



# The Genome Data Viewer from NCBI

Presenting public or custom data graphically under the context of genome annotation  
<https://www.ncbi.nlm.nih.gov/genome/gdv/>

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

## Overview

The Genome Data Viewer (GDV) is a web-based genome browser that offers graphical representations for more than 300 eukaryotic assemblies annotated by the NCBI Reference Sequence (RefSeq) project. Using GDV, you can visualize, explore, and analyze a selected genomic assembly and NCBI annotations at resolutions of the whole chromosome, nucleotide base level, and anywhere in between. You can visualize data of your interest either by selecting track sets organized to support different lines of inquiry or by picking individual tracks from a diverse collection. You also can upload files containing your own assembly annotations to view them alongside those offered by NCBI. GDV uses the graphical Sequence Viewer (SV) [1] as the basis for the genome display. The basic layout and control structure shared with other NCBI resources, such as the Variation Viewer [2], provides a unified interface.

## Access

The GDV landing page is the main entry point to browse and search the wide range of taxa and assemblies. On the page, an interactive tree (A) allows you to explore the scope of available organisms and their relationships to one another. A text box (B) above the tree allows you to browse and select assemblies for specific organisms or taxonomic groups, using either their common or scientific names. The knowledge panel (C) provides summary information for the selected organism. Within the knowledge panel, a drop-down menu (D) allows you to choose an assembly of interest when more than one assembly is available for a given organism. A text box (E) accepts location, gene or phenotype searches of the selected assembly for navigating to the corresponding genome position in the GDV display. You can access assembly-specific BLAST searches, general genome browsing, and several other assembly-related resources from the panel (F). Access to GDV is also available as browser links on record pages in several NCBI resources including Gene, Assembly, GEO.

For example, from the Assembly database, you can search with the fielded query term "assembly\_gdv[filter]" and combine it with an organism field-limited term (G) to see if your organism of interest has assemblies in GDV (H).

The screenshot shows the Genome Data Viewer interface. At the top, it says "U.S. National Library of Medicine" and "NCBI National Center for Biotechnology Information". The main heading is "Genome Data Viewer". Below this is a "Select organism" section with a tree view of various organisms. A text box (B) contains "Homo sapiens (human)". To the right, a "Knowledge panel" (C) displays details for the "Homo sapiens (human) genome", including search filters for "Location, gene or phenotype" and "Assembly" (GRCh38.p10). It also shows "Assembly details" such as Name, RefSeq accession, GenBank accession, and "Annotation details" like "Annotation Release 108" and "Release date 2016-06-06".

The screenshot shows search results for "Bos taurus" assemblies. It lists two items: "Bos taurus\_UMD\_3.1.1" and "Btau\_5.0.1". Each item includes details such as "Organism: Bos taurus (cattle)", "Intraspecific name: Breed: Hereford", "Sex: pooled male and female", "Submitter: Center for Bioinformatics and Computational Biology, University of Maryland", "Date: 2014/11/25", "Assembly level: Chromosome", "Genome representation: full", "RefSeq category: representative genome", "GenBank assembly accession: GCA\_000003055.5 (latest)", "RefSeq assembly accession: GCF\_000003055.6 (latest)", "Release type: Minor", and "IDs: 228231 [UID] 1400768 [GenBank] 1425528 [RefSeq]".

Access the data  
View the Genome

The screenshot shows the assembly details page for "Btau\_5.0.1". It includes "Display Settings" (Full Report), "Organism name: Bos taurus (cattle)", "Intraspecific name: Breed: Hereford", "Sex: pooled male and female", "BioSample: SAMN02898106", "Submitter: Cattle Genome Sequencing International Consortium", "Date: 2015/11/19", "Assembly level: Chromosome", "Genome representation: full", "GenBank assembly accession: GCA\_000003205.6 (latest)", "RefSeq assembly accession: GCF\_000003205.7 (latest)", "RefSeq assembly and GenBank assembly identical: yes", "WGS Project: AAF005", and "Assembly method: UMD Overlapper v. 2009: additional processing + PBJelly2 v. 14.9.9".

