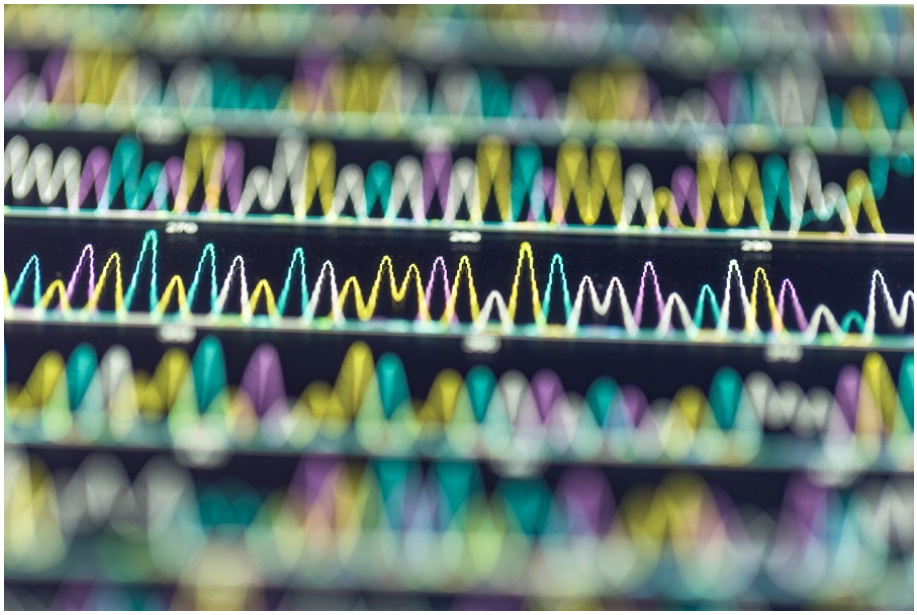
Evidence Summary

SEPTEMBER 2020 australiangenomics.org.au

Australian Genomics Data Management

This Program has been developing data standards and processes to capture and use clinical and genomic data from across Australian clinics and laboratories, to support the activities of Australian Genomics clinical Flagship studies.



Background

Using genomic technologies in healthcare brings unique challenges in technical and data infrastructure. Sequencing one patient's genome generates a huge amount of data - around 250GB with 3 billion base pairs, and 5 million genetic changes per patient.

To manage data of this scale and complexity, genomic medicine needs high-performance computing; data storage infrastructure; and local and national services for data access and exchange.

Making sense of this data, to determine which genetic changes are causing a patient's condition, is also an enormous challenge for clinicians. To better enable genomic diagnoses, it is vital that we have infrastructure to aggregate patient genomic data, combine it with meaningful clinical information, and make it available for clinical and research purposes.

Bringing together genomic and clinical information from across clinical sites and pathology labs in Australia's states and territories, however, brings operational, technical and regulatory challenges.

Project aims

The Program is coordinating a national approach to managing and analysing clinical and genomic information under four key areas:

1. A national genotype-phenotype database.
2. Standardisation of genomic and clinical data.
3. Infrastructure for storage and access of genomic data.
4. Improved clinical standards for interpreting genomic variants.

Key products

Australian Genomics Database

The clinical, demographic and survey data of all our participants is stored and managed in our database. Study data are collected and managed using REDCap electronic data capture tools[1] hosted at Murdoch Children's Research Institute. Our database integrates information collected from over 32 clinical recruitment sites across Australia.

Genomic Data Repository (GDR)

The Australian Genomics GDR is a cloud-based storage system, hosted at The University of Melbourne. The GDR stores genomic data from the Australian Genomics clinical Flagship studies and balances security, data integrity and access for researchers.

Data Access Agreements and Policies

Our participants provided consent to both national and international data sharing, for the benefit of healthcare. From this, a Data Governance Framework has been defined, which is guiding access principles and policies, agreements and terms of research access. Clinicians and researchers can request access to Australian Genomics datasets, which will be granted according to the level of data sensitivity; each participants' consent; and the researcher's Human Research Ethics Committee.

CTRL dynamic consent & DUO standard

Individualised consent is being sought through a trial implementation of our CTRL dynamic consent platform. Participant's consent choices about future research use of their data are being captured and integrated into the REDCap database, using a Global Alliance for Genomics and Health data use ontology (DUO) standard. This means each participant's consent preferences are captured in real-time and applied to cohorts for data sharing. Standardising data use conditions in this way eliminates ambiguity in data use permissions and streamlines complex data sharing processes.

Standardised Computer-Readable Clinical Data Tools

Our tools enable detailed clinical information from Flagship patients to be collected and translated into standardised codes (phenotype ontologies: HPO/SNOMED) and clinical data formats that are compatible with eHealth systems and diagnostic tools. This allows computerised search, discovery and coding, to power genomic research and clinical activities.

'Variant Atlas' Genotype-Phenotype Platform

Variant Atlas is an interactive genotype-phenotype data platform, hosted at the Garvan Institute. With this tool, researchers and clinicians can visualise aggregated Flagship data, query variants across the Flagships, and filter by key clinical features to explore and describe the different patient populations.

'Shariant' Platform for Interpreting Variants

Shariant is a centralised platform developed by Centre for Cancer Biology (an SA Pathology and UniSA Alliance) for Australian pathology laboratories to share curated variant classifications and detailed curation evidence in real-time. Labs are automatically notified of classification differences with other groups and can connect to international databases.

