Sarah Foster
Samy Kushwah
Gabriel Peer-Drake
Steven Miltenberger

How Stress Affects Meat Quality in Spanish Goats

I. Abstract

There are few studies that focus on the transcriptome of stress and how it affects meat quality, especially focusing on the species *Capra hircus*, known as spanish goats. Meat samples were collected from 54 Spanish goats that were randomly subjected to one of three treatments: 30 minutes of transportation, 180 minutes of transportation or the control was locked in a pen prior to slaughter. After the meat samples were collected, RNA sequencing was performed in order for analysis to be done looking at the effects of stress on the spanish goat's transcriptome. The precise genes responsible for meat quality that are influenced by Spanish goat's stress response is unknown. After collecting the RNA-seq, differential expression was run on the 3 groups and the gene counts in order to determine the significant genes that should be used in the analysis. In order to determine if there was a significant difference between the three groups the analysis techniques performed fell into three categories: enrichment analysis, clustering analysis, and statistical analysis. Through all three methods, the findings of this study showed less than significant levels of difference in gene expression between the control and stressed groups. Although our results affirmed the null hypothesis, there were results that continue to push for advancing research on Spanish goats as some major setbacks include a lack of published gene data for the species. This study has the potential to be a starting point for increasing interest in how stress affects the quality of products used for human consumption in goats and other organisms.

II. Introduction

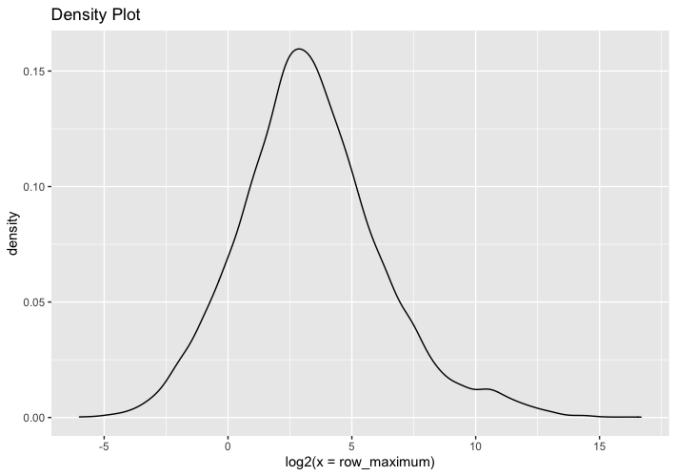
The meat quality of animals is conducive to the physical and psychological factors they experience prior to slaughter. The biological response to these factors at the genomic level is seldom the topic of study. The goal of this study is to determine whether stress influences the meat quality in Spanish goats by observing a difference in the transcriptome of goats that were stressed or not stressed. The data used was RNA-Seq data from Spanish Goats meat samples. The data contained 54 samples that included a control group and two experimental groups of stress for 180 minutes and 30 minutes. The data has 23,228 many different genes tested in all the samples. The approach we took to answer our question was to first perform differential expression analysis on our RNA-seq data. In assignment 2 we first took raw read counts, then used FPKM values in our analysis. We used a density plot, PCA plot, and TSNE plot to visualize the spread of our data. After, we performed differential expression and used a volcano plot to visualize the differentially expressed genes. Then we took genes that have a p value of less than .05 to use in our analysis. We then performed the gProfiler enrichment analysis and created a heatmap to visualize our data. In assignment 3, we performed four clustering methods on our most variable genes to find if any clusters emerged. Next we created another heatmap using those clusters. Then we ran statistical tests on our clusters to determine if there was significance to the groupings in the clusters that could answer our original question.

Multiple studies have sought to develop a relationship between the conditions livestock are subjected to prior to slaughter, and their subsequent meat quality. These studies generally conclude that psychological stress, such as excitement from transport, are attributed to the production of lactic acid, which is responsible for drier meat [1]. It is argued that all animals experience some level of stress prior to slaughter, which impacts the quality of the meat. The generally accepted explanation is that the magnitude of these negative effects is a function of the type, duration and intensity of these stressors [3]. While it is evident that animals experience stress during transportation to slaughter, stress factors that the animals explicitly experience are complex and non-quantifiable [6]. Due to the variation in the stress factors that affect meat quality, little development has been made as to the biological factors that are changed and their subsequent effect on meat quality [6][10]. The transcriptome of stress in Spanish goats serves to quantify the biological changes that are present due to the transportation of the goats prior to slaughter.

III. Methods

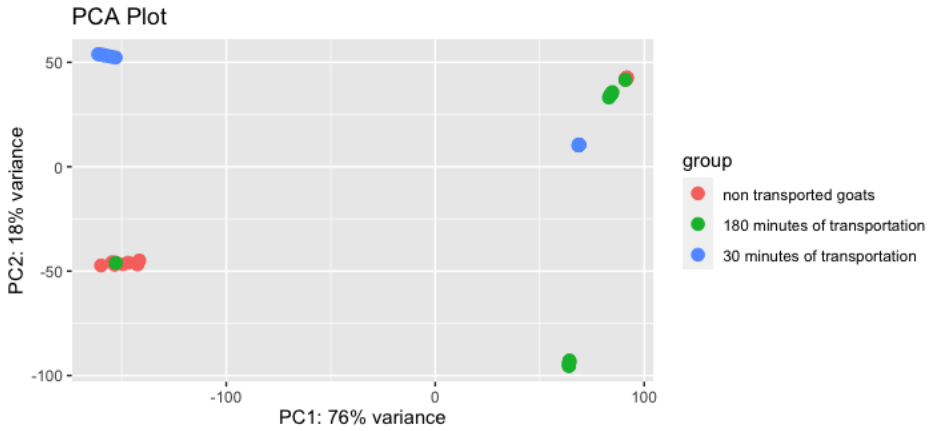
Visualization Methods:

