

Program documentation

D.P.

February 18, 2015

extract_tails.py Identifies tails from every .fastq file in a directory. Usage:

```
python extract_tails.py -d directory
```

The -d parameter sets the folder with the fastq files of interest. If the program opens every .fastq file in the current directory if the -d parameter is not set.

If the -s option is not set, the program expects MS2-tagged tRNA ends.

The targetted RNA end can be given in a file passed with the -s filename argument.

If the targetted RNA is the yeast three-hybrid RNA, the -y3h_mode option should be set.

The -blanks option is currently not functional.

The format of a tails file is as follows. For a tRNA, we expect, with X = some sequence: file format:

```
XCCA
XCC
XCA
XC
X
```

ect. Tails are searched for in the order given in the file. The X is the fixed sequence and must be present in all forms of the tail end. Since these are present in every tail output, they are not included in the output. The variations after the fixed sequence are the variable sequence.

The program will extract tRNA 3'end sequences and generate a .inserts file for each .fastq file, containing the tRNA end, tail and random heptamer for each tRNA end.

It will then read each .inserts file and remove duplicates, using the random heptamer. It generates a .inserts.noDups file that contains the tRNA end, tail and number of tRNA ends (with different heptamers) containing that end type. The .inserts.noDups file is sorted by tRNA end, and then by length of insert.

Expected output:

```
A,TTT,3
```

Meaning:

tRNA end,tail,number of unique heptamers.

fileInserts.py This is a class used by other programs.

analyzeTails.py USAGE: python analyzeTails.py [options] Outputs a dataForGraph.txt file. If the control is subtracted, the dataForGraph.txt file is put in a controlSubtracted folder.

Options are kept the same as previous versions.

Automatically calls make_figs.py after running.

Example call: `python analyzeTails.py -a -n -p -c -i folder_of_inserts_files`

-i is a folder of inserts files. By default, the ./inserts/ will be searched.

-o will only use tails with a CCA end.

-n will normalize to dataset size (per 1 million).

-s will subtract a control.

-c is a control .inserts file to subtract.

-p will output values as percents.

-a will convert negative enrichment vs the control to zero. That is, there is no negative enrichment allowed.

-m will average the negative control files. If this option is used, the negative control files should be given as a list following the other inputs, rather than with the -c option.

make_figs.py Makes figures from dataForGraph files. USAGE: python make_figs.py -f data_file Takes one argument, which is the data file to process. Outputs to a figs/ folder.