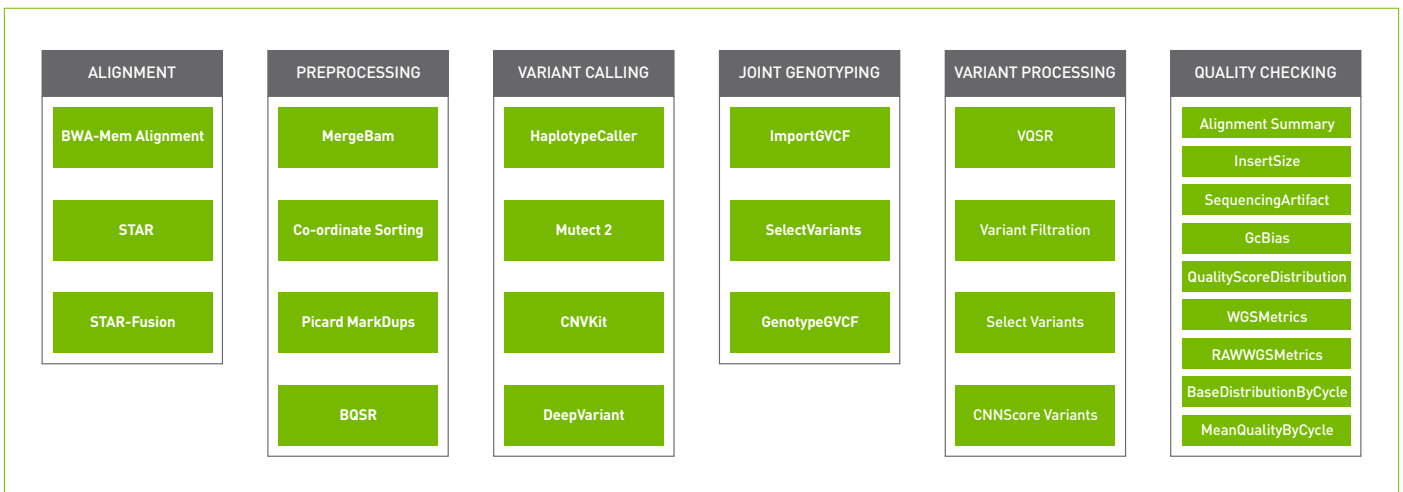# NVIDIA CLARA PARABRICKS PIPELINES FOR GENOMICS ANALYSIS

NVIDIA Clara Parabricks Pipelines provides 30-60 times faster secondary analyses of sequencer-generated FASTQ files to variant call files (VCFs). Additionally, Clara Parabricks Pipelines achieves equivalent results to that of common secondary analysis tools like GATK4 and DeepVariant, while increasing throughput significantly.

Providing Unmatched Secondary Analysis Performance and Throughput

Clara Parabricks Pipelines uses graphics processing units (GPUs) to accelerate secondary data processing. Current analyses take nearly 30 hours of computation on a 32-vCPU machine. Clara Parabricks Pipelines accelerates these computations, completing analyses in 20 minutes using 8 NVIDIA A100 GPUs, while implementing the exact same algorithms as GATK4. Clara Parabricks Pipelines can run the full GATK4 Best Practices, but it's also fully configurable. As a result, you can choose which steps and which versions of the pipeline to run.

NVIDIA Clara Parabricks Pipelines Accelerated Tools

ALIGNMENT	PREPROCESSING	VARIANT CALLING	JOINT GENOTYPING	VARIANT PROCESSING	QUALITY CHECKING
BWA-Mem Alignment	MergeBam	HaplotypeCaller	ImportGVCF	VQSR	Alignment Summary
STAR	Co-ordinate Sorting	Mutect 2	SelectVariants	Variant Filtration	InsertSize
STAR-Fusion	Picard MarkDups	CNVKit	GenotypeGVCF	Select Variants	SequencingArtifact
	BQSR	DeepVariant		CNNScore Variants	GcBias
					QualityScoreDistribution
					WGSMetrics
					RAWWGSMetrics
					BaseDistributionByCycle
					MeanQualityByCycle

The Parabricks Advantage

HIGH THROUGHPUT

On a single server, the software can process up to 40 whole genomes per day

30–60 TIMES FASTER

By running 30–60 times faster, Clara Parabricks Pipelines can reduce computing costs up to 50% compared to CPU-only solutions.

BETTER ACCURACY

No need to sacrifice accuracy for reducing turn-around time. Deep-learning based tools accelerated by Clara Parabricks Pipelines increase accuracy.

DETERMINISTIC REPRODUCTIBLE

Any configuration of the software on any platform generates the exact same results every time for a given input.

FLEXIBLE PIPELINE

Create customized-accelerated pipelines by connecting Clara Parabricks Pipelines in a configurable way.

AI AND ML INTEGRATION

GPU ecosystem is designed for deep learning and integrates with state of the art AI and ML frameworks and libraries instantly.

Reduced Computing Costs

By using GPU-accelerated computing, Clara Parabricks Pipelines can provide throughput comparable to nearly 40-50 CPU servers with one GPU server, reducing IT management overhead and operating costs (including power and cooling). Whether it's analysis or re-analysis of NGS data, Clara Parabricks Pipelines is rapidly evolving as a preferred solution, providing current pipelines with vast improvements in terms of efficiency, while also enabling user-driven customization.

For users on the cloud, computing costs are proportional to execution time. By reducing runtime by a factor of 30–60X, Parabricks reduces total computing costs by up to 50% compared to CPU-only solutions.
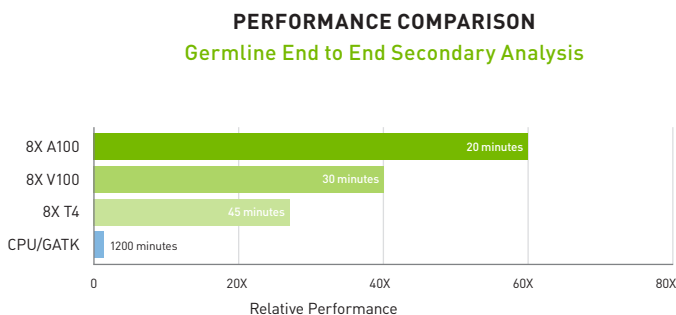
High-Performance Solution

The output BAM file after Alignment, Sorting, Marking Duplicates, and Apply BQSR stage is identical to the baseline. The baseline variant caller is nondeterministic and can generate slightly different results based on certain parameters. For this step, the result generated by the software is within 99.999% of the baseline execution.

NVIDIA GPU RECOMMENDATIONS	
Optimized for scale out performance	NVIDIA T4
Optimized for fastest turn around time	NVIDIA A100

SUPPORT FOR AN ANNUAL PER NODE LICENSE	
Full access to all pipelines in the NVIDIA Parabricks software suite	✓
No limitations on the number of genomes analyzed	✓

PERFORMANCE COMPARISON
Germline End to End Secondary Analysis

8X A100	20 minutes
8X V100	30 minutes
8X T4	45 minutes
CPU/GATK	1200 minutes

Relative Performance: 0, 20X, 40X, 60X, 80X

Learn more at **www.nvidia.com/clara-parabricks**