

SparseData

Mike Love

Max Planck Institute for Molecular Genetics, Berlin

December 12, 2012

motivation

UCSC Genome Browser on Human Feb. 2009 (GRCh37)

move <<< << < > >> >>> zoom in 1.5x 3x 10x (base)

chr10:118,018,279-118,126,340 108,062 bp. enter position, gene symbol or track name

chr10 (q25.3) p14 p13 12.1 10q21.3 23.1

Scale chr10: 118,030,000 118,040,000 118,050,000 118,060,000 118,070,000 118,080,000 118,090,000

UCSC Genes (RefSeq, UniProt, CCDS, Rfam, tRNAs & CDS) GFRF1 GFRF11 C19orf96

RefSeq Genes

Human mRNAs from GenBank

Human ESTs That Have Been Spliced

Layered H3K27Ac

Digital DNaseI Hypersensitivity Clusters

TF ChIP

Click on a feature for details. Click or drag in the base position track bars for track options. Drag side bars or labels up or down to reorder or right to new position.

track search default tracks default order hide all add custom tracks track hubs collapse all

Use drop-down controls below and press refresh to alter track options. Tracks with lots of items will automatically be displayed in more than one column.

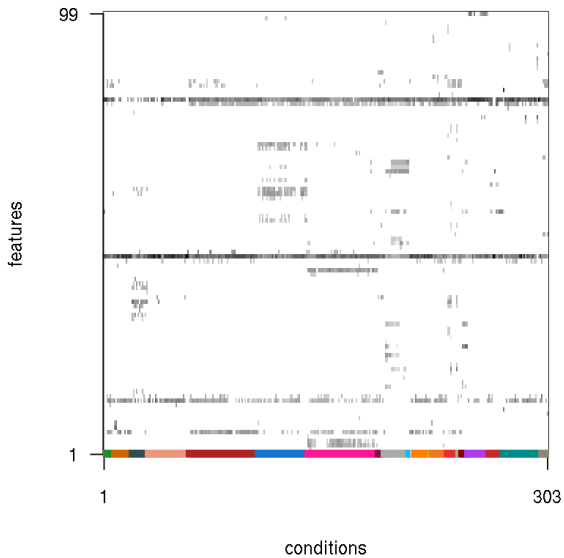
DNase Clusters item details

Items in Cluster: 25 of 148
 Cluster Score (out of 1000): 1000
 Position: [chr10:118076041-118076370](#)
 Band: 10q25.3
 Genomic Size: 330
[View DNA for this feature](#) (hg19/Human)

List of Items in Cluster

#	signal	cellType	treatment	replicate	lab
1	60.00	HAc	None	1	UW
2	27.00	HAc	None	2	UW
3	37.00	HA-h	None	1	UW
4	20.00	HA-h	None	2	UW
5	16.00	HBMEC	None	1	UW
6	11.00	HFF-Myc	None	1	UW
7	61.00	HRCEpiC	None	1	UW
8	56.00	HRCEpiC	None	2	UW
9	26.00	HRE	None	1	UW
10	15.00	HRE	None	2	UW
11	42.00	HRPEpiC	None	1	UW
12	37.00	HRPEpiC	None	2	UW

SparseDataSet



SparseData : basic statistics

- ▶ `sds <- calculateMeans(sds, nonzero.ratio=.1)`

SparseData : basic statistics

- ▶ `sds <- calculateMeans(sds, nonzero.ratio=.1)`
- ▶ `sds <- calculateTStats(sds)`

SparseData : basic statistics

- ▶ `sds <- calculateMeans(sds, nonzero.ratio=.1)`
- ▶ `sds <- calculateTStats(sds)`
- ▶ `fData(sds)[order(-tStats(sds)[["brain"]]),]`

SparseData : combine

- ▶ `z <- combine(x,y)`
- ▶ `z <- calculateMeans(z)`

SparseData : distances

- ▶ Pearson:

$$\frac{\langle \mathbf{x} - \bar{x}, \mathbf{y} - \bar{y} \rangle}{(n - 1)s_x s_y}$$

- ▶ Euclidean:

$$\|\mathbf{x} - \mathbf{y}\|$$

- ▶ cosine similarity:

$$\frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|}$$

- ▶ Pearson:

$$\frac{\langle \mathbf{x}, \mathbf{y} \rangle - n\bar{x}\bar{y}}{(n - 1)s_x s_y}$$

- ▶ Euclidean:

$$\sqrt{\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\langle \mathbf{x}, \mathbf{y} \rangle}$$