RNA-seq Analysis Report: Output File Summary

Shuye Pu

2025-07-08

Introduction

This report provides a detailed summary of the output files generated by the `RNAseq_enrichment_analysis.R` script.

The report is structured as follows: 1. Experimental design and pairwise comparisons. 2. Multi-factor differential expression analysis to model time and dose effects. 3. Standard single-factor (pairwise) differential expression analysis.

The outputs for each stage are described below.

1. Experimental Design

This section details the experimental design of the RNA-seq experiment. The experimental design is saved in the `design.tsv` file.

Table 1: Sample Design Overview

sample	condition	time	dose
NIR_1	NIR	0h	0Gy
NIR_2	NIR	0h	0Gy
NIR_3	NIR	0h	0Gy
2Gy24h_1	2Gy24h	24h	2Gy
2Gy24h_2	2Gy24h	24h	2Gy
2Gy24h_3	2Gy24h	24h	2Gy
2Gy6d_1	2Gy6d	6d	2Gy
2Gy6d_2	2Gy6d	6d	2Gy
2Gy6d_3	2Gy6d	6d	2Gy
10Gy24h_1	10Gy24h	24h	10Gy
10Gy24h_2	10Gy24h	24h	10Gy
10Gy24h_3	10Gy24h	24h	10Gy
10Gy6d_1	10Gy6d	6d	10Gy
10Gy6d_2	10Gy6d	6d	10Gy
10Gy6d_3	10Gy6d	6d	10Gy

The specific pairwise comparisons performed are listed in the table below.

Table 2: Pairwise Comparisons

Comparison	Condition 1 (Control/Base)	Condition 2 (Test)
2Gy24h_vs_NIR	NIR	2Gy24h
2Gy6d_vs_NIR	NIR	2Gy6d
10Gy24h_vs_NIR	NIR	10Gy24h
10Gy6d_vs_NIR	NIR	10Gy6d
2Gy6d_vs_2Gy24h	2Gy24h	2Gy6d
10Gy6d_vs_10Gy24h	10Gy24h	10Gy6d
10Gy24h_vs_2Gy24h	2Gy24h	10Gy24h
10Gy6d_vs_2Gy6d	2Gy6d	10Gy6d

2. Differential Expression and Enrichment Analysis Outputs

This section details the outputs from both the multi-factor and single-factor differential expression analyses with DESeq2.

2.1 Multi-Factor Analysis (Time & Dose)

This analysis models the effects of `time`, `dose`, and their `interaction`. Outputs are saved in the `multi_factor_analysis` directory.

Main Results Tables

- `time_effect_DESeq2_results.csv`: Full DESeq2 results for the main effect of time (6d vs. 24h).
- `dose_effect_DESeq2_results.csv`: Full DESeq2 results for the main effect of dose (10 Gy vs. 2 Gy).
- `time_dose_interaction_DESeq2_results.csv`: Full DESeq2 results for the interaction term.

Visualization and Plots

- **Figure 1:** `PCA_time_dose_effects.pdf`: A Principal Component Analysis (PCA) plot to visualize sample clustering based on experimental factors.

- **Figure 2:** `heatmap_top5000_variable_genes.pdf`: A heatmap of the top 5000 most variable genes across all samples.

- **Figures 3-5: Volcano Plots**: Volcano plots illustrating differential expression for each main effect and the interaction.

- **Figures 6-8: MA Plots**: MA plots showing log2 fold change versus mean normalized counts for each effect.

- **Figures 9-11: Gene Type Annotation Plots**: Bar plots showing the distribution of gene biotypes.

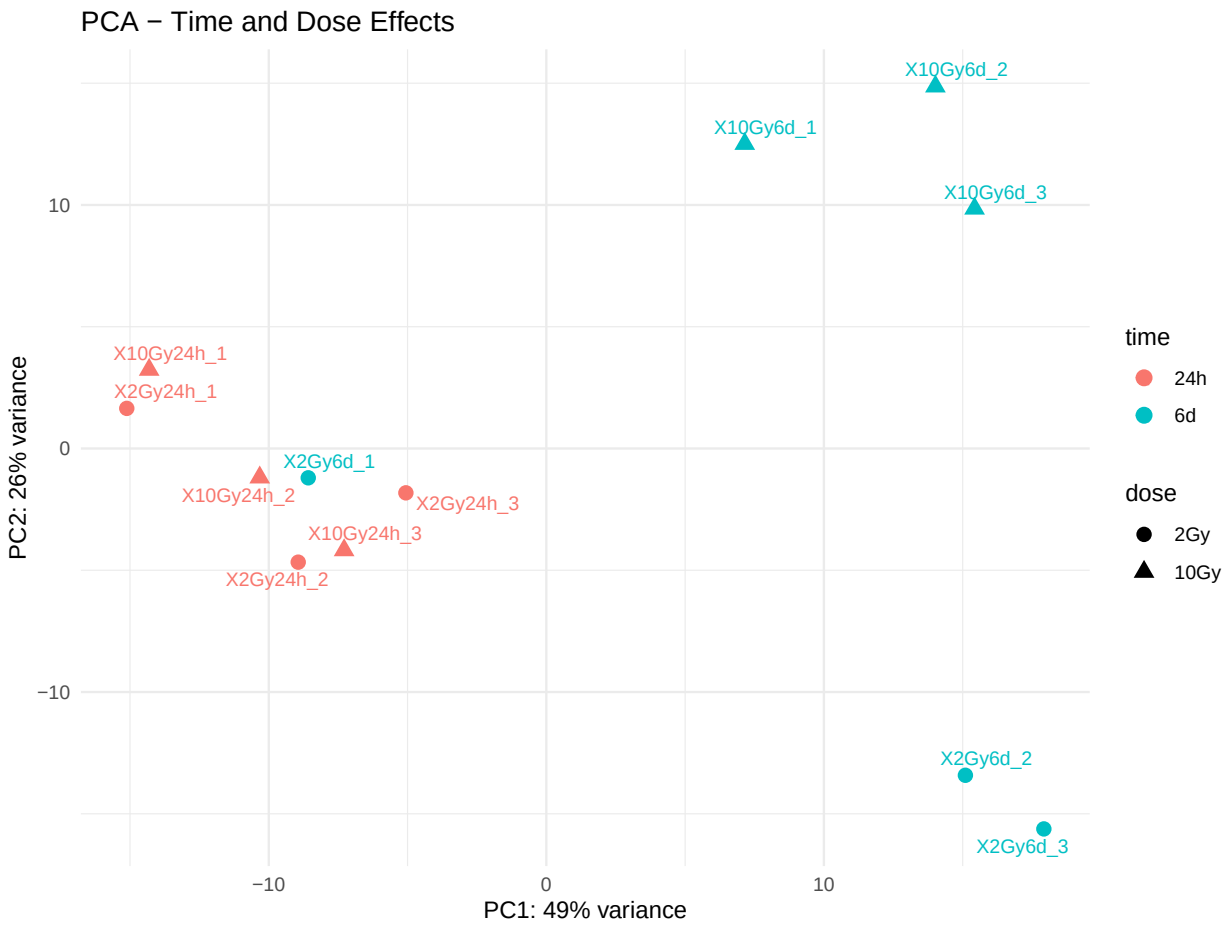

Figure 1: PCA plot of samples based on experimental factors.

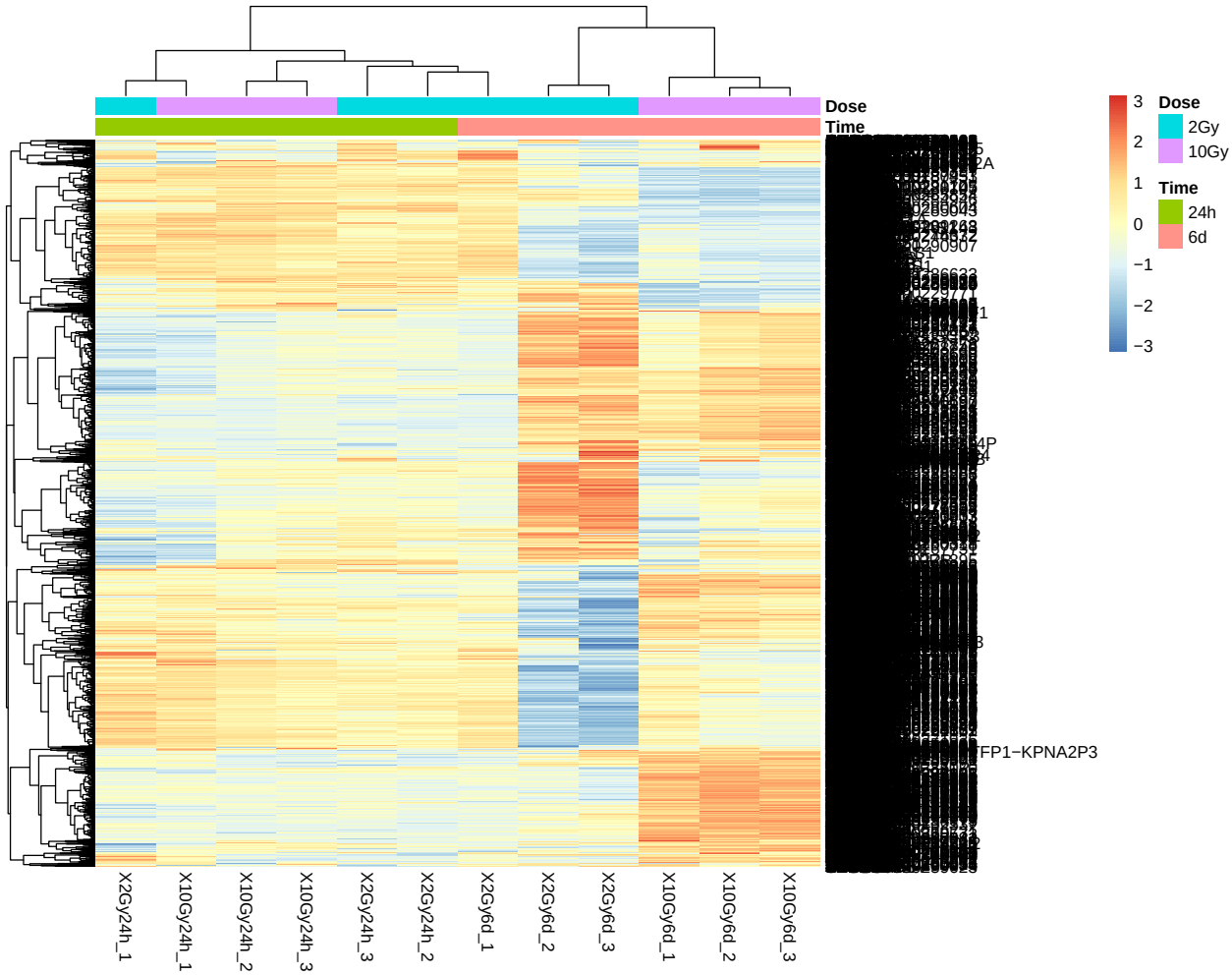

Figure 2: Heatmap of the top 5000 most variable genes.

