HPC pielietojums bioinformātikā

Edgars Liepa
2024

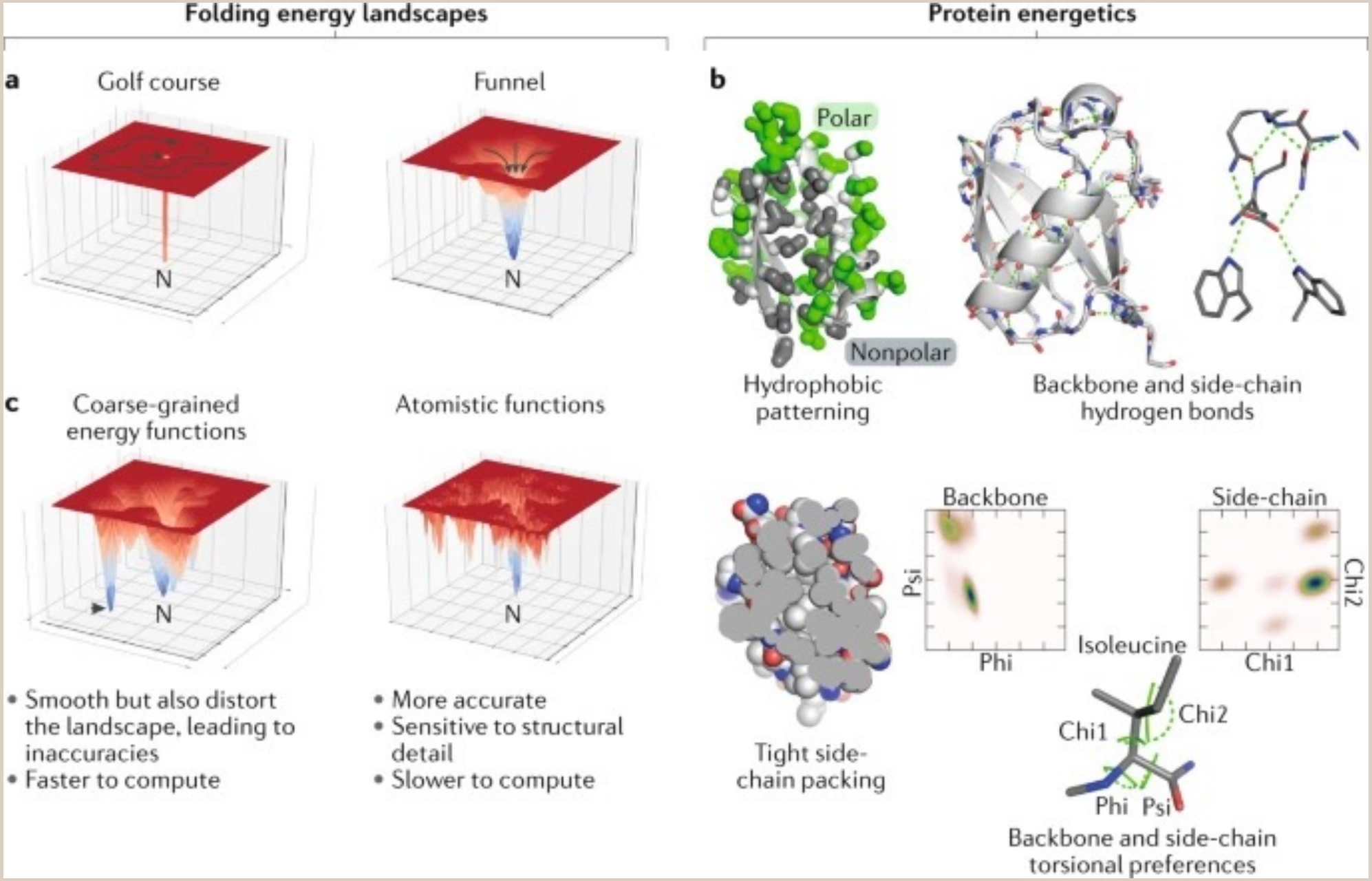
Par mani

- Bioinformātiķis LV BMC
  - D. Fridmaņa grupa
- PhD students LU BF

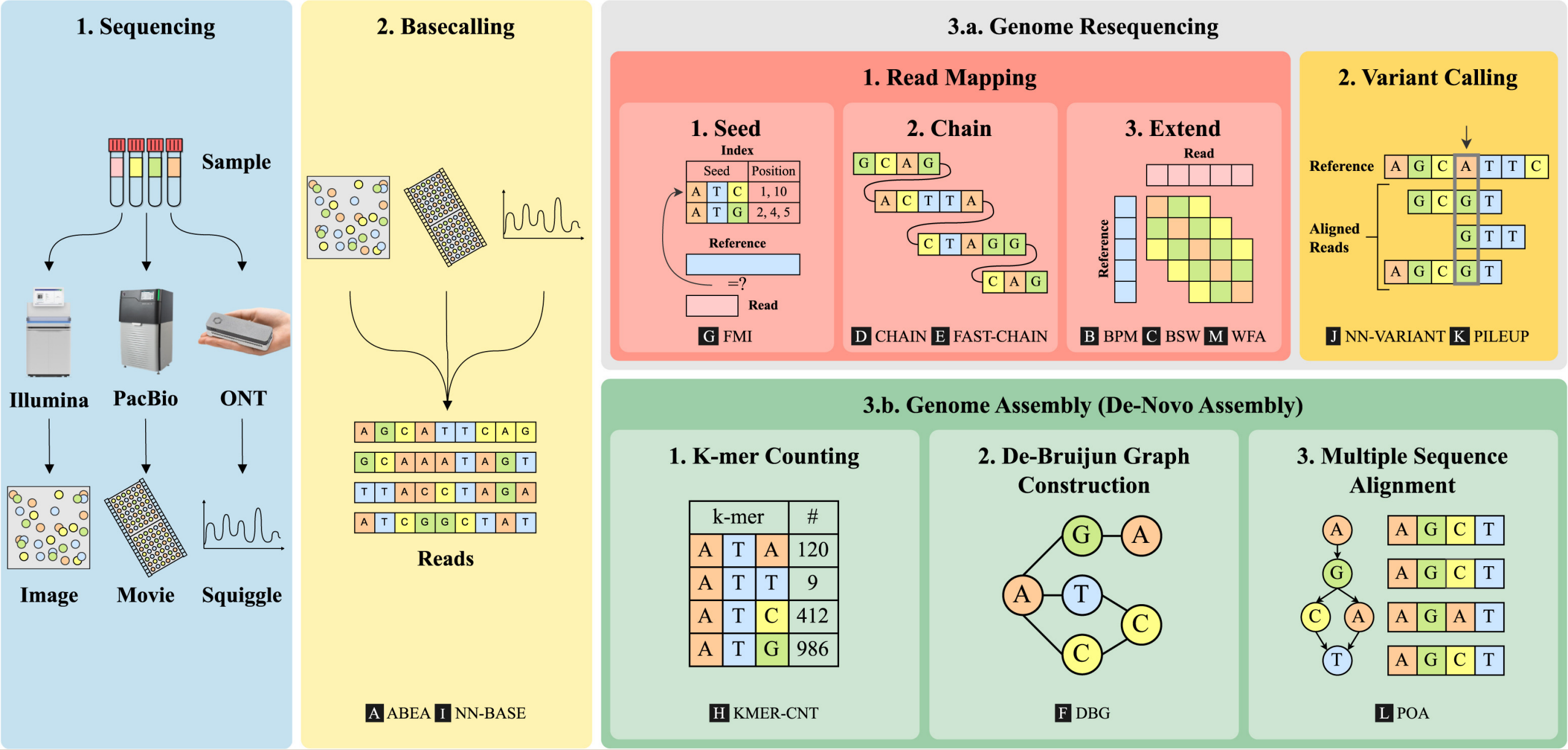
Web: edgarsliepa.lv

LinkedIn: linkedin.com/in/edgars-liepa/

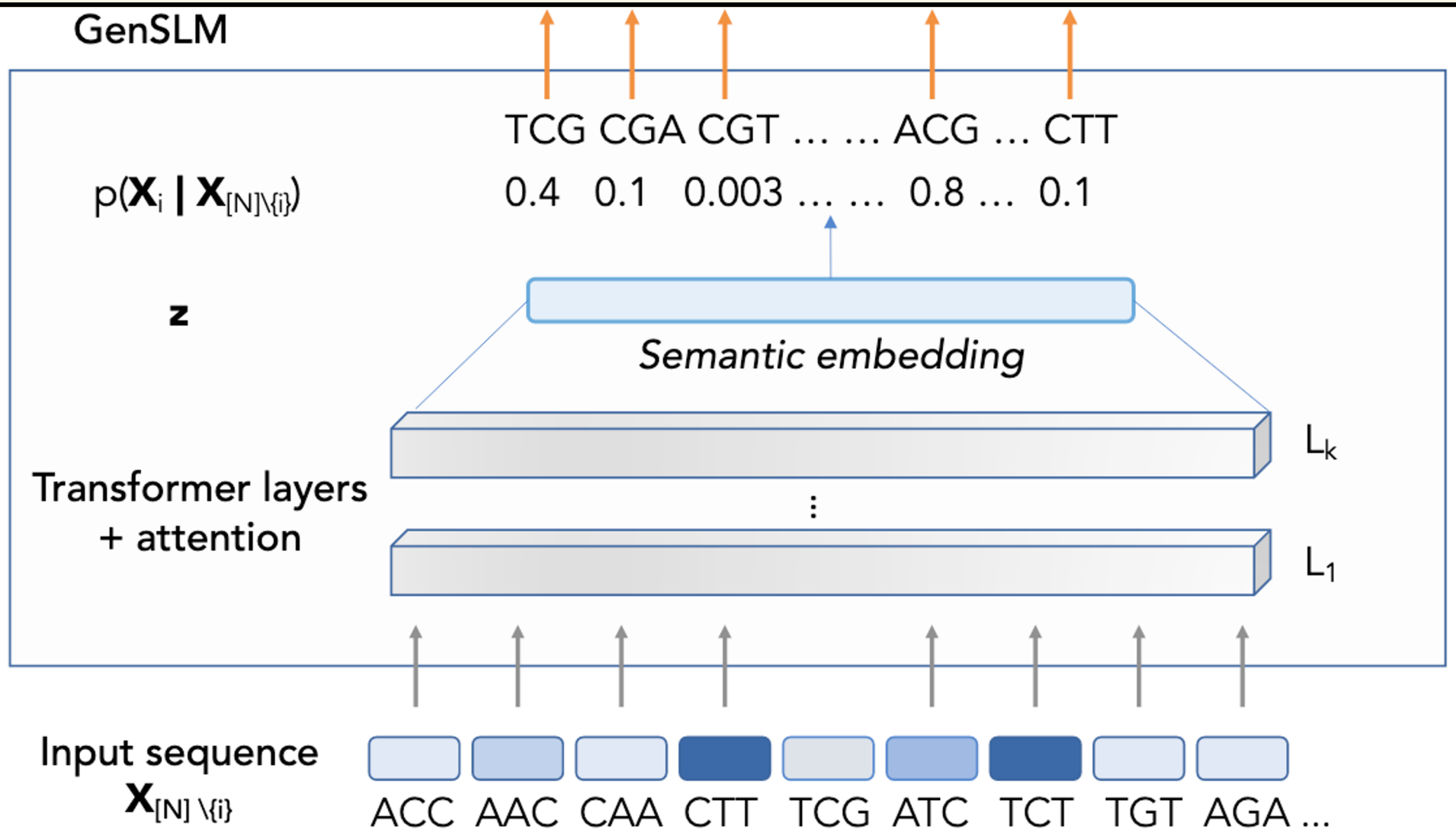
Github: github.com/EdgarsLiepa



McMaster, Benjamin, et al. "Can AlphaFold's breakthrough in protein structure help decode the fundamental principles of adaptive cellular immunity?." *Nature Methods* (2024): 1-11.



López-Villellas, Lorién, et al. "GenArchBench: A genomics benchmark suite for arm HPC processors." *Future Generation Computer Systems* 157 (2024): 313-329.


GenSLM
TCG CGA CGT … … ACG … CTT
$p(\mathbf{X}_i \mid \mathbf{X}_{[N]\setminus\{i\}})$
0.4 0.1 0.003 … … 0.8 … 0.1
z
Semantic embedding
Transformer layers
+ attention
L_k
L_1
Input sequence
$\mathbf{X}_{[N]\setminus\{i\}}$
ACC AAC CAA CTT TCG ATC TCT TGT AGA …

