

arnes 
povezujemo znanje



MREŽA ZNANJA

Ljubljana, 3.–5. december 2024



Pretok podatkov v klinični genetiki za redke bolezni

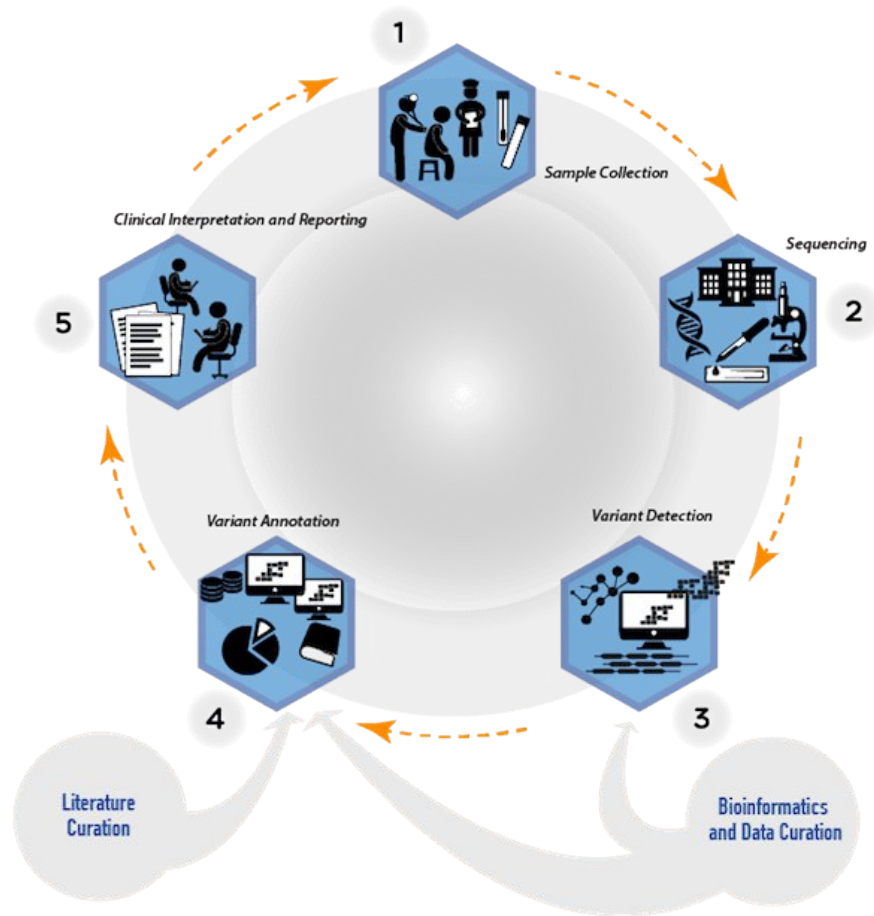
Blaž Vrhovšek

univerzitetni
klinični center ljubljana 
University Medical Centre Ljubljana

KISLD

Klinični inštitut za specialno
laboratorijsko diagnostiko

Standardni potek genetske diagnostike na osnovi tehnologije sekvenciranja nove generacije (NGS).



OSNOVE BIOINFORMATIKE

Kako izgleda DNK v digitalni obliki?