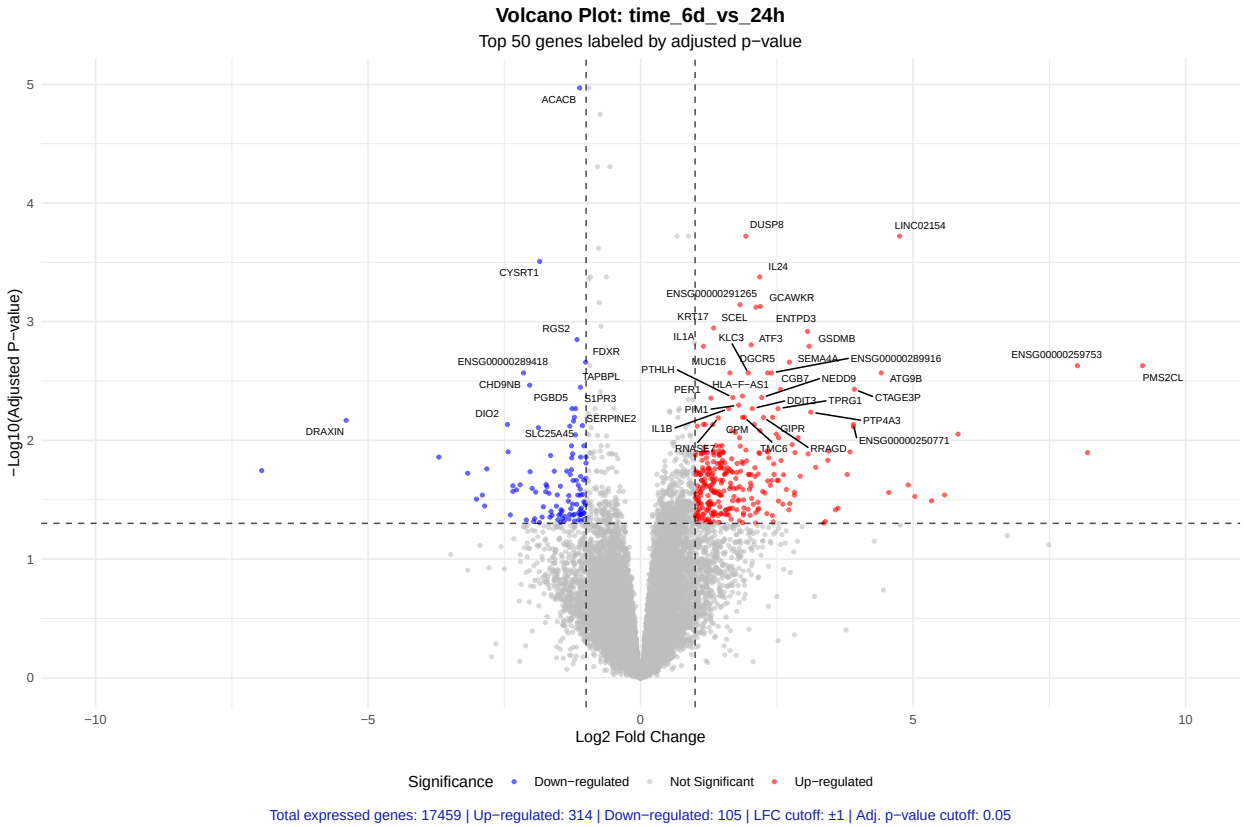

Figure 3: Volcano plot for the main effect of time (6d vs. 24h).

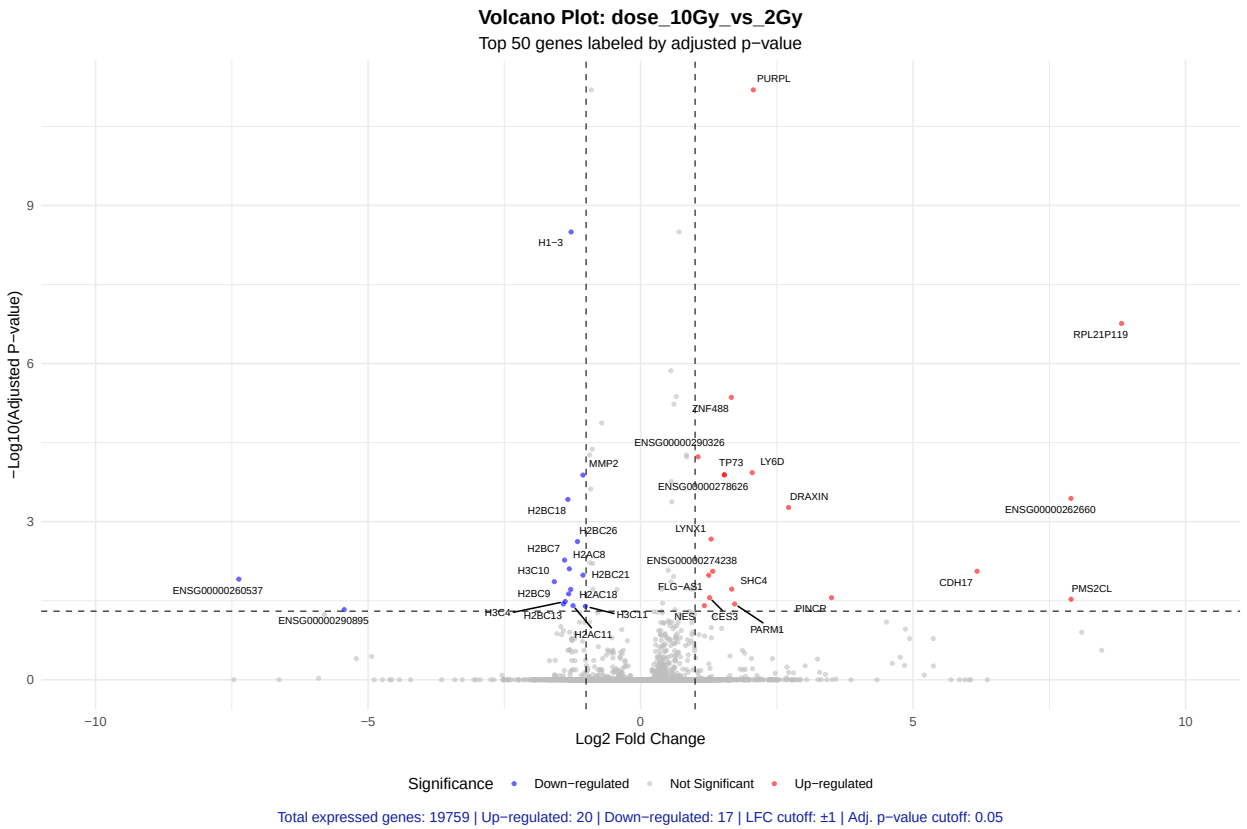

Figure 4: Volcano plot for the main effect of dose (10Gy vs. 2Gy).

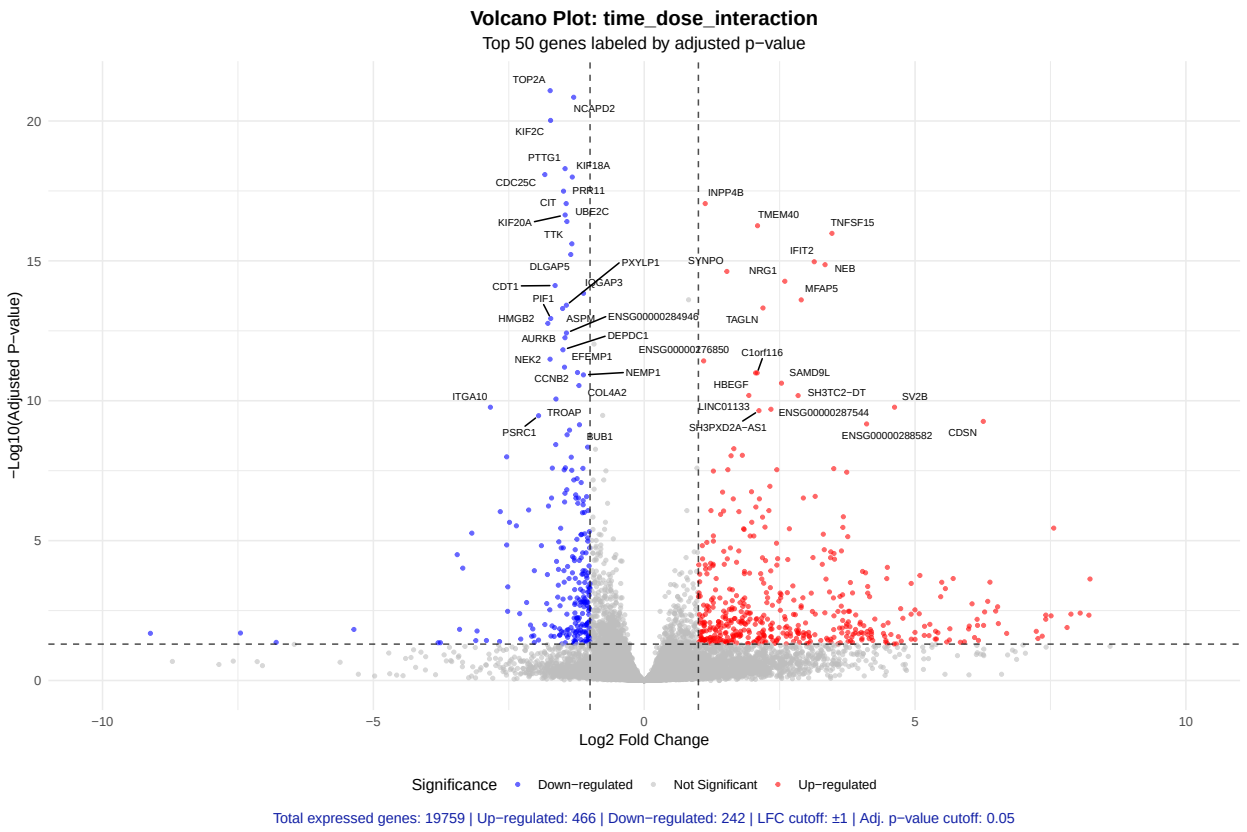

Figure 5: Volcano plot for the time-dose interaction effect.

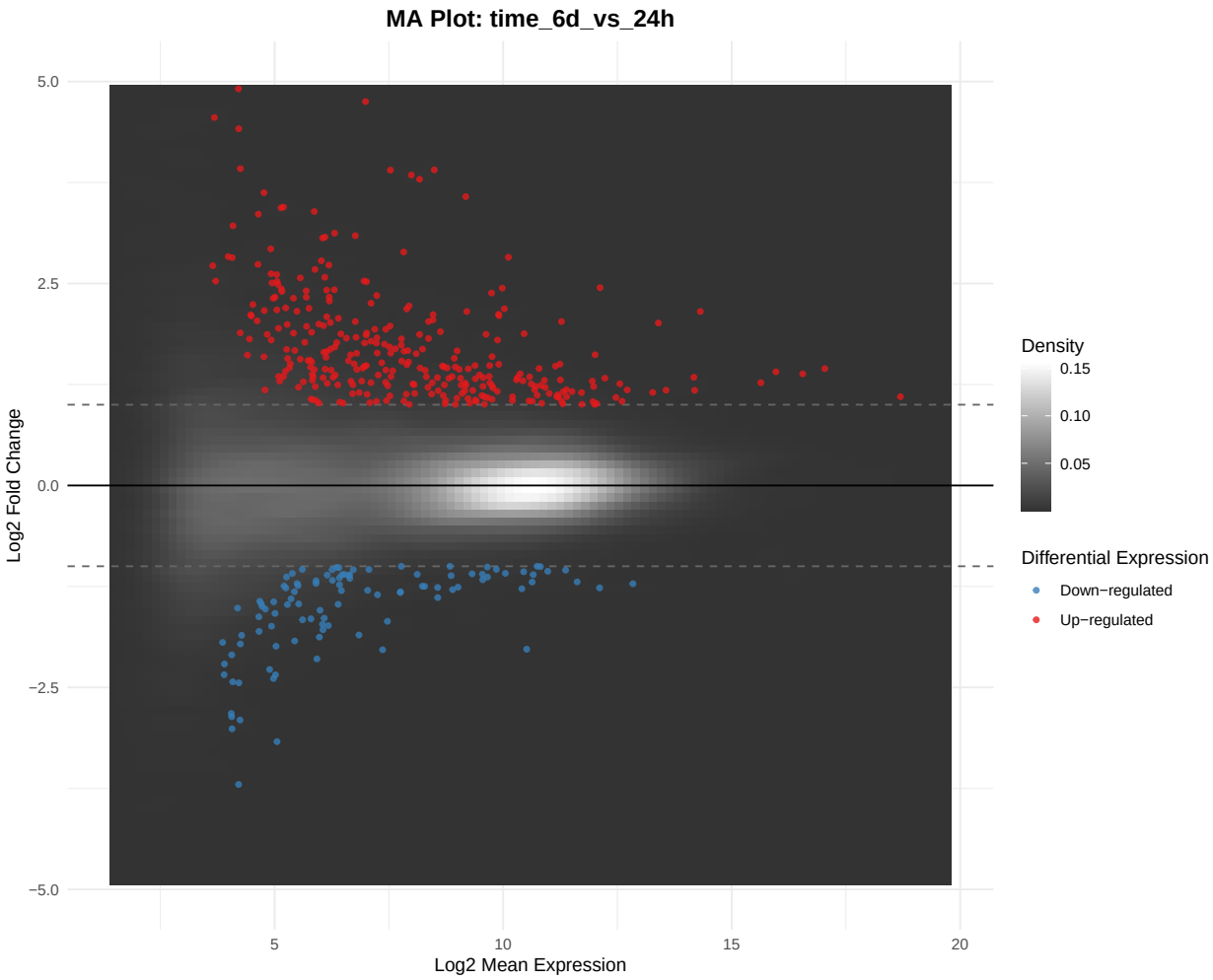

Figure 6: MA plot for the main effect of time.

