

**Widespread deregulation of gene expression in Pancreatic Ductal  
Adenocarcinoma is consistently present across all stages and  
facilitates early diagnosis.**

**2022**

**Supplementary Figures and Tables**

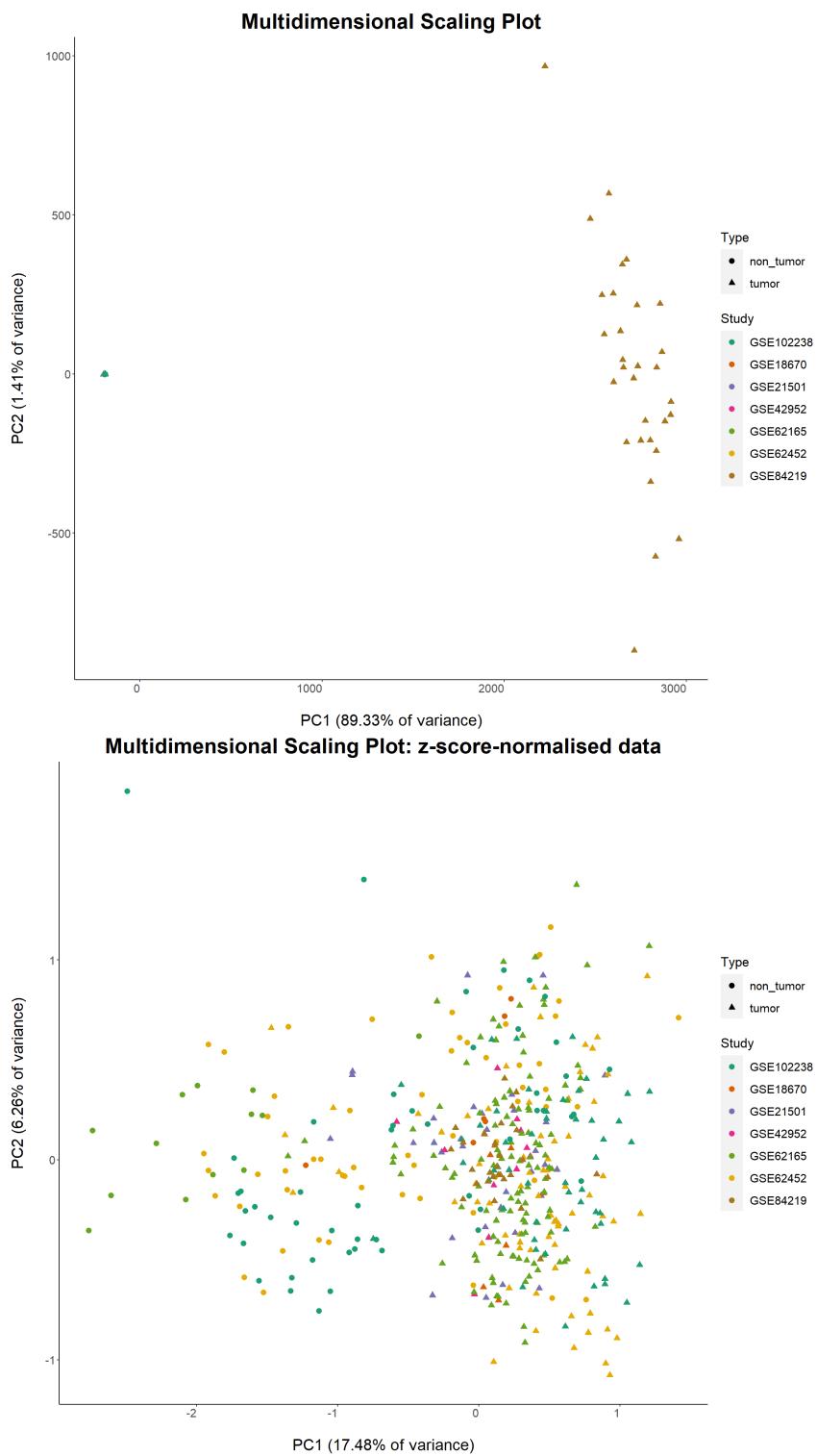
Table 1: Differential Gene Expression Analysis results: overlaps across stages

<b>General overlap</b>	<b>DEGs*</b>	<b>Stage 1</b>	<b>Stage 2</b>	<b>Stage 3</b>	<b>Stage 4</b>
<b>DEGs</b>	-	10891	16191	12934	11512
<b>Stage 1</b>	10891	-	10621	8597	8045
<b>Stage 2</b>	16191	10621	-	11437	10249
<b>Stage 3</b>	12934	8597	11437	-	9537
<b>Stage 4</b>	11512	8045	10249	9537	-
<b>Dichotomizers</b>	<b>DEGs</b>	<b>Stage 1</b>	<b>Stage 2</b>	<b>Stage 3</b>	<b>Stage 4</b>
<b>DEGs</b>	-	10891	16191	12934	11512
<b>Stage 1</b>	10891	-	273	2296	2847
<b>Stage 2</b>	16191	273	-	4756	5945
<b>Stage 3</b>	12934	2296	4756	-	3398
<b>Stage 4</b>	11512	2847	5945	3398	-

