AFCA analysis

1. **Overview of the Analysis Pipeline**
2. Data Source

The AnnData object (adata\_body\_filtered\_2.h5ad) containing single-cell RNA-seq data was read into R using the zellkonverter (readH5AD) and Seurat (as.Seurat) packages.

Each cell in this dataset is annotated with metadata, including:

* age (5, 30, or 70 days)
* batch
* sex
* afca\_annotation (cell type labels, e.g. “enterocyte of posterior adult midgut epithelium,” “intestinal stem cell,” etc.)

1. Genes of interest

The genes that are of particular interest to the analysis are the following: "Su(var)205","Su(var)3-9","G9a", "HP1b", "HP1c", "HP4", "HP5", "HP6", "ADD1", "Su(var)2-HP2", "Su(var)3-7", "Lam", "LamC", "LBR", "Kdm4A", "Kdm4B", "His2Av", "His3.3A", "His3.3B"

1. Pre-processing data

Initial preprocessing followed standard steps from the AFCA paper, including quality control measures to filter out doublets, remove ambient RNA contamination, exclude cells with high mitochondrial content, and retain genes expressed in a minimum number of cells. Subsequently, additional preprocessing steps were introduced: a unique sample identifier was generated for each individual fly to serve as a latent variable in the downstream MAST differential gene expression analysis; cell types with fewer than 100 cells per unique age group were excluded; and genes were filtered to include only those expressed in at least three cells per cell type, except for specific genes of interest which were retained irrespective of expression thresholds.

1. Differential Gene Expression Analysis (DGEA)

The analysis script loops over each unique cell type in afca\_annotation. Within a given cell type, cells are subsetted so that only that cell type is considered in the comparison. DGEA is performed using Seurat’s FindMarkers(...) function with the MAST test. The analysis compares 5-day-old samples (the “young group”) vs. each of the other ages found in that cell type subset. When running the MAST-based DGE analysis, batch and sex are included as latent variables (covariates) to account for potential confounding effects.

1. Combining and Correcting p-values

After each comparison (e.g., 5 vs 30, 5 vs 70, etc.), the resulting marker table is appended to a CSV file. An additional false-discovery-rate (FDR) with BH correction (p.adjust(..., method = "BH")) is performed within each cell type to yield an adjusted p-value column . The final consolidated table is saved as full\_combined\_markers.csv.

1. **Structure of the final Results**
2. The final marker table includes (among other columns):

* **avg\_log2FC**: The log-fold-change estimate (log2 scale) for 5-day-old vs. the other age. Positive values suggest higher expression in the 5-day-old group; negative values suggest lower expression in 5-day-old.
* **cell\_type**: Which cell type the comparison is associated with.
* **comparison**: The specific comparison made (e.g., 5 vs 70).
* **gene**: The gene name.
* **adj\_p\_val\_within\_cell\_type**: Additional BH/FDR correction performed per cell type.

1. In total, six cell types appear in the final table:

* enterocyte of posterior adult midgut epithelium
* intestinal stem cell
* enterocyte of anterior adult midgut epithelium
* enterocyte-like
* adult midgut enterocyte
* adult differentiating enterocyte

1. **Key Findings**

The differential gene expression analyses primarily revealed notable age-related transcriptional changes in the enterocytes of the posterior adult midgut epithelium. Among the genes of interest, significant changes were detected in comparisons between young (5-day-old) and older flies (30, 50, and 70 days), with genes such as LamC, His3.3A, His3.3B, His2Av, ADD1, Su(var)205, HP1 variants, and Su(var)3-9 showing modest yet consistent alterations in expression. In contrast, all other cell types analysed showed negligible or no significant expression changes for the selected genes.

1. **Overall Results**

The results section summarizes the differential gene expression (DGE) analyses conducted for each cell type, comparing gene expression between young (5-day-old) and older flies (30, 50, and 70 days). For each cell type and age comparison, the following metrics are assessed:

* **Number of Significant GOIs**: The number of genes of interest which were identified as being significantly expressed
* **Significant Genes & LogFC**: The specific significant genes along with their log-fold-change (logFC) values
* **Significant Percentage**: The percentage of tested genes meeting the adjusted p-value threshold of 0.1 out of the total number of genes analysed
* **Total Genes**: Total number of genes tested
* **Total Cells**: Total number of cells analysed
* **Sample Differentials**: The differences in the number of samples between the age groups
* **Cell Differentials**: The differences in cell counts between age groups (cell differentials).