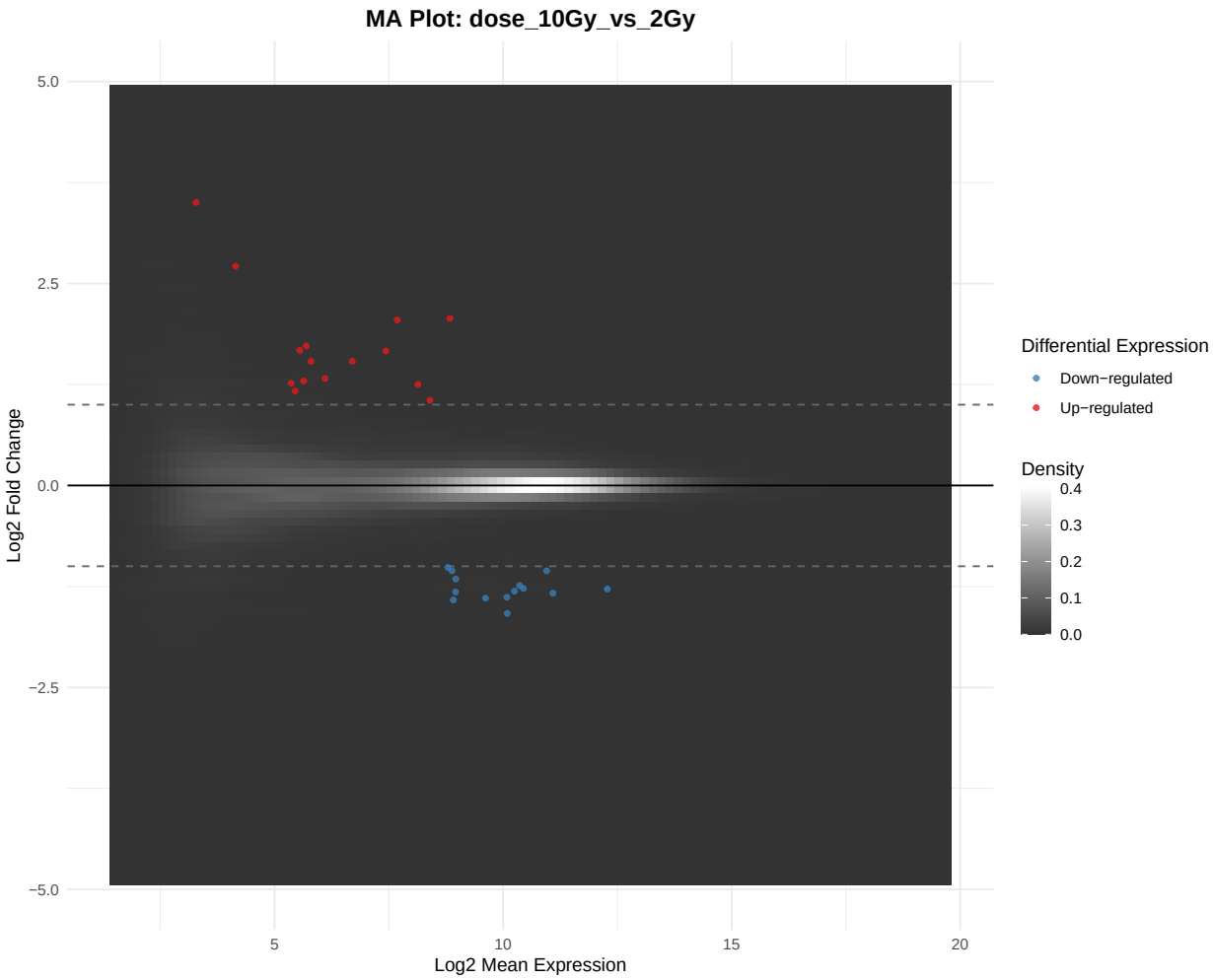

Figure 7: MA plot for the main effect of dose.

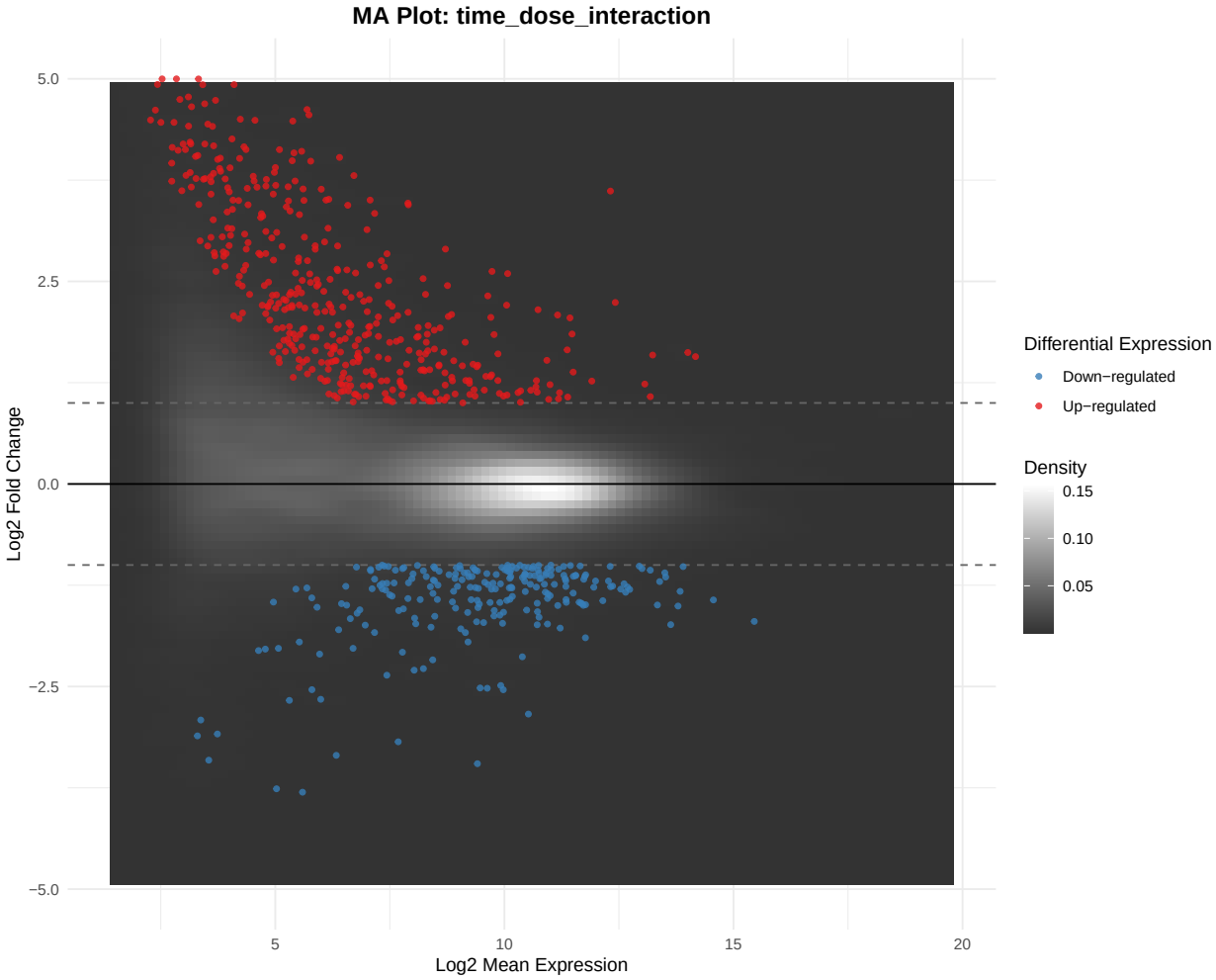

Figure 8: MA plot for the time-dose interaction effect.

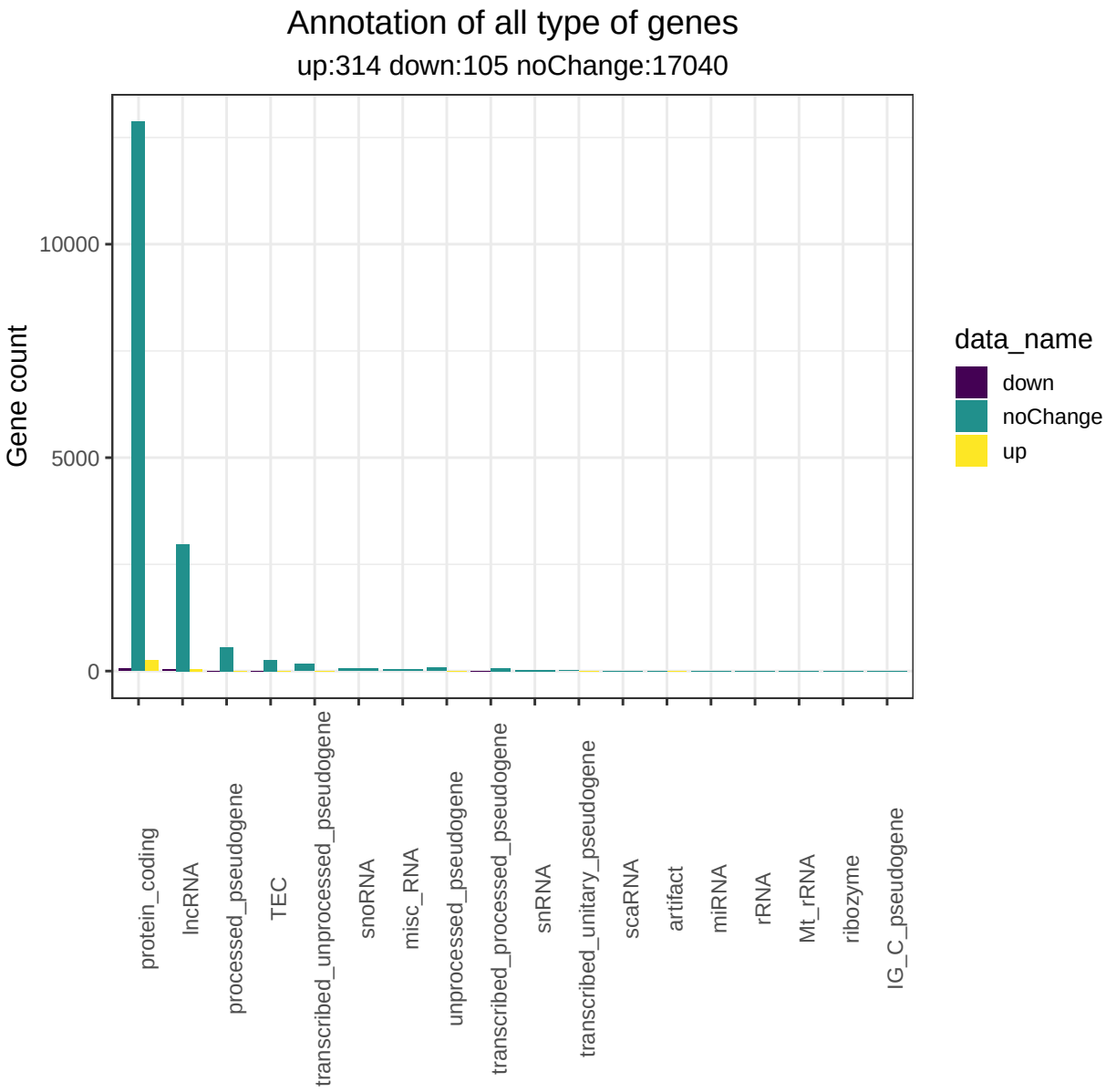

Figure 9: Gene biotype distribution for the time effect.

