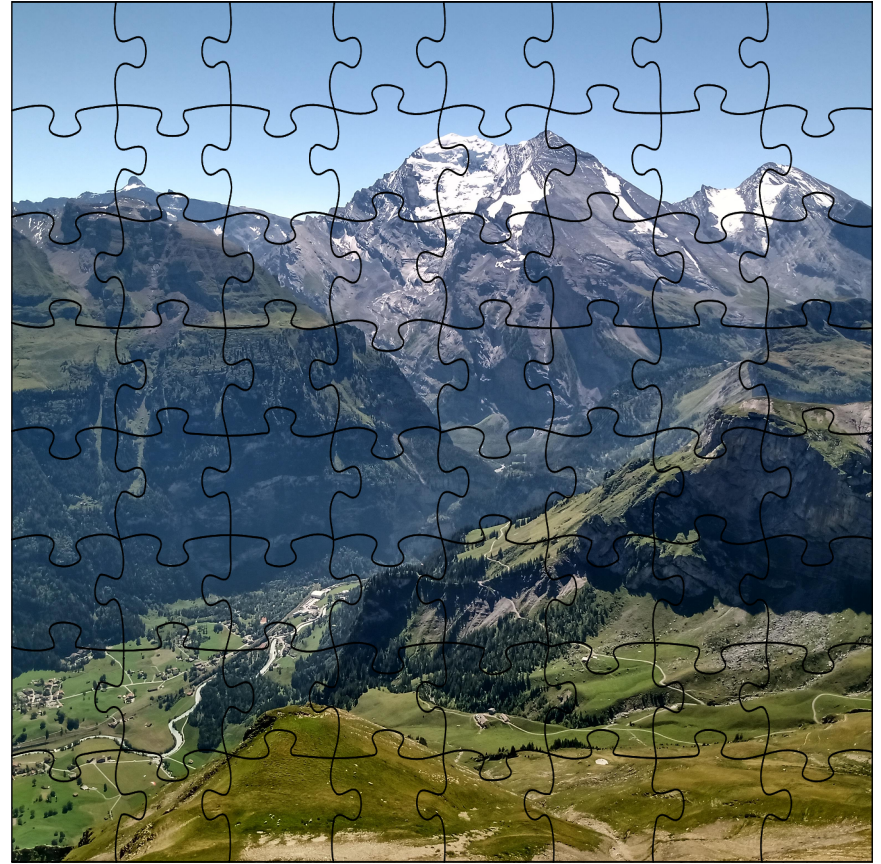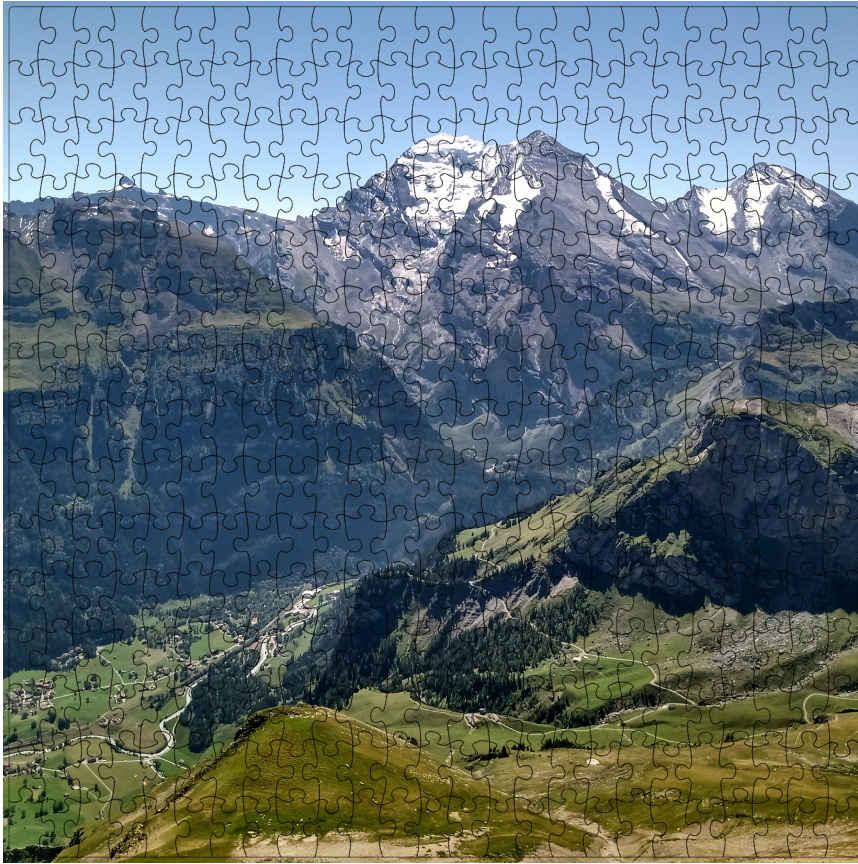# Long-read sequence analysis

