

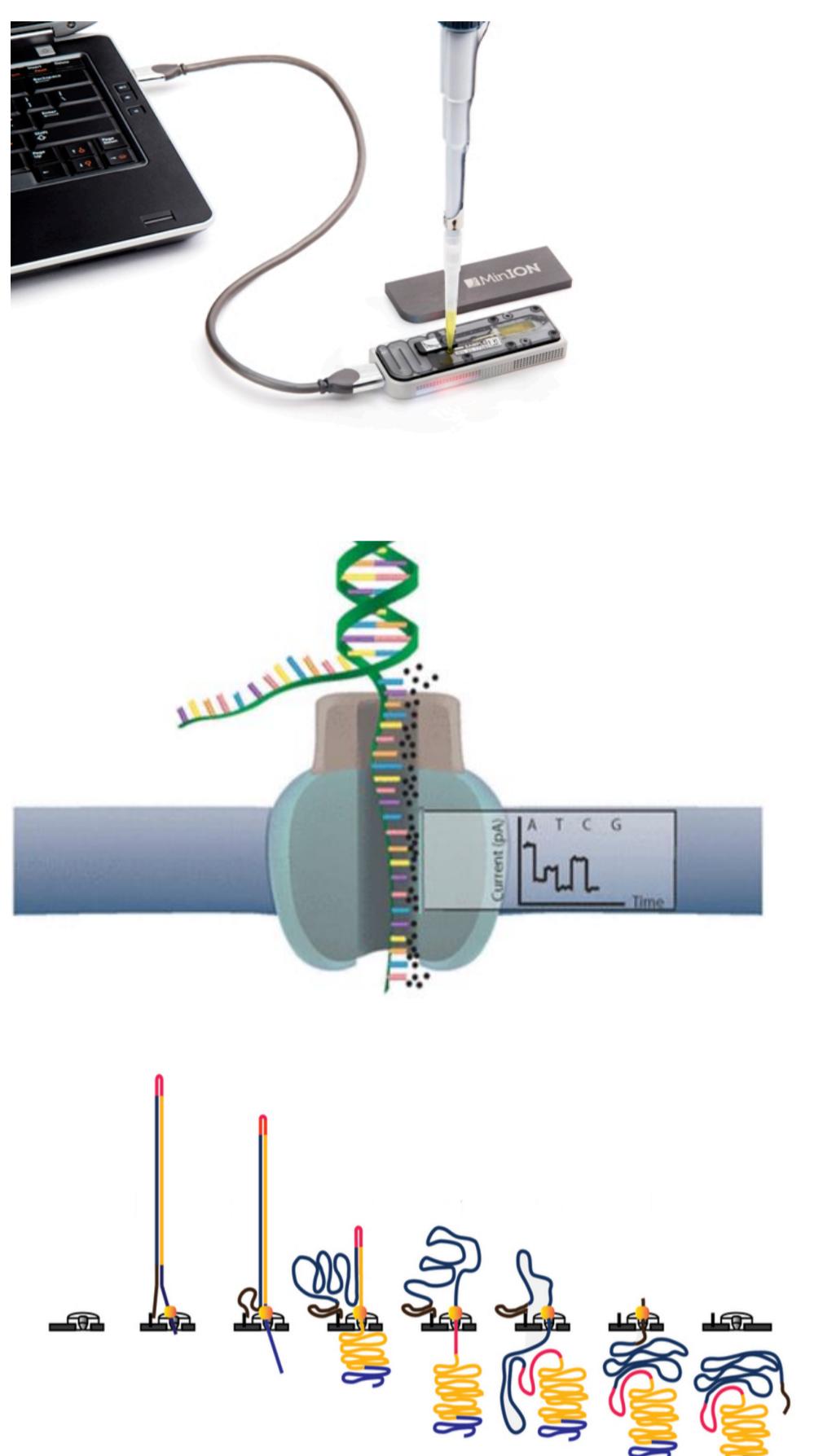
# MGMM - Nanopore Sequencing Mapper

## A new Bowtie-based Mapper for long reads with high error rate

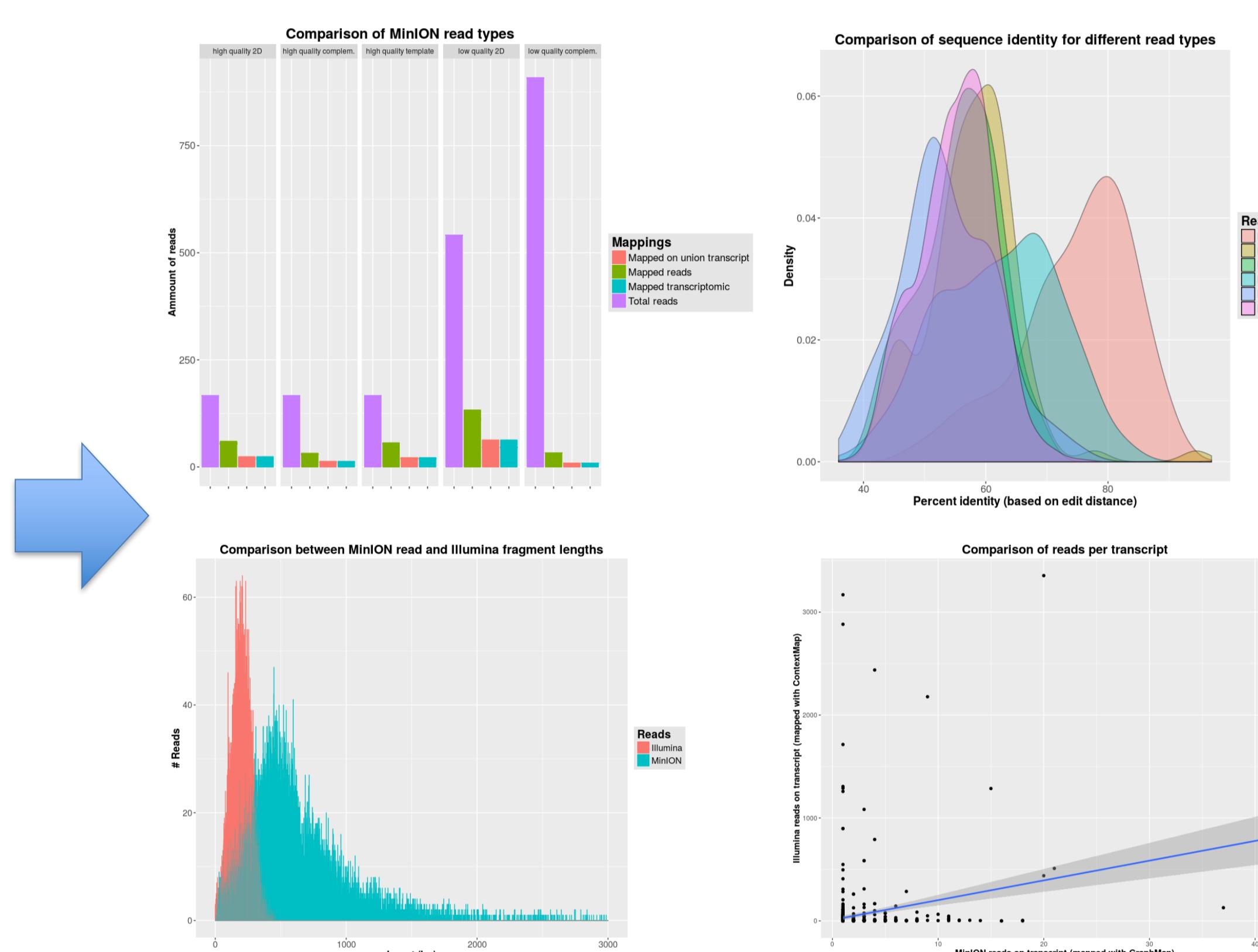
Ron Fechtner, Jonas Galli, Felix Offensperger, Florian Tichawa

Praktikum Genomorientierte Bioinformatik Ws 2016-2017

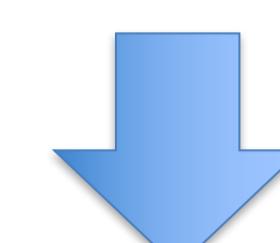
### Introduction in MinION reads and goal of the project



- Introduced Feb 2012 by Oxford Nanopore Technology (ONT)
- Real-time DNA, RNA / miRNA and Protein Sequencing
- Compact USB-Stick Format
- Low cost
- Fast library preparation
- A protein nanopore is set in an electrically-resistant polymer membrane
- Electric current generated through the nanopore when applying a potential across the membrane
- Passing molecule creates characteristic disruption in current, called the nanopore signal
- A passing molecule is composed of
  - Loading adaptor (Eases flow through pore)
  - Template strand
  - Hairpin adaptor (Connects strands)
  - Complement strand
  - Loading adaptor
- The information from both strands is used for the consensus sequence (2D read)



