THE HUMAN VARIOME PROJECT

sharing data • reducing disease

Coordinating Office Position Paper

AP03-2012: Minimal content for gene variant databases (LSDBs) ICO Comments

PP-ICO-08-2012

19 October, 2012

Human Variome Project International Ltd.
Level 2, 161 Barry St
Carlton South VIC 3053
AUSTRALIA

Prepared by:
Timothy D. Smith

Contact:
Timothy D. Smith
Level 2, 161 Barry Street
Carlton South VIC 3053
AUSTRALIA
+61 (0)3 8344 2096
tim@variome.org

Authorisation:

This Position Paper has been prepared by Timothy D. Smith and represents the official position of the Human Variome Project Coordinating Office only. It does not represent an official position of the Human Variome Project, its Consortium, Advisory Councils or International Scientific Advisory Committee.

-s-

Heather Howard

I. Context

The International Coordinating Office (ICO) received the attached Activity Proposal (AP) on the 18th of October, 2012. Pursuant to the Project's Standards Development Process (PD06-2011) we are forwarding it to the International Scientific Advisory Committee (ISAC) for referral to a Sponsoring Council (SC). Further pursuant to the Standards Development Process we have prepared the following comments on the AP for the ISAC and SC.

II. Comments

Background

Guidance around the minimal content requirements of gene/disease specific databases is probably the most requested information from the ICO. The development of an HVP Standard or Guideline in this are is therefore both timely and prudent.

A number of recommendations have been made in this area in the past, both stimulated by independent of the Human Variome Project. Charles Scriver, informed by his work with the PAHKb, first attempted to provide guidance to other databases (Scriver, Nowacki, & Lehväslaiho, 1999, 2000). Both Mireille Claustres (Claustres, Horaitis, Vanevski, & Cotton, 2002) and Christina Mitropolou (Mitropoulou, Webb, Mitropoulos, Brookes, & Patrinos, 2010) have provided summaries of the most ubiquitous fields across existing database instances, and other have provided guidelines for the creation and curation of these databases that touch on the fields and types of information necessary (Celli, Dalgleish, Vihinen, Taschner, & den Dunnen, 2012; Vihinen, den Dunnen, Dalgleish, & Cotton, 2012). Most notably, the Gen2Phen Consortium (http://www.gen2phen.org) have done a considerable amount of work creating data models for both genotype and phenotype data.

Despite this background work, there is not yet a recognised standard for minimal content requirements. The Gen2Phen work perhaps comes the closest to a standard, but this work has not had the benefit of a truly international consultation and review process, and is perhaps too technical to be useful for the majority of Human Variome Project Consortium members.

Scope of Work

We recommend that an HVP Working Group (WG) to develop minimum content requirements for gene/disease specific databases.

At first, this activity seems directly within the purview of the Gene/Disease Specific Database Advisory Council as it deals solely with content requirements for gene/disease specific databases. However, the ISAC, when making its determination as to the Sponsoring Council (SC) for this activity, should remember that HVP Country Nodes are required to pass information to gene/disease specific databases. As such, the interests of HVP Country Nodes may need to be represented on the SC, and the ISAC may be the better alternative.

The initial scope of work proposed in the AP (dot points 1 & 2), in our opinion, would be a sufficient amount of work to address via this activity in the first instance. While the merits of the additional 8 areas of investigation cannot be denied, and should be addressed by the Human Variome Project at

some point, the Sponsoring Council for this activity should take care that the chartered Working Group is not forced to sacrifice the quality and timeliness of their work by taking on too much.

The SC should also note that the activity "required copy right statements, databases policy document, etc." is being addressed, at least in part, by WG01: Disclaimer Statements on G/DSDB websites,[1] a working group of the Gene/Disease Specific Database Advisory Council.

The result of this activity should, in our opinion, be published as an HVP Standard. However, the WG may choose to include within the final standard, a number of components that carry less weight and could more appropriately be considered HVP Guidelines.

We support the intention of the AP submitter to see the result of the WG's efforts published in a peer-reviewed journal, but must caution the SC to ensure that such publication does not limit the Project's ability to publish, disseminate and create derivatives of the work during the course of its normal activities.

