Sub_0645: miRNA Normalization

John Steill

2018-03-05

Introduction

There are two goals for this analysis:

1. To evaluate the effect of rna concentration and pcr on expression estimation
2. Create normalization method:
   - C. Elegans spike-in
   - CPM vs tam_UQ

Unfortunately, the spike-in method failed. No reads were produced.

In evaluating the noise induced by lower concentrations and pcr amplification, we are mostly concerned with the abundances of the most prevalent miRNAs. We are interested in when the values are distorted to the extent that their ranks change. For this question, choice of normalization method won't matter: whether we divide by sum()/1e6 (cpm) or by a given quantile value, the ranks will be unaffected.

However, this data should provide insights into the best normalization method. As the base sample is the same for each trial, all differentially expressed miRNAs will be false positives. Thus we may want to minimize the FDR by choosing the method with the fewest DE miRNAs.

Raw Counts

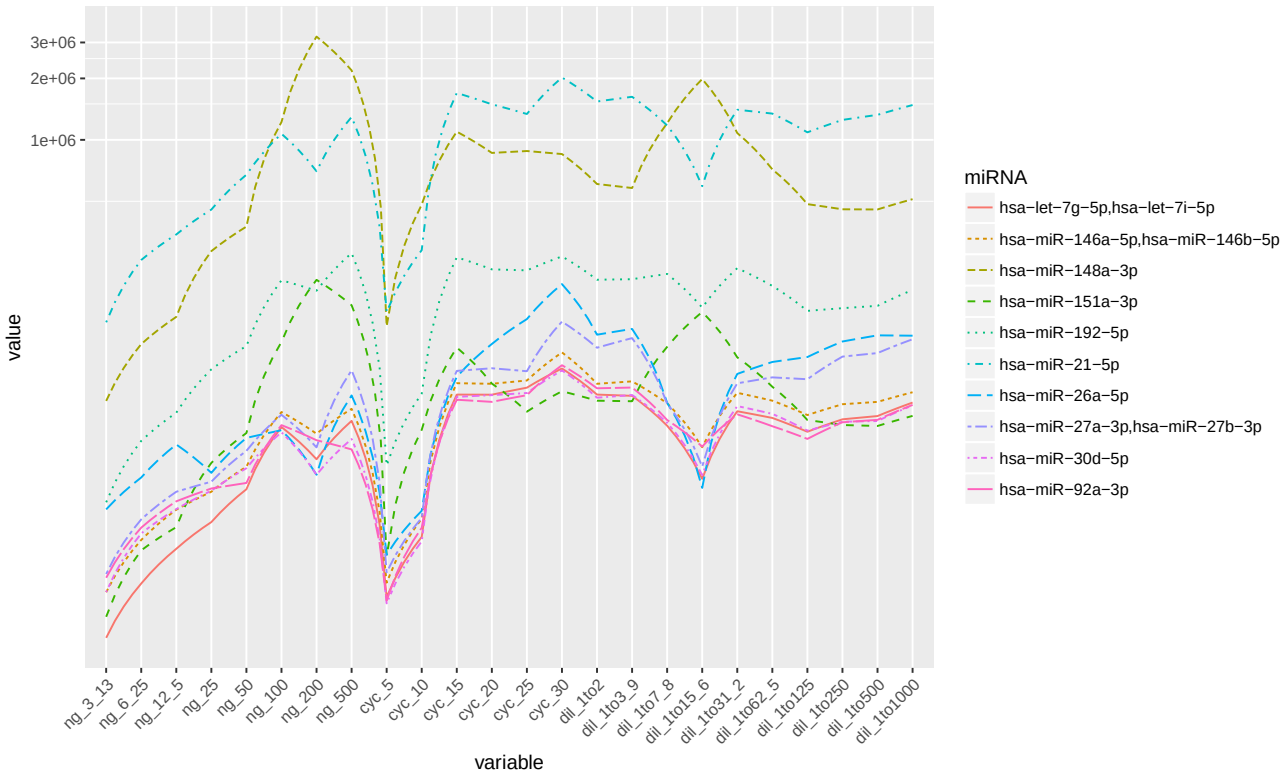

Figure 1: Raw Reads

We see some trends:

- Fewer total counts as the initial concentration drops
- Fewer total counts with lower cycle counts (saturating at ~15)
- Robust total counts in the dilution sequence, although the 1:15.6 appears to be an outlier.

