High-Performance Computing in Precision Medicine

Maksym Turkot

OUTLINE

01
INTRODUCTION
Motivation and importance

02
BACKGROUND
Genomic analysis, data generation, read mapping

03
INDEXING & MOVEMENT
Indexing acceleration and data movement

04
GPU ACCELERATION
NVIDIA Parabricks and benchmarks

05
CLOUD COMPUTING
Collaboration and performance

INTRODUCTION

Introduction

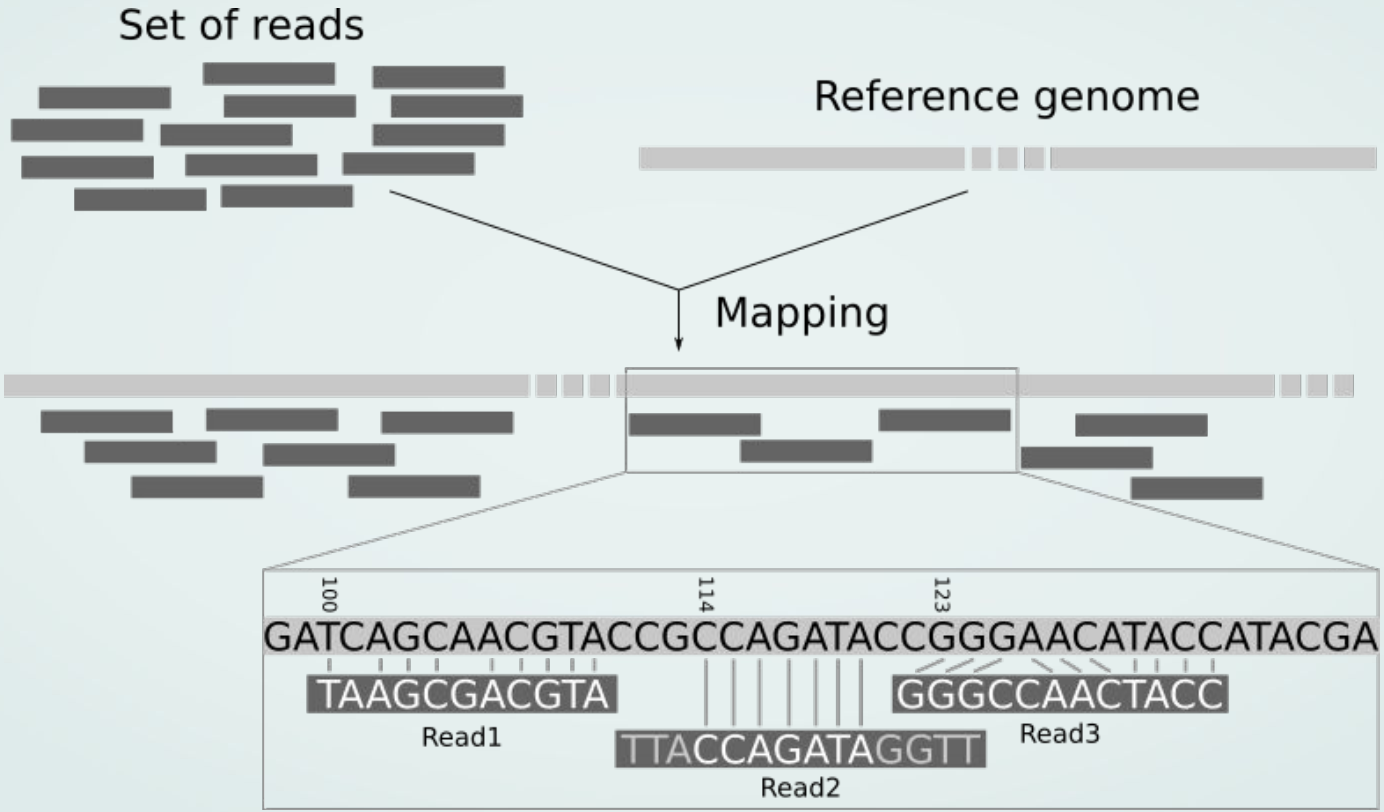
- HPC in Precision Medicine
- Shifting Bottleneck in Genomic Research
- Acceleration Techniques
  - Algorithmic (indexing)
  - Hardware (GPUs)
- Cloud Computing in Genomic Workflows

