DEVELOPMENT OF STANDARDS FOR THE
ACCREDITATION OF DNA SEQUENCE VARIATION
DATABASES

A final report by
The Royal College of Pathologists of Australasia

To the
Australian Department of Health

January 2015
V1.0
This page is intentionally blank
## GLOSSARY OF TERMS & ACRONYMS

<table>
<thead>
<tr>
<th>ACRONYM</th>
<th>DETAIL</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACMG</td>
<td>American College of Medical Genetics and Genomics</td>
</tr>
<tr>
<td>ASHG</td>
<td>American Society of Human Genetics</td>
</tr>
<tr>
<td>CAP</td>
<td>College of American Pathologists</td>
</tr>
<tr>
<td>CDC</td>
<td>Centers for Disease Control and Prevention</td>
</tr>
<tr>
<td>CLINGEN</td>
<td>NIH funded resource through ICCG</td>
</tr>
<tr>
<td>CLINVAR</td>
<td>ClinVar provides a freely accessible, public archive of reports of the relationships among human variations and phenotypes, with supporting evidence. Supported by National Center for Biotechnology Information (NCBI)</td>
</tr>
<tr>
<td>CLSI</td>
<td>Clinical and Laboratory Standards Insitute</td>
</tr>
<tr>
<td>COSMIC</td>
<td>Catalogue of Somatic Mutations in Cancer - genetic database</td>
</tr>
<tr>
<td>DECIPHER</td>
<td>Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources</td>
</tr>
<tr>
<td>DMuDB</td>
<td>The Diagnostic Mutation Database, NGRL, Manchester, UK</td>
</tr>
<tr>
<td>EMBL-EBI</td>
<td>The European Bioinformatics Institute - Part of the European Molecular Biology Laboratory</td>
</tr>
<tr>
<td>ESHG</td>
<td>European Society of Human Genetics</td>
</tr>
<tr>
<td>GA4GH</td>
<td>Global Alliance for Genomics and Health</td>
</tr>
<tr>
<td>GINA</td>
<td>The Genetic Information Nondiscrimination Act</td>
</tr>
<tr>
<td>HGNC</td>
<td>HUGO Gene Nomenclature Committee</td>
</tr>
<tr>
<td>HGSA</td>
<td>Human Genetics Society of Australasia</td>
</tr>
<tr>
<td>HIPAA</td>
<td>Health Insurance Portability and Accountability Act of 1996</td>
</tr>
<tr>
<td>HITECH</td>
<td>The Health Information Technology for Economic and Clinical Health Act</td>
</tr>
<tr>
<td>HVP</td>
<td>Human Variome Project</td>
</tr>
<tr>
<td>ICCG</td>
<td>International Collaboration for Clinical Genomics</td>
</tr>
<tr>
<td>ISO</td>
<td>International Standards Organisation</td>
</tr>
<tr>
<td>MAWSON</td>
<td>Online Genetic Familial Cancer Data Repository</td>
</tr>
<tr>
<td>NATA</td>
<td>National Association of Testing Authorities, Australia</td>
</tr>
<tr>
<td>NCOPP</td>
<td>National Coalition of Public Pathology</td>
</tr>
<tr>
<td>NGRL</td>
<td>National Genetics Reference Laboratory, Manchester, UK</td>
</tr>
<tr>
<td>NHGRI</td>
<td>National Human Genome Research Institute (NIH)</td>
</tr>
<tr>
<td>NIH</td>
<td>National Institutes of Health</td>
</tr>
<tr>
<td>NIST</td>
<td>National Institute of Standards and Technology</td>
</tr>
<tr>
<td>NPAAC</td>
<td>National Pathology Accreditation Advisory Council, Australia</td>
</tr>
<tr>
<td>Pathology Australia</td>
<td>Pathology Australia</td>
</tr>
<tr>
<td>PHG</td>
<td>Public Health Genomics Foundation - UK</td>
</tr>
<tr>
<td>RCPA</td>
<td>Royal College of Pathologists of Australasia</td>
</tr>
<tr>
<td>Wellcome Trust</td>
<td>The Human Genome</td>
</tr>
</tbody>
</table>
2 EXECUTIVE SUMMARY

The Purpose of this Final Report is to provide a description of work completed which addresses the reporting requirements of the Department of Health per the Deed of Variation No.2., executed on 25 August 2014 and summarised below.

- All work undertaken during the activity period of the project including an evaluation of the activity against the performance indicators
- A report on workshop and consultation
- The extent to which the activity achieved its goal to develop standards to accredit DNA sequence variant databases and how the results of this activity will be used to benefit pathology stakeholders
- Commentary on any issues, problems, or delays that the RCPA experienced in its performance of the activity, and an explanation of how the RCPA deal with the issues, problems or delays,
- The extent to which the activity achieved the Quality Use of Pathology Program Objectives.

THE STANDARDS DEVELOPMENT PROJECT

While there are numerous initiatives directed at the integration of emerging genomic technologies into mainstream clinical diagnostics, there are no specific standards or equivalent mechanisms which exist to guide the accreditation of DNA databases to ensure the accuracy and quality of uploaded data into any central repository to meet the needs of the clinical diagnostics environment.

An Australian national project led by the Royal College of Pathologists of Australasia (RCPA) in collaboration with the Human Genetics Society of Australasia (HGSAA), and the Human Variome Project (HVP), developed these standards for DNA sequence variant databases intended for clinical use. This project has been supported by an unrestricted grant from the Australian Department of Health’s Quality Use of Pathology Program (QUPP).

The standards are a broad reaching set of national standards that are sympathetic to the rapidly changing landscape of clinical genomics, and can be applied to both existing and future databases. The fundamental principle of the document is to provide regulatory oversight for DNA sequence variant databases intended to provide utility in clinical diagnostic service delivery ensuring that they are developed, curated, and maintained as safe, secure, and accurate repositories of genomic data. These standards are intended to complement existing laboratory standards and accreditation requirements, act as a guide to identify a quality database, assist the development of new databases, and improve existing databases that have evolved from a research environment.

Regulating the quality, accuracy, and clinical relevance of DNA sequence variant databases and the data held within them through the implementation of these standards will reduce the risk of aberrant or uninformative variants being reported, promote the sharing of clinical quality sequencing, and accelerate the delivery of accurate, actionable, and efficient clinical reports to improve patient management and outcomes.
This Final Report discusses:

- Key objectives, and outcomes
- Scope (and exclusions)
- Governance and controls
- Risks and risk mitigation
- Alignment with QUPP objectives
- Detailed description of the work undertaken to deliver a completed standards document including:
  - Background review including literature and global initiatives providing insight into current conditions
  - Standards Framework development
  - Standards document development
  - Steering committee, local and international consultation and workshop output; and its influence on the evolving standards document
  - Communications undertaken to promote awareness of the standards development
3 PROJECT STATEMENT

QUPP OBJECTIVES:
The goal of the QUPP is to achieve improvement in health and economic outcomes from the use of pathology in health care, through the pursuit of better practice amongst requesters (or referrers) and providers of pathology services and knowledgeable and engaged consumers.

The Quality Use of Pathology Program has three sub-programs and objectives:

- **Quality Consumer Services**: To develop and improve consumer-focused, accessible and coordinated services that promote informed choice and meet consumer needs.

- **Quality Referrals (Requesting or Ordering)**: To support referral practices that are informed and facilitated by best practice professional relationships and protocols between referrers and providers; that:
  - Are informed by evidence
  - Maximize health benefits, and
  - Inform and engage consumers.

- **Quality Pathology Practice**: To support professional practice standards that meet consumer and referrer needs and provide evidence-based, best practice, quality-assured services that are safe, cost effective, and efficient.

**PROJECT OBJECTIVE**:
To develop national standards that will assist laboratories to:

- Assess and ensure a DNA sequence variant data base has an appropriate level of functionality and houses data of sufficient quality to:
  - Be fit for their intended use by diagnostic laboratories for efficient, informative analysis and interpretation of genetic variants; to improve clinical decision making, counselling, and treatment planning,
  - Place “acceptable” data in the Human Variome Project Data Repository, and

- Seek accreditation of:
  - A laboratory’s in-house sequence variant database or
  - The use of external sequence variant database/s which are an integral part of their analysis and interpretation pipelines.

**PROJECT OUTCOME**:
To deliver a set of standards for Clinical DNA Sequence Variant Databases, this can be submitted to:

- The RCPA Board for approval; and
- The Australian Department of Health.

A long term outcome of this project is to seek to incorporate the standards as a Tier 4 NPAAC Reference Material to enable use in conjunction with existing NPAAC reference materials in the laboratory accreditation process via NATA to ensure evidence based, best practice; quality assured services that are safe, cost effective and efficient.
4 SCOPE

4.1 Purpose
To develop a set of national standards that are sympathetic to the rapidly changing landscape of genomics in the clinic to seek compliance by both existing and future databases. The fundamental principle of the document is to provide a standard for oversight for DNA sequence variant databases intended to:

- Provide utility in clinical diagnostic service delivery, and thereby ensure that they are developed, curated, and maintained as safe, secure, and reliable repositories of genomic data.
- Complement existing laboratory standards and accreditation requirements
- Align with global initiatives and guidelines in existence
- Act as a guide to identifying a clinical grade quality database for interpretation of clinical significance of sequence variants
- Act as a guide to laboratories establishing new databases, and improving existing databases that have evolved out of a research environment
- Set minimum requirements for clinical purposes within the boundaries of existing legislation both nationally and globally
- Be applicable to both in-house laboratory sequence variation databases (clinical service providers), locus and/or disease specific databases (LSDB’s), and population genomic databases
- Provide standards for laboratories to accredit in-house databases for clinical use

4.2 Meeting QUPP Objectives
The goal of the QUPP is to achieve improvement in health and economic outcomes from the use of pathology in health care, through the pursuit of better practice amongst requesters (or referrers) and providers of pathology services and knowledgeable and engaged consumers. This project addresses the Quality Use of Pathology Program “Quality Pathology Practice”, being to support professional practice standards that meet consumer and referrer needs and provide evidence-based, best practice, quality-assured services that are safe, cost effective and efficient.”

The development of the standards document aligns with the QUPP objectives as follows: Regulating the quality, accuracy and relevance of DNA sequence variation databases and the data held within will:

- Facilitate improved, streamlining of laboratory analysis and interpretation of variants into either benign, pathological, or variants of unknown or uncertain significance (VOUS)
- Reduce the risk of aberrant/uninformative/false reports being issued in a clinical environment,
- Promote and facilitate the sharing of clinical grade sequencing data
- Promote and demonstrate the feasibility of the inclusion of genomic information in clinical service delivery through expediting the informative reporting of actionable variants in a timely manner to practicing clinicians,
- Thereby improving the quality of patient management and/or outcome.
4.3 Exclusions

The standards will not include prescriptive requirements regarding:

- The amount of patient phenotype information that is to be provided. Clinical laboratories generally have a lack of control over the quality and volume of clinical information provided to them. The standards WILL however include what data should be entered into the database in the event it is provided to the laboratory, and standardisation of terminology used for phenotype (defined ontology).
- Ownership and / or location (physical server housing) of databases
- Implementation of the standards into the laboratory accreditation process
5 GOVERNANCE

5.1 Structure

The governance structure is represented in figure 1. Dotted lines indicated the communication and reporting requirements between the Department of Health, and the RCPA.

The Project Officer acted as the lead and overseer of the project conducting all administrative, logistic, communications, and other organisational activities as required.

Figure 1. Project governance structure
### 5.2 Roles and responsibilities

<table>
<thead>
<tr>
<th>ROLE</th>
<th>RESPONSIBILITIES</th>
</tr>
</thead>
</table>
| **Project Officer**               | • Develop an agreed Framework for the Standard for the Accreditation DNA Sequence Variations Databases consistent with the NPAAC Standards approach.  
                                      • Review Accreditation Standards for DNA Sequence Variation Data Bases (if any) and for databases storing health data at the international and national level.  
                                      • Consult widely with experts and stakeholders in DNA sequence data bases.  
                                      • Undertake a gap analysis of what occurs in Australia and what is occurring internationally.  
                                      • Develop Standards for Accreditation of DNA Sequence Variation Data Bases.  
                                      • Ensure the Standards are sent out for broad consultation.  
                                      • Assist Steering Committee to review and evaluate the Standards document as required.  
                                      • Ensure the Standards are recommended to RCPA Council for Approval as Policy document.  
                                      • Maintain records of project-related expenditure  
                                      • Provide Progress and final Reports for the Project for the Department of Health and Ageing |
| **Genetic Pathologist on staff 0.1FTE** | • Assist the Project Manager per above                                                                                                           |
| **Project Steering Committee**    | • TERMS OF REFERENCE (See Appendices Item 1.1)                                                                                                   |
| **Project Consultation Group**    | • Address key issues regarding database quality control including ISO standards, security, access, backup, patient consent and sharing data  
                                      • Facilitated discussion resulting in constructive input for feedback / response                                                             |
| **Manager**                       | • Oversee project officer (RCPA CEO)                                                                                                             |
5.3 Constraints

5.3.1 Reporting Requirements

- November 12 2013  Detailed activity work plan  Completed on time
- January 13 2014  First Performance Report  Completed on time
- May 13 2014  Second Performance Report  Completed on time
- November 24 2014  Draft Final Report  Completed on time
- January 19 2015  Final Report  Completed on time

A deed of variation was issued in August 2014 to reflect the activity end date of January 2015 which was not initially considered in the original deed of variation. This moved the Draft Final report from August to November 2014, and the Final report from September 2014 to January 2015. All other reporting requirements have been maintained, and have been met to date.

5.3.2 Activity Performance Indicators

- April 2014  Framework developed for development of standards
  - The initial framework was developed within this timeframe, reported in the second performance report in May 2014; and was submitted to Applied & Translational Genomics in April 2014, published July 2014.
  - It should be noted that as with any document under development, the framework was considered an evolving document, resulting in a second iteration being produced following the Consultation workshop conducted in August 2014
- May 2014  Literature review completed
  - The literature review is ongoing as new developments occur globally, however a comprehensive review was performed prior to May, and formed a body of discussion within the Applied & Translational Genomics manuscript published in July 2014.
  - The literature review continues for the duration of the project to ensure all relevant items and initiatives are not overlooked.
- July 2014  Report on workshops and consultations with experts
  - The consultation workshop was held in August 2014 due to limited availability of both steering committee members, and invited experts.
  - This workshop informed a heavily revised document which forms the basis for the final document to be submitted to the RCPA Board for approval.
- January 2015  Standards developed for DNA sequence variation databases
  - The Standards document was submitted to the RCPA Board at the December 2014 meeting, and was approved by the Board. The Standards were released on December 22 2014.

5.4 Controls

5.4.1 Steering Committee Terms of Reference

A Steering Committee terms of reference (TOR) was developed and circulated to all members in November 2013 to ensure they were aware of their role in, and responsibility to this project. The terms of reference may be found in Appendices Item 1.1.
5.4.2 Implementation Milestones [KPI’s] - Key Decision Points / Milestones
See 5.3.2 – Activity Performance Indicators

5.4.3 Contingency Planning (Risk Identification)

<table>
<thead>
<tr>
<th>The Risk</th>
<th>Source</th>
<th>Impact</th>
<th>Risk mitigation treatment</th>
<th>Likelihood¹</th>
<th>Consequence²</th>
<th>Risk rating²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficulty finding genetic pathologist</td>
<td>Workforce shortage</td>
<td>Project may be difficult to manage.</td>
<td>Remunerate</td>
<td>Possible</td>
<td>Moderate</td>
<td>Medium</td>
</tr>
<tr>
<td>Time constraints on participants</td>
<td>Workforce shortage, limited availability of stakeholders</td>
<td>Difficult to maintain timelines</td>
<td>Clear communication, minimisation of face to face requirements, frequent monitoring and revision of schedule as required</td>
<td>Almost certain</td>
<td>Moderate</td>
<td>High</td>
</tr>
<tr>
<td>Difficulty in agreeing to what to include in Standard.</td>
<td>Difference of professional opinion.</td>
<td>Project could not be completed.</td>
<td>Involve senior opinion leaders in the project and on the Steering Committee.</td>
<td>Possible</td>
<td>Major</td>
<td>Medium</td>
</tr>
<tr>
<td>Lack of acceptance by Fellows.</td>
<td>Difference of professional opinion.</td>
<td>Not desirable but need to have Standards.</td>
<td>Transparent communication of objectives, clear reasoning for the need for Standards, involvement of stakeholders throughout the development process</td>
<td>Possible</td>
<td>Minor</td>
<td>Medium</td>
</tr>
</tbody>
</table>

1 Likelihood: Rare, Unlikely, Possible, Likely, Almost Certain
2 Consequence: Insignificant, Minor, Moderate, Major, Severe
3 Risk Rating: Low, Medium, High, Extreme

5.5 Authority

The final standards must be approved and endorsed by the RCPA Board prior to their release as an RCPA standard.
5.6 Related activity/projects

- RCPA: The Pathology Information, Terminology and Units Standardisation (PITUS) Project – in particular genomics/informatics sub-committee.

- RCPA Informatics Advisory Committee
  - Project officer attended Informatics Advisory Committee meeting (July 3, 2014) to provide update and discuss project. Input received from Informatics Advisory Committee representatives.

- Human Variome Project (HVP)
  - The Australian Node of HVP feeds into the HVP International Network which promotes standardisation of databases and curation to facilitate responsible sharing of genomic and clinical data globally.

- Global Alliance for Genomics and Health (GA4GH)
  - More than 170 member organisations globally
  - Established to address security, regulatory and ethics, and clinical priorities which include the development of formal data models, application programming interface (API) implementations for submitting, exchanging, querying, and analyzing genomic data.

- DMuDB (The Diagnostic Mutation Database – UK)
  - A UK based NHS initiative to facilitate data sharing efforts, accessible by licence (with fee to maintain the database) to approved bodies directly involved in diagnostic / clinical patient care.

- ClinGen
  - A US based NIH funded initiative that is engaging the genomics community in data sharing efforts, develop the infrastructure and standards to support curation activities, and incorporate machine-learning approaches to speed the identification of clinically relevant variants. ClinGen has established ClinVar, a freely accessible public archive of reports of the relationships among human variations and phenotypes, with supporting evidence. ClinVar thus facilitates access to and communication about the relationships asserted between human variation and observed health status, and the history of that interpretation.
6 DETAILED WORK REPORT

6.1 Situation Analysis

It has now become routine practice to compare sequence variations identified during clinical genetic testing with variants recorded in a wide range of genetic variation databases as well as in the scientific literature to aid in understanding the potential clinical significance and determining a definitive diagnosis. Although numerous genetic variation databases already exist, there are few that meet the accuracy and reproducibility required for clinical diagnostics. Current databases are of variable quality and many contain errors in variant calls, non-standardised nomenclature, incomplete pathogenicity associations and limited phenotypic information linked to genomic data. These all represent limitations and risks to the quality of patient care. Based on current research experience of highly curated mutation data, the curation of databases to clinical standards is likely to require a substantial investment of time and effort.

The increasing ease of access to technologies such as massively parallel sequencing (MPS, also referred to as Next Generation Sequencing (NGS)) is producing increasing volumes of genomic data that needs to be recorded in an organised, accurate manner. The integrity of this stored data is critical as there becomes a greater demand for analysis and interpretation in clinical research and diagnostics, a task which now forms a substantial proportion of the genetic diagnostic workload.

