



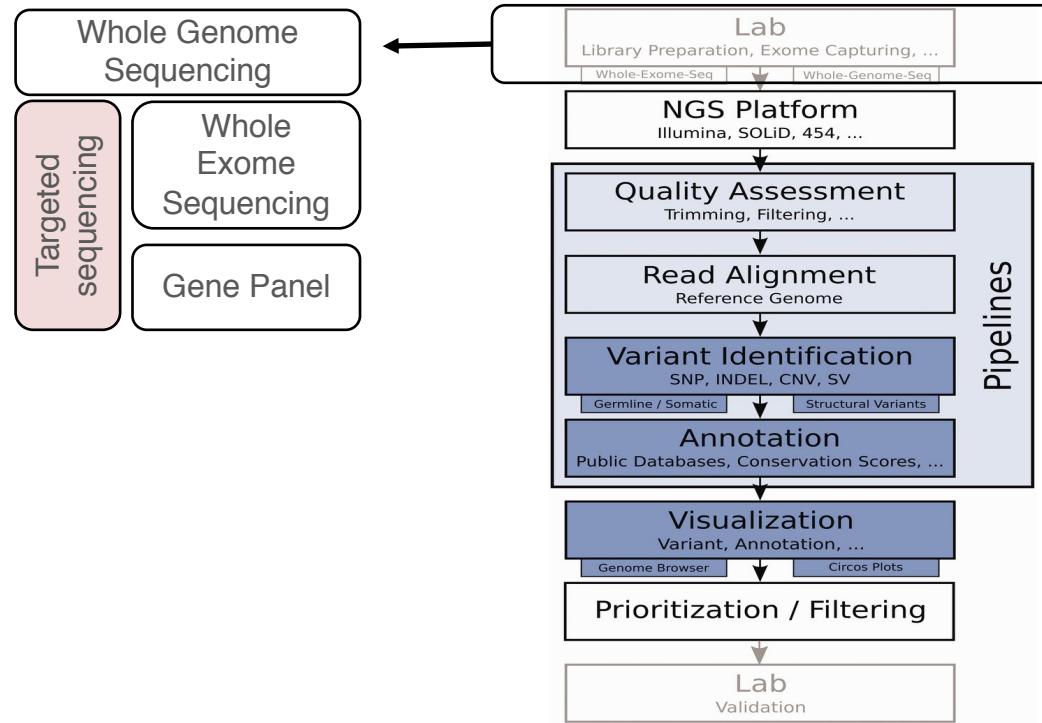
Introductory class: analysis workflow

Precision Oncology Course

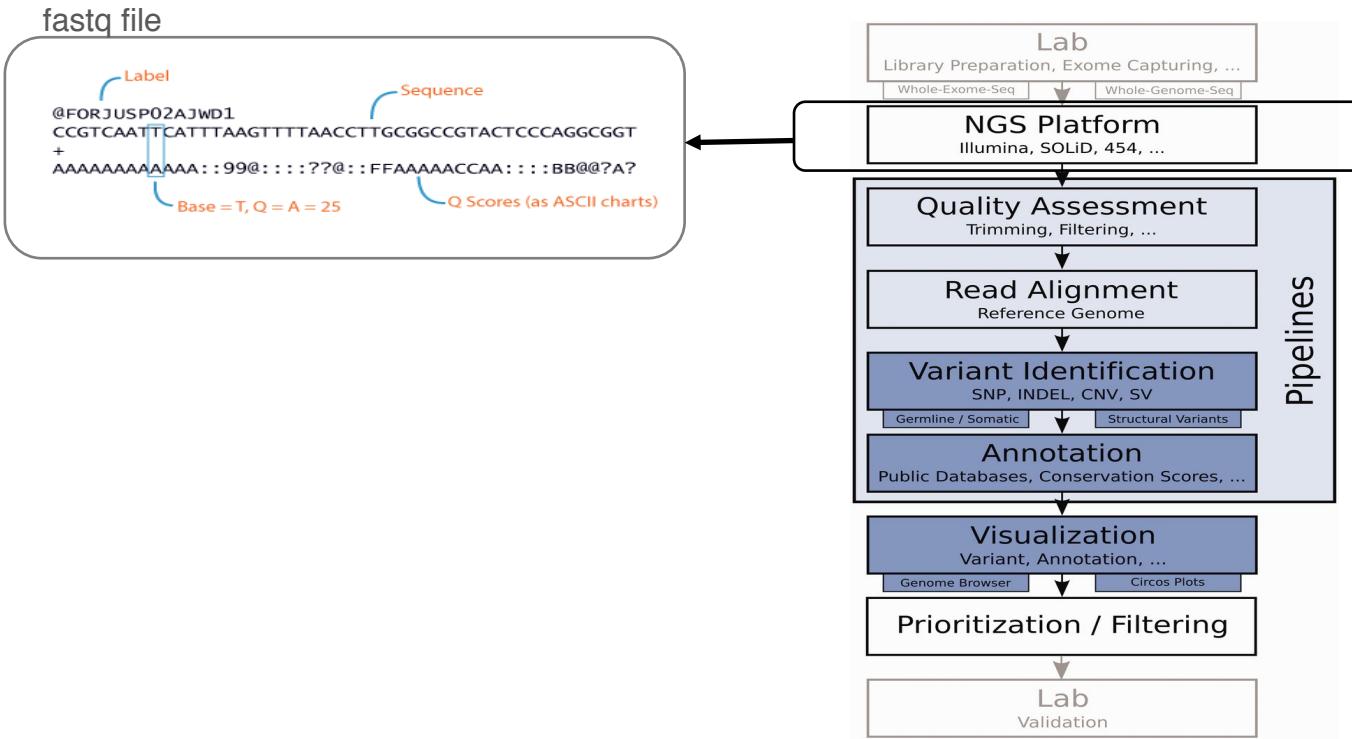
<https://cniobu.github.io/po/>

Analysis outline

