High expression SOP

**Standard as of:** December 21, 2018

**Note:** This is to be done after preprocessing. Preprocessing involves filtering for the 95th percentile of expression. As high expression includes all miRNAs, start with the raw data, but keep note of which samples were removed during preprocessing.

**miRNA**

1. Start with raw merged miRNA summary data
2. Remove samples that were marked as outliers in preprocessing
3. Relative frequency (RF) normalize
4. Find the median RF expression of each miRNA across all remaining samples
5. Remove star sequences (has ‘STAR’ suffix)
6. Get highest N expressed miRNAs where:

* N = [remaining miRNAs] \* [percentile]
* (round down)

**Cistrons**

1. Start with raw cistron data (clust\_summary)
2. Remove samples that were marked as outliers in preprocessing
3. Relative frequency (RF) normalize
4. Find the median RF expression of each cistron across all remaining samples
5. Get highest N expressed cistrons where:

* N = [remaining miRNAs] \* [percentile]
* (round down)