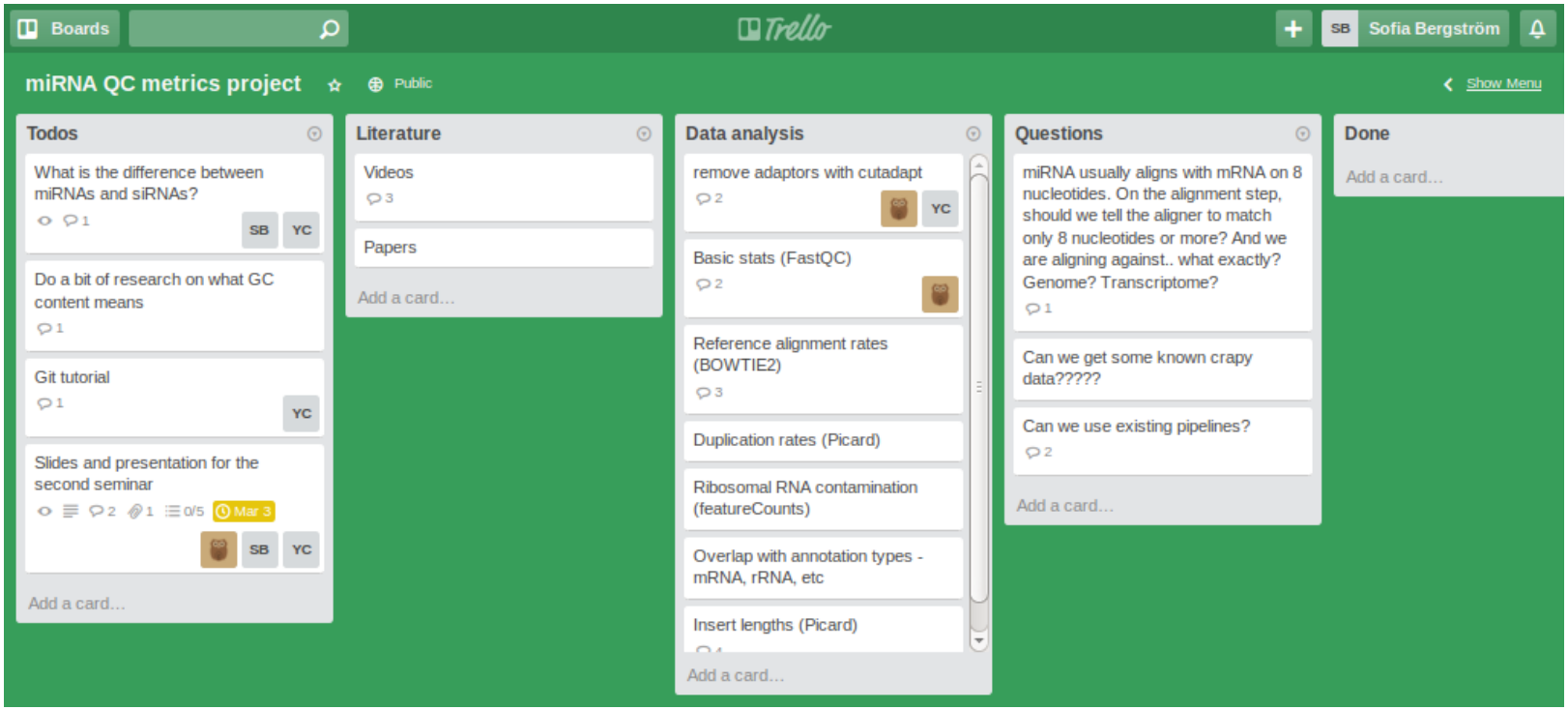
miRNA QC metrics

Guillermo Carrasco
Sofia Bergström
Yim Wing Chow

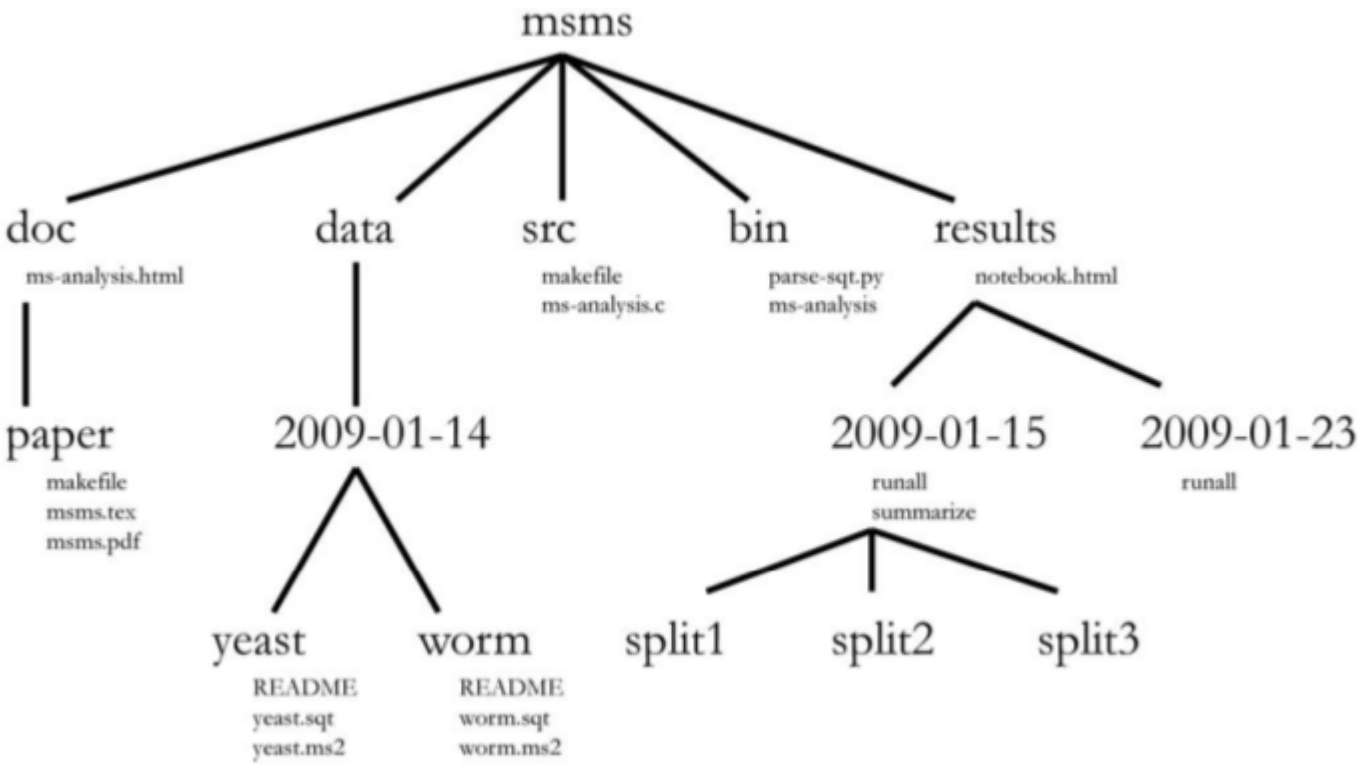
Project structure

Trello



Project structure

Github



Pipeline

https://github.com/SciLifeLab/miRNA_processing

- Remove adapters (**Cutadapt**)
- Basic stats (**FastQC**)
- Read alignment (**Bowtie**)
- Duplication rates (**Picard**)
- rRNA contamination (**featureCounts**)
- Overlap with annotation types; mRNA, rRNA
- Insert and alignment length (**Picard**)

