

SOLUTION BRIEF

Accelerate Genomics Analytics for Precision Medicine

Implement next-generation sequencing with Pure Storage®.

Genomics data is a critical tool for preventing diseases and designing individualized treatment plans in modern medical practices. With the advent of new analysis techniques like next-generation sequencing (NGS), precision medicine researchers can acquire exponential amounts of genomic data and define treatments in a fraction of the time, but only if their IT infrastructure keeps up. To keep pace with data growth and support concurrent sequencing workflows, pharma and biotech research firms are upgrading their IT infrastructures to deliver a quantum leap in speed, scale, and agility.

The Key to Modern Sequencing and Analytics

Genomic sequencing workflows start with the production of base-pair sequencing reads, and a single sequencer can produce 20 to 48 sequence reads per day, resulting in terabytes of data. Because most life-sciences organizations employ fleets of sequencing instruments, it is not uncommon to quickly accumulate petabytes of reads that are then assembled into whole genomes for biological data mining, visualization, and interpretation.

Assembling, mining, and interpreting sequences (called secondary and tertiary analysis) requires a high-performance compute environment built on a storage architecture that delivers low-latency IOPS and high throughput. In particular, secondary analysis demands high-performance storage to support metadata access and concurrency requirements. As the number of sequence reads grows, you need non-disruptive capacity scaling to prevent disruption of sequencing runs and scientific analysis.

Pure understands these needs and has a range of solutions designed to help your teams meet their unique requirements for speed, scale, and even purchasing agility—buy, lease, or use as a service—with unmatched simplicity and reliability.



Accelerate Analysis

Improve end-to-end sequencing performance and lower data latency - even up to 100s of millions of files.

Load genome sequence indexes in 1/3 the time.



Deliver Peace of Mind

Minimize downtime and disruption with snapshots, non-disruptive scaling, predictive support, and ransomware protection.



Increase Efficiency

Utility consumption pricing allows you to pay for only what they need and then scale on-demand for agility with greater budget control.

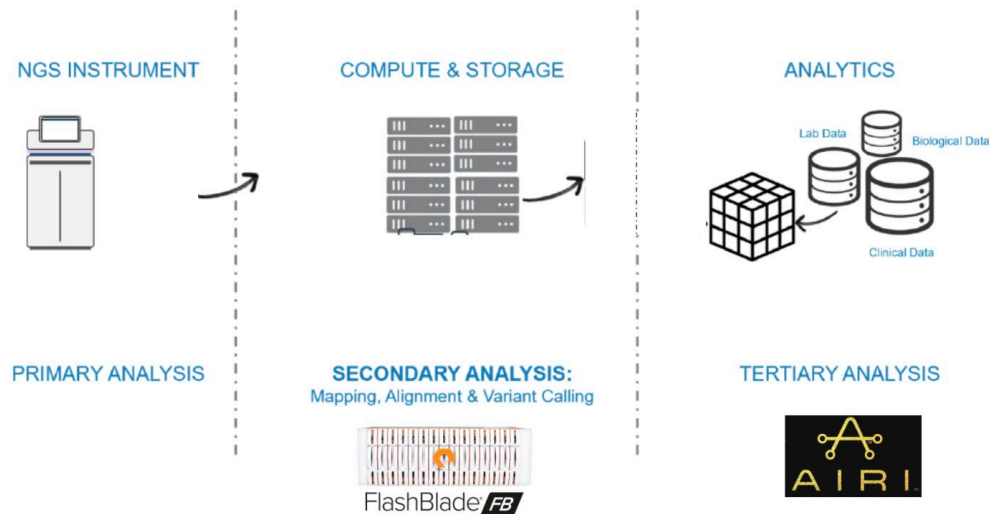


Figure 1. Next-generation sequencing workflow.

A Modern Data Experience™ for Genomics

At Pure Storage, we believe the best way to empower your researchers is through a Modern Data Experience. We built the modern data experience for genomics on four concepts embodied in our portfolio of advanced, all-flash solutions.

- Fast Matters:** The genomic sequencing and analysis workflows that drive biomedical discoveries demand storage that delivers high performance and low latency. We designed the FlashBlade™ array with this need in mind. Built on a custom, engineered flash architecture with independent scaling of performance and capacity, FlashBlade can deliver PBs of throughput and low latency access to hundreds of millions of files. For secondary and tertiary analytics processes including genome alignment and variant calling, FlashBlade architecture can accelerate processes by three to four times or more.
- Simple is Smart:** As concurrency increases and the amount of data from DNA and RNA sequencing grows exponentially, it is critical to have an architecture that can be easily scaled. Organizations must eliminate any complex configuration processes that delay access to researchers or result in downtime through human error. To help, FlashBlade automates internal network configuration and dynamically manages load balancing for better ongoing performance. To simplify support, Pure Storage offers AI-driven, proactive support as well as easiest-in-industry data protection functions like snapshots, disaster recovery, and ransomware protection. Combined, these functions identify issues before they are outages and protect you from business-crippling outages
- Cloud Everywhere:** The applications and processes used in life sciences are constantly evolving and IT infrastructures must evolve with them. To this end, we designed solutions to provide cloud-like economics and support containerized workloads. With Pure-as-a-Service, you can consume Pure Storage offerings via a true utility model for increased budget optimization and on-demand scaling. Support for containers, orchestration, and hybrid cloud configurations—via managed service providers and co-location providers—gives IT leaders the agility to change workflow design, resources provisioning, and where applications live without being locked into a specific location.

Subscription to Innovation: Pure Storage has consistently led the industry in our approach to customer relationships. When you buy from Pure Storage, you are entering into a relationship where we include controller upgrades, array software, and even flash storage upgrades so that you aren't burdened with hidden costs that might prevent needed upgrades to support the business. This ensures you have the most up-to-date tools you need to protect your genomics data and keep sequencing workflows running. And with a proven track record of non-disruptive upgrades, you can choose to update whenever you need without disrupting sequencing runs or disrupting time-sensitive, tertiary analysis.

AI-enabled Tertiary Analytics

With AIRI®, the world's first, full-stack AI-ready infrastructure, you can deliver the time-to-insight your scientists need to run tertiary analyses on genomic variations to understand their significance in the context of additional biological and clinical information. Jointly architected by Pure and NVIDIA, AIRI delivers a modern platform to accelerate end-to-end GPU workflows, delivering an out-of-box integrated solution built to accelerate all aspects of the data pipeline, from ingest to inference.

Additional Resources

- Learn more about [modern data protection](#) solutions.
- Get more information about [AIRI](#), [Pure 1](#), [Pure as-a-Service](#), and [Pure Professional Services](#).

Case Studies

- Learn how [UC Berkeley](#) researchers accelerate DNA sequencing, generate valuable new insights, and, in some cases, deliver faster answers to life-and-death questions.
- Barcelona's [Centre for Research in Agricultural Genomics](#) updates its HPC environment to improve speed, scalability, and stability.
- At [McMaster University](#) in Hamilton, Canada, scientists use power infectious disease and COVID-19 research.

purestorage.com

800.379.PURE

