Mahidol HPC-AI Cluster

Introducing NVIDIA Clara Parabricks

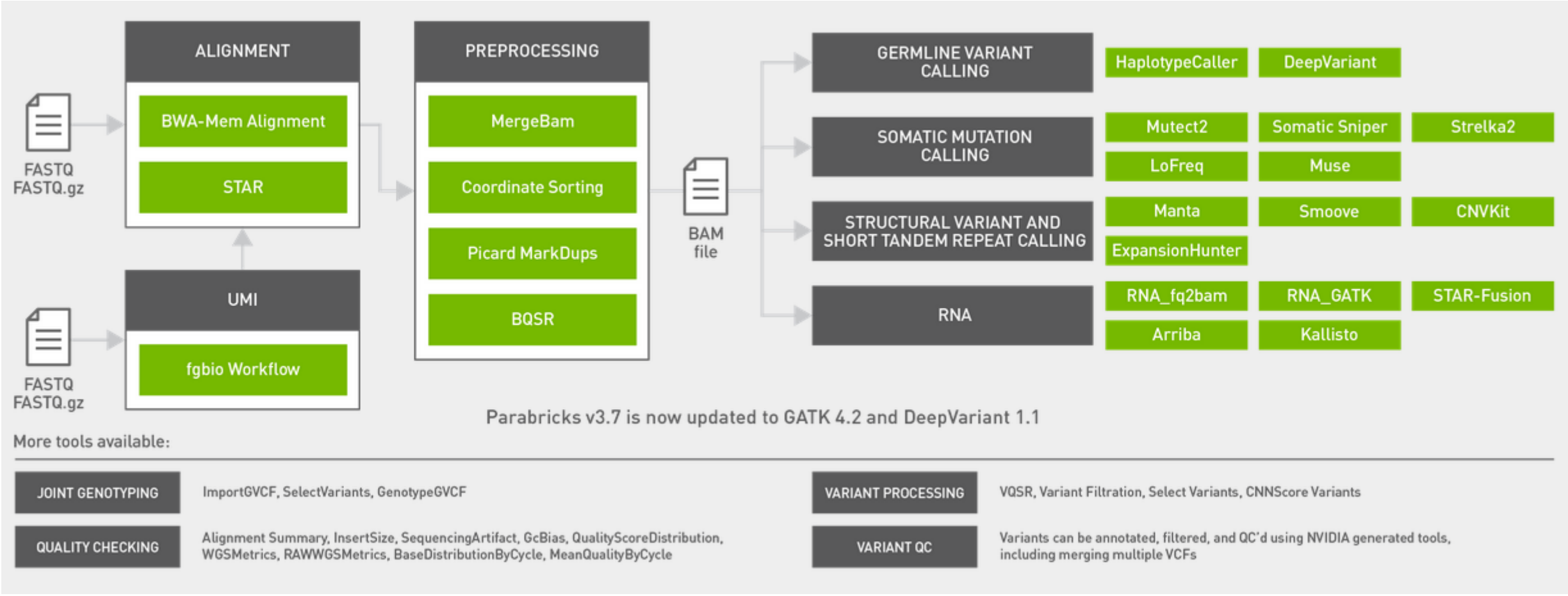
Harald Grove
harald.gro@mahidol.edu

Mahidol University

Accelerated Variant Calling

- NVIDIA has released a software to speed up variant calling
- The name is Clara Parabricks and it is using GPU's to massively parallelize variant calling pipelines and reduce runtimes.
- Processing a WGS dataset on a reasonably sized CPU-based cluster could easily take more than one day.
- With GPU acceleration, a single WGS dataset can be fully analyzed in less than 2 hours.

