Final Deliverables – Tanish Kumar

1. The paper began by collecting data from GSE32323, GSE 21510, and GSE 8671. The differentially expressed genes were then discovered using DeSeq2 and hypothesis testing. The paper also used DAVID for KEGG enrichment. They used STRING to generate the PPI networks. Finally, they used GEPIA and StarBase to get prognostic results for candidate genes.
2. The only thing I see missing is the quality control step of the analysis. They seemed to have jumped right into finding the DEGs. This suggests they already have assessed the quality of the data and verified that they are of good quality.
3. Heat Map and Volcano Plot
	1. 
	2. 
	3. 
4. Plots
	1. 
	2. 
5. Challenges and Things Learned
	1. Initially, I was not sure whether to include the LCM samples through my pipeline.
	2. I had to debate whether or not to remove the samples with bad quality.
	3. I had trouble generating the correlation matrix for the heat map but was able to troubleshoot.
	4. I tried using Enrichr instead of Metascape in the beginning however it did not give me the results I was looking for.
	5. Things Learned:
	6. How to use Metascape
	7. How to use affyQCReport
	8. Generating a heatmap using a correlation matrix and pheatmap
	9. Criteria for removing samples
	10. How to create a pipeline
6. I think it might be interesting to include using GEPIA and/or StarBase in future pipelines
7. Achievement Highlights
	1. I was able to run a dataset successfully through my chosen pipeline
	2. I was able to complete all of the final deliverables on time.
	3. I was able to troubleshoot through mentors and other interns.
8. Although the field of Bioinformatics is well-developed, there is still a large of amount of area left to explore!