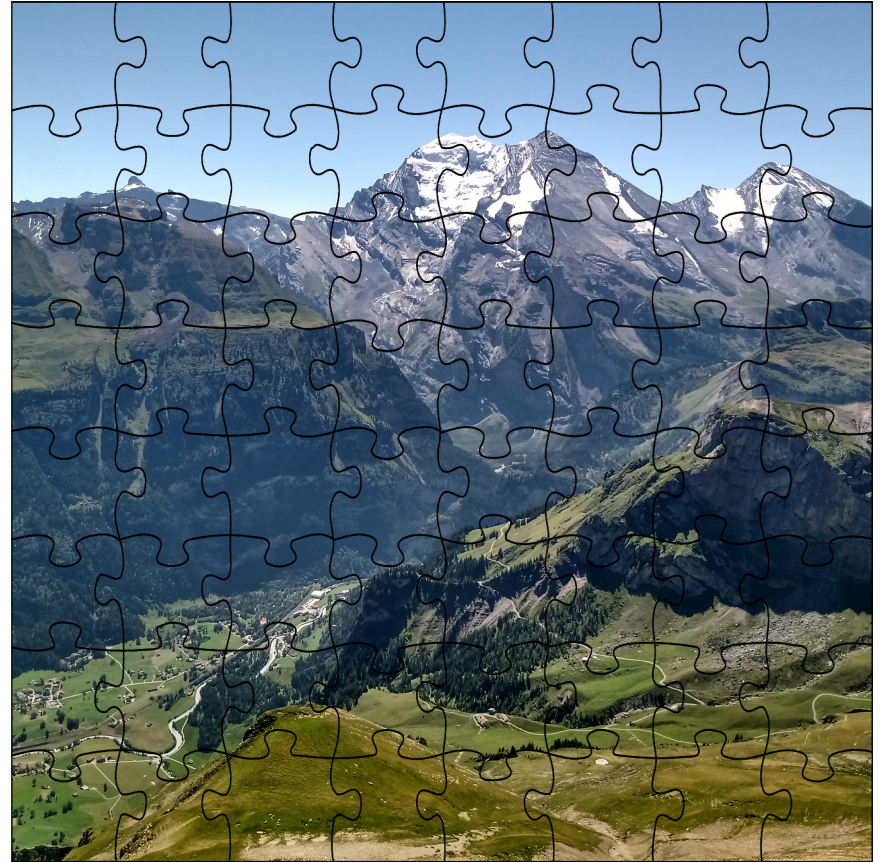
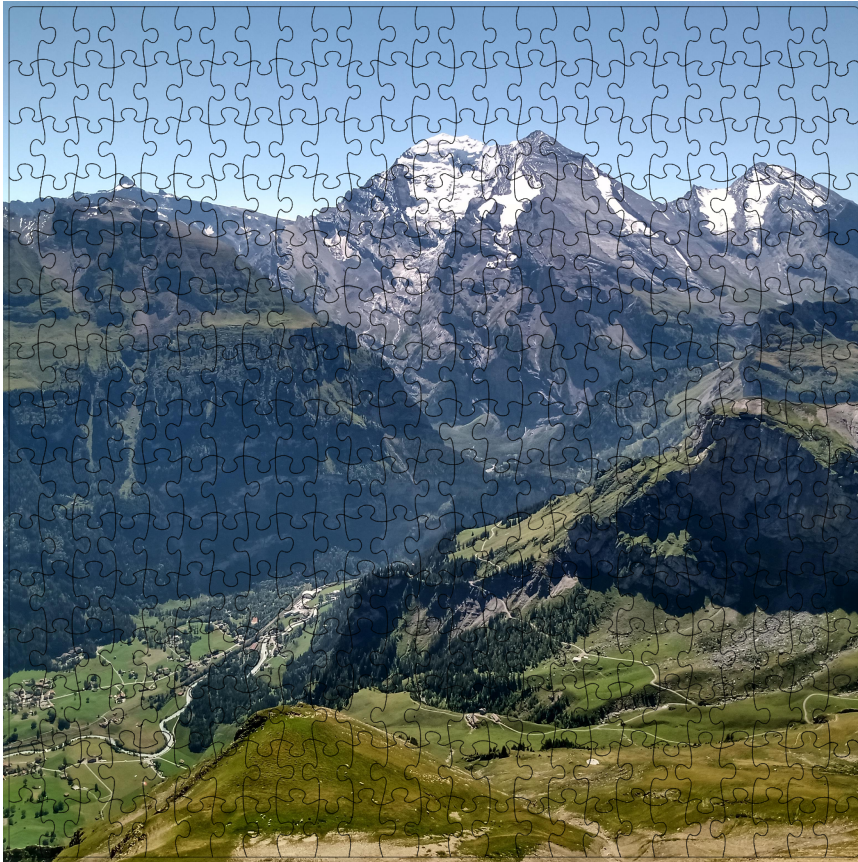


