dbVar: A Genomic Structural Variation Database

A collection of human genomic structural variation

https://www.ncbi.nlm.nih.gov/dbvar/

National Center for Biotechnology Information • National Library of Medicine • National Institutes of Health • Department of Health and Human Services

Scope and Access

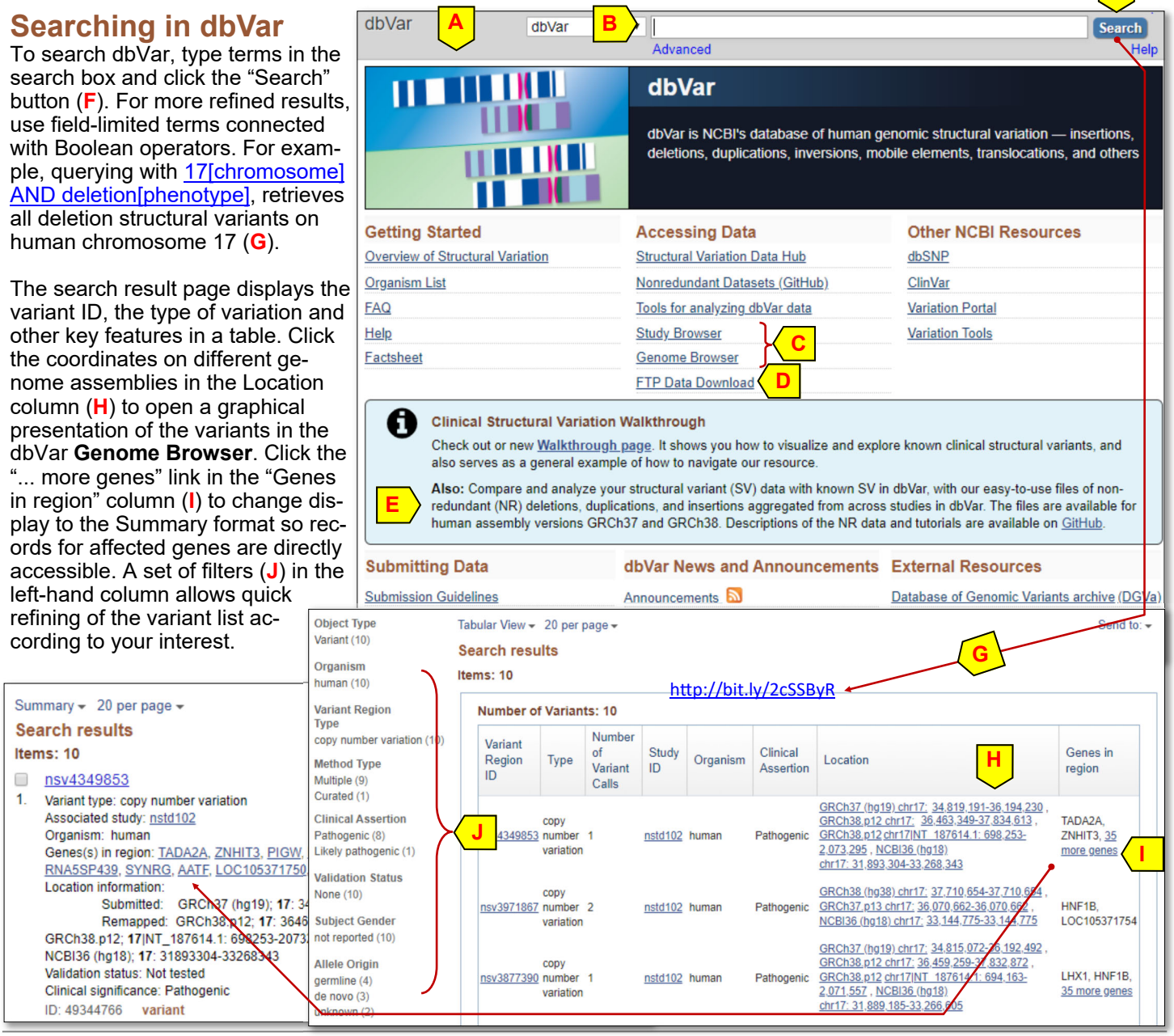
The NCBI dbVar Structural Variation database houses human genomic structural variants (SV) greater than 50 base pairs in length. From the dbVar homepage (**A**, www.ncbi.nlm.nih.gov/dbvar), you can search (**B**), browse (**C**), view and download variant data from over 150 studies, such as 1000 Genomes Phase 3 (estd219), Genome in a Bottle (nstd175), Clinical Structural Variants (nstd102), gnomAD (nstd166), etc. You can access the variants using the Study Browser or the graphical Genome Browser. Individual Study and Variant Pages include links to the raw data as well as to related information at other NCBI and external resources. Bulk data downloads are available by FTP at ftp.ncbi.nlm.nih.gov/pub/dbVar/data (**D**).

