Li QC Analysis

Vai Pathak

May 31, 2017

Contents

Li QC Analysis

This is a quick study to mainly see if the starting number of FFPE cores in a patient sample has any effect on the downstream QC metrics of the Affymetrix Clariom S microarray. So far 29 patient samples have been run on microarrays, each with each sample containing a different number of cores (mostly ranging from 1-3 cores).

1 Table Summary of Cores and Samples

Here is a summary table of the number of samples having 1 2 or 3 cores (with the total sum being 29)

.

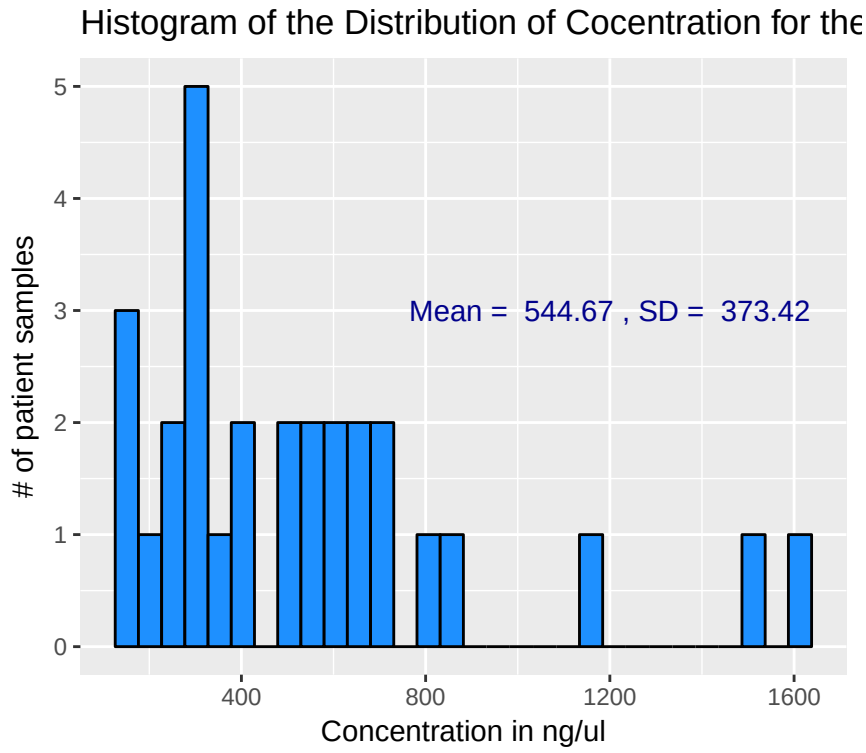
1	2	3	Sum
7	13	9	29

As seen in the table, there's a disproportionate number of samples and core numbers. There are 7 samples with 1 core, 13 samples with 2 cores, and 9 samples with 3 cores (for a total of 29 samples).