BACKGROUND

Obtaining Genomic Data

- Generation
  - Machine-based, sequencer
- Download
  - Prepared data
- Simulation
  - Mimic sequencer. Useful

Read Mapping

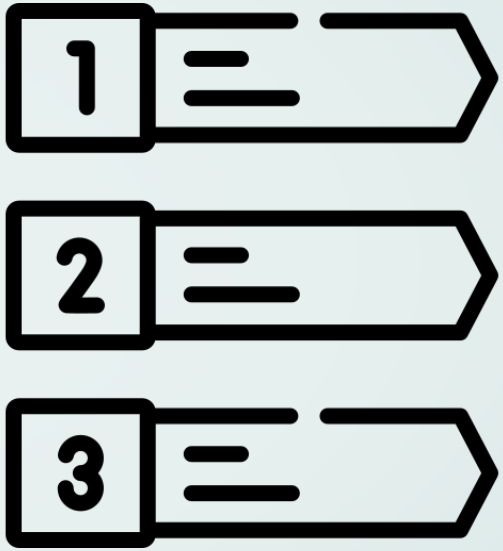


https://training.galaxyproject.org/training-material/topics/sequence-analysis/tutorials/mapping/tutorial.html

INDEXING & MOVEMENT

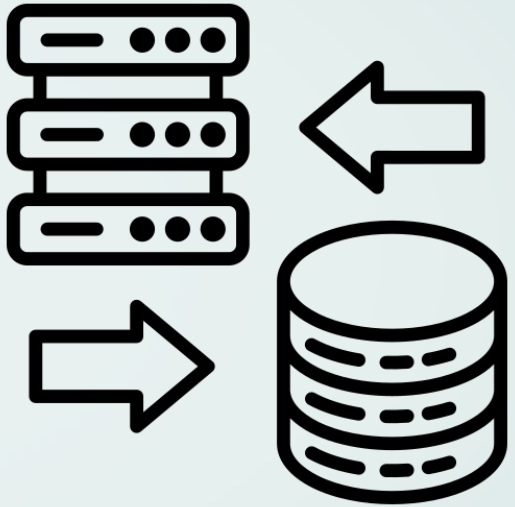
Indexing and Seeding Acceleration

- Indexing and Seeding
- Indexing Process
- Impact of Seed Characteristics
- Tradeoff and Improvements



Data Movement Reduction

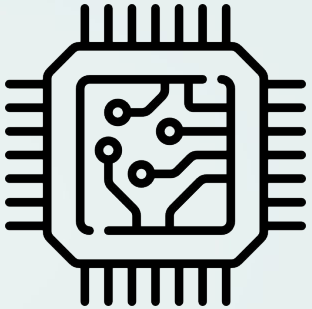
- Memory-Intensive Tasks
- Processing-in-Memory (PIM)
- Alternative Approaches
- 3D ReRAM-based Memory

