Computational Human Genomics Project, 2022-2023

Most genomic studies in the context of oncology require the characterization of somatic events assessed through the sequencing of bulk DNA tumor samples from patients' cohorts (i.e., multiple patients must be studied to answer specific biological/clinical questions). However, the proper analysis of single patients can provide informative data.

Starting from the two provided BAM files (i.e., tumor and control DNA seq from the same individual), identify all somatic events (e.g., SNV, CN), and define their clinical relevance. It is advisable to run quality checks, sample identity checks, visual inspections, and tumor population characterization. You are not required to use tools other than those utilized during the classes.

Notes:
If you need to determine the genotype, you can use ASEReadCounter, then compute allelic fraction (AF) and assign a genotype to each site considering the AF value (use the following thresholds: AA<0.2, 0.2≤AB≤0.8, BB>0.8).

Please note that the provided BAM files are limited in genomic size to allow for lighter processing. Despite this fact, some intermediary files may take up some GB of space. If you use the provided virtual machine, you may run out of space on disk: remember that some tools allow for setting a max depth parameter to avoid wasting too much space and time during the analysis.

Format of the required project report:
Please comply with the following: 4 pages max, font size 11, font type Arial, margins > 2cm, single line spacing. Include figures and related legends, min 2, max 4.
The report should include the following information: names of the students, project rationale (max 10 lines), computational workflow (including details to make it reproducible), relevant results with related interpretations, and if necessary, pitfalls and criticisms (max 10 lines).

Teamwork. Form groups of 3-4 students each.

Deadline: you need to submit your report 1 week before the exam date (yari.ciani@unitn.it; f.demichelis@unitn.it).