, M., Paten, B. "Cytosine Variant Calling with High-throughput Nanopore Sequencing." bioRxiv 047134, 2016.  
<https://doi.org/10.1101/047134>