

Accelerating groundbreaking multi-omics research.

How **Barcelona Supercomputing Center** selected Lenovo's Genomics Optimization and Scalability Tool running on a Lenovo ThinkSystem SR950 server, powered by 2nd Gen Intel® Xeon® Scalable processors, to evolve its multi-omics research capabilities, marking a new chapter of pioneering research.

Lenovo Infrastructure Solutions
for The Data-Centered



Lenovo

1

Background

Established in 2005, Barcelona Supercomputing Center (BSC) is Spain's leading national supercomputing resource. The organization runs and manages [MareNostrum](#), one of the most powerful supercomputers in Europe, and helps over 725 experts in computer sciences, life sciences, earth sciences, and engineering carry out ground-breaking research.

In its latest expansion drive, BSC has been increasing its support for computational and comparative geonomics research. By providing leading academics in these fields with the compute resources to fuel innovative projects, BSC hopes to help scientists build a deeper understanding of complex cellular processes, improve disease prevention, and chart the evolution of biological systems.

2

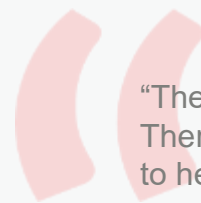
Challenge

Multi-omics analytics, a catchall term encompassing genomics, proteomics, transcriptomics, metabolomics, and microbiomics analyses, has revolutionized our understanding of the biological world. As a direct result, researchers and healthcare professionals are actively seeking powerful new ways to process such large and diverse datasets in a reasonable amount of time.

Integrating complex multi-omics datasets marks only the beginning of an exciting new stage of scientific discovery—deploying them at scale while reducing time-to-insight is the next goal in the field, which leading experts recognize will ultimately be driven by cutting-edge technologies.

Miguel Vazquez, Ph.D., Head of the Genome Informatics Unit at BSC, explains: “Genomics is a rapidly evolving field. To fuel discoveries, it is critical that research centers like BSC provide powerful computational resources to support the end users. And with the volume and complexity of the data increasing each year, tuning the pipelines to maximize the hardware resources and optimizing the hardware to the bioinformatics needs is critical to ensure that researchers can focus on their science and accelerate the pace of new insights at the MareNostrum.”

To this end, BSC set out to enhance the way it runs complex multi-omics workflows. While its in-house omics pipelines leveraging the Ruby Bioinformatics Toolkit (RBBT) have already helped BSC researchers make breakthroughs by increasing usability and repeatability, the workflows ran slowly, taking approximately 30 hours to process a single whole human genome.



“The future of genomics research will ultimately be driven by technology. Therefore, it is essential that we continuously optimize our HPC capabilities to help push the boundaries of scientific knowledge.”

Miguel Vazquez, Ph.D.

Head of the Genome Informatics Unit, Barcelona Supercomputing Center



Why Lenovo? World-leading expertise, next-generation technologies.

Having worked closely with Lenovo in the [original development of MareNostrum](#), BSC enlisted the help of its trusted technology partner to uncover new ways of optimizing its specialized multi-omics workflows.

“We’ve collaborated with Lenovo for many projects in the past and Lenovo technology has played a key role in much of the research carried out at BSC,” reflects Miguel Vazquez. “Lenovo is one of the leading providers of HPC solutions and we were keen to tap into their expertise, and get up to speed with the latest technical advances. With Lenovo’s support, we were confident that we could find smarter, more efficient, and more productive ways of running our multi-omics workloads.”



“Lenovo is a long-term partner of BSC. We’ve helped many scientists who come to the MareNostrum from institutions all over the country make impressive discoveries utilizing Lenovo HPC technology, and we look forward to continuing this work in the years ahead.”

Miguel Vazquez, Ph.D.

Head of the Genome Informatics Unit, Barcelona Supercomputing Center

Unlocking new efficiencies.