There are numerous initiatives and white papers, which discuss the steps needed to allow for responsible integration of emerging genomic technologies into mainstream clinical diagnostics, many of which touch on data quality and collection. A detailed analysis can be found in section 6.2.1 Review Outcomes.

Despite the initiatives and guidelines described in section 6.2.1 Review Outcomes, there are no specific standards or equivalent mechanisms which concentrate on guiding the accreditation of DNA sequence variation databases to ensure the accuracy, quality, and ongoing maintenance of uploaded data into any central repository to meet the needs of the clinical diagnostics environment.

This Australian national project led by the Royal College of Pathologists of Australasia (RCPA) in collaboration with the Human Genetics Society of Australasia (HGSa) and the Human Variome Project (HVP) has sought to fill this gap by developing standards for DNA sequence variation databases intended for use in the clinical environment. This project is being supported by the Australian Department of Health’s Quality Use of Pathology Program (QUPP).

The standards are a broad reaching set of national standards that are sympathetic to the rapidly changing landscape of genomics in the clinic to seek compliance by both existing and future databases. The fundamental principle of the document is to provide a standard for oversight for DNA sequence variation databases intended to provide utility in clinical diagnostic service delivery, and thereby ensure that they are developed, curated, and maintained as safe, secure, and accurate repositories of genomic data. They are intended to
complement existing laboratory standards and accreditation requirements, align with global initiatives and guidelines in existence, act as a guide to identify a quality database, establish new databases as well as improve existing databases that have evolved out of the research environment, and set minimum requirements for clinical purposes within the boundaries of existing legislation both nationally and globally.
6.2 Global Review

Per the Project Implementation Plan and work activity schedule, the Project has conducted a broad literature search and webcraw. A detailed review of Global activities / initiatives taking place in the space of DNA Sequence Variation Databases Standards has concluded that while there are numerous initiatives directed at the integration of emerging genomic technologies in to mainstream clinical diagnostics, there are currently no specific standards or equivalent mechanisms which exist to guide the accreditation of DNA databases to ensure the accuracy and quality of uploaded data into any central repository to meet the needs of the clinical diagnostics environment.

While there are no current standards that specifically address DNA Sequence Variation Database requirements that need to be met in a clinical setting;
- There are existing standards and guidelines for genetic testing which will impact the implementation of database standards.
- There are multiple bodies globally who are in the initial stages of considering the development of standards, which either specifically address or overlap the construction and utilisation of sequence variation databases for clinical diagnostic use. They are however in preliminary phases of development.

The reviews included:
- Databases currently operating, and accessed by genetic scientists, pathologists, and clinicians to facilitate patient diagnosis.
- Databases currently in development or planning which are specifically intended to be used in a clinical diagnostic setting to facilitate and inform patient management
- Associated policy development activities and initiatives to govern these databases.

The Steering Committee was informed of:
- Other government funded bodies who are working in the area of development of clinical grade genomic databases and / or associated policy development
  - E.g.: ICCG, NIH NHGRI policy, funded research and CLINVAR, Genomics England, PHG, EMBL-EBI, Wellcome Trust, and NGR.
- Other professional bodies who may also be considering policy development in this area
  - E.g.: ASHG, ACMG, ESHG,
- National/Global government funded databases currently in existence or under development
  - E.g.: DMuDB (NGRL, UK), CLINVAR (NIH, USA)
- Locus specific databases (LSDB) [generally run by institutions or consortiums]
  - E.g. MAWSON, InSiGHT, DECIPHER, COSMI
- Existing policies / standards / Acts that are inter-related to the area of clinical diagnostic databases such as health informatics, data sharing and security, privacy standards / laws, and existing nomenclature / ontologies
  - E.g.: HIPAA, HITECH, GINA; ISO, NIST, CLSI, HGNC, EMBL-EBI, and NPAAC.
A record of all relevant sites and documents viewed was held in an excel workbook, example worksheet can be found in Appendix Item 1.2.

In addition to the literature and webcrawl, the project officer reached out to key contacts in the US and EU seeking further information on activities in this area; as well as initiating relationships and discussions intended to ensure alignment long term.

In particular, feedback has been provided from:

- DMuDB, NGRL UK (Kathryn Robertson, Andrew Devereau)
- ICCG, CLINGEN project representative (Heidi Rehm)
- Global Alliance for Genomics and Health (GA4GH) Initiative – recent meeting held in UK, March 2014 (Mark Cowley), also represented by Kathryn North, and attended by Clara Gaff and Andrew Sinclair
- Tina Hambuch – serves on Clinical Genomics working groups for the Centers for Disease Control (USA), Association of Molecular Pathology, and Clinical Laboratory Standards Institute (CLSI), and various working groups, advisory committees in the area of clinical genomics in the USA. Tina provided an overview of all activities being undertaken in the USA, and their current status. It was felt that none of the current activities currently had a concentrated focus on this type of standard development
- CAP was engaged through Nazneen Aziz. Any activities in this area are in very early stages, and CAP was interested in keeping in close contact. Nazneen has since moved on, and a new contact is being sought within CAP and appropriate CAP committees.

Potential collaborations may be explored or engaged in to leverage synergies and promote alignment of the standards globally.

6.2.1 Review Outcomes

Those initiatives and papers specifically addressing the steps needed to allow for responsible integration of emerging genomic technologies into mainstream clinical diagnostics which touch on data quality, collection and sharing are discussed here.

6.2.1.1 INITIATIVES:

The Global Alliance for Genomics and Health (GA4GH) 2014

The Global Alliance for Genomics and Health (Global Alliance) is an international coalition, dedicated to improving human health by maximising the potential of genomic medicine through effective and responsible data sharing. The promise of genomic data to revolutionise biology and medicine depends critically on our ability to make comparisons across millions of human genome sequences, but this requires coordination across organizations, methods, diseases, and even countries. The members of the Global Alliance for Genomics and Health are working together to create interoperable approaches and catalyse initiatives that will help unlock the great potential of genomic data.

The Global Alliance for Genomics and Health is an independent, non-governmental alliance, made up of hundreds of world-leading organisations and individuals from across the world. The Global Alliance is focused on bringing together a diverse set of key stakeholders across regions.
and sectors, including leaders in healthcare and research, patient and disease advocacy, and life science and information technology.

Since its formation in 2013, the Global Alliance for Genomics and Health is leading the way to enable genomic and clinical data sharing. The Alliance’s Working Groups are producing high-impact deliverables to ensure such responsible sharing is possible, such as developing a Framework for Data Sharing to guide governance and research and a Genomics API to allow for the interoperable exchange of data. The Working Groups are also catalyzing key collaborative projects that aim to share real-world data, such as Matchmaker Exchange, Beacon Project, and BRCA Challenge.

**The UK 100,000 Genomes Project April 2013**
The British Society for Genetic Medicine (BSGM) made public the outcomes of the BSGM 100,000 Genome Group which made recommendations on the collaborative development of appropriate genomic standards and policies, promotion of data sharing and further development of the existing NHS Diagnostic Mutation Database (DMuDB) and DECIPHER database to be more readily usable for the clinical laboratory ix. This was followed up by reports on recommendations from United Kingdom appointed working groups x.

**eMERGE Network 2014**
Recent challenges being addressed by the eMERGE network and others include collection of phenotype data, the integration of genomic findings into electronic health records, and the current efforts to extend HL7 Version 2 vocabularies for exome and whole genome sequencing within the context of clinical workflows viii, ix.

**ClinGen / ClinVar, International Collaboration for Clinical Genomics (ICCG) 2013/14**
In September 2013, the National Human Genome Research Institute (NHGRI) and the Eunice Kennedy Shriver National Institute of Child Health and Development (NICHD) awarded USD25M to support a consortium of three groups to design and implement a framework for evaluating variants, and their role in patient care. This consortium is enabling access to this information through the NCBI ClinVar database. The International Collaboration for Clinical Genomics (ICCG) is a part of this project, and is intended to support data collection and sharing."
6.2.1.2 BEST PRACTICE POLICIES AND GUIDELINES

In addition to the white papers and initiatives, there are a growing number of best practice policies and guidelines addressing the responsible integration of genomics into a clinical environment. These include:

Royal College of Pathologists of Australasia (RCPA)

*Implementation of Massively Parallel Sequencing in Diagnostic Medical Genetic Testing May 2014*

In March 2011, the Royal College of Pathologists of Australasia (RCPA) held a workshop on the implementation of "next generation sequencing" in diagnostic laboratories across Australia. There was agreement that a guideline be developed for diagnostic laboratories wanting to implement next generation sequencing. The goal was to develop a tool that would assist such labs in identifying the issues to be addressed to ensure that the high quality of medical genetic testing across Australia is maintained. As yet, there is no accreditation standard for the implementation of next generation sequencing in Australian diagnostic laboratories. The resulting guideline document approved in May 2014 is the basis for such a standard in the future.

The guideline document includes sections addressing bioinformatics and IT infrastructure in which databases are touched on, and the standards document developed as a result of this project is intended to complement the existing guideline document, and inform future diagnostic medical genetic testing standards development.

American College of Medical Genetics and Genomics (ACMG)

*ACMG clinical laboratory standards for next generation sequencing. September 2013*

These standards refer briefly to data format, storage, and sharing in sections D3.4, F6, G2, and conclude with the following statement:

“Although key aspects of the clinical implementation of NGS technology have been addressed, additional recommendations regarding specific applications of the technology are still needed, and will continue to be needed as the field evolves. This will require an ongoing dialogue among those already engaged in this pursuit, those determining how to become involved in this new paradigm of molecular testing, and those who will be responsible for ordering and communicating NGS results to patients”.
Next-generation Sequencing: Standardisation of Clinical Testing (Nex-StoCT)
Workgroup Principles and Guidelines

Assuring the Quality of Next-Generation Sequencing in Clinical Laboratory Practice. November 2012 (Supplementary Guidelines, Section 4)

These guidelines represent initial steps to ensure that results from tests based on next-generation sequencing (NGS) are reliable and useful for clinical decision making. This group was convened by the US Centers for Disease Control and Prevention (CDC) which collaborated to define platform-independent approaches for establishing technical process elements of a quality management system to assure the analytical validity and compliance of NGS tests with existing regulatory and professional quality standards.

While they touched on the informatics pipeline, database elements were not thoroughly addressed.

Association for Clinical Genetic Science (ACGS, part of the British Society of Genetic Medicine (BSGM)) and The Dutch Society of Clinical Genetic Laboratory Specialists (VKGL)

Practice Guidelines for the Evaluation of Pathogenicity and the Reporting of Sequence Variants in Clinical Molecular Genetics. September 2013

The standards outlined in this document are a guide to assess the clinical significance of DNA sequence variants for situations where there is likely to be a clinical benefit. The authors and ratifying bodies acknowledged that the guidelines are aspirational and that the challenges of implementation may lead to further revision.

Variant databases including LSDBs are addressed in section 4.1, with a brief paragraph as follows:

“It is essential that LSDBs are used where available and that staff carrying out searches should be appropriately trained in the use of databases. LSDBs that contain references to the published literature should be used in preference to those that do not. It is highly desirable that all laboratories working within the public health sector have a policy of submitting all variants (excluding known – non-pathogenic polymorphisms) to the most appropriate LSDBs where this facility is available. One central, curated variant database to which reference alignments are attached would be ideal; however this is not currently available.”
The Clinical Molecular Genetics Society, UK (CMGS), part of the British Society of Genetic Medicine (BSGM)

*Practice Guidelines for Targeted Next Generation Sequencing Analysis and Interpretation. December 2012*

These guidelines aim to establish consensus standards for identifying and reporting mutations. They identify common elements for each part of the process and specify quality criteria that should be met or exceeded. Guidelines are described as either “Essential” or “Recommended”.

Annotation of variants is briefly addressed in section 4.5, while data storage is addressed in section 5. Guidelines from the Royal College of Pathologists and the Institute of Biomedical Science (UK) recommend that data and records pertaining to pathology tests are retained for a minimum of 25 years. Challenges of storage of next generation sequencing data are discussed briefly, and the guidelines states that it is essential practice to store the output file from the variant annotation step (e.g.: vcf file), and some laboratories may choose to also retain the FastQ, SAM or BAM files in order to re-analyse the read data in the future. A log of the informatics processing applied to the raw data to create the mapping files is also recommended.

“The targeted next generation sequencing services were introduced into the UK NHS in 2010 and whilst the laboratory and analytical protocols are both varied and rapidly evolving; these guidelines aim to describe the general principles that underlie the quality requirements of this technology. This document considers quality aspects of the whole process of targeted next generation sequencing and makes the assumption that the testing process takes place in an appropriate, accredited laboratory setting where routine aspects of good laboratory practice such as sample tracking and record keeping are in place. Local sequencing practices may vary both in terms of the targeting strategy, sequencing chemistry/hardware, analysis software and reporting of results. These guidelines are based on the principles established for Sanger sequencing (CMGS 2009)”

*National Collaborative Study of Dutch Genome Diagnostic Laboratories*

*Best Practice Guidelines for the Use of Next-Generation Sequencing Application in Genome Diagnostics: A National Collaborative Study of Dutch Genome Diagnostic Laboratories. October 2013*

The aim of this working group was to define best practice guidelines for implementing and validating NGS applications in a diagnostic setting, in line with the requirements of the international diagnostic accreditation standard (ISO 15189). These NGS guidelines and criteria describe key technical and process elements to assure analytical validity and compliance with regulatory and professional quality standards.

In this document, data storage and databases are discussed in a dedicated section. In summary:
“It is essential to store VCF files for an unlimited time, together with documentation on the pipeline version used and statistics on coverage.

We recommend storing FASTQ or alignment files for at least 1 year, keeping in mind that longer storage might be required for future data assessment.

We recommend creating national and international databases for the submission of identified and validated variants for diagnostic laboratory use only.”

NIH (USA) xvi

NIH Genomic Data Sharing Policy August 2014

To promote robust sharing of human and non-human data from a wide range of genomic research and to provide appropriate protections for research involving human data, the National Institutes of Health (NIH) issued the NIH Genomic Data Sharing Policy (GDS Policy) on August 27, 2014.

This document while directed towards NIH funded researchers does address data sharing, data repositories, privacy and confidentiality of data within repositories, and conditions for use of restricted and unrestricted access data.

The National Consortium for Data Science Data to Discovery: Genomes to Health xviii

The National Consortium for Data Science (USA) (www.data2discovery.org) was established in February 2013 as a public/private partnership of leading universities, governmental agencies, and businesses. The mission of the NCDS is to provide the foundation needed to advance the field of data science through research, education, and economic opportunity.

As part of its efforts to advance data science, the Consortium held the first annual Data Science Leadership Summit in April 2013. The theme of the summit was “Data to Discovery: Genomes to Health.”, the outcomes of which were published in 2014 Genomics was chosen as the initial theme because the field offers some of the greatest challenges, as well as some of the greatest promises, in data science. The purpose of the summit was to gather together leaders in data science and genomics to discuss current challenges in genomics, brainstorm for solutions, and reach consensus on the most appropriate and technologically advanced recommendations to address the challenges.

Over 70 leaders in data science and genomics gathered at the summit to discuss critical challenges in six areas of genomics, all of which touched on critical elements of databases, as summarised below.

- Data Provenance, Collection, and Management

Maintaining data provenance, without knowledge of how exactly data will be repurposed and reused, presents challenges to the collection and management of big data sets. These challenges are amplified by the technological complexities associated with data derived from
multiple sources and incompatible or legacy data management systems. Trained data and information specialists are required to manage all aspects of the life cycle of genomic data.

- **Delineation of Phenotypes**

  The characterization of rich and complex phenotypes in the absence of standardization of data elements and harmonization of existing data sets presents significant challenges in the delineation of phenotypes. The lack of adequate tools and techniques to extract phenotype data from large data sets further complicates this matter.

- **Adjudication of Genomic Variants**

  The adjudication of variants is complicated by the lack of standards for both phenotype and variant data, evolving expert consensus on the functional impact of variants, and an absence of high quality, non-proprietary, clinical-grade variant databases. There is also a need for new analytical tools to evaluate rare variants, epistasis, and epigenetic influences.

- **Biostatistics and Bioinformatics**

  Challenges in biostatistics and bioinformatics are amplified on a large scale due to inadequate statistical models and software, insufficient computer processing power or unacceptable time delays when running complex models, heterogeneous and incomplete data sets, limited tools to analyze rare sequence variants and mutations along entire pathways or systems, and limited adoption of federated distributed data systems to promote data integration and sharing. There is also a need for a workforce of skilled analysts who are trained to handle large data sets, as well as effective Clinical Decision Support systems to guide interpretation of analytical results.

- **Data Sharing**

  Data sharing is complicated by the risk of re-identification of de-identified genomic samples and the lack of policies and procedures to protect against genomic breaches and to ensure data provenance as data is repurposed and reused. Incentives are needed to promote data sharing and encourage compliance with policies and procedures related to privacy, security, and provenance.

- **Bioethics and the Law**

  Unresolved bioethical issues abound in genomics, including those related to incidental findings, disclosure of genetic test results, and privacy issues affecting sensitive populations. There also is a need for open discussions on the legal distinctions between physical property versus intellectual property versus informational property and between privacy versus confidentiality of genomic data.
National Human Genome Research Institute (NHGRI (USA))

*Guidelines for investigating causality of sequence variants in human disease. April 2014*\textsuperscript{viii}

This paper describes the challenges in reliably investigating the role of sequence variants in human disease, and approaches to evaluate the evidence supporting variant causality. It represents the conclusions of a working group of experts in genomic research, analysis and clinical diagnostic sequencing convened by the US National Human Genome Research Institute.

Their recommendations centre on five key areas: study design; gene-level implication; variant-level implication; publication and databases; and implications for clinical diagnosis.

Global Alliance for Genomics and Health (GA4GH)\textsuperscript{ix}

*Framework for Responsible Sharing of Genomic and Health Related Data September 2014*\textsuperscript{x}

This Framework is developed under the auspices of the Global Alliance for Genomics and Health. Its mission is to accelerate progress in human health by helping to establish a common Framework of harmonised approaches to enable effective and responsible sharing of genomic and clinical data and to catalyse data sharing projects that drive and demonstrate the value of data sharing.

This Framework provides guidance for the responsible sharing of human genomic and health-related data, including personal health data and other types of data that may have predictive power in relation to health. In particular, it highlights, and is guided by, Article 27 of the 1948 Universal Declaration of Human Rights. Article 27 guarantees the rights of every individual in the world “to share in scientific advancement and its benefits” (including to freely engage in responsible scientific inquiry), and at the same time “to the protection of the moral and material interests resulting from any scientific production of which [a person] is the author.”
6.2.1.3 LSDB SPECIFIC ESTABLISHED BEST PRACTICES & the Human Variome Project (HVP)

Further to these more recent NGS focussed documents and initiatives; collection of information related to genetic variation is not a new concept, with over 2,000 locus specific databases established with disease and/or gene specific variation information. There are currently no established ISO standards which govern sequence variation databases. There are however numerous de-facto standards and established best practicesxxi.

Elements of the ideal LSDB have been described by Cotton et.alxxii. These were adapted by the project officer over the course of this project to guide and support the development of the standards frameworkxxiii and can be found in documents presented at the British Society for Genetic Medicine Conference and Mutation Detection Conference in September 2014 (Appendices Items 1.5.3 and 1.5.4).

While the existing defacto standards aid with providing consistent formats, they are in part outdated as genomic data becomes more readily accessible and available.

