FI:IV110/IV114 Projekt z bioinformatiky a systémové biologie

Monika Čechová, Matej Lexa

 @biomonika, @matej_lexa

Interests

- Sex Chromosomes
- Satellite Biology and Heterochromatin
- Long reads and complete genomes
- Early Embryonic Development
- Reproductive Biology

Education

PhD Major in Biology, Minor in Statistics, 2020

Penn State, USA

MS in Bioinformatics, 2013

Masaryk University, Brno

BS in Applied Informatics, 2011

Masaryk University, Brno

My motivation:

Answering biological questions using next-generation sequencing data

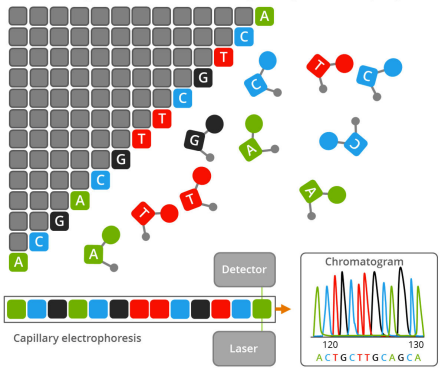
- Course requirements
- Project proposals
- Nanopore sequencing: *collaboration with the Institute of Biophysics of the Czech Academy of Sciences*
- Introduction to the Nanopore sequencing and its applications
- Hands-on exercise: species identification

Detekce kontaminací

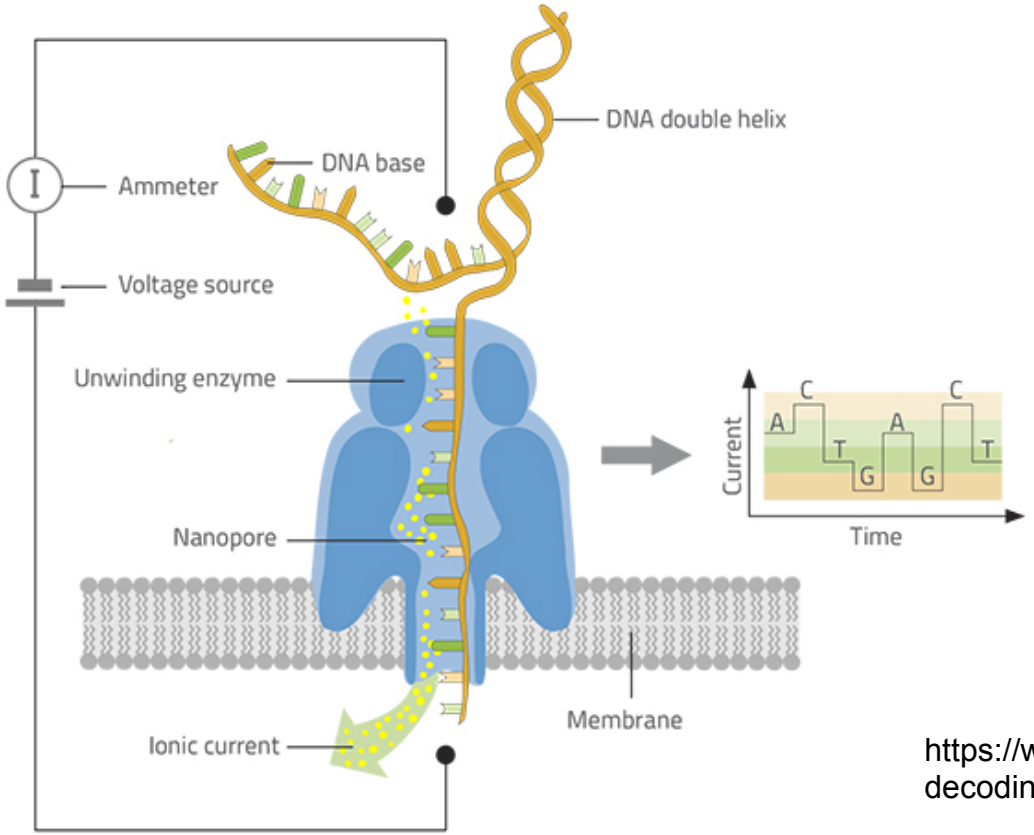
- Dr. Lucie Grodecká a Dr. Roman Hobza

Nanopore sequencing

https://www.gatc-biotech.com/fileadmin/_processed_/csm_Sanger_sequencing_illustration_small_898122494c.jpg



Nanopore



https://www.scienceinschool.org/content/decoding-dna-pocket-sized-sequencer

Nanopore

- Introduction (3 mins)
  - https://www.youtube.com/watch?v=GUb1TZvMWsw

https://twitter.com/MakovaLab

Real-time surveillance

Real time genomic surveillance of Ebola outbreak 2014-2015

05 Jun 2015

The current Ebola outbreak in West Africa is the largest ever recorded, with over 26,500 cases reported resulting in an estimated 11,000 deaths. Yet genomic surveillance of this outbreak has been patchy, hampered by understandable but vexing logistical, social, political and technical obstacles in securing and transporting samples for processing.