In 2018, dbVar introduced a new comprehensive set of non-redundant structural variants (NR set) consisting of unique insertions, duplications, and deletions. These compact files are suitable for use as references in the analysis of human structural variation. For example, you can use the dbVar NR set to filter and annotate other datasets in a broad range of applications including variant discovery and identifying rare and/or clinical variants. The dbVar NR set currently includes more than 2.5 million deletions, 1.3 million insertions, and 400 thousand duplications. The NR set is updated monthly as new variants are added to dbVar. You can find more information about the NR set, brief tutorials, and ways to access NR set FTP files from github at github.com/ncbi/dbvar/tree/master/Structural_Variant_Sets (**E**).

Searching in dbVar

To search dbVar, type terms in the search box and click the "Search" button (**F**). For more refined results, use field-limited terms connected with Boolean operators. For example, querying with 17[chromosome] AND deletion[phenotype], retrieves all deletion structural variants on human chromosome 17 (**G**).

The search result page displays the variant ID, the type of variation and other key features in a table. Click the coordinates on different genome assemblies in the Location column (**H**) to open a graphical presentation of the variants in the dbVar **Genome Browser**. Click the "... more genes" link in the "Genes in region" column (**I**) to change display to the Summary format so records for affected genes are directly accessible. A set of filters (**J**) in the left-hand column allows quick refining of the variant list according to your interest.