Density Plot



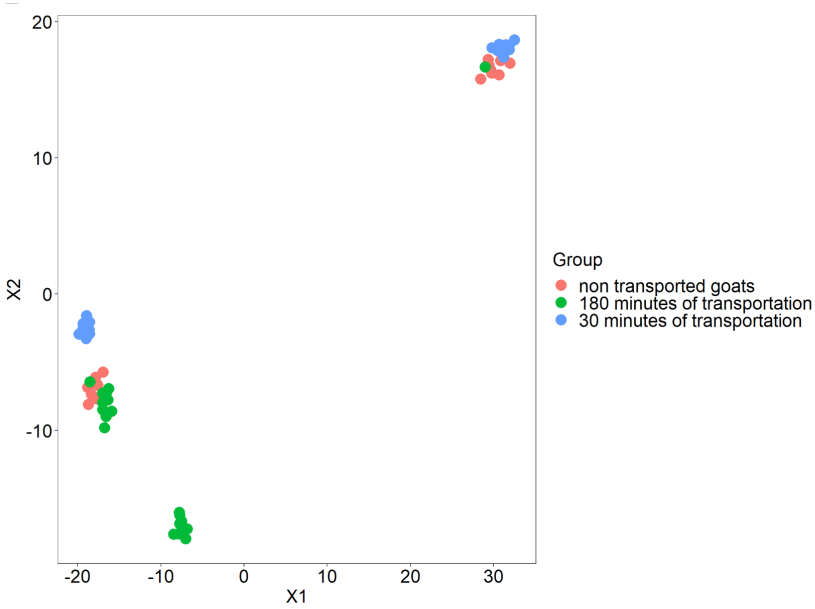
There is a lot of variation in the data as every gene has reads of 0 and higher reads for others. The density plot shows that the range of each gene's FPKM count was usually within the bounds of the density curve which starts around 2^{-5} and ends around 2^{12}. It appears that the largest amount of genes had ranges around 2^{3}. Something that is interesting about this plot is that even with normalizing the read counts into FPKM values there are still some genes with a very large gene count in comparison to the rest of the genes. Taking into account the log scale, genes in the 10-12 range have a considerably large read count in comparison to the majority of genes tested. This plot can be found in the file Assignment2.R in the github repository.

PCA Plot



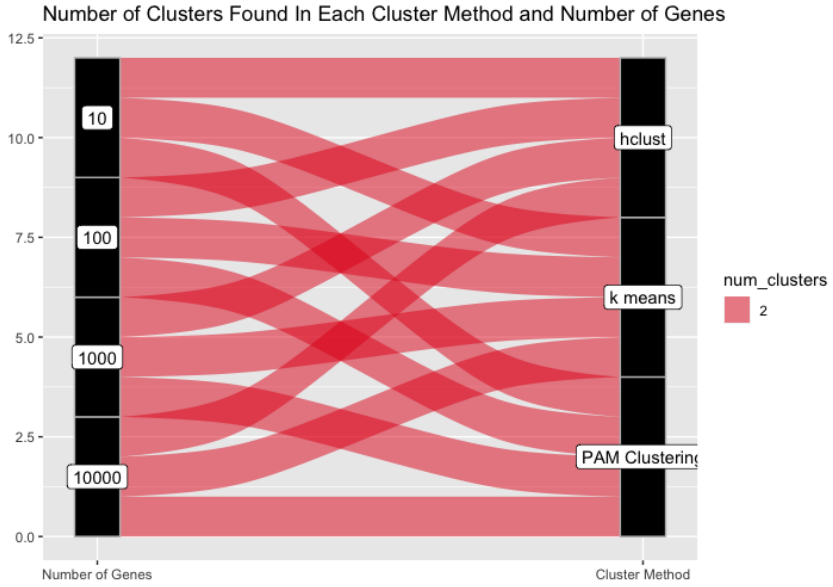
The PCA plot is used to reduce dimensionality of the data that we used. It created principal components, which are listed on the axes with the variance in the principal component. In our PCA plot, we can see that clustering of the groups is taking place based on the similarities in the sample, which is internally calculated and based off of the PCs. To highlight what the clustering means, the groups were distinguished using the metadata column for treatment.ch1, which included information about the amount of transportation for each group. The PCA plot tells us that most of the genes from their groups were clustered within their groups. For example, most of the non-transported goats have similarities and are mostly located around (-50,-50). The plot has specified parameters of a DESeqTransform data object and the intgroup parameter set to "treatment.ch1" which is used for grouping the data. This plot is good for the visualization of the variability of the original data. This plot can be found in the file PCAPlot.R in the github repository.

Tsne plot



The t-sne plot is also used to reduce dimensionality of data. In this plot, there are clusters of data based on only 2 components. It looks more at the local data trying to keep them in similar neighborhoods. The relevant information from this plot is the distance between the points in the plot. Similar attributes are grouped together. The data within each cluster is also similar. This is because it uses the neighboring technique to have points that were neighbors in the high dimensional data be neighbors in the low dimensional data as well. The parameters specified for this plot are a data frame and a label, which is treatment.ch1 groups like the PCA plot. This plot can be found in the file PCAPlot.R in the github repository.

