

Genomics with NVIDIA Clara Parabricks

Accelerate secondary analysis for next-generation sequencing

200 gigabytes

in a single human genome sequence or the space of about 200 copies of *Jaws*²

40

exabytes to store genome-sequence data generated worldwide by 2025²

6 billion

DNA letters in a cell as most cells have two copies of the genome²

A team of Stanford scientists recently set the first Guinness World Record for the fastest DNA sequencing technique, which was used to sequence a human genome in just 5 hours and 2 minutes.¹ Faster diagnoses means patients can recover more quickly and spend less on care.

Once a genome is sequenced, identifying differences between genomes is instrumental in assessing the response to diagnosis, treatment and disease prevention strategies. The secondary analysis phase of next generation sequencing (NGS) is the process of comparing an individual's genome to a DNA reference sequence to generate a list of variants. This process can take minutes or days, depending on the available software, computing and storage resources. When you're talking about the difference between life and death, a few days can be too long to wait. Having the secondary analysis resources to keep pace with the rate of raw NGS data generation is critical for preventing analysis backlogs.

That's why advanced computing technologies, such as High Performance Computing (HPC) and artificial intelligence (AI) have had such an impact on genomics and other life sciences fields. They grant researchers the power to work with vast amounts of data and get results quickly and with lower costs. While they have existed as separate technologies for many years, the fields are converging as the powerful, scalable compute, networking and storage that are provided by HPC are being used to empower data analytics and AI for genomics use cases. This convergence reshapes the ability of researchers to unravel the code of life.

But building an HPC system for genomics workloads can be challenging. Designing a system that provides the immense compute power and storage capacity required is a specialized task. Yet many organizations struggle to design their own HPC infrastructure, including complex integration and performance tuning, with little IT expertise or support.

Dell Technologies is dedicated to making HPC more accessible, with a portfolio of engineering-validated solutions designed to help researchers configure, deploy and manage HPC systems with speed, confidence and flexibility. System building blocks include the servers, storage, networking, software and services that have been proven in our labs and in customer deployments to meet workload requirements and customer outcomes.

¹ Hanae Armitage, Stanford Medicine, "[Fastest DNA sequencing technique helps undiagnosed patients find answers in mere hours](#)," January 12, 2022

² [National Human Genome Research Institute](#), accessed April 2, 2022

Resources

- [Validated Design](#)
- Engineering: hpcatdell.com and the [Dell InfoHub](#)
- Explore the [Dell Technologies HPC & AI Innovation Lab](#)
- Join the Dell HPC Community at DellHPC.org

Learn more

DellTechnologies.com/HPC

Genomics with NVIDIA Clara Parabricks

The Genomics for NVIDIA® Clara™ Parabricks solution uses a flexible and modular approach to HPC system design that leverages individual building blocks. These integrated, tested and tuned solutions include the resources required for NGS secondary analysis while providing an optimal balance of compute density, energy efficiency and performance.

The [Dell Technologies Validated Design for Genomics with NVIDIA Clara Parabricks](#) describes a modular, scale-out solution composed of NVIDIA Parabricks application software, Dell PowerEdge servers with NVIDIA GPUs, available with Dell PowerScale storage. Both Dell PowerEdge XE8545 and PowerEdge R7525 servers can process more than 20 50x WGS per day with NVIDIA Clara Parabricks.

In genomics, speed is essential, but so is compatibility with other analysis tools to enable comparability of results. The Parabricks germline analysis results are nearly identical to the well-known BWA-GATK Haplotype caller analysis from prior testing. Dell Technologies engineers also compared the Parabricks variant calling results to other toolsets like samtools/mpileup. These two different tools reached ~90% overall agreement for identified variants, and variations in many well-known genomic regions containing important genes agree more than 99%.

The team also provides configuration guidance based on NGS secondary analysis workloads. The options below serve as a starting point for a customized and engineering-validated HPC genomics solution.

Infrastructure compute	GPU compute	External storage	Networking	Software
PowerEdge R650	PowerEdge XE8545 with 4x NVIDIA A100 GPUs SXM4 with NVLink™ tested with R7525 servers with 2x NVIDIA A100 GPUs	PowerScale or HPC BeeGFS storage	NVIDIA QM8790 InfiniBand	NVIDIA Clara Parabricks Bright Cluster Manager (recommended)

If you're considering an HPC solution for NGS, Dell Technologies HPC experts are available to help you design an HPC solution for your specific needs. And [Dell Technologies Services](#) — ranging from consulting and education to deployment and support — are available when and where you need them.

Dell Technologies and NVIDIA

Dell Technologies and NVIDIA work together closely to deliver unprecedented acceleration and flexibility for AI, HPC and data analytics to help our customers tackle some of the world's toughest challenges.

