Given data: RNA-seq data from two house mouse (Mus musculus) tissues (Heart, Liver) across # tow sampling times (ZT0, ZT12), with a biological replicate for each tissue and sampling time, # resulting in a total of 16 paired-end FASTQ files.

---

1. Alignment Statistics Summary:

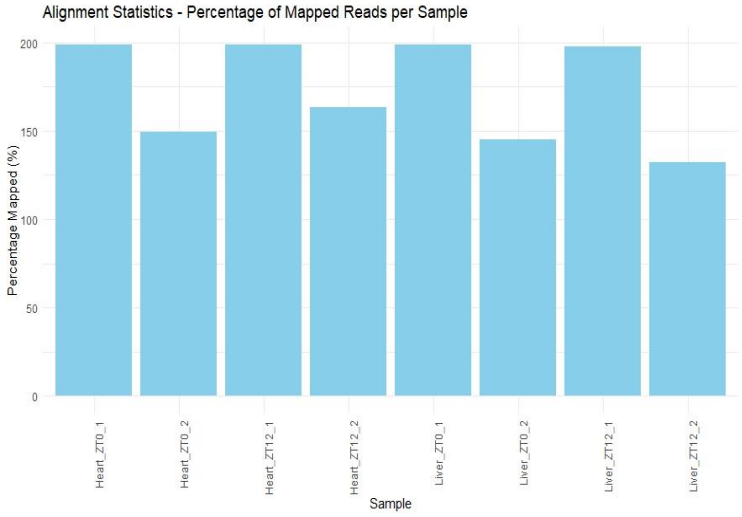


Fig. 1 Alignment Statistics, in brief the plot tells the percentage of read mapped to the genome for each sample, higher mapping rates resulting in good data

The percentage of mapped reads varies among samples, with some samples (like Heart_ZT0_1 and Liver_ZT12_1) showing higher mapping rates, potentially close to or above 200%, which may indicate a normalization scale difference or possible re-check for outliers. High mapping rates typically suggest good alignment quality and possibly good data quality, assuming consistent normalization

Lower percentages, as seen in samples like Heart_ZT0_2, Liver_ZT0_2, and Liver_ZT0_2, may indicate lower quality or potential issues in alignment for these specific samples or closer inspection to understand if biological variability, technical issues, or quality differences impacted them.

2. Principal Component Analysis (PCA) & Dispersion plot:

PCA can help visualize the variation across samples by grouping samples by tissue (Heart, Liver) and time (ZT0, ZT12).

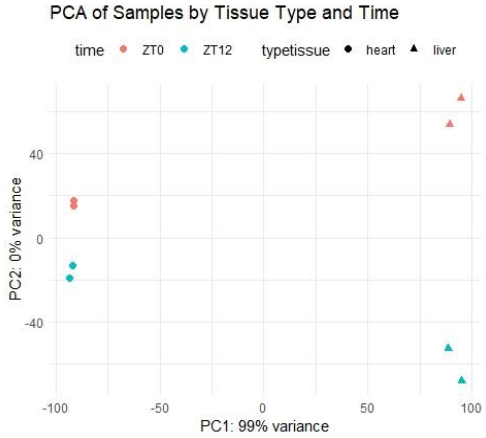
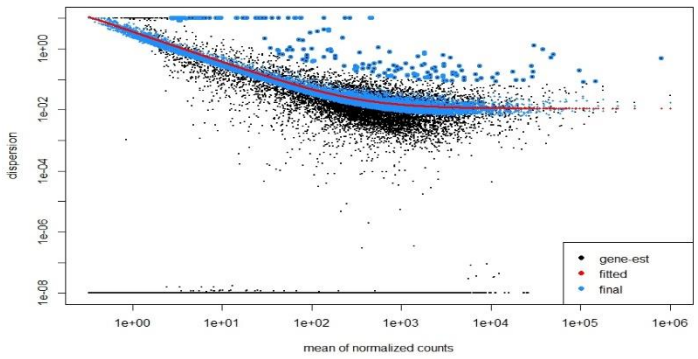


Fig. 2 PCA showing separation by tissue type, with heart and liver samples clustering distinctly, and by time (ZT0 vs. ZT12), which may separate samples within the tissue clusters. Biological replicates are expected to cluster closely, indicating reproducibility.

