

# MAGeCK Count Report

Wei Li

August 16, 2024

## Contents

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

## 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	Low_1	43718952	43718952	1.00
2	Low_1	45560656	45560656	1.00
3	Low_2	40553489	40553489	1.00
4	Low_2	42377934	42377934	1.00
5	Low_3	51495403	51495403	1.00
6	Low_3	54023585	54023585	1.00
7	High_1	37337429	37337429	1.00
8	High_1	38986079	38986079	1.00
9	High_2	37941677	37941677	1.00
10	High_2	39892307	39892307	1.00
11	High_3	44368672	44368672	1.00
12	High_3	46403720	46403720	1.00

Table 1: Summary of comparisons

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;

	Label	TotalsgRNA	ZeroCounts	GiniIndex
1	Low_1	21339	660	0.10
2	Low_1	21339	645	0.10
3	Low_2	21339	596	0.09
4	Low_2	21339	598	0.09
5	Low_3	21339	733	0.10
6	Low_3	21339	700	0.10
7	High_1	21339	736	0.10
8	High_1	21339	732	0.10
9	High_2	21339	734	0.11
10	High_2	21339	704	0.11
11	High_3	21339	555	0.09
12	High_3	21339	526	0.09

Table 2: Summary of comparisons

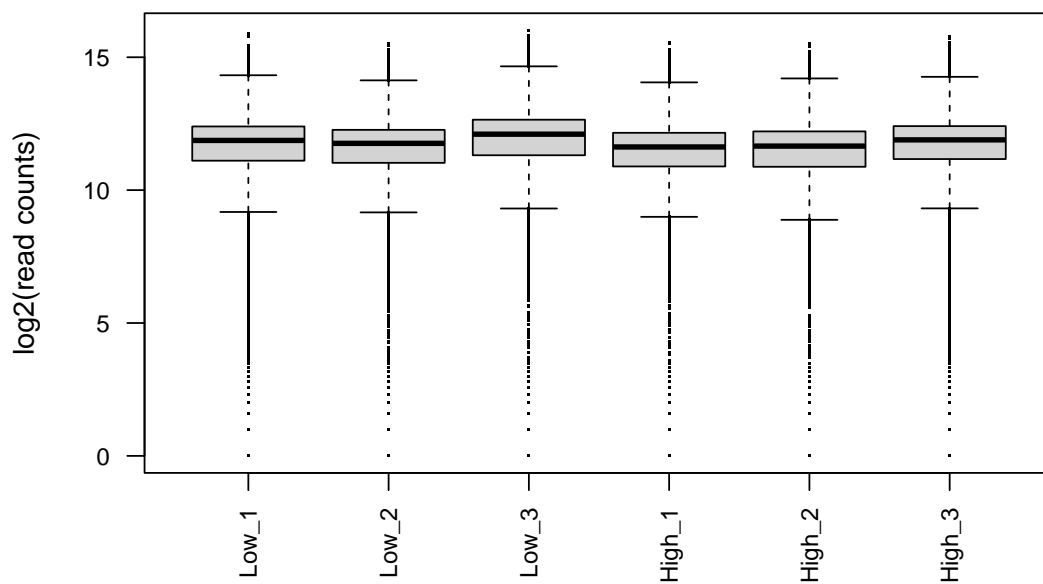
	File	Label
1	../hits.JH8105_1_S1_L001_R1_001.fastq.gz	Low_1
2	../hits.JH8105_1_S1_L002_R1_001.fastq.gz	Low_1
3	../hits.JH8105_2_S2_L001_R1_001.fastq.gz	Low_2
4	../hits.JH8105_2_S2_L002_R1_001.fastq.gz	Low_2
5	../hits.JH8105_3_S3_L001_R1_001.fastq.gz	Low_3
6	../hits.JH8105_3_S3_L002_R1_001.fastq.gz	Low_3
7	../hits.JH8105_4_S4_L001_R1_001.fastq.gz	High_1
8	../hits.JH8105_4_S4_L002_R1_001.fastq.gz	High_1
9	../hits.JH8105_5_S5_L001_R1_001.fastq.gz	High_2
10	../hits.JH8105_5_S5_L002_R1_001.fastq.gz	High_2
11	../hits.JH8105_6_S6_L001_R1_001.fastq.gz	High_3
12	../hits.JH8105_6_S6_L002_R1_001.fastq.gz	High_3

Table 3: Summary of samples

- **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.

