PURESTORAGE®

Pure Storage for Life Sciences

Building a modern storage foundation to accelerate genomic clinical and research workflows

A primer on how storage impacts genomic research potential

Executive summary

Genomics has reached an inflection point. Next Generation Sequencing (NGS) is fueling breakthroughs at unprecedented speed, generating 40 exabytes[1] of data each year. Scientists are sequencing millions of genomes worldwide[2] and every decoded genome adds billions of data points.[3] The business opportunity is massive: faster drug discovery, precision medicine, and global research collaboration. But the challenge is just as significant: most IT infrastructures were never designed to handle the scale, speed, or regulatory burden of global sequencing projects.

In this dossier, we'll discuss how organizations that continue to rely on legacy infrastructure are at risk for lost productivity, missed insights, escalating costs, and compliance violations. We'll also show how shifting to modern, enterprise data cloud technology from Pure Storage can accelerate sequencing and discovery, strengthen collaboration, and ensure compliance worldwide.

NGS is transforming genomics into a data-driven, precision process

Sequencing costs have fallen dramatically in recent years, pushing genomics research forward in ways never imaged before. The total expense of genomics is still huge, however. The big, diverse data sets that genomics thrives on requires cloud storage and compute power that older IT infrastructure can't support.

Legacy IT systems were built for "smaller science," not the high-throughput, global, collaborative world of genomics of today. This infrastructure relies on outdated formats and standards that aren't compatible with modern bioinformatics pipelines, APIs, or cloud-based workflows. Research data is kept in silos, with different labs, departments, or institutions maintaining their own separate systems.

This approach significantly hinders collaboration. Clinicians, researchers, and analysts can't seamlessly access shared information, limiting the pace of discovery and decision making. Sharing large files is difficult, and incompatible systems can further slow down critical workflows. Once a file is received, individuals must spend valuable time reformatting it so it's usable. In addition to hindering research, legacy IT is unlikely to meet global compliance standards such as HIPAA (US) and GDPR (Europe), leaving organizations with multinational projects exposed.

Leaders who invest early in infrastructure can move past these technological and compliance obstacles—capturing the advantage while others stall out. Pure Storage is redefining the genomics data storage experience and simplifying how organizations consume and interact with data. Pure Storage flash storage products accelerate research while protecting data assets and intellectual property.

Preserving and storing genomics is not just industry practice, it's the law in many regions. With AI, researchers can do more with stored data.

40-50% reduction in time
to synthesize and screen new drugs, saving up to **$26 billion** in research costs per year.[4]

Cut cost and time
with AI tools that sequence and analyze a whole human genome.

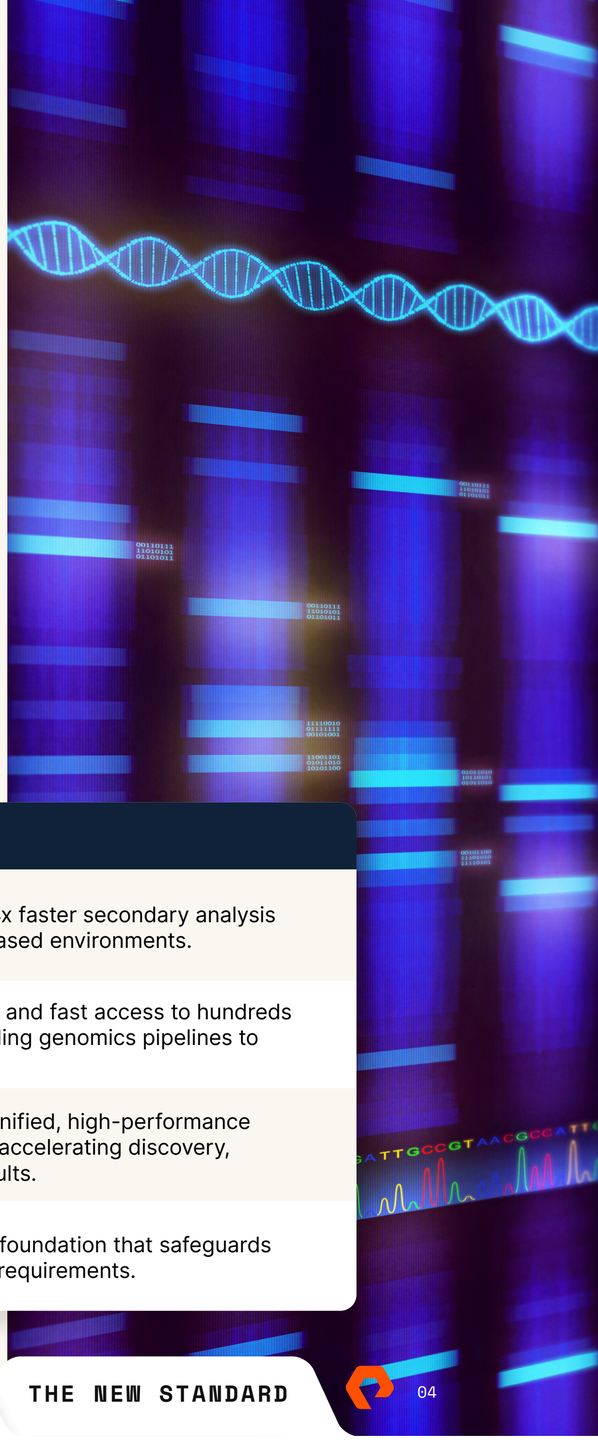
With AI, researchers can uncover patterns that humans alone may miss.