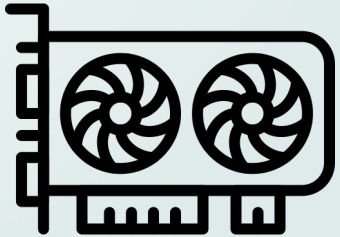


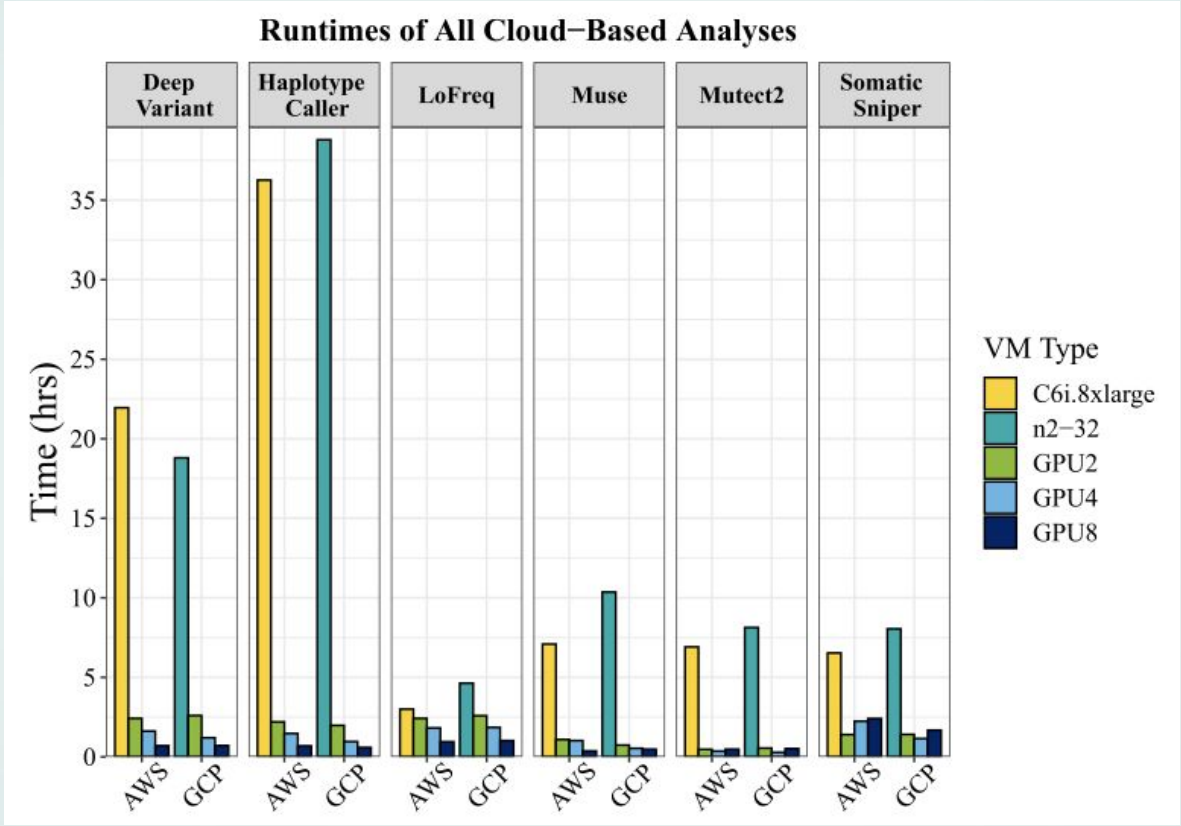
GPU ACCELERATION

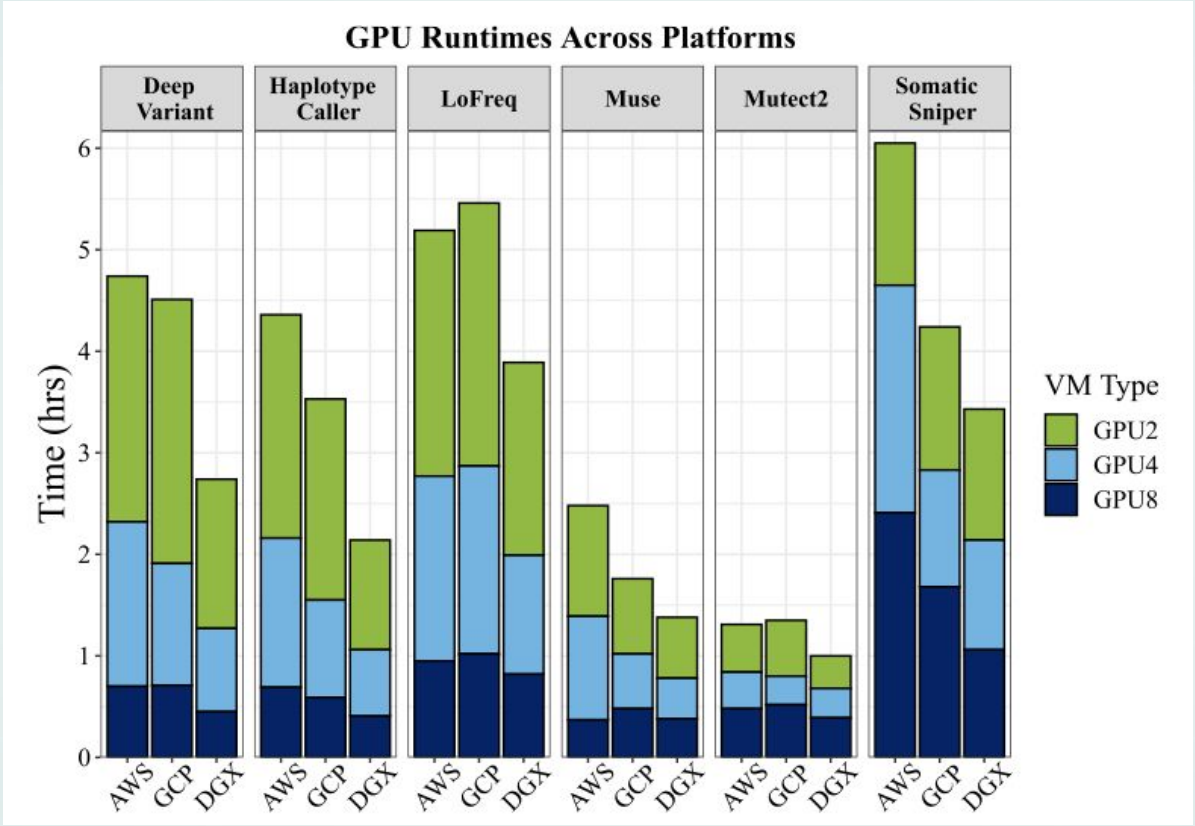
GPU Acceleration

- Data Analysis Bottleneck
- NVIDIA Parabricks
- Varying CPU performance
- Massive GPU acceleration





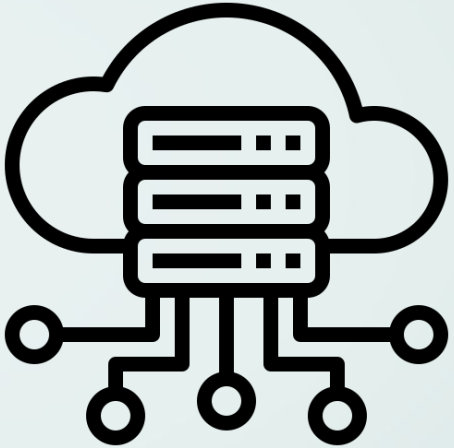

Runtimes of All Cloud−Based Analyses
Deep Variant
Haplotype Caller
LoFreq
Muse
Mutect2
Somatic Sniper
Time (hrs)
0
5
10
15
20
25
30
35
AWS
GCP
VM Type
C6i.8xlarge
n2−32
GPU2
GPU4
GPU8


GPU Runtimes Across Platforms
Deep Variant
Haplotype Caller
LoFreq
Muse
Mutect2
Somatic Sniper
Time (hrs)
0
1
2
3
4
5
6
AWS
GCP
DGX
VM Type
GPU2
GPU4
GPU8

CLOUD COMPUTING

Cloud Computing

- An Invaluable Tool
- Reproducibility & Accessibility
- Simplified setup
- Collaboration
- Advantage in genome work
- Scaling



References

[1] M. Alser, Z. Bing¨ ol, D. S. Cali, J. Kim, S. Ghose, C. Alkan, and O. Mutlu, “Accelerating genome analysis: A primer on an ongoing journey,” IEEE Micro, vol. 40, no. 5, pp. 65–75, 2020.

