



Contribution ID: 30

Type: **not specified**

Pretok podatkov v klinični genetiki za redke bolezni

Thursday, December 5, 2024 11:50 AM (20 minutes)

Redke bolezni so bolezni z incidenco manjšo od 1 na 2.500 ljudi, ki kumulativno prizadenejo 5 do 7 % populacije, kar znaša 120.000 ljudi le v Sloveniji. Pot do diagnoze je dolgotrajna, hkrati so možnosti zdravljenja omejene. Sekvenciranje humanega genoma postaja ena ključnih metod pri diagnostiki redkih bolezni. Na Kliničnem inštitutu za specialno laboratorijsko diagnostiko (KISLD), Pediatrične klinike UKCLJ, izvajamo genetsko diagnostiko na podatkih pridobljenih s tehnologijami dolgih sekvenc (Oxford Nanopore) in kratkih sekvenc (Illumina NovaSeq). Letno izvedemo približno 2.500 sekvenciranj celotnih eksomov (WES) in 200 sekvenciranj celotnega genoma (WGS), z 30% letnim prirastom. Gre za kompleksne genetske podatke, katerih velikost dosega do 1,2TB na pacienta. Analize teh podatkov potekajo v časovno potratnih večstopenjskih cevovodih, katerih vmesni rezultati porabljajo dodaten prostor. Procese je možno značilno pohitrit z visoko stopnjo paralelizacije. Za ta namen ima KISLD na voljo lokalni HPC s tremi računskimi vozlišči, od tega sta dve CPU-vozlišči in eno GPU-vozlišče. Čas obdelave za WGS je na delovni postaji z 256 CPU-nitmi približno 27 ur. Na HPC z dvema CPU vozliščema se ta čas skrajša na 17 ur, medtem ko z osmimi A100 grafičnimi procesorji (GPU) traja zgolj 1,5 ure. Izzivi ostajajo pri pretoku podatkov od sekvenatorjev do računskih vozlišč, avtomatizacija in sledenje specifičnih analiz, priklic in predstavitev rezultatov končnemu uporabniku ter hranjenje in varnostno kopiranje.

Presenter: VRHOVŠEK, Blaž (UKC Ljubljana)

Session Classification: Dan slovenskega superračunalniškega omrežja