NGS – variant analysis

Filtering and evaluation

GATK workflow

Yesterday

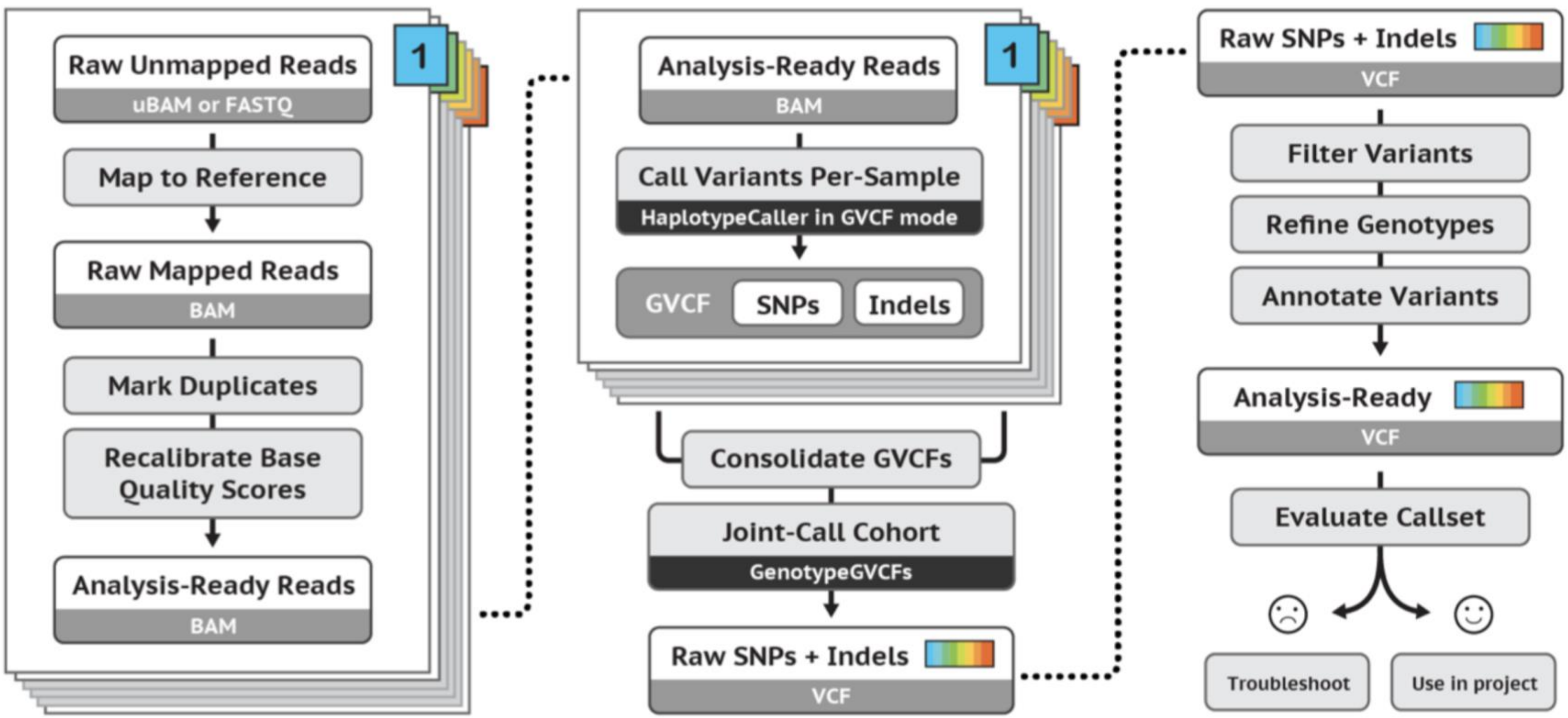


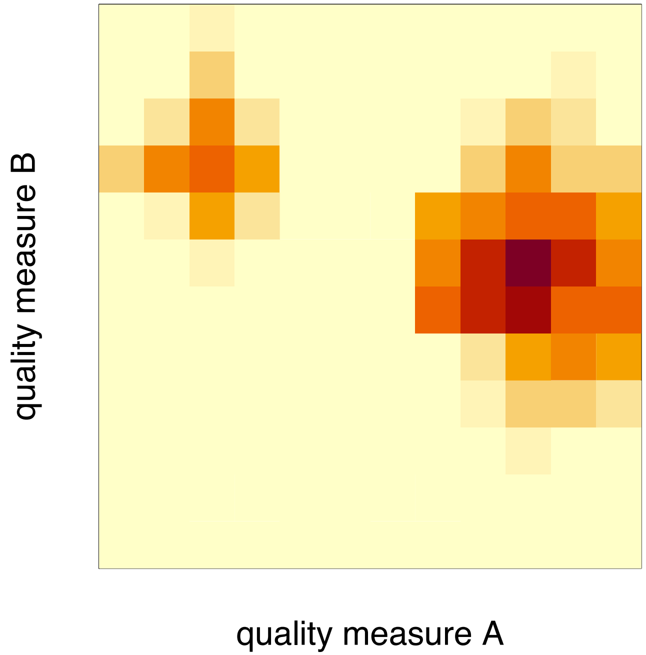
image source: https://gatk.broadinstitute.org

Important QC info

- Mapping quality
- Depth
- Strand-bias
- ..