- Few, long reads
- High error rate



- Demand for new mapping approach

- Enable transcriptomic mapping

### Methods

#### 1. Step: Creation of seeds

- Read is too long for single mapping
- Detection of seed length with best results
- Seeding with fix shift and fix length (5 and \*\*)
- Template and Complement
- End-to-end mapping with "Bowtie" allowing one Mismatch

#### 2. Step: Creation of overlapping seed-chains

- Order of mapping could be wrong
- Unification of overlapping seeds when in shift range
- All seed-chains on the same chromosome were collected and written in a *fastM* file

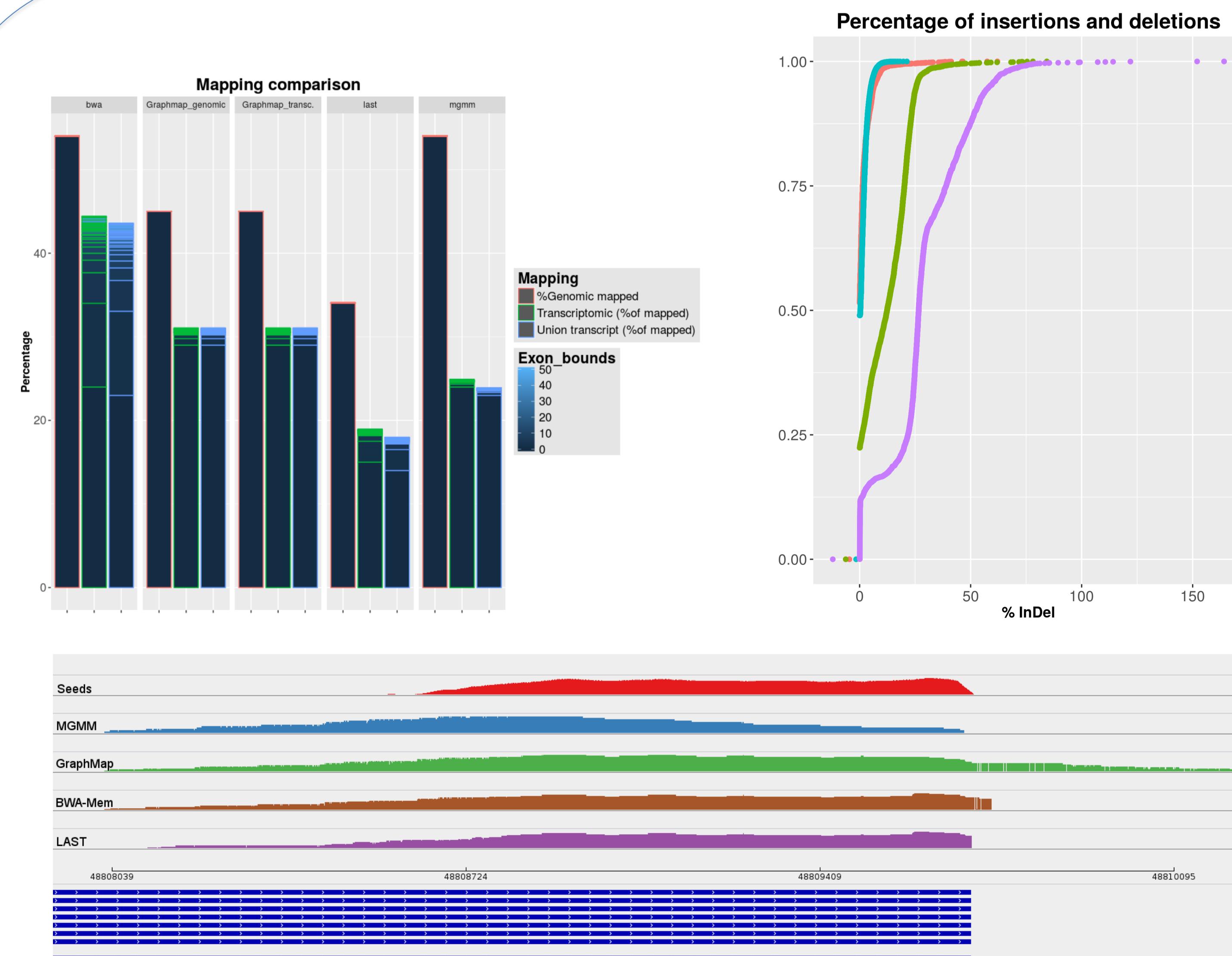