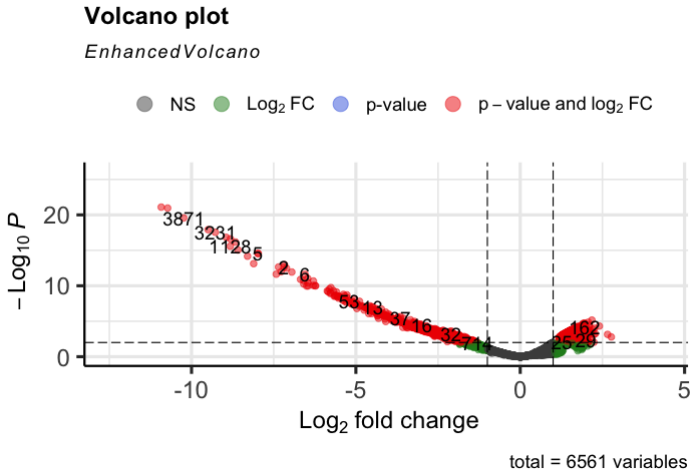
Alluvian Diagram



Rerunning all of the clustering techniques with 10, 100, 1000, and 10,000 genes did not yield a different amount of optimal clusters. All four of the clustering methods still had a preferred k = 2, which can be seen in the alluvial diagram below. One interesting observation in the hclust method was that as the number of genes were decreased, the smaller cluster of the two clusters had less samples within it. This is probably because using the less significant genes in clustering made it more difficult to differentiate the two cluster groups but as less and less genes were being used it became more obvious which samples were most similar to each other.

<u>Differential Analysis Methods:</u>

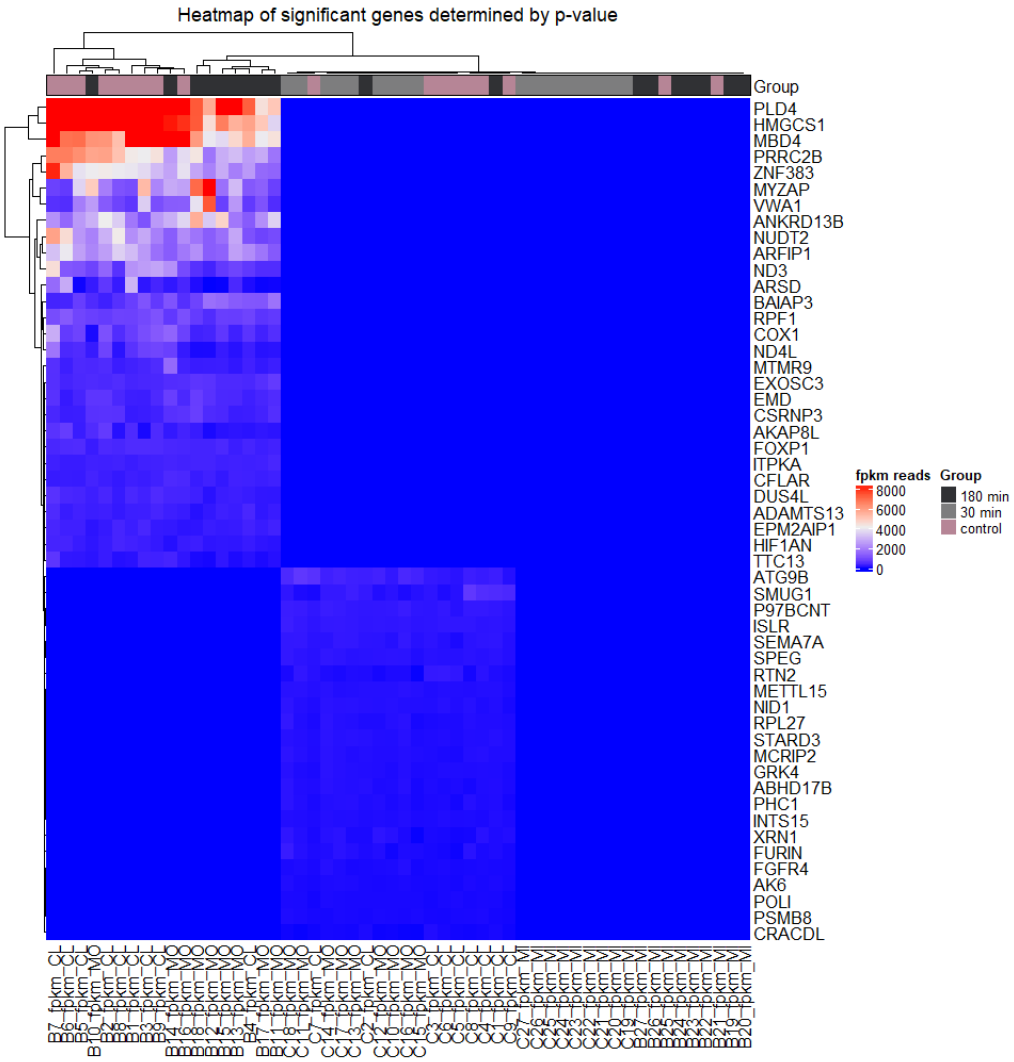
Volcano Plot



The volcano plot is a type of scatter plot. This plot is relevant in visualizing large magnitude changes. This plot puts together a measure of statistical significance with the magnitude of changes. Data that is higher up on the plot is meant to have more significance. This can be from the left and right side of the volcano. As we can see, there are a lot of genes in our data that are highly significant. Most of these significant genes are on the left, top side of the plot. These will be included in the differentially expressed table of genes. The parameters that were specified for this plot were a data frame, lab which is set to the rownames of the data frame, the x and y axis which is $\log_2$ fold change and $-\log_{10}P$ respectively, and pcutoff set to 0.01 where the horizontal dashed line is on the plot.

