MCORE Manual v05

Content

What is MCORE?

Multi-scale correlation evaluation (MCORE) is a method to robustly identify and compare patterns in deep sequencing data sets by calculating correlation functions. MCORE avoids assumptions on the shape or the dimension of enriched regions and evaluates all mapped sequencing reads without filtering. Therefore, MCORE is a powerful tool for the unbiased identification of relationships between chromatin features that can be mapped by deep-sequencing approaches in mouse cells.

The MCORE analysis is introduced and described in the following paper:

Molitor J, Mallm JP, Rippe K & Erdel F (2016). Retrieving the topology of chromatin domains from deep sequencing data with correlation functions.
doi: http://dx.doi.org/10.1101/054049
http://biorxiv.org/content/early/2016/05/18/054049

Download

http://malone.bioquant.uni-heidelberg.de/software/MCORE.html

Please note that the zip file with the program require a password that can be requested as described on that page.

Installing MCORE

The MCORE_java folder contains the executable program (MCORE.jar), the manual, the data folder with a test data set and a lib folder.

- copy the folder ‘data’ that contains test data to your home directory
- copy MCORE.jar and the lib folder to a place of your choice
- start MCORE by using a command line tool and typing:
    - java -Xms64m -Xmx4096m -jar MCORE.jar

Depending on the length of the genomic region to be analyzed more or less than 4 GB memory are required, which means that the –Xmx option has to be adjusted.

Data formats read by MCORE

Bed files

Requirements for the file:

- More than 5 columns
- Information about chromosome number, start, end, and strand

Example:

```
chr7    34319858        34319909        U0      0       +
chr9    25377952        25378003        U0      0       +
chr9    25862088        25862139        U0      0       +
chr1    138880978       138881029       U0      0       +
chr13   57453588        57453639        U0      0       +
chr7    113875620       113875671       U0      0       -
chr10   66381627        66381678        U0      0       -
chr4    45593508        45593559        U0      0       +
chrX    83288161        83288212        U0      0       +
chr1    99625536        99625587        U0      0       -
```

5mC

Requirements for the file:

- More than 4 columns
- Column 1: chr
- Column 2: start
- Column 4: total reads
- Column 5: converted reads

Example:

```
chr1    3000574 3000574 37      33
chr1    3000726 3000726 5       5
chr1    3000901 3000901 2       0
chr1    3001346 3001346 13      12
chr1    3001394 3001394 17      12
chr1    3001631 3001631 11      10
chr1    3002177 3002177 16      16
chr1    3002338 3002338 13      11
chr1    3002599 3002599 21      20
chr1    3002921 3002921 13      11
```

HiC

Requirements for the file:

- More than 6 columns
- Column 2: chr of first read
- Column 3: start of first read
- Column 4: strand of first read
- Column 5: chr of second read
- Column 6: start of second read
- Column 7: strand of second read

