



# NCBI Epigenomics

A portal to epigenetic maps and datasets from a variety of sources

<http://www.ncbi.nlm.nih.gov/epigenomics/>

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

## Scope and access

Epigenetics is an emerging field of study that examines global changes in the regulation of gene activity and expression that are not dependent on gene sequence. Epigenetic mechanisms include DNA methylation, post-translation modification of histone proteins and non-coding regulatory RNA. These mechanisms participate in the regulation of higher-order DNA structure and gene expression. We may consider the collection of epigenetic features across the genome to constitute the cellular 'epigenome'. The study of the epigenome, or epigenomics, refers to identifying what is literally 'on' the genome (the prefix *epi-* indicating above), and how these phenomena impact global gene expression, DNA-mediated processes, and subsequently, development.



The Epigenomics database at the National Center for Biotechnology Information (NCBI) is a resource that has been created to serve as a comprehensive public resource for whole-genome epigenetic data sets ([www.ncbi.nlm.nih.gov/epigenomics](http://www.ncbi.nlm.nih.gov/epigenomics)). We have constructed the new resource by selecting the subset of epigenetics-specific data from the general-purpose archive, the Gene Expression Omnibus (GEO), and then subjecting them to further review, annotation, and reorganization. Typically, raw data are processed and mapped to genomic coordinates to generate 'tracks' that are a visual representation of the data. These data tracks can be viewed using popular genome browsers or downloaded for local analysis.

## The Browser: An interface for navigating Epigenomics content

The Browser ([www.ncbi.nlm.nih.gov/epigenomics/browse](http://www.ncbi.nlm.nih.gov/epigenomics/browse)) provides a unique interface for navigating the experiments (or biological samples) that are in the Epigenomics Database. This tool has many navigation and display features. The database contents can be browsed in the context of experimental or biological sample data (A). The ability to use free text to filter records is also provided (B). Three pre-selected criteria allow easy filtering by species, biological source and/or the feature type assayed (C). The Configure button allows for customizing the display of various attributes (D). Records that are selected from the Browser interface, by using the check boxes in their corresponding rows, can be saved to the Clipboard or to a custom collection by using the Clipboard and Copy icons (E). Finally, epigenomic track data from selected samples can be viewed by clicking the View on Genome button (F) or downloaded to a local machine by clicking the Download button (G). The Download button also gives you the option of downloading the contents of the Browser window. A link to a thorough tutorial video regarding this interface is provided (H).

The screenshot shows the NCBI Epigenomics Browser interface. Annotations A through H highlight specific features:

- A**: The 'Browse' tab and 'Experiments'/'Samples' sub-tabs.
- B**: The 'Filters' section, including a text input for 'Containing word(s):' and a 'Filter' button.
- C**: The 'Feature Type' dropdown menu, showing options like H3K27me3, H3K36ac, H3K36me2, H3K36me3, H3K4ac, and H3K4me1.
- D**: The 'Configure' button in the top right corner.
- E**: The 'Clipboard' and 'Copy' icons in the 'My Collections' section.
- F**: The 'View on Genome' button.
- G**: The 'Download' button.
- H**: The 'YouTube Browser overview' link in the top right corner.

The main table displays a list of experiments with columns for Experiment ID, Feature Type, Assay Type, Species, Cell Type, Tissue Type, Cell Line, Cell Population, and Diff. State. The first five rows are selected, and the table shows 'All 20 items on this page are selected.'

Experiment ID	Feature Type	Assay Type	Species	Cell Type	Tissue Type	Cell Line	Cell Population	Diff. State
<input checked="" type="checkbox"/> <a href="#">ESX000000026</a>	H3K4me3	histone modification	Homo sapiens	embryonic stem cell		H1		undif
<input checked="" type="checkbox"/> <a href="#">ESX000000033</a>	H3K4me3	histone modification	Homo sapiens	embryonic stem cell		H1		undif
<input checked="" type="checkbox"/> <a href="#">ESX000000038</a>	H3K4me3	histone modification	Mus musculus	T lymphocyte	spleen/lymph node		CD4+	naiv
<input checked="" type="checkbox"/> <a href="#">ESX000000044</a>	H3K4me3	histone modification	Mus musculus	T lymphocyte			CD4+	T he
<input checked="" type="checkbox"/> <a href="#">ESX000000046</a>	H3K4me3	histone modification	Mus musculus	T lymphocyte			CD4+	T req