The reason the volcano plot is mostly towards the left is because the genes in this portion of the graph are downregulated. This means that these genes decrease the amount of a cellular component when responding to outside stimulus. This plot can be found in the file PCAPlot.R in the github repository.

Heatmap Assignment 2



The heatmap shown above displays the relationship between each sample (x-axis) and significantly expressed genes (y-axis). The relationship is presented by the color of a box that represents a sample, group pair. According to the legend on the left, dark red boxes indicate a high number of fpkm reads, while dark blue boxes indicate a low number. In this map, it is evident that despite the statistical significance of the displayed genes, (statistical significance determined through p-test), very few numbers of reads were present for goats that experienced 30 minutes of stress. Contrarily, the primary cluster of reads is associated with the control and 180 minutes of stress groups. Additionally, there is a faint cluster of reads in the bottom half of the map that appear mostly for goats that are in the 180 minutes of stress group. This could be attributed to the stress that the goats experienced altered the frequency that certain genes were read. This plot can be found in the file heatmap.R in the github repository.

Heatmap Assignment 3

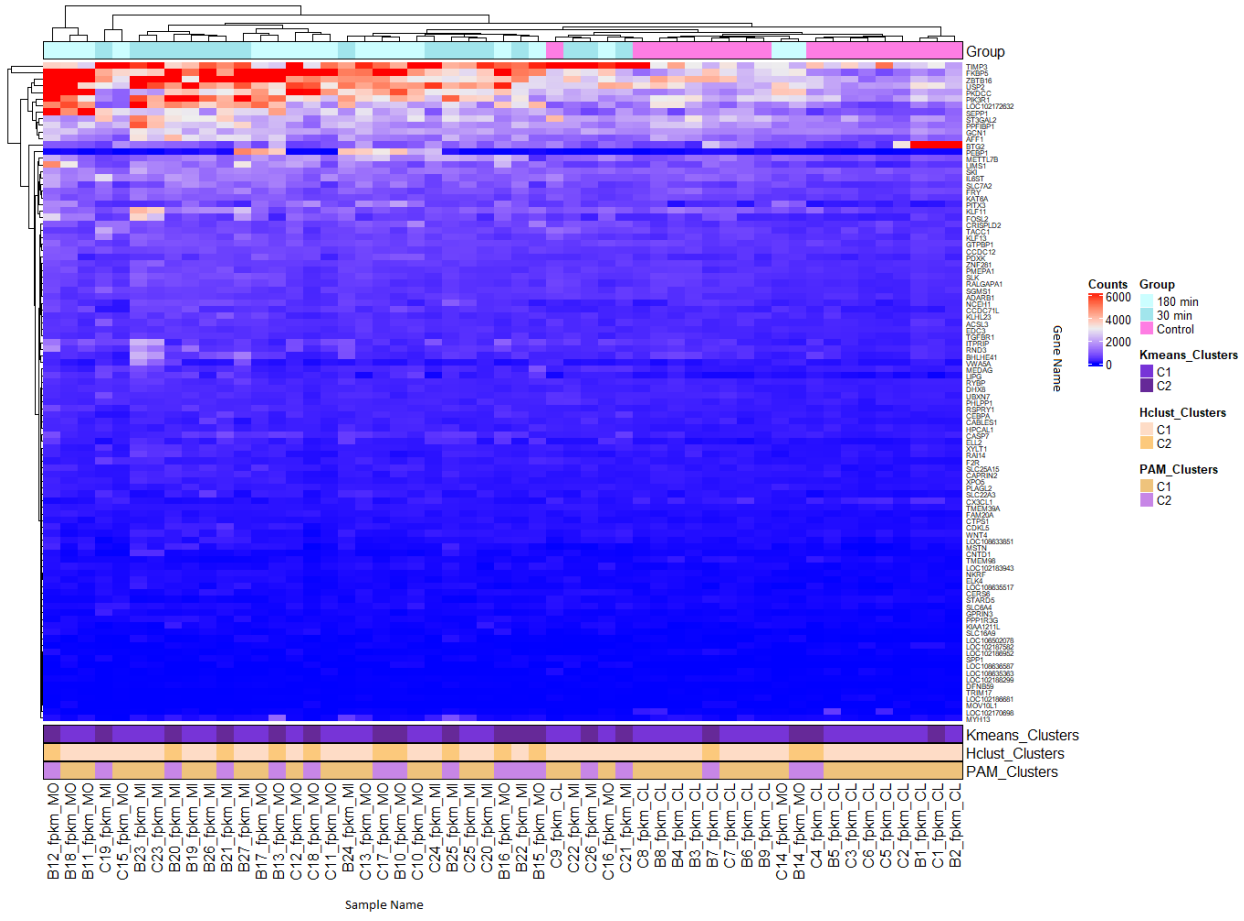


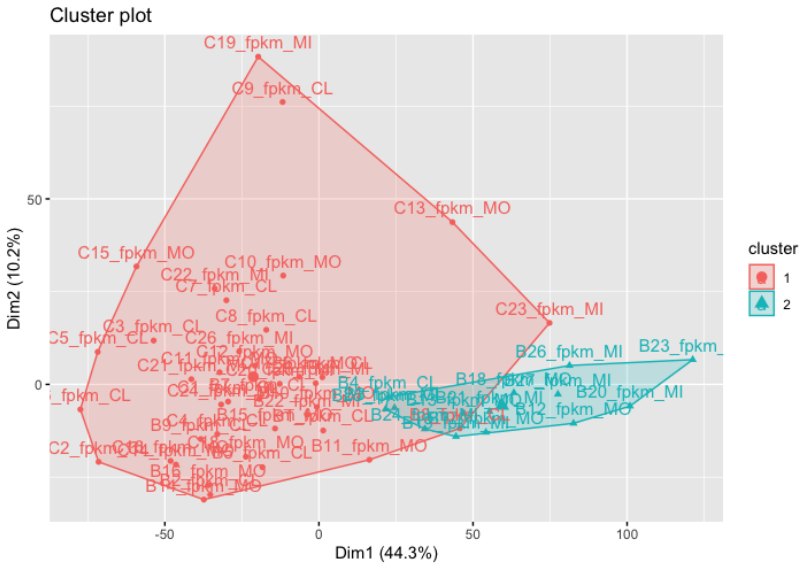
Image link for better resolution:
https://drive.google.com/file/d/1j4FiTycEYBkSEHDkr9F-x6r20rumqDp0/view?usp=sharing