With regard to guidelines for the establishment of locus specific databases (LSDBs), the Human Variome Project (HVP) has been collaborating with the Human Genome Variation Society, and the GEN2PHEN project, working towards standardising the way variation and pathogenicity data is presented.

In addition to these activities, Celli et.al. developed a supporting document describing curation of a gene variant database as first step to establishing guidelines for database curationxxiv; and Mitropolou et.al. explored the evolution and maturation of database content; the essential elements needing improvement to facilitate clinical usexxv. Both these documents have been drawn from to inform the development of the standards for clinical databases of genetic variants.

The HVP continues to promote global standards and guidelines which encourage the establishment and maintenance of quality-assured sequence variation data repositories.

6.2.1.4 EXISTING STANDARDS & LABORATORY ACCREDITATION MECHANISMS in AUSTRALIA

Accreditation of pathology laboratories for clinical service delivery in Australia is overseen by the National Pathology Accreditation Advisory Council (NPAAC). NPAAC is an agency within the Commonwealth (Federal) Department of Health. NPAAC plays a key role in ensuring the quality of Australian pathology services, and is responsible for the development and maintenance of standards and guidelines for pathology practicesxxvi.

The National Association of Testing Authorities (NATA) is the authority which provides independent assurance of technical competence in conjunction with the Royal College of Pathologists of Australasia (RCPA) through a proven network of best practice industry experts. NATA/RCPA provides assessment, accreditation, and training services to laboratories and technical facilities throughout Australia and internationallyxxvii. NATA audits against the standards and guidelines laid down by NPAAC.
Laboratories seeking eligibility for Federal government funding for medical tests are required to meet the specified quality standards as expressed by NPAAC in the context of the Australian pathology accreditation framework.

There are a number of specialised technical publications that specify requirements in laboratories undertaking specific areas of medical testing in addition to requirements for good medical practice in all pathology laboratories. The DNA Sequence Variation Database Standards under development are intended to be an adjunct to existing NPAAC standards and guidelines such as “Requirements for Medical Pathology Services” and “Requirements for the Retention of Laboratory Records and Diagnostic Material”.

When completed, the standards will be submitted for potential endorsement by the RCPA and HGSA boards, and will be made available as a tool for laboratories and NATA assessors alike to facilitate accreditation. Further, the RCPA will engage the NPAAC to seek their inclusion of these Standards in the Commonwealth Health Insurance (Accredited Pathology Laboratories) – Approval Principles 2002.

It is also recognised that there is a need to bridge a gap between the translational research environment and the clinical diagnostic environment, and therefore regulation of the use of data within the scope of the respective environments. To address this, in addition to the NATA/RCPA and NPAAC requirements, the Standards will encourage users to comply with the Australian Government National Health and Medical Research Council National Statement on Ethical Conduct in Human Research 2007 (Updated March 2014).
6.3 Standards Framework

The fundamental principle underpinning this document is that DNA sequence variant databases intended to provide utility in clinical diagnostic service delivery should be developed, curated, and maintained as safe, secure, and accurate repositories of genomic data.

These Standards have been developed with reference to current and proposed Australian regulations and standards from the International Organisation for Standardisation (ISO), including AS ISO 15189 Medical laboratories – Requirements for quality and competence.

In addition to these standards, existing NPAAC Requirements apply to all Laboratories seeking accreditation in Australia, and must be applied in conjunction with jurisdictional and other regulatory requirements. The standards should be used for guidance where accessing and or assessing databases outside the Australian accreditation framework.

In each section of the document, points deemed important for practice are identified as ‘Standards’ or ‘Commentaries’.

- A Standard is the minimum requirement for a procedure, method, staffing resource or facility that is required before a Laboratory can attain accreditation. Standards are printed in bold type and prefaced with an ‘S’ (e.g. S2.2). The word ‘must’ in each Standard within this document indicate a mandatory requirement for pathology practice. Commentary is provided to give clarification to the Standards as well as to provide examples and guidance on interpretation.

- Commentaries are prefaced with a ‘C’ (e.g. C1.2) and are placed where they add the most value. Commentaries may be normative or informative depending on both the content and the context of whether they are associated with a Standard or not. Note that when comments are expanding on a Standard or referring to other legislation, they assume the same status and importance as the Standards to which they are attached. As a general rule, where a Commentary contains the word ‘must’ then that Commentary is considered to be normative.

As a result of the global review to identify currently existing frameworks, standards, or guidelines which were likely to overlap with the work being conducted within the scope of this project; it was determined that while there are numerous initiatives directed at the integration of emerging genomic technologies in to mainstream clinical diagnostics, there are currently no specific standards or equivalent mechanisms which exist to guide the accreditation of DNA databases to ensure the accuracy and quality of uploaded data into any central repository to meet the needs of the clinical diagnostics environment.

A Steering Committee meeting Teleconference was held on February 06 2014 to discuss the literature and web reviews, and agree on a preliminary framework for the development of the standards.

Initially, the framework within which the standards were developed consisted of nine key elements (Table 1). These elements were used to address requirements in a systematic
# TABLE 1
Framework for development of standards for DNA sequence variation databases

<table>
<thead>
<tr>
<th>Framework Areas</th>
<th>Items Being Considered in Each of the Areas (Include, But Not Limited To)</th>
</tr>
</thead>
</table>
| **Purpose**     | • Scope of the database  
                  • Nature of information being held in the database  
                  • Quality parameters  
                  • Standard operating procedures |
| **Governance**  | • Custodian definition, accountability, and responsibility  
                  • Mechanisms for complaints, troubleshooting, auditing, and risk mitigation  
                  • Ethics committee, advisory board, and multidisciplinary team involvement  
                  • Sustainability, and contingency in case of demise  
                  • Compliance with jurisdictional legislations and or regulations |
| **Establishment** | • Principle hardware and software requirements including web interfacing, networking, infrastructure, storage, backup capabilities  
                  • Compatibility - external databases, electronic health/medical records (EMR/EHR), HL7 V2, SNOMED-CT, federated databases. |
| **Protection Privacy Security** | • Content of an Information Policy (such as how data are collected, used, disclosed, managed, administered, stored, and accessed)  
                  • Compliance with local Australian (Privacy Amendment Act 2012) and other jurisdictional legislation/regulation such as HIPAA.  
                  • Consent for storage of data, and use of data for diagnostic and or research purposes  
                  • Privacy, security through de-identification, data encryption, and protected access  
                  • Security breach management |
| **Content**     | • Data to be collected and submitted including but not limited to data structure, nomenclature and variant description, methodology used to detect the variant, orthogonal method verification, sequence quality data, reference genome, provenance of existing data, variant occurrences, inheritance information, phenotype, and clinical accreditation status of submitting laboratories. |
| **Functionality** | • Version control, modifications  
                  • Interrogation and return of information from external databases, linkage of variant occurrences and familial grouping  
                  • Mechanisms to track de-identified data to facilitate patient management. |
| **Currency of Information** | • Specific DNA database curation definition and requirements  
                  • Filtering and triaging variant calls, determination of relevance and inclusion  
                  • Quality controls and evaluation of level of confidence in accuracy  
                  • Maintaining relevance and accuracy of data,  
                  • Maintaining currency of genome builds and compatibility of variants recorded  
                  • Regular audits to assure quality of the database schema and data held within. |
| **Access & Sharing** | • Policy governing participation through access and sharing  
                  • Mechanisms for facilitating access and sharing through secure practices  
                  • User registration, and the clinical need to utilise the data  
                  • Communication between user and curator / custodian  
                  • Quality Control, auditing of access and sharing |
| **Professional Use** | • Standardising ontology within a database, or between federated databases  
                  • Variant classification, traceability of clinical reports, re-analysis  
                  • Skill sets, knowledge base, and experience required  
                  • Workforce training and development |
order with clearly defined and concise criteria. The draft standards document was presented to the consultation workshop attendees in this format. Over the course of the consultation workshop, the number of sections was revised down as it was felt that the contents of section 7 Currency of Information, 8 Access and Sharing, and 9 Professional Use, could be rolled in to previous sections of the document without disrupting the flow. In fact, it was felt the flow of the document would improve by changing the framework in this way. The resulting framework is as follows in Table 2, and is reflected in the final version of the standards document released in December 2014.

### TABLE 2
Revised Framework for development of standards for DNA sequence variation databases

<table>
<thead>
<tr>
<th>Framework Areas</th>
<th>Items Being Considered in Each of the Areas (Include, But Not Limited To)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose</td>
<td>• Essential elements</td>
</tr>
<tr>
<td></td>
<td>• Nature of use of database</td>
</tr>
<tr>
<td>Governance</td>
<td>• Custodian definition, accountability, and responsibility</td>
</tr>
<tr>
<td></td>
<td>• Curation, maintenance of relevance</td>
</tr>
<tr>
<td></td>
<td>• Mechanisms for complaints, troubleshooting, auditing, and risk mitigation</td>
</tr>
<tr>
<td></td>
<td>• Ethics committee, advisory board, and multidisciplinary team involvement</td>
</tr>
<tr>
<td></td>
<td>• Sustainability, and contingency in case of demise</td>
</tr>
<tr>
<td></td>
<td>• Compliance with jurisdictional legislations and or regulations</td>
</tr>
<tr>
<td>Establishment</td>
<td>• Principle hardware and software requirements including web interfacing, networking, infrastructure, storage, backup capabilities</td>
</tr>
<tr>
<td></td>
<td>• Compatibility - external databases, electronic health/medical records (EMR/EHR), HL7 V2, SNOMED-CT, federated databases.</td>
</tr>
<tr>
<td></td>
<td>• Quality</td>
</tr>
<tr>
<td></td>
<td>• Audit capabilities</td>
</tr>
<tr>
<td></td>
<td>• Backing up of database</td>
</tr>
<tr>
<td>Protection Privacy</td>
<td>• Content of an Information Policy (such as how data are collected, used, disclosed, managed, administered, stored, and accessed)</td>
</tr>
<tr>
<td>Security</td>
<td>• Access and Sharing Policy</td>
</tr>
<tr>
<td></td>
<td>• Compliance with local Australian (Privacy Amendment Act 2012) and other jurisdictional legislation/regulation such as HIPAA.</td>
</tr>
<tr>
<td></td>
<td>• Consent for storage of data, and use of data for diagnostic and or research purposes</td>
</tr>
<tr>
<td></td>
<td>• Security breach management</td>
</tr>
<tr>
<td>Content</td>
<td>• Minimum requirements of data to be collected and submitted including but not limited to data structure, nomenclature and variant description, methodology used to detect the variant, orthogonal method verification, sequence quality data, reference genome, provenance of existing data, variant occurrences, inheritance information, phenotype</td>
</tr>
<tr>
<td></td>
<td>• Analytical, clinical validity</td>
</tr>
<tr>
<td></td>
<td>• Laboratory accreditation</td>
</tr>
<tr>
<td></td>
<td>• Variant classification</td>
</tr>
<tr>
<td></td>
<td>• Use of terminology / ontology</td>
</tr>
<tr>
<td>Functionality</td>
<td>• Version control, modifications</td>
</tr>
<tr>
<td></td>
<td>• Usefulness of Interrogation and return of information from external databases, linkage of variant occurrences and familial grouping</td>
</tr>
<tr>
<td></td>
<td>• Mechanisms to track de-identified data to facilitate patient management.</td>
</tr>
<tr>
<td></td>
<td>• Mechanisms for sharing</td>
</tr>
<tr>
<td></td>
<td>• Summary reports, audits</td>
</tr>
</tbody>
</table>
6.4 Standards Document

The Standards document has undergone multiple iterations and revisions, the details of which are discussed above in section 6.3, and below in section 6.6.2.1.

The Final version which has been approved by the Board of The RCPA, and released via The RCPA website, the HGSA, and the HVP may be found at Appendix Item 1.7.

6.5 Stakeholders

Stakeholders in this process were identified and are listed below; and were considered during the consultation process.

6.5.1 Partners

- Human Variome Project, Australia Node
- Human Genetics Society of Australasia (HGSA)
- Department of Health and Aging (DoHA)

6.5.2 Other

- National Pathology Accreditation Advisory Council (NPAAC)
- National Association of Testing Authorities (NATA)
- National Health and Medical Research Council (NHMRC) – via Human Genetics Advisory Committee (HGAC)
- Genetic Laboratories and clinical services nationally: e.g.: VCGS, MCRI, PMCC
- Pathology Australia
- Industry: hardware, software, and service providers
- E-Health providers: e.g.: National e-Health Transition Authority (NeHTA)
- Health Informatics professional societies; and advisory committees
- Therapeutic Goods administration (TGA)

6.6 Communication Strategy

The Communication strategy aimed to promote awareness among clinical researchers, clinicians, pathologists, and laboratory scientists; and aimed to ensure that RCPA, HVP, and HGSA spokespersons and other key stakeholders had the appropriate resources and provision of a consistent message when discussing this project with colleagues and collaborators.

6.6.1 Public Awareness

6.6.1.1 Initial campaign

See Appendix Item 1.3 for initial public notice released September 2013.

A dedicated landing page was developed and made publicly available via the RCPA website http://www.rcpa.edu.au/Library/Practising-Pathology/DNASeqVar

Information and updates on this project were posted on the RCPA website and the landing page link. Public notices including a description of what the project was, why it was being undertaken, and any milestone and consultation information were provided to the HVP, and HGSA who included the information on their websites and notified their memberships. Additionally, the stakeholders listed above were notified of the project and its purpose.
6.6.1.2 Ongoing campaign

The project officer undertook an ongoing campaign to increase awareness and engagement amongst stakeholders involving but not limited to attendance at three conferences, and publication in a peer reviewed journal as follows:

- July 2014 – Manuscript accepted to Translational and Applied Genomics Appendix Item 1.5.1
- August 2014 – Attendance and oral presentation at Human Genetics Society of Australasia (HGSA), Appendix Item 1.5.2
- September 2014 – Attendance and oral presentation at Mutation Detection 2014, Appendix Item 1.5.3
- September 2014 – Attendance and poster presentation at British Society for Genetic Medicine (BSGM), Appendix Item 1.5.4 (please note, project officer personal trip to UK, volunteered time to attend and promote the standards under development)
- In progress March 1 2015 – Attendance and presentation of the released standards document at The RCPA’s Annual conference, Pathology Update

All these events provided opportunities to discuss with key experts and stakeholders, informing the development process and providing contacts to reach out to when seeking consultation, and promotion of alignment of standards globally.

6.6.2 Consultation

Target stakeholders were identified nationally and globally and were invited to respond with activities currently being undertaken by themselves or their colleagues.

Further to this, key stakeholders to be invited to the consultation workshop were identified and refined to ensure that those active in the area with a broad range of experience and perspectives could actively contribute to the development of the standards once an initial draft had been developed.

Following on from the Consultation Workshop (discussed below in section 6.4.2.1), the heavily revised standards document was released back to the consultation workshop attendees for review and feedback.

Following endorsement by the RCPA and HGSA, the standards document will be released for broad dissemination and consultation through the HVP International (HVPI), UNESCO (via HVPI), the Global Alliance for Genomics and Health (GA4GH), and existing global Pathology professional society networks.

6.6.2.1 Consultation Workshop

A Standards for the Accreditation of DNA Sequence Variation Databases Consultation Workshop was held on August 18 2014. The purpose of this workshop was to engage with professional stakeholders prior to consultation and communication with the broader medical and scientific community.

The agenda was designed to explore in depth key issues identified; and gather value add perspectives to intended to inform the evolution of the existing draft, and maximise a
positive response to the standards upon their release. The agenda may be found at Appendix Item 1.4

More than 30 experts in their fields attended the Consultation workshop on August 18 2014. Prior to the workshop, they were provided with the Translational & Applied Genomics peer reviewed article exploring the issues and framework around which the standards were being developed; and they were provided with the current version of the draft standards document (at the time).

The draft standards document was taken as read; and attendees were asked to come prepared to participate in active discussions, with their key areas of concern or key points prepared for debate.

At the commencement of the workshop, a general overview was presented by the project officer to set the scene, prior to the workshop attendees breaking out in to four focus groups led by steering committee members. The steering committee chair and the project officer floated amongst groups to facilitate difficult issues and keep the groups on topic and progressing throughout the day.

The smaller focus groups allowed for more active and efficient discussion intended to address any deficiencies, limitations, and or areas of concern in an open and transparent fashion with fellow workshop attendees.

Key points from each group were collated and presented to the entire workshop at key points throughout the day. These points were collated, taken under consideration, and incorporated in to a revised standards document by the project officer. This revised document was then circulated to the steering committee for comment prior to re-circulating to the workshop attendees for review and feedback. The consultation period closed on November 9, 2014. Late submissions were accepted prior to further revision by the project officer; and final review by the steering committee prior to submission to the RCPA Board for approval.

6.7 Steering Committee

Three meetings of the steering committee were held during the period; on October 30 2013, February 6 2014, and June 19 2014.

In addition to the steering committee meetings (two teleconferences, one face to face), the Chair met with the project officer by phone on several occasions out of session to discuss and agree on next steps, and agree on strategies to mitigate delays due to limited availability of steering committee members.

As the availability of steering committee members for face to face meeting caused delays to work in progress, this issue was mitigated by:

- Project Officer meeting with individual steering committee members in person and by phone as needed to ensure work continued to progress
• Project Officer scheduled and defining out of session work requirements with clearly communicated completion dates (done by email with phone follow up)
• In particular, the Steering committee had an out of session work schedule to complete prior to the face to face meeting in June, and prior to the Consultation Workshop held in August.

6.8 Department of Health Quality Pathology Section and Health Grants

The project officer managed all communications with the Quality Officer within the Quality Pathology section of the Department of Health, and with the Health Grants Division.

Department of Health representatives were invited to attend steering committee meetings, and attended the Consultation Workshop as delegates.