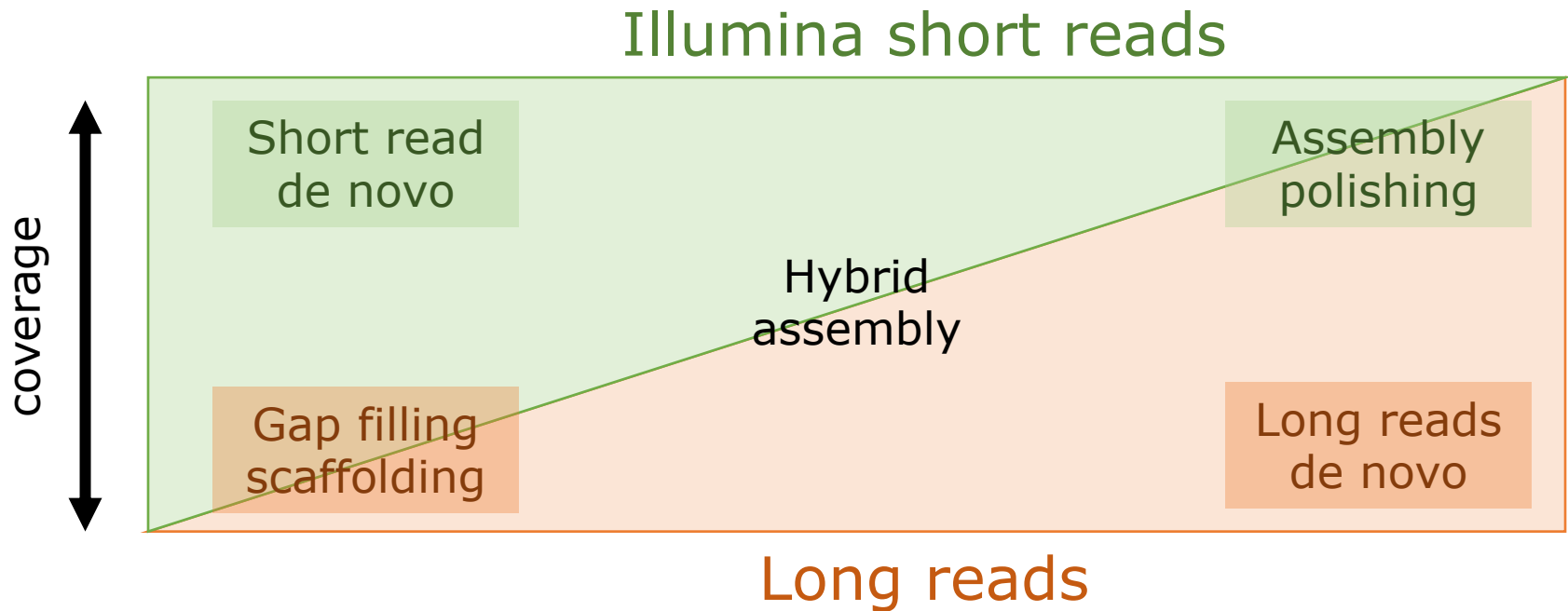**Applications**

Why long reads?