In this dataset, PC1 accounts for 99% of the variance, suggesting that most of the variation can be attributed to a single factor—likely the tissue type. The Heart and Liver samples are separated along PC1, indicating that tissue type is the primary driver of gene expression differences. PC2 accounts for much less of the

variance and may represent other sources of variation, such as sampling time (ZT0 vs. ZT12). For instance, within the Heart and Liver clusters, slight separation along PC2 could indicate differences due to sampling times (ZT0 and ZT12), although this effect is minimal compared to tissue type. The biological replicates for each tissue and time point cluster closely, further indicating good reproducibility within each group

Fig. 2 Dispersion plot implies genes with low dispersion estimates are shrunken towards the curve, and the more accurate, higher shrunken values are output for fitting of the model and differential expression testing.



This is a good fit for the DESeq2 model. You expect your data to generally scatter around the curve, with the dispersion decreasing with increasing mean expression levels[1]

2.Scatter Plot & Heatmap of Expression Levels:

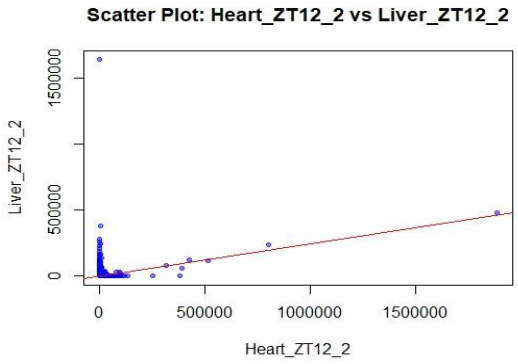


Fig. 3 In the given scatter plot, if there's a high correlation between samples, points should align along a diagonal line, indicating similar expression levels across both samples.

However, in this plot, most points are clustered around the axes, suggesting that the Heart and Liver samples at time point ZT12_2 have distinct expression profiles rather than consistent, reproducible patterns. This aligns with biological expectations, as Heart and Liver tissues are expected to express different sets of genes due to their distinct functions. Thus, the scatter plot shows low reproducibility between tissues (Heart vs. Liver) but would likely show a stronger correlation if comparing replicates within the same tissue and time.

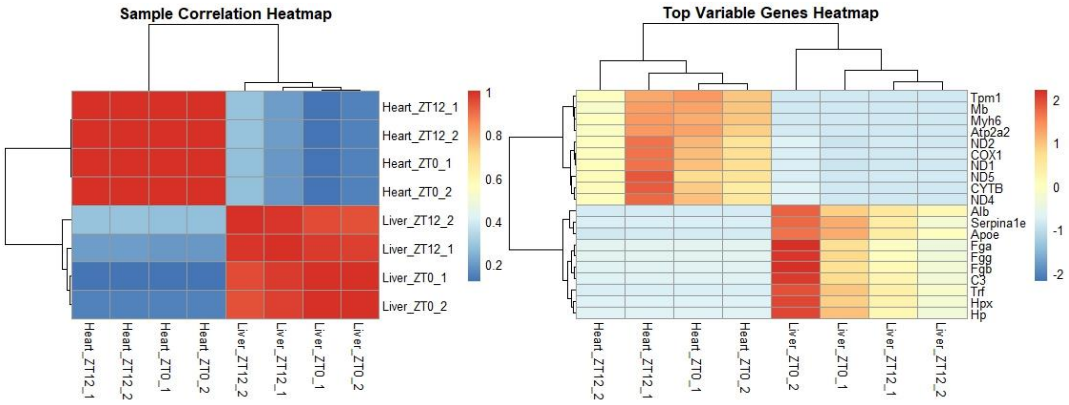


Fig. 5 Sample correlation and top variable genes cluster by tissue type and time, with biological replicates appearing together. This clustering shows that biological replicates are consistent, and variation is higher between different tissues and time points than between replicates of the same group.