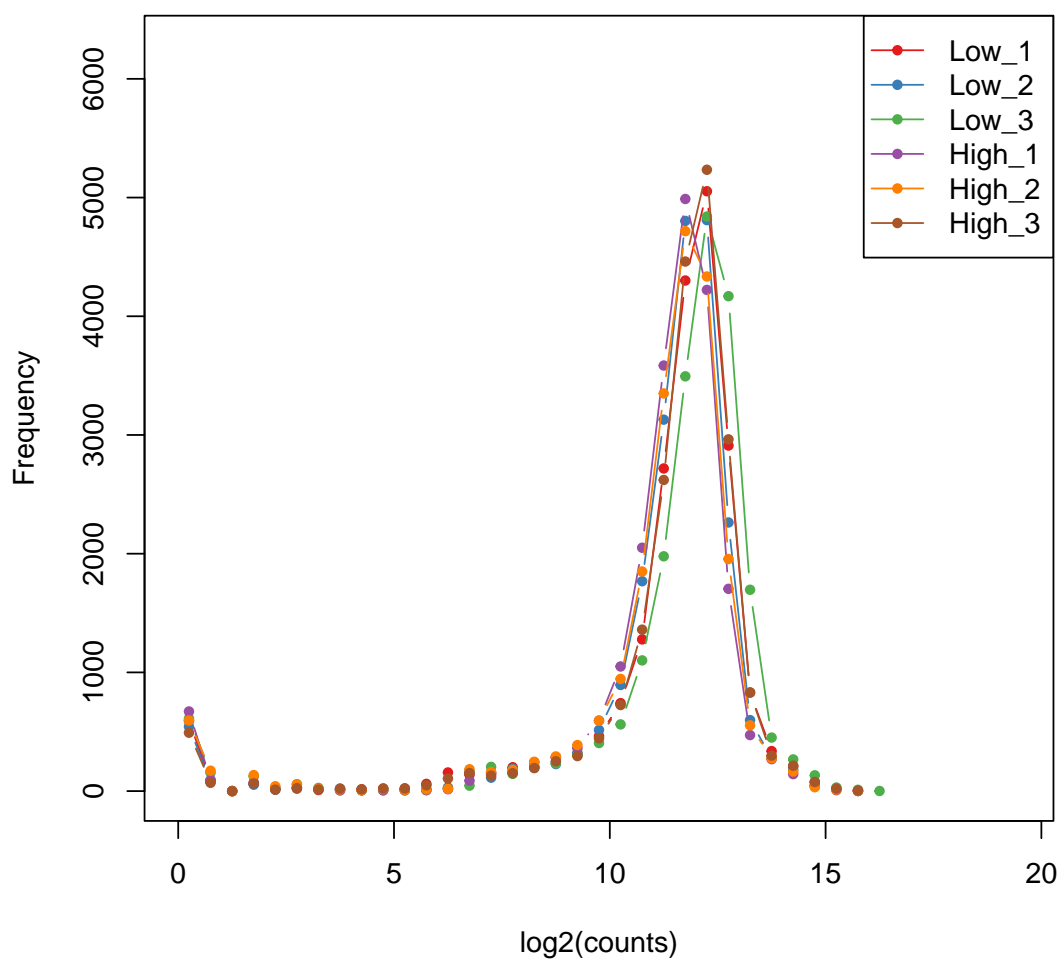
## 2 Normalized read count distribution of all samples

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



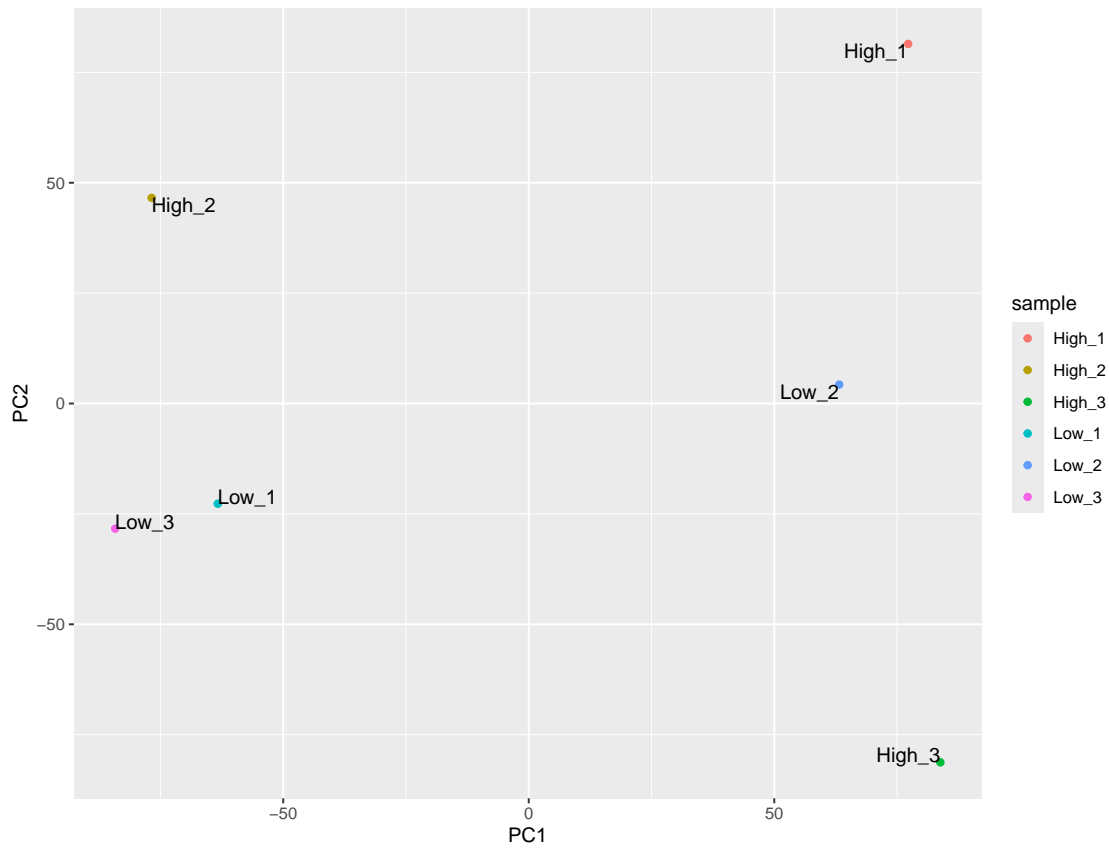
The following figure shows the histogram of median-normalized read counts in all samples.