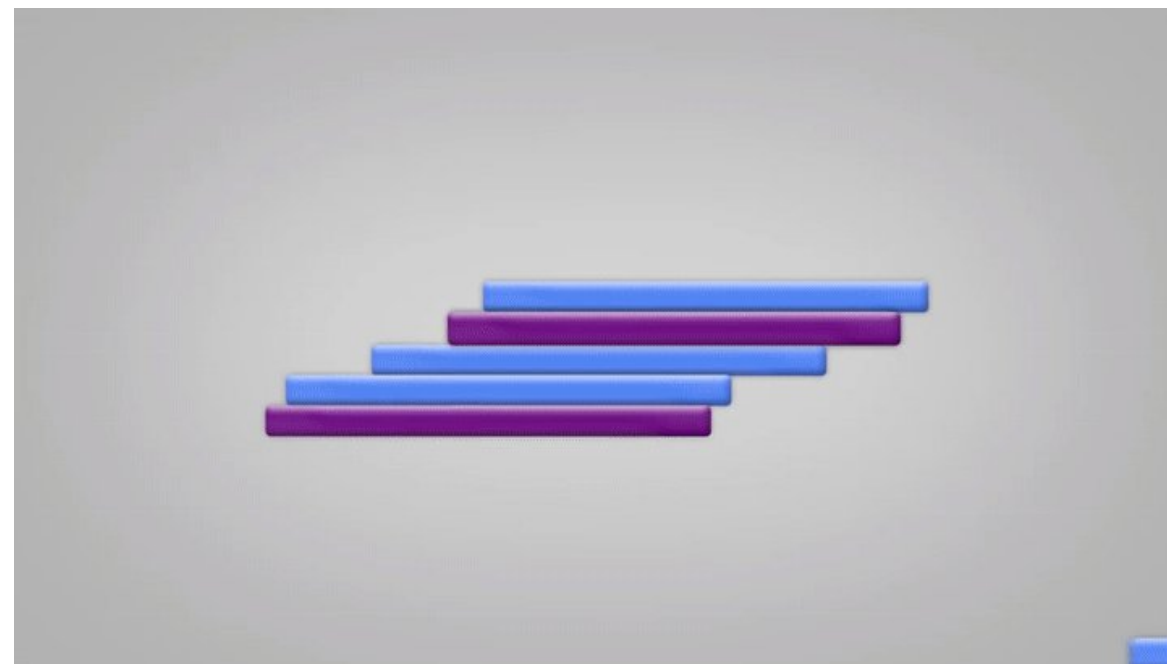
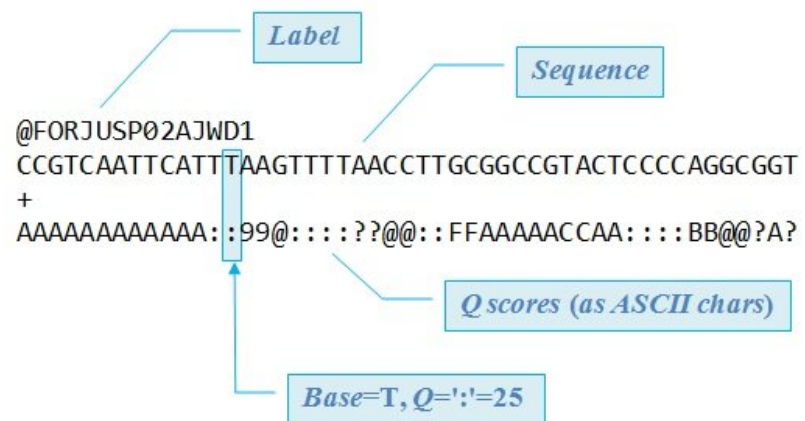
- Niz sestavljen iz A, C, G, T in N
- V praksi uporabljamo enak niz za vse osebe znotraj vrste
- Haploidni referenčni človeški genom sestavlja 3 milijarde milijarde nukleotidov ~ 3,2 GB
- Le ~1% je kodirajoče DNA, ali eksom

