

# MAGeCK Count Report

Wei Li

March 25, 2018

## Contents

1	Summary	1
2	Normalized read count distribution of all samples	3
3	Principle Component Analysis	5
4	Sample clustering	7

## 1 Summary

The statistics of comparisons are listed in Table 1 and Table 2. The corresponding fastq files in each row are listed in Table 3.

	Label	Reads	Mapped	Percentage
1	test1_fastq_gz	2500	1453	0.58
2	test2_fastq_gz	2500	1471	0.59

Table 1: Summary of comparisons

	Label	TotalsgRNA	ZeroCounts	GiniIndex
1	test1_fastq_gz	2550	1276	0.53
2	test2_fastq_gz	2550	1199	0.49

Table 2: Summary of comparisons

	File	Label
1	input_0.gz	test1_fastq_gz
2	input_1.gz	test2_fastq_gz

Table 3: Summary of samples

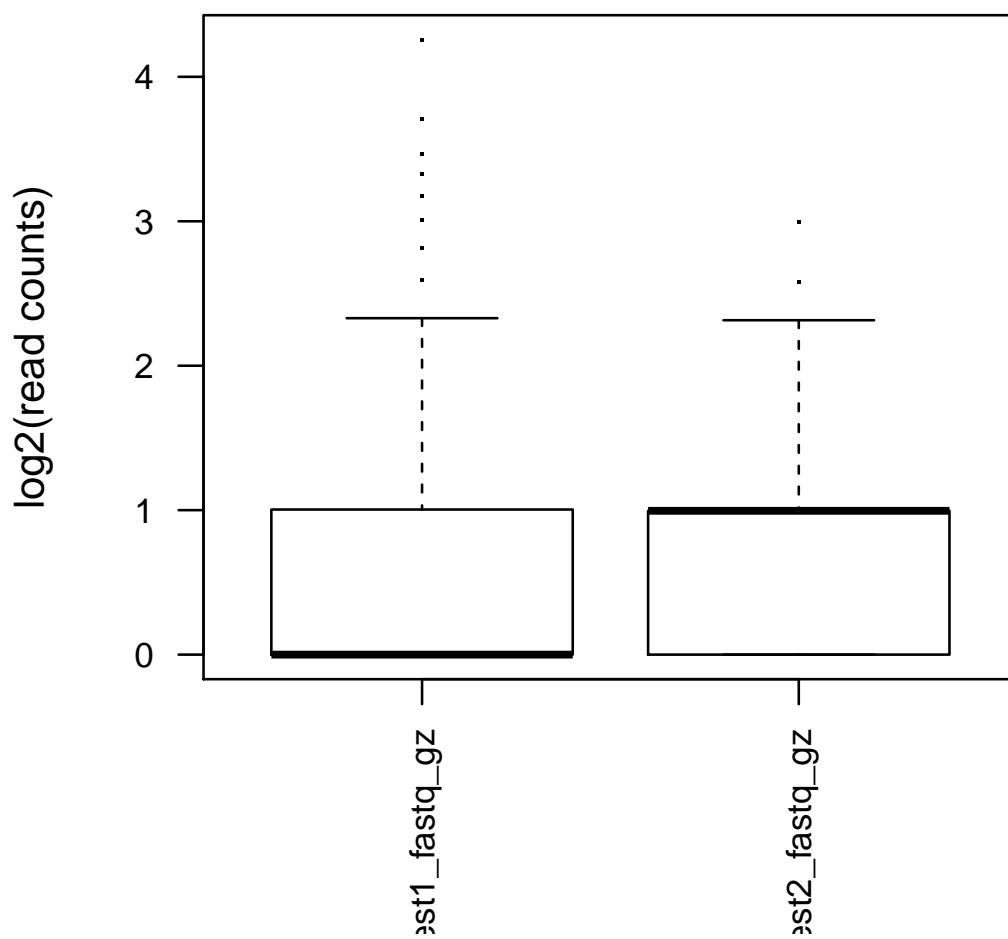
The meanings of the columns are as follows.

- **Row:** The row number in the table;
- **File:** The filename of fastq file;

- **Label:** Assigned label;
- **Reads:** The total read count in the fastq file;
- **Mapped:** Reads that can be mapped to gRNA library;
- **Percentage:** The percentage of mapped reads;
- **TotalsgRNAs:** The number of sgRNAs in the library;
- **ZeroCounts:** The number of sgRNA with 0 read counts;
- **GiniIndex:** The Gini Index of the read count distribution. Gini index can be used to measure the evenness of the read counts, and a smaller value means a more even distribution of the read counts.

