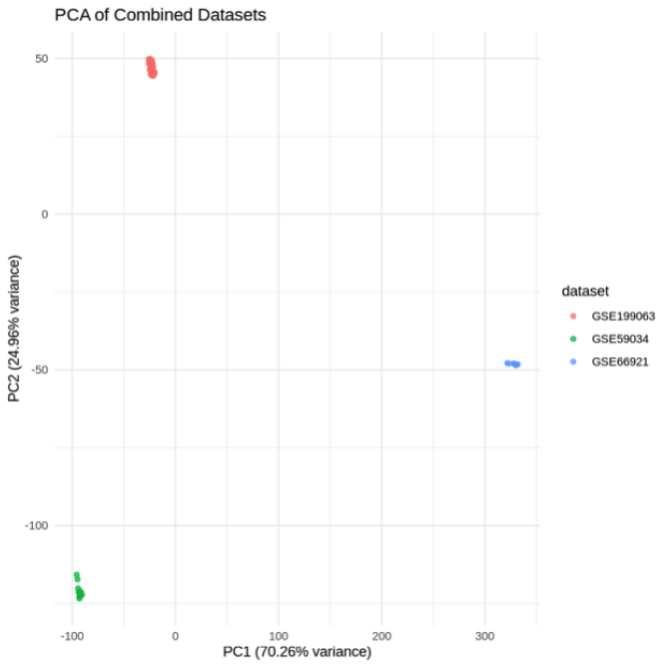
# Methodology - R code

## Second method (handling batch effect and combining all the samples into one)

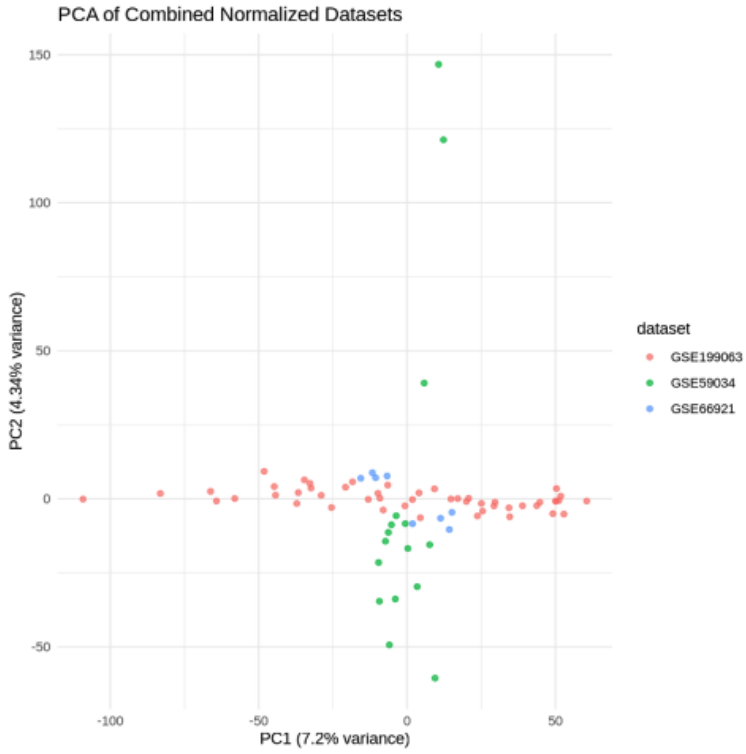
After noticing inconsistencies, we performed batch correction using the combat method and analyzed batch handling with PCA analysis. Then we combined all the normalized expression data from all the samples into one dataset and performed differential gene expression analysis to identify the top up-regulated and down-regulated genes after surgery.

To ensure that the technical variations from different microarray machines do not interfere with the results of the analysis, the ComBat function from the SVA package in R was used. ComBat is a method used to remove batch effects in microarray data by utilizing an empirical Bayes framework to adjust for both additive and multiplicative batch effects. It preserves biological variation while correcting for technical differences to help in integrating the data from different studies. Each filtered dataset was loaded and then converted to a matrix format to perform batch correction using the ComBat function from the SVA package. The distributions of the normalized datasets were visualized with a plot to verify the removal of the batch effect. A PCA analysis was performed on the datasets pre- and post-combat to see if it worked.