1. Enterocyte of Posterior Adult Midgut Epithelium

*Comparison: 5 vs 30 days*

* **Number of Significant Genes of Interest:** 8
* **Significant Genes and logFC**:

LamC: -0.50

His3.3A: 0.42

ADD1: -1.47

Su(var)3-9: -0.47

HP6: -0.73

HP1b: -0.77

HP1c: -0.81

His2Av: -0.41

* **Significance Percentage**: 44.98%
* **Total Genes Tested**: 8,759
* **Total Cells**: 570
* **Sample Differentials**: 11 vs 10
* **Cell Differentials**: 177 vs 393

*Comparison: 5 vs 50 days*

* **Number of Significant Genes of Interest**: 3
* **Significant Genes and logFC**:

His3.3B: -0.45

LamC: -0.60

His2Av: -0.53

* **Significance Percentage**: 8.78%
* **Total Genes Tested**: 8,721
* **Total Cells**: 318
* **Sample Differentials**: 11 vs 10
* **Cell Differentials**: 177 vs 141

*Comparison: 5 vs 70 days*

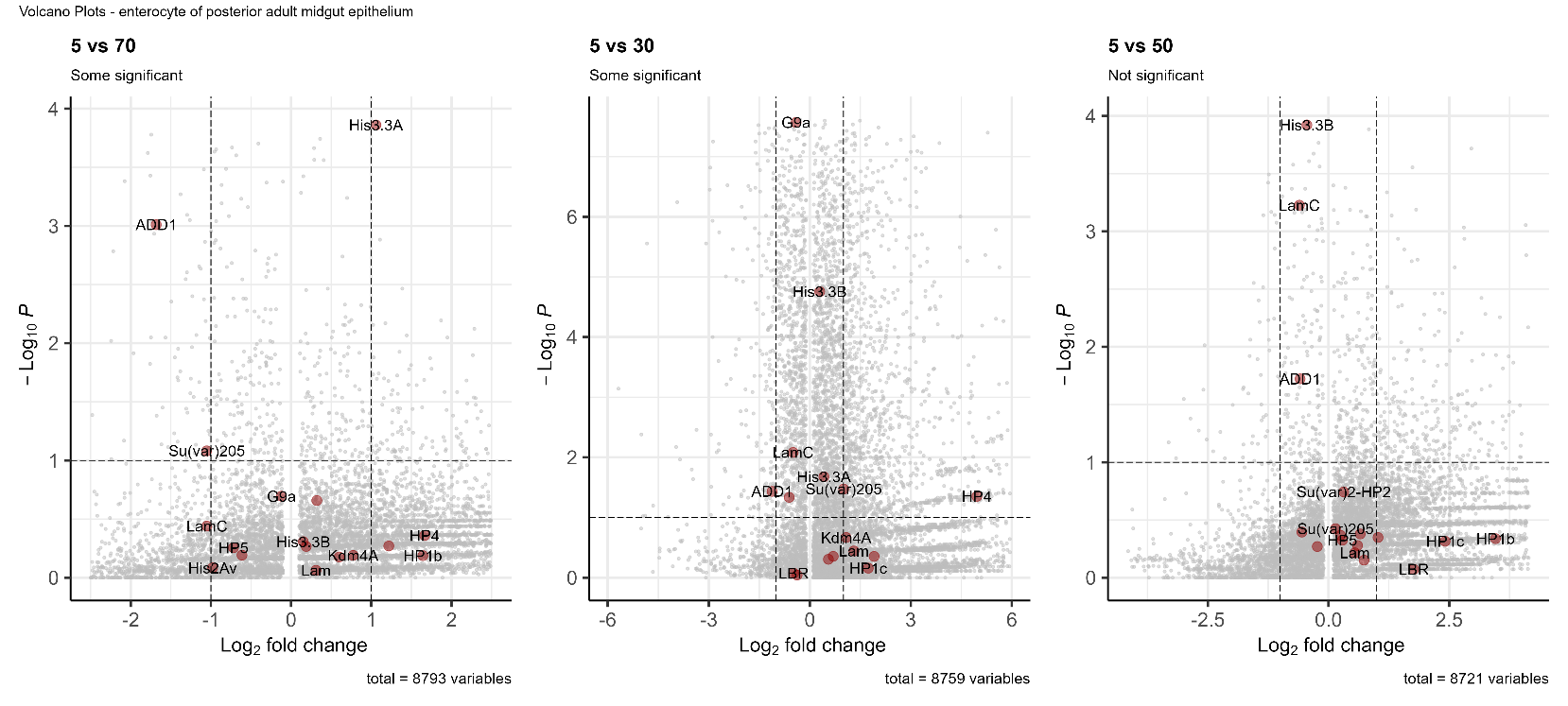
* **Number of Significant Genes of Interest**: 2
* **Significant Genes and logFC**:

His3.3A: 1.06

ADD1: -1.68

Su(var)205: -1.05

* **Significance Percentage**: 6.43%
* **Total Genes Tested**: 8,793
* **Total Cells**: 217
* **Sample Differentials**: 11 vs 9
* **Cell Differentials**: 177 vs 40



1. Intestinal Stem Cell

*Comparison: 5 vs 30 days*

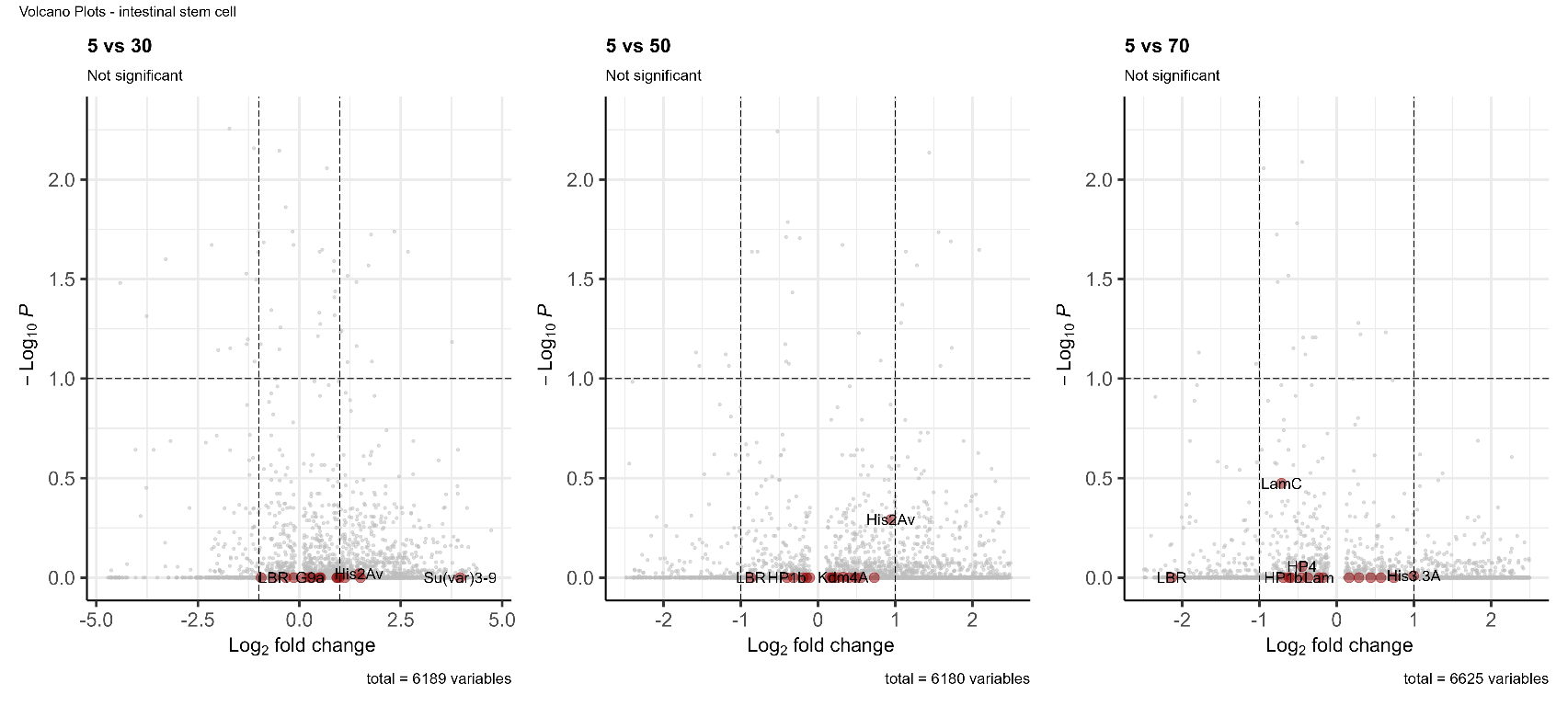
* **Number of Significant Genes of Interest:** 0
* **Significance Percentage**: 1.21%
* **Total Genes Tested**: 6,186
* **Total Cells**: 881
* **Sample Differentials**: 10 vs12
* **Cell Differentials**: 254 vs 627