The correlation heatmap provides a visual representation of the similarity between each pair of samples. In this heatmap, each cell represents the correlation coefficient between two samples, with red indicating high correlation and blue indicating low correlation. The samples appear to cluster based on tissue type and time, with biological replicates clustering closely together. For instance, Heart_ZT0_1 and Heart_ZT0_2 are highly correlated, as are Liver_ZT12_1 and Liver_ZT12_2, which indicates strong reproducibility within each group (tissue and time point). Patterns show up- or down-regulation trends across conditions, helping identify tissue-specific expression.

4. Differential Expression Analysis & Top DEGs Expression Patterns

Using paired contrasts, identify DEGs between tissues at each time point: Differentially Expressed Genes (DEGs) with Tissue-Specific, Time-Specific, and Interaction Effects Tissue time

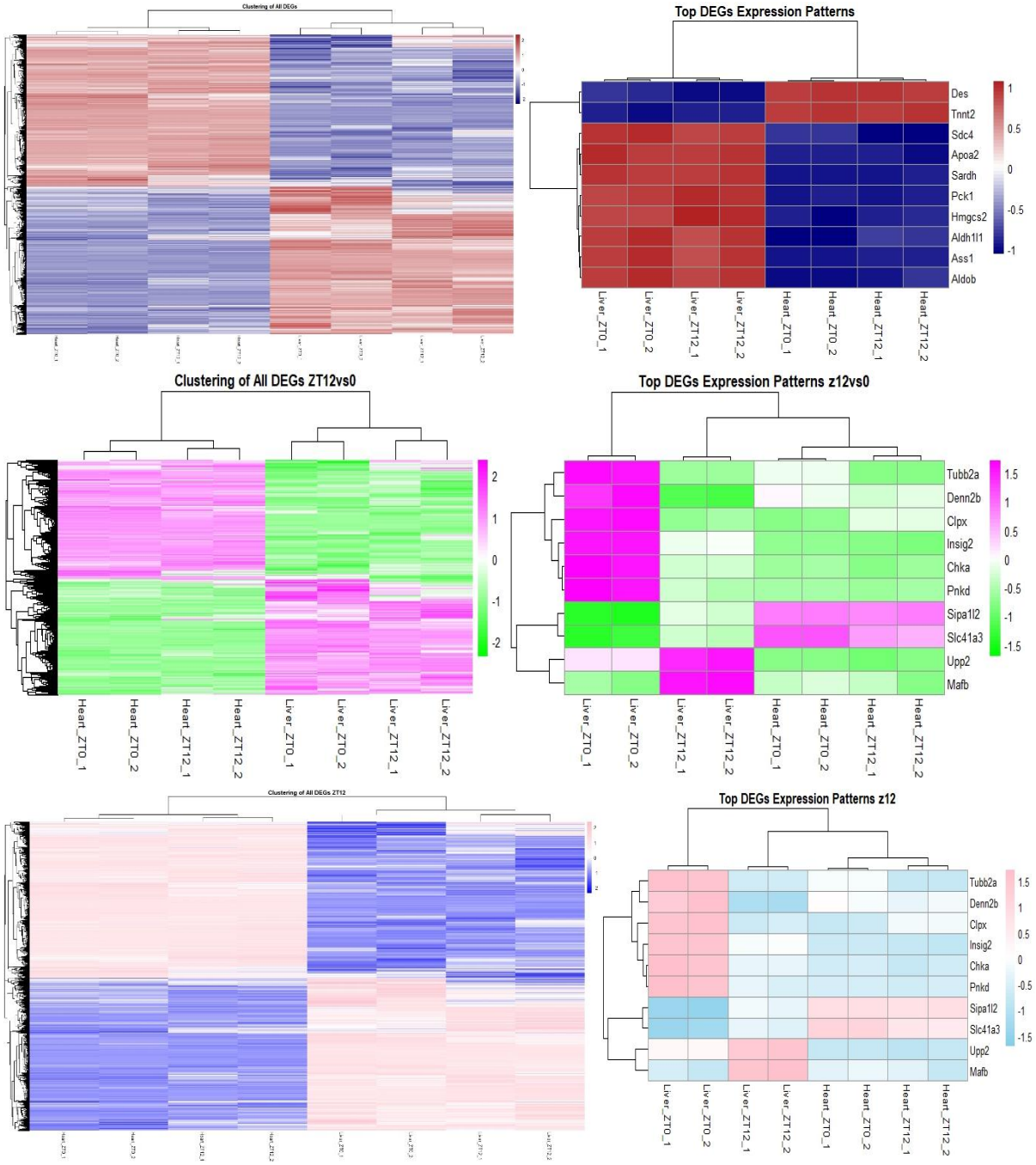


Fig. 6 The differences in gene expression between Heart and Liver tissues, between different sampling times (ZT0 and ZT12), and the interaction effect of tissue type with time. Here interaction specific and time specifc are similar.