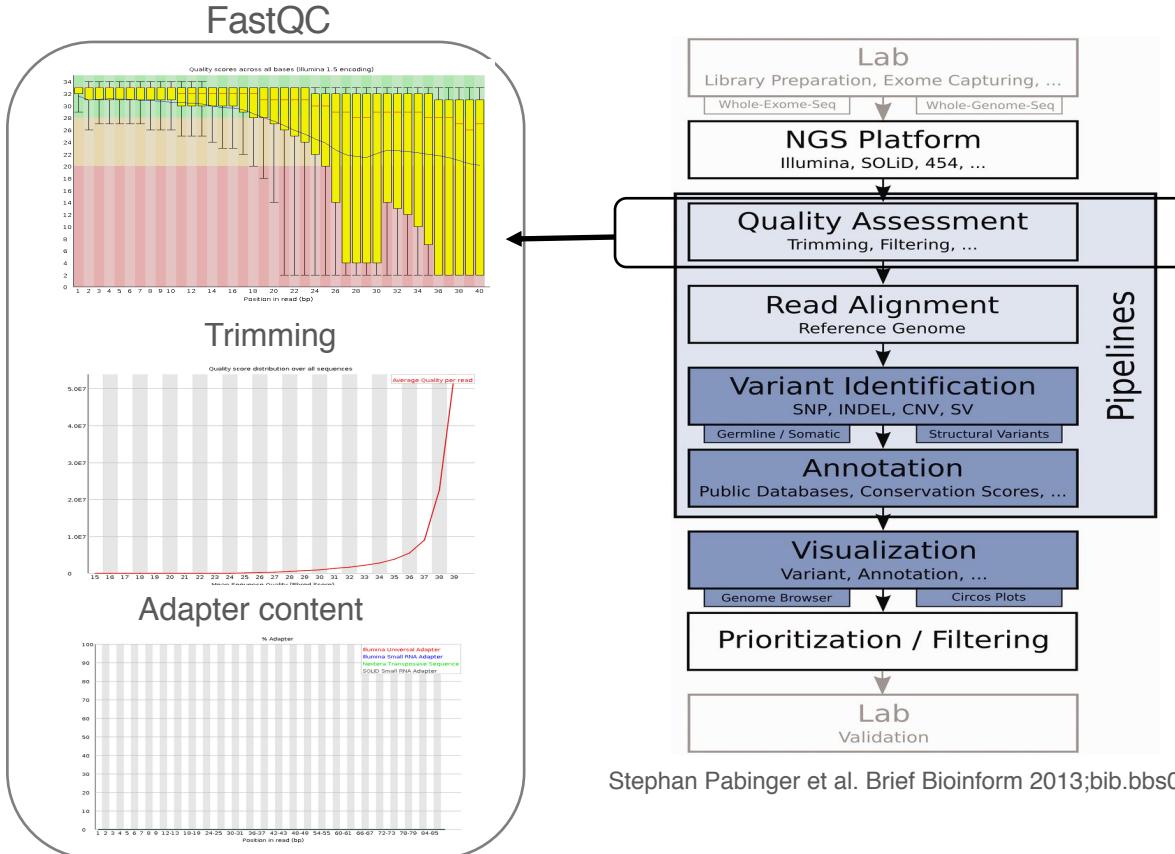
<https://cniobu.github.io/po/>



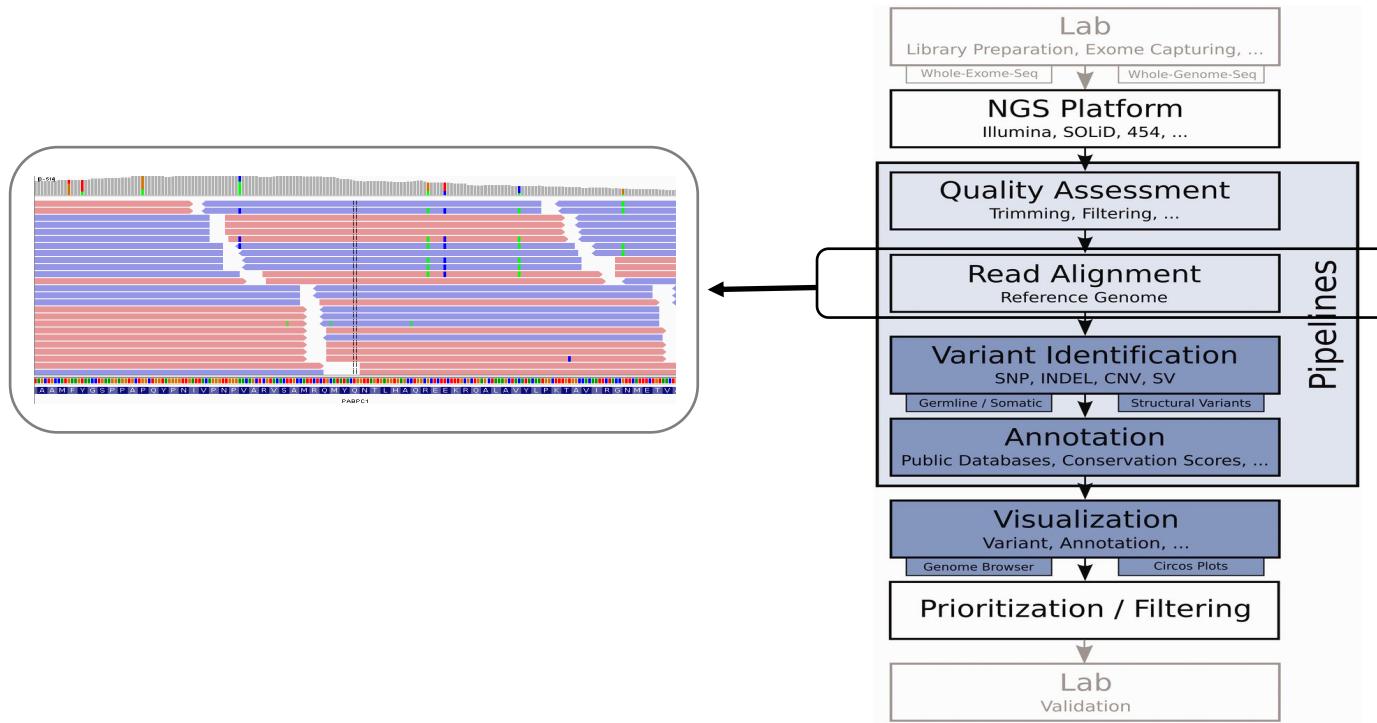
Analysis outline



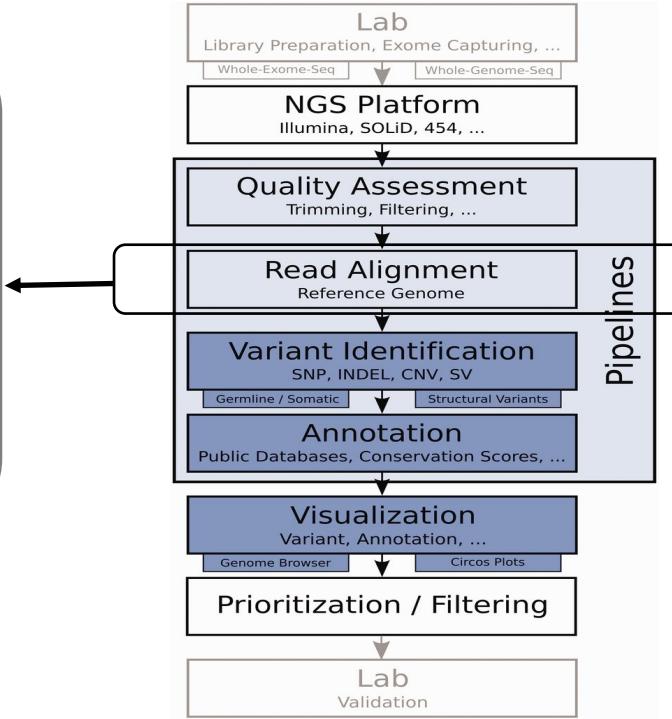
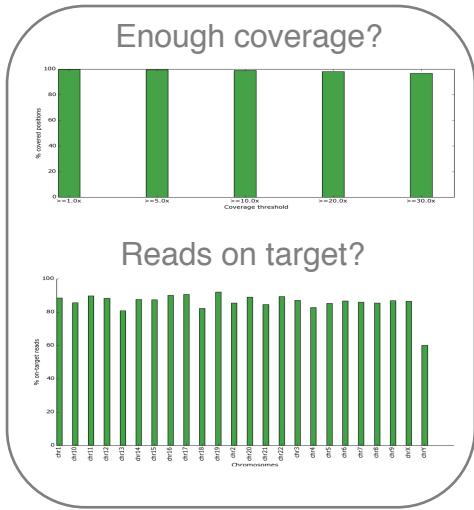
