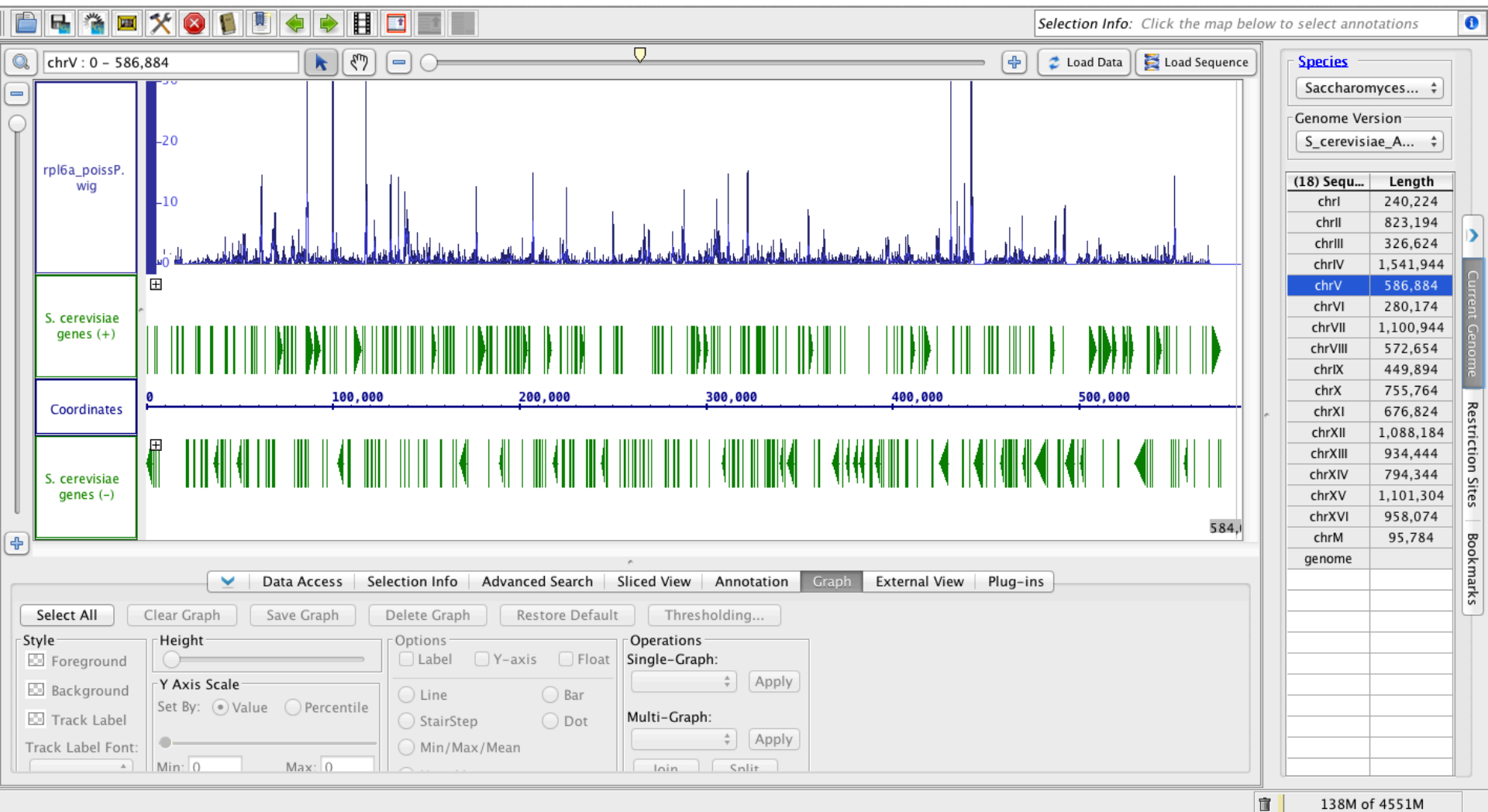




ChIP-seq analysis: Functional analysis







ChIP-seq analysis: Functional analysis

[Overview](#)[News](#)[Use GREAT](#)[Demo](#)[Video](#)[How to Cite](#)[Help](#)[Forum](#)[Bejerano Lab, Stanford University](#)

GREAT version 2.0.2 current (04/03/2012 to now)

GREAT predicts functions of cis-regulatory regions.

