

Zadatak 2

- Implementirati na programskom jeziku Python algoritam za variant calling zasnovan na odlučivanju po binomnoj raspodeli. Ulaz u algoritam je fajl u pileup formatu, a izlaz u VCF 4.2 formatu. Fajl u pileup formatu moguće je napraviti iz BAM fajla koristeći alat samtools pileup (10 poena).
- Uporediti rezultate implementiranog variant calling-a sa izlazom (VCF-om) dobijenim iz bcftools call alata. Kao test podatke koristiti (merged-normal.BAM fajl, referentni genom human_g1k_v37_decoy.fasta dostupni u CGC Public Files galeriji). Dati detaljan izveštaj u formi dijagrama: koliki je procenat varijanti koje se preklapaju (true positives), koliki procenat postoji u implementiranom rešenju, a ne postoji u bcftools call (false positives) i obrnuto, koliko postoji u bcftools call, a ne postoji u implementiranom rešenju (false negatives). Sračunati po definiciji precision, recall i f-score metrike iz dobijenih podataka (10 poena).
- Za svaku od funkcija u kodu, kao i za sam finalni algoritam napisati testove (5 poena).
- Pripremiti prezentaciju (Google slides ili power point) inicijalnog i optimizovanog algoritma, kao i samih rezultata (5 poena).
- Pripremiti video prezentaciju projekta (3 - 5 minuta trajanja) koja će biti dostupna na YouTube ili drugom on-line video servisu (10 poena).