Applications

- (Genome) assembly
- Variant analysis
- Transcriptome analysis
- Epigenetics
- Metagenomics

Assembly

- Reconstructing a DNA sequence out of fragmented sequences:
  - Better understand variations between and within species
  - Reduces resources and increases accuracy for many applications
- Longer reads:
  - Less assembly errors
  - Higher contiguity (longer contigs)
  - Lower computational resources required

Assembly

Illumina short reads

coverage

Short read
de novo

Assembly
polishing

Hybrid
assembly

Gap filling
scaffolding

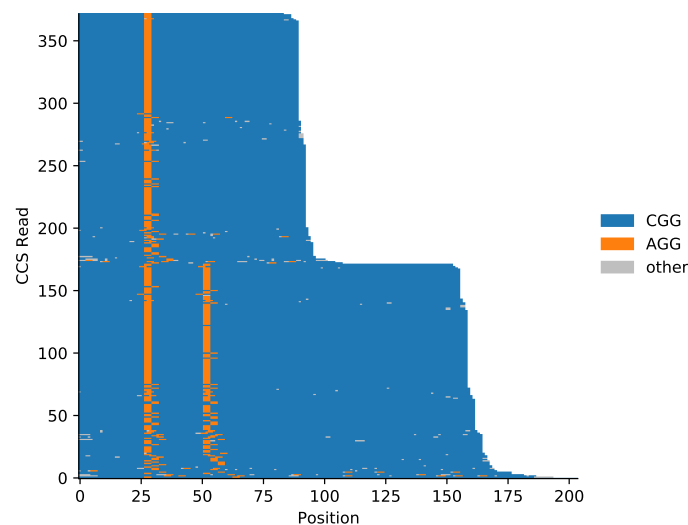Long reads
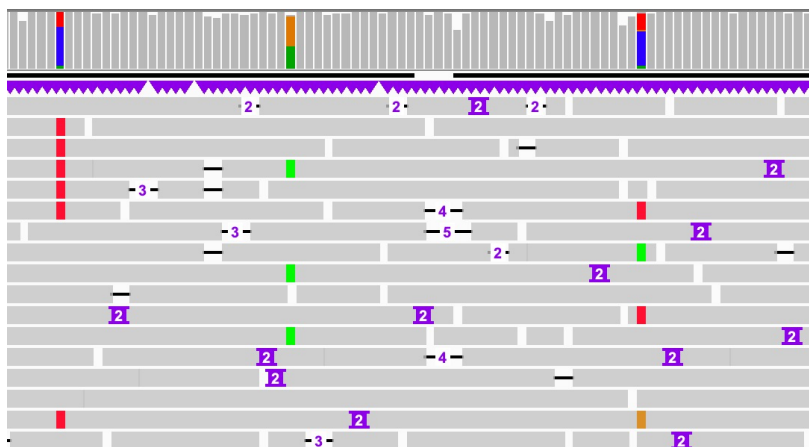de novo