# Long-read sequence analysis

Applications

# Why long reads?



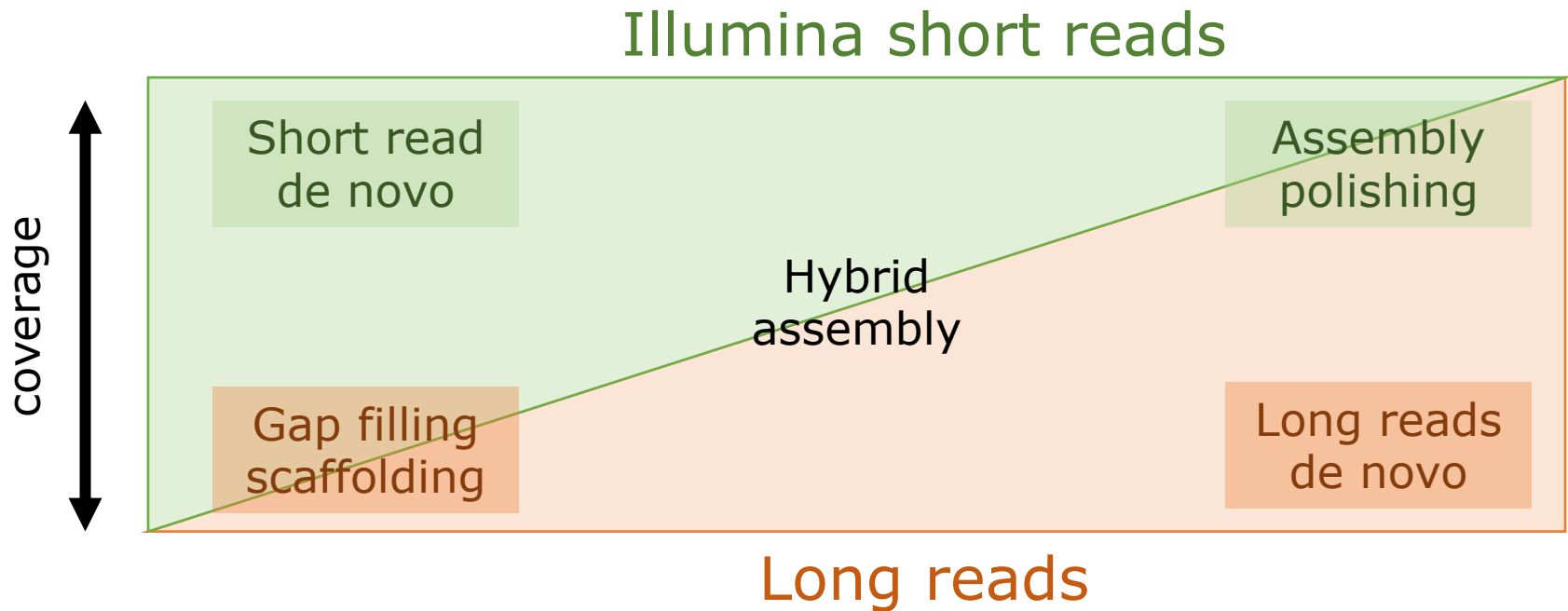
# Applications

- (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

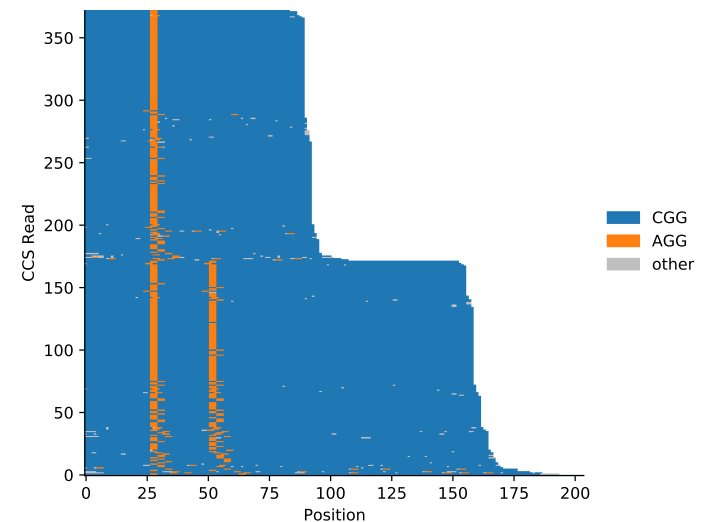
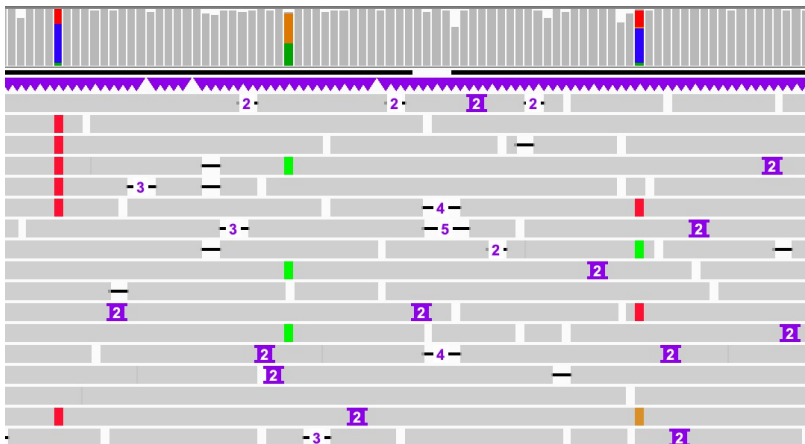