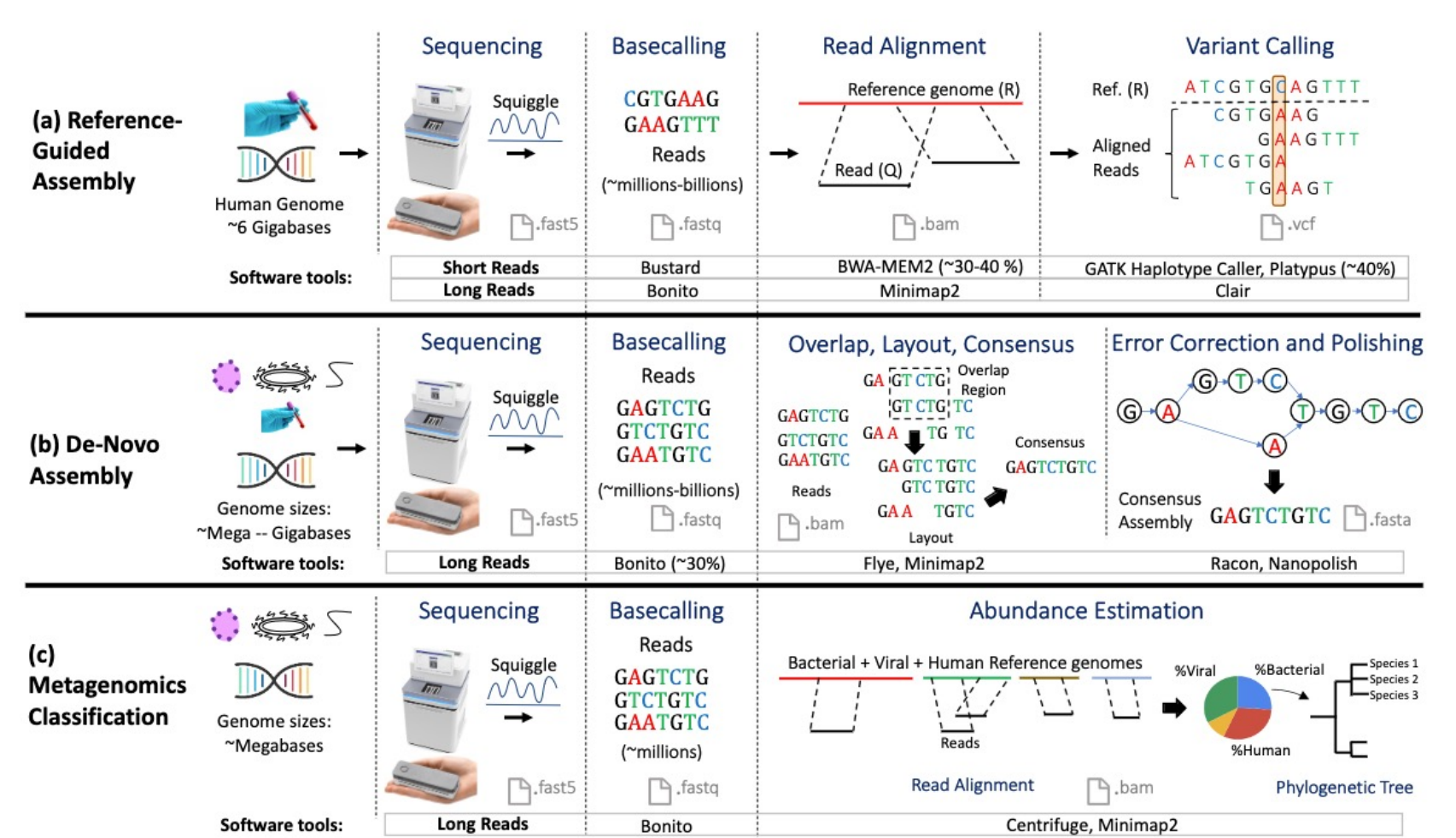
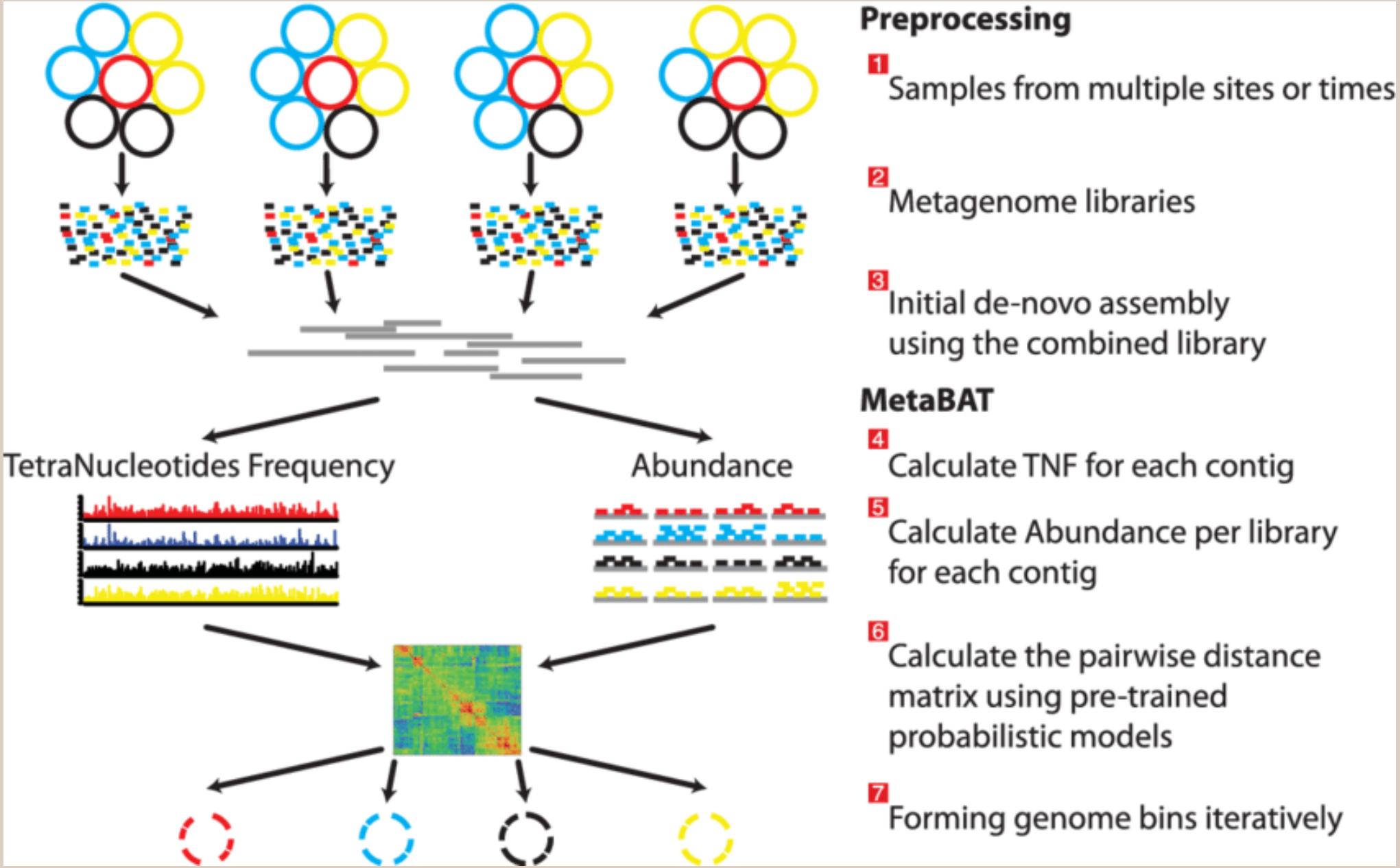
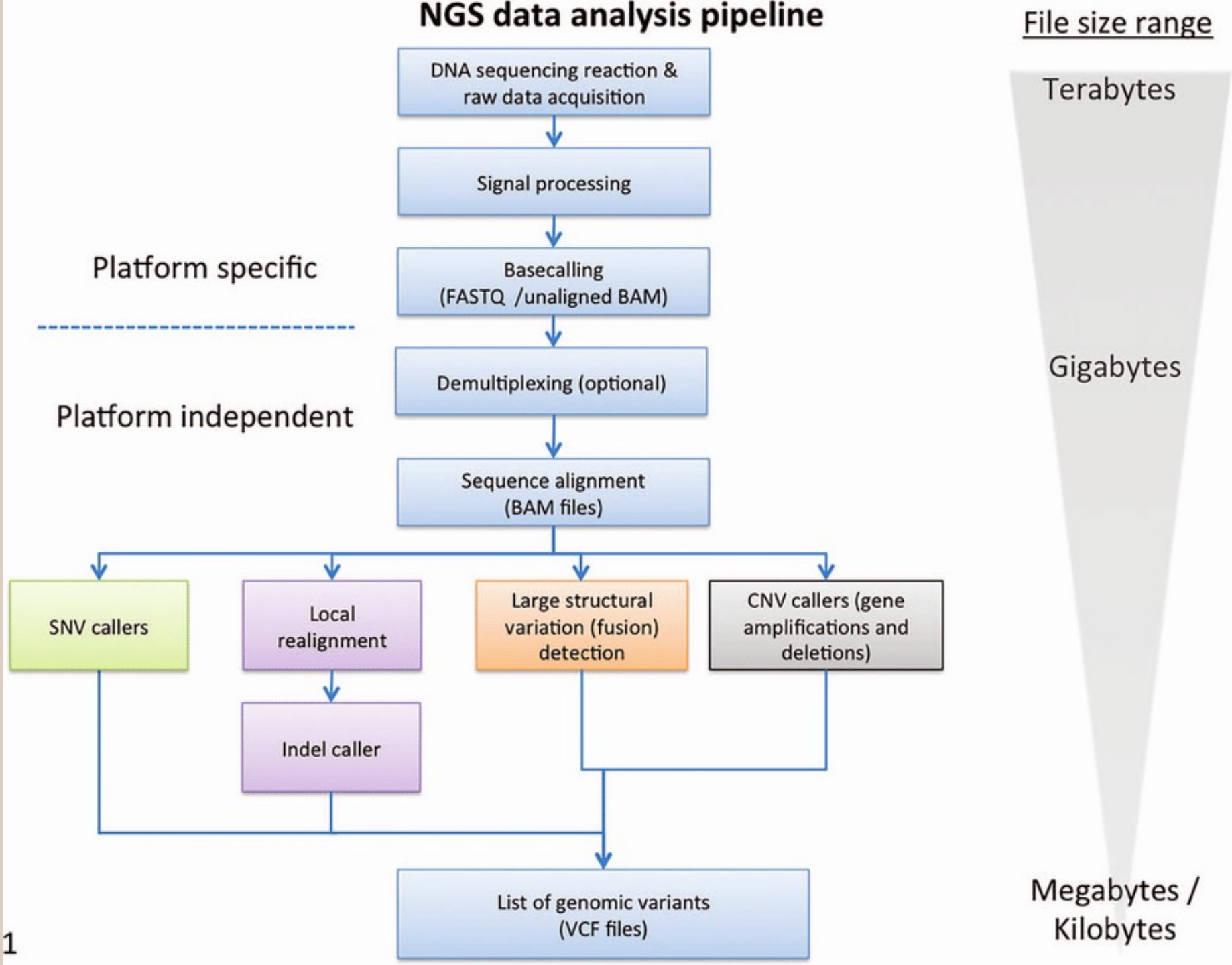