*Comparison: 5 vs 50 days*

* **Number of Significant Genes of Interest:** 0
* **Significance Percentage**: 0.81%
* **Total Genes Tested**: 6,180
* **Total Cells**: 692
* **Sample Differentials**: 10 vs12
* **Cell Differentials**: 254 vs 438

*Comparison: 5 vs 70 days*

* **Number of Significant Genes of Interest:** 0
* **Significance Percentage**: 0.38%
* **Total Genes Tested**: 6,625
* **Total Cells**: 1067
* **Sample Differentials**: 10 vs12
* **Cell Differentials**: 254 vs 813



1. Enterocyte of Anterior Adult Midgut Epithelium

*Comparison: 5 vs 30 days*

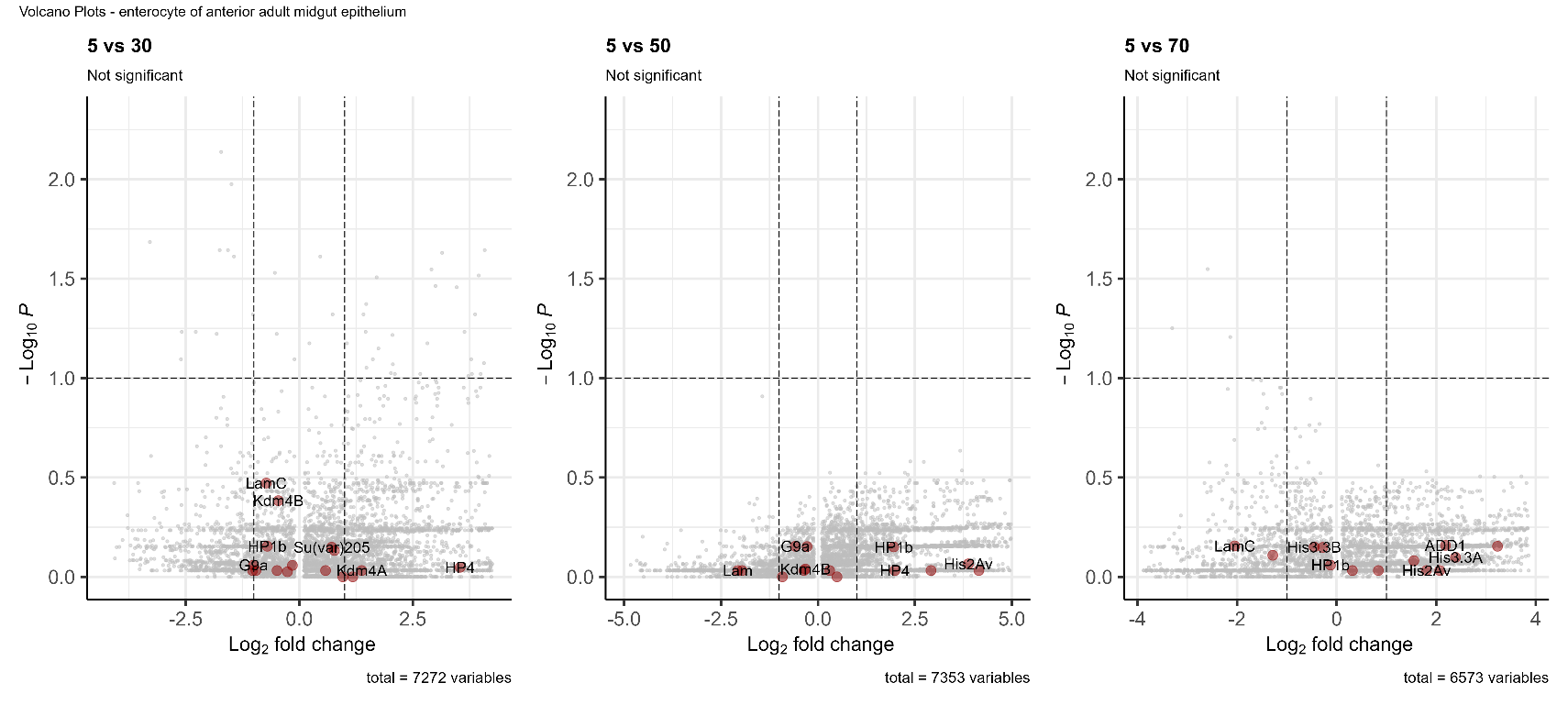
* **Number of Significant Genes of Interest:** 0
* **Significance Percentage**: 0.84%
* **Total Genes Tested**: 7,272
* **Total Cells**: 135
* **Sample Differentials**: 5 vs 7
* **Cell Differentials**: 23 vs 112