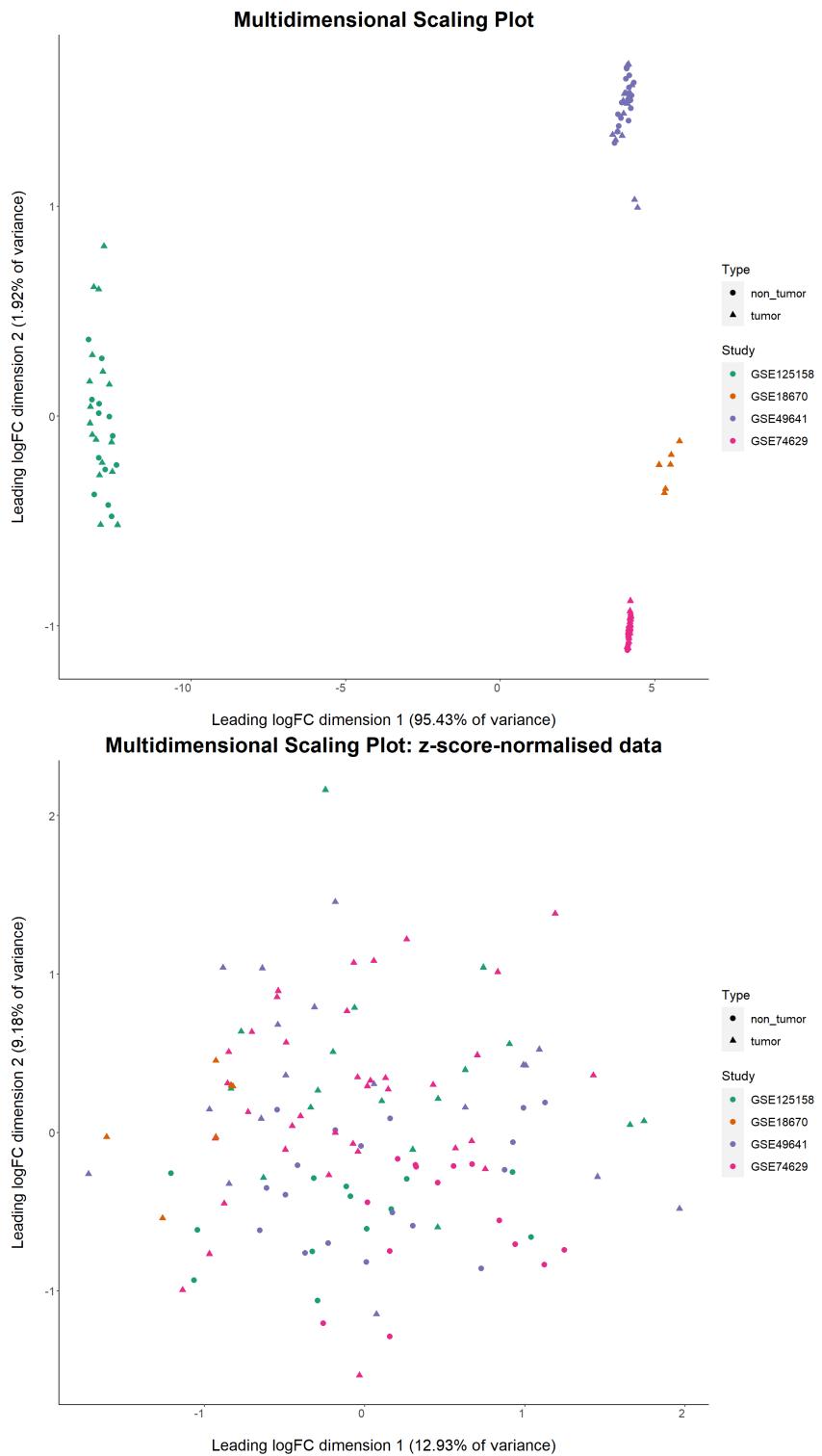
\*Differentially Expressed Genes

Table 2: Differentially expressed miRNAs and host genes

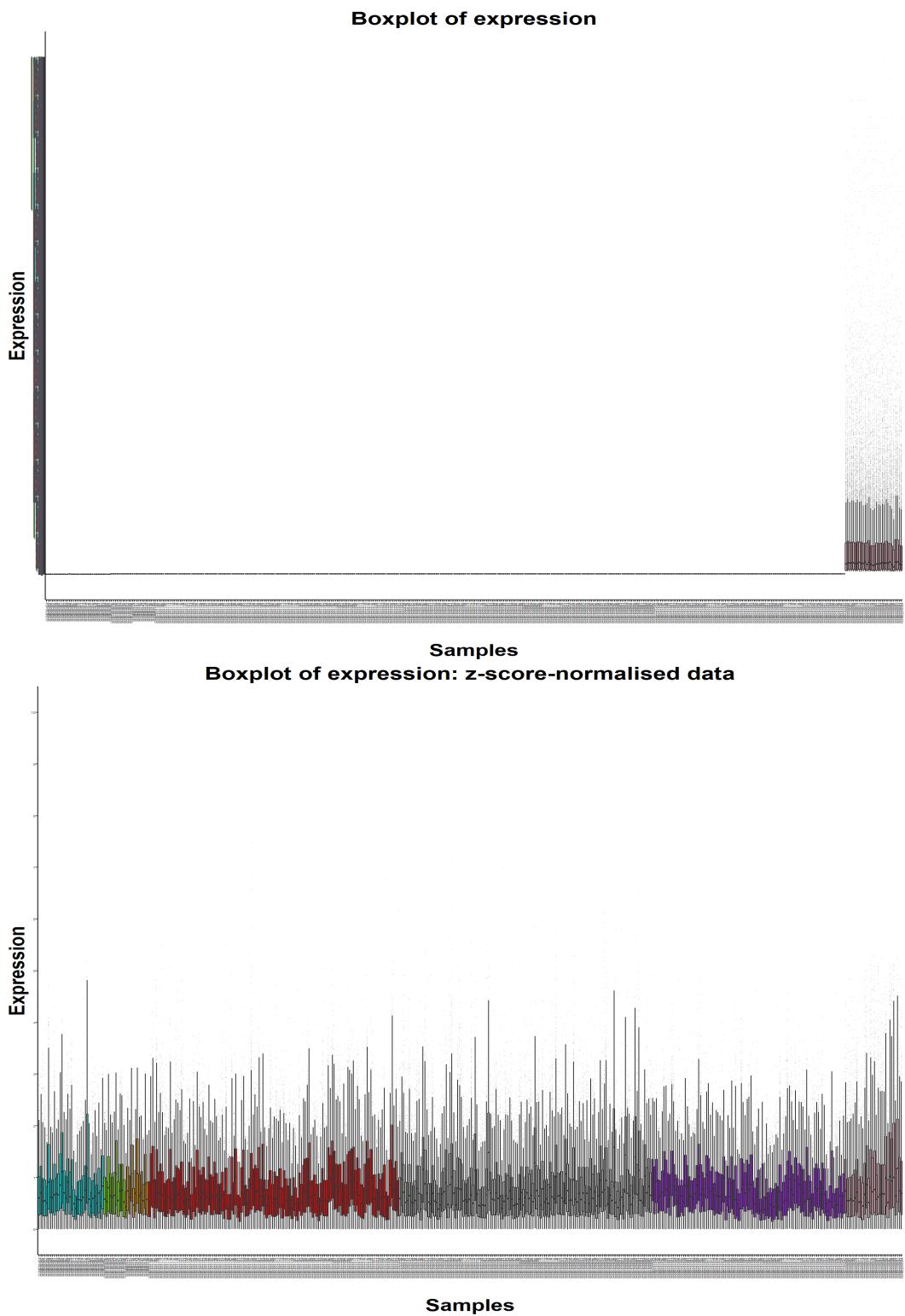
<b>Tumor</b>	<b>miRNA</b>	<b>miRNA plus host genes</b>
<b>Stage 1</b>	18	31
<b>Stage 2</b>	67	99
<b>Stage 3</b>	23	38
<b>Stage 4</b>	4	13
<b>Blood samples</b>	53	58



