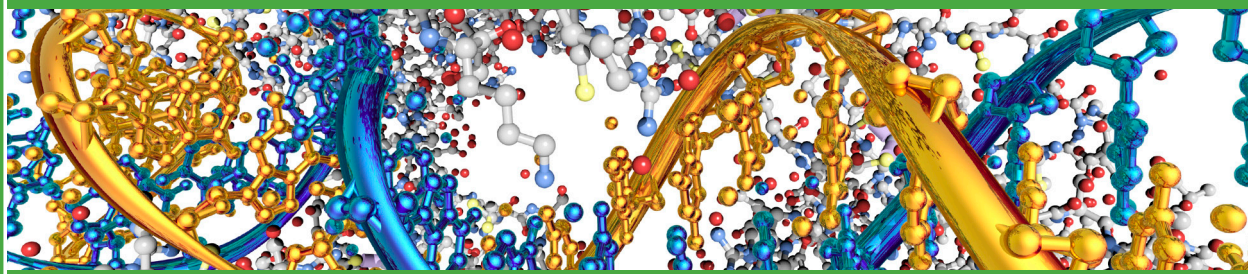


SGI Solutions Help Scientists Accelerate Cancer Research



Key Facts

Organization – Queensland Centre for Medical Genomics

Location – Brisbane, Australia

Applications – Data management

SGI Solutions – SGI® LiveArc™

SGI® Data Migration Facility (DMF)

SGI® InfiniteStorage 4500

SGI® Altix® XE

Tape library upgrades

Background

The University of Queensland's Institute for Molecular Bioscience (IMB) is internationally recognized as a leading center for molecular bioscience research. It was established in 2000 and is located in the Queensland Bioscience Precinct.

The IMB is a multidisciplinary research institute with 500 research staff and students and a range of strategic programs in mammalian systems biology, supported by some of the finest facilities in the world.

The major focus of IMB research is to improve human health with the development of new pharmaceuticals, cell therapies and diagnostics through the understanding of information contained in the genes, proteins and molecules of plants and animals.

Business Challenge

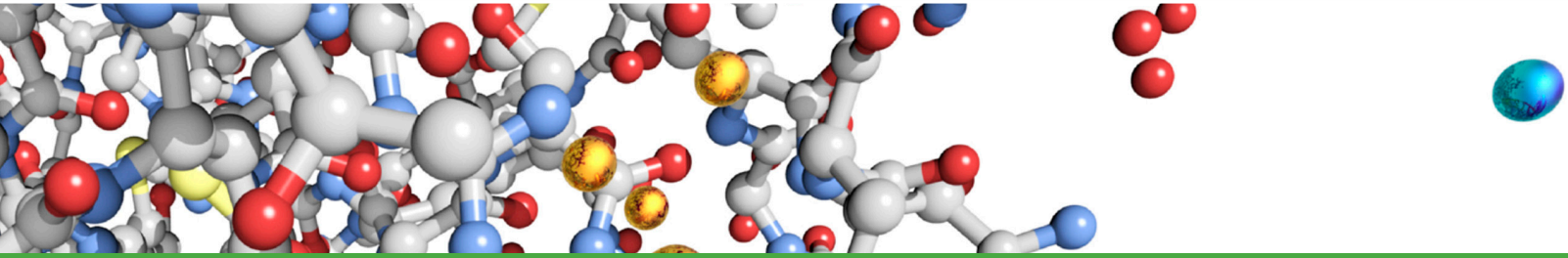
The QCMG is a member of the International Cancer Genome Consortium, whose members will together sequence the genetic codes of 25,000 tumors from 50 different types of cancer over the next five years.

Scientists at the QCMG are specifically studying pancreatic and ovarian tumors, two of the most common causes of cancer death in the developed world. Pancreatic cancer causes death within half a year after detection in the average patient. Ovarian cancer, while less deadly in its primary form, currently has no screening test and is therefore usually not discovered until it has spread, making treatment difficult.

QCMG's 11 ABI SOLiD genome sequencers produce over 5TB of summarized data per week. This data needs to be catalogued, archived, and routed to HPC systems and scratch storage for processing and transformation according to research requirements. Managing the volume of data and consequent workflow were major practical challenges for the QCMG.

"We need to keep the operations side of the QCMG lean so we can concentrate on research and that means automating as many of our workflows as possible and that's where we're looking to LiveArc¹. We have to manage sequencing, storage and computational resources and move raw and derived datasets from resource to resource to complete an analysis."

