Cancer variant analysis

COURSE INTRODUCTION

Flavio Lombardo

2025-12-12

Course etiquette

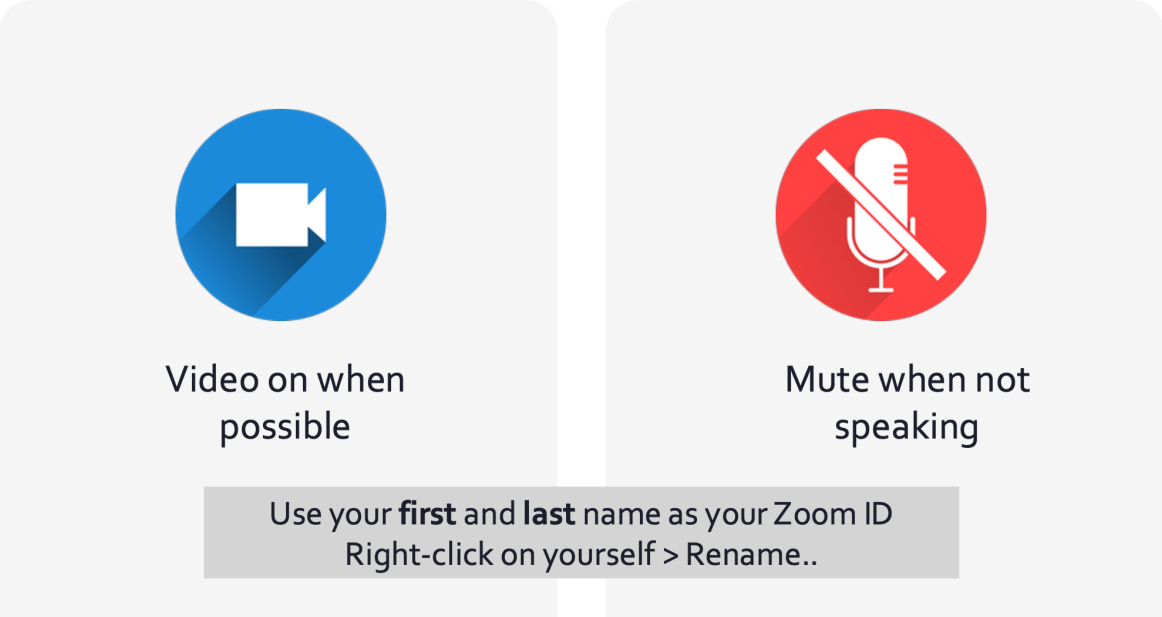


Figure 1: Etiquette

Video on when possible

Use your first and last name as your Zoom ID

- Right-click on yourself > Rename..

Mute when not speaking

Code of conduct

- Learning can only be done in a **safe environment**
- We abide to the ELIXIR **code of conduct** during the course
- If you witness unacceptable behaviour that can't be dealt with with the people involved in this course, contact the COC oversight group
- More info: here

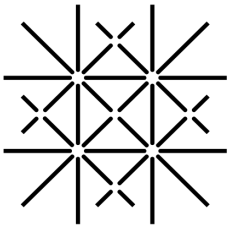
ELIXIR Code of Conduct Principles

Trainers/organizers/Helpers

- **Gregoire Rossier:** Training coordinator at SIB
- **Diana Marek:** Training coordinator at SIB
- **Geert Van Geest:** Training Project Manager at SIB
- **Flavio Lombardo:** Computational Biologist at the University Hospital Basel
- **Deepak Tanwar:** Bioinformaticis Specialist and Trainer at SIB

Swiss Institute of Bioinformatics

Universitätsspital Basel



Universität Basel

Learning outcomes

- Understand the difference between **germline** and **somatic** variants and the implication of computational analysis
- **Perform** a somatic **variant analysis** on a **paired** sample (tumor – normal) with GATK4
- Perform a somatic **variant annotation** with VEP and use the results to filter possible **high-impact mutations** in the cancer genome

Learning experiences

- Lectures
- Polls + quiz questions
- Exercises

Question

[Interactive poll or question would go here]

Communication

- Course website: https://sib-swiss.github.io/cancer-variants-training/
- Shared document
- Zulip

Variant calling

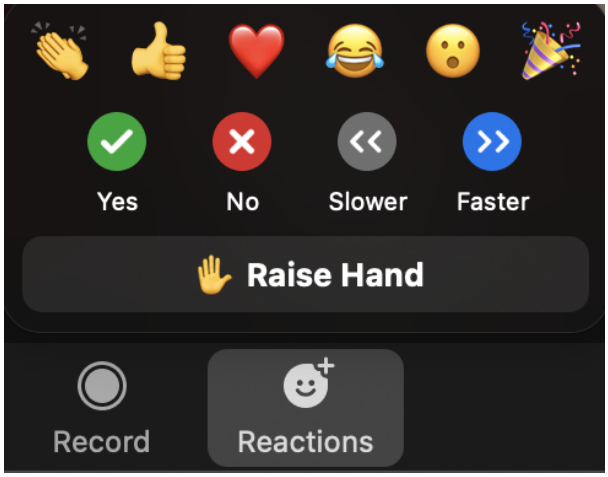
Short variant calling

Now that we have quality-controlled the BAM files, we can go ahead with the va
itself. For this, we used `mutect2` which is a somatic variant caller from GATK ba
`HaplotypeCaller`. In addition to the expected bam files and references genome
command requires some additional input:

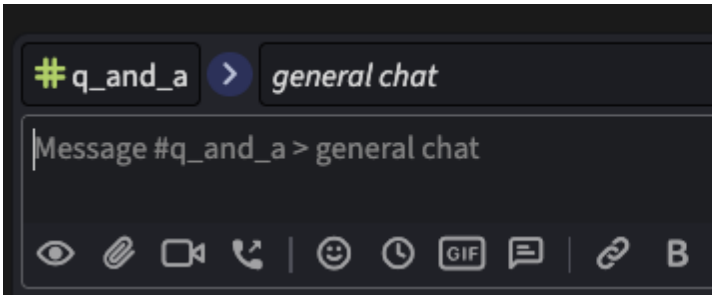
- `--intervals`: the intervals of our target regions. This is in the `interval_li`
- `-normal`: the sample name of the normal sample. So, the `SM` tag of the reac
normal sample.
- `--germline-resource`: know sites of germline variants with their allele frequ
population. These are used to estimate the confidence of a germline varian
normal sample.
- `--panel-of-normals`: A VCF generated from normal samples that contain s
known technical artifacts. Ideally, you create a PON from your own normal
this is typically recommended if you have more than 40 normal samples. Th
here, we use a pre-generated PON from the 1000 genomes project. More in
PON in this article.

Asking questions

- During lectures: zoom functionality – **but the chat will be ignored**



- During exercises: **#q_and_a channel on Zulip**



Introduction round

- Your name
- Where are you working
- Why are you joining this course

...And consider the social channel in Zulip!