Overview

Accurate and timely interpretation of genetic testing results is critical to translating genomics to clinical care. ClinVar supports the medical genetics community as a freely available, public archive of the relationships between medically important variants and phenotypes. It allows testing laboratories access to a broader set of clinical interpretations than they may have collected on their own, and the ClinVar data can be incorporated into their daily workflow. ClinVar is also available to individual users and organizations that want to incorporate it into their own applications. Data providers submit observed variants and make an assertion about the clinical significance of each variant with respect to a phenotype. The interpretation may be based on clinical testing, research, or the literature. Various types of evidence may also be provided to support the assertion. We continue to work closely with several genetic testing labs and other end users to refine the submission process and display of the data to maximize its utility for the clinical genetics community.

ClinVar adds value to submissions in several ways. Submissions for the same variant and phenotype pair from different submitters are aggregated, so that agreement or conflict in clinical significance is clear and evidence from different submitters can be viewed together. Accession numbers are assigned to individual submissions (SCV) and to aggregate records (RCV) to facilitate retrieval; version numbers allow tracking of updates to each record as submitters refine clinical interpretations over time. Both SCV and RCV records are given a review status which allows the user to evaluate the validity of each interpretation. ClinVar supports standardized descriptions of both variant and phenotype, by providing HGVS expressions at the genomic, cDNA, and protein level and phenotype terms reported in MedGen. The molecular consequence is predicted for variants within a coding region based on the effect of the sequence change on translation, and for others variants in a gene by reporting their location (UTR, splice site). Curation by NCBI staff may also add published allele names, citations, and links to the same variant in other databases.

Although ClinVar provides aggregation, standardization, and a central repository, the database is driven by submission of data from the clinical genetics community. ClinVar provides limited curation of variant and phenotype terms, but clinical interpretations are provided by submitters.

Data Submission

ClinVar welcomes submissions from clinical testing labs, research labs, locus-specific databases, expert panels and professional societies. The database has a flexible data model, so submissions may be minimal or very detailed. Two Excel spreadsheet templates (A) are available for simpler submissions, one aimed at more general variant-level data and one designed for more specific case-level data. More detailed submissions are submitted by XML; the xsd is available on the ftp site and a Data Dictionary (B), which defines data elements in ClinVar, is available from the home page. The minimal data required for submission is submitter information, type of study, a valid variant description (either HGVS or cytogenetic description) and a clinical assertion and/or phenotype. Utility may be increased further by providing supporting evidence, such as number of observations of the variant, allele frequency, co-segregations, and mode of inheritance for variant level data or affected status, presence of family history, bi-allelic variant occurrence, and ethnicity for case level data.
Data Access
For bulk download and analysis, the data is available on ClinVar’s ftp site (ftp://ftp.ncbi.nlm.nih.gov/pub/clinvar/) as VCF and XML.

The ClinVar website provides access to individual records. Users can search (A) for variants, phenotypes, genes, proteins, MIM numbers, variant IDs such as dbSNP or OMIM allelic variant IDs, molecular consequence as well as other data fields. Filters (B) can be used to restrict the results by gene, molecular consequence, clinical significance, variation type, review status, and method of review.

Search results provide variants reported to ClinVar. Each variation report includes general information about the allele(s) (C). Clinical significance, conditions reported for the variant, and a link to ClinVar’s aggregate (RCV) record (D), as well as genes affected (E) are given in the right hand column.

The variant report includes tabs to display the details of clinical assertions (F) and the evidence (G) for each individually submitted record (SCV).