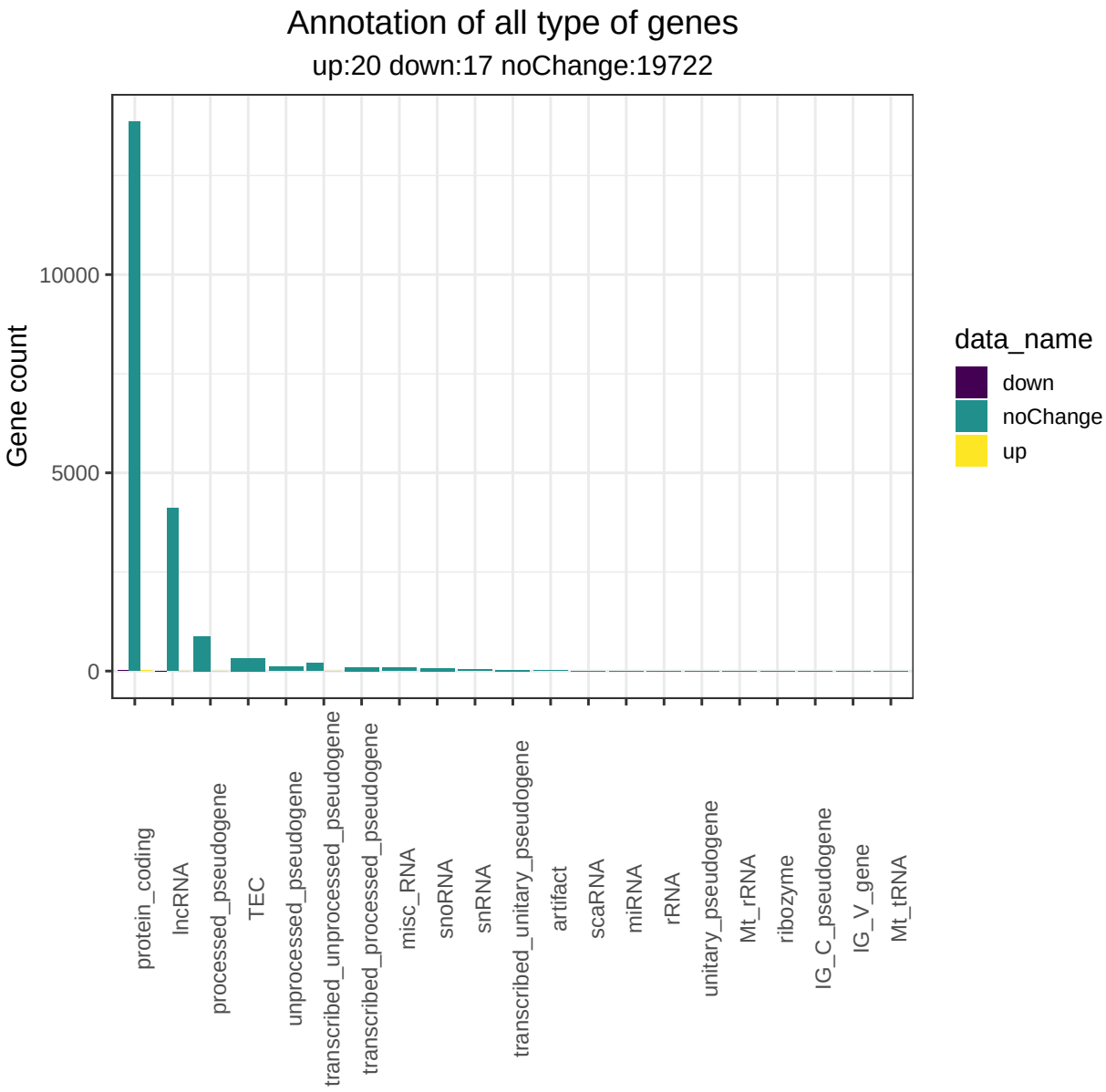

Figure 10: Gene biotype distribution for the dose effect.

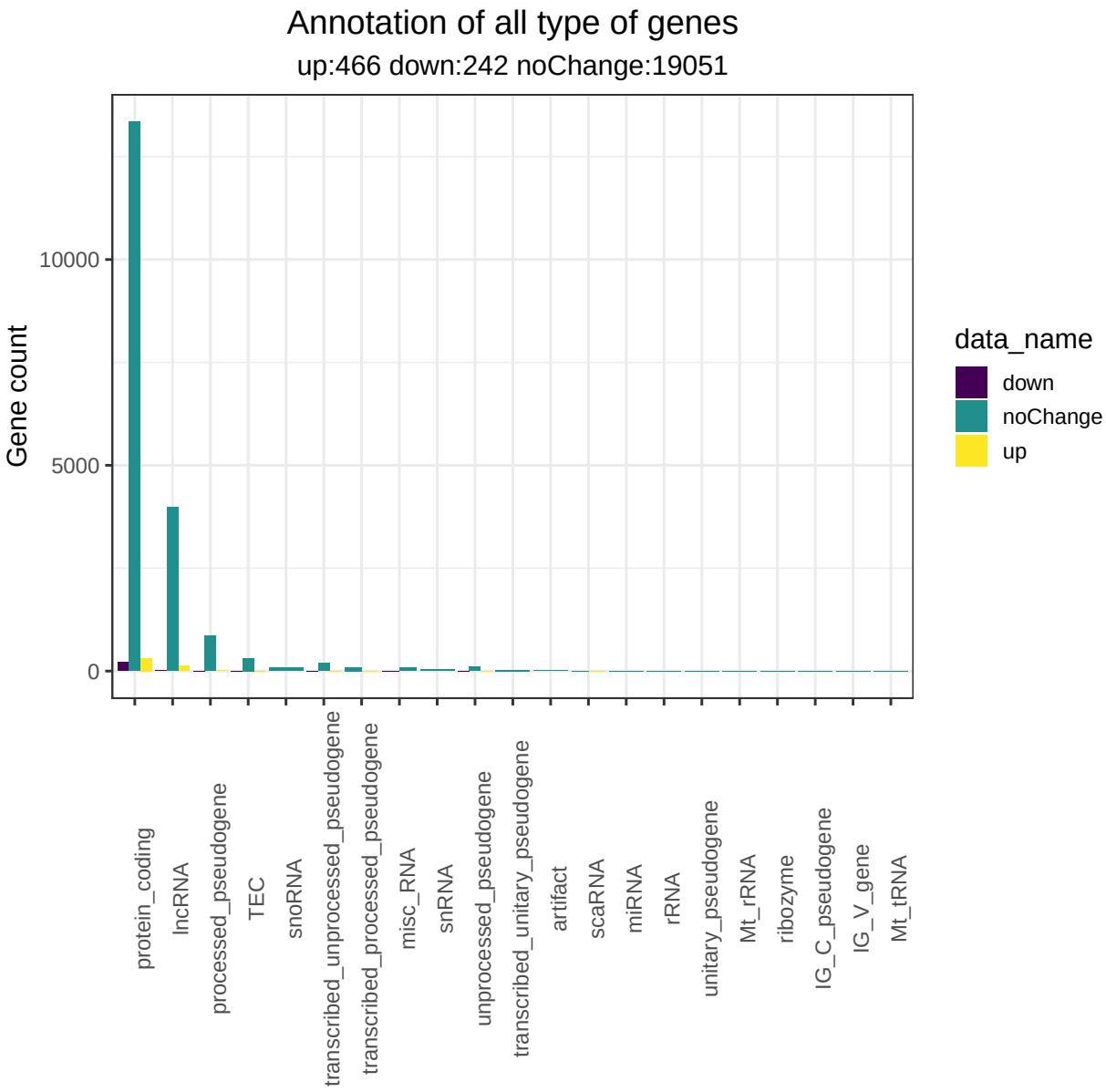

Figure 11: Gene biotype distribution for the interaction effect.

Enrichment Analysis (per effect)

For each effect (`Time_Effect_6d_vs_24h`, `Dose_Effect_10Gy_vs_2Gy`, `Time_Dose_Interaction`), the following enrichment files are generated:

GSEA (Gene Set Enrichment Analysis)

- **Figures 12-14:** `[effect]_BP_gsea_analysis.pdf`: Summary plots for GO:BP GSEA results. Details for each BP term are available in the `[effect]_BP_gsea_analysis_plot.pdf` files not shown in the report, please see the accompanying zip file for the full results.

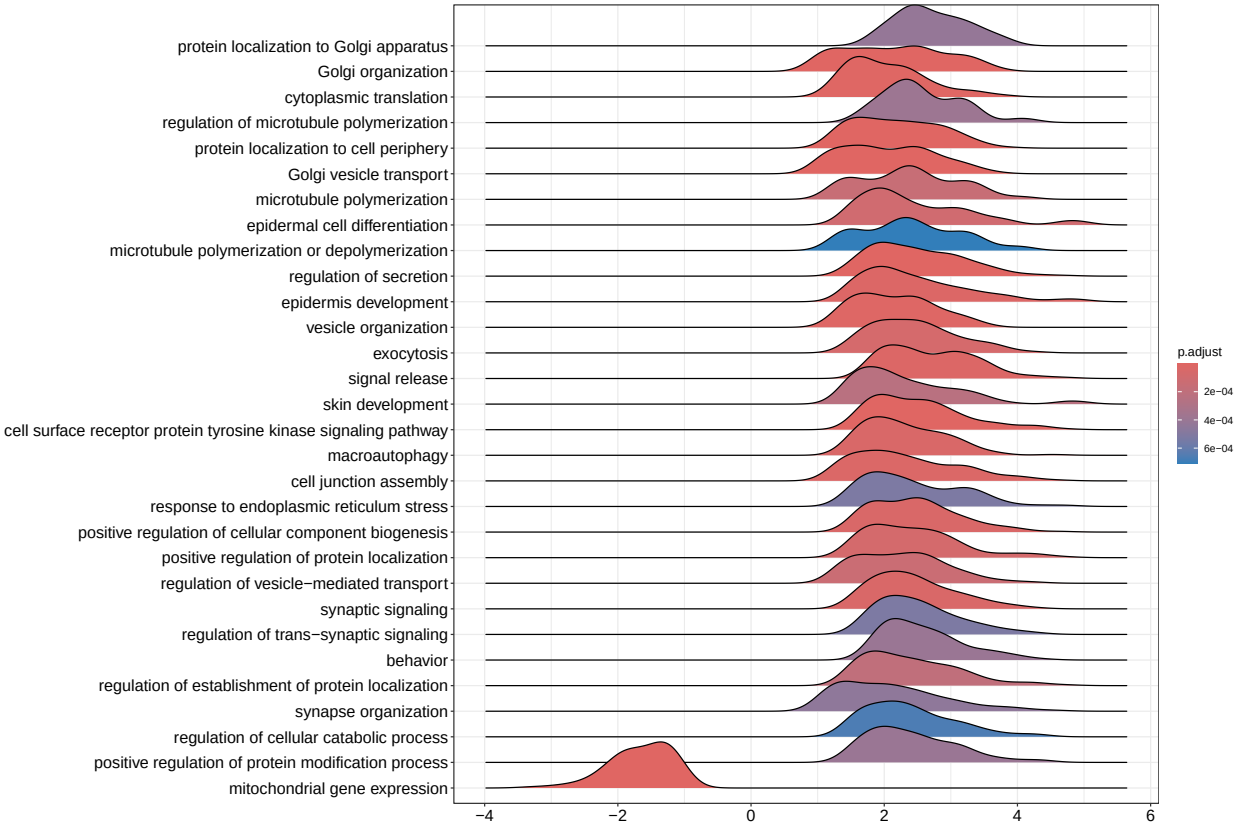

Figure 12: GSEA GO:BP summary for the time effect.