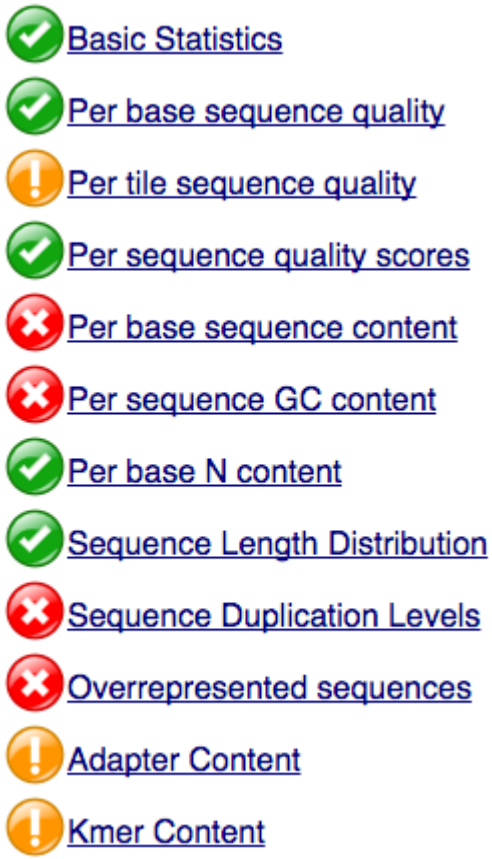
Pipeline

To be done:

- Automated search against miRBase
- Make it more easy to test different parameters
  - Options in configuration file?

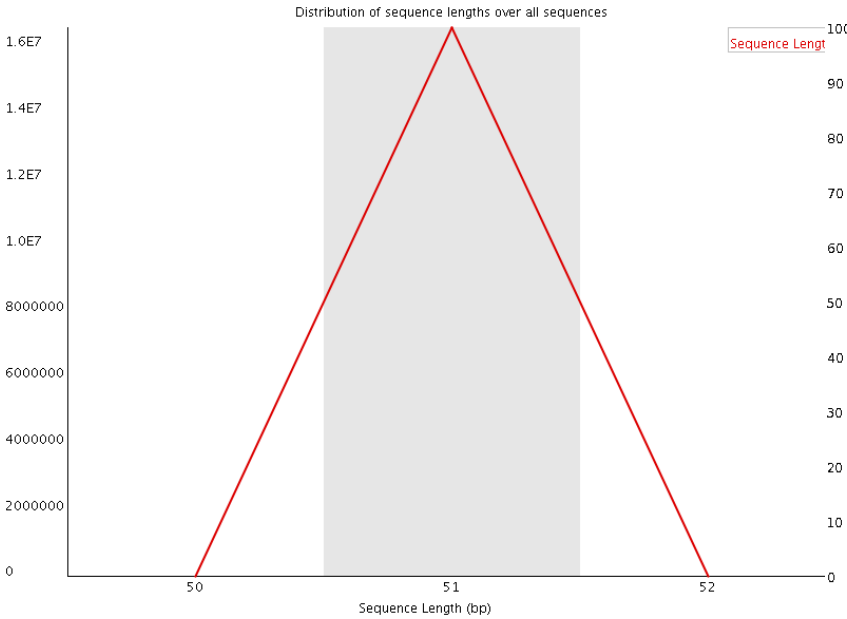
FastQC

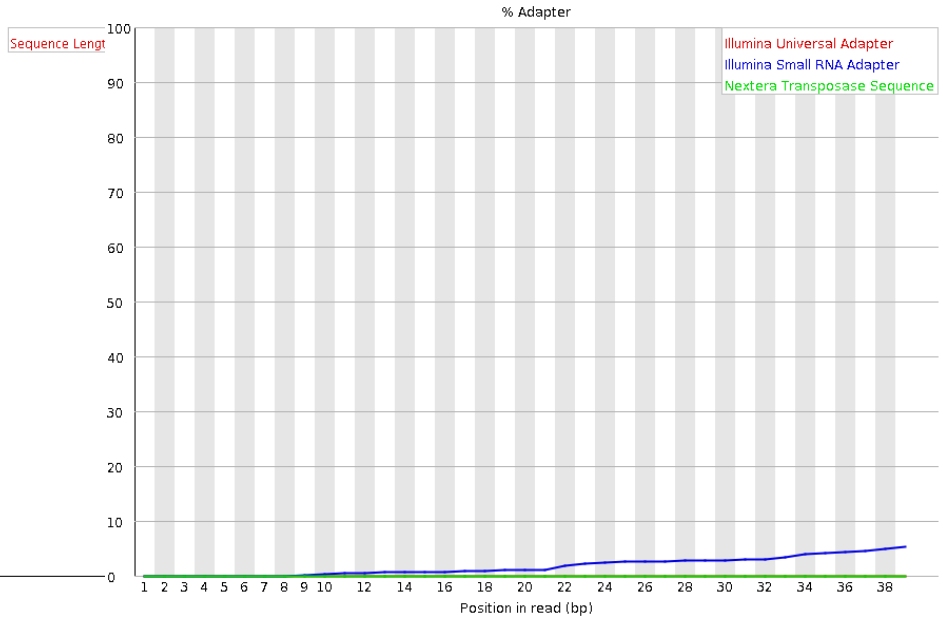
- FastQC is a quality control tool for NGS data