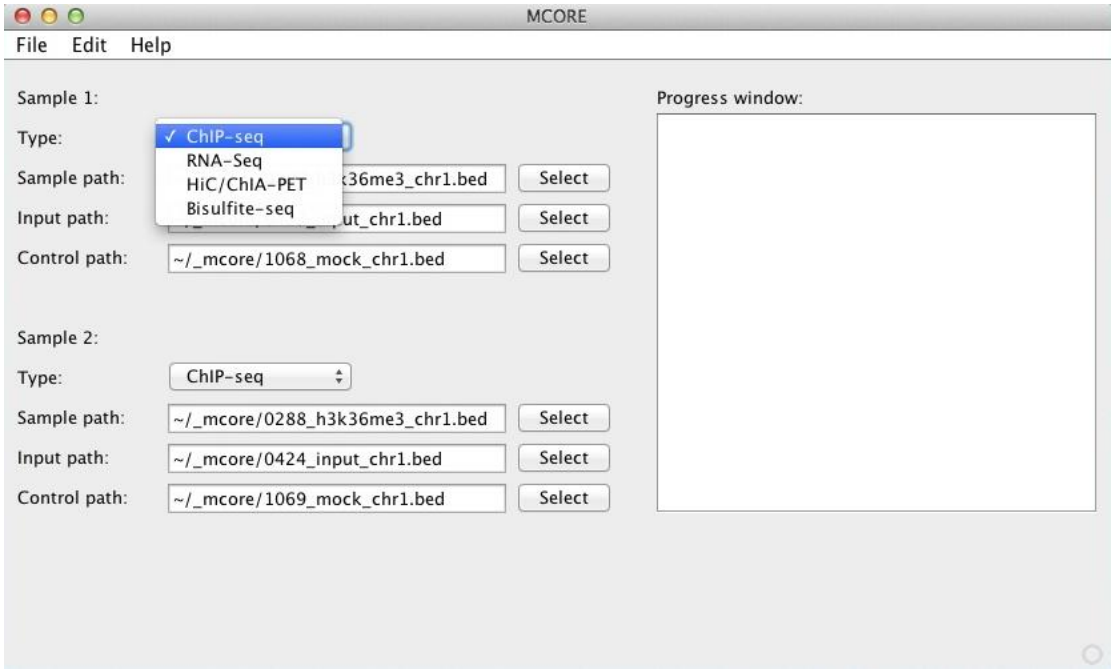
Example:

```
1104:16545:105833#AGTAAG        chr1    3000000 -       chr1    5404761 +
2208:8611:50989#AAATGA chr1     3000001 +       chr1    3000252 -
2307:15998:173700#GGTTGT        chr1    3000001 +       chr1    3000218 -
2206:8917:143447#0      chr1    3000001 -       chr1    3002321 +
2107:20574:193757#0     chr1    3000004 -       chr1    27496762        +
2108:14268:179950#GGTAGA        chr1    3000005 -       chr12   67441143        -
1201:5022:179023#..GG.A         chr1    3000008 -       chr3    107566242       -
2308:8270:100711#GGGTAT         chr1    3000009 -       chr17   76806821        +
2103:6453:156435#..C.GT         chr1    3000010 +       chr1    3000305 -
2302:13031:23416#......         chr1    3000010 -       chr2    48724064        -
```

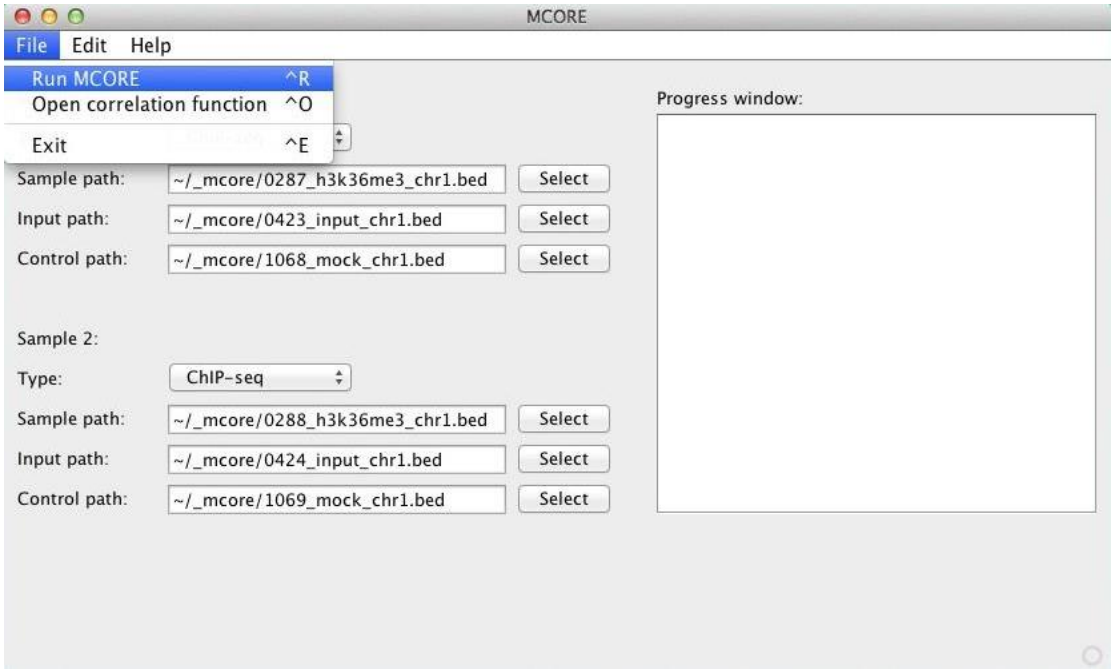
Calculations by MCORE

Run MCORE

Choose input files and select the type of data in the drop down menu.