# LR assembly software

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

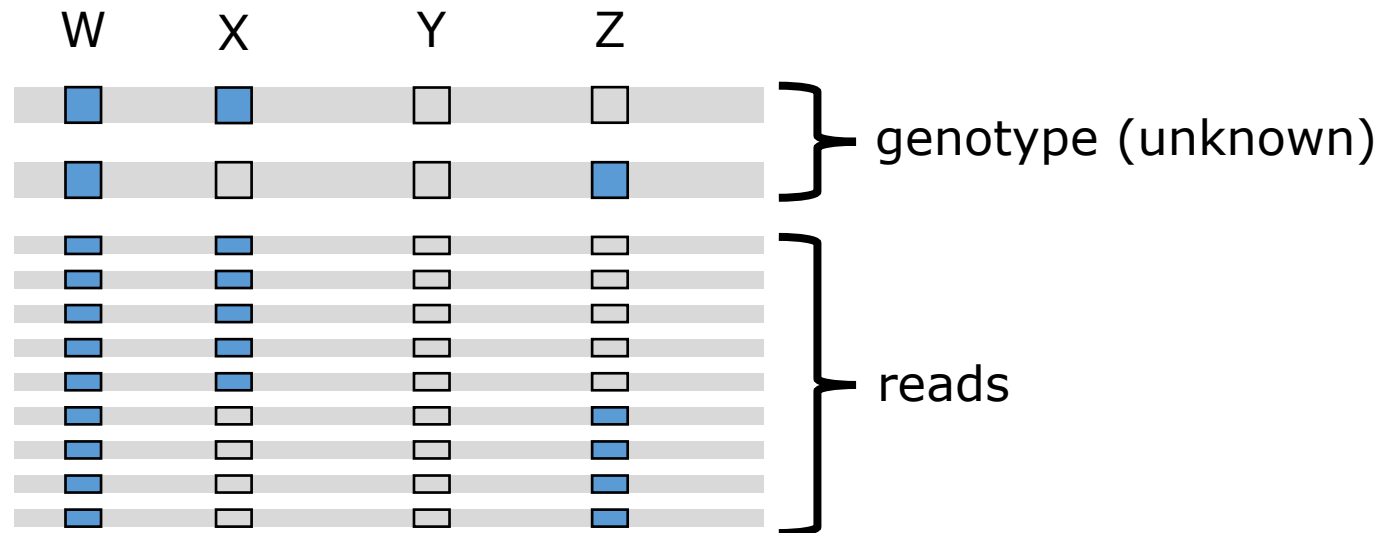
# Variant analysis

- 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: Whatsap

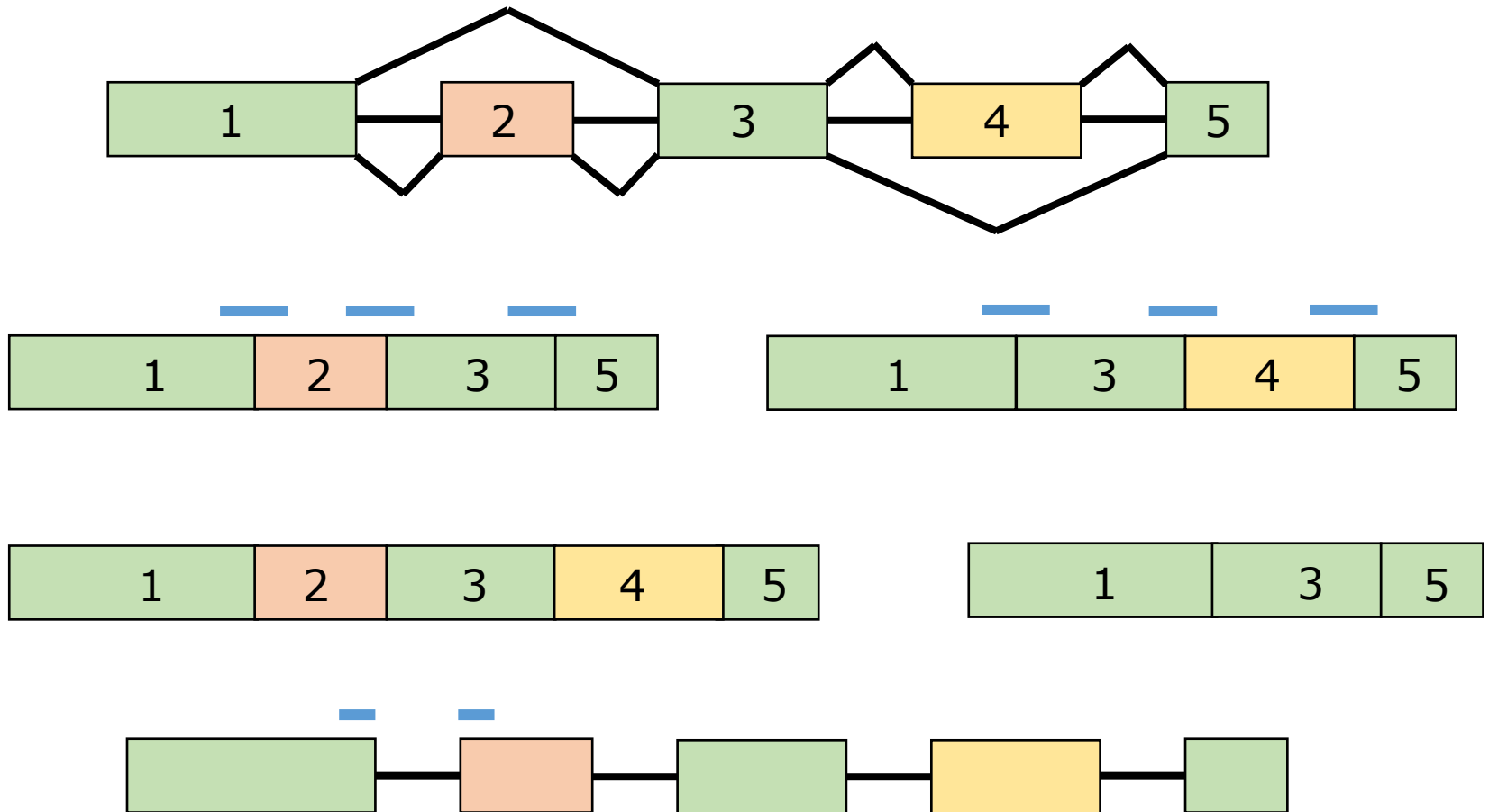




# Repeat expansion - project 2



# Transcriptome analysis

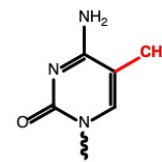
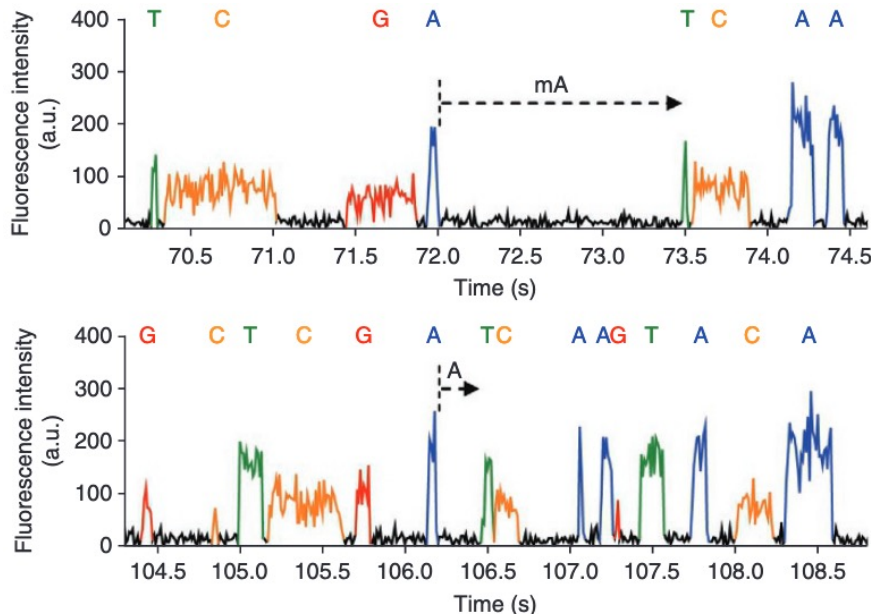


# Transcriptome analysis

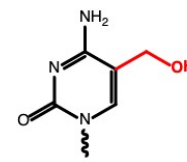