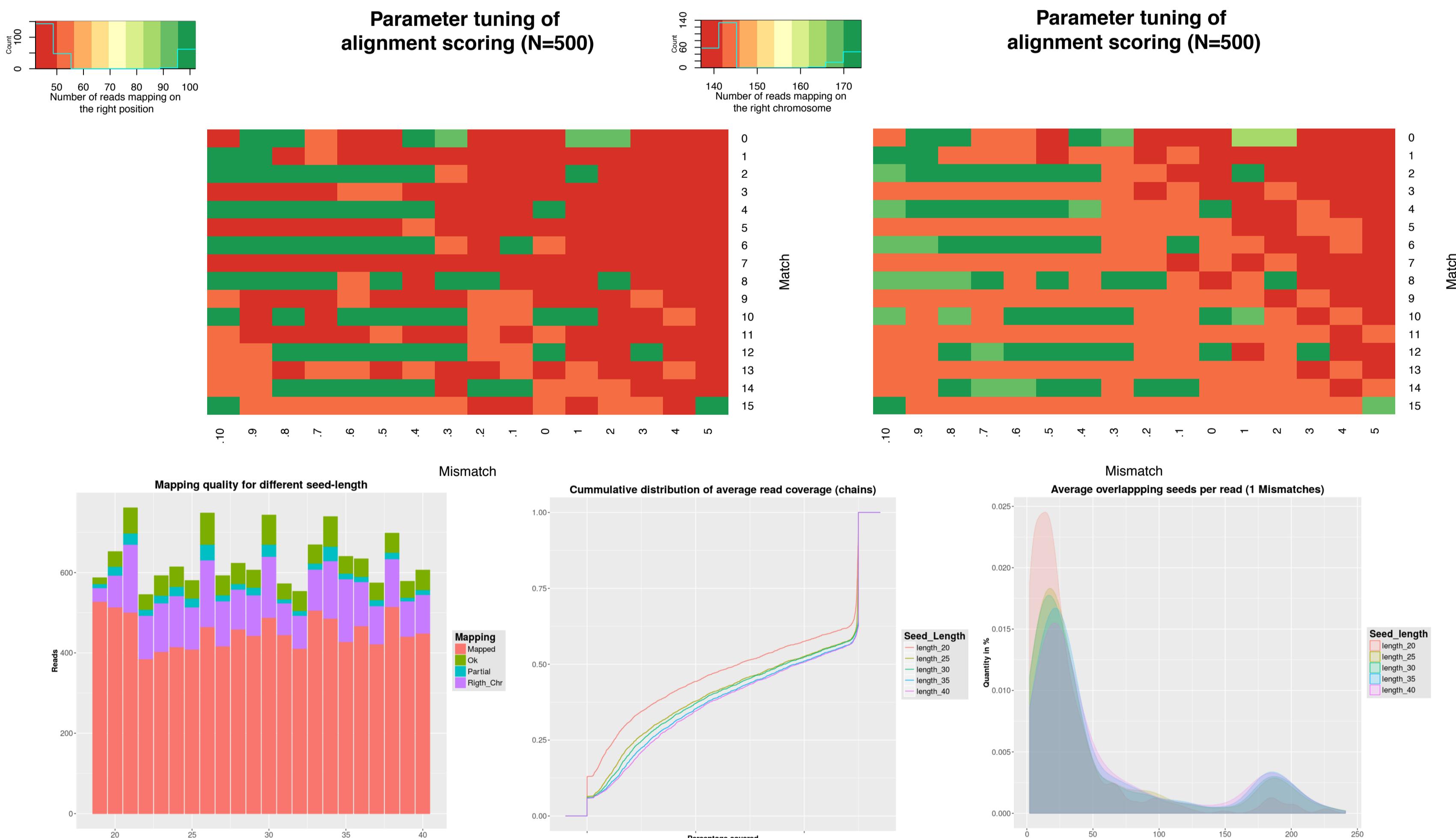
#### 3. Step: Pathing of missing read parts

- Gaps between seed-chains are too big for alignment
- Calculation of all possible exon / intron paths

#### 4. Step: Alignment

- Alignment of all previously detected paths
- Comparison of alignment scores and stitching of the best path combinations

### Detection of the right parameters



### Conclusion

- Seeding worked well, but filtering seemed too restrictive
- Parts without seeds could not be mapped
- Complex interaction between different parameters
- Writing runtime and memory efficient code is challenging