FastQC

First run on one sample. No adapter trimming



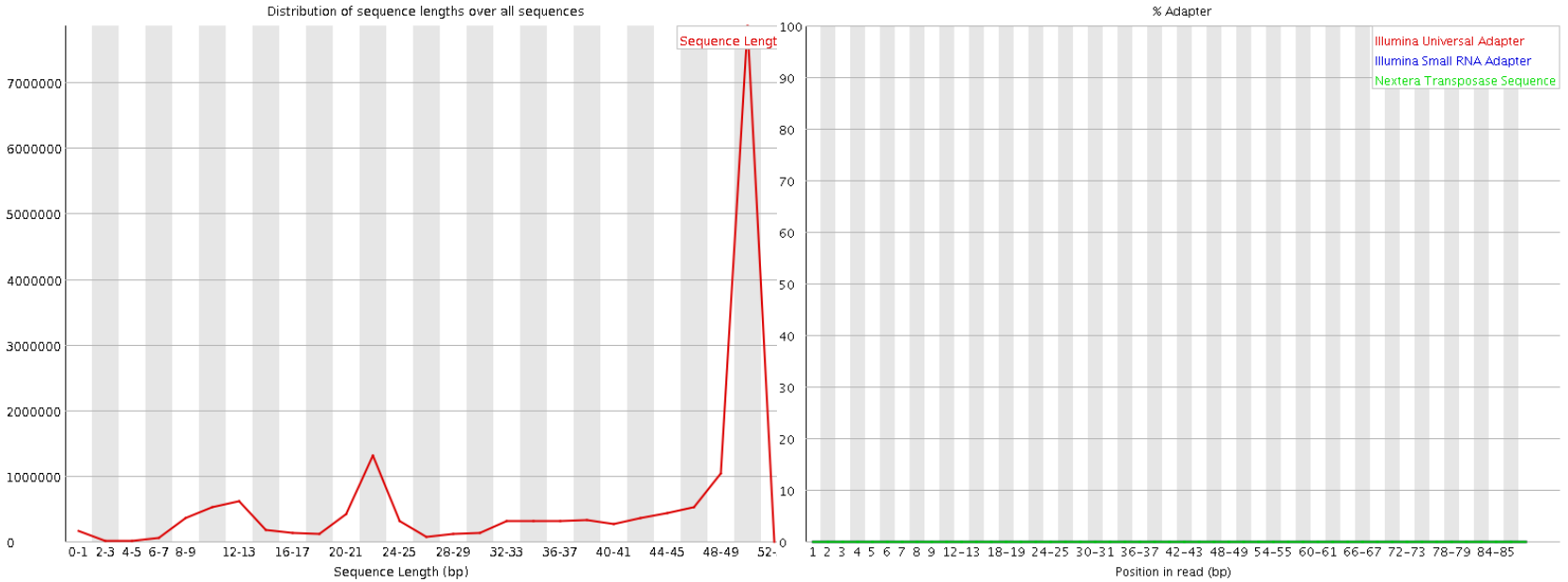


Length distribution

Adapter content

FastQC

First run on one sample. With adapter trimming



Length distribution

Adapter content

Project plan

- Week 10:
  - Small pipeline fixes and improvements
  - Continue testing/understanding tools
  - Try to run the analysis in all the samples
- Week 11:
  - Study and compare results
  - Try to find reasonable thresholds for different QC metrics
  - Poster preparation
- Week 12:
  - Conclusions
  - Finish the poster and presentation

Thank you!

Questions?

Guillermo’s diary

Yim’s diary

Sofia’s diary