- Full transcripts: less ambiguous identification + quantification of transcript isoforms
- Better gene model prediction (e.g. for genome annotation)
- ONT direct RNA-seq:
  - No PCR amplification
  - No reverse transcription
  - Base modification identification

# Epigenetics

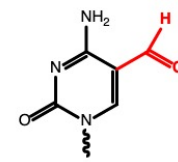
- Base modification in non-amplified libraries (Pacbio and ONT)
- No multiplexing
- Also possible in direct RNA-seq (ONT)



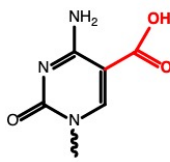
5-mC



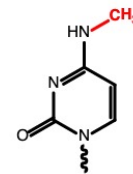
5-hmC



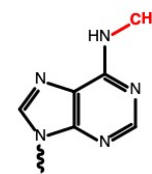
5-fC



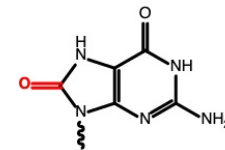
5-caC



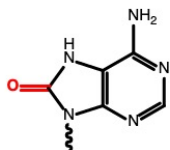
4-mC



6-mA



8-oxoG



8-oxoA

# Metagenomics

- Full length 16S genes
- Metagenome Assembled Genomes (MAGs)

# New possibilities

- Cas9 based targeting (both PacBio & ONT)
- Adaptive sampling (ONT)