We wanted to help address the gaps in our knowledge of viral evolution and to generate data for epidemiological use. So, in April, Josh Quick from my group went to Conakry, Guinea to establish proof-of-principle for portable nanopore sequencing. This was the most practical way we could rapidly establish a local sequencing lab in order to generate real-time information.

Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*

Joshua Quick †, Philip Ashton †, Szymon Calus, Carole Chatt, Savita Gossain, Jeremy Hawker, Satheesh Nair, Keith Neal, Kathy Nye, Tansy Peters, Elizabeth De Pinna, Esther Robinson, Keith Struthers, Mark Webber, Andrew Catto, Timothy J. Dallman, Peter Hawkey and Nicholas J. Loman

†Contributed equally

Genome Biology 2015 **16**:114
https://doi.org/10.1186/s13059-015-0677-2 | © Quick. 2015
Received: 25 February 2015 | Accepted: 14 May 2015 | Published: 30 May 2015

Sequencing the station: Investigation aims to identify unknown microbes in space

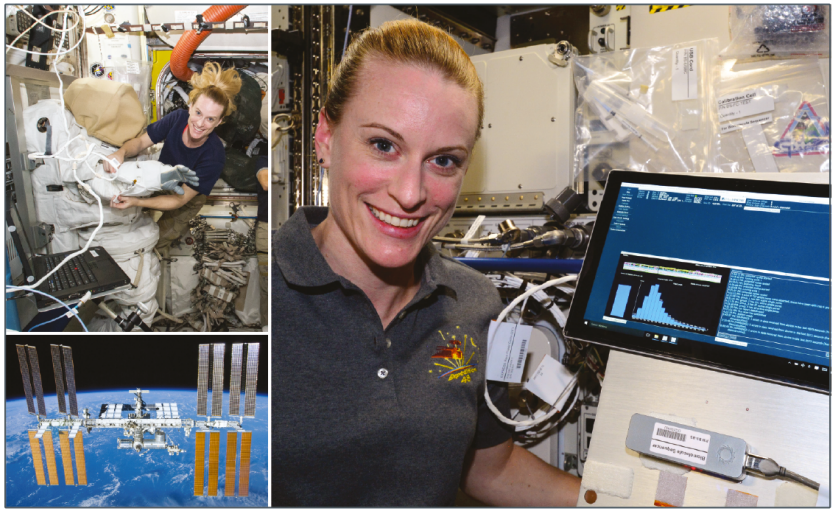


Fig. 1 Astronaut Kate Rubins on the ISS

https://phys.org/news/2017-04-sequencing-station-aims-unknown-microbes.html

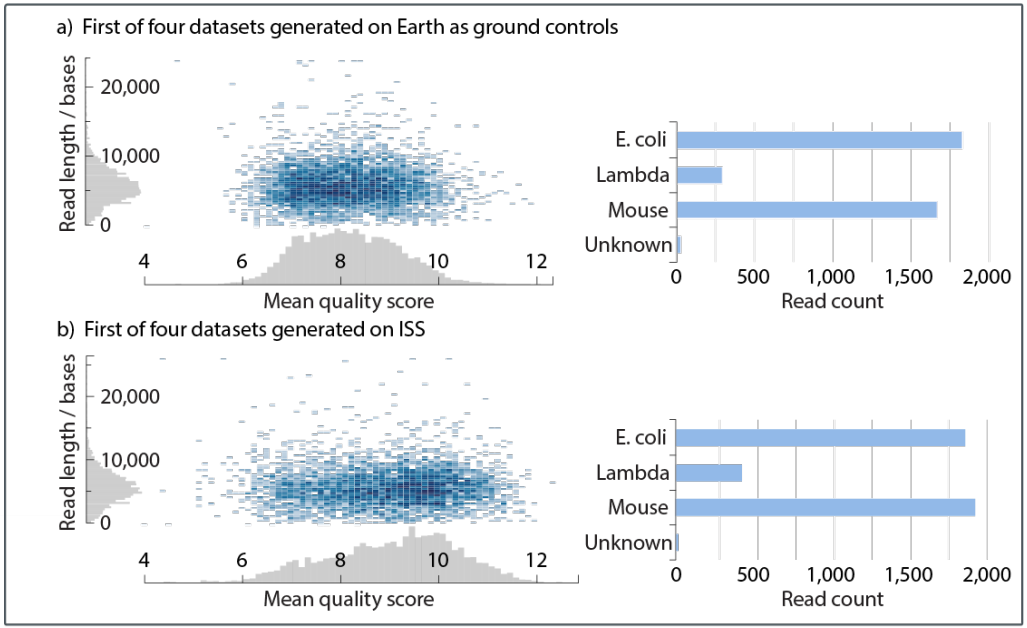
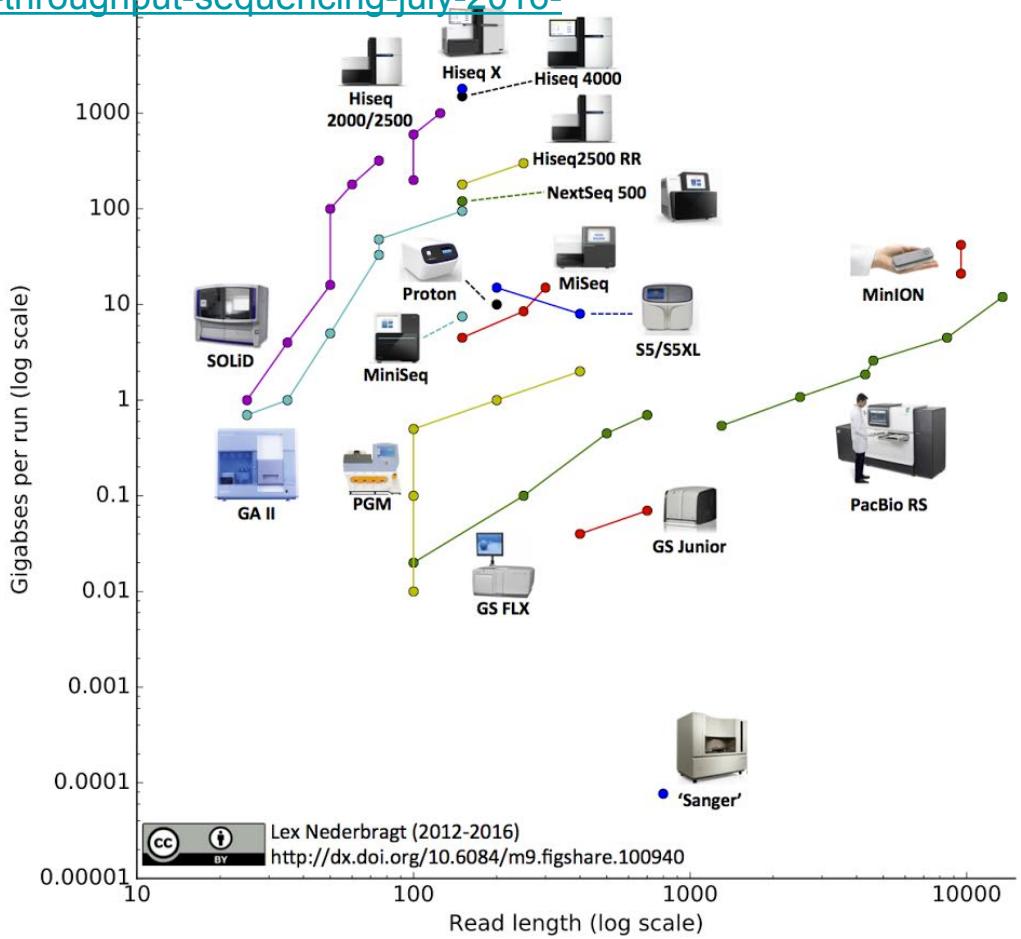


Fig. 2 Analysis workflow showing read quality for runs a) on Earth b) on the ISS

https://nanoporetech.com/resource-centre/posters/dna-sequencing-microgravity-international-space-station-iss-using-minion

https://flxlexblog.wordpress.com/2016/07/08/developments-in-high-throughput-sequencing-july-2016-edition/



Long-read sequencing

- Sample quality
- Library preparation (size selection, repair)
- Protocols
- Sequencing throughput
- Multiplexing

Budgeting is important: get yourself familiar with the cost of reagents and sequencing!