To create the heatmap we filtered the top 100 most significant differentially expressed genes from our initial matrix of 5,000. We filtered and determined significance based on the p-value. From there, in order to obtain reliable ordering in the multiple side bars included on the

plot, we ordered the data in our expression matrix alphabetically within each group, e.g. B1_CL, B2_CL… B19_MI, B20_MI. We then arranged the results of our cluster methods to align with this ordering. Doing this allows us to include the results of the cluster methods as an annotation sidebar in our heatmap. This annotation allowed us to view side by side the similarities and differences of the results of each method by comparing which sample belonged to which cluster. From this, we can see that across each method of clustering, C1 contains a higher volume of control samples as compared to C2, additionally, the samples in the C1 cluster have less reads on average compared to samples in the C2 cluster, which could indicate that the C1 cluster shows the significance of the difference between the control and non-control groups. This plot can be found in the file heatmap_assignment3.R in the github repository.

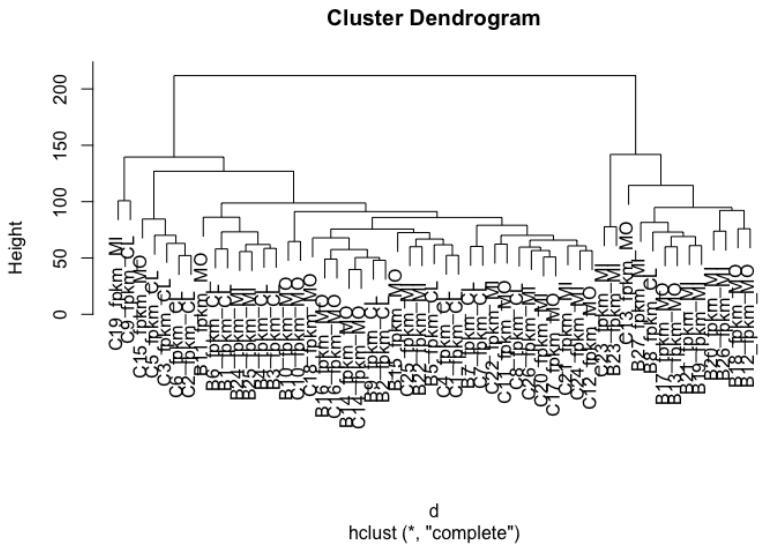
Clustering Methods:

K means



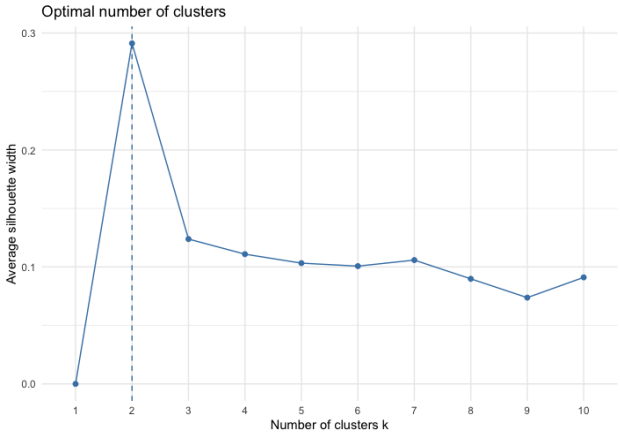
The k means method requires a k to be chosen. As k values were increased, the clusters became less defined and overlapped more, indicating that a higher cluster value did not give better results. Even as k was increased, there was still a large mixture of affected goats and control groups mixed in the clusters. As clusters were increased past two it was apparent that raising k did not improve the results, and that k = 2 was the best value to be chosen. It is interesting how one cluster is larger than the other, and how cluster 1 contains the majority of the control group while cluster 2 contains mostly affected goats with only one control goat in the cluster. This indicated that there may be a difference between the stressed goats and control goats. This plot can be found in the file Assignment3.R in the github repository.

