Package 'EPOM'

January 26, 2018

Type Package

Title EPOM for comparing tissue/cell types based on chromatin states

Version 0.1.0

Author Ruochen Jiang, Wei Vivian Li, Jessica Jingyi Li

Maintainer Ruochen Jiang <ruochenj@gmail.com>

Description It's package for the paper: Epigenome overlap measure (EPOM) for comparing tissue/ cell types based on chromatin states by Wei Vivian Li, Zahra S. Razaee and Jingyi Jessica Li (2016).
Including four steps:
1. transfer your bigwig file into 200 bp matrix.
2. Use anova to select regions that have significant difference in each cell type group.
3. Use t test to select enhancer regions for each cell type group.
4. Calculate EPOM score

License GPL

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1

Imports parallel

Suggests knitr, rmarkdown

VignetteBuilder knitr

R topics documented:

anova_select3 *Use ANOVA to select candidate associated regions that are significant*

Description

Step 2 in the EPOM method: use ANOVA to test whether a histone mark has the same group mean signals across different cell and tissue types.

Usage

```
anova_select3(cell_type, row_select, matrix_gen, state = "enhancer",
  alpha_1 = 1e-10, in_dir = NULL, save = TRUE, out_dir = NULL,
  cores = 16)
```

Arguments

cell_type	A vector indicating the cell type of each sample.
row_select	A vector indicating the row indicies corresponding to a certain state with respect to the original 200 bp matrix.
matrix_gen	A matrix which containing all the samples where the signal are correponding to the same bp interval. For example, if we have 127 available samples, the dimension of the matrix will be x*(3+127). x is the total length of 200 bp regions. First 3 columns are chrom name (eg. chr1), begin (eg. 201), width (eg. 200) and the other 127 are signals (eg. 0.04) of each sample.
alpha_1	A number indicating the significance level used in the ANOVA. The default is 1e-10, the larger this number is, the more regions will be selected.
in_dir	(Optional) Path for the .RData file to be read in.
save	A Boolean indicating whether the output should be saved.
out_dir	(Optional) Path for the output of the function.
cores	Indicate the number of cores you want to use in your server. The default value is 16.

Value

This function returns a list l (lower case "L") including (1) a vector called select_reg which indicates the indices for the regions in the original matrix. (2) a matrix called mat_anova which is the matrix after anova selection.

An RData file called "mat_anova4state.RData" (eg. "mat_anova425.RData" indicates the RData file after processing anova for state 25) includes the object l and l$select_reg corresponds to (1) and l$mat_anova corresponds to (2).

epom *Calculate the epom scores between every pair of samples.*

Description

This function outputs the epom scores for given parameters. It's package for the paper: Epigenome overlap measure (EPOM) for comparing tissue/cell types based on chromatin states by Wei Vivian Li, Zahra S. Razaee and Jingyi Jessica Li (2016).

Usage

```
epom(bw_list, bed_list, bd_header = FALSE, bed_sep = "\t", cell_type,
  histone_mark, state, state_name, in_dir = NULL, save = TRUE,
  out_dir = NULL, alpha_1 = 1e-10, cores = 16, alpha_2 = 0.01, m = 13)
```

Arguments

bw_list	A vector containing the full path and file name for all the big wig files.The length and order of bw_list, bed_list and bed_sep should be matched.
bed_list	A vector containing the full path and file name for all the bed files. The length and order of bw_list, bed_list and bed_sep should be matched.
bed_sep	This is used in reading in the bed file. The program uses read.table to read the bed files. This parameter corresponds to parameter sep in read.table. You can use ?read.table for further reference.
cell_type	A vector containing all the cell type information for each cell. The length and order of bw_list, bed_list and bed_sep should be matched.
histone_mark	A string indicating the current histone_mark you are processing
state	A vector indicating the chromatin states of interest.
state_name	A string indicating the name for the state. For example, "enhancer1".
in_dir	(Optional) Path of the .RData file to be read in.
save	A Boolean indicating whether the output should be saved.
out_dir	(Optional) Path of the output of the function.
cores	An integer indicating the number of cores to use in parallel computation. The default value is 16.
m	An integer indicating the significant values required for a region to be selected as the associated region for a cell type.
bed_head	This is used in reading in the bed file. The program uses read.table to read the bed files. This parameter corresponds to parameter header in read.table. You can use ?read.table for further reference.
bed_head	This is used in reading in the bed file. The program uses read.table to read the bed files. This parameter corresponds to parameter header in read.table. You can use ?read.table for further reference.
alpha1	A number indicating the significance level used in ANOVA. The default is 1e-10, the larger this number is, the more regions will be selected.
alpha2	A number indicating the significance level used in the t tests. The default is 0.01.