– John Pearson, Senior Bioinformatics Manager at QCMG



Technology Solution

SGI created a solution to provide automated handling and storage of data through each step of QCMG's process workflow. The key design principals for the solution are:

- To manage large amounts of raw sequence data as well as derived data from secondary and tertiary analyses and provide automated data and metadata handling, repository dispatch and lifecycle curation functions
- To provide a framework for further automation of other regularly repeated analysis processes
- To ensure the infrastructure provides flexibility for scientific endeavour without constraints from fixed or rigid processes, permitting the evolution of schemas, processes and system interactions

QCMG chose SGI LiveArc digital asset management for the management of data, metadata and workflow processes.

LiveArc is specifically designed to allow for faster discovery, one-step sharing and closer collaboration, enabling more efficient data management and reduced infrastructure costs. As data volumes continue to grow, so do the challenges for IT managers to cost-effectively provide higher levels of utilization across different data types and workflows.

LiveArc manages the QCMG metadata and data throughout its entire lifecycle, and is responsible for:

- Ingestion of data and metadata from the genome sequencers
- Automatic generation of tapes for transfer of data to collaborating ICGC institutions overseas
- Replication of metadata and data from the local data store to the University's highly available long term hierarchical data store, based on SGI Data Migration Facility (DMF)
- Sophisticated search facilities to allow researchers to select the required data as required for transformation
- Automatic creation of jobs and delivery of data to high speed, third party scratch storage for processing by the SGI High Performance Computing system
- Re-ingestion of the resulting secondary and tertiary metadata and data

Built on an extensible service-oriented architecture (SOA), LiveArc provides a multi-versioned asset management and revision control system with workflow and web serving in a single package. Not only does this simplify operation and administration, LiveArc's flexibility also enables it to support the integration of legacy applications and metadata types, and to easily evolve as workflow requirements grow.

This allows LiveArc to act as a common interface, bridging multiple legacy environments and overcoming incompatible data silos without the need to discard current applications or infrastructure.

At its core, LiveArc is a comprehensive platform used to rapidly develop and deploy workflow-specific digital asset management solutions.

Accessible via industry-standard protocols, LiveArc is platform neutral, and can be deployed on laptops, desktops and enterprise-class servers in all standard operating systems and file systems. Due to its unique and efficient binary XML object database, LiveArc enables orders of magnitude better performance, yet requires a much smaller memory and database footprint than competing solutions. Other benefits include near zero administration while retaining open standard data portability.

LiveArc is tightly integrated with SGI's DMF tier virtualization solution, which provides seamless online visibility to multiple types of disk and tape-based storage in a virtualized data pool. Adding LiveArc's ability to federate across multiple data repositories in parallel, the platform offers virtually limitless ability to scale in both data volume and file count.

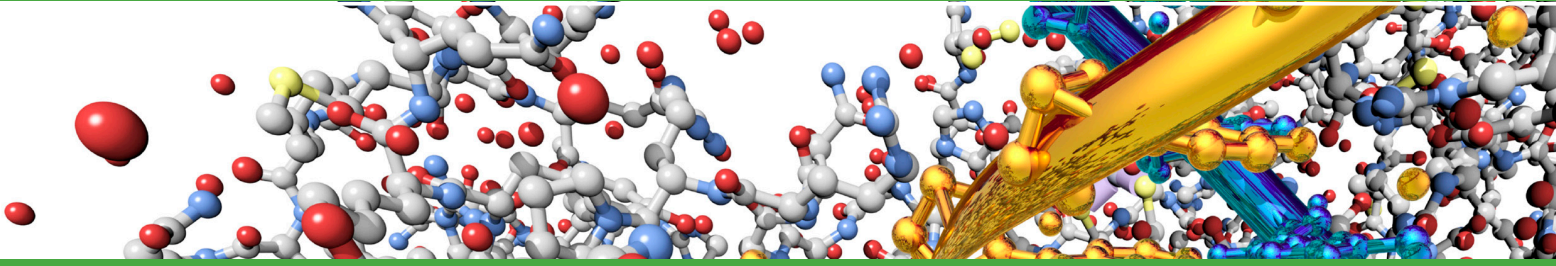
Business Results

The fundamental goal of the project is to use the combined resources of laboratories worldwide to create a map of the genetic changes that lead to cancer. The map will act as a huge information resource for medical researchers, allowing for more rapid and personal treatments for cancer sufferers.