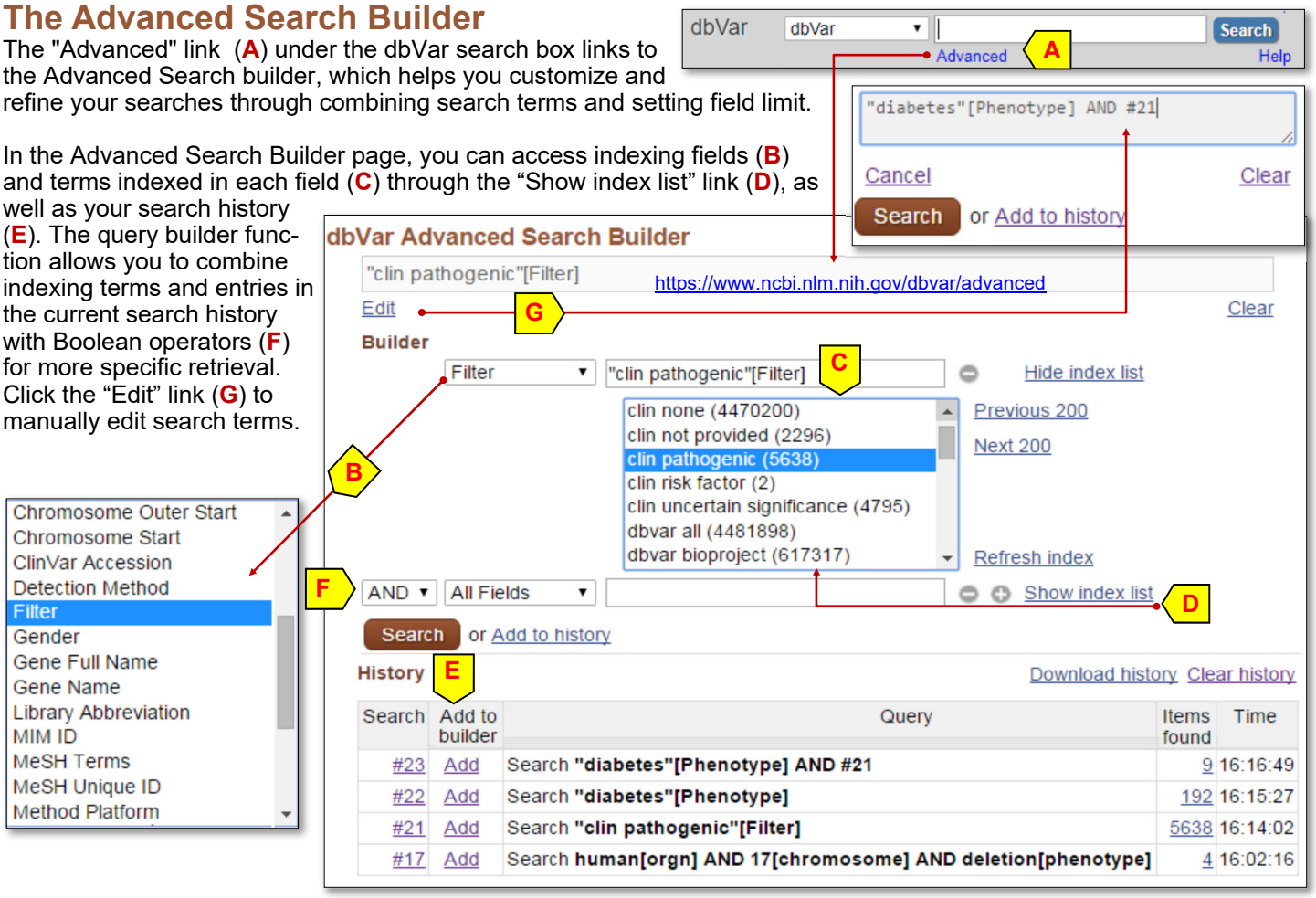


Contact: info@ncbi.nlm.nih.gov

The Advanced Search Builder

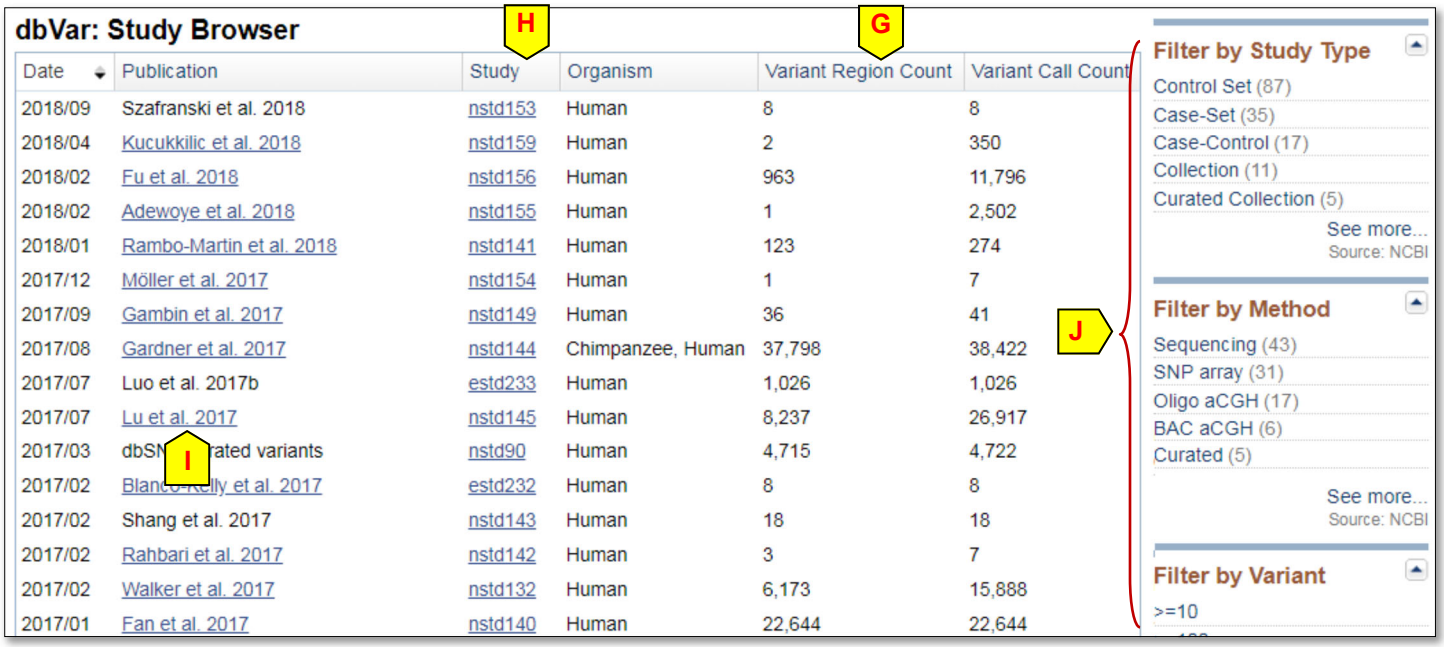
The "Advanced" link (**A**) under the dbVar search box links to the Advanced Search builder, which helps you customize and refine your searches through combining search terms and setting field limit.