2 The Distribution of the Concentration amongst all 29 samples.

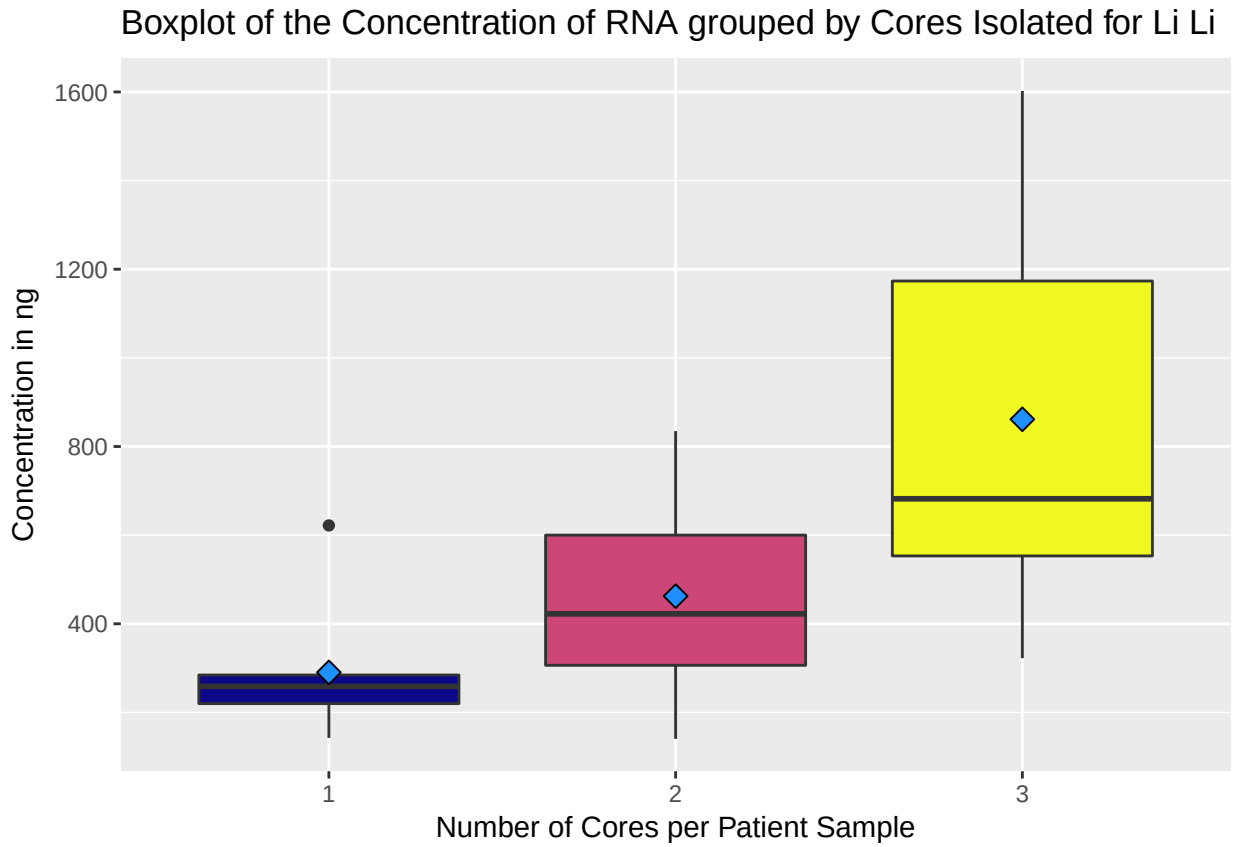
2.1 Histogram of the Concentrations

Here we can take a look at how the concentration of the samples look across the board.