- **Figures 15-17:** `[effect]_KEGG_gsea_analysis.pdf`: Summary plots for KEGG GSEA results. Details for each KEGG term are available in the `[effect]_KEGG_gsea_analysis_plot.pdf` files not shown in the report, please see the accompanying zip file for the full results.

- **Tables**:
  - `[effect]_BP_gsea_analysis.tab`: Table of GSEA results for GO Biological Process terms.
  - `[effect]_KEGG_gsea_analysis.tab`: Table of GSEA results for KEGG pathways.

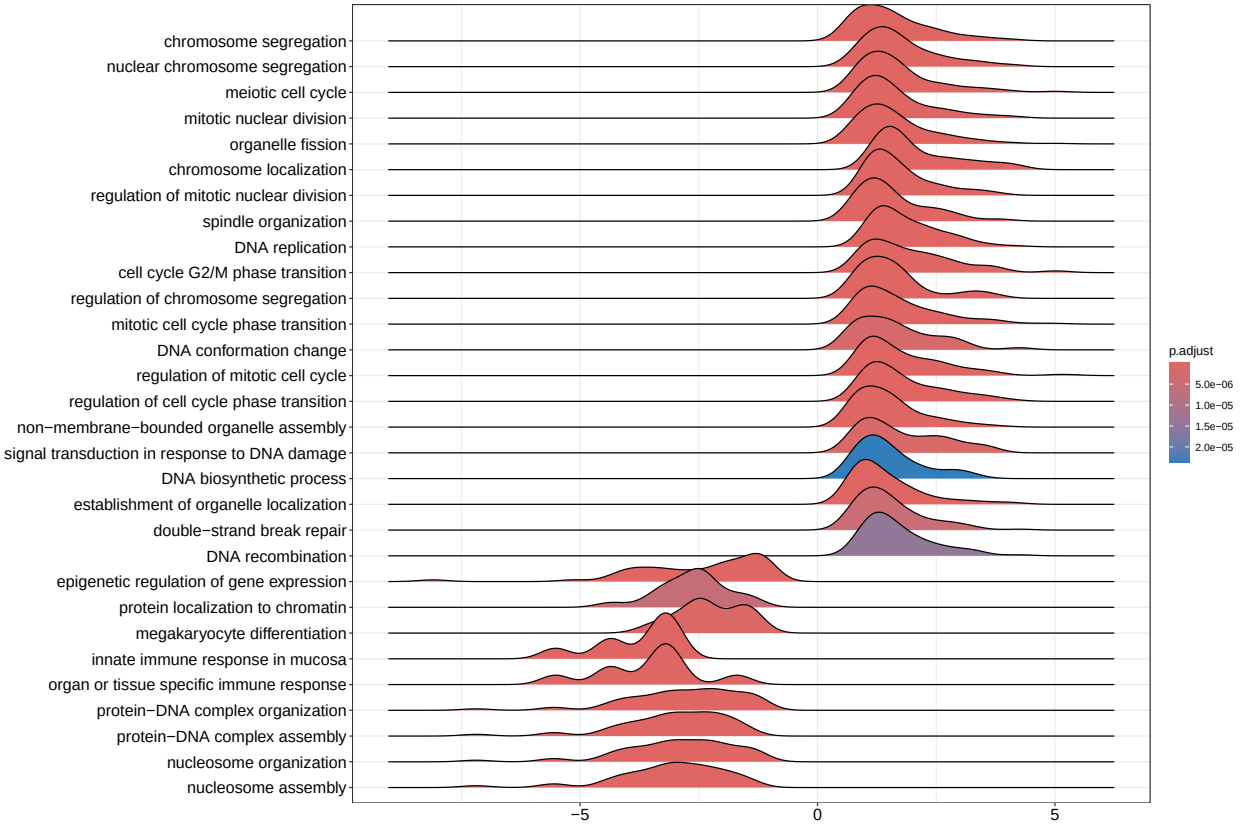

Figure 13: GSEA GO:BP summary for the dose effect.

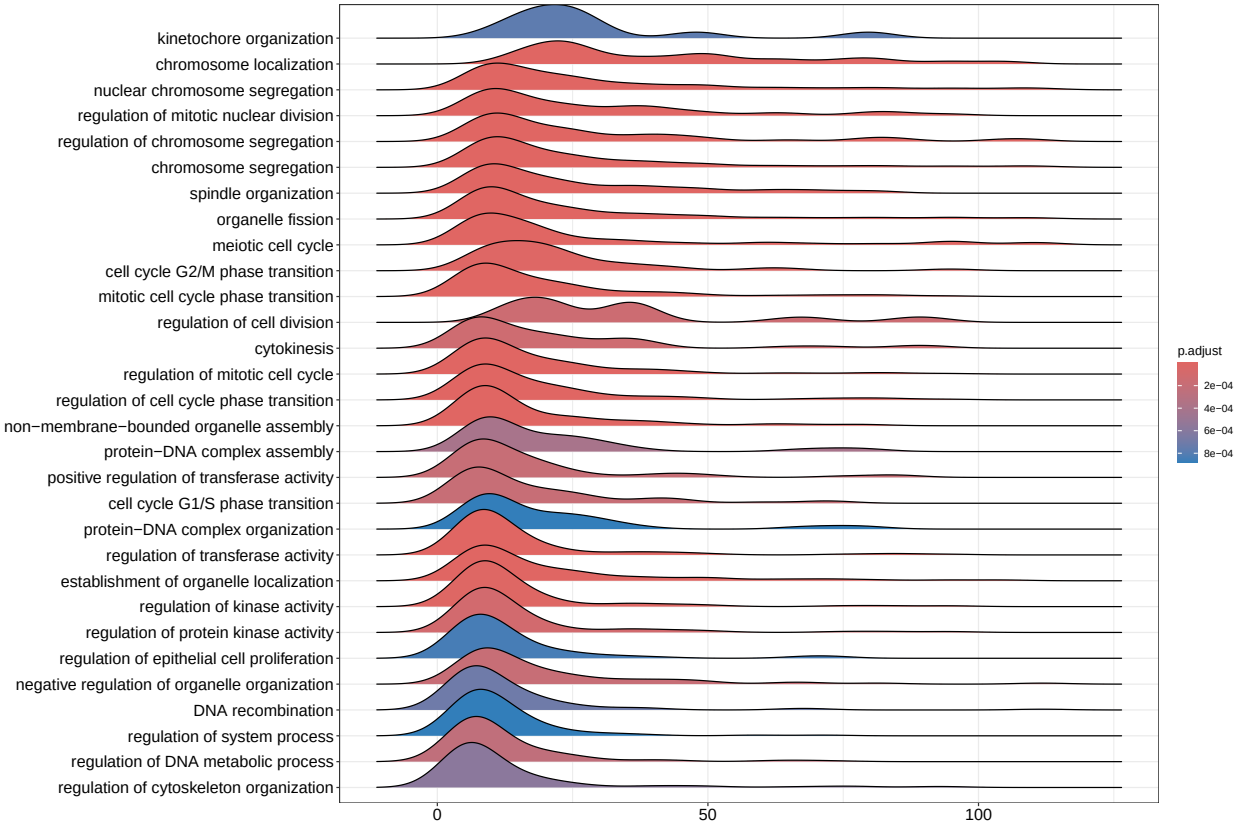

Figure 14: GSEA GO:BP summary for the interaction effect.

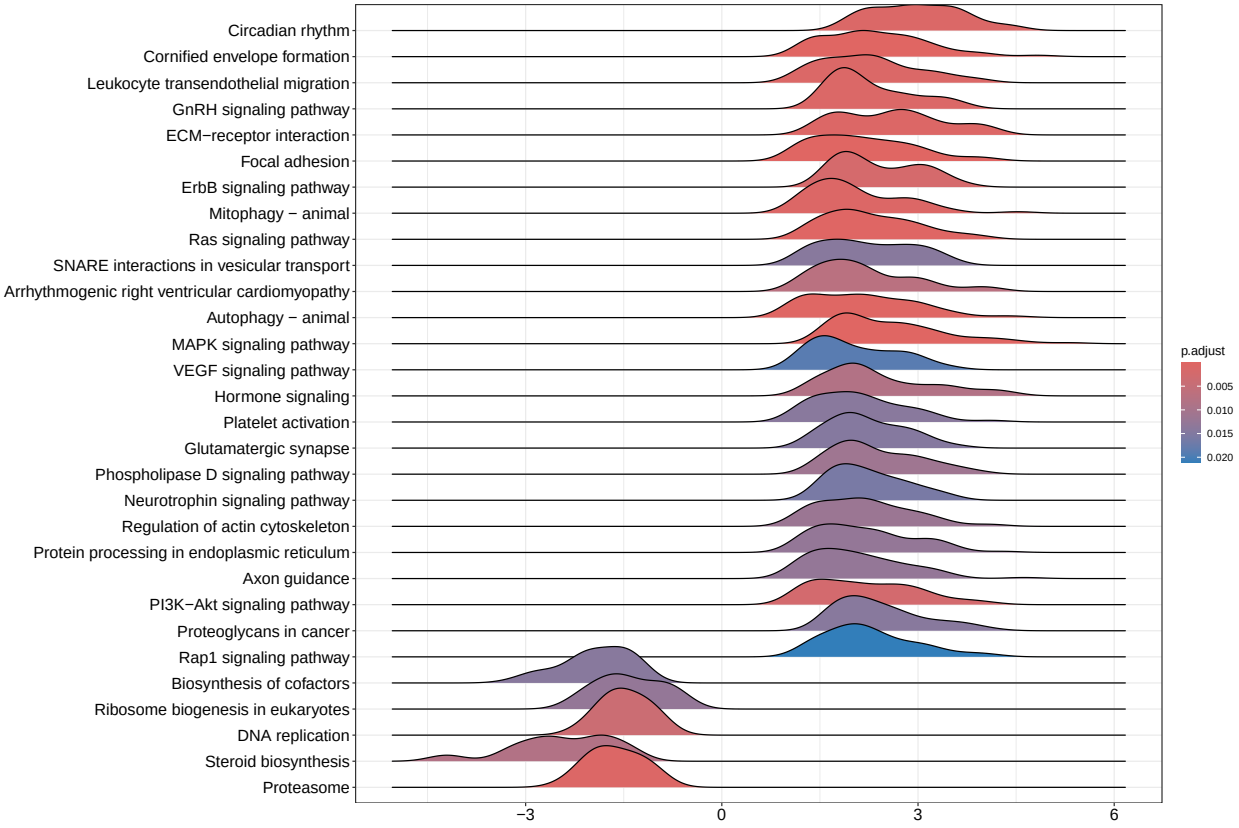

Figure 15: GSEA KEGG summary for the time effect.