Fig. 1. Common workflows in genomics

Subramaniyan, Arun, et al. "Genomicsbench: A benchmark suite for genomics." *2021 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS)*. IEEE, 2021.

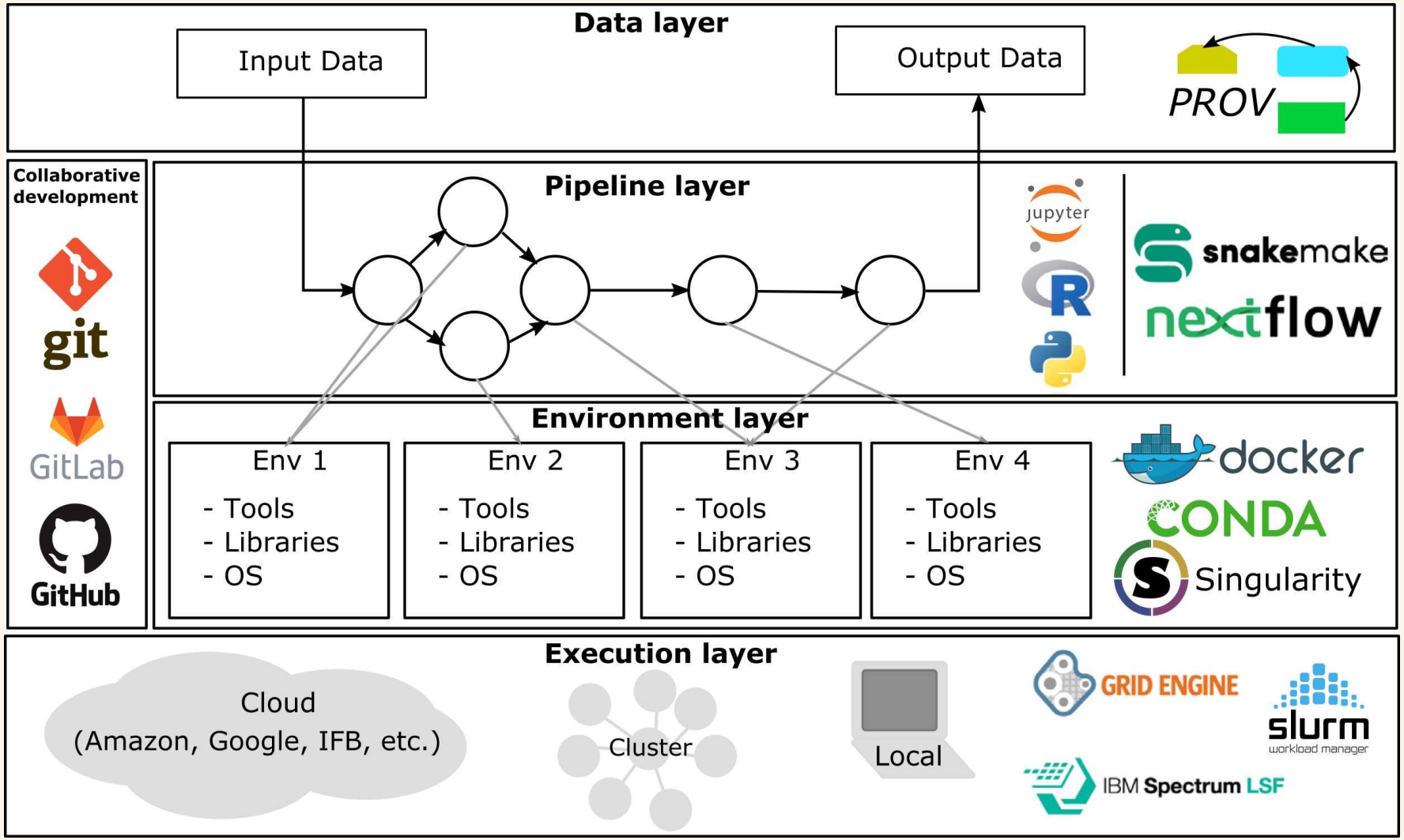


Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3:e1165 https://doi.org/10.7717/peerj.1165



