ARCITECTA®

QUEENSLAND CENTRE OF MEDICAL GENOMICS (QCMG)

OVERVIEW

The Queensland Centre of Medical Genomics (QCMG) has implemented Mediaflux to automate data workflows for over 500 research collaborators. Mediaflux has become the glue that holds QCMG's analytical pipeline together and keeps operations lean so that research remains the focus at QCMG.

The University of Queensland's Institute for Molecular Bioscience (IMB) is internationally recognised as a leading centre 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.

CASE STUDY:

QUEENSLAND CENTRE OF MEDICAL GENOMICS (QCMG)

THE CHALLENGE

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.

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

Scientists at the QCMG are specifically studying pancreatic and ovarian tumours, 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.

THE SOLUTION

Managing the volume of data and consequent workflow were major practical challenges for the QCMG. QCMG chose Mediaflux for the management of data, metadata and workflow processes associated with the 11 ABI SOLiD genome sequencers. Producing over 5TB of summarised data per week, Mediaflux is used to catalogue, archive, and route data to HPC systems and scratch storage for processing and transformation according to research requirements.

THE OUTCOME

Mediaflux 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's 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.

"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 Mediaflux. We have to manage sequencing, storage and computational resources and move raw and derived datasets from resource to resource to complete an analysis."

"Mediaflux 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 Mediaflux to be the glue that holds our analytical pipeline together."

- John Pearson, Senior Bioinformatics Manager at QCMG