

Prostate Cancer Heterogeneity and Assorted Chi-Squared Tests

Erle Holgersen

Department of Statistics and Actuarial Science
University of Waterloo

March 28th, 2017

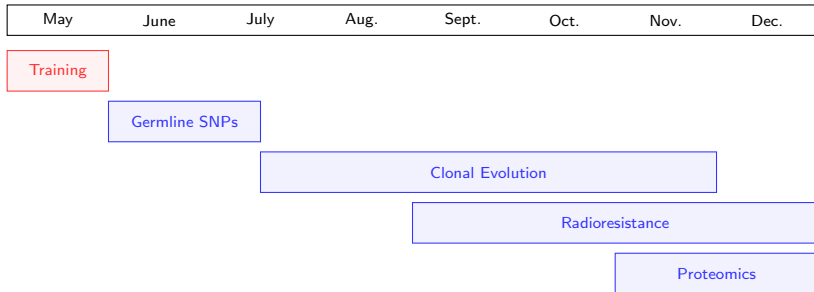
Introduction

- Worked in the Boutros Lab at OICR.
- Large bioinformatics lab, approx. 35 members.
- Mostly work on developing biomarkers for prostate cancer.
- Several different kinds of genomics data.
 - Copy number, SNVs, RNA, protein.



Projects

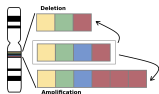
- Can broadly categorize my work into:
 - Data processing.
 - Statistical analyses.
 - Ad hoc statistical consulting.
- Worked on four projects.
 - They all involved prostate cancer, otherwise very different.



Radioresistance Project

- Samples from 11 radioresistant prostate tumours.
- Wanted to characterize how they differed from prostate tumours that have never been exposed to radiation.
- Different types of data:
 - copy number
 - methylation
 - miRNA

COPY NUMBER



METHYLATION



miRNA



Workflow

GENOMIC DATA

quality control → pre-processing → analysis → visualization

Workflow

COPY NUMBER

analysis → visualization

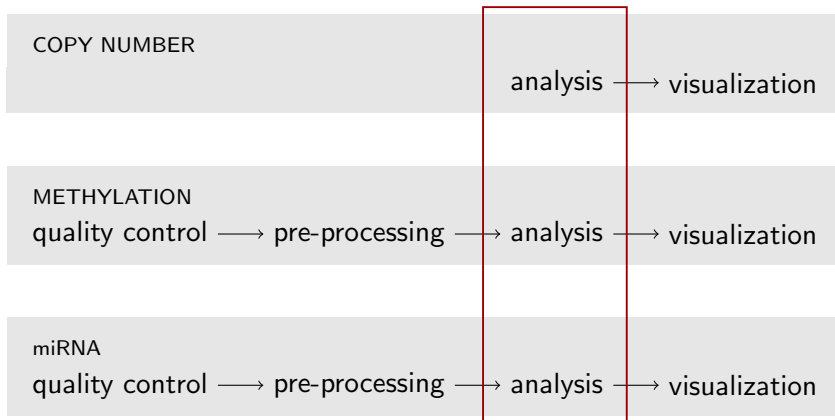
METHYLATION

quality control → pre-processing → analysis → visualization

miRNA

quality control → pre-processing → analysis → visualization

Workflow



Statistical Considerations

- Patients who suffered recurrence had more aggressive tumours to start with.
 - Wanted to adjust for tumour size and grade.
 - Not enough power to adjust for all relevant variables.
 - In some cases, no comparable variables.

Challenges

- Imperfect data.
 - Not everything can be fixed with more data cleaning.
 - At some point you have to accept it and decide to trust your downstream analyses anyways.

Challenges

- Imperfect data.
 - Not everything can be fixed with more data cleaning.
 - At some point you have to accept it and decide to trust your downstream analyses anyways.
- Can't do statistics in a vacuum.
 - Need to understand research questions to come up with sensible approaches.
 - Working on collaborative projects helped a lot.

Conclusion

- Learned way of thinking in school, picked up specific statistical techniques on the job.
 - Statistical tests often too specific to be taught in class.
- Learned to appreciate storytelling aspect of statistics.
 - Results have to be clear, otherwise they won't make it into the paper.
- Got much better at working with large codebases.