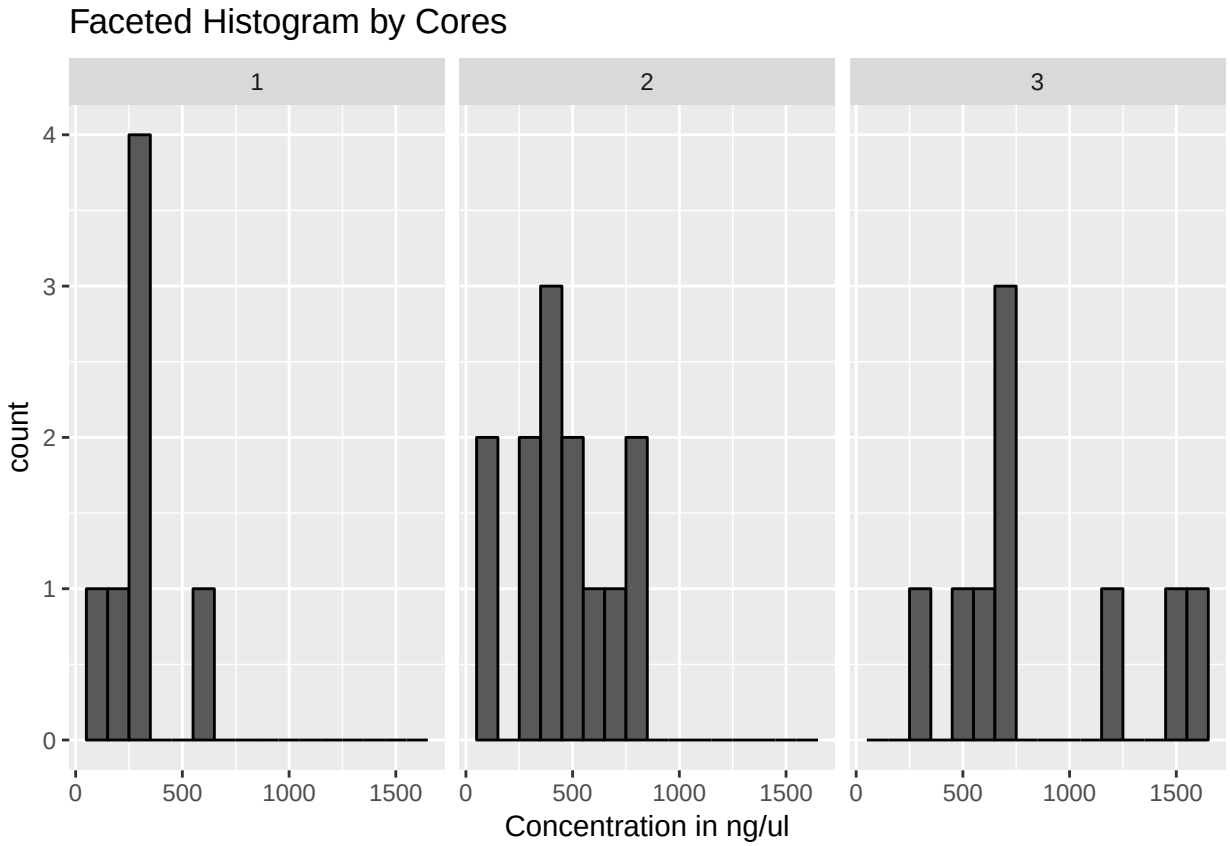


X shows the range of the concentration (in ng/ul) of each sample and Y gives the count (29 total samples). We can also see the average concentration is 544.67ng/ul - which is representative of all 29 samples regardless of their core numbers.

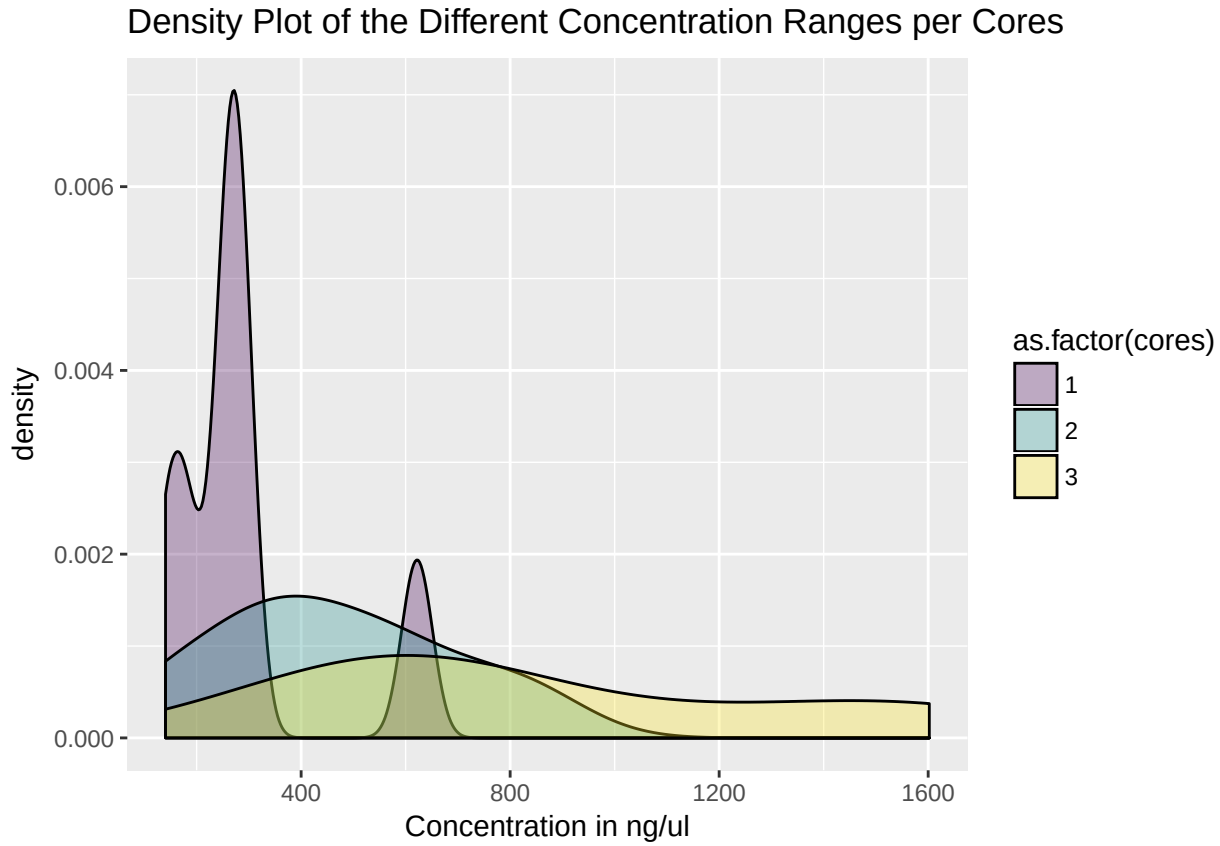
2.2 Boxplots, Density plots, and Histograms of the Cores and Concentrations