GGGTAAGATGCAAGATATTTCTGTGCACATCTTCAGATGAATTTCTTGTAGTGTGTG
TGTGTTTGTCTCACACATATGCGTGAAAGAAGAGTACATACACAGATCTCCTCAAAAAGGA
GGCAGCAAGCCCGTTCAAGAATGGGACTGAATACACCTGATGAGTGGTTACTTTCTGTC
TGCAAACATCTACTGATCATCTGTTAGGTGCAGGCCATGATCACAACAAAGACGAATAAG
ACACTACACTAGCCAGGGAGAGTCTCAAAAACAATAAATAAATAAATAAATAAATAAATAA
TCCAGTCATGGGTACAAAGCTAAGGAGTGACAAATCCCTCTTGGAGTTAGGGGAGTCAGG
AAAAAGCTCTTAGCAGAATGTGTGCCTCTCGGCCGGGCGCAGCGGCTCACGCCTGTAATC
CCAGCACTTTGGGAGGCGAAGGCAGGCAGATCACCTGAGGTGCGGGAGTTCGAGACCAGTC
TGACCAACATGGTGAAACTCCATCTCTACTAAAAATACAAAATTAGCCAGGCGTGGTGGT
GCATGCCTGTAATCCCCGCTACTCGGGAGGCTGAGGAAGGAGAATCACTTGAACCAGGAA
GGTGGAGGTTGCAGTGTGCCAAGATCGCGCCATGGCACTCCAGCCTAGGCAACGAGGGTG
AACCAGGTCCAGGAAGAAGGTGCAAAGACAGCATTCCAGGTAAAAGAAACAGCTTGAACA
AAAAGTGTAGGGGAACCGCAAGCGGTCTTGAGTGTGAGGGTACAATCATCCTTGGGG
AAGTACTAGAAGAAAGAATGATAAACAGAGGCCAGTTTGTAAAAAACTCAAAATTA
GCTAGGAGTTTGGACTTGTGGCAGGAATGAAATCCTTAGACCTGTGCTGTCCAATATGGT
AGCCACCAGGCACATGCAGCCACTGAGCACTTGAATGTGGATAGTCTGAATTGAGATGT
GCCATAAGTGTAAAATATGCACCAAATTTCAAAGGCTAGAAAAAAGAATGTAAAATATC
TTATTATTTATATTGATTACGTGCTAAAATAACCATATTTGGGATATACTGGATTTTAA
AAATATATCACTAATTTTCTGTTTCTTTTACTTTTAGAAATCACATATGTGACTTAA
ATATTTCTTTTCTTTTCTTTCTTCTCTCACTCAGCGTCTCTGTGATTCCAAAGAAATGAGTC
TCTGCTGTTTTTGGGCAGCAGATATCCTAGAATGGACTCTGACCTAAGCATCAAAATTA
TCATCATAACGTTATCATTTTATGGCCCTTCTTCTATATCTGGTAGCTTTTAAATGAT
GACCATGTAGATAATCTTTATTGTCCCTCTTTCAGCAGACGGTATTTTCTTATGCTACAG
TATGACTGCTAATAATACCTACACATGTTAGAACCATTCTGACTCCTCAAGAATCTCATT
TAACTCTTATTATCAGTGAATTTATCATCATCCCCTATTTTACATAAGGAAATGGGGTTA
GAAAGACCAATAACATTTTTTCAACATCAAAACACTAGCTTGGAGATCAAGCCCAGACTT
GGATCTGTCGTCTGAATTTCAAGCTTTTTGTTATTTATTGATATGTTTTGTTGTTTTCAT
GCAATAATGCAAATCTTAGCCCAAACATTTTGTAGTAGTACCAACTGTAAGTCACCTTA
TCTTCATACTTTGTCTTTATGTAACCTAAATTAGATCTGTTTTTGTACTGAGGGAAAA
ACAAGGGAATCTAACACTAACAGCCCGTAGTGTGTGGTCAACACTTTCGTTACTTTAGT
ATACATCACCCCAATTGTTTGTCTTACCACACACTTTGGAGTTAGGTAGTAGTATCTAT
TTTTACAAATAAGAAAACCCAGGCACAAAGGGTTGATTAGCAATTATCTTTGAAAAGC
CTGTAGTTGCTCATCTGAAGAAGTGACGGACCACCTCTTATTTAGTGGACAGACAGTAAC
TAGTTGAGAAGACAGGGGATTTTGTGGCGGAAAAAAAATTTATCAAAAGTCGTCTTCT
ATCAGGGAGTTTTATGAGAAACCTAGCTCCTCAGTTCACAGTGGGTAACGTAATTCA
TTCTAGGTCTGCGATATTTCTGCCTATCCATTTTGTAACTCTTCAATGCATTCCACAA
ATACCTAAGTATTTAATAATGGTGGTTTTTTTTTTTTTTTTTGCATCTATGAAGTTTTT
TCAAAATCTTTTTAAGTGACAAAATTTGTACATGTGTATCGCTCAATATTTCTAGTCGAC



