ENCODE Elements README - Candidate enhancers

The following 14 ENCODE cell lines were identified to be annotated in the prototype release, because there are both DNase-seq and H3K27ac ChIP-seq ENCODE data on them:

GM12878
H1-hESC
HeLa-S3
HepG2
HMEC
HSMM
HSMMtube
HUVEC
K562
NH-A
NHDF-Ad
NHEK
NHLF
Osteobl

Procedure for generating prototype tracks

1.   Distal and proximal DHS tracks:
    1.1.   Jan 2011 freeze DNase peaks were obtained from
            http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeAwgDnaseUniform/
    1.2.   All DNase peaks from the above 14 cell lines were merged to form
            non-overlapping DNase-hypersensitive regions.
    1.3.   The merged DNase-hypersensitive regions were separated into TSS-proximal
            and TSS-distal groups based on whether or not they intersected a 2000-bp
            window centered on any GENCODE TSS (V19; gencode.v19.TSS.notlow.gff).
    1.4.   For each proximal and distal DNase-hypersensitive region, the DNase peak with
            the maximum score was labeled as a "master peak". These master peaks are
            available as prototype tracks dnase_track_distal.bb and dnase_track_proximal.bb
    1.5.   Track details include names and number of cell lines in each merged
            DNase-hypersensitive regions.

2.   Distal H3K27ac:
    2.1.   H3K27ac signal files for the above 14 cell lines were obtained from
            http://encodedcc.stanford.edu/ftp/modENCODE_VS_ENCODE/Regulation/Human/signal/foldChange/

2.2. For each distal DNase master peak, the average H3K27ac signal in each of the 14 cell lines was calculated in a 1000-bp window around the center of the peak. This signal was converted to a percentile using the background distribution of H3K27ac signal in randomly chosen 1000-bp genomic regions that were outside all DNase peaks from the Jan 2011 freeze (not restricted to the 14 cell lines above) and ENCODE blacklisted regions.

2.3. DNase master peaks that have at least one cell line with H3K27ac signal > 95th percentile of background are reported in the track H3K27ac.bb. If there are multiple cell lines that fulfil the 95th percentile criteria, they are displayed as separate lines in the track.

2.4. Track details include actual H3K27ac percentile over background.

3. Distal and proximal TF binding:

3.1. Transcription factor (TF) ChIP-seq peaks were obtained from http://encodedcc.stanford.edu/ftp/modENCODE_VS_ENCODE/Regulation/Human/peakCalls/finalPk/

3.2. For each of the distal and proximal DNase master peaks, overlapping TF ChIP-seq peaks across all cell types (not restricted to above 14 cell lines) were identified. The TF peaks with maximum score in each master DNase peak is displayed in prototype tracks chipseq_track_distal.bb and chipseq_track_proximal.bb.

3.3. Track details include all names (with cell line information) of TFs whose peaks overlapped with the DNase master peak.