# PCA before and after combat




PCA before handling batch effect with comBat

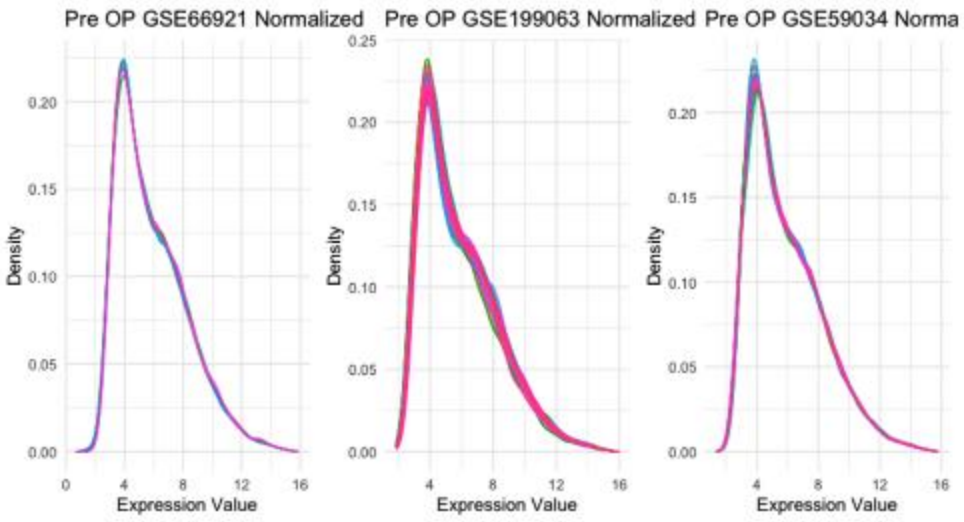



PC2 after handling batch effect with comBat

In the first PCA, samples from different datasets are clustered separately, indicating significant batch effects.

In the second image, PCA was performed on the same 3 datasets after using the CoBat method for handling batch effect. Samples from different datasets are more intermixed, indicating a reduction in batch effects but it has not improved completely therefore it was excluded from the final report.