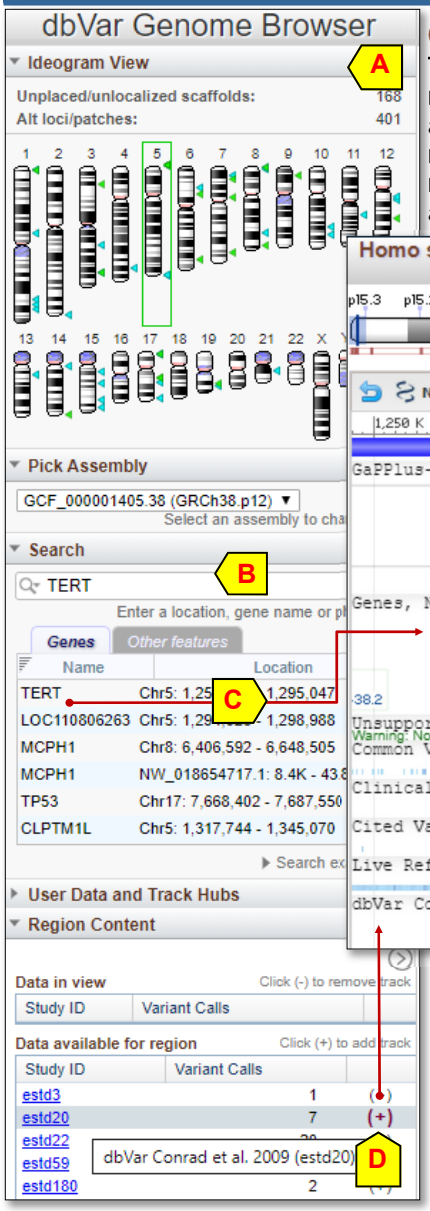
In the Advanced Search Builder page, you can access indexing fields (**B**) and terms indexed in each field (**C**) through the "Show index list" link (**D**), as well as your search history (**E**). The query builder function allows you to combine indexing terms and entries in the current search history with Boolean operators (**F**) for more specific retrieval. Click the "Edit" link (**G**) to manually edit search terms.