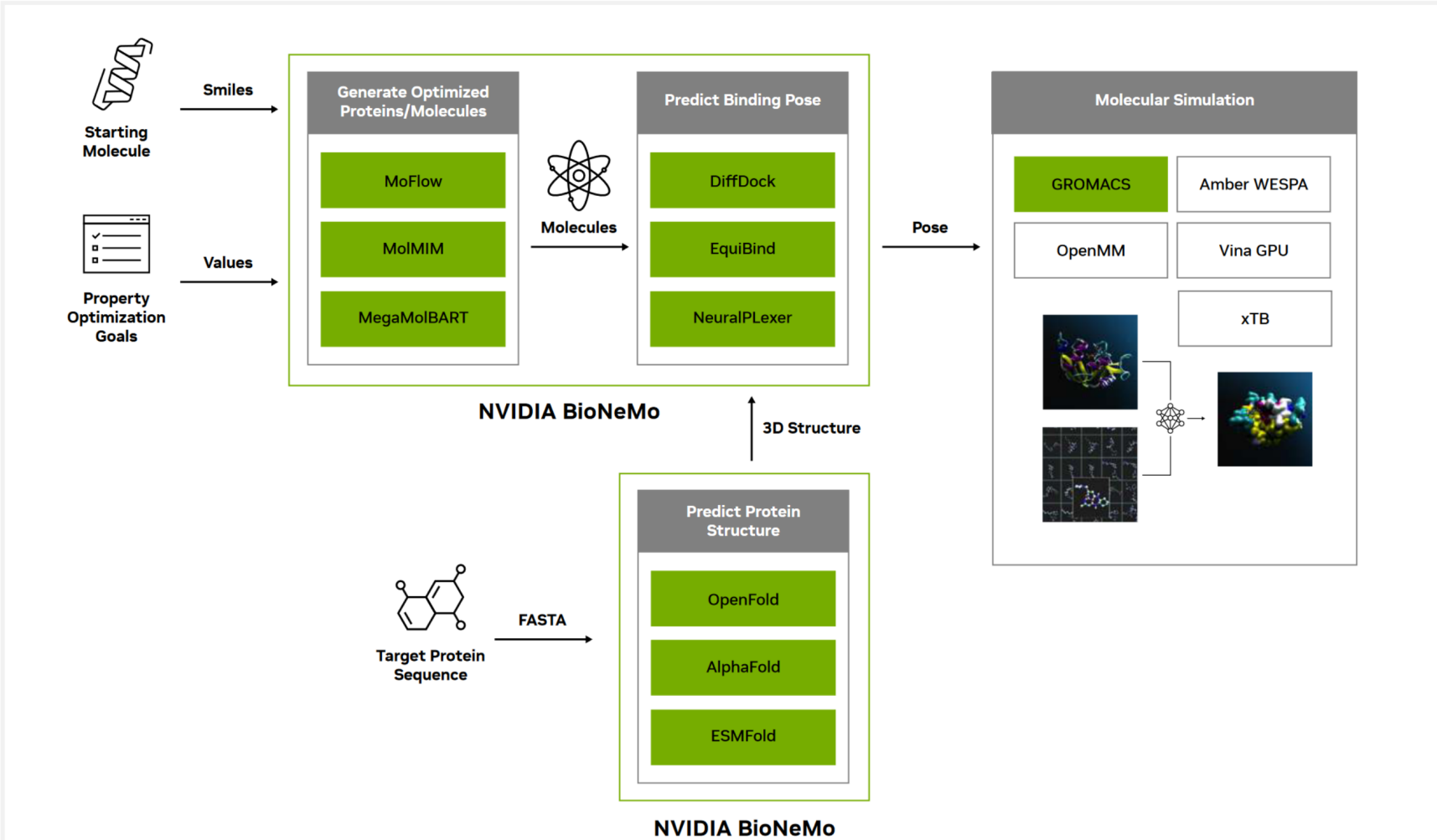
Roy, Somak, et al. "Next-generation sequencing informatics: challenges and strategies for implementation in a clinical environment." *Archives of pathology & laboratory medicine* 140.9 (2016): 958-975.

Zinātniskās darba plūsmas pārvaldības sistēma



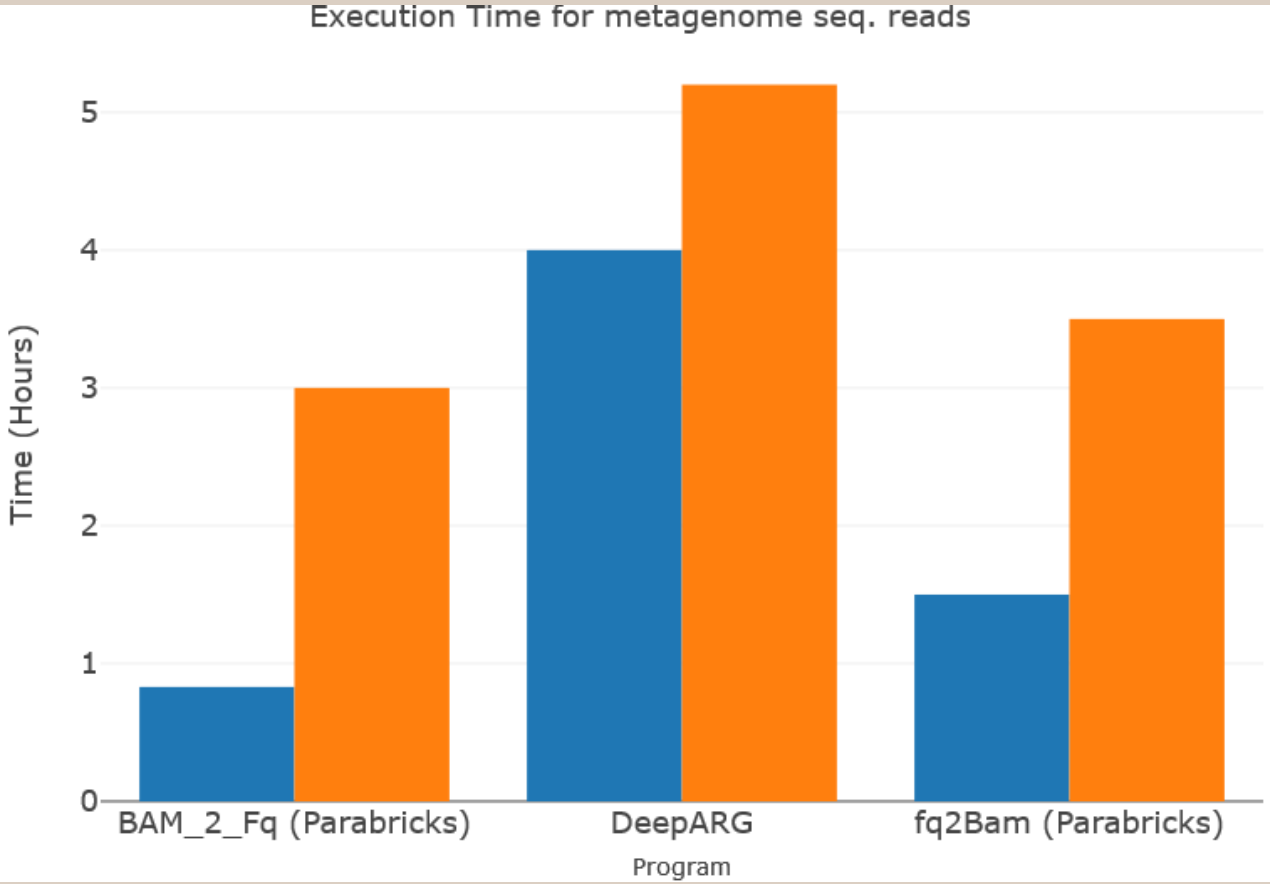
Djaffardjy, Marine, et al. "Developing and reusing bioinformatics data analysis pipelines using scientific workflow systems. " Computational and Structural Biotechnology Journal 21 (2023): 2075-2085.