Many coding genes are well annotated with their biological functions. Non-coding regions typically lack such annotation. GREAT assigns biological meaning to a set of non-coding genomic regions by analyzing the annotations of the nearby genes. Thus, it is particularly useful in studying cis functions of sets of non-coding genomic regions. Cis-regulatory regions can be identified via both experimental methods (e.g. [ChIP-seq](#)) and by computational methods (e.g. [comparative genomics](#)). For more see our [Nature Biotech Paper](#).

News

- Apr 3, 2012: GREAT version 2.0 [adds new annotations to human and mouse ontologies and visualization tools for data exploration](#).
- Feb 18, 2012: The [GREAT forums](#) are released, allowing increased user-to-user interaction

[More news items...](#)

Species Assembly

- ☐ Human: GRCh37 ([UCSC hg19, Feb/2009](#))
- ☐ Human: NCBI build 36.1 ([UCSC hg18, Mar/2006](#))
- ☐ Mouse: NCBI build 37 ([UCSC mm9, Jul/2007](#))
- ☐ Zebrafish: Wellcome Trust Zv9 ([danRer7, Jul/2010](#))  [Zebrafish CNE set](#)

[Can I use a different species or assembly?](#)



ChIP-seq analysis: Functional analysis

Associating genomic regions with genes

GREAT calculates statistics by associating genomic regions with nearby genes and applying the gene annotations to the regions. Association is a two step process. First, every gene is assigned a regulatory domain. Then, each genomic region is associated with all genes whose regulatory domain it overlaps.

☒ Basal plus extension



Proximal: kb upstream, kb downstream, plus Distal: up to kb

Gene regulatory domain definition: Each gene is assigned a basal regulatory domain of a minimum distance upstream and downstream of the TSS (regardless of other nearby genes). The gene regulatory domain is extended in both directions to the nearest gene's basal domain but no more than the maximum extension in one direction.

☐ Two nearest genes



within kb

Gene regulatory domain definition: Each gene is assigned a regulatory domain that extends in both directions to the nearest gene's TSS but no more than the maximum extension in one direction.

☐ Single nearest gene



within kb

Gene regulatory domain definition: Each gene is assigned a regulatory domain that extends in both directions to the midpoint between the gene's TSS and the nearest gene's TSS but no more than the maximum extension in one direction.

 Gene Transcription Start Site (TSS)

☒ Include curated regulatory domains

What are curated regulatory domains?



ChIP-seq analysis: Functional analysis

Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see [Using the Table Browser](#) for a description of the controls in this form, the [User's Guide](#) for general information and sample queries, and the OpenHelix Table Browser [tutorial](#) for a narrated presentation of the software features and usage. For more complex queries, you may want to use [Galaxy](#) or our [public MySQL server](#). To examine the biological function of your set through annotation enrichments, send the data to [GREAT](#). Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with these data. All tables can be downloaded in their entirety from the [Sequence and Annotation Downloads](#) page.

clade: Mammal
genome: Human
assembly: Feb. 2009 (GRCh37/hg19)

group: Genes and Gene Prediction Tracks
track: UCSC Genes

table: knownGene

region: ☒ genome ☐ ENCODE Pilot regions ☐ position chr21:33031597-33041570

identifiers (names/accessions):

filter:

intersection:

correlation:

output format: all fields from selected table Send output to ☐ [Galaxy](#) ☐ [GREAT](#)

output file: (leave blank to keep output in browser)

