THE HUMAN VARIOME PROJECT

sharing data · reducing disease

an NGO Official Partner of UNESCO

MINUTES

Gene/Disease Specific Database Advisory Council
Tuesday 19 May, 2015
1200-1300hrs UTC/GMT

MEMBERS

Present
Peter E M Taschner (Chair)
Arleen Auerbach
Nenad Blau
Raymond Dalgleish
Mary Fujiwara
Daniel Hampshire
Yves Sabbagh
Mauno Vihinen

ICO Staff
Timothy D. Smith
Helen Robinson

Apologies
Olubunmi K D Abel
Ammar Al-Chalabi
Stefan Aretz
Timothy Barret
David Baux
Jean-Pierre Bayley
Daniel Bichet
Nancy Braverman
Paola Carrera
Johan T den Dunnen
Rosemary Ekong
Pascal Escher
Marc Ferre
Bruce Gottlieb
Tamas Hegedus
Raoul Hennekam
Alex Hewitt
Ammar Husami
Sarah E A Leigh
Derek Lim
Finlay Macrae
Eamonn Maher
Etienne Mornet
Magali Olivier
Sue M Povey
Judith Anne Savige
Sarah Sim
Carli Tops
Ronald Trent
Richard van Wijk
Katarzyna Wertheim-Tysarowska
Tom Winder
Bing Yu
Martina Witsch-Baumgartner
Johannes Zschocke

1. **Welcome**

2. **Apologies**

   Apologies were noted as above.

3. **Welcome new members**

   No new members since the last meeting.

4. **Confirmation of minutes of previous meeting**

   The minutes of the previous meeting were confirmed.

5. **Matters arising from the previous meeting**

   a. **Increase awareness of journals to nomenclature and data sharing**

      Peter Taschner reminded the Council members to submit examples of incorrect uses of sequence variant nomenclature to the ICO to assist with ongoing efforts to encourage journals to mandate nomenclature use in published articles.

b. Policy recommendations on variant terminology and exon numbering – Raymond Dalgleish

Raymond Dalgleish spoke to the document (see Appendix A) sent previously to the Council. He noted that no Council members had so far responded to the proposal in a negative fashion. The Council discussed the proposal and the following additions were suggested:

- The Variation Ontology also recommends the use of the term "variation" and this should be noted in the proposal;
- The language should be clarified to make it clear that using exon numbers is optional.

Raymond Dalgleish undertook to update the proposal and submit to the Council for the next meeting.

6. Report from Chair

Peter Taschner reported on outcomes from a meeting he attended in London with the Global Alliance for Genomics and Health (GA4GH) where some issues with the methods used to describe variants on Ensembl transcripts were discussed. The major issue discussed was that most transcripts are specific to a particular version of Ensembl and this can lead to problems when different groups are using different versions. Work is still ongoing on a solution.

Peter Taschner also reported on further discssusions related to the HVP and GA4GH he had with Peter Goodhand at the International Mutation Detection meeting in Leiden. Peter reminded the Council that the HVP had previously sent a document to the GA4GH asking them to endorse the HVP's leadership in the area of genetic variation databases. No response to date has been received. Peter Gooodhand reported that they were aware of the communication but were unsure of which Working Group within the GA4GH was best placed to move this forward. He also noted that if anyone is aware of any missing areas of work within the GA4GH's remit that they should contact them to start a task team to address it. This could be conducted as a shared activity between HVP and GA4GH.

Peter Taschner raised the issue of alternative nomenclature schemes being developed in response to perceived limitations in the existing nomenclature scheme and in spite of groups like the America College of Medical Genetics and Genomics specifying the use of the HGVS nomenclature in their guidelines. He suggested that this is occurring because people are either unaware of the HGVS nomenclature either in its entirety or are unaware of the extent to which it can describe sequence variant changes. He is in discussion with one GA4GH task team to give a presentation to them about the nomenclature. He further reported that he is aware of some individuals who have expressed concern that the HGVS nomenclature allows the same variant to be described at different levels or on different transcripts, which means that one variant can have multiple names. Raymond Dalgleish informed the Council of a paper that had recently appeared on BioRxiv that proposes a new nomenclature scheme. The Council discussed this paper in the broader context of alternative nomenclature schemes and determined that member should comment on the paper, as that is a feature of BioRxiv, and offer to work with the authors to suggest improvements to the existing nomenclature scheme via the HVP/HGVS/HUGO Joint Sequence Variant Nomenclature Committee or the ISCN, rather than propose a new scheme. Further, concern was raised that if journals are unaware of the existence of the HGVS nomenclature then they may begin to require authors use alternative nomenclature schemes.

Helen Robinson suggested that a reason why people are reticent to suggest improvements to the existing nomenclature scheme may be due to the perceived informalness and opaqueness of the Joint Sequence Variant Nomenclature Committee's processes. She suggested that they might consider linking with UNESCO to provide a neutral home for the committee and some semblance of authority.

Arleen Auerbach asked what the status of the LRG initiative was as it has been silent for a while. Raymond Dalgleish responded that LRGs are still being created in response to requests from the community. Arleen Auerbach further inquired if the GA4GH had taken a similar position to the HVP on the use of LRGs. Peter Taschner responded that they had not yet formed a view on how to handle reference sequences.