The boxplot shows a grouping of concentrations by the number of cores. As expected - generally, the more cores we have, the higher concentration of RNA. Boxplots represent a five-number summary of the minimum, 25th percentile, median (black bars), 75th percentile, and the maximum. The blue diamonds represent the mean, which when different from the median, is representative of a skew in the distrubution.



The breakdown of the boxplot's range can be seen better in the histogram of concentration distribution, separated by the number of cores in the patient sample.

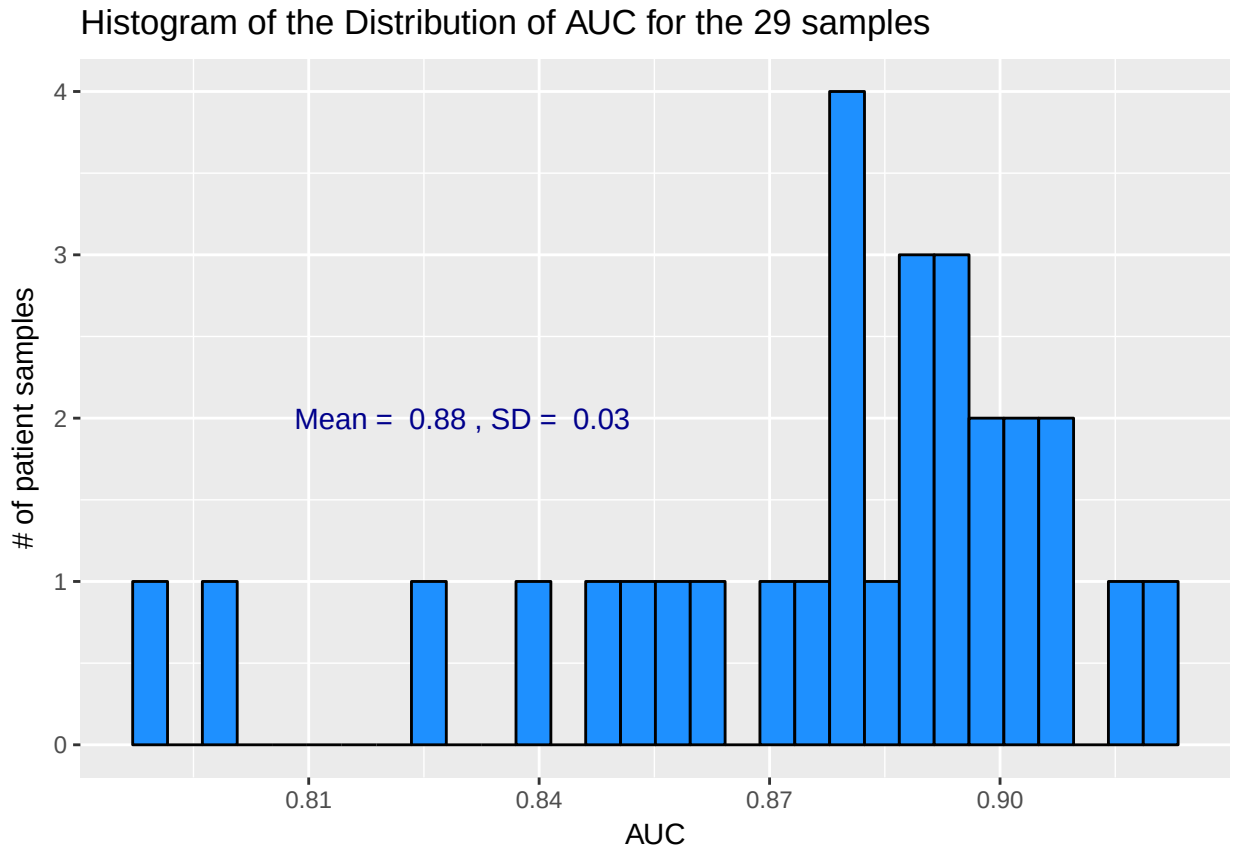


The density plot shows a more "fluid" visualization of the histograms for the distribution of the concentration of all three cores.

3 Taking a Look at the QC Metrics of the 29 Arrays

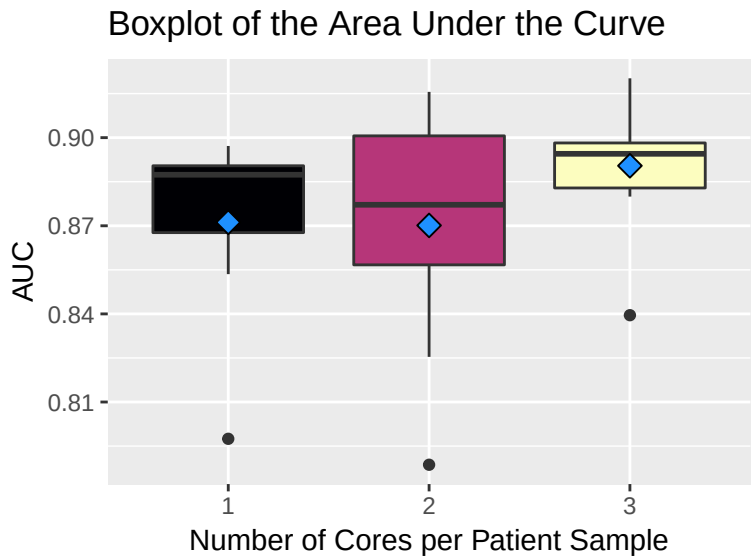
Finally we can take a look at the QC metrics of the microarray and see if there's any correlation between the number of cores and if it affects the QC. There are a variety of QC metrics measured by the microarray - the most critical being the "Area Under the Curve" (AUC).

The same graphs that were used in the concentration study, can also convey the distribution of the AUC curves for each core.