NVIDIA Clara Parabricks provides the following options for variant calling



https://www.nvidia.com/en-us/clara/genomics/

Predicted time savings by switching from CPU to GPU based processing

Somatic

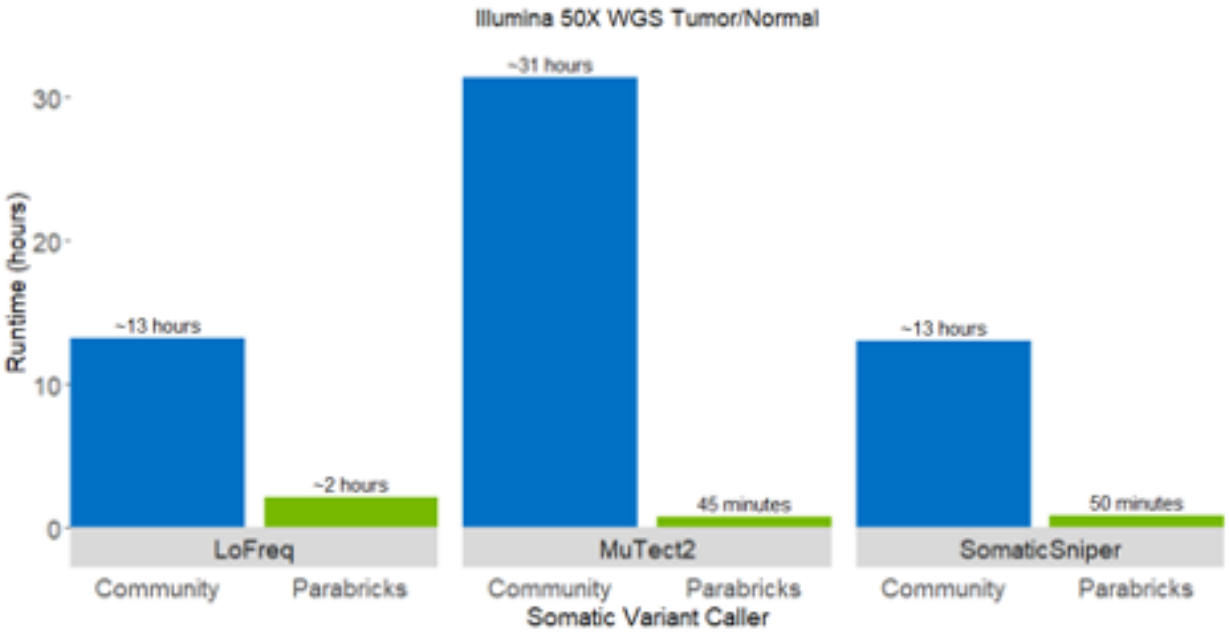


Figure 1: Analysis runtimes for open-source CPU-based somatic variant calling tools compared to GPU-accelerated NVIDIA Clara Parabricks. Relative to the community versions, NVIDIA Clara Parabricks accelerates LoFreq by 6x, SomaticSniper by 16x, and Mutect2 by 42x. These benchmarks were run on 50X WGS matched tumor-normal data from the SEQC-II benchmark set on 4x V100s.

Germline

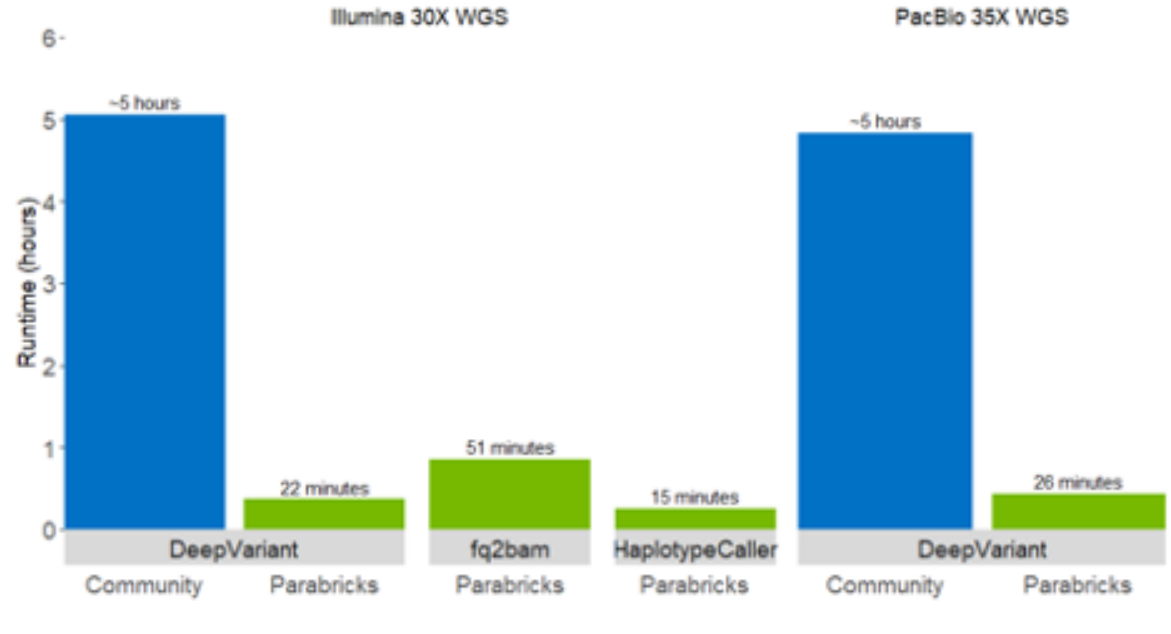


Figure 2: Runtimes for open-source DeepVariant (blue) and GPU-accelerated NVIDIA Clara Parabricks (green). Runtimes for 30X Illumina short read data are on the left; runtimes for PacBio 35X long read data are on the right. NVIDIA Clara Parabricks' DeepVariant is 10-15x faster than the open-source version (blue "DeepVariant" bars compared to green "DeepVariant" bars).

https://developer.nvidia.com/blog/gpu-accelerated-tools-added-to-nvidia-clara-parabricks-v3-6-for-cancer-and-germline-analyses/

Mahidol HPC-AI cluster

- Mahidol IT has set up an HPC cluster with 3 servers.
- **CPU specifications**
  - **ICT-HPC server specifications**
    - IP address: 10.134.1.9
    - Vendor id: AuthenticAMD
    - Model name: AMD EPYC 7742 64-Core Processor
    - Operating System: Ubuntu 20.04.2 LTS (focal)
    - CPUs: 256
    - Memory (RAM): 2TB
- **GPU specifications**
  - **ICT-HPC server specifications**
    - 1 x DGX A100 Server with 8 x NVIDIA A100-80GB GPUs
    - 2 x DGX A100 Server with 8 x NVIDIA A100-40GB GPUs

Getting started with Parabricks

- We have made a github repository
  - https://github.com/si-medbif/AI-MD-variant-calling
- The repository contains scripts, documentation and tutorials to get started running the parabricks pipelines with minimal previous knowledge