*Comparison: 5 vs 50 days*

* **Number of Significant Genes of Interest:** 0
* **Significance Percentage**: 0.00%
* **Total Genes Tested**: 7,353
* **Total Cells**: 61
* **Sample Differentials**: 5 vs 10
* **Cell Differentials**: 23 vs 28

*Comparison: 5 vs 70 days*

* **Number of Significant Genes of Interest:** 0
* **Significance Percentage**: 0.05%
* **Total Genes Tested**: 6,573
* **Total Cells**: 32
* **Sample Differentials**: 5 vs 4
* **Cell Differentials**: 23 vs 9



1. Enterocyte-like

*Comparison: 5 vs 30 days*

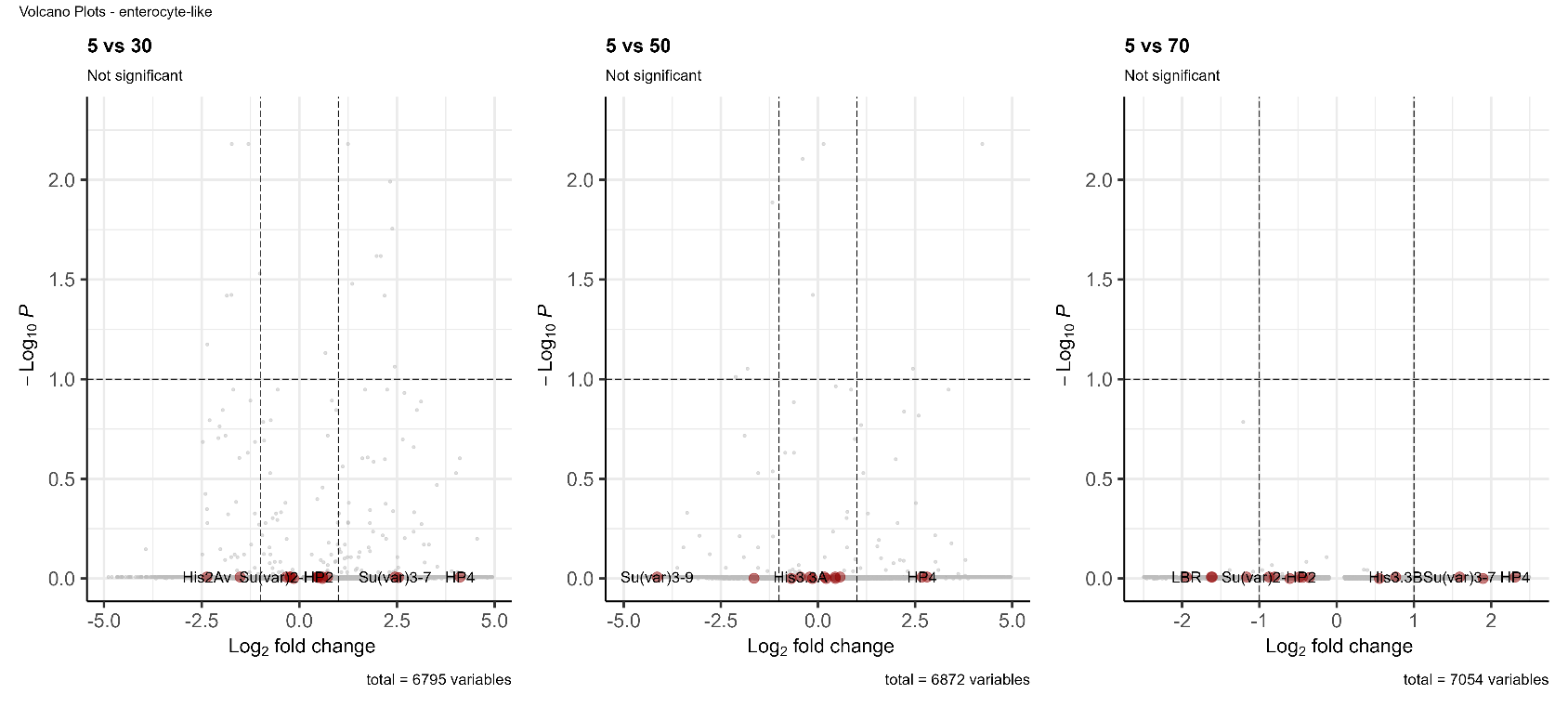
* **Number of Significant Genes of Interest:** 0
* **Significance Percentage**: 0.28%
* **Total Genes Tested**: 6,795
* **Total Cells**: 172
* **Sample Differentials**: 5 vs 8
* **Cell Differentials**: 39 vs 133

*Comparison: 5 vs 50 days*

* **Number of Significant Genes of Interest:** 0
* **Significance Percentage**: 0.12%
* **Total Genes Tested**: 6,872
* **Total Cells**: 88
* **Sample Differentials**: 5 vs 7
* **Cell Differentials**: 39 vs 49

*Comparison: 5 vs 70 days*

* **Number of Significant Genes of Interest:** 0
* **Significance Percentage**: 0.00%
* **Total Genes Tested**: 7,054
* **Total Cells**: 77
* **Sample Differentials**: 5 vs 9
* **Cell Differentials**: 39 vs 38



