

PROBLÉM 1 – DNA sekvence (jednotlivě)

Zjistěte co nejvíce informací o lidském genu BRCA1

- 1) jaké má identifikátory a synonyma (NCBI, Ensembl)
- 2) Kde se nachází v lidském genomu (UCSC Genome Browser)
- 3) Jakou má sekvenci
- 4) jaká je struktura genu (kolik exonů v jaké celkové délce?)
- 5) Jaké podobné sekvence jsou známy u savců a kterých, je něco podobného ve kvasinkách? (BLAST)

PROBLÉM 2 – DNA sekvence (skupinově)

Analyzujte genom a proteom koronaviru nového typu (<https://www.ncbi.nlm.nih.gov/sars-cov-2/>)

- 1) Kolik sekvencí je v NCBI? V jiných organismech kromě člověka?
- 2) Na jaké pozici je “spike protein” (označován jako S)?
- 3) Uložte si >20 sekvencí do souboru (formát FASTA) a vygenerujte mnohočetné zarovnání (multiple alignment).
- 4) Kolik pozic referenčního genomu v intervalu 200-29800 obsahuje v některé ze sekvencí mutaci oproti referenčnímu genomu?
- 5) Anotujte mutace podle toho jestli jsou v kódující nebo nekódující části genomu a podle procenta výskytu ve Vaší sbírce.

PROBLÉM 3 – DNA (sekvenační data)

- 1) Hledejte sekvenační experiment na lidské DNA s párovými “ready” (SRA, ENA)
- 2) Pro Vámi vybraný experiment, jaké jsou parametry (typ sekvenace, biologický materiál)?
- 3) Kolik RNA-seq experimentů je zveřejněno u knotovky bílé (silenka širolistá)?
- 4) Jaké jiné experimenty kromě DNA-seq a RNA-seq obsahují veřejné databáze?

PROBLÉM 4 – Proteiny (sekvence i struktura)

- 1) Jaká je sekvence a doménové složení lidského proteinu BRCA? (Uniprot)
- 2) S jakými jinými doménami se kombinují domény BRCA1? (Interpro, Pfam, BLAST)
- 2) Je známá struktura BRCA1? (PDB, Uniprot)
- 3) Jaké má struktura proteinu vlastnosti (sekundární struktura, příslušnost k rodině a pod.) CATH-SCOP?
- 4) Jaké podobné proteiny existují u jiných organismů? (Gene Ontology?)