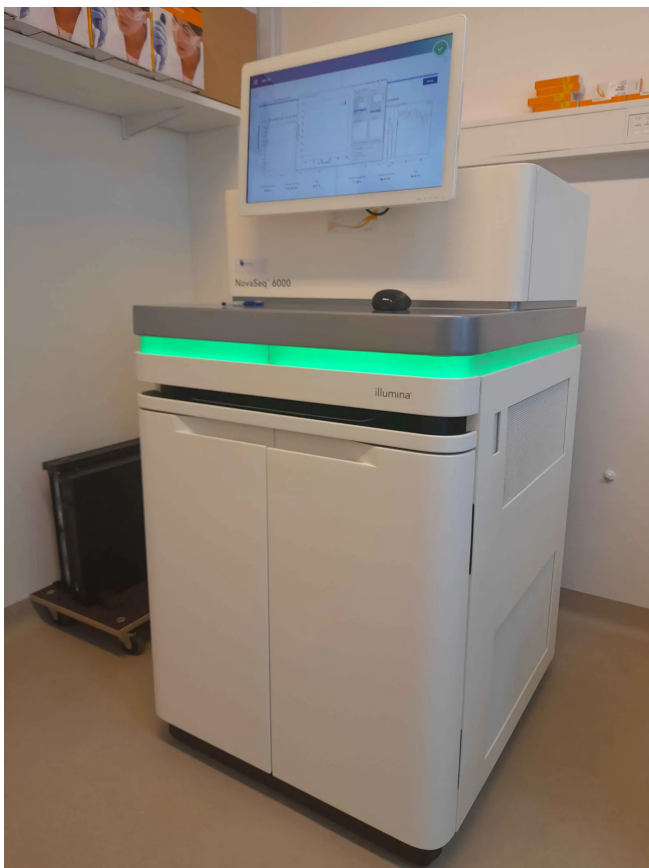
Kako "beremo" DNK na KISLD?

- DNK fragmentiramo in odčitamo na sekvencionatorjih, nato sestavimo na referenco



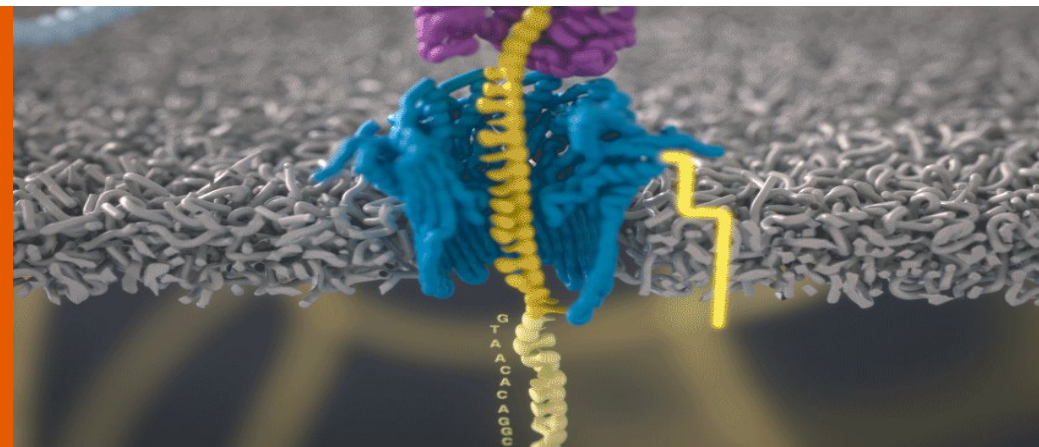
Kako "beremo" DNK na KISLD?

Sequencing



- Kratki odčitki z Illumina.
- ~ 100 GB surovih podatkov na gonom.
- Traja 1 – 2 dni, odvisno od velikosti "kasete".

Kako "beremo" DNK na KISLD?