7. Working Group Reports

a. WG06: Disease & Phenotype Descriptions in Gene/Disease Specific Databases – Peter Robinson

Peter Robinson reported, through the Chair, that he is involved in three separate projects that are developing phenotype descriptions, as part of his work with IrDiRC, the GA4GH and the HVP. Work is still continuing but there will be more to report in the next month.

b. WG08: Ethics Checklist for Gene/Disease Specific Database Curators and Submitters

The report from WG08 was received from Rosemary Ekong (see Appendix B) and noted in her absence. Peter Taschner encouraged Council members to respond to Rosemary's call for assistance in the report.

8. Gene/Disease Specific Database Activities

Dan Hampshire reported that the EAHAD Coagulation Factor Variant Databases are currently moving to an LOVD3 platform and that he would like to be able to create custom columns but does not have 'manager' level access. Peter Taschner took the issue under advisement.

9. Recommendations to the Scientific Advisory Committee

Peter Taschner asked the Committee for assistance with contacting journals (see Item 5a).

10. Other matters

No other matters.

11. Next Meetings

- Tuesday 7 July 2015 1200hrs GMT/UTC
- Tuesday 8 September 2015 1200hrs GMT/UTC
- Tuesday 10 November 2015 1200hrs GMT/UTC

Appendix A

HVP policy recommendations on variant terminology and exon numbering

Introduction:

Discussions within HVP and with other external parties have highlighted the need for HVP policies with respect to use of the word "variant" (and its many alternatives) and the issue of exon numbering. This document is intended as the basis for discussion within the Gene/Disease Specific Database Advisory Council of the HVP.

Variant terminology:

The terms variant, mutation and polymorphism are commonly used interchangeably, usually without proper regard for the context in which they are used. This leads to possible misunderstanding when the causation of disease is being discussed. Mutation is the process by which a change occurs in DNA and the term is widely used in the context of describing sequence changes in specific genes which are known to harbour disease-causing changes that result in well recognised heritable disorders, e.g. changes in the *HBB* gene resulting in thalassaemia.

The terms variant and polymorphism are more commonly used in the context of association studies, genome-wide or otherwise, where sequence variation might be demonstrated to make a small or large contribution to the heritability of a complex disorder. The term polymorphism has the added drawback that it is frequently defined as applying to sequence variants with a population frequency greater than 0.01.

These issues have been discussed in the literature [Cutting, 2014; Richards et al., 2015] and in both instances use of the term "variant" is recommended to describe a change to a DNA sequence relative to the reference genome sequence.

It is proposed that HVP should endorse the exclusive use of the term "variant" to describe all sequences changes irrespective of their contribution to phenotype. Mutation may be used in the correct sense of the word to describe the process by which variants arise. Use of the term polymorphism is deprecated.

Exon numbering:

Confusion often occurs when the exon or intron that harbours a sequence variant is discussed. The basis of this confusion lies in the fact that there is no single standard endorsed by any organisation or scientific body with respect to how exons should be numbered. In addition, there are many historical (legacy) gene-specific exon numbering, or naming, schemes. In some cases, historic exon designations may be non-numeric, thus precluding simple linear numbering as has been discussed by Dalgleish et al. [2010]. There is the added complication that, in genes with several mRNA splice variants, a variant in a particular exon of one mRNA variant might lie in an intron with respect to another mRNA variant.

Some would argue that it is unnecessary to specify the location of variants with respect to exons or introns (den Dunnen JT, personal communication) and there is indeed some merit in that view. However, the practice is unlikely to disappear as the habit of referring to exons when describing variants is well entrenched.

The coordinates of exon boundaries are annotated in GenBank RefSeqGene records and in LRG records as a standard feature. Both record types also allow for alternative numbering/naming schemes to exist in parallel with sequential numbering of exons. These alternative schemes are developed in consultation with genetics community stakeholders with expertise in the genes in question. The *COL1A1* gene RefSeqGene record

Appendix A

(NG_007400.1) and its equivalent LRG record (LRG_1) both note that there is an alternative numbering scheme in which the thirty-third exon is designated 33/34 and the remaining exons then number sequentially from 35 to 52.

It is proposed that HVP should endorse that variants may be described in terms of their location with respect to numbered or named exons. When locations are expressed in this way, the exon naming/numbering scheme must be described with respect to a default or alternative scheme that is described within the reference sequence for the gene in question.

Cutting GR. 2014. Annotating DNA variants is the next major goal for human genetics. Am J Hum Genet 94:5-10.

Dalgleish R, Flicek P, Cunningham F, Astashyn A, Tully RE, Proctor G, Chen Y, McLaren WM, Larsson P, Vaughan BW, Béroud C, Dobson G et al. 2010. Locus Reference Genomic sequences: an improved basis for describing human DNA variants. Genome Med 2:24.

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 10.1038/gim.2015.30 [Epub ahead of print].

Appendix B

Report from WG08: Ethics checklist for database curators

- 177 invitations were sent out for participation in the survey of curators.
- 27 responses were obtained at the close of the survey in February 2015.
- An initial overview of survey responses indicates that most curators who responded have applied some of the guidelines. A comprehensive analysis is yet to be concluded.
- Dr Mats Hansson was contacted (on the recommendation of Dr Peter Taschner) about the possibility of participating in WG08. He mentioned his wish to be involved, but also indicated that he could not do so due to his current work load.
- An intention to hold a meeting of WG08 in April did not proceed due to time constraints. One will be organised before the end of June.
- The 3rd member of WG08 (Dr Sarah Schlesinger) has not been able to participate due to personal reasons.
- It should be noted that WG08 effectively has 2 active members (Rosemary Ekong and Mauno Vihinen).
- The Working Group has not received any ethical issue scenarios from any member since the last meeting.