

## 1. Numster

Numster is a nucleosome mapping program using a hierarchical agglomerative clustering algorithm. It can be used for analyzing sequencing data from either MNase or hydroxyl radical treated samples. Software was written in C/C++ and tested on Debian 6.0.7 - Linux Kernel 2.6.32-5-amd64, with “Pthread” to support multi-threading.

This binary file is the 64-bit Linux executable version of numster.

Note: A command `chmod +x numster` may be needed to execute the unzipped numster file.

## 2. Usage

*numster arguments*

## 3. Mandatory arguments

- i: Path of input file. String value
- m: 1 for analyzing sequencing data from MNase treated samples. 2 for analyzing sequencing data from hydroxyl radical treated samples. Integer value of either 1 or 2
- o: Path of output file. String value

## 4. Optional arguments

- a: Distance threshold ( $\alpha$ ) of  $D_{Li}$  or  $D_{Ri}$  between adjacent level 1 clusters. Double value. Default 0.5
- b: Interquartile range value ( $\beta$ ) for noisy cluster removal. Double value. Default 0.5
- d: Euclidean distance threshold ( $\theta$ ) for conflicting templates. Double value. Default 0.1
- e: Distance ( $\epsilon$ ) to the mode tag of level 1 cluster. Integer value. Default 35 for -m 1. Default 3 for -m 2
- n: Interquartile range value ( $\eta$ ) for noisy nucleosome removal. Double value. Default 0.5
- t: Number of threads. Integer value. Default 4
- l: Left boundary of nucleosomal DNA range. Integer value. Default 80 for -m 1. Default 140 for -m 2
- r: Right boundary of nucleosomal DNA range. Integer value. Default 200 for -m 1. Default 180 for -m 2
- v: Maximum nucleosome overlapping rate. Double value. Default 0.3 (30% overlap)

## 5. Input file format

Tab-delimited text file contains the following columns

- Chr: Chromosome number. Integer type
- Start: Start location of tags. Integer type
- Str: F(forward strand) or R(reverse strand). Character type of either F or R
- Read: Number of tags starting at the location. Integer type

## 6. Input file example

Chr	Start	Str	Read
1	265	F	1
1	281	F	3
1	285	R	2

## 7. Testing numster

This software includes two test files, a MNase data file derived from Chromosome I of the Kaplan data (“test/mn\_tag.txt”) and a hydroxyl radical data file derived from Chromosome I of the Brogaard data (“test/ho\_tag.txt”).

```
numster -m 1 -t 1 -i test/mn_tag.txt -o test/mn_nuc.txt
```

- This command reads the “mn\_tag.txt” file under the “test” directory and runs the MNase nucleosome mapping method with one thread. The output of predicted nucleosomes can be found in the “mn\_nuc.txt” file under the “test” directory.

```
numster -m 2 -t 1 -i test/ho_tag.txt -o test/ho_nuc.txt
```

- This command reads the “ho\_tag.txt” file under the “test” directory and runs the hydroxyl radical nucleosome mapping method with one thread. The output of predicted nucleosomes can be found in the “ho\_nuc.txt” file under the “test” directory.

## 8. Usage examples

```
numster -m 1 -i input_file.txt -o output_file.txt
```

```
numster -m 2 -i input_file.txt -o output_file.txt
```

```
numster -m 1 -v 0.5 -i input_file.txt -o output_file.txt
```

```
numster -m 1 -t 2 -i input_file.txt -o output_file.txt
```