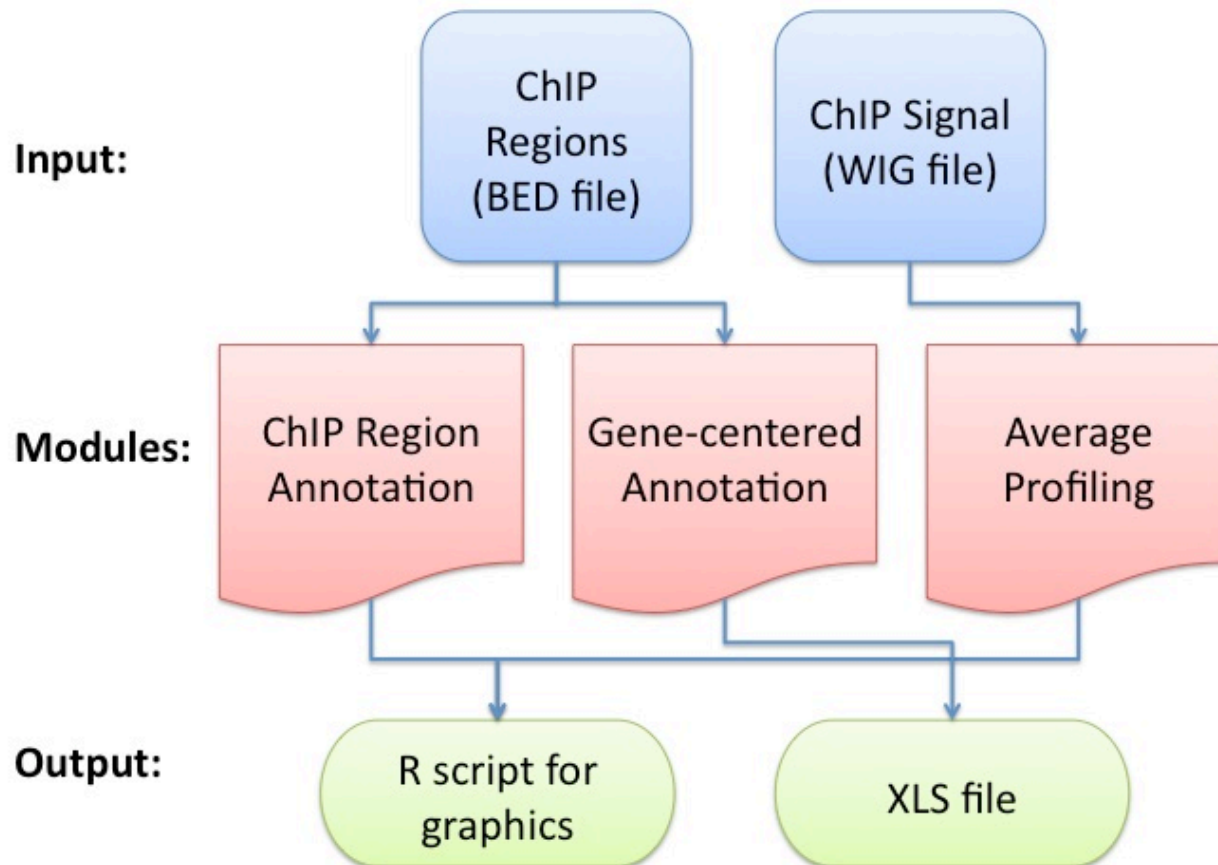
file type returned: ☒ plain text ☐ gzip compressed

To reset all user cart settings (including custom tracks), [click here](#).