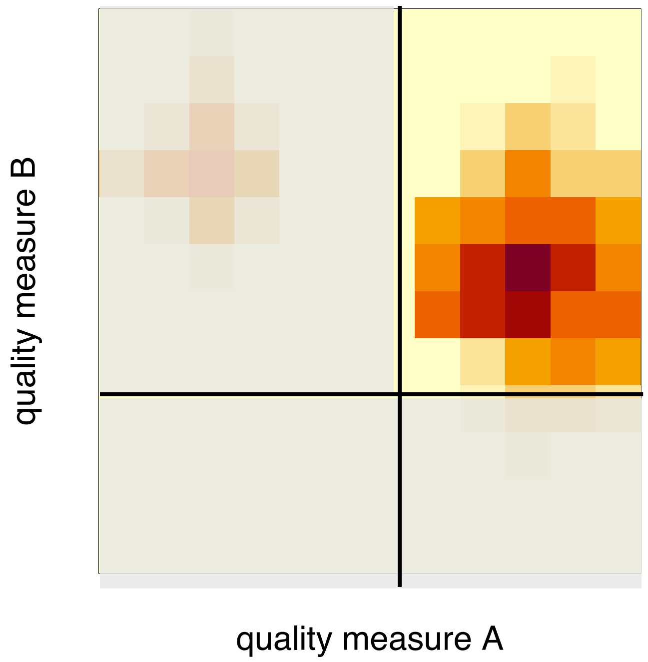
Filtering

- Hard filtering: take thresholds of each measure
- Machine-learning-based: VQSR



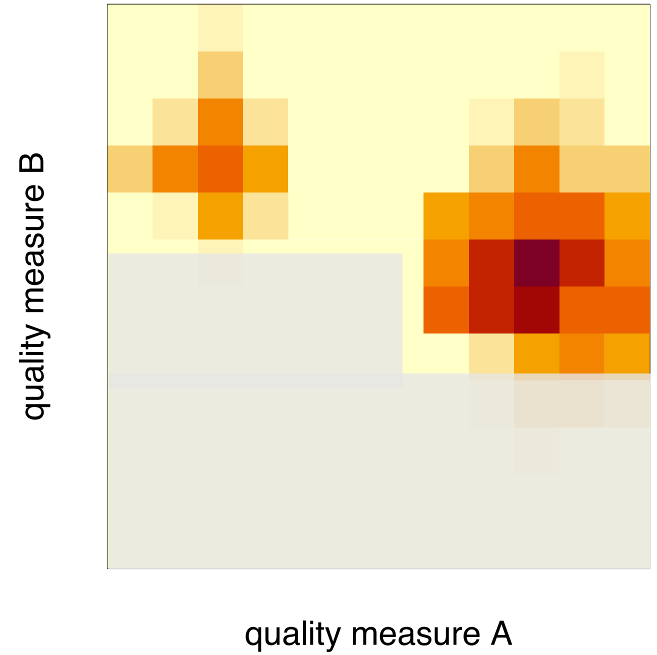
Filtering

- Hard filtering: take thresholds of each measure
- Machine-learning-based: VQSR



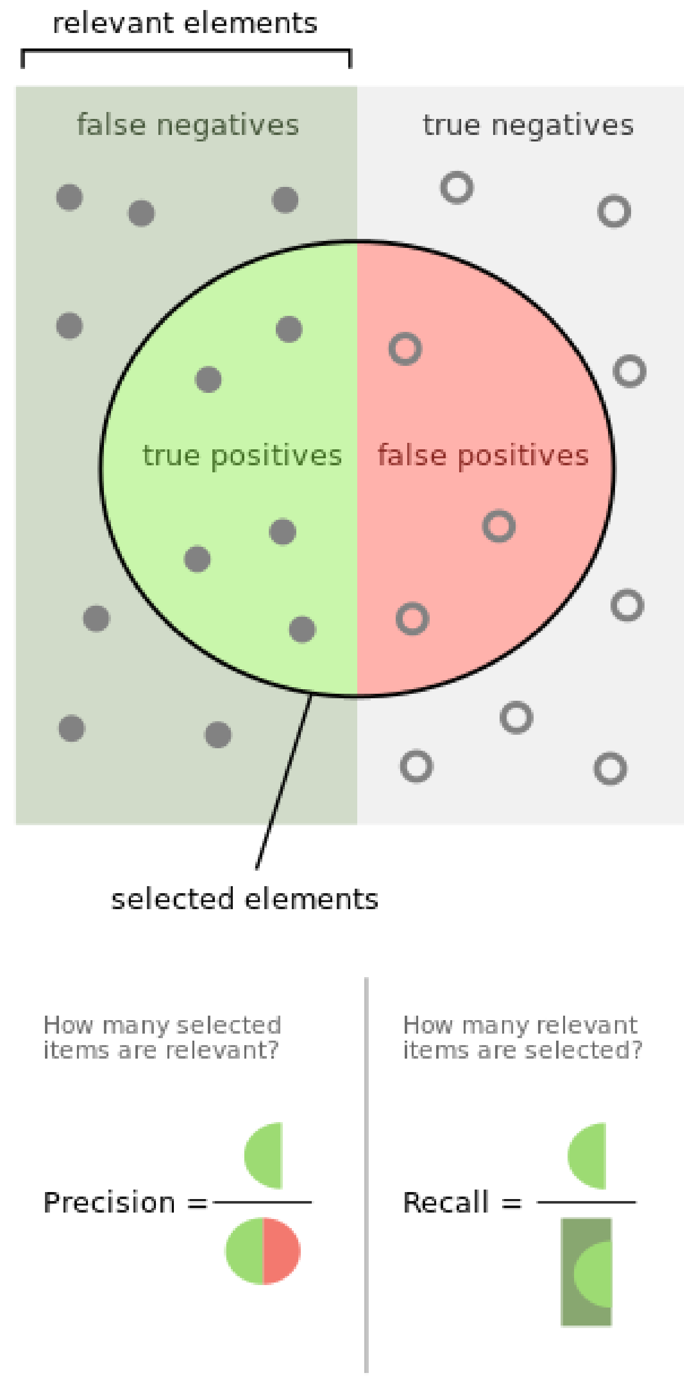
VQSR

- Better performance compared to hard filtering, but you will need:
  - Truth-set
  - ‘Large’ dataset (whole genome or 30 whole exomes)



Evaluation

- **Precision**: how many of the selected variants were true variants?
- **Recall**: how many of the true variants were selected?



Question 14

Evaluation

- **Precision**: how many of the selected variants were true variants?
- **Recall**: how many of the true variants were selected?