Spike-in Project

Carol Zhou

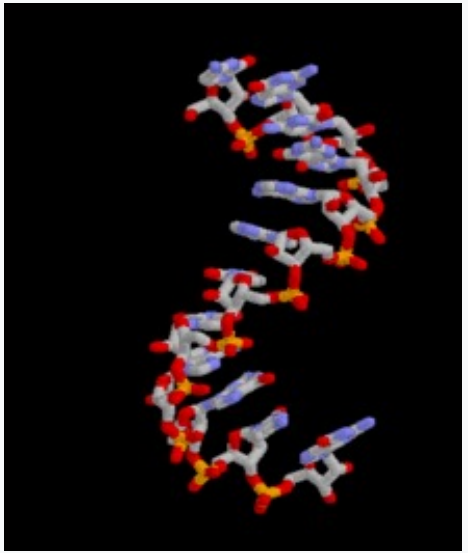


Table of contents

Introduction

RNA Spike-in

- An RNA transcript of known sequence.
  - It was first used to calibrate variations in RNA hybridization assays.
  - Now it is also adapted to be used in sequencing.
- It is a quality control.
  - Each sample is assigned and added a specific spike-in.
  - If the highest-read spike-in matches the assigned spike-in, the sample is prepared and sequenced correctly.

Questions

01

- What is the lower cutoff for the number of spike-in reads?
- Are there any spike-in aliquots with low concentrations?

02

- What is the upper cutoff for the read difference between spike-ins with the highest read and the second highest read?
- Is there any cross-contamination of spike-ins in a batch?

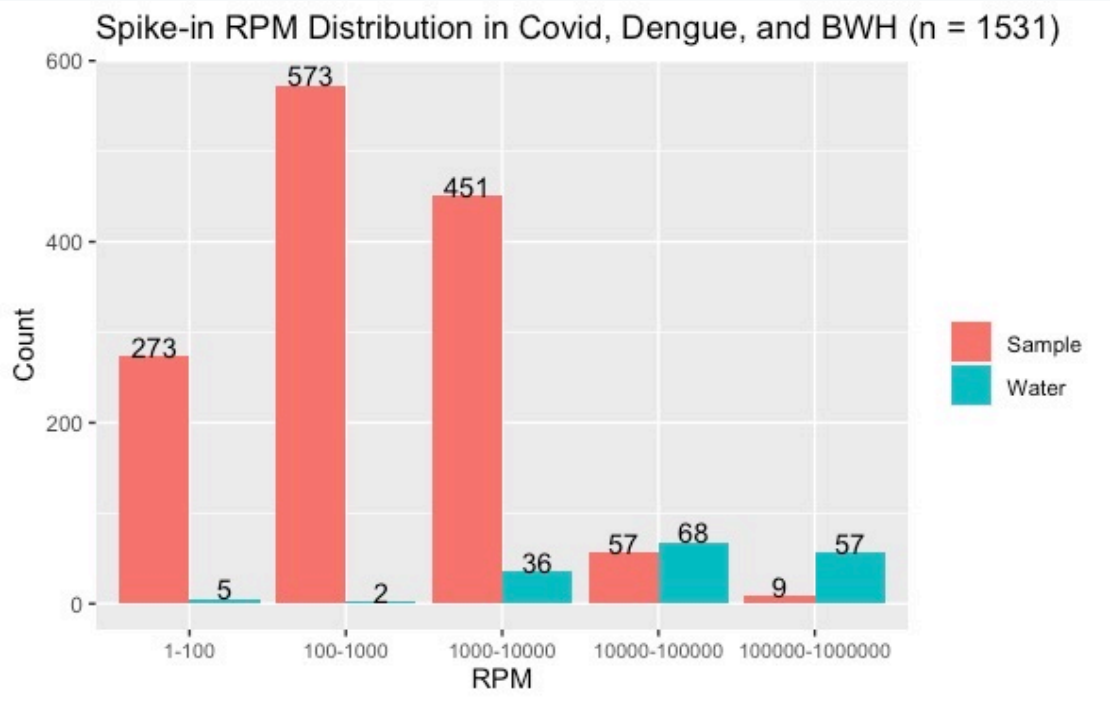
Methods

- 1606 COVID-19 (n = 1295), Dengue (n = 173), and BWH (n = 138) samples
- **R**ead **p**er **m**illion (**RPM**) = highest spike-in reads / total reads * 1000000
- **S**econd-highest to **h**ighest spike-in-read **r**atio (**SHR**) = second-highest spike-in reads / highest spike-in reads
    - Range: 0-1
- The spike-in confirmation status was re-calculated
- All analyses were done in R.