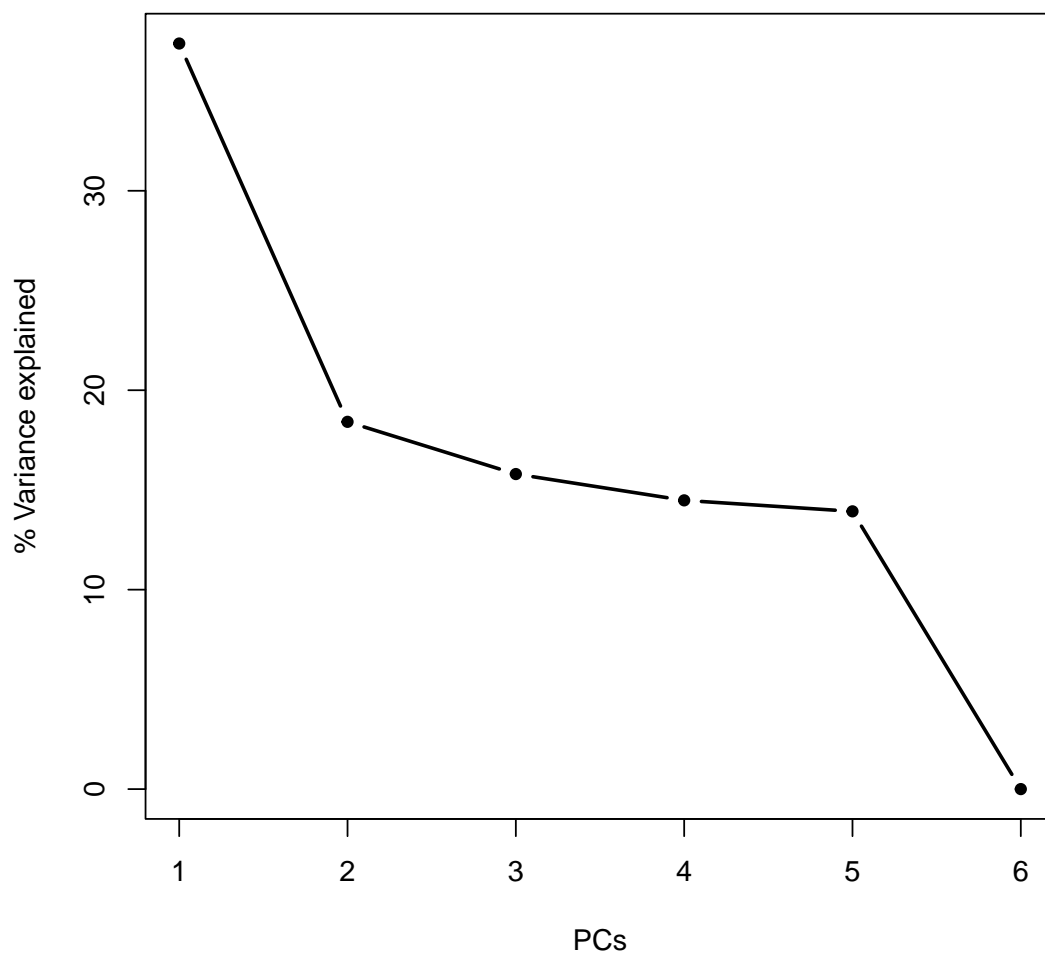
Distribution of read counts



### 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.





## 4 Sample clustering

The following figure shows the sample clustering result.