Analysis outline



Analysis outline



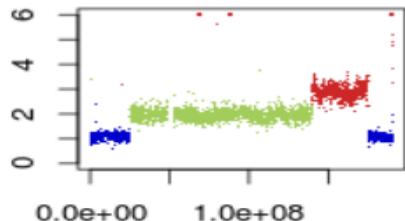
Analysis outline



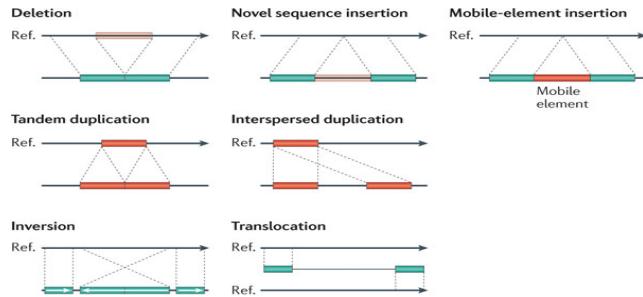
Analysis outline

Large-scale mutations

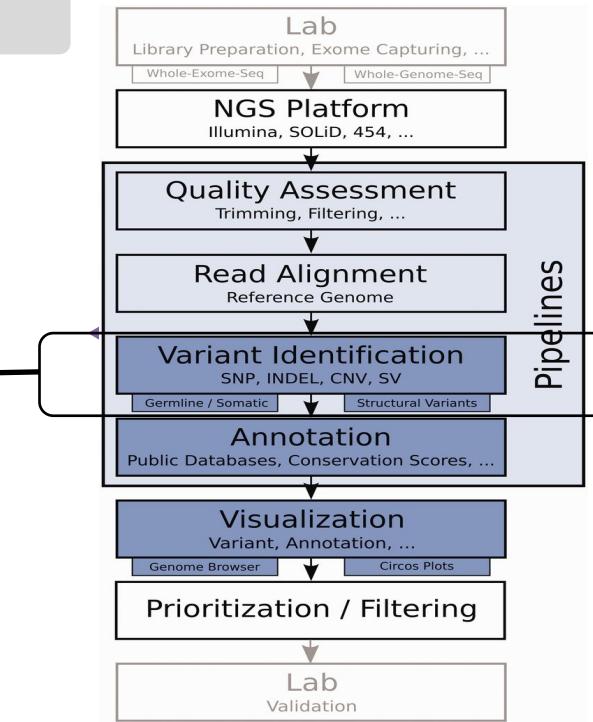
Copy Number Variants (CNV)



Structural Variation



Day 2



Stephan Pabinger et al. Brief Bioinform 2013;bib.bbs086

Analysis outline

Small-scale mutations

Single Nucleotide Variant (SNV)

