**Compare** “Results from JCI Insight” vs “Results derived from your own analysis”

**Identify** a data-set related to Monocytes in bulk RNA-seq and run the complete pipeline. Generate a report.

**Prepare an organized pipeline for RNA-seq.**

- Power point explaining the pipeline.   
- Example with R.  
   
**Extra mile:**

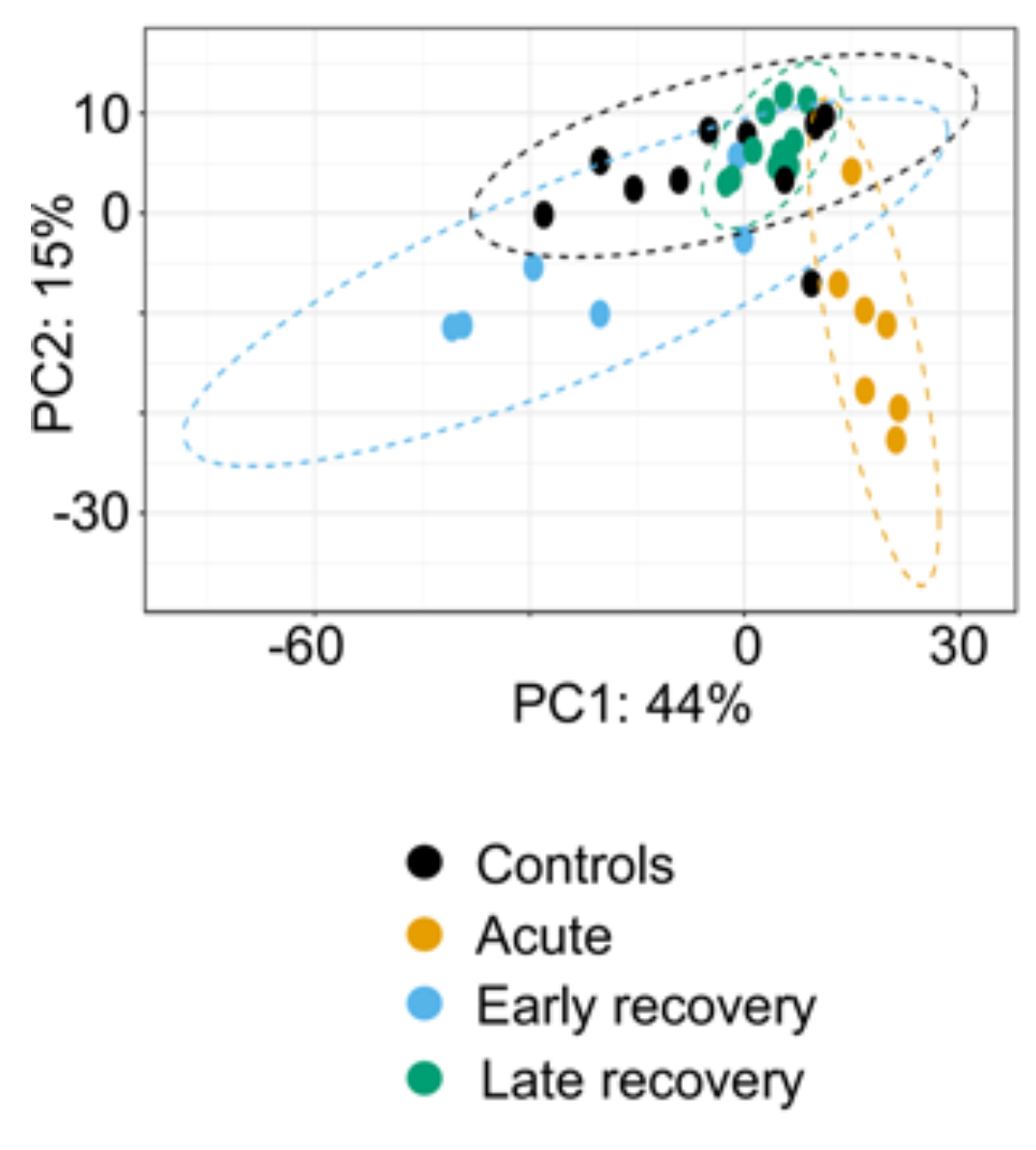
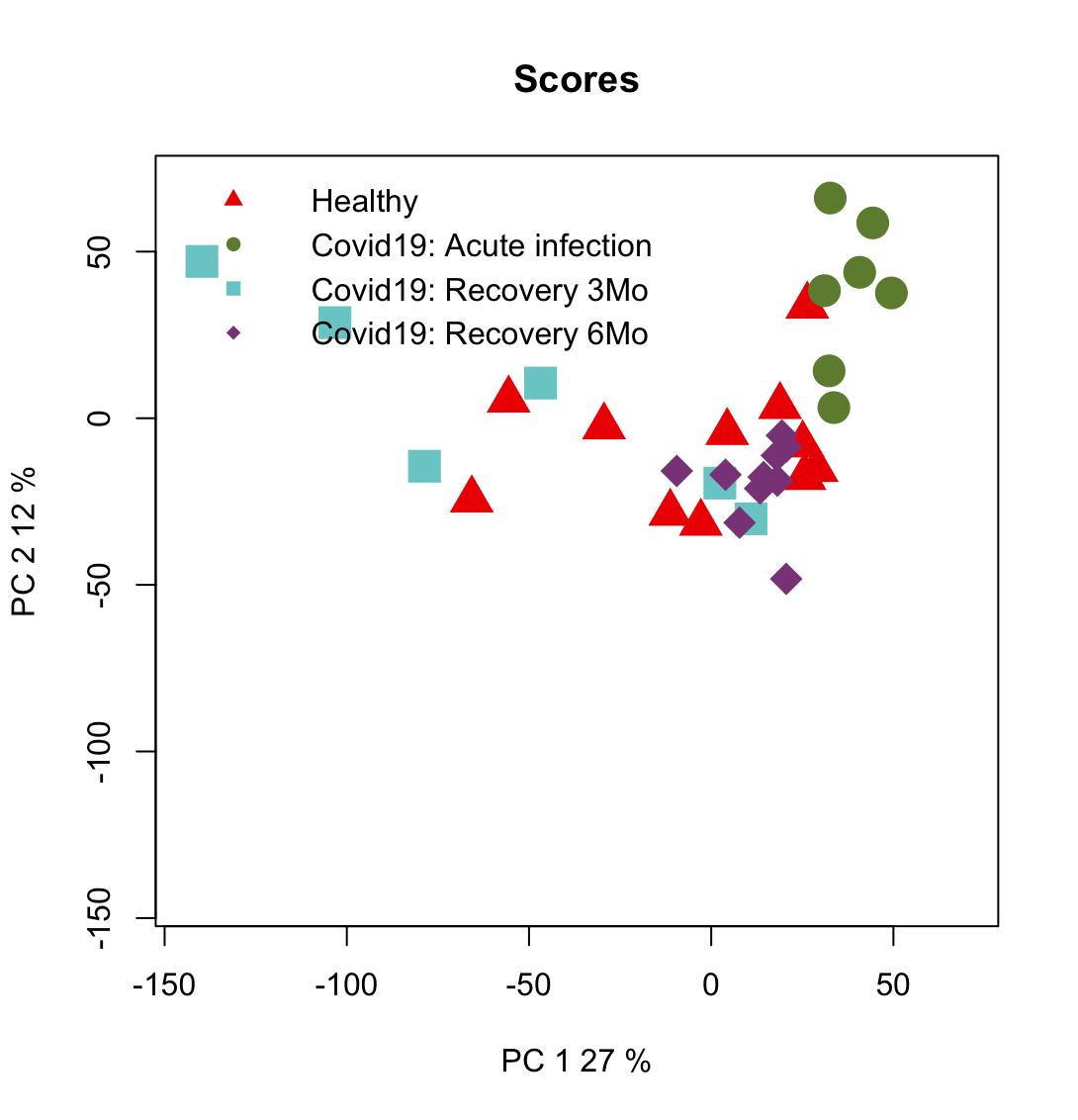
Process the fastq to count table from of the three data-sets. Re-do the analysis and generate a report:

- Code

- Outcomes of the analysis.

**Part 1: Compare** “Results from JCI Insight” vs “Results derived from your own analysis” - RNAseq results

1. **PCA Plots**



Our results show a comparable distribution of samples concerning their category. More notably different is the proportion of variance for the respective principal components. PC1 has 27% of the variance in our results to 44% in theirs, and 12% to their 15%. While unclear from the original study, our implementation of PCA is before any filtering or normalization. In contrast, the clearer signal in the original study’s figure might suggest their implementation of PCA was after filtering and normalization.

1. Differential Expression Analysis

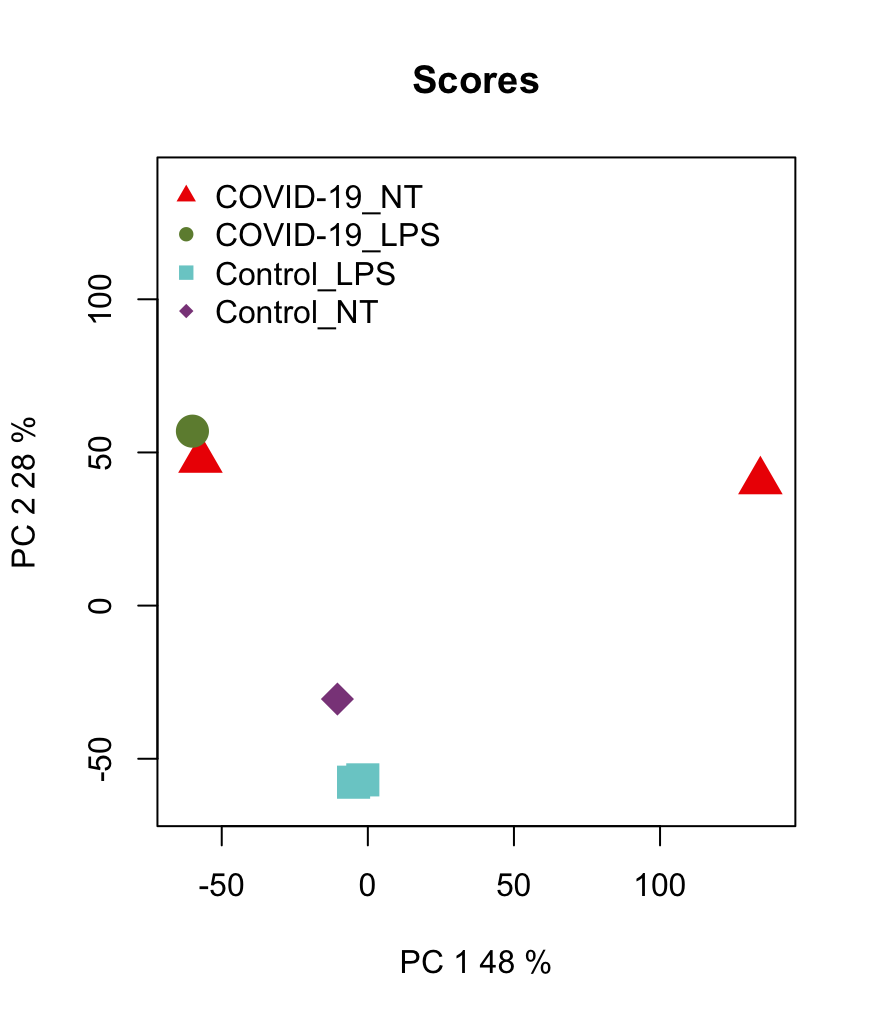
|  | **Original study** | **Our analysis** |
| --- | --- | --- |
| **Acute to control** | (339 genes) 184 up and 155 down | (669 genes) 257 up and 412 down |
| **Early to control** | (521 genes) 318 up and 203 down | (358 genes) 214 up and 144 down |
| **Late to control** | “Very few” | (12 genes) 2 up and 10 down |

Our approach seems to identify more DEGs for early relative to healthy than acute relative to healthy. Although, early relative to healthy has more upregulated compared to downregulated genes. We use shrinkage with method=”ashr”, and the same filtering as the original study (absLFC>1, adj.pval<0.05).

Part 2: **Identify** a data-set related to Monocytes in bulk RNA-seq and run the complete pipeline. Generate a report.

We process GSE247186, which is for myocytes in blood for healthy and COVID19 infected patients, some treated with TLR4 ligand and some not treated.

Using our pipeline from before, we get the following results.



Our results show that batch correction might be necessary since groups do not form distinct clusters.

After filtering, the results are very simple.

|  | **Total DEGs** | **Upregulated** | **Downregulated** |
| --- | --- | --- | --- |
| Infection (COVID19 vs Healthy) | 150 | 32 | 118 |
| Treatment (TLR4 vs No treatment) | 1 | 1 | None |

We see notable changes due to infection, but not due to treatment. However, this could be the result of the limited study, where only 6 samples were used.