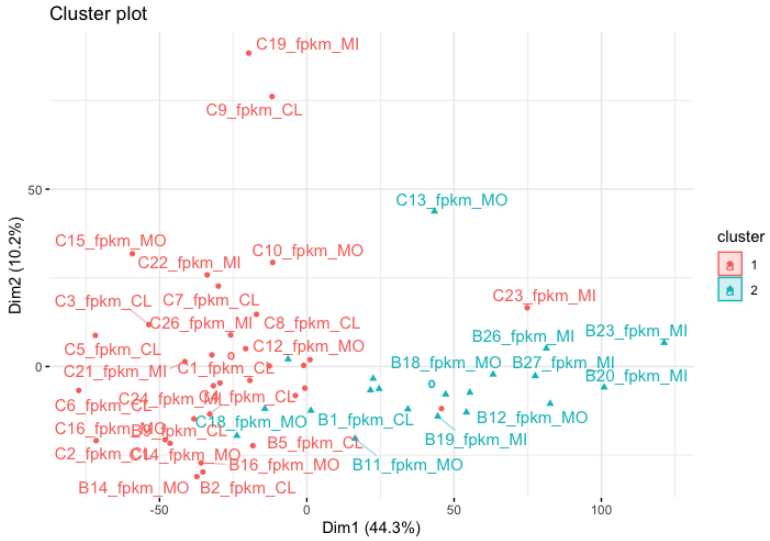
Hclust



This method does not require a k to be chosen, and it selected k = 2 to be the optimal amount of clusters. Despite there only being two main clusters, one of the interesting elements of hclust is how individual samples were grouped together within the dendrogram. It can be seen on a lower level that control goats and stressed goats are often grouped together within the clusters which signifies there may be a difference between the stressed and not stressed goats in terms of expression. This plot can be found in the file Assignment3.R in the github repository.

PAM Clustering

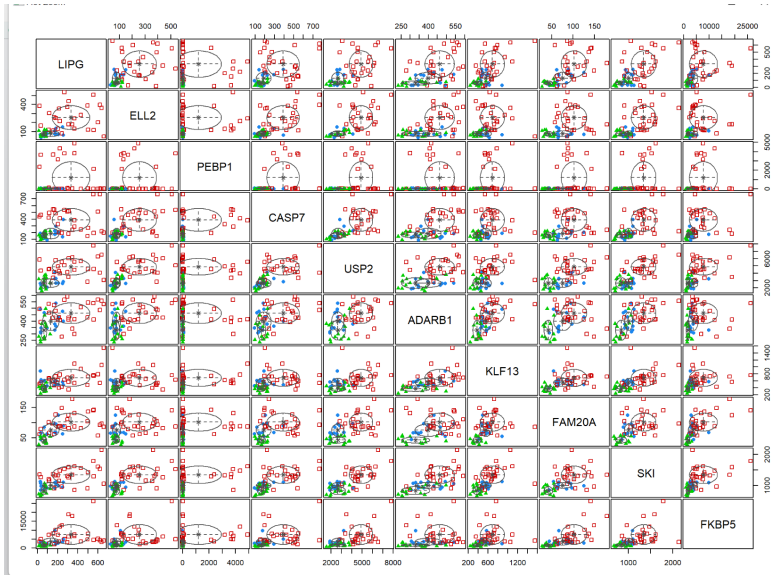


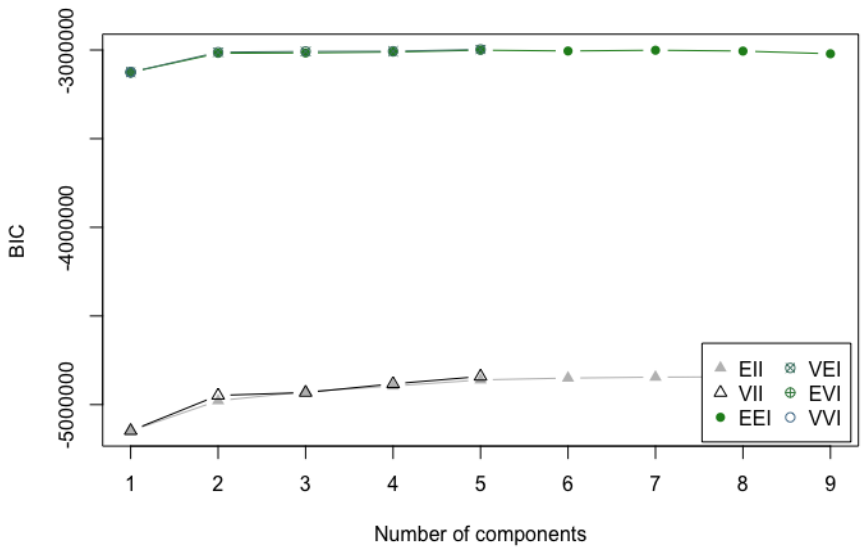


This method requires a k to be chosen, although it also provides a function that advises you on which k should be chosen. The optimal number of clusters was shown to be two, which is why k = 2 was chosen. Below is the graph that shows the different k values that were tested and the max at k = 2 is why this k was chosen.

The PAM cluster plot shows it is apparent that there is some overlap between the two clusters which makes it difficult to define when one begins and the other ends. This graph looks very similar to the k-means plot, but it is interesting that the samples in each cluster were not the same between both clustering techniques. This one has a similar structure as the last two methods, where there is a large dominating cluster with the majority of the control group and then the second, smaller cluster, has majority stressed goat groups. This plot can be found in the file Assignment3.R in the github repository.