All key reporting requirements were undertaken between the RCPA and the Department of Health through the project officer, and all requirements and expectations were met within the prescribed timeframes per Item 5.3.1.
## FINANCE

### 7.1.1 Financial Report

### Grant: Development of Standards for Accreditation of DNA Sequence Variation Data Bases

#### Statement of Income and Expenses

to January 2015

<table>
<thead>
<tr>
<th>Description</th>
<th>GST (excl)</th>
<th>GST</th>
<th>GST (incl)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Revenue</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inv 60552 - Funding agreement executed</td>
<td>115,000</td>
<td>11,500</td>
<td>126,500</td>
</tr>
<tr>
<td>Inv 61055 - First performance report</td>
<td>72,930</td>
<td>7,293</td>
<td>80,223</td>
</tr>
<tr>
<td>Inv 62423 - Second performance report</td>
<td>72,930</td>
<td>7,293</td>
<td>80,223</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>260,860</td>
<td>26,086</td>
<td>286,946</td>
</tr>
<tr>
<td><strong>Staffing Costs</strong></td>
<td>237,937</td>
<td>23,794</td>
<td>261,731</td>
</tr>
<tr>
<td>Project officer and employment on cost</td>
<td>217,898</td>
<td>21,790</td>
<td>239,688</td>
</tr>
<tr>
<td>Pathologist</td>
<td>20,039</td>
<td>2,004</td>
<td>22,043</td>
</tr>
<tr>
<td><strong>Steering Committee Costs</strong></td>
<td>3,173</td>
<td>317</td>
<td>3,490</td>
</tr>
<tr>
<td>Steering Committee meetings (4 face to face meetings)</td>
<td>2,868</td>
<td>287</td>
<td>3,155</td>
</tr>
<tr>
<td>Communications - Video Conferencing</td>
<td>304</td>
<td>30</td>
<td>335</td>
</tr>
<tr>
<td><strong>Travel and Conferencing</strong></td>
<td>20,853</td>
<td>2,085</td>
<td>22,938</td>
</tr>
<tr>
<td>1 Day Workshop for 30 People</td>
<td>3,378</td>
<td>338</td>
<td>3,716</td>
</tr>
<tr>
<td>Travel to attend the Workshop</td>
<td>11,183</td>
<td>1,118</td>
<td>12,301</td>
</tr>
<tr>
<td>Travel &amp; Accommodation for Project Officer (Melbourne Based)</td>
<td>6,291</td>
<td>629</td>
<td>6,920</td>
</tr>
<tr>
<td><strong>Administration Costs</strong></td>
<td>26,910</td>
<td>2,691</td>
<td>29,601</td>
</tr>
<tr>
<td>Project &amp; financial management</td>
<td>19,199</td>
<td>1,920</td>
<td>21,119</td>
</tr>
<tr>
<td>Financial audit</td>
<td>5,000</td>
<td>500</td>
<td>5,500</td>
</tr>
<tr>
<td>Equipment for the Project Officer</td>
<td>2,264</td>
<td>226</td>
<td>2,490</td>
</tr>
<tr>
<td>Telephone &amp; Internet Use</td>
<td>448</td>
<td>45</td>
<td>493</td>
</tr>
<tr>
<td><strong>Total Expenses</strong></td>
<td>288,873</td>
<td>28,887</td>
<td>317,761</td>
</tr>
<tr>
<td><strong>Net Profit/(loss)</strong></td>
<td>(28,013)</td>
<td>(2,801)</td>
<td>(30,815)</td>
</tr>
</tbody>
</table>
1 APPENDICES
1.1 Steering Committee Terms of Reference

The goal of this committee is to develop national standards to accredit DNA sequence variation data bases to ensure that data in the data bases are accurate and of sufficient quality to be

- Fit for their intended use by diagnostic laboratories for efficient, informative analysis and interpretation of genetic variants; to improve clinical decision making, counselling, and treatment planning.
- Placed in the Human Variome Project Data Repository

Membership:

- Committee members will be members of RCPA, and or HGSA, and or Human Variome Project (HVP) staff members.
- The Committee will include members from a cross section of the various disciplines involved in the development and utilisation of DNA sequence variation databases.
- Committee members will be recognized as having expertise in their discipline.
- Committee member appointment will base on their capacity to contribute to the work of the committee.
- Committee members must have an active interest in the development and or utilisation of DNA sequence variation databases [in clinical environment].
- The Committee may co-opt individuals with specific expertise to assist as required

Term of membership: 12 months or completion of project
Chairmanship: Appointed by RCPA CEO
Accountability: To RCPA Board of Directors
Appointment: By call for nominations

The Committee will:

- Foster the development of high professional standards in the development and provision of DNA sequence variation databases in clinical genomics environment
- Act as an expert reference group in matters pertaining to the project,
- Provide a forum for open communication and active discussion to guide development of standards applicable/appropriate to the broad spectrum of stakeholders involved in this field,
- Discuss and advise on a framework for the standards development for the accreditation of sequence variation databases
- Review existing guidelines / standards / literature pertaining to DNA Sequence variation database development; and make recommendations,
- Develop national standards following the agreed framework,
- Meet as required by Tele-Conference or face to face; to actively participate and contribute to the development of standards.

Process for endorsement of recommendations / standards developed:

- Must be endorsed by the RCPA Board of Directors, HVP, and HGSA Council prior to submission and or public release

The operation of the group is bound by the Policies and Procedures and Constitution of RCPA
### 1.2 Global Review Logs

<table>
<thead>
<tr>
<th>Body</th>
<th>Websites / Links</th>
</tr>
</thead>
<tbody>
<tr>
<td>NGRL Manchester UK</td>
<td>DMuDB Acceptable Use Policy</td>
</tr>
<tr>
<td>NCBI NIH NHGRI ICCG</td>
<td>ClinVar</td>
</tr>
<tr>
<td>Cardiff University</td>
<td>Human Gene Mutation Database HGMD</td>
</tr>
<tr>
<td>Wellcome Trust Sanger Institute</td>
<td>COSMIC</td>
</tr>
<tr>
<td>Wellcome Trust Sanger Institute DECIPHER (GRCh37)</td>
<td>Deciphering Developmental Disorders</td>
</tr>
<tr>
<td>Wellcome Trust Sanger Institute UK 10K rare genetic Variants in health and disease (2010-2013)</td>
<td>Data in EGA</td>
</tr>
<tr>
<td>EMBL-EBI, Wellcome Trust Genome Campus</td>
<td>Ensembl</td>
</tr>
<tr>
<td>EMBL-EBI, Wellcome Trust Genome Campus European Genome-phenome archive EGA</td>
<td></td>
</tr>
<tr>
<td>NIH supported</td>
<td>ENIGMA Evidence Based Network for the Interpretation of Germline Mutant Alleles</td>
</tr>
<tr>
<td>NIH NCBI</td>
<td>OMIM</td>
</tr>
<tr>
<td>NIH</td>
<td>Genetic Association Database</td>
</tr>
<tr>
<td>CHoP</td>
<td>MITOMAP</td>
</tr>
<tr>
<td>CHoP GWAS CENTRAL</td>
<td></td>
</tr>
<tr>
<td>University of South Australia</td>
<td>MAWSON</td>
</tr>
<tr>
<td>Institute of Medical Technology Bioinformatics, University of Tampere, Finland</td>
<td>MUTbaSe</td>
</tr>
<tr>
<td>INSERM</td>
<td>Universal Mutation Database (UMD)</td>
</tr>
<tr>
<td>Leiden University Medical Centre</td>
<td>LOVD</td>
</tr>
<tr>
<td>HGVS</td>
<td>LSDB LIST</td>
</tr>
<tr>
<td>CHW IRSF</td>
<td>RettBASE</td>
</tr>
<tr>
<td>InSiGHT</td>
<td>APC LOVD</td>
</tr>
<tr>
<td>Stanford Uni, NIH/NIGMS</td>
<td>PharmGKB</td>
</tr>
<tr>
<td>EU Commission</td>
<td>Gen2Phen</td>
</tr>
<tr>
<td>Max planck Institute for Molecular Genetics</td>
<td>Genomics Diagnostics Group&gt;1200 clinical cases by NGS</td>
</tr>
<tr>
<td>MCRI</td>
<td>POSSUM</td>
</tr>
<tr>
<td>CLARITAS CERNER</td>
<td>Cerner and Claritas Genomics will collaborate to advance personalised medicine</td>
</tr>
<tr>
<td>Ingenuity Systems</td>
<td>QIAGEN offers clinical labs early access to new Ingenuity solution</td>
</tr>
<tr>
<td>BODY/OWNER</td>
<td>Links / websites</td>
</tr>
<tr>
<td>---------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Health Level Seven (HL7) International</td>
<td>HL7 standards</td>
</tr>
<tr>
<td>Association for Pathology Informatics</td>
<td></td>
</tr>
<tr>
<td>National Institute of Standards and Technology (USA)</td>
<td>Data Informatics portal</td>
</tr>
<tr>
<td>National Institute of Standards and Technology (USA)</td>
<td>NIST Health Information Technology</td>
</tr>
<tr>
<td>National Institute of Standards and Technology (USA)</td>
<td>Health IT Standards Testing Infrastructure</td>
</tr>
<tr>
<td>AU govt funded Super Science initiative - lead - University of Melbourne</td>
<td>The National eResearch Collaboration Tools and Resources (NeCTAR) Project</td>
</tr>
<tr>
<td>EMBL Australia</td>
<td>The Bioinformatics Resource Australia-EMBL</td>
</tr>
<tr>
<td>Bioplatforms Australia</td>
<td>Services can be integrated into ‘omics based research projects or accessed specifically for tailored bioinformatics solutions</td>
</tr>
<tr>
<td>Centre for Comparative Genomics (WA)</td>
<td>Bioinformatics Research Laboratory</td>
</tr>
<tr>
<td><strong>Australian Access Federation (AAF)</strong></td>
<td>The <strong>Australian Access Federation</strong> provides the means of allowing a participating institution and/or a service provider to trust the information it receives from another participating institution. This provides seamless access to resources and secure communication by removing most of the roadblocks to collaboration and sharing at both the institutional and end user levels. Funded by NCRIS and Super Science</td>
</tr>
<tr>
<td>ITFoM</td>
<td>IT Future of Medicine</td>
</tr>
<tr>
<td>HeLEX</td>
<td>Centre for Health, Law, and Emerging Technologies</td>
</tr>
<tr>
<td>Bioshare</td>
<td>BioShareIT</td>
</tr>
<tr>
<td>Bioshare</td>
<td>Omics Connect Toolbox: A Management, Validation And Visualisation Toolkit For Complex Omics Data</td>
</tr>
<tr>
<td>International Medical Informatics Association (IMIA)</td>
<td>Trustworthy Reuse of Health Data: A Transnational perspective</td>
</tr>
<tr>
<td>Health Informatics Society of Australia</td>
<td></td>
</tr>
<tr>
<td>Health Information Management Association Australia</td>
<td></td>
</tr>
<tr>
<td>EXISTING LABORATORY STANDARDS / GUIDELINES</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>--</td>
</tr>
<tr>
<td>NATA</td>
<td><a href="http://www.nata.asn.au/">http://www.nata.asn.au/</a></td>
</tr>
<tr>
<td>ISO 13485, 15189 17025 13606</td>
<td><a href="http://www.iso.org/iso/home.htm">http://www.iso.org/iso/home.htm</a></td>
</tr>
<tr>
<td>CAP Molecular Pathology Checklist</td>
<td>CAP link</td>
</tr>
<tr>
<td>CLIA 42 CFR 493</td>
<td>CLIA link</td>
</tr>
<tr>
<td>RCPA - Implementation of MPS in Diagnostic Medical Genetic Testing</td>
<td>Pathwiki RCPA link</td>
</tr>
<tr>
<td>ACMG clinical laboratory Standards for NGS</td>
<td>Genetics in medicine 2013 15 733-747</td>
</tr>
<tr>
<td>Nex-StoCT Principles and Guidelines</td>
<td>guidelines link</td>
</tr>
<tr>
<td>AMP - Opportunities and Challenges</td>
<td>JMD link</td>
</tr>
<tr>
<td>VKGL</td>
<td>Dutch Best practice guidelines for NGS in genome Dx</td>
</tr>
<tr>
<td>Clinical Molecular Genetics Society</td>
<td>Targeted NGS practice Guidelines</td>
</tr>
<tr>
<td>NGRRL UK</td>
<td>Clinical Genetics Data Set Project (case study driven)</td>
</tr>
<tr>
<td>NGRRL UK - The Challenge of getting clinical data into databases</td>
<td>EuroGentest workshop - Jan 13</td>
</tr>
<tr>
<td>HGSA (example framework)</td>
<td>Clinical Genetics Services Standards Framework</td>
</tr>
<tr>
<td>Information Standards Board for Health and Social Care UK</td>
<td>overview of STDs devpt</td>
</tr>
<tr>
<td>ASHG</td>
<td>-</td>
</tr>
<tr>
<td>ESHG</td>
<td>-</td>
</tr>
</tbody>
</table>
1.3 Public Notice
Development of Standards for Accreditation of DNA Sequence Variation Data Bases

With the support of The Department of Health, through the Quality Use of Pathology Program (QUPP), a national project for the Development of Standards for Accreditation of DNA Sequence Variation Data Bases has been jointly initiated by the Royal College of Pathologists of Australasia (RCPA), and the Human Variome Project (HVP).

Background
There is a rapidly increasing volume, spectrum, and complexity of genetic tests emerging within diagnostic pathology laboratories. In particular, high throughput sequencing methods such as targeted panel, exome (WES), and whole genome sequencing (WGS), are producing an increasing quantity of genetic data requiring analysis and interpretation, forming a substantial proportion of the workload.

Currently, there is a plethora of online mutation databases to refer to, however there is a distinct lack of such databases that meet the stringent accuracy and reproducibility that the clinical diagnostic environment demands. Additionally, The current databases are “Fractured”, with varied access and sharing of the data within; and variable quality due to errors / inaccurate data posting, all of which is a clear risk to the quality of patient care. With more widespread, secure sharing of variants and associated phenotypes, the value of cumulative variant information will accelerate the delivery of accurate, actionable, and efficient clinical reports.

There are currently no standards or equivalent mechanisms for accreditation of databases to ensure the accuracy and quality of uploaded data into any central repository to meet the needs of the clinical diagnostics environment.

About the Project
The objective is to develop national standards to accredit DNA sequence variation data bases to ensure that data in the data bases are accurate and of sufficient quality to be

- Fit for their intended use by diagnostic laboratories for efficient, informative analysis and interpretation of genetic variants; to improve clinical decision making, counselling, and treatment planning.
- Placed in the Human Variome Project Data Repository

The outcome is to

- Deliver Standards for the Accreditation of DNA Sequence Variation Databases to RCPA Board of Directors for approval
- Submission of the approved Standards to the Australian Department of Health
- Acceptance of the standards by National Pathology Accreditation Advisory Council (NPAAC) to enable accreditation via the National Association of Testing Authorities / RCPA process

This will be achieved by engaging a steering committee with representation from RCPA, HVP, and HGSA to lead the project. The Steering committee in conjunction with an expert consultation working group and other stakeholders will address key issues to aid the development of the standards, to be completed in 2014.
1.4 Consultation Group Workshop

1.4.1 Agenda & Attendees

DEVELOPMENT OF STANDARDS FOR THE ACCREDITATION OF DNA SEQUENCE VARIATION DATABASES - RCPA CONSULTATION WORKSHOP

Date: Monday August 18 2014
Time: 09:30 – 16:30 AEST
Venue: Park Royal Melbourne Airport, Bendigo / Wangaratta Room, Level 5
Chair: Bruce Bennetts

<table>
<thead>
<tr>
<th>SCHEDULE</th>
<th>ITEM</th>
<th>COMMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00 - 09:30</td>
<td>Arrival - Tea, Coffee – Level 5 Foyer</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>09:30-09:45</td>
<td>1.0 Introduction</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.1 Welcome</td>
<td>Bruce Bennetts - Expectations, Run sheet for the day</td>
</tr>
<tr>
<td></td>
<td>1.2 Overview</td>
<td>Vanessa Tyrrell - Rationale behind development of standards</td>
</tr>
<tr>
<td>09:45-10:45</td>
<td>2.0 Work Activity</td>
<td>20 minutes to discuss each section &amp; record key points</td>
</tr>
<tr>
<td>09:45 – 10:00</td>
<td>Section 1</td>
<td>Purpose</td>
</tr>
<tr>
<td>10:00 – 10:15</td>
<td>Section 2</td>
<td>Governance</td>
</tr>
<tr>
<td>10:15 – 10:30</td>
<td>Section 3</td>
<td>Establishment</td>
</tr>
<tr>
<td>10:30 – 10:45</td>
<td>All debrief</td>
<td>Key feedback on the 3 sections from each group to all</td>
</tr>
<tr>
<td>10:45 – 11:15</td>
<td>Morning Tea – Level 5 Foyer</td>
<td></td>
</tr>
<tr>
<td>11:15 – 12:45</td>
<td>2.1 Breakout Sessions</td>
<td>25 minutes to discuss each section &amp; record key points</td>
</tr>
<tr>
<td>11:15 – 11:40</td>
<td>Section 4</td>
<td>Privacy, Protection, Security</td>
</tr>
<tr>
<td>11:40 – 12:05</td>
<td>Section 5</td>
<td>Contents</td>
</tr>
<tr>
<td>11:55 – 12:20</td>
<td>Section 6</td>
<td>Functionality</td>
</tr>
<tr>
<td>12:20 – 12:45</td>
<td>All - debrief</td>
<td>Key feedback on the 3 sections from each group to all</td>
</tr>
<tr>
<td>12:45 – 13:45</td>
<td>LUNCH – Airo Restaurant – Level 4</td>
<td></td>
</tr>
<tr>
<td>13:45 – 15:15</td>
<td>2.3 Breakout Sessions</td>
<td>25 minutes to discuss each section &amp; record key points</td>
</tr>
<tr>
<td>13:45 – 14:10</td>
<td>Section 7</td>
<td>Currency of Information</td>
</tr>
<tr>
<td>14:10 – 14:35</td>
<td>Section 8</td>
<td>Access and Sharing</td>
</tr>
<tr>
<td>14:35 – 15:00</td>
<td>Section 9</td>
<td>Professional Use and Management</td>
</tr>
<tr>
<td>15:00 – 15:15</td>
<td>All - debrief</td>
<td>Key feedback on the 3 sections from each group to all</td>
</tr>
<tr>
<td>15:15 – 15:45</td>
<td>Afternoon Tea – Level 5 Foyer</td>
<td></td>
</tr>
<tr>
<td>15:45 – 16:15</td>
<td>3. Wrap Up</td>
<td></td>
</tr>
<tr>
<td>16:15 – 16:30</td>
<td>3.1 Group Summary</td>
<td>Re-Cap Key points to address, Action Items</td>
</tr>
<tr>
<td></td>
<td>3.2 Meeting wrap up</td>
<td>Next steps, close of meeting</td>
</tr>
<tr>
<td>Attendees</td>
<td>Location</td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>---------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Bruce Bennetts</td>
<td>Chair</td>
<td></td>
</tr>
<tr>
<td>Arthur Hsu</td>
<td>Steering Committee Member, HVP Rep</td>
<td></td>
</tr>
<tr>
<td>Chiyan Lau</td>
<td>Steering Committee Member, FRCPA</td>
<td></td>
</tr>
<tr>
<td>Scott Mead</td>
<td>Steering Committee Member, FRCPA</td>
<td></td>
</tr>
<tr>
<td>Cliff Meldrum</td>
<td>Steering Committee Member, HGSA Rep</td>
<td></td>
</tr>
<tr>
<td>Graeme Suthers</td>
<td>Steering Committee Member, FRCPA</td>
<td></td>
</tr>
<tr>
<td>Graham Taylor</td>
<td>Steering Committee Member, HVP Rep</td>
<td></td>
</tr>
<tr>
<td>Ness Tyrrell</td>
<td>RCPA Project Officer, Committee Member</td>
<td></td>
</tr>
<tr>
<td>Debra Graves</td>
<td>RCPA CEO, Committee Member</td>
<td></td>
</tr>
<tr>
<td>Richard Allcock</td>
<td>UWA, Perth, WA</td>
<td></td>
</tr>
<tr>
<td>Damien Bruno</td>
<td>VCGS, VIC</td>
<td></td>
</tr>
<tr>
<td>Leslie Burnett</td>
<td>NSW Health, RCPA Informatics Advisory Committee</td>
<td></td>
</tr>
<tr>
<td>Mark Cowley</td>
<td>Garvan Institute of Medical Research, NSW</td>
<td></td>
</tr>
<tr>
<td>Mark Davis</td>
<td>PathWest, WA</td>
<td></td>
</tr>
<tr>
<td>Hugh Dawkins</td>
<td>OPHG, WA</td>
<td></td>
</tr>
<tr>
<td>Marcel Dinger</td>
<td>Garvan Institute of Medical Research, NSW</td>
<td></td>
</tr>
<tr>
<td>Ken Doig</td>
<td>PeterMac, VIC</td>
<td></td>
</tr>
<tr>
<td>Andrew Griffin</td>
<td>NATA</td>
<td></td>
</tr>
<tr>
<td>James Harraway</td>
<td>Mater Pathology, SNP, QLD</td>
<td></td>
</tr>
<tr>
<td>Alice Johnstone</td>
<td>BioGrid Australia, VIC</td>
<td></td>
</tr>
<tr>
<td>Karin Kassahn</td>
<td>SA Pathology, SA</td>
<td></td>
</tr>
<tr>
<td>Michael Legg</td>
<td>Chair, RCPA Informatics Advisory Committee</td>
<td></td>
</tr>
<tr>
<td>Finlay McRae</td>
<td>InSiGHT, Royal Melbourne Hospital, VIC</td>
<td></td>
</tr>
<tr>
<td>Suzanne Petrie</td>
<td>Department of Health, Canberra</td>
<td></td>
</tr>
<tr>
<td>John-Paul Plazzer</td>
<td>InSiGHT, Royal Melbourne Hospital, VIC</td>
<td></td>
</tr>
<tr>
<td>Attendees</td>
<td>Location</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>---------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Michael Ralston</td>
<td>NPAAC expert</td>
<td></td>
</tr>
<tr>
<td>Jan Stanek</td>
<td>Mawson Project, University of SA, SA</td>
<td></td>
</tr>
<tr>
<td>Markus Stumptner</td>
<td>University of SA, SA</td>
<td></td>
</tr>
<tr>
<td>Joanne Tester</td>
<td>Department of Health, Canberra</td>
<td></td>
</tr>
<tr>
<td>Natalie Thorne</td>
<td>Melbourne Genomics Health Alliance, WEHI, VIC</td>
<td></td>
</tr>
<tr>
<td>Heather Thorne</td>
<td>kConFab, VIC</td>
<td></td>
</tr>
<tr>
<td>Robyn Ward</td>
<td>Chair, HGAC, NHMRC</td>
<td></td>
</tr>
<tr>
<td>Andreas Zankl</td>
<td>Sydney Children’s Hospital Network, NSW</td>
<td></td>
</tr>
<tr>
<td>Tim Smith</td>
<td>HVP International</td>
<td></td>
</tr>
</tbody>
</table>
1.5 Communications

1.5.1 Translational & Applied Genomics July 2014

Article history:
Received 16 April 2014
Received in revised form 30 June 2014
Accepted 4 July 2014

Applied and Translational Genomics 3 (2014) 54-57
http://dx.doi.org/10.1016/j.atg.2014.07.002

This is an open access article under the CC BY-NC-ND license
(http://creativecommons.org/licenses/by-nc-nd/3.0/).