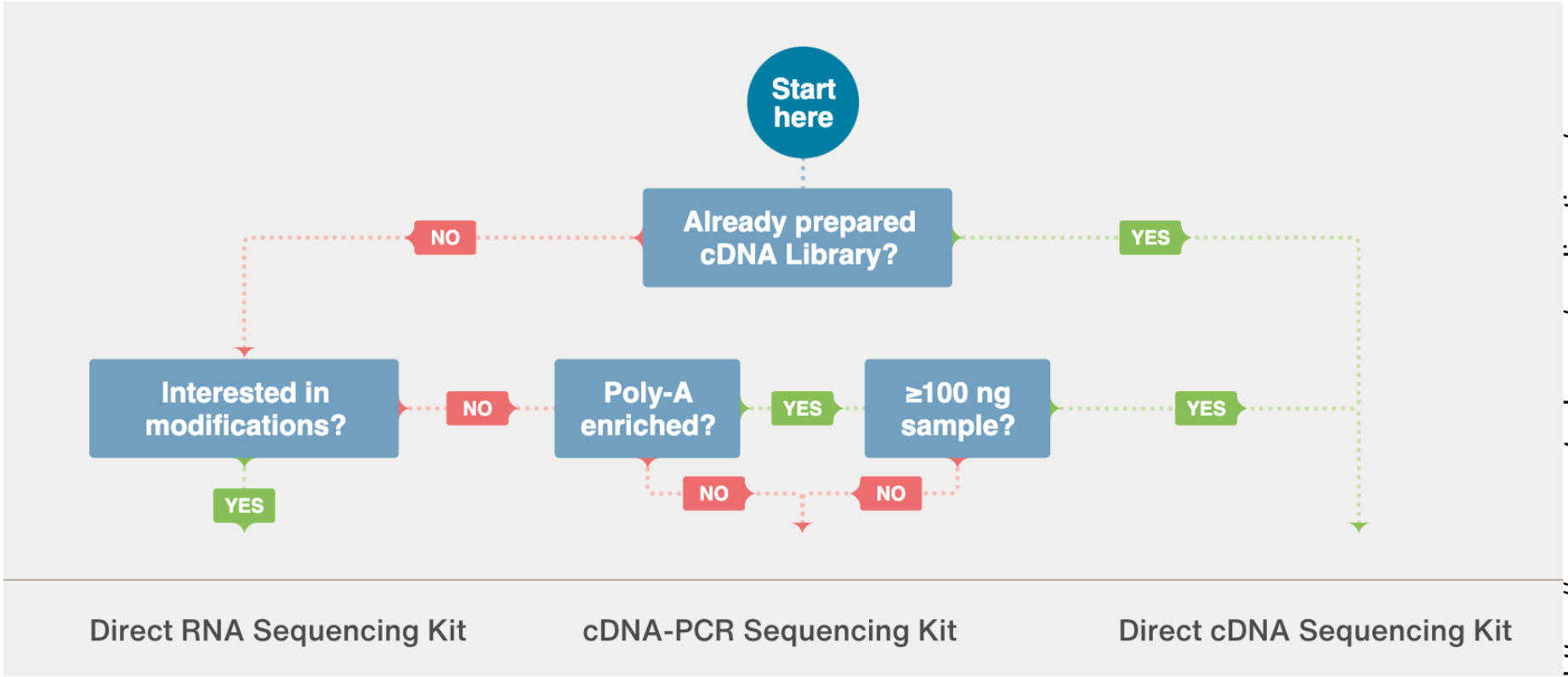
There’s always a trade-off.

Long-read applications

Long reads

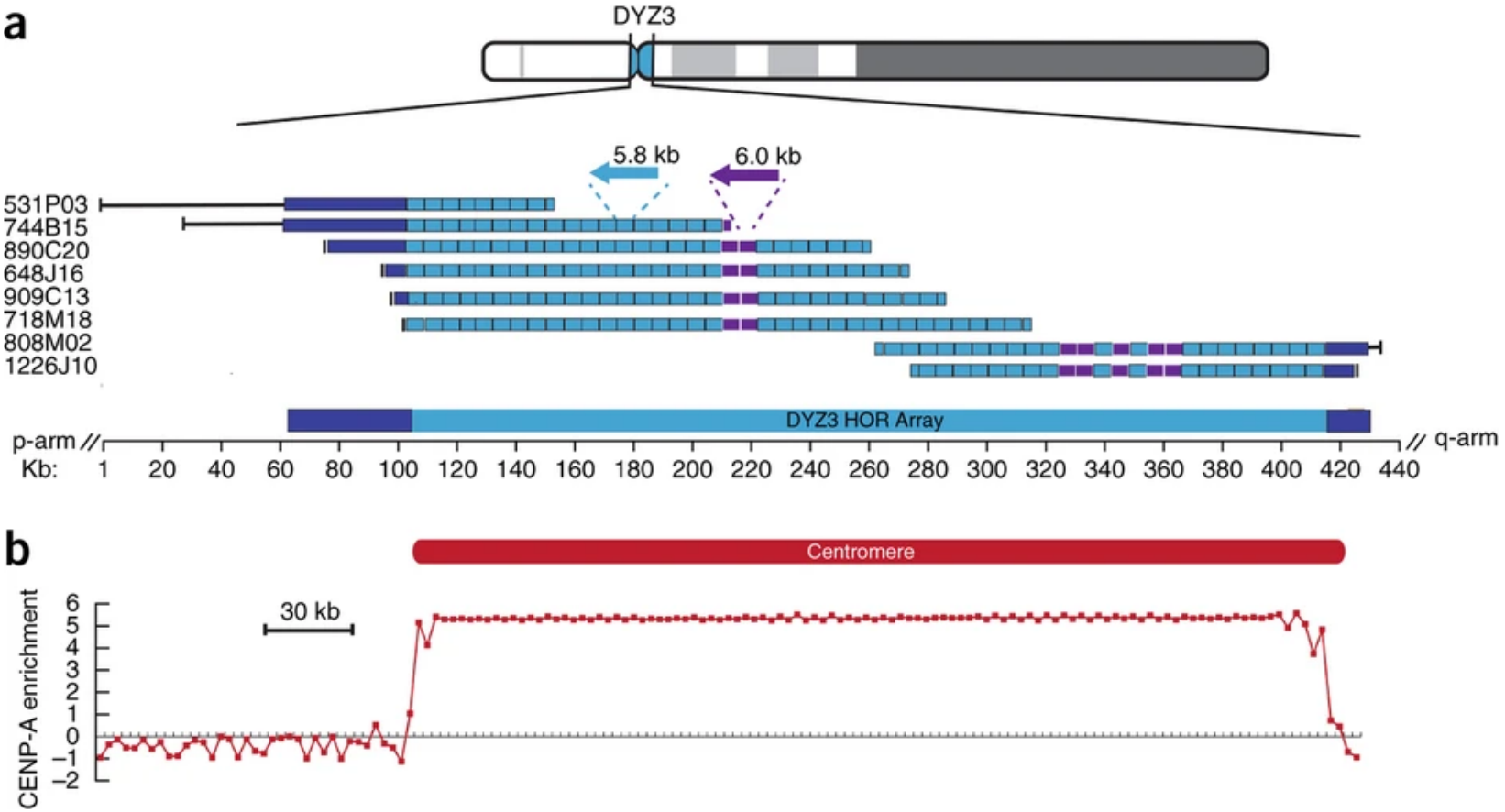
1. **Transcriptome (isoforms)**
2. **Assembly (PacBio HiFi + ultra-long Nanopore reads)**
3. **Epigenetic modifications**