More on our data program at australiangenomics.org.au/data

Potential impact

Unlocking the power of genomic information for clinical benefit requires complex patient data to be gathered and analysed in large numbers.

Our Data Program enables this through a national clinical genomics infrastructure to capture, store and securely share patient data on a large scale.

This will ensure clinicians and researchers have access to the information needed to **power clinical genomic discoveries**, and better **obtain genomic diagnoses**.

Conclusion

The Australian Genomics Data Management Program has developed data standards, tools and infrastructure to support its clinical Flagship studies, operating nationally.

Australian patients, clinical teams, hospitals, pathology labs and researchers are actively supporting and contributing to our national data sharing and management efforts.

This Program demonstrates technical, regulatory, and operational feasibility of a co-ordinated, national approach to capture and use clinical genomic data from across Australia.

Reference

1. PA Harris, R Taylor, R Thielke, J Payne, N Gonzalez, JG. Conde, Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support, J Biomed Inform. 2009 Apr; 42(2):377-81. Available here.

Acknowledgement

Published 22 October 2019. **Updated September 2020**.
Australian Genomics Health Alliance is funded by NHMRC grant 1113531 and the Medical Research Future Fund.

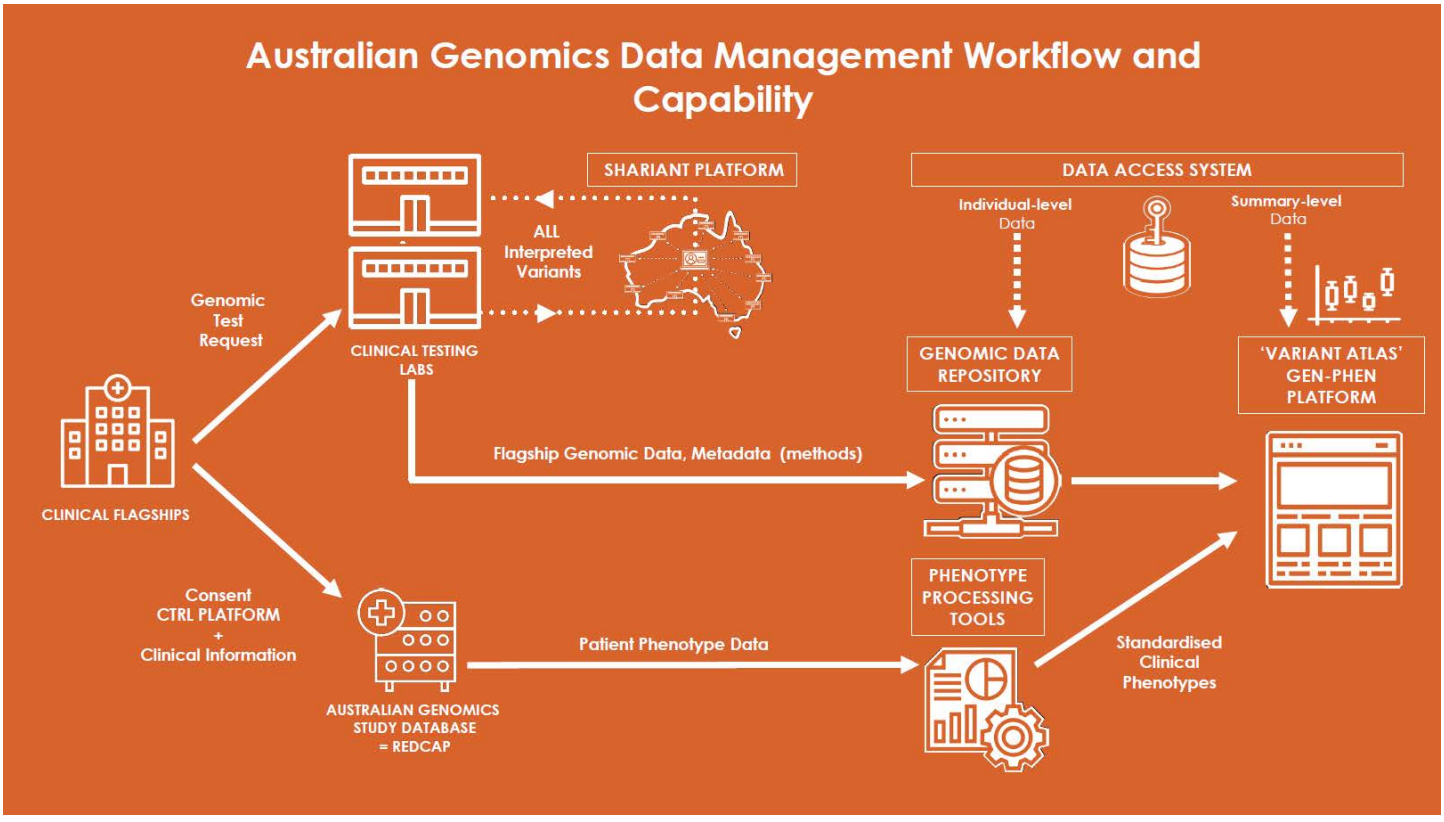


Figure: Australian Genomics Data Management Program. *A cohesive infrastructure for collecting, storing, and sharing clinical and genomic data from the national, clinical Flagship studies*.