Quality Standards for DNA Sequence Variation Databases to Improve Clinical Management Under Development in Australia


1) Western Sydney Genetics Program, The Children’s Hospital at Westmead, Westmead, NSW 2145 AUSTRALIA
2) SDS Pathology, North Ryde, NSW 2113, AUSTRALIA
3) Genomic Medicine, Department of Pathology, University of Melbourne, Melbourne, VIC 3010 AUSTRALIA
4) Royal College of Pathologists of Australasia, 207 Albion Street, Surry Hills, NSW 2010 AUSTRALIA
5) South Eastern Sydney Laboratory Services, Prince of Wales Hospital, Randwick, NSW 2031 AUSTRALIA
6) Hunter Area Pathology Service, John Hunter Hospital, New Lambton Heights, NSW 2305 AUSTRALIA
7) SA Clinical Genetics Service, SA Pathology, Adelaide, SA 5000 AUSTRALIA; and Department of Paediatrics, University of Adelaide, SA 5008.
8) Human Variome Project International, Level 5, 234 Queensberry Street, University of Melbourne, VIC 3010 AUSTRALIA
Quality Standards for Databases of DNA Sequence Variants

Vanessa Tyrrell

Overview

- Background
  - Review
- Rationale for development of standards
- Standards Framework
  - Examples
- Implementation
  - Challenges, broad adoption
- Summary

Background

- Current databases variable quality
  - Variant call errors
  - Non standard nomenclature
  - Incomplete / inaccurate pathogenicity association
  - Limited linked genotype:phenotype data
- RCPA, HGSA, HVP initiative
  - Develop standards for clinical environment to address the limitations and risk to patient care
Global Review
Initiatives

- Global Alliance for Genomics and Health (GA4GH)
  - Framework for responsible sharing of genomic and health related data
- NCDS –
  - Data to Discovery: Genomes to Health
- BSGM –
  - 100,000 Genome Group, DMuDB, DECIPHER
- ICCG –
  - ClinGen - ClinVar

Global Review
Best Practice Guidelines

- Address clinical workflows using next generation sequencing techniques
  - ACMG
  - BSGM (CMGS, ACGS)
  - VKGL
  - NIH Genomic Data Sharing Policy
  - NHGRI

Global Review
Best Practice Guidelines

- Address clinical workflows using next generation sequencing techniques
  - ACMG
  - BSGM (CMGS, ACGS)
  - VKGL
  - NIH Genomic Data Sharing Policy
  - NHGRI
Rationale for development

- No specific guidelines to guide and ensure quality, ongoing maintenance
- Provide oversight for databases intended for clinical diagnostic service delivery
- Complement existing standards
- Align with global initiatives
- Guide to
  - identify a quality database
  - establishment of new databases

Standards Framework

1. Purpose
   - Functional elements, nature of use of database
2. Governance
   - Sustainability
   - Compliance
3. Establishment of DB
   - Principle requirements for functional DB
4. Privacy Protection, Security
   - Compliance
   - Information policy
   - Access of security
5. Content
   - Minimum requirements for data submitted
   - Analytical, clinical validity
   - Laboratory accreditation
6. Functionality
   - Usefulness
   - Ease of access
7. Currency of Information
   - Question
   - Monitoring, relevance of data
8. Access and sharing
   - Policy
   - Mechanisms
   - User registration
   - Auditing of access and sharing
9. Professional Use and Management
   - Terminology, ontology
   - Variant classification
   - Data re-analysis
   - Skills, experience

Standards Layout

- Standard
  - Minimum requirement
  - “must”
  - $1.1
- Commentary
  - Provide clarity and guidance on interpretation of standard
  - Normative: “must”
  - Informative: explanatory discussion, and or examples.
  - C1.1(1)
Standards Layout

• Standard
  – Minimum requirement
  – “must”
  – S1.1
• Commentary
  – Provide clarity and guidance on interpretation of standard
  – Normative: “must”
  – Informative: explanatory discussion, and or examples.
  – C1.1(i)

2. Governance

S2.2 A database custodian [ownership] must be identified, who is accountable and responsible for the operation of the database, and who is accountable to a clinical institution and/or regulatory body.

C2.2(i) The custodian is the entity with responsibility and accountability for the operation of the database. This may be a position, committee, or board of the host institution (rather than a specifically named individual) which can ensure transparency and business continuity.

C2.2(ii) The custodian should have sufficient authority to carry out or delegate actions as required.

C2.2(iii) The custodian must ensure adequate curation and management of the data such that it is secure, accurate and subject to regular review.

6. Functionality

S6.1 The database must have extensive and flexible search capabilities.

C6.1(i) A database must be able to search by a specific variant, and gene
C6.1(ii) A database should have orthogonal search capabilities to increase filtering capabilities.
C6.1(iii) Searchable fields should include: gene, protein, disease, phenotype, and quality metrics (when available).
Implementation

• Broad consultation
• Adjunct to existing NPAAC standards
  — Tool for laboratories to facilitate accreditation
  — Tool for local labs to judge integrity of a database outside of their jurisdiction
• Overseas / Global implementation
  — Applicable more broadly
  — HVP
  — GA4GH

Conclusion

• National standards
  — Workable in changing environment
• Fundamental principle
  — Provide minimum requirements for databases intended to be utilised in clinical environment
  — Provide tool to judge quality of third party databases
• Framework provide structure & flow to standards
• NPAAC style format – tool for accreditation
• Seeking broad acceptance and adoption
• Applied & Translational Genomics July 2014

Thank you

Acknowledgements

Steering Committee
Bruce Bennetto (Chair)
Melody Coombes
Arthur Hsu
Chien Cau
Scott Mead
Cliff Meldrum
Graeme Suthers
Graham Taylor

Additional Support
Leslie Burnett
Michael Legg
Michael Raison

Funding Support
Department of Health
Quality Use of Pathology Program (QUPP)
1.5.3 Mutation Detection September 2014

Quality Standards for Clinical Databases of DNA Sequence Variants

Vanessa Tyrrell

Overview

• Background
  – Review
• Rationale for development of standards
• The ideal clinical database
• Standards framework
  – Criteria, examples
• Implementation
  – Challenges, adoption
• Summary

Background

• Current databases variable quality
  – Variant call errors
  – Non standard nomenclature
  – Incomplete / inaccurate pathogenicity association
  – Limited linked genotype:phenotype data
• RCPA, HGSA, HVP initiative
  – Develop standards for clinical environment to address the limitations and risk to patient care
Global Review

Initiatives

- **Global Alliance for Genomics and Health (GA4GH)**
  - Framework for responsible sharing of genomic and health related data
  - API’s
- **NCDS**
  - Data to Discovery: Genomes to Health
- **BSGM**
  - 100,000 Genome Group, DMuDB, DECIPHER
- **ICCG**
  - ClinGen - ClinVar

Global Review

- **Best practice guidelines**
  - Address clinical workflows using next generation sequencing techniques
    - ACMG, BSGM (CMGS, ACGS), VRGL, NIH Genomic Data Sharing Policy, NHGRI
  - Recommendations LSDB’s, curation
    - Multiple papers guiding towards a universal standard (1999-2012)
      - Scriver, Claustres, Vihinen, Cotton, Celli, Mitropoulou, den Dunnen, et.al.

Rationale for development of clinical standards

- Provide oversight for databases intended for clinical diagnostic service delivery
- Complement existing standards
- Align with global initiatives
- Guide to
  - identify a quality database
  - establishment of new databases
  - Set minimum requirements
The ideal clinical database

- Governance
  - Ethics and compliance
  - Custodianship and intellectual property
- Secure, protection of privacy, with controlled access and sharing
  - Aggregated data at a minimum
  - Clinical/HR-4 Database to aid patient management and outcomes
- Permanent
  - Records identified for sustainability [ongoing funding]
- Standardised for facilitating sharing, federation, consistency in reporting
  - Sensitive database
  - Data submission and data fields within database
- Nomenclature, terminology, ontology
- Genotype/Phenotype association information
- Curation meeting predetermined guidelines
  - Evaluation of new submitters/registration applications
  - Regular audits of data submission, database quality, security
  - Regular review and updating of contents
- Standards defined
- Compliance with standards via recognised accreditation process

APPENDIX: The Role of the CURATOR

Database features & content

- Database Features
  - General presentation,
  - Locus specific information,
  - Database structure
- Database Content
  - Data collection,
  - Summary table of variants,
  - Database querying
Database Content

Database Querying

56.1 The database must have flexible search capabilities

C6.1(i) Search capabilities should be customizable to allow for multiple types of queries including orthogonal queries to increase filtering capabilities.

C6.1(ii) Examples of searchable fields include specific variant, gene, alias [gene, disease], disease, phenotype, protein.

Database Structure

Summary table / downloadable

56.5 The database must be able to generate summary reports for viewers

C6.5(i) Summary data of information held in the database should be provided to the users of the database. This may include (but not limited to):

- Genes described, samples entered (single entities), coding mutations, papers cited, unique variants, fusion genes, genomic rearrangements, whole genomes, whole exomes, copy number anomalies

C6.5(ii) Summary reports may be customised according to context of the reports and needs of the end users

Implementation

- Consultation
- Adjunct to existing NPAAC standards
  - Tool for laboratories to facilitate accreditation
  - Tool for local labs to judge integrity of a database outside of their jurisdiction
- Overseas / Global implementation
  - Applicable more broadly
  - HVP
  - GA4GH
Conclusion

- **National standards**
  - Workable in changing environment
- **Fundamental principle**
  - Provide minimum requirements for databases intended to be utilised in clinical environment
  - Provide tool to judge quality of third party databases
- **Revised framework, and NPAAC format**
  - Improved structure & flow to standards
  - Tool for accreditation

Thank you

Acknowledgements

**Steering Committee**
Bruce Bennetts (Chair)
Melody Ceramins
Arthur Hsu
Chi yen Lee
Scott Mead
Cliff Meldrum
Greame Suthers
Graham Taylor

**Consultation Workshop Participants**

**Funding Support**
Department of Health
Quality Use of Pathology Program (QUPP)
1.5.4 BSGM September 2014

QUALITY STANDARDS FOR DATABASES OF DNA SEQUENCE VARIANTS

Bruce Bennetts, Melody Caraminos, Arthur Hsu, Chiyan Lau, R. Scott Mead, Cliff Meldrum, Graeme Suthers, Graham Taylor, Vanessa Tyrrell

1) Western Sydney Genetics Program, The Children’s Hospital at Westmead, New South Wales, Australia
2) RCPA Pathology, North Ryde, New South Wales, Australia
3) Department of Pathology, University of Melbourne, Melbourne, Victoria, Australia
4) Royal College of Pathologists of Australasia, Sunny Hills, New South Wales, Australia www.rcpa.edu.au
5) South Eastern Sydney Laboratory Services, Prince of Wales Hospital, Randwick, New South Wales, Australia
6) Hunter Area Pathology Service, John Hunter Hospital, New Lambton Heights, New South Wales, Australia
7) SA Clinical Genetic Service, SA Pathology, Adelaide, South Australia

1. Rationale

It is routine practice to compare sequence variations identified during clinical genetic testing with variants recorded in a wide range of genetic variation databases to aid in understanding the potential clinical significance and determining a definitive diagnosis.

Although numerous databases exist, few meet the accuracy and reproducibility requirements for clinical diagnostics. Current databases are of variable quality, contain errors in variant calls, use non-standardised nomenclature, and contain limited phenotypic information linked to genomic data. These represent limitations and risks to the quality of patient care.

2. The Ideal Clinical Database

Governance
- Ethics and compliance
- Custodianship
- Intellectual Property
- Secure, protection of privacy, with controlled access and sharing
- Aggregated data accessible
- Clin / EHR Database to aid patient management
- Permanent Records
- Site identified for sustainability of records with secure, ongoing funding

Content standardised in compatible databases to facilitate sharing, federating, consistency in reporting
- Standardised data submission, data fields within databases
- Use of standardised nomenclature, terminology, ontology
- Genotype Phenotype association information
- Curation meeting predetermined guidelines
- Evaluation of new submitters, registration applicants
- Regular audits of data submission
- Database quality, security
- Regular reviewer and updating of contents
- Standards defined, with global reach
- Compliance with standards via recognised accreditation process

3. Standards for Clinical Databases of Genetic Variants Framework

Addresses databases features and content such as those highlighted by Mitropoulou et al. in a systematic order, with clearly defined criteria.

1) Purpose
- Essential elements of the nature of use
- Governance
- Custodianship, maintenance of relevance, ethics, compliance, intellectual property, sustainability
- Establishment of databases
- Requirements for functional databases, quality, back up of databases
- Privacy, Protection, Security
- Information, access, and sharing policies
- Content
- Minimum data requirements, analytical & clinical validity, variant classification, nomenclature
- Functionality
- Search capabilities, summary reports, mechanisms of sharing, audits

APPENDIX: Detailed description of the role of the Curator

4. Standards Format

The standards are formatted per National Pathology Accreditation Advisory Council (NPAC) reference materials.

Standard (6.X): Minimum requirement, mandatory

Compliant (6.X.1): To ensure that all records are effectively permanent. (Mandatory)

Content (6.X.2): Complete audit trail should be visible to reviewers. (Mandatory)

Database (6.1): The database must have flexible search capabilities.

Search capabilities should be customizable to allow for multiple types of queries including orthogonal queries to increase filtering capabilities (Informativ)

Examples of searchable fields include specific variant, gene, allele, disease, disease phenotypes, protein (Informative)

5. Implementation & Conclusion

Broads collabtation is being undertaken to ensure standards are workable within the current evolving environment. A recent consultation workshop with a broad cross section of key stakeholders has facilitated the maturation of the draft standards document.

Reporting standards are intended to be an adjacent to existing NPAC standards to be used by laboratories to facilitate accreditation through an approved regulatory body, and judge integrity of databases housed overseas.

Given the reach of databases, and growing demand to meet clinical needs globally, these standards are likely to be applicable in other countries.

Reflecting standards will be promoted globally through the Human Variome Project International (HVP), and the Global Alliance for Genomics and Health (GA4GH) partnerships.