Gaussian Mixture Model



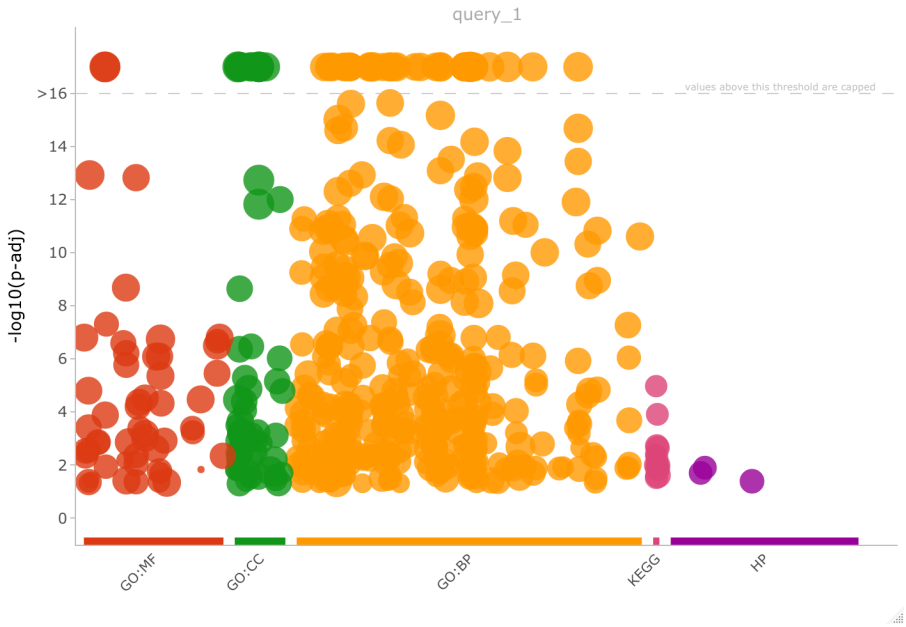


This method did not require a k to be chosen. The Gaussian Mixture Model (GMM) has many different Gaussians that all use k for the number of clusters in the dataset. The methods have variable names such as EII, VEI, VII, and so forth. The graph below is made from the mclust function that does the functionality for GMM. In this plot, the Bayesian Information Criterion (BIC) values are chosen and displayed. BIC values are a criterion for model selection and are used to get the number of clusters. The reason BIC values were chosen and not "classification" type, which shows pair clustering, is because there were too many genes to each have a pair plot. Above is a plot using the "classification" type for 10 genes when I tested my code. This is a crowded plot and having more than 10 genes would make the plot too large to display, so plotting BIC values is more optimal.

After analyzing the BIC values plotted, it was observed that as each number of genes were tested, the data became less and less readable, since there were a lot more genes we were looking at. The plot for 10 genes had a lot more distinct cluster models lining up that matched up to 2, the optimal number of components. As the values grew, the lineups of the clusters did not seem not as distinct and seemed like it was displaying only 1 clustering model. It is harder to see in larger datasets, so for this plot, the matrix with 100 gene samples is displayed. A lot of clusters all align at the point 2 while at other points there are less clustering methods present. Therefore, 2 is the best number of clusters that comes out of the dataset. This plot can be found in the file Assignment3pt2.R in the github repository.

<u>Enrichment Analysis Methods:</u>

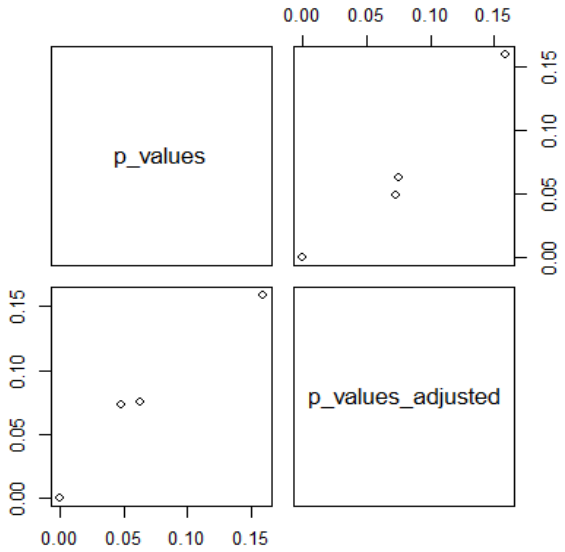
gProfiler



Enrichment analysis was performed using gProfiler mainly due to the fact that it had our species included in its package. After performing the analysis a Manhattan plot was used to see where the genes had representation in online databases. It appears that the genes were found mostly in the BP(Biological Processes) ontology and also had a high concentration among MF(Molecular Functions) ontology and CC(Cellular Component) ontology. Something interesting about this graph is the quantity of genes that were linked. Since goats are a less common species to study it was surprising how many of the genes were linked to different ontologies. This plot can be found in the file PCAPlot.R in the github repository.

Statistics

	Tests	p_values	p_values_adjusted
1	Samples/Kmeans	1.590e-01	1.5900e-01
2	Samples/Hclust	4.875e-02	7.3125e-02
3	Samples/PAM	6.262e-02	7.5144e-02
4	Kmeans/Hclust	7.152e-09	1.4304e-08
5	Kmeans/PAM	3.036e-12	1.8216e-11
6	Hclust/PAM	1.540e-09	4.6200e-09



To get the statistical test results for multiple variable hypothesis testing we ran Chi-Square tests of independence on our sample groupings compared to the clusters found using different clustering methods, as well as comparisons of cluster methods to other cluster methods. In doing so, we gathered raw p values as well as p values that had been adjusted for multiple hypothesis testing. As seen in the statistical test results table below, when compared to our samples' original groupings, only the Hclust clustering method resulted in a p value with an acceptable statistical significance, however once adjusted for multiple hypothesis testing, none of the cluster methods produced a statistically significant result. On the other hand, when comparing the clustering methods to other clustering methods, we can see that all of their p values and adjusted p values fall within the acceptable range of statistical significance. This tells us that the original grouping for our samples may not accurately indicate differences in DNA, but there is some characteristic that separates the goats into two groups.

We created the enrichment plot below using the p values and adjusted p values we found while carrying out our Chi-Square analysis. The enrichment plot is useful in helping us visualize the range of values in the pairs of statistical test results between all metadata and clustering results we collected. From the plot, we can see that there is a relatively linear trend between the pairs of p values and adjusted p values found from the different comparisons between our sample groups and calculated clusters. This plot can be found in the file Statistics.R in the github repository.