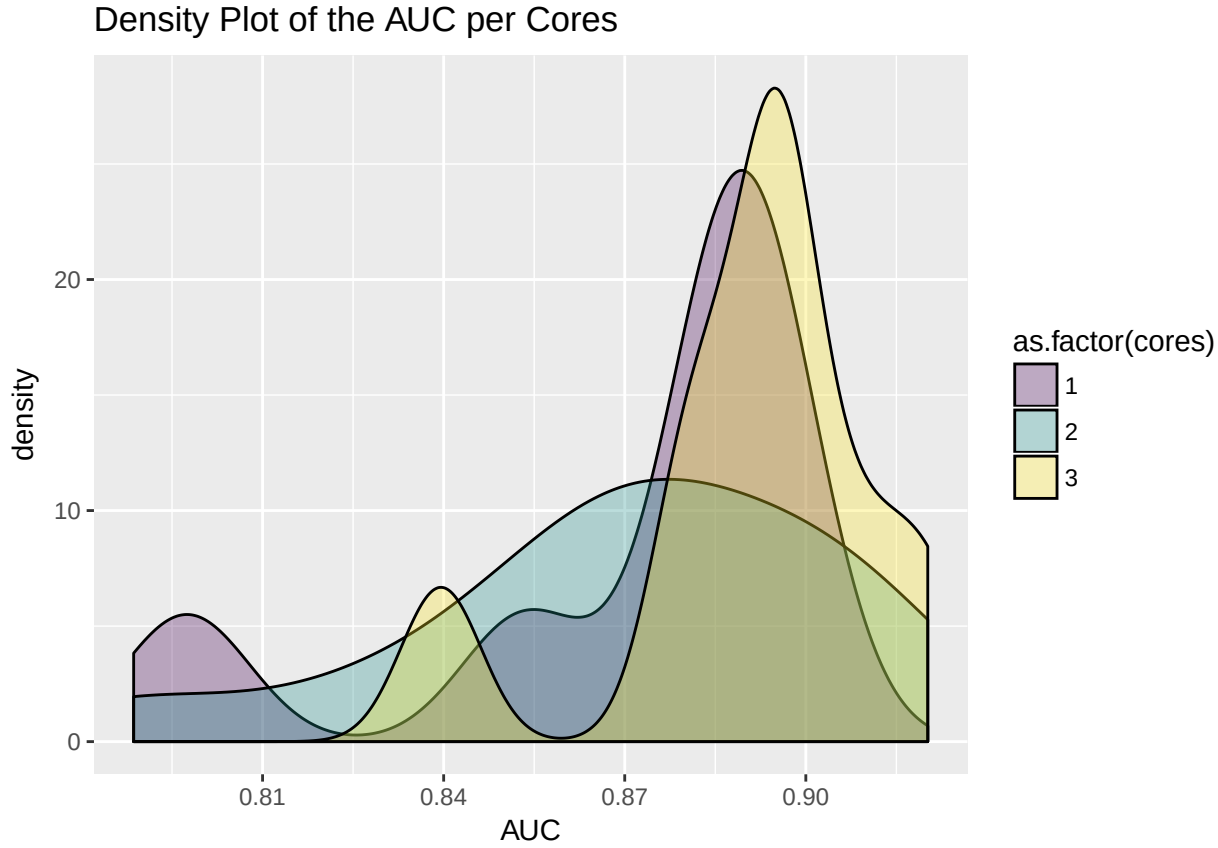


The mean AUC shows .88, which impies a good quality signal despite samples coming from formulin-fixed paraffin embedded tissue.

We cand also see how the AUC looks by cores:

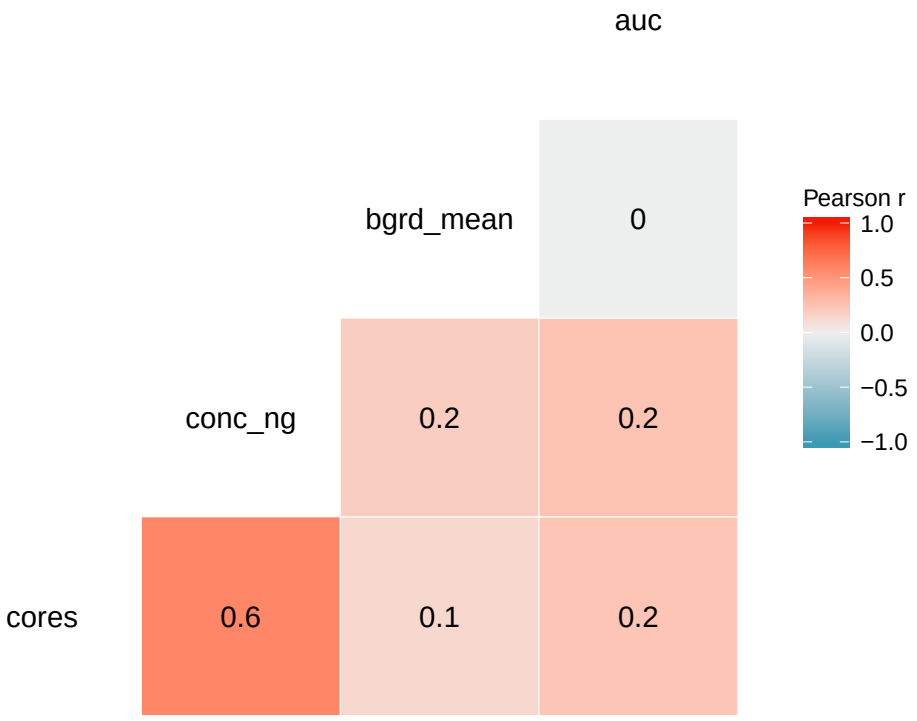


Looking at the boxplots, there doesn't seem to be a sizeable significant correlation between the the amount of cores and the AUC. We can note that the two samples that had an AUC under 0.81 both happened to occur from samples that had 1-2 cores, but more samples would be needed to affirm a significant correlation. The density plot agrees with the boxplot:



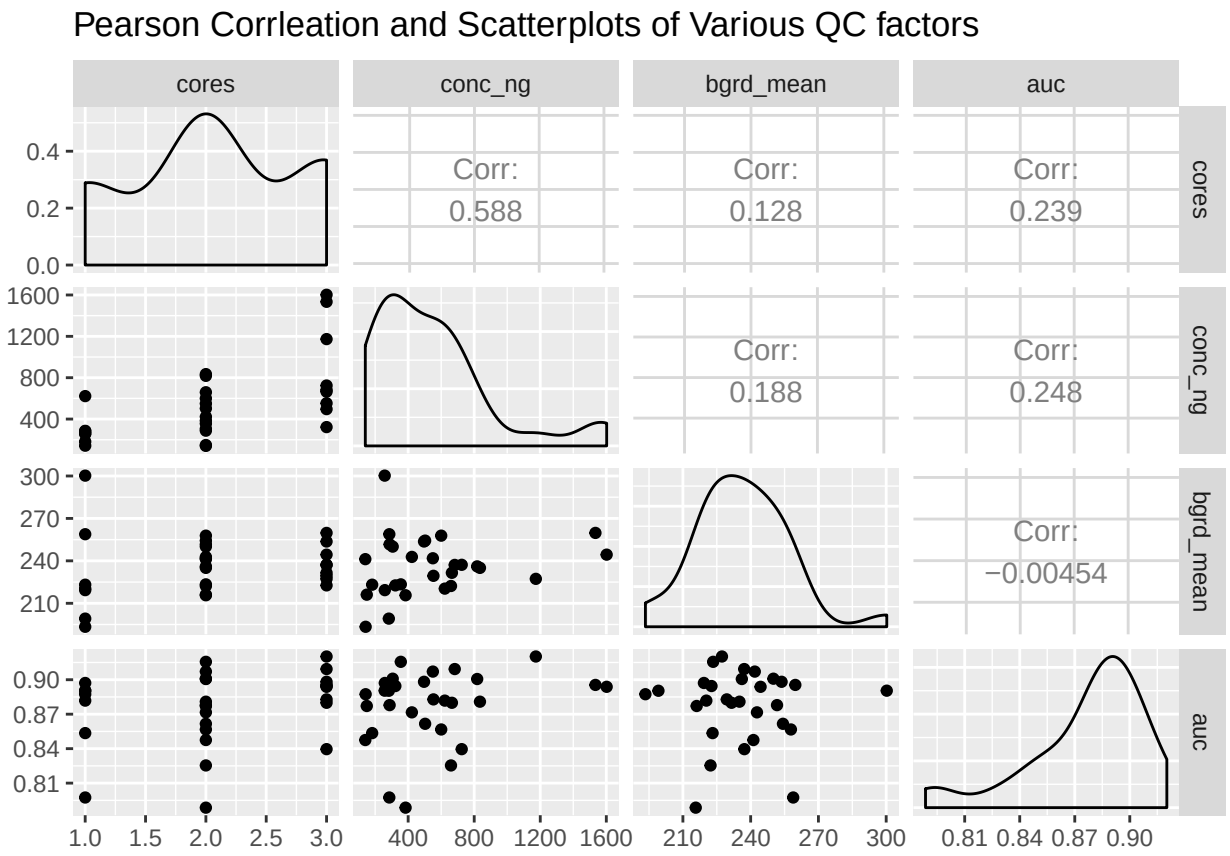
3.1 QC Metric Pearson Correlation

Finally we can check the Pearson correlation with a few of the major QC parameters to see if there is any correlation between the number of cores added and the QC metric. Here, we will take a look at AUC, concentration, and background:



This heatmap mainly shows that there's a slight positive correlation between the amount of cores and the concentration with a Pearson correlation of 0.6. However there doesn't seem to be a significant enough correlation between the number of cores and how it affects the QC metrics (Pearson correlation of 0.2).

