Manual

Part I: Installation

Dependent tools:

- ✓ MCL: https://micans.org/mcl/
- ✓ BLAST: http://blast.ncbi.nlm.nih.gov
- ✓ QGRS: http://bioinformatics.ramapo.edu/QGRS/downloads.php
- ✓ Mfold: http://unafold.rna.albany.edu/?q=mfold
  Or ViennaRNA: http://www.tbi.univie.ac.at/RNA/

Part II: Inputs

Library sequences: at least 1 initial library and 2 enriched libraries are needed, and the sequenced sequences (without primers) are formatted in .txt files. One .txt file for each round.

Example: R2.txt

TGTTGTTTTTTCTGTGTTTTTGTTTGTGTGTGTTTTT
AAGCCCTTTCTCTGTCTAAGCAAACAAGGGGGGATTTGCA
ATCTTTGCTCTTCGGTTGTGTTTCGTTGTGGTTTTTTTT
TAGTTAATTGTTTTTCCTTTACTTTTACTTTATTTTTTTC
TTTGTTCGTGGTTGTGGTTTTTGTTTTTGTCTTTGGTTTT
GGCCCGCGCATTCCTTCCTATCCGGCCGTTTCTTCTGGTG
GTGTACATTCCACTCAAGTTTTGGGATGTCCCGTCTTTGC
CTTTTTTGTTGGTTTTTGGTTTTTTTTTTTTTTTTGTT
CTGCCAGTTGTCTAAGCGCGGCTCTCTTACGCGCACCTGC
GTTGTGTGTTTGTTTTTGTTTTGTTGTTTTGTTTTGTTGG
ATTTTGTTTGTTGTGTTTCTTTTTTTTTTTTTTTTTTTT
……

Primers: .txt file with primers sequences (first line: the forward primer; second line: the backward primer)

Example: Primer.txt

TTCAGCACTCCACGCATAGC
CCTATGCGTGCTACCGTGAA

List file: which contain the round name of used libraries

Example: Rlist

R4
R6
R9
R12

Part III: Run

The whole procedure of SMART-Aptamer is consisting of 4 steps as follow:

✓ Step 1: Calculate distributions of k-mer frequencies.

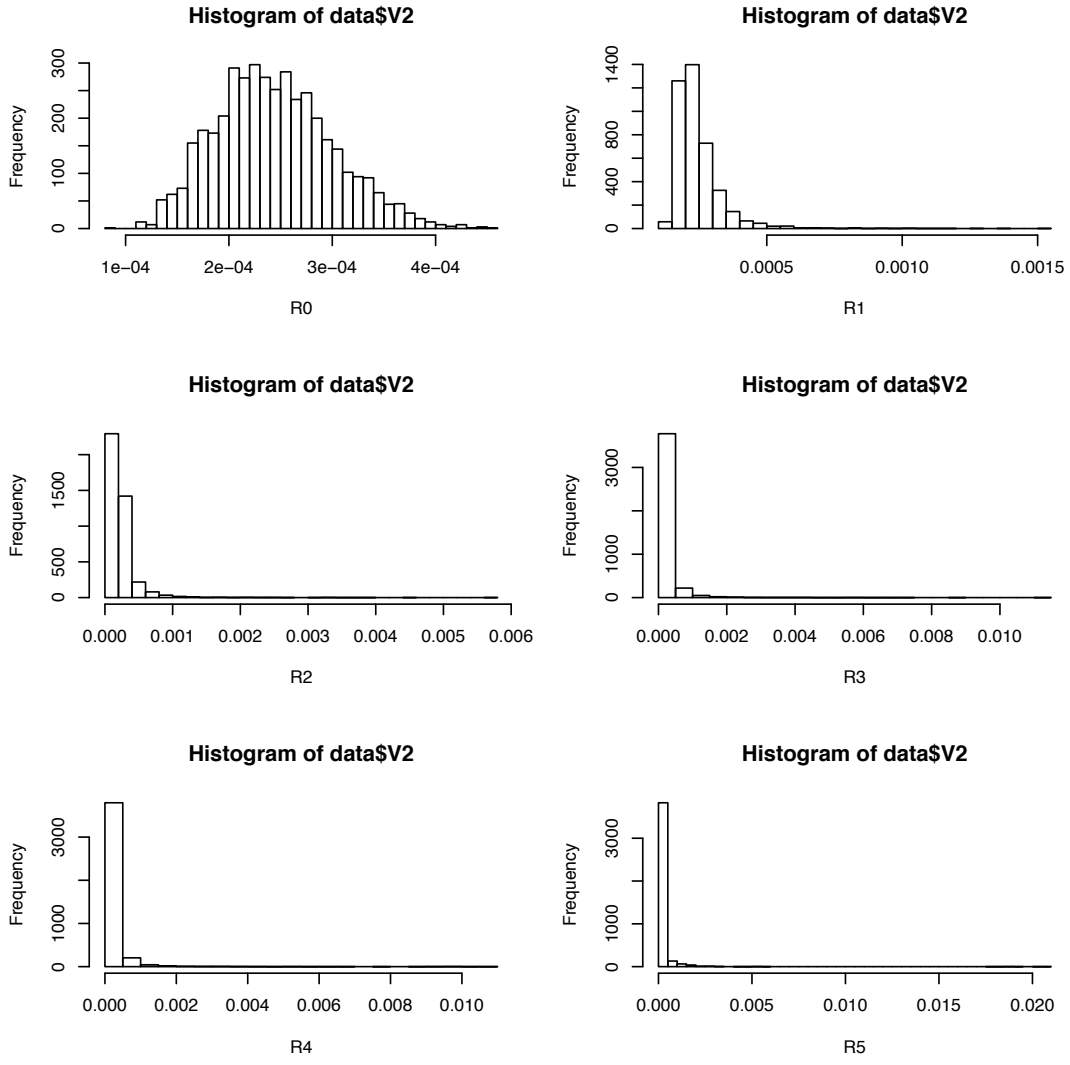
Example: Find_score -k 6 -t 35 -q 0.995 -c 'R4' -f input/Rlist -i input -d /home/songjiajia/test_data/WenKu/my_software/SMART-Apta -o output

