Effects of DNA adducts on nanopore sequencing

Yrjö Koski[1,2], Divyesh Patel[2,3], Lauri Aaltonen[2,4,5], Biswajyoti Sahu[2,3,6], Kimmo Palin[2,4,5], Esa Pitkänen[1,2,5]

1: Institute for Molecular Medicine Finland (FIMM), University of Helsinki
2: Applied Tumor Genomics Research Program, Faculty of Medicine, University of Helsinki
3: Centre for Molecular Medicine Norway (NCMM), University of Oslo
4: Department of Medical and Clinical Genetics, University of Helsinki
5: iCAN Digital Precision Cancer Medicine Flagship, Finland
6: Department of Cancer Genetics, Institute for Cancer Research, Norway

Institute for Molecular Medicine Finland
Nordic EMBL Partnership for Molecular Medicine
HiLIFE UNIT

Background and motivation

- DNA adducts are covalently bound lesions in the DNA that can give rise to mutations that can lead to cancer
- Mass spectrometry-based methods can be used to detect different adducts in a DNA sample simultaneously, but the sequence context of adducts is lost
- Nanopore sequencing can accurately detect DNA modifications at single-nucleotide precision and can also be applied to DNA adduct detection [1]

Research aims:

- Detect DNA adducts with nanopore sequencing
- Investigate effects of adducts at base-pair resolution
- Understanding the interplay of DNA damage and mutations

Materials and methods

- We prepared a novel dataset by treating *S. cerevisiae* DNA with four different adduct-forming compounds (Figure 1)
- We analyzed nanopore sequencing data from three distinct perspectives (Figure 2)
- We performed distributional testing on 9-mer statistics between the treated and control samples
- We ran motif discovery analysis with STREME [2] for significant 9-mers to examine putative binding motifs of adducts