1. RNA sequencing with Nanopore



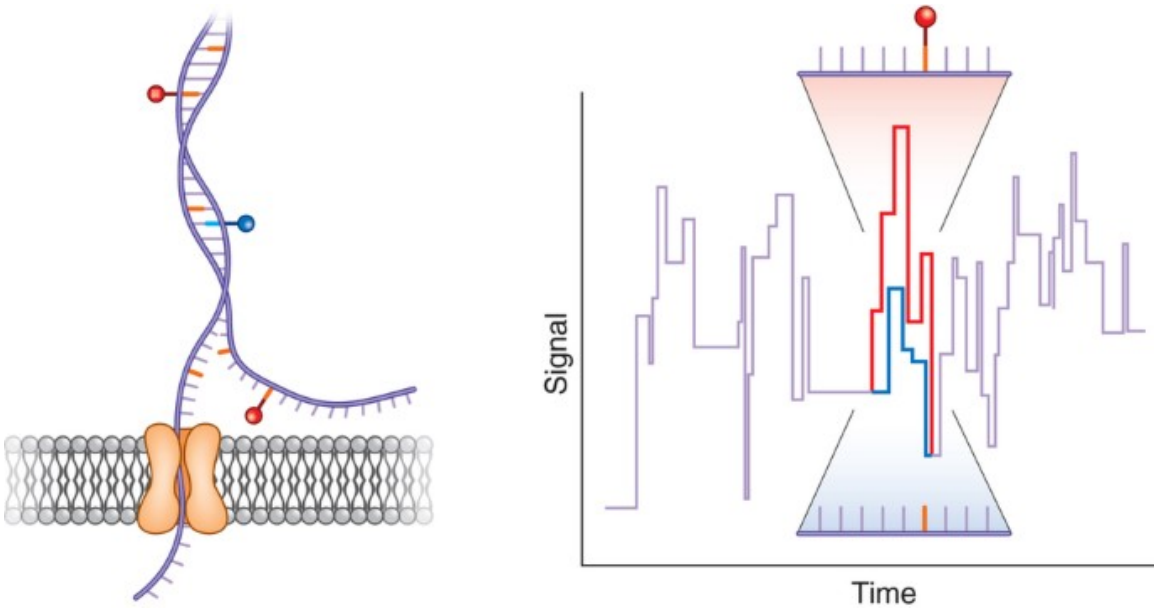
https://nanoporetech.com/applications/rna-sequencing

2. Linear assembly of the RP11 Y centromere



3. Nanopore sequencing meets epigenetics

Figure 1 | DNA methylation can be read out directly by nanopore sequencing. Single-stranded DNA alters ionic current in a sequence-dependent manner as it passes through a pore (left), with methylated bases highlighted with red and blue tags. The raw current signal (right) indicates small changes due to methylation that a new set of algorithms can robustly interpret.