The heatmaps for DEGs expression patterns highlight clustering based on expression similarity. Samples that cluster together have similar expression patterns. In the heatmap showing tissue-specific DEGs, you can observe distinct clusters for Heart and Liver, indicating that gene expression patterns are tissue-specific. For example, Heart samples cluster separately from Liver samples, which is expected as different tissues express distinct gene sets. In the time-specific heatmap (ZT12 vs. ZT0), you can see clustering that indicates genes affected by time within each tissue, though the effect of time appears less prominent than tissue type. The ZT12 vs. ZT0 heatmap shows a list of top DEGs, and clustering patterns indicate time-dependent gene regulation within tissues, reflecting circadian or time-based biological processes.

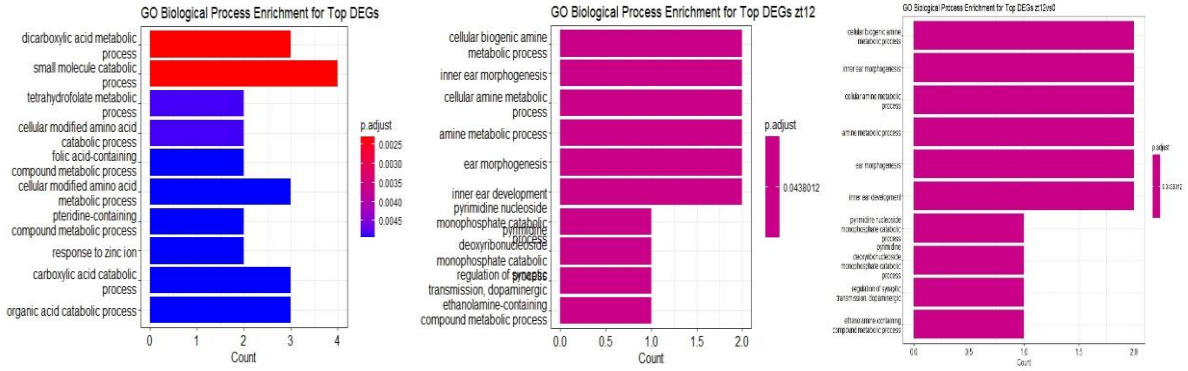


Fig.9 Biological process top DEG

In the heatmap of top DEGs between Heart and Liver, genes like Tubb2a, Insig2, and Upp2 show specific expression levels in either Heart or Liver, which reflects their roles in tissue-specific functions. Genes highly expressed in the Heart may be involved in cardiac-specific pathways, such as muscle contraction or energy metabolism, while those in the Liver are likely related to metabolic processing and detoxification. For genes affected by time, such as Sipa1l2 and Chka, there might be a role in circadian regulation or in processes that fluctuate based on time, like hormone release or metabolic activities.