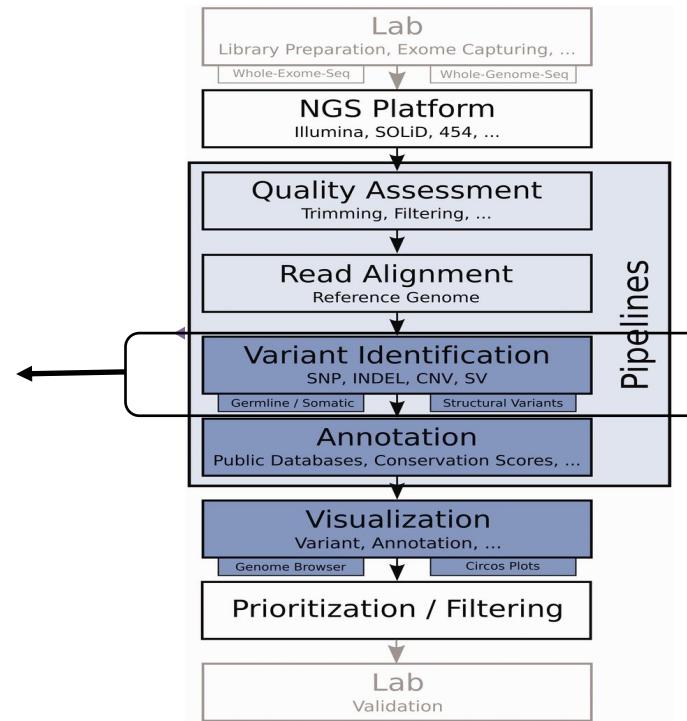
Reference: AACGGCCTGTAAC
Alternative: AACGGCCAGTAAC

Insertion

Reference: AACGGCCTGTAAC
Alternative: AACGGCCAGCTAAC

Deletion

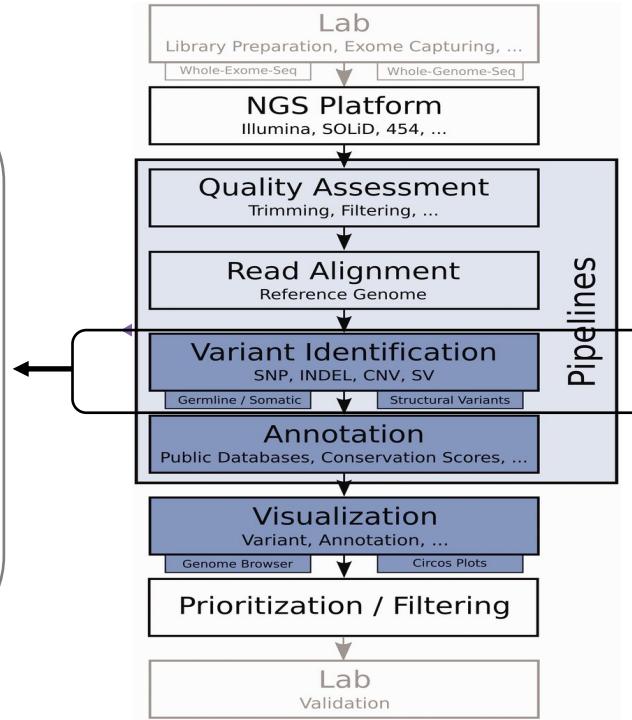
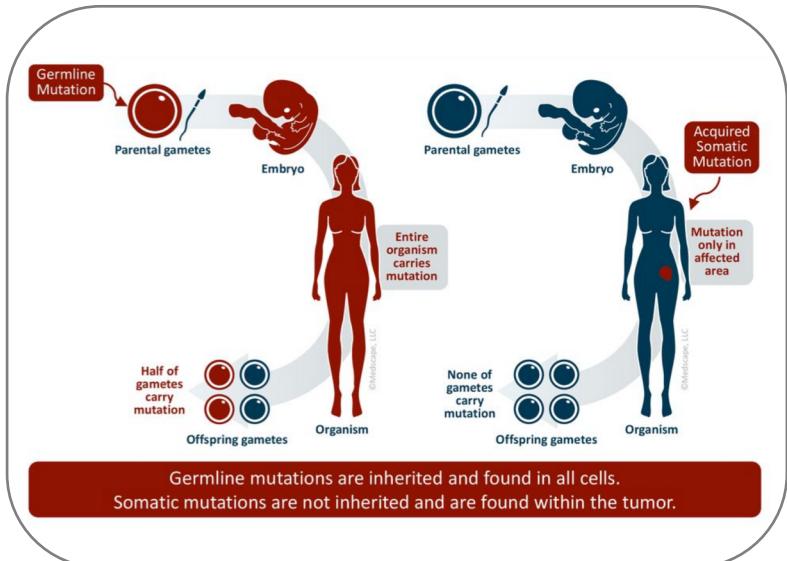
Reference: AACGGCCTGTAAC
Alternative: AACGGCC-GTAAC