CPM Normalized Counts

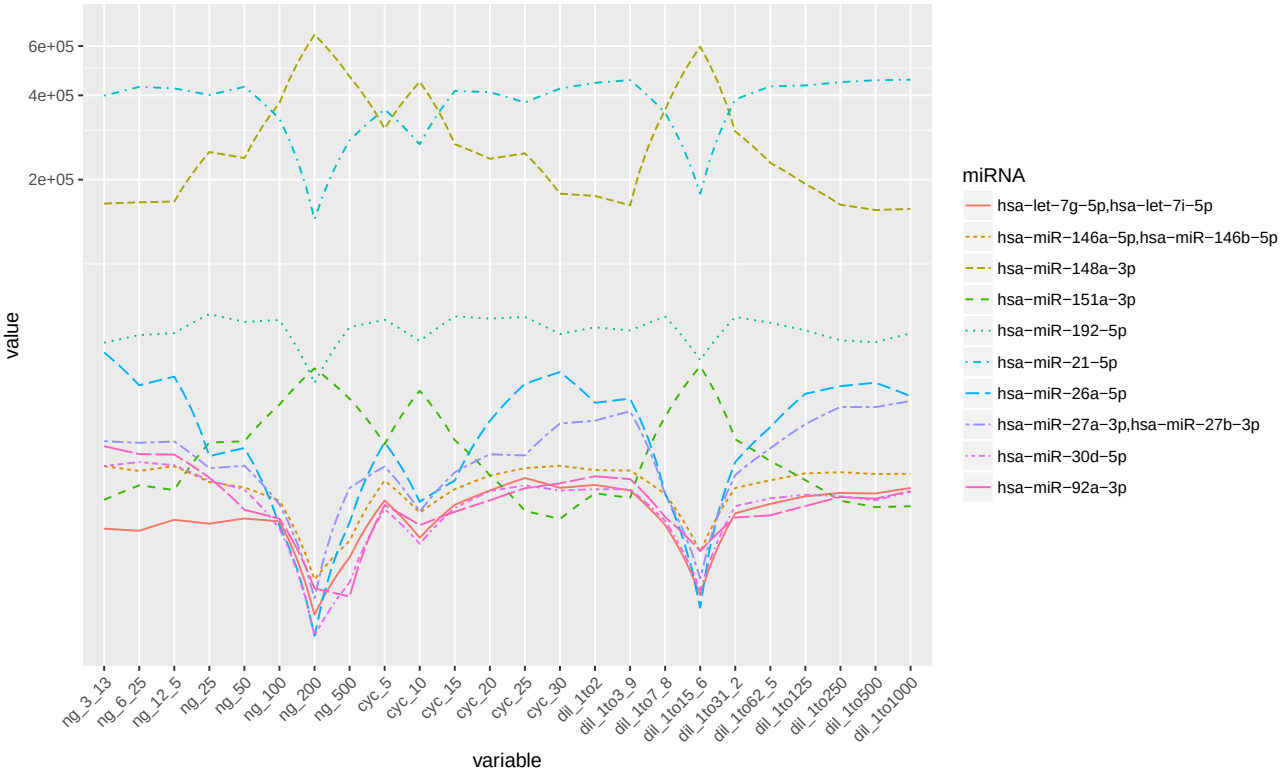

Figure 2: CPM

What I find interesting here is that the relative rank of two species, 148a-3p & 151a-3p vary quite a bit in the range of supposed stability: 100-500ng and dilutions of 7.8 through 31.2. Are there two outliers?