## Visualizing epigenomic data using Sequence Viewer

A popular approach to analyzing epigenomic data is to display the results as ‘tracks’ on a genome browser. Tracks represent continuous-value data aligned against the genome; enriched regions are depicted as peaks in the track. After selecting one or more samples from the Epigenomics Sample Browser (previous page), click the “View on Genome” button in the Sample browser toolbar to redirect to the a page with viewing options and select NCBI Sequence Viewer.

The screenshot displays the NCBI Sequence Viewer interface for the NANOG gene. The interface is divided into several sections:

- Search and Navigation:** A search box (A) contains "NANOG". Below it, a list of search results (B) shows "NANOG" and "NANOGP8". A "Species" dropdown (D) is set to "Human (hg19)". A "Gene or location" search box (B) also contains "NANOG".
- Chromosome and Gene Information:** A chromosome selection bar (C) shows chromosomes 1 through 22, X, and Y. The "NANOG" gene is highlighted. A "View at UCSC" button (F) is visible.
- Data Tracks:** The main display area shows several tracks:
  - ChIP-Seq analysis of H3K4me3 in human H1 cells; 314G7AAXX090305-5-5
  - ChIP-Seq Analysis of H3K27me3 in hESC H1 Cells, 30CMAAAXX081006-2-5
  - UCSD IMR90 Cell Line Histone H3K4me3 Library LL221 EA Release 2
  - UCSD IMR90 Cell Line Histone H3K27me3 Library LL223 EA Release 2
- Configuration Panel (H):** A "Configure Page" panel is open, showing "Active Tracks" (I) with checkboxes for Sequence, Variation, Epigenomics, Genes/Products, Alignments, and Features. The "Track Settings: Sequence" panel (I) is also visible, showing a description of the track and a "Show Label" checkbox.
- Buttons:** A "YouTube View on genome overview" button (G) is in the top right. A "Configure" button (J) is at the bottom right of the configuration panel.

The Sequence Viewer display of epigenomic data (shown above) offers many navigation and display features. Navigate to a gene or chromosome location by entering it the “Gene or Location” box (A). The results of the search are populated in the box below the query box (B). Alternatively, you can select to view an entire chromosome by clicking on the respective buttons above the NCBI Sequence Viewer interface (C). Autocomplete functionality has been incorporated and as you enter a gene name into the query box a list of possible matches is displayed. The query searches the NCBI Gene database and the list represents these results. If a specific gene is searched for and found, the gene is highlighted in the query result box. If you have selected data tracks from multiple species or assemblies you can use the “Species” drop down menu to select the species and genome assembly of interest (D). The sequence viewer window will show the gene of interest with the data tracks that had been previously selected. Other tracks are also displayed by default including a track showing CpG islands and cited clinical variants. If no data tracks had been selected previously a set of default tracks that have been manually curated will be displayed (E). Tracks can also be viewed on the UCSC Genome Browser. Clicking on the “View on UCSC” button will load your selected tracks to the UCSC browser (F). If you have specified a gene or location on the Epigenomics View page, the UCSC browser will display the data at the same location. A link to a thorough tutorial video regarding the genome viewer interface is provided (G).

Below the data track display, there is a configuration panel that provides customization options for the sequence viewer interface (H). The “Active tracks” tab lists the tracks that are currently displayed (I). This will include the data tracks selected from epigenomics, and other tracks such as genes, translations, genomic variations, etc. Tracks in this tab can be re-ordered, selected, or deselected which impacts what is being displayed. The “Epigenomics” tab specifically lists the tracks that have been selected from the Epigenomics database. After making changes such as selecting, deselecting or reordering tracks, clicking on the “Configure” button will reload the sequence viewer with these changes (J).