dbVar Study Browser

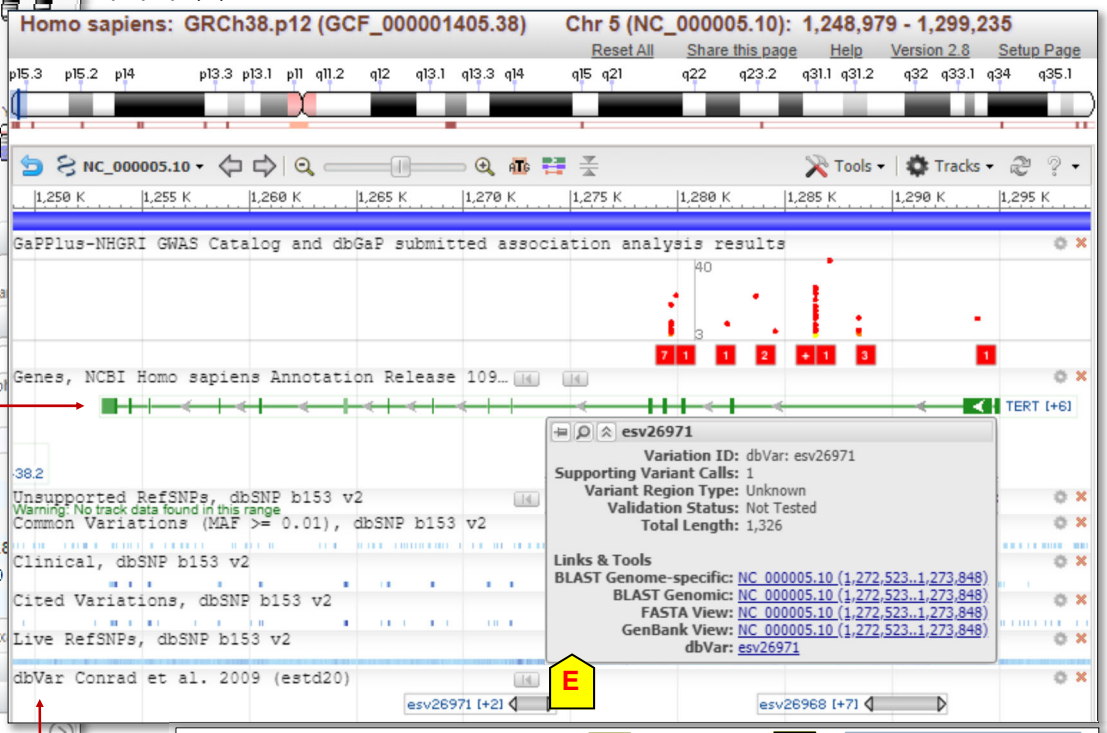
The dbVar homepage provides links to the Study Browser page, where you can browse by study IDs. The Study Browser table (below) sums up the available studies in a table. Clicking a column header, such as the "Variant Region Count" (**G**), sorts the list according to the value in that column so you can locate studies with certain characteristics. Identifiers for individual studies in the Study column (**H**) links to detailed display of specific studies, while the citations in the Publication column (**I**) link to relevant articles in PubMed. You can use criteria filters in the right-hand column (**J**) to narrow the list of studies displayed.

