Description

This track, produced as part of the ENCODE Project, contains deep sequencing DNase data that will be used to identify sites where regulatory factors bind to the genome (*footprints*).

Footprinting is a technique used to define the DNA sequences that interact with and bind DNA-binding proteins, such as transcription factors, zinc-finger proteins, hormone-receptor complexes, and other chromatin-modulating factors like CTCF. The technique depends upon the strength and tight nature of protein-DNA interactions. In their native chromatin state, DNA sequences that interact directly with DNA-binding proteins are relatively protected from DNA-degrading endonucleases, while the exposed/unbound portions are readily degraded by such endonucleases. A massively parallel next-generation sequencing technique to define the DNase hypersensitive sites in the genome was adopted. The DNase samples were sequenced using next-generation sequencing machines to significantly higher depths of 300-fold or greater. This produces a base-pair level resolution of the DNase susceptibility maps of the native chromatin state. These base-pair resolution maps represent and are dependent upon the nature and the specificity of interaction of the DNA with the regulatory/modulatory proteins binding at specific loci in the genome; thus they represent the native chromatin state of the genome under investigation. The deep sequencing approach has been used to define the footprint landscape of the genome by identifying DNA motifs that interact with known or novel DNA-binding proteins.

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.

For each cell type, this track contains the following views:

HotSpots
: DNaseI hypersensitive zones identified using the HotSpot algorithm.

Peaks
: DNaseI hypersensitive sites (DHSs) identified as signal peaks within FDR 1.0% hypersensitive zones.

Signal
: Per-base count of sequence reads whose 5' end (corresponding to a DNaseI-induced DNA cut) coincides with the given position.

Raw Signal
: The density of tags mapping within a 150 bp sliding window (at a 20 bp step across the genome).

NOTE: The names of the signal views in this track are reversed from conventions used in other ENCODE tracks, where the less processed signal is termed 'Raw'.

DNaseI sensitivity is shown as the absolute density of *in vivo* cleavage sites across the genome mapped using the Digital DNaseI methodology (see below).

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

Methods

Cells were grown according to the approved ENCODE cell culture protocols. Digital DNaseI was performed by DNaseI digestion of intact nuclei, followed by isolation of DNaseI 'double-hit' fragments as described in Sabo *et al*. (2006), and direct sequencing of fragment ends (which correspond to *in vivo* DNaseI cleavage sites) using the Solexa platform (27 bp reads). High-quality reads were mapped to the GRCh37/hg19 human genome using Bowtie 0.12.5 (Eland was used to map to NCBI36/hg18); only unique mappings were kept. DNaseI sensitivity is directly reflected in raw tag density (*Signal*), which is shown in the track as density of tags mapping within a 150 bp sliding window (at a 20 bp step across the genome). DNaseI hypersensitive zones (*HotSpots*) were identified using the HotSpot algorithm described in Sabo *et al*. (2004). False discovery rate thresholds of 1.0% (FDR 1.0%) were computed for each cell type by applying the HotSpot algorithm to an equivalent number of random uniquely mapping 36-mers. DNaseI hypersensitive sites (DHSs or *Peaks*) were identified as signal peaks within 1.0% (FDR 1.0%) hypersensitive zones using a peak-finding algorithm. Only DNase Solexa libraries from unique cell types producing the highest quality data, as defined by Percent Tags in Hotspots (PTIH ~40%), were designated for deep sequencing to a depth of over 200 million tags.

Verification

Results were validated by conventional DNaseI hypersensitivity assays using end-labeling/Southern blotting methods.

Images and their associated mappings can be found in the supplemental data.

Release Notes

This is Release 4 (August 2012) of this track, which includes 10 new experiments across 8 cell lines.

A number of previously released Peaks have been replaced by updated versions. The affected database tables and files include 'V2' in the name, and metadata is marked with "submittedDataVersion=V2", followed by the reason for replacement, "Fixed bug in peak calls that artificially reduced the number of peaks".

Previous versions of files are available for download from the FTP site.

Credits

These data were generated by the UW ENCODE group.

Contact: Richard Sandstrom

References

Sabo PJ, Hawrylycz M, Wallace JC, Humbert R, Yu M, Shafer A, Kawamoto J, Hall R, Mack J, Dorschner MO *et al*. Discovery of functional noncoding elements by digital analysis of chromatin structure. *Proc Natl Acad Sci U S A.* 2004 Nov 30;101(48):16837-42.

Sabo PJ, Kuehn MS, Thurman R, Johnson BE, Johnson EM, Cao H, Yu M, Rosenzweig E, Goldy J, Haydock A *et al*. Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nat Methods*. 2006 Jul;3(7):511-8.

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, above. The full data release policy for ENCODE is available here.