LEIDEN UNIVERSITY MEDICAL CENTER

Next generation sequence data analysis

Jeroen F. J. Laros

Leiden Genome Technology Center

Department of Human Genetics

Center for Human and Clinical Genetics

Introduction

NGS data analysis

In this practical, we will do two types of analyses:

- Exome sequencing data analysis.
- CAGE expression analysis.

Introduction

NGS data analysis

In this practical, we will do two types of analyses:

- Exome sequencing data analysis.
- CAGE expression analysis.

Cap analysis gene expression (CAGE):

- Isolate mRNA.
- Extract a small part of the beginning of the transcript.
- Sequence.
- Analysis consist mainly of counting mapped sequences in a specific region.

Some notes

Region files

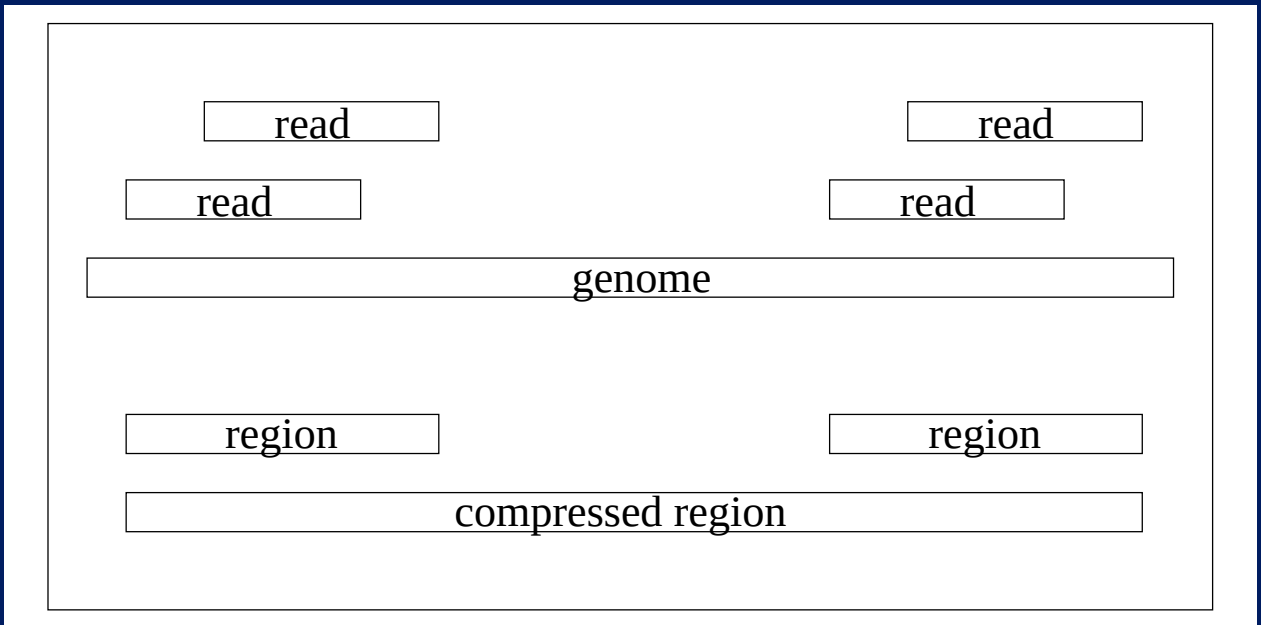

Figure 1 : Region files explained.

Terminology:

- Overlapping reads are combined to a “region”.
- Close by regions are combined to a “compressed region”.

Galaxy

Galaxy

We will use the Galaxy graphical user interface:

- Wrapper for command line utilities.
- User friendly.
- Point and click.
- Workflows.
  - Save all the steps you did in your analysis.
  - Rerun the entire analysis on a new dataset.
  - Share your workflow with other people.
  - ...

