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. While epigenetics typically refers to the study of a single gene or sets of genes, epigenomics refers to a more global analysis of epigenetic changes across an entire genome. Aberrant epigenetic mechanisms are involved in the development of many diseases, including cancer.

The NIH Roadmap Epigenomics Mapping Consortium was launched with the goal of producing a public resource of human epigenomic data to catalyze basic biology and disease-oriented research. The Consortium leverages experimental pipelines built around next-generation sequencing technologies to map DNA methylation, histone modifications, chromatin accessibility and small RNA transcripts in stem cells and primary ex vivo tissues selected to represent the normal counterparts of tissues and organ systems frequently involved in human disease. More information about the Roadmap Epigenomics project is available at: http://www.roadmapepigenomics.org/

Accessing Roadmap Epigenomics data

Data from the Roadmap Epigenomics project are hosted at NCBI’s Gene Expression Omnibus (GEO) database. Project data are listed at http://www.ncbi.nlm.nih.gov/geo/roadmap/epigenomics/. The data listings page includes multiple features. There are three tabs that allow navigation between sample holdings, study records and a graphical interface for easy visualization of experimental data. The Sample tab has multiple features including links to original GEO records (A), search box for filtering the table for records containing specific keywords (B), View Track check boxes that allow selection of data to display on the genome browser. The option to use NCBI’s Sequence Viewer or the UCSC Genome Browser is provided (C). Links to download the original sequence data (fastq format), alignment data (bam format) and trace data (wig format) are provided (D).

On the project data listings page, the Matrix tab conveniently displays current sample holdings at GEO and provides a birds eye view of samples and assays that are the focal points of the Roadmap Epigenomics Project, and includes: A sample/experiment matrix with samples grouped by mapping center (submitting institution). Biological sample is indicated on the left hand side of the Matrix, and epigenetic feature assayed is indicated on the top axis (E). The number of accessions for each cell is shown (F). A View Tracks tool is also available. This enables the display of selected tracks in a sequence viewer. NCBI Sequence Viewer and the UCSC Genome Browser are both supported (G).
NCBI Epigenomics
Roadmap Epigenomics project data is also being hosted at NCBI’s new dedicated Epigenomics database at [http://www.ncbi.nlm.nih.gov/epigenomics/](http://www.ncbi.nlm.nih.gov/epigenomics/). This resource enables users to explore Roadmap Epigenomics data alongside hundreds of other richly-annotated epigenomic datasets. The NCBI Epigenomics home page provides a portal to the Roadmap Epigenomics project. Links are provided to take you to the Roadmap Project page at NIH, the Mapping Consortium page, the Human Epigenome Atlas, which is hosted at Baylor College of Medicine (the Epigenomic Data Analysis and Coordination Center). Links also are provided for the GEO Roadmap data listings page and data access policies.

Visualizing epigenomics data on NCBI’s 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 tracks from the Roadmap Epigenomics data listings table (previous page), click the “View Track” button in the column header to redirect to the NCBI Sequence Viewer.

The Sequence Viewer has many navigation and display features, including the ability to go to a specific location either by entering an gene symbol or genomic coordinates into the search box (A). The queried gene, and the surrounding locus, is displayed in the lower ‘Gene models’ track (B). Corresponding epigenomic results for that locus are displayed in the middle (blue) tracks. The track labels of the example to the right indicate that histone H3 trimethylated at lysine 4 (H3K4me3) ChIP-Seq results in H1 cells (upper track) and IMR90 cells (lower track) are being displayed. Peaks in the tracks indicate areas of the genome that are enriched for this histone modification (C). Additionally, the slider/zoom tool can be used to expand or narrow the viewing region (D). A viewing window can also be defined by resizing and/or repositioning the slider tab in the overview panel (E).