Tam_UQ Normalized Counts

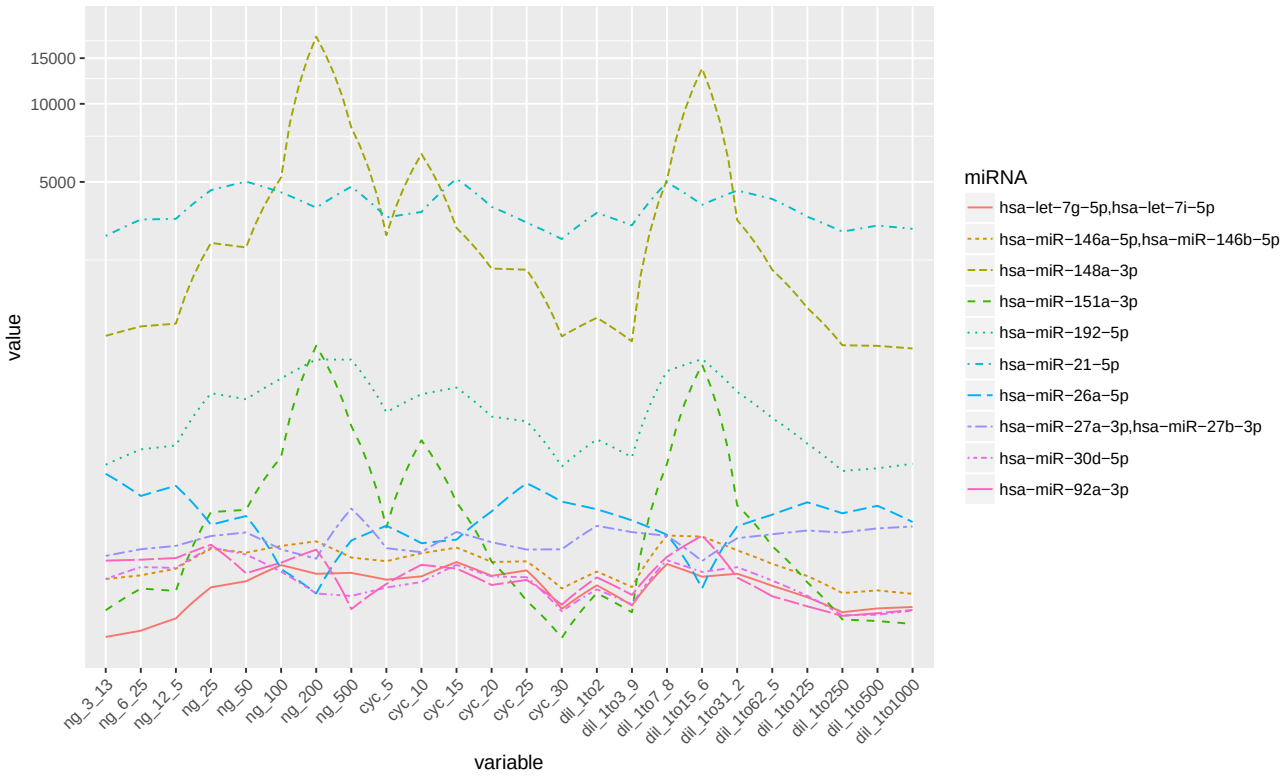

Figure 3: Tam UQ

The function of upper quantile normalization is shown by the smoothing of the highest counts.

Required Amount Analysis: Spearman Correlation

It satisfies common sense that the very highest expressed miRNAs will be relatively unaffected by initial concentration.By doing a spearman correlation, we can examine the sensitivity of species which are express at lower levels.

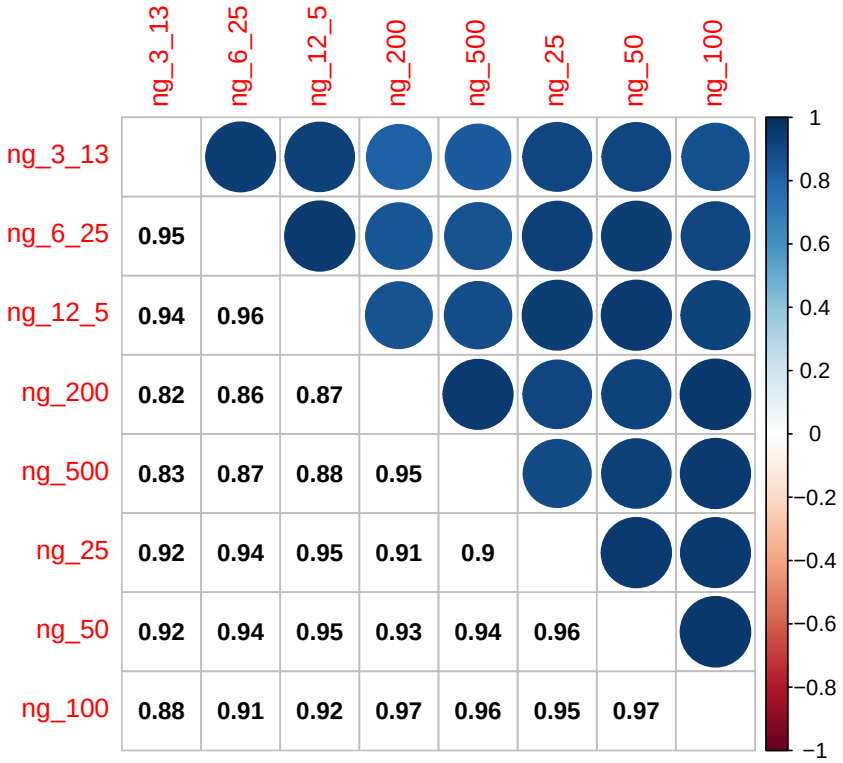

Figure 4: Tam UQ Spearman Correlation; 400 top miRNA

TO-DO

1. Normalization Analysis
   - Choose a region of interest (avoid outlying conditions, omit outliers?)
   - Choose a DE algorithm
   - Investigate which normalization gives the minimal DE set.
2. Revisit Spike-in's which failed.
3. ?

References