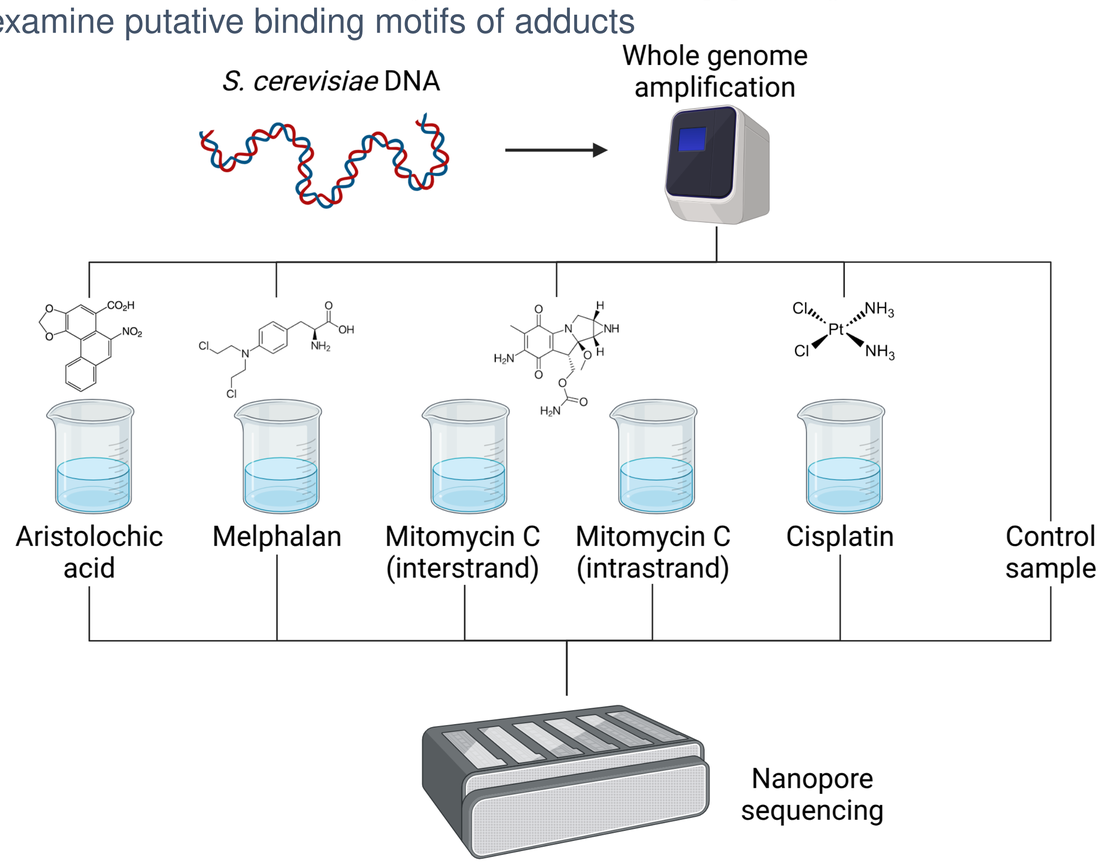

Figure 1: Experimental setup

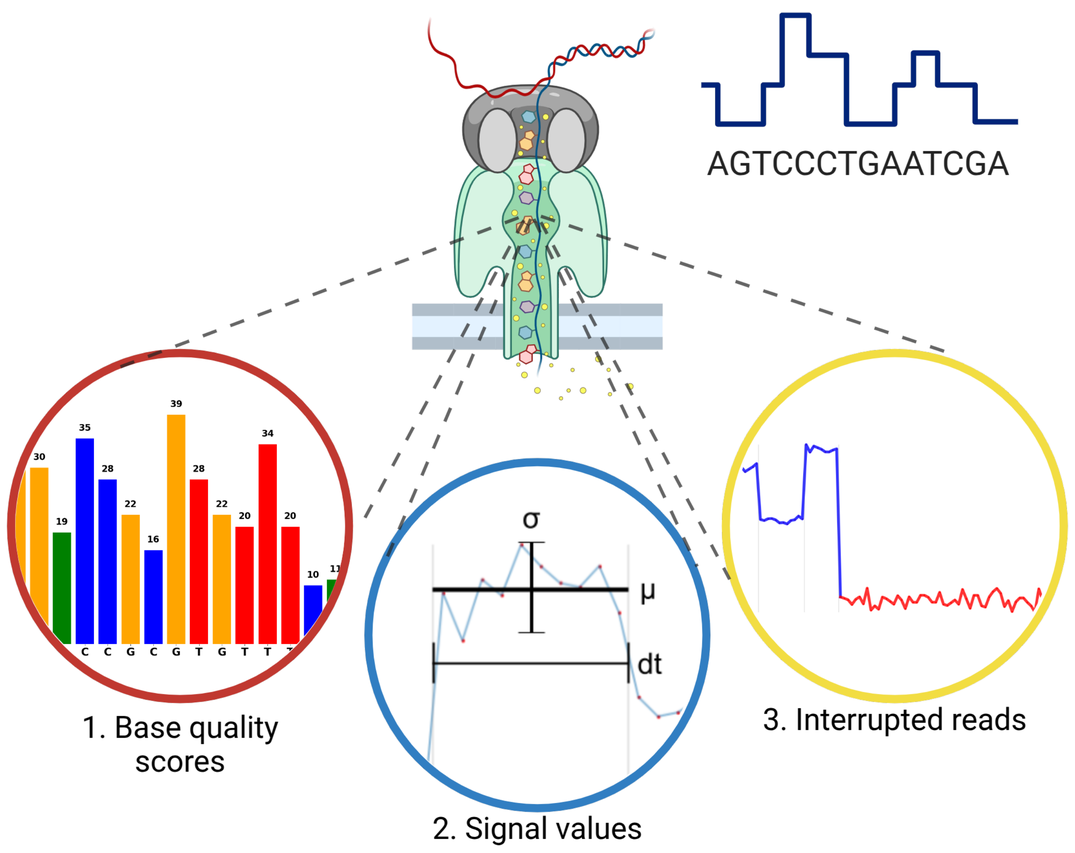

Figure 2: Summary of different data-analysis approaches

[1] I. Nookaew et al. Detection and Discrimination of DNA Adducts Differing in Size, Regiochemistry, and Functional Group by Nanopore Sequencing, *Chemical Research in Toxicology* (2020)

[2] T. Bailey, STREME: accurate and versatile sequence motif discovery, *Bioinformatics* (2021)

Results

- Aristolochic acid and melphalan treatment showed significant effects on base quality scores, signal value statistics and read interruptions (Table 1)
- The position of an adduct can have different effect on the ionic current based on its position inside the nanopore (Figure 3)
- Adducts can induce sequence-specific read interruptions (Figure 3)
- No significant motifs associated with mitomycin C adducts. This might be due to small amount of adducts in the samples

Sample	Quality score	Windowed mean error	Windowed std error	Signal mean	Signal std	Signal dwell time	Signal shifted dwell time	Interrupted reads
Aristolochic acid	✓	✓	✓	✓	✓	✓	✗	✓
Melphalan	✓	✓	✓	✓	✓	✓	✓	✓
Mitomycin C interstrand	✗	✗	✗	✗	✗	✗	✗	✗
Mitomycin C intrastrand	✗	✗	✗	✗	✗	✗	✗	✗
Cisplatin	✗	✓	✓	✓	✓	✗	✗	✗
Cisplatin size-selected	✗	✗	✗	✓	✓	✗	✗	✗