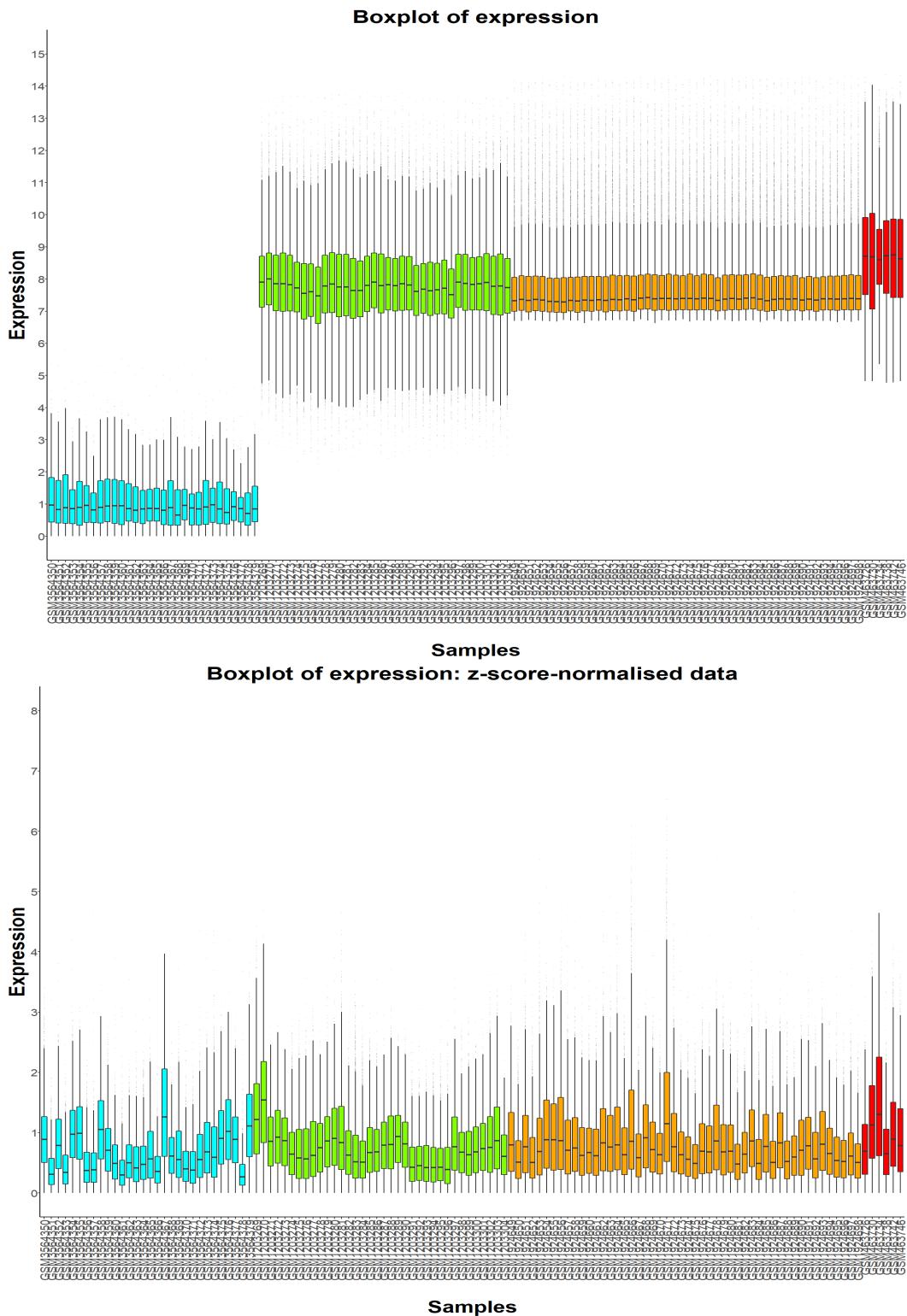
**Figure 1: Multidimensional scaling plots: tumor tissue samples.** A plot of the first two principal components from the original global expression matrix (top) and the matrix of z-score transformed values (bottom). After z-score normalisation samples are on a similar scale and comparable to one another even across studies, i.e. no batch effect is driving the observed variance. In the case of the global expression matrix, samples from the GSE84219 series are considerably different from the rest of the samples.



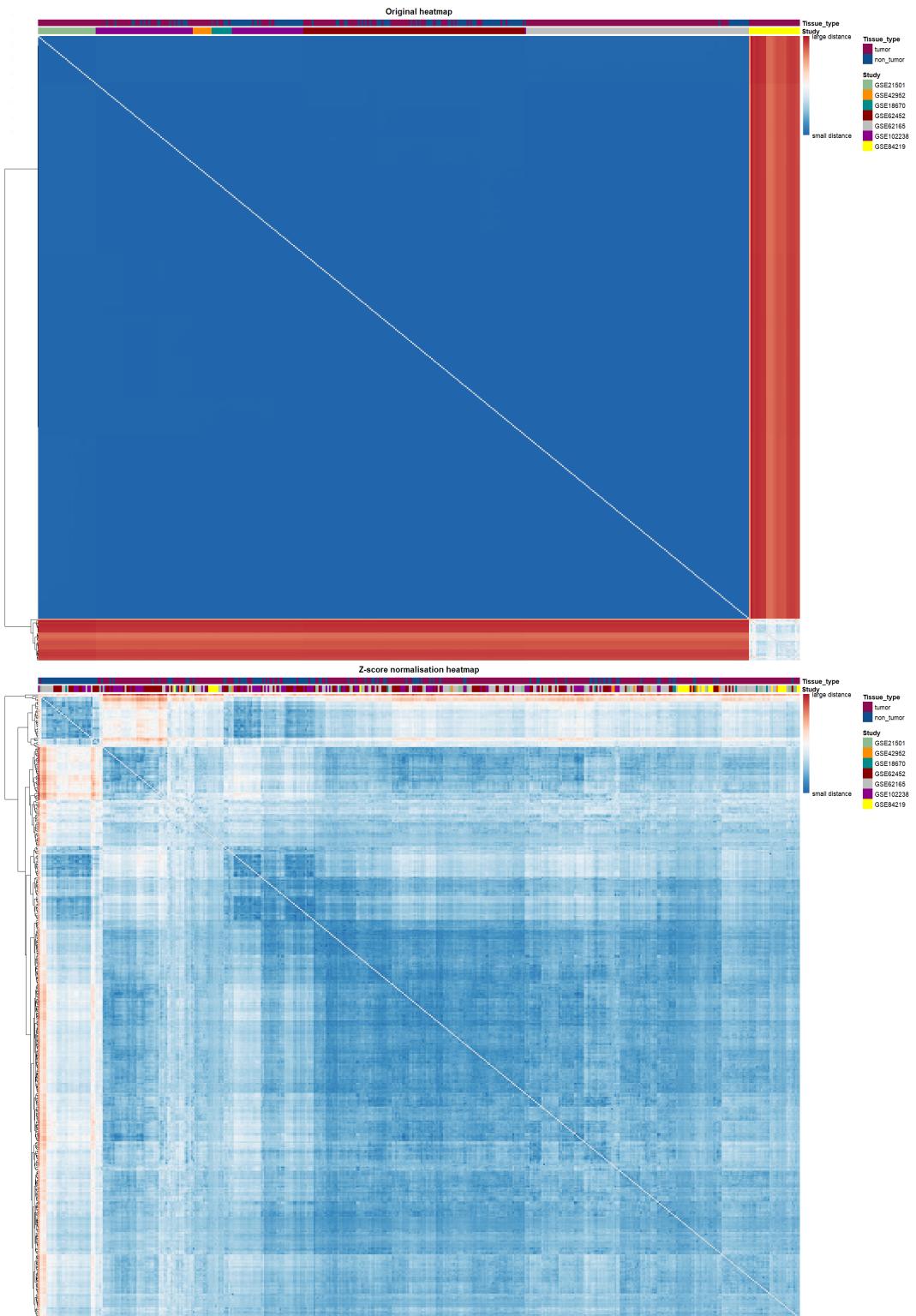
**Figure 2: Multidimensional scaling plots: blood samples.** A plot of the first two principal components from the original global expression matrix (top) and the matrix of z-score transformed values (bottom). After z-score normalisation samples are on a similar scale and comparable to one another even across studies, i.e. no batch effect is driving the observed variance. In the case of the global expression matrix, samples from the GSE125158 series are considerably different from the rest of the samples.