Schatz, M. Nanopore sequencing meets epigenetics. *Nat Methods* 14, 347–348 (2017). https://doi.org/10.1038/nmeth.4240

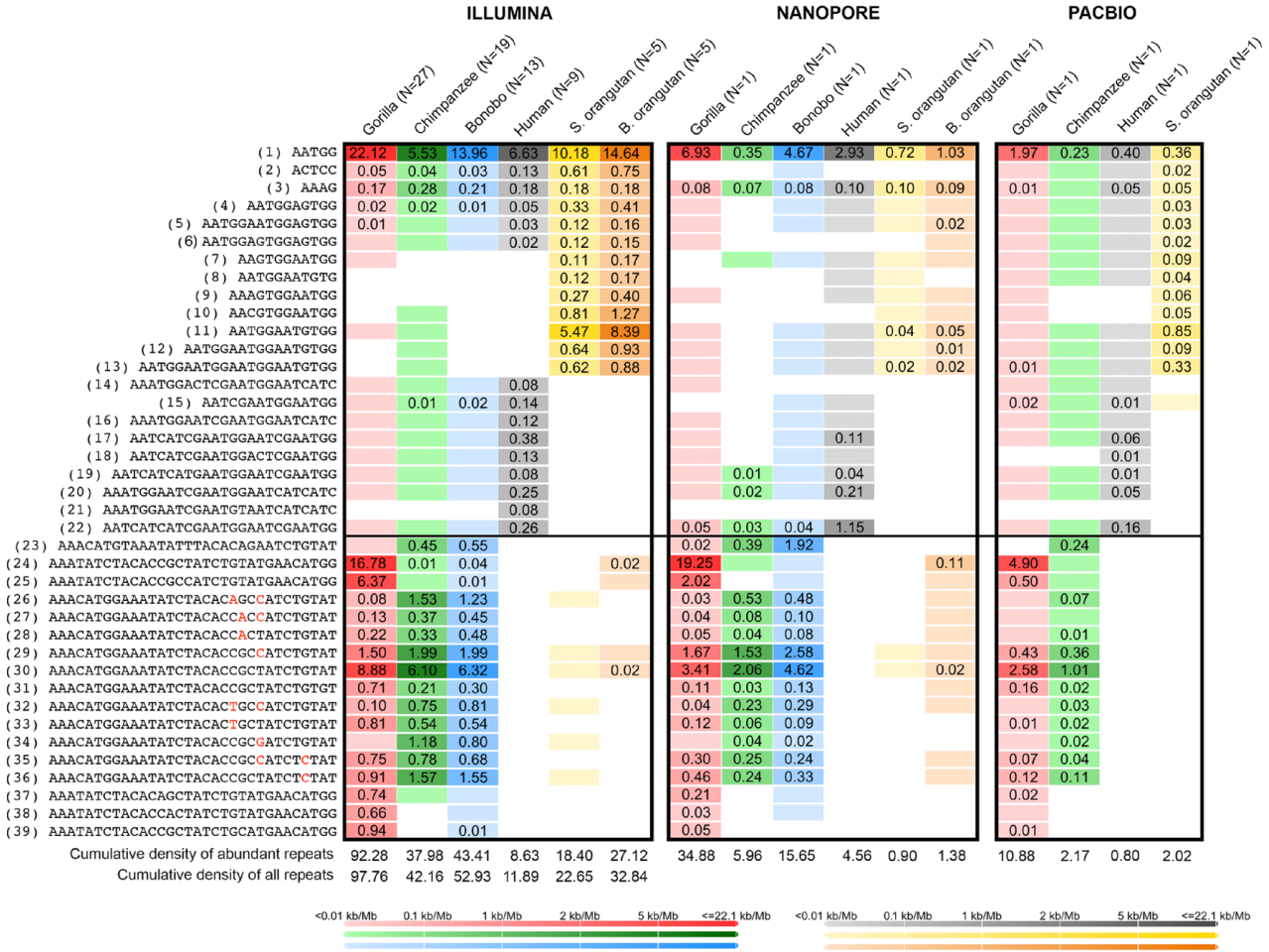
Error profiles

Table S1. Observed error rates in PacBio and Nanopore alignments. m=matches, mm=mismatches, io=insertion open, ix=insertion extend, do=deletion open, dx=deletion extend. Overall error rates are measured as 1 - m/(m+mm+io+ix+do+dx).