The case for eliminating shadow IT

Researchers solve problems—but sometimes, the solution can create more issues. For example, when researchers need storage, they may opt to buy a low-cost, off-the-shelf solution instead of calling IT. This DIY approach creates “shadow IT,” or technology that isn’t under IT’s control. What they don’t realize is that selecting a storage solution goes beyond capacity.

Lab-grade storage needs to offer superior speed, scalability, and resilience to support large volumes of data. Equally important are recovery metrics: minimizing research delays requires rapid restoration of data pipelines (low Recovery Time Objective, or RTO), while ensuring zero data loss during processing (tight Recovery Point Objective, or RPO) is essential in regulated environments governed by FDA, HIPAA, and GDPR standards.

Shadow IT cannot begin to meet these requirements. Plus, researchers are unlikely to keep up with critical updates, potentially creating security issues as well. It’s imperative that organizations eliminate these DIY storage solutions and bring storage and data oversight back under IT control.



Here are some common assumptions that lead teams to shadow IT and the issues that can arise:

Researchers say:	Issues:	Pure Storage:
“We only buy what we need, when we need it. That feels cost efficient and avoids waiting for long procurement cycles.”	Fragmented silos that slow down pipelines and inflate hidden costs. Eventually, a full replacement will be required.	Enables teams to achieve up to 24x faster secondary analysis as compared to traditional, disk-based environments.
“Off-the-shelf drives are affordable and easy to setup. They seem fine for storing raw data.”	Off-the-shelf storage can’t scale to meet the needs of genomic research. Time to results suffer, meaning slower discoveries and missed business opportunities.	Supports petabytes of throughput and fast access to hundreds of millions of files, enabling sprawling genomics pipelines to run unhindered.
“We need to share data across teams quickly. Moving files between drives or cloud shares feels like a workable solution.”	Shadow IT solutions can cause version mismatches, pipeline failures and data bottlenecks, slowing research and eroding trust across labs.	Consolidates research data on a unified, high-performance platform, eliminating bottlenecks, accelerating discovery, and protecting the integrity of results.
“We know data regulations matter, but our priority is just making sure the science keeps moving. We assume IT will handle the compliance side.”	Shadow IT lacks audit trails and governance controls, exposing the organization to HIPAA/GDPR violations, reputation damage, and regulatory fines.	Provides a secure, governed data foundation that safeguards data and meets HIPAA and GDPR requirements.

The importance of storage infrastructure within your lab

Genomics is not just about biology anymore; it's also a big data industry. As genomics data continues to grow at a historic pace, it's critical that organizations master the data with cloud, AI, and the right infrastructure to set the pace for the future of research. There's endless potential for research breakthroughs, but massive computing power and storage will be required to achieve them.

The right storage solution can:

Accelerate research cycles

High-performance storage capable of handling diverse workloads and high throughput can accelerate sequencing and analysis pipelines by up to 24x[5], reducing turnaround time from samples to insight.