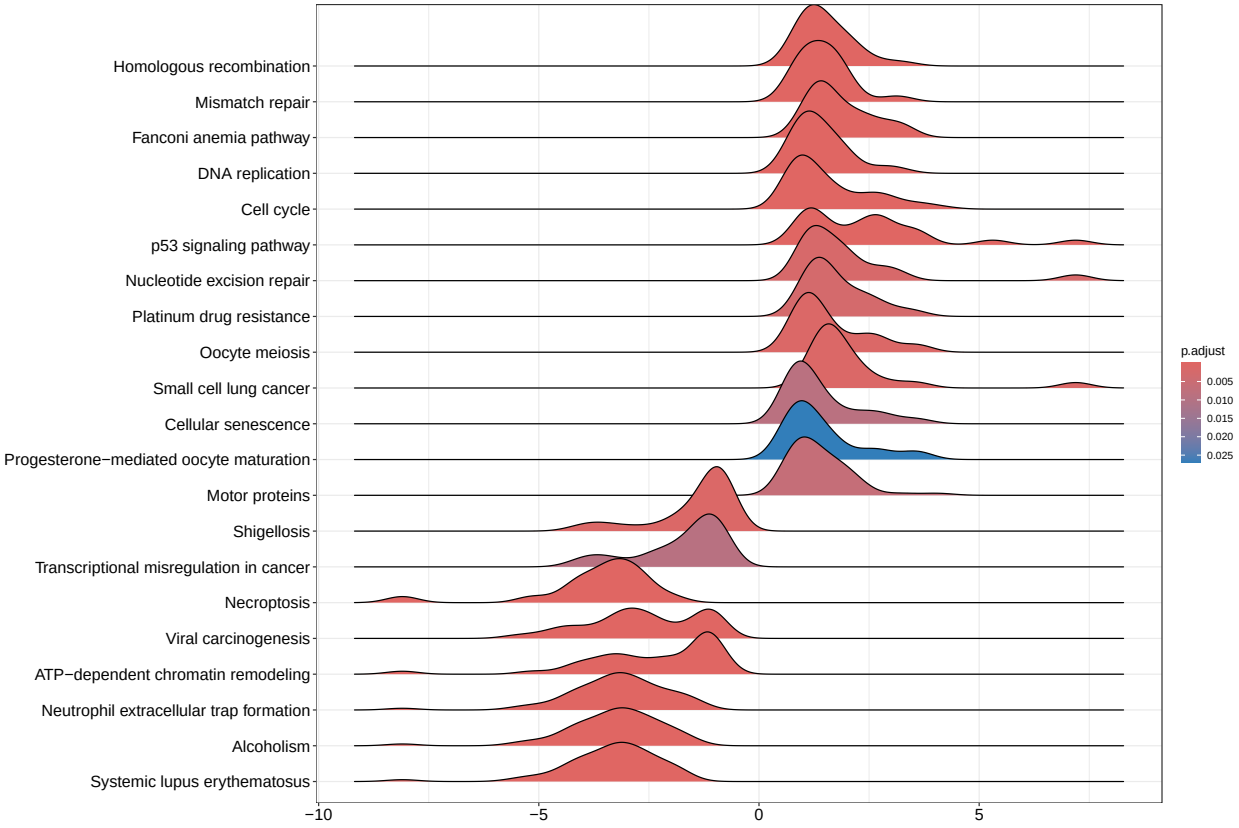

Figure 16: GSEA KEGG summary for the dose effect.

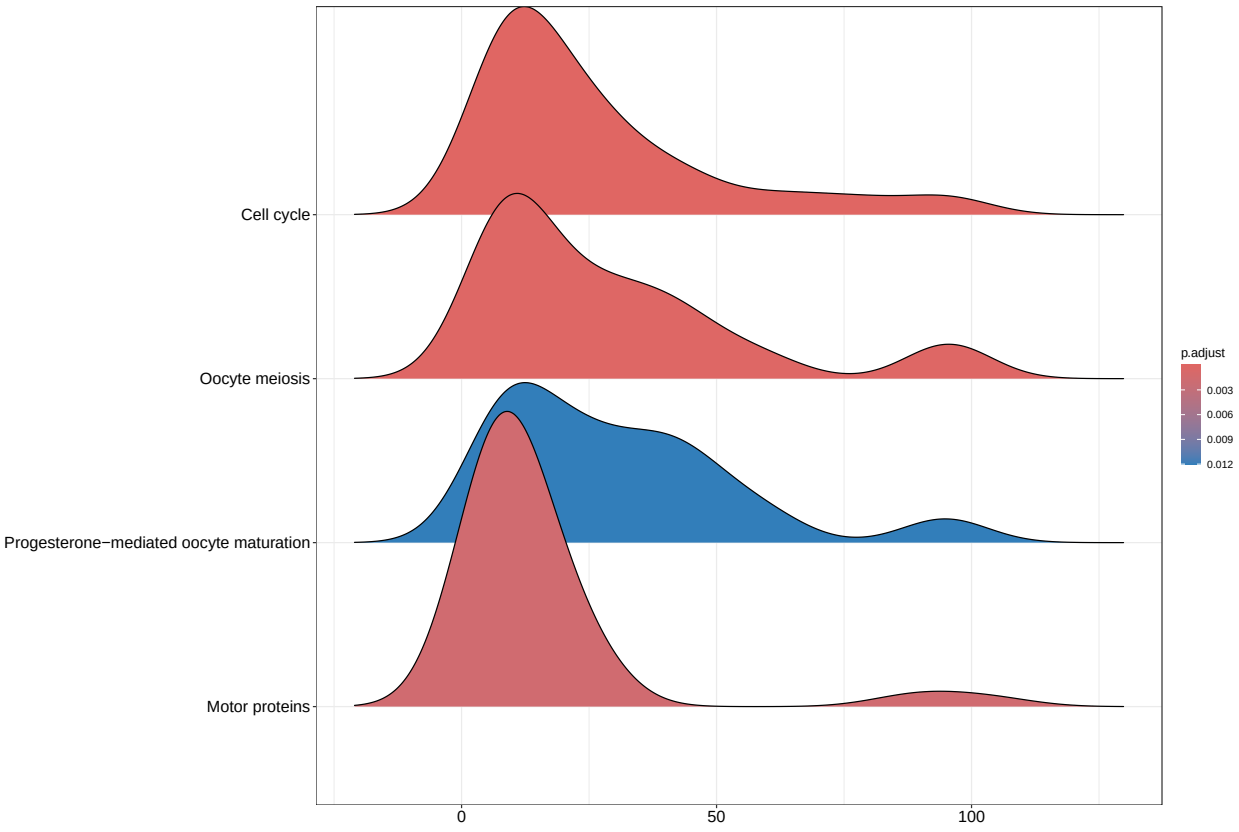

Figure 17: GSEA KEGG summary for the interaction effect.

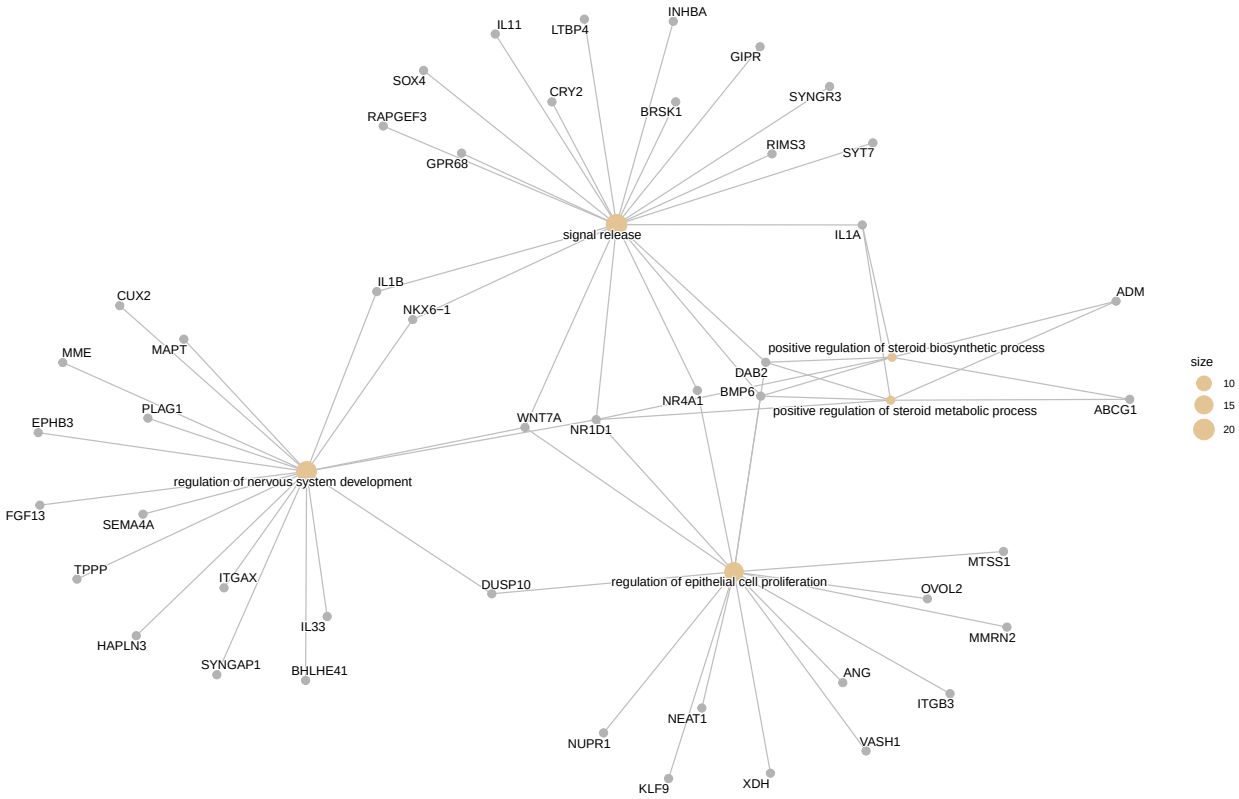

Figure 18: ORA GO:BP for up-regulated genes in the time effect.

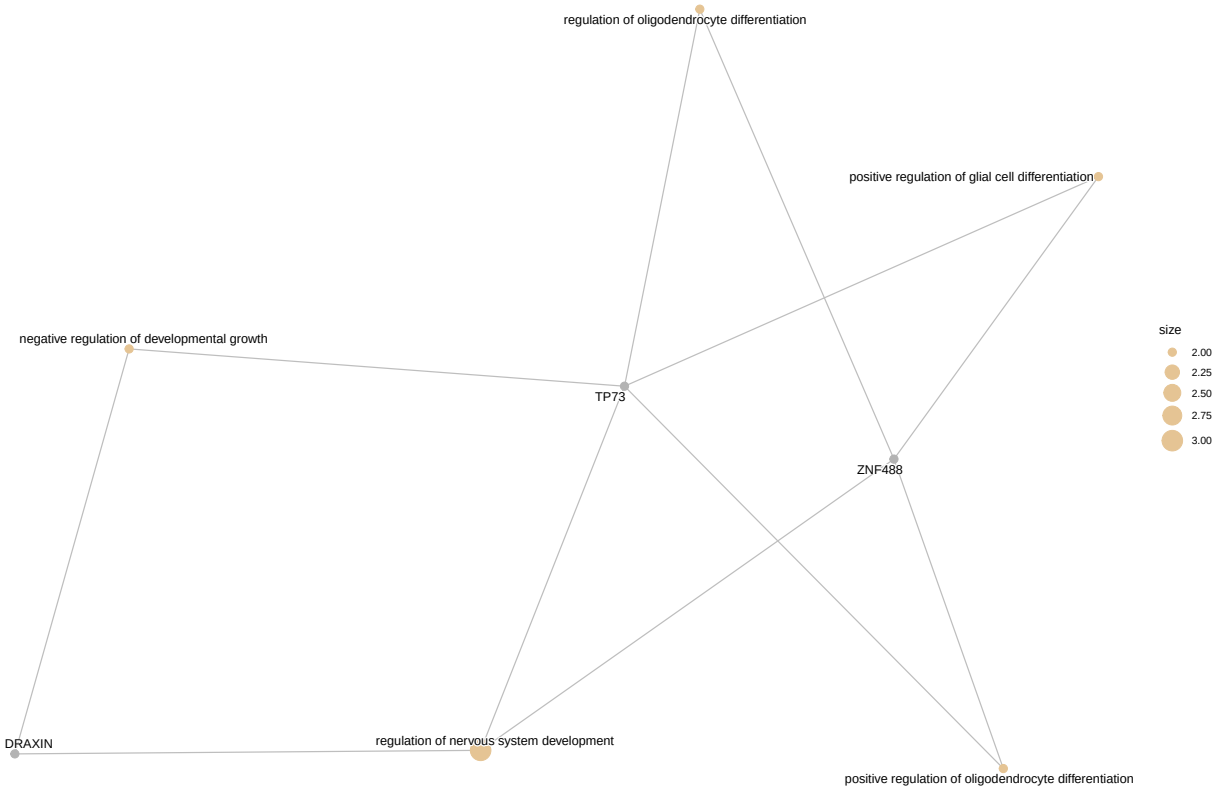

Figure 19: ORA GO:BP for up-regulated genes in the dose effect.