ChIP-seq analysis: Functional analysis

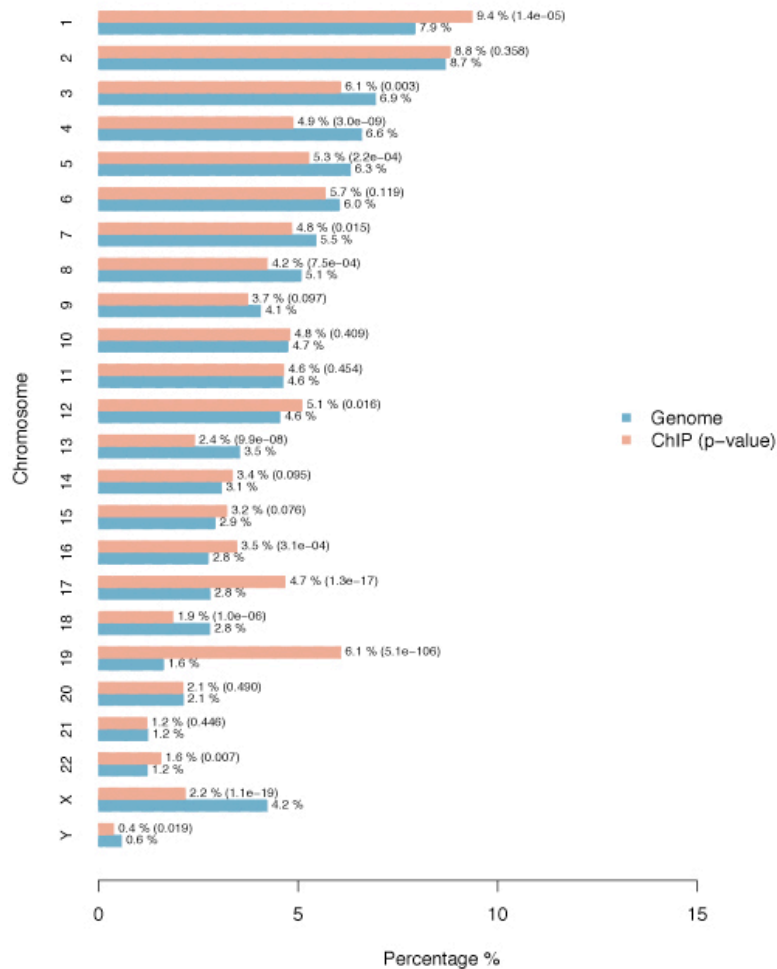
CEAS (Cis-regulatory Element Annotation System)



