Quality Standards for Databases of DNA Sequence Variants

Bruce Bennetts[1], Melody Caramins[2], Arthur Hsu[3], Chiyan Lau[4], R. Scott Mead[5], Cliff Meldrum[6], Graeme Suthers[7], Graham Taylor[3], **Vanessa Tyrrell[4]**

1) Western Sydney Genetics Program, The Children's Hospital at Westmead, New South Wales, Australia
2) SDS Pathology, North Ryde, New South Wales, Australia
3) Department of Pathology, University of Melbourne, Melbourne, Victoria, Australia
4) Royal College of Pathologists of Australasia, Surry 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 Genetics Service, SA Pathology, Adelaide, South Australia, 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 required 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.

The Royal College of Pathologists of Australasia (RCPA) in collaboration with the Human Genetics Society of Australasia (HGSA), and the Human Variome Project (HVP) is developing standards for DNA sequence variation databases intended for use in the Australian clinical environment.

The fundamental principle underpinning the standards document is that 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. The document is intended to:

- Essential elements Complement existing laboratory standards and accreditation requirements,
- Align with global initiatives in existence (e.g.: GA4GH, ClinGen, DMuDB,
- Act as a guide to identify a quality database, to establish new, or improve existing databases, and
- Set minimum requirements for clinical purposes within the boundaries of

2. The Ideal Clinical Database[i]

Governance

- Ethics and compliance
- Custodianship
- Intellectual Property

Secure, protection of privacy, with controlled access and sharing

- Aggregated data accessible at a minimum
- Clinic / 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 review 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 database features and content such as those highlighted by Mitropoulou et.al[ii] in a systematic order, with clearly defined criteria.

1) Purpose

- Essential elements of the nature of use

2) Governance

- Custodianship, maintenance of relevance, ethics, compliance, intellectual property, sustainability

3) Establishment of databases

- Requirements for functional databases, quality, back up of database.

4) Privacy, Protection, Security

- Information, access, and sharing policies

5) Content

- Minimum data requirements, analytical & clinical validity, variant classification, nomenclatures

6) 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 (NPAAC) reference materials[iii].

Standard (SX.X): Minimum requirement, mandatory

Commentary (CX.X): Give clarification to, provide examples and guidance on interpreting the standard.

Normative = prescriptive, mandatory,

Informative = assist in interpretation of the standard.

5. Standards Examples

S3.3 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.

C3.3(i) There must be a complete audit trail of changes to any record to ensure that all records are effectively permanent. [Normative]

C3.3(ii) The complete audit trail should be visible to viewers. [Informative]

S6.1 The database must have flexible search capabilities

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

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

6. Implementation & Conclusion

Broad consultation 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.

Resulting standards are intended to be an **adjunct to existing NPAAC 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.

Resulting standards will be promoted globally through the **Human Variome Project International** (HVPI), 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/Practising-Pathology/DNASeqVar** or email **vanessat@rcpa.edu.au**