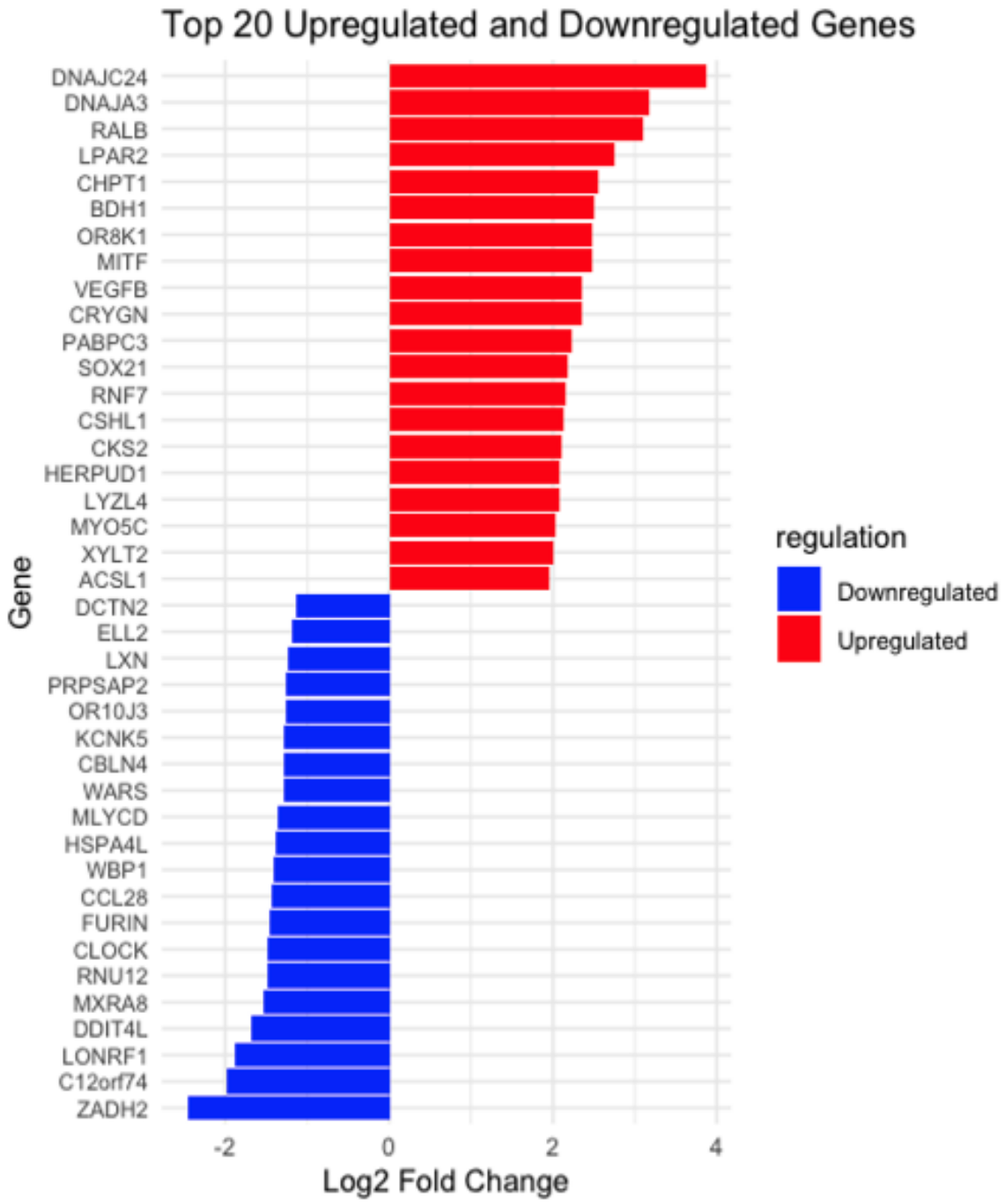
, it

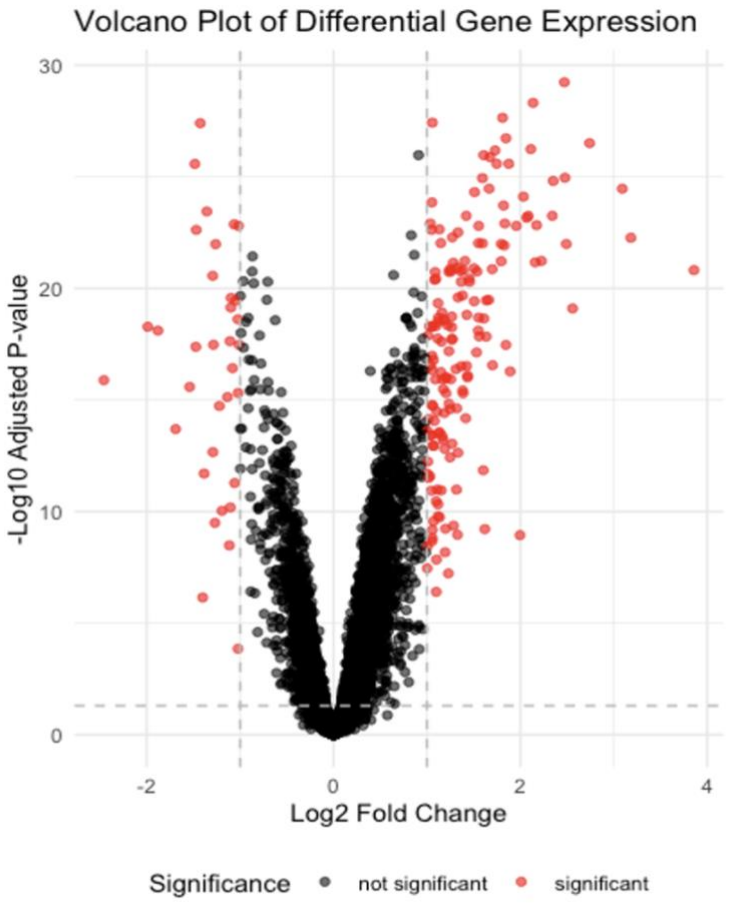



These plots were generated after correcting for batch effects to verify if the process was successful. By examining the image, the datasets appear to be properly normalized, as they show similar ranges and distributions of expression values. This suggests that the normalization was effective in making the datasets comparable, despite coming from different studies with different numbers of samples.