Technology	Error Rate	m	mm	io	ix	do	dx
PacBio	14.90%	203521495440	4039815573	15516830204	5868396512	8622457071	1709970892
		85.06%	1.69%	6.48%	2.45%	3.60%	0.71%
ON	16.10%	17603088986	965110092	492506288	299904008	702350550	919627709
		83.89%	4.60%	2.35%	1.43%	3.35%	4.38%

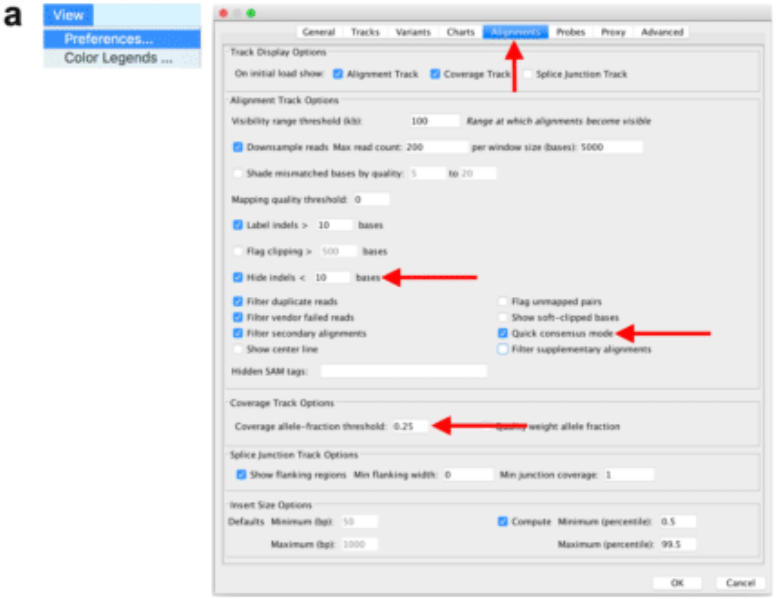
RS Harris, **M Cechova**, KD Makova. *Noise-Cancelling Repeat Finder: Uncovering tandem repeats in error-prone long-read sequencing data. BIOINFORMATICS*, 2019.

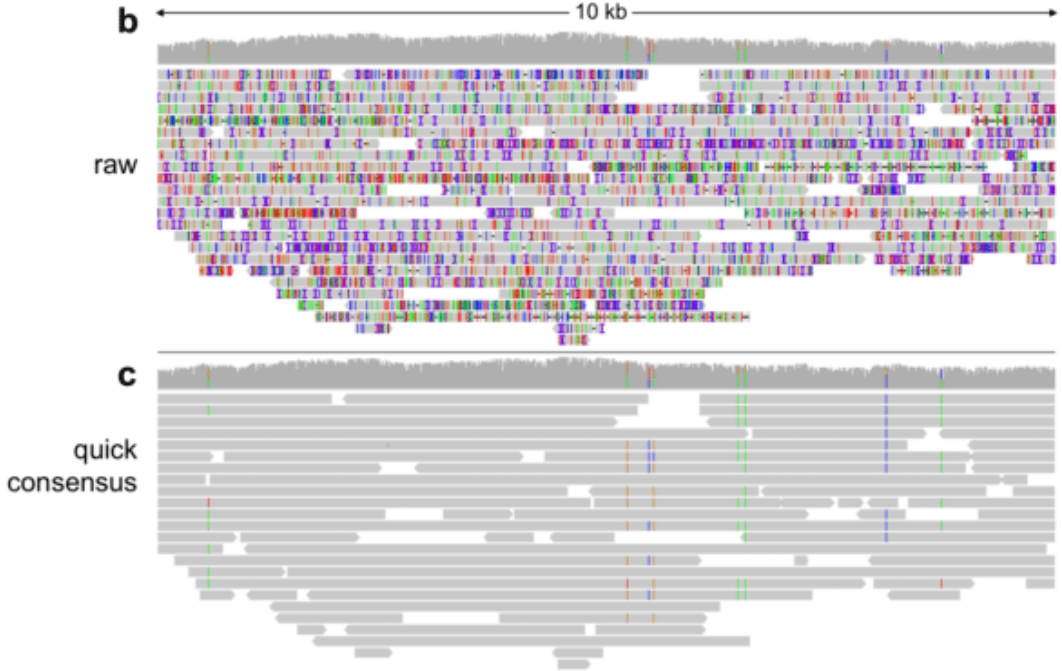
We demonstrated that orthogonal technologies (such as Illumina, Nanopore, and PacBio) are generally concordant in distinguishing between highly and lowly abundant repeated motifs.