Analysis outline

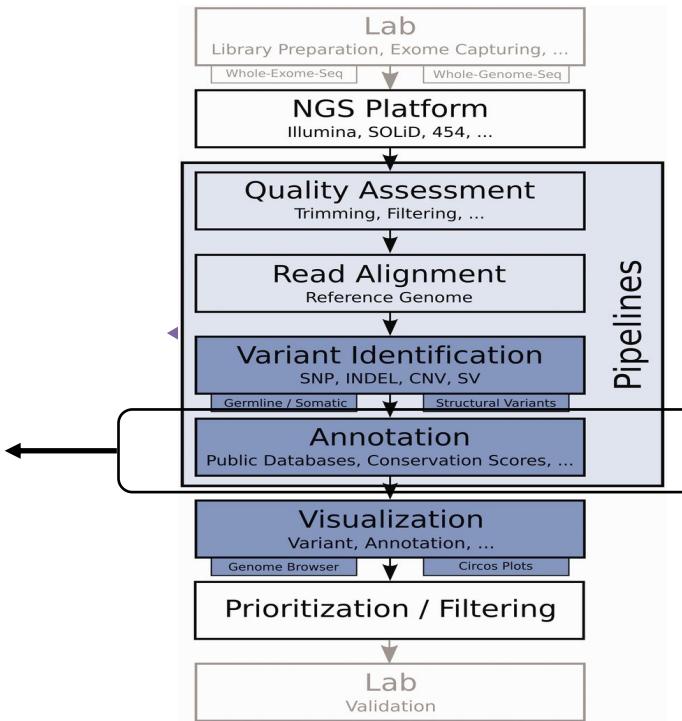
Day 2

Variant calling Whole exome/Panel analysis



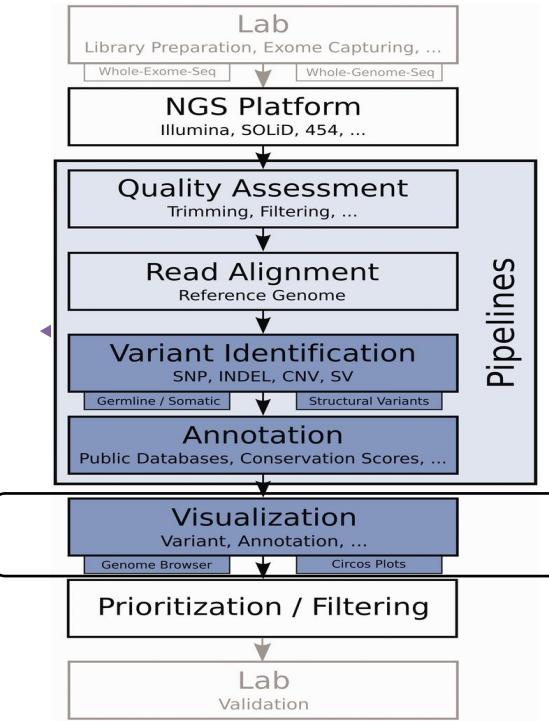
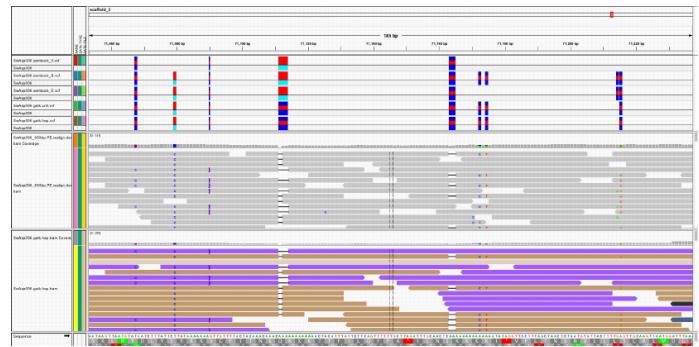
Stephan Pabinger et al. Brief Bioinform 2013;bib.bbs086