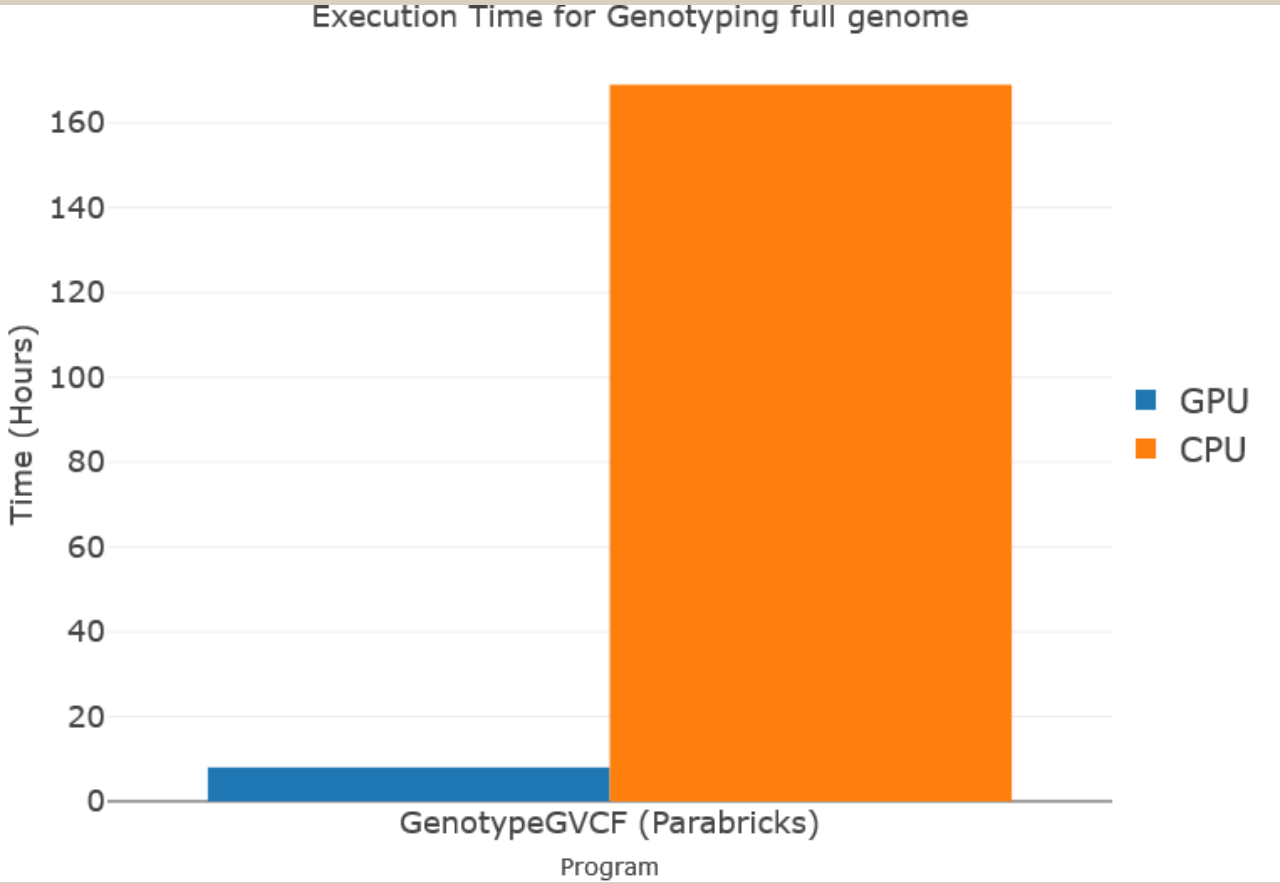
Small Molecule Discovery
Discover drugs faster with small-molecule virtual screening accelerated by molecular generative AI modelsa