Long reads

LR assembly software

- Microbial:
  - HGAP
  - Trycycler

- Large(r) genomes:
  - Shasta (ONT only)
  - Raven
  - wtdbg2
  - Falcon (PacBio only)
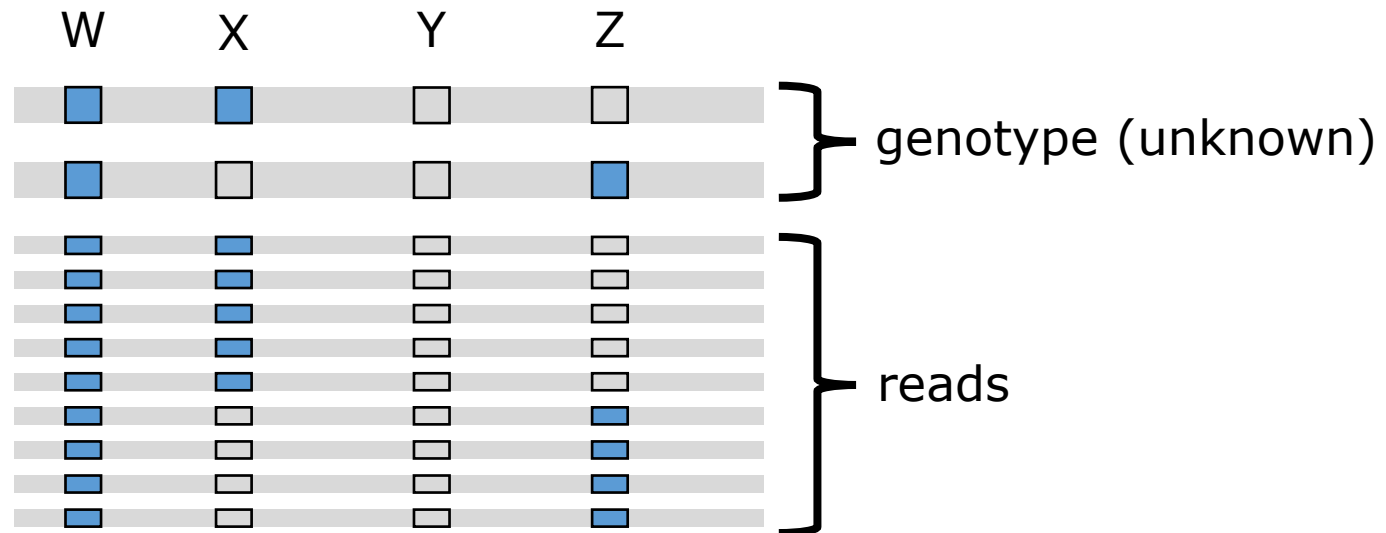 - Canu
  - Flye

Variant analysis

- Low accuracy long reads: not particularly good at variant analysis

- Major applications
  - Structural variation/large repeats
  - Phasing of variants