Differences between ZT0 and ZT12 may also indicate genes responding to day-night cycles, which is significant in tissues like the Liver that are involved in metabolic cycles.

5. Volcano Plots

Used volcano plots to visualize DEGs for each comparison (Heart vs. Liver at ZT0 and ZT12):

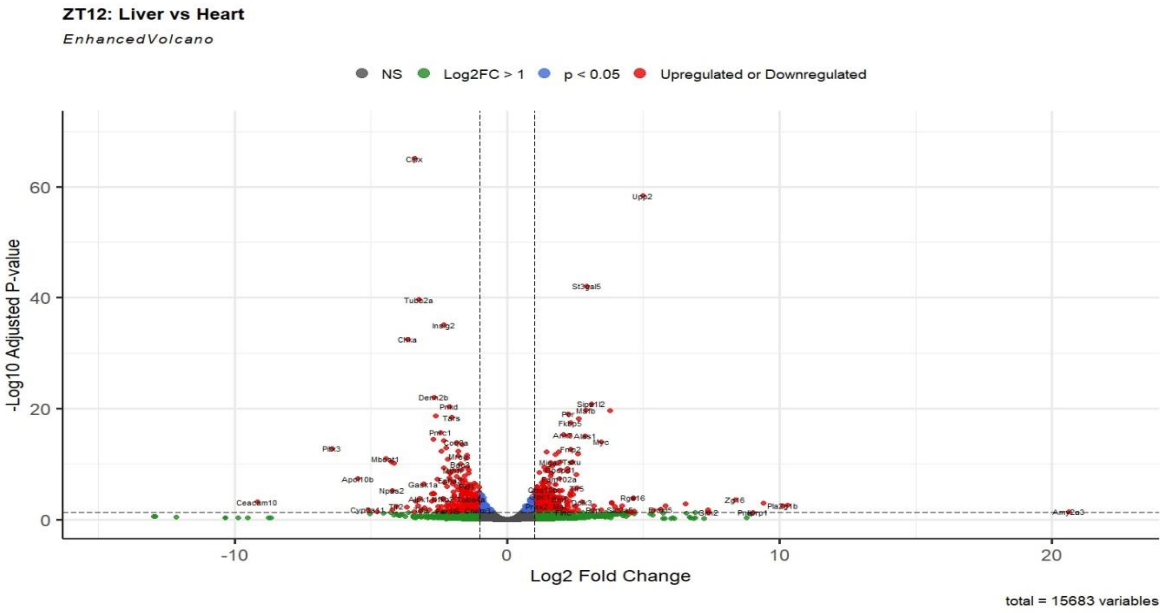


Fig. 7 The volcano plot for the ZT12 time point comparing Liver and Heart samples displays the Log2 Fold Change on the x-axis and the -log10 Adjusted P-value on the y-axis. Points in the upper left and right quadrants represent significant DEGs with high confidence, where red indicates upregulated genes (positive Log2 Fold Change) in one tissue compared to the other, while green indicates downregulated genes

The significant DEGs with high Log2 Fold Changes indicate genes that are either highly expressed in the Liver relative to the Heart or vice versa, at the ZT12 time point. The presence of several significantly upregulated and downregulated genes shows a distinct tissue-specific expression pattern, as expected between Heart and Liver.

