LUND UNIVERSITY
Faculty of Medicine

25.4.2014

Human Variome Project
Activity Proposal Submission

Department of Experimental Medical Science
Professor Mauno Vihinen

Variation Ontology, HVP Recommended system application

I hereby apply Recommend system status for Variation Ontology (VariO), which is a systematic approach for describing and annotating in databases effects, consequences and mechanisms of variations.

Please find attached a concise description of the ontology.

Yours sincerely,

Mauno Vihinen
Professor of Medical Structural Biology
Lund University

Postal address	*Visiting address*	*Telephone*	*Fax*	*E-mail*	*Website*
BMC D11 SE-22 184 Lund Sweden	Sölvegatan 19 Lund	+46 72 526 0022	+46 46 2228439	mauno.vihinen@med.lu.se	http://structure.bmc.lu.se

VariO and its features

VariO was developed for annotation of effects, consequences and mechanisms of variations. The major levels for description are the three molecules, namely DNA, RNA and protein. They are further divided to four major sublevels: *variation type*, *function*, *structure*, and *property*, and further up to eight sublevels. VariO annotation summarizes existing knowledge about a variation and its effects and formalizes it so that computational analyses are efficient. The annotations should be made on as many levels as possible. VariO annotations are made in reference to normal states, which vary for each data item including e.g. reference sequences, wild type properties, and activities.

The terms have a clear hierarchy and the organization of terms for DNA, RNA and protein has a similar layout whenever appropriate. The terms for the three molecular levels are consistent and related terms are used for related features at different levels. The hierarchy of the terms has been designed to allow for a versatile and flexible annotation.

VariO is intended for the description of all kind of variations and situations. Terms with genetic origin describe changes either in DNA or inherited from it to RNA and protein levels. The non-genetic terms, called variations emerging at RNA or protein level, are for either biological or artificial modifications that originate at RNA or protein level.

To keep the ontology compact, modifier attributes form the fourth major level. These terms are used to modify the meanings of terms at the other levels. For example, quantity terms are used to modify other terms when the effect is increased, decreased or missing quantity of the parameter, or when it is not changed. Instead of having separate terms for describing increase or decrease existing terms can be modified with attribute terms.

VariO terms should be combined with other systematics and ontologies. The Evidence Ontology (http://www.evidenceontology.org/, ECO) terms are used to describe the methods with which the annotations were obtained.

VariO aims at describing any effect, consequence and mechanism, at any organism. The variations can be of genetic or non-genetic origin. The size of the variation does not matter, anything ranging from nucleotide or amino acid changes to chromosome or genome duplications can be annotated. However, the annotations are position based, even if the position means e.g. an entire chromosome.

An annotation tool is available to help curators to annotate their databases easily as well as to provide related ECO evidence details for VariO annotations. The current version of the annotation tool, called VariOtator and available on the website, provides variation type annotations on all three molecular levels based on HGVS nomenclature which can be generated by Mutalyzer tool. The tool can at the moment make also function and structure annotations and the property features will follow.

Databases annotated with VariO terms will allow interoperability, collection and analysis of cases simultaneously from a single or multiple databases with simple or even very complex queries. There are several applications, including e.g. generation of benchmark datasets for different effects and mechanisms to test the performance of prediction tools, search for certain types of variations over several genes or proteins, and the integration of heterogeneous information sources for extensive analysis and interpretation of variations and their effects. VariO is most suitable for annotation of effects of variations in LSDBs, central variation and sequence databases. It can equally well be applied to specific variation databases whether devoted to diseases, mechanisms or effects.

Once databases have VariO annotations available queries can be made within and cross databases to find cases of interest such as variation types in certain protein structural elements, variants with certain

kind of functional effects, or for example disease-causing variants in protein-protein interaction interfaces. The possibilities are limitless as exemplified by GO annotations which have already been used in 5000 publications in numerous different ways.

The extent to which the system is currently being utilized

VariO is a new ontology. There are currently a number of user who are either in the process of starting to utilize VariO or who are seriously considering it.

The first user is VariBench, variation dataset database (http://structure.bmc.lu.se/VariBench/).

It will be implemented into over 130 databases at Lund University including IDbases (http://structure.bmc.lu.se/idbase/), KinMutBase (http://structure.bmc.lu.se/idbase/KinMutBase/) and SH2base (http://structure.bmc.lu.se/idbase/SH2base/). The annotation is on-going and will be released along with several other improvements to the databases.

We have possibility to generate automatically *variation type* annotations for entries in all LOVD databases.

VariOtator annotation tool works together with Mutalyzer (https://mutalyzer.nl/), variation naming tool.

We have agreement with ProTherm (http://www.abren.net/protherm/) and ProNIT (http://www.abren.net/pronit/pronit.php). These are databases thermodynamic properties of proteins and protein nucleic acid interactions).

We are principal agreement with FINDbase (http://www.abren.net/pronit/pronit.php) variation allele database.

Negotiations are on-going with CafeVariome (http://www.cafevariome.org/).

In addition we are discussing with some major databases, but are not allowed to reveal their identity.

The HVP Standards and Guidelines the system is compliant with

VariO is compliant with HGVS nomenclature and in fact works together with Mutalyzer to generate *variation type* annotations.

The access and license terms and conditions under which the system is made available

VariO is freely available. It is distributed under Open Bioinformatics Ontologies (OBO) OBOFoundry principles http://www.obofoundry.org/wiki/index.php/OBO_Foundry_Principles.

VariO publications

Vihinen, M. (2014) Variation Ontology for annotation of variation effects and mechanisms. *Genome Res.* **24**, 356-364.
Vihinen, M. (2014) Variation Ontology: Annotator guide. *J. Biomed. Semantics* **5**:9.