MAGeCK.Count

This command collects sgRNA read count information from fastq files. The output count tables can be used directly in the Mageck test command.

Author
Clarence Mah;Mesirov Lab

Contact
ckmah@ucsd.edu

Algorithm version
0.5.5

Task Type
CRISPR

CPU Type
Any

Operating System
Any

Language
Python, C

References

Li, W., Xu, H., Xiao, T., Cong, L., Love, M. I., Zhang, F., ... Mesirov, J. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biology 2014 15:12*, *15*(12), 819–823. https://doi.org/10.1126/SCIENCE.1231143

Input Files

1. Fastq files

Sample fastq/fastq.gz files. Each group will be considered as a separate sample while each fastq in a group indicates technical replicates of the same sample. Use the “Group name” field to indicate sample names.

2. sgRNA list

When starting from fastq files, MAGeCK needs to know the sgRNA sequence and its targeting gene. The sgRNA library file can be provided either in .txt format or in .csv format. There are three columns in the library file: the sgRNA ID, the sequence, and the gene it is targeting.

The contents of each column are as follows:

Column	Content
id	The sgRNA ID
sequence	The sgRNA sequence
gene	The sgRNA's target gene

Output Files

1. Count Table

The sgRNA read count file should list the names of the sgRNA, the gene it is targeting, followed by the read counts in each sample. Each item should be separated by the tab ('\t'). A header line is optional. For example in the studies of T. Wang et al. Science 2014, there are 4 CRISPR screening samples, and they are labeled as: HL60.initial, KBM7.initial, HL60.final, KBM7.final. Here are a few lines of the read count file:

sgRNA	gene	HL60.intial	KBM7.initial	HL60.final	KBM7.final
A1CF_m52595977	A1CF	213	274	883	175
A1CF_m52596017	A1CF	294	412	1554	891
A1CF_m52596056	A1CF	421	368	566	59
A1CF_m52603842	A1CF	274	243	314	55
A1CF_m52603847	A1CF	0	50	145	66

2. Log file

This file includes the logging information during the execution. For count command, it will list some basic statistics of the dataset at the end, including the number of reads, the number of reads mapped to the library, the number of zero-count sgRNAs, etc.

3. Intermediate files
These files will be automatically deleted after the completion of each command. To keep these files, use the "keep intermediate files" option during the execution.

Example Data

http:/github.com/ckmah/mageck_test/blob/master/data/test1.fastq
http:/github.com/ckmah/mageck_test/blob/master/data/test2.fastq
http:/github.com/ckmah/mageck_test/blob/master/data/library.txt

Requirements

MAGeCK can be run on either Mac or Linux system. Since MAGeCK is written in Python and C, Python 2.7 (>2.7) and a C compiler is needed.

Module Parameters

*required

Name	**Description**
Fastq files*	Sample fastq/fastq.gz files separated by space; use comma (,) to indicate technical replicates of the same sample. For example, "sample1_replicate1.fastq,sample1_replicate2.fastq sample2_replicate1.fastq,sample2_replicate2.fastq" indicates two samples with 2 technical replicates for each sample.
sgRNA list	A file containing list of sgRNA names, the sequences and target genes, either in .txt or in .csv format. If this file is not provided, mageck will count all possible sgRNAs in the fastq.
Output prefix	The prefix of the output file(s). Default sample1.
Trim 5 prime	Length of trimming the 5' of the reads. Default 0.
sgRNA length	Length of the sgRNA. The the program will automatically determine the sgRNA length from library file; so only use this if you turn on the “save unmapped reads” option.
Count n	Count sgRNAs with Ns. By default, sgRNAs containing Ns will be discarded.
Save unmapped reads	Save unmapped reads to file.
Reverse complement	Reverse complement the sequences in library for read mapping.

Keep intermediate files	Keep intermediate files.
Normalization method	Method for normalization, including "none" (no normalization), "median" (median normalization, default), "total" (normalization by total read counts), "control" (normalization by control sgRNAs specified by the control sgRNA option).
Control sgRNA	A list of control sgRNAs for normalization and for generating the null distribution of RRA.