Description

This track is produced as part of the ENCODE Project. This track shows genome-wide assessment of DNA replication timing in cell lines using the sequencing-based "Repli-seq" methodology (see below). Replication timing is known to be an important feature for epigenetic control of gene expression that usually operates at a higher-order level than at the level of specific genes. For each experiment (*cell line, replicate)*, replication timing was ascertained by the isolation and sequencing of newly replicated DNA from six cell cycle fractions: G1/G1b, S1, S2, S3, S4, G2 (six fraction profile). Replication patterns are visualized as a continuous function based on sequencing tag density (*Percentage-normalized Signal*) and as a wavelet-smoothed transform of the six fraction profile (*Wavelet-smoothed Signal*). Replication peaks corresponding to replication initiation zones (*Peaks*) and valleys corresponding to replication termination zones (*Valleys*) were determined from local maxima and minima, respectively, in the wavelet-smoothed signal data. A measure of relative copy number at each genomic location (*Summed Densities*) was determined by summing the normalized tag density values of each cell cycle fraction at that location (equals one replicated genome equivalent).

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:

Percentage-normalized Signal
: Replication signal at 1 kb intervals as a percentage of normalized +/-25 kb tag densities for all cell cycle fractions (G1/G1b, S1, S2, S3, S4, G2).

Wavelet-smoothed Signal
: Wavelet-smoothed transform of the six fraction profile that is a weighted average of the percentage-normalized signals such that earlier replication has higher values.

Peaks
: Local maxima in the wavelet-smoothed signal data corresponding to replication initiation (replication origin) zones.

Valleys
: Local minima in the wavelet-smoothed signal data corresponding to replication termination zones.

Summed Densities
: A measure of relative copy number at each genomic location that is the sum of normalized tag densities for each cell cycle fraction.

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. Repli-seq was performed as described by Hansen *et al.* (2010). Briefly, newly replicated DNA was labeled *in vivo* with a pulse of 5-bromo-2-deoxyuridine (BrdU), cells were fractionated into six different parts of the cell cycle by flow cytometry according to DNA content, cell cycle fractionated DNA was sonicated and an anti-BrdU monoclonal antibody was used to isolate the newly replicating DNA. Fragment ends were sequenced using the Illumina Genome Analyzer II or HiSeq platforms (36 bp reads). Some experiments (BJ, K562, BG02ES, GM06990) were originally performed and mapped to an earlier version of the human reference genome **NCBI36/hg18** (Hansen *et al.*, 2010) and were remapped to the more recent reference genome **GRCh37/hg19**.

Uniquely mapping high-quality reads were mapped to the genome minus the Y chromosome. Replication signals within each six cell cycle fraction were derived from the density of sequence tags mapping within a 50 kb sliding window (stepped 1 kb across the genome); these densities were normalized to 4 million tags per genome. To avoid variation due to copy number or sequence bias, cell cycle-specific replication signals at each location were determined as a percentage of the sum of the six normalized tag density signals (*Percentage-normalized Signal*).

To transform the six fraction replication signals into one track (*Wavelet-smoothed Signal*), the percentage-normalized signals at each location were used to calculate a weighted average value based on the average DNA content of each fraction according to flow cytometry [higher values correspond to earlier replication;
formula=(0.917*G1b)+(0.750*S1)+(0.583*S2)+(0.417*S3)+(0.250*S4)+(0*G2)]. These weighted average data were smoothed by wavelet transformation [J7 level, corresponding to a scale of 128 kb; see Thurman *et al.* (2007)].

Replication initiation zones were flagged by determining local maxima in the wavelet-smoothed data (*Peaks*) and, similarly, replication termination zones were flagged by local minima (*Valleys*).

The sum of the 4 million normalized replication tag densities correspond to replication of one genome and can, therefore, be used as a measure of relative genomic copy number (*Summed Densities*). This is useful for evaluation of unusual replication patterns, such as "biphasic" ones where replication has both early and late components [as described by Hansen *et al.* (2010)].

Verification

Data were verified by determining replication time with a PCR-based examination of replication for particular loci in addition to sequencing biological replicates, as described by Hansen *et al.* (2010).

Credits

These data were generated by the UW ENCODE group.

Contact: Richard Sandstrom

References

Hansen RS, Thomas S, Sandstrom R, Canfield TK, Thurman RE, Weaver M, Dorschner MO, Gartler SM, Stamatoyannopoulos JA. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc Natl Acad Sci U S A*. 2010 Jan 5;107(1):139-44.

Thurman RE, Day N, Noble WS, Stamatoyannopoulos JA. Identification of higher-order functional domains in the human ENCODE regions. *Genome Res*. 2007 Jun;17(6):917-27.

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.