\*\*\*\* To keep this file current, add a Date when something is completed and we can keep items still ongoing at the top of the list \*\*\*\*\*\*

- Consider Quantile normalization or a trimmed mean normalization to help decrease replicate bias do to normalization (suggestion from Tim Sackman FAS RC)

Also consider looking at the raw counts of top transcripts in Run3 are their any that are super high and would bias the dataset

- Use EdgeR to show genes that are upregulated >2 fold between certain samples. Ideally I would like to choose Myoblasts, Satellite cells, DMSO Mefs, AdMyoD Mefs then presumable the hits (a couple of them). For right now the consistent ones would be D24 and H24. Would we be using the Grep command for this? Then moving on we could create Heat maps with hierarchical clustering - It would be good to see as we would both predict the TCL samples should cluster more similarly with each other then the SCMyo samples

11.12.15 Trying to get rid of batch effects – create identities for our samples

- TCL3 when analyzed on it’s own versus the SCMyo samples does actually look OK in terms of showing the positive controls. Strange that it overlaps with the SC samples when we add the other TCL samples (1,2,5)

- Last time we chatted (Nov 5th) we were in the process of summing all the reads per gene across the samples then generating an MDS plot from this. We called this section “Summarize Data Mag’s Proposal”. We never finished this.

11.12.15 Awaiting response from Mag on this topic

- Where do we stand with our Targeted gene list being applied to our data for an MDS plot. This seemed to be quite beneficial when used for the AdMyoD fibroblast replicates before.

\*\*\*\*\* Completed \*\*\*\*\*

~~3) Lets add the real compound names as our column names. The file Real\_Col\_Names.txt is attached to this email.~~

~~11.11.15 Complete added code to Reptile~~