Another iteration of the heatap can be seen here with scatterplots and the Pearson correlation combined in one figure:



All in all, this is a very preliminary look at how the number of cores affect the downstream effects since there are only a few samples. There is an arguably positive correlation with the number of cores and the concentration of RNA, although there is also the issue of core size where 1 core could be larger than a sample that has 2 cores due to its overall surface area - or the question of what constitutes 1 or two cores (see Figure 1). Other batch effects during sample processing could also cause an issue and may remain to be seen when further samples have been processed. Even with the 2 samples with 1-2 cores that had a low AUC value, it remains to be seen if there is a significant correlation with the number of cores and the downstream affects on the array.

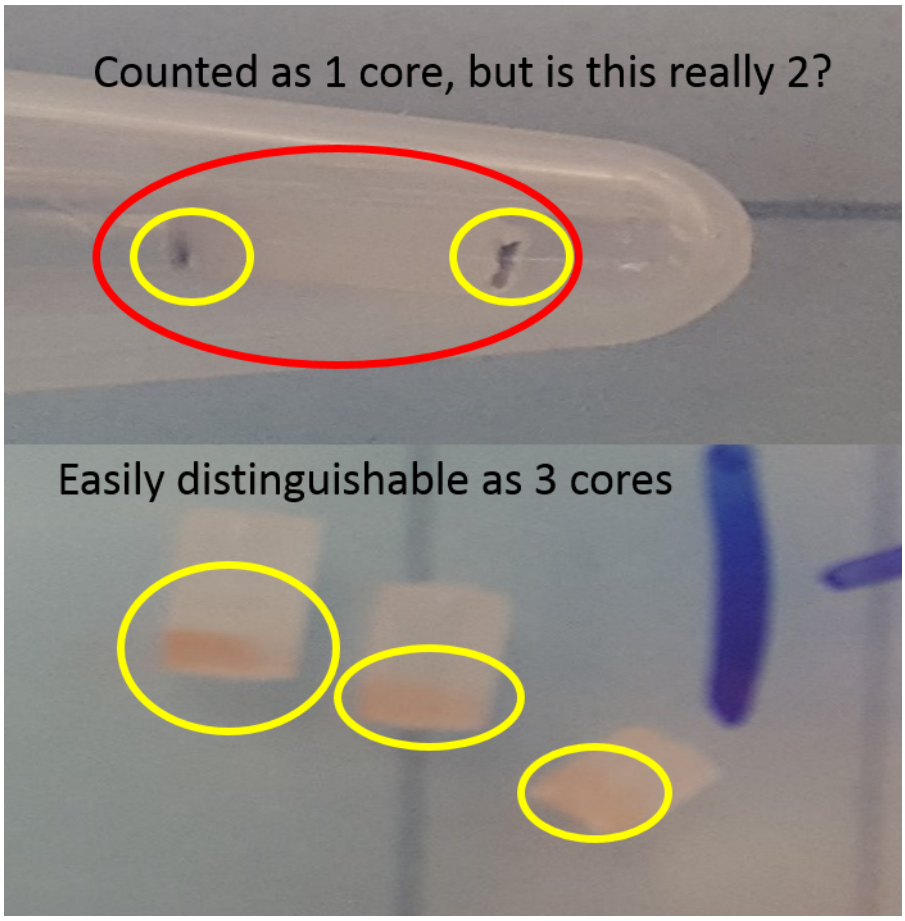


Figure 1: As shown, cores have different surface areas and sizes. Since core size is not generally normalized, it's difficult to come to a definitive conclusion simply based on the amount of cores per sample - though some slight correlation was found, particularly in the concentration.