http://galaxy.psu.edu

Galaxy

Galaxy

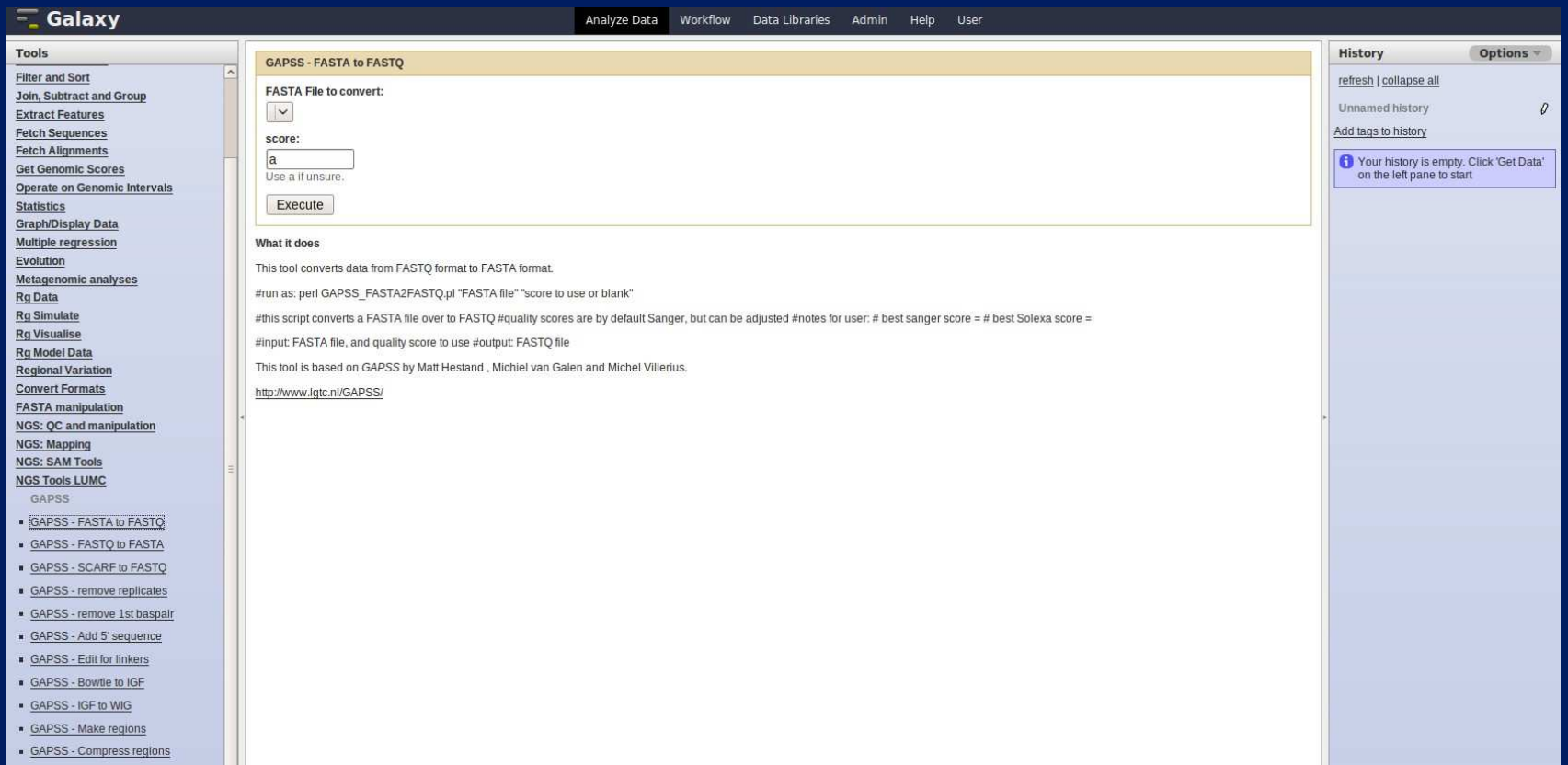

Figure 2 : Galaxy main user interface

http://galaxy.nbic.nl

Galaxy

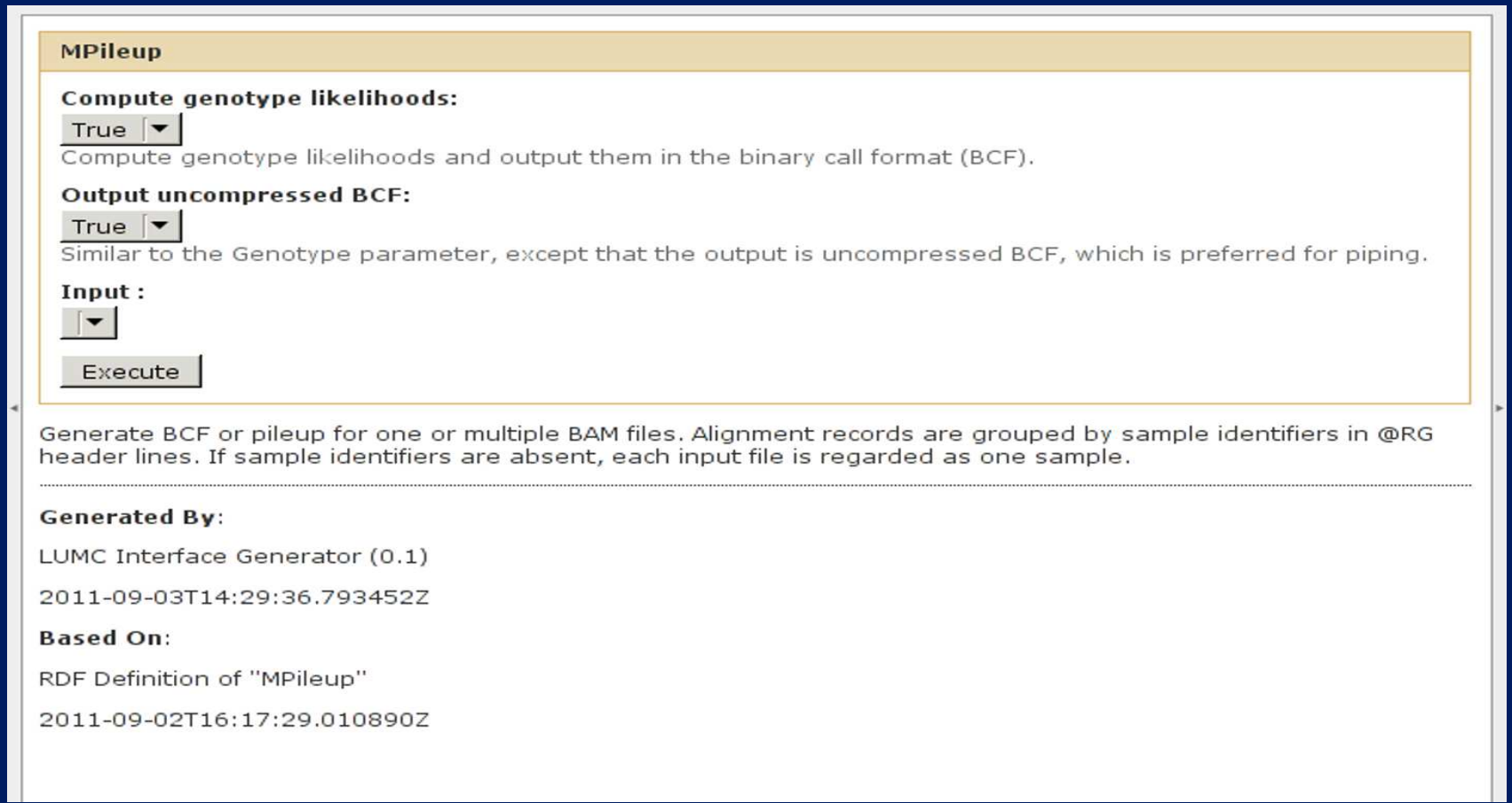

Figure 3 : User friendly interface with Galaxy

Galaxy

Galaxy icons

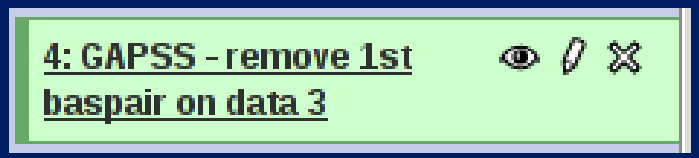

Figure 4 : Default view.

The top icons:

- Eye: view.
- Pencil: edit (rename).
- Cross: delete.

Click on the title for a more detailed view.

