



LEIDEN UNIVERSITY MEDICAL CENTER

# **GAPSS and Galaxy**

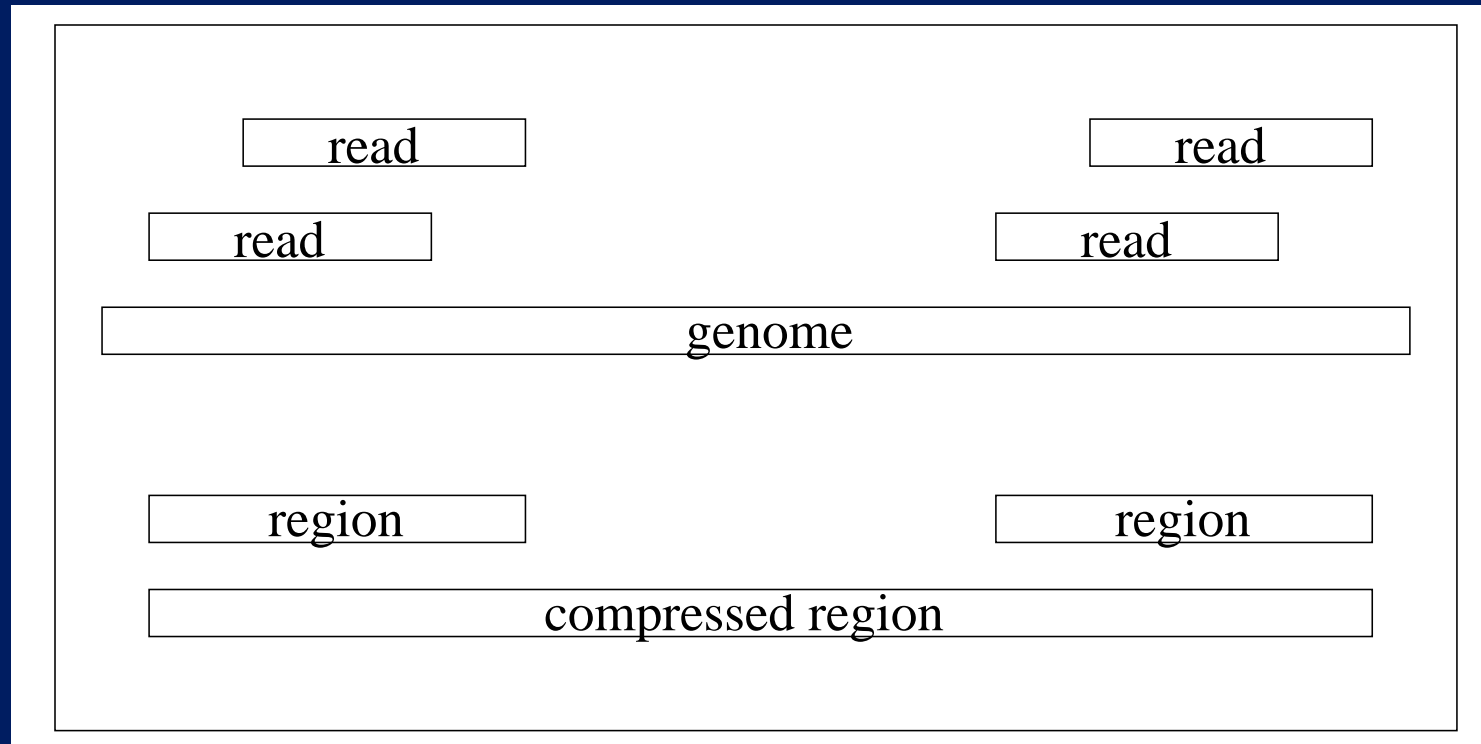
**Michiel van Galen and Jeroen Laros**  
**Department of Human Genetics**  
**Center for Human and Clinical Genetics**