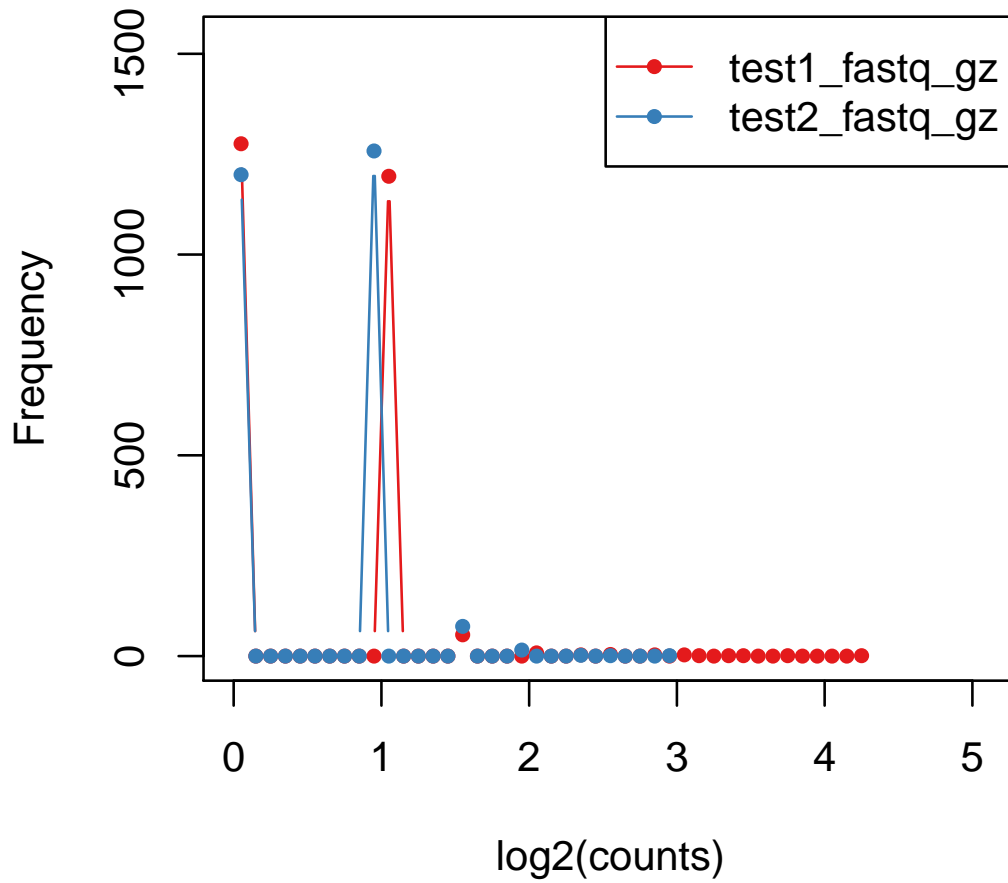
## 2 Normalized read count distribution of all samples

The following figure shows the distribution of median-normalized read counts in all samples.



The following figure shows the histogram of median-normalized read counts in all samples.