## Display Layout and Controls

The GDV browser display has a control panel (A) and a graphical panel (B). In the control panel, the “Ideogram View” (C) depicts individual chromosomes along with statistics of unlocalized/unplaced scaffolds and Alternate loci for the assembly. The boxed chromosome (D) is displayed in the graphical panel. The “Search” box (E) accepts input queries of location, gene and phenotype. GDV displays search results in a table below the search box and presents them on the ideogram as arrows. The “User Data and Track Hubs” section (F) accepts input annotation data for the displayed assembly and supports a wide range of file formats, such as remote Track Hubs, user uploaded BED, GFF3, and VCF. The BLAST section (G) allows for input and display of BLAST results through their assigned Request ID (RID). Hovering over an alignment (H) highlights the display in the graphical panel. The “Add Tracks” (I) allows you to add various NCBI datasets to the display, e.g., accession of GEO, SRA, public dbGaP and dbVar, if they have alignment to the assembly in display. You can customize the control panel display by clicking in the headers to collapse or expand specific sections, or by dragging-and-dropping their headers to reorder the sections.

In the graphical panel, the “Exon Navigator” (J) at the top allows for quick navigation from one gene or exon to another, or direct navigation to an exon of interest for a selected gene and transcript. Click the double-arrows to jump from one gene to the next, and hover over any one of the circles in this tool to see details about the exon for a selected gene/transcript of interest.

The default GDV graphical display includes only a small subset of the available tracks. You can use the cascading menu under the “Tracks” button (K) in the graphical display menu bar to update the display to one of the pre-configured “Track Sets” or create a custom configuration using the “Tracks” >> “Configure tracks” dialog box to select a desired category of tracks, select from the list, then activate the selection by clicking the “Configure” button.

## Custom Tracks

Through GDV, you can access many pre-computed data tracks in addition to those for features from NCBI's genome annotation. Those additional tracks include assembly supporting data from the sequence databases, expression data from the GEO database, and nextgen sequencing exon support data from the SRA database. GDV organizes those tracks into sets and makes them readily available through the cascading menu under the "Tracks" dialog box (K, p. 2). To select custom tracks, use the "Configure Tracks" option. Select or remove tracks by toggling on/off the checkboxes, then click the "Configure" button to activate your selection. You can take advantage of functions provided by My NCBI [3] to save your customized track selection to your My NCBI account. When logged in to My NCBI, your saved track sets will appear as an option under "Tracks" when the same genomic assembly is loaded.

## Custom Data Upload

### Case 1. Upload Mapped Variants

Researchers often align nextgen reads to a genome, and further process the alignment to variant calls in VCF format for SNP identification. GDV allows you to upload VCF files generated by such process and compare them against existing data from NCBI.

This example demonstrates how you can upload a custom VCF and compare them with known variants from dbSNP by using the provided file ([chr21.vcf](#)) that was generated by aligning reads from a brain RNA-seq run ([SRR5938470](#)) to human chromosome 21 from the previous assembly (GRCh37.p13). See magic-BLAST video tutorial and fact-sheet for more details [4,5]. To reproduce the display, go to GDV

landing page, change the assembly to GRCh.p13, expand the "User Data and ..." section in the right-hand column, click the plus sign (+), choose "Add URL" and enter the URL: <ftp.ncbi.nlm.nih.gov/pub/factsheets/chr21.vcf>. Click the ideogram for chromosome 21 in the Ideogram View section (A) to see the added track. Right click to activate the context menu, and use the "Zoom on range" option to zoom in to the region from 17,186,860 to 17186,919 (B). You can see some called variants match known variant positions in dbSNP (C), and some appear to be novel allele (D). There are also potentially previously unknown variants (E).

The screenshot shows the Genome Data Viewer interface for Homo sapiens: GRCh37.p13 (GCF\_000001405.25) on Chromosome 21 (NC\_000021.8). The interface includes an Ideogram View on the left, a search bar, and a 'Your Data' section. A red arrow points to the 'Add URL' button in the 'Your Data' section, which is used to upload the custom VCF file. Another red arrow points to the 'Zoom on Range' option in the context menu, which is used to zoom into the region from 17,186,860 to 17186,919. A 'Zoom to range' dialog box is shown with the sequence range 17186860 - 17186919 entered.