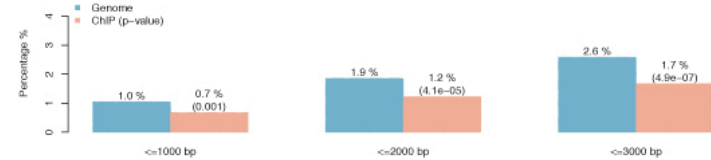


ChIP-seq analysis: Functional analysis

Chromosomal Distribution of ChIP Regions



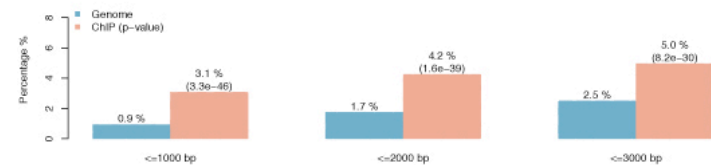
Promoter



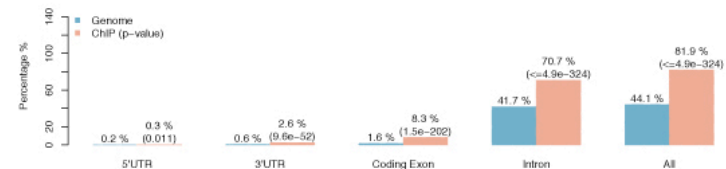
Bidirectional Promoter



Downstream



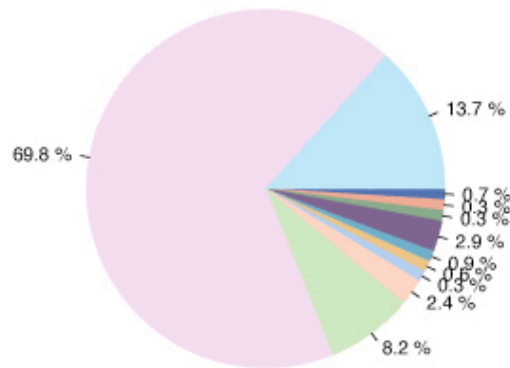
Gene





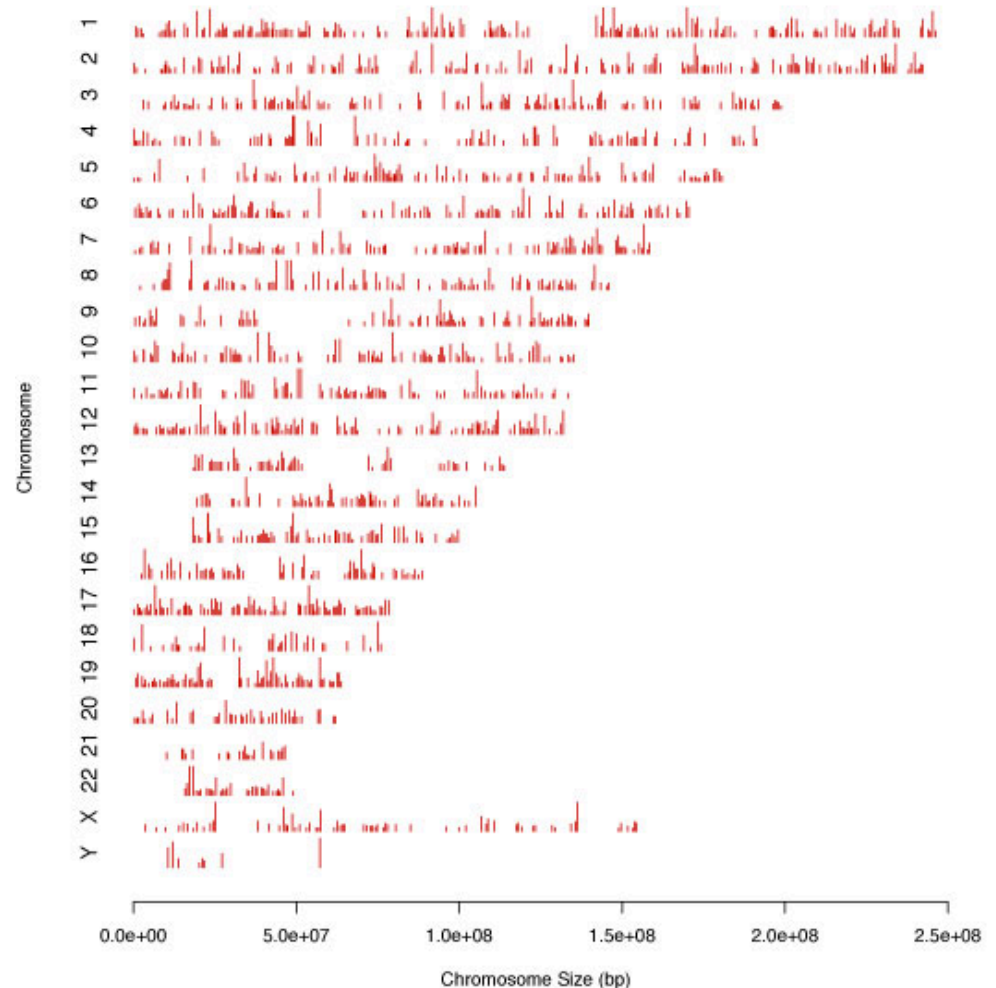
ChIP-seq analysis: Functional analysis

Distribution of ChIP Regions



- Promoter (<=1000 bp): 0.7 %
- Promoter (1000–2000 bp): 0.3 %
- Promoter (2000–3000 bp): 0.3 %
- Downstream (<=1000 bp): 2.9 %
- Downstream (1000–2000 bp): 0.9 %
- Downstream (2000–3000 bp): 0.6 %
- 5'UTR: 0.3 %
- 3'UTR: 2.4 %
- Coding exon: 8.2 %
- Intron: 69.8 %
- Distal intergenic: 13.7 %