Phasing

- Variant calling:
    - DeepVariant (Pacbio + ONT)
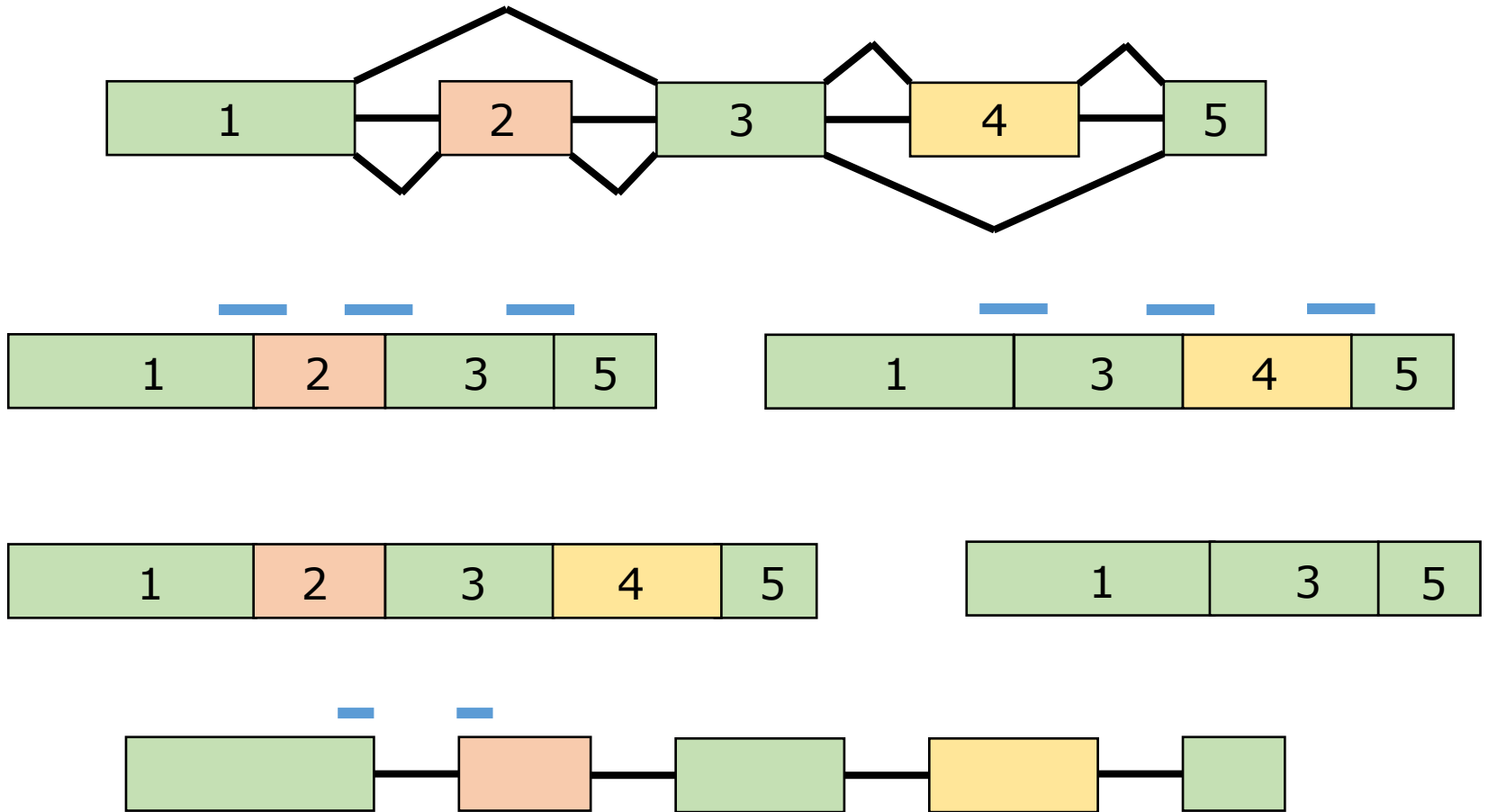   - Medaka (ONT)
- Phasing: Whatshap

# Repeat expansion - project 2

Transcriptome analysis

Transcriptome analysis

- Full transcripts: less ambiguous identification + quantification of transcript isoforms

- Better gene model prediction (e.g. for genome annotation)

- ONT direct RNA-seq:
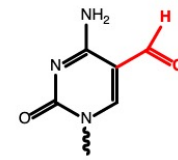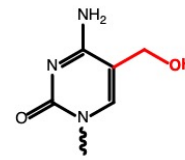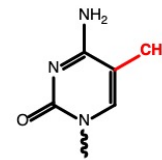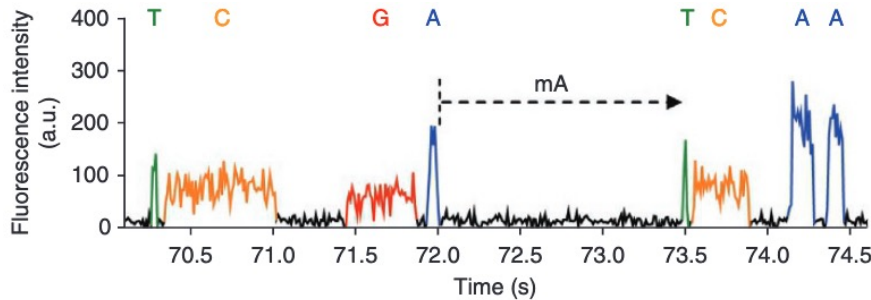 - No PCR amplification
  - No reverse transcription
  - Base modification identification