All of our code and analysis can be found at our Github Repository:
https://github.com/smiltenberger/SpanishGoats

IV. Results

We were able to find an answer to our original question of whether stress influenced meat quality in Spanish goats and whether or not there was a difference in the transcriptome of goats that were stressed or not stressed. We found that stress did not influence meat quality and that there was no difference in the goats' transcriptomes based on if they had been stressed or not.

The goat samples were all initially labeled as either control, stressed for 30 minutes or stressed for 180 minutes. Upon clustering analyses we expected to find 3 clusters that corresponded to these three different groups. Had our results fit this expectation, this would have indicated that the different amounts of stress the goats had been exposed to affected their DNA. However, in each clustering method that we used, we found that 2 clusters was the optimal number of clusters, not 3. Thus, these results do not correspond to our original prediction, but they do indicate there is a different way to group the goats based on their DNA sequences.

Among the clustering methods we performed, many of them (PAM, Hclust, K means) grouped a majority of the control group into one of the clusters, and the other cluster was a majority of stressed goats (with stress being either 30 minutes or 180 minutes). This is an unexpected result as we had assumed that there would be a more clear distinction between each of the three groups that samples were originally placed in (control, stressed 30 minutes, stressed 180 minutes). Additionally, it might have made more sense for the control group to be in a smaller, more distinct cluster than the stressed goats, but the clustering methods grouped most of them in the larger cluster, mixed in with stressed goats.

Another finding we were not expecting was the number of genes from our data that were linked to different ontologies when performing an enrichment analysis using gProfiler. Before this analysis, we were aware that goats are a less popular species to perform studies on and had assumed that there would be very few of our data's genes linked to onologies as a result. However, much to our surprise, there were a large number of genes linked to different ontologies like MF(Molecular Functions) and CC(Cellular Components), with an especially large number of genes linked to the BP(Biological Processes) ontology.

One of the weaknesses in our project was that we had issues with trying to do the ontology methods. Since goats do not have readily available databases in these ontologies we struggled to figure out these methods. This is why we were only able to do the gProfiler method after attempting other methods. To improve upon this weakness, if we were to do the project again, we would do more research about other data existing on our species before picking it to run analysis on. We would have chosen a different species since Spanish goats have less data. Another solution would be to choose some different analysis methods that have existing data with our species. This would require in our future work that we would need to do more research on what methods are usually used with our species and consult more past research to understand better how to approach analyzing species that are less researched.

In terms of bioethical issues, one of the recent ones to come to the forefront is the ethical treatment of animals used in experiments. This includes rearing animals for use in experiments and for killing them when unnecessary. While we did not avoid the issue of killing animals

without reason, the study was conducted so that the goats were treated with the best care where they were pasture raised and had all the food and water they needed. The goats were treated well during their life so there are no ethical concerns with the raising of the animals, but there are some concerns with the source of the goats and if they were bred with the intention of being killed for the experiment. In this case where the study was made to analyze the meat quality of goats this was an unavoidable outcome of the experiment.

V. Conclusion

Our original hypothesis was if stress has an effect on the transcriptome of spanish goats. Through our methods using visualization, differential expression, clustering analysis, enrichment analysis, our overall conclusion was that there was not a significant difference between the transcriptome of stressed goats and the control group. In assignment 2, using differential expression we were able to find differentially expressed genes and a large majority were connected to different gene ontologies. When taking these genes and using them for analysis in assignment three, we were better able to find the answer to our original question. When we ran all four clustering methods, we found that two clusters were optimal in each of the four methods. When more closely analyzing these clusters in our statistical analysis, it was found that there was not a significant difference between the two clusters. Our clusters also did not group the stressed goats and control group exactly, meaning that there was some overlap in differential expression in both groups. In conclusion, through all of our methods, the answer to our original hypothesis is that stress does not affect the transcriptome of goats.

References

1. Grandin, T. (1980). The effect of stress on livestock and meat quality prior to and during slaughter. International Journal for the Study of Animal Problems, 1(5), 313-337.
2. Miranda-de la Lama GC, Villarroel M, María GA. Livestock transport from the perspective of the pre-slaughter logistic chain: a review. Meat Sci. 2014;98:9–20.
3. Ferguson DM, Warner RD. Have we underestimated the impact of pre-slaughter stress on meat quality in ruminants? Meat Sci. 2008;80:12–9.
4. Odore R, D'Angelo A, Badino P, Bellino C, Pagliasso S, Re G. Road transportation affects blood hormone levels and lymphocyte glucocorticoid and β-adrenergic receptor concentrations in calves.
5. Lomiwes D, Farouk MM, Wu G, Young OA. The development of meat tenderness is likely to be compartmentalized by ultimate pH. Meat Sci. 2014;96:646–51.
6. Minka NS, Ayo JO. Physiological responses of food animals to road transportation stress. Afr J Biotechnol. 2010;9:6601–6613.
7. Gregory NG. Animal welfare at markets and during transport and slaughter. Meat Sci. 2008;80:2–11.
8. Loredo-Osti J, Sánchez López E, Barreras Serrano A, et al. An evaluation of environmental, intrinsic and pre- and post-slaughter risk factors associated to dark-cutting beef in a Federal Inspected Type slaughter plant. Meat Sci. 2019;150:85–92.
9. Kannan G., Kouakou B., Terrill T.H., Gelaye S. Endocrine, blood metabolite, and meat quality changes in goats as influenced by short-term, preslaughter stress. J. Anim. Sci. 2003;81:1499–1507.
10. Xing T., Gao F., Tume R.K., Zhou G., Xu X. Stress effects on meat quality: A mechanistic perspective. Compr. Rev. Food Sci. Food Saf. 2019;18:380–401.