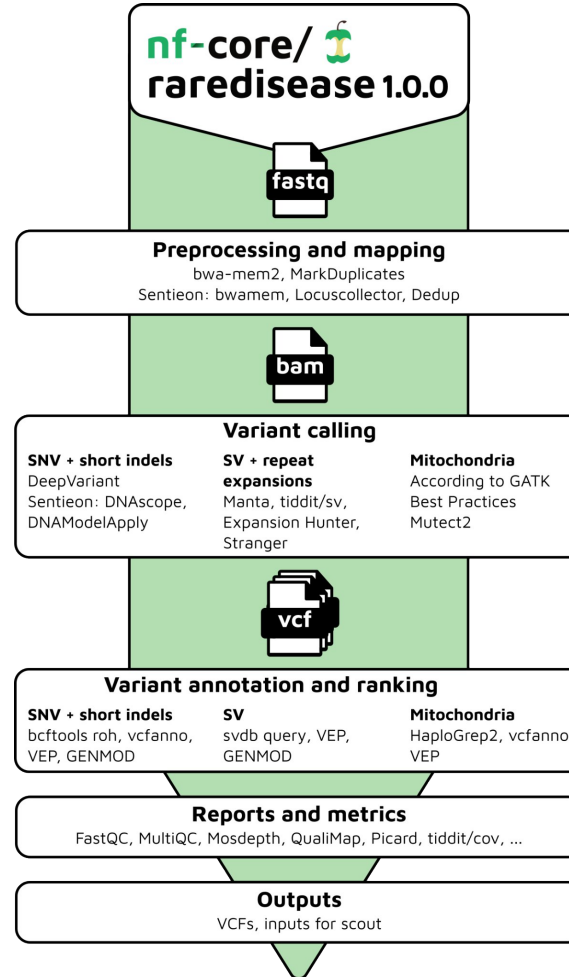
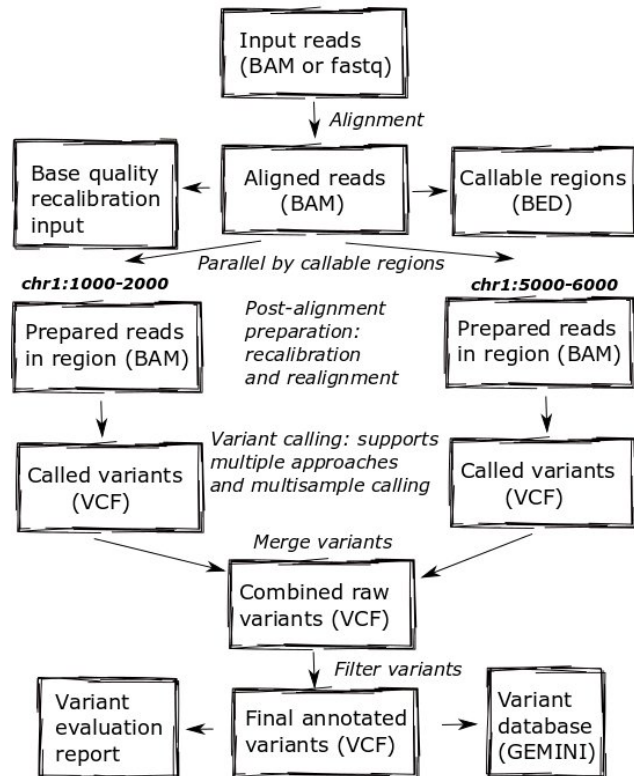
- Dolgi odčitki z Oxford Nanopore.
- ~ 1,2 TB surovih podatkov na genom.
- Traja ~3 dni.



Kako analiziramo DNK na KISLD?



