

Manual OccuPeak V2

Installation

MCR installation

OccuPeak is a Matlab based program. To be able to run the application the proper version of the MATLAB Compiler Runtime (MCR) has to be installed. For windows install the 32-bit version for R2012b and for Linux environments install the 64-bits version of 2013a. This can be done by clicking this [link](http://www.mathworks.com/products/compiler/mcr/) (<http://www.mathworks.com/products/compiler/mcr/>) and following the provided instructions.

OccuPeak package

OccuPeak is distributed as a self-extracting package. This package can be extracted to a location of your choice. Do not move the program to another directory because the program and the Genome subdirectory are required to remain at the same location. The Genome subdirectory lists text files containing chromosome lengths. For mouse genome version 9 (mm9), the file has been provided with the program package. When using data from other genomes, such files need to be placed in the Genome subdirectory and the OccuPeak program will detect them automatically.

Using OccuPeak

OccuPeak requires Sequence Alignment Map (SAM) files as input. These SAM files are the default output of the Bowtie mapping program and these files do not require to be sorted.

Settings

Log(ER) Threshold

Only peaks above this threshold will be included in the output BED file.

Log(ER) categories colors

In this control the color categories can be defined which will be included in the output BED file. Right click on the colored areas to change the color

Window size

The highly recommended default setting of the program uses windows per complete chromosome to build a noise model. The user can override this setting and choose a window size between 0.1 and 100 Mb to build the model.

Genome

Here you can select one of the files in the Genome directory of the program which is used to determine the size of the Chromosomes.

Output of OccuPeak

OccuPeak writes the identified peaks to a text file in BED format (<http://genome.ucsc.edu/FAQ/FAQformat.html#format1>). The header of this file contains information on the reconstructed fragment length and applied ER threshold. In the body of the file, the first three columns give chromosome, start coordinate and end coordinate of a region

that is identified as a peak. The fourth BED column contains the surface area of the peak in bp, calculated as the sum of the lengths of the overlapping tags in this region. The fifth column reports the corresponding log(ER) value. Columns 6, 7 and 8 are not used. Column 9 contains RGB values corresponding to user defined settings to distinguish different ER categories.

Save summary

Optionally the displayed summary per window can be saved as an Excel (.xls) file.

Trouble Shooting

The program does not give any results

- Check whether you have write permission at the location or you have enough disk space.
- The input file might be corrupted. This can be checked by opening the file in a text editor and check whether the last line is complete.