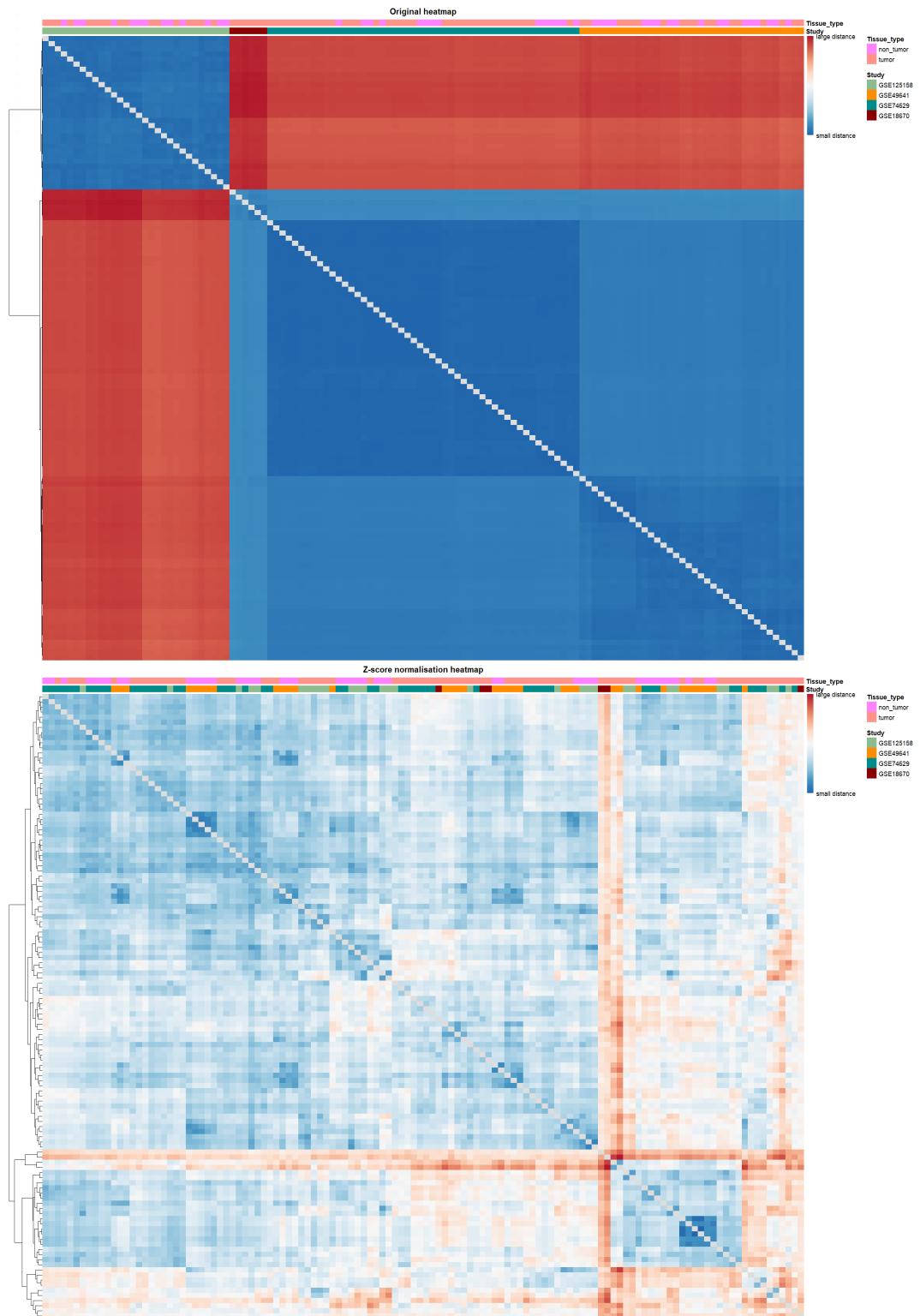
**Figure 3: Expression boxplots: tumor tissue samples.** Boxplots of expression for all tumor tissue samples, both tumor and normal samples. Un-normalized data are shown on the top plot and normalized data are shown on the bottom plot. Samples from the GSE84219 series are the main source of batch effects in un-normalized data. Upon normalization, samples are brought to a common scale.



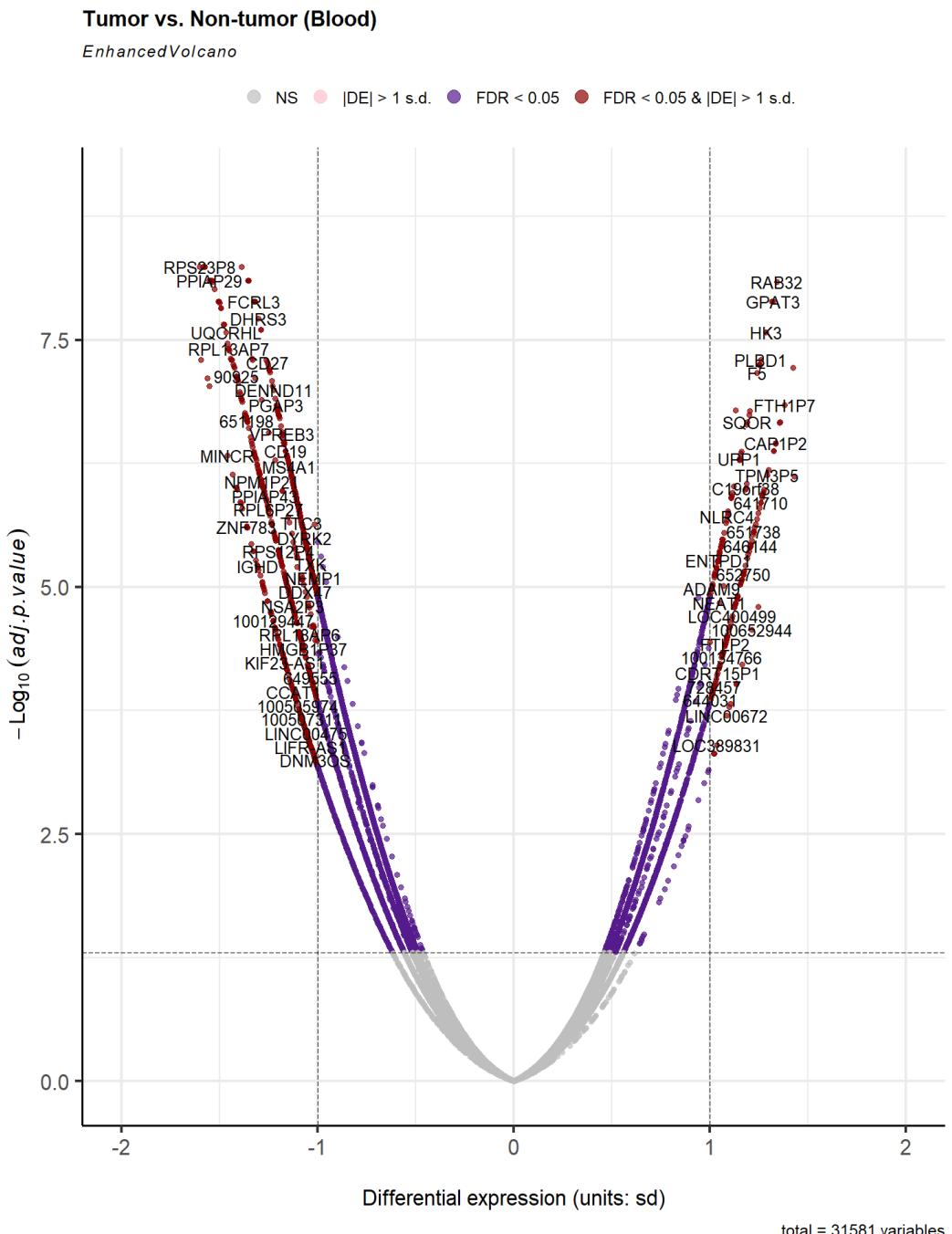
**Figure 4: Expression boxplots: blood samples.** Boxplots of expression for all blood samples, both tumor and normal samples. Un-normalized data are shown on the top plot and normalized data are shown on the bottom plot. Samples from the GSE125158 series are the main source of batch effects in un-normalized data. Upon normalization, samples are brought to a common scale.



