Scope and access
The Variation Reporter and its companion API are designed to expedite access to information about sequence variations at one or thousands of locations on a genome that may be available from various NCBI variation resources. It can be used to map and/or interpret variations identified in single individuals or genes, or variation calls produced by large sequencing projects. Variation Reporter generates a report for each submission. The report includes dbSNP and dbVar IDs when the submitted data match variants known to these resources. When the data are unknown (novel), the report includes a prediction of the molecular consequence of the variant, based on the NCBI annotation data for that target genome assembly.

The workflow of Variation Reporter is comprised of the following steps: selecting the source assembly, uploading user data, validating the uploaded data, submitting the mapping request, and displaying the returned results. Up to the first 100 entries from the input list will be displayed with the full results downloadable through a link in the results page.

Uploading input data
Available organism and genome assemblies are listed on the Variation Reporter startup page and this must be set to the assembly from which the input data was derived. Clicking the “+” button opens the data upload form, where input data can be pasted in or uploaded through the “Browse” button. Variation Report supports input in BED, HGVS, GVF and VCF formats. The example shown is in the preferred HGVS format. Clicking the “upload” button submits the dataset for format validation. A successfully validated dataset is added to the “Available data” list, which can be selected for submission using the “Submit for analysis” button.
The results page

A submitted request retrieves an intermediate page, which regularly updates the job status (A). The analysis results will be displayed once it is completed.

In the results page, the summary of input data and the analysis output are given in the upper left (B). Genome assembly used in the analysis is shown in the upper right (C).

Sample output for the first 100 entries found is presented as a table below (D). The complete table can be downloaded by clicking the "Download Report" diskette icon (E). The results table contains a lot of information. Before getting into the detailed content, one technical characteristic is worth mentioning up front: the results table contains one row for each transcript to which an input variant maps. If a given input variant maps to 3 transcripts, there will be three rows (F).

To correlate analysis results with input data, the result table lists the input local ID, chromosome position and alleles in the first three columns (G) with the following adjustment: the chromosomal coordinates are prefixed with the chromosomal record's accession version and hyperlinked to the display in the graphical sequence viewer (SV) display (detailed description on p. 3). Submitted alleles are converted to HGVS format.

rsID numbers matched by the input entry are listed in the NCBI id column hyperlinked to the actual full SNP report, followed by variation origin and minor allele frequency (MAF) calculated from the 1000Genomes genotype data (H). For variants with available data, the Clinical Information column contains links to popups listing corresponding entries in the ClinVar database (I). Some may also contain links to specific PubMed abstracts to provide additional details (J). Mappings to RNA transcripts and variation types are listed in the last two columns (K). Entries in the molecular Consequences column link to corresponding records from Sequence Ontology Project.
The results page (cont.)

A graphical sequence viewer display panel is shown below the summary table, which allows the interactive examination of the first mapped variant under the genomic context. This display is connected to the data table above. The default display shows the first entry; clicking a different variant in the table updates the graphical display to show that entry. In this display, the genomic position is highlighted with a marker. Since the entry shown matches an existing variation record, the rsID is used as the marker name (A). Information available from this display includes:

- The summary of mapped RefSNP entry with links to the actual record, through mousing over the red tick mark (B)
- The gene information with links to the entry in NCBI Gene database, through mousing over the gene bar (C)
- Details for the allele and its flanking sequences in relevant RefSeq records, through the marker context menu (D)
- Information from other mapped features, through the “Tracks” dialog box (E)
Specifying coordinates using HGVS for known or novel variations

Variation Reporter service takes input coordinates specified by various reference sequence-based HGVS names, such as NC_000001.10:g.17354297A>G or NM_003000.2:c.487T>C. It is necessary to point out that this tool is capable of handling coordinates that do not map to any existing variations deposited in dbSNP. This, plus the batch search function, makes Variation Reporter a great tool that complements variation data processing.

The mapping result below is from a hypothetical GRCh37p13-based HGVS input of NC_000008.10:g.19819723T>G and NC_000008.10:g.19819725A>G, both derived from rs328 for demonstration purposes. For the first variation, the highlighted rsID indicates that the input shares the same position with an existing variation rs75992339, but they differ in allele as explained by the tooltip (A). The second variation is a novel entry altogether since the “NCBI Id” is empty (B). The Transcript Allele and the Consequences (C) columns are populated to provide information in the functional evaluation of this variation.

The SV display (D) below the result table visually sums up the placement of the variation selected (the second novel entry). The display highlights the variation by an added marker (E) and places it in the context of annotated features. This variation is mapped within an exon (green bar) of the LPL gene (in the popup generated by hovering, F), next to an existing variation rs328 (G) with known clinical relevance and literature citations (H).

API access

For users with large datasets to analyze, the API access to Variation Reporter offers an alternative approach, which can be integrated into the workflow of a project. For users’ convenience, Perl and Java sample code is available online. Refer to the following page for usage guidelines, example code, and other descriptions of the service:


Relevant links

Youtube video demonstration: https://www.youtube.com/watch?v=tqe2qkynaAo
Homepage for dbSNP: https://www.ncbi.nlm.nih.gov/snp/
Homepage for dbVar: https://www.ncbi.nlm.nih.gov/dbvar/