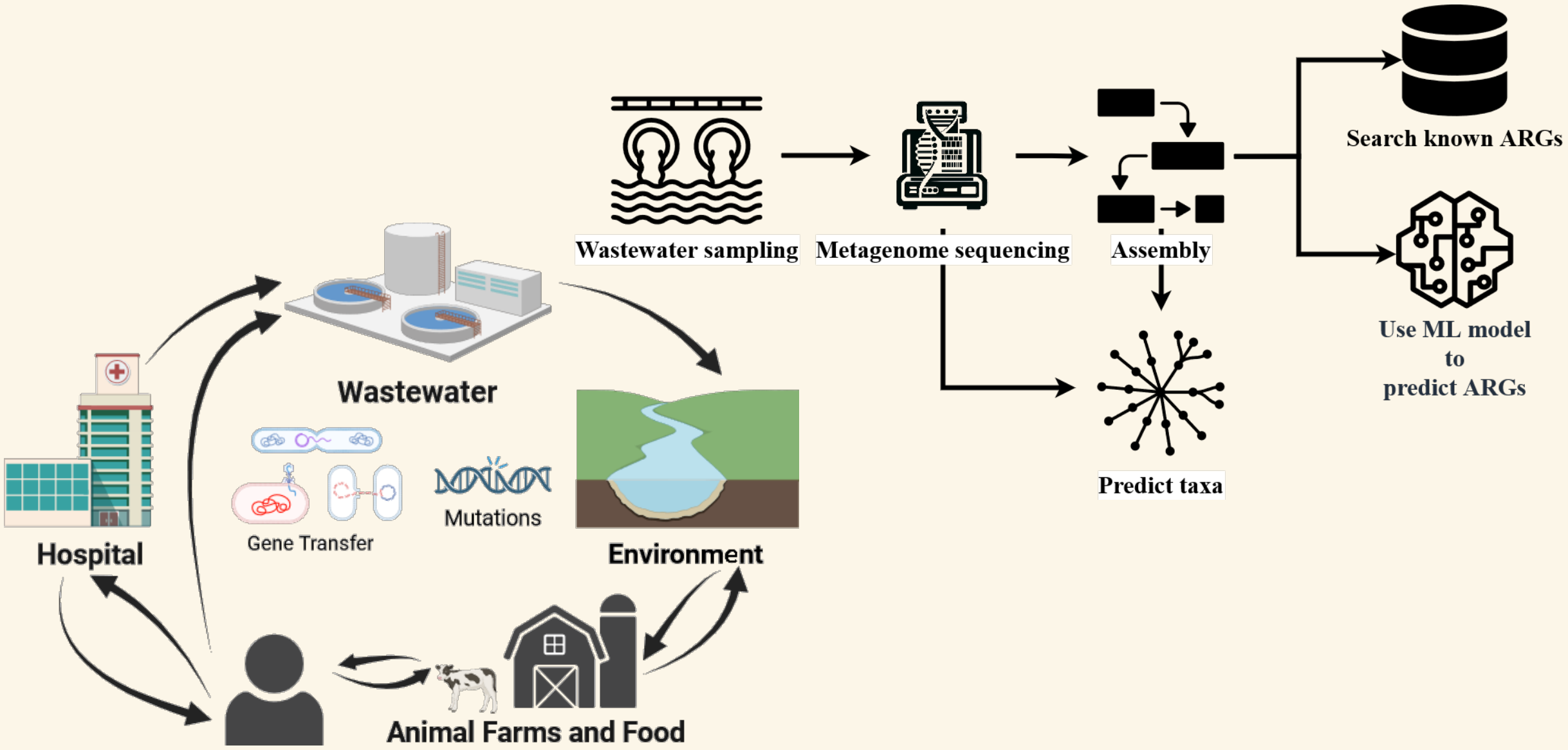


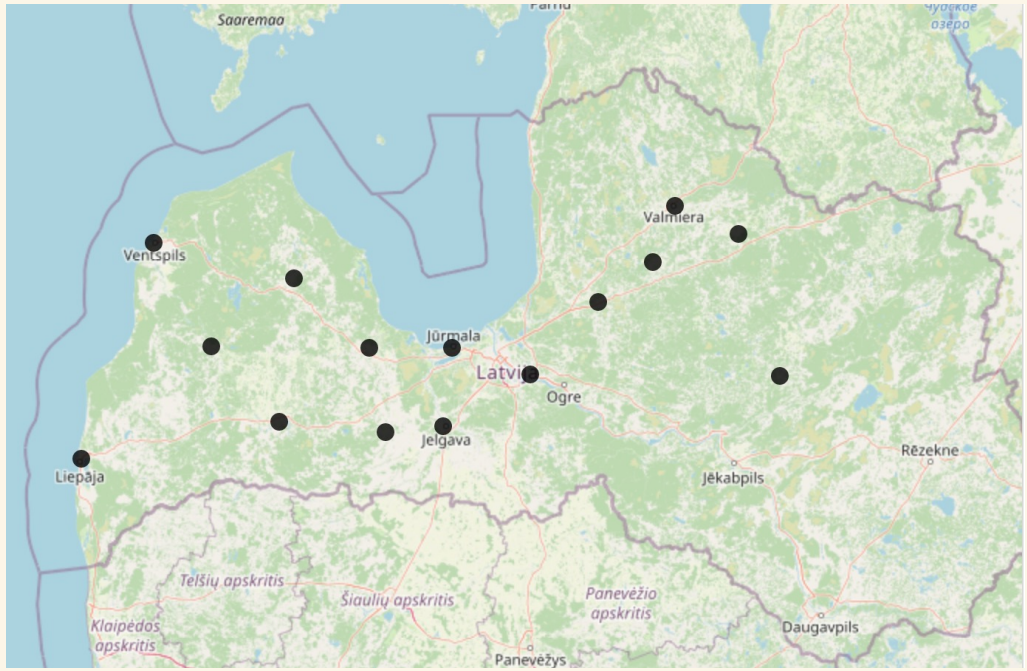
Attēls: https://resources.nvidia.com/en-us-hc-biopharma/hc-solution-overview-5