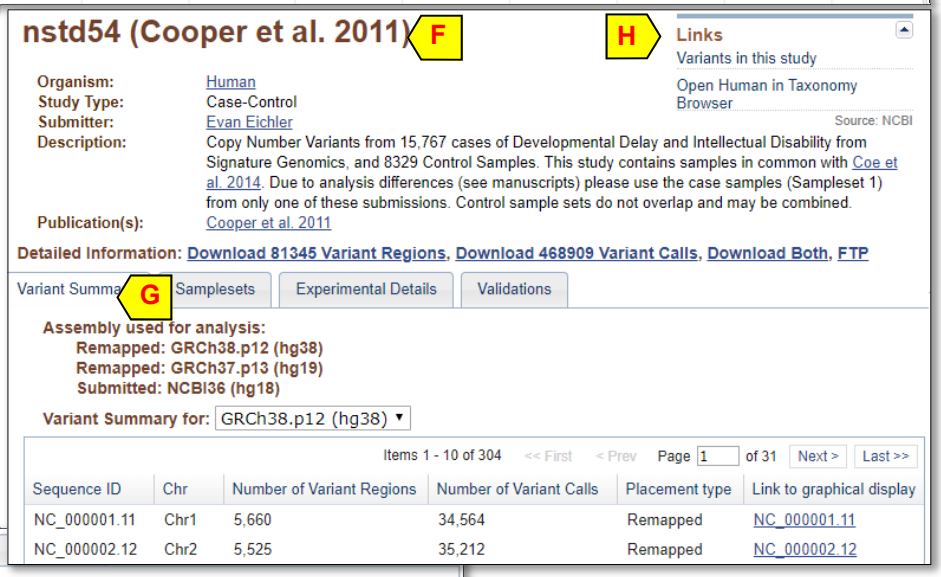




dbVar Genome Browser

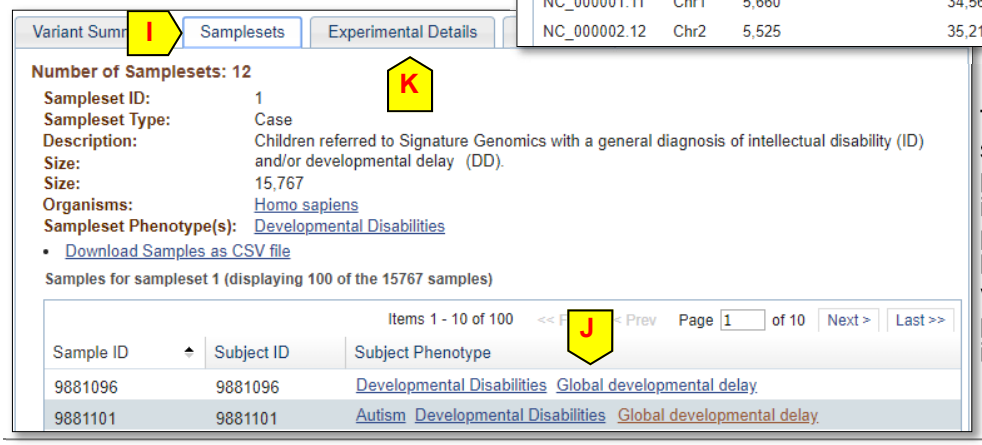
The Genome Browser (**A**) provides a way to locate reported variants for genomic regions of interest. You can specify a genomic region by using either coordinates or annotated features known to occur in these regions, such as gene names (**B**). Click a retrieved entry from the list to zoom to the feature (**C**), click the (+) sign (**D**) to the right of studies to add tracks for mapped variants. Hover a variant to see its detail in a popup (**E**).





dbVar Study Pages

The details of a specific study are shown in the Study Page (**F**). The default "Variant Summary" tab (**G**) lists variants mapped to different chromosomes. Variants identified by the study are available through the "Variant in this study" link (**H**).