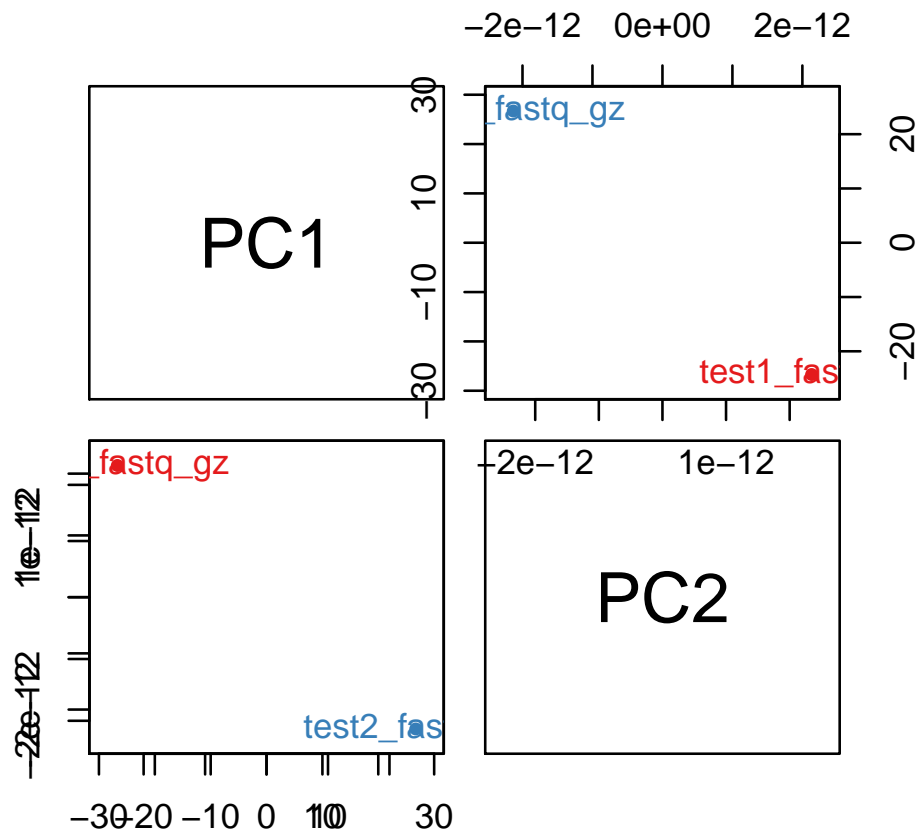
## Distribution of read counts

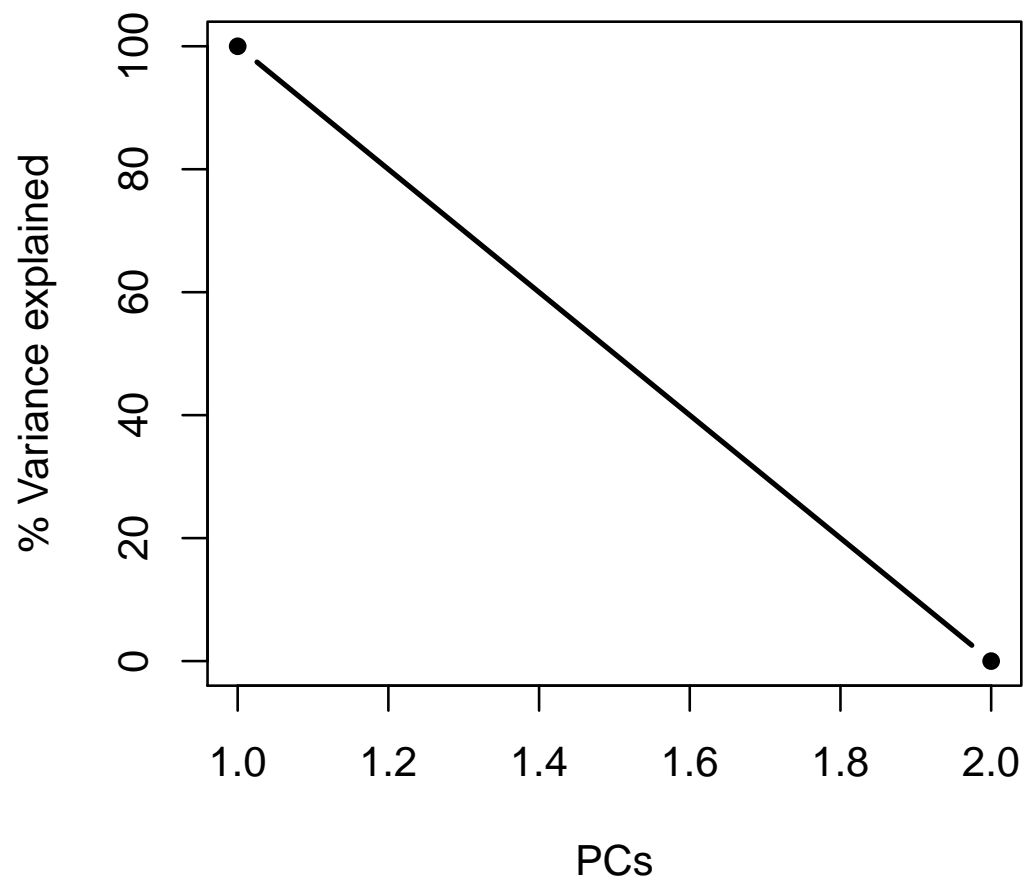


### 3 Principle Component Analysis

The following figure shows the first 2 principle components (PCs) from the Principle Component Analysis (PCA), and the percentage of variances explained by the top PCs.

#### First 2 principle components





## 4 Sample clustering

The following figure shows the sample clustering result.