Cechova et al., 2019
Harris et al., 2019

The visualization of sequencing errors

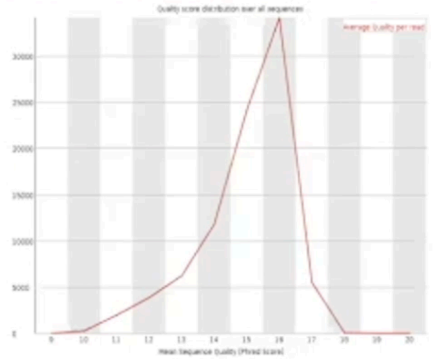


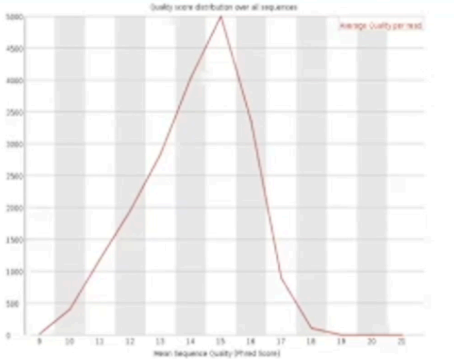


https://www.pacb.com/blog/igv-3-improves-support-pacbio-long-reads/

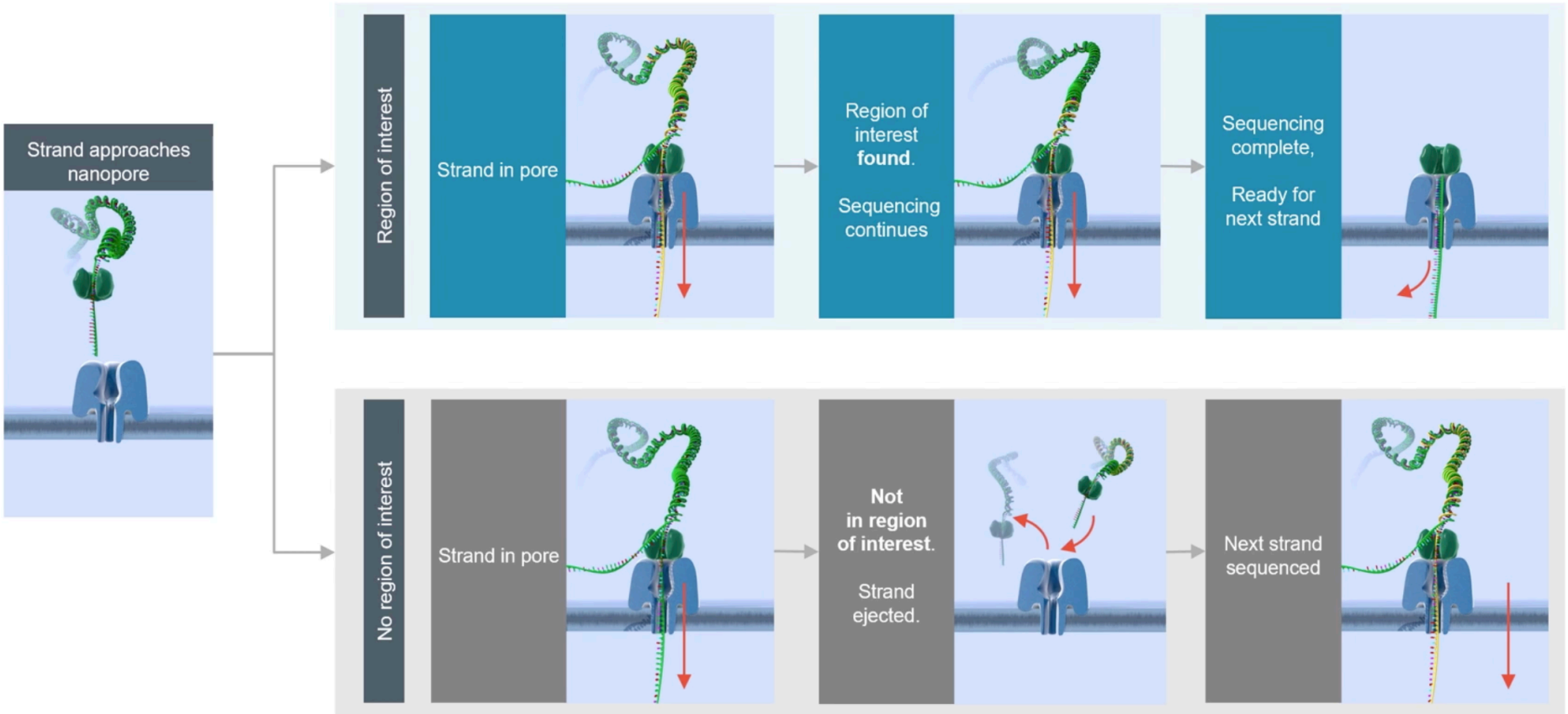
Metagenomics

- https://nanoporetech.com/resource-centre/snow-sledges-and-sequencing-grid-metagenomics-polar-expedition





Read until



https://nanoporetech.com/resource-centre/read-until-adaptive-sampling

Hands-on exercise: species identification

- Go to the folder: https://is.muni.cz/auth/el/fi/podzim2021/IV110/data/fast5/
- Download three files in .fast5 format
- Identify the species(s) of origin
- Upload your results into Odevzdávarna

Fast5 format: https://medium.com/@shiansu/a-look-at-the-nanopore-fast5-format-f711999e2ff6