Table 1: Association of discovered binding motifs with different variables

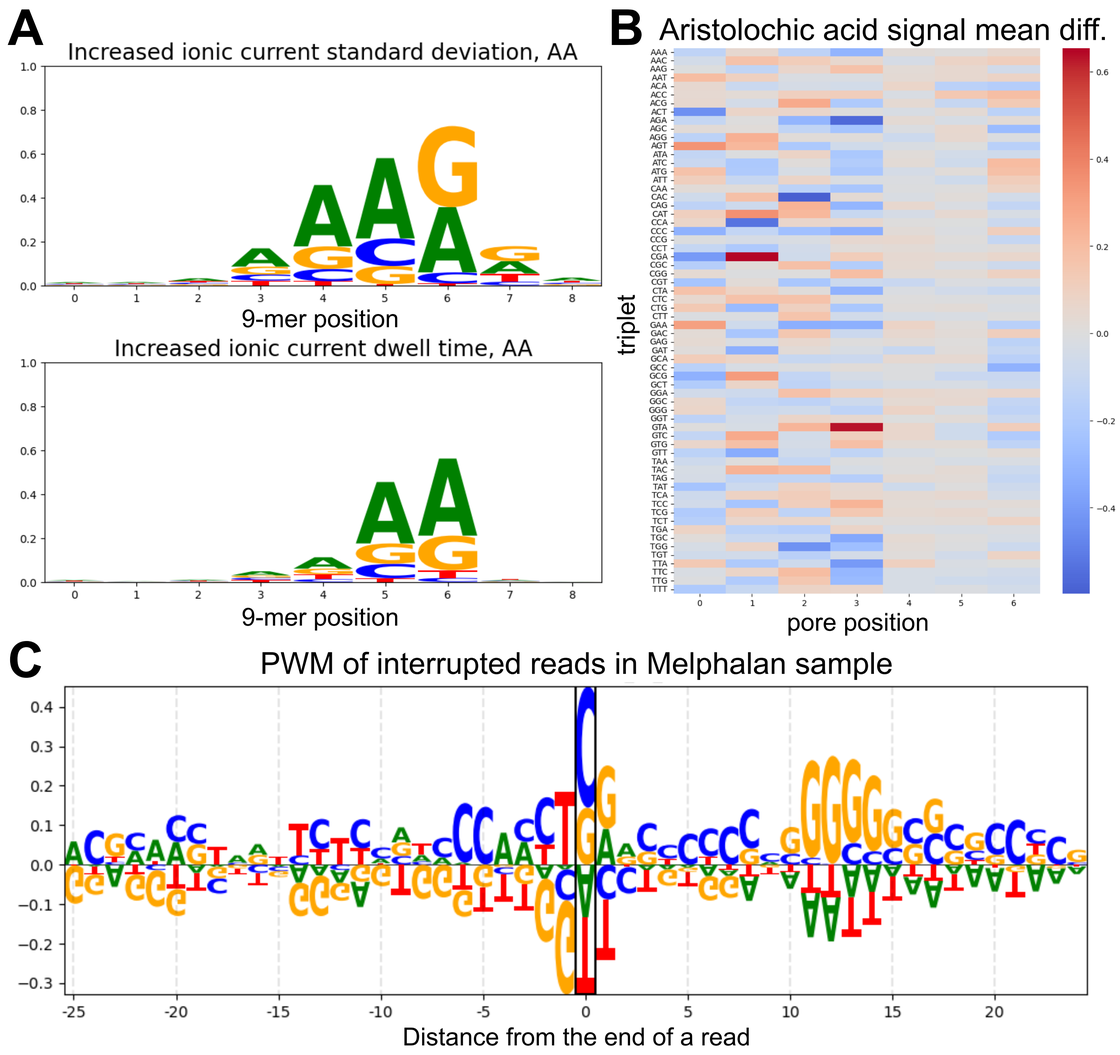

Figure 3: A) Binding motif of aristolochic acid associated with higher ionic current standard deviation and longer dwell times B) Signal mean difference for triplets and pore positions in aristolochic acid samples C) Sequence logo of interrupted reads in melphalan samples

Conclusions

- DNA adducts can affect base quality scores and ionic current in nanopore sequencing, as well as cause read interruptions
- Many data-analysis approaches should be considered when applying nanopore sequencing for adduct detection
- DNA adduct detection with nanopore sequencing is challenging when the fraction of adducted positions is low

Acknowledgements

Academy of Finland, iCAN Digital Precision Cancer Medicine Flagship, University of Helsinki, and CSC – Finnish IT center for science

Figures 1 and 2 were created with Biorender.com

ACADEMY OF FINLAND
CSC