Variant calling overview



epi2me-labs/wf-human-variation



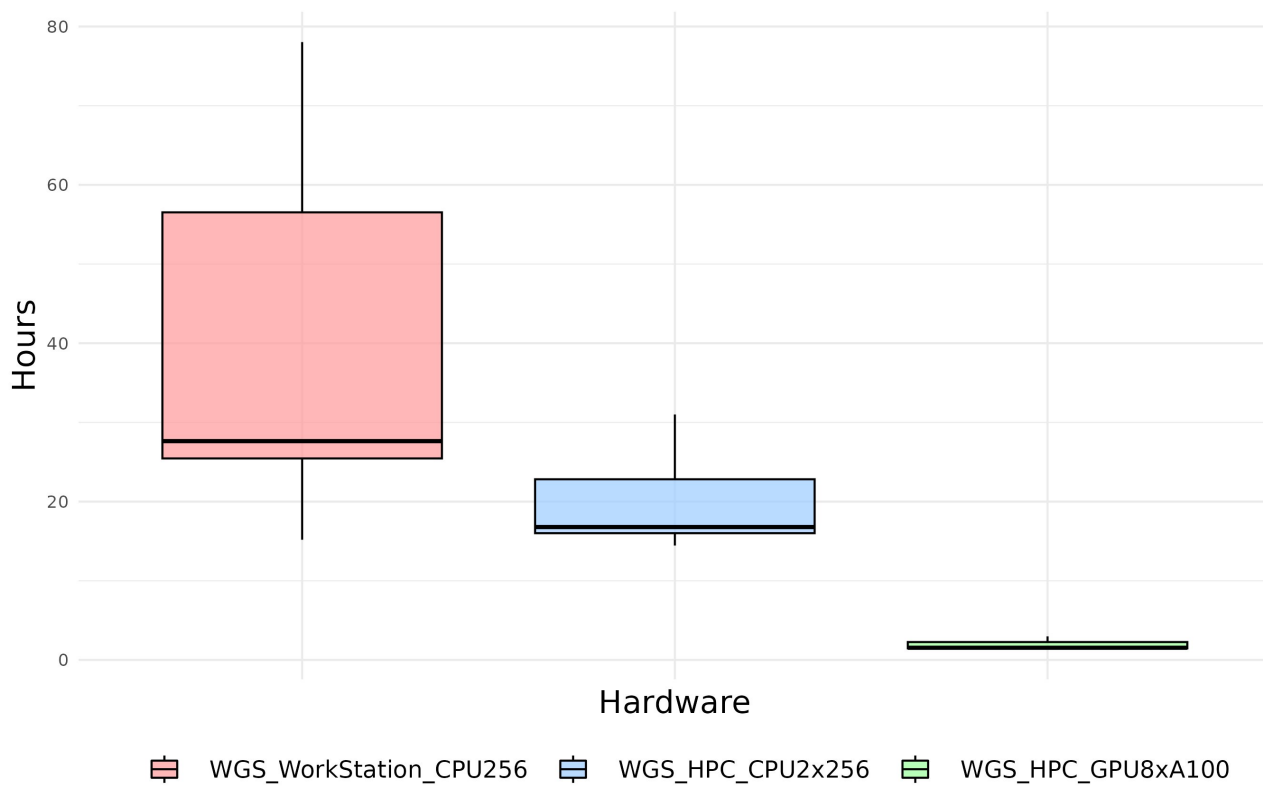
clara-parabricks-workflows/parabricks-nextflow





Kako hitro analiziramo DNK na KISLD?

Time Differences



- SuperMicro WorkStation z 256 niti
 - ▶ ~27 ur
- Lokalen HPC z 2x256 niti
 - ▶ ~17 ur
- Lokalen HPC z 8xA100
 - ▶ ~1,5 ur
- Prostor:
 - ▶ Dolgi odčitki x1 ~ +1 TB
 - ▶ Kratki odčitki x4 ~ +500 GB