Value

A matrix output the epom score.

References

Li, Wei Vivian, Zahra S. Razaee, and Jingyi Jessica Li. "Epigenome overlap measure (EPOM) for comparing tissue/cell types based on chromatin states." BMC genomics 17.1 (2016): S10.

---

epom_score *Calculate epom_score after the vector chosen by t_test*

Description

Step 4, Calculate epom_score after the vector chosen by t_test.

Usage

```
epom_score(l_t, cell_type, save = TRUE, out_dir = NULL)
```

Arguments

l_t	A list of length of chosen groups which have more than one element. Each element of the list is a matrix which have three columns: column one indicates the selected indicies, column two indicates the chr, colun hte
cell_type	A vector indicates the cell type of the ith gene.

Value

A matrix output the epom score.

---

index_num4state *Select the corresponding regions of a specific chromatin state.*

Description

Select the corresponding regions of a specific chromatin state.

Usage

```
index_num4state(gen_mat, bed, state)
```

Arguments

gen_mat	A matrix generated from matrix_bp_trans.
bed	A character specifying the path of the bed file that contains the bed file specify the corresponding regions for each state.
state	A character specifying the chromatin state of interest.

Value

A vector indicating the rows inside the input gen_mat matrix that correspond to state-region information of input bed file.

matrix_bp_trans *Transform BigWig file into a matrix*

Description

Step 1, transform BigWig files into a matrix by averaging the signals within a certain bandwidth.

Usage

```
matrix_bp_trans(bw, interval)
```

Arguments

bw	A BigWig file to be transformed.
interval	An integer specifying the bandwidth used to average the bigwig signals. We set it to 200 in order to match with the state information.

Value

A matrix with chrom name (eg. chr1), begin (eg. 201), width (eg. 200) and signal (there is only one column of signal in the output) after transformation.

t_select3 *Use t test to select associated regions of a certian cell type.*

Description

Step 3 in the EPOM method: use t test to identify associated regions of a certain cell type.

Usage

```
t_select3(select_reg, mat_anova, cell_type, state = "enhancer",
  alpha_2 = 0.01, m = 13, in_dir = NULL, save = TRUE, out_dir = NULL,
  cores = 16)
```

Arguments

select_reg	A vector indicating the index of rows corresponding to the orignal transformed matrix after selecing by ANOVA.
mat_anova	A matrix containing the signals only on regions selected in ANOVA (rows for regions and columns for samples).
cell_type	A character vector indicating the cell type each sample belongs to.
state	A vector indicating the chromatin states of interest.
m	An integer indicating the significant values required for a region to be selected as the associated region for a cell type.
in_dir	(Optional) Path for the .RData file to be read in.
save	A Boolean indicating whether the output should be saved.

out_dir	(Optional) Path for the output of the function.
cores	Indicate the number of cores you want to use in your server. The default value is 16.
alpha2	A number indicating the significance level used in the t tests. The default is 0.01.

Value

A list l that contains selected enhancer region corresponding to the cell types which have more than one corresponding genes.

An RData file called "enhancer_signal.RData" will be saved corresponding to the result of the t test, each row indicates how t-test results are significant for each cell type.

An RData file called "mat_t.RData" will be saved which includes the list l, where l[[i]] corresponds to the selected region of cell type i. Some of l[[i]] might be NULL if the cell type has only one gene corresponding to it.

Index