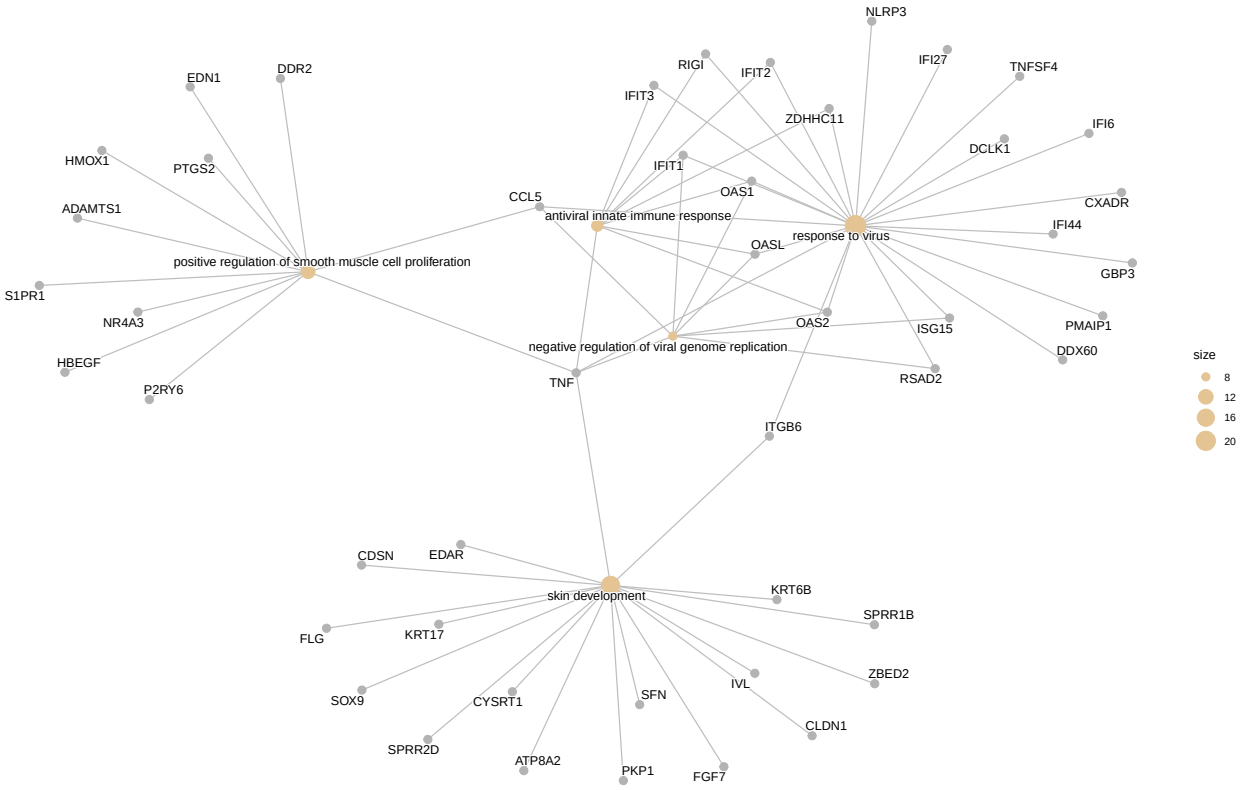

Figure 20: ORA GO:BP for up-regulated genes in the interaction effect.

Over-Representation Analysis (ORA)

- **Figures 18-20:** `GO_enrichment_[effect]upGenes_genes.pdf`: Summary plots for enriched GO:BP terms in the up-regulated gene set.

- **Figures 21-23:** `GO_enrichment_[effect]downGenes_genes.pdf`: Similar plots for the down-regulated gene set.

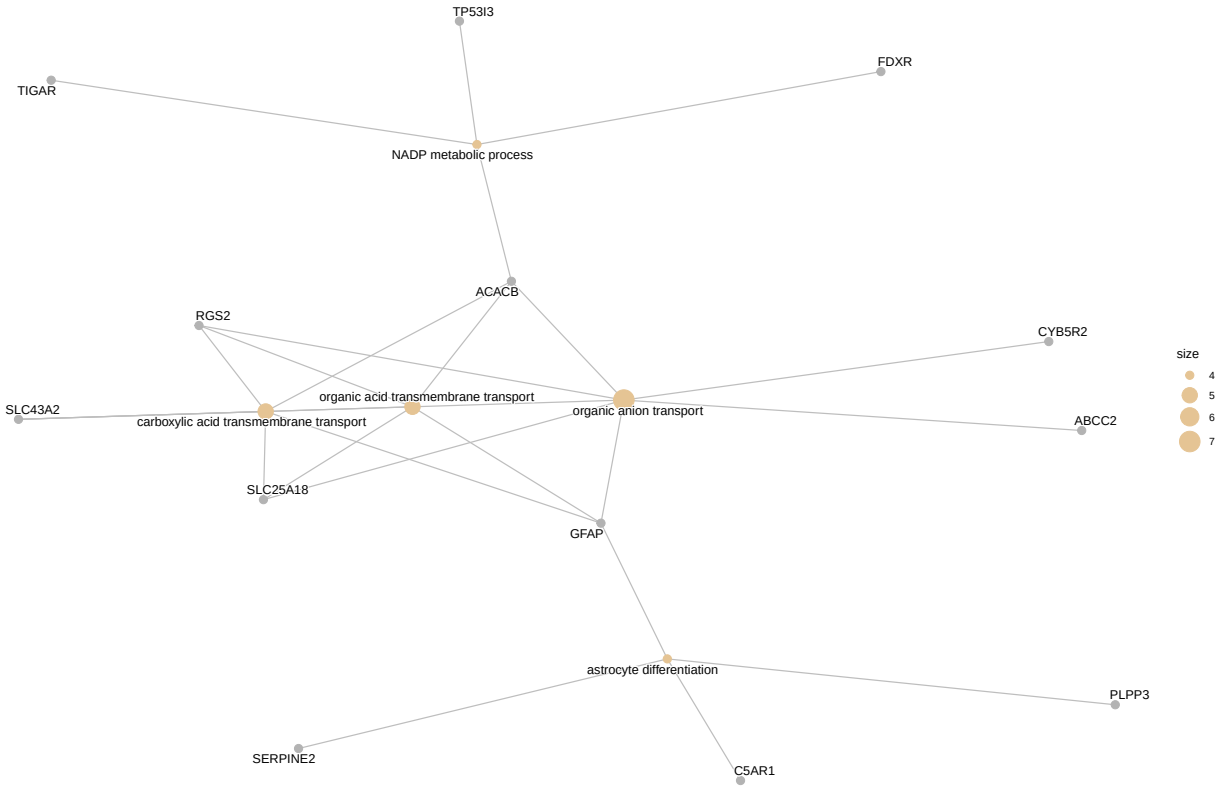

Figure 21: ORA GO:BP for down-regulated genes in the time effect.

- **Tables**:
  - `GO_enrichment_[effect]upGenes_genes.tab`: Table of enriched GO:BP terms for up-regulated genes.
  - `GO_enrichment_[effect]downGenes_genes.tab`: Table of enriched GO:BP terms for down-regulated genes.

2.2 Single-Factor (Pairwise) Analysis

This analysis performs standard pairwise comparisons between conditions as defined in `comparisons.tsv`. For each comparison (e.g., `2Gy24h_vs_NIR`), a set of output files is generated in a dedicated subdirectory within `single_factor_analysis`. The filenames follow the patterns described below.

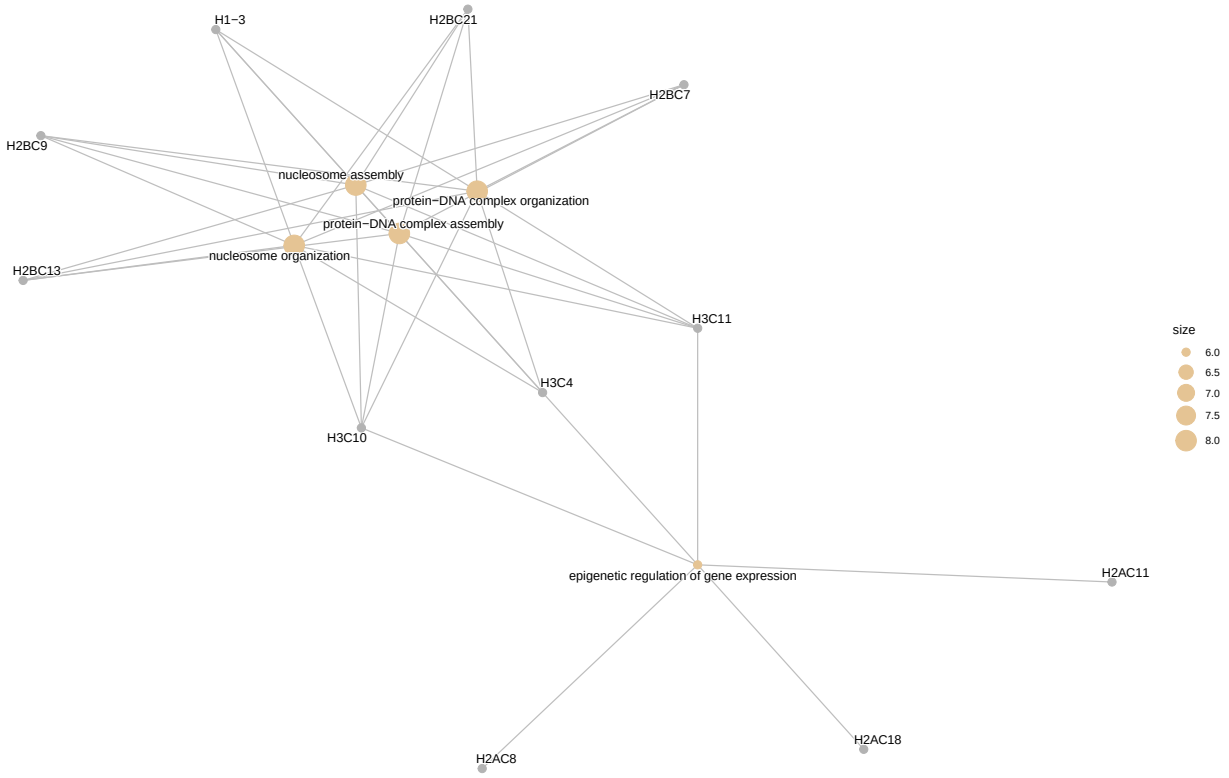

Figure 22: ORA GO:BP for down-regulated genes in the dose effect.

