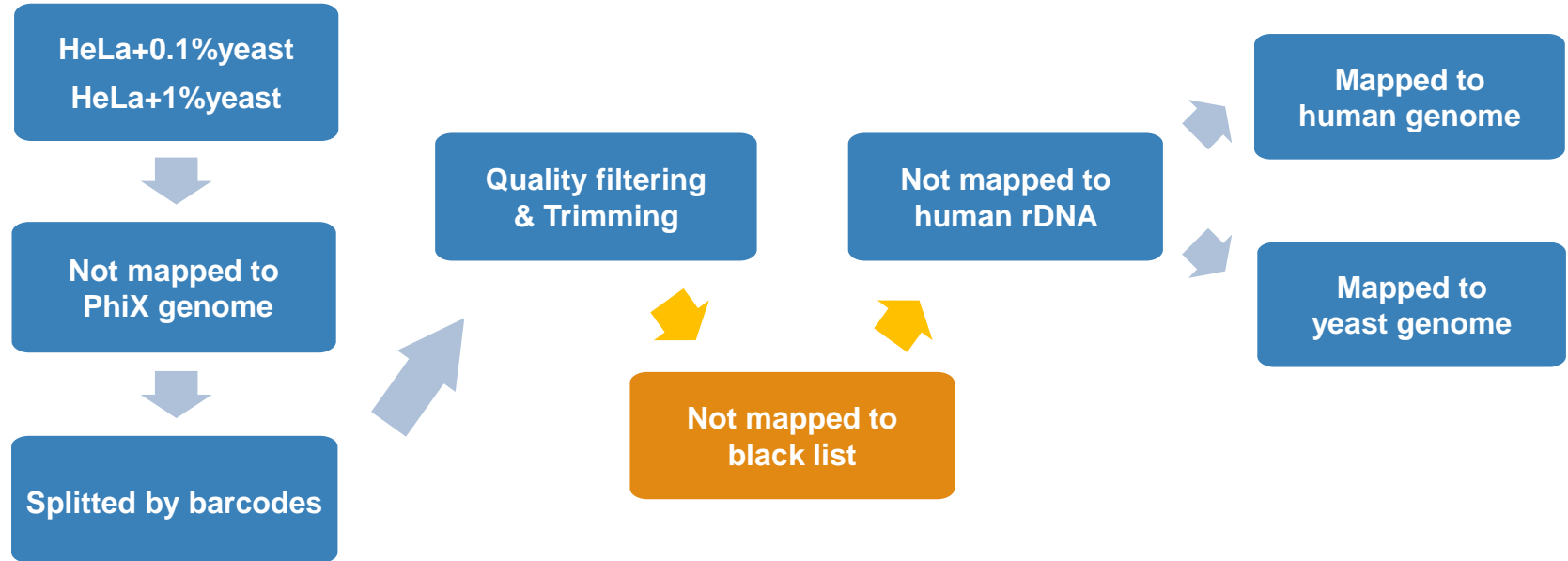


CAGE data analysis of HeLa-Yeast spike in

Enrichetta, Marco and Yunzhang
Supervisor: O. Emanuelsson

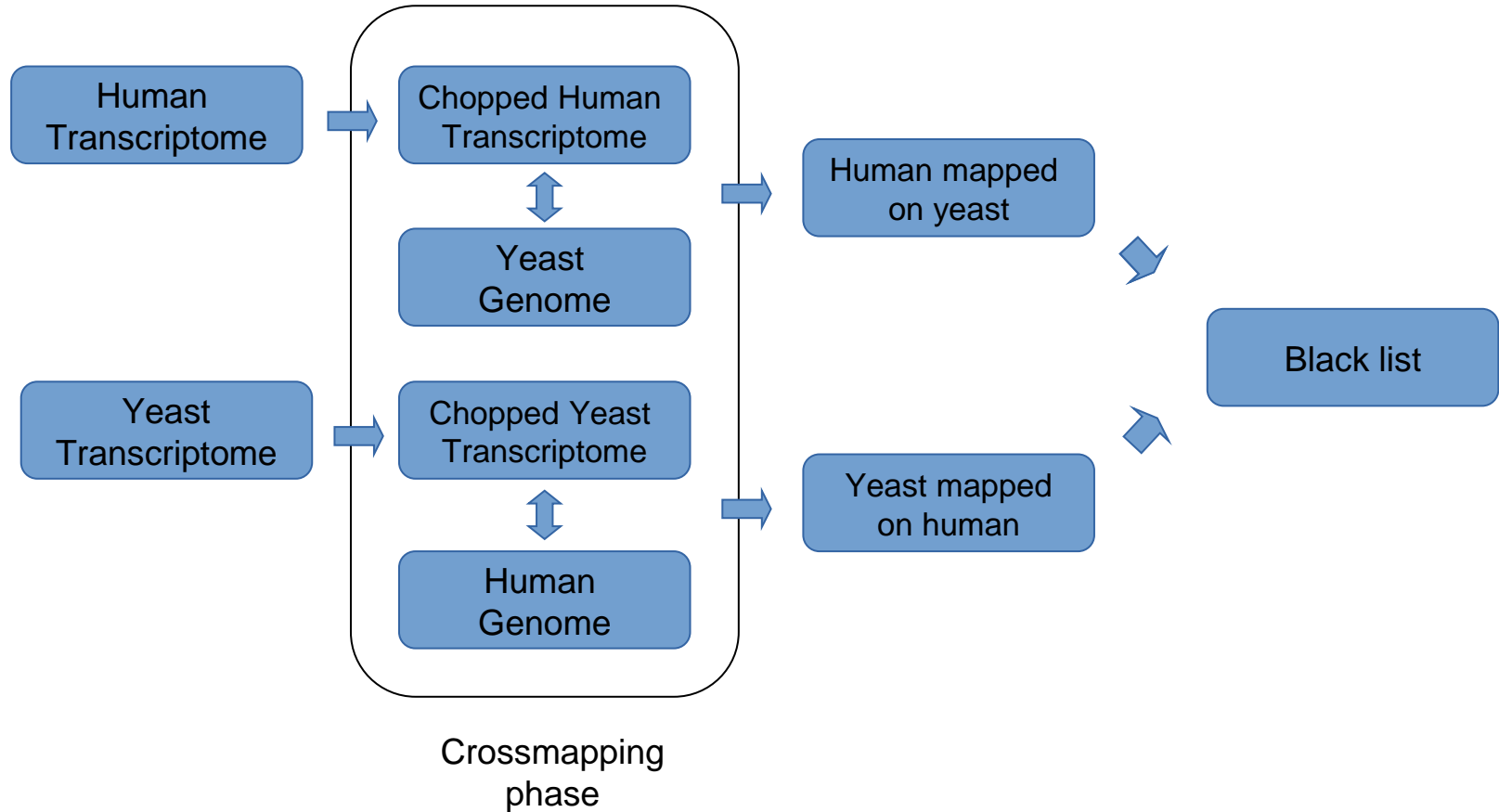
Data processing



Function of black list

Mapped to Human	Mapped to Yeast	Interpretation
TRUE	FALSE	Human transcripts
FALSE	TRUE	Yeast transcripts
TRUE	TRUE	Black List

Creating Human/Yeast black list



Problems

- When we run in upppmax, the huge amount of human fragmented sequences, that we created for the black list. We solved this splitting the job in sub-jobs.
- Some reads are not mapped to neither human nor yeast genome. They could be artefacts and are removed.

Best idea

- Create a black list of those regions in yeast and human that mapped in both and we will be able to remove this in order to increase the quality of the analysis.

Changes of plan

- We realize the possibility to have region mapped in both yeast and human, and since then we will not be able to define from which organism belongs, we are trying to remove this region making a black list.

Thank you for listening!