usages: Find_score ...

-k, the predefined length of k-mers (default: 6)
-t, threads (default: 1)
-q, the quantile that used to define the enriched k-mers (default: 0.995)
-c, the control round (default R0)
-f, library list (default Rlist)
-i, input directory where the sequenced library txt files located (default: input
-o, output directory (default: result)
-d, the SMART-Apta directory

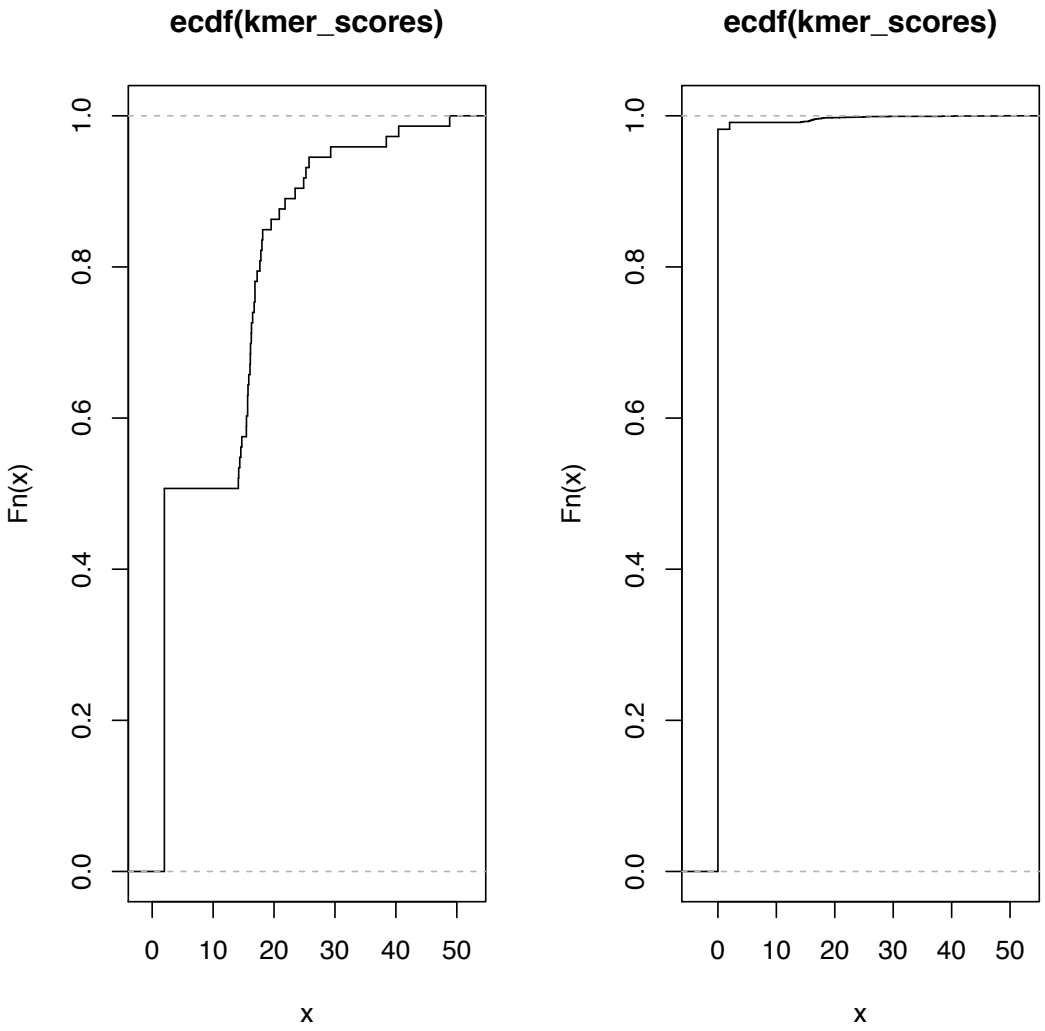
Important output files:

1) Kmer.pdf : can be used to understand the enrich status of each round



2) ecdf.pdf: used to find out the filter score cutoff value

for example: the following ecdf plots show that 10 can be a suitable cutoff value



✓ Step 2: find out the enriched motifs and filter data based on filter score

Example: SMART-motif -k 6 -t 35 -f input/Rlist -p 0.6 -s 2 -n 150000 -o output -i input

usages: SMART-motif ...

-k, the predefined length of k-mers (default: 6)

-t, threads (default: 1)

-f, library list (default Rlist)

-p, sequences with T bases > this cutoff are considered as t-rich sequences (0-1, default 0.6)

-s, the filter scores cutoff value used to filter sequences based on motif enrichment status (only need to set one between -s and -n, default 10)

-n, the total unique sequences should be left after filtering (set one between -s and -n, default 150000)

-o, output directory (default result)

-i, input directory (default input)

Important output files:

1) score_kmers.txt: the enriched motifs and their kscore

2) scores.txt: the calculated filter score for each aptamer
3) all_used_uniq.fasta: the sequences that kept after data filtering
4) all_used_info.txt: the frequency information of the kept sequences

✓ Step 3: cluster aptamers based on BLAST-MCL strategy

Example: SMART-cluster -t 35 -o output -i 1.5 -e 0.05 -p input/primer.fa

usages: SMART-cluster ...

-t, threads (default: 1)
-o, output directory (default result)
-i, inflation value for mcl algorithm (default 1.5)
-e, cutoff e-value after blast results (default 0.05)
-p, the primer file

Important output files:

1) aptamer_clusters: the aptamer families

✓ Step 4: Multidimensional assessment including the calculation of Kscore, Sscore and Fscore

Example 1: SMART-MDA-RNAfold -k 6 -t 35 -c 25 -o output -i input -r 'R6:R9' -d '+:+' -f input/Rlist

Example 2: SMART-MDA-RNAfold -k 6 -t 35 -c 25 -o output -i input -r 'R12' -d '+' -f input/Rlist

Example 3: SMART-MDA-RNAfold -k 6 -t 35 -c 25 -o output -i input -f input/Rlist

usages: SMART-MDA-RNAfold ...

-k, the predefined length of k-mers (default: 6)
-t, threads (default: 1)
-c, rescale energy parameters to a temperature of temp C. (default 37)
-o, output directory (default result)
-i input directory (default input)
-r rounds where the selection pressure has changed lead to a fixed family expansion/reduction trend; In case of multiple rounds, use ':' separate; For example, in this study ,we changed the selection pressure from 12th round SELEX, thus we set -r '12'
-d Use the +/- symbol to mark family expansion/reduction; In case of multiple rounds, use ':' separate
-f library list (default Rlist)

Important output files:

1) **result_aptamer.txt:** the final results, which contain the kscore, fscore, sscore and MDA-score and the representative sequence of each aptamer families. The outputs have been ranked by the MDA-score.
2) family_size_rounds.txt: Contains the size/per round of each aptamer family.

3) conflict_trends.txt: contain the kscore, fscore, sscore, MDA-score and the representative sequence of aptamer families (whose family expansion/ reduction trends conflict with the experimental design).