GPU runtime comparison on HPC

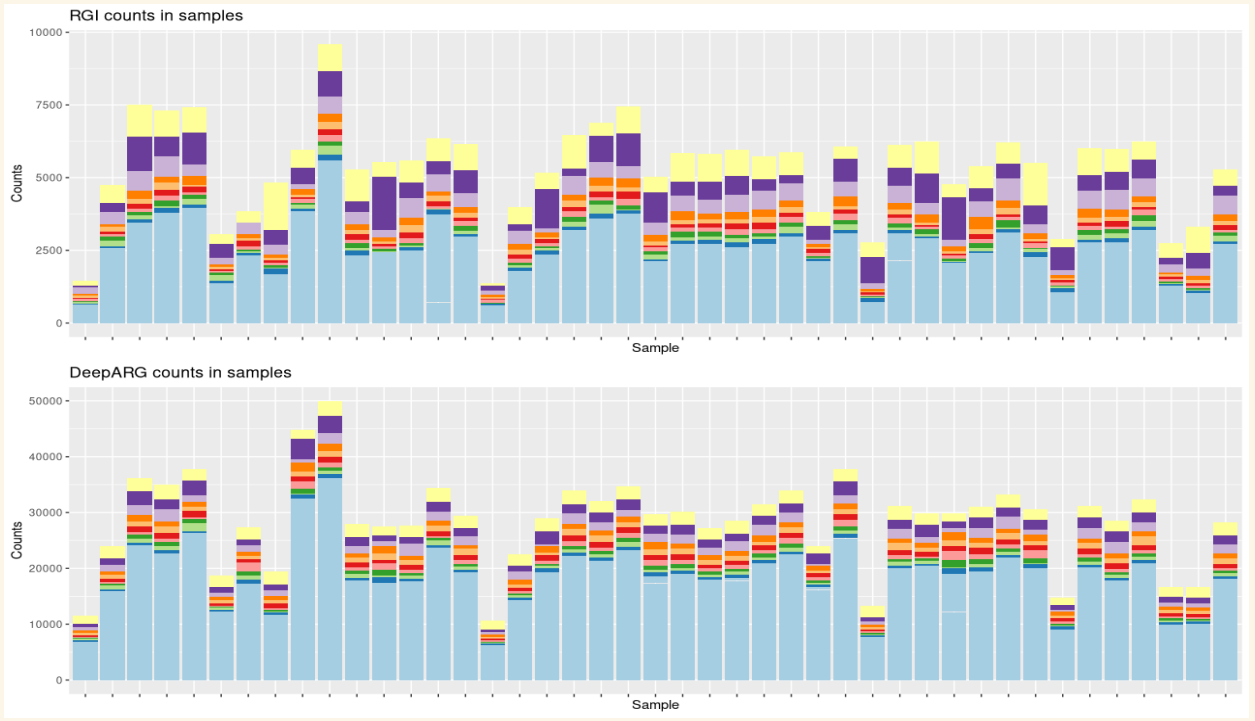




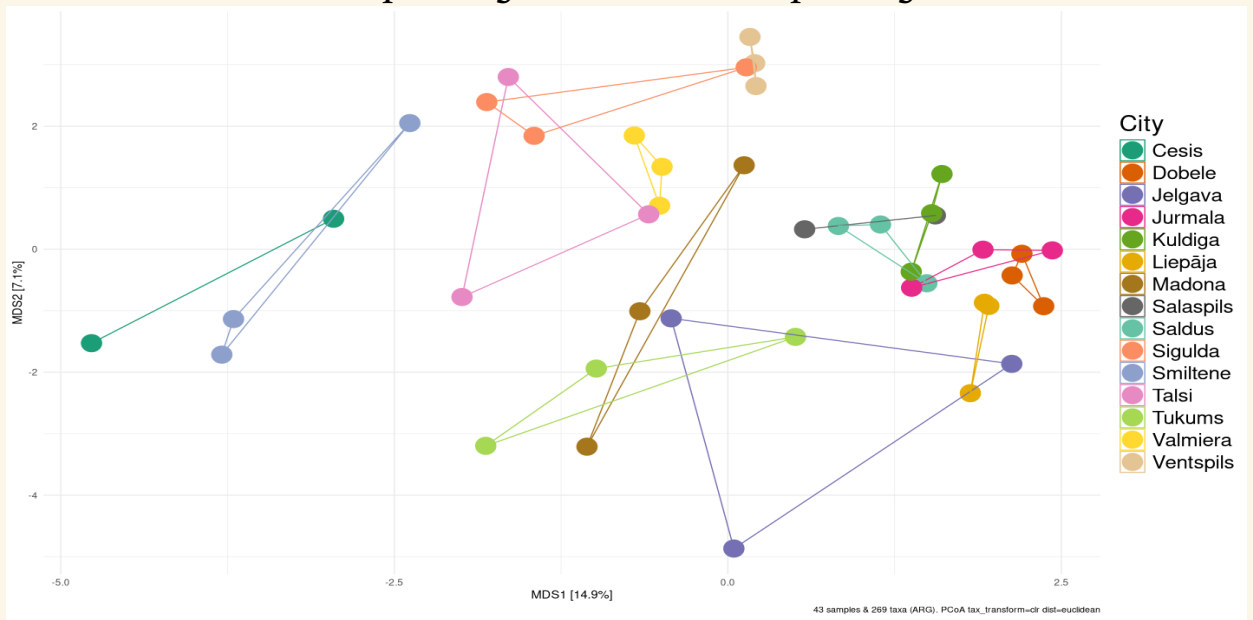

Wastewater sampling
Metagenome sequencing
Assembly
Search known ARGs
Use ML model
to
predict ARGs
Predict taxa
Wastewater
Hospital
Gene Transfer
Mutations
Environment
Animal Farms and Food



Att.: *Paraugu ievākšanas vietas no notekūdeņu stacijām Latvijas pilsētās*



Att.: *10 biežāk sastopamo gēnu daudzums paraugos*



Att.: *Rezistences gēnu beta diversitātes grafiks pa pilsētām*

Metagenoma sekvencēšana atklāj antibiotiku rezistento gēnu (ARG) un baktēriju sastāvu slimnīcas vidē.

Karbapenemāzi producējošas ***Klebsiella pneumoniae*** izolāti no:

- 6 pacientiem intensīvās terapijas nodaļās
- 10 izlietnēm intensīvās terapijas nodaļās
- 2 notekūdeņu paraugiem.

Metagenoma paraugi no:

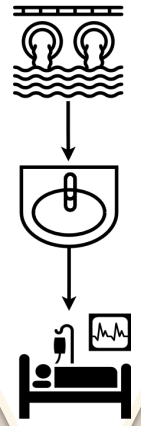
- 10 izlietnēm divās intensīvās terapijas nodaļās
- Slimnīcas ēkas notekūdeņu paraugi.

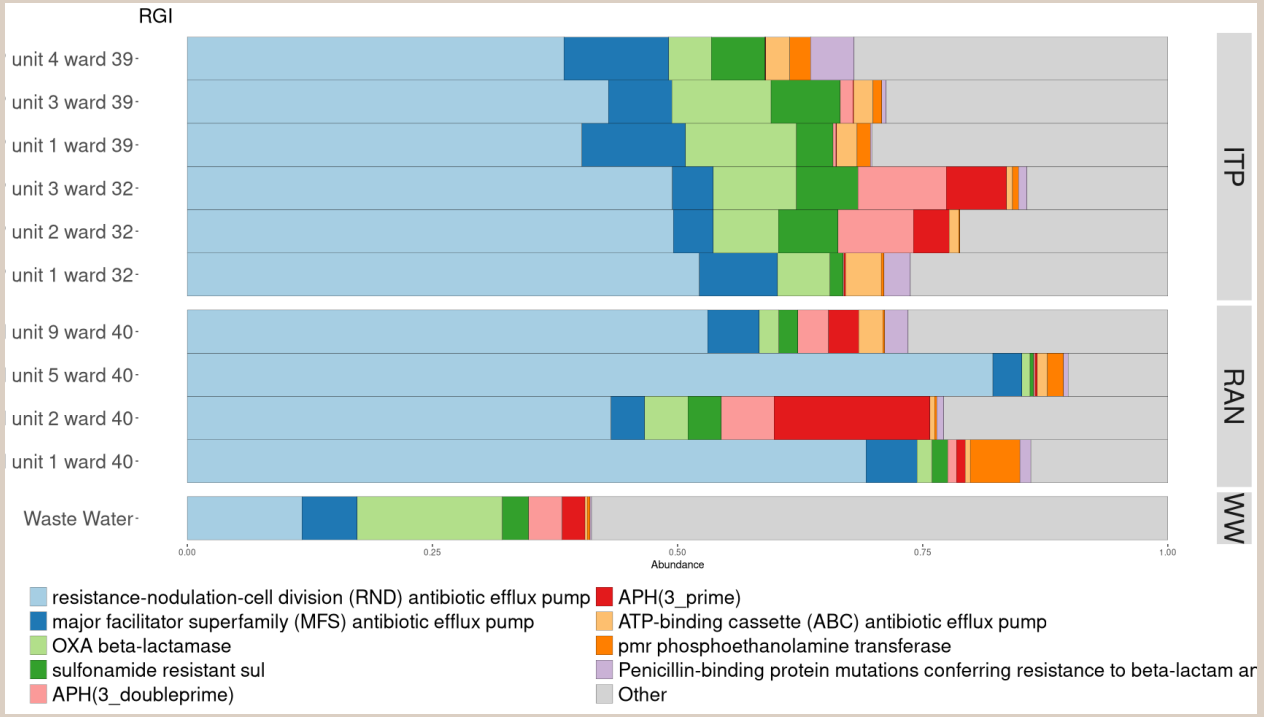
- Kopā identificēti **118 dažādi rezistences gēni**

- 94% no gēniem pacientu izolātos atrodami arī metagenoma datos.

- No **66 gēniem** pacientu izolātos **24** tika atrasti visos paraugos, **32 vismaz trīs paraugos**.

- 15 **karbapenemāzes rezistenci gēni** no kuriem 6 gēni tika atrasti visos izolētajos paraugos.





Attēls: *Biežāk sastopamās antibiotiku rezistences gēnu grupas slimnīcu nodaļās un notekūdeņos*

1) Noteiktie ARG parāda **visu baktēriju aizsardzības mehānismu kopumu**
2) Pacientos atrastajās baktēriju pilno genomu sekvencēs tika atrasti **tie paši gēnu veidi**, kas izlietnēs un notekūdeņos.
3) Pacientos atrastajās baktērijās novērojama līdzīgu rezistences mehānismu uzkrāšanās.
4) AI rīku pielietošana uzrāda salīdzināmus rezultātus,

Projekta Nr.: VPP-EM-BIOMEDICĪNA-2022/1-0001

Paldies par uzmanību