Whole genome sequencing at the Garvan Institute of Medical Research.

XENON Brings Affordable 24x7 Genome Sequencing Computing Capability to the **Garvan Institute** to Enhance Genomic Discovery

> *"XENON has brought enhanced performance computing to the highly specialised field of genomics.*
>
> *The whole point of technology is to enable new advancements for society as affordably as possible. When these gains are applied to the biomedical profession, we all benefit with every improvement, every day. "*
>
> *Dr. Warren Kaplan*
> *Garvan Chief of Informatics*

Background

The Garvan Institute of Medical Research is a world-renowned organisation whose mission is make significant contributions that will change the directions of science and medicine and have major impacts on human health through insights into cancer, neurological disorders, immune disorders, diabetes, obesity and osteoporosis.

"Remarkable advances have occurred in how genomics are contributing to this goal, and in how scientists have become able to read DNA over the last number of years," says Dr. Warren Kaplan, Garvan Chief of Informatics. Accelerating this pace is crucial to making key discoveries in biomedicine.

"The genome is an extraordinary vehicle with which to gain insights into the Institute's areas of disease research, and so genome sequencing comprises a fundamental component of this. We are now in the age of the cohort, and it's now

GARVAN
INSTITUTE
OF MEDICAL RESEARCH

XENON ®
High Performance Computing

Dr Mark Pinese, Mr Aaron Statham and Dr Warren Kaplan from the Garvan Institute's informatics and genomic research teams.

time to liberate cohorts with genomics, and technology is foundational to this," explains Dr. Kaplan.

Indeed, the world has moved from sequencing the first human genome that took ten years and $3bn dollars, to a time in which the Garvan Institute and others can now sequence around 40 genomes per day at a tiny fraction of this cost.

Challenge

These extraordinary gains are enabled not only by remarkable human talent, but also by highly sophisticated computational and storage technology. It's this combination that drives population scale genomics at the Garvan Institute, which can analyse thousands of genomes per year in a cost-effective and secure manner, for its own work as well as for third parties.

To put things in perspective, completing the analysis of a single human genome requires at least 200 gigabytes of storage and 650 CPU-hours. By way of example, a 1200 person-strong cohort currently being researched within Garvan required 79 CPU-years alone.

"And so genomics is no longer just about high performance computing, but increasingly about big data and data intensive computing," said Dr. Kaplan.

The challenge has now become to normalise the data across the cohort. This process is known as joint variant calling ("JVC"). The nature of the JVC computation is such that it is technically incredibly intensive, and

benefits from fast local storage in order to accelerate this analysis.

However, the Garvan Institute's previous computing infrastructure was pushed to the limit under such strenuous use by its 80 bioinformaticians. The time came twelve months ago when a new system was needed to keep up with the demand for computational resources.

The specific challenges were:
▶ how to use technology to provide the required capabilities and resources efficiently;
▶ how to interrogate the enormous amount of data the Garvan Institute collates;
▶ how to ensure the necessary privacy and security; and
▶ how to uncover the value in the data.

Solution

With these challenges in mind, the Garvan Institute went to market for a new HPC system. Dr. Kaplan consulted high-performance computing experts across Australia.

The group recommended Melbourne-headquartered HPC consultancy XENON Systems, a specialist in the field of HPC technology for academia, finance, and more recently, medical research.

XENON has developed a reputation for leading edge solutions through relationships with a variety of international technology companies. This gives XENON early access to new technologies and developments that allow it to continually push the boundaries of performance and capability.

The team at XENON designed a unique turnkey HPC system, which combines the high performance of Intel Xeon E5-2600v3 family CPUs with a local storage solution.

Ten high performance Intel DC P3500 NVMe SSD drives provide a total of 20 TB of local storage capacity for an ultra-high performance local storage subsystem, which alleviates the bottlenecks of traditional HDD or SSD based storage system. This design is delivering an incredible leap in performance and efficiency while balancing CPU and storage performance.

Since all applications run on a single compute node the network was been designed for

XENON
High Performance Computing

www.xenon.com.au
info@xenon.com.au
1300 888 030

"We were not aware that a machine like this existed until XENON designed it. "

good performance but not at the expense of the CPU and storage subsystems.

The proposed HPC system was also designed to support Hadoop/Spark workloads well as well as OpenStack for virtualized workloads.

Outcomes

"When we went to XENON, we had very specific workloads that we were hoping to support including JVC and other types of bioinformatics computations and codes. The computer cluster, which the Garvan Institute acquired from XENON, was designed specifically around very fast local storage based on Intel NVMe SSDs and the latest generation of Intel CPUs.

As a result, researchers at the Garvan Institute have now been able to apply the JVC method to 1200 genomes which is the first part of the Medical Genome Reference Bank (MGRB) (https://sgc.garvan.org.au/initiatives/mgrb) cohort of 4000. When finished, this will be the largest Australian genomic cohort.

The particular steps that we use the XENON cluster for, other systems simply cannot perform. We'd have to compromise in others way if we weren't using it. When we apply the JVC method, we take all 1200 genomes and look at them at the same time with the idea that we can more likely find a genetic variation. Otherwise, the compromise might be to do that kind of analysis in smaller chunks, say, 50 people. Larger variant calling makes it a lot easier at the end when we bring all analysis together. And that's essentially back to one of our core beliefs that we are in the age of the cohort. And this is what the XENON system is enabling," added Dr. Kaplan.

The system can be expanded easily by scaling out through the addition of compute nodes and switches. Additional storage systems and special systems for large memory requirements can be integrated into the infrastructure of the proposed system at any time.

Dr. Kaplan explains that the XENON solution provides the efficiency and affordability needed to deliver genomic sequencing on a large scale.

"Quite simply, with these kinds of demands if we'd gone to an external third party cloud service, we most certainly wouldn't be able to afford workloads running 24 x7 at this level."

"We were not aware that a machine like this existed until XENON designed it. XENON has brought high performance computing to the highly specialised field of genomics. The whole point of technology is to enable new advancements for society as affordably as possible. When these gains are applied to the biomedical profession, we all benefit with every improvement, every day," concluded Dr. Kaplan.

About The Garvan Institute

The Garvan Institute of Medical Research is one of Australia's largest medical research institutions and is at the forefront of next-generation genomic sequencing in Australia. Garvan's main research areas are: cancer, diabetes and metabolism, immunology and inflammation, osteoporosis and bone biology, genomics and epigenetics, and neuroscience. Garvan's mission is to make significant contributions to medical science that will change the directions of science and medicine and have major impacts on human health. In 2012, Garvan established the Kinghorn Centre for Clinical Genomics, Australia's first purpose-built facility for undertaking clinical-grade genome sequencing and large-scale research projects.

About XENON Systems

XENON is an Australian leader in High Performance Computing solutions. XENON's innovative products and technology are designed to tackle the most data intensive and complex visualisation challenges, allowing its clients to focus on breaking new ground in their respective fields.

Whether it's high performance computing, network design, server and storage solutions or visual workstation technology being sought, it has delivered tailored solutions in a variety of demanding environments including science, defence, manufacturing, precision medicine, finance, broadcasting, motor racing, education and telecommunications.

Recognised and trusted as a partner to achieve the extraordinary, its talent lies in applying new thinking and ideas to create pioneering solutions that uniquely address client's needs.

XENON
High Performance Computing