

CisModScan (Linux or OS executable) user manual and examples

Put CisModScanU.tar/CisModScanOS.tar file in a Linux/OS directory and use "tar xvf CisModScanU.tar" or "tar xvf CisModScanOS.tar" to unpack the file. Then you will find an executable file (CisModScanU or CisModScanOS) in the same folder. In the following, we will use CisModScanOS as an illustration (it is the same for CisModScanU).

Command line

Type ./CisModScanOS, you will see a simple user manual. To run CisModScan, an input sequence file and a weight matrix file must be specified. An example command is

```
./CisModScanOS -i cm_testdata.txt -o cmscan_testresult -K 2 -L 200 -M testPWM.txt  
-x 0.001
```

Under this command, we are running CisModScan on a data set cm_testdata.txt (which can be downloaded in the CisModScan webpage). The input data set should be written in FASTA format and each line in the file should contain no more than 10K characters. The input weight matrix file (-M) contains the known weight matrices for the motifs (please download testPWM.txt in the CisModScan webpage). In this file, each weight matrix starts from a head line of ">motif width" and the following rows contains the weights of A,C,G,T in each position separated by tab. For instance,

```
>8  
0.521  0.146  0.236  0.097  
0.000  0.993  0.000  0.007  
0.736  0.049  0.000  0.215  
0.014  0.000  0.979  0.007  
0.118  0.507  0.007  0.368  
0.097  0.222  0.000  0.681  
0.000  0.000  0.986  0.014  
0.208  0.583  0.035  0.174
```

This motif is of width 8, and the weights of the first position is 0.521 for A, 0.146 for C, 0.236 for G, and 0.097 for T. (The weights are not necessarily normalized.)

The program will search modules containing binding sites of two different motifs (K=2) and the module length L = 200. The output modules will be written in the file cmscan_testresult _module.txt (-o) and the motif sites will be reported to the files cmscan_testresult _motif1.txt and cmscan_testresult _motif2.txt, respectively. The format of these results are similar to those used in CisModule. (Please see next section for details). By this setting, CisModScan will search both strands of sequences (default setting). If you specify "-r 0", then only forward strand will be searched.

Output

1) Format for module information:

One module will be represented by one line. For example:

```
">ZK721.2 [769,970] -2(775)+2(822)-1(946)"
```

This means that in the sequence ZK721.2, there exists one module located between positions 769 and 970. This module contains one backward site for motif 2 (-2) at position 775, one forward site for motif 2 (+2) at position 822, and one backward site for motif 1 at position 946.

2) Format for motif information:

Each motif is represented by a weight matrix followed by its sites in the sequences.

For example:

```
">R10H f 367 0.606
```

```
GTGCTCCCCAGAG"
```

This item says that in the forward strand of the sequence R10H at position 367 there is one motif site GTGCTCCCCAGAG whose posterior probability of being sampled as a motif site is 0.606.