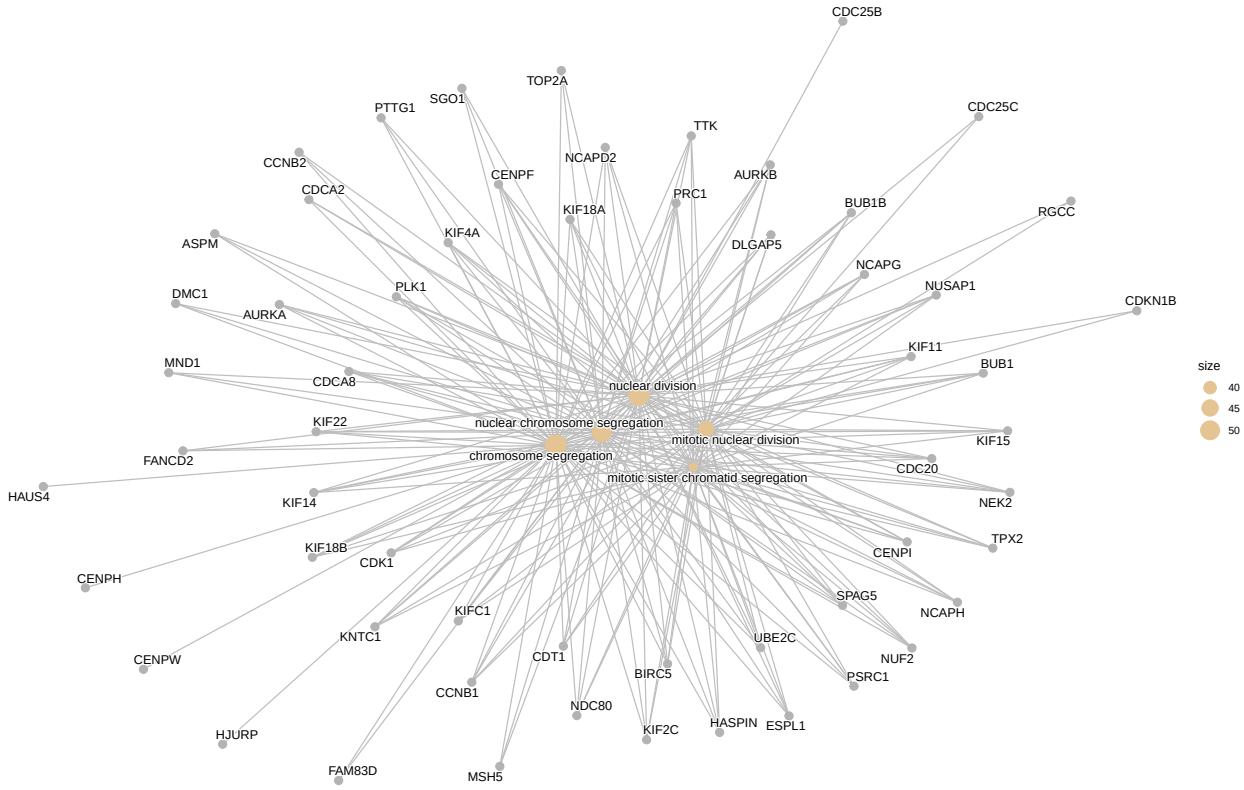

Figure 23: ORA GO:BP for down-regulated genes in the interaction effect.

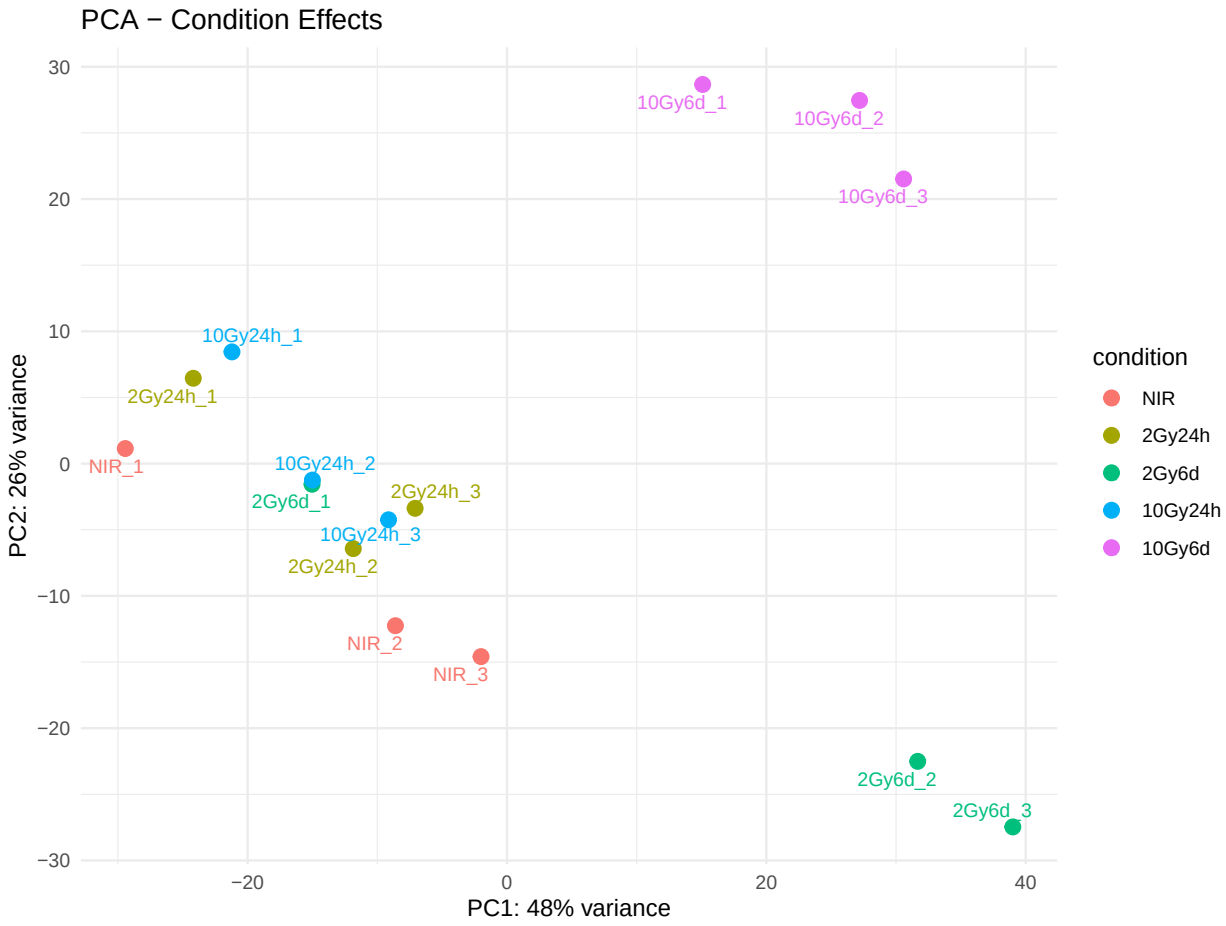

Figure 24: PCA plot of samples based on experimental factors.

- **Figure 24: PCA_condition_effects.pdf**: A PCA plot highlighting the effect of different conditions.
- **Figure 25: Heatmap_top5000_variable_genes.pdf**: A heatmap of the top 5000 most variable genes across all samples.

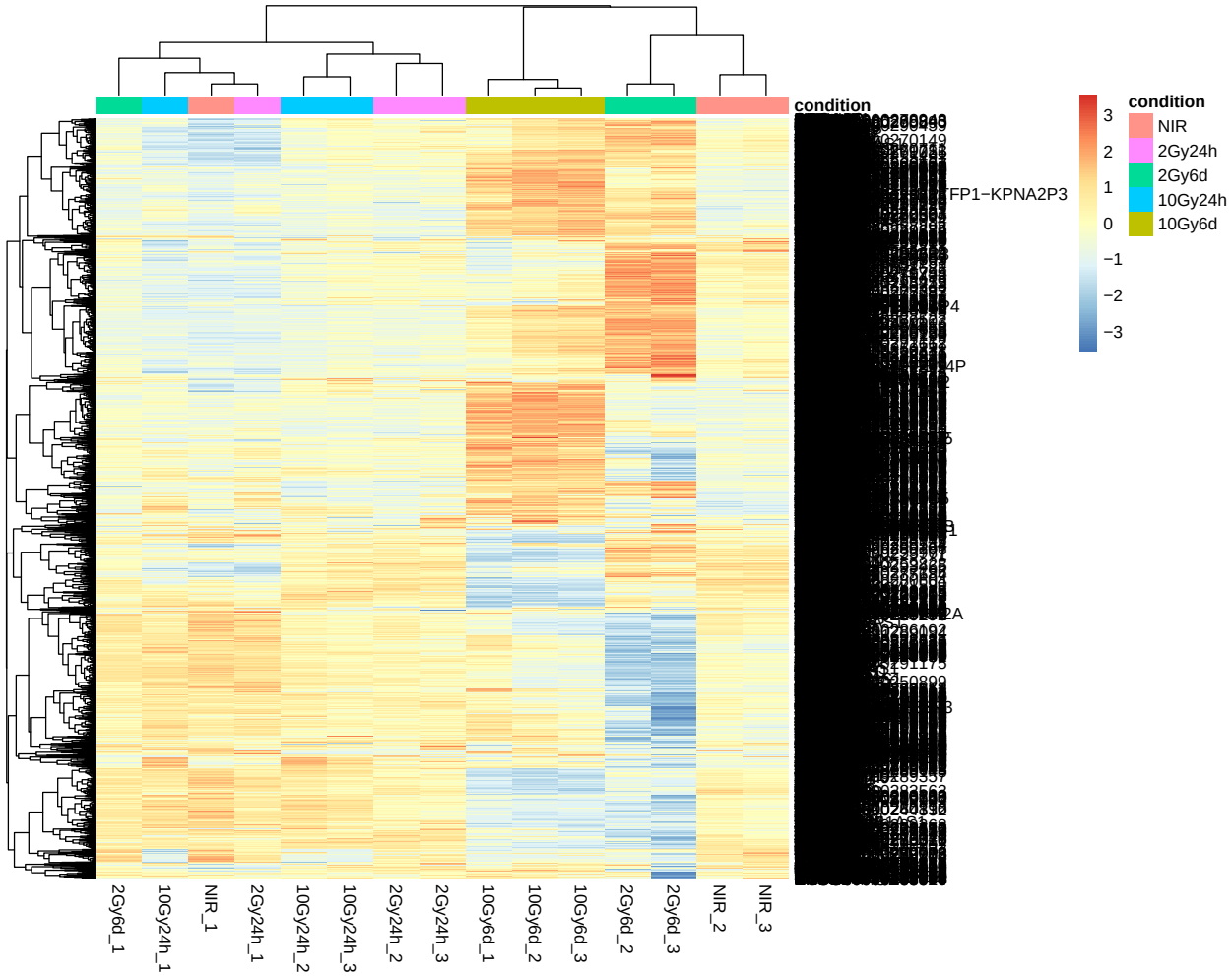

Figure 25: Heatmap of the top 5000 most variable genes.

Differential Expression Outputs (per comparison)

- `[comparison]_DESeq2_results.csv`: The complete DESeq2 results table.
- `[comparison]_volcano_plot.pdf`: A volcano plot highlighting significant genes.
- `[comparison]_MA_plot.pdf`: An MA plot for visualizing differential expression.

As their are too many comparisons, plots are not included in this report. Please see the accompanying zipped folder for all plots.

Enrichment Analysis Outputs (per comparison)

The same set of GSEA and ORA files as described in the **Multi-Factor Analysis** section are also generated for each pairwise comparison, using the comparison name (e.g., `2Gy24h_vs_NIR`) as the `[comparison]`

placeholder in the filenames. As their are too many comparisons, plots are not included in this report. Please see the accompanying zipped folder for all plots.

END