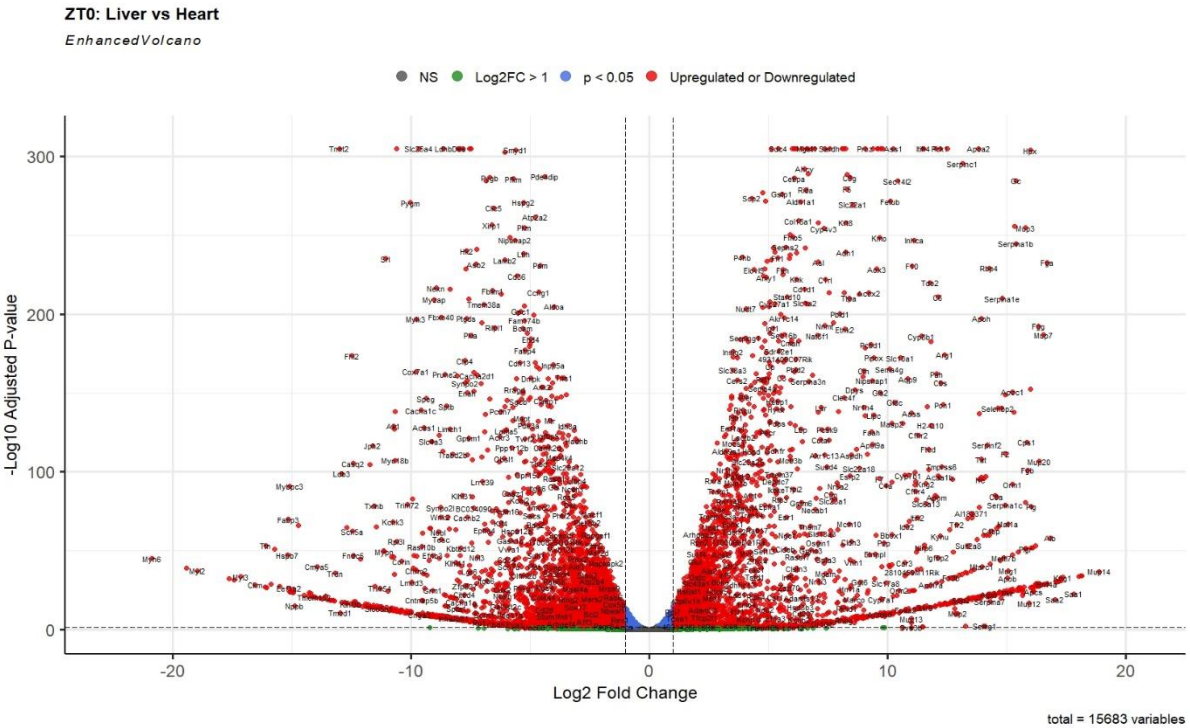


Fig. 8 The volcano plot at ZT0 between Liver and Heart would show a similar layout, highlighting DEGs specific to tissue type (Heart vs. Liver) at the ZT0 time point.

These plots allow for a quick assessment of which genes are significantly different in expression between tissues at each time point and help in identifying tissue-specific DEGs. Genes with significant fold changes (log2FoldChange > 1 or < -1 and adjusted p-value < 0.05) represent potential tissue-specific biomarkers.

6. Functional Enrichment Analysis (GO Analysis)

These analyses aim to identify the biological processes, molecular functions, and pathways that are significantly enriched among the differentially expressed genes (DEGs) in each comparison group. Bubble plots visually display the top enriched pathways, with bubble size indicating gene count, and color representing the adjusted p-value (statistical significance). Significant GO terms include "purine-containing compound metabolic process," "muscle cell differentiation," and "actin filament organization." These processes reflect tissue-specific functions: muscle-related processes align with the Heart's function, while metabolic processes relate to the Liver's role in managing metabolites, energy precursors, and detoxification.

Time-specific DEGs show enrichment in pathways such as "PI3K-Akt signaling pathway," "Apoptosis," and "Lipid metabolism," which may indicate circadian control over pathways related to cell survival, apoptosis, and metabolic activity. These pathways highlight that specific cellular processes, especially those connected to metabolism and cellular signaling, may be

