



# Sanger Institute Accelerates DNA Analysis With BlueArc Storage Systems

**“Dealing with DNA structure means terabytes of information, which need to be viewed and stored quickly and continuously. The Titan 2000 system allows us to perform tasks efficiently and reliably.”**

**“Our core mission is open access to data. We sequence it, publish it, and the world uses it.”**

**“Having a system like Titan that can handle many simultaneous clients is very useful. We have had millions of files on our systems in the past and seen that lead to problems. Now, with Titan, those issues have disappeared.”**

**“Being a non profit organization means cost considerations are vital. BlueArc’s Titan 2000 allowed us to use our existing technology infrastructure and established a platform we can build on over time.”**

**–Dave Holland,  
Senior Unix Systems Administrator  
for The Sanger Institute**

## Summary

The Wellcome Trust Sanger Institute, a genome research institute based in the United Kingdom, relies on BlueArc’s Titan 2000 storage system to aid invaluable research into the DNA sequencing process of a wide variety of species and organisms. Research into DNA variations allows the organization to find new ways to diagnose, treat, and prevent thousands of disorders that affect people worldwide. BlueArc storage systems give the Sanger Institute a reliable infrastructure that can be accessed by hundreds of users simultaneously, giving them the data they need to rapidly and effectively research new treatments.

## The Customer

The Wellcome Trust Sanger Institute is a genome research center set up in 1992 by the Wellcome Trust and the Medical Research Council in order to further knowledge of genomes, and in particular to play a substantial role in the sequencing and interpretation of the human genome. The Institute made the largest single contribution to the public human genome sequence. This information will underpin research on human biology and disease in this century and beyond.

The core mission of the Sanger Institute is to provide open access to DNA data and promote the study of and access to DNA for research purposes. Every project they create or are involved in is available on the Web for anyone to download at any time. Storing all this data and having it available quickly is vital to the organization’s mission of enabling progress in the understanding and treatment of disease.

## The Challenge

The Sanger Institute leverages powerful DNA sequencing technology to align and break down DNA sequences and compare organisms. To best analyze this highly complex and valuable data, the Institute was looking for a single scalable storage solution that could handle many researchers accessing the DNA information simultaneously. Up to one thousand clients would be taxing the organization’s data sequencing and storage systems, placing a major burden on the load balancing software that managed access to stored data. To meet the requirements of this environment, the Institute needed to implement a storage solution architected to deliver uptime through crunch time, and integrate with the company’s existing network infrastructure.

## The Solution

The Sanger Institute selected BlueArc’s Titan 2000 as it passed every performance requirement, offered scalability in terms of capacity and simultaneous users, and easily integrated into the organization’s existing technology infrastructure, acting as a NAS gateway to the Institute’s existing SAN environment. Following a rigorous testing process, the Institute approached the various components of BlueArc’s offering and tried to find weaknesses, through disconnecting

networks, failing disks and disconnecting fibre channel connections. Throughout all the tests, the Titan 2000 remained stable and continued to serve data, proving itself capable of dealing with any potential disaster.

### The Results

For the Sanger Institute, availability of data is key to its livelihood and reputation. The Institute's core mission, to deliver open access to research data, making it available to the world's scientists for further study, requires network storage performance and capacity that do not slow down under peak load, while also offering the capability to handle millions of small files generated from genome sequencing and DNA comparisons. Additionally, as system failure simply was not an option, with users anticipated to run demanding applications against the infrastructure around the clock, the selected solution needed to deliver more than a simple storage repository, but instead, operate as a mission-critical element of the organization's research infrastructure.

"For us, reliability is the most important factor in deploying storage," said Dave Holland, Senior Unix Systems Administrator for the Sanger Institute. "If you have 1,000 people performing analysis, having the infrastructure fail underneath is embarrassing to say the least. BlueArc passed every test we threw at it, and delivers strong performance."

Like many other storage buyers dissatisfied with frequently being asked to replace quickly outdated equipment, the Institute also wanted to future-proof their investment, eliminating the need to frequently replace obsolete equipment to deal with the ever-growing volume of data they would produce. The Titan 2000's inherent modular structure, as well as its ability to support capacities upwards of 512 terabytes on a single node, would enable the Institute to grow their data utilization without concern as to whether the storage system would be capable of keeping up.

In addition, the Institute, in an effort to best utilize equipment in house, was keen to utilize the organization's existing data storage from a pre-existing storage area network. BlueArc's Titan 2000 offered the ability to integrate well with the existing SAN, front-ending the disk on site.

"At the Sanger Institute, we try to do a technology refresh every two to three years," said Holland. "We looked at new NAS platforms, and had requirements to use our existing SAN storage."

BlueArc showed the willingness to partner with the Institute, undergoing a battery of tests to ensure compatibility with on-site equipment, that performance would be there under peak load, and the system would be resilient, even in times of highest demand. With hundreds of blade servers working in concert to study millions of DNA base pairs for sequencing and analysis, any hiccup in the process could impact time to results. Titan's hardware architecture, as well as integrated software for security, reliability and file system capacity were key to bringing peace of mind to the organization.

"You might have up to a thousand clients looking for load distribution software, and having a single repository for this information is ideal," said Holland. "Having a system like Titan that can handle many simultaneous clients is very useful. We have had millions of files on our systems in the past and seen that lead to problems. Now, with Titan, those issues have disappeared."

### The Conclusion

The Sanger Institute promotes the study and research of DNA, and the organization's research can lead to revolutionary new ways to diagnose, treat, and potentially prevent thousands of disorders that affect people worldwide. In addition to providing clues to understanding human biology, learning about nonhuman organisms' DNA sequences can lead to an understanding of their natural capabilities that can be applied toward solving challenges in health care, agriculture, energy production, environmental remediation, and carbon sequestration.

The Sanger Institute's far-reaching, challenging and important mission requires a technology infrastructure with highest reliability, scalability and performance. Science is most successful when collaboration is enabled and the best minds in the world can access not only their own data, but also the cumulative findings of thousands. BlueArc's Titan 2000 system enables the Sanger Institute to share valuable information and provide open access to research without having to worry about back-end technology issues, allowing them to do exactly what they do best — science.

*\*\*The Sanger Institute does not endorse commercial products.*



**BlueArc Corporation**  
Corporate Headquarters  
50 Rio Robles Drive  
San Jose, CA 95134  
t 408 576 6600  
f 408 576 6601  
www.bluearc.com

**BlueArc UK Ltd.**  
European Headquarters  
Queensgate House  
Cookham Road  
Bracknell RG12 1RB, United Kingdom  
t +44 (0) 1344 408 200  
f +44 (0) 1344 408 202