**Figure 5: Sample distance heatmap: tumor tissue samples.** Heatmaps of high-dimensional L1 distance between tumor tissue samples, both tumor and normal. Un-normalized data are shown on the top plot and normalized data are shown on the bottom plot. Samples from the GSE84219 series are the main source of batch effects in un-normalized data. Upon normalization, distances between samples are reduced and brought to a tighter scale.



**Figure 6: Sample distance heatmap: blood samples.** Heatmaps of high-dimensional L1-distance between blood samples, both tumor and normal. Un-normalized data are shown on the top plot and normalized data are shown on the bottom plot. Samples from the GSE125158 series are the main source of batch effects in un-normalized data. Upon normalization, distances between samples are reduced and brought to a tighter scale.



**Figure 7: Volcano plot for DGEA results on blood samples.** The negative logarithm of the adjusted *p*-value of each gene is plotted on the y-axis and the expression change (responders/non-responders) of each gene (measured in standard deviations) is plotted on the x-axis. Vertical and horizontal dashed lines define expression change and *p*-value thresholds to indicate the most notable results. Expression change threshold was set to 1 and adjusted *p*-value threshold was set to 0.05. Depending on the partition of the plot each gene falls into, the colour of the dot changes as indicated by the legend at the top. Gray: NS (not significant), pink: passes expression change threshold, purple: passes *p*-value threshold, red: passes both thresholds.

Venn diagram of sig. diff. expressed miRNAs (+ host genes)

Venn diagram of sig. diff. expressed miRNAs (+ host genes)

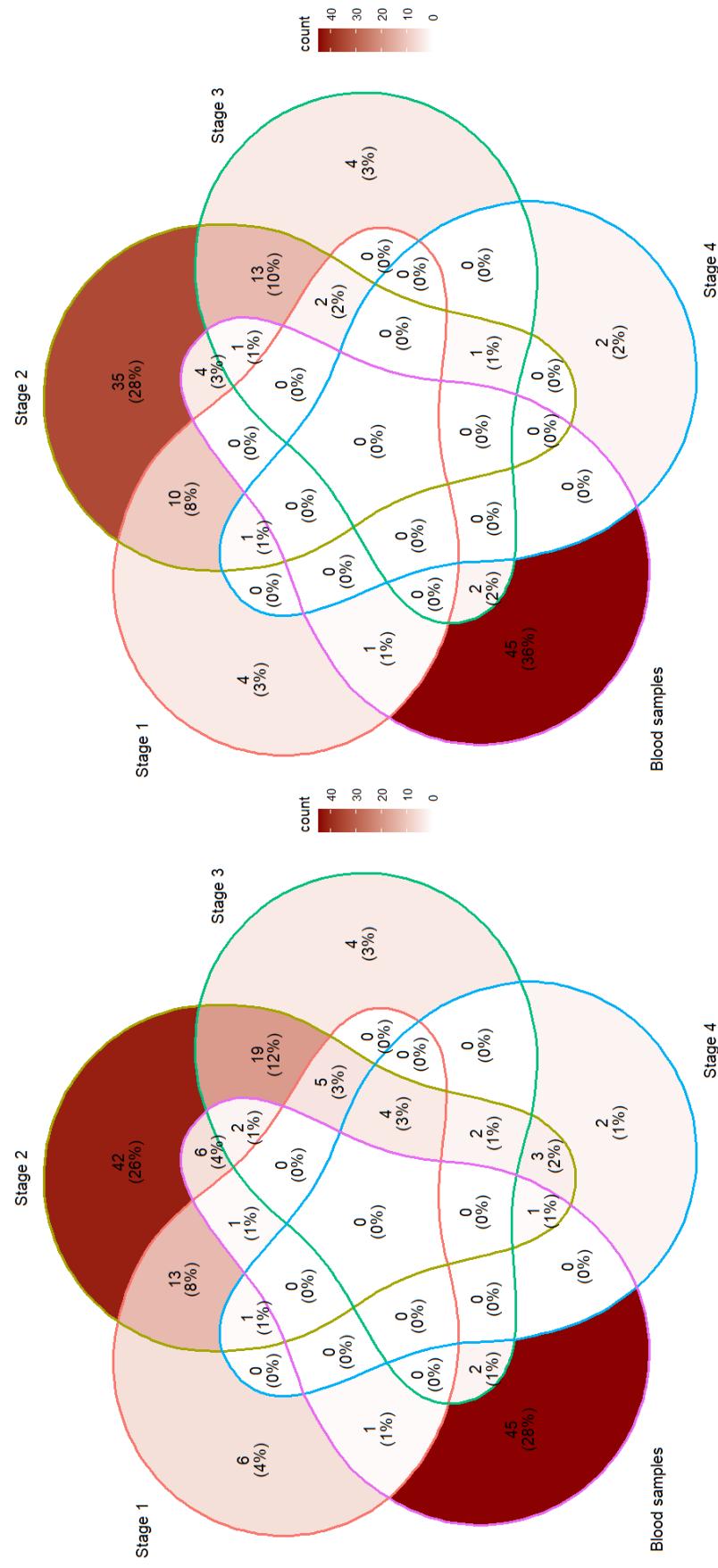
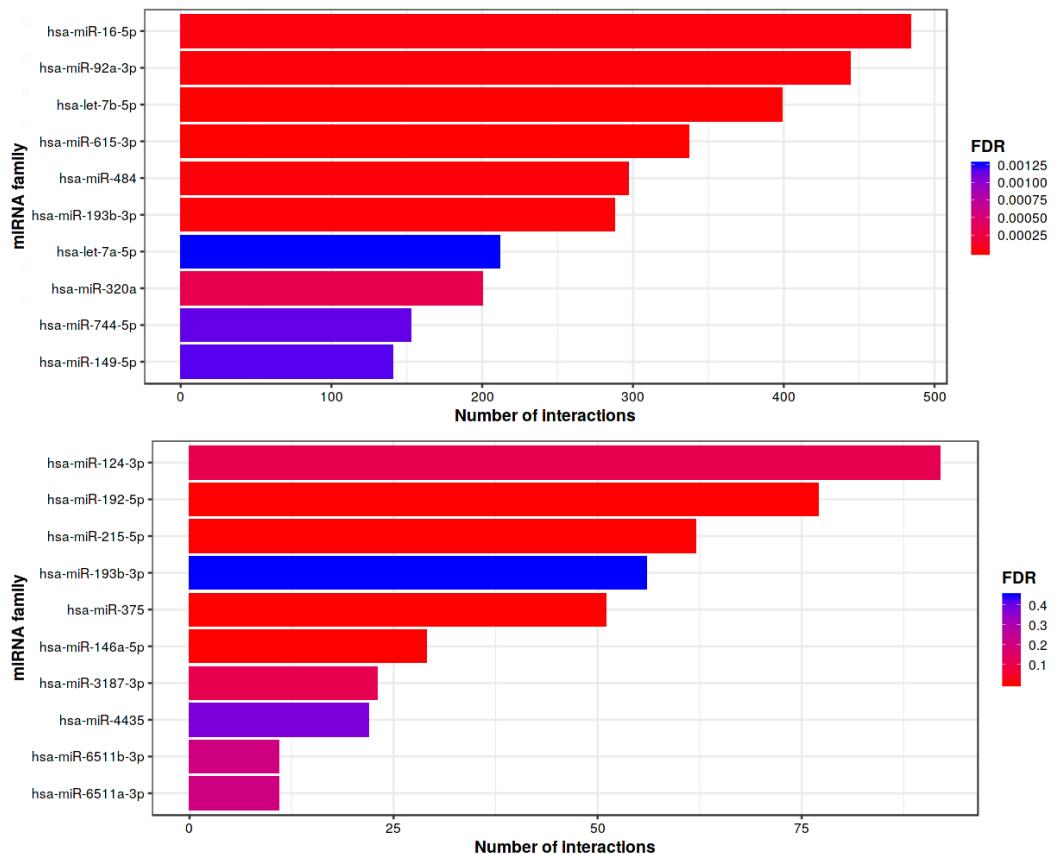
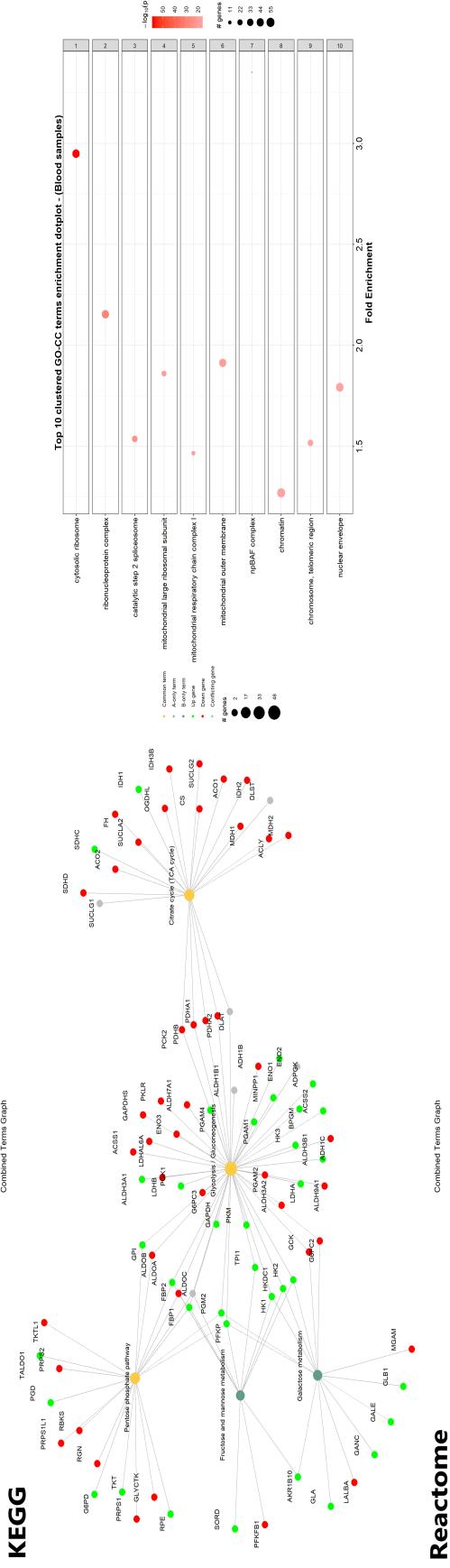