Epigenetics

- Base modification in non-amplified libraries (Pacbio and ONT)
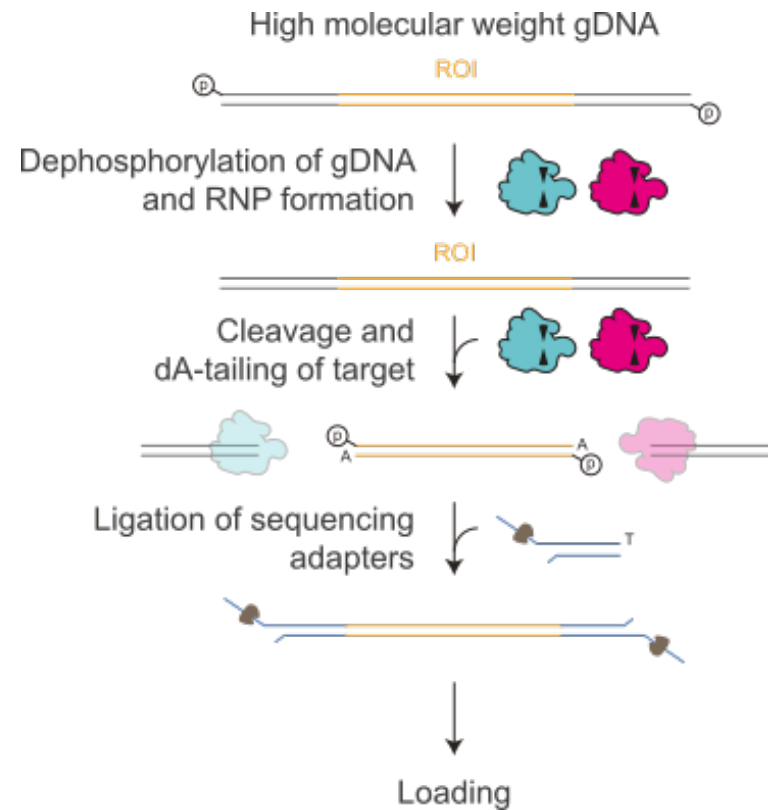- No multiplexing
- Also possible in direct RNA-seq (ONT)

12
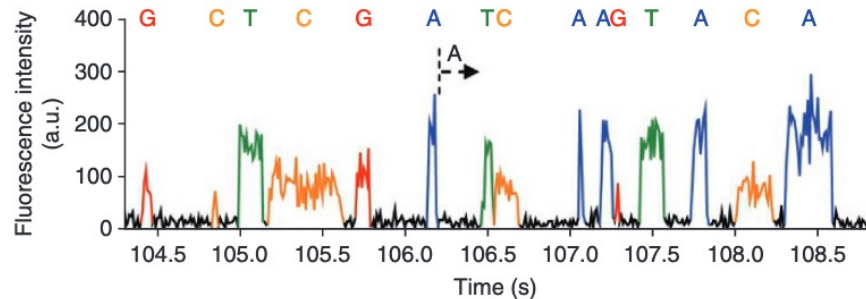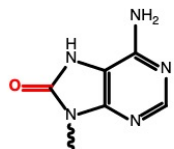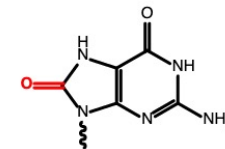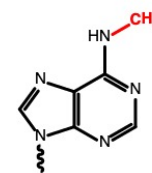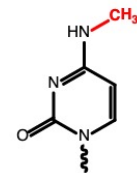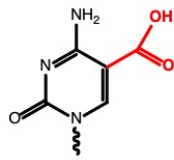
Metagenomics

- Full length 16S genes: better classification (PacBio CCS)

- Metagenome Assembled Genomes (MAGs) -> `flye --meta`

New possibilities

- Cas9 based targeting (both PacBio & ONT): also base modifications

- Adaptive sampling (ONT)

High molecular weight gDNA

Dephosphorylation of gDNA and RNP formation

Cleavage and dA-tailing of target

Ligation of sequencing adapters

Loading

https://genomebiology.biomedcentral.com/articles/10.1186/s13059-021-02582-x
https://www.nature.com/articles/s42003-021-01648-7

image: https://store.nanoporetech.com/catalog/product/view/id/461/s/cas9-sequencing-kit/category/28/

14