Ureditev KISLD HPC



nextflow



CISLD

Clinical Institute of Special
Laboratory Diagnostics



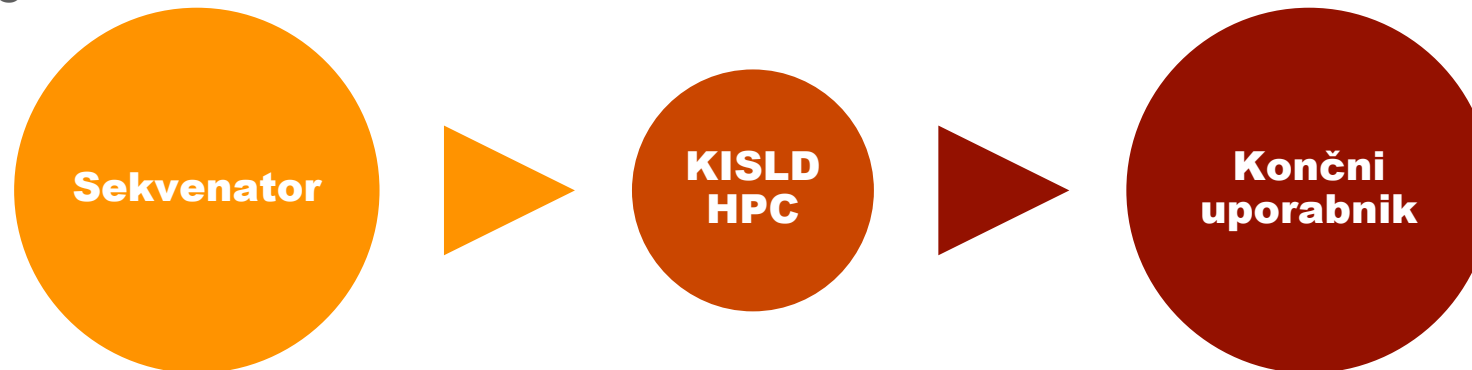
Podatkovni tok

Klinično podatki ostanejo znotraj lokalne mreže.

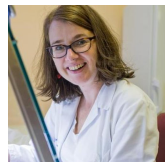
Možnost direktnega prenosa na HPC

Izzivi:

- Avtomatizacija in sledenje specifičnih analiz.
- Priklic in predstavitev rezultatov končnemu uporabniku.
- Hranjenje in varnostno kopiranje.



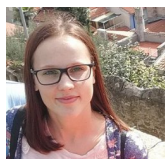
- 2.500 sekvenciranj celotnih eksomov
- 200 sekvenciranj celotnega genoma
- 30% letni prirast



dr. Maruša Debeljak



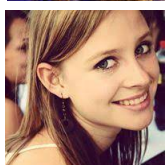
dr. Barbara Jenko Bizjan



Tamara Grgić



Ana Grom



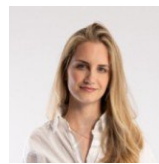
Barbara Slapnik



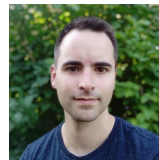
dr. Barbara Čugalj Kern



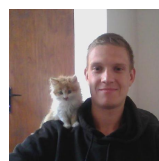
Ana Markež



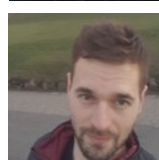
Karolina Mužina



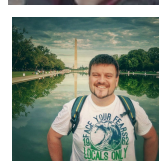
dr. Tine Tesovnik



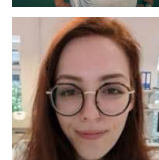
Blaž Vrhovšek



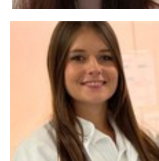
dr. Robert Šket



dr. Jernej Kovač



dr. Klementina Črepinšek



Anja Cerovšek



Financiranje

Projekt EuroCC 2 financira Evropska unija. Financiran je s sredstvi Skupnega podjetja za evropsko visokozmogljivo računalništvo (EuroHPC JU) ter Nemčije, Bolgarije, Avstrije, Hrvaške, Cipra, Češke republike, Danske, Estonije, Finske, Grčije, Madžarske, Irske, Italije, Litve, Latvije, Poljske, Portugalske, Romunije, Slovenije, Španije, Švedske, Francije, Nizozemske, Belgije, Luksemburga, Slovaške, Norveške, Turčije, Republike Severne Makedonije, Islandije, Črne gore in Srbije v okviru sporazuma o dodelitvi sredstev št. 101101903.

Delovanje Nacionalnega kompetenčnega centra SLING sofinancira Ministrstvo za visoko šolstvo, znanost in inovacije.

Medijski sponzor

**Računalniške
novice**

www.racunalniske-novice.com



**Co-funded by
the European Union**



REPUBLIKA SLOVENIJA
**MINISTRSTVO ZA VISOKO ŠOLSTVO,
ZNANOST IN INOVACIJE**