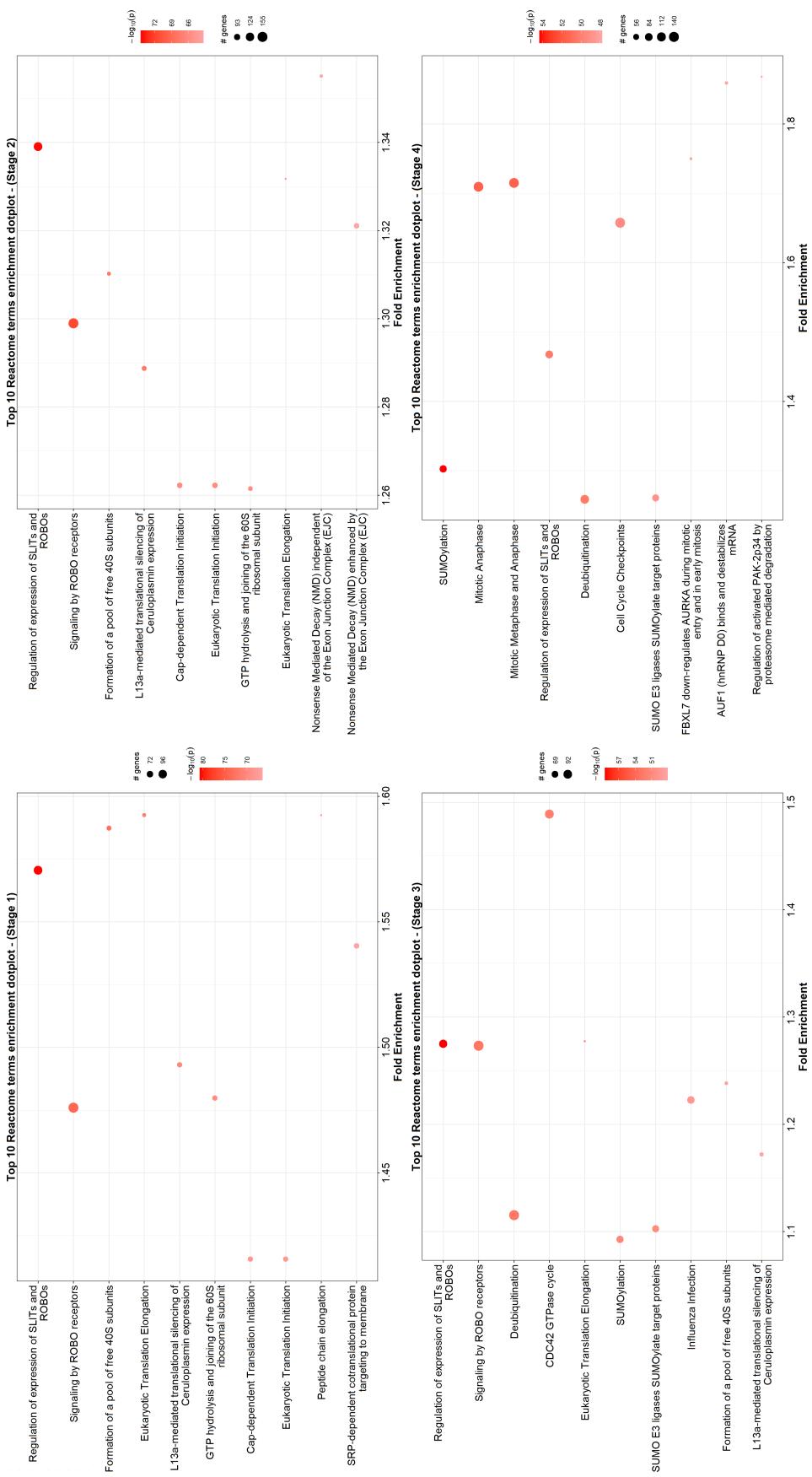
Figure 8: Venn diagram of differentially expressed miRNAs and host genes across tumor stages and types of samples. Differentially expressed miRNAs and host genes on the left; differentially expressed miRNAs excluding host genes on the right. Blood samples come with a relatively unique set of differentially expressed miRNAs with only one miRNA as a unique overlap with stage 1: miR-21. 35 unique differentially expressed miRNAs are found in stage 2.