The detailed screenshot shows a zoomed-in view of the genomic region from 17,186,860 to 17,186,910. The top track shows the DNA sequence: C C A G G C T G A G G T G G T C T C A G A T G G A G A T G A G G A A C T T T G T T G G G A A C T G A A G C A A A G G T G A G G T C C G A C T C C A C C A G A G T C T A C C T C T A C T C C T T G A A C A A C C C T T G A C T T C G T T T C C A C. Below the sequence are tracks for Genes, NCBI Homo sapiens Annotation Release 105, dbSNP Build 150 (Homo sapiens Annotation Release 105) all data, Cited Variants, dbSNP Build 150 (Homo sapiens Annotation Release 105), and SRR5938470\_chr21\_variants. Callouts A, B, C, D, and E are placed on the tracks to highlight specific features and variants.

Variation ID	Local ID	Alleles	Allele Length	Variant Call Type	Validation Status
snp3377	snp3377	G/A	1	Single nucleotide variant	Not Tested
snp3381	snp3381	T/C	1	Single nucleotide variant	Not Tested
snp3384	snp3384	A/G	1	Single nucleotide variant	Not Tested

## Custom Data Upload (cont.)

### Example 2. Uploading Remapped Methylation Marker Data

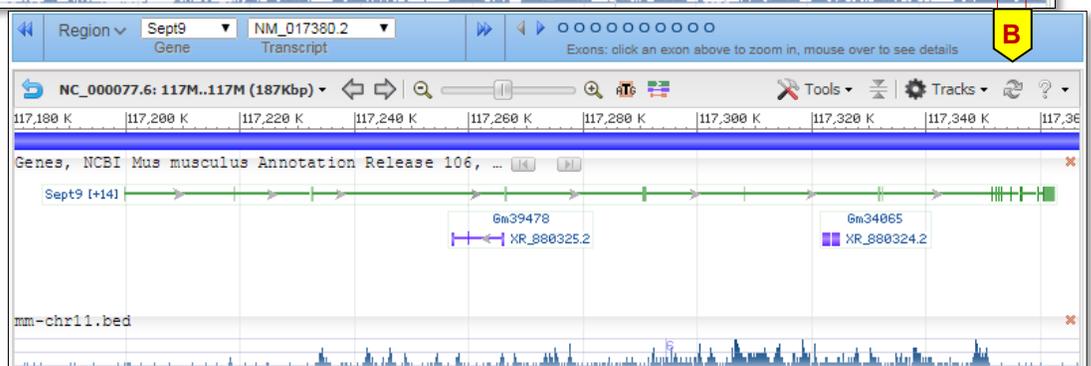
You may have data that were mapped on old genome assemblies. Such data can't be directly presented in GDV. However, you can work around this limit by transferring the mapping data to the new genome assembly using the NCBI Remap service [6], and then uploading the re-mapped data to the current assembly in GDV.

Due to large the amount of data, the example uses only a subset of the BED file from one specific sample, [GSM523698](#) that was mapped to chromosome 11. To generate the example BED file, we first remapped the data from the earlier mm8 mouse assembly to the current assembly GRCm38, then use the following steps to upload the resulting datasets and activate the display.

- Search with "mouse" in the GDV landing page and click "Browse genome" to load the mouse genome
- Use the "User Data and ..." >> "+" >> Add URL clicks to activate file upload widget, then copy and paste the following URL  
<ftp.ncbi.nlm.nih.gov/pub/factsheets/mm-chr11.bed>
- If the file is already downloaded to local disk, use "User Data and ..." >> "+" >> Files clicks instead, then load the file using "Choose Files" button
- Click on the ideogram of Chromosome 11 to see the track display (A)



You can load multiple datasets for direct side-by-side comparison, and zoom in to specific regions for detailed examination. For example, you can see the detailed methylation marks for the region of the Sept9 region (B) on the assembly.



## Technical Support and References

Please email questions and bug reports to [info@ncbi.nlm.nih.gov](mailto:info@ncbi.nlm.nih.gov). Refer to the following documents have more detail on related NCBI services.

1. Graphical Sequence Viewer Factsheet. <https://go.usa.gov/xRv4D>
2. Variation Viewer Factsheet. <https://go.usa.gov/xRv4k>
3. My NCBI factsheet. <https://go.usa.gov/xRv49>
4. Introducing Magic-BLAST. <https://www.youtube.com/watch?v=LrOHT73czZw>
5. Magic-BLAST factsheet. <https://go.usa.gov/xRAFj>
6. NCBI Remap Service. <https://go.usa.gov/xRv45>