Possible Working Group Members

The ICO does not wish to recommend any specific individuals for this activity's WG, but would strongly recommend that the SC strongly consider including representatives of Gen2Phen who have experience in this field. We would also recommend that the MIBBI Project (Minimum Information for Biological and Biomedical Investigations) be approached to collaborate on this activity.

Available Assistance from the ICO

The ICO should be able to provide the assistance requested in the AP.

III. Summary of Recommendations

We recommend:

- The ISAC strongly consider acting as the Sponsoring Council for this activity
- The SC charter a WG to carry out the scope of work recommended in this paper
- The resultant work be published as an HVP Standard
- The resultant work be jointly published in an academic journal

IV. References

Celli, J., Dalgleish, R., Vihinen, M., Taschner, P. E. M., & den Dunnen, J. T. (2012). Curating gene variant databases (LSDBs): toward a universal standard. *Human mutation*, *33*(2), 291–7. doi:10.1002/humu.21626

Claustres, M., Horaitis, O., Vanevski, M., & Cotton, R. G. H. (2002). Time for a unified system of mutation description and reporting: a review of locus-specific mutation databases. *Genome research*, *12*(5), 680–8. doi:10.1101/gr.217702

[1] See GDSDBAC03-2012: WG01: Disclaimer Statements on GDSDB Websites—Charter

Mitropoulou, C., Webb, A. J., Mitropoulos, K., Brookes, A. J., & Patrinos, G. P. (2010). Locus-specific database domain and data content analysis: evolution and content maturation toward clinical use. *Human mutation*, *31*(10), 1109–16. doi:10.1002/humu.21332

Scriver, C. R., Nowacki, P. M., & Lehväslaiho, H. (1999). Guidelines and recommendations for content, structure, and deployment of mutation databases. *Human mutation*, *13*(5), 344–50. doi:10.1002/(SICI)1098-1004(1999)13:5<344::AID-HUMU2>3.0.CO;2-U

Scriver, C. R., Nowacki, P. M., & Lehväslaiho, H. (2000). Guidelines and recommendations for content, structure, and deployment of mutation databases: II. Journey in progress. *Human mutation*, *15*(1), 13–5. doi:10.1002/(SICI)1098-1004(200001)15:1<13::AID-HUMU5>3.0.CO;2-Y

Vihinen, M., den Dunnen, J. T., Dalgleish, R., & Cotton, R. G. H. (2012). Guidelines for establishing locus specific databases. *Human mutation*, *33*(2), 298–305. doi:10.1002/humu.21646

Activity Proposal

Minimal content for gene variant databases (LSDBs)

Need

Quality of gene variant databases (LSDBs) is highly desired, and heavily discussed. To indicate their quality, curators of gene variant databases (LSDBs) would like to state that they follow the HVP/HGVS recommendations regarding data content and display. However, such recommendations do not yet exist.

Existing recommendations would also help gene variant database software developers, ensuring that their design includes all minimal requirements.

Scope

The "***Minimal content for gene variant databases (LSDBs)***" should cover at least;

- data fields to be included describing the variant(s) detected, method used, individuals analysed and associated phenotype.
- standards to follow for data entry & storage

It should be discussed whether the recommendations should be extended to include;

- minimal homepage description of the background and purpose of the database
- minimal requirements for back-ups and data safety
- minimal summary data that should be displayed
- minimal data displays that should be offered
- minimal query options that should be offered
- required copy right statements, databases policy document, etc.
- data curation
- additional recommended data fields

Since these recommendations are new, allowing existing databases to cope with the new demands, it seems appropriate to come to a set of "minimal requirements" and a set of "highly desired requirements". Databases can then choose to follow either the minimal or extended requirements.

Plan of action

Assign a group of experts (3-5) that write a draft of the document suggested under "Expected deliverables", give HGVS/HVP members a chance to comment on the proposal and finally complete the document and submit it for publication.

Resources required

Time, administrative support from HVP, a room to meet & discuss during an upcoming meeting (first within the group of experts, next with HGVS/HVP members).

Expected deliverables

A document describing the "***Minimal content for gene variant databases (LSDBs)***". Preferably the document will be published in a peer-reviewed scientific journal.

Recommendation

The "***Minimal content for gene variant databases (LSDBs)***" should become the (minimal) standard for gene variant databases and allow them to state that they follow the HVP recommendations.