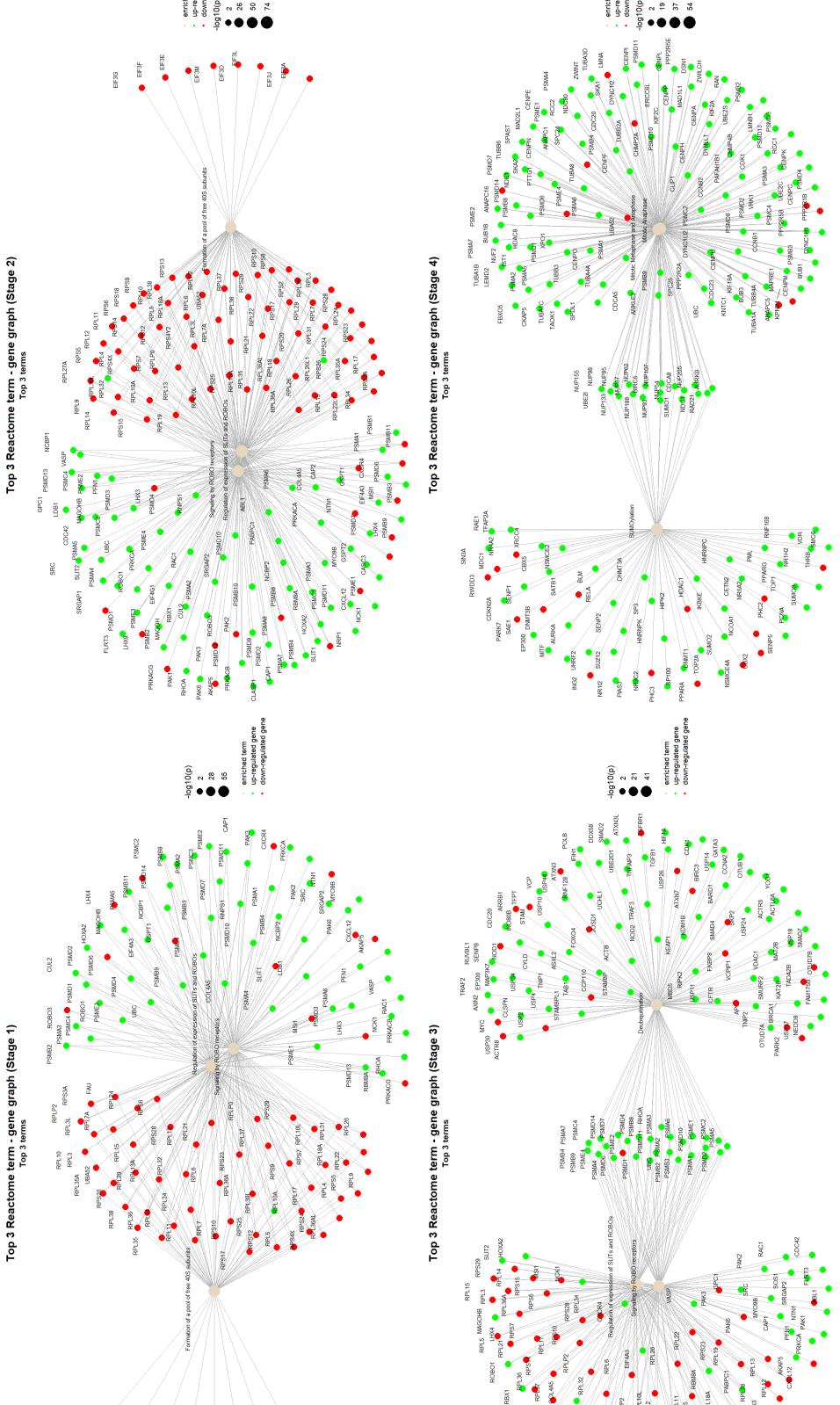
**Figure 9: Mienturnet® miRNA enrichment results for blood samples (top) and the 820-gene signature derived from DGEA.** No overlap exists between tissue stage-specific miRNA enrichment results and the results from blood samples. miR-375 and miR-192-5p are enriched when the 820-gene signature is used as input, similarly to the tissue stage-specific enrichment results.



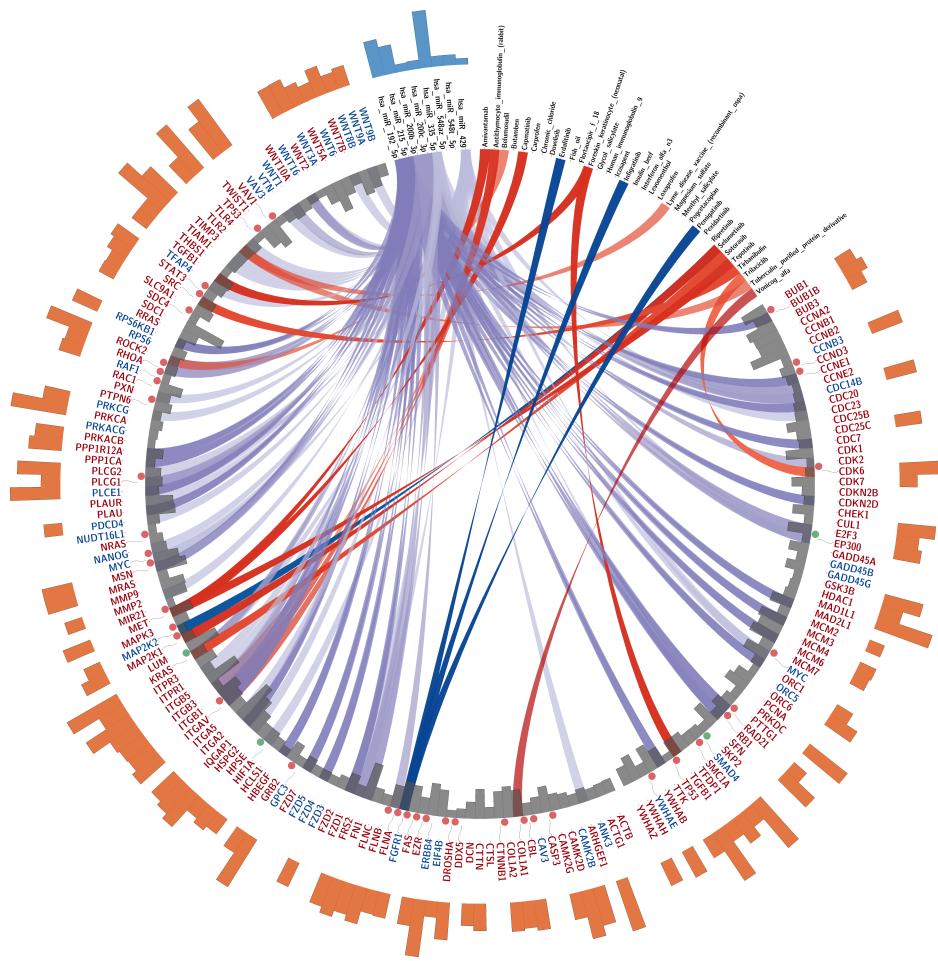
**Figure 10: Blood samples vs. Stage 1: KEGG pathways, Reactome processes and GO cellular components.** The top left plot shows the top five common KEGG terms between blood samples and stage 1 and how similarly deregulated these are. The bottom left plot provides the same comparison, but for Reactome terms. Yellow nodes are common enriched terms. Dark gray nodes are stage-1-only terms and lighter gray nodes are blood-sample-only terms. Similarly, red and green dots are common DEGs between the two groups and gray dots are conflicting genes. Top right plot is the clustered dotplot of the top 10 enriched GO cellular components in blood samples; bottom right plot is the clustered dotplot of the top 10 enriched GO cellular components in stage 1. The size of each dot is proportional to the number of DEGs that are in the corresponding gene set. Color of the dots is proportional to the negative logarithm of the adjusted  $p$ -value of each set's enrichment. The x-axis plots the enrichment score of each term.



