HwGS: Raw Reads to Variant Calls in Less Than an Hour
Mehrzad Samadi, Ankit Sethia, and Timothy Harkins (Ann Arbor, MI)

**Introduction**

- 35-50 times faster
- 100% reproducible and deterministic
- 40 genomes/day using 1 server
- 100% similar result as GATK4
- $3 per genome

**Pipeline**

- Alignment
- Preprocess
- Variant Caller
- Postprocess

**Performance**

- Execution time (Minutes)
  - CPU Server
  - 8 GPUs

**Third Party Case Study I**

fastq to vcf

- Speedup over CPU-only solution with 1,2,4,8 GPUs
- Accuracy compared to CPU-only solution

**Third Party Case Study II**

fastq to gvcf with 4 GPU for each sample

- Execution time (Minutes)
  - Min
  - Average
  - Max

**Deep Learning in Genomics**

- Google Brain team developed DeepVariant, deep-learning based variant caller.
- 3-fold error reduction compared to GATK4
- Variant Caller can be trained
- Variant calling phase can use GPUs to accelerate but data preparation and post-processing limits the performance
- We reimplemented all steps to use GPUs and improve the performance by more than 20X
- 40X whole genome in 25 mins