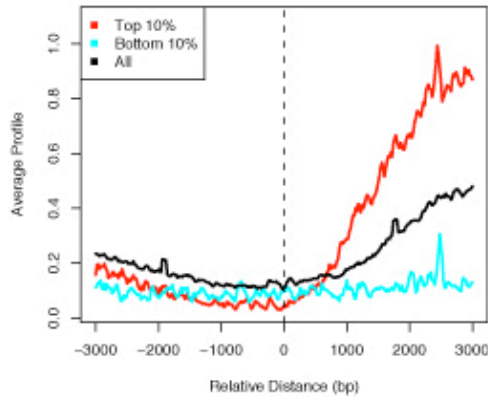
ChIP Regions (Peaks) over Chromosomes



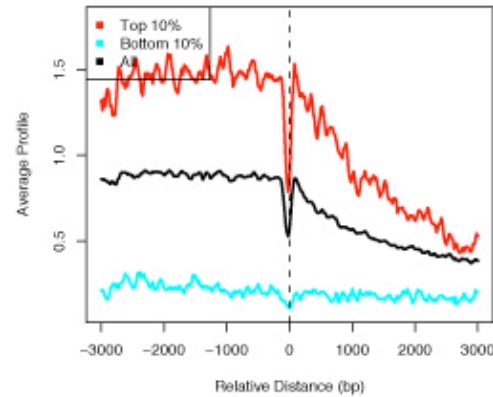


ChIP-seq analysis: Functional analysis

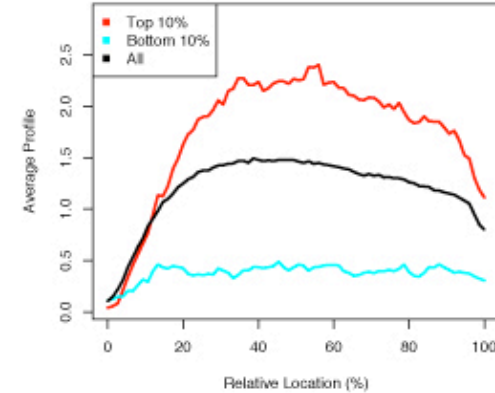
Average Profiles near TSS



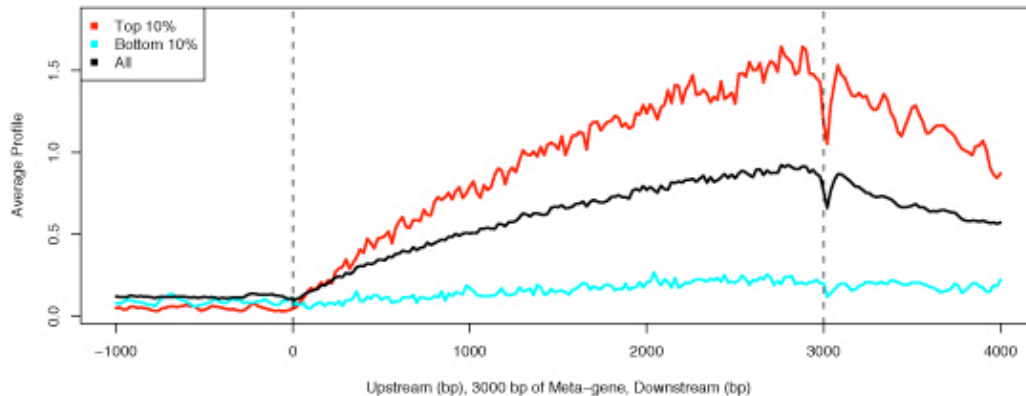
Average Profiles near TTS



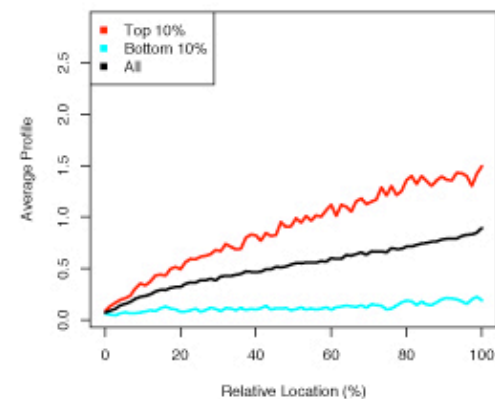
Average Concatenated Exon Profiles



Average Gene Profiles



Average Concatenated Intron Profiles





ChIP-seq analysis: Functional analysis

- `ceas --name=H3K36me3_ceas`
- `--pf-res=20`
- `--gn-groups=top10.txt,bottom10.txt`
- `--gn-group-names='Top 10%,Bottom 10%'`
- `-g ./hg18/refGene`
- `-b H3K36me3_MACS_pval1e-5_peaks.bed`
- `-w H3K36me3.wig`



ChIP-seq analysis: Functional analysis

Sitepro

Input:

ChIP
Regions
(BED file)

ChIP Signal
(WIG file)

ChIP Signal
(WIG file)



ChIP-seq analysis: Functional analysis

Gene Cluster 3.0

File loaded

Job name

Dataset has Rows Columns

Filter Data Adjust Data Hierarchical k-Means SOMs PCA

Filter Genes

☐ % Present >=

☐ SD (Gene Vector)

☐ At least observations with abs(Val) >=

☐ MaxVal - MinVal >=