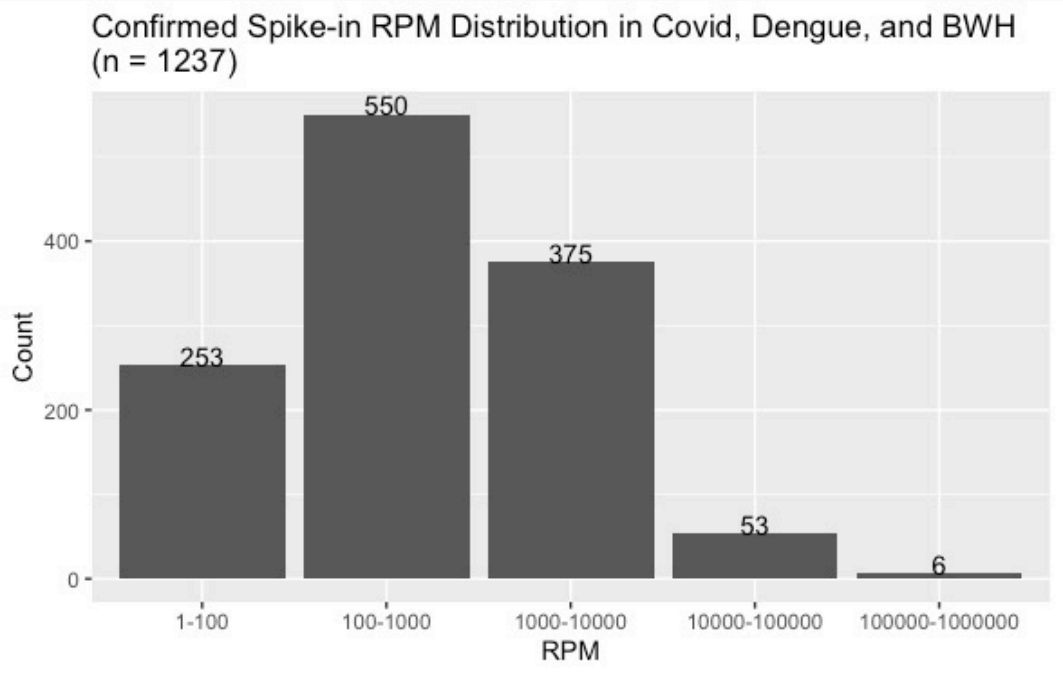
Results

Finding a lower cutoff for the RPM values

Samples (n = 1363) vs Water (n = 168) RPM



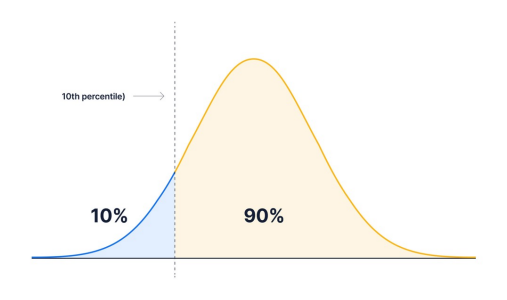
RPM Distribution



RPM Range	Percentage
1-100	20.45%
100-1000	44.46%
1000-10000	30.32%
10000-100000	4.28%
100000-1000000	0.49%

74.78%

RPM Distribution



RPM Range	Percentage
1-100	20.54%
100-1000	44.42%
1000-10000	30.4%
10000-100000	4.24%
100000-1000000	0.41%