For more information, and to follow the progress of the Standards Project, go to http://www.rcpa.edu.au/library/DiseasePathology/ DNASeq/s or email vanessa@rcpa.edu.au

Development of Standards for the Accreditation of DNA Sequence Variation Database − 53 | Page

January 2015 FINAL REPORT
## 1.6 Steering Committee Details

### 1.6.1 Project Officer

<table>
<thead>
<tr>
<th>PROJECT OFFICER</th>
<th>BACKGROUND / EXPERTISE</th>
<th>RESPONSIBILITIES</th>
</tr>
</thead>
</table>
| Vanessa (Ness) Tyrrell | A senior health professional with over 23 years broad-based experience in the health sector, providing commercial operations, business development, training and development expertise. Highly skilled in technical, public and private genetic pathology health services delivery, clinical liaison, medical device, biotechnology, and genomics. Extensive board and committee experience.  
- Fellow HGSA Genetics (cytogenetics), BAppSc Biomed, MBA, ARCPA  
- Over sixteen (16) years working in and managing diagnostic genetic pathology laboratories. Experienced in operations, business management, clinical liaison, QAP, NATA, and ISO accreditation of clinical laboratories.  
- Over four (4) years experience managing Molecular DNA laboratory including diagnostic and relationship molecular testing services.  
- Experience with TGA IVD applications  
- Extensive Public speaking and education – corporate board level, seminars, and invited speaker at conferences (clinical, scientific, regional)  
- Clinical validation project development and supervision  
- Relationship management and comprehensive internal and external stakeholder engagement across global market  
- Development and execution of strategic and operational plans and projects  
- Current Member NHMRC HGAC 2013-2015 Triennium |  
- Develop an agreed Framework for the Standard for the Accreditation DNA Sequence Variations Databases consistent with the NPAAC Standards approach.  
- Review Accreditation Standards for DNA Sequence Variation Data Bases (if any) and for databases storing health data at the international and national level.  
- Consult widely with experts and stakeholders in DNA sequence databases.  
- Undertake a gap analysis of what occurs in Australia and what is occurring internationally.  
- Develop Standards for Accreditation of DNA Sequence Variation Data Bases.  
- Ensure the Standards are sent out for broad consultation.  
- Assist the Steering Committee to review and evaluate and make changes to the Standard as required.  
- Ensure the Standards are recommended to RCPA Council for Approval as Policy document.  
- Maintain records of project-related expenditure  
- Provide Progress and final Reports for the Project for the Department of Health |
### 1.6.2 Steering Committee Members

<table>
<thead>
<tr>
<th>NAME</th>
<th>BACKGROUND / EXPERTISE</th>
<th>RESPONSIBILITIES</th>
</tr>
</thead>
</table>
| Chiyan Lau     | • Recent fellow of the RCPA (FRCPA), PhD Molecular Bioscience, MBBS  
• Special interest in molecular genetics and cytogenetics; and the adaptation and development of databases in healthcare and pathology  
• Previously worked as Genetic Pathology Registrar  
• Currently Medical Advisor with the TGA;  
• Member of the RCPA’s Informatics Advisory committee and Genetics Advisory Committee.                                                                 | • Project Pathologist |
| Bruce Bennettts| • Molecular Geneticist  
• Head, Molecular Genetics, Western Sydney Genetics Program Experienced NPAAC Committee Member                                                                                                                                  | • Chair |
| Melody Caramins| • FRCPA (genetics), Foundation FFSc RCPA, PhD Genetics, and BMed  
• National Head of Genetics for the Primary Healthcare/SDS Pathology group  
• Conjoint senior lecturer at the University of New South Wales.  
• Former acting Clinical Director of Genetics Laboratory at South Eastern Area Laboratory Services (SEALS).  
• Prior to specialising in genetics spent 12 years as a clinician with a focus in surgical and acute care.  
• Chair, RCPA Genetics Advisory Committee  
• Focus on implementation and clinical application of novel technologies in the diagnostic genetic testing.  
• Professional activities have included involvement in drafting national best practice guidelines for diagnostic laboratories implementing next-generation sequencing and chromosomal microarray testing.                                                                 | • Member,  
• RCPA |
| Arthur Hsu     | • PhD Engineering  
• Senior Research Fellow Genomic Medicine University Of Melbourne Department of Pathology  
• HVP Australia Node  
• Bioinformatics & Database expertise (>10yrs), more recently in next generation sequencing data analysis of cancer genome and transcriptome.                                                                 | • Member |
| Scott Mead     | • Recent fellow of the RCPA (FRCPA), PhD Molecular Medicine  
• Special interest in cancer genetics and integration of translational clinical research in to mainstream healthcare  
• Genetic Pathologist at the Garvan Institute of Medical Research Kinghorn Translational Cancer Centre, and SydPath;  
• Investigating clinical application of personal cancer genome sequencing, and targeted oncology testing                                                                 | • Member,  
• RCPA |
<table>
<thead>
<tr>
<th>NAME</th>
<th>BACKGROUND / EXPERTISE</th>
<th>RESPONSIBILITIES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graeme Suthers</td>
<td>• FRCPA Genetics, PhD Molecular Genetics, FRACP, MBBS, FAICD</td>
<td>• Member,</td>
</tr>
<tr>
<td></td>
<td>• Broad experience in healthcare delivery and clinical research, genetic disorders and genetic testing.</td>
<td>• RCPA</td>
</tr>
<tr>
<td></td>
<td>• Development of software solutions for specialty genetic services such as Kintrak,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Economic evaluation of healthcare interventions.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Head South Australian Clinical Genetics Service</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Past Chair RCPA Genetics Advisory Committee</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Genetics Advisor to MSAC</td>
<td></td>
</tr>
<tr>
<td>Graham Taylor</td>
<td>• Herman Professor of Genomic Medicine in the department of Pathology at the University of Melbourne, and Director of the Australian Node of the Human Variome Project.</td>
<td>• Member,</td>
</tr>
<tr>
<td></td>
<td>• Highly regarded in his role of implementing genomic technology in a healthcare research and service setting in UK; First laboratory in Europe to introduce Massively Parallel Sequencing (MPS/NGS) in a diagnostic setting and has reported the majority of the diagnostic assays performed thus far in the UK.</td>
<td>• HVP Representative</td>
</tr>
<tr>
<td></td>
<td>• Former Head of Cancer-Research UK’s Genomic Services and established the Mutation Detection Facility at the UK’s Imperial Cancer Research Fund.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Past Chair of UKNEQAS Special Advisory Group for Next Generation Sequencing QA</td>
<td></td>
</tr>
<tr>
<td>Cliff Meldrum</td>
<td>• Deputy Director Molecular Medicine Hunter Area Pathology Service, HGSAs representative</td>
<td>• Member,</td>
</tr>
<tr>
<td></td>
<td>• Fellow Faculty of Science RCPA</td>
<td>• HGSA representative</td>
</tr>
<tr>
<td></td>
<td>• Past Chair Molecular Genetics Society of Australasia, a special interest group of the HGSA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Expertise in utilising NGS in clinical service delivery.</td>
<td></td>
</tr>
<tr>
<td>Debra Graves</td>
<td>• RCPA CEO (appointed 1999)</td>
<td>• Member,</td>
</tr>
<tr>
<td></td>
<td>• MBBS, MHA, FRACMA, AASHSM, FAICD</td>
<td>• Oversee Project Officer</td>
</tr>
<tr>
<td>Suzanne Petrie</td>
<td>• Department of Health. Acting Director, Pathology Quality Section</td>
<td>• Member,</td>
</tr>
<tr>
<td></td>
<td>• Expertise in health administration and service delivery</td>
<td>• Department of Health Representative</td>
</tr>
</tbody>
</table>
1.7 Standards Document – V1.0 Released December 2014

THE ROYAL COLLEGE OF PATHOLOGISTS OF AUSTRALASIA (RCPA)

Standards for clinical databases of genetic variants.

1st Edition 2014 (version 1.0)
Disclaimer

The Royal College of Pathologists of Australasia ("the College") has developed these Standards for clinical databases of genetic variants to assist in ensuring the quality, accuracy, security, and utility of DNA variant databases used for reporting for clinical purposes.

While there are indicators of ‘minimum requirements’ (Standards) and ‘recommendations’ (Commentary), the Standards are a first edition and have not been through a full cycle of use, review and refinement.

Therefore, in this edition, the inclusion of “standards” is provided as an indication of the opinion of the expert authoring group, but should not be regarded as definitive or as widely accepted peer professional opinion. Specifically, these terms do not carry regulatory weight with regard to laboratory accreditation. The use of these standards and guidelines is subject to the health professional’s judgement in each individual case.

The College makes all reasonable efforts to ensure the quality and accuracy of the Standards and to update the Standards regularly. However subject to any warranties, terms or conditions which may be implied by law and which cannot be excluded, the Standards are provided on an “as is” basis. The College does not warrant or represent that the Standards are complete, accurate, error-free, or up to date. The Standards do not constitute medical or professional advice. Users should obtain appropriate medical or professional advice, or, where appropriately qualified, exercise their own professional judgement relevant to their own particular circumstances. Users are responsible for evaluating the suitability, accuracy, currency, completeness and fitness for purpose of the Standards.

Except as set out in this paragraph, the College excludes: (i) all warranties, terms and conditions relating in any way to; and (ii) all liability (including for negligence) in respect of any loss or damage (including direct, special, indirect or consequential loss or damage, loss of revenue, loss of expectation, unavailability of systems, loss of data, personal injury or property damage) arising in any way from or in connection with the Standards and Guidelines or any use thereof. Where any statute implies any term, condition or warranty in connection with the provision or use of the Standards and Guidelines, and that statute prohibits the exclusion of that term, condition or warranty, then such term, condition or warranty is not excluded. To the extent permitted by law, the College's liability under or for breach of any such term, condition or warranty is limited to the resupply or replacement of services or goods.
## REVISION HISTORY

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Document</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>January 2014</td>
<td>Draft Framework</td>
<td>Vanessa Tyrrell</td>
</tr>
<tr>
<td>0.1</td>
<td>February 2014</td>
<td>Draft Framework</td>
<td>Vanessa Tyrrell</td>
</tr>
<tr>
<td>0.2</td>
<td>March 2014</td>
<td>Draft Standards</td>
<td>Vanessa Tyrrell</td>
</tr>
<tr>
<td>0.3</td>
<td>April 09 2014</td>
<td>Draft Standards</td>
<td>Vanessa Tyrrell</td>
</tr>
<tr>
<td>0.4</td>
<td>May 05 2014</td>
<td>Draft Standards</td>
<td>Vanessa Tyrrell</td>
</tr>
<tr>
<td>0.5</td>
<td>May 08 2014</td>
<td>Draft Standards</td>
<td>Vanessa Tyrrell</td>
</tr>
<tr>
<td>0.6</td>
<td>May 29 2014</td>
<td>Draft Standards</td>
<td>Vanessa Tyrrell</td>
</tr>
<tr>
<td>0.7</td>
<td>June 12 2014</td>
<td>Draft Standards</td>
<td>Vanessa Tyrrell</td>
</tr>
<tr>
<td>0.8</td>
<td>June 16 2014</td>
<td>Draft Standards</td>
<td>Vanessa Tyrrell</td>
</tr>
<tr>
<td>0.9</td>
<td>June 30 2014</td>
<td>Draft Standards</td>
<td>Vanessa Tyrrell</td>
</tr>
<tr>
<td>0.10</td>
<td>July 09 2014</td>
<td>Draft Standards</td>
<td>Vanessa Tyrrell</td>
</tr>
<tr>
<td>0.11</td>
<td>July 31 2014</td>
<td>Draft Standards</td>
<td>Vanessa Tyrrell</td>
</tr>
<tr>
<td>0.12</td>
<td>August 11 2014</td>
<td>Draft Standards</td>
<td>Vanessa Tyrrell</td>
</tr>
<tr>
<td>0.13</td>
<td>August 18 2014</td>
<td>Draft Standards</td>
<td>Vanessa Tyrrell</td>
</tr>
<tr>
<td>0.14</td>
<td>September 20 2014</td>
<td>Draft Standards</td>
<td>Vanessa Tyrrell</td>
</tr>
<tr>
<td>0.15</td>
<td>October 2014</td>
<td>Draft Standards</td>
<td>Vanessa Tyrrell</td>
</tr>
<tr>
<td>0.16</td>
<td>October 23 2014</td>
<td>Draft Standards</td>
<td>Vanessa Tyrrell</td>
</tr>
<tr>
<td>0.17</td>
<td>November 28 2014</td>
<td>Draft Standards</td>
<td>Vanessa Tyrrell</td>
</tr>
<tr>
<td>0.18</td>
<td>December 09 2014</td>
<td>Final Draft</td>
<td>Vanessa Tyrrell</td>
</tr>
<tr>
<td>1.0</td>
<td>December 18 2014</td>
<td>First edition for release</td>
<td>Vanessa Tyrrell</td>
</tr>
</tbody>
</table>
Contents

Definitions  62
Scope    65
Introduction 66
Background 67
1       Purpose   68
2       Governance 70
3   Establishment of Databases    73
4   Privacy, Confidentiality, Ethics, and Data Security 75
5       Content   82
6       Functionality   87
Appendix    90
1       The Curator 90
## Definitions

| Custodian | 
|---|---|
| **Normative**
(as applied to commentaries and appendices) | Prescriptive or mandatory and the material carries the same weight as the Standards to which it is attached. |
| **Informative**
(as applied to commentaries and appendices) | The material is presented to assist in the application or interpretation of the Standards to which it is attached. |

### Sequence variant
The entry (which would normally be the logical record) within a DNA variant database.

### Cloud

### Clinical

| **Clinical Grade Sequencing Data** | DNA sequencing performed in an accredited facility to defined quality standards, undertaken to inform a particular clinical indication |
| **Curation** | The activity of managing and promoting the use of data from its point of creation, to ensure it is fit for contemporary purpose, and available for discovery and re-use. |

### Database
A computer structure that houses a collection of related data.

### Database Management System (DBMS)
Determines the data model, storage, maintenance and retrieval of data, security and other functions necessary to use the database.

### DNA variant database
The structured collation of records of variations in DNA sequence or structure identified in patients or subjects versus a specified reference sequence.

### Genetic Variant
An alteration in the DNA sequence compared to a reference sequence, the significance of which is often unclear.

### Healthcare database
A specific class of database, the primary use case of which is to store data for use in healthcare environments for the clinical management of patients. Such databases typically hold personally identifiable and protected health information. Examples include health information management systems and Electronic Health Records (EHR).

### Identity
The whole of the characteristics of a document or a record that uniquely identify it and distinguish it from any other document or

---

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrity</td>
<td>The quality of being complete and unaltered in all essential respects. With identity, a component of authenticity.</td>
</tr>
<tr>
<td>Knowledge database</td>
<td>A collection of information about the data stored in the database; expressing what we know about a particular piece of data (e.g.: variant is the data and the information is what we know about pathogenicity, inheritance patterns, population distribution, etc.)</td>
</tr>
<tr>
<td>Deidentified data</td>
<td>Data reasonably disconnected from the identity of a person according to the requirements of the Privacy Amendment Act 2012 or equivalent</td>
</tr>
<tr>
<td>NPAAC</td>
<td>National Pathology Accreditation Advisory Council (give ref or URL)</td>
</tr>
<tr>
<td>Orthogonal search</td>
<td>The combination of two or more searches whereby the method of search was independent of the previous method/s of search</td>
</tr>
<tr>
<td>Ontology</td>
<td>(using a shared vocabulary to denote the types, properties, and interrelationships of concepts within a domain)</td>
</tr>
<tr>
<td>Patient identifier</td>
<td>A string (alphanumeric, numeric or alphabetic) that identifies a patient unambiguously to the data submitter but maintains anonymity of the Patient to all other users of the database</td>
</tr>
<tr>
<td>Personal Information</td>
<td>Definition of “personal information” in the Australian context can be found in the Privacy Amendment Act 2012, and includes all information or opinion about an individual who is identified or reasonably identifiable, whether such information/opinion is true or not.</td>
</tr>
<tr>
<td>Preservation</td>
<td>An activity within archiving in which specific items of data are maintained over time so that they can still be accessed and understood through changes in technology.</td>
</tr>
</tbody>
</table>

---

2 InterPARES 2 Terminology Database.  
http://www.interpares.org/ip2/ip2_terminology_db.cfm

3 InterPARES 2 Terminology Database.  
http://www.interpares.org/ip2/ip2_terminology_db.cfm

This page is intentionally blank
Scope

Purpose – This document presents a broad set of standards for sequence variant databases used for clinical purposes. It complements existing NPAAC reference materials for laboratory accreditation, and

- Is applicable to all databases of genetic variants used for clinical purposes
- Provides a benchmark for the structure of such databases, including standards for ontologies and minimum content requirements.
- Provides standards for the tools which manipulate the data in such databases.
- Assists laboratory professionals in identifying databases of appropriate quality for clinical purposes.
- Set standards for data sharing for clinical purposes: including bi-directional data transfer, interfacing, and other collaborative methods within the boundaries of existing privacy laws

Benefits – Patient care will be improved by

- encouraging the collation of curated information about DNA variants identified in patient care
- facilitating the accurate interpretation of analytical results
- enabling the sharing of curated data between laboratories, thereby developing a broader repository of data to inform clinical interpretation
- improving the efficiency of interpretation and timely reporting to clinicians.

Exclusions – This document does not detail requirements regarding:

- The phenotypic information that is to be stored. Clinical laboratories generally lack control over the quality and volume of clinical information about a patient provided to them. This document notes the phenotype data fields which should be entered if it is available, and recommends the implementation of standard terms to describe phenotypes i.e. a defined ontology.

- The ownership or physical location of databases.

- The implementation of these standards.
**Introduction**

This document has been developed by the Royal College of Pathologists of Australasia (RCPA) in collaboration with the Human Variome Project (HVP), and the Human Genetics Society of Australasia (HGSA). It presents a set of standards to be used in conjunction with other reference materials (listed below) to promote the quality, accuracy, security, and utility of DNA variant databases used for clinical purposes.

This document is not designed for databases which
- are used for other purposes such as research or public health repositories, or
- electronic health records which include open identification of patients and act as record of their management.

The fundamental principle underpinning this document is that DNA sequence variant databases intended for use in clinical diagnostic testing should be developed, curated, and maintained as safe, secure, and accurate repositories of genomic data.

These Standards have been developed with reference to current and proposed Australian regulations and standards from the International Organisation for Standardisation (ISO), including *AS ISO 15189 Medical laboratories – Requirements for quality and competence.* The standards should be used for guidance where accessing and or assessing databases outside the Australian accreditation framework.

In addition to these standards, existing NPAAC Requirements apply to all Laboratories seeking accreditation for medical testing in Australia, and must be applied in conjunction with jurisdictional and other regulatory requirements.

In each section of the document, points deemed important for practice are identified as ‘Standards’ or ‘Commentaries’.

- A Standard is the minimum requirement for a procedure, method, staffing resource or facility that is required before a Laboratory or other accreditable entity can attain accreditation – Standards are printed in bold type and prefaced with an ‘S’ (e.g. S2.2). The word ‘must’ in each Standard within this document indicates a mandatory requirement for practice. Commentary is provided to give clarification to the Standards as well as to provide examples and guidance on interpretation.

- Commentaries are prefaced with a ‘C’ (e.g. C1.2) and are placed where they add the most value. Commentaries may be normative or informative depending on both the content and the context of whether they are associated with a Standard or not. Note that when comments are expanding on a Standard or referring to other legislation, they assume the same status and importance as the Standards to which they are attached. As a general rule, where a Commentary contains the word ‘must’ then that Commentary is considered to be normative.

Please note that any Appendices attached to this document may be either normative or informative in nature and should be considered to be an integral part of this document.
Background

It has become routine practice to compare a DNA variant identified during clinical testing with the description and interpretation of variants recorded in databases, and using this information to guide clinical interpretation of the patient’s variant. Although numerous DNA variant databases already exist, there are few that meet the accuracy and reliability required for clinical diagnostics. Current databases are of variable quality and may contain errors in variant calls, non-standardised nomenclature, incomplete pathogenicity associations and limited phenotypic information linked to genomic data. These all represent limitations and risks to the quality of pathology reporting and to patient care.

The increasing use of genomic technologies such as massively parallel sequencing is producing increasing volumes of data that need to be recorded, interpreted, and shared. This provides an additional risk of propagating errors, so that an incorrect or incomplete database entry is used to interpret other database entries or reports, which are in turn compromised.

With the growing interdependence of databases for clinical reporting, the integrity of these databases becomes a critical issue.

The Standards development project

There are numerous initiatives directed at the integration of genomic technologies into mainstream clinical diagnostics, however there are no specific standards or equivalent mechanisms to assure the quality or guide the accreditation of DNA variant databases. An Australian project led by the Royal College of Pathologists of Australasia (RCPA) in collaboration with the Human Genetics Society of Australasia (HGSA), and the Human Variome Project (HVP), developed these standards for DNA variant databases intended for clinical use. This project was supported by an unrestricted grant from the Australian Department of Health’s Quality Use of Pathology Program (QUPP).

The standards are a broad reaching set of standards that are sympathetic to the rapidly changing landscape of clinical genomics, and that can be applied to assess extant databases and to guide the development of new databases. The fundamental goal of the document is to provide a quality framework for the oversight of DNA variant databases. These standards complement existing laboratory standards and accreditation requirements, act as a guide to identify a quality database, assist the development of new databases, and in improving existing databases that have been developed in non-clinical environments.

Maintaining the quality, accuracy, and clinical relevance of DNA variant databases will reduce the risk of misinterpretation and inappropriate reporting of variants, promote the sharing of data which can be trusted for clinical use, and accelerate the delivery of actionable clinical reports to improve patient care.

---

1 Purpose

These standards are intended to be a high level, generic set of standards which are applicable to sequence variant databases serving clinical purposes. In order to establish the context in which a sequence variant database may be utilised, the purpose must be clearly articulated. It is important for database users to know what is available within the database, and what to expect from the database.

S1.1 The intended purpose of the database must be clearly defined and documented, and be made available through appropriate media such as internally controlled documentation and, if publicly accessible, on the database website.