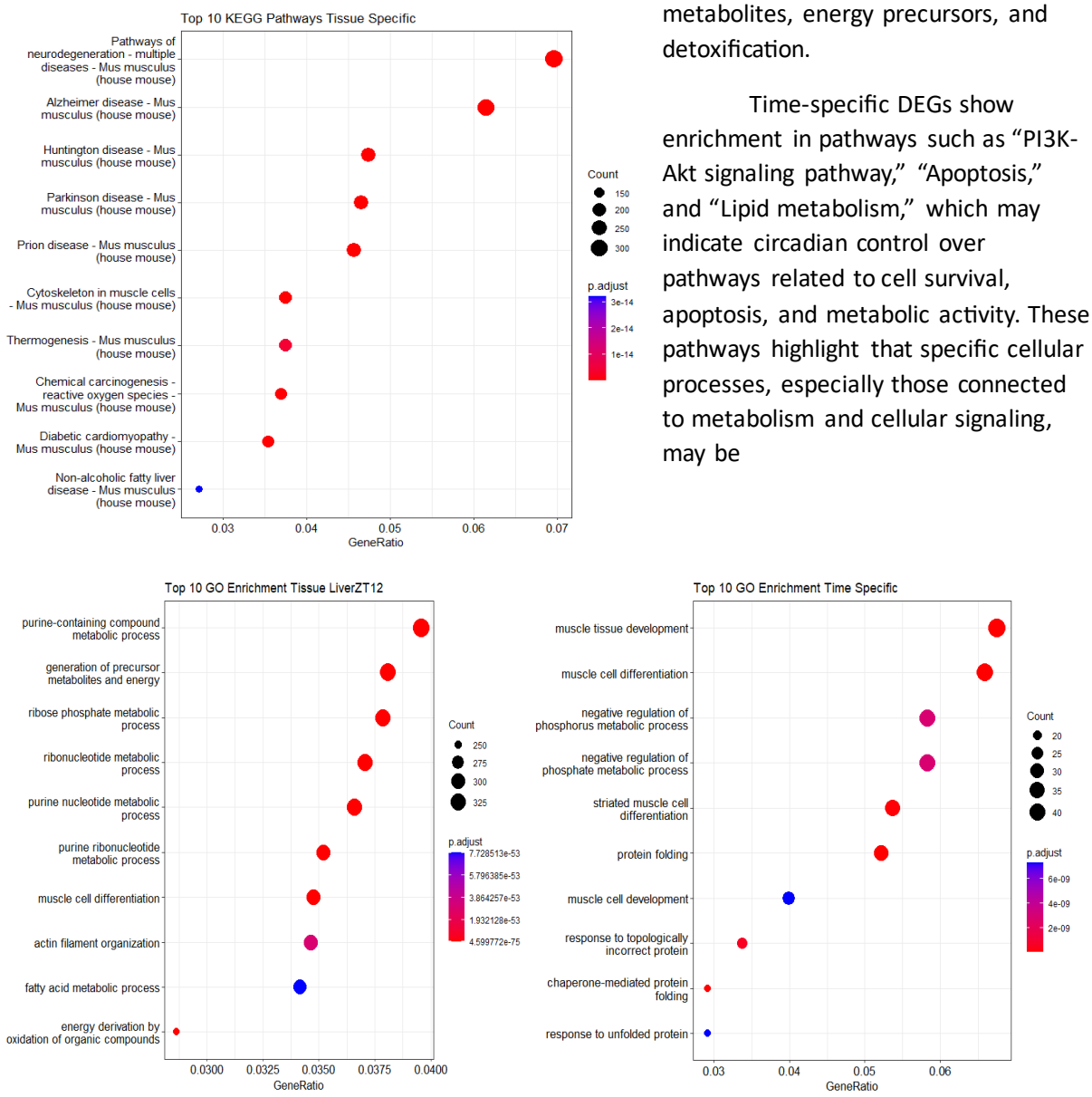


Fig. 9 Gene Ontology pathway enrichment bubble plots across various groups

References:

https://hbctraining.github.io/DGE_workshop/lessons/04_DGE_DESeq2_analysis.html [1]
https://www.sciencedirect.com/science/article/pii/S2001037024000424#sec0010 [2]