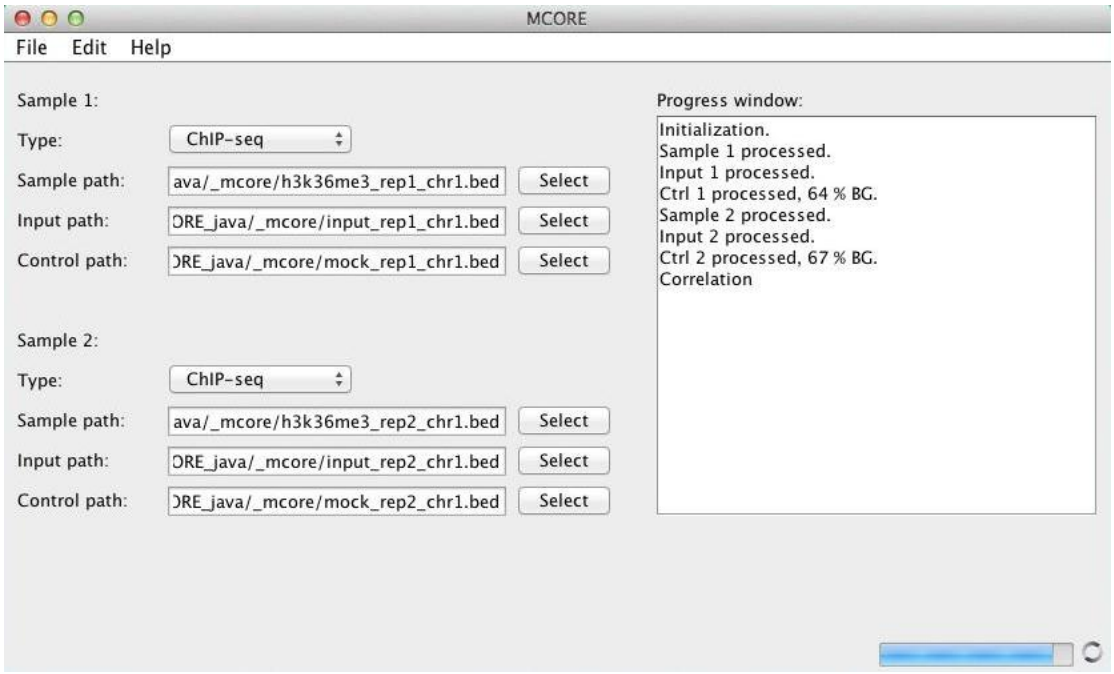


Start calculation by selecting File → Run MCORE.



