Coordinating Office Position Paper

Ap06-2013: Diagnostic Grade VCF Specification ICO Comments

PP-ICO-10-2013

13 August, 2013

Human Variome Project International Ltd.
Level 5, 234 Queensberry Street
University of Melbourne VIC 3010
AUSTRALIA

Prepared by:
Timothy D. Smith

Contact:
Timothy D. Smith
Level 5, 234 Queensberry Street
University of Melbourne VIC 3010
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 1st of August, 2013, fro Human Variome Project Consortium member Graham Taylor. 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

There is a shift towards the utilisation of massively parallel sequencing (MPS) techniques in diagnostic and clinical contexts across the world. One of the key goals of the Human Variome Project is to capture the data generated during the course of this routine genetic testing and make it freely and openly available through a network of genetic data repositories so that it can be reused for clinical and research purposes. As the AP notes, MPS technologies regularly make use of the variant call format file specification to document the variants called during sequencing. It seems logical that VCF files will become a common tool in testing pipelines in labs and clinics over the next few years, and as such, will become (if it has not already) a de-facto standard for sharing data between laboratories, clinics and wider data sharing infrastructure. LOVD3 has already implemented a VCF import function to ingest variant.

However, the current VCF file specification, without extension, has severe limitations that impact its utility in the diagnostic and clinical context. The AP proposes an activity to address these limitations by extending the current VCF file specification to include information on the provenance and reliability of variant calls. The ICO believes that this is an idea worthy of the Project's attention, and we note that there is already active interest in this specific proposal and the idea more generally, from a number of groups and organisations. These include, but are not limited to, Illumina and their gVCF specification (https://sites.google.com/site/gvcftools/home/about-gvcf) and the Centers for Disease Control and Prevention's Next Generation Sequencing: Standardization of Clinical Testing (Nex-StoCT) Working Groups (http://www.cdc.gov/osels/lspppo/Genetic_Testing_Quality_Practices/Nex-StoCT.html). The ICO is also pleased to note that the HGVS has given over some time in their upcoming meeting at ASHG in October to discuss this issue.

Scope of Work

We recommend that an HVP Working Group (WG) be chartered, with the International Scientific Advisory Committee acting as SC, to develop a "diagnostic grade" VCF specification that will be interoperable with HVGS mutation nomenclature, HUGO gene nomenclature and compatible with mutation databases. Such a specification should be published as an HVP Standard.

The AP calls for the WG to "help to develop software tools and databases compliant with recommended standards," but in our opinion, this would entail a large amount of work that is too far-reaching to be achieved within the Human Variome Project's Standards Development Process.

We would recommend that the SC charter the WG to assess the current level of open-source software offerings for manipulating VCF files. The SC may wish to instruct the WG to assist developers, where possible, to implement support for the new specification. The level of such support should be decided by the SC. Of particular interest for the community would be for the WG to establish the availability of tools that could automate the process of the conversion of variants recorded in VCF files to HGVS nomenclature and vice-versa.

It should be noted that the VCF file specification is not the only specification for describing sequence variation and associated information. Other specifications, most notably, VarioML, currently under review as an HVP Recommended System, serves a similar function. The SC may wish to consider specifically asking the WG to take notice of these other file formats and ensure that any extensions proposed to the VCF file specification could be mirrored in the other specifications, to ensure cross-functionality.

The AP further calls for the Project to seek ISO or IEE recognition of the standard. While this is a lofty goal, we note that the time required to secure such recognition may extend well beyond the two-year timeframe for HVP Standard development. Our recommendation is that seeking such recognition is a more appropriate action for the SC to take once the HVP Standard or Guideline has been developed, rather than form part of the charter for the WG.

Possible Working Group Members

The ICO has no recommendations to make regarding possible WG members beyond a strong recommendation that all known existing groups working on similar endeavours be approached to collaborate, if not join the WG outright.

Available Assistance from the ICO

The ICO is able to provide the following assistance to the WG and the ensuing Interest Group:

- secretarial assistance;
- face-to-face meeting organisation and venues;
- phone and video conferencing facilities;
- an email list.

The AP submitter has indicated that resources are required for travel, meetings and pilot software development. While the ICO can assist the WG in securing venues for meetings during other large conferences (e.g. ASHG, ESHG, HUGO, etc), the ICO is unable to assist WG members with travel funding. Further, the ICO has no discretionary resources at this time to contribute towards pilot software development.[1]

III. Summary of Recommendations

We recommend:

- The ISAC act as Sponsoring Council for this activity

---

[1] If such an activity is specifically included in the charter of the Working Group

- The ISAC charter a WG to develop a "diagnostic grade" VCF specification that will be interoperable with HVGS mutation nomenclature, HUGO gene nomenclature and compatible with mutation databases
- The resultant work be published as an HVP Standard
- The WG should the current level of open-source software offerings for manipulating VCF files
- ISO or IEEE recognition of the resultant HVP Standard should only be sought by the SC, if they deem it appropriate, and not by the WG

Attachment A

Developing a consensus standard for the variant call file format that is fit for diagnostic use.

Case of need

The need is to develop an agreed file format based around the existing VCF standard that can be used for the population of diagnostic grade variant databases with minimal human intervention.

Human genetic variation is routinely documented in the clinic using HGVS nomenclature. Variation identified in genome-scale studies is recorded using the Variant Call File (VCF) format developed by the 1,000 genomes project. Both systems use the human genome reference sequence as their fundamental reference. Neither system logs the provenance of the variant calling, nor carries robust estimates of the reliability of the variant calls.

Scope of activity

Conduct a technical, scientific and clinical review of the current status of variant calling nomenclature in research and diagnostics. Describe an ideal variant calling format. Identify of gaps in current provision, make recommendations regarding format, software tools and database specifications.

Plan of action

Convene HPV Working Group. Prioritise target groups for participation, e.g. database leads, national professional bodies and colleges, HGVS, NCBI, CDC, EBI HVP statement of intent and open invitation to join the working group together with targeted invitations to priority participants. Introduce at the HGVS meeting in Boston 2013. Work towards a written review and recommendations over 18 month period with a view to publication in peer-reviewed journal. Assess and help to develop software tools and databases compliant with recommended standards. Seek ISO or IEEE recognition of the standard.

Resources

Travel, pilot software development and meeting budget

Deliverables

Consensus diagnostic VCF standard (dVCF) specification and publication
Demonstrator tools to construct dVCF from lab pipelines
Demonstrator database to upload and flag dVCF files

The final document should be intended to constitute an HVP guideline.