# Description

This track displays maps of chromatin state generated by the Broad/MGH ENCODE group using ChIP-seq. Chemical modifications (methylation, acetylation) to the histone proteins present in chromatin influence gene expression by changing how accessible the chromatin is to transcription.

The ChIP-seq method involves first using formaldehyde to cross-link histones and other DNA-associated proteins to genomic DNA within cells. The cross-linked chromatin is subsequently extracted, mechanically sheared, and immunoprecipitated using specific antibodies. After reversal of cross-links, the immunoprecipitated DNA is sequenced and mapped to the human reference genome. The relative enrichment of each antibody-target (epitope) across the genome is inferred from the density of mapped fragments.

# Display Conventions and Configuration

This track is a multi-view composite track that contains multiple data types (*views*). For each view, there are multiple subtracks that display individually on the browser. Instructions for configuring multi-view tracks are here. ENCODE tracks typically contain one or more of the following views:

*Peaks*
: Regions of statistically significant signal enrichment. The score associated with each enriched interval is the mean signal value across the interval. (Note that a broad region with moderate enrichment may deviate from the background more significantly than a short region with high signal.)

*Signal*
: Density graph (wiggle) of signal enrichment. At each base-pair position the density is calculated as the number of sequenced tags overlapping a 25 bp window centered at that position.

Peaks and signals displayed in this track are the results of pooled replicates. The raw sequence and alignment files for each replicate are available for download.

Metadata for a particular subtrack can be found by clicking the down arrow in the list of subtracks.

# Methods

***ChIP-seq***: Cells were grown according to the approved ENCODE cell culture protocols. Cells were fixed in 1% formaldehyde and resuspended in lysis buffer. Chromatin was sheared to 200-700 bp using a Diagenode Bioruptor. Solubilized chromatin was immunoprecipitated with antibodies against each of the histone antibodies listed above. Antibody-chromatin complexes were pulled-down using protein A-sepharose (or anti-IgM-conjugated agarose for RNA polymerase II), washed and then eluted. After cross-link reversal and proteinase K treatment, immunoprecipitated DNA was extracted with phenol-chloroform, ethanol precipitated, treated with RNAse and purified. One to ten nanograms of DNA were end-repaired, adapter-ligated and sequenced by Illumina Genome Analyzers as recommended by the manufacturer.

***Alignment***: Sequence reads from each IP experiment were aligned to the human reference genome (**GRCh37/hg19**) using MAQ with default parameters, except '`-C 11`' and '`-H output_file`', which outputs up to 11 additional best matches for each read (if any are found) to a file. This information was used to filter out any read that had more than 10 best matches on the genome. Note: It is likely that instances where multiple reads align to the same position and with the same orientation are due to enhanced PCR amplification of a single DNA fragment. No attempt has been made, however, to

remove such artifacts from the data, following ENCODE practices.

***Signal***: Fragment densities were computed by counting the number of reads overlapping each 25 bp bin along the genome. Densities were computed using igvtools count with default parameters (in particular, '`-w 25`' to set window size of 25 bp and '`-f mean`' to report the mean value across the window), except for '`-e`' set to extend the reads to 200 bp, and the *.wig* output was converted to *bigWig* using `wigToBigWig` from the UCSC Kent software package.

***Peaks***: Discrete intervals of ChIP-seq fragment enrichment were identified using Scripture, a scan statistics approach, under the assumption of uniform background signal. All data sets where processed with '`-task chip`', and with '`-windows 100,200,500,1000,5000,10000,100000`'. (No mask file nor the '`-trim`' option have been used.) The resulting called segments were then further filtered to remove intervals that are significantly enriched only because they contain smaller enriched intervals within them. This post-processing step has been implemented using `Matlab`. The use of the post-processing step allowed very large enriched intervals (of the order of Mbps for H3K27me3, for instance) to be detected, as well as much smaller intervals, without the need to tailor the parameters of `Scripture` based on prior expectations.

# Release Notes

Release 1 (Feb 2011) of this track consists of a remapping of all previously released experiments to the human reference genome **GRCh37/hg19** (these data were previously mapped to **NCBI36/hg18**; please see the Release Notes section of the hg18 Broad Histone track for information on the NCBI36/hg18 releases of the data). New experiments for five additional cell lines (HeLa-S3, HSMMtube, NH-A, NHDF-Ad and Osteobl) and three additional antibody targets (H2A.Z, H3K9me3 and H3K79me2) have been included in this release and are not found in **NCBI36/hg18**.

# Credits

The ChIP-seq data were generated at the Broad Institute and in the Bradley E. Bernstein lab at the Massachusetts General Hospital/Harvard Medical School. Contact: Noam Shoresh.

Data generation and analysis was supported by funds from the NHGRI, the Burroughs Wellcome Fund, Massachusetts General Hospital and the Broad Institute.

# References

Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, McMahon S, Karlsson EK, Kulbokas EJ 3rd, Gingeras TR *et al.* Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*. 2005 Jan 28;120(2):169-81.

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*. 2006 Apr 21;125(2):315-26.

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*. 2011 May 5;473(7345):43-9.

Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, Fan L, Koziol MJ, Gnirke A, Nusbaum C *et al.* Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat Biotechnol*. 2010 May;28(5):503-10.

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP *et al*. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*. 2007 Aug 2;448(7153):553-60.

# Publications

Ram O, Goren A, Amit I, Shoresh N, Yosef N, Ernst J, Kellis M, Gymrek M, Issner R, Coyne M *et al*. Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. *Cell*. 2011 Dec 23;147(7):1628-39.

# Data Release Policy

Data users may freely use ENCODE data, but may not, without prior consent, submit publications that use an unpublished ENCODE dataset until nine months following the release of the dataset. This date is listed in the *Restricted Until* column on the track configuration page and the download page. The full data release policy for ENCODE is available here.