C1.1 The elements which characterise a DNA variant database must include:

- The context in which the database is intended to be used e.g. clinical diagnostic, clinical research, or clinical theranostic purposes, and whether access will be restricted to particular users (in-house, password protected) or in the public domain.

- The nature of the information included in the database. This description may include:
  - disease specific information,
  - gene specific information,
  - phenotype description),
  - whether the database contains deidentified data only (knowledge database), personal data (Healthcare database), or both related to submitted variants.
  - germline or somatic data, or both

- There must be a clear distinction between a database for use by in-house staff only i.e. all users are accountable to the custodian, versus a public database that may be used by people who are not accountable to the custodian (See S2.2). The requirements for databases in these two settings differ. If a database is transitioned from being in-house to being public, there must be a review of all aspects of the database operation to ensure that the different requirements are met.

- The basic limitations of the database including, but not limited to, criteria for inclusion and exclusion of data, the types of data included, the level of curation undertaken, and the mode and level of access that is facilitated.

- The technical and administrative functions that the database custodian has implemented to ensure the integrity of the data held by the database.

Examples of databases with well defined purpose include:

- COSMIC: [http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/about](http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/about)
- DMuDB: [http://www.ngrl.org.uk/Manchester/projects/dmudb](http://www.ngrl.org.uk/Manchester/projects/dmudb)
- Decipher:  
  [http://www.decipher.sanger.ac.uk](http://www.decipher.sanger.ac.uk)
- BIC:  

**S1.2** There must be a clearly defined procedure for regular review of the purpose and use of the database.

**C1.2** Regularity of the review of the purpose of the database should be stated, and the date when the review of the purpose and use of the database was last addressed must be readily accessible.

**S1.3** There must be a clearly defined procedure for reviewing the quality processes to ensure they are appropriate for any revision of the purpose and use of the database.

**C1.3** Quality parameter audits must be readily available and provide a clear description of the frequency of the audit cycle, what is undertaken, the degree to which the audit parameters have been achieved (i.e. were the minimum specifications achieved and/or exceeded), and actions arising from the outcomes of each audit.
2 Governance

S2.1 The governance structure for the operation of and access to the database must be clearly defined and be readily accessible

C2.1 The governance structure for the database must address oversight of all aspects of the database including privacy, secure access and sharing, content, quality, accuracy, curation, and clinical utility.

S2.2 A database custodian [ownership] must be identified, this custodian being accountable and responsible for the operation of the database, and who is accountable to a clinical institution or incorporated entity

C2.2(i) The custodian must ensure adequate curation and management of the data such that it is secure, accurate and subject to regular review.

C2.2(ii) The custodian must have sufficient authority to carry out or delegate actions as required.

C2.2(iii) The custodian is the entity with responsibility and accountability for the operation of the database. This may be a position, committee, board of the host institution, or incorporated entity which can ensure transparency, and business continuity.

S2.3 The custodian must appoint an appropriately qualified person as curator whose specific role is to oversee management of the database and be accountable to the custodian. The curator must have the requisite authority to carry out and/or delegate the tasks required.

C2.3(i) The custodian must ensure the curator has the appropriate skillset, knowledgebase, and experience to oversee the content and operations of the database (See Appendix – Curation); and provide access to continuing professional development.

C2.3(ii) the custodian and curator may delegate responsibilities to other parties, but such delegation must incur accountability to the curator or custodian.

C2.3(iii) In instances where the database is small, the custodian may also be the curator. It should be noted however that this does not necessarily constitute good governance.

C2.3(iv) If a multidisciplinary committee is appointed to assist the curator, the terms of reference of the committee, its constituent members (by position and expertise, not necessarily by name), and operational aspects such as meeting schedules and decision analytics models should be clearly defined and readily available for reference.

C2.3(v) Other approaches to curation may be considered, however accountability and responsibility to demonstrate efficiency and quality of the curation process reside with the custodian.
S2.4 The custodian must ensure that an appropriate policy regarding intellectual property held in the database is in place, that the policy is readily accessible to current and potential users of the database, and that implementation of the policy is audited.

   C2.4(i) In the case of a public database, a copyright notice and method of citation may be included to protect material rights.

   C2.4(ii) In the case of a public database, a disclaimer notice may be displayed to limit or exclude liability. Acknowledgement of this disclaimer may be a requirement for access to and use of the database.

S2.5 The curator must ensure that data sets submitted to the database comply with relevant professional standards and/or privacy legislation.

   C2.5(i) There must be a defined mechanism to facilitate communication between the authorised user and the curator. This should include the ability to provide free text information which may be periodically reviewed. This should be included in audit processes.

   C2.5(ii) There should be defined procedures relating to both clinical, technical, and regulatory issues; with regular quality audits to minimise occurrence or recurrence of operational issues identified by users or operators.

S2.6 When associating variant records in another or multiple databases, the external database/s must be audited prior to use.

   C2.6(i) An audit should be performed on external databases prior to utilisation of any data/information from it to reduce the risk of introducing errors into the database.

   C2.6(ii) Results or certificates of audits of a database conducted by a trusted third party may be accepted as evidence to minimise repetition and work burden.

   C2.6(iii) To increase efficiency and accuracy of information sourced from external databases, automation of the interrogation function to mine and update most recent data in the database is recommended. Such processes must be subject to control and regular audits.

2.7 The custodian must ensure there is appropriate ethical oversight of the database through an appointed ethics committee

   C2.7(i) The ethics committee may be a dedicated ethics committee, an institutional committee, or hosted by a professional society / organisation.

   C2.7(ii) Useful reference materials include:
• National Pathology Accreditation Advisory Council Requirements for Medical Pathology Services (First Edition 2013)\textsuperscript{6}
• The NHMRC National Statement of Ethical Conduct in Human Research\textsuperscript{7}.

S2.8 The custodian must ensure there is a procedure to follow in the event that the database no longer meets the stated purpose, or closes.

C2.8(i) The demise of a database may include abandonment, falling in to disuse due to reduced relevance, or being closed / discontinued due to obsolescence.

C2.8(ii) There should be a clear and detailed policy for transfer of data to ensure provision of continuity of access to data in the event of demise of a database. This may occur due to loss of funding to maintain the database, loss or change of custodianship, or force majeure.

C2.8(iii) There should be a clear and detailed policy for destruction of data in the event the database is closed because it no longer meets a need or is obsolete.

C2.7(iv) The RCPA Guideline *Privacy Guidelines – Managing Healthcare Information in Laboratories*\textsuperscript{8} discusses privacy principles related to pathologists and their laboratories. It also addresses the application of these principles when a pathology practice faces a change in business circumstance or closure.


\textsuperscript{7} National Statement on Ethical Conduct in Human Research 2007 (Updated March 2014). The National Health and Medical Research Council, the Australian Research Council and the Australian Vice-Chancellors’ Committee. Commonwealth of Australia, Canberra.

\textsuperscript{8} \url{http://www.rcpa.edu.au/getattachment/a631a573-0d07-4bd4-ba67-cfe545618dd1/Managing-Privacy-Information-in-Laboratories.aspx}
3 Establishment of Databases

This section outlines the basic requirements to ensure a fully functional and efficient database which is capable of maintaining data integrity in a secure environment.

S3.1 The infrastructure and storage capabilities of the custodian institution must be fully functional.

C3.1(i) The authenticity of the data must be maintained. The authenticity of a digital record refers to its trustworthiness i.e. that it is what it purports to be and it is free from tampering or corruption. Authenticity has two components:

- integrity: the quality of being complete and unaltered in all essential aspects, and
- identity: the characteristics of a record that uniquely identify it and distinguish it from any other record.

Any unintended change to a record or its identifiers as a result of storage, retrieval, processing and operation, including malicious intent, unexpected hardware failure, is a failure of data authenticity.

C3.1(ii) The underlying technical infrastructure used to implement the database must be capable of supporting the functionality required by this standard.

- A spreadsheet is not considered to be sufficient means of storing data.
- Examples which may meet the requirements include:
  - SQLite3, Microsoft SQL Server, PostgreSQL, MySQL, MongoDB, Apache Cassandra.

C3.1(iii) If the database is designed for public access, there should be a web interface to enable efficient access by users, and facilitate data gathering, sharing and report retrieval.

C3.1(vi) Data should be exportable and compatible with other data repositories used in healthcare to allow for efficient data sharing.

S3.2 Any modifications or updates must be new records, version controlled, and linked to the initially created record.

S3.3 Complete provenance information for all records must be stored within the database to ensure that records are effectively permanent and the state of any record at any point in time can be viewed.

C3.3(i) Provenance information should indicate:

- the origin;
- intermediate source(s); and
- complete modification history of the data.
C3.3(ii) The provenance information should be visible to all users, with the exception that any free text which might identify the patient should only be visible to authorised users (see 3.7).

S3.4 There must be a policy regarding audit of the database. This policy must be readily available, together with the last date on which the audit was performed.

C3.4(i) There must be a complete audit trail of changes to any record to ensure that all records are effectively permanent.

C3.4(ii) The complete audit trail should be visible to viewers, with the exception that any free text which might identify a patient should only be visible to authorised users (see S3.7).

C3.4(iii) A spreadsheet will not support the journaling requirements for audit trails and is therefore not an appropriate form of media for use as a database.

S3.5 The database backup must contain all required information to reinstate the database with minimal reconstruction in the event of a catastrophic failure.

C3.5(i) The data backup policy must ensure that the hardware, software, geographic location of redundant dataset/s, network accessibility, and personnel responsible for data backup are consistent with a high level of protection of patient privacy and confidentiality.

S3.6 The database must be backed up at regular intervals to minimise loss of data and the required reconstruction in the event of a catastrophic failure.

C3.6(i) There must be a policy which specifies the frequency and type of backups, together with regular audits to ensure that this policy is implemented.

C3.6(ii) Backups should be in a form which can be reloaded in to a database with minimal effort.

C3.6(iii) A three tier back up system should be employed
(a) The original database (current in daily use)
(b) A separate local copy via network or manual
(c) A separate offsite copy (i.e.: a separate location such as an encrypted portable hard drive or cloud based solution)

S3.7 To minimise the risk of inadvertent disclosure of private information, free text must not be included in a record that may be accessed by users without appropriate access authorisation.

C3.7(i) Free text data that is submitted must be reviewed by the curator to exclude personal identifiers before a record is made public.
4 Privacy, Confidentiality, Ethics, and Data Security

There are significant ethical, legal, and social issues that must be considered and handled responsibly when developing, operating and de-commissioning a sequence variant database. These issues relate to concepts of privacy and confidentiality for both patients and health system workers, the right of autonomy for individual patients and the related right to make decisions about the way information about them and their health care is used and disclosed, as well as broader societal concerns regarding the public interest and the benefits that can be derived from the use of genetic variation information. The responsible handling of these issues in Australia is mandated by a complex mix of state, territory and commonwealth legislation—often informed by international declarations and treaties from bodies such as the OECD, APEC and UNESCO—regulation, advice from the Office of the Australian Information Commissioner, relevant professional practice standards and professional codes of ethics.

Exactly which components of the above mix apply to the development, operation and de-commissioning of any one database is dependent on the jurisdiction under which the database custodian operates, whether the database custodian is a health service provider and if it operates in the public or private sector, the annual turnover of the database custodian, the primary use for which the data included in the database was collected, the intended use or uses of the database—i.e. for clinical or research purposes—and whether the information stored in the database is considered “personal information” under the Privacy Act 1988 (Cth). The RCPA Guidelines The Ethical and Legal Issues in Relation to the Use of Human Tissue and Test Results in Australia and Managing Privacy Information in Laboratories provide a more in-depth discussion of how these issues are controlled and regulated in Australia for accredited pathology laboratories.

The development, operation and de-commissioning of databases where the database custodian is located in a country other than Australia (international databases) will be regulated under the legislative and regulatory requirements of that country. Importantly, if an Australian entity transfers “personal information” about any Australian individual to international databases, then the Australian entity is responsible for ensuring that the international databases only use or disclose that information in accordance with the Australian Privacy Principles or be “subject to a law, or binding scheme, that has the effect of protecting the information in a way that, overall, is at least substantially similar to the way in which the Australian Privacy Principles protect the information.”

This document covers sequence variant databases used for clinical purposes and, as such, the Standards and Commentaries in this section should be considered to only apply to such databases. The Standards and Commentaries may not be applicable to databases that are used for research purposes.

---

10 http://www.rcpa.edu.au/getattachment/b52a239d-c5da-4f9c-8670-c65b14380e8f/Ethical-Legal-Issues-Use-Human-Tissue-Test-Results.aspx
S4.1 The database custodian must comply with relevant local legislation, regulations and professional practice standards in all aspects of the development, operation and de-commissioning of the database.

C4.1(i) The collection, storage, use and disclosure of all information must comply with all legislation and regulations that deal with the privacy and confidentiality of:

- the patients from whom the data is derived; and
- the laboratories, clinicians and laboratory staff who are submitting the data.

The RCPA Guideline Privacy Guidelines – Managing Healthcare Information in Laboratories\textsuperscript{13} discusses privacy principles related to pathologists and their laboratories.

C4.1(ii) In deciding what legislative and regulatory requirements must be met, a decision must be taken as to whether any information collected, stored, used or disclosed by the database custodian would constitute “personal information” under the Privacy Act 1988 (Cth). The Act defines personal information as information or an opinion about an identified individual, or an individual who is reasonably identifiable:

- whether the information or opinion is true or not; and
- whether the information or opinion is recorded in a material form or not. \textsuperscript{14}

The Act only applies to personal information. Guidance exists for what constitutes personal information under the Act. Whether an individual can be identified or is reasonably identifiable depends on context and circumstances. While it may be technically possible for an agency or organisation to identify individuals from information it holds, it may not be practical to do so. For example, logistics or legislation may prevent such linkage. In these circumstances, individuals are not ‘reasonably identifiable’. Whether an individual is reasonably identifiable from certain information requires a consideration of the cost, difficulty, practicality and likelihood that the information will be linked in such a way as to identify him or her. \textsuperscript{15}

De-identified information is not ‘personal information.’\textsuperscript{16} The Office of the Australian Information Commissioner provides guidance on what constitutes de-identification of data.

\textsuperscript{13}http://www.rcpa.edu.au/getattachment/a631a573-0d07-4bd4-ba67-cfe545618dd1/Managing-Privacy-Information-in-Laboratories.aspx
\textsuperscript{14}Privacy Act 1988 (Cth) s 6 (definition of ‘personal information’).
\textsuperscript{15}Explanatory Memorandum, Privacy Amendment (Enhancing Privacy Protection) Bill 2012 (Cth) 61.
De-identification involves removing or altering information that identifies an individual or is reasonably likely to do so. Generally, de-identification includes two steps:

- removing personal identifiers, such as an individual’s name, address, date of birth or other identifying information, and
- removing or altering other information that may allow an individual to be identified, for example, because of a rare characteristic of the individual, or a combination of unique or remarkable characteristics that enable identification.

De-identification may not altogether remove the risk that an individual can be re-identified. There may, for example, be a possibility that another dataset or other information could be matched with the de-identified information. The risk of re-identification must be actively assessed and managed to mitigate this risk. Relevant factors to consider when determining whether information has been effectively de-identified could include the cost, difficulty, practicality and likelihood of re-identification.

NHMRC Guidelines approved under Section 95A of the Privacy Act 1988 (the Guidelines) provide a framework to ensure privacy protection of health information that is collected, used or disclosed in the conduct of research and the compilation or analysis of statistics, relevant to public health or public safety, and in the conduct of health service management activities. The Guidelines form part of compliance requirements under the Australian Privacy Principles established in the Privacy Act 1988 (Cth).

C4.1(iii) If data are being submitted from outside the jurisdiction of the custodian, the requirements of other jurisdictions should be considered e.g. HIPAA, GINA.

C4.1(iv) If the database is stored in the Cloud, the regulatory requirements of the jurisdiction in which the custodian is located must be met.

C4.1(v) Where the purpose of the database includes medical research, the custodian must ensure that the management of the database also complies with the NHMRC national statement on ethical conduct in human research or local equivalent.

S4.2 The database must have a readily accessible policy regarding the management of information that reflects the purpose of the database.

C4.2(i) Disclosure in the context of this document means authorised access to data by a third party where the data remains in the control of the database custodian or the sharing of data with a third party by the

---

17 Ibid B.54-55 page 12.
18 Guidelines approved Under Section 95A of the Privacy Act 1988, National Health and Medical Research Council, Commonwealth of Australia, March 2014.
19 Health Insurance Portability and Accountability Act.
20 Genetic Information Non-Discrimination Act.
21 NHMRC national statement on ethical conduct in human research.
database custodian whereby the database custodian relinquished control of the data.

C4.2(ii) The information policy must include, but not be limited to, descriptions of how the data are collected, stored, used and disclosed by the database custodian.

C4.2(iii) If the database collects, stores, uses or discloses personal information, the information policy must address how the personal information is managed.

The Australian Privacy Principles 22 contain guidance on what information such a policy must contain.

C4.2(iv) The information policy should include information on how informed consent is collected for the collection, storage, use and disclosure of the data included in the database.


C4.2(v) If a determination is made that informed consent is not required for the collection, storage, use, and/or disclosure of personal information, the information policy must contain information on who made the determination and the reasoning used to justify the determination.

The Privacy Act 1988 (Cth) includes provisions for exceptions to the requirement to collect informed consent for the collection of personal information, including when such information is collected for:

- the compilation or analysis of statistics relevant to public health or public safety;
- the management, funding or monitoring of a health service.24

C4.2(vi) The information policy should address the manner of any data de-identification employed, as appropriate to the database. This should include a justification of the effectiveness of the de-identification techniques employed and an assessment of the risk of unauthorised re-identification of patient data. The policy should address data that are shared externally, e.g. by upload to a central data repository, to other databases, or other third-party users.

C4.2(vii) The International Code of Conduct for Genomic and Health Related Data Sharing, developed by the Regulatory and Ethics Working Group, Global Alliance for Genomics and Health (GA4GH), provides a

---

24 Privacy Act 1988 (Cth) s 16B(2)(a)
principled and practical framework for the disclosure of genomic and health-related data.\(^{25}\)

**S4.3 There must be mechanisms in place to control disclosure of data held in the data repository.**

- **C4.3(i)** The benefits of open sharing of data must be weighed against the risks to the privacy and confidentiality of the patients from whom the data in the database is derived. Where these risks outweigh the benefits, due consideration should be given to adequate deidentification of data and access to data should be through a register of approved users, rather than full open access. This should be driven by the clinical need to utilise the data for clinical or translational research and or clinical diagnostic purposes. Refer to S4.4 for descriptions of database user tiers.

- **C4.3(ii)** Methods to ensure confidentiality should take into account the nature of the entity with whom data is being shared, the nature of the data being shared, and the use that the data will be put to. A risk-based approach, following current best practice guidelines should be taken.

**S4.4 The database must be protected to ensure data security and to protect privacy and confidentiality of individuals.**

- **C4.4(i)** Users should only be permitted to gain access to the information that they are entitled to view. The management of access involves issues of computing security that lie beyond the scope of this document, and will involve liaison between the curator and the host institution’s IT management. The National eHealth Transition Authority (NEHTA) has identified a number of standards that are pertinent to this issue (see management for Australian Clinical Quality Registries\(^{26}\)).