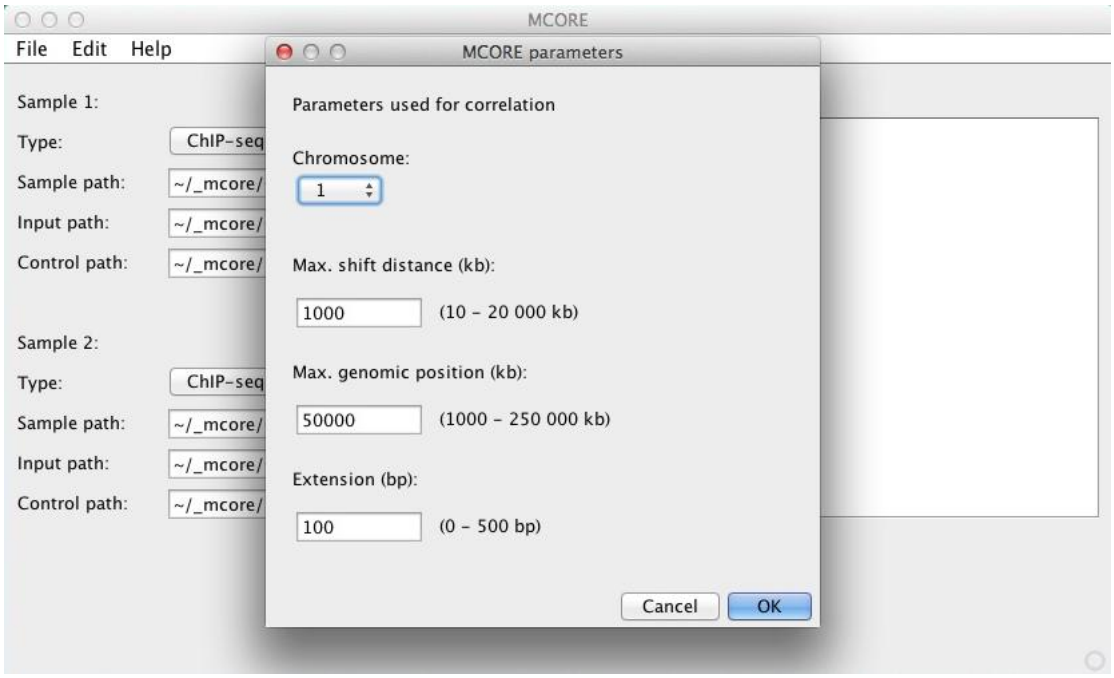
Progress window

The progress of the calculation is shown in the progress window. The weighting factor for the subtracted control sample is given as BG in %.



Optional: Select Parameters for calculation

Click on Edit → Parameters to open the Parameter window.



Select the following parameters

- chromosome (default: chr1)
- maximum shift distance for the correlation function (default: 1000 kb)

- maximum genomic position below which reads are considered (default: 50 000 kb)
- extension, by which the reads should be elongated to account for fragment size after sample preparation (default: 100 bp).

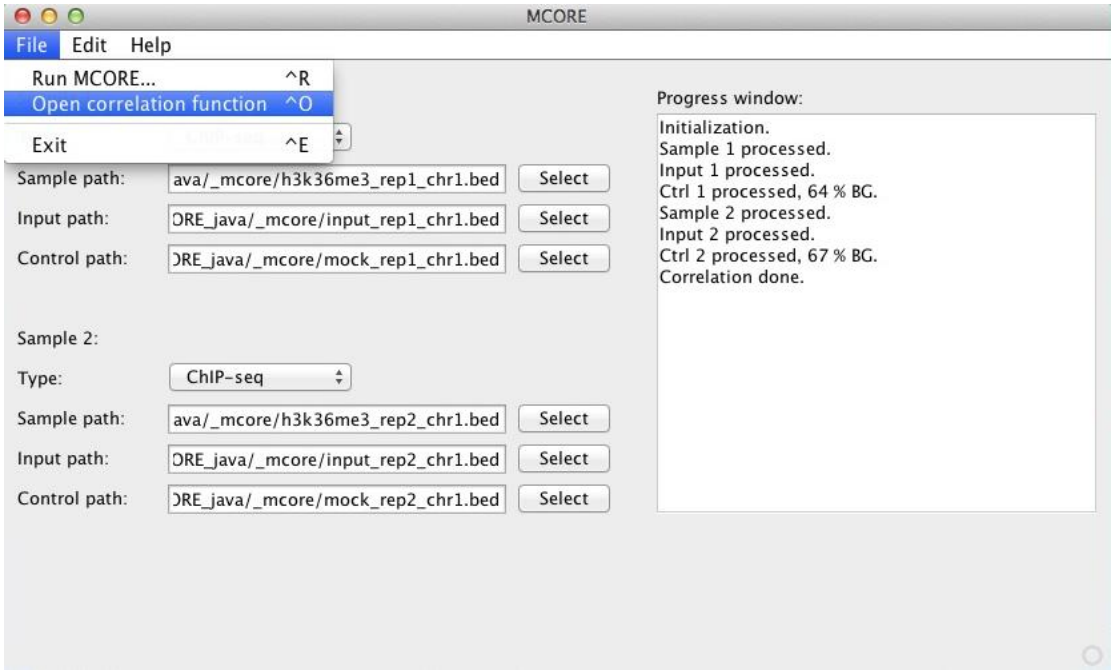
Optional: Select default paths

Click on Edit → Default paths to open the Default paths window.