Galaxy

Galaxy icons

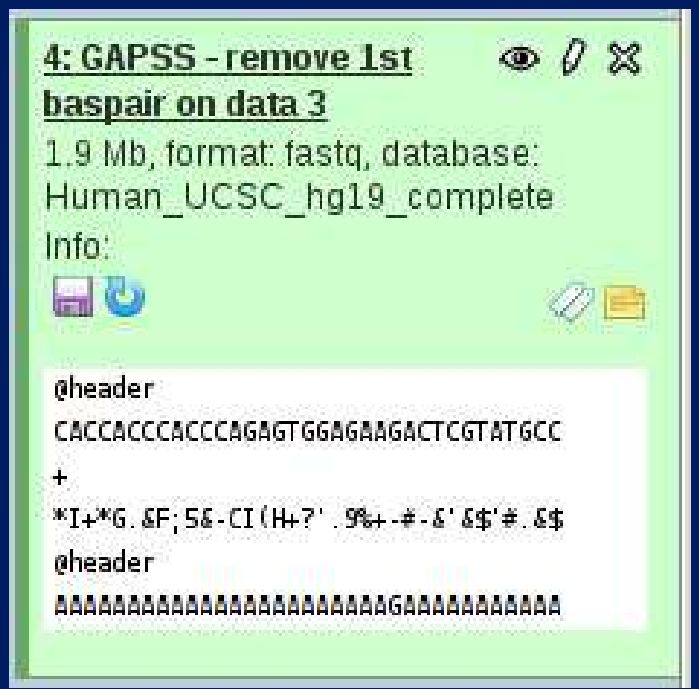

Figure 5 : Detailed view.

- Diskette: save.
- Blue looping arrow: rerun.

Galaxy

Outline of the practical

1. Create a SNP file from a FASTQ file.
2. CAGE analysis.
3. Workflows.
    - Rerun the SNP or CAGE analysis with no effort.

Questions?

See the handouts for the practical.

https://humgenprojects.lumc.nl/trac/humgenprojects/wiki/NGS_workshop