Analysis outline



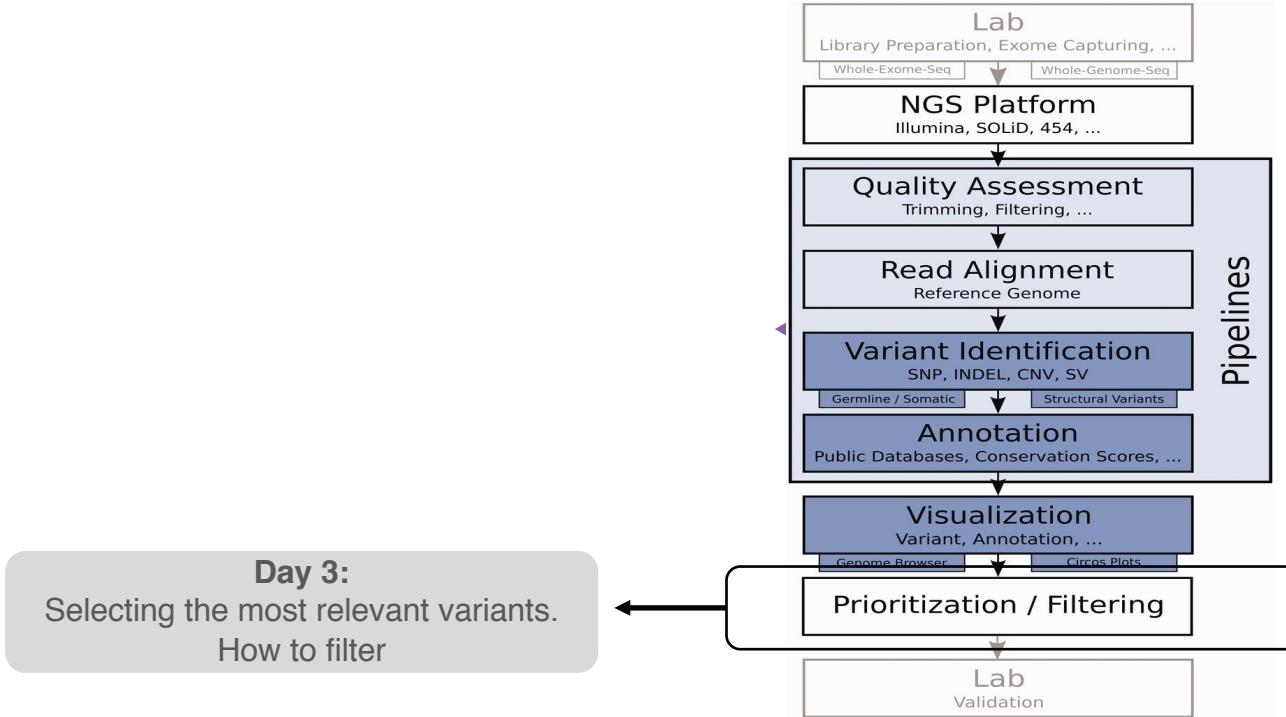
Analysis outline

Day 2 Variant calling Whole exome/Panel analysis

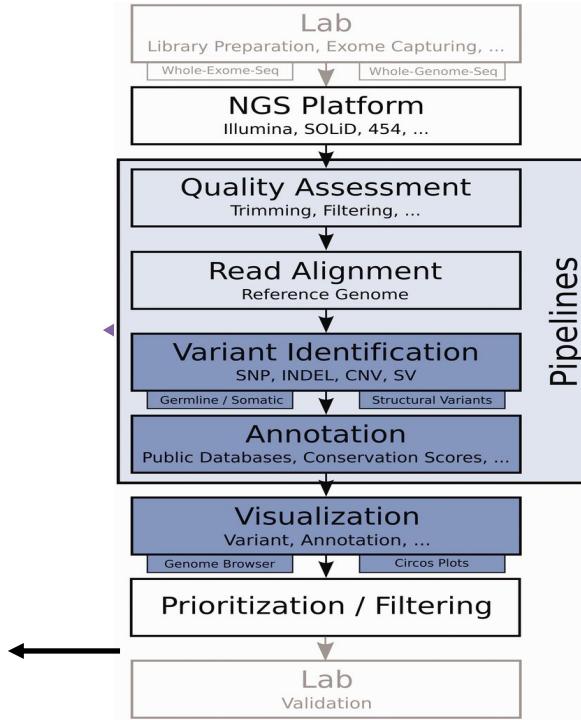


Stephan Pabinger et al. Brief Bioinform 2013;bib.bbs086

Analysis outline



Analysis outline



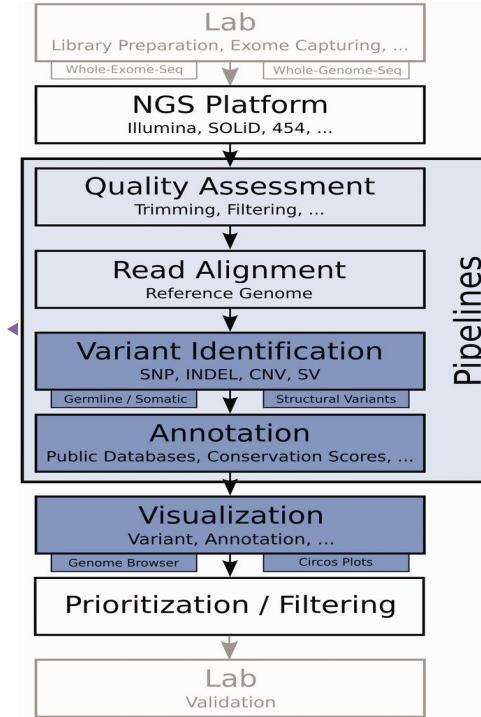
Day 4: Interpretation of the results
- Personalized therapy proposals
- Cancer Genomics resources

Stephan Pabinger et al. Brief Bioinform 2013;bib.bbs086

Analysis outline

Day 5

Case studies in Personalized cancer medicine



Stephan Pabinger et al. Brief Bioinform 2013;bib.bbs086



https://gitlab.com/bu_cnio/varca

Snakemake pipeline implements the **GATK** best-practices workflow for calling small genomic variants.

CNIO Bioinformatics Unit > varca

varca Project ID: 12486221

126 Commits 2 Branches 0 Tags 389 KB Files 8.4 MB Storage

A Snakemake pipeline based on the GATK best-practices workflow

[pipeline](#) [failed](#) [snakemake](#) v5.1.5

master varca / + History Find file Web IDE Clone

Update config-example.yaml Elena Piñeiro authored 1 week ago 835d4584

README MIT License CI/CD configuration

Name	Last commit	Last update
.test	Add VEP for variant annotation	1 year ago
envs	Fix BAM index file redundancy	3 weeks ago
img	Add logo to README	1 year ago
report	Add VCF to report.	3 years ago
rules	fix handling of contigs with asterisks in the name	1 week ago

```
$ mkdir varca
$ cd varca
$ git clone https://gitlab.com/bu_cnio/varca.git .
Cloning into '.'...
remote: Enumerating objects: 853, done.
remote: Counting objects: 100% (128/128), done.
remote: Compressing objects: 100% (74/74), done.
remote: Total 853 (delta 76), reused 103 (delta 53), pack-reused 725
Receiving objects: 100% (853/853), 1.47 MiB | 0 bytes/s, done.