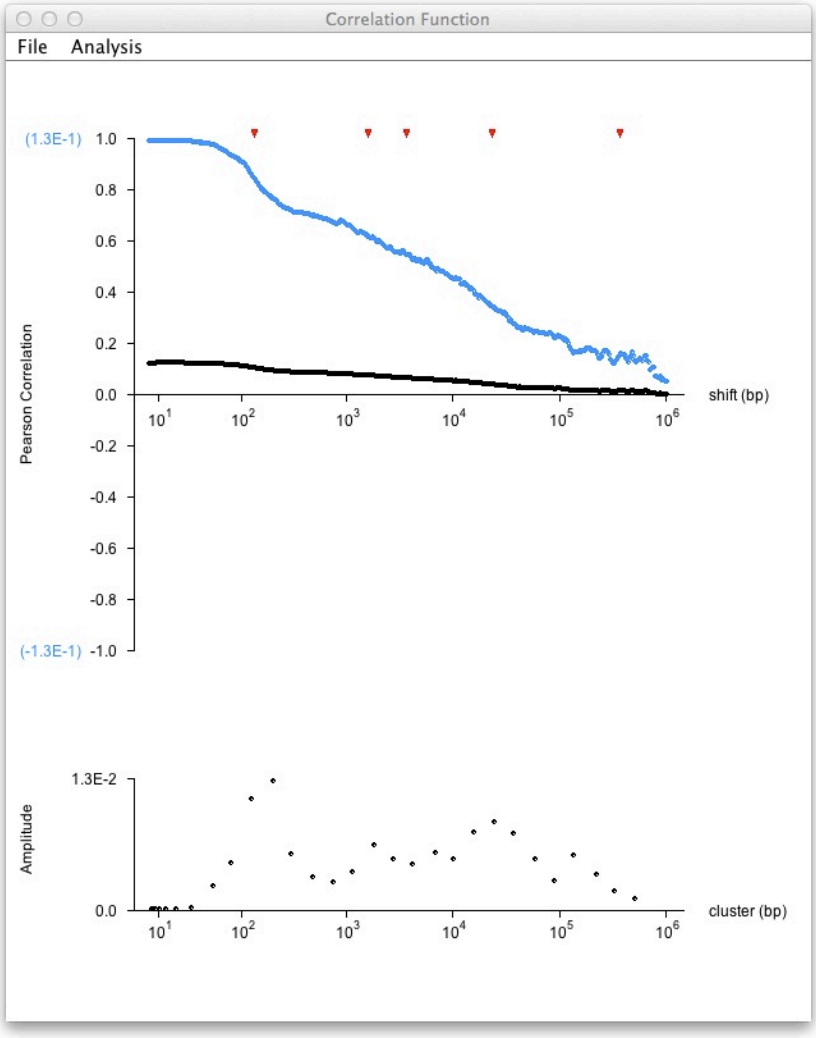
For quick access of input and output data, default input and storage paths for correlation functions, fit results and occupancy profiles can be inserted.

Optional: Open correlation functions

Click on File → Open correlation function to load a previously calculated function.