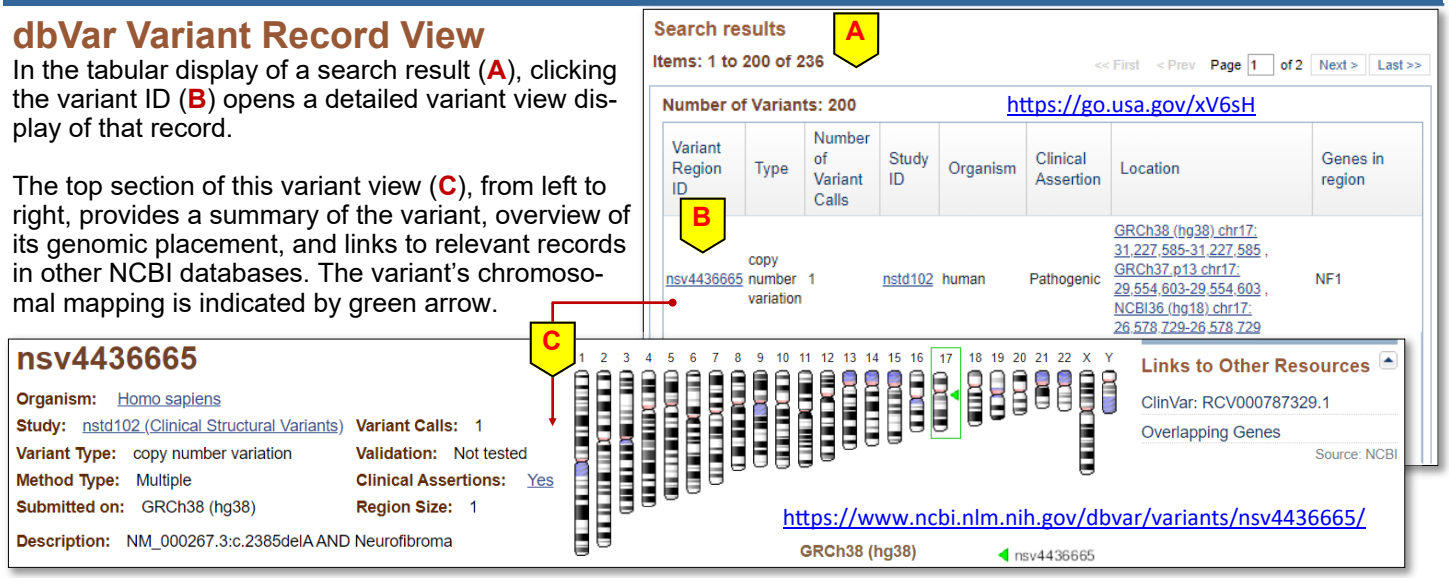


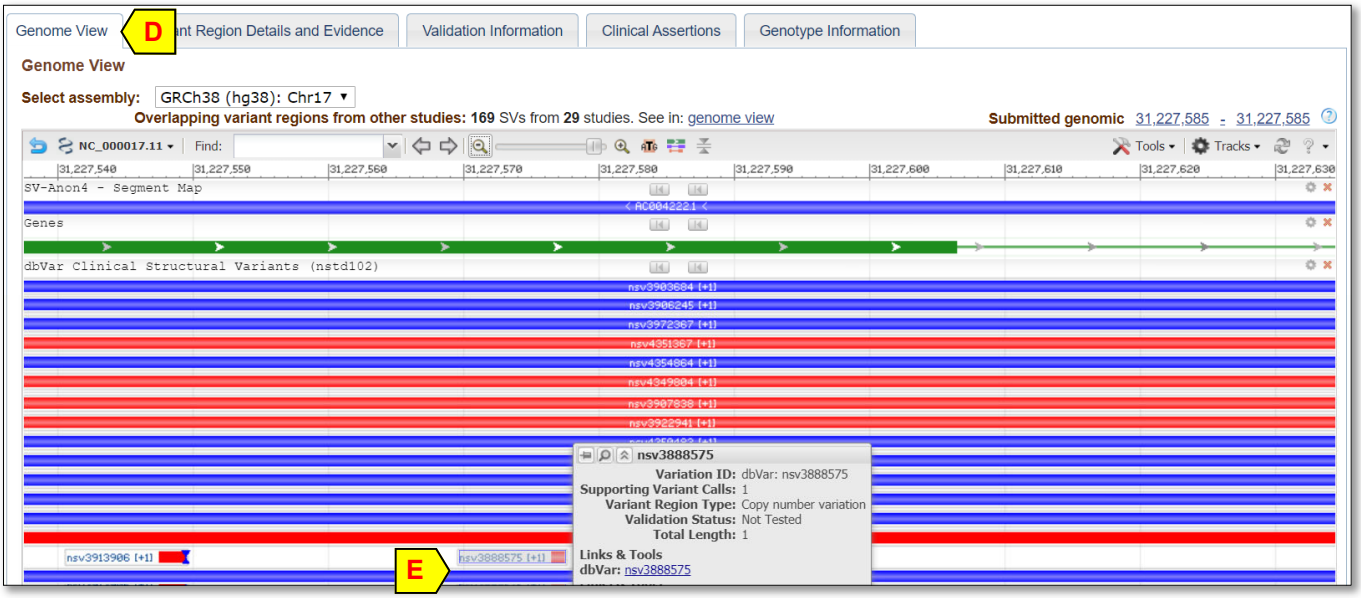
The "Samplesets" tab (**I**) displays samples used in the study. Information provided includes the IDs of the participating subject and their demonstrated phenotypes (**J**). The "Experimental Details" tab (**K**, details not shown) provides methodologies and experimental platforms used when such information is made available by the submitters.