Resolving deltas: 100% (545/545), done.
Checking connectivity... done.
```



https://gitlab.com/bu_cnio/varca

Snakemake pipeline implements the **GATK** best-practices workflow for calling small genomic variants.



The Snakemake workflow management system: **tool** to create reproducible and scalable data analyses.

- Ensures the correct order of execution of each workflow step
- **Modularity:** each step only executed when required by downstream steps
- **Parallelization:** as much as the available resources
- **Reproducibility:** each workflow step also defines its own isolated software environment using **conda**

Koster J and Rahmann S, 2012



https://gitlab.com/bu_cnio/varca

Snakemake pipeline implements the **GATK** best-practices workflow for calling small genomic variants.



"Package, dependency and environment management for any language - Python, R, Ruby, Lua, Scala, Java, JavaScript, C/ C++, FORTRAN, and more."

- Package and environment **management system**
- **Open source**
- Runs on Windows, macOS and Linux
- **Installs, runs and updates** packages and their dependencies
- Easily creates, saves, loads and switches between **environments**



https://gitlab.com/bu_cnio/varca

Snakemake pipeline implements the **GATK** best-practices workflow for calling small genomic variants.

A genomic analysis toolkit focused on variant discovery



- Industry **standard** for identifying SNPs and indels in germline DNA and RNA-seq data
- Currently expanding to somatic short variant calling, copy number and structural variations
- Includes variant callers, processing and quality control toolkits...
- Primarily tools to process **exomes and whole genomes** generated with Illumina sequencing technology



Thanks!