To help BSC optimize its multi-omics workloads, Lenovo invited the organization to the Lenovo Innovation Center in Stuttgart, Germany, and provided access to a GOAST Plus System built on a Lenovo ThinkSystem SR950 server and leveraging 2nd Gen Intel® Xeon® Scalable processors.

Miguel Vazquez comments: “Collaborating with experts at the Lenovo Innovation Center, we adopted the bioinformatics optimizations Lenovo found for GATK workflows into our RBBT multi-omics workflows and ran them on the GOAST system, which has been tuned at the hardware level to enhance bioinformatics execution.”

The collaboration with the Lenovo Life Sciences team gave BSC the opportunity to take the lessons learned to transform the way it will run omics analyses at the MareNostrum supercomputer, as Miguel Vazquez explains: “We recognized, for instance, just how important it was to run samples in throughput. In a production environment it’s critical to set up the secondary omics analytics to process samples steadily and in batches to make better use of the hardware resources available in the cluster, rather than using all node resources to run one sample at a time. Learning to run samples in throughput while squeezing the most out of the cluster resources is one of the best ways we found to increase computational efficiency.”



“Our collaboration with Lenovo using GOAST helped us unlock new ways to boost the efficiency of omics workflows.”

Miguel Vazquez, Ph.D.,
Head of the Genome Informatics Unit,
Barcelona Supercomputing Center

3

Results

With the Lenovo GOAST system, BSC was able to reduce execution times for multi-omics analyses, increase throughput capacity, and accelerate time-to-insight.

Miguel Vazquez comments: “With the GOAST Plus system running on the Lenovo ThinkSystem SR950 server, we were able to reduce the time taken to process a single 30x germline variant calling sample from 30 hours to just 45 minutes per sample, that’s 40 times faster. What’s more, a 50x whole exome run in a single node can take about 2 hours; on the optimized GOAST machine it can be brought down to 1.5 minutes per sample, which is 80 times faster.”¹

While these impressive improvements can partly be attributed to the high-performance Lenovo ThinkSystem SR950 server and 2nd Gen Intel® Xeon® Scalable processors used, BSC anticipates that it will be able to achieve similar results when it runs these multi-omics workflows on MareNostrum—thanks to optimizations the team made based on their experience using the Lenovo GOAST system.

“With Lenovo GOAST, we got a taste of just how much we could increase throughput capacity,” explains Miguel Vazquez. “For instance, the solution has the potential to help us process 32 whole genomes per node each day, or 351,000 whole exome samples annually.² Achieving similar throughput levels at the MareNostrum would significantly increase the number of research projects that we can support and help scientists get their results even faster.”

He concludes: “By implementing the optimizations we’ve learned from the Lenovo GOAST system, we will be even better equipped to help academics advance our understanding of the world around us.”



- ✓ 80 times faster processing of whole exomes
- ✓ 40 times faster processing of whole genomes
- ✓ ~351,000 whole exomes and ~12,000 whole genome samples can be processed per node per year

^{1,2} Data provided by the Genome Informatics Unit, Barcelona Supercomputing Center.



“Thanks to our strong collaboration with Lenovo, we are well positioned to keep pace with the latest advances in genomics research and lead the way for multi-omics analyses in Europe.”

Miguel Vazquez, Ph.D.

Head of the Genome Informatics Unit, Barcelona Supercomputing Center

What will you do with Lenovo ThinkSystem solutions?

The Data-Centered enable cutting-edge research and development
with Lenovo smarter infrastructure solutions, powered by
2nd Gen Intel® Xeon® Scalable processors.

[Explore Lenovo ThinkSystem Server Solutions](#)



Lenovo and the Lenovo logo are trademarks or registered trademarks of Lenovo.

Intel, the Intel logo, the Intel Inside logo and Xeon are trademarks of Intel Corporation or its subsidiaries.

Other company, product and service names may be trademarks or service marks of others.

© Lenovo 2021. All rights reserved.