[2] A. Taylor-Weiner, F. Aguet, N. J. Haradhvala, S. Gosai, S. Anand, J. Kim, K. Ardlie, E. M. Van Allen, and G. Getz, “Scaling computational genomics to millions of individuals with GPUs,” Genome Biology, vol. 20, pp. 1–5, 2019.

[3] K. A. O’Connell, Z. B. Yosufzai, R. A. Campbell, C. J. Lobb, H. T. Engelken, L. M. Gorrell, T. B. Carlson, J. J. Catana, D. Mikdadi, V. R. Bonazzi et al., “Accelerating genomic workflows using NVIDIA Parabricks,” BMC Bioinformatics, vol. 24, no. 1, p. 221, 2023.

[4] B. Langmead and A. Nellore, “Cloud computing for genomic data analysis and collaboration,” Nature Reviews Genetics, vol. 19, no. 4, pp. 208–219, 2018.

[5] M. Alser, J. Lindegger, C. Firtina, N. Almadhoun, H. Mao, G. Singh, J. Gomez-Luna, and O. Mutlu, “From molecules to genomic variations: Accelerating genome analysis via intelligent algorithms and architectures,” Computational and Structural Biotechnology Journal, vol. 20, pp. 4579–4599, 2022.

Thank You!

Question time

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**