Lower costs

A modern storage solution can greatly reduce operational overhead and research costs as well as strengthen stability with high mean time to data loss (MTTDL)/ high mean time to value (MTTF). The result: increased ROI.

Scale seamlessly to meet demand

Genomics data is doubling every 7 months.[6] Purpose-built storage can grow seamlessly to meet demand, no additional hardware required.

Enable secure, compliant data sharing

Modern storage solutions with built-in governance (HIPAA, GDPR, FDA-readiness) allow researchers to collaborate without risking violations.

What the impact of modern storage infrastructure looks like:

Up to 50% reduction in costs
per genome vs. shadow IT solutions

Process ~33% more genomes
per system, per day, enabling faster breakthroughs

Up to 60% more ROI
on storage investment through full utilization of GPUs and CPUs

What happens next
and what to watch for

Change doesn't always happen all at once. But the signs that it's time to re-evaluate are often right in front of you.

Use these questions to gather insights from clinical research, finance, and compliance teams:

Research head:

Are pipeline slowdowns affecting timelines?

High-performance storage accelerates analysis cycles by up to 24x.

CIO:

Are you balancing storage growth and budget constraints across workflows?

Modular storage platforms can reduce cost per genome by up to 50%.

Principal scientist:

How many genomes/day/node are you achieving? What limits that?

Purpose-built storage enables ~33% more genomes per day per node.

Compliance officer:

How are you ensuring data is preserved for 6+ years?

Flash-based storage ensures genome data secure, and audit-ready for 6+ years of regulatory retention.

The bottom line?

Now is the time to reexamine what your infrastructure is doing today and whether it can keep pace with where genomic research is going.

The difference between what you have **vs. what you actually need**

Legacy architecture and shadow IT solutions create issues that hold back research and burn budget. To truly optimize genomics research, organizations need a storage solution that can keep up with demand. It begins with treating storage as a strategic lever for speed, scale, cost and compliance.

Here are some legacy assumptions about storage, and how they are holding back genomic research teams:

Legacy assumption:	Reality:	Strategic infrastructure built with Pure Storage:
Storage is just a place to park data; all that matters is that we have enough of it	It's not just a container, it's also a strategic enabler of research	Flash-based storage drives speed, collaboration, and compliance natively in the data layer in genomics pipeline
Scaling up means adding hardware racks and the associated costs	Although this is true for physical, on-site storage, cloud-based solutions are just that, in the cloud	Future-proof your genomics platform without overprovisioning; flash storage enables you to start small and scale up confidently without expanding your footprint, energy use, or cooling needs
Cold data isn't valuable, there's no reason to keep it in active storage	Re-analysis of archived genomes can fuel new insights that would otherwise be missed	Stop managing data silos and start unlocking data as a strategic asset; enabling researchers to analyze data with fast recalls and no delays
Sharing data with another lab requires either physically shipping it or transferring it between sites	Modern platforms provide near-instant replicas and shared namespaces to enable real-time collaboration	Ensure real-time availability of high-volume datasets, enabling global team collaboration with Pure Storage, which provides cross-site data access without data duplication or delays

Where to go from here

In genomics, speed and accuracy are the difference between a breakthrough and a missed opportunity. Replacing legacy and shadow IT with an enterprise data cloud solution enables IT to scale up without adding physical infrastructure, reducing costs and increasing tech investment ROI. The right storage solution can enable AI insights and accelerate research while ensuring compliance with key governance. Bring IT out of the shadows and into the future of science with Pure Storage.

To learn more about how the right storage can accelerate genomics research, visit **Pure Storage**.

purestorage.com

800.379.PURE

PURESTORAGE®

1: Pure Storage life sciences website, Power Genomics Solutions, 2025.
2: Pure Storage ebook, Powering Data-driven Innovation in Life Sciences, 2022.
3: Lauren J. Young, IEEE, Genomic Data Growing Faster Than Twitter and YouTube, July 2015.
4: Kevin Gawora, Fact of the Week: Intelligence Can Save Pharmaceutical Companies Almost $54 Billion in R&D Costs Each Year, December 2020.
5. Pure Storage, Accelerate Genomics Analytics for Precision Medicine, 2023.
6. IDC, IDC Analyst Brief: Genomics Data Storage Solutions, 2025.