**Figure 11: Top 10 Reactome term dotplot for stages 1-4.** The size of each dot is proportional to the number of DEGs that are in the corresponding gene set. Color of the dots is proportional to the negative logarithm of the adjusted  $p$ -value of each set's enrichment. The x-axis plots the enrichment score of each term.



**Figure 12: Top 3 Reactome term-gene graphs for stages 1-4.** Beige nodes are the enriched terms. Up-regulated genes are colored green and down-regulated genes are colored red. The size of the nodes is proportional to the negative logarithm of the adjusted enrichment  $p$ -value. Top right: stage 1; top left: stage 2; bottom right: stage 3; bottom left: stage 4.



**Figure 13: Circos plot of various drugs and miRNAs vs. genes from cell cycle and proteoglycans in cancer pathways (Stage 1).** Drug-gene and miRNA-gene interactions from the DrugBank and miRTarBase databases respectively.

**1) Labels.**

- 1. Drugs:** Drugs belonging to various categories are drawn with black labels at the top right arc.
- 2. Genes:** Genes from the KEGG Cell cycle pathway are drawn in purple at the bottom right arc (smaller gene track) and genes from the KEGG proteoglycans in cancer pathway are drawn in purple at the left side (larger gene track).
- 3. miRNA:** Black labels on a small arc at the top of the plot represent miRNAs that interact with these genes.

**2) Plots.**

- 1. Links:** Red ribbons represent pharmacogenetic links between drugs and up-regulated genes. Blue ribbons indicate pharmacogenetic links between drugs and down-regulated genes. Links for genes with lower DGEA adjusted  $p$ -values and larger absolute expression changes are plotted on top of others and given darker colors. miRNA-gene interactions are drawn in purple ribbons. Darker colors and overlaid ribbons indicate lower DGEA adjusted  $p$ -values and higher absolute expression shifts in genes, and lower miRNA enrichment adjusted  $p$ -values for miRNAs.
- 2. Inner circle histogram:** The inwards-oriented histogram (gray) illustrates the DGEA adjusted  $p$ -value for each gene.
- 3. Cancer drivers glyph track:** Colored dots indicate global cancer drivers (orange) and pancreas-specific cancer drivers (green).
- 4. Connector track:** Lines which connect gene labels to the corresponding glyphs.
- 5. Mutation frequency:** The orange histogram in the outer layer gives an estimate of the COSMIC mutation frequency of each gene in pancreatic cancer samples.
- 6. miRNA degree histogram:** A histogram that displays the number of genes (from our full list of genes) each miRNA interacts with.

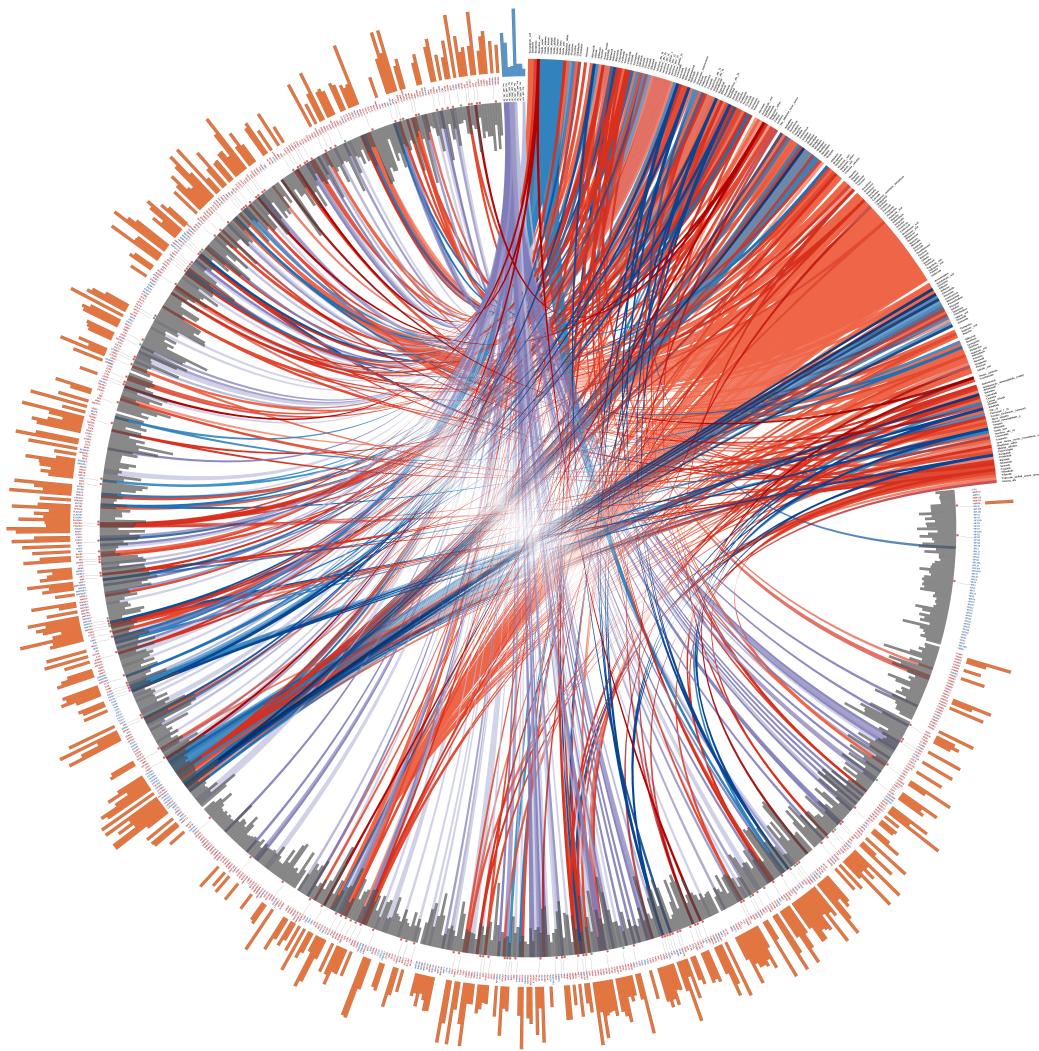


Figure 14: **Circos plot of all drugs and KEGG pathways (top 10) for Stage 1.** Elements are drawn as described in the previous figure. This figure can be used to narrow down drug categories and pathways that one is interested to investigate further