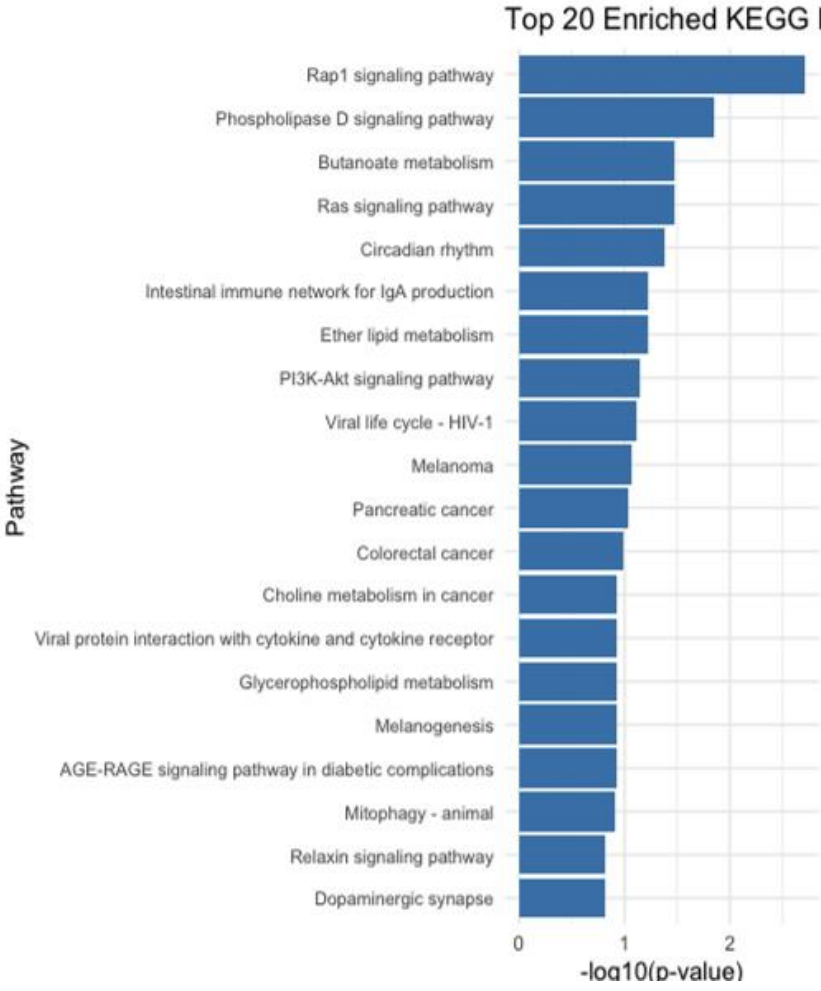
However, after reviewing the results of the top upregulated and downregulated genes, we noticed that there are no genes comparable across all three datasets. This indicates that even if the combat method performed well, it did so at the expense of knocking out all the common genes that were differentially expressed across all three datasets.

## Second Method - Combined dataset for Differential Gene Expression






This plot provides a clear visualisation of the differential gene expression results. The red points represent significantly differentially expressed genes, while the black points are not significant. The dashed lines indicate the thresholds for significance.



## Results of the differential gene expression and pathway analysis for the combined dataset (second method)

These upregulated genes demonstrate increased expression levels in the post-operative condition compared to the pre-operative state. While the downregulated genes show decreased expression in the post-operative condition compared to the pre-operative state.

The differential gene expression analysis identified 167 significantly upregulated genes, accounting for 2.99% of the total genes. The top upregulated gene is DNAJC24, a heat shock protein, which exhibited a log2 fold change of 3.86, corresponding to approximately a 14.5-fold increase in expression. On the other hand, 34 significantly downregulated genes were identified, representing 0.61% of the total genes. The top downregulated gene is ZADH2, Zinc-binding alcohol dehydrogenase, involved in metabolic processes, with a log2 fold change of -2.46.

The upregulated genes in this graph are primarily involved in stress response, protein folding, cell signaling, and membrane synthesis. Their upregulation indicates an active cellular response to the surgical intervention, aimed at protecting cells, promoting survival, and facilitating tissue repair.

The downregulated genes are mainly involved in lipid metabolism, protein degradation, mTOR signaling, and cell adhesion. Their downregulation suggests a strategic reduction in certain metabolic and cellular processes, possibly to conserve energy and resources during the recovery phase.

The enriched pathways identified in the analysis are the Rap1 signaling pathway, Phospholipase D (PLD) signaling pathway, and Ras signaling pathway. These pathways are crucial in regulating insulin-stimulated glucose uptake, lipid metabolism, and adipocyte differentiation. Dysregulation of these pathways can lead to insulin resistance, altered adipocyte function, and inflammation, all of which are key features of metabolic disorders such as obesity, type 2 diabetes, and metabolic syndrome. The findings suggest that these signaling cascades play a significant role in the metabolic adaptations associated with bariatric surgery.