- **C4.4(ii)** The database **must** comply with relevant health information systems and security standards including:
  - Health informatics – Functional and structural roles
    - ISO/DIS21298 (draft international standard 2014)

- **C4.4(iii)** Mechanisms should include password protected access, data encryption, licensing/certification for access, application for access, and access audit trails.

- **C4.4(iv)** Physical access to the underlying technical infrastructure on which the database is hosted **must** be controlled to ensure security of information.

---

\(^{25}\) **International code of conduct for genomic and health-related data sharing.** *The HUGO Journal* June 14 2014, 8:1 doi:10.1186/1877-6566-8-1

C4.4(v) The database must have a defined mechanism for managing appropriate access to different types of information held in the database.

For example, a four tier access level strategy could be considered for any database which operates beyond the scope of an individual laboratory, thereby facilitating the protection of the database contents.

(a) **Level 1: Unregistered Viewer**: This user has access to only collated information with no details regarding the patient sample, or laboratory submitting the information.

(b) **Level 2: Registered Viewer**: this user has access to individual reports of the variant, including patient identifier and laboratory submitting the information. However, this information is read-only.

(c) **Level 3: Registered Submitter**: this user has access to read-only information (as for a registered viewer) and is also able to submit information from a specific laboratory.

(d) **Level 4: Curator**: access to all information, and the ability to annotate certain fields.

(e) **DBAdmin**: Administrative and operational access (IT staff).

S4.5 **The privacy of laboratories, clinicians and laboratory staff who are submitting information to the database must be maintained.**


C4.5(ii) Where it is common practice to list submitters to a database, authorisation to publicly list the submitter must be obtained from each individual submitter.

S4.6 **The database custodian must ensure that the development, operation and management of the database complies with the information policy.**

C4.6(i) As part of the user registration process, a user should explicitly acknowledge they have read and understood the policy and a method of recording their acceptance of the obligations it details should be available.

C4.6(ii) The information policy should include how a breach in the terms and conditions of use of the database will be dealt with.

---

C4.6(iii) A registered user is defined as described in C4.4(v) above.

C4.6(iv) Audits of a database conducted by DBAdmin users should include:
- Review of access logs
- User management: information about the registrations to the database
- Database (usage) statistics: information about the number of contributors and contributions to the database
- Log in events and registration types

C4.6(v) Risk Analysis must be conducted periodically as a component of a systems test after hardware or software modifications or upgrades to identify and remove vulnerability to any threats or weaknesses identified. More information regarding quality systems can be found in the NPAAC document, Requirements for Medical Pathology Services (First Edition 2013).

S4.7 There must be a documented procedure/policy readily available to be followed in the event of a security breach or unauthorized disclosure of information.

C4.7(i) This policy should address the technical, operational, legal, and ethical consequences of such a breach. Such a breach may carry legal or professional obligations to report the breach to clinicians, patients, the host institution, contributing laboratories, and regulatory authorities.

C4.7(ii) The Office of the Australian Information Commissioner has produced a useful guide to data breach notifications.

---

5 Content

The scope of the data to be captured and maintained should be clearly articulated to encourage consistency across databases. It is essential there be adequate information provided to enable an external user to review and use the information with confidence in making clinical decisions.

S5.1 The means of submission of data to databases must include the use of commonly used data exchange formats

C5.1(i) The supported formats should be clearly stated and publicised through appropriate mechanisms such as internally documented procedures or on a database website

C5.1(ii) Submission using standard formats reduces the risk of corruption of data during upload into the database, and facilitates sharing of data, and federating of databases.

S5.2 Each record in the database must include the following data:

- The variant described using a recognised (and specified) nomenclature that uniquely identifies the variant. This must be referenced to the sequence stipulated by the database or precisely state which reference has been used.
- The zygosity state must be provided if known.
- For variants described using genomic coordinates, the reference Genome Build must be stated.
- The methodology of variant detection
- The reason for testing must be provided in the context of relevance to the database.
- If the submission includes a statement regarding clinical interpretation or significance, the basis of this statement must be provided. This may include reference to other published sources or to unpublished studies by the submitting laboratory.

C5.2(i) The Primary descriptor of the variant should be provided in HGVS nomenclature with reference to a genome sequence (genomic coordinates). Secondary descriptors (non-standard aliases or reference to a transcript sequence) may be included. Mapping tools and conversions must be identified.

C5.2(ii) Established legacy nomenclatures which are difficult to change may be used where conversion to HGVS is not possible, or difficult e.g. HLA haplotypes. The nomenclature being used must be clearly stated.

C5.2(iii) Where possible, information regarding the frequency of the variant in the tested population or a control group should be submitted.

C5.2(iv) Where available additional information should be provided e.g. protein function prediction, splicing abnormality prediction, literature
evidence, familial studies on variant co-segregation with disease, or other relevant evidence.

C5.2(v) A reason for testing may be:

- Diagnostic test in affected person
- Predictive testing in an unaffected/affected person at high risk of a specific mutation
- Predisposition testing for a particular disease in an unaffected/affected person who is not at high risk of a specific mutation
- Segregation analysis to assist with pathogenicity assessment
- Screening test for many disorders in a person who is not at high risk of specific disease.
- Theranostic testing in an affected person to guide therapeutic decisions.

C5.2(vi) Where available, the clinical phenotype and supporting multidisciplinary evidence should be provided. This may include:

- Patient history and diagnosis
- Inheritance information – This should include the number of affected and unaffected individuals tested for the variant – suspected mode of inheritance, consanguinity, ethnicity, gender, age at diagnosis.
- carrier status
- pathogenicity
- Relevant non-genetic pathology results
- Relevant non-genetic medical results

C5.2(vii) If phenotype data are submitted together with genotype data, the phenotypic information would preferably be reviewed by a relevant multidisciplinary team (MDT) or clinical service specialising in the disease.

S5.3 Standardised terminology and a recognised international ontology must be used within the database. The selected standard should be clearly stated and made available through appropriate media such as internally controlled operating procedures or on the database website.

C5.3(i) If phenotype information is provided in the database, the ontology system being used must be stated (e.g. SNOMED CT, Human Phenotype Ontology (HPO), etc.)

S5.4 There must be clearly defined guidelines for the classification of variants. Any pathogenicity classification must provide detailed information describing how and why the classification has been made.
C5.4(i) The criteria for classification of pathogenicity should be evidence based, clearly stated, and available through appropriate media such as internally controlled operating procedures, or on a database website. NPAAC Requirements for Medical Testing of Human Nucleic Acids\(^{30}\) and Requirements for Medical Pathology Services \(^{31}\), or local equivalents, should be referred to regarding reporting decisions and classifications.

An example of this is the InSiGHT classification criteria for mismatch repair genes [www.insight-group.org/criteria](http://www.insight-group.org/criteria).

C5.4(ii) If pathogenicity is determined by the submitter, it **must** be stated how the pathogenicity was determined. There may be multiple fields in the database where this information is recorded.

C5.4(iii) The database **must** flag where there are inconsistencies in the database, with a mechanism to resolve the discrepancies (e.g.: same variant submitted twice or more with different pathogenicity classifications). See Section 7 for more detail.

C5.4(iv) A description of “research” undertaken to reach the conclusion including citation of any peer reviewed papers should be included to enable the database user to make an informed professional judgment about the pathogenicity classification with a certain level of confidence.

C5.4(v) The level of confidence in the classification may be included.

C5.4(vi) The classification of pathogenicity should be described within the purpose of the database. This may include the reason for testing, and what pathogenicity means in what context (e.g.: describing the database as an LSDB versus a generic genome database).

**S5.5** Each patient, and each family, must have a unique identifier applied.

C5.5(i) The unique identifiers should be system generated.

C5.5(ii) These unique identifiers are required to flag the frequency and co-occurrence of variants. For example, this gives the database the ability to flag to the user that there are multiple variants in different genes in a single individual, or that a single variant is in a number of related (familial) or unrelated individuals.

C5.5(iii) The system generated identifiers also provide a mechanism which allows the submitting laboratory to identify the individual patient data within the database submitted by that laboratory. This is to enable future updating and correction of information which may impact patient management. It is neither necessary nor appropriate that any

---

\(^{30}\) National Pathology Accreditation Advisory Council Requirements for testing of human nucleic acids (second edition 2013) Commonwealth of Australia

user other than the submitter be able to make this association. This restriction on identifying a patient applies equally to the curator and custodian as to other registered and unregistered viewers of the database.

C5.5(iv) There must be a mechanism by which unique patient and family identifiers can be updated by the curator. This is necessary in the event that a submitter realises that multiple instances of a variant identified in supposedly unrelated people are actually from members of the same family, or that an individual’s results have been submitted multiple times to the database as independent events.

C5.5(v) A submitter should have a mechanism to relate the database-generated unique identifier to their own in-house medical records.

S5.6 The accreditation status of the laboratory which performed the analysis must be stated.

C5.6(i) The definition of an accredited laboratory in the context of the database must be clearly stated and made available on the database website interface or (in the case of a private or restricted database) in a documented policy.

S5.7 The analytical validity of the report must be clearly indicated, and documented in sufficient detail to enable assessment by a viewer.

C5.7(i) For each report of a variant, the following information must be included:

- A measure of the quality of the variant call. In a tightly controlled environment (such as a validated test in a laboratory accredited for clinical diagnostic service delivery), it may be sufficient to have a statement regarding quality which applies to the entire dataset. In a less controlled environment (such as an RUO or translational test in a research or unaccredited translational laboratory), it may be necessary to have a statement regarding quality for each reported instance of a variant.
- The consistency [accuracy of the nomenclature] of the variant events held in the database must be demonstrated, and the provenance of the data must be defined. This is a joint responsibility of the submitter and the curator.
- Indicate whether orthogonal method verification or previously validated test was performed. This is intended to lend more integrity to the data – if confirmed, or already validated, the user is likely to feel more confident utilising this information clinically than if it has not been confirmed by an alternative method or run as a validated test.

C5.7(ii) For further guidance regarding analytical validity parameters in medical testing, refer to The NPAAC reference material, Requirements for the development and use of in-house in vitro
In relation to genomic sequencing, refer to Assuring the Quality of Next Generation Sequencing in Clinical Laboratory Practice Working Principles and Guidelines\(^{33}\) developed by the Next Generation Sequencing Standardisation of Clinical Testing (Nex-StoCT) Workgroup.

**S5.8** The clinical validity of the report must be clearly indicated and documented in sufficient detail to enable assessment by a viewer.

**C5.8(i)** For each report of a variant, the following information must be included:

- Any records considered valuable in defining provenance. This may include diagnostic records, peer reviewed papers, research reports, confirmation of variant by other methods
- Clearly indicate whether the consequences were experimentally determined or only theoretically deduced.
- When changes in patients with a recessive disease are described, it should be clear in which combination (phase) the changes were found.

**C5.8(ii)** Further guidance regarding clinical validity parameters in NGS based genetic tests – can be found at: reference/s for guidance on clinical validity parameters

**5.9** It must be stated clearly when a review of data interpretation has taken place.

**C5.9(i)** There should be a clearly defined policy regarding data re-analysis, and this should be made available through appropriate media such as internally controlled operating procedures or on a database website. Expectations for re-analysis and re-interpretation of data should be managed against laboratory/database resources and priorities.

---


\(^{33}\) Gargis,A. et.al. Assuring the quality of next-generation sequencing in clinical laboratory practice.; Supplement 1, Nature Biotechnology, Vol30, No.11, November 2012
6 Functionality

A database is only as useful as the information contained within, and the ease of access to the information in a relevant, efficient and informative fashion. Ease of access includes the visual appearance and navigational qualities of a database and website interface, as well as quality and usefulness of interrogation capabilities.

S6.1 The database must have flexible search capabilities, including the ability to search the content of each record over time

C6.1(i) Search capabilities should be customisable to allow for multiple types of queries including orthogonal queries to increase filtering capabilities.

C6.1(ii) Examples of searchable fields include specific variant, gene name, alias (gene, disease), disease type and classification, phenotype, protein, codon, ethnic group, geographic location, author or citation.

C6.1(iii) Probabilistic search capabilities (“fuzzy” searches) may also be desirable

S6.2 The database must be capable of associating variant records from another or multiple databases complying with the requirement to audit such external databases prior to data import per S2.6.

C6.2(i) The database should support the ability to import good data to enhance the usefulness of a database.

S6.3 The database must have the functionality to allow for tracking for regular review and updates, and aggregate information in a version controlled manner.

C6.3(i) It is desirable that the database is able to track updates for entries and retrieve or receive data from participating sources automatically. In the event that this is not possible, the database should be capable of batch uploading from tables or spreadsheets to maintain most current information. This should include key metadata which defines the provenance and flags the status of the variant/s reported; noting any correction of nomenclature errors.

S6.4 The database must be able to account for the number of occurrences of the same variant in the same individual or in the same family.

C6.4(i) The database should have a means of flagging variants that have been reported in one individual and/or one family versus many unrelated individuals. This is needed to distinguish between rare/isolated events versus common events, and minimise “double” counting of the same variant event.

S6.5 The database must be able to generate summary reports for viewers
C6.5(i) Summary data of information held in the database should be provided to the users of the database. This may include (but not limited to) the number of:

- Genes described
- Samples entered (single entities)
- Coding mutations
- Papers cited, authors
- Unique Variants
- Fusion genes
- Genomic rearrangements
- Whole Genomes
- Whole Exomes
- Copy Number anomalies
- Mutation maps
- Graphics, tools, location of a variant in a gene

C6.5(ii) Summary reports may be customised according to context of the reports and needs of the end users, and be downloadable. For example: COSMIC Keyword search using the term “Lung” provides a summary report, and selection of the Primary site provides a customisable report including further internal links to more detailed information.

S6.6 The database must be able to support the transition of existing data to newer version of the Human Genome reference build as they become available.

C6.6(i) As newer versions of the Human Genome Reference build are released, the existing database content should be re-mapped to the new reference within a reasonable time frame.

C6.6(ii) The old variant description should be listed with the new variant description to allow searching using any version.

C6.6(ii) The conversion should be automated where possible. Automated transition modules are available (e.g.: the batch liftOver tool by UCSC, or Remap by NCBI). The module which is used must be accredited or optimised and validated by the laboratory. This is to ensure the robustness of the program and harm minimisation (such as corruption of existing data).

C6.6(iii) Any conversion should be clearly indicated to alert users.

S6.7 The database must be able to identify incorrect and inconsistent data entries.

---

34 http://chromium.liacs.nl/LOVD2/colon_cancer/variants_statistics.php
http://asia.ensembl.org/Homo_sapiens/Gene/Variation_Gene/Table?db=core;g=ENSG00000134982;r=5:112707498-112846239
C6.7(i) This should be an automated process with notification of incorrect or inconsistent records to the curator for review and correction where required.

C6.7(ii) The database should have a means of flagging duplicate entries of the same variant event.

S6.8 **Mechanisms for access and sharing between data repositories must be supported.**

C6.8(i) There should be mechanisms for importing data and exporting data for inclusion in an external database in a compatible format.

C6.8(ii) The type of data that can be shared should be clearly stated. Sharing of aggregated data may be a more readily accepted method for sharing data.

C6.8(iii) Mechanisms for supporting useful links such as clinical content / information websites may be included.

C6.9(iv) A citation list provided for individual or aggregated data should be included.
Appendix

1 The Curator

Curation is the activity of managing and promoting the use of data from its point of creation, to ensure it is fit for contemporary purpose, and available for discovery and re-use\textsuperscript{35}. It involves the selection, preservation, maintenance, collection, and archiving of data in order to establish, maintain, and add value to repositories for current and future use.

Challenges of DNA data repository curation include:
- The increasing rate of creation of data sets as MPS reduces in cost and increases in output
- Standardising terminology and ontology within a database/set of data/federated databases
- Filtering and triaging variant calls and evaluating level of confidence in accuracy
- Maintaining relevance, and accuracy of data within the database
- Maintaining currency of genome build and compatibility of variants recorded to current / updated genome builds (i.e. correct alignment of sequence to reference)
- Facilitating access and sharing through secure links
- Compatibility of database schema with external and/or federated databases

Good curation means checking provenance of the variants and any associated information: was it from an accredited source, is the evidence robust, and is there any metadata to support the variant identification. What, if any, is the evidence for pathogenicity?

Exemplars of well curated databases include DMuDB, ClinVar, and COSMIC (Catalogue of Somatic Mutations in Cancer)

When a database has been established, the Curator’s role includes the undertaking or delegation of the following tasks:

Curation:
- Provide general information about the database contents and functionality
- Evaluate and register new submitters
- Curate new submissions
  - ensure standards that are set for data collection / submission are met
    - formatting of submitted data to be uploaded to the database
    - ensure information within the database is accurate, up to date, and accessible
    - understand the ways in which genetic variant information are presented / stored (e.g.: nomenclature used), and utilise automated tools (such as MutaLyser) to perform variant curation, data formatting, and related tasks

\textsuperscript{35} Lord, Philip, and Alison Macdonald. e-Science Curation Report: Data curation for e-Science in the UK: an audit to establish requirements for future curation and provision. Digital Archiving Consultancy Limited, 2003.
• Ensure the consistency [accuracy of the nomenclature] of the variant events held in the database are demonstrated. If this is not possible, provenance of the data must be defined and ensured
  ▪ Assist any committee which is in place to classify or approve classification of genetic variants prior to inclusion in database (for example: providing supporting data, and arranging virtual / face to face meetings for discussion
    o Check URLs for currency
    o Prevent duplication of entries; and linking of entries within the database or between databases
    o Extract relevant information from published literature (data mining) and other resources. Upload this data as required and or organise in a manner which allows them to be used for the classification of variants
    o ensure mechanisms for separating public and non public data are maintained and comply with local privacy laws
    o review and approve any Locus Reference Genomic (LRG) sequences
• Undertake audits according to predefined schedule
  o Review and action any database content audit outcomes
• Update data entries (enter new publications, conclusions on functional consequences)
• Update variant descriptions as they evolve
• Promote the database
  o Apply the concept of microattribution to acknowledge database submitters for their contributions
  o Collaboration with other groups, national and global, to encourage and facilitate the sharing of data
  o Contact database submitters to inform them of other parties which are interested in their data / similar patient cohorts
  o Analysis of database statistics for presentations, grants, and papers.
• Publish regular updates including
  o Summary reports of database contents
  o Frequency of database updates, and last update,
  o audit outcomes,
  o notifications of corrections and or removal of data
• Ensure contingency plans for emergencies (back-up curators/ administrators, database transfer/archive) are in place

**Maintenance** (The following may be performed by the curator or delegated to a database administrator; however curator is responsible for overseeing):
• Server administration
• Maintain server and data security (operating system, firewall, back-ups, etc.)
• Assist with technical development of the database
  o Update software platform regularly (bug fixes, latest functionality)
• Review and action any database infrastructure audit outcomes
2 REFERENCES


vi Burn, J., Douglas, A. (2013) Delivering the 100,000 Genome Project in the NHS, April 2013. Downloaded January 29 2014


x http://www.iccg.org/about-the-iccg/clingen/
Development of Standards for the Accreditation of DNA Sequence Variation Database – 93

January 2015 FINAL REPORT


xviii Tyrrell et.al., Quality Standards for Databases of DNA Sequence Variants BSGM, Mutation Detection September 2013, Appendices X.XX


xx Mitropoulou, et.al., Locus-Specific Database Domain and Data Content Analysis: Evolution and Content Maturation Toward Clinical Use. Hum Mut 31(10), 1109-1116, 2010


xxii http://www.nata.asn.au