Percentile	RPM
20	94.28
15	61.85
10	32.56
5	12.41
3	7.24

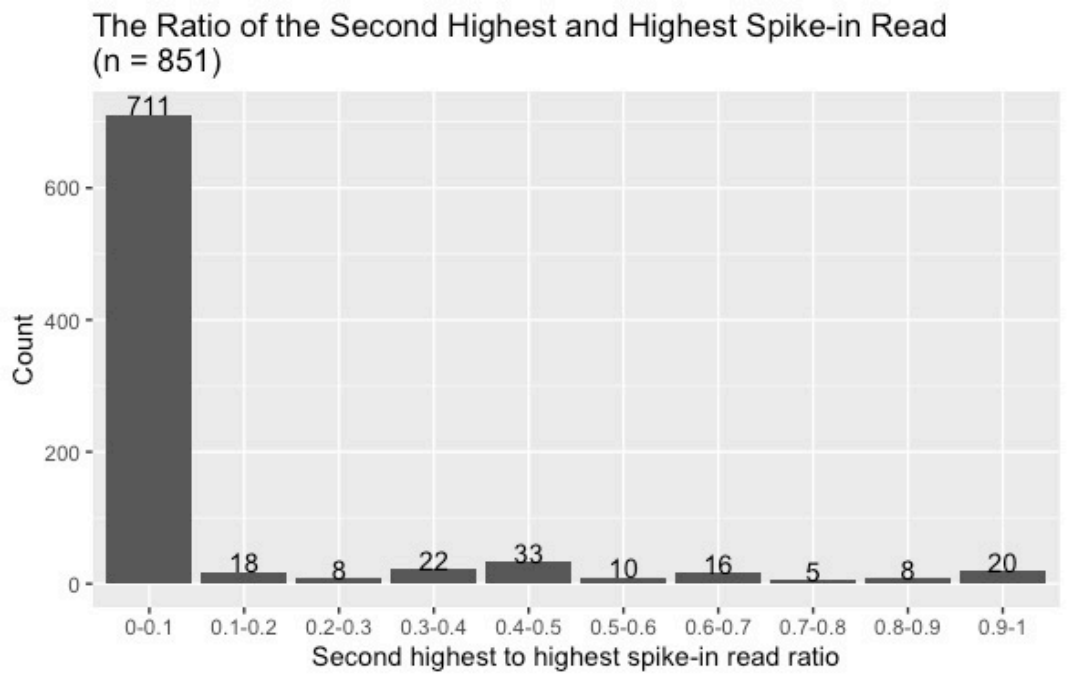
Spike-ins with 1-100 RPM

Assigned Spike-ins	Low RPM Count	Total Count	Percentage
ERCC-00023	24	27	88.89%
ERCC-00077	16	20	80.00%
ERCC-00054	18	23	78.26%
ERCC-00067	12	24	50.00%

Results

Finding an upper cutoff for the SHR values

SHR Distribution



SHR Range	Count	Percentage
0-0.1	711	83.55%
≥ 0.1	140	16.45%

Percentile	SHR
80	0.05
85	0.16
90	0.43
95	0.63
97	0. 84

Spike-ins with > 0.43 SHR

Sequenced Date	Virus	Count	Total Count	Percentage
2022-12-01	Dengue	15	18	83.33%
2023-02-13	Dengue	12	15	80.00%
2023-02-02	Dengue	13	21	61.90%
2022-12-08	Dengue	11	18	61.11%

Summary

- The lower cutoff for spike-in RPM
  - 12.41 – 100
  - 12.41 is the lowest cutoff
- The upper cutoff for spike-in SHR
  - 0.16 – 0.43
  - 0.43 is the highest cutoff

R Function

confirm_spikein3.0.1

Description

A function used for calculating whether the expected spike-ins are confirmed and other spike-ins detected and their reads.

Usage

confirm_spikein3.0.1(spikein_dnanexus, spikein_ssss, spikein_sampleid)

Arguments

spikein_dnanexus	A count_summary.tsv file downloaded from DNAnexus
spikein_ssss	A csv file that contains information for the corresponding samples from SSSSS on Teams
spikein_sampleid	A csv file that contains only sample IDs for the corresponding samples from SSSSS on Teams

Instruction

General → Current lab members folders → Carol → Spike-in Project → R

Citation

https://en.wikipedia.org/wiki/RNA_spike-in#
https://www.thermofisher.com/order/catalog/product/4456740