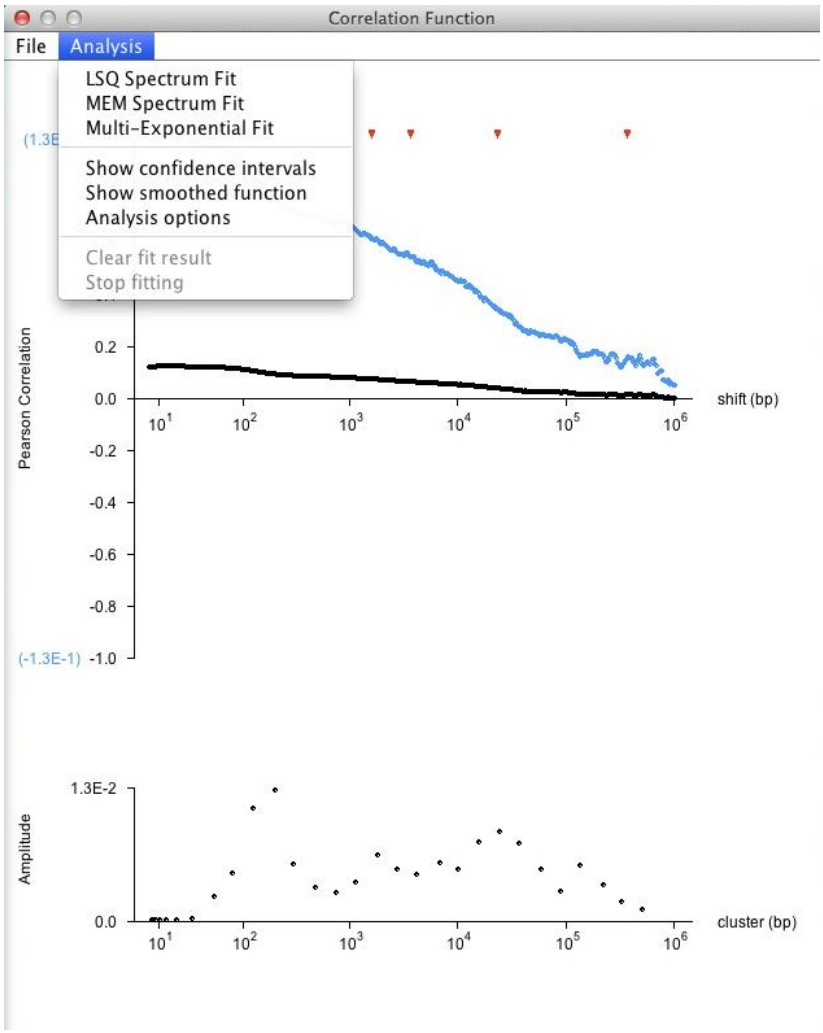
Correlation function



The calculated correlation function (black) is shown in the plot window. The blue curve corresponds to the rescaled correlation function (see blue scale). The red arrowheads above the curve mark inflection points corresponding to local maxima of the slope. Below the curve the decay spectrum is shown. Points in the spectrum reflect the abundance of domains on the respective scale.

Analysis of the correlation function

Analysis options for the correlation function are listed in the Analysis menu of the plot dialog.

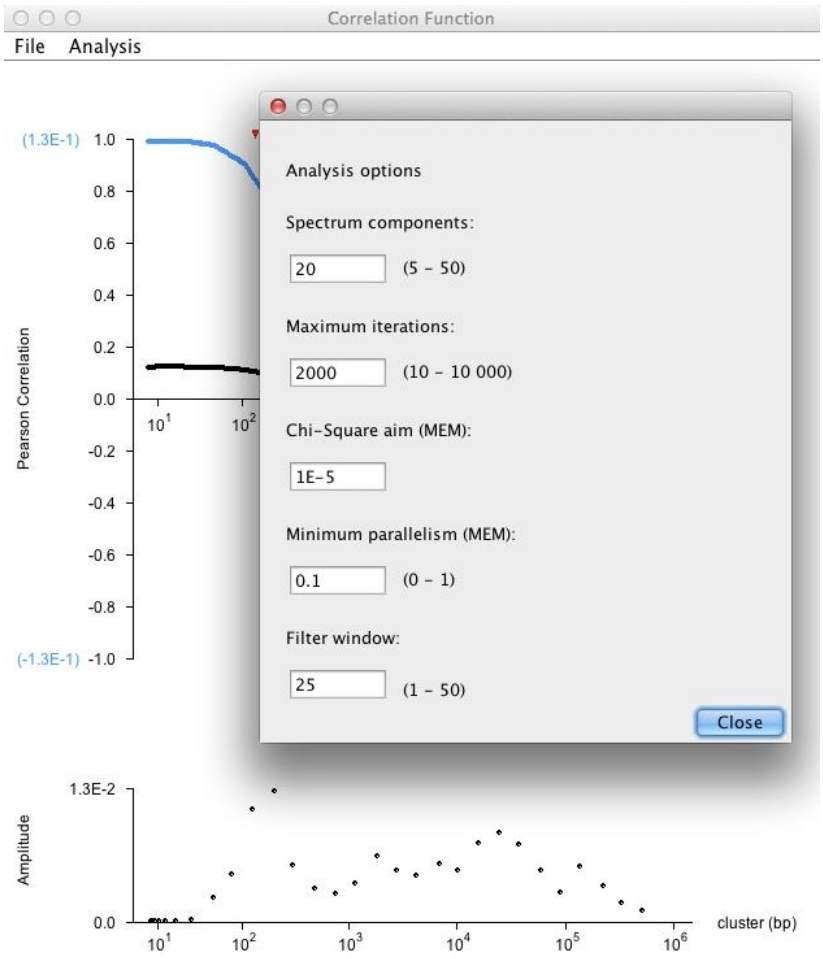


The decay spectrum of the correlation function is automatically analyzed by numerically determining inflection points that reflect shift distances with maximum slope. Additionally, the exponential decay spectrum according to a Gardner transformation is calculated. Both methods are complementary approaches to obtain the decay length distribution of the function, which correspond to domain sizes of chromatin features. Optionally, the exponential decay spectrum can be fitted by a least-squares (LSQ) algorithm or by a maximum entropy method (MEM), and the correlation function can be fitted to a model function as described in the Methods section of the paper. The fit algorithms can be stopped and the current fit result can be cleared to switch back to the Gardner transformation. Further, a smoothed correlation curve and the standard error can be displayed.

