THE HUMAN VARIOME PROJECT

HVP/DA/003-01/EN

2015-12-11

G/DSDBAC

DRAFT FOR APPROVAL

VARIANT TERMINOLOGY AND EXON NUMBERING

Notice

This document is not an HVP Standard or Guideline. This document is a Draft for Approval that has been distributed for community consultation and review. This document is subject to change without notice. **USE AT YOUR OWN RISK!** Because this is an unapproved draft, this document must not be utilized for any conformance/compliance purposes.

Authors

Raymond Dalgleish,
Mauno Vihinen

Editor

Timothy D. Smith

Published by:
Human Variome Project International Limited
Level 1, 200 Berkeley Street, The University of Melbourne VIC 3010, AUSTRALIA

Published 2015-12-11. Printed in Australia.

PDF: ISBN <<PDF_ISBN>>
Print: ISBN <<PRINT_ISBN>>

Contents

Foreword

The Human Variome Project is an international consortium of researchers, policy makers and healthcare professionals committed to the free and open collection, curation, interpretation and sharing of genomic knowledge.

The Human Variome Project Consortium envisions a world where the availability of and access to genetic variation information is not an impediment to diagnosis and treatment; where the burden of genetic disease on the human population is significantly decreased; and where the sharing of genetic variation information is standard clinical practice.

To facilitate worldwide and interoperable sharing of genomic knowledge, the Human Variome Project Consortium produces Standards and Guidelines. HVP Standards are those systems, procedures and technologies that the Human Variome Project Consortium has determined shall be used by the community. These carry more weight than the less prescriptive HVP Guidelines, which cover those systems, procedures and technologies that the Human Variome Project Consortium has determined would be beneficial for the community to adopt.

HVP Standards and Guidelines are central to supporting the work of the Human Variome Project Consortium and cover a wide range of fields and disciplines, from ethics to nomenclature, data transfer protocols to collection protocols for clinical data. They can be thought of as both technical manuals and scientific documents, and while the impact of HVP Standards and Guidelines differ, they are both generated in a similar fashion.

HVP Standards and Guidelines make the collection, curation and sharing of information more efficient and reliable by establishing consistent protocols that can be universally understood. They facilitate interconnection of and interoperability between different systems.

HVP Standards and Guidelines represent a consensus of the Human Variome Project Consortium, each member of which has had the opportunity to participate in the development and review of each standard and guideline. In addition, as every effort is made to include all interests in the activity, HVP Standards and Guidelines can be considered to be representative of all interests concerned within the scope of each Standard or Guideline.

The Human Variome Project defines consensus as significant agreement between all affected parties covered by the scope of the standard or guideline. Consensus requires that all views and objections be considered, and that a concerted effort be made toward their resolution.

More information on the Human Variome Project is available at the Project's website (http://www.humanvariomeproject.org/). Procedures for the development of HVP Standards and Guidelines can be found in *PD06-2011: Standards Development Process*, available at http://short.variome.org/PD06-2011.

This Document

This document has been prepared the Human Variome Project Gene/Disease Specific Database Advisory Council.

An Exposure Draft of this Document was released to the Human Variome Project Consortium on 2015-09-15.

Important Notice

HVP Standards and Guidelines are not intended to replace or substitute for any applicable legislation or regulation in any jurisdiction, or any institutional policy or funding agreement that a genetic variation information resource is operating under. Implementers of HVP Standards and Guidelines are responsible for determining and complying with all appropriate ethical and cultural protection practices and all applicable laws, regulations, policies and agreements.

Introduction

Discussions within the Human Variome Project and with other external parties have highlighted the need for guidelines with respect to use of the word "variant" (and its several alternatives) and the issue of exon numbering.

1 Scope

This document details the terminology that should be employed to refer to the concept of changes in the sequence of DNA in English text. It further details the manner in which exons should be numbered, if such numbering is necessary.

This document does not specify how specific changes in the sequence of DNA should be named or described.

2 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 to the sequence of 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 limitation 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; Vihinen, 2015) and in all instances use of the term "variant" is recommended to describe a change to a DNA sequence relative to the reference genome sequence. Further, the Variation Ontology (VariO), a systematic framework for description of variation effects, consequences and mechanisms utilises the term "variation" (Vihinen, 2014).

The term "variant" should be used 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.

3 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 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 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 (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.

Variants may be described in terms of their location relative to numbered or named exons, but doing so is optional. It is acknowledged that some variants will be better suited than others for their location to be described in this way.

When locations are so expressed, the exon naming/numbering scheme must be described with respect to a default or alternative scheme that is described within the reference sequence record for the gene or transcript in question.

It is further acknowledged that particular variants might lie in different exons in different transcripts of the same gene. Further, such variants might be exonic with respect to one transcript and intronic with respect to another. Absolute clarity is required in the description of the location of variants relative to exons.

4 Bibliography

Cutting GR. 2014. Annotating DNA variants is the next major goal for human genetics. Am J Hum Genet 94(1):5-10.

Dalgleish R, Flicek P, Cunningham F, Astashyn A, Tully RE, Proctor G, Chen Y, McLaren WM, Larsson P, Vaughan BW and others. 2010. Locus Reference Genomic sequences: an improved basis for describing human DNA variants. Genome Medicine 2(4):24.

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E and others. 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 17(5):405-24.

Vihinen M. 2014. Variation Ontology for annotation of variation effects and mechanisms. Genome Res 24(2):356-64.

Vihinen M. 2015. Muddled genetic terms miss and mess the message. Trends in Genetics 31(8):423-425.