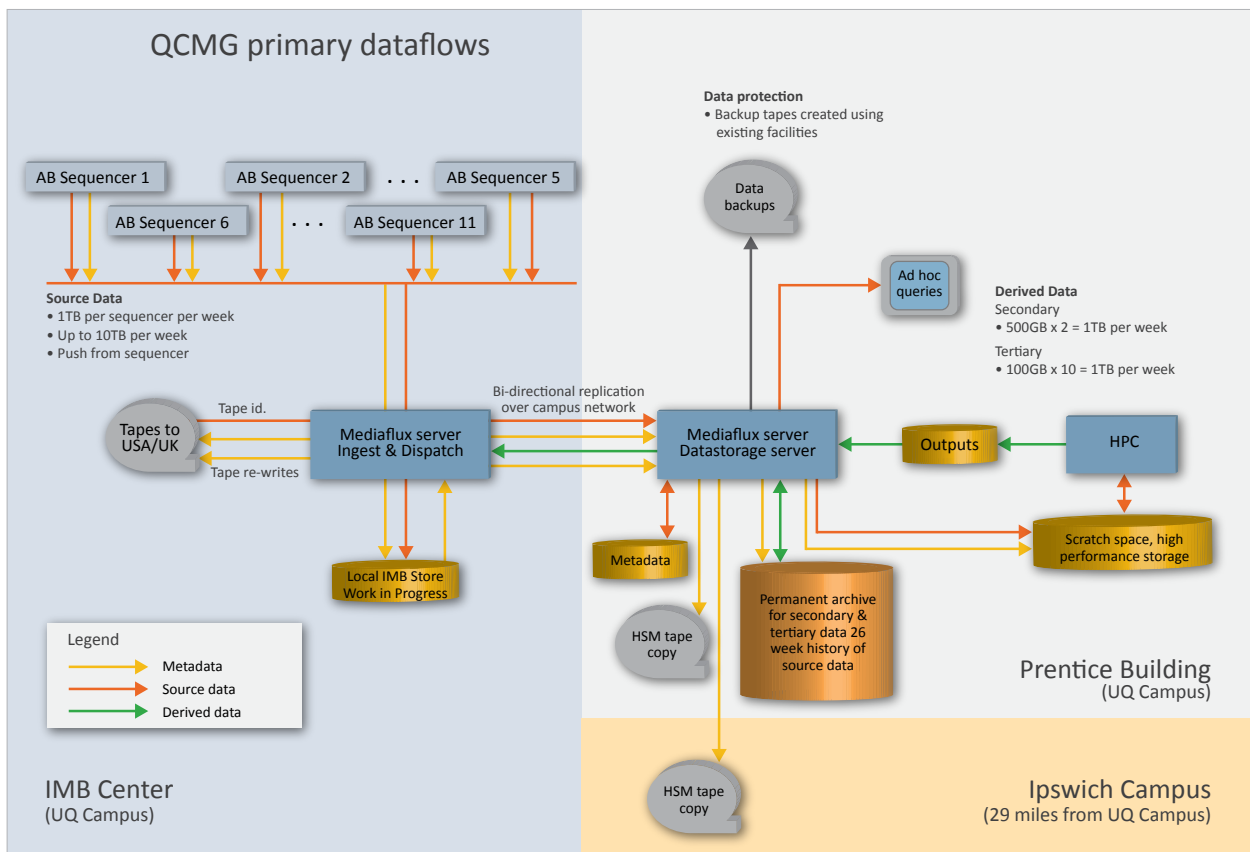
The QCMG is already producing results, with the first genetic data for two pancreatic tumors uploaded to the ICGC web portal. These data will be available for researchers all over the world to access and use.

¹LiveArc is a result of a partnership between SGI and Arcitecta Pty Ltd of Australia. SGI has deployed LiveArc in media, science and research institutions, government, and academia in the Asia Pacific region under the name Mediaflux™. LiveArc is now available worldwide exclusively from SGI, through SGI Professional Services. Mediaflux is a trademark of Arcitecta Pty Limited in Australia.

“LiveArc will allow us to automate and verify data transfers and archiving, query our LIMS to determine the appropriate type of analysis based on the run type and then trigger our cluster-based analysis tools and generate reports. We’re looking to LiveArc to be the glue that holds our analytical pipeline together,” explained Pearson.



The data is now undergoing interpretation at QCMG, and changes in the genetic code have been identified. It is as yet unknown whether the changes are responsible for disease. The next step is to run tests in the lab to confirm if these genetic changes have a biological effect. This will be done for every tumor sequenced over the five-year project.



Corporate Office
 46600 Landing Parkway
 Fremont, CA 94538
 tel 510.933.8300
 fax 408.321.0293
 www.sgi.com

North America +1 800.800.7441
 Latin America +55 11.5185.2860
 Europe +44 118.912.7500
 Asia Pacific +61 2.9448.1463