Optional: Analysis options

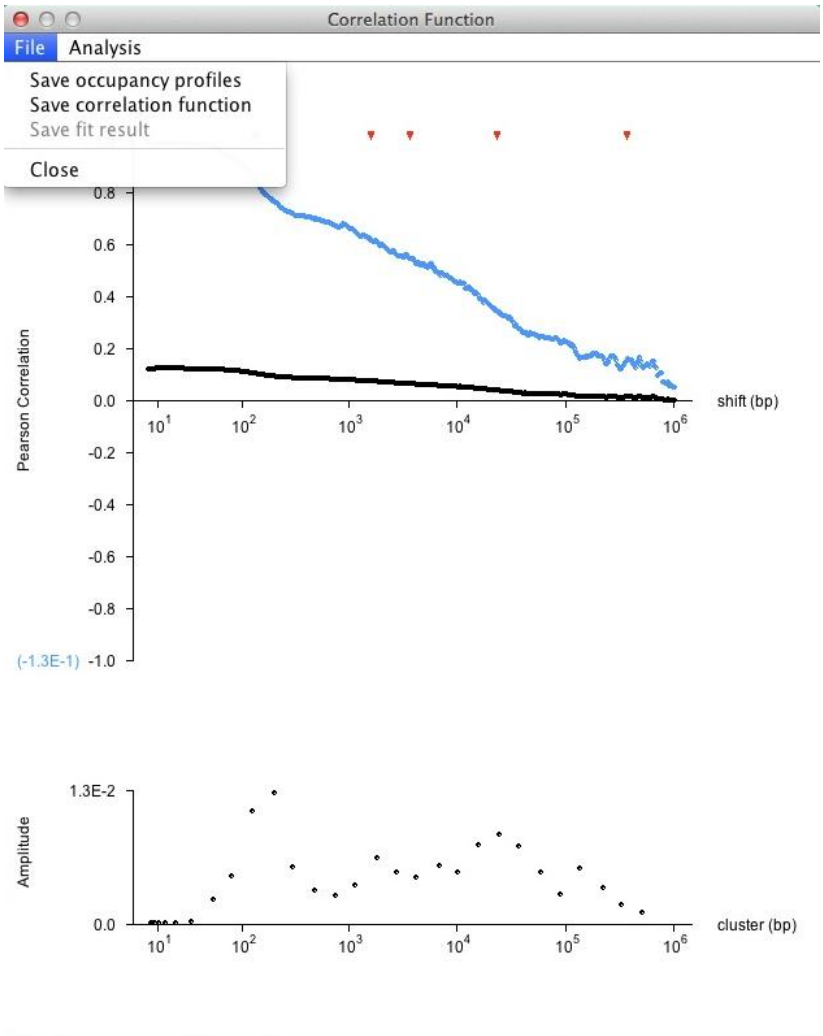
In the analysis options menu of the plot dialog, the parameters for the various analysis options of the correlation function can be adjusted. These include the number of components that are shown in the fitted spectra:

- the maximum number of fit iterations
- the fit quality the MEM algorithm aims at (Chi-Square aim)
- the minimum parallelism between the gradients of entropy and fit quality that is tolerated during fitting
- the window size for the Savitzky-Golay filter that is used to calculate the smoothed correlation function, which serves as a basis for finding inflection points and calculating the decay spectrum according to the Gardner transformation. For noisy functions, it is advisable to increase the window size for the filter.



Optional: Save results

Options for saving normalized occupancy profiles, correlation functions and fit results are found in the File menu.



Frequently asked questions

What can I do if below the progress window the progress bar is not shown and the progress circle is not running?

MCORE aborted its calculation. This is mostly due to insufficient memory. You will get an error massage on the command line (OutOfMemoryError). You can assign more memory or reduce the length of the maximal genomic regions to be read in Edit/MCORE parameters.

Why can't I start a new calculation after the fitting?

The correlation function window has to be closed before a new calculation can be run.

How can I retrieve information on the samples analyzed in the active correlation function window?

The sample names as well as selected parameters and results can be found in the output file.

Support & feedback

For question and feedback please contact mcore@dkfz.de.