1. Adult Midgut Enterocyte

*Comparison: 5 vs 30 days*

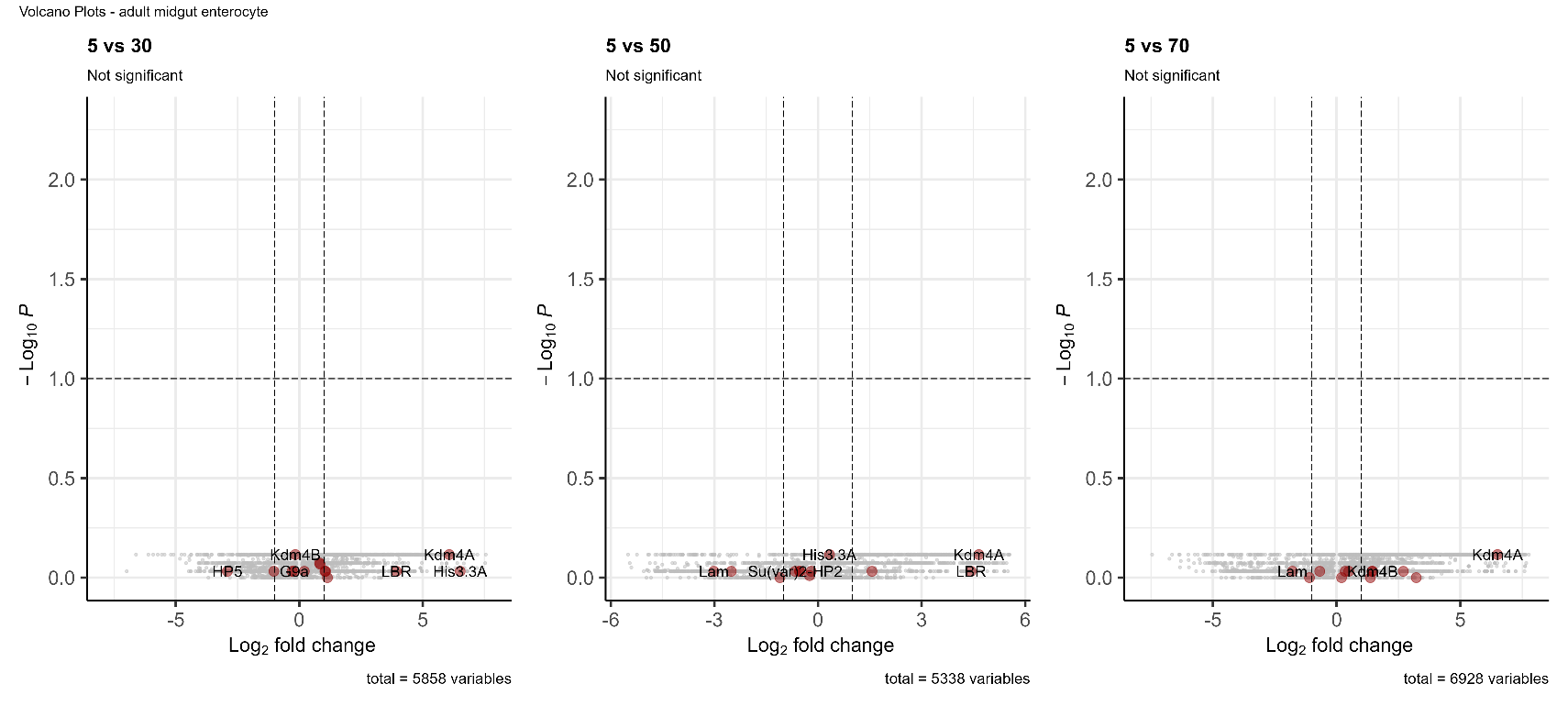
* **Number of Significant Genes of Interest:** 0
* **Significance Percentage**: 0.00%
* **Total Genes Tested**: 5,858
* **Total Cells**: 42
* **Sample Differentials**: 5 vs 10
* **Cell Differentials**: 10 vs 32

*Comparison: 5 vs 50 days*

* **Number of Significant Genes of Interest:** 0
* **Significance Percentage**: 0.00%
* **Total Genes Tested**: 5,338
* **Total Cells**: 22
* **Sample Differentials**: 5 vs 6
* **Cell Differentials**: 10 vs 12

*Comparison: 5 vs 70 days*

* **Number of Significant Genes of Interest:** 0
* **Significance Percentage**: 0.00%
* **Total Genes Tested**: 6,928
* **Total Cells**: 53
* **Sample Differentials**: 5 vs 9
* **Cell Differentials**: 10 vs 43



1. Adult Differentiating Enterocyte

*Comparison: 5 vs 30 days*

* **Number of Significant Genes of Interest:** 0
* **Significance Percentage**: 0.00%
* **Total Genes Tested**: 7,417
* **Total Cells**: 41
* **Sample Differentials**: 6 vs 5
* **Cell Differentials**: 34 vs 7

*Comparison: 5 vs 70 days*

* **Number of Significant Genes of Interest:** 0
* **Significance Percentage**: 0.04%
* **Total Genes Tested**: 8,114
* **Total Cells**: 65
* **Sample Differentials**: 6 vs 8
* **Cell Differentials**: 34 vs 31