## General Analysis Pipeline for Second-Generation Sequencers.

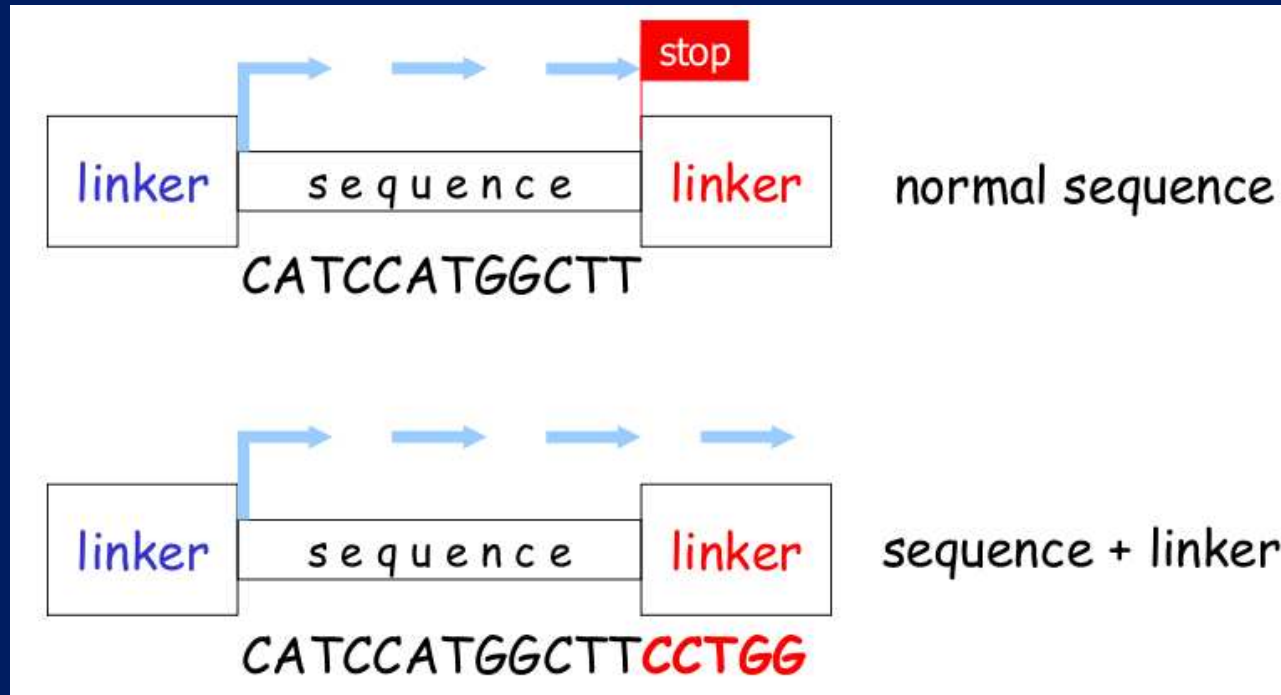
- Uses FASTA, FASTQ or scarf files as input.
- Generates a report file.
  - Number of tags in the input.
  - Number of tags aligned.
  - ...
- Generates UCSC genome browser tracks.
- Generates files with annotation of regionally clustered tags.
- Generates a SNP report.
- ...

# Region files



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

## Removing linkers



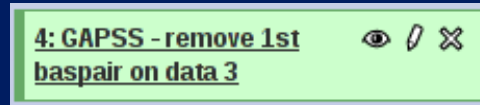
- A linker sequence can be found if there were more sequencing cycles than the length of the sequence.
- Linkers must be removed to align properly.

- 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>

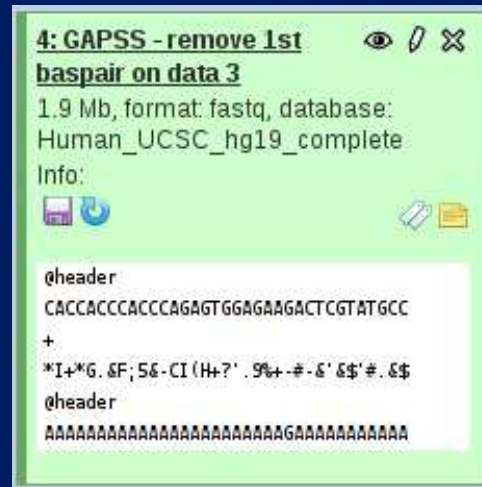
<http://galaxy.nbic.nl/>

## Galaxy icons



- Eye: view.
- Pencil: edit (rename).
- Cross: delete.
  
- Click on the title for a more detailed view.

# Galaxy icons



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

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



See the handouts for the practical.