dbVar Variant Record View

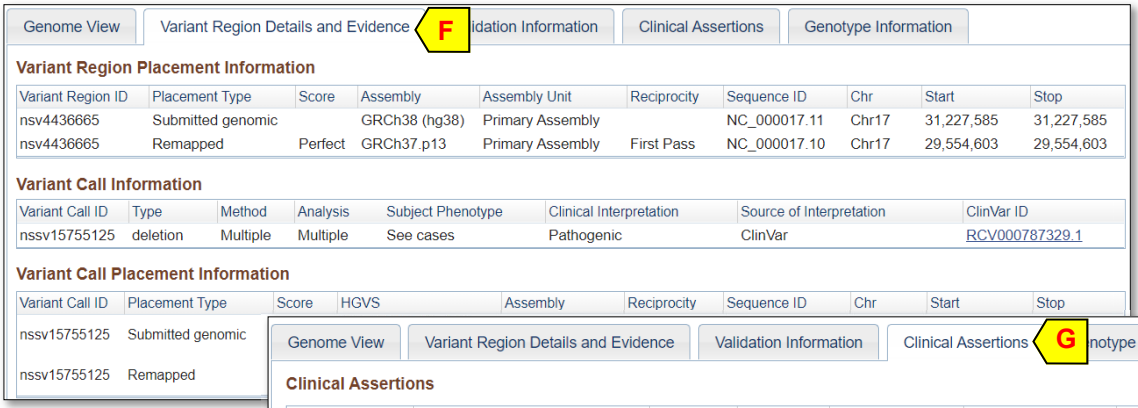
In the tabular display of a search result (**A**), clicking the variant ID (**B**) opens a detailed variant view display of that record.

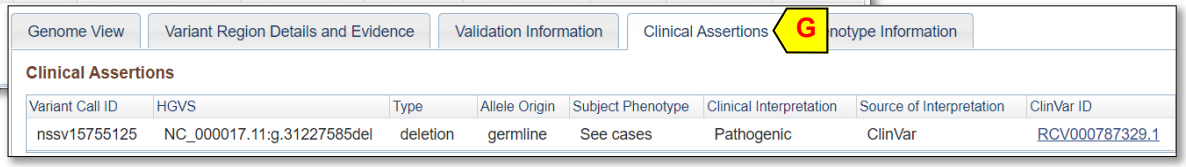
The top section of this variant view (**C**), from left to right, provides a summary of the variant, overview of its genomic placement, and links to relevant records in other NCBI databases. The variant's chromosomal mapping is indicated by green arrow.





Tabs underneath the summary provide other relevant details. Specifically, the "Genome View" (**D**) presents the variant graphically in its genomic context. Hovering the mouse pointer over a variant brings out its summary in a popup (**E**) The "Variant Region Details and Evidence" (**F**) provides detailed genomic coordinates on different genome builds. The "Clinical Assertions" tab (**G**) provides information on the phenotype and clinical interpretation when data is available.





References

1. Lappalainen I, et. al. (2013) dbVar and DGVa: public archives for genomic structural variation. Nucleic Acids Res. 41 (D1):D936-41. PMID: 23193291.
2. Sneddon TP and Church DM. (2012) Online resources for genomic structural variation. Methods Mol Biol. 838:273-89. PMID: 22228017.
3. dbVar help documentation: www.ncbi.nlm.nih.gov/dbvar/content/help/